ON THE PERCEPTION OF THE NEUTRAL TONE IN TAIWAN MANDARIN

Karen Huang

University of Hawai'i at Mānoa, USA
huangk@hawaii.edu

ABSTRACT

The neutral tone in Taiwan Mandarin (TM) behaves differently from that of Standard Mandarin. Many neutral-tone syllables in TM are not reduced and have a low pitch target regardless of the preceding tone. These neutral-tone syllables are therefore acoustically similar to Tone 3 syllables in TM.

This paper investigates how TM listeners distinguish Tone 3 (low tone) from the neutral tone. The result shows that the end pitch is the primary cue. When the end pitch is higher than -1 z-score (the pitch ranges between -2 to 2 z-score), the listeners are more likely to perceive the stimulus as a neutral tone. Also, a convex pitch contour is more likely to be perceived as a neutral tone. Lastly, when the pitch information is ambiguous, the listeners rely on the vowel quality or the phonation type to distinguish the pairs. The results suggest that the neutral tone in TM slowly reaches to a mid-low target.

Keywords: speech perception, speech prosody, tone

1. INTRODUCTION

In Taiwan Mandarin (TM) the neutral-tone syllables behave differently compared to Standard Mandarin (SM). In SM, an unstressed syllable reduces [3] and is shorter in duration [1, 7, 8]. The tone of the unstressed syllable varies with the preceding tone [1, 7, 8], so the pitch of the unstressed syllable is determined by the four possible lexical (full) tones /H, LH, L, HL/ (high, rising, dipping, falling) of the preceding syllable. These unstressed syllables are often referred to as having the "fifth tone" or a "neutral tone" /Ø/ [1].

However, in TM, unstressed syllables are less frequent [3, 6, 11]. TM is often described as more syllable-timed than SM [6]. Further, it seems that the rhythmic difference even affects the remaining neutral-tone syllables. Many prescriptive neutral-tone syllables in TM have a low pitch target regardless of the preceding tone [5]. Consecutive neutral-tone syllables in TM each have their own pitch targets unlike in SM, where the pitch gradually moves toward a mid target over neutral-tone sequences [2]. The neutral tone in TM behaves like a lexical tone rather than a neutral tone. Also, our pilot study found that the neutral-tone syllables in TM generally are not shortened or reduced. With a fairly low pitch target and rather long duration, these neutral-tone syllables are acoustically similar to Tone 3 syllables, which are mostly low-falling in TM [4]. Some neutral-tone syllables seemed to merge with Tone 3, but some seemed to be able to be recognized. One question that needs to be asked, then, is what the perceptual cues are for the TM listeners to distinguish the two tones if they could distinguish them.

A previous perception study [4] found that TM listeners rely on the falling portion of the syllables to identify Tone 3. With manipulation of the pitches, Fon, et al. showed that low initial pitch and steep pitch fall are important cues for the TM listeners to identify Tone 3.

Previous perception studies on neutral tone were only conducted on SM. Lin [9] found that duration is the main cue to distinguish the neutral tone from the full tones. The Initial pitch of the neutral tone also takes an important role [9, 10]. Intensity is not an important cue, and pitch contour plays a minor role because the neutral-tone syllables are usually short [9].

In TM, the durations of the neutral-tone syllables are generally not shorter than the full-tone syllables. Therefore duration cannot be the primary cue. Also, since intensity is not a major cue in SM, it is unlikely to be an important cue in TM, where the syllables seem to be more even. As a result, the perceptual cues used to distinguish the neutral tone from the low tone are probably pitch-related.

For this paper, I conducted a perception test using re-synthesized speech to investigate the pitch cues used by TM listeners to distinguish a neutral-tone syllable (with a low pitch target) from a Tone 3 (low tone) syllable.

2. METHOD

2.1. Subjects

The subjects were 40 Taiwan Mandarin speakers 18-55 years old without any speaking or hearing impairment, who had all grown up in Taiwan and resided in Taiwan for the past five years. Their parents were all native speakers of Taiwanese.

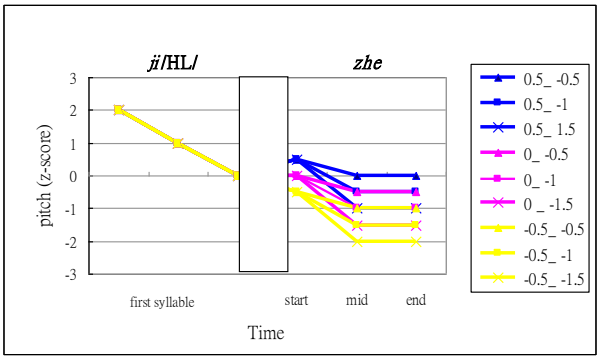
2.2. Tasks

Subjects were asked to choose between the minimal pair of full-full tones and full-neutral tones. Subjects were asked to listen to disyllabic words and to identify what the words are. The word was played in a frame sentence *qing-shuo X X ba-ci* 'please say X X eight times' twice. Listeners were presented with two options: the word with full-full tones and the word with full-neutral tones. Both words were written in Chinese characters without any phonetic spellings. Subjects were asked to circle the word they heard.

2.3. Materials

The words *ji-zhe* /HL-L/ 'reporter' and /HL-Ø/ 'remembering', and *wei-zhe* /LH-L/ 'violator' and /LH-Ø/ 'surrounding' produced by a TM female speaker with a clear voice were selected. Both the full-full and the full-neutral productions were included in order to test the possible vowel/voice qualities' effects. The durations were controlled to be the same, the average between the pairs, because the duration did not seem to be an acoustic cue for the TM listeners based on our pilot study.

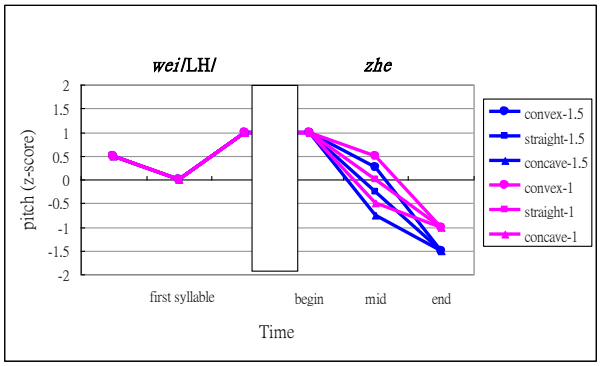
Figure 1: Schematized illustrations of synthesized stimuli *ji-zhe.*



The pitch contours of *-zhe* in *ji-zhe* /HL-L/ 'reporter' and /HL-Ø/ 'remembering' were re-synthesized into 9 pitch contours. We varied the starting pitch and the initial pitch fall because they are both cues to identify Tone 3 in TM [4]. Z-score was used for pitch height because it could more accurately reflect all the TM speakers. The pitch of TM speakers usually ranges from -2 to 2 z-score. As shown in Figure 1, three starting pitches of the second syllables were chosen: +0.5, 0, and -0.5 z-score. The pitch falls of the first half also varied between 0.5, 1, and 1.5 z-scores. In total, 2 (words of /HL-L/ and /HL-Ø/)* 3 (starting pitch)*3 (pitch falls) = 18 stimuli were created.

As for *wei-zhe*, the pitch contours of *-zhe* in /LH-L/ 'violator' and /LH-Ø/ 'surrounding' were re-synthesized into 6 different contours, varying the end pitch (-1 or -1.5 z-score) and the pitch-contour shapes (convex, straight, concave) that reached to the low end pitch (See Figure 2) because in the pilot study, these two features seem to best characterize the differences between the two tones. The pitch-contour shapes were schematized by controlling the midpoint pitches: midpoint+0.5 z-score (convex), midpoint (straight), and midpoint-0.5 z-score (concave) varying the speed of the pitch fall. 2 (words of /LH-L/ and /LH-Ø/) *2 (end pitches) *3 (shapes) = 12 stimuli of *wei-zhe* were created.

Figure 2: Schematized illustrations of synthesized stimuli *wei-zhe.*



12 varied *wei-zhe* tokens plus 18 varied *ji-zhe* tokens were tested in sentences. The 12+18=30 tested sentences were randomly mixed with 15 fillers produced by the same female speaker. The filler questions were tonal minimal pairs that differ only in lexical tones such as *zhi-yuan* /H-LH/ 'to support' vs. /LH-LH/ 'employee'. Listeners were expected to answer the fillers without difficulty.

2.4. Procedures

The recording included a 150 millisecond (ms) beep, then 150ms silence, and then the sentence, followed by a 300ms silence. The sentence was then repeated and after that there was a 3-second silence for the subjects to choose the answer. Every 15 questions, there was a short break. There

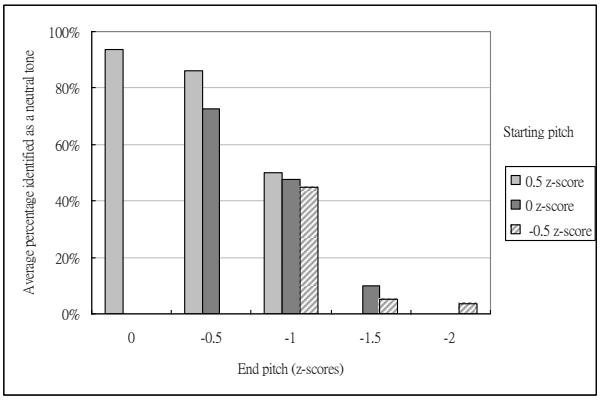
was a practice section before the test with two trials. Trial questions were produced by the same speaker, to familiarize the listeners with her voice.

3. RESULTS AND DISCUSSION

3.1. The *ji-zhe* pair

Unlike what we expected, the results show that the starting pitch and the pitch fall were not the determining factors. The results suggest that the end pitch was a critical acoustic cue for distinguishing the neutral tone from the full tone.

Figure 3: The result of *ji-zhe* by end pitch.



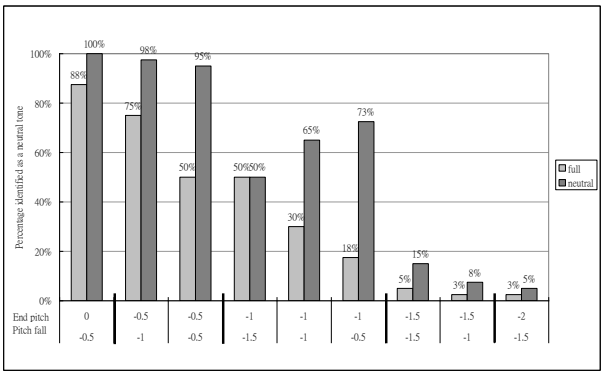
As shown in Figure 3, the end pitches (x-axis) affected the way the *ji-zhe* stimulus was perceived (y-axis). When the end pitch was higher than -1 z-score, the stimuli were identified as having a neutral tone. When the end pitch was lower than -1 z-score, the stimuli were identified as having a low tone. The listeners were confused when the end pitch was at -1 z-score regardless of the starting pitches or pitch falls.

However, aside from the end pitch, the original utterances that were used to re-synthesize the *ji-zhe* stimuli also seemed to be an important factor. As shown in Figure 4, the x-axis represents the 9 different pitch contours in the order of the end pitch (primary) and the pitch fall (secondary). With the pitch contours and the duration being equal, the vowel qualities or phonation types of the original (un-altered) utterances seemed to have an effect on the result.

Almost all the stimuli re-synthesized from the /HL-Ø/ segment (dark-shaded bars) had higher percentages of being identified as a neutral tone. The effect was especially obvious when the end pitch was -0.5 or -1 z-score. The differences between the original productions were statistically significant ($p<0.01$) except for the contour with -1 end pitch, -1.5 pitch fall. This shows that when the pitch cues were ambiguous, listeners used segmental features such as vowel quality or phonation type to distinguish the pair.

Furthermore, for these five pitch contours (end pitch: -0.5 and -1 z-score), the pitch fall seemed to play an important role too. With the end pitch being the same, when the initial pitch fall was smaller, the listeners were more likely to notice the vowel/voicing differences between the /HL-L/ and the /HL-Ø/ *ji-zhe*. The pitch fall's influence on perception was significant on the originally /HL-L/ tokens (end pitch -0.5: $p<0.05$, end pitch -1: $p<0.01$). That is, the vowel/voicing features of the /HL-L/ *ji-zhe* were more perceptible when the pitch is flatter in the mid-low pitch range.

Figure 4: The result of *ji-zhe* by end pitch, pitch fall, and original utterances.



3.2. The *wei-zhe* pair

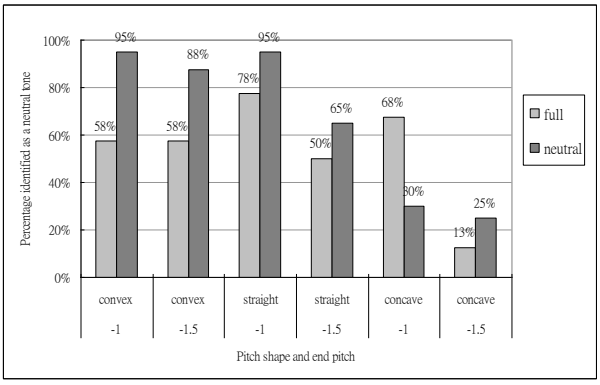
The results of the 12 *wei-zhe* stimuli are shown in Figure 5. The pitch shape and the end pitch seemed to influence how the stimuli were perceived. When the end pitch was low (-1.5 z-score) and the pitch shape was concave, the stimuli were generally identified as having a low tone (13% and 25%). The other pitch contours were either identified as having a neutral tone or were confusing.

The original utterances also played an important role. The stimuli that were re-synthesized from *wei-zhe* /LH-Ø/ (dark-shaded bars) tended to have higher percentages of identification as neutral tone compared to the originally /LH-L/ stimuli. The only exception was concave pitch shape with -1 end pitch.

When the pitch shape was convex, the end pitch did not seem to be a factor. The differences between the two original source tones were however statistically significant ($p<0.01$). The convex pitch shapes were generally perceived as a neutral tone for the originally /LH-Ø/ *wei-zhe*

(98% and 88%), but the originally /LH-L/ *wei-zhe* stimuli were ambiguous (both 58%).

Figure 5: The results of the *wei-zhe* stimuli.



When the pitch shape was straight, the end pitch was important to the perception of tone. When the end pitch was -1 z-score, both of the stimuli from /LH-L/ and /LH-Ø/ were identified as having a neutral tone. Listeners were much more confident with the stimulus re-synthesized from /LH-Ø/ (95% vs. 78%). On the other hand, with the straight pitch contour and -1.5 z-score end pitch, both of the stimuli were confusing (50% and 65%).The differences between the two end pitches were significant for both the originally /LH-L/ tokens ($p<0.05$) and the originally /LH-Ø/ tokens ($p<0.01$).

When the pitch shape was concave, the -1.5 z-score end pitch caused the stimuli to be perceived as a low tone regardless of the original segments (13% and 25%). The result of the concave shape with the end pitch -1 z-score is rather surprising as the /LH-L/ segment (68%) was perceived as a neutral tone more than the /LH-Ø/ segment (30%); the differences were significant ($p<0.05$).

4. CONCLUSION

In summary, the end pitch appears to be the most important acoustic cue in distinguishing low from neutral tone. As shown in both the *wei-zhe* and the *ji-zhe* experiments, when the end pitch was lower than -1 z-score, the stimulus was more likely to be identified as having a low tone. The results from the *ji-zhe* stimuli further showed that the perceptual boundary between Tone 3 and the neutral tone was -1 z-score, as the stimuli was more likely to be recognized as having a neutral tone when the end pitch was higher than -1 z-score. On the other hand, the pitch contour is also very important. The results from the *wei-zhe* stimuli showed that a concave pitch contour (falling more sharply in the beginning) is also necessary in order to be identified as a low tone. This result conforms to the finding of Fon, et al. [4].

When the initial pitch fall was not rapid, the pitch contour was ambiguous. As a result, the listeners utilized vowel quality or phonation type as a cue. This can be seen in the results of the *ji-zhe* stimuli with -1 z-score end pitch and the *wei-zhe* stimuli with convex or straight contours.

Even though the neutral tone in TM has a low target, the target is not as low as the low tone. Also, the speed reaching to the target is not as fast as the low tone. The acoustic features of the neutral tone are very different from those of Standard Mandarin. 'Neutral tone' in Taiwan Mandarin should be characterized as Tone 5 with the pitch of mid-low.

5. REFERENCES

[1] Chao, Y.R. 1968. *A Grammar of Spoken Chinese.* Berkeley and Los Angeles: University of California.

[2] Chen, Y., Xu, Y. 2006. Production of weak elements in speech-evidence from F0 patterns of neutral tone in standard Chinese. *Phonetica* 63, 47-75.

[3] Duanmu, S. 2007. *The Phonology of Standard Chinese.* Oxford University Press.

[4] Fon, J., Chiang, W.Y., Cheung, H. 2004. Production and perception of the two dipping tones (tone 2 and tone 3) in Taiwan Mandarin. *Journal of Chinese Linguistics* 32, 249-281.

[5] Huang, K. To be appeared. An acoustic study of the neutral tone in Taiwan Mandarin. *Proceedings of the 15th College-Wide Conference for Graduate Students in Languages, Linguistics, and Literature*.

[6] Kubler, C. 1985. The influence of Southern Min on the Mandarin of Taiwan. *Anthropological Linguistics* 27, 156-176.

[7] Lee, W.S., Zee, E. 2008. Prosodic characteristics of the neutral Tone in Beijing Mandarin. *Journal of Chinese Linguistics* 36, 1-29.

[8] Lin, M., Yan, J. 1980. Beijinghua qingsheng de shengxue xingzhi (Acoustic characteristics of neutral tone in Beijing Mandarin). *Fangyan* 3, 166-178.

[9] Lin, T. 1985. Tantao Beijinghua qingyin xingzhi de chubu shiyan (On neutral tone in Beijing Mandarin). In Hu, S. (ed.), *Beijing Yuyin Shiyanlu* (Working papers in experimental phonetics). Beijing: Beijing Daxue chubanshe (Peking University Press), 1-26.

[10] Lin, T., Wang, W.S.Y. 1984. Shengdiao ganzhi wenti (Perception of tones). *Zhongguo Yuyan Xuebao* 2, 59-69.

[11] Swihart, D.A.W. 2003. The two Mandarins: Putonghua and Guoyu. *Journal of the Chinese Language Teachers Association* 38, 103-118.