Tones of Beijing dialect since 1900 and their evolution

Evidence from early recordings

Tianheng Wang
Nankai University

Since the discovery of real-time tone evolution in Thai and Chinese dialects, the existence of circular tone shift has been largely confirmed. The clockwise shift (high > rising > low > falling > high) has been found to be more prevalent than the counterclockwise shift (high > falling > low > rising > high). This study extracted tone patterns from five early recordings of citation tones in Beijing dialect dating from 1900 to the 1930s, and performed a quantitative acoustic analysis of tone evolution of Beijing dialect by comparing tonal differences between early recordings, early experimental results, and modern data. A clockwise tone shift in Beijing dialect since 1900 is confirmed: T1, high convex falling > high level; T2, high convex rising > mid concave rising; T3, low rising > low falling-rising, with intensified creakiness; and T4, lower falling > higher falling. By integrating these results with earlier documentation, this study uncovers a nearly completed circle of the fast-paced tone shift: the tone value of each tone has almost shifted to the value of its downstream tone in the clockwise direction over approximately 200 years. Two possible motivations for tone evolution in Beijing dialect are identified: truncation-like effects coupled with hypocorrection, and push/drag effects between neighboring tones. These findings provide concrete evidence for clockwise tone shift and contribute to a better understanding of tone evolution.

Keywords: tone evolution, Beijing dialect, early recordings, clockwise tone shift

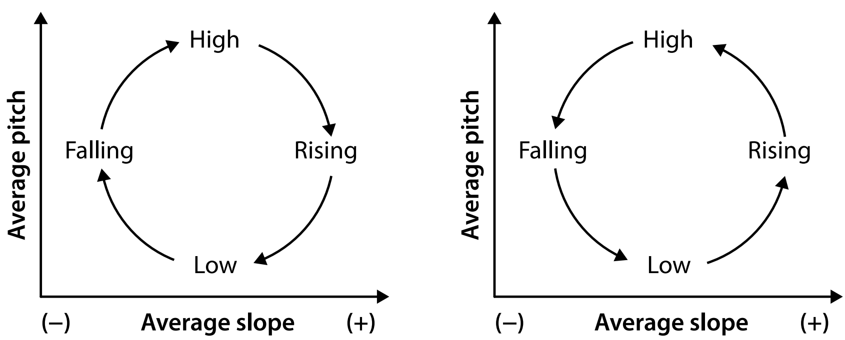
https://doi.org/10.1075/lali.00243.wan | Published online: 5 August 2025
Language and Linguistics 26:4 (2025), pp. 752–788. ISSN 1606-822X | E-ISSN 2309-5067

1. Introduction

1.1 Hypotheses on tone evolution

Through examining the differences between early documentation and modern fieldwork, linguists have become aware that speech sounds can evolve at an astonishing rate. Tones may evolve as rapidly as vowels and consonants, but our understanding of the typology and the mechanism of tone change is still limited compared to that of vowels and consonants (Endo 2015; Pittayaporn 2018; Yang & Xu 2019).

Wang's (1967) formalization of the circular tone sandhi rules in Xiamen Min with his newly proposed distinctive features of tones led to early attempts to analyze tone evolution and reconstruct tone values of earlier Chinese dialects based on modern dialects, e.g., tonal reconstructions of Proto-Southern Min (Ting 1982) and of Proto-Mandarin (Hirayama 1984). Moreover, Hirayama (1984) formally proposed the circulation hypothesis of tone chain shift: high > falling > low > rising > high, which Zhu et al. (2015) termed the counterclockwise shift (Figure 1b).[1] Mei (1977) also proposed an equivalent circular chain shift for the tone evolution of Beijing Mandarin since the 16th century.



a. b.

Figure 1. Clockwise tone shift (a) and counterclockwise tone shift (b) according to the tone pattern (tonal space) visualization scheme in §2.2.3

1. The use of the terms *clockwise* and *counterclockwise* is based on the convention that the positive direction of the average slope axis (Figure 1) is to the right (e.g. Zhu et al. 2015; Shen 2018; Wang 2021). If the positive direction is to the left (e.g. Hirayama 1998; Endo 2015), the meanings of these two terms are swapped.

However, the counterclockwise direction, despite serving as the foundation of Hirayama's later works on tone evolution and reconstruction (e.g. Hirayama 1998), was largely questioned by the groundbreaking discovery of tone change in Bangkok Thai. Pittayaporn (2007) and Zhu et al. (2015) both revealed that the tones of Bangkok Thai have undergone a clockwise shift over the past 100 years, which can be formulized as high > rising > low > falling > high (Figure 1a). The clockwise tendency is also found in Chinese dialects, e.g., Chongqing dialect (Liang & Meng 2013; Ming & Zhang 2015). Regardless of direction, the general hypothesis of circular tone shift is largely confirmed.

Subsequently, Endo (2015) investigated the tone evolution through dozens of Chinese dialects spanning 150 years of documentation and found multiple instances suggesting that tone shifts can occur in opposite directions in different dialects. Zhu & Li (2016) also suggested indirect evidence of the coexistence of bidirectional tone shifts in Hakka, based on dialect comparison. To further investigate the bidirectionality of tone change, Yang & Xu (2019) reviewed 52 tone change studies in East and Southeast Asia and revealed that changes in the clockwise direction are the most common, accounting for 56% of the reported tone changes, while changes in the counterclockwise direction account for only 9% of the cases.

Numerous mechanisms have been proposed to explain the motivation and the directional bias of tone change (Pittayaporn 2007; Ming & Zhang 2015; Zhu & Li 2016; Pittayaporn 2018; Yang & Xu 2019; Burroni & Kirby 2023), while some of them are conflicting. Among these theories, truncation-like effects have been observed in certain languages, interpreted as the premature termination of tonal target approximation under time pressure (Yang & Xu 2019), or the reduction of the final tonal element at faster rates when it conflicts with the onset of the next tone (Burroni & Kirby 2023). These effects can generate tone changes in the clockwise direction as termed in this paper. However, planning and implementation mechanisms for tone remain inadequately understood, leaving numerous questions unanswered.

1.2 Tone evolution in Beijing dialect

In contrast to general conjectures and discussions regarding the directionality of tone evolution, the trend of tone evolution in Beijing dialect can be carefully examined in detail due to its unique position among Chinese dialects. The findings are expected to be as fruitful as the study of Bangkok Thai.

The Beijing dialect has become the *de facto* common language (i.e. *lingua franca*) throughout China since the 1840s (Xiao Chen 2018), so there has been a significant amount of documentation of Beijing dialect by foreigners since the mid-19th century. By summarizing early literature, Endo (2015) and Xiao Chen (2018) discovered that for citation tones, T1 (*yīnpíng* 陰平, marked by a macron: *ā*) once had two variants of high level (55/44) and high falling (53/441), but now only the level variant survives; T2 (*yángpíng* 陽平, marked by an acute accent: *á*) remains high rising (e.g. 35); T3 (*shǎng* 上, marked by a caron: *ǎ*) has evolved from low rising to low falling-rising (13/113 > 214/313); and T4 (*qù* 去, marked by a grave accent: *à*) has undergone a change from low falling to full falling (31 > 51). Moreover, Zeng (2021) noticed that T2 in Shu's (1930) recording was evidently higher than today's T2.

By simplifying and plotting these discovered tone changes (Figure 2), an incomplete clockwise shift can be clearly observed. However, these findings are based only on early literature and one additional recording of citation tones (Shu 1930), thereby lacking sufficient acoustic data to ascertain the details of these tone changes. This study seeks to fill in the details by incorporating four newly noticed early recordings of citation tones, enabling a comprehensive acoustic analysis and a comparison between early and modern data.

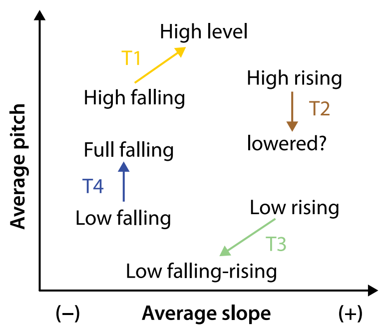


Figure 2. An illustration of tone shifts in Beijing dialect from the mid-19th century to the present, as discovered by previous studies (Endo 2015; Xiao Chen 2018; Zeng 2021)

2. Methods

2.1 Materials

2.1.1 Early recordings

During the 1900 Paris Exposition universelle, Léon Azoulay recorded a large collection of speech and music from all over the world. Among them, there are recordings of citation tones of Beijing dialect (Azoulay 1900) beginning with the announcement *Chinois, dialecte de Pékin* 'Chinese, Beijing dialect', after which prepared syllables were read in the order of the 4 tones. The speaker was Wenhui 文惠 (c. 1877–?), a Beijing native who was sent to study in France by the Qing government (Jia 2024). His demonstration of tones was recorded twice by Azoulay, on July 17 and October 15, 1900 (Jia 2024), with the second take including more syllables (Table 1).

Table 1. Number of syllables uttered in each recording*

	T1 to T5	T1 to T4	T2 to T4	Total
Azoulay (1900), 1st take		7 × 4 (28)		28 (28)
Azoulay (1900), 2nd take		75 × 4 (297)		300 (297)
Wang (1920)	20 × 5 (0)	45 × 4 (148)	1 × 3 (3)	283 (151)
Chao (1922a)	13 × 5 (0)	24 × 4 (95)	1 × 3 (3)	289 (98)
Shu (1930)		20 × 4 (80)		80 (80)
Pai (1933)		35 × 4 (140)	1 × 3 (3)	143 (143)

* Syllables are grouped by the available tones for each syllable. Numbers in parentheses are the number of syllables actually used in this paper. In particular, the syllable *er* is only available under T2 to T4 (i.e. no word exists for T1 *ēr*), so *er* is separated into a new column "T2 to T4".

In the 1920s, as part of the promotion of the Old National Pronunciation (*Lǎo Guóyīn* 老國音), two recordings (Wang 1920; Chao 1922a) of National Pronunciation courses were released. The Old National Pronunciation features 5 tones: T1 *yīnpíng* 陰平, T2 *yángpíng* 陽平, T3 *shǎng* 上, T4 *qù* 去, and T5 *rù* 入. While a standard for actual tone values was never established during the entire Old National Pronunciation era, both recordings utilized the tone values of T1 to T4 as spoken in Beijing (Li 1934: 101–107), so they serve as valuable materials. In their demonstration of citation tones, there are both syllables that can and cannot be combined with T5; only syllables without T5 tones are used in this study (Table 1).

Pu Wang 王璞 (1875–1929), the speaker of the first recording (Wang 1920), was from Wanping 宛平, a town situated 16 km southwest of the center of Beijing. Although the dialect in Wanping belongs to the Beijing urban dialect,[2] one may argue that a slight tonal difference exists. However, Wang was a pioneering figure in teaching and promoting the Beijing pronunciation (Li 1934: 25, 35), and his demonstration should therefore be considered a good specimen.

2. My brief field investigation among middle-aged and elderly residents in Wanping has affirmed that their accents do not show discernible differences from those of individuals of the same ages in central Beijing, except that the zero onset can be realized as [ŋ] by elder Wanping residents.

The second recording (Chao 1922a) features Yuen Ren Chao 趙元任 (1892–1982) as the speaker. Strictly speaking, Chao is not an ideal subject for studying Beijing dialect as he was not originally from Beijing, and his Beijing dialect, although spoken since childhood at home, is not authentic (Chao 1971). Nevertheless, the existence of his recording is valuable due to the scarcity of the recordings of citation tones.

At the same time, Lao She 老舍 (1899–1966) recorded the *Chinese* volume of the Linguaphone Language Courses (Shu 1930)[3] in London, including demonstrations of citation tones. Born and raised in a Manchu family in Beijing, he was an authentic Beijinger.

In 1932, the New National Pronunciation (*Xīn Guóyīn* 新國音), whose pronunciation was entirely based on the Beijing dialect, replaced the old one. What followed was the record (Pai 1933) by Ti-chou Pai 白滌洲 (1900–1934), who was also an authentic Beijinger of Mongolian ethnicity.

The Paris Expo recordings were on wax cylinders, while the rest were on phonograph records. They were digitized through specific projects (see References and Appendix B for details). Some syllables are inaudible under noise and have been discarded (Table 1). Additionally, I examine tones in connected speech from two other early materials of Beijing dialect, Li & Zhang (1909)[4] and Wang (1956),[5] as a supplement.

2.1.2 Early experiments

In addition to the recordings, I also use five early studies that mapped the pitch contours of Beijing dialect as references for comparison.

One of the earliest plots of Beijing tones was Karlgren (1915: 253–259), whose subject was Lien-teng Ts'i, an employee at *Teou-fou Kong-sseu* ('Tofu Company') near Paris, who was born and raised in a Beijinger family in Beijing.

Chao (1922c) reported the semi-quantitative result of his experiments on tones of several Chinese dialects, including Beijing dialect.

Meanwhile, the famous work of Liu (1925) recorded detailed pitch contours of Beijing dialect, with Mien Tcheng 陳綿 (1900–1968) as the speaker. Although

3. The release date was not specified in the record box set. Xiao Chen (2018) and Bickers (1994) attribute the release year to 1930, and this dating is adopted in this paper.

4. The speaker in the recording examined in this study is Deyang Li 李德錫 (1881–1926), a Xiangsheng 相聲 artist from Beijing. While previous sources suggest that Li & Zhang (1909) was recorded and released in 1908, recent research based on newspapers from the time confirms that it was recorded and released in 1909 instead (Wang & Du 2016).

5. This is a documentary film about the Shenyang trials of 1956. The speaker in the film examined in this study is Puyi Aisin-Gioro 愛新覺羅·溥儀 (1906–1967), the last emperor of China.

Tcheng's family was originally from Fujian, Tcheng was born and raised in Beijing. Lin (1992) certified his pronunciation as authentic Beijing dialect. For each tone in T1 to T4, six pitch contours were measured (Liu 1925: Pl. XI). Characters in Middle Chinese *rù* tone were also measured, but were found to have disappeared in Beijing dialect and merged into T1 to T4 (Liu 1925: 82–86). Therefore, I shall not include the results of *rù* tone characters.

Pai (1934) adopted the same method as Liu (1925) and collected his own pronunciation of the 4 tones. The posthumously published work does not include original pitch contours, but only a schematic representation of the average of all contours using lines. However, it is sufficient for reference purposes.

Finally, Obata & Tesima (1934) recorded the pitch contours of 4 speakers of Beijing dialect, among whom Xiangyin Bao 包象寅, a Chinese teacher at Tokyo University of Foreign Studies, was the oldest and spoke the purest Beijing dialect. I shall extract Bao's data from figures of pitch contours in Obata & Tesima (1934).

Table 2. Number and unit of pitch contours of Beijing tones in early studies

	Number	**Unit**
Karlgren (1915)	1 × 4	1 / Hz
Chao (1922c)	1 × 4	Semitones
Liu (1925)	6 × 4	Log Hz
Pai (1934)	1 × 4 (averaged)	Semitones
Obata & Tesima (1934)	2 × 4	Hz

In addition, Bradley (1915) and Bröring (1927: 50–51) also recorded the pitch contours of Beijing tones (see Endo 2015), but their results were too imprecise to be included. Later, Lin conducted tests on two speakers in 1965 (Bao & Lin 2014: 175) and their average pitch contours were similar to Chao (1922c), and are therefore not included for brevity.

2.1.3 Modern data

In order to reflect the tone characteristics of the modern Beijing dialect, I incorporate two additional data sources as a complement to the earlier data.

One material is from Shi (2007), who extracted citation tones of Beijing dialect in the series *Xiandai Hanyu Fangyan Yinku* (現代漢語方言音庫 'Phonetic Database of Modern Chinese Dialects') (Lin et al. 1998). The speaker (and one of the authors) of the Beijing dialect volume, Yimin Zhou 周一民, is a Beijing native born in 1950.

Further material is from Sanders & Shi (2003), a speech corpus of monosyllables spoken by Beijing speakers collected in 2003. I extract pitch data of four randomly selected male speakers from the corpus. They were born between 1979 and 1983.

All speakers in §2.1 are male. Although gender consistency enhances the comparability between recordings, it also underscores the lack of female subjects in the historical recordings. Therefore, we only have an incomplete picture of what Beijing dialect sounded like in earlier times.

2.2 Data processing and analysis

2.2.1 Extraction

The raw audio remastered from early recordings is with more or less noise (e.g., Figure 3a), which does not distort pitch information (fundamental and harmonics), but can interfere with the fundamental frequency extraction process. Therefore, I initially applied noise reduction in Adobe Audition. Noise samples and reduction levels were carefully chosen to minimize the loss of linguistic information while greatly reducing the background noise (Figure 3b). Subsequently, I extracted the fundamental frequency (F_0) of each syllable in all materials in Praat (Boersma & Weenink 2022) using the built-in autocorrelation method. Finally, I evenly distributed 30 measurement points over the duration of each syllable and extracted 30 F_0 values per syllable. Duration data were also collected.

For data from early experiments with Beijing tones, I first traced the contours in the scanned image into vectorized curves in CorelDRAW, and then extracted pitch data (in Hz or semitones; see Table 2) of 30 measurement points on the curves using the Pillow library (version 9.2.0)[6] in Python.

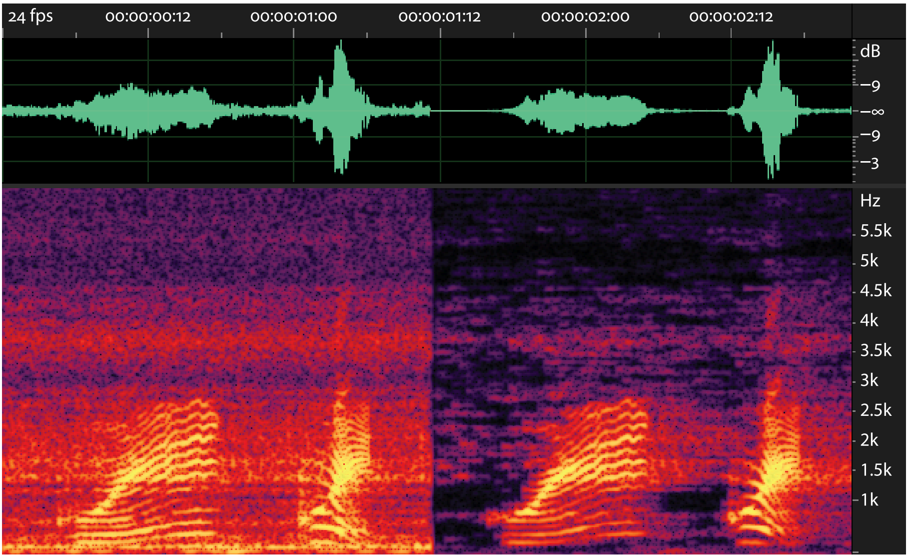
2.2.2 Normalization

To eliminate individual variations and enable between-speaker comparison, normalization of the F_0 is needed. In this study, I use the T-value (Shi 1986) for normalization. The original definition of the T-value is in (1).

(1) $$T_x = \frac{\lg F_{0,x} - \lg F_{0,\min}}{\lg F_{0,\max} - \lg F_{0,\min}} \times 5$$

Where $F_{0,x}$ denotes the average fundamental frequency of measurement point x for a speaker; $F_{0,\max}$ and $F_{0,\min}$ denote the upper and the lower limits of average frequencies of all measurement points for the speaker, respectively; and T_x, the

6. Available at Zenodo: ⟨https://doi.org/10.5281/zenodo.596518⟩.



a. b.

Figure 3. Spectrogram of *guǐ* and *guì* from Azoulay (1900) before (a) and after (b) noise reduction

resulting T-value, represents the reference tonal value on the five-point scale for point *x*. The calculated T-value ranges from 0 to 5.

When performing statistics on large samples of multiple speakers, the standard deviation of each measurement point is often added/subtracted to the average F_0, e.g. (2), in order to prevent the extremum points from being compressed, since a tone varies and occupies a band rather than a zero-width line in the tonal space (Shi & Wang 2006a; Liang & Meng 2013).

(2) $$T_x = \frac{\lg F_{0,x} - \lg(F_{0,\min} - \sigma_{\min})}{\lg(F_{0,\max} + \sigma_{\max}) - \lg(F_{0,\min} - \sigma_{\min})} \times 5$$

Where σ_{max} and σ_{min} represent the standard deviation of F_0 values at the maximum and minimum measurement points, respectively.

However, this adjustment is not effective in this study since each early recording was uttered by only one speaker. Thus, I introduce (3) as an approximation of (2), by fixing the standard deviation to 0.5 and the range of T-values from 0.5 to 4.5.[7]

7. The approximation is based on the fact that the averages of standard deviation in T-value are generally around 0.4 or 0.5 for a large sample of speakers (e.g. Shi & Wang 2006a; Liang & Meng 2013).

(3) $$T_x = \frac{\lg F_{0,x} - \lg F_{0,\min}}{\lg F_{0,\max} - \lg F_{0,\min}} \times 4 + 0.5$$

Pitch contours from early experiments are also converted into 30 T-values using (3), to ensure its comparability to T-values extracted from recordings.

2.2.3 Visualization

In East and Southeast Asia languages, average pitch and pitch slope across a syllable are identified as two primary factors in the perception of the tone of the syllable (Gandour 1983). Convexity is also an essential factor in characterizing a tone (Andruski & Costello 2004). Therefore, average pitch, average slope, and convexity can be used to depict a tone.

Based on these three parameters, calculated as in (4) in this study, I adapt the quantitative visualization schemes of tone patterns by Wang (2021) and Shen (2018) to plot the tone evolution directly. Duration is also a valuable factor, but for simplicity it is omitted from the visualization and will be presented separately.

(4) $$\text{Average pitch} = \frac{\sum_{i=1}^{n} T_i}{n}$$

$$\text{Average slope} = \frac{T_{n-1} + T_n}{2} - \frac{T_1 + T_2}{2}$$

$$\text{Convexity} = \frac{\sum_{i=1}^{n} T_i}{n} - \frac{T_1 + T_2 + T_{n-1} + T_n}{4}$$

Where T_i is the calculated T-value at measurement point i, and n is the total number of measurement points ($n=30$ in this study). When using the start or end point, I choose the average T-value of the first/last two points to avoid random fluctuations at syllable boundaries. Convexity follows the definition in Wang (2021). For a better understanding, I graphically represent these three parameters in Figure 4.

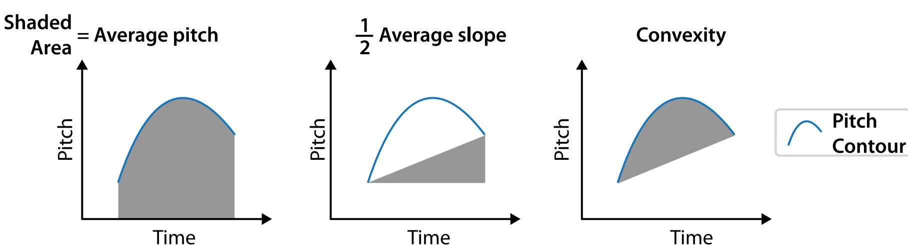


Figure 4. Graphical representations for the definitions of average pitch, (half of) average slope, and convexity

Hirayama (1998), Endo (2015), Zhu & Li (2016), and Pengfei Chen (2018) have also introduced similar visualizations of tone patterns using average pitch and slope, but without quantization.

3. Results

3.1 Pitch contours

We present the normalized pitch contours of the 4 tones across all materials in Figure 5, as well as the relative durations of them in Figure 6.

To maintain consistency and comparability, I selected the minimum of T3 as the minimum of the tone range across all studies, except Wang (1920), even if the minimum of T4 was lower in some cases. In Azoulay (1900) and Wang (1920), this behavior of T4 may result from reading syllables in groups from T1 to T4, causing the T4 syllable, being the last in the group, to be prolonged and fall deeper. For Wang (1920), I selected the first quartile of the minimums of T3 and T4, since his T4 was too low. Similarly, T4 was also excluded when selecting the maximum of the tone range for Obata & Tesima (1934), as they mentioned that T4 was over-produced.

Several recent studies (e.g., Kuang 2017; Sun & Shih 2021) regarding citation tones of Beijing dialect also reported tone patterns consistent with those observed in Sanders & Shi (2003) in this study. Therefore, Sanders & Shi (2003) is representative of modern Beijing tones.[8]

8. However, this is somewhat questionable for the contemporary Beijing tones, as we suspect that there have been slight tone changes over the last 20 years. See §5.

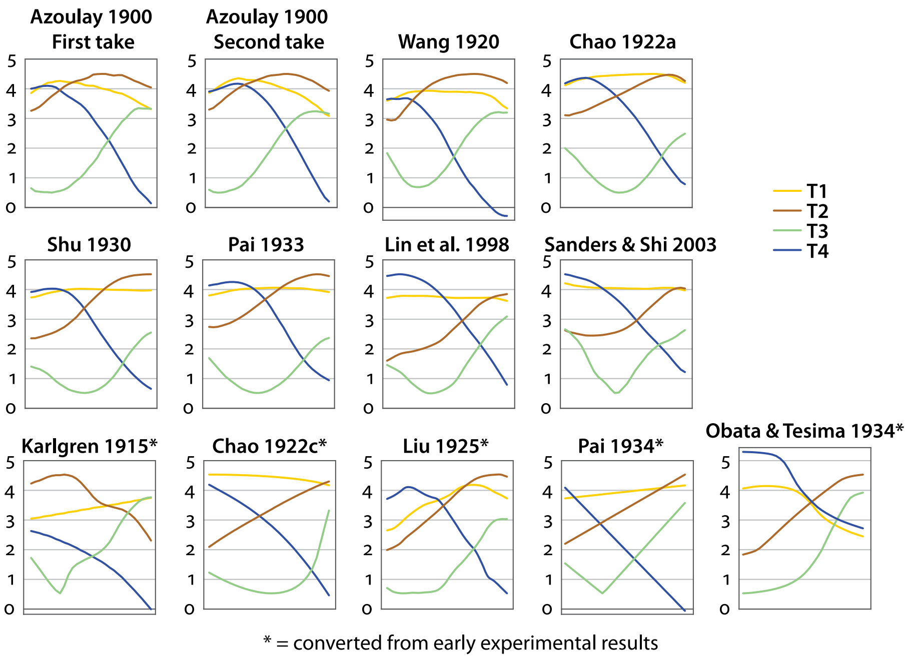


Figure 5. Pitch contours (in T-value) extracted from all recordings and converted from early experimental results

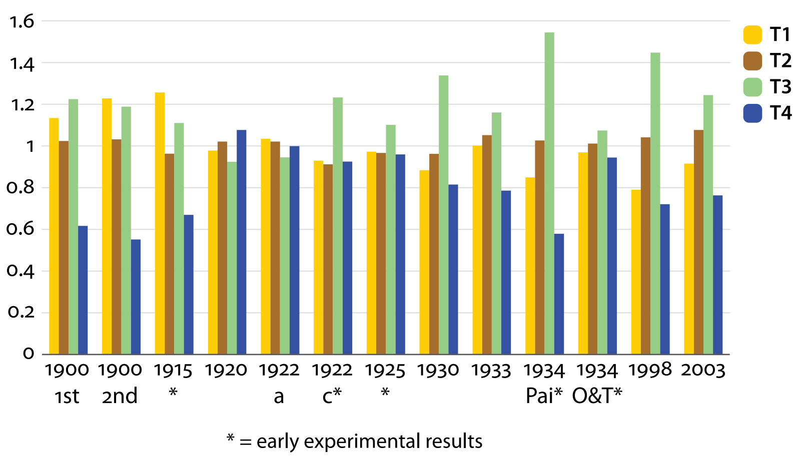


Figure 6. Ratio of the average duration of each tone to the average of the average durations of all tones[9]

9. Note that Wang (1920) and Chao (1922a) read syllables to a fixed beat, so their duration data are uninformative. For absolute durations, see Appendix A.

3.2 Tone patterns

Before examining the individual tones in detail, we also visualize all tones in Figure 7 to better demonstrate the tone pattern differences among these recordings.

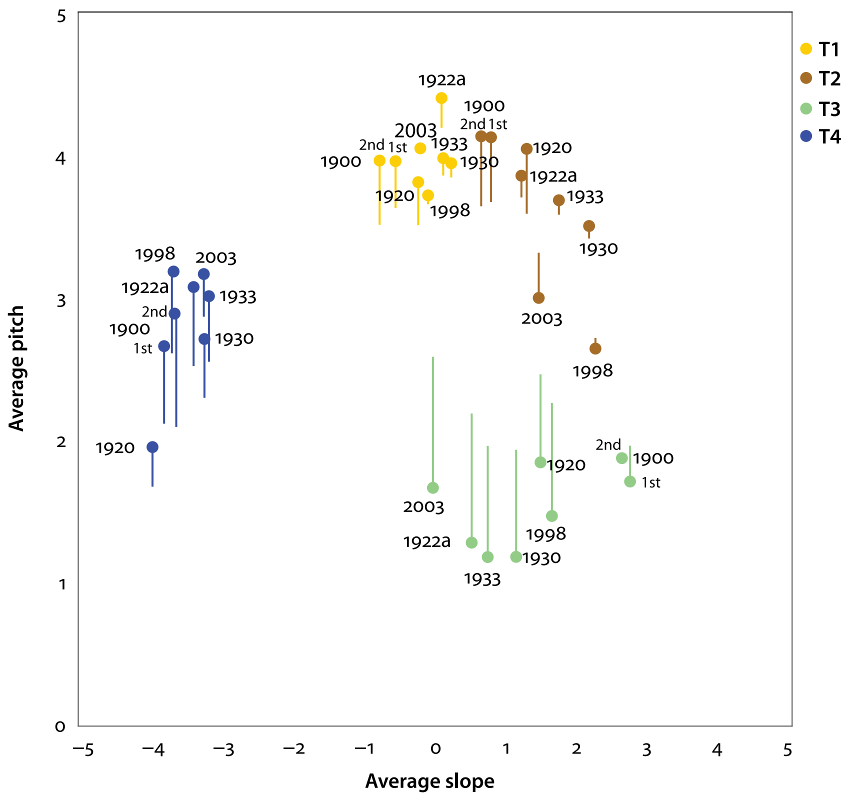


Figure 7. Visualization of tone patterns of all recordings[10]

3.3 Examination of each tone

To facilitate the observation and discussion of tone evolution, we categorized the speakers into four groups according to date of birth (Table 3).

10. The length and direction of bars stretching from the dots represent convexity, as defined in §2.2.3. A bar pointing downward indicates a convex tone, and a bar pointing upward indicates a concave tone. All values are in T-value, or equivalently, the five-point scale.

Table 3. Groups of speakers

	Birth year of speakers	Recordings
Group 1875	1875–c. 1877	Azoulay (1900), Wang (1920)
Group 1900	1892–1900	Chao (1922a), Shu (1930), Pai (1933)
Group 1950	1950	Lin et al. (1998)
Group 1980	1979–1983	Sanders & Shi (2003)

3.3.1 T1

The speaker of Azoulay (1900) pronounced T1 as a slightly falling tone, while other speakers pronounced T1 as a level tone in citation form (slight pitch bends at the beginning and the end can be ignored). This difference is also clearly shown in Figure 7, where the two points of Azoulay (1900) are the leftmost in T1. Edkins (1857: 17; 1864: 98) described both the falling and the level variants of T1.

In order to examine the variation of T1, we here introduce a concept, falling excursion, defined as the average of the two highest pitches (located near the beginning of the contour of T1) minus the average of the two lowest pitches (at the end of the contour) across 30 measurement points, in semitones.

Figure 8 presents the detailed variation of falling excursions of T1 by the speaker of Azoulay (1900), showing that his T1 is realized as a continuum from a near-level tone to a falling tone. More specifically, this can be described as a convex falling tone or level-falling tone, where the degree of falling is largely free to vary.

The falling variant of T1 disappeared in citation form in our materials later than Azoulay (1900), except for Obata & Tesima (1934). However, it had not yet disappeared in connected speech in the early 20th century. Shu (1930) made extensive use of the (convex) falling form of T1 in speech, although his T1 was fairly level in citation form. One may argue that this is simply due to anticipatory assimilation when T1 is followed by T2 or T3, both of which start lower than T1. However, the existence of such an anticipatory assimilation effect in Beijing dialect is controversial, as disputed by Sun & Shih (2021); furthermore, a few tokens of Shu's T1 preceding another T1 of the same height are still found to be falling. Besides, Shu's T1 frequently became an explicit falling tone when followed by a neutral tone, as already noticed by Endo (2015) and Xiao Chen (2018). Figure 9 provides a good example of these two variants, where two T1s (*xiān* and *jīn*) become falling before neutral tones and one T1 (*sān*) is falling before another T1.

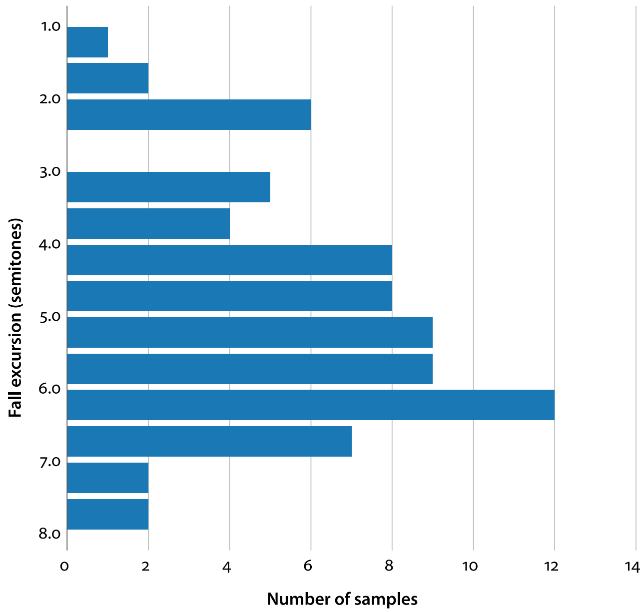


Figure 8. Histogram of fall excursions of 75 samples of T1 in Azoulay (1900), second take[11]

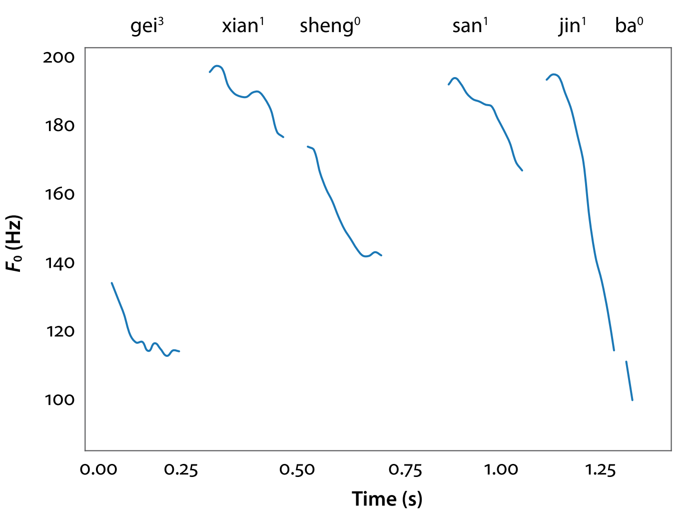


Figure 9. Pitch contours of *gěi xiānsheng sānjīn ba* 給先生三斤吧 'I shall give you three catties' from Lesson 17 of Shu (1930)

11. Lower bars indicate deeper falls. The mean is 4.99 semitones and the median is 5.16 semitones. By comparison, fall excursions of T1 by Wang (1920) have a mean of 2.51 semitones and a median of 2.22 semitones.

3.3.2 T2

Following the speaker grouping in Table 3, we can observe a clear evolutionary trajectory of T2 in Figure 10.

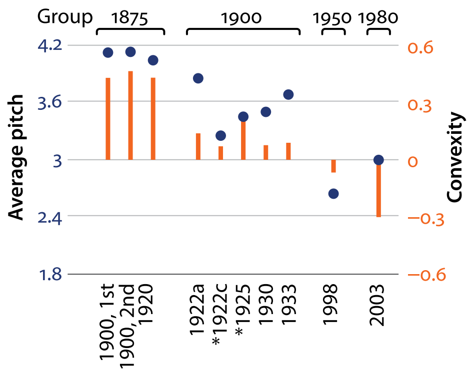


Figure 10. Average pitch (blue dots) and convexity (orange bars) of T2 of 4 groups in T-value

T2 of Group 1875 is obviously a convex rising tone, similar to the horizontal flip of the convex falling T1 of Azoulay (1900). This is different from T2 of Group 1980, which is a concave rising tone, frequently reported in the literature (e.g. Shi & Wang 2006a; Bao & Lin 2014: 174). As intermediates, Group 1900 exhibits a slightly convex T2, while Group 1950 (although containing only one speaker) exhibits a slightly concave T2.

Along with the decrease in convexity, the average height of T2 also decreases across four groups. T2 used to be the highest tone as in Group 1875, as Karlgren (1918: 31) noted "*Hia p'ing shêng* [T2] is generally somewhat higher than *shang p'ing shêng* [T1]". In later groups, the average pitch of T2 is lower than that of T1. However, except for Chao (1922a), the endpoint of T2 remains higher than T1 in Group 1900, as Zeng (2021) noted for Shu (1930).

The synchronous declining of average pitch and convexity of T2 is evident, e.g. 455 > 355 > 35 > 335 > 224.[12] Notably, the average pitch of T2 of Group 1980 is more conservative than that of Group 1950.

We also note the falling end of T2 in Group 1875, and furthermore, T2 with the falling part as the main part in Karlgren (1915: 257). Edkins (1857: 16; 1864: 98–99) described two variants of T2, which can be transcribed as 35 and 353 (Endo 2015). Karlgren (1918: 24) further elaborated that:

12. In this paper, I use three-digit tone values to describe convex/concave tones, not implying longer duration. For example, 355 indicates a convex 35, while 335 indicates a concave 35, etc.

> Many speakers produce it [T2] as a high-keyed falling tone (opposed to the *k'ü sheng* [T4] as a low-keyed falling one). [...] Others, perhaps the majority, produce it as a rapidly rising tone. [...] This surprising double nature of a tone inside the very city of Peking has been already observed by Edkins.

Additionally, we notice that in recitation-styled verses in Li & Zhang (1909), T2 is often clearly rising-falling. The syllable *chún* in Figure 11 is a distinct example, resembling Karlgren's (1915: 257) contour.

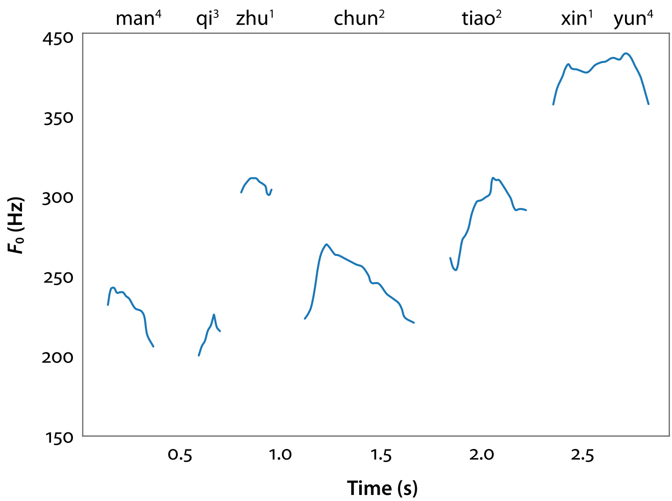


Figure 11. Pitch contours of *mànqǐ zhūchún tiáo xīnyùn* 慢啓朱唇調新韻 'slowly open red lips and sing a new melody' from Li & Zhang (1909)[13]

This rising-falling *chún* might be a result of recitation, which requires an elongation and declination at the end of the tone (described as an intonational particle by Chao 1968: 813–814); alternatively, it could also be a conservative form preserved during recitation. Given Karlgren and Edkins's familiarity with colloquial Beijing tones at the time, the rising-falling form they documented could be authentic, as an earlier citation form of T2.

3.3.3 T3

The speaker of Azoulay (1900) pronounced T3 as a pure rising tone with little or no concavity, while others pronounced T3 as a falling-rising tone, or termed as the dipping tone (Figure 12). This difference can also be seen in Figure 7 by com-

13. The word *xīnyùn* was pronounced in falsetto, only for jocular effect in the performance, not a characteristic of linguistic tones.

paring Azoulay's convexity bars with other bars of T3. T3 tested by Liu (1925) and Obata & Tesima (1934) was also without falling onsets. Wang (1979) has noted the difference between the results of Liu's T3 and modern T3.

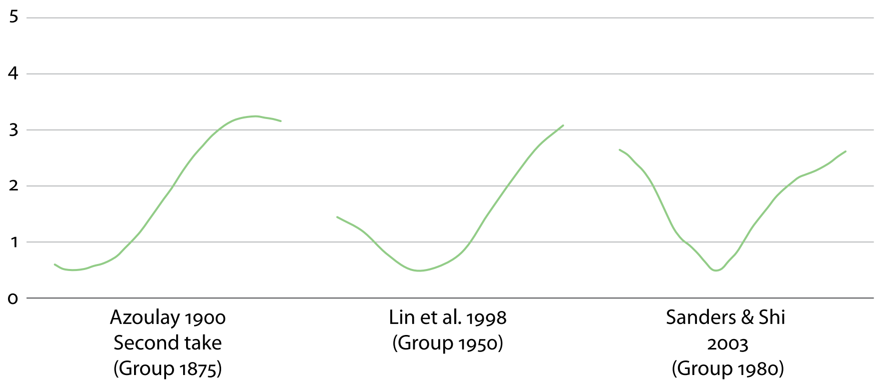


Figure 12. Three representative configurations of T3 in citation forms during the evolution excerpted from Figure 5

In the second take of Azoulay (1900), a total of 74 T3 syllables were uttered. They are not all pure rising: eleven of the 53 syllables with voiceless initials have an obvious falling F_0 at vowel onset, which can be attributed to the F_0 raising effect by voiceless consonants (Xu & Xu 2021), while none of the 21 syllables with voiced initials or zero initial have an obvious falling F_0 at voice onset. Therefore, the pure rising form is indeed the primary variant for Azoulay (1900).

Group 1980 also differs from earlier groups, since the beginning of their T3 has a higher pitch, approximately equal to the pitch at the end, while T3 of earlier groups ends in a higher pitch than it begins (Figure 12). Shi & Wang's (2006b) apparent-time study (i.e. by comparing the speech of individuals of different ages) showed results consistent with this study: T3 by old Beijing residents born before 1966 has the end higher than the beginning, while T3 by those born after 1966 has the end lower than the beginning.

The evolution of the citation form of T3 becomes clear: low rising > low concave rising > falling-rising, or 14 > 114 > 214 > 213 > 313. As T2 becomes lower and the beginning of T3 becomes higher, T2 and T3 are finally realized with the same pitch height at the onset (e.g. Groups 1950 and 1980).

We have also observed that in recordings of Li & Zhang (1909) and Puyi (Wang 1956), T3 can still be realized as low rising before other tones in a sentence (Figure 13). This is in stark contrast to the present-day realization of T3 as "half-T3" 21 before other tones, unless followed by a pause (Chao 1968: 27; Duanmu 2007: 236–238).

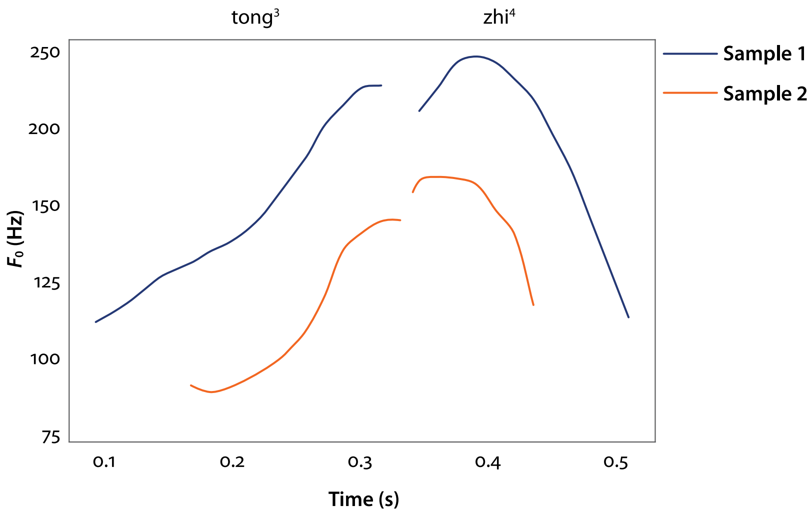


Figure 13. Pitch contours of two samples of *tǒngzhì* 統治 'to rule' by Puyi in Wang (1956)[14]

Modern scholars often suggest that the "essence" of T3 is low (low level by Wang 1979, and Shi & Ran 2011; low dipping by Zhu et al. 2012; see Cao & Wei 2016 for a broader review). In contrast, prior to 1900, the essence of T3 should not be low, as pure rising variants did appear in speech. I claim that the target of T3 has changed from rising to low since 1900. Therefore, "rising > low" could summarize the changes in both citation form and underlying form.

While our investigation primarily focuses on pitch height and contour, duration and phonation also play significant roles in characterizing T3. There is a notable increase in the duration of T3 accompanying the shift from rising to falling-rising (Figure 6 and Appendix A): in Groups 1950 and 1980, as well as in Shu (1930), the duration of T3 is significantly longer (> 50 ms) compared to the other tones; however, T3 in Azoulay (1900) does not exhibit a distinct longer duration. This aligns with the general trend that complex contour tones are often associated with longer durations (Hyman 2007).[15]

The creakiness in today's T3 is so intense that the F_0 contour often drops dramatically and disconnects in the middle part (Zhu et al. 2012; Kuang 2017).[16]

14. The second syllable *zhì* is partially unstressed, but not in the neutral tone.

15. Many thanks to the two anonymous reviewers for pointing out the importance of duration.

16. Less commonly, T3 can also be produced with breathy voice or slack voice, aside from creaky voice (Duanmu 2007: 228–233; Cao & Wei 2016; Kuang 2017), but such variants did not appear in our samples.

However, in the recordings of Groups 1875 and 1900, T3 is only occasionally accompanied by creaky voice, and the harmonics never disconnect, indicating that the creakiness is much weaker than that of today. Among all the monosyllabic demonstrations in Group 1950, only two out of 52 instances of T3 exhibit disconnected harmonics. In contrast, in samples of Group 1980, 10 out of 20 T3 syllables have harmonics completely disconnected in the middle of the syllable in the narrowband spectrogram.[17] In this case, the phonation type of the middle part is pure *creak* rather than *creaky voice* as distinguished by Laver (1980: 122–126).

It can be seen that the creakiness of T3 has progressively intensified, and the phonation is shifting from creaky voice to creak.[18] The trend of increasing creakiness was also observed in the apparent-time study by Liu et al. (2016): younger speakers are more likely to adopt creakiness in T3.

3.3.4 T4

The evolution of T4 is much simpler compared to that of the other three tones: its overall pitch keeps getting higher (Figure 14). The end of T4 reaches the lower bound of the tone range in Groups 1875 and 1900, whereas in Group 1980 it is between 1 and 2 by T-value, corresponding to 2 on the five-point scale. The beginning of T4 reaches the upper bound in Groups 1950 and 1980, but not in two earlier groups. We can formulate the evolution as 41 > 51 > 52. Its convexity remains unchanged (see Figure 7).

Earlier literature often described T4 as a low falling tone (e.g. Meadows 1847: 63; Rudy 1874: 6), transcribed as 31 by Endo (2015) and Xiao Chen (2018). Although the citation form of T4 of Group 1875 is not that low, low variants of T4 can be commonly heard in early recordings. For example, two T4 syllables *jùshì* in Li & Zhang (1909) (Figure 15) are considerably lower than T1 and T2.

Edkins (1857: 16–17; 1864: 98) also described a low falling-rising variant of T4, which can be transcribed as 313 (Endo 2015). I did not observe this variant in isolation throughout the recordings, but a tone sandhi is not rare: when two T4s

17. Note that the overall pitch (the tone range) of Group 1980 is also lower than that of previous groups (Appendix A), which can additionally aggravate the creakiness of their T3. Furthermore, limitations of early recording equipment also necessitated higher speaking efforts, which can lead to higher pitch and less creakiness. However, these factors are relatively minor and are unlikely to fully account for the substantial differences in creakiness between recordings.

18. The mid-glottal stop (a short period of complete glottal closure in the middle of the syllable, resulting in a moment of complete silence), which is an extreme product of creakiness amplification (Zhu et al. 2012), can even be heard in the middle of today's Beijing T3. Mid-glottal stop variants are also present in the corpus of Sanders & Shi (2003), but extremely rare, and the syllable samples used in this study do not encompass such variants. Therefore, I do not address the mid-glottal stop variant in the main text.

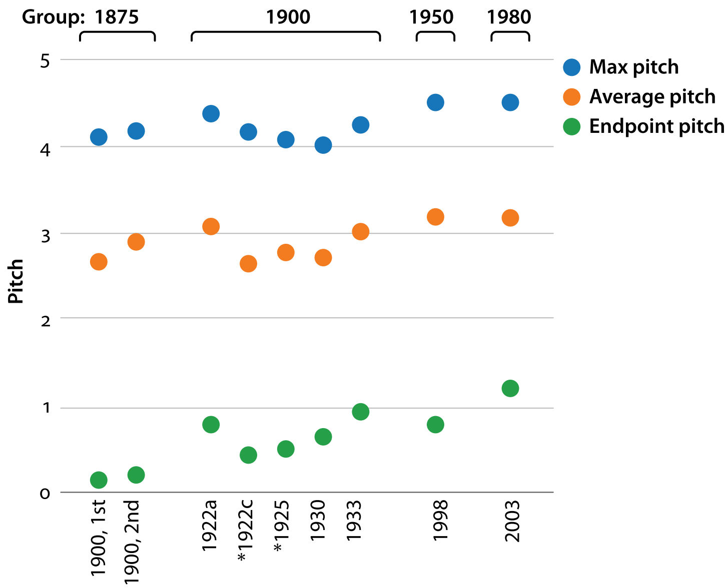


Figure 14. Max pitch, average pitch, and endpoint pitch of T4 of 4 groups in T-value[19]

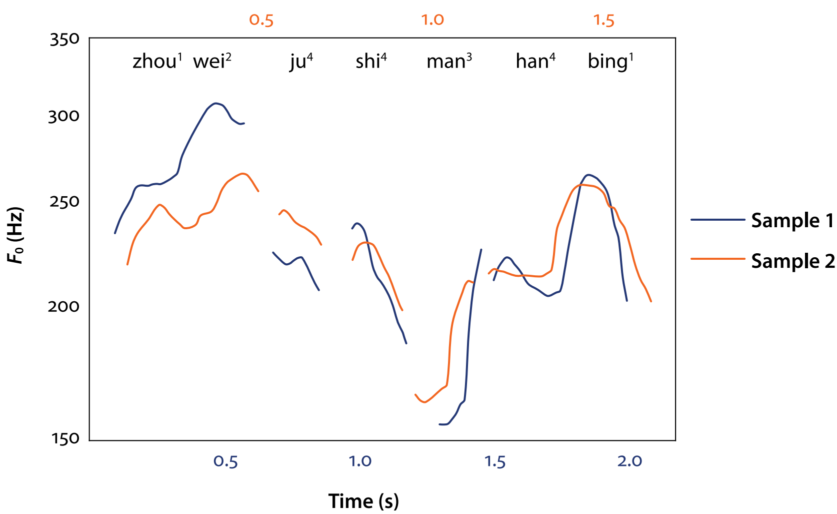


Figure 15. Pitch contours of two samples of *zhōuwéi jùshì Mǎn-Hàn bīng* 周圍俱是滿漢兵 'there are all Manchu and Han soldiers around' from Li & Zhang (1909)

19. Wang (1920) is removed due to his overlengthened T4.

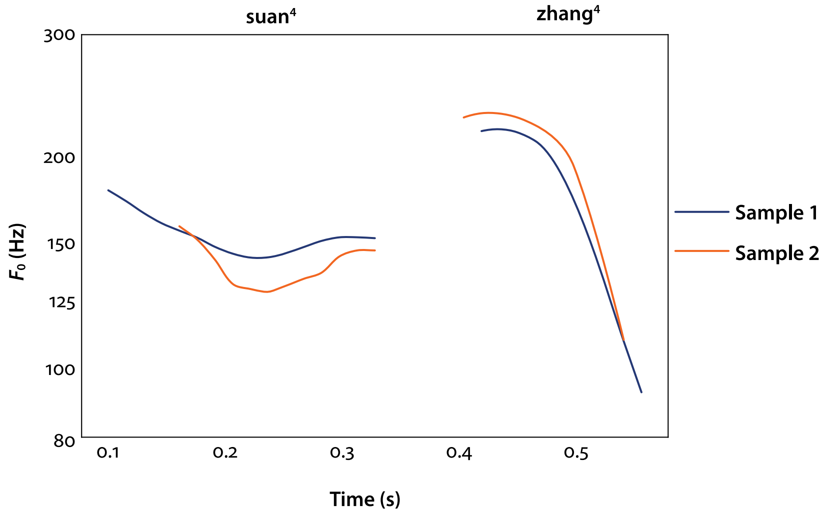


Figure 16. Pitch contours of two samples of *suànzhàng* 算賬 'to reckon account' from Lesson 14 of Shu (1930)

are adjacent and both syllables are stressed, the first T4 becomes an incomplete falling-rising tone, e.g. *suànzhàng* in Shu (1930) (Figure 16). This sandhi form may be a remnant of Edkins's 313, or it can also be interpreted as the first T4 with a higher end assimilated by the second T4. This sandhi is retained and evolves in the informal register of the Beijing dialect, and the sandhi form of the first T4 has now become a rising tone, similar or identical to T2 (Dong 2010; Higashi & Wang 2017), which can be interpreted as the height of the end of the first T4 entirely assimilated by the second T4 (Higashi & Wang 2017).

3.4 Summary of the evolution

With the results and analysis presented above, we can add arrows of the identified directions of tone evolution to the visualization (Figure 17). Moreover, we supplement the change pathways with dashed arrows based on our interpretation of earlier descriptions: (a) T1 was with greater decline before Azoulay (1900); (b) the 313 form of T4 recorded by Edkins (1857: 16–17; 1864: 98) was an earlier form. The tracks of the clockwise shift almost form a closed circle now (Figure 17). This provides a quantitative complement to Figure 2.

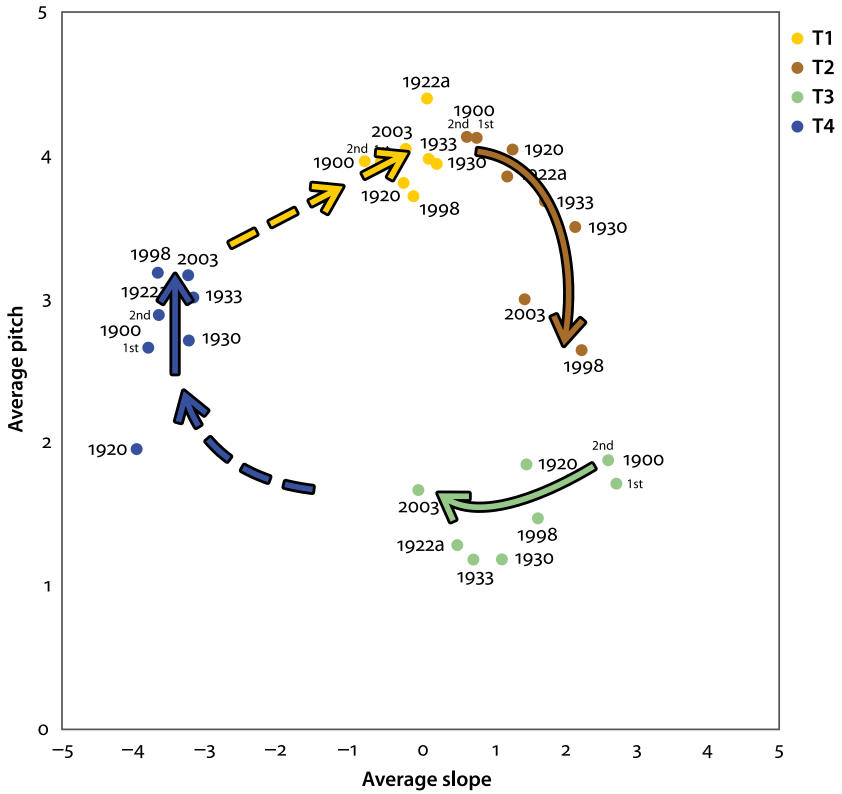


Figure 17. Evolution of the tone pattern of Beijing dialect since 1900

4. Discussion

4.1 Mechanisms of tone evolution

The evolution of each tone in the Beijing dialect, as observed in §3.3, can be summarized as follows: T1, HHM > HH;[20] T2, 455 > 355 > 35 > 335 > 224; T3, 14 > 114 > 214 > 213 > 313; and T4, (313 > 31 >) 41 > 51 > 52, with earlier literature-based forms included in parentheses. Based on these findings, I recognize two possible motivations contributing to the tone evolution of Beijing dialect from dual perspectives: truncation-like effects coupled with hypocorrection, and push/drag effects between neighboring tones. It is noteworthy that there is no consen-

20. H = high; M = mid. The height of the level part of T1 is generally between 4 and 5 on the five-point scale (Figure 5). To avoid bias in selecting numerals, I opt to represent T1 using H's and M's instead of numerals on the five-point scale. Similar to Footnote 12, HHM does not indicate a longer tone but denotes a high convex falling tone.

sus on the interpretation of the tone truncation-like process (see the end of §1.1), hence I describe the mechanism as "truncation-like effects", instead of explicitly as "truncation".

4.1.1 Truncation-like effects and hypocorrection

Edkins's (1864: 98) documentation of T1 serves as our entry point to the truncation-like process. He described a complementary distribution: T1 is high falling in citation form or at the end of a sentence, and becomes high level in the middle of a sentence (see also Endo 2015). The complementary distribution may not be absolute, as our observation is that the two variants were largely in free variation in 1900 in citation form (Figure 8), but the tendency here is reasonable; i.e. T1 tends to be realized as HHM in citation form or at the end of a sentence, and HH in the middle of a sentence.

In connected speech, the underlying tone contour tends to simplify itself by eliminating a tonal element at the syllable margin (Lin 1998), with the final element being the most likely to be deleted (Pittayaporn 2007, 2018; Yang & Xu 2019; Burroni & Kirby 2023). Therefore, HHM > HH tends to occur in the middle of a sentence, especially at a faster speech rate. Eventually, the frequently occurring truncated variant, HH, became the underlying target of younger generations through hypocorrection, wherein younger listeners reinterpreted the truncated variant as originally intended and solidified it (Ohala 1981; 1993). Additionally, the paradigm leveling between the citation form and the variant in speech could also be attributed to the preference for Paradigm Uniformity (Steriade 2000).

Skeptics might reversely suspect that the underlying form of T1 was HH in 1900 and the falling tail at the end was accessional and non-phonemic. Indeed, a similar phenomenon is observed in dialects such as Rongcheng 榮成 Mandarin and Qiyang 祁陽 Xiang, where, as a sublinguistic behavior, all tones have obligatory accessional falling tails in citation form (Lai et al. 2017). However, such suspicion is probably unfounded for Beijing since the HHM variant did exist in the middle of a sentence and can be noticeable as the example in Figure 9.

This suspicion resembles the hypocorrection process, in which younger generations failed to normalize the falling tail as a part of the underlying form but rather perceived it as unintended (e.g. misinterpreted as a product of certain intonation or a boundary marker), leading to its deletion. Chao (1922b: 22) actually reported the falling variant but seems to have made such a misperception:

> 'In Beijing, there is another witty pronunciation of T1, which is a little bit like T4. This is not considered a proper T1 sound in the National Pronunciation.' (My translations, from the original Chinese script: "京城裏另有一種俏皮的讀法，把「陰」聲念的有一點去聲意味，這不算是國語的陰聲。")

It appears that Chao, who did not grow up in Beijing (Chao 1971), failed to realize that the falling variant was a legitimate and conservative variant and misinterpreted it as "witty" instead.

For T2, truncation-like effects can account for the removal of the final "5", and subsequently, speakers may need to initiate T2 from a relatively lower pitch to retain its rising characteristic, leading to a shift from higher rising to lower rising.

For T3, the change can also be attributed to truncation-like effects akin to those observed in T2: coda deletion and onset bending. The accompanying hypocorrection also plays an important role. During the 20th century, as the truncated "half-T3" frequently appears before other tones (Chao 1968: 27; Duanmu 2007: 236–238), younger generations may have perceived the rising end of T3 as an additional "release" or "regress to the middle" instead of an underlying pitch target, and tended to weaken it. Likewise, Shi & Ran (2011) regarded the rising end of T3 as a syllable boundary marker. While their interpretation remains controversial (Cao & Wei 2016), it illustrates the potential details of such hypocorrection.

For T4, the literature-documented shift 313 > 31 and the 20th-century shift 51 > 52 represent two truncation-like processes. The intermediate shift 31 > 51 might potentially be a consequence of onset bending resulting from truncation-like effects or be attributed to push/drag effects discussed in the next section.

4.1.2 Push/drag effects

Besides truncation-like effects arising from the articulatory perspective, I also recognize push/drag effects within the tonal system as a secondary motivating factor for tone evolution.

Just as contrastive vowels in a vowel system are typically dispersed in the vowel space (e.g. /i/, /u/, and /a/ for a 3-vowel system) to maintain maximal perceptual contrast (Liljencrants & Lindblom 1972; Flemming 2004), contrastive tones similarly tend to retain as much distinction as possible in the tonal space. The tone pattern of the Beijing dialect today is a good synchronic example, where the features of the 4 tones can be regarded as being quasi-symmetrically dispersed along the four directions of the two main axes in the tonal space: high–rising–low–falling (Shi & Wang 2006a; Duanmu 2007: 236–238).

Throughout the evolution, this tendency might lead to a preference towards variants that are more distinct from the other tonemes in the system, namely push/drag effects between neighboring tones, or what Pittayaporn (2018) referred to as the systemic bias of the avoidance of similar tones. Given that the tone evolution trajectory of the Beijing dialect follows a circular chain, it may be difficult to identify a specific trigger or initiator of the chain shift and distinguish between pushing and dragging based on current evidence. Hence, I employ the generalized term, push/drag effects.

Nonetheless, we start the discussion from T1. As reflected in the early recordings, T1 and T2 were very close to each other from the dimensions of average pitch and average slope, particularly in 1900 (Figure 7 or Figure 17). From the dimension of convexity, distinguishing between two convex tones (as in 1900) also presents difficulties. This scenario could potentially motivate a push/drag chain shift to enhance perceptual distinctiveness between T1 and T2. Similarly, push/drag effects exist between T1 and T4 to maintain perceptual distinctiveness between these two tones, which shared the falling feature in 1900 or the high feature in later periods. Consequently, alongside the truncation of T1 itself, the shift from high falling to high level in T1 may also be attributed to the push effect from the raising T4 or the drag effect from the lowering T2. Likewise, the lowering of T2, dipping of T3, and raising of T4 may also be influenced respectively by push effects from their upstream tones or drag effects from their downstream tones.

4.1.3 Mechanism summary and bidirectional possibilities

The truncation-like effects and push/drag effects may have operated synergistically or alternately, producing a clockwise shift of the tone pattern. I posit that these motivations apply not only to the Beijing dialect, but also likely to many other Chinese dialects and other tonal languages. Meanwhile, I must acknowledge that further evidence, particularly real-time (i.e. detecting sound change through time as it progresses) evidence, is needed to substantiate the motivations I have conjectured, especially to explore the internal mechanisms of the truncation-like process and the microscopic progression of the chain shift.

Furthermore, it is essential to acknowledge that external factors, though maybe minor, can influence evolution. These factors include social class differences in Beijing, dialect contact, the National Language Movement, the promotion of Putonghua (Standard Chinese), etc. Such factors remain to be further explored since it is impossible to observe them from only recordings.

It must be noted that a small number of dialects/languages observed have undergone tone changes in the counterclockwise direction (Yang & Xu 2019). This suggests that there must be other (perhaps minor) mechanisms besides those in the Beijing dialect.

Returning to Hirayama's (1984) hypothesis that the vocal folds slacken at the end of the syllable and induce the pitch to regress to the middle (e.g. 55 > 53 and 11 > 13), we found that it contradicts the truncation-like explanation above of final tonal element deletion (e.g. 53 > 55 and 13 > 11). This raises bidirectional possibilities: What is the "fate" of the final tonal element – regression or deletion?

One single answer probably does not exist. As an analogy, we can look at the lenition of consonants. A consonant may lenite by two paths: sonorization (e.g. t > ɾ) and loss of place/laryngeal properties (e.g. t > ʔ or t^h > t) (Szigetvári 2008). Regression is similar to sonorization, while deletion is similar to the loss of place of articulation (a.k.a. debuccalization). Since both paths of lenition are common for consonants, I provisionally admit that both answers for the final tonal element are possible. However, the hypothetical counterclockwise shift lacks reliable real-time instances like Beijing Mandarin or Bangkok Thai, and is therefore far from being vigorously established.

4.1.4 Pitch–phonation interactions in T3

In addition to the contour changes and their directionality, the increasing creakiness of T3 in the Beijing dialect also warrants a discussion. I propose that the increasing creakiness of T3 is also a result of increasing lowness, as creaky phonations are associated with the lower register (Laver 1980: 122–126; Zhu et al. 2012).[21] Cao & Wei (2016) even conjectured that creakiness may be gradually superseding the essence of T3, while conflicting findings emerged from other studies (Zheng 2006; Kuang 2017), suggesting that creakiness is merely a surface-level phenomenon induced by lower pitch. Nevertheless, the increasing creakiness may also have exacerbated the concavity of T3 in turn.

The interaction between "extreme" pitch and "extreme" phonations is similar to the interaction between high vowels and spirantization (fricativization) (Zhu 2004; Faytak 2014). The acquisition of irregular phonation in Beijing T3 and the spirantization of high vowels in some languages are both results of chain shifts. Extreme target produces extreme sound quality.

4.2 Speed of tone evolution

The tone evolution is remarkably fast, compared to what one might initially expect. In the tonal space, the angular coordinates of the 4 tones have rotated nearly a quarter of the circle in around 200 years (assuming a starting point slightly earlier than Edkins 1857). In other words, the tone value of each tone has shifted to the value of its downstream tone after merely 200 years. We can turn Figure 17 back into a sketch (Figure 18) to better illustrate this.

Here I postulate a potential relationship between the astonishing speed and the great degree of freedom of the tonal space when occupied by only 4 tones. The

21. The breathiness or slackness, less commonly presented in T3, is also associated with the lower register (Zhu et al. 2012).

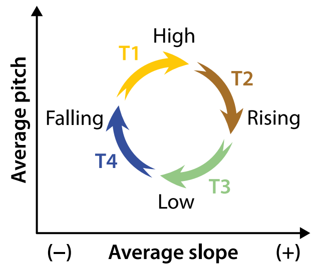


Figure 18. An illustration of the clockwise tone shift in the Beijing dialect in around 200 years

ample space allows each tone to exhibit a diverse range of variation, which might accelerate evolution under the biases of the two mechanisms.

It is also natural for tones to evolve at different speeds in different dialects (Hirayama 1984; Endo 2015). Among the dialects in the surrounding areas of Beijing that have similar tone patterns to the Beijing dialect, some tones are more advanced than their counterparts in Beijing following the clockwise shift, while some are more conservative (Zeng 2021).[22] Deeper research into the tones in the surrounding areas could shed light on the historical evolution of Beijing Mandarin.

Finally, an intriguing phenomenon is that some southern variants of Putonghua, such as Cantonese-accented Mandarin (Jin et al. 2008) and Changsha-accented Mandarin (Xu 2019), exhibit a tone pattern that is farther ahead of the Beijing dialect in the direction of the clockwise shift. The reason why they are ahead requires further exploration, and may also provide valuable insights into the tone evolution.

4.3 Value and enlightenment of early recordings

The study of sound change can be approached from two perspectives: apparent-time and real-time. The same applies to the study of tone change, e.g. Liang & Meng's (2013) study of tone change in Chongqing dialect is apparent-time, while Ming & Zhang's (2015) study is real-time.

Labov (1994: 111–112) suggests that apparent-time studies may underestimate the rate of sound change, since older speakers tend to exhibit communal change. This could also apply to tone changes. For example, T2 in Shu (1930) was

22. Tone patterns of some Tianjin suburban dialects even resemble that of the early Beijing dialect, e.g. the tone values of the 4 tones in Yangliuqing 楊柳青: 43, 45, 113, and 31 (Zeng 2021).

noticeably high, but in his 1966 recording,[23] it became almost indistinguishable from T2 in modern Putonghua. Surprisingly, some conservative segmental features in Shu (1930), e.g. pronouncing the neutral tone particle *de* 的 as *di* instead of reducing its vowel to schwa (*e*) as in modern Putonghua, persisted in his 1966 recording. This further suggests that tones may be more variable than segmentals.

Early recordings are invaluable resources for real-time studies of language evolution. As more collections are digitized and made available online, linguists have access to a great number of early recordings. It is not hard to identify some phonetic differences between these recordings and contemporary speech, indicating possible sound changes. Azoulay's (1900) recordings of citation tones of Beijing dialect are particularly valuable, as they enable the quantitative analysis of tone evolution since 1900.

4.4 Value and enlightenment of tone evolution study

Experimental results from early recordings confirm the existence of the fast clockwise tone shift in Beijing. The details and mechanisms in this study are of great significance for both diachronic and synchronic studies of tones. Additionally, this study provides new starting points and possibilities for the tonal reconstruction of Early Mandarin.

While the reconstruction of Early Mandarin is not the focus of this study, our results would prompt a reexamination of some previous works. Some attempts to reconstruct the tones of Early Northern Mandarin yielded results similar to modern Beijing tones (see review in Zhang 2010). However, I consider such results implausible given the tonal differences between contemporary and 1900 Beijing. Hirayama (1984) and Zhengzhang (2013: 207) suggested that the creakiness of T3 in Beijing is a relic of their reconstructed glottalization of T3 in Proto-Mandarin or Old Chinese. Yet, their proposals regarding Beijing dialect also seem incorrect, since it is likely that the creakiness of T3 was newly developed from lowness since the 19th century.

Some phonologists attribute the T3 sandhi 214 + 214 → 35 + 214 to low tone dissimilation (e.g. Lin 1998) or the Obligatory Contour Principle (see review in Duanmu 2007: 255). However, T3 sandhi has been documented as early as the 16th century (Mei 1977) and these explanations cannot be extrapolated to earlier stages given that T3 may not have been low historically. Thus, other explanations for the emergence of T3 sandhi must be explored.

---

23. An NHK interview with Lao She, remastered and published on the CD accompanying Shu (1999).

When confronted with inconsistent tone documents, linguists should not always attribute the problem to the inaccuracy and subjectivity of tone transcriptions. Instead, inconsistencies may indicate tone evolution. We must bear in mind that tones can evolve quickly before drawing any conclusions regarding tones.

5. Conclusion and prospects

I presented a comprehensive and quantitative analysis of the tone evolution in the Beijing dialect since 1900, and verified the clockwise tone shift in the Beijing dialect that has occurred at a rapid pace, with each tone nearly completing a quarter-circle shift in around 200 years. From articulatory and systemic perspectives, I identified two possible mechanisms of tone change: truncation-like effects coupled with hypocorrection, and push/drag effects between neighboring tones. The findings are valuable for both synchronic and historical linguistics, as well as for both Chinese dialects and other tonal languages.

Further research is needed to gain a deeper understanding of tone evolution in the Beijing dialect and in general. It is important to acknowledge that the suggestions in this paper are only based on the four-tone system of the Beijing dialect. A tonal system with more tones may provide a different evolutionary picture. The tone evolution still contains many unknowns. Thus, we call for further research on the tone evolution of more dialects/languages from both apparent-time and real-time perspectives, and covering phonation, tone sandhi, and tone of polysyllabic words whenever possible.

The contemporary data used in this study was collected in 2003. We can expect that tones in Beijing have undergone new subtle changes over the past two decades, and the clockwise tone shift may be further aided by tonal experiments on young generations today. Furthermore, the historical data is very limited, only recordings of five male speakers are available in this study. We believe that continued collection of tonal data over time will eventually make tone change self-evident, and the findings on tone change in contemporary literature will serve as a starting point for future research.

Acknowledgements

An earlier version of this paper was presented at the 15th Phonetic Conference of China (PCC 2023), Shenzhen, and the 2023 Annual Conference of the Linguistic Society of Tianjin, Tianjin. I thank the audience for their insightful feedback. I wish to thank Qibin Ran, the two anonymous reviewers, and the editor of *Language and Linguistics* for their valuable suggestions and comments which led to significant improvements of this paper. I am also grateful to Shengkai Zhou for her helpful perspectives on the visualization of tone patterns.

List of abbreviations

F_0 Fundamental frequency
H High
M Mid
T1 Tone 1
T2 Tone 2
T3 Tone 3
T4 Tone 4

Appendix A. Absolute duration and absolute pitch

	T1 (ms)	T2 (ms)	T3 (ms)	T4 (ms)	Pitch (Hz)
Azoulay (1900), 1st take	494.8	415.8	479.0	222.4	217.1
Azoulay (1900), 2nd take	375.6	339.0	405.6	204.3	297.7
Karlgren (1915)*	300.0	230.0	265.0	160.0	195.1
Wang (1920)	337.3	352.4	318.8	371.4	273.3
Chao (1922a)	296.7	292.9	271.1	286.5	211.0
Chao (1922c)*	/	/	/	/	140.4
Liu (1925)*	/	/	/	/	119.4
Shu (1930)	235.3	256.1	356.2	217.0	181.6
Pai (1933)	264.3	277.5	306.1	207.3	172.8
Pai (1934)*	198.0	239.2	359.8	135.0	193.3
Obata & Tesima (1934)*	436.0	455.0	483.0	425.0	204.0
Lin et al. (1998)	205.5	270.6	376.2	187.4	153.5
Sanders & Shi (2003)	292.0	343.6	396.9	243.5	139.6

Columns “T1” to “T4” provide the average duration of each tone. The last column provides the average of the average pitch of each tone. Chao (1922c) did not include a time scale. Liu (1925) included redundant segments of pitch contours in his figures, so it is hard to determine the absolute duration of a tone. Early experiments are marked by asterisks.

Appendix B. Supplementary data

Raw data (e.g., character, pinyin, duration, and pitch) of all samples of citation tones extracted or converted in this study, along with links to the digitized materials and details about the recordings, are available at Zenodo: ⟨https://doi.org/10.5281/zenodo.15192724⟩.

References

doi Andruski, Jean E. & Costello, James. 2004. Using polynomial equations to model pitch contour shape in lexical tones: An example from Green Mong. *Journal of the International Phonetic Association* 34(2). 125–140.

Azoulay, Léon (ed.). 1900. *Exposition universelle de Paris, 1900*. (https://archives.crem-cnrs.fr/archives/collections/CNRSMH_I_1900_001/) (Accessed 2022-01-07.) (Unpublished cylinder recordings digitized by the Centre de recherche en ethnomusicologie, Laboratoire d'ethnologie et de sociologie comparative.)

Bao, Huaiqiao & Lin, Maocan (eds.). 2014. *Shiyan yuyinxue gaiyao*. Revised edn. Beijing: Peking University Press.

Bickers, Robert A. 1994. New light on Lao She, London, and the London Missionary Society, 1921–1929. *Modern Chinese Literature* 8(1-2). 21–39.

Boersma, Paul & Weenink, David. 2022. *Praat: Doing phonetics by computer* (Version 6.2.09). (http://www.praat.org) (Accessed 2022-02-17.)

doi Bradley, Cornelius Beach. 1915. The tone-accents of two Chinese dialects. *Journal of the American Oriental Society* 35. 199–206.

doi Bröring, Theodor. 1927. *Laut und Ton in Süd-Schantung: Mit Anhang: Die Töne in Nordostschantung, Peking, Sötshuän, Shanghai, Amoy und Canton* (Veröffentlichungen des Seminars für Sprache und Kultur Chinas an der Hamburgischen Universität 2). Hamburg: L. Friederichsen & Co.

doi Burroni, Francesco & Kirby, James. 2023. Speakers adaptively plan and execute f0 trajectories under rate changes: Evidence from Thai contour tones. In Dong, Minghui (ed.), *Proceedings of the Second International Conference on Tone and Intonation (TAI 2023)*, 49–53. Singapore: ISCA Archive.

doi Cao, Wen & Wei, Liping. 2016. What is the form of Chinese third tone?: In honor of Professor Lin Tao (1921–2006). *Huawen Jiaoxue yu Yanjiu* 2016(4). 1–14.

Chao, Yuen Ren. 1922a. *Guoyu liushengji pian*. Shanghai: The Commercial Press. (https://dl.lib.ntu.edu.tw/s/78rpm/item/884038) (Accessed 2022-09-29.) (Disc 4.) (Phonograph records digitized by the National Taiwan University Library.)

Chao, Yuen Ren. 1922b. *Guoyu liushengpian keben*. Shanghai: The Commercial Press. (Textbook for Chao 1922a.)

Chao, Yuen Ren. 1922c. The methods for investigation of the intonation of the Chinese language. *Kexue* 7(9). 871–882.

Chao, Yuen Ren. 1968. *A grammar of spoken Chinese*. Berkeley: University of California Press.

doi Chao, Yuen Ren. 1971. My linguistic autobiography. *Bulletin of the Institute of History and Philology* 43(3). 303–317.

Chen, Pengfei. 2018. Circular chain shifts of tones and convergence of citation tones — A reexamination on the origin of three-tone system in Shandong and Hebei dialects after a model on circular chain shifts of tones. In Editorial Committee of *Lishi yuyanxue yanjiu*, Institute of Linguistics, CASS (ed.), *Lishi yuyanxue yanjiu*, vol. 12, 27–43. Beijing: The Commercial Press.

Chen, Xiao. 2018. Tonal variations of the Beijing dialect in late-Qing times and the early years of the Republic of China. *Zhongguo Yuwen* 2018(4). 435–445.

Dong, Jianjiao. 2010. Further exploration into the tone sandhi pattern of Qusheng (the fourth tone) in Beijing dialect. *Yuyan Yanjiu Jikan* 7. 164–172.

Duanmu, San. 2007. *The phonology of Standard Chinese*. 2nd edn. Oxford: Oxford University Press.

Edkins, Joseph. 1857. *A grammar of the Chinese colloquial language, commonly called the Mandarin dialect*. 1st edn. Shanghai: London Mission Press.

Edkins, Joseph. 1864. *Progressive lessons in the Chinese spoken language; with lists of common words and phrases, and an appendix containing the laws of tones in the Peking dialect*. 2nd edn. Shanghai: Presbyterian Mission Press.

Endo, Mitsuaki. 2015. Jin 150-nian lai Hanyu gezhong fangyan li de shengdiao yanbian guocheng: Yi Aiyuese de miaoxie wei chufa dian. In Endo, Mitsuaki & Ishizaki, Hiroshi (eds.), *Xiandai Hanyu de lishi yanjiu*, 199–228. Hangzhou: Zhejiang University Press.

Faytak, Matthew. 2014. High vowel fricativization and chain shift. *UC Berkeley PhonologyLab Annual Report* 10. 52–100.

Flemming, Edward. 2004. Contrast and perceptual distinctiveness. In Hayes, Bruce & Kirchner, Robert & Steriade, Donca (eds.), *Phonetically based phonology*, 232–276. Cambridge: Cambridge University Press.

Gandour, Jack. 1983. Tone perception in Far Eastern languages. *Journal of Phonetics* 11(2). 149–175.

Higashi, Takahiro & Wang, Yunjia. 2017. Cause of Tone 4 sandhi in Beijing Mandarin and its relationship with lexical stress. *Language and Linguistics* 18(3). 430–451.

Hirayama, Hisao. 1984. A genealogical classification of the tone value systems of Mandarin Chinese — With reference to the Hebei dialects. *Gengo Kenkyu* 86. 33–53.

Hirayama, Hisao. 1998. A discussion about the neutral tone development of Shāndōng dialects from the viewpoint of tone sandhi. *Fangyan* 1998(1). 7–13.

Hyman, Larry M. 2007. Universals of tone rules: 30 years later. In Riad, Tomas & Gussenhoven, Carlos (eds.), *Tones and tunes, volume 1: Typological studies in word and sentence prosody* (Phonology and Phonetics 12), 1–34. Berlin: De Gruyter Mouton.

Jia, Junyuan. 2024. La Chine entendue par l'Occident: Les enregistrements sonores de la Chine à l'Exposition universelle de 1900. *Artefact* 21. 171–196.

Jin, Jian & Hu, Weixiang & Wang, Xia & Xiong, Ziyu. 2008. A comparative study on tone realization in Cantonese-accented Mandarin and standard Mandarin. In Barbosa, Plínio A. & Madureira, Sandra & Reis, César (eds.), *Proceedings of the Speech Prosody 2008*, 305–308. Campinas: RG Editores.

Karlgren, Bernhard. 1915. *Études sur la Phonologie Chinoise* (Archives d'Études Orientales 15), vol. 1. Leyde: E. J. Brill.

Karlgren, Bernhard. 1918. *A Mandarin phonetic reader in the Pekinese dialect: With an introductory essay on the pronunciation* (Archives d'Études Orientales 13). Stockholm: P. A. Norstedt & Söner.

Kuang, Jianjing. 2017. Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice. *The Journal of the Acoustical Society of America* 142(3). 1693–1706.

Labov, William. 1994. *Principles of linguistic change, volume 1: Internal factors* (Language in Society 20). Oxford: Blackwell.

Lai, Wei & Xu, Xiaoying & Chen, Limei & Zhu, Xiaonong. 2017. Rongcheng aodiao: Jiangsheng zuowei bibei tezheng. *Dongfang Yuyanxue* 16. 18–26.

Laver, John. 1980. *The phonetic description of voice quality* (Cambridge Studies in Linguistics 31). Cambridge: Cambridge University Press.

Li, Deyang & Zhang, Dequan. 1909. *Dengmi yinyu*. Paris: Pathé Records. (Phonograph record remastered in the 2007 documentary *Xiangsheng Dashi*, Episode 1, produced by China Central Television.)

Li, Jiinshi. 1934. *Gwoyeu Yunndonq shyy gang*. Shanghai: The Commercial Press.

Liang, Lei & Meng, Xiaolin. 2013. The tone variation of Chongqing dialect. *Language and Linguistics* 14(5). 929–959.

doi Liljencrants, Johan & Lindblom, Björn. 1972. Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language* 48(4). 839–862.

Lin, Hua. 1998. "Tonal elements" and tone sandhi in Putonghua. *Zhongguo Yuwen* 1998(1). 31–39.

Lin, Tao. 1992. Additional interpretation of the 'redistribution of entering tone among other tones'. In Editorial Committee of *Yuyanxue luncong*, Department of Chinese Language and Literature, Peking University (ed.), *Yuyanxue luncong*, vol. 17, 3–18. Beijing: The Commercial Press.

Lin, Tao & Zhou, Yimin & Cai, Wenlan. 1998. *Beijinghua yindang* (Xiandai Hanyu Fangyan Yinku 1). Shanghai: Shanghai Education Publishing House. (Includes sound recordings on the accompanying tape.)

Liu, Fu. 1925. *Étude expérimentale sur les tons du Chinois*. Paris: Les Belles Lettres.

Liu, Zhijing & Lin, Ju & Zhang, Jinsong & Zhang, Weijia. 2016. The effect of age and gender on creaky voice in Tone 3 of Beijing dialect. *Zhongguo Yuyin Xuebao* 6. 32–37.

Meadows, Thomas Taylor. 1847. *Desultory notes on the government and people of China, and on the Chinese language; illustrated with a sketch of the province of Kwang-Tûng, shewing its division into departments and districts*. London: W. H. Allen and Co.

Mei, Tsu-lin. 1977. Tones and tone sandhi in 16th century Mandarin. *Journal of Chinese Linguistics* 5(2). 237–260.

doi Ming, Mao-xiu & Zhang, Xian-cheng. 2015. The patterns and evolution mechanism of tone value of Chinese dialects. *Xinan Daxue Xuebao (Shehui Kexue Ban)* 41(4). 145–155.

doi Obata, Jûichi & Tesima, Takehiko. 1934. Shina-go no butsuri onseigakuteki kenkyū: Shisei no seishitsu. *Nippon Sugaku-Buturigakkwaishi* 8(1). 1–10.

Ohala, John J. 1981. The listener as a source of sound change. In Masek, Carrie S. & Hendrick, Roberta A. & Miller, Mary Frances (eds.), *Papers from the Parasession on Language and Behavior: Chicago Linguistic Society, May 1–2, 1981*, 178–203. Chicago: Chicago Linguistic Society.

Ohala, John J. 1993. The phonetics of sound change. In Jones, Charles (ed.), *Historical linguistics: Problems and perspectives*, 237–278. New York: Longman.

Pai, Ti-chou. 1933. *Biaozhun guoyin*. Shanghai: Zhonghua Book Company. (https://dl.lib.ntu.edu.tw/s/78rpm/item/953708) (Accessed 2022-09-29.) (Disc 3.) (Phonograph records partially digitized by the National Taiwan University Library.)

Pai, Ti-chou. 1934. Beijingyu shengdiao ji bianhua. (Manuscript. Partially published in Luo, Changpei & Wang, Jun. 1957. *Putong yuyinxue gangyao*. Beijing: Science Press.)

Pittayaporn, Pittayawat. 2007. Directionality of tone change. In Trouvain, Jürgen & Barry, William John (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences, ICPhS XVI, 6–10 August 2007, Saarbrücken, Germany*, 1421–1424. Saarbrücken: Universität des Saarlandes.

doi Pittayaporn, Pittayawat. 2018. Phonetic and systemic biases in tonal contour changes in Bangkok Thai. In Kubozono, Haruo & Giriko, Mikio (eds.), *Tonal change and neutralization* (Phonology and Phonetics 27), 249–278. Boston: De Gruyter Mouton.

Rudy, Charles. 1874. *The Chinese Mandarin language after Ollendorff's new method of learning languages*, vol. 1. Geneva: H. Georg.

Sanders, Robert & Shi, Feng. 2003. *Hanyu Yuyin Shujuku*. (Unpublished speech corpus.)

Shen, Ruiqing. 2018. Modelling tonal chain shifts in tonal acoustic space. (Poster presented at the 17th Australasian International Conference on Speech Science and Technology, Sydney, 4–7 December 2018.) (https://scholarbank.nus.edu.sg/handle/10635/163047) (Accessed 2022-10-27.)

Shi, Feng. 1986. Analysis of tone sandhi in bisyllables of the Tianjin dialect. *Yuyan Yanjiu* 1986(1). 77–90.

Shi, Feng & Ran, Qibin. 2011. A comment on "Mandarin tone perception: A report on a low level tone". *Zhongguo Yuwen* 2011(6). 550–555.

Shi, Feng & Wang, Ping. 2006a. A statistic analysis of the tones in Beijing Mandarin. *Zhongguo Yuwen* 2006(1). 33–40.

doi Shi, Feng & Wang, Ping. 2006b. A statistic analysis of tone groups in Beijing Mandarin. *Dangdai Yuyanxue* 8(4). 324–333.

doi Shi, Shaowei. 2007. *"Xiandai Hanyu Fangyan Yinku" danzidiao shiyan yanjiu*. Nanjing: Nanjing Normal University. (Master's thesis.)

Shu, Chien Chun. 1930. *Chinese* (Linguaphone Language Courses). London: Linguaphone Institute. (https://mp.weixin.qq.com/s/2D4Po9luYczDZCjGiB6X8Q) (Accessed 2022-09-29.) (Phonograph records digitized by the Beijing Classics and Old Records Digital Publishing Project.)

Shu, Ji. 1999. *Lao She jiangyan ji*. Beijing: SDX Joint Publishing Company. (Includes a sound recording on the accompanying CD.)

Steriade, Donca. 2000. Paradigm uniformity and the phonetics-phonology boundary. In Broe, Michael B. & Pierrehumbert, Janet B. (eds.), *Papers in laboratory phonology V: Acquisition and the lexicon*, 313–334. Cambridge: Cambridge University Press.

doi Sun, Yan & Shih, Chilin. 2021. Boundary-conditioned anticipatory tonal coarticulation in Standard Mandarin. *Journal of Phonetics* 84. 1–27. (Article 101018.)

doi Szigetvári, Péter. 2008. Two directions for lenition. In Carvalho, Joaquim Brandão de & Scheer, Tobias & Ségéral, Philippe (eds.), *Lenition and fortition* (Studies in Generative Grammar 99), 561–592. Berlin: De Gruyter Mouton.

doi Ting, Pang-hsin. 1982. Some aspects of tonal development in Chinese dialects. *Bulletin of the Institute of History and Philology, Academia Sinica* 53(4). 629–644.

Wang, Gang & Du, Junmin. 2016. Baidai gongsi zai Zhongguo de faduan. In Zhu, Hengfu & Nie, Shengzhe (eds.), *Zhonghua yishu luncong*, vol. 17, 3–50. Shanghai: Shanghai University Press.

Wang, Li. 1979. Xiandai Hanyu yuyin fenxi zhong de jige wenti. *Zhongguo Yuwen* 1979(4). 281–286.

Wang, Pu. 1920. *Chinese national phonetic record*. Shanghai: Zhonghua Book Company. (https://dl.lib.ntu.edu.tw/s/78rpm/item/953728) (Accessed 2022-09-27.) (Disc 3.) (Phonograph records partially digitized by the National Taiwan University Library.)

doi Wang, Tianheng. 2021. A novel approach to visualization of tone value pattern. (Poster presented at the 14th Phonetic Conference of China, Lanzhou, 18–20 July 2021.)

doi Wang, William S-Y. 1967. Phonological features of tone. *International Journal of American Linguistics* 33(2). 93–105.

Wang, Yonghong (dir.). 1956. *Zhengyi de shenpan*. Beijing: Central Newsreel and Documentary Film Studio. (Documentary film.)

Xu, Chenzi. 2019. Characterising intonation in Plastic Mandarin using polynomial modelling. In Calhoun, Sasha & Escudero, Paola & Tabain, Marija & Warren, Paul (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, 696–700. Canberra: Australasian Speech Science and Technology Association Inc.

doi Xu, Yi & Xu, Anqi. 2021. Consonantal F_0 perturbation in American English involves multiple mechanisms. *The Journal of the Acoustical Society of America* 149(4). 2877–2895.

doi Yang, Cathryn & Xu, Yi. 2019. Crosslinguistic trends in tone change: A review of tone change studies in East and Southeast Asia. *Diachronica* 36(3). 417–459.

Zeng, Xiaoyu. 2021. On the internal differences of Beijing dialect single-tone in the late Qing dynasty and the early Republic of China. *Hanyu Yuyanxue* 2021(1). 142–155.

Zhang, Yu-lai. 2010. Tone categories and tone pitches of Chinese in the time of *Zhongyuan yinyun*. *Guhanyu Yanjiu* 2010(2). 11–25.

Zheng, Xiaoju. 2006. Voice quality variation with tone and focus in Mandarin. In Belotel-Grenié, Agnès & Grenié, Michel (eds.), *Proceedings of the 2nd International Symposium on Tonal Aspects of Languages (TAL 2006)*, 132–136. La Rochelle: Université de La Rochelle.

Zhengzhang, Shangfang. 2013. *Old Chinese phonology*. 2nd edn. Shanghai: Shanghai Educational Publishing House.

Zhu, Xiaonong. 2004. Sound changes of high vowels in Chinese dialects. *Zhongguo Yuwen* 2004(5). 440–451.

doi Zhu, Xiaonong & Li, Fei. 2016. Tone chain-shifts in Meizhou Hakka: A case study with the evolutionary comparative method. *Yuwen Yanjiu* 2016(4). 1–8.

Zhu, Xiaonong & Lin, Qing & Saengthong, Pratchaya. 2015. Thai tones: Tonotypes and their clockwise chain shift. *Minzu Yuwen* 2015(4). 3–18.

Zhu, Xiaonong & Zhang, Ting & Yi, Li. 2012. A classification of dipping tones. *Zhongguo Yuwen* 2012(5). 420–436.

Address for correspondence

Tianheng Wang
School of Liberal Arts
Nankai University
Fansun Building, 94 Weijin Rd.
Tianjin 300071
China

tw1701@nyu.edu
https://orcid.org/0009-0001-1306-7349

Publication history

Date received: 7 May 2023
Date accepted: 17 April 2024
Published online: 5 August 2025