Duration Reflexes of Syllable Structure in Mandarin

Fei Wu (Shanghai International Studies University) & Michael Kenstowicz (MIT)

(final manuscript of paper published in Lingua 164 (2015) 87-99)

1. Introduction

The structure of the Standard Chinese (Mandarin) syllable has been a subject of some controversy ever since the introduction of formal models of this prosodic constituent. Major questions concern the presence or absence of duration contrasts based on the number of segmental constituents, the affiliation of the prenuclear vocoid/glide, and the existence of onsetless syllables. We focus on the first two issues in this paper and leave the third for future research.

As a canonical monosyllabic language, most free-standing morphemes in Mandarin are composed of a single syllable (though there are clear preferences for disyllables in certain prosodic contexts; see Feng 2001, Duanmu 2012). There are no initial or final consonant clusters. Morphemic shapes include CV, CVN, CVG, CGV, CGVG, and CGVN, where C = consonant, G = glide /j, w, ɥ/, N = /n, ŋ/, V = /i, y, u, ə, a/, the latter two with several allophonic realizations. Given that each segment occupies some timing slot in the sense of Clements and Keyser (1983), we expect the following durational hierarchy, other things being equal: CV < CVG, CVN, CGV < CGVG, CGVN. However, CV syllables in Mandarin are typically analyzed as bimoraic, with a long vowel (Duanmu 2007:82). A number of factors motivate this analysis. First, the vowel of *bā* 'eight' sounds longer than the one in *bān* 'class'. Second, CV syllables fully contrast the four tones parallel to the segmentally more complex CVN, CVG, CGV, as shown by the famous suite of *mā* 'mother', *má* 'hemp', *mǎ* 'horse', *mà* 'scold' compared to *pēng* 'impulsive', *péng* 'friend', *pěng* 'to clasp', *pèng* 'to touch', and to *bāo* 'to cover', *báo* 'thin', *bǎo* 'jewel', *bào* 'to announce'. Finally, many languages impose a bimoraic minimality requirement on free-standing prosodic words and so it would not be surprising that Mandarin does as well.

Scholars such as Duanmu (2000, 2007) treat the contrast of toneless vs. tone-bearing syllables in Mandarin as a stress difference. Since each free-standing morpheme is tone-bearing, it must also be stressed under this view. Moreover, in many languages stressed syllables are bimoraic (cf. OT's Stress-to-Weight constraint, Kager 1999:268) and so this factor provides additional motivation for equating CV with CVN, CVG, CGV, etc. prosodically. In his discussion of the bimoraic syllable, Duanmu (2007:90) remarks "one would expect that all heavy syllables have similar durations." He goes on to say that "This seems to be the case in controlled environments, such as read speech in a carrier sentence". The principal sources cited are Howie (1976) and Lin and Yan (1980). The former study is based on the speech of a single individual (136 words in a carrier sentence) while the latter focused on unstressed syllables. Hence, a more systematic exploration of the durational properties of the various

Mandarin stressed syllables employing multiple speakers is warranted in order to corroborate the bimoraic status of CV syllables.

Another phonological contrast that has a significant phonetic duration correlate is tone. Various studies have found that syllables with tone 3 are longest. Duanmu (2007:90) cites Woo's (1969) finding that "Standard Chinese syllables with the third tone have an extra tone feature H in prepausal positions, and their average rhyme duration is about 50% longer than that of nonfinal syllables." As far as we can determine, Woo's remark was based on a few data points taken from her own speech. Published studies with more information on this point include Ho (1976) and Xu (1997). Both of these focused on the co-articulatory effects of the four tones in sentential contexts but also report duration measures for isolation forms. Ho's study included a substantial number of CV words from five Beijing speakers while Xu's was restricted to six repetitions of the *mā*, *má*, *mǎ*, *mà* suite produced by eight male Beijing speakers. The average syllable durations are shown in (1). The numbers from Ho's study are estimated based on the first chart shown in this paper.

(1) Average syllable durations (ms) under the four Mandarin tones in two previous studies

	Tone 1 (55)	Tone 2 (35)	Tone 3 (314)	Tone 4 (51)
Ho (1976)	260	300	400	195
Xu (1997)	247	273	349	214

So here as well a study with a greater variety of syllable shapes is called for in order to better determine the durational profiles of the four-way tonal contrast.

Another, more vexed question has concerned the syllabic affiliation of the prenuclear glide. Virtually every conceivable attachment has been proposed based on different types of evidence including judgments of syllable rhyme, speech errors, language games, loanword adaptation, as well as phonotactic restrictions. They include treating the glide as the first portion of a nuclear diphthong (Wang and Chang 2001), as a sister constituent of the rhyme parallel to the nuclear vowel and coda (Chao 1968), as a specifier c-commanding the entire rhyme (van de Weijer and Zhang 2008), as the immediate constituent of a flat syllable structure with no internal grouping (Yip 2003), or as a component of the onset (Bao 2000, Lin 2002). See van de Weijer and Zhang (2008) for a recent review of these analyses. Perhaps most controversial is Duanmu's (2000, 2007) proposal that the medial glide is realized as a secondary articulation on the onset consonant. One of his arguments (p. 80) is that "the presence of G does not increase the length of the syllable in any appreciable way. Rather, CVX and CGVX are more or less similar in duration." Again Howie's (1976) study with one speaker is cited as support for this analysis. Also, van de Weijer and Zhang (2008) mention an unpublished acoustic study by Lu (2005) who is reported to have found that CGVX syllables were reliably longer than CVX by an average of (only) 7 ms, $p = 0.012$. They do not indicate the range of X in the stimuli nor how many speakers were recorded or how many data points were collected in Lu’s (2005) experiment. They take this difference as a challenge to Duanmu’s secondary articulation analysis of the glide.

A final instance where appeal to phonetic duration has been crucially invoked in the analysis of Mandarin is Hsieh’s (2012) treatment of the fronting and raising of the low vowel in syllables with the underlying shape /jan/ and /Cjan/, which are realized as [jen] and [Cjen]: e.g. /ján/ -> [jén] ‘salt’, /tjān/ -> [tjēn] ‘sky’. This low-vowel raising process is found in a number of other Sinitic languages besides Mandarin. What is puzzling about this process (or an equivalent phonotactic constraint *Cjan) is that both the preceding palatal glide and the following coronal nasal must be present in order for the change (or restriction) to occur: cf. /tān/ -> [tān] ‘to carry on shoulders’ and /jà/ -> [jà] ‘duck’. Inspired by Flemming’s (2001, 2003) analysis of the fronting of back vowels between coronal consonants in Cantonese, Hsieh sees low-vowel raising as motivated by two factors: undershoot of the low vowel target under presumed time pressure in the closed syllable environment combined with assimilation to the front tongue-body position of the palatal glide. Crucial to this analysis is the presumption that /a/ is significantly shorter in a closed syllable and that as a result the vowel is articulated with a higher tongue-body position. As evidence for the latter point, Hsieh cites Lee and Zee’s (2003) IPA sketch of Standard Chinese, who report that “in closed syllables [a] = [a̝]” based on the speech of one 26 year old female speaker from Beijing. Since no other duration data on Mandarin were available, Hsieh (2012) references his own unpublished study of 10 male speakers of Taiwanese Southern Min and Zee’s (2003) study of Cantonese, both of which find substantial shortening of the nuclear vowel in a closed syllable, as seen in (2).

(2) Vowel duration (in ms) in various syllabic contexts in two Chinese languages

		Taiwanese Southern Min	Cantonese
Long	CV:	210	350
Half-long	CV.N	133	200
	VV.T		200
Short	CVT	92	120

Thus, there is ample motivation for a comparable investigation of the duration of the nuclear vowel in open vs. closed syllables of Mandarin and its effect on the realization of the low vowel.

A related question concerns the effects of the prenuclear glide and the coda consonant on the relative backness of the nuclear vowel. It is well known that the low vowel has a relatively front allophone before the coronal coda [n] and a back allophone [ɑ] before the velar nasal [ŋ] (Mou 2006, Lin 2007:153). Lin (2002) discusses the relative strength of the prenuclear and postnuclear glides on the realization of the mid vowel archiphoneme in various Mandarin dialects, noting than in Standard Mandarin the coda glide predominates: /tuəi/ ‘correct’ > [twej] and /tiəu/ ‘throw’ > [tjow]. In the realization of /Cuan/ syllables there is a conflict between the fronting effect demanded by the coda [n] and the backing effect of the prenuclear glide [w] comparable to the conflict Lin observes for the mid vowel nucleus. Will the predominance of the coda seen in the realization of mid-vowel CGVG syllables carry over to low-vowel /Cuan/ syllables? To the best of our knowledge, this question has not been addressed.

In sum, analyses of four different aspects of Mandarin phonology centered on the syllable have relied on evidence from their phonetic duration correlates. But the supporting phonetic studies typically are based on the speech of a single speaker and employ different methods of data collection and analysis. There is thus good motivation to conduct a more controlled investigation with a larger number of speakers and syllable types to see to what extent the claims reviewed above are supported by empirical evidence. Accordingly, our study addresses the questions detailed in (3).

(3)

a. Will the duration of the syllable remain stable as its structure is changed from CV to CVN, CGV, and CGVN as predicted by the bimoraic hypothesis?
b. Will the syllable duration hierarchy of tone 3 > tone 2 > tone 1 > tone 4 reported by the studies referenced in (1) for CV hold across the different syllable shapes?
c. If the Mandarin syllable has a more or less constant duration, how will the duration of the nuclear vowel differ as the syllable shape is altered from CV to CVN, CGV, CGVN?
d. Will there be a difference between the effects of a pre-nuclear glide vs. a post-nuclear coda on the duration of the nuclear vowel that could shed light on the structural position of the glide in the syllable?
e. Will the F1 value of the nuclear vowel that reflects its height decrease as the syllable structure is changed to the maximal CGVN, as assumed in the undershoot analysis of /tjān/ -> tjēn?
f. In the /Cuan/ maximal syllable will the front vowel allophone demanded by the coda [n] predominate over the coarticulatory backing of prenuclear [u]?

2. Materials and Methods

In order to address these questions, we constructed a corpus of 48 monosyllabic words of Standard Chinese. The words in the corpus were balanced for syllable type (CV, CVN, CGV, CGVN), tone (1, 2, 3, 4), and segment. The vowel, coda nasal, and prenuclear glide were held constant as [a], [n], and [u], respectively. Onset consonants were primarily drawn from the voiceless unaspirated stops (pinyin b, d, g). But since the Mandarin lexicon contains many gaps in the combination of tones and segments (see van de Weijer 2012 for recent discussion), some additional onsets were added, in lieu of introducing nonsense syllables, which we felt would be too problematic for our speakers to associate with the proper tone. See the Appendix for the corpus. We recruited five female Standard Chinese speakers ranging in age from 23 to 35. Two were from Shanghai, one was from Hunan province, one was from Hubei province, and one was from Taiwan. The speakers pronounced a randomized list containing each target word in Chinese script in isolation and then in a frame sentence *Tā zǒng shì bǎ ____ dú cuò* 'He/she is always reading X incorrectly.' Each speaker recorded the entire list of 48 isolation plus sentence-framed words twice in succession in a sound-insulated recording booth, resulting in a total of 960 words (48 words x 2 contexts x 2 trials x 5 speakers). The speakers were paid a nominal fee for their participation in the experiment. The recordings were made with a head-mounted Shure SM10A Unidirectional Head-Worn Dynamic Microphone and USB Pre 2 Preamp at a sampling rate of 44.1 kHz, 16 bits.

Our data was analyzed using Praat version 5.3.39 (Boersma and Weenink 1992-2013). Text grids were assigned for the entire target word and its individual segments. Segmentation was based on a comparison of the waveform and the formant displays in the spectrograms. For CGV and CGVN words, the glide portion almost always had a steady state F1 followed by a rise to the steady state for the nuclear vowel (see an example in Figure 1). Accordingly, we segmented the glide into a vocoid [u] portion followed by a GV transition, starting from when F1 began to climb in the spectrogram to when it approached the plateau for the nuclear vowel [a]. Praat scripts were used to collect the duration of each segment as well as the entire word. These duration measures were normalized with Z-scores in order to accommodate individual differences in speech rates. We also retrieved F1 and F2 measures for the mid point of the vowel and the prenuclear glide. They were normalized on the bark scale. Regression analyses and plots were produced in R (Bates and Maechler 2013).

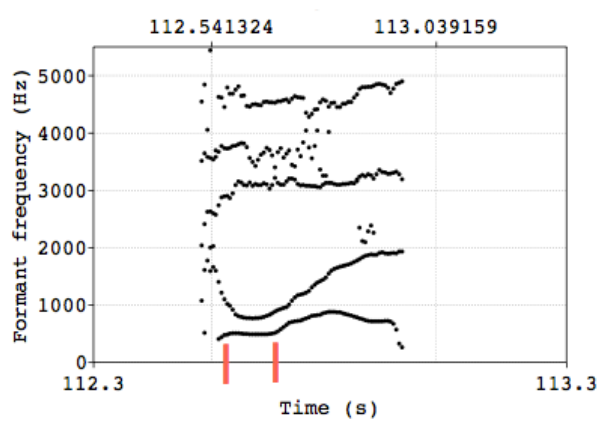

Figure 1 Praat drawing of vowel formants in duān 'hold' with bars showing [u] vocoid

3. Results

3.1 Word Duration

The bar charts in Figure 2 show the duration of the words as a function of syllable shape and tone. Syllable duration appears to align with the number of phonological segments, with CV syllables shorter than CVN and CGV, which in turn are shorter than CGVN. The tonal effect on duration mirrors the order found by the studies mentioned in (1) above: 3 > 2 > 1 > 4.

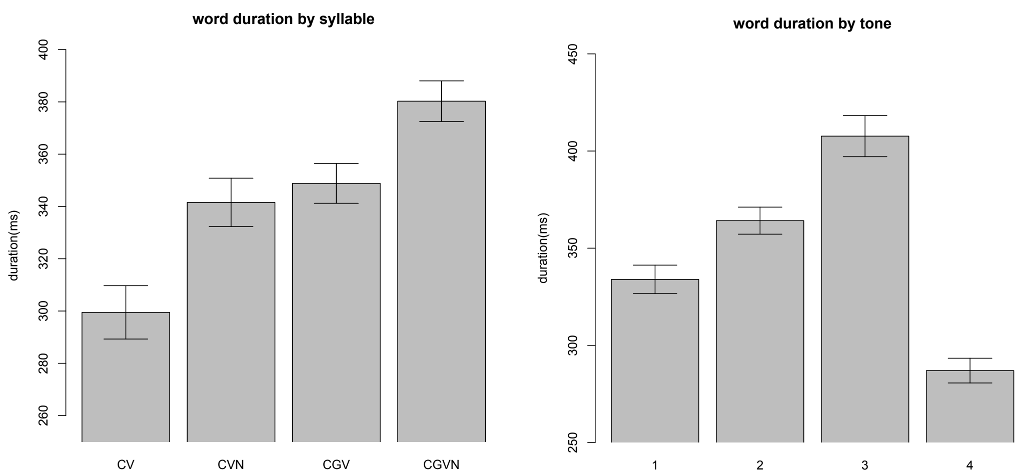

Figure 2 Mean duration of target words grouped by syllable type (left) and tone (right). Error bars are 95% confidence intervals.

In order to assess the statistical significance of these differences, a mixed-effects linear regression analysis was run in R. To accommodate individual differences in speech rate, word durations were normalized as Z-scores based on the mean and standard deviation for each speaker in each of the two contexts: word in isolation and word inside the carrier sentence. Speaker, word, and trial (repetition) were set as random factors, while syllable shape (CV, CVN, CGV, CGVN) and tone as well as their interactions were the predictor variables, and normalized word duration was the dependent variable. The variables were coded with CV and tone 1 as the baseline. The test returned the results in (4). The p values were calculated by R's pvals.fnc with MCMC.

(4) Results of a mixed-effects linear regression

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.78736	0.31934	-2.466	-2.466

syllCVN	0.24066	0.24069	1.000	0.1755
syllCGV	0.41162	0.23743	1.734	0.0831
syllCGVN	0.65005	0.22294	2.916	0.0036
tone2	0.13810	0.10599	1.303	0.1927
tone3	0.71947	0.10599	6.788	0.0000
tone4	-0.53842	0.10599	-5.080	0.0000

Relative to the baseline CV tone 1, only the CGVN syllable shape was significantly longer. This continued to be the only significant difference under multiple comparisons among the syllable shapes using Tukey's HSD and also held true when the data was broken down into isolation versus sentential contexts. We can thus conclude that it is legitimate to claim that CV syllables are not significantly different in duration from both CVN and CGV, a finding that is consistent with the bimoraic hypothesis reviewed in section 1. Also, CGVN syllables exceed CVN syllables by c. 40 ms, a much greater difference than the 7 ms. Lu (2005) reports for CGVX vs. CVX.

As far as tone is concerned, syllables with tone 3 were significantly longer and those with tone 4 significantly shorter relative to the baseline with tone 1. This effect was robust both in magnitude and reliability. There was only one significant interaction (between CVN and tone 2, $\beta = .65$, $t = 2.54$, $p = 0.0111$). This result replicates and generalizes to other syllable shapes the findings of Ho (1976) and Xu (1997) mentioned in (1) above for CV syllables. The table in (5) reports the means and standard deviations as well as the ratios of the syllable durations relative to the four tones in different contexts: overall, isolation, and frame sentence.

(5) Durations in ms of the four Mandarin tones in different contexts (mean, st dev, ratio)

Context	Tone 1	Tone 2	Tone 3	Tone 4
overall	333 (107)	364 (100)	407 (162)	286 (88)
	.81	.89	1.0	.70
isolation	390 (90)	413 (66)	497 (123)	313 (87)
	.78	.83	1.0	.62
sentence	276 (93)	312 (104)	316 (146)	259 (81)
	.87	.98	1.0	.81

It is interesting that the duration difference between tones 2 and 3 essentially disappears in the sentential context. This is presumably because the final rise of tone 3 is truncated or missing in the sentential context, as observed by most analyses of Mandarin tone (e.g. Yip 2002). Related to this point are the findings by Shih (1989) and Fon (1999) that in the Taiwanese variety of Standard Chinese (Guoyo) the final rise of tone 3 is either missing or compressed in isolation forms. Since one of our subjects is a speaker of this dialect, we were curious to see if this effect showed up in her speech. The table in (6) shows that this is indeed the case. For this speaker's speech, the durational difference between tones 2 and 3 disappears in the isolation form but both still remain significantly longer than tone 4. The word durations overall are shorter compared to the Mainland speakers and may reflect a stress difference.

(6) Syllable duration in ms (st dev) and ratios of four Mandarin tones in isolation form for Mainland vs. Taiwanese speakers

Dialect /tone	Tone 1	Tone 2	Tone 3	Tone 4
PRC	394 (79)	422 (66)	505 (97)	315 (82)
	.78	.83	1.0	.62
Taiwan	332 (62)	376 (45)	367 (62)	263 (68)
	.90	1.02	1.0	.71

Another point worthy of mention is that the falling tone 4 has the shortest duration. While falling tones are cross-linguistically typically shorter than rising tones (Ohala 1979, Sundberg 1979, Silverman 1997, Zhang 2004:158), the contrast with level tone 1 is striking since other things being equal contour tones are typically longer than level tones. The motivation for this difference remains unexplained as far as we know. Tone 4 does not seem to be accompanied by a voice quality difference that would compensate for its short duration.

As observed by a reviewer, the onset consonants in our corpus include continuants, whose duration was found by Zhuo (1986) to exceed that of stops, and which thus may introduce an experimental confound. In order to address this point, we ran the same regression analysis as in (4) over the 29 words in our corpus with /b, d, g/ onsets, with the results in (7). The overall patterns of significance among the factors and their levels are comparable to those in (4). The coefficients are smaller suggesting that stop onsets are indeed shorter than continuants. Also, the higher t-values provide a tighter fit for the syllable shape factor.

(7) Results of a mixed-effects linear regression on stop-initial words

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.73999	0.21511	-3.440	0.0006
syllCVN	0.12957	0.09087	1.426	0.1544
syllCGV	0.18342	0.09241	1.985	0.0476
syllCGVN	0.40163	0.09423	4.262	0.0000
tone 2	0.17394	0.11689	1.488	0.1372
tone 3	0.53049	0.08835	6.005	0.0000
tone 4	-0.44522	0.08520	-5.225	0.0000

3.2 Nucleus duration

Given that there are no significant differences in the duration of CVN and CGV in comparison to CV, we can ask what are the relative contributions of the nasal coda and the prenuclear glide to the duration of the syllable? If the glide is realized as a constituent of the syllable's onset (either as a sister to the consonant or as a secondary articulation), the duration of the nuclear vowel should not be significantly altered, given that onset consonants do not generally affect the phonological weight of the syllable (Hayes 1995). One the other hand, if the glide is part of the syllable's rhyme we would expect a decrease in the duration of the vowel comparable to that found with a coda nasal. Figure 3 shows the relative proportions of the vowel durations in the various syllable contexts. While on average the nucleus of CVN is shorter than CGV, the difference is not very large.

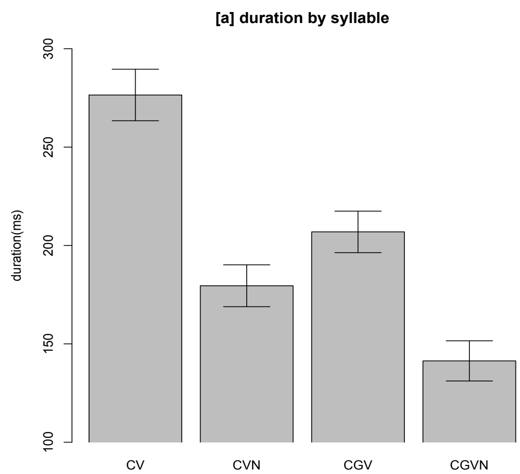

Figure 3 Mean duration of vowel nucleus in four syllable types. Error bars are 95% confidence intervals.

We ran a mixed-effects linear regression analysis to investigate this point further. The duration measures of the vowel were normalized as ratios of the word duration. Speaker, word, and trial (repetition) were set as random factors, while syllable shape (CV, CVN, CGV, CGVN) and tone as well as their interactions were the predictor variables, and the nuclear vowel's duration ratio relative to the word was the dependent variable. The variables were coded with the baseline set to CV and tone 1. As expected (8), CVN, CGV, and CGVN each were significantly shorter than CV. However, multiple comparisons with Tukey HSD revealed that when tone is held constant, there was no significant duration difference between the CVN and CGV syllables except in the case of tone 2. Differences between all of the other syllable types were significant.

(8) Results of a mixed-effects linear regression

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.927485	0.039757	23.329	0.0000

syllCVN	-0.410919	0.043340	-9.481	0.0000
syllCGV	-0.350877	0.043129	-8.136	0.0000
syllCGVN	-0.586644	0.041081	-14.280	0.0000
tone2	0.012341	0.018293	0.675	0.5001
tone3	-0.021257	0.018293	-1.162	0.2455
tone4	0.020888	0.018293	1.142	0.2538

Multiple comparisons of CGV and CVN syllables by tone

	Estimate	Std. Error	z value	Pr(>|z|)
CGV.1 - CVN.1	6.004e-02	4.125e-02	1.456	0.9771
CGV.2 - CVN.2	1.478e-01	4.357e-02	3.394	0.0390
CGV.3 - CVN.3	1.893e-02	4.230e-02	0.448	1.0000
CGV.4 - CVN.4	5.601e-02	4.282e-02	1.308	0.9918

Another point worth making is that the duration ratio of the nucleus is smallest in CGVN indicating that the effects of the glide and coda nasal seen separately in CGV and CVN combine in CGVN. To investigate this point further, we ran another regression analysis with a backwards difference coding scheme that allows one to check the significance of the differences between each level of a factor. As shown in (9), the difference between CV and CVN was significant, that between CVN and CGV was not, but that between CGV and CGVN was.

(9) Results of a mixed-effects linear regression

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.5903755	0.0278170	21.224	0.0000
CV - CVN	-0.4109187	0.0433402	-9.481	0.0000
CVN - CGV	0.0600416	0.0412460	1.456	0.1458
CGV - CGVN	-0.2357668	0.0388654	-6.066	0.0000

Thus, it appears that the prenuclear glide influences the duration of the nuclear vowel in a fashion indistinguishable from the coda nasal. This makes sense if both are components of the syllable's rhyme, where the phonologically relevant aspect of the duration of the syllable is normally calculated. It would not be expected if the glide were a constituent of the onset, especially a secondary articulation.

3.3 Nucleus duration and timbre

We now turn to the effect of syllable shape and duration on the height of the vowel and its location in F1/F2 space. Figure 4 is a plot of the data. The /a/ of Standard Mandarin is normally analyzed with a front allophone [a] before coda [n] (Lin 2007). This effect is apparent in the plot where CVN syllables have larger F2 values, indicating a more front vowel. CGV syllables present the opposite profile, showing the backing effect of the labial glide on the vowel. In terms of F1, CGVN syllables have more data points concentrated in the upper part of the plot, indicating smaller F1 values and suggesting a higher tongue body position. The CV syllables are more scattered in the F2 dimension, but they show a concentration with larger F1 values and hence a presumed lower tongue body position.

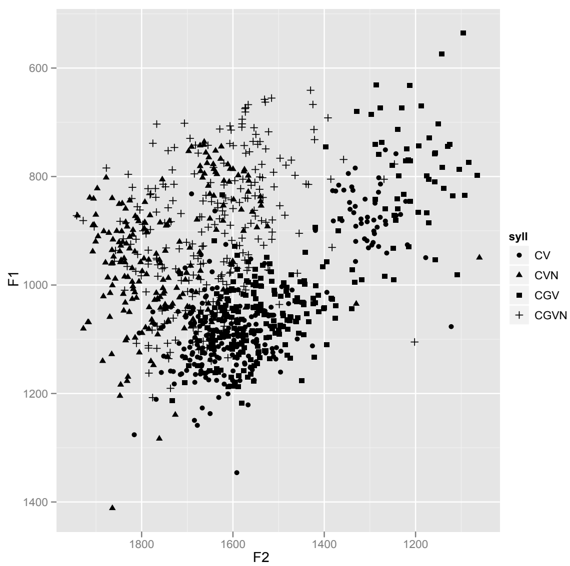

Figure 4 F1 x F2 plot of low vowel nucleus in Hz in four syllable contexts

The table in (10) indicates the means and standard deviations for our data. The raw scores were converted to the bark scale using the formula due to Traunmüller (1990). The resultant averages and standard deviations were converted back to the more familiar Hz scale with the calculator found at http://www2.ling.su.se/staff/hartmut/umrechnung.htm.

(10) Mean and standard deviation of F1 and F2 in bark and Hz for nucleus [a] in varying syllable contexts

	CV	CVN	CGV	CGVN
F1 (bark)	8.81 (.668)	8.24 (.715)	8.51 (.846)	7.84 (.795)
F2 (bark)	11.32 (.657)	12.07 (.457)	10.93 (.746)	11.68 (.497)
F1 (Hz)	1048 (91.7)	953 (95.5)	996 (106)	890 (101)
F2 (Hz)	1552 (90.8)	1737 (74.9)	1463 (98.0)	1639 (77.3)

We recall from section 1 that one of the crucial assumptions in Hsieh's (2012) analysis of the Mandarin raising and fronting of the low vowel seen in /tjān/ -> [tjēn] 'sky' was that the vowel was shorter in this context compared to /Can/ and /Cja/ syllables. Our data allow an indirect test of this assumption. Although the prenuclear glide in our data is a back vocoid, it should have the same effect on the duration of the following vowel as a front vocoid would. Granted this assumption, we can ask "are the different syllable contexts significantly correlated with the height of the vowel as reflected in F1?" In order to pursue this question, we ran another regression analysis. The formant values were again normalized on the bark scale using the formula due to Traunmüller (1990). The dependent variable was the normalized F1 of the nuclear vowel and the predictor variable was our four syllable shapes. Speaker, word, and trial were set as random factors. The factor of syllable shape was backwards difference coded to allow us to see the significance of the differences between the means at each level of this variable. The result is tabulated in (11).

(11) Results of a mixed-effects linear regression

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	8.35318	0.30250	27.613	0.0000
CV vs. CGV	-0.32111	0.10841	-2.962	0.0031
CGV vs. CVN	-0.23658	0.09375	-2.523	0.0118
CVN vs. CGVN	-0.39006	0.08529	-4.573	0.0000

We see that each level is significantly different from the preceding one, with the last being the greatest in magnitude and significance. Thus, Hsieh's (2012) assumption that the nuclear vowel in the CGVN syllable (where G = [j] and N = [n]) is shorter than in CGV or CVN, is consistent with our data. The plot in Figure 5 shows only a modest correlation between raw duration and F1 (r^2 = .21). Thus, the major factor in defining the low vowel raising is the CGVN syllable structure, suggesting that the phenomenon has been phonologized into a categorical effect rather than being simply the byproduct of articulatory undershoot.

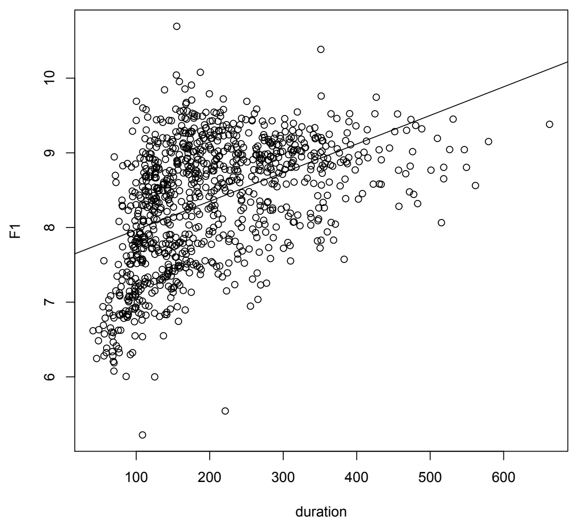

Figure 5 Plot of F1 (Hz) relative to duration (ms) for low vowel nucleus

A reviewer challenged our assumption that the duration effects of the labial glide on the following vowel would be informative of and comparable to those of the palatal glide based on the results of Gick's (2003) EMMA study of the palatal and labial glides in English. In this experiment Gick found that the labial constriction gesture of [w] in aw#a, aw#ha, and #wa nonsense sequences was more consonantal compared to its more vocalic tongue dorsum gesture as well as to the tongue dorsum gesture of the palatal glide [j] in aj#a, aj#h, and #ja. Gick related this articulatory difference between [w] and [j] in English to a phonological one. The [w] is banned from a syllable onset with a labial consonant while [j] patterns as an element of the nucleus, chiefly in the [ju] diphthong of *cute*, *pure*, etc. The reviewer notes a comparable phonotactic restriction in Mandarin against *pw*, *mw*, etc. sequences and on this basis suggests that [w]'s location in the syllable may well be different from [j] in Mandarin too. This is an intriguing observation but in the end we don't believe it bars us from seeing the nuclear vowel duration in Cuan syllables as informative about Cian syllables. First, Gick's experiment did not examine the glides in post-consonantal CGa contexts, which is what at issue here. Second, while there is a ban on *pw* sequences in English, in many American dialects there is a parallel ban against alveolar plus palatal glide sequences seen in words like *tune* ([tu], *[tju]), *suit* ([su], *[sju]), *nude* ([nu], *[nju]), etc. Hence, phonotactic restrictions are not directly informative of syllable structure affiliation. Third, even if we carry over the onset [w] vs. rhyme [j] analysis from English to Mandarin, the palatal glide should then induce even greater compression on the nuclear vowel than the labial glide and thus would still support Hsieh's proposal that /tjān/ > [tjēn] is due to the phonologization of phonetic undershoot due to time pressure.

A related point is that Standard Chinese also has low vowel raising in GVN syllables as seen in /ján/ > jén 'salt'. The glide is presumably in the syllable onset. If this is true then the nuclear vowel should be longer than it is in CGVN syllables and hence this in turn might put it out of the range of the undershoot analysis, which targets low vowels with the shortest duration. Our study did not include any GVN syllables and so we cannot comment on the nuclear vowel duration in this context. However, it is worth observing that in some languages an onset glide is more consonantal and produced with more constriction compared to a glide in the rhyme. If this were true for Standard Chinese then in articulating a GVN syllable the tongue body would have to traverse a greater distance to reach the low vowel compared to a glide in CGVN syllables. In other words, while more time to reach the low-vowel target may be available in GVN syllables this factor may be offset by a larger trajectory that could suffice to trigger undershoot of the low vowel. An analysis employing weighted constraints along the lines of Flemming (2003) and Cho (2011) may be able to shed light on this question and should be pursued in future research.

Another point worth making is that relative to the baseline CV, the prenuclear glide and the coda nasal [n] have opposing effects on F2: the latter pulls the vowel forward (increased F2) while the former lowers F2 through coarticulation with the back round vocoid. These differences are apparent in (10) above. In the conflicting CGVN context, the fronting effect of the coda nasal predominates (at least at vowel midpoint) since the nuclear vowel's F2 mean of 1643 Hz lies closer to CVN's 1742 Hz than to CGV's 1471 Hz and falls above the baseline CV's 1559 Hz. A mixed effects regression analysis found significantly different normalized F2 values relative to the baseline CV. The results appear in (12).

(12) Results of a mixed-effects linear regression

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	11.32048	0.24045	47.08	0.0000
syllCVN	0.74233	0.09117	8.14	0.0000
syllCGV	-0.40152	0.09949	-4.04	0.0001
syllCGVN	0.37255	0.09271	4.02	0.0001

All six comparisons with Tukey's HSD among the three levels of the syllable shape factor were significant at the 0.01 level or greater (13).

(13) multiple comparisons of syllable shape effects on F2

	Estimate	Std. Error	z value	Pr(>|z|)
CVN – CV	0.74233	0.09117	8.142	<0.001
CGV - CV	-0.40152	0.09949	-4.036	<0.001
CGVN – CV	0.37255	0.09271	4.018	<0.001
CGV - CVN	-1.14386	0.08542	-13.391	<0.001

CGVN - CVN	-0.36978	0.07742	-4.776	< 0.001
CGVN - CGV	0.77407	0.08706	8.891	< 0.001

This result coincides with another front-back vowel asymmetry in Mandarin, which was discussed by Lin (2002) in her study of the realization of the mid vowel archiphoneme in the context of adjacent glides. The situation in Standard Mandarin is shown in (14) (Lin 2002:303).

(14)	/piə/	[pje]		'don't'
	/kuə/	[kwo]		'wok'
	/pəi/	[pej]		'cup'
	/kəu/	[kow]		'dog'
	/tuəi/	[twej]	*[twøj]	'correct'
	/tiəu/	[tjow]	*[tjøw]	'throw'
	/yə/	[ɥe]	*[ɥø]	'moon'
	/çyə/	[çɥe]	*[çɥø]	'learn'

The nuclear /ə/ assimilates the backness (and rounding) of both a preceding and a following glide. In case both factors are present and they conflict in backness then the coda glide determines the outcome. Lin models this state of affairs in terms of a segmental markedness hierarchy *ø > *e, *o > *ə along with two harmony constraints imposing agreement for the feature [back]: one for the VC rhyme and the other for the CV demisyllable. As noted earlier, Lin (2002) assumes that the prenuclear glide is a constituent of the onset and thus two different prosodic domains must be postulated: rhyme harmony for the VG case and CV harmony for the prenuclear glide. In syllables with both glides such as /tuəi/and /tiəu/, under Lin's analysis there is no conflict since progressive assimilation is arbitrarily restricted to an open syllable. Given this stipulation, the backness of the vowel is determined by the coda glide in CGVG syllables, resulting in [twej] and [tjow]. But in view of our results suggesting that the prenuclear glide is in the syllable rhyme, a distinction must be drawn between progressive and regressive assimilation. The fact that the coda glide determines the outcome might be viewed as support for the c-command analysis of the prenuclear glide proposed by van de Weijer and Zhang (2008). On this view, the coda and nucleus form a syllabic constituent to the exclusion of the prenuclear glide. However, there does not seem to be any general tendency across the Mandarin dialects for the coda glide to predominate over the prenuclear glide in determining the backness of the mid vowel nucleus. In a survey of 160 Mandarin dialects, Lin (2002) finds a small bias for progressive assimilation over regressive assimilation for the back glide; but the inverse tendency was found for the front glide, as seen in (15).

(15) Number of Mandarin dialects showing assimilation of mid vowel nucleus to a preceding vs. a following glide (based on Lin 2002:311-312)

N	/əu/	/uə/	N	/əi/	/iə/
25	ow	wə	11	ej	jə
41	əw	wo	8	əj	je

An interesting topic for future research would be to survey the same dialects to see whether the implementation of /CGaN/ syllables with a front or back nuclear allophone aligns with the resolution of the vowel as front or back in CGVG syllables with conflicting glides. Lin (2002:305) states that in /CGəG/ structures the prenuclear glide is never decisive. As we have seen in (10) and (14) above, for Standard Mandarin resolution of the front-back conflict for the low and mid vowel nuclei in CGaN and CGəG structures coincides. But this may just be a coincidence.

4. Summary and Conclusion

This paper reviewed four different aspects of the Mandarin syllable in which previous research has invoked phonetic duration in an analysis of its internal phonological structure. These were its bimoraic structure, the effect of tone in isolation vs. sentential contexts, the affiliation of the prenuclear glide as a constituent of the rhyme or onset, and the raising of the low vowel in the context of /Cjan/ syllables. Our study was motivated by the fact that most of the prior investigations employed a single subject or a small number of data points. We reported the results of an analysis of the durational properties of 48 words balanced for syllable shape and tone that were spoken in isolation and sentential contexts by five female native Mandarin speakers. Our results are summarized as follows. First, we found statistical evidence that supported the bimoraic status of the CV syllable. Specifically, we did not find a significant difference in the duration of CVN and CGV syllables relative to CV. Second, we found a statistically significant duration hierarchy as a function of tone: tone 3 was the longest, tone 4 the shortest, while tone 1 and tone 2 were not significantly different. This duration hierarchy held across different syllable shapes. This finding coincides with two previous studies (with a smaller number of subjects or syllable shapes). Third, concerning the status of the prenuclear glide, we found that it shortened the duration of the nuclear vowel in a manner comparable to the effect of a coda nasal and took this as support for the analyses that parse this segment as a constituent of the syllable's rhyme. Finally, we had two results concerning the allophones of the low nuclear vowel. First, we found that the vowel was significantly shortened in CGVN syllables, which we interpreted as support for Hsieh's (2012) analysis of the raising of the low vowel as target undershoot due to time pressure, though we noted that there was no clear direct correlation between raw duration and F1. Second, we found that in the /Cuan/ syllable, the vowel was more front relative to baseline /Ca/, indicating that regressive assimilation to the coda nasal predominates progressive assimilation to the prenuclear glide. This finding correlates with the observation of Lin (2002) that in Standard Mandarin /CGəG/ syllables with conflicting front and back glides, the coda determines the character of the nuclear vowel. It was noted that this correlation between the behavior of the low and mid vowel nuclei should be investigated for other Mandarin dialects to determine if the state of affairs in the standard language is a coincidence or reflects a deeper principle of Mandarin syllable structure.

Acknowledgements

The research discussed in this article was conducted during the 2013-2014 academic year while the first author was a Visiting Scholar in the MIT Linguistics Department. Her visit was funded by the Shanghai Municipal Education Commission (SMEC), whose support is gratefully acknowledged. We are also grateful to our three anonymous reviewers as well as to Feng-fan Hsieh and Chingting Chuang for helpful comments.

References

Bao, Zhiming, 2000. Syllabic constituency and sub-syllabic processes. Journal of East Asian Linguistics 9, 287–313.

Bates, Douglas and Martin Maechler. 2013. Lme4: linear mixed effects models using S4 classes. R package version 0.999999-2. Online: http://cran.rproject.org/web/packages/lme4/index.html.

Boersma, Paul and David Weenink. 1992-2011. Praat: doing phonetics by computer. Version 5.2.21, http://www.praat.org

Chao, Yuen Ren., 1968. A Grammar of Spoken Chinese, third ed., 1985. University of California Press, Berkeley.

Cho, Hyesun. 2011. The timing of phrase-initial tones in Seoul Korean: a weighted constraint model. Phonology 28, 293-330.

Clements, George N. and Samuel J. Keyser. 1983. CV Phonology. MIT Press Cambridge, MA.

Duanmu, San. 2000. The Phonology of Standard Chinese. Oxford University Press, Oxford.

Duanmu, San. 2007. The Phonology of Standard Chinese. Oxford University Press, Oxford. (second edition)

Duanmu, San. 2012. Word-length preferences in Chinese: a corpus study. Journal of East Asian Linguistics 21, 89-114.

Feng, L. 1985. Duration of initials, finals and tones in Beijing dialects (in Chinese). Working Papers in Experimental Phonetics (Peking University Press), pp. 131-195.

Feng, Shengli. 2001. The multidimensional properties of “word” in Chinese. Contemporary Linguistics 3, 161-74.
Flemming, Edward. 2001. Scalar and categorical phenomena in a unified model of phonetics and phonology. Phonology 18, 7-44.
Flemming, Edward. 2003. The relationship between coronal place and vowel backness. Phonology 20, 335-373.
Fon, Janice, Chiang, Wen-Yu. 1999. What does Chao have to say about tones? A case study of Taiwan Mandarin. Journal of Chinese Linguistics 27, 15-37.
Gick, Bryan. 2003. Articulatory correlates of ambisyllabicity in English glides and liquids. In: John Local, et al. (Eds), Phonetic Interpretation: Papers in Laboratory Phonology VI, Cambridge University Press, Cambridge, pp. 223-239.
Hayes, Bruce. 1995. Metrical Stress Theory: Principles and Case Studies. University of Chicago Press, Chicago.
Ho, Aichen. 1976. The acoustic variation of Mandarin tones. Phonetica 33, 353-67.
Howie, John. 1976. Acoustical studies of Mandarin vowels and tones. Cambridge University Press, London.
Hsieh, Feng-fan. 2012. Low vowel raising in Sinitic languages: assimilation, reduction, or both? Language and Linguistics 13, 583-623.
Kager, René. 1999. Optimality Theory. Cambridge University Press, Cambridge.
Lee, Wai-Sum, and Eric Zee. 2003. Standard Chinese (Beijing). Journal of the International Phonetic Association 33, 109-112.
Lin, Yen-Hwei. 2002. Mid vowel assimilation across Mandarin dialects. Journal of East Asian Linguistics 11, 303-347.
Lin, Yen-Hwei. 2007. The Sounds of Chinese. Cambridge University Press, Cambridge.
Lin, Maocan and Jingzhu Yan. 1980. Beijinghua qingsheng de shengxue xingzhi [Acoustic characteristics of neutral tone in Beijing Mandarin]. Fangyan 3, 166-178.
Lu, D. 2005. The duration of CGVX and CVX syllables in Mandarin. Ms., University of Illinois at Urbana-Champaign.
Luo, Changpei, and Bianming Zhou. 1975. Xiamen Yinxi ji qi Yinyun Shengdiao zhi Gouzao yu Xingzhi [The Phonetics and Phonology of the Amoy Dialect]. Taipei: Hsiang-sheng Publishing House.
Mou, Xiaomin. 2006. Vowel modification by nasal codas in Standard Chinese: a study in the framework of the distinctive feature theory. MIT Ph.D. dissertation, Cambridge, MA.
Ohala, John. 1979. The production of tone. In: V. A. Fromkin (Ed.), Tone: a linguistic survey. Academic Press, New York, pp. 5-39.
Shih, Chilin. 1989. Tone and intonation in Mandarin. Working Papers of the Cornell Phonetics Laboratory 3, 83-109.
Silverman, Daniel. 1997. Tone sandhi in Comaltepec Chinantec. Language 73, 473-492.
Sundberg, Johan. 1979. Maximum speed of pitch changes in singers and untrained subjects.

Journal of Phonetics 7, 71-79.
Traunmüller, H. 1990. Analytical expressions for the tonotopic sensory scale. The Journal of the Acoustical Society of America 88, 97–91.
Van de Weijer, Jeroen. 2012. Using local conjunction to discover constraints: The case of Mandarin Chinese. In: Botma, B., Noske, R. (Eds.), Phonological Explorations: Empirical, Theoretical and Diachronic Issues. Mouton de Gruyter, Berlin and New York, pp. 255-264.
Van de Weijer, Jeroen and Jishen Zhang. 2008. An X-bar approach to the syllable structure of Mandarin. Lingua 118, 1416 -1428.
Xu, Yi. 1997. Contextual tonal variations in Mandarin. Journal of Phonetics 25, 61-83.
Wang, H.S., Chang, C., 2001. On the status of the prenuclear glides in Mandarin Chinese. Language and Linguistics 2, 243–260.
Woo, Nancy. 1969. Prosody and Phonology. MIT Ph.D. dissertation, Cambridge, MA.
Wu, Zongji. 1986. The Spectrographic Album of Mono-Syllables of Standard Chinese. Social Science Press, Beijing.
Yip, Moira. 2002. Tone. Cambridge University Press, Cambridge.
Yip, Moira. 2003. Casting doubt on the Onset-Rime distinction. Lingua 113, 779–816.
Zee, Eric. 2003. Frequency analysis of the vowels in Cantonese from 50 male and 50 female speakers. Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona: Universitat Autònoma de Barcelona, pp. 1117-1120.
Zhang, Jie. 2004. The role of contrast-specific and language-specific phonetics in contour tone distribution. Hayes, Bruce, Robert Kirchner, and Donca Steriade, eds. Phonetically Based Phonology. Cambridge University Press, Cambridge, pp. 157-90.

Appendix

number	word	gloss	syllable	tone
1	bā	bar	CV	1
2	gā	gamma ray	CV	1
3	dā	hook	CV	1
4	bá	draw	CV	2
5	gá	quack	CV	2
6	dá	reach	CV	2
7	dǎ	hit	CV	3
8	bǎ	hold	CV	3
9	gǎ	lovely	CV	3
10	dà	big	CV	4
11	bà	dad	CV	4
12	gà	embarrass	CV	4
13	gān	dry	CVN	1

14	bān	move	CVN	1
15	dān	pill	CVN	1
16	nán	men	CVN	2
17	hán	surname	CVN	2
18	zán	we	CVN	2
19	dǎn	courage	CVN	3
20	lǎn	lazy	CVN	3
21	bàn	conduct	CVN	4
22	gàn	do	CVN	4
23	dàn	egg	CVN	4
24	zhuā	catch	CGV	1
25	huā	flower	CGV	1
26	guā	melon	CGV	1
27	huá	China	CGV	2
28	guá	kind of sound	CGV	2
29	huá	slippery	CGV	2
30	huá	stroke	CGV	2
31	zhuǎ	claw	CGV	3
32	guǎ	skin	CGV	3
33	guǎ	widow	CGV	3
34	guǎ	widow	CGV	3
35	huà	words	CGV	4
36	guà	hang	CGV	4
37	guān	close	CGVN	1
38	zuān	dig	CGVN	1
39	duān	hold	CGVN	1
40	luán	mountains	CGVN	2
41	luán	mountains	CGVN	2
42	huán	return	CGVN	2
43	guǎn	manage	CGVN	3
44	duǎn	short	CGVN	3
45	huǎn	slowly	CGVN	3
46	duàn	paragraph	CGVN	4
47	guàn	pot	CGVN	4
48	luàn	wild	CGVN	4