A Production and Perception Study on Tonal Neutralization in Nanchang

Jiang Liu
Department of Linguistics, University of Kansas, USA
liujiang@ku.edu

Abstract

In a production study of tonal contrasts in lexically stressed but grammatically stressless syllables in Nanchang, a Gan dialect spoken in the city of Nanchang, Jiangxi province, we found that tonal neutralization only occurs in lexically stressless syllables. We argue that the main phonetic ground for such a tonal contrast distribution lies in the rhyme duration difference between syllables with and without lexical stress, namely, lexically stressless syllables have shorter rhyme duration than lexically stressed but grammatically stressless syllables, and the shorter the rhyme duration of a syllable is the fewer tonal contrasts the syllable allows. In terms of perception, we found that different tonal contrasts indeed become neutralized in lexically stressless syllables. However, the neutralization pattern at the perception level is not the same as the one at the production level due to word specific effects.

Index Terms: tonal neutralization, production, perception, lexical stress, grammatical stress, word specific effects, Nanchang

1. Introduction

1.1 The phonetic ground for tonal contrast distribution

This paper investigates the relation between tonal contrasts and syllables’ sonorous rhyme duration that is affected by phonological factors such as stress in Nanchang Chinese, a tone language spoken in Southeast China. There are five lexical tones in Nanchang, transcribed on a five point scale as 42 (yin ping), 24 (yang ping), 45 (yin qu), 213 (shang), and 21 (yang qu). There are two types of stress in Nanchang—lexical stress and grammatical stress. The durational properties of sonorous rhymes in syllables that have different stresses are examined. The tonal contrasts the syllables with different stresses carry are then investigated. This is in order to see whether there exists a relation between the durational properties of syllables and their ability to carry tonal contrasts. By studying the relationship between phonological contrasts (tonal contrasts in this case) and phonetic properties, it is possible to go beyond the typological generalizations about what syllables may carry more phonological contrasts and provide further explanation of why syllables with certain phonological features (stress in this case) carry more phonological contrasts.

According to Zhang’s typological survey of contour tone distribution, the rhyme duration is the crucial factor that licenses contour tones (e.g., rising or falling tone) [8]. Thus, if the rhyme duration of a syllable is affected by phonological parameters such as syllable type or stress, then the tonal contrast the syllable may be affected as well. For example, in standard Thai, CVR syllables (‘C’-Consonant; ‘V’-Vowel; ‘R’-Sonorous consonant) have richer tone-bearing possibilities than CV-O. In particular, CV-O (‘V’-‘long vowel; ‘O’-Obstruent) in Thai cannot host LH or M tones, whereas CVR can host any of the five phonemic tones of the language (H, M, L, HL, LH). In contrast, Navajo shows the opposite tonal contrast distribution: CV: O can host any phonemic tone (H, L, HL, LH), but CVR cannot host HL or LH.

To explain this type of language-specific difference, Zhang proposed that what licenses contour tones is a combination of length and sonority: vowels make better contour hosts than sonorant consonants, but at equal sonority levels, the longer sonorous rhyme is the better carrier. In Zhang’s Navajo data, the rhyme in CVR and the V: portion of CV:O are very close in duration. Since the sonority of V is greater than that of R, it implies CV:O is a better tone carrier than CVR, and the phonology bears this out: CV:O can host more contours. In contrast, in Thai, long vowels are dramatically shorter in closed syllables. As a result, Thai CV:O has considerably less sonorous rhyme duration than CVR, and the difference is enough to compensate for the CVR’s inferior sonority status. The comparison of the tonal contrast difference between Thai and Navajo provides crucial evidence that the degree of shortening in closed syllables is the source of their tonal contrast difference. If rhyme duration really matters for tonal contrast distribution, then when the syllable type and sonority of syllables are the same, stresses that have durational correlate are expected to have influence on tonal contrast as well. Following [8], the current study further studies the durational properties of syllables with different stresses when everything else being equal. We examine whether different stress types in Nanchang have durational correlates, and if so, what happens to the tonal contrasts in syllables both with and without such stress. Following the production study, we conducted a perception study to investigate whether any tonal neutralization found in the acoustic study is mapped onto the perception level.

1.2 Lexical stress and grammatical stress in Nanchang

In Nanchang, certain syllables are lexically stressless, known as qing sheng. Apart from lexical stress, Nanchang, like Standard Chinese, also has grammatical stress. The grammatical stress is introduced as a result of certain grammatical structures, for example, Verb+Noun [V N] disyllabic phrases have the right syllable more stressed than the left one and Noun+Noun [N N] disyllabic compounds have the left syllable more stressed than the right one. The existence of grammatical stress in Standard Chinese has been argued for through different properties of word length in [V N] and [N N], namely, [V N] phrases allow [1 1], [1 2] and [2 2] groupings but not [2 1] (the digit corresponds to the number of syllables of the syntactic category), while [N N] compounds allow [1 1], [2 1] and [2 2] but not [1 2]. This is known as ‘non-head’ stress rule [2]. Nanchang Chinese mirrors Standard Chinese in grammatical stress properties.

A quite recent phonetic study on durational correlate to grammatical stress in Mandarin Chinese is conducted by [3] in which the N in [V N] disyllabic phrases was found to have significantly longer rhyme duration than the N in non-[V N] phrases. The rhyme duration of N in [V N] was not longer than V in [V N] though. The result suggests that the existence...
of grammatical stress in [V N] argued by [2] is reflected by the rhyme duration difference between \( N \) and \( V \) in [V N] phrases. In [3], the authors studied the f0 realization on \( N \) in [V N] disyllables and non-[V N] disyllables. The result showed that there was no tonal neutralization in [V N] or non-[V N] structure, rather, the f0 range on \( N \) in [V N] is wider than that in non-[V N].

The reason we choose Nanchang as the target language for studying durational correlates to different stress types is that we want to examine the durational property of both lexical stress and grammatical stress cross-linguistically. Moreover, any tonal neutralization found in the syllables may tell how the contrast among underlying tones is reduced, namely, are the underlying tones still acoustically different from each other but drastically different from their citation forms? Or, are certain tones merged?

1.3 Research questions

1) Does lexical stress have durational correlates?
2) Does grammatical stress have durational correlates?
3) Are there any tonal contrast reductions in lexically stressless syllables and lexically stressed but grammatically stressless syllables?
4) If tonal contrasts among certain lexical tones are neutralized in a certain type of syllables acoustically, will such tonal neutralization be mapped onto native Nanchang speakers’ perception?

The results from the production study and perception study are reported in sections 2 and 3, respectively.

2. Production study

The production study consisted of two parts. The first part was a durational study on syllables with different types of stress. The second part was a tonal contrast study that examined the tonal contrasts in syllables with and without lexical stress.

2.1 Durational study of syllables with different types of stress

The rhyme durational properties of syllables with three types of stress were examined.

2.1.1 Method

To study the durational correlates of lexical stress and grammatical stress, a wordlist that included \([N N]\) (e.g., 货车 fo45 tsʰa42 ‘cargo truck’), \([V N]\) (e.g., 洗车 ci213 tsʰa42 ‘to wash a car’), and lexically stressless disyllabic words (e.g., 泥巴 ni45 pa0 (42) ‘mud’) was constructed. The underlying tones and the rhymes of the second syllables in the three types of syllables were controlled. In this way, the durational property of different stress types can be examined in the second syllables of \([N N]\), \([V N]\), and lexically stressless disyllabic words. For an \([N N]\) compound, the second syllable bears lexical stress but not grammatical stress (grammatical stress is on the first \( N \)). For a \([V N]\) phrase, the second syllable bears both lexical stress and grammatical stress. For a lexically stressless word, the second syllable does not bear lexical stress. The lexically stressless words used in the current study cannot be analyzed as having a \([V N]\) structure. Therefore, the second syllables in these words do not bear grammatical stress either (e.g., 随嫁 ka45 tsʰaŋ42 (42) ‘dowery’). Table 1 illustrates the stress types the second syllable bears in Nanchang:

<table>
<thead>
<tr>
<th>Target syllable</th>
<th>Lexical Stress</th>
<th>Grammatical Stress</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma ) 2</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Syllable structure: CV</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noun + Noun</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Verb + Noun</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Lexically Stressless</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

In total, 12 word triplets (syllables with three stress types) were used for the recording. All words were embedded in a carrier sentence in order to eliminate final lengthening. Ten native speakers of Nanchang (5 females, 5 males) participated in the recording. After the recording, the rhyme durations of the second syllables of each token were measured using Praat [1].

2.1.2 Results

A one-way repeated measure ANOVA shows that the rhyme durations in different stress conditions are significantly different from each other, as indicated by the asterisks in Figure 1. The mean rhyme durations of the three types of stresses in CV syllables are illustrated in the following graph:

Figure 1. Mean rhyme durations of the second syllable with a CV structure in \([V N]\), \([NN]\), and LS (‘Lexically Stressless’), *: \( p<.05 \), ***: \( p<.001 \).

Although there is a significant durational difference among the three types of stress, the effect sizes are different. In CV syllables, the effect size for the difference between the grammatically stressed condition \( \sigma \) 2 in \([V N]\) and lexically stressed but grammatically stressless condition \( \sigma \) 2 in \([N N]\) is 0.47 whereas the effect size for the difference between the lexically stressed condition \( \sigma \) 2 in \([N N]\) and lexically stressless condition \( \sigma \) 2 in Lexically stressless disyllables) is 0.91. In general, the effect size over 0.5 is considered as a large effect. From the duration study, it can be concluded that both grammatical stress and lexical stress have durational correlates. But the rhyme duration difference between syllables with both grammatical and lexical stress and syllables with only lexical stress is considerably smaller than that between syllables with only lexical stress and syllables without any stress.

2.2 A tonal contrast study of lexically stressed but grammatically stressless syllables vs. lexically stressless syllables

With the findings from the durational study, an acoustic study on tone was conducted to compare the tonal contrasts realized
on lexically stressed but grammatically stressless and lexically stressless syllables. The reason for why we did not compare the tonal contrasts in syllables only with lexical stress to the ones in syllables with both lexical stress and grammatical stress is that it is unlikely there will be any tonal contrast reduction in syllables with either stress type based on [3], though some tonal register feature difference may exist between the syllables with two different stress statuses.

2.2.1. Method

The wordlist for this study included 100 disyllabic words covering 25 tonal combinations. For each tonal combination, there were two words where the second syllables were lexically stressless and two words where the second syllables were lexically stressed but grammatically stressless. The second syllables in both types of words were homophones. The examples in Table 2 provide a quartet for the tonal combination 42+42:

Table 2: An example of word pairs that include lexically stressless second syllables vs. lexically stressed but grammatically stressless second syllables

<table>
<thead>
<tr>
<th>Lexically stressless 2nd syllable</th>
<th>Lexically stressed but grammatically stressless 2nd syllable</th>
</tr>
</thead>
<tbody>
<tr>
<td>冤家 yon42 ka0 (42)</td>
<td>私家 sɿv42 ka42</td>
</tr>
<tr>
<td>'enemy'</td>
<td>'private'</td>
</tr>
<tr>
<td>亲家 tɕin42 ka0 (42)</td>
<td>三家 san42 ka42</td>
</tr>
<tr>
<td>'parents of son/daughter in law'</td>
<td>'three families'</td>
</tr>
</tbody>
</table>

‘0’ in the transcription means lexically stressless. ‘42’ in the brackets next to ‘0’ indicates the morphone’s underlying tone. The words in the left column are labeled as having ‘lexically stressless’ second syllables. This is based on the author’s judgment. The words in the right column are words with an [N N] structure where the second syllables are lexically stressed. Hence, this forms a word pair where one word is lexically stressless while the other is lexically stressed, but both words have the same underlying tonal combination. For the tonal combination of 42+42, there were two word pairs. There were 50 word pairs for the 25 tonal combinations all together.

In the wordlist, 38 lexically stressless disyllabic words that were not in the Nanchang Dialectal Dictionary were added. Thus, to make sure that they really are lexically stressless disyllabic words, a web survey was designed to refine the selection of the lexically stressless words. In the web survey, the recordings of all 38 lexically stressless syllables made by the author were posted online. The second syllable in each disyllable was recorded with either a lexically stressed pronunciation or a lexically stressless pronunciation by the author. Thus, each word had two pronunciations. Eighteen native speakers of Nanchang participated in the online survey by rating the naturalness of the recorded words. Perl scripts were used to process the input and to record the answers to the naturalness judgments made by the speakers. To record the data, a data file was used to record the number of responses to each choice for each stimulus. For example, if 10 people selected the first choice, 6 people selected the second, 4 people selected the third and so on, the data file recorded the exact number of people who made the choice for that particular stimulus. A sketch of the data file is shown below:

As mentioned above, each stimulus had two pronunciations, one with a lexically stressless second syllable and one with a lexically stressed second syllable. If the native speakers only accepted the stressless pronunciation of a word then they would choose ‘very good’ or ‘good’ most of the time for the stressless pronunciation and choose ‘bad’ or ‘very bad’ for its stressed pronunciation. Each choice was assigned with a value rating from 5 to 1 (e.g., ‘very good’ is 5, ‘very bad’ is 1, etc.). For a real lexically stressless word, the stressless pronunciations were expected to have a much higher value than the stressed pronunciations. This means the word was widely recognized by the native speakers as lexically stressless.

For each stimulus, the value of the stressed pronunciation was subtracted from the value of the stressless pronunciation and then divided by 18 (the number of participants). If the obtained value was greater than or equal to 1, then the word was used as a lexically stressless stimulus. If the obtained value was less than 1, then the word was dropped from the lexically stressless wordlist. Twenty one lexically stressless words were selected from the web survey. Together with their corresponding lexically stressed words, they covered 17 out of 25 tonal combinations. Among the 17 tonal combinations, lexical tones that appeared in the second syllables included all five lexical tones in Nanchang.

All of the selected disyllabic words were embedded in a carrier sentences for recording. 10 native speakers of Nanchang participated in the recording. Each word was read twice. F0 of the tones carried by syllables with and without lexical stress was measured. F0 extraction was made by using Yi Xu’s TimeNormalize script. In each token, time normalized f0 values at every 10% of the duration were extracted.

2.2.2 Results

Ten native speakers’ f0 values for each lexical tone were averaged. The average pitch tracks of different lexical tones in the lexically stressed but grammatically stressless syllables are illustrated in Figure 2:

![Figure 2: Average f0 curves of the five lexical tones realized on syllables with lexical stress.](Image)

As can be seen in Figure 2, the five lexical tones produced by speakers in lexically stressed but grammatically stressless syllables were quite different from each other. In order to describe these differences, a two-way Repeated Measures ANOVA with Huynh-Feldt corrections was conducted, with Tone and Point as independent variables. The Tone variable has five levels—Tone42, Tone21, Tone45, Tone24 and Tone213. A significant main effect of the variable Tone would indicate that the two f0 curves representing the tones have different average pitches. The Point variable has eleven levels, representing the eleven points where f0 data were taken. A significant interaction between Tone and Point would indicate that the two curves have different tone shapes. This method of
Comparing two f0 curves has been used by Peng [4] and Zhang & Lai [9]. The result shows a highly significant main effect of Tone (average f0) and Tone shape (interaction between Tone and Points). Pairwise comparisons for each tonal contrast (e.g., 42 vs. 45) shows all tonal contrasts differ in terms of both average f0 and tone shape.

Turning to the tonal contrasts in syllables without lexical stress, the average f0 curves for each underlying tone on the lexically stressless syllables are illustrated in Figure 3.

From Figure 3, it can be seen that all lexical tones become falling tones. Two-way Repeated-Measures ANOVA with Huynh-Feldt corrections still shows a significant main effect of Tone and Point. However, pairwise comparisons show that Tones 42, 45 and 21 are the same in terms of average f0 and tone shape whereas 24 and 213 are the same in terms of average f0 and tone shape. It therefore seems that tonal neutralization has occurred in lexically stressless syllables, at least acoustically. This result indicates that the tonal contrasts of the five lexical tones are reduced to a contrast between two tones on lexically stressless syllables.

For the tonal contrast study, as a different group of word stimuli was used from the durational study, the rhyme duration of the lexically stressed syllables and lexically stressless syllables was measured again. The rhyme duration of the lexically stressless syllables was again significantly shorter of lexically stressless syllables was measured again. The rhyme duration of the syllables turns out to be a consistent predictor of lexically stressless syllables.

For the tonal contrast study, as a different group of word stimuli was used from the production study, in the perception study, we aim to examine whether native speakers of Nanchang can still perceive the difference between lexically stressless syllables with different underlying tones.

### 3. Perception study

Since we found that Tones 42, 45 and 21 were merged and Tones 24 and 213 were merged on lexically stressless syllables in production, in the perception study, we aim to examine whether native speakers of Nanchang can still perceive the difference between lexically stressless syllables with different underlying tones.

#### 3.1 Method

For a perception study on sound neutralization, ideally, a list of minimal pairs needs to be used as stimuli. For instance, a list of Chinese disyllabic words that differed in the underlying tone on the first syllable (Tone2 vs. Tone3) were used in the study of Mandarin Chinese tonal contrast neutralization caused by tone sandhi in word initial position [5]. However, for the perception study on tonal contrast neutralization in Nanchang lexically stressless syllables, it was very difficult to find a sufficient number of minimal pairs that differed in the underlying tone on lexically stressless syllables. The number of lexically stressless disyllabic words listed in the Nanchang Dictionary is limited, let alone the number of minimal pairs that can be found among lexically stressless words. Therefore, we opted to find stimuli for the current perception study by looking up near minimal disyllabic word pairs in the Nanchang Dictionary: the second syllables (lexically stressless syllables) were controlled to be segmentally identical, but with different underlying tones whereas the first syllables (the lexically stressed syllables) were not controlled either segmentally or suprasegmentally. An example of such a near minimal pair is given as follows:

(1a) pa42 tsoŋ0 (213)    (1b) ka45 tsoŋ0 (42)

‘slap’    ‘dowery’

The second syllables in both words are lexically stressless.

In the example above, both words have lexically stressless second syllables that only differ in their underlying tones. The first syllables are always lexically stressed but with different segments and underlying tones. The syllable type of the lexically stressless syllables is CV(R). In the Nanchang Dictionary, 26 near minimal word pairs were found, as used as stimuli for the perception study. Given that our interest lies in understanding whether native Nanchang speakers can differentiate five different underlying tones in lexically stressless syllables, ideally, we need to make 10 tonal contrasts for perception, which consist of all pairwise comparisons for the different underlying tones. However, the 26 near minimal word pairs only covered 7 tonal contrasts. This is illustrated in Table 3:

<table>
<thead>
<tr>
<th>Tones of σ 2 (lexically stressless) in word-2</th>
<th>Tones of σ 2 (lexically stressless) in word-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>42</td>
<td>42</td>
</tr>
<tr>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>213</td>
<td>213</td>
</tr>
</tbody>
</table>

A native speaker of Nanchang who did not participate in the previous production study recorded all the near minimal pairs. With near minimal pairs, it is impossible to conduct an identification task, which was used in the Mandarin Chinese tone sandhi neutralization study [4]. Therefore, a discrimination task was used instead, which told us whether native Nanchang speakers could differentiate lexically stressless syllables that only differed in terms of underlying tones (e.g., tsoŋ0 (213) vs. tsoŋ0 (42)). The paradigm we used for the discrimination task was the ABX/AXB paradigm. In this paradigm, subjects need to decide whether the target X is the same as or similar to the stimuli A or B. The only difference between ABX and AXB is the ordering of the three stimuli. The advantage of the ABX/AXB paradigm is that the subjects do not need to know the nature or the names of the stimuli. This characteristic fits the purpose of our current tonal perception study better: we only test whether subjects can discriminate two lexical tones in the lexically stressless syllables without referring to any tonal category. The software we used was Paradigm (version Beta 6.0, Perception Research Systems 2010). In one block of the discrimination experiment, ABX was used, where subjects listened to a near minimal word pair (one disyllabic word followed by another). After playing the two disyllabic words to the subjects, the second syllable of either the first disyllabic word or of the second disyllabic word was played to the subjects. The task for the subjects was to decide whether the monosyllabic was from the
first word or from the second word and then use the mouse to click the corresponding word icon on a computer screen. The ISI (Inter Stimuli Interval) is 800ms. No time limit was set for the subjects to make the decision. The ITI (Inter Trial Interval) was 1s. To reduce the recency effect, AXB was used in another block of the discrimination experiment. Practice trials were given before the real experiment. Twelve native speakers of Nanchang participated.

3.2 Results

The discrimination accuracy rate results for different tonal contrasts in ABX and AXB are reported together in Table 4. This is because a t-test showed that the accuracy rates in the ABX and AXB blocks were not significantly different from each other (t(12)=4.38, p>.05). The accuracy rate was calculated as averaging the percentage of correct responses out of 12 native speakers’ responses for each word pair and then averaging across word pairs’ correct discrimination percentages for each tonal contrast. A higher accuracy rate indicates that the word pairs can be discriminated by more native speakers of Nanchang.

Table 4. Accuracy rates for different tonal contrasts by word pairs.

<table>
<thead>
<tr>
<th>Tones of Syll2 (lexically stressless) in word-2</th>
<th>42</th>
<th>45</th>
<th>21</th>
<th>24</th>
<th>213</th>
</tr>
</thead>
<tbody>
<tr>
<td>42</td>
<td>58.3%(7)</td>
<td>66.7%(6)</td>
<td>79.5%(2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>50.3%(3)</td>
<td>83%(1)</td>
<td>50.0%(1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>70.8%(6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>70.8%(6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>213</td>
<td>70.8%(6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The number in parentheses indicates the number of near minimal pairs used for a particular tonal contrast. From Table 4, we can clearly see that the accuracy rates are numerically different across different tonal contrasts. Regardless of whether it is the 42-45-21 group or the 24-213 group in which the tones were found to be merged in the previous production study, no consistent accuracy rates were found across the tonal contrasts in the discrimination result. Such accuracy rate variance for different tone pairs suggests that the degree of discrimination for different tonal contrasts varies. However, the difference is not likely due to the tonal contrast condition but rather to word specific effects on tonal neutralization in lexically stressless syllables. The reason is that under those tonal contrasts with relatively high average accuracy rate (e.g., the 24-213 contrast), there were both word pairs that could be distinguished by native Nanchang listeners and word pairs that could not be distinguished based on a $\chi^2$ test. Even for tonal contrasts with relatively low average accuracy rates (e.g., the 42-21 contrast), there still existed both word pairs that can be distinguished and word pairs that cannot be distinguished. Table 5 shows the average accuracy rate together with the number of word pairs that could be distinguished for each tonal contrast.

Table 5. Accuracy rates together with the number of word pairs can be distinguished by native Nanchang listeners for different tonal contrasts.

<table>
<thead>
<tr>
<th>Tones of Syll2 in word-1</th>
<th>42</th>
<th>45</th>
<th>21</th>
<th>24</th>
<th>213</th>
</tr>
</thead>
<tbody>
<tr>
<td>42</td>
<td>58.3% (0/7)</td>
<td>66.7% (2/6)</td>
<td>79.5% (1/2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>50.3% (0/3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>83% (1/1)</td>
<td>50.0% (0/1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>213</td>
<td></td>
<td>70.8% (3/6)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The number in the parentheses indicates the proportion of the near minimal pairs for each tonal contrast that could be distinguished by native Nanchang listeners (e.g., for 42-21, two out of six near minimal word pairs could be distinguished by the listeners). Table 5 shows that in total there were 7 near minimal word pairs that could be distinguished by native Nanchang listeners. The uneven distribution of the distinguishable word pairs helps explain the accuracy rate difference across tonal contrasts. Since the word pairs that could be discriminated are accompanied by a high accuracy rate, those word pairs raise the overall accuracy rate for the tonal contrast. However, for those tonal contrasts with a high overall accuracy rate, it is not the case that all near minimal pairs could be discriminated by the native listeners. Thus, it suggests that there is a word specific effect on the discrimination. An alternative explanation for the uneven distribution of distinguishable word pairs across different tonal contrasts is that tonal coarticulation effect varies across different word pairs since the tones of the first syllables in the near minimal pairs were not controlled. However, we argue that tonal coarticulation effect may not be the reason for the following reason. Not only for the distinguishable near minimal pairs, but also other near minimal pairs had different tones precede the target tones. However, native speakers’ discrimination was much poorer for the other near minimal pairs.

The average pitch track results of the five underlying tones on lexically stressless syllables also show a sign of the word specific effect on the production of tones. Figure 4 illustrates the average pitch tracks across tokens for each underlying tone.

Figure 4. Average f0 curves in lexically stressless CV(R) syllables used in the perception study

The average f0 curves of the five underlying tones in Figure 4 do not look the same as the average f0 pattern that we found in the production study, namely, 42-45-21 were merged and 213-24 were merged in terms of both average f0 and tone shape. It seems that a larger average pitch track difference existed in the lexically stressless syllables used for the
perception study. This larger difference may have come from the pitch differences between lexically stressless syllables of certain word pairs. The acoustic differences in these word pairs may affect the discrimination task results.

4. Discussion

The production study on tonal neutralization in Nanchang showed a tonal neutralization pattern in which five lexical tones were reduced to two tones on lexically stressless syllables. Linking this tonal neutralization pattern to rhyme duration shortening in the lexically stressless syllables, we argued that the tonal contrast distribution in Nanchang is related to the rhyme duration, which is affected by the lexical stress status. Later in the perception study, we found a word specific effect on the discrimination of tones in lexically stressless syllables that in several word pairs can be distinguished by native Nanchang listeners but the majority of the word pairs cannot be distinguished. Interestingly, in the current study, we found that the tonal neutralization pattern in the lexically stressless syllables we used for the perception study is different from the one we found in the production study. In this section, we discuss the relationship between the tonal merging pattern in the production study and the word specific effect in the perception study under the framework of certain speech production models proposed in the literature.

Using the speech production model proposed by Levelt et al. [4], we may treat the phonetic implementation of phonological features in a sequential fashion. First, a lexeme is retrieved from the lexicon. Second, the lexeme is phonologically encoded in a phonological buffer. Then the phonologically encoded lexeme provides an input to the phonetic implementation module to compute the degree and timing of articulatory gestures. In terms of tone production, we may consider both the rhyme duration and f0 curves realized in the syllables are computed by the phonetic implementation rule after the prosodic structure of a syllable is available in the phonological buffer (e.g., the feature [stress] is specified, the tonal sequence is specified, etc.). If this model can be applied to tone production, then we would expect the surface f0 curves of different tones to pattern together purely based on the phonological structure because in a modular feed-forward model such as Levelt’s, the categorical form of the lexeme determines the phonetic outcome entirely. This is actually true for our production data where the underlying tone 42, 45 and 21 were neutralized in the lexically stressless syllable context and 24 and 213 were merged. However, the mismatch between the production and perception of tonal neutralization suggests certain words may differ from other words in terms of tonal reduction. Such difference makes native speakers be able to perceive the difference between different tones in the lexically stressless syllables. Therefore, we conclude rhyme duration indeed correlates to tonal contrasts’ licensing but the detailed mechanism of tonal reduction is not only affected by phonological feature such as [stress], it may be influenced by other linguistic factors as well such as word frequency, stylistics and so on.

5. References


documented a vowel shift in progress in Quebecois French. She found that a particular group of words failed to shift despite exhibiting the phonological sequences targeted in the change. These words were a group of semantic associates, representing organs of the church, the military, and the schools. Yaeger-Dror was not able to identify any phonological properties shared by these words that distinguished them from words that did undergo the shift. Similarly, the tonal neutralization in certain lexically stressless disyllables cannot be completely predicted by the phonological context and the underlying tonal categories. Many factors other than phonological context (e.g., word frequency, pragmatics/social stylistics, etc.) may help predict the phonetic outcome.

In sum, the production study showed an overall trend of tonal neutralization in the lexically stressless syllables due to rhyme shortening, namely, tones 42, 45 and 21 are merged whereas 24 and 213 are merged. However, the mismatch between the production and perception of tonal neutralization suggests certain words may differ from other words in terms of tonal reduction. Such difference makes native speakers be able to perceive the difference between different tones in the lexically stressless syllables. Therefore, we conclude rhyme duration indeed correlates to tonal contrasts’ licensing but the detailed mechanism of tonal reduction is not only affected by phonological feature such as [stress], it may be influenced by other linguistic factors as well such as word frequency, stylistics and so on.