

Alexandra Christine de Aguiar¹ 
Ana Carolina Constantini² 
Ronei Marcos de Moraes³ 
Anna Alice Almeida⁴ 

Keywords

Voice
Emotion
Speech Acoustics
Prosody
Emotion Recognition in Voice

Descritores

Voz
Emoção
Acústica da Fala
Prosódia
Reconhecimento da Emoção na Voz

Correspondence address:

Anna Alice Almeida
Departamento de Fonoaudiologia,
Centro de Ciências da Saúde,
Universidade Federal da Paraíba –
UFPB
Cidade Universitária, Campus I,
Castelo Branco, João Pessoa (PB),
Brasil, CEP: 58051-900.
E-mail: anna_alice@uol.com.br

Received: April 28, 2024

Accepted: December 02, 2024

Editor: Vanessa Veis Ribeiro.

Acoustic-prosodic measures discriminate the emotions of Brazilian portuguese speakers

Medidas acústico-prosódicas discriminam as emoções de falantes do português brasileiro

ABSTRACT

Purpose: To verify if there is a difference in acoustic-prosodic measures in different emotional states of speakers of Brazilian Portuguese (BP). **Methods:** The data sample consisted of 182 audio signals produced by actors (professionals or students), from the semi-spontaneous speech task “Look at the blue plane” in the various emotions (joy, sadness, fear, anger, surprise, disgust) and neutral emission. Values were extracted from acoustic-prosodic measures of duration, fundamental frequency and intensity of the various emotions. The Friedman comparison test was used to verify whether these measures are able to discriminate emotions. **Results:** The prosodic-acoustic analysis revealed significant variations between emotions. The disgust emotion stood out for having the highest rate of utterance, with higher values of duration. In contrast, the joy exhibited a more accelerated speech, with lower values of duration and greater intensity. Sadness and fear were marked by lower intensity and lower frequencies, and fear presented the lowest positive asymmetry values of z-score and z-smoothed, with less elongation of the segments. Anger was highlighted by the higher vocal intensity, while surprise recorded the highest values of fundamental frequency. **Conclusion:** The acoustic-prosodic measures proved to be effective tools for differentiating emotions in CP speakers. These parameters have great potential to discern different emotional states, broaden knowledge about vocal expressiveness and open possibilities for emotion recognition technologies with applications in artificial intelligence and mental health.

RESUMO

Objetivo: Verificar se há diferença de medidas acústico-prosódicas em diferentes estados emocionais de falantes do português brasileiro (PB). **Métodos:** A amostra de dados consistiu em 182 sinais de áudio produzidos por atores (profissionais ou estudantes), a partir da tarefa de fala semi-espontânea “Olha lá o avião azul” nas variadas emoções (alegria, tristeza, medo, raiva, surpresa, nojo) e emissão neutra. Foram extraídos valores das medidas acústico-prosódicas de duração, frequência fundamental e intensidade das variadas emoções. Utilizou-se o teste de comparação de Friedman para verificar se essas medidas são capazes de discriminar as emoções. **Resultados:** A análise acústico-prosódica revelou variações significativas entre as emoções. A emoção nojo destacou-se por apresentar a maior taxa de elocução, com valores mais altos de duração. Em contraste, a alegria exibiu uma fala mais acelerada, com menores valores de duração e maior intensidade. A tristeza e o medo foram marcados por menor intensidade e frequências mais baixas, sendo que o medo apresentou os menores valores de assimetria positiva de z-score e z-suavizado, com menor alongamento dos segmentos. A raiva se sobressaiu pela maior intensidade vocal, enquanto a surpresa registrou os valores mais altos de frequência fundamental. **Conclusão:** As medidas acústico-prosódicas demonstraram ser ferramentas eficazes para diferenciar emoções em falantes do PB. Esses parâmetros têm grande potencial para discernir diferentes estados emocionais, ampliam o conhecimento sobre a expressividade vocal e abrem possibilidades para tecnologias de reconhecimento de emoções, com aplicações em inteligência artificial e saúde mental.

Study conducted at Universidade Federal da Paraíba – UFPB - João Pessoa (PB), Brasil.

¹ Programa de Pós-graduação em Modelos de Decisão e Saúde, Universidade Federal da Paraíba – UFPB - João Pessoa (PB), Brasil.

² Departamento de Desenvolvimento Humano e Reabilitação, Universidade Estadual de Campinas – UNICAMP - Campinas (SP), Brasil.

³ Departamento de Estatística, Universidade Federal da Paraíba – UFPB - João Pessoa (PB), Brasil.

⁴ Departamento de Fonoaudiologia, Universidade Federal da Paraíba – UFPB - João Pessoa (PB), Brasil.

Financial support: National Council for Scientific and Technological Development (CNPq). Process nº 434508/2018-7.

Conflict of interests: nothing to declare.

Data Availability: Research data is available in the body of the article.



This is an Open Access article distributed under the terms of the Creative Commons Attribution license (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

The interface between voice and language is related to the point of intersection between vocal production and linguistic expression. It involves the connection between the ability to produce vocal sounds and the ability to use these sounds in a structured and meaningful way, following specific linguistic rules⁽¹⁾. The voice is used to convey language, that is, to express words and meanings according to the rules and linguistic conventions. It can differentiate the emotional state expressing personality traits, feelings, physical and mental health status, among others⁽²⁻⁴⁾.

Language influences the way we use our voice. The grammatical structure, vocabulary and linguistic patterns determine how we organize and express our ideas from the voice⁽⁵⁾. Therefore, voice and language are intrinsically related, and their interaction is critical for effective communication and emotional expression.

Several studies have contributed to the fact that some authors proposed that six basic emotions (happiness/joy, fear, anger, sadness, disgust/disgust, surprise) should be universally recognized in the face by human beings, for presenting specific configurations, expressed in a similar way in different cultures^(6,7). These emotions when combined generate a spectrum of emotional states.

Each emotional state causes changes in the vocal tract and momentarily alter the physiology of voice production, which interferes with breathing control, vertical positioning of the larynx, in the relative relaxation of vocal folds and in the positioning and relaxation of pharyngeal and tongue muscles, which may result in voice modification⁽⁸⁾. These variations of the human voice, when an individual experiences a certain emotional state, can involve both aspects of vocal quality and related attributes of prosody⁽⁹⁾. This can be defined as a set of speech properties, which is beyond the segment level, and is usually studied from the analysis of three classical phonetic-acoustic parameters: duration, fundamental frequency (f_0) and intensity⁽⁹⁻¹¹⁾. For some authors, prosody results from the coupling of syntactic, semantic and discursive information and the constraints of a speech production system^(12,13). The prosodic variations transmit information to the significance of expressiveness and mark the characteristics of a vocal dynamic, but little is known about this information in emotions for Brazilian Portuguese (BP).

Some voice banks were developed with the variation of emotions in different languages/cultures, such as the Berlin Database of Emotional Speech (EMO-DB)⁽¹⁴⁾ Interactive Emotional Dyadic Motion Capture (IEMOCAP)⁽¹⁵⁾ Sustained Emotionally colored Machine-human Interaction using Nonverbal Expression (SEMAINE)⁽¹⁶⁾ and Remote Collaborative and Affective interactions (RECOLA)⁽¹⁷⁾. The main characteristics that stand out in the differentiation of emotions in these data sets are the acoustic and prosodic characteristics, such as pitch, energy and duration. In addition, cepstral coefficients such as the Mel-Frequency Cepstral Coefficients (MFCC) are important measures to optimize emotional identification⁽¹⁸⁾.

Recently, a voice bank in the various emotions was developed from native speakers of BP, the EMOVOX-BR⁽¹⁹⁾. There was the validation of this bank through the perceptive-auditory judgment of expert judges. This study indicated that acoustic

aspects, such as pitch and loudness variation, were essential for the differentiation of emotions, in addition to their valence and power⁽¹⁹⁾.

The analysis of acoustic-prosodic aspects of emotions is an area of growing interest in speech and communication sciences, offering new perspectives on how emotions are expressed and perceived through voice⁽²⁰⁻²²⁾, and therefore the present study seeks to collaborate with the identification of vocal parameters and speech specific to each emotional state, as well as recognize the voice as a biological signal rich in information that can be instrumental in detecting emotional patterns.

The investigation of emotional prosody involves both how acoustic modulations are produced and how they are perceived, providing a broader understanding of the relationship between linguistic and emotional processes in human communication⁽²³⁾, going beyond the semantic content of words. Understanding these elements is essential to unravel the complexity of human communication, in which the emotional dimension significantly influences social interaction. In addition to its theoretical impact, the study of emotional prosody presents practical applications in advanced technologies for emotion recognition, artificial intelligence, and clinical diagnoses of communication disorders and mental health^(21,22).

These findings have the potential to boost the development of human-machine interaction systems and automatic voice emotion detection, since prosody involves elements that carry important information about the emotional state of the speaker⁽²³⁾. The expression of emotions affects these parameters in a distinct way, allowing machine learning models to use these variations to identify and classify emotions accurately⁽²⁴⁾. This ability is essential to create systems capable of adjusting their responses according to the emotional state of the user, promoting more human and efficient interactions.

These advances have fundamental applications for a wide range of industries such as call centers, voice recognition apps, web movies and mobile communication. It is believed that the analysis of prosodic-acoustic parameters could reveal phonic differences between the emotional variations of BP, contributing to the development of systems of synthesis and speech recognition more adapted to the local language and culture.

Given the above, questions arise: Is it possible to discriminate emotions from acoustic-prosodic measures in BP speakers? What prosodic aspects characterize the different emotional states? Are there differences in duration, frequency and intensity between emotions? Thus, the objective of this study was to verify if there is difference in acoustic-prosodic measures in different emotional states of BP speakers.

METHODS

This is an observational and cross-sectional research, evaluated and approved by the Research Ethics Committee of the Health Sciences Center of a higher education institution in Brazil, under number 3.304.419/2018.

The data set for analysis was composed of 182 sound signals, produced by 26 professional actors and Brazilian students, of both sexes, with an average age of 27 years, living in the states

of Paraíba, São Paulo, Rio Grande do Sul, Ceará, Roraima, Mato Grosso and the Federal District, which belong to the Brazilian Voice Bank in the Variations of Emotions - EMOVOX-BR⁽¹⁹⁾.

For the construction of EMOVOX-BR⁽¹⁹⁾ voice samples were recorded from native speakers of BP expressing different emotions, such as joy, sadness, fear, anger, surprise, disgust and neutral emission. Participants received detailed instructions on the recording procedure and then performed voice collection. Three different speech tasks were recorded: 1) extended vowel emission /ε/, 2) automatic counting from 1 to 10 and 3) semi-spontaneous speech of the phrase “Look at the blue plane.”, which is part of the Consensus Auditory Perceptual Evaluation of Voice - CAPE-V⁽²⁵⁾. Each participant performed these tasks in the six basic emotions. Which generated a total of 1,638 sound signals.

For validation, we chose to use the phrase “Look at the blue plane.” collected through smartphone. All audios were submitted to the signal-to-noise ratio (SNR) analysis and obtained values higher than the reference standard, that is, the value of SNR equal or greater than 30dB. We selected 182 audio signals for the stage of perceptive-auditory judgment. These vocal samples were evaluated by speech-language judges, who obtained high levels of precision in the identification of the six basic emotions.

This study analyzed the acoustic-prosodic measures of the same 182 audios used in the validation of EMOVOX-BR, which was performed from the perceptive-auditory judgment by expert judges. While the validation study of EMOVOX-BR emphasized the precision of human perception to identify emotions, this work deepens in the quantitative analysis of emotional prosody, mainly in the behavior of acoustic parameters of duration, f_0 and intensity. These parameters are considered the most robust elements for speaker discrimination and offer a technical and measurable perspective of the emotional aspects of voice^(10,13).

The duration of an emotion can vary from short moments to long periods, depending on the intensity of the emotional stimulus, the person's ability to regulate emotions and the context in which the emotion occurs⁽²⁰⁾. Duration is essential to identify variations in the speech cadence, enunciation structure and temporal organization of phrases⁽²⁶⁾.

For duration analysis, all speech samples were manually segmented into Vowel-Vowel (unit VV) units, syllable-sized units that comprise a segment from the beginning of a vowel to the beginning of the immediately following vowel, including the consonants between them⁽²⁷⁾.

The speech task used was segmented into four parts, being: [Al] [auav] [iaNU] [az]. To extract this measure, considering the calculation of the normalized duration of VV units, we used the script SG Detector⁽²⁸⁾. The script presents a reference table with averages and standard deviations of the phonic segments for the BP to calculate the duration value, z-score and z-score smoothed units VV throughout the statement, which generates a segmentation of the phrasal groups of statements. Segmentation is done by calculating the standard deviation of the duration averages of VV units, which are normalized by the z-score calculation.

The z-score value indicates the number of standard deviations from the average of an information point that in this case would be the variations of the BP, that is, it is a proportion of the number of standard deviations below or above the BP, which means a

gross score. A z-score is called standard score and may well be placed on a common scatter curve and extends from - 3 to + 3 deviations⁽²⁹⁾.

The z-score values softened allow to attenuate local variations of duration due to the fall of duration in units VV post-tonic and/or duration of headphones very different from the durations ratio of the BP⁽²⁷⁾. This value corresponds to the smoothed of five points applied to the z-score data sequence, which allows observing more precisely the duration prominences. The aim is to verify how much the duration values obtained in the corpus of EMOVOX-BR⁽¹⁹⁾ varied according to the intrinsic durations table for the phones of the BP.

When considering the number of speech segment units VV divided by the total sum of their duration, we obtain the rate of elocution⁽³⁰⁾. It influences the perception of speech rhythm, with slower rates associated with syllabation, elongation of final sounds and pauses, while faster rates tend to reduce these phenomena⁽²⁸⁾. The relationship between the duration of the emission and the rate of utterance was explored to observe the variation in speech speed according to different emotions.

The f_0 , measured in Hertz (Hz), is defined by the number of vibrations produced by vocal folds per second and is directly related to pitch perception⁽³¹⁾. This parameter allows a detailed analysis of the intonation and tonal variation along utterances, providing information about vocal control and quality⁽³²⁾. For the analysis of f_0 were measured the mean, maximum, minimum, standard deviation (f_0 sd) and variation (f_0 range) in each statement.

Intensity, measured in decibels (dB), is associated with the strength or degree of emotional activation experienced by the speaker and is related to the vocal energy employed during speech⁽³¹⁾. Its measurement helps to understand the energy of the emission and the modulation of force throughout the speech, contributing to studies of prosody and vocal expressiveness⁽¹³⁾. For the intensity analysis, the mean, minimum and maximum values were extracted for each signal and, subsequently, the data were compared to each emotion. The f_0 and intensity measurements were extracted with the help of the VoxMore plug-in⁽³³⁾.

This study investigates whether the acoustic-prosodic measures of duration, f_0 and intensity vary significantly between different emotions such as joy, sadness, fear, anger, surprise, disgust and neutral emission in BP speakers. Based on the literature, which demonstrates that emotions influence acoustic parameters in a distinct way⁽²⁰⁻²²⁾, the objective is to test these variations using a set of data from native speakers.

The PRAAT program, version 5.4.04, was used for data extraction. The Friedman test with post hoc of Dunn was selected to verify whether these differences in acoustic parameters between emotions were statistically significant, contributing to a better understanding of the relationship between the acoustic-prosodic aspects and emotional expression, after the results extracted and identified the groups that differed, a descriptive analysis of the variables was performed to identify which prosodic parameter and emotion that stood out in the group. All analyses were performed using the software Statistical Package for Social Sciences (SPSS), version 24, and the significance level of 5% was used.

RESULTS

The prosody analysis included data extracted according to segmented VV units, the contrast of the variations of f_0 and intensity after analysis.

There was a significant difference in the duration of segments according to emotion (Table 1). In relation to the total duration of the emission, the highest value was found in the disgust emotion (340.63 ms), which presented the highest rate of elocution. On the other hand, joy emotion had the lowest duration value (264.73 ms), corresponding to the lowest rate of elocution, which indicates a direct relationship between the total time of emission and the speed of speech in the different emotions.

The longest duration in all four parts of the VV units was again found in the disgust emotion when analyzing the duration of each segment individually (A1 - 220.88 ms; auav - 605.15 ms; iaNU - 239.92 ms; az - 296.58 ms). The shorter duration per segment significantly differentiated all emotions from each other. The shorter duration of the first segment was observed in fear emotion (A1 - 156.08 ms), that of the second segment was shown in joy (auav - 437.71 ms), that of the third in surprise (iaNU - 198.88 ms) and that of the last segment in sadness (az - 224.19 ms). There was no significant difference in the third segment [iaNU] in the varied emotions in all parameters related to duration (Table 1).

There was a significant difference when comparing the z-score and z-smoothed values of the first, second and last segment of the z-score in the various emotions (Table 2).

Table 2 presents the disgust emotion with the highest positive asymmetry values of z-score and z-smoothed. It was found that fear emotion has the lowest positive asymmetry values among all emotions. It is possible to observe negative values of z-score and z-smoothed in the third segment [iaNU] in the emotions joy, fear, sadness, surprise, anger and neutral emission. Already in the disgust emotion there was an elongation of the segments compared to the reference values (Table 2).

There was significance in the comparison of all acoustic measures related to voice frequency in the various emotions analyzed. It was found that the highest average f_0 was in surprise emotion (284.43 Hz), followed by joy (268.54 Hz), and the lowest in sadness (160.91 Hz) and neutral (163.44 Hz). The highest maximum frequency was presented in surprise emotion (360.36 Hz) and the lowest minimum frequency was marked by sadness (77.91 Hz). The fear emotion had the highest f_0 range (178.11 Hz) and the lowest in the sadness emotion (116.73 Hz) (Table 3).

There were comparisons of the independent samples of mean, minimum and maximum intensity significant in the various emotions (Table 4). The highest peak intensity was in anger (82.35 dB), followed by joy (66.79 dB), and the lowest in fear (35.5 dB) and sadness (36.78 dB). There was a statistically significant difference in the comparison of the average intensity, with the highest recorded intensity in the emotion of joy (66.79 dB) and the lowest in the emotion of sadness (60.23 dB).

Chart 1 shows a synthesis of the prosodic variations present in the various emotions, within the context of BP speakers, that is, the characteristics that most mark each emotion and differentiate them from the others.

Table 1. Comparison of duration parameters for each VV unit of the statement “Look at the blue plane”, produced in different emotions

Duration (ms)	Emotion							p-value
	Joy	Fear	Sadness	Anger	Surprise	Disgust	Neuter	
Mean	264.73	266.81	284.83	270.32	276.7	340.63	274.83	0.024*
A1	172.92	156.08	170.73	175.04	174.81	220.88	168.5	0.001*
Auav	437.73	481.81	528.46	466.27	469.04	605.15	495.69	0.066*
iaNU	212.46	204.27	215.92	205.23	198.88	239.92	206.77	0.317
az	238.81	225.08	224.19	234.73	264.08	296.58	228.35	0.000*

Caption: Friedman test; *Significant values ($p < 0.05$); ms = milliseconds

Table 2. Comparison of z-score and smoothed z-score parameters for each VV unit of the statement “Look at the blue plane”, produced in the various emotions

		Emotion							P-value
		Joy	Fear	Sadness	Anger	Surprise	Disgust	Neuter	
Z-score	Mean	0.95	0.74	1.23	0.98	1.10	2.74	0.99	0.014*
	A1	3.37	2.52	3.26	3.48	3.47	5.79	3.15	0.024*
	auav	0.22	0.85	1.54	0.62	0.98	2.67	1.06	0.079*
	iaNU	-0.63	-0.99	-0.48	-0.95	-1.22	0.56	-0.88	0.317
	az	0.83	0.55	0.56	0.75	1.19	1.95	0.63	0.001*
Smoothed Z-score	Mean	0.94	0.78	1.28	1.00	1.13	2.8	1.03	0.015*
	A1	2.32	1.97	2.69	2.53	2.64	4.75	2.45	0.002*
	auav	1.27	1.41	2.11	1.57	1.81	3.71	1.75	0.028*
	iaNU	-0.17	-0.31	0.13	-0.29	-0.3	1.26	-0.19	0.174
	az	0.34	0.05	0.20	0.19	0.39	1.49	0.13	0.055*

Caption: Friedman test; Significant values ($p < 0.05$)

Table 3. Comparison of f_0 values for the various emotions (average, minimum, maximum and range)

f_0 (Hz)	Emotion							P-value
	Joy	Fear	Sadness	Anger	Surprise	Disgust	Neuter	
Mean	268.54	228.29	160.91	213.23	284.43	188.47	163.44	0.000*
min	166.42	102.05	77.91	137.10	183.61	135.15	80.91	0.000*
max	344.10	281.86	203.80	287.79	360.36	269.94	194.65	0.000*
range	174.67	178.11	116.73	145.79	176.75	134.79	122.88	0.000*

Caption: Friedman test; *Significant values ($p < 0.05$); f_0 = fundamental frequency; Hz = hertz

Table 4. Comparison of intensity values for the various emotions (average, minimum and maximum).

Intensity	Emotion							P-value
	Joy	Fear	Sadness	Anger	Surprise	Disgust	Neuter	
Mean	66.79	63.8	60.23	65.58	65.08	61.91	61.28	0.001*
min	55.29	35.5	36.78	54.87	47.86	43.12	41.89	0.001*
max	80.11	79.64	77.02	82.35	79.72	77.77	79.8	0.001*

Caption: Friedman test; *Significant values ($p < 0.05$); dB = decibel

Chart 1. Prosodic variations in different emotions in Brazilian Portuguese

Emotions	Acoustic-prosodic measures		
	Duration	Frequency	Intensity
Joy	↓	↑	↑
Anger	ND	ND	↑
Surprise	ND	↑	ND
Sadness	ND	↓	↓
Fear	↓	ND	↓
Disgust	↑	ND	ND
Neuter	ND	↓	ND

Caption: ↑ = higher; ↓ = less; ND= not discriminated

DISCUSSION

The characteristics of the acoustic-prosodic parameters present variations that reflect on the expression of emotions. The knowledge of these measures contributes to the construction of the definition of common emotional variations in vocal signals, can assist in clinical diagnostic aspects and favor the creation of models for recognition of emotions from the voice⁽³⁴⁾.

Several banks of voices that incorporate emotional variation were developed, covering populations of actors in different languages and cultures⁽¹⁴⁻¹⁷⁾. However, these banks of voices usually focus only on the analysis of traditional acoustic variations and little is known about other types of measures, such as: deepening in the acoustic-prosodic and perceptual on the part of speech therapists and in the impact of judgment by hearing experts, as well as the possibility of finding recognition of voice patterns common to each of the emotions.

The inclusion of prosodic measures in the analysis of emotional voice variation in a data set validated by expert judges is fundamental for a comprehensive and accurate analysis of human emotional expression⁽³⁵⁾. This increases the validity of data and improves the ability to generalize to different contexts and populations, as well as improving the effectiveness and robustness of automatic emotion recognition models. Thus, it can contribute to a more holistic understanding of human

emotional processes and how they are expressed and perceived through speech⁽³⁶⁾.

Prosody has its dimension based on the suprasegmental aspects of speech, which relate to variations in duration, frequency and intensity^(10,37). The present study sought to explore whether there are differences in prosodic variations in different emotions from the voice, the data reported in this study confirmed that these parameters are important factors to define emotions through the objective analysis of vocal signals. Of the nineteen parameters that presented significant results, twelve concern the rhythmic structure of the varieties studied, four refer to f_0 and three to intensity, which shows that the three classical parameters of prosody study present distinctions between the various emotions.

There was a difference between the duration measures in the speech task in the various emotions. The disgust presented higher values in the average duration, as well as in all segmentations of VV units, that is, it was the emotion with the highest rate of speech elocution. As for the shorter average duration, the joy emotion presented lower values, so it is the emotion with more accelerated speech speed. Changes in speech speed alter the phonetic characteristics of the signal, thus becoming a parameter capable of differentiating emotions⁽³⁸⁾.

When analyzing the parameter of shorter duration per segment, it was observed that each segment presented shorter duration in

different emotions, varied between fear, joy, surprise and sadness. Speakers of a language do not always express emotions in the same way, with the same levels of activation⁽³⁹⁾. Each emotional state can be defined as a linear combination of some axes, such as activation (or excitement) and power (or power). Activation measures the individual's degree of excitement in expressing emotion. Potency refers to the strength of emotion⁽⁴⁰⁾. Then, each of the basic emotions can generate different levels of activation and power in the speaker, depending on its manifestation can cause differentiation in speech speed.

For the data obtained from z-score and z-smoothed, significant differences were observed in relation to the speech variations of BP between emotions. The disgust emotion presented higher values of positive asymmetry. This fact indicates that speakers increase the duration of some VV units during the emission of this emotion. It means that in the disgust there were more elongations in the duration of the units, which results in the higher positive value asymmetry, and thus relates to a more slowed speech⁽²⁷⁾. Fear was the emotion with lower values of positive asymmetry among all emotions, that is, there are less elongations in the duration of units.

As for the negative values of z-score and z-smoothed found in the third segment [iaNU] in the emotions joy, fear, sadness, anger, surprise and neutral emission. These values below the reference standards mean that there was a shortening in the duration of the segments for these emotions⁽²⁷⁾. The acceleration of the segment [iaNU] was a strategy used in the emission of emotions, so it did not present a significant difference in the comparison. In the case of emotion disgust, there was an elongation of the segments compared to the reference values, thus confirming the results found for the elocution rate.

Studies developed in different languages reveal that the faster utterance is associated with emotions of greater excitement. On the other hand, a slower speech is usually linked to states of reflection or calm. In tonal languages, such as Mandarin, the rate of utterance not only affects prosody, but can also alter the meaning of words, being crucial for the correct interpretation of the speaker's intentions⁽⁴¹⁾. Cultural and regional variations play an important role in the modulation of speech speed, besides indicating emotions, the rate of elocution may reflect specific cultural and linguistic norms⁽⁴¹⁻⁴³⁾. Therefore, duration analysis offers a deeper understanding of emotional and communicative patterns, varying according to language and cultural context.

It is found that the average and maximum values of f_0 allow to distinguish the surprise emotion and joy, which positions them between the upper bands of f_0 . These emotions are characterized by a greater tonal variation, typical of emotional states of higher excitement or positive⁽⁴⁴⁾, as opposed to sadness and neutral, which have low minimum and average values of f_0 . Joy is a positive valence emotion and surprise is a bivalent emotion. The f_0 characteristics discriminate emotions in the valence dimension with greater precision⁽⁴⁵⁾. For EMOVOX-BR⁽¹⁹⁾, the expert judges indicated that the surprise of this bank was a positive valence emotion. This fact justifies, therefore, the emotions of positive valence have as characteristic variation of pitch in ascending curve. The f_0 is directly related to laryngeal function, therefore, changes in intonation and airflow produced

by the vocal tract due to emotional state can be identified in the acoustic-prosodic analysis of the vocal signal⁽⁴⁶⁾. The variation of f_0 is important in fear, surprise and joy, because they move away from sadness, disgust and neutral.

The variation in pitch along the speech was essential to identify emotions of different valences. The intonation pattern reflects subtle emotional changes that would not be detected only by intensity or duration. Studies in emotional prosody performed in different languages often include intonation to identify complex emotional states⁽⁴¹⁻⁴³⁾. Therefore, the means and variation of f_0 were important to differentiate emotions.

The intensity parameter is related to the amplitude of the sound wave. The highest average intensity was recorded in joy emotion and the lowest in sadness emotion. The maximum intensity recorded was in anger, then joy and the smallest in fear and sadness. Vocal intensity is one of the main parameters that guide listeners in classification⁽⁴⁷⁾. According to the expert judges' assessment for the construction of EMOVOX-BR⁽¹⁹⁾, anger was the emotion with the highest percentage of success in identifying emotions and with the lowest percentage of success the fear emotion. According to the literature, anger emotion generates a greater impact on the identification of the emotions of the interlocutor, because for its production higher levels of energy are used, and it is related to changes in larynx positioning, speech speed and intensity^(8,48).

Previous studies have shown significant variations in parameters such as f_0 , intensity and duration, which are modulated differently according to the type of emotion expressed^(19,49). However, these variations can be influenced by factors such as the methodology adopted, the language spoken and even the cultural context of the participants⁽¹⁹⁾. The lack of standardization in the analysis protocols between studies, both in terms of equipment used, speech task used, and analysis techniques, hinders the understanding of emotional prosody in different contexts.

In this sense, it would be interesting for future studies to adopt a greater uniformity in their approaches, using emotion databases validated by specialized judges with sensitive speech and vocal measures to differentiate emotions⁽⁵⁰⁾. In addition, the inclusion of machine learning models, which have been successfully applied in emotional prosody studies, can contribute to identify more consistent and universal patterns in acoustic characteristics related to emotions. This standardization would allow a more direct comparison between studies, as well as increasing the applicability of results in areas such as artificial intelligence and automatic emotion recognition⁽⁴⁹⁾.

In general, the suprasegmental aspects of speech, such as temporal (duration) and dynamic characteristics (intensity and f_0) play a crucial role in the differentiation of basic emotions. It is observed that the prosodic characteristics highlight each emotion clearly. Finally, it can be inferred that, through the analysis of acoustic-prosodic signals, it is possible to identify emotional variations in native speakers of BP. These findings broaden the understanding of emotional communication and offer new perspectives for the development of future research and technological applications, with emphasis on areas such as automatic recognition of emotions and clinical interventions. The evidence obtained reinforces the role of prosody as an essential tool in understanding emotional dynamics in human communication.

CONCLUSION

It is possible to discriminate emotions from acoustic-prosodic measures in BP speakers. The acoustic-prosodic measures of f_0 , duration and intensity are sensitive to differentiate the various emotions.

The disgust is the one that best differed from the other emotions with higher rate of elocution, longer duration in all segments analyzed. Joy has a lower rate of utterance and higher average intensity. The fear emotion is the emotion with greater variability of f_0 , as well as presents lower values of stretches in the duration of the units. The emotion sadness is the emotion with lower values of average of f_0 , variability of f_0 and intensity. The anger emotion presents greater energy in production, with maximum intensity recorded. The surprise is the emotion with higher average of f_0 and with record of higher maximum frequency.

REFERENCES

- González Torre I, Luque B, Lacasa L, Luque J, Hernández-Fernández A. Emergence of linguistic laws in human voice. *Sci Rep*. 2017;7(1):43862. <http://doi.org/10.1038/srep43862>. PMID:28272418.
- Costa DB, Lopes LW, Silva EG, Cunha GMS, Almeida LNA, Almeida AAF. Fatores de risco e emocionais na voz de professores com e sem queixas vocais. *Rev CEFAC*. 2013;15(4):1001-10. <http://doi.org/10.1590/S1516-18462013000400030>.
- Cowen AS, Elfenbein HA, Laukka P, Keltner D. Mapping 24 emotions conveyed by brief human vocalization. *Am Psychol*. 2019;74(6):698-712. <http://doi.org/10.1037/amp0000399>. PMID:30570267.
- Barbosa IK, Behlau M, Lima-Silva MF, Almeida LN, Farias H, Almeida AA. Voice symptoms, perceived voice control, and common mental disorders in elementary school teachers. *J Voice*. 2021;35(1):158.e1-7. <http://doi.org/10.1016/j.jvoice.2019.07.018>. PMID:31416748.
- Alves CRST, Mastella V. Linguagem e comunicação na contemporaneidade. Cruz Alta: Ilustração; 2020.
- Ekman P. An argument for basic emotions. *Cogn Emotion*. 1992;6(3-4):169-200. <http://doi.org/10.1080/02699939208411068>.
- Wang Y, Zhu Z, Chen B, Fang F. Perceptual learning and recognition confusion reveal the underlying relationships among the six basic emotions. *Cogn Emotion*. 2019;33(4):754-67. <http://doi.org/10.1080/02699931.2018.1491831>. PMID:29962270.
- Yao X, Bai W, Ren Y, Liu X, Hui Z. Exploration of glottal characteristics and the vocal folds behavior for the speech under emotion. *Neurocomputing*. 2020;410:328-41. <http://doi.org/10.1016/j.neucom.2020.06.010>.
- Cohen AS, Hong SL, Guevara A. Understanding emotional expression using prosodic analysis of natural speech: refining the methodology. *J Behav Ther Exp Psychiatry*. 2010;41(2):150-7. <http://doi.org/10.1016/j.jbtep.2009.11.008>. PMID:20022000.
- Santos AJ, Rothe-Neves R, Pacheco V, Baldow VS. Emotional speech prosody: how readers of different educational levels process pragmatic aspects of reading aloud. *DELTA*. 2022;38(3):1-31. <https://doi.org/10.1590/1678-460X202258945>.
- Wagner M, Watson DG. Experimental and theoretical advances in prosody: a review. *Lang Cogn Process*. 2010;25(7-9):905-45. <http://doi.org/10.1080/01690961003589492>. PMID:22096264.
- Watson D, Gibson E. The relationship between intonational phrasing and syntactic structure in language production. *Lang Cogn Process*. 2010;25(5):713-55. <http://doi.org/10.1080/01690960444000070>.
- Arvaniti A. The phonetics of prosody. In: Aronoff M, Chen Y, Cutler C, editors. *Oxford research encyclopedia of linguistics*. Oxford: Oxford University Press; 2020. <http://doi.org/10.1093/acrefore/9780199384655.013.411>.
- Burkhardt F, Paeschke A, Rolfes M, Sendmeier W, Weiss B. A database of German emotional speech. In: 9th European Conference on Speech Communication and Technology (INTERSPEECH); 2005 Sep 4-8; Lisbon, Portugal. Proceedings. Los Alamitos, CA: IEEE/ISCA; 2005. p. 1517-20.
- Busso C, Bulut M, Lee CC, Kazemzadeh A, Mower E, Kim S, et al. IEMOCAP: Interactive Emotional Dyadic Motion Capture Database. *Lang Resour Eval*. 2008;42(4):335-59. <http://doi.org/10.1007/s10579-008-9076-6>.
- McKeown G, Valstar M, Cowie R, Pantic M, Schroder M. The SEMAINE database: annotated multimodal records of emotionally colored conversations between a person and a limited agent. *IEEE Trans Affect Comput*. 2012;3(1):5-17. <http://doi.org/10.1109/T-AFFC.2011.20>.
- Ringeval F, Sonderegger A, Sauer J, Lalanne D. Introducing the recola multimodal corpus of remote collaborative and affective interactions. In: 10th IEEE Int Conf Workshops Autom Face Gesture Recognit (FG); 2013; Shanghai, China. Proceedings. New York: IEEE; 2013. p. 1-8. <http://doi.org/10.1109/FG.2013.6553805>.
- Shinde AS, Patil VV. Speech emotion recognition system: a review. In: 4th International Conference on Advances in Science and Technology (ICAST 2021); 2021; Bahir Dar, Ethiopia. Proceedings. New York: SSRN; 2021. p. 1-6. <http://doi.org/10.2139/ssrn.3869462>.
- Lima HMO, Almeida AAF, Almeida LNA. Elaboração e validação do Banco de Vozes Brasileiro nas Variações das Emoções (EMOVOX-BR). In: 30º Congresso Brasileiro de Fonoaudiologia; 2022; João Pessoa. Anais. São Paulo: Sociedade Brasileira de Fonoaudiologia; 2022. p. 4298-302. (vol. 1).
- Larrouy-Maestri P, Poeppel D, Pell MD. The sound of emotional prosody: Nearly 3 decades of research and future directions. *Perspect Psychol Sci*. 2023 PMID:38232303.
- Oh C, Morris R, Wang X, Raskin MS. Analysis of emotional prosody as a tool for differential diagnosis of cognitive impairments: a pilot research. *Front Psychol*. 2023;14:1129406. <http://doi.org/10.3389/fpsyg.2023.1129406>. PMID:37425151.
- Filippa M, Lima D, Grandjean A, Labbé C, Coll SY, Gentaz E, et al. Emotional prosody recognition enhances and progressively complexifies from childhood to adolescence. *Sci Rep*. 2022;12(1):17144. <http://doi.org/10.1038/s41598-022-21554-0>. PMID:36229474.
- Silva W, Barbosa PA. Perception of emotional prosody: investigating the relation between the discrete and dimensional approaches to emotions. *Rev Estud Linguagem*. 2017;25(3):1075-102. <http://doi.org/10.17851/2237-2083.25.3.1075-1103>.
- Lausen A, Hammerschmidt K. Emotion recognition and confidence ratings predicted by vocal stimulus type and prosodic parameters. *Humanit Soc Sci Commun*. 2020;7(1):2. <http://doi.org/10.1057/s41599-020-0499-z>.
- Behlau M, Rocha B, Englert M, Madazio G. Validation of the Brazilian Portuguese CAPE-V instrument: br CAPE-V for auditory-perceptual analysis. *J Voice*. 2020;36(4):586.e15-20. <http://doi.org/10.1016/j.jvoice.2020.07.007>. PMID:32811691.
- Fox A. Prosody features and prosodic structure. Oxford: Oxford University Press; 2000. <http://doi.org/10.1093/oso/9780198237853.001.0001>.
- Constantini AC, Barbosa PA. Prosodic characteristics of different varieties of Brazilian Portuguese. *Rev Bras Criminol*. 2015;4(3):44-53. <http://doi.org/10.15260/rbc.v4i3.103>.
- Barbosa PA. *Inscursões em torno de ritmo da fala*. Campinas: Editora Pontes; 2006.
- Sterne JA, Kirkwood BR. *Essential medical statistics*. 2nd ed. Hoboken: Oxford Blackwell Science; 2003.
- Costa LMO, Martins-Reis VO, Celeste LC. Metodologias de análise da velocidade de fala: um estudo piloto. *CoDAS*. 2016;28(1):41-5. <http://doi.org/10.1590/2317-1782/20162015039>. PMID:27074188.
- Lopes LW, Alves JN, Evangelista DS, França FP, Vieira VJD, Lima-Silva MFB, et al. Acurácia das medidas acústicas tradicionais e formânticas na avaliação da qualidade vocal. *CoDAS*. 2018;30(5):e20170282. <http://doi.org/10.1590/2317-1782/20182017282>. PMID:30365651.
- Barbosa PA, Madureira S. *Manual de fonética acústica experimental*. São Paulo: Cortez; 2015.

33. Abreu SR, Moraes RM, Martins PN, Lopes LW. VOXMORE: artefato tecnológico para auxiliar a avaliação acústica da voz no processo ensino-aprendizagem e prática clínica. *CoDAS*. 2023;35(6):e20220166. <http://doi.org/10.1590/2317-1782/20232022166en>. PMID:37909540.
34. Silva LJ Jr, Barbosa PA. Speech rhythm of English as L2: an investigation of prosodic variables on the production of Brazilian Portuguese speakers. *J Speech Sci*. 2020;8(2):37-57. <http://doi.org/10.20396/joss.v8i2.14996>.
35. Moriarty P, Vigeant M, Wolf R, Gilmore R, Cole P. Creation and characterization of an emotional speech database. *J Acoust Soc Am*. 2018;143:1869. <http://doi.org/10.1121/1.5036133>.
36. Ekberg M, Stavrinou G, Andin J, Stenfelt S, Dahlström Ö. Acoustic features distinguishing emotions in Swedish speech. *J Voice*. 2023. Ahead of print. <http://doi.org/10.1016/j.jvoice.2023.03.010>. PMID:37045739.
37. Lehiste I. *Suprasegmentals*. Cambridge: MIT Press; 1970.
38. Almeida ANS, Oliveira M Jr, Almeida RAS. A velocidade de fala como pista acústica da emoção básica de raiva. *Rev Diadorim*. 2015;17(2):198-211. <http://doi.org/10.35520/diadorim.2015.v17n2a4076>.
39. Scherer KR. A cross-cultural investigation of emotion inferences from voice and speech: Implications for speech technology. In: 6th ICSLP; 2000; Beijing. Proceedings. Berlin: ISCAArchive; 2000. p. 379-82. <http://doi.org/10.21437/ICSLP.2000-287>.
40. Goudbeek M, Scherer K. Beyond arousal: valence and potency/control cues in the vocal expression of emotion. *J Acoust Soc Am*. 2010;128(3):1322-36. <http://doi.org/10.1121/1.3466853>. PMID:20815467.
41. Liu P, Pell MD. Processing emotional prosody in Mandarin Chinese: a cross-language comparison. In: International Conference on Speech Prosody 2014; 2014; Dublin, Ireland. Proceedings. Berlin: ISCAArchive; 2014. p. 95-9. <http://doi.org/10.21437/SpeechProsody.2014-7>.
42. Nunes VG. Contribuições sobre as características prosódicas de interrogativas totais neutras produzidas por sergipanos. In: Freitag RMK, Lucente L, editores. *Prosódia da fala: pesquisa e ensino*. São Paulo: Blucher; 2017. p. 145-62. <http://doi.org/10.5151/9788580392593-09>.
43. Muñetón-Ayala M, De Vega M, Ochoa-Gómez JF, Beltrán D. The brain dynamics of syllable duration and semantic predictability in Spanish. *Brain Sci*. 2022;12(4):458. <http://doi.org/10.3390/brainsci12040458>. PMID:35447989.
44. Kaur J, Juglan K, Sharma V. Role of acoustic cues in conveying emotion in speech. *J Forensic Sci Crim Invest*. 2018;11(1). <http://doi.org/10.19080/JFSCI.2018.11.555803>.
45. Busso C, Rahman T. Unveiling the acoustic properties that describe the valence dimension. In: Thirteenth Annual Conference of the International Speech Communication Association; 2012; Portland, OR, USA. Proceedings. Berlin: ISCAArchive; 2012. p. 1179-82. <http://doi.org/10.21437/Interspeech.2012-124>.
46. Lopes LW, Cavalcante DP, Costa PO. Intensidade do desvio vocal: integração de dados perceptivo-auditivos e acústicos em pacientes disfônicos. *CoDAS*. 2014;26(5):382-8. <http://doi.org/10.1590/2317-1782/20142013033>. PMID:25388071.
47. Barbosa PA. Aspectos de produção e percepção de estilos de elocução profissionais e não profissionais em quatro línguas. In: Freitag RMK, Lucente L, editores. *Prosódia da fala: pesquisa e ensino*. São Paulo: Blucher; 2017. p. 44-59. <http://doi.org/10.5151/9788580392593-03>.
48. Ververidis D, Kotropoulos C. Emotional speech recognition: resources, features, and methods. *Speech Commun*. 2006;48(9):1162-81. <http://doi.org/10.1016/j.specom.2006.04.003>.
49. Pervaiz M, Khan TA. Emotion recognition from speech using prosodic and linguistic features. *Int J Adv Comput Sci Appl*. 2016;7(8):84-9. <http://doi.org/10.14569/IJACSA.2016.070813>.
50. Swain M, Routray A, Kabisatpathy P. Databases, features and classifiers for speech emotion recognition: a review. *Int J Speech Technol*. 2018;21(1):93-120. <http://doi.org/10.1007/s10772-018-9491-z>.

Author contributions

ACA: conceptualization, methodology, statistical analysis, writing-original draft preparation, writing-review & editing; ACC: data curation, investigation, writing-review & editing; RMM: statistical analysis, writing-original draft preparation, writing-review & editing; AAA: conceptualization, discussion of clinical implications, writing-original draft preparation, writing-review & editing.