行政院國家科學委員會專題研究計畫 成果報告

地域及語體對台灣地區國語一聲調闊之影響
研究成果報告(精簡版)

計 畫 類 別 ：個別型
計 畫 編 號 ：NSC 95-2411-H-002-046-
執 行 期 間 ：95 年 08 月 01 日至 96 年 09 月 30 日
執 行 單 位 ：國立臺灣大學語言學研究所

計 畫 主 持 人：馮怡蓁

計畫參與人員：博士班研究生-兼任助理：陳萱芳
碩士班研究生-兼任助理：林欣怡、黃宜萱、莊育穎、吳怡臻

報 告 附 件 ：出席國際會議研究心得報告及發表論文

處 理 方 式 ：本計畫可公開查詢

中 華 民 國 96 年 12 月 27 日

行政院國家科學委員會補助專題研究計畫 ■ 成 果 報 告
□期中進度報告

地域及語體對台灣地區國語一聲調闊之影響

計畫類別：■ 個別型計畫　　□ 整合型計畫
計畫編號：NSC 95－2411－H－002－046－
執行期間：2006 年 08 月 01 日至 2007 年 09 月 30 日

計畫主持人：馮怡蓁
共同主持人：無
計畫參與人員：陳萱芳、林欣怡、黃宜萱、莊育穎及吳怡臻

成果報告類型(依經費核定清單規定繳交)：■精簡報告　□完整報告

本成果報告包括以下應繳交之附件：
□赴國外出差或研習心得報告一份
□赴大陸地區出差或研習心得報告一份
■出席國際學術會議心得報告及發表之論文各一份
□國際合作研究計畫國外研究報告書一份

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢
□涉及專利或其他智慧財產權，□一年□二年後可公開查詢

執行單位：國立台灣大學語言學研究所

中　華　民　國　九十六　年　十二　月　二十七　日

一、 中文摘要

本計劃探討方言差異影響台灣地區國語之調閾與音閾下降的情形。我們總共徵求了12位語言使用者，其中六位來自北部方言區(「標準」方言)，六立來自中部方言區(「非標準」方言)，以進行一項閱讀作業。一聲的目標字皆放置於裝載句中，共有三種不同的句位——句首、句中與句末。應變數則量測目標字之基頻極大值。結果顯示，北部方言一聲的調閾高於中部方言，而且此種現象男性較女性明顯。音閾下降的強度亦受性別與方言影響。中部男性語言使用者音閾下降坡度較北部男性為緩，而中部女性語言使用者則傾向以提高起始音高的方式，使其音閾下降坡度較北部女性為陡，顯示女性使用者較男性對於自身所使用之方言發音有自覺，因此較易以較大的音閾來過度矯正其非標準口音。

關鍵詞：國語、聲調調閾、一聲

二、 英文摘要

This study investigated how dialectal variations can influence the realization of tonal register and declination pattern in Taiwan Mandarin. Twelve speakers, six from the Northern dialect (standard) and six from the Central dialect (nonstandard), were recruited for a reading task. Target syllables of Tone 1 (T1) were embedded in a carrier sentence in three different sentential positions, initial, medial, and final. The dependent measure was the F_0 maximum of the target syllables. Results showed that the Northern dialect had higher T1 register than the Central dialect, with the effect being more prominent in males than females. The magnitude of declination was also a function of gender and dialect. Central male speakers produced milder declination slopes than their Northern counterparts while Central female speakers tended to have steeper slopes instead by raising the initial starting pitch, indicating that in the Central dialect, female speakers were more aware of their vernacular speech status compared to male speakers and tended to hypercorrect their nonstandard accent by using a larger pitch range.

Keywords: Guoyu, tonal range, Tone 1

三、 前言 (Introduction)

The linguistic situation in Taiwan is anything but monolithic. Although the official language is Mandarin, nearly 80% of the population is ethnically Min, who speaks a variant of Southern Min (Chen, 1989; Cheng, 1985). Therefore, Min acts as a powerful substrate language for Taiwan Mandarin.

Both Mandarin and Min are tone languages. Mandarin has four lexical tones, which are high level (Tone 1), mid dipping (Tone 2), low dipping/falling (Tone 3), and high-falling (Tone 4) (Chao, 1968; Fon & Chiang, 1999; Fon, Chiang, & Cheung, 2004), while Min has five long tones, which are high level (Tone 1), high falling (Tone 2), low falling (Tone 3), mid dipping (Tone 5), and mid level (Tone 7). Although the two languages have comparable tonal number and categories, the tonal range used in Min is somewhat lower than that in Mandarin (Chen, 2005; Hsu, 2006).

Due to the Mandarin-only language policy enforced by the government between 1945 and 1990, the relative statuses of the two languages have become unequal (Huang, 1993). Mandarin is promoted as a high language and is used extensively in public domains and formal contexts while Min is demoted as a low language and is often limited to private domains and informal contexts. In addition, the degree of bilingual proficiency is also geographically imbalanced. As Taipei is the political and economic center of the country, more people became monolingual Mandarin speakers or unequal Mandarin-Min bilingual speakers who are dominant in Mandarin. On the other hand, in cities that are more down south, there are more people that are equal bilingual speakers of Mandarin and Min, or even unequal bilinguals that are dominant in Min.

Therefore, in this study, we would like to investigate whether there is any difference in the tonal range by Mandarin speakers of different dialects. Specifically, the Northern dialect, which is the standard variety, and the Central dialect, a nonstandard variety, were chosen. The Northern dialect is spoken in the Taipei Metropolitan area while the Central dialect is spoken in the Taichung Metropolitan area.

四、研究目的 (Aims)

There are three specific aims in this study. First of all, we would like to explore possible dialectal differences in tonal register. If the low tonal range of Min could be carried over to that of Mandarin, one would expect this effect to be stronger in the Central dialect than the Northern dialect, as there are more fluent Min speakers in the former area than the latter. Therefore, the tonal register of the Central dialect should be lower than that of the Northern

dialect.

Secondly, if tonal register is indeed lower in nonstandard varieties, then one would suspect this difference to also influence the declination pattern, since declination exerts differential effects on high and low tonal targets, being more prominent on the former than the latter (Ladd, 1988). In other words, one would expect the topline declination of the Northern variety to be steeper than that of the Central variety, since the latter has a more restricted and lower tonal register.

Thirdly, we would like to explore whether gender would also influence tonal realization with regards to register and declination patterns in the two dialects. Sociolinguistic studies have demonstrated that women on average are more likely to conform to standard linguistic forms than men (Trudgill, 2000). If so, we would expect women in the Central dialect to be closer to those in the Northern variety than their male counterparts.

五、 文獻探討 (Literature review)

In the early 1970s, acoustic measurements of the four tones in Mandarin are fairly consistent with their names (Howie, 1976). Tone 1 (T1), being a high-level tone, is on average 150 Hz, Tone 2 (T2), a high-rising tone, ends at about 150 Hz, Tone 3 (T3), a low-dipping tone, starts at around 135 Hz, and Tone 4 (T4), a high-falling tone, starts at about 150 Hz.[1] In other words, acoustic measurements match perfectly with the descriptive names. T4 starts high, at a pitch height that is equivalent to the average pitch height of T1 and the final pitch height of T2.

However, Shih's (1988) results showed the matching relationship between acoustic measurements and descriptive names seems to have changed.[2] Of the four tones in Mandarin, the initial of T4 is the highest, with an average of 290 Hz, and T1 and T2, although termed high-level and high-falling, respectively, do not live up to their fame. The average pitch height of T1 is only 260 Hz, and the final pitch height of t2 is only 210 Hz. At around the same time, Tseng (1990) also did a series of studies on Mandarin tones. Results showed that the beginning of T4 and the end of T2 are about 245 Hz and 250 Hz, respectively.[3] On the other hand, the average pitch of T1 is 215 Hz and the beginning pitch of T3 is 135 Hz. Fon (1997) also studied a Taipei female Min-origin speaker whose native language is Mandarin and found that the beginning pitch of T4 is still the highest, around 255 Hz, while the average of T1 is 240 Hz, the ending pitch of T2 is 220 Hz, and the beginning pitch of T3 is 220 Hz.

[1] Howie (1976) studied a male speaker.
[2] Shen (1988) studied a female speaker.
[3] Tseng (1990) studied a female speaker.

From above studies, one can conclude that within ten to twenty years after Howie's (1976) study, the tonal range of high tones in Mandarin, such as T1 and T2, is shifted lower. However, individual differences still exist. Therefore, we have suggested that revision should be made for Taiwan Mandarin using Chao's five-point tone letter system (1956; 1968). Instead of the original 55, 35, 214, and 51, it should become 44, 323, 312, and 42, to match the acoustic realization (Fon & Chiang, 1999; Fon, 1997).

Since acoustic measurements are very time-consuming, not many studies included large amount of data to compare tonal range variations in the Mandarin tonal system.[4] Fon and Hsu (2007) studied five young Min-origin Mandarin speakers (ages between 20 and 25) in a reading task to compare T2 and T3 variations in various syllable structures and sentence positions. Results showed that aside from sentence-final accented positions in which the ending pitch of T2 is somewhat higher than the initial of T3 (110Hz : 100Hz), tones in isolation and in other positions have comparable pitch height at these two reference points (around 120 Hz), indicating that the ending pitch of T2 has already lowered to the mid pitch range, and there is a need to revise the descriptive name of "high-rising" tone. S. Hsu (2006) collected 24 college students' read speech, and also found that Mandarin-Min bilingual speakers tend to have lower tonal range for T2 than Mandarin monolingual speakers. The former group is inclined to realize T2 in the low pitch range, accounting for 21% of its total T2, while the latter group only realized 11% of T2 in the low pitch range.

In addition, H. Hsu (2006) also recorded 48 Mandarin speakers' read speech in order to investigate the relationship among T4 range variation, age, and ethnic group. Results showed that in general, elder speakers (ages 45 to 54) have wider tonal ranges than younger ones (ages 15 to 24) (40 Hz : 18 Hz). Interestingly, there are also variations among the elder speakers. Those whose parents are ethnically non-Min, or whose father is non-Min and mother is Min tend to have a wider tonal range than those whose parents are both Min (45 Hz : 33 Hz). However, such an interaction effect was not found in the young group, indicating that the ethnic difference in Mandarin tones is diminishing.

六、 研究方法 (Methods)

1. 受試者 (Subjects)

Twelve participants between ages 19 and 24 took part in this study. Half of them were from the Taipei Metropolitan area (the Northern dialect), and half of them were from the Taichung Metropolitan area (the Central dialect). Within each dialect group, there was an equal gender split. All of the subjects were Mandarin-Min bilinguals that were ethnically Min,

[4] Previous studies usually included only one subject.

but the Taipei speakers could not speak Min fluently.

2. 實驗刺激 (Stimuli)

Twenty-seven Tone 1 (T1) syllables representative of Mandarin phonotactics were chosen as stimuli, including 6 voiceless obstruent-initials (e.g., *han*^{1} [xan] 'charmingly naive'), 15 sonorant-initials (e.g., *la*^{1} [la] 'pull'), and 6 vowel-initials (e.g., *wu*^{1} [u] 'black'). T1 syllables were chosen because they contain only high tonal targets and thus potential register and declination differences could be most clearly observed without being clouded by the presence of low tones. The relatively high pitch also guarantees modal voice and thus successful pitch extraction. Syllables were placed in three comparable carrier sentences, so that they occurred in sentence-initial, -medial, and -final positions (Table 1). Carrier sentences were designed so that syllables immediately before the target in medial and final positions ended mid, which would also be the starting point for the initial position according to the PENTA model (Xu, 2005). In total, 27 (stimuli) × 3 (positions) = 81 sentences were recorded.

Table 1: Carrier sentences used in this study.

	Carrier Sentence
Initial	***'X'*** *zhe*^{4} *ge*^{0} *zi*^{4} *hen*^{3} *nan*^{2} *nian*^{4}.
	'X' this word is very hard to read.
Medial	*zhe*^{4} *ge*^{0} ***'X'*** *zi*^{4} *hen*^{3} *nan*^{2} *nian*^{4}
	This word 'X' is very hard to read.
Final	*zhe*^{4} *ge*^{0} *zi*^{4} *shi*^{4} *nian*^{4} ***'X'***
	This syllable indeed reads 'X'

3. 實驗儀器 (Equipment)

Recordings were done using a SONY PCM-M1 Digital Audio Recorder with Maxell R-64 DA 60 min DAT tapes and a SHURE SM10A head-mounted microphone.

4. 實驗過程 (Equipment)

The experiment was conducted in a quiet room. Speakers were asked to read aloud the semi-randomized stimuli using natural intonation at a normal rate. The whole process took about 15 minutes. The original recordings had a sampling rate of 48 kHz, which were

subsequently downsampled to 16 bit 22050 kHz using Adobe Audition 1.5.

5. 實驗分析方法 (Analysis)

The recordings were hand-labeled using Praat 4.6 (Boersma & Weenink, 2007). A Praat script was written for automatic pitch extraction on the voiced portion of the syllable, which is considered the measurable domain for tones (Chao, 1956; 1968; Wang, 1967; Xu, 1998). For obstruent-initial syllables, the starting point of a tone was determined by the onset of the voice bar after the obstruent, which was voiceless in this study. For the rest, the starting point began from the onset of the syllable, as the whole syllable was voiced. The ending point was always the offset of the voice bar. Occasional syllable-initial or -final glottalized portions caused by voice fry were not included for pitch extraction. Extracted pitch tracks were later hand-checked and hand-corrected for doubling and halving through pitch period calculation, and were interpolated and smoothed using Praat functions afterwards. A second Praat script was written to extract F_0 maximum for the target stimuli and the syllable *zhe*^{4} 'this', which, being a high-falling tone, acted as a reference point for the sentence.

七、 結果 (Results)

Subjects were asked to rate their Mandarin and Min fluency on a scale of 1 to 7. Since Mandarin was their native and most frequently used language, Min fluency was measured by dividing Min fluency scores by Mandarin fluency scores. As shown in Figure 1, Taipei speakers in general had lower degrees of Min fluency than Taichung speakers, with female speakers being even less fluent than males. On the other hand, not much gender difference was found in Taichung speakers. Both males and females were fairly fluent in both languages.

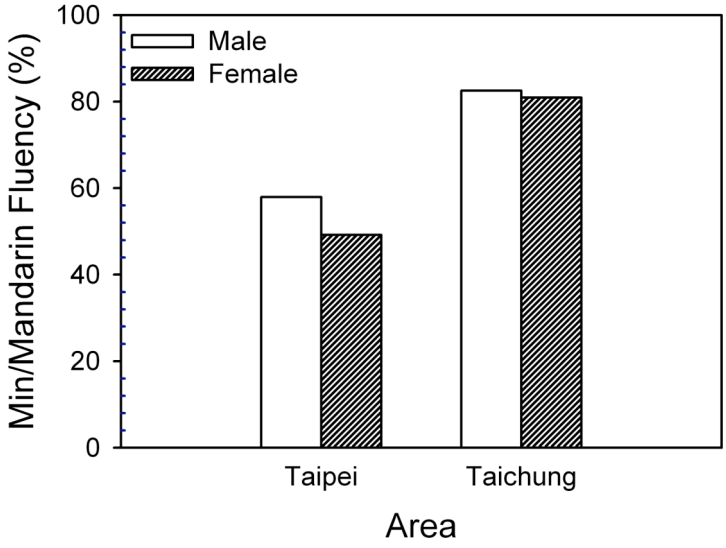


Figure 1: A bar graph of self-rated Min fluency relative to subjects' Mandarin fluency.

1. 「這」與目標字之基頻極大值 (F_0 maximum of *zhe*^{4} and the target syllable)

Figure 2 shows scatter plots of the F_0 maximum of *zhe*^{4} and the target syllables for the two dialects of Mandarin in sentence-initial, -medial, and -final positions. The cluster in the upper right corner in each graph represents female data and the cluster in the lower left corner represents male data. As can be seen from the figure, Taichung male speakers used lower F_0 maximum than Taipei speakers for both *zhe*^{4} and target syllables, indicating an overall lowering in the tonal register. On the other hand, for female speakers, the distinction was not as clear. The overall pitch range for *zhe*^{4} was fairly similar for the two dialect groups, with Taichung speakers being even slightly higher than Taipei ones. This was especially prominent in sentence-medial positions. However, for target syllables, Taipei speakers were still higher in pitch than those of Taichung.

A Gender (2) × Dialect (2) × Position (3) × Syllable (*zhe*^{4} *vs.* target) four-way mixed ANOVA was performed to confirm the above observations. The between-factors were Gender and Dialect. Results showed that all of the main effects were significant [Gender: $F(1, 318) = 18.94$, $p < .0001$, $\hat{\eta}^2 = .96$; Dialect: $F(1, 318) = 18.68$, $p < .0001$, $\hat{\eta}^2 = .06$; Position: $F(2, 636) = 172.60$, $p < .0001$, $\hat{\eta}^2 = .35$; Syllable: $F(1, 318) = 1498.8$, $p < .0001$, $\hat{\eta}^2 = .83$]. Five of the two-way interactions were also significant [Position × Gender: $F(2, 636) = 9.61$, p

< .0001, $\hat{\eta}^2$ = .03 ; Position × Dialect: $F(2, 636) = 4.79$, $p < .01$, $\hat{\eta}^2$ = .02; Syllable × Gender: $F(1, 318) = 348.59$, $p < .0001$, $\hat{\eta}^2$ = .52; Syllable × Dialect: $F(1, 318) = 33.72$, $p < .0001$, $\hat{\eta}^2$ = .10; Position × Syllable: $F(1.97, 624.89) = 665.60$, $p < .0001$, $\hat{\eta}^2$ = .68]. In addition, all of the three-way interactions involving Syllable were significant [Syllable × Gender × Dialect: $F(1, 318) = 68.42$, $p < .0001$, $\hat{\eta}^2$ = .18; Syllable × Gender × Position: F(1.97, 624.89) = 128.69, $p < .0001$, $\hat{\eta}^2$ = .29; Syllable × Dialect × Position: F(1.97, 624.89) = 15.45, $p < .0001$, $\hat{\eta}^2$ = .05]. Finally, the four-way interaction was also significant [$F(1.97, 624.89) = 29.90$, $p < .0001$, $\hat{\eta}^2$ = .09].

Post hoc independent *t*-tests regarding the Dialect effect indicated that the F_0 maximum of Taipei male speakers was significantly higher than their Taichung counterparts across both syllable types and sentential positions ($p < .0001$). However, this was not the case for female speakers. In the initial position, there was no difference between Taipei and Taichung speakers in *zhe*^{4} or the target syllables. In the medial and final positions, target syllables of Taipei speakers were significantly higher than those of Taichung (Medial: $p < .0001$; Final: $p < .05$), while for *zhe*^{4}, Taichung speakers were in turn higher than Taipei speakers (Medial: $p < .0001$; Final: $p < .05$).

2. 頂線音闕下降 (Topline declination)

Figure 3 shows the topline declination pattern of male and female speakers in the two dialect groups. For male speakers, the declination trend was steeper in the Northern dialect in sentence-medial and -final positions. Female speakers demonstrated an exactly opposite trend, with the Central dialect showing a steeper topline than the Northern variety, especially in sentence-medial positions.

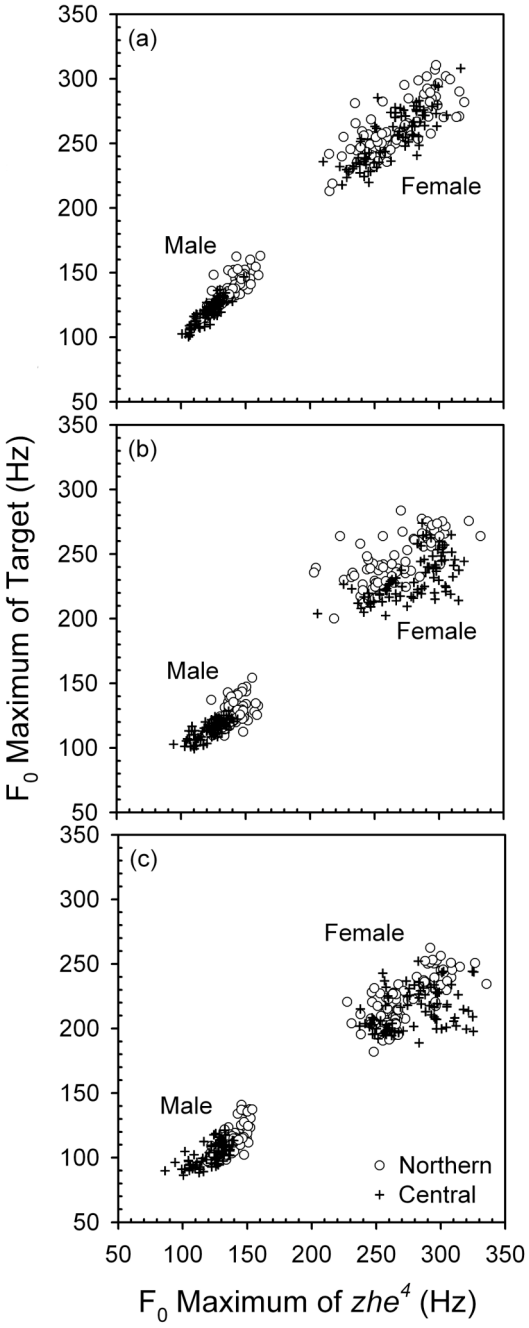


Figure 2: Scatter plots of F_0 maxima of zhe^4 and target syllables in sentence (a) -initial, (b) -medial, and (c) -final positions.

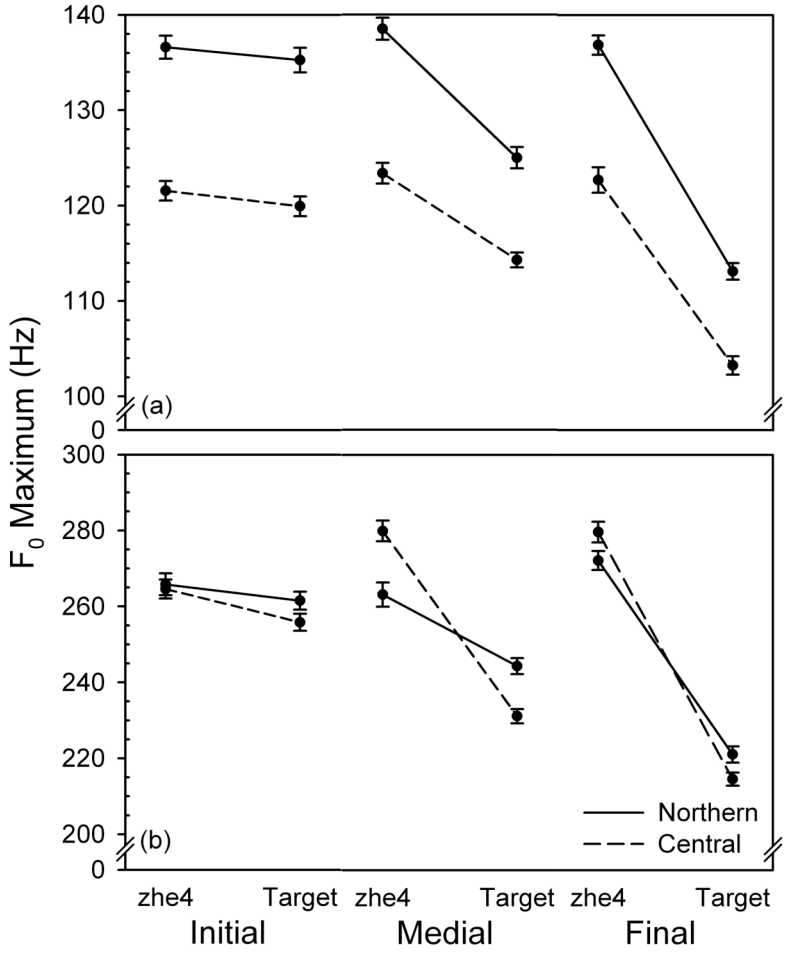


Figure 3: Mean F_0 maxima of zhe[4] and target syllables of (a) male and (b) female speakers.

In order to confirm what was observed above, a Gender (2) × Dialect (2) × Position (3) three-way mixed ANOVA was conducted on degree of declination, defined by F_0 maximum difference between the reference *zhe*[4] and the target syllable. Results showed that all of the main effects were significant [Gender: $F(1, 318) = 348.91$, $p < .0001$, $\hat{\eta}^2 = .52$; Dialect: $F(1, 318) = 33.59$, $p < .0001$, $\hat{\eta}^2 = .10$; Position: $F(1.97, 625.11) = 665.13$, $p < .0001$, $\hat{\eta}^2 = .68$]. All of the two-way interactions were also significant [Position × Gender: $F(1.97, 625.11) = 128.56$, $p < .0001$, $\hat{\eta}^2 = .29$; Position × Dialect: $F(1.97, 625.11) = 15.52$, $p < .0001$, $\hat{\eta}^2 = .05$; Gender × Dialect: $F(1, 318) = 68.61$, $p < .0001$, $\hat{\eta}^2 = .18$]. The three-way interaction was significant as well [$F(1.97, 625.11) = 19.92$, $p < .0001$, $\hat{\eta}^2 = .09$].

Post hoc independent *t*-tests regarding the Dialect effect showed that F_0 drop was much

larger in Taipei male speakers than those of Taichung in medial and final positions, the difference being greater in the latter than the former position (Medial: $p < .0001$; Final: $p = .001$). For female speakers, the difference was also significant in medial and final positions ($p < .0001$), but the trend was the opposite of male speakers. Taichung females tended to have a larger F_0 drop than Taipei females, with the difference being greater in the sentence-medial than the final position. In the initial position, however, both genders show no dialectal differences.

八、 討論 (Discussion)

Results in this study showed that variation did exist between the two dialects of Taiwan Mandarin for T1. In general, the Central dialect tended to have a lower tonal register than the Northern dialect. This is consistent with our predictions. Since Taichung speakers were more fluent in Min, their Mandarin tonal register would more likely be influenced and became lower in pitch. However, such an effect was more prominent in male than in female speakers. In female speakers, only sentence-medial and -final positions showed such an effect, while in male speakers, all three positions showed the same trend. The gender differences could not have been due to differential levels of Min proficiency, as speakers from the Central dialect had approximately the same level of Min fluency, regardless of gender. In other words, what was underlying the difference between male and female speakers was more likely to be a pure gender issue. Female speakers were more sensitive to the differences between standard and nonstandard forms and were more likely to conform themselves to the social norm. As a consequence, they were less inclined to show obvious regional traits.

With regard to the declination pattern, dialectal differences were also found. For both genders, only sentence-medial and -final positions showed reliable differences. However, the direction of the effect was exactly opposite for the two genders. Male Taichung speakers had a milder declination topline than their Taipei counterparts, while female Taichung speakers had a steeper declination topline instead. The pattern of male speakers was in line with our predictions. Since tonal register was lower in the Taichung dialect, and since higher tonal ranges were more elastic than lower ones (Ladd, 1988), the declination range would naturally

be more restricted and thus the slope of the topline would be shallower. However, in female speakers, the Central dialect showed steeper slopes than the Northern dialect. If this reverse pattern could also be attributed to differential gender sensitivity to the linguistic norm, then female Taichung speakers were actually counteracting regional characteristics by over-correction. Interestingly, this was done not by raising the overall pitch range of the sentence, using the same mechanism employed by their Taipei counterparts, but was instead achieved by only raising the initial starting point, that of *zhe*^{4}, demonstrating a partial-raising of the tonal range.

九、 結論 (Conclusion)

This study showed that dialectal differences existed in the tonal range and declination pattern of Taiwan Mandarin, with the nonstandard dialect demonstrating a phonetically lower high tone, and the declination slope shallower than the standard variety. It was assumed that such a difference could be attributed to differential influences from the substrate language Min. Fulfilling sociolinguistic predictions, female speakers showed a lesser degree of such dialectal differences and were more likely to counteract regional characteristics by over-corrections. In order to further confirm this trend, one plans to extend the scope of the study to include the Southern dialect. If such dialectal differences were indeed due to influences from Min, then one should be able to see the same pattern in the Southern dialect, perhaps even more so, as Min is even more commonly used in the area.

十、 參考文獻 (References)

CHAO, Y. R. (1956). Tone, intonation, singsong, chanting, recitative, tonal composition, and atonal composition in Chinese. In M. Halle (Ed.), *For Roman Jakobson* (pp. 52-59). The Hague: Mouton.

CHAO, Y. R. (1968). *A Grammar of Spoken Chinese*. Berkeley: University of California Press.

CHEN, S. H. (2005). The effects of tones on speaking frequency and intensity ranges in Mandarin and Min dialects. *Journal of the Acoustical Society of America*, 117, 3225-3230.

CHEN, Y.-D. (1989). *Taiwan de Kejiaren [Hakka on Taiwan]*. Taipei: Taiyuan Press.

CHENG, R. L. (1985). A comparison of Taiwanese, Taiwan Mandarin, and Peking Mandarin. *Language*, 61, 352-377.

FON, J., & CHIANG, W.-Y. (1999). What does Chao have to say about tones? -A case study of Taiwan Mandarin. *Journal of Chinese Linguistics*, 27, 15-37.

FON, J., CHIANG, W.-Y., & CHEUNG, H. (2004). Production and perception of two dipping tones (T2 and T3) in Taiwan Mandarin. *Journal of Chinese Linguistics*, 32, 249-280.

FON, J., & HSU, H. (2007). Positional and phonotactic effects on the realization of dipping tones in Taiwan Mandarin. In C. Gussenhoven and T. Riad (Eds.), *Phonology and phonetics, Tones and tunes: Phonetic and behavioural studies in word and sentence prosody* (pp. 239-269). Berlin: Mouton de Gruyter.

HOWIE, J. M. (1976). *Acoustical Studies of Mandarin Vowels and Tones*. Cambridge: Cambridge University Press.

HUANG, S. (1993). *Yuyan, shehui yu zuqun yishi--Taiwan yuyan shehuixue de yanjiu [Language, society, and ethnic identity--Studies in language sociology in Taiwan]*. Taipei: Crane.

LADD, D. R. (1988). Declination 'reset' and the hierarchical organization of utterances. *The Journal of the Acoustical Society of America*, 84, 530-544.

SHIH, C.-L. (1988). Tone and intonation in Mandarin. *Working Papers of the Cornell Phonetics Laboratory*, 3, 83-109.

TRUDGILL, P. (2000). *Sociolinguistics: an introduction to language and society*. New York: Penguin Books.

TSENG, C.-Y. (1990). *An Acoustic Phonetic Study on Tones in Mandarin Chinese*. Taipei: Institute of History & Philology, Academia Sinica.

WANG, W., S-Y. (1967). Phonological features of tone. *International Journal of American Linguistics*, 33, 93-105.

XU, Y. (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica: International Journal of Speech Science*, 55, 179-203.

XU, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech Communication*, 46, 220-251.

十一、 與執行本計畫相關之著作

Fon, J.*, Hsu, H.-j., Huang, Y.-H., & Chen, S. (2007). The effect of onset and position in the realization of Tone 1 in two dialects of Taiwan Mandarin. *Proceedings of the 16th International Congress of Phonetic Sciences*, 1297–1300. [請見附錄]

Huang, Y.-H. & **Fon, J.*** (submitted). Dialectal variations in tonal register and declination

pattern of Taiwan Mandarin. *Proceedings of the 4th International Conference on Speech Prosody*, Cambinas, Brazil.

十二、 計畫成果自評

1. 研究內容與原計劃相符程度

原計劃內容乃為探討國人對於國語中一聲調闊下降的情形。研究內容與原計劃大致上相符。

2. 達成預期目標情況

原計劃預計完成之目標包含：閱讀相關典籍、購買儀器設備、訓練實驗執行助理、完成實驗刺激挑選、執行實驗、完成Praat程式寫作、擷取相關數據、圖表製作、統計分析、結果撰寫及發表。本研究計劃大致上達成預期目標。不過，由於受限於經費，故將原本擬進行之六個實驗，刪減為二個（實驗一與實驗四），為本計劃在現實考量之下妥協之處，所幸所得成果仍可大致展現原計劃所欲達成之目標。

3. 研究成果之學術或應用價值

就學術研究而言，此一計劃的研究成果，可以使我們更進一步地了解到台灣地區國語一聲的調闊變化，及調闊下降的成因與分布狀況。同時，我們也盼能因此拋磚引玉，吸引更多的學者對於國語特有的現象，進行研究，讓我們對於這個演變中的台灣語言，有更精確的掌握與認識。

在應用方面，如前所提，本計劃的研究成果對於目前正如火如荼進行研發的語音科技，亦有其一定的貢獻。語音科技若要讓所產生出來的語句更自然，判讀的語句正確率更高，建立非正式口吻之閱讀語料庫是不可或缺的一環。許多北方官話的商業化語音科技產品，皆以大陸普通話為模型，於建立北方官話的語音系統後，再調整一些特定參數，將系統轉換成為較適合國語的模型。由於本計劃的目的之一即是在建立國語一聲調闊下降的分布模型，因此所得研究成果將可用以改良目前的國語語音科技，使其更自然、辨識率更高。目前中國大陸因2008年奧運之故，已於數年前將語言訊號處理視為國家發展重點之一。我們若不急起直追，五年之後，將很難與其爭鋒。

4. 是否適合在學術期刊發表或申請專利

本計劃適合於語音學相關期刊發表，目前部份研究成果已陸續投稿於知名期刊。本人亦擬於未來二三年內，將其餘成果整理發表，並繼續從事相關研究，以便能夠有一系列更完整的研究成果呈現。

5. 主要發現或其他有關價值等

本研究主要發現有二：其一、就聲調而言，台灣地區國語已與大陸的普通話不

同，自成一個系統，有其特有的聲調調型。一聲的調闊隨著台語的影響程度而下降，並有地域與性別上的不同。其二、台灣地區國語歷經數十年來的演變，不僅已標準化，且已發展出各地方言特色。北部方言與中部方言呈現調闊不同的現象，對於四聲之間的區辨性，將有極大的影響，值得近一步探討。雖然北部方言由於是標準方言的關係，語言變動的情形較為緩慢，但是隨著中南部人口大量移入北部，標準方言亦可能因此有較劇烈的變化，是值得持續觀察的現象。

附錄

THE EFFECT OF ONSET AND POSITION IN THE REALIZATION OF TONE 1 IN TWO DIALECTS OF TAIWAN MANDARIN

Janice Fon,[1] Hui-ju Hsu,[2] Yi-Hsuan Huang,[1] & Sally Chen[1]

[1]Graduate Institute of Linguistics
National Taiwan University

[2]Dept. Appl. Linguistics & Language Studies
Chung Yuan Christian University

{jfon, r94142011, d93142002}@ntu.edu.tw,[1] hsuhj@cycu.edu.tw[2]

ABSTRACT

This study investigates how onset and sentence positioning affect the realization of Tone 1 in two dialects of Taiwan Mandarin. Results showed that the central dialect was higher in register when placed in isolation, but lower when placed in a sentential context. When there was a tonal mismatch, coarticulatory effects were more robust in the northern dialect, implying that speakers of the central dialect (nonstandard) might be more self-conscious about the standard-vernacular distinction than those of the northern dialect (standard), and overcorrection tended to occur. The effect of onset type was also significant but fairly localized. Obstruent-initial syllables had higher initial pitch than sonorant ones. The declination effect was also significant, the rate of which being higher in the central variety. In addition, sentential stress tended to raise the sentence-final H targets in both varieties. However, the PENTA model was not fully supported.

Keywords: high tone, Taipei Mandarin, Taichung Mandarin, dialectal variation, tonal realization

1. INTRODUCTION

Taiwan Mandarin (TM) is the official language of Taiwan and is genetically related to Mainland Mandarin (MM), the official language of Mainland China. However, due to almost 60 years of political separation between the two places, the two Mandarins have developed independently so that dialectal variations are obvious to speakers of either variety [6].

Political division is not the only cause for the divergence of the two dialects, however. Ethnic distributions are also different. About 73-80% of the population in Taiwan is Southern Min, who speaks a variant of Mainland Southern Min, and this Min is therefore a powerful substrate language for TM [5, 6].

Both Mandarin varieties have four tones, traditionally termed Tone 1 (T1), Tone 2 (T2), Tone 3 (T3), and Tone 4 (T4), which are realized as high level, mid dipping, mid-low dipping, and high falling, respectively, with T2 having an allophonic variant of mid rising in MM, and T3 a mid-low fall in both varieties [7, 16]. Although the phonological categories of the four tones are the same between the two dialects, the phonetic realizations are somewhat different. Specifically, tonal registers of TM T2 and T3 are much lower and narrower than those of the MM variety [8]. This discrepancy is presumably due to the influence from Min, which seems to prefer a lower frequency range [4, 11]. This study thus planned to see if such a lowering effect is also affecting T1, which is a high tone.

2. AIMS OF THE STUDY

There are four specific aims in this study. First of all, we would like to explore possible dialectal differences in T1 realization. If the degree of Min influence is negatively correlated with tonal register [8, 11], one would expect the tonal targets of TM varieties that are more influenced by Min (*i.e.*, the nonstandard varieties) to be lower than those that are not as influenced (*i.e.*, the standard variety).

Secondly, we would like to see if sentential T1 demonstrates a similar interaction with stress as the other tones. Fon & Hsu [8] showed that when T2 and T3 are placed in sentence-final positions, H targets are realized higher and L targets lower than what would be expected from pure declination. We suspected that the exaggeration in the realization of stress might be due to a sentence-final stress rule [2, 13]. Therefore, we would like to see if such a trend could also be observed in T1. If so, then sentence-final T1s should be realized higher than what would be predicted by the neutral topline.

The third aim is to investigate possible effects of syllable structure on T1 realization. According to Hombert, Ohala, and colleagues [9, 10, 12], voiceless obstruents impose a slight pitch-raising effect on the F_0 values. However, this was not found in the realization of T2 and T3 [8]. One possible reason might be the constraints imposed by contour tones. Therefore, we would like to see if level T1 is also impervious to such effects.

Finally, according to the PENTA model [17], the default tonal register in utterance-initial positions should be mid unless otherwise specified (p. 240). Our previous findings [8] could not find affirmative evidence for this claim for TM T2. Thus, we would like to see if such pattern could be observed in T1. The model would predict T1 to be always realized as a mid-to-high rise in isolation and in sentence-initial positions, and as a low-to-high rise in other sentence-internal mismatch positions, as both would be considered as tonal mismatch cases in the PENTA model.

3. METHODS

3.1. Participants

Six subjects between ages 19 and 24 participated in the study. Half were from Taipei (the northern dialect), and half were from Taichung (the central dialect). All of them were ethnically Min, but the Taipei speakers could not speak Min fluently. As this study is still in progress, more subjects will be included when the project is complete.

3.2. Stimuli

27 T1 syllables representative of Mandarin phonotactics were chosen as stimuli, including 6 voiceless obstruent-initials (e.g., [xan] 'charmingly naive'), 15 sonorant-initials (e.g., [la] 'pull'), and 6 vowel-initials (e.g., [u] 'black'). Syllables were also placed in three comparable carrier sentences, so that they occurred in sentence-initial, -medial, and -final positions. Carrier sentences were designed so that syllables immediately before the target in medial and final positions ended mid-low and comparable tonal target clashes would occur in all three positions. In total, 27 (stimuli) × 3 (positions) = 81 sentences were recorded.

3.3. Equipment

Recordings were done using a SONY PCM-M1 Digital Audio Recorder with Maxell R-64 DA 60 min DAT tapes and a SHURE SM10A head-mounted microphone.

3.4. Procedure

Speakers were seated in a quiet room and asked to read out loud the semi-randomized stimuli using natural intonation at a normal rate. The whole process took about 15 minutes. The original recordings had a sampling rate of 48 kHz, which were subsequently downsampled to 16 bit 22050 kHz using Cool Edit Pro 2.00.

3.5. Analyses

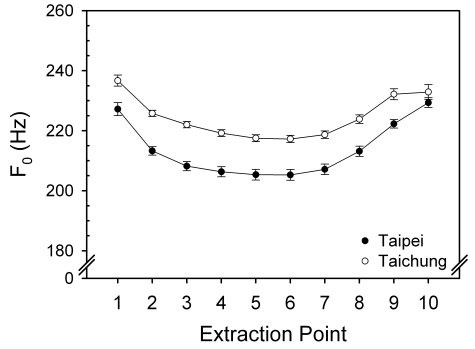
The recordings were hand-labeled using Praat 4.4 [1]. A Praat script was written for automatic pitch extraction on the voiced portion of each syllable, which is considered the measurable domain for tones [2, 3, 14, 15]. For obstruent-initial syllables, the starting point of a tone was determined by the onset of the voice bar after the obstruent, which was voiceless in this study. For the rest, the starting point began from the onset of the syllable, as the whole syllable was voiced. The ending point was always the offset of the voice bar. Occasional syllable-initial or -final glottalized portions caused by voice fry were not included for pitch extraction. Extracted pitch tracks were hand-checked and hand-corrected for doubling and halving through pitch period calculation, and were interpolated and smoothed using Praat functions afterwards. A second Praat script was written to extract pitch reference points at ten equal time points.

4. RESULTS

4.1. T1 in isolation

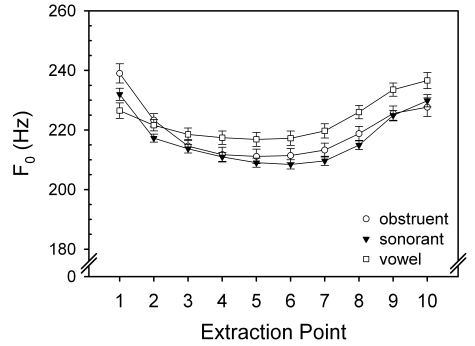
The average F_0 for Taipei speakers was 215.05 Hz while that for Taichung speakers was 225.10 Hz. A Dialect (2) × Onset (3) × Extraction (10) three-way mixed ANOVA was performed to test the effect of dialect and onset. Results showed that all of the main effects were significant [Dialect: $F(1,156) = 22.89$, $p < .0001$, $\hat{\eta}^2 = .13$; Onset: $F(2, 156) = 3.81$, $p < .05$, $\hat{\eta}^2 = .05$; Extraction: $F(2.76, 430.44) = 143.49$, $p < .0001$, $\hat{\eta}^2 = .48$]. Two of the two-way interaction effects involving Extraction were also significant [Dialect × Extraction: $F(2.76, 430.44) = 5.08$, $p < .01$, $\hat{\eta}^2 = .03$; Onset × Extraction: $F(5.52, 430.44) = 8.21$, $p < .0001$, $\hat{\eta}^2 = .10$]. The three-way interaction was not significant (Figures 1 & 2).

Figure 1: Time-normalized F_0 trajectories of isolated T1 in two dialects averaged across onset types.



Post hoc independent *t*-tests regarding the interaction effect of Dialect and Extraction showed that except for the final extraction point, all other extractions were significant ($p < .01$ for Point 1, and $p < .001$ for others). In addition, post hoc pairwise comparisons showed that for northern TM, Points 1, 9, and 10 were the highest in pitch, Points 2 and 8 were the second highest, the remaining points were the lowest ($p < .0001$). For central TM, Points 1, 9, and 10 were still the highest in pitch, Points 2, 3, and 8 were the next highest, and the remaining points were the lowest ($p < .0001$).

Figure 2: Time-normalized F_0 trajectories of T1 in isolation with regards to onset types.

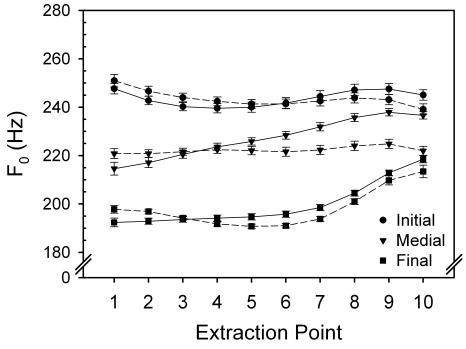


Post hoc one-way ANOVAs regarding the interaction effect of Onset and Extraction showed that Onset was only significantly different at Points 1 and 7 [Point 1: $F(2, 78) = 5.99$, $p < .01$, $\hat{\eta}^2 = .13$; Point 7: $F(2, 78) = 3.47$, $p < .05$. $\hat{\eta}^2 = .08$]. For Point 1, post hoc Tukey-*b* tests showed that obstruent-initial syllables were higher than vowel- and sonorant-initial ones ($p < .05$). For Point 7, post hoc pairwise comparisons showed that vowel-initial syllables were the highest while sonorant-initial ones were the lowest ($p < .05$).

4.2. T1 in context

The average F_0 values for Taipei speakers were 245.65 Hz, 230.20 Hz, and, 200.71 Hz for the three positions, respectively, while those for Taichung speakers were 244.26 Hz, 224.24 Hz, and 198.93 Hz, respectively. A Dialect (2) × Position (3) × Extraction (10) three-way mixed ANOVA was performed to test the effect of dialect and sentential positioning. Since Onset did not seem to have a very robust effect on isolated T1, it was excluded in the following analyses. Results showed that two of the main effects were significant [Position: $F(2, 320) = 730.78$, $p < .0001$, $\hat{\eta}^2 = .82$; Extraction: $F(1.88, 300.22) = 74.60$, $p < .0001$, $\hat{\eta}^2 = .32$]. Two of the two-way interactions involving Extraction were also significant [Dialect × Extraction: $F(1.88, 300.22) = 23.67$, $p < .0001$, $\hat{\eta}^2 = .13$; Position × Extraction: $F(3.36, 537.44) = 73.66$, $p < .0001$, $\hat{\eta}^2 = .32$]. So was the three-way interaction [$F(3.36, 537.44) = 5.80$, $p < .001$, $\hat{\eta}^2 = .03$] (Figure 3).

Figure 3: Time-normalized F_0 trajectories of T1 in context. Solid lines represent Taipei speakers, and dashed lines represent Taichung speakers.



Regarding the declination effect, post hoc pairwise comparisons showed that for northern TM, all sentence-initial extraction points were higher than sentence-medial ones, which were in turn higher than sentence-final ones. However, the difference between the latter two positions were much larger than the former two, especially in the final portions of the tone ($p < .001$ between initial and medial Point 9's, $p < .01$ between initial and medial Point 10's, and $p < .0001$ for others). For central TM, the overall trend was still the same. Sentence-initial extracts were the highest, and sentence-final ones were the lowest. However, the difference between the former two was much larger than that between the latter two for the final portion of the tone ($p < .05$ for medial and final Point 10's, and $p < .0001$ for others).

As for dialectal differences, post hoc pairwise

comparisons showed that for sentence-medial positions, Taipei T1 was significantly higher than that of Taichung starting from Point 6 to the end of the tone ($p < .01$ for Point 6, $p < .001$ for Point 7, and $p < .0001$ for others). For sentence-final positions, Taipei T1 was lower than Taichung T1 at Point 1 ($p < .05$), but was significantly higher from Points 5 to 8 ($p < .05$ for Points 5 & 8, and $p < .01$ for Points 6 & 7). No difference was found in the initial position.

With regards to tonal contours, Taipei and Taichung initial T1s and Taichung medial T1s were fairly level. Taipei medial and final T1s were rising, while Taichung final T1s were dipping.

5. DISCUSSION

Results in this study showed that variations did exist between the two varieties of TM. In terms of pitch register, the direction went as predicted in sentence-medial and -final positions. Taichung T1 was indeed lower, and the rate of declination faster. However, syllables in isolation showed an opposite trend, in which Taipei T1 was lower. Since reading isolated syllables is more unnatural and thus more formal than reading sentences, we hypothesized that the Taichung speakers, speaking a non-standard dialect, might be unconsciously over-correcting themselves in a more formal register, but were unable to do so in a more relaxed one.

The effect of sentential stress in raising sentence-final H tonal targets was also supported, as can be shown by the bigger rise towards the end of the syllable. The effect of sentential stress thus affects not only contour tones, but also level ones.

Different onset types did have an effect on the realization of tones. Obstruent-initial syllables had slightly higher pitch than sonorant-initial ones. However, the effect was fairly small and localized.

Finally, isolated and sentence-initial T1s were not realized as a rise, which contradicted Xu's [17] claim of a default mid tonal register. The only contours that conformed to the PENTA model were the medial and final Taipei T1s. However, Taichung tones did not show this effect.

6. CONCLUSION

This study shows that dialectal differences can affect realization of Mandarin T1. Phonotactic composition, while significant, imposes only minor effects. More studies will be needed in order to understand the actual mechanism underlying tonal realization.

7. REFERENCES

[1] Boersma, P., Weenink, D. 2006. *Praat: doing phonetics by computer* version 4.4.18. http://www.praat.org/.

[2] Chao, Y.R. 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.

[3] Chao, Y.R. 1956. Tone, intonation, singsong, chanting, recitative, tonal composition, and atonal composition in Chinese. In: Halle, M. (ed), *For Roman Jakobson*, The Hague: Mouton, 52-59.

[4] Chen, S.H. 2005. The effects of tones on speaking frequency and intensity ranges in Mandarin and Min dialects. *Journal of the Acoustical Society of America* 117(5), 3225-3230.

[5] Chen, Y.-D. 1989. *Taiwan de Kejiaren [Hakka on Taiwan]*. Taipei: Taiyuan Press.

[6] Cheng, R.L. 1985. A comparison of Taiwanese, Taiwan Mandarin, and Peking Mandarin. *Language* 61(2), 352-377.

[7] Fon, J., Chiang, W.-Y. 1999. What does Chao have to say about tones? -A case study of Taiwan Mandarin. *Journal of Chinese Linguistics* 27(1), 15-37.

[8] Fon, J., Hsu, H. 2007. Positional and phonotactic effects on the realization of dipping tones in Taiwan Mandarin. In: Gussenhoven, C. and Riad, T. (eds), *Phonology and phonetics, Tones and tunes: Phonetic and behavioural studies in word and sentence prosody*. Vol. 2, 239-269.

[9] Hombert, J.-M. 1978. Consonant types, vowel quality, and tone. In: Fromkin, V. A. (ed), *Tone: A Linguistic Survey*, New York: Academic Press, 77-111.

[10] Hombert, J.-M., Ohala, J.J., Ewan, W.G. 1979. Phonetic explanation for the development of tones. *Language* 55(1), 37-58.

[11] Hsu, S.-y. 2006. *The tonal variation of Mandarin Tone 2 in Taiwan: A phonetic study on Taiwanese-Taiwan Mandarin bilinguals and Taiwan Mandarin monolinguals*. National Chengchi University.

[12] Ohala, J.J. 1973. The physiology of tone. In: Hyman, L. M. (ed), *Consonant Types and Tone: Southern California Occasional Papers in Linguistics*. Vol. 1, Los Angeles: University of Southern California, 2-14.

[13] Peng, S.-h., Chan, M.K.M., Tseng, C.-y., Huang, T., Lee, O.J., Beckman, E.M. 2005. Towards a pan-Mandarin system for prosodic transcription. In: Jun, S.-A. (ed), *Prosodic Typology: The phonology of intonation and phrasing*, New York: Oxford University Press, 230-270.

[14] Wang, W., S-Y. 1967. Phonological features of tone. *International Journal of American Linguistics* 33(2), 93 - 105.

[15] Xu, Y. 1998. Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica: International Journal of Speech Science* 55(4), 179-203.

[16] Xu, Y. 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics* 25(1), 61-83.

[17] Xu, Y. 2005. Speech melody as articulatorily implemented communicative functions. *Speech Communication* 46(3-4), 220-251.

出席國際學術會議心得報告

計畫編號	95-2411-H-002-046-
計畫名稱	地域及語體對台灣地區國語一聲調闊之影響
出國人員姓名 服務機關及職稱	馮怡蓁
會議時間地點	96/8/6-10 德國薩爾布魯根市
會議名稱	第十六屆國際語音科學會議
發表論文題目	(1) The effect of onset and position in the realization of Tone 1 in two dialects of Taiwan Mandarin (2) The effect of acquisition order and word relatedness on codeswitching costs in balanced bilingual speakers (3) The effect of incredulity and particle on the intonation of yes/no questions in Taiwan Mandarin (4) The effects of phonetic distance, learning context and learner proficiency on L2 perception of English liquids

一、參加會議經過

本次會議共為期五天，自8/6至8/10。本次的語音相關主題包括言談對話、口語韻律、語音與非語言訊號的鏈結、發聲語音學、聽辨語音學、生理及病理語音學、語音及多語言訊號處理、語音傳播、語言／方言辨識、語音科技之應用與其評估等等。內容相當豐富而多元，亦有許多資深語音學者如Ann Cutler、John Local、Chih-Lin Shih等與會，對於年輕後輩有相當多的啟發與鼓勵。

二、與會心得

國際語音科學會議一向以鼓勵學者整合語音、心理及電腦科技等相關領域，以創新而嚴謹的研究方法研究韻律相關課題。此次參加會議，本人與其他與會者有相當多的接觸與討論，著實獲益良多。本人此次所發表的四篇論文，亦得到許多知名相關學者中肯的建議，為日後研究的課題，提供許多寶貴的方向。

THE EFFECT OF ONSET AND POSITION IN THE REALIZATION OF TONE 1 IN TWO DIALECTS OF TAIWAN MANDARIN

Janice Fon,[1] Hui-ju Hsu,[2] Yi-Hsuan Huang,[1] & Sally Chen[1]

[1]Graduate Institute of Linguistics
National Taiwan University

[2]Dept. Appl. Linguistics & Language Studies
Chung Yuan Christian University

{jfon, r94142011, d93142002}@ntu.edu.tw,[1] hsuhj@cycu.edu.tw[2]

ABSTRACT

This study investigates how onset and sentence positioning affect the realization of Tone 1 in two dialects of Taiwan Mandarin. Results showed that the central dialect was higher in register when placed in isolation, but lower when placed in a sentential context. When there was a tonal mismatch, coarticulatory effects were more robust in the northern dialect, implying that speakers of the central dialect (nonstandard) might be more self-conscious about the standard-vernacular distinction than those of the northern dialect (standard), and overcorrection tended to occur. The effect of onset type was also significant but fairly localized. Obstruent-initial syllables had higher initial pitch than sonorant ones. The declination effect was also significant, the rate of which being higher in the central variety. In addition, sentential stress tended to raise the sentence-final H targets in both varieties. However, the PENTA model was not fully supported.

Keywords: high tone, Taipei Mandarin, Taichung Mandarin, dialectal variation, tonal realization

1. INTRODUCTION

Taiwan Mandarin (TM) is the official language of Taiwan and is genetically related to Mainland Mandarin (MM), the official language of Mainland China. However, due to almost 60 years of political separation between the two places, the two Mandarins have developed independently so that dialectal variations are obvious to speakers of either variety [6].

Political division is not the only cause for the divergence of the two dialects, however. Ethnic distributions are also different. About 73-80% of the population in Taiwan is Southern Min, who speaks a variant of Mainland Southern Min, and this Min is therefore a powerful substrate language for TM [5, 6].

Both Mandarin varieties have four tones, traditionally termed Tone 1 (T1), Tone 2 (T2), Tone 3 (T3), and Tone 4 (T4), which are realized as high level, mid dipping, mid-low dipping, and high falling, respectively, with T2 having an allophonic variant of mid rising in MM, and T3 a mid-low fall in both varieties [7, 16]. Although the phonological categories of the four tones are the same between the two dialects, the phonetic realizations are somewhat different. Specifically, tonal registers of TM T2 and T3 are much lower and narrower than those of the MM variety [8]. This discrepancy is presumably due to the influence from Min, which seems to prefer a lower frequency range [4, 11]. This study thus planned to see if such a lowering effect is also affecting T1, which is a high tone.

2. AIMS OF THE STUDY

There are four specific aims in this study. First of all, we would like to explore possible dialectal differences in T1 realization. If the degree of Min influence is negatively correlated with tonal register [8, 11], one would expect the tonal targets of TM varieties that are more influenced by Min (*i.e.*, the nonstandard varieties) to be lower than those that are not as influenced (*i.e.*, the standard variety).

Secondly, we would like to see if sentential T1 demonstrates a similar interaction with stress as the other tones. Fon & Hsu [8] showed that when T2 and T3 are placed in sentence-final positions, H targets are realized higher and L targets lower than what would be expected from pure declination. We suspected that the exaggeration in the realization of stress might be due to a sentence-final stress rule [2, 13]. Therefore, we would like to see if such a trend could also be observed in T1. If so, then sentence-final T1s should be realized higher than what would be predicted by the neutral topline.

The third aim is to investigate possible effects of syllable structure on T1 realization. According to Hombert, Ohala, and colleagues [9, 10, 12], voiceless obstruents impose a slight pitch-raising effect on the F_0 values. However, this was not found in the realization of T2 and T3 [8]. One possible reason might be the constraints imposed by contour tones. Therefore, we would like to see if level T1 is also impervious to such effects.

Finally, according to the PENTA model [17], the default tonal register in utterance-initial positions should be mid unless otherwise specified (p. 240). Our previous findings [8] could not find affirmative evidence for this claim for TM T2. Thus, we would like to see if such pattern could be observed in T1. The model would predict T1 to be always realized as a mid-to-high rise in isolation and in sentence-initial positions, and as a low-to-high rise in other sentence-internal mismatch positions, as both would be considered as tonal mismatch cases in the PENTA model.

3. METHODS

3.1. Participants

Six subjects between ages 19 and 24 participated in the study. Half were from Taipei (the northern dialect), and half were from Taichung (the central dialect). All of them were ethnically Min, but the Taipei speakers could not speak Min fluently. As this study is still in progress, more subjects will be included when the project is complete.

3.2. Stimuli

27 T1 syllables representative of Mandarin phonotactics were chosen as stimuli, including 6 voiceless obstruent-initials (e.g., [xan] 'charmingly naive'), 15 sonorant-initials (e.g., [la] 'pull'), and 6 vowel-initials (e.g., [u] 'black'). Syllables were also placed in three comparable carrier sentences, so that they occurred in sentence-initial, -medial, and -final positions. Carrier sentences were designed so that syllables immediately before the target in medial and final positions ended mid-low and comparable tonal target clashes would occur in all three positions. In total, 27 (stimuli) × 3 (positions) = 81 sentences were recorded.

3.3. Equipment

Recordings were done using a SONY PCM-M1 Digital Audio Recorder with Maxell R-64 DA 60 min DAT tapes and a SHURE SM10A head-mounted microphone.

3.4. Procedure

Speakers were seated in a quiet room and asked to read out loud the semi-randomized stimuli using natural intonation at a normal rate. The whole process took about 15 minutes. The original recordings had a sampling rate of 48 kHz, which were subsequently downsampled to 16 bit 22050 kHz using Cool Edit Pro 2.00.

3.5. Analyses

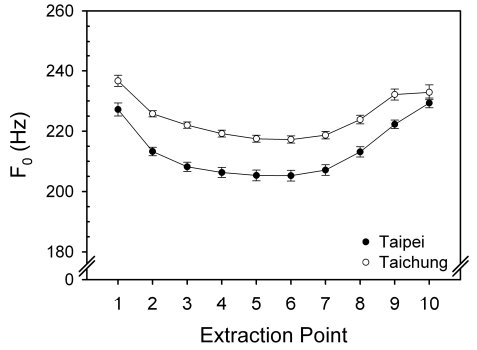
The recordings were hand-labeled using Praat 4.4 [1]. A Praat script was written for automatic pitch extraction on the voiced portion of each syllable, which is considered the measurable domain for tones [2, 3, 14, 15]. For obstruent-initial syllables, the starting point of a tone was determined by the onset of the voice bar after the obstruent, which was voiceless in this study. For the rest, the starting point began from the onset of the syllable, as the whole syllable was voiced. The ending point was always the offset of the voice bar. Occasional syllable-initial or -final glottalized portions caused by voice fry were not included for pitch extraction. Extracted pitch tracks were hand-checked and hand-corrected for doubling and halving through pitch period calculation, and were interpolated and smoothed using Praat functions afterwards. A second Praat script was written to extract pitch reference points at ten equal time points.

4. RESULTS

4.1. T1 in isolation

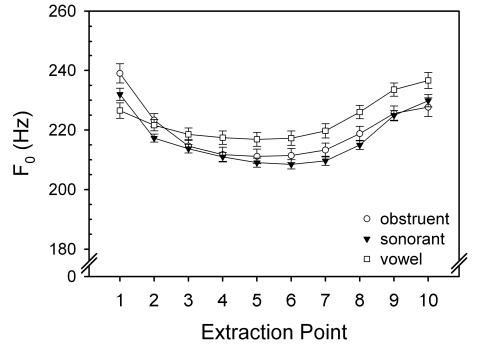
The average F_0 for Taipei speakers was 215.05 Hz while that for Taichung speakers was 225.10 Hz. A Dialect (2) × Onset (3) × Extraction (10) three-way mixed ANOVA was performed to test the effect of dialect and onset. Results showed that all of the main effects were significant [Dialect: $F(1,156) = 22.89$, $p < .0001$, $\hat{\eta}^2 = .13$; Onset: $F(2, 156) = 3.81$, $p < .05$, $\hat{\eta}^2 = .05$; Extraction: $F(2.76, 430.44) = 143.49$, $p < .0001$, $\hat{\eta}^2 = .48$]. Two of the two-way interaction effects involving Extraction were also significant [Dialect × Extraction: $F(2.76, 430.44) = 5.08$, $p < .01$, $\hat{\eta}^2 = .03$; Onset × Extraction: $F(5.52, 430.44) = 8.21$, $p < .0001$, $\hat{\eta}^2 = .10$]. The three-way interaction was not significant (Figures 1 & 2).

Figure 1: Time-normalized F_0 trajectories of isolated T1 in two dialects averaged across onset types.



Post hoc independent *t*-tests regarding the interaction effect of Dialect and Extraction showed that except for the final extraction point, all other extractions were significant ($p < .01$ for Point 1, and $p < .001$ for others). In addition, post hoc pairwise comparisons showed that for northern TM, Points 1, 9, and 10 were the highest in pitch, Points 2 and 8 were the second highest, the remaining points were the lowest ($p < .0001$). For central TM, Points 1, 9, and 10 were still the highest in pitch, Points 2, 3, and 8 were the next highest, and the remaining points were the lowest ($p < .0001$).

Figure 2: Time-normalized F_0 trajectories of T1 in isolation with regards to onset types.

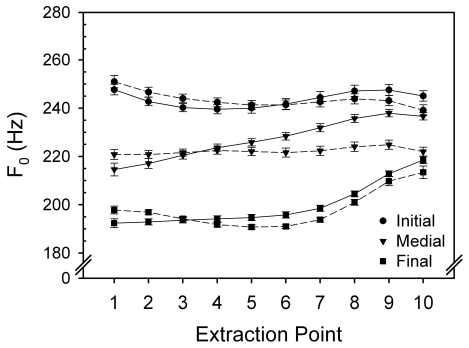


Post hoc one-way ANOVAs regarding the interaction effect of Onset and Extraction showed that Onset was only significantly different at Points 1 and 7 [Point 1: $F(2, 78) = 5.99$, $p < .01$, $\hat{\eta}^2 = .13$; Point 7: $F(2, 78) = 3.47$, $p < .05$. $\hat{\eta}^2 = .08$]. For Point 1, post hoc Tukey-*b* tests showed that obstruent-initial syllables were higher than vowel- and sonorant-initial ones ($p < .05$). For Point 7, post hoc pairwise comparisons showed that vowel-initial syllables were the highest while sonorant-initial ones were the lowest ($p < .05$).

4.2. T1 in context

The average F_0 values for Taipei speakers were 245.65 Hz, 230.20 Hz, and, 200.71 Hz for the three positions, respectively, while those for Taichung speakers were 244.26 Hz, 224.24 Hz, and 198.93 Hz, respectively. A Dialect (2) × Position (3) × Extraction (10) three-way mixed ANOVA was performed to test the effect of dialect and sentential positioning. Since Onset did not seem to have a very robust effect on isolated T1, it was excluded in the following analyses. Results showed that two of the main effects were significant [Position: $F(2, 320) = 730.78$, $p < .0001$, $\hat{\eta}^2 = .82$; Extraction: $F(1.88, 300.22) = 74.60$, $p < .0001$, $\hat{\eta}^2 = .32$]. Two of the two-way interactions involving Extraction were also significant [Dialect × Extraction: $F(1.88, 300.22) = 23.67$, $p < .0001$, $\hat{\eta}^2 = .13$; Position × Extraction: $F(3.36, 537.44) = 73.66$, $p < .0001$, $\hat{\eta}^2 = .32$]. So was the three-way interaction [$F(3.36, 537.44) = 5.80$, $p < .001$, $\hat{\eta}^2 = .03$] (Figure 3).

Figure 3: Time-normalized F_0 trajectories of T1 in context. Solid lines represent Taipei speakers, and dashed lines represent Taichung speakers.



Regarding the declination effect, post hoc pairwise comparisons showed that for northern TM, all sentence-initial extraction points were higher than sentence-medial ones, which were in turn higher than sentence-final ones. However, the difference between the latter two positions were much larger than the former two, especially in the final portions of the tone ($p < .001$ between initial and medial Point 9's, $p < .01$ between initial and medial Point 10's, and $p < .0001$ for others). For central TM, the overall trend was still the same. Sentence-initial extracts were the highest, and sentence-final ones were the lowest. However, the difference between the former two was much larger than that between the latter two for the final portion of the tone ($p < .05$ for medial and final Point 10's, and $p < .0001$ for others).

As for dialectal differences, post hoc pairwise

comparisons showed that for sentence-medial positions, Taipei T1 was significantly higher than that of Taichung starting from Point 6 to the end of the tone ($p < .01$ for Point 6, $p < .001$ for Point 7, and $p < .0001$ for others). For sentence-final positions, Taipei T1 was lower than Taichung T1 at Point 1 ($p < .05$), but was significantly higher from Points 5 to 8 ($p < .05$ for Points 5 & 8, and $p < .01$ for Points 6 & 7). No difference was found in the initial position.

With regards to tonal contours, Taipei and Taichung initial T1s and Taichung medial T1s were fairly level. Taipei medial and final T1s were rising, while Taichung final T1s were dipping.

5. DISCUSSION

Results in this study showed that variations did exist between the two varieties of TM. In terms of pitch register, the direction went as predicted in sentence-medial and -final positions. Taichung T1 was indeed lower, and the rate of declination faster. However, syllables in isolation showed an opposite trend, in which Taipei T1 was lower. Since reading isolated syllables is more unnatural and thus more formal than reading sentences, we hypothesized that the Taichung speakers, speaking a non-standard dialect, might be unconsciously over-correcting themselves in a more formal register, but were unable to do so in a more relaxed one.

The effect of sentential stress in raising sentence-final H tonal targets was also supported, as can be shown by the bigger rise towards the end of the syllable. The effect of sentential stress thus affects not only contour tones, but also level ones.

Different onset types did have an effect on the realization of tones. Obstruent-initial syllables had slightly higher pitch than sonorant-initial ones. However, the effect was fairly small and localized.

Finally, isolated and sentence-initial T1s were not realized as a rise, which contradicted Xu's [17] claim of a default mid tonal register. The only contours that conformed to the PENTA model were the medial and final Taipei T1s. However, Taichung tones did not show this effect.

6. CONCLUSION

This study shows that dialectal differences can affect realization of Mandarin T1. Phonotactic composition, while significant, imposes only minor effects. More studies will be needed in order to understand the actual mechanism underlying tonal realization.

7. REFERENCES

[1] Boersma, P., Weenink, D. 2006. *Praat: doing phonetics by computer* version 4.4.18. http://www.praat.org/.

[2] Chao, Y.R. 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.

[3] Chao, Y.R. 1956. Tone, intonation, singsong, chanting, recitative, tonal composition, and atonal composition in Chinese. In: Halle, M. (ed), *For Roman Jakobson*, The Hague: Mouton, 52-59.

[4] Chen, S.H. 2005. The effects of tones on speaking frequency and intensity ranges in Mandarin and Min dialects. *Journal of the Acoustical Society of America* 117(5), 3225-3230.

[5] Chen, Y.-D. 1989. *Taiwan de Kejiaren [Hakka on Taiwan]*. Taipei: Taiyuan Press.

[6] Cheng, R.L. 1985. A comparison of Taiwanese, Taiwan Mandarin, and Peking Mandarin. *Language* 61(2), 352-377.

[7] Fon, J., Chiang, W.-Y. 1999. What does Chao have to say about tones? -A case study of Taiwan Mandarin. *Journal of Chinese Linguistics* 27(1), 15-37.

[8] Fon, J., Hsu, H. 2007. Positional and phonotactic effects on the realization of dipping tones in Taiwan Mandarin. In: Gussenhoven, C. and Riad, T. (eds), *Phonology and phonetics, Tones and tunes: Phonetic and behavioural studies in word and sentence prosody*. Vol. 2, 239-269.

[9] Hombert, J.-M. 1978. Consonant types, vowel quality, and tone. In: Fromkin, V. A. (ed), *Tone: A Linguistic Survey*, New York: Academic Press, 77-111.

[10] Hombert, J.-M., Ohala, J.J., Ewan, W.G. 1979. Phonetic explanation for the development of tones. *Language* 55(1), 37-58.

[11] Hsu, S.-y. 2006. *The tonal variation of Mandarin Tone 2 in Taiwan: A phonetic study on Taiwanese-Taiwan Mandarin bilinguals and Taiwan Mandarin monolinguals*. National Chengchi University.

[12] Ohala, J.J. 1973. The physiology of tone. In: Hyman, L. M. (ed), *Consonant Types and Tone: Southern California Occasional Papers in Linguistics*. Vol. 1, Los Angeles: University of Southern California, 2-14.

[13] Peng, S.-h., Chan, M.K.M., Tseng, C.-y., Huang, T., Lee, O.J., Beckman, E.M. 2005. Towards a pan-Mandarin system for prosodic transcription. In: Jun, S.-A. (ed), *Prosodic Typology: The phonology of intonation and phrasing*, New York: Oxford University Press, 230-270.

[14] Wang, W., S-Y. 1967. Phonological features of tone. *International Journal of American Linguistics* 33(2), 93 - 105.

[15] Xu, Y. 1998. Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica: International Journal of Speech Science* 55(4), 179-203.

[16] Xu, Y. 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics* 25(1), 61-83.

[17] Xu, Y. 2005. Speech melody as articulatorily implemented communicative functions. *Speech Communication* 46(3-4), 220-251.

THE EFFECT OF ACQUISITION ORDER AND WORD RELATEDNESS ON CODE-SWITCHING COSTS IN BALANCED BILNGUAL SPEAKERS

Yi-Hsuan Huang and Janice Fon

Graduate Institute of Linguistics, National Taiwan University
{r94142011, jfon@ntu.edu.tw}

ABSTRACT

This paper aims to explore the effect of acquisition order and word-relatedness on code-switching costs in bilingual speakers. 38 Mandarin-Min bilinguals performed a picture-naming task, in which hand-drawn pictures were color-coded for the two languages, Mandarin and Min, and switching points were pre-determined but variable. Results showed that naming latencies of cognates were in general shorter than non-cognates, and Mandarin stimuli were also shorter than Min. Min non-cognates were especially difficult for subjects. Code-switched trials incurred longer latencies in subjects, but only in those who acquired both languages at the same time, contrary to what was predicted by the Inhibitory Control Model.

Keywords: code-switching, bilingual lexicon, Inhibitory Control Model, Mandarin-Southern Min

1. INTRODUCTION

Diglossia is a common phenomenon that exists among the bilinguals in Taiwan. Li and Lee [6] mentioned that compared with Hong Kong, individual bilingualism in Taiwan is more widespread largely as a result of the National Language Movement that promoted Mandarin as a High language, while Southern Min, Hakka, and the Aboriginal languages remained Low languages. Southern Min is the most widespread dialect and the second most widely used language variety in Taiwan. Code-switching between Mandarin and Southern Min in daily life is very common, with the matrix language being either language depending on the speaker, the context, and the genre [5, 7].

From a psycholinguistic perspective, Green [3] proposed the Inhibitory Control Model (ICM) to account for problems that bilingual speakers encounter when selectively attending to only one of two languages. Time costs incur when speakers switch between two languages as a result of the change of task schema. Furthermore, processing costs in code-switched materials differ according to various degrees of activation and suppression in L1 and L2 respectively. When one speaks L1, L1 would be activated and L2 inhibited. Since L2 is presumably a weaker language, it is less activated and is thus easier to be suppressed. Therefore, when one is switching from L1 to L2, the suppression of L2 in the first language schema when speaking L1 does not incur a large carryover effect on the next language schema (L2). However, when L2 is switched to L1, in which more suppression of L1 is required, greater switch costs would occur. Using a picture-naming paradigm, Meuter and Allport [9] found that the response latencies of language switching trials were longer than nonswitching trials in unbalanced bilinguals. More importantly, switching to a more dominant L1 took more time than switching to a weaker L2.

2. AIMS OF THE STUDY

There are three specific aims in this study. The first is to examine whether the imbalanced switching cost could also be shown in Mandarin-Min bilinguals. Based on Meuter and Allport [9] and the ICM [3], it would be reasonable to predict that switching from Min to Mandarin is slower than switching from Mandarin to Min since Mandarin is a relatively more dominant language. However, the special linguistic ecology in Taiwan tends to create bilinguals that are more balanced than the ones in Meuter and Allport's study [9]. It is thus unclear whether similar results could be replicated on balanced bilinguals of Mandarin-Min.

Secondly, we would also like to see whether the order of acquisition would affect code-switching costs. Due to the diglossic situation in Taiwan,

most Mandarin-Min bilinguals acquired Min earlier than Mandarin, the former being acquired at home while the latter at school. However, as Mandarin is the official language, some parents tend to use both languages at home so as to give children a "head start". As a consequence, there are also bilinguals that acquire the two languages simultaneously. Therefore, the second purpose of this study is to tease apart the effects caused by order of acquisition and those by language dominancy. If the status of being an L1 is special, even for nondominant languages, then one should find lower costs for code-switching in bilinguals that acquired both languages simultaneously. Finally, we would also like to examine the effect of word-relatedness. Meuter and Allport [9] used numbers in English and other European languages as their stimuli, which are composed of cognates sharing phonological similarities. However, it is unclear whether the same pattern could be replicated in non-cognate words. Hence, the third purpose is to explore the difference in code-switching between cognate and noncognate words. If the incurred switching cost is due mainly to words that are phonologically and morphologically similar, then one should find a lower switching cost for cognate than noncognate words.

3. METHODS

3.1. Participants

38 Mandarin-Min bilinguals (mean age = 23.3) took part in the study. Half acquired both languages simultaneously (the "Simultaneously-Acquired Group") and half acquired Min prior to Mandarin (the "Min-First Group"). No speaker in the two groups was exposed to Catonese or other dialects at their earlier age. Participants self-evaluated their language proficiency using a Likert scale ranging from 1 (very disfluent) to 7 (very fluent) .The average rating for Mandarin was 6.64 while that for Min was 5.38. The relative proficiency of participants in the two languages in both groups was similar. Though participants viewed themselves as highly-proficient bilinguals, for all participants, the proficiency level of Mandarin was respectively greater than that of Min. Participants spoke Mandarin more often than Min, which was reflected by language usage frequency survey filled up by participants. We are aware of that the level of proficiency will be more reliable if separate language proficiency raters are provided for two groups. However, since the stimuli in the experiment are names of pictures, it is difficult to determine the language proficiency objectively via single words uttered by participants. To sum up, as for the dominance of two languages, based on participants' level of relative proficiency and the frequency of language use, both groups of speakers spoke Mandarin as their more dominant language. Although there are inter-subject differences for each group, the general trend of the language dominancy is similar. Data from two speakers in the Min-First Group was deleted. One was due to an unpredictable technical problem. The other was because that a subject registered in the Min-first group but his language background mismatched the requirement of "acquiring Min prior to Mandarin" of the group. Data from one speaker in the Simultaneously-acquired Group was deleted because of self-report stutter in the language background questionnaire. Finally, there were 18 participants in the Siumultaneously-acquired Group and 17 in the Min-first Group.

3.2. Stimuli

40 target stimuli of everyday objects were chosen and were presented in cartoons. Half of the stimuli formed cognates in Mandarin and Min (e.g., [Mandarin] *jing*^{3}*-cha*^{2} *vs.* [Min] *keng*^{3}*-chhat*^{4} 'police') and the other half were noncognates (e.g., [Mandarin] *sha*^{1}*-fa*^{1} *VS.* [Min] *phong*^{3}*-i*^{2} 'sofa'). An additional 40 stimuli were also chosen to serve as primes. All of the line drawings were color-coded in light blue and white, the former representing Min, and the latter representing Mandarin. All drawings were pre-tested using a naming task to make sure they are representative enough of the stimuli in both languages. In total, there were 40 (targets) × 2 (colors) + 40 (stimuli) × 2 (colors) = 160 stimuli.

3.3. Equipment

E-prime 1.1 and its accompanying PST button box Model #200a were used to collect the reaction time data. A SONY MRD 7520 head-mounted microphone and a BurnIt CDR830 CD-ROM Burner were used for recording.

3.4. Procedure

Participants were seated in a quiet room before a computer monitor, and were asked to name the presented line drawings in the designated language. Instructions were provided by Mandarin. Each trial consisted of a prime and a target. In total, there were 40 (pairs) × 4 (color combinations: WW, WB, BW, BB) = 160 trials. WW and BB were the "non-switching" trials, in which the prime and the target were coded in the same color (Figure 1), while WB and BW were the "switching trials", in which the prime and the target were coded in different colors (Figure 2). The 160 trials were further divided into four equal groups, which contained 10 trials of each color combination. Participants were randomly assigned into four groups so that each person only saw each prime-target pair once. The order of presentation was randomized by E-prime for each participant.

Figure 1: An example of the structure of a non code-switching list (Prime: white, Mandarin; Target: white, Southern Min). Both fixation periods were 2 sec long and both the prime and the target were 4 sec long. The response window was also 4 sec long.

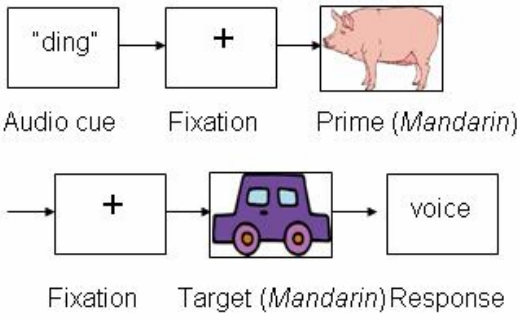
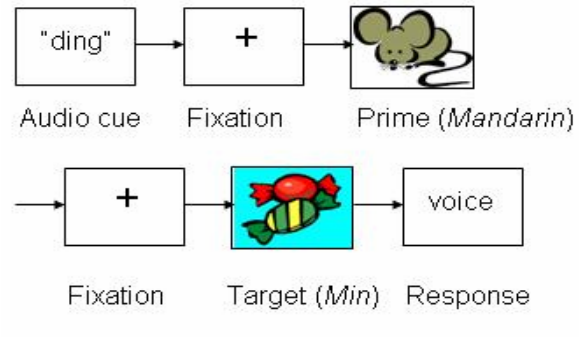


Figure 2: An example of the structure of a code-switching list (Prime: white, Mandarin; Target: blue, Southern Min). Fixation time and response window were the same as Figure 1.



3.5. Measurement

Reaction time was measured from the onset of stimuli by E-prime. Data collected by voice key was double-checked and hand-corrected based on the additionally recorded soundfiles using Praat. Responses were counted as correct only when the pictures were named correctly in the target language. The average correct rate was 92.23%. Latencies that were more than three standard deviations away from the mean were planned to be excluded. However, no data was deleted according to this criterion.

4. RESULTS

A Switch (2) × Language (2) × Cognate (2) × Order (2) four-way ANOVA was performed. Results showed that three of the main effects were significant [Language: $F(1,1277)= 75.34$, $p < 0.0001$, $\hat{\eta}^2 = .06$; Cognate: $F(1,1277)= 76.99$, $p < 0.0001$, $\hat{\eta}^2 = .06$; Order: $F(1,1277)= 4.50$, $p < 0.05$, $\hat{\eta}^2 = .004$]. The effect of Switch was also near significant [$F(1,1277)= 2.91$, $p = 0.09$, $\hat{\eta}^2 = .002$]. Two of the two-way interactions were also significant [Language × Cognate: $F(1,1277)= 15.15$, $p < 0.001$, $\hat{\eta}^2 = .01$; Switch × Order: $F(1,1277)= 5.47$, $p < 0.05$, $\hat{\eta}^2 = .004$]. No other higher-level interactions were found.

Post hoc independent *t* tests regarding the Language × Cognate interaction showed that cognates were in general faster than non-cognates [Mandarin: $t(644.66) = -4.45$, $p < .0001$; Min: $t(453.73) = -7.19$, $p < .0001$], and Mandarin stimuli were always faster than Min [Cognate: $t(598.45) = -4.22$, $p < .0001$; Non-cognates: $t(452.98) = -7.10$, $p < .0001$]. However, as shown in Figure 3, Min non-cognates were much slower than other types of stimuli.

As for the interaction of Switch × Order, post hoc independent *t* tests showed that RT for code-switched items were longer for subjects who acquired both languages simultaneously, as shown in Figure 4 [$t(633.24) = 3.46$, $p < .001$]. For non-switched items, there was no difference in acquisition order. Comparing across items, it was also interesting to find that code-switching incurred a cost only in subjects who acquired both languages at the same time [$t(660) = 2.91$, $p < .01$]. No difference was found in the Min-first subjects.

Figure 3: Overall RTs (in ms) for Mandarin and Min stimuli for cognates and non-cognates.

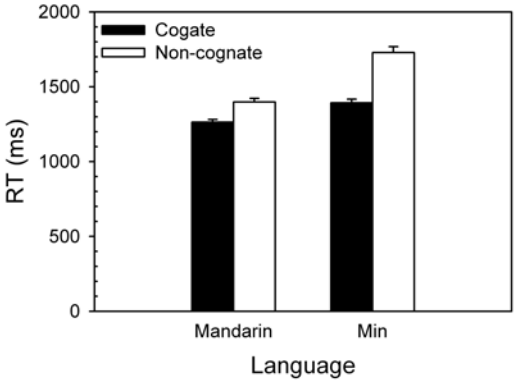
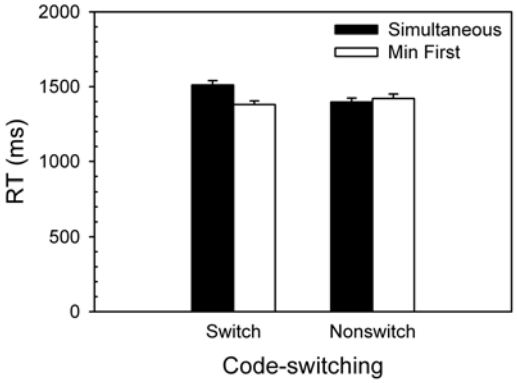


Figure 4: Overall RTs (in ms) regarding switching pattern and acquisition order.



5. DISCUSSION

The Inhibitory Control Model [3] predicts that time costs are incurred when bilinguals code-switch. However, this was only supported in our study for subjects who acquired both languages simultaneously. Code-switching time costs were not found for speakers acquiring Min prior to Mandarin. This interaction was interesting, as the two groups of subjects did not differ in their self-rated language proficiency. In other words, the order of acquisition might also affect the efficiency of lexical encoding, and thus code-switching latency. More studies would be needed in order to understand this effect.

The unequal switching costs found in Meuter and Allport [9] was not replicated in this study. Mandarin stimuli were in general responded faster than Min ones. This may have to do with the special language ecology in Taiwan. As Min is a language dominant in private domains and Mandarin a dominant language in public domains, the distinction of L1 and L2 in Taiwan might not have the same implication as that in Europe (e.g., English and French [8]). In other words, a later-acquired language is not necessarily a weaker language in Taiwan. Another possible underlying cause for the discrepancy might lie in the design. Meuter and Allport [9] used a within-subjects design while we used a between-subject one. Further studies will be needed in order to gain a clearer understanding of this matter.

The effect on word-relatedness is interesting. Cognates in general better facilitated code-switching than non-cognates. One possible reason is that the activation of phonological information affects the process of lexical selection [1], which causes the relationship between lexical items and phonological forms become bidirectional and interactive. The cognate status of word not only activates lexical representations in one language, but also spreads some activation to its corresponding segments, making the retrieval of the lexical item much easier. Non-cognates that were coded in Min were especially in disadvantage.

6. REFERENCES

[1] Costa, Albert., M. Santesteban & A. Cano. 2005. On the facilitatory effects of cognate words in bilingual speech production. *Brain and Language*, 94, 94-103.

[2] Dussias, Paola. E. 2001. Psycholinguistic complexity in codeswitching. *The International Journal of Bilingualism* 5 (1), 87-100.

[3] Green, D. W.1989. Control, activation, and resourse. *Brain and Language* 27, 210–223.

[4] Green, W.D. 1993. Toward a model of L2 comprehension and production . In: Schreuder, R., Weltens, B. (eds), *The Bilingual Lexicon.* Oxford: John Benjamins Publishing Co., 249-278.

[5] Huang, S. 2000. Language, identity and conflict: A Taiwanese study. *International Journal of the Sociology of Language*, 143, 139-149

[6] Li, D. C. S., Lee, S. 2006. *Bilingualism in East Asia*. In Bhatia, T.K., William, C. R. (eds), *The Handbook of Bilingualism*. London: Blackwell, 462-506

[7] Liao, C. C., 2000. Changing dominant language use and ethic equality in Taiwan since 1987. *International Journal of the Sociology of Language*, 143, 165-182.

[8] Meuter, R.F.I., 1994. *Language switching in naming tasks.* Unpublished Ph.D. thesis. University of Oxford, U.K.

[9] Meuter, R.F.I., Allport, A. 1999. Bilingual language switching in naming: Asymmetrical costs in language selection. *Journal of Memory and Language* 40, 25-40.

[10] Myers-Scotton, C., Jake, Janice.L. 1995. Matching lemmas in a bilingual language competence and production model: evidence from intrasentential code switching. *Linguistics* 33, 981-1024.

THE EFFECT OF INCREDULITY AND PARTICLE ON THE INTONATION OF YES/NO QUESTIONS IN TAIWAN MANDARIN

Yu-Ying Chuang, Yi-Hsuan Huang, & Janice Fon

Graduate Institute of Linguistics, National Taiwan University
{r95142001, r94142011, jfon}@ntu.edu.tw

ABSTRACT

This study explored the effect of incredulity and particle on the intonation of yes/no questions in Taiwan Mandarin. Two types of questions were examined – ones with and without the question particle *ma*. Results showed that to convey incredulity, the overall pitch would be raised and enlarged. Moreover, questions without particles were significantly higher in pitch and larger in pitch range than questions with particles. This thus led to a conclusion that the degree of incredulity being expressed in questions with *ma* might not be as great as that in questions without *ma*.

Keywords: incredulity, yes/no question intonation, question particle *ma*, pitch height, pitch range.

1. INTRODUCTION

Many languages have more than one way of constructing yes/no questions. In English, the most common construction is by switching the order of the subject and the verb of the sentence (VSO). However, in informal contexts, one could also use a statement construction (SVO) with a question intonation to achieve the same function.

On the other hand, in Mandarin, there are at least three ways of constructing yes/no questions. The most common way is by adding the question particle *ma* to the sentence-final position of the statement. Another way is by morphologically transforming the verb into an A-not-A construction (e.g., *ni*^{3} *yao*^{4}-*bu*^{2}-*yao*^{4} 'you want-not-want,' meaning 'do you want to?'). Finally, one can also use a statement construction with a question intonation, as in English. However, unlike English, such a construction is marked, and always conveys a tone of incredulity.

Intonations of yes/no questions in Mandarin have been widely studied [3, 7, 8]. Shen [9] showed that yes/no questions with and without *ma* are overall higher in pitch than the corresponding statements, while the A-not-A yes/no questions end as low as the corresponding statements.

For questions without *ma*, the results corresponded nicely with those found in English [4, 11] and Korean [5], in which the enlargement of the pitch range is associated with both the production and perception of incredulous intonations. However, as for questions with *ma*, studies do not always agree. Sung and Chiang [10] showed that questions with *ma* end high in pitch, which cause a rise towards the end of the sentence. In contrast, Chiang [3] showed that such questions end low, resulting in downward pitch contours. In other words, Sung and Chiang [10] corresponded better with Shen [9]. The discrepancy between Chiang [3] and Shen [9] could be easily explained away by possible dialectal differences, as the former studied Taiwan Mandarin, and the latter Mainland Mandarin. However, the discrepancy between Sung and Chiang [10] and Chiang [3] could not be as easily dismissed, as both studied the Taiwan Mandarin variety.

We suspected that differential elicitation methods might be the reason for the discrepancies in the results. Chiang [3] asked subjects to read sentences without contexts, while Sung and Chiang [10] provided subjects with felicitous discourse contexts for target elicitation. A closer inspection of the contexts used showed that most of the intended targets are incredulous yes/no questions. On the other hand, since Chiang [3] did not provide contexts for subjects, we suspected that it was the default/neutral questions that might have been elicited.

2. AIMS OF THE STUDY

There are two specific aims in this study. The first is to study the effect of incredulity on the realization of yes/no question intonation. If incredulity is the reason for discrepancies in yes/no questions with *ma* in previous studies, then one would expect to find similar intonation patterns in questions with and without *ma* under incredulous contexts, as questions without *ma* are interpreted as incredulous questions by default.

Secondly, we would like to study the effect of the final particle *ma* on incredulous questions. As questions with *ma* are much more common than those without in Mandarin, one suspected that the degree of incredulity might also be different. Considering questions with *ma* indicating a more unmarked situation, one would therefore predict that incredulous questions with *ma* might have a lower pitch height than those without *ma*.

3. METHOD

3.1. Materials

Stimuli were four two-syllable proper names with identical adjacent tones – *Ou*^{1}*la*^{1}, *Liu*^{2}*min*^{2}, *Li*^{3}*mei*^{3}, and *Ye*^{4}*na*^{4}. Each name was assigned a distinct cartoon character. All syllables were sonorants to facilitate pitch extraction. These four names were inserted into a carrier sentence, *ta*^{1} *shi*^{4} _____. 'S/he is_____.' There were four conditions, including incredulous questions without *ma* (IQ), incredulous questions with *ma* (IQm), neutral questions with *ma* (NQm), and neutral statements (NS). Appropriate punctuation marks were used to facilitate production (Table 1). Each sentence along with its corresponding cartoon character was printed on A4 paper. Additional contexts were given verbally for each condition. Incredulity was created by mismatching names and cartoon characters (Table 2). In total there were 4 (names) × 4 (conditions) = 16 stimuli.

Table 1: Stimuli used for elicitation.

Condition	Carrier Sentence
IQ	***ta*^{1} *shi*^{4} _____!?** 'Is he/she _____!?'
IQm	***ta*^{1} *shi*^{4} _____ *ma*^{0}!?** 'Is he/she _____!?'
NQm	***ta*^{1} *shi*^{4} _____ *ma*^{0}?** 'Is he/she _____?'
NS	***ta*^{1} *shi*^{4} _____.** 'He/She is _____.'

3.2. Subjects

Subjects were ten female native speakers of Taiwan Mandarin, aged from 18 to 30.

3.3. Equipment

A PCM-M1 DAT recorder and SONY MDR-7502 dynamic stereo headphones were used for recording.

3.4. Procedure

Subjects were seated in a sound-treated room. They were introduced to the four cartoon characters and their names, and were asked to read the sentences after given the verbal conditions. The sentences were blocked by condition (NQm, NS, IQ, IQm) (Table 2). Within each block, stimuli were randomized in the order of (T1, T4, T2, T3 for NQm and NS, and T2, T3, T1, T4 for IQ and IQm). The four blocks were repeated once. In total, there were 16 (stimuli) × 10 (subjects) × 2 (repetitions) = 320 sentences.

Table 2: The provided contexts for elicitation.

Stage	Provided Context
Prep.	Four cartoon characters and four names, were shown to the subjects.
NQm	"Now since you do not know which name belongs to which character, you will have to ask me one by one." ***ta*^{1} *shi*^{4} *Ye*^{4}*na*^{4} *ma*^{0}?** 'Is she Yena?' 
NS	"Now I will let you know every character's name, and please tell me who they are." ***ta*^{1} *shi*^{4} *Ye*^{4}*na*^{4}.** 'She is Yena.'
IQ	"After a while someone came to you and told you that in fact it is this character that is called *Ye*^{4}*na*^{4}. You were very surprised, and said, …" ***ta*^{1} *shi*^{4} *Ye*^{4}*na*^{4}!?** 'Is she Yena!?' 
IQm	"Now here comes another person. He told you again that the girl with short hair is indeed *Ye*^{4}*na*^{4}. So you were still surprised, and said, …" ***ta*^{1} *shi*^{4} *Ye*^{4}*na*^{4} *ma*^{0}!?** 'Is she Yena!?'

3.5. Judgment test

The judgment test was conducted to only select sentences that succeeded in expressing incredulity. The 320 sentences were divided into four groups respectively, constituting eight groups. Each group had 40 sentences containing every speaker's utterance of all four conditions and all four tones. These 40 sentences in each group were arranged in random order. The interstimulus interval was 4-sec.

24 proficient Mandarin speakers other than the subjects above, aged from 18 to 30, participated in this experiment. They were divided into eight groups. Their task was to decide whether the given sentence was incredulous or not. Each sentence could only be listened to once. No play-back was allowed. Each sentence was judged by three people.

If the judgment of one particular sentence was correct, then that sentence got one point. Only sentences scored above two were included for further data analyses. Data of one speaker were entirely excluded due to her heavy Taiwanese Mandarin accent. In the end, 233 sentences were subjected to statistical analyses (NQm – 43; NS – 72; IQ – 58; IQm – 60).

3.6. Data analyses

Pitch extraction and measurement were done by *Praat* (version 4.5.15). F_0 of eligible sentences were extracted and hand-checked. The two syllables of the proper name in the stimuli was the focus of this study. For *Ou*^{1}*la*^{1}, the initial (H) and final F_0 values (H) were extracted. For *Liu*^{2}*min*^{2}, the initial (L) and final F_0 values (H) of the two rises were extracted. For *Li*^{3}*mei*^{3}, due to a tone sandhi rule of changing the first T3 to a T2 (resulting in *Li*^{2}*mei*^{3}), we measured the initial (L) and the final F_0 (H) of the rise of *li*^{3} and the initial (H) and the final F_0 (L) of the fall of *mei*^{3} (the final rise was often omitted in Taiwan Mandarin). For *Ye*^{4}*na*^{4}, the initial (H) and the final F_0 (L) of the two falls were measured.

4. RESULTS

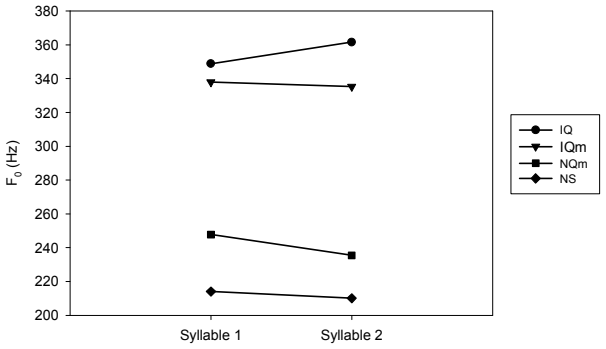
4.1. Pitch height

Two planned statistical analyses were executed. The first was a *Ma* (2) × Syllable position (2) two-way repeated-measures on the H-tone pitch extractions of the bisyllabic proper names to test the effect of the presence of particle *ma* on pitch height for incredulous sentences. Results showed that the main effect of *Ma* was significant [$F(1, 24) = 10.20$, $p < 0.01$, $\hat{\eta}^2 = .30$]. The interaction was also significant [$F(1, 24) = 8.29$, $p < .01$, $\hat{\eta}^2 = .26$].

As in Figure 1, post hoc paired *t* tests showed that pitch was higher in incredulous sentences without *ma* [Syllable 1: $t(24) = 2.57$, $p < .05$; Syllable 2: $t(24) = 3.44$, $p < .01$]. Across syllable positions, Syllable 1 was near-significantly higher than Syllable 2 in incredulous questions without *ma* [$t(32) = -1.82$, $p = .07$]. No Position difference was found for questions with *ma*.

The second analysis was a Incredulity (2) × Syllable Position (2) two-way repeated-measures on pitch height for questions with *ma*. There was a main effect of Incredulity [$F(1, 23) = 175.14$, $p < 0.0001$, $\hat{\eta}^2 = .88$]. The main effect of Position was also near-significant [$F(1, 23) = 3.90$, $p = 0.06$, $\hat{\eta}^2 = .15$] (Figure 1). Post hoc pairwise comparisons using Bonferroni's adjustments showed that incredulous sentences were significantly higher in pitch than neutral ones ($p < .0001$). In addition, Syllable 1 was marginally significantly higher than Syllable 2 ($p = .06$).

Figure 1: The mean highest F_0 values of different sentence types in two positions. The NS (neutral statement) serves as a baseline for comparison.

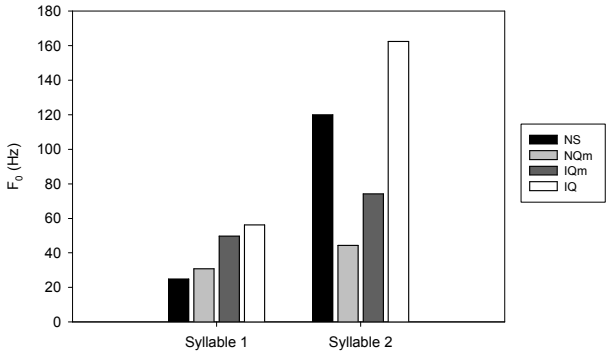


4.2. Pitch range

Figure 2 showed the pitch range of the three contour tones in two syllable positions. Pitch range was calculated by subtraction of pitch extraction at the lowest point of each syllable from its value at the highest point. (Since T1 is level tone, its pitch range was not included.) A *Ma* (2) × Position (2) two-way repeated-measured ANOVA was conducted to examine the effects of particle *ma* on pitch range of incredulous questions. Both of the main effects were significant [*Ma*: $F(1, 17) = 16.05$, $p < 0.001$, $\hat{\eta}^2 = .49$; Position: $F(1, 17) = 21.48$, $p < 0.001$, $\hat{\eta}^2 = .56$]. The interaction effect was also significant [$F(1, 17) = 11.95$, $p < 0.01$, $\hat{\eta}^2 = .41$]. Post hoc paired *t* tests showed that Syllable 2 always had a wider pitch range than Syllable 1 [w/ *ma*: $t(17) = -2.39$, $p < .05$; w/o *ma*: $t(25) = -5.47$, $p < .0001$]. On the other hand, the effect of *ma* was only affecting Syllable 2 [$t(17) = 3.94$, $p < .01$]. No difference was found in Syllable 1.

To explore the influence of incredulity, a similar Incredulity (2) × Syllable Position (2) two-way repeated-measures was conducted on questions with *ma*. Results showed that both of the main effects were significant [Incredulity: $F(1, 16)= 28.68$, $p < 0.0001$, $\hat{\eta}^2 = .64$; Position: $F(1, 16)= 4.64$, $p < 0.05$, $\hat{\eta}^2 = .22$]. The interaction effect was marginally significant [$F(1, 16)= 4.20$, $p = 0.06$, $\hat{\eta}^2 = .21$]. Post hoc paired *t*-test showed that incredulous sentences had significantly wider pitch ranges than neutral ones, regardless of positions [Syllable 1: $t(16) = -2.86$, $p < .05$; Syllable 2: $t(16) = -5.91$, $p < .0001$]. However, the pitch range was much larger in Syllable 2. In addition, Syllable 2 also had a wider pitch range than Syllable 1 for neutral questions [$t(25) = -2.34$, $p < .05$].

Figure 2: The mean pitch range of three contour tones in the two positions for four types of sentences.



5. DISCUSSION

Question intonations in Mandarin, as shown in this study, indeed pattern differently from statement intonations. In general, yes/no questions, with *ma* or without *ma*, in either neutral or incredulous condition, were higher in F_0 than statements. However, the patterning of pitch also varied among different types of questions. It was found that incredulous questions had higher F_0 contours than neutral questions. Further, incredulous questions without *ma* were even higher in F_0 than those with *ma*. This order held true regardless of the tone type.

Similar results were also found in the analyses of pitch range. Incredulous questions had a larger pitch range than neutral ones. In addition, pitch ranges of questions without *ma* were larger than those with *ma*. Taking the two results together, we suggested that higher pitch and larger pitch range were salient features of incredulous intonation. Since larger pitch range tends to indicate greater speaker's involvement [1, 6], this might be why incredulous sentences were wider in pitch range and higher in pitch. Along the same line, we also found questions without *ma* to be wider in pitch range and higher in pitch than questions with *ma*. As the latter is the default form of interrogatives in Mandarin and the former the marked form, there might also be a difference in the degree of incredulity between the two. The results in our data seemed to support this premise.

Another interesting finding was the unequal widening of pitch range in the two syllable positions. Across tones, pitch range of the second tone was significantly larger than that of the first one. We suspected that this might have to do with the word-final stress rule in Mandarin, which specified that stress tends to fall on the last syllable of words [2]. Further analyses should be included to further clarify this point.

6. REFERENCES

[1] Bolinger, D. 1989. *Intonation and its uses: melody in grammar and discourse.* London: Edward Arnold.

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

[3] Chiang, W.-Y. 1997. Guo^{2}yu^3 han^4 tai^{2}yu^3 dai^{4}you^3 yi^{2}wen^4 yu^{3}zhu^{4}ci^2 zhi^1 ju^{4}zi^0 de^0 yu^{3}diao^4 yan^{2}jiu^4 (Studies on intonations of interrogatives with particles in Taiwan Mandarin and Taiwanese), *Yin^{1}yun^4 lun^{4}cong^2* (*Phonology Studies*) 6, 827-864. Taipei: Student.

[4] Hirschberg, J., Ward, G. 1992. The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in English. *Journal of Phonetics* 20, 241-251.

[5] Jun, S.-A., Oh, M. 1996. A prosodic analysis of three types of wh-phrases in Korean. *Language and Speech* 39(1), 37-61.

[6] Ladd, D. R. *Intonational phonology*. Cambridge: Cambridge University Press.

[7] Liu, F., Xu, Y. 2005. Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica* 62, 70-87.

[8] Shen, X.-N. S. 1989. Interplay of the four citation tones and intonation in Mandarin Chinese. *Journal of Chinese Linguistics* 17(1), 61-73.

[9] Shen, X.-N. S. 1990. The prosody of Mandarin Chinese. Berkeley: University of California Press.

[10] Sung, L.-H., Chiang, W.-Y. 2006. When a neutral tone is no longer neutral: a study on the tonal behavior of the question particle *ma* in Taiwan Mandarin. Presented at the *14th Annual Conference of the International Association of Chinese Linguistics & 10th International Symposium on Chinese Languages and Linguistics Joint Meeting*, Taipei, Taiwan, May 25-29.

[11] Ward, G., Hirschberg, J. 1988. Intonation and propositional attitude: the pragmatics of L* + H L H%. *Proceedings of the fifth Eastern States Conference on Linguistics*, 512-22.

THE EFFECTS OF PHONETIC DISTANCE, LEARNING CONTEXT AND LEARNER PROFICIENCY ON L2 PERCEPTION OF ENGLISH LIQUIDS

Sally Chen and Janice Fon

National Taiwan University
{d93142002,jfon}@ntu.edu.tw

ABSTRACT

This study aims to investigate the effects of phonetic distance, learning context and learner proficiency on L2 perception of English liquids. Reaction time difference between the pre- and post-tests was analyzed. Results showed that the natural context induced the most progress for participants of a lower L2 proficiency level, while no preference was shown for those of a higher proficiency level. In general, L2 learners showed more progress for liquids occurring in novel phonotactic structures. Phone effect was significant only when L2 learners of lower proficiency perceived liquids in the singleton position.

Keywords: L2 perception, phonetic distance, learning context, SLM, PAM.

1. INTRODUCTION

Perception of unfamiliar L2 sounds has always been an obstacle that learners of the language have to strive to conquer. Both the speech learning model (SLM) [9, 10] and the perceptual assimilation model (PAM) [2, 4, 5] predict the degree of success listeners have in perceiving nonnative sounds is based on the perceived phonetic distance between L1 and L2 sounds. The two models differ in that the SLM focuses on the aspect of L2 learning, while the PAM is established on the fact that some foreign sounds are harder to perceive than others [12].

Regarding language contrasts, the SLM proposes that the greater the perceived dissimilarity between an L2 sound and its closest L1 counterpart, the more likely a new category is to be formed for the L2 sound [9, 10, 12]. Modeling from another perspective, the PAM proposes that sounds in a foreign language are perceived in accordance with their similarities to the closest native language sounds articulatorily [2, 4, 5, 12]. When two sounds are assimilated to two different categories, the perceptual performance is expected to be good. However, when two sounds are assimilated to one single category, the performance will be unsatisfactory [16]. In addition, based on a logic similar to the Native Language Magnet theory [14, 15], when two sounds are assimilated to the same category, it is believed that in cases where only one of the two sounds is regarded as a good exemplar of that target category, the performance will be much better than in those where both sounds are regarded as comparable candidates for the same target category [12].

2. AIMS OF THE STUDY

There are three specific aims of this study. The first is to examine the effect of phonetic distance. The two English liquids, /l/ and /r/, impose differential levels of difficulty on Mandarin speakers [7]. The former has a closer counterpart in Mandarin than the latter ([l̪] *vs.* [z̪]). According to the SLM, this would predict that L2 listeners are more likely to form perceptual categories for /r/ than for /l/. On the other hand, according to the PAM, both categories should be perceived equally well. This study investigates which model better interprets the perception of English liquids by Mandarin speakers of English.

The second aim is to investigate the effect of learning contexts. In Taiwan, the audiolingual method is used very often. Students learn new words and sentences by repeating after the instructor, often without contexts. However, in recent years, more and more EFL instructors, especially those outside the regular school system, are using some combination of language immersion strategies [6], believing that this can promote greater success in L2 learning. Therefore, it would be interesting to see whether the presentation method of the stimuli affects L2 sound perception.

Finally, this study looks into the effect of speaker proficiency. As listening is not a skill that

is much emphasized in Taiwan's mandatory education system, it is often hard to assess accurately students' listening comprehension abilities. However, it is worth investigating whether differences in general language proficiency affect L2 sound perception after a short-term training period.

3. METHOD

3.1. Participants

135 participants were recruited for this study, including 27 native speakers of English (with mean age of 22.89 years old and standard deviation of 4.31) and 108 EFL college students (with mean age of 21.00 years old and standard deviation of 1.46) with Mandarin as their native language. Half of the EFL learners were of high proficiency level and the other half were of low proficiency level as determined by the General English Proficiency Test (GEPT), a national English proficiency test held annually by the Language Training and Testing Center. The high proficiency level group passed the high-intermediate level of the GEPT while the low proficiency level group did not.

3.2. Materials

56 monosyllabic words were chosen from *the 1000 Basic English Vocabularies for Elementary and Junior High School Students* [8] announced by the Ministry of Education in Taiwan. These words either begin with an /l/ or /r/ (e.g. *last* and *rain*) or a Cl- or Cr- cluster [e.g. *climbed* and *broke*]. Another 56 phonotactically-matched monosyllabic pseudo words were also included as fillers (e.g., *laped* and *reim*). In total, there were 56 (targets) + 56 (fillers) = 112 stimuli. All materials were recorded by a female speaker of American English. Recordings were done in a sound-treated room with a sampling rate of 48 kHz and, were later downsampled to 16 bit 22050 kHz using Adobe Audition 2.0.

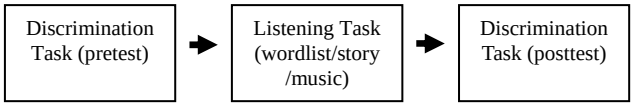
3.3. Equipment

A SONY DAT PCM-M1 digital recorder along with a Maxell DM120 tape and a SHURE SM10A-CN head-mounted microphone were used for recording stimuli. E-Prime 1.0 and its accompanying PST Serial Response Box (SR Box) Model 200a were used for the perception tasks.

3.4. Procedure

Participants were seated in a quiet room and were given a three-block perception task, as shown in Figure 1.

Figure 1: Procedure of the experiment.



The first block was a 2-alternative-force-choice (2AFC) speeded perception task on target words. Participants were instructed to respond by pressing the corresponding button on the SR Box as accurately and quickly as possible. The block lasted about 5 minutes.

The second block was a listening task, in which participants were randomly assigned to one of the three audio conditions, i.e. wordlist, story, and music (Table 1). The wordlist condition contained all 56 target words (inter-stimulus interval = 3s), and the story condition was a tailor-made children's story containing all 56 target stimuli. All other words in the story were carefully chosen so as to avoid any /l/ and /r/ sounds. The music condition was a piece of classical music and served as a control. All three conditions lasted for 3 minutes and 30 seconds. Half of the L2 speakers listened once (Short version) while the other half listened twice (Long version). Native speakers only listened to the short version to avoid boredom.

Table 1: Excerpts of the 3 audio conditions in the listening task.

Audio condition	Excerpt of the content
Wordlist	**Black, round, live, group, place, love, rice, like, read, play, dream**….
Story	My name is Andy. I am a mouse but I am big and **black** and **round**. I **live** with my mom and a **group** of mice in a nice **place**. We **love** to eat **rice**. We **like** to **read** books and **play** games. Each day I **dream** of catching a cat….
Music	(Handel's *Le Rejouissance*)

The third block was again a 2AFC speeded perception task, as in Block 1. However, in this posttest, both target words and fillers were included. The block lasted about 10 minutes.

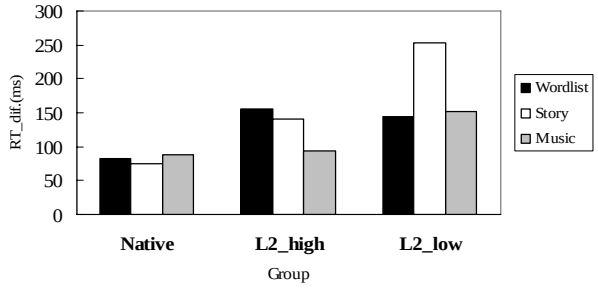
4. RESULTS

4.1. General contrast

A Proficiency (3) × Task (3) × Position (2) × Phone (2) four-way ANOVA was performed for the Short version on RT difference between the pre- and post- tests to examine the effect of listening conditions. Results showed that except for Phone, all the other main effects were significant [Proficiency: F (2, 4285) = 41.23, p < .0001; Task: F (2, 4285) = 8.39, p < .001; Position: F (1, 4285) = 38.69, p < .001]. The two-way interaction between Proficiency and Task was also significant [F (4, 4285) = 8.53, p < .0001]. The three-way interaction concerning Group, Position, and Phone was marginally significant [F (2, 4285) = 2.85, p = .06]. The four-way interaction was not significant.

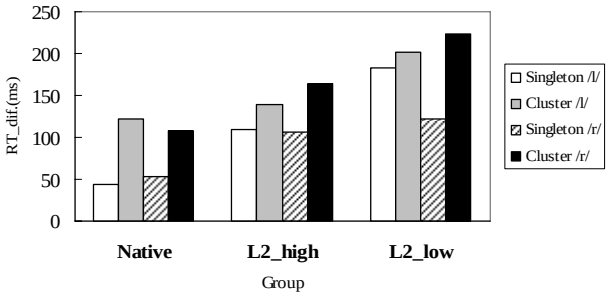
Figure 2 shows the interaction between L2 proficiency levels and different audio contexts in the listening task. Higher bars indicate more improvement. Given that native speakers were already fast in the pretest, they made relatively less progress between the pretest and the posttest. Post-hoc analyses of the two-way interaction showed that there was no significant difference among the three tasks for the native speakers. For L2 speakers of higher English proficiency, listening to music yielded the least progress (p < .01), while for those of lower English proficiency, the story condition yielded the best results (p < .0001).

Figure 2: Interaction between English proficiency and the audio tasks assigned.



Regarding the three-way interaction, as shown in Figure 3, Position is significant for all three proficiency groups. More progress was found for liquids in the cluster position [p < .01]. Phone contrast was significant only when L2 learners of lower proficiency perceived liquids in the singleton position: the sound /l/ was perceived with more progress than the sound /r/ between the pre- and post-tests [p < .05].

Figure 3: RT difference between /l/ and /r/ in different positions across 3 proficiency groups.

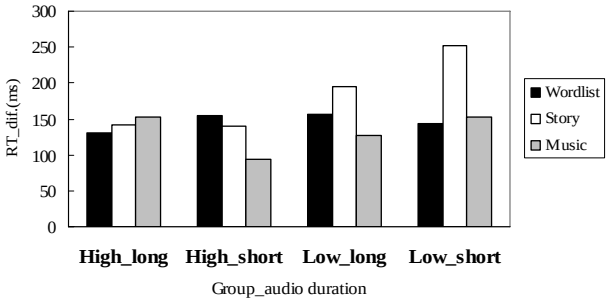


4.2. Durational contrast

A Proficiency (2) × Task (3) × Duration (2) × Position (2) × Phone (2) five-way ANOVA was performed on nonnative speakers to investigate whether the duration of audio input affected the RT difference between two perception tests. Results showed that Duration had a 2-way interaction with Proficiency [F (1, 5579) = 4.28, p < .05] and a three-way interaction with Proficiency and Task [F (2, 5579) = 4.66, p < .01].

Post-hoc analyses showed that for L2 speakers with a higher proficiency, Duration and Task had interactions: As shown in Figure 4, those who listened to the Short version in the listening task made more progress after given the "wordlist" or the "story" in the listening block (p < .01). On the other hand, for the lower proficiency group, Duration and Task effects were found significant. Longer listening times reduced their progress. In addition, those who listened to the "story" in the listening block showed the most progress, as previously mentioned in 4.1.

Figure 4: Interaction among English proficiency, Duration, and Task.



5. DISCUSSION

Unlike Japanese EFL speakers [1, 11], the two different liquid sounds were not perceived differently for Mandarin EFL learners; however, their interaction with phonotactic structures (singleton *vs.* cluster) did conform to the SLM in that for novel combinations, such as liquids in consonant clusters, L2 speakers were able to make more progress (same trend as native speakers) after being exposed to the listening task for a few minutes. In addition, the fact that those of lower L2 proficiency level made more progress on /l/ – the phone considered as a good exemplar to their L1 – supported the PAM.

Results in this study also showed that different learning contexts did facilitate L2 speakers' perception of L2 sounds to different extents. On top of that, L2 speakers' proficiency level also played a role – those of a higher proficiency level made comparable progress regardless of whether the listening material was natural or mundane; while those of a lower L2 proficiency level required that the material be interesting enough to keep them attentive. One explanation is that the basic linguistic unit for speech processing could be lexical items, instead of segments. As the lexical restructuring model proposes, as one becomes more experienced and gains better understanding of the relationship among lexical items, one gradually acquires better segmental representations [3, 13]. Considering that L2 speakers with a lower proficiency might not be as experienced with L2 words as those with a higher proficiency, providing a temporary artificial setting in which they could more easily define the distinctions among the stimulus words (such as the "story") thus boosted their performance on the differentiation of the two liquid sounds.

In sum, results of this study clearly indicated that all fundamental differences between L1 and L2 phonological systems need to be taken into consideration when designing EFL training materials. Further studies are required to claim whether the most fine-grained L2 perception unit should be of the segmental or the lexical level.

6. REFERENCES

[1] Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: the case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics 32*, 233-250.

[2] Best, C. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research*, 171-204. Timonium, MD: York Press.

[3] Bradlow, A. R. & Pisoni, D. B. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *Journal of the Acoustical Society of America* 106 (4), 2074-2085.

[4] Best, C. T., McRoberts, G. W., Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listeners' native phonological system. *Journal of the Acoustical Society of America* 109 (2), 775-792.

[5] Best, C. T., McRoberts, G. W, Sithole, N. (1988). The phonological basis of perceptual loss for non-native contrasts: maintenance of discrimination among Zulu clicks by English speaking adults and infants. *Journal of Experimental Psychology, Human Perception Performance* 14, 345-360.

[6] Chan, Y. -j., Lin, S. -H. (2001). The empirical study of whole language teaching in English class of junior high. *Journal of Taiwan Normal University* 46 (2), 233-253.

[7] Chen, S., Fon, J. (2005). To play or to pray?–The influences of initial exposure age and environment on perception of Mandarin speakers of English on devoiced liquids [Abstract]. *Journal of the Acoustical Society of America*, 118 (3), Pt. 2, 1898-1899.

[8] Dept. of Elementary Education, Ministry of Education, Taiwan. http://www.boe.tyc.edu.tw/1000words/%B3%CC%B0%F2%A5%BB%A4@%A4d%A6r%B5%FC.htm

[9] Flege, J. E. (1995). Second Language Speech Learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research*. 233-277. Timonium, MD: York Press.

[10] Flege, J. E., Birdsong, D., Bialystok, E., Mack, M., Sung, H., Tsukada, K. (2006). Degree of foreign accent in English sentences produced by Korean children and adults. *Journal of Phonetics* 34, 153-175.

[11] Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "l" and "r". *Neuropsychologia* 9, 317-323.

[12] Guion, S. G., Flege, J. E., Akahane-Yamada, R., Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America* 107 (5), 2711-2724.

[13] Imai, S., Walley, A. C., Flege, J. E. (2005). Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners. Journal of Acoustical Society of America 117 (2), 896-907.

[14] Kuhl, P. K. (1993). Innate Predictions and the effect of experience in speech perception: the native language magnet theory. In B. de Boysson-Bardies et al. (eds.), *Developmental neurocognition: Speech and face processing in the first year of life*, 259-274. Dordrecht: Kluwer Academic Publishers.

[15] Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Science*, 97, 11850-11857.

[16] Sebastian-Galles, N., Soto-Faraco. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition* 72, 111-123.