# A PRODUCTION AND PERCEPTION STUDY ON TONAL NEUTRALIZATION IN NANCHANG CHINESE

**Jiang Liu and Jie Zhang**
*The University of Kansas, USA*

ABSTRACT

In a production study of tonal contrasts in lexically stressed but grammatically stressless syllables vs. lexically stressless syllables in Nanchang, a Gan dialect spoken in the city of Nanchang, Jiangxi province, we found that tonal neutralization only occurred in lexically stressless syllables. We argue that the main phonetic ground for such a tonal contrast distribution lies in the rime duration difference between syllables with and without lexical stress, namely, lexically stressless syllables have shorter rime duration than lexically stressed but grammatically stressless syllables, and the shorter the rime duration of a syllable is, the fewer tonal contrasts the syllable allows. Linear mixed-effect models showed that the effect size of lexical stress on tonal neutralization varied across different lexical items, suggesting that there was a word-specific effect towards tonal neutralization in lexically stressless syllables. In terms of perception, we found that different tonal contrasts became neutralized in most lexically stressless syllables. However, tonal neutralization did not occur in a handful of lexical items due to the word-specific effect.

SUBJECT KEYWORDS
tonal neutralization, production, perception, lexical stress, grammatical stress, linear mixed-effect models, word-specific effect, Nanchang

## 1. INTRODUCTION

This paper investigates the relation between tonal contrasts and syllables' sonorous rime duration that is affected by 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)(Li 1995).There are two types of stress in Nanchang—lexical stress and grammatical stress. The durational properties of sonorous rimes in syllables that have different stresses are examined. The tonal contrasts the syllables with different stresses carry are then investigated to see whether there is 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, we intend 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. Another focus of the current study is to examine whether the relation between the sonorous rime duration and the tonal contrasts (if there is any) is mediated by lexical item differences, in other words, whether the potential effect of stress on tonal contrasts' licensing is across the board or word-specific.

According to Zhang's typological survey of contour tone distribution, the rime duration is the crucial factor that permits contour tones (e.g., rising or falling tone) (Zhang 2002). Thus, if the rime duration of a syllable is affected by phonological parameters such as syllable type or stress, then the tonal contrasts on 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 carry 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 rime is the better carrier. In Zhang's Navajo data, the rime 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 considerably shorter in closed syllables. As a result, Thai CV:O has considerably less sonorous rime 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 rime duration really matters for tonal contrast distribution, then when the syllable type and sonority of syllables are the same, stresses that have durational correlates are expected to have an influence on tonal contrasts as well.

Following Zhang (2002), the current research further studies the durational properties of syllables with different stresses when everything else is 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. We also extend Zhang's thesis to include any possible word-specific effect.

## 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 more stress on the final syllable while Noun+Noun [N N] disyllabic compounds have more stress on the initial syllable. 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 the 'non-head' stress rule (Duanmu 2007). For example, in Standard Chinese, (1c) is not allowed as N cannot have more syllables than V in [V N] whereas (2b) is not allowed as the second N cannot have more syllables than the first N in [N N].

| | | | |
|---|---|---|---|
| (1) [Verb | Noun] | (2) [Noun | Noun] |
| 'plant' | 'garlic' | 'coal' | 'store' |
| (a) zhong-zhi | da-suan | (a) mei-tan | shang-dian |
| (b) zhong | da-suan | *(b) mei | shang-dian |
| (c) zhong | suan | (c) mei | dian |
| *(d) zhong-zhi | suan | (d) mei-tan | dian |

Nanchang Chinese mirrors Standard Chinese in grammatical stress properties, thus, we argue that grammatical stress exists in Nanchang as well.

## 1.3 DURATIONAL CORRELATES OF LEXICAL STRESS AND GRAMMATICAL STRESS

It has been found that lexically stressless syllables to be much shorter than lexically stressed syllables in standard Chinese. On average, the sonorous rime of the light syllables is only 61% of the length of the sonorous rime in full syllables (Chen and Xu 2006).

A recent phonetic study on the durational correlate for grammatical stress in Mandarin Chinese showed that the N in [V N] disyllabic phrases had a significantly longer rime duration than the N in non-[V N] phrases. The rime duration of N in [V N] was not longer than V in [V N], however (Lai, Sui, and Yuan 2010). The result suggests that the existence of grammatical stress in [V N] is reflected by the rime duration difference between N in [V N] and N in non-[V N], rather than being reflected as a longer rime duration of N than V in [V N]. The f0 realization of different lexical tones on N in [V N] disyllables and non-[V N] disyllables was examined as well. 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] was wider than that in non-[V N].

Based on the acoustic study of lexical stress and grammatical stress in Mandarin Chinese, we expect syllables' rime duration to be affected by both lexical and grammatical stress in Nanchang Chinese. Moreover, we will examine whether the tonal contrasts are affected by the change of rime duration in syllables with different stresses. If any tonal neutralization occurs due to rime duration shortening, we will examine how the contrast among underlying tones is reduced. For example, are the underlying tones still acoustically different from each other even though they are realized differently from their citation forms? Are certain tones merged?

## 1.4 VARIABILITY OF PHONETIC IMPLEMENTATION OF LEXEMES

As more and more word-specific phonetics has been documented in recent years (see Pierrehumbert 2002 for a review), the current study also examines whether the expected stress effect on rime duration and tonal contrasts are across the board or variable among different lexical items. In terms of speech production, a phonetic implementation system relates lexemes (i.e., phonological representations of words) to the time course of phonetic parameters in particular speech acts. Because of distributional properties, sociostylistics, word frequency and other factors, the phonetic realization of the same phonological features may vary in different lexical items. Word-specific phonetic variability is not new cross-linguistically. Yaeger-Dror and Kemp (1992) and Yaeger-Dror (1996) documented a vowel shift in progress in Quebecois French. They found that a particular group of words failed to shift despite the fact that they exhibited 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 which did undergo the shift. A more revealing case for word specific effects on tone production comes from Pingxiang, a Gan dialect spoken in Jiangxi province. Pingxiang has four lexical tones—yinping (13), yangping (44), shang (35) and qu (11). Wei (1999) reported two types of tone sandhi that occur in lexically stressless syllables. In the first type, the surface form of the sandhi tone is purely determined by its corresponding underlying tone, namely, yinping (13) and yangping (44) become 44 whereas shang (35) becomes 45 and qu (11) becomes 1. This type of tone sandhi is called open-class tone sandhi ('guangyongshi 广用式'). It applies to words without any affixes, especially content words. In the second type of tone sandhi, the surface form of the sandhi tone is purely determined by the preceding tone. The sandhi tone preceded by yinping (13), yangping (44), and qu (11) is 5 whereas the sandhi tone preceded by shang(35) is 4. This type of tone sandhi only applies to words with suffixes '-zi', '-zai', and '-gu', and is thus called close-class tone sandhi ('zhuanyongshi 专用式'). Comparing these two types of tone sandhi, we can see the sandhi tones that occur in the suffixes seem to be shorter than the ones that occur in other lexically stressless syllables. Interestingly, 10 years later, Wei (2000) reported a new set of words that do not have suffixes, but show close-class tone sandhi. For example, the second tone in content words such as ‘lau35 su35’(mouse), ‘kœ35 tɕi35’ (a type of Chinese medicine) and ‘tʂo35 tʂu35’ (arm) undergoes close-class tone sandhi to become 4. In other words, the sandhi tone in the content words is not determined by its underlying tone. Based on this new finding, Wei (2000) claimed that the condition for close-class tone sandhi is no longer limited to words with suffixes.

In the current study, we examine whether the effect size of stress on syllables' rime duration and tonal contrasts is consistent across different lexical items. To examine the potential word specific effect on rime duration and tonal contrast, we used linear mixed-effect models to examine whether there is any random effect of item, which is a sign of word-specific effect (see more details in Section 2.1.2).

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.

## 1.5RESEARCH QUESTIONS

1) Does lexical stress have durational correlates in Nanchang?
2) Does grammatical stress have durational correlates in Nanchang?
3) Is the effect of stress on rime duration consistent for all lexical items?
4) Is there any tonal contrast reduction in lexically stressless syllables or lexically stressed but grammatically stressless syllables?
5) Is the effect of stress on tonal contrast the same for all lexical items?
6) 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 rime 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 ɕi213 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 rimes of the second syllables in the three types of words 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., 嫁妆 ka45 tsɔŋ0 (42) ‘dowry’) (see Appendix I for the full word list). Table 1 illustrates the stress types the second syllable bears in Nanchang:

**Table 1.** Stress types on the second syllable in [N N], [V N] and lexically stressless disyllabic words.

| Target syllable: σ2<br>Syllable structure: CV | Lexical Stress | Grammatical Stress |
|---|---|---|
| Noun + Noun | + | - |
| Verb + Noun | + | + |
| Lexically Stressless | - | - |

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 rime durations of the second syllables of each token were measured using Praat (Boersma and Weenink 2003).

2.1.2 RESULTS

All data were analyzed using R (R Development Core Team, 2009) and the R packages lme4 (Bates and Maechler 2009) and language R (Baayen 2008). We analyzed the data by using linear mixed-effect models, which take into account the random effect of subjects and items in the analysis. This is important as recent studies have shown that different individuals behave differently for the same task and the same manipulation has different effect sizes on the stimuli used in the experiment. For example, in terms of the random effect of subjects, in a lexical decision experiment, some subjects showed familiarization effect, making responses faster and faster throughout the experiment whereas some subjects showed a fatigue effect, making responses slower and slower during the experiment (Baayen 2008). In terms of the random effect of items, in a reading latency study, certain English nouns caused much longer latency for non-native English speakers even though they knew those nouns in a vocabulary test (Clark 1973). In the current study, since we are interested in whether the effect of stress on rime duration is consistent across all lexical items, we compared two models: in the first model, we included stress as the fixed effect (the factor that was manipulated) and only subject as the random effect; in the second model, the fixed effect was the same but we included both subject and item as random effects. The comparison between the two models can inform us whether there is a significant random effect of items/words. If so, it will suggest that the effect size of stress on rime duration is different across different lexical items.

Using lexically stressless syllables and lexically stressed but grammatically stressless syllables as the baselines respectively,[1] both linear mixed-effect models showed that the rime durations in different stress conditions were significantly different from each other, as indicated by the asterisks in Figure 1. The mean rime durations of the three types of stresses in CV syllables are illustrated in Figure 1:

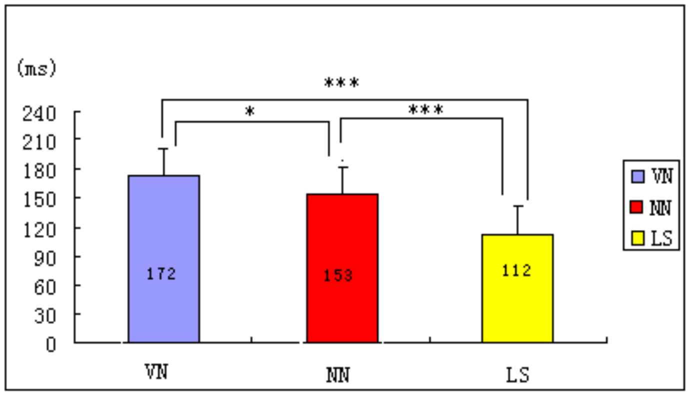


Figure 1. Mean rime durations of the second syllables with a CV structure in [VN], [NN], and LS ('Lexically Stressless'), *: p<.05, ***: p<.001.

The results showed that the rime duration of syllables with both grammatical and lexical stress was significantly longer than that of syllables with only lexical stress, which in turn was significantly longer than that of syllables without either lexical stress or grammatical stress. The results suggested that both grammatical stress and lexical stress had durational correlates in Nanchang. Although there was a significant durational difference among the three types of stress, the effect sizes were different. The standardized effect size *r* for the difference between the grammatically stressed condition (σ2 in [V N]) and grammatically stressless condition (σ2 in [N N]) was 0.47 whereas *r* for the difference between the lexically stressed condition (σ2 in [N N]) and lexically stressless condition (σ2 in Lexically stressless disyllables) was 0.91. In general, the effect size over 0.5 is considered as a large effect. Therefore, we argue that lexical stress in Nanchang has a more robust durational correlate.

To assess the random effect of words, we performed a likelihood ratio test comparing the model with subject as the only random effect to the model with both subject and item as random effects. The two models significantly differed from each other (MCMC estimated $p < .001$). The result suggested that the effect size of stress on the rime duration was different across different words.

## 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. We did not compare the tonal contrasts in syllables with only lexical stress to those in syllables with both lexical and grammatical stress as it is unlikely there will be any tonal contrast reduction in syllables with either stress type based on Lai, Sui, and Yuan (2010).

### 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 in which the second syllables were lexically stressless and two words in which 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 and lexically stressed but grammatically stressless second syllables.

| 42+42 | |
|---|---|
| Lexically stressless 2nd syllable | Lexically stressed but grammatically stressless 2nd syllable |
| 冤家 yon42 *ka0 (42)*<br>*'enemy'*<br>亲家 tɕʰin42 *ka0 (42)*<br>*'parents of son/daughter in law'* | 私家 sɿ42 *ka42*<br>*'private'*<br>三家 san42 *ka42*<br>*'three families'* |

(Note: /tɕ/ is a palatal fricative; /ɿ/ is an apical vowel /i/)

'0' in the transcription indicates lexical stresslessness. '42' in the brackets next to '0' indicates the morpheme's underlying tone, which is the tone realized in citation form. The words in the left column are labeled as having 'lexically stressless' second syllables. This is based on the first 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 (Li 1995) were added. 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 disyllables made by the first 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 participants selected the first choice, 6 participants selected the second, 4 participants selected the third and so on, the data file recorded the exact number of participants who made the choice for that particular stimulus. A sketch of the data file is shown below:

| | 'verygood', | 'good', | 'ok', | 'bad', | 'verybad' |
|---|---|---|---|---|---|
| Stimulus1 | 10, | 6, | 4, | 0, | 0 |

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 (5 = 'very good', 1 = 'very bad'). For a real lexically stressless word, the stressless pronunciation was expected to have a much higher value than the stressed pronunciation.

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 (see Appendix II for the full disyllable list).

All selected disyllabic words were embedded in carrier sentences for the recording. Ten 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:

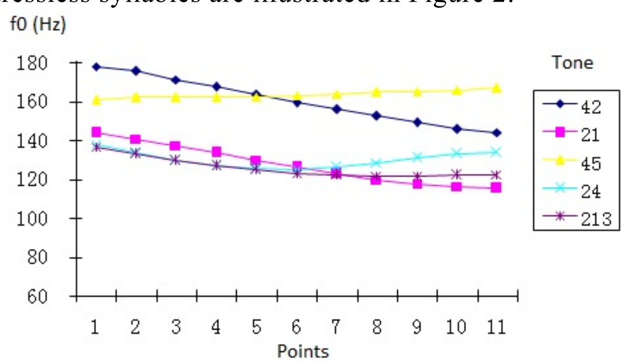


**Figure 2.** Average f0 curves for the five lexical tones on syllables with lexical stress.

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—Tones 42, 21, 45, 24, and 213. 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 2000) and (Zhang and Lai 2010). The result showed 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) showed all tonal contrasts differed in terms of both average f0 and tone shape.

In order to examine whether the random effect of words is significant, for each tonal contrast, we reran the analysis by using linear mixed-effect models with and without item as a random effect. For each tonal contrast, the linear mixed effect model showed that there was a main effect of Tone and a significant interaction between Tone and Point, similarly to the repeated measures ANOVA. In terms of the random effect of words, only the tonal contrast 42 vs. 45 had a significant effect (MCMC estimated p <.05), suggesting that the effect size of Tone and Tone by Point interaction was different across the words for this particular tonal contrast.

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:

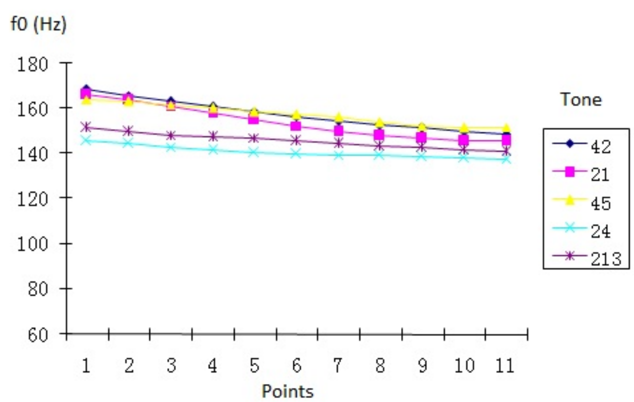


**Figure 3.** Average f0 curves forthe five underlying tones on lexically stressless syllables.

From Figure 3, it can be seen that all lexical tones became falling tones. A two-way Repeated-Measures ANOVA with Huynh-Feldt corrections still showed a significant main effect of Tone and Point. However, pairwise comparisons showed that Tones 42, 45 and 21 were identical in terms of average f0 and tone shape whereas 24 and 213 were identical in terms of average f0 and tone shape. It therefore seemed that tonal neutralization occurred in lexically stressless syllables. This result indicated that the tonal contrasts of the five lexical tones were reduced to a contrast between two tones on lexically stressless syllables.

To examine whether the random effect of words was significant, for each tonal contrast, we reran the analysis by using linear-mixed effect models with and without item as a random effect. The linear mixed effect models showed the same result as the repeated measures ANOVA in that there was no main effect of Tone or Tone by Point interaction for the tonal contrasts among 42, 45 and 21 and the tonal contrast between 24 and 213. In terms of the random effect of words, there were significant effects for tonal contrasts 42 vs. 21 (MCMC estimated $p < .001$), 45 vs. 21 (MCMC estimated $p < .05$), 24 vs. 21 (MCMC estimated $p < .05$), 42 vs. 213 (MCMC estimated $p < .01$) and 24 vs. 213 (MCMC estimated $p < .01$). The significant random effects for these tonal contrasts indicated a strong word-specific effect towards tonal neutralization in lexically stressless syllables. In other words, even though the aggregate results showed certain tonal contrasts were neutralized, the tonal neutralization was mediated by the random effect of words.

The tonal contrast study shows a stronger word specific effect on the tonal realization in lexically stressless syllables relative to that in lexically stressed but grammatically stressless syllables. The different degrees of tonal neutralization in lexically stressless syllables due to the word specific effect echo the random effect of words on rime duration in the durational study. It is possible that the source of the different degrees of tonal neutralization comes from the different degrees of rime shortening in lexically stressless syllables due to word specificity. Following the production study, we conducted a perception study to examine whether tonal neutralization occurs at the perception level, and whether any neutralization is word-specific as well.

## 3. PERCEPTION STUDY

Given that we found that Tones 42, 45 and 21 were merged and Tones 24 and 213 were merged on lexically stressless syllables in production and the degrees of tonal neutralization varied across the words, 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, moreover, we want to examine whether any word-specific effect occurs to tonal neutralization at the perception level as well.

3.1 METHOD

For a perception study on sound neutralization, ideally, a list of minimal pairs should be used as stimuli. For instance, a list of Chinese disyllabic words that differed in the underlying tone on the first syllable (Tone 2 vs. Tone 3) were used in the study of Mandarin Chinese tonal contrast neutralization caused by tone sandhi in word initial position (Peng 2000). 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 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 in (3):

(3) a. pa42 tsoŋ0 (213) b. ka45 tsoŋ0 (42)
巴掌 嫁妆
'slap' 'dowry'
(The second syllables in both words are lexically stressless.)

In (3), both words have lexically stressless second syllables that only differ in their underlying tones. The first syllables are always lexically stressed but have 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 (see Appendix III for the full word list) and 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 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 3.** The number of near minimal pairs with lexically stressless syllables for each tonal contrast.

| | | Syll2 (lexically stressless) in word2 | | | | |
|---|---|---|---|---|---|---|
| Syll2 (lexically stressless) in word1 | | 42 | 45 | 21 | 24 | 213 |
| | 42 | | 7 | 6 | 0 | 2 |
| | 45 | | | 3 | 0 | 0 |
| | 21 | | | | 1 | 1 |
| | 24 | | | | | 6 |
| | 213 | | | | | |

A native speaker of Nanchang who did not participate in the previous production study recorded all near minimal pairs. With near minimal pairs, it is impossible to conduct an identification task. Therefore, a discrimination task was used instead, which informed 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 stimulus 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 are only interested in testing whether subjects can discriminate two lexical tones in the lexically stressless syllables, not what tonal categories they belong to. 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 the two disyllabic words were played to the subjects, the second syllable of either the first or second disyllabic word was played to the subjects. The task for the subjects was to decide whether the monosyllable was from the first or the second word and then use the mouse to click on the corresponding word icon on a computer screen. The Inter Stimuli Interval (ISI) was 800ms. No time limit was set for the subjects to make the decision. The Inter Trial Interval (ITI) 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 in the discrimination task.

3.2RESULTS

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 ($t(12)=.438$, $p>.05$). For each tonal contrast, the accuracy rate was calculated by first averaging the correct discrimination percentages of the 12 speakers' responses for each word pair and then averaging across the word pairs. 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.

| Tones of Syll2 (lexically stressless) in word-1 | Tones of Syll2 (lexically stressless) in word-2 | | | | | |
|---|---|---|---|---|---|---|
| | | 42 | 45 | 21 | 24 | 213 |
| | 42 | | 58.3%(7) | 66.7%(6) | | 79.5%(2) |
| | 45 | | | 50.3%(3) | | |
| | 21 | | | | 83%(1) | 50.0%(1) |
| | 24 | | | | | 70.8%(6) |
| | 213 | | | | | |

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 were numerically different across different tonal contrasts. Regardless of whether it is the 42-45-21 group or the 24-213 group, where the tones were found to be merged in the production study, no consistent accuracy rates were found across the tonal contrasts in the discrimination result. Such accuracy rate variance for different tone pairs suggested that the degree of discrimination for different tonal contrasts varied. However, the difference was not likely due to the tonal contrast condition but rather to the word-specific effect on tonal neutralization in lexically stressless syllables. The reason is that under those tonal contrasts with a 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 those that could not be distinguished based on a $\chi^2$ test.[2] Even for tonal contrasts with relatively low average accuracy rates (e.g., the 42-21 contrast), there still existed both word pairs that could be distinguished and word pairs that could not be distinguished. Table 5 shows the average accuracy rates 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.

| Tones of syll2 (lexically stressless) in word-1 | Tones of syll2 (lexically stressless) in word-2 | | | | | |
|---|---|---|---|---|---|---|
| | | 42 | 45 | 21 | 24 | 213 |
| | 42 | | 58.3%(0/7) | 66.7%(2/6) | | 79.5%(1/2) |
| | 45 | | | 50.3%(0/3) | | |
| | 21 | | | | 83%(1/1) | 50.0%(0/1) |
| | 24 | | | | | 70.8%(3/6) |
| | 213 | | | | | |

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 were accompanied by a high accuracy rate, those word pairs raised the overall accuracy rate for the tonal contrast. However, for those tonal contrasts with a high overall accuracy rate, it was not the case that all near minimal pairs could be discriminated by the native listeners. Thus, it suggested that there was 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 varied across different word pairs as the tones of the first syllables in the near minimal pairs were not controlled. However, we argue against this alternative as different tones precede the target tones not only for distinguishable near minimal pairs, but also for non-distinguishable near minimal pairs. Therefore, it seems that certain word pairs were simply more difficult to distinguish than other pairs regardless of the different preceding tones.

The average pitch track results of the five underlying tones on lexically stressless syllables also showed 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.

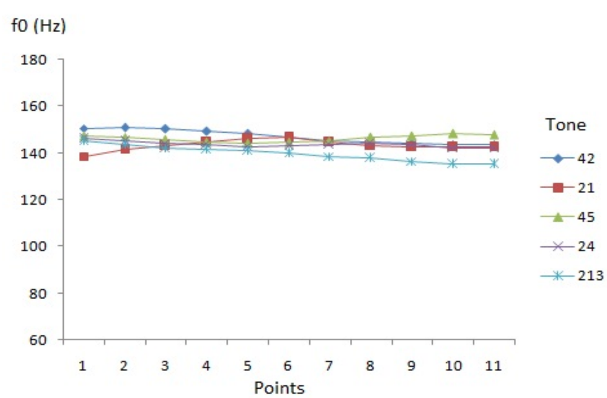


**Figure 4.** Average f0 curves on lexically stressless CV(R) syllables used in the perception study.

Under visual inspection, the average f0 curves for the five underlying tones in Figure 4 were not exactly 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. Nevertheless, the result was quite different from the tonal contrasts in lexically stressed but grammatically stressless syllables where all five tones were different from each other in terms of average f0 and tone shape. It seems that different tones still underwent tonal neutralization in the lexically stressless syllables used for the perception study. Based on the random effect of words on rime duration and tonal contrasts we found in the previous production study, we argue that the tones of certain words used in the perception study did not become reduced as much as the tones in other words. That is to say, syllables in certain words may resist undergoing tonal reduction to various extents even if they are lexically stressless. Therefore, a larger acoustic difference between tones in certain word pairs may make listeners be able to distinguish the different underlying tones. In order to confirm the potential word-specific effect on the realization of lexical tones in the lexically stressless syllables, we conducted a multiple regression analysis to find what acoustic differences among the word pairs contributed to the discrimination difference between different word pairs.

## 4. MULTIPLE REGRESSION ANALYSIS OF THE TONAL DISCRIMINATION RESULTS

Acoustic differences existed between the lexically stressless syllables for virtually all near minimal word pairs as the word tokens from which the lexically stressless syllables were extracted were recorded in different carrier sentences. It is reasonable to believe that for the 7 distinguishable near minimal word pairs, there must exist more salient acoustic differences between the lexically stressless syllables relative to the acoustic differences between indistinguishable lexically stressless syllables. Therefore, we first quantified the acoustic differences between the tones that surfaced on the lexically stressless syllables in every near minimal word pair in terms of pitch onset, mean f0, rime duration and pitch direction. Then we used the acoustic differences as predictors for the discrimination accuracy rates of the word pairs in a hierarchical multiple regression analysis to decide which acoustic parameters were the significant predictors for the accuracy rates. This allows us to test whether the word specific effect on tonal discrimination was manifested via the differences of certain acoustic parameters that existed in the word pairs.

We included four acoustic parameters as the predictors in the multiple regression analysis—pitch onset, mean f0, pitch direction and rime duration. These acoustic parameters were selected from previous tone perception studies in which the pitch direction was found to weigh more for tone identification (Gandour 1978, 1981). In our case, including these acoustic parameters will inform us whether the same acoustic parameter is the most salient cue for tone discrimination in the lexically stressless syllables. The acoustic difference between the lexically stressless syllables in the near minimal word pairs in terms of the four acoustic parameters were quantified as follows. First, the rime duration of the lexically stressless syllables were measured using a Praat script by Mietta. Then the rime duration difference between two lexically stressless syllables can be calculated. Second, the maximum, minimum and mean f0 of the tones on the lexically stressless syllables were measured by using a Praat script by Hirst (2011). The contour feature of a tone was judged by the first author. Based on the contour feature, the pitch direction values of tones were calculated as follows:

(4) a. Pitch direction (falling tone)=MaxF0-MinF0
b. Pitch direction (rising tone)=MinF0-MaxF0

In this way, the pitch direction value of a falling tone would be a positive value whereas the pitch direction value of a rising tone would be a negative value. For level tones, it was difficult to judge whether a tone was completely level or with a slight rise or fall, but in the pitch measurement, the pitch direction value would be very close to zero. In our calculation for a tone that looked like a level tone, we referred to the pitch onset and pitch offset values in the 11 time normalized pitch values measured by Yi Xu's Praat script to decide whether the tone was slightly rising or falling. Based on this judgment, we decided whether we should use (4a) or (4b) to calculate the pitch direction value. The pitch direction difference between two tones would be quantified as the absolute value of the difference between the tones' pitch direction values:

(5). Pitch direction difference between two tones

ΔPitch direction =| Pitch direction (Tone X) - Pitch direction (Tone Y) |

There is one near minimal word pair in which the tone in one lexically stressless syllable was judged to be a dipping tone (falling then rising) and the tone of the other lexically stressless syllable was judged as a falling tone. In this case, instead of using the Max F0 and Min F0 for calculating the direction value difference we used the f0 value at pitch onset, turning point and pitch offset f0 values for the calculation. We split the complex contour tone into two parts. The pitch direction value of the falling part was calculated by subtracting the lowest f0 value out of the 11 time normalized pitch values (the f0 value at the turning point) from the pitch onset f0 whereas the pitch direction value of the rising part was calculated by subtracting the f0 value at the pitch offset from the f0 value at the turning point. For its falling tone counterpart, we split the tone into two falling parts as well in order to make a comparison with the dipping tone's pitch direction value. The pitch direction values for the two parts were calculated by subtracting the f0 value proportionally to the dipping tone's turning point from the pitch onset f0 value for the first part and subtracting the pitch offset f0 value from the f0 value at the same time point as the dipping tone's turning point for the second part. The pitch direction difference between the two tones was then the sum of the pitch direction difference of the two parts. After the measurements and comparisons, the acoustic differences between lexically stressless syllables in the 26 near minimal word pairs were quantified. The acoustic differences in terms of the four parameters and the accuracy rate for each near minimal word pair are illustrated in the following table:

**Table 6.** Acoustic differences between each near minimal word pair used for the discrimination task in terms of four parameters and the percentage of native speakers of Nanchang who correctly discriminated the word pair out of 12 speakers.

| Stimulus Pair | Word 1 | Word 2 | Duration difference (ms) | Mean f0 difference (Hz) | Tone Direction difference (Hz) | Onset Pitch difference (Hz) | Perception Accuracy Rate (%) |
|---|---|---|---|---|---|---|---|
| 1 | 嫁妆 | 胜仗 | 52 | 23 | 28 | 36.75 | 58 |
| 2 | 剪刀 | 捡到 | 75 | 9 | 41 | 3.96 | 42 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 3 | 东西 | 把戏 | 78 | | | | 67 |
| 4 | 棉花 | 计划 | 16 | 33 | 5 | 30.95 | 75 |
| 5 | 官司 | 道士 | 35 | | | | 50 |
| 6 | 雷锋 | 裁缝 | 22 | 20 | 48 | 47.17 | 100 |
| 7 | 姐夫 | 师父 | 10 | 12 | 8 | 10.56 | 42 |
| 8 | 东西 | 关系 | 6 | | | | 50 |
| 9 | 经费 | 机会 | 52 | 2 | 2 | 0.72 | 42 |
| 10 | 收益 | 手艺 | 55 | 10 | 18 | 3.95 | 42 |
| 11 | 机器 | 肚脐 | 26 | 12 | 52 | 29.24 | 100 |
| 12 | 宽敞 | 肥肠 | 107 | 5 | 45 | 10.42 | 92 |
| 13 | 云彩 | 招财 | 7 | 18 | 30 | 32.84 | 42 |
| 14 | 姐夫 | 马虎 | 23 | | | 8.22 | 67 |
| 15 | 筲箕 | 挂记 | 5 | 12 | 49 | 64.52 | 67 |
| 16 | 比方 | 作坊 | 46 | 4 | 7 | 3.86 | 58 |
| 17 | 记性 | 良心 | 16 | 7 | 30 | 12.36 | 58 |
| 18 | 母舅 | 泥鳅 | 22 | 1 | 45 | 16.32 | 83 |
| 19 | 高粱 | 较量 | 51 | 28 | 30 | 13.12 | 67 |
| 20 | 算盘 | 背叛 | 21 | 4 | 6 | 0.43 | 58 |
| 21 | 套鞋 | 螃蟹 | 40 | 20 | 39 | 12.64 | 83 |
| 22 | 母舅 | 要求 | 31 | 29 | 43 | 8.79 | 83 |
| 23 | 凉拌 | 算盘 | 4 | 8 | 40 | 6.08 | 50 |
| 24 | 欺负 | 马虎 | 6 | 12 | 3 | 1.66 | 50 |
| 25 | 巴掌 | 嫁妆 | 75 | 24 | 47 | 33.67 | 92 |
| 26 | 伙计 | 筲箕 | 1 | 8 | 77 | 49.62 | 58 |

(Note: Word pairs #6, 11, 12, 18, 21, 22 and 25 were the ones that could be distinguished by native speakers of Nanchang. Pitch values could not be extracted in the second syllables for word pairs #3, 5, 8 and 14.)

In the multiple regression analysis, the four acoustic parameters mentioned above were used as the predictors for the accuracy rates of the discrimination for the near minimal word pairs. Based on previous studies on tone perception (Gandour 1978, 1981), we entered the four predictors in steps from the least important acoustic parameter for tone identification to the most important. Hierarchical multiple regression analysis informs us whether adding a predictor significantly improves the regression model and explains significant amount of variance in the model. In this way, we can find what acoustic parameter robustly improves the prediction of the tonal discrimination accuracy rates and thus can be interpreted as helping tonal discrimination. Also, hierarchical multiple regression will tell us what acoustic parameter is not salient enough for the tonal discrimination. The order we entered the predictor was rime duration, pitch onset, mean f0, and pitch direction. In the hierarchical regression model, we found pitch onset and pitch direction to be highly significantly correlated ($r=.577$, $p<.01$). In the model, the part correlation coefficient for pitch onset was much smaller than the zero correlation coefficient (zero correlation coefficient: 0.358; part correlation coefficient: 0.049) whereas the pitch direction had similar part and zero correlation coefficients (zero correlation coefficient: 0.495; part correlation coefficient: 0.392). These results indicated that pitch onset was mediated by some other factor in terms of predicting the discrimination accuracy rates whereas pitch direction was not mediated by any other factors. The diagnosis of multicollinearity showed that the tolerance value for pitch onset was 0.504, which is much smaller than 1.0, whereas other predictors had tolerance values fairly close to 1.0. Based on the diagnosis in the regression model, we argue that pitch onset is a redundant predictor in the model and thus needs to be removed as it is highly correlated to another predictor — pitch direction — in the model, causing the multicollinearity problem. The reason for arguing that pitch onset is mediated by tone direction but not other variables such as mean f0 is that the correlation between pitch onset and mean f0 is low ($r=.354$) and it is not significant. In the end, we decided to use only three acoustic parameters as the predictors: rime duration, mean f0, and pitch direction.

The hierarchical regression result of the three predictors is summarized in the following table. The order we entered the predictor was rime duration, mean f0, and pitch direction, as shown in each step in Table 7.

**Table 7.** Hierarchical regression result of three predictors: rime duration, mean f0, and pitch direction.

| | B | SE B | β (standardized B) | $R^2$ | $\Delta R^2$ |
|---|---|---|---|---|---|
| **Dependent variable:** | **Accuracy rate of tonal discrimination in lexically stressless syllables (percentage of native speakers of Nanchang who correctly discriminated the word pair out of 12 speakers)** | | | | |
| **Step 1** | | | | | |
| Rime duration | .147 | .157 | .205 | .042 | .042 |
| **Step 2** | | | | | |
| Rime duration | .139 | .139 | .194 | .282* | .240* |
| Mean f0 | .489 | .194 | .490 | | |
| **Step 3** | | | | | |
| Rime duration | .149 | .126 | .209 | .445* | .162* |
| Mean f0 | .430 | .177 | .431 | | |
| Pitch direction | .801 | .349 | .408 | | |

Table 7 shows that the regression models in step 2 and step 3 significantly predict the discrimination accuracy rates but step 1 does not. In step 1, $R^2$=.042, F(1, 20) =0.878, p=.36. In step 2, $R^2$=.282, F(2, 19) =3.739, p<.05. In step 3, $R^2$=.445, F(3, 18)=4.808, p<.05. Also, adding mean f0 and pitch direction as predictors significantly improved the model's ability to predict the discrimination accuracy rates in step 2 and step 3 respectively. In step 2, $\Delta R^2$=.24, ΔF(1, 19)=6.363, p<.05. In step 3, $\Delta R^2$=.162, ΔF(1, 18)=5.267, p<.05. After adding all three acoustic parameters as predictors in step 3, mean f0 and pitch direction turned out to be significant predictors: mean f0 (SE B=.177, t(18)=2.428, p<.05); pitch direction (SE B=.349, t(18)=2.295, p<.05). Rime duration was a non-significant predictor. Therefore, mean f0 and pitch direction difference affected the discrimination accuracy rate. Back to the perception result, those 7 near minimal word pairs that could be distinguished by native Nanchang listeners had a mean f0 difference of 15.85 Hz and a mean of pitch direction difference of 45.6 Hz whereas the rest of the near minimal word pairs had a mean f0 difference 12 Hz and a mean of pitch direction difference 25 Hz. Though no statistics could be used to test whether the two word pair groups had significant differences in mean f0 and pitch direction, the tendency suggests a trend that the 7 distinguishable word pairs had larger differences than the non-distinguishable pairs.

To sum up the results of the perception study, the tonal neutralization pattern found in the production study, namely, 42-45-21 were merged and 24-213 were merged in the lexically stressless syllables, did not map onto the perception as the tonal contrast within the 42-45-21 group could still be perceived in some lexically stressless disyllables and the same was true for the 24-213 group. We argue that the mismatch between production and perception in terms of tonal neutralization is due to a word-specific effect. This effect arises from the larger acoustic differences in those distinguishable lexically stressless syllables. A hierarchical regression analysis shows that the incomplete tonal neutralization at the perception level can be accounted for by certain lexically stressless syllables having larger mean f0 and pitch direction differences than other lexically stressless syllables. Therefore, we argue that the word-specific effect on tonal neutralization at the perception level is manifested via the larger acoustic differences in certain lexical pairs.

## 5.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 rime duration shortening in the lexically stressless syllables, we argued that the tonal contrast distribution in Nanchang is related to rime duration, which is affected by the lexical stress status. However, the effect of stress on both rime duration and tonal contrasts varied among the words, suggesting a word-specific effect. In the perception study, we also found a word-specific effect on the discrimination of tones in lexically stressless syllables in that several word pairs could be distinguished by native Nanchang listeners but the majority of the word pairs could not be distinguished. In this section, we discuss the phonetic implementation of lexemes (or phonological representation of words) under the framework of the speech production model and exemplar model in the literature, which sheds light on the word-specific effect on the tonal neutralization patterns in Nanchang.

Using the speech production model proposed by (Levelt, Roelofs, and Meyer 1999), 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 that both the rime duration and f0 curves realized on the syllables are computed by the phonetic implementation rule after the prosodic structure of a syllable becomes 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, as in a modular feed-forward model such as Levelt's, the categorical form of the lexeme determines the phonetic outcome entirely. This is 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 neutralized in the same context. In other words, the categories (the underlying tones) determine the f0 configuration in the prosodically weak position, namely, the lexically stressless syllables. Levelt et al.'s speech production model essentially reflects a decoding scheme that maps the phonological features onto the detailed motor gestures. However, it ignores how the phonological representation is built. The phonological representation is inevitably linked to lexical items or words. A word can appear in various contexts with certain frequency spoken in a certain manner or sociostylistics. All these pieces of lexical information contribute to how the phonological representation is implemented phonetically (Pierrehumbert 2002). In addition to the categorical feature (e.g., stressed vs. unstressed), factors such as contextual predictability in Jurafsky's terms (Jurafsky, Alan, and Cynthia 2002), word frequency in Pierrehumbert's terms and pragmatics/social stylistics all potentially contribute to the final realization of the categorical feature. The exemplar model that includes both categorical or abstract features and detailed distributional information of the lexical items seems to be more robust in terms of accounting for the word-specific effect.

In sum, the production study showed an overall trend of tonal neutralization in the lexically stressless syllables due to rime shortening, namely, tones 42, 45 and 21 are merged whereas 24 and 213 are merged in both production and perception. However, there is a word-specific effect towards the tonal neutralization at both production and perception level. Therefore, we conclude that rime duration is indeed correlated with tonal contrasts' licensing but the detailed mechanism of tonal reduction is not only affected by phonological feature such as [stress], but also influenced by other lexical information such as word frequency and contextual predictability.

Finally, we need to stress that this paper does not aim to find the exact causes to the word-specific effect on rime duration and tonal realization. Mixed-effect models can inform us by-item differences from the mean, but it has the limitation of only being able to show that certain items have a larger variance than others numerically, but not being able to detect specific outliers via rigid statistical criteria. Moreover, even if we could find the words that behave differently from other items in terms of rime duration and tonal realization, we do not have the frequency information for this particular dialect to correlate with the acoustic results. Therefore, we have to admit that the factors that cause word-specific effects need further and more thorough studies in the future.

## NOTES

1. The linear mixed-effect model does not provide post-hoc comparison as ANOVA. It only compares the condition to the baseline. Therefore, in order to make pairwise comparisons between all levels of an independent variable, the baseline needs to be changed.

2. $\chi^2$ test is a non-parametric sign test for the likelihood of the occurrence of some event. In our case, for df=1 (two tokens for a word pair) p=.05, the critical value $\alpha$=3.84in the $\chi^2$ table. Only if 10 out of 12 listeners correctly identified the disyllable from which the target lexically stressless monosyllable was extracted is the $\chi^2$ value for the word pair over the critical value.Therefore, we only counted a near minimal pair as distinguishableby native Nanchang listeners if 10 out 12 listeners could correctly identify the source of the stressless syllable.

# REFERENCES


Bates, D.M., and Maechler, M. 2009. lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-32.

Baayen, R.H. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

Boersma, Paul, and David Weenink. *Praat: doing phonetics by computer [Computer program]* 2003. Available from http://www.praat.org/.

Clark, H. H. 1973. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. Journal of Verbal Learning and Verbal Behavior, 12, 335–359.

Chen, Yiya, and Xu Yi. 2006. "Production of weak elements in speech — Evidence from f0 patterns of neutral tone in standard Chinese." *Phonetica* no. 63:47-75.

Duanmu, San. 2007. *The Phonology of Standard Chinese*. 2nd ed. Oxford: Oxford University Press.

Gandour, J. 1978. Perceived dimensions of thirteen tones: a multidimensional scaling investigation. Phonetica, 35, 169-179.

.1983. Tone perception in Far Eastern languages. Journal of Phonetics, 11, 149-175.

Hirst, D. 2011. Praat script: http://uk.groups.yahoo.com/group/praat-users/message/4915

Jurafsky, D., B. Alan, and G. Cynthia. 2002. "The Role of the Lemma in Form Variation." In *Papers in Laboratory Phonology* edited by Carlos Gussenhoven and Natasha Warner, pp. 1-34. Berlin/New York: Mouton de Gruyter.

Lai, C. , Y. Sui, and J. Yuan. 2010. A corpus study of the prosody of polysyllabic words in Mandarin Chinese. Paper presentedat Speech Prosody.

Levelt, W. J., A. Roelofs, and A. S. Meyer. 1999. "A theory of lexical access in speech production." *Behav Brain Sci* no. 22 (1):1-38; discussion 38-75.

Li, Rong.李荣. 1995. *南昌方言词典*. 苏州: 江苏教育出版社.

Peng, Shuhui. 2000. "Lexical vs. 'phonological' representation of Mandarin sandhi tones." In *Acquisition and the lexicon*, edited by M. B. Broe and J. B. Pierrehumbert, pp. 152-167.

Pierrehumbert, J. 2002. "Word-specific phonetics." In *Laboratory Phonology*,pp.101-139. Berlin: Mouton de Gruyter.

Wei, Gangqiang.魏刚强. 2000. "调值的轻声和调类的轻声." *方言* 2000 年第一期:20-29 页.

Wei, Gangqiang.魏钢强. 1999. *萍乡方言志*.语文出版社.

Yaeger-Dror, M. 1996. "Phonetic evidence for the evolution of lexical classes: The case of a Montreal French vowel shift." In *Towards a social science of language*, edited by. G. Guy, C. Feagin, J. Baugh, and D. Schiffrin, pp. 263-287. Philadelphia: Benjamins.

Yaeger-Dror, M., and W. Kemp. 1992. "Lexical classes in Montreal French." *Language and Speech* 35:251-293.

Zhang, Jie. 2002. *The effects of duration and sonority on contour tone distribution: Typological survey and formal analysis*. New York: Routledge.
Zhang, Jie and Yu-Wen Lai. 2010. "Testing the role of phonetic knowledge in Mandarin tone sandhi." *Phonology* 27 (1): 153-201.

# 对南昌话中声调中和的发音和感知研究

刘江，张杰
美国堪萨斯大学

提要
在对江西省内赣方言南昌话中的声调对立的声学实验中，我们发现声调中和只出现在轻声音节中而没有出现在没有语法重音的非轻声音节中。我们认为这种声调对立分布的不同是由于音节韵母时长的不同造成的，即轻声音节韵母时长明显短于没有语法重音的非轻声音节韵母时长。而时长越短的音节所能负载的声调类别越有限，即出现声调中和现象。混合效应模型显示词重音对不同轻生词中声调中和的影响不同，表明个体词汇差别对轻声词的中声调中和是有影响的。从感知角度研究，我们发现声调中和确实发生在大多数轻声音节中，但是由于个体词汇差异，声调中和在某一些轻声音节中并没有发生。

关键词
声调中和，发音研究，感知研究，轻声，语法重音，混合效应模型, 个体词汇差异，南昌话

**<u>APPEDIX I: Wordlist for the durational study</u>**

CV-L-G: CV syllable with both lexical stress and grammatical stress
CV-L: CV syllable with lexical stress but without grammatical stress
CV-NoStress: Lexically stressless CV syllable

| Vowel | CV-L-G | CV-L | CV-NoStress |
|---|---|---|---|
| a | 开花 kʰai42 fa42<br>‘to blossom’ | 鲜花 ɕiɛn42 fa42<br>‘fresh flower’ | 泥巴 ni45 pa0 (42)<br>‘mud’ |
| | 洗车 ɕi213 tsʰa42<br>‘to wash a car’ | 货车 fo45 tsʰa42<br>‘cargo truck’ | 哑巴 ŋa213 pa0 (42)<br>‘mute’ |
| | 驮骂 tʰo42 ma21<br>‘to be scolded’ | 脏话 tsoŋ42 fa21<br>‘obscene language’ | 芝麻 tsɿ42 ma0 (45)<br>‘sesame seeds’ |
| | 卖画 mai21 fa21<br>‘to sell paintings’ | 公社 kuŋ42 sa21<br>‘commune’ | 篱笆 li45 pa0 (42)<br>‘fence’ |
| i | 养鸡 ioŋ213 tɕi42<br>‘to raise chicken’ | 土鸡 tʰu213 tɕi42<br>‘free range chicken’ | 徒弟 t’u24 tʰi0 (42)<br>‘apprentice’ |
| | 开机 kʰai42 tɕi42<br>‘to turn on a machine’ | 云梯 yun45 tʰi42<br>‘cloud ladder’ | 楼梯 lɛu45 tʰi0 (42)<br>‘stairs’ |
| | 扫地 sau213 tʰi21<br>‘to sweep the floor’ | 纸币 tsi213 pʰi21<br>‘paper notes’ | 把戏 pa213 ɕi0 (21)<br>‘tricks’ |
| | 拖地 tʰo42 tʰi21<br>‘to mop the floor’ | 假币 ka213 pi21<br>‘counterfeit money’ | 玻璃 po21 li0 (21)<br>‘glass’ |
| u | 迁都 tɕʰiɛn42 tu42<br>‘to move the capital’ | 草菇 tsʰau213 ku42<br>‘grass mushroom’ | 姐夫 tɕia213 fu0 (42)<br>‘brother in law’ |
| | 收租 ɕiu42 tsu42<br>‘to collect rent’ | 房租 foŋ45 tsu42<br>‘house rent’ | 蘑菇 mo45 ku0(42)<br>‘mushroom’ |
| | 扫墓 sau213 mu21<br>‘to visit the cemetery’ | 支部 tsɿ42 p’u21<br>‘branch unit’ | 欺负 tɕʰi42 fu0 (21)<br>‘to bully’ |
| | 修路 ɕiu42 lu21<br>‘to repair roads’ | 继父 tɕi21 fu21<br>‘stepfather’ | 坟墓 fɨn24 mu0 (21)<br>‘tomb’ |

**<u>APPEDIX II: Wordlist for the tonal contrast study</u>**

Note: For each tonal combination, the word on the left has a lexically stressless second syllable and the word on the right has a lexically stressed but grammatically stressless second syllable. After the web survey, 17 word pairs were used in the recording. The words in boxes are the selected words.

| **42+42** | | | **42+24** | |
|---|---|---|---|---|
| $\boxed{\text{冤家}}$ yon42 ka0 (42)<br>‘enemy’<br>亲家 tɕʰin42 ka0 (42)<br>‘in-law’ | $\boxed{\text{私家}}$ sɿ42 ka42<br>‘private’<br>三家 san42 ka42<br>‘three families’ | | 价钱 ka42 tɕʰiɛn0<br>(24) ‘price’<br>$\boxed{\text{招牌}}$ tsɛu42 pʰai0<br>(24) ‘brand’ | 脏钱 tsoŋ42 tɕʰiɛn24<br>‘dirty money’<br>$\boxed{\text{金牌}}$ tɕin42 pʰai24<br>‘gold medal’ |
| **24+42** | | | **24+24** | |
| 徒弟 tʰu24 tʰi0 (42)<br>‘prentice’<br>$\boxed{\text{棉花}}$ miɛn24 fa0 (42)<br>‘cotton’ | 唐弟 tʰoŋ24 tʰi42<br>‘cousin’<br>$\boxed{\text{桃花}}$ tʰau24 fa42<br>‘blossom tree’ | | 葡萄 pʰu24 tʰau0 (24)<br>‘grape’<br>强强 tɕʰioŋ24 tɕʰioŋ0<br>(24) a nick name | 甜桃 tʰiɛn24 tʰau24<br>‘sweet peach’<br>城墙 tsʰɛn24 tɕʰioŋ24<br>‘city wall’ |
| **45+42** | | | **45+24** | |
| $\boxed{\text{嫁妆}}$ ka45 tsoŋ0(42)<br>‘dowry’<br>$\boxed{\text{神经}}$ sɨn45 tɕin0 (42)<br>‘nerve’ | $\boxed{\text{浓妆}}$ luŋ45 tsoŋ42<br>‘thick make-up’<br>$\boxed{\text{半斤}}$ pɵn45 tɕin42<br>‘half half-kilogram’ | | $\boxed{\text{来头}}$ lai45 tʰɛu0 (24)<br>‘background’<br>名堂 min45 tʰoŋ0<br>(24) ‘matters’ | $\boxed{\text{蛇头}}$ sa45 tʰɛu24<br>‘head of a snake’<br>鱼塘 nie45 tʰoŋ24<br>‘fish pound’ |
| **213+42** | | | **213+24** | |
| $\boxed{\text{姐夫}}$ tɕia213 fu0 (42)<br>‘brother in law’<br>点心 tiɛn213 ɕin0<br>(42) ‘snacks’ | $\boxed{\text{马夫}}$ ma213 fu42<br>‘hostler’<br>菜心 tsʰai213 ɕin42<br>‘heart of bok choy’ | | $\boxed{\text{本钱}}$ pɨn213 tɕʰiɛn0<br>(24) ‘financial<br>capital’<br>枕头 tsɨn213 tʰɛu0<br>(24) ‘pillow’ | $\boxed{\text{假钱}}$ ka213 tɕʰiɛn24<br>‘false money’<br>狗头 kiɛu213 tʰɛu24<br>‘head of a dog’ |
| **21+42** | | | **21+24** | |
| $\boxed{\text{地方}}$ tʰi21 foŋ0 (42)<br>‘place’ | $\boxed{\text{后方}}$ hɛu21 foŋ42<br>‘rear area’ | | $\boxed{\text{后头}}$ hɛu21 tʰɛu0(24)<br>‘rear side’ | $\boxed{\text{大头}}$ tʰai21 tʰɛu24<br>‘big head’ |
| 寿星 sɨu21 ɕin0 (42)<br>‘god of long life’ | 外星 uai21 ɕin42<br>‘extra-terrestrial’ | | $\boxed{\text{外头}}$ uai21 tʰɛu0(24)<br>‘outside’ | $\boxed{\text{树头}}$ ɕy21 tʰɛu24<br>‘top of a tree’ |

| **42+45** | | | **42+213** | |
|---|---|---|---|---|
| 聪明 tsʰuŋ42 miaŋ0 (45) ‘clever’<br>$\boxed{\text{功劳}}$ kuŋ42 lau0 (45) ‘contribution’ | 刁民 tiau42 min45 ‘cunning citizens’<br>$\boxed{\text{天牢}}$ tʰiɛn42 lau45 ‘prison’ | | $\boxed{\text{虾米}}$ ha42 mi0 (213) ‘dry shrimp’<br>$\boxed{\text{跟斗}}$ kiɛn42 tɛu0 (213) ‘tumble’ | $\boxed{\text{生米}}$ saŋ42 mi213 ‘raw rice’<br>$\boxed{\text{三斗}}$ san42 tɛu213 ‘three baskets’ |
| **24+45** | | | **24+213** | |
| 猴年 hou24 niɛn0 (45) ‘year of the monkey’<br>丰年 fɤŋ24 niɛn0 (45) ‘harvest year’ | 前年 tɕʰiɛn24 niɛn45 ‘the year before’<br>前年 tɕʰiɛn24 niɛn45 ‘the year before’ | | 寒气 hən24 tɕʰi0 (213) ‘cold air’<br>潮气 tsʰɛu24 tɕʰi0 (213) ‘humidity’ | 铜器 tʰuŋ24 tɕʰi213 ‘copper containers’<br>陶器 tau24 tɕʰi213 ‘pottery’ |
| **45+45** | | | **45+213** | |
| 记性 tɕi45 ɕin0 (45) ‘memory’<br>代价 tai45 ka0(45) ‘cost’ | 人性 nin45 ɕin45 ‘human nature’<br>油价 iu45 ka45 ‘price of oil’ | | 神气 sɨn45 tɕʰi0 (213) ‘arrogant’<br>$\boxed{\text{凉快}}$ lioŋ45 kʰuai0 (213) ‘coolness’ | 人气 nin45 tɕʰi213 ‘popularity’<br>$\boxed{\text{最快}}$ tsui45 kʰuai213 ‘the fastest’ |
| **213+45** | | | **213+213** | |
| 仔细 tsɿ213 ɕi0 (45) ‘careful’<br>$\boxed{\text{把戏}}$ pa213 ɕi0 (45) ‘trick’ | 好细 hau213 ɕi45 ‘so fine’<br>$\boxed{\text{好戏}}$ hau213 ɕi45 ‘good drama’ | | 手气 sɨu213 tɕʰi0 (213) ‘luck’<br>$\boxed{\text{韭菜}}$ tɕiu213 tsʰai0 (213) ‘leeks’ | 老气 lau213 tɕʰi213 ‘passe’<br>$\boxed{\text{好菜}}$ hau213 tsʰai213 ‘nice dish’ |
| **21+45** | | | **21+213** | |
| 面相 miɛn21ɕioŋ0 (45) ‘appearance’ | 外向 uai21 ɕioŋ0 (45) ‘extroversion’ | | $\boxed{\text{义气}}$ ɲi21 tɕʰi0 (213) ‘brotherhood’<br>运气 yn21 tɕʰi0 (213) ‘luck’ | $\boxed{\text{大气}}$ tʰai21 tɕʰi213 ‘grand’<br>利器 li21 tɕʰi213 ‘sharp instrument’ |

| 42+21 | | | 24+21 | |
|---|---|---|---|---|
| 机会 tɕi42 fɨi0 (21)<br>'chance'<br>欺负 tɕi42 fu0 (21)<br>'to bully' | 帮会 poŋ42 fɨi21<br>'gang'<br>生父 saŋ42 fu21<br>'biological father' | | 筹划 tsʰɨu24 fa0 (21)<br>'plan'<br>闲话 ɕiɛn24 fa0 (21)<br>'gossip' | 图画 tʰu24 fa21<br>'painting'<br>行话 hoŋ24 fa21<br>'jargon' |
| **45+21** | | | **213+21** | |
| 毛病 mau45 pʰiaŋ0 (21) 'bad habit'<br>计划 tɕi45 fa0 (21)<br>'plan' | 痨病 lau45 pʰiaŋ21<br>'phthisis'<br>年画 niɛn45 fa21<br>'new year's painting' | | 水稻 sui213 tʰau0 (21) 'water rice'<br>水分 sui213 fɨn0 (21)<br>'water content' | 小道 ɕiɛu213 tʰau0 (21) 'small road'<br>两份 lioŋ213 fɨn21<br>'two copies' |
| **21+21** | | | | |
| 便饭 pʰiɛn21 fan0 (21) 'simple meal'<br>肾病 sɨn21 pʰiaŋ0 (21) 'kidney disease' | 剩饭 sɨn21 fan0 (21)<br>'leftover'<br>重病 tsʰuŋ21 pʰiaŋ21<br>'serious illness' | | | |

## Appendix III: Wordlist for the the perception study

Note: The target syllables are the second syllables, which are lexically stressless. The tonal contrast (e.g., 42–45) indicates the underlying tones in two lexically stressless syllables, which are segmentally identical. The italicized word pairs in parentheses are the word pairs that can be distinguished by native Nanchang speakers.

Words used in the ABX discrimination task:

| **42–45** | **42–21** | **45–21** |
|---|---|---|
| 嫁 45 妆 42–胜 45 仗 45<br>ka45 tsoŋ42–seŋ45 tsoŋ45 | 棉 45 花 42–计 45 划 21<br>mian45 fa42–tɕi45 fa21 | 经 42 费 45–机 42 会 21<br>tɕi42 fɨi45– tɕi42 fɨi21 |
| 剪 213 刀 42–捡 213 到 45<br>kan213 tau42–kan213 tau45 | *(裁 24 缝 21–雷 45 锋 42<br>tsai24 foŋ21–lui45 foŋ21)* | 收 42 益 45–手 213 艺 21<br>sou42 i45– sou213 i21 |
| 东 45 西 42–把 213 戏 45<br>toŋ45 ɕi42–pa213 ɕi45 | 姐 213 夫 42–师 42 父 21<br>tɕia213 fu42–si42 fu21 | |
| 筲 42 箕 42–挂 42 记 45<br>tsuo42 tɕi42–kua42 tɕi45 | | |
| **213–24** | **213–42** | |
| *(机 42 器 213–肚 21 脐 24<br>tɕi42 tɕ$^{h}$i213–t$^{h}$u21 tɕ$^{h}$i24)* | *(巴 42 掌 213–嫁 45 妆 42<br>pa42 tsoŋ213–ka45 tsoŋ42)* | |
| *(宽 42 敞 213–肥 45 肠 24<br>kuan42 ts$^{h}$oŋ213–fɨi45 ts$^{h}$oŋ24)* | | |
| 云 45 彩 213–招 42 财 24<br>iun45 ts$^{h}$ai213–tsau42 ts$^{h}$ai24 | | |

Words used in the AXB discrimination task:

| **42–45** | **42–21** | **45–21** |
|---|---|---|
| 比 213 方 42–作 42 坊 45<br>pi213 foŋ42–tsuo42 foŋ45 | *(泥 45 鳅 42–母 213 舅 21<br>ni45 tɕiu42– mu213 tɕiu21)* | 高 42 粱 45–较 45 量 21<br>kau42 lioŋ45–kau45 lioŋ21 |
| 良 24 心 42–记 45 性 45<br>lioŋ24 ɕin42– tɕi45 ɕin45 | 官 42 司 42–道 21 士 21<br>huan42 si42–t$^{h}$au21 si21 | |
| 筲 42 箕 42–伙 213 计 45<br>f tsuo42 tɕi42– o213 tɕi45 | 东 45 西 42–关 42 系 21<br>toŋ45 ɕi42–kuan42 ɕi21 | |
| **213–24** | **213–42** | **21–24** |
| 背 45 叛 213–算 45 盘 24<br>pei45 p$^{h}$an213–soŋ45 p$^{h}$an24 | 马 213 虎 213–姐 213 夫 42<br>ma213 fu213–tɕi213 fu42 | *(母 213 舅 21–要 42 求 24<br>mu213 tɕiu21–iau42 tɕiu24)* |
| *(螃 24 蟹 213–套 213 鞋 24<br>p$^{h}$oŋ24 hai213–t$^{h}$au213 hai24)* | | 凉 45 拌 21–算 45 盘 24<br>lioŋ45 p$^{h}$oŋ21–suan45 p$^{h}$oŋ24 |
| **21–213** | | |
| 欺 42 负 21–马 213 虎 213<br>tɕ$^{h}$i42 fu21–ma213 fu213 | | |