PERCEPTION OF COARTICULATED TONES IN TAIWAN MANDARIN AND TAIWAN SOUTHERN MIN

Huang, Po-Hsuan[1]
Chiu, Chenhao[2]

Graduate Institute of Linguistics, National Taiwan University[1, 2]
Neurobiology and Cognitive Science Center, National Taiwan University[2]

r09142003@ntu.edu.tw[1]
chenhaochiu@ntu.edu.tw[2]

ABSTRACT

Lexical tones in tone languages have been found to coarticulate with ambient tones, and speakers are observed to normalize for such coarticulation. Beijing Mandarin is one of the examples. In languages with larger tone inventories, however, the recoverability of the target tones is lower, and faithful comprehension of coarticulated tones based solely on normalization may be less likely. In this study, perception of coarticulated tones in Taiwan Mandarin and Taiwan Southern Min is investigated. It was revealed that compared with Taiwan Mandarin, Taiwan Southern Min tones induced much less normalizing effects. Crucially, Taiwan Southern Min was found to have stricter tone boundaries than Taiwan Mandarin, presumably allowing for effective tone perception under tonal coariculation despite the larger tone inventory and a lower degree of normalization. This study sheds light on the perception strategies of coarticulated tones in languages with tone inventories of different sizes and the nature of tonal coarticulation.

Keywords: tone perception, normalization, tonal coarticulation, Taiwan Mandarin, Taiwan Southern Min.

1. INTRODUCTION

Lexical tones in tone languages are ideally distinguished with their tone values and shapes. In real-world communication, however, the production and perception of tones is highly variable, subject to factors including prosody [1], syllable duration [2], and tonal coarticulation [1, 3, 4, 5, 6, 7]. In this study, we focus on the last factor, i.e., tonal coarticulation, where a target tone's pitch value and/or pitch contour varies owing to preceding and/or following tones, to the extent that it might be acoustically similar to another lexical tone. Specifically, we focus on the perception of coarticulated tones in Taiwan Mandarin (TM) and Taiwan Southern Min (TSM).

Lexical tones are known to coarticulate with ambient tones [1, 3, 4, 5, 6, 7, 8], and such coarticulation has been found to be perceptible to the speakers [1, 4, 6, 9], and could influence tone perception [4, 6, 9]. For example, in Beijing Mandarin (BM), [4] found that changing ambient tones could affect the identification accuracies of the target tones. The possible confusion of tone perception caused by such variation has been found to be dealt with by means of normalization in BM [4, 9], where the target tones were perceived as higher when preceded by lower tones, and vice versa. This mechanism therefore serves to aid the perception of coarticulated tones in BM.

However, while such normalization is possible in BM and presumably TM, which have only four lexical tones, normalizing effects as strong would be less probable in languages with larger tone inventories, such as TSM, with the lower recoverability [10] led by the larger number of potential target tones.

Logically, there could be three types of languages pertaining the perception of coarticulated tones. One would have no tonal coarticulation, and therefore does not require any further mechanism. The other two, with their tones coarticulated, would have to develop strategies to deal with such coarticulation. One possibility is to have looser tone boundaries, and for the listeners to accept coarticulated tones as other tones, and then retrieve the speakers' intended tones through normalization. The other is to have stricter boundaries, and reject coarticulated tones as other lexical tones. These languages are referred to as Type III, I and II languages hereafter, as illustrated in Fig. 1. Under this account, BM, and likely TM, would be Type I languages. A complete reliance on normalization to resolve the possible confusion of coarticulated tones is however less likely for TSM, since normalization in a language with a large

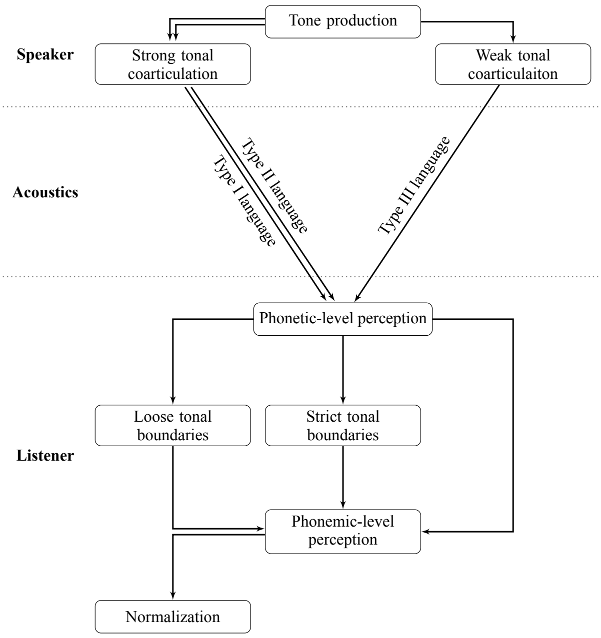

Figure 1: Three kinds of languages with reagrd to tone perception under tonal coarticulation.

tone inventory would be more difficult as mentioned previously. On the other hand, it has also been found that while TSM has more lexical tones than TM, it has identical magnitude of tonal coarticulation as TM [11]. On this account, one may predict that TSM is more likely a Type II language, which has to maintain a stricter tone boundary in order for the coarticulated tones to not be perceived as other tones.

Since previous studies on the perception of coarticulated tones in TM and TSM are sparse and inconsistent, and no direct comparison between the two languages could be found, this study aims to provide a systematic comparison of the two languages. Crucially, this study seeks to find out:

- whether linguistic differences exist between the degrees of normalization for coarticulated tones in TM and TSM.
- if TSM has weaker normalizing effect, how else TSM speakers could cope with the higher probability of perceptual confusion.

2. PERCEPTION OF COARTICULATED TONES IN PREVIOUS STUDIES

2.1. Perception of coarticulated tones in BM

In general, while no studies of TM could be found, past studies generally found coarticulated tones to be perceptible by the listeners in BM. In [4], swapping of ambient tones was found to affect the identification of target tones. Target tones were perceived with higher accuracies in compatible positions than in conflicting ones. Coarticulated tones were therefore hypothesized to be perceived through normalization. A more direct result supporting such stipulation was borne out in [9], where the low-tone-falling-tone boundaries were found to be higher when the ambiguous target tones were preceded by lower F0 offsets, and vice versa.

2.2. Perception of coarticulated tones in TSM

Tonal coarticulation was also found to be perceivable in TSM, and certain normalizing effects seemed to be present. In [1], subjects could predict truncated target tones with its ambient tones above chance rates, and in [6], target tones were found to be perceived as lower when preceded by higher F0 offsets, and vice versa. The audios were however not processed, and therefore the acoustic cues other than F0 might have given hints to the subjects in these two studies.

2.3. Sectional summary

While normalizing effects seemed to be found in both BM and TSM, no previous research of TM was conducted. Crucially, a quantifiable and consistent design would be required to make a direct comparison between the two languages.

3. METHODS

Two experiments were conducted in this study. The first was a replication of [9], where the magnitude of normalization of tonal coarticulation was measured. The second was a word-non-word test, aiming to examine the subjects' strictness on tone boundaries.

3.1. Experiment 1

Experiment 1 measures the normalization effect for tonal coarticulation in TM and TSM. Following the design of [9], a continuum from low-level tone to falling tone preceded by tones of different pitch levels of offsets were synthesized into disyllabic stimuli, and subjects were asked to judge whether they heard a low-level tone or a falling tone. If normalization was at work, different pitch levels of preceding offsets should render different thresholds of low-level tone/falling tone judgement.

3.1.1. Participants

This study recruited 43 Taiwanese college students as participants (25 females; 20–27 y.o., mean=21.93). 15 of them were native speakers of TM. These subjects were not speakers of TSM. 28 of them were native speakers of TSM (and also TM). Among these 28 participants, 11 were advanced TSM speakers.

Both Experiments 1 and 2 had a Mandarin version and a Southern Min version. The monolingual group participated in the Mandarin versions. The bilingual group participated in both the Southern Min and Mandarin versions.

3.1.2. Stimuli and procedure

3 Pairs of TM disyllabic words and 4 pairs of TSM words were chosen for each of the two versions. These pairs were minimal pairs that differed only in the tones of the second syllables. Within each pair, the first syllables were the same and served as the primes, while the second syllables were the targets, and were a low-level tone and a falling tone, respectively. The pairs were 55+21/55+51, 35+21/35+51, 51+21/51+51 in TM version and 33+21/33+51, 55+21/55+51, 51+21/51+51, 21+21/21+51 in TSM version.

The audios used for syntheses were produced by two native speakers. The F0 values of the primes were then synthesized into five-level tone scales with the equation of [12], with the f_{min} being 100 Hz:

$$scale = \frac{1}{2}(39.86 \times \log(\frac{f_i}{f_{min}})) + 1$$

The F0 values of the targets were divided into 10 F0 levels on Bark's scale from 0.9 to 1.9. The durations of all syllables were 0.3s, and the mean intensities were scaled to be the same. These are illustrated in Fig. 2.

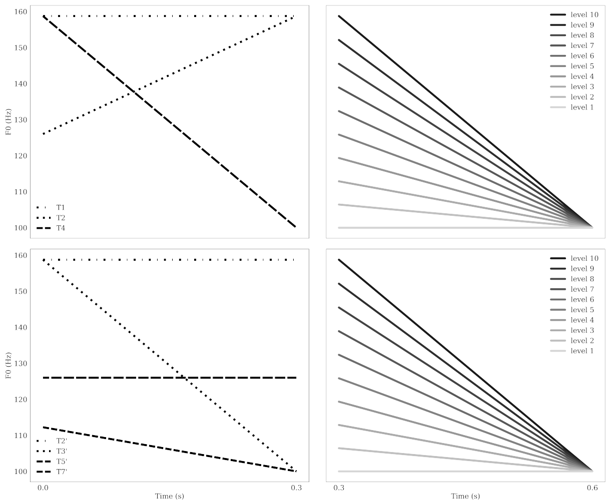

Figure 2: Illustration of synthesized stimuli.

Each stimuli were randomly shuffled, and repeated five times. In each trial, subjects saw an ear picture and were asked to judge whether they heard the low tone or the falling tone.

3.1.3. Analyses

The responses of the subjects were calculated as percentages of falling tone responses, and then fitted through GAMMs. The looseness of the fitted splines was taken as indicator of the magnitudes of perceptual compensation for tonal coarticulation.

3.2. Experiment 2

Experiment 2 was conducted to see whether TM speakers and TSM speakers had different acceptance ranges of falling and low-level tones.

3.2.1. Stimuli

In each version, 10 disyllabic words were used. All their first syllables were the high-level (55) tone. Half of them had as their second syllables the falling (51) tone, and another half the low-level (21) tone. The words were checked to have no minimal pair counterparts; that is, the words in the 55+21 group would be non-words if the second syllables were changed into a falling tone, and vice versa. The target tones were a continuum from low-level to falling tones. The targets and the preceding high-level tones were synthesized as in Experiment 1. For each level, there were 10 repetitions, randomly presented. In each trial, subjects saw the intended words, and were asked to decide whether the tokens were good production of the intended words or not.

3.2.2. Analyses

The subjects' responses were calculated as percentages and taken as acceptance rates of each of the 10 levels. The first derivatives of the acceptance rates were then fitted through GAMMs and examined with simple t-tests.

The maximum of the first derivatives was taken as indicators of the rate of acceptance climb, and the level on which this maximum point falls was determined as the threshold of tone acceptance. Both larger maximum "velocity" (first derivatives) of acceptance climb and a threshold closer to the two ends of the continuum (i.e., a lower threshold level for the low tone, and a higher threshold level for the falling tone) would indicate a stricter tone boundary, and vice versa.

4. RESULTS AND DISCUSSIONS

4.1. Normalizing effects for coarticulated tone in TM and TSM

GAMM results of Experiment 1 showed that the distances between the fitted splines of the monolingual group's TM falling tone responses (the maximum distance being 2.78 level) were significantly wider than those in TSM (the maximum distance being 1.53 level), as in Fig. 3, suggesting TM had stronger normalizing effects than TSM.

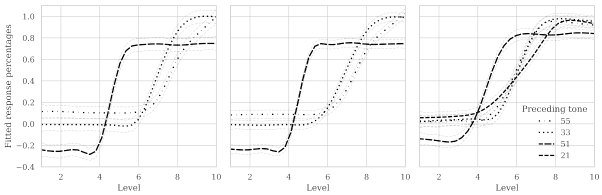

Figure 3: GAMM fitted falling tone response percentages in Experiment 2 (left: Mandarin (monolingual); middle: Mandarin (bilingual); right: Southern Min).

This is in accordance with our prediction. Since TSM has 7 lexical tones, while TM has only 4, retrieving the intended tone by the speaker would be much more difficult in the former. While in TM, a simple one-to-one mapping could be made during normalization, in TSM, multiple possibilities could exist. The lower level of retrievability may lead TSM listeners to rely less on normalization.

4.2. Tone boundaries in TM and TSM

Experiment 2 showed that the falling tone acceptance rates in TSM had substantially larger maximum of first derivatives of the acceptance rates (p=.019*), and marginally significantly higher levels of thresholds (p=.059), as in Fig. 4. This

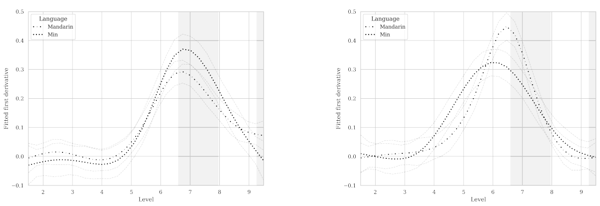

Figure 4: GAMM fitted first derivatives of acceptance rates in Experiment 3 (left: low tone; right: falling tone). The gray areas indicate intervals of significant differences.

pattern was more significant on the advanced bilingual group (p=.005***/.091**, respectively). This pattern, however, was not seen for the low tone, where the maximums of first derivatives of the acceptance rates were higher in TM than in TSM (p=.018*), and the levels of the thresholds were not significantly different (p=.974).

The results suggested that the boundary was stricter in TSM for the falling tone, while for the low tone, this pattern was absent. Nevertheless, such absence of a stricter boundary for the low tone in TSM was likely due to the tone sandhi rules, where the TSM low tone would become the falling tone. Though the stimuli was isolated disyllabic words, and the subjects were reminded that the words were produced in isolation and therefore should not undergo sandhi changes, several subjects reported feeling both the low tone versions and falling tone versions to be acceptable for them. This might explain why the boundaries for the low tone were not significantly stricter for TSM subjects as seen in the falling tone. The results thus indicated stricter tone boundaries in TSM than in TM.

4.3. Perception strategies of coarticulated tones in TM and TSM and the nature of tonal coarticulation

The results in the two experiments indicated that while TM, like BM in past studies, behaved like a Type I language that could have accepted coarticulated tones as other lexical tones, and then retrieved the targets through normalization, TSM did not have as strong normalization, and could have perceived coarticulated tones as the intended tones owing to the stricter tone boundaries.

Our results thus suggest that tonal coarticulation may be a universal constraint that cannot be voluntarily contolled. One could imagine effective perception to be difficult under tonal coarticulation in a language with a large tone inventory, with the higher possibility of confusion and the lower reliability of normalization. As seen, TSM listeners had to keep stricter tone boundaries to attain faithful comprehension. The reason why TSM does not choose to resolve such confusion with less tonal coarticulation may be that it simply is not voluntary and may well exist in all tone languages. This, however, is beyond the scope of this study and will require further typological research.

5. CONCLUSIONS

This study is the first to investigate the perception of coarticulated tones in TM, and to make quantifiable comparison between languages with different sizes of tone inventories. We hope to shed light on the interaction of production and perception, and to provide further clues pertaining the nature of tonal coarticulation.

6. REFERENCES

[1] S.-H. Peng, "Production and perception of Taiwanese tones in different tonal and prosodic contexts," *Journal of Phonetics*, vol. 25, no. 3, pp. 371–400, 1997.

[2] Y. Xu and M. Wang, "Organizing syllables into groups—Evidence from F0 and duration patterns in Mandarin," *Journal of Phonetics*, vol. 37, no. 4, pp. 502–520, 2009.

[3] X. S. Shen, "Tonal coarticulation in mandarin," *Journal of Phonetics*, vol. 18, no. 2, pp. 281–295, 1990.

[4] Y. Xu, "Production and perception of coarticulated tones," *Journal of the Acoustical Society of America*, vol. 95, no. 4, pp. 2240–2253, 1994.

[5] ——, "Contextual tonal variations in Mandarin," *Journal of Phonetics*, vol. 25, no. 1, pp. 61–83, 1997.

[6] H. Wang, "The prosodic effects on Taiwan Min tones," *Language and Linguistics*, vol. 3, pp. 839–852, 2002.

[7] Y.-C. Chang and F.-F. Hsieh, "Tonal coarticulation in Malaysian Hokkien: A typological anomaly?" *The Linguistic Review*, vol. 29, pp. 37–73, 2012.

[8] Y. R. Chao, *A Grammar of Spoken Chinese*. Berkeley: University of California Press, 1968.

[9] H. Zhang, H. Ding, and W.-S. Lee, "The influence of preceding speech and nonspeech contexts on Mandarin tone identification," *Journal of Phonetics*, vol. 93, p. 101154, 2022.

[10] E. Flemming, *The Grammar of Coarticulation*. Paris: L'Harmattan, 2011.

[11] B.-X. Huang and C. Chiu, "Tonal coarticulation in Taiwan Mandarin and Taiwan Southern min," unpublished.

[12] J. Fon and W.-Y. Chiang, "What does Chao have to say about tones?—A case study of Taiwan Mandarin," *Journal of Chinese Linguistics*, vol. 27, no. 1, pp. 13–37, 1999.