

Artigos originais

Avaliação de qualidade vocal: abordagem metodológica para análise de dados perceptivos

*Vocal quality assessment: methodological approach for a perceptive data analysis*Maria Fabiana Bonfim de Lima Silva⁽¹⁾Sandra Madureira⁽²⁾Luiz Carlos Rusilo⁽³⁾Zuleica Camargo⁽²⁾

⁽¹⁾ Universidade Federal da Paraíba (UFPB), João Pessoa, Paraíba, Brasil.

⁽²⁾ Pontifícia Universidade Católica de São Paulo (PUCSP), São Paulo, SP, Brasil.

⁽³⁾ Instituto de Ciência e Tecnologia da Universidade Federal de Alfenas (UNIFAL), Alfenas, Minas Gerais, Brasil.

Trabalho realizado no Curso de Fonoaudiologia da Universidade Federal da Paraíba – UFPB – João Pessoa (PB), Brasil e no Laboratório Integrado de Análise Acústica e Cognição (LIAAC)- PUC-SP, São Paulo, Brasil.

Conflito de interesses: inexistente

Recebido em: 30/01/2017

Aceito em: 14/09/2017

Endereço para correspondência:

Maria Fabiana Bonfim de Lima Silva
Departamento de Fonoaudiologia, Centro de Ciências da Saúde, Universidade Federal da Paraíba - Cidade Universitária, Campus I, Castelo Branco
CEP: 58051-900 - João Pessoa, Paraíba, Brasil
E-mail: fbl_fono@yahoo.com.br

RESUMO

Objetivo: apresentar uma abordagem metodológica para interpretação de julgamentos perceptivos de qualidade vocal por um grupo de juízes que utilizou o roteiro *Vocal Profile Analysis Scheme*.

Métodos: estudo transversal realizado a partir de 90 amostras de fala de 25 professoras da rede pública com distúrbio de voz e/ou alteração laringea. Previamente ao julgamento perceptivo, foram realizadas três tarefas de percepção para a escolha das amostras que foram apresentadas a cinco juízes por meio do *script Experiment MFC 3.2 (software PRAAT)*. A seguir, foi aplicada uma sequência de testes com base em abordagens sucessivas do comportamento inter e intrajuízes. Os dados foram tratados por meio de análise estatística (testes de *Cochran* e *Snedecor*).

Resultados: com relação à análise do desempenho dos juízes foi possível definir aqueles que apresentaram melhores resultados em termos de confiabilidade e de proximidade de análises ao juiz mais experiente, com exclusão de um. Os resultados da análise de *cluster* também permitiram compor o perfil de qualidade vocal do grupo de falantes estudado.

Conclusões: a proposta de abordagem metodológica permitiu definir os juízes, cujos julgamentos pautaram-se por conhecimentos fonéticos, bem como traçar o perfil de qualidade vocal do grupo de amostras analisadas.

Descritores: Avaliação; Voz; Docentes; Distúrbios da Voz; Percepção da Fala

ABSTRACT

Purpose: to present a methodological approach for interpreting perceptual judgments of vocal quality by a group of evaluators who used the script *Vocal Profile Analysis Scheme*.

Methods: cross-sectional study based on 90 speech samples from 25 female teachers with voice disorders and/or laryngeal changes. Prior to the perceptual judgment, we performed three perceptual tasks to select samples which were then presented to five evaluators using the *Experiment script MFC 3.2 (software PRAAT)*. Next, we applied a sequence of tests based on successive approaches to inter- and intra-evaluator behavior. Data were treated by statistical analysis (*Cochran* and *Selenor* tests).

Results: with respect to the analysis of the evaluators' performance, it was possible to define those that presented the best results, in terms of reliability and proximity of analyses, compared to the most experienced evaluator, excluding one. The results of the cluster analysis also allowed designing a voice quality profile of the group of speakers studied.

Conclusions: the proposal of a methodological approach allowed defining evaluators whose judgments were based on phonetic knowledge, and drawing a vocal quality profile of the group of samples analyzed.

Keywords: Evaluation; Voice; Teachers; Voice Disorders; Speech Perception

INTRODUÇÃO

A avaliação perceptiva da voz é um dos procedimentos mais antigos e amplamente utilizados na avaliação e no diagnóstico de distúrbios da voz¹⁻³, sendo que a eficácia de seus resultados depende fortemente da experiência do avaliador^{1,4-6}. Apesar de figurar como padrão ouro de avaliação vocal, há menção constante a possíveis interferências oriundas da subjetividade dos juizes, da falta de confiabilidade dos julgamentos, da variedade de métodos de avaliação, das inconsistências dos instrumentos e da falta de padronização da terminologia utilizada^{1,4,7-10}.

A busca pela superação destas limitações apontadas pauta-se em diferentes estratégias, tais como apresentação de estímulos-âncora, treinamento e calibração dos juizes, repetição de estímulos, aplicação de *scripts* para aleatorização da ordem de apresentação de amostras de voz (enquanto programação aplicável a *softwares* de código aberto) e abordagem estatística das respostas dos julgamentos perceptivos^{6,11-13}. Nesta perspectiva, dentre os instrumentos de avaliação disponíveis para uso clínico, há poucos roteiros e escalas que se fundamentem em modelos teóricos como a teoria fonética, como é o caso do *Vocal Profile Analysis Scheme (VPAS)*¹⁴.

Vale salientar que o uso dos instrumentos das Ciências Fonéticas na Clínica Fonoaudiológica tem contribuído com a identificação detalhada das estruturas do aparelho fonador em casos de alterações de fala^{15,16} bem como pistas do controle da fala, no processo de aquisição da linguagem de crianças com e sem deficiência auditiva (DA), desde idades precoces¹⁷⁻¹⁹. Além disso, o uso desses instrumentos pode oferecer possibilidades de caracterização de sons das línguas e de variantes linguísticas²⁰.

O roteiro *VPAS*, e sua adaptação para o português brasileiro *VPAS-PB*²¹, detalha a ocorrência de diversos ajustes de qualidade vocal nos planos fonatório, articulatorio, e de tensão, bem como de elementos da dinâmica vocal (*pitch*, *loudness*, uso de pausas, taxa de elocução e suporte respiratório), sob a perspectiva da teoria fonética. A aplicação do roteiro *VPAS* resulta num perfil de qualidade vocal da amostra em questão. Como exemplo, pode-se ter uma amostra cujo perfil de qualidade vocal caracteriza-se pela combinação de ajustes de mandíbula fechada (em grau 1), de laringe elevada (em grau 2) e de hiperfunção laríngea (em grau 2).

Parte da complexidade referida pelos clínicos no contato inicial com o roteiro *VPAS* reside nos princípios

teóricos que o sustentam. Os princípios de compatibilidade e de interdependência referem-se à relação entre os ajustes de qualidade vocal: - o primeiro aborda as ações que são fisiologicamente compatíveis ou incompatíveis entre si; - o segundo, por sua vez, foca em ações que são fisiologicamente interdependentes. Um terceiro princípio, da susceptibilidade, refere-se à relação entre ajustes e segmentos (vocálicos e consonantais), ou seja, como os ajustes de qualidade vocal afetam os segmentos na cadeia de fala. Neste último princípio, um segmento (vogal e consoante) pode ser suscetível à interferência de um ajuste; isto é, reflete o grau de vulnerabilidade dos segmentos em relação aos ajustes, principalmente aqueles segmentos tidos como “chave” para a detecção dos eventos da qualidade vocal²². Dessa forma, quando os ajustes apresentam características não compartilhadas pelo segmento, este último torna-se mais susceptível à influência dos primeiros.

Outro aspecto a se considerar no campo da avaliação perceptiva da qualidade vocal refere-se à demanda por adoção de grupo de examinadores/juizes, especialmente para abordar a questão da subjetividade implícita aos testes de percepção. Neste ponto, a questão que se aplica ao universo das pesquisas que utilizam o roteiro *VPAS* refere-se à demanda por estabelecimento do perfil vocal, a partir de julgamentos de vários juizes. Ou seja, o resultado final da avaliação de cada amostra vocal deve ser considerado à luz das várias apreciações realizadas pelos juizes de forma individualizada, levando à definição do(s) ajuste(s) de qualidade vocal (e de sua graduação de manifestação) que configuram o perfil de qualidade vocal do ponto de vista fonético.

A escassez de pesquisas que fundamentem um método de análise dos julgamentos perceptivos da qualidade vocal baseado em modelo fonético e em procedimentos de tratamento estatístico, que permita estimar os julgamentos mais próximos entre os juizes (em pares) bem como verificar o juiz com maior confiabilidade na resposta a instrumento de análise que apresenta uma escala de várias dimensões, justifica o interesse por este trabalho. Além disso, a abordagem a discussão dos princípios, dos procedimentos, e, especialmente, das possibilidades e das limitações da aplicação da análise perceptiva em rotinas de atendimento clínico e de ambientes de pesquisa estimulam um terreno fértil que visa promover reflexões a respeito da natureza e da fundamentação teórica dos protocolos de avaliação e de descrição perceptiva de voz

utilizada no contexto científico e na clínica, bem como sua relação ao histórico vocal do falante.

Deve-se considerar também que, em várias pesquisas, a descrição de dados perceptivos de qualidade vocal é uma etapa de análise e que, muitas vezes, tais achados serão correspondidos a dados acústicos e/ou fisiológicos. Dessa forma, a abordagem que permita que os dados de julgamentos perceptivos resultem em informações passíveis de análise estatística é uma demanda atual.

O presente estudo tem como objetivo apresentar uma abordagem metodológica para a interpretação de julgamentos perceptivos de qualidade vocal por um grupo de juízes que utilizou o roteiro *Vocal Profile Analysis Scheme- VPAS*, adaptado para o português brasileiro²¹.

MÉTODOS

Esta pesquisa foi aprovada pelo Comitê de Ética em Pesquisa com Seres Humanos da Universidade Federal da Paraíba - UFPB, sob o processo de número 298/2008. O *corpus* da pesquisa foi composto por amostras de um banco de dados com 54 vozes de professores. O professor foi eleito para participar da presente pesquisa uma vez que é o profissional da voz que mais apresenta distúrbios da voz e que mais procura atendimento fonoaudiológico³. Os critérios de inclusão foram: ser professor, do sexo feminino, com distúrbio de voz e alteração laríngea (por meio das informações perceptivas e do diagnóstico otorrinolaringológico) e que tiveram as amostras gravadas em três estilos de fala. Assim, com base nesses critérios restaram 33 professores.

Os registros em áudio foram realizados no ambiente de trabalho dos professores, em intervalos de aulas, e constaram das seguintes tarefas: fala semiespontânea (situação de entrevista), fala semiespontânea (simulação de aula) e fala lida²³. A opção por diferentes estilos de fala (tarefas) reside em variações já referidas nos ajustes de qualidade vocal e nos aspectos de dinâmica vocal²⁴.

A tarefa de leitura da presente pesquisa contemplou um trecho de texto padronizado²³. Além disso, foi coletado um trecho de fala semiespontânea-entrevista (FSE) a partir da primeira pergunta da entrevista: Que fatores você acha que interferem na voz? Por quê? conforme proposta dos autores². Finalmente, a simulação de aula consistiu na emissão de um trecho de aula com temática de escolha do professor (sem

tempo determinado na resposta), a partir da solicitação do examinador.

As amostras de fala foram registradas em sala silenciosa, com uso de microfone *headset Plantronics GameCom PRO 1* a uma distância de aproximadamente 15 cm da comissura labial direita, acoplado a notebook HP *Pavillion ZE 4920 CEL M330 1.4G* por meio do *software SoundForge 7.0*, na frequência de amostragem 22050 Hz, 16 bits, extensão "wav".

Todas as 99 amostras de 33 professoras foram submetidas a três tarefas de percepção, realizadas por distintos grupos de juízes. A partir dos resultados destas tarefas foi selecionado um conjunto de amostras que passou a figurar como *corpus* da presente pesquisa, detalhado na sequência do texto. Os objetivos e métodos, bem como os resultados de cada uma destas tarefas, são sintetizados na Figura 1.

Após a consideração dos resultados das três tarefas de percepção, foram excluídas amostras de oito professores. Assim, o *corpus* do experimento de percepção de qualidade vocal constou de 90 amostras de fala, originadas de 25 professores: 25 fala lida, 25 fala semiespontânea – entrevista (FSE), 25 semiespontânea – simulação de aula (FSA) e 15 repetições de algumas destas amostras utilizadas para abordagem da fidedignidade das respostas dos juízes, num total de 20% de repetição aleatorizada de amostras do *corpus*²⁵. Estas amostras foram editadas em trechos com aproximadamente de 20 segundos de duração, extraídas das gravações das 03 tarefas de fala.

Dessa forma, as 90 amostras foram etiquetadas enquanto enunciados referentes à fala lida (FL), semiespontânea – entrevista (FSE), semiespontânea - simulação de aula (FSA) e analisadas do ponto de vista perceptivo-auditivo (roteiro *VPAS-PB*) no *software PRAAT*, com auxílio do *script Experiment MFC 3.2*, versão 5143 (disponível em: <http://www.fon.hum.uva.nl/praat/>).

O *script Experiment MFC 3.2* foi utilizado como uma ferramenta para aleatorizar os estímulos a serem apresentados a cada um dos cinco juízes (J1 a J5) para a tarefa de julgamento da qualidade vocal com motivação fonética. Na primeira tela do experimento de percepção, foi apresentada a instrução do teste. Nas outras janelas, acionadas pelo juiz, foram apresentados os 90 estímulos sonoros.

A duração do experimento correspondeu a aproximadamente quatro horas por juiz, distribuídas em quatro sessões, em dias distintos, com duração de uma hora cada, e intervalos (pausas) de cinco minutos,

Tarefas	Objetivo	Método	Resultados
TP1	Investigar a influência da natureza das amostras de fala (em tarefas de FL, FSE e FSA) nos julgamentos pelo VPAS-PB.	Seleção de 08 amostras de fala (04 de FL e 04 de FSE) de 04 professoras, todas com queixa vocal, sendo que 02 apresentavam diagnóstico otorrinolaringológico de alteração laríngea. Tais amostras foram editadas, aleatorizadas e submetidas a 03 juízes experientes no uso do roteiro VPAS-PB.	A manifestação dos ajustes de qualidade vocal e dos elementos de dinâmica vocal no roteiro VPAS-PB variou em função da tarefa de fala. Impacto no planejamento da pesquisa: inclusão dos 03 estilos de fala (FL, FSE e FSA).
TP2	Definir a duração das amostras a serem utilizadas no experimento.	Seleção de 04 amostras de fala de 02 professoras com alteração laríngea em 02 tarefas de fala, as quais foram editadas em durações diferenciadas (20s, 30s e 40s). Participaram deste experimento 10 juízes que foram solicitados a julgar a qualidade vocal (roteiro VPAS-PB) em formulário específico.	A duração das amostras não teve impacto nos julgamentos perceptivos de qualidade vocal. Impacto no planejamento da pesquisa: duração das amostras de fala fixada em 20s.
TP3	Selecionar as gravações com melhor qualidade de áudio por meio da medida de relação sinal-ruído (SNR).	Seleção de 36 amostras: em 24 foram adicionados 02 níveis de ruído (no ponto médio e máximo da onda sonora). As amostras foram analisadas por 08 juízes com níveis de experiência variados no uso do roteiro VPAS-PB e por 02 juízes experientes, considerados como referência.	Apontou compatibilidade das respostas dos juízes de vários graus de experiência, com leve aumento de número de erros, proporcional ao aumento do nível de ruído na amostra. Impacto no planejamento da pesquisa: exclusão de estímulos com SNR < 2

Legenda: FL = fala lida, FSE = fala semiespontânea entrevista, FSA = fala simulação de aula, TP1 = tarefa de percepção 1, TP2 = tarefa de percepção 2 e a TP3 = tarefa de percepção 3.

Figura 1. Tarefas de percepção efetuadas na etapa de planejamento dos procedimentos de seleção de amostras para a análise perceptiva

após apresentação de conjunto de dez amostras, para descanso auditivo.

A seleção do grupo de juízes pautou-se na formação em Fonética e na experiência na aplicação do roteiro

VPAS-PB. Optou-se por reunir juízes com graus de experiência e de formação diferenciados, a fim de se discutir as interferências dessas variáveis no domínio da avaliação de qualidade vocal, conforme a Figura 2.

Juiz	Tempo de formação profissional (anos)	Formação	Tempo de uso do roteiro VPAS-PB (meses)	Tempo de formação no roteiro VPAS-PB (meses)	Participação no Workshop – roteiro VPAS-PB (anterior à aplicação do experimento)
J1	4	FONO com mestrado em Fonoaudiologia	18	18	Sim
J2	10	FONO com mestrado em Linguística	36	18	Sim
J3	24	FONO com mestrado em Linguística	18	18	Sim
J4	23	FONO com doutorado em Linguística	156	156	Sim
J5	28	FONO doutorando em Linguística	6	6	Não

Legenda: FONO = Fonoaudiólogo

Figura 2. Caracterização dos juízes participantes do experimento de percepção da qualidade vocal com motivação fonética

Para uniformização de procedimentos e de respostas a eventuais dúvidas, os cinco juízes foram convidados para participação no *Workshop* intitulado “Roteiro VPAS-PB: treinamento auditivo”, com duração de 15 horas. O Juiz 5 não participou desse *Workshop* e utiliza rotineiramente a avaliação perceptiva da voz em sua prática clínica, porém com base em outros instrumentos de avaliação.

A abordagem inicial dos julgamentos de J1 a J5 pautou-se na concordância e na confiabilidade inter- e intrajuízes no uso roteiro VPAS-PB, e em testes seriados, nos quais, gradativamente, foram estabelecidos escores e classificação entre os juízes, de forma que aqueles mais discrepantes em suas análises foram eliminados e os testes reaplicados no grupo restante. Para tanto, foram utilizados os testes de *Cochran* para homogeneidade de variâncias e *Snedecor*, com níveis de confiança em 95%, (*software Excel*) para definição da confiabilidade intra- e interjuízes (inclusive aos pares), verificando-se que todos os juízes, exceto um, apresentavam congruência significativa entre si. Deste modo, os julgamentos realizados pelo juiz incongruente foram excluídos em procedimento cego (ou seja, o analista dos dados não conhecia os juízes, tendo acesso apenas à planilha de julgamentos). Na sequência, os juízes foram classificados em função de seus escores de confiabilidade intrajulgamentos e de sua congruência aos demais examinadores.

Além disso, na etapa final, após avaliação de todos os juízes, os julgamentos foram comparados ao juiz adotado como referência, a partir da classificação gerada pela presente análise.

Os critérios utilizados para seleção do juiz referência foram: formação específica, tempo de uso e de formação do roteiro VPAS, participação do *workshop* sobre o uso deste roteiro bem como congruência e confiabilidade inter e intrajulgamentos.

Determinados os parâmetros estatísticos sobre os valores dos julgamentos, foi traçado o perfil de qualidade vocal de cada amostra do banco de dados. O perfil foi estabelecido a partir dos valores médios dos dados das análises dos juízes congruentes entre si, determinando-se um julgamento composto por valores esperados, a partir da análise estatística univariada (cômputo de intervalos de confiança).

Os valores de distâncias e produtos relativos de julgamentos interjuízes (em pares) para os resultados das avaliações perceptivas permitiu estimar os julgamentos mais próximos entre os juízes em pares.

A distância relativa é uma medida de dissimilaridade relativa entre os juízes, definida pela diferença de postos entre os juízes, ou seja, a diferença entre o *ranking* relativo de cada um dos juízes comparados dois a dois, sendo que o *ranking*, por sua vez, é definido como a colocação de cada juiz segundo sua média de acertos de julgamento. Uma vez que os juízes foram ranqueados de 1 a 5, as distâncias relativas podem assumir os valores 1 a 4, sendo que maiores valores indicam maior dissimilaridade entre os juízes comparados.

O produto relativo é um índice composto, construído pela multiplicação dos valores de *ranking* dos respectivos juízes em comparação dois a dois, pela distância relativa entre ambos, como definida no parágrafo anterior, indicando a qualidade de ambos bem como sua proximidade relativa. Quanto menor o valor desse índice, melhor a composição formada pelos dois juízes comparados.

A comparação dos dados de julgamentos intra-juízes com base nos valores de congruência (índices de acertos) permitiu estimar o juiz mais experiente e congruente nas análises.

RESULTADOS

Com relação à análise do comportamento inicial dos juízes (teste *Cochran*- análise interjuízes e teste *Snedecor* – análise interjuízes e intrajuízes), J5 foi excluído da etapa seguinte de reaplicação do teste, justamente por apresentar a maior variância em análise intra e interjuízes, menor tempo de uso e formação do roteiro VPAS bem como não participou do treinamento sobre o uso deste roteiro. Este teste foi reaplicado até que se atingiram as duas menores variâncias do grupo, que representaram os dois juízes com maior confiabilidade em termos de julgamento por meio do VPAS-PB: J2 e J4.

Os índices de acertos e intervalos de confiança (limites inferior e superior), com base na análise intra-juízes, são apresentados nas Tabelas 1 e 2. Os acertos foram considerados em função da abrangência dos resultados nos julgamentos perceptivos.

Os índices de acertos e intervalos de confiança (limites superior e inferior), com base na análise inter-juízes, são apresentados nas Tabelas 3 e 4. Os acertos foram considerados em função da abrangência de resultados nos julgamentos perceptivos.

Tabela 1. Valores de congruência (% acerto) de julgamentos intrajuízes (J1 a J5) para os resultados totais de julgamentos

	Limite Inferior	Acerto	Limite Superior
J1	90,93	92,72	94,51
J2	92,59	94,20	95,81
J3	90,93	92,72	94,51
J4	94,14	95,56	96,97
J5	87,40	89,51	91,62

Legenda: Resultados totais - todos os ajustes no roteiro VPAS-PB (graus 0 a 6)

Tabela 2. Valores de congruência (% acerto) de julgamentos intrajuízes (J1 a J5) para os resultados não nulos de julgamentos

	Limite Inferior	Acerto	Limite Superior
J1	48,64	56,93	65,23
J2	43,22	53,00	62,78
J3	36,52	45,87	55,23
J4	58,86	67,57	76,28
J5	57,69	63,83	69,97

Legenda: Resultados não nulos - os ajustes julgados no roteiro VPAS-PB (graus 1 a 6)

Tabela 3. Valores de congruência (% acerto) de julgamentos interjuízes para os resultados totais de julgamentos

	Limite Inferior	Acerto	Limite Superior
J1-J2	81,20	82,37	83,54
J1-J3	80,29	81,48	82,68
J1-J4	79,22	80,44	81,67
J1-J5	68,31	69,73	71,14
J2-J3	84,07	85,16	86,26
J2-J4	82,79	83,93	85,06
J2-J5	70,82	72,20	73,58
J3-J4	82,26	83,41	84,55
J3-J5	69,79	71,19	72,58
J4-J5	70,57	71,95	73,33

Legenda: Resultados totais - todos os ajustes no roteiro VPAS-PB (graus 0 a 6)

Tabela 4. Valores de congruência (% acerto) de julgamentos interjuízes para os resultados não nulos de julgamentos

Pares de juízes	Limite Inferior	Acerto	Limite Superior
J1-J2	8,09	10,19	12,29
J1-J3	7,63	9,64	11,65
J1-J4	8,49	10,51	12,53
J1-J5	8,58	10,18	11,79
J2-J3	8,00	10,30	12,60
J2-J4	7,43	9,58	11,73
J2-J5	8,06	9,70	11,35
J3-J4	7,44	9,56	11,67
J3-J5	6,54	8,04	9,53
J4-J5	9,32	11,04	12,76

Legenda: Resultados não nulos - os ajustes julgados no roteiro VPAS-PB (graus 1 a 6)

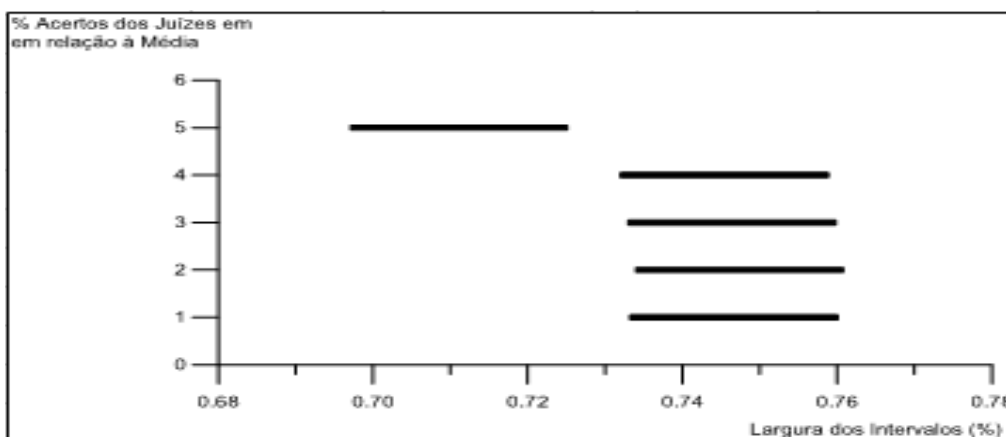
Tabela 5. Distâncias e produtos relativos dos julgamentos interjuízes

Pares de juízes	Distâncias relativas	Produtos relativos
J1-J2	1	6
J1-J3	1	12
J1-J4	2	6
J1-J5	2	30
J2-J3	2	16
J2-J4	1	2
J2-J5	3	30
J3-J4	3	12
J3-J5	1	20
J4-J5	4	20

A Tabela 5 apresenta as distâncias e produtos relativos de julgamentos interjuízes (em pares) para os resultados de julgamentos perceptivos.

A abordagem de distâncias e de produtos relativos permitiu estimar que os julgamentos foram mais próximos entre os juízes J2-J4, J1-J2, J1-J4 e J3-J5. Da comparação dos dados de julgamentos intrajuízes, destaca-se que J4 apresentou o maior índice de acertos em resultados totais, seguido de J2, e o par J1-J3. Por sua vez, J4 foi considerado o juiz mais experiente e congruente nas análises. Vale ressaltar que o grupo de juízes não revelou um comportamento absolutamente similar em todas as etapas, quando considerada

a abordagem interjuízes. Entretanto, foram consistentes em seus julgamentos para amostras repetidas na análise intrajuízes, de maneira que suas contribuições aos julgamentos puderam ser consideradas para composição do perfil de qualidade vocal. Neste aspecto, os resultados das respostas para a repetição de estímulos revelaram homogeneidade entre os quatro juízes (J1 a J4) e segregação do J5 (Figura 3). Diante da bateria de dados apresentados, os juízes foram qualificados de acordo com sua experiência no *VPAS-PB* na seguinte ordem decrescente: J4, J2, J1, J3 e J5.

**Figura 3.** Distribuição dos intervalos de confiança nos julgamentos com base nas médias de cada juiz (J1 a J5)

O perfil de qualidade vocal e de elementos de dinâmica vocal traçado para este estudo contemplou o conjunto de análises de quatro juízes (a partir das abordagens inter- e intrajuízes) que atingiram nível de distribuição de variâncias e intervalos de confiança dos julgamentos realizados similares entre si.

Destacaram-se, na composição de perfil de qualidade vocal do grupo estudado, em ordem decrescente de ocorrência: hiperfunção laríngea, voz áspera, laringe elevada, hiperfunção do trato vocal, mandíbula fechada, constrição faríngea, corpo de língua elevado e soproidade. Quanto aos aspectos da dinâmica

vocal, em ordem decrescente, destacaram-se: suporte respiratório inadequado, diminuição da variabilidade de *pitch*, *pitch* habitual elevado, *loudness* habitual elevado, taxa de elocução rápida e variabilidade de *loudness* aumentado.

DISCUSSÃO

Na prática fonoaudiológica da área de voz, a avaliação perceptiva da qualidade vocal é considerada padrão ouro²⁶. Ainda que alguns pesquisadores classifiquem-no como um método subjetivo, inconstante, com grande variabilidade terminológica, destaca-se que a avaliação perceptiva depende da formação e da experiência do avaliador, assim como de sua condição de atenção ao longo do procedimento²⁷⁻²⁹.

São escassos os estudos que apresentam uma abordagem metodológica para análise dos julgamentos perceptivos da qualidade vocal, fundamentado em modelo fonético¹⁴, bem como em procedimentos de tratamento estatístico para consideração de julgamentos de vários juízes. A presente pesquisa buscou apresentar uma abordagem metodológica para desenvolver um experimento de avaliação perceptiva de qualidade vocal em amostras de professores com distúrbio de voz e/ou alteração laríngea, por meio do roteiro *Vocal Profile Analysis Scheme* (VPAS-PB) e para avaliar o desempenho de um grupo de juízes. A opção por compor um grupo de juízes com experiências variadas em termos de tempo de exposição ao roteiro, do histórico de formação e da participação em treinamento prévio à aplicação da tarefa de percepção de qualidade vocal nesta pesquisa teve como objetivo justamente discutir as demandas de formação de juízes para o método em questão.

A questão da experiência de juízes tem sido amplamente debatida na literatura, especialmente quanto à possível subjetividade das análises³⁰. Neste estudo, foi possível resgatar o tempo de formação e a atuação de juízes, de acordo com a classificação estabelecida em termos da sua experiência pelos dados de análise estatística dos resultados dos julgamentos dos juízes com o uso do roteiro VPAS-PB. O grau de experiência esteve relacionado ao tempo de vivência (de uso) do VPAS-PB, em ordem decrescente: treze anos (J4); três anos (J2); um ano e seis meses (J1 e J3); seis meses (J5).

Na amostra de cinco juízes pode-se identificar fatores importantes na definição da experiência dos juízes: o tempo de convivência com o instrumento VPAS-PB, formação específica em abordagem fonética

da qualidade vocal, participação do *workshop* sobre o uso deste roteiro bem como congruência e confiabilidade inter e intrajulgamentos. Os procedimentos estatísticos adotados, desconhecidas as características subjetivas dos sujeitos (juízes) quando de sua aplicação, permitiram estabelecer uma escala de experiência dos juízes que foram congruentes com os aspectos da formação e do tempo de atividade em avaliação fonética da qualidade vocal (VPAS-PB) assim como a participação dos juízes no treinamento com o uso do VPAS-PB. Vale ressaltar que a aplicação do roteiro com os juízes deu-se depois deste treinamento (*Workshop*), buscando-se atingir um nível de calibração com auxílio de estímulos-âncora, procedimento esse também defendido por autores que exploram a complexidade dos experimentos de percepção de qualidade vocal¹². Dessa forma, o Juiz 5 foi excluído uma vez que tinha o menor tempo de formação e uso do roteiro VPAS-PB, não participou do *Workshop* bem como apresentou maior variância em análise intra e interjuízes, sendo considerado o juiz incongruente.

A opção pelo método de interpretação dos achados de percepção do grupo inicial de cinco juízes também se apresentou como um desafio, especialmente no que tange a complexidade de considerar a singularidade de análises individuais na busca por um “consenso” ou por uma análise que reflita a opinião de um grupo. A fim de se discutirem as particularidades e, especialmente, a complexidade de análises de base perceptiva por grupos de juízes, optou-se por não trabalhar com análises de consenso ou de avaliação de fidedignidade de respostas de juízes, que levassem à opção por um dos juízes. Diante da demanda de discussão das vantagens e desvantagens da adoção de modelo fonético de descrição da qualidade vocal¹⁴, decidiu-se estudar de forma mais detalhada o conjunto dos julgamentos perceptivos dos cinco juízes iniciais, até que fosse possível definir uma bateria de testes que permitisse a definição do perfil de qualidade vocal do grupo de amostras de vozes analisadas.

Após exploração global dos julgamentos e a análise do comportamento geral dos juízes, foi possível desenvolver a sequência de testes, que resultou na definição dos juízes, cujos julgamentos pautaram-se nos princípios do modelo fonético.

Vale salientar que esta avaliação não teve como intuito qualificar juízes em termos de suas habilidades perceptivas, mas de qualificar e estimar o seu desempenho em termos da tarefa proposta, considerando-se sua consistência de respostas para apresentações dos

mesmos estímulos em momentos distintos da análise, o que caracterizou a análise intrajuízes.

Outro ponto de destaque refere-se ao fato de que os quatro juízes cujas análises compuseram o perfil da média de julgamentos de qualidade vocal do conjunto de amostras estudadas não tiveram um comportamento similar quando analisados numa abordagem interjuízes; porém, foram consistentes em seus julgamentos para amostras repetidas. Desta forma, apesar de o grupo não ser absolutamente homogêneo, seus julgamentos mostram-se consistentes em momentos distintos. Esses achados foram semelhantes aos encontrados em outra pesquisa envolvendo grupo de avaliadores estudantes de Fonoaudiologia, no qual houve uma concordância das respostas intra-avaliadores em relação à avaliação das juízas (fonoaudiólogas com especialização em voz)⁶.

As informações de análises intrajuízes também foram interessantes para trazer à tona a comparação de um grupo de juízes aos julgamentos de um juiz com maior experiência no instrumento, que poderia ser considerado como juiz-referência, procedimento utilizado em outras etapas de estudos com base em julgamentos do VPAS-PB de dois examinadores experientes neste roteiro¹¹. Em um destes estudos, autores¹¹ relataram resultados relevantes decorrentes do treinamento de um grupo de 16 juízes (14 alunos do primeiro ano do curso de especialização em voz e de dois juízes restantes eram professores do curso de Fonoaudiologia e Linguística com experiência no uso do VPAS) ao investigar a validade e o consenso entre os examinadores do uso do VPAS. Um dos pontos destacados foi à falta de consenso entre os juízes participantes quanto ao grupo de ajustes fonatórios e suas possíveis combinações, o que segundo os autores¹¹ revelou a falta de sistematização dos métodos de avaliação vocal com base auditiva e de familiaridade com o referido modelo. Tais dados foram compatíveis com a presente pesquisa uma vez que na análise interjuízes encontrou-se algumas discrepâncias nos julgamentos, o que reforçam que a ampliação do período de formação, ou, mais precisamente, a atualização constante e o trabalho contínuo junto aos juízes, torna-se fundamental para se criar um grupo coesamente qualificado para conduzir análises fonéticas da qualidade vocal.

Os cuidados nos procedimentos dessa pesquisa para análise perceptiva da qualidade vocal, quanto à formação e vivência no roteiro VPAS do grupo de juízes e a comparação dos julgamentos deste grupo

aos julgamentos de um juiz com maior experiência, trazem à tona a complexidade inerente aos estudos que enfocam respostas de juízes, em variadas modalidades de percepção. No tocante à percepção de qualidade vocal, destaca-se a crítica à forma como a análise estatística vem sendo conduzida em muitos estudos, em que algumas correlações assinaladas podem ser efeitos de artefatos dos testes e de particularidades das amostras¹². Na presente investigação, as várias etapas de estudo do comportamento dos juízes e de abordagens sucessivas em termos intrínsecos (abordagem intrajuiz - consistência entre repetições de tarefa) e extrínsecos (abordagem intrajuízes - em relação aos demais juízes, ou mais precisamente, a cada um dos demais juízes) foram contempladas. Esta nova proposta de abordagem estatística apresentada no presente estudo pode tornar-se uma contribuição para a continuidade da exploração dos estudos de percepção auditiva da qualidade vocal³¹.

Ressalta-se que a análise da concordância interjuízes e intrajuízes é um fator fundamental para conferir confiabilidade da avaliação perceptiva da voz³². Tal concordância pode aumentar conforme a experiência e a formação em análises de alterações vocais, e é influenciada por fatores como fadiga, lapsos de atenção e equívocos durante a avaliação^{27,33,34}, além da própria concepção e estruturação do experimento de percepção.

Neste ponto da discussão, pode-se afirmar que os dados coletados reforçam que o tempo de formação no método é fundamental. Os dados de abordagem interjuízes, em que algumas discrepâncias ocorreram, reforçam que a ampliação do período de formação, ou, mais precisamente, a atualização constante e o trabalho contínuo junto aos juízes torna-se fundamental para se criar um grupo coesamente qualificado para conduzir análises fonéticas da qualidade vocal, a qual, ainda que possa ser considerada subjetiva, ou seja, sem um padrão objetivo e extrínseco, pode ser replicada com o treino. Outro ponto a se indicar para delineamento de futuros estudos no tema refere-se à ampliação do grupo de juízes.

Os achados encontrados reforçam o caráter multidimensional da qualidade vocal e a complexidade que envolve os julgamentos perceptivos deste fenômeno, bem como a demanda por aplicação de treinamento e experimentos de percepção para seleção dos juízes.

A questão da multidimensionalidade da voz reside no fato de que a qualidade vocal emerge de uma combinação de ações, de forma que não se pode

analisar a qualidade vocal com base num parâmetro. O roteiro VPAS apresenta-se como alternativa a abordar os aspectos de natureza fonatória, de tensão muscular e de atividade supralaríngea. Para tanto, requer, como em outras modalidades de roteiros de análise, formação, familiaridade e treinamento no uso do instrumento. Tal situação leva os pesquisadores que adotam o referido roteiro a adotar várias etapas de composição e seleção de seu(s) examinador(es)^{12,30,35-37}. Com a devida formação fonética, o avaliador passa a ser capaz de julgar a qualidade sonora proeminente na fala do indivíduo. Foi este o caminho trilhado, em que a adoção do referencial das Ciências Fonéticas propiciou a condição de detalhamento de eventos relacionados à qualidade vocal no grupo de professores avaliados com distúrbios de voz.

CONCLUSÃO

Com base na análise perceptiva dos 4 juízes congruentes foi traçado o perfil médio de qualidade vocal do grupo estudado (professoras da rede pública com distúrbio de voz e/ou alteração laríngea), sendo os ajustes mais frequentes neste grupo, em ordem decrescente de ocorrência, foram: ajustes de hiperfunção laríngea, voz áspera, laringe elevada, hiperfunção do trato vocal, mandíbula fechada, constrição faríngea, corpo de língua elevado e escape de ar.

Quanto aos aspectos da dinâmica vocal, em ordem decrescente de frequência, destacaram-se: suporte respiratório inadequado, diminuição da variabilidade de *pitch*, *pitch* habitual elevado, *loudness* habitual elevado, taxa de elocução rápida e variabilidade de *loudness* diminuída.

A proposta de abordagem metodológica desenvolvida para avaliar o desempenho do grupo de juízes na avaliação das qualidades de voz nesta pesquisa mostrou-se adequada, uma vez que a bateria de testes proposta permitiu definir os juízes, cujos julgamentos pautaram-se nos princípios fonéticos, bem como traçar o perfil médio de qualidade vocal do grupo de amostras de vozes analisadas.

REFERÊNCIAS

1. De Bodt MS, Van De Heyning PH, Wuyts FL, Lambrechts L. The perceptual evaluation of voice disorders. *Acta Otorhinolaryngol. Belg.* 1996;50(4):283-91.
2. Simões-Zenari M, Latorre MRDO. Changes in behavior associated to the use of voice after a speech therapy intervention with professionals of child day care centers. *Pró-Fono R Atual Cient.* 2008;20(1):61-6.
3. Lima-Silva MFB, Ferreira LP, Oliveira IB, Andrada e Silva MA, Ghirardi ACAM. Voice disorders in teachers: self-report, auditory-perceptive assessment of voice and vocal fold assessment. *Rev Soc Bras Fonoaudiol.* 2012;17(4):391-7.
4. Blaustein S, Bar A. Reliability of perceptual voice assessment. *J Commun Disord.* 1983;16(2):157-61.
5. Webb A, Carding PN, Deary IJ, Mackenzie K, Steen N, Wilson JA. The reliability of three perceptual evaluation scales for dysphonia. *Eur Arch Otorhinolaryngol.* 2004;261(8):429-34.
6. Silva RSA, Simões-Zenari M, Nemr NK. Impact of auditory training for perceptual assessment of voice executed by undergraduate students in Speech Language Pathology. *J Soc Bras Fonoaudiol.* 2012;24(1):19-25.
7. Kreiman J, Gerratt BR, Precoda K. Listener experience and perception of voice quality. *Speech Hear Res.* 1990;33(1):103-15.
8. Kreiman J, Gerratt BR, Precoda K, Berke GS. Individual differences in voice quality perception. *J Speech Hear Res.* 1992;35(3):512-20.
9. Kreiman J, Gerratt B, Kempster G, Erman A, Berke GS. Perceptual evaluation of voice quality: review, tutorial, and framework for future research. *J Speech Hear Res.* 1993;36(1):21-40.
10. Behlau M, Hogikyan ND, Gasparini G. Quality of life and voice: study of a Brazilian population using the voice-related quality of life measure. *Folia Phoniatr Logop.* 2007;59(6):286-96.
11. Camargo ZA, Madureira S. Avaliação vocal sob a perspectiva fonética: investigação preliminar. *Distúrb Comun.* 2008b;20(1):77-96.
12. Kreiman J, Gerratt B. Measuring voice quality. In: Kent R, Ball M. (Org). *Voice quality measurement.* San Diego: Singular Publishing; 2000. p. 73-99.
13. Valentim AF, Côrtes MG, Gama AC. Spectrographic analysis of the voice: effect of visual training on the reliability of evaluation. *Rev Soc Bras Fonoaudiol.* 2010;15(3):335-42.
14. Laver J. Phonetic evaluation of voice quality. In: Kent MJ, Martin JB (org). *Voice Quality Measurement.* San Diego: Singular, 2000.
15. Benninguer MS. Quality of the voice literature: what is there and what is missing. *J Voice.* 2011;25(6):647-52.

16. Pessoa-Almeida NA, Novaes BCC, Camargo Z. Dados perceptivo-auditivos e acústicos como indicadores prosódicos da fala em criança com deficiência auditiva. In: Camargo ZA (org). *Fonética Clínica: vinte anos de LIAAC*. São Paulo: Pulso, 2016. p.81-104.
17. Bailly G. Learning to speak. Sensori-motor control of speech movements. *Speech Communication*. 1997;22(1):251-67.
18. Meier RP, McGarvin G, Zakia RA, Willerman R. Silent mandibular oscillations in vocal babbling. *Phonetica*. 1997;54(4):153-71.
19. Buder EH, Choma LB, Oller DK, Robinson RB. Vibratory regime classification of infant phonation. *J Voice*. 2007;22(5):553-64.
20. Lima MFB, Camargo ZA, Ferreira LP, Madureira S. Qualidade vocal e formantes das vogais de falantes adultos da cidade de João Pessoa. *Rev. CEFAC*. 2007;9(1):99-109.
21. Camargo ZA, Madureira S. Voice quality analysis from a phonetic perspective: Voice Profile Analysis Scheme Profile for Brazilian Portuguese (BP-VPAS). In: *Fourth Conference on Speech Prosody*; 2008; Campinas, BR. São Paulo: Capes, Fapesp, CNPq, 2008a, v.1: 14.
22. Mackenzie-Beck J. Perceptual analysis of voice quality: the place of vocal profile analysis. In: *Hardcastle WJ, Mackenzie-Beck J (orgs). A figure of speech: a festschrift for John Laver*. Lawrence Erlbrum Associates: Mahwah; 2005. p. 285-322.
23. Camargo ZA, Madureira S, Tsuji DH. Analysis of dysphonic voices based on the interpretation of acoustic, physiological and perceptual data. In: *16th International Seminar on Speech Production Proceedings*; 2003; Sidney, Austrália. Sidney: Speech Production; 2003.
24. Lima MFB, Madureira S, Camargo ZA. Avaliação fonética de qualidade vocal em diferentes estilos de fala (semi-espontânea e leitura). In: *Anais do 17^a Congresso Brasileiro de Fonoaudiologia e 1^o Congresso Ibero-Americano de Fonoaudiologia*; 2009; Bahia, Brasil. São Paulo: Sociedade Brasileira de Fonoaudiologia; 2009. p. 1814.
25. Guirardello EB. Adaptação cultural e validação do instrumento Demandas de Atenção Dirigida. *Rev Esc Enferm USP*. 2005;39(1):77-84.
26. Köhle J, Camargo Z, Nemr K. Análise perceptivo-auditiva da qualidade vocal de indivíduos submetidos a laringectomias parciais verticais pela auto-avaliação dos indivíduos e pela avaliação fonoaudiológica. *Rev. CEFAC*. 2004;6(1):67-76.
27. Bele IV. Reability in perceptual analysis of voice quality. *J Voice*. 2005;19(4):555-73.
28. Sellars C, Stanton AE, McConnachie A, Dunnet CP, Chapman LM, Bucknall CE et al. Reliability of perceptions of voice quality: evidence from a problem asthma clinic population. *J Laryngol Otol*. 2009;123(7):755-63.
29. Oates J. Auditory-perceptual evaluation of disordered voice quality: pros, cons and future directions. *Folia Phoniatr Logop*. 2009;61(1):49-56.
30. Kreiman J, Gerratt B. The perceptual structure of pathologic voice quality. *Journal of the Acoustical Society of America*. 1996;100(3):1787-95.
31. Kreiman J, Sidtis D. *Foundations of voice studies: an interdisciplinary approach to voice production and perception*. Wiley-Blackwell: Malden, 2011.
32. Gama AC, Santos LL, Sanches NA, Côrtes MG, Bassi IB. Studying the effect of spectrogram visual support of in the auditory-perceptive voice evaluation reliability. *Rev. CEFAC*. 2011;13(2):314-21.
33. Eadie TL, Baylor CR. The effect of perceptual training on inexperienced listeners' judgments of dysphonic voice. *J Voice*. 2006;20(4):527-44.
34. Carding PN, Wilson JA, Mackenzie K, Deary IJ. Measuring voice outcomes: state of the science review. *J Laryngol Otol*. 2009;123(8):823-9.
35. Kreiman J. Listening to voices: theory and practice in voice perception research. In: *Johnson K, Mullennix JW (orgs). Talker variability in speech processing*. San Diego; 1997. p. 85-108.
36. Kreiman J, Gerratt B. Categorical judgments of vocal quality. Presented at the 134th Meeting of the Acoustical Society of America. San Diego; 1997.
37. Camargo ZA, Madureira S. The acoustic analysis of speech samples designed for the Voice Profile Analysis Scheme for Brazilian Portuguese (BP-VPAS): long term f0 and intensity measures. In: *Proceedings of the third ISCA Tutorial and research workshop on Experimental Linguistics*; 2010; Athens, Greece. Athens: International Speech Communication Association; 2010: 33-6.