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

Context Effects on Tone and Intonation Processing in Mandarin

Min Liu [1, 2], *Yiya Chen* [1, 2], *Niels O. Schiller* [1, 2]

[1] Leiden University Center for Linguistics, Leiden University, the Netherlands
[2] Leiden Institute for Brain and Cognition, Leiden University, the Netherlands
{m.liu, yiya.chen, n.o.schiller}@hum.leidenuniv.nl

Abstract

This study investigated how Mandarin listeners process tone and intonation when the F_0 encodings of the lexical tone and intonation are in conflict or in congruency and the role context plays during these processes. Tone and intonation identification experiments were conducted within neutral vs. constraining semantic contexts. Tone identification was much easier than intonation identification irrespective of contexts. Participants could perceive tones accurately and quickly in both question and statement intonation. However, intonation identification was greatly deteriorated within the neutral semantic context. Questions ending with a rising tone and a falling tone were equally difficult to identify. In a constraining semantic context, questions ending with a falling tone were much better identified. Thus, top-down information provided by the constraining semantic context does play an important role in disentangling intonation information from tone information.

Index Terms: tone, intonation, Mandarin, context

1. Introduction

Mandarin is a tonal language. At the lexical level, F_0 is employed to differentiate the four lexical tones. At the sentential level, F_0 is also used to convey post-lexical information, like intonation types. Mandarin has a rising tone (T2) and a falling tone (T4). Furthermore, question intonation in Mandarin is realized as an upward trend of the F_0 contour while statement intonation is realized as a downward trend [1-3]. This brings up the question of how tone and intonation are processed when their F_0 encodings are in conflict or in congruency.

Evidence from previous production studies showed that intonation-induced F_0 largely affects the F_0 height rather than the F_0 contour of lexical tones [4-6]. However, few studies have tested the effect of intonation on tone identification from a perceptual point of view. Would tone identification interfere with intonation processing?

As for the effect of tone on intonation identification, Yuan [7] discovered that Mandarin listeners identified questions ending with T4 better than questions ending with T2. Xu and Mok [8] replicated the asymmetrical results in Mandarin. However, in a follow-up study using low-pass filtered speech [9], the results were reversed, where it was found that Mandarin listeners had better identification of questions ending with T2 than questions ending with T4. Note that in low-pass filtered speech, not only semantic contexts, but also lexical information was removed. It therefore remains open whether the reversed identification pattern was exclusively due to semantic contexts.

To address this issue, this study was designed to investigate whether intonation identification differs between final T2 and T4 sentences as a function of semantic contexts. Two experiments were conducted. Experiment 1 aimed to examine the processing of tone and intonation in neutral semantic context, and Experiment 2 in constraining semantic context.

2. Experiment 1

2.1. Method

2.1.1. Materials

Forty monosyllabic word pairs with minimal tonal contrast (T2 vs. T4) were selected. Each minimal T2_T4 word pair contains words of comparable word frequency, phonological neighborhood density, and syntactic word category. To avoid any word frequency effect, only frequent words with more than 4,500 occurrences in a corpus of 193 million words were used [10]. All the critical words occurred in the final position of a five-syllable carrier sentence, i.e., *ta1 gang1gang1 shuo1 X* ('She just said X'), produced with either a statement (S) or a question intonation (Q). The carrier sentence is semantically meaningful but offered neutral semantic information to the target stimuli and will thus be referred to as the neutral semantic context hereafter. In total, 160 target sentences (40 Syllables × 2 Tones × 2 Intonations) were designed. Together with 240 filler sentences, which possess the same carrier but different critical words as to segmental or tonal elements (e.g. T1/ T3), a 400-sentence corpus was constructed.

2.1.2. Recording and Stimuli Preparation

Four native Mandarin speakers (2 females, 2 males), born and raised in Beijing, were recruited to record the sentences. The recordings took place in a soundproof recording booth at the Phonetics Lab of Leiden University. Sentences were randomly presented to the speakers and recorded at 16-bit resolution and a sampling rate of 44.1 kHz. To eliminate paralinguistic information, speakers were instructed to avoid any exaggerated emotional prosody during the recording.

One female speaker's recordings were chosen for the perception experiment based on the acoustic results, which showed comparable F_0 realization of tone and intonation to a prior study [11] and were therefore taken as the prototypical patterns for the perception study. The amplitude of all the sentences was normalized in PRAAT.

2.1.3. Participants

Eighteen native speakers of Mandarin (10 females, 8 males) from Northern China were paid to participate in the experiment. They were undergraduate or graduate students at Beijing Language and Culture University, between 19 and 27 years old ($M \pm SD$: 23.6 ± 2.3). None of them had received any formal musical training or had reported any speech or hearing disorders.

 doi: 10.21437/SpeechProsody.2016-217

2.1.4. Procedure

Participants were tested in a sound-attenuated room. Four hundred sentences (including 160 targets and 240 fillers) were randomly presented using E-Prime 2.0. Half of the target sentences were used for the tone identification task; the other half was used for the intonation identification task. The tasks were randomly allocated from trial to trial.

The experiment consisted of a practice session and four experiment sessions. The practice session contained 12 trials. Each experiment session contained 100 trials. Between two sessions there was a 3-minute break. An experimental trial started with a 100 ms warning beep, followed by a 300 ms pause. After that an auditory sentence was presented while a visual task interface appeared on the screen. Participants had to carry out either the tone identification task (whether the final tone is T2 or T4) or the intonation identification task (whether the sentence is a question or statement) from the onset of an auditory sentence until 2 seconds after the offset of the sentence as quickly and as accurately as possible. The interstimulus interval was 500 ms. Instructions were given visually on screen and orally by the experimenter beforehand.

2.1.5. Data Analysis

Previous studies on intonation perception typically report Identification Rate (IR) only [7-9]. In this study, in addition to IR, Reaction Time (RT) was included as a dependent variable, as RT serves as a good indicator of the degree of easiness of a perceptual decision: the easier a perceptual decision is, the shorter the RT is [12]. In our study, IR was defined as the percentage of correct identification of tone in the tone identification task, and as the percentage of correct identification of intonation in the intonation identification task. RT was defined as the response time relative to the onset of the last syllable for correct responses. To normalize the distribution, raw RTs were transformed using the natural logarithm.

Statistical analyses were carried out with the package *lme4* [13] in R [14]. Analysis of Response (Correct or Incorrect) was performed using binomial logistic regression models and analysis of RT was performed using linear mixed-effects regression models. The models included Task, Tone, Intonation, and their interactions as fixed factors, and Subjects and Items as random factors. The fixed factors were added in a stepwise fashion and their effects on model fits were evaluated via model comparisons based on log-likelihood ratios. Note that although trial-by-trial dependency was considered, Trial did not significantly improve the model fit, and was therefore excluded in the final model.

2.2. Results

2.2.1. Response

Figure 1 presents the identification rates of the four experimental conditions in the tone identification task (indicated by Tone) and the intonation identification task (indicated by Intonation). Each experimental condition is a combination of the levels of the factors Tone (T2, T4) and Intonation (Q, S); for example, QT2 refers to the condition of questions ending with T2.

To test whether tone and intonation are processed differently, we examined the effect of Task first. Results showed a significant main effect of Task ($\chi^2(1) = 91.42$, $p < 0.05$) and a Task × Intonation interaction ($\chi^2(1) = 19.28$, $p < 0.05$) on the odds of correct responses over incorrect responses.

Separate models for subset data of different intonation types revealed an interesting asymmetry between question and statement intonation. Specifically, in question sentences, the identification rate of the tone identification task was much higher than that of the intonation identification task ($\beta = 3.89$, $z = 10.08$, $p < 0.05$), but not in statement sentences, where near-ceiling level of identification was observed in both tasks.

Separate models were also constructed for subset data of different tasks. For the tone identification task, results showed no effect of Tone, Intonation or their interaction (all *p*s > .05). Thus the identification rate for each condition did not differ from each other. Regardless of intonation types, T2 and T4 were mostly correctly identified (with well above 90% identification rates). This suggests that the identity of lexical tone was not hindered by the intonation information. With respect to the intonation identification task, a significant main effect of Intonation was found ($\chi^2(1) = 83.59$, $p < 0.05$). Question intonation tended to be much more difficult to identify than statement intonation across final tone identities.

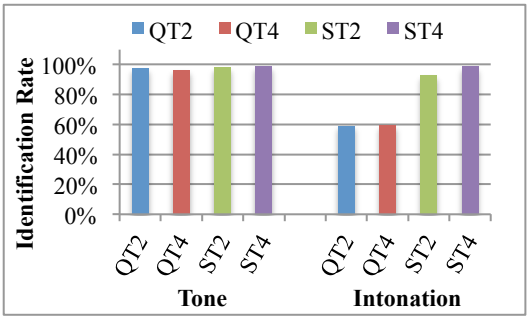

Figure 1: *Neutral semantic context: IRs for different tasks.*

2.2.2. Reaction Time

Figure 2 presents the average RTs for each experimental condition under different tasks. The error bars represent the 95% confidence interval of the means. The overall results revealed a significant main effect of Task ($\chi^2(1) = 110.52$, $p < 0.05$), indicating that the final tone was identified faster than the intonation. Other factors such as Tone, Intonation and the interaction of Tone × Intonation also reached significance (all *p*s < 0.05).

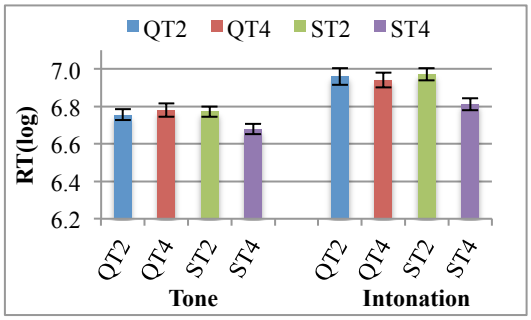

Figure 2: *Neutral semantic context: average RTs with 95% CI for different tasks.*

Separate models for subset data of different intonations showed no RT differences between the T2 and T4 conditions in question intonation across task types ($\beta = 0.01$, $t = 0.49$, $p > 0.05$), whereas a much shorter RT was observed for the T4

than the T2 condition in statement intonation irrespective of task types ($\beta = -0.13$, $t = -7.58$, $p < 0.05$).

Separate models were also constructed for subset data of different tones, confirming that there was no Intonation effect for the T2 conditions, but there was a significant effect of Intonation for the T4 conditions, with shorter RTs for statement sentences (ST4) than for question sentences (QT4) regardless of task types ($\beta = -0.13$, $t = -6.73$, $p < 0.05$).

Overall, tone identification almost reached a ceiling level across all experimental conditions. However, the identification of intonation displayed strong biases towards statement intonation. Moreover, reaction time for intonation identification was much longer than for tone identification. Taken together, it seems that in a neutral semantic context, participants had great difficulty perceiving question intonation. This is in line with previous studies [7-9]. Nevertheless, different from [7-8], no intonation perceptual difference was found for questions ending with T2 vs. T4.

3. Experiment 2

Results of Experiment 1 showed that in the neutral semantic context, question intonation processing is challenging, regardless of the final lexical tone identity. Since highly constraining semantic contexts have been shown to facilitate tone processing [15], the question arises as to whether a highly constraining semantic context contributes to intonation processing. Experiment 2 was designed to tap into this effect.

3.1. Method

3.1.1. Materials

To avoid learning effects from Experiment 1, an additional set of 40 syllables in combination with tone (T2 or T4) was selected for Experiment 2. The critical syllables are the second syllable of frequent disyllabic words, which were embedded in the final position of various ten-syllable natural sentences. This sentence context was verified to provide sufficient constraint to the final syllable in a pretest and will be referred to as the constraining semantic context hereafter. All the sentences were produced either with a statement or a question intonation, yielding another 160 target sentences (40 Syllables × 2 Tones × 2 Intonations). Like in Experiment 1, 240 sentences were included as fillers.

3.1.2. Recording and Stimuli Preparation, Participants, Procedure and Data Analysis

Recording and stimuli preparation, participants, procedure and data analysis were the same as in Experiment 1. The same speaker's recordings were selected. Experiment 2 was run after Experiment 1 over the same group of participants.

3.2. Results

3.2.1. Response

Figure 3 presents the IRs of all experimental conditions under different tasks in the constraining semantic context. As in Experiment 1, we found a significant main effect of Task ($\chi^2(1) = 32.59$, $p < 0.05$) and a two-way interaction of Task × Intonation ($\chi^2(1) = 28.63$, $p < 0.05$) in Experiment 2. The tone identification task showed a better performance (with well above 90% identification rates) than the intonation identification task in question sentences ($\beta = 3.59$, $z = 8.25$, $p < 0.05$). In statement sentences, however, the tone identification task and the intonation identification task were equally well performed ($\beta = -0.97$, $z = -1.14$, $p > 0.05$).

Separate models were constructed for subset data of different tasks. For the tone identification task, a significant effect of Intonation was found for T4 ($p < 0.05$), with better identification of T4 in statements than in questions. Results of the intonation identification task showed a significant main effect of Intonation ($\chi^2(1) = 89.91$, $p < 0.05$) and a significant main effect of Tone ($\chi^2(1) = 4.64$, $p < 0.05$). No significant interaction was found ($\chi^2(1) = 1.69$, $p > 0.05$). Specifically, intonation (both question and statement) identification was more difficult in sentences ending with T2 than in those ending with T4. Also, question intonation was more difficult to identify than statement intonation across final tone types.

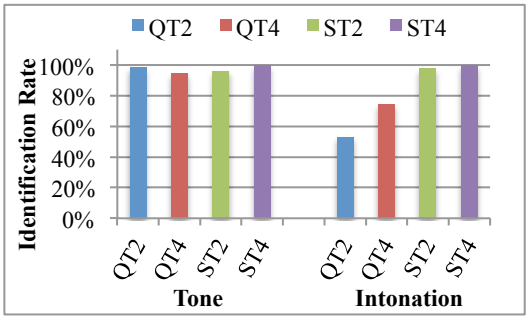

Figure 3: *Constraining semantic context: IRs for different tasks.*

3.2.2. Reaction Time

Figure 4 presents the average RTs for each experimental condition under different tasks in the constraining semantic context. The error bars represent the 95% confidence interval of the means. The overall analyses showed a significant interaction of Task × Intonation ($\chi^2(1) = 10.52$, $p < 0.05$). Participants were much faster in the tone identification task than in the intonation identification task under the question sentence conditions ($\beta = -0.16$, $t = -2.74$, $p < 0.05$), but not under the statement sentence conditions ($\beta = 0.04$, $t = 0.88$, $p > 0.05$). With a constraining semantic context, RTs in the intonation identification task for statements ending with T2 decreased to such a degree that it even became shorter than RTs for the same condition in the tone identification task.

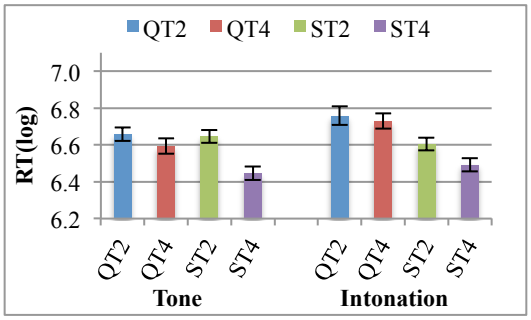

Figure 4: *Constraining semantic context: average RTs with 95% CI for different tasks.*

Separate analyses were performed for subset data of different tasks. For the tone identification task, there showed a significant interaction of Tone × Intonation ($\chi^2(1) = 6.73$, $p < 0.05$). Consistent with the neutral semantic context, in the

constraining semantic context, only in the statements, a shorter RT was observed for the identification of T4 than T2 ($\beta = -0.20$, $t = -4.53$, $p < 0.05$), and only in sentences ending with T4, a faster tone identification was found in questions than in statements ($\beta = -0.15$, $t = -3.7$, $p < 0.05$). Overall, ST4 had a significant advantage over the other conditions.

For the intonation identification task, there was a significant main effect of Intonation ($\chi^2(1) = 36.44$, $p < 0.05$), indicating a shorter RT for statement intonation identification than for question intonation identification under both T2 and T4 conditions. With a constraining semantic context, RTs for the statement intonation were greatly shortened. A significant interaction of Tone × Intonation was also found ($\chi^2(2) = 9.45$, $p < 0.05$). Only in statements, it took less time to identify the intonation in sentences with a final T4 than in those with a final T2 ($\beta = -0.12$, $t = -2.73$, $p < 0.05$).

To sum up, when given a constraining semantic context, participants still had difficulty perceiving question intonation, especially when the question intonation concurred with T2. However, question intonation identification did improve in questions ending with T4 if compared with the neutral semantic context (see below).

4. Experiment 1 vs. Experiment 2

To test whether tone and intonation perception by the same participants differed as a function of semantic contexts, we compared results from Experiments 1 and 2 in this section.

With respect to identification rate, in both neutral and constraining semantic contexts, tone identification yielded rather high IR across the four experimental conditions. About intonation identification, IR of QT2 dropped from 59.2% in the neutral semantic context to 53.1% in the constraining semantic context. In contrast, IR of the other three conditions increased in the constraining semantic context compared with their neutral semantic context counterparts (QT4: 74.7% vs. 59.7%; ST2: 97.8% vs. 92.8%; ST4: 100% vs. 98.6%).

As for RT, context showed a significant effect on the response time to identify tone and intonation ($\chi^2(1) = 146.29$, $p < 0.05$). The constraining semantic context played a significant role in speeding up the identification of both tone and intonation across the experimental conditions. It shortened RTs to a larger degree in the intonation identification task than in the tone identification task.

5. General Discussion

To address the question of how top-down information provided by semantic contexts affects tone and intonation processing in Mandarin when the F_0 encodings of the lexical tone and intonation are in conflict or in congruency, we examined the identification of tone and intonation in both neutral and constraining semantic contexts. Our results demonstrated that tone identification does not interfere with intonation processing irrespective of semantic contexts, whereas intonation identification, particularly question intonation, is susceptible to the final tone identity and is greatly deteriorated in the neutral semantic context.

In our study, the overall performance of the tone identification was better than that of the intonation identification regardless of semantic contexts. Evidence was found not only from the identification rate, but also from the reaction time, which was exclusively reported in this study. Intonation identification was shown to take more time than tone identification regardless of the final tone types, suggesting that when pitch movements are used to convey post-lexical contrast, its identification becomes a much more difficult decision-making process [16]. The advantage of tone over intonation is probably because that a phonetic dimension (i.e. F_0) exploited for one function of the grammar (e.g. lexical tone) limits its effectiveness to cue a different function (e.g. intonation) in the same linguistic system [17].

Previous studies found reversed patterns of question intonation identification in questions ending with T2 and T4 in normal context [7-8] and in low-pass filtered context [9]. However, it is unclear whether the reversed pattern is due to differences of the two test contexts in semantic or in lexical information. The present study teased apart the effect of semantic contexts from the other factors by introducing the neutral vs. constraining semantic contexts. We found that neutral semantic context did pose greater difficulty to question intonation identification, compared to the constraining semantic context. In the former, questions ending with T2 and T4 were equally badly identified. In the latter, questions with a final T4 were better identified than those with a final T2, in line with the results in [7-8]. Recall that in low-pass filtered speech, questions ending with T2 even had a higher identification rate than questions ending with T4 [9]. It seems that the stronger the linguistic context is (constraining semantic context > neutral semantic context > low-pass filtered context), the better the identification of questions ending with T4. We infer that with less semantic information, the frequency code [18], which holds that high or rising pitch marks questions, and low or falling pitch marks statements, is more likely to be applied to intonation identification, resulting in relatively better identification of questions ending with T2. However, under no circumstance could listeners disentangle question intonation from T2 easily (53.1% vs. 59.2%). When more semantic information is given, questions ending with T4 tend to get more cues of question intonation than questions ending with T2. The reasons for this warrant further investigation.

In addition to the identification rate, constraining semantic context also speeded up tone identification as well as intonation identification compared to the neutral semantic context. It shortened RTs for intonation identification to a larger extent. The most noticeable effect of the constraining semantic context was in the identification of statement intonation when the statement ends with T2. This indicates that overall, with a constraining semantic context, there are more cues for statement than for question intonation.

6. Conclusion

To conclude, results of the two experiments reported here show that tone at the lexical level and intonation at the sentential level in Mandarin interact with each other, causing asymmetrical difficulty of pitch processing at sentential level. To disentangle intonation information from tone information more efficiently, not only acoustic cues, but also semantic contexts need to be taken into consideration. A constraining semantic context greatly improves question intonation identification, but mainly so in sentences with the lexical falling tone in the final position.

7. Acknowledgements

We thank the support from the Chinese Scholarship Council to ML and the European Research Council (ERC Starting Grant-206198) to YC.

8. References

[1] A. T. Ho, "Intonation Variation in a Mandarin Sentence for Three Expressions: Interrogative, Exclamatory and Declarative," *Phonetica,* vol. 34, no. 6, pp. 446–457, 1977.

[2] E. Gårding, "Speech Act and Tonal Pattern in Standard Chinese: Constancy and Variation," *Phonetica*, vol. 44, no. 1, pp. 13–29, 1987.

[3] F. Liu and Y. Xu, "Parallel Encoding of Focus and Interrogative Meaning in Mandarin Intonation," *Phonetica*, vol. 62, no. 2–4, pp. 70–87, 2005.

[4] X. S. Shen, *The Prosody of Mandarin Chinese*. Berkeley: University of California Press, 1989.

[5] Z. Wu, "A New Method of Intonation Analysis for Standard Chinese: Frequency Transposition Processing of Phrasal contours in a sentence," *Analysis, Perception and Processing of Spoken language.* Elsevier Science and Technology books, 1996.

[6] J. Cao, "Intonation Structure of Spoken Chineses: University and Specificity," *Report of Phonetic Research*, pp. 31–38, 2004.

[7] J. Yuan, "Perception of intonation in Mandarin Chinese," *Journal of the Acoustical Society of America,* vol. 130, no. 6, pp. 4063–4069, 2011.

[8] B. R. Xu and P. Mok, "Cross-linguistic Perception of Intonation by Mandarin and Cantonese Listeners," in *Speech Prosody 2012, May 22–26, Shanghai, China, Proceedings,* 2012, pp. 99–102.

[9] B. R. Xu and P. Mok, "Intonation Perception of Low-Pass Filtered Speech in Mandarin and Cantonese," in *TAL 2012 – The Third International Symposium on Tonal Aspect of Languages, May 26–29, Nanjing, China*, 2012.

[10] J. Da, "A corpus-based study of character and bigram frequencies in Chinese e-texts and its implications for Chinese language instruction," *Proceedings of the 4th International Conference on New Technologies in Teaching and Learning Chinese,* pp. 501–511, 2004.

[11] J. Yuan, "Mechanisms of Question Intonation in Mandarin," in *ISCSLP 2006, December 13-16, Singapore, Proceedings,* 2006, pp. 19-30.

[12] K. Schneider, G. Dogil, and B. Möbius, "Reaction Time and Decision Difficulty in the Perception of Intonation," in *INTERSPEECH 2011 – 12th Annual Conference of the International Speech Communication Association, August 27–32, Florence, Italy, Proceedings*, 2011, pp. 2221–2224.

[13] D. Bates, M. Maechler, B. Bolker and S. Walker, "lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–9," https://CRAN.R-project.org/package=lme4, 2015.

[14] R Core Team, "R: A language and environment for statistical computing. R Foundation for Statistical Computing," http://www.R-project.org/, Vienna, Austria. 2014.

[15] Y. Ye, & C. M. Connine, "Processing Spoken Chinese: The Role of Tone Information," *Language and Cognitive Processes,* vol. 14, no. 5-6, pp. 609–630, 1999.

[16] B. Braun and E. K. Johnson, "Question or tone 2? How language experience and linguistic function guide pitch processing," *Journal of Phonetics*, vol. 39, no. 4, pp. 585–594, 2011.

[17] J. Liang and V. J. Heuven, "Chinese tone and intonation perceived by L1 and L2 listeners," *Tones and Tunes, Volume 2: Experimental studies in word and sentence prosody*, vol. 12, no. 2, pp. 27–61, 2007.

[18] J. J. Ohala, "Cross-Language Use of Pitch: An Ethological View," *Phonetica*, vol. 40, no. 1, pp. 1–18, 1983.