

Marina Englert<sup>1,2</sup> 

Livia Lima<sup>2</sup> 

Ana Carolina Constantini<sup>2</sup> 

Ben Barsties v. Latoszek<sup>3</sup> 

Youri Maryn<sup>3</sup> 

Mara Behlau<sup>1,2</sup> 

### Keywords

Voice  
Voice Disorders  
Acoustics  
Evaluation  
Auditory Perception

### Descritores

Voz  
Distúrbios da Voz  
Acústica  
Avaliação  
Percepção Auditiva

### Correspondence address:

Marina Englert  
Universidade Federal de São Paulo –  
Unifesp  
Rua Machado Bittencourt, 361/1001,  
Vila Mariana, São Paulo (SP), Brasil,  
CEP: 04044-001.  
E-mail: marinaenglert@gmail.com

Received: April 12, 2018

Accepted: August 15, 2018

# Acoustic Voice Quality Index - AVQI for brazilian portuguese speakers: analysis of different speech material

## *Acoustic Voice Quality Index - AVQI para o português brasileiro: análise de diferentes materiais de fala*

### ABSTRACT

**Purpose:** This study aimed to verify the best speech material for the AVQI for Brazilian Portuguese language and identify the best validity results between the auditory perceptual judgment (APJ) and the AVQI score on different speech materials. **Methods:** We recorded voice samples of 50 individuals (dysphonic and vocally healthy) of several continuous speech (cs) variants (i.e., months of the year, numbers 1 to 20, and CAPE-V sentences) and attached the vowel /a/ in each case. The recorded samples were edited to three different durations of cs variants plus vowel: D1-total speech material; D2-customized speech material without voiceless parts; D3-pre-defined cut-off point speech material. These samples were submitted to three voice experts who judged the overall voice quality; and the AVQI analysis. AVQI's precision and concurrent validity were evaluated considering a Gmean threshold of  $G < 0.5$  and  $G < 0.68$ . **Results:** The concurrent validity of AVQI and APJ ranged from  $r = 0.482$  to  $r = 0.634$ . Numbers presented higher values for all durations. For  $G < 0.5$ , the best sensitivity and area under the ROC curve was for CAPE-V sentences at D3 (57.8%; 72%). For  $G < 0.68$ , numbers 1 to 10 had the best diagnostic accuracy and numbers 1 to 20 had the best sensitivity. **Conclusion:** Numbers from 1 to 10 had the best correlation results between APJ and AVQI. For  $G < 0.5$ , sentences had the best diagnostic accuracy; while for  $G < 0.68$ , numbers had the best diagnostic accuracy. Numbers are commonly used in the Brazilian clinic routine. According to the validity results and daily clinical practice in Brazil, we suggest the use of numbers as cs for concatenated voice samples of voice quality assessments.

### RESUMO

**Objetivo:** Verificar a melhor amostra de fala para validação do AVQI para o português-brasileiro; identificar o contexto de fala com melhor correlação perceptivo-acústica e que possui maior acurácia diagnóstica com o AVQI. **Método:** Gravações de 50 sujeitos (disfônicos e vocalmente saudáveis), incluindo: vogal/a/; meses do ano; números (1 a 20) e repetição das frases do CAPE-V. As amostras de fala foram editadas para conter três diferentes durações mais vogal: D1-fala completa; D2-fala com 3s de segmentos sonoros; D3-fala com ponto de corte pré-determinado. Três avaliadores realizaram a análise perceptivoauditiva (APA) das amostras combinadas em 3 contextos seguidos da vogal e deram um único escore do desvio vocal ( $G: 0$  a  $3$ ). Verificou-se qual estímulo de fala possuía melhor correlação perceptivo-acústica considerando o  $G_{\text{médio}}$ ; analisou-se qual estímulo possuía melhor acurácia diagnóstica considerando como presença ou ausência  $G < 0,5$  e  $G < 0,68$ . **Resultados:** A correlação perceptivo-acústica variou de  $r = 0,482$  a  $r = 0,634$  (Correlação de Spearman); números apresentou os valores mais elevados. O AVQI foi altamente específico e pouco sensível. Considerando  $G < 0,5$ , a melhor sensibilidade e valor da curva ROC foi para frases em D3 (0,578; 0,72). Considerando  $G < 0,68$ , houve boa acurácia diagnóstica para números de 1 a 10 e maior sensibilidade para números de 1 a 20. **Conclusão:** Melhor correlação perceptivo-acústica foi para números, 1 a 10. As frases do CAPE-V produziram melhor acurácia diagnóstica com  $G < 0,5$ , números apresentou elevada acurácia diagnóstica com  $G < 0,68$ . Números é bastante usual na clínica brasileira, logo, sugere-se seu uso para validação e análises do AVQI.

Study conducted at Universidade Federal de São Paulo – Unifesp in partnership with the Centro de Estudos da Voz – CEV, São Paulo (SP), Brasil; University of Antwerp as a pilot study of a doctoral thesis, Antwerp, Belgium.

<sup>1</sup> Universidade Federal de São Paulo – Unifesp - São Paulo (SP), Brasil.

<sup>2</sup> Centro de Estudos da Voz – CEV - São Paulo (SP), Brasil.

<sup>3</sup> University of Antwerp, Antwerp, Belgium.

**Financial support:** nothing to declare.

**Conflict of interests:** nothing to declare.



This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## INTRODUCTION

The voice quality evaluation is performed by several professionals and it is essential in the clinical routine of voice patients. This evaluation uses different protocols<sup>(1)</sup> that need to be applied by trained professionals with enough know how<sup>(2)</sup>. The protocols include several assessments such as: perceptual-auditory judgment, acoustics, laryngeal imaging, aerodynamics, and the self-evaluation. These different assessments provide subjective or objective measures; the perceptual-auditory judgment and acoustics are two commonly procedures<sup>(3,4)</sup>.

The perceptual-auditory judgment of the voice quality is mostly used in voice clinics<sup>(5)</sup>. It is considered to be subjective, since it suffers great influence from the experience and the listeners internal standards, which are formed throughout his life according to his judgment experiences. Aiming to reduce the subjectivity of this rating it is usually complemented by patient self-assessment and acoustic measures<sup>(4,6,7)</sup>.

The acoustic measures provide more objective data of several voice parameters such as: fundamental frequency, cepstral peak prominence, jitter, shimmer, harmonic-to-noise ratio, among others; however, it is more often than not one single acoustic measure<sup>(3)</sup>. Voice is an acoustic phenomenon that must be evaluated in a multidimensional matter. Thus, the extraction of single parameters seems to be insufficient to characterize the voice quality. Therefore, the interest in multiparametric acoustic models of overall voice quality is raising; two examples are the Cepstral Spectral Index of Dysphonia (CSID) and the Acoustic Vocal Quality Index (AVQI)<sup>(4,8,9)</sup>.

Both indices, the CSID and AVQI, consider several acoustic parameters to provide one single score for voice quality. For its analysis, the indices consider two vocal tasks: a sustained vowel (sv)- traditionally used in the acoustic analysis of the voice - and a continuous speech (cs) part, that provides more information about the real vocal use, although less common in acoustic voice quality analyses<sup>(10)</sup>.

The CSID evaluates cs and sv separately, thus, the software generates two scores for voice quality ranging from 0 to 100 points. This index runs on the commercial KayPENTAX<sup>(9)</sup> program: Analysis of Dysphonia in Speech and Voice.

On the other hand, the AVQI, runs on the freeware Praat using an individual designed script that quantifies the vocal deviation considering concatenated voice samples of cs and sv<sup>(11)</sup>. The AVQI Praat-script generates one single score ranging from 0 to 10 points by combining six acoustic measurements (i.e., the smoothed cepstral peak prominence, harmonics-to-noise ratio, shimmer percent, shimmer dB, general slope of the spectrum, and tilt of the regression line through the spectrum)<sup>(11)</sup>.

The AVQI was originally developed in Dutch but has already been validated in other languages (German, English, French, Finnish, Korean and Lithuanian<sup>(12-15)</sup>). The index presents strong correlation with the auditory-perceptual judgment (APJ), ranging from 0.794 to 0.929, and also a consistent diagnostic accuracy. It is worth mentioning that the AVQI validations uses the reading of a phonetically balanced text, such as "The Rainbow Passage"<sup>(16)</sup> or its equivalents, as the connected speech sample.

Although many voice clinics and researchers from various countries use a phonetically balanced text to analyze cs, traditionally, the Brazilian voice evaluation uses automatic speech (e.g., months of the year, counting numbers), repetition of sentences and/or a spontaneous speech sample for its analysis. There is no standardized and phonetically balanced text for such evaluation in the Brazilian Portuguese language. In addition, the general Brazilian population lacks of fluent reading due to high illiteracy rates (8%) and low schooling (52%), according to data from the *Pesquisa Nacional por Amostra de Domicílios* (PNAD) of 2015<sup>(17)</sup>.

In order to complement teaching, research and vocal evaluation in the Brazilian Speech-Language Pathology Clinic, the validation of an objective and robust measure is essential. Thus, aiming at the future validation of this index, the objective of the present research was to verify the best speech material for the AVQI for the Brazilian

Portuguese language and to identify which stimuli best correlates with the APJ and the AVQI score, and which stimuli has the best diagnostic accuracy.

## METHODS

This study was approved by the Committee for Ethics in Research under the protocol number 2.106.335, June 2017. All participants agreed to participate in the study and signed an informed consent term.

Voice samples of 50 individuals (mean age 40.3 years old; standard deviation: 16.99) were recorded. The participants were divided into 38 dysphonic (5 men and 33 women) and 12 vocally healthy (4 men and 8 women). The vocally healthy individuals had no vocal complaint and a VHI-10 score below 5 points. The 38 dysphonic patients presented various medical dysphonia diagnoses. The data recording was performed in several different speech language pathology services and vocal clinics; thus, the diagnosis considered the last medical and health report containing the patient's clinical history. Subsequently, the authors classified the diagnosis according to the Behlau et al.<sup>(18)</sup> classification system. Hence, the dysphonia diagnosis of the voice-disordered patients was as following: 20 patients with functional dysphonia, 14 patients with organic-functional dysphonia and 4 patients with organic dysphonia. The individuals had various background and different professions. This variable was not controlled in the present study.

### Voice sample

The individuals were instructed to speak aloud, the months of the year, from January to December, count the numbers from 1 to 20, repeat the six sentences of the CAPE-V protocol<sup>(19)</sup> and to sustain the vowel /a/ at comfortable pitch and loudness.

All recordings were performed at a soundproof booth using an AKG C420 head-mounted condenser microphone, digitized at a rate of 44kHz and 16 bits of resolution with the 174 AKG MPA V L + the Focusrite iTrack Solo using the Audacity program version 2.0.6. The same program was used to edit the vowel /a/

in order to achieve 3 seconds without voice onset and offset to avoid instabilities of the raise and decay moments<sup>(20)</sup>.

The cs samples of all individuals were edited on the Praat program in order to obtain different durations. This process occurred as following:

1. D1 (Duration 1) - Total continuous speech material

Months of the year, January to December (32 syllables), counting number 1 to 20 (42 syllables) and all the CAPE-V sentences (60 syllables).

2. D2 (Duration 2) - Customized continuous speech material

The customized duration was performed so that the voiced segments of the continuous speech had three 3 seconds, as the sustained vowel. The purpose of this customization was that the voiced segments of the continuous speech had the same duration of the sustained vowel<sup>(11)</sup>.

In the following paragraph, the steps are described to perform this customization:

2a: Extraction of all voiceless segments using the extraction Praat-script from Maryn et al.<sup>(8)</sup>; the first 3 seconds of the voice sample was analyzed;

2b: The original audio file was hand-marked so that the audio voice sample file would have 3 seconds of voice segments as determined on step 2a. This hand-marked cut-off point was determined using the spectrogram, pitch contour and auditory feedback. The duration of each sample was verified by running the AVQI script on the edited voice sample and the 3 seconds vowel /a/; this script extracts the voiceless segments of the speech and links it to the vowel sample. A tolerance margin of 0.1 seconds, below or above 3 seconds for the continuous speech, was accepted.

Table 1 presents the average numbers of syllables for each continuous speech voice sample after the customization.

3. D3 (Duration 3) - Pre-defined cut-off point for the continuous speech material 20 syllables for months of the year (january to august), 15 syllables for counting number (1 to 10) and 32 syllables for the CAPE-V sentences (3 sentences: “Érica tomou suco de pera e amora; Sônia sabe sambar sozinha; *Olha lá o avião azul*”, equivalent to the English sentences: The blue spot is on the key again; How hard did he hit him; We were away a year ago).

Table 2 presents the continuous speech duration considering the voiceless and voiced segments and the voiced segments.

**Auditory-perceptual judgment**

Three Brazilian e speech-language pathologist who are voice experts with a mean of 8.67 years of clinical experience (minimum of 6 and maximum of 10 years) rated the voice

**Table 1.** Mean, median, minimum, maximum and standard-deviation for each continuous speech voice sample after the customization process

|                           | Months of the year | Numbers | CAPE-V Sentences |
|---------------------------|--------------------|---------|------------------|
| <b>Mean</b>               | 18.02              | 18.16   | 26.22            |
| <b>Median</b>             | 17                 | 17      | 25               |
| <b>Minimum</b>            | 11                 | 7       | 16               |
| <b>Maximum</b>            | 29                 | 42      | 45               |
| <b>Standard-Deviation</b> | 4.33               | 6.90    | 5.14             |

**Table 2.** Continuous speech duration for each voice sample considering voiceless and voiced segments and voiced segments

|                           | Mean  | Median | Minimum | Maximum | Standard Deviation |
|---------------------------|-------|--------|---------|---------|--------------------|
| <b>Months of the year</b> |       |        |         |         |                    |
| D1 Voiceless and voice    | 11.61 | 11.57  | 6.16    | 21.8    | 3.24               |
| D1 Voiced                 | 5.18  | 4.98   | 3.18    | 9.12    | 1.24               |
| D2 Voiceless and voice    | 6.29  | 6.07   | 4.00    | 10.62   | 1.55               |
| D2 Voiced                 | 3.00  | 3.00   | 2.91    | 3.09    | 0.06               |
| D3 Voiceless and voice    | 7.51  | 7.41   | 4.11    | 14.54   | 2.17               |
| D3 Voiced                 | 3.47  | 3.35   | 1.80    | 6.06    | 0.88               |
| <b>Numbers</b>            |       |        |         |         |                    |
| D1 Voiceless and voice    | 17.76 | 17.05  | 7.98    | 32.42   | 5.50               |
| D1 Voiced                 | 6.10  | 5.90   | 3.09    | 13.14   | 1.83               |
| D2 Voiceless and voice    | 9.23  | 9.19   | 5.49    | 15.77   | 2.60               |
| D2 Voiced                 | 3.00  | 3.00   | 2.91    | 3.09    | 0.05               |
| D3 Voiceless and voice    | 8.68  | 8.29   | 3.71    | 16.31   | 2.87               |
| D3 Voiced                 | 2.85  | 2.73   | 1.39    | 5.67    | 0.84               |
| <b>CAPE-V Sentences</b>   |       |        |         |         |                    |
| D1 Voiceless and voice    | 22.00 | 21.42  | 15.84   | 26.92   | 2.62               |
| D1 Voiced                 | 6.86  | 6.96   | 3.99    | 11.43   | 1.24               |
| D2 Voiceless and voice    | 9.03  | 8.64   | 6.21    | 15.32   | 1.80               |
| D2 Voiced                 | 3.00  | 3.00   | 2.91    | 3.09    | 0.06               |
| D3 Voiceless and voice    | 11.32 | 11.25  | 7.88    | 13.81   | 1.33               |
| D3 Voiced                 | 3.89  | 3.99   | 2.11    | 6.84    | 0.80               |

**Caption:** D1 = Total continuous speech material; D2 = Customized continuous speech material; D3 = Pre-defined cut-off point speech material; Voiceless and voice = Voiced and voiceless segments; Voiced = Only voiced segments

samples. They evaluated the overall voice quality for each one of the different speech material (months of the year, counting numbers and the CAPE-V sentences) considering the total continuous speech (D1).

The final voice samples for APJ contained the concatenation of particular cs parts and three seconds of sv. Thus, there was one audio file with the continuous speech, 1 second of silence, and the vowel. A total of 3 contexts were obtained for the APJ: 1<sup>st</sup>, month of the year + 1 second of silence + vowel /a/; 2<sup>nd</sup>, numbers + 1 second of silence + vowel /a/ and 3<sup>rd</sup>, CAPE-V sentences + 1 second of silence + vowel /a/. Therefore, 3 hearing sessions were performed to conclude the APJ, in which one session per context were achieved. To minimize memory effects, at least 1-hour break was taken between each session.

The raters used the G from the GRBAS scale<sup>(21)</sup> to perform their analysis for each context; G represents the degree of hoarseness. A 4-point Likert scale was used, in which: 0 = clear voice/no hoarseness; 1 = slightly hoarse; 2 = moderately hoarse; 3 = severely hoarse. The judgments took place in a quiet environment and the raters used headphones. To analyze the intra-rater reliability, 10% of the sample were repeated.

Furthermore, the listeners were blinded regarding the identity and diagnosis of the voice samples. Additionally, anchor voices were presented to the raters before they began the analysis. Thus, a better reliability was expected<sup>(22)</sup>. The anchors were representative of G = 0, G = 1, G = 2 and G = 3 and specific for each context.

The intra and inter-rater reliability were assessed by the Cohen's Kappa coefficient (Ck) and the Fleiss Kappa coefficient (Fk), respectively. Acceptable values were observed for the intra-rater reliability for the different contexts: months of the year (Ck = 0.667 to 1.000); numbers (Ck = 1.000 for all evaluators) and CAPE-V sentences (Ck = 0.688 to 1.000). The inter-rater reliability also presented acceptable values: 0.5038 for months of the year, 0.5788 for numbers and 0.6386 for the CAPE-V sentences. All raters were reliable, thus, the APJ related to the AVQI considered the G mean score of the 3 listeners. Furthermore, the analysis of AVQI's accuracy, a cut-off point of G<0.5 and G<0.68 were used as presence or absence of dysphonia. The G<0.68 was also used as cut-off point in order to include the analysis of the counting numbers.

### AVQI analysis

The AVQI 03.01 version were used for acoustic analysis, it considers six acoustic parameters according to the formula<sup>(11)</sup>:

$$AVQI_{03.01} = \left[ \frac{4.152 - (0.177 * CPPs) - (0.006 * HNR) - (0.037 * Shim)}{(0.941 * ShdB) + (0.01 * Slope) + (0.093 * Tilt)} + \right] * 2.8902 \quad (1)$$

To run the script by Barsties and Maryn<sup>(11)</sup>, each continuous speech voice sample and its durations, D1, D2 and D3, were opened on the Praat Program (version 6.0.6) simultaneously with the audio file of the 3 seconds vowel /a/ for each individual.

There was a total of 9 cs parts with several durations for the AVQI analysis: months of the year + vowel /a/; numbers + vowel /a/; CAPE-V sentences + vowel /a/, for each one of the three durations.

For each one of the combinations of the cs and sv, a report with the AVQI score was generated. Hence, each individual had 9 different reports and therefore, 9 different AVQI scores; being 3 scores for each context and each duration (D1, D2 and D3). Thus, there was a total of 450 scores for the 50 recorded individuals.

### Statistical analysis

The software SPSS (version 23.0) and the MS-Excel (version MS-Office 2013) were used for the statistical analysis. The level of significance was set at 0.05 (5%).

Spearman rank-order correlation was used to assess which context presented the best concurrent validity, thus, the higher correlation with the APJ considering the overall vocal quality results of the G mean and the AVQI score. The area under the receiver operating characteristic curve (AROC) was used to determine which context had the best diagnostic accuracy; the thresholds used were G<0.5 and G<0.68 as presence or absence of vocal deviation.

## RESULTS

The context with best correlation between the APJ and the AVQI was counting numbers 1 to 10 (Table 3).

The contexts with stronger correlation was counting numbers and the CAPE-V sentences. Considering these both contexts, the durations of the customized speech material (D2) and the pre-defined cut-off point were different, however, the AVQI scores in both durations, were considered to be the same (Table 4).

As to the diagnostic accuracy, considering all contexts, the AVQI presented high specificity (values ranging from 0.625 to 1.000) and low sensitivity (values ranging from 0.413 to 0.881). For G<0.5, the D3 CAPE-V sentences presented the best sensitivity and AROC (0.72 e 0.578). On the other hand, this context presented the lowest AVQI threshold (1.1). No G<0.5 was found for numbers, thus, the diagnostic accuracy was analyzed using G<0.68. Numbers 1 to 10 presented a good diagnostic accuracy, the higher sensitivity was for Numbers 1 to 20, both durations with low threshold values (Table 5).

**Table 3.** AVQI concurrent validity for each duration and speech context

|                           | Spearman rank-order correlation coefficient (r) |
|---------------------------|---|
| <b>Months of the year</b> |   |
| D1                        | 0.4826  |
| D2                        | 0.5084  |
| D3                        | 0.5500  |
| <b>Numbers</b>            |   |
| D1                        | 0.5796  |
| D2                        | 0.6219  |
| D3                        | 0.6347  |
| <b>CAPE-V Sentences</b>   |   |
| D1                        | 0.5036  |
| D2                        | 0.5194  |
| D3                        | 0.5374  |

**Caption:** D1 = Total continuous speech material + 3 seconds vowel /a/; D2 = Customized continuous speech material + 3 seconds vowel /a/; D3 = Pre-defined cut-off point speech material + 3 seconds vowel /a/

**Table 4.** Comparison of the AVQI scores and duration of voiced-only segments between D2 and D3 for numbers and CAPE-V sentences

|                         | Median | Minimum | Maximum | p-value  |
|-------------------------|--------|---------|---------|----------|
| <b>AVQI score</b>       |        |         |         |          |
| <b>Numbers</b>          |        |         |         |          |
| D2                      | 1.00   | -0.04   | 8.89    | 0.6343   |
| D3                      | 0.885  | -1.68   | 8.80    |          |
| <b>CAPE-V Sentences</b> |        |         |         |          |
| D2                      | 1.105  | -1.30   | 9.06    | 0.8067   |
| D3                      | 1.10   | -1.04   | 9.02    |          |
| <b>Stimuli Duration</b> |        |         |         |          |
| <b>Numbers</b>          |        |         |         |          |
| D2                      | 3.00   | 2.91    | 3.09    | 0.0029*  |
| D3                      | 2.73   | 1.39    | 5.67    |          |
| <b>CAPE-V Sentences</b> |        |         |         |          |
| D2                      | 3.00   | 2.91    | 3.09    | <0.0001* |
| D3                      | 3.99   | 2.11    | 6.84    |          |

Wilcoxon Test

**Caption:** D2 = Customized continuous speech material + 3 seconds vowel /a/; D3 = Pre-defined cut-off point speech material + 3 seconds vowel /a/; \*Significant values ( $p \leq 0.050$ )

**Table 5.** AVQI ROC curve data considering  $G < 0.5$  and  $G < 0.68$ 

|                             | Area ROC | Threshold | Sensitivity | Specificity | LR+   | LR-   |
|-----------------------------|----------|-----------|-------------|-------------|-------|-------|
| <b>G &lt; 0.5</b>           |          |           |             |             |       |       |
| <b>Months of the year</b>   |          |           |             |             |       |       |
| D1                          | 0.603    | 1.88      | 0.478       | 1.000       | Inf   | 0.522 |
| D2                          | 0.649    | 1.42      | 0.413       | 1.000       | Inf   | 0.587 |
| D3                          | 0.693    | 1.16      | 0.500       | 1.000       | Inf   | 0.500 |
| <b>Numbers</b>              |          |           |             |             |       |       |
| NO $G < 0.5$ , NOT ANALYZED |          |           |             |             |       |       |
| <b>CAPE-V Sentences</b>     |          |           |             |             |       |       |
| D1                          | 0.698    | 1.53      | 0.556       | 1.000       | Inf   | 0.444 |
| D2                          | 0.689    | 1.19      | 0.533       | 1.000       | Inf   | 0.467 |
| D3                          | 0.72     | 1.1       | 0.578       | 1.000       | Inf   | 0.422 |
| <b>G &lt; 0.68</b>          |          |           |             |             |       |       |
| <b>Numbers</b>              |          |           |             |             |       |       |
| D1                          | 0.796    | 0.83      | 0.881       | 0.625       | 2.349 | 0.190 |
| D2                          | 0.814    | 0.31      | 0.857       | 0.750       | 3.428 | 0.190 |
| D3                          | 0.817    | 0.19      | 0.857       | 0.750       | 3.428 | 0.191 |
| <b>CAPE-V Sentences</b>     |          |           |             |             |       |       |
| D1                          | 0.715    | 1.67      | 0.561       | 1.000       | Inf   | 0.439 |
| D2                          | 0.745    | 1.19      | 0.585       | 1.000       | Inf   | 0.415 |
| D3                          | 0.757    | 1.22      | 0.585       | 1.000       | Inf   | 0.415 |

**Caption:** D1 = Total continuous speech material + 3 seconds vowel /a/; D2 = Customized continuous speech material + 3 seconds vowel /a/; D3 = Pre-defined cut-off point speech material + 3 seconds vowel /a/; ROC = Receiver-operating Curve; LR+ = likelihood ratio for a positive result; LR- = likelihood ratio for a negative result

## DISCUSSION

The AVQI is an index that based on a multiparametric acoustic construct, it considers many acoustic parameters to generate one single score for overall voice quality<sup>(4,8)</sup>. Hence, it has higher ecological validity.

The AVQI uses an individual designed Praat-script that considers both cs and sv to perform its acoustic analysis and generate one single score for the vocal quality<sup>(11)</sup>. This index

was validated in several different languages; because it uses continuous speech, linguistic differences may influence the AVQI accuracy. The validations found a good concurrent validity, ranging from 0.794 to 0.929, and also a good diagnostic accuracy<sup>(12-15)</sup>. The present study was the first which performed AVQI in the Brazilian Portuguese language.

Usually, the AVQI uses an outline of the reading of a text for the analysis of the continuous speech part. The Brazilian voice clinic does not have a standardized text for vocal evaluation.

Also, the general Brazilian population lacks of fluent reading, what could be an important limitation for the AVQI use in Brazil. Thus, in order to validate the index for the Brazilian Portuguese language, first, the continuous speech sample must be defined.

The present study selected three types of continuous speech that are commonly used in the voice evaluation for possible use on the AVQI validation process. The cs voice samples were: month of the year, counting number 1 to 20 and the repetition of the CAPE-V sentences.

The intra and inter-rater reliability for all contexts were acceptable; studies in other languages considered values of at least 0.53 for intra and values of at least 0.37 for inter-rater reliability as acceptable to correlate with the AVQI<sup>(4,8,14,15)</sup>.

Numbers presented the stronger correlation between the APJ and the AVQI score (Table 3), also, counting numbers is the most commonly used in the Brazilian voice evaluation<sup>(23-26)</sup>. Thus, the Brazilian evaluators listen more to number while evaluating the voice, therefore, they have more training with this stimulus, so, the evaluation becomes more precise and has greater chances of best correlating with the AVQI score, as an objective analysis. On the other hand, the CAPE-V is a standardized protocol, however, it is beginning to be more widespread and many Brazilian professionals still do not use it in their daily clinic routine.

In general, the AVQI uses a text for the cs part that enables the analysis of fluency and intonation, parameters that are considered to determine the index final score. Therefore, the best speech material for the AVQI is theoretically the spontaneous speech. To standardize the spontaneous speech in order to compare the scores among different individuals and in different evaluations moments of the same individual is quite challenging. On the other hand, counting numbers is quite habitual for most people and usually expressed in an automatic and natural way; what may explain its better outcome. The CAPE-V sentences, as well as being new for the individual, has no semantic meaning and are out of context, which may reflect on hesitation by the speaker while repeating them, making the speech sound less natural.

The best precision and concurrent validity for the AVQI 03.01 is when the voiced segments of the continuous speech reach 3 seconds, like the length of the vowel<sup>(27)</sup>. The best concurrent validity for counting numbers was for D3, 1 to 10 (Table 3) with a mean duration was 2.85 seconds; D2 had a longer duration with 3 seconds (Table 4). Numbers in D3 (i.e. 1 to 10) presented durations values closer to 3 seconds (Table 2), thus, it seems that counting 1 to 10 (15 syllables) is more related to the AVQI best precision and concurrent validity. Moreover, the AVQI scores for both durations were the same (Table 4). Differences in the continuous speech duration were also observed for the CAPE-V sentences, similarly, this difference did not reflect at a different AVQI score (Table 4).

The AVQI validation in Japanese considered 30 syllables as the most appropriate, since this length presented outcomes more similar to the customized length with 3 seconds of voiced segments<sup>(14)</sup>. For the Brazilian Portuguese language around 15 syllables might be more proper for the AVQI analysis.

The AVQI is an index with high specificity (Table 5), it correctly identifies vocally healthy individuals thus, it is quite

unlikely that someone with dysphonia is identified as vocally healthy. On the other hand, the AVQI presented limitations regarding its sensitivity, therefore, many individuals with vocal deviation could be identified as vocally healthy.

According to other studies about AVQI in different languages<sup>(8,13-15,28,29)</sup>, the present study showed lower sensitivity regarding the area under the ROC curve. Considering  $G < 0.5$ , the best sensitivity was for the D3 CAPE-V sentences. However, it also presented the lowest AVQI threshold value (Table 5).

No G score below 0.5 was obtained for numbers, therefore, to analyze this speech material, the cut-off point of  $G < 0.68$  was considered. It is noteworthy that this analysis was not observed in other publications with the AVQI. Nevertheless, the present study was a pilot study and numbers presented the best perceptual-acoustic correlation. For this reason, the analysis of the ROC curve was performed considering also the  $G < 0.68$  cut-off point. By means of this, the CAPE-V sentences (found to be the context with higher accuracy considering  $G < 0.5$ ) could be compared with numbers (found to be the context with higher perceptual-acoustic correlation) regarding to the ROC curve analysis.

The best diagnostic accuracy was found for numbers 1 to 10 for  $G < 0.68$ ; in addition, higher sensitivity values were found with extremely low thresholds (Table 5). These values are lower than the findings in other languages, Japanese = 0.915<sup>(14)</sup> and Belgium = 0.923<sup>(11)</sup>. However, the Brazilian outcomes with numbers might be more favorable for AVQI in the Brazilian Portuguese language.

The preliminary data of the Brazilian Portuguese language showed moderate validity results which were the lowest in comparison with other studies and languages<sup>(4,8,14,15)</sup>. Also, it shows that the AVQI for the Brazilian Portuguese does not seem to be useful as a screening or diagnostic tool. It is worth mentioning that the aim of this study was not to validate the AVQI, but rather to define which speech stimulus should be used in its validation, whose analyzes must deepen such aspects. However, a future study is essential to validate AVQI in the Brazilian-Portuguese language to assess the validity of AVQI for potential recommendation as a screening or diagnostic tool for voice clinics and research in Brazil as well.

It is suggested that the AVQI validation, and further analyzes with this index uses, in the Brazilian Portuguese, the continuous speech sample of counting numbers, since numbers presented the best perceptual-acoustic correlation and higher accuracy values. In addition, counting numbers presented an average value very similar to the one of the best AVQI accuracy and internal consistency and it is an easy cut-off point for editing. Also, counting numbers is a common adopted speech task for voice analysis in Brazilian studies<sup>(23-26)</sup>, thus, its use will enable more retrospective research using the AVQI, when validated.

## CONCLUSION

The stimulus that best correlated between the APJ and the AVQI is numbers 1 to 10. The CAPE-V sentences presented the best diagnostic accuracy considering  $G < 0.5$ ; numbers presented the best diagnostic accuracy considering  $G < 0.68$ . Numbers

is usual in the Brazilian clinic routine and it was the speech sample with values closer to those of the best AVQI precision and concurrent validity; thus, number is suggested as cs part for AVQI in the Brazilian Portuguese language.

## REFERENCES

1. Dejonckere PH, Crevier-Buchman L, Marie JP, Moerman M, Remacle M, Woisard V. Implementation of the European Laryngological Society (ELS) basic protocol for assessing voice treatment effect. *Rev Laryngol Otol Rhinol.* 2003;124(5):279-83. PMID:15144022.
2. Linder R, Albers AE, Hess M, Pöppel SJ, Schönweiler R. Artificial neural network-based classification to screen for dysphonia using psychoacoustic scaling of acoustic voice features. *J Voice.* 2008;22(2):155-63. <http://dx.doi.org/10.1016/j.jvoice.2006.09.003>. PMID:17074463.
3. 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. [http://dx.doi.org/10.1044/1058-0360\(2012/12-0014\)](http://dx.doi.org/10.1044/1058-0360(2012/12-0014)). PMID:23184134.
4. Barsties B, De Bodt M. Assessment of voice quality: current state-of-the-art. *Auris Nasus Larynx.* 2015;42(3):183-8. <http://dx.doi.org/10.1016/j.anl.2014.11.001>. PMID:25440411.
5. Oates J. Auditory-perceptual evaluation of disordered vocal quality: pros, cons and future directions. *Folia Phoniatr Logop.* 2009;61(1):49-56. <http://dx.doi.org/10.1159/000200768>. PMID:19204393.
6. Sellars C, Stanton AE, McConnachie A, Dunnet CP, Chapman LM, Bucknall CE, et al. Reliability of perceptions of voice quality: evidence from a problem asthma clinic population. *J Laryngol Otol.* 2009;123(7):755-63. <http://dx.doi.org/10.1017/S0022215109004605>. PMID:19250586.
7. Roy N, Mazin A, Awan SN. Automated acoustic analysis of task dependency in adductor spasmodic dysphonia versus muscle tension dysphonia. *Laryngoscope.* 2014;124(3):718-24. <http://dx.doi.org/10.1002/lary.24362>. PMID:23946147.
8. 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.* 2010;24(5):540-55. <http://dx.doi.org/10.1016/j.jvoice.2008.12.014>. PMID:19883993.
9. Awan SN, Roy N, Zhang D, Cohen SM. Validation of the Cepstral Spectral Index of Dysphonia (CSID) as a screening tool for voice disorders: development of clinical cutoff scores. *J Voice.* 2016;30(2):130-44. <http://dx.doi.org/10.1016/j.jvoice.2015.04.009>. PMID:26361215.
10. Carding PN, Wilson JA, MacKenzie K, Deary IJ. Measuring voice outcomes: state of the science review. *J Laryngol Otol.* 2009;123(8):823-9. <http://dx.doi.org/10.1017/S0022215109005398>. PMID:19454129.
11. Barsties B, Maryn Y. External Validation of the Acoustic Voice Quality Index version 03.01 with extended representativity. *Ann Otol Rhinol Laryngol.* 2016;125(7):571-83. <http://dx.doi.org/10.1177/0003489416636131>. PMID:26951063.
12. Reynolds V, Buckland A, Bailey J, Lipscombe J, Nathan E, Vijayasekaran S, et al. Objective assessment of pediatric voice disorders with the acoustic voice quality index. *J Voice.* 2012;26(5):672.e1-7. <http://dx.doi.org/10.1016/j.jvoice.2012.02.002>. PMID:22632794.
13. Maryn Y, De Bodt M, Barsties B, Roy N. The value of the acoustic voice quality index as a measure of dysphonia severity in subjects speaking different languages. *Eur Arch Otorhinolaryngol.* 2014;271(6):1609-19. PMID:24162765.
14. Hosokawa K, Barsties V, Latoszek B, Iwahashi T, Iwahashi M, Iwaki S, Kato C, et al. The Acoustic Voice Quality Index version 03.01 for the japanese-speaking population. *J Voice.* 2019;33(1):125.e1-12. <http://dx.doi.org/10.1016/j.jvoice.2017.10.003>. PMID:29153335.
15. Uloza V, Petrauskas T, Padervinskis E, Ulozaitė N, Barsties B, Maryn Y. Validation of the Acoustic Voice Quality Index in the Lithuanian language. *J Voice.* 2016;14(16):30071-6. PMID:27427182.
16. Fairbanks G. *Voice and articulation drillbook.* 2nd ed. New York: Harper & Row; 1960. p. 124-9.
17. IBGE: Instituto Brasileiro de Geografia e Estatística. Pesquisa Nacional por Amostra de Domicílios 2015: PNAD. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística; 2017 [citado em 2017 Abr 14]. Disponível em: <https://www.ibge.gov.br/>
18. Behlau M, Azevedo R, Pontes P. Conceito de voz normal e classificação das disfonias. In: Behlau M. *Voz: o livro do especialista.* 1. ed. Rio de Janeiro: Revinter; 2001. Capítulo 2; p. 53-79.
19. Behlau M. Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V), ASHA 2003 [Refletindo sobre o novo]. *Rev Soc Bras Fonoaudiol.* 2004;9:187-9.
20. Behlau M. *Uma análise das vogais do português brasileiro falado em São Paulo: perceptual, espectrográfica de formantes e computadorizada de frequência fundamental [dissertação].* São Paulo: Universidade Federal de São Paulo; 1984.
21. Hirano M. *Clinical examination of voice.* New York: Springer Verlag; 1981.
22. Awan SN, Lawson LL. The effect of anchor modality on the reliability of vocal severity ratings. *J Voice.* 2009;23(3):341-52. <http://dx.doi.org/10.1016/j.jvoice.2007.10.006>. PMID:18346869.
23. Santos AC, Borrego MC, Behlau M. Effect of direct and indirect voice training in speech-language pathology and audiology students. *CoDAS.* 2015;27(4):384-91. <http://dx.doi.org/10.1590/2317-1782/20152014232>. PMID:26398263.
24. Fadel CB, Dassie-Leite AP, Santos RS, Santos CG Jr, Dias CA, Sartori DJ. Efeitos imediatos do exercício de trato vocal semiocluído com Tubo LaxVox® em cantores. *CoDAS.* 2016;28(5):618-24. <http://dx.doi.org/10.1590/2317-1782/20162015168>. PMID:27849247.
25. Diaféria G, Madazio G, Pacheco C, Takaki PB, Behlau M. Clima de grupo na terapia vocal de pacientes com Doença de Parkinson. *CoDAS.* 2017;29(4):e20170051. <http://dx.doi.org/10.1590/2317-1782/20172017051>. PMID:28876371.
26. Moreira FS, Gama AC. Efeito do tempo de execução do exercício vocal sopra e som agudo na voz de mulheres. *CoDAS.* 2017;29(1):e20160005. <http://dx.doi.org/10.1590/2317-1782/20172016005>. PMID:28225849.
27. Barsties B, Maryn Y. The Improvement of internal consistency of the Acoustic Voice Quality Index. *Am J Otolaryngol.* 2015;36(5):647-56. <http://dx.doi.org/10.1016/j.amjoto.2015.04.012>. PMID:25964174.
28. Maryn Y, DeBodt M, Roy N. The Acoustic Voice Quality Index: toward improved treatment outcomes assessment in voice disorders. *J Commun Disord.* 2010;43(3):161-74. <http://dx.doi.org/10.1016/j.jcomdis.2009.12.004>. PMID:20080243.
29. Barsties B, Maryn Y. The Acoustic Voice Quality Index: toward expanded measurement of dysphonia severity in German subjects. *HNO.* 2012;60(8):715-20. <http://dx.doi.org/10.1007/s00106-012-2499-9>. PMID:22527525.

## Author contributions

*ME* was responsible for the study design, data collection, audio editing, data analysis and writing of the manuscript; *LL* and *ACC* were responsible for the study design, data analysis and the manuscript revision and correction; *BBL* and *YM* were responsible for the index construction, study design, data analysis and revision of the manuscript English version; *MB* was the advisor, responsible for the study design, data analysis and final revision of the manuscript.