

Amanda Feliciano¹ 
Mara Behlau¹ 
Jordana Balbinot¹ 
Samuel Ribeiro de Abreu² 
Vanessa Veis Ribeiro³ 
Leonardo Wanderley Lopes⁴ 

Keywords

Voice
Acoustic
Voice Disorders
Dysphonia
Diagnostic

Descritores

Voz
Acústica
Distúrbios da Voz
Disfonia
Diagnóstico

Correspondence address:

Leonardo Wanderley Lopes
Departamento de Fonoaudiologia,
Centro de Ciências da Saúde,
Universidade Federal da
Paraíba – UFPB
Cidade Universitária - Campus 1,
Bairro Castelo Branco, João Pessoa
(PB), Brasil, CEP: 58051-900.
E-mail: lwlopes@hotmail.com

Received: August 04, 2025

Accepted: December 15, 2025

Editor: Ana Carolina Constantini.

Accuracy and cut-off values of different acoustic measures in the sustained vowel task of Brazilian Portuguese speakers

Acurácia e ponto de corte de diferentes medidas acústicas na tarefa de vogal sustentada em falantes do português brasileiro

ABSTRACT

Purpose: To define cut-off values for acoustic measures extracted from the sustained vowel [a] task to discriminate between the presence and absence of vocal disorders in Brazilian Portuguese speakers. **Methods:** The sample consisted of 376 participants, 288 female and 88 male, with a mean age of 41.20 ± 14.04 years. All individuals underwent a laryngological evaluation and a voice recording session, during which a sustained vowel [a] sample was collected. Subsequently both auditory-perceptual and acoustic analyses of the voice were performed. Among the participants 277 were classified as having a vocal disorder and 99 as having no vocal disorder. A total of 44 acoustic measures were extracted and analyzed using ROC curves to determine the cut-off value for each measure, along with their performance metrics (sensitivity, specificity and AUC). **Results:** Cut-off values were established for 23 of the 44 extracted acoustic measures. Six measures showed robust and clinically relevant performance in discriminating between voices with and without vocal disorders in Brazilian Portuguese speakers. The measures with AUC values greater than 0.70 and balanced sensitivity and specificity above 70% were: Amplitude Variability Index (AVI) (≤ 0.91), Glottal-to-noise excitation (GNE)2000Hz (> 0.92), Spectral Noise Level (SNL)100-2600Hz (≤ 7.71), SNL100-3000Hz (≤ 3.38), SNL100-8000Hz (≤ 10.95) and SNL5100-8000Hz (≤ -1.84). **Conclusion:** Six acoustic measures demonstrated satisfactory performance in identifying the presence of vocal disorders in Brazilian Portuguese speakers. Ten measures showed good performance for voice screening purposes and seven for diagnostic confirmation of vocal disorders.

RESUMO

Objetivo: Definir os valores de corte das medidas acústicas extraídas da tarefa de vogal sustentada [a] para discriminação entre presença e ausência de distúrbio vocal em falantes do português brasileiro. **Método:** A amostra foi composta por 376 participantes, 288 do gênero feminino e 88 do gênero masculino, com idade média de $41,20 \pm 14,04$ anos. Os indivíduos foram submetidos ao exame de avaliação laringológica, e a uma gravação vocal, sendo coletada uma amostra de fala com a vogal [a] sustentada e posteriormente realizado o julgamento perceptivo auditivo e acústico da voz, dos quais 277 foram classificados com distúrbio vocal e 99 sem distúrbio vocal. Foram extraídas 44 medidas acústicas e analisadas por meio da curva ROC para identificação do ponto de corte de cada medida, bem como suas métricas de desempenho. **Resultados:** Foram definidos pontos de corte para 23 das 44 medidas acústicas inicialmente extraídas. Dentre elas, seis medidas demonstraram desempenho robusto e clinicamente relevante para discriminar vozes com e sem distúrbio vocal em falantes do português brasileiro. As medidas com valores de AUC superiores a 0,70 e sensibilidade e especificidade equilibradas acima de 70% foram: Amplitude Variability Index (AVI) ($\leq 0,91$), Glottal-to-noise excitation (GNE)2000Hz ($> 0,92$), Spectral Noise Level (SNL)100-2600Hz ($\leq 7,71$), SNL100-3000Hz ($\leq 3,38$), SNL100-8000Hz ($\leq 10,95$) e SNL5100-8000Hz ($\leq -1,84$). **Conclusão:** Seis medidas apresentaram desempenho satisfatório para identificar a presença de distúrbio vocal em falantes do português brasileiro. Dez medidas apresentaram bom desempenho para triagem vocal e sete para confirmação diagnóstica de distúrbio vocal.

Study conducted at Centro de Estudos da Voz – CEV - São Paulo (SP), Brasil.

¹ Centro de Estudos da Voz – CEV - São Paulo (SP), Brasil.

² Universidade Federal de Rondonópolis – UFR - Rondonópolis (MS), Brasil.

³ Departamento de Fonoaudiologia, Faculdade de Ciências e Tecnologias em Saúde, Universidade de Brasília – UnB – Brasília (DF), Brasil.

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

Financial support: nothing to declare.

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 multidimensional voice assessment procedures include acoustic analysis, auditory-perceptual judgment (APJ), aerodynamic analysis, structural and vibratory examination of the larynx, and vocal self-assessment⁽¹⁾. Acoustic analysis provides measures to quantify the characteristics of the vocal signal and perform a descriptive evaluation of the visual patterns of these signals⁽²⁾. This contributes to screening, diagnostic confirmation, and monitoring the effectiveness of therapeutic interventions in individuals with voice disorders^(3,4).

There are several acoustic measures for assessing the vocal signal. Among the conventional and most widely used measures are fundamental frequency (F0), jitter, shimmer, and harmonic-to-noise ratio (HNR)⁽¹⁾. Conventional measurements are particularly effective in mild vocal deviations. Spectral and cepstral measurements, more widely used and studied in recent years, result from the double transformation of the acoustic voice signal, revealing the degree of harmonic organization in the spectrum. They are effective for analyzing deviated voices, as they are strong predictors of vocal deviation⁽⁵⁾.

The clinical applicability of acoustic measurements depends on the definition of cutoff values, with consistent diagnostic performance appropriate to the target population. Few acoustic measures have defined cutoff values in Brazilian Portuguese (BP), which limits intersubject comparison and comparison of the patient's voice with reference values⁽⁶⁾, in dissonance with well-founded and conscious clinical practice.

Recently, cutoff values were established for two multiparametric indices, Acoustic Voice Quality Index (AVQI) and Acoustic Breathiness Index (ABI), to help identify vocal disorders in BP speakers⁽⁷⁾. With excellent diagnostic accuracy, the AVQI achieved a cutoff point of 1.33, sensitivity of 78.8%, and specificity of 90.6%. In turn, the ABI showed good diagnostic accuracy, with a cutoff point of 2.94, sensitivity of 75.3%, and specificity of 93.4%⁽⁷⁾. This highlights the strength of AVQI and ABI as objective tools for assessing and monitoring vocal disorders^(7,8).

Two other measures have stood out as objective parameters in vocal acoustic analysis, the *Cepstral Peak Prominence* (CPP) and the *Smoothed Cepstral Peak Prominence* (CPPS). They have a good correlation with APJ and high sensitivity in detecting vocal deviations, without depending on the extraction of the fundamental frequency. There are cutoff values to discriminate between voices with and without vocal disorders in BP⁽⁶⁾. The study identified that values below 28.15 decibels (dB) (vowel [ε]), 28.77 dB (vowel [a]) and 28.58 dB (Consensus Auditory-Perceptual Evaluation of Voice - CAPE-V sentence), and that for the CPPS, values below 16.42 dB (vowel [ε]), 17.02 dB (vowel [a]), and 11.30 dB (all CAPE-V sentences) are indicative of a vocal disorder⁽⁶⁾.

The cutoff point is derived from the best relationship between sensitivity and specificity. It is essential to identify measures with balanced performance that simultaneously present satisfactory values of sensitivity and specificity. These measures have the ability to correctly identify both individuals with vocal disorders (true positives) and individuals without the condition

(true negatives), reducing the occurrence of false diagnoses. This balance is generally represented by area under the curve (AUC) values greater than 0.70 in ROC analyses, indicating good discriminative performance⁽¹⁾.

The classification regarding the presence or absence of specific health conditions depends on the accuracy of the measurement or testing procedure. This accuracy reflects the agreement between the findings of a new test, called an index test, and those obtained through the most reliable existing method for detecting the health outcome in question, the reference standard⁽⁹⁾. Therefore, selecting an appropriate reference standard is imperative to validate the accuracy metrics of an emerging diagnostic test. When choosing a reference test, clinicians face challenges in establishing criteria for vocal normality due to its inherently multidimensional nature. Currently, there is no single reference method capable of accurately differentiating vocally healthy individuals from those with a vocal disorder⁽¹⁰⁾. Given the lack of robust scientific evidence and expert consensus on the detection of certain health conditions, alternative reference standards must be developed⁽⁹⁾.

Although APJ and self-assessment are traditionally used as preliminary methods to identify voice disorders⁽¹¹⁾, visual inspection of the larynx (including laryngoscopy and stroboscopy) is considered the gold standard for confirming the diagnostic categorization of voice disorders⁽¹⁰⁾. Synthesizing information from different procedures can construct an integrative reference criterion, derived from the correlation between the results of a visual assessment of the larynx (performed by a physician) and the APJ (performed by a speech-language pathologist), contributing to the evaluation of the discriminative capacity of acoustic measures in distinguishing between individuals with and without vocal disorders.

In clinical practice, different types of tasks can be used to identify vocal disorders. The task selected for acoustic analysis influences its outcome. The vowel [a] is widely used in acoustic and spectrographic vocal analysis due to its stable production and representativeness of vocal resonance. Recent studies indicate that the choice of vowel can significantly influence diagnostic accuracy, with [a] being the most internationally adopted because it facilitates the standardization of procedures and the comparison of results⁽¹²⁾. The predominance of the vowel [a] in acoustic protocols is associated with the ease of obtaining sustained speech samples with minimal articulatory interference, allowing for detailed spectrographic analysis of formants and vocal temporal characteristics⁽¹³⁾. However, the literature lacks certain cutoff values, and there is a need to identify and validate acoustic measures with robust diagnostic performance for the sustained vowel [a] task in BP speakers.

The aim of the present study was to define the cutoff values of acoustic measures extracted from the sustained vowel [a] for discrimination between the presence and absence of vocal disorders in BP speakers.

METHODS

This is a cross-sectional, retrospective, and descriptive study that followed the guidelines of the Standards for Reporting

Diagnostic Accuracy (STARD)⁽¹⁴⁾. The study was approved by the Institutional Research Ethics Committee under the number 02954612.9.0000.5188. All participants signed the Informed Consent Form.

In the study, the index test was the acoustic analysis of the sustained vowel [a], and the gold standard was the presence or absence of a vocal disorder, defined based on a clinical evaluation consisting of a laryngological examination and APJ. Although the index test assessed the voice (acoustic measures), the target condition to be detected was the presence of a vocal disorder. This study used results derived from the agreement between the visual examination of the larynx and the APJ results as the gold standard to determine the clinical outcome: the presence or absence of a vocal disorder.

The study sample was defined based on inclusion and exclusion criteria applied through procedures that ensured the analysis of the reference standard. Adults aged 18 to 65 years who had undergone an otolaryngological evaluation two weeks prior to data collection, with a confirmed diagnosis of vocal disorder, and without prior vocal treatment (therapy or surgery) before data collection were included. Participants excluded were those with cognitive or neurological disorders that prevented the use of recording procedures, professional voice users, and individuals who had already undergone formal vocal therapy or surgery in the head or neck region. In addition, exclusion criteria related to the quality of vocal signals were also applied, excluding voice samples that had a maximum phonation time of less than 5 seconds, the presence of a cutoff peak in the acoustic signal, and a signal-to-noise ratio (SNR) below 30 dB of sound pressure level (SPL)⁽¹¹⁾.

Since the research aimed to define the cutoff values of the acoustic measures extracted from the sustained vowel [a] task for discrimination between voices with and without vocal disorders in BP speakers, the eligibility criteria allowed for the inclusion of individuals in both groups. The group without vocal disorders consisted of individuals without vocal complaints, structural or functional laryngeal disorders, and without deviations in vocal quality as demonstrated by the APJ. The group with vocal disorders consisted of individuals with vocal complaints, structural or functional laryngeal disorders, and deviations in vocal quality detected by the APJ.

A sample size calculation based on population size (Equation 1) was performed to estimate how many participants were needed for each group. The parameters used were: population size with vocal disorders (N) in Brazil of 15,232,500 (203,100,000 x 0.075)^(15,16), margin of error (e) of 10%, confidence level of 95% (z=1.96), and population proportion of individuals with non-teaching vocal disorders (p) of 0.5. A sample size (n) of 97 participants per group was calculated.

$$Sample\ Size(n) = \frac{z^2 \cdot p(1-p)}{e^2} \div \left(1 + \frac{z^2 \cdot p(1-p)}{e^2 N} \right) \quad (1)$$

Being:

z = 1.96

p = 0.5

e = 0.1

N = 15,232,000 (7.5% of non-teaching individuals with dysphonia are found in a population of 203,100,000 in Brazil)

Initially, 391 individuals were included (304 women and 87 men). Of these, 15 were excluded based on the following criteria: three individuals with an inconclusive diagnosis of voice disorder to avoid inconsistency in the dataset and to maintain the clinical relevance of the study results; 12 individuals were voice professionals. Thus, the study sample consisted of 376 participants: 277 with voice disorder (217 women and 60 men) and 99 without voice disorder (71 women and 28 men), with a mean age of 41.20 ± 14.04 years.

The recruitment of participants with voice disorders was carried out from the waiting list of a university voice laboratory. All patients sought the service spontaneously or were referred by otolaryngologists. Participants without voice disorders were recruited from students and staff of the university where the research was conducted.

The vocal sample collection procedures followed the laboratory's routine initial vocal assessment protocol. All data were recorded digitally (voice signals) or in medical records (anamnesis and laryngeal examination report). Vocal samples were collected during the first clinical vocal assessment, before vocal therapy started. All participants answered a brief medical history interview and were referred for visual laryngeal examination by an otolaryngologist, with a request to return the report within 15 days.

The following equipment was used for recording: Fonoview software (version 4.5, CTS Informática), Dell all-in-one computer, Sennheiser E-835 cardioid unidirectional microphone (on stand), coupled to a Behringer U-Phoria UMC 204 preamplifier. The recordings were conducted in an acoustic booth with background noise below 20 dB SPL, a sampling rate of 44,000 Hz, and 16 bits. The microphone was positioned 10 cm from the speaker's mouth at a 45° angle. The distance was measured with a ruler before each task, and the participant was positioned at the rear edge of the booth to maintain the fixed distance. The participant was instructed not to move their head.

All participants recorded the sustained vowel [a] speech task at their self-selected habitual frequency and intensity. During recording, the signals were visually monitored using Fonoview software, observing vowel duration (minimum of 5 s) and the presence of peak clipping. When necessary, the recording was repeated to ensure its quality. After collection, participants were referred for an otolaryngological medical examination with videolaryngostroboscopy. These examinations were performed in the same laboratory and resulted in a written report.

For the APJ, vocal samples were presented to three speech-language pathologists, with over 20 years of experience in voice therapy. The APJ session took place in a quiet environment, with headphones connected to a laptop, at a comfortable volume, self-reported by the evaluators. For the APJ, the speech-language

pathologists listened to the vocal sample and then marked it on a Visual Analogue Scale (VAS) with a range between 0 and 100 mm. The VAS cutoff values⁽¹⁷⁾ were used to classify the voices according to the presence of vocal deviation and the overall degree (G). A score closer to 0 represents less deviation in vocal quality, and closer to 100 more deviant of the vocal quality⁽¹⁷⁾. Voices assessed with values ≤ 35.5 mm were considered without vocal disorder, while voices with values > 35.5 mm were classified as having a vocal disorder. At the end of the APJ session, 20% (n=76) of the vocal samples were randomly repeated to analyze intra-rater reliability using Cohen's kappa coefficient. We selected the results with the highest intra-rater reliability, which obtained a kappa coefficient of 0.89, indicating good agreement.

Before extracting the acoustic measurements analyzed in this study, we used the open-source software Praat⁽¹⁸⁾ (Paul Boersma and David Weenink, University of Amsterdam, Netherlands), version 6.2.10, to obtain the SNR of the vocal signals. The average SNR of the samples included in this research was 43.08 ± 8.02 dB SPL. This value is acceptable (SNR > 30 dB SPL) and guarantees the quality of the signals for the extraction of acoustic measurements⁽¹¹⁾. None of the vocal signals used in this study presented an RSR < 30 dB SPL. The measurements were automatically extracted using a script in the free software Praat version 6.2.10, VoxMore⁽¹⁹⁾. Forty-four acoustic measurements were extracted, including:

- Fundamental frequency statistics (F0) – eight measures: mean (f_0 mean), median (f_0 median), standard deviation (sdF0), first (f_0 q1) and third quartile (f_0 q3), minimum (f_0 min), maximum (f_0 max) and coefficient of variation (f_0 CV);
- Period statistics – three measures: period average (PER), period standard deviation (PSD), and natural logarithm of the period standard deviation (LNPSD);
- Short-term disturbance of F0 – five measures: local jitter, absolute jitter ($jitter_{ABS}$), jitter relative average perturbation ($jitter_{RAP}$), jitter period perturbation quotient ($jitter_{PPQ5}$) and jitter difference of differences in pitch ($jitter_{ddp}$);
- Short-term amplitude perturbation – seven measures: shimmer_{local}, shimmer_{dB}, shimmer_{APQ1}, shimmer_{APQ3}, shimmer_{APQ5}, shimmer_{APQ11} and shimmer_{DDA}, amplitude variability index (AVI);
- Spectral and cepstral measures – 16 measures: spectral decay, spectral falloff, six variations of spectral noise level (SNL_{100-2600Hz}, SNL_{100-3000Hz}, SNL_{100-5100Hz}, SNL_{100-8000Hz}, SNL_{2600-5100Hz} and SNL_{5100-8000Hz}), high-frequency noise (Hfno), difference between the amplitude of the first and second harmonic (H1H2), difference between the amplitude of the first formant and the amplitude of the first harmonic (H1A1), difference between the amplitude of the third formant and the amplitude of the

first harmonic (H1A3), spectral flatness of residue signal (SFR), standard deviation of the harmonic-to-noise ratio (HNRDP), harmonic-to-noise ratio Dejonckere (HNRD) and harmonic-to-noise ratio (HNR);

- Reverse filtering – one measure: Pitch amplitude (PA);
- Disturbances in waveform – four measures: autocorrelation and three variations of Glottal to Noise Excitation, including GNE_{1000Hz}, GNE_{2000Hz} and GNE_{3000Hz}.

To generate the ROC curve, the area under the ROC curve was calculated based on the best cutoff point for the analyzed acoustic measures of the sustained vowel, and its performance metrics were obtained. The AUC is a diagnostic performance measure of a test and can range from 0.5 to 1; the higher the value, the better the test performance, being classified as: excellent (> 0.90), good (0.80-0.90), acceptable (0.70-0.80), poor (0.60-0.70), and no acceptable discrimination capacity (< 0.60)⁽²⁰⁾.

In addition to AUC, the performance of the analyzed measures encompassed the positive predictive value (+PV), the negative predictive value (-PV), the positive likelihood ratio (+LR), and the negative likelihood ratio (-LR). In the present study, only the results of acoustic measures with accuracy, sensitivity, and/or specificity values ≥ 0.70 , considered satisfactory in this research, will be presented. Values less than 0.70 do not have acceptable discriminatory capacity to detect voices with and without vocal disorders²⁰. Thus, cutoff points and diagnostic accuracy indices (AUC, sensitivity, specificity, +LR, -LR, +PV, and -PV) were defined for only 23 of the 44 extracted and analyzed acoustic measures.

All analyses were performed using the Statistical Package for the Social Sciences (SPSS) software, version 2.0. The significance level adopted was 5%.

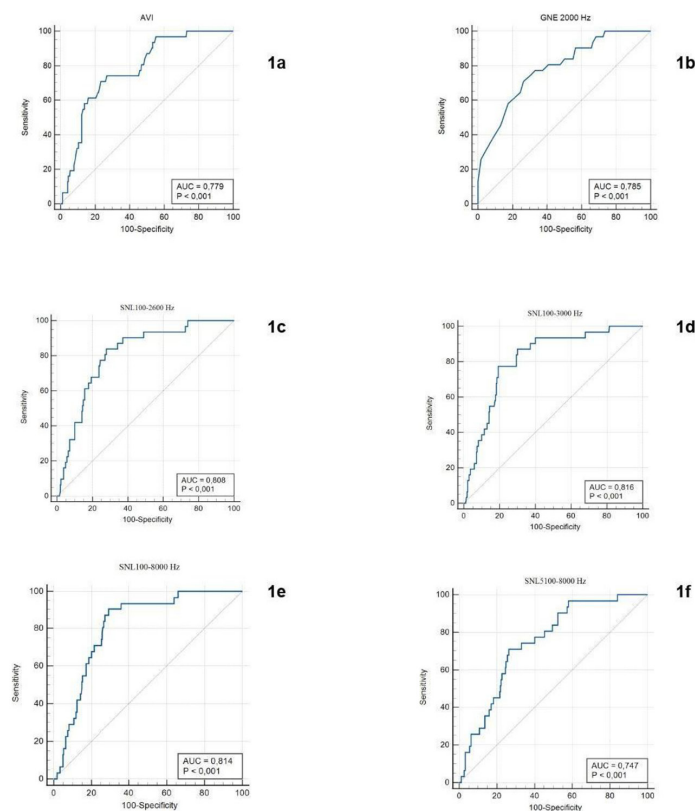
RESULTS

Table 1 was subdivided into three sections. It contains the cutoff values and performance metrics of the 23 acoustic measures that presented an AUC above 0.7, among the 44 analyzed in the study. Six acoustic measures showed satisfactory performance, with AUC > 0.70 , balanced sensitivity and specificity, and values greater than 70%: AVI (≤ 0.91), GNE_{2000Hz} (> 0.92), SNL_{100-2600Hz} (≤ 7.71), SNL_{100-3000Hz} (≤ 3.38), SNL_{100-8000Hz} (≤ 10.95), and SNL_{5100-8000Hz} (≤ -1.84) (Figure 1). Ten measures showed satisfactory sensitivity ($> 70\%$): GNE_{1000Hz} (> 0.92), HNR (> 20.73), LNPSD (≤ -9.58), PA (> 0.91), PSD (≤ 0.00), shimmer_{APQ5} (≤ 2.26), shimmer_{APQ11} (≤ 2.78), shimmer_{dB} (≤ 0.24), shimmer_{local} (≤ 2.78), and SNL_{100-5100Hz} (≤ 11.44). Seven measures showed satisfactory specificity ($> 70\%$): GNE_{3000Hz} (> 0.9), HNRD (> 29.78), Jitter_{local} (≤ 0.31), Jitter_{PPQ5} (≤ 0.18), Shimmer_{APQ3} (≤ 1.18), Shimmer_{DDA} (≤ 3.54), and SNL_{2600-5100Hz} (≤ 1.21).

Table 1. Cutoff point of the acoustic measurements extracted from the sustained vowel [a]

Acoustic Measures	Cutoff	AUC	Sensitivity (%)	Specificity(%)	+LR	-LR	+PV	-PV
AVI	≤ 0.91	0.779	70.97	76.49	3.02	0.38	25.9	95.8
GNE _{2000 Hz}	> 0.92	0.785	70.97	73.61	2.69	0.39	23.7	95.7
SNL _{100-2600 Hz}	≤ 7.71	0.808	83.87	72.12	3.01	0.22	25.7	97.5
SNL _{100-3000 Hz}	≤ 3.38	0.816	77.42	80.67	4	0.28	31.6	96.9
SNL _{100-8000 Hz}	≤ 10.95	0.814	90.32	70.63	3.08	0.14	26.2	98.4
SNL _{5100-8000 Hz}	≤ -1.84	0.747	70.97	73.61	2.69	0.39	23.7	95.7
GNE _{1000Hz}	> 0.92	0.76	83.87	54.65	1.85	0.3	17.6	96.7
HNR	> 20.73	0.837	90.32	64.68	2.56	0.15	22.8	98.3
LNPSD	≤ -9.58	0.722	93.55	48.88	1.83	0.13	17.5	98.5
PA	> 0.91	0.709	70.97	61.34	1.84	0.47	17.5	94.8
PSD	≤ 0.00	0.722	93.55	48.88	1.83	0.13	17.5	98.5
SHIMMER _{APQ 5}	≤ 2.26	0.779	93.55	51.12	1.91	0.13	18.1	98.6
SHIMMER _{APQ 11}	≤ 2.78	0.792	100	51.31	2.05	0	19.3	100
SHIMMER _{DB}	≤ 0.24	0.78	77.42	66.04	2.28	0.34	20.9	96.2
SHIMMER _{local}	≤ 2.78	0.776	77.42	65.67	2.26	0.34	20.7	96.2
SNL _{100-5100 Hz}	≤ 11.44	0.807	90.32	69.52	2.96	0.14	25.5	98.4
GNE _{3000 Hz}	> 0.9	0.76	64.52	82.16	3.62	0.43	29.4	95.3
HNR D	> 29.78	0.72	61.29	76.21	2.58	0.51	22.9	94.5
JITTER _{local}	≤ 0.31	0.712	67.74	70.9	2.33	0.46	21.2	95
JITTER _{PPQ5}	≤ 0.18	0.725	67.74	70.9	2.33	0.46	21.2	95
SHIMMER _{APQ 3}	≤ 1.18	0.765	64.52	74.63	2.54	0.48	22.7	94.8
SHIMMER _{DDA}	≤ 3.54	0.765	64.52	75	2.58	0.47	23	94.8
SNL _{2600-5100 HZ}	≤ 1.21	0.743	54.84	83.27	3.28	0.54	27.4	94.1

Caption: Cutoff: Cutoff value for the acoustic measure. AUC - Area Under the Curve. +LR: Positive likelihood ratio. Negative likelihood ratio. +PV: Positive predictive value. -PV: Negative predictive value. AVI - Amplitude Variability Index. GNE - Glottal-to-Noise Excitation - GNE_{1000Hz}, GNE_{2000Hz}, GNE_{3000Hz}. SNL - Spectral Noise Level - SNL_{100-2600Hz}, SNL_{100-3000Hz}, SNL_{100-5100Hz}, SNL_{100-8000Hz}, SNL_{2600-5100Hz}, SNL_{5100-8000Hz}. HNR - Harmonic-to-noise-ratio. LNPSD - Natural logarithm of the standard deviation of the period. PA - Pitch amplitude. PSD - Period standard deviation. Shimmer - shimmer_{APQ5}, shimmer_{APQ11}, shimmer_{dB}, shimmer_{local}, shimmer_{APQ3}, shimmer_{DDA}. HNR D - Harmonic-to-noise ratio Dejonckere. Local Jitter, Jitter_{PPQ5}. Jitter period perturbation quotient.

**Figure 1.** ROC curves of acoustic measures AVI (1a), GNE_{2000Hz} (1b), SNL_{100-2600Hz} (1c), SNL_{100-3000Hz} (1d), SNL_{100-8000Hz} (1e), and SNL_{5100-8000Hz} (1f) from sustained vowel [a] task

DISCUSSION

This study aimed to define the cutoff values of acoustic measures extracted from the sustained vowel [a] task for discrimination between the presence and absence of vocal disorders in BP speakers. Six acoustic measures obtained good performance in discriminating voices with and without vocal disorders, in addition to specific measures based on sensitivity or specificity, and their cutoff values can be used in clinical vocal practice.

Good accuracy is a combination that reflects the necessary balance between the precise detection of positive cases and the minimization of incorrect diagnoses, being fundamental for the reliable clinical applicability of the evaluated acoustic measures. A measure with high sensitivity has the capacity to correctly identify individuals with vocal disorders, the true positives, being a valuable instrument for screening for vocal disorders, as it minimizes the risk of not detecting true cases. High specificity indicates that the measures are effective in correctly identifying individuals who do not have vocal disorders, the negatives, and can be applied in diagnostic confirmation to avoid unnecessary treatment of individuals without vocal disorders⁽⁴⁾.

AVI measure is described in literature as having high sensitivity for predicting the presence of roughness⁽²¹⁾. In our research, AVI presented an AUC of 0.779, with a sensitivity of 70.97% and a specificity of 76.49%, suggesting acceptable performance (0.70-0.80) in distinguishing between groups with and without vocal disorders, and may have potential for use as a screening and evaluation tool. Although the +PV of this measure is low (25.9%), the -PV of 95.8% indicates that a negative result can be reliably used to exclude vocal disorders⁽²²⁾, however, a joint analysis with complementary measures is necessary for better decision-making security at the time of evaluation⁽²³⁾.

The measure GNE has proven relevant for voice acoustic analysis by estimating the relative amount of noise in sound signal, reflecting the efficiency of the glottal source compared to the turbulence of the vocal tract⁽³⁾. GNE can be useful in detecting vocal disorders, especially when combined with other acoustic measures⁽²⁴⁾. Among its variants, the GNE_{2000Hz} stands out in this study for presenting a more robust balance between diagnostic sensitivity and specificity. GNE_{2000Hz} achieved an AUC of 0.76, with a specificity of 73.61%, demonstrating better performance in correctly identifying individuals without vocal disorders. Clinically, the GNE_{2000Hz} has potential application in differentiating between efficient glottal vocal production and sound signals with a higher noise component, being particularly useful in detecting breathiness and glottal incompetence. Studies with GNE_{2000Hz} reinforce its applicability as a complementary index for vocal disorders detection, especially in multiparametric assessment protocols⁽²⁵⁾. Regarding the other variations of this measure, despite GNE_{1000Hz} having high sensitivity (83.87%), it demonstrates low specificity (54.65%), resulting in a higher rate of false positives. Despite GNE_{3000Hz} having high specificity (82.16%), it did not demonstrate an advantage in terms of sensitivity and diagnostic stability, which favors the clinical use of GNE_{2000Hz} as a more efficient acoustic measurement.

The acoustic measurement SNL aims to quantify the proportion of aperiodic energy (noise) present in vocal signal⁽²¹⁾.

The higher the SNL value, the greater the roughness degree⁽⁸⁾. Among the performed measures, SNL_{100-3000Hz} presented the highest AUC (0.816), combining a sensitivity of 77.42% and a specificity of 80.67%, providing a good balance for detecting individuals with vocal disorders, making it a viable option in clinical practice⁽²⁶⁾. Furthermore, it presented the highest +LR (4.00), suggesting that a positive result considerably increases the probability of a vocal disorder⁽²⁷⁾. SNL_{100-8000Hz} showed one of the highest sensitivities (90.32%), an AUC of 0.814, and excellent -PV, indicating that this measure has strong potential for detecting voice disorders individuals. SNL_{100-2600Hz} presented an AUC of 0.808, with a sensitivity of 83.87% and a specificity of 72.12%. Its -LR was 0.22, indicating that a negative result significantly reduces the probability of vocal disorder presence. The SNL_{5100-8000Hz} measure obtained an AUC of 0.747, with balanced specificity and sensitivity.

Among the measures analyzed, ten performed well in identifying individuals with vocal disorders, but failed to identify individuals without vocal disorders, indicating that they have greater applicability in high-risk populations⁽³⁾. These measures exhibited high sensitivity, and an AUC greater than 0.7, but low specificity. These are: GNE_{1000Hz}, HNR, LNPSD, PA, PSD, shimmer_{APQ5}, shimmer_{APQ11}, shimmer_{dB}, shimmer_{local}, and SNL_{100-5100Hz}.

Seven other measures performed well in identifying individuals without vocal disorders, but failed to identify individuals with vocal disorders. These measures are applicable in low-risk populations⁽³⁾, as they showed high specificity and AUC greater than 0.7, but low sensitivity. They are: GNE_{3000Hz}, HNRD, jitter_{local}, jitter_{PPQ5}, shimmer_{APQ3}, shimmer_{DDA}, and SNL_{2600-5100Hz}.

HNR parameter stands out as a measure that quantifies the proportion of noise relative to the proportion of harmonics in a vocal sample and can be of great value in differentiating between voices with and without vocal disorders⁽²⁸⁾. In the present study, the HNR measure obtained the highest AUC among those analyzed (0.837), in addition to high sensitivity (90.32%), suggesting that it is one of the best metrics for detecting the presence of vocal alterations, despite its moderate specificity (64.68%), which indicates a possible number of false positives. This phenomenon is common in highly sensitive tests and should be considered when choosing the best diagnostic approach⁽²⁷⁾. This finding suggests that HNR may be a promising measure for initial screenings where the priority is to correctly identify positive cases⁽²⁹⁾. Combining highly sensitive measurements (such as HNR) with others of high specificity (such as SNL_{2600-5100Hz}) can optimize diagnostic accuracy. The results highlight the importance of acoustic analysis in voice diagnosis and monitoring, emphasizing the need for a multimodal approach, combining different parameters to increase diagnostic accuracy^(1,25).

The definition of cutoff points for acoustic measurements represents an important advance for clinical practice by providing objective parameters that can assist in screening and differential diagnosis between voices with and without vocal disorders. However, it is crucial to emphasize that the lack of satisfactory diagnostic performance for some measurements does not imply their clinical inapplicability. Many of these measurements can be important in specific contexts, such as intra-subject

comparison, allowing the monitoring of vocal changes over time, evaluating responses to therapeutic interventions or vocal techniques, and identifying subtle variations that are not captured by purely categorical assessments. Therefore, the integration of multiple acoustic parameters, associated with other procedures of multidimensional voice assessment, favors a more comprehensive and personalized approach to voice management, valuing both robust diagnostic indicators and data that contribute to the continuous monitoring and improvement of vocal health⁽³⁰⁾.

Every scientific study, by its nature, operates under certain conditions and methodological choices that generate limitations. The present work is no exception, and the considerations that follow aim to offer a transparent analysis of these aspects, outlining the scope and boundaries of applicability of the results. However, although these limitations inform the generalization and context of the findings, they do not compromise the internal validity of the study or the robustness of its main conclusions, which remain relevant and scientifically valid within the population and methodological universe investigated.

The cross-sectional and descriptive design prevents the establishment of direct causal relationships or the evaluation of the stability of the cutoff points over time. Additionally, despite the significant sample size, the imbalance between the number of participants with and without vocal disorders, although reflecting clinical reality, makes it difficult to generalize the cutoff points. Therefore, the cutoff points of this study can be used as a reference in clinical practice, but studies with more comprehensive samples and populations with different prevalences of vocal disorders are necessary. The lack of a recognized gold standard for this purpose can be considered a limitation of the study. The classification of participants as having or not having vocal disorders was based on a combination of visual laryngeal examination, performed by a physician, and the APJ performed by speech-language pathologists. Although the high reliability of the APJ was ensured, its subjective nature may influence categorization. Taken together, these considerations delimit the scope of applicability of the results and highlight the complexity of vocal assessment, while simultaneously expanding and refining knowledge in this area.

As a clinical recommendation, the preferred set of measures in cases of suspected vocal disorder, derived from the sustained vowel [a] task in BP speakers, are: $SNL_{100-2600Hz}$, $SNL_{100-3000Hz}$, and $SNL_{100-8000Hz}$, since they exhibit sensitivity and specificity $\geq 80\%$. For screening purposes, the recommended measures, based on sensitivity $\geq 80\%$, are: GNE_{1000Hz} , HNR, LNPSD, PSD, Shimmer_{APQ5}, Shimmer_{APQ11}, and $SNL_{100-5100Hz}$. Finally, for diagnostic confirmation, the measures with specificity of $\geq 80\%$ include GNE_{3000Hz} , and $SNL_{2600-5100Hz}$.

CONCLUSION

In the present study, out of the 44 acoustic measures extracted from the sustained vowel [a] in BP speakers, 23 presented satisfactory performance metrics. The acoustic measures AVI, GNE_{2000Hz} , $SNL_{100-2600Hz}$, $SNL_{100-3000Hz}$, $SNL_{100-8000Hz}$, and $SNL_{5100-8000Hz}$ demonstrated acceptable discriminatory ability to differentiate

between voices with and without vocal disorders. The measures GNE_{1000Hz} , HNR, LNPSD, PA, PSD, shimmer_{APQ5}, shimmer_{APQ11}, shimmer_{dB}, shimmer_{local}, and $SNL_{100-5100Hz}$ presented a cutoff point with satisfactory sensitivity and can be used in clinical practice for screening purposes. The measures GNE_{3000Hz} , HNRD, Jitter_{local}, Jitter_{PPQ5}, shimmer_{APQ3}, shimmer_{DDA}, and $SNL_{2600-5100Hz}$ presented a cutoff point with satisfactory specificity and are applicable for diagnostic confirmation.

REFERENCES

- Englert M, Lopes L, Vieira V, Behlau M. Accuracy of Acoustic Voice Quality Index and its isolated acoustic measures to discriminate the severity of voice disorders. *J Voice*. 2022;36(4):582.e1-10. <https://doi.org/10.1016/j.jvoice.2020.08.010>. PMID:32873433.
- Jalalinajafabadi F, Gadepalli C, Jarchi D, Cheetham BMG. Acoustic analysis and digital signal processing for the assessment of voice quality. *Biomed Signal Process Control*. 2021;70:103018. <https://doi.org/10.1016/j.bspc.2021.103018>.
- 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. <https://doi.org/10.1590/2317-1782/20182017282>. PMID:30365651.
- Uloza V, Pribušis K, Ulozaite-Stanienė N, Petrauskas T, Damaševičius R, Maskeliūnas R. Accuracy analysis of the multiparametric acoustic voice indices, the VWI, AVQI, ABI, and DSI measures, in differentiating between normal and dysphonic voices. *J Clin Med*. 2023;13(1):99. <https://doi.org/10.3390/jcm13010099>. PMID:38202106.
- Nascimento RCPB, Sampaio MC. Os efeitos do gênero e da idade nas medidas acústicas tradicionais e cepstrais em vozes saudáveis e disfônicas: revisão de literatura sistemática [trabalho de conclusão de curso]. Salvador: Departamento de Fonoaudiologia, Universidade Federal da Bahia; 2022.
- Lopes LW, Abreu SR. Accuracy and cut-off values of cepstral measures in the clinical evaluation of Brazilian Portuguese speakers. *J Voice*. 2024;1997(24):142-5. <https://doi.org/10.1016/j.jvoice.2024.04.021>. PMID:38724311.
- Englert MT, Behlau M, Lucero JC. Validação do Acoustic Voice Quality Index (AVQI v.03.01) e do Acoustic Breathiness Index (ABI) para o português brasileiro [thesis]. São Paulo: Universidade Federal de São Paulo, Escola Paulista de Medicina; 2020.
- Latoszek BBV, Maryn Y, Gerrits E, De Bodt M. A metaanalysis: acoustic measurement of roughness and breathiness. *J Speech Lang Hear Res*. 2018;61(2):298-323. https://doi.org/10.1044/2017_JSLHR-S-16-0188. PMID:29392295.
- Rutjes AWS, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PMM. Evidence of bias and variation in diagnostic accuracy studies. *CMAJ*. 2006;174(4):469-76. <https://doi.org/10.1503/cmaj.050090>. PMID:16477057.
- Roy N, Barkmeier-Kraemer J, Eadie T, Sivasankar MP, Mehta D, Paul D, et al. Evidence-based clinical voice assessment: a systematic review. *Am J Speech Lang Pathol*. 2013;22(2):212-26. [https://doi.org/10.1044/1058-0360\(2012\)12-0014](https://doi.org/10.1044/1058-0360(2012)12-0014). PMID:23184134.
- Deliyski DD, Shaw HS, Evans MK. Adverse effects of environmental noise on acoustic voice quality measurements. *J Voice*. 2005;19(1):15-28. <https://doi.org/10.1016/j.jvoice.2004.07.003>. PMID:15766847.
- Maryn Y, Roy N, De Bodt M, Van Cauwenberge P, Corthals P. Acoustic voice analysis: does vowel choice influence the diagnostic value? *J Voice*. 2019;33(5):666.e1-9.
- Titze IR, Lemke J, Montequin D. Voicing and silence periods in continuous speech. *J Acoust Soc Am*. 2020;107(2):1065-76.
- Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al, and the STARD Group. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *Radiology*. 2015;277(3):826-32. <https://doi.org/10.1148/radiol.2015151516>. PMID:26509226.

15. Brasil. Instituto Brasileiro de Geografia e Estatística. Pesquisa Nacional por Amostra de Domicílios Contínua: resultados 2022. Rio de Janeiro: IBGE; 2022.
16. Behlau M, Zambon F, Guerrieri AC, Roy N. Epidemiology of voice disorders in teachers and nonteachers in Brazil: prevalence and adverse effects. *J Voice*. 2012;26(5):665.e9-18. <https://doi.org/10.1016/j.jvoice.2011.09.010>. PMID:22516316.
17. Yamasaki R, Madazio G, Leão SHS, Padovani M, Azevedo R, Behlau M. Auditory-perceptual evaluation of normal and dysphonic voices using the Voice Deviation Scale. *J Voice*. 2017;31(1):67-71. <https://doi.org/10.1016/j.jvoice.2016.01.004>. PMID:26873420.
18. Boersma P, Weenink D. Praat: doing phonetics by computer. Version 6.0.37 [software]. Amsterdam: University of Amsterdam; 2018.
19. Abreu SR, Moraes RM, Martins PN, Lopes LW. VoxMore: technological artifact to assist voice acoustic evaluation in the teaching-learning process and clinical practice. *CoDAS*. 2023;35(6):e20220166. <https://doi.org/10.1590/2317-1782/20232022166pt>. PMID:37909540.
20. Hosmer DW Jr, Lemeshow S. Applied logistic regression. 2nd ed. New York: Wiley; 2000. <https://doi.org/10.1002/0471722146>.
21. Abreu SR. Índice do grau geral do desvio vocal: desenvolvimento, avaliação e validação de um modelo de suporte à decisão para os falantes do português brasileiro [thesis]. João Pessoa: Universidade Federal da Paraíba; 2024.
22. Zhao L, Xu R, Zhu H, Zhang L. Predictive values of acoustic voice parameters for diagnosing vocal fold pathologies. *Clin Otolaryngol*. 2018;43(3):745-51.
23. Maryn Y, Corthals P, Van Cauwenberge P, Roy N, De Bodt M. Toward improved ecological validity in the acoustic measurement of overall voice quality: combining continuous speech and sustained vowels. *J Voice*. 2009;24(5):540-55. <https://doi.org/10.1016/j.jvoice.2008.12.014>. PMID:19883993.
24. Latoszek BBV, Mayer J, Watts CR, Lehnert B. Advances in clinical voice quality analysis with VOXplot. *J Clin Med*. 2023;12(14):4644. <https://doi.org/10.3390/jcm12144644>. PMID:37510759.
25. Nguyen DD, Novakovic D, Madill C. Voice disorder discrimination using vowel acoustic measures in female speakers. *Int J Lang Commun Disord*. 2024;59(5):2087-102. <https://doi.org/10.1111/1460-6984.13081>. PMID:38884559.
26. Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett*. 2006;27(8):861-74. <https://doi.org/10.1016/j.patrec.2005.10.010>.
27. Leite DRA. Desenvolvimento de um modelo de classificação da tipologia dos sinais vocais com base no deep learning [thesis]. João Pessoa: Universidade Federal da Paraíba; 2022.
28. Cappellari VM, Cielo CA. Características vocais acústicas de crianças pré-escolares. *Rev Bras Otorrinolaringol*. 2008;74(2):82-8. <https://doi.org/10.1590/S0034-72992008000200018>.
29. Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and using sensitivity, specificity and predictive values. *Indian J Ophthalmol*. 2008;56(1):45-50. <https://doi.org/10.4103/0301-4738.37595>. PMID:18158403.
30. Patel RR, Awan SN, Barkmeier-Kraemer J, Courey M, Deliyski D, Eadie T, et al. Recommended protocols for instrumental assessment of voice: American Speech-Language-Hearing Association Expert Panel to Develop a Protocol for Instrumental Assessment of Vocal Function. *Am J Speech Lang Pathol*. 2018;27(3):887-905. https://doi.org/10.1044/2018_AJSLP-17-0009. PMID:29955816.

Author contributions

AF: writing – original draft; MB: supervision; JB: writing – original draft; SRA: data curation; VVR: supervision; LWL: supervision.

Use of artificial intelligence-assisted technology

The authors declare that artificial intelligence tool was used exclusively for grammatical review and improvement of the Portuguese language during the preparation of the manuscript. The tool did not interfere with the scientific content, statistical analyses, or conclusions of the study. The entire final text was fully reviewed, validated, and approved by the authors, who assume full responsibility for the content presented.