frontiers | Frontiers in Communication

TYPE Original Research
PUBLISHED 06 February 2026
DOI 10.3389/fcomm.2026.1728758

OPEN ACCESS

EDITED BY
Donna Mae Erickson,
Haskins Laboratories, United States

REVIEWED BY
Tommaso Raso,
Minas Gerais State University, Brazil
Toshiyuki Sadanobu,
Kyoto University, Japan

*CORRESPONDENCE
Marisa Cruz
✉ marisac@edu.ulisboa.pt

RECEIVED 20 October 2025
REVISED 16 January 2026
ACCEPTED 19 January 2026
PUBLISHED 06 February 2026

CITATION
Santos D and Cruz M (2026) Perception of emotions across Portuguese varieties. *Front. Commun.* 11:1728758. doi: 10.3389/fcomm.2026.1728758

Perception of emotions across Portuguese varieties

Diana Santos and Marisa Cruz*

Centre of Linguistics, School of Arts and Humanities, University of Lisbon, Lisbon, Portugal

This study investigates the perception of emotional prosody across two major varieties of Portuguese—Brazilian (BP) and European (EP)—examining how differences in their intonational systems and cultural backgrounds shape emotion recognition. Building on theoretical frameworks of vocal emotion, an experiment was designed using acoustically controlled, acted stimuli expressing neutrality, happiness, sadness, and anger. Native listeners from both varieties completed a perception task measuring identification accuracy and reaction times. Results revealed a clear emotion hierarchy: neutrality and sadness were recognized most accurately and rapidly, whereas happiness and anger were frequently confused, indicating higher perceptual ambiguity. While both listener groups showed a native-variety advantage, EP participants demonstrated superior overall accuracy, attributed to greater exposure to BP media, and more consistent cue reliance. These findings highlight the interplay between universal acoustic cues, variety-specific phonological structuring, and cultural exposure in shaping emotional speech perception. This research contributes to prosodic typology in Portuguese, cross-variety speech perception, and the integration of linguistic and affective models, with implications for phonological theory, applied technologies, and communication in general.

KEYWORDS

Brazilian Portuguese, cross-linguistic perception, cultural exposure, emotional prosody, European Portuguese, intonation, prosodic typology, speech perception

1 Introduction

The human voice is one of the most powerful channels for emotional communication. Through prosodic modulations of pitch, intensity, duration, and rhythm, it conveys affective states that often transcend linguistic boundaries, enabling listeners to infer emotion even when semantic information is absent. Early evolutionary accounts such as Darwin's *The Expression of the Emotions in Man and Animals* (1872) viewed emotional expression as a biological adaptation that evolved to facilitate survival and social coordination. By vocalizing anger, fear, or joy, humans and other animals transmit signals that prepare receivers for appropriate behavioral responses. These principles laid the foundation for modern theories of emotion as biologically grounded yet culturally mediated phenomena.

Subsequent theorists expanded Darwin's framework. Plutchik (1982) proposed a psychoevolutionary model in which emotions represent adaptive responses to environmental challenges. He identified a core set of primary emotions—happiness, sadness, anger, fear, surprise, and disgust—that are universal across species and serve fundamental communicative and survival functions. Paul Ekman's empirical work in the second half of the twentieth century provided robust evidence supporting this universality. His cross-cultural studies demonstrated that people across distinct societies could recognize the same basic emotions from facial expressions, even in the absence of shared language or culture (Ekman, 1992; Ekman and Friesen, 1971). These findings,

later extended to vocal expression, established the notion that the acoustic parameters of the voice—such as pitch height, intensity, and temporal variation—encode biologically grounded affective information.

The study of vocal emotion has since then evolved into a multidisciplinary field integrating psychology, linguistics, and neuroscience. Among its most influential models is Scherer's *Component Process Model* (1986, 2003), which describes emotion as the outcome of recursive appraisal processes that modulate physiological subsystems like respiration, phonation, and articulation. These physiological changes are acoustically realized in speech, producing prosodic cues that listeners decode to infer the speaker's emotional state. Anger and happiness are typically characterized by higher mean fundamental frequency (F0), broader pitch range, and faster speech rate, reflecting heightened arousal, whereas sadness tends to display lower pitch, reduced intensity, and slower tempo, associated with physiological relaxation and low activation (Banse and Scherer, 1996; Castro and Lima, 2010).[1] These patterns are largely consistent across languages, suggesting the existence of universal acoustic correlates of emotion. Yet, despite their universality, emotional prosodies are not entirely invariant: each language's phonological and rhythmic structure constrains how these acoustic parameters are expressed and perceived.

Prosody is an intrinsic property of language that interacts with affective meaning. While certain acoustic dimensions correlate broadly with emotional categories, their perceptual salience varies depending on the phonological system. For instance, tonal languages such as Mandarin use pitch lexically to distinguish word meaning, limiting its flexibility as an emotional cue and prompting speakers to rely more heavily on intensity, tempo, or voice quality to express affect (Wang et al., 2018). This demonstrates that although emotional vocalization arises from biological mechanisms, its expression is filtered through linguistic systems, resulting in what Laukka and Elfenbein (2021) term "emotional dialects" —language-specific prosodic configurations shaped by both phonetic constraints and cultural norms. Consequently, understanding emotional communication through speech requires not only examining the acoustic correlates of emotion but also the phonological context in which they are embedded.

Portuguese offers a compelling case study for examining how linguistic and cultural factors intersect in emotional prosody. Brazilian Portuguese (BP) and European Portuguese (EP), though sharing a common grammar and lexicon, differ markedly in their prosodic organization. BP exhibits high tonal density, a wide pitch range, and frequent melodic variation, which give rise to a dynamic and rhythmically "chanted" intonation (Frota and Moraes, 2016; Frota et al., 2015a). EP, by contrast, presents fewer tonal events, a narrower pitch range, and longer prosodic phrases with flatter melodic contours (Frota and Vigário, 2000). Rhythmically, these varieties also differ, although exhibiting a mixed pattern: BP favors syllable- and mora-timed rhythm with frequent vowel epenthesis, while EP combines syllable- and stress-timed patterns characterized by vowel reduction and deletion (Barbosa, 2006; Frota and Vigário, 2000, 2001; Vigário, 2003). Such differences are not purely linguistic; they also shape how emotional cues are produced and interpreted. The high tonal density (and dynamic pitch movements) typical of BP might amplify the perceptual salience of high-arousal emotions such as happiness or anger, whereas EP's sparse tonal distribution (thus, more restrained intonation) may facilitate the recognition of lower-arousal emotions such as sadness or neutrality. These prosodic asymmetries invite investigation into whether listeners of each variety show distinct perceptual strategies and sensitivities to prosodic cues.

Cultural context further complicates this picture. Vocal emotion expression is not only a function of physiology and phonology but also of social convention. Cross-cultural psychology demonstrates that cultures differ in emotional "display rules" —socially learned norms governing the appropriateness and intensity of emotional expression (Ekman and Friesen, 1969; Matsumoto, 1990). Hofstede (2001) model of cultural dimensions, particularly individualism vs. collectivism, provides a framework for understanding such differences. Individualistic societies, which value autonomy and self-expression, tend to encourage open and frequent emotional display, especially for emotions signaling personal agency like anger or pride. Collectivist societies, which prioritize group harmony, often promote emotional moderation, and especially for negative or disruptive emotions. Although both Brazil and Portugal are Western societies sharing many cultural values, they differ in their relative positioning on this dimension: Brazil scores lower on individualism (38) than Portugal (63), suggesting stronger collectivist tendencies and more expressive social interaction norms (Hofstede, 2001). This cultural profile may correspond with BP's acoustically exuberant intonation, reflecting a greater tolerance for emotional expressivity in speech, whereas EP's melodic restraint may parallel its comparatively higher uncertainty avoidance and social formality.

An additional cultural variable influencing perception is asymmetrical cross-cultural exposure. Owing to Brazil's dominant media industry, Portuguese audiences are frequently exposed to Brazilian television, music, and cinema, whereas the reverse exposure is limited. This asymmetry could enhance the perceptual flexibility of EP listeners by increasing their familiarity with BP's intonational contours and emotional expressivity. In contrast, BP listeners' more limited exposure to EP prosody might hinder their ability to interpret its subtler acoustic cues. The combination of linguistic, cultural, and experiential factors thus creates an intricate landscape for examining emotional prosody within a single language continuum.

The present study investigates how these interrelated dimensions—universal acoustic cues, prosodic structure, and cultural exposure—jointly shape the perception of emotional prosody in Brazilian and European Portuguese. It examines whether listeners from both varieties differ in their ability to recognize emotions expressed through speech, whether their native prosodic systems influence emotional recognition accuracy

1 Importantly, a more detailed analysis of some acoustic cues, such as the F0 excursion, also allows to distinguish between hot anger, louder and with a higher F0, and cold anger (Bänziger and Scherer, 2005) or between quiet/passive sadness and active grief, the latter being louder and with a higher F0 (Scherer, 1979). These differences are not observed when only the overall F0 range is considered to the characterization of emotions.

and reaction time, and whether asymmetrical cultural exposure modulates cross-variety perception. Based on prior evidence, it is expected that listeners will show a native-variety advantage (Hypothesis 1), reflecting familiarity with their own intonational system; that BP's expressive prosody will favor recognition of high-arousal emotions, while EP's more restrained prosodic profile will facilitate identification of low-arousal states (Hypothesis 2)[2]; that Brazil's collectivist cultural orientation will lead to a higher rate in emotions recognition by BP listeners, outperforming EP listeners, culturally and linguistically more restrained (Hypothesis 3) and that EP listeners, owing to greater exposure to BP media, will outperform BP listeners in recognizing emotions produced in the non-native variety (Hypothesis 4).

By integrating evolutionary, linguistic, and cultural perspectives, this research aims to clarify how emotion perception operates within a shared but prosodically diversified linguistic system. It contributes to ongoing debates about the universality vs. specificity of emotional prosody, offering empirical evidence that emotional decoding in speech is neither purely biological nor purely conventional but emerges from the interplay of cross-linguistic regularities and variety-specific phonological and cultural patterns.

2 Materials and method

2.1 Participants

Twenty-eight adult native speakers of Portuguese participated in the perception experiment, divided evenly between the two major national varieties: fourteen speakers of Brazilian Portuguese (BP) and fourteen speakers of European Portuguese (EP). Participants were recruited through university mailing lists and social networks in Portugal and in Brazil. All were born and raised in monolingual households, reported no speech, hearing, or neurological disorders, and had normal or corrected-to-normal vision. The sample included 18 females and 10 males, aged between 20 and 35 years ($M = 26.1$, $SD = 3.9$). None of the participants had formal training in linguistics, phonetics, or music, minimizing potential expertise effects on prosodic sensitivity. All participants provided informed consent, previously validated by the Ethical Committee of the University of Lisbon and were compensated with a small token of appreciation for their time.

2.2 Stimuli

The speech stimuli consisted of three syntactically simple, semantically neutral declarative sentences: *Esta mesa é de madeira* ("This table is made of wood"), *As pessoas vão a concertos* ("People go to concerts"), and *O futebol é um desporto* ("Soccer is a sport"). These sentences were originally validated by Castro and Lima (2010) for use in research on emotional prosody. Their selection was motivated by three criteria: (a) syntactic and lexical simplicity, which reduces potential confounds arising from semantic processing; (b) equivalent phonological length, allowing direct comparison of duration and pitch contours across conditions; and (c) high lexical frequency, ensuring that all words were equally familiar to both Brazilian and European participants.

2.3 Stimuli recording

The sample for the recording of the stimuli consisted of two professional actresses, one Brazilian and one Portuguese, both female. Selection criteria included formal academic training in acting and proven ability to modulate emotional prosody. Female voices were chosen due to their acoustic properties, which are less susceptible to phenomena such as creaky voice.

The recording sessions, lasting approximately 30 min each, were conducted individually in the Phonetics and Phonology Laboratory at the University of Lisbon, in a semi-soundproof environment. Each actress produced the selected sentences in a neutral tone and in three emotional expressions (happiness, sadness, and anger), aiming for a natural yet clear emotional intonation. No specific instructions were given regarding pitch, pace, or intensity, although feedback was provided by the researcher to ensure authenticity. Recordings were made in.wav format using Pro Tools LE 5.1.1 software, a high-quality microphone, and an Apple Macintosh G4 computer, with a sampling rate of 41 kHz and 16-bit resolution.

2.4 Stimuli pre-processing

Audio segmentation was performed using Praat (Boersma and Weenink, 2022), to create individual files per sentence/emotion for consistent analysis. A total of 52 stimuli underwent acoustic analysis focusing on the following parameters: minimum, maximum, and mean F0 (Hz), mean intensity (dB), and speaking rate (i.e., number of phonological syllables divided per utterance total duration) in order to ensure that emotions produced fitted the acoustic properties pointed out in Castro and Lima (2010). Additionally, an intonational analysis was also performed, using P-ToBI (Frota et al., 2015b), a standardized transcription system for Portuguese prosody, to ensure adherence to tonal patterns described for each variety (Frota and Vigário, 2000; Frota et al., 2015a).

Based on acoustic and intonational consistency, a subset of 24 stimuli (12 per variety; 3 per emotion) was then selected to be used in the perception experiment. Table 1 displays the averaged acoustic[3] and prosodic characterizations for the selected stimuli only.

2 For the Hypothesis 2 formulation, we also took into account the Contrastive Analysis Hypothesis (Lado, 1957). Following this Hypothesis, we expected a result similar to the positive transfer observed in a second language acquisition process, i.e., the high linguistic proximity of an emotion to the prosodic characteristics of a given variety would facilitate the recognition of those emotions by that variety.

3 Although voice quality has been shown by prior research as a relevant cue to the emotions characterization, in the current study we were more focused on the role of the phonological grammar of each Portuguese variety to the perception of emotions, and voice quality does not play, as far as we know, any phonological role in these Portuguese varieties.

TABLE 1 Average acoustic patterns and prosodic characterizations of BP and EP.

Variety	Emotion	Min F0 (Hz)	Max F0 (Hz)	Mean F0 (Hz)	Mean intensity (dB)	Speaking rate (syll/sec)	Dominant pattern (%)
BP	Neutral	138.9	264.3	212.3	64.9	5.49	H+L* L%
	Happiness	153.4	401.4	317.5	78.0	4.59	H* HL%
	Sadness	141.7	255.4	205.2	59.3	4.28	H+L* L%
	Anger	143.0	365.7	279.9	77.0	4.93	H+! H* HL%
EP	Neutral	162.4	248.6	206.0	66.1	6.57	H+L* L%
	Happiness	191.2	387.0	305.9	75.2	5.27	H*+L L%
	Sadness	140.4	253.4	209.5	58.5	5.14	H* L%
	Anger	173.1	400.9	297.7	78.0	4.69	H*+L L%

2.5 Experimental design of the perception task and procedure

An overt emotion identification task was implemented in SuperLab 6.4 (Cedrus Corporation), consisting of a training block followed by a test block. The training block contained eight stimuli: one sentence per emotion (neutral, happiness, sadness, and anger) for each variety (BP, EP). Importantly, the sentences used in the training block were not used in the test block, ensuring that participants could not rely on memorization. The test block consisted of 16 different stimuli (eight per variety), each repeated three times in randomized order, totalizing 48 test trials. In both blocks, participants selected one of five response options—neutral, happiness, sadness, anger, or other—using pre-assigned keyboard keys. Responses were open-ended in the sense that no cues about correctness were provided, and no feedback was given at any stage of the experiment.

A schematic representation of the experimental design is presented in Figure 1, illustrating the sequence of the training and test phases, the stimulus distribution, and the exposure to both varieties.

All participants were exposed to both native and non-native stimuli, allowing for both within- and between-subject comparison analyses. Thus, the native variety of participants (BP/EP), the variety they perceived (native/non native), and the emotion type (neutral, happiness, sadness, anger, or other) were the independent variables considered in the analysis. As dependent variables, accuracy (i.e., correctness in the identification of emotions) and reaction times (in milliseconds) were measured.

Data collection was conducted remotely: EP participants completed the task in Lisbon, BP participants in Brazil.[4] In both cases, *SuperLab TaskPlayer* was temporarily installed on participants' own laptops to ensure controlled presentation and timing, as the task was set to run only once per participant. They were instructed to wear headphones while doing the task, and they were informed that they would be listening to acoustically modified sounds and that repetition was not allowed. At the end of the task, their responses were automatically stored in the *SuperLab* cloud.

2.6 Statistical analysis

The data analysis employed three key statistical approaches. Chi-square tests evaluated how emotion identification accuracy was influenced by participants' native variety (BP/EP), the variety being perceived (native/non-native), and the emotion type. Reaction times were analyzed with a Generalized Linear Mixed Model (GLMM), testing the fixed effects of native variety, perceived variety, emotion type, and response correctness, along with their critical interactions. Finally, a multinomial logistic regression was used to determine if accuracy could be predicted by acoustic/prosodic parameters (F0, intensity, speaking rate, and intonational pattern), also considering the main factors.

3 Results

3.1 Overall identification of emotions

The overall accuracy rate across all participants and conditions was 71.7%, indicating that participants were generally successful in identifying emotions based on prosodic cues. Table 2 summarizes the accuracy rates per emotion. The highest accuracy was observed for neutral productions (79.5%), followed by sadness (74.1%). Happiness exhibited the lowest accuracy rate (64.6%).

In order to assess how each intended emotion was perceived by participants we may conclude that happiness and anger triggered some difficulties, as 18.1% of happiness stimuli were misclassified as anger, and 16.2% of anger stimuli were misclassified as happiness. Neutral and sadness were less frequently confused with other emotions, thus suggesting that these two emotions are easily identifiable based on prosodic information (Table 3).

While the overall confusion matrix provides a broad overview of emotion identification, the relatively consistent rate of "Other" responses (ranging from 2.2 to 4.8%) warrants a deeper investigation. A critical question is whether these "Other" responses are distributed randomly across all stimuli or if they are

4 Because the researcher who collected the data was based in Lisbon, it was easier to limit the participants' recruitment to the Lisbon city, where Standard European Portuguese is spoken. In Brazil, the task was run in more than one city (e.g., Brasília, Rio de Janeiro, São Paulo), as it was not as easy to recruit the participants at distance. However, we believe this did not affect our results: 'Participant' was included in our statistical analysis as a random factor, and no significant effect was found ($p > 0.05$).

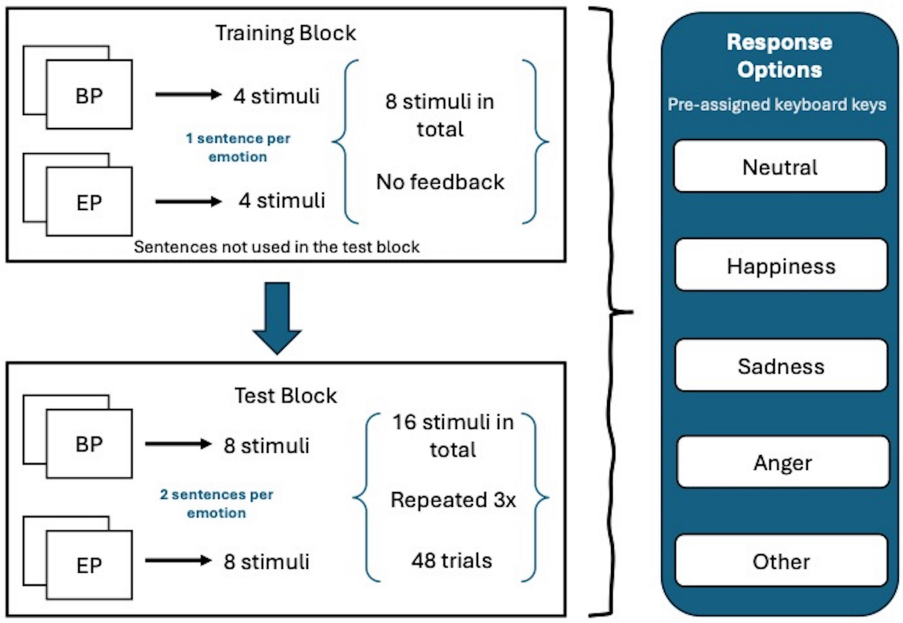

FIGURE 1
Schematic of the experimental design.

TABLE 2 Accuracy rates per emotion.

Emotion	Accuracy (%)
Neutral	79.5
Happiness	64.6
Sadness	74.1
Anger	68.5
Average Accuracy	**71,7**

The bold value highlights the overall average accuracy rate in the perception task.

concentrated on a few specific items. The analysis of the frequency of "Other" responses per individual stimulus revealed that the responses were not randomly distributed. Instead, they were highly concentrated on a very small subset of the stimuli. In particular, EP tokens of happiness were responsible for the largest amount of "Other" classifications. Importantly, this pattern was observed in participants from both varieties: BP listeners showed a peak of "Other" responses when perceiving EP happiness tokens (eight cases), and EP listeners likewise chose "Other" responses predominantly for EP happiness (seven cases). By contrast, BP tokens elicited far fewer "Other" responses for happiness, and the few misclassifications were spread more evenly across emotions. This cross-variety consistency indicates that the difficulty does not stem from a single listener group, but rather reflects particular properties of the EP happiness stimuli themselves, which in turn helps explain the lower identification accuracy for this emotion.

This finding strongly suggests that the "Other" category was not used as a mere "I don't know" option for a generally difficult task. Rather, it was a specific response triggered by particular stimuli that were perceived as acoustically anomalous or emotionally ambiguous. These stimuli likely contain prosodic contours or voice qualities that do not neatly align with the prototypical acoustic profiles of Neutral, Happiness, Sadness, or Anger for our participant pool. This could be due to factors such as the perceived intensity of the emotion (e.g., a happiness stimulus that sounded more like euphoria or hysteria), the presence of mixed emotional cues, or even idiosyncratic production features from the speaker that made the intended emotion less clear.

3.2 Mean accuracy rates by variety

When looking at the mean accuracy rates by native variety, i.e., BP vs. EP participants, independently of the variety under perception, we may observe that EP participants exhibited a slightly higher overall accuracy (74.1%) compared to BP participants (69.3%), suggesting that EP participants are better than BP ones in perceiving emotions. This seems to contradict our Hypothesis 3, according to which BP participants—due to Brazil's collectivist cultural orientation and a more expressive prosodic system—would outperform EP participants in recognizing emotions. Instead, the data show that EP participants achieved higher overall accuracy, suggesting that factors beyond cultural dimensions may be influencing emotion perception. One possible explanation is that EP participants, despite their prosodic restraint, may rely more consistently on specific acoustic cues, leading to more stable interpretations across stimuli. Alternatively, the stimuli themselves—particularly those produced in EP—may contain prosodic features that are less universally interpretable, thereby affecting BP participants' performance when perceiving EP stimuli. Another plausible factor is the high degree of exposure Portuguese citizens have to Brazilian Portuguese through cultural and media

TABLE 3 Confusion matrix showing how each intended emotion was perceived by participants.

Intended/perceived	Neutral	Happiness	Sadness	Anger	Other
Neutral	79.5	6.3	8.4	3.6	2.2
Happiness	8.2	64.6	4.3	**18.1**	4.8
Sadness	9.5	5.7	74.1	6.5	4.2
Anger	4.1	**16.2**	6.9	68.5	4.3

The bold values highlight the highest percentages of misclassifications, thus signaling the intended emotions that are more difficult to identify and how they were perceived.

TABLE 4 Accuracy perception of native and non-native stimulus.

Participant variety	Native stimulus (%)	Non-native stimulus (%)
BP listeners	72.8	55.4
EP listeners	66.67	57.8

channels. The presence of a large Brazilian community in Portugal, along with the widespread consumption of Brazilian soap operas, music, literature, and cinema (or even, more recently, YouTubers' contents), may contribute to a greater familiarity with BP prosodic patterns. This frequent contact could facilitate perceptual adaptation, thus confirming Hypothesis 4, allowing EP participants to more effectively decode emotional cues in BP speech, despite the prosodic differences between the varieties.

When inspecting the accuracy rates of BP and EP listeners considering the variety perceived, the results (illustrated in Table 4) reveal a clear trend: both BP and EP participants demonstrated higher accuracy when perceiving emotional stimuli produced in their own native variety. BP listeners achieved **72.8%** accuracy with BP stimuli, compared to **55.4%** with EP stimuli. Similarly, EP listeners reached **66.67%** accuracy with EP stimuli, vs. **57.8%** with BP stimuli. These findings suggest that native prosodic patterns facilitate emotion recognition, likely due to greater familiarity with the intonational contours, rhythm, and expressive norms of one's own variety. This aligns with Hypothesis 1, which posit that BP and EP participants perceive emotions differently, and supports the idea that intonational familiarity enhances perceptual accuracy.

To further explore how emotional perception varies across Portuguese varieties, Figure 2 breaks down accuracy rates by emotion, highlighting patterns in listeners performance considering their native variety.

Chi-square tests were conducted to determine if the accuracy of participants' responses was significantly related to (i) their native variety (BP vs. EP), (ii) the variety being perceived (native vs. non-native), and (iii) the type of emotion they were exposed to (neutral, happiness, sadness, and anger).

Our results show that participants' correct responses were significantly related to their native variety [$\chi^2(1) = 4.32, p < 0.05$], and to the emotion under perception [$\chi^2(3) = 93.87, p < 0.001$]. As we may observe in Figure 2, for both BP and EP listeners, neutral and sadness yielded the highest accuracy. Happiness and anger showed lower accuracy and higher confusion. This partially contradicts Hypothesis 2, which predicted that, influenced by their prosodic profiles, BP listeners would identify better high-arousal emotions whereas EP listeners would exhibit higher accuracy rates

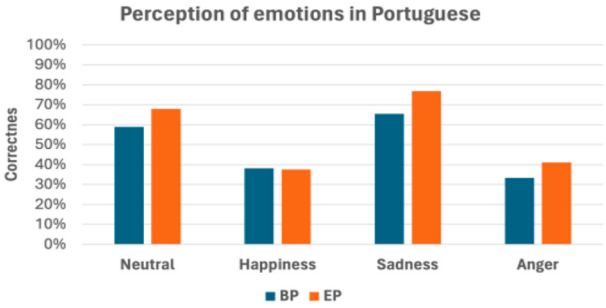

FIGURE 2
Perception of emotions between Portuguese varieties.

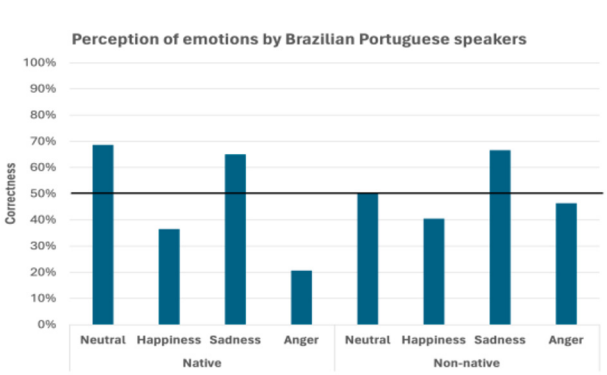

FIGURE 3
Perception of emotions by Brazilian Portuguese participants, produced by speakers from their native (BP) and non-native (EP) varieties.

in low-arousal emotions. Thus, Hypothesis 2 is only confirmed for EP participants.

The variety being perceived (native vs. non-native) did not play a statistically significant role in identification accuracy ($p > 0.05$). However, a detailed observation of the data suggests a different behavior between BP (Figure 3) and EP (Figure 4) participants.

EP participants demonstrated a better performance than BP participants in identifying emotions produced by non-native speakers, thus confirming Hypothesis 4, which predicted that EP participants would outperform BP participants in recognizing emotions produced in the non-native variety due to their higher exposure to BP. However, prosody also seems to play a crucial role: intonation in BP is more chanted than in EP (Frota and Vigário, 2000; Frota et al., 2015a), thus more salient. This also explains why

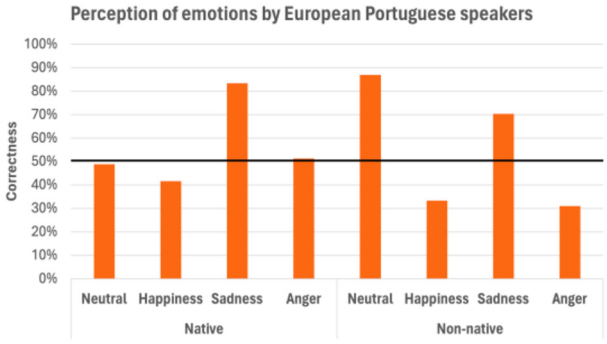

FIGURE 4
Perception of emotions by European Portuguese participants, produced by speakers from their native (EP) and non-native (BP) varieties.

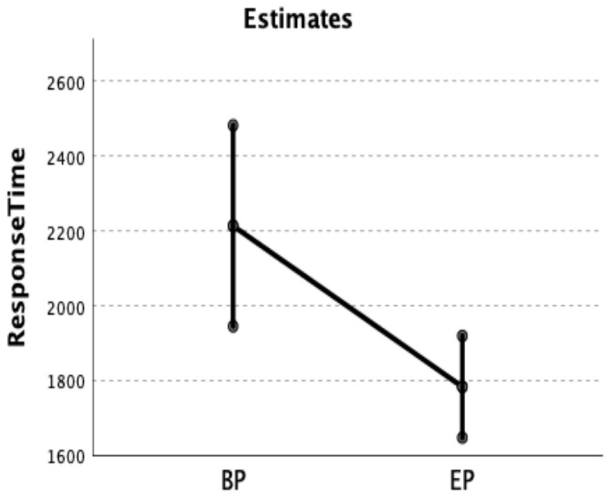

FIGURE 5
RTs considering the native variety of the participants.

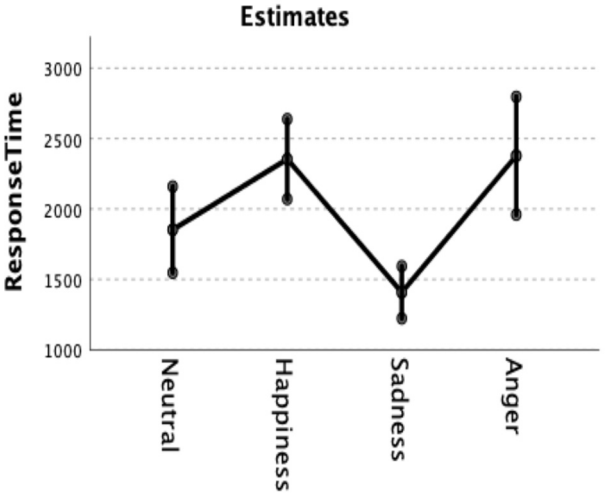

FIGURE 6
RTs considering each emotion under perception.

BP participants are better than the EP ones in identifying emotions produced by speakers from their native variety.

3.3 Reaction times

To investigate whether participants' reaction times (RTs) were influenced by linguistic background and stimulus characteristics, we conducted a Generalized Linear Mixed Model (GLMM). The model included RTs as the dependent variable and four fixed effects: (i) participants' native variety (BP vs. EP), (ii) the perceived variety (native vs. non-native), (iii) the type of emotion (neutral, happiness, sadness, and anger), and (iv) correctness of responses (correct vs. incorrect). In addition, interactions between correctness and emotion, as well as their modulation by native and perceived variety, were also tested.

Our results showed a significant effect of participants' native variety [F (1, 1300) = 8.41, $p < 0.01$], correctness [F (1, 1300) = 6.34, $p < 0.05$], and emotion under perception [F (3, 1300) = 13.10, $p < 0.001$]. The interaction correctness*emotion under perception shows a borderline effect [F (3, 1300) = 2.49, $p = 0.059$], and like for accuracy in the identification of emotions, the variety being perceived does not play a relevant role ($p > 0.05$) for RTs.

Considering firstly the effect of participants' native variety, indeed we may observe, in Figure 5, that, overall, EP participants responded faster (M = 1,784 ms) than the BP ones (M = 2,213 ms).

Considering now the effect of the emotion under perception, and as illustrated in Figure 6, sadness triggers the shortest RTs (M = 1,408 ms), whereas happiness and anger trigger the longest RTs (Ms = 2,354 and 2,379 ms, respectively).

These results are aligned with the accuracy rates observed in section 3.1, i.e., sadness exhibits the second highest accuracy rate (after the neutral emotion) and it is the fastest to be identified, and happiness and anger display the lowest accuracy rates and the longest RTs, thus being the most difficult ones to be perceived.

Looking now at the borderline effect of the interaction correctness*emotion under perception, illustrated in Figure 7, we may observe that the longest RTs were registered for incorrect responses given for happiness and anger, thus reinforcing the conclusion that these two emotions are the most difficult ones to perceive.

In the same line of thought, but in an opposite way, the shortest RTs were registered for correct responses given for sadness, thus highlighting the conclusion that this is the easiest emotion to be perceived. Overall, correct responses were produced faster (M = 1,799 ms) than incorrect ones (M = 2,197 ms), as expected (Schneider et al., 2011).

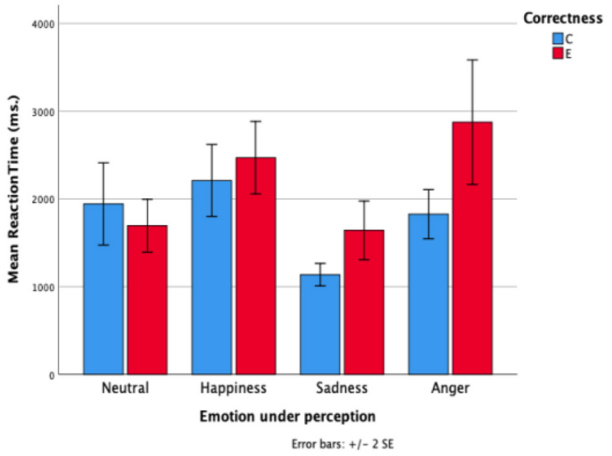

FIGURE 7
RTs considering the correctness of participants' responses for each emotion under perception.

3.4 Acoustic/prosodic properties predicting the perception results

A multinomial logistic regression was run with correctness as dependent variable, and min F0, max F0, mean F0, mean intensity, speaking rate, and dominant pattern as covariates. Since we also wanted to observe whether these predictors are influenced by the participants' native variety, the perceived variety or the emotion under perception, these three independent variables were also included as factors in the same analysis, and each of them was also included in the model in two-way interactions with each covariate.

The chosen method of regression was forward entry, meaning that an initial model is defined, containing only the constant (b0), and that then the software looks for the predictor that best predicts the outcome variable; in other words, it selects, from the whole list of inserted covariates and factors, the one that has the highest correlation with participants' correctness.

As a result, the model selected the interaction between the emotion under perception and the dominant pattern as the best predictor of participants' correctness, as illustrated in Table 5.

Thus, this means that, independently of the participants' native variety and of the variety under perception, participants' correctness is explained by the dominant pattern per emotion, i.e., although stimuli were segmentally masked, the intonational properties in the signal differ per emotion and this is a cue used by participants in the identification of emotions.

As we may observe in Table 6 the predicted values are not significantly different from the observed values. Thus, the model is a good fit of the data.

Indeed, when revisiting Table 1, we may observe that neutral and sadness emotions display the simplest contours (i.e., all-falling melodies with simple boundary tones), whereas happiness and anger display complex boundary tones (in BP) or more prominent nuclear pitch accents (in EP – H*+L), which are more marked in the EP intonational system (Frota, 2012), thus, probably more difficult to be identified.

4 Discussion

The primary objective of this investigation was to determine whether speakers of two closely related yet prosodically distinct varieties of Portuguese—BP and EP—exhibit divergent patterns in their perception of acted emotional prosody. A foundational question underpinning this research was whether the well-documented production differences between these varieties, such as BP's broader F0 range and more frequent use of rising contours compared to EP's more restrained and flat melodic patterns (Frota and Vigário, 2000; Frota et al., 2015a), would lead to corresponding differences in perceptual processing. The analysis was structured around two key metrics: the accuracy of emotion identification and the reaction times (RTs) associated with these judgements, which together provide a view of the underlying perceptual mechanisms. The findings reveal a complex interplay between native variety, exposure, and cognitive processing that shapes the perception of emotional intent.

The present investigation confirms that prosody is a robust carrier of emotional information, with an overall identification accuracy of 71.7%. This result aligns with previous studies on emotional prosody across languages (Pell et al., 2009), demonstrating that listeners can effectively decode emotional states from suprasegmental cues even when semantic information is masked.

The highest accuracy rates were observed for neutral (79.5%) and sadness (74.1%). According to the literature, these emotions display more stable and prototypical acoustic profiles—characterized by a narrower F0 range, slower speech rate, reduced pitch variability, and stable vocal qualities—findings that are supported by Banse and Scherer (1996), for languages such as English, or Colamarco and Moraes (2008) and Nunes et al. (2010) for BP and EP, respectively. These acoustic properties make neutral and sadness emotions less susceptible to misinterpretation. Conversely, happiness (64.6%) and anger (68.5%) showed the lowest accuracy. As discussed in the previous section, these two emotions are more complex from a melodic point of view frequently confused with one another (cf. results in section 3), and often perceived as ambiguous due to their prosodic characteristics—complex boundary tones in BP or marked nuclear accents in EP (Frota, 2012), confirming they are the most challenging emotions to perceive based on prosody alone. The confusion between these two emotions (18.1% of happiness stimuli perceived as anger, and 16.2% of anger as happiness) suggests they share key acoustic features, such as high pitch and increased intensity (Banse and Scherer, 1996; Colamarco and Moraes, 2008), which listeners struggle to disambiguate without other contextual cues.

This hierarchy of recognition, in which neutral and sadness are identified more accurately than happiness and anger, is consistent with cross-linguistic evidence. For example, Pell et al. (2009) found that across multiple languages (including English, German, and Mandarin), low-arousal emotions such as sadness and neutral states tend to yield higher recognition rates, while high-arousal emotions like happiness and anger are more prone to confusion. Similarly, Sauter et al. (2010) reported that the recognition of sadness and neutral affect showed greater cross-cultural stability compared to more dynamic emotions such as happiness, anger, or

TABLE 5 Summary of the multinomial regression model that predicts participants' correctness.

Step summary								
Model	Action	Effect(s)	Model fitting criteria		Effect selection tests			
			AIC	BIC	−2 log likelihood	Chi-square[a]	df	Sig.
0	Entered	Intercept, Min F0, Max F0, Mean F0, Mean intensity, Speaking rate, Dominant pattern, Native variety, Emotion under Perception, Perceived variety	220.161	282.602	196.161			
1	Entered	Emotion under perception*dominant pattern	170.456	243.304	142.456	53.704	2	0.000
Stepwise method: forward entry								

[a]The chi-square for entry is based on the likelihood ratio test.

TABLE 6 Chi-square goodness-of-fit test results for the regression model.

Goodness-of-fit			
	Chi-square	df	Sig.
Pearson	12.959	18	0.794
Deviance	12.979	18	0.793

fear. These findings suggest that certain acoustic cues associated with low-arousal emotions—such as slower tempo, reduced pitch range, and falling intonational patterns—are more universally interpretable, whereas high-arousal emotions share overlapping prosodic markers (e.g., higher F0 range and intensity), which increases ambiguity. By aligning these findings with the prosodic profile of BP and EP varieties of Portuguese, we hypothesized that BP's prosodic profile may enhance the perceptual salience of high-arousal emotions (such as happiness and anger), leading listeners to an easy and fast recognition of these emotions produced by BP speakers, and that, conversely, EP's prosodic profile, being characterized by a more restrained melody (Frota and Vigário, 2001), aligns more closely with low-arousal (such as sadness) or neutral emotional states, resulting in an easy and fast recognition of these emotions produced by EP speakers (Hypothesis 2). We observed that BP participants did not recognize happiness and anger more accurately, whereas EP participants performed better with neutral and sadness, thus partially confirming Hypothesis 2 (for EP participants only). The present results suggest that emotion perception is shaped not only by variety-specific intonational systems but also by the emotion type (i.e., its acoustic properties), as the relative ease of identifying emotions like sadness and neutrality seems to be related with their more stable and prototypical acoustic profiles.

The analysis of the "Other" category provided a crucial insight into the limits of prosodic categorization. Contrary to being a random "I don't know" response, its use was concentrated on a specific subset of stimuli: EP-produced happiness tokens. This pattern was consistent across listeners from both varieties, indicating that the issue lay not with the listeners' strategies but with the stimuli themselves. This concentration suggests that these specific EP happiness productions contained highly atypical or ambiguous acoustic properties[5]—potentially an extreme F0 range, a particular voice quality, or a complex contour that deviated from a prototypical "happiness" affect for our participants. When faced with a stimulus that did not cleanly match any of the four target categories, listeners systematically defaulted to the "Other" label rather than forcing a likely incorrect choice. This finding underscores the importance of considering intra-variety production variability in perceptual studies and highlights that certain emotional expressions can fall outside common perceptual categories, creating islands of ambiguity even within a familiar variety.

Although descriptively both groups were more accurate with stimuli from their own variety (BP: 72.8% native vs. 55.4% non-native; EP: 66.67% native vs. 57.8% non-native), this trend was not statistically significant, as the perceived variety factor did not reach significance ($p > 0.05$). Instead, accuracy varied with listeners' native variety (EP > BP) and with the emotion perceived, with neutral and sadness outperforming happiness and anger. It is important to note that, because the perceived variety was not a significant predictor of accuracy, our results cannot robustly support the classic native-language advantage effect described by Scherer et al. (2001) and Pell et al. (2009) as a definitive finding. Instead, the data may suggest a tendency in that direction, but any such effect in this experiment was not strong enough to be separated from random variance and was secondary to the significant effects of participant variety and emotion type.

A key finding of this study was the significant effect of participants' native variety on emotion recognition accuracy. Closer inspection of this effect, however, revealed a striking asymmetry that ran contrary to our initial predictions: EP participants demonstrated a significantly higher overall accuracy (74.1%) compared to BP participants (69.3%). This result, termed here the EP Superiority Effect, directly contradicts Hypothesis 3, which predicted that BP listeners, hypothetically aided by a more

5 As pointed out by one of the reviewers, there are several expressive strategies behind a given emotion. Thus, besides valence and arousal, dominance might also play a relevant role to the perception of emotions, being a possible explain for some of the responses of the "Other" type. This is left for future research, as this category of responses was not deeply analyzed here.

expressive prosodic system and cultural factors, would outperform EP listeners. This counterintuitive finding can be explained by a confluence of factors centered on asymmetrical exposure and its perceptual consequences. The primary explanation lies in the well-documented, unidirectional nature of media and cultural flow between Portugal and Brazil. EP speakers are extensively exposed to BP through a constant flow of media (Brazilian soap operas, music, movies, and digital content) and the presence of a remarkable Brazilian community in Portugal. This frequent and naturalistic contact constitutes a form of implicit perceptual training, thus aligning with Hypothesis 4. It likely enhances EP listeners' flexibility, allowing them to develop a broader and more adaptive "perceptual map" that can accommodate the wider F0 ranges and more dynamic contours characteristic of BP emotional prosody. This asymmetry as a contact effect finds parallels in other linguistic contexts. For instance, studies have shown that Dutch listeners often outperform German listeners in perceiving German emotional prosody, an advantage attributed to the Netherlands' greater exposure to German media (van Bezooijen and Gooskens, 1999). This suggests a clear case of asymmetrical intelligibility, where the variety with greater cultural penetration (BP) becomes more intelligible to speakers of the other variety (EP) than vice versa. This asymmetrical exposure likely does more than simply familiarize EP listeners with BP sounds; it may actively enhance their perceptual flexibility and cue-weighting strategies. Constant exposure to the more variable BP prosody could train EP listeners to attend to a broader set of acoustic parameters and to be more tolerant of acoustic deviation from their own variety's prototypes. This is akin to the perceptual benefits observed in bilinguals or musicians, where experience with multiple sound systems enhances auditory cognitive function (Neumann et al., 2024; Varnet et al., 2015). Consequently, EP listeners may develop a superior ability to deal with the acoustic variation in BP speech and focus on the core prosodic gestalt, leading to more efficient classification.

Beyond mere exposure, another factor may underpin the EP advantage. Namely, the concept of sociophonetic markedness (Kerswill and Williams, 2002; Trudgill, 1986) may be at play. The more expressive and variable nature of BP prosody, while highly effective within its own communicative context, may sometimes be perceived as acoustically "exaggerated" or less prototypical of a canonical emotional expression by listeners less familiar with it. This potential deviation from expected prosodic norms could hinder accurate recognition for BP stimuli when processed by EP listeners who, despite their exposure, may still hold EP productions as a subconscious baseline.

In summary, the EP Superiority Effect is likely not a reflection of an inherent perceptual deficit in BP listeners, but rather the outcome of an asymmetrical sociolinguistic environment that has trained EP listeners to navigate a wider range of prosodic variation, combined with differences in the inherent variability and cue weighting of the two systems themselves.

The Reaction Times (RT) data provide a complementary window into the underlying perceptual and cognitive processes, offering more than just a mirror of the accuracy findings but a deeper explanation for them. The robust pattern observed reinforces and enriches the conclusions drawn from accuracy alone and demonstrates a key universal aspect of emotional speech processing. The finding that EP participants responded faster overall than BP participants further reinforces the asymmetry observed in accuracy, providing additional evidence for an EP perceptual advantage. In line with the dual-process models of perception proposal (Schirmer and Kotz, 2006; Kahneman, 2011), this significant difference in processing speed suggests that for EP listeners, the task involved a more efficient and automatic processing of the emotional cues, likely facilitated by their extensive exposure to both varieties. In contrast, the longer RTs for BP listeners point to a higher cognitive load and a more effortful decoding process. Another point to consider is that the greater acoustic variability in BP productions might inherently increase the cognitive load for all listeners, including native BP speakers. The wider range of possible realizations for a single emotion category could require more complex cue integration and decision-making processes. This is consistent with our data showing that BP participants not only achieved lower overall accuracy but also exhibited significantly longer reaction times, suggesting a more effortful and less automatic perceptual process compared to their EP counterparts.

The emotion-specific RTs perfectly corroborate the accuracy data. This convergence of high accuracy with low RTs for sadness, and low accuracy with high RTs for happiness and anger, is not an isolated phenomenon but a well-established cross-linguistic pattern observed in the perception of emotional prosody across diverse languages (Liu and Pell, 2012; Pell and Skorup, 2008). This consistency reinforces the universality of the underlying cognitive mechanisms. This pattern can be explained through the lens of processing load and cue distinctiveness. Low-arousal emotions like sadness and neutral are typically cued by a highly stereotypical and acoustically distinct prosodic profile (e.g., low pitch, slow tempo, and falling contours). These canonical markers are consistently realized across speakers and varieties, allowing for rapid feature detection and classification with minimal cognitive effort. Conversely, the perception of high-arousal emotions like happiness and anger is inherently more cognitively demanding. Their acoustic signatures show significant overlap (e.g., high pitch, high intensity), creating direct competition during categorization and forcing the listener's cognitive system to resolve fine-grained cue differences. This ambiguity, combined with the potential for greater internal variability in their productions, directly increases decision time, resulting in the protracted RTs observed.

From an evolutionary perspective, this perceptual asymmetry aligns with the proposed communicative functions of different emotional states. Lower-arousal emotions such as sadness and neutrality are often acoustically marked by reductions in pitch dynamicity, energy, and speech rate (Juslin and Laukka, 2003; Scherer, 1986). These acoustic profiles are not only highly stable but also suggest a signal of lack of immediate threat or action, making them arguably easier to distinguish for a receiver. In contrast, high-arousal emotions like happiness and anger share a suite of urgency-related acoustic cues—including high pitch, increased intensity, and faster tempo—that evolved to rapidly capture attention and signal a need for immediate social response (Darwin, 1872). While highly effective as an alerting mechanism, this acoustic similarity in high-arousal states (Banse and Scherer, 1996; Sauter et al., 2010) can blur the distinctions between specific positive and negative valences, leading to the higher confusion rates observed in our study.

The analysis demonstrated that listeners' perception of emotion was not random but systematically guided by the intonational configurations most closely associated with each category. While Neutral and Sadness benefited from clear, stereotypical cues, the more complex melodic shapes of happiness and anger, particularly in EP, increased ambiguity and decision times. Additionally, the multinomial logistic regression analysis from section 3.4 yielded a crucial and clarifying finding: the interaction between Emotion and Dominant Intonational Pattern was the single best predictor of participants' accuracy. This result confirms that listeners were not guessing but were actively relying on the systematic prosodic gestalts—the specific phonological contours—embedded in the signal to make their decisions, even with all segmental information removed. This finding powerfully underscores that intonational structure is not merely an epiphenomenon of emotion but a core perceptual vehicle for its communication.

The findings of this study collectively argue for a model of emotional prosody perception that integrates universal biological influences with variety-specific phonological structuring. While the recognition hierarchy (e.g., sadness/neutral > happiness/anger) and the cognitive mechanisms (longer RTs for ambiguous stimuli) show cross-linguistic commonality, the precise instantiation of these processes is filtered through the listener's native intonational grammar. We propose that perception is not a direct mapping from acoustic cues to emotion, but a two-stage process: first, the auditory system parses the continuous acoustic signal into discrete, language-specific intonational categories (e.g., H+L, L+H). Second, these categorized phonological units are mapped onto emotional meanings, a process heavily influenced by the frequency and conventionality of these mappings within the variety. This explains the native variety advantage (familiarity with the system), the EP superiority effect (asymmetrical acquisition of a second prosodic system), and the supreme predictive power of the dominant intonational pattern. This model positions emotional prosody not as an exception to linguistic relativity but as a prime example of it.

This two-stage model finds strong support in and extends existing theoretical frameworks. The first stage aligns with the concept of "categorical perception" in prosody, as discussed by Ladd (2008) and others, which argues that listeners perceive intonational contours not as acoustic continua but as members of discrete, phonologically defined categories. Our data robustly confirm this; listeners did not respond to raw F0 or duration values in isolation but to the holistic pattern they formed. The work of Frota et al. (2015a,b) is paramount here, as their development of P_ToBI provides the precise phonological inventory—the set of possible categories like H+L or L+H—that BP and EP listeners use to parse the signal. The regression analysis proves that the accuracy of emotion identification is contingent on how well a stimulus instantiates one of these expected categories. A stimulus that is acoustically ambiguous between categories, like some of the EP happiness tokens that were frequently labeled "Other," leads to perceptual hesitation or failure, precisely because it cannot be cleanly categorized in the first stage of processing.

The second stage of our model, the mapping of phonological categories onto emotional meaning, is crucially mediated by the listener's linguistic and cultural experience. This directly refines Scherer's (2003) Brunswikian Lens Model. While Scherer posits a direct link between acoustic cues (proximal stimuli) and inferred emotion, our results demonstrate that this link is indirect and is gated by the phonological system. This variety-specific conventionalization echoes the findings of Prieto and Roseano (2010) in their cross-linguistic studies on question intonation, showing that the pragmatic meaning of tunes is language-specific. Our study demonstrates that this principle applies to the emotional domain.

5 Conclusion

This study set out to investigate the perception of emotional prosody across two major varieties of Portuguese, Brazilian (BP) and European (EP), by examining how their distinct intonational systems and sociocultural contexts shape emotion recognition. It is relevant to clarify that acted (or simulated) emotions were used as stimuli, and that our findings should be interpreted accordingly, as it is known that authentic and acted vocal emotion expressions exhibit acoustic differences (e.g., Jurgens et al., 2011). With this caveat in mind, our findings reveal a complex perceptual landscape governed by both universal principles and variety-specific conditioning.

At a universal level, we identified a robust emotion hierarchy: neutrality and sadness were recognized with significantly higher accuracy and speed than happiness and anger. This pattern, consistent with cross-linguistic evidence, underscores the role of stable, prototypical acoustic profiles in facilitating the rapid decoding of low-arousal states, while the shared high-arousal cues of happiness and anger create inherent perceptual ambiguity, leading to confusion and slower reaction times.

However, these universal tendencies were powerfully filtered through the lens of linguistic and cultural experience. Contrary to our initial hypotheses, the predicted native-variety advantage, while descriptively present, was not the dominant statistical effect. More strikingly, we observed an "EP Superiority Effect," where European Portuguese listeners outperformed their Brazilian counterparts in both overall accuracy and processing speed. This counterintuitive finding is best explained by the profound impact of asymmetrical cultural exposure. The extensive and unidirectional flow of Brazilian media into Portugal has equipped EP listeners with greater perceptual flexibility, allowing them to navigate the broader prosodic variability of BP more effectively than BP listeners can decode the more restrained patterns of EP. This highlights that exposure can fine-tune the human auditory cognitive system, enhancing the ability to interpret emotional signals from a different, albeit related, prosodic dialect.

Most critically, our results demonstrate that emotional prosody perception is a fundamentally phonological process. The finding that the interaction between emotion and its dominant intonational pattern was the single best predictor of accuracy confirms that listeners do not decode raw acoustics, but rather rely on categorized, language-specific prosodic gestalts. This led us to propose a two-stage model of perception: the continuous acoustic signal is first parsed into discrete intonational categories native to

the listener's variety, and these phonological units are then mapped onto emotional meanings.

In conclusion, this research bridges evolutionary emotion theory with prosodic typology, demonstrating that the perception of vocal emotions is a sophisticated psycholinguistic act. While universal acoustic and cognitive mechanisms provide a foundation, their interpretation is deeply constrained by the phonological grammar of the listener's native variety and fine-tuned by cultural and exposure factors. The results challenge purely universalist accounts of emotional prosody and underscore the necessity of integrating linguistic specificity into both theoretical models of affective communication and applied technologies such as cross-cultural speech emotion recognition systems. Future research should explore these dynamics using neuroimaging techniques to elucidate the neural correlates of this proposed two-stage model and incorporate spontaneous speech to further validate these findings in ecologically rich contexts.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

Ethics statement

The studies involving humans were approved by Comissão de Ética para a Investigação (Faculdade de Letras da Universidade de Lisboa). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

DS: Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft. MC: Conceptualization, Formal analysis, Funding acquisition, Methodology, Supervision, Writing – review & editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This research was funded by Fundação para a Ciência e a Teconologia (UID/214/2025, https://doi.org/10.54499/UID/00214/2025).

Acknowledgments

The authors would like to thank the participants and the actresses who recorded the stimuli, as well as the Phonetics and Phonology Lab where stimuli recordings took place. The comments and suggestions made by the reviewers are also deeply acknowledged.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

References

Banse, R., and Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *J. Pers. Soc. Psychol.* 70, 614–636. doi: 10.1037/0022-3514.70.3.614

Bänziger, T., and Scherer, K. R. (2005). The role of intonation in emotional expressions. *Speech Commun.* 46, 252–267. doi: 10.1016/j.specom.2005.02.016

Barbosa, P. A. (2006). *Incursões em torno do ritmo da fala* [Explorations around the Rhythm of Speech]. Campinas, SP: Pontes Editores.

Boersma, P., and Weenink, D. (2022). *Praat: Doing Phonetics by Computer* (Version 6.3.01) [Computer Software]. Available online at: http://www.praat.org/ (Accessed June 6, 2022).

Castro, S. L., and Lima, C. F. (2010). Recognizing emotions in spoken language: a validated set of Portuguese sentences and pseudosentences for research on emotional prosody. *Behav. Res. Methods* 42, 74–81. doi: 10.3758/BRM.42.1.74

Colamarco, M., and Moraes, J. A. (2008). Emotion expression in speech acts in Brazilian Portuguese: production and perception. *Proc. Speech Prosody* 2008, 717–720. doi: 10.21437/SpeechProsody.2008-159

Darwin, C. (1872). *The Expression of the Emotions in Man and Animals*. London: John Murray. doi: 10.1037/10001-000

Ekman, P. (1992). An argument for basic emotions. *Cogn. Emot.* 6, 169–200. doi: 10.1080/02699939208411068

Ekman, P., and Friesen, W. V. (1969). The repertoire of nonverbal behavior: categories, origins, usage, and coding. *Semiotica* 1, 49–98. doi: 10.1515/semi.1969.1.1.49

Ekman, P., and Friesen, W. V. (1971). Constants across cultures in the face and emotion. *J. Pers. Soc. Psychol.* 17, 124–129. doi: 10.1037/h0030377

Frota, S. (2012). "The intonational phonology of European Portuguese," in *Intonation in Romance*, eds S. Frota and P. Prieto (Oxford: Oxford University Press), 6–42. doi: 10.1093/acprof:oso/9780199567300.003.0002

Frota, S., Cruz, M., Svartman, F., Collischonn, G., Fonseca, A., Serra, C., et al. (2015a). "Intonational variation in Portuguese: European and Brazilian varieties," in *Intonation in Romance*, eds S. Frota and P. Prieto (Oxford: Oxford University Press), 235–283. doi: 10.1093/acprof:oso/9780199685332.003.0007

Frota, S., and Moraes, J. A. (2016). "Intonation in European and Brazilian Portuguese," in *The Handbook of Portuguese Linguistics*, eds W. L. Wetzels, J. Costa, and S. Menuzzi (Hoboken, NJ: Wiley Blackwell), 141–166. doi: 10.1002/9781118791844.ch9

Frota, S., Oliveira, P., Cruz, M., and Vigário, M. (2015b). *P-ToBI: Tools for the Transcription of Portuguese Prosody*. Lisboa: Laboratório de Fonética, CLUL/FLUL. Available online at: http://labfon.letras.ulisboa.pt/InAPoP/P-ToBI/.

Frota, S., and Vigário, M. (2000). "Aspetos de prosódia comparada: ritmo e entoação no PE e no PB [Aspects of comparative prosody: rhythm and intonation in EP and BP]," in *Actas do XV Encontro Nacional da Associação Portuguesa de Linguística, vol. 1*, eds. V. R. Castro P. Barbosa (Coimbra: APL), 533–555.

Frota, S., and Vigário, M. (2001). On the correlates of rhythmic distinctions: the European/Brazilian Portuguese case. *Probus* 13, 247–276. doi: 10.1515/prbs.2001.005

Hofstede, G. (2001). *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations across Nations*, 2nd Edn. Thousand Oaks, CA: Sage.

Jurgens, R., Hammerschmidt, K., and Fisher, J. (2011). Authentic and play-acted vocal emotion expressions reveal acoustic differences. *Front. Psychol.* 2:180. doi: 10.3389/fpsyg.2011.00180

Juslin, P. N., and Laukka, P. (2003). Communication of emotions in vocal expression and music performance: different channels, same code? *Psychol. Bull.* 129, 770–814. doi: 10.1037/0033-2909.129.5.770

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York, NY: Farrar, Straus and Giroux.

Kerswill, P., and Williams, A. (2002). ""Salience" as an explanatory factor in language change: evidence from dialect levelling in urban England," in *Language Change: The Interplay of Internal, External and Extra-Linguistic Factors*, eds M. C. Jones and E. Esch (Berlin, Germany: De Gruyter Mouton), 81–110. doi: 10.1515/9783110892598.81

Ladd, D. R. (2008). *Intonational Phonology*, 2nd Edn. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511808814

Lado, R. (1957). *Linguistics across Culture: Applied Linguistics for Teachers*. Ann Arbor, MI: University of Michigan Press.

Laukka, P., and Elfenbein, H. A. (2021). Cross-cultural emotion recognition and in-group advantage in vocal expression: a meta-analysis. *Emot. Rev.* 13, 3–11. doi: 10.1177/1754073919897295

Liu, P., and Pell, M. D. (2012). Recognizing vocal emotions in Mandarin Chinese: a validated database of Chinese vocal emotional stimuli. *Behav. Res. Methods* 44, 1042–1051. doi: 10.3758/s13428-012-0203-3

Matsumoto, D. (1990). Cultural similarities and differences in display rules. *Motiv. Emot.* 14, 195–214. doi: 10.1007/BF00995569

Neumann, C., Sares, A., Chelini, E., and Deroche, M. (2024). Roles of bilingualism and musicianship in resisting semantic or prosodic interference while recognizing emotion in sentences. *Bilingual. Lang. Cogn.* 27, 419–433. doi: 10.1017/S1366728923000573

Nunes, A., Coimbra, R.L., Teixeira, A. (2010). "Voice Quality of European Portuguese Emotional Speech," in *Computational Processing of the Portuguese Language. PROPOR 2010. Lecture Notes in Computer Science, vol 6001*, eds. T. A. S. Pardo, A. Branco, A. Klautau, R. Vieira, V. L. S. de Lima (Berlin; Heidelberg: Springer), 142–151. doi: 10.1007/978-3-642-12320-7_19

Pell, M. D., Monetta, L., Paulmann, S., and Kotz, S. A. (2009). Recognizing emotions in a foreign language. *J. Acoust. Soc. America* 125, 468–478. doi: 10.1007/s10919-008-0065-7

Pell, M. D., and Skorup, V. (2008). Implicit processing of emotional prosody in a foreign versus native language. *Speech Commun.* 50, 519–530. doi: 10.1016/j.specom.2008.03.006

Plutchik, R. (1982). A psychoevolutionary theory of emotions. *Soc. Sci. Inform.* 21, 529–553. doi: 10.1177/053901882021004003

Prieto, P., and Roseano, P. (eds.). (2010). *Transcription of Intonation of the Spanish Language*. Lincom Germany: Europa.

Sauter, D. A., Eisner, F., Ekman, P., and Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proc. Nat. Acad. Sci.* 107, 2408–2412. doi: 10.1073/pnas.0908239106

Scherer, K. R. (1979). "Nonlinguistic vocal indicators of emotion and psychopathology," in *Emotions in Personality and Psychopathology. Emotions, Personality, and Psychotherapy*, eds C. E. Izard (Boston, MA: Springer), 493–529. doi: 10.1007/978-1-4613-2892-6_18

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychol. Bull.* 99, 143–165. doi: 10.1037/0033-2909.99.2.143

Scherer, K. R. (2003). Vocal communication of emotion: a review of research paradigms. *Speech Commun.* 40, 227–256. doi: 10.1016/S0167-6393(02)00084-5

Scherer, K. R., Banse, R., and Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *J. Cross-Cult. Psychol.* 32, 76–92. doi: 10.1177/0022022101032001009

Schirmer, A., and Kotz, S. A. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends Cogn. Sci.* 10, 24–30. doi: 10.1016/j.tics.2005.11.009

Schneider, K., Dogil, G., and Möbius, B. (2011) Reaction time and decision difficulty in the perception of intonation. *Proc. Interspeech* 2011, 2221–2224. doi: 10.21437/Interspeech.2011-581

Trudgill, P. (1986). *Dialects in Contact*. Oxford: Basil Blackwell.

van Bezooijen, R., and Gooskens, C. (1999). Identification of language varieties: the contribution of different linguistic levels. *J. Lang. Soc. Psychol.* 18, 31–48. doi: 10.1177/0261927X99018001003

Varnet, L., Wang, T., Peter, C., Meunier, F., and Hoen, M. (2015). How musical expertise shapes speech perception: evidence from auditory classification images. *Sci. Rep.* 5:14489. doi: 10.1038/srep14489

Vigário, M. (2003). *The Prosodic Word in European Portuguese*. Berlin; New York, NY: Mouton de Gruyter. doi: 10.1515/9783110900927

Wang, T., Lee, Y.-C., and Ma, Q. (2018). Within and across-language comparison of vocal emotions in Mandarin and English. *Appl. Sci.* 8:2629. doi: 10.3390/app8122629