TEMPORAL SCOPE OF ARTICULATORY SLOWDOWN UNDER INFORMATIONAL FOCUS: DATA FROM ENGLISH AND MANDARIN

Yuyang Liu[1], Yichen Wang[2], Michael C. Stern[1], Benjamin M. Kramer[1], Jason A. Shaw[1]

[1]Yale University, [2]Michigan State University
yuyang.liu.yl2472@yale.edu, wangy176@msu.edu, michael.stern@yale.edu, ben.kramer@yale.edu, jason.shaw@yale.edu

ABSTRACT

Syllables produced under informational focus are longer than nonfocused syllables. Articulatorily, it is unclear whether this lengthening is the result of longer gestures or of less overlap between them and whether the articulatory basis of this effect varies across languages. We investigated these temporal aspects of focus production in English and Mandarin CV sequences using electromagnetic articulography, measuring four theoretically relevant intervals in nonfocused and focused conditions. Our results indicated that both English and Mandarin showed a temporal slowdown under focus, although the scope and magnitude of this slowdown varied across languages. We propose a generalization of the prosodic gesture model to account for the results.

Keywords: focus, Articulatory Phonology, prosodic gesture, electromagnetic articulography

1. INTRODUCTION

Crosslinguistically, focus has been shown to have a wide range of phonetic effects, impacting pitch, segment duration, and the magnitude of articulatory movements [1–7]. In this paper, we concentrate on how informational focus influences the timing of articulatory gestures in CV sequences. From acoustic studies, we know that focused syllables are longer than their closely controlled nonfocused counterparts in both English and Mandarin [2–5]. However, it is unclear what the articulatory basis of focus-modulated lengthening might be and whether it is uniform across languages. One possibility is that the consonant and vowel gestures of a CV sequence are pushed apart in time, resulting in less overlap and therefore longer syllable duration (cf. *localized hyperarticulation* [8]). Alternatively, it could be that focused syllables are longer because their constituent gestures are lengthened.

The Articulatory Phonology (AP; [9] et seq.) framework has mechanisms available to model each of the possibilities for focus realization listed above. Modeling prosody has received increased attention in AP [10], but there is no consensus on how to model focus. Here, we outline three possibilities. First, focus could make use of the same mechanism deployed to account for articulatory slowdown at phrase boundaries. Boundary-adjacent lengthening and strengthening have been modeled in AP using a prosodic gesture (π-gesture), which locally slows down all concurrently active constriction gestures [11–18]. Generalized to focus, the π-gesture model predicts that both consonants and vowels would show temporal slowdown if they are overlapped with a π-gesture. Another possibility is that the focus-induced lengthening is localized to individual gestures: either the consonant gesture, the vowel gesture, or both. This could be implemented by having focus modulate the gestural parameters, such as activation duration or stiffness [19], directly (e.g., [20]). A third possibility is that focus impacts the relative timing of consonant and vowel gestures, reducing overlap [8], which could be implemented in AP in a number of different ways [21]. This possibility might be less likely for Mandarin than for English, since Mandarin generally has less CV overlap [22, 23] (cf. [24–27] for English).

In order to assess these possibilities, we evaluate the effect of focus on four intervals in CV sequences, selected to capture the theoretically relevant aspects of intra- and intergestural timing referenced above. We evaluate the effect of focus in two languages, English and Mandarin. Both of these languages are known to have longer syllables under focus, but syllable lengthening may have different articulatory bases across languages.

2. METHODS

2.1. Participants

Acoustic and articulatory data were collected from 12 native speakers of (American) English (8 female, 4 male; ages 19–28, $\mu = 20.75$) and 12 native speakers of Mandarin (1 nonbinary, 7 female, 4

Item	Language	Condition	Stimuli	
/ni ma/	English	Nonfocused	Prompt	Is she a knee model client?
			Carrier	She's a knee model **representative**, not a knee model client.
		Focused	Prompt	Is she a knee surgeon?
			Carrier	She's a knee **model**, not a knee surgeon.
	Mandarin	Nonfocused	Prompt	Wǒ yīnggāi mà tā háishì mà nǐ? 'Should I scold him or scold you?'
			Carrier	Nǐ mà **tā** jiù xíng le, bié mà wǒ. 'Just scold him; don't scold me.'
		Focused	Prompt	Wǒ yīnggāi mà tā háishì dǎ tā? 'Should I scold him or hit him?'
			Carrier	Nǐ **mà** tā jiù xíng le, bié dǎ tā. 'Just scold him; don't hit him.'

Table 1: Examples of stimuli.
Note: Focused constituents are bolded, and target syllables are underlined.

male; ages 19–33, $\mu = 24$). No participants reported speech or hearing impairments.

2.2. Stimuli

We elicited productions of eight word-initial CV sequences in each language, in which the initial consonant was a bilabial — either /b/ or /m/ — and the vowel was either /a/ or /i/. Target syllables containing /a/ were immediately preceded by /i/, and those containing /i/ were immediately preceded by /a/. Mandarin target syllables bore a falling tone (T4) and were immediately preceded by a low tone (T3). Each target syllable was produced across two conditions: *focused*, in which the word containing the target syllable was informationally prominent in the carrier sentence, and *nonfocused*, in which the word containing the target syllable was not prominent. In order to encourage natural, communicative speech, a prompt accompanied each carrier sentence. Examples of prompts and carrier sentences are given in Table 1.

2.3. Procedure

Prompts and carrier sentences were presented to participants using E-Prime in a sound-attenuated laboratory. Each prompt was displayed on the screen and accompanied by an audio recording of the prompt. The prompt remained on screen for 5 seconds before the carrier sentence was presented. Participants were instructed to listen to the prompt and to read the carrier sentence that followed.

In total, each participant produced 128 tokens (8 items $\times$ 2 conditions $\times$ 8 repetitions) across 4 blocks of 32 items each. The first two blocks each consisted of 4 repetition cycles of stimuli in the nonfocused condition, and the last two blocks each consisted of 4 repetition cycles of stimuli in the focused condition. Within each repetition cycle, stimuli were presented in a randomized order.

The NDI Wave Speech Research System was used to record movements of nine sensors attached to the articulators and head at a sampling rate of 100 Hz. High-viscosity PeriAcryl was used to attach three sensors to the tongue: tongue tip (TT), tongue blade (TB), and tongue dorsum (TD), placed ~1 cm, ~3 cm, and ~5 cm from the tip of the tongue, respectively. In order to track movements of the jaw, one lower incisor (LI) sensor was attached to the hard tissue of the gum directly below the left incisor. Two sensors were attached at the vermillion border of the upper lip (UL) and lower lip (LL). Reference sensors were attached on the left and right mastoids and on the nasion. Measurements of the occlusal plane and a midsagittal palate trace were also collected. Acoustic data were collected using a Sennheiser shotgun microphone at a sampling rate of 22,050 Hz.

2.4. Analysis

Articulatory data were rotated to the occlusal plane and corrected for head movement computationally.[1] Articulatory gestural landmarks were parsed from the sensor trajectories using MVIEW, a MATLAB-based program for articulatory data visualization.[2] Gestures associated with /b/ and /m/ were extracted from measurements of lip aperture (LA), calculated as the Euclidean distance between the UL and LL sensors. Gestures associated with /a/ were parsed from the trajectory of the TD sensor. Gestures associated with /i/ were parsed from the trajectory of the TB sensor if this sensor was judged to be closest to the palate at the point of maximum /i/ constriction for a participant and from the TD sensor otherwise.

The following gestural landmarks entered into the calculation of the intervals for analysis: the *onset* of controlled movement, the achievement of *target*, the *point of minimal velocity*, the *release* from constriction, and the *offset* of controlled movement. These landmarks were parsed from the continuous trajectories with reference to the velocity signal. The onset and target of a gesture were measured as the timepoints at which the gesture's tangential velocity

Language	Condition	CV lag	C_{CLOS}	C_{OPEN}	V_{OPEN}
English	Nonfocused	36.02 (13.17)	74.39 (8.44)	86.89 (14.06)	146.90 (13.45)
	Focused	41.02 (17.82)	79.73 (10.35)	102.38 (14.21)	174.08 (19.07)
Mandarin	Nonfocused	38.18 (13.25)	81.38 (7.92)	81.58 (6.24)	135.57 (16.09)
	Focused	36.40 (10.91)	83.11 (7.67)	82.93 (10.55)	145.92 (10.96)

Table 2: Means (standard deviations) of key interval durations (ms).

Fixed effect		CV lag	C_{CLOS}	C_{OPEN}	V_{OPEN}
Focus estimate (ms)	English	5.52***	5.56***	16.87***	27.53***
	Mandarin	.51	1.39	2.34	9.25***
Focus + language χ^2		.28	2.78	7.70**	9.46**
Focus × language interaction χ^2		3.77	4.24*	33.22***	32.21***

Table 3: ANOVA results of LMER models.
Note: *** $p \leq .001$, ** $p \leq .01$, * $p \leq .05$.

in the movement towards constriction exceeded or sank below, respectively, a 20% threshold of a manually selected local velocity peak. The point of minimal velocity was measured as the timepoint of the velocity minimum. The release and offset of a gesture were measured as the timepoints at which the gesture's tangential velocity in the movement away from constriction exceeded or sank below, respectively, a 20% threshold of a manually selected local velocity peak. Figure 1 provides an example of the gestural landmarks defined above.

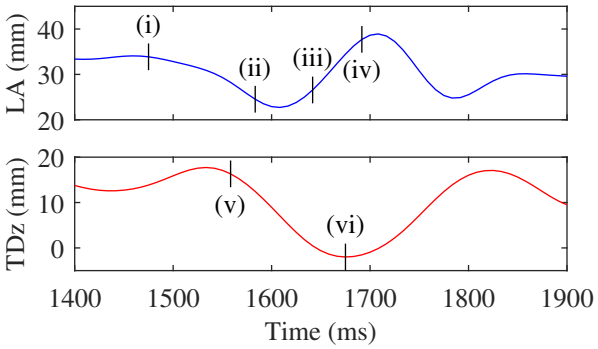


Figure 1: Gestural landmarks.
Note: (i) C onset, (ii) C target, (iii) C release, (iv) C offset, (v) V onset, (vi) V minimal velocity.

Out of the 3,072 tokens elicited, a total of 556 tokens (18.10%) were eliminated from analysis for the following reasons: data storage issues (18); disfluency (5); failure of the participant to produce contrastive focus on the informationally prominent syllable (155); or failure of the gesture parsing tool to extract the consonant or vowel gesture (378).

For all the remaining tokens, we calculated the following four intervals: (i) *CV lag* is defined as the interval from the consonant onset to the vowel onset; (ii) *consonant closing interval* (C_{CLOS}) is defined as the interval from the consonant onset to the consonant target; (iii) *consonant opening interval* (C_{OPEN}) is defined as the interval from the consonant release to the consonant offset; and (iv) *vowel opening interval* (V_{OPEN}) is defined as the interval from the onset to the point of minimal velocity of the vowel gesture. After calculating the intervals, 82 tokens, for which at least one interval was beyond three standard deviations from the mean, were excluded from analysis.

Linear mixed effects regression (LMER) models were fit to each of the intervals defined above [28].[3] We first assessed the effect of focus in each language separately. For each interval in each language, a baseline model included a fixed effect of vowel and random intercepts for subject and item. We then added focus as a fixed effect and compared it to the baseline model. We next fit a model to each interval in the combined data set, including a fixed effect of language. Finally, to statistically assess whether the effect of focus was uniform across languages, we added a fixed effect for the interaction between focus and language. For all these models, statistical significance was determined by a likelihood ratio test against a baseline model that excluded the fixed effect of interest.

3. RESULTS

Means and standard deviations of the durations of each key interval under nonfocused and focused conditions are given in Table 2. In both languages, the difference in the mean duration between the two conditions was the largest for V_{OPEN}. Additionally, for each interval, the effect of focus was larger in English than in Mandarin.

The top rows of Table 3 report the estimate for the fixed effect of focus for each interval and language

along with statistical significance. For English, focus had the biggest effect on V_{OPEN}, followed by C_{OPEN}, followed by C_{CLOS}, followed by CV lag. Mandarin followed the same numerical trend, but the effect was only significant for V_{OPEN}. The effect of language and the interaction between language and focus are reported in the bottom rows. Results show a main effect of language for C_{OPEN} and V_{OPEN} (longer for English) and confirm a significant interaction. The effect of focus is stronger in English for C_{CLOS}, C_{OPEN}, and V_{OPEN}.[4]

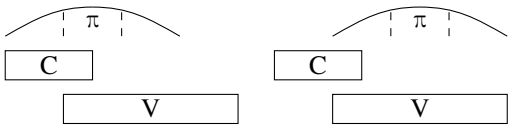
4. DISCUSSION

Our results showed that focus had a lengthening effect in English and Mandarin. In both languages, some intervals were longer in focused syllables than in nonfocused syllables, as expected from past work [2–7]. However, the locus and magnitude of the effect varied across languages. In English but not in Mandarin, focus had a significant effect on the closing and opening phases of the consonant gesture and on the relative timing between the consonant and vowel gestures. Additionally, while focus had a significant effect on the opening phase of the vowel gesture in both languages, it had a stronger effect in English than in Mandarin.

These data allow us to rule out two of the three possible theoretical accounts of the effect of focus on timing outlined in the introduction. It seems that focused syllables are not longer because of less overlap between gestures; there was no effect of focus on CV lag in Mandarin and only a small effect in English, similar in magnitude to the effect on the closing phase of the consonant gesture. We can therefore rule out localized hyperarticulation as an articulatory basis of focus lengthening. It also appears that focus is not directly modulating single gestures; at least in English, the temporal scope of focus encompasses the entire CV sequence.

The third possibility that we raised was the π-gesture model. To date, this has been primarily used for explaining the local lengthening and slowing of articulatory gestures at phrase boundaries [11–18]. The π-gesture is argued to have an extent in time but lack independent articulatory realization; it may only be realized vicariously through its effect on the concurrently active vocal tract constriction gestures [16]. Essentially, the π-gesture triggers a transgestural local slowdown in proportion to its activation level on all constriction gestures with which it overlaps [13, 14, 16]. This approach has promise for accounting for the effects of focus in these data, including the crosslinguistic variation observed.

We propose that the target of the temporal effects of focus is a continuously ramped π-gesture, whose activation level is the highest at the midpoint of the gesture and the lowest at its onset and offset. As a consequence of the focus-induced π-gesture, all concurrently active constriction gestures are slowed down and therefore lengthened, as is observed at phrase boundaries. To account for the language-specific effects that we observed, we posit that the π-gesture may be temporally aligned to different timepoints in the constriction dynamics in different languages: in English, the onset of the π-gesture is aligned to the onset of the syllable, whereas in Mandarin, it is aligned to the onset of the vowel gesture. This hypothesis is illustrated in Figure 2.



(a) English. **(b)** Mandarin.

Figure 2: Proposed gestural alignments.

As in Figure 2(a), in English, the π-gesture has its onset aligned to the onset of the syllable and is maximally activated at the opening phase of the vowel gesture, which partially overlaps with the closing phase of the consonant gesture. As a result, when focus is realized on the syllable, V_{OPEN} and C_{OPEN} get slowed down the most by the π-gesture, and C_{CLOS} less so. In Mandarin, on the other hand, the onset of the π-gesture is aligned to the onset of the vowel gesture, as in Figure 2(b). This way, since the consonant gesture hardly overlaps with the π-gesture, focus has no effect on its duration. In contrast, focus does have an effect on V_{OPEN} but not as strong as the effect found in English, because Mandarin π-gestures are not yet maximally activated at the opening phase of the vowel.

5. CONCLUSION

Past work has shown that informational focus has a lengthening effect in both English and Mandarin. We explored the articulatory basis of this effect, finding that focus increases the gestural duration of vowels (both languages) and onset consonants (English only). Both languages showed focus-driven articulatory lengthening, but its scope and magnitude differed. We argued that this pattern of results is consistent with a π-gesture account of focus. On this account, language variation in focus realization derives from the temporal alignment of the π-gesture.

6. REFERENCES

[1] D. R. Ladd, "Phonological features of intonational peaks," *Language*, vol. 59, no. 4, pp. 721–759, Dec. 1983.

[2] W. Cooper, S. Eady, and P. Mueller, "Acoustical aspects of contrastive stress in question-answer contexts," *J. Acoust. Soc. Am.*, vol. 77, no. 6, pp. 2142–2156, Jun. 1985.

[3] M. Beckman and J. Pierrehumbert, "Intonational structure in Japanese and English," *Phonol. Yearb.*, vol. 3, pp. 255–309, May 1986.

[4] Y. Xu, "Effects of tone and focus on the formation and alignment of f_0 contours," *J. Phon.*, vol. 27, no. 1, pp. 55–105, Jan. 1999.

[5] Y. Xu and C. X. Xu, "Phonetic realization of focus in English declarative intonation," *J. Phon.*, vol. 33, no. 2, pp. 159–197, Apr. 2005.

[6] S. Roessig and D. Mücke, "Modeling dimensions of prosodic prominence," *Front. Commun.*, vol. 4, p. 44, Sep. 2019.

[7] S. Roessig, B. Winter, and D. Mücke, "Tracing the phonetic space of prosodic focus marking," *Front. Artif. Intell.*, vol. 5, May 2022, Art. no. 842546.

[8] K. de Jong, "The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation," *J. Acoust. Soc. Am.*, vol. 97, no. 1, pp. 491–504, Jan. 1995.

[9] C. Browman and L. Goldstein, "Articulatory phonology: An overview," *Phonetica*, vol. 49, no. 3–4, pp. 155–180, May 1992.

[10] D. Byrd and J. Krivokapić, "Cracking prosody in Articulatory Phonology," *Annu. Rev. Linguist.*, vol. 7, no. 1, pp. 31–53, Jan. 2021.

[11] M. Beckman, J. Edwards, and J. Fletcher, "Prosodic structure and tempo in a sonority model of articulatory dynamics," in *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, G. Docherty and D. R. Ladd, Eds. Cambridge, U.K.: Cambridge Univ. Press, 1992, pp. 68–89.

[12] D. Byrd and E. Saltzman, "Intragestural dynamics of multiple prosodic boundaries," *J. Phon.*, vol. 26, no. 2, pp. 173–199, Apr. 1998.

[13] D. Byrd, "Articulatory vowel lengthening and coordination at phrasal junctures," *Phonetica*, vol. 57, no. 1, pp. 3–16, Mar. 2000.

[14] D. Byrd, A. Kaun, S. Narayanan, and E. Saltzman, "Phrasal signatures in articulation," in *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, M. Broe and J. Pierrehumbert, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2000, pp. 70–87.

[15] T. Cho, *The Effects of Prosody on Articulation in English*. New York, NY, USA: Routledge, 2002.

[16] D. Byrd and E. Saltzman, "The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening," *J. Phon.*, vol. 31, no. 2, pp. 149–180, Apr. 2003.

[17] T. Cho, "Manifestation of prosodic structure in articulatory variation: Evidence from lip kinematics in English," in *Laboratory Phonology 8*, L. Goldstein, D. Whalen, and C. Best, Eds. Berlin, Germany: Mouton de Gruyter, 2006, pp. 519–548.

[18] D. Byrd, J. Krivokapić, and S. Lee, "How far, how long: On the temporal scope of prosodic boundary effects," *J. Acoust. Soc. Am.*, vol. 120, no. 3, pp. 1589–1599, Sep. 2006.

[19] C. Browman and L. Goldstein, "Gestural specification using dynamically-defined articulatory structures," *J. Phon.*, vol. 18, no. 3, pp. 299–320, Jul. 1990.

[20] T. Cho, "Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /ɑ, i/ in English," *J. Acoust. Soc. Am.*, vol. 117, no. 6, pp. 3867–3878, Jun. 2005.

[21] E. Saltzman, H. Nam, J. Krivokapić, and L. Goldstein, "A task-dynamic toolkit for modeling the effects of prosodic structure on articulation," *Speech Prosody*, vol. 4, pp. 175–184, May 2008.

[22] M. Gao, "Mandarin tones: An Articulatory Phonology account," Ph.D. dissertation, Dept. Linguist., Yale Univ., New Haven, CT, 2008.

[23] M. Zhang, C. Geissler, and J. Shaw, "Gestural representations of tone in Mandarin: Evidence from timing alternations," *Int. Congr. Phon. Sci.*, vol. 19, pp. 1803–1807, Aug. 2019.

[24] A. Löfqvist and V. Gracco, "Interarticulator programming in VCV sequences: Lip and tongue movements," *J. Acoust. Soc. Am.*, vol. 105, no. 3, pp. 1864–1876, Mar. 1999.

[25] C. Browman and L. Goldstein, "Competing constraints on intergestural coordination and self-organization of phonological structures," *Bull. Commun. Parlée*, vol. 5, pp. 25–34, Dec. 2000.

[26] H. Nam, L. Goldstein, and E. Saltzman, "Self-organization of syllable structure: A coupled oscillator model," in *Approaches to Phonological Complexity*, F. Pellegrino, E. Marsico, I. Chitoran, and C. Coupé, Eds. Berlin, Germany: Mouton de Gruyter, 2009, pp. 297–328.

[27] S. Marin and M. Pouplier, "Temporal organization of complex onsets and codas in American English: Testing the predictions of a gestural coupling model," *Mot. Control*, vol. 14, no. 3, pp. 380–407, Jul. 2010.

[28] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using **lme4**," *J. Stat. Softw.*, vol. 67, no. 1, pp. 1–48, Oct. 2015.

[1] The MathWorks, Inc. *MATLAB version: 9.13.0 (R2022b)*. (2022). The MathWorks, Inc.

[2] M. Tiede. *MVIEW: Software for visualization and analysis of concurrently recorded movement data*. (2005). Haskins Laboratories.

[3] R Core Team. *R: A language and environment for statistical computing*. (2022). R Foundation for Statistical Computing.

[4] Data and complete statistical analysis are publicly available at https://osf.io/xkj72/.