

Perception of Tone Variation

–Evidence from the Varieties of Taiwan Mandarin

Chun-Mei Chen

Department of Linguistics
The University of Texas at Austin
chench@mail.utexas.edu

Abstract

This study focuses upon voice data from mass media to disclose the tone features of Taiwan Mandarin and to identify its acoustic and perceptual varieties. The study consists of perceptual judgment and acoustic analyses of spontaneous speech. Each candidate represents one unique kind of Taiwan Mandarin. Candidate Chen's speech represents the Mandarin variety which is heavily influenced by Taiwanese. Candidate Soong's speech represents the Mandarin variety of "Mainlanders", whose parents were born and raised in one of the provinces in Mainland China. Candidate Lien's speech represents the variety that is employed by bilingual speakers of Mandarin and Taiwanese. Based on perceptual judgments, the prosodic features of each candidate are structured. The acoustic analyses interpret some but not all of the perceptual judgments. Subtle tone variants were detected by perceptual judgments, rather than acoustic measurements.

1. Introduction

In this study, "Taiwan Mandarin" (TM) refers to three typical varieties¹ of Mandarin commonly spoken in Taiwan today, which include the Mandarin spoken by people whose first language is Taiwanese (Candidate Chen), Mainlanders in Taiwan (Candidate Soong), and the bilingual speakers of Mandarin and Taiwanese (Candidate Lien).

Tone varieties in Taiwan Mandarin have been broadly discussed in earlier work. Kubler (1985) has pointed out that with two important exceptions, the tones of Taiwan Mandarin are the same as those of standard Mandarin. The exceptions are: (1) the third tone usually has no rise; (2) the neutral tone occurs much less frequently in Taiwan Mandarin than in standard Mandarin. Chen (1991) conducted a survey in Taipei to see the shift and diffusion of tones. In Chen's (1991) survey, the mix of the second and third tones in Taiwan Mandarin occurs more frequently. Whether neutral tone is employed in Taiwan Mandarin has been a big issue in the past decade. Tseng (1999) found that the neutral tone in Taiwan Mandarin is more like the fifth tone. Its prosodic properties are low pitch with short duration, similar to an entering tone of Taiwanese. Feng's (1997) empirical study has shown that Taiwan Mandarin uses a different tonal system than what Chao (1968) proposed. Instead of having 55, 35, 214 and 51 (falling), the four tones are changed to 44, 323, 312, and 42. Pitch height and pitch range are lower and narrower than what Chao (1968) has reported. Tone 2 is not a rising tone, but dipping in nature. Instead of a five-scaled system, Taiwan Mandarin is sustained by a four-scaled interval. However, the

factor of varieties in Taiwan Mandarin has not been taken into account in Feng's (1997) study.

As far as perception is concerned, in Tseng's (1981) study of tones in Mandarin Chinese, she found that when forced to identify lexical tones produced in citation, listeners performed almost perfectly, which was shown in complementary perception studies. However, when forced to identify lexical tones edited from spontaneous speech, listeners' performance was relatively poor. It is clear that the actual production and perception of spontaneous speech is very likely a cognitive decision involving interacting linguistic/extra-linguistic factors. Borden *et al.* (1994) have pointed out the importance of context to speech perception. Contextual effect on perception is apparent in the recovery of suprasegmental information. Formants need not be of particular frequencies to be recognized as vowels, but they must bear a certain general relationship to one another.

On the other hand, native speaker's intuition does play a role in perception identification and judgment. Gussenhoven and Jacobs (1998) have proved the effect of native-speaker's intuitions in determining different sounds. Native speakers appear to make reference to the lexical representation when determining whether two phonetically different sounds are the same sound or two different sounds.

The goal of this study is to identify the perceptual cues of tone varieties in Taiwan Mandarin, based on the judgments of the Mandarin teachers, TM native speakers and American listeners. This study may have pedagogical implications for second language learners that there are different varieties of Taiwan Mandarin and that language learners have to adjust their listening comprehension of Taiwan Mandarin.

2. Between Acoustic and Perceptual Variance

2.1. Research methodology

The voice data was recorded from the speech of the three presidential candidates during presidential campaign in 2000. Sources of the voice data include the recordings of ETTV News, CTN News, TVBS-N and TVBS on Taiwan TV channels. The voice data were gathered from February 1st to February 22nd in 2000. Voice editing equipment is TASCAM MD Recorder/ Player. After we compiled the digital materials for the experiment, which contain 5128 tokens of segments with 1658 tokens of Chen's voice tokens, 1740 tokens of Soong's voice data, and 1730 tokens of Lien's voice data, Computerized Speech Lab (CSL) Model 4300B was employed to conduct acoustic measurements.

First perceptual judgment was conducted to survey speech prosody of the candidates, and to discover the patterns of Taiwan Mandarin as defined by ten experienced Mandarin

¹ Hakka and aboriginal languages are not discussed here.

teachers. Second perceptual judgment was conducted to investigate how varieties of tones in Taiwan Mandarin were perceived by American listeners and native Taiwan Mandarin speakers on the one hand, and to investigate the differences in the perception of Taiwan Mandarin between native Taiwan Mandarin speakers and American listeners on the other. Thirty-one subjects ages 18-30 participated in the second perceptual judgment.

2.2. Perception of the TM native speakers

All the Mandarin teachers participated in this study are female. None has a history of hearing or speech impairments. The average age of the Mandarin teachers is 35.3. They are experienced in teaching Mandarin, with an average of eleven years of teaching experience.

Based on quantitative and qualitative studies, it was found that Candidate Chen's second tone was perceived as the third tone. Though Soong had only few tone variants, the deviant neutral tone in the syllable /səŋ/ was noticed by half of the Mandarin teachers. Candidate Lien's tokens were perceived as having the most variations. The second tone was perceived as shifting to the first, third and neutral tones. Lien's falling tone (Tone 4) was perceived as the rising tone (Tone 2). We claim that this type of variants (4→2) is due to Lien's false realization of lexical tone, for some cognates in Mandarin second tone are pronounced as abrupt tone (Rù shēng, 'entering tone') in Taiwanese. Furthermore, his fourth tone is perceived to be the variants of the first, second, third, and neutral tones. All the Mandarin teachers detected the deviants of Lien's neutral tone.

The percentage perceived by the Mandarin teachers as correct tones is shown in Figure 1.

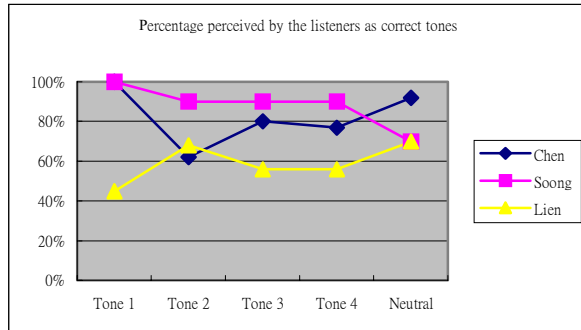


Figure 1: Percentage of the candidate's tokens perceived by the Mandarin teachers as correct tones

It is clear that Candidate Lien's tones were perceived as having more deviants than those of the other two. Note that both Chen's and Lien's second tones were perceived as merged into their third tones. This is exemplified as follows:

(1) Tone merger: Tone 2 → Tone 3

Standard Mandarin	Taiwan Mandarin	Gloss
[liou ¹]	(Chen) [liou ⁴]	'flow'
[tʂ ^h l ¹]	(Chen) [tʂ ^h l ⁴]	'hold'
[tci ¹]	(Lien) [tci ⁴]	'level'
[fu ¹]	(Lien) [fu ⁴]	'happiness'

This finding was confirmed by the second perception judgment. Sixty percent of the native TM listeners perceived Candidate Chen's second tone as third tone. Though it can be judged from the context that the token holds the fourth tone, none of the native TM listener perceived Lien's token [tʂ^hl⁴] as holding the fourth tone. Twenty-four percent of the native TM listeners perceived Lien's falling tone as rising tone.

2.3. Perception of the American learners

As we have predicted, most of the American listeners were not able to identify the tones when they were asked to identify the segments simultaneously. Nevertheless, two subjects perceived Candidate Chen's second tone as the third tone. This finding confirms the results obtained from the perceptions of Mandarin teachers and native TM listeners. Note that none of the American listeners correctly identified the fourth tone of Lien's tokens.

2.4. Tone scale of the three candidates

Based on the data obtained from acoustic measurements, the scales of the candidate's four tones are constructed. The tokens selected in this section are /tci¹/, /tci²/, /tci³/ and /tci⁴/, i.e., first tone (tone 1), second tone (tone 2), third tone (tone 3) and fourth tone (tone 4). Each token was extracted from the identical context of the three candidates' speech. There are total 36 tokens examined in this section. After the pitch in each token was measured, a logarithm with the base number 10 was employed to calculate the range on a five-scale system. The results are shown in Table 1, Table 2 and Table 3.

Table 1: Candidate Chen's Tone Scale

Tone	Pitch (Hz)	LOG10	Five-Scale Value
Tone 1	214.6	2.33163	44
Tone 2	170 – 228	2.231215 - 2.358316	14
Tone 3	173 – 156	2.238464 - 2.193357	21
Tone 4	262 – 210	2.418429 - 2.321423	53

Table 2: Candidate Soong's Tone Scale

Tone	Pitch (Hz)	LOG10	Five-Scale Value
Tone1	230	2.36157	55
Tone 2	156 – 185	2.193357 – 2.265996	24
Tone 3	148 – 132	2.169086 – 2.119696	21

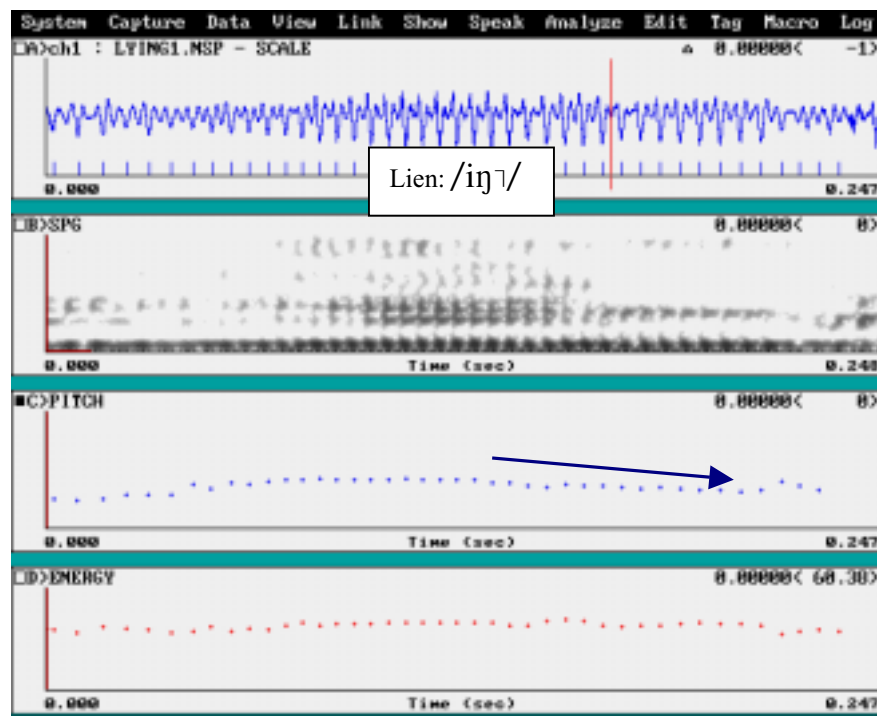


Figure 2: Lien's voice token /iŋ/

Tone 4	221 – 175	2.345178 – 2.242044	53
---------------	-----------	---------------------	-----------

Table 3: Candidate Lien's Tone Scale

Tone	Pitch (Hz)	LOG10	Five-Scale Value
Tone1	182	2.259275	33
Tone 2	157 – 177	2.195069 – 2.248219	13
Tone 3	164 – 157	2.213606 – 2.194977	11
Tone 4	201 – 177	2.303785 – 2.246857	53

As shown in Table 1, Table 2 and Table 3, the value of the three candidates' fourth tone is [53]. The falling tone of the three candidates is not sharp, as the slope is from the highest scale [5] to the two-lower scale [3]. As far as pitch height is concerned, Candidate Chen's speech presents the maximum pitch (262Hz), whereas Soong's speech presents the minimum pitch (132Hz). Candidate Lien's level tone value is the lowest whereas the value of Soong's level tone is the highest on the five-scale system. Candidate Chen's first tone confirms Feng's (1999) finding, in which the first tone in Taiwan Mandarin is changed to 44. Chen's second tone has three levels of transition, rising from scale [1] to [4], which presents more

rising property of the second tone than that described in earlier work. The third tone of the three candidates' voice tokens is judged to have the lowest pitch. In our study, Candidate Lien's third tone presents the property of a level tone, which confirms Kubler's (1985) claim that the third tone usually has no rise. Though the second tone of the three candidates does not reach the highest scale [5], Lien's second tone apparently presents both low and level properties, as it ranges from scale [1] to scale [3]. This narrow pitch range of Lien's tone might be perceived as variation.

Note that these tokens were selected from continuous speech, rather than individual citations, and that this has a bearing on how they were perceived by the listeners. The value of the four tones, thus, may be affected by adjacent syllables. The third tone presented here is only its falling transitions, rather than a complete falling-rising contour.

2.5. Comparison of acoustic and perceptual variance

In this section, we examine the tone variants of candidate's tokens. Note that the dot lines in the figure indicate pitch transitions.

Shown in Figure 2, there is slight descent in the second half of the pitch transition. It is surprising such a subtle transition was detected by the listeners and thus resulted in the variant being perceived as the fourth tone, though the token should hold the first tone (Tone 1>Tone 4). In Lien's token /tʂl-/ , we detect the level feature from the pitch contour. This phenomenon explains why the third tone of this token was perceived as the first tone.

As far as neutral tone is concerned, it is apparent that the property of neutral tone is not present in Lien's token /ʂəŋ/ ,

whereas there is a downstream movement in Soong's token. Interestingly, both Soong's and Lien's tokens were perceived as the variant of first tone. This finding indicates that Soong's realization of the neutral tone did not reach the perceptual categorical threshold of the listeners, to be perceived as a neutral tone, even though it has the property of the neutral tone based on acoustic measurements. In the token of neutral tone /mən/, Chen's and Soong's tokens exhibit the property of the second tone, which corresponds to the results obtained from the perceptual judgments. Lien's neutral tone, though presents the property of a level tone, was not perceived as variant. We conclude that the attention of the listeners and individual's perceptual categorization have effect on whether a token is perceived as a neutral tone.

2.6. Effect of speech rate

The duration and the syllables of each sentence were measured to examine the speed rate of each candidate. The rate of speech can be expressed by Equation (1):

$$\text{Rate} = \frac{\text{Syllable(Num)}}{\text{Duration (sec)}} \text{ (syllables per second)} \quad (1)$$

The pauses in each sentence were ignored here, because the voice data were extracted from continuous speech. The intervals between syllables are too short to be measured. The results are presented in Table 4.

Table 4: The rate of speech of the three candidates
(Unit: syllables per second)

Rate	Chen	Soong	Lien
Max Rate	4.08	4.60	5.65
Min Rate	1.91	2.43	1.88
Average	2.99	3.23	3.18
N	35	35	35
SD	0.60	0.47	0.88

As shown in Table 4, Candidate Soong's rate of speech is slight higher than the rate of the other two. This indicates Soong's speaking is faster. All of the three candidates uttered around 3 syllables per second, though Lien uttered a maximum of 5.65 syllables in one second whereas Chen's maximum was only 4.08 syllables. Soong's rate of speaking is also somewhat more consistent than the other two, as indicated by the lower standard deviation in the column. Nevertheless, without taking the context of the speech and other para-linguistic factors such as the occasions into consideration, the claims about the candidates' rate of speech in this study intend to be somewhat conservative. One point we can make here is that the effect of speech rate on the tone representation is not apparent. Nevertheless, given the fact that Lien's speech was perceived to have more tone variation, his speech is indeed slighter faster than Chen's. Rate of speech may also have effect on the number of tone variations obtained from perceptual judgments.

3. Discussion and Conclusion

The feature of the tone variant in Taiwan Mandarin is expressed in (1): Tone 2 → Tone 3. The perception of the second tone merged into third tone could be attributed to the low starting point of second tone in Candidate Chen's speech and Lien's speech, shown on the scale as [1]. This indicates the starting point of a tone may be the perceptual cue for the categorization. Categorical perception occurs when we perceived stimuli in terms of categories so that we can discriminate between stimuli from different categories but can't discriminate between stimuli in the same category. Based on this fact, the perception of tone variant must rely on both physical detection and psychological reality. As shown in our study, subtle tone transition was judged to be variants by both native Taiwan Mandarin speakers and American listeners, though the acoustic measurements do not provide strong evidence for the variants. On the other hand, attention of the listeners and individual's perceptual categorization also have effect on whether a token is perceived as a neutral tone.

Candidate Chen's first tone is changed to 44. His second tone has three levels of transition, rising from scale [1] to [4], which presents more rising property of the second tone than that described in earlier work. The third tone of the three candidates' voice tokens is judged to have the lowest pitch. In our study, Candidate Lien's third tone presents the property of a level tone. The third tone in the three varieties of Taiwan Mandarin has no rise. In other words, the third tone presented here is only its falling transitions, rather than a complete falling-rising contour. This is one of the important factors why the merger between second and third tone was perceived in the perception judgments.

4. References

- [1] Borden, Gloria J., Katherine S. Harris & Lawrence J. Raphael. 1994. *Speech Science Primer*. Physiology, Acoustics, and Perception of Speech. Third Edition. Baltimore: Williams & Wilkins.
- [2] Chao, Yuen-Ren. 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.
- [3] Chen, Chung-Yu. 1991. Shēngdiào de Zhuǎnbìan yǔ Kuòsàn [The Shift and Diffusion of the Tones]: The sampling of different generations in Taipei. *Journal of the Chinese Teachers Association*, Vol. 24, No.1: 69-99.
- [4] Feng, Yi-chen. 1997 "What are tones really like?--An acoustic-based study of Taiwan Mandarin Tones". M.A. thesis, National Taiwan University.
- [5] Gussenhoven, Carlos & Haike Jacobs. 1998. *Understanding Phonology*. Great Britain: Arnold. Co-published in USA by Oxford University Press, Inc..
- [6] Kubler, Cornelius C.. 1985. *The Development of Mandarin in Taiwan: A Case Study of Language Contact*. Taipei: Student Book Co., Ltd.
- [7] Tseng, Chin-Chin. 1999. *A contrastive analysis on the pronunciation between Mainland China and Taiwan: A case study of news reporters*. NSC Project: No 88-2411-H-003-017. Taipei: National Science Committee.
- [8] Tseng, Chiu-Yu. 1981. "An Acoustic Phonetic Study on Tones in Mandarin Chinese" Ph. D. diss., Brown University.