The contributions of prosody and semantic context in emotional speech processing

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

[1] School of Communication Sciences & Disorders, McGill University, Montréal, Canada
[2] Département de Réadaptation, Université Laval, Québec, Canada
[3] Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
marc.pell@mcgill.ca

Abstract

The present study examined the relative contributions of prosody and semantic context in the implicit processing of emotions from spoken language. In three separate tasks, we compared the degree to which *happy* and *sad* emotional prosody alone, emotional semantic context alone, and combined emotional prosody and semantic information would prime subsequent decisions about an emotionally congruent or incongruent facial expression. In all three tasks, we observed a congruency effect, whereby prosodic or semantic features of the prime facilitated decisions about emotionally-congruent faces. However, the extent of this priming was similar in the three tasks. Our results imply that prosody and semantic cues hold similar potential to activate emotion-related knowledge in memory when they are implicitly processed in speech, due to underlying connections in associative memory shared by prosody, semantics, and facial displays of emotion.

1. Introduction

In spoken language, emotional information can be communicated through the stress and intonation patterns in prosody, as well as through semantic meaning. Past studies have shown that emotions can be accurately recognized through speech prosody alone, devoid of meaningful semantic information (i.e., from “pseudo-utterances”) [10]. Conversely, listening to utterances containing an emotional semantic context can lead to varying responses depending on the presence or absence of congruent emotional prosody [5]. Recent neurocognitive studies employing event-related potentials (ERPs) have also advanced the idea that semantics and prosody are differentially processed at the neural level [6, 13]. However, few studies have looked at the relative contributions of each cue in emotional speech processing. Notably, to what extent do listeners incorporate information from prosody vs. semantic context when interpreting the emotional significance of a stimulus? Evidence from several lines of inquiry including studies with brain damaged patients [3] as well as using ERPs [6], suggests that the emotional semantic meaning may be the more salient cue, and the emotional significance of an utterance may be more accurately characterized through semantic context compared to prosody.

While studies investigating affective processing have often employed forced-choice emotion recognition paradigms [2], relatively few studies have looked at how emotional prosody and semantic context are processed implicitly, without conscious attention to specific emotion labels. One avenue for studying implicit emotional processing is through priming. Using priming paradigms, researchers can analyze how emotional meanings are implicitly activated through cues from different sensory channels. To date, researchers have demonstrated that decisions about an emotional stimulus are faster when preceded, or primed, by an emotionally congruent (vs. incongruent) cue [11], even when the prime and target stimuli are elicited from differing channels [4].

Recent research on priming using cues from emotional prosody has employed the Facial Affect Decision Task (FADT) [7, 8]. In the FADT, a short spoken sentence (prime) is presented followed by a facial expression (target). Similar to the lexical decision task, the face targets are either “true” emotional expressions (e.g., *happy*, *sad*) or facial grimaces that do not represent a discrete emotion. Participants respond “yes” or “no” as to whether each face target represents a true emotion. The absence of conscious verbal labeling, such as in emotion recognition and categorization tasks, ensures the emotional meaning is implicitly activated. Past studies using the FADT have shown that emotional information from prosody alone (i.e., pseudo-utterances) primes decisions about an emotionally congruent target face [7, 8]. That is, participants render a facial affect decision more rapidly when the implicitly processed emotion from the prosodic prime is congruent with the emotion expressed by the face target. However, the magnitude of this effect relative to primes with semantic information is unknown. Such comparisons may inform the strength and interaction of both cues in processing emotion from spoken language.

In the present study, we used the FADT to compare the relative strength of implicitly processed prosody, semantics, and both cues in tandem, in priming subsequent decisions about a congruent facial emotion. In three tasks, we manipulated the prime stimulus to contain emotional information from only prosody (Prosody Task), only semantic context (Semantic Task), and congruent prosody and semantic context (Prosody-Semantic Task), in order to facilitate comparisons of implicit emotional speech processing across conditions. We report here a comparison of the three tasks and further analyses of these data can be found elsewhere [9].

2. Methods

2.1. Participants

Fifty-two students (26 female) from McGill University, whose native language was Canadian English, participated in the study. Participants had a mean age of 23.7 years (*SD* = 5.7) with 15.6 (*SD* = 1.9) mean years of education.

2.2. Stimuli

The prime stimuli were short sentences (approximately 7-10 syllables in length) spoken in English by two female and two male speakers. These utterances were produced to express *happiness*, *sadness*, or *neutral* affect. In the Prosody Task, the prime stimuli were pseudo-utterances (e.g., *Someone migged the pazing.*) which contained appropriate emotional intonation, but no meaningful semantic context. The pseudo-utterances were created by replacing the content words in semantic sentences with meaningless but phonologically valid sounds, ensuring that while sounding similar to language, these sentences lacked semantic information. In the Semantic Task, the prime stimuli had a distinct, meaningful emotional semantic context (e.g., *They accepted me idea!*), but were spoken with neutral prosody. In the Prosody-Semantic Task, primes were spoken with congruent emotional prosody and semantic context.

The target stimuli were color photographs of three male and three female faces. Half of the face targets were "real" emotional expressions (*happy* or *sad*) and the other half were "grimaces" that involved movements of the face that did not convey a discrete emotion.

Prior to this study, both prime and target stimuli were perceptually validated through several pilot tests. Stimuli were judged by participants who did not take part in the current study. The emotionally intoned pseudo-utterances were correctly recognized at a minimum rate of 70% by 24 listeners, and the semantically meaningful sentences (presented in written format) at a minimum rate of 90%, by 20 raters. Furthermore, the sentences in the Semantic Task were judged to be prosodically neutral by 16 listeners on a 5-point positive-negative valence scale. Finally, the emotional face targets were recognized by 32 listeners at a minimum rate of 78% and the grimaces were recognized as not emotions at a minimum rate of 60%.

2.3. Experimental Task & Procedure

Each of the three tasks contained 144 trials with the appropriate prime stimuli paired with face targets. Within each task, each happy, sad, and neutral prime stimulus was paired with one happy face target, one sad face target, and two facial grimaces. For those trials consisting of "true" emotional face targets, the prime-target relationship was defined as congruent (*happy-happy* or *sad-sad*), incongruent (*sad-happy* or *happy-sad*), or neutral (*neutral-happy* or *neutral-sad*). Neutral trials were used primarily as filler items to prevent participants from engaging in strategic processing. Prime and target stimuli were displayed using Superlab presentation software on a laptop computer.

In a quiet testing room, participants passively listened to each prime stimulus through stereo headphones and subsequently responded yes/no whether the face target represented a real emotional expression. The face targets were always presented immediately after the end of the prime sentence. Participants were instructed to ignore the auditory stimulus and to focus solely on the facial judgment. Response times and accuracy in judging the face targets were recorded. Participants completed the three tasks in two sessions (two tasks during the first session, the third task during the second session) separated by a one week interval. The order the tasks were presented was counterbalanced across participants. Within each task, prime-target stimuli were divided into blocks, each containing a similar number of male/female and true/false face targets. The order that blocks were presented was also counterbalanced across participants. Upon the completion of all three tasks, participants were compensated $30 CAD.

3. Results

One male participant displayed abnormally high error rates (Prosody Task: 35.4%; Semantic Task: 33.3%; Prosody-Semantic Task: 33.3%) and these data were subsequently excluded from further analyses. To evaluate priming effects, analyses of response time data included only correct responses to real face targets. For these analyses, an additional six participants who had an error rate higher than 25% were excluded. Furthermore, to eliminate extreme values, response times less than 300 ms and greater than 2000 ms were eliminated from subsequent analyses. For each participant, values greater than two standard deviations from their mean were replaced by the value corresponding to two standard deviations.

3.1. Accuracy

Analysis of accuracy rates considered data from 51 participants. A 3 x 2 ANOVA was conducted with Task (Prosody, Semantic, Prosody-Semantic) and Prime-Target Relationship (congruent, incongruent) as repeated measures. The ANOVA revealed a significant main effect of Prime-Target Relationship, *F*(1, 50) = 6.57, *p* = .01. Participants responded more accurately when the emotion conveyed by the prime was congruent with the emotion expressed by the face (*M* = 91.7%), as compared to when the prime and target were incongruent (*M* = 87.6%). There was no significant main effect of Task, *F*(2, 100) = 1.27, *p* = .29, and no interaction between Task and Prime-Target Relationship, *F*(2, 100) = 1.54, *p* = .22.

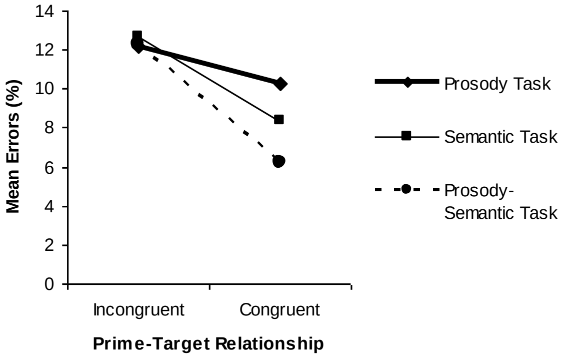

Figure 1: *Mean errors (%) by Task and Prime-Target Relationship*

3.2. Response Times

Analysis of response times considered data from 45 participants. A similar 3 x 2 ANOVA was conducted with Task (Prosody, Semantic, Prosody-Semantic) and Prime-Target Relationship (congruent, incongruent) as repeated measures. This analysis revealed a significant main effect of Prime-Target Relationship, *F*(1, 44) = 25.34, *p* < .001. Participants were overall faster to respond to emotionally congruent (*M* = 603 ms) compared to incongruent (*M* = 620 ms) prime-target pairs. Furthermore, there was a significant main effect of Task, *F*(2, 88) = 5.09, *p* = .01. Post hoc Tukey's HSD tests revealed that regardless of prime-target relationship, facial affect decisions in the Semantic Task (*M* = 598 ms) and Prosody-Semantic Task (*M* = 606 ms) were significantly faster than responses in the Prosody Task (*M* = 631 ms). There was no significant difference between the Semantic Task and the Prosody-Semantic Task. Additionally, there was no significant interaction between Task and Prime-Target Relationship, *F*(2, 88) = 1.21, *p* = .31.

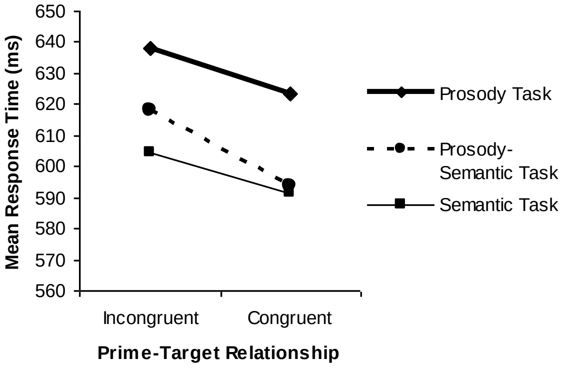

Figure 2: *Mean response time (ms) by Task and Prime-Target Relationship*

4. Discussion

In the present study, we attempted to clarify the relative contributions of prosody and semantic context in activating emotional information using the Facial Affect Decision Task. We varied the prime to contain only prosody, only semantic context, and combined prosody and semantic context, to investigate how these cues would differentially affect accuracy and response times to congruent and incongruent facial expressions. The results demonstrated that facial affect decisions were significantly faster and more accurate when the emotion conveyed by the prime was congruent with the facial emotion. Notably, there was no difference in this emotion congruency effect across tasks. This suggests that although prosody and semantics are fundamentally different speech cues, both enable the activation of emotional information that similarly speed reactions to a congruent emotional face.

The Prosody-Semantic Task included congruent prosody and semantic context that was used to evaluate whether the two cues in tandem would have additive or redundant effects. However, response times and accuracy was not significantly better with the presence of both prosody and semantic context. This result was somewhat surprising given prior findings have pointed towards an additive advantage of both cues together [1, 13]. However, such an advantage may be more apparent in emotion recognition tasks that require participants to consciously attend to the emotion and assign verbal labels, as presumably the two congruent cues would facilitate heightened confidence in the judgment. This advantage may not be as robust in implicit tasks such as the FADT used here.

Based on previous research, we had considered the possibility that semantic information may be a “stronger” cue than prosody [3, 6], and thus lead to a greater emotion congruency effect. We did find that both tasks with semantically meaningful information (Semantic Task, Prosody-Semantic Task) led to overall faster response times, regardless of congruency/incongruency between the prime and target. This finding suggests that the presence of a meaningful semantic context may generally speed response times. Alternatively, it must be considered that response times in the Prosody Task were slowed by the presence of pseudo-sentences as they may place unique demands on processes involved in lexical-grammatical processing [12]. However, this finding was not tied to the congruency or incongruency of the prime and target, implying that both cues are adequate in facilitating priming effects when the prime and target convey the same emotional meaning.

5. Conclusion

In general, we found that both prosody and semantic context are relevant cues for implicitly processing emotional information from speech and that both cues subsequently facilitate judgment of facial emotion. This finding highlights that underlying features of prosody, semantics, and faces are shared or common across channels, and that emotional information activated through one channel is important in processing information from a different channel. Importantly, prosody appears to be relatively similar in strength to semantic context in implicitly activating emotional information stored in memory, which guides responses to emotional facial expressions.

6. Acknowledgements

This research was funded by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (to M.D.P.) and by the German Research Foundation (DFG FOR 499 to S.A.K.). We thank Nicole Price and Catherine Knowles for help with data collection.

7. References

[1] Beaucousin, V., Lacheret, A., Turbeline, M.-R., Morel, M. I., Mazoyer, B. & Tzourio-Mazoyer, N. 2007. FMRI study of emotional speech comprehension. *Cerebral Cortex* 17(2). 339-352.

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

[3] Breitenstein, C., Daum, I. & Ackermann, H. 1998. Emotional processing following cortical and subcortical brain damage: contribution of the fronto-striatal circuitry. *Behavioural Neurology* 11. 29-42.

[4] 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.

[5] Mitchell, R. L. C., Elliot, 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.

[6] 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.

[7] Pell, M. D. 2005. Nonverbal emotion priming: evidence from the ‘facial affect decision task’. *Journal of Nonverbal Behavior* 29(1). 45-73.

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

[9] Pell, M. D., Jaywant., A., Monetta, L. & Kotz, S. A. In review. Emotional speech processing: Disentangling the effects of prosody and semantics.

[10] 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.

[11] Schirmer, A., Kotz, S.A. & Friderici, A. (2002). Sex differentiates the role of emotional prosody versus word processing. *Cognitive Brain Research, 14*, 228-233.

[12] Shuster, L. I. 2009. The effect of sublexical and lexical frequency on speech production: an fMRI investigation. *Brain and Language* 111(1). 66-72

[13] 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.