How amplitude influences Mandarin Chinese tone recognition in whisper

Patricia Peilin Yang
National Taiwan University
ckjpyang040@gmail.com

This study investigates whether intensity affects listeners' identification of whispered Mandarin tones. Two alternative forced choice identification experiments of whispered Mandarin tones were designed. In Experiment 1, twenty triplets of commonly used Mandarin Chinese bisyllabic words served as stimuli: two otherwise identical-sounding words of different tones and a word with completely different tones as a filler. The intensity level of the stimulus was set to 50, 60, and 70 dB, which were amplitudes determined according to a pilot study. Response accuracies across different intensity levels were analyzed. The results showed a tendency: the higher the amplitude, the better the listeners' recognition of whispered Mandarin tone. However, the accuracies across different intensity levels are insignificantly different, and Experiment 2 was then carried out. Four quartets of bisyllabic non-words were designed to match one of the four tones. The same intensity levels from Experiment 1 were employed. Each syllable was normalized to 480 ms. The results show that higher amplitudes only yield better tone recognition of tone 4. However, the absence of main effect for Intensity for either experiment suggests that amplitude may not be the only factor in determining the identification of whispered tone in Mandarin Chinese.
Keywords: whisper; amplitude; tone recognition; Mandarin Chinese

1 Introduction

As the vocal folds do not vibrate in whisper, there is no F0, the crucial acoustic correlate of tone. Accordingly, tone is difficult for people to recognize in whisper. Both Jensen (1958) and Gao (2002) argued that semantic context plays an important role in tone perception in whisper. Tones of a pair of otherwise identical-sounding words in whispered Mandarin could not be easily distinguished from each other unless there is a context as a cue. In other words, if words are whispered in isolation, it could be a hard task to identify their tones.

Some researchers have focused on intensity and speech identification in either phonated or whispered speech. Zhang et al (2007) found that the sound intensity levels of five speech modes, including whispered, soft, neutral, loud and shouted, were greatly varied, and the mean amplitude could represent the speech mode. Whisper, which possessed the lowest intensity, increased the difficulty of the speaker identification test and led to the worst performance of subjects.

As for perceptual cues in whisper, while Liu et al. (2004) proposed that duration helps recognition of whispered tones, Chang et al. (2007) argued that since speakers do not exaggerate intensity differences or durations among tones in whispered speech, there is no information available about duration or intensity for listeners to identify tones in whisper as is the case in normal speech. In Tartter (1989), it was found that the intensity of burst was one of the differences between voiced and voiceless stops in whisper for the latter have double-spiked bursts. This could thus serve as a perceptual cue in distinguishing voicing in whispered speech. Even though intensity can be a cue, when amplitude and duration are both available, Mandarin listeners do not show a preference between using amplitude and duration to recognize tone in whisper (Li et al., 2015). Then if one of two hints is controlled, will listeners be forced to use the other as a basis for judgement? As such, the role of amplitude in whispered tone recognition remains unknown.

While the above-mentioned studies have contributed to the understanding of tone recognition in whisper, none of them focused on its relationship with intensity. This paper attempts to probe into the role amplitude plays in distinguishing tones in whispered Chinese words by addressing one research question: Does amplitude affect listeners' tone recognition of isolated Mandarin Chinese words in whisper? The hypothesis is that the higher the amplitude, the better the tone recognition. Two perception experiments were designed to test people's performance on recognizing tones of Chinese words at different amplitudes in whisper.

2 Experiment 1

2.1 Methods

2.1.1 Participants

Ten (7 male; 3 female) native speakers aged 21 to 25 of Taiwan Mandarin participated in the study. All speakers were naïve to the purposes of the study.

2.1.2 Materials

The materials were twenty triplets of commonly used Mandarin Chinese bisyllabic words whose amplitudes were normalized (Table 1). Each triplet was composed of three bisyllabic words: two were identical in segments, but not tone, and the third was a filler word. The fillers were words that never cause any lexical confusion in terms of tone. The materials were recorded by a male native speaker of Mandarin Chinese in the phonetics laboratory at National Taiwan University and then normalized to 50, 60, and 70 dB, which were amplitudes decided according to a pilot, in which three female native speakers of Mandarin Chinese joined and which was conducted in the same way as the experiment.

Table 1. Sixty bisyllabic words used as materials for Experiment 1.

20 Triplets of Commonly Used Mandarin Chinese Bisyllabic Words		
Otherwise identical-sounding words of different tones		Filler
睡覺 *Shuìjiào* "sleep"	水餃 *Shuǐjiǎo* "dumpling"	吃飯 *Chīfàn* "to eat"
研究 *Yánjiù* "research"	菸酒 *Yānjiǔ* "tobacco and liquor"	哲學 *Zhéxué* "philosophy"
失眠 *Shīmián* "insomnia"	市面 *Shìmiàn* "market"	網路 *Wǎnglù* "network"
符號 *Fúhào* "sign"	富豪 *Fùháo* "mogul"	手錶 *Shǒubiǎo* "watch"
話筒 *Huàtǒng* "microphone"	花童 *Huātóng* "page boy"	水壺 *Shuǐhú* "kettle"
香蕉 *Xiāngjiāo* "banana"	橡膠 *Xiàngjiāo* "rubber"	瓢蟲 *Piáochóng* "ladybug"
看書 *Kànshū* "to read"	砍樹 *Kǎnshù* "to cut trees"	道路 *Dàolù* "way"
亞洲 *Yàzhōu* "Asia"	壓軸 *Yāzhóu* "finale"	書本 *Shūběn* "book"
天空 *Tiānkōng* "sky"	填空 *Tiánkòng* "to fill in the blanks"	朋友 *Péngyǒu* "friend"
發展 *Fāzhǎn* "development"	罰站 *Fázhàn* "to stand as punishment"	單車 *Dānchē* "bicycle"
上海 *Shànghǎi* "Shanghai"	傷害 *Shānghài* "to hurt"	天氣 *Tiānqì* "weather"
國民 *Guómín* "people"	過敏 *Guòmǐn* "allergy"	畫面 *Huàmiàn* "picture"
鉛筆 *Qiānbǐ* "pencil"	錢幣 *Qiánbì* "coin"	遵守 *Zūnshǒu* "to obey"

港幣 *Gǎngbì* “Hong Kong dollar”	鋼筆 *Gāngbǐ* “pen”	冷氣 *Lěngqì* “air conditioner”
鳳梨 *Fènglí* “pineapple”	鋒利 *Fēnglì* “sharp”	鋼琴 *Gāngqín* “piano”
約分 *Yuēfēn* “reduction of a fraction”	月份 *Yuèfèn* “month”	總統 *Zǒngtǒng* “president”
威脅 *Wēixié* “threat”	猥褻 Wěixiè “obscene”	牛肉 *Niúròu* “beef”
緣故 *Yuángù* “reason”	遠古 *Yuǎngǔ* “ancient”	忙碌 *Mánglù* “busy”
國中 *Guózhōng* “junior high school”	過重 *Guòzhòng* “overweight”	標題 *Biāotí* “title”
貿易 *Màoyì* “trade”	毛衣 *Máoyī* “sweater”	青菜 *Qīngcài* “vegetable”

2.1.3 Procedure

A two-alternative forced-choice task was designed. The stimuli were randomly delivered through PsychoPy and the participants were instructed to identify the whispered tone combination. Two tone combinations were shown in Zhuyin tone markers and visually displayed on the left and right sides of the monitor in front of them. The participants were asked to respond by pressing “p” or “q” buttons as quickly and accurately as possible within a time window of 4 seconds. For example, if they heard “*Guózhōng*,” “ ˊ - ” and “ ˋ ˋ ” would be shown on the monitor, and participant should choose “ ˊ - .” The whole process would repeat once again after a short break as block 1 and block 2. Thus, a total of 360 trials = 3 stimuli * 20 triplets * 3 levels of amplitude * 2 blocks was included.

2.2 Results

Mean accuracy and standard deviation combining blocks 1 and 2 are presented in Table 2, with the highest accuracy of each subject marked in bold print. The mean accuracies at 50, 60, and 70 dB are 55% (SD = 5.6%), 57% (SD = 6.8%), and 58.8% (SD = 7.4%), respectively. Across the ten participants, six reached their personal highest accuracy score in response to 70 dB stimuli, four to 60 dB and two to 50 dB, with two people having the same score for both 60 dB and 70 dB.

Table 2. Individual subject's accuracies of tone recognition at different amplitudes in Experiment 1.

Subject	**Mean accuracy**	**50 dB**	**60 dB**	**70 dB**
1	48%	**50%**	46%	48.7%
2	52.5%	52.5%	51%	**53.7%**
3	55%	**56%**	55%	55%
4	59%	56%	**61%**	60%
5	65%	57.5%	67.5%	**71%**
6	55.8%	52.5%	**57.5%**	**57.5%**
7	52.9%	48.7%	**55%**	**55%**
8	51.6%	50%	51%	**53.7%**
9	66%	66%	62.5%	**71%**
10	63%	62.5%	**65%**	62.5%
Mean	57%	55%	57%	58.8%
SD	6.1%	5.6%	6.8%	7.4%

2.3 Discussion

The results of Experiment 1 show a tendency: the higher the amplitude, the better participants perform in tone recognition. In terms of individuals, there is a trend for the participants to reach their personal best accuracy scores at the higher amplitudes; in terms of the whole, the mean accuracy gets higher as the intensity level increases. Thus, it is shown that the subjects performed better when the amplitude was higher, suggesting an influential role of intensity level in whispered tone recognition. While the results reveal the tendency between the high stimulus intensity and the high identification accuracy, the differences between the accuracy across the three amplitudes were not significant. Between 50 and 60 dB, $p = 0.49$; between 60 and 70 dB, $p = 0.61$; between 60 and 70 dB, $p = 0.23$. Between the two blocks, $p = 0.32$. For all tests, $df = 9$, $a = 0.05$.

The current design only controlled two variables: the response time allowed (4 seconds for each stimulus) and the intensity levels (50, 60, and 70 dB). The insignificant differences may have resulted from several uncontrolled variables, including vowel quality, combination of tones, order of stimuli, and speech rate. The vowel qualities differ from stimulus to stimulus; possible tone combinations were not exploited in the current design; the speech rates of each stimulus sound were similar but not identical. All of the uncontrolled variables may also affect the results of the experiment. As such, it is unlikely that amplitude is the one and only factor that influences people's tone identification in whisper.

3 Experiment 2

To further explore which factors may influence tone identification in whisper, Experiment 2 was designed and conducted to examine the identification of tones at different intensity levels again. Different from Experiment 1, Experiment 2 focused on people's tone perception of the second syllable in whispered bisyllabic words, with amplitudes, durations, and tone combinations controlled.

3.1 Methods

3.1.1 Participants

Another 15 native speakers (5 male; 10 female) of Taiwan Mandarin were recruited (aged 18 to 24). All speakers were naïve to the purposes of the study.

3.1.2 Materials

The materials contain four quartets of Mandarin Chinese bisyllabic non-words to avoid lexical processing. The sound combinations include /tipa/, /tapi/, /p^hitu/, and /p^huti/. Each quartet has one sound combination. For each stimulus of each quartet, the first syllable was always matched with Tone 1 whereas the second syllable was one of the four lexical tones (Tones 1 - 4). The first two quartets were designed to investigate the influence of vowel height in whispered tone recognition (with the contrast between [i] and [a] in the syllables), and the latter two, vowel frontness with the contrast between [i] and [u]. The materials were recorded by a male native speaker of Mandarin Chinese in the phonetics laboratory at National Taiwan University, and then normalized to 50, 60, and 70 dB, and 480 ms for each syllable.

3.1.3 Procedure

All setups were identical to those in Experiment 1 except that four possible tone combinations were provided this time. Participants were asked to identify the correct tone combination from four options by pressing one of the 1, 3, 7, 9 keys on the number pad. For example, if they heard "/tipa/," " - - ," " - ´ ," " - ˇ ," and " - ` " would be shown on the monitor, and they should choose " - ´ ." The whole process would repeat once again after a short break as block 1 and block 2. Therefore, there is a total of 192 trials = 4 sound combinations * 4 tone combinations * 3 levels of amplitude * 2 repetitions in one block * 2 blocks.

3.2 Results

Combining blocks 1 and 2, mean accuracy and standard deviation are presented in Table 3, with the highest accuracy of each subject marked in bold print. The mean accuracies at 50, 60, and 70 dB are 31% (SD = 11.6%), 32% (SD = 11.3%), and 34% (SD = 10.6%), respectively. Across the fifteen participants, six reach their personal highest accuracy score in response to 70 dB stimuli, five to 60 dB and five to 50 dB, with one person having the same score for both 60 dB and 70 dB.

Table 3. Individual subject's accuracies of tone recognition at different amplitudes in Experiment 2.

Subject	**Mean accuracy**	**50 dB**	**60 dB**	**70 dB**
1	45.3%	**51.6%**	39.1%	45.3%
2	29.7%	**32.8%**	29.7%	26.6%
3	30.7%	28.1%	26.6%	**37.5%**
4	32.8%	29.7%	32.8%	**35.9%**
5	44.8%	40.6%	**48.4%**	45.3%
6	12.5%	7.8%	12.5%	**17.2%**
7	26.0%	26.6%	**28.1%**	23.4%
8	44.3%	**45.3%**	43.8%	43.8%
9	52.1%	46.9%	**56.3%**	53.1%
10	33.9%	26.6%	32.8%	**42.2%**
11	28.7%	**31.3%**	26.6%	28.1%
12	24.5%	17.2%	26.6%	**29.7%**
13	21.9%	**26.6%**	18.8%	20.3%
14	31.8%	26.6%	**39.1%**	29.7%
15	25.0%	21.9%	**26.6%**	**26.6%**
Mean	32%	31%	32%	34%
SD	10.5%	11.6%	11.3%	10.6%

An ANOVA with sphericity corrections was conducted to examine the effect of tone and intensity. No effects for Tone and Intensity were found (both $p > .05$), whereas the interaction between Tone and Intensity was significant ($p < .01$). The post-hoc analyses (Tukey HSD) reported that the Tone*Intensity interaction was driven by the significant rise for Tone 4 from 50 to 70 dB ($p < .001$).

The accuracies of each tone at each amplitude are shown in Figure 1. Among the four tones, only Tone 4 shows the expected pattern: the higher the amplitude, the higher the accuracy.

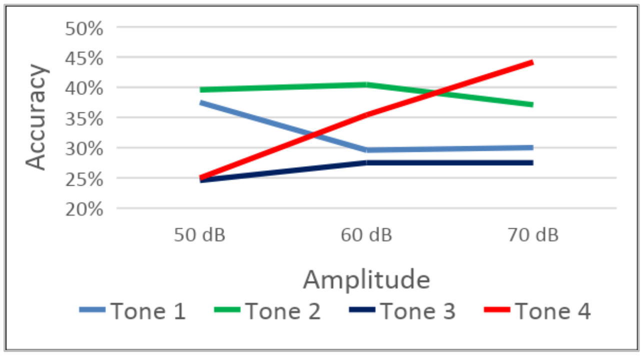

Figure 1. Accuracies of each tone at each amplitude in Experiment 2.

Furthermore, the proportions of each tone in response to each target tone at different intensity levels are also shown in Figures 2 to 4. The highest proportion was marked in bold print. At 50 dB, when the target tone was Tone 1 and 2, the responses that matched the correct tones have the largest response proportion; when Tone 3 and 4 were presented, the largest response proportion for Tone 1 was reported. At 60 dB, when the target tone was Tone 2 and 4, the tone-matched responses have the largest response proportion; when Tone 1 and 3 were presented, the largest response proportion for Tone 2 and 1 were reported. At 70 dB, the highest proportion was found for each tone-matched response.

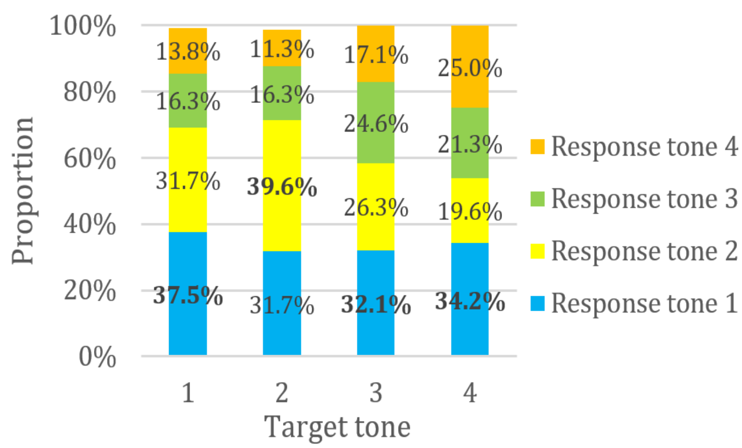

Figure 2. Proportion of each response tone in each target tone at 50 dB.

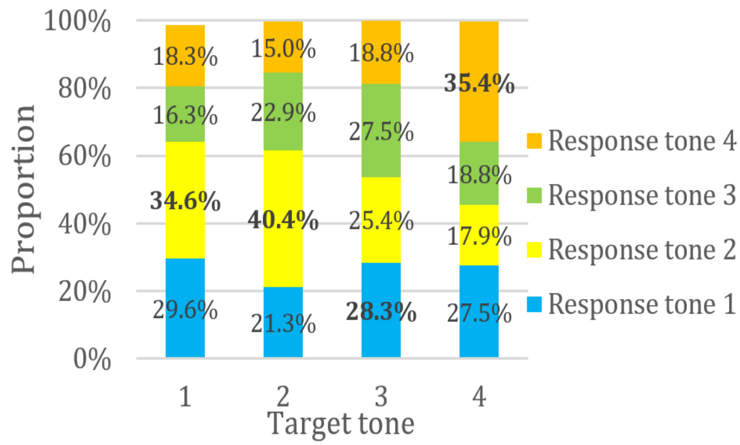

Figure 3. Proportion of each response tone in each target tone at 60 dB.

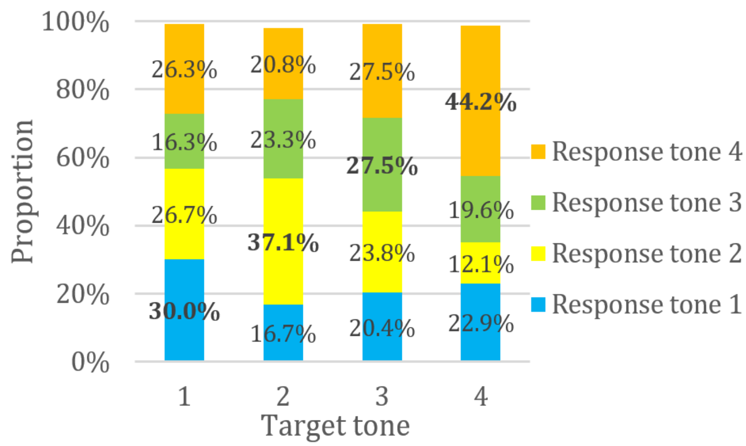

Figure 4. Proportion of each response tone in each target tone at 70 dB.

3.3 Discussion

The results of Experiment 2 show the tendency as Experiment 1: the higher the amplitude, the better participants perform in tone recognition. Similar to Experiment 1, there is a trend for the participants to reach their personal best accuracy scores at the higher amplitudes, and the overall mean accuracy gets higher as the intensity level increases. Figure 2 to 4 also show that the response proportion is positively matched with the target tone when the intensity of the stimulus increases. Collectively, the results suggest that the amplitude appears to be influential in whispered tone identification.

As revealed in the results, no main effect for Intensity was found. Possible factors include the number of response choices, normalized duration, and the use of non-word stimuli. Four options were provided in this experiment, which were twice more than those in Experiment1. This could potentially cause identification difficulty. Besides, as duration serves as a hint in tone recognition (Liu et al., 2004), the lack of duration difference among syllables could also make the task more challenging. In addition, in Experiment 1, the identification might be associated with lexical decision while Experiment 2 only provided non-words, which involved only low-level tone identification. These changes could potentially be responsible for the lower accuracy reported and the absence of main effect for Intensity.

The interaction between Tone and Intensity indicates that the performance of one tone or one intensity is different from that of others. As revealed in Figure 1, the different performance was associated with Tone 4, confirmed with the post-hoc analyses (Tukey HSD). Only Tone 4 shows the obvious tendency: the higher the amplitude, the higher the accuracy of tone recognition.

4 General Discussion and Conclusion

This research probes into how amplitude influences tone recognition in whispered Mandarin Chinese. As there is no vibration of the vocal folds in whisper, tone is hard to recognize. Jensen (1958) also argued that if there is no context as a cue, tones of a pair of otherwise identical-sounding words in whispered Mandarin could not be easily distinguished from each other. Besides, duration helps recognition of whispered tones (Liu et al., 2004), while the role of amplitude in whispered tone recognition remains unknown.

This present study proposed one research question: Does amplitude affect listeners in tone recognition of whispered Mandarin Chinese? Two perception experiments were conducted. In Experiment 1, twenty triplets of commonly used Mandarin Chinese bisyllabic words served as stimuli: two otherwise identical-sounding words of different tones and a word with completely different tones as a filler. The intensity level of the stimulus was set to 50, 60, and 70 dB, which were amplitudes determined according to a pilot study. Response accuracies across different intensity levels were analyzed. The results showed a tendency: the higher the amplitude, the better the listeners in whispered Mandarin tone recognition. However, the accuracies across different intensity levels are insignificantly different. Experiment 2 was designed to examine whether the identification of tones has different sensitivity to the intensity level in whisper. Four quartets of non-words were designed to match one of the four tones. The same intensity levels from Experiment 1 were employed and duration of each syllable was set to 480 ms. The results show that higher amplitudes only yield better tone recognition of tone 4, which is a surprising fact and can bring further investigation. However, the absence of main effect for Intensity for either experiment suggests that amplitude may not be the only factor in determining the identification of whispered tone in Mandarin Chinese.

Acknowledgements

I would like to express my sincere gratitude to Dr. Chenhao Chiu for his invaluable guidance and constructive comments throughout this research. I am also grateful to my family and roommates for their strong support. Thank you also to my fellows at the Graduate Institute of Linguistics at National Taiwan University for their practical suggestions and help. My appreciation also goes to the participants at Northwest Linguistics Conference 2019 for their insightful discussions and the WPLC editors for helpful comments. Any weaknesses are entirely mine.

References

Chang, C. B., & Yao, Y. (2007). Tone production in whispered Mandarin. *UC Berkeley PhonLab Annual Report*, *3*(3), 326-329.

Gao, M. (2002). *Tones in whispered Chinese: Articulatory features and perceptual cues* (Master's thesis). University of Victoria, BC, Canada.

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

Li, J., Ma, Q., Wang, T., & Xu, Y. (2015). Perceptual cues of whispered tones: are they really special? In *Sixteenth Annual Conference of the International Speech Communication Association*.

Liu, S., & Samuel, A. G. (2004). Perception of Mandarin lexical tones when F0 information is neutralized. *Language and Speech*, *47*(2), 109-138.

Tartter, V. C. (1989). What's in a whisper? *The Journal of the Acoustical Society of America*, *86*(5), 1678-1683.

Zhang, C., & Hansen, J. H. (2007). Analysis and classification of speech mode: whispered through shouted. In *Eighth Annual Conference of the International Speech Communication Association*.