

Categorizing Emotion in Spoken Language: An Analysis of Semantic and Prosodic Contributions to Emotional Communication

Janine K. Fitzpatrick (jfitzpat@connect.carleton.ca)

Institute of Cognitive Science, Carleton University
Ottawa, ON

John Logan (john_logan@carleton.ca)

Institute of Cognitive Science and Department of Psychology, Carleton University
Ottawa, ON

Abstract

The current study aimed to replicate and expand upon research conducted by Bagley, Abramowitz and Kosson (2009) to examine categorization of emotional sentences among non-psychopathic individuals. 36 monolingual English-speaking undergraduate participants categorized spoken English sentences (produced with neutral prosody but containing semantic cues to emotion) and French sentences (produced with appropriate prosody but with no semantic cues to emotion) into one of five emotion categories: *happiness, sadness, anger, fear, or neutral*. By isolating the semantic and prosodic information available to listeners, we determined that categorization accuracy was higher among sentences expressing anger in the prosodic condition. Accuracy was higher among sentences expressing all other emotions in the semantic condition. Overall, the lowest categorization accuracy was found for sentences expressing fear in the prosodic condition. Across all emotion categories and both presentation conditions, reaction time was longest for sentences expressing fear in the prosodic condition. Although all participants in the current study had normative scores on the Self-Report Psychopathy Scale, those with relatively high scores displayed lower categorization accuracy for semantic sentences expressing happiness, anger and fear than lower-scoring participants. An extension of the current study comparing this normative sample to a group of individuals with psychopathy will need to account for possible implications of subclinical psychopathic characteristics on vocal affect categorization accuracy.

Keywords: emotion; language processing; psychopathy; SRP-III.

We are able to identify the emotional content of spoken language based on two types of cues: semantic cues, the content and meaning of what is being said, and prosodic cues, the patterns of pitch, amplitude and duration of speech associated with particular emotions. The current study was conducted to investigate the contributions of semantic and prosodic information towards listeners' ability to identify the emotions expressed in spoken sentences. Currently, little is known about emotion categorization in the broader field of cognitive science; the present study was designed to add to the relatively limited knowledge in this area. In addition, the emotion categorization data collected in the present study will provide a normative comparison to data from

individuals with psychopathic characteristics, a group known to have deficits in processing emotion (e.g., Blair, Mitchell, & Blair, 2005).

The aim of the current study was to replicate and expand upon research by Bagley, Abramowitz and Kosson (2009), who investigated emotional language processing in individuals with primary and secondary psychopathy compared to individuals without psychopathy. Their experimental design involved isolating the semantic and prosodic cues present in spoken language by presenting monolingual English listeners with English sentences spoken neutrally (i.e., semantic cues present but minimal prosodic cues) and Bulgarian sentences spoken with appropriate prosody (i.e., prosodic cues present but minimal semantic cues). These two conditions will be referred to as the 'semantic' and 'prosodic' conditions, respectively. The vocal affect identification task included sentences corresponding to the following emotional categories: happiness, sadness, anger, surprise and neutral content. Fearful sentences were not included in their study. Because clear deficits have been found among individuals with psychopathy when identifying fear from verbal cues (see Blair et al., 2002; Blair, Budhani, Colledge, & Scott, 2005), the inclusion of this emotion category is a natural next step in this research. Among non-psychopathic individuals, Bagley et al. found that categorization accuracy was higher in the semantic condition for sentences expressing happiness, sadness, and surprise, while neutral sentences and those expressing anger displayed higher categorization accuracy in the prosodic condition. Psychopathic individuals classified sentences less accurately than non-psychopathic individuals in the semantic condition, while differences in classification in the prosodic condition approached significance. Further, participants with middle scores on the PCL-R (Psychopathy Checklist-Revised, an instrument designed to assess psychopathy; Hare, 2003) were less accurate at identifying happiness in the semantic condition. The authors propose that even subclinical levels of psychopathic characteristics may interfere with an individual's ability to process semantic cues for happiness, while deficits in processing prosodic cues may only be found in individuals with the full psychopathy syndrome.

Bagley et al. (2009) examined how the accuracy of emotion identification from prosodic cues in Bulgarian; the prosodic condition in the current study was comprised of French sentences. It is important to determine whether the prosodic cues from another language would yield similar results of identification accuracy among monolingual English speakers. Further, although all participants included in the present study scored within the normative range of Self-Report Psychopathy Scale (SRP-III; Williams, Paulhus, & Hare, 2007), we investigated whether participants' accuracy at identifying emotions correlated with their score on the SRP-III. We also added a measure of reaction time (RT) to examine its relationship with categorization accuracy.

Also of interest was whether the discrepancy between identification of fearful sentences is as marked between psychopathic and nonpsychopathic individuals as it is for identification of other emotions. In a systematic review of empirical literature concerning emotion recognition from spoken language, Scherer, Johnstone and Klasmeyer (2003) found that the acoustic cues reported to be associated with fear are unclear. A variety of acoustic parameters (such as number and duration of pauses and F0 range) indicate inconsistent empirical findings for recognition of fearful utterances. Many prosodic cues depend on speaker-specific factors such as age and gender, but the results for fear indicate more variability across findings than any other emotion category included in the review. Scherer et al.'s results, together with the results of a pilot version of the current study in which non-psychopathic listeners also displayed lower categorization accuracy for sentences expressing fear presented in the prosodic condition, suggest that people may generally find it more difficult to recognize fear in language without a semantic context than other emotions.

The central hypotheses of the current study were as follows: First, based on the findings from Bagley et al. (2009), we hypothesized that categorization accuracy would be significantly higher for sentences expressing happiness and sadness presented in the semantic condition, and for sentences expressing anger in the prosodic condition. We predicted that categorization accuracy would be lower for sentences expressing fear than all other sentences, particularly in the prosodic condition. Second, we hypothesized that participants with higher scores on the SRP-III would display significantly lower categorization accuracy among sentences expressing fear in both semantic and prosodic conditions than lower-scoring participants. Finally, we hypothesized that reaction time would display an inverse relationship with categorization accuracy: RT would be longer in the prosodic condition, particularly for sentences expressing fear, than the semantic condition. RT would also be longer among participants with higher scores on the SRP-III than lower-scoring participants.

Method

Participants

Forty-five Carleton University undergraduate students participated in this study, 19 males and 26 females. Participant age ranged from 18-35 years ($M_{age} = 20.2$ years, $SD = 3.0$). Participants were recruited through an online database run by the Carleton University Psychology Department and received course credit for their participation in the study. 92% of participants were monolingual English speakers, while the remaining 8% were native English speakers who spoke a second language other than French. All participants had little to no proficiency in French comprehension and production, as measured by a Language Experience Questionnaire in which participants described their exposure to the French language over their lifetimes. In addition, participants completed the Self-Report Psychopathy Scale (SRP-III, Williams, Paulhus, & Hare, 2007) to ensure they represented a sample of non-psychopaths. Both the Language Experience Questionnaire and the SRP-III were presented via computer.

Materials and Design

The experimental stimuli were comprised of 93 sentences recorded by four speakers. All sentences were recorded using Praat software (Boersma & Weenick, 2010) using a headset microphone in a sound attenuated booth. The sentences represented five different emotion categories: happiness (e.g.: *All my wishes came true that day*), sadness (e.g.: *I had no money to buy Christmas gifts*), anger (e.g.: *He just smashed my new car*), fear (e.g.: *I hope they don't find me here*), and neutral (e.g.: *It's time to fill the bird feeder*). Fluently bilingual speakers, two males and two females, recorded the sentences in English and French using written scripts. They recorded English sentences with neutral prosody and French sentences with prosody appropriate to the emotion the sentences conveyed. Listeners who do not understand spoken French should not be able to understand what is being said in the French sentences, and so they would need to rely on the non-semantic cues to determine the emotion being conveyed by the speaker. Thus, there were two conditions for listeners: the English sentences comprised the semantic condition, in which only the content of the sentences was available to listeners, while the French sentences comprised the prosodic condition, in which listeners only had access to prosodic factors such as pitch contour and speech rate.

The experimental task was conducted using a 5 (emotion) x 4 (speakers) x 2 (conditions: semantic vs. prosodic) repeated measures design. Participants were presented with four sentence lists counterbalanced by affect category and speaker, with each list containing all 93 sentences presented in random order. Two of the four lists were presented in the prosodic condition, and two in the semantic condition; both lists in each language were presented together so that participants heard all the sentences in one condition followed by all the sentences in the other. In total, each participant heard 372 sentences.

Procedure

Participants listened to the sentences, presented via headphones, and indicated the affect category to which they thought the sentence belonged using a scale presented on the computer screen in which a number on the keyboard corresponded with a particular emotion. After categorizing the sentence, participants rated the degree to which they thought the sentence conveyed the emotion they had indicated, referred to as the 'quality' of the sentence, on a 7-point Likert scale (1 = *low quality*, 4 = *moderate quality*, 7 = *high quality*).

All participants completed a 10-sentence practice block to familiarize themselves with the scales used in the task and to adjust the volume of the auditory stimuli. The sentences presented during the practice block were not used during the experimental task. Participants were then presented all sentences from the four lists, taking a 5-minute break after completing the second list. All of the sentences in the first two lists were presented in the same condition, and the sentences in the third and fourth list were presented in the other language (e.g.: lists 1 and 2 were semantic, lists 3 and 4 were prosodic). This ensured that participants completed the semantic and prosodic conditions of the task without interruption. Sentences in each list were presented in a different randomized order to each participant.

Results

SRP-III results were analyzed before proceeding with analysis of the experimental task data. The mean overall score for all males was 160, $SD = 27$, and the mean overall score for females was 135, $SD = 20$. One male participant and one female participant scored above the high normative cutoff of two standard deviations above their gender's mean; their experimental data were excluded from further analyses. Among the remaining 43 participants, the mean overall score for males was 157, $SD = 24$, and the mean overall score for females was 133, $SD = 18$. Male scores ranged from 98 to 194, and female scores ranged from 103 to 171.

Overall, the mean proportion categorization accuracy was .58 across all emotion categories, $SD = .16$. This is significantly higher than chance levels of accuracy, $\chi^2(4, n = 43) = 431.51, p = .00$. An accuracy cutoff criterion was set at .40, twice the level of chance. Data from seven participants whose overall accuracy was less than .40 were excluded from further analysis. Further, seven sentences were removed from the final data set because they displayed a mean categorization accuracy of more than two standard deviations below the means of their respective emotion and condition groups. One sentence in each emotion category was removed from the semantic condition, and a sentence expressing happiness and one expressing sadness were removed from the prosodic condition. Among remaining participants and sentences, mean categorization accuracy was .64, $SD = .11$. Figure 1 displays the mean proportion of sentences accurately categorized by participants as a function of the emotion expressed by the sentence and the

condition in which they were presented (semantic or prosodic). Generally, participants were better at categorizing the emotion of the sentences they heard in the semantic (English) condition than the prosodic (French) condition. All means reported represent the mean proportion of accurate responses. Only the sentences conveying anger were more accurately categorized in the prosodic condition ($M_{\text{prosodic}} = .75, SD = .16$; $M_{\text{semantic}} = .68, SD = .24$). Sentences conveying fear showed the largest discrepancy between categorization accuracy in the semantic and prosodic conditions ($M_{\text{semantic}} = .74, SD = .21$; $M_{\text{prosodic}} = .23, SD = .17$).

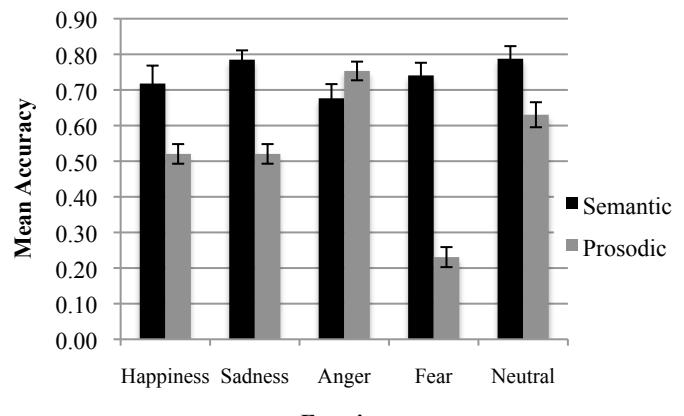


Figure 1. Mean accuracy for sentence categorization by emotion in semantic and prosodic conditions. Error bars represent ± 1 standard error of the mean.

The relationship between sex and categorization accuracy of sentences in all emotions and conditions was not significant. A repeated-measures analysis of variance (ANOVA) was conducted to determine the effects of the sentences' intended emotion and condition on categorization accuracy. Significance was set at $p = .05$ for all tests. The results indicate significant main effects of both emotion and condition on categorization accuracy, $F(2.77, 97.09) = 23.59$ and $F(1, 35) = 26.43$, respectively. Contrasts revealed that categorization accuracy for sentences expressing happiness, $F(1, 35) = 5.11, r = .36$, and sentences expressing fear, $F(1, 35) = 59.58, r = .79$, were significantly lower than accuracy of neutral sentences. Sentences in the prosodic condition, $F(1, 35) = 26.43, r = .66$, displayed significantly lower mean accuracy than sentences in the semantic condition.

There was a significant interaction between emotion and condition on accuracy, $F(4, 140) = 43.82$. To further investigate this interaction, contrasts were performed to compare all emotion categories to their baseline (neutral) and sentences in the prosodic condition to those in the semantic condition. The contrasts revealed significant interactions when comparing sentences expressing anger, $F(1, 35) = 24.00, r = .64$; and fear, $F(1, 35) = 46.85, r = .76$, to neutral sentences. These effects reflect that, compared to neutral sentences, categorization accuracy of sentences

Because males and females were grouped into high and low-score categories based on different median scores, regression analyses were performed to determine the relationship between SRP-III category (high or low), sex and categorization accuracy (see Table 4).

Table 4
Categorization accuracy for all sentences by SRP-III score and sex

Variable	B (SE)	OR	95% CI	
			Lower	Upper
Constant	.33 (.03)	1.39		
SRP-III	.43 (.04)	1.54**	1.43	1.66
Sex	.07 (.04)	1.08	1.00	1.16

Note. OR = odds ratio; CI = confidence interval. ** $p < .01$.

When controlling for the influence of sex, SRP-III score is a significant predictor of overall categorization accuracy, with low scoring participants 1.5 times more likely to respond correctly than high scoring participants. Sex was not found to be a significant predictor of categorization accuracy when controlling for SRP-III score.

Discussion

The accuracy results generally support our hypotheses: Categorization accuracy was higher in the semantic condition for all emotion categories except for sentences expressing anger, which were more accurately categorized in the prosodic condition. Scherer et al. (2003) describe several acoustic parameters associated with utterances expressing anger (compared to neutral utterances), including increased F0 mean and range, higher mean voice source intensity (dB), and increased frequency of accented syllables. Detailed analysis of acoustic profiles of each sentence used in the present study was conducted as part of a separate project and will not be discussed further.

Categorization accuracy was significantly lower among sentences expressing fear in the prosodic condition than all other sentences. This is in accordance with our hypotheses as well as Bagley et al.'s (2009) results. Participants with higher SRP-III scores displayed significantly lower accuracy than low-scoring participants when identifying fear in sentences presented in the semantic condition, but not the prosodic condition. However, because overall accuracy for sentences expressing fear in the prosodic condition was significantly lower than all other sentences, group comparisons between high and low-scoring participants in this category are likely not as meaningful as group comparisons for other emotions. All participants were deemed to be non-psychopathic based on their scores. An extension of this study will examine accuracy data among individuals with psychopathy; data indicating that a normative population displays lower categorization accuracy for prosodic sentences expressing fear will be useful when interpreting the results of the individuals with psychopathy. However, discrepancies between categorization accuracy among subclinical individuals with

relatively high and low SRP-III scores will need to be taken into account when comparing normative accuracy levels between normative and psychopathic populations. Vassileva, Kosson, Abramowitz, and Conrod (2005) describe two distinct subgroups of psychopathy: primary psychopathy, characterized by higher scores on interpersonal and affective items on the PCL-R (Hare, 2003) and the Interpersonal Measure of Psychopathy (IM-P, an additional measure of the personality core of psychopathy; Kosson, Stuerwak, Forth, & Kirkhart, 1997); and secondary psychopathy, characterized by higher scores on the antisocial items on the PCL-R and increased severity of alcohol and drug dependence. Kosson et al. (2009) treated psychopathy as a heterogeneous construct and examined differences between primary and secondary psychopaths in vocal affect recognition; results from the current study indicate that 'normative' individuals may need to be treated as a heterogeneous group as well. Subclinical levels of psychopathic characteristics may be implicated in difficulties in emotional sentence processing, and research designed to compare categorization differences between psychopathic and non-psychopathic individuals will need to take this relationship into account.

Several theories have emerged concerning emotional response and regulation among psychopathic individuals. The dysfunctional fear hypothesis suggests that individuals with psychopathy show less aversive reactions to punishment than non-psychopathic individuals (see Blair et al., 2005). Lykken (1957) demonstrated that psychopathic individuals demonstrated less avoidance of punished responses (a harmless but painful electric shock) in a maze-learning task and less galvanic skin response reactivity to a conditioned stimulus associated with shock than controls. These results may provide some rationale concerning psychopathic individuals' impairment of fear recognition; if they are less adept at learning and responding to fear responses on their own, they may also be less likely to recognize signals of fear in others.

The current experimental task relies upon discrimination instead of recognition of emotions, thereby reducing its ecological validity. One risk of isolating semantic and prosodic cues is the potential use of response bias in the face of ambiguity, as discussed by Johnstone and Scherer (2000): When presented with ambiguous cues in experimental stimuli, participants may be more likely favour one response over another. This preference may reflect a response heuristic not otherwise utilized during emotional processing in a natural setting. After examining response data, we excluded several sentences from analysis because they displayed significantly lower categorization accuracy than the other sentences in their emotion category. Among the excluded sentences, several were semantically ambiguous and/or context-specific (e.g.: *Use your signal when you're switching lanes!* was intended to express anger, but when heard in the semantic condition with neutral prosody it could easily be interpreted as a neutral sentence). Indeed, many of these semantically ambiguous sentences received

more neutral categorizations than sentences with clearer meaning. The degree to which each sentence accurately reflects its intended emotion can be drawn from further examination of confusion data, which will be helpful when deciding which sentences to use in future iterations of this study.

The stimuli used in the present study were intended to simulate emotion in speech. Although obtaining voice samples from professional or lay actors has been the preferred method in the field, Scherer (2003) suggests that actors may miss the more subtle cues of natural emotional speech in favour of obvious, stereotypical ones. Scherer et al. (2003) describe the push and pull effects inherent in the expression of emotion: Push effects are physiological changes that characterize emotional responses in speech production, while pull effects reflect the notion that vocalization is often regulated and monitored in order to fit into conventional expression norms. For example, in social groups in which expression of anger is deemed unattractive, pull effects would cause the inhibition of some of the characteristic cues of anger expression in order to better fit convention. Thus, while the use of actor portrayals may exaggerate certain cues of emotion expression and provide 'stereotypical' voice samples, it is presumed that pull effects will be less influential in the simulated setting than in samples derived from natural vocal expression and induced emotions.

In the semantic condition of encoding, actors were told to speak in a neutral voice in order to minimize any prosodic cues available for interpretation by the listeners. It is nearly impossible to ensure the actor's voice is completely neutral. Researchers wishing to further study the disparate contributions of semantic and prosodic information to emotional speech processing should examine the relative advantages and drawbacks of natural vs. synthetic speech samples (see Scherer, 2003).

Because sentence length was not standardized across speaker, emotion and condition, it is difficult to interpret the results of RT analysis. In order to effectively compare RT to accuracy data, alternative strategies for dealing with this variability in sentence duration would need to be adopted. Further analysis can be conducted by removing individual sentence length from each RT value to determine whether observed patterns still hold.

The present study was conducted to examine the effects of isolating semantic and prosodic speech cues on emotional language processing among a normative sample of listeners. The results indicate that semantic cues may be more heavily implicated in categorization accuracy for sentences expressing happiness, sadness and fear, while listeners may use more prosodic cues to identify sentences expressing anger. In a broad sense, the present study is an example of extending categorization research beyond the categorization of physical objects into more abstract types of events. Emotion research is an area that cognitive science has tended to avoid. By employing empirical methods in the study of emotion, this lack of attention can be remediated.

In addition to the theoretical benefits of exploring the categorization of emotions, the present work also has clinical implications. By studying categorization data from a clinically psychopathic population, we may learn more about the nature of the deficits associated with emotional processing in psychopathic individuals.

Acknowledgments

We are indebted to David Kosson for the stimuli used in this study, and to Erin Deboot for her help with designing and running the experimental task.

References

Bagley, A. D., Abramowitz, C.S., & Kosson, D.S. (2009). Vocal affect recognition and psychopathy: Converging findings across traditional and cluster analytic approaches to assessing the construct. *Journal of Abnormal Psychology, 118* (2), 388-398.

Blair, R.J.R., Budhani, S., Colledge, E., & Scott, S. (2005). Deafness to fear in boys with psychopathic tendencies. *Journal of Child Psychology and Psychiatry, 46* (3), 327-336.

Blair, J., Mitchell, D., & Blair, K. (2005). *The psychopath: Emotion and the brain*. Oxford: Blackwell Publishing.

Blair, R.J.R., Mitchell, D.G.V., Richell, R.A., Kelly, S., & Leonard, A. (2002). Turning a deaf ear to fear: Impaired recognition of vocal affect in psychopathic individuals. *Journal of Abnormal Psychology, 111* (4), 682-686.

Boersma, P., & Weenick, D. (2010). Praat: Doing phonetics by computer (Version 5.2.05) [Computer program]. Retrieved from <http://www.praat.org/>

Hare, R. D. (2003). *The Hare Psychopathy Checklist – Revised*, 2nd Edition. Toronto: Multi-Health Systems.

Johnstone, T., & Scherer, K. R. (2000). Vocal communication of emotion. In M. Lewis and J. Haviland (Eds.), *The Handbook of Emotion*. New York: Guilford.

Kosson, D. S., Stuerwald, B. L., Forth, A. E., & Kirkhart, K. J. (1997). A new method for assessing the interpersonal behavior of psychopathic individuals: Preliminary validation studies. *Psychological Assessment, 9*, 89-101.

Lykken, D. T. (1957). A study of anxiety in the sociopathic personality. *Journal of Abnormal and Social Psychology, 55*, 6-10.

Scherer, K. R., Johnstone, T., & Klasmeyer, G. (2003). Vocal expression of emotion. In R. J. Davidson, K. R. Scherer, and H. Goldsmith (Eds.), *Handbook of the Affective Sciences*. New York and Oxford: Oxford University Press.

Vassileva, J., Kosson, D. S., Abramowitz, C., & Conrod, P. (2005). Psychopathy versus psychopathies in classification of criminal offenders. *Legal and Criminological Psychology, 10*, 27-43.

Williams, K., Paulhus D., & Hare, R.D. (2007). Capturing the four-factor structure of psychopathy in college students via self-report. *Journal of Personality Assessment, 88* (2), 205-219.