CoDAS
ISSN 2317-1782 (Online version)

Artigo Original
Original Article

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

Effect of synthesized voice anchors on auditory-perceptual voice evaluation

Priscila Campos Martins dos Santos[1]
Maurílio Nunes Vieira[2]
João Pedro Hallack Sansão[3]
Ana Cristina Côrtes Gama[1]

Descritores

Voz
Distúrbios da Voz
Qualidade da Voz
Disfonia
Percepção Auditiva
Treinamento da Voz

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 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.

Keywords

Voice
Voice Disorders
Voice Quality
Dysphonia
Auditory Perception
Voice Training

ABSTRACT

Purpose: To analyze if the use of synthesized voice anchor emissions in auditory-perceptual evaluation improves intra- and inter-rater agreement. **Methods**: This is a quantitative study. Thirty-two inexperienced evaluators were selected and performed two activities on a Programming Interface created by the authors: Active Calibrator Activity — auditory-perceptual evaluation of the roughness and breathiness parameters as 0–no deviation, 1–slight deviation, 2–moderate deviation, or 3–intense deviation of 25 voices with the support of anchored emissions of synthesized voices; and Inactive Calibrator Activity — auditory-perceptual evaluation of these same voices without the support of anchored vocal emissions. The voices were randomized for each activity, and the order of the activities was drawn randomly for each evaluator. The second activity was performed 15 days after the first. The Kappa coefficient was used to analyze intra- and inter-rater agreement, and the confidence interval (CI) was employed to compare concordances. **Results**: Inter-rater agreement was higher for the intense degree of the breathiness parameter in the Active Calibrator Activity when compared to the Inactive Calibrator Activity, as well as the intra-rater agreement of the roughness parameter. **Conclusion**: Use of anchor emissions of synthesized voices directly in the evaluation improves intra- and inter-rater agreement in auditory-perceptual voice analysis.

Endereço para correspondência:
Priscila Campos Martins dos Santos
Departamento de Fonoaudiologia da Faculdade de Medicina da Universidade Federal de Minas Gerais – UFMG
Av. Professor Alfredo Balena, 190, sala 249, Santa Efigênia. Belo Horizonte (MG), Brasil, CEP: 30130-100.
E-mail: priscila.fonoaudiologia@gmail.com

Recebido em: Agosto 13, 2019

Aceito em: Março 25, 2020

Trabalho realizado na Faculdade de Medicina, Universidade Federal de Minas Gerais – UFMG - Belo Horizonte (MG), Brasil

[1] Departamento de Fonoaudiologia, Faculdade de Medicina, Universidade Federal de Minas Gerais – UFMG - Belo Horizonte (MG), Brasil.

[2] Departamento de Engenharia Eletrônica, Escola de Engenharia, Universidade Federal de Minas Gerais – UFMG - Belo Horizonte (MG), Brasil.

[3] Departamento de Tecnologia em Engenharia Civil, Computação, Automação, Telemática e Humanidades, Universidade Federal de São João Del Rei – UFSJ - Ouro Branco (MG), Brasil.

Fonte de financiamento: Fundação de Amparo à Pesquisa do Estado de Minas Gerais – Fapemig (APQ-02594-15) e Conselho Nacional de Desenvolvimento Científico e Tecnológico-Brasil – CNPq (nº309108/2019-5).

Conflito de interesses: nada a declarar.

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[1]. 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[2-5]. Essa subjetividade, desvantagem da análise perceptivo-auditiva, gera alta variabilidade na concordância intra e interavaliadores, prejudicando a confiabilidade dessa avaliação[6-8].

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[8,9]. 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[10-12]. 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[7].

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[6,8,13]. Uma pesquisa realizada com avaliadores inexperientes[13], mostrou que o uso de emissões vocais âncoras no treinamento melhorou a concordância intra e interavaliadores na avaliação pós treinamento.

Ao compararem o uso de âncoras ao método de pareamento no treinamento de avaliadores experientes, pesquisadores observaram que os dois métodos facilitaram a avaliação perceptivo-auditiva, mostrando uma melhora significativa na precisão da avaliação após o treinamento[8]. 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 analisaram, em outro estudo[6], 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.

Avaliadores inexperientes apresentaram o mesmo grau de concordância intra e interavaliadores que os avaliadores experientes em estudo que utilizou estímulos âncoras sintetizados em dois tipos diferentes de treinamento: um graduando os estímulos vocais segundo a magnitude do desvio, da mais alterada para a menos alterada; e outro organizando os estímulos vocais em categorias segundo o grau de desvio[14].

Diante do exposto, as emissões vocais âncoras têm sido frequentemente associados ao treinamento perceptivo-auditivo para posterior análise de seu efeito na avaliação vocal[9,10]. No entanto, poucos estudos analisam o uso de emissões vocais âncoras diretamente na avaliação da voz[11,15]. 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[15], 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[8,16]. Portanto, o objetivo do presente estudo foi analisar se a utilização de emissões âncoras de vozes sintetizadas na avaliação perceptivo-auditiva melhora a concordância intra e interavaliador.

MÉTODO

A presente pesquisa foi aprovada pelo Comitê de Ética em Pesquisa (COEP) sob o parecer de número 920866. 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 32 avaliadores realizaram as atividades de avaliação perceptivo-auditiva de 30 emissões vocais.

Foram criadas pelas pesquisadoras duas atividades para avaliação perceptivo-auditiva e disponibilizadas em um aplicativo, construído pelas pesquisadoras para o desenvolvimento do presente estudo e disponibilizado apenas para os participantes do mesmo no momento da coleta. 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 soprosidade (B), sendo que considerou-se como rugosidade a qualidade vocal que apresentasse qualquer irregularidade perceptível durante a produção vocal, e como soprosidade a qualidade vocal com escape de ar audível durante a produção da voz[17].

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[18-20].

Cada atividade será descrita a seguir.

Atividade Calibrador Ativo

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

Processo

Durante essa atividade cada voz foi avaliada primeiramente segundo o parâmetro R e, em seguida, quanto ao parâmetro B. Para isso, os avaliadores foram orientados a realizar os seguintes procedimentos: 1. Escutar a voz natural a ser avaliada; 2. Escutar as emissões âncoras de vozes sintetizadas para cada grau do parâmetro R; 3. Novamente escutar a voz a ser avaliada; 4. Digitar no espaço em frente ao ícone "grau de rugosidade" o número correspondente ao grau de classificação da voz para o parâmetro R, sendo 0 – ausência de desvio, 1 – desvio leve, 2 – desvio moderado ou 3 – desvio intenso (Figura 2). Repetiram os mesmos procedimentos para classificar a mesma voz quanto ao parâmetro B.

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

Seleção das emissões vocais para avaliação

Para compor a amostra de vozes naturais a 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

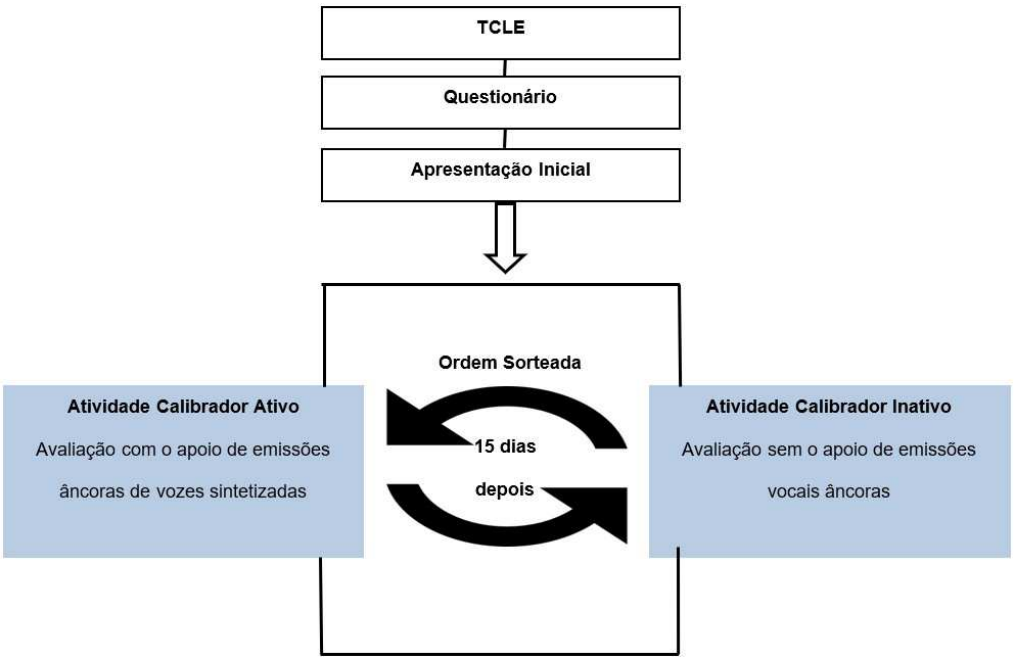

Figura 1. Fluxograma do Calibrador Auditivo

Figura 2. Atividade Calibrador Ativo do aplicativo

parâmetros recebeu quatro emissões vocais a fim de se alcançar o N previamente encontrado por meio de cálculo amostral, com o total de 25 vozes. Para definir o parâmetro e grau que receberia uma amostra a mais foi realizado um sorteio, sendo selecionado o grau leve do parâmetro soprosidade. Acrescentou-se 20% das vozes com o objetivo de analisar a concordância intra-avaliador, totalizando 30 emissões vocais. Os avaliadores não sabiam quantas emissões havia em cada grau, bem como não sabiam que havia vozes duplicadas.

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

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. A manipulação dessas medidas conferiu às vozes características de rugosidade ou soprosidade. 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[21].

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[22], conforme apresentado na Tabela 1.

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.

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

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

Teste estatístico: Curva de Roc

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

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

Média da marcação realizada por avaliadores em mm na Escala Visual Analógica relativo à naturalidade das vozes

Atividade Calibrador Inativo

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

Processo

Durante essa atividade cada voz também foi avaliada primeiramente segundo o parâmetro R e, em seguida, quanto ao parâmetro B. Novamente os avaliadores foram orientados a realizar os seguintes procedimentos: 1. Escutar a voz natural a ser avaliada; 2. Digitar no espaço em frente ao ícone "grau de rugosidade" o número correspondente ao grau de classificação da voz para o parâmetro R, sendo 0 – ausência de desvio, 1 – desvio leve, 2 – desvio moderado ou 3 – desvio intenso. Repetiram os mesmos procedimentos para classificar a mesma voz quanto ao parâmetro B.

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.

Para a coleta foram disponibilizados horários em laboratórios de informática localizados em diferentes prédios da instituição de ensino, a fim de facilitar a participação dos alunos dos períodos iniciais do curso de Fonoaudiologia como avaliadores, uma vez que estes realizam aulas em prédios diferentes e em período integral. Os avaliadores realizaram as tarefas fora do horário de aula, comparecendo aos laboratórios exclusivamente para realização das atividades da pesquisa. Foi realizado agendamento prévio com os participantes a fim de garantir que cada avaliador teria um computador à sua disposição, onde realizaria as atividades individualmente acessando o aplicativo pelo navegador Internet Explore. Um dos pesquisadores acompanhou os avaliadores fornecendo orientações prévias à realização das atividades, mas sem intervir em sua execução. Foi utilizado fone de ouvido supra-auricular modelo *Multilaser Vibe Headphone* estéreo durante todos os procedimentos. Os avaliadores podiam escutar as vozes quantas vezes julgassem necessário, desde que respeitassem a ordem dos procedimentos.

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.

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 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, ao observarmos os IC (Tabela 3) pode-se verificar uma tendência de aumento da concordância interavaliadores para os graus 0, 1 e 2 do parâmetro R e de diminuição da mesma para o grau 3 deste mesmo parâmetro na Atividade Calibrador Ativo – com emissões âncoras de vozes sintetizadas, 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).

Quanto à soprosidade, não houve diferença ao observarmos os IC (Tabela 4) dos graus 0, 1 e 2. Porém, também é possível verificar uma tendência a aumento da concordância interavaliadores 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 soprosidade mostrou-se estatisticamente maior na Atividade Calibrador Ativo quando comparada à Atividade

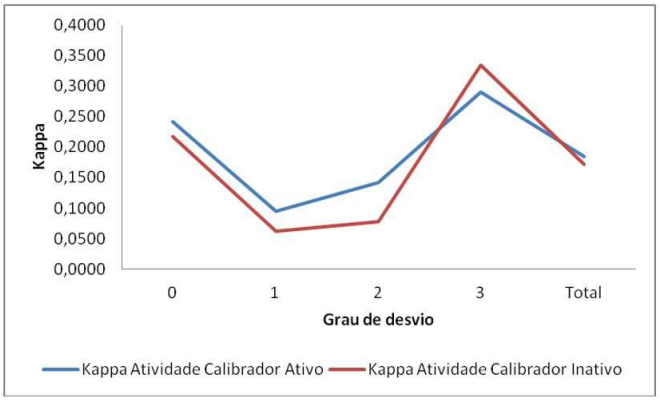

Figura 3. Comparação entre a concordância interavaliadores na Atividade Calibrador Ativo - com emissões âncoras de vozes sintetizadas, e Atividade Calibrador Inativo - sem emissões vocais âncoras, para cada grau de desvio quanto ao parâmetro Rugosidade, por meio do coeficiente Kappa ponderado

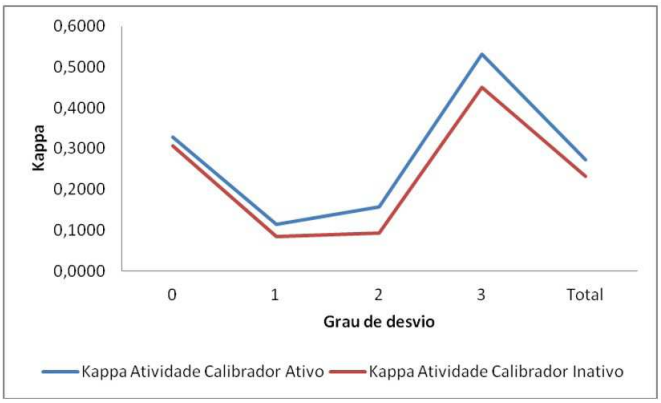

Figura 4. Comparação entre a concordância interavaliadores na Atividade Calibrador Ativo – com emissões âncoras de vozes sintetizadas, e Atividade Calibrador Inativo - sem emissões vocais âncoras, para cada grau de desvio quanto ao parâmetro Soprosidade, por meio do coeficiente Kappa ponderado

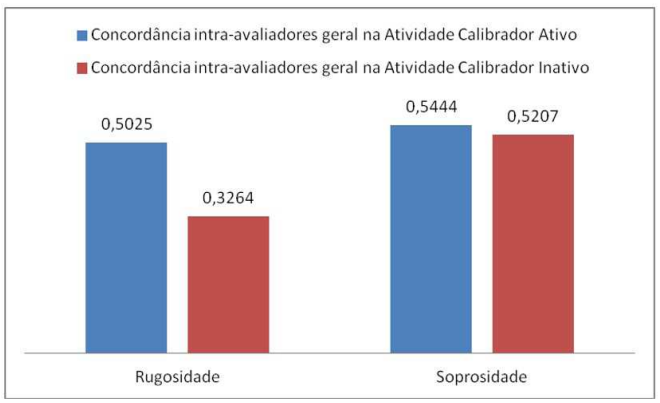

Figura 5. Comparação entre a concordância intra-avaliadores na Atividade Calibrador Ativo – com emissões âncoras de vozes sintetizadas, e Atividade Calibrador Inativo - sem emissões vocais âncoras, para os parâmetros Rugosidade e Soprosidade, por meio do coeficiente Kappa ponderado

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
Total	0,1846	0,1346	0,2346	0,1724	0,1216	0,2232

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

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
Total	0,2738	0,2293	0,3183	0,2313	0,1842	0,2784

Para análise estatística foi considerado o coeficiente Kappa ponderado e o intervalo de confiança (IC)

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 Soprosidade, por meio do coeficiente Kappa

	Atividade Calibrador Ativo			Atividade Calibrador Inativo		
	Kappa	IC		Kappa	IC	
Rugosidade	0,5025	0,4862	0,5188	0,3264	0,3105	0,3423
Soprosidade	0,5444	0,5284	0,5604	0,5207	0,5047	0,5367

Para análise estatística foi considerado o coeficiente ponderado Kappa e o intervalo de confiança (IC)

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).

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 (Tabela 5 e Figura 5).

DISCUSSÃO

No presente estudo optou-se pelo uso de vozes sintetizadas como âncoras. 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[4,7,9-12,23]. O uso de vozes sintetizadas permite a escuta de cada parâmetro vocal isoladamente durante a avaliação, facilitando a percepção dos mesmos[7]. Optou-se, ainda, 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 previamente realizada a classificação da naturalidade das vozes para cada um dos parâmetros, a fim de garantir que as vozes com maior naturalidade fossem selecionadas para o presente estudo. 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 soprosidade (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[24]. 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[24] 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.

Estudo em que emissões âncoras foram utilizadas diretamente na avaliação perceptivo auditiva da voz[11] selecionou 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[15] 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.

No presente estudo, a concordância interavaliadores para o parâmetro rugosidade apontou uma tendência a aumentar na Atividade Calibrador Ativo – com emissões âncoras de vozes sintetizadas, para os graus 0, 1 e 2 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 ao observarmos os IC. O resultado corrobora a literatura[15] que aponta uma concordância interavaliadores significativamente maior para rugosidade em análise realizada por avaliadores experientes 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. A literatura[25] aponta que quanto maior o grau de desvio vocal, maior a confiabilidade da avaliação. No entanto, no presente estudo o grau 3 do parâmetro R mostrou uma tendência a ser menor na Atividade Calibrador Ativo quanto comparado à Atividade Calibrador Inativo. Esse achado pode estar relacionado à complexidade do parâmetro R[19], que envolve diferentes qualidades vocais, como rouquidão, aspereza, crepitação e bitonalidade, o que pode ter favorecido a diferente percepção entre os avaliadores quanto ao parâmetro e contribuído para a redução da concordância entre eles.

Quanto à soprosidade, no presente estudo não houve diferença ao observarmos os IC (Tabela 4) dos graus 0, 1 e 2. Porém, também é possível verificar uma tendência ao aumento da concordância interavaliadores na Atividade Calibrador Ativo quando comparada à Atividade Calibrador Inativo. A concordância interavaliadores para o grau 3 de soprosidade mostrou-se estatisticamente maior na Atividade Calibrador Ativo. Não foram encontrados estudos na literatura em que foram utilizadas emissões âncoras de vozes sintetizadas diretamente na avaliação do parâmetro soprosidade. 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[13]. Ainda segundo a literatura[25] desvios vocais intensos favorecem uma maior concordância interavaliadores, o que corrobora este achado.

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[15], 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. Este achado mostra ainda que, apesar da discordância na percepção do parâmetro R entre os avaliadores,

o uso da âncora favorece a estabilização de padrões internos, aumentando a concordância intra-avaliadores.

No presente estudo houve também uma tendência ao aumento da concordância intra-avaliadores na Atividade Calibrador Ativo para o parâmetro soprosidade, apesar de não ser observada diferença. Estudo em que este mesmo parâmetro foi avaliado após o treinamento com emissão vocal âncora, verificou uma tendência ao aumento da concordância intra-avaliadores[13], embora também não tenha sido observada diferença. O uso de tarefa de fala encadeada associada à vogal sustentada poderia beneficiar na percepção deste parâmetro auxiliando no aumento da concordância intra-avaliadores, uma vez que, segundo a literatura[26] a soprosidade é mais facilmente identificada na fala encadeada que na vogal sustentada.

No presente estudo verificou-se, pela classificação do coeficiente Kappa[27], uma concordância interavaliador pequena para o parâmetro R e regular para o parâmetro B, sendo ainda observada uma concordância intra-avaliador moderada para os dois parâmetros. Ou seja, a concordância intra-avaliador foi maior que a concordância interavaliador para os dois parâmetros, achado que corrobora a literatura[26].

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[28]. 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 entanto, outros estudos mostram que a concordância na avaliação perceptivo-auditiva é maior para avaliadores experientes, devido ao padrão interno previamente desenvolvido. Estudo anterior[11] apontou que avaliadores experientes apresentaram menor variabilidade da concordância na avaliação com apoio da emissão vocal âncora. Em um segundo estudo[29], avaliadores experientes apresentaram melhor habilidade para classificar vozes humanas e sintetizadas. Outro estudo[28], apontou o impacto positivo da experiência dos avaliadores na concordância interavaliadores da análise perceptivo-auditiva da voz. Outra pesquisa[30] mostrou ainda que indivíduos experientes na análise perceptivo-auditiva da voz parecem apresentar mais facilidade em utilizar estratégias de aprendizagem para melhorar sua *performance* na avaliação vocal, mostrando que a experiência profissional influencia de modo positivo essa análise. Diante disso, ressalta-se a importância da realização de outros estudos com emissões âncoras de vozes sintetizadas na avaliação perceptivo-auditiva com avaliadores experientes.

Estudo[22] aponta que avaliadores podem ser mais críticos na avaliação de parâmetros isolados que na avaliação do grau geral da qualidade vocal. No entanto, é importante ressaltar que, pela maioria das escalas usadas na clínica e em pesquisas fonoaudiológicas na área da voz, é realizada, além da avaliação do grau geral da qualidade vocal, a avaliação dos parâmetros de forma isolada. Sendo assim, o uso de instrumentos que facilitem a percepção dos parâmetros isolados por meio de emissões âncoras, podem ser facilitadores no processo de aprendizagem durante a formação acadêmica em Fonoaudiologia, bem como podem auxiliar no aumento da concordância intra e interavaliadores, melhorando a confiabilidade desta avaliação.

Sugere-se o aprimoramento do uso de emissões âncoras na avaliação perceptivo-auditiva da voz a partir de ajustes em estudos futuros, como a utilização da tarefa de fala encadeada além da vogal sustentada, definição de parâmetros mais complexos, como a rugosidade, assim como a seleção de avaliadores experientes e sua aplicação a uma quantidade maior de participantes, a fim de favorecer o aumento da concordância para graus e parâmetros não observados no presente estudo.

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. No entanto, sugere-se que ajustes sejam realizados em estudos futuros a fim de aprimorar o uso de emissões âncoras e favorecer tanto o ensino quanto a prática clínica da avaliação perceptivo-auditiva da voz.

AGRADECIMENTOS

Ao apoio da Fundação de Amparo à Pesquisa do Estado de Minas Gerais – Fapemig (APQ-02594-15) e do Conselho Nacional de Desenvolvimento Científico e Tecnológico-Brasil – CNPq (nº309108/2019-5).

REFERÊNCIAS

1. Oates J. Auditory-perceptual evaluation of disordered vocal quality: pros, cons and future directions. Folia Phoniatr Logop. 2009;61(1):49-56. http://dx.doi.org/10.1159/000200768. PMid:19204393.
2. Behlau M. Voz: o livro do especialista. Vol. 1. Rio de Janeiro, RJ: Revinter; 2001.
3. Kreiman J, Gerratt BR, Ito M. When and why listeners disagree in voice quality assessment tasks. J Acoust Soc Am. 2007;122(4):2354-64. http://dx.doi.org/10.1121/1.2770547. PMid:17902870.
4. Solomon NP, Helou LB, Stojadinovic A. Clinical versus laboratory ratings of voice using the CAPE-V. J Voice. 2011;25(1):e7-14. http://dx.doi.org/10.1016/j.jvoice.2009.10.007. PMid:20430573.
5. Chaves CR, Campbell M, Côrtes Gama AC. The influence of native language on auditory-perceptual evaluation of vocal samples completed by brazilian and Canadian SLPs. J Voice. 2017;31(2):258.e1-5. http://dx.doi.org/10.1016/j.jvoice.2016.05.021. PMid:27427162.
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(1):111-26. http://dx.doi.org/10.1044/1092-4388(2002/009). PMid:14748643.
7. 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):1091-101. http://dx.doi.org/10.1121/1.1500753. PMid:12243157.
8. Chan KMK, Yiu EML. A comparison of two perceptual voice evaluation training programs for naive listeners. J Voice. 2006;20(2):229-41. http://dx.doi.org/10.1016/j.jvoice.2005.03.007. PMid:16139475.
9. dos Santos PCM, Vieira MN, Sansão JPH, Gama ACC. Effect of auditory-perceptual training with natural voice anchors on vocal quality evaluation. J Voice. 2017;33(2):220-5. http://dx.doi.org/10.1016/j.jvoice.2017.10.020. PMid:29331406.

10. Awan SN, Lawson LL. The effect of anchor modality on the reliability of vocal severity ratings. J Voice. 2009;23(3):341-52. http://dx.doi.org/10.1016/j.jvoice.2007.10.006. PMid:18346869.

11. Eadie TL, Kapsner-Smith M. The effect of listener experience and anchors on judgments of dysphonia. J Speech Lang Hear Res. 2011;54(2):430-47. http://dx.doi.org/10.1044/1092-4388(2010/09-0205). PMid:20884782.

12. Sofranko JL, Prosek RA. The effect of the levels and types of experience on judgment of synthesized voice quality. J Voice. 2014;28(1):24-35. http://dx.doi.org/10.1016/j.jvoice.2013.06.001. PMid:24119637.

13. Eadie TL, Baylor CR. The effect of perceptual training on inexperienced listeners' judgments of dysphonic voice. J Voice. 2006;20(4):527-44. http://dx.doi.org/10.1016/j.jvoice.2005.08.007. PMid:16324823.

14. Gurlekian JA, Torre HM, Vaccari ME. Comparison of two perceptual methods for the evaluation of vowel perturbation produced by jitter. J Voice. 2016;30(4):506.E1-8. http://dx.doi.org/10.1016/j.jvoice.2015.05.009. PMID: 26106070.

15. Gerratt BR, Kreiman J, Antonanzas-Barroso N, Berke GS. Comparing internal and external standards in voice quality judgments. J Speech Hear Res. 1993;36(1):14-20. http://dx.doi.org/10.1044/jshr.3601.14. PMid:8450655.

16. Goldstone RL. Perceptual learning. Annu Rev Psychol. 1998;49(1):585-612. http://dx.doi.org/10.1146/annurev.psych.49.1.585. PMid:9496632.

17. Hirano M. Clinical examination of voice. New York: Springer Verlag; 1981.

18. 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 Speech Lang Pathol. 2010;19(3):248-58. http://dx.doi.org/10.1044/1058-0360(2010/09-0012). PMid:20484704.

19. 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. http://dx.doi.org/10.1590/S2179-64912012000100005. PMid:22460368.

20. Brinca L, Batista AP, Tavares AI, Pinto PN, 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. http://dx.doi.org/10.1016/j.jvoice.2015.01.007. PMid:25795348.

21. Vieira MN, Sansão JPH, Yehia HC. Measurement of signal-to-noise ratio in dysphonic voices by image processing of spectrograms. Speech Communication. 2014;61-62:17-32. http://dx.doi.org/10.1016/j.specom.2014.04.001.

22. Baravieira PB, Brasolotto AG, Montagnoli AN, Silvério KCA, Yamasaki R, Behlau M. Análise perceptivo-auditiva de vozes rugosas e soprosas: correspondência entre a escala visual analógica e a escala numérica. CoDAS. 2016;28(2):163-7. http://dx.doi.org/10.1590/2317-1782/20162015098. PMid:27191880.

23. Kreiman J, Gerratt BR, Kempster GB, Erman A, Berke GS. Perceptual evaluation of voice quality: review, tutorial, and a framework for future research. J Speech Hear Res. 1993;36(1):21-40. http://dx.doi.org/10.1044/jshr.3601.21. PMid:8450660.

24. 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. http://dx.doi.org/10.1016/j.jvoice.2015.07.017. PMid:26337775.

25. Eadie T, Sroka A, Wright DR, Merati A. Does knowledge of medical diagnosis bias auditory-perceptual judgments of dysphonia? J Voice. 2011;25(4):420-9. http://dx.doi.org/10.1016/j.jvoice.2009.12.009. PMid:20347262.

26. Law T, Kim JH, Lee KY, Tang EC, Lam JH, van Hasselt AC, et al. Comparison of Rater's reliability on perceptual evaluation of different types of voice sample. J Voice. 2012;26(5):666.e13-21. http://dx.doi.org/10.1016/j.jvoice.2011.08.003. PMid:22243971.

27. Altman DG. Some common problems in medical research. In: Altman DG. Practical statistics for medical research. London: Chapman and Hall; 1991.

28. Oliveira SB, Gama ACC, Chaves AR. Interferência do tempo de experiência na concordância da análise perceptivo-auditiva de vozes. Distúrb Comun. 2016;28(3):415-22.

29. Englert M, Madazio G, Gielow I, Lucero J, Behlau M. Perceptual error analysis of human an synthesized voices. J Voice. 2016;31(4): 516.E5-18. https://doi.org/10.1016/j.jvoice.2016.12.015.

30. Englert M, Madazio G, Gielow I, Lucero J, Behlau M. Influência do fator de aprendizagem na análise perceptivo-auditiva. CoDAS. 2018;30(3):e20170107. http://dx.doi.org/10.1590/2317-1782/20182017107. PMid:29898037.

Contribuição dos autores

Os autores PCMS, ACCG, MNV e JPHS conceberam e planejaram o projeto, assim como analisaram e interpretaram os dados e revisaram criticamente o conteúdo do manuscrito.