Interspeech 2024
1-5 September 2024, Kos, Greece

Form and Function in Prosodic Representation: In the Case of 'ma' in Tianjin Mandarin

Tianqi Geng, Hui Feng

School of Foreign Languages Tianjin University
fredageng@163.com, fenghui@tju.edu.cn

Abstract

This study investigates the acoustic features of "ma" in Tianjin Mandarin, analyzing its functions as deixis, pronoun, modal particle, and discourse marker. Eight native speakers participated in the research. Results revealed distinct prosodic patterns, with "ma" as a modal particle exhibiting the shortest duration, weakest intensity, and narrowest pitch range. As a discourse marker, "ma" displays the widest pitch range. As a deixis, "ma" shares similar acoustic characteristics with it as a pronoun but exhibits a more stable distribution of primary acoustic features. Clustering identifies two distinct pitch patterns – high-falling and low-falling tones, demonstrating various realizations speakers may choose under different functions. This study highlights the correlation between the form and function of the prosodic representation of "ma" in Tianjin Mandarin.

Index Terms: Prosodic Representation, Form and Function, Tianjin Mandarin, "ma"

1. Introduction

There is still considerable opportunity for investigation, though nowadays, linguists are increasingly interested in the relationship between form and function. One is that most of the exploration of prosodic form and function is focused on Indo-European languages [1][2][3]. What's more, research on Chinese form and function is primarily out of teaching requirements [4][5][6].

The usages and features of the "ma" in Mandarin Chinese, Beijing Mandarin, and Tianjin Mandarin are pretty different regarding phonologic characters, intonation patterns, and semantic functions.

In Mandarin Chinese, "ma" is realized as a neutral tone and functions as a modal particle with two primary purposes – to express modal meanings [7][8][9]or to serve as a discourse marker [10]. Modal particles in Chinese tend to adhere to their previous syllable [7], while discourse markers function as independent intonation units [11][12] and can sometimes be identified by phonological shortening [13].

In Beijing Mandarin, "ma" is defined as a high-level tone with a tone value of 35 when it appears at the end of declarative sentences, used as a modal particle. This modal particle is characterized as low, flat, slowly rising, and slowing down[14]. Furthermore, in the Beijing dialect, "ma" can only occur in the phrase "gan^4ma^4 *what for*" [7].

Tianjin Mandarin is one of the Northern Mandarin dialects, which contains four lexical tones similar to the four in Standard Chinese but different in pitch values. In Tianjin Mandarin, "ma" is realized as a falling tone [8][15] and serves multiple functions. First, it can function as a modal particle, indicating that something is taken for granted or used as an exclamation to express surprise or unexpectedness [15]. Secondly, "ma" serves as the counterpart of "$shen^2ma^0$ *what*" in Mandarin, functioning as both a pronoun and a deixis [7]. Lastly, "ma" also forms specific discourse markers like "nei^4 ma^4 *well*" in Tianjin Mandarin.

Due to the elusive nature of modal particles and the profound connection between their usages and their pragmatic meanings, there is no formal basis in syntactic structure to grasp their substantive meanings. In addition, as for other categories, previous research has often explored the form and meaning from semantic, pragmatic, and syntactic perspectives but has rarely delved into phonetic characteristics.

The Tianjin Mandarin particle "ma" is unique in its ability to assume multiple functions. This study aims to (1) investigate primary acoustic parameters, (2) figure out the prosodic features that can distinguish between different functions, and (3) describe the possible prosodic realizations of various functions. Taking the Tianjin Mandarin "ma" as an example, this study explores the correspondence between form and function from a prosodic perspective.

2. Method

2.1. Materials

This designed material includes four categories: deixis, pronoun, modal particle, and discourse marker. The preceding and followed contexts of target words are considered, controlling the potential impact of phonetic environments on the target words. In total, the stimuli include 218 sentences consisting of 245 target words. Samples of materials are shown in Table 1.

2.2. Participants

A total of 8 female participants, aged between 18 and 60, were recruited. All participants were born and raised in Tianjin and frequently engaged in dialectal communication in their daily interactions. At least two generations of their families are from the urban area of Tianjin, and they never left Tianjin for more than three months. All their pronunciations passed the pre-pronunciation test, with authentic dialects and no unclear speech.

2.3. Procedure

Each participant was required to read materials presented on a screen at a natural and fluent speed in a quiet room. The recording equipment was a Plantronics Blackwir3220 Series

10.21437/Interspeech.2024-1909

headset microphone connected to a computer. Praat [16], set at a sampling of 44.1 kHz and 16-bit resolution, was used to record the audio samples and stored them as mono tracks.

Table 1： *Sample of Materials*

Word Category	Stimuli
Deixis	Ta^{21}men hao^{24}xiang^{53} zai^{53} ting^{21} **ma** sheng^{21}.
	(It seems they are listening to some sound.)
	Ta^{21}men hao^{24}xiang^{53} zai^{53} liao^{45} **ma** ren^{45}.
	(It seems they are talking about some people.)
	Ta^{21}men hao^{24}xiang^{53} zai^{53} mai^{24} **ma** shui^{24}.
	(It seems they are purchasing some beverages.)
	Ta^{21}men hao^{24}xiang^{53} zai^{53} wen^{53} **ma** shi^{53}.
	(It seems they are asking some questions.)
Pronoun	Ta^{21}men hao^{24}xiang^{53} zai^{53} shuo^{21} **ma**.
	(It seems they are speaking about something.)
	Ta^{21}men hao^{24}xiang^{53} zai^{53} liao^{45} **ma**.
	(It seems they are chatting about something.)
	Ta^{21}men hao^{24}xiang^{53} zai^{53} jiang^{24} **ma**.
	(It seems they are talking about something.)
	Ta^{21}men hao^{24}xiang^{53} zai^{53} wen^{53} **ma**.
	(It seems they are asking about something.)
Modal Particle	Ni^{24} he^{45} wo^{24} yi^{53}qi^{24} cai^{21} **ma**!
	(Guess with me, please!)
	Ni^{24} he^{45} wo^{24} yi^{53}qi^{24} lai^{45} **ma**!
	(Come with me, please!)
	Ni^{24} he^{45} wo^{24} yi^{53}qi^{24} xie^{24} **ma**!
	(Come with me, please!)
	Ni^{24} he^{45} wo^{24} yi^{53}qi^{24} zuo^{53} **ma**!
	(Do it with me, please!)
Discourse Marker	Nin^{45} ya, na^{53}**ma**, nin^{45} xian^{21} xie^{21}huir^{53}.
	(Well, you can rest for a while.)

2.4. Data Extraction

After recording, all the audio materials are automatically segmented by xSegmenter [17] and manually corrected in Praat. f_0 (in Herz) contour over vowels of the target word, together with the minimum pitch, maximum pitch, duration, and mean intensity, are extracted through the Praat script. Min-max normalization was used to normalize duration, mean intensity, and pitch-related parameters. Specifically, pitch was normalized on a scale from 0 to 5, resembling the tonal system of Tianjin Mandarin. In a well-designed corpus with normalized speaker data, one-way ANOVA and Tukey's HSD *post-hoc* tests were applied for data analysis. f_0 contours were obtained by taking 10 points (in Hertz) at even intervals from each target word. To differentiate f_0 contours produced under different functions, the 10-dimensional f_0 points should be visualized in a 2-dimensional figure. Therefore, the t-SNE algorithm was performed. After that, K-means was used to cluster these 2D points of f_0 contours and help find the specific prosodic representation.

3. Results and Discussion

3.1. Duration

The violin plot is used to present the distribution of normalized duration across the four categories – deixis, pronoun, modal particles, and discourse marker. Each violin plot depicts the data distribution within a specific category, with wider areas indicating higher data density. The median values for each category are denoted by black dots within the violins. To enhance clarity, outliers beyond 1.5 times the interquartile range (IQR) are not displayed in the violin plot but are included in the analysis.

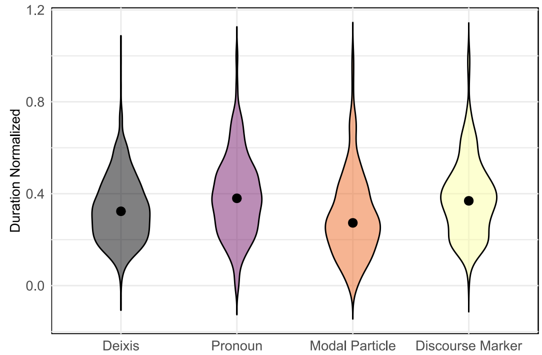

Figure 1: *Violin Plot of normalized duration*

Figure 1 shows that in Tianjin Mandarin, when "ma" functions as a modal particle, it exhibits the shortest duration, as indicated by the lowest median point and the most concentrated distribution of the violin plot in the lower segment.

The result of the one-way ANOVA test revealed a significant effect of categories on duration (F (3, 1830) = 21.73, p < 0.001). To further compare between groups to identify which specific pairs of groups exhibit significant differences, as well as to quantify the magnitude of these differences and their statistical significance level, Tukey's HSD test [19] is used in this study, and results are shown in Table 2.

Table 2: *Tukey's HSD Post-Hoc Test Results for Multiple Comparisons of Duration*

Comparison	diff	lwr	upr	p-value
Pron-Deixis	18.675	11.087	26.262	0.000
MP-Deixis	-4.973	-14.169	4.224	0.506
0DM-Deixis	13.832	2.867	24.797	0.007
MP-Pron	-23.647	-32.604	-14.690	0.000
DM-Pron	-4.843	-15.608	5.922	0.654
DM-MP	18.805	6.851	30.758	0.000

Notes: Prom-pronoun; MP-modal particle; DM-discourse marker

The p-values in Table 2 indicate that duration can effectively differentiate between most pairs. However, no significant difference was observed between pronoun and discourse marker (p > 0.05) as well as modal particle and deixis (p > 0.05), implying a respective potential similarity in their prosodic patterns concerning duration.

3.2. Intensity

According to Figure 2, "ma" as a modal particle signifies the lowest intensity, as evidenced by the lowest median point and the most constrained distribution in the lower half of its violin plot. Conversely, when "ma" functions as deixis, pronoun, or discourse marker, it demonstrates noticeably stronger intensity. Notably, "ma" exhibits more stable shapes as a deixis and a discourse marker, as indicated by the violin plots being most constrained at the upper half under these categories. However, its representation as a pronoun is characterized by a narrow and elongated distribution, suggesting a different pattern of prosodic form.

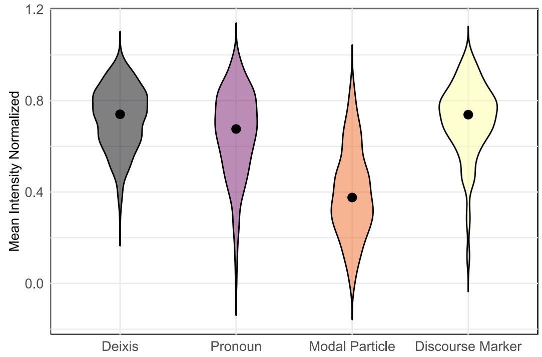

Figure 2: *Violin plot of normalized mean intensity*

A one-way ANOVA test is performed. It demonstrated a significant effect of word categories on speakers' mean intensity of "ma" in Tianjin Mandarin (F (3, 1830) = 242.6, p < 0.001). Subsequent *post-hoc* tests using Tukey's HSD revealed significant differences among the word categories, as shown in Table 3.

Table 3: *Tukey's HSD Post-Hoc Test Results for Multiple Comparisons of Intensity*

Comparison	diff	lwr	upr	p-value
Pron-Deixis	-2.274	-3.147	-1.401	0.000
MP-Deixis	-10.547	-11.605	-9.489	0.000
DM-Deixis	-0.227	-1.488	1.035	0.967
MP-Pron	-8.274	-9.304	-7.243	0.000
DM-Pron	2.047	0.809	3.285	0.000
DM-MP	10.321	8.946	11.696	0.000

The p-values in Table 3 show that intensity can be used as a distinguishing feature between nearly each pair of categories. However, the comparison between discourse marker and deixis did not reveal a statistically significant difference, suggesting a potential similarity in their prosodic forms in mean intensity.

3.3. Pitch

3.3.1. Pitch Range

As shown in Figure 3, "ma" displays the narrowest pitch range when it serves as a modal particle, as evidenced by the lowest median point and the most constrained distribution in the lower half of its violin plot.

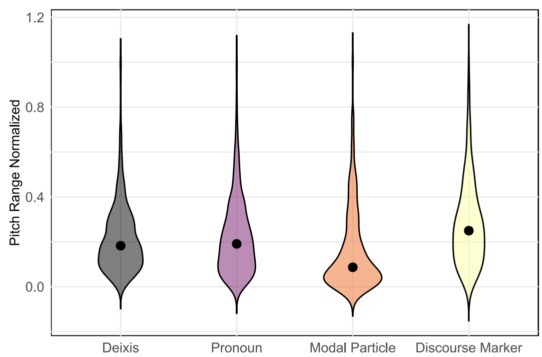

Figure 3: *Violin plot of normalized pitch range*

When "ma" functions as a discourse marker, the speakers' pitch range dispersion is reflected in its violin plot, which appears to be the most elongated shape.

The results of one-way ANOVA revealed a significant effect of word categories on pitch range (F (3, 1830) = 19.03, p < 0.001). Post-hoc tests using Tukey's HSD, as shown in Table 4, indicated significant differences between multiple pairs of word categories (p < 0.05), suggesting that the word categories variable significantly influences the dependent variable.

P-values in Table 4 display that pitch range is a discriminative factor across almost all category pairs. Nonetheless, the comparison between pronoun and deixis did not yield a statistically meaningful distinction, indicating a possible similarity in their prosodic characteristics related to pitch range.

Table 4: *Tukey's HSD Post-Hoc Test Results for Multiple Comparisons of Pitch Range*

Comparison	diff	lwr	upr	p-value
Pron-Deixis	2.224	-3.378	7.827	0.737
MP-Deixis	-11.597	-18.387	-4.806	0.000
DM-Deixis	13.420	5.323	21.516	0.000
MP-Pron	-13.821	-20.435	-7.207	0.000
DM-Pron	11.195	3.246	19.144	0.002
DM-MP	25.016	16.190	33.842	0.000

3.3.2. Pitch Contour

In this analysis, t-SNE (t-distributed Stochastic Neighbor Embedding) [19] was applied to perform two-dimensional (2D) dimensionality reduction, with a perplexity parameter set to 30 and a theta parameter set to 0.5. The resulting 2D coordinates of the data points represent the positions of the data points in the 2D space. After that, the K-means clustering algorithm was applied to the 2D coordinates obtained from t-SNE dimensionality reduction. The algorithm grouped the data points into two clusters based on their spatial proximity in the reduced feature space. Compared with results with a higher number of clusters, this two-cluster result presents a more convincing Silhouette Coefficient (BSS. TSS ~ 0.587). Furthermore, the silhouette scores for the data points generated by the K-means clustering algorithm were calculated, resulting in an overall silhouette score of 0.55. It indicates the effectiveness of K-means clustering on this dataset.

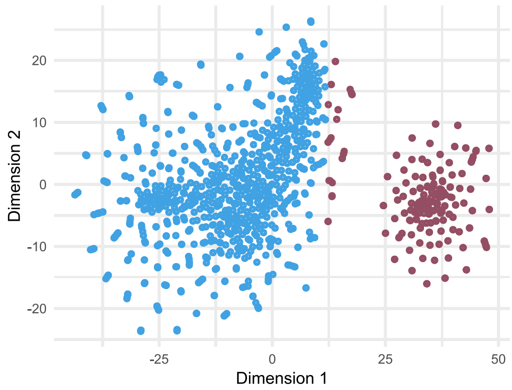

Figure 4: *K-means Clustering of 2D f_0*

From Figure 4, it can be observed that k-means successfully divided the f_0 contour into two clusters, indicating that the eight speakers exhibited two distinct prosodic patterns when producing "ma" with different functions.

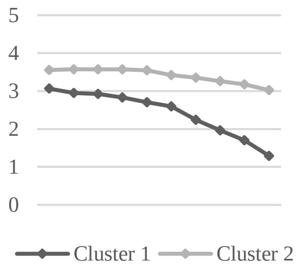

Figure 5: *Cluster f_0 contour*

To further analyze these two prosodic patterns, the average f_0 values were computed based on the clustering results using cleaned data, and the results were plotted as shown in Figure 5, representing the two types of f_0 contour. The f_0 contour of Cluster 1 exhibits a narrower pitch range and a gentler pitch slope. According to previous studies, it represents a high-falling tone in Tianjin Mandarin [20]. In contrast, the f_0 contour in Cluster 2 displays a broader pitch range and a steeper pitch slope, which can be described as a low-falling tone [20] with a tone value of 31.

Table 5: *K-means Result Distribution*

	Cluster 1	Cluster 2
Deixis	32%	68%
Pron	43%	57%
MP	68%	32%
DM	40%	60%
SUM	44%	56%

Table 5 represents the distribution of K-means results. Both pitch patterns occur across the four functions of "ma." This suggests that while different functions statistically impact speakers' prosodic patterns, the semantic functions and syntactic positions within each category also greatly influence them, leading to this relatively even distribution.

4. Conclusion

This study investigated the acoustic features of the word "ma" in Tianjin Mandarin, focusing on its functions as a deixis, pronoun, modal particle, and discourse marker. The findings contribute to understanding the correlation between form and function in prosodic representation.

The findings of this research show that in Tianjin Mandarin, when "ma" functions as a modal particle, its prosodic features include the shortest duration, weakest intensity, and narrowest pitch range compared with "ma" under other functions. This aligns with previous research on the acoustic characteristics of modal particles, which commonly exhibit a neutral tone.

When "ma" functions as a discourse marker, it exhibits the widest pitch range in Tianjin Mandarin. This may be attributed to the variability in its position within the phrase, as "ma" can occur at both the beginning and the end of a phrase (e.g., "ma^4 wan^2 yir^4 *what?*" & "nei^4 ma^4 *well*"). Further investigation is needed to determine whether syntactic or semantic function differences drive this positional variation.

When "ma" functions as deixis and pronoun, it exhibits similar acoustic features, namely longer duration, higher intensity, and broader pitch range, compared to when it serves as a modal particle. However, the difference lies in the more stable distribution of these three primary acoustic features observed when "ma" functions as deixis. This stability may stem from the fact that "ma" as deixis often functions as an attributive in the sentence, leading to more stable prosodic patterns together with the nouns it modifies. On the other hand, when "ma" functions as a pronoun, it can take on various syntactic roles, such as subject or object, which may introduce further nuances in its prosodic characteristics that warrant further investigation.

The statistical analysis demonstrated a significant association between the three selected acoustic features and the functions of "ma" in Tianjin Mandarin. Results from Tukey's HSD *post-hoc* test indicated that "ma" with the same function could be further subdivided based on specific semantic functions and syntactic position, which may also lead to variations in its prosodic features. Through dimensionality reduction and clustering analysis, the paper identified two pitch patterns of "ma" produced by native speakers: one resembling the high-falling tone and the other resembling the low-falling tone characteristic of Tianjin Mandarin.

The results explain the correspondence between form and function from a phonetic and prosodic perspective. In future research, the classification method should be further refined, and specific syntactic positions and semantic functions within each category should be delineated for detailed discussion. In addition, when analyzing acoustic features, it is possible to expand the scope of the study from individual words to the entire sentence and to figure out the significant differences in its functions by comparing the acoustic characteristics of "ma" within the context of complete sentences.

5. Acknowledgment

The Key Laboratory of Linguistics, Chinese Academy of Social Sciences (Project #2024SYZH001), supports this research.

6. References

[1] C. Gussenhoven, "Intonation and interpretation: phonetics and phonology," *Speech Prosody*, Apr. 2002, doi: 10.21437/speechprosody.2002-7.

[2] M. Steedman, "Structure and intonation," *Language*, vol. 67, no. 2, p. 260, Jun. 1991, doi: 10.2307/415107.

[3] G. R. Guy and J. Vonwiller, "The meaning of an intonation in Australian English," *Australian Journal of Linguistics*, vol. 4, no. 1, pp. 1–17, Jun. 1984, doi: 10.1080/07268608408599317.

[4] Z. Wu., "The balance between form and meaning in language teaching: A study of task-based approach.," *Research on Chinese as a Second Language*, 0, pp. 62–71, 2005.

[5] M. Jiang., "Form of second language vocabulary acquisition: Meaning mapping: Controversies and reflections," *Contemporary Linguistics*, vol. 9, no. 1, pp. 52–67, Jan. 2007. Available: http://www.cqvip.com/QK/82143X/200701/24244066.html

[6] X. Shen., "The characteristics of Chinese meaning from the form and meaning," *Chinese Language Learning*, no. 03, pp. 33–36, 1990.

[7] Y. Chao, *A grammar of spoken Chinese*, vol. 3. 1965.

[8] S. Lv, *"Eight hundred words in modern Chinese."* 1983.
[9] L. Wang, *"Chinese Modern Grammar."* 1985.
[10] X. Cui, "The modal meaning of Chinese modal word '~ ma,'" *Language Teaching and Linguistic Studies*, no. 04, pp. 60–68, 2019.
[11] D. Schiffrin, *Discourse markers*. 1987. doi: 10.1017/cbo9780511611841.
[12] L. J. Brinton, *Pragmatic markers in English: Grammaticalization and discourse functions*. 1996. [Online]. Available: https://ci.nii.ac.jp/ncid/BA27897488
[13] G. Han, *Pragmatic Discourse Marker: Toward bi-directional optimization in communication*. Southeast University Press, 2008.
[14] Z. Xiong and M. Lin, Eds., *Interrogative and non-interrogative usage of modal particle "ma0". National Conference on Man-Machine Speech Communication (NCMMSC7)*, 2003.
[15] S. Qi, *The study of Tianjin Grammar*. Shanghai Jiao Tong University Press, 2020.
[16] P. Boersma, "Praat: Doing phonetics by computer," *Ear And Hearing*, vol. 32, no. 2, p. 266, Mar. 2011, doi: 10.1097/aud.0b013e31821473f7.
[17] Z. Xiong, "xSegmenter: Automatic segment segmentation and annotation tool," *Chinese Journal of Phonetics*, no. 01, pp. 27–34, 2019.
[18] H. Abdi and L. J. Williams, "Tukey's honestly significant difference (HSD) test," *Encyclopedia of Research Design*, vol. 3, no. 1, pp. 1–5, 2010.
[19] A. C. Belkina, C. O. Ciccolella, R. Anno, R. Halpert, J. Špidlen, and J. Snyder-Cappione, "Automated optimized parameters for T-distributed stochastic neighbor embedding improve visualization and analysis of large datasets," *Nature Communications*, vol. 10, no. 1, Nov. 2019, doi: 10.1038/s41467-019-13055-y.
[20] Q. Li, Y. Chen, and Z. Xiong, "Tianjin Mandarin," *Journal of the International Phonetic Association*, vol. 49, no. 1, pp. 109–128, Jul. 2017, doi: 10.1017/s0025100317000287.