Pitch and Phonation Type Perception in Wenzhou Dialect Tone

Xu Xiaoying¹, Liu Xuefei², Tao Jianhua³, Che Hao⁴

¹,²PSC Center/School of Chinese Language & Literature, Beijing Normal University, China
¹,²,³,⁴NLPR, Institute of Automation, Chinese Academy of Sciences, Beijing, China
Xuxiaoying2000@bnu.edu.cn, liuxuefei1234@163.com, {jhtao, hche}@nlpr.ia.ac.cn

Abstract

Wenzhou dialect is characterized by a large tonal inventory and tones which combine pitch contour and phonation type cues. In this paper, a perception experiment was conducted to investigate identification cues of the eight tones in Wenzhou. The results of the experiment show that: I) Not all native speakers can identify tones using pitch contour alone. II) It was difficult for participants to recognize tone 1. III) Some stimuli, especially tone 4, could carry a wide range of F0 contours and still be identified as belonging to the original tonal category; many other tones, however, were perceived as belonging to other categories after only slight modifications to F0. IV) Phonation type impacts the identification of some tone categories. Based on the results, the phonological organization of Wenzhou’s tone system is examined and the possibility of reconstructing the Middle Chinese tone system is discussed.

Index Terms: Wenzhou Wu dialect, tone, pitch, phonation type, perception

1. Introduction

The lexical tones and tone sandhi patterns of the Wenzhou dialect of Wu Chinese (hereafter Wenzhou) have been studied in detail [6, 10, 11]. In early Chinese phonological descriptions, Wenzhou was generally considered to have eight lexical tones [18, 5] whose main distinctive features are F0 and the voicing characteristic of the initial consonant.

Recently, attention has been drawn to the role of phonation type in the phonetic realization of tone, and accordingly phonation type has been used to reconstruct the Middle Chinese (MC) tone system and the phonetic characteristics of initial consonants (which are considered to have had an extremely close connection with the MC tone system) [19, 20].

While there exists a large amount of previous work focusing on Wenzhou tone, and several studies have investigated acoustic cues in tone identification—including phonation type [19, 20]—the acoustic cues that are used to determine tone features are usually based on auditory impressions or experimental evidence and are rarely shown to be perceptually relevant to native speakers/listeners [3]. As a language characterized by large tonal inventory and tones which combine pitch, phonatory and consonant voicing et alia as perceptual cues, the phonological organization of Wenzhou tone is still open to debate, and no studies so far have investigated listeners’ perception of mixed pitch/phonation-type tones.

In this paper, a perception experiment was conducted to verify the perceptual cues for identifying the eight tones of Wenzhou. The result of the experiments show that: I) Not all native speakers can identify tones using pitch contour alone. II) It was difficult for participants to recognize tone 1. III) Some stimuli, especially tone 4, could carry a wide range of F0 contours and still be identified as belonging to the original tonal category; many other tones, however, were perceived as belonging to other categories with only slight modifications to F0. IV) Phonation type impacts the identification of some tone categories. Below, based on the results, the phonological organization of Wenzhou tone system is examined and the implications for reconstructing the Middle Chinese tonal system are discussed.

The rest of the paper is organized as follows. Section 2 reviews the relevant literature. Section 3 introduces pitch and phonation type in Wenzhou tone. The perception experiment is presented in Section 4. Section 5 discusses the results. Finally, Section 6 concludes the paper and lays down directions for our future research.

2. Prior research

2.1 Previous research on acoustic and perception properties of tone

In past decades, the issue of tonal representation and tone features has been important in the phonological field. Mostly, the features that have been proposed are based on contour, pitch height, register [6, 10, 11, 1, 8, 9], and duration [1, 8, 9, 13]. Recently, phonation type has also been proposed to be an important feature of tonal representation in certain languages [2, 3]. Most researchers agree that a set of phonetically grounded universal tone features defining pitch contour and tone height can be used to capture phonological tone process [4, 6, 7, 14, 15, 16], but some scholars have begun to argue for a flexible phonetic realization of phonological categories [3].

Concerning research methods, though the acoustic cues that are used to determine tone features are usually based on auditory impressions or experimental evidence, the perception method had been used in some work. [2] examines acoustic data to address the question of what makes tones clear when tonal identity is determined by a complex of acoustic cues including phonation type in addition to pitch height. [3] describes experimental research on the perception of Vietnamese tone and challenges previous models of Vietnamese tone features.

2.2 Previous research on tone in Wenzhou

Wenzhou is in the southern area of Zhejiang province. The tone and sandhi system of Wenzhou is well studied. Wenzhou is described as having a highly complex tone system that combines pitch contour with the voicing feature of the initial consonant [17], a complex tone sandhi system [6, 10, 11], as well as complex stress patterns in prosodic words [18].

Recently, Zhu Xiaonong has suggested that attention must be paid to the role of phonation type when analyzing the tone system of Wenzhou [19, 20, 21]. He outlines a tone analysis model which divides phonation type and pitch into different
levels [21], and argues that the acoustic properties traditionally associated with the difference between voiced and voiceless initial consonants are actually related to breathy and modal phonation type. Furthermore, according to [19], falsetto phonation (假声, jiasheng) is associated with the shangsheng "rising tone" (上声) tone in Wenzhou. [19] proposed that Middle Chinese shangsheng, which was traditionally described as sharp and raised, was likely pronounced with falsetto phonation. Furthermore, based on evidence from slack (驰声, chisheng) phonation in yang (阳) tone syllables in Wenzhou, [21] suggests a new reconstruction of phonation type in the initial voiced/voiceless stops of MC. Though the acoustic properties of isolated monosyllables and disyllabic words have been studied extensively, more research, especially on the variation of phonation types across different contexts and sociolinguistic registers is still needed. Moreover, no studies are available on Wenzhou tone identification, thus we know very little about the perception of Wenzhou tone.

### 3. The acoustic properties of lexical tone in Wenzhou

The Wu dialects are a group of Chinese dialects which are traditionally considered to be differentiated from other dialects by their use of voiced and voiceless initial consonants and their complex tone sandhi. In this section, the tonal acoustics of a male and a female native speaker are analyzed.

#### 3.1. Minimal tone set for Wenzhou lexical tone

According to traditional descriptions, there are eight tones in single syllables of Wenzhou, with each tone defined by an audibly distinguishable pitch and duration shape.

Figure 1 and Figure 2 show the tonal acoustics of a female native speaker in her forties and a male native speaker in his seventies, respectively. Both are from Pingyang (平阳) County in Wenzhou.

Comparing Figure 1 with Figure 2, we can find that except for tone 3, the configurations of the pitch contour are quite similar among the speakers. Table 1 shows the minimal tone set for Wenzhou Wu dialect used by the speakers.

<table>
<thead>
<tr>
<th>Tones</th>
<th>Qìéyùn (Qiè yùn)</th>
<th>Example</th>
<th>Character</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tone 1</td>
<td>yìnpíng 阴平</td>
<td>/tʂ33/</td>
<td>单</td>
<td>single</td>
</tr>
<tr>
<td>Tone 2</td>
<td>yángpíng 阳平</td>
<td>/tʂ31/</td>
<td>谈</td>
<td>talk</td>
</tr>
<tr>
<td>Tone 3</td>
<td>yínshāng 阴上</td>
<td>/tɿ454/</td>
<td>胆</td>
<td>courage</td>
</tr>
<tr>
<td>Tone 4</td>
<td>yángshāng 阳上</td>
<td>/tɿ313/</td>
<td>胆</td>
<td>courage</td>
</tr>
<tr>
<td>Tone 5</td>
<td>yìngu 阴去</td>
<td>/tɿ41/</td>
<td>旦</td>
<td>morning</td>
</tr>
</tbody>
</table>

#### 3.2. Phonation types for Wenzhou lexical tones

Three phonation types have been described for Wenzhou: modal, breathy and glottal [19]. Glottal phonation type is illustrated clearly in tone 3 of the female speaker in figure 1, but not for the male speaker in figure 2. This indicates that glottal voice is beginning to exhibit sociolinguistic variation in contemporary Wenzhou Wu dialect.

Modals and breathy tones are distinguished partly by acoustic correlates of phonation type. Three cues can be examined: difference in first and second harmonic amplitude (H1-H2), jitter and shimmer. In our paper, only the most typical cue—the difference in first and second harmonic amplitudes (H1-H2)—will be discussed.

Difference in first and second harmonic amplitudes is correlated with the open quotient (OQ), or the ratio of time in each glottal cycle during which the vocal folds are open. [12] provides an illustration in which difference in H1-H2 amplitudes in the glottal waveform for a male voice changes from -5dB when the OQ is 30% (creaky phonation) to +5dB when the OQ is 70% (breathy phonation). [2] concludes that H1-H2 values above 9dB appear to provide an excellent cue to the presence of breathy voice. But exceptions still exist, and the correlation between phonation type and the difference in first and second harmonic amplitudes is not entirely consistent [21].

H1 and H2 amplitudes were measured by hand from a spectrum in Praat. The spectrums and the H1, H2 values are shown in figure 3 and figure 4:

| Tone 6 | yànggu 阳去 | /tɿ31/ | 但 | but |
| Tone 7 | yínrú 前入 | /tɿ323/ | 搭 | construct |
| Tone 8 | yángru 阳入 | /tɿ212/ | 达 | arrive |

“Qièyùn” is given in reference to the ancient “qièyùn 《切韵》” categorizations. The column “Example” gives instances of each tone with phonemic transcription according to auditory impressions; pitch contour realizations are described using Chao tone letters, where “1” means low pitch and “5” means high pitch. “Character” and “Gloss” give the words in Chinese orthography and their meanings, respectively.

Arranged according to similarities in pitch contour, the eight tones of Wenzhou are: low level (tone 6), mid level (tone 1), high falling (tone 5), low falling (tone 2), high rising-falling (tone 3), low rising-falling (tone 4), low falling-rising (tone 8) and mid falling-rising (tone 7). Generally, there are no obvious mean F0 distinctions between these two speakers.
Comparing the female speaker with the male speaker, the female speaker retains falsetto phonation and the glottal stop in the tone 3 syllable, but the male speaker doesn’t. For tone 2, 4, 6, and 8, the male speaker shows a more typical breathy phonation than the female speaker, both in auditory impressions and H1-H2 measurements.

The results of the auditory analysis show that the distinction in the voicing of initial consonants recorded in traditional phonological documents is indeed a difference in phonation type, which is consistent with the description in [19, 20]. The results of the acoustic analysis illustrate that variation in phonation type exists between the two speakers, and that measurements of H1-H2 are not completely consistent with auditory impression.

4. Methods: perception experiment

4.1. Recordings

The syllables “tɕʰ”, “tɕ” and “tɕʰ” pronounced in isolation, were recorded in a quiet room for each of Wenzhou’s eight tone categories. The target words are listed in Table 1. The speaker is a male native speaker in his seventies. Figure 2 and Figure 4 illustrate the pitch and phona tory properties of the natural stimuli used in our experiment.

4.2. Stimuli synthesis

Besides the natural stimuli described in the previous section, five groups of stimuli were synthesized in Praat. The first group of stimuli was manipulated by removing the segmental properties of the natural stimuli and replacing them with a synthesized hum voice. The other four groups (2–5) are made up of four perceptually similar pairs which consist of tone 1 and tone 6, tone 2 and tone 5, tone 3 and tone 4, tone 7 and tone 8 respectively. The tones in each pair were matched to have similar pitch contour configuration, relatively similar auditory impression (according to native listeners), and divergent phonatory properties.

We designed a pitch contour continuum between the two tones in each pair. The synthesized stimuli retain their original segmental properties and just the F0 cues are modified. Different methods were used according to the nature of distinctions in each pair.

All the stimuli were not modified to neutralize duration because the duration differences in the unmodified stimuli are not obvious. In order to ensure the stimuli sounded natural, they were evaluated by a native speaker of Wenzhou before their use in the experiment. The native speaker confirmed that the F0 manipulation had little effect on naturalness

4.2.1. Hum stimuli with the F0 properties retained

For removing the segmental properties of the natural stimuli, after using the command “Extract visible pitch contour” on each of the eight natural stimuli in Praat, we used the command “To Sound (hum)”, thus producing eight tokens, one for each of the eight lexical tones, with segmental information removed. Each of these eight tokens was recognized twice in the experiment.

4.2.2. Stimuli with a level contour

The first experimental pair is tone 1 and tone 6, which are level contour tones with F0 of the last pitch point of 122Hz (tone 1) and 111Hz (tone 6). The Tone 1 to tone 6 F0 continuum was manipulated in Praat according to the
following steps: 1) Perform “To Manipulation...” on the opened wave file, then 2) click “Multiply pitch frequencies...” under “Pitch” in the Manipulation editor window. 3) Use a factor of 1.05 to raise the F0 and 0.95 to lower it. 4) Repeat this process 10 times in each direction (higher and lower) to yield 21 tokens. The value of the last pitch point of the interpolated stimuli will be shown in figure 9 when discussing the perception results.

The modified range of tone 1 is illustrated in figure 4 as follows:

Figure 4: Reducing F0 from 122Hz to 73Hz for tone 1 stimulus. (Grey line shows the original pitch contour)

4.2.3 Stimuli with a falling contour

The second perception pair is the tone 2 and tone 5, which are falling contour tones with the first F0 point of 122Hz and 213Hz respectively, and the second and third pitch point in stylized pitch form is very similar. The Tone 1 to tone 6 F0 continuum were manipulated according to the following steps: 1) perform “To Manipulation...” on the opened wave file, then 2) click “Stylize pitch...” in the Manipulation editor window’s “Pitch” menu. 3) After getting three stylized pitch points, we only modified the first pitch point of tone 2 and tone 5 for the continua. At each step in the continuum we reduced tone 2 by 10Hz and raised tone 5 by 10 Hz.

The ranges of modification are illustrated in Figure 5, below.

Figure 5: Manipulation of the first F0 pitch point for tone 2 and tone 5. (Grey lines show the original pitch contour)

For this perception pair, we manipulated the first pitch point in 10 steps from the original tone 2 stimulus, starting at 122Hz and ending at the target pitch value of 212Hz; for tone 5, 12 steps were created from an original starting F0 of 213Hz to the target pitch value of 93Hz.

4.2.4 Stimuli with a falling-rising contour

The third perception pair is tone 7 and tone 8, which are falling-rising contour tones with the F0 of the lowest pitch point of 113Hz and 90Hz respectively. The manipulation method used in this pair is same as the first perception pair shown in section 4.2.2. So the pitch contour of each tone is retained and pitch is multiplied by a factor of 1.05 or 0.95 for each step. Resultingly, for tone 7, F0 at the lowest point ranged from 72Hz to 109Hz, and for tone 8, lowest F0 ranged from 90Hz to 177Hz.

The modified ranges are illustrated in Figure 6.

Figure 6: Manipulations of tone 7 and tone 8. (Grey lines show the original pitch contour)

4.2.5 Stimuli with a rising-falling contour

The forth perception pair are tone 3 and tone 4, both rising-falling tones. Due to the auditory impressions of this pair being rather different, we manipulated two groups of stimuli. For the first group of stimuli, the manipulating method used is same as the first pair, as discussed in section 4.2.2. The F0 contour of the syllable ws retained and pitch was multiplied by factors of 1.05 or 0.95. For tone 3, the highest pitch point was modified from 195Hz in the original to 60Hz in the last step of the continuum, and for tone 4 the F0 peak was eventually raised from 159Hz to 247Hz, as well as reduced to 69Hz. The second group of stimuli, only had two stimuli. We copied the F0 contours of tone 3 to tone 4 and tone 4 to tone 3 manually in the manipulation window in Praat. The result is shown in Figure 7.

Figure 7: Manipulated versions of tone 3 and tone 4. (Grey lines show the original pitch contour)

4.3 Subjects

3 listeners born between 1940 and 1969 were selected for the experiment. They were born and raised in Pingyang district and they are bilingual speakers of Wenzhou dialect and Putonghua.

4.4 Procedure

All subjects carried out a forced-choice identification task presented in Praat on a laptop computer. They listened to three repetitions of each of the 135 stimuli in isolation, for a total of 405 (3x135) randomized tokens. They identified the tone of the stimuli by clicking on one of eight boxes containing the words from the minimal set given in Table 1. Subjects were allow to take a break after each block of 20 tokens. All subjects had a short practice run (10 tokens) before doing the actual experiment.

5 Results of perception experiment

5.1 Identification of hum voice stimuli

The results on tone identification in hum voice stimuli are shown in Figure 8.
In figure 8, Axe X means the perception tone type, and axe Y means the 16 modified tokens, due that each tone has two tokens, so the 1 and 2 in axe Y means modified tone 1. Individual divergence exists in recognizing the hum voice stimuli. Among the three informants, informant 2 can identify all the stimuli correctly, informant 1 recognized three stimuli (tone 2, 4, 7) wrongly, and informant 3 did not finish the perception task, citing its extreme difficulty for her. The results illustrate that not all native speakers can recognize tone categories by using pitch contour alone.

5.2 Identification of the synthesis stimuli

Figures 9-11 show the perception results of the natural and synthesized stimulus continua. The vertical axis illustrates the tone types which the informants assigned to the stimulus. The numbers in the upper line of the horizontal axis show the F0 value of the reference points referred to above, while the lower line shows the tone type of the segment that the stimuli was originally based on.

5.2.1 Identification of the tone 1 and tone 6 pair

The perception results for the tone 1 and 6 pair are given in Figure 9, below.

Figure 9: Perception results for tone 1 and tone 6.

Tone 1 stimuli (that is, stimuli modified based on the original tone 1 syllable) show the lowest consensus among all stimuli. Besides all three informants judging the original tone 1 stimuli as tone 8, the other points with consensus are 95Hz as tone 6 and 110Hz as tone 1. We can’t find any patterns in these perception results. We observe that it is difficult to recognize tone 1 stimuli in the course of an experiment.

Among the modified stimuli, tone 6 stimuli were recognized as tone 6 only at 81Hz, 91Hz and 117Hz. At 100Hz and 106Hz, they were judged as tone 5, and when the pitch was raised higher than 122 Hz, they were recognized as tone 5 and tone 1.

5.2.2 Identification of the tone 2 and tone 5 pair

The perception results for tones 2 and 5 are given in Figure 10.

Figure 10: Perception result of Tone 2 and tone 5 pair.

The majority of tone 2 and tone 5 stimuli were recognized as tone 5. Tone 2 stimuli were recognized as tone 5 when the pitch of the first point was raised to between 142Hz and 192Hz (the unmodified value being 122Hz). The tone 2 stimuli were also recognized as tone 5 when the first point was raised to 202Hz and 213Hz, but at this level some categorizations as tone 1 also occurred. For tone 5 stimuli, when the pitch of the first point went below 143Hz, a divergence in the identification results appeared, with some steps going to tone 4 and tone 6, and several going to tone 2, especially 93 and 103 Hz, which are quite close to the unmodified F0 for tone 2 (122Hz).

5.2.3 Identification of the tone 7 and tone 8 pair

The perception results for the tone 7 and 8 pair are given in Figure 11.

Figure 11: Perception results for tone 7 and tone 8.

The stimuli of tone 8 can be identified as tone 8 quite well. With the lowest pitch point raising from 95Hz in the unmodified stimulus to 177Hz, most of the tone 8 stimuli were identified as tone 8. Only the stimuli with 121Hz and 127Hz were recognized as Tone 7. Stimuli based on tone 7 were judged as tone 7 at the 103Hz and 109Hz steps, which are relatively near the 117Hz of the natural stimulus. When the lowest pitch points were lowered to 89Hz, 85Hz and 80Hz, they were judged as tone 6, which is a level tone with a F0 of 111Hz for that speaker. When the lowest pitch points went below 76Hz, they were judged as tone 7, tone 8 and even tone 1.

5.2.4 Identification of the tone 3 and tone 4 pair

The perception results for the tone 3 and 4 pair are given in Figure 12.

Figure 12: Perception result of Tone 3 and tone 4 pair.

Only when the highest pitch point was at 165Hz, 174Hz or 183Hz were tone 3 stimuli recognized stably as tone 3. When the highest pitch points go below 157Hz, they were judged as tone 4 at 157Hz and 149Hz, and other steps below...
142Hz elicited identifications in all tonal categories except for tone 7.

Tone 4 has a very wide range of F0 for which identification as tone 4 is stable. No matter if the pitch contour is raised (from the original 159Hz to 247Hz) or reduced (from the original 159Hz to 63Hz), the majority of the stimuli based on tone 4 were judged as tone 4, with a few of them recognized as tone 6 (at 101Hz and 105Hz) and tone 8 (at 69Hz, 73Hz, 101Hz, and 105Hz).

Except for a few identifications as tone 3, tone 4 stimuli were only identified as tones in the yung (陽) category, i.e. tones 2, 4, 6 and 8, which are all characterized by breathy phonation. Tone 3 stimuli, meanwhile, were identified as belonging to tones in both yung and yin (陰), encompassing tones 1, 3, 5 and 7 categories.

6 Discussion and Conclusion

Though the perception experiment of this paper is quite preliminary, we obtained some interesting results: I) Not all native speakers can identify tones using pitch contour alone. II) It was difficult for participants to recognize tone 1. III) Some stimuli could carry a wide range of F0 contours and still be identified as belonging to the original tonal category; many other tones, however, were perceived as belonging to other categories after only slight modifications to F0. For example, it was difficult for tone 2 stimuli to be recognized as tone 2, while tone 4 has a very wide range of F0 values recognized as tone 4. Similarly, tone 8 stimuli can be identified as tone 8 quite well, with acceptable F0 minima ranging from 95Hz in the unmodified stimulus to 177Hz, in which range most of the stimuli were identified as tone 8. IV) Phonation type impacts the identification of some tone categories, such as in tone 4.

These results can shed light on two research areas: I) The reconsideration of traditional citation tone description. The results of our work verify that tonal representation includes not only pitch contour, pitch height/register, and duration, but also phonation type, which must be incorporated into accounts of tonal phonology in Wenzhou. Thus our results are not consistent with the conclusion that a set of phonetically grounded universal tone features including only pitch contour and tone height can be used to capture tonal phonology. We agree, rather, with the suggestion that there exist a flexible phonetic realization of phonological categories [3]. II) The reconsideration of synchronic variability in Chinese tone. Previous research has proved that some phonation types, such as falsetto and breathy, do exist in certain dialects and the auditory impression of these phonation types is very similar to those of the original 159Hz to 63Hz, the majority of the stimuli based on tone 4 were judged as tone 4, with a few of them recognized as tone 6 (at 101Hz and 105Hz) and tone 8 (at 69Hz, 73Hz, 101Hz, and 105Hz). For example, it was difficult for participants to recognize tone 4. III) Some previous research has proved that some phonation types, such as falsetto and breathy, do exist in certain dialects and the auditory impression of these phonation types is very similar to those of.

In the future work, we will pay attention to several issues: I) It has been shown that besides pitch and phonation type, duration and intensity can be important components of East Asian tone (Halle,1994; Xu, 1991). We will factor out duration and intensity in the future work. II) As regards acoustic properties, more syllables with different segmental content, utterances in different intonational contexts and different sociolinguistic registers will be examined. III) In the acoustic analysis in this work, we only used H1-H2 measurements to examine the acoustic properties of the phonation type. More parameters will be examined in future.

7 Acknowledgements

The work was supported by Beijing Normal University (004-127028), the Fundamental Research Funds for the Central University (2010105565004GK), the State Language Commission 12th Five-Year language research programme (YB125-41), the National Science Foundation of China (No. 60873160, 61011140075 and 90820303) and China-Singapore Institute of Digital Media (CSIDM).

8 References