

Recognizing emotions in spoken language: A validated set of Portuguese sentences and pseudosentences for research on emotional prosody

SÃO LUÍS CASTRO AND CÉSAR F. LIMA
University of Porto, Porto, Portugal

A set of semantically neutral sentences and derived pseudosentences was produced by two native European Portuguese speakers varying emotional prosody in order to portray anger, disgust, fear, happiness, sadness, surprise, and neutrality. Accuracy rates and reaction times in a forced-choice identification of these emotions as well as intensity judgments were collected from 80 participants, and a database was constructed with the utterances reaching satisfactory accuracy (190 sentences and 178 pseudosentences). High accuracy (mean correct of 75% for sentences and 71% for pseudosentences), rapid recognition, and high-intensity judgments were obtained for all the portrayed emotional qualities. Sentences and pseudosentences elicited similar accuracy and intensity rates, but participants responded to pseudosentences faster than they did to sentences. This database is a useful tool for research on emotional prosody, including cross-language studies and studies involving Portuguese-speaking participants, and it may be useful for clinical purposes in the assessment of brain-damaged patients. The database is available for download from <http://brm.psychonomic-journals.org/content/supplemental>.

The ability to correctly perceive the emotional state of a speaker is crucial for successful communication. Perceptual cues related to the tone of voice, such as changes in pitch, loudness, duration, and voice quality, collectively referred to as *emotional prosody*, are an important source of information that listeners use to infer how the speaker feels (e.g., Belin, Fecteau, & Bédard, 2004; Schirmer & Kotz, 2006). For instance, how we realize that someone is angry, scared, or surprised depends to a large extent on how the tone of voice sounds. Several studies have shown that the vocal expression of emotions is associated with specific acoustic profiles (for a review, see Juslin & Laukka, 2003), and, in recent years, behavioral (e.g., Laukka, 2005; Laukka & Juslin, 2007; Paulmann, Pell, & Kotz, 2008b), neuropsychological (e.g., Adolphs, Damasio, & Tranel, 2002; Adolphs, Tranel, & Damasio, 2001; Pell & Leonard, 2003; Ross & Monnot, 2008), neuroimaging (e.g., Grandjean et al., 2005; Mitchell, Elliott, Barry, Cruttenden, & Woodruff, 2003; Mitchell & Ross, 2008), and electrophysiological (e.g., Paulmann & Kotz, 2008; Paulmann, Pell, & Kotz, 2008a) studies have been conducted in order to determine which cognitive and neural mechanisms underlie the processing of emotional prosody. Well-devised and validated stimuli are a prerequisite to conducting this research (e.g., Pell, 2002; Ross, Thompson, & Yenkosky, 1997). Because such stimuli have at least some degree of language specificity, validated sets

are needed for different languages. With this article, we provide a database of validated stimuli in European Portuguese for research on emotional prosody.

Different types of vocal stimuli, such as sentences, single words, or monosyllabic utterances, have been used to shed light on the neurocognitive foundations of emotional prosody. Issues that have to be dealt with when devising such stimuli include the interplay of prosody with semantics, the linguistic complexity of the utterances, and language specificity. The potential interplay between the expressiveness embodied in prosody proper and the semantic content carried by words or sentences has been addressed experimentally by manipulating both aspects orthogonally and then comparing the conditions in which semantics and prosody are emotionally congruent or incongruent. For example, Mitchell et al. (2003) have done this in a study that used functional magnetic resonance imaging to determine the neural correlates of emotional prosody processing. They have identified moderately right-lateralized responses in the lateral temporal lobes (middle and superior temporal gyri) for congruent as well as for incongruent conditions. These two conditions also induced differential activations in the medial frontal gyrus; incongruent prosody activated fewer brain regions than congruent prosody did, which, according to the authors, might reflect an attentional effect of participants “disengaging themselves from prosodic processing cognitions” (Mitchell et al.,

S. L. Castro, slcastro@fpce.up.pt



2003, p. 1419) in order to deal with the unexpected conflict between prosody and semantics. Furthermore, behavioral results showed that emotional prosody is better identified in the congruent condition than in the incongruent one. Given that trade-off effects are bound to occur when semantics are congruent, or incongruent, with prosody, it appears preferable—unless one is interested in addressing the role of semantics in prosody processing—to resort to vocal stimuli that express emotional qualities in emotionally neutral carrier words and sentences. This is, in fact, a very common procedure reported in the literature on emotional prosody (e.g., Adolphs et al., 2002; Adolphs et al., 2001; Wildgruber et al., 2005).

Another strategy that has been used to control for the influence of semantics on prosody is to manipulate the linguistic complexity of the utterances. This can be achieved either by removing semantics through filtering procedures that preserve only suprasegmental features (e.g., Kotz et al., 2003) or by comparing stimuli with different degrees of linguistic content, as has been done in studies using the Aprosodia Battery (e.g., Mitchell & Ross, 2008; Ross & Monnot, 2008). The Aprosodia Battery (Ross et al., 1997) employs six intonations—angry, disinterested, happy, sad, surprised, and neutral—in three conditions of verbal complexity. In the first condition, the intonations are expressed in short sentences with emotionally neutral semantic content (such as “the door is ajar”). In the second condition, the intonations are expressed in utterances composed only with “ba”s, and in the third condition they consist of a prolonged “aaaaahhhh.” Results from healthy participants have shown that accuracy was similarly high in the three conditions, about 70% correct, thus indicating that the absence of semantic content was not detrimental to the recognition of emotional prosody (Mitchell & Ross, 2008; Ross & Monnot, 2008). These studies have also shown that, although both the left and the right lateral temporal lobes are recruited by emotional prosody, the contribution of the left hemisphere is modulated by the linguistic complexity of the stimuli: It got smaller as linguistic complexity decreased. This has led the authors to conclude that emotional prosody is a right-lateralized function, the contribution of the left hemisphere being probably more related with processing linguistic information, such as phonetic-segmental characteristics and lexico-semantic content. Indeed, a neurocognitive dissociation between semantic and syntactic processes versus prosodic processes has been postulated by Friederici and Alter (2004) in their dynamic dual-pathway model of language functions. According to this model, a left-lateralized temporo-frontal network supports syntactic and semantic processes, and the right hemisphere is dominant for prosodic processes, depending on the linguistic demands of the stimuli or task (the higher these are, the greater is the involvement of the left hemisphere). Perhaps a more natural way to reduce the linguistic complexity of the stimuli (neither through filtering, nor by using repetitive syllables or monosyllables) is to use meaningless speech and pseudosentences—that is, sentences composed of pseudowords together with a few function words (e.g., Bach et al., 2008; Pell, 2002). The lexico-semantic content of these utterances is sub-

stantially reduced, because pseudowords have no meaning, yet they afford a good language-like quality because both phonetic-segmental and suprasegmental features of speech are present. Pseudosentences have elicited high accuracy rates in forced-choice classification, with agreement levels above 80% for some emotions (Pell, 2002). They have also been used successfully in neuropsychological research. For example, Pell and colleagues (Dara, Monetta, & Pell, 2008; Pell & Leonard, 2003) showed that patients with Parkinson’s disease had impaired performance when compared with healthy controls. Altogether, these findings indicate that vocal stimuli of relatively low linguistic complexity are well suited to engage prosodic processing.

The third aspect to be considered is language specificity. Unlike for other emotional stimuli, such as facial expressions (e.g., Ekman & Friesen, 1978), pictures (e.g., Lang, Bradley, & Cuthbert, 2005), and music (Vieillard et al., 2008), for emotional prosody, it is necessary to devise stimuli that take into account how a particular language is understood by its listeners. The main goal of the present study was to devise a set of vocal stimuli in European Portuguese for research on emotional prosody, a research tool that may be useful not only for research with Portuguese-speaking participants, but also for cross-language studies. Cross-language comparisons are essential in determining whether the recognition of emotions in voice is universal or rather is shaped by language and culture (Pell, Monetta, Paulmann, & Kotz, 2009; Pell & Skorup, 2008; Thompson & Balkwill, 2006). We selected two types of vocal stimuli that appear well suited to focus on prosody and to put few demands on semantic and syntactic processing: simple sentences with emotionally neutral content and derived pseudosentences. Because these stimuli afford a reasonable degree of phonetic and phonological differentiation, they may be especially well suited to explore the role of language-general versus language-specific processes in relation to emotional prosody. The stimuli were produced by two Portuguese native speakers in a neutral intonation and in six other tones of voice intended to express the basic emotions that are usually examined in studies of prosody and facial expression: anger, disgust, fear, happiness, sadness, and surprise. This set of stimuli was validated by a group of 80 participants who performed emotion-identification and intensity-judgment tasks. Reaction times (RTs) were also collected. A secondary goal of the experiment was to compare sentences and pseudosentences on accuracy rates, RTs, and intensity judgments. Since prosodically relevant acoustic properties such as intensity, duration, and fundamental frequency were equivalent across sentences and pseudosentences, we hypothesized that both types of stimuli would pose similar demands on the resources required for prosodic processing and thus would yield similar accuracy and intensity judgments. However, we predicted that RTs for pseudosentences would be faster than those for sentences because pseudosentences present less semantic information to be processed. Although sentences and pseudosentences were equivalent in phonetic-segmental complexity, we predicted that the near absence

of lexico-semantic content in the pseudosentences would decrease the demands on the resources that are responsible for semantic processing. The lack of semantic content in the pseudosentences might enhance response speed by decreasing the processing cost due to the integration of emotional prosody with linguistic content (Schirmer & Kotz, 2006).

METHOD

Recording

Participants. We conducted pilot tests with several candidates and selected 2 women (mean age = 18 years), who were native speakers of European Portuguese, to record the stimuli. Both had formal musical training that included singing lessons and were selected on the basis of their ability to modulate emotional prosody.

Materials. A set of 16 short Portuguese sentences and 16 pseudosentences was devised. The complete list can be inspected in the Appendix. The sentences were syntactically simple and composed of high-frequency words; they had no emotional content, as assessed by three independent judges. Mean length was 8 syllables ($SD = 1.3$, range = 6–11). The pseudosentences were derived from the sentences, yet respected the syntax, morphology, and phonotactics of Portuguese. The procedure was the following: One to three phonemes of the content words of the sentences (nouns, adjectives, and verbs carrying salient semantic information) were replaced, such that pseudowords were formed. Vowels were replaced by vowels and consonants by consonants, and the syllabic structure and stress of the original word were kept. In this manner, a sentence was transfigured into a pseudosentence. For example, *O quadro está na parede* (“the painting is on the wall”) was transformed into *O quadro está na pafêne*. Since the pseudosentences retained the phonological properties of Portuguese, they were highly language-like; that is, they were “nonsensical but an appropriate carrier of suprasegmental information” (Pell, 2002, p. 500).

Procedure. Two separate recording sessions were conducted, one for each speaker, each lasting about 4 h. Both speakers were asked to produce the sentences and pseudosentences in a neutral manner and in the six emotional expressions (anger, disgust, fear, happiness, sadness, and surprise). A list of the 16 sentences and corresponding pseudosentences was given to them 3 days before the recording sessions so that they could familiarize themselves with the materials. They were also informed that they would be asked to produce each sentence and pseudosentence in seven different “tones of voice” (neutral, angry, happy, etc.) and were asked to try these out by producing some of them as if they were feeling angry, happy, and so forth. In the recording session proper, they were instructed to produce the sentences and pseudosentences with the clearest, but still natural, target emotional tone. The general instruction was again: “Say this as if you were feeling [the target emotion].” In order to enhance expressiveness, the authors (S.L.C. and C.F.L.) coached the speakers by providing them with a corresponding scenario (e.g., “imagine that you are very happy; you have just been given fantastically good news. Think of a situation where that really happened, put yourself in the mood, smile! [pause] Done? OK, then let’s speak out these sentences in a happy manner!”). No specific instructions concerning how to achieve the emotional tones were given (high- or low-pitched voice, slow or quick paced, etc.). The authors also provided feedback and fine-tuning of the expressions with particular regard for naturalness and perceived spontaneity. When necessary, a stimulus was produced several times until being qualitatively evaluated as a good instance of the intended expression. The speakers produced subtypes of some emotions more than others: Anger was produced mainly in hot rather than cold form, sadness as quiet sadness rather than despair, and fear as a sustained or milder state rather than panic (Banse & Scherer, 1996; Scherer, Johnstone, & Klasmeyer, 2003).

The stimuli were recorded in the sound-insulated booth of the Speech Laboratory at the University of Porto Psychology Department, using Pro Tools LE version 5.1.1 software and a high-quality microphone attached to an Apple Macintosh G4 computer. Digitization was done at a 48-kHz sampling rate and 16-bit resolution. Individual files were prepared with the best productions of each sentence and pseudosentence from each speaker according to the judgment of the two authors and informed, in case of doubt or disagreement, by judgments from a small number of lab colleagues. The sound files were normalized for average amplitude using Sound Studio version 3.5.5, and 50 msec of silence were added at the beginning and end of the utterance. A total of 448 individual files/stimuli (32 sentences/pseudosentences \times 7 intonations \times 2 speakers) was then submitted to the validation procedure.

Validation

Participants. We recruited 80 University of Porto undergraduate students (72 females; mean age = 21.8 years, $SD = 6.1$), who participated in exchange for course credit. All were native speakers of European Portuguese and reported no hearing impairments or speech disorders, no major psychiatric or neurological illnesses, and no head trauma or substance abuse. They were divided into four groups of 20 participants, each of which judged either the sentences or the pseudosentences produced by one of the speakers.

Procedure. Participants were tested individually in one experimental session lasting about 25 min. There were four between-subjects conditions, resulting from the combination of two speakers with two types of stimuli. Each participant evaluated 112 stimuli—that is, all the sentences, or pseudosentences, produced by one speaker (16×7). In each condition, the stimuli were pseudorandomized and were divided into two blocks of 56 trials, which were counterbalanced across participants. Participants were told that they would listen to simple sentences or to meaningless pseudosentences that had been produced in one of seven different intonations: neutral, happy, sad, angry, fearful, disgusted, and surprised (*neutro, alegre, triste, zangado, assustado, repulsivo, and surpreendido*, respectively). The stimuli were presented via high-quality headphones. Participants were asked to perform two independent consecutive judgments for each stimulus. First, they had to identify as fast and accurately as possible which of those expressions characterized the stimulus (forced choice), and then they had to make an intensity judgment evaluating how well the expression was displayed—that is, how representative the stimulus was of the chosen category. The responses were given by pressing prespecified buttons on a 7-button response pad, model RB-730 from Cedrus Corporation. Seven labels for the expression categories (in the above order) were affixed below the buttons of the response pad, and the numbers 1 (*low intensity*) to 7 (*high intensity*) were affixed above the buttons. Participants were instructed to respond with the index finger of their dominant hand and to keep the hand in the same position, at the center of the response pad, throughout the session.

The presentation of the stimuli and the recording of RTs and accuracy were controlled with SuperLab V4.0 (Abboud, Schultz, & Zeitlin, 2006) running on a Macintosh PowerBook G4 computer. A trial consisted of the following events: a plus sign (+) acting as a warning signal appeared on the center of the screen for 500 msec; the stimulus was then presented auditorily while the fixation point remained visible; after the participant identified the emotion, the 7-point scale appeared on screen, prompting the participant to make the intensity judgment; after the response, there was a 3,000-msec ISI with a blank screen. In order to familiarize the participants with the task and to adjust headphone volume, the session started with four practice trials, after which the first block was presented. There was a short pause before the second block. RTs were measured from stimulus onset until the first buttonpress (corresponding to the identification response).

Selection. For each stimulus, the mean percentage of correct identifications of the target emotion was computed using the judg-

Table 1
**Mean Fundamental Frequency (f_0), f_0 Variation (SD),
 and Duration of the Sentences and Pseudosentences**

Emotion Category	f_0 (Hz)		f_0 Variation (SD)		Duration (msec)	
	Sentences	Pseudosentences	Sentences	Pseudosentences	Sentences	Pseudosentences
Neutrality	211	214	38	42	1,477	1,484
Happiness	357	350	81	83	1,479	1,513
Sadness	206	205	38	39	1,532	1,530
Anger	313	307	67	62	1,266	1,277
Fear	302	317	37	43	1,445	1,457
Disgust	279	276	67	66	1,634	1,580
Surprise	332	323	71	80	1,435	1,446

Note—Values are averaged across speakers.

ments from 20 subjects. The stimuli that reached a satisfactory level of identification accuracy were selected for inclusion in the database. Only stimuli fulfilling the conditions of an accuracy rate of at least 43% correct, three times the level of chance, and of less than 43% rating in any of the other categories were considered satisfactory to be included. This double criterion was used instead of one based solely on a higher consensus on the target emotion (such as the 60% consensus in Pell, 2002), in order to have a varied set of stimuli in the database, including some that fell within a range of relative ambiguity because these may be useful in order to manipulate this variable. From the initial set of stimuli, 80 were excluded and 368 were selected. The mean number of stimuli per emotion was 26 ($SD = 6.4$). These were included in the database and were submitted to the further analyses that are presented below.

RESULTS AND DISCUSSION

This database for research on emotional prosody consists of 190 sentences and 178 pseudosentences spoken in European Portuguese and portraying a neutral state and the six basic emotions of anger, disgust, fear, happiness, sadness, and surprise. Detailed acoustic and perceptual characteristics of each stimulus can be found in the supplementary materials. It and the database itself are available for download at brm.psychonomic-journals.org/content/supplemental.

Acoustic Analysis

For each stimulus in the database, mean fundamental frequency (f_0), f_0 variation (standard deviation), and duration were measured using Praat (www.praat.org). These acoustic measures are presented separately for each emotion category in Table 1. Sentences did not differ from pseudosentences in any of the measures ($F_s < 1$). However, emotions differed from one another in mean f_0 , f_0 variation, and duration, as revealed by one-way ANOVAs [respectively, $F(6,361) = 286.94, p < .0001, \eta_p^2 = .83$; $F(6,361) = 54.11, p < .0001, \eta_p^2 = .47$; and $F(6,361) = 12.09, p < .0001, \eta_p^2 = .17$]. These differences are, in general, a replication of findings by Banse and Scherer (1996), Juslin and Laukka (2003), and Pell and Skorup (2008). For instance, happy intonations had high mean f_0 , high f_0 variation, and relatively short duration; whereas sad intonations were relatively longer and had low mean f_0 and f_0 variation, and fearful intonations had high mean f_0 but low f_0 variation. To test whether these differences in f_0 and duration could predict the classification of the

emotion categories, a discriminant analysis was calculated taking the three acoustic measures as independent variables. The acoustic measures were indeed significant predictors of the category membership of the stimuli [Wilks's $\lambda = .11$; $F(18,1015) = 66.94, p < .0001$]; they correctly classified, on average, 57% of the stimuli (disgust, 77%; fear, 66%; happiness, 65%; neutrality, 58%; sadness, 53%; anger, 49%; and surprise, 44%—all well above the 14% chance level).

Recognition Accuracy

Interparticipant reliability in the identification of the intended expression was very high for both speakers and stimulus types: Cronbach's α for speaker M.L.A. was .98 in sentences and .97 in pseudosentences; for speaker M.S., it was .95 in sentences and .96 in pseudosentences. For each stimulus, we computed the number of correct identifications of the intended expression (true positives or sensitivity) and the number of correct rejections of all the remaining nonintended expressions (true negatives or specificity). These rates for individual stimuli were then averaged across speakers by expression category, separately for sentences and pseudosentences (see Table 2). The percent of correct identifications of the intended expression, or sensitivity, is shown in the diagonal cells of Table 2; it reached high rates, ranging from a maximum of 88% for neutrality to a minimum of 50% for disgust, both in sentences. Correct rejections of the nonintended expressions, or specificity, are shown in italics. They were even higher, ranging from 99% for fear to 91% for surprise. High sensitivity indicates that the expressions were correctly recognized when they were presented. The high specificity rates further indicate that the response labels were correctly used when they were selected for response—that is, a specific label tended to be used only when the corresponding category had been presented.

We examined the variation in responses by computing an ANOVA on correct identifications with expression category and stimulus type as between-subjects factors. Tukey HSD tests were used for all the post hoc comparisons. This analysis confirmed that some expressions were better identified than others were [main effect of category: $F(6,354) = 51.14, p < .0001, \eta_p^2 = .46$]. Surprise, sadness, and neutrality yielded the best accuracy (above 80%; $ps < .0001$), followed by anger (76%) and

Table 2
Distribution (%) of the Responses Given to Each of the Intended Expressions
in Sentences and Pseudosentences

Intended Expression	<i>n</i>	Response						
		Neutral	Happy	Sad	Angry	Fearful	Disgusted	Surprised
Sentences								
Neutrality	30	88	0	9	2	0	1	1
Happiness	29	1	75	1	2	1	2	19
Sadness	31	11	1	84	3	1	0	0
Anger	32	8	3	0	77	2	6	3
Fear	24	3	2	5	1	65	2	23
Disgust	12	8	22	1	4	2	50	15
Surprise	32	5	3	1	1	3	1	87
Correct rejection rate (specificity)		94	97	97	98	99	98	91
Pseudosentences								
Neutrality	30	83	1	13	3	0	0	1
Happiness	17	2	59	2	3	4	3	27
Sadness	28	13	2	82	2	1	1	0
Anger	31	4	2	1	74	4	10	4
Fear	17	4	4	7	1	56	5	24
Disgust	23	2	17	0	5	2	60	13
Surprise	32	4	4	0	1	3	2	85
Correct rejection rate (specificity)		95	95	96	98	98	96	91

Note—Recognition accuracy (sensitivity) is indicated in bold. Values are averaged across speakers.

happiness (67%), whereas fear (60%) and disgust (55%) were the least well recognized ($ps < .05$). Disgust was particularly difficult to recognize, which had already been indicated by the fact that fewer of these stimuli reached the criteria for inclusion in the database (only 12/32 sentences and 23/32 pseudosentences from the original set). The difficulty in recognizing disgust has also been observed in other studies of emotional prosody (Adolphs et al., 2002; Banse & Scherer, 1996; Scherer, Banse, Wallbott, & Goldbeck, 1991; Thompson & Balkwill, 2006; for an exception, see Pell, 2002). Scherer et al. (1991) interpreted it as evidence that the natural expression of disgust tends to consist of brief affective bursts or vocal emblems, such as “yuck!” rather than of a tone of voice imparting over longer stretches of speech, such as whole sentences. Consistent with this view, disgust was in fact one of the better recognized (81%) categories of affective bursts in the recent study by Belin, Fillion-Bilodeau, and Gosselin (2008). Another reason why disgust may not be easily recognized from emotional prosody is that in natural communication this emotional quality is likely to be expressed in the face rather than through voice. In our data, disgust was confused with happiness and surprise (see Table 2 by rows for confusion results), a confusion that is probably linked with the fact that these three expressions were the ones with higher f_0 variation (see Table 1). f_0 variation was relatively low for sad, neutral, and fearful expressions (around 40 Hz), but relatively high for happiness and surprise (80 Hz and 76 Hz, respectively), as well as for disgust and anger (67 and 65 Hz, respectively). These two expressions, however, had very different durations: anger was the shortest, but disgust was the longest of all, and this may explain why they were not confused with each other.

Among the remaining expressions, the pattern of confusions is also associated with similarities in the acoustic

profiles. Happiness was confounded with surprise, and both types of utterances had high mean f_0 (>330 Hz) and f_0 variation (>70 Hz). Sadness was confused with neutrality, both with low mean f_0 and f_0 variation. Fear was confounded with surprise, which were very similar in duration and mean f_0 . Again, these results are consistent with previous reports: The mislabeling of happiness and fear with surprise has also been found by Adolphs et al. (2002) and that of sadness with neutrality has been found by Pell (2002).

With respect to the comparison between sentences and pseudosentences, the ANOVA revealed a significant main effect of stimulus type [$F(1,354) = 9.05, p < .005, \eta_p^2 = .02$], which was qualified by an interaction with emotion category [$F(6,354) = 4.15, p < .001, \eta_p^2 = .07$]. Emotions were similarly well recognized in sentences and in pseudosentences ($ps > .4$), with the sole exception of happiness, which was better recognized in sentences. Overall, this result is consistent with our expectation that both types of stimuli would elicit similar accuracy. The difference found for happiness is unexpected, given that care was taken to ensure that sentences and pseudosentences were as close as possible in terms of gross acoustical profile.

Summing up, mean identification accuracy was quite high for both sentences (75% correct) and pseudosentences (71%). This level of accuracy is similar to what has been found in studies using the English language. For example, around 81% was reported for 5 emotions in neutral sentences (Adolphs et al., 2002), around 78% for 6 emotions in pseudosentences (Pell, 2002), and around 55% for 14 emotions in sentences and meaningless speech (Banse & Scherer, 1996). Thus, pseudosentences and sentences with emotionally neutral content produced with different intended emotional expressions are an effective means of communicating emotions and, consequently, are good stimuli for use in the study of emotional prosody.

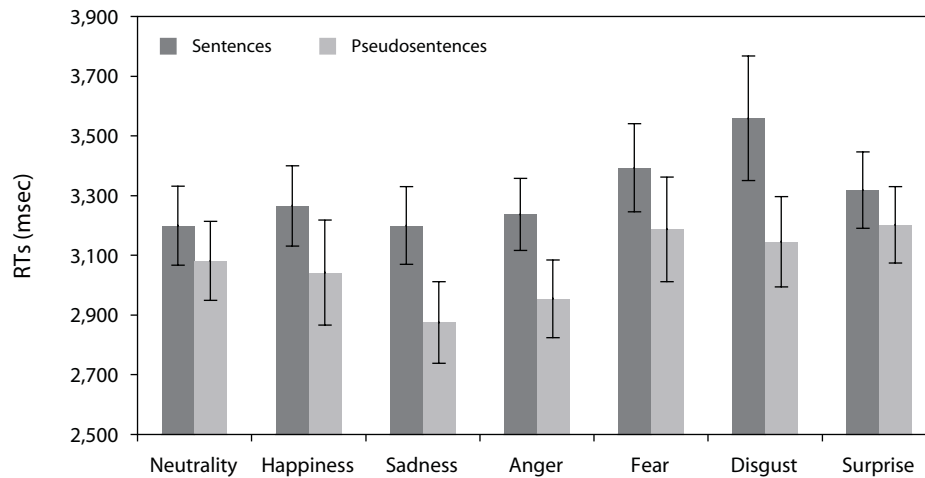


Figure 1. Mean RTs to correctly identify each expression in sentences and pseudosentences. Standard error bars are shown.

RTs

The average time taken to correctly identify the target expressions was computed for each stimulus, and RTs more than 3 *SDs* from the grand mean were excluded from further RT analyses. An ANOVA was computed with expression category and stimulus type as between-subjects factors. As can be observed in Figure 1, expressions were recognized more rapidly in pseudosentences (3,070 msec) than in sentences [3,310 msec; main effect of stimulus type, $F(1,354) = 9.3, p < .005, \eta_p^2 = .03$], irrespective of expression category ($F < 1$ for the interaction). This advantage of pseudosentences was expected on the basis of their lack of semantic content (see the introduction). The longer response latency for sentences might be the behavioral marker of the time needed to integrate prosody with semantics. Although neutral and irrelevant to the task, the semantic information contained in the sentences was probably processed by left-lateralized neurocognitive systems that are involved in semantic processing (Friederici & Alter, 2004; Mitchell & Ross, 2008; Ross & Monnot, 2008), and its availability was likely to require the

integration of linguistic content with affective prosody, as proposed by Schirmer and Kotz (2006). Since lexico-semantic processing was dramatically decreased and no integration was needed with prosody, the responses could be produced faster for prosody-only utterances than for utterances with semantic content.

Even though some categories were responded to faster than others (anger and sadness faster; surprise, disgust, and fear slower), the effect of category was not significant ($F < 1$). Interestingly, in a study of emotion recognition in facial expressions, Palermo and Coltheart (2004) observed that, among the six major emotions, RTs were slowest for fearful expressions and fastest for happy expressions. However, RTs for faces were much faster than were ours for emotional prosody.

Intensity Rates

As we did for RTs, we computed average intensity rates for each stimulus that was correctly identified, and we calculated an ANOVA with expression category and stimulus type as between-subjects factors. Sentences and

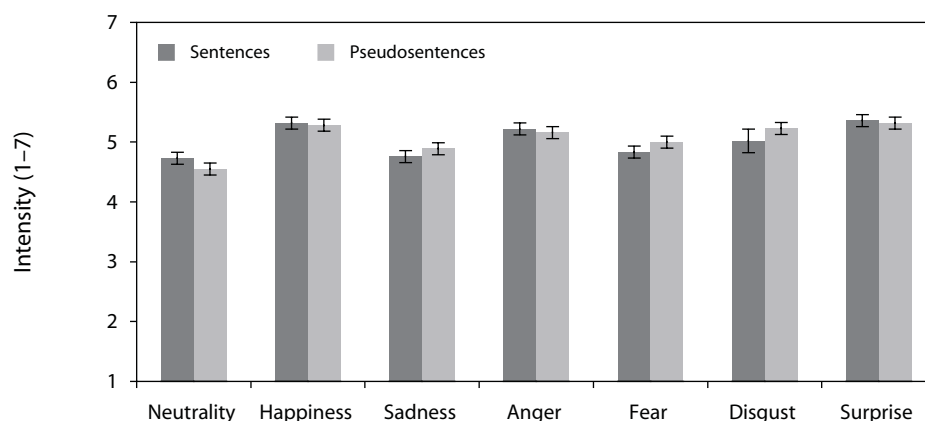


Figure 2. Mean intensity rates by emotional expression for sentences and pseudosentences. Standard error bars are shown.

pseudosentences were rated as being equally good exemplars of the different emotion categories, with an average magnitude of 5 on a 7-point scale ($F < 1$ for main effect of stimulus type; see Figure 2). As also shown in Figure 2, the intensity ratings varied across categories [main effect of category, $F(6,354) = 11.68, p < .0001, \eta_p^2 = .17; F < 1$ for the interaction]. Happy (5.3), angry (5.2), disgusted (5.1), and surprised (5.3) expressions were rated similarly ($ps > .2$) and were more representative than were neutral (4.6), sad (4.8), and fearful expressions (4.9; $ps < .05$), which did not differ from each other ($ps > .2$). The similarity of intensity ratings for sentences and pseudosentences is a confirmation that both types of stimuli were effective in their ability to communicate emotions through prosody.

CONCLUSION

In the present study, we devised and validated a set of sentences and pseudosentences in European Portuguese for research on emotional prosody. Two speakers produced 16 sentences and 16 pseudosentences derived from them in seven different intonations—angry, disgusted, fearful, happy, sad, surprised, and neutral—and the validation procedure included acoustic analysis and three perceptual measures: recognition accuracy, RT, and intensity. The resulting database, which we have made available for the research community, contains 190 sentences with emotionally neutral semantic content and 178 pseudosentences. The acoustic profiles found for the different emotions are in good agreement with the previous literature (Banse & Scherer, 1996; Juslin & Laukka, 2003) and are similar for sentences and pseudosentences. These stimuli elicited high recognition rates (well above chance level), were recognized rapidly, and were judged to be portraying the intended emotions clearly.

It was also shown that the sentences and pseudosentences were similarly effective in expressing emotions through prosody, in that both were recognized with similar accuracy and judged to be of similar intensity. This suggests that the neurocognitive mechanisms underlying emotional prosody are equally effective in processing utterances with neutral lexico-semantic content as utterances without it. The presence of lexico-semantic information might even have had a processing cost, as attested by the longer latencies in responding to the semantically neutral sentences as compared with the pseudosentences. Altogether, these findings are consistent with the view that prosody processing is segregated from linguistic semantic processing, probably because both depend on partially dissociated neural mechanisms (Adolphs et al., 2002; Friederici & Alter, 2004; Ross & Monnot, 2008; Ross et al., 1997).

The database presented here may be useful for behavioral, neuropsychological, neuroimaging, and electrophysiological research on prosody, language, and communication. Specifically, it may be useful for cross-language studies aiming to determine how our ability to recognize emotions in speech is shaped by universal or language-specific factors (e.g., Pell et al., 2009) or for inclusion

in neuropsychological tests for the clinical assessment of brain-damaged and neuropsychiatric patients.

AUTHOR NOTE

This work was supported by Grants PTDC/PSI/66641/2006 and SFRH/BD/39306/2007 from the Portuguese Foundation for Science and Technology and Grant No29/08 from the Bial Foundation. We thank Marc D. Pell and one anonymous reviewer for valuable comments on an earlier version of the manuscript. A special thank you goes to our two speakers, Maria Luís Almeida and Mariana Schöner, who generously agreed to produce the stimuli and granted permission to use them in this database. Address correspondence to S. L. Castro, Faculty of Psychology and Education, University of Porto, Rua Doutor Manuel Pereira da Silva, P 4200–392 Porto, Portugal (e-mail: slcastro@fpce.up.pt).

REFERENCES

- ABBOUD, H., SCHULTZ, W., & ZEITLIN, V. (2006). SuperLab (Version 4.0) [Computer software]. San Pedro, CA: Cedrus Corporation.
- ADOLPHS, R., DAMASIO, H., & TRANEL, D. (2002). Neural systems for recognition of emotional prosody: A 3-D lesion study. *Emotion, 2*, 23-51. doi:10.1037//1528-3542.2.1.23
- ADOLPHS, R., TRANEL, D., & DAMASIO, H. (2001). Emotion recognition from faces and prosody following temporal lobectomy. *Neuropsychology, 15*, 396-404. doi:10.1037//0894-4105.15.3.396
- BACH, D. R., GRANDJEAN, D., SANDER, D., HERDENER, M., STRIK, W. K., & SEIFRITZ, E. (2008). The effect of appraisal level on processing of emotional prosody in meaningless speech. *NeuroImage, 42*, 919-927. doi:10.1016/j.neuroimage.2008.05.034
- BANSE, R., & SCHERER, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality & Social Psychology, 70*, 614-636.
- BELIN, P., FECTEAU, S., & BÉDARD, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences, 8*, 129-135. doi:10.1016/j.tics.2004.01.008
- BELIN, P., FILLION-BILODEAU, S., & GOSSELIN, F. (2008). The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods, 40*, 531-539. doi:10.3758/BRM.40.2.531
- DARA, C., MONETTA, L., & PELL, M. D. (2008). Vocal emotion processing in Parkinson's disease: Reduced sensitivity to negative emotions. *Brain Research, 1188*, 100-111. doi:10.1016/j.brainres.2007.10.034
- EKMAN, P., & FRIESEN, W. V. (1978). *Facial action coding system: Investigator's guide*. Palo Alto, CA: Consulting Psychologists Press.
- FRIEDERICI, A. D., & ALTER, K. (2004). Lateralization of auditory language functions: A dynamic dual pathway model. *Brain & Language, 89*, 267-276. doi:10.1016/S0093-934X(03)00351-1
- GRANDJEAN, D., SANDER, D., POURTOIS, G., SCHWARTZ, S., SEGHER, M. L., SCHERER, K. R., & VUILLEUMIER, P. (2005). The voices of wrath: Brain responses to angry prosody in meaningless speech. *Nature Neuroscience, 8*, 145-146.
- JUSLIN, P. N., & LAUKKA, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129*, 770-814. doi:10.1037/0033-2909.129.5.770
- KOTZ, S. A., MEYER, M., ALTER, K., BESSON, M., VON CRAMON, D. Y., & FRIEDERICI, A. D. (2003). On the lateralization of emotional prosody: An event-related functional MR investigation. *Brain & Language, 86*, 366-376. doi:10.1016/S0093-934X(02)00532-1
- LANG, P. J., BRADLEY, M. M., & CUTHBERT, B. N. (2005). *International Affective Picture System (IAPS): Affective ratings of pictures and instruction manual* (Tech. Rep. No. A-6). Gainesville: University of Florida.
- LAUKKA, P. (2005). Categorical perception of vocal emotion expressions. *Emotion, 5*, 277-295. doi:10.1037/1528-3542.5.3.277
- LAUKKA, P., & JUSLIN, P. N. (2007). Similar patterns of age-related differences in emotion recognition from speech and music. *Motivation & Emotion, 31*, 182-191. doi:10.1007/s11031-007-9063-z
- 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. doi:10.1016/S0028-3932(03)00017-4
- 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
- PALERMO, R., & COLTHEART, M. (2004). Photographs of facial expression: Accuracy, response times, and ratings of intensity. *Behavior Research Methods, Instruments, & Computers*, **36**, 634-638.
- 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 & Language*, **105**, 59-69. doi:10.1016/j.bandl.2007.11.005
- PAULMANN, S., PELL, M. D., & KOTZ, S. A. (2008a). 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
- PAULMANN, S., PELL, M. D., & KOTZ, S. A. (2008b). How aging affects the recognition of emotional speech. *Brain & Language*, **104**, 262-269. doi:10.1016/j.bandl.2007.03.002
- PELL, M. D. (2002). Evaluation of nonverbal emotion in face and voice: Some preliminary findings on a new battery of tests. *Brain & Cognition*, **48**, 499-504. doi:10.1006/brcg.2001.1406
- PELL, M. D., & LEONARD, C. L. (2003). Processing emotional tone from speech in Parkinson's disease: A role for the basal ganglia. *Cognitive, Affective, & Behavioral Neuroscience*, **3**, 275-288. doi:10.3758/CABN.3.4.275
- PELL, M. D., MONETTA, L., PAULMANN, S., & KOTZ, S. A. (2009). Recognizing emotions in a foreign language. *Journal of Nonverbal Behavior*, **33**, 107-120. doi:10.1007/s10919-008-0065-7.
- PELL, M. D., & SKORUP, V. (2008). Implicit processing of emotional prosody in a foreign versus native language. *Speech Communication*, **50**, 519-530. doi:10.1016/j.specom.2008.03.006
- ROSS, E. D., & MONNOT, M. (2008). Neurology of affective prosody and its functional-anatomic organization in right hemisphere. *Brain & Language*, **104**, 51-74. doi:10.1016/j.bandl.2007.04.007
- ROSS, E. D., THOMPSON, R. D., & YENKOSKY, J. (1997). Lateralization of affective prosody in brain and the callosal integration of hemispheric language functions. *Brain & Language*, **56**, 27-54. doi:10.1006/brln.1997.1731
- SCHERER, K. R., BANSE, R., WALLBOTT, H. G., & GOLDBECK, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation & Emotion*, **15**, 123-148. doi:10.1007/BF00995674
- SCHERER, K. R., JOHNSTONE, T., & KLASMEYER, G. (2003). Vocal expression of emotion. In R. J. Davison, K. Scherer, & H. H. Goldsmith (Eds.), *Handbook of the affective sciences* (pp. 433-456). New York: Oxford University Press.
- SCHIRMER, A., & KOTZ, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, **10**, 24-30. doi:10.1016/j.tics.2005.11.009
- THOMPSON, W. F., & BALKWILL, L.-L. (2006). Decoding speech prosody in five languages. *Semiotica*, **158**, 407-424. doi:10.1515/SEM.2006.017
- VEILLARD, S., PERETZ, I., GOSSELIN, N., KHALFA, S., GAGNON, L., & BOUCHARD, B. (2008). Happy, sad, scary and peaceful musical excerpts for research on emotions. *Cognition & Emotion*, **22**, 720-752. doi:10.1080/02699930701503567
- WILDGRUBER, D., RIECKER, A., HERTRICH, I., ERB, M., GRODD, W., ETHOFER, T., & ACKERMANN, H. (2005). Identification of emotional intonation evaluated by fMRI. *NeuroImage*, **24**, 1233-1241. doi:10.1016/j.neuroimage.2004.10.034

SUPPLEMENTAL MATERIALS

The sentence and pseudosentence stimuli discussed in this article, as well as an Excel spreadsheet listing their characteristics, may be downloaded from <http://brm.psychonomic-journals.org/content/supplemental>.

APPENDIX

Sentences and Pseudosentences Included in the Database

Sentence (Translation)	Pseudosentence
Esta mesa é de madeira. (<i>This table is of wood.</i>)	Esta dêpa é de faneira.
O rádio está ligado. (<i>The radio is switched on.</i>)	O dárrio está guilado.
Aquele livro é de história. (<i>That book is of history.</i>)	Aquele jicro é de hisbólia.
A Terra é um planeta. (<i>The Earth is a planet.</i>)	A Pirra é um flanelo.
O cão trouxe a bola. (<i>The dog brought the ball.</i>)	O lão droube a nóma.
Ele chega amanhã. (<i>He arrives tomorrow.</i>)	Ele chena aguinhã.
Esta roupa é colorida. (<i>These clothes are colorful.</i>)	Esta souda é lacoripa.
Os jardins têm flores. (<i>The gardens have flowers.</i>)	Os bartins têm pléres.
As pessoas vão a concertos. (<i>People go to concerts.</i>)	As semoas vão a cambêrtos.
Há árvores na floresta. (<i>There are trees in the forest.</i>)	Há árjuques na plurisca.
Os tigres são selvagens. (<i>Tigers are wild animals.</i>)	Os lagres são siltávens.
O quadro está na parede. (<i>The painting is on the wall.</i>)	O juadre está na pafêne.
Alguém fechou as janelas. (<i>Someone closed the windows.</i>)	Alguém belhou as jalétas.
Os jovens ouvem música. (<i>Young people listen to music.</i>)	Os dófens mavam tézica.
O futebol é um desporto. (<i>Football is a sport.</i>)	O dutebel é um nesforpo.
Ela viajou de comboio. (<i>She traveled by train.</i>)	Ela jiavou de lantóio.

(Manuscript received March 3, 2009;
revision accepted for publication June 6, 2009.)