

EmoLexPros: A database of prosodically incongruent emotional words in European Portuguese

Rui Gomes*  / São Luís Castro*  / Susana Silva* 

* Faculdade de Psicologia e Ciências da Educação, Universidade do Porto, Porto, Portugal

Abstract: In this study, we validated a set of audio stimuli addressing emotional lexico-prosodic congruence. Positive, neutral and negative words classified as such in European Portuguese lexical databases were recorded with matching and mismatching emotional prosodic patterns (positive, neutral and negative). The produced set was submitted to behavioral validation, where each utterance was rated for lexico-prosodic match. Word pairs showing maximal differences across match and mismatch versions were selected ($n=55$) and subjected to prosody-related acoustic validation. The three types of prosodic patterns, positive, negative and neutral, differed in F0, amplitude, duration and timbral parameters. Finally, items were catalogued for their psycholinguistic properties. The present validated set may be used for emotional prosody research in European Portuguese. Since it is compatible with passive listening tasks, it is particularly relevant to research and clinical practice engaging populations who are limited in their ability to provide explicit responses.

Keywords: Emotional prosody, Incongruence detection, Speech processing, European Portuguese.

Introduction

In everyday social interactions, humans rely on multiple sources of information to successfully communicate with one another. Visual, auditory, and contextual information provide fundamental cues to unveil the meaning, intentions and emotional state of others. Emotional prosody is crucial to emotional communication in speech, and it refers to the modulations of vocal acoustic parameters to convey emotions (Banse & Scherer, 1996). Different combinations of pitch (fundamental frequency), amplitude, speech rate and timbre index different emotional states. For example, a fast-paced utterance with ascending pitch contour and bright timbre may convey a joyful, positive-valence, emotional state; a slow, low-pitched utterance with an *opaque* timbre is associated with boredom or sadness (negative valence); and an utterance with a flat pitch contour, short duration and rapid speech rate may convey emotional *neutrality* (neutral valence).

Prosody is not the sole conveyer of emotional content, as emotions can also be communicated through the semantics of spoken words. In real world communication, the interplay between prosodic and lexical channels is ever-present, each channel alone having the capacity to modify the emotional meaning of an utterance (Kotz & Paulmann, 2007; Lin et al., 2020). The study of emotional prosody and its interplay with lexical content has received considerable attention in the past decades (see, for instance, Bağ, 2016; Paulmann, 2016), and neurocognitive models of vocal emotion comprehension have been proposed (Brück et al., 2011; Pell & Kotz, 2011; Schirmer & Kotz, 2006; Wright et al., 2018). These models converge on the idea that the input follows three

Correspondence concerning this article should be addressed to: Susana Silva, Faculdade de Psicologia e Ciências da Educação, Universidade do Porto, Rua Alfredo Allen, 4200-135 Porto, Portugal. E-mail: susanamsilva@fpce.up.pt

major stages or subprocesses, ranging from (1) low-level acoustic analysis to (2) integrated emotional representations that become available to (3) cognitive judgement. Cognitive judgement represents the highest processing level, where the listener's mind-brain uses both lexical and prosodic meaning to access the speaker's emotions.

Prosodic and lexical information channels are processed by the mind-brain independently at the first stages of the processing chain (Schirmer & Kotz, 2006; see also Hickok & Poeppel, 2007). Therefore, *lexico-prosodic integration* is a major component of the highest-level subprocess – cognitive judgement. At this stage, listeners become able to perform two tasks: (i) judging lexico-prosodic congruence, and (ii) labeling emotions explicitly (see Brück et al., 2011; Schirmer & Kotz, 2006). The former task does not require participants to label the emotions portrayed in an attentive way, and thus may be addressed with implicit tasks. Tasks of lexico-prosodic congruence judgement typically use conflicting cues of information, such as the word “sad” spoken with a positive intonation (Bostanov & Kotchoubey, 2004; Mitchell et al., 2003; Schirmer & Kotz, 2003). Different experimental settings have been used, such as passive hearing (e.g., Paulmann & Kotz, 2008), Stroop tasks (e.g., Schirmer & Kotz, 2003) or other interference-based tasks, e.g., Lin et al., 2020; Paulmann & Pell, 2011). When studying lexico-prosodic integration in emotional prosody research, it is important to have language-specific stimulus sets addressing conflicting prosodic and lexical cues.

Experimental manipulations of lexico-prosodic congruence can be achieved based on, at least, two theoretical approaches. One is a categorical approach, where basic emotion categories such as joy, sadness, anger, etc., are considered. In this case, incongruence arises from combining a lexical item that conveys a given emotion with a prosodic pattern that characterizes a different emotion, such as the word “joy” spoken with surprised prosody, or the word “sad” with angry prosody. The other approach is a dimensional one, focusing on bipolar dimensions that characterize emotional contents as positive/pleasant vs. negative/unpleasant (valence), or associated with high vs. low physiological activation (arousal). Here, incongruence occurs when the lexical item is combined with a prosodic pattern pertaining to the opposite side of the bipolar axis (e.g., a positive word with a negative prosody). A major advance in the debate on whether emotions are represented as categories or dimensions (e.g., Cowen et al., 2019; for a review, see Hamann, 2012) has recently been made by Giordano et al. (2021), who presented MEG and fMRI evidence for a conciliatory view: vocal emotional processing would first be represented in a categorical manner and then become progressively incorporated into a valence classification (negative, neutral or positive). Given that lexico-prosodic incongruence judgements are expected to occur at a later stage in the emotion recognition process (Schirmer & Kotz, 2006), it is reasonable to expect that the mind-brain will respond to incongruences based on the representations that are available at that stage – i.e., dimensional ones.

Another important remark must be made regarding the status of lexical and prosodic *neutrality*. At the lexical level, neutral words have properties that differentiate them from negative and positive ones: they are processed more slowly (Kousta et al., 2009) and tend to show specific psycholinguistic properties such as higher scores of concreteness and lower scores of imageability (e.g., Kousta et al., 2011; Soares et al., 2016). Thus, along with positivity and negativity, *lexical neutrality* seems to be well-represented in the mind-brain, with multiple studies acknowledging the relevance of this triadic contrast at the lexico-semantic level (e.g., Kousta et al., 2011; Pinheiro et al., 2016). At the prosodic level, the dyadic approach to valence (positive-negative) has been dominant in the literature, but empirical findings have shown that emotion-specific prosodic patterns are discriminated from neutral ones in the context of congruent lexico-prosodic combinations (e.g., Paulmann & Kotz, 2008). Such evidence raises the possibility of three distinct prosodic valence segments – positive, negative and neutral. Our approach intended to further explore the validity of this triadic approach.

Therefore, in the present study, we opted for a dimensional approach – specifically, a triadic (negative, neutral, positive), valence-based approach to emotional prosody.

Rationale behind stimulus creation and validation

Validated emotional prosody stimulus materials available for research in European Portuguese (EP) are not suitable to address lexico-prosodic integration. To our knowledge, available materials are limited to a set of emotion-specific prosodic patterns in sentences and pseudo-sentences (Lima & Castro, 2010), and a set of valence-specific (liking vs. disliking) prosodic patterns used on neutral words (Filipe et al., 2015). These materials are suitable for studying the explicit labelling of emotions (e.g., Castro & Lima, 2014; Correia et al., 2019), but, given the absence of conflicting prosodic vs. lexical cues in these stimuli, lexico-prosodic integration is left out. Due to lack of materials to study emotional lexico-prosodic integration in European Portuguese, we aimed to bridge that gap in the present work. To that end, we generated and validated an auditory set of emotional words spoken with congruent vs. incongruent prosodic pattern (e.g., the word *alegre* was read both with a congruent/happy and an incongruent/sad prosodic pattern).

To manipulate lexico-prosodic congruence based on valence, we used three lexical-valence levels: positive, negative, and neutral words. For each of these classes, we created prosodic matches (congruence) and prosodic mismatches (incongruence): prosodic matches for positive, negative, and neutral words corresponded to positive, negative and neutral prosodic patterns, respectively; prosodic mismatches for positive and negative words were negative and positive prosodic patterns, respectively; prosodic mismatches for neutral words included two types – positive and negative prosodic patterns (Figure 1). When defining the neutral match as a neutral word with neutral prosody, we assumed that – paralleling lexical neutrality – prosodic neutrality is psychologically real and may be seen as the central segment of a valence continuum.

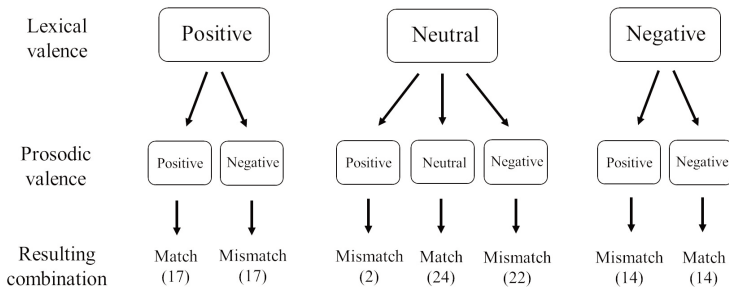


Figure 1. Schematic representation of the validated stimuli. A total of 110 audio files (55 pairs) reflected optimal lexico-prosodic contrast

The creation and validation of the stimuli went through the following stages. First, we sought for a published lexical stimulus set specifically organized into positive, negative and neutral valence categories where valence categories had shown experimental effects (Pineiro et al., 2016), from which we extracted 90 words. We checked the emotional valence of the lexical items in the Portuguese databases ANEW (Soares et al., 2011) and EMOTAIX PT (Costa, 2012), which originated the exclusion of some items. The remaining subset was read aloud by a trained speaker with congruent and incongruent prosodies. The utterances were then submitted to ratings of lexico-prosodic congruence by a group of participants, and a second selection was made.

The prosodic patterns of the resulting selection were validated acoustically for the intended positive, negative, and neutral emotional prosodies. Following classic findings on the acoustic features that seem to be most relevant to differentiate emotions (Banse & Scherer, 1996; Goudbeek & Scherer, 2010; Juslin & Laukka, 2003; Lima & Castro, 2010; Pell & Skorup, 2008), we focused on duration (in ms), mean fundamental frequency (F0, measured in Hz, representing the vibration frequency of the vocal folds), F0 variation (measured as the standard deviation of F0), amplitude

variation (standard deviation) and amplitude range. In addition to basic acoustic descriptors, we compared valence levels according to timbral parameters. Timbral parameters refer to spectral dimensions associated with tone color (e.g., Dowling & Harwood, 1986) or perceptual quality of sounds (Handel, 1995). The acoustic substrate of timbre comprises the relation between the partials of a complex sound, the distribution of energy across the spectrum, as well as the spectral envelope. Timbral measures have not been considered as standard acoustic descriptors of emotionality in the literature (see, e.g., Juslin & Laukka, 2003; Lima & Castro, 2010; Pell & Skrup, 2008; Preti et al., 2015). Recently, Tursunov and colleagues (2019) proposed a set of timbral acoustic descriptors which more accurately distinguished the valence of emotions. These descriptors include: spectral centroid, representing the spectral center of gravity; spectral spread, representing the spread around the mean value; spectral crest, or the crest factor of the spectrum, translating the ratio of peak-average power; inharmonicity, a measure of how partial portions of the signal differ from purely harmonic frequencies; noisiness, the degree to which the signal is harmonic or not; tristimulus, a measure conceived to define the first harmonics within the spectrum, based on three energy ratios.

Once acoustic validation was finished, lexical items were catalogued for their psycholinguistic properties according to previously validated norms.

The outcome was a validated set of 110 words (55 congruent and 55 incongruent) that can be used for emotion and language research in European Portuguese. These steps are detailed below.

Methods

Selection of lexical items

As Pinheiro et al.'s (2016) lexical database contained positive, negative and neutral valence words – the same triadic view we intended to follow –, we used it to select 90 words, 30 of each valence level, that we then translated from English (the original language of the dataset) into Portuguese. Contrary to English, Portuguese has grammatical gender inflections (masculine, feminine and invariable wordforms). While about 40% of the translated words remained gender-invariable (e.g., “excellent” corresponds to *excelente*, invariable in Portuguese), the remaining did not; for example, “satisfied” corresponds to the gender-marked wordforms *satisfeita* (feminine) and *satisfeito* (masculine) in Portuguese. In these cases, to balance grammatical gender in our stimuli, half was translated into the masculine form (e.g., “fantastic” was translated to *fantástico*) and the other half to the feminine form (e.g., “furious” was translated to *furiosa*).

To ensure that the translated words had the same valence as the original ones, we confirmed their valence in the two largest emotional-word databases that have been validated for European Portuguese: ANEW (Soares et al., 2011; original version, Bradley & Lang, 1999) and EMOTAIX PT (Costa, 2012; original version, Piolat & Bannour, 2009). ANEW presents mean valence scores and standard deviations for 1034 emotional words. Scores range from 1 to 9, with 5 representing a valence cutoff under which words would be unpleasant (negative valence), and above which they would be pleasant (positive valence). In EMOTAIX PT, 3983 words are classified according to three hierarchical valence category systems (basic, supra and super categories). Supra categories are the most relevant to us, since they correspond to positive valence (including benevolence, well-being, and composure) vs. negative valence (malevolence, ill-being, and anxiety). Super ($n=8$) and basic categories ($n=50$) are further subdivisions of supra categories, with more specificity (for instance, the word *horror* belongs to the super category of anxiety, the supra category of fear, and the basic category of dread). In addition, EMOTAIX includes non-hierarchical categories standing outside the valence poles established by supra categories. These valence-unrelated categories include surprise, unspecified emotions, and impassiveness.

The inclusion of words in our set was based on the following criteria: Positive and negative words were kept if they were present in at least one of the two databases with the expected valence. Specifically, they should be part of EMOTAIX (as positive/negative words when positive/negative valence was intended) or display mean valence scores in ANEW lower or higher than central values ($mean \pm SD$, 5 ± 1.5) for negative and positive words, respectively. Given that EMOTAIX and ANEW do not include a neutral label nor a numerical definition of neutral valence, we kept as neutral words those which were absent from both datasets, or which were included in EMOTAIX as *unspecified emotion* and absent from ANEW, or present in ANEW with a valence score within $\pm 1 SD$ of the mean and absent from EMOTAIX. For example, the word *casual* (“casual”) was absent from EMOTAIX and present in ANEW with a valence score of 5.18, which is within the range of the mean $\pm 1 SD$, and therefore was taken as a neutral word. This would ensure that the words selected as neutral would fall around a *central zone* of a dimensional valence continuum. Finally, if a translated word was not present in the Portuguese databases, but a word with the same root morpheme was, we kept it in our dataset; for example, we kept the adjective *horroroso* because the noun *horror* was included in ANEW. After the exclusion of 6 translated words (3 positive, and 3 negative), the resulting set containing 84 words was recorded by the first author (RG) (for details, see the dedicated spreadsheet in the supplementary material).

Recording

Audio files were recorded by a 24-year-old male speaker (RG), with instructions being provided by the remaining authors. The selected set of lexical items was recorded twice: first with valence-congruent prosody ($n=84$, 27 positive words with positive prosody, 27 negative words with negative prosody and 30 neutral words with neutral prosody), and then with valence-incongruent prosody ($n=114$, 27 positive words with negative and with positive prosody, 30 neutral words with negative and with positive prosody). For congruent stimuli, the speaker embodied an emotional state that would justify the use of that word to characterize a person/object/situation, not portraying irony or sarcasm, e.g., the word *triste* (“sad”) with a sad prosody, or the word *alegre* (“happy”) with a happy prosody; for the incongruent ones, he used negative prosody for positive words, positive prosody for negative words, and both positive and negative prosody for neutral words. No rigid suggestions on the specific prosodic pattern to be used were given to the reader. Therefore, there was some variability in this domain. For example, the reader has employed both *joyful*, *surprised* or *excited* as positive prosodic patterns, and *sad*, *disgusted*, or *enraged* as negative prosodic patterns according to his discretion. Recordings were made in a sound-insulated booth at a sampling rate of 48-KHz and 16-bit resolution, using a high-quality cardioid microphone, an audio interface and version 2.3.3 of Audacity® software (<https://www.audacityteam.org/>).

Files were RMS-normalized to +70 dB, and silences removed using Praat (<http://www.fon.hum.uva.nl/praat>). The 198 stimuli were then presented to listeners to evaluate whether the meaning suited the prosodic pattern.

Behavioral validation of lexico-prosodic incongruence

Sixty-eight healthy adults (51 women, mean age=22.81 years, $SD=5.63$) listened to the recorded set of words and were asked to rate their lexico-prosodic congruence using a 5-point Likert-type scale, ranging from 1 (*Totally inadequate to meaning*) to 5 (*Totally adequate to meaning*). Data collection was made in an acoustically shielded room using Presentation® 21.1 (Neurobehavioral Systems, <https://www.neurobs.com/>). Participants listened to stimuli with high-quality loudspeakers, and responses were recorded using a 5-button response box. All participants engaged in five training trials before the experiment. Each trial was composed of a fixation cross (200 ms), the audio file,

and then a question displayed on the computer screen (“Does the intonation match the meaning?”). The order of trials was pseudorandomized, assuring that consecutive occurrence of stimuli of the same experimental type (match & mismatch) did not exceed three, and two versions (A and B) were created by switching the order of the first and second halves. Half the participants listened to version A, and the other half to version B.

Acoustic validation of prosodic valence

To validate the valence-related prosodic labels of our stimuli (positive, negative and neutral patterns), we started by grouping the 110 items (55 pairs) according to prosodic valence, disregarding congruence. This resulted in 33 positive (17 lexically congruent and 16 lexically incongruent), 24 neutral (all of them lexically congruent), and 53 negative (14 congruent and 39 incongruent) utterances (Figure 1). We then extracted basic acoustic features for each utterance using Praat (<http://www.fon.hum.uva.nl/praat/>) and computed average values per prosodic class.

Concerning timbral measures, the computation of minimum, maximum and median values for spectral centroid, spectral spread, spectral crest, inharmonicity, noisiness, and tristimulus was made with the Timbre Toolbox (Peeters et al., 2011) for MATLAB (www.mathworks.com).

Psycholinguistic properties

Finally, we characterized the psycholinguistic properties of the validated lexical items ($n=55$) using Minho Word Pool (Soares et al., 2016). This database includes three subjective measures for 3800 words in European Portuguese: imageability, capturing how easy it is for a word to evoke mental images, concreteness, measuring the extent to which a word reflects experienced situations, places or objects, and subjective frequency, representing a self-estimate of the relative frequency of word usage by an individual. Together with objective measures such as objective frequency (a measure of the actual frequency of word occurrence based on large lexical corpora) and word length, these are important measures to consider in psycholinguistic experiments, given that they might explain on its own a considerable portion of variance (Soares et al., 2016).

The search criteria in Minho Word Pool database were similar to what was described in the methods section, that is, whenever the same exact word was not found, we looked for the opposite grammatical gender, or words with the same root morpheme.

Statistical analysis and results

Behavioral validation

For each item, participant-averaged ratings were computed for congruent stimulus (e.g., one positive word with positive prosody) and its incongruent counterpart (the same word with negative prosody) for lexically positive, negative and neutral words. Differences greater than 2 points (e.g., 5 vs. 2.9, 4 vs. 1.9, 3 vs. 0.9) likely indicated relevant changes in the congruency of each word, and so we used this difference as a cutoff to distinguish between high-contrasting and low-contrasting pairs. As lexically neutral words had two incongruent counterparts (spoken with positive and negative prosody), we chose the one with the largest congruent-incongruent difference score. Using the cutoffs above, we arrived at 55 high-contrasting pairs (17 based on positive words, 24 based on neutral and 14 on negative, see Figure 1, Table 1 and supplementary materials) and 59 low-contrasting pairs (10 based on positive words, 36 on neutral and 13 on negative).

Table 1

List of validated words

Positive		Neutral		Negative	
Incongruence score	Word	Incongruence score	Word	Incongruence score	Word
3.25	Confiante	3.21	Azul	2.79	Egoísta
3.23	Alegre	3.08	Usual	2.71	Errada
3.08	Brincalhona	3.02	Regular	2.62	Inútil
2.85	Adorável	2.95	Direta	2.61	Furiosa
2.76	Esplêndida	2.89	Anual	2.59	Sangrento
2.71	Fantástico	2.87	Central	2.52	Horroroso
2.69	Formidável	2.80	Ampla	2.47	Triste
2.65	Elegante	2.77	Roxo	2.43	Terrível
2.65	Livre	2.74	Próximo	2.38	Cruel
2.61	Excelente	2.60	Diário	2.35	Rude
2.60	Encantador	2.52	Ligeira	2.33	Morto
2.54	Satisfeito	2.47	Alta	2.23	Desagradável
2.50	Gentil	2.44	Privada	2.21	Enfurecida
2.46	Amável	2.44	Constante	2.19	Falhada
2.39	Amigável	2.37	Compacta		
2.33	Viva	2.29	Casual		
2.21	Honesta	2.29	Leigo		
		2.25	Negrilo		
		2.21	Perto		
		2.20	Relacionada		
		2.17	Calado		
		2.16	Aberto		
		2.13	Básico		
		2.13	Média		

Note. Incongruence score – the difference between congruent and incongruent averaged ratings per item. Higher values represent stronger lexico-prosodic incongruence. For the English words, please consult the supplementary materials.

To check the cross-subject consistency in rates for high-contrasting pairs, paired samples *t*-tests comparing congruent and incongruent versions were run on the ratings for high- and low-contrasting pairs. For high-contrasting pairs, ratings for lexically congruent prosody items ($M=4.34$, $SD=0.39$) were significantly different from lexically incongruent prosody items [$M=1.80$, $SD=0.41$, $t(67)=36.48$, $p<.001$, Cohen's $d=4.42$]. For low-contrasting pairs, results were similar: lexically congruent prosody items ($M=3.76$, $SD=.48$) differed significantly from lexically incongruent ones [$M=2.60$, $SD=.48$, $t(67)=19.51$, $p<.001$, Cohen's $d=2.37$], but the effect size was smaller. These results lend support to the initial selection based on average ratings.

Acoustic validation

For basic acoustic measures (Table 2), we ran independent samples *t*-tests to investigate whether positive, negative and neutral utterances were differentiated, with Bonferroni corrections made to handle multiple comparisons (Table 3). Compared to neutral, positive utterances had significantly longer duration, higher fundamental frequency (F0), greater F0 variation and lower amplitude variation, with amplitude range not showing statistical differences. When compared to negative utterances, positive ones had significantly higher F0 and F0 variation, but duration, amplitude variation and range were similar. Finally, compared to negative, neutral utterances had significantly shorter duration and larger amplitude variation and range, but similar F0 and F0 variation. A summary of the analysis is presented in Table 4. In the present data, amplitude variation and range differed in negative-positive, and positive-neutral valence contrasts. However, no differences emerged in the negative-neutral contrast.

Table 2

Basic acoustic features for the validated set of words, grouped by emotional prosody

Prosody	Duration (ms)		F0 (Hz)		F0 Variation (SD)		Amp SD		Amp R	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Positive	1179	196	159.62	20.44	47.84	15.27	0.11	0.03	0.38	0.17
Neutral	739	164	137.47	16.29	32.14	25.15	0.13	0.02	0.45	0.16
Negative	1259	229	136.43	18.38	28.42	17.35	0.11	0.03	0.94	0.27

Note. $n=110$ (33 positive, 24 neutral, 53 negative prosodies). F0 – Fundamental Frequency; Amp SD – Amplitude variation, or the standard deviation of the Amplitude; Amp R – Amplitude range (Amp R=AmpMax – AmpMin).

Table 3

*Independent samples *t*-tests for the basic acoustic parameters between positive, neutral and negative utterances*

Pair		<i>t</i>	<i>df</i>	<i>p</i> _{bonf}	Cohen's <i>d</i>
Positive vs. Neutral	Duration	8.970	55	<.001	2.406
	F0	4.386	55	<.001	1.177
	F0 Variation	2.925	55	.025	0.785
	Amplitude <i>SD</i>	-2.792	55	.035	-0.749
	Amplitude Range	-1.677	55	.495	-0.450
Positive vs. Negative	Duration	-1.675	84	.49	2.406
	F0	5.451	84	<.001	1.177
	F0 Variation	5.280	84	<.001	0.785
	Amplitude <i>SD</i>	0.621	84	1	-0.749
	Amplitude Range	1.682	84	1	-0.450
Neutral vs. Negative	Duration	-10.043	75	<.001 ^a	-2.471
	F0	0.240	75	1	0.059
	<i>f</i> 0 Variation	0.754	75	1	0.185
	Amplitude <i>SD</i>	3.719	75	<.001	0.915
	Amplitude Range	4.301	75	<.001	1.058

Note. *p* values are Bonferroni-corrected. ^a Levene's test is significant ($p<.05$) suggesting a violation of the equal variance assumption.

Table 4

Acoustic features that showed statistically significant differences in the valence domain

	Positive vs. neutral	Positive vs. negative	Neutral vs. negative
Basic descriptors	Duration	F0	Duration
	F0	F0 variation	Amplitude variation
	F0 variation		Amplitude range
	Amplitude variation		
Timbre descriptors	Spectral centroid (min, max)	Spectral crest (max [*])	Tristim 2 (max)
	Spectral spread (max)		
	Tristim 3 (max)		

Note. Asterisks stand for marginally significant results ($p<.10$).

Concerning timbral measures, we calculated independent samples *t*-tests with Bonferroni corrections for all the statistical measures to compare positive and neutral, positive and negative, as well as neutral and negative utterances (Table 5). Spectral centroid, spectral spread and tristimulus were effective in discriminating between positive and neutral words. Differences between positive and negative words were captured by spectral crest, although with marginal results. Finally, neutral utterances differed from negative ones only in tristimulus values. A table with the extracted timbral measures can be found in supplementary materials (2.4. Acoustic Measures).

In sum, both basic acoustic measures and timbral descriptors provided evidence in favor of prosodic contrasts between positive and neutral, positive and negative, and negative and neutral valence pairs (see Table 4 for a summary of the relevant acoustic features for each pair). Therefore, the intended emotional prosody of the present set of utterances was validated.

Table 5

Timbre-based comparison between pairs of prosodic valence

Parameter		Positive vs. neutral			Positive vs. negative			Neutral vs. negative		
		<i>t</i>	<i>p_{bonf}</i>	Cohen's <i>d</i>	<i>t</i>	<i>p_{bonf}</i>	Cohen's <i>d</i>	<i>t</i>	<i>p_{bonf}</i>	Cohen's <i>d</i>
Spectral centroid	Min	3.52	.021	0.94	1.34	1	0.30	-2.23	.62	-0.56
	Max	3.96	.005	1.06	1.77	1	0.39	-2.39	.46	-0.59
	Med	1.40	1	0.38	1.21	1	0.27	-0.21	1	-0.05
Spectral Crest	Min	-0.13	1	-0.03	-0.81	1	-0.18	-0.62	1	-0.15
	Max	2.66	.24	0.71	2.97	<u>.096</u>	0.66	0.30	1	0.07
	Med	-0.20	1	-0.05	2.17	.79	0.48	2.17	.79	0.53
Spectral Spread	Min	2.51	.36	0.67	0.98	1	0.22	-1.01	1	-0.25
	Max	3.48	.024	0.93	1.79	1	0.40	-2.55	.31	-0.63
	Med	1.87	1	0.50	0.83	1	0.18	-0.97	1	-0.24
InHarm	Min	-2.24	.69	-0.60	-1.04	1	-0.23	0.78	1	0.19
	Max	0.70	1	0.19	-0.69	1	-0.15	-1.25	1	-0.31
	Med	-1.79	1	-0.48	-0.20	1	-0.04	1.83	1	0.45
Noisiness	Min	0.13	1	0.03	-1.39	1	-0.31	-1.56	1	-0.38
	Max	1.38	1	0.37	0.18	1	0.04	-1.44	1	-0.35
	Med	-0.39	1	-0.11	1.20	1	0.27	1.40	1	0.34
Tristim 1	Min	-0.46	1	-0.12	0.003	1	<0.001	0.49	1	0.12
	Max	2.38	.504	0.64	0.43	1	0.10	-2.66	.22	-0.66
	Med	1.58	1	0.42	-0.22	1	-0.05	-1.68	1	-0.41
Tristim 2	Min	-2.19	.792	-0.59	-0.51	1	-0.11	2.74	.19	0.67
	Max	2.88	.144	0.77	-0.83	1	-0.18	-3.89	.005	-0.96
	Med	-0.69	1	-0.19	-0.33	1	-0.07	0.41	1	0.10
Tristim 3	Min	-1.07	1	-0.29	0.21	1	0.05	1.53	1	0.38
	Max	3.28	.048	0.88	1.81	1	0.40	-1.57	1	-0.39
	Med	-2.16	.864	-0.58	-0.60	1	-0.13	1.62	1	0.40

Note. Displayed *p* values are Bonferroni-corrected. Parameters in bold represent statistically significant results ($p < .05$). Underlined parameters refer to marginally significant results ($p < .10$). InHarm – Inharmonicity; Tristim – Tristimulus.

Psycholinguistic properties of validated lexical items

Table 6 displays a summary of psycholinguistic properties for the validated set of words according to semantic valence (see supplementary material's file 4. Psycholinguistic Properties for detailed data).

Table 6

Psycholinguistic properties according to Minho Word Pool

Semantic valence	Objective Freq		Subjective Freq		Concreteness		Imgeability		WL_1		WL_syll	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Positive	22.77	29.5	5.25	0.56	3.34	0.42	3.99	0.55	8.06	2.08	3.47	0.8
Neutral	93.03	102.34	4.84	0.78	3.94	0.82	3.81	0.98	6.04	1.67	2.78	0.74
Negative	13.34	14.68	4.26	0.78	3.47	0.80	4.10	0.69	7.21	2.19	3.21	1.05

Note. 6 missings. Objective Freq – Objective frequency; Subjective Freq – Subjective frequency; WL_1 – Word length (no. letters); WL_syll – Word length (no. syllables).

The main purpose of this level of description was cataloguing our stimulus set.

For exploratory purposes, we ran independent samples *t*-tests comparing positive vs. neutral, positive vs. negative, and neutral vs. negative words (Figure 2).

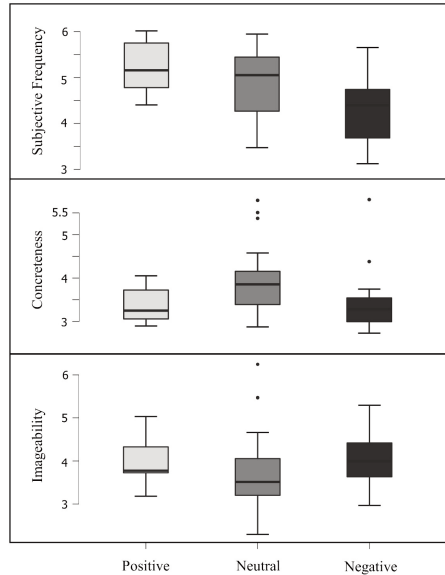


Figure 2. Subjective psycholinguistic measures for the selected positive, neutral and negative words

Positive and negative words did not differ significantly in any measure ($ps > .28$) except in subjective frequency [$t(27) = 3.93$, $p < .001$, Cohen's $d = 1.46$]: positive words were subjectively more frequent than negative ones (Table 6). This may reveal a bias towards attributing positive *color* to our lexicon (Hepach, 2011), even though, in fact, the objective frequency does not differ. Compared to neutral words, positive words had significantly higher objective frequency ($p = .015$), concreteness ($p = .024$), and word length ($p = .002$ for number of letters, $p = .007$ for number of syllables), and marginally significant higher subjective frequency ($p = .098$), but no differences in imageability ($p = .54$). Finally, negative words had lower subjective frequency ($p = .039$), lower objective frequency ($p = .007$), and marginally significant higher word length (number of letters, $p = .079$) than neutral words, but no differences in concreteness ($p = .102$), imageability ($p = .362$), and word length (syllables, $p = .15$).

In sum, apart from subjective frequency, where positive words scored higher than neutral, and neutral higher than negative (Figure 2), we saw no obvious pattern of psycholinguistic contrast across the three valence levels. Unlike Altarriba and Bauer (2004) and Soares et al. (2016), we found no differences in concreteness and imageability between positive and negative words.

Discussion and conclusion

To make up for the absence of materials to study emotional lexico-prosodic integration in European Portuguese, we validated a set of utterances comprising positive, negative and neutral words spoken with congruent or incongruent emotional prosody organized into three valence levels, positive, negative and neutral. The validation procedure included lexical selection,

behavioral validation of congruence, acoustic validation of differences across prosodic patterns, as well as psycholinguistic classification of words. The validated set elicited consistent and large differences in behavioral ratings for congruent vs. incongruent versions. Prosodic patterns for positive, negative and neutral utterances differed in several acoustic parameters, and the characteristics of these patterns were partly consistent with previous research findings (see below).

Notes on acoustic validation results

Though all three prosody-related emotional contrasts (positive-neutral, positive-negative, negative-neutral) showed significant differences in, at least, two out of five basic acoustic parameters (Table 4), parameters were not equally relevant across contrasts. For instance, duration was irrelevant to the positive vs. negative contrast, but relevant to the other two. This is in line with the literature (Banse & Scherer, 1996; Goudbeek & Scherer, 2010; Lausen & Hammerschmidt, 2020; Montenegro & Maravillas, 2015).

On the other hand, our data suggest that F0 parameters are not relevant for contrasting negative and neutral prosodic patterns, which contrasts with Preti et al. (2015). However, it should be noted, that, unlike us, Preti and colleagues (2015) operationalized negative valence as anger only, which can be a source of error on its own.

Among the timbral measures suggested by Tursunov et al. (2019), spectral descriptors seemed particularly relevant. In addition, timbral descriptors were clearly more helpful to distinguish positive from neutral utterances than the other pairs (positive-negative and negative-neutral). The irrelevance of timbral measures to the positive-negative comparison contrasts with the results from of Tursunov and colleagues (2019). One explanation may be that may be that Tursunov and colleagues (2019) included neutral emotions in the negative valence group. It is, thus, possible that the positive-negative contrasts in timbre that they saw reflected, at least partly, the positive-neutral differences we observed in our data. This strengthens the idea that, instead of grouping neutral with negative prosody, it may be useful to view neutral prosody as a valence level of its own.

Limitations and implications for future research

In this study, we adopted triadic views on lexical valence and prosodic valence, meaning that in both cases positive, negative and neutral qualities were considered. While this triadic view of lexical valence seems relatively consensual (e.g., Kousta et al., 2011; Pinheiro et al., 2016), the same does not hold for prosodic valence, where the concept of neutral prosody tends to not be integrated into valence analyses. Although this lack of comparability with extant research may be viewed as limitation of the present study, our triadic view on prosody gained some support from our data. Our results suggest that neutral prosody might indeed have a separate representation in the brain: behavioral results demonstrated that neutral words with neutral prosody show higher congruence scores than neutral words with negative prosody and, though less marked, with positive prosody. This finding leads us to propose that future valence-oriented prosody research would benefit from considering the neutral level. Furthermore, it suggests that studies using neutral lexical content combined with emotional prosody as a way to minimize lexical interference (e.g., Lima & Castro, 2010; Paulmann & Kotz, 2008) may not be entirely free from confounds, as the combination of lexical neutrality with emotional prosody might be a source of incongruence.

Besides behavioral validation, the idea of neutral prosodic valence was supported by our acoustic analysis of prosodic patterns: neutral utterances showed significant differences compared to positive and negative ones, not only in durational, F0, and amplitude-related values (traditional descriptors), but also in timbral measures. The observed differences in traditional descriptors –

which were also present in the comparison between positive and negative prosody – stress the relevance of these measures for signaling emotion (e.g., Goudbeek & Scherer, 2010; Juslin & Laukka, 2003; Lima & Castro, 2010). Such descriptors seem useful to distinguish not only between emotion categories (sadness, joy, etc.), but also to distinguish between valence classes (positive, negative, neutral).

A central contribution of our study for the neurocognition of emotion research field lies in the fact that our validated stimulus set offers a tool to address the cognition-related level of emotional processing in European Portuguese: emotional lexico-prosodic integration, assessed via lexico-prosodic incongruence detection. Besides the numerous applications that come with the possibility of addressing this substage of emotional prosody processing, our materials provide new opportunities to investigate the relations between the two substages of high-level cognitive processing in this domain: lexico-prosodic integration and explicit emotion recognition. Furthermore, it can be used to address later, highly cognitive emotional prosody subprocesses with no need to engage explicit emotion labelling (emotion recognition tasks, as in Lima & Castro, 2010). Given that lexico-prosodic incongruence triggers an implicit response (e.g., Paulmann & Kotz, 2008), this set is also appropriate for implicit tasks, which widens the experimental scope of possibilities in emotional prosody research. Also, given that the present set is composed of words and not sentences, it may be more suitable to ERP studies in that it allows for finer time-course parametrizations such as precise trigger positioning.

Our validated set also contributes to developments in clinical applications: as emotional prosody can be used to address the deterioration of neurocognitive systems (e.g., Lima et al., 2013; see also Bak, 2016), this set may serve as a tool to assess the integrity of vocal emotional processing in novel ways, namely without requiring patients to consciously recognize the emotions being conveyed. It may, furthermore, be used to address selective impairments of integration *vs.* emotional recognition, where one, but not the other, cognitive mechanism might be affected.

In sum, the present set of stimuli is relevant for various fields of research where emotion, language and its relations play a role.

Declaration of conflicting of interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Availability of data and materials

Supplementary materials, including all generated datasets and the final stimulus set are available in the OSF repository at: https://osf.io/t9z6g/?view_only=a2d2f057f25849a0a088a1cb4ed11697

Funding

This study was supported by Fundação para a Ciência e a Tecnologia (FCT) (UIDB/PSI/00050/2020) to the Center for Psychology at University of Porto.

Author contributions

Conceptualization: RG, SS; Data curation: SS; Formal analysis: RG; Methodology: SS; Project administration: SLC; Supervision: SLC, SS; Writing – Original draft: RG; Writing – Review and editing: RG, SLC, SS.

All authors read and approved the final manuscript.

References

- Altarriba, J., & Bauer, L. (2004). The distinctiveness of emotion concepts: A comparison between emotion, abstract, and concrete words. *The American Journal of Psychology*, *117*(3), 389-410. <https://doi.org/10.2307/4149007>
- Bak, H. (2016). The state of emotional prosody research – A meta-analysis. In R. Heredia & A. Cieslicka (Eds.), *Emotional prosody processing for non-native english speakers: Towards an integrative emotion paradigm* (pp. 79-105). Springer International Publishing. <https://doi.org/10.1007/978-3-319-44042-2>
- Banse, R., & Scherer, K. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*(3), 614-636. <https://doi.org/10.1037/0022-3514.70.3.614>
- Bostanov, V., & Kotchoubey, B. (2004). Recognition of affective prosody: Continuous wavelet measures of event-related brain potentials to emotional exclamations. *Psychophysiology*, *41*(2), 259-268. <https://doi.org/10.1111/j.1469-8986.2003.00142.x>
- Bradley, M. M., & Lang, P. J. (1999). *Affective Norms for English Words (ANEW): Instruction manual and affective ratings (Technical Report C-1)*. The Center for Research in Psychophysiology, University of Florida.
- Brück, C., Kreifelts, B., & Wildgruber, D. (2011). Emotional voices in context: A neurobiological model of multimodal affective information processing. *Physics of Life Reviews*, *8*(4), 383-403. <https://doi.org/10.1016/j.plrev.2011.10.002>
- Castro, S., & Lima, C. (2014). Age and musical expertise influence emotion recognition in music. *Music Perception*, *32*(2), 125-142. <https://doi.org/10.1525/mp.2014.32.2.125>
- Correia, A., Branco, P., Martins, M., Reis, A., Martins, N., Castro, S., & Lima, C. (2019). Resting-state connectivity reveals a role for sensorimotor systems in vocal emotional processing in children. *Neuroimage*, *201*, 116052. <https://doi.org/10.1016/j.neuroimage.2019.116052>
- Costa, S. (2012). *Adaptação e teste de uma base lexical de palavras emocionais para o português europeu: EMOTAIX.PT* [Dissertação de mestrado não publicada, Faculdade de Psicologia e de Ciências da Educação, Universidade do Porto, Porto, Portugal]. https://sigarra.up.pt/fpceup/pt/pub_geral.pub_view?pi_pub_base_id=30055
- Cowen, A., Laukka, P., Elfénbein, H., Liu, R., & Keltner, D. (2019). The primacy of categories in the recognition of 12 emotions in speech prosody across two cultures. *Nature Human Behaviour*, *3*(4), 369-382. <https://doi.org/10.1038/s41562-019-0533-6>
- Dowling, W. J., & Harwood, D. L. (1986). *Music Cognition*. Academic Press.
- Filipe, M., Branco, P., Frota, S., Castro, S., & Vicente, S. (2015). Affective prosody in European Portuguese: Perceptual and acoustic characterization of one-word utterances. *Speech Communication*, *67*, 58-64. <https://doi.org/10.1016/j.specom.2014.09.007>
- Giordano, B., Whiting, C., Kriegeskorte, N., Kotz, S., Gross, J., & Belin, P. (2021). The representational dynamics of perceived voice emotions evolve from categories to dimensions. *Nature Human Behaviour*. <https://doi.org/10.1038/s41562-021-01073-0>
- Goudbeek, M., & Scherer, K. (2010). Beyond arousal: Valence and potency/control cues in the vocal expression of emotion. *The Journal of the Acoustical Society of America*, *128*(3), 1322. <https://doi.org/10.1121/1.3466853>

- Hamann, S. (2012). Mapping discrete and dimensional emotions onto the brain: Controversies and consensus. *Trends in Cognitive Sciences*, 16(9), 458-466. <https://doi.org/10.1016/j.tics.2012.07.006>
- Handel, S. (1995). Timbre perception and auditory object identification. In B. C. J. Moore (Ed.), *Handbook of perception and cognition Hearing* (2nd ed., pp. 425-461). Academic Press. <https://doi.org/10.1016/B978-012505626-7/50014-5>
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393-402. <https://doi.org/10.1038/nrn2113>
- Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code?. *Psychological Bulletin*, 129(5), 770-814. <https://doi.org/10.1037/0033-2909.129.5.770>
- Kotz, S., & Paulmann, S. (2007). When emotional prosody and semantics dance cheek to cheek: ERP evidence. *Brain Research*, 1151, 107-118. <https://doi.org/10.1016/j.brainres.2007.03.015>
- Kousta, S., Vinson, D., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition*, 112(3), 473-481. <https://doi.org/10.1016/j.cognition.2009.06.007>
- Kousta, S., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General*, 140(1), 14-34. <https://doi.org/10.1037/a0021446>
- Lausen, A., & Hammerschmidt, K. (2020). Emotion recognition and confidence ratings predicted by vocal stimulus type and prosodic parameters. *Humanities and Social Sciences Communications*, 7(1). <https://doi.org/10.1057/s41599-020-0499-z>
- Lima, C. F., & Castro, S. L. (2010). Recognizing emotions in spoken language: A validated set of Portuguese sentences and pseudosentences for research on emotional prosody. *Behavior Research Methods*, 42(1), 74-81.
- Lima, C. F., Garrett, C., & Castro, S. (2013). Not all sounds sound the same: Parkinson's disease affects differently emotion processing in music and in speech prosody. *Journal of Clinical and Experimental Neuropsychology*, 35(4), 373-392. <https://doi.org/10.1080/13803395.2013.776518>
- Lin, Y., Ding, H., & Zhang, Y. (2020). Prosody dominates over semantics in emotion word processing: Evidence from cross-channel and cross-modal stroop effects. *Journal of Speech, Language, and Hearing Research*, 63(3), 896-912. https://doi.org/10.1044/2020_jslhr-19-00258
- Mitchell, R., Elliott, R., Barry, M., Cruttenden, A., & Woodruff, P. (2003). The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. *Neuropsychologia*, 41(10), 1410-1421. [https://doi.org/10.1016/s0028-3932\(03\)00017-4](https://doi.org/10.1016/s0028-3932(03)00017-4)
- Montenegro, C. S., & Maravillas, E. A. (2015). Acoustic-prosodic recognition of emotion in speech. Paper presented at the 8th IEEE International Conference Humanoid, Nanotechnology, Information Technology Communication and Control, Environment and Management (HNICEM). Cebu, Philippines. <https://doi.org/10.1109/HNICEM.2015.7393229>
- Paulmann, S. (2016). Neurocognition of prosody. In G. Hickok & S. L. Small (Eds.), *Neurobiology of language* (pp. 1109-120). Academic Press. <https://doi.org/10.1016/B978-0-12-407794-2.00088-2>
- Paulmann, S., & Kotz, S. (2008). Early emotional prosody perception based on different speaker voices. *Neuroreport*, 19(2), 209-213. <https://doi.org/10.1097/wnr.0b013e3282f454db>
- Paulmann, S., & Pell, M. (2011). Is there an advantage for recognizing multi-modal emotional stimuli?. *Motivation and Emotion*, 35(2), 192-201. <https://doi.org/10.1007/s11031-011-9206-0>
- Peeters, G., Giordano, B., Susini, P., Misdariis, N., & McAdams, S. (2011). The timbre toolbox: Extracting audio descriptors from musical signals. *The Journal of the Acoustical Society of America*, 130(5), 2902-2916. <https://doi.org/10.1121/1.3642604>

- Pell, M., & Kotz, S. (2011). On the time course of vocal emotion recognition. *Plos ONE*, 6(11), e27256. <https://doi.org/10.1371/journal.pone.0027256>
- Pell, M. D., & Skorup, V. (2008). Implicit processing of emotional prosody in a foreign versus native language. *Speech Communication*, 50(6), 519-530. <https://doi.org/10.1016/j.specom.2008.03.006>
- Pinheiro, A., Rezaii, N., Nestor, P., Rauber, A., Spencer, K., & Niznikiewicz, M. (2016). Did you or I say pretty, rude or brief? An ERP study of the effects of speaker's identity on emotional word processing. *Brain and Language*, 153-154, 38-49. <https://doi.org/10.1016/j.bandl.2015.12.003>
- Piolat, A., & Bannour, R. (2009). EMOTAIX: Un scénario de tropes pour l'identification automatisée du lexique émotionnel et affectif. *L'Année Psychologique*, 109, 655-698. <https://doi.org/10.4074/S0003503309004047>
- Preti, E., Suttora, C., & Richetin, J. (2015). Can you hear what I feel? A validated prosodic set of angry, happy, and neutral Italian pseudowords. *Behavior Research Methods*, 48(1), 259-271. <https://doi.org/10.3758/s13428-015-0570-7>
- Schirmer, A., & Kotz, S. (2003). ERP evidence for a sex-specific stroop effect in emotional speech. *Journal of Cognitive Neuroscience*, 15(8), 1135-1148. <https://doi.org/10.1162/089892903322598102>
- Schirmer, A., & Kotz, S. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, 10(1), 24-30. <https://doi.org/10.1016/j.tics.2005.11.009>
- Soares, A., Comesaña, M., Pinheiro, A., Simões, A., & Frade, C. (2011). The adaptation of the Affective Norms for English Words (ANEW) for European Portuguese. *Behavior Research Methods*, 44(1), 256-269. <https://doi.org/10.3758/s13428-011-0131-7>
- Soares, A., Costa, A., Machado, J., Comesaña, M., & Oliveira, H. (2016). The Minho Word Pool: Norms for imageability, concreteness, and subjective frequency for 3.800 Portuguese words. *Behavior Research Methods*, 49(3), 1065-1081. <https://doi.org/10.3758/s13428-016-0767-4>
- Tursunov, A., Kwon, S., & Pang, H. (2019). Discriminating emotions in the valence dimension from speech using timbre features. *Applied Sciences*, 9(12), 2470. <https://doi.org/10.3390/app9122470>
- Wright, A., Saxena, S., Sheppard, S., & Hillis, A. (2018). Selective impairments in components of affective prosody in neurologically impaired individuals. *Brain and Cognition*, 124, 29-36. <https://doi.org/10.1016/j.bandc.2018.04.00>

EmoLexPros: Uma base de dados de palavras emocionais prosodicamente incongruentes em português europeu

Resumo: No presente estudo, validou-se uma base de dados de estímulos áudio contendo informação léxico-prosódica congruente e incongruente. Palavras lexicalmente positivas, neutras e negativas, classificadas como tal em bases de dados portuguesas, foram gravadas com diferentes padrões prosódicos, formando pares congruentes e incongruentes. O conjunto produzido foi submetido a validação comportamental, onde cada palavra foi avaliada quanto à sua congruência léxico-prosódica. Os pares de palavras que mostraram a maior diferença entre as versões congruentes e incongruentes ($n=55$) foram posteriormente submetidos a análise acústica. Os três tipos de padrões prosódicos utilizados – positivo, neutro e negativo – apresentaram diferenças na frequência fundamental (F0), na amplitude, na duração e em parâmetros tímbricos. Por fim, os itens selecionados foram catalogados tendo em conta as suas propriedades psicolinguísticas. O presente conjunto de estímulos pode ser utilizado em estudos de prosódia emocional em português europeu, bem como junto de populações clínicas.

Palavras-chave: Prosódia emocional, Detecção de incongruências, Processamento da fala.

