Articulatory timing in Hindi CV sequences

Shihao Du[1], Indranil Dutta[2], Adamantios I. Gafos[1]

[1]*Universtität Potsdam, Potsdam, Germany*
[2]*Jadavpur University, Kolkata, India*

shihao.du@uni-potsdam.de, indranildutta.lnl@jadavpuruniversity.in, gafos@uni-potsdam.de

Abstract

This short report presents some preliminary results from electromagnetic articulography (EMA) recordings of Hindi Consonant-Vowel (CV) sequences. We specifically asked if and how articulatory timing in CV, quantified by the interval from V-target to C-offset, is modulated by consonant phonation, consonant place of articulation and vowel quality. Results show that vowel height and frontness and C place of articulation exert significant effects on CV timing, whereas C phonation (voicing and aspiration) has no significant effect on the interval we chose to quantify CV timing here. Potential explanations of the disparity between vowel-related and consonant-related effects are suggested.

Keywords*: speech production, Consonant-Vowel articulatory timing, Hindi*

1. Introduction

In stop-vowel sequences of Hindi, we compared Consonant-Vowel (CV) timing for different CVs where the consonant was one of /b/, /p/, /bʱ/, /d/, /t/, /dʱ/, /tʰ/ and the vowel was one of /iː, ɪ, uː, ʊ, eː, e, oː, o, aː/. Not much is known about how phonation and place of articulation of the consonant affects articulatory timing in CV sequences. Likewise, much is yet to be documented on how properties of the vowel affect its timing with its preceding consonant. Early studies such as Ostry et al. (1983) and Löfqvist and Gracco (1997) on English did report on a possible consonant voicing effect (but whether the effect was due to voicing or aspiration could not be determined given the language) and effects related to vowel quality on the kinematics of the consonantal gestures, but how these features jointly affect CV timing remains largely unknown. For consonant sequences, Bombien and Hoole (2013) show that the temporal distance between the oral constrictions in German stop-liquid sequences (e.g., [gl] versus [kl]) varies systematically as a function of stop voicing. The former shows about 21 ± 2 ms more overlap than the latter. Whereas these studies focus on effects of voicing on the timing of the C oral gestures in CC sequences, our study focuses on the timing between the oral gestures of C and V in CV sequences. Our motivation is the same as that in Bombien and Hoole (2013) who note that "the coordination of supra-laryngeal articulations with respect to laryngeal specification is an area of speech production research which so far has received only limited attention and is far from being understood" (Bombien & Hoole 2013: p. 539). Hindi offers an ideal case study in this respect. In the CV context, where C is a stop, consonants exhibit a four-way contrast (in alveolar, retroflex, and velar stops; labial stops show primarily a three-way contrast, as /pʰ/ is realized increasingly as [f]), with the full suite of voiced unaspirated, voiced aspirated (also known as breathy), voiceless unaspirated, and voiceless aspirated stops.

We give an example of how the lack of knowledge in this domain has hindered theory development and evaluation. Browman and Goldstein (1988) first observed that when adding a consonant to the start of a syllable, from [pa] to [spa], the temporal organization of the whole changes such that [p], [a] timing in [spa] is different from that in [pa]. The gestures of [p], [a] seem to slide closer to one another in [spa] than in [pa]. It was hypothesized that the vowel onset in such sequences is synchronous with the center of the prevocalic consonantism (be it a single [p] or an [sp]) and specifically with the midpoint of the consonantal closure intervals of all consonants (Browman & Goldstein 1988: p. 150; see also Honorof & Browman 1995, Figure 1, p. 552). As the American English stop in an [s]-stop cluster before a vowel is not aspirated (but the lone voiceless stop is), such a comparison implies a potential confound (see also Katz 2012) due to the phonation (presence versus absence of the aspiration gesture) which may independently affect vowel timing. Perhaps a more appropriate comparison would be to consider the timing of the vowel in relation to the prevocalic consonantism in [s]-stop-vowel versus single voiced stop-vowel sequences, because in both the stop is not aspirated; this is still imperfect, however, because of the presence of the /s/ which makes it impossible to decide whether any differences are exclusively due to the phonation of the stop (because /s/ also implicates tongue movement just like the vowel following the stop, it may be that whatever requirements /s/ imposes on tongue body control, these have an influence on the timing of the subsequent vowel which also implicates the tongue body). In any case, the facts are simply not known here. An ever more appropriate comparison would be to compare the timing of the vowel in relation to the prevocalic C in single, not aspirated stop-vowel sequences versus single aspirated vowel sequences but the former are not available in English.

Consider furthermore the fact that typically segments are ensembles of gestures. In defining the notion of inter-segmental coordination, which gestures from the segments so coordinated are to be related to one another? Is the glottal opening gesture of a [t] or the velic lowering gesture of an [m] eligible for entering in a coordination relation with other segments? In Gafos (2002), inter-segmental coordination was defined by making reference to notion of 'head' of a segment: "Two segments S1, S2 are coordinated with some coordination relation λ, /S1 λ S2/, if the head gestures of these segments are coordinated as in λ" (Gafos 2002: 284), where coordination was operationalized by specifying that one landmark from the first and another from the second gesture are aligned in time (synchronized). The head gesture of a segment is the gesture of the oral task variable of that segment (Browman & Goldstein 1986; Saltzman 1986). This can be motivated on a number of reasons. Theoretical precedent in feature-geometric representations pointed to the key role of the oral gesture of a segment (Sagey 1986; Halle 1995). Kingston's (1985) work on "articulatory binding" proceeded from the fact that contrastive laryngeal articulations tend to be bound to the release of oral stops. Steriade (1993; 1994) formulated a theory of representations which directly encoded so-called "anchor" positions of oral closure and release to explain facts about possible segments with contrastive laryngeal and velic specifications. It was on the backdrop of these proposals that oral gestures were assumed to drive segment-to-segment coordination. Finally, the data Gafos (2002) aimed to account

for indicated that laryngeal or velic gestures did not enter into the phonological and morphological effects that provided the core argument for a grammar of gestural coordination in that study. Thus, identity avoidance effects were observed for adjacent segments with identical oral gestures (e.g., [d-t]) but not so for identical velic or laryngeal gestures. An [n-m] or a [t]-[k] sequence did not trigger identity avoidance effects even though these are sequences of two identical velic lowering and laryngeal gestures respectively. It was on the basis of such facts that inter-segmental coordination relations were proposed to be stated by reference to the oral gestures of the segments so coordinated, with the intra-segmental laryngeal or velic gestures following suit by maintaining their segment-internal relation to the head gesture of their segment (i.e., when the oral gestures slide apart, their corresponding velic gestures slide along with them). If inter-segmental coordination in CV sequences is not mediated by laryngeal specifications of the consonant, this implies that CV timing in Hindi should not be modulated by the phonation characteristics of the C (voiceless unaspirated, voiceless aspirated, voiced unaspirated, and voiced unaspirated stops). It is thus clear that further theory evaluation and development rely crucially on a better understanding of the facts regarding the role of consonant phonation and place of articulation on the timing of the oral gestures in CV sequences.

Recently, intervals delineated by landmarks on CV sequences have been examined in works that aim to assess the extent to which inter-segmental coordination can be expressed in terms of synchronicity relations among landmarks. For instance, Shaw and Chen (2019) demonstrated on basis of Mandarin CV sequences consisted of labial consonants (/m/ and /p/) and back rounded vowels (/ou/, /u/, /uo/) that the lag from V-target to C-offset has a mean of zero, representing close synchrony of the two landmarks. In another study along the same lines, Kramer et al. (2023) report the mean and standard deviation of four intervals (C-onset to V-onset, V-onset to C-target, C-target to V-target, V-target to C-offset) on the basis of eight word-initial CV sequences in American English and Mandarin, where the initial consonant is either /b/ or /m/ and the vowel is either low back /ɑ/ or high front /i/. Out of the four intervals examined in Kramer et al. (2023), V-target to C-offset was the one with a mean closest to zero (implying near synchronicity of the two landmarks). Similarly, Durvasula and Wang (2023) examined whether it is V-onset or V-target that is aligned to some landmark within the prevocalic consonantal gesture in five American English words (*back*, *fiber*, *make*, *much*, *people*) with a word-initial labial obstruent-vowel sequence and reported that V-target was consistently aligned with the C-offset. In the current work on Hindi, we adopt the V-target to C-offset interval to quantify CV timing and examine how consonant phonation, place of articulation, and vowel quality modulate this interval.

2. Methods

Electromagnetic articulography data were collected from 2 native male speakers of Hindi aged from 22 to 23, who reported no hearing or other health issues. The speakers produced 63 target words beginning with CV sequences where the consonant was either /b/, /p/, /bʱ/, /d/, /t/, /dʱ/, or /tʰ/ (aspirated /pʰ/ is not included because in Hindi it underwent fricativization and is realized contemporarily as [f]) and the vowel was one of / iː, ɪ, uː, ʊ, eː, e, oː, o, aː/ (the consonants and vowels were fully crossed, such that each consonant is paired with all nine vowels; i.e., 7 consonants × 9 vowels = 63 CV sequences). The target words were all embedded in the carrier phrase *Ramā* ___ *bolī* (translation: 'Ramā said ___'), in which the target appears at a phrase-medial and prosodically neutral position. Each phrase was repeated 10 times by each speaker in a random order, yielding 1260 tokens in total (63 target words × 10 repetitions × 2 speakers). The Carstens AG501 device was used to record movements of 10 sensors attached to the speech organs and head at a sampling rate of 1250 Hz. For the current study, the movements of the sensor attached to the tongue dorsum (TD) was used to identify vowel gestures, the sensor attached to the tongue tip (TT) for the gestures associated with the alveolar consonants, and the Euclidean distance between the sensors attached at the vermillion border of the upper and lower lips (UL and LL) for the bilabial consonant gestures. Articulatory gestures were parsed manually using the matlab-based software MVIEW (Tiede 2005). Temporal landmarks were identified using a 20% peak velocity threshold. Out of the elicited 1260 tokens, 77 tokens (6.11%) were eliminated because of data storage failure or failure of gestural parsing. For each of the remaining tokens, the temporal distance from consonant offset and vowel target was computed to assess landmark synchrony in CV sequences. A linear-mixed effects model was fitted to the data with the synchrony measure as the dependent variable and consonant voicing (voiced vs. voiceless), aspiration (aspirated vs. un-aspirated), place (alveolars vs. labials), vowel height (high vs. low vs. mid), vowel frontness / roundness (back / rounded vs. non-back / unrounded), and vowel length (long vs. short) as fixed effects (all sum-coded). Random intercepts for speakers and items were also included.

3. Results

We first set out to assess the extent to which pairs of landmarks drawn from the vowel and the consonant, such as the landmarks V-target and C-offset, show synchrony. **Table 1** below lists means and standard deviations for the intervals C-onset to V-onset, V-onset to C-target, C-target to V-target, V-target to C-offset as well as the inter-plateau interval (C-release to V-target) and the interval from C-opening peak velocity (PV) to V-target. It can be seen that the V-target to C-offset interval has a mean of 8.54 ms in our Hindi dataset, which is the mean closest to zero among all the tested intervals, indicating near synchrony. This result is in line with findings from other recent work (Shaw & Chen 2019; Kramer et al. 2023; Durvasula & Wang 2023) which suggests that V-target and C-offset may be (near) synchronous in CV timing.

Table 1: *Means and standard deviations of six intervals delineated by a landmark on the consonant and a landmark on the vowel in Hindi CV sequences.*

Interval	Mean (ms)	SD (ms)
C-onset to V-onset	136.13	48.69
V-onset to C-target	62.81	45.43
C-target to V-target	156.07	41.50
C-release to V-target	106.10	36.60
C-opening-PV to V-target	53.55	37.87
V-target to C-offset	8.54	44.66

We then assessed how consonant phonation, place of articulation, and vowel quality modulate the duration of this interval. **Figure 1** presents density plots of the V-target to C-offset interval as a function of the six fixed effects (consonant voicing, aspiration, place, vowel height, frontness / backness, and length). The model had an intercept of 3.14 ms, indicating that the vowel target occurs on average approximately 3 ms before the consonant offset. An anova test applied to the linear-mixed effects model revealed that consonant place, vowel height and frontness / backness had significant effects on the synchrony measure (p-value < 0.0001 for all three; F-value = 24.50, 24.83, and 17.01 respectively), whereas the effects of

consonant voicing, aspiration, and vowel length did not reach significance (*p*-value = 0.17, 0.56, and 0.20 respectively; *F*-value = 1.86, 0.33, and 1.65 respectively). For the significant effects post-hoc pairwise comparisons were implemented using the R package *emmeans* (Lenth et al. 2023). For consonant place, the comparisons indicate that the two landmarks are 12.9 ms farther apart when C place is alveolar versus labial. In terms of vowel height, the synchrony measure was 14.8 ms shorter in high compared to mid vowels and 26.2 ms shorter in low compared to mid vowels. Finally, with regard to frontness / backness, back rounded vowels had 10.9 ms longer lag than non-back unrounded vowels.

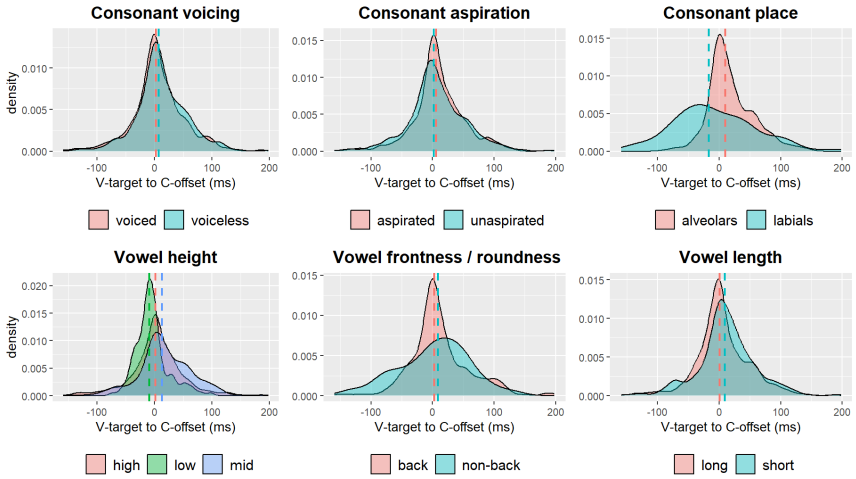

Figure 1: *Distribution of the V-target to C-offset interval across subjects as a function of consonant voicing, aspiration, place, vowel height, frontness / roundness, and length. Vertical lines are the medians in each group.*

Table 2: *Significant effects of consonant and vowel-related factors on gestural kinematics of the consonantal closing and opening movements. Forward slashes denote the absence of significant effects. Asterisks denote the level of statistical significance for each effect in terms of p-value.*

C movement	Kinematic measure	Consonant-related	Vowel-related
Closing movement	displacement	Place*** Aspiration***	Height*** Frontness*
	peak velocity	Place*** Aspiration***	Height***
	stiffness	Place*** Voicing**	Frontness***
Opening movement	displacement	/	Height*** Frontness**
	peak velocity	/	Height ***
	stiffness	Place**	Frontness***

4. Discussion and conclusion

A main result emerging from our data is that CV timing, as quantified by the interval from V-target to C-offset, is more sensitive to vowel quality (vowel height and frontness) than to consonant phonation (voicing and aspiration). Why may this be so? Early studies on English CV sequences (Ostry et al. 1983, Löfqvist and Gracco 1997) reported robust effects of vowel quality on the consonant's kinematics, with any effects of consonant voicing being place-specific or not consistent across subjects. Thus, Löfqvist and Gracco (1997) reported no consistent voicing effect in labial consonant-initial CVs (their stimuli consist of only labials), whereas Ostry et al. (1983) reported such an effect on C displacement and peak velocity in the opening and closing movements for velar consonant-initial CVs (their stimuli consist of only velars). To assess if and how these results on differential effects of consonant and vowel properties on the consonant's kinematics also extend to Hindi's more elaborate system of phonation contrasts, we fitted the model described in the Methods section to our data with six kinematic measures from the consonantal gesture as the dependent variable: displacement, peak velocity, and stiffness of the closing and opening movements. In **Table 2** below, we summarize the significant effects for each kinematic measure grouped by whether they are related to the consonant or the vowel.

It can be seen that while the kinematics of the consonantal closing movement are modulated by both consonant and vowel-related factors, those of the opening movement are almost exclusively vowel-sensitive and immune to consonant phonation. Therefore, effects related to consonant phonation (i.e., voicing and aspiration) on gestural kinematics are not only limited compared to vocalic effects in terms of their number (3 significant aspiration and voicing effects vs. 8 significant height

and frontness effects), but also highly localized on the consonantal closing movement as opposed to the opening movement. Since CV timing mainly concerns the transition between C and V, which mostly encompasses the C opening and V closing movement, the lack of consonantal effects on the kinematics of the consonantal opening movement may be the reason why CV timing is insensitive to consonant phonation as revealed by our results on CV landmark synchronicity shown above.

In conclusion, it has been found that vowel height and frontness and C place of articulation exert significant effects on the interval from V-target to C-offset, whereas C phonation of the initial stop has no significant effect. We sought to explain this finding by demonstrating, in an extension of earlier work on English, that while vowel quality significantly affects movements towards and away from the C constriction, effects of C phonation are confined to the kinematics of the closing movement alone. That is, such effects are absent in the opening movement, which is the one directly involved in the transition between the C and the V. This may then explain the presence of vowel quality effects and the absence of consonant phonation effects in CV timing. Of course, our preliminary results are limited, given the choice to quantify CV timing in the specific way chosen here, which is motivated by recent work reporting on this interval (Shaw & Chen 2019; Kramer et al. 2023; Durvasula & Wang 2023).

We note that despite the fact that vowel quality significantly affects extent of landmark synchrony, the V-target to C-offset interval also shows a relatively high standard deviation as documented in **Table 1** (though not the highest as in the results reported by Kramer et al. 2023), implying that it is not the most stable interval. The more extensive set of CV sequences examined in our data compared to earlier work brings out the specificity of such descriptive statistics (on interval variability) as a function of segmental composition. In turn, these results indicate that taking grand means of these intervals across all CV sequences may not be appropriate given the significant effects of V quality in our data. Moreover, the fact that V-target to C-offset interval is relatively variable both in our data as well as in the data from Kramer et al. (2023) hints at the insufficiency of considering synchrony alone as the sole basis of inter-gestural coordination (as noted in Kramer et al. 2023).

5. Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 317633480 – SFB 1287.

6. References

Bombien, L., & Hoole, P. (2013). Articulatory overlap as a function of voicing in French and German consonant clusters. *The Journal of the Acoustical Society of America*, *134*(1), 539–550. https://doi.org/10.1121/1.4807510

Browman, C. P., & Browman, L. M. (1986). Towards an Articulatory Phonology. *Phonology Yearbook*, *3*, 219–252. JSTOR.

Browman, C. P., & Goldstein, L. M. (1988). Some notes on syllable structure in Articulatory Phonology. *Phonetica*, *45*(2–4), 140–155. https://doi.org/10.1159/000261823

Durvasula, K., & Wang, Y. (2023). Revisiting CV timing with a new technique to identify inter-gestural proportional timing. *Conference Proceedings of the 20th International Congress of Phonetic Sciences*, 2284–2288. https://drive.google.com/file/d/15U2l2y4_-9lyZAgmiccQYXYj9zBi_CAu/view

Gafos, A. I. (2002). A grammar of gestural coordination. *Natural Language & Linguistic Theory*, 20, 269–337.

Halle, M. (1995). Feature geometry and feature spreading'. *Linguistic Inquiry* 26, 1–46.

Honorof, D. N. and Browman, C. P. (1995). The center or edge: how are consonant clusters organized with respect to the vowel. In Elenius, K. and Branderud, P. (Eds). Proceedings of the 13th ICPhS, Stockholm, Sweden, volume 3, pages 552-555.

Katz, J. (2012). Compression effects in English. *Journal of Phonetics*, *40*(3), 390–402. https://doi.org/10.1016/j.wocn.2012.02.004

Kingston, J. (1985). The phonetics and phonology of the timing of oral and glottal events. [Ph.D. dissertation]. University of California, Berkeley.

Kramer, B. M., Stern, M. C., Wang, Y., Liu, Y., & Shaw, J. A. (2023). Synchrony and stability of articulatory landmarks in English and Mandarin CV sequences. *Conference Proceedings of the 20th International Congress of Phonetic Sciences*, 1022–1026. https://drive.google.com/file/d/15U2l2y4_-9lyZAgmiccQYXYj9zBi_CAu/view

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2023). *emmeans: Estimated marginal means, aka least-squares means (Version 1.8.9)* [Computer software]. Retrieved from: https://cran.r-project.org/web/packages/emmeans/index.html

Löfqvist, A., & Gracco, V. L. (1997). Lip and jaw kinematics in bilabial stop consonant production. *Journal of Speech, Language, and Hearing Research*, *40*(4), 877–893. https://doi.org/10.1044/jslhr.4004.877

Nam, H., & Saltzman, E. (2003). A competitive, coupled oscillator model of syllable structure. Proceedings of the 15th International Congress of Phonetic Sciences, 2003, 2253–2256.

Ostry, D. J., Keller, E., & Parush, A. (1983). Similarities in the control of the speech articulators and the limbs: kinematics of tongue dorsum movement in speech. *Journal of Experimental Psychology: Human Perception and Performance*, *9*(4), 622–636. https://doi.org/10.1037/0096-1523.9.4.622

Saltzman, E. (1986). Task dynamic coordination of the speech articulators: a preliminary model. In H. Heuer & C. Fromm (Eds.), *Generation and Modulation of Action Patterns* (pp. 129–144). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-71476-4_10

Sagey, E. (1986). The representation of features and relations in non-linear phonology. [Ph.D. dissertation]. MIT. [Published 1991, Garland, New York.]

Shaw, J. A., & Chen, W. (2019). Spatially conditioned speech timing: evidence and implications. *Frontiers in Psychology*, 10, 2726. https://doi.org/10.3389/fpsyg.2019.02726

Steriade, D. (1993). Closure, release, and nasal contours', in Huffman, M. K. and Krakow, R. A. (Eds.), *Phonetics and Phonology 5: Nasals, Nasalization, and the Velum, Academic Press*, New York, pp. 401–470.

Steriade, D. (1994). Complex onsets as single segments: the Mazateco pattern', in Cole, J. and Kisseberth, C. (Eds.), *Perspectives in Phonology*, CSLI, Stanford, pp. 203–291.

Tiede, M. (2005). MVIEW: Software for visualization and analysis of concurrently recorded movement data. Haskins Laboratories. New Haven, CT.