


Marina Englert^{1,2} 

Livia Lima² 

Ana Carolina Constantini² 

Ben Barsties v. Latoszek³ 

Youri Maryn³ 

Mara Behlau^{1,2} 

Descritores

Voz
Distúrbios da Voz
Acústica
Avaliação
Percepção Auditiva

Keywords

Voice
Voice Disorders
Acoustics
Evaluation
Auditory Perception

Endereço para correspondência:

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

Recebido em: Abril 12, 2018

Aceito em: Agosto 15, 2018

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

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

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 Gmédio; 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.

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.

Trabalho realizado na Universidade Federal de São Paulo – Unifesp em parceria com o Centro de Estudos da Voz – CEV, São Paulo (SP), Brasil; University of Antwerp como pesquisa piloto de uma tese de doutorado, Antuérpia, Bélgica.

¹ Universidade Federal de São Paulo – Unifesp - São Paulo (SP), Brasil.

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

³ University of Antwerp, Antwerp, Belgium.

Fonte de financiamento: nada a declarar.

Conflito de interesses: nada a declarar.



Este é um artigo publicado em acesso aberto (Open Access) sob a licença Creative Commons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

INTRODUÇÃO

A avaliação da qualidade da voz é feita por diversos profissionais, sendo rotina básica na clínica vocal. Diferentes protocolos são utilizados para sua avaliação⁽¹⁾, sendo necessário que profissionais que forem aplicá-los tenham treinamento prévio e certa *expertise* para tal⁽²⁾. Os protocolos contemplam a avaliação perceptivo-auditiva, avaliação acústica, avaliação visual, laríngea, avaliação de medidas aerodinâmicas e autoavaliação do paciente. Percebe-se que essas avaliações fornecem medidas subjetivas ou objetivas; duas comumente utilizadas são: avaliação perceptivo-auditiva e a avaliação acústica^(3,4).

A avaliação perceptivo-auditiva da voz é a mais comumente utilizada na clínica vocal⁽⁵⁾. Ela é considerada subjetiva, uma vez que sofre grande influência da experiência e sistema de referências internos que o ouvinte forma ao longo da vida de acordo com as experiências sonoras que ele vivencia, sendo assim, comumente é complementada pela autoavaliação vocal e por medidas acústicas^(4,6,7).

As medidas acústicas, apesar de fornecerem dados supostamente mais objetivos de parâmetros vocais diversos, tais como: frequência fundamental, *jitter*, *shimmer*, proporção harmônico-ruído, são compostas por uma única medida acústica⁽³⁾. A voz é um fenômeno acústico e deve ser avaliada de modo multidimensional e a extração de parâmetros isolados parece ser insuficiente para realmente caracterizá-la. Assim, vem crescendo o interesse por modelos acústicos multiparamétricos, tais como o *Cepstral Spectral Index of Dysphonia* (CSID) e o *Acoustic Vocal Quality Index* (AVQI)^(4,8,9).

Tanto o CSID quanto o AVQI consideram diversos parâmetros acústicos para fornecer um único escore para qualidade vocal. Ambos consideram a vogal sustentada – tradicionalmente já utilizada na análise acústica da voz - e a fala encadeada, que, embora forneça maiores informações do real uso da voz na fala, é menos comum nas análises acústicas voltadas para a avaliação da qualidade vocal⁽¹⁰⁾.

O CSID avalia separadamente a fala e a vogal, dessa forma, o *software* gera dois escores para a qualidade vocal que varia de 0 a 100 pontos. Esse índice é comercializado com o programa *Analysis of Dysphonia in Speech and Voice* da KayPENTAX⁽⁹⁾.

O AVQI, por sua vez, utiliza um *script* no programa Praat – um *software* livre que pode ser baixado pela internet - para quantificar a intensidade de desvio da qualidade vogal, considerando concomitantemente a fala encadeada e a vogal sustentada⁽¹¹⁾. Ele combina seis medidas acústicas, a saber: *Smoothed ceptral peak prominence CPPs*; *Harmonic-to-noise ratio HNR*; *Shimmer local Shim*; *Shimmer local dB ShdB*; *General slope of the spectrum Slope* e *Tilt of the regression line through the pectrum Tilt* para fornecer um único escore de 0 a 10 pontos⁽¹¹⁾.

O AVQI, originalmente desenvolvido em holandês, já foi validado para outros idiomas (alemão, inglês, francês, finlandês, coreano e lituano⁽¹²⁻¹⁵⁾), apresentando forte correlação com a análise perceptivo-auditiva, que varia entre 0,794 e 0,929, e precisão diagnóstica consistente. Vale ressaltar que a amostra de fala utilizada no AVQI e nessas validações é a leitura de

um texto foneticamente balanceado, como o “*The Rainbow Passagem*”⁽¹⁶⁾ ou seus equivalentes.

No Brasil, a avaliação da qualidade vocal da fala encadeada utiliza, tradicionalmente, sequências automáticas (meses do ano, contagem de números), repetição de frases e/ou amostra de fala espontânea, não havendo no português brasileiro um texto padronizado e foneticamente balanceado para tal avaliação. Ainda, há questões relacionadas ao nível de leitura dos sujeitos, uma vez que a população brasileira, em sua grande maioria, não é fluente para realizar a leitura de um texto, devido à grande taxa de analfabetismo (8%) e baixa escolaridade (52%), segundo dados da Pesquisa Nacional por Amostra de Domicílios (PNAD) para 2015⁽¹⁷⁾.

Existe a necessidade da validação de medidas objetivas e robustas, tais como o AVQI, para complementar o ensino, a pesquisa e a avaliação vocal na clínica fonoaudiológica brasileira. Assim, visando a futura validação desse índice, o objetivo da presente pesquisa foi verificar qual a melhor amostra de fala encadeada para ser utilizada na validação do AVQI para o português brasileiro e identificar o contexto de fala com melhor correlação perceptivo-acústica e que possui maior acurácia diagnóstica com o AVQI.

MÉTODO

O projeto de pesquisa foi aprovado pelo Comitê de Ética em Pesquisa sob o número do parecer consubstanciado 2.106.335, de junho de 2017. Todos os sujeitos envolvidos concordaram em participar do estudo e assinaram o termo de consentimento livre e esclarecido.

Foram gravadas amostras de fala de 50 sujeitos (idade média de 40,3 anos de idade; desvio padrão: 16,99), 38 disfônicos (5 homens e 33 mulheres) e 12 vocalmente saudáveis (4 homens e 8 mulheres). Os sujeitos vocalmente saudáveis não tinham queixa vocal e possuíam escore do IDV-10 abaixo de 5 pontos. Os 38 sujeitos disfônicos possuíam diagnósticos médicos de disfonia variados. A coleta foi realizada em diversos serviços fonoaudiológicos e clínicas vocais, dessa forma, considerou-se o último laudo médico para definir o diagnóstico. Posteriormente, os autores classificaram os diagnósticos de acordo com a classificação de Behlau et al.⁽¹⁸⁾. Havia na amostra: 20 sujeitos com disfonia funcional, 14 sujeitos com disfonia organofuncional e 4 sujeitos com disfonia orgânica. Os sujeitos possuíam diferentes profissões, não sendo essa variável controlada.

Amostra de fala

Os sujeitos foram instruídos a falar, em frequência e intensidade habituais, os meses do ano, de janeiro a dezembro, contar os números de 1 a 20, repetir as seis frases do protocolo CAPE-V⁽¹⁹⁾ e a vogal /a/ sustentada.

Todas as gravações foram realizadas em cabina acústica com o microfone de cabeça AKG C420 e digitalizadas em 44 kHz e 16 bits pelo programa *Audacity*, versão 2.0.6. A gravação da vogal /a/ sustentada foi editada no mesmo programa, eliminando-se o início e final da emissão, para evitar as instabilidades dos instantes de *raise* e *decay*⁽²⁰⁾, mantendo-se 3 segundos de duração.

As amostras de fala encadeada foram editadas no Praat via *script* do AVQI para que diferentes durações fossem obtidas. Esse processo ocorreu da seguinte forma:

1. D1 (Duração 1) - Amostra de fala completa

Meses do ano de janeiro a dezembro (32 sílabas), contagem de números de 1 a 20 (42 sílabas) e todas as frases do CAPE-V (60 sílabas).

2. D2 (Duração 2) - Amostra de fala com ponto de corte customizado

A customização foi feita a fim de que apenas os segmentos sonoros atingissem 3 segundos de duração, como a vogal sustentada. O objetivo dessa customização é que a parte sonora da fala encadeada tenha a mesma duração da vogal sustentada.

Passos para a edição de cada amostra de fala encadeada:

2a: Extração de todos os segmentos surdos do estímulo de fala utilizando o *script* de extração específico do Praat elaborado por Maryn et al.⁽⁸⁾;

2b: Edição do arquivo de áudio original, a fim de que, ao se extrair os segmentos surdos, o arquivo atingisse 3 segundos de extensão; tal ponto foi determinado por meio da espectrografia, marcação da frequência e *feedback* auditivo. A customização do ponto de corte com tolerância de mais ou menos 0,1 segundo foi confirmada ao analisar os estímulos de áudio com uso do *script* do AVQI que, após extração dos seguimentos surdos, concatena a amostra de fala encadeada e de vogal, extraindo os segmentos surdos da fala.

A Tabela 1 apresenta o número médio de sílabas para cada amostra de fala encadeada após customização.

3. D3 (Duração 3) - Amostra de fala com ponto de corte pré-definido

Para meses do ano, ficaram definidas 20 sílabas (janeiro a agosto); para contagem de números, 15 sílabas (1 a 10); e para frases do CAPE-V, 32 sílabas (“Érica tomou suco de pera e amora; Sônia sabe sambar sozinha; Olha lá o avião azul”).

A Tabela 2 mostra dados da duração das amostras de fala encadeada considerando os segmentos surdos e sonoros e apenas os sonoros para todas as durações.

Tabela 1. Média, mediana, mínimo, máximo e desvio padrão do número de sílabas para cada amostra de fala encadeada após customização em D2

	Meses do Ano	Números	Frases do CAPE-V
Média	18,02	18,16	26,22
Mediana	17	17	25
Mínimo	11	7	16
Máximo	29	42	45
Desvio padrão	4,33	6,90	5,14

Análise perceptivo-auditiva

A avaliação perceptivo-auditiva (APA) do grau geral da qualidade vocal foi feita por 3 fonoaudiólogos brasileiros especialistas em voz, com média de 8,67 anos (mínimo de 6, máximo de 10 anos) de experiência clínica. Eles avaliaram as vozes para todas as amostras

Tabela 2. Duração do estímulo de fala para cada amostra de fala encadeada considerando os segmentos surdos e sonoros e apenas os segmentos sonoros

	Média	Mediana	Mínimo	Máximo	Desvio Padrão
Meses do ano					
D1 Completo	11,61	11,57	6,16	21,8	3,24
D1 Sonoro	5,18	4,98	3,18	9,12	1,24
D2 Completo	6,29	6,07	4,00	10,62	1,55
D2 Sonoro	3,00	3,00	2,91	3,09	0,06
D3 Completo	7,51	7,41	4,11	14,54	2,17
D3 Sonoro	3,47	3,35	1,80	6,06	0,88
Números					
D1 Completo	17,76	17,05	7,98	32,42	5,50
D1 Sonoro	6,10	5,90	3,09	13,14	1,83
D2 Completo	9,23	9,19	5,49	15,77	2,60
D2 Sonoro	3,00	3,00	2,91	3,09	0,05
D3 Completo	8,68	8,29	3,71	16,31	2,87
D3 Sonoro	2,85	2,73	1,39	5,67	0,84
Frases do CAPE-V					
D1 Completo	22,00	21,42	15,84	26,92	2,62
D1 Sonoro	6,86	6,96	3,99	11,43	1,24
D2 Completo	9,03	8,64	6,21	15,32	1,80
D2 Sonoro	3,00	3,00	2,91	3,09	0,06
D3 Completo	11,32	11,25	7,88	13,81	1,33
D3 Sonoro	3,89	3,99	2,11	6,84	0,80

Legenda: D1 = Amostra de fala completa; D2 = Amostra de fala customizada com 3 segundos de segmentos sonoros; D3 = Amostra de fala com ponto de corte; Completo = Segmentos surdos e sonoros da fala; Sonoro = Apenas segmentos sonoros da fala

da fala encadeada (meses do ano, números de 1 a 20 e frases do CAPE-V) considerando as amostras de fala completas (D1).

As amostras de fala encadeada foram combinadas com a amostra da vogal sustentada por 3 segundos em um mesmo arquivo de áudio e com um segundo de silêncio entre eles. Dessa forma, obtiveram-se 3 contextos para a análise perceptivo-auditiva: 1º, meses do ano + 1 segundo de silêncio + vogal /a/; 2º, números + 1 segundo de silêncio + vogal /a/; e 3º, frases do CAPE-V + 1 segundo de silêncio + vogal /a/. Assim, houve 3 sessões de escuta para concluir a análise perceptivo-auditiva, uma sessão para a análise de cada contexto. Para minimizar efeitos de memória auditiva, houve um intervalo de no mínimo 1 hora entre uma sessão e outra.

Os avaliadores utilizaram o G da escala GRBAS⁽²¹⁾ para realizar a avaliação de cada contexto. O G se refere à qualidade vocal geral da voz em uma escala de *likert* de 4 pontos, a saber: 0 = sem alteração vocal, 1 = alteração vocal leve, 2 = alteração vocal moderada e 3 = alteração vocal intensa. A análise perceptivo-auditiva foi realizada em ambiente silencioso com uso de fone de ouvido. Os avaliadores poderiam repetir as vozes quantas vezes desejassem para concluir a análise e eram cegos quanto ao diagnóstico e quanto à presença ou ausência de disfonia. Repetiram-se 10% das amostras de voz para mensuração da confiabilidade intrassujeito.

A fim de garantir uma melhor confiabilidade da avaliação perceptivoauditiva, os avaliadores ouviram vozes âncoras referentes a cada contexto antes de iniciar a avaliação⁽²²⁾ e poderiam consultá-las quantas vezes desejassem durante a análise. Utilizaram-se estímulos âncoras representativos de G = 0, G = 1, G = 2 e G = 3 para cada contexto utilizado.

Para a análise da confiabilidade da avaliação perceptivo-auditiva, intra e intersujeitos, utilizaram-se os Coeficiente Kappa (Ck) e o Coeficiente Fleiss Kappa (Fk), respectivamente. A confiabilidade intrassujeitos apresentou valores aceitáveis para os diferentes contextos de fala: meses do ano (Ck = 0,667 a 1,000); números (Ck = 1,000 para todos os avaliadores) e frases do CAPE-V (Ck = 0,688 a 1,000). A confiabilidade intersujeitos também teve valores aceitáveis, sendo de 0,5038 para meses do ano, 0,5788 para números e de 0,6386 para frases do CAPE-V. Como os três avaliadores foram confiáveis, as análises da APA em relação ao AVQI consideraram o escore médio de G dos 3 avaliadores, utilizando os cortes de G<0,5 e G<0,68 como presença e ausência de disfonia para posterior análise da acurácia.

Análise do AVQI

A versão 03.01 do AVQI foi utilizada para realizar a análise acústica. Essa versão do índice trata os sinais acústicos no programa *Praat* utilizando um *script*. A análise considera seis parâmetros acústicos (*Smoothed cepstral peak prominence CPPs*; *Harmonic-to-noise ratio HNR*; *Shimmer local Shim*; *Shimmer local dB ShdB*; *General slope of the spectrum Slope*; *Tilt of the regression line through the pectrum Tilt*) segundo a fórmula⁽¹¹⁾:

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

Para utilizar esse *script*, abriu-se separadamente no Praat (versão 6.0.6) cada amostra de fala encadeada, com as respectivas durações, D1, D2 e D3, simultaneamente com um outro arquivo de áudio contendo a vogal /a/ já editada com três segundos de duração. Sendo assim, havia no Praat dois arquivos de áudio, um referente à amostra de fala encadeada e outro referente à vogal /a/ com 3 segundos de duração. Esses arquivos de fala encadeada e de vogal sustentada foram renomeados como “cs” e “sv” respectivamente, conforme orientação do *script* do AVQI.

Ao todo, obtiveram-se 9 contextos de fala para análise do AVQI: meses do ano + vogal /a/; números + vogal /a/; e frases do CAPE-V + vogal /a/, para cada uma das durações.

O *script*, ao analisar cada contexto, contendo as amostras da fala encadeada e a vogal sustentada, produziu um arquivo final com o escore do AVQI. Assim sendo, cada sujeito teve 9 escores, 3 escores para cada contexto de fala referente a cada uma das durações (D1, D2 e D3), totalizando 450 escores para 50 sujeitos.

Análise estatística

Para a análise estatística, foram utilizados os *softwares* SPSS versão 23.0, MS-Excel, em sua versão do MS-Office 2013. O nível de significância foi de 0,05 (5%). A Correlação de Spearman foi utilizada para identificar qual estímulo de fala correlacionou-se melhor entre a análise perceptivo-auditiva, considerando o G médio dos 3 avaliadores, e o escore do AVQI. A Curva ROC foi utilizada para determinar qual estímulo de fala possuía melhor acuracidade diagnóstica e utilizaram-se os cortes de G<0,5 e G<0,68 como presença ou ausência para essa análise.

RESULTADOS

A amostra de fala que melhor se correlacionou com a análise perceptivo-auditiva e o escore do AVQI foi a contagem de números, sendo mais forte para números de 1 a 10 (Tabela 3).

Para os contextos que apresentaram melhor correlação (contagem de números e repetição das frases do CAPE-V), apesar de haver diferença entre a duração dos segmentos sonoros na edição customizada (D2) e na edição pré-definida

Tabela 3. Acurácia do AVQI conforme o contexto de fala

	Correlação de Spearman
Meses do ano	
D1	0,4826
D2	0,5084
D3	0,5500
Números	
D1	0,5796
D2	0,6219
D3	0,6347
Frases do CAPE-V	
D1	0,5036
D2	0,5194
D3	0,5374

Legenda: D1 = Amostra de fala completa + 3 segundos vogal /a/; D2 = Amostra de fala customizada com 3 segundos de segmentos sonoros + 3 segundos vogal /a/; D3 = Amostra de fala com ponto de corte + 3 segundos vogal /a/

(D3), essa diferença não se refletiu nos escores do AVQI, que foram iguais (Tabela 4).

Em relação à acurácia diagnóstica, considerando todos os contextos, o AVQI mostrou ser um índice altamente específico (valores de 0,625 a 1,000) e pouco sensível (valores de 0,413 a 0,881). Considerando o corte de $G < 0,5$, a melhor sensibilidade e valor da curva ROC foi para frases do CAPE-V em D3 (0,72 e 0,578).

Entretanto, esses estímulos de fala apresentaram os menores valores de corte do AVQI (1,1). Não foi possível analisar números, uma vez que a média do escore de G foi maior que 0,5, sendo assim, a análise foi feita considerando como corte $G < 0,68$; houve boa acurácia diagnóstica para números de 1 a 10, a maior sensibilidade foi para números de 1 a 20, ambos os contextos com baixos valores de nota de corte (Tabela 5).

Tabela 4. Comparação dos escores do AVQI e da duração dos segmentos sonoros entre D2 e D3 para os contextos de fala: números e frases do CAPE-V

	Mediana	Mínimo	Máximo	p-valor
Escore do AVQI				
Números				
D2	1,00	-0,04	8,89	0,6343
D3	0,885	-1,68	8,80	
Frases do CAPE-V				
D2	1,105	-1,30	9,06	0,8067
D3	1,10	-1,04	9,02	
Duração do Estímulo				
Números				
D2	3,00	2,91	3,09	0,0029*
D3	2,73	1,39	5,67	
Frases do CAPE-V				
D2	3,00	2,91	3,09	<0,0001*
D3	3,99	2,11	6,84	

Teste de Wilcoxon

Legenda: D2 = Amostra de fala customizada com 3 segundos de segmentos sonoros + 3 segundos vogal /a/; D3 = Amostra de fala com ponto de corte + 3 segundos vogal /a/; *Valores significativos $p < 0,05$

Tabela 5. Dados da Curva ROC do AVQI considerando $G < 0,5$ e $G < 0,68$

	Área Curva ROC	Ponto de Corte	Sensibilidade	Especificidade	LR+	LR-
G < 0,5						
Meses do ano						
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
Números						
NENHUM $G < 0,5$, NÃO ANALISADO						
Frases do CAPE-V						
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
G < 0,68						
Números						
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
Frases do CAPE-V						
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

Legenda: D1 = Amostra de fala completa + 3 segundos vogal /a/; D2 = Amostra de fala customizada com 3 segundos de segmentos sonoros + 3 segundos vogal /a/; D3 = Amostra de fala com ponto de corte + 3 segundos vogal /a/; ROC = Receiver-operating Curve; LR+ = razão de likelihood para resultados positivos; LR- = razão de likelihood para resultados negativos

DISCUSSÃO

O AVQI é um índice que fornece uma medida acústica multiparamétrica, ou seja, considera diversos parâmetros para fornecer um escore único da qualidade vocal^(4,8). Dessa forma, tem melhores chances de refletir de modo mais preciso a qualidade vocal do indivíduo.

Para realizar essa análise e fornecer um escore único, o AVQI utiliza um *script* no programa Praat considerando fala encadeada e vogal sustentada⁽¹¹⁾. Esse índice foi validado em diferentes idiomas. Ao utilizar fala encadeada como uma das amostras a serem inseridas no contexto para análise, evidentemente o código linguístico pode ter influência na acurácia da medida. As validações já realizadas possuem boa correlação com a análise perceptivo-auditiva, variando de 0,794 a 0,929, e boa sensibilidade diagnóstica⁽¹²⁻¹⁵⁾. Nenhum estudo foi ainda realizado com português brasileiro.

A fim de validar o AVQI para o português brasileiro, é necessário primeiramente definir a amostra de fala encadeada a ser utilizada, uma vez que no Brasil não há um texto foneticamente balanceado comumente usado na avaliação vocal. Além disso, a grande parte da população apresenta baixa fluência de leitura, o que pode ser limitante na utilização desse índice.

A presente pesquisa selecionou três amostras de fala encadeada comumente utilizadas na avaliação vocal para potencialmente serem utilizadas na análise do AVQI – meses do ano, números de 1 a 20 e as frases do protocolo CAPE-V.

A confiabilidade intra e interavaliadores para todos os contextos de fala apresentou valores aceitáveis; isso permitiu que as próximas análises fossem realizadas sem necessidade de excluir nenhum dos avaliadores, uma vez que outras pesquisas com o AVQI consideraram aceitáveis valores a partir de 0,53 e de 0,37 para a confiabilidade intra e intersujeito, respectivamente^(4,8,14,15).

O contexto de fala que melhor se correlacionou com a análise perceptivo-auditiva e o escore do AVQI foi números (Tabela 3), que também parece ser o mais utilizado na clínica fonoaudiológica brasileira⁽²³⁻²⁶⁾. Assim, acredita-se que, como os fonoaudiólogos estão mais habituados a avaliar a qualidade vocal com essa amostra de fala, há mais treino com esse tipo de estímulo, logo, a avaliação fica mais precisa e tem maiores chances de uma melhor correlação com o escore do AVQI, que é uma análise objetiva. Apesar do protocolo CAPE-V ser padronizado e estar cada vez mais difundido, muitos profissionais ainda não o utilizam em sua rotina clínica.

O AVQI comumente utiliza a leitura de um texto, que permite avaliar melhor fluência e entonação, parâmetros considerados na análise para definir o escore final do índice acústico. Neste sentido, a fala espontânea seria a melhor amostra de fala para análise do AVQI, contudo é muito difícil padronizá-la para comparar os escores de diferentes sujeitos. A contagem de números, por sua vez, é algo bastante habitual para as pessoas, sendo, em geral, emitida de forma automática e mais natural, o que pode ter influenciado na vantagem obtida. Já as frases do CAPE-V, além de serem novas para os sujeitos, não fazem sentido semântico e não possuem um contexto, o que possibilita uma hesitação em sua repetição, gerando uma fala menos natural.

Para uso no AVQI versão 03.01, recomenda-se que a duração da fala encadeada, sem os segmentos surdos, seja de 3 segundos, igual ao tempo da vogal sustentada, uma vez que essa duração

apresenta melhor precisão e consistência interna com o índice⁽²⁷⁾. Na contagem de números, a melhor correlação foi em D3, de 1 a 10 (Tabela 3), cuja duração média foi de 2,85 segundos. Em D2, esse tempo foi maior (Tabela 4), 3 segundos. Números em D3, ou seja, de 1 a 10, foi a amostra com valores mais próximos de 3 segundos (Tabela 2), portanto, parece que a contagem de 1 a 10 (15 sílabas) se aproxima mais da melhor precisão e consistência interna do AVQI; ainda, não houve diferença entre os escores do AVQI para ambas as durações (Tabela 4). A diferença entre as durações de fala encadeada também foi observada para as frases do CAPE-V e também não se refletiu em diferentes escores do AVQI (Tabela 4).

A validação do AVQI para o japonês considerou 30 sílabas como o mais apropriado para uso do AVQI, uma vez que essa duração foi a que mais se aproximou do resultado obtido com o ponto de corte customizado de 3 segundos de duração de segmentos sonoros⁽¹⁴⁾. No Brasil parece que o ideal para validação são 15 sílabas.

O AVQI é um índice específico (Tabela 5), o que mostra que ele detecta corretamente os sujeitos vocalmente saudáveis, sendo pouco provável que alguém com disфония seja classificado como vocalmente saudável. Em contrapartida, o AVQI mostrou limitações na sensibilidade, dessa forma, muitos sujeitos com alteração vocal são identificados como vocalmente saudáveis.

Os valores obtidos na curva ROC são diferentes dos encontrados em outras validações que apresentaram sensibilidade elevada^(8,13-15,28,29). Considerando $G < 0,5$, a melhor sensibilidade, ainda que muito mais baixa que em outros países, foi para frases do CAPE-V em D3. Entretanto, apresentou o menor valor de corte do AVQI (Tabela 5).

Como não houve no contexto de números média de G inferior a 0,5, considerou-se para análise da acurácia diagnóstica $G < 0,68$; essa conduta não foi observada em nenhuma das outras validações do AVQI até o presente momento. No entanto, por se tratar de um estudo-piloto e ter-se observado a melhor correlação perceptivo-acústica com números, essa pesquisa optou por realizar a análise com outro ponto de corte a fim de comparar a acurácia das frases do CAPE-V, contexto com melhor acurácia considerando $G < 0,5$, com contagem de números, contexto com melhor correlação perceptivo-acústica.

Considerando $G < 0,68$, observou-se boa acurácia diagnóstica para números de 1 a 10; houve, ainda, maiores valores de sensibilidade, apesar de valores de corte extremamente baixos (Tabela 5). Apesar de esses valores ainda serem inferiores às validações de outros países, Japão = 0,915⁽¹⁴⁾ e Bélgica = 0,923⁽¹¹⁾, esses resultados parecem mais promissores para o futuro uso do indicador no Brasil.

Os dados preliminares da presente pesquisa mostram que o AVQI para o português brasileiro parece ser pouco útil como um instrumento de triagem ou de diagnóstico. Vale ressaltar que o objetivo desse estudo não foi validar o AVQI e sim definir qual estímulo de fala será usado em sua validação cujas análises deverão aprofundar esses aspectos.

Sugerimos que a validação e demais análises com o AVQI utilize, no português brasileiro, a amostra de fala encadeada de contagem de números, uma vez que foi a de melhor correlação com a análise perceptivo-auditiva e com melhores valores de acurácia. Além disso, a contagem de números apresentou um valor médio bem próximo ao de melhor precisão e consistência

interna do AVQI e um fácil ponto de corte para edição. Como a contagem é uma amostra de fala comumente adotada em análises perceptivo-auditivas em pesquisas brasileiras⁽²³⁻²⁶⁾, seu uso possibilitará que mais pesquisas retrospectivas sejam realizadas utilizando esse índice, quando validado.

CONCLUSÃO

A amostra de fala encadeada com melhor correlação perceptivo-acústica foi a contagem de número de 1 a 10. As frases do CAPE-V produziram uma melhor acurácia diagnóstica com $G < 0,5$; números apresentou elevada acurácia diagnóstica com $G < 0,68$. Considerando-se que a contagem de números é bastante usual na clínica fonoaudiológica brasileira e que essa amostra de fala foi a que mais se aproximou do valor de melhor precisão e consistência do AVQI, sugere-se que esta seja a amostra utilizada para a futura validação do AVQI, assim como as demais análises utilizando esse índice.

REFERÊNCIAS

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, Dassist-Leite AP, Santos RS, Santos CG Jr, Dias CA, Sartori DJ. Efeitos imediatos do exercício de trato vocal semiocluido 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.

Contribuição dos autores

ME foi responsável pela concepção do estudo, coleta, tabulação, edição dos áudios, análise dos dados e elaboração do manuscrito; LL e ACC foram responsáveis pela concepção do estudo, análise dos dados e revisão do manuscrito; BBL e YM foram responsáveis pela construção do índice, pela concepção do estudo, análise dos dados; MB foi responsável pela proposta e concepção do estudo, análise dos dados e revisão final do manuscrito.