Crosslinguistic Influence on the Gestural Dynamics of Focus Prosody in Native Mandarin Learners of L2 English

Yichen Wang[1], Benjamin Kramer[2], Noah Macey[2], Michael Stern[2], Yuyang Liu[2], Jason Shaw[2]

[1]*Michigan State University (USA),* [2]*Yale University (USA)*
wangy176@msu.edu, ben.kramer@yale.edu, noah.macey@yale.edu,
michael.stern@yale.edu, yuyang.liu.yl2472@yale.edu, Jason.shaw@yale.edu

1. Introduction. Studies on L2 prosody have found evidence for the influence of L1 on L2 prosody production across a range of language pairs [1–5]. L2 speakers often exhibit prosodic patterns—accentual peaks, stress patterns and tonal contours—that can be attributed to L1 influence (*Transfer Hypothesis*; [6]). However, little has been done to capture such patterns formally. Using the parameters of Articulatory Phonology (AP; [7]), we explore how L2 prosody relates to L1 and to the target language. We investigate this issue in the kinematics of L2 focus production, considering highly proficient L2 learners of English from Mandarin L1. In both of these languages, as in others, syllables under informational focus tend to be longer in duration than unfocused syllables [8–13]. In AP, increased duration can be explained by a π-gesture [14]. The temporal scope of the π-gesture may vary across languages, possibly in ways that could be obscured by acoustic measures of syllable duration. We therefore collected kinematic data using Electromagnetic Articulography and analyzed sub-intervals of the syllable based on articulatory landmarks. Focused and unfocused syllables produced by L1 Mandarin and L1 English speakers established baseline patterns of π-gesture alignment in each language, against which we assessed L2 focus prosody.

2. Methods. Participants were 12 native speakers of Mandarin Chinese who were also L2 speakers of English and 12 native speakers of American English. Materials included eight word-initial CV sequences, consisting of /b/ or /m/ followed by /ɑ/ or /i/ in each language. All Mandarin target syllables carried a falling tone (T4). Each was produced in a *focused* condition, in which it was located in an informationally prominent position in the sentence, and a *non-focused* one. Carrier sentences were preceded by a context presented auditorily and orthographically. Nine sensors attached to the articulators and head were tracked using the NDI Wave Speech Research System at a sampling rate of 100 Hz. Acoustic data were recorded concurrently. Sensors were attached to the tongue tip, tongue blade, tongue dorsum, lower incisor, upper lip, and lower lip. Reference sensors were attached on the nasion or bridge of the nose, as well as the right and left mastoids. After computationally correcting for head movements, target C and V gestures were parsed from sensor trajectories in MVIEW [15]. We used the lips to parse labial consonants and the tongue body for vowels. Gestural landmarks were used to defined four key intervals: *CV lag* (consonant onset to vowel onset), *consonant closing interval* (C_{CLOS}; consonant onset to consonant target), *consonant opening interval* (C_{OPEN}; consonant release to consonant offset), and *vowel opening interval* (V_{OPEN}; vowel onset to point of minimum vowel velocity). A total of 3612 tokens entered into the analysis.

3. Results. Table 1 presents means and standard deviations of the target intervals for L1 English, L1 Mandarin, and L2 English, organized from the beginning of the syllable (left) to the end of the syllable (right). Most of the intervals were longer in syllables produced in the *focused* condition than in syllables produced in the *non-focused* condition. We fit linear mixed effects regression (LMER) models to the four sub-intervals (Table 2). Models included random intercepts for item and subject, a control fixed factor for vowel, and an experimental fixed factor for focus. Results indicated a significant effect of focus on all intervals in L1 English, whereby each interval was longer in the *focused* condition than in the *non-focused* condition. The effect increased in magnitude over time: smallest for CV lag, progressively larger for C_{CLOS}, C_{OPEN}, and V_{OPEN}. For Mandarin, the model indicated a significant lengthening effect of focus only on the final interval V_{OPEN}, and this effect was smaller than the effect in English (9.25 in Mandarin, 27.53 in English). For L2 English, the effect of focus was significant for the last two intervals

C_{OPEN} and V_{OPEN}. Notably, the sizes of the effects in L2 English were intermediate between the sizes of the effects in L1 English and Mandarin.

Language	Condition	CV lag	C_{CLOS}	C_{OPEN}	V_{OPEN}
L1 English	non-focused	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)
L1 Mandarin	non-focused	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)
L2 English	non-focused	43.34 (13.43)	83.01 (5.37)	96.10 (12.69)	142.25 (15.77)
	focused	46.55 (17.57)	84.12 (10.54)	103.71 (11.90)	159.33 (15.46)

Table 1. Means (standard deviations) of key interval durations (ms).

Focus estimate (ms) & significance	CV lag	C_{CLOS}	C_{OPEN}	V_{OPEN}
L1 English	5.52***	5.56***	16.87***	27.53***
L1 Mandarin	0.51✗	1.39✗	2.34✗	9.25***
L2 English	4.32✗	0.81✗	7.10**	15.66***

Table 2. Focus estimates and ANOVA results of LMER models.
Note: ✗ $p \leq 1$, . $p \leq .1$, * $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$.

4. Discussion. The results indicate that temporal slowdown occurred in both English and Mandarin under focus. However, the two languages exhibited different patterns in the realization of focus. Since the slowdown started earlier in English than in Mandarin, we posit an earlier onset of the π-gesture. In English, the π-gesture aligns to the start of the syllable, ramping up from the onset of the syllable to maximum slowdown (π-gesture peak) at the vowel opening. In Mandarin, the π-gesture is aligned later in the syllable, possibly coupled to the start of the vowel.

The L2 speakers of English showed an intermediate pattern. The π-gesture seemed to exert its influence later in the syllable than in L1 English but earlier than in L1 Mandarin. One possible interpretation is that L1 speakers of Mandarin maintain a coupling relation between the vowel and the π-gesture even as they acquire English, adding a new coupling relation between the π-gesture and the consonant. In this way, the intermediate pattern could surface because the π-gesture is competitively coupled to both the consonant and the vowel gesture resulting in a "C-center effect" [16]), reflecting influences of both languages. The competition account can also be formalized in Dynamic Field Theory (DFT; [17, 18]). In a DFT planning model, a dynamic neural field governing the timing of the π-gesture within the syllable could stabilize under input from L1 and L2 at an intermediate value (cf. [19, 20]).

References

[1] Chen, Y., Robb, M. P., Gilbert, H. R., & Lerman, J. W. (2001). A study of sentence stress production in mandarin speakers of american english. *The Journal of the Acoustical Society of America*, *109*(4).

[2] Juffs, A. (1990). Tone, syllable structure and interlanguage phonology: Chinese learners' stress errors.

[3] Li, A., & Post, B. (2014). L2 acquisition of prosodic properties of speech rhythm: Evidence from l1 mandarin and german learners of english. *Studies in Second Language Acquisition*, *36*(2).

[4] Hosseini, A. (2013). L1 interference in l2 prosody: Contrastive focus in japanese and persian. *Journal of Logic Language and Information*, *11*.

[5] Li, P., Zhang, F., Yu, A., & Zhao, X. (2020). Language history questionnaire (lhq3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition*, *23*(5).

[6] Clark, H. (1987). The stress accent in adult second language acquisition.

[7] Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, *49*(3-4).

[8] Cooper, W., Eady, S., & Mueller, P. (1985). Acoustical aspects of contrastive stress in question-answer contexts. *J. Acoust. Soc. Am.*, *77*(6).

[9] Beckman, M., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonol. Yearb.*, *3*.

[10] Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f_0 contours. *J. Phon.*, *27*(1).

[11] Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *J. Phon.*, *33*(2).

[12] Roessig, S., & Mücke, D. (2019). Modeling dimensions of prosodic prominence. *Front. Commun.*, *4*(44).

[13] Roessig, S., Winter, B., & Mücke, D. (2022). Tracing the phonetic space of prosodic focus marking [Art. no. 842546]. *Front. Artif. Intell.*, *5*.

[14] Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, *31*(2).

[15] Tiede, M. (2005). Mview: Software for visualization and analysis of concurrently recorded movement data. *New Haven, CT: Haskins Laboratories*.

[16] Browman, C. P., & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica*, *45*(2-4).

[17] Schöner, G., & Spencer, J. P. (2016). *Dynamic thinking: A primer on dynamic field theory*. Oxford University Press.

[18] Erlhagen, W., & Schöner, G. (2002). Dynamic field theory of movement preparation. *Psychological review*, *109*(3).

[19] Stern, M. C., & Shaw, J. A. (2022). Neural inhibition during speech planning contributes to contrastive hyperarticulation. *arXiv preprint arXiv:2209.12278*.

[20] Stern, M. C., Chaturvedi, M., & Shaw, J. A. (2022). A dynamic neural field model of phonetic trace effects in speech errors. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*(44).