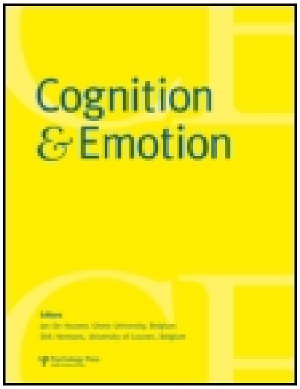


Cognition & Emotion
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/pcem20

Emotional speech processing: Disentangling the effects of prosody and semantic cues

Marc D. Pell [a], Abhishek Jaywant [a], Laura Monetta [b] & Sonja A. Kotz [c]

[a] McGill University, Montréal, Canada

[b] Université Laval, Montréal, Canada

[c] Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Available online: 10 Nov 2010

To cite this article: Marc D. Pell, Abhishek Jaywant, Laura Monetta & Sonja A. Kotz (2011): Emotional speech processing: Disentangling the effects of prosody and semantic cues, Cognition & Emotion, 25:5, 834-853

To link to this article: http://dx.doi.org/10.1080/02699931.2010.516915

COGNITION AND EMOTION
2011, 25 (5), 834–853

Psychology Press
Taylor & Francis Group

Emotional speech processing: Disentangling the effects of prosody and semantic cues

Marc D. Pell[1], Abhishek Jaywant[1], Laura Monetta[2], and Sonja A. Kotz[3]

[1]McGill University, Montréal, Canada
[2]Université Laval, Montréal, Canada
[3]Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

To inform how emotions in speech are implicitly processed and registered in memory, we compared how emotional prosody, emotional semantics, and both cues in tandem prime decisions about conjoined emotional faces. Fifty-two participants rendered facial affect decisions (Pell, 2005a), indicating whether a target face represented an emotion (happiness or sadness) or not (a facial grimace), after passively listening to happy, sad, or neutral prime utterances. Emotional information from primes was conveyed by: (1) prosody only; (2) semantic cues only; or (3) combined prosody and semantic cues. Results indicated that prosody, semantics, and combined prosody–semantic cues facilitate emotional decisions about target faces in an emotion-congruent manner. However, the *magnitude* of priming did not vary across tasks. Our findings highlight that emotional meanings of prosody and semantic cues are systematically registered during speech processing, but with similar effects on associative knowledge about emotions, which is presumably shared by prosody, semantics, and faces.

Keywords: Speech perception; Emotions; Audio-visual priming; Vocal cues; Facial expression.

INTRODUCTION

Spoken language contains more than "language" or linguistic information; discourse provides a context where speakers realise a variety of social goals by manipulating language, prosody, and other nonverbal cues in tandem. In speech, lexical-semantic information (hereafter, referred to simply as "semantics" or "semantic cues") is inextricably tied to prosodic information, as meanings encoded in these two communication channels exploit distinct auditory characteristics of the same physical event. In some contexts, information in one speech channel (e.g., prosody) serves to modify or negate meanings conveyed in the other channel (e.g., semantics), as frequently occurs when speakers convey irony (e.g., Cheang & Pell, 2008). In other contexts, the underlying

Correspondence should be addressed to: Marc D. Pell, McGill University, School of Communication Sciences & Disorders, 1266 avenue des Pins Ouest, Montréal, Québec H3G 1A8, Canada. E-mail: marc.pell@mcgill.ca

This research was funded by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (to MDP) and by the German Research Foundation (DFG FOR-499 to SAK).

We would like to thank Nicole Price and Catherine Knowles for their extensive help with data collection and organisation.

http://www.psypress.com/cogemotion
DOI:10.1080/02699931.2010.516915

meaning encoded by prosody and semantics is meant to be congruent, although intended meanings may be more salient to listeners in only one of these channels. Effective monitoring of prosody and semantic cues during spoken language processing is therefore essential to interpret the intended meanings of a speaker, their true beliefs, attitudes, and emotions (Caffi & Janney, 1994; Tanenhaus & Brown-Schmidt, 2008).

A very common situation where speakers use both prosody and semantic cues to reinforce shared meanings is to communicate emotion (Grandjean, Banziger, & Scherer, 2006; Wilson & Wharton, 2006). The past decade has been marked by a surge of interest in how humans process emotion, although studies of emotional speech are still limited. With respect to prosody, acoustic-perceptual studies show that listeners attend to relative changes in pitch, loudness, timing, and voice quality attributes over the course of an utterance to recognise a speaker's emotion (Banse & Scherer, 1996). There is clear evidence that a speaker's emotion can be accurately recognised from prosody independent of semantic cues, either when listening to emotionally intoned but semantically anomalous "pseudo-utterances" (Pell, Paulmann, Dara, Alasseri & Kotz, 2009b; Scherer, Banse, Wallbott, & Goldbeck, 1991) or when listening to a foreign language (Pell, Monetta, Paulmann, & Kotz, 2009a; Thompson & Balkwill, 2006). Even when the semantic content of utterances is emotionally neutral or nonsensical, data show that emotional meanings of prosody are activated in a rapid and involuntary manner by listeners (de Gelder & Vroomen, 2000; Hietanen, Leppanen, Illi, & Surakka, 2004; Kitayama & Ishii, 2002; Kotz & Paulmann, 2007; Paulmann & Kotz, 2008; Pell, 2005a, 2005b; Pourtois, de Gelder, Vroomen, Rossion, & Crommelinck, 2000).

Similarly, research implies that processing the emotional semantic context of utterances (e.g., "That man has a knife!") is associated with differential responses when listening to speech with or without corresponding emotional prosody (Beaucousin et al., 2007; Kotz & Paulmann, 2007; Mitchell, Elliott, Barry, Cruttenden, & Woodruff, 2003). An obvious methodological constraint of studying auditory language processing, which is highlighted by this work, is that semantic effects cannot be empirically evaluated in the complete absence of prosodic cues; rather, researchers simply attempt to mitigate the influence of prosody by using computer speech or utterances spoken in a "neutral" tone. Nonetheless, it seems likely that the neurocognitive mechanisms for processing information from emotional semantic cues versus emotional prosody are dissociable; for example, functional neuroimaging data show that the processing of non-emotional speech implicates neural networks mainly in the left hemisphere (Friederici, 2002; Hagoort, Hald, Bastiaansen, & Petersson, 2004), whereas sentences with a clear emotional context appear to induce additional right-sided involvement, with or without emotional prosody (Beaucousin et al., 2007; Borod et al., 1998; Buchanan et al., 2000; Mitchell et al., 2003; see Schirmer & Kotz, 2006, for an overview). Research using event-related brain potentials (ERPs) has also reported distinct lateralisation and ERP responses when processing emotional semantics, prosody, or both in tandem (Paulmann & Kotz, 2008; Wambacq & Jerger, 2004). These findings underscore that emotional semantics and emotional prosody in speech are treated differently, at least in terms of the neurocognitive mechanisms involved.

However, empirical data that inform the relative *degree* to which listeners harness prosody versus semantic cues, or information from both channels in combination, to activate and retrieve emotional meanings during ongoing speech processing, are scarce. That is, the relative potential of each information channel to register knowledge about a speaker's emotion while listening to speech is poorly understood, although there are suggestions that the channels do vary in their importance, for example, due to cultural preferences (Ishii, Reyes, & Kitayama, 2003; Kitayama & Ishii, 2002). Based on studies of "off-line" sentence processing, the idea that emotional semantic information takes precedence over emotional prosody, or somehow assumes greater weight in this analysis, is implied by several lines of inquiry. First, it is ubiquitously shown that when (healthy or brain damaged) listeners are asked to name a speaker's emotion in a forced-choice response format, they are significantly more accurate when presented with utterances that contain emotional semantic cues rather than only prosodic cues (Dara, Monetta, & Pell, 2008; Johnson, Emde, Scherer, & Klinnert, 1986; Pell, 2006; Pell & Baum, 1997). Moreover, in experiments where the emotional meaning conveyed by semantic cues and prosody is manipulated to conflict within the same utterance, listeners tend to judge the speaker's emotion based on the semantic meaning (Bowers, Coslett, Bauer, Speedie, & Heilman, 1987; Breitenstein, Daum, & Ackermann, 1998). These results imply that emotional information derived from semantics is somehow stronger than that conveyed by prosody, at least when this issue is investigated using forced-choice emotion identification tasks (but cf. Morton & Trehub, 2001).

Unfortunately, these traditional off-line approaches do not sensitively index the degree to which prosody and semantic information activate knowledge about a speaker's emotion as speech unfolds in real time. To address this concern, other researchers have investigated the interaction between emotional prosody and semantics in tasks where these cues are processed in a more implicit manner. For example, several researchers have presented interference paradigms (e.g., Stroop, dichotic listening) in which emotional words are spoken with a congruent or incongruent prosody, and participants must attend to either the prosody or the semantic emotion (Grimshaw, 1998; Kitayama & Howard, 1994; Kitayama & Ishii, 2002; Morais & Ladavas, 1987; Schirmer & Kotz, 2003; Techentin, Voyer, & Klein, 2009). In general, these data suggest that when prosody and semantic cues are presented in an unattended channel, they can each influence emotional decisions about the stimulus, irrespective of attentional focus; however, semantic information seems to hold greater potential to interfere with conscious judgements of emotional prosody than vice versa.

In a recent ERP study that involved implicit processing of emotional cues, Kotz and Paulmann (2007) presented emotional sentences, which were cross-spliced to violate expectancies generated by emotional prosody alone, or by combined prosody and emotional semantic cues, while listeners monitored for the presence of a target word in the sentence. The authors reported that processing emotional prosody alone, when compared to combined prosody/semantic cues to emotion, is associated with a distinct time-course when ERP data were analysed in each condition. Of particular note, ERP responses in the combined prosody–semantic condition appeared to be guided largely by the *semantic* content of German utterances rather than by the prosody, although both sets of cues were present. These results further imply that emotional semantics override emotional prosody when these two sources of information interact in real time (Kotz & Paulmann, 2007; Paulmann & Kotz, 2008). Accordingly, there are now several sources of evidence, using both off-line and on-line methodologies, which suggest that the impact of emotional prosody and semantics is not entirely uniform when listeners process emotional information from speech.

Emotional meanings in memory: Further evidence from priming

An important undercurrent of the research summarised, which is not always explicitly addressed by researchers, refers to the cognitive apparatus by which stored knowledge about emotions is activated and retrieved from different types of emotional cues. While opinions vary, it can be argued that emotional meanings are perceived and represented categorically in an associative memory network (Bower, 1981; Carroll & Young, 2005; Ekman, 1992; Laukka, 2005; Niedenthal, Halberstadt, & Innes-Ker, 1999; Niedenthal, Halberstadt, & Setterlund, 1997), and that these concepts can be accessed by cues from various sensory modalities (Mahon & Caramazza, 2008). Within this framework, researchers have analysed priming effects as a way to infer how humans implicitly process different types of emotional expressions (encoded by prosody, semantic cues, faces, etc.), allowing us to gauge how different emotional events activate the emotional conceptual store. This literature has uncovered increasing evidence of "emotion congruency" effects in information processing: decisions about an emotional target stimulus are systematically enhanced when preceded by a prime word, sentence, facial expression, or picture that represents emotionally *congruent* versus incongruent features in memory (e.g., Carroll & Young, 2005; Hsu, Hetrick, & Pessoa, 2008; Niedenthal, 1990; Niedenthal et al., 1997). Although not the focus of this report, congruency effects referring to the specific emotional meanings of two events may be viewed as related, albeit distinct, from general "affective priming" effects reported in the wider literature, which are linked to the positive–negative valence of the events (see Fazio, 2001, for a discussion of affective priming).

Only a small number of studies have presented spoken utterances, as opposed to isolated words or faces, to investigate emotional congruency effects in priming studies. Wurm, Vakoch, Strasser, Calin-Jageman, and Ross (2001) presented emotionally intoned English utterances conveying happiness, disgust, or fear, and then required participants to make a lexical decision about the sentence-final word of the utterance (which was emotionally congruent or incongruent with the sentence prosody). They found that the sentence prosody primed lexical decisions about emotionally congruent, sentence-final words, at least in certain conditions (when listeners allocated attention to the prosody). Similar results have been reported by Schirmer, Kotz, and Friederici (2002, 2005), who found that semantically neutral utterances produced with a happy or sad prosody primed lexical decisions about emotionally congruent target words in German. These findings imply that prosodic cues in a sentence activate or "amplify" emotion-related knowledge in memory, which enhances cognitive processing of a related word target (Kitayama & Howard, 1994; Wurm et al., 2001). However, the extent to which prosody allows access to emotional knowledge in memory when compared to semantic cues during speech processing cannot be inferred from these data.

Using an adaptation of the lexical decision task—the Facial Affect Decision Task, or FADT—recent studies by Pell and colleagues (Pell, 2005a, 2005b; Pell & Skorup, 2008) have tested whether prosodic attributes of emotional pseudo-utterances prime decisions about a target facial expression presented at sentence offset. In the FADT, as in the lexical decision task, participants render a yes/no decision about whether a face represents a "real" emotional meaning stored in memory (e.g., happiness, sadness) or does not represent an emotion (e.g., a facial grimace whose physical configuration does not convey a basic emotion). Performance is evaluated in the context of emotionally congruent (i.e., matching emotion in the speech prime and face target) versus incongruent prosodic cues presented in the prime stimulus. Studies using this approach provide further evidence that emotional prosody is implicitly processed by listeners, whether listening to pseudo-utterances (Pell, 2005a, 2005b) or to a foreign language (Pell & Skorup, 2008), and that the meanings activated by prosody facilitate facial affect decisions in an emotion-congruent manner. However, to date only emotional pseudo-utterances devoid of semantic information have been presented in the FADT; again, this restricts present claims as to how emotional prosody and not semantic information accesses emotion-related knowledge in memory. The prospect that emotional face targets may be differentially primed by congruent prosody, congruent semantic cues, or both cue sources in combination has not yet been tested.

The present study

The purpose of this study was to establish that both prosody *and* semantic cues prime decisions about an emotionally congruent target face ("congruency effect"), and to compare the *magnitude* of cross-modal priming produced by isolated prosodic cues, isolated semantic cues, and a combination of prosody and semantic cues conveying the same emotional meaning in speech. Our study involved three distinct tasks in which only features of the prime stimulus were manipulated; listeners were exposed to emotionally meaningful cues conveyed only by prosody (Prosody Task), only by semantics (Semantic Task), or by emotionally congruent prosody and semantic cues combined (Prosody–Semantic Task). Following previous research (Pell, 2005b; Schirmer et al., 2002), we restricted our purview to a relatively uncontroversial set of emotions (happiness and sadness), to focus our findings on comparisons among the three task conditions. These comparisons will help clarify whether emotional semantics are indeed more powerful than prosody during the (implicit) processing of emotional speech. This will also allow new insights into the presumed operation of the conceptual memory store as it relates to emotional prosody and semantics, such as how conceptual information about emotions is accessed and how it interacts in memory, depending on the nature of the cue(s) present. Simultaneously, our data will shed new light on potential cross-modal interactions in the processing of communicative displays of emotion encoded by different cues in the auditory and visual modalities (e.g., Carroll & Young, 2005; de Gelder & Vroomen, 2000).

Based on the literature, we predicted that listeners would implicitly evaluate emotional speech information conveyed by both prosodic and semantic cues, yielding significant emotion congruency (priming) effects in all three tasks. The relative extent of priming observed in each task could not be predicted with certainty. Given suggestions that semantic cues take precedence over prosody in many speech-processing environments, it is possible that facial affect decisions will be enhanced (i.e., faster and/or more accurate) in the tasks where semantic cues are available to guide emotional processing (i.e., Semantic Task and Prosody–Semantic Task) when compared to the Prosody Task. Furthermore, it can be reasoned that listening to combined prosody and semantic cues to emotion might have an *additive* or amplification effect, yielding increased priming in the Prosody–Semantic condition (e.g., Wurm et al., 2001). Alternatively, semantic cues might largely override prosodic cues in the Prosody–Semantic Task (Kotz & Paulmann, 2007; Paulmann & Kotz, 2008), revealing similar priming patterns in the Semantic and Prosody–Semantic Tasks, when compared to the Prosody Task.

METHODS

Participants

Fifty-two adults volunteered for the study by responding to an electronic campus advertisement. All participants were undergraduate or graduate students at McGill University who spoke (Canadian) English as their native language. Participants averaged 23.7 years in age ($SD = 5.7$) with a mean education of 15.6 years ($SD = 1.9$). An equal number of male ($n = 26$) and female ($n = 26$) participants completed all experimental tasks. All participants reported good hearing and had normal or corrected-to-normal vision as verified by the examiner at study onset. All procedures were ethically approved by a McGill University review panel in accordance with the Helsinki Declaration.

Materials/stimulus validation procedures

The materials used in the study were emotional sentences (as the primes) and static facial expressions of emotion (as the targets), which were paired for cross-modal presentation according to the Facial Affect Decision Task (FADT). All prime and target displays were constructed and perceptually validated in an earlier norming study, and many of these stimuli have been used in previous experiments (Paulmann & Pell, 2010; Pell, 2005a, 2005b; Pell & Skorup, 2008; Pell et al., 2009b). A summary of the stimulus validation procedures, and the procedure for selecting stimuli for the current study, is provided below. For detailed commentary on the construction and assumptions of the FADT, consult Pell (2005a).

(a) Utterance primes

The prime stimuli were short, spoken sentences of approximately 7–10 syllables in length produced by two male and two female English speakers. Each task included 12 distinct tokens (three tokens per speaker), which expressed one of three emotions: happy, sad, and neutral (3 emotions × 12 items = 36 primes per task). The manner in which prosody and/or semantic features of the prime conveyed each of the three emotions was manipulated uniquely for each task, resulting in three utterance prime types: emotionally intoned but semantically anomalous pseudo-utterances (Prosody Task); neutrally intoned utterances with an emotional semantic context (Semantic Task); and emotionally intoned utterances with congruent emotional semantics (Prosody–Semantic Task).

Stimulus elicitation and recording procedure. All prime stimuli presented in the three tasks were constructed and recorded at the same time, and then extensively piloted to ensure that the intended emotional meaning could be accurately recognised by English participants prior to this study. Details about stimulus construction and validation are summarised briefly here, but can be consulted in full elsewhere (Pell et al., 2009b). Prior to eliciting the recordings, a list of 30 simple utterances associated with an emotional semantic context were constructed for each emotion (e.g., for sad: "I have nothing to hope for"). A comparable set of pseudo-utterances was then constructed by replacing the content words of the original utterances with sound strings that were phonologically licensed in English but semantically meaningless (e.g., "I marlipped the tovity"). Pseudo-utterances retained certain grammatical elements, which made them highly "language-like", and which ensured that the speakers could produce them with ease to express emotions through prosody. During recording, the speakers first produced the utterances with combined prosody and semantic cues to express each of the emotions, and then produced the corresponding pseudo-utterance in the same emotional tone. In a later recording session, the speakers were instructed to produce the same list of utterances with emotional semantic cues in a way that should sound emotionally neutral to listeners, or like a "reporter" would speak. All stimuli were recorded onto digital audiotape, saved and edited as.wav files on a PC, and then piloted to gather necessary data to ensure that only utterances that encoded the intended emotional meaning through prosody and/or semantic cues were selected for the present investigation.

Prime validation procedure. To establish that prosodic attributes of the pseudo-utterance primes strongly conveyed the emotions of interest, the recordings were presented to a group of 24 young English listeners in a larger emotion-recognition experiment involving 840 exemplars of 7 different emotions (Pell et al., 2009b). During a separate testing session, 20 of these listeners also judged the utterances with congruent emotional prosody and emotional semantic cues using an identical procedure. Based on the outcome of these perceptual experiments, only happy, sad, and neutral utterances recognised correctly by 70% of the listener group, a rate of approximately five times chance (14%) in the validation study, were selected for the Prosody Task and Prosody–Semantic Task, respectively.

For the stimuli with combined prosody and semantic cues, one could argue that listeners in our pilot group may have been using only prosody or semantic cues to recognise the intended emotion; as a further precaution, we therefore presented the utterances with an emotional semantic context a second time in written format to a new group of 20 English participants. All items selected for the present study were recognised for the intended happy, sad, or neutral emotion by a minimum of 90% of this group (in a 7-choice task). Finally, to ensure that utterance primes recorded for the Semantic Task were truly produced with a neutral prosody, these recordings were randomised in a final perceptual task with utterances that conveyed a broader range of emotional meanings through the prosody. The stimuli were then judged by 16 new young adults who were instructed to rate the positive–negative valence of each utterance based solely on the speaker's "tone of voice" (where − 2 was *very negative* sounding and + 2 was *very positive* sounding). All items chosen for the Semantic Task of this study obtained mean group valence ratings between − 0.5 and + 0.5 as determined by these pilot data, suggesting that the speaker's prosody was relatively neutral in tone for each of the selected utterances. Mean emotion recognition rates for the prime stimuli selected for each of our tasks are provided in Table 1, according to the respective validation procedure for each task.

(b) Face Targets

The target stimuli were colour photographs of an encoder's unobstructed facial expression, hair, and shoulders posed by three male and three female actors (Pell, 2002). Half of the face targets were expressions that represented a happy ($n = 12$) or sad ($n = 12$) emotion (YES trials), whereas half of the targets were facial grimaces posed by the same actors that did not represent an emotion ($n = 24$, NO trials). Here, the term "grimace" refers to any expression that involves movements of the brow, mouth, jaw, and lips and that does not lead to the recognition of an emotion based on our previous norming data (see below). For each of the six actors, two distinct exemplars of each "real" emotion and four unique grimaces were chosen for the study (all target items have been successfully used in our previous studies; see Table 2 for examples). All face stimuli were saved as bitmaps measuring 17.1 × 17.1 cm when presented on a computer screen. The exact same face targets were presented in each of the three tasks.

Target validation procedure. The face targets were also validated to establish their emotional status and meaning; all happy and sad facial expressions chosen for the study were correctly recognised by at least 78% of a group of 32 young adults in a forced-choice emotion-recognition study (Pell, 2002). Mean recognition rates for happy and sad face targets are also shown in Table 1. The same study established that all grimace expressions were not recognised as an emotion by at least 60% of the validation group (most typically, these expressions were described as "silly" by raters). Elsewhere, we have shown that the grimace expressions elicit distinct responses from "real" emotional expressions in three different ERP components during facial expression decoding (Paulmann & Pell, 2009). Thus, while it is likely true that many of the grimace faces possessed certain affective characteristics, there is strong evidence that these features do not symbolise discrete emotions, and that participants readily reject these exemplars as representing an emotion when instructed to make a facial affect decision (Pell, 2005a, 2005b; Pell & Skorup, 2008).

Experimental task design

For each task, individual prime stimuli (i.e., pseudo-utterances, neutrally intoned utterances with emotional semantics, or emotionally intoned utterances with congruent semantics) were paired with individual face targets to construct a series of

Table 1. *Mean and standard deviation recognition rates for all prime and face stimuli*

Stimulus type	*Emotion*	*Prosody task*[a]	*Semantic task*[b]	*Prosody–Semantic task*[c]
Prime sentence	Happy	0.88 (0.10)	0.98 (0.04)	0.97 (0.05)
	Sad	0.94 (0.07)	0.98 (0.05)	0.97 (0.03)
	Neutral	0.85 (0.07)	0.99 (0.03)	0.96 (0.03)
Face target[d]	Happy	0.99 (0.01)	0.99 (0.01)	0.99 (0.01)
	Sad	0.93 (0.06)	0.93 (0.06)	0.93 (0.06)

Notes: [a]Participants ($n = 24$) judged the emotion of auditory pseudo-sentences with emotional prosody but no emotional semantic cues. [b]Participants ($n = 20$) judged the emotion of visually presented sentences with an emotional semantic context but no prosodic cues. [c]Participants ($n = 20$) judged the emotion of auditory sentences with emotionally congruent semantic and prosodic cues. [d]Participants ($n = 32$) judged the emotion of visually presented emotional facial expressions.

144 trials. The same principles were applied to each task: each of the 12 happy, sad, and neutral prime utterances was paired separately with one exemplar of each YES target (i.e., one happy and one sad face) and two distinct exemplars of each NO target (grimace). This resulted in 72 YES trials ending in an emotional face (3 prime types × 2 face emotions × 12 items) and 72 NO trials ending in a grimace (3 prime types × 24 items). Prime–target pairings always involved a speaker/actor of the same sex (male–male, female–female). Once a specific prime item was paired with a particular YES face target, the same item was always matched with a grimace posed by the same actor. This matching process was first undertaken for the Prosody Task using the pseudo-utterances as primes; after all trials in this task were assigned correctly, the Semantic and Prosody–Semantic Tasks were constructed by simply substituting the appropriate prime utterance type produced by the same speaker, which had already been matched with face targets. Overall, each utterance prime appeared 4 times per task in unique combination with YES ($n = 2$) and NO ($n = 2$) face targets. Each face appeared 3 times in the experiment, once following a happy, sad, and neutral prime produced by the same speaker. The same prime and target stimulus was always assigned to different presentation blocks to minimise repetition effects in our data.

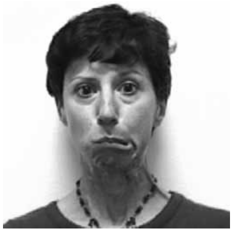
For the critical subset of YES trials, which ended in facial expressions of basic emotion (happy, sad), the emotional relationship of the prime–target pair could be defined in three ways: *congruent* trials were composed of primes and targets with the same emotional meaning (happy–happy or sad–sad, $n = 24$/task); *incongruent* trials consisted of prime–target pairs with conflicting emotions (happy–sad or sad–happy, $n = 24$/task); and *neutral* trials included a prime that was emotionally neutral in both its prosody and semantic cues (neutral–happy or neutral–sad, $n = 24$). It should be noted that neutral trials were included in the experiment primarily as filler items and to reduce the proportion of congruent trials in each task, which was 16.7% of all trials. This design was meant to limit strategic processing of emotional primes by listeners (Neely, Keefe, & Ross, 1989). A summary of the experimental task design, with examples of the prime and target stimuli, is furnished by Table 2.

Procedure

The experiment was administered in a quiet laboratory setting. Each participant completed all three tasks during two testing sessions with a one-week interval between each session. Two of the tasks were completed during the first session and the remaining task was completed during the second session. The presentation order for the three tasks was fully counterbalanced within the listener group to further avoid repetition or learning effects. Each task consisted of 144 experimental trials divided into six presentation blocks (24 trials/block). Block presentation order was also counterbalanced across participants. Trials were assigned in a quasi-random manner to blocks with the rule that each block should contain a relatively equal ratio of YES and NO trials, involving stimuli posed by a roughly equal ratio of female and male actors. In addition, the same prime or target stimulus was never assigned twice to the same block.

Experimental tasks were presented on a portable computer using Superlab presentation software (Cedrus, USA). The emotional prime utterances were played through high-quality stereo headphones with manual volume adjustment, and the facial targets were presented in the centre of the computer screen, viewed from a distance of approximately 45 cm. Participants were instructed that they would always hear a sentence before each face appeared, but that their goal was to focus strictly on the face to judge whether or not it represented an emotional expression (YES/NO response). Participants were encouraged to judge the face as accurately and quickly as possible by pushing one of two buttons on a Cedrus response pad. For the Prosody Task only, participants were informed that the (pseudo) sentences they would hear were

Table 2. *Summary of the experimental design and examples of the prime and target stimuli, by task*

	Prime features			*Target features*
Task	*Sentence type*	*Prosody*	*Semantics*	*Face type*
Prosody	*Pseudo-sentence*			*Emotional*
	They pannifered the moser. I tropped for swinty gowers. She kuvelled the noralind.	Happy Sad Neutral	Anomalous	
Semantic	*English sentence*			
	I had such a great weekend. My friends have all moved away. The filing cabinet is grey.	Neutral	Happy Sad Neutral	*Grimace* 
Prosody–Semantic	*English sentence*			
	I had such a great weekend. My friends have all moved away. The filing cabinet is grey.	Happy Sad Neutral	Happy Sad Neutral	

not supposed to make sense, and that they should focus their attention on the face.

Each trial was always composed of the following events: a visual fixation marker (350 ms); a 500 ms silent interval; the emotional prime utterance (presented at a comfortable listening level); and the face target presented immediately at the offset of the prime utterance. Accuracy and response latencies to the face were recorded by the computer. At the beginning of each task, two practice blocks ensured that participants understood the task and could render affect decisions about faces that did or did not convey an emotion, first in the absence of a preceding prime stimulus (Practice block 1) and then with a preceding prime resembling those in the corresponding task (Practice block 2). During the first practice block, which presented only the face targets, participants received feedback from the computer, first in relation to their accuracy (*Correct* or *Incorrect*) and then to their response speed (*Please respond faster*). After the second testing session, participants were compensated $30 CDN for their involvement.

RESULTS

The overall error rate across tasks for the 52 participants was 7.6% of all trials ($SD = 7.0$). As noted before (Pell, 2005a, 2005b), error tendencies in response to YES trials ($M = 9.5\%$, $SD = 12.5$) were greater overall than to NO trials ($M = 4.9\%$, $SD = 6.4$). This error-rate pattern showed little difference for the Prosody Task, Semantic Task, and Prosody–Semantic Task for either YES trials (10.9%, 10.3%, 9.1%) or NO trials (4.4%, 5.3%, 4.7%), respectively. These data exemplify that participants could accurately execute facial affect decisions (i.e., discriminate facial expressions that do or do not represent an emotion) and that participants adhered to task goals throughout the experiment. Exceptionally, one male participant demonstrated a high error rate that exceeded 33% across tasks (Prosody = 35.4%, Semantic = 33.3%, Prosody–Semantic = 33.3%); data for this participant were excluded from further analyses.

For the latency measures, only responses corresponding to *correct* facial affect decisions about real facial expressions of emotion (YES trials) were of interest for evaluating congruency priming effects in the study. To ensure that all participants contributed a reliable number of correct responses to each prime–target condition, following Pell (2005a), six of the remaining 51 participants (3 male, 3 female) who committed more than 25% errors to YES trials across the three tasks were removed from the latency analyses (mean YES errors across tasks for the six excluded participants = 39.7%, error range = 25.5–55.1%). Latency measures were then processed to normalise the individual subject distributions by eliminating extreme values and outliers. Latencies greater than 2000 ms and less than 300 ms were removed from the data for each task (0.2% of total values). At a second stage, individual latencies greater than two standard deviations from the conditional mean for a given participant were replaced with the value equal to two standard deviations in the appropriate direction (4.6% of total values). In total, all analyses of accuracy considered data for 51 participants and all analyses of latency measures considered data for 45 participants. To analyse accuracy and response times to YES trials across tasks, we conducted separate 3 × 2 × 3 analyses of variance (ANOVAs) with repeated measures on Task (Prosody Task, Semantic Task, Prosody–Semantic Task), Target (happy, sad) and Prime (happy, sad, neutral), which were elaborated by Tukey's HSD post hoc comparisons of significant main and interactive effects ($p < .05$). The mean accuracy and latencies to judge face targets preceded by a happy, sad, or neutral utterance in each of the three tasks are summarised in Table 3.

Accuracy

For accuracy, the 3 × 2 × 3 ANOVA produced main effects of Target, $F(1, 50) = 51.04$, $MSE = 0.11$, $p < .001$, and Prime, $F(2, 100) = 4.45$, $MSE = 0.01$, $p = .01$. In addition, there was a significant Target × Prime interaction, $F(2, 100) = 5.02$, $MSE = 0.04$, $p < .01$. Post hoc Tukey's HSD tests on the interaction demonstrated that, regardless of task, participants responded significantly more accurately to sad faces when they were preceded by a congruent sad prime (86%) than an emotionally incongruent happy prime (79%)—neutral primes (81%) did not differ from either happy or sad primes. By contrast, when judging happy faces, there was no difference in facial affect decision accuracy when these targets were preceded by happy (99%), sad (95%), and neutral (98%) prime utterances. Additionally, when the prime and target were congruent, participants responded with significantly greater accuracy in the happy–happy condition (99%) when compared to the sad–sad condition (86%). Although one of our hypotheses was that priming effects produced by different emotional speech cues could affect the magnitude of priming across the three tasks, there was no main effect of Task, nor any significant interactions involving Task, Target, and/or Prime type (all Fs < 2.82, ps $> .05$). The effect of pairing emotionally congruent versus incongruent prime–target events on error patterns in the three tasks is displayed in Figure 1a.

Although our primary focus was on the YES trials that ended in a meaningful emotional facial expression, we also analysed accuracy for NO trials that ended in a facial grimace that participants had to reject as a valid exemplar of a basic emotion. A 3 × 3 ANOVA was performed using the independent variables of Task (Prosody Task, Semantic Task, Prosody–Semantic Task) and Prime (happy, sad, neutral). This analysis yielded

Table 3. *Mean accuracy (% correct) and response latencies (milliseconds) as a function of task, prime emotion, and face target (standard deviations in parentheses;* ***bold figures*** *refer to emotionally congruent prime–target pairs)*

		Prosody task			*Semantic task*			*Prosody–Semantic task*		
		Prime emotion			*Prime emotion*			*Prime emotion*		
Measure	*Face target*	*Happy*	*Sad*	*Neutral*	*Happy*	*Sad*	*Neutral*	*Happy*	*Sad*	*Neutral*
Accuracy	Happy	**98.7 (6.1)**	96.1 (14.9)	98.5 (4.36)	**99.5 (2.0)**	95.6 (18.3)	98.0 (6.1)	**99.0 (4.0)**	94.6 (20.1)	97.7 (9.7)
	Sad	76.3 (25.2)	**84.8 (18.6)**	80.1 (24.6)	78.9 (27.6)	**83.7 (22.4)**	80.7 (27.5)	80.7 (24.2)	**88.6 (16.7)**	84.6 (22.2)
	Grimace	95.9 (8.1)	94.3 (9.2)	95.9 (7.5)	95.1 (8.1)	94.6 (7.3)	94.5 (8.5)	96.1 (6.6)	95.1 (8.4)	94.8 (9.2)
Latency	Happy	**556 (99)**	562 (92)	564 (104)	**525 (88)**	550 (104)	544 (98)	**527 (84)**	550 (93)	554 (98)
	Sad	714 (116)	**691 (103)**	707 (127)	659 (126)	**657 (126)**	663 (118)	685 (133)	**660 (100)**	668 (121)
	Grimace	630 (94)	647 (102)	636 (107)	613 (120)	628 (117)	609 (127)	594 (99)	606 (90)	615 (103)

no significant main effects of Task or Prime, and no interaction of these factors for NO trials (all Fs < 3.04, ps $> .05$).

Latencies

A similar $3 \times 2 \times 3$ (Task × Target × Prime) ANOVA was performed on the response times to correct YES target decisions to test for priming effects on this measure. This analysis yielded a significant main effect of Target, $F = 263.01$, $MSE = 13058$, $p < .001$, and Task, $F = 5.32$, $MSE = 14669$, $p < .01$. Post hoc Tukey's HSD test ($p < .05$) on the Target main effect revealed that participants responded significantly faster to happy faces (548 ms) compared to sad faces (678 ms), regardless of the Prime emotion and the Task. Post hoc inspection of the Task main effect demonstrated that irrespective of the Prime and Target emotion, participants rendered facial affect decisions significantly slower in the Prosody Task (632 ms) when compared to both the Semantic Task (600 ms) and the Prosody–Semantic Task (607 ms), which did not differ.

Additionally, there was a significant interaction of Prime and Target for response latencies, $F = 14.56$, $MSE = 1507$, $p < .001$. Post hoc Tukey's tests revealed the following significant effects: (1) when judging happy faces, participants were significantly faster after hearing a congruent happy prime (536 ms) than a sad (554 ms) or neutral prime (554 ms), regardless of the task; (2) when judging sad faces, participants were significantly faster after hearing a congruent sad utterance prime (669 ms) than a happy prime (686 ms), irrespective of the task—neither sad primes nor happy primes differed from neutral primes (679 ms); (3) Participants responded faster in the happy–happy congruent prime–target condition when compared to the sad–sad congruent prime–target condition, again regardless of the task (536 ms vs. 669 ms across tasks, respectively). Again, Task did not significantly interact with either Target or Prime when the latency data were examined (all Fs < 2.42, all ps $> .05$). The impact of prime–target emotional congruency on

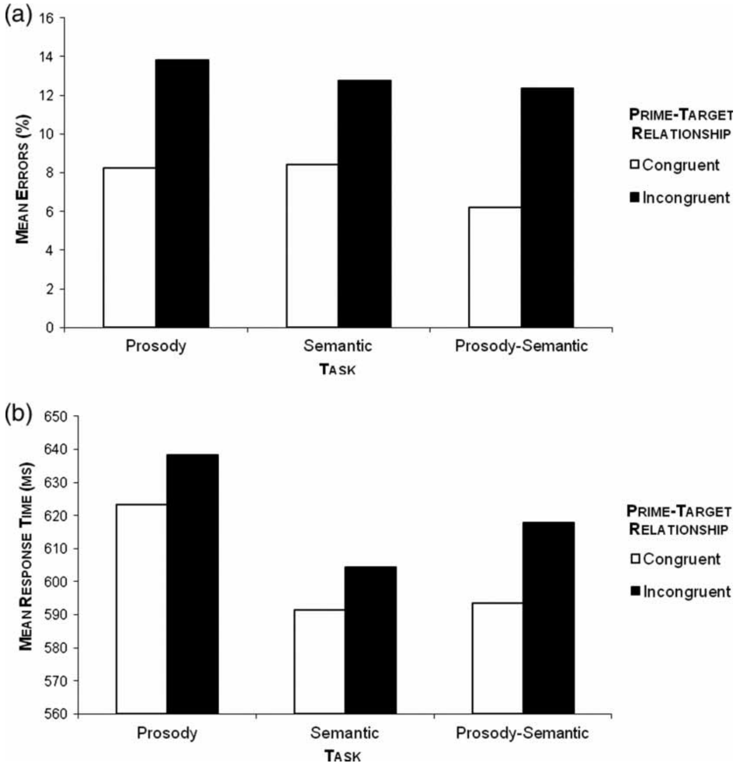


Figure 1. *Effects of emotion congruency on mean (a) error rates and (b) response latencies, as a function of task and prime–target relationship (emotionally congruent vs. incongruent).*

response times in each task is illustrated in Figure 1b.

Finally, a 3×3 (Task × Prime) ANOVA was conducted on the response times to correct NO trial responses. This analysis revealed a significant main effect of Prime, $F = 7.89$, $MSE = 906$, $p < .001$, and Task, $F = 4.46$, $MSE = 8254$, $p = .01$. Furthermore, there was a significant Task × Prime interaction, $F = 4.48$, $MSE = 609$, $p < .01$. Post hoc Tukey's HSD tests revealed the following patterns of interest: (1) no systematic effects in judging grimaces as a function of the emotion of the prime utterance; (2) regardless of the prime emotion, responses to grimaces were always significantly slowest in the Prosody Task, compared to both the Semantic Task and Prosody–Semantic Task; (3) for happy and sad (but not neutral) primes, responses were significantly slower in the Semantic Task when compared to the Prosody–Semantic Task.

DISCUSSION

This study investigated the nature of emotional speech processing when prosody, semantic cues, and combined prosodic and semantic cues were processed in a relatively implicit manner, based on evidence of cross-modal priming of a conjoined face. Our results uncovered a robust emotion-congruency effect in each of our three tasks: facial affect decisions were significantly faster overall when prosodic, semantic, or prosodic–semantic primes were emotionally *congruent* with the target face (as well as more accurate overall when sad primes and targets were paired). Thus, even when participants were instructed to focus on the target face, and when expectations about the prime–target relationship were mitigated in the experiment, our data establish that the happy or sad meaning encoded by prosody, semantics, or both cues in tandem was implicitly registered in memory by listeners, influencing subsequent decisions about the representational status of facial expressions in an emotion-congruent manner. When viewed generally, these findings fit with previous claims that emotional speech cues are processed implicitly and automatically by listeners (de Gelder, Bocker, Tuomainen, Hensen, & Vroomen, 1999; Pourtois et al., 2000; Vroomen, Driver, & de Gelder, 2001), irrespective of whether emotions are encoded by prosody or semantic cues.

In the Prosody Task, the observation that emotional prosody alone successfully primes decisions about a related target face replicates our previous work in which pseudo-utterances conveying happiness, sadness, and/or anger and surprise were presented (Pell, 2005a, 2005b; Pell & Skorup, 2008). Researchers have argued that emotional prosody is encoded in memory in the form of expression "prototypes" that refer to basic emotions (e.g., Laukka, 2005; Pell et al., 2009a), and which share underlying features with corresponding facial expressions of emotion (de Gelder et al., 1999; de Gelder & Vroomen, 2000; Massaro & Egan, 1996). Our results add further to these claims by demonstrating that prosodic features of semantically anomalous utterances influence emotional face processing in a systematic and emotion-specific manner, as has been shown in similar research (de Gelder & Vroomen, 2000; Massaro & Egan, 1996). Importantly, parallel results in the Semantic Task extend previous data by showing that emotional semantic cues expressing happiness or sadness also prime judgements of facial expressions in an emotion-congruent manner, independent of prosody. While there is a wide array of studies that have elicited priming effects using visually displayed words with emotional semantics (e.g., Innes-Ker & Niedenthal, 2002; Preston & Stansfield, 2008; Zhang, Guo, Lawson, & Jiang, 2006), this is one of the first to show that spoken sentences with an emotional semantic context systematically prime emotional face processing in a category-specific manner. Since our findings indicate robust cross-modal priming effects in each of our three tasks, it can be argued that critical aspects of emotional information processing are shared among the prosodic, semantic, and facial channels (Borod et al., 2000; Bowers, Bauer, & Heilman, 1993).

As raised earlier, a methodological constraint of our design is that the effects of semantics in spoken language could not be evaluated in the complete absence of prosody, even when speakers adopt a neutral or non-emotional tone of voice (as was the case in our Semantic Task; see also Vingerhoets, Berckmoes, & Stroobant, 2003). In fact, combining happy or sad emotional semantics with neutral prosody could very well present a conflict situation to listeners, if the neutral tone of the utterance is perceived as incompatible with the emotional interpretation of the semantics, introducing a potential confound in the Semantic Task. However, we observed a reliable emotion-congruency effect independent of Task, and all other performance measures were similar in these two conditions. This implies that neutral prosody did not meaningfully interfere with the processing of emotional semantic meanings, which dictated the congruency effect in the Semantic task. To some extent, this result could partially confirm the hypothesis that semantic information overrides prosody when these cues are processed concurrently and conflict (Astesano, Besson, & Alter, 2004; Kotz & Paulmann, 2007; Paulmann & Kotz, 2008), or at least that prosody can be ignored when it conflicts with the semantic message and is neutral or unemotional in nature.

The design of the present experiment included one positively valenced emotion (happy) and one negatively valenced emotion (sad). While not the primary focus of this study, this design permits some exploration of valence effects in our data. If one examines the impact of presenting different facial expressions on our measures, it is obvious that the overall accuracy and speed of behavioural responses was markedly better when participants judged happy versus sad face targets, in each of the three tasks. This pattern has been shown repeatedly in the FADT (Pell, 2005a, 2005b; Pell & Skorup, 2008) and other implicit emotion processing tasks (Rossell & Nobre, 2004; Wurm et al., 2001), and can be explained by the underlying positive–negative valence of the face targets. These tendencies are likely mediated by an automatic attention bias that temporarily interrupts the cognitive analysis of negative events such as faces (Dimberg & Öhman, 1996; Mogg & Bradley, 1999), which were the focus of conscious attention in our experiment (see Pell, 2005b, for a detailed discussion).

The fact that happy faces could be judged more accurately and quickly than sad faces in our various task conditions could explain why there was no significant priming in accuracy for happy prime–target pairs; for accuracy measures, these faces are associated with a near ceiling effect (99% for happy primes, 95% for sad, 98% for neutral), which permitted little sensitivity to effects of a prime stimulus when accuracy rates are analysed (see Leppänen & Hietanen, 2004, for a discussion of the "happy face advantage" in behavioural performance). In the latency data, we did observe subtle differences in how happy, sad, and neutral primes influenced facial affect decisions; for example, neutral primes led to significantly slower responses than happy primes when participants judged happy face targets, whereas neutral primes did not differ from sad or happy primes in their effect on sad face targets. However, despite the variable effects of presenting neutral utterances as primes, our data evince a clear emotion congruency effect that was present in the majority of conditions, independent of stimulus valence or other task variables. As such, the remaining discussion can proceed largely independent of knowledge of how the valence of primes and face targets influenced performance measures in our data.

On the relative contributions of emotional prosody and semantics

By presenting utterance primes with prosody, semantic cues, or combined prosody and semantic emotional cues, one of our main objectives was to compare the *extent* of priming observed among our three tasks. Based on the literature, we hypothesised that semantic cues would be stronger, or somehow more salient to listeners, than prosodic cues (e.g., Breitenstein et al., 1998; Ishii et al., 2003; Johnson et al., 1986; Paulmann & Kotz, 2008; Pell & Baum, 1997), yielding greater priming in the two semantic-related tasks. In contrast, our data show that response times to emotional faces were equally facilitated by congruent prosodic or semantic cues; following cross-task comparisons, we found no evidence that the priming effect was stronger in the presence of emotional semantics versus prosodic information. Rather, it appears that the prosody and semantic channels activate emotional category meanings such as happiness or sadness in a similar manner, and that this information leads to comparable levels of facilitation of a congruent emotional face, at least when assessed through a priming paradigm as was presented here.

Our results do demonstrate that semantic information affected certain responses, as evidenced by a significant Task main effect in our overall analysis. Regardless of whether prime–target pairs were congruent or incongruent, the two tasks that presented meaningful semantic cues (i.e., Semantic and Prosody–Semantic tasks) were associated with significantly faster facial affect decision response times overall, when compared to the Prosody Task. This pattern suggests that primes with meaningful semantic information may have facilitated facial affect decisions, irrespective of the congruency of the prime–target events. However, it is more plausible that response times in the Prosody Task were generally slowed or inhibited by the presence of semantically anomalous pseudo-utterances. Despite the fact that pseudo-utterances are routinely used to evaluate the effects of emotional prosody independent of semantics (e.g., Paulmann & Kotz, 2008; Scherer et al., 1991), these stimuli likely increase cognitive demands at the stage of lexical-semantic and grammatical encoding (Pell et al., 2009a; Pell & Skorup, 2008; Shuster, 2009). This interference would tend to slow overall response times in the Prosody Task versus the two semantic-related tasks, as observed, although in a way that is independent of emotional meaning resolution from the pseudo-utterances or the semantically well-formed utterances. Further evidence for this explanation can be seen in the analysis of "NO" trials, where response times were slowest in the Prosody Task, even when the target was a facial grimace unrelated to the emotional prime.

Regardless, it must be emphasised that the extent of cross-modal priming produced by prosody, semantics, and combined prosodic-semantic stimuli was highly comparable across tasks. Unfortunately, our methods do not furnish details on the time-course for activating emotional knowledge encoded by prosody versus semantic cues, which may well have differed between the two speech channels. Research indicates that emotional meanings of prosody are encoded continuously in the speech signal and are detected very quickly by listeners, perhaps within 200–400 ms after speech onset (Paulmann & Pell, 2010). On the other hand, emotional semantic meanings are determined locally according to the linguistic context and prevailing language structure (e.g., Eckstein & Friederici, 2006); this allows the possibility that emotional semantic meanings are activated at varying time points in the utterance, whereas emotional prosody is detected earlier (see Kotz & Paulmann, 2007, for a discussion). Because facial affect decisions in our study were always executed *after* the end of the prime sentence, it is possible that emotional meanings activated by prosody versus semantic cues were present at different time points, and to varying degrees, although they yielded equal priming of a face when measured at the sentence offset. In light of the unique, dynamic interplay of prosodic and semantic cues in speech, future studies could better control the linguistic structure of prime sentences to test whether meaning activations produced by prosody versus semantic cues vary according to the location of emotional cues within the utterance.

The effect of combined emotional prosody and emotional semantics

Another question addressed by our study was whether combined prosodic-semantic cues facilitate emotional processing beyond that observed by prosody or semantic cues alone. Some research implies that congruent prosody and semantic information interacts in an additive manner, allowing the stimulus to be recognised more quickly and accurately due to increased salience of its meaning (Beaucousin et al., 2007; Nygaard & Queen, 2008; Wambacq & Jerger, 2004). In the present study, responses to grimaces in the "NO" trials were indeed fastest when preceded by a prime that contained both prosodic and semantic information. However, as argued above, our priming results ("YES" trials) demonstrated little evidence of an additive mechanism, since congruency effects on facial decision latencies were no different when primed by prosodic-semantic cues than by prosody or semantics alone. Thus, it can be said that the simultaneous presence of congruent prosody and semantic information did not yield a detectable advantage in emotional processing of a conjoined emotional face stimulus, when compared to the processing of either prosody or semantics alone.

These conclusions appear to contrast with studies that have measured *explicit* recognition of emotional displays, using identification or discrimination tasks (Beaucousin et al., 2007; Morais & Ladavas, 1987; Pell, 2006; Techentin et al., 2009; Wambacq & Jerger, 2004). In these investigations, there is strong evidence that a presumably "enriched" speech stimulus containing both emotional prosody and semantic information promotes increased confidence, and quicker/more accurate behavioural responses, about the emotion conveyed. However, note that in the FADT and many other tasks of relatively *implicit* emotion processing, participants are not required to consciously retrieve or apply verbal labels referring to either the prime or target emotion (Pell, 2005a). One can therefore speculate that when emotional priming effects are measured, there is an activation threshold beyond which the addition of congruent prosody or semantic information does not yield detectable benefits in behavioural performance. However, facilitation produced by a multimodal versus unimodal stimulus may be detected at very early stages of emotional processing when investigated using ERPs (see Paulmann, Jessen, & Kotz, 2009, for recent data); also, an additive advantage of processing two cue sources in tandem often becomes apparent in many explicit tasks that require emotion categorisation, or other forms of conscious attention to emotional details registered in memory. Situations in which emotional cues conflict in their meaning, or when one speech channel is ambiguous, could also serve to direct attentional resources to emotion-related knowledge, leading to differential effects of prosody versus semantic cues, even when the stimuli are first processed implicitly (Massaro & Egan, 1996; Nygaard & Lunders, 2002).

On the representation and processing of emotion across channels

Based on the priming data reported here and elsewhere (Schirmer et al., 2002, 2005; Wurm et al., 2001), and a growing fMRI literature on the audio-visual integration of emotional displays (e.g., Johnstone, van Reekum, Oakes, & Davidson, 2006; Kreifelts, Ethofer, Shiozawa, Grodd, & Wildgruber, 2009), it seems that prosodic, semantic, and facial expressions of emotion share underlying conceptual knowledge and/or neurocognitive mechanisms, which promote cross-modal interactions during emotional information processing (Carroll & Young, 2005). As such, one can speculate on how emotional concepts may be organised in associative memory. Much research suggests that basic emotions are represented categorically as units in memory (e.g., Bower, 1981; Hansen & Shantz, 1995; Niedenthal et al., 1997, 1999), although aspects of a given emotion category are likely represented in diverse brain regions as a function of different information-processing channels used to access these details (Mahon & Caramazza, 2008, 2009). As sensorimotor features are believed critical to the organisation and processing of conceptual knowledge, each emotion concept can be accessed in a partially distinct manner via associated auditory or visual cues (Mahon & Caramazza, 2008). For example, within the concept of "happiness" may exist highly correlated cross-modal features such as the percept of a smiling face, a voice characterised by an elevated and expanded pitch range, and words such as "great", "delightful", or "lucky". Although encoded in distinct communication channels, these cues would be mentally grouped together because they are associated with and/or evoke feelings of happiness (Niedenthal et al., 1999), whether through innate mechanisms and/or learned experience.Even if participants are not required to consciously attend to or name the emotion conveyed, an implicitly processed happy prosody, or semantic cues associated with happiness, would presumably activate the concept of happiness in emotional memory. Following Scherer (1986) and others, we assume that basic emotional meanings can be the end product of a series of automatic stimulus appraisals which quickly evaluate the valence, potency, relevance, etc., of the stimulus before semantic meanings are assigned. As observed in our study, the subsequent presentation of an emotion-congruent cue in a *different* channel (i.e., a smiling face) would be rapidly integrated and facilitate judgements of a face target as an instance of the same emotion, because the concept of happiness is already pre-activated in memory. This process is similar to a spreading activation, network model of memory in which emotions are represented as "nodes" associatively linked to related sensory events and information (e.g., Bower, 1981, 1987). Activation spreads along these links when a given emotional category is activated, facilitating access to or "priming" conceptually related cues. Instances in which the two cues refer to mismatching categories (e.g., a happy prosody and an angry face) may lead to more effortful processing in many communicative settings, until a coherent or "pragmatically appropriate" emotional meaning is determined from the broader situational context. In some cases, cross-modal discrepancies registered in emotional memory act as a principle cue for interpreting humour, irony, and sarcastic intent (Attardo, Eisterhold, Hay, & Poggi, 2003; Cheang & Pell, 2008).

Since our data suggest that prosody and semantic information reliably activate conceptual knowledge about emotions, although largely to the same extent, it appears that listening to short utterances which contain unambiguous prosody or semantic cues about a speaker's emotion is sufficient to activate corresponding emotion concepts; encountering multiple sources of congruent information does not serve to strengthen the unit's activation in associative memory when the automatic effects of this knowledge on a related stimulus are measured in the manner conducted here. However, as argued earlier, this does not mean that the strength of activations within the emotion network does not fluctuate according to temporal properties of an utterance (Kotz & Paulmann, 2007), especially at early stages of emotional salience detection (Paulmann et al., 2009), and/or for stimuli which promote greater emotional ambiguity than the items we presented here (Massaro & Egan, 1996). Future research could address these issues in an informative manner to elaborate how emotional meanings are registered and change over the course of an utterance and in extended discourse.

Conclusions

Our results lend credence to the notion that emotional prosody, emotional faces, and emotional semantic cues engage an overlapping neurocognitive processing system, and that representative features encoded in each of these communication channels have access to shared conceptual information stored in memory (Borod et al., 2000; Bowers et al., 1993; Kreifelts et al., 2009). Although we restricted our scope to two relatively uncontroversial emotions (happiness and sadness), it is possible that the strength of emotional cues inherently present in each speech channel varies as a function of the emotion type, and future research will benefit by investigating a larger set of emotions (e.g., anger, fear, disgust, surprise). At a methodological level, our study supports the use of the FADT as a means for indexing cues in the lexical-semantic channel, as well as the prosodic channel, for documenting cross-modal effects during emotional processing. Future studies should address the implication of cultural factors on emotional processing, as accumulating evidence suggests that the relative weight given to prosody or semantics depends on a listener's cultural orientation (Ishii et al., 2003, Kitayama & Ishii, 2002). New studies which provide psychophysiological (Kotz & Paulmann, 2007; Paulmann & Kotz, 2008) and neuroimaging (Kreifelts, Ethofer, Grodd, Erb, & Wildgruber, 2007; Kreifelts et al., 2009) data on how the brain integrates emotional information across channels will also continue to be highly informative. Converging findings from these diverse investigative domains will bring us closer to understanding how humans integrate and interact with words, voices, and faces when processing socially relevant emotional cues.

Manuscript received 1 December 2009
Revised manuscript received 10 June 2010
Manuscript accepted 15 June 2010
First published online 10 November 2010

REFERENCES

Astesano, C., Besson, M., & Alter, K. (2004). Brain potentials during semantic and prosodic processing in French. *Cognitive Brain Research*, *18*, 172–184.

Attardo, S., Eisterhold, J., Hay, J., & Poggi, I. (2003). Multimodal markers of irony and sarcasm. *Humor—International Journal of Humor Research*, *16*(2), 243–260.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*(3), 614–636.

Beaucousin, V., Lacheret, A., Turbelin, M.-R., Morel, M. L., Mazoyer, B., & Tzourio-Mazoyer, N. (2007). fMRI study of emotional speech comprehension. *Cerebral Cortex*, *17*(2), 339–352.

Borod, J. C., Cicero, B. A., Obler, L. K., Welkowitz, J., Erhan, H. M., Santschi, C., et al. (1998). Right hemisphere emotional perception: Evidence across multiple channels. *Neuropsychology*, *12*(3), 446–458.

Borod, J. C., Pick, L. H., Hall, S., Sliwinski, M., Madigan, N., Obler, L. K., et al. (2000). Relationships among facial, prosodic, and lexical channels of emotional perceptual processing. *Cognition and Emotion*, *14*(2), 193–211.

Bower, G. H. (1981). Mood and memory. *American Psychologist*, *36*(2), 129–148.

Bower, G. H. (1987). Commentary on mood and memory. *Behaviour Research and Therapy*, *25*(6), 443–455.

Bowers, D., Bauer, R., & Heilman, K. (1993). The nonverbal affect lexicon: Theoretical perspectives from neuropsychological studies of affect perception. *Neuropsychology*, *7*, 433–444.

Bowers, D., Coslett, H. B., Bauer, R. M., Speedie, L. J., & Heilman, K. M. (1987). Comprehension of emotional prosody following unilateral hemispheric lesions: Processing defect versus distraction defect. *Neuropsychologia*, *25*(2), 317–328.

Breitenstein, C., Daum, I., & Ackermann, H. (1998). Emotional processing following cortical and subcortical brain damage: Contribution of the frontostriatal circuitry. *Behavioural Neurology*, *11*, 29–42.

Buchanan, T. W., Lutz, K., Mirzazade, S., Specht, K., Jon Shah, N., Zilles, K., et al. (2000). Recognition of emotional prosody and verbal components of spoken language: An fMRI study. *Cognitive Brain Research*, *9*, 227–238.

Caffi, C., & Janney, R. W. (1994). Toward a pragmatics of emotive communication. *Journal of Pragmatics*, *22*, 325–373.

Carroll, N. C., & Young, A. W. (2005). Priming of emotion recognition. *The Quarterly Journal of Experimental Psychology*, *58A*(7), 1173–1197.

Cheang, H. S., & Pell, M. D. (2008). The sound of sarcasm. *Speech Communication*, *50*, 366–381.

Dara, C., Monetta, L., & Pell, M. D. (2008). Vocal emotion processing in Parkinson's disease: Reduced sensitivity to negative emotions. *Brain Research*, *1188*, 100–111.

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. *Neuroscience Letters*, *260*, 133–136.

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

Dimberg, U., & Öhman, A. (1996). Beyond the wrath: Psychophysiological responses to facial stimuli. *Motivation and Emotion*, *20*(2), 149–182.

Eckstein, K., & Friederici, A. D. (2006). It's early: Event-related potential evidence for initial interaction of syntax and prosody in speech comprehension. *Journal of Cognitive Neuroscience*, *18*(10), 1696–1711.

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

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

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, *6*(2), 78–84.

Grandjean, D., Banziger, T., & Scherer, K. R. (2006). Intonation as an interface between language and affect. *Progress in Brain Research*, *156*, 235–268.

Grimshaw, G. M. (1998). Integration and interference in the cerebral hemispheres: Relations with hemispheric specialization. *Brain and Cognition*, *36*(2), 108–127.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*, 438–441.

Hansen, C., & Shantz, C. (1995). Emotion-specific priming: Congruence effects on affect and recognition across negative emotions. *Personality and Social Psychology Bulletin*, *21*, 548–557.

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. *European Journal of Cognitive Psychology*, *16*(6), 769–790.

Hsu, S.-M., Hetrick, W. P., & Pessoa, L. (2008). Depth of facial expression processing depends on stimulus visibility: Behavioral and electrophysiological evidence of priming effects. *Cognitive, Affective Behavioral Neuroscience*, *8*(3), 282–292.

Innes-Ker, A., & Niedenthal, P. (2002). Emotion concepts and emotional states in social judgment and categorization. *Journal of Personality and Social Psychology*, *83*, 804–816.

Ishii, K., Reyes, J. A., & Kitayama, S. (2003). Spontaneous attention to word content versus emotional tone: Differences among three cultures. *Psychological Science*, *14*(1), 39–46.

Johnson, W. F., Emde, R. N., Scherer, K. R., & Klinnert, M. D. (1986). Recognition of emotion from vocal cues. *Archives of General Psychiatry*, *43*, 280–283.

Johnstone, T., van Reekum, C., Oakes, T., & Davidson, R. J. (2006). The voice of emotion: An fMRI study of neural responses to angry and happy vocal expressions. *Social, Cognitive, and Affective Neuroscience*, *1*, 242–249.

Kitayama, S., & Howard, S. (1994). Affective regulation of perception and comprehension: Amplification and semantic priming. In P. M. Niedenthal & S. Kitayama (Eds.), *The heart's eye: Emotional influences in perception and attention* (pp. 41–65). San Diego, CA: Academic Press.

Kitayama, S., & Ishii, K. (2002). Word and voice: Spontaneous attention to emotional utterances in two languages. *Cognition and Emotion*, *16*(1), 29–59.

Kotz, S. A., & Paulmann, S. (2007). When emotional prosody and semantics dance cheek to cheek: ERP evidence. *Brain Research*, *115*, 107–118.

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.

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

Leppänen, J. M., & Hietanen, J. K. (2004). Positive facial expressions are recognized faster than negative facial expressions, but why? *Psychological Research*, *69*, 22–29.

Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology – Paris*, *102*, 59–70.

Mahon, B. Z., & Caramazza, A. (2009). Concepts and categories: A cognitive neuropsychological perspective. *Annual Review of Psychology*, *60*, 27–51.

Massaro, D., & Egan, P. (1996). Perceiving affect from the voice and the face. *Psychonomic Bulletin and Review*, *3*(2), 215–221.

Mitchell, R. L. C., Elliott, R., Barry, M., Cruttenden, A., & Woodruff, P. W. R. (2003). The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. *Neuropsychologia*, *41*(10), 1410–1421.

Mogg, K., & Bradley, B. P. (1999). Orienting of attention to threatening facial expressions presented under conditions of restricted awareness. *Cognition and Emotion*, *13*(6), 713–740.

Morais, J., & Ladavas, E. (1987). Hemispheric interactions in the recognition of words and emotional intonations. *Cognition and Emotion*, *1*(1), 89–100.

Morton, J., & Trehub, S. (2001). Children's understanding of emotion in speech. *Child Development*, *72*, 834–843.

Neely, J. H., Keefe, D. E., & Ross, K. L. (1989). Semantic priming in the lexical decision task: Roles of prospective prime-generated expectancies and retrospective semantic matching. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *15*(6), 1003–1019.

Niedenthal, P. M. (1990). Implicit perception of affective information. *Journal of Experimental Social Psychology*, *26*, 505–527.

Niedenthal, P. M., Halberstadt, J. B., & Innes-Ker, A. H. (1999). Emotional response categorization. *Psychological Review*, *106*(2), 337–361.

Niedenthal, P. M., Halberstadt, J. B., & Setterlund, M. B. (1997). Being happy and seeing "happy": Emotional state mediates visual word recognition. *Cognition and Emotion*, *11*(4), 403–432.

Nygaard, L., & Lunders, E. (2002). Resolution of lexical ambiguity by emotional tone of voice. *Memory & Cognition*, *30*(4), 583–593.

Nygaard, L. C., & Queen, J. S. (2008). Communicating emotion: Linking affective prosody and word meaning. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(4), 1017–1030.

Paulmann, S., Jessen, S., & Kotz, S. A. (2009). Investigating the multimodal nature of human communication: Insights from ERPs. *Journal of Psychophysiology*, *23*(2), 63–76.

Paulmann, S., & Kotz, S. A. (2008). An ERP investigation on the temporal dynamics of emotional prosody and emotional semantics in pseudo- and lexical-sentence context. *Brain and Language*, *105*, 59–69.

Paulmann, S., & Pell, M. D. (2009). Decoding facial expressions 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? *Cognitive Affective and Behavioral Neuroscience*, *10*(2), 230–242.

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

Pell, M. D. (2005a). Nonverbal emotion priming: evidence from the "facial affect decision task". *Journal of Nonverbal Behavior*, *29*(1), 45–73.

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

Pell, M. D. (2006). Cerebral mechanisms for understanding emotional prosody in speech. *Brain and Language*, *96*(2), 221–234.

Pell, M. D., & Baum, S. R. (1997). The ability to perceive and comprehend intonation in linguistic and affective contexts by brain-damaged adults. *Brain and Language*, *57*(1), 80–99.

Pell, M. D., Monetta, L., Paulmann, S., & Kotz, S. A. (2009a). Recognizing emotions in a foreign language. *Journal of Nonverbal Behaviour*, *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. *Journal of Phonetics*, *37*, 417–435.

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

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.

Preston, S., & Stansfield, R. B. (2008). I know how you feel: Task-irrelevant facial expressions are spontaneously processed at a semantic level. *Cognitive, Affective Behavioral Neuroscience*, *8*, 54–64.

Rossell, S. L., & Nobre, A. (2004). Semantic priming of different affective categories. *Emotion*, *4*(4), 354–363.

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, *99*(2), 143–165.

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

Schirmer, A., & Kotz, S. A. (2003). ERP evidence for a sex-specific Stroop effect in emotional speech. *Journal of Cognitive Neuroscience*, *15*(8), 1135–1148.

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, *10*, 24–30.

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.

Shuster, L. I. (2009). The effect of sublexical and lexical frequency on speech production: An fMRI investigation. *Brain & Language*, *111*(1), 66–72.

Tanenhaus, M. K., & Brown-Schmidt, S. (2008). Language processing in the natural world. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*, 1105–1122.

Techentin, C., Voyer, D., & Klein, R. M. (2009). Between- and within-ear congruency and laterality effects in an auditory semantic/emotional prosody conflict task. *Brain and Cognition*, *70*, 201–208.

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

Vingerhoets, G., Berckmoes, C., & Stroobant, N. (2003). Cerebral hemodynamics during discrimination of prosodic and semantic emotion in speech studied by transcranial Doppler ultrasonography. *Neuropsychology*, *17*(1), 93–99.

Vroomen, J., Driver, J., & de Gelder, B. (2001). Is cross-modal integration of emotional expressions independent of attentional resources? *Cognitive, Affective Behavioral Neuroscience*, *1*, 382–387.

Wambacq, I. J. A., & Jerger, J. F. (2004). Processing of affective prosody and lexical-semantics in spoken utterances as differentiated by event-related potentials. *Cognitive Brain Research*, *20*(3), 427–437.

Wilson, D., & Wharton, T. (2006). Relevance and prosody. *Journal of Pragmatics*, *38*, 1559–1579.

Wurm, L. H., Vakoch, D. A., Strasser, M. R., Calin-Jageman, R., & Ross, S. E. (2001). Speech perception and vocal expression of emotion. *Cognition and Emotion*, *15*(6), 831–852.