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.
© 2011 Elsevier B.V. All rights reserved.

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

دریافت فوری ← متن کامل مقاله

ISIArticles
مرجع مقالات تخصصی ایران

✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
✓ امکان دانلود نسخه ترجمه شده مقالات
✓ پذیرش سفارش ترجمه تخصصی
✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
✓ دانلود فوری مقاله پس از پرداخت آنلاین
✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات