What did you say just now, bitterness or wife?
An ERP study on the interaction between tone, intonation and context in Cantonese Chinese

Carmen Kung [1], *Dorothee J. Chwilla* [1], *Carlos Gussenhoven* [2,3], *Sara Bögels*[1], *Herbert Schriefers* [1]

[1]Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour
[2]Radboud University Nijmegen, Centre of Language Studies [3]Queen Mary University of London.
c.kung@donders.ru.nl

Abstract

Previous studies on Cantonese Chinese showed that rising question intonation contours on low-toned words lead to frequent misperceptions of the tones. Here we explored the processing consequences of this interaction between tone and intonation by comparing the processing and identification of monosyllabic critical words at the end of questions and statements, using a tone identification task, and ERPs as an online measure of speech comprehension. Experiment 1 yielded higher error rates for the identification of low tones at the end of questions and a larger N400-P600 pattern, reflecting processing difficulty and reanalysis, compared to other conditions. In Experiment 2, we investigated the effect of immediate lexical context on the tone by intonation interaction. Increasing contextual constraints led to a reduction in errors and the disappearance of the P600 effect. These results indicate that there is an immediate interaction between tone, intonation, and context in online speech comprehension. The difference in performance and activation patterns between the two experiments highlights the significance of context in understanding a tone language, like Cantonese-Chinese.

Index Terms: ERP, speech comprehension, speech perception, tone, intonation, context, reanalysis, Cantonese Chinese

1. Introduction

Intonation is a universal feature of languages that allows speakers to use fundamental frequency (F_0) variation to mark phrase structures and express discoursal meaning [1]. For example, in English, questions and statements are signalled by different intonation contours. In addition to realising intonation contours, F_0 variation is also exploited in tone languages to convey lexical information [2]. For example, in a tone language such as Cantonese Chinese, the syllable /fu/ means 'husband' if it is pronounced with the high-level tone (55), 'bitterness' if it is pronounced with the high-rising tone (25), 'wealth' with the mid-level tone (33), 'symbol' with the low-falling tone (21), 'wife' with a low-rising tone (23), and 'negative' with a low-level tone (22). Since intonation and lexical tone are both realised in terms of F_0 variation and since particular pitch movements may express both a lexical tone and an intonational meaning, the two functions may be hard to disentangle whenever there is a conflict between the direction of F_0 movement of the lexical tone and that of the intonation pattern (e.g. [3]). This is supported by the findings of Ma and colleagues [3], who showed that perception of lexical tones can be strongly influenced by the addition of a question intonation in Cantonese Chinese. Instead of the canonical form, the low tones (21, 23, 22) were misperceived as the high-rising tone (25) between 62% and 78.5% of the cases. This indicates that the conflicting information provided by tone and intonation can have a severe impact on tone perception.

As information from the tone of a word is important for lexical identification and processing in tone languages (e.g. [4]), tone misperception caused by intonation can profoundly disturb lexical processing. However, so far little is known about the online processing of conflicting information yielded by tone and intonation. The aim of the current study is to explore the online processing consequences of this interaction by comparing the processing and identification of monosyllabic critical words at the end of questions versus statements in Cantonese Chinese, using a tone identification task, and ERPs as an online measure of speech comprehension.

The error patterns in [3] indicated a qualitative difference between the effect of intonation contour on the identification of the high-mid tones and on the low tones. Therefore, we divided the six contrastive tones into two groups, high-mid tones (55, 25, 33) and low tones (21, 23, 22). Crossing the factor Lexical Tone (low, high-mid) with the factor Intonation (question, statement) results in a 2 x 2 design (see Table 1).

Table 1
Design of the present study with examples.

Conditions		Examples
Intonation	Tone	
Question	High-mid	(1) nei^{35} t^{h}ɐu^{21} sin^{55} tɐp^{33} **fu$^{25-6}$** "You just answered **bitter**?"
	Low	(2) nei^{35} t^{h}ɐu^{21} sin^{55} tɐp^{33} **fu$^{22-6}$** "You just answered **negative**?"
Statement	High-mid	(3) ŋɔ^{35} t^{h}ɐu^{21} sin^{55} tɐp^{33} **fu^{25}** "I just answered **bitter.**"
	Low	(4) ŋɔ^{35} t^{h}ɐu^{21} sin^{55} tɐp^{33} **fu^{22}** "I just answered **negative.**"

Note. The superscripted numbers represent five relative pitch levels, ranging from 1 the lowest to 5 the highest. The strings of numbers represent the pitch contour. The tone letter -6 represents the rising tone of the question intonation.

Given that no previous ERP studies investigated the interaction between tone and intonation, this study is of an exploratory nature. Two comparisons were made to study to what extent the interaction between tone and intonation affects online speech comprehension in Cantonese. First, we examined the effect of the conflicting F_0 information on speech comprehension. This is accomplished by comparing questions and statements ending with low tones. Second, we compared questions and statements ending with high-mid tones (as a control condition) for studying the extent to which question intonation affects speech processing when the canonical F_0 shape of lexical tones is not, or less drastically, affected by the rising contour of the question intonation.

We expected that the tonal and intonational F0 information accompanying words with low tones at the end of questions would elicit two distinct lexical representations: one with the canonical low tone and one with the intonation-induced high-rising tone. The conflict between the two possible lexical representations could trigger a P600-effect, which has been proposed to reflect a reanalysis to check for perceptual errors (e.g. [5]). We might also observe an N400-effect, which is generally taken to signal difficulties during (semantic) processes of information integration (e.g. [6, 7]).

2. Experiment 1

2.1. Methods

2.1.1. Participants and Materials

Twenty right-handed native Cantonese-Chinese speakers (15 female, 5 male, aged 21-57) participated in the experiment. Four sets of six monosyllabic words (which differ only in tone) were derived from the following root syllables: /jɐn/, /si/, /søy/ and /fu/. All the resulting 24 target words were located at the end of the carrier sentences because intonation-induced F_0 variation is the largest at this location ([3, 8]). These carrier sentences were semantically neutral and were produced with either a question intonation (see (1) and (2) in Table 1) or statement intonation (see (3) and (4) in Table 1). This resulted in 24 questions and 24 statements.

2.1.2. Task and procedure

Before the experiment, participants were told that they were going to listen to randomly presented questions and statements over a headphone. They were also instructed to pay attention to the sentence-final word and that they had to identify this word after the presentation of the sentence in a speeded tone identification task. For the task, participants were given six possible choices, which corresponded to the minimal tonal sextuplets based on each of the four syllables. Participants had to choose one of the six words as quickly as possible within a five-second time limit. Accuracy of the response was measured.

2.1.3. EEG recording and analysis

EEG was recorded from 25 tin electrodes, which were a subset of the international 10% system. Three midline and 22 lateral electrodes were used. The electrode montage and the procedure for data preprocessing was the same as the one used in earlier auditory ERP studies conducted in our lab [9]. All trials were time-locked to the sentence-final critical word. ERPs were measured from 150 ms preceding the stimulus to 1000 ms after the onset of the critical word.

In order to explore the onset and the duration of the effects, time-course analyses were performed on the mean amplitude in consecutive 50-ms time windows starting from stimulus onset. These time-course analyses were performed with Intonation (questions, statements), Tone (high-mid, low), Hemisphere (left, right), and Electrodes (AF7/8, F7/8, FT7/8, F3/4, FC3/4, T7/8, C3/4, CP5/6, P7/8, P3/4, PO7/8) as within-subject factors. Because of the increased chance of Type-I errors associated with the large number of analyses, only effects that reached significance in two or more consecutive time windows were considered significant. The analyses revealed significant Intonation × Electrode interactions between 350 ms and 550 ms (*p*s < .05) as well as an Intonation × Tone × Hemisphere × Electrode interaction between 200 and 550 ms (*p*s < .05).

2.2. Results

Since the perceptual accuracy for the high-level tone was at ceiling, we used Wilcoxon signed rank tests to compare the perceptual accuracy for each tone across conditions yielded by the tone identification task. Low tones (21, 23 and 22) at the end of questions elicited significantly more errors than the same tones at the end of statements (all *p*s <.001). In contrast, there was no significant difference in accuracy rate between the high-mid tones (55, 25 and 33) at the end of questions and the same tones at the end of statements (all *p*s >.05).

Due to the qualitative difference in accuracy rate between low tones in questions and the other conditions, we analysed the error patterns for these low tones by calculating the percentage of misidentifications of each of the tones as another tone out of the total number of errors for that tone. Most of the identification errors reflect a misperception of the target tones as the high-rising tone (Tone 21, 23 and 22 were misperceived as tone 25 at a rate of 78.7%, 74.0% and 53.6% respectively). In conclusion, the results revealed that low tones in questions had the highest error rate and that majority of these tones were misperceived as the high-rising tone.

Figure 1 shows the grand average waveforms for one left posterior electrode (P3). Based on the time-course analyses, the N400 effect was quantified in a 200-350 ms time window. Separate midline and lateral analyses were computed for the low tones and for the high-mid tones in this time window. For the low tones, the main effect of Intonation for the midline sites (F (1, 19) = 3.00, p = .099) and for the lateral sites (F (1, 19) = 3.60, p = .073) approached significance. The analysis for the lateral sites yielded a three-way Intonation by Hemisphere by Electrodes interaction (F (10, 190) = 4.39, p < .01). Follow-up analyses indicated a larger N400-like effect for the low tones in questions than in statements at the left central (FC3, C3), temporal (T7), and parietal sites (CP5, P3, P7: all *p*s < .05). The midline and lateral analyses did not show any main effect of Intonation for the high-mid tones (all *F*s < 1).

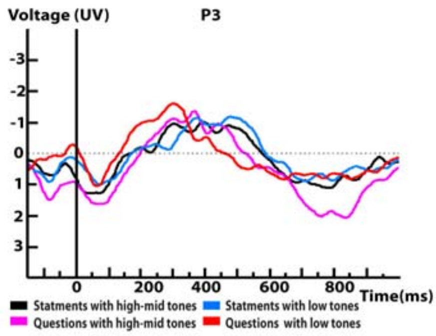

Figure 1: Grand average waveforms for electrode P3 time-locked to the onset of the sentence-final critical word in Experiment 1.

Based on the time-course analyses, the P600-effect was quantified using a 350-550 ms time window. Midline and lateral analyses were performed for the low and high-mid tones. For the low tones, the main effect of Intonation was significant in both the midline (F (1, 19) = 5.40, p < .05) and the lateral analyses (F (1, 19) = 6.03, p < .05). The lateral analysis revealed an interaction of Intonation by Hemisphere by Electrodes (F (10,190) = 5.17, p = .001). Single-site analyses indicated a larger P600-like effect for low tones in questions than in statements at the frontal (F3, F4), central (FC3, FC4, C3, C4), and parietal sites (CP5, CP6, P3, P7, PO7: all *p*s < .05). For the high-mid tones, no main effect was found in the midline and lateral analyses (both *F*s < 1.3). In sum, the biphasic N400-P600 pattern of low tones in questions reflected that these low tones are more difficult to process than the same tones in statements and the high-mid tones in general.

3. Experiment 2

Experiment 1 showed that conflicting F_0 information yielded by tone and intonation give rise to processing difficulties as reflected by high error rates and the biphasic N400-P600 pattern. In daily life, this can lead to huge misunderstanding and can severely impede communication. In order to circumvent this problem, people may have to rely on other cues. One possible candidate is context information. Ye and Connine [10] showed that a highly-constraining semantic context (e.g. idioms) can pre-activate the lexical representation of the forthcoming word and the associated tone information, and in turn, facilitate the processing of tone information. This raises an interesting question: can this pre-activation of tonal information by lexical context counteract the effect of intonation on tone perception? This question was addressed in Experiment 2. The aim of Experiment 2 was to investigate if an immediate lexical context, in addition to the interaction of tone and intonation affects online speech comprehension. The immediate lexical context was created by embedding the critical word in a compound composed of two monosyllabic words. The first member of the compound creates a highly-constraining lexical context that elicits a stronger expectation for the second member (the critical word) as compared to a homophone of the critical word but with another tone. The design of Experiment 2 was the same as that of Experiment 1, except that the critical words were now embedded in two-word compounds situated at the end of questions as in example (5), and statements as in example (6).

(5) nei^{23} t^{h}ɐu^{21} sin^{55}tɐp^{33} p^{h}ou-**fu$^{22-6}$** (**fu$^{25-6}$**)?
"You just answered **aspiration (embrace-bitter)**?"

(6) ŋɔ^{23} t^{h}ɐu^{21} sin^{55} tɐp^{33} p^{h}ou-**fu$^{22-6}$**
"I just answered **aspiration**"

3.1. Methods

The participants, procedure and EEG data acquisition were the same as in Experiment 1. The materials were the same as in Experiment 1, except for the word-stimuli. In Experiment 2, 24 bi-syllabic compounds were used as word stimuli. The second member of the bi-syllabic compounds consisted of the 24 single-word stimuli used in Experiment 1 and the first member consisted of another 24 monosyllabic words (see examples (5) and (6)). As in Experiment 1, the compound stimuli were always situated at the end of questions and statements in Experiment 2. The EEG data were preprocessed and analysed using the same procedure as in Experiment 1. The exploratory time-course analyses revealed Intonation × Electrode interactions between 150 ms and 350 ms ($ps < .05$) as well as a four-way interaction (Intonation × Tone × Hemisphere × Electrode) between 50 ms and 350 ms ($ps < .05$) and between 450 and 650 ms ($ps < .05$).

3.2. Results

In order to test if there was a difference in the percentages of correct identification of tones between Experiment 1 (when the tones were presented alone) and Experiment 2 (when the tones were embedded in compounds), a series of Wilcoxon tests were conducted. These comparisons revealed significant differences for all low tones (21, 23, 22; $ps < .01$) at the end of questions, but not for the high-mid tones (all $ps > .01$). At the end of statements, the low tones (23, 22) had a higher accuracy rate in Experiment 2 than in Experiment 1 (all $ps < .008$). Compared to Experiment 1, significant improvement in the identification of low tones in questions was found in Experiment 2.

Figure 2 shows the grand average waveforms time-locked to the sentence-final critical words at P3. Based on the time-course analyses, the N400 effect was quantified in a 100-350 ms window. Separate midline and lateral analyses were performed for the low tones and for the high-mid tones. For the low tones, a main effect of Intonation was found for the midline sites (F (1, 19) = 7.10, $p < .05$) but not for the lateral sites ($F < 1.2$). The analyses for the lateral sites yielded a three-way interaction of Intonation by Hemisphere by Electrodes (F (10,190) = 8.48, $p < .01$). Single-site analyses indicated a larger N400-like effect for low tones with a question intonation compared with the same tones with a statement intonation at the left centroparietal sites (C3, CP5, P3, P7, PO7: all $ps < .05$).

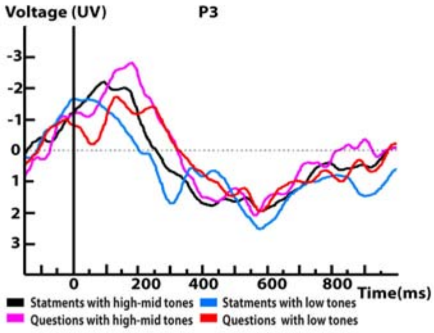

Figure 2: Grand average waveforms for electrode P3 time-locked to the onset of the sentence-final critical word in Experiment 2.

For the high-mid tones, no main effect of Intonation was found for the midline sites (F (1,19) = 2.12, $p > .05$) but an interaction between Intonation and Electrodes was present (F (1, 38) = 8.63, $p = .001$). Single site analyses demonstrated an N400-effect at Pz but not at other sites. Likewise, no main effect of Intonation for the lateral sites ($F < 1$) but an Intonation × Electrode interaction, (F (10,190) = 11.56, $p < .001$), and an Intonation × Hemisphere × Electrode interaction were found (F (10,190) = 13.01, $p < .001$). Single site analyses revealed a larger N400-like effect for high-mid tones with a question intonation compared with the same tones with a statement intonation at frontal (AF7, AF8, F7) and parietal sites (P3, P4, P7, P8, PO7, PO8: all $ps < .05$).

The P600-effect was quantified in the 450-650 ms time window based on the time-course analyses. Separate midline and lateral analyses were performed for the low and high-mid tones. For the low tones, no effect of Intonation was observed (all $Fs < 1$). The analysis for lateral sites revealed an interaction of Intonation × Hemisphere × Electrodes (F (10,190) = 3.18, $p = .01$). However, single-site analyses did not yield any reliable differences. For the high-mid tones, no effect of Intonation was found for midline and lateral sites (both $Fs < 1.5$.). In sum, a larger N400-effect for questions than statements was found in Experiment 2 but this time in the absence of a P600-effect.

4. General discussion

Experiments 1 and 2 provide evidence for the simultaneous use of tone, intonation and context information during online speech comprehension. In Experiment 1, we focused on whether tone perception is affected by intonation when F_0 information of tone and intonation becomes available at the same time; and if so, how this affects lexical processing. In Experiment 2, we investigated if the immediate lexical context also plays a vital part in online speech comprehension.

Consistent with earlier findings [3], we found that tone perception is influenced by intonation contour when tone and intonation provide conflicting F_0 information in a semantically neutral context (Experiment 1). The reasons for the poor perception may be that listeners use the F_0 contour as the major acoustic cue for tone perception and that the F_0 contour of these low tones was altered by the rising intonation contour into a rising contour [3]. This is reflected by the misperception of the target low tone as the high-rising tone in 53.6%-78% of the cases and by the much more accurate perception of the high-mid tones at the same sentence position.

The present study builds on previous studies, which showed that intonation influences tone perception and that tone information is immediately accessed and used during lexical processing [4, 7]. Our goal was to further examine whether, and if so, how a tone-language speaker's brain deals with conflicting information due to the diverging F_0 movements of the lexical tone and the intonation contour. The main finding of Experiment 1 was that low tones with a question intonation (the condition which yields a maximal conflict between the information provided by the tone and intonation) elicit a biphasic N400-P600 pattern. No such pattern was observed for the other conditions. Consistent with previous ERP studies on Chinese dialects [4, 7], the present ERP results support the position that conflicting F_0 information provided by tone and intonation is immediately accessed and used for speech processing. This was signalled by an N400-effect elicited by low tones at the end of questions with an onset as early as 74 ms. The N400-effect here is interpreted to reflect a failure in processing the available lexical information (e.g. [6, 7]). This can be caused by the activation of two equally plausible lexical representations (the word with the target tone and the homophone but with a high-rising tone) by the F_0 information provided by low tones in questions. This failure leads to a reanalysis of the input to check for the possibility of a perceptual error, which is signalled by the P600-effect (e.g. [5]).

In Experiment 2, we expanded on Experiment 1 and showed that an immediate highly-constraining lexical context can modulate the interaction between tone and intonation in speech comprehension. In terms of lexical identification, there was a significant improvement in the accuracy rate, especially for low-toned words at the end of questions. The misperception rate of the low tones as the high-rising tone was also clearly reduced even though the F_0 information provided by tone and intonation was still as conflicting as it was in Experiment 1. In terms of the ERPs, the presence of an immediate context in Experiment 2 give rise to an N400-like effect but no P600-effect.

The difference in the findings between Experiments 1 and 2 can be explained by the role of context information in Chinese speech comprehension. Comprehension in Cantonese, and other Chinese dialects, relies more strongly on contextual information than non-tonal Indo-European languages due to the large number of homophones in the language [11] (on average, 7.6 Chinese characters share the same syllable, and on top, 2.95 characters share the same tone [12]). In addition, processing of tone information depends on the immediate input from the lexical context by means of pre-activating the tone information of the forthcoming word [10]. Thus, it is plausible that in Cantonese, context information immediately modulates speech processing once the information becomes available because context information is required to disambiguate homophones and to arrive at the correct interpretation of the word. This can explain the sizeable improvement in tone perception in Experiment 2.

The above explanations can also account for the differences in the ERP results between the two experiments. First of all, compared to the condition where there is no lexical context, it can be more effortful to process the combined F0 information of tone and intonation in a lexical context because there are more sources of information which have to be integrated. In addition to processing the F_0 information from tone and intonation, the combined F_0 information also has to be matched with the tonal information pre-activated by the preceding lexical context. This additional processing effort could be reflected by the larger N400-effect found for both high-mid tones and low tones in questions relative to the same tones in statements. Furthermore, the availability of context information at an early stage of lexical processing could explain the absence of a P600-effect when a lexical context was present. Even though it may still be more effortful to process the conflicting F_0 information from tone and intonation, the conflicting F_0 information did not lead to an integration failure, which in turn would have triggered a reanalysis (signalled by P600-effect) in this case, because the immediate lexical context intervened to provide additional information for selecting the target lexical tone. The present findings are consistent with the claim that there is a strong context dependency during online speech comprehension in Cantonese Chinese (e.g. [4, 11])

5. Conclusion

Two experiments were conducted to investigate to what extent the online interplay between tone, intonation and context has an effect on speech comprehension in Cantonese Chinese. The results of Experiment 1 indicated that the interaction between tone and intonation can affect lexical identification and speech processing. The effect is particularly strong when the interaction yields conflicting F_0 information. This was reflected by the high error rates for the low tones in questions and by the biphasic N400-P600 pattern elicited by these low tones. In Experiment 2, we showed that an immediate lexical context can modulate the effect of the interaction between intonation and tone on lexical processing and identification. This is indicated by the significant reduction in the number of errors for low tones with a question intonation and the absence of a P600-effect for these tones. The results highlight the importance of context during speech comprehension in a tone language.

6. References

[1] Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

[2] Laver, J. (1994). *Principles of Phonetics*. Cambridge: Cambridge University Press.

[3] Ma, J. K-Y, Ciocca, V., and Whitehill, T. (2006). Effect of intonation on Cantonese lexical tones. *JASA,* 320, 3978-3987.

[4] Schirmer, A., Tang, S. L., Penney, T. B., Gunter, T. C., and Chen, H. C. (2005) Brain responses to segmentally and tonally induced semantic violations in Cantonese. *Journal of Cognitive Neuroscience*, 17, 1-12.

[5] Van de Meerendonk, N., Kolk, H. H. J., Vissers, C. T. W. M., and Chwilla, D. J. (2010). Monitoring in language perception: Mild and strong conflicts elicit different ERP Patterns. *Journal of Cognitive Neuroscience,* 22, 67-82.

[6] Chwilla, D. J., Kolk, H. H. J., and Mulder, G. (2000). Mediated priming in the lexical decision task: Evidence from event-related potentials and reaction time. *Journal of Memory and Language,* 42, 314-341.

[7] Li, X., Yang, Y., and Hagoort, P. (2008). Pitch accent and lexical tone processing in Chinese discourse comprehension: an ERP study. *Brain Research*, 1222,192-200.

[8] Flynn-Chang, C-Y. (2003) *Intonation in Cantonese*. München : Lincom Europa.

[9] Bögels, S., Schriefers, H., Vonk, W., Chwilla, D. J., and Kerhkofs, R. (in press). The interplay between prosody and syntax in sentence processing: The case of subject- and object-control verbs. *Journal of Cognitive Neuroscience*.

[10] Ye, Y., and Connine, C. M. (1999). Processing Spoken Chinese: The Role of Tone Information. *Language and Cognitive Processes*, 14, 609-630.

[11] Chen, H.-C. (1992). Reading comprehension in Chinese: Implications from character reading times. In H.-C. Chen & O. Tzeng (Eds.) *Language processing in Chinese* (pp. 175-205). Amsterdam: North-Holland.

[12] Zhang, L. Y., & Zhang, S. Y. (1987). *Common Chinese characters pronounced according to Mandarin and Cantonese.* Hong Kong: Zhong Hua.