TONE CHANGE IN MANDARIN: COMPARING ACOUSTIC ANALYSES FROM THE EARLY 20th CENTURY TO TODAY

Cathryn Yang
Payap University & SIL International
cathryn_yang@sil.org

ABSTRACT

This real-time study of Standard Mandarin tones presents a new acoustic analysis of Yuen Ren Chao's 1922 archival recordings and compares the results to studies from the early 21st century. Three positional variants (citation, final, and non-final) are compared across time, and Tone 2 and Tone 3 are shown to have undergone subtle changes to their contours. Two phonetic mechanisms induce contextual variants that, over time, spread to more environments and result in contour change. For both tones, carryover effects of a high preceding offset leads to raising of f0 onset and rightward sliding of f0 turning point. For Tone 3, the final rise is truncated due to the time pressure on f0 movement, which leads to contour reduction. Overall, these findings support Pittayaporn's phonetically-grounded model of tone change, which connects synchronic tone variation with diachronic tone change.

Keywords: Tone change, real-time study, Mandarin, archival recordings, phonetically-grounded sound change

1. INTRODUCTION

Phonetic tone change (i.e., change over time of a tone's f0 height and/or contour) has not received much scholarly attention and may be one of the least understood types of sound change [1]. Lack of empirical studies makes it difficult to answer basic questions, e.g., is there directionality in tone change? Real-time studies could provide valuable evidence for inferring the direction of tone change [2], but few such studies exist, e.g., [3] on Bangkok Thai. This study investigates tone change in Standard (Beijing) Mandarin in real time, by conducting a new acoustic analysis of archival recordings from 1922 [4] and comparing the results to acoustic studies from 2006 [5] and 2018 [6].

Early 20th century acoustic studies of Mandarin tone such as [7] and [8] focused exclusively on the citation form (i.e., the surface form in monosyllables uttered in isolation), without consideration of positional variants. However, tone shapes are known to vary according to the segmental and prosodic environment, including the position within an utterance [9].

In this study, a new analysis of phonograph recordings of Yuen Ren Chao (赵元任) from 1922 [4] enables comparison of three positional variants: citation form, the first syllable in a disyllabic word (non-final form), and the second syllable in a disyllabic word (final form). The 1922 data afford examination of the synchronic effects on tone contour of the preceding and following tone. Comparing contextual variants from 1922 to those seen in the early 21st century allows us to track which variants have changed in their distribution, that is, spread beyond their original conditioning environment. This study thus presents an opportunity to see links between synchronic variation and diachronic change, following the approach proposed in [3].

Mandarin has four lexical tones (Figure 1) and a neutral tone in unstressed syllables (not included in the present study) [10]. In Chao pitch numbers [11], with "1" representing low and "5" high, Tone 1 (T1) is usually transcribed as 55; Tone 2 (T2) as 35; Tone 3 (T3) as 214; and Tone 4 (T4) as 51 [10]. The above transcriptions correspond well with the 1922 tone system in Figure 1 (left panel), but less well with the system in 2006 [5] (Figure 1, right panel). T1 and T4 citation forms have not changed substantially, but the contours of T2 and T3 have changed noticeably. This study focuses on describing their changes.

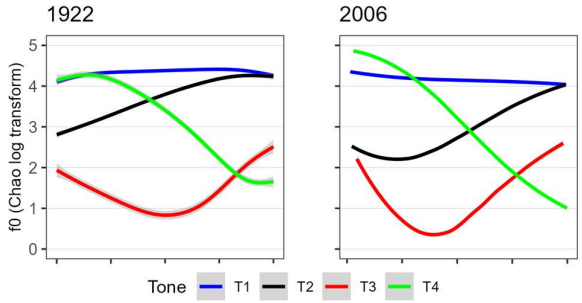

Figure 1: Left panel: Mandarin citation tones in 1922, n=187; Right panel: Mandarin citation tones in 2006, n= 2080 (adapted from [12] with permission).

2. DATA AND METHOD

Four acoustic studies, two from the early 20th century and two from the early 21st century, are compared with the new analysis from 1922. [7] from 1915 and [8] from 1934 are some of the earliest acoustic studies on Standard Mandarin citation tone. The early 21st century acoustic studies selected for comparison are

[5] from 2006 and [6] from 2018 on Beijing Mandarin. Standard Mandarin tone values are based on Beijing Mandarin pronunciation [13].

In the 2006 study [5], 10 monosyllabic words were recorded for each of the four tones. The sample included 52 speakers (balanced for gender) across a wide age range. To enhance comparability, f0 data from the 30–40-year age group (7 speakers) were selected for comparison with Chao, who was 32 at the time of the 1922 recordings. In the 2018 study [6], 50 speakers (balanced for gender; mean age = 20.9 years) were recorded reading 16 disyllabic words (1 combination of each tone pair); thus, the dataset was balanced for preceding and following tone.

In both 21st century studies, f0 was measured at 9 equidistant time points throughout the rime. Inter-speaker variation of f0 range was normalized by logarithmic transform according to the equation in (1), where *x* represents the observed f0 value, and *max* and *min* represent speaker-specific maximum and minimum f0 values [5]:

$$T = 5 * \frac{(lg\ x - lg\ min)}{(lg\ max - lg\ min)} \quad (1)$$

The subject of the present study's new analysis is Yuen Ren Chao, well-known linguist and composer. Chao was born in 1892 and spent his early childhood in northern China; as a child, he learned several Chinese lects including Mandarin [14]. A gifted scholar, Chao taught an early form of Standard Mandarin at Harvard in 1922 [13]. This early version of the national standard, replaced by the 1930's, included a fifth tone (reflex of the *ru* tone) from Nanjing Mandarin, but the other four tones followed Beijing Mandarin pronunciation [13].

The 1922 phonograph recordings, originally produced to accompany the textbook [4], were downloaded from [15]. 200 monosyllabic words (citation form) and 234 disyllabic words were selected and annotated in Praat [16]. Tokens with pitch measurement errors or with the fifth or neutral tone were excluded, leaving a total of 549 tokens. f0 of the rhyme was extracted every 10 msec. Annotated recordings and the R [17] code used to produce the relevant figures are available at [18]. f0 was normalized according the function in (1) to facilitate comparison with the 21st century studies.

The 1922 disyllabic data was not balanced for preceding and following tone: there were more tokens of the T2-T2 sequence compared to other sequences. To enhance comparability with the 2018 data, which was balanced for prosodic context, 28 T2 tokens were excluded, resulting in 9-15 tokens for each sequence. The one exception was the T2-T3 sequence, which had only 4 example words. For T3, there were 7-14 tokens for each preceding and following environment (except T2-T3).

3. RESULTS

3.1. Tone 2

T2 citation form has subtly but visibly changed over the past 100 years. In 1915 [7] and in 1934 [8] (Figure 2), T2 appears as a mid rising tone that starts its rise within the first 10% of the rhyme; there is no dip. Results from the 1922 data (Figure 3, left panel) show a similar rising contour. The similarity between the studies' results implies that the no-dip variant was standard in the early 20th century, not simply an idiosyncrasy of Chao.

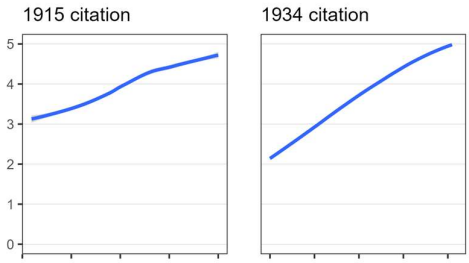

Figure 2: Left: T2 citation form in 1915, redrawn from [7] and 1934, redrawn from [8] (© The Physical Society of Japan).

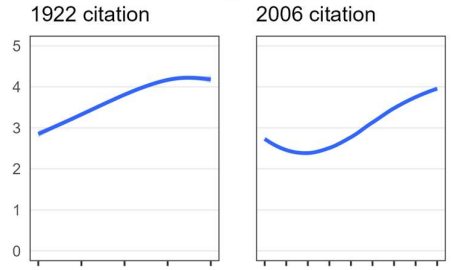

Figure 3: T2 citation form in 1922 (left, n=40); and in 2006 (right, n=70, redrawn from [5]).

In contrast, there is a noticeable difference between the early 20th century and the early 21st century (Figure 3 right panel). In the 2006 study [5], the first third of the f0 trajectory entails a small fall before the rise begins. A similar contour is seen in [19] from 1997 (not pictured here), which has a slight dip and a f0 turning point at around 25% into the syllable.

Besides citation form, Chao's 1922 recordings of disyllables afford the opportunity to compare contours in other positions. Figure 4 compares non-final and final contours in 1922 with those of the 2018 study [6]. In 2018, the f0 onset is higher and the rise begins later (at approximately 40% relative time point) compared to the early 20th century—similar to the change observed for citation tone.

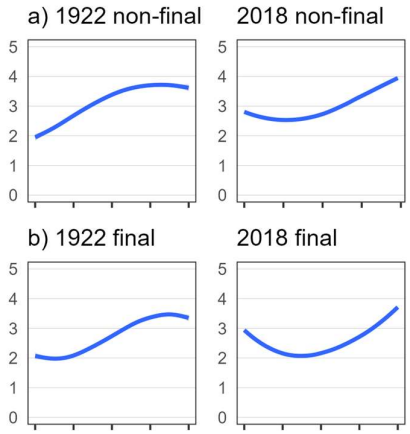

Figure 4: a) T2 non-final in 1922 (n=42) and in 2018 (n=400); b) T2 final in 1922 (n=50) and in 2018 (n=400). 2018 plots adapted from [12] with permission.

Where did the innovative dipping variant of originate from? Figure 5 compares T2 final forms in different prosodic contexts: after a tone with a high f0 offset (that is, T1, T2, and T3 sandhi form) vs. after T4, which has a low f0 offset. In the high context, T2 f0 onset is raised slightly compared to the low context; this raising is an assimilatory carryover effect of the same type reported in [19]. Also, there appears a short recovery period that results in a later f0 turning point. It is this rightward-shifted variant that has spread outside of its original conditioning environment and now appears in the 21st century citation and non-final forms.

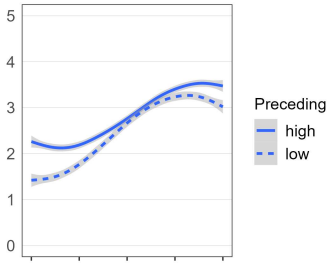

Figure 5: T2 in final position in 1922 (n=42); line type represents pitch height of the preceding tone's offset.

3.2. Tone 3

In the majority of tone descriptions from the mid-19th century to 1915, T3 is described as low rising, but between 1915 and the 1930's, descriptions vary, some reporting low rising and others low dipping [20]. Since the 1940's, most descriptions are low dipping [20]. Based on the above trends, Endo [21] suggests that T3 citation form changed from low rising to low falling-rising sometime in the early 20th century. The comparison of Figure 6 with Figure 7 supports Endo's [21] proposal. In the 1915 and 1934 studies, T3 has a "short level run" [7] and then rises—there is no dip. However, in 1922 (Figure 7 left panel), f0 onset is slightly higher (at Chao pitch number 2 rather than 1), and this raising results in an early dip before the rise begins (similar to the change in T2).

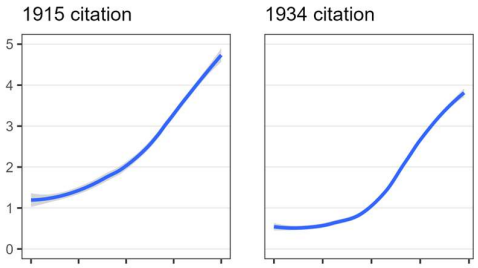

Figure 6: Left: T3 citation form in 1915, redrawn from [7] and 1934, redrawn from [8] (© The Physical Society of Japan).

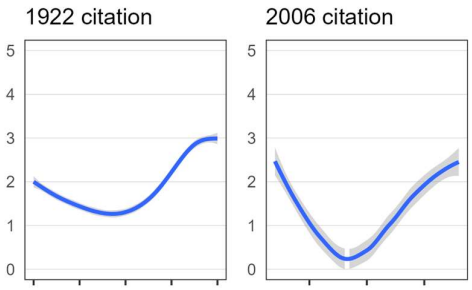

Figure 7: T3 citation form in 1922 (left, n=70) and in 2006 (right, n=70, redrawn from [5]).

The citation form in 1922 and 2006 (Figure 7 right panel) both have a dipping contour, though T3's f0 minimum in 2006 appears considerably lower. The current data is insufficient to confirm whether there was a true lowering of f0 minimum over time, but since the same normalization function was used across the compared studies, this potential change warrants further investigation, along with the potential connection between lower f0 minimum and the development of creaky voice as a phonetic correlate of T3 [22].

Differences are also apparent in non-final and final forms (Figure 8). Chao [4] describes the low falling non-final form as the "half T3" variant, because only the first half of the contour is present. Contour simplification in non-final position is often due to shorter duration [23]. In 1922, mean duration of non-final form is 78% of that of the final form. Shorter duration leads to increased time pressure on f0 [24]. The 1922 non-final form shows both target undershoot (raised f0 minimum) and truncation (cut-off of the rise) compared to citation form. The half T3 variant appears only in non-final form in 1922, but has spread to final form in 2018. Xu's [19] study of disyllabic sequences in 1997 report a similar spreading of the non-final allotone to final position.

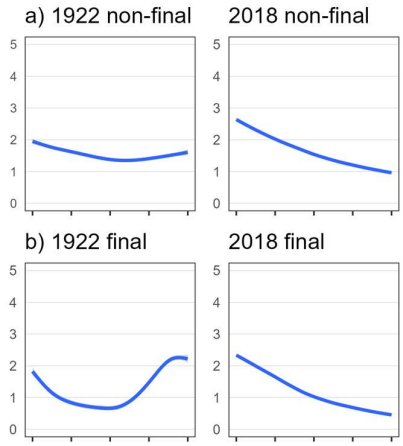

Figure 8: a) T3 in non-final position in 1922 (n=37) and in 2018 (n=350); b) in final position in 1922 (n=34) and in 2018 (n=400, adapted from [12] with permission).

In addition, T3 f0 onset has raised over time: below Chao level 2 in 1922 but consistently higher than level 2 in 2018. In 1922, T3 raised onset is observed after a preceding high offset, as seen in Figure 9. This variation is similar to the results for T2, implying that the same conditioning factors, i.e., preceding tone context, affected both T2 and T3.

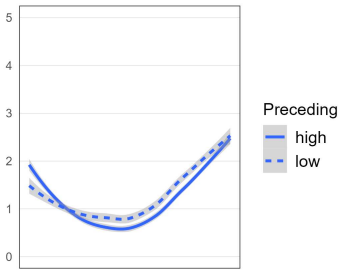

Figure 9: T3 in final position in 1922 (n=34).

4. DISCUSSION

The observed changes in Mandarin show similarity to the changes reported for Bangkok Thai [3], seen in Figure 10. Over the past 60 years, Thai Tone 4 (High) has gradually developed a slight dip or level period before its rise, similar to Mandarin T2. Thai Tone 5 (Rising) is gradually losing its final rise, similar to Mandarin T3. These similarities are not shared innovations or caused by language contact; instead, it is likely that the same phonetic biases are at play in both languages.

From a typological perspective, the observed changes follow several crosslinguistic trends in tone change: 1) raising of f0 onsets that are below mid, 2) rightward sliding of f0 turning points, and 3) reduction of low rising contour [25]. Raising of low f0 onsets is a well-attested carryover effect of high preceding offset in both Thai [26] and Mandarin [19]. Carryover effects also result in a short recovery period before the initial tone target is reached, causing a later temporal location of the f0 turning point. In the case of T3, shorter duration in the non-final position results in truncation of the rise. This study shows that the variants with raised f0 onset, later alignment of f0 turning points and/or reduced contour have spread outside their original conditioning environment. The change over time is a result of these variants' change in distribution.

What phonetic biases underlie these changes? Pittayaporn [3] points to time pressure on f0 contour as a generator of synchronic variants that may then spread to other environments. Xu [27] identifies two time constraints on f0: 1) minimum duration for f0 movement and 2) synchronization of the tone with the segmental string. In other words, it takes a certain amount of time to make f0 transitions, but there is also a limit on how much time is available, depending on the tonal domain. These articulatory constraints are responsible for time pressure on f0 movement, and their diachronic consequences can be seen in crosslinguistic trends in tone change.

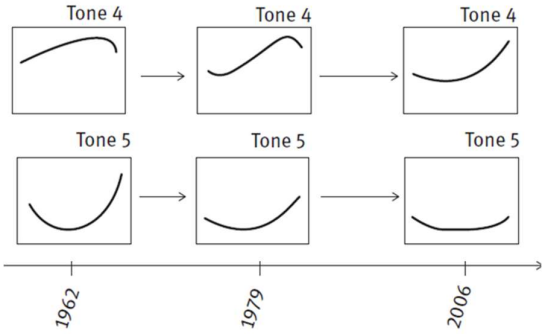

Figure 10: Bangkok Thai citation tone over the past 60 years. Adapted from [3] with permission.

5. CONCLUSION

This study conducts tonal acoustic analysis on Yuen Ren Chao's recordings from 1922 and compares the analysis to acoustic studies from 1915 [8], 1934 [9], 2006 [6] and 2018 [7]. The differences between the 20th century studies and the 21st century studies for T2 and T3 are interpreted as reflecting tone change over time; of course, more data from the early 20th century is needed to confirm this. In the 1922 data, certain prosodic environments condition a raised f0 onset and later temporal location of the f0 turning point and/or a truncated contour. The raised-onset, rightward-shifted and/or truncated variants spread beyond the original conditioning environment by the early 21st century, resulting in contour change. This study thus reveals a link between synchronic variation and diachronic change, evidence that supports Pittayaporn's [3] phonetically-grounded tone change model.

6. ACKNOWLEDGEMENTS

Sincere thanks to Elissa Low for segmentation and labelling. This project was supported by SIL International.

7. REFERENCES

[1] E. W. Campbell, "Why is tone change still poorly understood, and how might documentation of less-studied tone languages help?," in *Historical linguistics and endangered languages: Exploring diversity in language change*, P. Epps, D. Law, and N. Pat-El, Eds., Routledge, 2021, pp. 15–40.

[2] J. Tillery and G. Bailey, "Approaches to real time in dialectology and sociolinguistics," *World Englishes*, vol. 22, no. 4, pp. 351–365, 2003, doi: 10.1111/j.1467-971X.2003.00305.x.

[3] P. Pittayaporn, "Phonetic and systemic biases in tonal contour changes in Bangkok Thai," in *Tonal change and neutralization*, H. Kubozono and M. Giriko, Eds., in Phonology and Phonetics, vol. 27. Berlin: Mouton de Gruyter, 2018, pp. 249–278.

[4] Y. R. Chao, *Guoyu liushengpian keben (國語留聲片課本) [A textbook of the national language with gramophone records]*. Shanghai: Commercial Press, 1922.

[5] F. Shi and P. Wang, "Beijing hua danziyin shengdiao de fenzu tongji fenxi (北京话单字音声调的分组统计分析) [A statistical analysis of tone groups in Beijing Mandarin]," *Dangdai Yuyanxue*, vol. 8, no. 4, pp. 324–333, 2006.

[6] Y. Zhang, "Beijing Putonghua de shengdiao shiyan fenxi (北京普通话的声 调实验分析) [An experimental analysis of the tones of Putonghua in Beijing]," PhD dissertation, Beijing Language and Culture University, Beijing, 2018.

[7] C. B. Bradley, "The tone-accents of two Chinese dialects," *J. Am. Orient. Soc.*, vol. 35, pp. 199–206, 1915, doi: 10.2307/592645.

[8] S. Obata and T. Toshima, "Shinago no butsuri oto koe manabu teki kenkyū: (sore ichi) shi koe no seishitsu (支那語の物理音聲學的研究 (其一) 四聲の性質) [Acoustic study of Chinese (Part 1: Tone)]," *J. Math. Phys. Soc. Jpn. 日本数学物理学会誌*, vol. 8, no. 1, pp. 1–10, 1934, doi: 10.11429/subutsukaishi1927.8.1.

[9] C. DiCanio, J. Benn, and R. Castillo García, "Disentangling the effects of position and utterance-level declination on the production of complex tones in Yoloxóchitl Mixtec," *Lang. Speech*, vol. 64, no. 3, pp. 515–557, 2021, doi: 10.1177/0023830920939132.

[10] Y. R. Chao, *A grammar of spoken Chinese*. Berkeley: University of California Press, 1968.

[11] Y. R. Chao, "ə sistim əv 'toun-letəz,'" *Maître Phon.*, vol. 8 (45), no. 30, pp. 24–27, 1930.

[12] F. Shi, "Shilun Putonghua shengdiao de bendiao: Jiantan wuduzhi jidiaofa de xingzhi (试论普通话声调的本调: 兼谈五度值记调法的性质) [On the essential tones in Standard Chinese: Concurrent discussion on the nature of five-degree method]," *Int. J. Chin. Linguist.*, vol. 7, no. 1, pp. 140–157, Jun. 2020, doi: 10.1075/ijchl.00008.shi.

[13] Y. R. Chao, "What is correct Chinese?," *J. Am. Orient. Soc.*, vol. 81, no. 3, pp. 171–177, 1961, doi: 10.2307/595651.

[14] R. Levenson, "Chinese linguist, phonologist, composer and author, Yuen Ren Chao," 1974. Accessed: Oct. 29, 2022. [Online]. Available: https://oac.cdlib.org/view?docId=hb8779p27v&query=&brand=calisphere

[15] "Lao guoyin xuexi (老国音学习) [Study of the Old Standard Mandarin pronunciation]," 2022. https://zh.wikiversity.org/wiki/老國音學習 (accessed Oct. 18, 2022).

[16] P. Boersma and D. Weenick, "Praat: doing phonetics by computer." 2022. [Online]. Available: https://www.praat.org

[17] R Core Team, "R: A language and environment for statistical computing." R Foundation for Statistical Computing, Vienna, Austria, 2020. [Online]. Available: https://www.R-project.org

[18] Yang, Cathryn, "CAYang-blip/Tone-change: v1.0." Zenodo, Jan. 2023. doi: 10.5281/zenodo.7509125.

[19] Y. Xu, "Contextual tonal variations in Mandarin," *J. Phon.*, vol. 25, no. 1, pp. 61–83, Jan. 1997, doi: 10.1006/jpho.1996.0034.

[20] X. Chen, "Qing zhonghouqi zhi Minguo chuqi Beijing hua de shengdiao bianhua (清中后期至民国初期北京话的声调变化) (Tone change in Beijing dialect from mid-late Qing to early Republic)," *Zhongguo Yuwen*, vol. 2018, no. 04, pp. 435–445, 2018.

[21] M. Endo, "Bidirectional change in tone: Evidence from Chinese," *Econ. Rev.*, vol. 12, pp. 1–29, 2020.

[22] J. Kuang, "Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice," *J. Acoust. Soc. Am.*, vol. 142, pp. 1693–1706, Sep. 2017, doi: 10.1121/1.5003649.

[23] J. Zhang, *The effects of duration and sonority on contour tone distribution — A typological survey and formal analysis*. London: Routledge, 2002.

[24] J. Hanssen, J. Peters, and C. Gussenhoven, "Responses to time pressure on phrase-final melodies in varieties of Dutch and West Frisian," *J. Phon.*, vol. 93, p. 101150, Jul. 2022, doi: 10.1016/j.wocn.2022.101150.

[25] C. Yang and Y. Xu, "Crosslinguistic trends in tone change: A review of tone change studies in East and Southeast Asia," *Diachronica*, vol. 36, no. 3, pp. 417–459, 2019.

[26] J. Gandour, S. Potisuk, S. Dechongkit, and S. Ponglorpisit, "Tonal coarticulation in Thai," *J. Phon.*, vol. 22, pp. 477–492, 1994.

[27] Y. Xu, "Timing and coordination in tone and intonation—An articulatory-functional perspective," *Lingua*, vol. 119, no. 6, pp. 906–927, Jun. 2009, doi: 10.1016/j.lingua.2007.09.015.