Research Article

Prosody and Semantics Are Separate but Not Separable Channels in the Perception of Emotional Speech: Test for Rating of Emotions in Speech

Boaz M. Ben-David,[a,b,c] **Namita Multani,**[b] **Vered Shakuf,**[a]
Frank Rudzicz,[b,c] **and Pascal H. H. M. van Lieshout**[b,c,d,e,f]

Purpose: Our aim is to explore the complex interplay of prosody (tone of speech) and semantics (verbal content) in the perception of discrete emotions in speech.
Method: We implement a novel tool, the Test for Rating of Emotions in Speech. Eighty native English speakers were presented with spoken sentences made of different combinations of 5 discrete emotions (anger, fear, happiness, sadness, and neutral) presented in prosody and semantics. Listeners were asked to rate the sentence as a whole, integrating both speech channels, or to focus on one channel only (prosody or semantics).
Results: We observed supremacy of congruency, failure of selective attention, and prosodic dominance. Supremacy of

congruency means that a sentence that presents the same emotion in both speech channels was rated highest; failure of selective attention means that listeners were unable to selectively attend to one channel when instructed; and prosodic dominance means that prosodic information plays a larger role than semantics in processing emotional speech.
Conclusions: Emotional prosody and semantics are separate but not separable channels, and it is difficult to perceive one without the influence of the other. Our findings indicate that the Test for Rating of Emotions in Speech can reveal specific aspects in the processing of emotional speech and may in the future prove useful for understanding emotion-processing deficits in individuals with pathologies.

T he identification of emotions in speech is essential to spoken communication. An impaired ability to identify and describe emotions has been associated with lower quality of life (Joukamaa, Saarijärvi, Muuriaisniemi, & Salokangas, 1996), significant feelings of depression, and reduced relationship well-being (Carton, Kessler, & Pape, 1999). In order to partake effectively in a

social dialogue, one has to understand the verbal expression of emotions in both their semantics (lexical content of the words) and prosody (tone of speech, i.e., rhythm, stress, and intonation). The ability to decipher this complex interplay of prosody and semantics, and to selectively attend to one or the other, is an essential element of daily communication (Mitchell & Ross, 2008).

Given its central role in adaptive social interactions, a growing body of research focuses on the perception of emotions in speech (Ben-David, van Lieshout, & Leszcz, 2011; Kotz & Paulmann, 2007; Paulmann, Pell, & Kotz, 2008; Pell, Paulmann, Dara, Alasseri, & Kotz, 2009; for a review, see Juslin & Laukka, 2003). However, as Pell, Jaywant, Monetta, and Kotz (2011) recently indicated, "empirical data that inform the relative degree to which listeners harness prosody versus semantic cues, or information from both channels in combination, to activate and retrieve emotional meanings during ongoing speech processing, are scarce" (p. 835). The current study presents a novel comprehensive paradigm, the Test for Rating of Emotions in Speech (T-RES), that examines the interplay between

[a]Baruch Ivcher School of Psychology, Interdisciplinary Center (IDC), Herzliya, Israel
[b]University of Toronto, Ontario, Canada
[c]Speech-Language Pathology, Toronto Rehabilitation Institute, University of Toronto, Ontario, Canada
[d]University of Toronto Mississauga, Ontario, Canada
[e]Institute of Biomaterials and Biomedical Engineering, University of Toronto, Ontario, Canada
[f]Rehabilitation Sciences Institute, University of Toronto, Ontario, Canada

Correspondence to Boaz M. Ben-David: boaz.ben.david@idc.ac.il

prosody and semantics and its impact on the perception of emotions in spoken sentences.

Perception of Spoken Emotions

Difficulties identifying emotions in speech can arise from failures in understanding the emotional meaning expressed in one of two speech channels, semantics (e.g., as in people with brain injury; Cicero et al., 1999) or prosody (e.g., as in older adults; Dupuis & Pichora-Fuller, 2010). Isolated impairment of one channel may be linked to the notion that different networks in the brain are activated in the processing of prosody and semantics (Bowers, Bauer, & Heilman, 1993; Ethofer et al., 2006; Grimshaw, Séguin, & Godfrey, 2009; Mitchell, Elliott, Barry, Cruttenden, & Woodruff, 2003; Wildgruber, Ackermann, Kreifelts, & Ethofer, 2006). In the alternative, incorrect identification may arise from a biased integration of the two sources, giving one speech channel a more dominant role than intended by the speaker (Paulmann et al., 2008). Whether listeners are more attuned to the prosody (prosodic dominance) or to the semantics (semantic dominance) has thus far not been conclusively answered, as described in the following sections.

Prosodic Dominance

An early foray into the interaction of semantics and prosody by Mehrabian and Wiener (1967) supports prosodic dominance in the perception of emotions in speech. Listeners were presented with positive, negative, and neutral single words, spoken with positive, negative, and neutral prosody, creating congruent combinations (e.g., negative words spoken with negative prosody), incongruent combinations (e.g., negative words spoken with positive prosody), and neutral combinations (e.g., negative words spoken with a neutral prosody). Emotional ratings were affected by both speech channels. However, when listeners were asked to attend to both, prosody had a larger impact than semantics. Similar results were obtained when listeners were asked to focus on one channel only.

Over the 45 years since that study, there is only scant evidence that directly supports prosodic dominance in spoken emotions. However, several recent studies have demonstrated the impact of prosody on the processing and interpretation of affective spoken language. For example, Jacob, Brück, Plewnia, and Wildgruber (2014; see also Morton & Trehub, 2001) showed that listeners rely on prosody to disambiguate the emotional message of an incongruent combination of semantics and prosody in spoken words. Nygaard and Lunders (2002) found that prosodic cues help resolve the ambiguity of the semantics of single spoken words (using homophones, e.g., *die–dye*). Prosody has also been found to speed up naming responses to spoken words, suggesting that it is processed even when listeners are attending to semantics (Nygaard & Queen, 2008). A recent study by Roche, Petters, and Dale (2014) provides more indirect support for prosodic dominance. Despite large talker variability, listeners were able to identify and categorize the talkers' intent on the basis of only prosodic

information. On the converse, other research suggests semantic dominance in the perception of emotions in speech, as discussed next.

Semantic Dominance

A series of studies by Kitayama, Ishii, and their colleagues (Ishii, Reyes, & Kitayama, 2003; Kitayama & Ishii, 2002) highlights the primacy of semantics in Western cultures. In a paradigm similar to that of Mehrabian and Wiener's original study (1967), spoken words with incongruent semantics and prosody slowed down prosodic categorization (as compared to categorization of congruent combinations) to a larger extent than semantic categorization. In other words, it was more difficult for listeners to ignore the irrelevant information presented in the semantic domain (when asked to focus on the prosody) than to ignore the prosody (when asked to focus on the semantics). The authors concluded that members of Western cultures (the majority of participants in published studies, as well as the population tested in the current study; Jones, 2010) are "attentionally biased toward verbal content" (Ishii et al., 2003, p. 44). Other studies examining emotional (Kotz & Paulmann, 2007; Paulmann & Kotz, 2008) and nonemotional (Astésano, Besson, & Alter, 2004) prosody and semantics have shown similar results—the rating of prosody was affected by the semantics to a larger extent than vice versa.

Integration of Semantic and Prosodic Content in Speech

There is clearly no conclusive evidence on the dominance of semantics versus prosody in speech. In contrast, there is ample evidence of the supremacy of congruency. That is, congruent combinations of semantics and prosody are processed faster, with greater accuracy, and yield higher ratings than neutral or conflicting ones (Beaucousin et al., 2007; Mitchell, 2006; Nygaard & Queen, 2008; Wurm, Vakoch, Strasser, Calin-Jageman, & Ross, 2001). For example, in Mehrabian and Wiener's original study (1967), the highest ratings of positivity were recorded when both prosody and semantics were positive. In a similar vein, Ishii et al. (2003) reported faster categorization responses for congruent combinations than for incongruent ones. This effect was maintained regardless of whether listeners were asked to focus on the prosody or on the semantics. The difference between congruent and incongruent combinations is further supported by neuroimaging studies, suggesting that congruent combinations of semantics and prosody (redundancy of information) are processed earlier (Paulmann, Jessen, & Kotz, 2009) and in a different manner than incongruent ones (Wambacq & Jerger, 2004; Wittfoth et al., 2010).

This supremacy of congruency could arise from the advantage of presenting the same information in both channels (for a discussion, see Ben-David & Algom, 2009). An examination of the data presented by Ishii et al. (2003, Table 6, p. 44) shows more accurate and faster identification responses when the same emotion is presented in both channels than when it is presented only in one channel, for both Westerners and members of Asian cultures. On the other hand, Pell and colleagues (Pell et al., 2011; Schwartz

& Pell, 2012) have shown that simple redundancy may not be the only factor in explaining the advantage of congruency. They found that the effect of priming of emotions (the impact of a passive exposure to spoken emotions on performance in a following task—categorizing emotional facial expression) was not increased by presenting the primed emotion in both channels rather than in only one. One possible reason for this inconsistency may be related to the use of different methodologies, as highlighted in the following section.

Assessment of Spoken Emotions

Often, the assessment of the respective roles of prosody and semantics in the perception of emotions in speech has been conducted with spoken single words, using two discrete emotions (one positive and one negative). For example, Ishii et al. (2003) presented listeners with pleasant (*pretty*) and unpleasant (*sore*) words spoken in two vocal tones (harsh and constricted or smooth and round). Listeners were asked to judge the words as pleasant or unpleasant on the basis of the prosody or the meaning. Natural speech clearly utilizes longer utterances than single words, as well as a wider variety of emotions (cf. Paulmann & Pell, 2011). Indeed, current research conducted in the field uses full utterances (e.g., Kotz, Dengler, & Wittfoth, 2015; Schwartz & Pell, 2012). Therefore, our focus will be on paradigms that use complete sentences rather than single words.

There are several paradigms for the assessment of processing of emotions in spoken sentences. Two commonly used tools are the Florida Affect Battery (FAB; Bowers, Blonder, & Heilman, 1999) and the Diagnostic Analysis of Nonverbal Accuracy, Version 2 (DANVA2; Nowicki & Duke, 1994). The FAB measures the processing of emotions in facial and prosodic domains under various task demands. With respect to semantics, there are only congruent and incongruent combinations of semantics and prosody. The DANVA2 was designed as a screening tool, to help identify individuals who have trouble understanding nonverbal emotional information. It measures individual differences in the ability to decipher cues such as facial expression, posture, gesture, and prosody across four emotions (anger, happy, fear, and sad), using a single emotionally ambiguous lexical sentence.

Both the FAB and DANVA2 are highly valuable for examining nonverbal abilities to process emotion, but they are somewhat limited in their ability to reveal the full complexity of the interaction of emotions conveyed in both semantics and prosody as used in everyday speech. To our knowledge, the sentences used in these tests were not equated for their linguistic characteristics, such as average word frequency and number of syllables, across the emotional categories. These characteristics have been found to affect the time course of cognitive processes in following spoken language, as shown in the visual word paradigm (Ben-David, Nguyen, & van Lieshout, 2011; Tanenhaus, Magnuson, Dahan, & Chambers, 2000). Moreover, the FAB and DANVA2 present unequal combinations of emotions conveyed by the prosodic and semantic channels, resulting in different set sizes of emotions presented on each channel. For example, the FAB presents two different prosodies, whereas semantics vary across more than two discrete emotions. Psychophysical analysis indicates that this type of unequal stimuli combination can create a bias in responses (Melara & Algom, 2003). In addition, both tests lack baseline-neutral combinations (e.g., a semantically neutral sentence spoken with emotional prosodies) to gauge performance on one channel with limited interference from the other.

As a final point, these two standard tests use a forced-choice categorization response, where the listener is asked to choose which of the predefined emotions best describes the utterance. This restrictive mode of response simplifies analysis and use. However, it provides a limited vantage point, because no information can be obtained on the options rejected by the listener. For example, in a forced-choice paradigm the listener may respond *sad* to semantically angry sentences spoken with sad prosody. No information is available on the listener's perception of anger. Apart from these two tools, various studies have generated their own sets of spoken sentences (e.g., Pell, Monetta, Paulmann, & Kotz, 2009).

The Current Study

The goal of the current study is to explore the complex interplay of prosody and semantics in the perception of discrete emotions in spoken sentences using a novel tool, the Test for Rating of Emotions in Speech (T-RES). This tool was designed to complement existing tools using a rating scale rather than a forced-choice paradigm. For the T-RES, we created a new set of validated spoken sentences, equated for linguistic characteristics, that present an equal number of discrete emotions (anger, fear, sad, happy, and neutral) for each of the two speech channels (Ben-David, Thayapararajah, & van Lieshout, 2013; Ben-David, van Lieshout, & Leszcz, 2011). The T-RES is composed of several subtests, measuring the extent to which each channel (prosody and semantics) affects the rating of the other and how the two combine to generate a perception of emotions in spoken language. Data were collected from a sample of 80 young adults to provide important reference values, and results were examined in light of the prosodic and semantic dominance hypotheses, specifically focusing on the mechanism underlying the supremacy for congruent spoken sentences.

To sum, the aim of this study is to validate and test the efficacy of a new tool: the T-RES. We implement this novel tool to explore the processes underlying emotional speech perception addressing three main questions: (a) Selective attention: Can listeners completely ignore one of the speech channels and selectively attend to the other? (b) Channel dominance: Which of the two speech channels, prosody (the tone of voice) or semantics (the lexical content of the words), plays a larger role in the decoding of emotions in speech? (c) Supremacy of congruency: Are congruent spoken sentences, which present the same emotion in both speech channels, rated higher on emotional scales than

any other combination? In general, across the themes, we will discuss the nature of the relationship between the two channels and how they integrate to generate the perception of emotional speech.

Method

Participants

Eighty young adults (52 female, 28 male; M = 19.1 years, SD = 1.4), all undergraduates at the University of Toronto Mississauga, participated in this study. They received either a course credit or $10 per hour for their participation. All participants were native English speakers, as assessed by a self-report and a minimum score of 9/20 on the Mill Hill Vocabulary Scale (Raven, 1965)—corresponding to typical vocabulary levels of native English speakers (Ben-David, Nguyen, & van Lieshout, 2011; Ben-David & Schneider, 2010; Ben-David, Tse, & Schneider, 2012; Ben-David, van Lieshout, & Leszcz, 2011)—with a mean score of 12.4/20 (SD = 2.1/20). A random half of the participants also completed the vocabulary test of the Wechsler Adult Intelligence Scale–Fourth Edition (Wechsler, 2008) and achieved an average score of 41.7/60 (SD = 6.8/60), again representing typical vocabulary levels for native English speakers in their age range (M = 36.0/60, SD = 12.7/60; Ardila, 2007). A self-report questionnaire was used to ensure that all participants had good health and no history of speech, language, or hearing problems. All participants had pure-tone air-conduction thresholds within clinically normal limits for their age group, from 0.25 to 3 kHz for both ears (≤20 dB HL).

Stimuli and Apparatus

The semantic (lexical) stimuli consisted of 50 sentences, with 10 corresponding to each of the following semantic categories: anger, fear, happy, sad, and neutral. The four emotional categories were chosen because they are commonly used in the literature (for a review, see Zupan, Neumann, Babbage, & Willer, 2009) and have been found to be easily recognized and distinguished in both prosody and semantics (Breitenstein, Daum, & Ackermann, 1998; Laukka, 2003; Scherer, 2003; Scherer, Banse, & Wallbott, 2001). In a previous study conducted in our lab (Ben-David, van Lieshout, & Leszcz, 2011), these sentences were found to be reliably associated with a discrete semantic category (see Table A1 in Appendix A). These sentences were also matched on the linguistic characteristics of their content words across the five semantic categories (frequency of usage in English, phonologic neighborhood density, and number of syllables; see Table A2 in Appendix A).

In a follow-up study (Ben-David, Thayapararajah, & van Lieshout, 2013), these semantic sentences were recorded by a trained professional female actor, a native (Canadian) English speaker, using a sampling rate of 24414 Hz. Each sentence was recorded three times in each of the five different prosodies (anger, fear, happy, sad, and neutral) to generate a set of 750 recorded sentences. Digital audio files were equated with respect to their root-mean-square amplitude.

From these spoken sentences, a subset of 50 spoken sentences was selected on the basis of the perceived high quality of prosodic information. Next, after the recorded sentences were delexicalized by digital acoustic filters, they were further validated as good representations of their intended emotional category as presented by the prosody without the impact of their semantic content (for details, see Ben-David, Thayapararajah, & van Lieshout, 2013). A complete description of the stimuli, their linguistic characteristics (equated across emotional categories), and their associated testing and validation (ensuring they present a discrete emotion in both semantics and prosody) is provided in our prior work (semantics: Ben-David, van Lieshout, & Leszcz, 2011; prosody: Ben-David, Thayapararajah, & van Lieshout, 2013).

Two subsets of 25 spoken sentences were taken from this resource. Each set consisted of five sentences in each of the five discrete categories—anger, fear, happiness, sadness, and neutral—such that no semantic content was repeated. In a single set, each emotional semantic category is represented once in each of the tested emotional prosodies, generating a 5 × 5 (semantics × prosody) matrix, as presented in Table 1. However, we deemed spoken sentences that carry neutral information in both semantics and prosody to be uninformative, especially because the experimental task involves rating of emotion. To save time and to reduce confusion, we removed these sentences, leaving a final set of 48 spoken sentences (24 in each subset) that all present emotional information in at least one channel.

Procedure

Participants were tested individually, in a sound-attenuating booth and seated in front of a 17-in. flat color monitor, wearing a headset. An experimental session consisted of three rating tasks—General-rating, Semantics-rating, and Prosody-rating—separated by short breaks. Each rating task was made up of four emotion-rating blocks—anger-rating, fear-rating, sadness-rating, and happiness-rating—comprising 12 experimental blocks per session. In each block, participants were asked to listen to a spoken sentence and rate how much they agreed that the speaker conveyed a predefined emotion, using a 6-point Likert scale ranging from *strongly disagree* (1) to *strongly agree* (6)—for example, "How much do you agree that the speaker is _____?" (angry, fearful, sad, or happy). The participants responded by pressing the respective number key (1–6) on a standard PC keyboard. Each block consisted of 48 trials, making for 576 trials per session. An emotion-rating block began with the presentation of an instruction slide, followed by two practice trials. Practice trials were followed by a reminder of the instructions, and experimental trials were initiated by the participant. No feedback was provided throughout the practice and experimental trials, because the T-RES gauges the listener's subjective perception of emotion (i.e., there are no "right" or "wrong" answers).

In the General-rating task (Set 1 in Table 1), listeners were asked to imagine that they were listening to a person over the phone (thus eliminating visual cues) and rate

Table 1. The 24 sentences for the General-rating task (Set 1) and 24 sentences for the Semantics- and Prosody-rating tasks (Set 2).

Lexical content	Prosody				
	Anger	**Fear**	**Sad**	**Happy**	**Neutral**
Anger	1. You over-charged me for that. 2. This is infuriating.	1. Stop what you're doing and listen to me. 2. Go to hell.	1. Quiet, this is a library. 2. Get out of my room.	1. Don't waste my time. 2. I am very angry.	1. Do not push your luck. 2. I'm sick of you being late.
Fear	1. I can hear footsteps in the night. 2. I can hear a sharp scream from behind.	1. Watch out for that tiger. 2. Run for your life.	1. I can't see the bear but I can hear it. 2. I smell the gas leaking from the stove.	1. The fire is spreading to the gas pipe. 2. Look out there's a car coming.	1. It's about to explode. 2. He has a knife.
Sad	1. I'm going to a funeral. 2. This song makes me cry.	1. Gray clouds make me feel gloomy. 2. This scene makes him feel blue.	1. This is a sad moment. 2. I am so lonely.	1. I've been crying all day. 2. My best friend is moving away.	1. The weather is depressing. 2. My pet died today.
Happy	1. Congratulations, you're hired. 2. This is my favorite part.	1. Good job, the crowd really loves you. 2. I got promoted in my job.	1. Great, you got first place. 2. This is the happiest day of my life.	1. I won an award. 2. This food tastes very good.	1. It's a beautiful day outside. 2. I feel wonderful today.
Neutral	1. He stands on the deck. 2. Digital clocks are common.	1. Lots of bins are in the room. 2. A bag is in the room.	1. Four drawers are in the cabinet. 2. This is a garbage can.	1. Red pipes are metallic. 2. Our body is made of water.	

the emotion expressed by the speaker, on the basis of the spoken sentence as a whole. In the Semantics-rating task (Set 2 in Table 1), listeners were asked to rate exclusively the emotion conveyed by the semantic content, ignoring the prosody of the spoken sentence. In the Prosody-rating task, the same set of sentences as in the Semantics-rating task was presented (Set 2), but listeners were asked to attend exclusively to the prosody, ignoring the semantic content of the utterance. For example, consider the semantically happy sentence "I feel wonderful today" spoken with an angry prosody. In the General-rating task, participants were asked to rate the spoken sentence without specific instructions to attend to one speech channel or the other. In the Semantics-rating task, they were asked to specifically attend to the (happy) content of the utterance; in the Prosody-rating task, the (angry) prosodic content.

To control for a possible effect of presentation order, three methods were utilized. First, the order of the trials in each block was pseudorandomized, ensuring that no emotion was presented in the same channel (prosody or semantics) more than twice in a row. Second, in each task the order of the four emotion-rating blocks was counterbalanced across participants using a Latin-square procedure (Grant, 1948) but maintained across all three rating tasks for an individual participant. As a final point, each experimental session started with the General-rating task, to prevent biasing the listeners to pay attention to a specific channel. For a randomly selected half of the participants, this was followed by the Semantics-rating task and then the Prosody-rating task. For the other half, this order was reversed. The latter two counterbalancing procedures required the use of eight experimental groups (10 participants per group), as illustrated in Table 2.

Tested Measures

We define two variables for the analysis: Rating of Attended Channel (RAC) and Rating of Ignored Channel (RIC). RAC measures the influence of the attended channel on rating. In particular, it indicates whether listeners accurately identify the emotion presented in the attended channel. Channel dominance, prosodic or semantic, will be manifested in larger RAC scores for the dominant channel. RIC

Table 2. The order of emotion-rating blocks.

Number of participants	Order of emotion-rating blocks			
20	Anger	Sad	Fear	Happy
20	Fear	Anger	Happy	Sad
20	Happy	Fear	Sad	Anger
20	Sad	Happy	Anger	Fear

Note. This order was preserved in all the three tasks (General-, Prosody-, and Semantics-rating tasks). For all participants, the General-rating task was presented first. For half of the participants in each order, this was followed by the Semantics-rating task and then the Prosody-rating task; for the other half, the Prosody-rating task was presented second and the Semantics-rating task last.

measures the extent to which listeners can selectively ignore the emotional information presented in the to-be-ignored channel while rating the target channel. To be specific, if listeners can process speech channels separately and independently, RIC should be zero. If listeners are processing spoken emotions in an integrative manner, RIC should be larger than zero.

As a formal matter, RAC is calculated as the difference between the average ratings of sentences that present the rated emotion in the attended channel versus sentences that do not present the rated emotion in the attended channel. The variable $RAC_{prosody}$ measures the extent to which listeners identify the emotion conveyed in the prosody, whereas $RAC_{semantics}$ measures the extent to which listeners can identify semantics. In general,

$$RAC_{prosody_x} = prosody_x rating \left| prosody_x - \frac{1}{3} \sum_y prosody_x rating \right| prosody_y$$

(1)

$$RAC_{semantics_x} = semantics_x rating \left| semantics_x - \frac{1}{3} \sum_y semantics_x rating \right| semantics_y,$$

(2)

where x represents one of the four rated emotions (anger, fear, sad, or happy) and y represents the other three emotions that are not x. For example, in the Semantics-rating task, if x represents anger, y represents the sum of the scores for fear, sad, and happy semantics.

In contrast, RIC measures the difference between average ratings of sentences that present the rated emotion in the to-be-ignored channel and sentences that do not present the rated emotion in either channel:

$$RIC_{prosody_x} = semantics_x rating \left| prosody_x - \frac{1}{3} \sum_y semantics_x rating \right| prosody_y$$

(3)

$$RIC_{semantics_x} = prosody_x rating \left| semantics_x - \frac{1}{3} \sum_y prosody_x rating \right| semantics_y.$$

(4)

If listeners are able to selectively ignore prosody, $RIC_{prosody} = 0$; and if they are able to ignore semantics, $RIC_{semantics} = 0$. On the other hand, when RIC > 0, $RIC_{prosody}$ measures the extent to which listeners fail to selectively ignore the prosody, whereas $RIC_{semantics}$ measures the extent to which listeners fail to selectively ignore the semantics.

Statistical Analysis

All of the following analyses were made of omnibus repeated measures analyses of variance (ANOVAs), with average rating as the dependent variable and rated emotion (anger, fear, sad, or happy) as a within-participant variable, as well as one or two other tested within-participant variables as specified in each test. ANOVAs also included two between-participant variables of order: order of rated emotion and order of target channel. Because none of the order variables was found to yield any significant effect ($F < 1$), they will not be further discussed. These tests were conducted separately for baseline sentences, conveying an emotion in only one channel (and held neutral on the other), and for emotional sentences, conveying emotions in both channels (not necessarily the same one). This procedure was implemented to avoid a scaling bias in statistical analyses, because baseline sentences received substantially lower ratings than emotional ones.

Prosody- and Semantics-Rating Tasks

In the Prosody-rating and Semantics-rating tasks, we conducted the ANOVA with the following tested variables: target channel (attending to the semantics or the prosody) and RAC. This was paralleled with an ANOVA using target channel and RIC as the tested variables.

General-Rating Task

Baseline sentences. For baseline sentences, we compared ratings of the rated emotion when it was presented in the prosody or the semantics. To that end, the tested variables were channel (whether the rated emotion was presented in the semantics or the prosody) and RAC (comparing the first and second components of Equations 1 or 2).

Emotional sentences. For emotional sentences, we tested semantic and prosodic dominance directly, as well as the influence of congruency. As a first step, we used channel combination as the tested variable, comparing four combinations: (a) rated emotion appears in both channels, (b) rated emotion appears in the prosody channel, (c) rated emotion appears in the semantics channel, and (d) rated emotion is absent from either channel. The second step included six planned post hoc (Bonferroni-corrected) ANOVA tests of each possible pairing of the four channel combinations.

Emotional versus baseline sentences: Congruency supremacy. As discussed in the introduction, the supremacy of congruency can be attributed to the advantage of presenting the same information in both channels, as reflected in a redundancy source. To test this source separately for prosody and semantics, separate ANOVAs were used, examining whether congruent emotional sentences were rated higher than all other sentences.

Results

The 12 panels of Figure 1 present the average ratings of each combination of channels in all four emotion-rating blocks, separately for the three rating tasks (Semantics-, Prosody-, and General-rating tasks).

Prosody- and Semantics-Rating Tasks

Baseline Sentences

RAC. Baseline sentences were examined to verify that listeners could correctly identify the rated emotion in the target channel when no other emotion was expressed in the to-be-ignored channel. For example, for baseline $RAC_{prosody}$, Figure 1B (anger-prosody rating) shows that a semantically neutral sentence with angry prosody was rated substantially higher on expressing a prosody of anger (5.8) than the average ratings obtained for semantically neutral sentences spoken with fear, sad, and happy prosodies (1.5, 1.8, and 1.4; $M = 1.6$; $RAC_{prosody} = 4.2$). Indeed, the omnibus ANOVA of baseline sentences' ratings shows a main effect of RAC ($M = 4.01$), $F(1, 72) = 2,961$, $p < .001$. The effect size of RAC is strong, $\eta_p^2 = .976$. There was no main effect for target channel ($F < 1$) and no significant interaction of the two.

In sum, it is clear that listeners were able to identify the emotions presented in baseline spoken sentences, both in the prosody and in the semantics. The lack of an interaction indicates that listeners were equally good at identifying emotions in both channels.

RIC. Here we measure the extent of failures in selective attention as reflected in ratings of baseline sentences. Figure 1J (happy-semantics rating) shows an example for baseline $RIC_{prosody}$ measuring the ability to ignore the (happy) prosody. In particular, it compares a neutral-semantics, happy-prosody sentence (3.4) with the average ratings of neutral-semantics sentences carrying anger, fear, and sad prosody (2.4, 1.8, 2.2; $M = 2.1$). Because none of these sentences presents happy semantics, their happy-semantics ratings should not differ. In other words, if listeners can selectively ignore the prosody, then $RIC_{prosody} = 0$. In this example (happy-semantics rating), $RIC_{prosody} = 1.3$, indicating that listeners were not able to ignore the happy prosody when asked to focus on the semantics. However, in equivalent baseline happy-prosody ratings (Figure 1K) this was not the case, and $RIC_{semantics} = 0$, suggesting that listeners were able to selectively ignore the happy semantics when rating the prosody.

The analysis of RIC in baseline sentences shows that listeners were generally unable to inhibit the information presented in the to-be-ignored channel, $F(1, 72) = 86.7$, $p < .001$, $\eta_p^2 = .55$ ($RIC = 0.41$). As seen in the foregoing examples, these failures of selective attention were more extensive when listeners were asked to ignore the prosody ($RIC_{prosody}$) than when they were asked to ignore the semantics ($RIC_{semantics}$), showing a significant RIC × Target Channel interaction, $F(1, 72) = 52.95$, $p < .001$, $\eta_p^2 = .42$ ($RIC_{prosody} = 0.75$ vs. $RIC_{semantics} = 0.08$). These results indicate that it was more difficult to inhibit the prosodic information than the semantics, providing support for the prosodic-dominance hypothesis.

Emotional Sentences

Figure 2 provides a graphic representation of the averages for ratings of emotional sentences in the three tasks,

Figure 1. Average ratings on a 6-point Likert scale for all participants ($N = 80$) of each combination of channels in all four emotion-rating blocks, separately for the three rating tasks (Semantics-, Prosody-, and General-rating tasks). Columns present the different emotional prosodies, and rows present the different emotional semantics. Cells are shaded to facilitate interpretation, with different shades of gray varying from a filled (black) cell, indicating the maximum score of 6 (highest agreement that the spoken sentence conveyed the rated emotion), to a blank (white) cell, indicating the minimum score of 1 (lowest agreement). Cells representing sentences that carry the rated emotion in either the prosody (column) or the semantics (row) are indicated in italics. Finally, baseline spoken sentences (sentences that carry a neutral content in one channel) are separated from emotional spoken sentences (sentences that carry emotional content in both channels).

LEXICAL CONTENT

A. Semantics-rating: Anger (PROSODY)

	A	F	S	H	N
A	5.9	5.7	4.8	5.7	5.6
F	2.5	2.0	2.2	2.0	2.3
S	2.1	1.5	2.0	2.8	2.5
H	1.5	1.1	1.2	1.1	1.1
N	1.9	1.7	1.8	1.2	

B. Prosody-rating: Anger (PROSODY)

	A	F	S	H	N
A	5.8	2.6	1.7	1.1	1.9
F	4.9	1.9	1.6	1.3	1.7
S	5.8	1.6	1.5	1.2	1.7
H	5.6	1.7	1.9	1.1	1.9
N	5.8	1.5	1.8	1.4	

C. General-rating: Anger (PROSODY)

	A	F	S	H	N
A	5.4	3.3	2.1	1.9	2.0
F	4.6	1.7	1.7	1.8	1.7
S	5.3	2.0	1.8	2.0	1.8
H	4.9	1.8	2.7	1.1	1.9
N	5.8	1.9	1.9	1.7	1.6

D. Semantics-rating: Fear

	A	F	S	H	N
A	1.6	2.4	2.5	1.8	1.7
F	5.1	5.8	5.4	5.4	5.7
S	1.7	2.3	2.6	2.3	2.0
H	1.2	1.3	1.2	1.1	1.1
N	1.2	2.5	1.5	1.4	

E. Prosody-rating: Fear

	A	F	S	H	N
A	1.5	4.1	2.6	1.2	1.8
F	3.0	5.4	3.8	1.2	1.6
S	1.5	3.8	2.2	1.3	1.4
H	1.7	4.0	1.7	1.2	1.5
N	1.6	5.2	2.3	1.2	

F. General-rating: Fear

	A	F	S	H	N
A	1.5	4.4	2.2	1.3	1.8
F	3.2	5.3	5.2	3.6	1.8
S	2.1	3.7	2.1	1.3	1.5
H	1.7	2.9	1.8	1.1	1.3
N	2.0	4.4	2.9	1.2	1.5

G. Semantics-rating: Sad

	A	F	S	H	N
A	1.9	2.7	3.5	1.8	2.4
F	1.9	2.2	2.3	1.9	1.9
S	5.3	4.6	5.8	5.4	5.8
H	1.3	1.2	1.3	1.2	1.2
N	1.4	1.7	1.6	1.2	

H. Prosody-rating: Sad

	A	F	S	H	N
A	1.5	3.2	5.8	1.2	2.6
F	1.5	2.0	5.2	1.2	2.6
S	1.8	4.4	5.8	1.3	2.8
H	1.5	4.3	5.7	1.2	3.0
N	1.6	3.0	5.8	1.3	

I. General-rating: Sad

	A	F	S	H	N
A	1.7	4.5	5.5	1.2	2.1
F	1.6	2.0	4.9	1.8	2.4
S	2.6	2.8	5.6	2.4	3.5
H	2.2	3.4	5.3	1.1	2.9
N	1.7	3.9	5.4	1.2	2.0

J. Semantics-rating: Happy

	A	F	S	H	N
A	1.1	1.2	1.3	1.2	1.3
F	1.4	1.3	1.3	1.3	1.2
S	1.3	1.5	1.1	1.4	1.1
H	5.4	5.7	5.7	5.7	5.7
N	2.4	1.8	2.2	3.4	

K. Prosody-rating: Happy

	A	F	S	H	N
A	1.2	2.0	1.1	5.5	1.6
F	1.5	1.8	1.5	5.7	1.4
S	1.1	1.8	1.1	5.6	1.3
H	1.5	2.1	1.2	5.7	1.4
N	1.2	1.7	1.2	5.7	

L. General-rating: Happy

|	A	F	S	H	N
A	1.4	1.4	1.2	4.0	1.8
F	1.7	1.9	1.3	3.7	1.5
S	1.3	2.8	1.2	4.1	1.4
H	1.9	2.6	1.5	5.7	1.8
N	1.2	1.6	1.2	5.0	1.7

General-rating (Figure 2A), Prosody-rating (Figure 2B), and Semantics-rating (Figure 2C).

RIC. Here we examine the extent of failures in selective attention when both the target channel and the to-be-ignored channel carry emotional content. As a first step, we confirmed that listeners rated congruent emotional sentences (where the rated emotion is presented in both channels) differently than sentences that conveyed the rated emotion in only the target channel. For example, in Figure 1A we compared anger-semantics ratings of a sentence that presents anger in both channels (anger semantics spoken with anger prosody: 5.9) with the average anger-semantics ratings of sentences that present anger in the semantics with a different prosodic emotion (anger semantics with fear, sad, and happy prosodies: 5.7, 4.8, 5.7; $M = 5.4$). For anger, $RIC_{prosody} = 0.5$. Again, if listeners were able to completely ignore the prosody, $RIC_{prosody}$ should be equal to 0. In a similar vein, one can observe anger $RIC_{semantics} = 0.4$ for anger-prosody rating (Figure 1B). For a graphic representation of anger $RIC_{prosody}$ (indicating failure to selectively ignore the prosody), turn to the leftmost bars in Figure 2C (the difference between the black and the light-gray bars), and for anger

Figure 2. Graphic description of ratings of emotional sentences in the General-rating (Panel A), Prosody-rating (Panel B), and Semantics-rating (Panel C) tasks, for the four emotion-rating tasks. Black bars represent average ratings of congruent trials (the rated emotion appears in both channels); dark-gray bars, prosody trials (the rated emotion appears only in the prosody); light-gray bars, semantics trials (the rated emotion appears only in the semantics); and white bars, rated-emotion-absent trials (the rated emotion does not appear in either the semantics or the prosody).
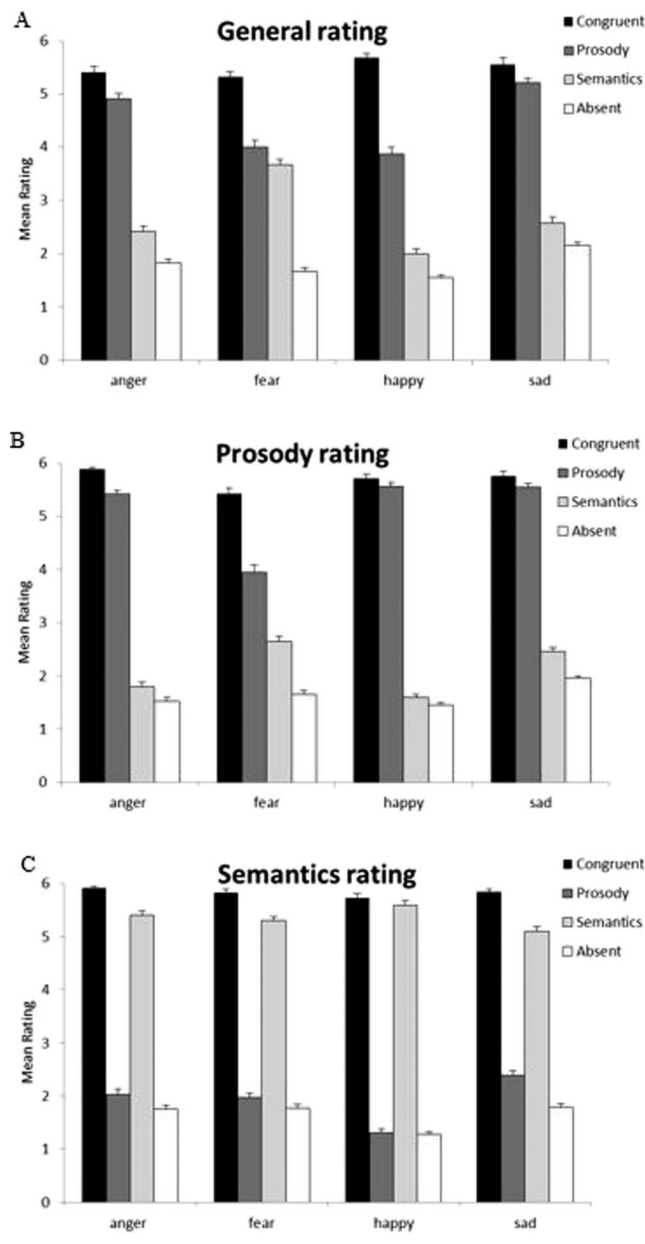
There was no significant difference between $RIC_{semantic}$ and $RIC_{prosody}$ ratings, $F(1, 72) = 1.6$, $p = .2$. The main effect of RIC can thus be interpreted as evidence supporting the supremacy of congruency. That is, congruent sentences that present the rated emotion in both speech channels were rated higher than incongruent sentences that carried the rated emotion in only the target channel and another emotion in the to-be-ignored channel.

As the second step, we tested the impact of the to-be-ignored channel on sentences that did not present the rated emotion in the target channel. For example, in Figure 1A (anger-semantics rating), anger $RIC_{prosody}$ is the difference between the average ratings of sentences that present anger in only the prosody (with fear, sad, and happy semantics: 2.5, 2.1, 1.5; $M = 2.0$) and sentences that do not present anger in either dimension (fear, sad, and happy semantics with a fear, sad, or happy prosody: 2.0, 2.2, 2.0, 1.5, 2.0, 2.8, 1.1, 1.2, 1.1; $M = 1.8$; $RIC_{prosody} = 0.2$). Because none of these sentences present anger semantics, their anger-semantics ratings should not differ ($RIC_{prosody} = 0$). For a graphic representation, examine the leftmost bars of Figure 2: For $RIC_{prosody}$, see Figure 2C (the difference between the dark-gray and the white bars); for $RIC_{semantics}$, Figure 2B (the difference between the light-gray and the white bars).

The ANOVA shows significant failures of selective attention, across target channels, $F(1, 72) = 109.7$, $p < .001$, $\eta_p^2 = .60$ (RIC = 0.38). Notably, prosodic dominance is indicated, with larger failures of selective attention when the listeners were asked to ignore the prosody ($RIC_{prosody} = 0.48$) than the semantics ($RIC_{semantics} = 0.28$), $F(1, 72) = 12.6$, $p < .001$, $\eta_p^2 = .15$.

General-Rating Task

This section compares the relative impact of the information conveyed in the prosody and in the semantics, when listeners are asked to attend to both channels simultaneously and rate each sentence as a whole. The data are presented in the leftmost column of Figure 1 and graphically in Figure 2A.

Baseline Sentences

RAC. We compared the role of prosody and semantics in ratings of baseline sentences. In particular, we compared $RAC_{prosody}$ with $RAC_{semantics}$ when the other channel conveyed neutral content. For example, in Figure 1C (general anger ratings), baseline anger $RAC_{prosody}$ is the difference between the ratings of a semantically neutral sentence with angry prosody (5.8) and the average ratings of semantically neutral sentences with fear, sad, and happy prosodies (1.8, 2.7, 1.1; $M = 1.8$; $RAC_{prosody} = 4.0$). Baseline anger $RAC_{semantics}$ is the difference between a semantically angry sentence with neutral prosody (2.0) and the average of semantically fear, sad, and happy sentences with neutral prosody (1.7, 1.8, 1.9; $M = 1.8$; $RAC_{semantics} = 0.2$). In this example with the anger emotion, there is clear evidence supporting prosodic dominance ($RAC_{prosody} > RAC_{semantics}$).

$RIC_{semantics}$ (indicating failure to selectively ignore the semantics), the leftmost bars in Figure 2B (the difference between the black and the dark-gray bars).

In the ANOVA, a significant failure of selective attention is evident, $F(1, 72) = 148.3$, $p < .001$, $\eta_p^2 = .67$ (RIC = 0.53).

The overall ANOVA echoes this example. A main effect of RAC was observed, $F(1, 72) = 1,247.4$, $p < .001$, $\eta_p^2 = .95$ (RAC = 1.9), alongside a main effect of channel, $F(1,72) = 400.4$, $p < .001$, $\eta_p^2 = .85$. The two effects interacted significantly, $F(1, 72) = 446.2$, $p < .001$, $\eta_p^2 = .86$, with $\text{RAC}_{\text{prosody}} = 3.3$ almost 7 times as large as $\text{RAC}_{\text{semantics}} = 0.5$. In other words, prosodic emotional information had a substantially larger impact than semantics on ratings.

Emotional Sentences
Prosodic advantage and channel integration. In this section, we directly compare prosodic and semantic dominance and confirm congruency supremacy when listeners are asked to rate emotional sentences as a whole. For example, for general anger rating in Figure 1C, we found that congruent anger ratings (5.4) were greater than the average ratings of anger-prosody sentences ($M = 4.9$), which were in turn greater than the average ratings of anger-semantics sentences ($M = 2.4$), followed up by the average ratings of sentences that did not present anger in either channel ($M = 1.8$). For a graphical representation, this is the gradient observed in the leftmost bars of Figure 2A, with the black, dark-gray, light-gray, and white bars.

Thus it appears that for general anger ratings, we see both the supremacy of congruency and prosodic dominance. In the overall ANOVA, we compared the four channel combinations: congruent trials (the rated emotion appears in both channels), prosody trials (the rated emotion appears in only the prosody), semantics trials (the rated emotion appears in only the semantics), and the rated-emotion-absent trials (the rated emotion does not appear in either the semantics or the prosody). We found these four combinations to differ significantly, $F(3, 216) = 724.5$, $p < .001$, $\eta_p^2 = .91$.

In a series of post hoc analyses, the following results were obtained: (a) supremacy of congruency: Sentences that present the rated emotion in both channels were rated higher than sentences that present that emotion in only the prosody (5.5 vs. 4.4), $F(1, 72) = 172.6$, $p < .001$, $\eta_p^2 = .71$; (b) prosodic dominance: Listeners were influenced by the prosodic information to a larger extent than by the semantic information (4.4 vs. 2.7), $F(1, 72) = 196.9$, $p < .001$, $\eta_p^2 = .73$; and (c) semantic information has a significant impact on rating (2.7 vs. 1.8 for semantics and rated-emotion-absent trials), $F(1, 72) = 244.8$, $p < .001$, $\eta_p^2 = .77$. In other words, in the General-rating task, listeners derived emotional information from the prosody and the semantics, but prosodic emotional information had a substantially larger impact on their emotional ratings.

Congruency supremacy: Redundancy of information. Here we examine directly the role of redundancy of information (two compatible sources rather than one source of information) in the supremacy of congruency when listeners are asked to rate the sentence as a whole. For example, turn to the general happiness rating (Figure 1L). For prosody, we compare the congruent happy sentence (5.7) with the baseline happy-prosody sentence (5.5). For semantics, we compare the same congruent happy sentence (5.7) with the baseline happy-semantics sentence (1.8). In other words, when the rated emotion appears in one channel (and the other channel is neutral), to what extent does adding the same emotion in the other channel increase ratings?

In the ANOVA for prosody, there was a small but significant effect of Redundancy (5.5 vs. 5.2 for congruent and prosody baseline trials), $F(1, 72) = 17.5$, $p < .001$, $\eta_p^2 = .18$. For semantics, there was a markedly large effect of Redundancy (5.5 vs. 2.3), $F(1, 72) = 756.6$, $p < .001$, $\eta_p^2 = .91$. In sum, this analysis further supports the theory of prosodic dominance. When the rated emotion is presented in the prosody, adding the same emotion in the semantics has a mild redundancy effect. However, when the rated emotion is presented only in the semantics, adding the same emotion in the prosody introduces a large redundancy effect ($\eta_p^2 = .18$ vs. .91).

Discrete-Emotion Analysis

In all of the ANOVAs so far described, the rated emotion (anger, fear, sad, or happy) was a within-participant variable (as indicated in the Statistical Analysis subsection earlier). In the majority of these analyses, we found the rated emotion to interact with the main tested effect (for each test). In a series of post hoc ANOVAs, one for each of the rated emotions, we found that the rated emotion had an impact on only the extent of the tested effects, but it did not change the general trend.

For example, examine Figure 2A, graphically describing general ratings for emotional sentences on the four rated emotions. For all rated emotions, the same gradient is demonstrated, from congruent trials (black bars) to prosody trials (dark gray) to semantics trials (light gray) to rated-emotion-absent trials (white). Figure 2A shows that rated emotions have an impact merely on the extent of this trend, with sad ratings presenting the largest effect for prosody and fear ratings presenting the largest effect for semantics. This is supported by the post hoc tests: (a) Supremacy of congruency (congruent trials being rated higher than prosody trials) was significant for all rated emotions, $F(1, 72) > 13$, $p < .001$, except sad (not significant after Bonferroni correction), with the largest effect for happy ratings (5.3 vs 3.7); (b) prosodic dominance (prosody trials being rated higher than semantics trials) was significant for all rated emotions, $F(1, 72) > 106$, $p < .001$, except fear, with the largest effect for sad ratings (5.2 vs. 2.6); and (c) semantic impact (semantics trials being rated higher than rated-emotion-absent trials) was significant for all four rated emotions, $F(1, 72) > 17$, $p < .001$, with the highest effect for fear ratings (4.0 vs. 1.7).

In Appendix B we present the full discrete-emotion analysis for all of the tests conducted in this study. The results across all analyses are very clear—in each test, the general effect is observed in almost all rated emotions, but several changes between emotions can be found. The most notable changes are that the impact of prosody is the highest for sad ratings, and the impact of semantics is the highest for fear ratings.

Correlation

Table 3 shows that scores on the General-rating task are highly similar to scores on the Prosody-rating task. That is, when listeners are asked to focus on the prosody, their ratings are not dissimilar from when they are asked to rate the spoken sentence as a whole. To measure this directly, we compared the correlations of ratings between the General-rating task and the other two tasks.

First, we calculated Pearson correlation coefficients for each of the combinations of semantics and prosody (16 excluding baseline sentences) and for the four emotional rating blocks separately (64 correlations for Prosody- vs. General-rating and 64 for Semantics- vs. General-rating tasks). In other words, across each row in Figure 1, we calculated the correlation coefficients for their respective cells. Next, we averaged these correlation coefficients across the 16 combinations, for each emotional rating block. These averaged coefficients are presented in Table 3 for each of the emotional rating scales. All Pearson correlation coefficients are significantly different from chance at $p < .01$. Across rated emotions, scores on the General-rating task correlated more strongly with prosodic ratings than with semantic ratings, especially when the rated emotions were anger and happy. These correlations support a prosodic dominance in rating of spoken emotions, as found in the previous analyses.

General Discussion

The aim of this study was to explore the complex interplay of prosody and semantics in the perception of discrete emotions in spoken sentences using a novel tool, the T-RES. Eighty native English-speaking listeners were asked to rate well-controlled spoken sentences on emotional scales, relating to the prosody, the semantics, or the combination of the two. Our findings highlight the following main trends: (a) supremacy of congruency—a sentence that presents an emotion in both channels was rated highest on its emotional scale; (b) failure of selective attention—when instructed, listeners were not able to selectively attend to one channel while completely ignoring the other; and (c) prosodic dominance—in rating an emotion in speech, prosodic information plays a larger role than semantics.

Efficacy of the T-RES

This study presents the T-RES, a novel test designed to expose the full complexity of the interaction of the two speech channels, semantics and prosody. In the T-RES, listeners are presented with spoken sentences that carry emotions in the semantics, prosody, or both, and asked to rate them on three tasks. In the Prosody- and Semantics-rating tasks, listeners are asked to selectively attend to only one channel (prosody or semantics), ignoring the other. In the General-rating task, listeners rate the sentence as a whole, using the information in both channels. The T-RES complements the arsenal of tools currently available (FAB, DANVA2) and other more recent tools (Pell, Monetta, et al., 2009) that are very useful for measuring individual variance but less effective at revealing the integration between semantic and prosodic channels. The stimuli used were found to be distinctive in conveying their corresponding emotions, equated on main linguistic characteristics (Ben-David, Thayapararajah, & van Lieshout, 2013; Ben-David, van Lieshout, & Leszcz, 2011), and presented in a design that avoids a set-size and covariate bias (Algom, Chajut, & Lev, 2004; Melara & Algom, 2003). Moreover, the T-RES is unique in using a rating task rather than forced choice. These methodological features were found to be fruitful in providing an in-depth understanding of the processes underlying the perception of emotions in speech, as discussed in the next sections.

Some validation of the tool comes from the analysis of baseline sentences, presenting an emotional content in one channel while the other is held neutral. The ability of listeners to successfully identify the emotion conveyed by the semantics or by the prosody was gauged by the RAC measure (rating of attended channel). Baseline RAC was significant and high for both prosody and semantics. These results show that it was very easy for listeners to identify the emotion presented in one channel when no possible interference was presented in the other. A closer inspection of the results of the General-rating task shows similar results. Sentences that carry the rated emotion in at least one channel were rated higher than sentences that do not carry that emotion in either channel. This was true even in examining incongruent emotional trials that present the rated emotion in one channel but a different emotion in the other channel. Taken together, the results indicate that the emotions as presented in the T-RES are easily recognized by participants. We have tested this paradigm with other groups of participants that vary in age and language background (see, e.g., Ben-David, Multani, Durham, Green, & van Lieshout, 2014; Ben-David, Multani, Shakuf, & van Lieshout, 2013; Shakuf & Ben-David, 2014). Its administration is simple, and participant debriefing shows that the paradigm was clear and easy to follow.

Table 3. Average correlation coefficients between ratings on the General-rating task and the Semantics- and Prosody-rating tasks across each of the four nonneutral emotions.

Emotion	General- versus Semantics-rating	General- versus Prosody-rating
Anger	.16	.30
Fear	.25	.33
Happy	.13	.22
Sad	.18	.20

Failure of Selective Attention

Our results show that listeners were not able to selectively attend to one channel and ignore the other. When they were specifically instructed to focus on and rate only one channel, the information in the to-be-ignored channel still had a significant effect on their ratings. This was indicated in highly significant RIC scores in the Prosody- and Semantics-rating tasks with emotional sentences, and even

when the target channel was held neutral in baseline sentences. It appears that listeners regarded neutral prosody and neutral semantics as carrying a modicum of the emotion conveyed in the to-be-ignored channel. For example, the semantically neutral sentence "Red pipes are metallic" was rated as moderately semantically happy when it was spoken with a happy prosody. This general failure of selective attention is in line with the evidence found in the literature. In the introduction, we discussed studies that presented single spoken words and asked listeners to focus on only one speech channel, ignoring the other (Ishii et al., 2003; Kitayama & Ishii, 2002; Mehrabian & Wiener, 1967). Taking these studies together, they show significant failures of selective attention to semantics or prosody. Our study adds a significant extension to these findings, following current trends in the literature and also showing evidence for failures in selective attention in spoken sentences.

Further support for our findings comes from studies indicating that the meaning of emotional prosody and semantics is processed even when it is task irrelevant. For instance, evidence from auditory versions of the emotional Stroop task demonstrate the effect of negative prosody and semantics (Bertels, Kolinsky, Pietrons, & Morais, 2011; Egloff & Schmukle, 2004) on performance when it is irrelevant to the task. In a similar vein, emotional prosody (regardless of its specific emotional meaning) has been found to speed up the identification of semantically congruent emotional words (Wurm et al., 2001) even when listeners were asked to focus solely on the semantics. It is notable that these studies found failures of selective attention only when one channel (semantics or prosody) was held constant throughout the experimental block (e.g., all sentences carried anger semantics), which may generate a context effect or expectations (Kitayama, 1996). Our data add to the literature by demonstrating the robust impact of spoken emotions, with failures of selective attention recorded even when the emotional meaning of the to-be-ignored channels and attended varied from trial to trial.

A possible theoretical framework for these results is provided by Craik and Lockhart (1972). Their classic *levels of processing* theory describes a gradient of processing, starting from the basic physical attributes, followed by the nominal level (where the name is processed), to the level of meaning as the highest level of processing (for a discussion on a "flexible" hierarchy of processing, see Craik, 2002). Adopting this model, the failures of selective attention in our data indicate processing of the to-be-ignored dimension—meaning. As a final point, failures of selective attention were more pronounced when listeners were asked to ignore the information conveyed in the prosodic domain than in the semantics one. This hints toward prosodic dominance, as discussed next.

Prosodic Dominance

Across all analyses, there is overwhelming evidence to show that prosodic information plays a much larger role than semantics in processing emotion in speech. This can be demonstrated by ratings of baseline sentences in the General-rating task. Neutral-semantics sentences spoken with emotional prosody were rated much higher than emotional-semantics sentences spoken with a neutral prosody (5.3 vs. 2.3). It is notable that the average of 5.3 indicates very high agreement that baseline sentences convey the prosodic emotion (on a 1–6 Likert scale), whereas the 2.3 average score does not convey the same agreement with semantics. In other words, to express a specific emotion in speech, it is sufficient to add an emotional prosody to neutral semantics. To a large extent, this is confirmed in Table 3, which demonstrates the high similarity of General-rating scores to Prosody-rating scores. However, adding emotional semantics to neutral prosody does only a little to express a specific emotion. Examining general ratings of emotionally incongruent sentences replicates these results—prosody was rated much higher than semantics (4.4 vs. 2.7). The same conclusion can be drawn from the analysis of the Prosody- and Semantics-rating tasks. Failures of selective attention were significant for both channels, but they were much more substantial when listeners had to ignore the prosodic information.

This asymmetry of emotional speech channels may be related to the arbitrariness of sound-to-meaning allocation in semantics (Brown, 1973), where any pattern of sound can refer to any kind of discrete emotional meaning. However, prosodic physical attributes are more strongly associated with emotions. For example, listeners speaking different languages coming from nine countries (in Europe, Asia, and North America) were able to infer emotion from prosodic emotional speech segments spoken by German actors (Scherer et al., 2001), even though they did not speak the language. These and other studies support universal inference rules for emotional prosodic information. Indeed, in a review of emotional speech synthesis, Schröder (2001) concluded that "prosody rules are at the heart of automatically generated emotional expressivity" (p. 562).

The prosodic dominance indicated in the current study stands somewhat in contrast with Ishii and Kitayama's findings. In their studies, semantics played a larger role than prosody in the processing of spoken emotion for members of Western cultures (Ishii et al., 2003; Kitayama & Ishii, 2002). It is possible that the reason for this difference stems from the items used in our study and the outcome measure tested. Whereas Kitayama and Ishii tested latency for emotional categorization, our dependent measure was the extent of emotional ratings. Perhaps it is easier and quicker to identify semantics, engendering faster latencies, but prosody carries a higher value in the overall rating of the emotion for a given utterance. Moreover, prosodic indicators may be more difficult to derive from single-word utterances (for the time course of prosody detection, see Kotz & Paulmann, 2007; Paulmann & Kotz, 2008). In the T-RES we used spoken sentences rather than the single words used by Kitayama and Ishii. These issues should be addressed in future studies.

Supremacy of Congruency

Across all our analyses and tasks, another common finding emerges—a clear advantage for congruent combinations

of semantics and prosody. In selective-attention tasks, congruent spoken sentences received the highest ratings, even when the listener was explicitly asked to ignore the irrelevant channel. In a similar fashion, when listeners were asked to attend to both channels in the General-rating task, congruent sentences received the highest ratings. These findings are in line with existing literature on the perception of emotions in speech, demonstrating faster processing of congruent combinations as well as greater accuracy and higher ratings (Beaucousin et al., 2007; Mitchell, 2006; Nygaard & Queen, 2008; Wurm et al., 2001). Data from studies also confirm the special role of congruency, showing a unique brain signature (Kotz & Paulmann, 2007; Paulmann et al., 2009; Paulmann & Kotz, 2008).

A notable feature of our paradigm is the inclusion of baseline sentences. These emotion-neutral combinations provide a unique perspective on the possible sources of the supremacy of congruency. Congruent combinations entail a redundancy of information—one channel has enough information to elicit the emotional rating, yet the redundancy of information improves ratings. However, redundancy by itself has not always been found to engender an advantage (Pell et al., 2011). The supremacy of congruency may also arise from the lack of interference from the conflicting channel, as occurs in incongruent combinations (for a discussion, see Ben-David & Algom, 2009). On a further examination of the data in the General-rating task, we can find some evidence for this latter assumption. On average, incongruent sentences that present the rated emotion on one channel and a different emotion on the other were rated lower than baseline sentences, $F(1, 78) = 6.2$, $p < .05$, $\eta_p^2 = .07$. In other words, incongruence by itself has a small but significant impact on emotional ratings.

Our findings correspond with Garner's theory on selective attention to attributes of multidimensional stimuli. In a seminal work, Garner and Felfoldy (1970), distinguished between two types of stimulus dimensions: *integral* dimensions (i.e., dimensions that are perceived united and cannot be perceived without the other) and *separable* dimensions (i.e., dimensions that are perceived as unrelated and are easily pulled apart from one another; see a discussion in Melara & Algom, 2003). Integral dimensions are indicated by a facilitation caused by redundancy of information and failures of selective attention (even if only for one dimension; see Ben-David & Schneider, 2009, 2010). They are usually taken to have some common meaning or hidden relationship (e.g., Stroop or Gestalt). Separable dimensions, on the other hand, are indicated by no advantage for redundancy or toll for incongruence. These dimensions share no common meaning (e.g., color and shape—a green triangle and a blue square). Introducing this terminology to our study, the two emotional speech channels can be taken as integral dimensions. The two are perceived as one and cannot be pulled apart, even when the listeners are specifically asked to do so.

Discrete-Emotion Effects

Examining the discrete emotions separately, we observe the same general effects (in almost all cases), namely

supremacy of congruency, failures of selective attention, and prosodic dominance. This suggests that the effects are not emotion specific but may represent a general quality of perception of emotions in spoken language. The rated emotion was found to yield a difference in the extent of the tested variables, where sad ratings present the largest effect for prosody and fear ratings present the largest effect for semantics. In particular, turn back to Figure 2A, graphically presenting general ratings of emotional sentences. We note a gradient of the effect of prosody (dark-gray bars)—sad (5.2) to anger (4.9) to fear (4.0) to happy (3.9)—and of the effect of semantics (light-gray bars)—fear (3.7) to sad (2.6) to anger (2.4) to happy (2.0). It is interesting that a highly similar gradient is noted in Paulmann and Pell's study (2011, Figure 3). They found the hit-rate gradient for correct identification of emotions presented in the prosody and in the semantics to decline from sad to anger to happy.

One cannot discount the option that the gradient found in the current study may merely reflect the quality of the semantic content of the sentences and the prosody as produced by the actress. However, as indicated in our previous studies (Ben-David, van Lieshout, & Leszcz, 2011; Ben-David, Thayapararajah, & van Lieshout, 2013) both the semantics and the prosody of these spoken sentences were tested extensively and found to be good representatives of their emotional categories. Indeed, note the high ratings when listeners were asked to focus solely on one channel (Figure 2B, dark-gray bars, and 2C, light-gray bars). Moreover, the gradient noted in our data bears a resemblance to Paulmann and Pell's data. Our data thus echoes their conclusion that "some emotions are recognized systematically better … during emotional communication" (2011, p. 200). As a final point, a recent study by Roche et al. (2014) also found variability in processing of different spoken emotions (in prosody).

Caveat and Future Studies

The novel paradigm presented in this study has noted limitations. First, the T-RES uses neutral prosody for a baseline condition. We acknowledge that no prosody can be considered completely emotionless (Vingerhoets, Berckmoes, & Stroobant, 2003), but both behavioral (Pell et al., 2011) and neuroimaging data (Schirmer & Kotz, 2006) support the claim that neutrality is different from other emotional meaning. Second, T-RES stimuli are recorded by a trained professional actress, rather than using natural emotional scenes. Despite the artificial nature of this condition, it provides better control over the variability in the acoustics of the recorded material and yields a more intense and prototypical expression of the specific emotion. This controlled condition adheres to a standard procedure in the literature (Mitchell, 2006). It may be especially relevant for testing specific populations, reducing variability in stimuli with participants who are characterized by high variability in their performance due to variations in sensory and/or cognitive changes (e.g., older adults: Ben-David & Schneider, 2009; people with Alzheimer's disease: Ben-David, Tewari, Shakuf, & van Lieshout, 2014; people with traumatic

brain injury: Ben-David, Nguyen, & van Lieshout, 2011). Third, as noted by Kitayama and Ishii (Ishii et al., 2003; Kitayama & Ishii, 2002), the perception of emotions in speech may be culturally bound. In that case, future studies should examine the validity of our main conclusions when testing other English-speaking cultures and other languages and cultures entirely (see Icht & Ben-David, 2014, 2015). Fourth, to allow for a fine-grained analysis of the data, the experimental procedure takes about 1 hr. To facilitate the process and adapt it to clinical and experimental use with other populations, a shortened version of the paradigm might prove useful; this is currently being investigated in our lab. As a final point, difficulties in identification of emotions in speech (see Bagby, Parker, & Taylor, 1994) present a challenge for many clinical populations and lead to social isolation and depression (Honkalampi, Hintikka, Tanskanen, Lehtonen, & Viinamäki, 2000). For example, in traumatic brain injury, these difficulties have been found to be highly related to a reduced quality of life (Henry, Phillips, Crawford, Theodorou, & Summers, 2006). These findings led the authors to conclude that locating the potential sources of these problems in patients is an "urgent priority." We hope that the T-RES can be used to augment existing tools and assist in discerning between the possible sources of difficulties in identification of emotions in spoken language in various populations.

Summary and Implications

The current study presents a novel tool, the T-RES, for testing the identification of emotions in spoken language. With 80 healthy young North American English–speaking undergraduate-student participants, our results support the validity of the tool in assessing the complex interplay between the prosodic and semantic channels of speech and their impact on rating of emotions. Our analyses lead to three main conclusions: (a) supremacy of congruency—presenting the same emotion in both the prosodic and semantic channels inflates its emotional value; (b) prosody and semantics are integral, and it is difficult to perceive one without the influence of the other; and (c) prosodic dominance—prosody has a larger impact on emotional rating of speech than semantics. We believe that these conclusions can be used in both clinical and experimental settings to improve communication techniques. We hope that this novel tool in future research may provide a path to better understanding of the difficulties in processing of emotional speech with populations with pathologies.

Acknowledgments

References

Algom, D., Chajut, E., & Lev, S. (2004). A rational look at the emotional Stroop phenomenon: A generic slowdown, not a Stroop effect. *Journal of Experimental Psychology: General, 133,* 323–338. doi:10.1037/0096-3445.133.3.323

Ardila, A. (2007). Normal aging increases cognitive heterogeneity: Analysis of dispersion in WAIS-III scores across age. *Archives of Clinical Neuropsychology, 22,* 1003–1011.

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

Bagby, R. M., Parker, J. D. A., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia Scale—I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research, 38,* 23–32.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English Lexicon Project. *Behavioral Research Methods, 39,* 445–459.

Beaucousin, V., Lacheret, A., Turbelin, M.-R., Morel, M., Mazoyer, B., & Tzourio-Mazoyer, N. (2007). fMRI study of emotional speech comprehension. *Cerebral Cortex, 17,* 339–352. doi:10.1093/cercor/bhj151

Ben-David, B. M., & Algom, D. (2009). Species of redundancy in visual target detection. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 958–976. doi:10.1037/a0014511

Ben-David, B. M., Multani, N., Durham, N. A.-M., Green, R., & van Lieshout, P. H. H. M. (2014, July). *A new tool for assessing impairments of identification of emotions in spoken language following traumatic brain injury—the Test of Rating of Emotions in Speech (T-RES)*. Paper presented at the 2014 Mid-Year Meeting of the International Neuropsychological Society, Jerusalem, Israel.

Ben-David, B. M., Multani, N., Shakuf, V., & van Lieshout, P. H. H. M. (2013, June). *Inter-generational communication breakdown: Age-related differences in processing of emotions in speech*. Paper presented at the 20th World Congress of the International Association of Gerontology and Geriatrics, Seoul, South Korea.

Ben-David, B. M., Nguyen, L. L. T., & van Lieshout, P. H. H. M. (2011). Stroop effects in persons with traumatic brain injury: Selective attention, speed of processing, or color-naming? A meta-analysis. *Journal of International Neuropsychological Society, 17,* 354–363. doi:10.1017/S135561771000175X

Ben-David, B. M., & Schneider, B. A. (2009). A sensory origin for color-word Stroop effects in aging: A meta-analysis. *Aging, Neuropsychology, and Cognition, 16,* 505–534. doi:10.1080/13825580902855862

Ben-David, B. M., & Schneider, B. A. (2010). A sensory origin for color-word Stroop effects in aging: Simulating age-related changes in color-vision mimics age-related changes in Stroop. *Aging, Neuropsychology, and Cognition, 17,* 730–746. doi:10.1080/13825585.2010.510553

Ben-David, B. M., Tewari, A., Shakuf, V., & van Lieshout, P. H. H. M. (2014). Stroop effects in Alzheimer's disease: Selective attention, speed of processing, or color-naming? A meta-analysis. *Journal of Alzheimer's Disease, 38,* 923–938.

Ben-David, B. M., Thayapararajah, A., & van Lieshout, P. H. H. M. (2013). A resource of validated digital audio recordings to assess identification of emotion in spoken language after a brain injury. *Brain Injury, 27,* 248–250. doi:10.3109/02699052.2012.740648

Ben-David, B. M., Tse, V. Y. Y., & Schneider, B. A. (2012). Does it take older adults longer than younger adults to perceptually

segregate a speech target from a background masker? *Hearing Research, 290,* 55–63. doi: 10.1016/j.heares.2012.04.022

Ben-David, B. M., van Lieshout, P. H. H. M., & Leszcz, T. (2011). A resource of validated affective and neutral sentences to assess identification of emotion in spoken language after a brain injury. *Brain Injury, 25,* 206–220. doi:10.3109/02699052.2010.536197

Bertels, J., Kolinsky, R., Pietrons, E., & Morais, J. (2011). Long-lasting attentional influence of negative and taboo words in an auditory variant of the emotional Stroop task. *Emotion, 11,* 29–37. doi:10.1037/a0022017

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

Bowers, D., Blonder, L. X., & Heilman, K. M. (1999). *Florida Affect Battery: A manual.* Gainesville, FL: Center for Neuropsychological Studies.

Breitenstein, C., Daum, I., & Ackermann, H. (1998). Emotional processing following cortical and subcortical brain damage: Contribution of the fronto-striatal circuitry. *Behavioural Neurology, 11*(1), 29–42.

Brown, R. (1973). *A first language: The early stages.* Cambridge, MA: Harvard University Press.

Carton, J. S., Kessler, E. A., & Pape, C. L. (1999). Nonverbal decoding skills and relationship well-being in adults. *Journal of Nonverbal Behavior, 23,* 91–100.

Cicero, B. A., Borod, J. C., Santschi, C., Erhan, H. M., Obler, L. K., Agosti, R. M., . . . Grunwald, I. S. (1999). Emotional versus non-emotional lexical perception in patients with right and left brain damage. *Neuropsychiatry, Neuropsychology, & Behavioral Neurology, 12,* 255–264.

Craik, F. I. M. (2002). Levels of processing: Past, present … and future? *Memory, 10,* 305–318. doi:10.1080/09658210244000135

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671–684.

Dupuis, K., & Pichora-Fuller, M. K. (2010). Use of affective prosody by young and older adults. *Psychology and Aging, 25,* 16–29. doi:10.1037/a0018777

Egloff, B., & Schmukle, S. C. (2004). Gender differences in implicit and explicit anxiety measures. *Personality and Individual Differences, 36,* 1807–1815.

Ethofer, T., Anders, S., Erb, M., Herbert, C., Wiethoff, S., Kissler, J., . . . Wildgruber, D. (2006). Cerebral pathways in processing of affective prosody: A dynamic causal modeling study. *Neuro-Image, 30,* 580–587. doi:10.1016/j.neuroimage.2005.09.059

Garner, W. R., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology, 1,* 225–241.

Grant, D. A. (1948). The latin square principle in the design and analysis of psychological experiments. *Psychological Bulletin, 45,* 427–442.

Grimshaw, G. M., Séguin, J. A., & Godfrey, H. K. (2009). Once more with feeling: The effects of emotional prosody on hemispheric specialisation for linguistic processing. *Journal of Neurolinguistics, 22,* 313–326.

Henry, J. D., Phillips, L. H., Crawford, J. R., Theodorou, G., & Summers, F. (2006). Cognitive and psychosocial correlates of alexithymia following traumatic brain injury. *Neuropsychologia, 44,* 62–72.

Honkalampi, K., Hintikka, J., Tanskanen, A., Lehtonen, J., & Viinamäki, H. (2000). Depression is strongly associated with alexithymia in the general population. *Journal of Psychosomatic Research, 48,* 99–104.

Icht, M., & Ben-David, B. M. (2014). Oral-diadochokinesis rates across languages: English and Hebrew norms. *Journal of Communication Disorders, 48,* 27–37.

Icht, M., & Ben-David, B. M. (2015). Oral-diadochokinetic rates for Hebrew-speaking school-age children: Real words vs. non-words repetition. *Clinical Linguistics & Phonetics, 29,* 102–114.

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

Jacob, H., Brück, C., Plewnia, C., & Wildgruber, D. (2014). Cerebral processing of prosodic emotional signals: Evaluation of a network model using rTMS. *PloS ONE, 9*(8), e105509.

Jones, D. (2010, June 25). A WEIRD view of human nature skews psychologists' studies. *Science, 328*(5986), 1627.

Joukamaa, M., Saarijärvi, S., Muuriaisniemi, M.-L., & Salokangas, R. K. R. (1996). Alexithymia in a normal elderly population. *Comprehensive Psychiatry, 37,* 144–147.

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

Kitayama, S. (1996). Remembrance of emotional speech: Improvement and impairment of incidental verbal memory by emotional voice. *Journal of Experimental Social Psychology, 32,* 289–308.

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

Kotz, S. A., Dengler, R., & Wittfoth, M. (2015). Valence-specific conflict moderation in the dorso-medial PFC and the caudate head in emotional speech. *Social Cognitive and Affective Neuroscience, 10*(2), 165–171.

Kotz, S. A., & Paulmann, S. (2007). When emotional prosody and semantics dance cheek to cheek: ERP evidence. *Brain Research, 1151,* 107–118. doi:10.1016/j.brainres.2007.03.015

Laukka, P. (2003). Categorical perception of emotion in vocal expression. *Annals of the New York Academy of Sciences, 1000,* 283–287.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers, 28*(2), 203–208.

Mehrabian, A., & Wiener, M. (1967). Decoding of inconsistent communications. *Journal of Personality and Social Psychology, 6,* 109–114.

Melara, R. D., & Algom, D. (2003). Driven by information: A tectonic theory of Stroop effects. *Psychological Review, 110,* 422–471.

Mitchell, R. L. C. (2006). How does the brain mediate interpretation of incongruent auditory emotions? The neural response to prosody in the presence of conflicting lexico-semantic cues. *European Journal of Neuroscience, 24,* 3611–3618. doi:10.1111/j.1460-9568.2006.05231.x

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

Mitchell, R. L. C., & Ross, E. D. (2008). fMRI evidence for the effect of verbal complexity on lateralisation of the neural response associated with decoding prosodic emotion. *Neuropsychologia, 46,* 2880–2887. doi:10.1016/j.neuropsychologia.2008.05.024

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

Nowicki, S., Jr., & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: The Diagnostic

Analysis of Nonverbal Accuracy scale. *Journal of Nonverbal Behavior, 18,* 9–35.

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

Nygaard, L. C., & Queen, J. S. (2008). Communicating emotion: Linking affective prosody and word meaning. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 1017–1030. doi:10.1037/0096-1523.34.4.1017

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

Paulmann, S., & Kotz, S. A. (2008). An ERP investigation on the temporal dynamics of emotional prosody and emotional semantics in pseudo- and lexical-sentence context. *Brain and Language, 105,* 59–69. doi:10.1016/j.bandl.2007.11.005

Paulmann, S., & Pell, M. D. (2011). Is there an advantage for recognizing multi-modal emotional stimuli? *Motivation and Emotion, 35,* 192–201.

Paulmann, S., Pell, M. D., & Kotz, S. A. (2008). Functional contributions of the basal ganglia to emotional prosody: Evidence from ERPs. *Brain Research, 1217,* 171–178. doi:10.1016/j.brainres.2008.04.032

Pell, M. D., Jaywant, A., Monetta, L., & Kotz, S. A. (2011). Emotional speech processing: Disentangling the effects of prosody and semantic cues. *Cognition & Emotion, 25,* 834–853. doi:10.1080/02699931.2010.516915

Pell, M. D., Monetta, L., Paulmann, S., & Kotz, S. A. (2009). Recognizing emotions in a foreign language. *Journal of Nonverbal Behavior, 33,* 107–120.

Pell, M. D., Paulmann, S., Dara, C., Alasseri, A., & Kotz, S. A. (2009). Factors in the recognition of vocally expressed emotions: A comparison of four languages. *Journal of Phonetics, 37,* 417–435.

Raven, J. C. (1965). *Guide to using the Colored Progressive Matrices Sets A, Ab, and B.* Dumfries, Scotland: Griem.

Roche, J. M., Petters, B., & Dale, R. (2014). "Your tone says it all": The processing and interpretation of affective language. *Speech Communication, 66,* 47–64.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 40,* 227–256.

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology, 32,* 76–92.

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

Schröder, M. (2001). Emotional speech synthesis: a review. In P. Dalsgaard, B. Lindberg, H. Benner, & Z.-H. Tan (Eds.), *Eurospeech 2001 Scandinavia: 7th European Conference on Speech Communication and Technology* (pp. 561–564). Baixas, France: International Speech Communication Association.

Schwartz, R., & Pell, M. D. (2012). Emotional speech processing at the intersection of prosody and semantics. *PLoS ONE, 7*(10), e47279.

Shakuf, V., & Ben-David, B. M. (2014, February). *Who is afraid of the bear? Hebrew test for recognition of emotions in spoken language for the diagnosis and evaluation of abilities and disabilities.* Paper presented at the 50th Annual Meeting of the Israeli Speech, Hearing and Language Association, Tel Aviv, Israel.

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research, 29,* 557–580.

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

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,* 427–437. doi:10.1016/j.cogbrainres.2004.03.015

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale–Fourth Edition.* San Antonio, TX: Pearson.

Wildgruber, D., Ackermann, H., Kreifelts, B., & Ethofer, T. (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. *Progress in Brain Research, 156,* 249–268. doi:10.1016/S0079-6123(06)56013-3

Wittfoth, M., Schröder, C., Schardt, D. M., Dengler, R., Heinze, H.-J., & Kotz, S. A. (2010). On emotional conflict: Interference resolution of happy and angry prosody reveals valence-specific effects. *Cerebral Cortex, 20,* 383–392. doi:10.1093/cercor/bhp106

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.

Zupan, B., Neumann, D., Babbage, D. R., & Willer, B. (2009). The importance of vocal affect to bimodal processing of emotion: Implications for individuals with traumatic brain injury. *Journal of Communication Disorders, 42,* 1–17.

Appendix A

Table A1. Means (and *SD*s) of ratings of the semantics of the 50 sentences used in our study, on four emotional scales, by a group of native English speakers.

Emotional categories	N	Ratings on emotional scales			
		Anger	Fear	Happiness	Sadness
Anger	10	**5.3 (0.3)**	1.9 (0.2)	1.4 (0.2)	2.1 (0.4)
Fear	10	2.2 (0.3)	**5.5 (0.2)**	1.5 (0.2)	1.8 (0.3)
Happiness	10	1.1 (0.1)	1.2 (0.1)	**5.6 (0.2)**	1.2 (0.1)
Sadness	10	2.0 (0.3)	2.1 (0.3)	1.3 (0.2)	**5.4 (0.2)**
Neutral	10	1.3 (0.1)	1.3 (0.1)	2.2 (0.1)	1.3 (0.1)

Note. Participants were asked to read the printed sentences. Ratings were on a 6-point Likert scale, ranging from 1 (*completely disagree* that the sentence conveys the relevant emotion) to 6 (*completely agree*). Boldface data cells present average ratings of affective sentences on their corresponding emotional scales. Table from Ben-David, Thayapararajah, and van Lieshout (2013). Copyright © Taylor & Francis LLC. Reprinted with permission of Taylor & Francis LLC (http://www.tandfonline.com); non-exclusive English rights only.

Table A2. Means (and *SD*s) of linguistic characteristics of the 50 sentences used in our study.

Semantic categories	N	Linguistic characteristics		
		Number of syllables	Frequency (HAL)	Phonologic neighborhood
Anger	10	6.30 (1.9)	10.55 (1.8)	14.45 (12.4)
Fear	10	7.00 (2.0)	10.06 (1.4)	14.48 (10.5)
Happiness	10	7.00 (1.4)	10.14 (1.9)	9.03 (3.4)
Sadness	10	6.50 (1.3)	10.06 (1.4)	11.36 (8.6)
Neutral	10	6.70 (0.9)	10.03 (1.1)	13.64 (5.8)

Note. HAL = Hyperspace Analogue to Language (Lund & Burgess, 1996). Linguistic characteristics are taken from the English Lexicon Project database (Balota et al., 2007). Table from Ben-David, Thayapararajah, and van Lieshout (2013). Copyright © Taylor & Francis LLC. Reprinted with permission of Taylor & Francis LLC (http://www.tandfonline.com); non-exclusive English rights only.

Appendix B

Effects for Discrete Emotions

Semantics- and Prosody-Rating Tasks

RAC in Baseline Sentences

We found a Rated Emotion (anger, fear, sad, happy) × RAC interaction, $F(3, 216) = 21.86$, $p < .001$, $\eta_p^2 = .23$. Separate post hoc ANOVAs (Bonferroni corrected) for each rated emotion show that the main effect of RAC was significant for all four rated emotions, $F(1, 72) > 1,210$, all ps < .0001, but to a different degree, with the largest effect for happiness ratings (4.4) and the lowest for fear ratings (3.8).

RIC in Baseline Sentences

We note a triple Rated Emotion (anger, fear, sad, happy) × RIC × Target Channel interaction, $F(3, 77) = 10.6$, $p < .001$, $\eta_p^2 = .29$. In all four post hoc tests, for each rated emotion, we found the same trend, with nominally larger RIC when the listener was asked to ignore the prosody. However, the Target Channel × RIC interaction was only significant (after Bonferroni correction) for two rated emotions: happy, $F(1, 72) = 53.3$, $p < .001$, $\eta_p^2 = .43$, and fear, $F(1, 72) = 39.28$, $p < .001$, $\eta_p^2 = .35$.

RIC in Emotional Sentences

The failure of selective attention to the target channel interacted with the rated emotion, $F(3, 216) = 28.38$, $p < .001$, $\eta_p^2 = .28$. In post hoc tests (one for each rated emotion), we found a significant effect of RIC for all rated emotions, $F(1, 72) = 53.3$, $p < .001$, $\eta_p^2 = .43$, except happy (the effect was not significant after Bonferroni correction), $F(1, 72) = 4.2$.

General-Rating Task

RAC in Baseline Sentences

The prosodic dominance indicated in the general analysis was found to interact with the rated emotion, $F(3, 70) = 22.34$, $p < .001$, $\eta_p^2 = .49$. The four post hoc tests indicate that the RAC × Target Channel interaction was significant for all rated emotions, $F(1, 72) > 110$, all ps < .001, but varied in extent, with the most extreme difference for anger ratings ($RAC_{prosody} = 4.0$ vs. $RAC_{semantics} = 0.2$) and the least extreme for sad ratings ($RAC_{prosody} = 3.2$ vs. $RAC_{semantics} = 1.1$). (For prosodic advantage and dimensional integration in emotional sentences, see the Discrete-Emotion Analysis section.)

Congruency Supremacy: Redundancy of Information

In the general ANOVA for prosody and for semantics, we found a Rated Emotion × Redundancy interaction—prosody: $F(3, 70) = 17.04$, $p < .001$, $\eta_p^2 = .42$; semantics: $F(3, 70) = 24.36$, $p < .001$, $\eta_p^2 = .51$. For prosody (where congruent combinations rated higher than semantically neutral sentences with the rated emotion in the prosody), examining the four rated emotions separately, we find the supremacy of congruency significant for happy and fear ratings, $F(1, 72) > 18.3$, $p < .001$, but not for sad ratings, $F < 1$ (same direction), or anger ratings, $F(1, 72) = 12.2$, $p = .001$ (opposite direction, with anger prosody rated 0.3 higher than congruent sentences). For semantics, the supremacy of congruency was found in all four rated emotions, $F(1, 72) > 83$, $p = .001$, with the largest effect for happy ratings (3.9).