Speech Prosody 2016
31 May - 3 Jun 2106, Boston, USA

Interactions of Tone and Intonation in Whispered Mandarin

Li Jiao [1], *Yi Xu*[2]

[1] School of Foreign Languages, Tongji University, Shanghai, China
[2] Department of Speech, Hearing and Phonetic Sciences, University College London, London, UK
jennyjiao8758@126.com, yi.xu@ucl.ac.uk

Abstract

A previous study has found that whispered Mandarin, though still allowing listeners to perceive tones to a certain degree, does not carry acoustic cues that are special to whispered tones. That conclusion, however, was based on data from only one speaker. The present study attempted to verify the earlier finding with data from more speakers, with an additional goal to find out if there are acoustic cues to intonation in whispered Mandarin and whether they interact with tonal cues. Twelve Mandarin speakers produced tonal as well as intonational contrasts in both phonated and whispered speech. Acoustic analyses found that whispered questions had longer duration, greater intensity and shallower spectral tilt than statements. However, a perception experiment with 20 native listeners showed a strong bias toward hearing statement in whispers, so that questions were identified well below chance. Thus the acoustic properties in whispers were countering each other as cues to intonation. There was also an interaction of tone and intonation in whispers in that Tone 2 and question help each other while Tone 4 and question hinder each other in their perceptual identification. Overall, therefore, there do not seem to be special perceptual cues to whispered intonation either.

Index Terms: Whispered Mandarin, perceptual cues of whispered intonation, interactions of tone and intonation

1. Introduction

In a previous study, we found that whispered Mandarin carries sufficient acoustic cues to allow reasonably good perception of tones produced in isolation [1]. However, although tone had significant effects on duration, intensity and spectral tilt in whispers, there were no interactions between phonation and tone, indicating that the tonal cues in whispered utterances were already in phonated speech. In addition, perceptual results of tone identification showed no evidence of enhancing these cues in whispers. Therefore, we concluded that there were no special perceptual cues for whispered tones in Mandarin. The acoustic analysis of that study, however, was done only on one female speaker. There is therefore a need for acoustic analysis of whispered tones by more speakers.

The absence of F_0 in whispered speech affects not only tones, but also intonation, which is also mainly carried by pitch in phonated speech. A previous study shows that statement and question could still be identified well above chance in whispered Dutch, and there were acoustic cues (the second formant, the first formant and intensity) correlating to high and low boundary tones [2]. However, there is yet no study to our knowledge of whispered intonation in a tone language such as Mandarin. In this regard, there is also a question as to whether there is any interaction between tone and intonation. In particular, Tone 2, a rising tone, and Tone 4, a falling tone, may have interesting interactions with statement and question intonations. For example, a previous perception study has found that question intonation is easier for native speakers to identify if a sentence ends with Tone 4, but more difficult to identify if it ends with Tone 2 [3]. In contrast, a recent ERP study shows that words with a low lexical tone, at the end of a question leads to a processing conflict [4]. There is therefore a need to compare identification of both tone in different intonations and intonation in various tonal conditions in phonated utterances. Furthermore, the same questions can be asked about whispered speech where prosodic information is highly degraded.

The present study is therefore an extension of [1], with the inclusion of more speakers and examination of the interaction of tone and intonation in both phonated and whispered Mandarin. Experiment 1 is a comparison of tonal and intonational cues in whispered and phonated speech. Experiment 2 examines the use of these cues for the perception of tone and intonation from both phonation types.

2. Production Experiment

2.1. Participants and speaking materials

Together with the two previous speakers from University of Oxford [1], ten more native speakers of Mandarin from Tongji University (12 speakers in total: 6 males and 6 females, mean age = 20.3 years) took part in the recording.

The reading list was a subset of the list used in [1]. The new list (Table 1) consisted of five sets of syllables with vowel or glide onsets. All of them were in 4-way tonal contrast (hereafter T1-T4). The other controlled factors were intonation (statement, question) and phonation (phonated, whispered). In total, 1920 tone tokens (5 syllables × 4 tones × 2 intonations × 2 phonations × 2 repetitions × 12 speakers) were recorded.

Table 1. *A list of selected syllables for acoustic and perceptual studies.*

Tone \ Vowel		a	e	i	u	y
T1	Character	啊	婀	衣	乌	迂
	Pinyin	*ā*	*ē*	*yī*	*wū*	*yū*
	Glossary	*oh*	*graceful*	*clothes*	*black*	*winding*
T2	Character	啊	鹅	姨	无	鱼
	Pinyin	*á*	*é*	*yí*	*wú*	*yú*
	Glossary	*eh*	*goose*	*aunt*	*nothing*	*fish*
T3	Character	啊	恶	椅	五	雨
	Pinyin	*ǎ*	*ě*	*yǐ*	*wǔ*	*yǔ*
	Glossary	*what*	*nausea*	*chair*	*five*	*rain*
T4	Character	啊	饿	意	物	玉
	Pinyin	*à*	*è*	*yì*	*wù*	*yù*
	Glossary	*ah*	*hungry*	*meaning*	*thing*	*jade*

doi: 10.21437/SpeechProsody.2016-20

2.2. Recording Procedures

All these syllables were recorded without a carrier and in two sentence intonations: statement and question. In each recording session, two speakers (a male and a female) sat side by side in the booth and performed a dialogue, with one saying the monosyllabic word as a question, and the other saying the same word as an answer. They then rotated their question-answer roles between trials. The recordings were done with an Audio-Technica AT4031 microphone in Oxford and a Neumann U87 microphone in Tongji. Each microphone was on a stand between the two speakers, with a distance of 15 cm from each. The stimuli (characters and corresponding pinyin) were presented on a screen inside the sound booth. The experimenter monitored the recording and controlled the progression of the recording outside the booth. The input volume of the recording was set to be the same for phonated and whispered registers, and was neither too loud for the phonated register nor too soft for the whispered register [1]. In Oxford, the sounds were recorded onto a Compact Disk by a CD recorder (HHB CDR-850) at 44.1 kHz and 16 bits resolution, and then re-recorded into a PC using a Sound Blaster analogue to digital conversion. In Tongji, the audio was recorded by Pro Tools 8.1 and saved in .wav form at a rate of 24 bits and a sampling frequency of 44.1 kHz.

2.3. Measurements

The analysis was done with a modified version of ProsodyPro, a Praat script for large-scale prosody analysis [5]. With the script we obtained the following measurements: Duration (ms), Intensity (dB), Spectral center of gravity (COG), Hammarberg index (Difference between the energy in the 0-2kHz and 2-5kHz bands [6]), Energy below 500 Hz and Energy below 1000 Hz.

2.4. Analysis and Results

In the analysis, we tried to address the following questions: 1) Are the acoustic patterns of the twelve speakers consistent with those of one speaker in [1]? 2) Does phonation type influence intonations in production? 3) Is there any acoustic interaction among tone, intonation and phonation?

Table 2. *Repeated Measures ANOVAs of significant acoustic data from 12 speakers.*

Measurements	Variables	DF	F-Value	P-Value
duration	phonation	1, 11	96.66	< 0.0001
	tone	3, 33	46.65	< 0.0001
	intonation	1, 11	17.12	< 0.005
	intonation*tone	3, 33	30.16	< 0.0001
	phonation*tone	3, 33	3.53	0.03
intensity	phonation	1, 11	586.61	< 0.0001
	tone	3, 33	18.86	< 0.0001
	intonation	1, 11	16.92	< 0.005
	intonation*tone	3, 33	3.05	0.04
	phonation*tone	3, 33	14.39	< 0.0001
Hammarberg index	phonation	1, 11	60.7	< 0.0001
	intonation*phonation	1, 11	9.56	0.01
	phonation*tone	3, 33	4.14	0.01
COG	phonation	1, 11	57.47	< 0.0001
	intonation	1, 11	10.11	0.01
	intonation*phonation	1, 11	8.23	0.02
energy below 500 Hz	phonation	1, 11	87.34	< 0.0001
	tone	3, 33	3.68	0.02
	intonation	1, 11	12.94	< 0.005
	phonation*tone	3, 33	3.76	0.02
energy below 1000 Hz	phonation	1, 11	72.88	< 0.0001
	tone	3, 33	3.1	0.04
	intonation*phonation	1, 11	5.88	0.03

To answer the first question, we compared results between [1], which involved only one speaker, and the current study with 12 speakers in total. Note that in [1] we did normal ANOVAs because there was only one speaker. Here with twelve speakers, Repeated Measures ANOVAs were performed, which had greater statistical power.

Results from Table 2 showed that there is a main effect of phonation on all measurements, which is consistent with [1] (duration:whispered>phonated; intensity: phonated>whispered; Hammarberg Index: phonated>whispered; energy below 500 Hz: phonated>whispered; energy below 1000 Hz: phonated>whispered). There are also significant effects of tone on duration (T3>T2>T1>T4) and intensity (T4>T1>T2>T3), which is also consistent with [1]. In terms of spectral tilt, tone had significant main effects on energy below 500 Hz (T2>T1>T4>T3) and below 1000 Hz (T2>T1>T3>T4). In [1], however, there were significant main effects of tone on Hammarberg Index (T3>T2>T1>T4) and energy below 500 Hz (T3>T2>T1>T4). A separate set of Repeated Measures ANOVAs on whispers only, however, showed a significant effect of tone only on duration [($F(3,33) = 18.00$, $p < 0.0001$); T3>T1>T4>T2], which is again consistent with [1].

Furthermore, different from [1], there are significant interactions between phonation and tone in terms of duration, intensity, Hammarberg Index and energy below 500 Hz, as can be seen in the interaction plots in Fig. 1. We found that differences among the four tones were much smaller in whispered than in phonated conditions, which implies that the acoustic cues of tones in phonated utterances may have been slightly weakened in whispers. It is consistent with previous findings in [1] and [7].

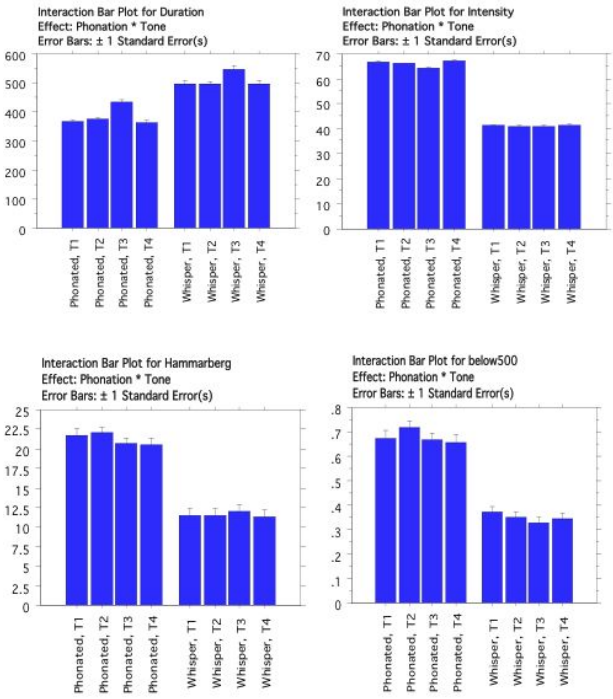

Figure 1: *Interaction between phonation and tone for duration (upper left), intensity (upper right), Hammarberg Index (lower left) and energy below 500 Hz (lower right) in the current study.*

With regard to the second question, Table 2 shows that there is a main effect of intonation on duration (question>statement), intensity (question>statement), COG (question>statement) and energy below 500Hz (statement>question). However, there are interactions between intonation and phonation only in terms of

Hammarberg Index (phonated: question>statement; whispered: statement>question), COG (phonated and whispered: question>statement) and energy below 1000Hz (phonated: question>statement; whispered: statement>question). We then performed a separate set of Repeated Measures ANOVAs on whispers only, and found a main effect of intonation on duration ($F(1,11) = 14.28$, $p = 0.003$) (question>statement); intensity ($F(1,11) = 6.14$, $p = 0.03$) (question>statement); Hammarberg Index ($F(1,11) = 6.10$, $p = 0.03$) (statement>question) and energy below 500Hz ($F(1,11) = 7.53$, $p = 0.02$) (statement>question). Therefore, in whispers, questions had longer duration, greater intensity and shallower spectral tilt than statements. But whether these acoustic cues were really helpful would need perceptual evidence.

As for the last question, Table 2 further shows that there is an interaction between intonation and tone only in duration [(statement: T3>T2>T1>T4; question: T3>T4>T2>T1); (T1-T4: question>statement)] and intensity [(statement and question: T4>T1>T2>T3); (T1-T4: question>statement)]. As shown in Fig. 2, durations (left) of all tones in questions are significantly longer than those in statements. In particular, Tone 4 is the shortest in statements but the second longest in questions. Although intensities (right) of all tones in questions are higher than in statements, the patterns across the four tones are the same in the two intonation conditions. Here again, we did a separate set of Repeated Measures ANOVAs on whispers. Results show that there is an interaction between intonation and tone only in terms of duration ($F(3,33) = 15.73$, $p < 0.0001$), because there were different cross-tone duration patterns in statement (T3>T1>T2>T4) and in questions (T3>T4>T1>T2). In other words, duration of T4 was greatly lengthened in questions. But again, whether this "enhancement" would help identification awaits perception results. Additionally, durations of questions were longer than those of statements in all tones.

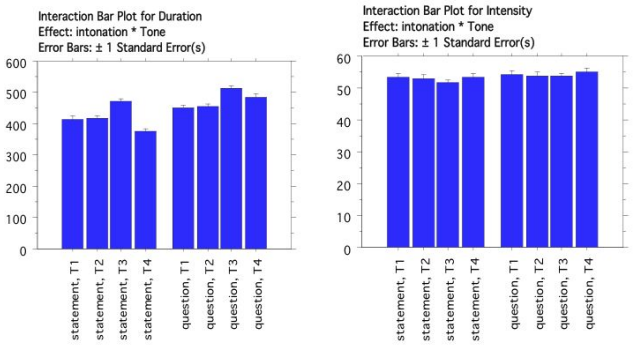

Figure 2: *Interaction between intonation and tone for duration (left) and intensity (right).*

3. Perception Experiment

Following [1], we further investigated how well Mandarin listeners identify tones produced in different intonations from both phonated and whispered speech, and how well they identify intonations in different tones. Our aim was to know whether the acoustic cues of intonation and tone found in experiment 1 would help or hinder their identification.

3.1. Participants

Twenty native speakers of Mandarin (10 males and 10 females) living in China participated as subjects. All were undergraduate or graduate students with an age range of 18-27 (mean = 20.3 years). They had no self-reported speech and hearing disorders.

3.2. Stimuli

From Table 1, we selected phonated and whispered syllables (/e/, /i/ and /y/) in four tones and two intonations by the female Mandarin speaker from the Oxford group as stimuli. There were two identification tasks. Task 1 was for tone and Task 2 for intonation. For each task, each listener went through a total of 96 trials (3 syllables × 4 tones × 2 intonations × 2 phonations × 2 repetitions).

3.3. Procedure

The experiments were run in a quiet room in Tongji University, Shanghai. Subjects wore Sennheiser PC166 headphones, and were seated comfortably in front of a Dell computer (OPTIPLEX 390). At the beginning of each session, they received instructions and had a short round of practice.

The tests were run with an ExperimentMFC script in Praat [8]. In each trial of Task 1, the subject heard an utterance, and saw on the screen four Chinese characters of the corresponding syllables with four tones. They then pressed the button with the character closest to what they had heard. In Task 2, the choices shown on the screen were "statement" and "question". Each sound was played only once. All the subjects did Task 1 first, but half of them heard the phonated utterances first, while the other half heard the whispered ones first.

3.4. Results

3.4.1. Perception of whispered tones

The identification rates of tone and intonation were examined separately on phonated and whispered stimuli in a set of three-way Repeated Measures ANOVAs. In the following report the effects of vowel will not be discussed, although vowel was one of the independent variables in the analysis. Note that the chance level for tone identification is 25%.

In the phonated condition, there were significant main effects of intonation ($F(1,19) = 15.43$, $p = 0.0009$) and tone ($F(3,57) = 8.27$, $p = 0.0001$). Tones were more easily identified in statements (98%) than in questions (94%). And they were perceived differently [T2 (99%) > T1 (98%) > T4 (95%) > T3 (92%)]. There was an interaction between intonation and tone ($F(3,57) = 4.87$, $p = 0.004$), due to the difference in correct responses to tones in statement [T1 (99%) = T2 (99%) >T3 (98%) >T4 (97%)] and in question [T2 (98.3%) > T1 (97.5%) > T4 (94%) > T3 (85%)] (Fig. 3. left).

In whispers, there was no significant difference between statements and questions. There was a main effect of tone ($F(3,57) = 53.4$, $p < 0.0001$). But the ranking of identification rates [T3 (84%) > T4 (60%) > T2 (32%) > T1 (23%)] was reversed from that in the phonated condition mentioned above. Similarly, there was an interaction between intonation and tone ($F(3,57) = 15.3$, $p < 0.0001$), with the identification rate of T4 dropping by 25% while that of T2 rising by 20% from statement to question (Fig. 3. right).

3.4.2. Perception of whispered intonation

Same as for tone, two separate sets of three-way Repeated Measures ANOVAs on intonation identification were conducted for the two phonation types. Note that here the chance level is 50%.

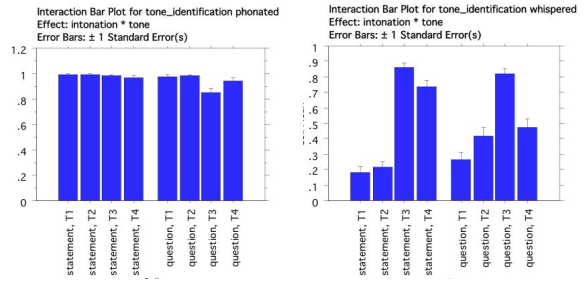

Figure 3: *Interaction between intonation and tone for identification rates of tones in phonated utterances (left) and in whispers (right).*

In phonated speech, there was only a main effect of tone ($F(3,57) = 9.6$, $p < 0.0001$). The difference between identifications of statement (85%) and question (72%) was not significant. There was an interaction between intonation and tone ($F(3,57) = 48$, $p < 0.0001$), which can be seen in Fig. 4: statements [T4 (98.3%) > T1 (97.5%) > T3 (89%) > T2 (53%)] versus questions [T2 (91%) > T4 (80%) > T3 (77%) > T1 (40%)]. Note in particular that questions were particularly hard to identify in phonated T1 (Fig. 4. left).

In whispered speech, there was a main effect of intonation ($F(1,19) = 41.8$, $p < 0.0001$), but not tone. Identification rate was 85% for statements but only 36% for question, which shows a strong bias toward hearing statement in whispers. Also there was an interaction between intonation and tone ($F(3,57) = 6.8$, $p = 0.0005$). As shown in Fig. 4 right, the overall identification rates were T1 (91%) > T4 (90%) > T2 (81%) > T3 (78%) for statements and T2 (46%) > T3 (40%) > T1 (32%) > T4 (26%) for questions. In all four tones questions in whispers were hard to identify (all below chance). They were best perceived in T2 but worst perceived in T4 (Fig. 4. right).

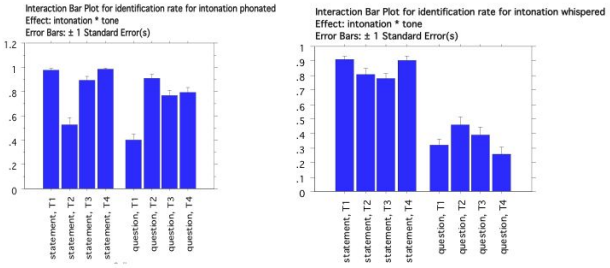

Figure 4: *Interaction between intonation and tone for identification rates of intonations in phonated utterances (left) and in whispers (right).*

4. General discussion and conclusions

The goal of this study is to both verify the general finding of a previous study [1], and examine the interaction of tone and intonation in phonated and whispered speech through both acoustic analysis and perceptual identification.

Like in [1], we found significant main effects of both phonation (duration, intensity, Hammarberg Index, energy below 500 Hz and 1000 Hz) and tone (duration, intensity, energy below 500 Hz and 1000 Hz). When whispered speech was analyzed separately, a significant effect of tone was found only on duration, just like in [1]. Also, acoustic differences of tones were smaller in whispers than in phonated speech.These results have verified our earlier finding that there are no special perceptual cues to whispered tones [1].

In terms of intonation, we found that whispered questions had longer duration, greater intensity and shallower spectral tilt than statements. But perception results showed a bias toward statement (85%), as recognition rate for questions was well below chance (36%). Therefore, the acoustic differences due to intonations in whispers did not provide helpful cues, as they were likely to have countered each other.

Finally, we found interesting interactions of tone and intonation in whispers. Acoustically, T4 had greatly increased duration in whispered question, but its identification rate actually dropped significantly for questions. T2 showed no special acoustic cues in whispered questions in terms of the measurements analyzed in this study, but its identification rate in questions increased significantly. In contrast, although questions had greater duration and intensity than statements in both T2 and T4, they were best perceived in T2 but worst in T4 in whispers. So it seems that T2 and question helped each other while T4 and question hindered each other in their perceptual identification. Overall, therefore, there do not seem to be special perceptual cues to whispered intonation either.

One may question the reliability of the perceptual results given that they were based on stimuli from only one speaker. So we compared the ANOVA results of the acoustic analysis on this speaker with those of the Repeated Measures ANOVA on all the 12 speakers. The significant results turned out to be all in the same directions. It is therefore likely that stimuli from one or more of the other speakers would have generated similar perceptual results.

One limitation of the current study is the lack of formant analysis on all 12 speakers. The main reason was the difficulty of accurate formant extraction from whispered speech. So one direction of future work is to develop an effective method of formant analysis on whispers. Also there is a need to investigate if there is any effect of vowel on tone and intonation in whispered speech.

In summary, the acoustic analysis of twelve Mandarin speakers in this study has confirmed previous finding of lack of specially developed perceptual cues for whispered tones. Furthermore, acoustic analysis of both phonated and whispered intonation also found no special cues for the perception of whispered intonation, as the few significant acoustic differences between whispered and phonated intonation appeared to have hampered rather than enhanced the correct identification of question intonation in whispers.

5. Acknowledgements

This work was supported by the International Exchange Program for Graduate Students, Tongji University (No. 201601020). The authors would like to thank Professor Qiuwu Ma, Professor Daniel Hirst, Professor Jie Liang and Dr. Ting Wang for their help in Tongji University. Also we are very grateful of participants in University of Oxford and Tongji University. Our special thanks go to the two anonymous reviewers' comments and suggestion. And we are indebted to Dr. Marjoleine Sloos and Mr. Chunan Qiu for their statistical help although we did not present these results in this paper.

6. References

[1] L. Jiao, Q. W. Ma, T. Wang, and Y. Xu, "Perceptual cues of whispered tones: Are they really special?" in *INTERSPEECH 2015, September 6–10, Dresden, Germany, Proceedings*, 2015, pp. 2361–2365.

[2] W.F.L. Heeren and V.J.Van Heuven, "Perception and production of boundary tones in whispered Dutch," in *INTERSPEECH 2009, September 6–10, Brighton, UK, Proceedings*, 2009, pp.2411–2414.

[3] J. H. Yuan, "Perception of intonation in Mandarin Chinese," *Journal of Acoustical Society of America*, vol. 130 (6), pp.4063-4069, 2011.

[4] C. Kung, D. J. Chwilla and H. Schriefers, "The interaction of lexical tone, intonation and semantic context in on-line spoken word recognition: An ERP study on Cantonese Chinese," *Neuropsychologia*, vol. 53, pp. 293–309,2014.

[5] Y. Xu, "ProsodyPro — A tool for large-scale systematic prosody analysis," *in Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013), August 30, Aix-en-Provence, France*, 2013, pp. 7-10.

[6] B. Hammarberg, B. Fritzell, J. Gauffin, *et al.*, "Perceptual and acoustic correlates of abnormal voice qualities," *Acta Otolaryngologica*, vol. 90, pp. 441-451, 1980.

[7] C. Chang and Y. Yao, "Tone production in whispered Mandarin," in the *16th International Congress of Phonetic Sciences, August 6-10, Saarbrücken, Germany, Proceedings*, 2007, pp. 1085-1088.

[8] P. Boersma, "Praat, a system for doing phonetics by computer," *Glot International*, vol. 5, pp. 341-345, 2001.