

# Task and Sex Modulate the Brain Response to Emotional Incongruity in Asian Listeners

Annett Schirmer  
University of Georgia

Ming Lui  
Northwestern University

Burkhard Maess  
Max Planck Institute for Human Cognitive and Brain Sciences

Nicolas Escoffier  
University of Georgia

Mandy Chan and Trevor B. Penney  
Chinese University of Hong Kong

In order to recognize banter or sarcasm in social interactions, listeners must integrate verbal and vocal emotional expressions. Here, we investigated event-related potential correlates of this integration in Asian listeners. We presented emotional words spoken with congruous or incongruous emotional prosody. When listeners classified word meaning as positive or negative and ignored prosody, incongruous trials elicited a larger late positivity than congruous trials in women but not in men. Sex differences were absent when listeners evaluated the congruence between word meaning and emotional prosody. The similarity of these results to those obtained in Western listeners suggests that sex differences in emotional speech processing depend on attentional focus and may reflect culturally independent mechanisms.

*Keywords:* Cantonese, Chinese, gender, affective, tone, voice, vocal, N400, sad, happy, lateralization

Understanding the emotions encoded in speech is a complex skill that requires listeners to perceive and integrate verbal and prosodic information. Prosody, which refers to acoustic properties of speech such as melody, amplitude, and temporal structure, can convey both linguistic and emotional information. It is presumed that although both aspects of prosodic information rely on the same acoustic parameters, the encoding stages that follow initial acoustic analysis are lateralized differently in the brain (Schirmer, Alter, Kotz, & Friederici, 2001; Schirmer & Kotz, 2006). With respect to the perception of emotional prosody, both brain lesion (Blonder, Bowers, & Heilman, 1991; Van Lancker & Fromkin, 1973; Van Lancker, 1980) and functional imaging research (Buchanan et al., 2000; Gandour et al., 2003; George et al., 1996) suggest a right hemisphere lateralization. Additionally, subcortical structures such as the basal ganglia (Kotz et al., 2003; Pell & Leonard, 2003) have been implicated. Language processing, on the

other hand, seems to rely more strongly on the left hemisphere (Blonder et al., 1991; Broca, 1865; De Renzi, Liotti, & Nichelli, 1987; Kuperberg et al., 2000; Taubner, Rayner, & Heilman, 1999; Thompson-Schill, D'Esposito, & Kan, 1999), although certain language processes such as the drawing of inferences, joke comprehension, and the evaluation of affective content also involve right hemisphere structures (Brownell, Michel, Powelson, & Gardner, 1983, 1986; Atchley, Ilardi, & Enloe, 2003).

In addition to the neuroanatomical underpinnings of spoken language processing, temporal aspects have been of interest. Event-related potentials (ERPs) have been particularly useful for determining when in time listeners retrieve word information from memory and integrate this information into the preceding sentence context. There is general agreement that these retrieval and integration processes are reflected by the negative ERP component termed the N400. The N400 peaks approximately 400 ms following word onset and is larger for words that are incongruous in a given sentence context compared to congruous (Hahne & Friederici, 2002; Friederici, Pfeifer, & Hahne, 1993; Kutas & Hillyard, 1980, 1984; Kutas, Van Petten, & Besson, 1988; Van Petten, Coulson, Rubin, Plante, & Parks, 1999). Based on these observations one can conclude that sentence context modulates word processing approximately 400 ms following word onset.

Recently, researchers have begun to investigate when and how emotional prosody modulates word processing. Similar to what was found for sentence context, this research revealed a larger N400 for words presented with incongruous compared to congruous emotional prosody (e.g., "succeeded" spoken with angry compared to happy prosody; Schirmer, Kotz, & Friederici, 2002;

---

Annett Schirmer and Nicolas Escoffier, Department of Psychology, University of Georgia, Athens, Georgia; Ming Lui, Department of Psychology, Northwestern University, Evanston, Illinois; Burkhard Maess, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany; Mandy Chan and Trevor B. Penney, Department of Psychology, Chinese University of Hong Kong, Hong Kong, China.

This project was supported by Grant 4322/01H awarded to T.B. Penney by the Research Grants Council of the Hong Kong SAR.

Correspondence concerning this article should be addressed to Dr. Annett Schirmer, Department of Psychology, Faculty of Arts and Social Sciences, University of Singapore, Block AS6, Level 3, 11 Law Link, Singapore. 117570. E-mail: psysa@nus.edu.sg

Schirmer & Kotz, 2003). Furthermore, vocal emotional exclamations (e.g., “ooh” spoken with a happy voice) that follow words with incongruous (e.g., “disgust”) compared to congruous emotional valence (e.g., “pleasure”) elicit an N400-like effect (Bostanov & Kotchoubey, 2004). Additionally, incongruity between word valence and emotional prosody has been found to activate the inferior frontal gyrus more strongly than congruity (Schirmer, Zysset, Kotz, & von Cramon, 2004). Given that analysis of semantic incongruity draws on the inferior frontal gyrus (Kang, Constable, Gore, & Avrutin, 1999; Kiehl, Laurens, & Liddle, 2002; Kuperberg et al., 2000; Newman, Pancheva, Ozawa, Neville, & Ullman, 2001; Ni et al., 2000), these results suggest a common mechanism for the integration of purely semantic information and the integration of semantic and emotional-prosodic information.

A further question with respect to emotional speech processing is how the listeners’ sex and the attention that listeners allocate to emotional prosody modulate the integration of semantic and emotional-prosodic information. Schirmer and colleagues (Schirmer, Kotz, & Friederici, 2002) addressed this question in a cross-modal priming study that required participants to perform lexical decisions to positive and negative visual target words that followed a sentence spoken with congruous or incongruous emotional prosody. Only female participants responded slower and showed a larger N400 to words that followed a sentence with incongruous compared to congruous emotional prosody. A similar effect in male participants appeared only if the interval between sentence offset and target onset was increased (Schirmer, Kotz, & Friederici, 2002) or they were asked to judge the emotional congruence between sentence prosody and target valence (Schirmer, Kotz, & Friederici, 2005). Subsequent work using individually presented spoken words revealed similar results. Words spoken

with incongruous emotional prosody elicited a larger N400 (see Figure 1, Schirmer & Kotz, 2003) and larger activity in the inferior frontal gyrus (Schirmer, Zysset, Kotz, & von Cramon, 2004) than words with congruous emotional prosody. Moreover, when asked to evaluate word meaning while ignoring prosody, this effect was present in women but not in men. Together, these findings suggest that women integrate emotional prosody more readily into language processing than men, but that men and women may show comparable emotional-prosodic processing when emotional prosody is relevant to the task at hand. This conclusion is in accord with behavioral reports suggesting a greater sensitivity to nonverbal emotional cues in women than in men (Hall, 1978). However, it is unclear whether it applies to male and female listeners from different cultures because the observed effects of attention and sex on emotional speech processing may be specific for the German culture and/or language for which they have been reported.

Cross-cultural research by Kitayama and Ishii (2002; see also Ishii, Reyes, & Kitayama, 2003) suggests that this is a relevant concern. The authors found that emotional prosody influenced the emotional evaluation of words in Japanese, but not in American listeners. These findings suggest that culture modulates the use of emotional prosody. Moreover, they fit well with the idea that Asian and Western cultures differ with respect to the role of contextual information in communication. Compared to Western cultures, Asian cultures are considered to express less information verbally and to rely more on contextual cues when interpreting communication (Hall, 1976). Moreover, because of the more restrictive emotional display rules (Matsumoto, 1990; Matsumoto, Kasri, & Kookan, 1999), the contextual cues provided by Asian interaction partners are very subtle so that great sensitivity is required for their perception and interpretation. As this sensitivity is expected of both male and female interaction partners, sex differences in emotional-prosodic processing observed in a Western culture (i.e., German) may be absent in Asians. If true, this would suggest that cultural background rather than the listeners’ sex modulates the perception and integration of emotional prosody during language processing. If, however, Asian men and women differ with respect to emotional-prosodic processing, then culturally independent (i.e., biological) mechanisms may underlie the observed sex differences.

We addressed this question by examining ERP correlates of emotional speech processing in Cantonese listeners. Cantonese is a Chinese dialect spoken in the southern provinces of China, as well as Hong Kong and Macau. It is a tone language comprising six lexical tones that are used to disambiguate the meaning of otherwise identical words. In the present study, Hong Kong Cantonese listeners were presented with spoken words that always carried the proper lexical tone. However, on half of the trials, the emotional prosody of the speaker was incongruous with the emotional meaning of the word (e.g., “success” spoken with sad prosody) while for the remaining trials emotional prosody and word meaning were congruous. In Experiment 1, participants evaluated the emotional meaning of each word while ignoring its emotional prosody. If culture independent mechanisms underlie the processing and integration of emotional prosody, Experiment 1 should replicate the findings observed in Western listeners. That is to say, women, but not men, should show a larger N400 for words with incongruous compared to congruous emotional prosody.

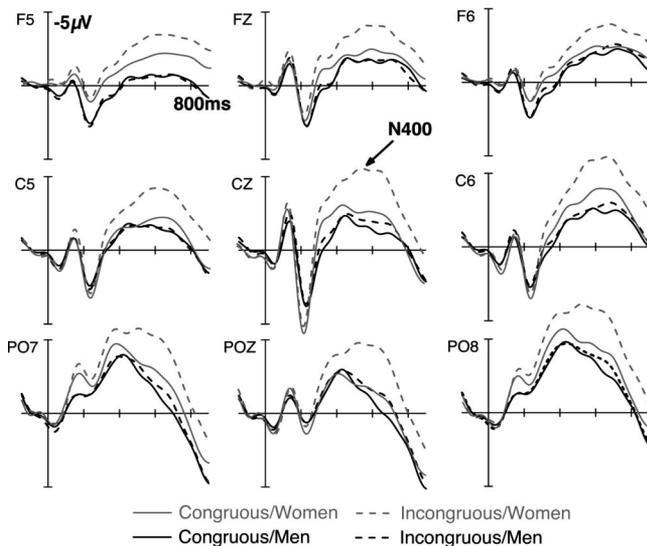


Figure 1. ERP results from Schirmer & Kotz, 2003 (reprinted with permission). Negativity is plotted upwards. When words were evaluated for their emotional meaning, incongruous emotional prosody enhanced the N400 amplitude relative to congruous emotional prosody in women but not in men. This effect was significant between 350 and 900 ms following word onset.

## Experiment 1

### Methods

**Subjects.** Fifty-three Chinese University of Hong Kong (CUHK) students were invited to participate in Experiment 1. Five participants were excluded because of excessive eye-movement artifacts in the electroencephalogram (EEG). Half of the remaining participants were male with a mean age of 20.67 years ( $SD$  1.13). Female participants had a mean age of 20.58 years ( $SD$  0.83). All participants were right-handed, native Cantonese speakers who grew up in Hong Kong. They were compensated at a rate of HK\$50 an hour for their time.

**Material.** The stimulus material consisted of 68 positive and 68 negative disyllabic Cantonese verbs. All words were nonhomophones. Word valence, word familiarity, and word type were assessed in a separate rating study by 40 CUHK undergraduate students, none of whom participated in the subsequent ERP experiment. These participants rated all words on a 5-point scale ( $-2 =$  very negative,  $-1 =$  negative,  $0 =$  neutral,  $+1 =$  positive,  $+2 =$  very positive) for emotional valence. The rating of positive words ( $M + 1.27$ ;  $SD$  0.19) differed significantly from the rating of negative words ( $M - 1.28$ ;  $SD$  0.22), whereas the absolute rating values were identical. As there is no complete word frequency database for the Cantonese language, we used a word familiarity rating to assess the frequency of words in everyday usage. Word familiarity was rated on a six-point-scale ( $1 =$  extremely seldom,  $2 =$  very seldom,  $3 =$  sometimes,  $4 =$  quite often,  $5 =$  very often,  $6 =$  extremely often heard or spoken). Positive words ( $M$  3.45,  $SD$  0.64) were rated equally familiar as negative words ( $M$  3.61,  $SD$  0.58). A word type scale was constructed to assess whether the word stimuli were perceived as verbs or nouns. In Cantonese many words can be used as either verbs or nouns and the rating ensured that the probability of listeners perceiving a verb was comparable for positive (75.6%,  $SD$  0.12) and negative words (75.87%,  $SD$  0.14). Finally, we controlled for abstractness. Words were considered concrete when they referred to object manipulation (e.g., to write), whereas words were considered abstract when they referred to psychological states (e.g., to think; Perani et al., 1999). Twenty-two positive and 25 negative words were considered concrete. Forty-eight positive and 45 negative words were considered abstract.

A female native Cantonese speaker with drama experience produced all words once with happy and once with sad prosody. The words were recorded with a DAT recorder and digitized at a 16-bit per 44.1 KHz sampling rate. The recorded stimuli were presented to 40 CUHK students, none of whom participated in the preceding rating study or the subsequent ERP experiment. Listeners were asked to recite and write down each word and to rate its emotional-prosodic expression on a 5-point scale ( $1 =$  very sad,  $2 =$  sad,  $3 =$  neutral,  $4 =$  happy,  $5 =$  very happy). The mean ratings for happy (3.76,  $SD = .33$ ) and sad prosody (1.68,  $SD = 0.2$ ) were significantly different. Words that could not be recited or that were neutral in prosody were re-recorded by the female speaker. The mean duration of happily spoken positive words was 621 ms ( $SD = 125$ ) and that of negative words was 605 ms ( $SD = 134$  ms). The mean duration of sadly spoken positive words was 977 ms ( $SD = 146$ ) and that of negative words was 961 ms ( $SD = 128$ ).

**Procedure.** Participants were seated in a chair facing a computer monitor at a distance of 100 cm. Words were presented via headphones. Participants were instructed to evaluate the emotional meaning of each word as quickly and as accurately as possible while ignoring its emotional prosody. They gave their responses by pressing one of two buttons on a hand-held response pad. Button (i.e., positive vs. negative) to response hand assignment was counterbalanced across participants. Three-hundred milliseconds (ms) before each word, a fixation cross appeared in the middle of the monitor and remained onscreen for 2.8 seconds. Participants were asked to fixate the cross and to prevent eye movements and eye-blinks for as long as they saw the cross. The stimulus onset asynchrony (SOA) was equally distributed across values of 3,000, 3,100, and 3,200 ms. Each

participant completed 272 trials (68 positive words with happy and sad prosody, 68 negative words with happy and sad prosody) which included the repetition of the same word spoken once with congruous and spoken once with incongruous emotional prosody. This allowed us to counterbalance words across congruous and incongruous conditions and to increase the signal to noise ratio. One shortcoming of this approach is that memory effects may have influenced our results. However, we controlled for this possibility by including word repetition as a factor in the statistical analysis (see below). Trials were presented in four blocks. The order of trials was pseudo-randomized with the constraint that words were not repeated within one half of the experiment. In order to familiarize participants with the task, the experimental session started with twelve practice trials.

**Electrophysiological recordings.** The EEG was recorded from 50 electrodes with a bandpass between DC and 70 Hz and digitized at a 256 Hz sampling rate. The reference electrode was placed on the nose tip. In order to control for horizontal and vertical eye movements, a bipolar electrooculogram was recorded using four electrodes. Electrode resistance was kept below 5 kOhm. ERP averages were computed with a 150 ms baseline and a 2,000 ms ERP time window. Trials containing eye blinks or movement artifacts as well as trials that elicited an incorrect response were omitted from the data analysis. This resulted in a mean of 54.4 trials ( $SD$  8.62) per condition (congruous—first presentation; incongruous—first presentation; congruous—second presentation; incongruous—second presentation) in women and 51.4 trials ( $SD$  10.84) in men. Grand averages were smoothed with a 10 Hz low pass filter for illustration only.

**Data analysis.** Behavioral and EEG measures for the two experiments were analyzed separately. For EEG measures, an exploratory statistical analysis was conducted with 50 ms time window ANOVAs with repetition (first presentation/second presentation), congruence (congruous/incongruous), hemisphere (left/right), and site (anterior/posterior) as repeated measures factors and sex as a between-subjects factor. The factors hemisphere and site were created by grouping the electrodes as follows: left anterior—FP1, AF3, F7, F5, F3, FT7, FC5, FC3, FC1; left-posterior—TP7, CP5, CP3, CP1, P7, P5, P3, PO3, O1; right anterior—FP2, AF4, F8, F6, F4, FT8, FC6, FC4, FC2; right posterior—TP8, CP6, CP4, CP2, P8, P6, P4, PO4, O2. Electrode recordings within these groups were averaged to derive a representative, mean electrode for each cortical region. Because a large number of analysis time windows is associated with an increased likelihood of Type I errors, only effects that reached statistical significance in more than two consecutive time windows were considered reliable. The family-wise error rate for this correction procedure was estimated with a bootstrap technique (Efron & Tibshirani, 1993), which revealed a probability of 0.0056 for obtaining more than two consecutive false positive tests.

Significant effects of the initial analysis were followed up with ANOVAs, which collapsed statistically significant consecutive time windows into one larger time window. Only significant main effects and interactions from these follow-up analyses are reported. Main effects and interactions were further investigated with ANOVAs. Type I errors associated with multiple comparisons were controlled by using Tukey's studentized range test for means where appropriate.

### Results of Experiment 1

**Behavioral results.** Reaction time and accuracy for the word valence judgments are presented in Figure 2. To analyze effects of emotional-prosodic congruence, separate ANOVAs treating congruence and repetition as repeated measures factors and sex as a between subjects factor were conducted for reaction time and accuracy. The analysis of reaction time revealed a main effect of congruence [ $F(1, 46) = 24.8, p < .0001$ ] and repetition [ $F(1, 46) = 65, p < .0001$ ] while the congruence\*repetition interaction was nonsignificant ( $p > .1$ ). Congruous presentations were responded to faster than incongruous presentations. Stimulus repetition facilitated word valence judgments. The analysis of accuracy revealed a main effect of congruence indicating that participants responded more

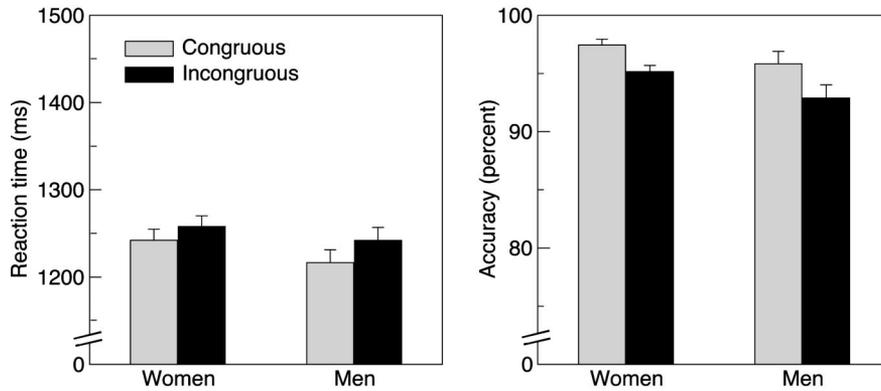


Figure 2. Mean reaction time and accuracy rates (with standard error bars) for the word valence judgment in Experiment 1. Both men and woman responded faster and more accurately to words with congruous compared to incongruous emotional prosody. There was a tendency for women to be more accurate than men.

accurately to congruous compared to incongruous presentations [ $F(1, 46) = 25.5, p < .0001$ ]. This effect was modulated by repetition, as indicated by a significant congruence\*repetition interaction [ $F(1, 46) = 4, p = .05$ ]. Follow-up analysis of the congruence\*repetition interaction showed that the congruence effect was only marginally significant during first presentations [ $F(1, 46) = 3.5, p = .07$ ], but significant during second presentations [ $F(1, 46) = 39.8, p < .0001$ , Tukey  $p < .05$ ].

**ERP results.** ERP traces for the congruous and the incongruous condition are presented in Figure 3. Word presentations elicited a late negativity that peaked approximately 750 ms following stimulus onset. In contrast to previous studies with Western listeners, the amplitude of this negativity failed to differ between words with congruous and incongruous emotional prosody. A subsequent positivity was larger for words with

incongruous compared to congruous emotional prosody. As in Western listeners, this effect seemed present in women, but not in men.

The results of the exploratory, 50 ms time window analysis of the ERP data are presented in Figure 4. ERPs were more negative over anterior compared to posterior electrodes in the 0 to 450 ms [ $F(1, 46) = 48, p < .0001$ ], 550 to 1450 ms [ $F(1, 46) = 25.7, p < .0001$ ], and 1700 to 2000 ms time range [ $F(1, 46) = 9.9, p < .01$ ]. The hemisphere\*sex interaction [ $F(1, 46) = 6.2, p < .05$ ] was significant between 300 and 600 ms following word onset. The time window of this interaction covers the slope of the late negativity, which was lateralized to the right hemisphere in women [ $F(1, 23) = 13.7, p < .01$ , Tukey  $p < .05$ ], but not in men ( $p > .1$ ). After 600 ms, there was a main effect of hemisphere [400 to 1250 ms,  $F(1, 46) = 14.1, p < .001$ ; 1650 to 1850 ms,  $F(1, 46) = 6, p < .05$ ], whereas the

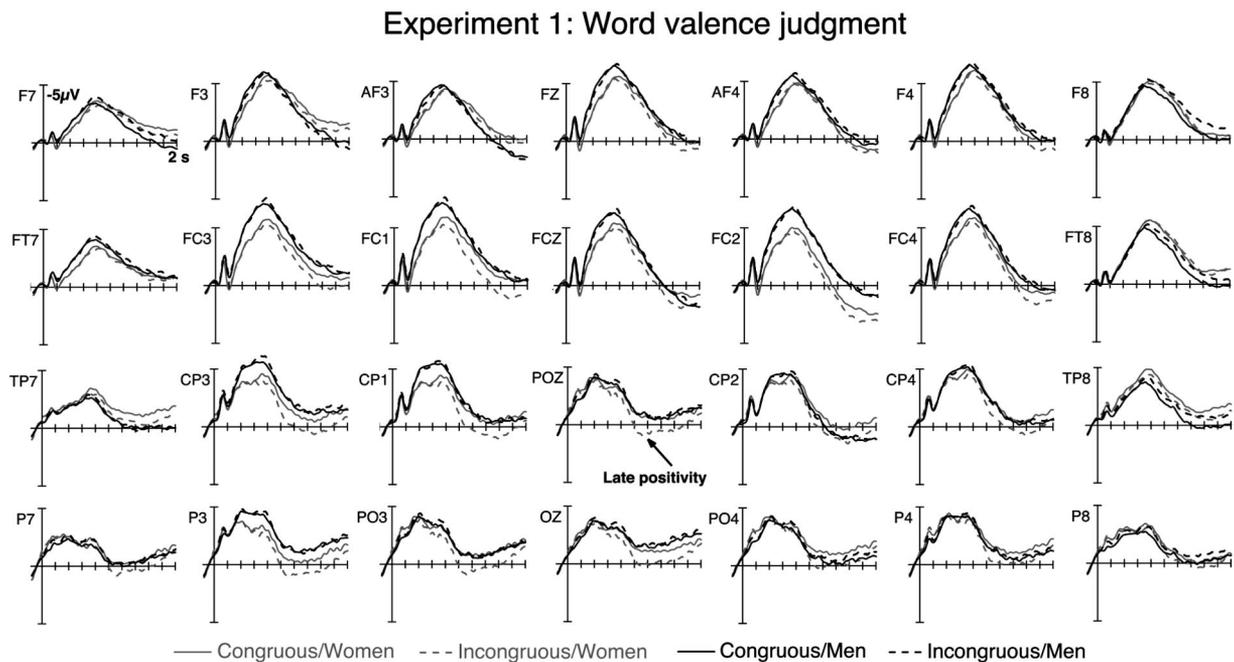


Figure 3. ERPs time locked to the onset of words presented in Experiment 1. Negativity is plotted upwards. Between 1150 and 1900 ms following word onset, women, but not men, showed a larger late positivity for words with incongruous compared to words with congruous emotional prosody.

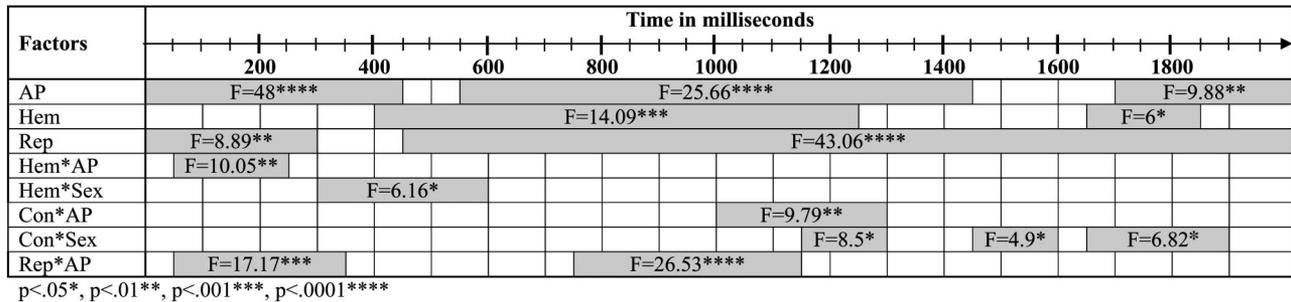


Figure 4. Effects that reached significance in the exploratory 50-ms time window statistical analysis for Experiment 1. The far left column lists significant main effects and interactions. The remaining columns illustrate the time range of significant main effects and interactions in milliseconds. For example, the congruence by sex interaction was significant between 1,150 ms and 1,300 ms. The ANOVA conducted on mean voltages of the 1.15- to 1.3-s time window revealed an  $F$ -value of 8.5 for this interaction.

hemisphere\*sex interaction was nonsignificant ( $p > .1$ ). Thus, starting at about 600 ms both men and women showed a right-hemisphere lateralization of the late negativity and the subsequent positivity. An interaction of the factors congruence and sex modulated the positivity between 1150 and 1300 ms [ $F(1, 46) = 6.9, p < .01$ ], between 1450 and 1600 ms [ $F(1, 46) = 6.1, p < .05$ ], and between 1650 and 1900 ms [ $F(1, 46) = 7.7; p < .05$ ]. Analysis of the congruence effect for each level of sex revealed that women [ $F_s(1, 23) = 5.3, 5.3, 5.7$ , respectively, Tukey all  $ps < .05$ ], but not men (all  $ps > .1$ ), showed a significantly larger positivity in response to words with incongruous compared to congruous emotional prosody across all three time windows. These congruence effects were unaffected by repetition. Moreover, repetition reduced the amplitude of the late negativity and increased the amplitude of the late positivity for both emotionally congruous and incongruous words [0 to 300 ms,  $F(1, 46) = 8.9, p < .01$ ; 450 to 2000 ms,  $F(1, 46) = 43.1, p < .0001$ ].

## Discussion

Experiment 1 examined when Cantonese listeners integrate verbal and vocal emotional information and whether this integration is modulated by the listeners' sex. ERP-results indicated that initial word processing was unaffected by congruence between emotional prosody and word valence in both men and women. A late posteriorly distributed positivity that peaked approximately 1,500 ms following word onset was larger for incongruous compared to congruous emotional prosody in women but not in men. Behaviorally, both sexes responded slower to words with incongruous compared to congruous emotional prosody. Word repetition had little to no influence on these effects. Moreover, repetition increased differences in accuracy, but not reaction time, between words with congruous and incongruous emotional prosody. This increase may be because of an influence of emotional congruence on mnemonic processes. However, this interpretation is called into question by the finding that word repetition failed to modulate ERP congruence effects. Thus, one may speculate that, during the course of the experiment, participants' behavioral decisions became less resistant to influences of task-irrelevant emotional prosody.

The present findings differ in several respects from what has been reported in listeners with a Western culture and language background (e.g., Schirmer & Kotz, 2003). First, the congruence of emotional prosody failed to modulate the N400 amplitude. Moreover, the negativity elicited by words in the present study

shared few characteristics with the N400 previously reported for auditory word processing in Western listeners. In contrast to the centro-parietal distribution known for the N400 in Western listeners (Hagoort & Brown, 2000; Hahne & Friederici, 2002; Holcomb & Anderson, 1993; Friederici et al., 1993), the present negativity was maximal over right anterior regions. Moreover, its maximum occurred at 750 ms rather than at 400 ms following word onset. Nevertheless, we believe that both the present negativity and the classical N400 reflect similar processes, and that differences between the two can be explained by language features, which we will describe in more detail in the General Discussion.

A second difference between the present and previous studies of emotional speech processing concerns the latency of the congruence effect. Rather than modulating the N400-like negativity, emotional congruence modulated a subsequent positive ERP component that peaked approximately 1,500 ms following word onset. This late positivity was larger when prosody and word meaning were incongruous compared to congruous. Given an average word length of 790 ms, the long latency of this effect suggests that words had already been processed for their meaning before they were integrated with emotional prosody. Based on these results one might conclude that emotional prosody plays a subordinate role in Cantonese language processing. However, such a conclusion may seem in conflict with the presumed high sensitivity to nonverbal cues in Asian listeners. We propose that it is not. Previous cross-cultural research on emotional processing relied on behavioral measures. In accord with this research (Kitayama & Ishii, 2002), Cantonese listeners in the present study were highly sensitive to emotional prosody as their reaction time increased when emotional prosody failed to match word meaning. Furthermore, the late positivity effect in the ERP does not imply that listeners were late in accessing the emotional information expressed through prosody, rather it indicates that they were late in using this information for language processing. In other words, Cantonese listeners may have accessed emotional prosody early, but the processing of this information occurred independently from language processing up until several hundred milliseconds following word offset. We assume that the delayed integration of emotional prosody and word meaning is associated with the high significance of linguistic prosody or tone for Cantonese word processing. Because Cantonese listeners have to rely on linguistic tone for word compre-

hension, they may postpone the matching of a speaker's emotional prosody with the verbal message.

Beside these differences between the present and previous work on emotional speech processing, there is one important commonality. In both Western and Asian culture, women seem to be more susceptible to influences from emotional prosody during language processing. The present emotional-prosody congruence effect in the ERPs was significant in female listeners only. This suggests that compared to men women directed more processing resources to the integration of emotional prosody and word meaning. Note that this does not necessarily mean that men and women differed in how they accessed prosodically conveyed emotional meaning. On the contrary, the behavioral results indicate that this was comparable as both sexes showed longer response latencies and lower accuracy for incongruous compared to congruous trials. These behavioral effects likely reflect a response conflict for incongruous trials, where emotional prosody and word valence may have activated both the correct and the incorrect response. Such a conflict has been reported previously and has been found to be independent of the listeners' sex (Schirmer & Kotz, 2003; Schirmer et al., 2004).

To date, sex differences in the integration of emotional prosody into language processing have been reported only when either emotional prosody or word meaning were task irrelevant (Schirmer et al., 2002, 2004; Schirmer & Kotz, 2003). In contrast, a cross-modal priming study indicated that men and women use emotional prosody in a similar way when asked to evaluate the congruence between prosody and word meaning (Schirmer, Kotz, & Friederici, 2005). This suggests that attention can facilitate the integration of verbal and nonverbal information in communication and eliminate existing sex differences. We conducted a second experiment to test whether attention has a similar effect on emotional speech processing in a paradigm that uses individually presented spoken words and in listeners from an Asian culture. We hypothesized that if we instructed participants to evaluate the congruence between prosody and word meaning, then both men and women would show an integration effect in the ERP.

## Experiment 2

### Method

**Participants.** For Experiment 2, 36 CUHK students were invited. Four participants were excluded due to excessive eye-movement artifacts in the EEG. Half of the remaining participants were male with a mean age of 20.12 ( $SD$  1.09). Female participants had a mean age of 20.5 ( $SD$  0.89). All participants were right-handed, native Cantonese speakers who grew up in Hong Kong. They were compensated at a rate of HK\$50 an hour for their time.

**Procedure.** The stimulus material, electrophysiological recordings, and data analyses were identical to Experiment 1. The procedure was different in that participants were asked to indicate whether the emotional prosody of a word and its emotional meaning were congruous or incongruous. They responded by pressing the left or right button on a response box. Button assignment was counterbalanced across participants. As in Experiment 1, trials that elicited an incorrect response as well as trial containing eye blinks or movement artifacts were omitted from the data analysis. This resulted in a mean of 52.0 trials ( $SD$  9.59) per condition (congruous—first presentation; incongruous—first presentation; congruous—second presentation; incongruous—second presentation) in women and 46.81 trials ( $SD$  11.64) in men.

### Results of Experiment 2

**Behavioral results.** Reaction time and accuracy for the word valence judgments are presented in Figure 5. To analyze effects of emotional-prosodic congruence, separate ANOVAs with congruence and repetition as repeated measures factors and sex as a between subjects factor were conducted for reaction time and accuracy. Analysis of reaction time revealed a sex main effect [ $F(1, 30) = 4.34, p < .05$ ]. Women responded faster than men. Furthermore, both sexes responded faster to congruous than to incongruous trials [congruence:  $F(1, 30) = 59.28, p < .0001$ ] and to repeated compared to first presentations [repetition:  $F(1, 30) = 77.05, p < .0001$ ]. Additionally, a significant congruence\*repetition interaction [ $F(1, 30) = 5.19, p < .05$ ] indicated that the reaction time advantage for congruous over incongruous trials was more pronounced during first [ $F(1, 30) = 40.45, p < .0001$ , Tukey  $p < .05$ ] compared to repeated presentations [ $F(1, 30) = 29.37, p < .0001$ , Tukey  $p < .05$ ]. Analysis of accuracy failed to reveal a main effect of sex and congruence ( $ps > .1$ ). How-

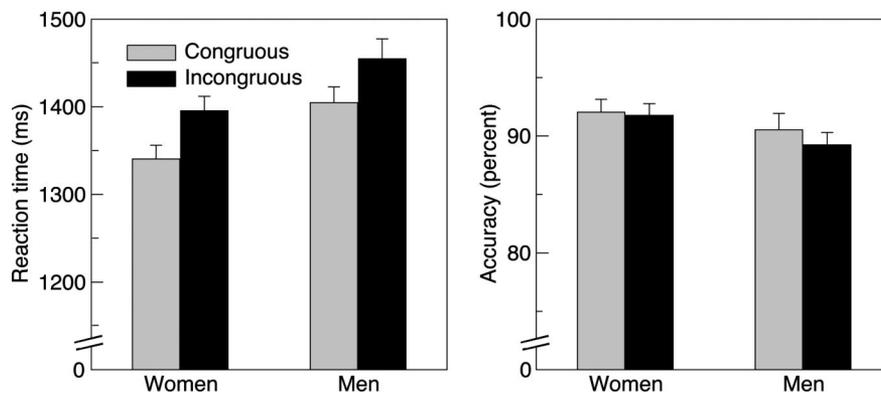


Figure 5. Mean reaction time and accuracy rates (with standard error bars) for the congruence judgment in Experiment 2. Both men and women responded faster and more accurately to words with congruous compared to incongruous emotional prosody. Women responded faster than men.

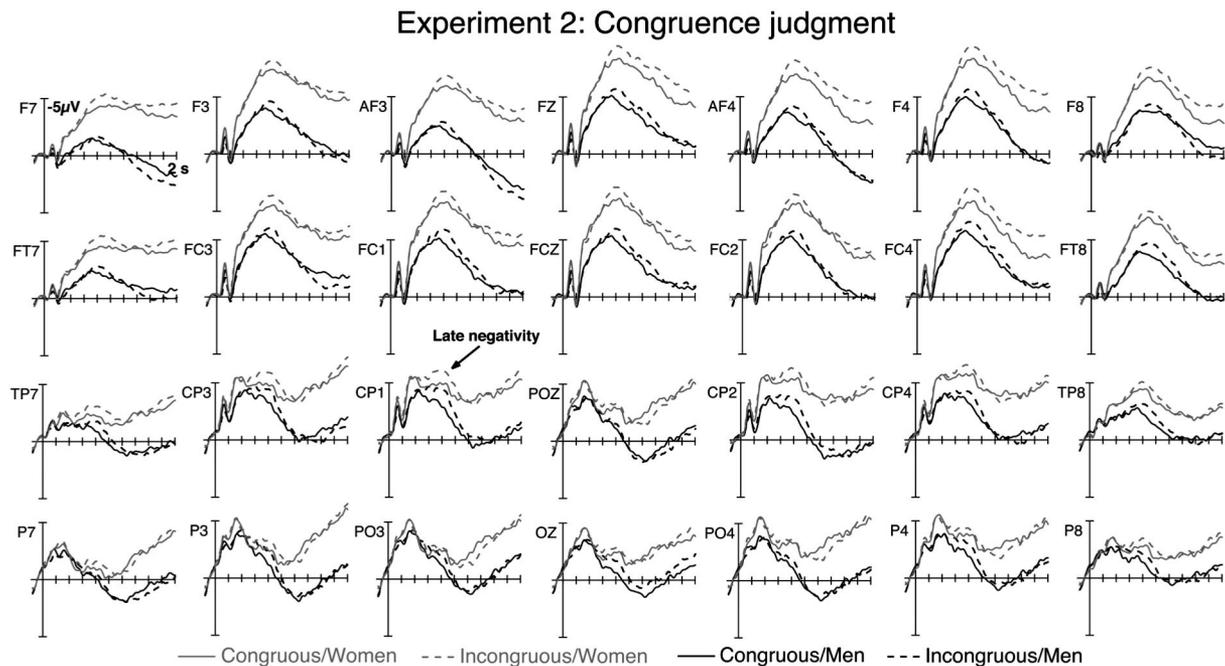
ever, similar to reaction time, there was a significant main effect of repetition [ $F(1, 30) = 46.88, p < .0001$ ] and a congruence\* repetition interaction [ $F(1, 30) = 13.38, p < .01$ ]. Follow-up comparisons indicated that at first presentations participants were tendentially more accurate to congruous compared to incongruous trials [ $F(1, 30) = 2.95, p < .09$ ] and that this difference reached significance for second presentations [ $F(1, 30) = 12.94, p < .01, \text{Tukey } p < .05$ ].

**ERP results.** ERP traces for the congruous and the incongruous conditions are presented in Figure 6. As in Experiment 1, word presentations elicited a late negativity that peaked approximately 750 ms following stimulus onset and had a right-frontal distribution. In contrast to Experiment 1, the amplitude of this negativity differed between words with congruous and incongruous emotional prosody in both men and women.

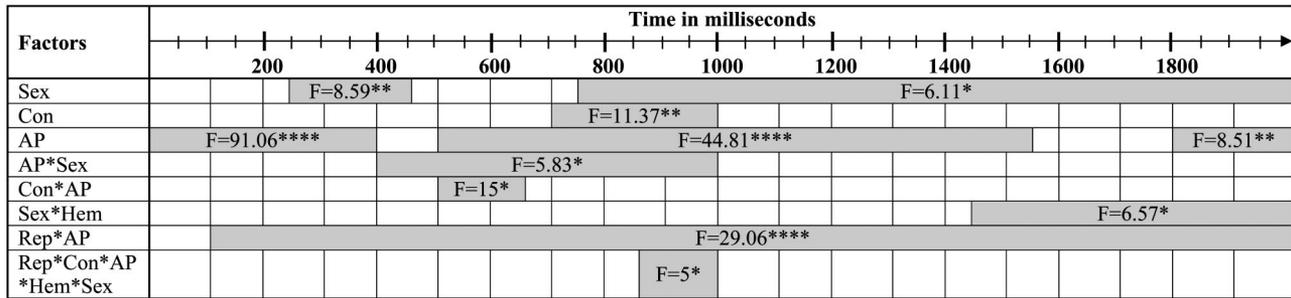
The results of the exploratory statistical analysis of the ERP data are presented in Figure 7. This analysis revealed a significant congruence\*AP interaction [ $F(1, 30) = 14.96, p < .001$ ] between 500 and 650 ms. Post hoc comparisons indicated that the congruence effect was significant over posterior electrodes [ $F(1, 30) = 5.8, p < .05, \text{Tukey } p < .05$ ] and was nonsignificant over anterior electrodes ( $p > .1$ ). Between 700 and 1,000 ms, the congruence effect was significant independently of AP [ $F(1, 30) = 11.37, p < .01$ ]. During both the 500 to 650 ms and the 700 to 1,000 ms time intervals, incongruous trials elicited a significantly larger negativity than congruous trials. Starting at 850 ms and continuing until 1,000 ms following word onset, the congruence effect was modified by repetition, hemisphere, AP, and sex [ $F(1, 30) = 5.92, p < .05$ ]. Follow-up analysis revealed a significant repetition\* congruence\*hemisphere\*AP interaction in women [ $F(1, 30) = 5.4, p < .05$ ], but not in men ( $p > .1$ ). Moreover, in women

Congruence interacted with Hemisphere and AP during second [ $F(1, 15) = 5.64, p < .05$ ] but not first presentations ( $p > .1$ ). During second presentations, the effect was largest over right anterior electrode sites [anterior:  $F(1, 15) = 4.53, p = .05$ ; anterior-left:  $F(1, 15) = 6.95, p < .05, \text{Tukey } p < .05$ ; anterior-right:  $F(1, 15) = 20.9, p < .001, \text{Tukey } p < .05$ ; posterior:  $p > .1$ ]. The factor AP reached significance between 0 and 400 ms [ $F(1, 30) = 91.06, p < .0001$ ], 500 and 1550 ms [ $F(1, 30) = 44.81, p < .0001$ ], and between 1,800 and 2,000 ms [ $F(1, 30) = 8.51, p < .01$ ] following word onset. The negativity was larger and the late positivity smaller over anterior compared to posterior electrode sites. The factor AP interacted with sex between 400 and 1000 ms [ $F(1, 30) = 5.83, p < .05$ ]. Over anterior [ $F(1, 30) = 8.16, p < .01, \text{Tukey } p < .05$ ], but not posterior sites ( $p > .1$ ), the negativity was larger in women than in men. Between 1,450 and 2,000 ms sex interacted with hemisphere [ $F(1, 30) = 6.57, p < .05$ ]. During this time window, the ERP was more positive in men than in women and this difference was larger over left [ $F(1, 30) = 11.2, p < .01, \text{Tukey } p < .05$ ] compared to right hemisphere recording sites [ $F(1, 30) = 6.5, p < .05, \text{Tukey } p < .05$ ]. Sex differences were also present between 250 and 450 ms [ $F(1, 30) = 8.59, p < .01$ ] and between 750 and 2,000 ms [ $F(1, 30) = 6.11, p < .05$ ] indicating more negative potentials in women compared to men.

Word repetitions were associated with a more positive ERP between 100 and 2000 ms [repetition\*AP:  $F(1, 30) = 29.06, p < .0001$ ]. Moreover, this effect was larger over posterior [ $F(1, 30) = 26.11, p < .0001$ ] compared to anterior sites [ $F(1, 30) = 10.99, p < .01$ ]. No other interactions with the factor Repetition reached significance.



**Figure 6.** ERPs time locked to the onset of words presented in Experiment 2. Negativity is plotted upwards. Between 500 and 1000 ms following word onset, both men and women showed a larger negativity to words with incongruous compared to congruous emotional prosody.



p<.05\*, p<.01\*\*, p<.001\*\*\*, p<.0001\*\*\*\*

Figure 7. Effects that reached significance in the exploratory 50-ms time window statistical analysis for Experiment 2. The far left column lists significant main effects and interaction. The remaining columns illustrate the time range of significant main effects and interactions in seconds. For example, the Sex main effect was significant between 0.25 and 0.45 s. The ANOVA conducted on mean voltages of the 0.25- to 0.45-s time window revealed an  $F$ -value of 8.59 for this main effect.

## Discussion

Experiment 2 tested whether directing listeners' attention to emotional prosody would eliminate sex differences in the integration of verbally and vocally expressed emotions. To this end, participants were required to judge the congruence between word meaning and emotional prosody. Reaction time and accuracy indicated that this task was more difficult than the word valence judgment employed in Experiment 1. Participants responded approximately 200 ms later and made 5% more errors. Moreover, in contrast to Experiment 1, Experiment 2 revealed significant sex differences in reaction time: women responded approximately 50 ms faster than men. Given that women showed generally larger negative amplitudes than men in Experiment 2, women may have engaged more processing resources than men when performing the task, which could explain the female behavioral advantage observed in the present study.

As in Experiment 1, congruence between word meaning and emotional prosody modulated the behavioral and ERP measures. Congruous trials were responded to faster and more accurately than incongruous trials. While for reaction time, this congruence effect was reduced with stimulus repetition, repetition increased the congruence effect for accuracy suggesting a speed-accuracy trade-off presumably associated with a change in the response criterion during the course of the experiment. In the ERP, congruous trials elicited a smaller late negativity than incongruous trials suggesting a reduction in processing effort. Unlike the behavioral measures, the ERP congruence effect was quite comparable for first and second stimulus presentations with the exception that its scalp distribution was more pronounced over right-anterior regions at second compared to first presentations in women. This may indicate that women's processing strategy slightly changed over the course of the experiment in that they relied more strongly on information processing in cortical sources detected at right-anterior sites. As the ERP effects of emotional congruence were comparable in men and women, other than these small topographical differences, we assume that, as in German listeners (Schirmer, Kotz, & Friederici, 2005), attention modulates sex differences in emotional speech processing in Cantonese listeners. If attention is directed at both word meaning and emotional prosody, men and women integrate this information in a similar way. These findings

suggest similarities in the processing of emotional speech between Western and Asian cultures and between intonation and tone languages. Nevertheless, the present results also suggest differences. For example, ERP effects of emotional-prosodic congruence were observed later than in German listeners. Moreover, depending on attentional focus, they were present for different ERP components. Below we discuss the potential determining factors and functional implications of these findings.

## General Discussion

Previous work in German listeners revealed that emotional prosody is integrated with verbal information approximately 400 ms following word onset and that women are more likely than men to show this integration when vocal information is task irrelevant. The present study addressed the question whether the way men and women use emotional prosody for language processing is comparable across languages and cultures. To this end, we investigated when Cantonese male and female listeners integrate vocal and verbal emotional information and whether this integration is modulated by attention. This study was designed to closely match previous work in German listeners. Specifically, we aimed at replicating the study published by Schirmer and Kotz (2003). There are a few methodological differences between the earlier and the present experiments that will be mentioned briefly before the discussion of the results. First, the original German study included emotional and neutral words, whereas the present study used emotional words only. The reason for this is that a follow-up study in German listeners showed that accuracy is better when only emotional words are presented while brain correlates of the integration of vocal and verbal information and respective sex differences persist (Schirmer et al., 2002, 2004; Schirmer, Kotz, & Friederici, 2005). Second, the original German study employed a prosodic and a verbal emotional judgment within the same participants. As only the verbal emotional judgment revealed significant ERP differences between words with congruous and incongruous emotional prosody, only this task was adopted here. Third, the present study differed from Schirmer and Kotz (2003) in that sad instead of angry prosody was used. This difference is attributable to the fact that none of the speakers tested initially was good at conveying anger. In contrast, speaker ratings indicated that sadness

was conveyed convincingly and we therefore decided to use this emotion. This decision is supported by previous work in German listeners showing that sad prosody successfully elicits integration effects reflected by the N400 (Schirmer et al., 2002, 2005). Fourth, owing to differences in mean number of syllables and speaker tempo the average duration of spoken words was slightly shorter in the present (790 ms) study than in the Schirmer and Kotz study (2003; 823 ms). Finally, stimuli in the present study were presented via head phones whereas loud speakers were used by Schirmer and Kotz. Although all other aspects of the present and the previous German study were comparable and it is unlikely that the reported methodological differences can explain the present results, future cross-cultural work should aim to increase the comparability of findings from different cultural groups by cross-validating experimental paradigm and stimuli (Ishii et al., 2003).

### *Language Features Modulate Language Related ERP Components*

The N400 is the most intensively studied language related ERP component. Typically, an N400 is elicited when participants read or hear a word. Depending on whether words are presented visually or auditorily, the N400 has a more central or centro-parietal scalp distribution, respectively (e.g., Gomes, Ritter, Tarter, Vaughan Jr., & Rosen, 1997). However, in either case its peak latency is approximately 400 ms. With respect to the functional underpinnings of the N400, most researchers agree that it reflects the retrieval of word information from long-term memory and the integration of this information into ongoing discourse. Moreover, an increase in N400 amplitude is thought to reflect more effortful retrieval and integration processes (Kutas & Federmeier, 2000; Van Berkum, Zwitserlood, Hagoort, & Brown, 2003). Frequently, an N400 is followed by a late positive component (LPC), which has been found to be larger in amplitude if a word is discordant with semantic or syntactic constraints (Kutas & Hillyard, 1980; Hahne & Friederici, 2002; Münte, Heinze, Matzke, Wieringa, & Johannes, 1998; Van den Brink, Brown, & Hagoort, 2001). This language related LPC resembles the P3b, which is an ERP component elicited by rare categorical deviants. The latency of the P3b is thought to reflect stimulus evaluation time (McCarthy & Donchin, 1981; Picton, 1992), whereas modulations in amplitude are thought to depend on factors such as task relevance or subjective and objective stimulus probability (Johnson, 1986). Similarities between the P3b and the LPC elicited by words have been reported with respect to morphology and the influence of stimulus probability on P3b/LPC amplitude (Coulson, King, & Kutas, 1998). Therefore, some researchers proposed that the LPC elicited by words and the P3b are functionally related (Coulson, King, & Kutas, 1998).

In accordance with previous non-Cantonese language work, the present study revealed a negativity followed by a positivity for the processing of single words in Cantonese. However, both components differ from earlier reports in several respects. The observed negativity had a frontal scalp distribution, peaked 750 ms following word onset, and reached a maximum amplitude of approximately 8 microvolts. Given the centro-parietal scalp distribution of the classical N400 and the N400 elicited in the context of emotional word judgments in German (Schirmer & Kotz, 2003; Figure 1), the frontal topography of this negativity is surprising. However,

anterior scalp distributions have been observed in previous N400 studies (e.g., Van Petten & Riefelder, 1995), which suggests that the sources that mediate this component may vary. Although it is still unclear what causes variations in N400 scalp distribution in Western languages, we speculate that the right anterior distribution observed here may be explained by language characteristics of Cantonese, such as the use of tone. This is in accord with two recent Cantonese sentence processing studies that revealed a right anterior N400-like negativity (Schirmer, Tang, Penney, Gunter, & Chen, 2005; Schön, Schirmer, Penney, Ng, & Besson, submitted). Language characteristics may also explain latency and amplitude differences between the present negativity and the N400 observed in a comparable study conducted in German listeners (Schirmer & Kotz, 2003). In contrast to German, Cantonese is a tone language and comprises a large number of words that differ only in tone. Additionally, many words—although tonally specified—have multiple meanings. Given this high degree of homophony, Cantonese listeners rely heavily on context when interpreting speech (Chen, 1992; Chen, Cheung, Tang, & Wong, 2000). Thus, if context is not available, word processing may be compromised by the large number of competing meanings and, therefore, relatively effortful. In the present study, we tried to reduce processing effort by presenting two-syllable rather than one-syllable words. Nonetheless, ambiguity present at the first syllable may have interfered with and slowed down word processing thereby increasing the latency and amplitude of language related ERP components.

A further factor that might have modulated the latency and amplitude of language related ERP components in the present study is the participants' task. Perhaps the present emotional judgment elicited processes in addition to the ones reflected by the N400 and these processes overlapped in the observed negativity. Findings from Schirmer and Kotz (2003) suggest that this is a possibility. In their study, the same stimuli elicited an early (300 ms) or a late (500 ms) N400 depending on whether participants judged prosodic or word valence, respectively. Thus, perhaps the more difficult word valence judgment recruited processing resources that were not required during the easier prosodic judgment. Moreover, these processes may have overlapped in producing the negativity referred to as the N400. A similar assumption can be formulated for the present results. In order to judge word valence and emotional-prosodic congruence, participants may have recruited memory retrieval and evaluation processes in addition to the ones involved in simple prosodic decisions. These processes may have increased the latency and amplitude of language related ERP components.

### *Attention Modulates the Integration of Emotional Prosody and Word Meaning*

Comparison of the results of Experiments 1 and 2 indicates that attention to emotional prosody determines when in time (female) Cantonese listeners integrate emotional prosody and word meaning. In Experiment 1, where attention was directed at word meaning only, an LPC peaking around 1,500 ms following word onset differentiated between words with congruous and incongruous emotional prosody. The LPC to words with incongruous emotional prosody was larger than the LPC to words with congruous emotional prosody. A similar effect has been reported for the processing of purely prosodic violations in nontone languages (e.g., sen-

tence prosody changing from a question to a statement; Astesano, Besson, & Alter, 2004). Both the latter and the present LPC effect may reflect subjective stimulus probability such that listeners perceived incongruous prosody as relatively rare and unusual. Moreover, given that in the present study words had an average duration of 790 ms, the latency of the observed LPC effect suggests that listeners integrated emotional prosody after a word had been retrieved from memory. Thus, the retrieval of word meaning from memory was unaffected by emotional prosody. In Experiment 2, emotional prosody modulated the amplitude of an N400-like negativity. Between 500 and 1,000 ms following word onset, words with incongruous emotional prosody elicited a significantly larger negativity than words with congruous emotional prosody. Although having a shorter latency than the congruence effect in Experiment 1, this congruence effect nevertheless peaked after words had been fully presented and may thus reflect processes that follow word identification. This contrasts with the results from German listeners who integrate emotional prosody and word meaning as early as 300 ms following word onset, when only little word information is available (Schirmer & Kotz, 2003).

Together, both experiments indicate that emotional prosody plays a subordinate role during word processing in Cantonese. One might argue that this is attributable to the fact that speaker pitch in Cantonese primarily codes lexical information such that its role in conveying emotional information is restricted relative to nontone languages such as German. As a consequence, Cantonese listeners may have fewer prosodic cues available for emotion recognition (Vance, 1976). However, there are other cues besides pitch that Cantonese listeners can use. Moreover, physiological parameters that change with speaker arousal (e.g., blood pressure, muscle tension) modulate spoken utterances via their impact on the vocal production system (Scherer, 1989). Thus, as in other languages, vocal quality, intensity, and speaker tempo, among other cues, are informative of speaker emotion in Cantonese. Moreover, researchers have argued that the availability of these cues should be comparable across languages (Scherer, Banse, & Wallbott, 2001). Based on this and the present finding that Cantonese listeners identify the congruency between vocal and verbal emotional information with 90 percent accuracy we conclude that vocal emotional information in Cantonese is sufficient for emotion recognition. Furthermore, we believe that the delayed ERP integration effects observed in the present study cannot be attributed to difficulties in processing vocal emotions. Rather, they may reflect a delayed use of vocal emotional information for language processing. Moreover, because of the significance of linguistic prosody or tone in Cantonese, initially, word processing resources may be bound at a lexical level so that the integration of emotional prosody is delayed relative to the use of tone.

### *Sex Differences in Emotional Speech Processing*

Recent research in German listeners indicates that men and women differ in the integration of emotional prosody and word meaning when emotional prosody is unattended (Schirmer et al., 2002; Schirmer & Kotz, 2003; Schirmer, Kotz, & Friederici, 2005). The present study was conducted to determine whether sex

differences in emotional speech processing are specific to Western culture. To this end, we investigated emotional speech processing in Cantonese, an Asian tone language. Our results indicate that as for Germans, Cantonese men and women differ in the integration of emotional prosody and word meaning. When emotional prosody was unattended, ERP effects of emotional incongruity were more pronounced in women than in men. This sex difference was eliminated under the instruction to attend to both emotional prosody and word meaning. These findings suggest that, compared to male Cantonese listeners, female Cantonese listeners direct more processing resources towards the integration of vocal and verbal emotional information when vocal information is task-irrelevant. As a consequence, female listeners might be better able to recognize banter or sarcasm—where vocal and verbal emotional information is in conflict. Given the similarities between these findings and earlier findings from German listeners, it seems that the observed sex differences may be relatively independent of cultural and language background.

This tentative conclusion coincides with a number of behavioral reports that suggest a higher sensitivity to nonverbal signals in women than in men (for a review see Hall, 1978) as well as more recent electrophysiological evidence for women being more likely than men to discriminate between emotional and neutral tone of voice when vocal utterances are unattended (Schirmer, Striano, & Friederici, 2005). One reason women might be more sensitive to tone of voice is because it helps them maintain social relationships. This interpretation agrees with the finding that, compared to men, women are more socially oriented and more likely to define themselves in relational terms (Clancy & Dollinger, 1993; Dollinger, Preson, O'Brien, & DiLalla, 1996; Kashima et al., 2004). That this is true across different cultures is suggested by the work of Kashima and colleagues who found that women from Australia, the United States, Japan, and Korea describe themselves as more emotionally related toward others than do men (Kashima et al., 1995). Even though this cross-cultural evidence does not contradict social accounts of sex differences such as the division of labor and the social institution of task-oriented male and female behavior (Kashima et al., 1995), it may be viewed in the light of evidence that sex differences in social orientation are biologically modulated, for example by hormones such as oxytocin and estrogen (Taylor et al., 2000).

### *Conclusions*

Although vocal emotional expression and recognition are largely comparable across languages and cultures (Scherer, Banse, & Wallbott, 2001), differences exist as to when and how listeners use vocal emotional information for speech processing. These differences may be determined by language characteristics such as the linguistic functions of prosody (e.g., lexical tone) or the degree of homophony. Furthermore, there appear to be interindividual differences in the integration of vocal and verbal emotional information that show across different languages and cultures. Specifically, women seem more sensitive to incongruities between vocal and verbal emotional information. This may enhance their perception of banter and sarcasm in social interactions and likely reflects the significance of social interactions for women and men, respectively.

## References

- Astesano, C., Besson, M., & Alter, K. (2004). Brain potentials during semantic and prosodic processing in French. *Cognitive Brain Research, 18*, 172–184.
- Atchley, R. A., Ilardi, S. S., & Enloe, A. (2003). Hemispheric asymmetry in the processing of emotional content in word meanings: The effect of current and past depression. *Brain and Language, 84*, 105–119.
- Blonder, L. X., Bowers, D., & Heilman, K. M. (1991). The role of the right hemisphere in emotional communication. *Brain, 114*, 1115–1127.
- Bostanov, V., & Kotchoubey, B. (2004). Recognition of affective prosody: Continuous wavelet measures of event-related brain potentials to emotional exclamations. *Psychophysiology, 41*, 259–268.
- Broca, P. A. (1865). Sur la faculté du langage articulé. *Bulletin de la Société Anatomie de Paris, 6*, 493–494.
- Brownell, H., Michel, D., Powelson, J., & Gardner, H. (1983). Surprise but not coherence: Sensitivity to verbal humor in right-hemisphere patients. *Brain and Language, 18*, 20–27.
- Brownell, H., Michel, D., Powelson, J., & Gardner, H. (1986). Inference deficits in right brain-damaged patients. *Brain and Language, 27*, 310–321.
- Buchanan, T. W., Lutz, K., Mirzazade, S., Specht, K., Shah, N. J., Zilles, K., et al. (2000). Recognition of emotional prosody and verbal components of spoken language: An fMRI study. *Cognitive Brain Research, 9*, 227–238.
- 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.
- Chen, H.-C., Cheung, H., Tang, S. L., & Wong, Y. T. (2000). Effects of antecedent order and semantic context on Chinese pronoun resolution. *Memory and Cognition, 28*, 427–438.
- Clancy, S. M., & Dollinger, S. J. (1993). Photographic depictions of the self: Gender and age differences in social connectedness. *Sex Roles, 29*, 477–495.
- Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes, 13*, 21–58.
- De Renzi, E., Liotti, M., & Nichelli, P. (1987). Semantic amnesia with preservation of autobiographic memory. A case report. *Cortex, 23*, 575–597.
- Dollinger, S. J., Preston, L. A., O'Brien, S. P., & DiLalla, D. L. (1996). Individuality and relatedness of the self: An autophotographic study. *Journal of Personality and Social Psychology, 71*, 1268–1278.
- Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the bootstrap*. New York: Chapman & Hall.
- Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research, 1*, 183–192.
- Gandour, J., Wong, D., Dziedzic, M., Lowe, M., Tong, Y., & Li, X. (2003). A cross-linguistic fMRI study of perception of intonation and emotion in Chinese. *Human Brain Mapping, 18*, 149–157.
- George, M. S., Parekh, P. I., Rosinsky, N., Ketter, T. A., Kimbrell, T. A., Heilman, K. M., et al. (1996). Understanding emotional prosody activates right hemisphere regions. *Archives of Neurology, 53*, 665–670.
- Gomes, H., Ritter, W., Tarter, V. C., Vaughan Jr., H. G., & Rosen, J. J. (1997). Lexical processing of visually and auditorily presented nouns and verbs: Evidence from reaction time and N400 priming data. *Cognitive Brain Research, 6*, 121–134.
- Hagoort, P., & Brown, C. (2000). ERP effects of listening to speech: Semantic ERP effects. *Neuropsychologia, 38*, 1518–1530.
- Hahne, A., & Friederici, A. D. (2002). Differential task effects on semantic and syntactic processes as revealed by ERPs. *Cognitive Brain Research, 13*, 339–356.
- Hall, E. T. (1976). *Beyond culture*. New York: Doubleday.
- Hall, J. A. (1978). Gender effects in decoding nonverbal cues. *Psychological Bulletin, 4*, 845–857.
- Holcomb, P. J., & Anderson, J. E. (1993). Cross-modal semantic priming: A time-course analysis using event-related brain potentials. *Language and Cognitive Processes, 8*, 379–411.
- Ishii, K., Reyes, J. A., & S. Kitayama (2003). Spontaneous attention to word content versus emotional tone. *Psychological Science, 14*, 39–46.
- Johnson, R. (1986). A triarchic model of P300 amplitude. *Psychophysiology, 23*, 367–384.
- Kang, A. M., Constable, R. T., Gore, J. C., Avrutin, S. (1999). An event-related fMRI study of implicit phrase-level syntactic and semantic processing. *NeuroImage, 10*, 55–61.
- Kashima, Y., Kokubo, T., Kashima, E. S., Boxall, D., Yamaguchi, S., & Macrae, K. (2004). Culture and self: Are there within-culture differences in self between metropolitan areas and regional cities? *Personality and Social Psychology Bulletin, 30*, 816–823.
- Kashima, Y., Yamaguchi, S., Kim, U., Choi, S. C., Gelfand, M. J., & Yuki, M. (1995). Culture, gender, and self: A perspective from individualism-collectivism research. *Journal of Personality and Social Psychology, 69*, 925–937.
- Kiehl, K. A., Laurens, K. R., & Liddle, P. F. (2002). Reading anomalous sentences: An event-related fMRI study of semantic processing. *NeuroImage, 17*, 842–850.
- Kitayama, S., & Ishii, K. (2002). Word and voice: Spontaneous attention to emotional utterances in two languages. *Cognition and Emotion, 16*, 29–59.
- Kotz, S. A., Meyer, M., Alter, K., Besson, M., von Cramon, D. Y., & Friederici, A. D. (2003). On the lateralization of emotional prosody: An event-related functional MR investigation. *Brain and Language, 86*, 366–376.
- Kuperberg, G. R., McGuire, P. K., Bullmore, E. T., Brammer, M. J., Rabe-Hesketh, S., Wright, I. C., et al. (2000). Common and distinct neural substrates for pragmatic, semantic, and syntactic processing of spoken sentences: An fMRI study. *Journal of Cognitive Neuroscience, 12*, 321–341.
- Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*, 463–470.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*, 203–205.
- Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*, 161–163.
- Kutas, M., Van Petten, C., & Besson, M. (1988). Event-related potential asymmetries during the reading of sentences. *Electroencephalography and Clinical Neurophysiology, 69*, 218–233.
- Münte, T. F., Heinze, H.-J., Matzke, M., Wieringa, B. M., & Johannes, S. (1998). Brain potentials and syntactic violations revisited: No evidence for the specificity of the syntactic positive shift. *Neuropsychologia, 36*, 217–226.
- Matsumoto, D. (1990). Cultural similarities and differences in display rules. *Motivation and Emotion, 14*, 195–214.
- Matsumoto, D., Kasri, F., & Kookan, K. (1999). American-Japanese cultural differences in judgements of expression intensity and subjective experience. *Cognition and Emotion, 13*, 201–218.
- McCarthy, G., & Donchin, E. (1981). A metric for thought: A comparison of P300 latency and reaction time. *Science, 211*, 77–80.
- Newman, A. J., Pancheva, R., Ozawa, K., Neville, H. J., & Ullman, M. T. (2001). An event-related fMRI study of syntactic and semantic violations. *Journal of Psycholinguistic Research, 30*, 339–364.
- Ni, W., Constable, R. T., Mencl, W. E., Pugh, K. R., Fulbright, R. K., Shaywitz, S. E., et al. (2000). An event-related neuroimaging study distinguishing form and content in sentence processing. *Journal of Cognitive Neuroscience, 12*, 120–133.
- Pell, M. D., & Leonard, C. L. (2003). Processing emotional tone from

- speech in Parkinson's disease: A role for the basal ganglia. *Cognitive and Affective Behavioral Neuroscience*, 3, 275–288.
- Perani, D., Cappa, S. F., Schnur, T., Tettamanti, M., Collina, S., Rosa, M. M., et al. (1999). The neural correlates of verb and noun processing: A PET study. *Brain*, 122, 2337–2344.
- Picton, T. W. (1992). The P300 wave of the human event-related potential. *Journal of Clinical Neurophysiology*, 9, 456–479.
- Schön, D., Schirmer, A., Ng, E., Penney, T., & Besson, M. (submitted). Effects of phonemic and tone changes on semantic processing in Cantonese: A behavioral and ERP study.
- Scherer, K. R. (1989). Vocal correlates of emotional arousal and affective disturbance. In H. Wagner & A. Manstead (eds.), *Handbook of social psychophysiology* (pp. 165–197). Oxford, UK: Wiley & Sons.
- Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion interferences from vocal expressions correlate across languages and cultures. *Journal of Cross-Cultural Psychology*, 32, 76–92.
- Schirmer, A., Alter, K., Kotz, S. A., & Friederici, A. D. (2001). Lateralization of prosody during language production: A lesion study. *Brain and Language*, 76, 1–17.
- Schirmer, A., & Kotz, S. A. (2003). ERP evidence for a gender specific Stroop effect in emotional speech. *Journal of Cognitive Neuroscience*, 15, 1135–1148.
- Schirmer, A. & Kotz, S. A. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, 10, 24–30.
- Schirmer, A., Kotz, S. A., & Friederici, A. D. (2002). Sex differentiates the role of emotional prosody during word processing. *Cognitive Brain Research*, 14, 228–233.
- Schirmer, A., Kotz, S. A., & Friederici, A. D. (2005). On the role of attention for the processing of emotions in speech: Sex differences revisited. *Cognitive Brain Research*, 24, 442–452.
- Schirmer, A., Striano, T., & Friederici, A. D. (2005). Sex differences in the pre-attentive processing of vocal emotional expressions. *Neuroreport*, 16, 635–639.
- Schirmer, A., Tang, S. L., Penney, T. B., Gunter, C. T., & Chen, H. C. (2005). Brain responses to segmentally and tonally induced semantic violations in Cantonese. *Journal of Cognitive Neuroscience*, 17, 1–12.
- Schirmer, A., Zysset, S., Kotz, S. A., & von Cramon, D. Y. (2004). Gender differences in the activation of inferior frontal cortex during emotional speech perception. *NeuroImage*, 21, 1114–1123.
- Taubner, R. W., Rayner, A. M., & Heilman, K. M. (1999). Frontal-ocular aphasia. *Brain and Language*, 70, 240–261.
- Taylor, S. E., Cousino Klein, L., Lewis, B. P., Gruenewald, T. L., Gurung, R. A. R., & Updegraff, J. A. (2000). Biobehavioral responses to stress in females: Tend-and-befriend, not fight-or-flight. *Psychological Review*, 107, 411–429.
- Thompson-Schill, S. L., D'Esposito, M., & Kan, I. P. (1999). Effects of repetition and competition on activity in left prefrontal cortex during verb generation. *Neuron*, 23, 513–522.
- Van Berkum, J. J. A., Zwitserlood, P., Hagoort, P., & Brown, C. M. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research*, 17, 701–718.
- Vance, T. (1976). An experimental investigation of tone and intonation in Cantonese. *Phonetica*, 33, 368–392.
- Van den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience*, 13, 967–985.
- Van Lancker, D. (1980). Cerebral lateralization of pitch cues in the linguistic signal. *International Journal of Human Communication*, 13, 227–277.
- Van Lancker, D., & Fromkin, V. A. (1973). Hemispheric specialization for pitch and 'tone': Evidence from Thai. *Journal of Phonetics*, 1, 101–109.
- Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 394–417.
- Van Petten, C., & Rheinfelder, H. (1995). Conceptual relationships between spoken words and environmental sounds: Event-related brain potential measures. *Neuropsychologia*, 33, 485–508.

Received February 17, 2005

Revision received February 10, 2006

Accepted March 7, 2006 ■