Chinese tone and intonation perceived by L1 and L2 listeners

Liang, J.; Heuven, V.J. van; Gussenhoven C., Riad T.

Chinese tone and intonation perceived by L1 and L2 listeners[1]

Jie Liang

Vincent J. van Heuven

Phonetics Laboratory
Leiden University Centre for Linguistics
The Netherlands

Submitted to *Tone and Intonation,* 2005

Published in C. Gussenhoven & T. Riad (Eds.), *Tones and Tunes, Volume 2: Experimental studies in word and sentence prosody.* Vol. 12-2. Phonology and Phonetics (pp. 27-61). Berlin/NewYork: Mouton de Gruyter, 2007.

Authors' address

Phonetics Laboratory, Leiden University Centre for Linguistics
Cleveringaplaats 1, PO Box 9515
2300 RA Leiden, The Netherlands

phone: +31 71 527 2319
fax: +31 71 527 2615
e-mail: j.liang@let.leidenuniv.nl
v.j.j.p.van.heuven@let.leidenuniv.nl

Abstract

Pitch is used in Chinese both at the word level to differentiate between four (or more) lexical tones and at the sentence level. A functional view claims that if some phonetic dimension is exploited in one area of the grammar (including the phonology) it will not be used to the same extent in another part of the phonology. We addressed the issue by asking how well learners of Chinese as a second language (L2) perceive pitch variations at word and sentence-levels. We assume that if there is competition between these two uses of pitch, we should expect L2 speakers whose native language (L1) is a tone language to be more sensitive to lexical tones and less sensitive to sentence intonation than L2 speakers whose L1 is a non-tone language.

Eighty-one Chinese-speaking (Beijing dialect) subjects participated in two laboratory-controlled perception experiments, on lexical tone and sentence intonation (statement vs. question), in which we manipulated the intonation contour of an entire utterance as well as that of the final syllable. We recruited one group of L1 listeners (native Beijing speakers) and three groups of L2 listeners, i.e. two groups with a tone-language L1 (Nantong and Changsha dialect) and one group with non-tone language L1 (Uygur, an Altaic language).

Results revealed that in perceiving tones and intonation in Beijing dialect, relative to L1 performance, L2 listeners with a non-tone-language L1 (Uygur) proved less sensitive to lexical tones but were at the same time more sensitive to intonation than L2 listeners with a tone-language L1. These results support the functional view, i.e., if some phonetic dimension is exploited in one area of the grammar, e.g. lexical tones, it will not be used to the same extent in another part of the phonology, e.g. sentence intonation.

1. Introduction

Pitch is used in Chinese both at the word level to differentiate between four (or more) lexical tones and at the sentence level to signal – among other things – differences in sentence type such as statement versus question. Cross-linguistically there is a clear tendency for languages to make a choice for either word tones or sentence intonation. That is to say, if a language such as Chinese has a substantial inventory of word tones, it will exploit the pitch dimension less for the signalling of contrasts at the sentence level. In this connection, it is interesting that it has often been observed that Chinese possesses a sentence-final question particle (*ma*), which may signal interrogativity along with any pitch differences. This, then, is a functional view that claims that if some phonetic dimension is exploited in one area of the grammar (e.g. the phonology), it will not be used to the same extent in another part of the grammar (see, for instance, Ross, Edmondson, and Seibert 1986; Gandour et al. 1995; Seddoh 2002).

There is a substantial body of experimental results on the question how well speakers of non-tone languages, such as English and Dutch, learn to speak and recognize lexical tones, mostly in Mandarin – which is the most widely spoken (and taught) tone language in the world. Results typically indicate that non-tone learners have poorly defined cross-overs in tone identification tasks, and show considerable inter-speaker variability in toneme boundaries (e.g. Leather 1987). There have been virtually no studies addressing the question of native-language interference in the acquisition of lexical tones by speakers of another tone language. A very recent literature survey (Wang, Jongman, and Sereno 2005) mentions two studies that touch upon this question, but only tangentially. Gandour (1983) studied perceptual similarity between pairs of natural tones in Mandarin, Cantonese, Taiwanese and Thai, but the task did not involve the perceptual identification of tones. Similarily, Vakoch and Wurm (1996) presented natural tokens of Mandarin and Cantonese tones to Mandarin, Cantonese and English listeners in a discrimination task. Their results indicate that English listeners discriminate the tones less successfully than listeners who speak a tone language. We are not aware of studies addressing tone identification by non-native learners with or without a tone-language L1.

Our questions therefore are as follows:

(1) How well do learners of Chinese as a second language (L2) perceive pitch variations at the word level and at the sentence-level in comparison with L1 listeners of Chinese?
(2) Are L2 learners of Chinese differentially sensitive to linguistic levels (lexical tone versus intonation)?
(3) Are L2 listeners with a non-tone L1 less sensitive to lexical tones but more sensitive to sentence intonation than L2 listeners with a tone L1?

In our study, we set up two laboratory-controlled perception experiments on lexical tone and sentence intonation (statement vs. question), respectively, in which we manipulated the intonation contour of the whole utterance as well as the (non-lexical) boundary tone in the sentence-final syllable. As for subjects, we recruited 30 Beijing dialect speakers as the control sample, and compared their performance with that of four groups of non-native learners of Beijing dialect (Mandarin). The L2 groups differed in terms of native-language type (tone vs. non-tone language). Moreover, within the group of learners with a tone-language background a further division was made depending on the complexity of the native tone system (four, six or seven-tone inventory). This set-up will allow us to distinguish between several possibilities. We may find that all listeners show the same behavioral pattern irrespective of their L1 background. Alternatively, results may differ depending on the L1 type (tone versus non-tone) and/or the complexity of the tone system. If the latter results should obtain, the hypothesis that lexical tone and intonation perception compete for resources may be tested. We predict that, relative to L1 listeners, learners with a tone-language L1 should be more sensitive to lexical tones and less to intonation than non-tone L1 listeners.

2. Methods

We set up an experiment with native language, stimulus type and task as the principal factors. Each factor will be explained in a subsection.

2.1. Native language.

Our study involved four groups of listeners, who all had the Chinese nationality but differed in their native language.

(1) Listeners of the **Beijing** dialect, which is very similar to the Standard language (Mandarin). The 30 Beijing native speakers were aged between 21 and 70, average 40, 17 male and 13 female. All of them had at least twelve years of normal education.
(2) Listeners speaking the seven-tone **Nantong** dialect of Chinese. The eight Nantong dialect listeners were construction workers, aged 22−48, who had just spent their first year in Beijing and learned Beijing dialect through natural exposure to the language.[2]
(3) Listeners speaking the six-tone **Changsha** dialect of Chinese. The thirteen Changsha listeners were first-year students at Hunan University, aged 19−21, who also intended to take the HSK[3], which test is required for teacher certification in China. They had learned Standard Chinese either through formal training or self-study.
(4) Listeners living in the People's Republic of China speaking **Uygur**, a Tungusic (Altaic) non-tone language which is not related to Chinese. The thirty Uygur listeners were students of Xinjiang Normal University, aged 18−20[4], who had had just finished a one-year intensive training course in Standard Chinese (Beijing dialect) and intended to pass the HSK, which is obligatory for them if they wish to continue their higher education in China.

In total, eighty-one Chinese volunteers participated. All of them were right-handed and reported no speech or hearing disorders. Composition of the listener groups is summarized in Table 1.[5]

Table 1. Listener types.

Language family		Language	Number of word tones	Number of subjects
Sinitic	Mandarin	Beijing	4	30
		Nantong	7	8
	Xiang	Changsha	6	13
Altaic	Turkic	Uygur	0	30

2.2. Stimuli

Two types of stimuli were used in the experiments for the lexical (word tone) as well as for the post-lexical (intonation) part. Recordings used in the experiments were from a male native speaker of Beijing dialect (Standard Chinese), experienced in sound recording. All the recordings were made in a sound-insulated booth at the Phonetics Laboratory, Leiden University, on digital audio tape (DAT) using a Sennheiser MKH-416 unidirectional microphone, then transferred to computer memory and downsampled to 16 kHz (16 bit amplitude resolution) using the Praat speech processing package (Boersma and Weenink 1996).

For the lexical part, four frequent lexical words were used, covering the Beijing four-tone system, /ma^1 ma^2 ma^3 ma^4/ ‘mother, hemp, horse, scold’, respectively. In connected speech, some of the lexical tones may change their tonal category in context. For example, the low tone in the Beijing dialect changes into a rising tone when followed by another low tone. When that happens, the derived rising tone is perceptually indistinguishable from the lexical rising tone (Wang and Li 1967). This kind of categorical tone shift due to tonal context is usually referred to as ‘tone sandhi’. In order to control the context effect, the lexical-tone stimuli were recorded by reading aloud a list of the four words in a carrier sentence and in isolation, respectively, four times for each. In all there were 40 stimuli for lexical tone identification, i.e. 4 (tone) × 5 (repetition) × 2 (in a carrier sentence and isolation).

For the post-lexical (intonation) part, a single question utterance was used: /kai^1 sun^1 jiŋ^1 kʰai^1 fei^1 tɕi^1/, ‘must Sun Ying fly flying machine?’ meaning ‘Is it Sun Ying’s turn to fly the plane?’ The basic utterance contains high level tones (Tone 1) only, in order to reduce the lexical tonal influence. The sentence can be interpreted as either a statement or a question depending on the intonation. Only tokens with question intonation were elicited. The speaker was instructed to ‘say’ the sentence written with a question mark as naturally as possible.

Studies have shown that intonation is perceived in terms of specific fundamental frequency (F0) patterns (Garding and Abramson 1965; ‘t Hart, Collier, and Cohen 1990), which involve the same acoustic parameter that is employed in the signaling of lexical tone (Pike 1948; Wang 1972; Howie 1976). Previous acoustic and perception studies on question intonation of Standard Chinese have shown that besides the use of question particles, the main acoustic correlate of the statement~question contrast is the fundamental frequency (Howie 1976; Tseng 1981; Coster and Kratochvil 1984). Since our sentences did not contain any lexical (including particles) and/or syntactic question markers, they are optimally suited for studying the role of intonation in signalling the difference between statement and question.

Using the Praat software the pitch contour of one question utterance was stylized with seven pivot points connected by straight lines in a log-frequency (semitone, ST) by linear time representation. It is generally accepted that the pitch variation in the final syllable signals the statement ~ question contrast in Standard Chinese (Tsao 1967; Gårding 1985, 1987; Kratochvil 1998; Yuan, Shih, and Kochanski 2002) but there is disagreement on the issue whether the overall pitch level of the utterance provides a subsidiary cue to the contrast (De Francis 1963; Kratochvil 1998) or not (e.g. Tsao 1967). Kratochvil (1998), in fact, described the question intonation of Beijing Dialect as accompanied by a raising of the overall pitch level by about 50 Hz. We perceptually tested the relative contributions of the overall pitch and of the final fall/rise to the statement ~ question contrast. To this effect, a two-dimensional stimulus continuum was generated (through PSOLA resynthesis, cf. e.g. Moulines and Verhelst 1995) from the question utterance, by systematically combining seven overall pitch levels and seven different boundary tones. The overall pitch levels were decremented by 1-ST steps, which adequately covered the range we established in the production of question and statement tokens by our speaker. The seven boundary tones were generated by decrementing the terminal pitch of the question token in six steps of .5 ST. Again, the 3.5-ST difference between the highest and lowest terminal frequencies in the boundary tone continuum corresponds closely to what we found in our speaker's production. By manipulating only the frequency values while leaving the time coordinates unaltered, 49 different pitch patterns were generated according to the following schema (Figure 1).

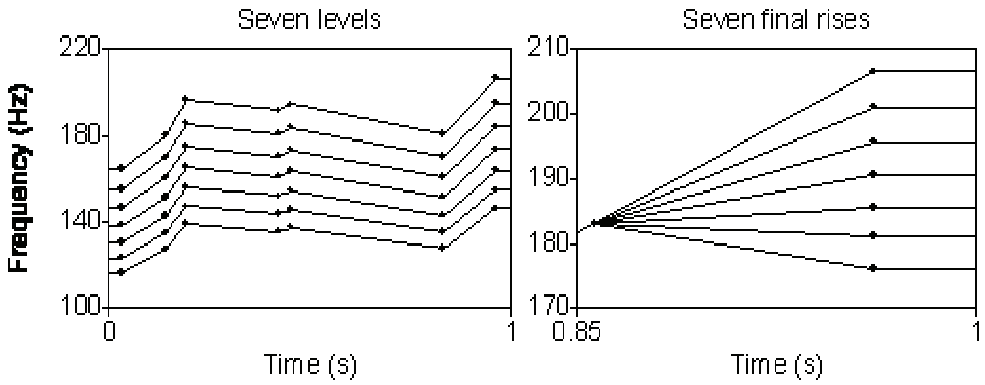


Figure 1. Overview of the manipulations of pitch level (7 steps) and boundary tone (7 steps). Orthogonal variation of level and boundary tone yields 49 different stimulus types

2.3. Tasks and procedure

Considering that lexical tone stimuli contained naturally produced tokens only, we ran the two experiments in a fixed order, such that the lexical-tone experiment always preceded the intonation test (with potentially ambiguous, hybrid exemplars). The stimuli were randomized and presented by computer binaurally over headphones (Sony MDR-V3) at a comfortable listening level. Listeners were tested one at a time in a quiet room. A specially designed keyboard was used with either four buttons marked with the corresponding Chinese characters (in font 72) for the tone identification, or two buttons marked with '。' and '?', i.e. the punctuation marks that are used in Chinese dialects for statement and question.[6] Before each part of the experiment specific instructions were given both orally and in writing. Each experimental task was preceded by a short practice session, with four practice trials for the tone identification task and seven for the intonation task. Each experiment lasted about 20 minutes. Decisions made and reaction times (measured from the offset of the stimulus, precision 1 ms) were stored in the computer. Stimulus presentation and response collection were controlled by E-prime software.[7]

In the **lexical tone identification experiment**, the subjects decided which of four words they had heard by pressing one of the four buttons on the response keyboard each time they heard a stimulus. In the first stage of the experiment they were asked to avoid errors (no time pressure). In the second stage the subjects were instructed not only to avoid errors but also to perform the task as quickly as they could manage (time pressure). The four buttons were evenly spaced across the top row on the response box; a black key for continue/start was located in the center of the bottom row. The 40 lexical-tone stimuli, 20 in a carrier sentence and 20 in isolation were presented to the listeners twice in two blocks (once without time pressure and a second time with time pressure).

When there was no time pressure, there was a fixed 3000-ms (isolated targets) or 5000-ms (targets in carrier) inter-stimulus interval (ISI) after the offset of the stimulus, irrespective of the reaction time. When no response was given within the ISI, the subject timed out, and the next stimulus was presented. In the sections with time pressure imposed on the listener, the next stimulus started 1000 ms (in isolation) or 2000 ms (in a carrier sentence) after the response. The shorter ISI in the time-pressure condition prompted the subjects to speed up their reaction time. In the debriefing after the experiment, subjects confirmed that they had felt pressured; the effects of pressure are also apparent from the results: not only were reaction times much faster, but also was the number of timed-out responses higher (even though the ISI had not been reduced).

In the **intonation experiment**, the listeners were required to choose between statement and question by pressing one of the two buttons marked with '。' or '?' each time they heard a stimulus. As before, the subjects performed their task without and with time pressure. The 49 stimuli were presented twice in two blocks in random order, i.e. once without and once with time pressure. Every stimulus was preceded by a fixed reference sound, which was a single token of the neutral vowel (schwa) spoken by the same speaker who recorded the stimuli, and PSOLA resynthesized at 119 Hz, which is the lowest pitch used in the experiment, i.e. the onset pitch of the lowest resynthesized version.[8] The reference precursor preceded the test sentence by 500 ms (offset to onset). In the absence of time pressure, there were fixed 6000-ms ISIs between stimuli (offset to onset), irrespective of the reaction time. When time pressure was imposed, the next stimulus was presented 3000 ms after the previous response.

3. Results

In this section, we report the results obtained with the types of listener for lexical tone identification and post-lexical identification of sentence type. We will first analyze the data collected for the perception of the lexical tone contrast. In the second part we will complement this with the corresponding analysis of the sentence intonation contrast. It should be noted beforehand that these two parts are rather different conceptually. In the matter of the lexical tones, we can determine whether the listener correctly identified the tone pattern as it was intended by the speaker of the utterance. Also, we may determine how much time it took the listener to correctly decide on the tonal category. Since the stimuli were presented for identification once with and once without emphasis on speed of response (time pressure), we predict that listeners traded accuracy for speed when time pressure was on, i.e., were prepared to gamble in the case of ambiguous stimuli in order to gain speed. In the identification of sentence melody (question ~ statement), all the stimuli were derived through parametric manipulation from a single question utterance. There is no pre-given norm here of what would be a statement and what counts as a question. So in this section of the data we will just analyze the percentage of 'question' responses and the decision latencies; we expect longer latencies as the choice between the response alternatives is more evenly balanced (which would be a sign of ambiguity in the stimulus). It is unclear, in the identification of sentence melody, how time pressure should affect the quality of the response.

3.1. Lexical tone identification

A four-way repeated measures analysis of variance (RM-ANOVA) with type of listener (Beijing, Nantong, Changsha and Uygur) as a between-subject factor, and with stimulus tone (Tone 1 to Tone 4), context (in carrier sentence vs. in isolation) and time pressure (on vs. off) as within-subject factors was carried out on the lexical tone responses. It shows that, overall, **percent correct** identification was significantly different for the four types of listener [$F(3, 77) = 65.4$], for context [$F(1, 77) = 24.1$], and for tone [$F(3, 223) = 7.9$] ($p < .001$ in all cases). Significance was also found for the interactions of type of listener × context, $F(3, 77) = 40.5$ ($p < .001$), and of listener type × tone [$F(8.7, 223) = 12.5$] ($p < .001$).[9] Scheffé post-hoc tests divided the listener type into two groups, i.e. (i) Uygur, (ii) Nantong, Beijing and Changsha.

A similar four-way RM ANOVA was carried out on the reaction time of the lexical-tone responses. It shows that, overall, reaction time was significantly different for the various types of listener [$F(3, 77) = 38.5$ ($p < .001$)], context [$F(1, 77) = 57.1$ ($p < .001$)], time pressure [$F(1, 77) = 416.3$ ($p < .001$)] and for tone [$F(2.9, 225.8) = 4.2$ ($p = .006$)]. Significant effects were found for second-order interactions of listener type × context [$F(3, 77) = 19.4$ ($p < .001$)], listener type × time pressure [$F(3, 77) = 24.9$ ($p < .001$)], context × time pressure [$F(1, 77) = 28.9$ ($p < .001$)], context × tone [$F(3, 231) = 9.0$ ($p < .001$)], time pressure × tone [$F(2.9, 226.6) = 4.5$ ($p = .005$)], and listener type × tone [$F(8.8, 225.8) = 2.4$ ($p = .014$)]. Also, three-way interactions proved significant for listener type × context × time pressure [$F(3, 77) = 13.7$ ($p < .001$)]. Scheffé post-hoc tests divided the listener type into two groups: (i) Changsha, Beijing and Nantong; (ii) Uygur.

Correct tone identification (percent) and the associated reaction time (ms) are presented in Figure 2 as a function of tone type broken down by the four types of listener.

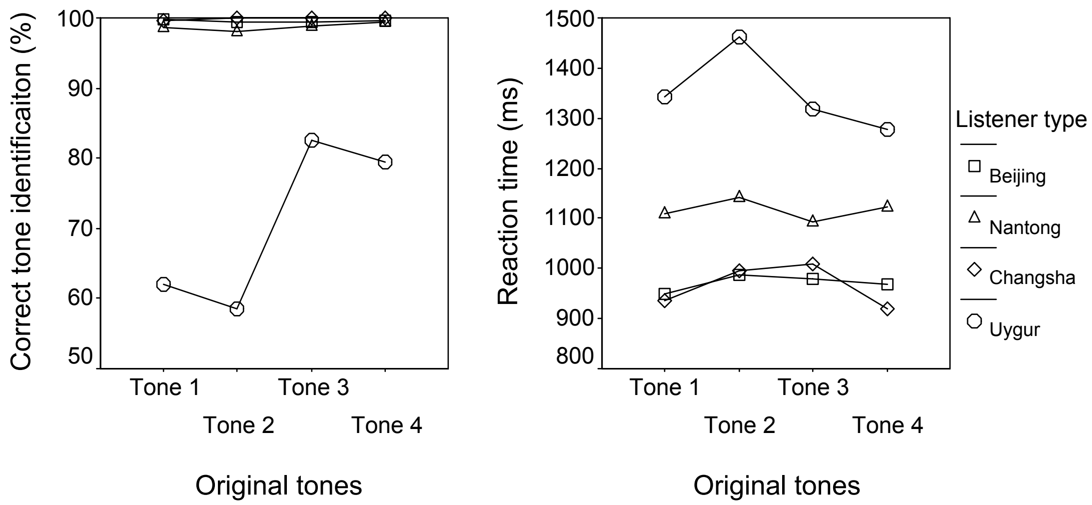


Figure 2. Correct tone identification (percent) and associated reaction time (ms) as a function of tone type broken down by four types of listener.

Figure 2 shows that the Uygur group was much poorer in tone identification, and much slower in their reaction time relative to the tone-language groups. In order to have a clear picture of the distribution of these data, we also plotted percent correct by the four speaker groups against the associated reaction time broken down by presence versus absence of time pressure for tones in carrier sentences and in isolation in Figure 3.

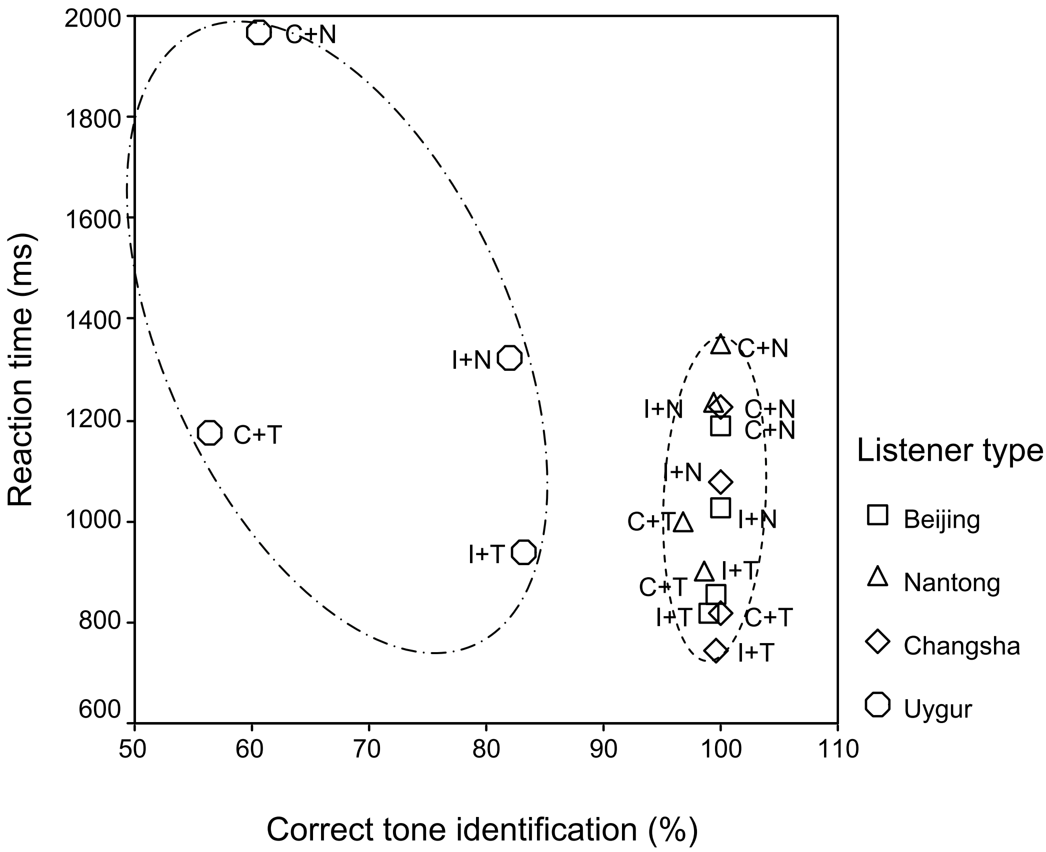


Figure 3. Percent correct identification of lexical tone plotted against reaction time (ms) for four types of listener (differing in native language background) in two time-pressure conditions (no time pressure 'N' versus time pressure 'T') and two contexts (in carrier sentence 'C' versus isolation 'I'). Ellipses were drawn around the centroids by hand.

Figure 3 shows the following effects:

(i) **Language background**. Uygur non-tone-language listeners (dash-dotted line) are separate from the tone-language groups.

(ii) **Time pressure**. In each type of listener time pressure leads to considerably shorter reaction time but has little effect on percent correct. Post-hoc tests above show that the Uygur groups were significantly different from the tone-language groups, while the three tone-language groups were not significantly different from each other.

(iii) **Context**. All listeners took longer when identifying tones in the carrier sentence than in isolation. Importantly, however, embedding the target in a carrier did not affect percent correct for any tone-language group, but it clearly reduced percent correct for the non-tone-language Uygur listeners.

From the above findings, we conclude that the contexted targets created a problem for Uygur listeners, such that their overall percent correct tone recognition is substantially poorer and takes much more time than that of the tone-language listeners.

As a last stage in the analysis of lexical tone identification let us break down the results by the four separate tones. Figure 4 plots percent correct identification for each of the four tones broken down by listener type. The presentation is broken down further into four panels, one for each combination of context (carrier sentence, isolation) and time pressure (with, without).

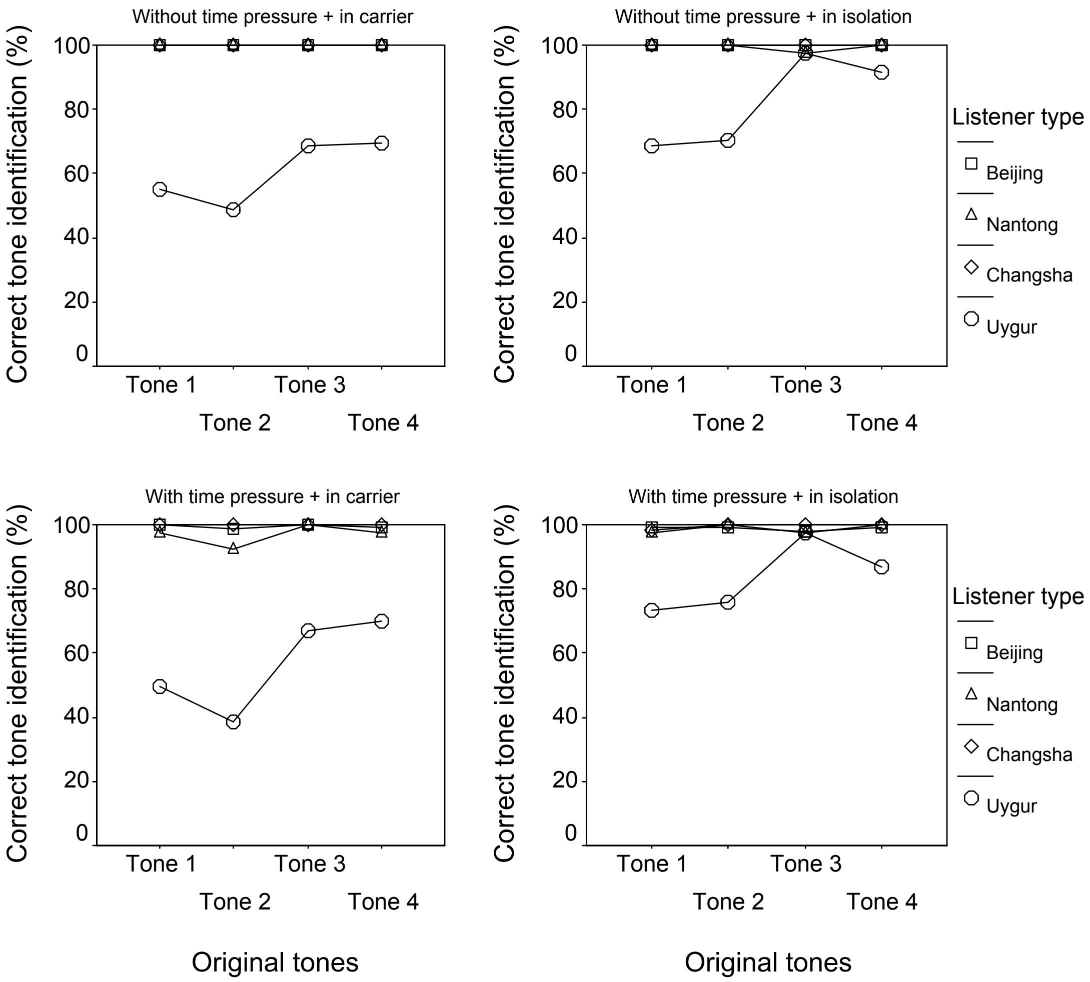


Figure 4. Correct tone identification (percent) as a function of tone type broken down by four types of listener: one for each combination of context (in carrier sentences, isolation) and time pressure (with, without).

From Figure 4 we can see the following:

(i) Beijing and Changsha listeners identified the tones almost perfectly with no less than 98.5 percent correct.
(ii) The Nantong group was less perfect, with 92.5% correct Tone-2 identification.
(iii) The Uygur group was the poorest, with percent correct ranging from 40 to 98.

Clearly, then, being a native speaker of some other Chinese tone language affords a listener to perform as well in the Beijing tone identification task as the native listeners of the Beijing dialect themselves. There is no point in analyzing the perceptual confusion structure for these three groups of listeners, as their level of performance is nearly perfect. However, a tone-confusion analysis makes sense for the Uygur listeners. Therefore, table 2 presents the lexical-tone confusion for Uygur listeners broken down by context (carrier vs. isolation) and time pressure.

Table 2. Confusion matrix of tone identification by Uygur listeners broken down by presence versus absence of time pressure for tones in carrier sentences and in isolation. Bold numbers are correct responses.

Listener type	Time pressure	Tones	Responses							
			in carrier sentence				in isolation			
			Tone 1	Tone 2	Tone 3	Tone 4	Tone 1	Tone 2	Tone 3	Tone 4
Uygur	No	Tone 1	**55.2**	24.8	4.8	15.2	**68.5**	21.5	5.4	4.7
		Tone 2	23.3	**48.6**	17.1	11.0	22.1	**70.5**	1.3	6.0
		Tone 3	16.0	12.0	**68.7**	3.3		2.0	**97.3**	0.7
		Tone 4	5.5	17.2	7.6	**69.7**	2.7	4.7	1.3	**91.3**
	Yes	Tone 1	**49.6**	30.2	12.2	7.9	**73.3**	16.7		10.0
		Tone 2	38.6	**38.6**	7.6	15.2	18.1	**75.8**		6.0
		Tone 3	8.1	21.5	**67.1**	3.4	1.3		**97.3**	1.3
		Tone 4	6.0	10.7	13.4	**69.8**	4.7	8.7		**86.7**

There is considerable confusion for Uygur listeners, especially when the tones were presented in carrier sentences. The largest confusion occurred between Tone 1 and Tone 2: Tone 1 tokens were misidentified as Tone 2 in 30% and 25% with and without time pressure, respectively, while 39% and 23% of Tone 2 were misidentified as Tone1.

In order to come to an overall characterization of the performance of the four listener groups, we determined Receiver Operating Characteristic (ROC) curves for each individual listener in each of the four groups (Beijing, Nantong, Changsha and Uygur) with and without time pressure, and broken down further for presence versus absence of context. Since ROC analysis applies to dichotomies only, the responses were analyzed in terms of four binary oppositions: Tone-1 vs. not Tone-2, Tone-2 vs. not Tone-2, Tone-3 vs. non Tone-3, and Tone 4 vs. not Tone-4. We then computed d'-values as a measure of detectability of each lexical tone among its competitors (with standard correction for perfect scores and zero false alarms).

Figure 5 plots the d'-scores averaged over the four tones, but broken down by context (carrier sentence, isolation), time pressure (on, off) and by context (in carrier sentence: left panel; in isolation: right panel).

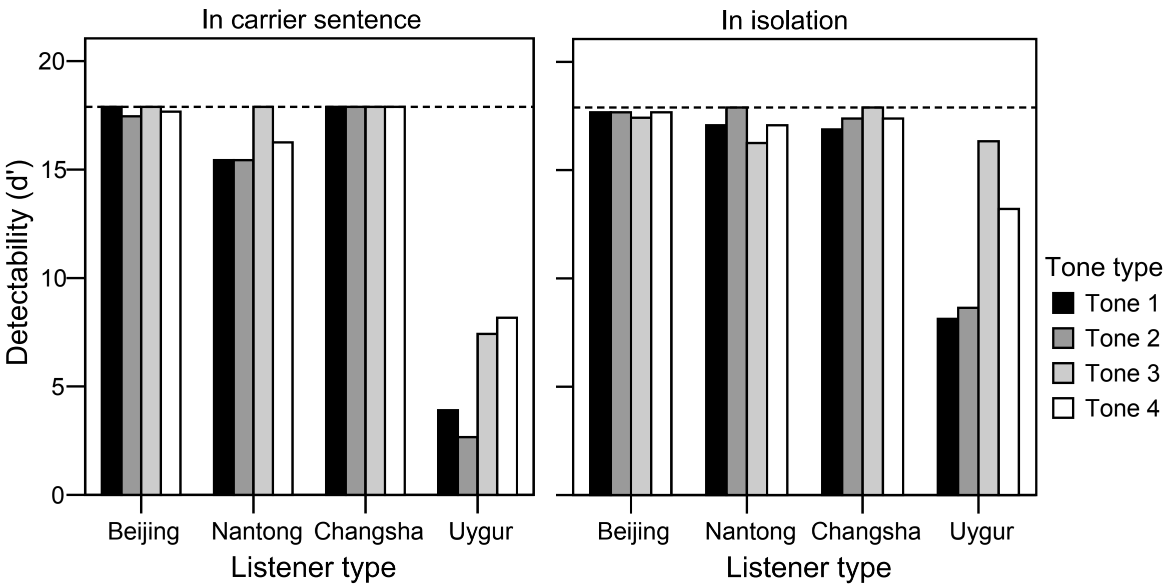


Figure 5. Detectability measure (d') for listener type broken down by tone type (Tone 1, Tone 2, Tone 3 and Tone 4). Maximum d' = 17.89

The d'-values were submitted to a similar RM-ANOVA as described above. Significant effects were still found for listener type [$F(3, 77) = 92.3$ ($p < .001$)], time pressure [$F(1, 77) = 8.8$ ($p = .004$)], context [$F(1, 77) = 22.8$ ($p < .001$)] and tone [$F(2.9, 222.9) = 10$ ($p < .001$)]. Significant effects were also found for two second-order interactions, i.e. context × listener type [$F(3, 77) = 37.4$ ($p < .001$)] and tone × listener type [$F(8.7, 222.9) = 12.6$ ($p < .001$)]. One third-order interaction, viz. context × tone type × listener type [$F(8.4, 216.5) = 2.4$ ($p = .016$)] remains significant as well. Scheffé post-hoc tests separated the Uygur listeners from the tone-language groups, who did not differ from each other ($\alpha = .01$).

To summarize, the results show that the tone listeners were significantly more sensitive to the lexical tones than the non-tone listeners (Uygurs) overall, even though the Uygurs discriminated between the four tones much better than chance. When the tones were presented in context, the Uygurs' performance deteriorated much more than that of the tone-language groups. The Uygur listeners came closest to the tone-language groups when listening to citation forms, especially Tone 3; performance was intermediate for Tone 4 in isolation, and poorest with contexted Tone 2 and Tone 1.

3.2. Post-lexical identification (statement vs. question intonation)

In this section, we will analyze question responses to the two pitch manipulations in detail, i.e. the final rises and the overall pitch levels, in order to investigate the perception of question intonation across the two groups of listeners. When we analyze the effect of manipulation of the boundary tone, we average the data over the seven overall pitch levels. Conversely, when analyzing the effect of overall pitch level, we average over the seven final rises. This procedure reduces the number of data points by a factor 7. This procedure was adopted after we ascertained that the effects of boundary tone and overall pitch level were essentially independent. Although a four-way RM-ANOVA – with final rise, overall level and time pressure as within-subject factors and listener type as a between-subjects factor –

indicated a small interaction between rise and level, F(28.6, 2205) = 4.0 (p < .001), it accounts for a mere 4 percent of the total variability in perceived percent questions (as opposed to 70% and 23% for the main effects of final rise and overall pitch level, respectively). Moreover, visual inspection (see appendix) reveals that the weak interaction is only found with the native speakers (the Beijing listeners) responding without time pressure. In this condition only, there is a tendency for the effect of overall pitch level to be smaller for the lowest boundary tones than elsewhere along the continuum but this does not seem to invalidate our assumption of independence for the two stimulus factors.

3.2.1. Question responses to manipulation of final pitch movement

Figure 6 presents the percentage of question responses, and the associated reaction time, for the four types of listener as a function of the pitch manipulation of the final rise. Positive X-values denote a final rise, negative values are falls; '0 ST' indicates that the final syllable is spoken at a monotone.

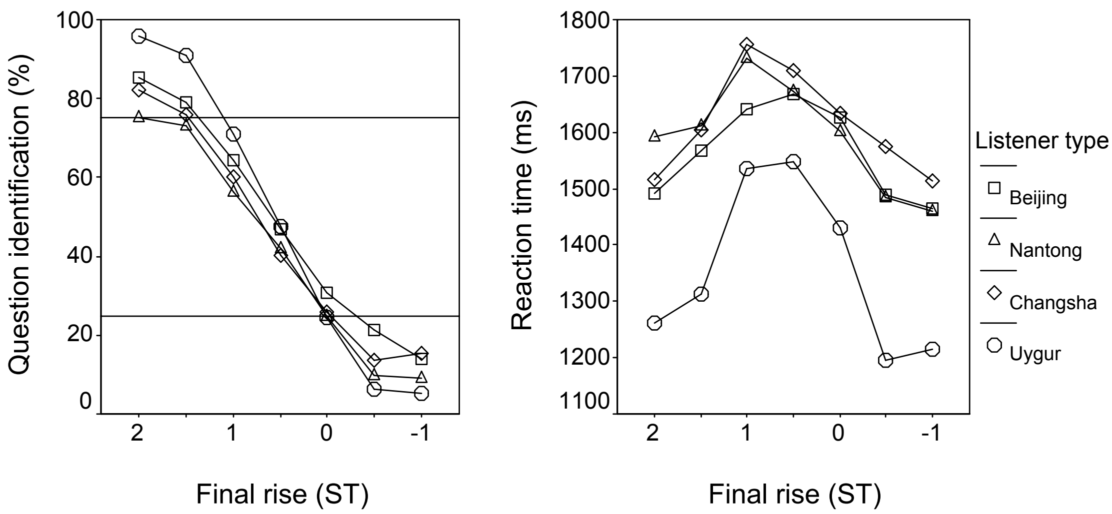


Figure 6. Question identification (percent, in the left panel) and associated reaction time (ms, in the right panel) as a function of excursion size of final boundary tone (semitones) for listeners from four languages spoken in China.

Percent question decisions. A three-way RM-ANOVA with listener type as a between-subject factor, and final rise and time pressure as within-subject factors was carried out on the sentence-type responses (statement = 0, question = 1). A large significant effect was obtained for final rise [F(2.9, 227) = 289 (p < .001)], but no significance was found for either listener type [F(3, 77) = 1.3 (p = .27)] or time pressure [F(1, 77) = .047 (p = .83)]. The effect of final rise is due to well-defined cross-overs from question to statement for all four groups of listeners. Typically, final rises are perceived as question markers, falls and the final monotone yield a clear majority of statement judgments. It is only the small-sized rises (between 0.5 and 1.0 ST) that are ambiguous between question and statement. The effect of listener group is not interesting to our analytical problem, as it might only reveal slight differences in response bias.

Significance was also found for the interaction of listener type × final rise [$F(8.8, 227) = 3.6$ ($p < .001$)] and final rise × time pressure [$F(4.8, 366) = 2.8$ ($p = .02$)] but not for listener type × time pressure [$F(3, 77) = .22$ ($p = .89$)]. There was no significant third-order interaction. The significant interaction of listener type × final rise is crucial, since it reveals differences in sensitivity, i.e., the psychometric function is clearly steeper for Uygurs than, for instance, for Changsha listeners.

We tested this by first computing for each listener a boundary width, which was defined as the distance in step-size units separating the 25% and 75% cross-over points as determined by probit interpolation. The steepness (slope coefficient) of the cross-over was then calculated by dividing the boundary width into 50. Thus, a slope coefficient of 10 would indicate that increasing the terminal frequency of the final rise by 1 semitone (i.e. 2 steps) yields an increment of the number of ‘question’ responses by 10 percentage points. More generally, the higher the slope coefficient, the steeper the cross-over and the more sensitive the listener. The results are listed in Table 3.

Table 3. Slope measure (N, mean, SD) for manipulation of terminal boundary pitch for four groups of listeners broken down by task performance with and without time pressure.

Listener type	*N*	*Without time pressure*		*With time pressure*	
		Mean	*SD*	*Mean*	*SD*
Beijing	30	20.9	12.9	23.1	20.9
Nantong	8	27.1	15.4	16.4	10.2
Changsha	13	16.4	9.5	22.9	16.9
Uygur	30	42.1	29.1	32.8	23.9

From Table 3 we can see that Uygur listeners have the steepest slopes (42.1), which indicates that these listeners are more sensitive to the changes in the final rise than the three tone-language groups. A two-way RM-ANOVA on the slope measures with listener type as a between-subject factor and time pressure as a within-subject factor show significant effects for listener type [$F(3, 77) = 5.0$ ($p = .003$)] and not for time pressure [$F(1, 77) = 1.2$ ($p = .28$)]. However, the second-order interaction between listener type and time pressure was found significant [$F(3, 77) = 3.0$ ($p = .034$)]. Scheffé post-hoc tests show that the Uygur group was significantly different from the other tone-language groups, which did not significantly differ among themselves ($\alpha = .01$).

Decision time. A similar analysis was carried out on the reaction time for sentence-type responses. Here significance was found for listener type [$F(3, 77) = 6.7$ ($p < .001$)], final rise [$F(4.4, 340) = 27.4$ ($p < .001$)] and for time pressure [$F(1, 77) = 165.2$ ($p < .001$)]. Significance was also found for the second-order interaction final rise × time pressure [$F(4.8, 367) = 5.8$ ($p < .001$)], but not for the interactions of listener type × final rise [$F(13.2, 340) = 1.5$ ($p = .103$)] and listener type × time pressure [$F(3, 77) = 1.6$ ($p = .206$)]. Scheffé post-hoc tests show that the Uygur group was significantly different from the other tone-language groups (Beijing, Nantong and Changsha), which did not significantly differ among themselves ($\alpha = .01$), i.e. 1357 vs. 1570, 1595 and 1624 ms respectively ($\alpha = .01$).

The effect of final rise seems to follow in a straightforward fashion from the ambiguity of the stimulus. The clearest example is seen in the case of Uygur listeners. The two most extreme rise values are unambiguous exemplars of questions, yielding fast decisions. Likewise, the two final falls (negative rise values) are unambiguous tokens of statements, with fast decision times. Only the three middle values are ambiguous, with question responses between 25 and 75%. Here the decisions are delayed by some 100 to 200 ms. The same effect, though less extreme, is seen for the other listener groups.

Figure 7 is a further breakdown (by time pressure) of the left-hand panel of Figure 6. The effect of time pressure as such is, predictably, that responses are a lot faster (over 400 ms on average) when time pressure is applied. However, without time pressure the listeners are especially slow when having to decide on ambiguous stimuli. When the need arises to respond quickly, they save relatively more time on ambiguous stimuli than on unambiguous ones.

There is a further complication in the data. When there is no time pressure, the difference in reaction time between slow and fast listener groups is ca. 440 ms. When time pressure is on, the difference between the extreme groups is reduced to 300 ms. Consequently, we would expect there to be greater overlap between the four groups such that the groups are less distinct when performing under time pressure. However, the reverse proves to be the case. The statistical separation between the four listener groups is clearly better when subjects perform under time pressure than when not. Also, the order among the four groups changes considerably. Uygurs are always the fastest group, regardless of time pressure. Without time pressure the three groups of tone-language listeners are relatively close together and only Changsha listeners are significantly slower than the others.

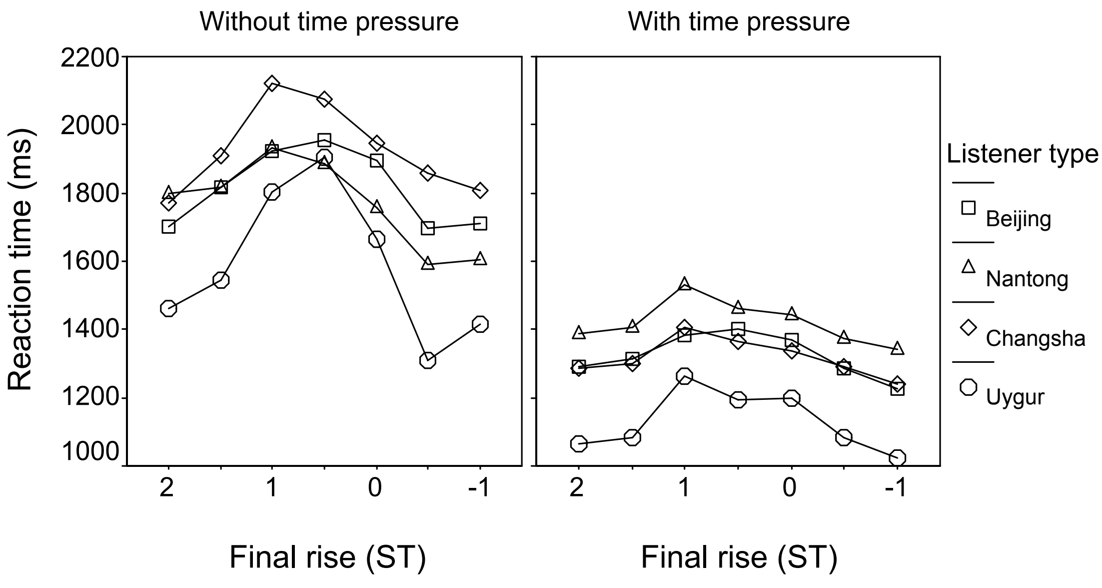


Figure 7. Reaction time (ms) for question identification without time pressure (left panel) and with time pressure (right panel) as a function of excursion size of final boundary tone (semi-tones) for listeners from four languages spoken in China.

3.2.2. Question responses to manipulation of overall pitch level

Now let us consider the sentence-type responses and their associated reaction time from the four types of listener as a function of the manipulation of pitch level. The data are presented in Figure 8.

We ran a similar three-way RM-ANOVA on the sentence-type responses as we did with final manipulation, with listener type as a between-subject factor and with overall pitch level and time pressure as within-subject factors. A significant effect was revealed for pitch level [$F(3.5, 272.9) = 54.18$ ($p < .001$)], but not for listener type [$F(3, 77) = 1.35$ ($p = .27$)] nor for time pressure [$F(1, 77) = .05$ ($p = .83$)]. The interactions of listener type × pitch level [$F(10.0, 272.9) = 4.30$ ($p < .001$)] and of time pressure × pitch level [$F(6, 462) = 7.41$ ($p < .001$)], however, were significant.

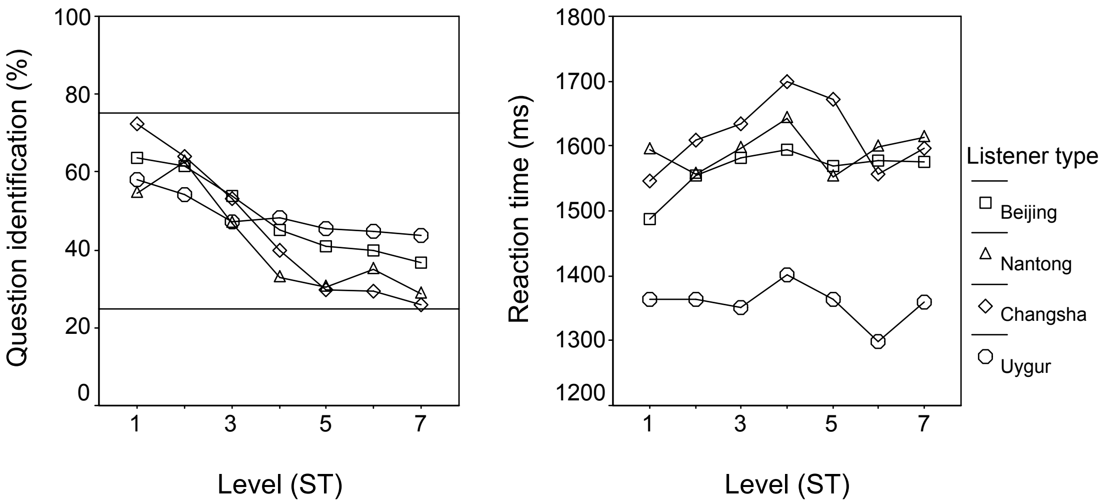


Figure 8. Question identification (percent, in the left panel) and associated reaction time (ms, in the right panel) as a function of excursion size of pitch level (semitones) for listeners from four languages spoken in China.

Again, the main effect of listener type is irrelevant (see above). The effect of pitch level is as would be expected: higher overall pitch levels yield more question responses. However, in contradistinction to manipulating the final rise, the manipulation of overall level fails to effect a well-defined cross-over, at least when the responses are averaged across all four listener groups (left panel). In fact, a full cross-over (from 78 to 20% question responses) is obtained only for Changsha listeners, and only when responding without time pressure.

In Table 4, we present slope measures for manipulation of overall pitch level for the four groups of listeners with and without time pressure.

Table 4. Slope measure (N, mean, SD) for manipulation of overall pitch level for four groups of listeners broken down by task performance with and without time pressure.

Listener type	*N*	*Without time pressure*		*With time pressure*	
		Mean	*SD*	*Mean*	*SD*
Beijing	30	7.7	6.6	2.0	9.9
Nantong	8	9.8	12.1	5.6	4.7
Changsha	13	13.5	9.9	8.9	12.1
Uygur	30	2.7	2.4	1.7	2.5

Table 4 shows that the steepest slope coefficient (13.5) is obtained, when no time pressure is applied, by the Changsha group, closely followed by the Nantong and Beijing groups, while the Uygur listeners have the flattest curves. A two-way RM-ANOVA on the slope measures with listener type as a between-subject factor and time pressure as a within-subject factor shows significant effects for listener type [F(3, 77) = 6.6 (p < .001)] and time pressure [F(1, 77) = 14.9 (p < .005)]. No significant effect was found for the second-order interaction. Scheffé post-hoc tests divided listener type into three subsets, (i) Uygur, Beijing, (ii) Beijing, Nantong, and (iii) Nantong, Changsha (α = .01).

Decision times. These are the same data as before but with one factor (final rise) replaced by another (pitch level). Therefore, the RM-ANOVA on the reaction times showed, predictably, significant effects for listener type [F(3, 77) = 6.75 (p < .001)], for time pressure [F(1, 77) = 164.74 (p < .001)] and for pitch level [F (5.8, 443) = 3.44 (p = .003)]. Again, Scheffé post-hoc tests divided the Uygur listeners from the rest of the other three tone-language groups.

In order to better understand the interaction between listener group and time pressure, a further breakdown of the results is presented in Figure 9, which plots percent question responses (upper panels) and the associated reaction times (lower panels) separately for absence (left) and presence (right) of time pressure.

Crucially, without time pressure the listener groups come closer to realizing a cross-over than when performing under time pressure. Also, in the absence of time pressure, the relationship between sentence-type responses and reaction time is as predicted (and also as found in the analysis above). Again, when the stimulus is ambiguous between question and statement, the reaction time is slow, but when the categorization is (reasonably) clear, the decisions are faster. This relationship is found for all listener groups. With time pressure applied, however, all response curves are shallow or completely level, and local maxima in the reaction-time curves are largely absent.

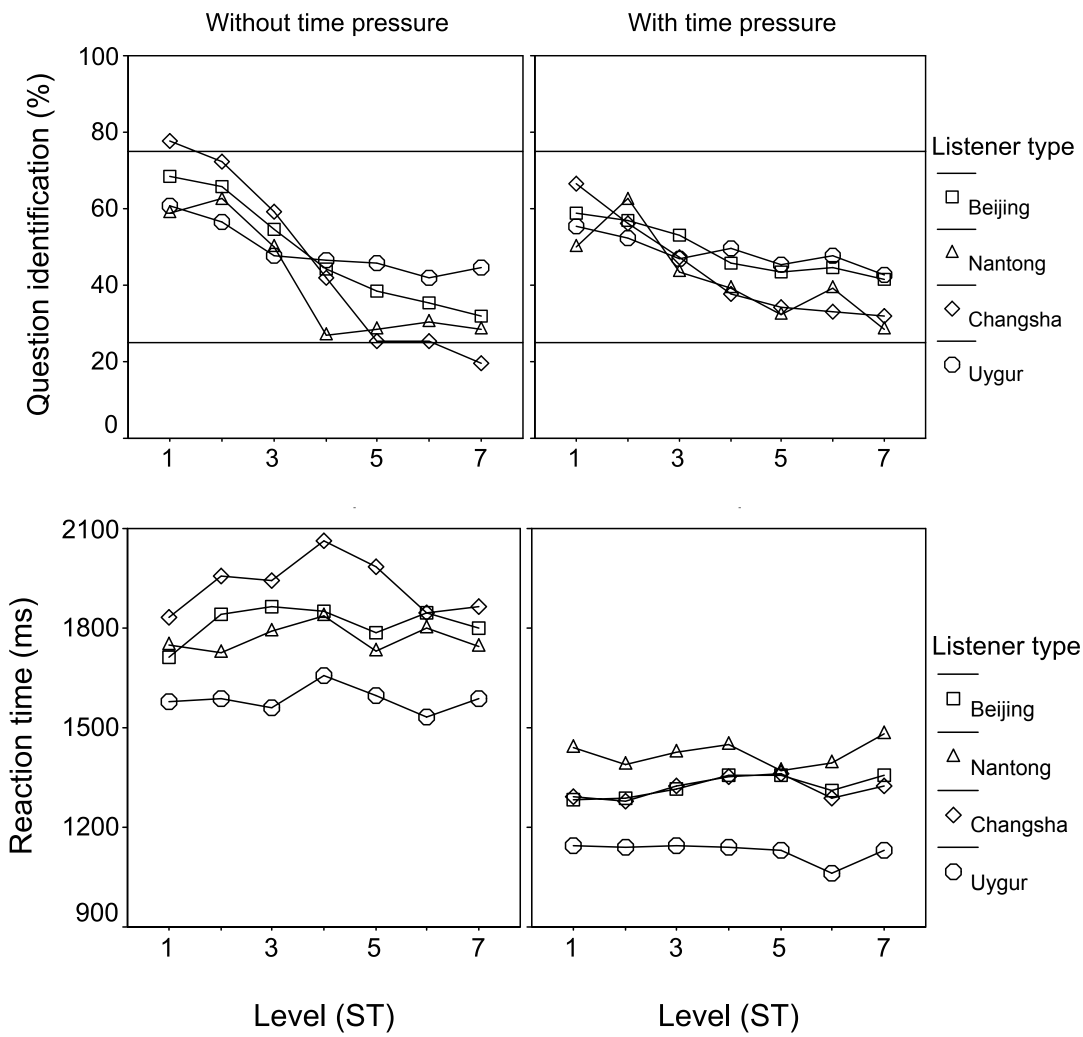


Figure 9. Question identification (percent, upper panels) and their associated reaction time (ms, lower panels) as a function of pitch level broken down by four types of listener, with and without time pressure (left and right panels, respectively).

To summarize the above, we have plotted the slope values of the four groups of listeners separated by manipulation type in two panels (right: final rise; left: overall pitch level) and broken down by time pressure (yes vs. no) in Figure 10.

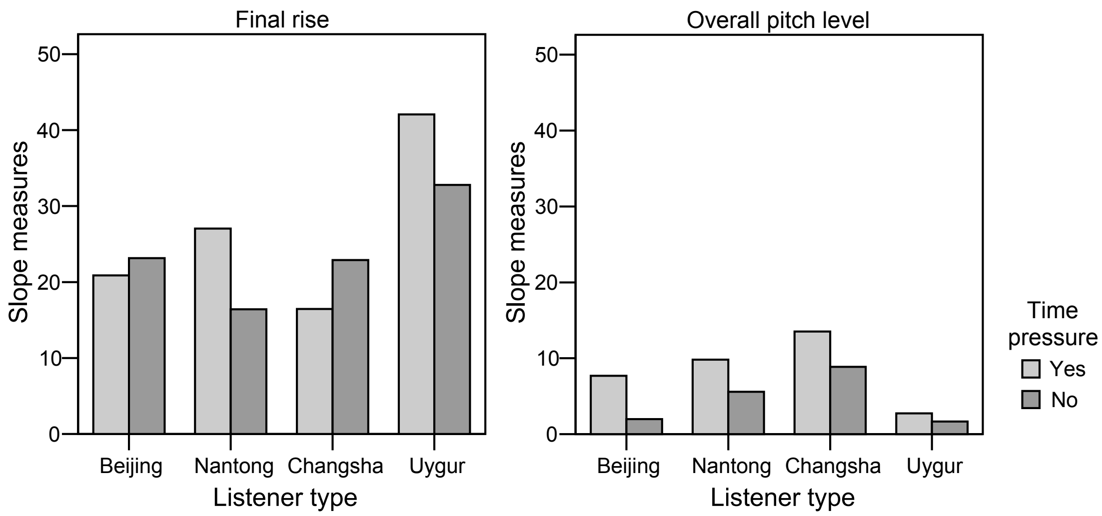


Figure 10. Slope measures as a function of listener type separated by manipulation type (final rise, right; overall pitch level, left) and broken down by time pressure (On/Yes vs. Off/No).

A three-way RM-ANOVA with listener type as a between-subject factor and manipulation type (final vs. level) and time pressure (on vs. off) as a within-subjects factors was carried out on the slope values. It yielded significant effects for listener type [F(3, 77) = 3.1 (p = .031)], manipulation type [F(1, 77) = 45.8 (p < .001)] and time pressure [F(1, 77) = 8.7 (p = .004)]. Significance was also found for the second-order interaction of listener type × manipulation type [F(3, 77) = 6.5 (p = .001)] and for the third-order interaction listener type × time pressure × manipulation [F(3, 77) = 3.1 (p = .030)]. Scheffé post-hoc tests show that the Beijing group significantly differed from the Uygur group only (α = .01).

From Figure 10, we can see that manipulating the final rise has a much stronger effect on the perception of sentence type than manipulation of overall pitch level, regardless of time pressure. Among the four groups, the Uygurs were most sensitive to the final rise and least to overall pitch level, with or without time pressure. The tone-L2 listeners are more sensitive to pitch level as a question cue than the native Beijing listeners, both with and without time pressure.

3.3. Comparison of identification in lexical and sentence intonation

In the preceding sections we presented the results for perception of lexical tone and of sentence intonation separately. We will now attempt to evaluate the differences between Beijing L1 and L2 listeners by taking into account the results from both types of test simultaneously. As we explained before, the two perception tests cannot be compared directly due to the fact that one (lexical tone) can be evaluated in terms of correct or incorrect, but the other (intonation) has no pre-given norm. This difference should be borne in mind when considering the following results.

In Figure 11 we have plotted the four listener groups performing with and without time pressure in a two-dimensional plane defined by their sensitivity to the final rise as a question cue (vertically) and the detectability of the lexical tones (d'-values, horizontally).

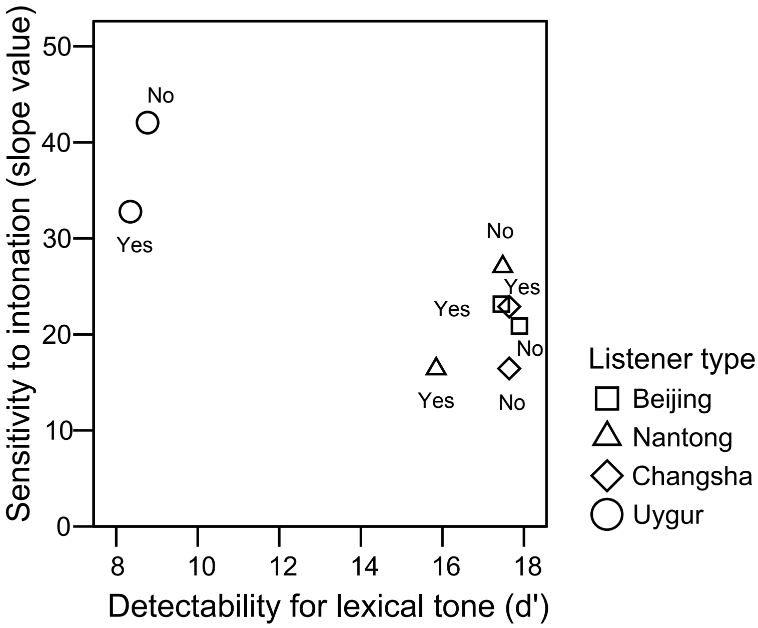


Figure 11. Sensitivity to final rise as cue for question intonation (slope coefficient) plotted against detectability based on percent of lexical tone identification (d prime) broken down by listener type and time pressure(No: no time pressure; Yes: with time pressure).

Figure 11 shows a negative correlation between sensitivity to intonation and successful lexical-tone identification. The Uygur group with a non-tone language background has relatively poor lexical-tone identification but is highly sensitive to manipulation of the final rise as a question cue; for the tone-language listeners the situation is the reverse: these are much more successful at lexical-tone identification but rather insensitive to the final boundary pitch as a question cue at the sentence level.

4. General discussion

4.1. Summary of main findings

Before we proceed to drawing any conclusions, let us first summarize, in Table 5, what we have found in our experiments. In this table a plus or minus sign indicates that the L2 listeners displayed a significantly better or poorer task performance, respectively, than the Beijing controls, in terms of tone discrimination (d'), and speed thereof (ms) or in sensitivity to manipulation of pitch (boundary rise, overall pitch) in the identification of sentence type. Cells are empty when no significant difference could be established by the RM-ANOVA.

Table 5. Summary of findings for perception of lexical tone and sentence intonation: task performance of L1 listeners (Beijing) versus L2 listeners (Nantong, Changsha, Uygur). A plus or minus sign indicates significantly better or poorer task performance, respectively, by L2 listeners than by Beijing controls.

Stimulus type			L2 listeners re. Beijing		
			Nantong	Changsha	Uygur
Lexical tone	In carrier	Detectability (d′)			−
		Decision time			−
	Citation	Detectability (d′)			−
		Decision time			−
Sentence intonation	Final rise	Sensitivity (slope)			+
		Decision time			+
	Level	Sensitivity (slope)		+	
		Decision time			+

The tone-language listener groups correctly identify lexical tones in Beijing dialect as often as Beijing listeners themselves, and they are as insensitive to pitch manipulations at the sentence level as the Beijing group. Additionally, a further analysis of tone perception by the Uygurs shows a significant difference between carrier sentences vs. citation forms; this difference is not manifest in table 5, but was found earlier in figure 3. Next, we will discuss the results concerning the perception of lexical tone and sentence intonation by Chinese L1 and L2 speakers.

4.2. Lexical tone deficit exhibited by L2 speakers

For all four groups of listeners, whether having a native language with or without lexical tone phenomena, identifying tones presented in context took considerably more time than identifying the tones in citation forms. For the non-tone Uygur listeners, however, there was also a substantial increase in identification errors for tones presented in context; this effect simply did not arise with any of the tone-language groups: context slowed them down but did not lead to more errors. The canonical shapes of the four tones of Beijing change considerably when preceded and/or followed by other tones in context. We assume that listeners with a tone-language background apply some phonological rule in order to relate the surface tone to its canonical form. Applying this rule is a more or less error-free process but it takes time. For the non-tone Uygur listeners the application of the tone-sandhi rules is not automatic and error-free, leading not only to delay but also to errors.

The finding that the listeners are able to factor out the effect of tonal context is consistent with previous studies on Chinese lexical tone identification in context (Xu 1994). For Uygur listeners, however, undoing the effect of tone sandhi is a problem (as it was shown to be for American listeners, cf. Broselow, Hurtig, and Ringen, 1987). Typically, Uygur listeners learned the tones in citation forms first (in supervised learning) but had to acquire the tone sandhi rules later through natural exposure outside the classroom. Of course, the tone listeners simultaneously acquired the tones both in their sandhi form as well as in citation form. Even if they are L2 speakers of Beijing dialect, they could still decode tones in connected speech as well as they did in citation forms. In conclusion, we have provided evidence for lexical tone deficit in tone-language learning that is only experienced by L2 speakers with a non-tone L1 and the positive transfer of tones, both in citation forms and in context, with L2 learners with a tone-language background. At the moment, we cannot determine the main factor that governs the positive transfer exhibited by our L2 subjects, i.e. Nantong and Changsha speakers. It may be due to the exposure to radio or television in Standard Chinese (which resembles the Beijing dialect) or to their tonal experience in their mother tongue. Further investigation is needed to separate the two factors[10].

4.3. Intonation perception by Chinese speakers

In non-tone languages such as English or Dutch, questions typically end on a rising pitch and statements have falling terminal pitch. In Chinese, however, the difference between statement and question intonation is more complicated because of the interaction of tone and intonation. For instance, question intonation with Tone 2 (rising tone) in sentence-final position has a rising shape but it has falling pitch with Tone 4 (falling tone).

Since the lexical tonal modification in Chinese employs the same dimension (i.e. pitch) as intonation does, it tends to blur the visibility of the latter to the point that the very existence of intonation in the Beijing dialect (and other Chinese dialects) has been questioned (Kratochvil 1998).

Our findings suggest that F0 is primarily perceived at the lexical level for listeners of a tone language but at the sentence level for listeners of a non-tone language.[11] Our results are in line with Seddoh's hypothesis (2002), i.e. the linguistic typological differences between tone and non-tone languages suggest that native speakers of tone languages may be less able, relative to native speakers of intonation languages, to use intonation to signal different sentence types. Our perceptual results are in line with cross-linguistic production data from tone languages (e.g. Ewe, Chinese and Thai) and intonation languages (e.g. English, French and Russian) (cf. Gandour et al. 1995; Ross et al. 1986).

Gandour et al. (2003) investigated the identification of statement versus question intonation in Chinese with Chinese L1 and English L2 listeners, using natural tokens only. The Chinese L1 listeners were more accurate in identifying sentence type than the English non-native listeners, but curiously enough, the English listeners were significantly faster. Table 6 combines the findings by Gandour et al. and our own. It contrasts the performance of Chinese listeners with that of non-tone listeners (English, Uygur); the better language type is indicated by a '+'-sign, the counterpart is given a '–'.

Table 6. Comparison of task performance by Chinese versus English/Uygur listeners when judging sentence type from Chinese intonation in natural and resynthesized stimuli.

Type of stimuli in Chinese	Task/measure	Tone-language Chinese	Non-tone language English / Uygur
Natural speech, four utterances	Accuracy	+	–
	Reaction time	–	+
Manipulated speech, one basic utterance	Sensitivity	–	+
	Reaction time	–	+

From Table 6, we see that the non-tone-language listeners are always faster than the tone-language listeners regardless of the type of stimuli (natural or manipulated), and they are more sensitive to intonation. We suggest that listeners with a non-tone language background give processing priority to sentence melody – leading to fast decision times. Speakers of a tone language have learnt to divide their resources between two competing tasks: listening for pitch cues bearing on word meanings (lexical tone) as well as monitoring pitch at the sentence level. This would account for the longer processing times as well as the diminished sensitivity to pitch cues at the sentence level for Chinese listeners. Chinese listeners, of course, are still better at identifying the sentence melody as intended by a fellow-Chinese speaker, simply because they know the system better than foreign listeners do.

More than likely the Uygur listeners used the intonational system of the Uygur language when deciding whether the target utterance was a statement or a question. We are not aware of any published studies on the melodic marking of sentence type in Uygur but it seems safe to assume that this language uses the semi-universal association of high terminal pitch for question and low pitch for statement (see also Gussenhoven 2004). The exact implementation of the categorical distinction, however, differs from one language to the next (Bolinger 1978; Ladd 1981), which would account for the fact that the category boundary between statement and question is not the same for our Uygur listeners as for the Beijing listeners. For all this, the finding that the cross-overs from statement to question are steeper, and the identifications are faster for the Uygur listeners, indicates that they were less bothered by the lexical tones than the tone-language listeners were. This finding concurs with recent cross-linguistic perceptual studies by Gussenhoven and Chen (2000) and Chen (2005). These studies, too, show that Chinese listeners use differences in sentence melody less effectively than listeners of non-tone languages (Dutch and Hungarian) when having to interpret nonsense stimuli (meaningless CVCVCV sequences) as either a statement or a question.

In line with this, we conclude that, once free of lexical-tone interference, listeners become more sensitive to intonation. This is tantamount to saying that lexical-tone interference is the primary factor in reducing the sensitivity to pitch cues at the sentence level in listeners with a tone-language background.

4.4. Lateralization of lexical tone and sentence intonation

Imaging studies on pitch perception of healthy speakers show that pitch processing is influenced by language experience (Gandour et al. 2000, 2002, 2003, 2004; Hsieh et al. 2001). Their findings show that for Chinese listeners prosodic functions at the word level predominantly engage the left hemisphere (LH); prosodic functions at the sentence level elicit bilateral activation, with a right hemisphere (RH) preference. In contrast, listening to the same Chinese stimuli, English listeners show bilateral activation for lexical tone perception but predominantly right-sided activation for intonation.

Several hypotheses have been forwarded concerning the neurobiological mechanism underlying speech perception. The hypotheses involve either acoustics or linguistics, or both, as the source of hemispheric lateralization. Hypotheses involving an acoustically driven mechanism are called cue dependent. These claim that speech processing is subserved by neurobiological mechanisms specialized for particular aspects of the acoustic signal (e.g. the absolute size of a temporal domain) to account for the hemispheric specialization of local and global processing, irrespective of communicative or linguistic relevance as suggested by the cue-dependent hypothesis (e.g. Behrens 1989; Peoppel 2001). Hypotheses focusing on linguistic mechanisms are called task dependent. These claim that unique, neural mechanisms are recruited for the speech domain as suggested by the task-dependent or domain-specific effects (e.g. Van Lancker 1980; Ross and Mesulam 1979). Hybrid views engaging both linguistic and acoustic processing claim the existence of dynamic interaction between the two hemispheres, i.e., during spoken-language comprehension processes in the left and the right hemispheres are assumed to interact dynamically. There are two models especially worth mentioning in this context, (i) a comprehensive model of speech perception that is mediated primarily by RH regions for complex-sound analysis but which is lateralized to task-dependent regions in the LH when language processing is required (Gandour et al. 2004), and (ii) a dynamic dual-pathway model of auditory language comprehension that claims that segmental, lexical (i.e. tone) and syntactic information is processed in the LH, but suprasegmental sentence-level information (i.e. intonation) in the RH; this model additionally claims that the involvement of the LH increases as either the stimulus or the task is more linguistic in nature (Friederici and Alter 2004).

First, our findings cannot be explained by the cue-dependent hypothesis, which depends crucially on the absolute size of temporal domain of the speech, e.g., the left hemisphere preferentially extracts information from short 20−50 ms temporal-integration windows while the right hemisphere homologues preferentially extract information from long 150−250 ms integration windows (Hickok and Poeppel 2000; Poeppel 2001). In our stimuli the time domain of the manipulation of sentence intonation was either the syllable (word) in the case of the final rise (150 ms) or the entire sentence (six syllables, roughly 1000 ms) in the case of overall pitch level. Nevertheless, Uygur listeners only showed lexical tone deficit but their perception did not differ from that of the tone-language groups in intonation identification, and in fact they were significantly faster – regardless of the temporal difference.[12]

Second, our findings cannot be explained by a task-dependent hypothesis, either, because the selective impairment that occurred in lexical-tone perception indicates that the areas necessary for lexical tone processing are not the same; therefore the processing of lexical tone and sentence intonation does not happen in one single module. However, our findings are compatible with hybrid hypotheses claiming the existence of dynamic interaction between the two hemispheres. The differential patterns between tone-language and non-tone-language groups (i.e. greater sensitivity to lexical tone but reduced sensitivity to intonation) can be accounted for if the perception of speech prosody is mediated by different areas or in separate modules (Gandour et al. 2004).

In conclusion, then, this study demonstrates that only L2 speakers with a non-tone L1 exhibit lexical tone impairment in learning a tone language, but the same speakers show better sensitivity to sentence intonation than L2 speakers with a tone L1. Our results, finally, support the functional view, i.e., if some phonetic dimension is exploited in one area of the grammar, e.g. lexical tones, it will not be used to the same extent in another part of the phonology, e.g. sentence intonation (Ross et al. 1986; Gandour et al. 1995; Seddoh 2002).

Notes:

1. This article is the full version of a paper first presented at the International Conference on Tone and Intonation, in Santorini, Greece, 9-11 September 2004. We gratefully acknowledge a travel grant by the Leiden University Fund (LUF) for subsidizing the first author's trip to the conference.

2. In fact, some passive knowledge of Beijing dialect may have been acquired through watching television in Mandarin before the Nantong and Changsha speakers moved to the capital.

3. HSK stands for China's Hanyu Shuiping Kaoshi 'Chinese Proficiency Test', a standardized test at the state level designed and developed by the HSK Center to measure the Chinese proficiency of foreigners, overseas Chinese and students from Chinese national minorities (http://www.hsk.org.cn).

4. The Xinjiang Uygur Autonomous Region is an area where bilingual language policy is in practice. Television programs and news papers, for instance, are issued in both Standard Chinese and in Uygur. The majority of the Uygur children receive their pre-university education in Uygur and take the university entrance examination in the Uygur language. When they are accepted by universities, they receive a one-year compulsory language training in Standard Chinese and they are required to pass Band 5 of the HSK before being allowed to continue their university education as all the courses at the university level are offered in Standard Chinese.

5. The authors thank Prof. Ning Chunyan (in Changsha), Prof. Li Bing (in Changsha), Associate Prof. Guo Weidong (in Ulumqi), and Mr. Li Gang (in Beijing) for their kind help in recruiting subjects. We also thank all our subjects for participating in the experiments.

6. The keyboard was designed and built by J.J.A. Pacilly at the Universiteit Leiden Phonetics Laboratory.

7. The E-prime script for stimulus presentation and response collection was written by J.J.A. Pacilly

8. The reference signal was a sequence of fourteen identical periods sampled from the middle of a natural schwa token. The artificial intensity contour comprised a 40-ms rise time, 50-ms steady state, and a 40-ms decay time.

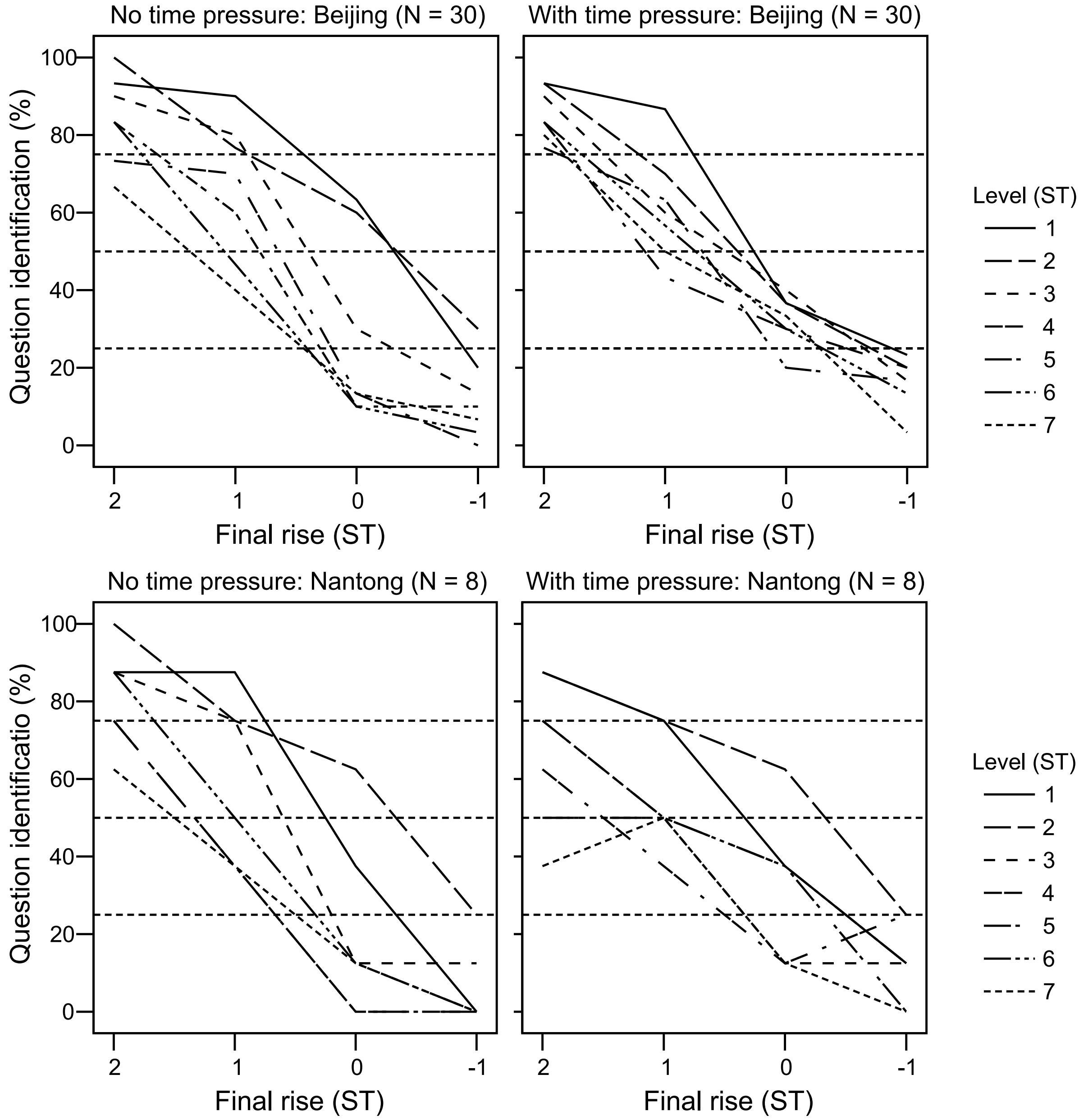
9. Degrees of freedom were always Huyhn-Feldt corrected, which may yield non-integer values for interaction terms.

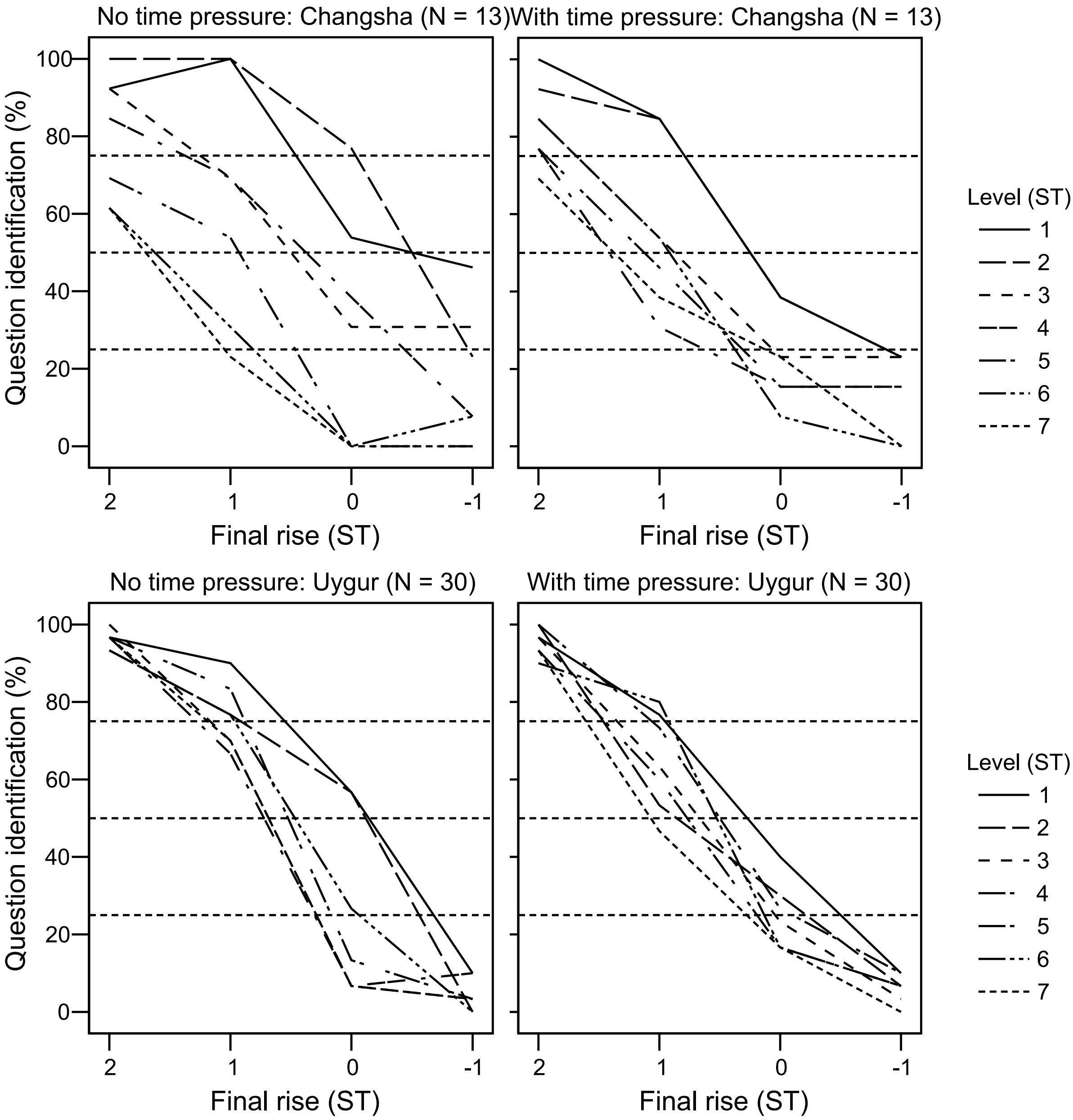
10. Here we thank Yi Xu for comments and valuable suggestions.

11. This conclusion is speculative to some extent. It may be the case that the processing of pitch at the sentence level may still get priority if the full range of linguistic functions is tested, i.e. not just the signalling of sentence type but also the marking of focus and phrase boundaries (cf. Xu and Xu, 2005).

12. It may be objected that the temporal domain of signalling the sentence type spans the entire duration of the utterance, since it was found, for instance for Danish (Thorsen 1980) and Dutch (van Heuven and Haan 2000) that questions are globally characterized by a gradually rising pitch from beginning of the sentence onwards. However, we know from other research where such global pitch cues for question (vs. statement) were pitted against the local cue in the final rise, the latter cue is decisive. No combination of global cues could ever reverse the percept that was evoked by the final boundary cue (van Heuven and Haan 2002).

Appendix: full breakdown of perception of sentence intonation.





Percent question responses as a function of final boundary tone and of overall pitch level, broken down by type of listener (Beijing, Nantong, Changsha and Uygur, top-to-bottom) and by time pressure (absent, left versus present, right).

References

Behrens, Susan J. 1989 Characterizing sentence intonation in a right hemisphere damaged population. *Brain and Language* 37: 181–200.

Bolinger, Dwight, 1978 Intonation Across Languages. In: Greenberg, Joseph H. (ed.), *Universals of Human Language. Vol II: Phonology*, 471–524. Stanford:Stanford University Press.

Boersma, Paul and David Weenink 1996 *Praat, a system for doing phonetics by computer.* Report 132. University of Amsterdam: Institute of Phonetic Sciences.

Broselow, Ellen, Richard R. Hurtig and Catherine Ringen 1987 The perception of second language prosody. In: Georgette Ioup and Steven H. Weinberger (eds.), *Inter-language phonology, the acquisition of second language sound system,* 350–362. Cambridge: Newbury House.

Chen, Aoju 2005 Paralinguistic meaning of intonation: universal or language-specific? Ph.D. dissertation, Nijmegen University.

Coster, D. and Paul Kratochvil 1984 Tone and stress discrimination in normal Beijing speech. In: Beverly Hong (ed.), *New Papers on Chinese Language Use,* 119–132. Canberra: Australia National University.

De Francis, John F., 1963 *Beginning Chinese*. New heaven, Connecticut: Yale University Press.

Friederici, Angela D. and Kai Alter 2004 Lateralization of auditory language functions: a dynamic dual pathway model. *Brain and Language* 89: 267–276.

Gandour, Jack 1983 Perception of tone in Far Eastern languages. *Journal of Phonetics* 11: 149–175.

Gandour, Jack, Mario Dzemidzic, Donald Wong, Mark Lowe, Yunxia Tong, Li Hsieh, Nakarin Satthamnuwong and Joseph Lurito 2003 Temporal integration of speech prosody is shaped by language experience: An fMRI study. *Brain and Language* 84: 318–336.

Gandour, Jack, Jennifer Larsen, Sumalee Dechongkit, Sovit Ponglorpisit and Fuangfa Khunadorn 1995 Speech prosody in affective contexts in Thai patients with right hemisphere lesions. *Brain and Language* 51: 422–443.

Gandour, Jack, Yunxia Tong, Donald Wong, Thomas Talavage, Mario Dzemidzic, Yisheng Xu, Xiaojian Li and Mark Lowe 2004 Hemispheric roles in the perception of speech prosody. *NeuroImage* 23: 344–357.

Gandour, Jack., Donald Wong, Li Hsieh, Bret Weinzapfel, Diana Van Lancker and Gary Hutchins D. 2000 A crosslinguistic PET study of tone perception. *Journal of Cognitive Neuroscience* 12: 207–222.

Gandour, Jack, Donald Wong, Mark Lowe, Mario Dzemidzic, Nakarin Satthamnuwong, Yunxia Tong and Joseph Lurito 2002 Neural circuitry underlying perception of duration depends on language experience. *Brain and Language* 83: 268–290.

Gårding, Eva 1985 Constancy and variation in Standard Chinese tonal patterns. *Lund University Working Papers* 28: 19–51.

Gårding, Eva 1987 Speech act and tonal pattern in Standard Chinese: constancy and variation. *Phonetica* 44: 13–29.

Gårding, Eva and Arthur S. Abramson 1965 A study of the perception of some American English intonation contours. *Studia Linguistica* 19: 61–79.

Gussenhoven, Carlos, 2004 *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

Gussenhoven, Carlos, and Aoju Chen 2000 Universal and language-specific effects in the perception of question intonation. In: Yuan, Baozong, Taiyi Huang, Xiaofang Tang (eds.), *Proceedings of the 6th International Conference on Spoken Language Processing ICSLP 2000*, 2: 91–94. Beijing: China Military Friendship Publish.

Hickok, S. Gregory , and David Poeppel 2000 Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences* 4: 131–138.

Howie , M. John, 1976 *Acoustical studies of Mandarin vowels and tones*. London: Cambridge University Press.

Hsieh, Li, Jack Gandour, Donald Wong and Gary D. Hutchins 2001 Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. *Brain and Language* 76: 227–252.

Kratochvil, Paul 1998 Intonation in Beijing Chinese. In: Daniel Hirst and Albert Di Cristo (eds.), *Intonation Systems: A Survey of Twenty Languages*, 417–431. Cambridge: Cambridge University Press.

Ladd, Robert D. 1981 On intonational universals. In: Terry Myers, John Laver and John Anderson (eds.), *The Cognitive Representation of Speech*, 389–397. Amsterdam: North Holland Publishing.

Leather, Jonathan 1987 F0 pattern interference in the perceptual acquisition of second language tone. In: Allan James and Jonathan Leather (eds.), *Sound patterns in second language acquisition,* 59–81. Dordrecht: Foris.

Lee, Yuh-Shiow, Douglas Vakoch and Lee H. Wurm 1996 Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research* 25: 527–544.

Moulines, Eric, and Werner Verhelst 1995 Time-domain and frequency-domain techniques for prosodic modification of Speech. In: Kleijn, W. Bastiaan and Kuldip K. Paliwal (eds.), *Speech Coding and Synthesis,* 519–555. Amsterdam: Elsevier Science.

Pike , Kenneth L. 1948 *Tone languages.* Ann Arbor: University of Michigan Press.

Poeppel, David 2001 Pure word deafness and the bilateral processing of the speech code. *Cognitive Science* 25: 679–693.

Ross, Elliot D., and M-Marsel Mesulam 1979 Dominant language functions of the right hemisphere? Prosody and emotional gesturing. *Archives of Neurology* 36: 144–48.

Ross, D. Elliot, Herold A. Edmondson and Burton G. Seibert 1986 The effect of affect on various acoustic measures of prosody in tone and non-tone languages: A comparison based on computer analysis of voice. *Journal of Phonetics* 14: 283–302.

Seddoh, S. Amebu 2002 How discrete or independent are “affective prosody” and “linguistic prosody”? *Aphasiology* 16: 683–692.

‘t Hart, Johan, René Collier, and Antonie Cohen 1990 *A perceptual study of intonation*. Cambridge: Cambridge University Press.

Thorsen, Nina G. 1980 A study of the perception of sentence intonation − Evidence from Danish. *Journal of the Acoustical Society of America* 67: 1014−1030.

Tsao, Wen Yen, 1967 Question in Chinese. *Journal of Chinese Language Teachers’ Association* 2: 15–26.

Tseng, Chiu-yu, 1981 An acoustic phonetic study on tones in Mandarin Chinese. Ph.D. dissertation, Brown University.

Van Heuven, Vincent J., and Judith Haan 2000 Phonetic correlates of statement versus question intonation in Dutch. In: Botinis, Antonis (ed.), *Intonation: Analysis, Modelling and Technology*, 119−144. Dordrecht: Kluwer.

Van Heuven, Vincent J., and Judith Haan 2002 Temporal development of interrogativity cues in Dutch. In Gussenhoven, Carlos and Warner, Natasha. (eds.), *Papers in Laboratory Phonology VII*, 61−86. Berlin: Mouton de Gruyter.

Van Lancker, Diana, 1980 Cerebral lateralization of pitch cues in the linguistic signal. *Papers in Linguistics* 13: 201–277.

Wang, William S.-Y., 1972 The many uses of F_0. In: Albert Valdman (ed.), *Papers in Linguistics and Phonetics to the Memory of Pierre Delattre,* 487–503. The Hague: Mouton.

Wang, William S.-Y., and Kung-pu Li 1967 Tone 3 in Pekinese. *Journal of Speech Hearing Research* 10: 629–636.

Wang, Yue, Joan A. Sereno and Allard Jongman in press L2 acquisition and processing of Mandarin tone. In: Ping, Li, Elizabeth Bates, Li Hai Tan and Ovid J.L. Tzeng, (eds.), *The Handbook of East Asian Psycholinguistics*. Cambridge: Cambridge University Press.

Xu, Yi, 1994 Production and perception of coarticulated tones. *Journal of the Acoustical Society of America* 95: 2240–2253.

Xu, Yi, and Ching X. Xu 2005 Phonetic realization of focus in English declarative intonation. *Journal of Phonetics* 33: 159–197.

Yuan, Jiahong, Chilin Shih and Greg P. Kochanski 2002 Comparison of declarative and interrogative intonation in Chinese. In: Bernard Bel and Isabelle Marlien (eds.), *Proceedings of the Speech Prosody 2002 Conference*, 711–714. Aix-en-Provence: Laboratoire: Parole et Langage.