

Syllabic reduction in Mandarin and English speech

L. Ann Burchfield and Ann R. Bradlow

Department of Linguistics, Northwestern University, 2016 Sheridan Road, Evanston, Illinois 60208 ann@u.northwestern.edu, abradlow@northwestern.edu

Abstract: This study examined whether language specific properties may lead to cross-language differences in the degree of phonetic reduction. Rates of syllabic reduction (defined here as reduction in which the number of syllables pronounced is less than expected based on canonical form) in English and Mandarin were compared. The rate of syllabic reduction was higher in Mandarin than English. Regardless of language, open syllables participated in reduction more often than closed syllables. The prevalence of open syllables was higher in Mandarin than English, and this phonotactic difference could account for Mandarin's higher rate of syllabic reduction.

© 2014 Acoustical Society of America **PACS numbers:** 43.70.Kv, 43.70.Fq [AL]

Date Received: January 28, 2014 Date Accepted: April 21, 2014

1. Introduction

Phonetic reduction refers to many kinds of deviation of a given word's pronunciation in running speech from canonical or careful pronunciations of the word. These forms may be reduced to such a degree that they are not recognizable in isolation, but they can easily be recognized in context (Ernestus et al., 2002; Xu, 1994). The same word may also be realized in many different ways depending on the degree of reduction (Johnson, 2004). Phonetic reduction complicates the task of speech recognition by removing the possibility of simple correspondence between acoustic input and underlying form. Determining what factors affect the probability of certain kinds of reduction could help in determining why reduction takes the form it does and how listeners are able to reconstruct full forms. If certain kinds of reduction are more common in specific environments, listeners may use information about the observed or expected phonetic environment to reconstruct full forms.

One way to determine what factors may contribute to reduction is cross-language comparison. Factors that have been shown to interact with rates of phonetic reduction in single-language studies also vary cross-linguistically, raising the possibility that rates of phonetic reduction overall may vary cross-linguistically. For example, languages differ in their permissible syllable structures, with some languages severely limiting possible coda consonants and studies of reduction in Mandarin (Tseng 2005; Cheng and Xu, 2009) have found open syllables more likely to undergo phonetic reduction than closed syllables. Thus, languages that favor coda-less syllables may show overall higher rates of phonetic reduction than languages that freely allow syllable codas. The principle here is that phonetic reduction may arise from language-general phenomena, such that the rate of occurrence of particular types of phonetic reduction within a language and across languages may be determined by general properties of the sound shape of the forms that are most likely to undergo reduction.

We explore this possibility by comparing rates of syllabic reduction across English and Mandarin, which have very different syllable-level phonotactics. We define syllabic reduction as any case in which the number of acoustic syllables realized in a word is less than the expected number of syllables based on the word's canonical form

(following the definition of "massive reduction" in Johnson, 2004). Syllabic reduction is a documented phenomenon in both English (Johnson, 2004) and Mandarin (Cheng and Xu, 2009; Tseng, 2005), and differences in syllabic structure make this comparison between English and Mandarin particularly interesting. English allows complex syllable structures with clusters of up to three consonants in onset position and four in coda position (e.g., /strenk θ s/). Mandarin, on the other hand, allows only a single consonant in onset position and codas are restricted to /n/ or /n/ (Tseng, 2005). In part because of this highly restricted syllable structure, syllable boundaries in Mandarin can be determined with higher accuracy based on an undifferentiated stream of phonemes than they can in English (Chen et al., 2007). This could lead to a higher rate of syllabic reduction in Mandarin because of a reduced need for acoustic cues to determine syllable boundaries. Another consequence of Mandarin's simpler syllable structure is that approximately 75% of syllables are open in Mandarin (Tseng, 2005) while in English only about 40% are open (Delattre and Olsen, 1969). If the observed tendency for open syllables to participate in syllabic reduction at higher rates in Mandarin (Tseng 2005; Cheng and Xu, 2009) is due to a language-general intrinsic vulnerability to reduction in open syllables, we would expect to see more syllabic reduction in Mandarin than English because of its higher proportion of open syllables.

However, another property of Mandarin points toward a different possible outcome. Recent work (e.g., O'Seaghdha et al., 2010) suggests that syllables, rather than phonemes, are the basic units of phonological planning in Mandarin. Moreover, Mandarin's morphosyllabic writing system and nearly one-to-one syllable-to-morpheme correspondence also support the notion of a special status for syllables in Mandarin. These syllable-based features of Mandarin morpho-phonology may exert sufficient pressure toward syllable preservation to counteract the general tendency for open syllables to undergo syllabic reduction, in which case we might observe less syllabic reduction in Mandarin than in English.

A third possibility is that all languages have a similar degree of redundancy at the syllable level and that speakers can let the demands of a given communicative task (e.g., importance of clarity vs speed, familiarity with the interlocutor, predictability of the utterance, word frequency, etc.) guide the amount of reduction permitted to surface. In this case, we would expect English and Mandarin to show similar degrees of syllabic reduction in speech produced under comparable circumstances. Under this view, rates of syllabic reduction may be affected by various factors within a single language (e.g., syllable structure, status of a particular syllable as a morpheme, etc.), but these effects might differ in magnitude. For example, English and Mandarin might both show higher rates of reduction in open than closed syllables, but the absolute rates for each kind of syllable could differ across languages, allowing the overall rate of syllabic reduction to remain similar in both languages.

The present study will address these issues by comparing rates of syllabic reduction in English and Mandarin in three speaking styles: Spontaneous speech, speech read in a natural style (read-plain style), and speech read in a careful, hyperarticulated style (read-clear style). The read-clear speech is included in order to provide a baseline condition closer to the speakers' underlying representations of the words spoken (for more information about clear speech, see Smiljanic and Bradlow, 2009; Uchanski, 2005) while the read-plain speech uses the same material produced in a more natural style. These two styles will form the core of the comparison because they are produced from the same script, but vary only in terms of degree of hyperarticulation, which is expected to affect rates of reduction. The use of scripted speech also allows reduction to be measured in terms of deviation from an externally specified target, avoiding the problem of making potentially erroneous assumptions about what a talker's internal targets may have been. Basing these scripts on the spontaneous speech, rather than a pre-existing text, allows the specified targets to be natural for the talker and comparable across languages without relying on translations. This allows us to avoid concerns about any potential differences in writing style or average word frequency across materials for the two languages which might arise due to idiosyncrasies of an individual translator.

2. Experiment

Eight native speakers of Mandarin (2 female) and eight native speakers of American English (6 female) participated in this study. Native Mandarin-speaking participants ranged in age from 23 to 27 yr (average age 23.9). All were bilingual with English as the L2, but dominant in Mandarin and raised in Mainland China. These participants began study of English at an average age of 9.5 yr and had lived in the United States for an average of 0.8 yr (range 0 to 2 yr). Three subjects reported experience with a third language (Spanish, Japanese, and Cantonese), but none reported using these languages regularly. Native English speaking participants ranged in age from 18 to 22 yr (average age 19.4 yr). All participants had some experience with one or more additional languages, but only two reported using another language regularly (Spanish in both cases). All participants were paid for their participation.

The set of speech recordings analyzed in this study consisted of three repetitions of two discourses by each participant. First, each participant produced two spontaneous narratives in his or her native language in response to pictorial prompts taken from Mayer (1974a) and Mayer (1974b). The same two prompts were used for both the Mandarin and English recordings. The Mandarin spontaneous narratives were transcribed by a native speaker of Mandarin in simplified Chinese characters, and the English spontaneous narratives were transcribed by a native speaker of English in standard English orthography with appropriate punctuation and capitalization, as judged by the transcriber. Speech errors and major disfluencies, including restarts and filled pauses, were removed in the versions of the transcriptions presented to participants for subsequent recordings.

The same subjects returned to the laboratory for a second recording session after the transcriptions were completed. In this session, subjects were asked to read each transcription of their own spontaneous narratives twice: Once with instructions to read as naturally as possible (read-plain style), and once as though speaking to someone having difficulty understanding (read-clear style). For all recordings, participants spoke into a Shure SM81 condenser microphone and were recorded using a Marantz PMD 670 flash recorder in a sound attenuated booth.

For each recording, four measures were taken: Number of acoustic syllables, total duration of speech, number of orthographic syllables, and articulation rate. Portions of the digital speech files not included in the transcription (e.g., disfluencies, restarts) were excluded from analysis. In the case of a reading error on the part of the participant or a transcription error on the part of the transcriber(s) of the spontaneous speech recording in the read-plain or read-clear recordings, the erroneous portions were removed from analysis in all three recordings so that the same materials would be analyzed across all three speaking styles. Twelve syllables out of a total of 1952 were excluded for one of these reasons. Periods of silence or non-speech sounds (e.g., breaths, coughs, lip smacks) lasting more than 100 ms were also excluded from analysis.

A Praat script adapted from De Jong and Wempe (2009), with minor modifications for ease of use across platforms and options for adjusting the format of output, was used to calculate the location and number of acoustic syllable nuclei (defined as voiced acoustic intensity peaks which are above the median intensity for the sample and are followed by an intensity dip), for each recording. To determine the orthographic syllable count in the English transcriptions, we used an algorithm which counts syllable nuclei based on the orthographic conventions of English, including rules accounting for contextual variation (e.g., not counting word-final "-e" except in specific contexts, such as the sequence "-ble"). See Kendall (2013) for full details. For Mandarin, the orthographic syllable count was obtained by counting the number of characters in each transcription. Articulation rate for each passage was calculated as

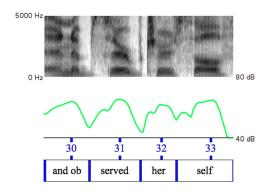


Fig. 1. (Color online) The phrase "and observed herself" with intensity peaks (acoustic syllable nuclei) marked on the first tier and orthographic syllables marked inside acoustic syllable boundaries on the second tier. The first two orthographic syllables would be classified as "participating in reduction" because they are each one of two orthographic syllables contained in a single acoustic syllable.

the number of orthographic syllables per second of speech duration (excluding errors, silences, and non-speech).

A subset of the recorded material from each of the three recordings by each participant was subjected to the critical analysis in which each acoustic syllable in the digital speech file was aligned with the orthographic transcript. For this analysis, the first fifteen seconds of speech from each of the two discourses in the spontaneous speech style were selected. Then, for each talker, the portions of the read-clear and read-plain recordings containing the same lexical material as the section selected from the spontaneous speech were selected for this analysis. On average, 15s contained 62 syllables and represented 23% of the discourses, which ranged in length from 20 to 151 s (average 66 s).

The selected recording portions were then segmented into individual acoustic syllables. For each intensity peak which was counted as an acoustic syllable nucleus, the associated intensity minima were located and marked in a PRAAT textgrid file. These minima served as the boundaries between acoustic syllables. Each acoustic syllable was then labeled with the orthographic syllable or syllables it contained, based on information in the transcript. An orthographic syllable was defined as participating in syllabic reduction if it was one of two or more orthographic syllables contained within a single acoustic syllable. Each orthographic syllable was coded for whether it was open or closed and whether it participated in reduction. The amount of reduction in each passage was then expressed as the proportion of orthographic syllables participating in reduction.³ An example of a phrase aligned using this method is depicted in Fig. 1.

3. Results

From the data in Table 1, we can see several notable trends. First of all, Mandarin contained a much higher proportion of open syllables (73%) than English (38%). These rates are comparable to those found in previous work [approximately 40% open syllables for English (Delattre and Olsen, 1969) and 75% open syllables for Mandarin (Tseng, 2005)]. Second, open syllables participated in reduction more often than closed syllables in all styles across languages. The average difference between rates of syllabic reduction in open and closed syllables was 6.5% (average across all styles in both languages). Third, Mandarin also showed more reduction than English in all speaking styles, with an average difference of 6% (average across all styles and both syllable types). Additionally, read-clear speech had less reduction than both read-plain and spontaneous speech, which shared similar levels of reduction. Finally, articulation rate was slightly higher in Mandarin than in English.

A linear mixed effects model was constructed with rate of syllabic reduction as the dependent variable. Fixed effects included language (English and Mandarin),

Table 1. Proportion of open and closed syllables in the materials from each language, and rates of reduction and articulation (calculated as the average of the articulation rates for each passage for a given language/style combination) for all syllable types and speaking styles for both languages.

Language and Style	% open syllables	% closed syllables	Reduction: open syllables	Reduction: closed syllables	Reduction: Overall	Avg. artic rate (syllables/second)
Mandarin						
Read-clear	72%	28%	26%	17%	23%	5.20
Read-plain			29%	27%	28%	5.97
Spontaneous English			30%	23%	29%	5.49
Read-clear	38%	62%	17%	12%	13%	4.19
Read-plain			28%	24%	25%	5.39
Spontaneous			30%	19%	24%	4.99

speaking style (read-clear and read-plain), and syllable type (open and closed), the interactions of these three variables, and articulation rate (taken over each whole passage, as described above). Style included both styles of read speech (clear and plain), but not the spontaneous speech because the read-clear and read-plain styles share the same script and thus the same externally prompted target syllables, making them more directly comparable to each other than to the spontaneous speech where there was no direct control over the participant's intended targets. Syllable type was included because of its relationship with syllabic reduction in Mandarin in previous work (Tseng, 2005; Cheng and Xu, 2009) and to test whether the syllable type differences between English and Mandarin can account for the observed cross-language differences in syllabic reduction. Speaking rate was included because of its expected relationship with reduction and its difference across the two language samples. Language was residualized to account for its correlation (r = -0.353, df = 4045, p < 0.0001) with syllable type, and style was residualized to account for its correlation with speaking rate (r = 0.493, df = 4045, p < 0.0001). For all subjects, a random intercept and random slopes for articulation rate and speaking style were included. All p-values for the model reported here were determined through model comparison.

Significant predictors of rate of syllabic reduction included syllable type and speaking rate (p < 0.05). Open syllables predicted more reduction (β = -0.309, z = -3.087, p < 0.01), as did faster speaking rates (β = 0.400, z = 5.599, p < 0.001). The interaction of style and syllable-type was marginally significant (β = 0.381, z = 1.96, p = 0.052). No other two- or three-way interactions reached significance.

Critically for the present study, language was not a significant predictor in this model, but syllable type was. Moreover, the fact that language did not enter into any interactions suggests that the difference across languages in rate of syllabic reduction can be accounted for by the higher frequency of open syllables in Mandarin. There was no significant effect of style when it was residualized to take into account its correlation with articulation rate. This suggests that a faster articulation rate, independent of other factors, predicted higher rates of reduction, a finding also supported in Cheng and Xu (2009, 2013). This main effect of articulation rate suggests that the well-established reduction of rate in clear speech relative to plain speech (e.g., Picheny et al., 1986) accounted for the variation in syllabic reduction rate across styles. Finally, the marginal style by syllable type interaction arose from the larger difference in reduction rates between open and closed syllables in the read-clear speaking style (7%) than in the read-plain speaking style (3%).

In summary, open syllables showed reduction more often than closed syllables (main effect of syllable type), and reduction increased with speaking rate (main effect

of articulation rate), while language and speaking style did not make independent contributions to reduction rates. Thus, the higher prevalence of open syllables in Mandarin contributes to its higher rate of syllabic reduction. Articulation rate contributed to the variation as well but was not able to account for it alone. Additionally, the difference between reduction rates in open and closed syllables was more pronounced in the read-clear style than in the read-plain style.

4. Discussion

This study aimed to determine how cross-linguistic differences affect phonetic reduction at the syllable level. The findings support the view that the rate of occurrence of syllabic reduction within a language and across languages is determined by basic properties of the sound shape of the forms that undergo reduction. Specifically, open syllables appear to be intrinsically more prone to reduction than closed syllables. Critically for the aim of this study, Mandarin showed overall more syllabic reduction than English, but this language effect was not significant when syllable type was included in the statistical model. This suggests that differences in the relative proportions of open and closed syllables across these languages contribute to the observed difference in rates of syllabic reduction. Additionally, the lack of a language by syllable type interaction suggests that rates of reduction in open and closed syllables behave similarly in English and Mandarin, further supporting the view that basic properties of a syllable determine its susceptibility to reduction. If this is the case, we would also expect other languages with high proportions of open syllables (e.g., Japanese) to show relatively high rates of syllabic reduction, a prediction which could be tested in future work.

The strength of this effect of syllable type on syllabic reduction varies by speaking style, with clear speech showing a greater difference between syllable reduction rates for open versus closed syllables than plain speech. One possible reason for this is that clear speech-related hyperarticulation is more likely to result in a strong acoustic cue to a syllable boundary for closed syllables. A hyperarticulated coda consonant might be likely to result in full closure of the vocal tract and thus a sharp drop in intensity at the syllable boundary. A hyperarticulated vowel at the end of an open syllable, on the other hand, might have a longer duration or an expanded vowel space, which would not necessarily result in stronger cues to syllable boundaries and therefore not result in a change to the syllable count.

The finding that syllabic reduction follows similar patterns in Mandarin and English is somewhat surprising given that several properties of Mandarin might suggest that it would be less prone to syllabic reduction than English. The apparent role of syllables as basic speech planning units (O'Seaghdha et al., 2010) and nearly one-to-one syllable-morpheme correspondence could plausibly result in lower susceptibility to syllabic reduction. In addition, the higher prevalence of open syllables in Mandarin might provide pressure to mitigate the difference between reduction rates in open and closed syllables in order to avoid higher overall rates of syllabic reduction, but this does not appear to occur. Instead, basic properties of the syllables themselves determine the rate of syllabic reduction.

Acknowledgments

We are grateful to Fan Gao for the Mandarin transcriptions, Chun Liang Chan and Kelsey Mok for technical support, and Angela Cooper and Patrick C. M. Wong for helpful suggestions. Work supported by Grant R01-DC005794 from NIH-NIDCD and an Advanced Cognitive Science Fellowship from Northwestern University.

References and links

¹The smaller sample was used due to the time-consuming, manual (rather than automatic) nature of aligning the transcriptions with the acoustic syllables, a procedure that was necessary for determining the separate rates of participation in reduction for open and closed syllables.

²The locations of the minima were determined using an internally developed PRAAT script which detected

minima in the intensity of the speech signal and marked their locations. In cases where the script detected minima associated with secondary peaks that were not considered syllable nuclei, the extraneous minima were removed from the output.

³For example, a reduction rate of 30% for a transcript of 100 orthographic syllables would mean that 30 of the syllables in the transcript reduced to half as many, i.e., 15, acoustic syllables in the acoustic signal, yielding a total of 85 acoustic syllables (70 that did not participate in reduction plus 15 that resulted from the reduction of the other 30).

Chen, T. M., Dell, G. S., and Chen, J. Y. (2007). "A cross-linguistic study of phonological units: Syllables emerge from the statistics of Mandarin Chinese, but not from the statistics of English," Chin. J. Psychol. 39(2), 137–144.

Cheng, C., and Xu, Y. (2009). "Extreme reductions: Contraction of disyllables into monosyllables in Taiwan Mandarin," in *Proceedings of Interspeech 2009* (ISCA, Brighton, UK), pp. 456–459.

Cheng, C., and Xu, Y. (2013). "Articulatory limit and extreme segmental reduction in Taiwan Mandarin," J. Acoust. Soc. Am. 134(6), 4481–4495.

De Jong, N. H., and Wempe, T. (2009). "Praat script to detect syllable nuclei and measure speech rate automatically," Behav. Res. Meth. 41(2), 385–390.

Delattre, P., and Olsen, C. (1969). "Syllabic features and phonic impression in English, German, French, and Spanish," Lingua 22, 160–175.

Ernestus, M., Baayen, H., and Schreuder, R. (2002). "The recognition of reduced word forms," Brain Lang. 81, 162–173.

Johnson, K. (2004). "Massive reduction in conversational American English," in *Spontaneous Speech: Data and Analysis. Proceedings of the 1st Session of the 10th International Symposium*, edited by K.

Yoneyama and K. Maekawa (The National International Institute for Japanese Language, Tokyo, Japan), pp. 29–54.

Kendall, T. (2013). Speech Rate, Pause, and Language Variation: Studies in Corpus Sociophonetics (Palgrave Macmillan, Basingstoke, UK), available at http://ncslaap.lib.ncsu.edu/speechrateandpause/ (Last viewed January 28, 2014).

Mayer, M. (1974a). Two Moral Tales (Four Winds Press, New York).

Mayer, M. (1974b). Two More Moral Tales (Four Winds Press, New York).

O'Seaghdha, P. G., Chen, J. Y., and Chen, T. M. (2010). "Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English," Cognition 115(2), 282–302.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (1986). "Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech." J. Speech Hear. Res. 29(4), 434–446.

Smiljanic, R., and Bradlow, A. R. (2009). "Speaking and hearing clearly: Talker and listener factors in speaking style changes," Linguist. Lang. Compass 3(1), 236–264.

Tseng, S. (2005). "Syllable contractions in a Mandarin conversational dialogue corpus," Int. J. Corpus Linguist. 10(1), 63–83.

Uchanski, R. M. (2005). "Clear Speech," in *The Handbook of Speech Perception*, edited by D. B. Pisoni and R. Remez (Blackwell, Malden, MA).

Xu, Y. (1994). "Production and perception of coarticulated tones," J. Acoust. Soc. Am. 95(4), 2240–2253.