Context matters for tone and intonation processing in Mandarin

Liu, M.; Chen, Y.; Schiller, N.O.

Article

Language *and* Speech

Context Matters for Tone and Intonation Processing in Mandarin

Language *and* Speech
2022, Vol. 65(1) 52–72

DOI: 10.1177/0023830920986174
journals.sagepub.com/home/las
SAGE

Min Liu
College of Chinese Language and Culture & Institute of Applied Linguistics, Jinan University, China

Yiya Chen

Niels O. Schiller
Leiden University Centre for Linguistics & Leiden Institute for Brain and Cognition, Leiden University, the Netherlands

Abstract

In tonal languages such as Mandarin, both lexical tone and sentence intonation are primarily signaled by F0. Their F0 encodings are sometimes in conflict and sometimes in congruency. The present study investigated how tone and intonation, with F0 encodings in conflict or in congruency, are processed and how semantic context may affect their processing. To this end, tone and intonation identification experiments were conducted in both semantically neutral and constraining contexts. Results showed that the overall performance of tone identification was better than that of intonation. Specifically, tone identification was seldom affected by intonation information irrespective of semantic contexts. However, intonation identification, particularly question intonation, was susceptible to the final lexical tone identity and affected by the semantic context. In the semantically neutral context, questions ending with a rising tone and a falling tone were equally difficult to identify. In the semantically constraining context, questions ending with a falling tone were much better identified than those ending with a rising tone. This perceptual asymmetry suggests that top-down information provided by the semantically constraining context can play a facilitating role for listeners to disentangle intonational information from tonal information, but mainly in sentences with the lexical falling tone in the final position.

Keywords

Tone, intonation, Mandarin, neutral context, constraining context

Introduction[1]

Different languages may have different ways of marking questions. One common way of marking questions in various languages is with the use of syntactic means, including changing word order

Corresponding author:
Min Liu, College of Chinese Language and Culture & Institute of Applied Linguistics, Jinan University, Room 422, Main building, Shougouling Road 377, Tianhe District, Guangzhou, Guangdong 510610, China.
Email: nwliumin@gmail.com

(see, e.g., Dewaele, 1999 for French; Durrell, 2011 for German; Quirk, Greenbaum, & Leech, 1972 for English), employing question words (see, e.g., Dornisch, 1998 for Polish; Koutsoudas, 1968 for English; Rojina, 2004 for Russian), or adding interrogative particles (see, e.g., Chao, 1968 for Mandarin; Kuong, 2008 for Cantonese; Tsuchihashi, 1983 for Japanese). Another way frequently adopted across languages to signal questions is via prosodic means, known as intonation. In fact, intonation may be the only means to distinguish questions from statements in syntactically-unmarked yes-no questions (Ultan, 1978; Vaissière, 2008), which usually expresses pragmatic meanings such as surprise, incredulity, or confirmation-seeking (Lee, 2005). In such cases, to express question-statement contrasts, a prominent feature of intonation is its modulation of F0 at the sentential level. However, F0 is not only recruited to convey post-lexical intonation information but also used to distinguish lexical meanings in many tonal languages such as Cantonese and Mandarin.

Tone and intonation are commonly considered as two significant prosodic features of Cantonese and Mandarin speech. Although other acoustic correlates (such as duration, intensity, and voice quality) have also been shown to contribute to cue tonal and intonational contrasts (Kuang, 2017; Whalen & Xu, 1992; Xu, 2009; Yu & Lam, 2014; Yuan, 2006), F0 has been identified as the primary acoustic correlate of both tone and intonation in Cantonese (Fok-Chan, 1974) and Mandarin (Shen, 1985; Wu, 1982; Xu & Wang, 2001).

At the lexical level, there are six phonemically distinct tones in Cantonese. They differ in F0 contours, with T1 having a high-level contour (55)[2], T2 a mid-rising contour (25), T3 a mid-level contour (33), T4 a low-falling contour (21), T5 a low-rising contour (23), and T6 a low-level contour (22). At the sentential level, question intonation in Cantonese is marked by a salient F0 rising in the final syllable of question sentences (Ma et al., 2006). The dual functions of F0 inevitably lead to the interaction of tone and intonation in Cantonese. When occurring sentence-finally in questions, all Cantonese tones obtain a rising tail superimposed onto their canonical F0 contour. As a result, the F0 contour of the low tones (21, 23, 22) in questions resembles that of the mid-rising tone (25).

Mandarin has four lexical tones, with T1 having a high-level contour (55), T2 a mid-rising contour (35), T3 a low-dipping (214), and T4 a high-falling contour (51). At the sentential level, question intonation in Mandarin is generally realized as an upward trend of the F0 contour while statement intonation is realized as a downward trend (Gårding, 1987; Ho, 1977; Liu & Xu, 2005). The dual functions of F0 in Mandarin also result in the interaction of tone and intonation. One may have noticed that when a statement ends with a falling tone (T4) or a question ends with a rising tone (T2), the F0 encodings of the final lexical tone and sentence intonation are in congruency. However, when a statement ends with a rising tone (T2) or a question ends with a falling tone (T4), the F0 encodings of the final lexical tone and sentence intonation are in conflict.

Previous production studies have shown that in Mandarin, intonation-induced F0 primarily affects the F0 height rather than the F0 contour of lexical tones (Cao, 2004; Shen, 1989; Wu,1996). The F0 height of lexical tones increases in questions. However, there is a controversy about the temporal scope in which intonation exerts its effects. Two alternative theories have been proposed: the *global-rising theory* holds that there is a global F0 rising of sentences in questions compared to statements (Ho, 1977; Shen, 1989), whereas the *final-rising theory* claims that the F0 rising in questions is more pronounced towards the end of sentences at the local level (Kratochvil, 1998; Liu & Xu, 2005; Peng et al., 2005; Xu, 2005). Different from these production studies, Liang and Van Heuven (2007) provided perceptual evidence on this issue. They found that manipulating the final rise has a much stronger effect on the perception of intonation type than manipulation of the overall pitch level, which indicates that in Mandarin, the F0 of the final tone is more important than that of the whole sentence for intonation perception.

The aforementioned interactions between tone and intonation in Cantonese and Mandarin regarding their acoustic characteristics may cause pitch processing difficulties in speech perception. In Cantonese, the rising tail superimposed by question intonation onto all sentence-final tones alters the tonal contours of the low tones (21, 23, 22). Consequently, these low tones at the final position of questions run into the risk of being misperceived as the mid-rising tone (25) (Fok-Chan, 1974; Kung et al., 2014; Ma et al., 2006). Intonation identification, in contrast, remains largely accurate (Ma et al., 2011). In Mandarin, intonation-induced F0 has little effect on tone perception and tone identity is maintained in question intonation (Connell et al., 1983). However, there is a significant effect of the final tone on intonation perception. Yuan (2011) found that in Mandarin, questions ending with T4 were easier to identify than questions ending with T2. This is interesting considering that in the former, the F0 encodings of question intonation and the final T4 are in conflict, whereas in the latter, the F0 encodings of question intonation and the final T2 are in congruency. In other words, an asymmetrical intonation perception pattern has been observed for different F0 encodings of question intonation and final lexical tone. A similar asymmetrical pattern of perception was also reported in Xu and Mok (2012a). However, in a follow-up study using low-pass filtered speech (Xu & Mok, 2012b), the pattern was reversed. Mandarin listeners were found to be better at identifying questions ending with T2 than questions ending with T4. These reversed perception patterns might result from many factors, such as prosodic features and lexical intelligibility, among which a potentially very important factor is sentence context.

Sentence context has been shown to play a non-negligible role in language comprehension. It helps comprehenders resolve the identity of lexically ambiguous words (i.e., words with multiple meanings; see Simpson, 1994 for a review). In adverse listening conditions, sentence context helps listeners compensate for noisy or degraded speech input (Patro & Mendel, 2016; Sheldon et al., 2008). Moreover, sentence context has been consistently reported to facilitate language processing, reflected in, for example, reduced processing time or attenuated neural activity (N400) of a word in a highly constraining context versus a weakly constraining context (see Van Petten & Luka, 2012 for a review). Several accounts have been proposed for such a facilitation effect. The *integration account* holds that the facilitation effect can be ascribed to the ease of integration of incoming lexical and semantic information of a word with the on-going discourse context (Hagoort et al., 2004), whereas the *prediction account* presumes that comprehenders use context to pre-activate specific upcoming items which facilitates the form and semantic processing of the target words (Bornkessel-Schlesewsky & Schlesewsky, 2019; DeLong et al., 2005, Wlotko & Federmeier, 2015; for further details, see reviews in, for example, Federmeier, 2007; Kuperberg & Jaeger, 2016; Kutas et al., 2011). Efforts have been made to dissociate the contextual facilitation effect due to ease of integration from that due to prediction of an upcoming word, although more recent evidence suggests that both mechanisms may be at play (Nieuwland et al., 2020). Li and colleagues (2020) simultaneously modeled the integration and the prediction processes by measuring ERP responses at transitive verbs in Chinese sentences with SVO (subject-verb-object) order. They found an inverse correlation between the N400 amplitudes at the verbs and the predictability of their following object nouns, which lends strong evidence to the pre-activation effect. They also showed that lexical tonal variation in different Mandarin Chinese dialects modulates the effects of lexical prediction for listeners with different dialectal experiences.

It is important to note that context-dependent predictive processing has been argued to be present at multiple levels of linguistic representation, such as semantic (Altmann & Kamide, 1999; Federmeier & Kutas, 1999; Van Petten et al., 1999), syntactic (Van Berkum et al., 2005; Wicha et al., 2003), and phonological (for segments, see Allopenna et al., 1998; DeLong et al., 2005; for prosody, see Bishop, 2012; Buxó-Lugo & Watson, 2016; Cole et al., 2010) information. Of particular relevance to the present paper are studies that reported the effect of semantic

context on tone and intonation processing. In Mandarin, Ye and Connine (1999) investigated tone processing with the target syllables occurring in sentence-final position in a semantically highly constraining context (i.e., idiomatic context) and a semantically neutral context (i.e., a carrier sentence with neutral semantic meaning). They found that the constraining semantic context considerably facilitated the processing of tone. In Cantonese, as mentioned earlier, low tones are easily misperceived as the mid-rising tone in the final positions of questions (Ma et al., 2006). When embedding the low and mid-rising tone words sentence-finally in a semantically neutral context versus a semantically strong biasing context (i.e., a disyllabic word context), Kung et al. (2014) found that the latter context led to much better lexical identification performance for words with a low tone at the end of questions than the former context. This led them to conclude that semantic context plays a major role in disentangling tonal information from intonational information in Cantonese when tone and intonation interact.

In contrast to the tone processing difficulty in questions in Cantonese, the interaction of tone and intonation in Mandarin leads to intonation processing difficulty (Xu & Mok, 2012a, 2012b; Yuan, 2011). This contrast invites further research on the potential typology of the interaction between tone and intonation in tonal languages. Moreover, while we know that context facilitates tone processing in Mandarin, the specific role of context, in particular its role in intonation processing and in disentangling intonation from tone processing, remains unclear. Therefore, the present study was designed to investigate how tone and intonation are processed in Mandarin as a function of semantic context when F0 encodings of the final lexical tone and sentence intonation are in conflict or in congruency. Two semantic contexts were constructed to address this issue: a semantically neutral context and a semantically constraining context. In each semantic context, tone and intonation identification experiments were performed using the same design with the same group of participants, allowing for a direct systematic comparison of tone versus intonation identification. The resulting measurements included the commonly-reported response accuracy, as well as an additional measurement, reaction time, which measures the amount of time participants need to identify tone/intonation. There has been a long history of psycholinguistic research which shows convincingly that reaction time serves as a good indicator of the degree of difficulty of a perceptual decision: the more difficult a perceptual decision is, the longer the reaction time (Donders, 1969; Luce, 1986). We expect that reaction time, together with the response accuracy, would reveal the pitch processing difficulties entailed in Mandarin listeners' judgement of tonal and intonational features.

We hypothesized that Mandarin listeners would encounter more difficulties in intonation processing than tone processing. However, a semantically constraining context may help Mandarin listeners disentangle intonational information from tonal information, and ease the intonation processing difficulties. This should be reflected in higher response accuracy and shorter reaction times for question intonation identification in the semantically constraining context than in the semantically neutral context.

2 Method

2.1 Materials

2.1.1 Materials in the semantically neutral context. Forty monosyllabic word pairs with minimal tonal contrast (T2 vs. T4) and otherwise identical segments were selected. All the pairs of words occurred in the final position of a five-syllable carrier sentence, that is, *ta1 gang1gang1 shuo1 X* (English translation: "She just said X"), produced with either a statement (S) or a question (Q) intonation. Note that only high-level tones were contained in the carrier sentence. This is to avoid

Table 1. An example of the experimental stimuli in the semantically neutral context.

Conditions		Examples				
Tone	Intonation					
Tone2	Statement	Characters	她	刚刚	说	**X(财)。**
		Pinyin	ta1	gang1gang1	shuo1	**cai2**
		IPA	[tʰa1]	[kaŋ1kaŋ1]	[ʂuo1]	**[tsʰai2]**
		English	She	just	said	**money.**
Tone2	Question	Characters	她	刚刚	说	**X(财)?**
		Pinyin	ta1	gang1gang1	shuo1	**cai2**
		IPA	[tʰa1]	[kaŋ1kaŋ1]	[ʂuo1]	**[tsʰai2]**
		English	She	just	said	**money?**
Tone4	Statement	Characters	她	刚刚	说	**X(菜)。**
		Pinyin	ta1	gang1gang1	shuo1	**cai4**
		IPA	[tʰa1]	[kaŋ1kaŋ1]	[ʂuo1]	**[tsʰai4]**
		English	She	just	said	**vegetable.**
Tone4	Question	Characters	她	刚刚	说	**X(菜)?**
		Pinyin	ta1	gang1gang1	shuo1	**cai4**
		IPA	[tʰa1]	[kaŋ1kaŋ1]	[ʂuo1]	**[tsʰai4]**
		English	She	just	said	**vegetable?**

Note. The critical syllables are in bold.

the down-step effect and to minimize the contribution of tone to the observed F0 movement (Shih, 2000). The carrier sentence was semantically meaningful but offered neutral semantic information to the target stimuli and will thus be referred to as the semantically neutral context hereafter.

For each minimal pair of T2–T4 words, their word frequencies and homophone densities were comparable, and their syntactic word categories were the same. To avoid any word frequency effect, frequent words were selected only when they have both more than 4,500 occurrences in a corpus of 193 million words (Da, 2004) and a log10 word frequency above 3.0 in the SUBTLEX-CH frequency list (Cai & Brysbaert, 2010). The average log10 word frequencies were 3.67 (SD = 0.64) for T2 and 3.95 (SD = 0.60) for T4 words. Following Ziegler et al. (2000) and Chen et al. (2009), homophone density was defined as the number of homophone mates of a word, that is, words that contain exactly the same phonetic segments and lexical tones. We ensured that T2 words ($M \pm SD$: 15.05 ± 10.65) had similar homophone densities as their T4 equivalents ($M \pm SD$: 15.05 ± 11.25). As for the syntactic word category, the 40 word pairs comprised mainly pairs of nouns (32), but pairs of verbs (6) and adjectives (2) were also included to guarantee a sufficient number of stimuli.

In total, 160 target sentences (40 Syllables × 2 Tones × 2 Intonations) were constructed (see Table 1 for an example). In addition, 240 filler sentences were included for the perception experiment. The filler sentences possess the same carrier but different critical syllables in terms of either segmental composition or lexical tone (e.g., T1/T3).

2.1.2 Materials in the semantically constraining context. To avoid learning effects from the experimental stimuli in the semantically neutral context, an additional set of 40 monosyllables in combination with tone (T2 or T4) was selected. Each minimal pair of T2–T4 monosyllables were the second syllables of two disyllabic words with comparable word frequency. The disyllabic words were then embedded in the final position of various nine- or ten-syllable natural sentences. This

Table 2. An example of the experimental stimuli in the semantically constraining context.

Conditions		Examples						
Tone	Intonation							
Tone2	Statement	Characters	这家	旅馆	有	三十	间	客**房**。
		Pinyin	zhe4jia1	lv3guan3	you3	san1shi2	jian1	ke4**fang2**
		IPA	[tʂɤ4tɕia1]	[ly3kuan3]	[iou3]	[san1ʂ1̩2]	[tɕiæn1]	[kʰɤ4**faŋ2**]
		English	This hotel has thirty guest rooms.					
Tone2	Question	Characters	这家	旅馆	有	三十	间	客**房**？
		Pinyin	zhe4jia1	lv3guan3	you3	san1shi2	jian1	ke4**fang2**
		IPA	[tʂɤ4tɕia1]	[ly3kuan3]	[iou3]	[san1ʂ1̩2]	[tɕiæn1]	[kʰɤ4**faŋ2**]
		English	This hotel has thirty guest rooms**?**					
Tone4	Statement	Characters	海瑞	故居	将	向	游人	开**放**。
		Pinyin	Hai3 Rui4	gu4ju1	jiang1	xiang4	you2ren2	kai1**fang4**
		IPA	[xai3 ʐuei4]	[ku4tɕy1]	[tɕiɑŋ1]	[ɕiɑŋ4]	[iou2ʐən2]	[kʰai1**faŋ4**]
		English	Hai Rui's former residence will be open to visitors.					
Tone4	Question	Characters	海瑞	故居	将	向	游人	开**放**？
		Pinyin	Hai3 Rui4	gu4ju1	jiang1	xiang4	you2ren2	kai1**fang4**
		IPA	[xai3 ʐuei4]	[ku4tɕy1]	[tɕiɑŋ1]	[ɕiɑŋ4]	[iou2ʐən2]	[kʰai1**faŋ4**]
		English	Hai Rui's former residence will be open to visitors?					

Note. The critical syllables are in bold.

sentence context was verified to provide sufficient constraint on the final syllable and will be referred to as the semantically constraining context hereafter.

According to the SUBTLEX-CH frequency list (Cai & Brysbaert, 2010), the average log10 word frequencies were 2.49 (SD = 0.48) for the disyllabic words ending in T2 and 2.71(SD = 0.67) for those ending in T4. The reason for us using the disyllabic word as part of the sentence context frame is that it is the predominant word type in Mandarin, and most often used in natural sentences (Duanmu, 2007). Furthermore, previous studies (Xu & Mok, 2012a, 2012b; Yuan, 2011) have embedded disyllabic words sentence-finally in their studies. Our similar set-up thus enables a comparison of results.

We conducted a cloze probability pretest to verify that our sentence context provides sufficient constraint on the final syllable. The cloze probability of a word is the percentage of participants who offer that word as a completion for a sentence of which the final word is missing (Van Petten & Luka, 2012). Thirty native Mandarin speakers participated in the pretest. They were visually presented with the sentences but without the final syllable, and were asked to provide the most likely syllable that fits the given sentence frame. Consequently, each final syllable had a cloze probability of at least 70%.

As in the semantically neutral context, all the sentences in the semantically constraining context were produced with either a statement (S) or a question (Q) intonation, yielding another 160 target sentences (40 Syllables × 2 Tones × 2 Intonations, see Table 2 for an example). Fillers (240 sentences) were also included in the experiment.

2.2 *Recording and stimuli preparation*

A female native speaker of Mandarin, born and raised in Beijing, recorded the sentences in a soundproof recording booth at the Phonetics Laboratory of Leiden University. Sentences were

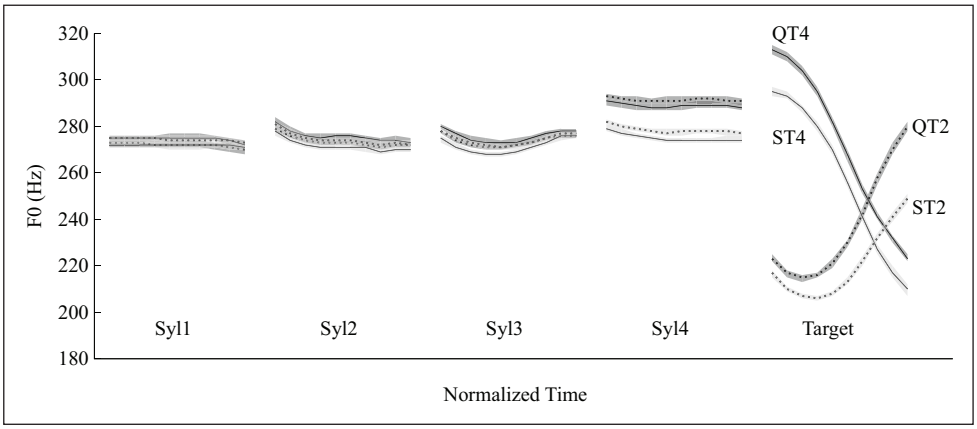

Figure 1. F0 contours of the T2 and T4 syllables in question versus statement intonation, produced within the frame sentence "*ta1 gang1gang1 shuo1*" with all syllables containing T1 (reproduced from Liu et al., 2016). The F0 value of each syllable was averaged over 40 stimulus tokens, with each syllable represented by 10 equally distanced F0 values taken from the rhyme part of the time-normalized syllable. Solid lines indicate the mean F0 of T4 syllables in question intonation (dark solid) and in statement intonation (light solid). Dotted lines indicate the mean F0 of T2 syllables in question intonation (dark dotted) and in statement intonation (light dotted). Grey areas indicate ±1 *SD* of the mean.

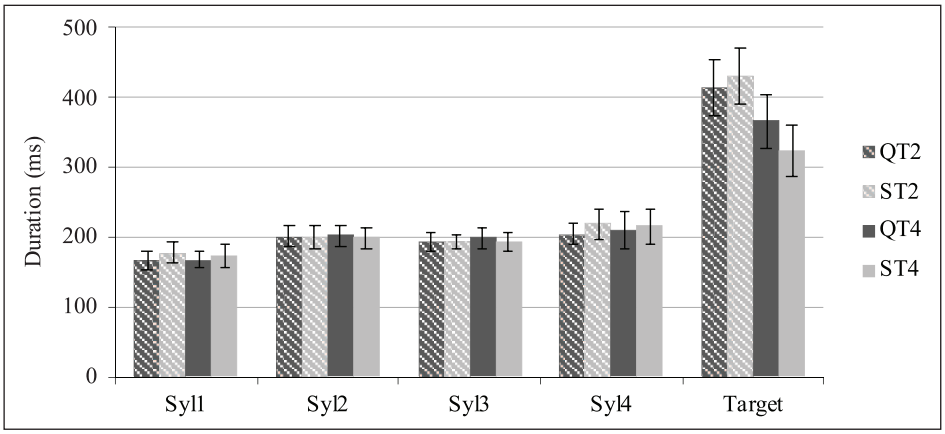

Figure 2. Duration mean ± 1 *SD* for each syllable of the sentences under four experimental conditions (adapted from Liu et al., 2016).

randomly presented to the speaker using an HTML JavaScript and recorded with a Sennheiser MKH416T microphone at 16-bit resolution and a sampling rate of 44.1 kHz using Adobe Audition 2.0. To eliminate paralinguistic information, the speaker was instructed to avoid any exaggerated emotional prosody during the recording.

This female speaker's recordings were clear and consistent. We analyzed the acoustic realization of the stimuli in the semantically neutral context. The acoustic results (see Figures 1 and 2, and also Liu et al., 2016 for details) of the recordings showed comparable F0 realization of tone and intonation to a prior study (Yuan, 2006) and were taken as the representative patterns of question

and statement intonation in Mandarin, and therefore suitable for the perception study. In the subsequent perception experiment, the amplitude of all the sentences was normalized in Praat (Boersma & Weenink, 2015).

All the stimuli were manually annotated in Praat (Boersma & Weenink, 2015). A custom-made script was then used to extract F0 and duration values. Gross errors in F0 extractions were manually corrected. Statistical analyses for each parameter were carried out using the paired *t*-test in R version 4.0.1 (R Core Team, 2020). Results of the acoustic analysis of the speech files in the semantically neutral context showed that over our stimulus sentences (containing five syllables), the mean F0 of the first three syllables and the duration of the first four syllables revealed no differences between the two types of intonation for sentences ending with both T2 and T4 (all ps $>$ 0.05). However, intonation affected the F0 of the fourth syllable significantly: the mean F0 of the level tone was higher in questions than in statements. This holds for both the T2 condition, $t(39) = 7.22$, $p < 0.001$, Cohen's $d = 1.64$, and the T4 condition, $t(39) = 7.17$, $p < 0.001$, Cohen's $d = 1.64$.

Regarding the critical syllable (i.e., the fifth syllable of the stimulus sentences), questions with a final T2 had a significantly wider F0 range than their statement counterparts, $t(39) = 9.87$, $p < 0.001$, Cohen's $d = 1.90$. The minimum F0 showed only a slight increase, $t(39) = 6.37$, $p < 0.001$, Cohen's $d = 1.27$, while the increase of the maximum F0 was very salient, $t(39) = 11.47$, $p < 0.001$, Cohen's $d = 2.30$, resulting in a sharper final rising trend in questions than in statements. For questions with a final T4, an overall higher F0 contour was observed relative to the statements with a final T4 condition. Despite a comparable F0 range between the two conditions, $t(39) = 1.24$, $p = 0.22$, Cohen's $d = 0.30$, both the maximum F0, $t(39) = 8.65$, $p < 0.001$, Cohen's $d = 1.40$ and the minimum F0, $t(39) = 4.48$, $p < 0.001$, Cohen's $d = 0.80$, of the high-falling tone (T4) were significantly higher in questions than in statements. Concerning duration, final T4 syllables in question intonation were significantly longer than those in statement intonation, $t(39) = 6.73$, $p < 0.001$, Cohen's $d = 1.11$. Final T2 syllables, however, tended to be relatively shorter in question intonation than in statement intonation, $t(39) = -3.10$, $p < 0.01$, Cohen's $d = -0.41$.

Our data are consistent with previous findings (Cao, 2004; Wu, 1996; Yuan, 2006). A consensus has emerged that in Mandarin, the pitch contour of T2 and T4 is maintained in both question and statement intonations, but the pitch height differs between the two types of intonation across final tone identities. In addition to F0, duration also seems to play a role in distinguishing between question and statement intonation. Considering the temporal scope of the interaction of tone and intonation, our data support neither the global-rising nor the strictly local final-rising theory of question intonation, but they are more in line with the final-rising theory given that F0 does not increase significantly before the penultimate syllable.

To further verify the validity of the intonation patterns perceptually, a panel of five phonetically trained researchers who are native speakers of Mandarin was asked to evaluate the typicality of the intonation of the sentences on a five-point scale (1 = "very typical statement"; 5 = "very typical question"). The evaluation was conducted for sentences in both the semantically neutral and constraining contexts, but in different sessions. While all tokens produced by our speaker were included in the perception experiment, only tokens identified as typical of their corresponding intonation category (i.e., score $\leqslant 1.5$ for statements and score $\geqslant 3.5$ for questions) by at least three out of the five researchers were selected for the data analysis reported below. The average typicality rating score for the final set of selected stimuli with a semantically neutral context in the data analysis was 1.1 for statements and 4.5 for questions, resulting in an exclusion of 13.1% of the data points. Likewise, the average typicality rating score for the remaining selected stimuli with a semantically constraining context in the data analysis was 1.0 for statements and 4.4 for questions. In this context, 13.8% of the data points were excluded.

2.3 Participants

Eighteen native speakers of Mandarin (10 female, 8 male) from Northern China were paid to participate in the experiment. All the selected participants achieved the 1B level in the Putonghua Shuiping Ceshi (Mandarin Proficiency Test), indicating that they have native proficiency in Mandarin without regional accents. 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 reported any speech or hearing disorders. Informed consent was obtained from all the participants before the experiment.

2.4 Procedure

Participants were tested individually in a sound-attenuated room in two different sessions for the semantically neutral context and the semantically constraining context. Note that the session for the semantically constraining context was always ran after the session for the semantically neutral context for each participant, with a long break in-between. At each session, 400 sentence trials (including 160 targets and 240 fillers) were randomly presented to the participants using the E-Prime 2.0 software through headphones (AKG K242HD) at a comfortable listening level. Instructions were given both visually on screen and orally by the experimenter in Mandarin before the experiment.

Each experimental session included a practice block and four experimental blocks. The practice block contained 12 trials. Each experimental block contained 100 trials. Between two blocks, there was a short 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 were requested to carry out either a tone identification task or an intonation identification task as quickly and accurately as possible. For each test session, half of the trials contained the tone identification task while the other half contained the intonation identification task; the task varied randomly from trial to trial. Task types were indicated by the tone and intonation marks in Mandarin Pinyin system, that is, the official romanization system for Mandarin, which all participants knew very well. For example, when "ˊ ˋ" marks ("ˊ" stands for T2; "ˋ" stands for T4) appeared on the screen, participants were asked to identify whether the final tone of the sentence was T2 or T4. When the "。？" marks appeared on the screen ("。" stands for statement intonation; "？" stands for question intonation), they were asked to identify whether the sentence bore a statement or question intonation. Listeners were given up to 2 seconds after the offset of the sentence to respond. No participants reported difficulty in understanding the tasks. The inter-stimulus interval was 500 ms.

2.5 Data analysis

Previous studies on intonation perception have typically only reported response accuracy (Xu & Mok, 2012a, 2012b; Yuan, 2011). In this study, in addition to response accuracy, reaction time was included as a dependent variable. Response accuracy was defined as the percentage of correct identification of tone in the tone identification task and the percentage of correct identification of intonation in the intonation identification task. Reaction time was defined as the response time relative to the last syllable's onset for correct responses.

Statistical analyses were carried out with the package *lme4* (Bates et al., 2015) in R version 4.0.1 (R Core Team, 2020). Analysis of response accuracy was performed using binomial logistic regression models, and analysis of reaction time was performed using linear mixed-effects

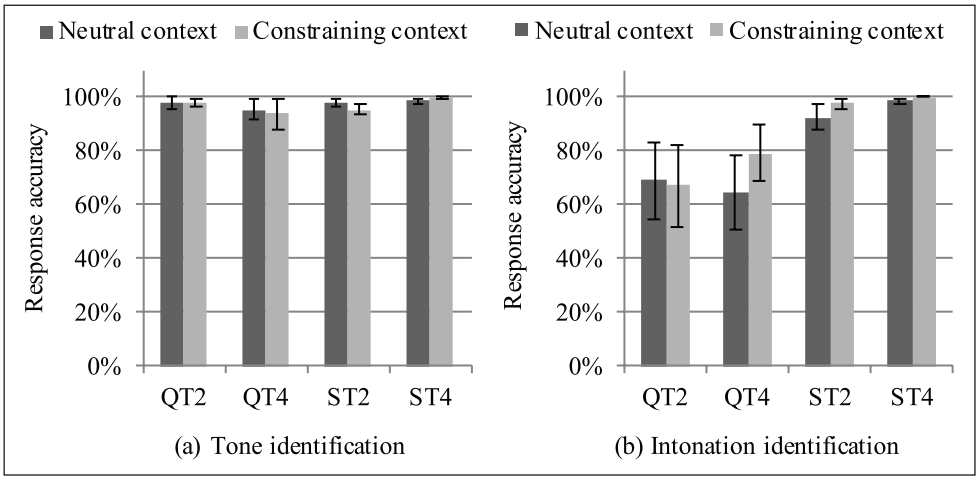

Figure 3. Response accuracy of each experimental condition in the semantically neutral context (dark grey bars) and in the semantically constraining context (light grey bars) for the tone identification task (a) and the intonation identification task (b). The error bars represent the corresponding 95% confidence interval of the means across participants.

regression models. Note that we analyzed the raw reaction time, as it yielded models with normally distributed residuals when excluding outliers. Trials of the semantically neutral context and those of the semantically constraining context were merged into one dataset for model constructions. All models first included random intercepts of subjects and items. The fixed effects of Context (Semantically neutral context, Semantically constraining context), Task (Tone identification, Intonation identification), Tone (T2, T4), Intonation (Q, S) and their interactions were then added in a stepwise fashion. All effects on model fits were evaluated via model comparisons based on log-likelihood ratios. Finally, the random slopes of the significant fixed effects on subjects and items were added if model fits were significantly improved. For models of reaction time, trials with residuals from the overall model exceeding 2.5 absolute standardized deviations were considered as outliers and removed from further analysis. We also considered trial-by-trial dependency in model constructions. However, it did not significantly improve the model fit, and was therefore excluded in the final model.

3 Results

3.1 Response accuracy

Figure 3 presents the response accuracy of each experimental condition in the semantically neutral and constraining contexts for the tone identification task (a) and the intonation identification task (b). The error bars represent the corresponding 95% confidence interval of the means across participants.

The overall analyses of response accuracy showed a significant main effect of Task, $\chi^2(1) = 51.65$, $p < 0.001$, and a significant main effect of Intonation, $\chi^2(1) = 126.27$, $p < 0.001$. A two-way interaction of Context × Task, $\chi^2(1) = 7.75$, $p = 0.005$, and a three-way interaction of Context × Task × Intonation, $\chi^2(3) = 37.31$, $p < 0.001$, were also significant.

As shown in Figure 3(a), tone identification was almost at ceiling level across the experimental conditions. The very few incorrect responses were likely motor-related errors as a result of the

speed requirement of the task. No main effect of Context or any interaction of Context with other factors was found for the tone identification task (all ps > 0.05). It seems that tone identification was not affected by intonation information, irrespective of semantic contexts.

For the intonation identification task (see Figure 3(b)), we found a significant main effect of Context, $\chi^2(1) = 5.48$, $p = 0.02$, and a significant three-way interaction of Context × Tone × Intonation, $\chi^2(2) = 6.17$, $p = 0.045$. Further analyses for subset data of different intonations confirmed the main effect of Context in statements across final lexical tone identities, $\chi^2(1) = 10.45$, $p = 0.001$, and in questions ending with T4, $\chi^2(1) = 8.06$, $p = 0.005$, but not in questions ending with T2, $\chi^2(1) = 0.07$, $p = 0.79$, which suggests that the response accuracy of intonation in the former three conditions increased in the semantically constraining context compared to their semantically neutral counterparts, ST2: 97.7% versus 92.7%; ST4: 100% versus 98.6%; QT4: 79.0% versus 64.5%. In questions ending with T2, the response accuracy of question intonation in the semantically constraining context was inclined to decrease if compared to that in the semantically neutral context, QT2: 67.2% versus 69.2%.

Separate models were also constructed for subset data of different contexts for the intonation identification task. In the semantically neutral context, we found a significant main effect of Intonation, $\chi^2(1) = 76.55$, $p < 0.001$, and a significant interaction of Tone × Intonation, $\chi^2(1) = 12.19$, $p < 0.001$. Overall, question intonation tended to be more difficult to identify than statement intonation regardless of the final lexical tone identities. Specifically, statement intonation was more accurately identified in statement sentences ending with T4 than in those ending with T2, $\beta = 1.55$, $z = 2.08$, $p = 0.04$, whereas question intonation was equally difficult to identify in question sentences ending with T2 and T4, $\beta = -0.42$, $z = -0.80$, $p = 0.42$. In the semantically constraining context, we found a significant main effect of Intonation, $\chi^2(1) = 71.97$, $p < 0.001$, and a marginally significant interaction of Tone × Intonation, $\chi^2(1) = 3.69$, $p = 0.05$. Importantly, question intonation was more accurately identified in questions ending with T4 than in those ending with T2, $\beta = 1.06$, $z = 2.01$, $p = 0.045$.

To sum up, the identity of lexical tone was not hindered by intonation information irrespective of semantic contexts. Tone identification reached almost ceiling levels across all experimental conditions. In contrast, Intonation identification, particularly question intonation identification, was much less accurate. The identification of intonation was susceptible to the final lexical tone identity and affected by the semantic context. In a semantically constraining context, the response accuracy for intonation increased relative to those in the semantically neutral context except in questions ending with a rising T2, where a slight decrease of response accuracy was found. Consequently, in the semantically neutral context, questions ending with T2 and T4 were equally poorly identified. In the semantically constraining context, questions ending with T4 were more accurately identified than questions ending with T2.

3.2 Reaction time

Figure 4 presents the average reaction time of each experimental condition in the semantically neutral and constraining contexts for the tone identification task (a) and the intonation identification task (b). The error bars represent the corresponding 95% confidence interval of the means across participants.

We identified 126 trials (2.76% data points) as outliers and removed them from the analyses. The overall analyses of the remaining data points showed significant main effects of all fixed factors on reaction time: Context, $\chi^2(1) = 146.41$, $p < 0.001$, Task, $\chi^2(1) = 42.30$, $p < 0.001$, Tone, $\chi^2(1) = 25.31$, $p < 0.001$, and Intonation, $\chi^2(1) = 44.13$, $p < 0.001$. Moreover, two-way interactions of Context × Task, $\chi^2(1) = 33.63$, $p < 0.001$, Context × Intonation, $\chi^2(1) = 5.96$, $p = 0.015$,

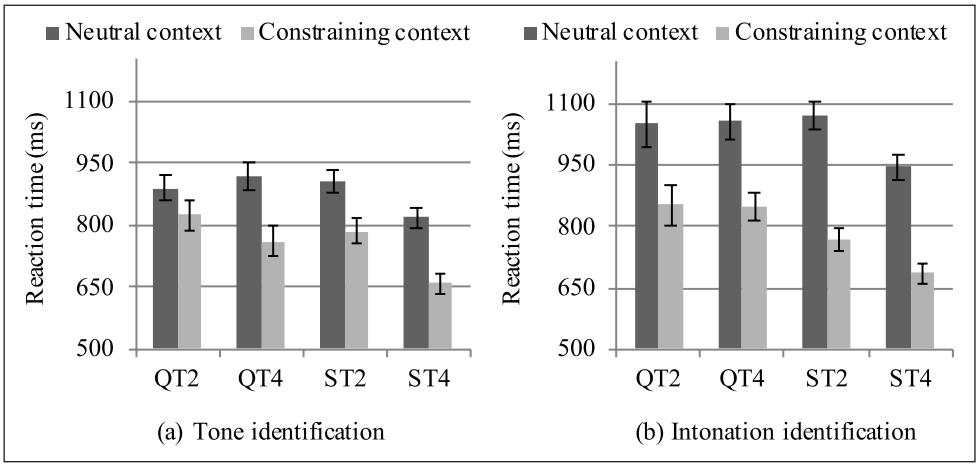

Figure 4. Average reaction time of each experimental condition in the semantically neutral context (dark grey bars) and in the semantically constraining context (light grey bars) for the tone identification task (a) and the intonation identification task (b). The error bars represent the corresponding 95% confidence interval of the means across participants.

and Tone × Intonation, $\chi^2(1) = 23.10$, $p < 0.001$, as well as the three-way interaction of Context × Tone × Intonation, $\chi^2(2) = 3.22$, $p = 0.02$, were also significant.

We will first report the effects of Task. In the semantically neutral context, there was a significant main effect of Task, $\chi^2(1) = 72.40$, $p < 0.001$, and no interactions of Task with other factors were found. This suggests that the final lexical tone was overall identified faster than sentence intonation across final lexical tones and sentence intonations. In the semantically constraining context, there was a significant two-way interaction of Task × Intonation, $\chi^2(1) = 4.59$, $p = 0.03$. Participants were faster in identifying the final lexical tone than sentence intonation in question sentences. $\beta = -70.22$, $t = -2.94$, $p = 0.003$, but not in statement sentences, $\beta = -4.79$, $t = -0.20$, $p = 0.84$, where comparable reaction times were recorded to identify the final lexical tone and sentence intonation.

The effects of Context were then explored by constructing separate models for subset data of different tasks. Results for the tone identification task showed a significant main effect of Context, $\chi^2(1) = 54.41$, $p < 0.001$, and a significant interaction of Context × Tone, $\chi^2(1) = 4.93$, $p = 0.03$. Clearly, the reaction time to identify the final lexical tone was considerably shorter in the semantically constraining context relative to that in the semantically neutral context across all experimental conditions (see Figure 4(a)). Furthermore, the effect of Context was greater for T4 identification compared to its T2 equivalent regardless of the intonation types, as evidenced by the larger reaction time difference between the two semantic contexts for T4 than for T2. This suggests that participants benefited more from the semantically constraining context when identifying T4 as compared to T2. Likewise, results for the intonation identification task revealed a significant main effect of Context, $\chi^2(1) = 120.41$, $p < 0.001$, and a significant interaction of Context × Intonation, $\chi^2(1) = 7.45$, $p = 0.006$. Intonation identification was generally faster in the semantically constraining context than in the semantically neutral context across conditions. However, semantic context did not affect the identification of different types of intonation to the same degree. The semantically constraining context seemed to contribute more to the identification of statement intonation than question intonation regardless of the final lexical tone identities.

Finally, it should be noted that the analyses for the subset of different tasks revealed a significant two-way interaction of Tone × Intonation in both the tone identification task, $\chi^2(1) = 11.38$, $p < 0.001$, and the intonation identification task, $\chi^2(1) = 11.34$, $p < 0.001$. An investigation of the interaction effect revealed shorter reaction time to identify statement intonation in statements ending with T4 relative to statements ending with T2. Still, when identifying question intonation, no reaction time difference was found between questions ending with T2 and those ending with T4 in both tasks.

To sum up, context considerably affected the reaction time needed to identify the final lexical tone and sentence intonation. The semantically constraining context played a significant role in speeding up the identification of both final lexical tone and sentence intonation across the experimental conditions. It shortened the reaction time more in identifying intonation than tone, in identifying T4 than T2, and also in identifying statement intonation than question intonation. More specifically, in the semantically neutral context, the final lexical tone was identified faster than sentence intonation across all lexical tones and both intonation types. In the semantically constraining context, the identification of final lexical tone was faster than that of the sentence intonation in question sentences, but not in statement sentences. Reaction time needed to identify question intonation did not differ between the questions ending with T2 and T4 conditions. In both conditions, their reaction time was shortened in the semantically constraining context than in the semantically neutral context.

Taking together the results of response accuracy and reaction time, it can be seen that tone identification maintained its near-ceiling-level identification accuracy across all experimental conditions in both semantic contexts, and notably with shorter reaction time in the semantically constraining context than in the semantically neutral context. Intonation identification, particularly question intonation identification, however, was susceptible to the final lexical tone identity and also affected by the semantic context. In the semantically neutral context, questions ending with a rising T2 and a falling T4 were equally poorly identified with relatively long reaction times. In the semantically constraining context, questions ending with a falling T4 were more accurately identified than questions ending with a rising T2, and with considerably shorter reaction times than in the semantically neutral context.

4 General discussion

To address the question of how top-down information provided by semantic contexts affects tone and intonation processing in Mandarin when F0 encodings of the final lexical tone and sentence intonation are in conflict or in congruency, we examined the identification of tone and intonation in both semantically neutral and constraining contexts. Our results demonstrated that in Mandarin, tone identification was seldom affected by intonation information irrespective of semantic contexts, whereas intonation identification, particularly question intonation, was susceptible to the final tone identity and affected by the semantic context. A semantically constraining context considerably improved question intonation identification, but mainly so in question sentences with the lexical falling T4 in the final position.

In our study, the overall performance of tone identification was better than that of intonation identification regardless of semantic contexts in Mandarin. Evidence was found not only from the response accuracy results, but also from the reaction time patterns. Intonation identification took more time than tone identification irrespective of the final lexical tone identities, presumably because intonation processing requires integrating more information over a longer period of time than lexical tone processing. It seems that in Mandarin, when pitch movements are used to convey post-lexical contrast, its identification becomes a much more difficult decision-making process

(Braun & Johnson, 2011). The advantage of tone over intonation is probably because a phonetic dimension (i.e., F0) exploited for one function of the grammar (e.g., lexical tone) would, to some extent, limit its effectiveness to cue a different function (e.g., intonation) in the same linguistic system (Liang & Van Heuven, 2007; Nolan, 2006; Torreira et al., 2014). Probably as a compensation mechanism, tonal languages have been found to develop more syntactic devices to express sentence-level meanings typically encoded by intonation than non-tonal languages (Torreira et al., 2014). With respect to Mandarin, as argued in Gussenhoven and Van de Ven (2020), it is perhaps not coincidental that Mandarin has developed rich sentence-final question particles to express interrogativity.

For intonation identification, previous studies found reversed patterns of question intonation identification in questions ending with rising T2 and falling T4 in normal context (Xu & Mok, 2012a; Yuan, 2011) versus in low-pass filtered context (Xu & Mok, 2012b). It is unclear whether the reversed patterns are due to the different test contexts in semantic or other information. The present study teased apart the effect of semantic context from the other factors by introducing the semantically neutral versus constraining contexts. We found that the semantically neutral context posed greater difficulty to question intonation identification, compared to the semantically constraining context. In the former, questions ending with T2 and T4 were equally poorly identified. In the latter, questions ending with T4 were better identified than those ending with T2, in line with the results in Xu and Mok (2012a) and Yuan (2011). Recall that in low-pass filtered speech, questions ending with T2 even had a higher response accuracy than questions ending with T4 (Xu & Mok, 2012b). Therefore, it seems that context plays a role in question intonation identification. The stronger and more informative the linguistic context is (i.e., semantically constraining context $>$ semantically neutral context $>$ low-pass filtered context), the better the identification of questions ending with T4. The opposite pattern was observed for questions ending with T2, with better identification of question intonation for weaker and less informative linguistic context. We infer that with less semantic information, the frequency code (Gussenhoven, 2004; Morton, 1994; Ohala, 1983), which holds that high or rising pitch marks questions while 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 (69.2% vs. 67.2%). Afterall, the pitch encodings of T2 (rising) and question intonation (rising) are in congruency. Whether the rising of final T2 in questions is tone-related or intonation-induced is ambiguous in nature. Compared to questions ending with T2, the conflicting pitch encodings of T4 (falling) and question intonation (rising) makes semantic information more useful in the questions ending with T4 condition, which is confirmed by the facilitation effect of the semantically constraining context.

If response accuracy speaks for context effects only in question intonation identification where overt processing difficulties occur, reaction time lends stronger support to the effects of context in a broader sense. A semantically constraining context speeded up not only intonation identification, but also tone identification compared to the semantically neutral context. It shortened reaction times for intonation identification to a larger extent.

In the semantically neutral context, very limited information except bottom-up acoustic information was available to the participants for them to identify tone and intonation. With one and the same carrier sentence for all the final target syllables, participants had no knowledge of what the last syllable would be until they actually heard it. It was therefore not possible for participants to have predicted the tone of the last syllable before it actually occurred. From the standpoint of information theory (Shannon, 1948), the entropy (i.e., the surprisal of a word in a context, which indexes the amount of information a word conveys) of the upcoming final syllable was very high. Evidence has shown that a word's entropy is an important predictor of its processing cost (Frank, 2013; Hale, 2001;

Levy, 2008). The larger the entropy of a word, the more processing cost associated with the word. It thus required more resources to process the incoming information in the semantic neutral context, leading to lower response accuracy and longer reaction time. In contrast, in the semantically constraining context, the final target syllables had lower entropy due to their high predictability. Participants could easily predict the target syllables based on the available context information prior to hearing the syllables. Such syllable prediction allows for the pre-activation of both the segment and tone of the syllable (Ye & Connine, 1999). Therefore, it is not surprising that tone identification was faster in the semantically constraining context relative to the semantically neutral context. With fewer processing resources taken up by the tone identification task, participants could devote more attention and processing resources to identifying the intonation of the sentences, which in turn led to the overall improved response accuracy and shorter reaction time in the intonation identification. One other possible explanation for the better intonation identification is that in the semantically constraining context, participants might be able to better "subtract" the tone influence on the F0 contour given the tone's context-induced pre-activation, and hence they are better at decoding the subtle pitch cue for intonation identification.

Though a semantically constraining context considerably shortened reaction times to identify both the final lexical tone and sentence intonation, it did not entirely resolve the processing difficulty in intonation identification. Participants still had difficulty identifying question intonation, especially when the question intonation concurred with T2. The processing difficulty of question intonation here reflects actual difficulties in teasing question intonation apart from surface pitch information (which also signals tone information), rather than simple difficulties in meta-linguistic judgment. This is supported by evidence from an ERP study (Liu et al., 2016), in which Mandarin sentences in a semantically neutral context (the same as the sentences with a semantically neutral context in this study) were auditorily presented to native listeners and their electrophysiological responses were recorded. The participants were asked to pay special attention to the final lexical tone and sentence intonation but were not asked to make any motor-related responses or metalinguistic judgments. A clear P3b effect was found for questions ending with T4 relative to statements ending with T4, whereas no ERP effect was found for questions ending with T2 relative to statements ending with T2. The ERP results thus suggest that the question-statement contrast in the T4 conditions is easier to categorize, whereas categorization of the question-statement contrast in the T2 conditions is much more demanding for Mandarin native listeners. Comparing questions ending with T4 to statements ending with T4, the former tends to be more difficult and requires more processing effort as evidenced by the P3b effect in questions ending with T4.

In our study, we intermixed tone and intonation identification tasks in our design. One might wonder whether such mixture have increased the difficulty of tasks and resulted in task switching costs. It is worth bearing in mind that that our study attempted to look into the interaction of tone and intonation in Mandarin, so we had to make sure that both tone and intonation information were attended to when participants heard the stimuli. If tone identification and intonation identification tasks were presented in separate blocks, participants might have selectively tuned to one aspect of information and neglected the other. Therefore, we deliberately intermixed tone and intonation identification tasks with the task varied randomly from trial to trial in our design. We found that switching tasks from one trial to the next did not affect the response accuracy of either task compared to the no task-switching conditions. However, when the task was switched from tone identification to intonation identification, reaction time for the intonation identification was significantly longer compared to the corresponding reaction time for the intonation identification with a preceding intonation identification task. The reverse direction did not lead to a similar effect; switching from the intonation identification task to the tone identification task did not affect the reaction time for tone identification significantly, in comparison to the condition with two consecutive tone

identification trials. In short, there were indeed task-switching costs in terms of reaction time, but mainly when the task was switched from tone identification to intonation identification. This, in a sense, is in line with our finding that tone identification is overall easier than intonation identification.

One remaining question is why the question intonation identification demonstrated large individual differences among participants in both questions ending with T2 and T4 across the semantically neutral and constraining contexts. Some participants were able to identify almost all the question intonation without difficulty, whereas others could only identify a few question intonations correctly. It might be that the question-statement contrast in Mandarin is not expressed or perceived categorically, as has been found in Zhumadian Mandarin (Gussenhoven & Van de Ven, 2020), but further research is needed before we can make any strong claim. The challenge here in Mandarin intonation processing echoes the fundamental challenge facing prosody perception in general, which partly, if not all, comes down to the ambiguity arising from the continuous and variable nature of the prosodic signal itself (Tanenhaus et al., 2015).

Last but not least, the present study on tone and intonation processing in Mandarin may contribute to the potential typology of the interaction between tone and intonation in tonal languages. Simply comparing the results here with those Cantonese studies has demonstrated different mechanisms of tone and intonation interaction. In Mandarin, the interaction of tone and intonation leads to difficulties in intonation processing, whereas in Cantonese, it is tone processing rather than intonation processing that is problematic for native listeners (Kung et al., 2014). It seems that in tonal languages, when tone and intonation interact, whether tone or intonation causes processing difficulties can be language-dependent. In Mandarin, each tone is characterized by a distinctive F0 contour. The cue of tone is too robust to be affected by intonation. In Cantonese, tone identity can be easily distorted by intonation due to the crowded distribution of tones in a limited space. When tone and intonation compete for F0 cues, one function of F0 often has to give way to the other. In Mandarin, intonation gives way to tone. In Cantonese, tone gives way to intonation. Nevertheless, the pitch processing difficulties (tone or intonation) in both languages can be, in part, resolved via top-down information provided by a constraining semantic context.

5 Conclusion

The results reported here show that tone at the lexical level and intonation at the sentential level in Mandarin interact with each other, causing an asymmetrical difficulty of pitch processing at the sentential level. To disentangle intonational information from tonal information more efficiently, listeners not only tune to acoustic cues, but also rely on predictions based on the semantic context. A semantically constraining context considerably improves question intonation identification, but mainly so in sentences with the lexical falling tone in the final position.

Acknowledgements

We are very grateful to our editor and the anonymous reviewers for their comments on earlier versions of this paper.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the Guangdong Planning Office of Philosophy and Social Science for the Grant-GD19YYY06 to ML, the European Research Council for the ERC Starting Grant-206198 and the Netherlands Organization for Scientific Research for the Vici Grant-VI.C.181.040 to YC.

ORCID iDs

Min Liu https://orcid.org/0000-0002-5547-7581

Niels O. Schiller https://orcid.org/0000-0002-0392-7608

Notes

1. Portions of this work were presented at the Speech Prosody 2016 conference in Boston, USA, May 31–June 3, 2016.
2. Tone values in both Cantonese and Mandarin are transcribed using a 5-point scale notation system according to Chao (1968); each tone is described by the initial and the end point of the pitch level.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419-439. http://dx.doi.org/10.1006/jmla.1997.2558

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition: International Journal of Cognitive Science*, *73*(3), 247–264. http://dx.doi.org/10.1016/S0010-0277(99)00059-1

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://CRAN.R-project.org/pa//ckage=lme4

Bishop, J. (2012). Information structural expectations in the perception of prosodic prominence. In G. Elordieta & P. Prieto (Eds.), *Prosody and meaning* (pp. 239–270). De Gruyter Mouton.

Boersma, P., & Weenink, D. (2015). Praat: Doing phonetics by computer [Computer program]. Version 5.4.21. http://www.praat.org/

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2019). Toward a neurobiologically plausible model of language-related, negative event-related potentials. *Frontiers in Psychology*, *10*. https://www.frontiersin.org/article/10.3389/fpsyg.2019.00298

Braun, B., & Johnson, E. K. (2011). Question or tone 2? How language experience and linguistic function guide pitch processing. *Journal of Phonetics*, *39*(4), 585–594. https://doi.org/10.1016/j.wocn.2011.06.002

Buxó-Lugo, A., & Watson, D. G. (2016). Evidence for the influence of syntax on prosodic parsing. *Journal of Memory and Language*, *90*, 1–13. https://doi.org/10.1016/j.jml.2016.03.001

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS ONE*, *5*(6), e10729. 10.1371/journal.pone.0010729

Cao, J. (2004). Intonation structure of spoken Chinese: Universality and specificity. *Report of Phonetic Research*, 31–38.

Chao, Y. R. (1968). *A grammar of spoken Chinese*. University of California Press.

Chen, H.-C., Vaid, J., & Wu, J.-T. (2009). Homophone density and phonological frequency in Chinese word recognition. *Language and Cognitive Processes*, *24*(7–8), 967–982. https://doi.org/10.1080/01690960902804515

Cole, J., Mo, Y., & Hasegawa-Johnson, M. (2010). Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology*, *1*(2), 425–452. https://doi.org/10.1515/labphon.2010.022

Connell, B. A., Hogan, J. T., & Rozsypal, A. J. (1983). Experimental evidence of interaction between tone and intonation in Mandarin Chinese. *Journal of Phonetics*, *11*(4), 337–351.

Da, J. (2004). A corpus-based study of character and bigram frequencies in Chinese e-texts and its implications for Chinese language instruction. In P. Zhang, T. Xie, & J. Xu (Eds.), *Proceedings of 4th International Conference on New Technologies in Teaching and Learning Chinese* (pp. 501–511). The Tsinghua University Press.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117–1121. http://www.nature.com/neuro/journal/v8/n8/suppinfo/nn1504_S1.html

Dewaele, J.-M. (1999). Word order variation in interrogative structures of native and non-native French. *International Journal of Applied Linguistics*, *123*, 161–180.

Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica*, *30*, 412–431. https://doi.org/10.1016/0001-6918(69)90065-1

Dornisch, E. (1998). *Multiple-wh-questions in Polish: The interactions between wh-phrases and clitics*. (PhD Dissertation), Cornell University.

Duanmu, S. (2007). *The phonology of Standard Chinese*. Oxford University Press.

Durrell, M. (2011). *Hammer's German grammar and usage (fifth edition)*. Routledge.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491–505. https://doi.org/10.1111/j.1469-8986.2007.00531.x

Federmeier, K. D., & Kutas, M. (1999). Right words and left words: Electrophysiological evidence for hemispheric differences in meaning processing. *Cognitive Brain Research*, *8*(3), 373–392. https://doi.org/10.1016/S0926-6410(99)00036-1

Fok-Chan, Y.-Y. (1974). *A perceptual study of tones in Cantonese*. University of Hong Kong Press.

Frank, S. L. (2013). Uncertainty reduction as a measure of cognitive load in sentence comprehension. *Topics in Cognitive Science*, *5*(3), 475–494. https://doi.org10.1111/tops.12025

Gårding, E. (1987). Speech act and tonal pattern in Standard Chinese: Constancy and variation. *Phonetica: International Journal of Speech Science*, *44*(1), 13–29. https://doi.org/10.1159/000261776

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge University Press.

Gussenhoven, C., & Van de Ven, M. (2020). Categorical perception of lexical tone contrasts and gradient perception of the statement–question intonation contrast in Zhumadian Mandarin. *Language and Cognition*, 1–35. https://doi.org/10.1017/langcog.2020.14

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*(5669), 438–441. https://doi.org/10.1126/science.1095455

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. In *NAACL '01: Second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies 2001* (pp. 1–8). Association for Computational Linguistics.

Ho, A. T. (1977). Intonation variation in a Mandarin sentence for three expressions: Interrogative, exclamatory and declarative. *Phonetica: International Journal of Speech Science*, *34*(6), 446–457. https://doi.org/10.1159/000259916

Koutsoudas, A. (1968). On wh-words in English. *Journal of Linguistics*, *4*(2), 267–273. https://doi.org/10.1017/S0022226700001912

Kratochvil, P. (1998). Intonation in Beijing Chinese. In D. Hirst & A. D. Cristo (Eds.), *Intonation systems: A survey of twenty languages* (pp. 417–431). Cambridge University Press.

Kuang, J. (2017). Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice. *The Journal of the Acoustical Society of America*, *142*(3), 1693–1706. https://doi.org/10.1121/1.5003649

Kung, C., Chwilla, D. J., & Schriefers, H. (2014). The interaction of lexical tone, intonation and semantic context in on-line spoken word recognition: An ERP study on Cantonese Chinese. *Neuropsychologia*, *53*, 293–309. https://doi.org/10.1016/j.neuropsychologia.2013.11.020

Kuong, I.-K. J. (2008). Yes/no question particles revisited: The grammatical functions of mo4, me1, and maa3. In M. K. M. Chan, & H. Kang (Eds.), *Proceedings of the 20th North American Conference on Chinese Linguistics* (pp. 715–733). The Ohio State University.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the brain: Using our past to generate a future* (pp. 190–207). Oxford Scholarship Online.

Lee, O. J. (2005). *The prosody of questions in Beijing Mandarin*. The Ohio State University.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition: International Journal of Cognitive Science*, *106*(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Li, X., Ren, G., Zheng, Y., & Chen, Y. (2020). How does dialectal experience modulate anticipatory speech processing? *Journal of Memory and Language*, *115*, 104169. https://doi.org/10.1016/j.jml.2020.104169

Liang, J., & Van Heuven, V. J. (2007). Chinese tone and intonation perceived by L1 and L2 listeners. In C. Gussenhoven & T. Riad (Eds.), *Tones and tune: Experimental studies in word and sentence prosody* (Vol. 12-2, pp. 27–61). Mouton de Gruyter.

Liu, F., & Xu, Y. (2005). Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica: International Journal of Speech Science*, *62*(2–4), 70–87. https://doi.org/10.1159/000090090

Liu, M., Chen, Y., & Schiller, N. O. (2016). Online processing of tone and intonation in Mandarin: Evidence from ERPs. *Neuropsychologia*, *91*, 307–317. https://doi.org/10.1016/j.neuropsychologia.2016.08.025

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. Oxford University Press.

Ma, J. K.-Y., Ciocca, V., & Whitehill, T. L. (2006). Effect of intonation on Cantonese lexical tones. *The Journal of the Acoustical Society of America*, *120*(6), 3978–3987. https://doi.org/10.1121/1.2363927

Ma, J. K., Ciocca, V., & Whitehill, T. L. (2011). The perception of intonation questions and statements in Cantonese. *The Journal of the Acoustical Society of America*, *129*(2), 1012–1023. https://doi.org/10.1121/1.3531840

Morton, E. S. (1994). Sound symbolism and its role in non-human vertebrate communication. In L. Hinton, J. Nichols, & J. J. Ohala (Eds.), *Sound symbolism* (pp. 348–365). Cambridge University Press.

Nieuwland, M. S., Barr, D. J., Bartolozzi, F., Busch-Moreno, S., Darley, E., Donaldson, D. I., . . .Von Grebmer Zu Wolfsthurn, S. (2020). Dissociable effects of prediction and integration during language comprehension: Evidence from a large-scale study using brain potentials. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1791), 20180522. https://doi.org/10.1098/rstb.2018.0522

Nolan, F. (2006). Intonation. In B. Aarts & A. McMahon (Eds.), *The handbook of English linguistics* (pp. 433–457). Blackwell Publishing.

Ohala, J. J. (1983). Cross-language use of pitch: An ethological view. *Phonetica: International Journal of Speech Science*, *40*(1), 1–18. https://doi.org/10.1159/000261678

Patro, C., & Mendel, L. L. (2016). Role of contextual cues on the perception of spectrally reduced interrupted speech. *The Journal of the Acoustical Society of America*, *140*(2), 1336–1345. https://doi.org/10.1121/1.4961450

Peng, S.-h., Chan, M. K. M., Tseng, C.-y., Huang, T., Lee, O. J., & Beckman, M. E. (2005). Towards a pan-Mandarin system for prosodic transcription. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 230–270). Oxford University Press.

Quirk, R., Greenbaum, S., & Leech, G. (1972). *A grammar of contemporary English*. Longman.

R Core Team (2020). R: A language and environment for statistical computing. In *R Foundation for Statistical Computing*. Vienna, Austria. http://www.R-project.org/

Rojina, N. (2004). The acquisition of wh-questions in Russian. *Nordlyd: Tromsø University Working Papers on Language & Linguistics*, 32(1), 68–87. https://doi.org/10.7557/12.59

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(4), 623–656. https://doi.org/10.1002/j.1538-7305.1948.tb00917.x

Sheldon, S., Pichora-Fuller, M. K., & Schneider, B. A. (2008). Priming and sentence context support listening to noise-vocoded speech by younger and older adults. *The Journal of the Acoustical Society of America*, *123*(1), 489–499. https://doi.org/10.1121/1.2783762

Shen, J. (1985). Beijinghua shengdiao de yinyu he yudiao [The range of tones and intonation in Mandarin]. In T. Lin & L. Wang (Eds.), *Experimental analyses on Beijing Mandarin* (pp. 73–125). Peking University Press [in Chinese].

Shen, X. S. (1989). *The prosody of Mandarin Chinese*. University of California Press.

Shih, C. (2000). A declination model of Mandarin Chinese. In A. Botinis (Ed.), *Intonation: Analysis, modelling and technology* (pp. 243–268). Kluwer Academic Publishers.

Simpson, G. B. (1994). Context and the processing of ambiguous words. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (Vol. 22, pp. 359–374). Academic Press.

Tanenhaus, M. K., Kurumada, C., & Brown, M. (2015). Prosody and intention recognition. In L. Frazier & E. Gibson (Eds.), *Explicit and implicit prosody in sentence processing: Studies in honor of Janet Dean Fodor* (pp. 99–118). Springer International Publishing.

Torreira, F., Roberts, S. G., & Hammarström, H. (2014). Functional trade-off between lexical tone and intonation: Typological evidence from polar-question marking. In C. Gussenhoven, Y. Chen, & D. Dediu (Eds.), *Proceedings of the 4th International Symposium on Tonal Aspects of Language* (pp. 100–103). International Speech Communication Association.

Tsuchihashi, M. (1983). The speech act continuum: An investigation of Japanese sentence final particles. *Journal of Pragmatics*, *7*(4), 361–387. https://doi.org/10.1016/0378-2166(83)90024-3

Ultan, R. (1978). Interrogative systems. In J. H. Greenberg (Ed.), *Universals of human language* (Vol. 4, pp. 211–248). Stanford University Press.

Vaissière, J. (2008). Perception of intonation. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 236–263). Blackwell Publishing Ltd.

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 443–467. https://doi.org/10.1037/0278-7393.31.3.443

Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(2), 394–417. https://doi.org/10.1037/0278-7393.25.2.394

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176–190. https://doi.org/10.1016/j.ijpsycho.2011.09.015

Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica: International Journal of Speech Science*, *49*, 25–47.

Wicha, N. Y. Y., Bates, E. A., Moreno, E. M., & Kutas, M. (2003). Potato not Pope: Human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters*, *346*(3), 165–168.

Wlotko, E. W., & Federmeier, K. D. (2015). Time for prediction? The effect of presentation rate on predictive sentence comprehension during word-by-word reading. *Cortex*, *68*, 20–32. https://doi.org/10.1016/j.cortex.2015.03.014

Wu, Z. (1982). Putonghua yuju zhong de shengdiao bianhua [Tonal changes in Mandarin discourses]. *ZHONGGUO YUWEN*, *171*(6), 439–449 [in Chinese].

Wu, Z. (1996). A new method of intonation analysis for Standard Chinese: Frequency transposition processing of phrasal contours in a sentence. In G. Fant et al. (Ed.), *Analysis, perception and processing of spoken language* (pp. 255–268). Elsevier Science B.V.

Xu, B. R., & Mok, P. (2012a). Cross-linguistic perception of intonation by Mandarin and Cantonese listeners. In Q. Ma , H . Ding, & D. Hirst (Eds.), *Proceedings of the 6th International Conference on Speech Prosody 2012* (pp. 99–102). Tongji University Press.

Xu, B. R., & Mok, P. (2012b). Intonation perception of low-pass filtered speech in Mandarin and Cantonese. *Paper presented at the Third International Symposium on Tonal Aspect of Languages*, Nanjing, China, 26-29 May, 2012.

Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech Communication*, *46*(3–4), 220–251. https://doi.org/10.1016/j.specom.2005.02.014

Xu, Y. (2009). Timing and coordination in tone and intonation—An articulatory-functional perspective. *Lingua: International Review of General Linguistics*, *119*(6), 906–927. https://doi.org/10.1016/j.lingua.2007.09.015

Xu, Y., & Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication*, *33*(4), 319–337. https://doi.org/10.1016/s0167-6393(00)00063-7

Ye, Y., & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes*, *14*(5–6), 609–630. https://doi.org/10.1080/016909699386202

Yu, K. M., & Lam, H. W. (2014). The role of creaky voice in Cantonese tonal perception. *The Journal of the Acoustical Society of America*, *136*(3), 1320–1333. https://doi.org/10.1121/1.4887462

Yuan, J. (2006). Mechanisms of question intonation in Mandarin. In Q. Huo, B. Ma, E.-S. Chng, & H. Li (Eds.), *Chinese Spoken Language Processing, ISCSLP 2006* (pp. 19–30). Springer, Berlin, Heidelberg.

Yuan, J. (2011). Perception of intonation in Mandarin Chinese. *The Journal of the Acoustical Society of America*, *130*(6), 4063–4069. https://doi.org/10.1121/1.3651818

Ziegler, J. C., Tan, L. H., Perry, C., & Montant, M. (2000). Phonology matters: The phonological frequency effect in written Chinese. *Psychological Science*, *11*(3), 234–248.