Floating Boundary Tone: Production and Perception of Syntactically Unmarked Polar Question in Tianjin Mandarin

Cong Zhang[1,2], Aditi Lahiri[1]

[1] **University of Oxford, Oxford, United Kingdom**

[2] **Radboud University, Nijmegen, Netherlands**

Email:

cong.zhang@ru.nl (corresponding author)

aditi.lahiri@ling-phil.ac.uk

Funding

Competing interests

The authors have no competing interests to declare.

Floating Boundary Tone: Production and Perception of Syntactically Unmarked Polar Question in Tianjin Mandarin

Abstract

The present study investigates the intonational tune of syntactically unmarked polar question in Tianjin Mandarin. The central query focused on the relevance of high boundary tones in a language which already has lexical tonal contrasts. A production study was conducted to examine the phonological features of the syntactically unmarked polar question (a.k.a intonational YES/NO question) tune and comparing it with the statement tune. The results show a significant register lift H_R and a high floating boundary tone $H̥_I$. A tune identification task then further verifies whether the two prosodic features concluded from the production are used in perception. The results showed that both the register difference and the boundary tone made a difference in native speakers' perception in discriminating questions from statements.

Keywords: Tianjin Mandarin; tone; intonation; floating boundary tone; polar question

1.   Introduction

Alternating pitch accents and boundary tones are common methods of asking polar questions in non-tonal languages. Furthermore, intuitively one expects a rise towards the end of an intonational boundary. However, in languages with lexical tones, intonational tune change is much restricted by the pitch contours and registers of lexical tones, since both lexical tones and intonational tunes are intrinsically linked with pitch modulation. This study examines the tones and tunes in Tianjin Mandarin, a tonal variety of northern Chinese, to investigate the interaction between lexical tones and intonational tunes.

Tianjin Mandarin, as well as many other varieties of Mandarin, has two ways of asking a polar question (a.k.a. yes/no question, hereinafter "*YNQ*"): One type is syntactically marked – with a question particle "-ma" added to the end of an utterance (hereinafter "*ma-Q*"); the other has no syntactic markings – the question is asked through modulating the intonation of the utterance (hereinafter "*IntQ*"). Since Tianjin Mandarin is a tonal language, and IntQs only differ from the statements in terms of the intonation, how do the lexical tones interact with the IntQ tune?

Tianjin Mandarin has four underlying lexical tones, which are shown in Figure 1: Tone 1 is low and slightly falling; Tone 2 is high and slightly rising; T3 is low and rising, and Tone 4 is high and falling. Note that the pairs of tones are evenly distributed in terms of registers and tonal contours; that is, there are two pairs of rising (LH) and falling (HL) tones, the pairs differing in their overall register. While it is not unusual for lexical tones to have different registers in other Sino-Tibetan tonal languages, or even in southern Chinese such as Shanghainese or Cantonese, it has not been reported that register is a factor in Northern Mandarin varieties. However, the results and analysis from previous studies (Li et al., 2017; F. Shi, 1987; J. Zhang & Liu, 2011) all have presented four symmetrically distributed lexical tones in Tianjin Mandarin. To represent the differences in register phonologically, we will use the following notation where register is indicated by a superscript ^{TR}. "T" suggests the **T**ones (HL or LH) and "R" represents the **r**egister. The four tones are:

$$T1 = {}^{LR}(HL)$$
$$T2 = {}^{HR}(LH)$$
$$T3 = {}^{LR}(LH)$$
$$T4 = {}^{HR}(HL)$$

Since the lexical tones in Tianjin Mandarin are well-balanced in terms of the number and location of Hs and Ls, Tianjin Mandarin is ideal for the investigation of intonation in tonal languages.
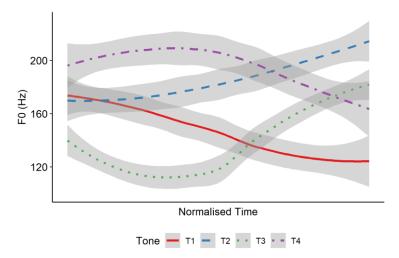
Figure 1. *Four Tianjin citation tones. The lines report the mean f0 values (Hz), with 95% interval shown by the grey area. The red solid line represents the low-register falling tone, Tone 1: ^{LR}(HL); the blue dashed line represents the high-register rising tone, Tone 2: ^{HR}(LH); the dotted green line represents low-register rising tone, Tone 3: ^{LR}(LH); and the dot-dash line represents the high-register falling tone, Tone 4: ^{HR}(HL).*

We investigated the features of Tianjin Mandarin IntQ tune from the perspectives of both production and perception. We first conducted a production study to compare between the statement tune and the IntQ tune, by examining the phonetic details to account for the phonological features. Then, perception data were collected to investigate how well listeners can identify the statement tune and IntQ tune. The perception study made use of the conclusions drawn from the production study to investigate of the crucial cues for the perception of the tunes.

The production of IntQ tune in Tianjin Mandarin is analysed under the Autosegmental-Metrical (hereinafter "AM", Ladd, 1996, 2008; Pierrehumbert, 1980) framework by way of examining the potential boundary tones and pitch accents phonologically. For clarity of representation, we follow Hayes & Lahiri (1991) in using H_I and L_I to represent high and low boundary tones at the intonational phrase level, which are effectively the same as the notation of H% and L% in AM terminology. In this study, we did not only aim to describe the IntQ tune in Tianijin Mandarin, but, more importantly, aim to discuss how various intonational features, such as boundary tone, were used in a tonal language.

1.1 Polar question tune

YNQ tune is frequently associated with a high or rising end, by linguists and by laymen. This association has been proposed as a linguistic universal based on different reasons and from different perspectives (Gussenhoven, 2002; Ohala, 1983, 1984). However, whether such a "universal" exists at all and to what extent it is true remains a question. Bolinger (1978) reported a few languages that did not use high pitch in questions, although he made one of the earliest universalist claims. Rialland (2007) further surveyed YNQ tunes in 78 African languages, and found 11 different YNQ markers, including five high-pitched markers and six non-high pitched ones. A substantial number of languages did not associate high pitch with question prosody. One of the markers was a final HL boundary tone, also described as rise-fall. It is not only found

in African languages, but also observed in Bengali [L*+H$_I$L$_I$] (Hayes & Lahiri, 1991). Other types of falling or low boundary tone have also been discovered to mark IntQ in languages such as Greek [L* L+H- L%] (Arvaniti et al., 2006), Puerto Rican Spanish [!H*+L% / H+L*+L%] (Armstrong, 2012), and Bari Italian [L+H*/ H+L* + L-L%] (Grice & Savino, 2003).

Hyman and Monaka (2008) summarised major tone-intonation interaction in tonal languages into three categories: accommodation, submission, and avoidance. "Accommodation", i.e. "peaceful coexistence", refers to where boundary tones are attached. "Submission" refers to when lexical tones "surrender" to intonation. The third type is "avoidance", where intonation avoids lexical tones and exploits changes such as pitch range and pitch interval or makes use of additional particles. Such non-boundary related phenomena are used in a number of languages. For example, in Wolof (non-tonal, Rialland & Robert, 2001), Hausa (tonal, Lindau, 1986), and Danish (non-tonal, Grønnum, 1998), reduction of downdrift can be used for making an IntQ. Among contour-tone languages such as Chinese, Thai, and Vietnamese, register change and the extent of rise or fall of the lexical tones are commonly seen. In Vietnamese, "greater high tone movement" is found in IntQ tune (Dung et al., 1998, p. 405); in Thai, "narrower pitch range" is observed in IntQ tune (Luksaneeyanawin, 1998, p. 385). Mandarin has similar results as these contour-tone languages. In the following section, the tunes in Mandarin Chinese will be reviewed in greater detail. While pitch is a main indicator of intonation, it is not the only cue. Cahill (2013) examines polar questions in six Ghanaian languages and suggests that Adele and Safaliba make use of final vowel lengthening, and Chumburng employs breathy termination. It is therefore important to examine the temporal information and other potential important cues in our study.

1.2   Polar question tunes in Mandarin varieties

Although there has not been earlier research on intonation in Tianjin Mandarin, previous studies of other varieties of Mandarin offer a general phonetic account of intonation in tonal languages. However, due to the wide variety of research targets, the results were not directly comparable with each other: results were reported in terms of slopes or grids (Gårding, 1984, 1987; Zeng et al., 2004), top/bottom lines (J. Shen, 1985), sentence-level contours (X. S. Shen, 1990), phonetic details (Liu & Xu, 2005), and local phenomena (Lee, 2005; H. Lin, 2006; M. Lin, 2004; X. S. Shen, 1990). These different results and approaches have made it difficult to make a good generalisation about what exactly a Mandarin polar question tune should be like.

Some studies of Standard Mandarin (a.k.a *Putonghua*) argue that the intonation of IntQ has a rising slope, whereas that of ma-Q does not necessarily rise. X. S. Shen (1990) found that both types of polar questions had higher pitches at the beginning than their statement counterparts: the beginning and the end were both raised for IntQs (Tune 2); while for ma-Qs (Tune 3), the end was almost only as high as the beginning of the corresponding statement tune. Ni & Kawai (2004) experimented with the same set of materials and reached the same conclusion as X. S. Shen (1990). Gårding (1984, 1987) analysed intonation using "grids" and "pivots". She generalised the YNQ tune in Standard Mandarin as a rising "grid". However, Shih (1988) contended that one uniformed grid failed to explain all the variations brought by the different lexical tones in Mandarin. Liu and Xu (2005) elicited different types YNQs with sequences of

syllables of the same lexical tones in Standard Mandarin and concluded that there was exponential or double-exponential raising for the questions and argued that accelerated final rise was merely a part of the global pitch raise.

While the idea of rising intonation for IntQ prevails, some studies did not fully support this claim. In Shi's (1980) early impressionistic description of Mandarin statement and IntQ tunes, IntQ was not described as rising; instead, IntQ was depicted as having both a higher beginning and a higher ending than statements. Shen (1985) observed that, in Beijing Mandarin, the Mandarin variety that Standard Mandarin is based on, the IntQ's bottom line got raised, while the topline did not – this was different from IntQs in such non-tonal languages as English or Russian. Zeng, Martin, and Boulakia (2004) showed that IntQs had a rising slope in Standard Mandarin; however, it was not significant for the two rising tones, Tone 2 (LH) and Tone 3 (LLH), which indicated that the IntQ tune did not rise much at all. This reflects that Standard Mandarin, with three lexical tones ending in H and only one in L, sometimes is more representative of the lexical tones which ends in H. These studies presented different pictures for whether the overall f0 was raised or the pitch span was expanded, which motivated the examination of both mean f0 and f0 extrema in the current study.

Previous research also reported different local phenomena for the intonation of Mandarin polar questions. Lin (2006, 2004) examined IntQs in a telephone conversation corpus and found that the registers of the starting and ending points of the last non-neutral lexical tones were higher. He thus believed that the boundary tone in Standard Mandarin is the most crucial component for IntQ tune. Lee (2005) elicited conversation data and concluded that the boundary tone at the last NP of IntQs got raised in Beijing Mandarin. X. S. Shen (1990) proposed that the first few syllables of YNQs – both IntQ and ma-Q – had higher boundary tones than their corresponding statement intonations.

2.   Production study

2.1   Methods

2.1.1   Materials

As shown in Table 1, twelve monosyllabic words in isolation, which consisted of three syllables ([ma], [mi], [maʊ]) with four different tones [T1: ^{LR}(HL); T2: ^{HR}(LH); T3: ^{LR}(LH); T4: ^{HR}(HL)], were recorded as both statements and IntQs. The statement tune for isolated monosyllables was the same as the citation lexical tones. All monosyllabic words in this study could be used naturally as a question. We restricted ourselves to sonorant monosyllabic words in this study, since we have no prior knowledge about intonation in Tianjin Mandarin, and the only way to avoid any influence from other lexical-level prosodic phenomena such as tone sandhi was to use monosyllabic words.

Table 1. *Monosyllabic words*

SYLLABLES	LR(HL) (TONE 1)	HR(LH) (TONE 2)	LR(LH) (TONE 3)	HR(HL) (TONE 4)
[ma]	妈 'mother'	麻 'hemp'	马 'horse'	骂 'scolding'
[maʊ]	猫 'cat'	毛 'fur'	铆 'rivet'	帽 'hat'
[mi]	咪 'meow'	迷 'riddle'	米 'rice'	蜜 'honey'

2.1.2 Participants and recordings

Six native speakers of Tianjin Mandarin (three male: M101-M103; three female: F101-F103) were recorded. All speakers were born and raised in the city area of Tianjin, and spoke Tianjin Mandarin on a daily basis. The age range of the informants was between 22 and 50. All were capable of speaking Standard Mandarin in formal situations, though with different degrees of Tianjin accents. None of them reported any speech disorders. All speakers received financial compensation for their time.

Recordings were made in a quiet room at Nankai University in Tianjin, China. We recorded mono-channel .wav files using a Rode NT-USB Microphone with Audacity (AudacityTeam, 2016) onto a PC, at a sampling rate of 44.1 kHz, 16 bit. The speakers were given time to briefly familiarise themselves with the materials before the experiment started. The test materials were presented as Chinese characters placed in the middle of the screen in a Microsoft PowerPoint presentation. To avoid any potential list effect, each utterance was placed on a separate slide. Each token had three discontinuous occurrences. The characters were presented with a full stop or a question mark. The informants were asked to produce a statement as naturally as they could when seeing a Chinese full stop "。" and an IntQ when seeing a Chinese question mark "？". For the IntQs, the speakers were asked to produce the utterance such that they would expect an answer "yes" or "no". The experimenter controlled the pace of the experiment, leaving a short pause after each utterance.

2.1.3 Data annotation and extraction

All data were manually segmented into onsets and rhymes in Praat (Boersma & Weenink, 2021), since the tone-bearing unit in Mandarin is commonly recognised as the rhyme (Yip, 1989). F0 values were extracted with a range of 70 Hz to 500 Hz for female speakers, and 60 Hz to 350 Hz for male speakers in Praat. Two least creaky instances out of three repetitions were selected. Obstruents and the creakiness of the LH tone often cause errors and undetectable pitch values, which is frequently an issue in both Tianjin Mandarin and Standard Mandarin prosody research (Chao, 1968; Davison, 1991). Together with the artefacts detected during visual inspection, all such data were marked as missing data points during the analyses. The values of f0 range were also inspected to make sure the maxima and minima were not tracking artefacts and were representative of the f0 range.

All F0 data were converted to ERB (Equivalent Rectangular Bandwidth) units for statistical analysis. ERB is a scale that was derived from human auditory selectivity and

is more representative of human perception than Hertz (Nolan, 2003). It also makes the data from male speakers and female speakers more comparable (Daly & Warren, 2001). The formula used here is as follows (Hermans & van Gestel, 1991):

$$\text{ERB} = 16.7 \log_{10}(1 + \frac{f}{165.4})$$

2.1.4 Data analysis

The data were analysed in two steps. In the first step, we examined mean values in the traditional way to explore and generalise the properties of the two tunes. Then a functional Principal Component Analysis (hereinafter "*fPCA*"), a dimension reduction algorithm, was used to further examine the major components abstracted from the two tunes. More details about fPCA will be introduced in Section 2.1.4.2.

2.1.4.1 Analysis of mean values

The mean values of f0, f0 range, maximum f0, minimum f0, and duration were analysed. F0 scaling was investigated by calculating the mean f0 values to examine the overall register change; maximum and minimum f0 values were further analysed to understand whether the register was raised or expanded. The presence of boundary tone was concluded preliminarily from visual inspection, and then verified by analysing the f0 range change. The type of boundary tone can also be inferred from the f0 range data. Pirate plots (generated using "yaRrr" package in R, Phillips, 2017) were used to visualise the data by lexical tone.

Linear mixed effect models were built using the *lmer4* package (Bates et al., 2015) in R (R Core Team, 2021) to predict the dependent variables (mean f0, f0 range, maximum f0, minimum f0, and duration). The full model was built with the following fixed effects: TUNE (Statement, Question), TONE_SHAPE (HL, LH), TONE REGISTER (L, H), the interaction between TONE_SHAPE and TUNE, and the interaction between TONE_REGISTER and TUNE. The random intercepts included SPEAKER (6 speakers) and ITEM (24 unique tokens), and TUNE was included as by- SUBJECT and by- ITEM random slopes. The random slopes were removed when the models did not converge or reported singularity. Note that TONE_SHAPE and TONE REGISTER were used instead of the lexical tones to examine the influence of specific features of the lexical tones. However, these two factors combined equated the four lexical tones. The p-values of the models were calculated using the Satterthwaite approximation in the *lmerTest* package. The same model fitting procedures were applied to the onsets and rhymes separately, since rhymes are the tone-bearing segments in Tianjin Mandarin. The final models were reported together with the relevant results.

2.1.4.2 Functional Principal Component analysis

FPCA has been an increasingly popular method in linguistics research. It has been successfully applied in a range of prosodic studies (Asano & Gubian, 2018; Chen & Boves, 2018; Cheng et al., 2013; Lohfink et al., 2019). In this study, we followed the processing and analysis procedures in Gubian et al. (2015). F0 contours were first smoothed with b-splines. This useful procedure erased outliers caused by micro-prosodic effect which were irrelevant to the phonology of intonation. The smoothed contours were then submitted to the fPCA and the Principal Component (PC) scores

were generated. In general, fPCA reduces the dimensionality of a dataset while preserving as much of the data's variation as possible. The PCs represents the major features of the dataset. In this study, the PCs could inform us of the distinctive difference between the two tunes. Moreover, fPCA allows landmark registration, which is a process to align the meaningful points in a curve. In this study, the rhyme onset was used as a landmark so the non-tone-bearing onset and tone-bearing rhyme could be clearly separated in an f0 curve.

FPCA is a method that shares the same spirit as the AM framework theory, i.e., only retaining the distinctive events in the intonation contours. However, after all fPCA is only a mathematical algorithm and cannot distinguish phonologically meaningful events from phonetically important trends. It is ultimately still dependent on the researchers' analysis and interpretation to decide what the PC scores capture. It is therefore beneficial to combine the traditional mean value analysis with the fPCA analysis to cross-validate the results.

Linear mixed effect models were later fitted for the PC scores to check whether the PC scores of the two tunes were significantly different. The whole dataset was separated by tones to focus on the tune distinction. The fixed effects and random effects were the same as the models reported in the last section.

2.2   Results and discussion

Figure 2 show instances from Speaker F103 (female) for all four tones [T1: ^{LR}(HL); T2: ^{HR}(LH); T3: ^{LR}(LH); T4: ^{HR}(HL)] in both statement and IntQ tunes. Observable differences are minimal between the figures on the left (statement tune) and the ones on the right (IntQ tune). Differences in register and in the height of the right boundary are present in some instances in Figure 2. Since the instances were the least creaky instances in the data and were not produced as minimal pairs, it was difficult to generalise by visual inspection. We resort to the more detailed analyses for the generalisation of the tune features.
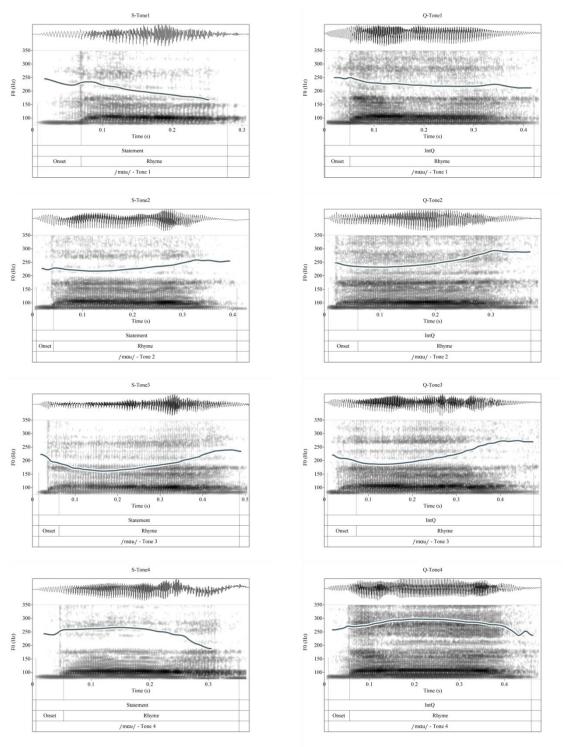
Figure 2. *Representational f0 contours for each tone with statement tune (left) and IntQ tune (right) by Speaker F103.*

2.2.1   Results of mean value analysis

2.2.1.1 Mean f0

For onset, the significant predictors were TUNE ($\chi^2(1)$ = 39.44, p < 0.001 ***), TONE_REGISTER ($\chi^2(1)$ = 187.9, p < 0.001 ***) and TONE_SHAPE ($\chi^2(1)$ = 548.46, p < 0.001 ***). The mean F0 values of the IntQ tune (shown as "Q" in all figures) were consistently higher in the IntQ tune than their statement counterpart in all tones. The interaction between TONE_SHAPE and TUNE indicated that the difference between the two tunes for the rising tones (T2: ^{HR}(LH), T3: ^{LR}(LH)) differed from the falling tones (T1: ^{LR}(HL), T4: ^{HR}(HL)). Figure 3A clearly displays the interaction: the rising tones have smaller differences than the falling tones, which suggests that the IntQ did not rise as much in the rising tones compared with the falling tones at the beginning part of the utterances.

The significant predictors for rhyme data were TUNE ($\chi^2(1)$ = 42.38, p < 0.001 ***), TONE_REGISTER ($\chi^2(1)$ = 749.61, p < 0.001 ***), and TONE_SHAPE ($\chi^2(1)$ = 11.03, p < 0.001 ***). In Figure 3B, the mean f0 values of the IntQ rhymes were all significantly higher than the values from statements, and the differences were constant across different tones. In contrast to the onset result, the lack of significance in the interaction indicated that the tone-bearing unit were modified differently.
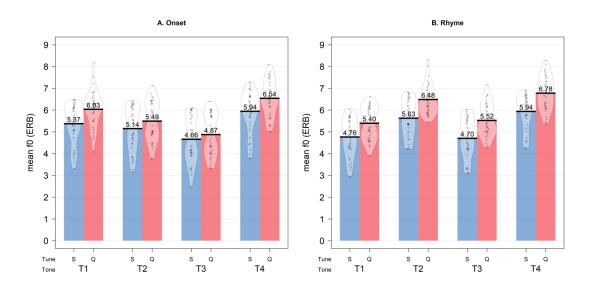

Figure 3: *Pirate plots for mean f0 values (A: onset; B: rhyme) by tune and grouped by lexical tone. The bars depict the mean f0 values (unit: ERB) of the two tunes (Blue: Statement; Red: IntQ). The jitter points and the violin shapes show the data distribution.*

2.2.1.2 Maximum and minimum f0

Consistent with the mean f0 results, both maximum f0 values and minimum f0 values in onset and in rhyme were significantly predicted by TUNE, TONE_REGISTER, and TONE_SHAPE. Detailed values are presented in Table 2. Figure 4

The range of maximum f0 for the statements was 5.50 ERB to 6.42 ERB, while that for IntQs was 6.09 ERB to 7.53 ERB; minimum f0 values ranged from 3.51 ERB to 5.10 ERB for the statements, while the range for IntQs was between 3.90ERB and 5.94 ERB. The maximum f0 of IntQs was higher than the statements by 0.99 ERB on average; while the minimum was higher by 0.75 ERB on average. These results suggested that the IntQ tune was higher than the statement tune in terms of both the f0 maximum and the f0 minimum.

Table 2. Results of the fixed effects for maximum f0 and minimum f0

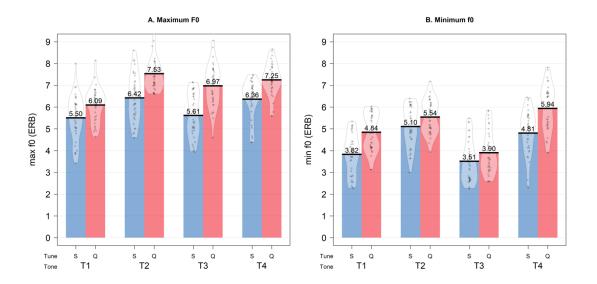| effect	Onset		Rhyme	
	Maxf0	**Minf0**	**Maxf0**	**Minf0**
TONE_REGISTER	$\chi^2(1) = 223.04$ $p < 0.001$ ***	$\chi^2(1) = 121.17$ $p < 0.001$ ***	$\chi^2(1) = 239.23$ $p < 0.001$ ***	$\chi^2(1) = 273.49$ $p < 0.001$ ***
TONE_SHAPE	$\chi^2(1) = 583.11$ $p < 0.001$ ***	$\chi^2(1) = 356.7$ $p < 0.001$ ***	$\chi^2(1) = 37.05$ $p < 0.001$ ***	$\chi^2(1) = 17.82$ $p < 0.001$ ***
TUNE	$\chi^2(1) = 25.06$ $p < 0.001$ ***	$\chi^2(1) = 63.36$ $p < 0.001$ ***	$\chi^2(1) = 46.22$ $p < 0.001$ ***	$\chi^2(1) = 85.47$ $p < 0.001$ ***
TONE_REGISTER: TUNE	$\chi^2(1) = 0.6$ $p = 0.44$	$\chi^2(1) = 0.58$ $p = 0.45$	$\chi^2(1) = 0.05$ $p = 0.82.$	$\chi^2(1) = 0.25$ $p = 0.62$
TONE_SHAPE: TUNE	$\chi^2(1) = 21.41$ $p < 0.001$ ***	$\chi^2(1) = 12.98$ $p < 0.001$ ***	$\chi^2(1) = 20.66$ $p < 0.001$ ***	$\chi^2(1) = 17.04$ $p < 0.001$ ***

Figure 4. *Pirate plots for mean maximum (A) and mean minimum (B) f0 values in rhymes by tune and grouped by lexical tone. The bars depict the f0 values (unit: ERB) of the two tunes (Blue: Statement; Red: IntQ). The jitter points and the violin shapes show the data distribution.*

2.2.1.3 F0 range

Pitch excursion was investigated through examining f0 range. Different from the mean f0 and f0 extrema data previously presented, f0 exhibited very different patterns between onset and rhymes. None of the fixed effects was significant in predicting the f0 range for the two tunes. As shown in Figure 5A, the differences between the tunes or between the tones were marginal. However, for rhyme, the significant fixed effects were: TUNE ($\chi^2(1) = 10.2$, p = 0.02 *), TONE_REGISTER ($\chi^2(1) = 5.06$, p = 0.03 *) and TONE_SHAPE ($\chi^2(1) = 68.67$, p < 0.001 ***). The interaction between TONE_SHAPE and TUNE was also significant ($\chi^2(1) = 47.66$, p < 0.001 ***). These results indicated that the two tunes have significantly different f0 range; and the f0 range changed systematically according to the registers and shapes of the lexical tones. The significant interaction between tone shape and tune again suggested that the rising tones (T2: ^{HR}(LH), T3: ^{LR}(LH)) and the falling tones (T1: ^{LR}(HL), T4: ^{HR}(HL)) exhibited different patterns. Figure 5B displays the differences: for the two falling tones, the f0 range of the IntQ tune is smaller than that of the statement tune; while for the two rising tones, the f0 range of the IntQs is larger than that of the statements. These results suggest that the falling tones did not fall as much in the IntQs as they did in their declarative counterparts; instead, the falls were ended higher. On the contrary, the rising tones rose more in the IntQ tune than in the statement tune.
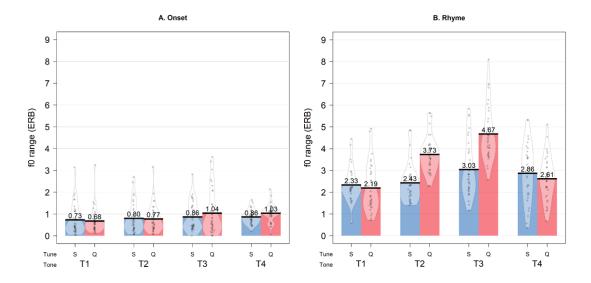

Figure 5. *Pirate plots for mean f0 range values (A: onset; B: rhyme) by tune and grouped by lexical tone. The bars depict the mean f0 range values (unit: ERB) of the two tunes (Blue: Statement; Red: IntQ). The jitter points and the violin shapes show the data distribution.*

2.2.1.4 Duration

The only significant fixed effect for predicting onset duration was TUNE ($\chi^2(1) = 8.19$, $p < 0.001$ ***). The onsets in IntQ tune (mean = 59.8 ms, sd = 20.4 ms) was significantly shorter than the statement tune counterparts (mean = 65.7 ms, sd = 24.5 ms). As shown in Figure 6, the differences were marginal. However, in the rhyme data, all fixed effects and interactions were significant: TUNE ($\chi^2(1) = 9.33$, p = 0.02 *); TONE_REGISTER ($\chi^2(1) = 68.99$, p < 0.001 ***); TONE_SHAPE ($\chi^2(1) = 203.84$, p < 0.001 ***); TUNE:TONE_REGISTER ($\chi^2(1) = 5.44$, p = 0.02 *); TUNE:TONE_SHAPE ($\chi^2(1) = 30.79$, p < 0.001 ***). The duration of rhymes, however, were longer in IntQs (mean = 324.1 ms, sd = 71.4 ms) than in statements (mean =290.4 ms, sd = 74.1 ms). The interactions were plotted in as effect plots (generated with "effects" package in R, Fox & Weisberg, 2018). The TUNE:TONE_REGISTER interaction suggested that the durational difference between the two tunes differed in the L-register tones (T1, T3) from the H-register tones (T2, T4). The L-register tones had longer duration in statements than the H-register tones; and the increase in the IntQs of the L-register tones were also larger than in the H-register tones. The interaction of TUNE:TONE_SHAPE demonstrated the difference between the rising tones (T2, T3) and the falling tones (T1, T4), with the falling tones having much bigger increase from statement to IntQ. These trends are also supported by the values in Figure 6.
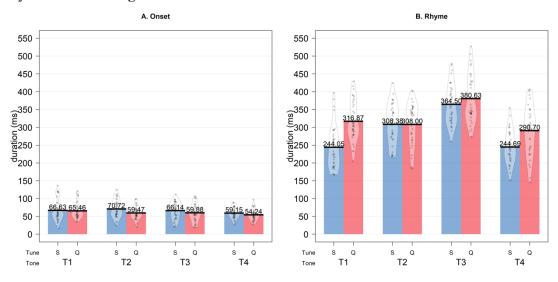

Figure 6. *Pirate plots for mean duration values (A: onset; B: rhyme) by tune and grouped by lexical tone. The bars depict the mean duration values (unit: ERB) of the two tunes (Blue: Statement; Red: IntQ). The jitter points and the violin shapes show the data distribution.*
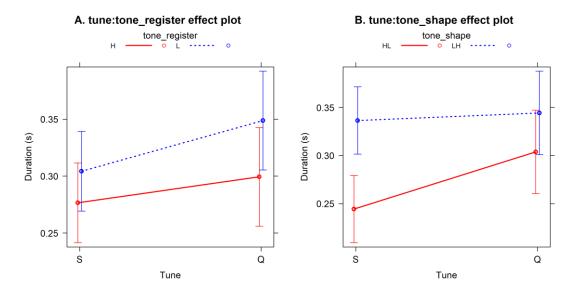
Figure 7. *Effects plots of the interaction in the rhyme duration model. A shows the interaction between* TUNE *and* TONE_REGISTER; *B shows the effect of the interaction between* TUNE *and* TONE_SHAPE.

2.2.2 Results of fPCA analysis

Figure 8 displays average curves for the two tunes in each lexical tone across the entire dataset. The statement contours are consistent with the citation tones presented in Figure 1, and the IntQ contours demonstrate the findings of what were previously reported in the mean value analysis section, including higher register of the IntQ tune, and higher endings for all four tones with IntQ tune. In the next few subsections, the results of the PC scores will be reported by each lexical tone. For ease of comparison and presentation, the order of the tones reported is T1, T4 (two falling tones); and then T2, and T3 (two rising tones).
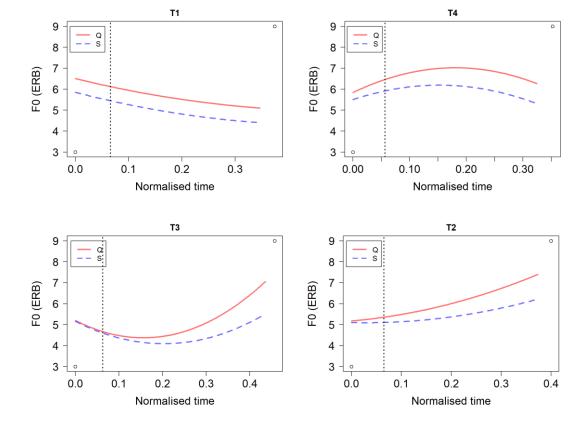
Figure 8. *Average curves for statement (dashed line) and IntQ (solid line) tunes grouped by lexical tones. The four lexical tones are T1: ^{LR}(HL); T2: ^{HR}(LH); T3: ^{LR}(LH); T4: ^{HR}(HL). The tones were arranged by tone shape vertically (top: HL; bottom: LH) and by tone register horizontally (left: L; right: H). The vertical dotted line represents the landmark registration site, i.e. onset of rhyme in the current study.*

2.2.2.1 Tone 1: ^{LR}(HL)

Since this dataset contains speech materials that were short and tightly controlled, two PCs were sufficient to capture 100% of the variance (cf. Asano & Gubian, 2018, in which three PCs were reported). For T1, PC1 explained 89.8% of the variance and PC2 accounted for 10.2% of the variance. Figure 9A displays the PC1 and PC2 curves separately.

In each PC figure, the dark black curve shows the average curve of all the data in the current dataset. For instance, in Figure 9A, the black solid line shows the average of the data including both statement data and IntQ data for Tone 1. The x-axis is the normalised time after landmark registration; and the y-axis is the f0 values in ERB. Note there is a third dimension shown with the darkness of the curve colours: the variance of PC scores by their standard deviation (SD) – the darkest purple adds one SD of the PC score to the average curve and the darkest blue represents subtracting one SD of the PC score from the average curve. The information in the boxplots can be matched up with the PC curves plot. The boxplots show the z-scored f0:

$$scaled\ f0 = (f0 - mean) / SD$$

Since the mean values for the PC scores were 0, the scaled PC scores in the boxplots show 25th percentile, median, and 75th percentile of the number of SDs. For instance, in Figure 9B, the solid lines in the boxes show that the number of SDs PC1 has for the statement tune is -0.25 times of the standard deviation of PC1 scores, while that for the IntQ tune is 0.609 times of the standard deviation of PC1 scores. Therefore, the PC1 curve for the statement tune is roughly corresponding to the 4th blue line from the bottom, while the IntQ tune resembles the second purple line from the top. The average curves in Figure 8 demonstrate how the two PC curves can be combined into one curve. This method of interpretation is suitable for all following PC curve figures and boxplots.

The PC curves show the major differences between the two tunes. The PC1 curves in Figure 9A exhibit a register difference between the two tunes; and the PC2 curve show a slope difference between the tunes. According to the median values of the number of SDs of PC1 shown in Figure 9B, the IntQ tune's PC1 scores are generally higher than those of the statement tune, which suggests that the IntQ tune resembles the purple lines more, while the statement tune resembles the blue lines more. Therefore, the IntQ tune demonstrates a higher register than the statement tune. Figure 9C shows that the IntQ tune has a higher median score of PC2 SD than statement as well. Therefore, the IntQ tune corresponds to the curves with a gentler slope than the statement tune in Figure 9A. The results from the linear mixed effect models reported that both PC1 and PC2 scores were significantly higher for the IntQ tune than the statement tune (the main effect of TUNE in PC1: $\chi^2(1) = 99.58$, $p < 0.001$ ***; PC2: $\chi^2(1) = 15.7$, $p < 0.001$ ***). While the PC2 curves showed a lower starting point for the IntQ tune, the difference signalled is the slope difference, since the average curves of the two tunes for Tone 1 in Figure 8 are not lower than the statement tune.
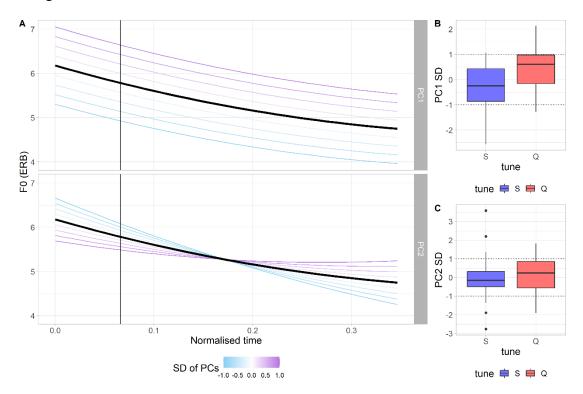

Figure 9. *PC curves and boxplots showing PC1 and PC2 scores for Tone 1 ^{LR}(HL). A: PC curves. X-axis represents registered time in seconds after normalisation; y-axis*

2.2.2.2 Tone 4: ^{HR}(HL)

Since Tone 4 is also a falling tone, the PC curves exhibit almost the same patterns with the ones for Tone 1. PC1 explained 87.6% of the variance in the dataset, and PC2, 12.4%. Figure 10A also shows a register difference and a slope difference between the two tunes. According to Figure 10B and C, IntQ tune has higher values for PC1 and PC2 which are both above 0, while statement tune has lower values and both were below 0 (the main effect of TUNE in PC1: $\chi^2(1) = 105.18$, $p < 0.001$ ***; PC2: $\chi^2(1) = 15.7$, $p < 0.001$ ***). Therefore, the IntQ corresponds to the purple lines and statement, the blue lines. In Figure 10B, PC1 demonstrates a higher register for the IntQ tune for Tone 4 data. Figure 10C shows that the slopes of the IntQ PC2 curve was gentler than the statement PC2 curve, especially after the highest peak of the lexical tone where the intonation influenced the lexical tone curve most. Yet, although the IntQ tune of both Tone 1 and Tone 4 have higher ending parts than their statement counterparts, IntQ curves with Tone 1 and Tone 4 still end with a local L tone, which keeps the falling tones falling.

Figure 10. *PC curves and boxplots showing PC1 and PC2 scores for Tone 4 ^{HR}(HL). A: PC curves. X-axis represents registered time in seconds after normalisation; y-axis shows the f0 values in ERB; the darkness of the colours represents adding (purple) or subtracting (blue) the number of SDs to/from the PC scores from the average line (black). B: Boxplot showing scaled PC1 scores (the number of SDs of the PC1*

2.2.2.3 Tone 2: ^{HR}(LH)

The PC1 scores explained 90.9% of the variance in the Tone 2 data, and PC2, 9.1%. Figure 11A once again shows a register difference and a slope difference between the two tunes: the IntQ tune has a higher register and steeper slope, as shown in Figure 11B and C (the main effect of TUNE in PC1: $\chi^2(1) = 94.56$, $p < 0.001$ ***; PC2: $\chi^2(1) = 56.81$, $p < 0.001$ ***).
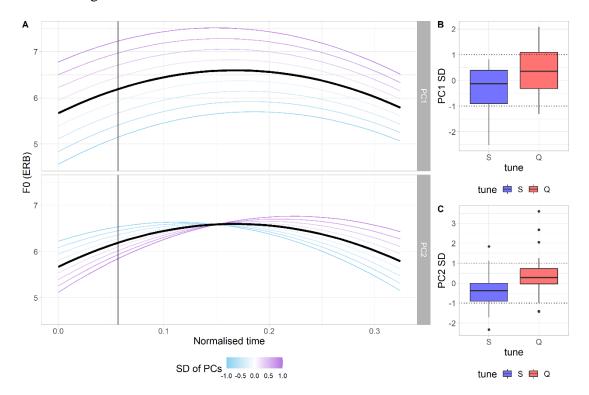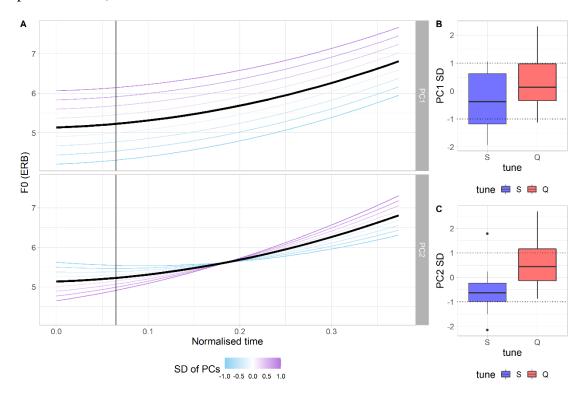

Figure 11. *PC curves and boxplots showing PC1 and PC2 scores for T2: ^{HR}(LH). A: PC curves. X-axis represents registered time in seconds after normalisation; y-axis shows the f0 values in ERB; the darkness of the colours represents adding (purple) or subtracting (blue) the number of SDs to/from the PC scores from the average line (black). B: Boxplot showing scaled PC1 scores (the number of SDs of the PC1 scores). C: Boxplot showing scaled PC2 scores (the number of SDs of the PC2 scores).*

2.2.2.4 Tone 3: ^{LR}(LH)

The PC1 scores accounted for 84.1% of the variance in the Tone 3 data, and PC2, 15.9%. Figure 12A shows almost the same pattern as Figure 11 – the IntQ tune has a higher register than statement tune (Figure 12B, the main effect of TUNE in PC1: $\chi^2(1) = 66.92$, $p < 0.001$ ***) and has steeper slopes than the statement tune especially after the lowest f0 point (Figure 12C, the main effect of TUNE in PC2: $\chi^2(1) = 34.92$, $p < 0.001$ ***). Similar to Tone 2, the IntQ tune for Tone 3 also ends high phonetically.
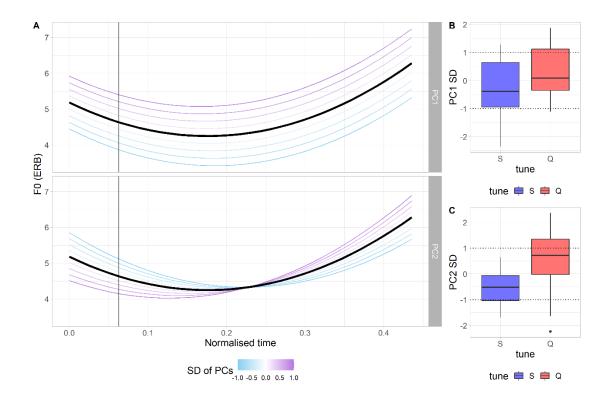
Figure 12. *PC curves and boxplots showing PC1 and PC2 scores for T3: ^{LR}(LH). A: PC curves. X-axis represents registered time in seconds after normalisation; y-axis shows the f0 values in ERB; the darkness of the colours represents adding (purple) or subtracting (blue) the number of SDs to/from the PC scores from the average line (black). B: Boxplot showing scaled PC1 scores (the number of SDs of the PC1 scores). C: Boxplot showing scaled PC2 scores (the number of SDs of the PC2 scores).*

2.3   Discussion of production results

2.3.3   Higher register: H$_R$

The results of mean f0, maximum f0, minimum f0, as well as the PC1s of all tones in the fPCA analysis showed a clear register difference between the IntQ tune and a statement tune, with the register of IntQ tune being higher than the statement tune. We use the symbol of H$_R$ to represent the high register, same as what we used for the high lexical tone register. In Hyman's (2008) categorisation, this type of tone-tune interaction belongs to the "avoidance" type, where the IntQ does not modify the tonal change. Note that Hyman's classification refers to languages in which non-tonal changes are the sole changes, while in Tianjin Mandarin, except for the register change, tonal modification co-occurs.

Previous studies on standard Mandarin polar questions reported mixed results about whether the register was higher in general or only the bottom or top line was raised. Our study showed support for the whole register being lifted, including both the lower limit and upper limit. Register raising is not a rare phenomenon in questions in tonal

language questions. For instance, in Xhosa, a tonal Bantu language, a raised register at the beginning of sentences indicates a question tune (Jones et al., 1998). Thai and Vietnamese, both being contour-tone languages, also have a similar register lift in their question tunes (Dung et al., 1998; Luksaneeyanawin, 1998). A number of studies on Standard Mandarin have also shown that global register raising is very important in Mandarin questions (J. Shen, 1992; Shih, 2000).

2.3.4 H floating boundary tone: H̥$_I$

In the mean value analysis, the f0 range data showed that the falling lexical tones (T1 and T4) had a smaller f0 range for the IntQ than statements; while the two rising tones (T2 and T3) had a larger range with IntQ tune than with statement tune. The PC2 curves in the fPCA analysis also showed that the IntQ tune had a gentler slope than the statement tune towards the end of the utterance for the two falling tones, and a steeper slope in the IntQ tune than the statement tune for the rising tones. These results indicate that the fall in the IntQ tune was not permitted to fall as much as in the statement tune, and the rising tones were encouraged to rise higher in the IntQ tune than in the statement tune. This boundary phenomenon is a phonological H̥$_I$ tone rather than a simple phonetic rise/ fall, since a phonetic rise/ fall would result in a same phonetic realisation for all four lexical tones – the IntQ tune for all the four tones would end with either a H or a L tone. In the case of the Tianjin Mandarin, although the IntQ tune ends higher than the statement tune, it keeps the mixture of ending Hs and Ls from the original lexical tone contours. Based on these results, we propose that it is a H floating boundary tone that is at the end of the IntQ to deter the falls from falling and facilitate the rises with rising. We adopt the floating tone symbol from previous lexical tone literature, which is a small circle under the tone as "T̥", e.g., "H̥" or "L̥". To make it clear that it is a boundary tone at an Intonational Phrase (IP) level, we follow Hayes and Lahiri (1991) in using a subscript I to symbolise the IP boundary tone. Therefore, our production data showed that the IntQ in Tianjin Mandarin features a high floating boundary tone, H̥$_I$. We continue our discussion on the concept of boundary tone and floating tone in order to elaborate the reason for the proposal of the term and the necessity of separating floating boundary tones from the regular boundary tones.

2.3.4.1 Boundary tone

The concept of "boundary tone" is often used in the intonation literature. The definition of "boundary tone" was made explicit in Ladd (1996, p. 103): "The H% boundary tone always indicates a final rise, which typically takes place at the very end of the phrase or utterance." This suggests that if a H$_I$ boundary tone (same with H%) exists, the end of the prosodic unit is phonetically high; if the boundary tone is a L%, then the boundary physically falls low. Most existing studies in intonation research follow this criterion.

Such definition of a boundary tone can be applied to tonal languages without modification. As shown in Figure 13 (schematised based on Ma et al., 2006, fig. 1), in Cantonese, a clear high boundary tone presents both on the rising tone (Tone 25 on the left) and the falling tone (Tone 21 on the right).
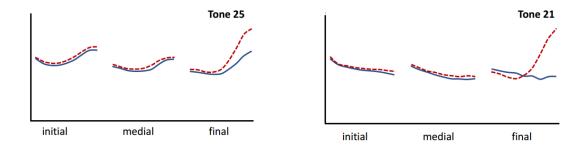
Figure 13. *Cantonese questions (red dotted lines) and statements (blue solid lines).*

However, the literature of boundary tones in Standard Mandarin has divided opinions. One view posits that boundary tone is not necessary to be included in the phonological representation in Mandarin (Wang & Shi, 2010; Yuan, 2011; Yuan & Shih, 2004); the other side argues that boundary tone is essential (Flemming & Nie, 2016; Ho, 1977; M. Lin, 2006, 2004). However, this is not actually a division in term of research results, but reflects different choices of the definitions of "boundary tone". For example, Lin (2004) argued that there was a H boundary tone in the interrogative tune, as shown as the red broken line in Figure 14. However, the boundary tone for the falling lexical tone (T4) remained a L tone despite of the raise from the statements. Therefore, the boundary tones claimed by these studies are not actual H% boundary tones and would be beneficial in having a separate category for this type of boundary tones.
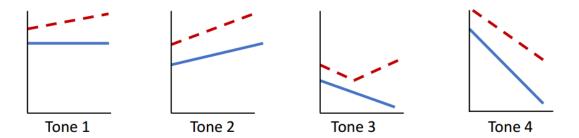

Figure 14. *Boundary tones in Standard Mandarin. Schematised based on Lin (2004, fig. 6)*

Similar to Standard Mandarin, Tianjin Mandarin does not possess any boundary tones that can be easily detected from the f0 contours. Instead, it has an invisible anchor point that modifies the lexical tone, which is very similar to the concept of a floating tone in tonal literature.

2.3.4.2 Floating Tone

Hyman and Tadadjeu (1976, pp. 60–61) divided floating tones into two categories – grammatical floating tones and lexical floating tones. Grammatical floating tones act as "tonal morphemes" (Welmers, 1959). In such cases, "… a tone can exist without a TBU, and vice versa. When a morpheme consists solely of a tone, it is referred to as a tonal morpheme" (Hyman, 2011). Lexical floating tones, however, do not serve any

grammatical function. The floating tones are often manifested as upstep/downstep of the neighbouring tones.

We propose the term "floating boundary tone" since this type of floating boundary tone has the *function* of a grammatical floating tone since it constitutes an essential part of the prosodic grammar and changes the sentence types; however, it does not serve any morphological or syntactical function as in the tonal literature. It has the phonological *effect* of a lexical floating tone on the neighbouring tones – upstepping the final part of the final lexical tone. For instance, in Tianjin Mandarin, the floating H boundary tone functionally contributes to the formation of an IntQ tune, and phonologically attaches to the boundary, and plays a role in either enhancing the rise or blocking the fall of an existing lexical tone, which itself is manifested at the boundary. Thus, for an LH lexical tone, the H_I enhances the final H, while for the HL tones, the floating tone prevents the final L from lowering further and thus keeping the HL tone shape, rather than attaching an H to the boundary and produce an HLH contour. This type of tone-tune interaction is what Hyman and Monaka (2008) described as "submission", where lexical tones "surrender" to intonation.

The most important reason for separating the floating boundary tones from regular boundary tones is that both types of boundary tones can occur in the same language. Zhang (2018a, 2018b) has demonstrated how an actual L boundary tone can be attached to the lexical tones to form the chanted call tune in Tianjin Mandarin. In Hyman and Monaka's (2008) categorisation, this type of tone-tune interaction is "accommodation", where intonation tones can be added to the lexical tones.

Separating floating boundary tones from regular boundary tones is not only important in Tianjin Mandarin, it is also useful in many other Mandarin varieties, including Standard Mandarin (Flemming & Nie, 2016), Sichuan Mandarin, etc. For instance, in Sichuan Mandarin, the lexical tones go through different changes in questions, as shown in (1) (Chang, 1958, cited from Ladd 2008: 159):

(1) Changes of the lexical tones of final syllables in questions:
high-rising: remains high and often ends higher than usual, ^{HR}(LH) → ^{HR}(LH+)
low-falling: becomes low level, ^{LR}(HL) → ^{LR}(HH)
high-falling: becomes high level, ^{HR}(HL) → ^{HR}(HH)
low-falling-rising: becomes low rising, ^{LR}(HLH) → ^{LR}(LH)

The changes to the first three tones are identical with the changes in Tianjin Mandarin: the rising tone rises higher in questions than in statements, and the falling tones do not fall in question as much as in statements. The IntQ tune in Sichuan Mandarin clearly also presents a floating H boundary tone at the IP boundary.

Floating tones can also explain phenomena in other languages. In a contour-tone language, Thai, the question tune seems to behave in a similar way as in Tianjin Mandarin. Luksaneeyanawin (1998, p. 386) presented schematised figures of declarative tune and interrogative tune in Thai. A floating tone analysis can explain the differences between the two tunes across all lexical tones, especially for the Mid tone and the Low tone. In non-contour-tone languages, such as the African tonal languages mentioned in the Section 1, the reduction of downdrift in African question (Rialland,

2007) may be as a result of floating tones. In a pitch-accent language, Danish, the intonation contours of statement and questions are largely the same with a potential floating tone at the end to raise the pitch, based on the schematised results presented in Grønnum (1998, p. 143) .

2.3.5   Lengthening in IntQ Tune

Apart from the tonal differences between the two tunes, we also investigated the durational difference, which often co-occurs with the change of tune. The duration differences between IntQs and statements in Tianjin Mandarin are statistically significant – not only across different tunes, but also systematically different across different lexical tones. Yuan (2012) studied the duration of sentence-final syllable in IntQ and statements in Standard Mandarin. He found that in sentence-final position, Tone 4 (falling tone, HL) was longer in IntQ than in statements; while Tone 2 (rising tone, LH) did not change significantly. A study on Cantonese by Ma, Ciocca, & Whitehill (2006) also had similar results: questions had longer durations than statements at sentence-final position. In some non-tonal languages, questions were also observed to have longer durations than statements, especially with the last part of the utterances; for example the last segment in Italian (Cangemi & D'Imperio, 2013) and last vowel in Galician-Asturian (Muñiz Cachón et al., 2006). Our duration results exhibited the same pattern as these studies, especially Yuan's on Standard Mandarin: IntQs are longer than statements; the difference between the durations of IntQs and statements manifest more significantly in falling lexical tones than rising lexical tones. In both Cantonese (Ma et al., 2011) and Italian (Cangemi & D'Imperio, 2013), follow-up perception experiments were conducted to investigate whether duration was a cue for perceiving the difference between IntQ and statements: in both languages, duration was not a useful cue. These experiments left whether durational difference should be included in the phonological description of tunes unclear. Moreover, the great variability of duration within each tune in the current study (152 ms to 478 ms for rhymes in statements, and 146 ms to 527 ms for rhymes in IntQ), together with the reult that the average difference is merely 34 ms, also made it difficult to conclude that duration is a phonologically meaningful factor in distinguishing IntQs from statements. In contrast, lengthening is a phonological feature of the chanted call tune in Tianjin Mandarin and the average difference between the lengths of the monosyllabic rhymes with the chanted call tune and the statement tune is 321.1 ms – six times that of the difference between an IntQ and a statement. Therefore, with the evidence we have so far, we do not consider lengthening as a part of the IntQ tune in Tianjin Mandarin.

2.3.6   Summary of production study results

In summary, we concluded that the Tianjin Mandarin intonational polar question tune is a raised register and a H floating boundary tone, i.e. ^{HR} H_I. The tunes also systematically differ by the lexical tone register and the lexical tone shape. In order to examine how register and boundary tone are used in the perception, and how lexical tones interact with the intonation in perception, we conducted a tune identification experiment for investigation.

3.   Perception Study

While the production study presented clear distinction between the two tunes, the differences could be subtle even for native speakers. The IntQ tune is often regarded as

one of the most basic and salient intonation tunes in various languages that allow such YNQ, since IntQ is syntactically identical to its declarative counterpart, and intonation is the only cue to suggest that it is a question. However, it is not always the case that such questions can be perceived without failure, especially when a language uses pitch for various acoustic properties. In tonal languages such as Tianjin Mandarin, as shown in the production study, the differences between IntQs and statements are subtle. Communication failures sometimes occur when an IntQ tune is used, even between native speakers, and even when the speaker and hearer are familiar with each other's speech. In this experiment, we will investigate how lexical tones interact with intonational question tune. To be specific, the following questions are examined:

i. How well are IntQs perceived by native Tianjin speakers?
ii. Does lexical tone play a role in intonation identification? If yes, is there a tonal bias?
iii. Are the register cue and floating boundary tone cue from the production study used in perception?

A tune identification task was designed to investigate these issues. Natural stimuli (i.e. not artificially manipulated and synthesised) were used in the experiments to investigate native speakers' ability to identify real utterances.

3.1   Literature on tune identification

In non-tonal languages, the perception of intonational tunes is rarely studied with the aim of finding out whether the native speakers can perceive a certain intonation, since it is fairly straightforward and there is hardly any reason for native speakers not being able to recognise the tunes. Tune identification is therefore more commonly used with resynthesised materials to test certain hypotheses. For instance, in order to study whether the durational cue is used in the identification of tunes, Cangemi and D'imperio (2013) created a nine-step continuum using two base factors, i.e. F0 and duration.

Ni and Kawai (2004) generated a continuum to study the relationship between the perception of questions and the F0 height of the final syllable in Standard Mandarin. They stimuli sentences contained four, five, or nine syllables. They discovered that for Tone 2 (LH Tone), the higher the f0 peaks were, the easier it was for the native speakers to consider them as questions; for Tone 4 (HL Tone), the utterance was only regarded as a question when the whole tone was placed at a higher register.

However, contrary results were found in a series of studies on Standard Mandarin intonation identification (Yuan, 2011; Yuan & Shih, 2004). In Yuan (2011), a corpus of 130 sentence (an 8-syllable long sentence was given as an example) that varied in terms of the tones of the last syllables, the tunes (question tune or statement tune), as well as the focus location (initial, medial, final), were used as stimuli. The corpus had 1040 stimuli in total, which were recorded by 8 speakers. The results from these studies demonstrated that when the last syllable of an utterance carried a Tone 4, it would be easier for the utterances to be correctly identified as a question. On the contrary, those with Tone-2 final syllables were the most difficult ones to identify. Liu et al. (2021) examined the identification of Standard Mandarin intonation with semantically neutral context and semantically constraining context. They found that, in syntactically unmarked polar questions, T4 (HL) had slightly higher identification accuracy in

semantically constraining context, but had marginally lower accuracy in non-constraining context than T2 (LH).

Xu and Mok (2012) studied the cross-linguistic perception of four types of sentences in Cantonese and in Standard Mandarin by Cantonese speakers and Mandarin speakers respectively. The four types of sentences were complete statements, cut-off statements (with the final syllable being cut off), complete questions, and cut-off questions (with the final syllable being cut off). Cantonese speakers had the most problems with identifying questions with the Cantonese Tone 1 (H Tone) in complete sentences, although the accuracy was still above 90%. For the cut-off questions, the identification was much more difficult. The tones with higher register gained higher identification rate than the ones with lower register. The Mandarin tune identification task by native Mandarin speakers had the same results as Yuan's (2011): T4 (HL Tone) had higher accuracy than T2 (LH Tone), in both complete and cut-off sentences. The cut-off sentences achieved the same level of accuracy as the complete sentences in Mandarin. They therefore concluded that in Mandarin it was the global contour that mattered the most in question identification.

Ma et al. (2011) also investigated the perception of statements and IntQs in Cantonese. They varied the last syllable of a sentence. The results from this study showed that Tone 33 (mid-level Tone) had the highest accuracy, while Tone 55 and 25 had the lowest accuracy (note: the other tones in Cantonese are 21, 23, 22). The results partly supported Yuan (2011) in that the tones ending with H were the most difficult to distinguish in Cantonese as well.

The above studies present diverging results. In some cases, a high tone facilitated the identification of the question, while in some other cases, T4, which is a falling tone, was easier to be identified as a question. Nevertheless, in general it appears that a higher register leads to a preference for identifying a question. None of the studies systematically compared production data to perception. Since we have identified the IntQ tune as $H^R \underset{\circ}{H}_I$ in our production study, we aimed to test whether these were phonologically significant cues for the identification of the IntQ and statement tunes with a tune identification task.

3.2 Methodology

3.2.1 Participants

A total of 28 native Tianjin speakers (age range: 20-28, mean = 20.25), including 13 female and 15 male speakers, participated in the perception experiment. The participants were university students at Tianjin University, Nankai University, and Tianjin College of Commerce. None of them reported any hearing loss or speech disorder. They were screened by a pre-test in which they were asked to produce a few utterances in Tianjin Mandarin to ensure they could use Tianjin Mandarin natively. The participants were not the same participants as the informants in the previous production. All participants received payment for their participation.

3.2.2 Stimuli

The stimuli used in the experiments were all from the previous production study (Table 3) by one female speaker (F103) and one male speaker (M101). The total number of stimuli is 96: 3 monosyllables ([ma], [mau], [mi]) * 4 lexical tones * 2 tunes (IntQ, Statement) * 2 speakers (male speaker, female speaker) * 2 repetitions = 96 stimuli. To evaluate the suitability of the chosen stimuli, three native speakers participated in a pilot identification task without limitation on the response time or the number of times they could listen to the stimuli. All chosen stimuli had at least two correct responses.

3.2.3 Procedures

The stimuli were programmed in the experimental software SPLICE (Reetz & Kleinmann, 2008). The participants heard three bleeps at the beginning of the experiment as a signal for the start of the experiment. For each stimulus, the participants heard two 200 ms bleeps separated by a 100 ms pause. Then a 300 ms pause was inserted before the onset of the audio stimulus. The time given to participants for deciding whether a stimulus was a question or a statement and pressing the button was 1500 ms from the offset of the audio stimulus. The reaction time was recorded from the onset of the audio stimulus.

The experiments were conducted in a quiet classroom at Tianjin University. The auditory stimuli were played through individual closed-ear headphones (Sennheiser PX200 stereo headphones). The participants were asked to decide whether each stimulus was a question or a statement. They made their choices on two-button handsets with two buttons labelled as "陈述。" ("statement" in Chinese, with a Chinese full-stop) above the left button, and "疑问？" ("question" in Chinese, with a Chinese question mark) above the right button. The stickers were switched for left-handed participants to keep the "question" choice on their dominant-hand side. The participants were explicitly instructed to press the buttons with their thumbs and make their responses as accurate and as fast as possible. A practice block containing non-experiment fillers were used at the beginning of the experiment to familiarise the participants with the task.

3.2.4 Data analysis

Responses that were made to stimuli that contained the syllable /ma/ were excluded in all three experiments. In Mandarin, multiple characters share the same pronunciation. One frequent token of /ma/ with Tone 4 is 骂 ("to scold"); however, in Tianjin Mandarin, the dialectal wh- word 嘛 ("what") also has the same pronunciation. It thus created an opposite trend against the other rhymes (Figure 15 for an illustration of Tone-4 monosyllabic words with different rhymes – *a*, *ao*, and *i*). This trend shows that the semantic meaning of a word is much more influential in the identification of a question than the intonational cues. To prevent the influence from the semantics of *ma*, data of all *ma* instances (32 trials in total), including all four tones across two tunes (as shown as shaded in Table 3), were excluded.
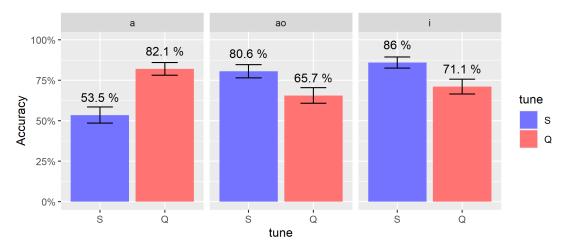
Figure 15. *Accuracy of responses for trials with Tone 4. The bars depict the accuracy rate for the two tunes (Blue: Statement; Red: IntQ) grouped by the rhymes of stimuli: a [a], ao [au], and i [i].*

Accuracy was analysed with generalised linear mixed models, using the *lme4* package (Bates et al., 2015) in R (R Core Team, 2021). To analyse the accuracy of the responses, binomial regression models were constructed with the ACCURACY (Correct, Incorrect) as the dependent measure. Similar to the models built in the production study, the full model was built with the following fixed effects: TUNE (Statement, Question), TONE (T1, T2, T3, T4), and the interaction between TONE and TUNE. The random intercepts included PARTICIPANTS and ITEM, and TUNE was included as by- PARTICIPANTS and by-ITEM random slopes. The random slopes were removed when the models did not converge or reported singularity.

Reaction time data were calculated using the reaction time recorded by SPLICE with the duration of the audio stimulus subtracted from it. Early responses that were made before voice onset were excluded from the analysis, since the participants made the responses before hearing the audio stimuli. The modelled reaction time were the reaction time for the correct trials only. The reaction time data were analysed with linear mixed effect models, with the same main factors and random factors as in the accuracy models. Post-hoc *TukeyHSD* tests were then run to test the effects of different tones on each tune to better understand the results. The generalised linear mixed models were evaluated using the *Anova* function in the *car* package and the linear mixed effect models were evaluated using the anova function in the *lmerTest* package. In these analysis, the lexical tones were used directly as the main effect instead of using the tone shape and tone register, in order to better understand the effect of each lexical tone.

To further examine the roles of tone shape and tone register in the identification process of each tune, the data were further analysed using decision tree using the *party* package in R (Hothorn et al., 2006). One model was created for each tune with the full dataset for the tune. The decision tree method provided the most likely approximation of what the identification process was by classifying the accuracy and reaction time results with the two lexical tone features (tone register and tone shape). This analysis helps illustrate the tone-tune interaction and cross validate the results from the perception study.

3.3   Results

3.3.5   Accuracy

The accuracy was significantly predicted by TONE ($\chi_2(3) = 49.2$, $p < 0.001$ ***), and the interaction between TONE and TUNE ($\chi_2(3) = 111.2$, $p < 0.001$ ***), but not TUNE alone ($\chi_2(1) = 1.7$, $p = 0.198$, n.s.). Figure 16A and B show the accuracy rate of the identification of the statement tune and IntQ tune respectively. The bars ranked the accuracy categorised by the lexical tones from high to low. The *TukeyHSD* test reported the following significance values as labelled in Figure 16A and B.
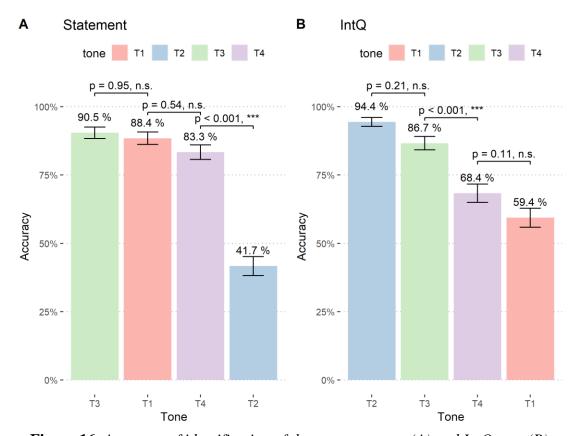

Figure 16. *Accuracy of identification of the statement tune (A) and IntQ tune (B), categorised by lexical tones. The tones are displayed by accuracy from high to low.*

To summarise the results and simplify the presentation, in (2) and (3) below, the results were presented using the almost equal sign "≈" for symbolising that the two elements connected did not differ significantly from each other; the greater-than sign ">" and lesser-than sign "<" were used to indicate significant differences. For the statement tune identification in (2), the differences between T3 and T1, and between T1 and T4 did not reach statistical significance. However, the accuracy for T2 statements were significantly lower than the other three tones: 41.7% (in Figure 16A), which was even lower than the lowest in IntQ tune (T1 Figure 16B: 59.4%), and was also lower than chance level. As we see in (2), T2, which has both a rising contour and a high register, is significantly different from the others. Moreover, descriptively (as noted above), the tones with the H register (T4 and T2) pattern together as stimulating low identification rate, and the tones with L register (T3, T1) generate higher identification rate.

For the IntQ stimuli in (3), a significant difference between T3 and T4 was reported, while T2 and T3, T4 and T1 were comparable. The most accurately identified stimuli carried lexical tone T2, while the least accurate were utterances with T1. The difference between T1 and T2 was very large, T2 being identified correctly in most instances while the identification of T1 was only near chance level. These results suggest that the LH rising tone shape (T2, T3) have higher accuracy than the HL tones (T4, T1).

(2) S: T3 ^{LR}(LH) ≈ T1 ^{LR}(HL) ≈ T4 ^{HR}(HL) > T2 ^{HR}(LH)

(3) IntQ: T2 ^{HR}(LH) ≈ T3 ^{LR}(LH) > T4 ^{HR}(HL) ≈ T1 ^{LR}(HL)

It is important to note that there is a clear and perhaps predictable interaction between the lexical tones and the identification of the IntQ tune. From the production study, we know that the IntQ tune H^R $H̥_I$ has an overall raising effect, particularly towards the boundary. It is conceivable that despite the overall rise, an HL tone, whatever maybe its register, is harder to identify as a question than an LH tone. We now turn to latency measures.

3.3.6 Reaction Time

Recall that we also recorded the reaction times of the all the responses. All main effects were significant factors that affected the reaction time of the responses: TUNE ($\chi^2(1) = 45.12$, $p < 0.001$ ***), TONE ($\chi^2(3) = 7.2$, $p < 0.001$ ***), and the interaction between TUNE and TONE ($\chi^2(3) = 21.57$, $p < 0.001$ ***).

Figure 17 shows the reaction time results for the correct trials from short to long. The reaction time data exhibit a mirror pattern of the accuracy data. The reaction time data for the statements in Figure 17A shows that T2 needs a significantly longer reaction time than any other tones. This result converged with the accuracy results: T2 had the lowest accuracy but longest reaction time, indicating that the participants had great difficulty with identifying statement tune with T2. Further, as shown in (4), the reaction time for identifying statement was also influenced by the register of the lexical tones.

For the IntQ tune in Figure 17B, T1 had the longest reaction time of 611.16 ms, while T2 was much shorter (457.99 ms). Post hoc Tukey Test did not show any significant differences between neighbouring tones. Only T1 and T2, T4 and T2 differed significantly, as shown in (5), which again showed mirror pattern with the accuracy.

(4) S: T3 ^{LR}(LH) ≈ T1 ^{LR}(HL) ≈ T4 ^{HR}(HL) < T2 ^{HR}(LH)

(5) IntQ: T2 ^{HR}(LH) ≈ T3 ^{LR}(LH) > T4 ^{HR}(HL) ≈ T1 ^{LR}(HL)
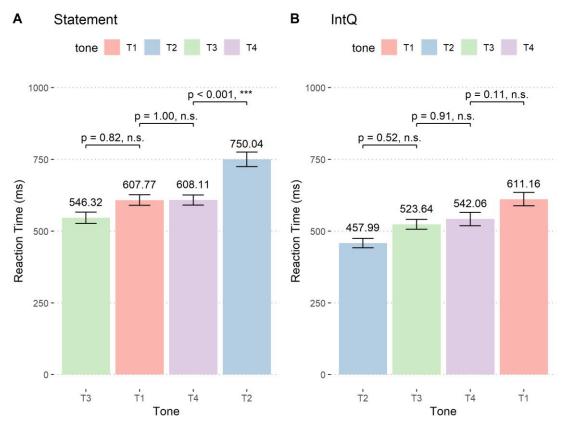
Figure 17. Reaction time for correctly identified trials *of the statement tune (A) and IntQ tune (B), categorised by lexical tones. The tones are displayed by reaction time from short to long.*

3.3.7 Effect of tone register and tone shape: a decision Tree analysis

The production study concluded that the lexical tone register and lexical tone shape were significant factors that influenced the phonetics of the tunes. The above perception results also presented patterns where the accuracy and reaction times pattern by lexical tone properties. Figure 18 shows the procedures for categorising the correct and incorrect responses in the identification experiment results. As shown in Figure 18A, in statement, tone register was the first cue used to divide the data: if the tone register is a L, no other cues is needed; if the tone register is a H, the cue of tone shape is then used. Figure 18B illustrates how IntQ responses were made. In contrast to the statement tune, the first cue used in identifying the IntQ tune is tone shape: only the tone shape information is needed when the tones are HL tones; if the tones are LH tones, the tone register information is then needed to categorise the data. The decision trees for the reaction time data are shown in Figure 19. The order of the cues used for the reaction time data is the same as what are used in the accuracy data: tone register first and then tone shape for the statement tune, and the other way round for the IntQ tune.
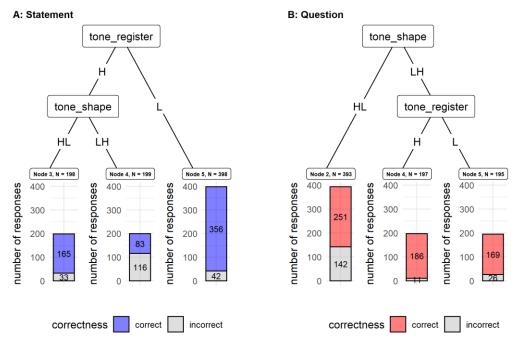
Figure 18. *Decision tree for (A) statement tune and (B) IntQ tune, based on accuracy. Node number indicates the order of the nodes used in the model; N is the number of responses explained in the node; the number of correct and incorrect responses explained in the node are colour-coded in the bar plots.*
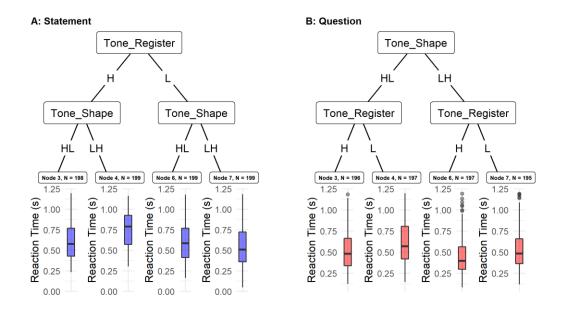

Figure 19. *Decision tree for (A) statement tune and (B) IntQ tune, based on reaction time. Node number indicates the order of the nodes used in the model; N is the number of responses explained in the node; the 25th percentile, median, and 75th percentile of the reaction time data explained in the node are illustrated wih the boxplots.*

3.4 Discussion of Perception Results

Table 3. Summary of perception results

	Accuracy	Reaction Time
Q	T2^{HR}(LH)≈T3^{LR}(LH)>T4^{HR}(HL)≈T1^{LR}(HL)	T2^{HR}(LH)≈T3^{LR}(LH)<T4^{HR}(HL)≈T1^{LR}(HL)
S	T3^{LR}(LH)≈T1^{LR}(L)≈T4^{HR}(HL)>T2^{HR}(LH)	T3^{LR}(LH)≈T1^{LR}(L)≈T4^{HR}(HL)<T2^{HR}(LH)

Table 3 presents a summary of the perception results. The interaction between TONE and TUNE exhibited an interesting general pattern: in IntQ tune, the rising tones (T2, T3) were identified both more accurately and faster than the falling tones (T4, T1); while in statements, the low tones (T3, T1) were better identified than the high tones (T4, T2). The decision tree results, as illustrated by Figure 20B and Figure 19B, also showed that for the IntQ tune, tone shape was the first cue used modelling the identification process. The tone shape of HL alone can account for the majority of the responses in the accuracy data.

For IntQ tune, T2 ^{HR}(LH), with both high register and high ending tone, was correctly identified as questions most of the time. T3[^{LR}(LH)] ends with a H tone too; yet, the L register of T3 became a distractor in IntQ identification. Therefore, although the accuracy of T3 was significantly above chance, it was not as high as T2. The decision tree results in Figure 21B also shows that Node 5, i.e. LH tone shape and L tone register, generated more incorrect responses than Node 4 (LH tone shape and H tone register). The HL-L nodes (Node 4 and Node 7) in Figure 19B also showed longer reaction time than their neighbour nodes.

From these results, it is evident that the listeners were trying to look for a H ending tone for the IntQs. These results are consistent with the production data, in which the IntQ tune was found to have a floating H̥$_I$ boundary tone associated with the IP boundary, despite the fact that the floating tone did not have a clear phonetic realisation. During the identification process, the listeners searched for this information from the audio stimuli. A H ending from the lexical tones are confusing during the process of IntQ tune identification.

In contrast, for the statements, the crucial cue was not the right boundary but the register. The L register facilitated the identification of a statement. T1 and T3 both have an initial L, so they both achieved the highest accuracy. When the initial tone did not help with identification, such as T4 and T2, the ending tone interfered with the identification. T2 ends with a H tone that interferes with the identification, therefore the accuracy is extremely low. T4, although it does not have a facilitating L register, it does not have an interfering H tone at the end either. The decision tree results in Figure 22A and Figure 19A further supported this analysis. Figure 23A showed that if the register is L, the vast majority of the data are correct responses. The responses with a distracting H register and a distracting LH tone shape had the most incorrect responses (Node 4, Figure 24A) and the longest reaction time (Node 4, Figure 19A). In the production study, IntQ tune was observed to have a higher register than the corresponding statement tune.

The results here show that the register information is also used during the idetification process.

In summary, the results suggest that during the identification of IntQ tune, the floating H̥I is what the listeners search for; and the L register interferes with the identification. While the identification of statements regards the L register as the most crucial cue; the ending H interferes with the identification.

These results are not consistent with some of the findings in Standard Mandarin. Two Mandarin studies (Xu & Mok, 2012; Yuan, 2011) both found the IntQ tune of Tone 4 (HL Tone) to be the easiest to identify while that of the Tone 2 (LH Tone) was the most difficult. The results of the current study are completely opposite in that the high rising tone was the easiest to identify as the IntQ tune. However, the results from the current study do support the findings in Ni and Kawai (2004) that a higher pitch peak for Tone 2 would indicate a question tune and a higher register facilitates with the identification of question tunes with the falling Tone 4. Although the Mandarin results in Xu and Mok (2012) are different from the results of the experiments from our results, their Cantonese results saw some similarities with our findings that the tones with higher register provided more information for the identification of questions with cut-off endings. Due to the very small number of comparable studies, it is difficult to draw any meaningful cross-linguistic conclusions based on the implications of some similarities and differences between these experiments. The perception, especially identification, of intonation tunes definitely needs to be studied more in-depth.

4.   Conclusions

We investigated the differences between the IntQ tune and statement tune in Tianjin Mandarin in both production and perception. It also addresses the issue of tone-tune interaction by examining how the features of lexical tones, i.e. tone register and tone shape, influence the production and identification of the tunes. The results of the production were analysed using traditional mean value analysis and fPCA. Both analyses corroborate each other. The results of the IntQs and statements were reported and discussed in terms of register, boundary tone, and duration by examining acoustic parameters of mean pitch, tonal scaling, tonal alignment, F0 range, and duration. The results are summarised below:
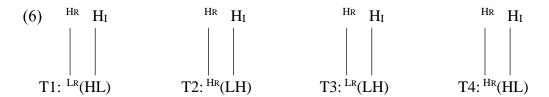
H_R Register: Both the onsets and the rhymes were higher with IntQ tune than with statement tune. This suggests that the register of IntQs is higher than statements. Both maxima and minima were higher in IntQs than in statements. This indicates that the register was lifted, instead of merely being expanded. In Tianjin Mandarin, register is not only used as a phonological means for distinguishing lexical tones, but also plays a significant role for questions.

H̥I Boundary tone: There was a floating H̥I boundary tone at the right edge of the IP. The F0 range of the onset did not differ between IntQ and statements. F0 range of the rhymes displayed an interesting pattern: the IntQ F0 range of falling tones (Tone 1 and Tone 4) showed a decrease from statements, while that of rising tones (Tone 2 and Tone 3) showed an increase. This suggests that the fall was deterred, while the rise was facilitated by a floating H̥I boundary tone.

Longer duration: On the utterance level (also the whole syllable), IntQs' duration was longer than that of statements. The duration of the rhymes was also longer in IntQs than in statements. The onsets were slightly longer in statements than in IntQs. However, further evidence from psycholinguistic experiments is needed to determine whether the duration is a phonologically meaningful cue.

For the perception study, the accuracy of the identification and the reaction time were analysed for the tone-tune interaction; then the usage of tone register and tone shape in the process of tune identification was analysed using a decision tree method. The perception results showed that the listeners utilise both an ending H tone and the register information for identification, which shows that Tianjin Mandarin speakers use the same cues for both production and perception for the IntQ tune.

Therefore, the final phonological representation of the tune is: ^{HR} H_I. The association with the four lexical tones is as in (6), where the high register ^{HR} of the tune associates with the whole utterance and manifest as a raised register for the lexical tone register. The final high floating boundary tone $\overset{\circ}{H}_I$ associates with the lexical tone and modifies the lexical tone.

(6)

^{HR} H_I	^{HR} H_I	^{HR} H_I	^{HR} H_I
T1: ^{LR}(HL)	T2: ^{HR}(LH)	T3: ^{LR}(LH)	T4: ^{HR}(HL)

Since lexical tones and intonation are so tightly intertwined with each other in a tonal language, this study is only a first step towards the exploratory investigation of the tune of IntQ, or Tianjin Mandarin intonation in general. We acknowledge that there are many unanswered questions, such as whether there is a pitch accent, what part of the floating boundary tone modify in longer utterances, how do other lexical-level phenomena interact with sentence intonation (see C. Zhang, 2019 for a preliminary discussion). For future investigations, we would like to examine longer utterance materials, with larger pools of speakers, and using less controlled data with varied tasks, to have a more well-rounded account for Tianjin Mandarin intonational polar question tune.

References

Armstrong, M. E. (2012). *The development of yes-no question intonation in Puerto Rican Spanish* [Doctoral Thesis, The Ohio State University]. http://rave.ohiolink.edu/etdc/view?acc_num=osu1345565869

Arvaniti, A., Ladd, D. R., & Mennen, I. (2006). Phonetic effects of focus and "tonal crowding" in intonation: Evidence from Greek polar questions. *Speech Communication*, *48*(6), 667–696. https://doi.org/10.1016/j.specom.2005.09.012

Asano, Y., & Gubian, M. (2018). "Excuse meeee!!": (Mis)coordination of lexical and

paralinguistic prosody in L2 hyperarticulation. *Speech Communication*, *99*(May 2017), 183–200. https://doi.org/10.1016/j.specom.2017.12.011

AudacityTeam. (2016). *Audacity* (2.1.2).

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software1*, *1*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Boersma, P., & Weenink, D. (2021). *Praat: doing phonetics by computer* (6.1.08). http://www.praat.org/

Bolinger, D. (1978). Intonation across languages. *Universals of Human Language*, *2*, 471–524.

Cahill, M. C. (2013). Polar question intonation in five Ghanaian languages. *LSA Annual Meeting Extended Abstracts*, *4*.

Cangemi, F., & D'Imperio, M. (2013). Tempo and the perception of sentence modality. *Laboratory Phonology*, *4*(1), 191–219. https://doi.org/10.1515/lp-2013-0008

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

Chen, A., & Boves, L. (2018). What's in a word: Sounding sarcastic in British English. *Journal of the International Phonetic Association*, *48*(1), 57–76. https://doi.org/10.1017/S0025100318000038

Cheng, C., Chen, J. Y., & Gubian, M. (2013). Are mandarin sandhi tone 3 and tone 2 the same or different? The results of functional data analysis. *27th Pacific Asia Conference on Language, Information, and Computation, PACLIC 27*, *4*, 296–301.

Daly, N., & Warren, P. (2001). Pitching it differently in New Zealand English: Speaker sex and intonation patterns. *Journal of Sociolinguistics*, *5*(1), 85–96. https://doi.org/10.1111/1467-9481.00139

Davison, D. S. (1991). An acoustic study of so-called creaky voice in Tianjin Mandarin. *UCLA Working Papers in Phonetics*, 50–57. https://escholarship.org/uc/item/8j00q68c.pdf#page=54

Dung, Đ. T., Huong, T. T., & Boulakia, G. (1998). Intonation in Vietnamese. In D. Hirst & A. Di Cristo (Eds.), *Intonation Systems: A Survey of Twenty Languages* (pp. 398–420). Cambridge University Press.

Flemming, E., & Nie, H. (2016). Are there boundary tones in Mandarin Chinese echo questions? *Linguistics Association of Great Britain Annual Meeting 2016*.

Fox, J., & Weisberg, S. (2018). *An R companion to applied regression* (3rd editio). Sage publications.

Gårding, E. (1984). Chinese and Swedish in a generative model of intonation. *Nordic Prosody*, *111*, 79–91.

Gårding, E. (1987). Speech Act and Tonal Pattern in Standard Chinese: Constancy and Variation. *Phonetica*, *44*(1), 13–29. http://www.karger.com/DOI/10.1159/000261776

Grice, M., & Savino, M. (2003). Question type and information structure in Italian. *Proceedings of Prosodic Interface 2003*.

Grønnum, N. (1998). Intonation in Danish. In *Intonation Systems: A Survey of Twenty Languages* (pp. 131–151). Cambridge University Press.

Gubian, M., Torreira, F., & Boves, L. (2015). Using Functional Data Analysis for investigating multidimensional dynamic phonetic contrasts. *Journal of Phonetics*, *49*, 16–40. https://doi.org/10.1016/J.WOCN.2014.10.001

Gussenhoven, C. (2002). Intonation and interpretation: Phonetics and Phonology. *Speech Prosody 2002: Proceedings of the First International Conference on Speech Prosody*, 47–57. http://www.isca-speech.org/archive/sp2002/sp02_047.pdf

Hayes, B., & Lahiri, A. (1991). Bengali intonational phonology. *Natural Language & Linguistic Theory*, 47–96. http://link.springer.com/article/10.1007/BF00133326

Hermans, D., & van Gestel, J. C. (1991). The frequency scale of speech intonation. *The Journal of the Acoustical Society of America*, *90*(1), 97–102.

Ho, A. T. (1977). Intonation Variation in a Mandarin Sentence for Three Expressions: Interrogative, Exclamatory and Declarative. *Phonetica*, *34*(6), 446–457. https://doi.org/10.1159/000259916

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, *15*(3), 651–674.

Hyman, L. M. (2011). The Representation of Tone. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell Companion to Phonology*.

Hyman, L. M., & Monaka, K. C. (2008). Tonal and Non-tonal Intonation in Shekgalagar. *UC Berkeley PhonLab Annual Report*, *4*, 269–288. https://doi.org/https://doi.org/10.5070/P74429729r

Hyman, L. M., & Tadadjeu, M. (1976). Floating tones in Mbam-Nkam. *Studies in Bantu Tonology: Southern California Occasional Papers in Linguistics*, *3*, 59–111.

Jones, J., Louw, J., & Roux, J. C. (1998). Perceptual Experiments on Queclaratives in Xhosa. *South African Journal of Linguistics*, *16*(sup36), 19–32. https://doi.org/10.1080/10118063.1998.9724400

Ladd, D. R. (1996). *Intonational phonology*. Cambridge University Press.

Ladd, D. R. (2008). *Intonational Phonology* (2nd ed.). Cambridge University Press.

Lee, O. J. (2005). *The prosody of questions in Beijing Mandarin* [Doctoral Thesis, Ohio

State University]. http://rave.ohiolink.edu/etdc/view?acc_num=osu1122332580

Li, Q., Chen, Y., & Xiong, Z. (2017). Tianjin Mandarin. *Journal of the International Phonetic Association*, 1–20. https://doi.org/10.1017/S0025100317000287

Lin, M. (2006). Yiwen he chenshu yuqi yu bianjiediao [Interrogative vs. declarative and the boundary tone in Standard Chinese]. *Zhongguo Yuwen*, *4*.

Lin, M. (2004). On production and perception of boundary tone in Chinese intonation. *Proceedings of the International Symposium on Tonal Aspects of Languages*, 125–130. http://www.isca-speech.org/archive_open/tal2004/tal4_125.pdf

Lindau, M. (1986). Testing a model of intonation in a tone language. *Journal of the Acoustical Society of America*, *80*(3), 757–764. https://doi.org/10.1121/1.393950

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

Liu, M., Chen, Y., & Schiller, N. O. (2021). Context Matters for Tone and Intonation Processing in Mandarin. *Language and Speech*. https://doi.org/10.1177/0023830920986174

Lohfink, G., Katsika, A., Arvaniti, A., Sasha, C., Paola, E., Marija, T., & Paul, W. (2019). Variability and category overlap in the realization of intonation. *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*.

Luksaneeyanawin, S. (1998). Intonation in Thai. In D. Hirst & A. Di Cristo (Eds.), *Intonation Systems: A Survey of Twenty Languages* (pp. 376–394). Cambridge University Press.

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

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

Muñiz Cachón, C., Rodríguez, R. G., Gómez, L. D., & Pedrero, M. A. (2006). Prosodia gallego-asturiana en enunciaos SVO. *Revista de Filoloxía Asturiana*, *6*, 335–349.

Ni, J., & Kawai, H. (2004). Pitch targets anchor Chinese tone and intonation patterns. *Speech Prosody*, 95–98. http://www.isca-speech.org/archive_open/sp2004/index.html

Nolan, F. (2003). Intonational equivalence: an experimental evaluation of pitch scales. *Proceedings of the 15th International Congress of Phonetic Sciences*, 2–5. http://www.let.uu.nl/LOT/GraduateProgram/LotSchools/Summerschool2005/course descriptions/nolan_icphs2003.pdf

Ohala, J. J. (1983). Cross-language use of pitch: an ethological view. *Phonetica*, *40*(1), 1–18.

Ohala, J. J. (1984). An ethological perspective on common cross-Language utilization of F 0 of voice. *Phonetica*, *41*(1), 1–16. https://doi.org/10.1159/000261706

Phillips, N. D. (2017). Yarrr! The pirate's guide to R. *APS Observer*, *30*(3).

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation* [Doctoral Thesis]. MIT. Distributed 1988, Indiana University Linguistics Club.

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.r-project.org/

Reetz, H., & Kleinmann, A. (2008). Multi-subject hardware for experiment control and precise reaction time measurement. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 1489–1492).

Rialland, A. (2007). Question prosody: an African perspective. In T. Riad & C. Gussenhoven (Eds.), *Tones and Tunes: Typological Studies in Word and Sentence Prosody* (Vol. 1). Walter de Gruyter.

Rialland, A., & Robert, S. (2001). The Intonational System of Wolof. *Linguistics*, *39*(375), 893–940. https://doi.org/https://doi.org/10.1515/ling.2001.038

Shen, J. (1985). Beijinghua shengdiao de yinyu he yudiao [The register and intonation of Beijing Mandarin]. In T. Lin & L. Wang (Eds.), *Beijing Yuyin Shiyanlu [Reports of Beijing Phonetic Experiments]* (Vol. 73, pp. 1–21). Peking University Press.

Shen, J. (1992). Hanyu yudiao moxing chuyi (On Chinese intonation model). *Yuwen Yanjiu*, *45*, 16–24.

Shen, X. S. (1990). *The Prosody of Mandarin Chinese* (Vol. 118). Univ of California Press.

Shi, F. (1987). Tianjin fangyan danzidiao shiyan fenxi [Analysis of monosyllabic tones in Tianjin dialect]. In *Yuyinxue Tanwei [Research in Phonetics]*. Peking University Press.

Shi, P. (1980). Sizhong juzi de yudiao bianhua [The intonational change of four types of sentences]. *Yuyan Jiaoxue Yu Yanjiu*, *2*, 71–81.

Shih, C. (1988). Tone and intonation in Mandarin. *Working Papers of the Cornell Phonetics Laboratory*, *3*, 83–109.

Shih, C. (2000). A declination model of Mandarin Chinese. In A. Botinis (Ed.), *Intonation: Analysis, Modelling and Technology* (pp. 243--268). Kluwer Academic Publishers. http://link.springer.com/chapter/10.1007/978-94-011-4317-2_11

Wang, P., & Shi, F. (2010). Hanyu Beijinghua yiwenju yudiao de qifu du (The degree of rising and falling of Beijing Mandarin questions). *Nankai Linguistics*, *2*.

Welmers, W. E. (1959). Tonemics, Morphotonemics, and Tonal Morphemes. *General Linguistics*, *4*, 1–9.

Xu, B. R., & Mok, P. P. K. (2012). Cross-linguistic Perception of Intonation by Mandarin and Cantonese Listeners. *Proceedings of Speech Prosody*, 99–102.

Yip, M. (1989). Contour tones. *Phonology*, *6*(01), 149–174. https://doi.org/10.1017/S095267570000097X

Yuan, J. (2011). Perception of intonation in Mandarin Chinese. *The Journal of the Acoustical Society of America*, *130*(1), 4063–4069. https://doi.org/10.1007/s13398-014-0173-7.2

Yuan, J. (2012). The effects of speaking rate and intonation on the duration of tones in Mandarin Chinese. *Proceedings of the 6th International Conference on Speech Prosody (SP2012)*, 532–534.

Yuan, J., & Shih, C. (2004). Confusability of Chinese intonation. *Speech Prosody 2004, International Conference*. http://www.isca-speech.org/archive_open/sp2004/sp04_131.html

Zeng, X., Martin, P., & Boulakia, G. (2004). Tones and intonation in declarative and interrogative sentences in Mandarin. *TAL-2004*, 235-238. http://www.isca-speech.org/archive_open/tal2004/tal4_235.html

Zhang, C. (2018a). *Tianjin Mandarin Tones and Tunes* [Doctoral Thesis]. University of Oxford.

Zhang, C. (2018b). Chanted Call Tune in Tianjin Mandarin: Disyllabic Calls. *9th International Conference on Speech Prosody 2018*, 522–526. https://doi.org/10.21437/SpeechProsody.2018-106

Zhang, C. (2019). Stacking and Unstacking Prosodies : The Production and Perception of Sentence Prosody in a Tonal Language. *Proceedings of the 19th ICPhS*. https://icphs2019.org/icphs2019-fullpapers/pdf/full-paper_930.pdf

Zhang, J., & Liu, J. (2011). Tone sandhi and tonal coarticulation in Tianjin Chinese. *Phonetica*, *68*(3), 161–191. https://doi.org/10.1159/000333387