Sex differences in the preattentive processing of vocal emotional expressions

Annett Schirmer,[1,CA] Tricia Striano[2] and Angela D. Friederici[3]

[1]Department of Psychology, University of Georgia, Athens, GA 30602-3013, USA; [2]Max Planck Institute for Evolutionary Anthropology, Germany; [3]Max Planck Institute for Human Cognitive and Brain Sciences, Germany

[CA,2]Corresponding Author and Address: schirmer@uga.edu

Emotional tone of voice indicates the significance of a spoken utterance. We asked whether listeners recognize this significance even when attending to something else and whether men and women differ in this regard. To answer these questions, we presented emotionally or neutrally spoken syllables as standards and deviants in a mismatch negativity paradigm. Independent of the listeners' sex, deviants elicited a mismatch negativity in the scalp-recorded event-related potential as an indicator of preattentive acoustic change detection. Only women, however, showed a larger mismatch negativity to emotional than to neutral deviants. Thus, even though both sexes detect change in voice preattentively, only women recruit additional processing resources when the change in voice is one of emotional valence. *NeuroReport* 16:635–639 © 2005 Lippincott Williams & Wilkins.

Key words: Affective prosody; Arousal; Emotional prosody; Emotional valence; Gender differences; Lateralization

INTRODUCTION

Spoken language relies on the vocal expression system, which serves in most mammals as a physiologically mediated indicator of emotional arousal [1]. As such, spoken language consists of both a linguistic component and an emotional component. The emotional component is referred to as emotional prosody and can be described by modulations in the acoustic parameters of speech, such as intensity, rate and fundamental frequency ($F0$). A specific profile of these parameters can be indicative of a specific emotion. For example, hot anger is characterized by a relatively high intensity, speech rate and mean $F0$, whereas sadness is characterized by low intensity, speech rate and mean $F0$ [2]. Because of its linkage with emotional arousal, emotional prosody or emotional vocalization, in general, reflect the significance of an internal or external event and therefore has important communicative functions. For example, angry or fearful vocalizations in response to an aggressor will more likely attract the attention of others than neutral vocalizations. This signalling function, which has also been reported in primates [3,4], kept its significance throughout human ontogeny and the evolution of speech. Moreover, it plays an important role in everyday communication and has the potential to overwrite a conflicting verbal message (e.g. sarcasm).

Studies that investigated the recognition of emotional prosody frequently reported differences between men and women. As in the domain of facial emotion recognition, women identify a speaker's emotional state more accurately and faster than men [5–7]. Additionally, recent event-related potential (ERP) studies found sex differences in the sensitivity to emotional prosody. Compared with men, women made use of emotional prosody at an earlier point in time during word processing [8]. Furthermore, women, but not men, integrated emotional prosody into word processing when emotional prosody was task-irrelevant [6]. These findings are in congruence with proposals of sex differences in social orientation. According to these proposals, women, more than men, define themselves in relational terms [9–11]. Thus, information about the emotional state of others may be more relevant for women and consequently processed more readily than in men.

The present study investigated whether listeners pick up the significance of emotional prosody when speech is unattended background noise. Moreover, we were interested in whether womens' enhanced sensitivity to emotional prosody reflects automatized processes. If true, sex differences should show also for preattentive emotional-prosodic processing. Specifically, women should be more likely than men to recruit additional processing resources for emotional prosody compared with neutral prosody when speech is unattended. To test these predictions we employed a mismatch negativity paradigm. In this paradigm, listeners generally read a book or watch a silent movie while a sequence of acoustic events is presented. In this sequence, a standard stimulus (e.g. 1000 Hz tone) is repeated and occasionally interrupted by a deviant stimulus (e.g. 1032 Hz tone). Subtracting standards from deviants in the ERP reveals a negative component peaking at around 200 ms following stimulus onset [12]. This component has been termed mismatch negativity (MMN) and is thought to reflect a preattentive memory-based comparison by which listeners detect changes in their environment [13]. This comparison mechanism is determined by both the salience of acoustic change [14] and the expertise that listeners have in processing a specific auditory stimulus [15]. The easier a

stimulus is to discriminate the earlier and larger is the MMN [16]. Given the signalling function of emotional prosody [1], we hypothesized that an emotionally spoken deviant would be more salient and discriminated more easily than a neutrally spoken deviant. Therefore, the emotional deviant should elicit a larger and earlier MMN than the neutral deviant. On the basis of the finding of sex differences in emotion recognition [7] and the use of emotional prosody during language processing [6,8], we furthermore predicted that women would be more likely than men to process emotional prosody preattentively. Accordingly, MMN differences between emotional and neutral utterances were expected to be more pronounced in women than in men.

MATERIALS AND METHOD

Participants: Eighty-two participants were invited for the study. One was excluded from data analysis because of technical problems during the electroencephalogram (EEG) recording. Another participant was excluded because of drift artifacts in the EEG. Forty of the remaining participants were men with a mean age of 24.8 (SD 2.9) years. Women were on average 24.3 (sd 2.3) years old. All had normal hearing and vision or corrected-to-normal vision. They received 6 Euros per hour for participation in the study. The experiments were conducted with the understanding and the written consent of each participant, and an ethical committee formally approved the experiments.

Stimulus material and design: A female speaker produced the syllables 'dada' several times with angry and neutral prosody. Another female speaker produced the syllables 'dada' with happy and neutral prosody. Each set of syllables was rated for emotionality by a group of six listeners (three men). For the angry/neutral set, listeners classified each stimulus as either very angry, angry, neutral or emotional (with the emotion not being anger). For the happy/neutral set listeners classified each stimulus as either very happy, happy, neutral or emotional (with the emotion not being happiness). Two syllables that had been uniquely identified as 'very angry' and 'neutral' were selected from the angry/neutral set. Two syllables that had been uniquely identified as 'very happy' and 'neutral' were selected from the happy/neutral set. Emotional and neutral syllables were equally long (angry/neutral: 557 ms; happy/neutral: 573 ms) and loud (angry/neutral: 67 dB max, 56 dB mean; happy/neutral: 72 dB max, 64 dB mean) but differed with respect to $F0$ and other frequency formants. One experimental group, consisting of 20 men and 20 women, listened to the angry/neutral stimuli. The other group, consisting of the remaining participants, listened to the happy/neutral stimuli. An experimental session involved of two blocks with 1050 standards ($p=0.875$) and 150 deviants ($p=0.125$) in each block. The emotional and the neutral stimuli served as standards and deviants in two separate blocks. In one block, neutral syllables were presented as standards and emotional syllables as deviants. In the other block, emotional syllables were presented as standards and neutral syllables as deviants. Presenting the same syllables as both standards and deviants allowed us to investigate the MMN by comparing physically identical stimuli, thereby reducing the influence of acoustic differences between the neutral and the emotional conditions on MMN amplitude. Block order

was counterbalanced across participants. The syllable onset to syllable onset interval in each block was 1200 ms.

Procedure: Participants were seated in a comfortable chair facing a monitor that showed a silent, self-selected movie with subtitles. Syllables were presented over loudspeakers positioned to the left and right of the monitor. The EEG was recorded from 23 electrodes positioned according to the modified expanded 10–20 system. Electrodes were referenced to linked mastoids. Two bipolar recordings were used to control for horizontal and vertical eye movements. The EEG was filtered offline with a 0.5–15 Hz bandpass filter.

Data analysis: To determine whether standards differed significantly from deviants in the time window of the MMN across conditions and groups, we conducted an overall ANOVA for mean voltages between 100 and 300 ms following stimulus onset. This ANOVA included frequency (standard/deviant), prosody (emotional/neutral), electrode (all electrodes) as repeated measures factors and group (anger/happiness) and sex as between-participants factors. To further specify MMN amplitude and latency, we subtracted standards from deviants and determined the negative maximum within the 100–300 ms time range. MMN amplitude and latency were then subjected to an ANOVA with prosody (emotional/neutral), electrode site (anterior/central/posterior) and laterality (left/middle/right) as repeated measures factors and group (anger/happiness) and sex as between-participants factors. MMN topography was explored using the electrodes F7, Fz, F8, C7, Cz, C8, P7, Pz and P8, which entered the factors electrode site (anterior/central/posterior) and laterality (left/middle/right) according to their position on the scalp.

As indicated above, we used each stimulus as standard and deviant in order to avoid that physical differences between standard and deviant could confound MMN amplitude. In other words, we compared emotionally spoken deviants with emotionally spoken standards and neutrally spoken deviants with neutrally spoken standards. However, we were unable to control for the preceding stimulus, which was a neutrally spoken syllable for the emotional deviant and an emotionally spoken syllable for the neutral deviant. Thus, the direction of change was different in both cases and the associated acoustic effects may have influenced the MMN amplitude independently of emotional salience. Research by Näätänen [16] addressed the possible effect of direction of change on MMN amplitude. He found that regardless of whether a deviant is louder or softer than the standard, as long as the physical distance is identical in both cases the MMN is identical also. Therefore, we assume that purely acoustic effects associated with the direction of change are unlikely to modify MMN amplitude in the present design. Additionally, sex differences have been reported in the processing of emotional information but not in the perception of acoustic change. Thus, if women show a larger MMN to emotional than to neutral syllables while men do not, it is likely that this effect reflects the perception of emotionally salient information.

RESULTS

The ERPs elicited to standards and deviants are illustrated in Fig. 1. Overall statistical analysis revealed a frequency

main effect [$F(1,76)=14.51$, $p<0.001$] indicating that listeners responded differently to standards and deviants and showed a significant MMN. An interaction involving frequency [$F(22,1672)=2.65$, $p<0.0001$], furthermore, suggests that the MMN was modulated by prosody, electrode, group, and sex. We conducted an analysis of MMN amplitudes and latencies to further specify the influence of these factors (see Fig. 2). For peak amplitudes, this analysis revealed a main effect of prosody [$F(1,75)=4.84$, $p<0.05$] and a prosody by sex interaction [$F(1,75)=5.63$, $p<0.05$]. Separate analyses for men and women indicated that women [$F(1,38)=7.34$, $p<0.01$], but not men ($F<1$), showed larger MMN amplitudes for emotional than for neutral deviants. Additionally, a prosody by electrode site by laterality interaction [$F(4,300)=3.38$, $p<0.05$] showed that this prosody effect was localized over anterior middle [$F(1,75)=6.06$, $p<0.05$], anterior right [$F(1,75)=5.07$, $p<0.05$] and central right [$F(1,75)=10.75$, $p<0.01$] electrode sites. Analysis of the peak latency measures revealed a main effect of prosody [$F(1,75)=44.27$, $p<0.0001$] that was modulated by group [$F(1,75)=33.07$, $p<0.0001$]. The emotional deviant elicited an earlier MMN than the neutral deviant, in participants who listened to syllables with happy and neutral prosody [$F(1,38)=64.4$, $p<0.0001$], but not in participants who listened to syllables with angry and neutral prosody ($F<1$).

DISCUSSION

The present study investigated whether listeners process emotional prosody preattentively and whether men and women differ in this regard. With respect to our first question, we found that emotional prosody modulated the MMN, an ERP correlate of preattentive acoustic change detection. MMN amplitude was larger for angry and happy deviants relative to neutral deviants. This effect was lateralized to the right hemisphere, which accords with the presumed role of the right hemisphere in emotional processing [17,18]. Emotional prosody also modulated MMN latency. Happy deviants elicited a shorter MMN latency relative to neutral deviants, while there was no difference in MMN latency between angry and neutral deviants. Given that happiness and anger are both high arousal emotions, but differ in valence, these findings suggest differential effects of arousal and valence on the MMN. Arousal seems to increase MMN amplitude and thus the processing resources engaged in change detection. In contrast, valence seems to modulate MMN latency and thus the temporal constraints of change detection. The arousal effect is in congruence with the assumption that emotionally arousing vocalizations have a signalling function and are more likely to attract the attention of the listener than neutral vocalizations [1]. Given previous evidence for a correlation between MMN peak latency and response time [19], the present valence effect associated with MMN peak
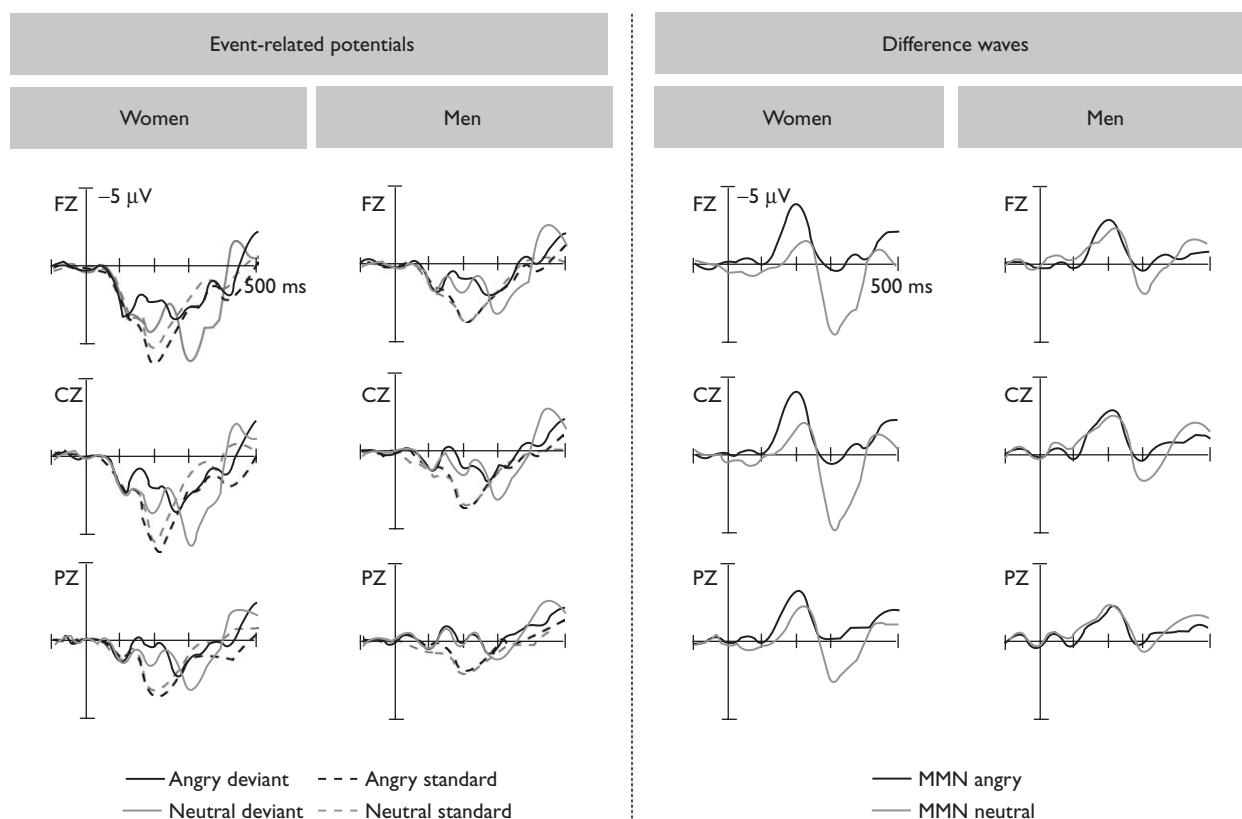

Fig. 1. Event-related potentials for angry standard and deviant syllables and neutral standard and deviant syllables are illustrated on the left. Difference waves obtained by subtracting angry standards from angry deviants and neutral standards from neutral deviants are illustrated on the right.
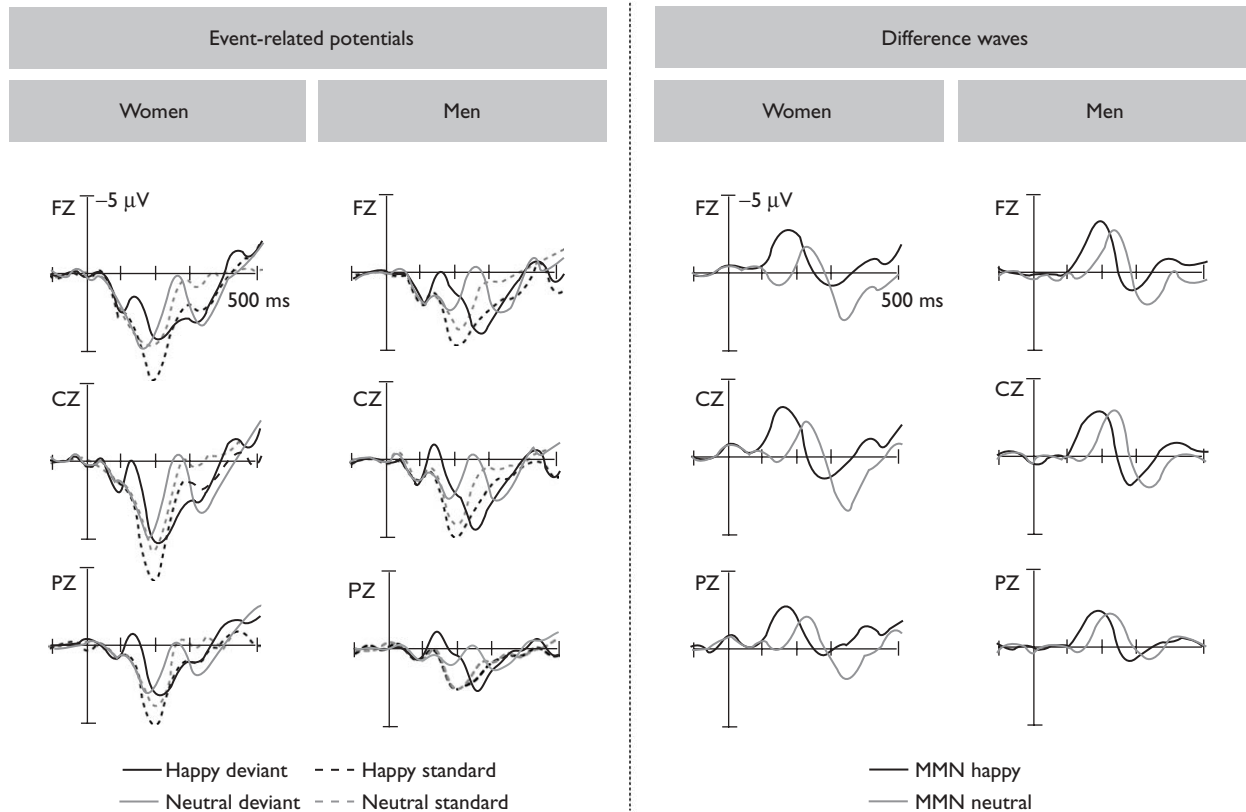
Fig. 2. Event-related potentials for happy standard and deviant syllables and neutral standard and deviant syllables are illustrated on the left. Difference waves obtained by subtracting happy standards from happy deviants and neutral standards from neutral deviants are illustrated on the right.

latency accords with previous research, revealing faster reaction times to positive than to neutral or negative stimuli [8,20].

Most interesting with respect to our predictions, however, is the finding that both sexes show an influence of emotional prosody on MMN latency, whereas only women show an influence of emotional prosody on MMN amplitude. This suggests that both sexes were sensitive to emotional valence, but that only women responded to the emotional arousal associated with happy and angry prosody. Given the correlation of MMN amplitude and discrimination ability [21], these findings accord with women's superior emotion recognition [7], and greater susceptibility to influences from emotional prosody on word processing [6]. Additionally, they are in congruence with reports of a higher relevance of social information for women than for men. According to these reports, women, more than men, define themselves in relational terms [9–11]. The higher relevance of social relationships makes women more dependent on the emotional state of others, which may enhance their processing of vocal emotional expressions. A further aspect that may contribute to the observed sex differences is women's role in childcare. In this context, there is evidence for a larger psychophysiological response to infant crying in women than in men [22]. Moreover, women have been found to identify emotional expressions in infants faster and more accurately than men [23]. These findings led researchers to speculate that, through evolution, selective pressure in-

creased sensitivity to emotional cues in women, who are the primary caretaker of offspring [23]. The present finding of women showing enhanced emotional processing, even when emotional information is unattended, accords with this view.

CONCLUSION
One can conclude that sex differences in the processing of emotional expressions do not merely reflect strategic differences between men and women. Rather, these sex differences seem to emerge at a relatively early, automatized stage of information processing.

REFERENCES
1. Scherer KR. Vocal correlates of emotional arousal and affective disturbance. In: Wagner H, Manstead A (eds). *Handbook of Social Psychophysiology*. Oxford: Wiley & Sons; 1989. pp. 165–197.
2. Banse R, Scherer KR. Acoustic profiles in vocal emotion expression. *J Pers Soc Psychol* 1996; **70**:614–636.
3. Scherer KR, Kappas A. Vocal affect expression in primates. In: Todt D, Goedeking P, Newman E (eds). *Primate Vocal Communication*. Heidelberg: Springer; 1988. pp. 171–194.
4. Hauser MD.*The Evolution of Communication*. Cambridge: MIT Press; 1997.
5. Rahman Q, Wilson GD, Abrahams S. Sex, sexual orientation, and identification of positive and negative facial affect. *Brain Cogn* 2004; **54**:179–185.

6. Schirmer A, Kotz SA. ERP evidence for a gender specific Stroop effect in emotional speech. *J Cogn Neurosci* 2003; **15**:1135–1148.

7. Hall JA. Gender effects in decoding nonverbal cues. *Psychol Bull* 1978; **4**:845–857.

8. Schirmer A, Kotz SA, Friederici AD. Sex differentiates the role of emotional prosody during word processing. *Cogn Brain Res* 2002; **14**:228–233.

9. Clancy SM, Dollinger SJ. Photographic depictions of the self: gender and age differences in social connectedness. *Sex Roles* 1993; **29**:477–495.

10. Dollinger SJ, Preston LA, O'Brien SP, DiLalla DL. Individuality and relatedness of the self: an autophotographic study. *J Pers Soc Psychol* 1996; **71**:1268–1278.

11. Kashima Y, Kokubo T, Kashima ES, Boxall D, Yamaguchi S, Macrae K. Culture and self: are there within-culture differences in self between metropolitan areas and regional cities? *Pers Soc Psychol Bull* 2004; **30**:816–823.

12. Sams M, Paavilainen P, Alho K, Näätänen R. Auditory frequency discrimination and event-related potentials. *Electroencephalogr Clin Neurophysiol* 1985; **62**:437–448.

13. Näätänen R, Paavilainen P, Alho K, Reinikainen K, Sams M. Do event-related potentials reveal the mechanism of the auditory sensory memory in the human brain? *Neurosci Lett* 1989; **98**:217–221.

14. Tiitinen H, May K, Reinikainen K, Näätänen R. Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature* 1994; **372**:90–92.

15. Näätänen R, Lehtokoski A, Lennes M, Cheour M, Huotilainen M, Iivonen A *et al.* Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 1997; **385**:432–434.

16. Näätänen R. *Attention and Brain Function*. Hillsdale: Lawrence Erlbaum Associates; 1992.

17. Buchanan TW, Lutz K, Mirzazade S, Specht K, Shah NJ, Zilles K *et al.* Recognition of emotional prosody and verbal components of spoken language: an fMRI study. *Cognitive Brain Research* 2000; **9**:227–238.

18. Gandour J, Wong D, Dzemidzic M, Lowe M, Tong Y, Li X. A cross-linguistic fMRI study of perception of intonation and emotion in Chinese. *Hum Brain Map* 2003; **18**:149–157.

19. Novak G, Ritter W, Vaughan HG Jr. Mismatch detection and the latency of temporal judgements. *Psychophysiology* 1992; **29**:398–411.

20. Wentura D. Activation and inhibition of affective information: evidence for negative priming in the evaluation task. *Cogn Emotion* 1999; **13**:65–91.

21. Näätänen R, Winkler I. The concept of auditory stimulus representation in cognitive neuroscience. *Psychol Bull* 1999; **125**:826–859.

22. Furedy JJ, Fleming AS, Ruble D, Scher H, Daly J, Day D *et al.* Sex differences in small-magnitude heart-rate responses to sexual and infant-related stimuli: a psychophysiological approach. *Physiol Behav* 1989; **46**:903–905.

23. Babchuk WA, Hames RB, Thompson RA. Sex differences in the recognition of infant facial expressions of emotion: the primary caretaker hypothesis. *Ethol Sociobiol* 1986; **6**:89–101.