VOWEL COUPLING IN MANDARIN SYLLABLE CONTRACTION

David Cheng-Huan Li

University of Southern California, USA
lidc@usc.edu

ABSTRACT

Syllable contraction involves reducing a sequence of two or more syllables to fewer. The present study examines the phenomenon under the framework of Articulatory Phonology in which speech is composed of gestures. The experiment reported here tests the hypothesis that Mandarin contracted syllables result from a transition from anti-phase to the more stable in-phase coordination between vowel gestures of successive syllables. This is tested by comparing the F2 of disyllabic function words, which are more likely to be contracted, to non-function words. Results indicate that the hypothesized switch in coupling mode can explain the phenomenon of syllable contraction in Taiwan Mandarin.

Keywords: syllable contraction, articulatory phonology, coupling mode, language production

1. INTRODUCTION

Among many Chinese dialects, there is a special type of pronunciation or segmental reduction that occurs frequently in spontaneous speech. The phenomenon involves a sequence of two or more syllables reduced to a lesser number of syllables and has subsequently been known as "syllable contraction" e.g. [3, 4].

1.1. Phonology, articulation, and syllable contraction

There have been several proposals that examined the phenomenon of Chinese syllable contraction. Chung (1997) proposes the edge-in model to account for the phenomenon and suggests that contraction occurs according to the rule stated in (1). In this model, the left edge of the first syllable and the right edge of the second syllable are kept in the contracted syllable. Crucially, the result of contraction should adhere to the structure of CVX and the target of syllable contraction must obey the rules of Chinese phonotactics.

(1) syllable contraction CVX + CVX => CVX

Offering a different perspective on syllable contraction, Cheng and Xu [2] argue that the underlying mechanism of syllable contraction is the gradient undershoot of the articulatory target. Using nonsense Mandarin disyllabic sequences, they demonstrated that duration is the most determining factor for contraction. Cheng and Xu argue that when duration of articulation decreases beyond a certain threshold due to time pressure, the execution of the articulatory movement toward a consonantal target eventually becomes difficult.

1.2. Consonant coda [m] in Mandarin

Despite the explanatory power of the edge-in model and gradient undershoot hypothesis, both approaches remain unsatisfactory. As Tseng [10] noted in her corpus study, she observed many contracted syllables that did not obey the rule in (1). That is, while the consonant [m] is never allowed in the coda position in Mandarin, many contracted syllables in her study have a clear coda consonant [m]. Similarly, Cheng and Xu's account also does not predict the presence of such contracted forms. To help explain the presence of these forms, the current study aims to investigate the phenomenon from a different perspective—Articulatory Phonology.

1.3. Articulatory phonology and syllable contraction

In Articulatory Phonology, gestures are the basic units of combination [1]. According to this view, speech is composed of an ensemble of gestures overlapped temporally, which are coupled by a system of planning oscillators, or clocks [5]. In this model, the oscillators would entrain to two stable modes of inter-gestural coordination, in-phase—i.e. occurring simultaneously, and anti-phase—i.e. occurring sequentially.

While the both the in-phase and anti-phase mode are intrinsically accessible, research has suggested that an in-phase mode is more stable than an anti-phase one. As seen in finger tapping experiments (e.g. [7]) and speech studies (e.g.[6, 8]), in-phase coupling remains stable as the rate of production increases but anti-phase coupling

becomes unstable and spontaneously switch to an in-phase mode.

The phenomenon of syllable contraction can be analyzed in terms of change in coupling relations. For instance, when the Mandarin disyllabic word, 'we', is not contracted, its pronunciation, [uomɤn], may be represented gesturally as in Figure 1. In contrast, when the word is contracted, gestures may become more in-phase with respect to each other. In cases of extreme contraction as shown in Figure 2, gestures for [m] in [mɤn] (velum opening, bilabial stop) may become in-phase with the gestures for [n] (velum opening and alveolar stop). When this occurs, the fact the gestures for lips and tongue tip overlapping at the same time produces a sound that is perceived as an [m], resulting in the sound [uom] for 'we' (as attested in the corpus study of [10]). Crucially, this account of in-phase coupling can explain the presence of consonant coda [m] in Mandarin.

Figure 1: Gestural representation for non-contracted 'we'.

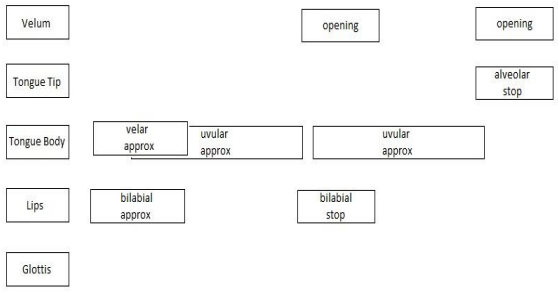
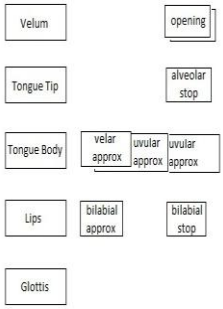


Figure 2: Gestural representation for contracted 'we'.



To test whether change in gestural coordination occurs in contracted syllables of Taiwan Mandarin, this study examined the coordination of vowel gestures in function words, which more likely to be contracted (e.g. [9]), to non-function words. It is predicted that the successive vowels of a disyllabic function word are more in-phase with respect to each other, even if they are not reduced to a single contracted syllable. If the word contains a front and a back vowel (or vice versa) in the successive syllables, the strength of in-phase coupling will determine the amount of coarticulation of the back and front vowel, which is measurable via a change in the second formant (F2) frequencies. Therefore, F2 frequency values are predicted to decrease compared to the non-contracted form if a disyllabic word with the vowel backness order is front-to-back, and F2 frequency to increase if back-to-front.

2. METHODS

2.1. Subjects

Four native speakers (1 male and 3 female) of Mandarin from Taiwan with no hearing or language impairment participated in the study.

2.2. Stimuli

Seven sets of mostly disyllabic words were selected (for details see Appendix). There are two word types within each set: contracted candidates and non-contracted candidates. Contracted candidates are either conjunctions or adverbs and non-contracted candidates were either verbs or nouns. The backness of successive vowels in the disyllabic word is either (a) front-to-back or (b) back-to-front.

To tease apart the effect of coarticulation from possible undershoot due to temporal truncation, there are two controls within each set: words that have vowels that do not differ in backness in the successive vowels. Therefore, there are four words in each set: (1) contracted target, (2) non-contracted target, (3) contracted control, and (4) non-contracted control.

All targets and controls have the form of (C1)V1(C2)V2 (where C1 and C2 are optional). All V1's within a set are identical, and C1's and C2's are maximally matched. The targets and controls are placed in the middle of a carrier sentence at the beginning of a prosodic phrase.

Each carrier sentence is always the second part of a dialogue. An example with the target *chufei*, 'unless', is illustrated in (2). Each stimulus is set up to be like a conversation between two individuals to simulate a real conversation.

(2) ni gen ni pengyou dasuan sheme shihou yao hui jia
"When do you plan on going home with your friends?"
wo deng bu xiaqu xiang yao xian zou **chufei** bisai keyi mashang jieshu
"I don't want to wait unless the game will end soon."

2.3. Procedure

Subjects were shown one stimulus at a time on a computer screen. They were allowed to rehearse the stimulus and were instructed to pretend they were the person providing the second part of the dialogue. The first part of the dialogue was played back to them, and then they began reading the second part of the dialogue. This process of playback and sentence production was repeated six times (or more should production errors occur).

2.4. Acoustic measurements

The acoustic measurements made included second formant (F2) frequency values at the midpoint of V1 and the durations of syllables. Durations of the contracted syllables is calculated by combining durations of (C1)V1(C2)V2, where applicable. The following criteria are used for labeling by Praat:

1. Consonants onset: If the consonant is a fricative, the onset was defined as the first appearance of aperiodic noise (> 2500 Hz). When it is not clear where the aperiodic noise begins, the onset was defined by the onset of a 10 ms frame during which a decrease of 150 Hz in F2 occurred. If the consonant is a stop or affricate, the onset was defined by the substantial decrease in amplitude. Set 7 was excluded from further analysis due to problems in segmentation involving [x].

2. Consonant offset: Regardless the consonant, the offset of the consonant is defined as the onset of the vowel.

3. Vowel onset and offset: The onset and offset were defined as the onset and offset of periodicity.

4. V1 midpoint: The point half way between the onset and offset of V1 was the midpoint

3. RESULTS

3.1. Durations

Since contracted syllables are expected to be shorter, contracted targets and controls should be of shorter durations. A RM ANOVA was performed on subject means of syllable duration for Contraction (contracted vs. non-contracted). As expected, the test revealed a main effect of Contraction ($F(1, 3) = 16.13$, $p < .05$).

3.2. Contracted vs. non-contracted target F2

If the vowels in the targets are coarticulated due to in-phase coupling, F2 is expected to decrease in the contracted form if the backness of the successive vowels is front-to-back and the F2 to increase if back-to-front.

When a RM ANOVA was conducted, there was an significant interaction of Contraction and Vowel Order (back-to-front vs. front-to-back; $F(1, 3) = 15.04$, $p < .05$), with more coarticulation in the contracted form.

3.3. Target vs. control F2

Although there was an interaction of Contraction and Vowel Order, the effect observed could be a mere difference between contracted and non-contracted targets but not due to *coarticulation*. To investigate its effect, targets need to be compared with controls that vary in the same way in terms of contraction but do not contain vowels that would lead to F2 changes due to coarticulation.

Figure 3a: Average F2 frequency values of targets and controls when vowel backness is back-to-front.

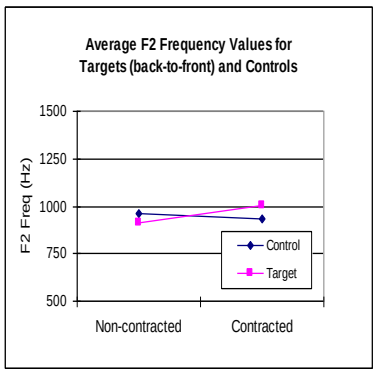


Figure 3b: Average F2 frequency values of targets and controls when vowel backness is front-to-back

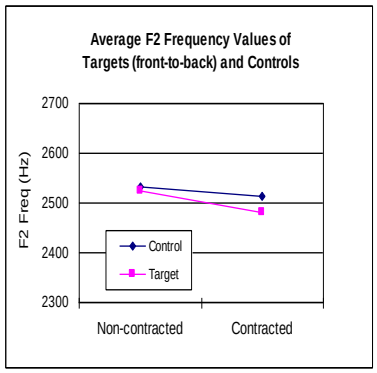


Figure 3 illustrates the comparison of targets to the controls. When the successive vowel backness is back-to-front, notice that F2 for targets increased in the contracted form while F2 for controls decreased. A RM ANOVA resulted a marginal interaction of Contraction and Vowel Environment (targets vs. controls; ($F(1, 3) = 7.42$, $p = .07$). When the successive vowel backness is front-to-back, F2 for

targets appeared to decrease to a greater extent compared to the controls. A RM ANOVA resulted in a significant interaction ($F(1, 3) = 18.42$, $p < .05$).

3.4. Within set comparison

A within-set comparison of targets and controls could help explain the source of variance. RM ANOVAs showed that there was a significant interaction in Set 1, Set 2, Set 4, and marginal in Set 5 while no effect was found in Set 3, and Set 6.

4. DISCUSSION

The analysis presented in this paper provides the initial evidence that disyllabic contractions in Taiwan Mandarin may involve change in coupling modes as hypothesized. This study first established that the contracted syllables used in the study were significantly shorter than the non-contracted syllables. Then, when the vowel backness order was front-to-back, the trend was that, among targets, F2 was lower in the contracted form. The reverse trend was also true when the vowel order was back-to-front.

To alleviate concerns that the change in F2 values was due to contraction but not coarticulation, targets were compared to controls that had successive vowels that are identical in backness. Statistical analyses revealed that the targets with vowel order of front-to-back had a significant change in F2 values due to coarticulation while the effect was marginal for targets with vowel order of back-to-front. To sort out the source of variance, individual sets of data were also analyzed: three (Set 1, 2, 4) out of six sets of data analyzed had a significant change of F2, one set (Set 5) had a marginal change.

The remaining two sets (Set 3, 6) did not have a significant change of F2. This may be due to too much noise with the small sample, but it may also reflect the fact that even at a faster rate of production, gestural reorganization does not always occur. Even though in-phase coupling is more stable, other factors, such as the word frequency, may exert additional influence on inter-gestrual coordination. Moreover, the choice of consonant for one of these two sets, Set 6, may also contribute the values of F2. In Set 6, the contracted target is [it^{h}ou] while the non-contracted target is [pit^{h}ou]. The F2 value at the mid-point of the vowel may be influenced by the presence of the labial gesture especially when the vowel gesture is shortened due to contraction.

In summary, the present study finds that the coarticulation pattern in Mandarin syllable contraction is consistent with the hypothesis of change in coupling mode. This account has the power to explain attested patterns that were not accounted for by other proposals, such as the presence of nasal coda [m]. As the present study suggests, syllable contraction in Mandarin is the consequence of change in coupling mode because of the stability of the in-phase relations.

5. REFERENCES

[1] Browman, C., Goldstein, L. 1995. Dynamics and articulatory phonology. In Port, R., van Gelder, T. (eds.), *Mind as Motion: Dynamics, Behavior, and Cognition*. Cambridge: MIT Press, 175-193.

[2] Cheng, C., Xu, Y. 2009. Extreme reductions: Contraction of disyllables into monosyllables in Taiwan Mandarin. *Interspeech*, 456-459.

[3] Cheng, R.L. 1985. Sub-syllabic morphemes in Taiwanese. *J. Chinese Linguistics* 13(1), 12-43.

[4] Chung, R.F. 1997. Syllable contraction in Chinese. *Proc. 3rd Conf. on Chinese Languages and Linguistics*, 199-235.

[5] Goldstein, L., Byrd, D., Saltzman, E. 2006. The role of vocal tract gestural action units in understanding the evolution of phonology. In Arbib, M. (ed.), *From Action to Language: The Mirror Neuron System.* Cambridge: Cambridge University Press, 215-249.

[6] de Jong, K.J., Lim, B.J., Nagao, K.. 2002. Phase transitions in a repetitive speech task as gestural recomposition. *IULC Working Papers*, online 2.

[7] Kelso, J. 1984. Phase transitions and critical behavior in human bimanual coordination. *Am. J. Physiology – Regulatory, Integrative, and Comparative Physiology*, 246(6), R1000-R1004.

[8] Parrell, B. 2009. Rate conditioned variability in Western Andalusian Spanish aspiration. *J Acoust. Soc. Am.* 125(4), 2570.

[9] Tseng, S.C. 2001. Highlighting utterances in Chinese spoken discourse. *Proc. 5th Pacific Asia Conf. on Language, Information and Computation*, 163-174.

[10] Tseng, S.C. 2005. Contracted syllables in Mandarin: Evidence from spontaneous conversations. *Language and Linguistics* 6(1), 155-182.

6. APPENDIX

	Type 1	Type 2	Type 3	Type 4
Set 1 IPA	pupi	upi	pukuo	fukuo
Set 2 IPA	tʂ^{h}ufei	t^{h}ufei	tʂupu	tʂup^{h}u
Set 3 IPA	xɤpi	xɤp^{h}it ɕu	xɤk^{h}u	xɤku
Set 4 IPA	iʂou	iʂou	itsai	it^{h}saitɑʊ
Set 5 IPA	itu	it^{h}u	i ɕ	i ɕkuən
Set 6 IPA	it^{h}ou	pit^{h}ou	it ɕ	it ɕ
Set 7 IPA	t ɕxu	t ʂ^{h}ixu	it ʂ^{h}i	it ʂ^{h}i

Type 1= Contracted target Type 2 = Non-contracted target
Type 3= Contracted control Type 4 = Non-contracted control