

**UNIVERSIDADE FEDERAL DE MINAS GERAIS**

Faculdade de Medicina

**EMISSÕES VOCAIS ÂNCORAS EM UM SÍTIO DE TREINAMENTO E NA  
AVALIAÇÃO PERCEPTIVO-AUDITIVA DE VOZES: ANÁLISE DA EFETIVIDADE**

Priscila Campos Martins dos Santos

Belo Horizonte

2018

Priscila Campos Martins dos Santos

**EMISSÕES VOCAIS ÂNCORAS EM UM SÍTIO DE TREINAMENTO E NA  
AVALIAÇÃO PERCEPTIVO-AUDITIVA DE VOZES: ANÁLISE DA EFETIVIDADE**

Trabalho apresentado à banca de defesa de  
dissertação do curso de Mestrado em Ciências  
Fonoaudiológicas da Faculdade de Medicina  
da Universidade Federal de Minas Gerais.

Orientadora: Prof<sup>a</sup>. Ana Cristina Côrtes Gama

Coorientador: Prof. Maurílio Nunes Vieira

Belo Horizonte

2018

S237e Santos, Priscila Campos Martins dos.  
Emissões vocais âncoras em um sítio de treinamento e na avaliação perceptivo-auditiva de vozes [manuscrito]: análise da efetividade. / Priscila Campos Martins dos Santos. - - Belo Horizonte: 2018. 45f.: il.  
Orientador (a): Ana Cristina Côrtes Gama.  
Coorientador (a): Maurílio Nunes Vieira.  
Área de concentração: Funcionalidade e saúde da comunicação humana.  
Dissertação (mestrado): Universidade Federal de Minas Gerais, Faculdade de Medicina.

1. Voz. 2. Qualidade da Voz. 3. Treinamento da Voz. 4. Disfonia. 5. Percepção Auditiva. 6. Dissertações Acadêmicas. I. Gama, Ana Cristina Côrtes. II. Vieira, Maurílio Nunes. III. Universidade Federal de Minas Gerais, Faculdade de Medicina. IV. Título.

NLM: WV 501

## **UNIVERSIDADE FEDERAL DE MINAS GERAIS**

*Reitor: Prof. Jaime Arturo Ramírez*

*Vice-Reitora: Prof<sup>a</sup>. Sandra Regina Goulart Almeida*

*Pró-Reitor de Pós-Graduação: Prof. Rodrigo Antônio de Paiva Duarte*

*Pró-Reitora de Pesquisa: Prof<sup>a</sup>. Adelina Martha dos Reis*

### **FACULDADE DE MEDICINA**

Diretor da Faculdade de Medicina: Prof. Tarcizo Afonso Nunes

Vice-Diretor da Faculdade de Medicina: Prof. Humberto José Alves

Coordenadora do Centro de Pós-Graduação: Prof. Luiz Armando Cunha De Marco

Subcoordenador: Prof. Edson Samesima Tatsuo

### **PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS FONOAUDIOLÓGICAS**

Coordenadora: Prof<sup>a</sup>. Stela Maris Aguiar Lemos

Subcoordenadora: Prof<sup>a</sup>. Amélia Augusta de Lima Friche

### **COLEGIADO**

Prof<sup>a</sup>. Sirley Alves da Silva Carvalho- Titular  
Prof<sup>a</sup>. Luciana Macedo de Resende- Suplente

Prof<sup>a</sup>. Letícia Caldas Teixeira- Titular  
Prof<sup>a</sup>. Ana Cristina Cortes Gama- Suplente

Prof<sup>a</sup>. Stela Maris Aguiar Lemos- Titular  
Prof<sup>a</sup>. Adriane Mesquita de Medeiros- Suplente

Prof<sup>a</sup>. Andréa Rodrigues Motta- Titular  
Prof<sup>a</sup>. Helena Gonçalves Becker- Suplente

Prof<sup>a</sup>. Amélia Augusta de Lima Friche- Titular  
Prof<sup>a</sup>. Patrícia Cotta Mancini- Suplente

Daniele Veloso de Castro Ferreira– discente titular  
Thalita Evaristo Couto Dias – discente suplente

## **Declaração de Defesa**

Prof<sup>a</sup>. Ana Cristina Côrtes Gama –UFMG

Prof. Maurílio Nunes Vieira –UFMG

Prof<sup>a</sup>. Letícia Caldas Teixeira - UFMG

Dra. Iara Barreto Bassi – FHEMIG - HJXXIII

Prof<sup>a</sup>. Glaucya Maria Vicente Madazio – Centro de Estudos da Voz

## **Resumo da Dissertação / Descritores**

### **Introdução**

A análise perceptivo-auditiva da voz humana é a principal forma de avaliação na clínica vocal e nas pesquisas fonoaudiológicas na área da voz. Várias são suas vantagens: permite descrições perceptivas que abrangem diversos parâmetros vocais, é um método rápido, indolor e confortável ao paciente, e, além disso, não depende de equipamentos, gerando um baixo custo. Porém, a avaliação perceptivo-auditiva da voz depende de vários padrões internos, o que provavelmente justifica a alta variabilidade na concordância intra e interavaliadores e torna essa avaliação subjetiva.

Algumas tarefas são apontadas como fatores que podem otimizar o resultado da avaliação perceptivo-auditiva: utilização de escalas padronizadas, de emissões vocais âncoras, de emissões vocais sintetizadas e o próprio treinamento auditivo.

### **Objetivo**

Analisar o efeito do treinamento perceptivo-auditivo com emissões âncoras de vozes naturais na concordância interavaliadores da avaliação da qualidade vocal. Analisar o efeito da avaliação perceptivo-auditivo com emissões âncoras de vozes sintetizadas na concordância intra e interavaliadores da avaliação da qualidade vocal.

### **Métodos**

Trata-se de um estudo de natureza quantitativa.

Este projeto foi estruturado em dois blocos. No primeiro bloco, desenvolveu-se um sítio de treinamento composto por emissões âncoras de vozes naturais, em que os parâmetros da escala GRBASI (R – rugosidade, B – soprosidade, A – astenia, S – tensão e I – instabilidade) foram treinados e avaliados. Foram selecionados 15 avaliadores não experientes para realizar o treinamento e a avaliação vocal, sujeitos de ambos os sexos com idade entre 18 e 28 anos. Este bloco foi composto por duas Atividades: Atividade Controle e Atividade Treinamento Auditivo. Os mesmos avaliadores realizaram as duas atividades, com um intervalo de 15 dias entre elas, sendo que a ordem de realização das atividades foi sorteada para cada participante.

Na Atividade Controle, os avaliadores deveriam, na primeira etapa, avaliar 57 vozes classificando-as quanto ao grau de desvio (0 – ausência de desvio, 1 – desvio leve, 2 – desvio moderado e 3 – desvio intenso) de cada parâmetro da escala GRBASI. Na segunda etapa, deveriam realizar a leitura de um texto. Por fim, na terceira etapa, deveriam realizar a mesma avaliação realizada na primeira etapa. As vozes eram aleatorizadas a cada etapa.

Na Atividade Treinamento Auditivo, a primeira e terceira etapas eram iguais à da Atividade Controle. Na segunda etapa, os parâmetros RBASI foram treinados separadamente. No treinamento, os avaliadores deviam escutar a voz a ser avaliada, escutar a emissão vocal âncora e, novamente, escutar a emissão vocal a ser avaliada e classificá-lo quanto ao grau de desvio do parâmetro em treinamento. Para cada parâmetro, foram avaliadas quatro vozes, exceto para o parâmetro A, em que foram avaliadas duas vozes, totalizando 56 vozes avaliadas no treinamento. As âncoras eram compostas por vozes naturais.

No segundo bloco, foi desenvolvido um Calibrador Auditivo, composto por duas atividades em que os parâmetros R e B da escala GRBASI foram classificados como 0 – ausência de desvio, 1 – desvio leve, 2 – desvio moderado ou 3 – desvio intenso. Trinta e dois avaliadores inexperientes, sujeitos de ambos os sexos com idade a partir de 18 anos, foram selecionados para este bloco e realizaram as duas atividades, com um intervalo de 15 dias entre elas. A ordem de realização das atividades foi sorteada para cada participante.

Na Atividade Calibrador Auditivo, 25 vozes foram analisadas com o apoio de emissões âncoras de vozes sintetizadas. Os avaliadores escutaram a voz natural a ser avaliada, em seguida escutaram as emissões âncoras de vozes sintetizadas para cada grau do parâmetro R, e novamente escutaram a voz a ser avaliada, classificando-a em 0 – ausência de desvio, 1 – desvio leve, 2 – desvio moderado ou 3 – desvio intenso para este parâmetro. Repetiram o processo para classificar a mesma voz quanto ao parâmetro B.

Na Atividade Calibrador Inativo, as mesmas vozes foram avaliadas, porém, sem o apoio das emissões vocais âncoras.

Para análise da concordância intra e interavaliadores foi utilizado o coeficiente AC1 de Gwet e Kappa. Para comparação entre as concordâncias foram utilizados o intervalo de confiança (IC) e a variação percentual da concordância. Para realizar a análise estatística foi utilizado o *software* R versão 3.3.1 Em todas as análises foi considerado um nível de significância de 5%.

## **Resultados**

Como resultado do primeiro bloco da pesquisa, verificou-se que houve uma diminuição da concordância interavaliadores na Atividade Controle para os

parâmetros RBS e I. Na Atividade Treinamento Auditivo com emissões âncoras de vozes naturais, a concordância interavaliadores aumentou para todos os parâmetros, exceto para astenia, mas essa melhora não atingiu significância estatística. Como resultado do segundo bloco da pesquisa, verificou-se que a concordância interavaliadores foi significativamente maior para o grau intenso do parâmetro soprosidade na Atividade Calibrador Ativo quando comparada à Atividade Calibrador Inativo, assim como a concordância intra-avaliadores do parâmetro R.

### **Conclusão**

O treinamento associado a emissões âncoras de vozes naturais sugerem um aumento da concordância interavaliadores na avaliação perceptivo-auditiva da voz, indicando que novas referências internas foram estabelecidas.

O uso de emissões âncoras de vozes sintetizadas na avaliação vocal melhora a concordância intra e interavaliadores na análise perceptivo-auditiva da voz, podendo ser usados na clínica fonoaudiológica.

**Descritores:** voz, qualidade da voz, disfonia, treinamento da voz, percepção auditiva.

## **Abstract / Keywords**

### **Introduction**

The auditory-perceptual analysis of the human voice is the main way of evaluation in vocal clinical and speech-language-pathology research. There are several advantages to this: it allows perceptual descriptions that cover several vocal parameters, it is a quick, painless and comfortable method to the patient, and besides, it does not depend on equipment, generating a low cost. However, the auditory-perceptual assessment of voice depends on several internal standards, which probably justifies the high variability in intra and inter-rater agreement and makes this evaluation subjective.

Some tasks are pointed out as factors that can optimize the result of auditory-perceptual evaluation: the use of scales, anchor stimuli, synthesized stimuli and auditory training.

### **Purpose**

To analyze the effect of auditory-perceptual training with anchor stimuli of natural voices in the inter-rater agreement of vocal evaluation. To analyze the effect of auditory-perceptual evaluation with anchor stimuli of synthesized voices in the intra and inter-rater agreement of vocal evaluation.

### **Methods**

This is a quantitative study.

This project was structured in two blocks. In the first block, it was developed a training site composed of anchors of natural voices, in which the GRBASI (R -

roughness, B - breathiness, A - asthenia, S - tension and I - instability) parameters scale were trained and evaluated. Fifteen inexperienced evaluators were selected to perform training and vocal assessment, subjects of both genders aged between 18 and 28 years old. This block was composed of two Activities: Control Activity and Auditory Training Activity. The same evaluators performed the two activities, with an interval of 15 days between them, and the order of accomplishment of activities was drawn for each participant.

In the Control Activity, the evaluators should, in the first stage, evaluate 57 voices classifying them according to the degree of deviation (0 – no deviation, 1 – light deviation, 2 – moderate deviation or 3 – intense deviation) of each GRBASI parameter scale. In the second stage, they should read a text. Finally, in the third stage, they should carry out the same evaluation of the first stage. Voices were randomized at each stage.

In the Auditory Training Activity, the first and third stages were the same as the Control Activity. In the second stage, the RBASI parameters were trained separately. In the training, the evaluators should listen to the voice to be evaluated, listen to the anchor stimulus and again listen to the stimulus to be evaluated and classify it as to the degree of deviation of the parameter under training. For each parameter, four voices were evaluated, except for the parameter A, in which two voices were evaluated, totaling 56 voices. The anchors stimuli were composed by natural voices.

In the second block, an Auditory Calibrator was developed, consisting of two activities in which the roughness and breathiness parameters of the GRBASI scale were classified as 0 – no deviation, 1 – light deviation, 2 – moderate deviation or 3 – intense deviation. Thirty two inexperienced evaluators, subjects of both genders aged

from 18 years old, were selected for this block and performed the two activities, with an interval of 15 days between them. The order of accomplishment of the activities was drawn for each participant.

In Evaluation Activity, 25 voices were evaluated with the support of anchored stimuli of synthesized voices. The evaluators should listen to the natural voice to be evaluated, listen to the anchors of synthesized voices for each degree of the R parameter, and, again, listen to the voice being evaluated, classifying it as 0 – no deviation, 1 – light deviation, 2 – moderate deviation or 3 – intense deviation for this parameter. They repeated the process to classify the same voice for the breathiness parameter (B).

In the Control Activity the same voices were evaluated, however, without the support of the anchor stimuli.

For the analysis of the intra and inter-rater agreement, the AC1 coefficient of Gwet and Kappa was used, and to compare the concordances, the confidence interval (CI) and the percentage variation of agreement were used. Statistical analysis was performed using R version 3.3.1 software. A significance level of 5% was considered for all analyzes.

## **Results**

As a result of the first block of this research, it was found that there was a decrease of inter-rater agreement in the Control Activity for the RBS and I parameters. In the Auditory Training Activity with natural anchor stimuli, the inter-rater agreement increased for all parameters except for asthenia, but this improvement did not reach statistical significance. As a result of the second block of this research, it was verified that the inter-rater agreement was significantly higher for

the intense degree of the breathiness parameter in Evaluation Activity when compared to Control Activity, as well as the intra-rater agreement of the roughness parameter.

## **Conclusion**

Training associated with natural voice anchor stimuli suggests an increase in inter-rater agreement in the auditory-perceptual evaluation of voice, indicating that new internal references have been established.

The use of vocal anchors stimuli synthesized in the vocal evaluation seems to improve the intra and inter-rater agreement in auditory-perceptual analysis of the voice.

**Keywords:** voice, voice quality, dysphonia, voice training, auditory perception.

## Sumário

<b>Considerações Iniciais.....</b>	<b>15</b>
<b>Objetivos.....</b>	<b>22</b>
Objetivo Geral.....	22
Objetivos Específicos.....	22
<b>Métodos.....</b>	<b>23</b>
<b>Resultados e Discussão .....</b>	<b>43</b>
Artigo 1.....	44
<i>Summary</i> .....	44
<i>Introduction</i> .....	44
<i>Materials and Methods</i> .....	45
<i>Results</i> .....	47
<i>Discussion</i> .....	47
<i>Conclusion</i> .....	49
<i>References</i> .....	49
Artigo 2.....	50
<i>Folha de Rosto</i> .....	50
<i>Resumo</i> .....	51
<i>Introdução</i> .....	52
<i>Métodos</i> .....	56
<i>Resultados</i> .....	65
<i>Discussão</i> .....	69
<i>Conclusão</i> .....	74
<i>Referências</i> .....	75

<b>Considerações Finais</b> .....	79
<b>Anexos / Apêndices</b> .....	81
Anexo 1 - Aprovação do COEP.....	81
Anexo 2 - Resolução nº01/2015, de 26 de março de 2015 - Programa de Pós-Graduação em Ciências Fonoaudiológicas.....	82
Anexo 3 - Normas do periódico <i>Journal of Speech, Language, and Hearing Research</i> (JSLHR).....	84

A voz pode ser medida objetivamente de várias formas, sendo a avaliação perceptivo-auditiva a análise mais utilizada na clínica fonoaudiológica vocal. Um dos principais aspectos que traz o paciente à clínica vocal é a alteração da voz que o próprio paciente ou outros percebem auditivamente, ou seja, as mudanças de natureza perceptiva na qualidade vocal<sup>(1)</sup>. Na avaliação perceptivo-auditiva a qualidade vocal é classificada segundo a impressão que um ouvinte tem da voz de um falante como um todo, por isso, acredita-se que a percepção auditiva de uma voz é de natureza holística e integrativa<sup>(2)</sup>.

Diversas são as vantagens da análise perceptivo-auditiva, dentre elas o fato desta avaliação permitir que sejam realizadas descrições perceptivas que abrangem diversos parâmetros vocais. Suas vantagens também estão diretamente relacionadas às limitações das abordagens instrumentais da mensuração da voz – enquanto a avaliação perceptivo-auditiva possui como instrumento básico a audição e avalia diversos parâmetros da voz, as abordagens instrumentais da mensuração da voz, como a avaliação acústica, avaliam o sinal sonoro e necessitam de diversos instrumentos e cuidados, que vão desde às condições de registro das vozes, os protocolos e softwares de gravação, até a análise do sinal sonoro<sup>(1)</sup>. Por fim, a análise perceptivo-auditiva é vantajosa por tratar-se de um método rápido quando comparado a outros métodos, confortável ao paciente e de baixo custo<sup>(3)</sup>.

Apesar de ser a avaliação padrão na clínica vocal, há muitas críticas à sua subjetividade e às suas limitações<sup>(1)</sup>. Embora as medidas da avaliação perceptivo-auditiva sejam fáceis de serem obtidas e não necessitem de muita instrumentalização, sofrem baixa sensibilidade, confiabilidade e concordância entre avaliadores<sup>(4)</sup>. A confiabilidade das avaliações auditivas da qualidade vocal é

questão central na pesquisa de voz, por apresentar uma alta variabilidade da concordância intra e interavaliadores. Muito se tem pesquisado sobre o assunto a fim de garantir a confiabilidade dessa avaliação<sup>(1)</sup>.

A avaliação perceptivo-auditiva da voz depende de vários padrões internos, o que provavelmente justifica essa alta variabilidade da confiabilidade dessas avaliações<sup>(5)</sup>. O fato de as vozes não serem homogêneas, ou seja, frequentemente serem caracterizadas pela presença de mais de um parâmetro perceptivo-auditivo concomitantemente, como rugosidade e sopro, sopro e tensão, ou este três parâmetros principais juntos, impacta na confiabilidade da avaliação perceptivo-auditiva<sup>(3,5-7)</sup>. Além disso, estudos revelaram que os seguintes fatores podem interferir na confiabilidade da avaliação: experiência pessoal, profissional e treinamentos anteriores do avaliador; suas preferências; a tarefa de fala, erros casuais, além de outros aspectos<sup>(1,5,8)</sup>.

Algumas tarefas são apontadas como fatores que podem otimizar o resultado da avaliação perceptivo-auditiva: utilização de escalas, de emissões vocais âncoras, de emissões vocais sintetizadas e o próprio treinamento auditivo<sup>(7)</sup>.

O uso de escalas de avaliação perceptivo-auditiva reduz potencialmente a variabilidade e as inconsistências dessa análise<sup>(9)</sup>. Estudos anteriores<sup>(10,11,12)</sup> apontaram o amplo uso da escala GRBAS<sup>(13)</sup> ou GRBAS<sup>(14)</sup> na clínica e pesquisa fonoaudiológica. A GRBAS<sup>(13)</sup> é uma escala composta por cinco parâmetros: grau geral de desvio vocal (G), rugosidade (R) – qualquer irregularidade perceptível durante a produção vocal, sopro (B) – qualquer escape de ar audível durante a produção vocal, astenia (A) – percepção de fraqueza vocal, e tensão (S) – percepção de esforço vocal. Posteriormente, foi acrescentado à escala o parâmetro instabilidade (I) – qualquer flutuação na emissão vocal<sup>(14)</sup>. É uma escala numérica de

quatro pontos em que cada parâmetro deve ser avaliado de acordo com a intensidade do desvio em: ausência de desvio (0), desvio leve (1), desvio moderado (2) ou desvio intenso (3). No parâmetro G deve ser avaliado o grau geral do desvio vocal em: voz neutra (0), desvio leve (1), desvio moderado (2) ou desvio intenso (3)<sup>(13-14)</sup>.

Para aumentar a confiabilidade da avaliação perceptivo-auditiva, alguns pesquisadores sugerem o uso de emissões vocais âncoras controladas<sup>(5,8,15-17)</sup>. As âncoras são emissões vocais pré-definidas e selecionadas como representante de um determinado tipo e / ou grau de desvio e pode ser natural ou sintetizada.

As âncoras compostas por emissões vocais naturais são vozes humanas pré selecionadas, em concordância por mais de um avaliador, como representantes de cada grau de cada parâmetro avaliado. Seu uso tem sido apontado como uma estratégia para melhorar a concordância intra e interavaliadores, apesar de suas limitações, como o fato da maioria das vozes serem caracterizadas por mais de um parâmetro perceptivo-auditivo concomitantemente, ou que dificulta ou, por vezes, até mesmo impossibilita a seleção de vozes humanas âncoras que possuam apenas o parâmetro que está representando<sup>(15-17)</sup>.

O uso de emissões vocais sintetizadas é outra alternativa para otimizar o resultado da avaliação perceptivo-auditiva, uma vez que possibilita a manipulação dos seus parâmetros acústicos conforme desejar ou necessitar, permitindo a análise de cada parâmetro vocal separadamente<sup>(5)</sup>. Apesar do uso promissor de vozes sintetizadas, essa ainda não é uma prática comum, devido à dificuldade de produzir vozes que sejam consideradas naturais pelo ouvinte. Porém, estudo apontou alta qualidade, ou seja, naturalidade, das amostras de vozes sintetizadas quando comparadas com as vozes humanas<sup>(18)</sup>.

A literatura indica que o treinamento auditivo também aumenta a confiabilidade da avaliação perceptivo-auditiva<sup>(8,15)</sup>. Segundo pesquisas na área<sup>(19)</sup>, os estímulos de impressão e de diferenciação são dois processos que estão envolvidos no aprendizado perceptivo. Os estímulos de impressão são formados pela exposição repetida ao estímulo, desenvolvendo detectores ou padrões internos nos avaliadores. Já o estímulo de diferenciação refere-se à distinção entre os parâmetros por meio da apresentação de cada parâmetro vocal separadamente, o que é possível com a voz sintetizada. Outros pesquisadores<sup>(20,21)</sup> argumentam também que quando é fornecido *feedback* em um treinamento, esse processo de aprendizagem de diferenciação é acelerado<sup>(6)</sup>.

Diante disso, este projeto foi estruturado em dois blocos. No primeiro bloco, desenvolveu-se um sítio de treinamento composto por emissões âncoras de vozes naturais, em que os parâmetros da escala GRBASI foram treinados e avaliados. No segundo bloco desenvolveu um Calibrador Auditivo, composto por duas atividades em que os parâmetros R e B da escala GRBASI foram avaliados. Numa atividade, a avaliação vocal foi realizada sem o apoio de emissões vocais âncoras e, noutra, a mesma avaliação foi realizada com o apoio de emissões âncoras de vozes sintetizadas.

O presente estudo foi aprovado pelo Comitê de Ética em Pesquisa (COEP-UFMG) da Universidade Federal de Minas Gerais, sob o número CAAE 37872314.2.0000.5149 (Anexo 1). Todos os indivíduos avaliadores foram informados quanto aos objetivos e procedimentos do estudo e assinaram, mediante leitura prévia e esclarecimento de dúvidas, o termo de consentimento livre e esclarecido.

Os resultados encontrados neste projeto de pesquisa, de acordo com a Resolução nº 01/2015 de março de 2015 (Anexo 2), do programa de Mestrado em Ciência Fonoaudiológicas, serão apresentados em formato de dois artigos científicos, a saber:

O artigo 1 intitulado: *Effect of Auditory-Perceptual Training with Natural Voice Anchors on Vocal Quality Evaluation*. Aceito para publicação pelo periódico *Journal of Voice* em 26 de outubro de 2017, e disponível *in press*.

O artigo 2 intitulado: Efeito de emissões âncoras de vozes sintetizadas na avaliação perceptivo-auditiva da voz. Será submetido à Revista *Journal of Speech, Language, and Hearing Research* (JSLHR), e foi estruturado de acordo com as normas do periódico (Anexo 3).

### **Referências Bibliográficas**

1. Behlau M, organizadora. *Voz: o livro do especialista*. Rio de Janeiro: Revinter; 2001.
2. Zraick RI, Wendel K, Smith-Olinde L. The Effect os Speaking Task on Perceptual Judgment of the Severity os Dysphonic Voice. *J Voice*, 2005;19(4):574-581.
3. Oates J. Auditory-perceptual evaluation of disordered vocal quality—pros, cons and future directions. *Folia Phoniatr Logop*. 2009;61:49–56.
4. Patel S, Shrivastav R. Perception of Dysphonic Vocal Quality: Some Thoughts and Research Update: Perspectives on Voice and Voice Disorders. *ASHA Divison 3*. 2007;17(2):3-6.

5. Yiu EML, Murdoch B, Hird K, Lau P. Perception of synthesized voice quality in connected speech by Cantonese speakers. *J Acoust Soc Am.* 2002;112(3 Pt 1):1091–1101.
6. Chan KMK, Yiu EML. The effects of anchors and training on the reliability of perceptual voice evaluation. *J Speech Lang Hear Res.* 2002;45:111–126.
7. Chan KMK, Yiu EML. A comparison of two perceptual voice evaluation training programs for naive listeners. *J Voice.* 2006;20:229–241.
8. Solomon NP, Helou LB, Stojadinovic A. Clinical versus laboratory ratings of voice using the CAPE-V. *J Voice.* 2011;25:e7–e14.
9. Nemr K, Simões-Zenari M, Cordeiro GF, Tsuji D, Ogawa AI, Ubrig MT, et al. GRBAS and Cape-V Scales: high reliability and consensus when applied at different times. *J Voice.* 2012;26(6):812e17-22.
10. Vieira MN, Sansão JPH, Yehia HC. Measurement of signal-to-noise ratio in dysphonic voices by image processing of spectrograms. *Speech Commun.* 2014;61-62:17-32.
11. Silva RSA, Simões-Zenari M, Nemr NK. Impacto de treinamento auditivo na avaliação perceptivo-auditiva da voz realizada por estudantes de Fonoaudiologia. *J Soc Bras Fonoaudiol.* 2012;24(1):19-25.
12. Chaves CR, Campbell M, Gama ACC. The influence of native language on auditory-perceptual evaluation of vocal samples completed by brazilian and Canadian SLPs. *J Voice.* 2016; in press.
13. Hirano M. *Clinical examination of voice.* New York: Springer-Verlag; 1981.
14. Dejonckere PH, Leback J. Acoustic, perceptual, aerodynamic and anatomical correlations in voice pathology. *ORL J Otorhinolaryngol Relat Spec.* 1996;58:326-32.

15. Awan SN, Lawson LL. The effect of anchor modality on the reliability of vocal severity ratings. *J Voice*. 2009;23:341–345.
16. Eadie TL, Kapsner-Smith M. The effect of listener experience and anchors on judgments of dysphonia. *J Speech Hear Res*. 2011;54:430–447.
17. Sofranko JL, Prosek RA. The effect of the levels and types of experience on judgment of synthesized voice quality. *J Voice*. 2014;28:24–35.
18. Englert M, Madazio G, Gielow I, Lucero J, Behlau M. Perceptual error identification of human and synthesized voices. *J Voice*. 2016;30(5):639.e17-23.
19. Goldstone RL. Perceptual learning. *Ann Rev Psychol*. 1998;49:585-612.
20. Gibson EJ. *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts. 1969.
21. Hall G. *Perceptual and associative learning*. Oxford, UK: Clarendon. 1991.

**Objetivo geral**

Analisar o efeito das emissões âncoras no treinamento e na avaliação perceptivo-auditiva de vozes.

**Objetivos específicos**

- Analisar se a utilização de **emissões âncoras de vozes naturais no treinamento** melhoram a concordância interavaliador da análise perceptivo-auditiva;
- Avaliar se o uso de **emissões âncoras de vozes sintetizadas na avaliação** melhoram a concordância intra e interavaliador da análise perceptivo-auditiva.

A presente pesquisa foi aprovada pelo Comitê de Ética em Pesquisa (COEP) da Universidade Federal de Minas Gerais sob o parecer CAAE – 37872314.2.0000.5149. Trata-se de um estudo de natureza quantitativa.

O estudo foi organizado em dois blocos:

1. Sítio de treinamento perceptivo-auditivo;
2. Calibrador auditivo.

### **Bloco 1: Sítio de treinamento perceptivo-auditivo**

As atividades deste bloco foram elaboradas a fim de analisar o efeito do **treinamento perceptivo-auditivo com emissões âncoras de vozes naturais** na concordância interavaliadores da avaliação vocal. Para compor este bloco foi criado e disponibilizado pelas pesquisadoras um sítio de treinamento e avaliação perceptivo-auditiva de vozes naturais, descrito posteriormente. Antes de iniciar as atividades, os avaliadores deviam ler o Termo de Consentimento Livre e Esclarecido (TCLE) e, caso concordassem, deveriam selecionar a opção “Aceito” para prosseguir na participação da pesquisa. Em seguida, responderam a um breve questionário (Figura 1) fornecendo dados sobre sua experiência em treinamento auditivo, idade, profissão e acuidade auditiva; e receberam uma apresentação inicial da pesquisa (Figura 2).

Login

# Treinamento Perceptivo-Auditivo

## Questionário

\*Obrigatório

**Gênero \***

Masculino  
 Feminino

**Idade \***

**Profissão \***

Fonoaudiólogo  
 Músico  
 Estudante  
 Outro:

Figura 1 – Questionário.

Login / Registre-se

# Treinamento Perceptivo-Auditivo

Seja bem vindo ao sítio de treinamento perceptivo-auditivo de vozes!

A pesquisa será realizada por meio de duas atividades: Atividade Controle e Atividade Treinamento Auditivo. Você deverá escolher qual atividade será feita nesse momento, e deverá fazer a outra atividade exatamente daqui a 15 dias.

Você passará pelas seguintes fases:

- 1º. Atividade Controle, contendo as seguintes etapas:
  - Avaliação pré-intervalo. Nessa etapa você deverá realizar a classificação das vozes quanto aos parâmetros e ao grau de desvio que ela corresponde.
  - Intervalo de 20 minutos. Durante todo esse intervalo você deverá realizar a leitura de um texto.
  - Avaliação pós-intervalo. Você deverá novamente, após o intervalo, realizar a classificação das vozes quanto aos parâmetros e ao grau de desvio que ela corresponde.
- 2º. Atividade Treinamento Auditivo, contendo as seguintes etapas:
  - Avaliação pré-treinamento. Nessa etapa você deverá realizar a classificação das vozes quanto aos parâmetros e ao grau de desvio que ela corresponde.
  - Treinamento. Nessa etapa cada parâmetro será treinado individualmente. Você deverá classificar as vozes segundo o grau de desvio para o parâmetro treinado.
  - Avaliação pós-treinamento. Você deverá novamente, após o treinamento, realizar a classificação das vozes quanto aos parâmetros e ao grau de desvio que ela corresponde.

Haverá um intervalo de 15 dias entre as duas atividades. A ordem de realização das atividades será sorteada pelo pesquisador. Selecione abaixo a atividade que foi sorteada para você fazer nessa fase.

Todo o treinamento deverá ser realizado utilizando um fone de ouvido supra-auricular (ficam sobre os ouvidos e são ligados por uma cordão).


 Este site foi criado por WIX.com. Crie seu site GRÁTIS >>

Figura 2 – Apresentação inicial da pesquisa.

Após esse processo inicial, os avaliadores realizaram as duas atividades do sítio: “Atividade treinamento auditivo” e “Atividade controle”. A ordem de realização das atividades foi sorteada para cada participante, sendo que a segunda atividade foi realizada exatamente 15 dias após a primeira. É possível observar na literatura o uso de um intervalo de pelo menos uma semana entre atividades de avaliação, a fim de evitar qualquer memorização<sup>(1,2,3)</sup>. Cada atividade foi composta por três etapas (Figura 3).

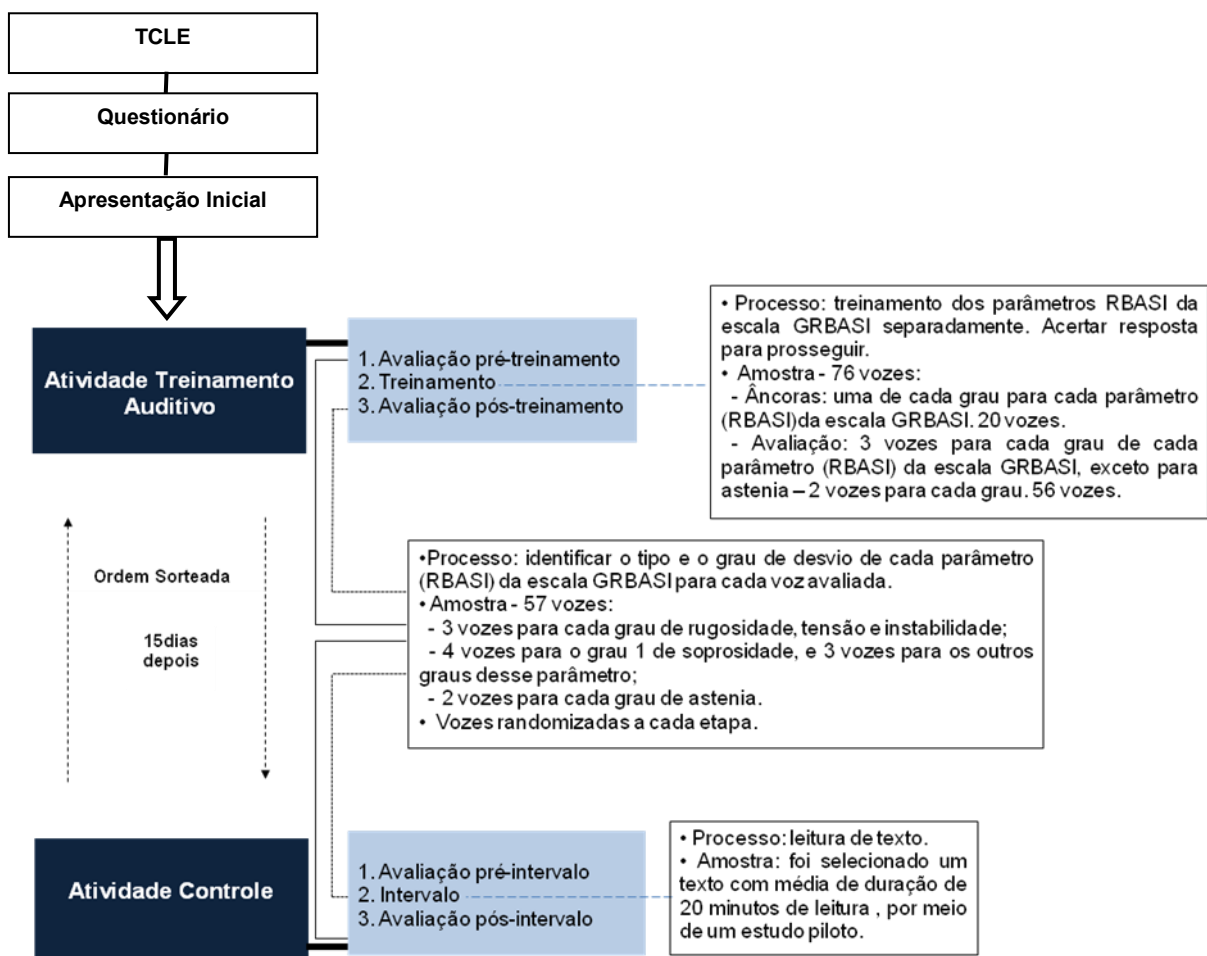


Figura 3 – Fluxograma do sítio de treinamento perceptivo-auditivo da voz.

Cada atividade será descrita a seguir.

## 1. Atividade Treinamento Auditivo

Foi estruturada três etapas, descritas abaixo, sendo que, uma delas consiste em um treinamento com o uso emissões âncoras de vozes naturais.

### 1.1. Processo.

Essa atividade foi estruturada em três etapas.

A) Avaliação pré-treinamento: nessa etapa os avaliadores escutaram as vozes naturais e as classificaram segundo os parâmetros predominantes rugosidade (R), soprosidade (B), astenia (A), tensão (S) e instabilidade (I) em 0 – ausência de desvio, 1 – desvio leve, 2 – desvio moderado ou 3 – desvio intenso (Figura 4).

**ATIVIDADE TREINAMENTO AUDITIVO**

### Avaliação Pré-Treinamento

Classifique as vozes segundo os parâmetros e o grau de desvio. Lembre-se que as vozes são humanas, e por isso, podem apresentar mais de um parâmetro.

Rugosidade - Qualquer irregularidade perceptível durante a produção vocal  
Soprosidade - Qualquer escape de ar audível durante a produção vocal  
Astenia - Percepção de fraqueza vocal  
Tensão - Percepção de esforço vocal  
Instabilidade - Qualquer flutuação na emissão vocal

**\*Obrigatório**

Voz 1 \*

	0 - Neutro	1 - Leve	2 - Moderado	3 - Intenso
Rugosidade	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Soprosidade	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Astenia	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Tensão	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Instabilidade	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Voz 2 \*

0 - Neutro 1 - Leve 2 - Moderado 3 - Intenso

Este site foi criado por WIX.com. Crie seu site GRÁTIS >>

Figura 4 – Etapa Avaliação pré-treinamento da Atividade Treinamento Auditivo

B) Treinamento: foram apresentados separadamente os treinamentos dos parâmetros predominantes R, B, A, S e I. Os avaliadores escutaram as emissões de vozes naturais a serem avaliados e, posteriormente, as emissões vocais âncoras, também compostos por vozes naturais. Classificaram então as vozes avaliadas segundo a intensidade do desvio do parâmetro em treinamento. Caso a resposta estivesse incorreta, os avaliadores deveriam fazer novas tentativas até que a resposta estivesse correta (Figura 5).



Figura 5 – Etapa Treinamento da Atividade Treinamento Auditivo

C) Avaliação pós-treinamento: essa etapa foi realizada da mesma forma que a etapa avaliação pré-treinamento.

A definição escrita dos parâmetros e dos graus de desvio vocal foi disponibilizada durante todas as etapas da Atividade Treinamento Auditivo: rugosidade – qualquer irregularidade perceptível durante a produção vocal; soprosidade – qualquer escape de ar audível durante a produção vocal; astenia – percepção de fraqueza vocal; tensão – percepção de esforço vocal; instabilidade – qualquer flutuação na emissão vocal.

Os avaliadores podiam escutar as vozes quantas vezes julgassem necessário em todas as etapas. Foi utilizado fone de ouvido supra-auricular modelo *Multilaser Vibe Headphone* estéreo durante todos os procedimentos. Embora as avaliações tenham sido realizadas em uma sala com vários computadores, ou seja, vários avaliadores no mesmo momento, e sob a supervisão dos pesquisadores, foram realizadas individualmente. Os pesquisadores conduziram os avaliadores, mas não podiam interferir na execução das atividades, a fim de se observar puramente o efeito do treinamento.

## 1.2. Seleção das emissões vocais para avaliação.

Para compor a amostra de vozes naturais para avaliação nas etapas avaliação pré e pós-treinamento e treinamento, utilizou-se o banco de vozes do ambulatório da Universidade Federal de Minas Gerais, formado por 381 vozes, amostras da emissão da vogal /a/ sustentada de forma habitual, de indivíduos de ambos os gêneros com idade a partir de 18 anos. Todos os avaliadores assinaram o TCLE.

Duas avaliadoras, fonoaudiólogas, especialistas em voz, com mais de cinco anos de experiência em avaliação perceptivo-auditiva, analisaram individualmente as vozes, utilizando o fone de ouvido supra-auricular modelo *Multilaser Vibe*

*Headphone* estéreo. Classificaram as vozes conforme o parâmetro predominante (R, B, A, S ou I) em 0 – ausência de desvio, 1 – desvio leve, 2 – desvio moderado ou 3 – desvio intenso, por meio da escala GRBASI.

Foram considerados os seguintes critérios de inclusão: vozes naturais de sujeitos do sexo feminino e masculino, com idade a partir de 18 anos, com um parâmetro predominante de variados graus de desvio vocal; vozes que apresentaram a mesma classificação pelas duas avaliadoras.

Foram selecionados para as etapas avaliação pré e pós-treinamento três emissões vocais para cada grau dos parâmetros predominantes R, B, S e I, sendo que um grau de um dos parâmetros recebeu quatro emissões, totalizando 49 vozes. Para o parâmetro predominante A, foram selecionados duas emissões vocais para cada grau, totalizando oito vozes. A menor prevalência do parâmetro astenia justifica o menor número de emissões vocais. Sendo assim, o total de vozes para essas etapas do Aplicativo N foi de 57. Acrescentou-se 20% das vozes para posterior análise da concordância intra-avaliador, totalizando 68 vozes.

Para a etapa treinamento, foram selecionados três emissões vocais para cada grau – ausência de desvio, desvio leve, desvio moderado e desvio intenso, de cada parâmetro – R, B, S, I. Para o parâmetro astenia, foram selecionadas duas emissões vocais de cada grau, totalizando 56 vozes para essa etapa.

Em todas as etapas do sítio, assim como na publicação da presente pesquisa, as vozes foram identificadas por números.

### 1.3. Seleção das emissões vocais âncoras para o treinamento.

A amostra das emissões âncoras foi composta por vozes naturais. Utilizou-se o mesmo banco de vozes usado para seleção da amostra das emissões vocais a

serem avaliados nas etapas avaliação pré e pós-treinamento e treinamento. As vozes foram previamente classificadas quanto ao parâmetro predominante – R, B, A, S e I em ausência de desvio, desvio leve, desvio moderado e desvio intenso, por meio da escala GRBASI, por duas fonoaudiólogas, especialistas em voz, com mais de cinco anos de experiência em avaliação perceptivo-auditiva. A análise foi feita individualmente, utilizando o fone de ouvido supra-auricular modelo *Multilaser Vibe Headphone* estéreo.

Foram considerados os seguintes critérios de inclusão: vozes naturais de sujeitos do sexo feminino e masculino, com idade a partir de 18 anos, com um parâmetro predominante de variados graus de desvio vocal; vozes que apresentaram a mesma classificação pelas duas avaliadoras.

Foi selecionada como emissão vocal âncora uma amostra para cada grau de desvio de cada parâmetro. Sendo assim, o total de emissões âncoras para a etapa treinamento foi de 20 vozes.

## **2. Atividade Controle**

### **2.1. Processo.**

Optou-se pela inclusão dessa atividade para permitir a análise puramente do efeito do treinamento, eliminando a possibilidade de uma melhora na concordância devido ao acaso. Essa atividade também foi estruturada em três etapas.

- A) Avaliação pré-intervalo: foi realizada da mesma forma que a etapa avaliação pré-treinamento da Atividade Treinamento Auditivo.
- B) Intervalo: nessa etapa o participante realizou a leitura de um texto.

C) Avaliação pós-intervalo: foi realizada da mesma forma que a etapa avaliação pré-intervalo.

A definição escrita dos parâmetros e dos graus de desvio vocal foi disponibilizada durante as etapas de avaliação pré e pós-intervalo da Atividade Controle.

## 2.2. Seleção das emissões vocais para avaliação.

Foram utilizados nas etapas avaliação pré e pós-intervalo as mesmas emissões vocais usados nas etapas avaliação pré e pós-treinamento da Atividade Treinamento Auditivo. Em cada etapa as vozes foram aleatorizadas.

## 2.3. Seleção do texto para a etapa Intervalo.

Estabeleceu-se que a etapa intervalo da Atividade controle deveria ter a mesma média de duração de tempo que a etapa treinamento da Atividade treinamento auditivo. Para isso, foi realizado um estudo piloto com três sujeitos sem experiência prévia em avaliação perceptivo-auditiva da voz, por meio do qual realizaram a etapa treinamento, sendo calculada a média de duração de tempo do treinamento de 21 minutos (DP=4,02min).

Solicitou-se aos mesmos sujeitos que realizassem a leitura de um texto previamente selecionado pelas pesquisadoras. O tempo foi cronometrado, encontrando-se uma média de 20 minutos (DP=2,07min) para leitura do texto. Essa etapa foi realizada nos dois aplicativos. Todos os avaliadores assinaram o TCLE.

Cada atividade teve duração aproximada de uma hora.

## Seleção dos avaliadores

Foi determinado 15 avaliadores, considerando 57 observações (vozes a serem avaliadas) e 20 variáveis (parâmetros R, B, A, S e I com os graus de desvio ausência de desvio, desvio leve, desvio moderado e desvio intenso), por meio do Kappa proposto por Fleiss<sup>(4)</sup>, considerando poder estatístico de 80% e nível de significância de 5%. Foram selecionados 15 avaliadores não experientes para avaliação vocal, estudantes dos primeiros períodos do curso de graduação em Fonoaudiologia, sujeitos de ambos os sexos com idade entre 18 e 28 anos. Os mesmos avaliadores realizaram as duas atividades: “Atividade treinamento auditivo” e “Atividade controle”. Foram considerados os seguintes critérios de inclusão: relatar ausência de perda auditiva, aceitar participar da pesquisa por meio da concordância com o TCLE via *web*, responder ao questionário inicial, participar de todas as etapas das duas atividades que compõem o sítio. A pesquisa foi realizada por 22 avaliadores, sendo que sete destes foram excluídos por não realizarem alguma das atividades, totalizando 15 avaliadores.

Em nenhum momento os avaliadores foram ou serão identificados. A identidade será preservada, sendo identificados por números.

## **Bloco 2: Calibrador Auditivo**

As atividades deste bloco foram elaboradas a fim de analisar o efeito da **avaliação perceptivo-auditivo com apoio de emissões âncoras de vozes sintetizadas** na concordância intra e interavaliadores da avaliação vocal. Este bloco recebeu o nome de Calibrador Auditivo (Figura 6) devido à proposta de que as

emissões vocais âncoras fossem usadas como referências para a classificação das vozes em avaliação, assim como as emissões vocais são oferecidas a um músico por um afinador no processo de afinação do instrumento musical. Para compor este bloco foram criados e disponibilizados pelas pesquisadoras duas atividades para avaliação perceptivo-auditiva de vozes, descritas posteriormente. Assim como no Bloco 1 do projeto, antes de iniciar as atividades os avaliadores deviam ler o Termo de Consentimento Livre e Esclarecido (TCLE) e, caso concordassem, deveriam selecionar a opção “Aceito” para prosseguir na participação da pesquisa. Em seguida, responderam a um breve questionário fornecendo dados sobre sua experiência em treinamento auditivo, idade, profissão e acuidade auditiva; e receberam uma apresentação inicial da pesquisa.

## CALIBRADOR AUDITIVO

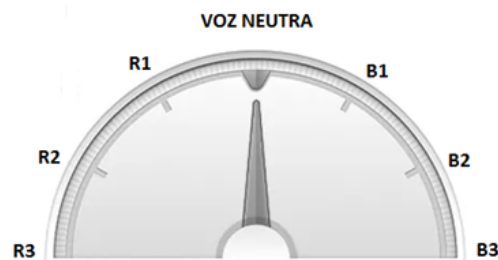


Figura 6 – Página inicial do Calibrador Auditivo

Após apresentação inicial, os avaliadores realizaram duas atividades: Atividade Calibrador Ativo – referente à análise de vozes com apoio de emissões

âncoras de vozes sintetizadas, e Atividade Calibrador Inativo – referente à análise de vozes sem apoio de emissões vocais âncoras. A ordem de realização das atividades foi sorteada para cada participante, sendo que a segunda atividade foi realizada exatamente 15 dias após a primeira (Figura 7). É possível observar na literatura o uso de um intervalo de pelo menos uma semana entre atividades de avaliação, a fim de evitar qualquer memorização<sup>(1,2,3)</sup>.

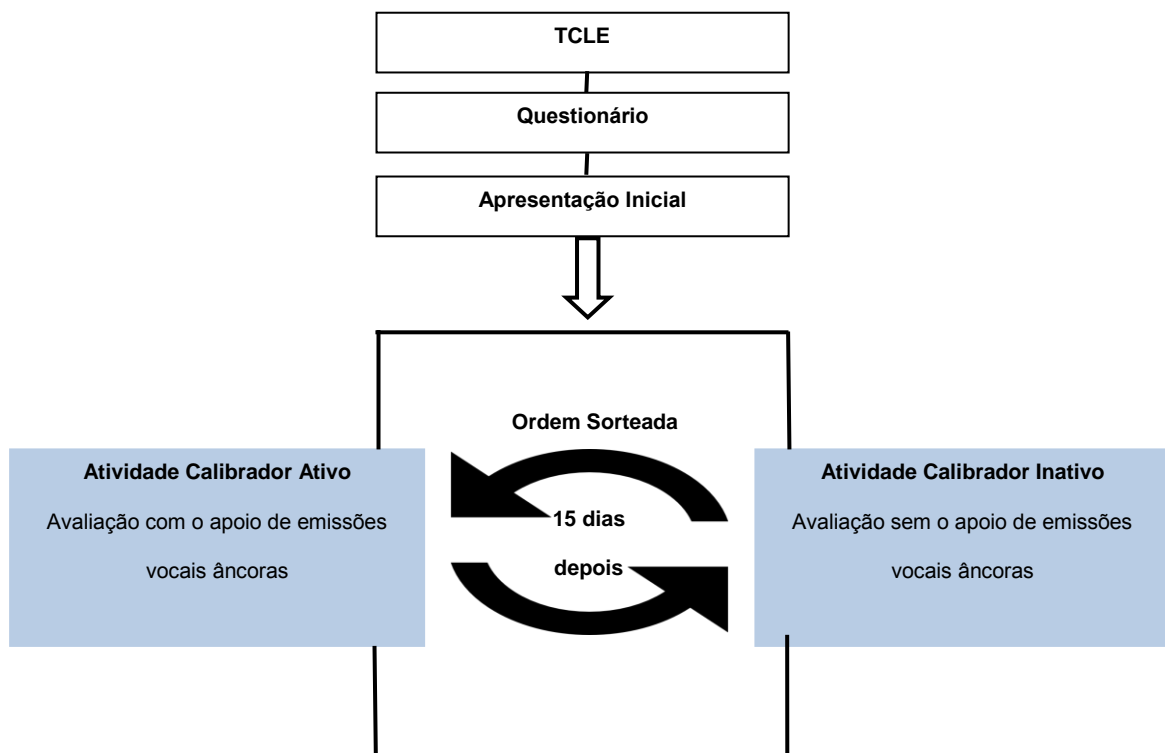


Figura 7 – Fluxograma do Calibrador Auditivo.

Cada atividade será descrita a seguir.

### 1. Atividade Calibrador Ativo

A atividade que utilizou emissões âncoras de vozes sintetizadas na avaliação perceptivo-auditiva foi chamada de Atividade Calibrador Ativo.

## 1.1. Processo

Durante essa atividade os avaliadores escutam a voz natural a ser avaliada, em seguida escutam as emissões âncoras de vozes sintetizadas para cada grau do parâmetro rugosidade (R), e novamente escutam a voz a ser avaliada, classificando-a em 0 – ausência de desvio, 1 – desvio leve, 2 – desvio moderado ou 3 – desvio intenso para este parâmetro. Repetiram o processo para classificar a mesma voz quanto ao parâmetro soprosidade (B) (Figura 8).

**Atividade Calibrador Ativo**  
**VOZ 1**

---

**Rugosidade**

Qualquer irregularidade perceptível durante a produção vocal.

  
**R0**


  
**R1**


  
**R2**

  
**R3**

**Soprosidade**


Qualquer escape de ar audível durante a produção vocal.

  
**B0**

  
**B1**

  
**B2**


  
**B3**



Grau de rugosidade: \_\_\_\_\_

Grau de soprosidade: \_\_\_\_\_

---



**Figura 8** – Atividade Calibrador Ativo.

A definição escrita dos parâmetros foi disponibilizada durante todas as etapas da Atividade Calibrador Ativo.

## 1.2. Seleção das emissões vocais para avaliação.

Para compor a amostra de vozes naturais a serem avaliadas, utilizou-se o banco de vozes do ambulatório da Universidade Federal de Minas Gerais, formado por 381 vozes, amostras da emissão da vogal /a/ sustentada de forma habitual, de indivíduos de ambos os gêneros com idade a partir de 18 anos. Todos os avaliadores assinaram o TCLE.

Duas avaliadoras, fonoaudiólogas, especialistas em voz, com mais de cinco anos de experiência em avaliação perceptivo-auditiva, analisaram individualmente as vozes, utilizando o fone de ouvido supra-auricular modelo *Multilaser Vibe Headphone* estéreo. Classificaram as vozes considerando-se a R e B da escala GRBASI, conforme o parâmetro predominante, em 0 – ausência de desvio, 1 – desvio leve, 2 – desvio moderado ou 3 – desvio intenso.

Foram considerados os seguintes critérios de inclusão: vozes naturais de sujeitos do sexo feminino e masculino, com idade a partir de 18 anos, com um parâmetro predominante de variados graus de desvio vocal; vozes que apresentaram a mesma classificação pelas duas avaliadoras.

Foram selecionados três emissões vocais para cada grau dos parâmetros predominantes R e B, sendo que um grau do parâmetro soproside recebeu quatro emissões, totalizando 25 vozes. Acrescentou-se 20% das vozes com o objetivo de analisar a concordância intra-avaliador.

Em todas as etapas do aplicativo, assim como na publicação da presente pesquisa, as vozes foram identificadas por números.

### 1.3. Seleção das emissões vocais âncoras.

A amostra das emissões vocais âncoras foi composta por vozes sintetizadas. Para a construção das vozes sintetizadas neutras (N) ou contendo como parâmetro

R ou B com diferentes graus de desvio vocal utilizou-se como fonte (fluxo glótico) um modelo paramétrico que permite o controle da frequência fundamental, de *jitter*, de *shimmer* e da relação sinal ruído. Como filtro, utilizou-se um trato vocal que modela a vogal /a/, extraído de voz natural por técnica de predição linear. As emissões vocais foram construídas por um engenheiro, totalizando 300 vozes sintetizadas<sup>(5)</sup>.

Para a análise do grau de naturalidade, e da qualidade das vozes sintetizadas, foram selecionados três avaliadores, fonoaudiólogos com mais de cinco anos de experiência em avaliação vocal, que realizou individualmente a análise perceptivo-auditiva da naturalidade das vozes (relacionado ao quanto o ouvinte percebe a voz como humana) por meio de uma escala visual analógica (EVA); em seguida foi feita a classificação da qualidade vocal como N (voz neutra: ausência de desvio vocal), R (rugosa: relacionado à percepção de irregularidade durante a produção vocal) ou B (soprosa: escape de ar audível durante a produção vocal), e da mensuração, também por meio de uma EVA, do grau de desvio vocal. Os valores encontrados para o desvio vocal das vozes classificadas como R ou B por meio da EVA, foram convertidos segundo sugerido pela literatura<sup>(6)</sup>:

Grau de desvio	Correlação da classificação do desvio vocal pela escala visual analógica e escala numérica	
	Rugosa (mm)	Soprosa (mm)
<b>Neutro</b>	0 – 8,5	0 – 8,5
<b>Leve</b>	8,5 – 28,5	8,5 – 33,5
<b>Moderado</b>	28,5 – 59,5	33,5 – 52,5
<b>Intenso</b>	A partir de 59,5	A partir de 52,5

Quadro 1 – Correlação da classificação do desvio vocal pela escala visual analógica e escala numérica.

Foram selecionadas como âncoras as vozes sintetizadas de diferentes graus de desvio, para cada parâmetro, classificadas com maior naturalidade por pelo menos dois avaliadores, conforme descrito no Quadro 2.

Grau de desvio	Classificação da naturalidade das vozes (mm)		
	Neutra	Rugosa	Soprosa
_____	97,3	_____	_____
<b>Leve</b>	_____	56	86
<b>Moderado</b>	_____	41	60
<b>Intenso</b>	_____	37	40

Quadro 2 – Média do grau de naturalidade das vozes sintetizadas para cada parâmetro perceptivo-auditivo selecionado para a amostra.

A amostra das emissões âncoras foi composta por uma voz de cada grau – ausência de desvio, desvio leve, moderado e intenso, de cada parâmetro – R e B, totalizando oito vozes.

## 2. Atividade Calibrador Inativo

A atividade que não utilizou emissões âncoras de vozes sintetizadas na avaliação foi chamada de Atividade Calibrador Inativo.

### 2.1. Processo.

Durante essa atividade, os avaliadores escutaram a voz natural a ser avaliada e em seguida a classificaram em 0 – ausência de desvio, 1 – desvio leve, 2 – desvio moderado ou 3 – desvio intenso quanto ao parâmetro R. Repetiram o processo para classificar a mesma voz quanto ao parâmetro B (Figura 9).

# Atividade Calibrador Inativo

## VOZ 1

---

<b>Rugosidade</b> Qualquer irregularidade perceptível durante a produção vocal.	<b>Soprosidade</b> Qualquer escape de ar audível durante a produção vocal.
R0   R1   R2   R3	B0   B1   B2   B3

🔊

Grau de rugosidade: _____
Grau de soprosidade: _____

---



 

Figura 9 – Atividade Calibrador Inativo.

### 2.2. Seleção das emissões vocais para avaliação.

Foram utilizadas na Atividade Calibrador Inativo as mesmas emissões vocais usados na Atividade Calibrador Ativo. Em cada etapa as vozes foram aleatorizadas.

Cada atividade teve duração aproximada de vinte minutos. Os avaliadores podiam escutar as vozes quantas vezes julgassem necessário, desde que respeitassem o comando fornecido – escutar a voz a ser avaliada, em seguida as emissões vocais âncoras do parâmetro R, e novamente a voz à ser avaliada, repetindo o processo ao classificar quanto ao parâmetro B. Foi utilizado fone de ouvido supra-auricular modelo *Multilaser Vibe Headphone* estéreo durante todos os procedimentos e as avaliações foram realizadas individualmente.

### Seleção dos avaliadores

Foi determinado 32 avaliadores em um cálculo amostral considerando 25 observações (vozes a serem avaliadas) e 8 variáveis (parâmetros R e B com ausência de desvio, desvio leve, moderado e intenso), por meio do teste Kappa proposto por Fleiss<sup>(4)</sup>, considerando poder estatístico de 80% e nível de significância de 5%. Foram selecionados 32 avaliadores não experientes para avaliação vocal, estudantes dos primeiros períodos do curso de Graduação em Fonoaudiologia, sujeitos de ambos os sexos com idade entre 18 e 28 anos. Os mesmos avaliadores realizaram as duas atividades. Foram considerados os seguintes critérios de inclusão: aceitar participar da pesquisa por meio da concordância com o TCLE, responder ao questionário inicial, participar de todas as atividades, relatar ausência de perda auditiva. A pesquisa foi realizada por 32 avaliadores. É importante ressaltar que os avaliadores do Bloco 2 não foram os mesmos que participaram do Bloco 1.

Em nenhum momento os avaliadores foram ou serão identificados. A identidade será preservada, sendo identificados por números.

### **Análise Estatística**

Todos os dados e procedimentos realizados pelos avaliadores nos dois blocos do projeto (aceitação do TCLE, preenchimento do questionário e realização de todas as atividades) foram salvos para posterior análise dos dados. As respostas dos questionários, assim como os resultados das análises, foram inseridos em um banco de dados, no programa Microsoft Office Excel 2007.

Para análise da concordância intra e interavaliadores foi utilizado o coeficiente AC1 de Gwet<sup>(7)</sup> e Kappa, e para comparação entre as concordâncias foram

utilizados o intervalo de confiança (IC) e a variação percentual da concordância. Apesar do coeficiente Kappa ser amplamente utilizado para análise da concordância em avaliações perceptivo-auditivas da voz, o que pode ser observado na literatura<sup>(4)</sup>, é uma medida que não é corrigida pela chance de acordo e, por ser um coeficiente de acordo, não diferencia as várias fontes e tipos de desacordo<sup>(4,7)</sup>. Diante disso, um novo coeficiente tem sido proposto para análise, o AC1 de Gwet<sup>(4,7,8)</sup>. Este coeficiente, assim como o Kappa, varia entre 0 e 1 e, quanto mais próximo de 1, maior a concordância e menor a probabilidade da concordância acontecer ao acaso. Ao contrário do Kappa, o AC1 considera tanto a chance de acordo quanto os erros de classificação, o que permite analisar a probabilidade dos avaliadores concordarem eliminando o acordo por acaso<sup>(7,8)</sup>.

Para realizar a análise estatística foi utilizado o *software* R versão 3.3.1 Em todas as análises foi considerado um nível de significância de 5%.

### **Referências Bibliográficas**

1. Silva RSA, Simões-Zenari M, Nembr NK. Impacto de treinamento auditivo na avaliação perceptivo-auditiva da voz realizada por estudantes de Fonoaudiologia. J Soc Bras Fonoaudiol. 2012;24(1):19-25.
2. Helou LB, Solomon NP, Henry LR, Coppit GL, Howard RS, Stojadinovic A. The role of listener experience on Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) ratings of postthyroidectomy voice. Am J Pathol Speech Lang. 2010;19(3):248-258.
3. Brinca L, Batista AP, Tavares AI, Pinto PN e Araújo L. The effect of anchors and training on the reliability of voice quality ratings for different types of speech stimuli. J Voice. 2015;29(6):776.e7-14.

4. Wongpakaran N, Wongpakaran T, Wedding D, & Gwet, K L A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *BMC Med Res Methodol*, 2013;13(1):1.
5. Vieira MN, Sansão JPH, Yehia HC. Measurement of signal-to-noise ratio in dysphonic voices by image processing of spectrograms. *Speech Commun*. 2014;61-62:17-32.
6. Baravieira et al. Análise perceptivo-auditiva de vozes rugosas e soprosas: correspondência entre a escala visual analógica e a escala numérica. *CoDAS* 2016;28(2):163-167.
7. Gwet KL. Testing the difference of correlated agreement coefficients for statistical significance. *Educ. Psychol. Meas*. 2016;76(4):609-637.
8. Haley DT, Thomas P, Petre M, Roeck AD. Using a new inter-rater reliability statistic. *Technical Report*. 2008;15.

Os resultados serão apresentados em formato de artigos, intitulados:

- Santos PCM, Vieira MN, Sansão JPH, Gama ACC. ***Effect of auditory-perceptual training with natural voice anchors on vocal quality evaluation.*** *Journal of Voice.* 2017 (*in press*).
- **Efeito de emissões âncoras de vozes sintetizadas na avaliação perceptivo-auditiva da voz.** O artigo será submetido ao *Journal of Speech, Language, and Hearing Research (JSLHR)*.

## ARTICLE IN PRESS

## Effect of Auditory-Perceptual Training With Natural Voice Anchors on Vocal Quality Evaluation

\*Priscila Campos Martins dos Santos, †Maurílio Nunes Vieira, ‡João Pedro Hallack Sansão, and \*Ana Cristina Côrtes Gama, \*†Belo Horizonte and ‡São João Del Rey, Minas Gerais

**Summary: Purpose.** To analyze the effects of auditory-perceptual training with anchor stimuli of natural voices on inter-rater agreement during the assessment of vocal quality.

**Study design.** This is a quantitative nature study.

**Methods.** An auditory-perceptual training site was developed consisting of Programming Interface A, an auditory training activity, and Programming Interface B, a control activity. Each interface had three stages: pre-training/pre-interval evaluation, training/interval, and post-training/post-interval evaluation. Two experienced evaluators classified 381 voices according to the GRBASI scale (G-grade, R-roughness, B-breathiness, A-asthenia, S-strain, I-instability). Voices were selected that received the same evaluation by both evaluators: 57 voices for evaluation and 56 for training were selected, with varying degrees of deviation across parameters. Fifteen inexperienced evaluators were then selected. In the pre-, post-training, pre-, and postinterval stages, evaluators listened to the voices and classified them via the GRBASI scale. In the stage interval evaluators read a text. In the stage training each parameter was trained separately. Evaluators analyzed the degrees of deviation of the GRBASI parameters based on anchor stimuli, and could only advance after correctly classifying the voices. To quantify inter-rater agreement and provide statistical analyses, the ACI coefficient, confidence intervals, and percentage variation of agreement were employed.

**Results.** Except for the asthenia parameter, decreased agreement was observed in the control condition. Improved agreement was observed with auditory training, but this improvement did not achieve statistical significance.

**Conclusion.** Training with natural voice anchors suggest an increased inter-rater agreement during perceptual voice analysis, potentially indicating that new internal references were established.

**Key Words:** Voice–Voice quality–Dysphonia–Auditory perception–Voice training.

### INTRODUCTION

Although the voice can be measured objectively, perceptual changes in voice quality are likely the primary stimulus for patients to seek clinical care.<sup>1</sup> Vocal quality is frequently judged based on the impression that a listener has of the voice of a speaker as a whole; auditory-perception of a voice is both holistic and integrative.<sup>2</sup> Auditory-perceptual evaluation is classically employed to evaluate vocal quality and is frequently employed in the clinical setting as instrumental analyses do not recapitulate human perception.<sup>1</sup>

Auditory-perceptual analyses have several advantages, including evaluation of vocal quality, a perceptual phenomenon in response to an auditory stimulus. As a consequence, perceptual descriptions of voice are potentially intuitive and significant.<sup>2</sup> The advantages of perceptual analyses are also directly related to limitations of instrumental approaches to voice measurement, including the acoustic conditions of the recording environment, characteristics of hardware and software systems, analysis protocols employed, and individual variability in acoustic

parameters. Finally, auditory-perceptual analysis is fast, noninvasive, and inexpensive.<sup>2</sup>

However, auditory-perceptual evaluation has limitations. Primarily, it is subjective and based on internal standards to evaluate aberrant voices.<sup>1,3</sup> Although auditory-perceptual assessment is relatively easy and largely free from instrumentation, reliability and sensitivity may vary.<sup>3</sup> Increased variability in intra- and inter-rater agreement has long been a central issue in voice research.<sup>1</sup>

Recent studies outlined factors that may interfere with reliability of auditory-perceptual assessment such as the multidimensional characteristics of voice as well as auditory processing, speech task employed, and the internal standards of evaluators which can be reflective of their personal, professional experience, and previous training.<sup>1,4</sup> The fact that voices are inherently unstable and are often characterized by the presence of more than one aberrant parameter, such as roughness plus breathiness or breathiness plus tension confirmed the need for improved reliability of auditory-perceptual evaluation.<sup>5–8</sup>

Specifically, some tasks have been shown to hold promise to optimize auditory-perceptual evaluation, such as the use of anchor stimuli and auditory training.<sup>3</sup> Specifically, to increase the reliability of auditory-perceptual evaluation, recent research suggested the use of controlled anchored stimuli.<sup>3–5</sup> Anchors are predefined and selected to represent a determined type and/or degree of deviation. In addition, auditory training has also been shown to increase the reliability of auditory-perceptual evaluation, reducing variability and subjectivity of analyses.<sup>4,6–8</sup>

According to previous work, perceptual learning during auditory training is most effective when two key processes are

Accepted for publication October 26, 2017.

From the \*Department of Speech-language Pathology, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais; †Department of Electronic Engineering, Federal University of Minas Gerais, Belo Horizonte, Minas Gerais; and the ‡Department of Telecommunications Engineering and Mechatronics, Federal University of São João Del Rey, São João Del Rey, Minas Gerais.

Address correspondence and reprint requests to Priscila Campos Martins dos Santos, Universidade Federal de Minas Gerais, Av. Alfredo Balena, 190/249, Belo Horizonte, Minas Gerais, Brazil. E-mail: priscila.fonoaudiologia@gmail.com

Journal of Voice, Vol. ■■■, No. ■■■, pp. ■■■–■■■  
0892-1997

© 2017 The Voice Foundation. Published by Elsevier Inc. All rights reserved.  
<https://doi.org/10.1016/j.jvoice.2017.10.020>

involved: stimulus imprinting and differentiation.<sup>9</sup> Stimulus imprinting refers to the development of internal detectors or standards formed by repeated exposure to the stimulus. The form of detector is shaped by the impinging stimulus, and it justifies the term stimulus imprinting. These detectors increase the speed, accuracy, and general fluency with which the stimuli are processed.<sup>9</sup> Differentiation of stimuli allows the listener to learn to differentiate signals; that is, to identify each voice parameter independently. When feedback is provided during training, this differentiation occurs more rapidly.<sup>9</sup>

Therefore, standardization of training to increase sensitivity, reliability, and agreement of auditory-perceptual evaluation of voice is of great relevance for both research and clinical care. The objective of the present study was to analyze the effect of auditory-perceptual training with anchor stimuli of natural voices on inter-rater agreement. It was hypothesized that the inter-rater agreement during auditory-perceptual evaluation would improve after the training employing anchor stimuli of natural voices.

## MATERIALS AND METHODS

The current project was approved by the Research Ethics Committee (CAAE—37872314.2.0000.5149) at the Federal University of Minas Gerais. This is a quantitative nature study. All subjects provided written, informed consent before participation. Upon consent, subjects completed a brief questionnaire providing information regarding their experience with auditory-perceptual analysis including any previous training, as well as age and profession. Subjects were then randomly assigned to one of two programming interfaces.

### Programming interface

The programming interfaces are websites created with explicit instructions for performing each stage of auditory-perceptive voice evaluation. The evaluators received the link of programming interface and accessed it by the Google Chrome browser. No other

instruction was provided by the applicator, so that it could not influence the training. Two programming interfaces were created: Programming Interface A involved auditory training and Programming Interface B was the control activity. The second programming interface was performed 15 days after the first to avoid task memorization.<sup>6,10</sup> Each programming interface was composed of three stages (Figure 1).

### Programming interface A

This programming interface involved auditory training.

**Process.** This activity was structured in three stages:

- (A) Pretraining evaluation: in this stage, participants listened to voices and classified them according to the GRBASI scale (G-grade, R-roughness, B-breathiness, A-asthenia, S-strain, I-instability)<sup>11</sup> according the previously reported protocols.
- (B) Training: Training focused on the parameters R, B, A, S, and I presented separately. Participants listened to voice stimuli, and subsequently, anchors. Participants then rated the voices on the degree of parameter in training. If their response was incorrect, they were encouraged to repeat the task until correct. It was considered incorrect answers those with classification of degree of parameter in training different from the classification given by the evaluators in the selection of voices to be evaluated—just voices that obtained the same classification of degree of parameter in training by two experienced evaluators were selected.
- (C) Post-training evaluation: similar to the pretraining evaluation.

All participants were provided with written definitions of the parameters and degrees of vocal deviation during Programming Interface A. Analyses were completed individually, using

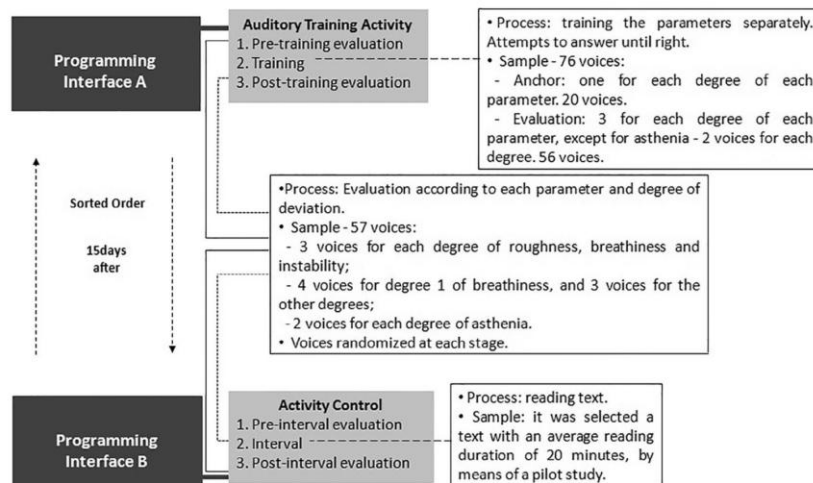


FIGURE 1. Experimental flow chart of auditory-perceptual training programming interfaces.

an over-ear headphone, in a silent classroom, which, however, was not acoustically treated. Each participant set the headphone at a comfortable volume for themselves and could hear voices as many times as necessary.

**Selection of vocal stimuli for evaluation.** The voice bank of the university clinic, composed of 381 samples of sustained /a/ from male and female adults was employed to compile the samples for evaluation. Two voice specialists with more than 5 years of experience in auditory-perceptual assessment analyzed the voices using an over-ear headphone, model Multilaser Vibe Stereo. Voice samples were classified according to the degree (0—neutral, 1—slight, 2—moderate, and 3—intense) of predominant parameter (R, B, A, S, or I).

Inclusion criteria considered were natural voices of female and male subjects, greater than 18 years of age, with a predominantly deviant vocal parameter. Voices were selected that obtained the same rating by two evaluators on the degree of predominant parameter.

Three stimuli for each parameter (R, B, S, and I) and severity were selected for pre- and post-training, and a grade of one of the parameters received four stimuli, totaling 49 samples. For asthenia, two stimuli were selected for each degree, totaling eight samples. The occurrence of asthenic voices in the bank used for research was reduced when compared with other parameters, which justified the decreased number of stimuli. Therefore, the total number of samples for Programming Interface A was 57.

For the training stage, three stimuli were selected for each degree of each parameter (R, B, S, I). For asthenia, two stimuli for each degree were selected, totaling 56 samples for this stage of Programming Interface A.

**Selection of anchors for training.** Anchor stimuli were derived from the same bank of voices used to select the stimuli to be evaluated in the pre- and post-training evaluation. The voices were previously classified by two voice specialists, with more than 5 years of experience in auditory-perceptual assessment, according to the degree (0—neutral, 1—slight, 2—moderate, and 3—intense) of predominant parameter (R, B, A, S, and I). Analyses were completed individually, using an over-ear headphone, model Multilaser Vibe Stereo.

The following inclusion criteria were considered: natural voices of women and men, 18 years of age or older, and with deviation degree 0—neutral, 1—slight, 2—moderate, and 3—intense for each predominant parameter. Voices were selected if they obtained the same rating by two evaluators on the degree of predominant parameter.

One sample was selected as an anchor for each degree of deviation for each parameter. Thus, the total number of anchor stimuli for training in Programming Interface A was 20.

#### **Programming interface B**

Programming Interface B is a control activity. This program includes the same stages as Program A. However, instead of training, the evaluators read a text, which is the control activity. It was included to allow for analysis of a training effect, eliminating the possibility of improved performance due to chance.

**Process.** Programming Interface B was also structured in three stages:

- (A) Preinterval evaluation: performed in the same way as the pretraining evaluation stage of Programming Interface A.
- (B) Interval: participant read text.
- (C) Postinterval evaluation: performed in the same way as the preinterval evaluation stage.

Similar to Programming Interface A, written definitions of the parameters and degrees of vocal deviation were provided to the participants during the pre- and postinterval evaluation of Programming Interface B. Analyses were completed individually, using an over-ear headphone, in a silent classroom, which, however, was not acoustically treated. Each participant set the headphone at a comfortable volume for themselves and could hear voices as many times as necessary.

**Selection of vocal stimuli to compose the task of evaluation.** The same stimuli employed in pre- and post-training of Programming Interface A were used in pre- and postinterval evaluation stages. The samples were randomized for each stage.

**Selection of the text to compose the interval step.** The interval stage of Programming Interface B was determined to be the same average duration as the training stage of Programming Interface A. Therefore, a pilot study was conducted with three subjects with no prior experience in auditory-perceptual evaluation of voice. These subjects completed the training stage in an average of 21 minutes (standard deviation [SD] = 4.02).

These same subjects were then asked to read a text previously selected by the researchers. The duration of this task was an average of 20 minutes (SD = 2.07).

Each programming interface lasted approximately 1 hour.

#### **Selection of evaluators**

The number of evaluators was determined from the Kappa Index proposed by Fleiss<sup>12</sup> with statistical power of 80% and significance level of 5%. These calculations confirmed that 15 evaluators of 57 observations (voices) and 20 variables (R, B, A, S, and I parameters with degrees neutral, slight, moderate, and intense) were adequate to achieve meaningful data.

Fifteen nonexperienced evaluators were selected from a pool of Speech Therapy graduate students, male and female, aged 18 and 28 years. The same evaluators participated in both programming interfaces.

The following inclusion criteria were considered: absence of a hearing loss complaint, agreed to participate in the research accepting the informed consent, responded to the initial questionnaire, and participated in all stages of two programming interfaces.

The order of activities was randomized.

### Statistical analysis

All data were archived for further analysis. Responses to the questionnaires, as well as the results of the sample analyses of the three stages in each programming interface, were inserted into Microsoft Excel. Gwet's AC1 coefficient<sup>12,13</sup> was used to analyze intra- and inter-rater concordance, and to compare pre- and post-training evaluation and pre- and postinterval evaluation concordances. Concordances between the two programming interfaces were calculated via confidence interval (CI) and percentage variation of agreement. All statistical analyses were performed using R (v3.3.1). A significance level of 5% was considered for all analyses.

### RESULTS

Twenty-two evaluators initially participated in the study. However, seven were excluded as they did not return to perform the second programming interface. In a brief questionnaire, all participants stated that they did not have previous experience or training in perceptual-auditory assessment of voice. All the participants declared to be undergraduate students of the speech-language pathology course between the first and third period, and had not received content on perceptual-auditory assessment of the voice. The participants' ages ranged from 17 to 28 years, with a mean of 20.06 years.

A mean duration of 58.33 minutes (SD = 13.41) was observed for completion of the auditory training activity, and 48.13 minutes (SD = 19.69) in the performance of control activity. In the pilot study, a mean duration of 21 minutes (SD = 4.02) for the training stage was estimated, and 20 minutes (SD = 2.07) for the interval stage. However, it was found that the evaluators had a mean duration of 12.87 minutes (SD = 3.99) in the training stage, and 9.73 minutes (SD = 6.21) in the interval stage. No difference was observed between inter-rater concordances for the control activity and auditory training activity (Table 1).

Concordance in the control activity decreased with regard to variation of inter-rater agreement in the pre- and post-training evaluation and pre and postinterval evaluation stages. Improved

agreement was observed following the auditory training activity, with exception of asthenia, which decreased in both groups (Table 1 and Figure 2).

The order of activities was randomized, and in this way, nine evaluators completed Programming Interface A first, and six evaluators performed the analysis of Programming Interface B initially (Table 2).

### DISCUSSION

Auditory-perceptual analysis is influenced by extrinsic and intrinsic factors. The main intrinsic factors identified in the literature are time of professional experience, type of professional training, and previous auditory training, as well as internal standards of the evaluators and their state of attention during analysis.<sup>3,6,14,15</sup> As for extrinsic factors, presentation of vocal stimuli, type of perceptive-auditory scale, parameters, and speech task have been described to alter auditory perceptual analysis.<sup>3,6,14,15</sup>

In the present study, we sought to control many of these factors to ensure that evaluators only experience training. As such, graduate students in speech therapy were selected as they had no previous experience in auditory-perceptual assessment of voice. Stimuli and differentiation with feedback were employed, which have been shown to facilitate learning.<sup>9</sup>

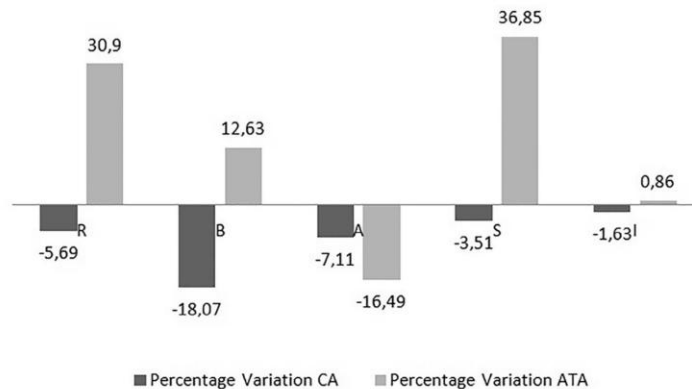
Anchor stimuli were also employed to control extrinsic factors.<sup>4,5</sup> Recent investigation<sup>7</sup> suggested that auditory anchors associated with auditory training yielded increased inter-rater agreement when compared with training without anchors or only with written anchors. The current study performed auditory training using auditory anchor stimuli in addition to written definitions of the parameters across severities. Previous work has shown that anchor stimuli potentially increased inter-rater agreement for the G, R, and B parameters.<sup>7,10,16</sup> In the present study, the G, R, B, A, S, and I parameters were evaluated using the GRBASI scale as a widely used scale in clinical and speech-language pathology research.<sup>6,14,16</sup> Sustained /a/ was also employed as this vowel is stable and consistent and has been associated with high inter-rater concordance.<sup>17,18</sup>

**TABLE 1.**  
Inter-rater Agreement and Percentage Variation of Inter-rater Agreement Between Pre- and Postinterval Evaluation and Pre and Post-training Evaluation in Control and Auditory Training Activities

		CA Pre	CA Pos	PV%	ATA Pre	ATA Pos	PV%
R	AC1 (%)	13.53	12.76	-5.69	15.60	20.42	30.90
	CI 95% (%)	-2.46 to 29.53	-3.07 to 28.58		-0.41 to 31.62	2.49 to 33.48	
B	AC1 (%)	25.07	20.54	-18.07	18.13	20.42	12.63
	CI 95% (%)	10.07-40.08	12.76 to 35.99		2.74-33.52	5.45-20.42	
A	AC1 (%)	25.75	23.92	-7.11	24.32	20.31	-16.49
	CI 95% (%)	10.84-40.65	8.93-38.91		9.68-38.97	5.60-35.03	
S	AC1 (%)	30.37	29.30	-3.51	19.83	27.14	36.85
	CI 95% (%)	16.61-44.13	15.73-42.87		6.12-33.54	13.41-40.87	
I	AC1 (%)	23.99	23.60	-1.63	22.04	22.23	0.86
	CI 95% (%)	8.79-39.20	8.53-38.67		7.02-37.07	6.85-37.61	

Notes: For statistical analysis, AC1 (first-order agreement coefficient), confidence interval, and percentage variation (the difference between the percentage values of the pre- and postinterval and pre and post-training evaluation stages) were considered.

Abbreviations: CA Pre, preinterval evaluation of control activity; CA Pos, postinterval evaluation of control activity; PV, percentage variation; ATA Pre, pretraining evaluation of auditory training activity; ATA Pos, post-training evaluation of auditory training activity; R, roughness; B, breathiness; A, asthenia; S, tension; I, instability.



**FIGURE 2.** Comparison of the percentage variation of inter-rater agreement between pre- and postinterval and pre- and posttraining evaluation in control and auditory training activities. *Abbreviations:* R, roughness; B, breathiness; A, asthenia; S, tension; I, instability; CA, control activity; ATA, auditory training activity.

Decreased inter-rater agreement was observed, which was expected as the evaluators were inexperienced. These data agree with findings from a recent study that verified that experience of the evaluator impacted inter-rater agreement.<sup>19</sup> However, decreased agreement may also be related to the state of evaluator attention, as has been reported in the literature.<sup>6</sup> The duration of activities in the current study was approximately 1 hour. For future studies, the pretraining evaluation could be performed at a different sitting to attempt to avoid this confound.

Another factor that may have influenced inter-rater agreement was the number of training sessions. Studies have shown an increase in agreement after approximately 10 training sessions.<sup>6,8</sup> Additional training sessions are suggested for future studies, as well as a post-training evaluation at a different sitting from the training to analyze the duration of the training effect. No statistically significant differences were observed between

the auditory training and control activities with regard to inter-rater agreement. However, upon *post hoc* review, decreased agreement was observed across all parameters in the postinterval stage. In the auditory training activity, an increase in agreement across all parameters was observed in the post-training stage, except for asthenia, for which there was no change in agreement (Figure 1). Based on these data, one may infer that the improved agreement following auditory training was the result of training. That is, training enhanced the acquisition of new internal references. However, this increased agreement was discreet. Clearly, further investigation is warranted regarding the relevant factors that interfere with auditory-perceptual evaluation to eventually standardize training, and consequently, increase reliability.

Challenges regarding the evaluation of asthenia has been reported across studies.<sup>14,16,20</sup> This decreased agreement may be related to the lower number of stimuli for this parameter, due to the lower occurrence of asthenic voice in the bank used for research, yielding greater inconsistency for this classification. The multidimensional characteristics of voice and the nature of auditory processing make auditory-perceptual evaluation challenging. The evaluation of voice depends on several internal standards, which likely underlie substantial variability in the reliability of auditory-perceptual evaluation.<sup>3,14,21</sup> The fact that voices are not stable and are often characterized by more than one parameter, such as roughness and breathiness, breathiness and tension, or these three parameters combined, also influences the reliability of auditory-perceptual evaluation.<sup>3,14,21</sup> Synthesized stimuli have been employed to optimize auditory-perceptual evaluation as it allows for precise control of acoustic properties. Synthetic stimuli also allow for analysis of each vocal parameter separately, which is not always possible as voices are typically characterized by more than one parameter.<sup>16,21</sup> In this regard, synthesized anchored stimuli in a structure called “binary tree” likely hold promise.<sup>16</sup> Based on these data, a clear need exists for further investigation into factors that influence auditory-perceptual evaluation with the ultimate goal of standardized training to ensure increased intra- and inter-rater agreement and reliability.

**TABLE 2.**  
The Order of Activities by the Evaluators

First Activity	Evaluators	Total
Programming Interface A	Evaluator 1	9 Evaluators
	Evaluator 2	
	Evaluator 4	
	Evaluator 6	
	Evaluator 9	
	Evaluator 10	
	Evaluator 12	
	Evaluator 13	
Programming Interface B	Evaluator 3	6 Evaluators
	Evaluator 5	
	Evaluator 7	
	Evaluator 8	
	Evaluator 11	
	Evaluator 15	

*Notes:* The order of activities was randomized. The evaluators were identified by numbers.

The order of activities was randomized. Nine evaluators performed Programming Interface A first, and six evaluators performed Programming Interface B first (Table 2). It was not possible to estimate if there is a statistically significant difference of inter-rater agreement between the different orders of programming interface realization, since the number of evaluators does not generate degree of freedom for this type of analysis through the coefficients AC1 and kappa, losing the statistical significance of analysis.<sup>13</sup>

### CONCLUSION

Training with natural voice anchors suggest an increased inter-rater agreement during perceptual voice analysis, potential indicating that new internal references were established.

### REFERENCES

- Behlau M. *Voz: o Livro do Especialista*. Rio de Janeiro: Revinter; 2001.
- Oates J. Auditory-perceptual evaluation of disordered vocal quality—pros, cons and future directions. *Folia Phoniatr Logop*. 2009;61:49–56.
- Eadie TL, Kapsner-Smith M. The effect of listener experience and anchors on judgments of dysphonia. *J Speech Hear Res*. 2011;54:430–447.
- Solomon NP, Helou LB, Stojadinovic A. Clinical versus laboratory ratings of voice using the CAPE-V. *J Voice*. 2011;25:e7–e14.
- Sofranko JL, Prosek RA. The effect of the levels and types of experience on judgment of synthesized voice quality. *J Voice*. 2014;28:24–35.
- Silva RSA, Simões-Zenari M, Nembr NK. Impacto de treinamento auditivo na avaliação perceptivo-auditiva da voz realizada por estudantes de Fonoaudiologia. *J Soc Bras Fonoaudiol*. 2012;24:19–25.
- Awan SN, Lawson LL. The effect of anchor modality on the reliability of vocal severity ratings. *J Voice*. 2009;23:341–352.
- Kreiman J, Gerratt BR. Comparing two methods for reducing variability in voice quality measurements. *J Speech Lang Hear Res*. 2011;54:803–812.
- Goldstone RL. Perceptual learning. *Ann Rev Psychol*. 1998;49:585–612.
- Brinca L, Batista AP, Tavares AI, et al. The effect of anchors and training on the reliability of voice quality ratings for different types of speech stimuli. *J Voice*. 2015;29:776, e7-e14.
- Hirano M. *Clinical Examination of Voice*. New York: Springer Verlag; 1981:81–84.
- Wongpakaran N, Wongpakaran T, Wedding D, et al. A comparison of Cohen's kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *BMC Med Res Methodol*. 2013;13:1.
- Gwet KL. Testing the difference of correlated agreement coefficients for statistical significance. *Educ Psychol Meas*. 2016;76:609–637.
- Chaves CR, Campbell M, Gama ACC. The influence of native language on auditory-perceptual evaluation of vocal samples completed by Brazilian and Canadian SLPs. *J Voice*. 2017;31:258.e1–258.e5.
- Lopes LW, Cavalcante DP, Costa PO. Severity of voice disorders: integration of perceptual and acoustic data in dysphonic patients. *Codas*. 2014;26:382–388.
- Vieira MN, Sansão JPH, Yehia HC. Measurement of signal-to-noise ratio in dysphonic voices by image processing of spectrograms. *Speech Commun*. 2014;61–62:17–32.
- Maryn Y, Roy N. Sustained vowels and continuous speech in the auditory-perceptual evaluation of dysphonia severity. *J Soc Bras Fonoaudiol*. 2012;24:107–112.
- Martins PC, Couto TE, Gama ACC. Avaliação perceptivo-auditiva do grau de desvio vocal: correlação entre a Escala Analógica Visual e escala numérica. *CoDAS*. 2015;27:279–284.
- Byron S, Gama ACC, Chaves CR. Interferência do tempo de experiência na concordância da análise perceptivo-auditiva. *Distúrb Comum*. 2016;28:415–422.
- Freitas SV, Pestana PM, Almeida V, et al. Audio-perceptual evaluation of Portuguese voice disorders—an inter and intra-judge reliability study. *J Voice*. 2014;28:210–215.
- Englert M, Madazio G, Gielow I, et al. Perceptual error identification of human and synthesized voices. *J Voice*. 2016;30:639, e17-e23.

**Effect of synthesized voice anchors on auditory-perceptual vocal evaluation**

**Efeito de emissões âncoras de vozes sintetizadas na avaliação perceptivo-  
auditiva da voz**

Priscila Campos Martins dos Santos – Programa de Pós-graduação em Ciências Fonoaudiológicas da Faculdade de Medicina da Universidade Federal de Minas Gerais – Belo Horizonte (MG), Brasil.

Maurílio Nunes Vieira - Departamento de Engenharia Eletrônica da Escola de Engenharia da Universidade Federal de Minas Gerais – Belo Horizonte (MG), Brasil.

João Pedro Hallack Sansão - Departamento das Engenharias de Telecomunicações e Mecatrônica da Universidade Federal de São João Del Rey - São João Del Rey (MG), Brasil.

Ana Cristina Côrtes Gama - Departamento de Fonoaudiologia da Faculdade de Medicina da Universidade Federal de Minas Gerais – Belo Horizonte (MG), Brasil.

Trabalho realizado no Programa de Pós-graduação em Ciências Fonoaudiológicas da Faculdade de Medicina da Universidade Federal de Minas Gerais - UFMG - Belo Horizonte (MG), Brasil, com bolsa concedida pela Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), sob o processo de número APQ-02594-15.

Endereço para correspondência: Priscila Campos Martins dos Santos - Av. Professor Alfredo Balena, 190 - Sala 251, Belo Horizonte (MG), Brasil, CEP: 30130-100. E-mail: priscila.fonoaudiologia@gmail.com.

## Resumo

**Objetivo:** Analisar se a utilização de emissões âncoras de vozes sintetizadas na avaliação perceptivo-auditiva melhora a concordância intra e interavaliador. **Método:** Trata-se de um estudo de natureza quantitativa. Foram selecionados 32 avaliadores inexperientes que realizaram, em um aplicativo criado pelos autores, duas atividades: Atividade Calibrador Ativo – avaliação perceptivo-auditiva dos parâmetros rugosidade e soprosidade como 0-ausência de desvio, 1-desvio leve, 2-desvio moderado ou 3-desvio intenso de 25 vozes com o apoio de emissões âncoras de vozes sintetizadas; e Atividade Calibrador Inativo – avaliação perceptivo-auditiva dessas mesmas vozes sem o apoio de emissões vocais âncoras. As vozes foram aleatorizadas em cada atividade, e a ordem de realização das atividades foi sorteada para cada avaliador, sendo que a segunda atividade foi realizada 15 dias após a primeira. Para análise da concordância intra e interavaliadores foi utilizado o coeficiente Kappa, e para comparação entre as concordâncias foi utilizado o intervalo de confiança (IC). **Resultados:** A concordância interavaliadores foi significativamente maior para o grau intenso do parâmetro soprosidade na Atividade Calibrador Ativo quando comparada à Atividade Calibrador Inativo, assim como a concordância intra-avaliadores do parâmetro rugosidade. **Conclusão:** O uso de emissões âncoras de vozes sintetizadas diretamente na avaliação melhora a concordância intra e interavaliadores na análise perceptivo-auditiva da voz, podendo ser utilizado na clínica fonoaudiológica.

## Introdução

A análise perceptivo-auditiva tem sido a principal ferramenta de avaliação da qualidade vocal nas clínicas e pesquisas fonoaudiológicas devido às suas vantagens: permite descrições perceptivas que abrangem diversos parâmetros vocais, é um método rápido, indolor e confortável ao paciente, e, além disso, não depende de equipamentos, gerando um baixo custo (Oates, 2009). Porém, a qualidade vocal caracterizada por mais de um parâmetro concomitantemente é um fator frequente e que torna essa avaliação complexa, uma vez que o avaliador precisa distinguir auditivamente os parâmetros em uma mesma voz e isolá-los para que possa analisá-los, podendo ser influenciado pelos seus padrões internos, construídos a partir de experiências e treinamentos prévios (Behlau, 2001; Kreiman, Gerrat, & Ito, 2007; Solomon, Helou, & Stojadinovic, 2011; Chaves, Campbell, & Gama, 2016). Essa subjetividade, desvantagem da análise perceptivo-auditiva, gera alta variabilidade na concordância intra e interavaliadores, prejudicando a confiabilidade dessa avaliação (Chan, & Yiu, 2002; Yiu, Murdoch, Hird, & Lau, 2002; Chan, & Yiu, 2006).

Diante disso, estudos recentes têm apontado o uso de emissões vocais âncoras em treinamentos perceptivo-auditivos da avaliação vocal como uma ferramenta para aumentar a confiabilidade dessa avaliação (Chan, & Yiu, 2006; Santos, Vieira, Sansão, & Gama, 2017). As emissões vocais âncoras são vozes selecionadas, em concordância por pelo menos dois avaliadores, para serem usadas como referência de um determinado parâmetro e grau de desvio vocal (Awan, & Lawson, 2009; Eadie, & Kapsner-Smith, 2011; Sofranko, & Prosek, 2014). As vozes usadas como âncoras podem ser naturais, ou seja, vozes humanas; ou

sintetizadas, que são vozes construídas a partir de cálculos matemáticos. A principal vantagem do uso da voz humana como emissões âncoras é a sua naturalidade. Porém, junto a essa naturalidade está associado o fato de que geralmente as vozes são caracterizadas por mais de um parâmetro concomitantemente, o que pode ser apontado como a principal desvantagem do uso deste tipo de emissão, uma vez que dificulta a classificação das vozes. Em contrapartida, apesar das emissões vocais sintetizadas apresentarem como desvantagem a característica de artificialidade das vozes, por vezes com traços robóticos e pouco naturais, sua principal vantagem é a possibilidade de manipulação dos parâmetros acústicos conforme desejar ou necessitar, possibilitando a análise de cada parâmetro vocal separadamente. Por isso, acredita-se que a emissão vocal sintetizada seria o tipo ideal para ser usado como âncora em treinamentos perceptivo-auditivos da voz (Yiu, Murdoch, Hird, & Lau, 2002).

Vários estudos têm usado a emissão âncora de voz sintetizada associada ao treinamento perceptivo-auditivo e analisado seu efeito na concordância intra e interavaliadores da avaliação da qualidade vocal (Chan, & Yiu, 2002; Chan, & Yiu, 2006; Eadie, & Baylor, 2006). Eadie, & Baylor (2006) desenvolveram uma pesquisa com avaliadores inexperientes, e mostraram que o uso de emissões vocais âncoras no treinamento melhorou a concordância intra e interavaliadores na avaliação pós-treinamento.

Chan, & Yiu (2006), ao compararem o uso de âncoras ao método de pareamento no treinamento de avaliadores experientes, observaram que os dois métodos facilitaram a avaliação perceptivo-auditiva, mostrando uma melhora significativa na precisão da avaliação após o treinamento. Contudo, perceberam que o uso de emissões vocais âncoras no treinamento permite que essa referência seja

memorizada e resgatada durante as tarefas de avaliação perceptivo-auditiva, por ser um método mais semelhante à avaliação da qualidade vocal que o método de pareamento.

Estes mesmos autores, em outro estudo (Chan, & Yiu, 2002), analisaram o efeito de emissões âncoras de vozes naturais e sintetizadas no treinamento perceptivo-auditivo, e apontaram que quando as âncoras são associadas ao treinamento estabilizam os padrões internos dos avaliadores, melhorando a concordância da avaliação. Concluíram ainda que as emissões âncoras de vozes sintetizadas mostraram-se mais confiáveis que as âncoras naturais.

Diante do exposto, as emissões vocais âncoras têm sido frequentemente associados ao treinamento perceptivo-auditivo para posterior análise de seu efeito na avaliação vocal (Awan, & Lawson, 2009; Santos, Vieira, Sansão, & Gama, 2017). Poucos estudos analisam o uso de emissões vocais âncoras diretamente na avaliação da voz (Gerrat, Kreiman, Antonanzas-Barroso, & Berke, 1993; Eadie, & Kapsner-Smith, 2011). É lícito supor que o uso dessas emissões âncoras durante a avaliação perceptivo-auditiva da voz eliminaria a necessidade de memorização prévia de vozes referências por meio de treinamentos anteriores ou periódicos, assim como diminuiria a influência dos padrões internos dos avaliadores na classificação vocal, uma vez que o avaliador teria uma emissão referência à sua disposição (Gerrat, Kreiman, Antonanzas-Barroso, & Berke, 1993), assim como um instrumentista usa os estímulos oferecidos por um afinador como referência ao afinar seu instrumento. A emissão âncora de voz sintetizada facilitaria a diferenciação dos parâmetros avaliados e dos seus respectivos graus de desvio, uma vez que permite a análise de um parâmetro isolado, o que geralmente não é possível com as âncoras de vozes humanas (Chan, & Yiu, 2006; Goldstone, 1998).

Portanto, o objetivo do presente estudo foi analisar a concordância intra e interavaliadores da avaliação perceptivo-auditiva utilizando emissões âncoras de vozes sintetizadas.

## Método

A presente pesquisa foi aprovada pelo Comitê de Ética em Pesquisa (COEP) da Universidade Federal de Minas Gerais sob o parecer CAAE-37872314.2.0000.5149. Trata-se de um estudo de natureza quantitativa.

Antes de iniciar as atividades, os avaliadores leram o Termo de Consentimento Livre e Esclarecido (TCLE) e selecionaram a opção “Aceito” para prosseguir na participação da pesquisa. Em seguida, responderam a um breve questionário fornecendo dados sobre sua experiência em treinamento auditivo e idade; e receberam uma apresentação inicial da pesquisa. Enfim, os avaliadores realizaram as atividades de avaliação perceptivo-auditiva das vozes.

Foram criadas pelas pesquisadoras duas atividades para avaliação perceptivo-auditiva e disponibilizadas em um aplicativo. Na Atividade Calibrador Ativo, os avaliadores avaliaram as vozes com o apoio de emissões âncoras de vozes sintetizadas, e na Atividade Calibrador Inativo, os avaliadores avaliaram as vozes sem apoio de emissões vocais âncoras. Em ambas atividades utilizou-se uma escala de quatro pontos (0 – ausência de desvio, 1 – grau de desvio leve, 2 – grau de desvio moderado e 3 – grau de desvio intenso) quanto aos parâmetros rugosidade (R) e soproidade (B), sendo que considerou-se como rugosidade a qualidade vocal que apresentasse qualquer irregularidade perceptível durante a produção vocal, e como soproidade a qualidade vocal com escape de ar audível durante a produção da voz (Hirano, 1981).

As atividades receberam o nome de Calibrador Auditivo, pois, a emissão âncora de voz sintetizada à disposição durante a avaliação perceptivo-auditiva se assemelha aos estímulos oferecidos por um afinador como referência para o

musicista ao afinar seu instrumento. Sendo assim, na atividade em que as emissões âncoras de vozes sintetizadas estão presentes o Calibrador está Ativo – Atividade Calibrador Ativo, enquanto na atividade em que as emissões âncoras de vozes sintetizadas estão ausentes o Calibrador está Inativo – Atividade Calibrador Inativo.

A ordem de realização das atividades foi sorteada para cada participante, sendo que a segunda atividade foi realizada exatamente 15 dias após a primeira (Figura 1). É possível observar na literatura o uso de um intervalo de pelo menos uma semana entre atividades de avaliação, a fim de evitar qualquer memorização (Helou et al., 2010; Silva, Simões-Zenari, & Nemr, 2012; Brinca, Batista, Tavares, Pinto, & Araujo, 2015). Todas as atividades da pesquisa foram realizadas no ano de 2017.

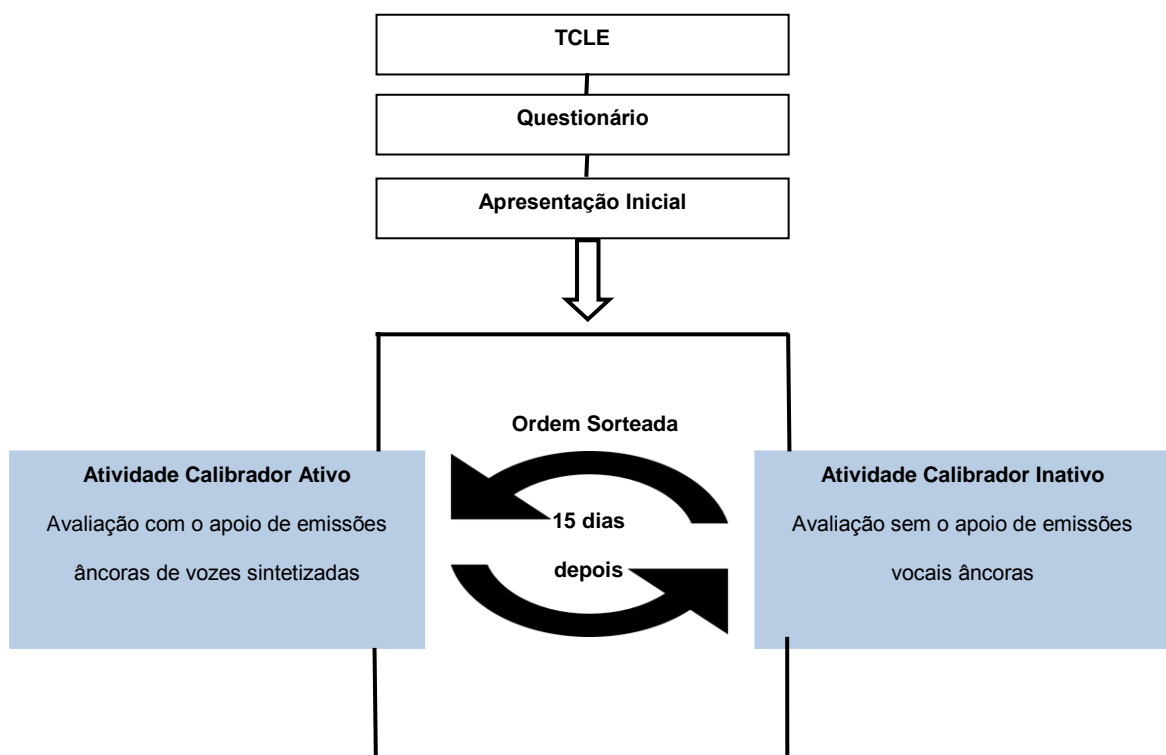


Figura 1 – Fluxograma do Calibrador Auditivo.

Cada atividade será descrita a seguir.

## 1. Atividade Calibrador Ativo

A atividade que utilizou emissões âncoras de vozes sintetizadas na avaliação perceptivo-auditiva foi chamada de Atividade Calibrador Ativo.

### 1.1. Processo.

Durante essa atividade os avaliadores escutaram a voz natural a ser avaliada, em seguida escutaram as emissões âncoras de vozes sintetizadas para cada grau do parâmetro R, e novamente escutaram a voz à ser avaliada, classificando-a em 0 – ausência de desvio, 1 – desvio leve, 2 – desvio moderado ou 3 – desvio intenso para este parâmetro. Repetiram o processo para classificar a mesma voz quanto ao parâmetro B (Figura 2).

**Atividade Calibrador Ativo**  
**VOZ 1**


---


**Rugosidade**

Qualquer irregularidade perceptível durante a produção vocal.

  
**R0**

  
**R1**


  
**R2**


  
**R3**

**Soprosidade**


Qualquer escape de ar audível durante a produção vocal.

  
**B0**

  
**B1**

  
**B2**

  
**B3**



Grau de rugosidade: \_\_\_\_\_

Grau de soprosidade: \_\_\_\_\_

---





Figura 2 – Atividade Calibrador Ativo do aplicativo.

A definição escrita dos parâmetros foi disponibilizada durante todas as etapas da Atividade Calibrador Ativo.

### 1.2. Seleção das emissões vocais para avaliação.

Para compor a amostra de vozes naturais à serem avaliadas, utilizou-se o banco de vozes do ambulatório de uma universidade, formado por 381 vozes, amostras da emissão da vogal /a/ sustentada de forma habitual, de indivíduos de ambos os gêneros com idade a partir de 18 anos. Duas avaliadoras, fonoaudiólogas, especialistas em voz, com mais de cinco anos de experiência em avaliação perceptivo-auditiva, analisaram individualmente as vozes, utilizando o fone de ouvido supra-auricular modelo *Multilaser Vibe Headphone* estéreo. Classificaram as vozes conforme o parâmetro predominante, R ou B, e o grau geral de desvio vocal (0 – ausência de desvio, 1 – desvio leve, 2 – desvio moderado, 3 – desvio intenso), por meio da escala GRBASI.

Foram considerados os seguintes critérios de inclusão: vozes naturais de sujeitos do sexo feminino e masculino, com idade a partir de 18 anos, com um parâmetro predominante de variados graus de desvio vocal; vozes que apresentaram a mesma classificação pelas duas avaliadoras.

Foram selecionadas três emissões vocais para cada grau dos parâmetros predominantes R e B, sendo que um grau de um dos parâmetros recebeu quatro emissões vocais, com o total de 25 vozes. Acrescentou-se 20% das vozes com o objetivo de analisar a concordância intra-avaliador, totalizando 27 vozes.

Em todas as etapas da pesquisa as vozes foram identificadas por números.

### 1.3. Seleção das emissões vocais âncoras para o treinamento.

A amostra das emissões vocais âncoras foi composta por vozes sintetizadas. Para a construção das vozes sintetizadas neutras (N) ou contendo o parâmetro R ou B com diferentes graus de desvio vocal utilizou-se como fonte (fluxo glótico) um modelo paramétrico que permite o controle da frequência fundamental, do *jitter*, do *shimmer* e da relação sinal ruído. Como filtro, utilizou-se um trato vocal que modela a vogal /a/, extraído de voz natural por técnica de predição linear. As emissões vocais foram construídos por um engenheiro, totalizando 300 vozes sintetizadas (Vieira, Sansão, & Yehia, 2014).

Para a análise do grau de naturalidade, e da qualidade das vozes sintetizadas, foram selecionados três avaliadores, fonoaudiólogos com mais de cinco anos de experiência em avaliação vocal, que realizaram individualmente a análise de cada voz em três aspectos. Primeiramente foi realizada uma análise perceptivo-auditiva da naturalidade da voz (relacionado ao quanto o ouvinte percebe a voz como humana), indicando em uma escala visual analógica (EVA) de 100mm o quanto consideravam aquela voz natural, sendo zero não natural e 10 indicando o máximo de naturalidade. Em seguida, a voz foi classificada como neutra, rugosa ou soprosa. Por fim, foi mensurado, também por meio de uma EVA de 100mm, o grau de desvio vocal para o parâmetro em que foi classificada anteriormente (R ou B). Os valores encontrados para o desvio vocal das vozes classificadas como R ou B por meio da EVA, foram convertidos segundo sugerido pela literatura (Baravieira, 2016), conforme apresentado na Tabela 1.

Grau de desvio	Correlação da classificação do desvio vocal pela escala visual analógica e escala numérica	
	Rugosa (mm)	Soprosa (mm)
Neutro	0 – 8,5	0 – 8,5
Leve	8,5 – 28,5	8,5 – 33,5
Moderado	28,5 – 59,5	33,5 – 52,5
Intenso	A partir de 59,5	A partir de 52,5

Tabela 1 – Correlação da classificação do desvio vocal pela escala visual analógica e escala numérica.

Foram selecionadas como âncoras as vozes sintetizadas de diferentes graus de desvio, para cada parâmetro, classificadas com maior naturalidade por pelo menos dois avaliadores. A amostra das emissões vocais âncoras foi composta por uma emissão de cada grau – ausência de desvio, desvio leve, moderado e intenso, de cada parâmetro – R e B, totalizando oito vozes.

As vozes neutras ou com menor desvio vocal foram classificadas com maior naturalidade para os dois parâmetros, diminuindo a naturalidade conforme aumenta o grau de desvio (Tabela 2). Para o parâmetro R, a voz classificada com ausência de desvio apresentou maior naturalidade, seguida das vozes classificadas com grau de desvio leve, moderado e intenso. Para o parâmetro B, a voz com grau de desvio leve foi classificada com maior naturalidade, seguida da voz com ausência de desvio e, posteriormente, com desvio moderado e intenso. As vozes selecionadas para os graus leve, moderado e intenso do parâmetro B apresentaram maior naturalidade que as vozes selecionadas para os mesmos graus de desvio do parâmetro R.

Grau de desvio	Classificação da naturalidade das vozes (mm)		
	Neutra	Rugosa	Soprosa
	97,3		
Leve		56	86
Moderado		41	60
Intenso		37	40

Tabela 2 – Média do grau de naturalidade das vozes sintetizadas para cada parâmetro perceptivo-auditivo selecionado para a amostra.

## 2. Atividade Calibrador Inativo

A atividade que não utilizou emissões âncoras de vozes sintetizadas na avaliação perceptivo-auditiva foi chamada de Atividade Calibrador Inativo.

### 2.1. Processo.

Durante essa atividade os avaliadores escutaram a voz natural a ser avaliada e em seguida a classificaram em 0 – ausência de desvio, 1 – desvio leve, 2 – desvio moderado ou 3 – desvio intenso quanto ao parâmetro R. Repetiram o processo para classificar a mesma voz quanto ao parâmetro B.

### 2.2. Seleção das emissões vocais para avaliação.

Foram utilizadas na Atividade Calibrador Inativo as mesmas emissões vocais usadas na Atividade Calibrador Ativo. Em cada atividade as vozes foram aleatorizadas.

Durante a coleta, o pesquisador que acompanhou os avaliadores observou que a Atividade Calibrador Inativo teve duração aproximada de vinte minutos, apesar do tempo não ter sido cronometrado. Observou ainda que a Atividade Calibrador

Ativo teve duração discretamente maior quando comparada a Atividade Calibrador Inativo. Os avaliadores podiam escutar as vozes quantas vezes julgassem necessário, desde que respeitassem o comando fornecido – escutar a voz a ser avaliada, em seguida as emissões vocais âncoras do parâmetro R, e novamente a voz a ser avaliada, repetindo o processo ao classificar quanto ao parâmetro B. Foi utilizado fone de ouvido supra-auricular modelo *Multilaser Vibe Headphone* estéreo durante todos os procedimentos e as avaliações foram realizadas individualmente.

### Seleção dos avaliadores

Para determinar a quantidade de 32 avaliadores foi realizado um cálculo amostral, considerando 25 observações (vozes a serem avaliadas) e oito variáveis (parâmetros R e B com ausência de desvio, desvio leve, moderado e intenso), por meio do teste Kappa proposto por Fleiss, com poder estatístico de 80% e nível de significância de 5%.

Foram selecionados 32 indivíduos para a avaliar as vozes, sendo 27 do sexo feminino e cinco do sexo masculino, estudantes do primeiro ao terceiro período do curso de graduação em Fonoaudiologia, sem experiência ou treinamento prévio em avaliação perceptivo-auditiva da voz, com idade de 17 a 24 anos (média = 19,66 anos). Foram considerados os seguintes critérios de inclusão: responder ao questionário inicial, participar de todas as atividades, não possuir experiência prévia em avaliação perceptivo-auditiva da voz, e ausência de queixas auditivas.

Em nenhum momento os avaliadores foram identificados.

Para análise da concordância intra e interavaliadores foi utilizado o coeficiente Kappa, e para comparação entre as concordâncias foi utilizado o intervalo de

confiança (IC). Para realizar a análise estatística foi utilizado o *software* Stata versão 12. Em todas as análises foi considerado um nível de significância de 5%.

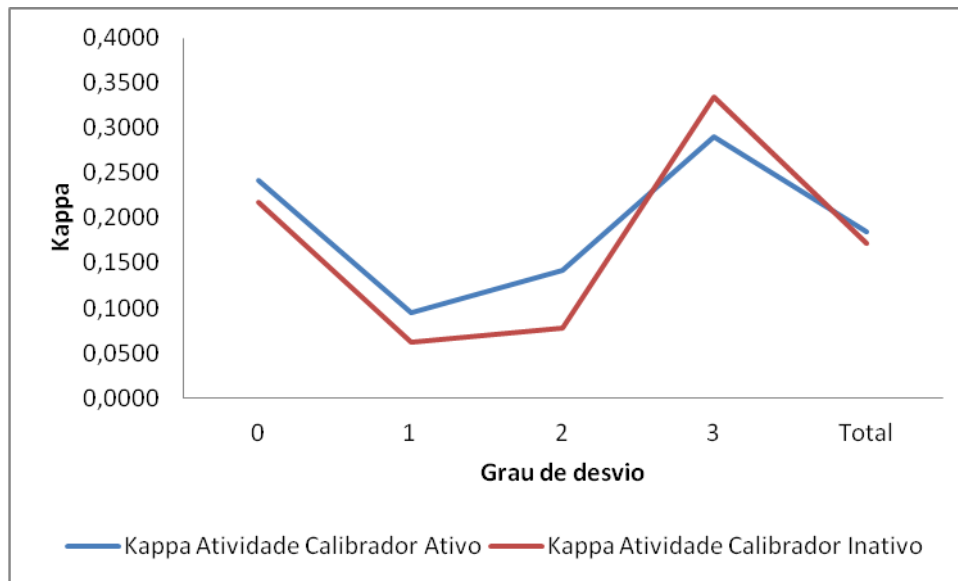
## Resultados

Apesar de não haver diferença estatística ao observarmos os IC (Tabela 3), pode-se verificar que a concordância interavaliadores foi maior na Atividade Calibrador Ativo – com emissões âncoras de vozes sintetizadas, para os graus 0, 1 e 2 e menor para o grau 3 do parâmetro R quando comparado à concordância na Atividade Calibrador Inativo – sem emissões vocais âncoras, para o mesmo parâmetro e graus de desvio (Tabela 3 e Figura 3).

Tabela 3 - Concordância interavaliadores da Atividade Calibrador Ativo – com emissões âncoras de vozes sintetizadas, e da Atividade Calibrador Inativo – sem emissões vocais âncoras, para cada grau de desvio quanto ao parâmetro Rugosidade, por meio do coeficiente Kappa.

Grau	Atividade Calibrador Ativo			Atividade Calibrador Inativo		
	Kappa	IC		Kappa	IC	
0	0,2412	0,1947	0,2877	0,2177	0,1698	0,2656
1	0,0943	0,0388	0,1498	0,0619	0,0044	0,1194
2	0,1421	0,0895	0,1947	0,0778	0,0213	0,1343
3	0,2898	0,2463	0,3333	0,3346	0,2938	0,3754
<b>Total</b>	0,1846	0,1346	0,2346	0,1724	0,1216	0,2232

Nota: para análise estatística foi considerado o coeficiente Kappa ponderado e o intervalo de confiança (IC).



**Figura 3 - Comparação entre a concordância interavaliadores na Atividade Calibrador Ativo - com emissões âncoras de vozes sintetizadas, e Atividade Calibrador Inativo - sem emissões vocais âncoras, para cada grau de desvio quanto ao parâmetro Rugosidade, por meio do coeficiente Kappa ponderado.**

Quanto à soproside, não houve diferença estatística ao observarmos os IC (Tabela 4) dos graus 0, 1 e 2. Porém, também é possível verificar que a concordância interavaliadores foi maior na Atividade Calibrador Ativo – com emissões âncoras de vozes sintetizadas, que na Atividade Calibrador Inativo – sem emissões vocais âncoras para estes graus. A concordância interavaliadores para o grau 3 de soproside mostrou-se estatisticamente maior na Atividade Calibrador Ativo quando comparada à Atividade Calibrador Inativo (Tabela 4 e Figura 4). Observa-se que a concordância interavaliadores foi maior para os graus 0 e 3 dos dois parâmetros avaliados (Figuras 3 e 4).

Tabela 4 - Concordância interavaliadores da Atividade Calibrador Ativo - com emissões âncoras de vozes sintetizadas, e da Atividade Calibrador Inativo – sem emissões vocais âncoras, para cada grau de desvio quanto ao parâmetro Soprosidade, por meio do coeficiente Kappa.

Grau	Atividade Calibrador Ativo			Atividade Calibrador Inativo		
	Kappa	IC		Kappa	IC	
0	0,3279	0,2867	0,3691	0,3060	0,2635	0,3485
1	0,1147	0,0604	0,1690	0,0850	0,0289	0,1411
2	0,1572	0,1055	0,2089	0,0927	0,0371	0,1483
3	0,5321	0,5034	0,5608	0,4498	0,4161	0,4835
<b>Total</b>	<b>0,2738</b>	<b>0,2293</b>	<b>0,3183</b>	<b>0,2313</b>	<b>0,1842</b>	<b>0,2784</b>

Nota: para análise estatística foi considerado o coeficiente Kappa ponderado e o intervalo de confiança (IC).

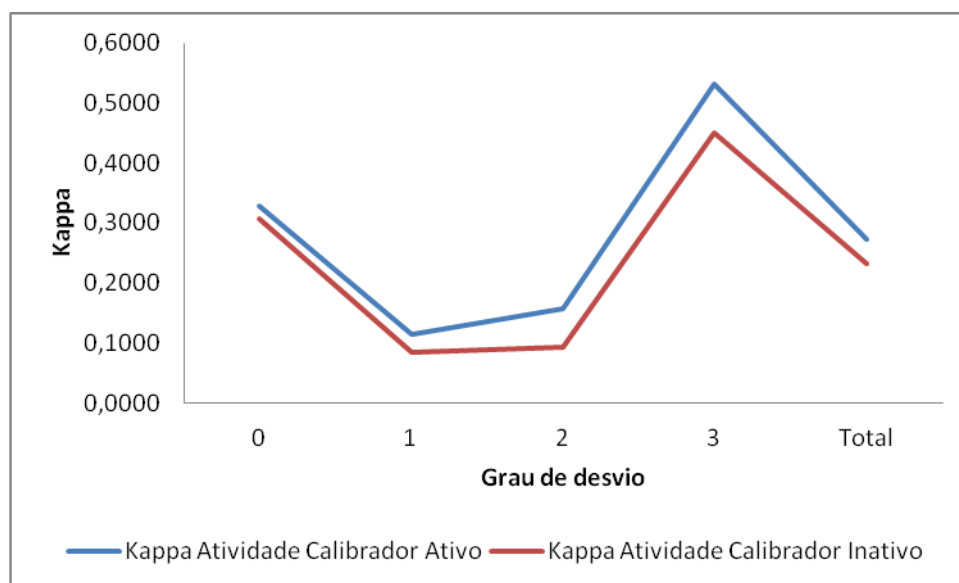


Figura 4 - Comparação entre a concordância interavaliadores na Atividade Calibrador Ativo – com emissões âncoras de vozes sintetizadas, e Atividade Calibrador Inativo - sem emissões vocais âncoras, para cada grau de desvio quanto ao parâmetro Soprosidade, por meio do coeficiente Kappa ponderado.

A concordância intra-avaliadores mostrou-se estatisticamente maior na Atividade Calibrador Ativo quando comparada à Atividade Calibrador Inativo para o parâmetro rugosidade (Tabela 5). Houve também uma maior concordância na Atividade Calibrador Ativo para o parâmetro soprosidade, apesar de não ser observada diferença estatística (Tabela 5 e Figura 5).

Tabela 5 - Concordância intra-avaliadores da Atividade Calibrador Ativo – com emissões âncoras de vozes sintetizadas, e da Atividade Calibrador Inativo – sem emissões vocais âncoras, quanto aos parâmetros Rugosidade e Soproidade, por meio do coeficiente Kappa.

	Atividade Calibrador Ativo			Atividade Calibrador Inativo		
	Kappa	IC		Kappa	IC	
<b>Rugosidade</b>	0,5025	0,4862	0,5188	0,3264	0,3105	0,3423
<b>Soproidade</b>	0,5444	0,5284	0,5604	0,5207	0,5047	0,5367

Nota: para análise estatística foi considerado o coeficiente ponderado Kappa e o intervalo de confiança (IC).

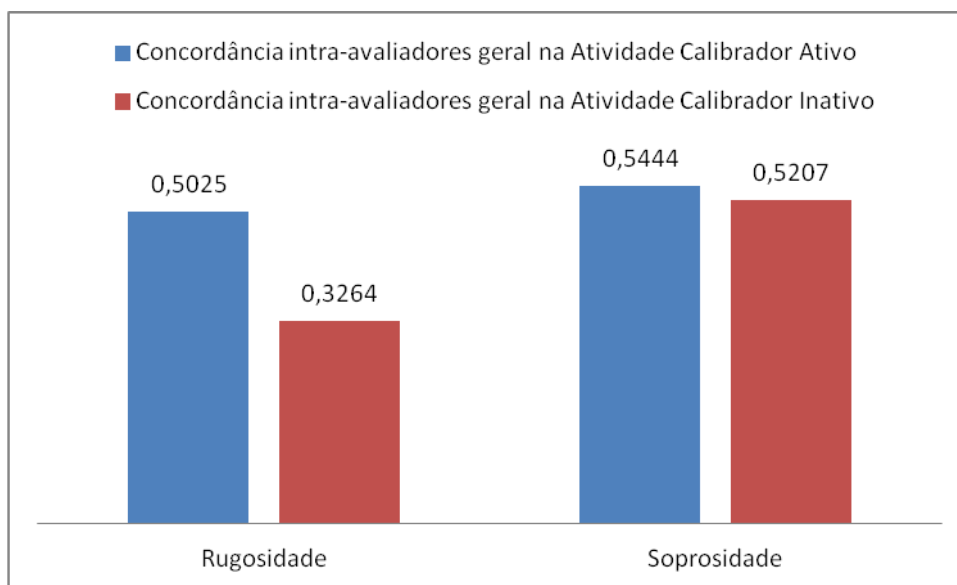


Figura 5 - Comparação entre a concordância intra-avaliadores na Atividade Calibrador Ativo – com emissões âncoras de vozes sintetizadas, e Atividade Calibrador Inativo - sem emissões vocais âncoras , para os parâmetros Rugosidade e Soproidade, por meio do coeficiente Kappa ponderado.

## Discussão

No presente estudo optou-se pelo uso de emissões âncoras de vozes sintetizadas. Pesquisas sugerem que pode-se reduzir a variabilidade na classificação da qualidade vocal substituindo os padrões internos instáveis dos ouvintes usando padrões externos, como as vozes âncoras, ou vozes de referência para diferentes qualidades vocais, podendo ser comparadas à amostra de voz a ser julgada (Kreiman, Gerratt, Kempster, Erman, & Berke, 1993; Yiu, Murdoch, Hird, & Lau, 2002; Awan, & Lawson, 2009; Eadie, & Kapsner-Smith, 2011; Solomon, Helou, & Stojadinovic, 2011; Sofranko, & Prosek, 2014; Santos, Vieira, Sansão, & Gama, 2017). Outro estudo (Canals-Fortuny, & Vila-Rovira, 2017) que analisou a influência de âncoras nas respostas a um questionário de qualidade de vida (VHI-10), apontou que os pacientes que responderam ao questionário utilizando as âncoras apresentaram menos dispersão e um coeficiente de variação menor do que aqueles que responderam ao questionário sem o apoio de emissões vocais âncoras.

Optou-se ainda, no presente estudo, pelo uso de vozes sintetizadas como âncoras, o que permitiu a escuta de cada parâmetro vocal isoladamente durante a avaliação (Yiu, Murdoch, Hird, & Lau, 2002); e pela seleção de avaliadores inexperientes, a fim de eliminar a influência de qualquer experiência ou treinamento prévio, assim como de padrões internos, possibilitando analisar puramente o efeito da âncora na avaliação.

Apesar do uso promissor de vozes sintetizadas, essa ainda não é uma prática comum, devido à dificuldade de produzir as vozes que sejam consideradas naturais pelo ouvinte. Por isso, para selecionar as vozes sintetizadas foi realizada a classificação da naturalidade das vozes para cada um dos parâmetros. Verificou-se

alta qualidade das amostras de vozes sintetizadas principalmente para os graus ausência de desvio e desvio leve para os parâmetros rugosidade (R) e soproidade (B), diminuindo a naturalidade conforme o grau do desvio vocal aumentou. Outro estudo apontou alta qualidade das vozes sintetizadas, mostrando maior acerto da classificação das vozes como sintetizada para graus mais intensos dos mesmos parâmetros (Englert, Gielow, Lucero & Behlau, 2016). As discrepâncias entre os estudos podem ser justificadas por questões metodológicas. Os estudos desenvolveram as vozes sintetizadas utilizando diferentes métodos matemáticos; enquanto a presente pesquisa analisou o grau de naturalidade, a literatura (Englert, Gielow, Lucero & Behlau, 2016) avaliou quais vozes, entre um banco de vozes humanas e sintetizadas, eram identificadas corretamente. A diferente forma de avaliar a naturalidade nos dois estudos provavelmente impactou os resultados. Estudos futuros são necessários para melhor compreensão da percepção auditiva de vozes sintetizadas, quando comparada com emissões vocais humanas.

Apenas dois artigos foram encontrados na literatura em que emissões vocais âncoras foram utilizados diretamente na avaliação perceptivo-auditiva da voz (Gerrat, Kreiman, Antonanzas-Barroso, & Berke, 1993; Eadie, & Kapsner-Smith, 2011). Um destes estudos (Eadie, & Kapsner-Smith, 2011) foi realizado com três grupos de avaliadores, incluindo avaliadores experientes e inexperientes. Os parâmetros avaliados foram o grau geral de desvio vocal e o esforço vocal e classificados como grau 1, 2 ou 3. Foi utilizada uma escala visual analógica (EVA) de 100mm para avaliação e emissões âncoras de vozes naturais. Dois grupos, compostos por avaliadores inexperientes e experientes, avaliaram as vozes em uma EVA primeiramente sem o apoio de emissões vocais âncoras e, posteriormente, com a âncora; e um terceiro grupo, grupo controle composto por avaliadores

inexperientes, realizou a avaliação apenas com o apoio de âncoras. A concordância intra e interavaliadores mostraram-se significativamente maior na avaliação com o apoio da emissão vocal âncora para os dois parâmetros avaliados.

Outro estudo (Gerrat, Kreiman, Antonanzas-Barroso, & Berke, 1993) realizado com âncoras na avaliação, utilizou emissões de vozes sintetizadas. Foi analisado apenas o parâmetro rugosidade por avaliadores experientes por meio de duas avaliações. Na primeira avaliação os avaliadores escutavam as vozes a serem avaliadas, sem apoio da emissão vocal âncora de voz sintetizada, e as classificava em uma escala de cinco pontos em que um indicava voz normal e cinco definia o grau intenso de rugosidade. Já na segunda avaliação, cada ponto da escala de cinco pontos era representado por uma voz sintetizada, emissão âncora. O participante deveria escutar as âncoras sintetizadas duas vezes e depois escutar a voz a ser avaliada e selecionar a emissão âncora de voz sintetizada com classificação mais semelhante à voz em avaliação. Os avaliadores podiam escutar as vozes quantas vezes julgassem necessário e foram instruídos a ignorar os outros desvios presentes na voz, concentrando-se apenas na rugosidade. Verificou-se alta concordância para as duas escalas. Porém, a concordância intra e interavaliadores foi significativamente maior na avaliação por meio da escala ancorada. O estudo mostrou ainda que dois avaliadores irão concordar significativamente melhor na escala ancorada do que na escala sem âncoras.

O tempo de experiência dos fonoaudiólogos impacta positivamente na concordância interavaliadores, sugerindo que a experiência nesta análise tende a uniformizar o processo de julgamento auditivo de vozes disfônicas (Oliveira, Gama, & Chaves, 2016). Foi possível verificar essa relação no presente estudo ao selecionar para a pesquisa avaliadores inexperientes e oferecer a eles as mesmas

referências de vozes para avaliação, verificando-se uma melhora na concordância interavaliadores na análise de vozes soprosas de grau intenso e na concordância intra-avaliador de vozes rugosas.

No presente estudo, a concordância interavaliadores para o parâmetro rugosidade foi maior na Atividade Calibrador Ativo – com emissões âncoras de vozes sintetizadas, para os graus 0, 1 e 2 e menor para o grau 3 do parâmetro R quando comparado à concordância na Atividade Calibrador Inativo – sem emissões vocais âncoras para o mesmo parâmetro e graus, apesar de não haver diferença estatística ao observarmos os IC. O resultado corrobora a literatura (Gerrat, Kreiman, Antonanzas-Barroso, & Berke, 1993) que aponta uma concordância interavaliadores significativamente maior para rugosidade em avaliação realizada com o apoio de emissões vocais âncoras quando comparada a avaliação sem âncoras, apesar do estudo não descrever a concordância por grau de desvio vocal para rugosidade.

Quanto à soproidade, no presente estudo não houve diferença estatística ao observarmos os IC (Tabela 4) dos graus 0, 1 e 2. Porém, também é possível verificar que a concordância interavaliadores foi maior na Atividade Calibrador Ativo, que na Atividade Calibrador Inativo. A concordância interavaliadores para o grau 3 de soproidade mostrou-se estatisticamente maior na Atividade Calibrador Ativo. Não foram encontrados estudos na literatura em que foram utilizados emissões âncoras de vozes sintetizadas diretamente na avaliação do parâmetro soproidade. Porém, um estudo em que este mesmo parâmetro foi avaliado após o treinamento com emissão vocal âncora, verificou-se o aumento significativo da concordância interavaliadores (Eadie, & Baylor, 2006).

A concordância intra-avaliadores mostrou-se estatisticamente maior na Atividade Calibrador Ativo quando comparada à concordância na Atividade

Calibrador Inativo para o parâmetro rugosidade no presente estudo. Este resultado corrobora a literatura (Gerrat, Kreiman, Antonanzas-Barroso, & Berke, 1993), que aponta uma concordância intra-avaliadores significativamente maior para rugosidade em avaliação realizada com o apoio de emissões vocais âncoras quando comparada a avaliação sem âncoras.

No presente estudo houve também uma maior concordância na Atividade Calibrador Ativo para o parâmetro soprosidade, apesar de não ser observada diferença estatística. Estudo em que este mesmo parâmetro foi avaliado após o treinamento com emissão vocal âncora, verificou-se o aumento da concordância intra-avaliadores (Eadie, & Baylor, 2006), embora a diferença não tenha sido estatisticamente significativa.

Por meio dos resultados encontrados, infere-se que o uso de emissões âncoras de vozes sintetizadas diretamente na avaliação melhorara a concordância intra e interavaliadores na análise perceptivo-auditiva da voz. Estudo anterior (Eadie, & Kapsner-Smith, 2011) apontou que avaliadores experientes apresentaram menor variabilidade da concordância na avaliação com apoio da emissão vocal âncora. Outro estudo (Oliveira, Gama, & Chaves, 2016) apontou o impacto positivo da experiência dos avaliadores na concordância interavaliadores da análise perceptivo-auditiva da voz. Diante disso, ressalta-se a importância da realização de outros estudos com emissões âncoras de vozes sintetizadas na avaliação perceptivo-auditiva com avaliadores experientes.

## **Conclusão**

A utilização de emissões âncoras de vozes sintetizadas na avaliação perceptivo-auditiva de vozes, melhora a concordância interavaliador na análise de vozes soprosas de grau intenso, e na concordância intra-avaliador de vozes rugosas, podendo ser usados na clínica fonoaudiológica.

## Referências

**Awan, S.N., & Lawson, L.L.** (2009). The effect of anchor modality on the reliability of vocal severity ratings. *Journal of Voice*, 23(3), 341–352.

**Baravieira, P.B., Brasolotto, A.G., Montagnoli, A.N., Silvério, K.C.A., Yamasaki, R., & Behlau M.** (2016). Análise perceptivo-auditiva de vozes rugosas e soprosas: correspondência entre a escala visual analógica e a escala numérica. *CoDAS*, 28(2), 163-167.

**Behlau, M.** (2001). *Voz: o livro do especialista*. Rio de Janeiro, RJ: Revinter.

**Brinca, L., Batista, A.P., Tavares, A.I., Pinto, P.N., & Araújo, L.** (2015). The effect of anchors and training on the reliability of voice quality ratings for different types of speech stimuli. *Journal of Voice*, 29(6), 776.e7-14.

**Canals-Fortuny, E., & Vila-Rovira, R.** (2017). Influence of self-generated anchors on the Voice Handicap Index-10 (VHI-10). *Journal of Voice*, 31(2), 255.e9-255.e12.

**Chan, K.M.K., & Yiu, E.M.L.** (2002). The effects of anchors and training on the reliability of perceptual voice evaluation. *Journal of Speech, Language and Hearing Research*, 45, 111–126.

**Chan, K.M.K., & Yiu, E.M.L.** (2006). A comparison of two perceptual voice evaluation training programs for naive listeners. *Journal of Voice*, 20(2), 229–241.

**Chaves, C.R., Campbell, M., & Gama, A.C.C.** (2016). The influence of native language on auditory-perceptual evaluation of vocal samples completed by Brazilian and Canadian SLPs. *Journal of Voice*. In press.

**Eadie, T.L., & Baylor, C.R.** (2006). The effect of perceptual training on inexperienced listeners' judgments of dysphonic voice. *Journal of Voice*, 20, 527-544.

**Eadie, T.L., & Kapsner-Smith, M.** (2011). The effect of listener experience and anchors on judgments of dysphonia. *Journal of Speech and Hearing Research*, 54, 430-447.

**Englert, M., Madazio, G., Gielow, I., Lucero, J., & Behlau, M.** (2016). Perceptual error identification of human and synthesized voices. *Journal of Voice*, 30(5), 639.e17-23.

**Gerratt, B.R., Kreiman, J., Antonanzas-Barroso, N., & Berke, G.S.** (1993). Comparing internal and external standards in voice quality judgments. *Journal of Speech Hearing and Research*, 36, 14-20.

**Goldstone, R.L.** (1998). Perceptual learning. *Annual Review of Psychology*, 49, 585-612.

**Helou, L.B., Solomon, N.P., Henry, L.R., Coppit, G.L., Howard, R.S., & Stojadinovic, A.** (2010). The role of listener experience on Consensus Auditory-

Perceptual Evaluation of Voice (CAPE-V) ratings of postthyroidectomy voice. *American Journal of Speech- Language Pathology*, 19(3), 248-258.

**Hirano, M.** (1981). *Clinical examination of voice*. New York: Springer-Verlag.

**Kreiman, J., Gerratt, B.R., Kempster, G.B., Erman, A., & Berke, G.S.** (1993). Perceptual evaluation of voice quality: review, tutorial, and a framework for future research. *Journal of Speech and Hearing Research*, 36, 21-40.

**Kreiman, J., Gerratt, B.R., & Ito, M.** (2007). When and why listeners disagree in voice quality assessment tasks. *Journal of the Acoustical Society of America*, 122(4), 2354-2364.

**Oates, J.** (2009). Auditory-perceptual evaluation of disordered vocal quality—pros, cons and future directions. *Folia Phoniatrica Logopaedica*, 61, 49–56.

**Oliveira, S.B., Gama, A.C.C., & Chaves, A.R.** (2016). Interferência do tempo de experiência na concordância da análise perceptivo-auditiva de vozes. *Distúrbios da Comunicação*, 28, 415-422.

**Santos, P.C.M., Vieira, M.N., Sansão, J.P.H., & Gama, A.C.C.** (2017). Effect of auditory-perceptual training with natural voice anchors on vocal quality evaluation. *Journal of Voice*. In press.

**Silva, R.S.A., Simões-Zenari, M., & Nemr, N.K.** (2012). Impacto de treinamento auditivo na avaliação perceptivo-auditiva da voz realizada por estudantes de Fonoaudiologia. *Jornal da Sociedade Brasileira de Fonoaudiologia*, 24(1), 19-25.

**Sofranko, J.L., & Prosek, R.A.** (2014). The effect of the levels and types of experience on judgment of synthesized voice quality. *Journal of Voice*, 28, 24–35.

**Solomon, N.P., Helou, L.B., & Stojadinovic, A.** (2011). Clinical versus laboratory ratings of voice using the CAPE-V. *Journal of Voice*, 25, e7–e14.

**Vieira, M.N., Sansão, J.P.H, & Yehia, H.C.** (2014). Measurement of signal-to-noise ratio in dysphonic voices by image processing of spectrograms. *Speech Communication*, 61, 17-32.

**Yiu, E.M.L., Murdoch, B., Hird, K., & Lau, P.** (2002). Perception of synthesized voice quality in connected speech by Cantonese speakers. *Journal of the Acoustical Society of America*, 112(3 Pt 1), 1091–1101.

É evidente a necessidade de melhorar a confiabilidade da avaliação perceptivo-auditiva da voz, diante do seu amplo uso na clínica e em pesquisas fonoaudiológicas. Conseqüentemente, é essencial que pesquisas sejam realizadas para estudar quais estratégias atuariam de forma mais eficaz para o aumento da concordância intra e interavaliadores.


O objetivo do primeiro bloco deste projeto foi analisar o efeito do treinamento perceptivo-auditivo com emissões âncoras de vozes naturais na concordância interavaliadores da avaliação vocal. Verificou-se que o treinamento perceptivo-auditivo com uso de emissões âncoras de vozes naturais sugere um aumento na concordância interavaliador na análise perceptivo-auditiva da voz, indicando que novas referências internas foram estabelecidas. Pesquisas futuras que avaliem o efeito do treinamento a médio e longo prazo são importantes para se analisar o quanto os treinamentos auditivos na avaliação perceptivo-auditiva pode interferir a longo prazo nos resultados desta avaliação. É importante ressaltar a necessidade da realização de treinamentos com menor tempo de duração, uma vez que o estado de atenção do avaliador tende a oscilar em atividades com longa duração, o que pode influenciar na confiabilidade da avaliação. Futuras pesquisas que avaliem o efeito do treinamento perceptivo-auditivo com o uso de emissões âncoras de vozes sintetizadas na avaliação perceptivo-auditiva são também importantes, já que podem gerar melhores resultados.

O objetivo do segundo bloco deste projeto foi analisar o efeito da avaliação perceptivo-auditiva com emissões âncoras de vozes sintetizadas na concordância intra e interavaliadores da avaliação vocal. Como resultado, verificou-se que o uso

de emissões âncoras de vozes sintetizadas na avaliação parece melhorar a concordância intra e interavaliadores na análise perceptivo-auditiva da voz. Ressalta-se a importância de estudos sobre o efeito do uso de emissões âncoras de vozes naturais e sintetizadas tanto no treinamento quanto na avaliação perceptivo-auditiva da voz com avaliadores experientes.

É essencial prosseguir com pesquisas para que se desenvolva um treinamento padronizado, ou, ainda, um instrumento de avaliação que aumente a concordância intra e interavaliadores, a fim de melhorar a confiabilidade da avaliação perceptivo-auditiva da voz.

ANEXO 1



UNIVERSIDADE FEDERAL DE MINAS GERAIS  
COMITÊ DE ÉTICA EM PESQUISA - COEP

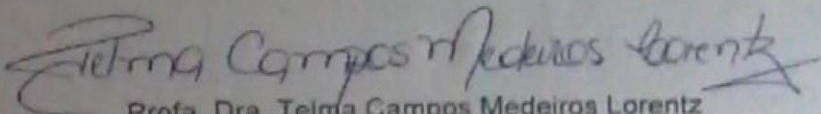
Projeto: CAAE – 37872314.2.0000.5149

Interessado(a): Profa. Ana Cristina Côrtes Gama  
Departamento de Fonoaudiologia  
Faculdade de Medicina - UFMG

**DECISÃO**

O Comitê de Ética em Pesquisa da UFMG – COEP aprovou, no dia 18 de dezembro de 2014, o projeto de pesquisa intitulado "Sítio de treinamento perceptivo-auditivo de vozes: análise da efetividade" bem como o Termo de Consentimento Livre e Esclarecido.

O relatório final ou parcial deverá ser encaminhado ao COEP um ano após o início do projeto através da Plataforma Brasil.

  
Prof. Dra. Telma Campos Medeiros Lorentz  
Coordenadora do COEP-UFMG

Av. Pádua Castanho, 667 - Unidade Administrativa II - 2º andar - Belo Horizonte - Minas Gerais - Cep 31270-901 - B11-MG  
Telefone: (51) 3409-4791 - e-mail: coep@ufmg.br

## ANEXO 2



FACULDADE DE MEDICINA  
CENTRO DE PÓS-GRADUAÇÃO

Av. Prof. Alfredo Balena 190/ sala 533  
Belo Horizonte – MG - CEP 30.130-100

Fone: (031) 3409.9641/ 3248.9640

E-mail: [cpg@medicina.ufmg.br](mailto:cpg@medicina.ufmg.br)

Curso de Pós-Graduação em Ciências Fonoaudiológicas



### Resolução nº01/2015, de 26 de março de 2015.

*Regulamenta o formato de dissertações do Curso de Pós-Graduação em Ciências Fonoaudiológicas da Faculdade de Medicina da UFMG*

O Colegiado do Programa de Pós-Graduação em Ciências Fonoaudiológicas, no uso de suas atribuições, e considerando a necessidade de regulamentar o formato das dissertações do Programa.

#### RESOLVE:

Art. 1º A dissertação de mestrado poderá ser elaborada no formato convencional e no formato de artigo.

Parágrafo único - O formato de artigo é considerado preferencial pelo colegiado do Programa.

Art. 2º O Colegiado do Curso de Pós-Graduação em Ciências Fonoaudiológicas propõe o seguinte roteiro para elaboração da dissertação no formato de artigo:

1. Capa
2. Folha de Rosto
3. Folha da Instituição
4. Declaração de Defesa
5. Resumo da dissertação/Descritores (1300 palavras/3 a 5 descritores)
6. Abstract/Keywords
7. Sumário
8. Introdução ou considerações iniciais: duas a três paginas com breve fundamentação teórica e/ou contextualização do tema cujos resultados serão apresentados sob formato de artigo ou artigos;
9. Objetivos: redigido da forma convencional (uma ou duas páginas);
10. Métodos: redigido da forma convencional e detalhado (se necessário);
11. Resultados e discussão: sob a forma de artigo ou artigos;
12. Conclusão ou considerações finais: até cinco paginas.
13. Anexos/Apêndices

Art. 3º O Colegiado do Curso de Pós-Graduação em Ciências Fonoaudiológicas propõe o seguinte roteiro para elaboração da dissertação no formato convencional:

1. Capa



2. Folha de Rosto
3. Folha da Instituição
4. Declaração de Defesa
5. Resumo da dissertação/Descritores (1300 palavras/3 a 5 descritores)
6. Abstract/Keywords
7. Sumário
8. Introdução;
9. Revisão da literatura;
10. Objetivos;
11. Métodos;
12. Resultados;
13. Discussão;
14. Conclusão;
15. Referências bibliográficas;
16. Anexos/Apêndices.

Art. 4º - Outros aspectos de formatação:

1. Referências bibliográficas: serão apresentadas após cada sessão da dissertação de acordo com as normas de Vancouver e conforme as recomendações específicas de cada periódico para os quais os artigos serão submetidos. 2. A dissertação de mestrado poderá conter os textos escritos na língua inglesa, de acordo com esta resolução.

Art. 5º. Os casos omissos e especiais serão decididos pelo Colegiado de Pós-Graduação.

Art. 6º. Esta Resolução entra em vigor na data de sua aprovação.

Ficam revogadas todas as disposições em contrário, em especial a Resolução 01/2014.

Resolução aprovada pelo Colegiado do Curso de Mestrado em  
Ciências Fonoaudiológicas em 26/03/2015.

Resolução aprovada pela Câmara de Pós-Graduação em 28/04/2015

Profa. Ana Cristina Côrtes Gama  
Coordenadora do Curso de Pós-Graduação em Ciências Fonoaudiológicas

# ASHA Journals Academy

### Manuscript Sections

Generally, scientific manuscripts should be organized as follows:

- Title page
- Abstract
- Introduction
- Methods
- Results
- Discussion
- Acknowledgments
- References
- Tables and Figures
- Appendices (optional)
- Supplemental information (optional)

Because scientific papers are organized in this way, readers know what to expect from each part of the paper and they can quickly locate specific information.

### Page Limit

A guideline of 40 pages (including title page, abstract, text, acknowledgments, references, appendixes, tables, and figures) is suggested as an upper limit for manuscript length for most manuscript types, but please note that this is just a general guideline. This page limit does not include supplemental materials. Longer manuscripts, particularly for critical reviews and extended data-based reports, will not be excluded from review, but the author(s) should be prepared to justify the length of the manuscript if requested to do so. Research Notes should be kept to 20 or fewer pages in length, inclusive of all pieces other than supplemental materials. There is not a prescribed limit on the number of tables or figures that can be included.

### **Title Page**

The title should be short and clear, yet provide a sufficient description of the work. As the title becomes the basis for most online searches, it should contain the key words describing the work presented. If your title is insufficient, people will have difficulty finding your article. The title page should also include a list of the authors and their affiliations (see Authorship Criteria and Guidelines for more information).

## **Abstract**

The abstract helps readers scan through lists of articles or search results and is essential for helping users decide whether to read the rest of the article or save it for future reference. As a result, abstracts must be brief but also informative enough to be genuinely useful. ASHA recommends that abstracts be 150–250 words. The size limit for what can be included in your submission is set above 300 words, but that is so that very detailed abstracts for specific types of studies can be accommodated (see, for example, the abstract for this [randomized controlled trial](#) reported according to the CONSORT framework). Regardless of the type of manuscript, abstracts must be structured using the following sections: Purpose: The Purpose section must include a concise statement of the specific purposes, questions addressed, and/or hypotheses tested. Lengthy descriptions of rationale are not necessary or desirable. Method: The Method section must describe characteristics and numbers of participants and provide information related to the design of the study (e.g., pre–post group study of treatment outcomes, randomized controlled trial, multiple baseline across behaviors; ethnographic study with qualitative analysis; prospective longitudinal study) and data collection methods. If the participants have been assigned randomly to study conditions, this must be noted explicitly, regardless of the design used. If the article is not data-based, information should be provided on the methods used to collect information (e.g., computerized database search), to summarize previously reported data and to organize the presentation and arguments (e.g., meta-analysis, narrative review). Results: The Results section should summarize findings as they apply directly to the stated purposes of the article. Statistical outcomes may be summarized, but no statistics other than effect sizes should be provided. This section may be omitted from articles that are not data-based. Conclusions: The Conclusions section must state specifically the extent to which the stated purposes of the article have been met. Comments on the generalizability of the results (i.e., external validity), needs for further research, and clinical implications often are highly desirable.

## **Introduction**

The introduction usually describes the theoretical background, indicates why the work is important, states a specific research question, and poses a specific hypothesis to be tested. This section should provide your statement of purpose and rationale.

## **Method**

The methods section must provide a clear and precise explanation of how you carried out the study and why specific experimental procedures were chosen. This section describes both the techniques and the overall experimental strategy used by the authors in order to address any questions the readers may have about the experimental design. The methods section must be written with enough information so that (1) the experiment could be repeated by others to determine if the results can be replicated and (2) the audience can judge the study's validity.

## **Results**

The results section contains the data collected during your study and is the heart of a scientific paper. The body of the results section is a text-based presentation of the key findings which includes references to each of the Tables and Figures. Much of the important information may be in the form of tables or graphs. The text should guide the reader through the results stressing the key results that provide the answers to the question(s) investigated.

### **Discussion**

The discussion section should explain what the results mean and how the results relate to other studies. This section interprets your findings, evaluates the hypotheses or research questions, discusses unexpected results, and ties the findings to the previous literature (discussed first in the Introduction). Any possible objections to the work and/or suggestions of areas for improvement in future research can be addressed in this section.

### **Acknowledgments**

Citation of grant or contract support of research must be given in an acknowledgments section at the end of the article (before the References). If any part of the research was supported by an institution not named on the title page, that institution should be acknowledged in this section. Individuals who assisted in the research may be acknowledged. Do not name individuals (editors and reviewers) who participated in the review process.

### **References**

All literature cited in the text, as well as test and assessment tools, ANSI and ISO standards, and specialized software, must be listed in this section. References should be listed alphabetically, then chronologically under each author. Journal names should be spelled out and italicized. Pay particular attention to accuracy and APA style for references cited in the text and listed in the references.

### **Tables and Figures**

Tables present lists of numbers or text in columns, each column having a title or label. Figures are visual presentations of results, including graphs, diagrams, photos, drawings, schematics, maps, etc. Each table or figure should appear on its own page (i.e., don't put more than one figure or table on the same page). Use arabic numerals to identify both tables and figures, and do not use suffix letters for complex tables. Instead, simplify complex tables by making two or more separate tables. Table titles and figure captions should be concise but explanatory. The reader should not have to refer to the text to decipher the information. Keep in mind the width of a column or page when designing tables and figures. In other words, consider whether legibility will be lost when reductions are made to fit a column or page width. Avoid "special effects" in figures (e.g., three-dimensional bar graphs) because they distort, rather than enhance, the data and distract the reader.

### **Optional Materials**

## Appendices

An appendix is an optional part of the paper that allows you to include detailed information that would be distracting in the main body of the article. Examples of items you might have in an appendix include lists of words, a questionnaire or tool used in the study, a detailed description of an apparatus used in the research, etc.

## Supplemental Material

Supplemental material is non-essential to understanding of the paper, but may present information that further enhances the article. Examples of the types of material and file formats accepted may be found on the Supplemental Material and Multimedia section of this page. Any files for supplemental materials should be submitted at the same time as the manuscript and will be subject to the normal peer review process. Please indicate clearly that the material is intended as supplementary, and be sure that it is referred to within the text of the manuscript. Also, please provide a concise (1- or 2-sentence) description for each file supplied.

## **Reference Style and Formatting**

The ASHA Journals use the *Publication Manual of the American Psychological Association* (6th ed.) for reference style and formatting. You can purchase your own copy of the APA Publication Manual directly through their website, or various other retailers. Please use the quick resources box to the right for various other online materials which may prove helpful when formatting your references.

## ***Supplemental Material and Multimedia***

### Supplemental Material

ASHA has partnered with Figshare to enable authors to automatically archive data and supporting materials in an open source, public repository when submitting an article to an ASHA journal. Figshare provides unlimited data storage for a variety of file formats. All content is assigned a permanent web link (DOI) so you and other authors can link directly to it from future papers.

You can easily upload supplemental files within the existing ScholarOne Manuscript submission workflow. This supplemental material can consist of any of the following:

- *text* (e.g., tables that are too lengthy for publication within the journal; equations and models; or program source code for presentation of experimental protocols or analysis of data)
- *images* (e.g., visual stimuli or alternative figures with data plotted on different time scales)
- *video* (e.g., instruction in classroom, or demonstrations of clinical or research protocols)
- *sound clips* (e.g., auditory stimuli)
- *data* (e.g., raw data for testing and evaluating models, or normative data)

If you plan to take advantage of this service, then your data and other supplemental material must be submitted before your article is accepted. If your article is accepted for publication, then all of your supplemental files are automatically deposited into ASHA journals' Figshare open source data

repository without charge. Figshare uses Creative Commons licenses for supplemental material hosted on their site. CC BY is the license used for most file types. CC0 is the standard license used for sharing data and databases. However, you can select another license to set access restrictions on your supplemental material if needed. Please review the explanation of [Creative Commons licenses](#) for more information.

### Multimedia

Authors who have video or animation files that they wish to submit with their article are encouraged to include links to these within the body of the article. This can be done in the same way as a figure or table by referring to the video or animation content and noting in the body text where it should be placed. All submitted files should be properly labeled so that they directly relate to the file's content. In order to ensure that your video, audio, or animation material is directly usable, please provide the files in one of our recommended file formats (see below). Video, audio, and animation files supplied will be published online in the electronic version of your article. Please supply "stills" with your files: you can choose any frame from the video or animation or create a separate image. For audio clips, you can select a thumbnail image that you feel is representative of the content of the audio clip. These will be used instead of standard icons and will personalize the link to your multimedia data. Because video, audio, and animation cannot be embedded in the print version of the journal, please provide text for both the electronic and the print version for the portions of the article that refer to this content. If the content being submitted is truly "supplementary" (not essential to the content of the article or only of supplementary interest to the reader), it can be included as Supplementary Content (i.e., accessible only electronically via an active link in the article). Multimedia files included as Supplementary Content should be referred to at an appropriate place in the text. If this is not done, any Supplementary Content will be referred to in an appendix without specifying exactly what it is. All submitted files should be properly labelled so that they directly relate to the file's content. This will ensure that the files are fully searchable by users. If the author does not hold copyright to the video, the author must obtain permission for the video to be published in the journal. This permission must be for unrestricted use in all print, online, and licensed versions of the journal. Please see our [guide on seeking permission](#) for more information.

### Acceptable File Formats

- Documents: plain text (.txt); HTML (.html, .htm); Microsoft Word, PowerPoint, or Excel (.docx, .pptx, .xlsx); PDF (.pdf); and XML (.xml)
- Images: JPEG (.jpg, .jpeg), PNG (.png), and TIF (.tif, .tiff)
- Audio: MPEG (.mp3) and WAV (.wav) files
- Video: MPEG (.mp4), Microsoft AVI Video (.avi), and Quicktime Video (.mov) files

If you are not sure what specifications to use to create a video, the following settings are recommended:

- Consider providing videos in .mp4 format to help ensure the broadest possible compatibility. If you will be providing a MOV or AVI video, we recommend using the H.264 codec.
- The aspect ratio should be consistent through the entire video, and should be either 16:9 or 4:3.
- The minimum resolution should meet or exceed 512×384 for 4:3 videos, or 512×288 for 16:9 videos.
- Consider providing a descriptive title or caption at the start of the video.
- If applicable, either provide a transcript of the video in an accompanying .txt file, or include open captions within the video file.

#### Image Recommendations

- Images should be no larger than 1800 pixels wide or high and should be easily viewed at 100% on screen in a standard browser.
- As .tif images are converted to PNG-24 format by our platform for display on article webpages, we recommend that .tif images be created using or converted to an RGB color space prior to submission.

Any files for supplemental materials should be submitted at the same time as the manuscript and will be subject to the normal peer review process. Please indicate clearly that the material is intended as supplementary, and be sure that it is referred to within the text of the manuscript. Also, please provide a concise (1- or 2-sentence) description for each file supplied. The material must be original content that has not been previously published. Where possible, the material will be copyedited. Please note: Recordings or images that involve identifiable participants require permission from those individuals. Please secure and provide that signed consent.

#### Consent for Publication

When publishing identifiable images, or video and audio recordings, from human research participants in ASHA journals, authors include a statement in the published paper affirming that they have obtained informed consent for publication of the images and/or recordings. The consent form should cover both image and voice of the person (if both are used); should specifically grant consent for submitting the recording for publication in a scientific journal; and should include online publishing if the recording and/or still images from it for purposes related to promotion of the published study.

All reasonable measures must be taken to protect patient anonymity. Black bars over the eyes are not acceptable means of anonymization. In certain cases, ASHA may insist upon obtaining evidence of informed consent from authors. Images and recordings without appropriate consent will be removed from publication.

**Disponível em:** <https://academy.pubs.asha.org/asha-journals-author-resource-center/manuscript-preparation/writing-and-formatting-your-manuscript/>