Paradigm Academic Press
Studies in Social Science & Humanities
ISSN 2709-7862
NOV. 2023 VOL.2, NO.11

Speech Rate and Sentence Length's Influence on Perception of Mandarin Tone Sandhi

Xueke Yang[1]

[1] University College London, London, the United Kingdom

Correspondence: Xueke Yang, University College London, London, the United Kingdom.

doi:10.56397/SSSH.2023.11.03

Abstract

It is well known that in Mandarin, the first low tone (T3) changes to a Rising tone when two low tones occur in an uninterrupted succession. It was discovered that Mandarin speakers could not distinguish the second tone from the third tone regardless of speech rate. People could not perceive the small pitch difference between the sandhi tone and the underlying tone 2 found in the production experiment. Speech rates at which the phrases were spoken did not affect the identification of the phrases. This study aims to investigate if the speech rate and sentence length really influence the perception of sandhi. A perception test is performed and 15 Mandarin speakers participated in the perception of the material. The identification differences under distinct speech rates and sentence lengths are significant. It is found that the sandhi tone is perceptually indistinguishable from tone 2, and speech rates and sentence length impact the identification of the sandhi.

Keywords: the third tone sandhi, speech rate, sentence length, perception

1. Introduction

1.1 An Introduction to Mandarin Chinese

Mandarin is a tonal language. Chao (1930) proposed a categorization of the total pitch range into five distinct levels, denoted as 1, 2, 3, 4, and 5, which aligned with the descriptors of low, half-low, mid, half-high, and high, respectively. Mandarin has four distinct tonal patterns. These tonal patterns are denoted by specific starting and ending points. The first tone is characterised by a steady high pitch (55), while the second tone exhibits a rising pitch (35). The third tone is distinguished by a falling and then rising pitch contour (214), while the fourth tone is characterised by a high falling pitch (51).

1.2 The Third Tone Sandhi

Chao (1948) defined tone change as the alteration in the actual pitch values of tones that occur when syllables are said consecutively. Tones undergo shape modifications due to the influence exerted by neighbouring tones. There exist three distinct versions of the third tone. According to Chao's (1948) findings, the Mandarin Tone 3 exhibited its whole tonal contour [214] exclusively when it was said in isolation or as the final word of a speech segment followed by a pause. When Tone 3 was succeeded by a tone that was not Tone 3, it was typically articulated with a low tone contour [21] and lacked a rising tail, which could be attributed to articulatory constraints. Finally, when two third tones were consecutively presented without interruption, the initial third tone underwent a transition to a second tone. This particular phonological occurrence was commonly referred to as the third tone sandhi. The term "tone sandhi" is employed to describe post-lexical tonal alterations that are influenced by a range of phonetic, morphological, and syntactic elements. The underlying mechanics of this phenomenon, however, continue to be mostly unknown at present.

There are different opinions on whether underlying the rising tone is entirely the same as the third tone sandhi or not. Through statistical analysis of 52 Beijingers' phonetic materials, Feng Shi and Wang Ping (2006) investigated the statistical characteristics of the intonation of Beijing bill characters, obtained the distribution trend of the internal tone variants of each tone position, and found some differences between the second tone and the third tone in Mandarin. From the data distribution in every measurement point, f0 of the underlying second tone was approaching the high pitch range, while most data of the third tone was concentrated at the bottom of the pitch range. The commonality of the second tone and the third tone was that there was a turning point in the middle of the syllable, but the turning point of the second tone was closer to the onset, while the third tone was closer to the middle of the syllable. The second difference was that the minimum of the third tone was lower than the second tone. The onset values between them were similar. For the second tone, the offsetf0 value was its characteristic, while the turning point was the characteristic of the third tone.

Regarding the acoustic difference between the underlying rising tone and third tone sandhi, Peng (2000) showed that not only the offsetf0 of the pitch contour but also the whole pitch height was lower than that of the rising tone. In spite of the acoustic difference, the contours of sandhi and rising tone were both influenced by speech rate. Peng found that their pitch ranges became smaller, and the rising slopes were shallower as the speech rate increased. Peng supposed that the reason why the tone targets were undershot at a fast speech rate was that the realization of a significant pitch increase appeared to be in conflict with the short syllable duration. However, listeners were not able to identify these differences among speech rates even when the speech rate was slow, and the syllable duration was long.

1.3 Perceptual Investigation

According to Wang and Li (1967), it was found that the sandhi tone exhibited perceptual indistinguishability from tone 2. A total of 130 pairs of test items were produced for the experiment. One member of the pair bore the second-third underlying tone sequence, while the other member bore the third-third sequence. The remaining properties, such as initials and finals, were identical. A total of fourteen individuals who were native speakers were enlisted to participate in the task of identifying the sequences. The participants' scores for accurately identifying the proper information varied between 49.2% and 54.2%. It was evident that T2 and T3S were perceptually indistinguishable.

The finding of Peng (2000) replicated the findings of the earlier perception study of Mandarin tone sandhi by Wang & Li (1967). Peng did a speech recognition experiment, analyzed the recognition results of the subjects, and found that the average listening accuracy of the first-name subjects was 50%, which showed that the listeners were guessing randomly and did not distinguish them. According to Peng (2000), "Mandarin speakers could not distinguish T23 from T33 regardless of speech rate. They could not perceive the small pitch difference between the sandhi tone and the underlying tone 2 found in the production experiment. Speech rates at which the phrases were spoken did not affect the identification of the phrases."

Speer (2008) investigated the time course of lexical ambiguity recognition caused by tonal triad by tracking the subjects' eye movements. They found that when they heard a T3S tone sequence, they looked first at the word with a T3S and vice versa. It was precisely because the subject was hesitant and uncertain about the tone sequence heard that he would look at another tone sequence; that was, the subject had difficulty distinguishing between T2 and T3S.

From the above literature discussion, T2 and T3S are still very easily confusing in normal speech communication. Precise acoustic measurements reveal some differences, but the ear can not distinguish them. This phenomenon, the mismatch between phonetic pronunciation and speech perception, is known as "class merging".

2. Perception Experiment

2.1 Method

A perception experiment was conducted to test whether T2 is perceptually distinguishable from T3S and whether speech rate and sentence length influence the distinction between them, following the results found in the production test. Based on the production results: Faster speech rate and the shorter sentence lead T3S tend to be more T2-like and less T3-like than ones in slow speech rate and long sentences, it is predicted that T2 and T3S may become neutralized in fast speech since the increasing speech rate makes the f0 difference between T2 and T3S more minor. We predict that listeners could not distinguish T3S from T2 on identification. The T3S may be biased toward identification as T3 in slow speech rate and long sentences and be biased toward identification as T2 in fast speech rate and short sentences.

Before the perception experiment, a preliminary experiment was run to ensure that the subjects could hear tones correctly. During the test, subjects were asked to choose the words they heard between the /tone 2 + tone 3/ phrase and the /tone 3 + tone 3/ phrase. In the main perception task, the stimuli were recordings from the production data, coming from 2 male and two female speakers, who were randomly selected. The materials were only statements containing T22, T23, T32, and T33, the same as the production test. Speakers read them at normal, fast, and slow speech rates with two repetitions.

The stimuli used in the perception test were 288 tokens (3 sentence types x 4 tones x 3 speech rates x 2 repetitions x 4 speakers). Tokens were placed in random order. Fifteen native Mandarin speakers (eleven females and four males) participated in the perception test. All of them were born and grew up in China. The experiment was run individually in a quiet room. Listeners sat before a computer, listening to the stimuli through headphones. There was a sheet listeners needed to finish. In the sheet, two Chinese characters with two different tones, such as “fan1 tu2” and “fan1 tu3”, responded to every recording. Listeners were required to make a judgement between the two target syllables by labelling them yellow.

2.2 Results

The results of the perception experiment are shown in Figure 1 and Figure 2. Figure 1 shows the average error rate among fifteen subjects in different tone conditions. Figure 2 indicates the averaged error rate at distinct speech rates and in different sentence groups only in the condition of T33.

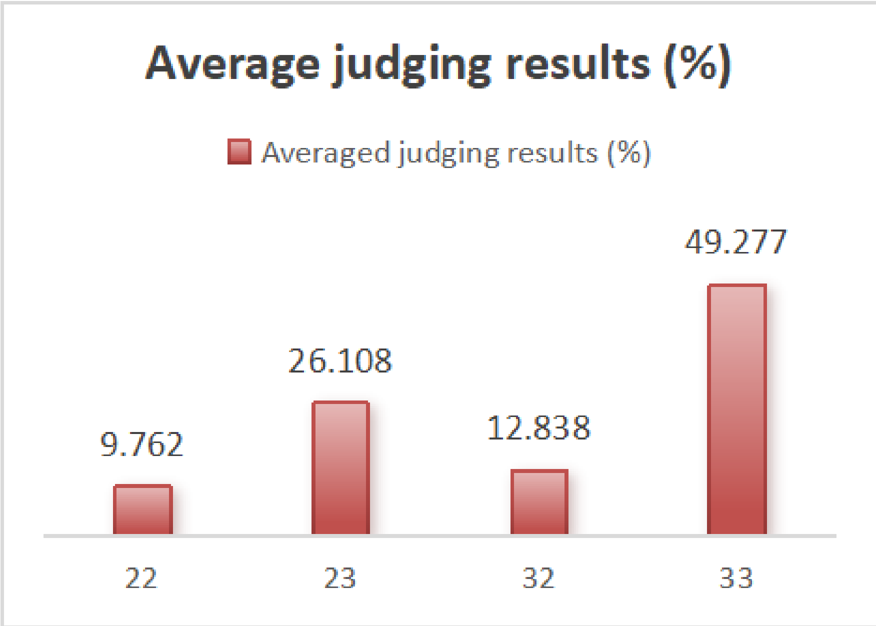

Figure 1. The average judging results of different tone conditions

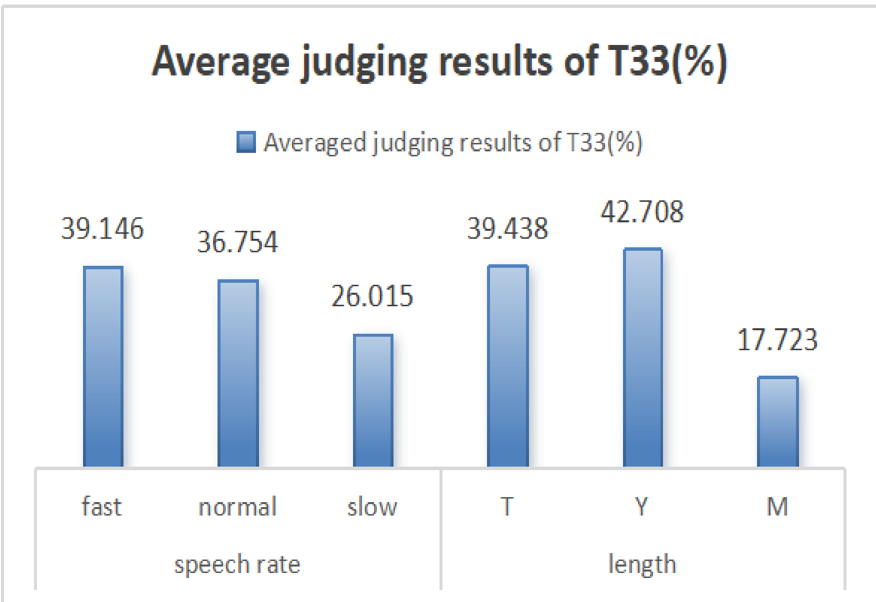

Figure 2. The average T33 judging results at different speech rate and sentences

A one-way ANOVA was performed, fitted with the predictor's tone types (T22, T23, T32 and T33) and the dependent variable error rate. The effect of the difference between tone types on identification was significant (SS=12606, df=3, SE=4202, F=20.121. p<.001). From Figure 1, the average percentage of wrong identifications in T33 is the largest (49.277%), and the average error rate in T23 was in the middle (26.108%), which was more extensive than that of T22 (9.8%) and T32 (12.8%). It is worth noting the low error rate of T23, which is different from the result of Peng (2000). The results suggest that listeners can identify T22 and T32 relatively accurately, but it is challenging to identify T23 and T33 precisely. It is indicated that most listeners need help to perceptually distinguish the nuance between the underlying second tone and the sandhied third tone.

Concerning sandhi T33 items, a two-way ANOVA was conducted, fitted with the predictor's speech rate (fast, normal, slow) and sentence length (T, Y, M). The dependent variable was the error rate in the T33 condition. The speech rate (SS=1271.6, df=2, SE=635.8, F=8.178, p=.001) and length (SS=4794.7, df=2, SE=2397.4, F=15.834, p < .001) was reported to show significant effects on identification results. In the T33 condition, the error rate at fast speed (39.146%) is almost the same as (only a little bit higher) that of normal speed (36.754%) but significantly higher than that at slow speed (26.015%). As can be seen, the result indicated that in the conditions of fast and normal speech rate, listeners were likely to make a wrong judgement that T3S was identified as T2, whereas, at slow speech rate, T3S tended to be identified as T3. Moreover, 42.708% of T3S syllables in the sentence group Y were identified as T2, a little higher than that of group T (39.438%) and much higher than that of group M (17.723%). This generally indicates that the longer the sentence, the more T3S is identified as T3.

In brief, the phrases spoken at a slow speech rate and in longer sentences significantly bias the decision of listeners toward T33. Phrases spoken in fast speech rate and shorter sentences were identified as T23 more frequently than as T33. That is, speech rate and sentence length did significantly contribute to differentiating T33 from T23.

3. Discussion

Our predictions for the perception task are that most listeners are unable to perceptually distinguish T3S and T2, and speech rate and sentence length have an effect on identification. Specifically, the T3S may be biased toward identification as T3 in slow speech rate and long sentences and be biased toward identification as T2 in fast speech rate and short sentences. The results are consistent with these predictions, corresponding to the finding in the production experiment that the application of the third tone sandhi rule decreases with slower speech rate and longer sentences.

Our finding was also in line with the discoveries of Wang and Li (1967) and Peng (2000), who found that the tone sequence T33 is homophonous with the sequence T23 in neutral speech, and native listeners could not perceptually distinguish T3S from T2. However, our findings on speech rate are in opposition to the findings of Peng (2000), whose study indicated that participants' answers were random and did not differ across different speech rates. In this study, it was shown that the identification of sandhi was influenced by speech rate, for the error rate decreased as the speech rate slowed.

Another point is the low error rate of T23, which is different from Peng (2000) where T23 was found to be biased toward being recognized as T33. In this perception test, the error rate of T33 is nearly 50%, while the error rate of T23 is about 25%, so listeners were more likely to identify T33 as T23 but less likely to identify T23 as T33 when the boundary strength is strong. There may be some reasons why T23 has a low error rate. The reason may be that the error rate of T23 should be originally as low as that of T22 and T32, but the existence of T32 stimuli and options may interfere with the identification of T23, especially at a fast speech rate. Listeners may identify the underlying tone of T2 they listened to is T3 when they think T3 sandhi applied here because they had realized that sandhi may happen in some sentences during the test. This hypothesis needs another experiment to be verified which suppresses the interference of T33 to get the true identification of T23. These controversies require further study in the future.

4. Conclusion

Through the experiment, it shows that in the conditions of fast and normal speech rate, listeners were likely to make a wrong judgement that T3S was identified as T2, whereas, at slow speech rate, T3S tended to be identified as T3. It also indicates that the longer the sentence, the more T3S is identified as T3. In summary, the perception study demonstrates that most listeners cannot distinguish the nuance between the underlying second tone and the third tone sandhi. Contrary to the previous research, speech rate and sentence length also have an effect on the perception results.

References

Chao, Y. R., (1968). *A Grammar of Spoken Chinese*. Berkeley: University of California Press.

Chen, Matthew Y., (2000). *Tone Sandhi: Patterns across Chinese Dialects*. Cambridge: Cambridge University

Press.

Cheng, Chin-chuan, (1968). Mandarin phonology, Ph.D. dissertation. University of Illinois.

Cheng, Chin-chuan, (1973). *A synchronic phonology of Mandarin Chinese*. The Hague: Mouton.

Ito, K., & Speer, S. R., (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of memory and language, 58*(2), 541-573.

Peng, S.-H., (2000). Lexical versus 'phonological' representations of Mandarin Sandhi tones. In *Papers in Laboratory Phonology V: Acquisition and the Lexicon*. M. B. Broe and J. B. Pierrehumbert. Cambridge: Cambridge University Press, pp. 152-167.

Wang, W. S., & Li, K.-P., (1967). Tone 3 in Pekinese. *Journal of Speech & Hearing Research, 10*(3), 629-636.

Xu, Y., (1986). Acoustic-phonetic characteristics of junctures in Mandarin Chinese [in Chinese]. *Journal of Chinese Linguistics*, (4), 353-360.