ERP Evidence for a Sex-Specific Stroop Effect in Emotional Speech

Annett Schirmer and Sonja A. Kotz

Abstract

■ The present study investigated the interaction of emotional prosody and word valence during emotional comprehension in men and women. In a prosody–word interference task, participants listened to positive, neutral, and negative words that were spoken with a happy, neutral, and angry prosody. Participants were asked to rate word valence while ignoring emotional prosody, or vice versa. Congruent stimuli were responded faster and more accurately as compared to incongruent emotional stimuli. This behavioral effect was more salient for the word valence task than for the prosodic task and was comparable between men and women. The event-related potentials (ERPs) revealed a smaller N400 amplitude for congruent as compared to emotionally incongruent stimuli. This ERP effect, however, was significant only for the word valence judgment and only for female listeners. The present data suggest that the word valence judgment was more difficult and more easily influenced by task-irrelevant emotional information than the prosodic task in both men and women. Furthermore, although emotional prosody and word valence may have a similar influence on an emotional judgment in both sexes, ERPs indicate sex differences in the underlying processing. Women, but not men, show an interaction between prosody and word valence during a semantic processing stage. ■

INTRODUCTION

In spoken language, emotions can be expressed on two separate levels. One level refers to the words in an utterance that convey emotional meaning, such as "sadness" or "happiness." Additionally, speech melody serves to communicate emotions. This second level, also referred to as emotional prosody, subsumes different acoustic parameters such as time structure, loudness, and fundamental frequency. These acoustic parameters of speech differ with respect to the emotion that is expressed by the speaker (Scherer, Banse, & Wallbott, 2001; Banse & Scherer, 1996). A happy utterance, for example, is usually spoken faster, louder, and with a higher fundamental frequency as compared to an utterance spoken in a sad emotional state.

How listeners derive emotions from words and prosody has been subject to a number of studies. With respect to word information, affective priming studies were used to explore the influence of emotionally valenced words on the processing of subsequent emotional words. When those words were emotionally congruent, evaluative judgments were faster and more accurate as compared to emotionally incongruent words (Fazio, Sanbonmatsu, Powel, & Kardes, 1986). Furthermore, neurophysiological processing characteristics of emotional words have been studied using event-related potentials (ERPs). For example, emotional as compared to neutral words elicit a larger positivity between 300 and 700 msec following stimulus onset regardless of whether participants count letters or evaluate the emotional content of a word (Naumann, Bartussek, Diedrich, & Laufer, 1992). Similar findings using emotional and neutral pictures have been interpreted as reflecting the affective significance of a particular stimulus (Kayser et al., 2000).

Studies investigating the processing of emotional prosody mainly focused on the underlying neuroanatomical correlates. Results from brain-damaged patients and normal participants suggest a right hemisphere dominance for emotional-prosodic processes at a cortical level (Starkstein, Federoff, Price, Leiguarda, & Robinson, 1994; Blonder, Bowers, & Heilman, 1991; Ley & Bryden, 1982), as well as a contribution of subcortical structures such as the basal ganglia (Kotz et al., in press; Pell, 1996).

Grimshaw (1998) conducted one of the few investigations of the interaction between emotions at the prosodic and at the word level as conveyed in speech. She presented the words "mad," "sad," and "glad" in three different emotional tones that either matched or mismatched the emotional valence of those words. In a stimulus identification task, subjects either listened to the emotional prosody or to the word content and responded by pressing one of three buttons. When listening to the prosody, incongruous trials were responded

Max Planck Institute of Cognitive Neuroscience

to more slowly and less accurately than congruous trials. No such interference was present when subjects concentrated on the word content. In accordance with assumptions about automaticity derived from other Stroop-like interference tasks (MacLeod & MacDonald, 2000; Cohen, Dunbar, & McClelland, 1990), this suggests that word content is processed more automatically than is emotional prosody. Kitayama and Ishii (2002) reached a similar conclusion. They asked a group of American listeners to indicate whether a word's prosody was positive or negative or whether a word's connotative meaning was positive or negative. As has been reported previously, participants were faster in their prosodic judgments when prosody and word meaning were congruent as compared to incongruent, whereas there was no such effect in the word valence judgment.

The present study further investigated the interaction of emotional prosody and word valence in a Stroop-like situation. Moreover, this study should determine whether the listener's sex would modulate this interaction. Previous work on emotional perception suggests that women are more attuned to nonverbal cues such as prosody and facial expression as compared to men (Rotter & Rotter, 1988; Hall, 1978). Hall (1978), for example, conducted a meta-analysis of over 70 studies on emotional comprehension and found a small but significant female advantage in the recognition of nonverbal emotional cues. Further support for sex differences in the role of emotional prosody has been reported recently by Schirmer, Kotz, and Friederici (2002). Cross-modal priming effects from prosody on word processing occurred with a short interstimulus interval (200 msec) in female participants and with a long interstimulus interval (750 msec) in male participants, suggesting that women integrate words into the emotional-prosodic context earlier than men do. That emotional prosody is recognized more accurately and influences speech processing earlier in women than in men might suggest sex differences in the reliance on prosodic information when judging the emotional content of speech. To test this latter assumption, the present study employed an emotional Stroop paradigm similar to Grimshaw (1998) and Kitayama and Ishii (2002). Participants listened to words that had a positive, neutral, or negative valence and were spoken with a happy, neutral, or angry prosody. Consequently, there were trials in which prosody and word information were congruent, and there were trials in which this information was incongruent. In one experimental block, participants made a word valence judgment, while ignoring emotional prosody. As the word valence has to be derived from the semantic meaning of a word, this task is referred to as the semantic task. In another experimental block, participants had to judge emotional

Table 1. RT and Accuracy ANOVA: Significant Effects

RT			*Accuracy*		
Source	*df*	*F Value*	*Source*	*df*	*F Value*
Sex	1,60	11.21**	Task	1,60	605.05****
Voice	1,60	9.06**	Task × Voice	1,60	13.07***
Task	1,60	375.92****	Prosody	2,120	14.43****
Task × Voice	1,60	3.35#	Prosody × Voice	2,120	13.07**
Prosody	2,120	94.25****	Word	2,120	9.78****
Prosody × Voice	2,120	56.78****	Task × Prosody	2,120	10.57****
Word	2,120	82.1****	Task × Prosody × Voice	2,120	6.3**
Task × Prosody	2,120	53.1****	Task × Word	2,120	13.11****
Task × Prosody × Voice	2,120	14.57****	Task × Word × Sex	2,120	2.53#
Task × Word	2,120	54.41****	Prosody × Word	4,240	21.82****
Prosody × Word	4,240	26.51****	Prosody × Word × Voice	4,240	4.86***
Prosody × Word × Voice	4,240	8.56****	Task × Prosody × Word	4,240	10.35****
Task × Prosody × Word	4,240	2.35*			

*$p < .05$.

**$p < .01$.

***$p < .001$.

****$p < .0001$.

#$p < .1$.

Table 2. RT *F* Values: Post Hoc Comparisons for the Prosody × Word Interaction as Modified by Task and Voice

		Prosodic Task			Semantic Task		
	df	*Positive Prosody*	*Neutral Prosody*	*Negative Prosody*	*Positive Prosody*	*Neutral Prosody*	*Negative Prosody*
Word	2,122	23.81****	<1	<1	77.38****	47.92****	66.81****
Neutral versus negative words	1,61	3.18#			75.04****	88.4****	101.58****
Positive versus negative words	1,61	23.78****			7.31**	3.05#	45.34****
Positive versus neutral words	1,61	47.37****			123.6****	44.14****	38.23****

		Female Speaker			Male Speaker		
	df	*Positive Prosody*	*Neutral Prosody*	*Negative Prosody*	*Positive Prosody*	*Neutral Prosody*	*Negative Prosody*
Word	2,60	21.48****	28.21****	31.35****	76.04****	9.73***	34.29****
Neutral versus negative words	1,30	21.99****	36.99****	65.59****	56.45****	22.76****	46.96****
Positive versus negative words	1,30	3.42#	<1	30.4****	29.0****	<1	9.24**
Positive versus neutral words	1,30	38.29****	39.89****	8.25**	113.9****	10.5**	34.16****

*p < .05.

**p < .01.

***p < .001.

****p < .0001.

#p < .1.

prosody as positive, neutral, or negative, while ignoring word content. Participants were asked to respond by pressing one of three buttons on a response box as quickly and as accurately as possible. In addition to the behavioral responses, we recorded the ERPs.

In accordance with previous behavioral findings (Kitayama & Ishii, 2002; Grimshaw, 1998), we expected an emotional Stroop effect to be reflected in longer reaction times (RTs) and lower accuracy rates for incongruous as compared to congruous trials. Furthermore, if word content is indeed processed more automatically than emotional prosody, there should be no or only a small difference between congruous and incongruous trials in the semantic task, whereas the prosodic task should show a clear emotional Stroop effect. Additionally, the strength of this emotional Stroop effect in both tasks might vary as a function of sex. Accordingly, if emotional prosody is more important and harder to ignore for women than for men, women should show a smaller emotional Stroop effect than men in the prosodic judgment. Furthermore, if the semantic task elicits differences between congruous and incongruous trials they should be stronger for women than for men. We hoped to obtain additional information about the interaction between emotional prosody and word valence from the ERPs. Because of their high temporal resolution, ERPs can reveal when during processing such an interaction occurs. Furthermore, there is evidence from the ERP literature on Stroop-like interference tasks that a negativity with a maximum around 400 msec post stimulus onset is sensitive to incongruous, task-irrelevant information (Greenham, Stelmack, Campbell, 2000; Liotti, Woldorff, Perez, & Mayberg, 2000; Rebai, Bernard, & Lannou, 1997). This negativity is larger on incongruous as compared to congruous trials. Based on these findings, we expected behavioral indices of interference in the present study to be accompanied by a similar ERP effect. Moreover, as the semantic processing of a word is reflected by a negativity, called the N400 (Kutas & Hillyard, 1980), which is elicited at a similar point intime as the Stroop-related negativity, we assumed that an influence of task-irrelevant information on this ERP component would reflect interference during a stimulus processing stage (e.g., Rebai et al., 1997).

In sum, we expected behavioral measures to indicate whether there are sex differences in the relative significance of prosody as compared to word valence for the comprehension of emotions in speech and that ERPs would indicate whether the influence that one type of information exerts on processing the other occurs during a semantic processing stage. A last question addressed in this study relates to the speaker's voice. Our previous finding of sex differences (Schirmer et al., 2002) was obtained using a stimulus set produced by only one female speaker. In the present study, we employed a male and a female speaker in order to be able to generalize any sex differences across different speakers and to demonstrate that they are independent of the speaker's sex.

RESULTS

Behavioral Data

RTs and accuracy data were entered into separate ANOVAs treating task (semantic/prosodic), word (positive/neutral/negative), and prosody (happy/neutral/angry) as repeated-measures factors and sex and voice (male/female) as between-subjects factors. Table 1 summarizes all marginally significant and significant effects for both ANOVAs. Interactions and main effects not listed in this table failed to reach a p value smaller than .1. As the present study aimed at specifying the interaction between prosody and semantics, only effects that reflect this interaction are discussed in the following text.

Statistical analysis of the RTs revealed a significant Task × Prosody × Word interaction (Table 1), suggesting that the Prosody × Word interaction was stronger for the semantic, $F(4,244) = 19.48$, $p < .0001$, than for the prosodic task, $F(4,244) = 9.41$, $p < .0001$ (Table 2). The analytical approach for examining the Prosody × Word interaction for each task was influenced by differences in word length across the prosodic conditions. Words with neutral prosody (719 msec) were spoken, on the average, more than 100 milliseconds faster than words with happy (863 msec) and angry prosody (886 msec). This difference in word length modulated response latencies especially for the semantic task (Figure 1). Therefore, we avoided a comparison between the prosodic conditions and analyzed the word effect for each level of prosody. During the semantic judgment, response latencies for happily and angrily spoken words were significantly shorter when word meaning was congruent as compared to incongruent. Neutrally spoken words elicited shorter response latencies when their meaning was emotional rather than neutral. During the prosodic judgment, happily spoken words were responded to faster when word valence was congruent as compared to incongruent. No significant effects occurred for words with neutral or angry prosody.

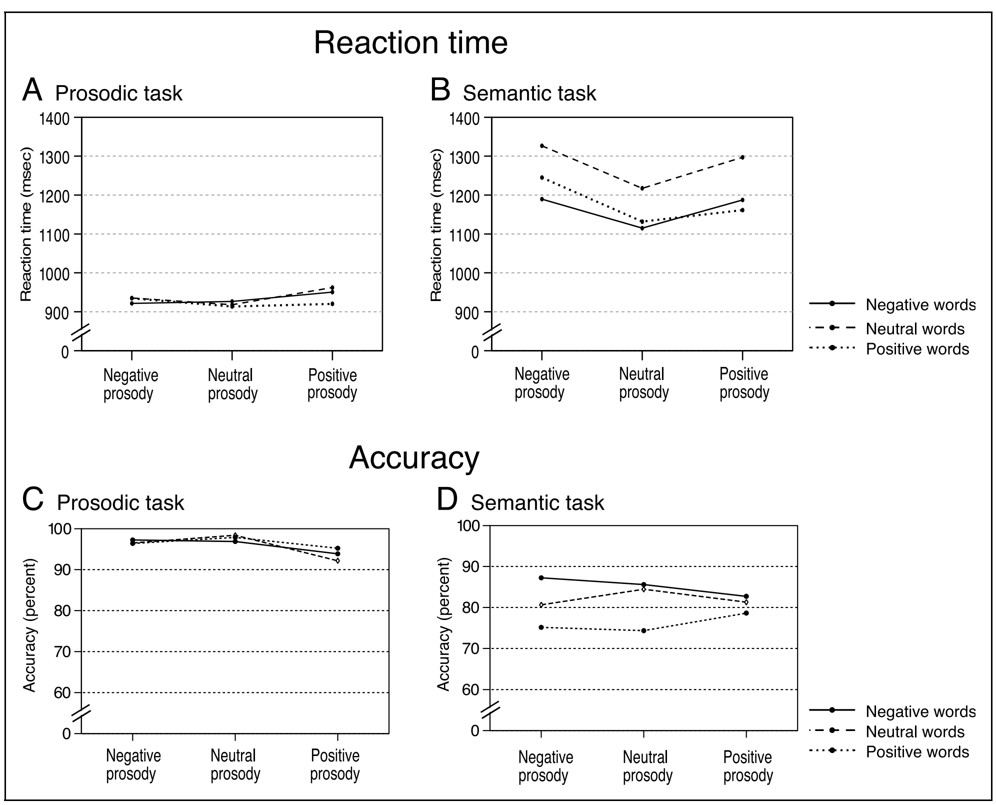
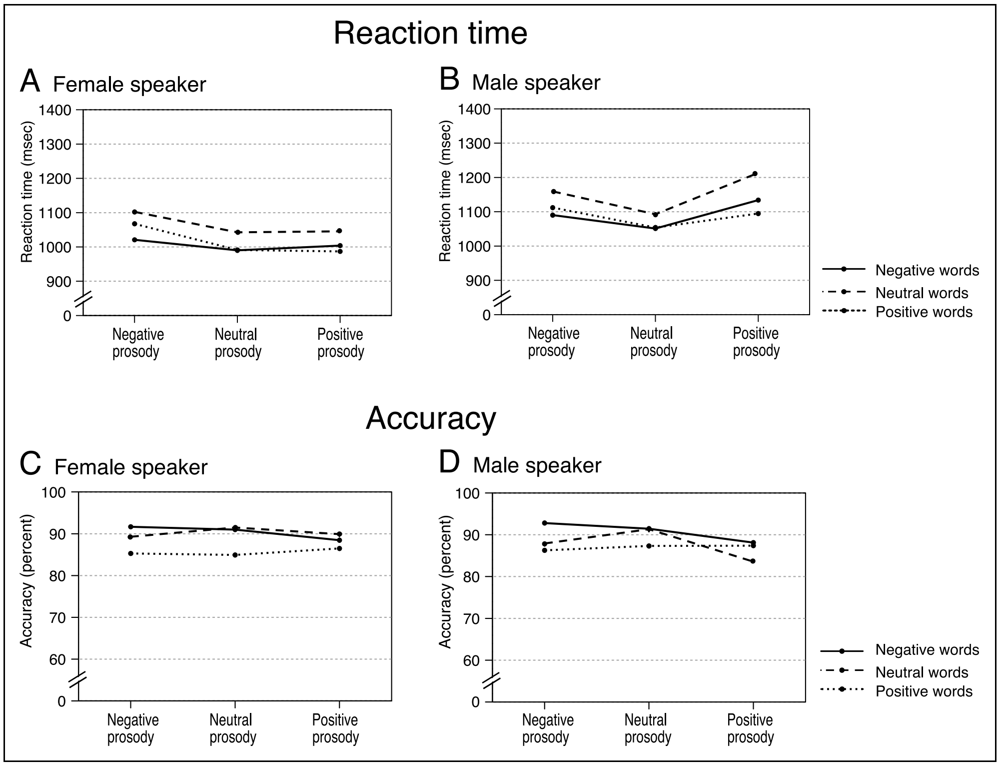


Figure 1. Reaction times for the prosodic task (A) and the semantic task (B) averaged across all participants. Percent correct for the prosodic task (C) and the semantic task (D) averaged across all participants.

Figure 2. Reaction times for all participants that listened to the female (A) and the male speaker (B) averaged across tasks. Percent correct for all participants that listened to the female (C) and the male speaker (D) averaged across tasks.



Additionally, the Prosody × Word interaction was modified by voice (Table 2; Figure 2). Post hoc comparisons suggested similar but stronger effects for the male as compared to the female voice.

The interaction between prosody and semantics was not affected by the sex of the listener. Across the different conditions, women responded faster than men (Figure 3).

Before examining the Stroop effect in the accuracy data, two other effects require consideration. First, accuracy was much higher for the prosodic as compared to the semantic task. For the prosodic task, emotional categories were relatively clear cut and marked by specific acoustic stimulus properties. Therefore, the emotional judgment was more consistent across subjects as compared to the semantic task, where emotional associations to a word had to be retrieved from long-term memory and may have varied due to personal experience. This difference in task difficulty between the prosodic and the semantic judgment may have contributed to task differences such that the prosody–semantics interaction must be interpreted with caution.

Second, although an independent stimulus evaluation with a different group of subjects indicated that valence strength was comparable for negative and positive words (see Methods), in the present study, participants recognized positive words less accurately than negative words. This has to be taken into account when examining the Prosody × Word interaction for the semantic and the prosodic task. Moreover, to avoid a comparison between the three semantic conditions, the analysis of this interaction was approached by examining the prosody effect for each level of word. During the semantic task, congruence between prosody and word meaning elicited higher accuracy than incongruence (Figure 1). In contrast, post hoc comparisons for the prosodic task revealed only weak or no interference effects (Figure 1; Table 3). Independent of emotional meaning, words with happy prosody were rated less accurately than words with neutral or angry prosody. Words with angry prosody were rated less accurately than words with neutral prosody, except when words carried a negative meaning in which case neutrally and angrily spoken words did not differ.

As for the RTs, the Prosody × Word interaction in the accuracy data was modified by voice. Again, the direction of significant effects was comparable for both speakers. However, male prosody had a stronger influence on neutral and negative words, whereas female prosody had a stronger influence on positive words (Table 3; Figure 2). In contrast to the RTs, accuracy measures failed to reveal a significant effect of sex.

ERP Data

As reported in the previous section, accuracy differed between the prosodic and the semantic task. Consequently, there were more trials available for the ERP analysis of the prosodic as compared to the semantic task (85.6% vs. 72.7% following artifact rejection). Moreover, visual inspection revealed different ERP waveforms for the two tasks. Therefore, statistical analyses were conducted separately for each task, and comparisons between the two tasks have to remain tentative.

To analyze the onset and offset of the prosody–semantics interaction, we conducted a series of exploratory ANOVAs for the mean amplitudes of 50-msec time windows starting from 0 to 900 msec post stimulus onset. Mean amplitudes of significant consecutive time windows were then averaged across the significant time interval and entered into a second ANOVA. ANOVAs treated prosody (happy/neutral/angry), word (positive/neutral/negative), hemisphere (left/right), and channel (electrode position 1–18) as repeated-measures factors and sex and voice (male/female) as between-subjects factors. As was done for the RTs, we narrowed the focus of this analysis to the interaction between prosodic and semantic information, but a summary of all significant and marginally significant effects is presented in Table 4. Main effects and interactions that are not listed in this table did not reach a p value smaller than .1.

While there was no significant Prosody × Word interaction for the prosodic task (Figure 4), the semantic task elicited such an interaction for three different time windows. First, the Prosody × Word × Sex interaction was significant between 350 and 650 msec, $F(4,240) = 4.4$, $p < .01$. Analysis of the Prosody × Word interaction with respect to sex revealed a significant effect in women, $F(4,120) = 7.52$, $p < .0001$, but not in men, $F < 1$. The Prosody × Word interaction in women was examined further by analyzing the word effect for each level of prosody. The word effect reached significance for happy and angry prosody, but not for neutral prosody. Happily and angrily spoken words elicited a larger negativity when word valence was incongruent as compared to congruent (see Figures 5 and 6, and Table 5).

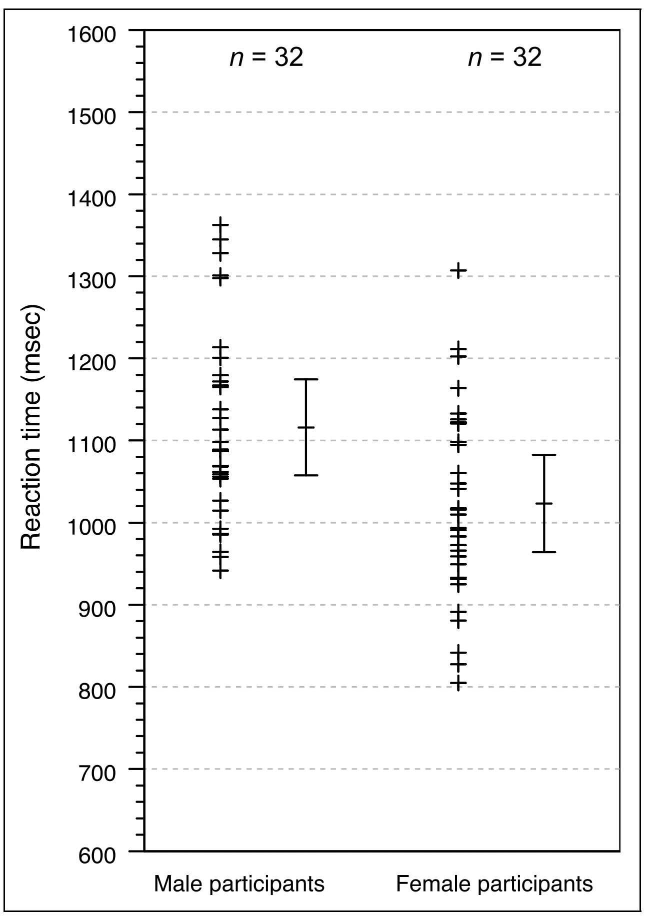


Figure 3. Reaction times for male and female participants averaged across tasks and conditions. Each cross represents the reaction time of one participant. Women responded significantly faster than men.

Second, there was a Prosody × Word × Voice interaction, $F(4,240) = 3.16$, $p < .05$, between 600 and 750 msec. Post hoc comparisons revealed a significant Prosody × Word interaction for the male, $F(4,120) = 8.1$, $p < .0001$, and the female speaker, $F(4,120) = 3.34$, $p < .05$. Furthermore, both speakers elicited a significant word main effect for happy, angry, and neutral prosody. For the female speaker, the word effect for happy and angry prosody was comparable in that incongruence elicited a more negative ERP than congruence (Table 5). For the male speaker, such an effect was present in happily, but not in angrily spoken words. Words with neutral prosody elicited a larger negativity when word valence was neutral as compared to positive (female speaker) or negative (female and male speaker).

The third time window that revealed a Prosody × Word × Sex interaction ranged from 750 to 900 msec, $F(4,240) = 3.85$, $p < .01$. Separate analyses for each sex indicated a significant interaction in women, $F(4,120) = 7.08$, $p < .0001$, but only a tendency toward an interaction in men, $F(4,120) = 2.0$, $p = .0987$. Again, the ERPs elicited to happily and angrily spoken words were more negative when word valence was incongruent as compared to congruent (Figure 6; Table 5). Neutrally spoken words elicited a more negative ERP to congruent (i.e., neutral words) as compared to incongruent trials (i.e., positive or negative words).

To determine whether the RT advantage found for our female participants explains the finding of sex differences in the ERP, a second ANOVA was performed with RTs as an additional factor. Participants were categorized as fast or slow responders. Nevertheless, the Prosody × Word × Sex interaction was significant for both the early [350–650 msec, $F(4,224) = 4.12$, $p < .01$] and the late time window [750–900 msec, $F(4,224) = 3.94$, $p < .01$], whereas the Prosody × Word × RT (both Fs < 1) and the Prosody × Word × Sex × RT interactions [350–650 msec, $F(4,224) = 1.81$, $p = .13$; 750–900 msec, $F < 1$] were not significant. Furthermore, separate ANOVAs for fast and slow responders for the early time window revealed a significant Prosody × Word × Sex interaction in both groups [fast, $F(4,112) = 2.79$, $p < .05$; slow, $F(4,112) = 3.4$, $p < .05$]. Post hoc comparisons indicated that women [fast, $F(4,56) = 3.89$, $p < .01$; slow, $F(4,56) = 4.27$, $p < .01$] but not men [fast, $F(4,56) = 1.61$, $p = .18$; slow, $F < 1$] showed an emotional Stroop effect. For the later time window, separate analyses for fast and slow

Table 3. Accuracy *F* Values: Post Hoc Comparisons for the Prosody × Word Interaction as Modified by Task and Voice

	df	*Prosodic Task* *Positive Words*	*Prosodic Task* *Neutral Words*	*Prosodic Task* *Negative Words*	*Semantic Task* *Positive Words*	*Semantic Task* *Neutral Words*	*Semantic Task* *Negative Words*
Prosody	2,122	8.05***	27.28****	9.74****	13.73****	9.84****	17.99****
Neutral versus negative prosody	1,61	7.18**	9.3**	<1	<1	20.56****	7.36**
Positive versus negative prosody	1,61	4.17#	27.37****	12.4***	15.18***	<1	28.26****
Positive versus neutral prosody	1,61	10.34**	32.92****	10.8**	25.82****	13.02***	12.62***

	df	*Female Speaker* *Positive Words*	*Female Speaker* *Neutral Words*	*Female Speaker* *Negative Words*	*Male Speaker* *Positive Words*	*Male Speaker* *Neutral Words*	*Male Speaker* *Negative Words*
Prosody	2,60	3.13*	4.32*	10.69****	<1	24.89****	14.02****
Neutral versus negative prosody	1,30	<1	8.42**	1.78		16.68***	3.78#
Positive versus negative prosody	1,30	5.64°	<1	14.65***		15.85***	20.05***
Positive versus neutral prosody	1,30	4.73#	4.79#	9.92**		34.28****	12.2**

For Bonferroni-adjusted single comparisons, °$p < .03$.

*$p < .03$.

**$p < .01$.

***$p < .001$.

****$p < .0001$.

#$p < .1$.

responders revealed only marginally significant Prosody × Word × Sex interactions [fast, $F(4,112) = 2.05$, $p = .092$; slow, $F(4,112) = 2.08$, $p = .088$], while the Prosody × Word interactions were significant [fast, $F(4,112) = 2.58$, $p < .05$; slow, $F(4,112) = 3.11$, $p < .05$]. Single comparisons, however, showed that women [fast, $F(4,56) = 3.71$, $p < .01$; slow, $F(4,56) = 3.89$, $p < .01$], but not men [fast, $F < 1$; slow, $F(4,56) = 1.17$, $p = .33$], contributed to this later interaction.

DISCUSSION

We conducted an emotional Stroop experiment to investigate the interaction between emotional prosody and word valence during speech processing. Participants listened to words that were either congruent or incongruent for the emotional information expressed at the prosodic and the semantic level. When participants were asked to indicate the valence of positive and negative words they responded faster and more accurately when prosody was congruent as compared to incongruent. A similar emotional Stroop effect, albeit smaller than in the semantic task and not as consistent, was elicited when participants judged prosodic valence. Additionally, there was ERP evidence for an emotional Stroop effect during the word valence, but not during the prosodic judgment. In contrast to the behavioral results, however, this ERP effect revealed sex differences as it was significant in women, but not in men. While the speaker's sex had some effect on the strength of the prosody–semantics interaction, it did not modulate the reported sex differences.

The present experiment was designed to answer three questions. First, we were interested in whether there are sex differences in the significance of emotional prosody and word valence for the comprehension of emotions in speech. Second, ERPs should reveal whether or not an interaction between emotional prosody and word valence occurs during a semantic processing stage. Finally, we asked whether any sex differences in the perception of emotional speech would be modulated by the speaker's sex.

The first question was derived from previous work on emotional processing that revealed women to be slightly better than men in the recognition of nonverbal emotional cues such as prosody (Hall, 1978). Furthermore, a recent cross-modal priming study suggests that women use emotional prosody earlier during speech processing than men (Schirmer et al., 2002). The present study was designed to determine whether, in addition to recognition accuracy and temporal processing characteristics, men and women differ in the influence of emotional prosody and word valence on the perception of emotions in speech. More specifically, we were interested in whether there are sex differences in the ability to ignore task-irrelevant emotional information such as prosody when making an emotional judgment. Behavioral measures indicate that this is not the case. Rather, they

Table 4. ERP *F* Values of the 50-msec Time Window Analysis

	df	*0 . . . 50*	*100*	*150*	*200*	*250*	*300*	*350*	*400*	*450*	*500*	*550*	*600*	*650*	*700*	*750*	*800*	*850 . . . 900*
Prosodic Task																		
Prosody × Channel × Hemisphere	34,2108													2.9****				
Prosody × Sex	2,120					3.23*												
Prosody × Voice	2,120																	4.73*
Prosody × Sex × Voice	2,120													3.79*				
Word	2,120													9.12***				
Word × Hemisphere × Voice	2,120															8.72***		
Semantic Task																		
Sex	1,60												5.99*					
Prosody × Channel	34,2040											12.97****						
Prosody × Voice	2,120									5.57**								
Prosody × Channel × Voice	34,2040															5.45****		
Word	2,120					8.41***												
Word × Channel	34,2040															19.24****		
Prosody × Word × Sex	4,240									4.40**								3.85**
Prosody × Word × Voice	4,240													3.16*				

*$p < .05$.

**$p < .01$.

***$p < .001$.

****$p < .0001$.

#$p < .1$.

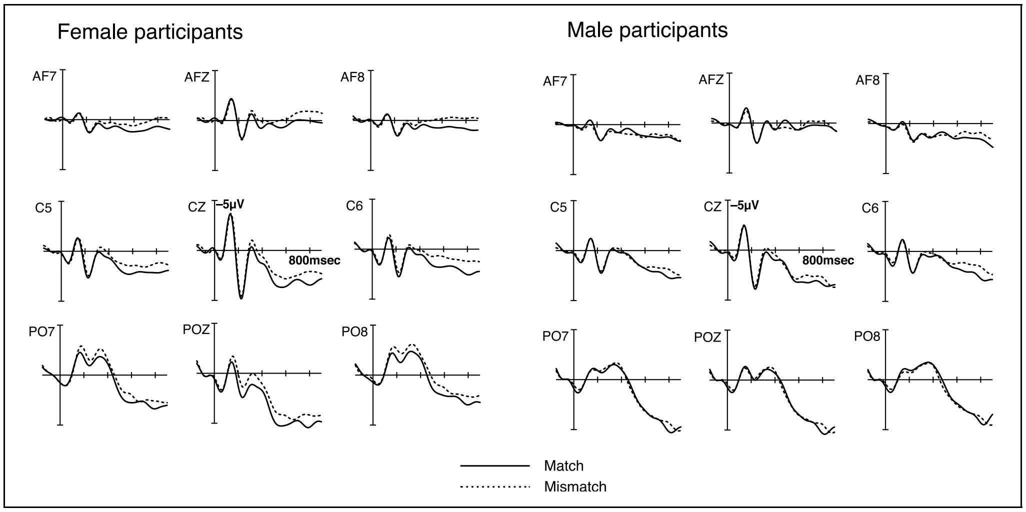


Figure 4. ERPs for the prosodic task. Positive words spoken with happy prosody and negative words spoken with angry prosody were averaged (i.e., match) and are represented by solid lines. Positive words spoken with angry prosody and negative words spoken with happy prosody were averaged (i.e., mismatch) and are represented by dotted lines. There was no significant difference between match and mismatch conditions.

suggest an equivalent role of prosody in male and female emotional comprehension. Both men and women showed a strong influence of prosodic information on the semantic judgment, whereas the influence of semantics on the prosodic judgment was comparatively small. Therefore, one can assume that when the prosodic expression is relatively easy to recognize, as was the case in the present study, prosodic expression influences an emotional judgment in men and women to the same degree. However, this might change when recognizing emotional prosody becomes more difficult, for example, when several speakers are involved or the prosodic expression is less clear.

The second question addressed the process that would be subject to interference in an emotional Stroop paradigm. Previous Stroop-like experiments suggest different possible loci for interference. Hock and Egeth (1970), for example, proposed that interference may arise during the perceptual encoding of a stimulus. Others place the locus of interference at a semantic processing stage (Rebai et al., 1997; Rayner & Springer, 1986; Seymour, 1977) or at the output stage caused by the parallel activation of the correct and the incorrect response (DeSoto, Fabiani, Geary, & Gratton, 2001; Masaki, Takasawa, & Yamazaki, 2000; Duncan-Johnson & Kopell, 1981). With respect to the present experiment, we expected a negativity elicited at around 400 msec following word onset to be indicative of interference at a semantic processing stage (Rebai et al., 1997). In accordance with this prediction, we found a negativity that started around 300 msec following word onset, and it was larger for emotionally incongruous as compared to congruous words. This effect showed a similar time course and scalp distribution as the N400 effect reported in the word processing literature (Holcomb & Anderson, 1993; Kutas & Hillyard, 1980). Therefore, it may reflect the integration of word meaning and prosody. However, another interpretation of the present N400 effect may be more appropriate. Given lexical integration is thought to be a fairly controlled process (Chwilla, Kolk, & Mulder, 2000; Brown & Hagoort, 1993), and listeners were instructed to ignore task-irrelevant prosodic information the present N400 effect may reflect the effort of keeping task-irrelevant information separate from processing task-relevant information.

In congruence with the behavioral results, the present N400 effect was elicited when participants evaluated the emotional meaning of a word, but not when they evaluated the prosodic expression. However, in contrast to the behavioral measures, this ERP effect differed between male and female listeners. Women, but not men, showed an interaction between prosody and se-

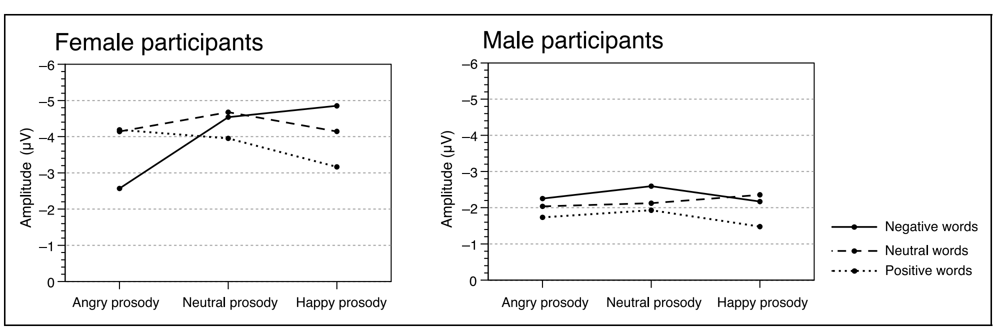


Figure 5. Mean amplitudes of the 350- to 650-msec time window in the ERPs for the semantic task. There was a significant interaction between prosody and word meaning in female but not in male participants.

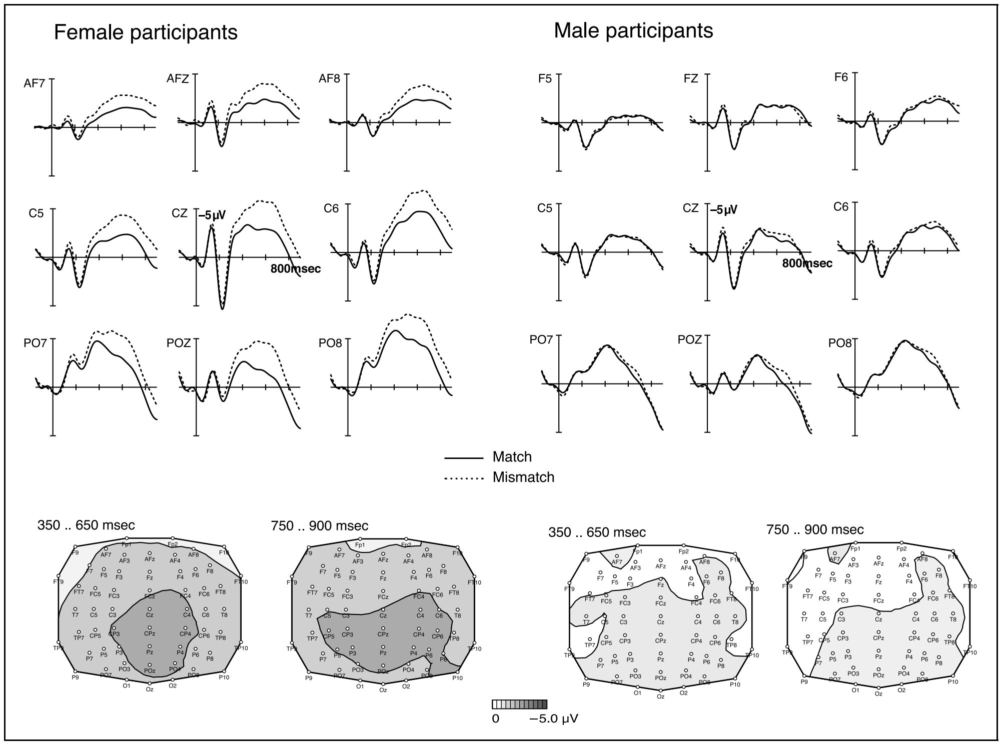


Figure 6. ERPs for the semantic task. Positive words spoken with happy prosody and negative words spoken with angry prosody were averaged (i.e., match) and are represented by solid lines. Positive words spoken with angry prosody and negative words spoken with happy prosody were averaged (i.e., mismatch) and are represented by dotted lines. There was a significant difference between match and mismatch conditions only in female participants. Scalp maps in the lower part of the figure illustrate the topographical distribution of the difference between match and mismatch conditions in the ERPs.

mantics as reflected in the N400. This divergence between behavioral and ERP measures is surprising and raises the question of why men showed interference from prosody in the former but not the latter measure. A tentative answer to this question is suggested by the different loci of interference reported in the literature. Men, in contrast to women, might be influenced by prosody at a stage that follows stimulus processing, namely response preparation. Parallel processing of emotional prosody and word valence might have triggered the activation of both the correct and the incorrect response for incongruent trials. Such a response conflict might have increased response latencies and judgment errors. Although this interpretation requires further probing with a design that allows examination of motor preparation processes, it receives preliminary support from a previous study on sex differences in emotional speech processing (Schirmer et al., 2002). In that study, women showed priming effects from emotional prosody on the processing of subsequent visual words with a shorter interstimulus interval than men. As this suggests that prosody exerts an earlier influence on word processing in women as compared to men, it seems reasonable to assume a similar processing difference between men and women in the present study. While women were influenced by prosody while processing the emotional meaning of a word, prosody in men might have exerted a later influence on response processes. An earlier interaction between word valence and emotional prosody in women might be due to an earlier access of emotional information. In accordance with this explanation, women performed both the prosodic and the semantic judgment approximately 100 msec faster than men. However, other explanations are also possible. For example, men and women may differ in whether they can separate semantic processes from emotional prosodic processing.

Finally, we examined whether the sex of the speaker modulates sex differences in emotional-prosodic processing. There is evidence for sender specific processing differences between men and women in the face recognition literature. Lewin and Herlitz (2002), for example, found that women have a better face recognition memory than men for female faces, whereas women do not differ from men in the recognition of male faces. Therefore, it is possible that sex differences in emotional processing are modulated by the sender's sex as well. To test this question, we employed a male and a female speaker in the present study. Except for minor differences, the behavioral and the ERP effects elicited by those speakers were comparable. Differences that were found between participants that listened to the male and the female speaker were restricted to the effect size of significant effects while the direction of those effects was comparable. For example, the RT difference between happily spoken positive and negative words was only marginal for participants that listened to the female speaker, but significant in participants that listened to the male speaker. In both cases, however, happily spoken negative words were responded to more slowly than happily spoken positive words. This suggests that there were differences in the salience of the three target emotions as expressed by the male and the female speaker and that these differences influenced the strength of the emotional Stroop effect. Most important, however, the differences between the participants that listened to the male and the female speaker occurred independently of the listener's sex. Moreover, the differences in the RTs and in the ERPs found between male and female listeners were not modulated by the speaker's sex. Therefore, one can conclude that sex differences in emotional speech processing occur regardless of whether participants listen to a male or a female voice.

Finally, it is necessary to examine the discrepancy between the present findings and earlier work concerning the significance of prosody and word valence for emotional speech comprehension. In two North American samples, a significant emotional Stroop effect was found only when participants were asked to judge emotional prosody but not when listening to word content (Kitayama & Ishii, 2002; Grimshaw, 1998). Interestingly, Kitayama and Ishii (2002) also tested a sample of Japanese listeners. In contrast to the two Western samples, they performed faster and more accurately in the word valence judgment when prosody and word meaning where congruous as compared to incongruous. In order to explain this discrepancy, the authors invoked the different role of nonverbal information such as prosody in so-called ''high contextual'' Asian cultures as compared to ''low contextual'' Western cultures (Kashima & Kashima, 1998; Hall, 1976). Moreover, Japanese listeners are thought to regard prosodic cues as more important and are, therefore, more easily influenced by prosody when judging the emotional meaning of a word. Furthermore, Kitayama and Ishii (2002) argued that in Western cultures, meaning is preferably expressed verbally and that, therefore, spontaneous attention is allocated to word content rather than prosody. The present findings, however, fail to support such an interpretation as they revealed an influence of prosody on judging word valence in a sample of Western (i.e., German) listeners. Moreover, this influence was even stronger than the effect word valence had on the prosodic judgment. Based on these discrepancies between the present findings and earlier work, it becomes clear that one cannot easily infer the significance of prosody or word valence based on an emotional Stroop task. Rather, specific details of the emotional Stroop paradigm are likely to modulate the obtained results. One factor influencing the effects of prosody on a word valence judgment is the salience of the vocally expressed emotion. As described earlier in the text, this can vary

Table 5. ERP *F* Values: Post hoc Comparisons for the Prosody × Word Interaction

		Women: Prosody × Word (350 ... 650)			*Women: Prosody × Word (750 ... 900)*		
	df	*Positive Prosody*	*Neutral Prosody*	*Negative Prosody*	*Positive Prosody*	*Neutral Prosody*	*Negative Prosody*
Word	2,60	7.21**	<1	14.62****	18.46****	8.89***	25.68****
Neutral versus negative words	1,30	<1		22.46****	12.68**	7.72**	56.98****
Positive versus negative words	1,30	20.1****		23.94****	11.17**	<1	14.24***
Positive versus neutral words	1,30	3.66#		<1	25.45****	18.29***	10.9**
		Female Voice: Prosody × Word (600 ... 750)			*Male Voice: Prosody × Word (600 ... 750)*		
	df	*Positive Prosody*	*Neutral Prosody*	*Negative Prosody*	*Positive Prosody*	*Neutral Prosody*	*Negative Prosody*
Word	2,60	19.85****	7.12**	3.75*	7.22**	11.29****	7.36**
Neutral versus negative words	1,30	5.79°	6.18°	4.95#	4.83#	<1	20.79****
Positive versus negative words	1,30	16.48***	<1	6.51°	3.38#	10.62**	<1
Positive versus neutral words	1,30	32.48****	13.81***	<1	11.32**	18.95****	3.97#

For Bonferroni-adjusted single comparisons, °$p < .03$.

*$p < .05$.

**$p < .01$.

***$p < .001$.

****$p < .0001$.

#$p < .1$.

within and across different speakers and modulates the interaction between emotional prosody and word valence. Another factor that might explain the differences between the present study and earlier work concerns the stimulus categories presented. In addition to emotionally valenced stimuli such as positive and negative words, we presented neutral stimuli. This might have resulted in increased task demands especially with respect to the semantic judgment. Rather than making a rapid decision between ''good'' and ''bad,'' participants may have checked each word for its emotional strength and whether or not this was different from neutral. This checking must have been more difficult for neutral as compared to positive and negative words, as neutral words elicited the longest RTs. Given this, the neutral condition could not act as a baseline for the discrimination between facilitative and inhibitory effects of task-irrelevant emotional information. Furthermore, the occurrence of neutral words within the present design contributed to the relatively stronger effects of prosody as compared to word valence on the processing of task-relevant emotional information. Together, the present findings and earlier work suggest that task difficulty modifies the emotional Stroop effect: The more difficult an emotional judgment (e.g., word valence) the more it is influenced by emotional information from other sources (e.g., prosody). Moreover, it seems that the interaction between emotional prosody and word valence is not fixed within a language, culture, or sex. Rather, we assume that this interaction is a flexible process that is modulated by situational factors such as the expressiveness of the speaker or the occurrence of neutral words together with emotional words. A further modulating factor that is relevant with respect to everyday speech is the relative timing of processing prosody as compared to semantics. While listeners can process prosodic emotions at sentence beginning, they sometimes have to wait until the end of the sentence to be certain about verbally expressed emotions. That preceding task-irrelevant information can enhance Stroop interference as compared to simultaneous presentation has been demonstrated by Glaser and Glaser (1982) and may apply to emotional processing as well. Therefore, emotional Stroop interference may also differ depending on what the relative timing of prosodic and semantic processing is.

In sum, the present study puts earlier work suggesting verbal information dominates emotional speech comprehension in Western listeners into perspective. Moreover, we found that emotional prosody is a powerful instrument for the expression of emotions that can, under certain circumstances, overshadow the effects of verbal emotions even in so called ''low contextual'' cultures such as German (Kashima & Kashima, 1998; Hall, 1976). Furthermore, the present findings contribute to the understanding of sex differences in emotional speech comprehension. Both men and women showed an emotional Stroop effect for the behavioral response, which suggests that both were sensitive to the emotions conveyed in prosody and word meaning. That only women showed an influence of prosody on semantic processing is taken as evidence that men and women differ in what they do with the encoded emotional information. More specifically, women seem to integrate emotional information from prosody more readily into ongoing semantic processing.

METHODS

Subjects

Seventy-one subjects were invited to participate in the experiment. Thirty-six listened to the stimulus material spoken by a female speaker. Because of excessive EEG artifacts, three of these subjects were excluded from the data analysis. Another subject was excluded because his behavioral responses were more than 2 standard deviations slower than the group mean. Sixteen of the remaining 32 subjects were women with a mean age of 22.9 (SD = 2.3). The 16 male participants had a mean age of 23.5 (SD = 2.9). Thirty-five out of 71 subjects listened to the stimulus material spoken by a male speaker. The error rate of two participants was too high to provide enough trials for the ERP analysis. Another subject was excluded from data analysis because of excessive EEG artifacts. Sixteen of the remaining 32 subjects were women with a mean age of 24.0 (SD = 4.1). The 16 male participants had a mean age of 24.1 (SD = 2.3).

Material

The stimulus material consisted of 74 positive, 74 neutral, and 74 negative verbs. Word valence was obtained in an earlier rating study. Twenty-three male and 23 female subjects rated all words on a five-point scale that ranged from 2 to −2 for emotional valence. Men and women did not differ in how emotional they rated the stimuli. Negative words had a mean valence of −1.12 (SD = 0.29), neutral words of 0.04 (SD = 0.23), and positive words of 1.05 (SD = 0.25). Both negative and positive words differed significantly from neutral. Furthermore, there was no significant difference in valence strength between positive and negative words: Positive words were rated just as emotional as negative words. Additionally, words were controlled for word frequency so that there was no difference between the three valence conditions Baayen, Piepenbrock, and van Rijn (1995). A female and a male native speaker of German produced all words with happy, neutral, and angry prosody. Words were taped with a DAT recorder and digitized at a 16-bit/44.1-kHz sampling rate. The stimulus material was divided into two lists. Each list contained 37 positive, 37 neutral, and 37 negative words spoken with a happy, neutral, or angry voice. Each list was presented under a prosodic and under a semantic instruction. Whenever a subject heard one list under a prosodic instruction he or she heard the other list under a semantic instruction. Half the participants listened to the stimulus material spoken by the female speaker and half the participants listened to the stimulus material spoken by the male speaker.

Procedure

Testing was carried out in a sound-proof and electrically shielded chamber. Subjects were seated in a comfortable chair facing a computer monitor at a distance of 1.15 m. An experimental session consisted of two blocks. In one block, subjects had to judge word valence as either positive, neutral or negative while ignoring the prosody. In another block, subjects had to judge emotional prosody as positive, neutral, or negative while ignoring word valence. Block order was counterbalanced across subjects. Responses were given by pressing one of three buttons of a response box as quickly and as accurately as possible. Subjects used the middle and the index finger of the right hand to press the right and the middle button; they used their left hand index finger to press the left button. Half the subjects pressed the left button for positive responses and the right button for negative responses; the remaining subjects had the reverse button assignment. Neutral responses were always assigned to the middle button. To reduce baseline artifacts, the intertrial intervals were 3000 msec, 3100 msec, and 3200 msec for a third of the trials each. Thirty-six practice trials preceded every experimental block.

ERP Recording and Data Analysis

The electroencephalogram (EEG) was recorded from 50 electrodes mounted in an elastic cap according to the International 10–20 system. The sampling rate was 250 Hz. 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 kΩ. ERP averages were computed with a 150-msec baseline and a 900-msec ERP time window. Trials containing eye blinks or movement artifacts were omitted from the data analysis. On the average, 8.4% of the trials were rejected for the semantic task and 10.46% of the trials were rejected for the prosodic task. Grand averages were smoothed with an 8-Hz low-pass filter for illustration only.

For statistical analysis, we first conducted 50-msec time window ANOVAs for each task with sex and voice as between-subject factors and prosody, word, hemisphere, and channel as within-subject factors. Left (F7, F5, F3, FT7, FC5, FC3, T7, C5, C3, TP7, CP5, CP3, P7, P5, P3, PO7, PO3, O1) and right hemisphere electrodes (F8, F6, F4, FT8, FC6, FC4, T8, C6, C4, TP8, CP6, CP4, P8, P6, P4, PO8, PO4, O2) constituted the factor hemisphere. The factor channel had 18 levels resulting from the 18 electrode positions over right and left hemispheres. Because of the increased likelihood of Type I errors associated with the large number of comparisons, only effects that reached significance in more than two consecutive time windows were considered significant. For significant effects, a second ANOVA was performed summarizing significant consecutive time windows into one larger time window. *p* Values of post hoc single comparisons were corrected using a modified Bonferroni procedure (see Keppel, 1991).

Acknowledgments

We would like to thank T. Penney, A.D. Friederici, C. Friedrich, M. Besson, and two anonymous reviewers for comments on the manuscript. This work was supported by the German Research Foundation DFG and the Leibniz Science Price awarded to A.D. Friederici.

Reprint requests should be send to Annett Schirmer, Max Planck Institute of Cognitive Neuroscience, Stephanstrasse 1a, 04103 Leipzig, Germany, or via e-mail: schirmer@cns.mpg.de.

REFERENCES

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1995). The CELEX Database. Nijmegen: Center for Lexical Information, Max Planck Institute for Psycholinguistics, CD-ROM.

Blonder, L. X., Bowers, D., & Heilman, K. M. (1991). The role of the right hemisphere in emotional communication. *Brain, 114,* 1115–1127.

Brown, C., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience, 5,* 34–44.

Chwilla, D. J., Kolk, H. J., & 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.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review, 97,* 332–361.

DeSoto, M. C., Fabiani, M., Geary, D. C., & Gratton, G. (2001). When in doubt, do it both ways: Brain evidence of the simultaneous activation of conflicting motor responses in a spatial Stroop task. *Journal of Cognitive Neuroscience, 13,* 523–536.

Duncan-Johnson, C. C., & Kopell, B. S. (1981). The Stroop effect: Brain potentials localize the source of interference. *Science, 214,* 938–940.

Fazio, R. H., Sanbonmatsu, D. M., Powel, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50,* 229–238.

Glaser, M. O., & Glaser, W. R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 875–894.

Greenham, S. L., Stelmack, R. M., & Campbell, K. B. (2000). Effects of attention and semantic relation on event-related

potentials in a picture–word naming task. *Biological Psychology, 55,* 79–104.
Grimshaw, G. M. (1998). Integration and interference in the cerebral hemispheres: Relations with hemispheric specialization. *Brain and Cognition, 36,* 108–127.
Hall, E. T. (1976). *Beyond culture.* New York: Doubleday.
Hall, J. A. (1978). Gender effects in decoding nonverbal cues. *Psychological Bulletin, 4,* 845–857.
Hock, H. S., & Egeth, H. (1970). Verbal interference with encoding in a perceptual classification task. *Journal of Experimental Psychology, 83,* 299–303.
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,* p1379–411.
Kashima, E. S., & Kashima, Y. (1998). Culture and language: The case of cultural dimensions and personal pronoun use. *Journal of Cross-Cultural Psychology, 29,* 461–486.
Kayser, J., Bruder, G. E., Tenke, C. E., Stewart, J. W., & Quitkin, F. M. (2000). Event-related potentials (ERPs) to hemifield presentations of emotional stimuli: Differences between depressed patients and healthy adults in P3 amplitude and asymmetry. *International Journal of Psychophysiology, 36,* 211–236.
Keppel, G. (1991). *Design and analysis: A researcher's handbook.* Englewood Cliffs, NJ: Prentice Hall.
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. (in press). On the lateralization of emotional prosody: An event-related functional MR investigation. *Brain and Language.*
Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207,* 203–205.
Lewin, C., & Herlitz, A. (2002). Sex differences in face recognition - women's faces make the difference. *Brain and Cognition, 50,* 121–128.
Ley, R. G., & Bryden, M. P. (1982). A dissoziation of left and right hemispheric effects for recognizing emotional tone and verbal content. *Brain and Cognition, 1,* 3–9.
Liotti, M., Wolldorf, M. G., Perez, R., Mayberg, H. S. (2000). An ERP study of the temporal course of the Stroop color–word interference effect. *Neuropsychologia, 38,* 701–711.
MacLeod, C. M., & MacDonald, P. A. (2000). Interdimensional interference in the Stroop effect: Uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences, 4,* 383–391.
Masaki, H., Takasawa, N., & Yamazaki, K. (2000). An electrophysiological study of the locus of the interference effect in a stimulus–response compatibility paradigm. *Psychophysiology, 37,* 464–472.
Naumann, E., Bartussek, D., Diedrich, O., & Laufer, M. E. (1992). Assessing cognitive and affective information processing functions of the brain by means of the late positive complex of the event-related potential. *Journal of Psychophysiology, 6,* 285–298.
Pell, M. (1996). On the perceptive prosodic loss in Parkinson's disease. *Cortex, 32,* 693–704.
Rayner, K., & Springer, C. J. (1986). Graphemic and semantic similarity effects in the picture–word interference task. *British Journal of Psychology, 77,* 207–222.
Rebai, M., Bernard, C., & Lannou, J. (1997). The Stroop test evokes a negative brain potential, the N400. *International Journal of Neuroscience, 91,* 85–94.
Rotter, N. G., & Rotter, G. S. (1988). Sex differences in the encoding and decoding of negative facial emotions. *Journal of Nonverbal Behavior, 12,* 139–148.
Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology, 32,* 76–92.
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.
Starkstein, S. E., Federoff, J. P., Price, T. R., Leiguarda R. C., & Robinson, R. G. (1994). Neuropsychological and neuroradiologic correlates of emotional prosody comprehension. *Neurology, 44,* 515–522.
Seymour, P. H. K. (1977). Conceptual encoding and locus of the Stroop effect. *Quarterly Journal of Experimental Psychology, 29,* 245–265.