The Effect of Tone Sandhi on Speech Perception of Taiwanese Falling Tones

Chia-Hsin Yeh, Yen-Hwei Lin

Department of Linguistics and Languages, Michigan State University, United States
chiahsin@msu.edu, liny@msu.edu

Abstract

This study investigates phonetic and lexical effects on speech perception of two Taiwanese tones: high-falling and low-falling, and suggests that tone sandhi also plays a role. A tonal identification task and a lexical task were conducted on 30 Taiwanese speakers. The results show that onset pitch and pitch drop are more crucial than offset pitch to the contrast of the two falling tones, and participants are more responsive to frequent words. In addition, the response frequency of the two categories is consistently lower than 90%, and that the high-falling responses significantly increase in the tone sandhi condition. This study verifies the psychological reality of multiple linguistic representations, including phonetic, phonological, lexical information, and suggests that a model of speech perception should consider these linguistic cues as processing fundaments.

Index Terms: falling tone, speech perception, Taiwanese, lexical effect, tone sandhi

1. Introduction

This study models after Yeh and Lin’s [1] attested effects of pitch perturbations and lexical frequency on falling tones in examining speech perception of two Taiwanese falling tones, high-falling tone (Tone-51) and low-falling tone (Tone-21). According to Yeh and Lin’s [1] results, the phonetic and lexical effects on the perceptual patterns of Hakka and Mandarin falling tones are quite similar. The similar results obtained from Hakka and Mandarin are predictable since the falling tone pairs are phonemic in the two languages. However, the two Taiwanese phonemic falling tones differ in that they interact in tone sandhi. Taiwanese is known for its complex tone sandhi system, as illustrated in (1), which exhibits an opaque chain shift [2]. For example, in non-final position, underlying Tone-21 (low-falling tone) becomes surface Tone-51 (high-falling tone), and underlying Tone-51 becomes surface Tone-33 (mid-rising tone). The tone sandhi system can potentially exert an influence on speech perception of Taiwanese falling tones and present different perceptual patterns from those in Hakka and Mandarin.

(1) Taiwanese Tone Circle [3]

Tone-51 ←→ Tone-33 ←→ Tone-21

Similarly, as Yeh and Tu [5] and Liu and Wang [6] indicate, the tone sandhi system in (1) can account for the confusion matrix of Taiwanese tonal errors. For example, mid level tone (Tone-33) is largely mispronounced as low-falling tone (Tone-21), but mostly misperceived as high level tone (Tone-55). Yeh and Tu [5] suggest that the production errors result from over-application of tone sandhi and the perception errors result from under-application of tone sandhi.

However, there are two confounding factors which may disclaim any tone sandhi effects on Taiwanese tonal processing: i.e. the effects of phonetic similarity and the psychological reality of Taiwanese tone circle. First, as Yeh and Tu [5] point out, tonal errors and their substitutes are usually phonetically similar to each other, and so are underlying and surface tones of tone sandhi. For instance, Moore and Jongman [7] show that Mandarin T2 and T3 (surface and underlying tones of Mandarin T3 sandhi) are phonetically similar in terms of pitch contour, Huang [4] teases apart the phonetic factor from the tone sandhi factor by showing that Mandarin speakers spent longer time in discriminating T2-T3 pairs than American English subjects, whereas the two Taiwanese studies [3, 6] do not separate these two factors. Second, the psychological reality of Taiwanese tone sandhi is questioned by several studies [2, 8, 9], which show that the application rate of Taiwanese tone sandhi is mostly below 30%. Although Chuang et al [10] find that adopting an alternative method may facilitate Taiwanese speakers’ application of tone sandhi, their results are slightly above the chance level, and far lower than the Mandarin results in Zhang and Lai [11]. These two potential confounds will be considered in our examination of how phonetic attributes and lexical frequency, as well as tone sandhi, influence the perceptual patterns of Taiwanese falling tones.

2. Research Questions

This study addresses three research questions as follows: first, what phonetic attributes affect speech perception of falling tones; second, whether or not there is a lexical effect on speech perception of falling tones; and third, whether or not tone sandhi plays a role in the perception of falling tones. The phonetic and lexical variables are confounding factors in the previous studies [3, 6], and are found to be crucial to the perception of falling tones by Yeh and Lin [1]. Therefore, the two factors are also examined, and are controlled to assess the tone sandhi effect.

2.1. Phonetic attributes and lexical frequency

According to Yeh and Lin [1], onset pitch, offset pitch, and pitch drop are three crucial phonetic attributes of falling tones in Mandarin and Hakka, and lexical frequency also plays a role in tonal perception. Onset pitch refers to the initial pitch peak; offset pitch refers to the ending pitch valley; and pitch drop refers to the difference between the peak and the valley.
Lexical frequency refers to both the actual word count in a given corpus and the subjective judgment of lexical familiarity. If these phonetic and lexical factors are crucial to the perception of falling tones, then the hypotheses are as follows:

(3) Hypotheses Based on Yeh and Lin [1]

a. Onset pitch perturbations will cause tonal confusion between high-falling and low-falling tones, and a perceptual boundary of the two categories will be found.

b. Offset pitch perturbations will cause tonal confusion between high-falling and low-falling tones, and a perceptual boundary of the two categories will be found.

c. Pitch drop perturbations will cause tonal confusion between high-falling and low-falling tones, and a perceptual boundary of the two categories will be found.

d. Frequent words will be more likely to be chosen, and a significant difference between responses of the frequent words and less frequent words will be found.

2.2. Psychological reality of tone sandhi

Before examining the effect of tone sandhi on speech perception, it is important to clarify how we define tone sandhi as a variable. According to previous studies [2, 9, 10], Taiwanese tone sandhi can be represented as allomorphic variations or opaque phonological patterns. Since Taiwanese tone sandhi is unproductive, Tsay and Myers [9] argue that both underlying and surface forms of sandhi tones are memorized, and are retrieved accordingly in different prosodic positions. However, Zhang et al [2] argue that it is the phonological opacity that makes Taiwanese tone sandhi unproductive and underlearned, and that Taiwanese tone sandhi is not represented differently from other sandhi rules. The representation issue as to whether Taiwanese tone sandhi is represented lexically or phonologically is beyond the scope of this study. We use the term “tone sandhi” in a general sense, and simply consider it to be part of speakers’ linguistic knowledge. Since both studies [2, 9] show that prosodic positions, sandhi or non-sandhi conditions, constitute the key to the application of Taiwanese tone sandhi, we predict that lexical responses in non-final position (the sandhi condition) are more susceptible to the effect of tone sandhi. As shown in (1) above, low-falling tone (Tone-21) becomes high-falling tone (Tone-51) in non-final position, so high-falling responses are predicted to increase significantly in the sandhi condition.

3. Methodology

In the experimental design, the three phonetic attributes (onset pitch, offset pitch, and actual pitch drop) of falling tones are chosen as stimulus variables, and the lexical factor is set up as response type in a tonal identification task and a lexical task.

3.1. Stimuli

This experiment has 30 stimuli total, 3 continua (onset pitch, offset pitch, and actual pitch drop) x 10 variations. Each stimulus is made of a /ti/ syllable, 380 ms in duration, and 80 dB in intensity, based on Yeh and Lin’s [11] stimulus setup. Two anonymous reviewers point out that high-falling tone is short in duration, and the 380 ms length may lead to perceptual biases against high-falling tone. Such a concern is alleviated by the findings that f0 is the main cue for tonal perception in Taiwanese [12] and most tonal languages [13], and that the longer duration did not disfavor high-falling responses in Yeh and Lin [1].

As to the 30 stimuli, the pitch perturbations of onset pitch height, offset pitch height, and pitch drop were synthesized into three continua respectively (onset pitch continuum, offset pitch continuum, and pitch drop continuum) by Praat version 5.1 [14]. In the onset pitch continuum, the f0 ranges of 10 stimuli are 100-90 Hz, 110-90 Hz, 120-90 Hz, etc. These 10 stimuli vary in onset pitch height and pitch drop (both actual and perceived differences), but not in offset pitch height. In the offset pitch continuum, the f0 ranges of 10 stimuli are 160-60 Hz, 160-70 Hz, 160-80 Hz, etc. These 10 stimuli vary in offset pitch height and pitch drop (both actual and perceived differences), but not in onset pitch height. In the pitch drop continuum, the f0 ranges of 10 stimuli are 120-60 Hz, 130-70 Hz, 140-80 Hz, etc. These 10 stimuli vary in onset pitch and offset pitch height, but not in actual pitch drop: i.e., the actual pitch drop difference is kept constant as 60 Hz although the perceived difference decreases from the low pitch register to the high one.

3.2. Participants

30 Taiwanese speakers (14 males, 16 females; mean age: 35.5 years old) were recruited from the Taipei, Taichung, and Kaohsiung areas in Taiwan. The three cities are chosen to neutralize dialectal differences. Among these participants, 13 speakers use the northern dialect; 12 speakers use the central dialect; 5 speakers use the southern dialect.

3.3. Tasks

Two perceptual tasks, a tonal identification task and a lexical task, were conducted. In each task, there are 120 trials total (30 stimuli x 4 repetitions). Before the identification task, a training session was given with three minimal pairs, such as high-falling tone in 府 hu51 of 政府 ‘government’ versus low-falling tone in 抵 hu21 of 未抵 ‘too late’ because Taiwanese speakers had never been taught formally the two tonal categories.

In the identification task, the participants were asked to identify if they heard a high-falling tone as hu-51 in the word of ‘government’ or a low-falling tone as hu-21 in the word of ‘too late’. They had to respond with ‘hu-51 of government’ or ‘hu-21 of too late’ in Taiwanese. In the lexical task, they were asked to recognize if they heard the Taiwanese word 抵 ti51 of 抵抗 ‘resistance’ or the Taiwanese word 智 ti21 of 心智 ‘mind’. They had to respond with ‘ti51 of resistance’ or ‘ti21 of mind’ in Taiwanese. The setup of lexical responses is essential, as Chuang et al [10] show that the phrase of ‘A of AB’ can significantly enhance application of tone sandhi. Although the two words, 抗 as in ‘resistance’ and 心 as in ‘mind’, have different tones, they are quite similar in pitch height. Peng [3] shows that listeners’ responses were largely based on the height of f0 contour, and the co-articulation effect on tonal perception was barely better than the chance level. Hence, we assume that these two words’ co-articulation effects are minimal. After the experiment, participants were asked to judge lexical familiarity of those lexical responses on a five-point scale, with 5 indicating the most familiar and 1 the least familiar.

4. Results

A perceptual boundary of each continuum was assessed to examine the three phonetic factors. Two-sample T-test was then conducted to compare the results of the two perceptual tasks. The comparison was made to examine the lexical and the potential tone sandhi effects on tonal perception.
4.1. Onset pitch continuum

The frequency rates of 10 onset pitch stimuli are shown in Figure 1. The x-axis of the figure refers to the 10 stimuli (from stimulus #1:100-90 Hz to stimulus #10:190-90 Hz) and the y-axis refers to the frequency rates of the low-falling tone. The higher the percentage of the x-axis, the more low-falling-tone responses are made. Otherwise, lower percentage refers to more high-falling-tone responses.

The results show that the frequency rates of both low-falling and high-falling responses are consistently lower than 90%, including those of the first and the last three stimuli, which are supposed to be typical variants of low-falling and high-falling tones respectively. In addition, there is a perceptual boundary (50% rate) of the two tones, as illustrated by the dashed line in Figure 1. The perceptual boundary is found around stimulus #6:150-90 Hz. As indicated by the dotted circle and lines, the boundaries of the two tasks are slightly different. For instance, the red dotted line (the lexical task) is on the right of the blue dotted line (the identification task). The difference shows that there are more low-falling responses in the lexical task.

The two-sample T-Test analysis was conducted to compare the frequency rates of both tasks: tonal identification task (IDN) and lexical task (LEX), and the analysis shows that there is a significant difference between tonal and lexical responses of only two stimuli, stimulus #6:150-90 Hz, \( t(58)=1.7829, p=0.0399<0.05 \), and stimulus #9:180-90 Hz, \( t(58)=1.7179, p=0.0456<0.05 \), but not of the others.

4.2. Offset pitch continuum

The frequency rates of 10 offset pitch stimuli, from stimulus #1:160-60 Hz to stimulus #10:160-150 Hz, are shown in Figure 2. The results show that the frequency rates of both low-falling and high-falling responses are mostly lower than 90%. There is a perceptual boundary of the two tones, as illustrated by the dashed line in Figure 3, although those stimuli with a higher pitch register, such as stimulus #8 to stimulus #10, slightly increases in low-falling responses. The perceptual boundary is found around stimulus #4:150-90 Hz. As indicated by the dotted circle and lines, the boundaries of the two tasks are slightly different, too. The red dotted line (the lexical task) is on the left of the blue dotted line (the identification task). The difference shows that there are more high-falling responses in the lexical task.

The two-sample T-Test analysis was conducted, and the analysis shows that there is a significant difference between tonal and lexical responses of two stimuli, stimulus #8:190-130 Hz, \( t(58)=1.9531, p=0.0278<0.05 \), and stimulus #9:200-140 Hz, the result of which is very close to the significant level, \( t(58)=1.4634, p=0.0744<0.10 \).

5. Discussion

The effects of three phonetic attributes, lexical frequency, and tone sandhi on speech perception of two Taiwanese falling tones are evaluated based on the current results. Generally, hypotheses (3a), (3c), and (3d) are verified, but hypothesis (3b) is rejected. The current results not only confirm the previously attested phonetic and lexical effects, but also support the effect of tone sandhi. These findings are discussed in detail below.

5.1. Effect of three phonetic attributes

The phonetic effect on speech perception of Taiwanese falling tones is verified by the findings of perceptual boundaries in
the onset pitch and pitch drop continua, as illustrated in Figure 1 and Figure 3. Onset pitch height and pitch drop are found to be two crucial phonetic correlates of Taiwanese falling tones. However, offset pitch height is found to be a trivial characteristic since perturbations of offset pitch do not decrease or increase the linguistic contrast between low-falling and high-falling tones, as indicated by the lack of a perceptual boundary in Figure 2. That is, the offset pitch perturbations do not lead to perceptual confusion of the two falling tones, but the onset pitch and pitch drop perturbations do. The findings are consistent with Yeh and Lin’s [1] results of Hakka and Mandarin falling tones, suggesting that cross-linguistically onset pitch and pitch drop contrast different types of falling tones, but offset pitch does not.

5.2. Effect of lexical frequency and tone sandhi

The lexical effect is verified by the findings of significantly different frequency rates between the identification task and the lexical task, as indicated by the significant levels in Figures 1, 2, and 3. According to our survey of subjective lexical frequency, low-falling tone in 聽 hu21 of 未 聽 ‘too late’ is more familiar than high-falling tone in 聽 hu51 of 政府 ‘government’ to the 30 Taiwanese participants, and the lexical response 聽 ti21 of 心智 ‘mind’ is more familiar than the word 聽 ti51 of 抵抗 ‘resistance’. The subjective judgments are consistent with the results of Chui and Lai’s [15] corpus study in which the ratio of occurrence frequency between the ti21 of ‘mind’ and ti51 of ‘resistance’ is 1:0 and that between hu21 of ‘too late’ and hu51 of ‘government’ is 2:0. Since the lexical response 聽 ti21 of ‘mind’ is more frequent and more familiar, it is more likely to be chosen, as shown by the difference of the IDN and LEX lines in Figure 1.

However, the lexical response of the more frequent ti21 of ‘mind’ is disfavored in the offset and pitch drop continua in those stimuli that are not typical variants of low-falling or high-falling tones. The typical low-falling tone consists of a lower pitch register and a smaller pitch drop, and the typical high-falling tone is made of higher onset pitch, lower offset pitch, and a larger pitch drop. Those stimuli with a higher pitch register and a smaller pitch drop, such as stimulus #8 to stimulus #10 in the offset pitch and pitch drop continua, are more likely to sound like the surface form of high-falling tone, i.e. Tone-55 derived from Tone-51 in (1), in non-final position. i.e. the tone sandhi condition. Although these ambiguous stimuli are predicted to be confusing and the responses of these stimuli are supposedly influenced by lexical frequency based on Yeh and Lin [1], the sandhi condition using the phrase of ‘A of AB’, such as ti51 of /ti51-ŋ/ [ti55-ŋ] ‘resistance’, as proposed by Chuang et al [10], favors the effect of tone sandhi over that of lexical frequency. The effect of tone sandhi is indicated by the significantly more high-falling responses in the lexical task in stimulus #8, #9, and #10 of the offset pitch and pitch drop continua in Figure 2 and Figure 3. In addition, the overall lower frequency rates of both low-falling and high-falling tones may also support the effect of tone sandhi, assuming that such effect leads to perceptual confusion, as argued by Yeh and Tu [5] and Liu and Wang [6]. The frequency rates in Figures 1, 2, and 3 are mostly lower than 90%. Although the percentage is above the chance level, it is lower than Yeh and Lin’s [1] results of Hakka and Mandarin and Zhu and Dodd’s [16] criteria of tonal acquisition.

6. Conclusion

This study confirms the previously attested phonetic and lexical effects on the perception of falling tones, and indicates the psychological reality of Taiwanese tone sandhi. Although this study does not directly address how Taiwanese tone sandhi is represented in speech processing, either allomorphic variations or phonological generalization, the effect of tone sandhi suggests that tone sandhi, together with phonetic and lexical cues, should be considered a processing fundamental.

7. Acknowledgements

This research project is funded by the 2010 Summer Support Fellowship from the College of Arts and Letter at Michigan State University. We thank José Benzí for his assistance with the methodology, and three TAI12 reviewers and Karthik Durvasula for suggestions and comments.

8. References