

AN INVESTIGATION OF AFFRICATE SIMPLIFICATION IN CONVERSATIONAL MANDARIN

Aletheia Cui & Taylor Jones

Department of Linguistics, University of Pennsylvania, USA
acui@sas.upenn.edu; tayjones@sas.upenn.edu

ABSTRACT

We investigate the reduction of the palatal affricate /tʃ/ both word-initially and word-medially in spoken Mandarin, using the CALLHOME corpus of informal telephone conversations. An overwhelming majority of the tokens shows reduction in both environments. The processes of reduction include voicing, deletion of obstruent acoustic cues, glide formation, and complete deletion of the segment. We found that word-initial tokens of /tʃ/ have a higher retention rate of affricate acoustic cues, and word-medial tokens have a higher rate of lenition to a glide. While word-initial /tʃ/ tokens always exhibit some consonantal feature, a number of word-medial tokens show complete deletion of /tʃ/. Additionally, the realizations of reduction were highly gradient. There were no immediately apparent regional patterns. There was no suggestion of change in apparent time or age grading effects.

Keywords: Mandarin, consonant lenition, syllable contraction, corpus phonetics

1. INTRODUCTION

Consonant reduction is a common cross-linguistic phenomenon both in synchronic variation and historical change. One of the most widely studied case of consonant lenition is the reduction of /s/ in Spanish. While /s/ has been described as having three phonetic variants [s], [h] and [Ø], detailed phonetic study has shown that /s/ reduction has a much larger range of variation [3, 7, 8].

There has been a number of studies on syllable reduction and contraction in Taiwanese Mandarin [4, 5, 8], and some processes of lenition have been described for ‘Cosmopolitan’ Mandarin among Beijing speakers in non-state-run businesses [10]. However, there has not yet been a systematic treatment of consonant lenition in spoken Mandarin, especially mainland Mandarin. We aim to fill this gap through the study of a single phoneme in multiple environments, and we will provide an account of its full range of possible realization.

Specifically, we investigate the behaviour of the palatal affricate /tʃ/ in the syllable *jia* (IPA: /tʃjǎ/) in different lexical contexts, and in the word *bi3jiao4*

(IPA: / pi˥ tʃjǎo˥/) ‘relatively’ in order to gain an understanding of word-level effects from consonant reduction. Two main questions are examined here:

1. What processes characterize the reduction of /tʃ/ (e.g., affricate simplification, voicing, reduction to a glide, deletion, consonant simplification)?
2. What is the rate of reduction or deletion by position (i.e., initial vs. medial)?
3. Are consonant lenition and syllable contraction aspects of the same phenomenon?

2. METHOD

2.1 Data

Tokens of /tʃ/ were collected from the CALLHOME Mandarin Chinese Speech Corpus [2]. The corpus comprises 120 unscripted telephone conversations among native speakers of Mandarin with family members or close friends. Calls were made from North America to various places in Mainland China. Calls lasted up to 30 minutes. For each conversation, a continuous portion lasting from 5 to 10 minutes has been transcribed in standard Chinese orthography. Demographic data, including age, gender, and years of education, are available for most speakers making the calls. There is no explicit record of the locations of the speakers, but the area codes are available for each phone call.

For the present study, we examined 327 tokens of *jia* and 139 tokens of *bi3jiao4*, a total of 466 tokens. Of the *jia* tokens, 60 are monosyllabic words. Most of remaining tokens are disyllabic words, with a few exceptions that are longer. Among these tokens, 90 have *jia* as the first syllable, and 177 have *jia* as a word-medial syllable.

2.2 Classification

Relevant tokens of *jia* and *bi3jiao4* were located through the transcription, and utterances containing these tokens were extracted from the corpus. Each token was then inspected individually in Praat [1] and classified according to the presence or absence of acoustic cues typically expected of affricate. A fully articulated [tʃ] should have a period of voiceless closure followed by a transient and

frication noise. The tokens were divided into subgroups based their realization of these features.

For tokens of *bi3jiao4* in which the affricate has reduced to a glide, the duration of the glide is recorded. Distinction between complete absence of the affricate and reduction to a glide was decided by individual inspection in Praat with recourse to intensity and duration, and was subsequently verified through statistical analysis performed in the R statistical computing language. In addition, the durations of the entire word were recorded for tokens of *bi3jiao4*.

Relevant speaker information, including gender, age, and geographic information based on area code were also documented for each token.

3. RESULTS AND DISCUSSION

Previous discussion of reduction in Mandarin has primarily treated reduction and lenition as a binary process whereby syllables are either contracted or not contracted [4, 5, 9]. We find that there is a much wider range of variation.

3.1 Types of Reduction

Many of the tokens show simplification of the affricate through the deletion of one or more acoustic cues (closure, transient, and frication). Voicing often occurs independently or in conjunction with affricate simplification and signifies a further weakening of the affricate. Voicing occurred intervocally, both in word medial positions as in *bi3jiao4*, and across word boundaries, as in *peng2you jia1* ‘friend’s house.’ When there is voicing without any simplification, the surface realization of the affricate is [dʒ]. The absence of the stop features, closure and transient noise, can result in a voiceless fricative [ç] or a voiced fricative [ʒ] (Figure 2). The deletion of the frication cue yields [t] or the voiced [r].

In addition to simplification, the obstruent features of the affricate are sometimes completely absent, resulting in a glide [j]. There are also cases in which the consonant deletes completely, leaving only the vowel. Figure 1 shows the range of possible realizations for /tʃ/.

3.2 Rates of Reduction

Our investigation found that the palatal affricate was reduced the majority of the time, in both word initial and medial environments. However, there are a few differences between the reduction rates in these two environments. Figure 3 summarizes the rates of reduction by lexical environment.

Figure 1. Schematic representation of lenition processes of /tʃ/

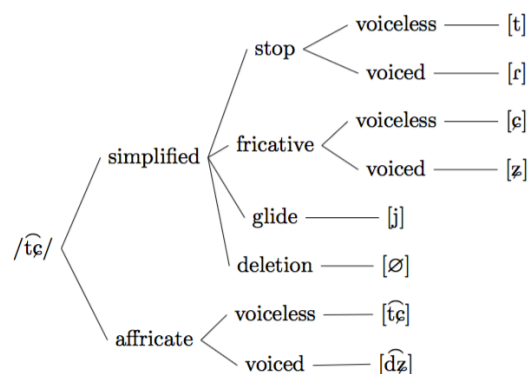
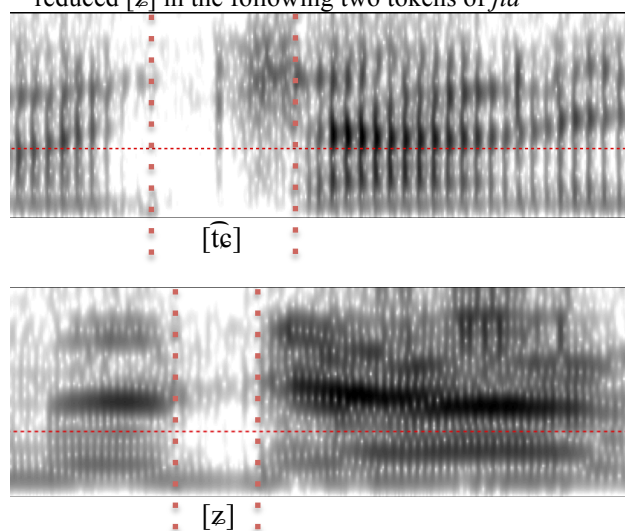


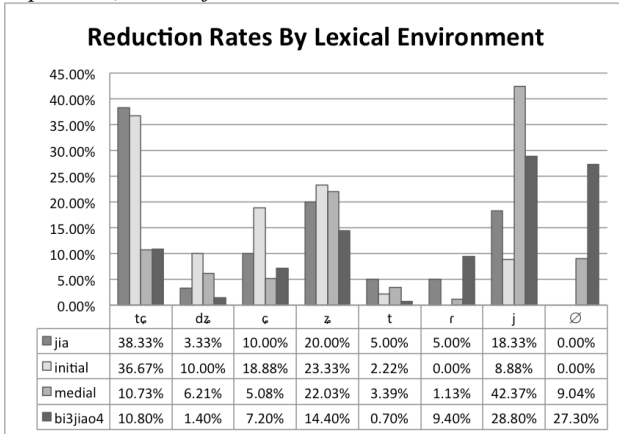
Figure 2. Compare a fully articulated [tʃ] and a reduced [ʒ] in the following two tokens of *jia*



Word-initial /tʃ/ shows a much greater rate of retention of affricate features. The affricate is articulated fully as [tʃ] for 38.33% of the monosyllabic tokens of *jia* and 36.67% of the tokens where *jia* is the first syllable. Word medially, it is pronounced as [tʃ] for only 10.73% of *jia* and 10.80% of *bi3jiao4*. Word initial /tʃ/ also show greater rates of reduction to fricatives and stops than to a glide, at a total of 40.00% for monosyllabic *jia* and 44.43% for word-initial *jia*. In comparison, the rates for word-medial *jia* and *bi3jiao4* are 31.63% and 31.70% respectively.

Word-medial /tʃ/ has a much higher rate of reduction to a glide [j], at 42.37% for word-medial *jia* and 28.80% for *bi3jiao4*. In contrast, reduction to [j] only occurs in 18.33% of the monosyllabic *jia* tokens and 8.88% of word-initial *jia* tokens. Additionally, the segment /tʃ/ is completely deleted in 9.04% of the word-medial *jia* tokens and 27.30% of the *bi3jiao4* tokens. Complete deletion does not occur in monosyllabic or word-initial *jia* tokens.

Figure 3. Percentages of reduction to each of the phonetic categories for *jia* as a monosyllabic word, *jia* in word-initial position, *jia* in word-medial position, and *bi3jiao4*.



3.3 Gradient Nature of Consonant Lenition

Although we have classified tokens of /tɕ/ into phonetic categories based on their acoustic cues, it is important to note that the realizations within each category also show a wide range of variation. For instance, for tokens that show voicing as part of the lenition process, the voicing may be present for only part of the segment or may last throughout the entire segment. Similarly, the reduction to [j] can vary greatly by duration and amplitude. Our finding that lenition is a gradient process parallel the work on the lenition of Spanish /s/ [3, 7, 8]. While dividing the tokens into categories can inform our understanding of consonant reduction, it is necessary to keep in mind that consonant lenition is far from a categorical phenomenon.

3.4 Disyllabic vs. Monosyllabic *bi3jiao4*

While there are clear tokens of fully articulated *bi3jiao4* as a disyllabic word, there is also a number of tokens in which the realization of *bi3jiao4* appears to be monosyllabic (Figure 4). These two extremes differ in duration by 95ms on average.

Much of the existing literature on reduction and lenition processes in Mandarin focuses on the contraction of disyllabic words to monosyllables. The presence of obstruent features in *bi3jiao4* tokens is a good indicator that the word has not contracted to a single syllable. However, the reduction to [j] presents a challenge. Since the vowel preceding /tɕ/ is /i/, it can be difficult to determine whether a homorganic glide is present as well. Both duration and change in intensity (Figure 5) were used to identify whether an intervocalic glide is present. This distinction was verified by performing Welch’s two-sample t-tests of word duration, comparing tokens labelled as monosyllabic with those labelled as disyllabic. Word duration between the two

categories was significantly (t-statistic: 4.6983, p-value <0.01).

We also compared monosyllabic tokens of *bi3jiao4* to underlying monosyllabic words *biao* (Figure 6). In total, 27 tokens of *biao* were found in the CALLHOME corpus and were measured for duration. We ran a Wilcoxon Rank-Sum test on the durations of monosyllabic *bi3jiao4* and underlying *biao*. While we cannot reject the null hypothesis that there is no difference between the two groups at the 0.01 level, we can reject it at the 0.05 level (W = 707, p-value = 0.01365). With an alpha of 0.05, the average duration of reduced *bi3jiao4* is actually shorter than the average *biao*. However, this test is not perfect. Most of the *biao* tokens are Tone 3; although the durational differences between Tone 3 and other tones are small in nonfinal positions, Tone 3 words can be significantly longer phrase-finally [10]. Nevertheless, the results are suggestive of the fact that extremely reduced tokens *bi3jiao4* have properties of being a monosyllable.

As discussed previously, there also exists an enormous range of variation in the reduction of *bi3jiao4* that is not well captured in terms of syllable contraction. Reduction to a monosyllable should best be understood as an extreme form of consonant lenition, and it stands at the end of a highly gradient spectrum.

Figure 4. Spectrograms showing an instance of disyllabic *bi3jiao4* and a token in which *bi3jiao4* has reduced to a single syllable

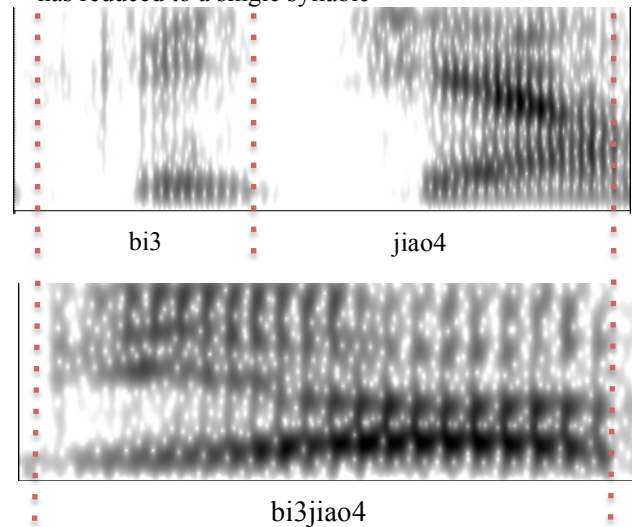


Figure 5. Intensity contour of a monosyllabic realization (left) versus disyllabic realization (right) of *bi3jiao4*.

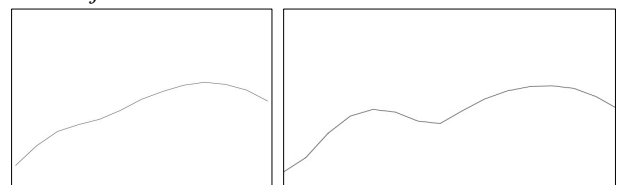
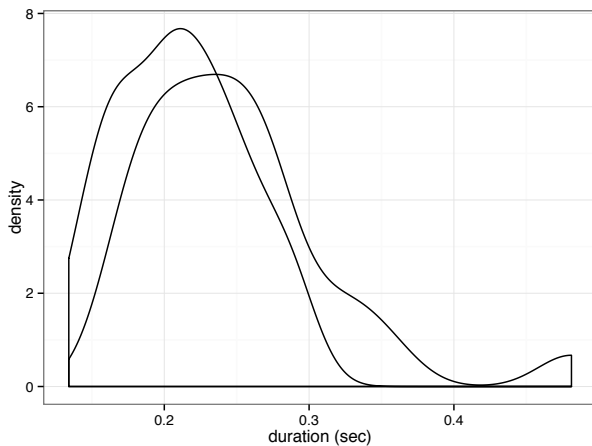


Figure 6. Word duration of *bi3jiao4* and *biao*



3.5 Geographic and age variation

Because the CALLHOME corpus has area code information, it is possible to geocode the data, albeit with a degree of uncertainty. There was no significant geographic variation. Nevertheless, there was a noticeable concentration of tokens of an alveolar in Beijing area codes. There was, however, not enough data to clearly identify this as a trend. We also found no significant variation by age.

4. CONCLUSION

In this paper, we present acoustic data on the lenition in Mandarin. We found that the reduction of $\widehat{t\epsilon}$ is very frequent in spontaneous speech, and $\widehat{t\epsilon}$ has a wide range of acoustic realizations. Our results show that syllable contraction does not provide an adequate account of lenition phenomena in Mandarin. While tokens of *bi3jiao4* can be classified as disyllabic or monosyllabic, there is a number of intermediate stages of reduction between the two oppositions.

Lenition is a gradient process in spoken Mandarin. The paucity of fully articulated tokens and the high variability among recorded tokens suggest that the lenition of $\widehat{t\epsilon}$ may be a change in progress. More data is needed in order to understand the status of lenited consonants in spoken Mandarin and the role of lenition in Mandarin phonology. It would be interesting to see whether other affricates undergo a similar process and show a comparable range of variability in pronunciation. Furthermore, it would be worth investigating whether the reduction to an alveolar tap is special in Beijing Mandarin.

There are a few potential problems in the present study. Since the acoustic information of the tokens were labelled manually, it is possible that there was inconsistency in the classification of the tokens. There could also be a certain amount of discrepancy between different labellers since subjective judgment was involved. In addition, the locations of the speakers, traced with the area codes, should be

taken with a grain of salt. There is no demographic data on the speaker's origin and linguistic background. More concrete data is needed for any conclusions.

Although lenition has been observed in spoken Mandarin, there have been few systematic examinations of what lenition actually encompasses. The Taiwanese studies provide evidence for word-internal reduction [4, 5, 9], but there is no comprehensive picture of the acoustic realization of lenition and the extent of lenition in Mandarin. This paper attempts to start filling the gap in our understanding of Mandarin lenition through an acoustic characterization of $\widehat{t\epsilon}$. Much future work is needed in this area of Mandarin phonetics and phonology.

5. ACKNOWLEDGEMENTS

We would like to thank Mark Liberman, Jianjing Kuang, and Meredith Tamminga for helpful comments and feedback.

6. REFERENCES

- [1] Boersma, Paul & Weenink, David (2015). Praat: doing phonetics by computer [Computer program]. Version 5.4.08, retrieved 24 March 2015 from <http://www.praat.org/>
- [2] Canavan, Alexandra, and George Zipperlen. CALLHOME Mandarin Chinese Speech LDC96S34. DVD. Philadelphia: Linguistic Data Consortium, 1996.
- [3] Cedergren, Henrietta Cecilia Jonas. The interplay of social and linguistic factors in Panama. PhD Dissertation, Cornell University, 1973.
- [4] Cheng, C., & Xu, Y. 2009. Extreme reductions: contraction of disyllables into monosyllables in Taiwan Mandarin. In *INTERSPEECH* (pp. 456-459).
- [5] Cheng, C., & Xu, Y. 2013. Articulatory limit and extreme segmental reduction in Taiwan Mandarin. *The Journal of the Acoustical Society of America*, 134(6), 4481-4495.
- [6] Duanmu, S. (2007). The phonology of standard Chinese. Oxford: Oxford University Press.
- [7] File-Muriel, R. J., & Brown, E. K. (2011). The gradient nature of s-lenition in Caleño Spanish. *Language Variation and Change*, 23(02), 223-243.
- [8] Terrell, T. D. (1979). Final /s/ in Cuban Spanish. *Hispania*, 599-612.
- [9] Tseng, S. C. 2005. Contracted syllables in Mandarin: Evidence from spontaneous conversations. *LANGUAGE AND LINGUISTICS-TAIPEI*, 6(1), 153.
- [10] Zhang, Q. (2006). Cosmopolitan Mandarin: Linguistic practice of Chinese. *Journal of Asian Pacific Communication*, 16(2), 215-235.