Available online at www.sciencedirect.com
ScienceDirect
Speech Communication 54 (2012) 1–10
SPEECH COMMUNICATION
www.elsevier.com/locate/specom

# Categorical processing of negative emotions from speech prosody

Abhishek Jaywant, Marc D. Pell *

*McGill University, School of Communication Sciences and Disorders, Montréal, Québec, Canada*

Received 17 December 2010; received in revised form 24 May 2011; accepted 25 May 2011
Available online 1 June 2011

**Abstract**

Everyday communication involves processing nonverbal emotional cues from auditory and visual stimuli. To characterize whether emotional meanings are processed with category-specificity from speech prosody and facial expressions, we employed a cross-modal priming task (the Facial Affect Decision Task; Pell, 2005a) using emotional stimuli with the same valence but that differed by emotion category. After listening to angry, sad, disgusted, or neutral vocal primes, subjects rendered a facial affect decision about an emotionally congruent or incongruent face target. Our results revealed that participants made fewer errors when judging face targets that conveyed the same emotion as the vocal prime, and responded significantly faster for most emotions (anger and sadness). Surprisingly, participants responded slower when the prime and target both conveyed disgust, perhaps due to attention biases for disgust-related stimuli. Our findings suggest that vocal emotional expressions with similar valence are processed with category specificity, and that discrete emotion knowledge implicitly affects the processing of emotional faces between sensory modalities.

*Keywords:* Prosody; Facial expressions; Emotion; Nonverbal cues; Priming; Category-specific processing

## 1. Introduction

In everyday human communication, emotions are conveyed through both vocal utterances and facial expressions. In these two modalities, both faces and speech prosody (i.e., the “melody” of language) are critical carriers of nonverbal emotional information. It is a ubiquitous finding in the scientific literature that specific configurations of facial features communicate emotional meanings that are reliably identified by observers (Borod et al., 2000; Ekman, 1992; Palermo and Coltheart, 2004; Tracy and Robins, 2008; Williams et al., 2009; Young et al., 1997). Similarly, acoustic markers of vocal prosody such as fundamental frequency, intensity, and speech rate also serve to convey emotional information (Banse and Scherer, 1996). Listeners are able to accurately recognize emotions from prosody even when speech cues lack relevant lexical-semantic content or are presented as a foreign language (Pell et al., 2009a,b; Pell and Skorup, 2008; Scherer et al., 1991).

Of particular relevance to the current investigation, nonverbal emotional information from the two modalities frequently interacts: emotional cues from the voice have the ability to influence processing of a conjoined face and vice-versa. For example, numerous studies on emotion recognition have shown that facial and/or vocal emotional cues that are not consciously attended to, still influence the processing of facial and/or vocal stimuli that are attended (de Gelder et al., 1999; de Gelder and Vroomen, 2000; Hietanen et al., 2004; Massaro and Egan, 1996; Pourtois et al., 2000; Vroomen et al., 2001). Evidence from event-related potentials (ERPs) suggests the integration of the two modalities occurs rapidly and may be mandatory (de Gelder et al., 1999; Pourtois et al., 2000). Another important source of evidence for cross-modal emotional processing comes from studies on priming. Recent research from Pell and colleagues has demonstrated that when primed with emotional information through speech prosody, subjects respond faster and more accurately to facial

* Corresponding author. Address: McGill University, School of Communication Sciences and Disorders, 1266 Avenue des Pins Ouest, Montréal, Québec, Canada H3G 1A8. Tel.: +1 514 398 4133; fax: +1 514 398 8123.
*E-mail address:* marc.pell@mcgill.ca (M.D. Pell).
*URL:* http://www.mcgill.ca/pell_lab (M.D. Pell).

0167-6393/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.specom.2011.05.011

displays that are emotionally congruent with the vocal prime (Pell,2005a,b; Pellet al., in press). Carroll and Young (2005) have also found evidence of cross-modal emotional priming using nonverbal sounds and facial expressions. These findings lend credence to the idea that aspects of emotional prosody share underlying features with facial expressions in associative memory, and that the mechanisms for processing emotional meanings from the two modalities overlap in their cognitive processing structure (Borod et al., 2000; Bowers et al., 1993) and the neural regions involved (Kreifelts et al., 2009).

Researchers have focused on different ways in which emotional knowledge is represented, processed, and how it interacts when emotional events are encountered in more than one information channel. For example, dimensional accounts of emotional processing (Gerber et al., 2008; Posner et al., 2005) have posited that affective information can be represented along underlying continua such as valence (positive–negative) and arousal (high–low). Several studies have shown that subjects are faster to respond to a target stimulus that shares the same valence as the prime, yielding affective priming (see Fazio (2001) for a review) (Hermans et al., 1994; Spruyt et al., 2007; Zhang et al., 2006). As summarized by Fazio (2001), the presentation of a word with a negative (positive) valence unconsciously and automatically facilitates rapid response times to target words with a congruent negative (positive) valence and this effect has been shown to occur both within and across sensory modalities using various priming tasks such as lexical decisions and explicit valence judgments.

In line with the idea of ‘basic’ emotions (Ekman, 1992), other researchers have investigated the influence of discrete emotion meanings on the perception and recognition of emotional displays, particularly for facial expressions (Ekman, 1992; Young et al., 1997; Batty and Taylor, 2003; Bimler and Kirkland, 2001; Etcoff and Magee, 1992; Krolak-Salmon et al., 2001; Levenson et al., 1990). For example, Young et al. (1997) demonstrated that continua of blended emotional faces are perceived as distinct prototypical emotional categories and that facial emotion discrimination was better across categories than within categories. Moreover, early visual processing of emotional face categories is associated with unique electrophysiological activity as measured by ERPs (Batty and Taylor, 2003). In the context of speech, Laukka (2005) constructed blended emotional utterances which varied along an acoustic continuum (e.g., fear–happy) and found that in a recognition task, the continua were perceived as distinct emotion categories with a sudden boundary. Moreover, subjects were better at discriminating between stimuli belonging to different emotion categories (compared to stimuli within a category) even though the physical differences between these stimuli were identical. This ability to more reliably discriminate stimuli across different categories compared to stimuli within a category while controlling for physical differences has been suggested to be a defining feature of categorical perception (Goldstone and Hendrickson, 2010). The findings of Laukka (2005) provide support for the presence of discrete emotion categories in the auditory modality in addition to facial expressions.

Further evidence that specific emotion categories can be communicated and recognized through speech, particularly prosody, comes from a recent study by Pell et al. (2009b). The authors looked at participants’ ability to recognize basic emotion categories through speech prosody in English, German, Hindi, and Arabic. When listening to utterances in their native language, participants could accurately discriminate between emotion categories and in particular, anger, sadness, and fear were well recognized. Importantly, acoustic analyses were also conducted and the results showed that specific acoustic markers of prosody (mean fundamental frequency, range of fundamental frequency, and speech rate) differentiated the emotion categories. For example, disgust was expressed with a low mean fundamental frequency and speech rate, anger with a wide range of fundamental frequency, and sadness with a low mean and range of fundamental frequency and low speech rate. Discriminant function analyses also suggested that emotion categories could be classified by their underlying acoustic properties. These results lend further support to the idea that discrete emotions can be communicated through speech, differentiated acoustically (in particular through pitch, speaking rate, and their interaction), and that the discrete properties of vocal emotion expressions play an important role in how prosody is recognized and how these meaning activations prime other emotional stimuli (Banse and Scherer, 1996; Pell, 2005a,b; Juslin and Laukka, 2003; Schirmer et al., 2005; Niedenthal and Setterlund, 1994).

Still, there is a dearth of information on how emotional speech prosody influences the processing of communicative displays in other modalities, such as the face, and whether these cross-modal effects occur with emotion-specificity. Since the vocal and facial channels are crucial in conveying affect, a better understanding of how emotions are perceived from both types of cues is important to advance our knowledge of how emotional knowledge is communicated in social and interpersonal contexts. Many studies on the relationship between vocal and facial emotional processing have compared only one positive and one negative valence emotion (often happy vs. sad), rendering it difficult to ascertain whether observed cross-modal influences are due to category-specific processing or more general valence-based effects (de Gelder and Vroomen, 2000; Vroomen et al., 2001; Pell, 2005b; Pell et al., in press; Schirmer et al., 2005) (see Pell (2005a) for data using additional emotion categories). Given the evidence for general valence-based effects that can occur across modalities (Fazio, 2001), an experimental design using emotions of the same valence that differ by category will allow us to infer whether voice–face interactions can occur with category specificity, beyond affective priming. In addition, certain emotion categories are seldom studied and may differ in their cross-modal interactions. For example, recent research has suggested that the emotion of disgust is unique in how it is processed acoustically (Pell et al., 2009b; Scherer et al., 1991; Pell and Kotz, in review), is particularly difficult to recognize from speech prosody (poorer accuracy, Pell et al., 2009b; Scherer et al., 1991), and may lead to attention biases (Cisler et al., 2009). New research is therefore needed to infer how specific emotion-based knowledge is activated and communicated through speech, especially when emotional information processing involves conjoined processing of socially relevant auditory and visual displays which begin to approximate natural language contexts.

In the present study, we used the Facial Affect Decision Task (FADT) – a cross-modal priming task – to investigate whether nonverbal emotional information from speech prosody primes decisions about related face targets in a category-specific manner. In the FADT, subjects hear a prime sentence and subsequently view a facial expression that either conveys an emotion (e.g., sadness) or does not convey an emotion (i.e., a "grimace" that involves facial movements that do not communicate an emotion; (Paulmann and Pell, 2009). Similar to the lexical decision task, subjects make a yes/no decision about each face target (*Does the facial expression represent an emotion?*). Emotional knowledge presumably shared across auditory and visual modalities (i.e., in "yes" targets) is activated implicitly, without drawing conscious attention to verbal labels associated with each category (Carroll and Young, 2005). The FADT has been used reliably in the past to index cross-modal priming effects related to happiness, sadness, anger, surprise, and fear (Pell, 2005a,b; Pell et al., in press; Paulmann and Pell, 2010). Here, we only presented emotion categories with a negative valence (sadness, disgust, and anger) in order to achieve a finer test of whether priming effects of prosody on a face are category-specific. While past studies have frequently used emotion categories with both positive and negative valences that render it difficult to disentangle the contribution of affective (valence-based) and category-specific interactions, the present design allowed us to test whether priming effects occurred uniquely for each emotion category, regardless of valence. We hypothesized that subjects would respond faster and/or more accurately to the presence of an emotional face when it was categorically congruent with the emotion communicated by the vocal prime. Although we expected a cross-modal congruency effect for all emotion categories of interest, a lack of previous data on disgust and reported biases for this emotion (Pell and Kotz, in review; Cisler et al., 2009) did not allow for exact predictions about this emotion and are thus somewhat exploratory.

## 2. Methods

### 2.1. Participants

Fifty native English speaking adults (25 male/25 female) were recruited from McGill University to participate in the study. Participants had a mean age of 21.5 years ($SD = 3.6$) and 15.0 mean years of education ($SD = 1.7$). All participants had self-reported normal hearing and normal or corrected-to-normal vision. Research procedures were ethically approved by a McGill University institutional review board following the Declaration of Helsinki.

### 2.2. Materials

The stimuli consisted of vocal prime utterances and static emotional face targets. For a comprehensive description of the construction and assumptions of the FADT see Pell (2005a). Stimulus features are described briefly here but additional details may be found elsewhere (Pell, 2002).

#### 2.2.1. Prime utterances

The vocal primes were short spoken sentences (5–10 syllables in length) constructed by taking a set of English sentences and replacing the content words with phonologically appropriate but semantically meaningless sounds arranged using valid grammatical/syntactic structure to ensure the perception of speech that sounded "language-like" (e.g., *I nestered the flugs*). These "pseudo-utterances" thus lacked meaningful emotional semantic content but effectively conveyed the emotion of interest through prosodic features of speech. Pseudo-sentences have been frequently used in the past to isolate emotional information to the vocal channel of speech, both in cross-modal priming tasks such as the FADT (Pell, 2005a,b; Pell et al., in press) and in other studies of vocal emotion recognition (Pell et al., 2009a; Scherer et al., 1991).

Ten native English speakers produced the pseudo-utterances to convey emotions of *anger, disgust,* and *sadness* as well as *neutrality*, which were recorded onto digital audiotape and subsequently transferred to a PC in .wav format. All prime utterances were first piloted in a separate, larger stimulus validation study to ensure the presence of the appropriate emotion (Pell, 2002). In that study, 24 native English speaking participants listened to and rated each stimulus in a forced-choice emotion recognition format. From this pilot study, vocal stimuli conveying *anger, disgust, sadness,* and *neutrality* ($n = 10$ each) were selected for the current experiment. These 40 pseudo-utterance primes were recognized correctly at a very high mean level of 86% (range = 80% for disgust versus 97% for sadness), where chance performance was 14%. Since speaking rate differences are a natural cue for differentiating emotions in the voice (Pell et al., 2009b), the selected prime utterances varied in average duration according to the emotion category (mean prime duration for: *anger* = 1618 ms; *disgust* = 2144 ms; *sadness* = 1789 ms; *neutrality* = 1739 ms). To gather data on valence and arousal characteristics of our primes, on a separate occasion the 40 selected stimuli were randomized with 20 pseudo-utterances produced by the same actors that conveyed joy or pleasant surprise, and each item was independently rated on a five-point scale to determine its perceived valence (−2 to +2) and speaker arousal (1–5). Valence and arousal features were assigned by 12 young raters who judged each affective dimension in a separate task.

The major perceptual and acoustic features of vocal expressions presented in this study are presented in Table 1 by emotion type. A one-factor MANOVA examined differences among the vocal emotion expressions (anger, disgust, sadness, neutral) in reference to the seven acoustic measures reported in Table 1: fundamental frequency (F0, Mean, Range), Amplitude (Mean, Range), Harmonics-to-noise Ratio (HNR, Mean, *SD*), and Speech Rate. The MANOVA was significant according to Wilks' $\Lambda(0.19)$, $F(21, 87) = 3.23$, $p < .00001$. Follow-up, univariate analyses demonstrated a significant effect of emotion in our prime stimuli for: F0Mean, $F(3, 36) = 4.40$, $p = .01$; HNR variation, $F(3, 36) = 3.35$, $p = .02$; and speech rate, $F(3, 36) = 11.42$, $p = .00002$. Tukey's HSD post hoc comparisons indicated that angry prime utterances were spoken with a significantly higher mean F0, and greater variability in HNR, than both disgust and neutral utterances (sad utterances did not differ from the other emotions for either measure). For speech rate, anger and neutral utterances were spoken more quickly than disgust, and neutral utterances were also significantly faster in rate than sad prime utterances.

### 2.2.2. Target faces

The target facial expressions were static 17.1 cm × 17.1 cm color photographs of eight encoders that included the face, hair, and shoulders. From a large set of exemplars, face targets conveying *anger*, *disgust,* and *sadness* ($n = 10$ each) were chosen for the present study. Like vocal primes, the selected facial expressions were validated as part of a larger corpus (Pell, 2002) and were recognized correctly at a mean rate of 96% by 32 raters using a forced-choice emotion recognition paradigm. Furthermore, 30 facial "grimaces" which do not express an emotion were selected for use in the "NO" trials of the FADT. All grimaces were not recognized as conveying basic emotions by a minimum of 60% of raters in the pilot study, who frequently described these expressions as "silly." While facial grimaces likely possess certain affective qualities, behavioral data (Pell, 2005a,b) as well as electrophysiological responses to these expressions (Paulmann and Pell, 2009) establish that grimaces are not recognized as conveying discrete emotional meanings and thus allow participants to successfully render facial affect decisions, similar to lexical decisions about visually-presented pseudo-words.

## 2.3. Experimental task design

The pseudo-utterance primes and face targets were paired to create 240 experimental trials: 120 YES trials ending in a "real" emotional face target and 120 NO trials ending in a facial grimace. Examples of facial expressions presented in YES versus NO trials are shown in Fig. 1. The experiment was balanced to contain an equal number of male and female primes and targets, and primes were always paired with targets of the same gender. For YES trials, the prime–target relationship was defined in three ways: (1) 30 congruent trials consisting of the same emotion being conveyed in the vocal prime and face target (i.e., 10 trials each of anger–anger, sadness–sadness*,* and disgust–disgust); (2) 60 incongruent trials where each face target was paired with two instances of a prime conveying an unrelated emotion (e.g., for angry face targets, incongruent trials contained prime–target relationships of sadness–anger and disgust–anger); (3) 30 neutral trials consisting of pseudo-utterance primes spoken with neutral intonation and paired with emotional face targets (i.e., 10 trials each of neutral–anger, neutral–sadness*,* and neutral–disgust). For NO trials, all 40 pseudo-utterance primes were paired with all 30 facial grimaces to create 120 trials. The 240 YES/NO trials were presented in six blocks of 40 trials. Prime–target pairs were pseudo-randomized within blocks but balanced to contain relatively equal numbers of YES and NO trials, male and female stimuli, and

Table 1
Major perceptual and physical parameters of the emotional stimuli presented in the experiment.

| Stimulus type/parameter | Emotion | | | |
|---|---|---|---|---|
| | Anger | Disgust | Sadness | Neutral |
| *Vocal primes (prosody)* | | | | |
| Recognition – (%) correct | 86 | 80 | 97 | 82 |
| Perceived valence (scale of −2 to +2) | −0.8 | −0.7 | −1.4 | 0.1 |
| Perceived intensity (scale of 1–5) | 3.4 | 2.7 | 2.7 | 1.8 |
| Fundamental frequency – mean (Hz) | 218 | 155 | 208 | 172 |
| Fundamental frequency – range (Max–Min, Hz) | 172 | 122 | 135 | 106 |
| Amplitude – mean (dB) | 61 | 57 | 55 | 57 |
| Amplitude – range (Max–Min, dB) | 38 | 38 | 35 | 34 |
| Harmonics-to-noise ratio – mean (dB) | 19.2 | 19.7 | 20.4 | 17.7 |
| Harmonics-to-noise ratio – variation (*SD*, dB) | 5.1 | 3.1 | 4.3 | 3.2 |
| Speaking rate (syllables/s) | 4.2 | 2.8 | 3.7 | 4.9 |
| *Face targets* | | | | |
| Recognition (%) correct | 96 | 97 | 95 | – |

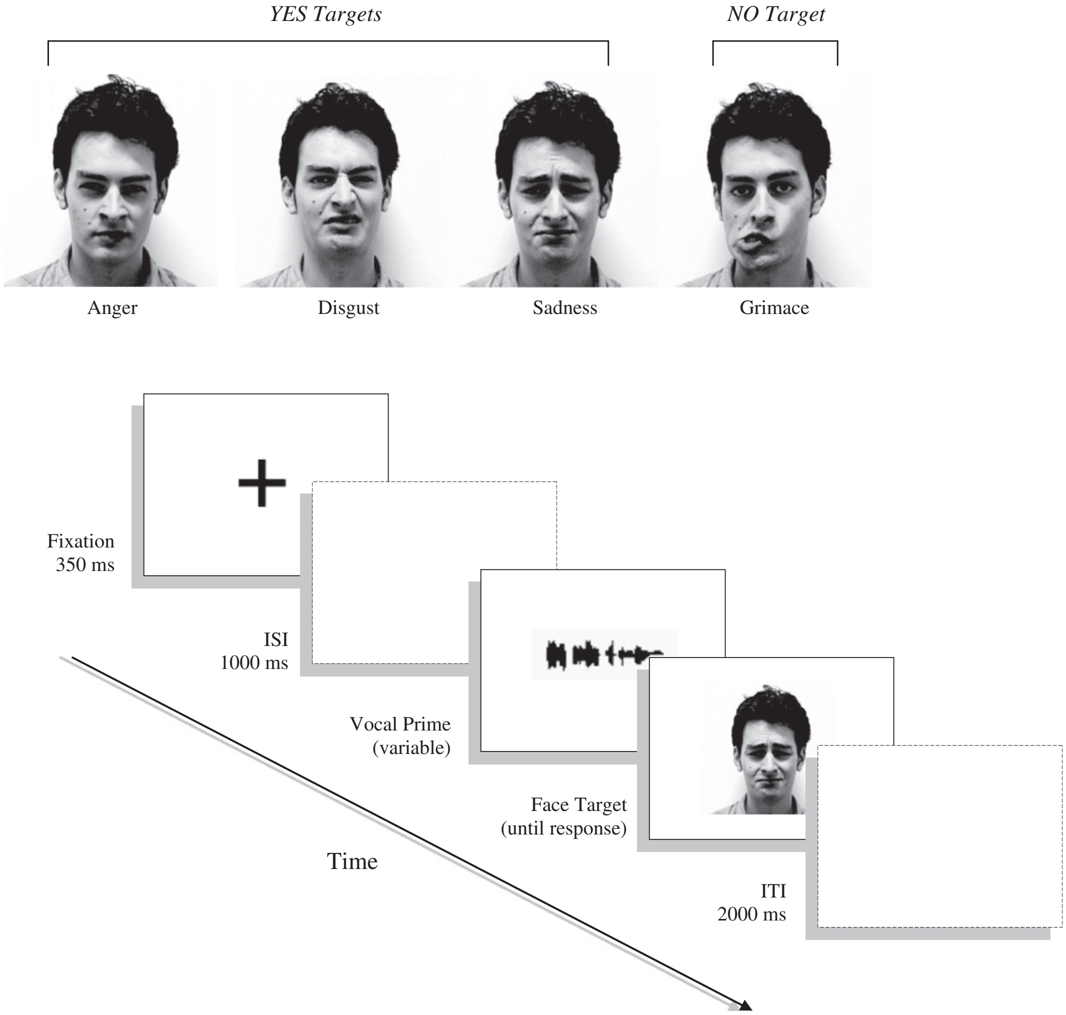

Fig. 1. Examples of face targets and illustration of a trial in the Facial Affect Decision Task. The face target contains either a discrete emotional expression (“Yes” response) or a facial grimace (“No” response).

congruent, neutral, and incongruent prime–target pairs. Identical prime utterances and target faces were assigned to separate blocks whenever possible to avoid repetition. The presentation order of the six blocks was counterbalanced across participants.

### 2.4. Procedure

Participants were tested in a quiet, sound-attenuated room. The experiment was run using SuperLab presentation software (Cedrus, USA) on a laptop computer. The vocal prime stimuli were presented through high-quality stereo headphones with the volume adjusted to a comfortable listening level. The face targets were viewed on the computer screen at a distance of about 45 cm. Participants were told that a sentence would be heard through the headphones prior to viewing a face, but that they should focus their attention on the face to judge whether it conveyed an emotion (YES response) or not (NO response), as quickly and accurately as possible. Responses were given using two buttons on a Cedrus response box and accuracy and reaction times were recorded by the computer.

Fig. 1 illustrates the construction of each trial. All trials started with the presentation of a fixation cross (350 ms) followed by a 1000 ms inter-stimulus interval. Next, the vocal pseudo-utterance prime was presented in its entirety (varying in mean duration from 1618–2144 ms); this was followed by the face target, which was always presented 1000 ms after the onset of the vocal stimulus (and therefore perceptually overlapped with the prime presentation). After participants rendered a facial affect decision, a 2000 ms interval was presented before the subsequent trial. Prior to the experimental trials, participants completed two practice blocks. In the first block, five emotional faces and five grimaces were presented without the prime and participants responded “yes” or “no” as to the presence of an emotion. In the second block, the same faces were presented preceded by congruent and incongruent vocal primes. Participants again responded and were provided with feedback on their accuracy (“Correct” or “Incorrect”), followed by their response speed if responses occurred longer than 1000 ms ("Please respond faster"). Feedback was only provided during the practice blocks. Upon completion of the study, each participant received $10 CAD.

## 3. Results

The overall error rate in the experiment was 14.9% ($SD = 10.7$), with a rate of 20.8% ($SD = 18.1$) for YES trials and 9.0% ($SD = 11.6$) for NO trials. Error rates for YES trials, while high, are consistent with those observed in previous studies for facial targets representing negative emotions (Pell and Skorup, 2008; Pell,2005a,b). One female participant who made a high number of errors for both YES and NO trials (28% and 40%, respectively) was excluded from all analyses on the basis that she may not have correctly understood the task. Following established procedures (Pell, 2005a,b; Pell et al., in press), analysis of errors and response times were performed separately on data for YES trials (i.e., those ending in a meaningful face target) to allow priming effects to be investigated according to the emotional relationship of the prime–target events. Analysis of response times were confined to data for *correct* YES target responses. To ensure that conditional means were based on a reliable number of correct responses, six additional participants with errors greater than 50% for YES trials were excluded for the response time analysis only. Thus, the error analysis considered data for 49 participants, whereas the response time analysis considered data for 43 participants. To eliminate outliers and extreme values, response times less than 300 ms and greater than 2000 ms were removed (2.3% of correct YES responses). Furthermore, individual responses over two standard deviations from each participant's conditional mean were replaced with the value equivalent to two standard deviations in the observed direction (4.3% of correct YES responses). The data were then analyzed using separate 3 × 3 repeated measures ANOVAs for errors and response times using the independent variables of face (*angry, disgusted, sad*) and prime–target relationship (*congruent, incongruent, neutral*).

Fig. 2a and b illustrates the mean errors and mean response times, respectively, as a function of face and prime–target relationship. For errors, the 3 × 3 ANOVA demonstrated a significant main effect of face, $F(2, 96) = 11.91$, $MSE = .03$, $p < .001$. Tukey's post hoc tests ($p < .05$) revealed that irrespective of the prime–target relationship, participants made significantly more errors to *disgusted* faces (24.9%) compared to *angry* (18.5%) and *sad* faces (15.8%), which did not significantly differ. Furthermore, there was a significant main effect of prime–target relationship, $F(2, 96) = 7.31$, $MSE = .03$, $p = .005$. Post hoc comparisons showed that participants made significantly less errors when the prime–target relationship was congruent (15.5%) when compared to both incongruent (20.4%) and neutral (23.3%). The interaction of face × prime–target relationship was not significant, $F(4, 192) = .30$, $MSE = .01$, *ns*.

For response times, the 3 × 3 ANOVA yielded a significant main effect of face, $F(2, 84) = 4.68$, $MSE = 9871$, $p = .02$, and prime–target relationship, $F(2, 84) = 6.25$, $MSE = 6766$, $p < .01$. Additionally, there was a significant interaction between face and prime–target relationship, $F(4, 168) = 9.49$, $MSE = 6287$, $p < .001$. Tukey's HSD post hoc tests ($p < .05$) on the interaction revealed the following effects: (1) participants responded significantly faster to *angry* and *sad* face targets when preceded by a congruent prime compared to an incongruent prime, although neutral primes acted differently (i.e., for *sad* targets, neutral differed significantly from congruent but not incongruent; for *angry* targets, neutral did not differ significantly from either incongruent or congruent; review Fig. 2b); (2) in contrast, participants responded significantly *slower* to *disgusted* face targets when listening to congruent primes compared to neutral primes (there was no difference between neutral and incongruent primes, or congruent and incongruent primes).[1]

To briefly show how different negative prime types may have influenced the processing of grimace (i.e., non-emotional) face targets, the errors and response times to NO trials for the same participants were entered into two one-way ANOVAs with repeated measures on prosody (angry, disgusted, sad, neutral). For errors, the main effect of prosody was significant, $F(3, 144) = 3.57$, $MSE = .002$, $p = .02$. Vocal primes conveying *anger* (9.7%) and *disgust* (9.7%) were accompanied by significantly more NO errors than when primes were *sad* (7.4%) or neutral (7.9%). In contrast, there was no significant influence of prosody on response times to NO trials, $F(3, 144) = 1.89$, $MSE = 3056$, *ns*.

## 4. Discussion

In the present study, we investigated whether speech prosody and facial expressions are processed with respect to their discrete emotional meanings and interact during information processing, using a cross-modal priming task. We found evidence for category-specific priming for three discrete emotions with a negative valence, as well as unique patterns for disgust (see below for a detailed discussion of this emotion). With respect to errors, our data furnish clear evidence that communicative displays are processed in reference to their emotional meanings: for each emotion of interest, participants made significantly fewer errors when judging a target face that expressed the same emotion as

[1] The 3 × 3 ANOVA on the response time data was re-run, including all outliers and extreme values, to ensure that the observed effects were not influenced by post-processing of our data. The observed effects of face, prime–target relationship, and interaction of these factors were again significant (all $F$'s > 4.17, $p$'s < .02). Post hoc comparisons revealed no differences in the interpretation of how the prime-target relationship influenced the speed of facial affect decisions.

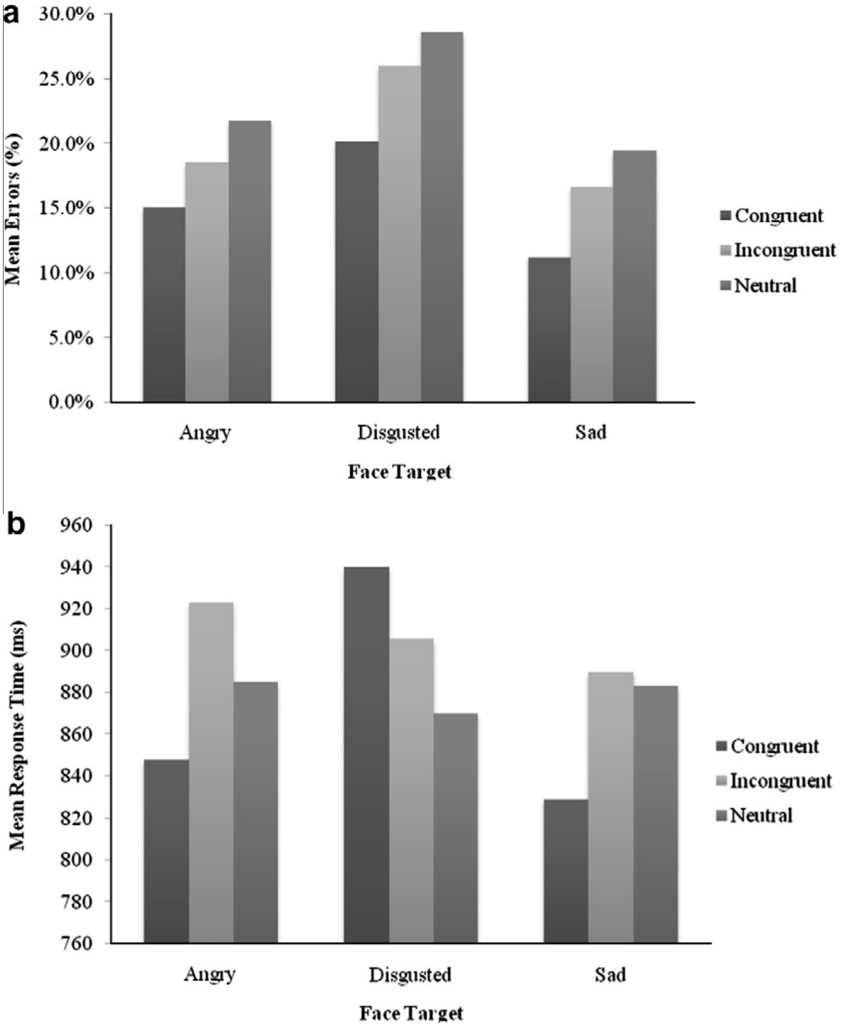

Fig. 2. Mean priming effects for (a) error rates and (b) response times, as a function of face target (angry, disgusted, sad) and prime–target relationship (congruent, incongruent, neutral).

the preceding vocal stimulus. Our response time data also showed that when judging sad and angry faces, participants responded significantly faster if they had previously heard a vocal prime expressing a categorically congruent emotion. These results indicate that vocal expressions of anger, sadness, and disgust – despite all having a negative valence – are processed in a conceptually unique manner that promotes judgments about categorically congruent facial expressions.

Our findings are in line with previous research on the representation and processing of discrete facial emotions (Ekman, 1992; Young et al., 1997; Batty and Taylor, 2003; Bimler and Kirkland, 2001; Etcoff and Magee, 1992; Krolak-Salmon et al., 2001; Levenson et al., 1990) as well as more recent research that has begun to focus on emotions expressed nonverbally through speech prosody (Laukka, 2005; Pell,2005a,b; Pell et al., in press). Studies on the categorical representation of emotions expressed through the auditory modality (Pell, 2005a; Laukka, 2005; Niedenthal and Setterlund, 1994) are less common in the literature; however, acoustic–perceptual studies strongly support the presence of distinct vocally-expressed emotion categories (Pell et al., 2009b; Juslin and Laukka, 2003) and the current investigation adds further evidence for the representation of discrete emotions inherent in the prosodic patterns of the speech signal, which are registered even when speech is devoid of meaningful lexical-semantic content such as in the pseudo-utterances used here. In particular, our study included emotion categories similar with respect to valence (i.e., negative), strengthening the conclusion that the priming effects demonstrated here were not due to the underlying affective dimension of valence, but rather characteristics unique to each emotion category of interest.

The present results are also convergent with those of Carroll and Young (2005) (Experiment 4), who used a similar priming task to demonstrate category-specific priming between facial expression primes and emotional word targets. Importantly, our study extends their results by showing that category-specific priming also occurs *across* modalities, between stimuli processed through the auditory (speech prosody) and visual (facial expressions) sensory channels. Even when the focus of the task was to attend only to the face and provide a facial affect decision, participants processed the emotional information conveyed by speech prosody and registered this knowledge in memory, which subsequently facilitated responses to related facial displays. Emotions represented and processed through the voice and the face may therefore share underlying conceptual features in associative memory, which facilitates information processing across modalities. The neurocognitive mechanisms for processing emotions from prosody and facial expressions are likely to be shared or linked (Borod et al., 2000; Pell et al., in press; Kreifelts et al., 2009), allowing for rapid access to category-specific information from different sensory inputs. Such a cognitive system that rapidly integrates emotional information from the voice and face would be critical for everyday human interactions, particularly nonverbal communication.

An important implication of our paradigm is that cross-modal interactions during emotional processing appear to occur implicitly. While participants had to explicitly identify (i.e., verbally name) emotion categories in a forced-choice format in Carroll and Young's (2005) paradigm, the present study involved an implicit yes/no judgment that presumably did not draw conscious attention to verbal labels associated with the emotion categories. Therefore, even when the focus of the task does not require categorical labeling of emotional stimuli, participants still encode emotional meanings specific to anger, sadness, and disgust from both auditory and visual nonverbal cues. A recent study that combined an implicit cross-modal priming task (the FADT) with ERP measures, revealed distinct responses to congruent vs. incongruent prime–target relationships in the N400 signal that is thought to be sensitive to contextual integration (Paulmann and Pell, 2010). Our results are in line with this research as well as previous studies that have also demonstrated implicit cross-modal emotional priming (de Gelder et al., 1999; de Gelder and Vroomen, 2000), providing evidence for the automaticity of the influence of vocally-processed emotions on facial expressions.

Our analysis of individual emotion categories allowed us to elucidate whether the nature of priming was similar for each emotion category (i.e., comparison of errors and response times to angry, sad, and disgusted face targets). Our analysis of errors showed that priming effects were similar across emotions whereby participants responded with fewer errors to congruent prime–target pairs (i.e., angry–angry, sad–sad, disgusted–disgusted). This result demonstrates that anger, sadness, and disgust are processed as conceptually unique emotion categories, both vocally though speech prosody, as well as through facial expressions. However, while our response time data showed facilitating emotion congruency effects for anger and sadness, this was not the case with disgust. In fact, participants responded *slower* when both the vocal prime and face target conveyed disgust (review Fig. 2), evincing a pattern opposite to our hypothesis. For disgust, this slowing of response time but improvement in accuracy in the congruent prime–target condition implies a speed/accuracy trade-off. Traditionally, disgust has been thought of as a "defensive" emotion that signifies revulsion or rejection of a particular stimulus or object (Cisler et al., 2009; Charash and McKay, 2002), which has prompted researchers to investigate the presence of an attention bias. In support of this notion, Charash and McKay (2002) primed participants with stories conveying disgust and then had them perform an emotional Stroop task using words with disgusted lexical-semantic meaning. Participants were slower to respond to disgusted words, suggesting an attention bias to disgust-related stimuli. A related study found impaired attention to and difficulty disengaging from disgusted words, although this depended on the task-relevance of the emotional stimuli (Cisler et al., 2009). In the present study, if participants had an automatic attention bias or difficulty disengaging from disgusted prime utterances, this may have led to a slowed response time to disgusted face targets, similar to findings using a Stroop task (Charash and McKay, 2002). However, the activation of disgust-specific knowledge by the vocal prime would presumably still have contributed to improved accuracy when judging whether or not a subsequently presented face conveyed an emotion, which would manifest as a speed/accuracy trade-off.

In studies by Charash and McKay (2002) and Cisler and colleagues (2009), the attention bias to disgust was additionally moderated by individual participants' degree of disgust sensitivity, suggesting that individual differences play a role in emotional processing and could be an important variable for future research in this domain. Furthermore, both studies found additional evidence for a similar attention bias to fear-related words and though we did not include fear as a category, it would be noteworthy to incorporate this emotion in future studies to investigate whether cross-modal priming effects with fearful stimuli are similar to those with disgust (see Paulmann and Pell (2010) for a recent example using fear).

Another reason for our findings related to disgust may be due to the manner in which disgust is most accurately communicated. Using vocal sentences as carriers of emotional information, studies have frequently demonstrated relatively poor accuracy and increased processing demands related to disgust recognition (Banse and Scherer, 1996; Pell et al., 2009b; Juslin and Laukka, 2001). By comparison, studies investigating the recognition of short nonverbal emotional vocalizations or "affective bursts" have demonstrated relatively higher recognition accuracy for disgust (Schroder, 2003; Simon-Thomas et al., 2009). In general, nonverbal emotional vocalizations differ from emotional speech in acoustic properties such as pitch contours that may enhance emotion recognition (see Scott et al. (2009) for an example). Disgust may therefore be more accurately or efficiently captured through emotional vocalizations (e.g., "*Ugh!*") compared to sentences (Banse and Scherer, 1996). Given that in the present study disgust was conveyed via pseudo-sentences with a mean recognition rate of 80% and a mean length of 2.14 s, it is possible that increased processing demands and less rapid recognition associated with this emotion from sentence-length speech contributed to the lack of response time facilitation, which may differ from studies using nonverbal vocalizations as primes. In support of this hypothesis, Carroll and Young (2005) showed priming effects using nonverbal vocalizations as primes (including "retching" sounds for disgust) and emotional words and faces as targets, although the authors did not compare the extent of priming for each emotion individually. Future studies should include nonverbal vocalizations of disgust using a similar priming paradigm to elucidate whether priming effects vary as a function of the stimulus type. While these hypotheses are only speculative, the current results still support the view of disgust as a unique emotion category (especially when considering error rates), although processing speed advantages appear to differ from anger and sadness and may vary based on the nature of the stimulus.

Finally, although the present study demonstrates emotion-specific effects in the processing of vocal and facial nonverbal cues, our data do not discount the relevance of underlying affective dimensions of emotional displays, such as valence and arousal. For example, Laukka et al. (2005) have shown that vocally expressed discrete emotions of happiness, anger, sadness, fear, and disgust are all characterized by unique combinations along the dimensions of valence, arousal, and potency (as perceived and rated by a group of university students and speech experts). Moreover, these dimensions were correlated with specific combinations of acoustic cues. In the present study, although our three emotions of interest were similar in terms of valence, the arousal levels of our stimuli differed. While it is possible that arousal was a factor influencing the priming results, both anger (relatively high arousal) and sadness (relatively low arousal) displayed response time facilitation. Moreover, unique priming effects were seen for disgust, despite having an arousal level equal to sad vocal primes (review Table 1). This suggests that arousal did not differentially affect the priming observed with these emotions. In a recent investigation on affective priming using high and low arousal stimuli while controlling for valence, Hinojosa et al. (2009) found that prime–target pairs that were congruent in terms of arousal did not demonstrate priming behaviorally (i.e. no facilitation of response times), although there were electrophysiological differences as measured by ERPs. Future studies should further investigate the role of arousal on cross-modal emotional priming. In general, differences along major affective dimensions of emotional displays are likely relevant in conveying subtle distinctions in meaning within emotion families (Ekman, 1992), which is likely to affect how emotional events interact in different sensory or communication channels. Nonetheless, it can be said that the activation of emotional knowledge from prosody and facial expressions, which provides a crucial mechanism for social information processing, involves processing the category-specific meanings of communicative displays (Paulmann and Pell, 2010).

## 5. Conclusion

Our findings argue that despite sharing the same valence, anger, sadness, and disgust are communicated and recognized as unique emotion categories. These results add to an existing body of knowledge that supports this notion using acoustic–perceptual techniques (Banse and Scherer, 1996; Juslin and Laukka, 2003; Pell et al., 2009b). More generally, our results contribute to a growing field of research on multi- and cross-modal processing of basic emotions that incorporates neuroimaging (Kreifelts et al., 2007, 2009), electrophysiological (Paulmann and Pell, 2009; Paulmann et al., 2009; Schirmer et al., 2002), and behavioral techniques (Paulmann et al., in press). Increasingly, integrative approaches to studying emotions will provide researchers with new views into how humans process, conceptualize, and communicate socially relevant emotional information that is a hallmark of everyday interpersonal interactions.

## Acknowledgements

This research was financially supported by the Natural Sciences and Engineering Research Council of Canada (Discovery grants competition) and by McGill University (William Dawson Scholar Award to MDP). We thank Catherine Knowles, Shoshana Gal, Laura Monetta, Pan Liu, and Hope Valeriote for their input and help with data collection and manuscript preparation.

## References

Banse, R., Scherer, K.R., 1996. Acoustic profiles in vocal emotion expression. J. Pers. Soc. Psychol. 70 (3), 614–636.

Batty, M., Taylor, M.J., 2003. Early processing of the six basic facial emotional expressions. Cogn. Brain Res. 17, 613–620.

Bimler, D., Kirkland, J., 2001. Categorical perception of facial expressions of emotion: evidence from multidimensional scaling. Cogn. Emotion 15 (5), 633–658.

Borod, J.C., Pick, L.H., Hall, S., Sliwinski, M., Madigan, N., Obler, L.K., Welkowitz, J., Canino, E., Erhan, H.M., Goral, M., Morrison, C., Tabert, M., 2000. Relationships among facial, prosodic, and lexical channels of emotional perceptual processing. Cognition Emotion 14 (2), 193–211.

Bowers, D., Bauer, R.M., Heilman, K.M., 1993. The nonverbal affect lexicon: theoretical perspectives from neuropsychological studies of affect perception. Neuropsychology 7, 433–444.

Carroll, N.C., Young, A.W., 2005. Priming of emotion recognition. Quart. J. Exp. Psychol. 58A (7), 1173–1197.

Charash, M., McKay, D., 2002. Attention bias for disgust. Anxiety Disorders 16, 529–541.

Cisler, J.M., Olatunji, B.O., Lohr, J.M., Williams, L.M., 2009. Attention bias differences between fear and disgust: implications for the role of disgust in disgust-related anxiety disorders. Cognition Emotion 23 (4), 675–687.

de Gelder, B., Vroomen, J., 2000. The perception of emotions by ear and by eye. Cognition Emotion 14 (3), 289–311.

de Gelder, B., Bocker, K.B.E., Tuomainen, J., Hensen, M., Vroomen, J., 1999. The combined perception of emotion from voice and face: early interaction revealed by human brain responses. Neurosci. Lett. 260, 133–136.

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

Etcoff, N.L., Magee, J.L., 1992. Categorical perception of facial expressions. Cognition 44, 227–240.

Fazio, R.H., 2001. On the automatic activation of associated evaluations: an overview. Cognition Emotion 15 (2), 115–141.

Gerber, A.J., Posner, J., Gorman, D., Colibazzi, T., Yu, S., Wang, Z., Kangarlu, A., Zhu, H., Russell, J., Peterson, B.S., 2008. An affective circumplex model of neural systems subserving valence, arousal, and cognitive overlay during the appraisal of emotional faces. Neuropsychologia 46, 2129–2139.

Goldstone, R.L., Hendrickson, A.T., 2010. Categorical perception. Wiley Interdiscipl. Rev.: Cogn. Sci. 1 (1), 69–78.

Hermans, D., De Houwer, J., Eelen, P., 1994. The affective priming effect: automatic activation of evaluative information in memory. Cognition Emotion 8 (6), 515–533.

Hietanen, J.K., Leppanen, J.M., Illi, M., Surakka, V., 2004. Evidence for the integration of audiovisual emotional information at the perceptual level of processing. Eur. J. Cogn. Psychol. 16 (6), 769–790.

Hinojosa, J.A., Carretie, L., Mendez-Bertolo, C., Miguez, A., Pozo, M.A., 2009. Arousal contributions to affective priming: electrophysiological correlates. Emotion 9 (2), 164–171.

Juslin, P.N., Laukka, P., 2001. Impact of intended emotional intensity on cue utilization and decoding accuracy in vocal expression of emotion. Emotion 1 (4), 381–412.

Juslin, P.N., Laukka, P., 2003. Communication of emotions in vocal expression and music performance: different channels, same code? Psychol. Bull. 129 (5), 770–814.

Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., Wildgruber, D., 2007. Audiovisual integration of emotional signals in voice and face: an event-related fMRI study. NeuroImage 37, 1445–1456.

Kreifelts, B., Ethofer, T., Shiozawa, T., Grodd, W., Wildgruber, D., 2009. Cerebral representation of non-verbal emotional perception: fMRI reveals audiovisual integration area between voice- and face-sensitive regions in the superior temporal sulcus. Neuropsychologia 47 (14), 3059–3066.

Krolak-Salmon, P., Fischer, C., Vighetto, A., Mauguiere, F., 2001. Processing of facial emotional expression: spatio-temporal data as assessed by scalp event-related potentials. Eur. J. Neurosci. 13, 987–994.

Laukka, P., 2005. Categorical perception of vocal emotion expressions. Emotion 5 (3), 277–295.

Laukka, P., Juslin, P., Bresin, R., 2005. A dimensional approach to vocal expression of emotion. Cognition Emotion 19 (5), 633–653.

Levenson, R.W., Ekman, P., Friesen, W.V., 1990. Voluntary facial action generates emotion-specific autonomic nervous system activity. Psychophysiol. 27 (4), 363–384.

Massaro, D., Egan, P., 1996. Perceiving affect from the voice and the face. Psychon. Bull. Rev. 3 (2), 215–221.

Niedenthal, P.M., Setterlund, M.B., 1994. Emotion congruence in perception. Pers. Soc. Psychol. Bull. 20, 401–411.

Palermo, R., Coltheart, M., 2004. Photographs of facial expression: accuracy, response times, and ratings of intensity. Behav. Res. Methods Instrum. Comput. 36 (4), 634–638.

Paulmann, S., Pell, M.D., 2009. Facial expression decoding as a function of emotional meaning status: ERP evidence. NeuroReport 20, 1603–1608.

Paulmann, S., Pell, M.D., 2010. Contextual influences of emotional speech prosody on face processing: how much is enough? Cogn. Affect. Behav. Neurosci. 10, 230–242.

Paulmann, S., Jessen, S., Kotz, S.A., 2009. Investigating the multimodal nature of human communication. J. Psychophysiol. 23 (2), 63–76.

Paulmann, S., Titone, D., Pell, M.D., in press. How emotional prosody guides your way: evidence from eye movements. Speech Comm.

Pell, M.D., 2002. Evaluation of nonverbal emotion in face and voice: some preliminary findings on a new battery of tests. Brain Cognition 48, 499–504.

Pell, M.D., 2005a. Nonverbal emotion priming: evidence from the ‘facial affect decision task’. J. Nonverbal Behav. 29 (1), 45–73.

Pell, M.D., 2005b. Prosody–face interactions in emotional processing as revealed by the facial affect decision task. J. Nonverbal Behav. 29 (4), 193–215.

Pell, M.D., Kotz, S.A., in review. On the timecourse of vocal emotion recognition.

Pell, M.D., Skorup, V., 2008. Implicit processing of emotional prosody in a foreign versus native language. Speech Comm. 50, 519–530.

Pell, M.D., Monetta, L., Paulmann, S., Kotz, S.A., 2009a. Recognizing emotions in a foreign language. J. Nonverbal Behav. 33 (2), 107–120.

Pell, M.D., Paulmann, S., Dara, C., Alasseri, A., Kotz, S.A., 2009b. Factors in the recognition of vocally expressed emotions: a comparison of four languages. J. Phonetics 37 (4), 417–435.

Pell, M.D., Jaywant, A., Monetta, L., Kotz, S.A., in press. Emotional speech processing: disentangling the effects of prosody and semantic cues. Cognition Emotion, doi:10.1080/02699931.2010.516915.

Posner, J., Russell, J.A., Peterson, B.S., 2005. The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. Develop. Psychopathol. 17, 715–734.

Pourtois, G., de Gelder, B., Vroomen, J., Rossion, B., Crommelinck, M., 2000. The time-course of intermodal binding between seeing and hearing affective information. NeuroReport 11 (6), 1329–1333.

Scherer, K.R., Banse, R., Wallbott, H.G., Goldbeck, T., 1991. Vocal cues in emotion encoding and decoding. Motiv. Emotion 15 (2), 123–148.

Schirmer, A., Kotz, S.A., Friederici, A.D., 2002. Sex differentiates the role of emotional prosody during word processing. Cogn. Brain Res. 14 (2), 228–233.

Schirmer, A., Kotz, S.A., Friederici, A., 2005. On the role of attention for the processing of emotions in speech: sex differences revisited. Cogn. Brain Res. 24, 442–452.

Schroder, M., 2003. Experimental study of affect bursts. Speech Comm. 40, 99–116.

Scott, S.K., Sauter, D., McGettigan, C., 2009. Brain mechanisms for processing perceived emotional vocalizations in humans. In: Brudzynski, S.M. (Ed.), Handbook of Mammalian Vocalization. Academic Press, Oxford, pp. 187–198.

Simon-Thomas, E.R., Keltner, D.J., Sauter, D., Sinicropi-Yao, L., Abramson, A., 2009. The voice conveys specific emotions: evidence from vocal burst displays. Emotion 9 (6), 838–846.

Spruyt, A., De Houwer, J., Hermans, D., Eelen, P., 2007. Affective priming of nonaffective semantic categorization responses. Exp. Psychol. 54 (1), 44–53.

Tracy, J.L., Robins, R.W., 2008. The automaticity of emotion recognition. Emotion 8 (1), 81–95.

Vroomen, J., Driver, J., de Gelder, B., 2001. Is cross-modal integration of emotional expressions independent of attentional resources. Cogn. Affect. Behav. Neurosci. 1, 382–387.

Williams, L.M., Mathersul, D., Palmer, D.M., Gur, R.C., Gur, R.E., Gordon, E., 2009. Explicit identification and implicit recognition of facial emotions: I. Age effects in males and females across 10 decades. J. Clin. Exp. Neuropsychol. 31 (3), 257–277.

Young, A., Rowland, D., Calder, A., Etcoff, N., Seth, A., Perrett, D., 1997. Facial expression megamix: tests of dimensional and category accounts of emotion recognition. Cognition 63, 271–313.

Zhang, Q., Lawson, A., Guo, C., Jiang, Y., 2006. Electrophysiological correlates of visual affective priming. Brain Res. Bull. 71, 316–323.