COPiL CAMBRIDGE OCCASIONAL PAPERS *in* LINGUISTICS
Volume **16**, Article 4: pp. 113–126, 2024 | ISSN: 2050-5949

The Role of the Duration Cue and Tonal Contour in Lexical Tone Recognition in Whispered Mandarin Chinese*

Yueke Zeng
University of Cambridge

1 Introduction

Whispering is a normal way of communication, with which people are able to reduce the audibility of the utterance they are producing while still being able to transmit enough information for the listeners to receive. In whispered speech, the glottis remains open, and the vocal folds do not vibrate, making the periodic voice source produced from the vibration absent, and thus no fundamental frequency is present in the utterance.

The absence of fundamental frequency might be expected to cause great difficulties in understanding whispered utterances of tonal languages like Mandarin Chinese, which rely heavily on the pitch contour to distinguish lexical tones. However, both native speakers' smooth whispered communication and academic investigations have proved this statement to be less plausible.

Early studies of whispered Mandarin have supported the identifiability of whispered Mandarin tones (Miller 1934), but researchers attribute this to the help of context, ruling out the possibility of comprehending genuinely whispered, context-free words. In contrast to this statement, Giet (1955) asserts that isolated words are identifiable in tones, though less clearly compared to the phonated ones, and changes in vowel color and the variation of airflow could be the cues compensating for the absence of fundamental frequency. This possibility is supported by Meyer-Eppler's (1957) spectrographic evidence on whispered German intonation.

A series of empirical studies on whispered tone perceptibility, making use of minimal pairs in tone, have also supported the idea of Giet, having Mandarin native speakers successfully identify whispered four tones with a rate from 62% (Wise & Chong 1957) to 85% (Jensen 1958). But the intelligibility of whispered tones in other tonal languages such as Thai (Abramson 1972) and Vietnamese (Miller 1961) turns out to be less promising, only presenting a rate of correctness slightly above chance, making context still an indispensable factor that can contribute to distinguishing different tones in whispered words and speech. To summarize from now, early research on whispered tone identification shows that despite the contribution that context can make, context-free whispered tone identification is still possible,

* Many thanks to Francis Nolan, my supervisor, who has provided enormous support for the completion of this paper. Also many thanks for all the participants who have contributed their precious time and efforts for the experiment.

©2024 Zeng
This is an open-access article distributed by the Section of Theoretical & Applied Linguistics, Faculty of Modern and Medieval Languages and Linguistics, University of Cambridge under the terms of a Creative Commons Non-Commercial License (creativecommons.org/licenses/by-nc/3.0).

though not perfectly clear, and this suggests the existence of acoustic correlates that can assist in identifying whispered tones without fundamental frequency.

A subsequent series of studies on whispered tone recognition in Mandarin Chinese have been attempting to examine the acoustic correlates that would affect the recognition of whispered tones. Jiao & Xu (2019) conducted a relatively comprehensive investigation on the acoustic cues in whispered tones, focusing on whether these cues already exist in phonated speech or not. Their production experiment covers cues like duration, intensity, spectral center of gravity, and formant frequencies, showing no significant phonetic enhancement in whispered speech compared to phonated speech. Their perception experiment, making use of human-produced whispered speech and amplitude-adjusting phonated speech in which fundamental frequency is removed, also shows that the non-f_0 cues for whispered tone perception already exist in phonated speech, as the recognition patterns for the two sets of speech are quite similar, with whispered Tone 3 and Tone 4 being well recognized and the other two tones just slightly above chance. Though Jiao & Xu's study has been convincing in showing no special cues to aid whispered tone perception, it does not delve into what the specific contributions of these cues are to whispered speech recognition: are they really useful for the listeners to tell apart different tones? How far they play a role in phonated as well as whispered tone perception? To unravel these questions, the recognition patterns of different Tones are actually an ideal starting point.

In their study, Jiao & Xu found a significantly higher recognition rate in Tone 3 and Tone 4, which is in line with many other experiments. Gao (2002) also found Tone 3 and Tone 4 the easiest to identify, with a rate of correctness of 90.8% and 74.7%, respectively. Kong & Zeng (2006) have a similar finding, with Tone 3 being the most identifiable one and Tone 1 the least; performance between Tone 2 and Tone 4 presents no significant difference. The result of McLoughlin, Xu & Song (2014) is slightly different, with Tone 3 being the most recognizable one, followed by Tone 2, and with Tone 4 and Tone 1 being the most difficult ones to identify. Though different in their ranking of the other three tones, all the above-mentioned studies have agreed that Tone 3 is the most identifiable one among the four lexical tones of Mandarin Chinese. To account for this, Jiao & Xu mentioned the extra-long duration of Tone 3, indicating that this could be an acoustic cue that helps the participants to identify it as Tone 3. Kong & Zeng take a similar way to explain their result: with Tone 3 being the longest one and Tone 1 or Tone 4 being the shortest, participants would tend to regard a longer duration as being Tone 3 while a shorter one being Tone 1 and Tone 4. They also mentioned another possibility that with Tone 3 being the most 'dynamic' one and Tone 1 being the 'static' one, the cues that could mimic the dynamic properties of the missing f_0 information would be most viable for Tone 3 and least for Tone 1, and this may contribute to the easier recognition of Tone 3. Still, another role that duration could play in whispered tone recognition has also been raised. According to Gao, although the duration cue does not make the primary contribution in whispered tone identification, it does partly assist the better recognition of Tone 3, the inherently longest tone, as this allows listeners to have a better chance to perceive the information, they need to decide

the tone. In his experiment on whispered Thai, Abramson (1972) also suggests that a longer duration could lead to better tone recognition.

Following the steps of previous research, the current study is going to discuss a possible reason for Tone 3's easy recognition in whispered Mandarin, focusing on the duration cue and the dipping tonal contour. By using the term 'tonal contour' instead of 'pitch contour', the current paper intends to emphasize the fact that whilst f_0 is absent, other dimensions, like amplitude, may survive to perform the contour pattern of different tones. With this experiment, the essay hopes to set a starting point to dive deeper into the exact function of different cues in whispered tone recognition.

2 Methodology

This study makes use of a perception experiment to explore the role that duration and tonal contour play in distinguishing whispered tones in Mandarin. The experiment is divided into two parts, with the first one focusing on the perception of the five monophthongs that can form a syllable independently in Mandarin: [a], [ɤ], [i], [u], and [y] (see Table 1 for the word list) and the second one focusing on the sets of disyllabic words that only differ in the tone of the first syllable (See Table 2 for the word list). The experiment is designed to investigate Mandarin native speakers' ability in identifying tones in context free whispered conditions, paying special attention to whether some tones are easier to recognize in whisper than others or not, and to test two of the possible explanations for Mandarin native speakers' divergent performance perceiving different whispered tones, as proposed by previous research: duration and tonal contour.

2.1 Materials

2.1.1 Vowels

The word list for vowels (Table 1) follows the one that Jiao & Xu used in their experiment. Considering the need to include a full tonal contour, using vowels to carry all the investigated tones suits best for the experiment purpose, and would lead to the fewest side effects when adjusting syllable duration. Audio of the phonated and whispered vowels in Table 1 will be used as stimuli to investigate Mandarin native speakers' ability to perceive whispered tones as well as to compare their performance in distinguishing the four tones. The audio was recorded in a sound booth by the researcher, a 21-year-old Mandarin native speaker who does not have an obvious accent when speaking standard Mandarin.

Furthermore, to find out the role of syllable duration in whispered tone perception, a Praat script was written to equalize the duration of all the whispered vowels to the average value of all the tokens (0.5709s). The duration of the whispered tokens is presented in Table 3, which shows that Tone 3 is the longest of the four tones in all the vowels, which is the same case as in phonated tones (Shi 1991, Liu 1926).

Tone		Vowel				
		a	ɤ	i	u	y
Tone 1 level	Character	啊	婀	衣	屋	迂
	Pinyin	ā	ē	yī	wū	yū
	Glossary	Oh	Graceful	Clothes	House	Winding
Tone 2 rising	Character	啊	鹅	姨	无	鱼
	Pinyin	á	é	yí	wú	yú
	Glossary	Eh	Goose	Aunt	Nothing	Fish
Tone 3 dipping	Character	啊	恶	椅	五	雨
	Pinyin	ǎ	ě	yǐ	wǔ	yǔ
	Glossary	What	Nausea	Chair	Five	Rain
Tone 4 falling	Character	啊	饿	意	物	玉
	Pinyin	à	è	yì	wù	yù
	Glossary	Ah	Hungry	Meaning	Thing	Jade

Table 1 List of vowels used in the perception experiment.

First syllable		Second Syllable			
		Tone 1 level	Tone 2 rising	Tone 3 dipping	Tone 4 falling
Tone 1 level	Character	鸡胸	基石	伊始	包被
	Pinyin	jī xiōng	jī shí	yī shǐ	bāo bèi
	Glossary	chicken's breast	base	begin	cover
Tone 2 rising	Character	吉凶	及时	遗史	薄被
	Pinyin	jí xiōng	jí shí	yí shǐ	báo bèi
	Glossary	good and bad luck	in time	leftover issue	thin quilt
Tone 3 dipping	Character	挤胸	几时	以使	宝贝
	Pinyin	jǐ xiōng	jǐ shí	yǐ shǐ	bǎo bèi
	Glossary	to squeeze breast	when	in order to	baby
Tone 4 falling	Character	继兄	纪实	亦使	报备
	Pinyin	jì xiōng	jì shí	yì shǐ	bào bèi
	Glossary	stepbrother	Documentary	also	report

Table 2 List of disyllabic words used in the perception experiment.

	a	ɤ	i	u	y	Average
Tone 1	0.695	0.573	0.632	0.500	0.485	0.577
Tone 2	0.577	0.574	0.542	0.499	0.548	0.548
Tone 3	0.716	0.626	0.649	0.638	0.723	0.670
Tone 4	0.461	0.489	0.488	0.449	0.449	0.467

Table 3 Duration of the whispered vowel tokens (in seconds).

The manipulated audios, referred to as duration-adjusted audios in the following paper, will be used as the experiment stimuli along with the original whispered and phonated audios to compare participants' correctness when judging the tone of different tokens.

2.1.2 Disyllabic words

Taking advantage of Tone 3's special sandhi pattern, phonated and whispered audio of the disyllabic words in Table 2 will be of great use to investigate the role that tonal contour plays in whispered tone perception.

Tone 3, the dipping tone in Mandarin, presents special sandhi patterns when placed in the first syllable of a disyllabic word. On the one hand, when placed in front of Tone 1, 2, or 4, the low dipping contour (described as '214' using Chao's (1930) five-level tone mark) will change into a low falling one (described as '21' using Chao numbers), which means the rising part of the dipping tone will not be pronounced when Tone 3 is followed by Tone 1, 2 or 4. On the other hand, when followed by another Tone 3, the first Tone 3 will have its dipping pattern altered into a rising one, which is similar to Tone 2 (described as '35' using Chao numbers). Research (Peng 2000, Wang & Li 1967) has suggested that Mandarin native speakers are not able to perceive the difference between an original Tone 2 and an altered Tone 3, but acoustic analysis suggested that there are differences between these two (Zee 1980). Therefore, whether the altered Tone 3 is exactly the same as Tone 2 or not is still a question that requires further discussion. To prevent their possible divergences from affecting the result, the word list includes the pair of 'Tone 3 + Tone 3' and 'Tone 2 + Tone 3' combination and will take the highly possible confusion into consideration when analyzing the result.

As Tone 3 will be changed into a rising or a low falling one when placed in the first syllable of the disyllabic words, it loses its distinctive dipping pattern and stops being the most dynamic tone of the four tones. In this case, according to the assumption of Kong & Zeng (2006) which suggests that being the most dynamic tone can contribute to Tone 3's high identifiability, Tone 3's rate of correctness when being whispered in the first syllable of a disyllabic word could be different

from that in isolated vowels, which may help present the role that tonal contour plays in whispered tone perception.

To sum up, phonated, whispered, and duration-adjusted whispered vowels as well as phonated and whispered disyllabic words will be utilized as the stimuli in the perception experiment to investigate Mandarin native speaker's performance in whispered tone perception.

2.2 Participants and experiment process

Participants, 11 females and 9 males, were recruited via the Internet. All of them are native speakers of Mandarin Chinese and none of them has self-reported hearing disorders.

To prevent participants from becoming familiar with the individualized acoustic habits belonging to the speaker in the phonated token, which could in turn facilitate participants' recognition of the corresponding whispered counterpart, the audio clips were divided into two balanced sets, with phonated and whispered clips of the same vowel or word necessarily belonging to different sets. All participants were randomly divided into two groups of 10 each and listened to two different sets of materials (see Appendix for the list of the stimuli that each group got). At the beginning of the experiment, participants had three trial questions for practice. The first part of the experiment of each group consisted of 30 vowels, 10 phonated, 10 whispered, and 10 whispered duration-adjusted. All the questions were randomized. Participants were required to decide the tone of the vowel in the clip they heard. The second part contained 16 words for each group, 8 phonated and 8 whispered. Participants were expected to listen to the clip and chose the word they had heard from the four options provided, which are a set of words that only vary in the tone of the first syllable.

3 Results and analysis

3.1 Overall performance of phonated and whispered tone recognition

Participants' responses to all the questions were collected and reorganized to compare their performance in identifying tones under different conditions (see results in Figure 1). Results for the whispered and phonated words were calculated using the average rate of all tokens, regardless of the tone of the second syllable. A more detailed analysis that looks into the performance in different tonal environments will be presented in the following passage.

According to Figure 1, phonated tokens are, without doubt, easily recognized by all the participants, except for Tone 2 and Tone 3 in words, which is caused by the confusion between 'Tone 2 + Tone 3' and 'Tone 3 + Tone 3' pair, as presented in Table 4, revealing that Mandarin native speakers are not able to tell apart the Tone 3 sandhi from an original Tone 2. Therefore, the primary results shown in Table 4 and Table 6 accept the identicality between 'Tone 2 + Tone 3' and 'Tone 3 + Tone 3', but the results rejecting this presumption are also indicated by the numbers in the parentheses.

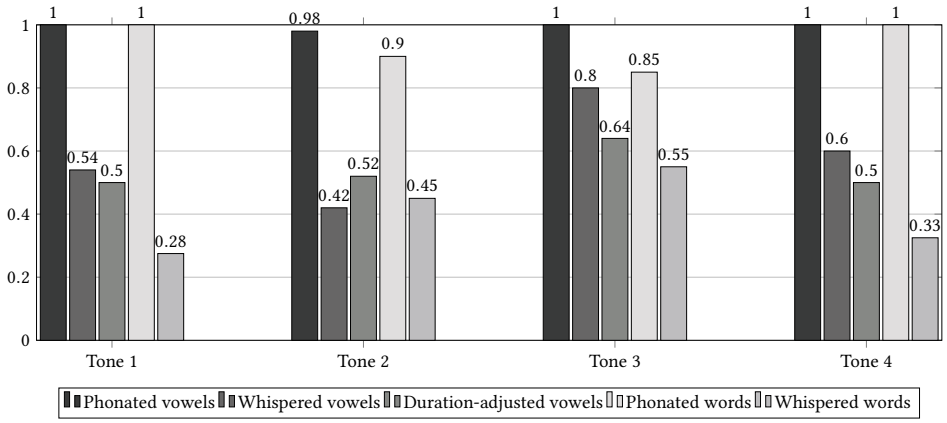

Figure 1 Rate of correctness in different tones.

First tone	Second tone			
	Tone1	Tone2	Tone3	Tone4
Tone1	1	1	1	1
Tone2	1	1	1 (0.6)[1]	1
Tone3	1	1	1 (0.4)	1
Tone4	1	1	1	1

Table 4 Rate of correctness of different tones in phonated words.

The whispered vowels, on the other hand, are shown to be more difficult to distinguish. Though the rate of correctness of the four tones is all comparatively promising compared to the chance rate of 25%, there is still a distance from the near 100% performance in the phonated tokens, and the difference between individuals' performance in the phonated and the whispered tokens is statistically significant for all four tones (ANOVA test shows $p < 0.001$ for all tones and $\alpha = 0.05$), which clearly supports a difficulty in identification in whispered speech. However, despite a lower performance in the other three tones, participants have a better performance in whispered Tone 3, with the rate of correctness being 80%, which is in line with other research's ranking results, suggesting that native speakers of Mandarin do perform better in whispered Tone 3 recognition.

[1] The 0.6 and 0.4 in the parenthesis stands for the rate of correctness calculated based on the assumption that the original Tone 2 and altered Tone 3 in the sandhi are actually different. However, the divergence from 0.5 suggests a slight bias to Tone 2, perhaps because of the match between the sandhi form and citation Tone 2, and mismatch to Tone 3.

3.2 The role of duration in whispered tone recognition

The result for the whispered duration-adjusted tokens helps to reveal the role that duration could play in tone identification. According to Table 3, by adjusting the duration to the average value of 0.5709s, most tokens of Tone 1 and 2 are little changed whereas T3 loses about 15% of its duration, and T4 gains about 22%. Comparing whispered and duration adjusted whispered results in Table 5, the adjustment does the most harm to T3 correct identification and second-most harm to T4 identification. This tendency could support the distinctive role that duration could play in tone identification. However, the correlation coefficient between the rate of duration change (calculated as $\frac{t_{\text{whispered}} - t_{\text{adjusted}}}{t_{\text{whispered}}}$, t stands for the duration of each token) and the rate of correctness change (calculated as $r_{\text{whispered}} - r_{\text{adjusted}}$, r stands for the rate of correctness of each token) does not present a statistically significant and strong correlation, with $r = 0.193$.

Moreover, to find out whether a longer duration can assist tone identification or not, two ANOVA tests were conducted to compare participants' rate of correctness in whispered and duration-adjusted vowels, one for those with the duration-adjusted token longer than the whispered ones and one for those with the duration-adjusted tokens being shorter. Both tests showed no significant difference between the two groups (for the first test, $p = 0.55091$, $\alpha = 0.05$; for the second, $p = 0.24494$, $\alpha = 0.05$). This result does not favor the assumption that the change in duration could assist or obstruct tone identification. To delve deeper into the role that duration can play, a confusion matrix was made to scrutinize participants' tendencies when identifying different tones (Table 5).

Comparing the confusion matrix between the whispered vowels and the duration-adjusted ones, the rate of correctness dropped for Tone 1, 3 and 4 (with a decrease of 4, 16 and 10 percentage points, respectively), while rising for Tone 2 (increased 10 percentage points). When the duration of all the vowels is adjusted to the average value of all tokens, there is a higher tendency for Tone 4, the shortest of all four tones, to be mis-regarded as Tone 3, the longest one, which might be evidence for the suggestion that people tend to regard a longer syllable as Tone 3, but this fails to explain both the fact that it is Tone 1 that is the closest one to the average value in duration and the increased accuracy in Tone 2. To sum up, according to the result in vowels, the role that duration could play in tone identification, though not totally absent, is rather limited. However, the result in this part does show some interesting patterns that are worth exploring, which will be discussed in the final part of this paper.

3.3 The role of tonal contour in whispered tone recognition

The rate of correctness in whispered words gives a hint to the role that tonal contour plays in Tone 3's easy identification. Compared to the whispered vowels the overall rate of correctness of the whispered disyllabic words (see Figure 1) dropped drastically in Tone 1, 3, and 4 (decreased 26.5, 25, and 27.5 percentage points, re-

Said	Heard			
	Tone 1	Tone 2	Tone 3	Tone 4
Tone 1	0.54	0.02	0.02	0.42
Tone 2	0.28	0.42	0.18	0.12
Tone 3	0.04	0.12	0.8	0.04
Tone 4	0.16	0.14	0.1	0.6

(a) Whispered vowels

Said	Heard			
	Tone 1	Tone 2	Tone 3	Tone 4
Tone 1	0.5	0.02	0.08	0.4
Tone 2	0.24	0.52	0.12	0.12
Tone 3	0.16	0.14	0.64	0.06
Tone 4	0.2	0.12	0.18	0.5

(b) Duration-adjusted vowels

Table 5 Confusion matrix of whispered vowels and duration-adjusted vowels.

First Tone	Second tone			
	Tone 1	Tone 2	Tone 3	Tone 4
Tone 1	0.4	0.3	0.4	0
Tone 2	0.4	0.7	0.2(0.4)	0.3
Tone 3	0.4	0.4	0.7(0.9)	0.5
Tone 4	0.3	0.6	0.3	0.1

Table 6 Rate of correctness of whispered words.

spectively). A closer scrutiny of the rate of correctness in different environments helps to find the source of this change.

According to Table 6, while being in the first syllable of a word does make the tone perceptibility drop a little, the tone of the second syllable contributes more. Tone 1 becomes difficult to identify when placed in front of Tone 4, which is the main reason for Tone 1's drop in whispered words. Whispered Tone 2 becomes more identifiable when being with another Tone 2, compensating for the comparatively slight drop when placed in other environments. The drop in Tone 3, on the other hand, comes from its presence in front of Tone 1, 2, and 4, suggesting that the loss of dipping pattern when placed before Tone 1, 2 and 4 does affect the identifiability of Tone 3. As for Tone 4, being placed with itself turns out to cause a major drop in the rate of correctness, but a Tone 2 after it makes the identification less difficult. To summarize the results gathered from this table of correctness, it was revealed that the dipping contour does contribute to a higher recognition rate of whispered Tone 3, as the loss of it when being placed in front of Tone 1, 2, and 4 in a whispered word causes correct identification to drop drastically. Phonated Tone 3 is not affected probably because f_0 is a relatively strong cue to help tone identification, but when it is absent in whispered speech, the remaining cues are not strong enough to sustain the identifiability. It's also worth noticing that the loss of dipping pattern is not the only way to account for the drop in Tone 3 recognition in words: being placed in the first syllable of a word itself might also obstruct the identification, as suggested by the correctness of the other four tones.

Also, the interaction between the tonal environment and whispered tone recognition is more than just the tonal sandhi happening in Tone 3; the other three tones also have interesting performances when being placed with certain tones, which will be further discussed in the next section.

4 Discussion and Conclusion

The current paper investigates Mandarin native speakers' ability to identify whispered tones and tries to testify to the role that syllable duration and tonal contour, two previously suggested reasons for the outperformance in Tone 3 perception, could play in Mandarin native speakers' whispered tone recognition. The result shows that Mandarin speakers are able to distinguish different whispered tones, though not 100% perfectly, and they tend to perform a lot better in Tone 3 compared to the other three tones. Following the steps of previous discussions, the current essay investigated the possible reason for Tone 3's better performance, focusing on duration and tonal contour. Results in the duration-adjusted tokens, compared to the whispered tokens, tend to disfavour duration being the decisive cue for tone identification, as neither the increase nor the decrease of duration brings about a significant difference in the correctness of tone identification, and the confusion matrix of duration-adjusted and whispered tokens presents patterns contradictory to the assumption that duration could be a decisive cue to tell apart different tones. Participants' performance in disyllabic words, on the other hand, could be evidence supporting the contribution of the dipping pattern to Tone 3's easy recognition, but not only losing the dipping pattern but also being followed by another tone would bring about the fall in identification accuracy, as suggested by participants' performance in the other three tones.

However, it is still worth noticing that even though the current essay treats duration and tonal contour as two comparable factors in whispered tone recognition, they are actually at different levels in the whole system: tonal contour is a feature of whispered syllables, which can be realized by a series of acoustic cues, with duration being one of them. Therefore, future investigation may be conducted with acoustic analysis on whispered speech to look for the cues that distinguish one tone from another and changes when the sandhi happens. Some detailed phenomena presented in the confusion matrix of the four tones under different tonal environments could shed light on the possible directions to handle this problem.

Table 5 showed that participants have a strong tendency to mistake Tone 1 for Tone 4, which both happens in whispered and duration-adjusted vowels, and Tone 1 turns out to be trickier to identify when being placed before a Tone 4, with participants misjudging it as a Tone 3 or a Tone 4 (see Table 7).[2] This might be a hint that the cues that mimic the function of fundamental frequency in whispered tone seem to also present a falling tendency when it comes to the end of a syllable, thus

[2] Graph (a) to (d) stands for the confusion matrix of the four tones when being said in the first syllable of a disyllabic word. The cells of each graph stands for the rate of being heard as the tone indicated by the column value while the row stands for the tone second syllable, the fixed one. The numbers in bold stands for the correct answers' rate of being chosen.

Placed before	Heard			
	Tone 1	Tone 2	Tone 3	Tone 4
Tone 1	**0.4**	0.2	0.1	0.3
Tone 2	**0.3**	0.4	0.2	0.1
Tone 3	**0.4**	0.2	0	0.4
Tone 4	**0**	0	0.6	0.4

(a) First syllable spoken as Tone 1

Placed before	Heard			
	Tone 1	Tone 2	Tone 3	Tone 4
Tone 1	0.3	**0.4**	0.1	0.2
Tone 2	0	**0.7**	0.1	0.2
Tone 3	0.4	**0.2**	0.2	0.2
Tone 4	0	**0.3**	0.6	0.1

(b) First syllable spoken as Tone 2

Placed before	Heard			
	Tone 1	Tone 2	Tone 3	Tone 4
Tone 1	0.4	0	**0.4**	0.2
Tone 2	0	0.1	**0.4**	0.5
Tone 3	0.1	0.2	**0.7**	0
Tone 4	0	0	**0.5**	0.5

(c) First syllable spoken as Tone 3

Placed before	Heard			
	Tone 1	Tone 2	Tone 3	Tone 4
Tone 1	0.1	0	0.6	**0.3**
Tone 2	0.1	0	0.3	**0.6**
Tone 3	0.2	0	0.5	**0.3**
Tone 4	0.1	0	0.8	**0.1**

(d) First syllable spoken as Tone 4

Table 7 Confusion matrix of the four tones when placed before different tones.

misleading participants to regard Tone 1, the level tone, as a falling one. Also, the confusion matrix of Tone 4 presents a high preference for Tone 3, even higher than Tone 4 itself, but it never gets mixed up with Tone 2. This shows that participants are precise at telling whether a tone is falling or rising, but might be less capable of distinguishing a low falling tone from a high falling one.

To sum up, the current essay investigates the role that syllable duration and tonal contour could play in whispered tone identification. With experimental evidence, the current research coincides with previous research in supporting that Mandarin native speakers are able to distinguish whispered tones even without context, and perform best in Tone 3 identification. The experiment also suggests that duration does play a limited role in whispered tone recognition, and the loss of dipping pattern of Tone 3 harms the whispered Tone 3 perception a lot. But there still remains much to be explained, such as participants being inclined to mistake whispered Tone 1 for Tone 4 and the logic that connects dipping pattern loss with difficulty to identify Tone 3. The answer to these questions could shed light on the exploration of deciding the alternative acoustic cues for whispered tone recognition when f_0 is absent.

Future studies on acoustic correlates that assist whispered tone recognition could possibly go in two directions: for one thing, analysis comparing the performance of certain acoustic cues like amplitude, formants, and spectral center of gravity in whispered audio along with perception experiments that investigate participants' ability to distinguish whispered tones when the above-mentioned cues are manipulated and messed up can help reveal the mechanism of whispered tone recognition acoustically; for another, this question can also be tackled from an articulatory view: using equipment like laryngograph and airflow meter, movement of speech

organs as well as the speed of airflow produced when whispering can also be measured and investigated to help build a more comprehensive understanding of whispered tone recognition.

References

Abramson, A. S. 1972. *Tonal experiments with whispered Thai.* 31–44. De Gruyter Mouton.

Chao, Y. R. 1930. ə sistim əv "toun-letəz". *Le Maître Phonétique* 8 (45)(30). 24–27. http://www.jstor.org/stable/44704341.

Gao, M. 2002. *Tones in whispered Chinese: articulatory features and perceptual cues*: Thesis.

Giet, F. 1955. Kann man in einer Tonsprache fluestern? *Lingua* 5. 372–381.

Jensen, M. K. 1958. Recognition of word tones in whispered speech. *Word* 14(2–3). 187–196.

Jiao, L & Y Xu. 2019. Whispered Mandarin has no production-enhanced cues for tone and intonation. *Lingua* 218. 24–37.

Kong, Y.-Y & F.-G Zeng. 2006. Temporal and spectral cues in Mandarin tone recognition. *The Journal of the Acoustical Society of America* 120(5). 2830–2840.

Liu, F. 1926. An experimental study of tones [四声实验录].

McLoughlin, I. V, Y Xu & Y Song. 2014. Tone confusion in spoken and whispered Mandarin Chinese. In *The 9th International Symposium on Chinese Spoken Language Processing*, 313–316. IEEE.

Meyer-Eppler, W. 1957. Realization of prosodic features in whispered speech. *The Journal of the Acoustical Society of America* 29(1). 104–106.

Miller, C. B. 1934. *An experimental-phonetic investigation of whispered conversation: considered from the linguistic point of view.* H. Pöppinghaus.

Miller, J. D. 1961. Word tone recognition in Vietnamese whispered speech. *Word* 17(1). 11–15.

Peng, S.-H. 2000. Lexical versus 'phonological' representations of Mandarin sandhi tones. *Papers in Laboratory Phonology V: Acquisition and the lexicon* 5. 152.

Shi, F. 1991. Tonal patterns in Beijing dialect [北京话的声调格局]. *Studies in Language and Linguistics [语言研究]* 17–24.

Wang, W. S & K.-p Li. 1967. Tone 3 in Pekinese. *Journal of Speech and Hearing Research* 10(3). 629–636.

Wise, C & L. P.-H Chong. 1957. Intelligibility of whispering in a tone language. *Journal of Speech and Hearing Disorders* 22(3). 335–338.

Zee, E. 1980. A spectrographic investigation of Mandarin tone sandhi. *UCLA Working Papers in Phonetics* 49(9). 98–116.

Appendix

Tone		Vowel				
		a	ɤ	i	u	y
Tone 1 level	Character	啊		衣		迂
	Pinyin	ā		yī		yū
	Glossary	Oh		Clothes		Winding
Tone 2 rising	Character		鹅		屋	
	Pinyin		é		wú	
	Glossary		Goose		Nothing	
Tone 3 dipping	Character	啊		椅		雨
	Pinyin	ǎ		yǐ		yǔ
	Glossary	What		Chair		Rain
Tone 4 falling	Character		饿		物	
	Pinyin		è		wù	
	Glossary		Hungry		Thing	

Vowel list for group 1

First syllable		Second Syllable			
		Tone 1 level	Tone 2 rising	Tone 3 dipping	Tone 4 falling
Tone 1 level	Character	鸡胸		伊始	
	Pinyin	jī xiōng		yī shǐ	
	Glossary	chicken's breast		begin	
Tone 2 rising	Character		及时		薄被
	Pinyin		jí shí		báo bèi
	Glossary		in time		thin quilt
Tone 3 dipping	Character	挤胸		以使	
	Pinyin	jǐ xiōng		yǐ shǐ	
	Glossary	to squeeze breast		in order to	
Tone 4 falling	Character		纪实		报备
	Pinyin		jì shí		bào bèi
	Glossary		Documentary		report

Word list for group 1

Tone		Vowel				
		a	ɤ	i	u	y
Tone 1 level	Character		婀		屋	
	Pinyin		ē		wū	
	Glossary		Graceful		House	
Tone 2 rising	Character	啊		姨		鱼
	Pinyin	á		yí		yú
	Glossary	Eh		Aunt		Fish
Tone 3 dipping	Character		恶		五	
	Pinyin		ě		wǔ	
	Glossary		Nausea		Five	
Tone 4 falling	Character	啊		意		玉
	Pinyin	à		yì		yù
	Glossary	Ah		Meaning		Jade

Vowel list for group 2

First syllable		Second Syllable			
		Tone 1 level	Tone 2 rising	Tone 3 dipping	Tone 4 falling
Tone 1 level	Character		基石		包被
	Pinyin		jī shí		bāo bèi
	Glossary		base		cover
Tone 2 rising	Character	吉凶		遗史	
	Pinyin	jí xiōng		yí shǐ	
	Glossary	good and bad luck		leftover issue	
Tone 3 dipping	Character		几时		宝贝
	Pinyin		jǐ shí		bǎo bèi
	Glossary		when		baby
Tone 4 falling	Character	继兄		亦使	
	Pinyin	jì xiōng		yì shǐ	
	Glossary	stepbrother		also	

Word list for group 2

Yueke Zeng
The University of Cambridge
yz872@cam.ac.uk