frontiers | Frontiers in Communication

ORIGINAL RESEARCH
published: 27 June 2022
doi: 10.3389/fcomm.2022.849132

Perception of Intonation on Neutral Tone in Mandarin

Yixin Zhang[1]*, *Elaine Schmidt*[1,2] and *Brechtje Post*[1]

[1] Phonetics Laboratory, Faculty of Modern and Medieval Languages and Linguistics (MMLL), University of Cambridge, Cambridge, United Kingdom, [2] Cambridge Assessment, University of Cambridge, Cambridge, United Kingdom

In Mandarin, lexical tone has been found to interact with intonational tone to influence intonation perception, with the falling T4 facilitating the perception of the statement/question contrast the most, and the rising T2 the least. However, in addition to the four citation tones T1-T4, Mandarin has "neutral tone" which marks weak, non-initial syllables that do not carry a citation tone. The prevailing view is that neutral tone is, in fact, phonologically toneless. It is unknown whether neutral tone can also affect intonation perception. However, it is reasonable to hypothesize that if neutral tone is indeed toneless, it cannot interact with intonational tone in the same way as citation tones do. We investigated this novel hypothesis with a perception experiment in which 22 Mandarin speakers had to determine whether disyllabic citation tone and neutral tone words were a question or statement. Results show that the identification of intonation contours is more accurate for neutral tone than for T2, and similarly accurate for neutral tone and T4, regardless of whether the neutral tone is intrinsic or derived. Furthermore, both T4 and neutral tone are realized with a reduced pitch range at a higher pitch level in questions, unlike T2, which is characterized by a slightly expanded pitch range and a higher pitch level. It is possible that intonation perception in Mandarin is facilitated by changes in the phonetic shapes of lexical tones brought by intonation rather than the phonological interaction between lexical tones and intonation. The importance of pitch changes to the intonation perception in Mandarin was further tested in a second perception experiment with the same 22 participants and disyllabic stimuli with manipulated pitch level and range. Results indicate that the use of pitch cues in intonation perception shows tone-specific differences, namely, pitch range is more important in signaling the question/statement contrast in utterances ending with T4 or neutral tone, while pitch level is the only perceptual cue to interrogativity for utterances ending in T2.

Keywords: neutral tone, Tone and intonation, lexical tone, intonation perception, Mandarin Chinese

OPEN ACCESS

Edited by:
Carlos Gussenhoven, Radboud University Nijmegen, Netherlands

Reviewed by:
Liquan Liu, Western Sydney University, Australia
Bettina Braun, University of Konstanz, Germany

***Correspondence:**
Yixin Zhang
yz510@cantab.net

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *05 January 2022*
Accepted: *18 May 2022*
Published: *27 June 2022*

Citation:
Zhang Y, Schmidt E and Post B (2022) Perception of Intonation on Neutral Tone in Mandarin. Front. Commun. 7:849132. doi: 10.3389/fcomm.2022.849132

INTRODUCTION

In tone languages, f_0 is used as the primary acoustic parameter for two important prosodic features, tone and intonation. At a lexical level, f_0 is employed to distinguish word meanings, and at an utterance level, it conveys intonational information such as discourse function (e.g., signaling questions or statements). Therefore, the realization of intonation in tone languages is more restricted than in non-tone languages. In tone languages, intonation is often realized through

a change in pitch register throughout the whole utterance, the insertion of boundary tones, register re-set, and/or through the suspension of downdrift (e.g., Shen, 1989, 1992b; Yuan et al., 2002). In other words, it seems that intonation interacts with rather than overrides lexical tones (Yip, 2002, p. 261). However, there are also tone-less elements in tone languages, like neutral tone in Mandarin. This raises the question addressed in this paper: How is intonation signaled when the syllables involved are phonologically toneless, and there are therefore no lexical tones to interact with it?

Neutral Tone in Mandarin

Mandarin Chinese is a tone language in which lexical tones are part of the phonological specification of morphemes, in addition to vowels and consonants. The majority of Mandarin morphemes are monosyllabic, and each either bears one of the four citation tones (CTs: T1, high-level tone, T2, mid-rising tone, T3, low-convex tone and T4, high-falling tone) or a neutral tone (NT). Syllables with NT are prosodically weak and cannot appear in word-initial positions or on their own, but must be attached to a syllable that carries a CT, whereby more than one NT-bearing syllable can be attached to the same preceding CT-bearing syllable. In this study, we focused on disyllabic words with a single NT. Henceforth, we will refer to words that contain a CT followed by an NT as "NT words". The f_0 realization of NT depends on the preceding CT: NT has a high-falling f_0 contour when following a high-level T1 or mid-rising T2, a high-level contour when following a low-dipping T3, and a mid or low falling contour when following a high-falling T4, and in addition, any following CT may also influence the realization of NT (Lin and Yan, 1980; Lin, 1983; Cao, 1986; Wang, 1996; Lee and Zee, 2014). However, some recent phonetic studies suggest that NT has a static mid target which is implemented with weak articulatory strength (Li, 2003; Chen and Xu, 2006).

From a morpho-phonological perspective, NT is not a homogeneous phenomenon either. Shen (1992a), for instance, proposed a three-way categorization of NT: toneless, detonic and atonic, based on their morphological status combined with their ability to be realized with CTs. Duanmu (2007, p. 248–250) instead identified NT as a stress phenomenon, categorizing Shen's toneless NT as associated with unstressed syllables, while all other tone-bearing syllables are stressed. Zhang (2018, 2021), by contrast, distinguishes two types of NT with different tonal representations but similar phonetic realizations in neutral utterances without narrow focus, based on a series of production and perception experiments: *Intrinsic NT* is carried by functional morphemes that have lost their etymological tone and is phonologically toneless; *Derived NT* is carried by notional morphemes which lose their CT on the surface in particular words when not in focus. In that account, a Derived NT is phonologically represented as the CT it is derived from in all its occurrences, regardless of its surface realization Thus, Derived NT is not phonologically toneless, unlike Intrinsic NT, and the two may therefore affect the production and perception of intonation in different ways. The only study investigating the realization of intonation on NT, however, focused exclusively on Intrinsic NT. It finds that like statements, questions are realized with a gradual f_0 declination when multiple NTs are pronounced in sequence at the end of an utterance, although the declination is not as steep as in statements. In contrast, the high-level T1s are realized with a slightly rising contour in questions (Liu and Xu, 2007). This suggests that in production at least, question intonation does not manifest itself more straightforwardly on NT than on CT. This raises two hitherto unanswered questions: (i) how are different intonation types realized on different types of NT, and (ii) how is intonation perceived on different types of NTs? This study focuses on the second question.

Intonation on Mandarin CTs

In Mandarin, two mechanisms for signaling question intonation have been identified for the final syllables of an utterance, an overall higher f_0 compared to statement intonation and a terminal rise. The implementation of these mechanisms is tone-dependent (Cao, 1986; Shen, 1989, 1992b; Yuan et al., 2002; Liu and Xu, 2005; Peng et al., 2005; Xu, 2005; Yuan, 2006, 2011). To be specific, in addition to raising their overall f_0, the high-level T1 becomes slightly rising, the mid-rising T2 and low-convex T3 have an expanded range, while the high-falling T4 is flattened as its final tonal target is raised (Yuan, 2004; Liu and Xu, 2005; Peng et al., 2005).

The perception of intonation in Mandarin has also been shown to be tone-dependent (Yuan, 2006, 2011; Ren et al., 2013). According to Yuan (2006, 2011), yes/no questions were easiest to identify in utterances ending with a falling T4, and hardest in utterances ending with a rising T2. In other words, the more saliently rising T2 was not necessarily interpreted as a question but led to greater bias toward statements. This finding, according to Yuan (2011), indicates that the phonological identity of tone "intervenes in the mapping of f_0 contours to intonational categories" (p. 19), and that hence tone and intonation interact at a phonological and linguistic level. Furthermore, in an electroencephalographic (EEG) oddball paradigm study using naturally produced monosyllabic stimuli which were controlled for differences in duration, Ren et al. (2013) found that the question-statement contrast elicits a clear mismatch negativity for T4-bearing syllables, but not for T2-bearing ones, indicating that the question-statement contrast is more salient on T4 than on T2. These findings suggest that the phonological identity of the utterance-final tone in a sentence determines the relative "ease" with which they are identified in perception.

However, in a more recent study, Liu et al. (2016) found that questions in utterances ending with T2 and T4 were equally difficult to identify while the identification of statements was difficult in sentences ending with T2 but not in those ending with T4. In other words, while the results of Liu et al. confirmed that there was tone-specific asymmetry in Mandarin intonation perception, they found it in statement perception. This is different from Yuan (2006, 2011) in which the asymmetry was found in question perception, because more questions on utterances ending with the rising T2 were misinterpreted as statements compared to utterances ending with the falling T4. An explanation suggested by Liu et al. (2016) to account for these potentially contradictory findings is that the realization of the question-final T4 used in Liu et al. (2016) differed from the f_0

contour used for T4 in Yuan's and Ren et al.'s studies in that the f_0 curve of the T4 in Liu et al. (2016) was not flattened as much as in the other two studies. This raises an alternative possibility that the ease of intonational perception in a tone language like Mandarin Chinese depends on the size of the difference between statement and question intonations of any given tone.

To summarize, if the phonological representations of lexical tones interfere with intonation perception as Yuan (2006, 2011) suggested, the perception of questions carried by Intrinsic NT syllables should differ from questions carried by CTs. Furthermore, the acoustic realization and interpretation of question intonation on phonologically toneless Intrinsic NT may also differ from phonologically specified Derived NT. We examine these possibilities in Experiment 1 by testing whether intonation is easier to perceive on intrinsic neutral tone because it is phonologically toneless than on CT and derived NT, because these are phonologically specified, at least in their underlying their forms. Following on from the findings of Experiment 1, Experiment 2 then investigates the relative contribution of different pitch cues to the perception of intonation type (in this case, question intonation).

EXPERIMENT 1

In Experiment 1, we investigated intonation perception (question vs. statement) in short utterances ending with Intrinsic NT, Derived NT, and T2 and T4 as the baseline citation form conditions.

H1. Since Intrinsic NT is phonologically toneless, the identification of intonation type for Intrinsic NT stimuli should be more accurate and faster compared to stimuli that are phonologically specified for tone (i.e., Derived NTs and CTs).

Methodology

Participants

Twenty-two Northern Mandarin speakers (6 males, 16 females) aged between 18 and 29 (mean age 23.7) participated in the experiment. All participants were current students at the Shanghai Jiao Tong University, Minhang Campus. None of the participants had lived in Shanghai for more than 3 years, as they all completed their pre-university education in the Huabei region and reported Northern Mandarin as the main language they used in school and at home. Therefore, the influence of the local Wu dialect in Shanghai on these participants is very limited. All of them were right-handed and none of them reported any hearing impairments. Informed consent was obtained prior to the experiment.

Stimuli

To test H1, we chose disyllabic stimuli in which the second syllables carried the target tone (i.e., Intrinsic NT, Derived NT, T2, or T4) and the first syllables carried T1. T1-T2 and T1-T4 words were chosen as the representative CT words, since previous studies found that intonation type was the hardest to identify in the case of utterances ending with T2 and the easiest in utterances ending with T4. The high-level T1 was chosen as the first syllable tone because it allowed for the most natural range of f_0 manipulations, and to keep the overall duration of the experiment to a reasonable time, no other CTs were used as the first syllable tone. For each of the four tone conditions, 32 stimulus words were used (in the Derived NT condition, 8 words were phonologically specified as T1+T1, 8 words as T1+T2, 8 words as T1+T3 and 8 words T1+T4), resulting in 128 items in total. Due to the limitation of the natural language, the second syllable of the items in the different tone conditions (Intrinsic NT, Derived NTs, T2 or T4) did not have the same segments, but the segmental complexity of the items (calculated by dividing the number of segments in the second syllable by the number of segments in the first syllable) was matched across the tone conditions. Furthermore, the NT items chosen here do not have minimal pairs carrying T2 or T4, and the T2 and T4 items did not exist in minimal pairs carrying NT (**Table 1**; **Appendix A**). Thirty-six disyllabic Mandarin words with T1 as the tone on the first syllable were added as fillers. Half of those had T1 as the second syllable tone and the other half had T3 as the second syllable tone.

The experimental items and the fillers were recorded by the first author, a native northern Mandarin speaker aged 27, and another female speaker of very similar age, education and language background. Both speakers hold the Level 1 (the top level) certificate of the National Mandarin Test for native speakers. Two speakers rather than one were recorded to ensure that participants could not just focus on the acoustic differences occurring within a single speaker's productions.

The recordings were made separately by each speaker in a quiet room with a Zoom H1 handy recorder at 96.000 Hz/26Bit. Stimuli that were not clear enough to allow for f_0 manipulation straightforwardly (e.g., due to co-articulation) were re-recorded. The naturalness of the stimuli was examined by the two speakers as well as a naive male northern Mandarin speaker by asking them to pick out the unnatural stimuli.

The recordings were then cross-spliced to neutralize the acoustic parameters of the preceding T1-bearing syllable between the two intonation conditions (declarative and interrogative) using Praat (Boersma and Weenink, 2021). For each recording, the initial syllable and second syllable were separated into two sound files. The pitch height, range and duration of the initial syllables were manipulated into an average pitch height, pitch range and duration of the statement and the question versions of the same stimulus word produced across all words by the same speaker. The second syllables were then spliced onto the manipulated initial syllable with the other intonation type, that is, the second syllables in statement intonation were attached to the initial syllables in question intonation of the same word. Equally, second syllables in question intonation were attached to the initial syllables in statement intonation of the same word and speaker. The intonation of the stimuli as discussed henceforth was determined by the intonation of the second syllable of the stimuli. The fillers were all manipulated in the same way as the stimulus words.

After cross-splicing, the average intensity of the stimuli was scaled to 75 dB. The digitally edited recordings were judged as natural by two native speakers who did not participate in the study. The stimuli were separated into two equal sets of

TABLE 1 | Examples of stimuli in each condition.

Tone	Word	Pinyin transcription	IPA transcription	Glossary
Intrinsic NT	鸽子	ge1zi0	/kɤ1 tsi0/	Pigeon
Derived NT from T1	孙家	sun1jia0(1)	/sʊn1 tɕia0 (1)/	The Sun's family
Derived NT from T2	敦实	dun1shi0(2)	/tun1 ʂi0 (2)/	Stoky
Derived NT from T3	家里	jia1li0(3)	/tɕia1 li0 (3)/	(At) home
Derived NT from T4	吃过	chi1guo0(4)	/tʂhi1 kuɔ (3)/	Have eaten
T2	清除	qing1chu2	/tɕhiŋ1 tʂhu2/	Delete
T4	捉住	zhuo1zhu1	/tʂuo1 tʂu4/	Get hold of

The numbers in Pinyin Transcription and IPA Transcription indicate tones (0 = NT) and the numbers in bracket indicate the phonological CTs of Derived NTs.

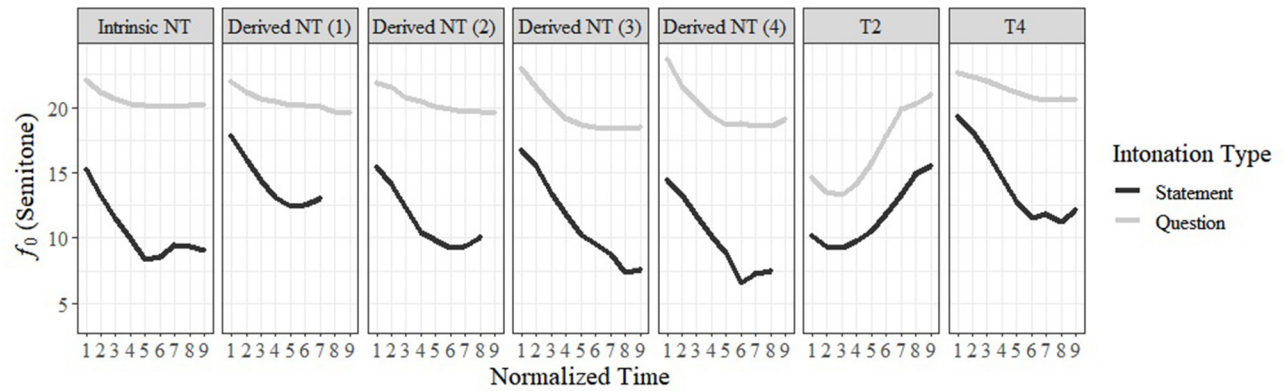

FIGURE 1 | Contours of the 2nd-syllable tone by tone and intonation (Numbers in brackets indicate the phonological tones of Derived NTs).

TABLE 2 | Average f_0 height and range of the second tones.

Tone	Intonation	f_0 Height (semitones)		Intonation	f_0 Range (semitones)	
		Average	SE		Average	SE
Intrinsic NT	Statement	14.96	0.06	Statement	8.65	0.06
	Question	20.5	0.02	Question	2.13	0.02
Derived NT (1)	Statement	13.08	0.38	Statement	11.43	0.72
	Question	20.41	0.23	Question	2.4	0.23
Derived NT (2)	Statement	12.33	0.39	Statement	11.46	0.62
	Question	19.62	0.32	Question	4.74	0.32
Derived NT (3)	Statement	11.81	0.33	Statement	12.72	0.62
	Question	19.86	0.31	Question	6.05	0.42
Derived NT (4)	Statement	11.3	0.32	Statement	7.79	0.45
	Question	20.47	0.17	Question	2.25	0.12
T2	Statement	11.61	0.01	Statement	6.75	0.02
	Question	16.4	0.02	Question	10.7	0.04
T4	Statement	14.43	0.02	Statement	10.45	0.04
	Question	21.35	0.02	Question	2.38	0.02

word-pairs with different intonations. Each set had half of the recordings of the stimulus words from one speaker who did the recording and the other half from the other. The order of the stimuli was pseudorandomized. Half the participants were tested with one set and half with the other.

To be better able to interpret the perception data, we conducted acoustic analyses of the stimuli after manipulation with Praat (Boersma and Weenink, 2021). Firstly, we analyzed the f_0s of the second syllables. A Praat script was applied to extract f_0 values (converted to semitones with 1 Hz as the reference value)

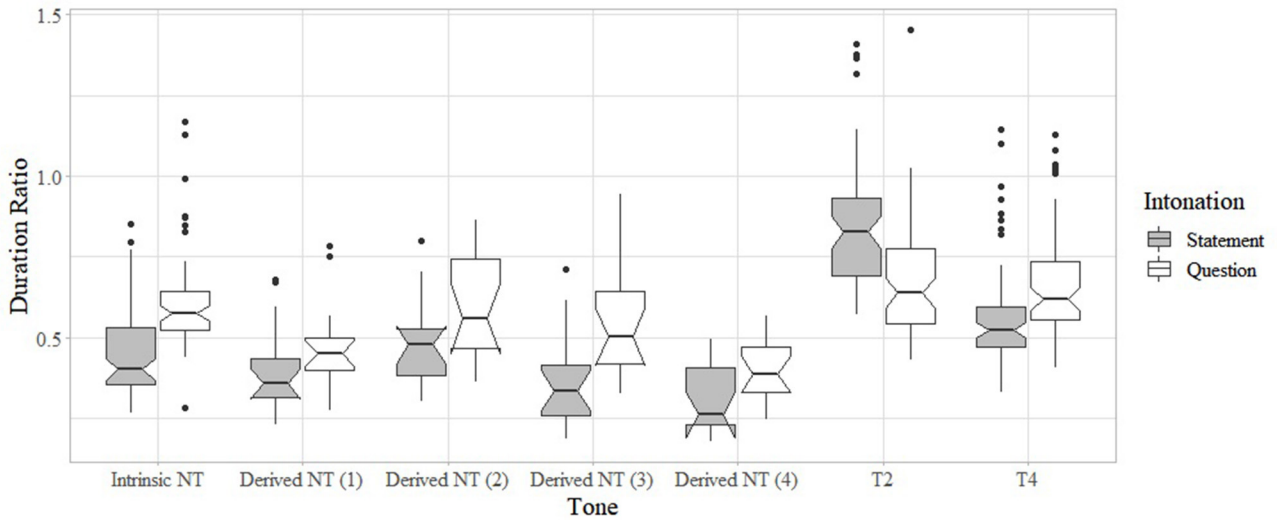

FIGURE 2 | Duration ratio by tone and intonation (Numbers in brackets indicate the phonological tones of Derived NTs).

TABLE 3 | Duration ratio and duration of the 2nd syllable by tone and intonation.

Tone	Intonation	Ratio		Intonation	Duration of the 2nd syllable (ms)	
		Average	SE		Average	SE
Intrinsic NT	Statement	0.46	0.00	Statement	181.52	1.95
	Question	0.61	0.01	Question	238.82	2.57
Derived NT (1)	Statement	0.40	0.03	Statement	242.26	20.19
	Question	0.47	0.04	Question	292.97	24.41
Derived NT (2)	Statement	0.48	0.04	Statement	278.78	23.23
	Question	0.60	0.05	Question	341.98	28.50
Derived NT (3)	Statement	0.36	0.03	Statement	203.86	16.99
	Question	0.54	0.05	Question	309.00	25.75
Derived NT (4)	Statement	0.31	0.03	Statement	200.13	16.68
	Question	0.40	0.03	Question	259.11	21.59
T2	Statement	0.84	0.01	Statement	355.09	3.95
	Question	0.68	0.01	Question	273.97	2.99
T4	Statement	0.57	0.01	Statement	235.17	2.53
	Question	0.67	0.01	Question	268.15	2.88

of the sonorous part in the second syllable of each stimulus. The f_0 contours were time-normalized by dividing the sonorous parts into 10 equal intervals, and f_0 values were extracted at each 10% step. Time-normalized rather than raw f_0-values were used to better illustrate any differences in pitch movement, range and height, as NTs and CTs differ in duration. The last value was excluded to reduce effects of final creakiness on f_0, and tokens with creakiness (i.e., no f_0 value extracted at a measure point) in more than 50% of the measured points were excluded from the f_0 analysis (thirty-eight statement Intrinsic NT and two statement T4 tokens; note that they were included in the perceptual experiment). We also analyzed the f_0 height and range (i.e., the difference between the minimum and maximum f_0 values) of the second syllables, and used linear-mixed effect (LME) models to evaluate the effects of Tone, Intonation, Speaker and their interactions on these two f_0 parameters. The model-building process is presented in detail in Section Data Analysis.

The f_0 contours realized on stimuli with <50% creakiness are illustrated in **Figure 1**. A clear difference between question and statement could be observed for all tones. **Figure 1** shows that question intonation raised the f_0 level in all tone stimuli, and that the range of the falling contour of Intrinsic NT was reduced. All other tones tested here show the same pattern except T2, which is only realized with raised pitch but slightly expanded range.

Analyses of average f_0 height and range confirmed these observations (**Table 2**). LME models showed that Tone, Intonation, Speaker, and the two-way interactions between them all had significant effects on the average f_0 height and range

TABLE 4 | Identification accuracy, hit rate (H, i.e., the identification accuracy of statement intonation), false alarm (FA), discriminability (A') and Bias (B''_D) in each tone condition.

Tone	Identification accuracy	Intonation	Identification accuracy	Hit rate (H)	False alarm (FA)	A''	B''_D
Intrinsic NT	95.48%	Statement	96.72%	96.72%	5.81%	0.98	0.29
		Question	94.19%				
Derived NT (1)	93.75%	Statement	95.45%	95.45%	7.95%	0.97	0.29
		Question	92.05%				
Derived NT (2)	92.90%	Statement	94.03%	94.03%	8.24%	0.96	0.17
		Question	91.76%				
Derived NT (3)	93.89%	Statement	94.89%	94.89%	7.10%	0.97	0.17
		Question	92.90%				
Derived NT (4)	94.03%	Statement	95.17%	95.17%	7.10%	0.97	0.20
		Question	92.90%				
T2	85.85%	Statement	86.77%	86.77%	15.00%	0.91	0.07
		Question	85.00%				
T4	93.17%	Statement	96.88%	96.88%	10.65%	0.96	0.57
		Question	89.35%				

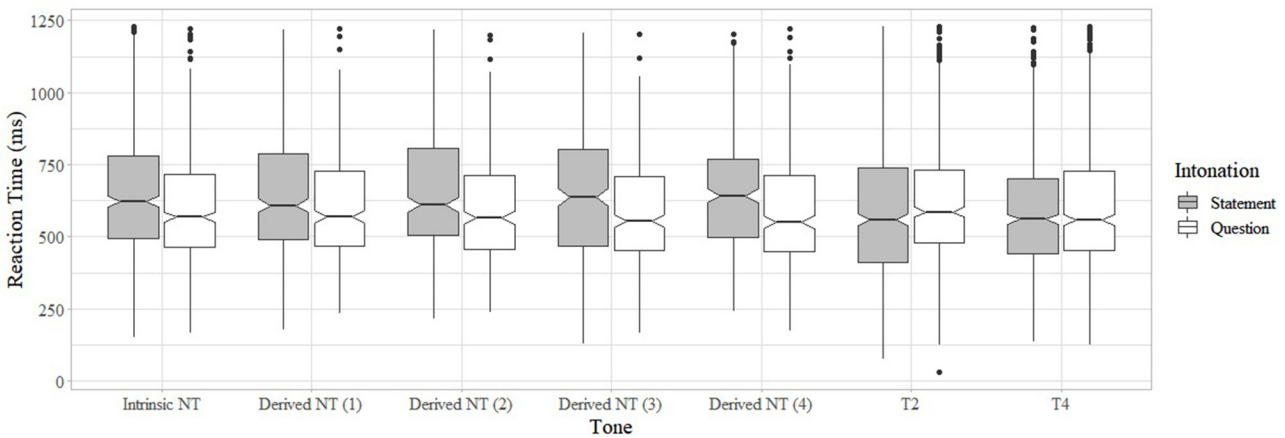

FIGURE 3 | Reaction time by tone and intonation (Numbers in brackets indicate the phonological tones of Derived NTs).

of the second tones, and the three-way interaction between Tone, Intonation and Speaker only affected average f_0 height ($ps < 0.0001$; for the full model, see **Supplementary Table 1** in **Appendix B**). Tukey *post-hoc* comparisons showed that the average f_0 height of questions was significantly higher than that of statements in all tone conditions ($ps < 0.001$). As to f_0 range, questions showed a significantly smaller f_0 range than statements in Intrinsic NT, Derived NT (3), Derived NT (4) and T4 ($ps < 0.001$). However, in T2, the pattern was reversed, namely, the pitch range for statements was significantly smaller than the pitch range for questions ($p < 0.001$). The interaction between Tone, Intonation and Speaker was significant due to the speakers consistently differing in their production of different tones in different intonation types. Since this is not relevant for our study, this will not be further presented.

Furthermore, the duration ratio for all stimuli in each tone and intonation condition was calculated (=duration of 2nd syllable/duration of the 1st syllable) and evaluated using an LME model. In general, we focused on the acoustic differences between the two intonation types within the same tone condition, rather than differences between tones with the same intonation, as different tones are already expected to have different f_0 and durational realizations.

In terms of duration ratio, the LME model showed that Tone, Intonation, Speaker, the interaction between Tone and Intonation, as well as the interaction between Tone and Speaker had significant effects on the duration ratio ($ps < 0.0001$; for the full model, see **Supplementary Table 2** in **Appendix B**). As can be seen in **Figure 2** and **Table 3**, stimuli with T2 also showed a reversed pattern to the other tones, namely, the duration ratio was smaller in T2 when it was a question, while it was larger when it was a statement all other tone conditions, and the same patterns were observed for the absolute duration of the 2nd syllable (all $ps < 0.001$).

TABLE 5 | Reaction time by tone and intonation.

Tone	Reaction time (ms)		Intonation	Reaction time (ms)		*Post-hoc* comparisons between statement and question
	Average	SE		Average	SE	
Intrinsic NT	626.76	6.06	Statement	652.72	12.07	$p < 0.005$
			Question	598.99	10.7	
Derived NT (1)	624.78	8.28	Statement	649.52	9.02	0.18
			Question	599.57	8.29	
Derived NT (2)	625.58	8.33	Statement	657.65	12.05	$p < 0.005$
			Question	592.25	10.9	
Derived NT (3)	617.95	8.32	Statement	648.76	12.65	$p < 0.05$
			Question	586.34	10.53	
Derived NT (4)	617.79	8.33	Statement	644.79	11.88	$p < 0.05$
			Question	589.85	11.84	
T2	602.28	6.63	Statement	583.44	10.1	0.14
			Question	620.05	8.53	
T4	593.83	6.23	Statement	588.46	8.8	0.99
			Question	599.81	8.86	

Procedure

The experiment was programmed in PsychoPy 3.0 (Peirce et al., 2019). Participants heard a manipulated recording of the stimuli (**Table 1**; **Appendix A**) while watching a screen on the experimental laptop which showed two horizontally arranged icons, "?" and "!" to record whether they heard a question or a statement.[1] Participants were asked to indicate their choice by pressing the keys on the keyboard labeled "?" or "!" after which the next trial automatically started. If no button was pressed within 3,000 ms, the next trial automatically started. The "?" was assigned to the left keyboard response button for half the participants and the right for the other half to avoid interference from handedness. The 42 trials with null results were treated as incorrect answers in the data analyses.

The experiment consisted of 256 test trials, 36 fillers and 32 trials of repeated words (324 trials in total) with two participant-controlled breaks available in between. The 32 trials of repeated words had randomly chosen words from the other stimulus set (eight words per tone condition) that appeared only once either in statement or in question intonation to prevent predictability, that is, the within-subject manipulation (i.e., each utterance is presented in both intonation conditions) may lead participants to choose the other response options for strategic reasons. Nine practice trials were given at the beginning to help participants familiarize themselves with the procedure. The whole experiment took about 40 min including instructions and practice trials.

[1]The symbol for a full stop "." was not used for statements to avoid the visual imbalance between "." and "?", and it was made clear in the instructions that "!" stood for "statement" rather than "exclamation". In Mandarin, "!" can be used to express strong emotions ranging from surprise, happiness to sadness and regrets in declarative sentences (The National Bureau of Quality Technical Supervision, 1996). This may have slowed down statement compared to question responses, but this is not relevant here, since such a bias would have applied across all tone conditions.

Data Analysis

Identification accuracy was calculated to measure how well an intonational function (i.e., statement vs. question) was recognized by the listeners. The response was considered accurate only if the intonation was identified as the same intonation that the speakers were asked to produce. The effects of Tone, Intonation and individual differences between Speakers on identification accuracy, a binary categorical variable (Accurate vs. Inaccurate), were evaluated by logistic mixed effects models, using glmer in the lmerTest package (Kuznetsova et al., 2017) in R (R Core Team, 2020). We assigned value 0 to Inaccurate and 1 to Accurate, and selected the optimal fixed structure by using stepwise comparisons from the most complex structure to the simplest and the optimal random effect structure according to the smallest Akaike Information Criterion (AIC). The anova() function served to compare different models to determine whether excluding factors from the analysis led to a better fit (Field et al., 2012). The details of the final models are presented in **Appendix B**.

According to Signal Detection Theory (see Macmillan and Creelman, 2004, for an introduction), identification involves not only the ability to discriminate between the two intonation conditions, but also the bias toward one of them in ambiguous situations. More specifically, Signal Detection Theory applies to the situation in which participants are asked to determine which of two categories (i.e., statement and question in our case) a stimulus belongs to. The task generates two measures of behavioral performance: the hit rate and the false alarm rate. In the present study, the response option of the statement was arbitrarily assigned to the signal, the question to the noise. Then, a hit (H) referred to when "the signal (statement) was presented and chosen" (i.e., the correct identification), a miss to when "the signal (statement) was presented but not chosen", a false alarm (F) to when "the noise (question) was presented but not chosen" and a correct rejection to when "the noise (question)

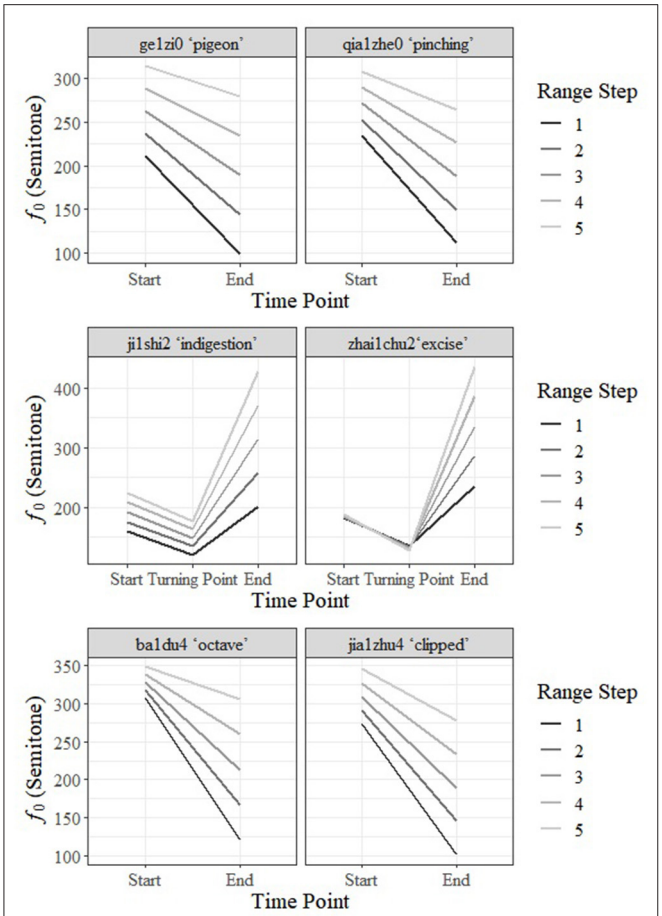

FIGURE 4 | Range steps of each stimulus at Height Step 1.

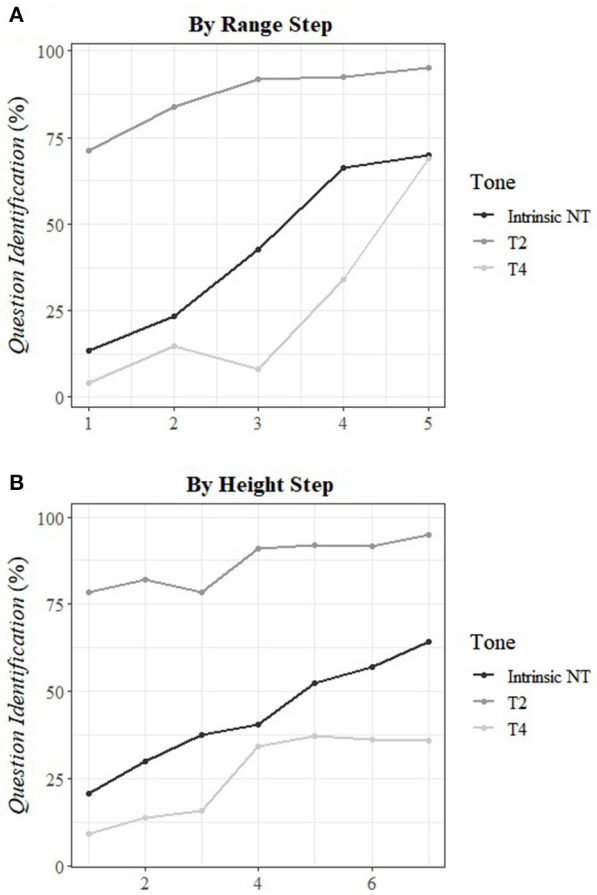

FIGURE 5 | Percentage of stimuli identified as questions (*question identified)* by Range Step **(A)** and Height Step **(B)**.

was presented and chosen". In studies using Signal Detection Theory, H and F are transformed into indices of identification sensitivity like A' based on statistical models, which indicates the discriminablility between the signal and the noise (Pollack and Norman, 1964; Smith, 1995; Zhang and Mueller, 2005). Calculated as (1), A' ranges between 0 and 1, 1 indicating maximum performance and 0.5 indicating chance performance (Zhang and Mueller, 2005, p. 207). The larger A' is, the better the perceptual result is.

$$A' = \begin{cases} 0.75 + \frac{H-F}{4} - F(1-H) & \text{when } F \le 0.5 \le H \\ 0.75 + \frac{H-F}{4} - \frac{F}{4H} & \text{when } F \le H < 0.5 \\ 0.75 + \frac{H-F}{4} - \frac{1\text{-}H}{4(1\text{-}F)} & \text{when } 0.5 < F \le H \end{cases} \quad (1)$$

The participants' response bias was indexed by B''_D, which correlates to the slope of the receiver operating characteristic function at the point of observation. B''_D was calculated following Pallier (2002) as (2) and ranges from −1 (maximum bias to the question) to 1 (maximum bias to statement). The absolute value of B''_D reflects the perceptual bias. The smaller it is, the better the perceptual result. We calculated A' and B''_D by tone condition.

$$B''_D = \frac{(1-H) \times (1-F) - H \times F}{(1-H) \times (1-F) + H \times F} \quad (2)$$

Reaction time (=the time of key-pressing minus the offset time of the auditory stimulus) was collected alongside as a measure of the difficulty of identifying intonation. Null results were excluded from the analyses here. Outliers in the reaction time data were removed following the Interquartile Rule (Tukey, 1977), and the effects of Tone, Intonation and Speakers on reaction time (a continuous numeric variable), were evaluated by linear mixed effect (LME) models. LME models were built through a similar process to the logistic mixed effect model but used lmer in the lmerTest package (Kuznetsova et al., 2017) in R (R Core Team, 2020). The details are presented with the results. To establish the LME models, the skewed data were transformed using square root transformation (Hothorn and Everitt, 2006).

TABLE 6 | Percentage of stimuli identified as question, by contour steps and height steps.

Tone	Contour step	Average (%)	SE (%)	Height step	Average (%)	SE (%)
Intrinsic NT	1 (Statement)	13.31	2.74	1 (The lowest)	20.68	7.29
T2		71.1	3.01		78.41	5.51
T4		4.06	2.16		9.09	4.81
Intrinsic NT	2	23.54	4.77	2	30	8.48
T2		83.93	3.83		82.05	6.75
T4		14.77	4.37		13.86	8.14
Intrinsic NT	3	42.7	7.48	3	37.5	12.28
T2		91.88	3.56		78.41	3.8
T4		8.28	2.93		15.91	9.04
Intrinsic NT	4	66.4	10.48	4	40.46	13.58
T2		92.53	3.28		90.91	3.39
T4		34.09	8.19		34.09	15.88
Intrinsic NT	5 (Question)	69.97	7.25	5	52.5	14.33
T2		94.97	2.22		92.05	5.01
T4		68.83	10.94		37.05	14.88
Intrinsic NT	-	6	57.04	12.47		
T2			91.59	4.97		
T4			36.14	16		
Intrinsic NT	-	7 (The highest)	64.09	15.13		
T2			94.77	4.14		
T4			35.91	17.77		

Results

Identification and Bias

The logistic regression model found that identification accuracy was significantly influenced by Tone ($p < 0.005$), Intonation ($p < 0.0001$), Speaker ($p < 0.0001$), and interactions between Tone and Intonation ($p < 0.01$) and between Speaker and Intonation ($p < 0.0005$) (for the full model, see **Supplementary Table 3** in **Appendix B**).

Tukey *post-hoc* comparisons showed that the identification accuracy of intonation for T2 was significantly lower than for all the other tones, namely, the four Derived NTs ($ps < 0.05$), Intrinsic NT ($p < 0.001$) and T4 ($p < 0.001$; **Table 4**). None of the accuracy differences between the other tones were significant, and neither were the differences among the four Derived NTs. When examined by intonation type, with regard to the identification of statement, the accuracy for T2 was significantly lower than that for Intrinsic NT ($p < 0.001$) and T4 ($p < 0.001$), but the other accuracy differences between tones were not significant. With regard to the identification of questions, the accuracy on T2 was significantly lower than that on Intrinsic NT ($p < 0.005$), while the differences between the other tones were not significant. To summarize, significant differences were mainly found between T2 and the other tones, especially between T2 and Intrinsic NT and T4, but not between the two types of NTs or between NTs and T4. Since there was no consistent difference in the identification of the same intonational contours produced by different speakers, the interaction between Intonation and Speaker is not relevant here and will thus not be further analyzed.

Discriminability and bias results showed that intonation on NTs and T4 was highly differentiable, more differentiable than intonation on T2 (**Table 4**). B''_D values were positive in all the conditions, suggesting that there was a bias toward statements, in line with previous findings (Yuan, 2006). However, the identification bias was larger in the T4 condition than in the other tone conditions, but smallest in the T2 condition.

Reaction Time

Significant effects of Intonation ($p < 0.001$), Tone ($p < 0.0001$) and the interaction between Tone and Intonation ($p < 0.0001$) on reaction time were found for reaction times (for the full model, see **Supplementary Table 4** in **Appendix B**). On average, the reaction time for question identification was significantly shorter than the reaction time for statement identification ($p < 0.001$). With regards to reaction time differences between tones, Tukey *post-hoc* comparisons showed that reaction time in the Intrinsic NT condition was significantly longer than that in T4 ($p < 0.01$), but the other differences between tone conditions were not significant.

When examined more closely, the reaction time differences between intonation types were only significant for Intrinsic NT ($p < 0.005$), Derived NT phonologically specified as T2 ($p < 0.005$), T3 ($p < 0.05$), and T4 ($p < 0.05$), but not for Derived NT phonologically specified as T1 or the two CTs (**Figure 3**; **Table 5**). When examined by intonation type, no significant differences were observed between tones with regard to question identification, but it took significantly longer to identify statements on Intrinsic NT and Derived NTs compared to T2 and T4 ($ps < 0.05$).

Discussion of Experiment 1

The present experiment examined the identification of intonation on Intrinsic NT and Derived NT in comparison to that of two CTs, the rising T2 and falling T4. The findings confirmed that intonation perception is easiest on T4 and hardest on T2, as has been found in previous studies (e.g., Yuan, 2011). However, the results did not confirm the hypothesis that the intonation type realized on Intrinsic NT is identified faster and more accurately than intonation on the other tones tested here, on the basis that it does not have phonologically specified tones that interact with intonation (H1). Instead, we found that the identification accuracy for Intrinsic NT was only higher than T2, but it was not significantly different from that for Derived NTs and T4. Moreover, there was similarly high discriminability (A') of intonation types in the Intrinsic NT, Derived NT and T4 conditions, higher than that in T2. In other words, Intrinsic NT patterned with the other falling tones (i.e., Derived NTs and T4) in accuracy, suggesting that the phonetic shape of the tonal contour provides the crucial explanatory information in tone-intonation interaction in Mandarin. There may be a ceiling effect in play as the identification accuracy in Intrinsic NT, Derived NT and T4 conditions was above 90% (Huang and Johnson, 2010).

In terms of identification bias, although all tones showed a bias toward statement interpretation in line with previous studies (e.g., Yuan, 2006, 2011; Liu et al., 2016), the largest bias was found for T4, and the smallest for T2. B''_D values in the NT conditions were all smaller than 0.3, and not comparable to the B''_D value of 0.57 in T4 ($B''_D = 0$, no bias; $B''_D = 1$, maximum bias to statement). It seems that although the identification accuracy for T4 was as high as for the NTs, it contained more bias toward statements. Also, although the intonation identification accuracy on T2 was more problematic, the smallest bias toward statements was found in the T2 condition, which indicates that the identification of question and statement were equally problematic, in line with Liu et al. (2016).

The bias results so far seemed to suggest that Intrinsic NT facilitates intonation perception in a more balanced way in comparison to T4. It is possible that in the absence of a phonological interaction between lexical tone and intonation in phonologically toneless Intrinsic NT syllables, intonation can somehow be better accommodated than in syllables with phonologically specified lexical tones. However, this interpretation fails to explain the low B''_D values found for Derived NTs. Derived NTs seem to have phonological tones which are assumed to interact with intonation just like T2 and T4. Moreover, we found that Derived NT phonologically specified as T4 did not enable a higher identification accuracy nor a less biased identification than Derived NTs with the other phonological tones, which also indicates that intonation perception for Derived NTs is not affected by the phonological identity of the tone. Note that high B''_D value found in T4 may in fact be affected by the fact that it was on a small number of misses and false alarms (Stanislaw and Todorov, 1999; Zhang and Mueller, 2005).

The reaction time results did not support H1 either. Normally, we would expect higher accuracy and shorter reaction times to indicate ease of identification, as was the case for T4 vs. T2, but against the hypothesis, intonation identification took significantly longer for NTs than T4. It is possible that the participants found it harder to identify the NT stimuli due to their weak surface realization and short duration, diminishing the salience of the relevant perceptual cues. It is also possible to attribute this finding to the very short duration of the NT-bearing syllables and the statements in general, because the key-pressing process will always need a certain amount of time.

Taken together, these findings show that any interaction that may take place at a phonological level between intonation and lexical tones cannot account for the intonation identification data analyzed here. Instead, the facilitative effect observed for T4 as opposed to T2, as well as the absence of a significant difference between T4 and both types of NT suggest that it is, in fact, the surface f_0 pattern that is of crucial importance here. More specifically, unlike T2, T4 and both types of NT all have a falling contour which is raised and flattened under question intonation, which makes their surface realizations quite unlike their realization in statement contexts. T2 also shows a slight shift in range and height, but otherwise, the contour is identical in the two intonation conditions.

EXPERIMENT 2

In Experiment 1, despite tone-specific differences, the raising of f_0 to signal question intonation was clearly found across all tone conditions as well as a changed f_0 range due to a further raising of the utterance-final targets (see **Figure 1** above). Liang and Heuven (2009) used a sentence made up of seven syllables carrying high-level T1 to investigate the relative weighting of these two cues in intonation perception. By manipulating the overall f_0 height of the utterance and the terminal f_0 height of the final syllable, they established that the f_0 rise in the utterance-final tone was a more important cue to question intonation than the overall height of the utterance. What Liang and Heuven (2009) could not fully investigate by using T1 syllables only is the potential effects of individual lexical tones on the perpetual cues to intonation type, especially the changed f_0 range of the utterance-final tones. Since question intonation changes the surface contour of question-final tones in a tone-specific manner, how cues to questions are weighted in perception may also vary between different utterance-final tones. More specifically, we hypothesized that:

H2: The change in pitch range is more important to the perception of intonation type on tones with falling contours (i.e., Neutral tone and T4) while changes in both pitch range and height are important cues to intonation perception in the rising T2.

Methodology

Participants

The same 22 participants that participated in Experiment 1 also took part in Experiment 2. Experiment 2 took place about 2 months after Experiment 1.

Stimuli

Two Intrinsic NT words, two T2 words and two T4 words from Experiment 1 were recorded by the first author as representative NT and CTs, and cross-spliced as in Experiment 1. We then manipulated the duration of the first and the second syllables of the recordings of the same stimulus word into the average duration of the two intonation conditions and scaled the intensity of the recordings at 75 dB. Then, we systematically manipulated the pitch height of the disyllabic stimuli and the pitch range of the second syllables using Praat (Boersma and Weenink, 2021). For each stimulus word, we created 3 height steps and 3 range steps with equal intervals between the question and the statement version, and also added 2 more extra height steps (i.e., one higher than the question and one lower than the statement). Height step 1 is the lowest and 7 is the highest while range step 1 is the statement range and 5 is the question range. For pitch height, to simplify the manipulation, we calculated the average f_0 of all six stimuli and rounded the number to create intervals. The average f_0 of the statement stimuli across tone conditions was 289.96 Hz (SE = 7.65 Hz), about 70.24 Hz lower than that of the question stimuli, 360.20 Hz (SE = 6.59 Hz). Therefore, we set Height step 1 of all stimuli at 245 Hz, and Height step 7 at 380 Hz, with an equal interval of 22.5 Hz in between. The manipulation of range at Step 1 is illustrated in **Figure 4**.

We manipulated the stimuli starting from both the statement and question recordings, resulting in 70 manipulations (5 range steps * 7 height steps * 2 source recordings) for each stimulus and 420 stimuli in total (70 steps manipulations * 2 stimulus words * 3 tones). Forty-eight stimuli from Experiment 1 were added as fillers without manipulation.

Procedure

The experiment was programmed in PsychoPy 3.0 (Peirce et al., 2019) and the procedure was the same as in Experiment 1 except that this time, no time limit was set for key-pressing, though participants were encouraged to give their answers as quickly as possible. This was because during piloting, participants reported that they were distracted by trying to observe the time limit. The participants took part in this experiment in a quiet room. The experiment consisted of nine practice trials which were the same as in Experiment 1, 420 experimental trials and 48 filler trials which were pseudo-randomized with two 5-min breaks. The whole experiment took about 40 min including instructions and practice trials.

Data Analysis

The analysis focused on intonation identification (i.e., question or statement). A binominal ordinary logistic regression model was first established to evaluate whether and how Tone, Pitch height, Pitch range and Original intonation (i.e., manipulated from the original recording of the statement or the question version) affected identification (Question vs. Statement) in each tone condition. Since the complexity of the model influences the degree of uncertainty (Babyak, 2004), we further split the data by Tone, and for each tone condition, a binominal ordinary logistic regression was established to evaluate the effects of Pitch height, Pitch range and Original intonation, and their interactions. The models were established through a process similar to the models established in Experiment 1 using glm in the lmerTest package (Kuznetsova et al., 2017) in R (R Core Team, 2020), and *post-hoc* comparisons were carried out using Tukey tests. The percentage of stimuli identified as questions (henceforth *question identification*) was also calculated by dividing the number of questions chosen by the total stimulus number in each tone and intonation combination in Intrinsic NT, T2, and T4 conditions to enable a visual description of the results.

Results

The binominal ordinary logistic regression model established on the whole dataset showed that Tone, Pitch height, Pitch range and the interactions between all variables (except Pitch range × Original intonation) had a significant effect on the identification of question intonation ($p < 0.0001$; for the full model see **Supplementary Table 1** in **Appendix C**). Tukey *post-hoc* comparisons showed that the perceptual results for Intrinsic NT (43.13% trials identified as question), T2 (86.88% trials identified as question) and T4 (26.01% trials identified as question) all differed significantly from each other ($ps < 0.0001$).

Binominal ordinary logistic models by tone condition demonstrated that in all three tone conditions, the effects of both Pitch height and Pitch range on intonation identification were significant ($ps < 0.0001$; for the full models and the interactions between the variables, see **Supplementary Table 2** in **Appendix C**). Moreover, in the Intrinsic NT condition, Original intonation also had a significant effect ($p < 0.0001$). Increases in the average pitch height as well as the manipulation of the range (to the question intonation) both led to more questions identified in all tone conditions, but the specific effects showed tone-specific patterns (**Figure 5**; **Table 6**).

When the **pitch range** of the stimuli became more question-like (i.e., step number increased, illustrated in **Figure 4**), more stimuli were perceived as a question, regardless of lexical tone. However, since there was already a high preference to question identification in the T2 condition at Range Step 1 (i.e., the statement range), the increase in *question identification* brought by changes in range in the T2 condition was restricted compared to the other two conditions. Specifically, differences in intonation identification were significant between the first 3 steps that were more statement-like (i.e., 1 vs. 2, 1 vs. 3 and 2 vs. 3, $ps < 0.0005$) but not the last 3, more question-like range steps (i.e., 3 vs. 4, 3 vs. 5 and 4 vs. 5). Nevertheless, at each range step, T2 stimuli were interpreted as a question more often than Intrinsic NT and T4 stimuli ($ps < 0.0001$).

In the Intrinsic NT and T4 conditions, while the question-like manipulation of pitch range led to a much larger increase in *question identification* than in the T2 condition, the trajectories were different. In the Intrinsic NT condition, there was a steady increase in *question identification* with the range steps. Tukey *post-hoc* comparisons demonstrated that the differences in intonation identification between all range steps in the Intrinsic NT condition were significant ($ps < 0.0001$) except that between Range Step 4 and 5, that is between the most question-like range and the question range. In the T4 condition, a steady increase in *question identification* was observed at Range

Step 4 and 5, but not for the first three more statement-like contours. It is worth mentioning, however, that the differences in intonation identification between all contour steps were statistically significant in the T4 condition ($ps < 0.005$). Although *question identification* at Range Step 2 was larger than that at Range Step 3, they were all far lower than chance, suggesting that slightly flattened contours still led to a preference for a statement interpretation in the T4 condition. The identification differences between Intrinsic NT and T4 at Range Step 3 and 4 were also significantly different ($ps < 0.005$).

The increase in **pitch height** also led to more stimuli being identified as a question, and again the increase was much larger in the Intrinsic NT and T4 conditions than in the T2 condition. Tukey *post-hoc* comparisons showed that in the Intrinsic NT condition, except for differences between adjacent steps (i.e., 1 vs. 2, 2 vs. 3, 3 vs. 4, 5 vs. 6, and 6 vs. 7), the identification differences between height steps were all statistically significant ($ps < 0.005$). In general, an increase in pitch height led to a gradual increase in *question identification* in the Intrinsic NT condition. In contrast, in the T4 condition, significant differences in intonation identification were found between Height Step 1 and Height Steps 4–7 ($ps < 0.01$), Height Step 2 and Height Steps 4–7 ($ps < 0.0001$) and Height Step 3 and Height Steps 5–7 ($ps < 0.0001$), but not within the higher height steps, namely, steps 4–7. In the T2 condition, from Height Step 3, the increasing pitch height seemed to play a predominant role, leading to over 90% *question identification*. The identification difference was significant between Height Step 1 and Height Steps 4–7 ($ps < 0.0001$), Height Step 2 and Height Steps 4–7 ($ps < 0.005$), and Height Step 3 and Height Steps 5–7 ($ps < 0.0001$). Again, at each height step, question interpretations were more frequent in the T2 condition than in the Intrinsic NT and T4 conditions ($ps < 0.0001$), and at Height Step 2, 3, 6, and 7, question interpretations were more frequent in the Intrinsic NT condition than in the T4 condition ($ps < 0.0005$).

Discussion of Experiment 2

In Experiment 2, we examined the influence of changes in pitch height and range, and the relative weighting of these cues in intonation perception in Mandarin, using disyllabic Intrinsic NT, T2, and T4 stimuli with T1 as the preceding tone. Although both pitch height and range played important roles in intonation perception, a general effect of the lexical tone on the weightings of the two cues in intonation perception was observed. In general, Intrinsic NT and T4, which were phonetically realized as similar falling contours, showed more similarity to each other than to rising T2. Pitch range played a more important role in question identification in Intrinsic NT and T4, while an increase in pitch height was the primary cue to question intonation in T2. This means that H2 is largely confirmed, though pitch range was a less important cue on T2 than expected. Since the difference in pitch range in T2 was quite small, this finding is not surprising. Unlike the results of Experiment 1, a clear preference for question interpretations was found in the T2 condition in the present experiment, regardless of range or height steps. F_0 manipulation only significantly influenced intonation perception at the first several range and height steps, namely, the more statement-like steps in the T2 condition. From Range Step 3 and Height Step 4 upwards, the percentage of stimuli identified as questions (*question identification)* became higher than 90%, which could be indicative of a ceiling effect. It also possible that the intervals between height steps were not large enough, but further enlargement of the intervals would have made the stimuli sound unnatural, as Height Step 7 in the present study already sounded very high and Height Step 1 very low. This marked preference for question interpretations in the T2 condition may be due to the lack of durational cues in this experiment, or simply, because a rising contour is interpreted as more question-like than other contours. In contrast, a preference for statement interpretations existed in the other two tone conditions, especially in the T4 condition (overall 56.82% in Intrinsic NT and 73.99% in T4), but it was not as strong as the preference for question interpretations in the T2 condition (overall 86.88%). In other words, intonation identification appeared to be especially difficult in utterances ending with a T2 rising contour. On the one hand, the expanded range of the final T2 and the rise in the overall pitch height both led to more stimuli identified as questions. On the other hand, though, the participants were not as sensitive to the more question-like manipulation of f_0 range in the T2 condition as in the other two conditions.

In both the Intrinsic NT and T4 conditions, pitch range played a more important role in intonation perception such that the pairwise identification differences between all range steps reached statistical significance, except Intrinsic NT Range Step 4 vs. 5. Nevertheless, the perception in the Intrinsic NT and T4 conditions also showed some interesting differences. The effects of question-like range manipulation and height were relatively gradual on *question identification* in the Intrinsic NT condition, but showed a sudden rise at step 4 of both the height and range manipulations in the T4 condition. Moreover, it seemed that a stronger flattening of the falling contour was required for a T4 word to be identified as a question than a NT word, as the facilitating effects were more clearly observed in the last two contour steps for T4, while they were already observed at lower steps for NT. In addition, raising pitch height alone did not lead to any preference for question interpretations in the T4 condition. Even when presented with the highest pitch step, participants still tended to identify T1-T4 words as statements unless the falling contour of T4 was reduced at its end, resulting in a reduced f_0 range. Therefore, although f_0 range played a rather important role in intonation perception in the two tone conditions with a phonetically falling movement, it seemed to weigh more heavily as a cue in the T4 condition than in the Intrinsic NT condition. At the same time, other subtle acoustic cues that we did not consider in the present experiment that were hidden in the original recordings (e.g., spectro-temporal differences in sonorous segments between statement and question, see for instance, Coath et al., 2005) might have played a role in the Intrinsic NT condition, since identification in the Intrinsic NT condition was affected by the version of the stimuli that were used for manipulation. In other words, participants showed more sensitivity to more types of cues in the Intrinsic NT condition than in the CT conditions.

To sum up, the findings of Experiment 2 suggested that in short utterances, changes in both overall pitch height and pitch range realized on utterance-final syllables were important cues in question intonation identification in Mandarin. However, the latter cue seemed to weigh more heavily in the identification of intonation for lexical tones that were realized as falling contours than for rising T2, probably because the combined cues made the difference between statements and questions particularly salient, while the intonational contrast is primarily signaled by a difference in height in the T2 condition.

GENERAL DISCUSSION

The present study investigated the perception of intonation type (question vs. statement) on Mandarin NT in comparison to representative CTs, the mid-rising T2 and the high-falling T4. Although Intrinsic NT and Derived NT differ from each other in their phonological representations on some accounts (e.g., Zhang, 2021) as discussed in the introduction, Experiment 1 showed that intonation identification in these two conditions was as highly accurate as in the T4 condition, which was significantly more accurate than that in the T2 condition. The acoustic analyses of the stimuli in Experiment 1 revealed that question intonation was always marked by a higher overall f_0 level, but this was accompanied by a decrease of the f_0 range for the falling contours of NTs and T4 (manifested as a higher ending of the falling contour), while the T2 rise only changed in that it became slightly steeper. Experiment 2 showed that pitch range was the most important cue in the T4 condition and also important in the Intrinsic NT condition, while pitch height played a role in the T2 condition. These results confirm that the reduced f_0 range on the surface plays a more important role in intonation perception in Mandarin NT words than any possible tone-intonation interaction at a phonological level. The present findings shed light on the intonation perception mechanisms in Mandarin as well as the phonetic targets of both types of NT.

That the phonetic tonal realization can be modified to such an extent may be due to the relatively simple tonal system of Mandarin. The flattening of the falling tone or the raising of a level tone would hardly lead to misidentification of lexical tones (Liu et al., 2016), but can be used to alert the listener that there is an intonational event going on. In syllable tone languages with a more complex tonal system like Cantonese, the effect of intonation on the realization of lexical tones often leads native listeners to misidentify them as other lexical tones rather than facilitating intonation perception (Kung et al., 2014). This difference in complexity may also explain why Cantonese (and middle ancient Chinese which had 4 tonal contours and 2 tonal registers) has retained a much richer inventory of modality particles than modern Mandarin, and why they are not reduced to the NT-bearing syllables of Mandarin.

The acoustic analysis of stimuli in Experiment 1 also showed that, despite their difference in phonological tonal representation, both types of NT surface with similar phonetic forms in declarative utterances. In addition, both Intrinsic NT and Derived NT maintain a slightly falling contour in questions as short as disyllabic words, suggesting that they may share a unique phonetic target, namely, a mid static target according to Chen and Xu (2006) that is different from the tonal targets found for the four CTs.

To conclude, the investigation of intonation perception on different types of NT in the present study allows us to attribute the tone-specific pattern found in Mandarin to the phonetic realization rather than the phonological interaction between lexical tone and intonation. It would be interesting to investigate to what extent our findings generalize to preceding tones other than T1 and other intonations, or longer utterances.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the Faculty of Modern and Medieval Languages and Linguistics (MMLL), University of Cambridge. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YZ, ES, and BP contributed to the conception of the study, experimental design, and manuscript preparation. YZ carried out the experiment and data analysis. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

YZ would like to thank the CHINA Scholarship COUNCIL (CSC) and Cambridge Trust for their doctoral scholarship.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2022.849132/full#supplementary-material

REFERENCES

Babyak, M. A. (2004). What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychos. Med.* 66, 411–421. doi: 10.1097/00006842-200405000-00021

Boersma, P., and Weenink, D. (2021). *Praat: Doing Phonetics By Computer [Computer program]*. Version 6.1.51. Available online at: http://www.praat.org/ (accessed July 22, 2021).

Cao, J. F. (1986). Acoustic features of Neutral Tone in Standard Mandarin. *Appl Phonetics* 4, 1–6.

Chen, Y., and Xu, Y. (2006). Production of weak elements in speech – evidence from f0 patterns of neutral tone in standard Chinese. *Phonetica* 63, 47–75. doi: 10.1159/000091406

Coath, M., Brader, J. M., Fusi, S., and Denham, S. L. (2005). Multiple views of the response of an ensemble of spectro-temporal features support concurrent classification of utterance, prosody, sex and speaker identity. *Network Comput. Neural Syst.* 16, 285–300. doi: 10.1080/09548980500290120

Field, A., Miles, J., and Field, Z. (2012). *Discovering Statistics Using R*. Great Britain: Sage Publications, Ltd, 958.

Hothorn, T., and Everitt, B. S. (2006). *A Handbook of Statistical Analyses Using R*. Boca Raton: CRC Press.

Huang, T., and Johnson, K. (2010). Language specificity in speech perception: Perception of Mandarin tones by native and nonnative listeners. *Phonetica*. 67, 243–267.

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

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. *J. Statist. Software* 82, 1–26. doi: 10.18637/jss.v082.i13

Lee, W.-S., and Zee, E. (2014). "Chinese phonetics," in *The Handbook of Chinese Linguistics*, eds C. J. Huang, Y. A. Li, and A. Simpson (Hoboken: John Wiley & Sons), 367–399.

Li, Z. (2003). *The phonetics and phonology of tone mapping in a constraint-based approach*. (Doctoral dissertation). Massachusetts Institute of Technology.

Liang, J., and Heuven, V. J. (2009). "Chinese tone and intonation perceived by L1 and L2 listeners," in *Experimental Studies in Word and Sentence Prosody* (De Gruyter Mouton), 27–62.

Lin, M. C., and Yan, J. Z. (1980). Acoustic features of Neutral Tone in Mandarin. *Dialects* 3, 166–178.

Lin, T. (1983). "A primary test on neutral tones in Beijing Mandarin," in *Phonetic experiments on Beijing Dialect*, eds T. Lin and L. J. Wang (Beijing: The Peking University Publishing House), 1–26.

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

Liu, F., and Xu, Y. (2007). "The neutral tone in question intonation in Mandarin," in *Eighth Annual Conference of the International Speech Communication Association*.

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

Macmillan, N. A., and Creelman, C. D. (2004). *Detection Theory: A User's Guide*. New York, NY: Psychology Press.

Pallier, C. (2002). *Computing Discriminability and Bias With the R Software*. Available online at: http://www.pallier.org/ressources/aprime/aprime (accessed April 22, 2022).

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., et al. (2019). PsychoPy2: experiments in behavior made easy. *Behav. Res. Methods*. 51, 195–203. doi: 10.3758/s13428-018-01193-y

Peng, S. H., Chan, M. K., Tseng, C. Y., Huang, T., Lee, O. J., and Beckman, M. E. (2005). "Towards a Pan-Mandarin system for prosodic transcription," in *Prosodic Typology: The Phonology of Intonation and Phrasing*, ed S. A. Jun (OUP Oxford), 230–270. doi: 10.1093/acprof:oso/9780199249633.003.0009

Pollack, I., and Norman, D. A. (1964). A non-parametric analysis of recognition experiments. *Psychonomic Sci.* 1, 125–126. doi,: 10.3758/BF03342823

R Core Team (2020). *R: A Language and Environment for Statistical Computing*.Vienna: R Foundation for Statistical Computing. Available online at: https://www.R-project.org/ (accessed April 22, 2022).

Ren, G. Q., Tang, Y. Y., Li, X. Q., and Sui, X. (2013). Pre-attentive processing of Mandarin tone and intonation: evidence from event-related potentials. *Funct. Brain Mapp. Endeav. Understand Working Brain* 6, 95–108. doi: 10.5772/56503

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

Shen, X. (1989). *The Prosody of Mandarin Chinese*. Berkeley: University of California Press, 9–30.

Shen, X. S. (1992b). Mandarin neutral tone revisited. *Acta Linguistica Hafniensia* 24, 273. doi: 10.1080/03740463.1992.10412273

Smith, W.D. (1995). Clarification of sensitivity measure A. *J. Math. Psychol.* 39, 82–89. doi: 10.1006/jmps.1995.1007

Stanislaw, H., and Todorov, N. (1999). Calculation of signal detection theory measures. *Behav. Res. Methods Instr. Comput.* 31, 137–149. doi: 10.3758/BF03207704

The National Bureau of Quality and Technical Supervision (1996). *National Standard of the People's Republic of China: The Usage of Punctuation*. Beijing: Standards Press of China.

Tukey, J. W. (1977). *Exploratory Data Analysis*. 2, 131–160. Available online at: http://theta.edu.pl/wp-content/uploads/2012/10/exploratorydataanalysis_tukey.pdf (accessed June 17, 2022).

Wang, J. (1996). "An acoustic study of the interaction between stressed and unstressed syllables in spoken Mandarin," in: *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96 (Vol. 3)*. Philadelphia: IEEE, 1616–1619.

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

Yip, M. (2002). *Tone*. Cambridge, MA: Cambridge University Press.

Yuan, J. (2004). "Perception of Mandarin intonation," in *2004 International Symposium on Chinese Spoken Language Processing* (Paris: IEEE). 45–48.

Yuan, J. (2006). "Mechanisms of question intonation in Mandarin," in *International Symposium on Chinese Spoken Language Processing*. Berlin, Heidelberg: Springer, 19–30.

Yuan, J. (2011). Perception of intonation in Mandarin Chinese. *J. Acoust. Soc. Am.* 130, 4063–4069. doi: 10.1121/1.3651818

Yuan, J., Shih, C., and Kochanski, G. P. (2002). "Comparison of declarative and interrogative intonation in Chinese," in *Speech Prosody 2002, International Conference* (Aix-en-Provence).

Zhang, J., and Mueller, S. T. (2005). A note on ROC analysis and non-parametric estimate of sensitivity. *Psychometrika* 70, 203–212. doi: 10.1007/s11336-003-1119-8

Zhang, Y. (2021). *Neutral tone in Mandarin: representation and interaction with utterance-level prosody*. (Doctoral dissertation). University of Cambridge, Cambridge, United Kingdom.

Zhang, Y. (2018). "Anticipatory dissimilation in (non-clitic) neutral tones in Mandarin," in: *Tone and Intonation in Europe 2018, International Conference* (Stockholm).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.