PROSODY–FACE INTERACTIONS IN EMOTIONAL PROCESSING AS REVEALED BY THE FACIAL AFFECT DECISION TASK

Marc D. Pell

ABSTRACT: Previous research employing the facial affect decision task (FADT) indicates that when listeners are exposed to semantically anomalous utterances produced in different emotional tones (*prosody)*, the emotional meaning of the prosody primes decisions about an emotionally *congruent* rather than incongruent facial expression (Pell, M. D., *Journal of Nonverbal Behavior, 29*, 45–73). This study undertook further development of the FADT by investigating the approximate *timecourse* of prosody–face interactions in nonverbal emotion processing. Participants executed facial affect decisions about *happy* and *sad* face targets after listening to utterance fragments produced in an emotionally related, unrelated, or *neutral* prosody, cut to 300, 600, or 1000 ms in duration. Results underscored that prosodic information enduring at least 600 ms was necessary to presumably activate shared emotion knowledge responsible for prosody–face congruity effects.

KEY WORDS: emotion processing; vocal expression; facial expression; speech communication; affective priming; social cognition.

Nonverbal displays furnish potent cues which guide human interactions and social-interpretive processes applied to these events. In particular, dynamic changes in speech *prosody* (or "voice tone") and facial movements are of central importance to understanding the emotions of communication partners. During most interpersonal events, the emotional content of prosody and of facial expressions is intercepted concurrently and tends to *cohere* in its intended value across channels, although cross-channel cues may be purposefully manipulated to conflict in service of communicative goals such as humor or sarcasm (Attardo, Eisterhold, Hay, & Poggi, 2003). The fact that prosodic and facial expressions of emotion frequently correlate in their underlying value renders the possibility that cognitive

Marc D. Pell is affiliated with the School of Communication Sciences and Disorders, McGill University, Montréal, Canada.

Address correspondence to Marc D. Pell, School of Communication Sciences and Disorders, McGill University, 1266 ave. des Pins ouest, Montréal, Que., Canada H3G 1A8, Canada; e-mail: marc.pell@mcgill.ca.

DOI: 10.1007/s10919-005-7720-z

mechanisms which assign significance to these events are highly sensitive to shared and/or discrepant associations activated by cues in each channel (de Gelder, Bocker, Tuomainen, Hensen, & Vroomen, 1999; Massaro & Egan, 1996; Pell, 2005). Moreover, prosody and facial expressions may share a common processing and/or representational substrate in the human brain (Borod, 2000; Bowers, Bower, & Heilman, 1993; Pourtois, Debatisse, Despland, & de Gelder, 2002), facilitating integration and ultimate comprehension of these distinct but calibrated sources of information.

The present investigation was motivated by an interest in how emotional attributes of speech prosody are implicitly processed during on-line speech perception, by referring to their effects on judgements about a conjoined static facial expression of emotion. There is mounting evidence that affective or emotional features of a perceptual event impact systematically on the processing of an adjoined stimulus that is related or unrelated with respect to one of the critical activated components; these congruity-based effects in information processing have been referred to as *affective* or *emotion priming* and are detectable under a variety of conditions (see Klauer & Musch, 2003 for a review). Investigations of emotion-specific bias on perceptual judgements imply the existence of a set of 'emotion concepts' in perceptual memory, which may be brought to bear on many forms of input (Bower, 1981; Niedenthal & Halberstadt, 1999). This knowledge, which has been shown to organize memory and influence categorization of stimuli in favor of specific emotion states such as *happy*, *sad*, or other 'basic' emotions (Halberstadt & Niedenthal, 2001; Innes-Ker & Niedenthal, 2002; Niedenthal & Halberstadt, 1995; Niedenthal, Halberstadt, & Setterlund, 1997), provides a mechanism for understanding emotion-congruent bias between communicative stimuli related by underlying emotion (Hansen & Shantz, 1995; Pell, 2005; Wurm, Vakoch, Strasser, Calin-Jageman, & Ross, 2001).

As biologically salient channels for communicating emotion, prosody and facial expressions may be especially prone to activate conceptual or 'prototypic' knowledge pertaining to discrete emotions (Etcoff & Magee, 1992; Young et al., 1997). In line with affective congruity effects in information processing, activation of this emotion-based knowledge holds the potential to influence cognitive operations applied to prosody or a conjoined face when the emotional processing system is engaged by these stimuli. A small but valuable group of studies has begun to qualify some of the operational influences of emotional prosody on an emotional face, or vice versa (de Gelder & Vroomen, 2000; Massaro & Egan, 1996). For example, Massaro and Egan presented computer-generated faces expressing a *happy, angry*, or *neutral* emotion to 15 undergraduate students, accompanied by the word ''please'' spoken by a male actor in one of the three emotional tones. Subjects identified the emotion of the paired event as ''happy'' or ''angry'' without instruction as to which stimulus they should attend to in making these judgements. The authors noted that emotional attributes of the face had a much larger effect on judgements than the prosody, although information from both modalities was influential; the impact of prosody on emotional decisions appeared to increase as defining features of the (dominant) face stimulus became more ambiguous (Massaro & Egan, 1996). These data argue that underlying emotional features of prosody, despite being somewhat impoverished (i.e., the single word *please*), were nonetheless processed and integrated to some extent with a concurrent face, influencing how decoders characterized the emotional event.

Work by de Gelder and colleagues (de Gelder & Vroomen, 2000; de Gelder et al., 1999; Pourtois, de Gelder, Vroomen, Rossion, & Crommelinck, 2000; Pourtois et al., 2002; Vroomen, Driver, & de Gelder, 2001) has greatly reinforced the idea that the emotional value of prosody and a concurrent face stimulus are combined across channels, with measurable consequences on evaluative judgements when either the face or the prosody serves as the target stimulus. Using behavioral and electrophysiological approaches, these studies have shown that sentences spoken in an emotionally congruent prosody facilitate evaluative (*happy/sad*) decisions and related responses tied to an emotionally ambiguous, ''morphed'' facial expression. Their data further argue that comparing the emotional components of prosody–face events occurs independent of the attentional focus of the task or its general cognitive load (i.e., in a dual task environment), implying that combining the two forms of input is relatively automatic or mandatory in emotional perceptual processing (de Gelder & Vroomen, 2000; Vroomen et al., 2001). Again, there was evidence in their findings that the impact of prosody on emotional judgements tends to increase with the level of emotional ambiguity of the related face (de Gelder & Vroomen, 2000; Vroomen et al., 2001; also Massaro & Egan, 1996).

For researchers interested in how prosody influences face processing when *unambiguous* displays of emotion are encountered, available studies are somewhat limited in their reliance on morphed or physically altered stimuli that have rendered the emotional meanings of interest ambiguous to the decoder. In the evaluation paradigm, at least, this variable seems crucial in how well researchers detect *prosodic* influences on a face (de Gelder & Vroomen, 2000; Massaro & Egan, 1996). Another methodological limitation of the evaluation task is that by requiring subjects to explicitly name the emotional identity of a prosody or face stimulus (e.g., *happy* vs. *sad*), this task limits the ability of researchers to register the array of affective and emotional details that may be activated by these events, and which may guide on-line interpretative processes. Finally, given known variability in how different ''encoders'' exploit nonverbal cues to convey discrete emotions, especially in the prosodic channel (Banse & Scherer, 1996; Pell, 2002; Wallbott, 1988), more research is needed that taps prosody–face interactions from stimuli produced by a *group* of encoders rather than just one, allowing data to be generalized with greater confidence.

Studying Prosody–Face Interactions through the Facial Affect Decision Task (FADT)

Recently, the facial affect decision task (FADT) was introduced as a tool for examining the interface between prosody and facial expressions during on-line emotional processing (Pell, 2004, 2005). On the assumption that facial expressions activate emotion concepts or prototypes in associative memory (Etcoff & Magee, 1992; Halberstadt & Niedenthal, 2001), properties of the cross-modal lexical decision task (Swinney, 1979) were harnessed to gauge the influences of an emotionally related or unrelated prosody on decisions about a conjoined face. In this paradigm, subjects are presented an emotionally inflected sentence followed by a static facial expression that represents an unambiguous exemplar of one of the basic emotions, or a facial ''grimace'' that does not convey a true or known emotional state. Participants judge whether the face represents a ''true'' expression of emotion (YES/NO response) in the absence of any requirement to comment *explicitly* on emotional attributes of either the prosody or the face stimulus. By manipulating the emotional value of the prime and target, the stimulus onset asynchrony (SOA) of the two events, or other critical parameters, the FADT can reveal areas of intersection during on-line processing of the prosody and face which may be traced to shared affective or emotion-specific details of these events activated implicitly during the task.

In an initial study, Pell (2005) observed robust indications that identifying the representative value of an emotional face was biased by the emotional relationship of a preceding prosodic stimulus to the target. Using prosody and face stimuli that unambiguously conveyed *happiness*, *(pleasant) surprise*, *anger*, and *sadness* which were exhaustively combined across emotions for presentation to 48 young adults, results established that participants were significantly more accurate *and* faster to render facial affect decisions when the face was preceded by an emotionally congruent rather than incongruent prosody (Pell, 2005). These findings were obtained at two distinct prime-target SOA's (300 and 1000 ms) and in spite of instructions to attend strictly to the face, adding to claims that the emotional components of prosody and face interact in an empirically detectable and perhaps involuntary manner (de Gelder & Vroomen, 2000; Massaro & Egan, 1996). The observed congruity effects were attributed to the activation of conceptual knowledge about emotion categories shared by prosody and face, details which would be necessary to affirm the representative status of the emotional face target (Pell, 2005). The importance of the facial channel for interpreting results in the FADT was also highlighted by this research, as *negative* faces always promoted a marked increase in errors and decision times irrespective of the emotional category of the face and/or preceding prosody. This latter finding was attributed to broader claims that negative and potentially ''threatening'' facial expressions promote interference during perceptual processing tasks such as the FADT by directing attention away from immediate task-relevant goals (Dimberg & Öhman, 1996; Mogg & Bradley, 1999; see Pell, 2005 for a discussion). However, further investigation of prosody–face interactions in the context of these initial findings is clearly warranted.

The Current Study

As a means to reinforce and advance ideas proposed in our initial study, one of the key issues that bears critically on how emotion congruity effects in the FADT are understood is *how much* prosodic information from the prime stimulus is necessary to index these cross-channel ties. This question was not informed by results of our initial FADT study in which prosodic prime stimuli were always relatively long in duration (greater than 1.2 s) and where there was perceptual overlap between the offset of the prosodic stimulus and the onset of the face target (see Pell, 2005). In the broader literature, it is widely held that vocally expressed emotions gain meaning through the processing of long-term changes in spectro-temporal patterns of speech (e.g., Banse & Scherer, 1996; Ladd, Silverman, Tolkmitt, Bergmann, & Scherer, 1985; Pell, 2001; Williams & Stevens, 1981). However, there is little consensus as to what level of exposure to these cues is needed to implicitly or explicitly identify discrete emotions from prosody. While there is evidence that certain affective features shared by prosody and face are integrated very rapidly (less than 300 ms) into perceptual processing (see de Gelder et al., 1999; Pourtois et al., 2000), the source of these early perceptual responses is unlikely to include detailed semantic information about nonverbal displays of emotion needed to judge the representational status of 'true' facial expressions, which appears to govern prosody–face congruity effects in the FADT (Pell, 2005). As a consequence, there is little instruction in the existing literature as to how long prosodic information must be monitored by listeners to activate emotional meanings responsible for prosody–face congruity effects.

The principal goal of this research was to specify the approximate timecourse of prosody–face interactions in the FADT by manipulating the duration of prosodic prime stimuli. This approach will permit more accurate inferences about the origins of cross-channel priming effects between prosody and face, and supply invaluable new data on how emotion knowledge is retrieved from prosodic stimuli over time. To concentrate on the role of prosody duration in the FADT, we employed a simplified task design involving only two emotions, *happy* and *sad*; these two well-studied emotions were successfully employed by Pell (2004) in a recent study which replicated the priming effects of emotional prosody on decision times about a related face when the prosodic primes were relatively long in duration (Pell, 2004, 2005). Here, prime stimuli used by Pell (2004) were cut into prosodic "fragments" measuring precisely 300, 600, or 1000 ms in duration to better qualify how much prosodic information may have been necessary for activating emotional meanings of the primes responsible for cross-channel effects. We expected that emotion congruity between prosody–face events would impact significantly on performance measures to the face in conditions of a relatively "long" prime-target SOA (600 and 1000 ms) which resemble those of our previous FADT experiments (Pell, 2004, 2005). However, existing data precluded firm predictions about the effects of a brief, 300 ms burst of prosodic information in this study, nor could it be predetermined what prosody duration condition(s) might be associated with maximal priming of facial affect decisions.

Method

Participants

Forty-five participants (23 female, 22 male) were recruited to the study through campus advertisements. All participants were native English speakers with good hearing and normal or corrected-to-normal vision.

Subjects within the group had a mean age of 24.4 years (±6.9) and had completed an average of 15.9 years (±2.2) of formal education.

Stimuli

Materials consisted of ''fragments'' of emotionally inflected utterances as the prime stimuli, matched with faces portraying either a 'true' or 'false' expression of emotion as the target stimuli. All prosodic and facial stimuli were selected from an existing database of exemplars, which were rigorously defined along key perceptual dimensions in a norming study (see Pell, 2002, for details of stimulus construction and validation). Face target stimuli were static, color photographs of an actor's unobstructed facial expression, hair, and shoulders (expressions were posed by three male and three female actors). 'True' face targets were expressions of *happiness* or *sadness* that had achieved a minimum of 80% group consensus about the intended target emotion in the validation study. 'False' emotional expressions were facial grimaces posed by the same encoders that were identified as not representing an emotion state by greater than 70% of decoders in the norming study. All face stimuli were saved as bitmaps measuring $17.1 \times 17.1\ cm^2$ when presented on a computer screen. Examples of 'true' and 'false' facial expressions are illustrated in Figure 1.

Prosodic prime stimuli consisted of ''fragments'' of semantically anomalous utterances (e.g., *Someone migged the pazing*) produced by the same actors to communicate a *happy*, *sad*, or *neutral* emotion through the use of vocal-prosodic cues. Only stimuli greater than 1.5 s in total duration and which had achieved a minimum consensus of 80% correct identification of the target emotion based on the whole utterance were selected for the construction of prosodic primes. Each selected item was cut and saved in the form of three prosodic fragments measuring 300, 600, and 1000 ms in duration from the *onset* of the utterance using Praat speech editing software. The decision to extract fragments from the onset of utterances was justified by our interest in how emotional meanings of prosody are registered and perhaps evolve during ongoing speech processing when novel utterances are first intercepted. Table 1 summarizes major perceptual and acoustic attributes of both the prosody and face stimuli selected for the study.

Although prosodic fragments were constructed from stimuli with well-recognized emotional properties when judged as whole utterances, a brief perceptual experiment was run to ascertain whether emotion information could be *explicitly* recognized from the new prosodic frag-

Figure 1. Examples of 'true' and 'false' facial expressions posed by one male and one female encoder.

ments as a function of their duration. (There was no *a priori* assumption that explicit recognition abilities would fully predict effects due to implicit registration of prosodic meanings in the FADT; these data were gathered for descriptive purposes only.) Following Pell's (2002) procedure for defining perceptual features of the base stimuli, the prosodic fragments measuring 300, 600, or 1000 ms in duration were entered into three separate auditory experiments, randomized with additional fragments of the same length which were prepared from stimuli conveying *anger, disgust,* and *pleasant surprise*. A group of 10 young listeners was asked to judge the emotion of each fragment (6-choice task). For experimental primes conveying a *happy*, *sad*, or *neutral* prosody, results indicated the following target recognition consensus among the 10 listeners, averaged across tokens in each prosody length condition: ''*300*'': 25%, 75%, 78%; ''*600*'': 50%, 80%, 80%; ''*1000*'': 70%, 90%, 78%. These findings confirmed, by and large, that the emotions of interest could be well identified in most conditions even when prosodic primes were confined to 300 ms in duration (except for *happy*) and that explicit recognition of all emotions tended to increase as a function of the prosody duration.

TABLE 1

Perceptual Characteristics of Prosody and Face Stimuli Selected for the Experiment

Stimulus type	Parameter	Stimulus Emotion		
		Happiness	Sadness	Neutral
Prosodic primes	Perceived target recognition[a] (%)	85	93	89
	Perceived target intensity[b] (0–5)	3.5	3.6	–
	Perceived valence[c] (−5 to +5)	+3.4	−2.8	−.1
	Fundamental frequency (*mean*[d])	.93	−.93	−.65
	Fundamental freq. (*variation*[d])	1.15	.43	.54
	Speaking rate (syllables/second)	5.1	4.2	4.4
Face targets	Perceived target recognition[a] (%)	100	94.8	–
	Perceived target intensity[b] (0–5)	3.8	3.6	–
	Perceived valence[c] (−5 to 5)	+2.9	−2.9	–

[a]Based on 26 young control subjects in an eight-choice forced identification task including an open response category, where chance performance=12.5%.

[b]Represents the perceived intensity of the intended target emotion when correctly identified by the 26 subjects in the forced-choice identification task.

[c]Reflects the perceived evaluation of the stimuli along an 11-point positive–negative continuum by a separate group of 14 healthy decoders.

[d]Utterance measures were first normalized in reference to each speaker and then averaged across speakers who produced exemplars of each emotion. For fundamental frequency *mean*, all utterances produced by a given speaker was transformed as follows: $F_{0,\text{norm}} = (F_{0i} - F_{0,\text{Mean}})/$SD, where F_{0i} is the observed utterance F_0 of the item in Hertz, $F_{0,\text{Mean}}$ is the speaker's typical mean F_0 averaged for a broad array of their utterances (Pell, 2002), and SD is the standard deviation. For fundamental frequency *variation*, the F_0 range for each utterance (maximum–minimum) was divided by the corresponding F_0 mean for that item and then averaged across speakers, by emotion.

Experimental Design

The FADT generates a YES or NO decision whereby the participant either confirms or rejects the status of the face target as a 'true' expression of emotion under different processing conditions. A total of 144 trials, representing an equal number of 'true' (YES trials, $n = 72$) and 'false' (NO trials, $n = 72$) facial expressions of emotion as the target, were randomized within a single experiment composed of five separate blocks. YES trials reflected one of three underlying relationships between the prosody and the matched face stimulus: *congruent* (happy–happy, sad–sad); *incongruent* (happy–sad, sad–happy); or *neutral* (neutral–happy, neutral–

sad). Twelve distinct prosody–face pairings were constructed for each target emotion within the congruent, incongruent, and neutral conditions for a total of 24 trials in each condition overall. To construct trials across the three YES conditions, each 'true' face exemplar was always matched with a *happy*, *sad*, and *neutral* prosody spoken by a member of the same sex. Similarly, the 72 NO trials were constructed by pairing each *happy*, *sad*, and *neutral* prime stimulus entered into YES trials with one of 12 facial grimaces posed by the same actors (again respecting the sex of the actor who posed the two expressions for a given trial).

According to these principles, the 300, 600, and 1000 ms fragment constructed from a given prosodic stimulus was always matched with the same face target but entered into three separate facial affect decision tasks (''300'', ''600'', and ''1000'' tasks, respectively). The three tasks were identical in all respects with the exception of the duration of prosodic primes. The composition of blocks within each task avoided exact stimulus repetitions to the greatest extent possible. The critical proportion of trials consisting of a prosodic prime matched with an emotionally congruent face target within each task was relatively low (.17).

Procedure

Subjects were invited to take part in a study of ''communication and emotion'' and each participant completed all three tasks, presented during two sessions using Superlab presentation software (Cedrus Corporation, USA). Two of the tasks were completed during the initial session and the third task was completed at least one week later to mitigate repetition and carryover effects. Presentation order of the three tasks was counterbalanced across subjects, although the ''300'' task was always completed during the first session (although not always first); this decision reflected our impression that 300 ms prosodic fragments bore the least resemblance to primes in the ''600'' and ''1000'' tasks, and were thus unlikely to promote carryover between tasks completed during the first session. Subjects were compensated CAD$20 for their participation following completion of the third task.

For each trial, subjects were instructed to judge whether the face target they viewed represented a 'true' expression of emotion by pressing a YES or NO button on a Cedrus 4-button response box. Subjects were encouraged to judge the facial expression as accurately but also as quickly as possible and both the accuracy and latency of these decisions were recorded. The experiment was run and all data recorded by a high-quality portable computer with a 37.5 cm viewing screen and millisecond timer.

Auditory stimuli were played at a comfortable listening level over adjustable stereo headphones, while face stimuli were viewed on the computer monitor from a distance of approximately 45 cm. Each trial consisted of: a centrally displayed visual fixation marker of 350 ms; a 500 ms pause; the prosodic stimulus; and the face target presented at the offset of the prosodic stimulus. Participants were instructed not to pay close attention to the auditory stimulus, which would sound like ''the beginning of sentences that do not make sense'', but rather to focus at all times on the nature of the face stimulus. When a button press was registered in response to the face, the visual stimulus was erased from the screen and a 2000 ms inter-trial interval signaled the onset of the next trial.

Two blocks of practice trials preceded each task to familiarize subjects with the notion of 'true' vs. 'false' emotional faces, first in isolation (practice block 1) and then in conjunction with prosodic fragments of the appropriate duration (practice block 2). In the practice blocks only, feedback on the accuracy and then speed of the subjects' response was provided automatically by the computer (''Incorrect''; ''Please respond faster''). Following practice, experimental blocks were presented in an established random order, which was varied across subjects. At no time during testing did the examiner comment on specific emotional qualities that may be associated with either the prosody or face stimulus.

Results

Overall error rates among the three tasks were relatively comparable at 11.4% (±12.4) for the ''*300*'' task, 9.7% (±10.6) for the ''*600*'' task, and 10.1% (±11.6) for the ''*1000*'' task. Errors for YES trials were predictably higher than those for NO trials in each task (*300*: YES = 15.9% (±19.8), NO = 6.3% (±11.1); *600*: YES = 13.5% (±16.0), NO = 5.9% (±11.1); *1000*: YES = 13.7% (±17.4), NO = 6.1% (±10.6)). Based on the overall error patterns, data for two male and one female participant who committed in excess of 33% errors for combined YES and NO trials, and who were likely to have misunderstood the task goals, were removed from further consideration. When judging 'true' emotional expressions (YES trials), eight additional subjects (five female, three male) made greater than 25% errors across the three tasks; following Pell (2005), data for these individuals were considered in all analyses of error patterns but were deemed too unreliable for inclusion in analyses of latencies which considered only *correct* responses to YES trials. For the remaining 34 subjects, latency data were combed for outliers and extreme values prior to analysis by eliminating values shorter than 300 ms or greater than 2000 ms and by setting individual subject latencies greater than 2 SD from the conditional mean at values equal to 2 SD in the appropriate direction. For the *''300''*, *''600''*, and *''1000''* tasks, respectively, these modifications affected .1%, .3%, and 0% of total values which were removed, and 4.2%, 4.4%, and 3.9% of total values which were replaced. Table 2 summarizes the group error and latency measures analyzed in the experiment as a function of the emotional value of the prime and target and the duration of the prosodic prime.

Error and latency patterns obtained for 'true' emotional targets were entered into separate $2\times2\times3$ ANOVAs with repeated measures on Face (happiness, sadness), Prosody (congruent, incongruent), and Prosody Duration (300, 600, 1000 ms). Data reflecting neutral prosody were omitted to focus these analyses on the interplay between prime duration and the prosody–face relationship. Tukey's post hoc comparisons ($p < .05$) were conducted to qualify the nature of significant main or interactive effects and effect size was computed, where prescribed, according to Rosnow and Rosenthal (2003).

For *errors*, there was a main effect of Face emotion on decision accuracy across tasks, F (1, 41) = 21.99, $p < .001$, $r = .59$; participants made notably fewer errors when judging *happy* than *sad* expressions as a 'true' exemplar of emotion. As well, there was a significant main effect of Prosody, F (1, 41) = 4.15, $p < .05$, $r = .30$, and a three-way interaction of Face×Prosody×Prosody Duration, F (2, 82) = 5.07, $p =$ .01. Post hoc analysis of the interaction revealed that whereas decision errors about *happy* faces did not vary according to the value and/or duration of prosodic primes, more accurate decisions about *sad* faces were observed following a congruent rather than an incongruent prosody that was presented for 600 ms in duration. Moreover, decisions about *sad* faces accompanied by a congruent *sad* prosody that endured for 600 or 1000 ms were significantly more accurate than when preceded by a congruent *sad* prosody measuring only 300 ms. Error rates for *sad* faces accompanied by an incongruent prime did not vary as a function of prime duration.

For *latencies*, there was evidence that *happy* faces were judged significantly faster than *sad* faces across tasks in the form of a Face main effect, F (1, 33) = 179.32, $p < .001$, $r = .92$. In addition, main effects for Prosody, F (1, 33) = 23.89, $p < .001$, $r = .65$, and Prosody Duration, F (2, 66) = 3.40, $p = .05$, on judgement latencies were observed, and facial decision latencies were influenced significantly according to the interaction of Prosody and Prosody Duration, F (2, 66) = 5.69, $p = .006$. Inspec-

TABLE 2

Mean Errors (%) and Latencies (ms) of Facial Affect Decisions when Preceded by Prosodic Fragments Measuring 300, 600, or 1000 ms in Duration, according to Emotional Attributes of the Prosodic Prime and Face Target

Measure	Emotion of face	Prosody duration	Emotion of prosody			Priming[a] (Incon-con)
			Happiness	Sadness	Neutral	
Errors	Happiness	300	2.8	3.8	3.6	1
		600	2.0	1.4	2.8	−.6
		1000	2.4	2.4	2.8	0
	Sadness	300	19.2	20.4	17.1	−1.2
		600	19.2	14.3	15.5	4.9
		1000	17.3	14.1	15.9	3.2
	'False'	300	5.0	5.0	5.0	–
		600	4.5	5.9	5.9	–
		1000	5.8	5.3	4.4	–
Latencies	Happiness	300	548	557	565	12
		600	533	568	570	32
		1000	571	589	602	18
	Sadness	300	704	699	710	5
		600	695	666	683	23
		1000	736	704	718	35
	'False'	300	653	656	672	–
		600	639	657	655	–
		1000	666	678	686	–

[a]To calculate the priming effect the congruent condition was always subtracted from the corresponding incongruent condition (ignoring data for neutral prosody).

tion of the interaction revealed that independent of the emotional category of the face, prosody–face congruity effects (i.e., *incongruent – congruent*) were only reliable when subjects were exposed to prosodic information enduring for 600 or 1000 ms but not fragments of 300 ms. Interestingly, whenever trials were composed of an emotionally *congruent*

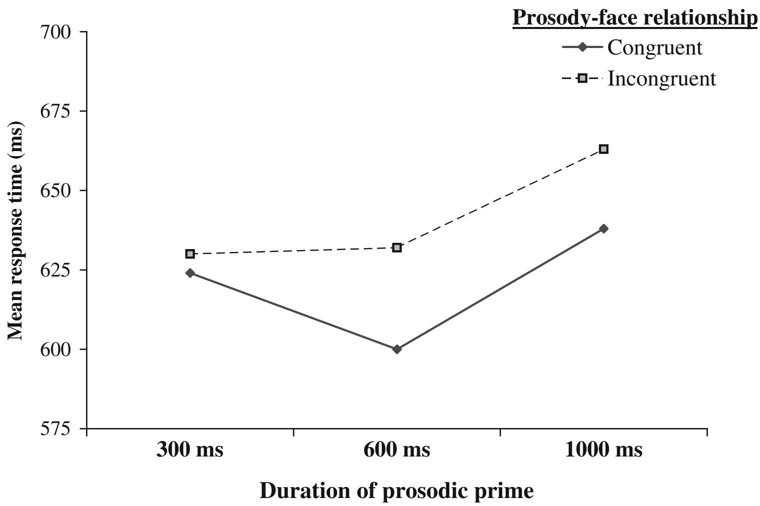

Figure 2. Response latencies to 'true' facial expressions as a function of the prosody–face relationship and the prosody duration.

prosody and face, prosodic fragments measuring 600 ms in duration yielded significantly shorter latencies (*M* = 600 ms) than when the prosodic prime measured 300 ms (*M* = 624 ms) or 1000 ms (*M* = 638 ms). By contrast, when trials presented emotionally conflicting prosody–face expressions, latencies following a 300 or 600 ms prosodic fragment (*M* = 630 and 632 ms, respectively) tended to be shorter than when a 1000 ms prosodic stimulus was presented (*M* = 663 ms). These relationships are depicted graphically in Figure 2.[1]

Discussion

Employing perceptually unambiguous displays of emotion posed by a group of actors, our results establish that conjoined prosody–face expressions associated with the same emotion tend to facilitate information processing of the face target, extending similar findings in the small literature on this topic (de Gelder & Vroomen, 2000; Massaro & Egan, 1996; Pell, 2005). These findings imply that emotional specifications of prosody are registered automatically by listeners during on-line speech processing, even when subjects are not instructed to pay attention to these stimuli or their emotional meanings (de Gelder & Vroomen, 2000; Vroomen et al., 2001). As such, the observed congruity effects are likely to reflect, at least in part, the *implicit* registration of emotional meanings from prosody which are shared to some extent by facial expressions of basic emotions (Borod, 2000; Bowers et al., 1993; Pell, 2005). Equally important, our new data indicate that prosodic expressions of discrete emotions such as *happy* or *sad* have a definite timecourse for their underlying specifications to be sufficiently activated in memory during on-line speech processing, moderating the extent to which emotional similarities between nonverbal expressions have an outcome on target decisions.

Effects of Prosody Duration on Emotion Processing

Evidence that prosody–face interactions were dictated by the duration of prosodic stimuli encountered by listeners was reflected in both the speed and accuracy of subsequent facial affect decisions. With respect to the latency measures, there were clear signs that emotion priming occurred only when 600 or 1000 ms of prosodic information preceded the face target, and not in the context of a 300 ms prosodic fragment. In fact, latencies pertaining to congruent versus incongruent prosody–face trials showed practically no differentiation in the task that restricted the duration of prosodic information to 300 ms (review Figure 2). The importance of prosody duration was further exemplified by subjects' accuracy for judging the status of face targets (especially for *sad* faces) which was facilitated by a congruent prosody presented for 600 or 1000 ms (but not 300). These results argue that underlying knowledge about emotional prosody was not sufficiently activated by a 300 ms sample of representative cue sequences, precluding cross-modal emotion bias on a related face in this particular environment.

There are strong reasons to expect that positive/negative *valence* attributes of prosody would be activated by a 300 ms prosodic stimulus (Bargh, Chaiken, Govender, & Pratto, 1992; Fazio, Sanbonmatsu, Powell, & Kardes, 1986; Hermans, de Houwer, & Eelen, 1994; Pourtois et al., 2000); as such, the lack of congruity priming in the "300" task suggests that prosody–face interactions are not fully governed by these basic affective properties of event valence, but rather, revolve around underlying emotion knowledge shared by emotional prosody and faces which is tapped when executing a facial affect decision (Niedenthal & Halberstadt, 1995; Pell, 2005). Based on the current findings, this knowledge appears to be present in memory if approximately 600 ms of prosodic information is encountered, yielding robust emotion congruity effects across channels, and these activations continue to impact on response latencies to a related face at even longer exposures to the prosodic stimulus (i.e., 1000 ms). This information on the timecourse for retrieving emotion knowledge about prosody is novel, and as well, serves to clarify initial findings of Pell (2005) who observed prosody–face congruity effects involving four distinct emotions at two separate SOAs (300 and 1000 ms); as speculated in that report, due to perceptual overlap between the prosody and face stimulus in the administration of our initial FADT task, the current results argue that any facilitative effects of the prosody on the face at the short (300 ms) SOA would have occurred sometime *after* the onset of the face, as subjects continued to register emotionally relevant prosodic cues in the utterance beyond the initial 300 ms segment (Pell, 2005).

It is potentially informative that when congruent prosody–face cues promoted correct decisions about the emotional status of the face, responses were optimally enhanced in speed in the ''600'' task when compared to either the ''300'' or ''1000'' task (which did not differ). Conflicting prosody–face expressions, while consistently slower than congruent trials at each prosody duration as one might expect, displayed minimal latency differences following a 300 or 600 ms fragment with significantly longer decision times when the prosodic fragment played for 1000 ms. Thus, when the effect of prosody duration is examined for both congruent and incongruent trials, data reveal that mean group latencies were always *delayed* in the condition where subjects received the *most* prosodic information (i.e., the ''1000'' task).

While further data are clearly needed to understand the effects of the prosody duration manipulation, the patterns described supply initial clues that processing of emotion knowledge from prosody may be optimized when approximately 600 ms of prosodic information is encountered by listeners. (Perhaps more accurately, congruity effects would appear to ''peak'' within the time window of 300–1000 ms of prosodic information, pending experiments which more finely manipulate the duration of the prosodic fragment.) One issue that is unresolved by our data is whether these emotional meanings would be more readily accessed from prosody if sentence ''fragments'' were taken from the middle or the end of spoken utterances, where representative cues to specific emotions may manifest differently in the vocal signal, affecting comprehension processes. In addition, although these data establish that emotion congruity between prosody and face has a continued benefit on cognitive performance over conflicting trials when longer (1000 ms) prosodic stimuli are presented (also Pell, 2005), exposure to these longer primes has a general cost on the speed of facial affect decisions irrespective of the emotional relationship of the prosody to the face. Identifying factors that seem to promote interference in face processing following relatively long prosodic fragments, and the exact point at which maximal facilitation of congruent prosody–face events is realized, are important questions for future applications of the FADT and other investigative approaches for studying the prosody–face interface.

Effects of Face on Emotion Processing in the FADT

Independent of the characteristics and potential associations of prosodic information presented to our subjects, a wide divergence was noted in how accurately subjects classified *happy* vs. *sad* faces as exemplars of 'true' emotions (2.7% vs. 17.0% overall errors, respectively). A corresponding departure in response latencies to *happy* (567 ms) vs. *sad* (702 ms) faces overall was observed in the same direction. These marked differences were not unexpected given our observation that facial expressions associated with a *negative* appraisal always took longer to judge and elicited more frequent errors in our earlier administration of the FADT which involved two positive and two negative emotion categories (Pell, 2005). These effects highlight that the FADT indexes some of the cognitive sequelae of underlying valence attributions of face events in addition to content-specific features of these expressions.[2]

Although face-related effects in the FADT have already been considered in some length (Pell, 2005), it merits emphasizing that impaired performance for *negative* vs. *positive* facial expressions is a highly stable pattern in all of our experiments involving the FADT. These highly resilient response tendencies due to the valence of facial expressions are not trivial as they are likely to reflect basic, ''hard-wired'' human responses to specific affective components of potentially aversive faces (Dimberg, 1991; Dimberg & Öhman, 1996; Mogg & Bradley, 1999; Öhman, 1987). Faces conveying *sadness* may have promoted interference in rendering a facial affect decision because underlying negative evaluations of these stimuli are initially conducive to threat and capture attention away from immediate task goals (Dimberg & Öhman, 1996; Mathews & Mackintosh, 1998; Pratto & John, 1991; van Honk, Tuiten, de Haan, van den Hout, & Stam, 2001). This interference tends to be far less for expressions of *sadness* than for *anger* and *fear* (Dimberg & Öhman, 1996; Pell, 2005).

In the cross-modal context of the FADT, one can surmise that influences of a negative face are introduced at a stage *prior* to processes that compare semantic properties of the prosody and face in detail (Massaro & Egan, 1996). This explanation fits our finding that a conjoined, negative prosody did not influence or amplify the magnitude of interference to negative faces in our response measures. Future research could usefully look at whether presenting prosody as the *target* or focus of attention in some variant of the FADT would lead to valence-related disruptions of similar magnitude to a target face, illuminating whether biological asymmetries exist in how cues to ''threat'' are communicated and registered in the two nonverbal channels.

Finally, when one considers accuracy performance in this study there was evidence that prosody–face congruity had a primary impact on decisions about *sad* but not *happy* facial expressions. Given the high number of errors promoted by *sad* face targets, this condition would have allowed greater sensitivity in our error measures to detect the overlapping features of prosody and face events in the context of *sad* but not *happy* facial expressions. *Happy* expressions are notoriously easy to identify in the face channel for normal participants (Calder et al., 2003; Kirouac & Dore, 1984; Wallbott, 1988) and recognition of these displays is comparatively resistant to acquired brain pathologies which tend to disrupt facial expression processing (Adolphs, Damasio, Tranel, & Damasio, 1996; Pell & Leonard, 2005). In contrast to our previous investigation which combined prosody and face stimuli representing two positive and two negative emotions—a more complex design which promoted nearly twice the proportion of errors in response to *happy* face targets (Pell, 2005)—the observed ceiling effect for *happy* faces here would have permitted little opportunity in the error analyses to index the relationship of the prosody on facial affect decisions, pending further examination of these task-dependent patterns.

Toward an Understanding of Prosody–Face Interactions in Natural Communication

While it is commonly held that emotional meanings from prosody are decoded from long-term cue variations involving (minimally) the speaking rate of an utterance, its mean pitch, and related variations in pitch and loudness over time (see Juslin & Laukka, 2003 for a review), available studies supply few clues about the time domain over which emotional meanings are activated by prosodic cue sequences pertaining to discrete emotions. As a starting point, our findings argue that the emotional specifications of prosody that drive congruity effects in the FADT are registered at some point between 300 and 600 ms into the prosodic stimulus, and that these memory activations include emotion-specific details about prosodic events (Pell, 2005). By nature, facial affect decisions require access to detailed or prototypical descriptions of known emotional expressions in the face, placing prosody–face interactions at this level of semantic detail. Our findings allow speculation that the timing for detecting emotion congruity/incongruity effects during nonverbal emotion processing bear certain similarities to that of well-documented and potentially related ''semantic'' effects observed during verbal processing of non-emotional stimuli (see Kutas & Federmeier, 2000).

Naturally, describing the origins of prosody–face interactions in the FADT at the level of emotion-specific category information (e.g., Niedenthal & Halberstadt, 1995, 1999) does not preclude earlier evaluations of prosody valence and other affective components of these primes such as arousal (Pell, 2005) which may have been fully appraised in our ''300'' condition. In line with component process models of emotional appraisal (e.g., Scherer, 1986), these initial evaluations may serve as the entry point for integrating prosody–face details when examined using other approaches that do not necessarily elicit a conceptual analysis of nonverbal communicative displays (de Gelder et al., 1999; Pourtois et al., 2000). Our data suggest that prosodic influences on facial affect decisions are not governed by these coarse affective dimensions, but rather, that they sensitively index the emotion-specific, socially-relevant details associated with nonverbal expressions in the two channels (Innes-Ker & Niedenthal, 2002; Niedenthal & Halberstadt, 1995).

In spite of instructions not to attend to prosodic primes, it is likely that some participants allocated conscious attention to the prosodic stimuli at times and that attentional strategies exert an important influence on the cross-modal integration of emotion displays (see Klauer & Musch, 2002, 2003, for a critical discussion). The importance of attentional factors on how emotional prosody is integrated with adjoined communicative events may also vary as a function of biological sex (Schirmer & Kotz, 2003; Schirmer, Zysset, Kotz, & von Cramon, 2004) or as a result of cultural influences (Kitayama & Ishii, 2002). During natural communicative interactions, one can speculate that interlocutors tend to distribute their attention in some measure to features of both prosody and face, a strategy that would tend to facilitate decisions about the emotive intentions of interlocutors whose nonverbal behaviors typically contain overlapping features across channels. When the input presents a conflict between nonverbal channels–for example, in the case of a sarcastic comment which presents a *cheerful* sounding prosody with an incongruent *neutral* or *negative* looking face, or in the case of intended deceit when these intentions ''leak out'' in only one channel—distributed attention to both prosody and face would also highlight discrepancies in the feature specifications available in memory for further processing and social analysis.

The FADT may prove especially useful in the investigation of these and other facets of communication that revolve around an underlying *mismatch* of nonverbal cues across channels. As well, the FADT is well positioned to advance knowledge of the on-line interactions between prosody, face, and lexical-semantic channels of emotion communication in healthy decoders (Borod, 2000) and will aid future descriptions of acquired impairments of emotional communication in brain-damaged populations who may exhibit explicit, but not implicit, impairments for recognizing emotional prosody (Pell, 1998; Pell & Leonard, 2003). Of greatest relevance here, continued attention to studying emotional prosody during *on-line* processing tasks will render new information on how the emotional meanings encoded by prosodic cue sequences are activated and interpreted during natural communication, a process which remains poorly understood.

Acknowledgments

I would like to thank Elmira Chan, Nicole Skelton, and Nicole Price for running the experiments, for their care in data organization, and for help in manuscript preparation. I am also grateful to Dr. Klaus Scherer for comments received on an earlier presentation of this research. This research was made possible by a William Dawson Scholar Award granted to the author by McGill University, and by operating funds received from the Natural Sciences and Engineering Research Council of Canada.

Notes

1. Two additional 3 × 3 ANOVAs with repeated measures on Prosody (*happy, sad, neutral*) and Prosody Duration (300, 600, 1000) were run on the error and latency responses to 'false' facial expressions (NO trials). There were no effects of the emotional value and/or duration of prosodic primes on errors to 'false' faces, all $F < 1.28$, $p > .28$. For latencies, shorter decision times were observed when 'false' expressions were preceded by *happy* rather than *sad* or *neutral* prosody [Prosody main effect, F (2, 66) = 11.94, $p < .001$], although no main or interactive effects of Prosody Duration were noted, both $F < 1.55$, $p > .22$. These patterns exemplify that the influence of emotional prime characteristics was largely restricted to 'true' expressions of facial emotion in the paradigm.

2. An alternative explanation for the face main effect is that performance measures in the experiment were biased by judgmental tendencies at a *post-lexical* stage of generating a YES/NO decision in the context of specific feature activations of the target and/or prime stimulus (Klauer & Musch, 2002; Musch & Klauer, 2001; Wentura, 2000). For example, it has been suggested that negative details associated with *sad* faces initially bias a negative, NO rather than a positive, YES response, promoting interference in accuracy and latency measures selectively for *sad* or other negative faces at the stage of response generation. This possibility was tested in a separate, unpublished experiment, which *inverted* the target–response relationship following methods of Pell (2004) and Wentura (2000), such that decisions about 'true' emotional faces mapped onto a NO response. The observation that *negative* face targets yielded markedly higher decision errors and latencies in the inverted paradigm remained unchanged, indicating that valence information about the face, and not the response category, was critical for understanding the face main effect.

References

Adolphs, R., Damasio, H., Tranel, D., & Damasio, A. R. (1996). Cortical systems for the recognition of emotion in facial expressions. *The Journal of Neuroscience*, *16*, 7678–7687.

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

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

Bargh, J., Chaiken, S., Govender, R., & Pratto, F. (1992). The generality of the automatic attitude activation effect. *Journal of Personality and Social Psychology*, *62*, 893–912.

Borod, J. (2000). Relationships among facial, prosodic, and lexical channels of emotional perceptual processing. *Cognition and Emotion*, *14*, 193–211.

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

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

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

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*, 289–311.

Dimberg, U., (1991). Emotional reactions to facial expressions: A case of automatic responding? *Psychophysiology*, *28*, S19.

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

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

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, *50*, 229–238.

Halberstadt, J., & Niedenthal, P. (2001). Effects of emotion concepts on perceptual memory for emotional expressions. *Journal of Personality and Social Psychology*, *81*, 587–598.

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

Hermans, D., DeHouwer, J., & Eelen, P. (1994). The affective priming effect: Automatic activation of evaluative information in memory. *Cognition and Emotion*, *8*, 515–533.

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.

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

Kirouac, G., & Dore, F. (1984). Accuracy and latency of judgment of facial expressions of emotion. *Perceptual Motor Skills*, *57*, 683–686.

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

Klauer, K. C., & Musch, J. (2002). Goal dependent and goal independent effects of irrelevant evaluations. *Personality and Social Psychology Bulletin*, *28*, 802–814.

Klauer, K., & Musch, J. (2003). Affective priming: Findings and theories. In J. Musch, & K. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 7–49). London: Erlbaum.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*, 463–470.

Ladd, D. R., Silverman, K., Tolkmitt, F., Bergmann, G., & Scherer, K. (1985). Evidence for the independent function of intonation contour type, voice quality, and F0 range in signaling speaker affect. *Journal of the Acoustical Society of America*, *78*, 435–444.

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

Mathews, A., & Mackintosh, B. (1998). A cognitive model of selective processing in anxiety. *Cognitive Therapy and Research*, *22*, 539–560.

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

Musch, J., & Klauer, K. C. (2001). Locational uncertainty moderates affective congruency effects in the evaluative decision task. *Cognition and Emotion*, *15*, 167–188.

Niedenthal, P. M., & Halberstadt, J. B. (1995). The acquisition and structure of emotional response categories. *The Psychology of Learning and Motivation*, *33*, 23–64.

Niedenthal, P., & Halberstadt, J. (1999). Emotional response categorization. *Psychological Review*, *106*, 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*, 403–432.

Öhman, A. (1987). The psychophysiology of emotion: An evolutionary-cognitive perspective. *Advances in Psychophysiology*, *2*, 79–127.

Pell, M. D. (1998). Recognition of prosody following unilateral brain lesion: Influence of functional and structural attributes of prosodic contours. *Neuropsychologia*, *36*, 701–715.

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

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. (2004). A method for on-line evaluation of emotional prosody in healthy and disordered populations. *Brain and Language*, *91*, 25–26.

Pell, M. D. (2005). Nonverbal emotion priming: Evidence from the 'Facial Affect Decision Task'. *Journal of Nonverbal Behavior*, *29*, 45–73.

Pell, M. D., & Leonard, C. (2003). Processing emotional tone from speech in Parkinson's disease: A role for the basal ganglia. *Cognitive, Affective, & Behavioral Neuroscience*, *3*, 275–288.

Pell, M. D., & Leonard, C. L. (2005). Facial expression decoding in early Parkinson's disease. *Cognitive Brain Research*, *23*, 327–340.

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*, 1329–1333.

Pourtois, G., Debatisse, D., Despland, P.-A., & de Gelder, B. (2002). Facial expressions modulate the time course of long latency auditory brain potentials. *Cognitive Brain Research*, *14*, 99–105.

Pratto, F., & John, O. P. (1991). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology*, *61*, 380–391.

Rosnow, R., & Rosenthal, R. (2003). Effect sizes for experimenting psychologists. *Canadian Journal of Experimental Psychology*, *57*, 221–237.

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

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

Schirmer, A., Zysset, S., Kotz, S., & von Cramon, D. Y. (2004). Gender differences in the activation of inferior frontal cortex during emotional speech perception. *NeuroImage*, *21*, 1114–1123.

Swinney, D. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, *18*, 645–659.

van Honk, J., Tuiten, A., de Haan, E., van den Hout, M., & Stam, H. (2001). Attentional biases for angry faces: Relationships to trait anger and anxiety. *Cognition and Emotion*, *15*, 279–297.

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.

Wallbott, H. G. (1988). Big girls don't frown, big boys don't cry: Gender differences of professional actors in communicating emotion via facial expression. *Journal of Nonverbal Behavior*, *12*, 98–106.

Wentura, D. (2000). Dissociative affective and associative priming effects in the lexical decision task: Yes versus no responses to word targets reveal evaluative judgement tendencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 456–469.

Williams, C., & Stevens, K. (1981). Vocal correlates of emotion states. In J. K. Darby (Ed.), *Speech evaluation in psychiatry* (pp. 221–240). New York: Grune & Stratton.

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*, 831–852.

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.