Available online at www.sciencedirect.com
ScienceDirect
Brain and Language 104 (2008) 262–269
Brain and Language
www.elsevier.com/locate/b&l

How aging affects the recognition of emotional speech

Silke Paulmann [a,*], Marc D. Pell [b], Sonja A. Kotz [a,c]

[a] Max Planck Institute for Human Cognitive and Brain Sciences, P.O. Box 500 355, 04303 Leipzig, Germany
[b] School of Communication Sciences and Disorders, McGill University, Montréal, Canada
[c] Day Care Clinic of Cognitive Neurology, University of Leipzig, Germany

Accepted 3 March 2007
Available online 10 April 2007

Abstract

To successfully infer a speaker's emotional state, diverse sources of emotional information need to be decoded. The present study explored to what extent emotional speech recognition of 'basic' emotions (*anger*, *disgust*, *fear*, *happiness*, *pleasant surprise*, *sadness*) differs between different sex (male/female) and age (young/middle-aged) groups in a behavioural experiment. Participants were asked to identify the emotional prosody of a sentence as accurately as possible. As a secondary goal, the perceptual findings were examined in relation to acoustic properties of the sentences presented. Findings indicate that emotion recognition rates differ between the different categories tested and that these patterns varied significantly as a function of age, but not of sex.

Keywords: Emotion; Prosody; Vocal expression; Aging; Sex

1. Introduction

Communicating and understanding emotions is central to human social interactions throughout the lifespan. From the first moments of life humans express their feelings (e.g., babies scream when feeling pain, hunger, fear, or happiness). In adulthood, individuals must decode the emotional expressions of others effectively or risk a breakdown in interpersonal communication. It is therefore of major social relevance to understand how emotions are encoded and decoded and to determine which factors influence these processes. As a contribution to this literature, the present study investigated the recognition[1] of 'basic' emotions from speech in two distinct age and sex cohorts.

Much of the literature on emotional expression has focused on emotional facial expression (e.g., Ekman, 1972, 1982, 1992). More recently, interest in *emotional speech*, by which we refer to both the verbal-semantic content and the prosodic realization of an emotional utterance, has increased (for a recent review see Juslin & Laukka, 2003). For instance, Banse and Scherer (1996) studied the recognition of emotional prosody and reported that accuracy rates varied significantly as a function of the emotion category; whereas *hot anger* was recognized with 78% accuracy, recognition rates for *panic fear*, *elation*, or *shame* were all below 40% although still larger than expected by chance (Banse & Scherer, 1996). Most studies of emotional prosody demonstrate recognition rates that are approximately four times higher than expected by chance (Pittam & Scherer, 1993), although not all emotional categories are recognized equally well. Expressions of *anger* and *sadness* are typically recognized more reliably from prosody than expressions of *fear* and *pleasant surprise* (Banse & Scherer, 1996; Johnstone & Scherer, 2000). The high recognition rates for these two emotions may be due to their distinctiveness at the acoustic level; a small literature on this topic indicates that *sadness* is usually expressed with low

* Corresponding author. Fax: +49 (0)341 9940 260.
E-mail address: paulmann@cbs.mpg.de (S. Paulmann).

[1] In order to stay coherent with the literature available we use the term emotion "recognition". However, we would like to point out that several authors (e.g., Banse & Scherer, 1996) have argued that most published recognition studies may have studied discrimination processes (choosing between a limited number of response alternatives) rather than true recognition processes. However, the likelihood of participants discriminating rather than recognizing an emotion is reduced with an increasing number of response alternatives that were applied in the current study.

0093-934X/$ - see front matter
doi:10.1016/j.bandl.2007.03.002

intensity, low speaking rate, and rather low pitch, while *anger* is conveyed with high intensity, fast speaking rate, and high pitch (Juslin & Laukka, 2003; Pell, 2001; Scherer, 1986; Williams & Stevens, 1972). The possibility that vocal expressions of these and perhaps other emotions are expressed in a 'universal' manner across languages is also being explored (e.g., Thompson & Balkwill, 2006).

Despite the growing research interest in emotional speech there are still several open questions. For example, it is under debate to what extent vocal emotion recognition differs between male and female listeners. Evidence for possible sex differences comes from studies on facial expression, where it has been suggested that women are more responsive to non-verbal cues than men (e.g., Briton & Hall, 1995). In addition, research on facial expression reveals perception differences between female and male participants which might extend to emotional speech processing; young women tend to rate male faces more positively than young men (e.g., van Strien & van Beek, 2000). To date, there is less direct evidence of sex differences in emotional speech perception. Some evidence suggests that women are better than men in identifying or discriminating emotional prosody (e.g., Bonebright, Thompson, & Leger, 1996; Schirmer, Kotz, & Friederici, 2002, 2005). Not all existing studies have reported this sex difference though (e.g., Fecteau, Armony, Joanette, & Belin, 2005; Orbelo, Grim, Talbott, & Ross, 2005; Raithel & Hielscher-Fastabend, 2004). Reasons for these diverse results are probably manifold, but varying task demands as well as the use of different methodologies and measurements across studies are likely candidates.

Further to possible sex effects in vocal emotion recognition, it has been noted that emotional speech recognition declines with age (e.g., Allen & Brosgole, 1993; Brosgole & Weismann, 1995; Kiss & Ennis, 2001; Orbelo et al., 2005). For example, Kiss and Ennis (2001) investigated the perception of emotional prosody in sentences with and without lexical content. Their results showed that younger participants significantly outperformed older participants irrespective of the emotional category tested (Kiss & Ennis, 2001). Moreover, Orbelo and colleagues reported an advantage for comprehending both emotional and attitudinal prosody in a group of young rather than older participants (Orbelo et al., 2005). Similarly, a recent study of *facial* emotion recognition (Calder et al., 2003) reported that older participants showed lower recognition rates for expressions of *fear* and *anger*, although not for *disgust*. It is often argued that age differences in emotional processing may not be domain specific, but rather domain general affecting, for example, attention or working memory (Filley & Cullum, 1994; Nicholas, Connor, Obler, & Albert, 1998). To reduce the potential effects of attention or working memory decline on emotional processing we thus opted to test middle-age participants in the current experiment. In comparison to previous studies in which the mean age of older participants was 75 years, our middle-aged-group comprised of participants with a mean age of 43 years.

In the context of this literature, the aim of the current study was to clarify whether emotional speech recognition differs as a function of sex and/or age. Male and female participants of two age cohorts (young and middle-aged) listened to sentences and then identified the meaning of the prosody in reference to six basic emotional categories (*anger*, *disgust*, *fear*, *happiness*, *pleasant surprise*, *sadness*) or a *neutral* category; the accuracy of each group was then compared. To build on previous work in the literature, we also investigated whether the recognition rates in different age and sex groups could be predicted by acoustic dimensions of the stimuli presented. Based on previous findings, we predicted that the ability to recognize emotional speech would decline with increasing age, showing lower accuracy rates for middle-aged as compared to young participants. The effect of sex on emotional speech recognition was less certain from the literature, although according to our design, an effect of sex on emotional speech recognition should lead to differences for both young and middle-aged participants irrespective of their age.

2. Materials and methods

2.1. Participants

Sixty-four native German speakers (age range: 18–50 years) participated in the experiment. Participants were divided into four equal age- and sex-controlled groups, resulting in two group dimensions. In the first dimension, the results of 16 young women and 16 middle-aged women were contrasted against the results of 16 young men and 16 middle-aged men (age range of both groups: 18–50 years). In the second dimension, recognition rates of emotional prosody of 16 young women and 16 young men (range 18–28 years) were compared to the performance of 16 middle-aged women and 16 middle-aged men (range: 38–50 years); (see Table 1 for mean age of participants). All participants had a similar educational background, normal or corrected-to-normal–vision, and no reported hearing impairment.

2.2. Stimulus material

The stimulus material consisted of 350 syntactically similar (SVO) sentences. The verb and noun of the sentences were controlled for word letter length, syllable length, word frequency, initial sounds, and plosive consonants. Prior to

Table 1
Participant information: the table displays the mean age in years (± standard deviation) for each sex and age group individually and grouped together

Participants	Female (years)	Male (years)	Mean (years)
Young	22.88 ± 1.89	24.0 ± 2.13	23.44 ± 2.06
Middle-aged	43.44 ± 2.50	41.81 ± 3.35	42.63 ± 3.02
Mean	33.16 ± 10.67	32.19 ± 9.46	

testing, sentences were grouped into one of the six basic emotional categories (*anger*, *disgust*, *fear*, *happy*, *pleasant surprise*, *sadness*) or into a semantically *neutral* category (50 sentences per category). Four German actors (two female, two male), two in each age cohort (young/middle-aged) were then asked to produce each sentence in the respective emotional prosody. This resulted in a total number of 1400 sentences.

Sentences were taped with a video camcorder (SONY Digital Video camera Recorder MiniDV DCR-TRV60E) attached to a high-quality clip-on microphone. The video material was digitized and the voice-track was separated from the visual-track. Within the current experiment, only the voice material was tested. The voice material was digitized at a 16-bit/44.1 kHz sampling rate and the amplitudes were normalized (with *CoolEdit* Version 2000).

2.3. Acoustic analysis

The stimulus material was acoustically analyzed using *Praat* (Boersma & Weenink, 2006). Primary acoustical measurements (mean pitch, intensity, duration) were entered in a series of one-way ANOVAs and results revealed significant differences across emotional categories for mean pitch ($F(6,343) = 744.18$, $p < 0.0001$), mean intensity ($F(6,343) = 1024.95$, $p < 0.0001$), and mean duration ($F(6,343) = 39.45$, $p < 0.0001$).

A discriminant analysis was performed to infer whether the stimuli contained detectable acoustic contrasts which might help listeners to correctly differentiate the intended emotion categories. In the analysis acoustic measurements served as independent variables whereas the dependent variable was the intended emotional category. The vast majority (83.3%) of the variance was accounted for the first function described by this discriminant analysis. Pooled-within-groups correlations between acoustics parameters and the first canonical discriminant function scores revealed that mean intensity drove the highest correlation ($r = .729$). Mean pitch had the largest pooled-within-groups correlation with the canonical discriminant function score ($r = .809$) in a second function that accounted for 15.5% of the variance. In a third function, that accounted for 1.2% of the variance, mean duration had the highest pooled-within-groups correlation with the canonical discriminant function score ($r = .989$). Fig. 1 illustrates how the canonical discriminant function scores for functions one and two separate the emotional categories for each sentence. As can be seen, the first two functions successfully separate sentences by emotional category. Classification results obtained from the discriminant analysis revealed that the model identified 87.1% of the sentences correctly. With the exception of *disgust*, the model correctly predicted category membership for all emotions above 80% correct (*anger*, 98%; *disgust*, 60%; *fear*, 82%; *happiness*, 100%; *neutral*, 90%; *pleasant surprise*, 100%; *sadness*, 80%).

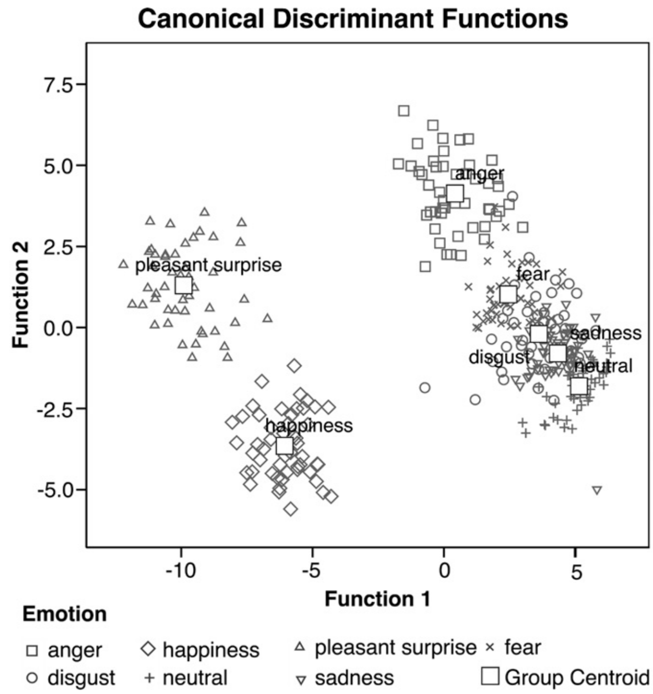

Fig. 1. Results of a discriminant feature analysis in which the intended emotional category was predicted by three acoustic parameters (mean pitch, intensity, duration). Each sentence is plotted according to its discrimination scores for the discriminant functions 1 (highest correlation with mean intensity) and 2 (highest correlation with mean pitch).

2.4. Procedure

Each participant was comfortably seated in a chair at a distance of approximately 60 cm from a computer screen with a seven-button panel placed before him/her. The sentences were presented via loudspeaker located 70 cm from the participant. The Event-related Run Time System (*ERTS*) (Beringer, 1993) was used to carry out the experimental task. Participants were instructed to listen to a sentence and to recognize the emotional prosodic category as quickly and accurately as possible. Secondly, participants were asked to judge the stimulus intensity as quickly and accurately as possible (results not reported here). The trial sequence was thus as follows: (1) presentation of a fixation cross for 200 ms; (2) clear screen for 100 ms; (3) acoustical presentation of a sentence with simultaneous presentation of a question mark on the screen requesting emotional prosody recognition; (4) a blank screen for 500 ms; (5) presentation of a number scale (++ + 0 − −−) requesting the intensity categorization; (6) inter-trial interval of 2000 ms. Response time was limited to 8000 ms to ensure spontaneous reactions.

To control the length of the experiment, trials were divided into four pseudo-randomized lists (350 sentences each). In each list, all 50 sentences of each emotional category were presented; however, lists differed with respect to which speaker had articulated the sentence. Each participant was presented with one of the four lists. The presentation of each speaker in each list was balanced as closely as possible

(12 vs. 13 sentences for each emotional category per speaker).

3. Results

3.1. Accuracy rates

3.1.1. Descriptive statistics

Overall, emotional prosody recognition was well above chance level (14%). As shown in Fig. 2, mean accuracy rates for *anger* and *neutral* vocalizations were highest, followed by *disgust*, *sadness*, and *happiness*. Expressions of *fear* and *pleasant surprise* were recognized the least accurately overall.

3.1.2. Analysis of variance (ANOVA)

Accuracy scores for the six emotion categories (+ *neutral*) were submitted to a $7 \times 2 \times 2$ ANOVA treating *Affect* (*angry*, *disgust*, *fear*, *happy*, *neutral*, *pleasant surprise*, *sadness*) as a repeated-measures factor and participant *Sex* (male/female) and *Age* (young/middle-aged) as between-subject factors. Effect size was estimated by omega-square (Ω^2), i.e., the coefficient of determination, which represents the proportion of variance in the dependent variable accounted for by the independent variable (interpreted in a similar manner as r^2). For between-subject designs, Ω^2 effect sizes greater than 0.138 are considered large effects, indices between 0.048 and 0.138 are considered medium effects, and values between 0.0009 and 0.048 are considered small effects (c.f. Olejnik & Algina, 2003).

No main effect of *Sex* was found ($p > .05$), but the *Age* effect was highly significant ($F(1,60) = 29.74$, $p < .0001$, $\Omega^2 = 0.31$). Emotion recognition rates were generally higher in younger participants than in middle-aged participants (73.22% vs. 61.82%). The main effect of *Affect* was significant ($F(6,360) = 82.67$, $p < .0001$, $\Omega^2 = 0.39$), although this was informed by a significant interaction of *Age* × *Affect* ($F(6,360) = 3.21$, $p < .05$, $\Omega^2 = 0.017$). Post hoc *t*-tests conducted for each emotion category revealed that young participants were significantly better at recognizing all emotion categories from prosody than middle-aged participants (all $p < .05$), except for *pleasant surprise* which did not differ as a function of age ($p > .05$) (see Fig. 2).

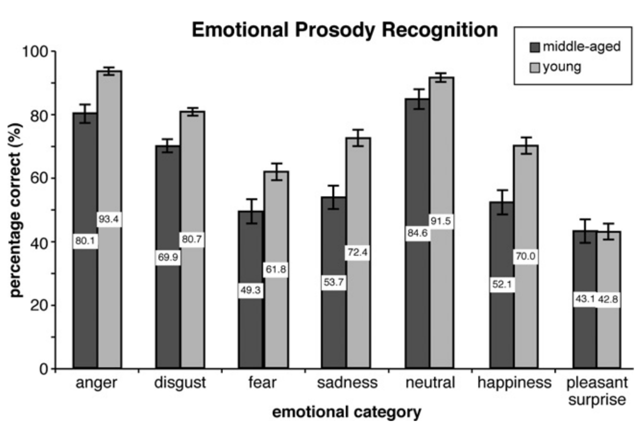

Fig. 2. Accuracy (in %) of emotional prosody recognition according to the age of participants. Bars show the correct responses for each intended emotional category and error bars represent the standard errors.

3.2. Error analysis

3.2.1. Chi-square tests for error matrices

Chi-square analyses for each error matrix indicated that participant errors were not equally distributed (all $p > .0001$). As can be seen in Table 2, listeners frequently confused *pleasant surprise* with *happy* sentences, and *fearful* with *sad* sentences. These patterns did not differ between the age and sex groups.

3.2.2. Distribution of false alarms

The distribution of incorrectly recognized emotional sentences split by age is illustrated in Fig. 3. Analyses revealed that middle-aged participants made more errors than young participants. Also, the figure shows that even though errors made by the two groups had a very similar distribution, middle-aged participants chose the categories *neutral* and *pleasant surprise* more often than young participants. Additional analyses revealed that middle-aged participants did not only make more misclassifications, but they had more time-outs (i.e., no hits during the given time-frame) than young participants (251 vs. 151).

3.2.3. Discriminant analysis by age group

To determine whether the overall *Age* effect was due to using acoustic cues differently, errors were entered into an additional discriminant analysis. Sentences were grouped according to their most frequent misclassification; those sentences that had equally frequent misclassifications were left out of the analyses (92 sentences for young, and 67 sentences for middle-aged participants). Results revealed that errors made by all participants could not be successfully predicted by the acoustic properties of the stimuli. However, the discriminant analysis for young participants showed higher prediction accuracy (32.2%) than the discriminant analysis for middle-aged participants (19.8%). For young participants, 96.6% of the variance was accounted for by the first function described by the discriminant analysis. Pooled-within correlations between the acoustics and the first canonical discriminant function scores revealed that mean intensity had the highest correlation ($r = .786$). Mean pitch had the largest pooled-within correlation with the canonical discriminant function score ($r = .994$) in the second function which accounted for 3.4% of the variance. For middle-aged participants, 84% of the variance was accounted for in the first function described by this discriminant analysis. Pooled-within correlations between the acoustics and the first canonical discriminant function scores revealed that mean duration had the highest correlation ($r = .835$). There were no significant correlations between acoustics and the second function described by this discriminant analysis which accounted for 16% of the variance.

Table 2
Error analysis: the matrix shows the confusion pattern for errors made in the emotional recognition task split by participant's age and sex

Group	Emotion	Intended emotion						
		Anger	Disgust	Fear	Happiness	Neutral	Pls. surp.	Sadness
Middle-aged	Anger	—	0.95	3.17	0.51	5.65	6.86	1.46
	Disgust	8.84	—	4.04	1.22	8.39	2.95	3.07
	Fear	9.34	4.60	—	0.33	11.37	9.47	13.02
	Happiness	0.51	0	0.51	—	19.91	25.36	0.13
	Neutral	0.88	0.32	0.38	7.88	—	4.60	0.57
	Pls. surp.	1.90	0.06	1.01	44.55	8.75	—	0.06
	Sadness	11.03	2.90	13.86	0.32	13.48	3.61	—
Young	Anger	—	0.44	1.26	0.63	2.07	1.13	0.50
	Disgust	6.92	—	3.24	0.95	3.17	1.08	2.86
	Fear	4.90	4.4	—	0.52	3.85	2.48	19.19
	Happiness	0.31	0	0.25	—	11.66	17.18	0.13
	Neutral	2.20	0.44	0.31	3.45	—	0.69	0.94
	Pls. surp.	2.27	0.06	0.63	50.94	2.71	—	0.13
	Sadness	6.41	3.55	8.69	0.30	7.11	0.63	—
Female	Anger	—	0.63	2.53	0.32	2.84	3.10	0.88
	Disgust	8.14	—	3.31	0.83	4.32	1.53	3.31
	Fear	7.73	4.22	—	0.26	5.13	5.33	18.65
	Happiness	0.51	0	0.44	—	16.55	19.39	01.3
	Neutral	1.39	0.38	0.25	5.23	—	1.64	1.07
	Pls. surp.	2.28	0.06	0.82	46.11	525	—	0.13
	Sadness	7.88	3.27	10.63	0.19	9.16	2.05	—
Male	Anger	—	0.76	1.89	0.82	4.86	4.86	1.07
	Disgust	7.61	—	3.97	1.34	7.23	2.50	2.62
	Fear	6.47	4.82	—	0.59	10.11	6.61	13.54
	Happiness	0.31	0	0.31	—	14.98	23.10	0.13
	Neutral	1.70	0.38	0.44	6.09	—	3.64	0.44
	Pls. surp.	1.89	0.06	0.82	49.40	6.18	—	0.06
	Sadness	9.51	3.19	11.88	0.51	11.37	2.17	—

Values are the mean incorrect answers (in %).

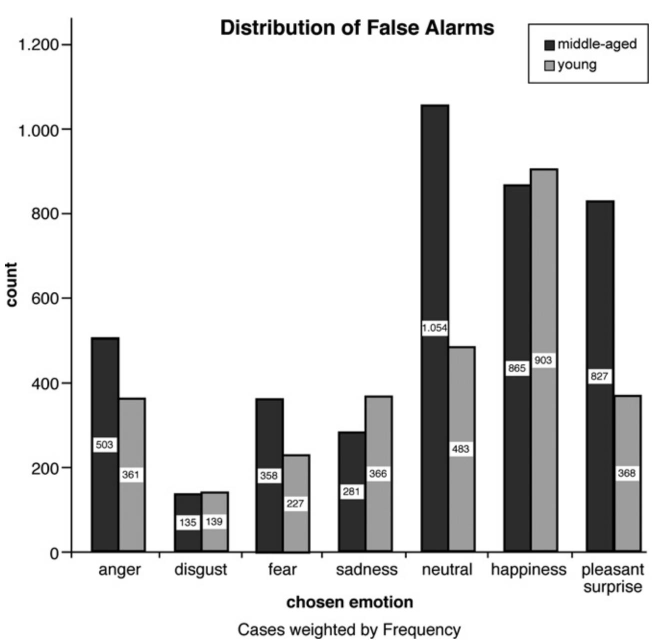

Fig. 3. The distribution of emotional error classifications across trials, split by age. Overall, 11,200 responses could be given by participants (50 sentences per emotional category × seven emotional categories × number of participants). The figure displays all errors for trials excluding time-outs.

4. Discussion

The present study investigated how the combined effects of age and sex influence emotional speech recognition. Taken together, our results confirmed that recognition accuracy rates vary significantly as a function of emotion category when listening to emotional speech. Averaging across emotions, an overall accuracy rate of 70% was obtained which is approximately five times higher than expected by chance in our task and in line with previous research (e.g., Scherer, Johnstone, & Klasmeyer, 2003). Interestingly, our findings further revealed an overall recognition advantage for young participants over middle-aged participants on this task with the exception of one emotion category (*pleasant surprise*). For example, we found that 37% of the sentences were classified incorrectly by middle-aged participants in contrast to only 26% classification errors by young participants. In contrast, no effects of sex on emotional speech recognition emerged from our data. These results will be discussed in sequence in the following paragraphs.

4.1. Emotional speech recognition

Prior to the study, acoustic analyses revealed that there were obvious differences in pitch, amplitude, and temporal attributes of our emotional utterances that might help listeners to correctly classify the intended emotional category. As well, a discriminant feature analysis showed that category membership for all emotional tokens was correctly predicted by the acoustic measures in excess of 80% with the exception of *disgust* (which still achieved 60% correct), suggesting that the perceptual stimuli were distinct at the acoustic level. Nonetheless, even though listeners could successfully decode basic emotions from speech samples at a high level, our findings reaffirm that some emotions are recognized much better than others (e.g., *anger* was recognized better than *pleasant surprise*; review Fig. 2).

Studies that demonstrate how well particular emotions are recognized from speech show rather mixed findings. Some of these inconsistencies could be due to different presentation modalities (e.g., with/without lexical context, recognition at word/sentence level). Some emotions may be more easily recognized if lexical information is available, and there is behavioural and electrophysiological evidence that semantic information cannot be ignored even if it is not in the focus of a task (Besson, Magne, & Schön, 2002; Pell, Kotz, Paulmann, & Alasseri, 2005). Banse and Scherer (1996) argue that some emotions are recognized more poorly in sentences because of the limited ecological validity of certain expressions in a sentential context. Differences in ranking emotional categories could also be due to biological factors which can influence emotion recognition (e.g., detecting *fear* is necessary to survive, or recognizing *anger* is necessary to perceive potential danger; e.g., Öhman & Mineka, 2001; Öhman, 2002; Williams & Mattingley, 2006).

Tasks that include emotion categories which tend to yield high confusion rates are likely to influence the relative ranking of emotions in recognition studies. The current results illustrate that recognition errors are not randomly distributed; for example, *pleasant surprise* was frequently mistaken for *happiness*, and *fear* was frequently mistaken for *sadness*. It has been argued that emotions that are acoustically similar are often misclassified (c.f., Banse & Scherer, 1996). This may be particularly the case when the semantic context is sufficiently ambiguous as to the intended emotional interpretation (e.g., confusion between *pleasant surprise* and *happiness*); this may explain some of the errors witnessed in the current study. Some researchers also argue that misclassifications frequently involve emotions of similar valence (e.g., *happiness* and *pleasant surprise*) and arousal (i.e., *anger* and *sadness* are less often confused than *anger* and *fear*; c.f., Scherer, 1986). Thus, it is likely that emotional category recognition varies between studies due to multiple reasons, such as the presentation modality, the ecological validity of stimulus materials, the different acoustic realization of stimuli, and possibly biological factors.

4.2. Influence of sex on emotional speech recognition

Our investigation failed to uncover significant evidence of sex-specific effects on emotional speech recognition. In the introduction, we put forward several reasons that could explain why some research has reported sex differences in emotional speech recognition (e.g., Briton & Hall, 1995; Bonebright et al., 1996; Grunwald et al., 1999; van Strien & van Beek, 2000), while others have not (e.g., Fecteau et al., 2005; Orbelo et al., 2005; Raithel & Hielscher-Fastabend, 2004). Based on the broader literature, one could speculate that strong sex differences are only detectable in young participants, and then only under certain task-related conditions; electrophysiological data gathered by Schirmer and colleagues suggest that women are primarily faster to process emotional prosody than men, but they do not necessarily engage in different processing strategies (e.g., Schirmer et al.,2002,2005). In addition, sex differences are not found if young participants are instructed to take emotional prosody into account when rendering perceptual judgments (Schirmer, Kotz, & Friederici, 2005). Our results are thus in line with the assumption that sex differences only occur under implicit emotional prosody processing situations. Another possibility is that sex differences are highly pronounced at the stage of *expressing* rather than recognizing emotional speech; in a recent study of emotional communication in three languages (English, German, Arabic), we noted that the sex of the speaker had a major impact on how vocal emotions were expressed, but that these differences were recognized similarly by male and female listeners of each language (Pell et al., 2005).

4.3. The influence of age on emotional speech recognition

The present results reveal a clear decline in emotional speech recognition with increasing age. This age effect is in line with general findings on emotion perception (e.g., Brosgole & Weismann, 1995; McDowell, Harrison, & Demaree, 1994; Prodan, Orbelo, & Ross, 2007) and with particular findings on the processing of emotional prosody (e.g., Allen & Brosgole, 1993; Kiss & Ennis, 2001; Orbelo, Testa, & Ross, 2003; Orbelo et al., 2005). Based on physiological evidence, women, but also men, undergo hormonal changes around 40 to 55 years of age that can affect cognitive processes (e.g., Everhardt, Demaree, & Shipley, 2006 for a recent review). A decline in attention (Filley & Cullum, 1994) and working memory (Nicholas et al., 1998) have each been associated with increasing age. For instance, Filley and Cullum (1994) evaluated the attention of participants who were 50–69 and 70–90 years of age (Filley & Cullum, 1994) and found that sustained attention starts to decline after the age of 70, while basic attention was not affected at all. Grunwald et al. (1999) found that lexical emotion perception was associated with an age-related decline in participants 60 years or older, but not in younger participants. Given that our middle-aged group was younger than most aging groups studied previously, it is unlikely that emotion recognition abilities were confounded by a reduction in other cognitive capacities, although this confound may be present in the wider literature. Rather, the effects of aging on our data are likely to be domain specific rather than domain general, i.e., due to true changes in the recognition of emotional prosody as a function of aging that appear to demonstrate a notably ‘‘early’’ decline in middle-aged adults in the auditory modality.

Age-related differences in emotion recognition can be accounted for in several ways. First, it has been said that differences in various cognitive domains between young and old participants may be due to specific age-related neuroanatomic changes (e.g., Filley & Cullum, 1994) and this may also be true for recognizing emotional speech. Some research suggests that emotion perception engages different neural networks in young and older adults (see Fecteau et al., 2005). For instance, Gunning-Dixon and colleagues (2003) investigated the influence of age on emotional face processing using a discrimination task; their results imply that young participants rely on a different cortical network to discriminate emotional faces (involving visual, frontal and limbic regions) than older participants (involving parietal, temporal and frontal regions). Along similar lines, some authors have argued that emotional prosody comprehension declines in older participants due to asymmetric, age-related changes in the right hemisphere (e.g., Orbelo et al., 2005).

Alternatively, age-related differences in emotional speech recognition could result from acoustic cues being used differently in young and older adults, a possibility that is implied by our data. Results of a discriminant analysis which attempted to predict group error patterns from acoustic properties of the stimuli showed that emotional misclassifications were better accounted for in the analysis for young participants when compared to the middle-aged participants (32.2% vs. 19.8%). Thus, it appears that the two age groups may have been using the acoustic cues of the stimuli differently or perhaps using different acoustic cues altogether, for example those that were not captured by the current measures. The fact that middle-aged participants responded with a greater frequency of delays in the recognition task (as evidenced by an increased number time-outs) may constitute further evidence that the middle-aged group experienced trouble using the acoustic input to categorize specific emotions.

In conclusion, our results establish that vocal expressions of the basic emotions can be recognized at levels that well exceed chance by both female and male listeners who are young or middle-aged. Although we found no evidence of sex effects in emotional speech recognition, the age of participants influenced emotion recognition from speech samples in a significant manner. In fact, our data clearly show that emotional speech recognition begins to decline already in participants who are middle-aged, potentially for several reasons (e.g., neurophysiological changes, differential use of acoustic cues, etc.). Future studies are needed to investigate these issues more closely by clarifying how emotional speech recognition declines in the latter half of the lifespan, preferably using a longitudinal design. Neuro-cognitive studies which specify changes in the underlying brain mechanisms that lead to a decline in emotional prosody recognition will also be valuable.

Acknowledgments

The authors thank Beate Günther for help with participant recruitment, Katharina Flach for help with data acquisition, Marie Desmarteau for help with the acoustic analyses, and Dorothee Geyer for useful comments on statistical analyses. This work was supported by the German Research Foundation (DFG FOR 499 to S.A. Kotz) and by the Natural Sciences and Engineering Research Council of Canada (to M.D. Pell).

References

Allen, R., & Brosgole, L. (1993). Facial and auditory affect recognition in senile geriatrics, the normal elderly and young adults. *International Journal of Neuroscience, 68*, 33–42.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 3*, 614–636.

Beringer, J. (1993). Experimental Run Time System (computer software). BeriSoft Corporation, Frankfurt/ Main, Germany.

Besson, M., Magne, C., & Schön, D. (2002). Emotional prosody: Sex differences in sensitivity to speech melody. *Trends in Cognitive Science, 6*, 405–407.

Boersma, P., & Weenink, D. (2006): Praat: Doing phonetics by computer (Version 4.1.13) [Computer program]. Retrieved from http://www.praat.org/, July 15, 2003.

Bonebright, T. L., Thompson, J. L., & Leger, D. W. (1996). Gender stereotypes in the expression and perception of vocal affect. *Sex Roles, 34*(5/6), 429–445.

Briton, N., & Hall, J. (1995). Beliefs about female and male nonverbal communication. *Sex Roles, 32*, 79–90.

Brosgole, L., & Weismann, J. (1995). Mood recognition across the ages. *International Journal of Neuroscience, 82*(3/4), 169–189.

Calder, A. J., Keane, J., Manly, T., Sprengelmeyer, R., Scott, S., Nimmo-Smith, I., et al. (2003). Facial expression recognition across the adult life span. *Neuropsychologia, 41*, 195–202.

Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion. In J. Cole (Ed.). *Nebraska Symposium on Motivation 1971* (Vol. 19, pp. 207–283). Lincoln, NE: University of Nebraska Press.

Ekman, P. (1982). *Emotion in the human face*. New York: Cambridge University Press.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6*, 169–200.

Everhardt, D. E., Demaree, H. A., & Shipley, A. J. (2006). Perception of Emotional Prosody: Moving toward a model that incorporates sex-related differences. *Behavioural and Cognitive Neuroscience Reviews, 5*(2), 92–102.

Fecteau, S., Armony, J. L., Joanette, Y., & Belin, P. (2005). Judgment of emotional nonlinguistic vocalizations: Age-related differences. *Applied Neuropsychology, 12*(1), 40–48.

Filley, C. M., & Cullum, C. M. (1994). Attention and vigilance functions in normal aging. *Applied Neuropsychology, 1*, 29–32.

Grunwald, I. S., Borod, J. C., Obler, L. K., Erhan, H. M., Pick, L. H., Welkowitz, J., et al. (1999). The Effects of Age and Gender on the Perception of Lexical Emotion. *Applied Neuropyschology, 6*(4), 226–238.

Gunning-Dixon, F. M., Gur, R. C., Perkins, A. C., Schroeder, L., Turner, T., Turetsky, B. I., et al. (2003). Age-related differences in brain activation during emotional face processing. *Neurobiology of Aging, 24*(2), 285–295.

Johnstone, T., & Scherer, K. R. (2000). Vocal communication of emotion. In Michael Lewis & M. Jeannette Haviland-Jones (Eds.), *Handbook of Emotion* (2nd ed., pp. 220–235). New York: Guilford press.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129*(5), 770–814.

Kiss, I., & Ennis, T. (2001). Age-related decline in perception of prosodic affect. *Applied Neuropsychology, 8*(4), 251–254.

McDowell, C. L., Harrison, D. W., & Demaree, H. A. (1994). Is right hemisphere decline in the perception of emotion a function of aging? *International Journal of Neuroscience, 79*, 1–11.

Nicholas, M., Connor, L., Obler, L. K., & Albert, M. L. (1998). Aging, language, and language disorders. In M. T. Sarno (Ed.), *Acquired aphasia* (pp. 413–450). San Diego, CA: Academic.

Öhman, A. (2002). Automaticity and the amygdala: Nonconscious responses to emotional faces. *Current Directions in Psychological Science, 11*, 62–66.

Öhman, A., & Mineka, S. (2001). Fear, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review, 108*(3), 483–522.

Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods, 8*, 434–447.

Orbelo, D. M., Grim, M. A., Talbott, R. E., & Ross, E. D. (2005). Impaired comprehension of affective prosody in elderly subjects is not predicted by age-related hearing loss or age-related cognitive decline. *Journal of Geriatric Psychiatry and Neurology, 18*, 25–32.

Orbelo, D. M., Testa, J. A., & Ross, E. D. (2003). Age-related impairments in comprehending affective prosody with comparison to brain-damaged subjects. *Journal of Geriatric Psychiatry and Neurology, 16*, 44–52.

Pell, M. D. (2001). Influence of emotion and focus location prosody in matched statements and questions. *Journal of the Acoustical Society of America, 109*(4), 1668–1680.

Pell, M. D., Kotz, S., Paulmann, S., & Alasseri, A. (2005). Recognition of basic emotions from speech prosody as a function of language and sex. Psychonomics, Toronto, Ontario, Abstracts of the Psychonomic Society 46th Annual Meeting, 10, pp. 97–98.

Pittam, J., & Scherer, K. R. (1993). Vocal expression and communication of emotion. In M. Lewis & J. M. Havliand (Eds.), *Handbook of emotions* (pp. 185–198). New York: Guilford Press.

Prodan, C. I., Orbelo, D. M., & Ross, E. D. (2007). Processing of facial blends of emotion: Support for right hemisphere cognitive aging. *Cortex, 2*, 196–206.

Raithel, V., & Hielscher-Fastabend, M. (2004). Emotional and linguistic perception of prosody. *Folia Phoniatrica et Logopaedica, 56*, 7–13.

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin, 99*, 143–165.

Scherer, K. R., Johnstone, T., & Klasmeyer, G. (2003). Vocal expression of emotion. In R. J. Davidson, H. Goldsmith, & K. R. Scherer (Eds.), *Handbook of affective sciences*. New York and Oxford: Oxford University Press.

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.

Thompson, W. F., & Balkwill, L. L. (2006). Decoding speech prosody in five languages. *Semiotica, 158*(1–4), 407–424.

van Strien, J. W., & van Beek, S. (2000). Ratings of emotion in laterally presented faces: Sex and handedness effects. *Brain and Cognition, 44*(3), 645–652.

Williams, M. A., & Mattingley, J. B. (2006). Do angry men get noticed? *Current Biology, 16*, R402–R404.

Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: Some acoustic correlates. *The Journal of the Acoustic Society of America, 52*(4), 1238–1250.