PLEASE NOTE:

This is the author’s version of the manuscript accepted for publication in *Speech Communication*. Changes resulting from the publishing process, namely editing, corrections, final formatting for printed or online publication, and other modifications resulting from quality control procedures, may have been subsequently added.

Affective Prosody in European Portuguese: Perceptual and Acoustic Characterization of One-
Word Utterances

Marisa G. Filipe¹, Paulo Branco¹, Sónia Frota², São Luís Castro¹, and Selene G. Vicente¹

¹ University of Porto
² University of Lisbon

Author Note

Marisa G. Filipe, Faculty of Psychology and Education Sciences, University of Porto, Porto, Portugal; Paulo Branco, Faculty of Psychology and Education Sciences, University of Porto, Porto, Portugal; Sónia Frota, Centro de Linguística, University of Lisbon, Lisbon, Portugal; São Luís Castro, Faculty of Psychology and Education Sciences, University of Porto, Porto, Portugal; and Selene G. Vicente, Faculty of Psychology and Education Sciences, Porto, University of Porto, Portugal.

This research was supported by the Portuguese Foundation for Science and Technology (PEst-C/PSI/00050/2011; SFRH/BD/64166/2009; PEst-OE/LIN/UI0214/2013).

Correspondence concerning this article should be addressed to Marisa Filipe or São Luís Castro, Laboratório de Fala, Faculty of Psychology and Education Sciences, University of Porto, Rua Alfredo Allen, 4200-135 Porto, Portugal. Telephone: +351 226 079 700. Fax: +351 226 079 725. E-mail: labfala@fpce.up.pt.
Abstract

A perceptual and acoustic characterization was provided on the expression of liking and disliking in the European Portuguese language. Thirty participants identified vocal patterns and judged the intensity of expressed affect in one-word items recorded by six untrained speakers. Listeners consistently associated vocal profiles with the two emotional patterns of liking and disliking. However, liking intonation was easier to recognize than disliking intonation. The feature most commonly associated with liking intonation was a wider and higher F0 pattern and a rising-falling contour. For disliking, the results revealed a flatter melodic pattern with a fall into the stressed syllable yielding a low plateau. In sum, both prosodic patterns showed different and consistent correlates.

*Keywords*: emotional prosody, perceptual features, acoustic features, liking, disliking
Emotional prosody refers to the variation in acoustic cues such as fundamental frequency (F0), amplitude (or intensity), timing, and voice quality during speech that is used to convey the emotional meaning of an utterance (e.g., Kotz & Paulmann, 2011). The role of vocal cues in the expression of feelings and affection, as well as their powerful effects on social interaction, has been recognized for many years (e.g., Darwin, 1872/1965). Indeed, beginning with the first moments of life, emotional prosody is commonly used by developing individuals to express and perceive emotions. This ability is crucial to communication, and it might be atypical or impaired in some clinical cases, such as in children with speech and language impairments (Wells & Peppé, 2003). Consequently, it is both socially and clinically relevant to understand how individuals realize, for instance, that someone is irritated, frightened, or pleased. Because the vocal expression of affection and emotions is associated with perceptual identification and acoustic profiles (for a review, see Juslin & Laukka, 2003), the present study examined perceptual and acoustic features involved in the vocal expression of liking and disliking in the European Portuguese language.

Emotional prosody is crucial to human communication. Although the communication of emotions through facial expression has received much attention, psychologists, linguists, psycholinguists, and neuroscientists have been increasingly focusing on the vocal communication of emotions (e.g., Elfenbein, Marsh, & Ambady, 2002). There is currently a significant body of knowledge on emotional vocal cues (e.g., Juslin & Laukka, 2003;
Scherer, 2003). Several main questions have already been addressed. First, several studies indicated that verbal comprehension is affected by prosody. For example, Nygaard and Lunders (2002) found that a listener’s selection of word meaning in homophones (e.g., die/dye) with and without emotional meaning was influenced by the emotional tone of voice (see also: Ishii, Reyes, & Kitayama, 2003; Kitayama & Ishii, 2002). Neuroimaging studies have found general dissociations between emotional prosody and linguistic processing. For instance, individuals with damage to the right hemisphere have difficulty in identifying emotional prosody, although their ability to extract the literal linguistic content of the spoken utterance remains intact (Adolphs, Damasio, & Tranel, 2002). Therefore, it is important to take into account both linguistic and prosodic content when determining the emotional perception of a word.

Another issue that has been addressed is the search for vocal profiles that are associated with particular emotions. Juslin and Laukka (2001, 2003), among others (e.g., Banse & Scherer, 1996; Hammerschmidt & Jürgens, 2007; Paulmann, Pell, & Kotz, 2008; Scherer, Banse, Wallbott, & Goldbeck, 1991), have shown that there are specific acoustic profiles associated with the vocal expression of different emotion categories. For instance, happiness is characterized by high, rising, and highly variable F0, fast speech rate, medium-high voice intensity, medium high-frequency energy, and fast voice onsets. In contrast, sadness can be characterized by low, falling, and not so variable F0, slow speech rate, low voice intensity, little high-frequency energy, and slow voice onsets/attacks. In fact, acoustic variables such as speech rate, voice intensity, voice quality, and F0 (mean, maximum, minimum, and deviation) can be powerful means to categorize emotions (Juslin & Madison, 1999), and acoustic tools (i.e., discriminant acoustic vocal features) have been proven efficient in profiling emotions and in assisting emotional recognition.

Research has also demonstrated that listeners reliably and accurately perceive
emotions in the human voice. Typically, listeners select the emotion that best describes the stimulus from a list of emotion labels (forced-choice recognition task), and then the percentage of correctly recognized stimuli per emotion is calculated. These perceptual classifications suggest that listeners detect emotion in voice with great accuracy (Castro & Lima, 2010; Juslin & Laukka, 2003; Pell, Paulmann, Dara, Alasseri, & Kotz, 2009; Scherer, 2003; Scherer, Johnstone, & Klasmeier, 2003; Sobin & Alpert, 1999). As a result, most studies on emotional prosody used emotional identification tasks and asked the participants to identify the emotion presented in speech samples (e.g., Banse & Scherer, 1996; Scherer, 1979). For example, Castro and Lima (2010), with adult speakers of Portuguese, showed that discrete emotions (e.g., anger, disgust, fear, happiness, sadness, surprise, and neutrality) were recognized well-above chance and were given high-intensity ratings. However, not all emotions seem to be decoded equally well. Banse and Scherer (1996) found that anger and sadness were decoded more reliably than joy and fear. Considering the high levels of accuracy in the identification of most emotions based on vocal parameters, as well as findings suggesting that vocal profiles are associated with particular emotions, the assumption of distinct profiles for emotions based on acoustic and perceptual features is strongly supported in the literature.

Recently, research has focused on how the vocal expression of emotions varies among languages and cultures. Although some researchers have suggested that there are innate, stable, and universal features (e.g., Juslin & Laukka, 2003), differences have also been reported. For example, liking and disliking have been characterized differently across languages. For British English, this distinction can be expressed in the use of differing tones, rise-fall for liking and fall-rise for disliking (Martínez-Castilla & Peppé, 2008). Also, Banse and Scherer (1996) demonstrated that English utterances produced with a wider and higher pitch range suggest positive affect. On the other hand, for Spanish, Martínez-Castilla and
Peppé (2008) found that the intonation feature most indicative of liking was an inverted U-shaped contour with an exaggerated pitch peak on the stressed syllable, whereas disliking was mostly expressed by a flat melodic contour with a slight fall on the stressed syllable. Thus, there may be differences between languages in the expression of affective valences. Research has not yet addressed this topic extensively, and issues about language differences remain unclear.

In sum, the vocal expression of affect is reliably associated with acoustic cues and elicits reliable listener identification, and researchers in the field have conducted acoustic and perceptual analyses to establish the vocal patterning of expressed emotions. However, important issues remain to be addressed. Further research is needed to answer the question of whether the recognition of affect differs across languages, and to what extent the vocal cues have some degree of language specificity. European Portuguese includes properties of both Romance and Germanic languages in its phonology and prosody (Frota, 2000; Frota, 2014; see also Duncan et al., 2013), posing challenging questions for the study of emotions, and no study so far has characterized liking and disliking vocal patterns for this language. The present study was designed to address the following issues: How well are liking and disliking recognized on the basis of prosody, and which intonation features are involved in the expression of liking and disliking in general, and specifically in European Portuguese (EP)?

Method

Participants

Six native speakers of EP (from the Porto region; three women and three men) between 22 and 47 years ($M = 31.83, SD = 10.22$) recorded the production data. None of these speakers had had any specific vocal training. Another 30 undergraduate students (24
women and 6 men) studying psychology at the University of Porto, with a mean age of 19.3 years ($SD = 0.95$) participated in the perceptual experiment. All were native speakers of EP with no hearing difficulties or speech disorders.

Materials

A set of six one-word items for training plus 22 one-word items for the acoustic and perceptual test (i.e., experimental condition) were selected, taking into account criteria for high frequency and age of acquisition equal or inferior to 5 years (Cameirão & Vicente, 2010; Gomes & Castro, 2003; the complete list is in the Appendix). The words consisted of food-related items (e.g., laranja, orange) with different lengths in number of syllables (from 1 to 3). Most words had penultimate stress, since this is the most common stress pattern in Portuguese. For training, 2 words for each word length were used; for the experimental items, there were 11 disyllables, 10 three-syllable-words, and 1 monosyllable.

The 28 items were recorded by the six speakers. Six separate recording sessions of about one hour each were conducted, one for each speaker. All speakers were asked to produce the items in two different affective tones: liking and disliking. The instructions were, (A) “Say the name of the food item as if you like it” and (B) “Say the name of the food item as if you don’t like it.” Stimuli were recorded with a high-quality microphone in a soundproof room using the speech station of the Speech Laboratory at the University of Porto. The software used was Pro Tools LE version 5.1.1 with a 48-kHz sampling rate and 16-bit resolution.

Post-processing of the data included slicing the recording sessions into individual sound files containing one-word utterances, and setting onset and end of sound files to 100 ms silence. A total of 12 training items (6 liking + 6 disliking), two from each speaker, were randomly selected, and then 264 experimental items (22 items x 2 patterns x 6 speakers) were chosen for the perceptual test and for acoustic analysis. These items were divided into two
blocks [Block A (132 items) + Block B (132 items)] with a one-minute pause between blocks and an inter-stimulus interval of 4 seconds. Item order was randomized.

Procedure

For the perceptual test, participants were individually tested in one experimental session that lasted about 20 minutes. The training block ($N = 12$) was presented first in order to familiarize the subject with the task and to avoid learning curve effects. It was followed by the two blocks of experimental trials. Reaction times and accuracy of the responses were recorded with SuperLab V4.0 (Abboud, Schultz, & Zeitlin, 2006). All items were presented through high-quality headphones. Participants were asked to perform two independent, consecutive judgments for each stimulus: (1) classify each stimulus as neutral, expressing liking, or expressing disliking (forced-choice task), and (2) grade the perceived intensity of the expressed emotion on a seven-point scale from 1, not intense, to 7, very intense.

Additionally, an acoustic analysis of F0 (F0 mean, F0 range, F0 maximum, F0 minimum, and F0 contour), intensity, and duration was conducted for the 264 experimental items using PRAAT (Boersma & Weenink, 2011). The parameters were chosen because they have been widely investigated in the expression of affective patterns. Also, we adopted the Autosegmental-Metrical Model of intonational phonology (Ladd, 2008) and the Tones and Break Indices framework (Beckman, Hirschberg, & Shattuck-Hufnagel, 2005), which have been used for several languages including Portuguese (Frota, 2014) in order to perform a phonological analysis of the nuclear contour patterns obtained in the data.

Results

Perceptual Analysis

The initial analysis explored the accuracy in identifying liking and disliking patterns. Accuracy measured in percent of correct identifications was calculated for each speaker (see
Table 1). Mean identification accuracy ranged from 73% to 91%, showing that participants were able to identify the affective patterns with great accuracy. Moreover, inter-participant consistency in the identification of the intended patterns was very high for every speaker: Cronbach’s alpha was 0.92. Regarding the intensity of the ratings, the vast majority of the items were classified as having 5 or more points on the seven-point scale. There was a moderate Pearson correlation between the percentage of correct identifications and the intensity of expressed affect ($r = .66; p < .001$), showing that the higher the perceived intensity, the higher the accuracy in identification.

Table 1 about here

We tested the normality of all of the variables (accuracy, F0 mean, F0 range, maximum F0, minimum F0, duration, and intensity) with the Kolmogorov-Smirnov Test, and the data were not normally distributed. We then inspected skewness and kurtosis, which are common criteria for assessing normality, and the resulting values fell under 1 for all of the variables, well below the values recommended as acceptable by Kline (2005). This led us to use parametric methods.

One of our interests was to examine the accuracy of the participants’ answers according to word length. Monosyllables had an overall accuracy of 79%, disyllables 82%, and 3-syllable words 85%. We found that word length had no influence on accuracy ($F (2, 251) = 1.83; ns.$). Regarding the affective patterns, liking items obtained about 92% of correct answers, and disliking items 75%. Analyses of variance (ANOVA) showed that liking items were significantly easier to recognize than disliking items ($F (1, 251) = 54.19; p < .001; \eta^2 = .178$). Finally, words produced by men (85% correct answers) were easier to recognize than the words produced by women (80% correct answers; $F (1, 251) = 4.52; p < .05; \eta^2 = .018$).
although with an extremely small effect size. In sum, our results demonstrated that accuracy was high, and liking prosody was more easily recognized. Variability in the vocal cues and individual differences involved in the expression of liking and disliking were analyzed next.

Acoustic Analysis

The mean and standard deviations of the acoustic measures — F0 mean, F0 range, maximum F0, minimum F0, and duration — are presented in Table 2. As expected, there were significant differences between women and men in all F0 measurements. To control for gender variance, we used z-transformed residuals of the acoustic parameter scores. As shown in Table 3, the z-scores independent of gender indicated a positive deviance towards the mean in the liking condition, whereas disliking had a negative deviance from the mean. Liking and disliking differed from one another in F0 mean, F0 range, F0 maximum, F0 minimum, and duration, as revealed by one-way ANOVAs and the robust test of equality of means with the Brown-Forsythe statistic ($F(1, 251) = 50.16; p < .001; \eta^2 = .161; F(1, 251) = 36.56; p < .001; \eta^2 = .123; F(1, 251) = 66.38; p < .001; \eta^2 = .202; F(1, 251) = 11.66; p = .001; \eta^2 = .043; F(1, 251) = 4.41 p < .037; \eta^2 = .017$, respectively).

Preliminary analyses revealed that there was no violation of the assumptions for normality, homocedasticity, linearity, and independence of errors. So, a standard multiple regression was used to assess the significance of the dummy variable “affective pattern” (i.e., liking vs. disliking) predicted by the z-scores of each acoustic parameter. The variables were entered in the model by the enter method. The model ($R^2 = 0.25, F(5, 257) = 17.188; p < .001$) that predicted the affective patterns was composed by duration ($b = -.108; \text{beta} = -.217$),
maximum F0 $b = .246; \beta = -.490$, mean F0 $b = -.058; \beta = -.115$, minimum F0 $b = -.079; \beta = -.158$, and F0 range $b = -.016; \beta = -.032$. The strongest predictor was F0 maximum, followed by duration.

As an illustration of F0 contours involved in the expression of like and dislike in EP, we plotted the mean F0 for one word. The word limão, lemon, was chosen for its all sonorant nature that provided a continuous F0 trace; it is a word with final stress. Liking and disliking contours differed from each other on the acoustic parameters analyzed in this study. As shown in Figures 1 and 2, in the case of this word the strongest indicator was F0 maximum: liking had a higher F0 peak on the stressed syllable than disliking. For disliking, there was a flatter F0 pattern with a fall on the stressed syllable. Figures 3 and 4 illustrate the typical phonological nuclear configuration of liking and disliking, according to the autosegmental-metrical framework\(^1\) for the analysis of intonation: the most indicative liking pattern is expressed by a high plateau on the stressed syllable followed by a fall (H* L%; 85% of the trials), whereas the most indicative disliking pattern is conveyed by a fall into the stressed syllable yielding a low plateau (H L* L%; 73% of the trials). Importantly, the stressed syllable is high in the liking pattern (H*) and low in the disliking pattern (L*). Regarding the second strongest parameter, duration, disliking was longer than liking.

\[^1\] The autosegmental-metrical framework uses High (H) and Low (L) tone labels to transcribe melodic targets observed on fundamental frequency curves. The label * is used for the tone associated with the stressed syllable, and the label % is used to signal a boundary tone.
Discussion

Given that research on affective prosody had not yet identified the vocal cues for liking and disliking in EP, the purpose of this paper was to describe and analyze the perceptual and acoustic correlates of these two patterns applied to one-word items. Our results showed that there were clear differences between the acoustic profiles of the liking and disliking patterns, and listeners were able to identify these patterns with high accuracy. In general, the intonation feature most frequently associated with liking was a wider and higher F0 pattern, with a peak on the stressed syllable, as well as an overall rising-falling prosodic contour (H* L%). For disliking, results showed a flatter melodic pattern with a fall into the stressed syllable which displays low pitch (H L* L%). Duration was also a strong indicator, at least in in one-word utterances: disliking was longer than liking. Furthermore, liking was more easily recognized than disliking.

The first research question of this study was how well liking and disliking are perceptually recognized on the basis of prosody. Our findings showed that individuals were able to identify these affective patterns with great accuracy, regardless of the speaker (73% to 91%). The degree of accuracy was comparable to what has been found in several studies on the identification of emotion categories using semantically neutral sentences spoken in different intonations. For instance, in Portuguese, Castro and Lima (2010) reported a mean correct of 75% for sentences and 71% for pseudosentences for the seven emotions studied. In English, the results were similar: Adolphs and colleagues (2002) reported 81% correct identifications of the five emotions studied (happy, sad, angry, afraid, and surprised). Moreover, our results suggested that the patterns of liking and disliking were not equally easy to decode: the negative pattern (disliking) was less well recognized compared to the positive one (liking). An explanation could be the use of different intonation features (85% of the trials presented the H* L% pattern for liking, whereas only 73% of the trials presented the H
L* L% pattern for disliking) that may justify some of the misidentified stimuli in the perceptual tests. However, a similar difference between negative and positive patterns has been suggested by several authors (e.g., Adolphs, Damasio, & Tranel, 2002; Banse & Scherer, 1996; Scherer et al., 1991). One possible explanation could be the fact that in natural conversation negative emotions (e.g., disgust) are probably expressed by face mimicking together with voice (Castro & Lima, 2010). Even if not all emotions have the same accuracy in identification, it is clear that humans identify emotions from speech with a very high degree of accuracy.

The second question addressed the characterization of the intonation features involved in the expression of liking and disliking in EP. Consistent with previous research in other languages (e.g., Banse & Scherer, 1996), the present experiment indicated that specific acoustic features were related to the two contrastive affective patterns studied here, and that listeners were able to reliably recognize the two patterns on the basis of vocal cues. In EP, the intonation feature most indicative of liking is a wider and higher F0 and a rising-falling contour, with high pitch on the stressed syllable. For disliking, the results showed a flatter melodic pattern with low pitch on the stressed syllable.

Finally, the last question was the cross-language comparison of the vocal profiles for affective patterns. The profile for EP is similar to English with respect to the intonation of liking — both show a rising-falling contour, but different from Spanish which shows a falling-rising contour. For disliking, Portuguese is similar to Spanish showing a flat pattern with a fall and low pitch on the stressed syllable, which is different from the fall-rise found in English (Martinez-Castilla & Peppé, 2008). Nevertheless, across languages, wider and higher pitch seems to be a general feature of liking patterns. These results suggest that the phonological contours linked to affect patterns may differ across languages (possibly, depending on the independent features of their specific intonation systems), whereas the
phonetic implementation of affect patterns seems to be universal (possibly along the lines of affective interpretations of biological codes in pitch range, as suggested in Gussenhoven 2004). Further cross-linguistic research is needed to determine whether these suggestions are on the right track.

On the whole, this study has contributed to the characterization of the acoustic features that underlie the affective patterns of liking and disliking. Particularly, this study has revealed that both patterns are expressed in a consistently different way, but also with possible differences between languages. Future studies should address important methodological issues that remain to be explored, including studying other acoustic factors that may have an impact on the perception of liking and disliking such as intensity or voice quality, using naturalistic recordings and continuous speech, and a large variety of items (for instance, shorter and longer utterances).

Because it is socially and clinically relevant to understand how emotions are vocally expressed, the present study also has clinical implications. The main contribution — establishing the perceptual and acoustic features involved in the expression of liking and disliking in EP — offers guidelines for the understanding of affective prosody in Portuguese. As affective prosody is widely accepted as a key deficit in clinical populations such as those with schizophrenia (e.g., Pijnenborg, Withaar, van den Bosch, & Brouwer, 2007), Alzheimer's disease (e.g., Taler, Baum, Chertkow, & Saumier, 2008), and autism (e.g., McCann & Peppé, 2003), therapists must know the normative profile in order to offer intervention, and researchers need to provide adequate assessment procedures. Therefore, the present study contributes to the characterization of affective prosody in EP with implications for its assessment and intervention in clinical populations.
Acknowledgments

This research was supported by the Portuguese Foundation for Science and Technology (PEst-C/PSI/IU0050/2011; SFRH/BD/64166/2009; PEst-OE/LIN/UI0214/2013).

The authors would like to thank Marisa Cruz of the Center of Linguistics of the University of Lisbon for her help with the figures.
References


Duncan, L. G., Castro, S. L., Defior, S., Seymour, P. H. K., Baillie, S., Leybaert, J., …Serrano, F. 2013. Phonological development in relation to native language and


Appendix

One-Word Utterances Included in the Training and Experimental Sessions

<table>
<thead>
<tr>
<th>Training Item (Translation)</th>
<th>Experimental Item (Translation)</th>
<th>Experimental Item (Translation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Noz (Nut)</td>
<td>1) Limão (Lemon)</td>
<td>12) Pão (Bread)</td>
</tr>
<tr>
<td>2) Figo (Fig)</td>
<td>2) Salada (Salad)</td>
<td>13) Maçã (Apple)</td>
</tr>
<tr>
<td>3) Pepino (Cucumber)</td>
<td>3) Ervilhas (Peas)</td>
<td>14) Iogurte (Yogurt)</td>
</tr>
<tr>
<td>4) Mel (Honey)</td>
<td>4) Ovo (Egg)</td>
<td>15) Sumo (Juice)</td>
</tr>
<tr>
<td>5) Café (Coffee)</td>
<td>5) Laranja (Orange)</td>
<td>16) Massa (Pasta)</td>
</tr>
<tr>
<td>6) Cebola (Onion)</td>
<td>6) Uvas (Grapes)</td>
<td>17) Salsicha (Sausage)</td>
</tr>
<tr>
<td>7) Bolacha (Cookie)</td>
<td>18) Banana (Banana)</td>
<td></td>
</tr>
<tr>
<td>8) Peixe (Fish)</td>
<td>19) Pêra (Pear)</td>
<td></td>
</tr>
<tr>
<td>9) Cenoura (Carrot)</td>
<td>20) Bolo (Cake)</td>
<td></td>
</tr>
<tr>
<td>10) Leite (Milk)</td>
<td>21) Queijo (Cheese)</td>
<td></td>
</tr>
<tr>
<td>11) Tomate (Tomato)</td>
<td>22) Gelado (Ice Cream)</td>
<td></td>
</tr>
</tbody>
</table>
Figure Caption
Figure 1. Female fundamental frequency contour. Mean F0 contour of the utterances produced by female speakers (N = 3) for the word limão (lemon), for liking and disliking. In the x-axis, time was normalized as a percentage of the duration of the word.
Figure 2. Male fundamental frequency contour. Mean F0 contour of the utterances produced by male speakers (N = 3) for the word limão (lemon), for liking and disliking. In the x-axis, time was normalized as a percentage of the duration of the word.
Figure 3. F0 contour of the word *Limão* (lemon) for liking. The label tiers indicate the tonal analysis and the orthographic transcription, respectively.
Figure 4. F0 contour of the word Limão (lemon) for disliking. The label tiers indicate the tonal analysis and the orthographic transcription, respectively.