Acoustic and Perceptual Study of Tones in Jin Chinese (Togtoh variety)

Yue Wang[1], *Wen Liu*[1,2]

[1] Center for Language Sciences, School of Literature, Shandong University, Jinan, China

[2] Chinese National Information Technology Research Institute, Northwest Minzu University, Lanzhou, China

yuewang0292922@163.com, liuwen3214@163.com

Abstract

Previous research has shown that tone is a complex linguistic concept in terms of speech production and perception. In this study, based on the fieldwork, tones in monosyllabic words from the dialect of Near-field Togtoh County were studied. This paper first extracts the F0 and duration of monosyllabic tones in this dialect and analyzes its pitch pattern. It demonstrates that the Togtoh variety has five tones for monosyllabic words: one level tone, two contour tones, and two falling tones. Secondly, from the perspective of speech perception, this study investigates the native speakers' distinctions between the two contour tones(Tone 1 and Tone 3) utilizing a discrimination test. It turns out that Tone 1 and Tone 3 cannot be completely distinguished. On this basis, the distinctive feature system of tones in the dialect of Near-field Togtoh County is established. Finally, this paper summarizes the acoustic and perceptual characteristics of monosyllabic tones.

Index Terms: Near-field Togtoh County Dialect, monosyllabic words, acoustic analysis, perceptual experiment, distinctive features

1. Introduction

Jin Chinese is a representative dialect of Chinese, and its "status" has been debated for many years. Among them, Inner Mongolian Jin Chinese is one part. For Inner Mongolian Jin Chinese, there has been a lack of systematic and detailed individual dialect documentation. To date, the following varieties (Huhhot, Zhuozishan, Xinghe, Liangcheng, Chahar) are available in the Zhanghu cluster. However, only the Qing Shui river variety in the Dabao cluster has been described [1]. So, there's a need to explore more language materials in this area.

Togtoh County is located in central the Inner Mongolia Autonomous Region, China. Located at the junction of Shanxi, Shaanxi, and Inner Mongolia Autonomous Region, the variety belongs to the Inner Mongolian Jin Chinese [2], subdivided into the Near-field and Far-field dialects. The Near-field dialect is the dominant dialect in the county and is similar to the Baotou dialect. Few studies have been concerned with this dialect, and there is almost no study on the tone system of Togtoh County, though some scattered records were reported in the *Togtoh county annals* and *A Study of Modern Jin Chinese*.

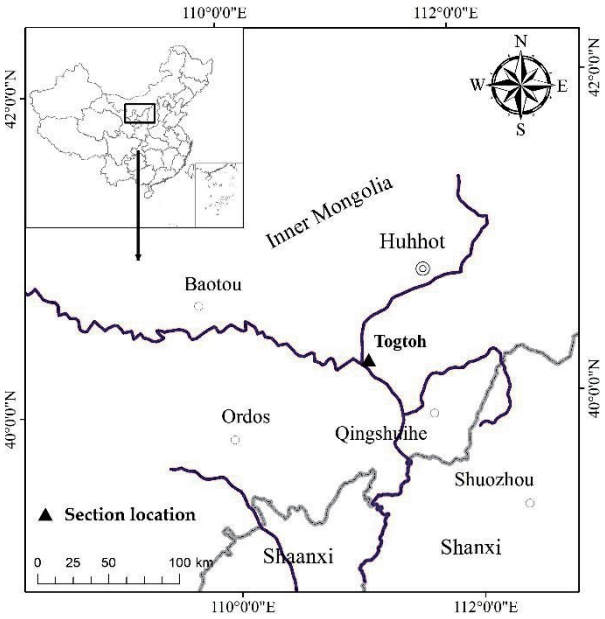

Figure 1: *Map showing the location of the junction of Shanxi, Shaanxi, and Inner Mongolia Autonomous Region. The triangle indicates the location of Togtoh.*

The difference between Inner Mongolian Jin Chinese mainly manifests in monosyllabic tones [3]. The tone is a linguistic concept. As Maddieson says, there are perhaps more difficulties in establishing phonological universals in the area of tone than in most of the other areas of phonology [4]. Therefore, it is necessary to discover more cases for discussing the status of Jin Chinese in Northern Mandarin and revealing some prominent features of Jin Chinese in Northern Mandarin.

Therefore, it is vital to conduct a detailed and scientific investigation of the dialect's facts and explore the universals of the tone. This study aimed to investigate the pitch pattern of tones in monosyllabic words of this dialect and its distinctive feature system preliminarily through a case study.

Correspondence to: W. Liu (liuwen3214@163.com).

2. Acoustic Analysis

2.1. Materials

Voiceless unaspirated plosives [p, t] and zero initials were selected as initials to avoid the consonants' effect on the tone. The finals were preferred monophthong [a, i, u]. It is also necessary to consider that samples are ideally free and common morphemes in the dialect [5]. Finally, after careful consideration, 15 words were selected for this dialect. The experimental materials are shown in Table 1.

Table 1：*monosyllabic words*

Tone categories	Sample words	Gloss	Sample words	Gloss	Sample words	Gloss
T1	pa	巴cling to	i	衣clothes	u	乌black
T2	pa	拔 pluck	i	移move	u	无no
T3	pa	把disposal	i	椅chair	u	五five
T4	pa	爸 father	i	意meaning	u	雾fog
T5	paʔ	八eight	tiəʔ	笛flute	tuəʔ	读read

The recording environment was an anechoic room with equipment, including a clip-on condenser microphone (*Audio-Technica AT9904*) and a laptop computer (*Lenovo Legion Y9000X 2020*). The recording software was *Adobe Audition CC 2018*, the recording channel is mono, and the sound file format is .wav, with a sampling rate of 44.1 kHz and 16-bit resolution. The signal analysis software was *Praat 6.1.29,* and the data analysis software was *Excel 2019* and *R 4.1.2*.

The 15 words were chosen as a monosyllabic-word group. Each word was pronounced naturally and comfortably four times, two seconds apart, without carrier sentences, and with a signal-to-noise ratio of at least 26 dB during recording.

2.2. Participants

Four speakers were recruited for the acoustic experiment (two of them were males aged 49 and 24 years, and the females were 46 and 25 years old respectively) were recruited for the acoustic experiment. They were all native speakers, and none had trouble hearing or speaking.

2.3. Procedure

The main procedures are as follows:

Step1: Determine the beginning and end of the tone. Extract F0 values of original tones in *Praat*. Each syllable's final stability was divided into ten parts (11 points).

Step2: Normalize the data. The data of the 11 points were averaged to acquire the mean F0 of all tokens under each tone category [6]. For more information, see Table 2, where the F0 average is Hertz, and the last line of the table shows duration (milliseconds).

Step3: Use the data from steps 1 and 2 to draw the tones. Get the F0 curve with R. Then, the F0 values were converted to semitones. Because the semitones are more consistent with the human sense of hearing [7] and are convenient for forming a physical-psychological mapping. The formula for the Hz to semitone method is:

$$n = 12 \times log_2 \frac{f_A}{f_{ref}} \text{ (reference frequency = 64.66Hz)}$$

Table 2:*F0 and duration*

Tones / Points	T1		T2		T3		T4		T5	
Gender / Points	M	F	M	F	M	F	M	F	M	F
1	108	215	115	232	98	204	151	286	125	242
2	104	206	114	233	94	198	146	282	120	237
3	102	202	114	234	91	191	144	279	119	234
4	102	200	114	234	86	183	142	276	118	232
5	103	201	114	233	89	180	139	272	117	230
6	105	204	114	232	90	181	136	268	116	229
7	108	210	113	231	93	188	133	262	115	227
8	112	217	113	231	97	198	130	257	114	225
9	117	225	113	230	102	208	125	251	112	223
10	122	232	112	230	105	214	120	244	110	222
11	127	238	111	230	108	216	116	236	108	220
Duration	283	356	403	437	314	368	161	179	82	86

2.4. Results

Based on the obtained semitone data, the tonal values of the five tones of the Near-field Togtoh County were established using the "five-level tone system [8]". Figure2 and Figure3 are the semitone charts with a 95% confidence interval (method: bootstrapping) after checking the normality of the data.

On the one hand, the tonal register for males ranges from 4.95 to 14.67 semitones, with 1.94 semitones per degree. However, the degree among any of the two levels is not equal. We still transcribe the tones with the help of native speakers' sense of hearing. Thus, the resulting tonal values of the five tones for males are 324 for Tone 1 (Yinping), 44 for Tone 2 (Yangping), 213 for Tone 3 (rising tone), 53 for Tone 4 (departure tone), and 43 for the Tone 5 (entering tone).

On the other hand, the tonal register for females ranges from 17.70 to 25.92 semitones, with 1.60 semitones between each degree. Same as for males, the resulting tonal values of the five tones for females are Yinping 324, Yangping 44, rising tone 213, departure tone 53, and entering tone 43.

According to Zhu [9], it is speculated that {53} is a variant of /52/ or /54/, which is an unlikely tonal category. However, Liu [10]developed a frequency table of Jin tonal types (top 25 in frequency ranking) from a statistical point of view in *The Theory of Chinese Tones*. We can know that the tonal value 53 as a falling contour occurs most frequently in the dialectal system of Jin Chinese, so the departure tone is still described as 53.

The entering tone is also falling, but the falling amplitude is only 1 degree. The duration is about 70 ms for both males and females, starting at the 4th degree and ending at the 3rd degree. When marked with the five-number tone system, it is recorded

as 43ʔ, because the Inner Mongolian Jin Chinese entry vowel character always closes with the glottal stop [-ʔ] [3].

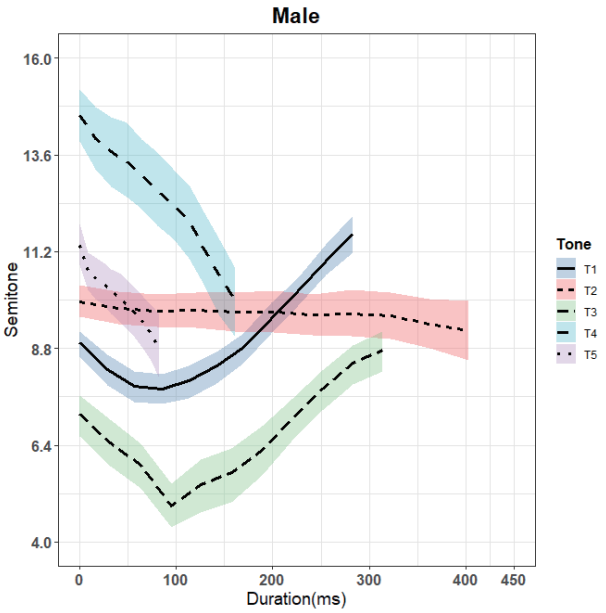

Figure 2: *semitone chart (male)*

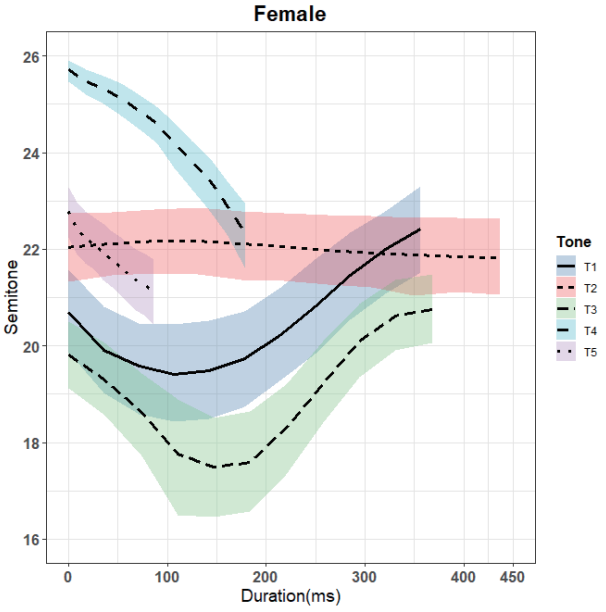

Figure 3: *semitone chart (female)*

It can be observed that the differences between males and females are reflected in the fundamental frequency values and tonal register and that the females' duration lasts longer than that of the male. But the confidence interval width for female speakers is more expansive than for males. This means data for men are relatively more reliable. Moreover, each tonal category is identical in the relative distribution of tonal values. Yinping and the rising tone are contour tones. Yangping is a high-level tone, departure tone is a high-falling tone, and entering tone is a slightly falling tone.

In summary, the pitch patterns of tones in monosyllabic words of Near-field Togtoh County were obtained. Two universal phonological rules can be summarized: 1) If a language has contour tones, it also has level tones. 2) A language with complex contours also has simple contours [4]. However, this dialect is unique between the two contour tones and needs to be confirmed by a perceptual study.

3. Perceptual Experiment

3.1. Materials and Methods

From the data for females, we could find overlaps between the starting points of Tone1 and Tone3. And during the investigation of *Dialect Word List*, there was an ambiguity in the production of Tone1 (324) and Tone3 (213). For example, the discrimination of "[piau^{324}] vs. [piau^{213}] (clock)" "[kũŋ^{324}] vs. [kũŋ^{213}] (common)" "[kɑ^{324}] vs. [kɑ^{213}] (steel)"'[dɑ^{324}] vs. [dɑ^{213}] (when)" "[tɕia^{324}] vs. [tɕia^{213}] (family)" and so on. So, what is the relationship between production and perception in Tone1 and Tone3? Do the production and perception have correlation and even consistency? Therefore, a simple experiment was designed to investigate the perceptual discrimination between the two tones. The experimental materials were derived from the monosyllabic word materials of male1 [pa], and the method was the AX paradigm which was developed from the ABX paradigm. Subjects were presented with a pair of stimuli, such as AA, AB, BA, and BB. A and B each represent a stimulus in the continuum, and the difference between AB and BA lies in the order of stimulus presentation. Stimuli were presented in pairs with a 500 milliseconds inter-stimulus interval (ISI) because 500ms can maximize differences in the performance of between-versus within-category discrimination [11]. The subject's experimental task was to discriminate whether the tone of the two tones heard was the same or different.

3.2. Participants

Thirteen native speakers (6M and 7F, mean age = 29.4 years, SD = 8.3 years) were recruited for the perceptual test. None of them reported any speech, language, or hearing difficulties.

3.3. Results

The results of the perceptual experiment are shown in Table 3.

Table 3: *dissimilarity matrix of T1 and T3*

	T1	**Correct**	**T3**	**Correct**
T1	0	100%	1	93.33%
T3	1	93.33%	0	100%

In this dissimilarity matrix, 11 speakers judged it entirely correctly. Two people judged it incorrectly, one speaker could not discriminate the T1-T3 continuum, and one speaker was unable to distinguish the T3-T1 continuum, so native speakers

could not fully discriminate the two contour tones 100%, suggesting a real correlation between production and perception in Tone1 and Tone3, as demonstrated by *the motor theory of speech perception* [12] and *the quantal theory of speech production* [13]. However, the exact reasons for this phenomenon still need further examination.

4. Distinctive Feature

Since the distinctive features were proposed [14], more and more scholars have started to use DF theory to study tones [15][16].

According to Zhu's theory, a general tonotype inventory includes four dimensions: register, length, height, and contour. This inventory can compare and classify tonal systems from various languages or dialects [17]. So in this part, from the perspective of linguistic typology, we use the distinctive features of Zhu's universal tonotype inventory as the reference to establishing the distinctive features system of the dialect of Near-field Togtoh County. The DF matrix is shown in Table 4.

From the DF matrix, we can find three types of contours in this dialect: level, dipping, and falling. In addition, it turns out that Tone 1 (324) and Tone 3 (213) are only different in the register, one belongs to Mid, and another is Low, which is probably one of the reasons why Tone 1 and Tone 3 are easily confused in the Near-region dialect of Togtoh County.

Table 4: *DF matrix*

		Contour	Level	Dipping	Falling	
Long	Mid	Tone categories	Mid high level	Front dipping	High falling	Mid slightly falling
		Tone value	44	324	53	
		Contour	+	+	+	+
		Falling	-	+	+	+
		Rising	-	+	-	-
		Straight			+	+
		$High_{front}$	-	-	+	-
		Low_{front}	-	-		
		$High_{back}$		+	-	+
	Low	Tone value		213		
Short	Mid	Tone value				43ʔ

5. Discussion

In the present study, we have added a new variety of Inner Mongolian Jin Chinese in the aspect of tones. We not only got the pitch patterns of tones in monosyllabic words but also built its distinctive feature system. It suggested five tones in Jin Chinese (Togtoh variety), one level tone, two contour tones, and two falling tones. Moreover, the four dimensions (including the register, length, height, and contour) could describe the distinctive features well. The results show that the tones in Togtoh are similar to the DaTong variety (Shanxi province) [18], which demonstrates the current strong position of the Jin Chinese in the evolution of Northern Mandarin. The phenomenon might have a close relationship with the continued entry of immigrants from Shanxi and Shaanxi provinces.

In addition, the two falling tones are easily distinguishable because of the phonation type of the glottal stop [-ʔ]. In contrast, the two contour tones are a confusable pair due to their similar concave shapes. In the acoustic study, we found overlaps between the starting points of Tone1 and Tone3 in female data, providing confusing acoustic clues. The present study has confirmed that the discrimination between Tone 1 and Tone 3 is ambiguous utilizing perceptual experiments.

Previous studies have documented Yinping mixing with Shangsheng [19][20], but no reliable acoustic and perceptual cues have been found. Further research is needed to identify possible causes of false discrimination, which can facilitate research into tone production and perception. It may be a better option to incorporate the phonation type into the system for distinctive features.

6. Conclusions

In this study, the monosyllabic tones from the dialect of Near-field Togtoh County were investigated. From the acoustic and perceptual study, it turns out that the dialect of Near-field Togtoh County has five tones for monosyllabic words: one level tone, two contour tones, and two falling tones. Meanwhile, by using the discrimination task, it is found that confusion between Tone 1 and Tone 3 does exist, which demonstrates that the perception between the two contour tones is correlative with the production. Last but not least, to correspond the continuous acoustic parameters to discrete cognitive categories, we also established a system of distinctive features for the dialect. However, the specific reasons for Tone 1 and Tone 3 still need more effort from more perspectives and methods.

7. Acknowledgements

This work was supported by National Social Science Foundation (22CYY022) and Future Plan for Young Scholars of Shandong University. We would like to give our thanks to all participants.

8. References

[1] M. Shen, "Re-classification / Re-distribution of Jin Group", *Fangyan,* vol.2, pp. 343-356,2006.

[2] Chinese Academy of Social Sciences and Australian Academy of the Humanities, *The language atlas of China*. Hong Kong:

Longman, 1987.
[3] J. Y. Hou, *A Study of Modern Jin Chinese*. Beijing: Commercial Press, 2010.
[4] I. Maddieson, "Universals of tone," in *Universals of human language*, vol. 2, J. H. Greenberg, C. A. Ferguson, and E. A. Moravcsik, Eds. Stanford, CA: Stanford University Press, 1978, pp. 335-365.
[5] X. N. Zhu, *Phonetics: New Concepts*. Beijing: Commercial Press, 2010.
[6] W. Liu, Z. H. Yang, and J. P. Kong, "An acoustic study of monosyllabic and disyllabic tone in Xinzhai Miao," *Minority Languages of China,* vol. 2, pp. 12–24, 2017.
[7] F. Liu, *Record of Experiments on the Four Tones*. Shanghai: Qunyi Publishing Press, 1924.
[8] Y.-R. Chao, "A system of tone letters," *Le maître phonétique,* vol. 45, pp. 24-27, 1930.
[9] X. N. Zhu, *Phonetic question and answer session*. Shanghai: Xuelin Press, 2018.
[10] L. L. Liu, *The Theory of Chinese Tones*. Nanjing: Nanjing Normal University Press, 2004.
[11] D. B. Pisoni, "Identification and discrimination of the relative onset time of two component tones: implications for voicing perception in stops," *The Journal of the Acoustical Society of America,* vol. 61, no. 5, pp. 1352-1361, Jun. 1977.
[12] A. M. Liberman and I. G. Mattingly, "The motor theory of speech perception revised," *Cognition,* vol. 21, no. 1, pp. 1-36, Oct. 1985.
[13] K. N. Stevens, "On the quantal nature of speech," *Journal of phonetics,* vol. 17, no. 1, pp. 3-45, 1989.
[14] R. Jakobson, C. G. Fant, and M. Halle, *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge: MIT Press, 1951.
[15] W. S. Wang, "Phonological features of tone," *International Journal of American Linguistics,* vol. 33, no. 2, pp. 93-105, Apr. 1967.
[16] Z. J. Wu, "A Preliminary Study of Distinctive Features and Their Correlates in Standard Chinese" *Studies of the Chinese Language* no. 05, pp. 321-327, 1980.
[17] X. N. Zhu, "Multiregisters and four levels: A new tonal model," *Journal of Chinese linguistics,* vol. 40, no. 1, pp. 1-17, 2012.
[18] M. Shen, " On the entering tone in Zhanghu cluster of the Jin group in Inner Mongolia", Studies of the Chinese Language, vol.5, pp.584-595+639-640,2022.
[19] C. P. Luo, *Northwest Dialect during Tang and The Five Dynasties.* Beijing: Commercial Press,2012.
[20] Q. S. Qiao, *A Study on the Phonetic History 100 of Jin Dialect*.Bejing：Zhonghua Book Company,2008.