CONSONANT-VOWEL COARTICULATION PATTERNS IN SWEDISH AND MANDARIN

Man Gao[1], Malin Svensson Lundmark[2]

[1]Dalarna University, [2]Lund University, [2]University of Southern Denmark
[1]mao@du.se, [2]malin.svensson_lundmark@ling.lu.se

ABSTRACT

This paper reports a cross linguistic study that compares the coarticulation patterns between consonant and vowel (CV) in Mandarin Chinese and Southern Swedish. Kinematic data were collected using the Electromagnetic Articulography (EMA) for both languages and were subjected to three types of CV time lag measurement, based on more or less equivalent landmarks on lips and tongue, and partially adopted in previous studies [1, 2, 3]. We found rather consistent CV coordination patterns in these two typologically different languages with both the velocity-based and the acceleration-based measurements on the lips and the tongue body. The most striking result to emerge from the data is the same effect of gender on the variation of CV coarticulation in both languages, which has not been reported previously. In addition, only when gender was added as a factor, did we find the language differences on the CV time lags.

Keywords: CV coarticulation, EMA, Mandarin Chinese, Southern Swedish, gender difference

1. INTRODUCTION

In the past half century, new methods and equipment have been developed to obtain a greater understanding of the temporal and spatial coordination patterns of various speech articulators. For example, some studies [1, 2, 4] have used the Electromagnetic Midsagittal Articulometer to collect 2D kinematic data for analyzing the interarticulator coordination in sequences of segment or in words that carry lexical tones. Other studies [3, 5] have used the ElectroMagnetic Articulography (EMA) to collect 3D kinematic data for analyzing the consonant and vowel coordination in pitch accents. It has been observed in these studies [2, 3, 5] that the word initial consonant and following vowel (CV) exhibit a strong articulatory overlap to a considerable extent.

Löfqvist and Gracco [1] recruited American English speakers to investigate the issue of coarticulation in the sequence of V1CV2, in which C is a labial stop. They found that the tongue articulator moves for the second vowel synchronously or even before the start of the lips close for the consonant. In a study which set out to examine the segment-to-segment and tone-to-segment temporal coordination patterns in Mandarin words, Gao [2] reported that the word-initial consonant and its following vowel are aligned in a constant manner, about 45 ms apart with the consonant articulator activated first. Most recently, detailed examination of the word initial C-V coordination in Swedish words by Svensson Lundmark et al. [3] reported both synchronous coarticulation as in [1] and coarticulation with a temporal lag as in [2], depending on the methodological choices such as landmarks measured and the participants.

While the articulatory overlap patterns may be a language-specific feature, the varied patterns observed in previous studies could be a result of lack of standardized measures. In the aforementioned studies, different landmarks were selected for computing the coarticulation. For example, Gao [2] measured the onset of consonant and vowel on the basis of velocity curve of the corresponding articulators, but the measurement in Löfqvist and Gracco [1] was based on the tangential velocity of the tongue body movement for vowel and the peak acceleration of the lips for consonant. Recent work on peak acceleration of the lip articulator shows that it is systematically timed with the acoustic segments across speakers [6]. While the nature of peak acceleration of the tongue body onset movement is yet to be investigated, it is not entirely clear whether the consonantal and the vocalic gestural onsets should be measured using equivalent landmarks on velocity or acceleration, or because of the different linguistic functions of the articulators, use non-equivalent landmarks.

This paper reports an initial attempt to investigate the consonant-vowel coarticulation patterns in Mandarin Chinese and Southern Swedish. Specifically, two issues will be addressed: first is to better understand the interarticulator coordination pattern in two typologically different languages. And the other issue is methodological: to evaluate three types of measurements with more or less equivalent landmarks, some of which have been used previously for computing the CV coarticulation on both sets of language data.

2. METHOD

The speech material is EMA data from two Mandarin (one female) and four Southern Swedish speakers (two female). It contains 474 syllables on the CV sequence /ma/: 46 tokens from Mandarin, and 428 from Swedish (which is part of a larger corpus of EMA data on 21 Swedish speakers, see e.g. [3]).

In Swedish, the target /ma/ receives the primary stress and is the first syllable in a disyllabic word. It's an open syllable which contains a bilabial consonant [m] and a vocalic nucleus [ɑː], followed by either a [l] or an [n] in the next syllable. In Accent 1 (a tonal fall) these are in turn followed by the suffix +en, while in Accent 2 (a tonal rise) by +ar. The target words are placed in statements preceded by leading questions, to ensure a non-focused elicitation.

In Mandarin the target syllable /ma/ can carry four lexical tones, which correspond to four different words. The target word is embedded in the statements with conflicting tonal environment, which are the same as the speech material used in [2]. The test syllable /ma/ of the two languages, occurs in disyllabic and monosyllabic words respectively, due to the different tone carrying units in Mandarin and Swedish. As they are both embedded in sentence-medial positions, and receive primary stress, this difference should not affect the comparison of CV coarticulation.

2.1. Procedure

All speakers were recorded at 250 Hz with EMA; a Carstens AG501 at the Lund University Humanities Laboratory. Audio was recorded simultaneously using an external condenser microphone (a t.bone EM 9600) at a sampling rate of 48 kHz. The Swedish speakers read leading and target sentences from a prompter in a random order, each set appearing eight times. The Mandarin speakers read a set of target sentences six times in random order.

The authors separately segmented the acoustic data of the two languages in Praat [7]. The textgrid files were later used in R [8] as reference time windows for collection of the articulatory data. An inter-annotation agreement (IAA) between the two authors was performed on 60 tokens randomly collected from the large Swedish corpus, showing a good agreement for segment boundaries between the two authors with a mean deviation of 2.4 ms, and 93.4% of tokens (213 of 228) within 10 ms [9].

2.1.1. Articulatory data

Articulatory data were collected from six sensors: two placed on the upper and the lower lip at the vermilion border, one on the lower incisor, and three sensors placed on the midline of the tongue. The first tongue sensor was placed on the tongue body where the participant made a bite mark after having stretched out his or her tongue as far as possible. The second tongue sensor was placed between the sensor at the back and a third tongue sensor, which was placed approximately 1 cm from the tongue tip. To correct for head movements three additional sensors were used: one behind each ear and one on the nose ridge. Post-processing was done in the Carstens software. Only the sensor positions of the lips and the sensor furthest back on the tongue body (TB) were further analyzed in this study.

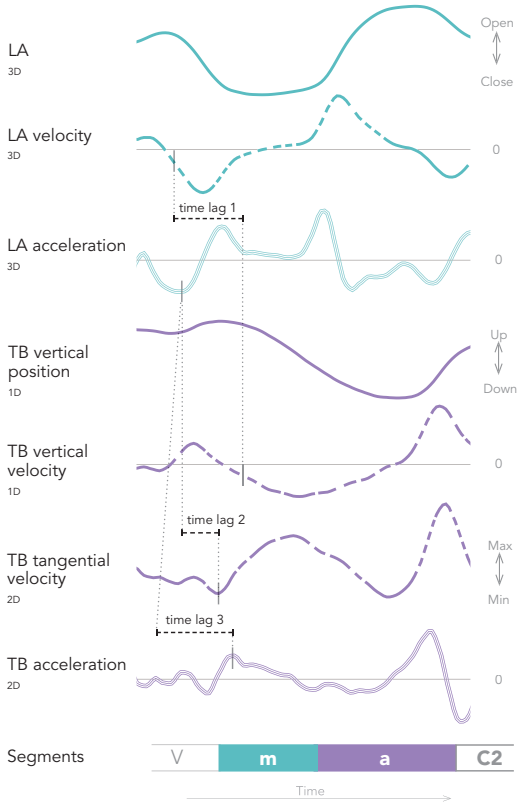

Figure 1: Lip aperture (LA) and tongue body (TB) landmarks used to calculate the CV time lags (Swedish word example). Positive time lags signify that TB onset follows LA onset, while negative time lags mean TB precedes LA.

2.1.2. Articulatory measurements

The data were then processed in R [8]. Lip aperture (LA) was calculated using the three-dimensional (3D) Euclidian distance between the sensors on the upper and the lower lips. The articulatory data was smoothed using locally weighted regression by the R function *loess* (low span 0.1). Articulatory landmarks were automatically collected using the textgrid files as reference time windows. CV time lags was calculated using the following velocity and acceleration landmarks at the onsets of the bilabial closing and the tongue movement (Fig. 1):

1) 20% threshold from 0-crossing to peak velocity of LA (3D), and of TB lowering (one-dimensional, 1D) (based on [2]);
2) peak acceleration of LA (3D) and the minimum tangential velocity of TB (two-dimensional, 2D) (based on [1]);
3) peak acceleration of LA (3D) and of TB (2D).

2.2. Statistical analysis

All statistical tests were run in R [8]. Generalized linear mixed models (GLMM) were run with speaker and tone/accent as random effects (random intercept). Language, and subsequently also gender, were set as fixed effects. Likelihood ratio tests were performed to evaluate added complexity (following [10]). The models were run using the lme4-package [11] and the lmer Test-package [12]. In addition, one-way ANOVA tests were also used to determine whether there were any statistically significant differences between the tones/accents within each language.

3. RESULTS

3.1. CV time lag measures

Results from the CV time lag 1 measure are similar to previously reported results [2, 3]: we find time lags of about 60 ms for Swedish speakers, and about 42 ms for Mandarin speakers (Table 1). Even though Figure 2a suggests otherwise, there is no statistically significant difference between the languages (t =-1.14, p =.297). The one-way ANOVA tests reveal no statistically significant differences among the four Mandarin tones or between the two Swedish word accents.

CV time lag 2, the time lag of the combined acceleration/velocity measurement, displays more or less synchronous timing for Swedish, and negative time lags of about -28 ms for Mandarin (Fig. 2b). The language difference is not statistically significant (t =-1.8, p =.112) (Table 1). Figure 2b indicates less varied time lags for some of the tones of the Mandarin speakers. The one-way ANOVA reveals no statistically significant difference between the four Mandarin tones. However, a significant difference is found between Swedish A1 and A2 (F (1,125) = 8.255, p <.01), which is similar to previous reported results [3].

Similarly, the third time lag measure (CV time lag 3, based on acceleration peaks) does not display any difference between the two languages (t =-0.8, p =.452) (Table 1). Figure 2c suggests less varied time lags for the Mandarin speakers only. The one-way ANOVA indicates statistically significant difference among the Mandarin tones (F (3,42) = 4.243, p <.05), and the post hoc Tukey test reveals a statistically significant difference between T1 and T3 (p <.01). Between the Swedish word accents the difference is marginally significant (F (1,116) = 3.352, p =.07).

		Estimate	SE	df	t-value	p-value
CV time lag 1	(Intercept) Sw	60.58	9.66	5.95	6.27	.001
	Language Ma	-19.19	16.80	6.06	-1.14	.297
CV time lag 2	(Intercept) Sw	2.48	9.53	7.56	0.26	.801
	Language Ma	-27.65	15.38	7.49	-1.80	.112
CV time lag 3	(Intercept) Sw	34.86	7.05	4.87	4.95	.005
	Language Ma	-9.17	11.44	6.34	-0.80	.452

Table 1: GLMM models on the CV time lags, with language as fixed effect.

3.2. Gender differences

Because of the large time lag variation observed for the Swedish speakers, ad-hoc one-way ANOVA tests were performed on gender variations within each language. Figure 3 displays the results on the combined gender and language approach: taking LA onset as the reference point, TB onset for male speakers precedes TB onset for female speakers. The ANOVA tests show that the temporal lags between the male and female speakers differ significantly at the level of p <.01 for all three measures. However, as the ANOVAs do not account for speaker or tone variability, ad-hoc GLMM models adding gender are also performed: complexity was only warranted for CV time lag 1, showing significant differences on gender and language (Table 2). Thus, language differences are found, but only for the velocity-based measure, and only when gender is added as a factor.

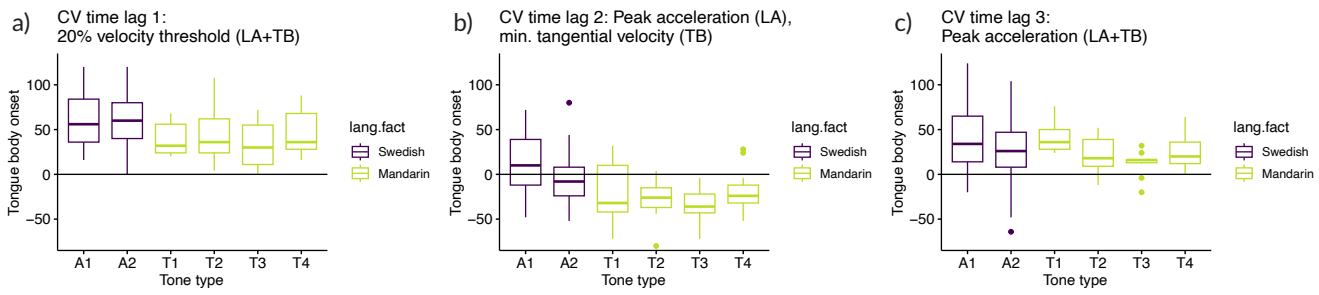

Figure 2: CV time lags (in ms): a) time lag 1, based on velocity; b) time lag 2, based on acceleration + velocity; c) time lag 3, based on acceleration. 0 marks LA onset, boxes marks TB onset. Swedish word accents: A1 and A2; Mandarin lexical tones: T1, T2, T3 and T4.

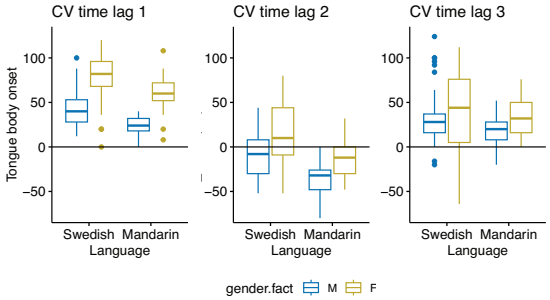

Figure 3: CV time lags 1-3 (in ms) divided according to gender and language. 0 marks LA onset, boxes marks TB onset.

		Estimate	SE	df	*t*	*p*
	(Intercept) Sw M	40.19	5.00	9.30	8.04	.000
CV time	Gender F	40.43	3.06	6.06	13.24	.000
lag 1	Language Ma	-17.50	7.64	12.60	-2.29	.039
	Gend F:Lang Ma	-3.63	5.94	167.08	-0.61	.541
CV time	(Intercept) M	-19.78	9.14	6.20	-2.16	.072
lag 2	Gender F	24.71	12.15	4.67	2.03	.102
CV time	(Intercept) M	26.57	7.16	6.39	3.71	.009
lag 3	Gender F	9.58	9.37	5.69	1.02	.348

Table 2: GLMM models on language and gender as fixed effects (complexity added only for CV time lag 1).

4. DISCUSSION AND CONCLUSION

The velocity-based (CV time lag 1) and the combined acceleration/velocity-based measurements (CV time lag 2), which were adopted repeatedly in previous studies [1, 2, 3, 13], yield similar results that are in line with previous work [2, 3], and further suggest no significant difference between the two languages. Since pitch variations are used in both languages to contrast lexical meanings, the possible effect of pitch on the CV coarticulation has also been examined. Our results have been, once again remarkably close to what were reported in previous studies, no effect of lexical tones is observed for Mandarin Chinese (time lag 1 and 2), and only time lag 2 reports an effect of the word accent for Swedish. CV time lag 2 includes the TB front-back dimensions, possibly yielding the observed difference between word accents, as previously reported in [3]. The use of CV time lag 3 was motivated by the missing standard measure of articulatory data. Although the acceleration-based measures did not show any significant difference between Swedish and Mandarin, only CV time lag 3 displayed differences between both the word accents and the lexical tones.

One surprising factor that was significantly associated with the CV timing is gender. For both languages, and all measurement types, speakers' gender was proven to affect the temporal lag in the same direction. However, when the speaker variation was taken into account, the gender differences were not as obvious. Only on the velocity-based measure (CV time lag 1), we found different patterns of co-articulation between Mandarin and Swedish. This suggests that CV coordination between the two languages only differ when comparing male and female speakers separately. The rather unexpected findings on gender could be due to the anatomical and physiological features of the speakers. There are obvious biological differences between sex, where generally speaking, male speakers have bigger head size than female speakers, thus they should also have bigger vocal tracts. Because the larger vocal tract is closely correlated with the velocities of tongue in a positive manner [14], subsequently, the shorter CV temporal lags observed with male speakers may be associated with the greater velocity of tongue body (responsible for vowel). Moreover, compensatory articulations as a consequence of anatomically differences [15], linked to biological differences between sex, or socially constructed roles of gender, could also play a part in gender-specific articulation (see e.g. [16]). Clearly this needs more investigation to establish the role that speakers' gender is playing when comparing CV coarticulation patterns.

CV time lag 1 is based on equivalent landmarks on velocity of LA and TB, which yield not only the largest gender difference, but also the longest time lags of the three measures. We presume the longer time lags are due to the movement characteristics of the lips and tongue (e.g., the lips being smaller and faster), notably, a difference not as evident for the acceleration landmarks in CV time lag 3. Peak acceleration (equated with adding force to the movement) of both lips and tongue (time lag 3) does not only display shorter time lags, but also, as already mentioned, differences between tones/accents in both languages. Although non-equivalent landmarks (time lag 2) can be motivated by different linguistic functions, and movement characteristics, of the lips and the tongue, both measures on equivalent landmarks seem to reveal significant differences between groups: the velocity-measured (time lag 1) differs between gender and language; the acceleration-based (time lag 3) between tones and word accents. However, such a hypothesis, linking systematic tendencies in CV timing to movement characteristics, needs to be confirmed with bigger sample size, specifically of the Mandarin data.

In summary, our results provide first evidence that the word-initial consonant and vowel may be coarticulated in a similar manner in typologically different languages. For the first time, the present data also show that speakers' gender is most likely to affect the CV coarticulation. However, this study has some limitations: besides the small sample size and un-balanced data sets, it does not consider the potential influence of the prosodic environment, which is due to the limited understanding of this variable in studies of articulatory coordination.

5. ACKNOWLEDGMENTS

This work was supported by an International Postdoc grant from the Swedish Research Council (Grant No. 2021-00334), and by an infrastructure grant from the Swedish Research Council (SWE-CLARIN, 2018–2024; Grant No. 2017-00626). The authors gratefully acknowledge the Lund University Humanities Lab.

6. REFERENCES

[1] Löfqvist, A., Gracco, V. L. 1999. Interarticulator programming in VCV sequences: Lip and tongue movements. *J. Acoust. Soc. Am.* 105(3), 1864–1876.

[2] Gao, Man. 2008. *Tonal alignment in Mandarin Chinese: An articulatory phonology account*. New Haven: Yale University Doctoral dissertation.

[3] Svensson Lundmark, M., Frid, J., Ambrazaitis, G., Schötz, S., 2021. Word-initial consonant–vowel coordination in a lexical pitch-accent language. *Phonetica* 78(5-6), 515–569.

[4] Mücke, D., Nam, H., Hermes, A., Goldstein, L. M. 2012. Coupling of tone and constriction gestures in pitch accents. In Hoole, P., Bombien, L., Pouplier, M., Mooshammer, C., Kühnert B. (eds.), *Consonant clusters and structural complexity.* Mouton de Gruyter, 205–230.

[5] Niemann, H., Grice, M., Mücke, D. 2014. Segmental and positional effects in tonal alignment: An articulatory approach. *Proc. 10th ISSP* Cologne, 285–288.

[6] Svensson Lundmark, M. 2023. Rapid movements at segment boundaries. *J. Acoust. Soc. Am.* 153 (3).

[7] Boersma, P., Weenink, D. 2018. Praat: Doing phonetics by computer [Computer software]. Version 6.0.37. Retrieved 3 February 2018 from http://www.praat.org/.

[8] R Core Team. 2015. R: A language and environment for statistical computing [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/.

[9] Machač, P., Skarnitzl, R. 2009. *Principles of phonetic segmentation*. Epocha.

[10] Wieling, M., Tiede, M. 2017. Quantitative identification of dialect-specific articulatory settings. *J. Acoust. Soc. Am.* 142(1), 389–394.

[11] Bates, D., Maechler, M., Bolker, B. M., Walker, S. C. 2015. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67(1). 1–48.

[12] Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. 2017. lmerTest package: Tests in linear mixed effects models. *J. Stat. Softw.* 82(13), 1–26.

[13] Zhang, M., Geissler, C., Shaw, J. 2019. Gestural representations of tone in Mandarin: Evidence from timing alternations. *Proc. 15th ICPhS* Melbourne, 1803–1807.

[14] Kuehn, D. P., Moll, K. L. 1976. A cineradiographic study of VC and CV articulatory velocities. *J. Phon.* 4(4), 303–320.

[15] Engwall, O., Delvaux, V., Metens, T. 2006. Interspeaker variation in the articulation of nasal vowels. *Proc. 7th ISSP* Ubatuba, 3–10.

[16] Foulkes, P., Docherty, G. 2006. The social life of phonetics and phonology. *J. Phon.* 34(4), 409–438.