

A qualidade de voz avaliada por meio de descritores impressionísticos e fonéticos: subsídios para o contexto forense

DOI: <http://dx.doi.org/10.21165/el.v54i3.3815>

Rafael Scarpelli¹

Renata Passetti²

Sandra Madureira³

Resumo

Este estudo avalia a relação entre descritores impressionísticos e fonéticos de qualidade de voz e discute as implicações para exames conduzidos na área de Fonética Forense envolvendo testemunhas auriculares. A metodologia compreendeu o julgamento de um conjunto de vozes por ouvintes com e sem treinamento em fonética e a avaliação dessas vozes por especialistas no uso do Vocal Profile Analysis (VPA). Análise estatística multivariada foi empregada para análise de relações entre os ajustes de qualidade de voz presentes no VPA (descritores fonéticos) e os adjetivos usados por ouvintes sem treinamento em fonética (descritores impressionísticos). Os resultados mostraram associações entre ajustes de qualidade de voz do VPA e adjetivos usados por ouvintes sem treinamento em fonética e lançam luz sobre a compreensão de como características de qualidade de voz impactam as percepções de ouvintes leigos e treinados, tema de interesse para a área de Fonética Forense.

Palavras-chave: qualidade de voz; Fonética Forense; Vocal Profile Analysis; descrição impressionística.

1 Pontifícia Universidade Católica de São Paulo (PUC-SP), São Paulo, São Paulo, Brasil; rafacuryscarpelli@gmail.com; <https://orcid.org/0000-0002-2062-721X>

2 Universidade Federal de São Carlos (UFSCar), São Carlos, São Paulo, Brasil; re.passetti@gmail.com; <https://orcid.org/0000-0002-1547-2831>

3 Pontifícia Universidade Católica de São Paulo (PUC-SP), São Paulo, São Paulo, Brasil; sandra.madureira.liaac@gmail.com; <https://orcid.org/0000-0001-8263-053X>

Voice quality described through impressionistic and phonetic labels: contributions to forensic application

Abstract

This study evaluates the relationship between impressionistic and phonetic descriptors of voice quality and discusses the implications for forensic phonetic investigation regarding earwitnesses. The methodology included the judgments of a set of voice samples by listeners with and without training in phonetics, as well as the assessment of these voices by experts using Vocal Profile Analysis (VPA). Multivariate statistical analysis was used to examine the relationships between the VPA settings (phonetic descriptors) and the adjectives used by untrained listeners (impressionistic descriptors). The results show associations between the VPA settings and adjectives used by untrained listeners and highlight how voice quality characteristics impact the perceptions of both lay and trained listeners – an area of interest in forensic phonetics.

Keywords: voice quality; Forensic Phonetics; Vocal Profile Analysis; impressionistic labels.

Introdução

Durante a investigação de um crime, é comum a coleta de depoimentos de testemunhas. Em algumas situações, elas podem oferecer detalhes sobre o crime ou ajudar a criar um retrato falado do criminoso; em outras podem descrever a voz do criminoso, como em crimes ocorridos por meio do telefone, envolvendo golpes e extorsões, por exemplo. Este trabalho aborda o cenário apresentado neste último exemplo, em que um “retrato” de uma voz é fornecido por ouvintes leigos. Procura-se entender a relação existente entre adjetivos fornecidos por ouvintes leigos e termos técnicos empregados por especialistas, como peritos e foneticistas, para caracterizar uma voz.

Desta forma, o objetivo deste trabalho é comparar avaliações impressionísticas de qualidade de voz, baseadas em descritores semânticos, com aquelas feitas por meio de descritores fonéticos. Busca-se compreender as relações entre termos leigos e técnicos para descrição da qualidade de voz e como esses mapeamentos terminológicos são organizados, bem como discutir as possíveis aplicações para exames conduzidos na área de Fonética Forense.

Fundamentação teórica

A Fonética Forense é uma área da fonética que aplica teorias, métodos e técnicas experimentais na perícia de voz (Jessen, 2008). A Comparação de Locutor (doravante

CL) é um dos principais exames conduzidos na área e tem como objetivo determinar a probabilidade de amostras de fala comparadas terem sido produzidas por um mesmo falante. Durante a condução dessa tarefa, o foneticista forense recorre a conhecimentos linguísticos, sobretudo fonéticos e sociofonéticos, para encontrar padrões semelhantes e/ou divergentes entre a amostra de fala questionada, associada ao criminoso, e a amostra de referência, associada a um suspeito.

A evolução do conhecimento e de técnicas de análise em Fonética Forense conferiu maior robustez e sistematicidade a investigações envolvendo amostras de voz como evidências de crimes. Dessa forma, foi possível evitar julgamentos que antes se baseavam apenas em depoimentos de testemunhas auriculares ou em análises sem o suporte de metodologias cientificamente fundamentadas.

Um caso de grande repercussão envolvendo o depoimento de uma testemunha auricular foi o sequestro do filho do aviador norte-americano Charles Lindbergh, ocorrido em março de 1932 em Nova Jersey. Durante as negociações do pagamento de resgate, o pai da vítima ouviu a voz do sequestrador. Passados 29 meses do ocorrido, Lindbergh afirmou ser de Bruno Hauptmann, um suspeito detido pela polícia, a voz ouvida durante as negociações e testemunhou contra ele durante o julgamento do caso (Eriksson, 2014).

O caso Lindbergh motivou estudos sobre o efeito do tempo sobre a memória de vozes não familiares e a confiabilidade dos depoimentos de testemunhas auriculares. Em uma revisão sobre a literatura envolvendo identificações oculares e auriculares, conduzida por Fraser (2019), a autora argumenta que depoimentos de testemunhas oculares possuem um nível de precisão questionável devido às altas taxas de erros e que depoimentos de testemunhas auriculares são ainda menos confiáveis, conforme corroborado por Kreiman e Sidtis (2011), recomendando cautela em seu uso nos tribunais.

A confiabilidade de identificações de vozes por ouvintes com treinamento (i.e. foneticistas forenses) e sem treinamento (i.e. ouvintes leigos) em fonética foi investigada por Schiller e Koster (1998). Neste experimento, os ouvintes deviam identificar as amostras pertencentes à voz-alvo em um conjunto com 108 amostras de fala de até oito segundos de duração. Os resultados mostraram o melhor desempenho de ouvintes com treinamento em fonética, cuja taxa de acerto foi de 98% e a de alarmes falsos foi de 1%⁴. Ouvintes sem treinamento apresentaram uma taxa de acerto de 92% e taxa de alarme falso de 2%. Embora os resultados de ambos os grupos de ouvintes tenham sido satisfatórios, esses percentuais revelam a maior precisão dos ouvintes com treinamento em fonética na identificação de vozes.

4 Os percentuais referentes a rejeições corretas não foram considerados na explanação acima.

Em estudo posterior, Koster *et al.* (2007) investigaram se ouvintes treinados teriam desempenho superior a ouvintes leigos na tarefa de identificação de ajustes de qualidade de voz de naturezas fonatórias (e.g. sopro, aspereza, crepitação etc.) e articulatórias (e.g. estiramento ou protusão labial, abertura ou fechamento de mandíbula etc.). Os especialistas apresentaram melhor desempenho que os ouvintes leigos, com taxas de erros de 17,8% e de 51,2%, respectivamente. Esses dados confirmam o melhor desempenho de ouvintes com treinamento em fonética em tarefas de identificação de vozes, como demonstrado no trabalho de Schiller e Koster (1998).

Ao investigar a relação entre termos usados por especialistas em fonética e rótulos impressionísticos dados por ouvintes leigos, Burns (2010) encontrou correlações entre descritores técnicos e aqueles dados por ouvintes leigos, como “Creaky” para descrever “Creaky voice” e “Falsetto” para falsete. No entanto, termos equivocados também foram registrados nas descrições dos ouvintes leigos, como o uso do termo “nasal” para vozes denasais. Além disso, esses ouvintes fizeram referência a vozes de personagens fictícios e de figuras públicas para descrever qualidades de voz, como “Sean Connery” para se referir ao ajuste de língua retroflexa e “Kermit, the Frog” para se referir ao ajuste de laringe elevada. Esses achados sublinham a importância de desenvolver metodologias que melhorem a comunicação entre especialistas e leigos na descrição e caracterização de vozes.

A presente pesquisa tem como objeto de análise a qualidade de voz, compreendida aqui em seu sentido amplo, conforme definido por Laver (1980). Para o linguista, a qualidade de voz resulta da combinação a longo prazo de ajustes de tensão, extensão e de configuração do trato vocal (Laver, 1968). Os ajustes correspondem às posturas articulatórias e fonatórias que se mantêm ao longo da fala e alteram a qualidade dos segmentos fônicos. Esses ajustes, que conferem “coloridos” às vozes, permitem que informações linguísticas e extralinguísticas, como sexo, idade perceptível, região de origem e estado emocional, possam ser acessadas (Madureira, 2020).

Por considerar a qualidade de voz de forma componencial, o modelo fonético de Laver (1980) permite que diferentes componentes de uma voz sejam registrados. Essa abordagem tem grande vantagem para a área forense, pois, pelo seu uso, torna-se possível, em uma perícia criminal, a comparação do conjunto de ajustes que compõem os perfis vocais do suspeito e do criminoso, auxiliando na identificação de semelhanças e divergências em termos de qualidade de voz.

O modelo fonético da qualidade de voz foi sistematizado por meio do Vocal Profile Analysis (doravante VPA, Laver *et al.*, 1991). Em sua versão mais recente⁵ (Laver; Mackenzie-Beck, 2007), constituem o sistema VPA 37 ajustes de qualidade de voz e 19 ajustes de dinâmica

⁵ A versão mais atual do sistema VPA (Laver; Mackenzie-Beck, 2007) está apresentada na seção Anexos.

vocal. Os ajustes de qualidade de voz informam sobre aspectos articulatórios, fonatórios, de tensão e de extensão do trato vocal, enquanto os ajustes de dinâmica vocal informam sobre os aspectos prosódicos e de organização temporal.

A análise da qualidade de voz por meio do sistema VPA é realizada em duas etapas: a primeira, em que se avalia o caráter de neutralidade ou não neutralidade dos 56 ajustes que o compõem, e a segunda, em que se atribui graus de 1 a 6 para os ajustes identificados como não neutros. É possível registrar ainda ajustes não neutros que se manifestam intermitentemente.

O sistema VPA foi utilizado por Scarpelli, Passetti e Madureira (2022) para investigar relações entre descritores técnicos e não técnicos de qualidade de voz no português brasileiro. Nesta pesquisa, adjetivos impressionísticos atribuídos por ouvintes leigos ao conjunto de vozes foram comparados a ajustes não neutros de qualidade de voz atribuídos a essas vozes por especialistas no uso do sistema VPA. Os termos impressionísticos se relacionam tanto a ajustes de configuração e tensão dos subsistemas laríngeo e supralaríngeo quanto a ajustes de dinâmica vocal. São exemplos dessas relações os empregos dos adjetivos “abafada”, associado tanto a um *pitch* médio abaixado quanto à posição recuada do corpo de língua, e “esganiçada”, associado à presença de tensão no trato vocal e a um *pitch* médio elevado.

Os estudos apresentados (Scarpelli et al, 2022; Burns, 2010) evidenciam a relação existente entre descritores técnicos e não técnicos na análise de vozes e a capacidade de ouvintes leigos descreverem uma voz com certo grau de precisão. Portanto, compreender as relações existentes entre termos impressionísticos e técnicos para a descrição de vozes é crucial para a Fonética Forense e outras áreas que têm a voz como objeto de análise. Os desafios são compreender como os mapeamentos entre termos impressionísticos e termos técnicos são organizados e avaliar a confiabilidade dos depoimentos de testemunhas leigas (Fraser, 2019; Jong *et al.*, 2015; Nolan; Grabe, 1996), a fim de se discutir as possíveis aplicações para exames conduzidos na área de Fonética Forense.

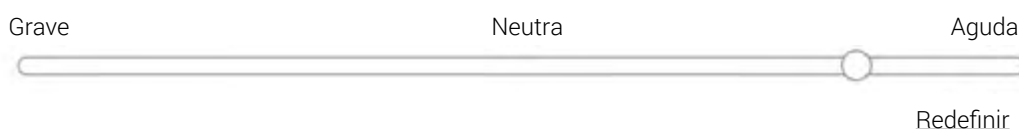
Metodologia

Os experimentos de natureza perceptiva conduzidos neste estudo foram realizados em três etapas: (1) experimento envolvendo julgamentos das características de vozes por juízes leigos, (2) experimento envolvendo julgamentos por juízes treinados com base nos qualificadores fornecidos pelos leigos e (3) avaliação fonético-auditiva das vozes por especialistas no uso do Vocal Profile Analysis (VPA).

Primeiro teste perceptivo

O primeiro teste perceptivo teve como objetivo avaliar a percepção de ouvintes leigos. O teste consistiu na seleção de adjetivos pelos ouvintes leigos, a partir de uma lista de pares de adjetivos. Os ouvintes foram instruídos a escolher os adjetivos da lista que consideravam mais adequados para descrever cada voz apresentada e uma pontuação referente ao grau de manifestação desses adjetivos. A lista foi apresentada em uma escala contínua de diferencial semântico, representada na Figura 1, contendo uma barra de deslizamento, o que permitia a movimentação em três níveis. Os participantes atribuíam pontuações de 1 a 3 para cada adjetivo, indicando o nível de manifestação na voz avaliada. O deslizador era posicionado mais próximo do adjetivo que melhor caracterizava a voz ouvida, conforme a pontuação de intensidade percebida.

Figura 1. Barra de deslizamento para uma voz considerada aguda com pontuação 2



Fonte: Elaboração própria

Os dez pares de adjetivos escolhidos para este teste são advindos da pesquisa de Scarpelli *et al.* (2022): Grave/Aguda, Tensa/Relaxada, Rouca/Límpida, Jovem/Idosa, Esganiçada/Suave, Entusiasmada/Desanimada, Rápida/Lenta, Segura/Insegura, Forte/Fraca e Abafada/Clara.

No primeiro teste perceptivo, participaram 58 ouvintes sem treinamento em fonética. Os participantes foram convidados por *e-mail*. Dos 58 participantes, 16 eram homens, e 42, mulheres. As idades variaram de 16 a 80 anos (desvio padrão de 18,3 anos). Os participantes eram de diversos estados, principalmente São Paulo (n = 48), com alguns de Minas Gerais (n = 4), Rio de Janeiro (n = 1), Tocantins (n = 1) e Amazonas (n = 1). A escolaridade variou de ensino médio incompleto a pós-graduação completa, e não houve divergências nas respostas devido à região de origem.

Para analisar a concordância entre as respostas dos participantes nesse primeiro teste perceptivo, aplicamos o teste Alpha de Cronbach (Cronbach, 1951), utilizando a linguagem R no *software* R Studio. O teste foi realizado com o auxílio do pacote Psych (Revelle, 2023). De acordo com as orientações do teste, valores acima de 0,65 são recomendados para indicar uma boa consistência interna. O coeficiente Alfa de Cronbach obtido para as respostas desses participantes foi de 0,67, indicando uma coerência moderada. O intervalo de confiança a 95% variou entre 0,61 e 0,69, evidenciando a consistência moderada das respostas dos participantes.

Segundo teste perceptivo

A segunda etapa do estudo compreendeu a aplicação de um teste perceptivo envolvendo filas de reconhecimento vocais, as quais foram geradas conforme as diretrizes apresentadas em Broeders e Van Amelvoort (1999). Essas filas eram compostas pelo mesmo conjunto de vozes do primeiro teste perceptivo, porém com cada amostra editada para ter a duração de 20 segundos. Cada fila de reconhecimento era composta por cinco amostras de fala, uma voz-alvo e quatro vozes distratoras. As amostras de fala foram organizadas aleatoriamente nas filas por meio do *script* CreateLineUp (Barbosa, 2021) Praat.

Cada fila de reconhecimento foi acompanhada pelos quatro adjetivos mais frequentes usados para descrever a voz-alvo, selecionados com base nos resultados do primeiro teste perceptivo. Esses adjetivos ajudaram a guiar os juízes na identificação das vozes-alvo nas filas de reconhecimento. O objetivo foi investigar se os adjetivos atribuídos pelos participantes do primeiro teste seriam úteis para o reconhecimento das vozes descritas. Participaram desta etapa 32 juízes com formação em fonética (linguistas, fonoaudiólogos e peritos de voz). A idade dos participantes variou entre 22 e 66 anos (desvio-padrão de 11,6 anos). A origem dos participantes variou entre São Paulo, com 13 participantes; Paraíba, com 5; Rio Grande do Norte, Ceará, Mato Grosso do Sul e Santa Catarina, com 2 cada um; e Rio Grande do Sul, Pernambuco, Goiás, Maranhão, Sergipe e Distrito Federal, com 1 participante cada. Todos os participantes possuíam pós-graduação, com exceção de um deles, que possuía apenas ensino superior.

Avaliação do conjunto de vozes por especialistas no VPA

As amostras de fala utilizadas no primeiro teste perceptivo também foram avaliadas por um grupo de três juízes treinados no uso do VPA, no caso, os autores deste trabalho. Cada juiz preencheu o VPA de forma independente, e, posteriormente, as avaliações individuais foram comparadas para discutir convergências e divergências. Esse processo permitiu a elaboração de um único perfil vocal para cada uma das 10 amostras de fala.

Análise estatística

A correlação entre os descritores impressionísticos – adjetivos atribuídos pelos ouvintes leigos – e técnicos da qualidade de voz – ajustes não neutros do VPA – foi investigada por meio do método estatístico Multiple Factor Analysis (MFA). Seu principal objetivo é correlacionar as variáveis qualitativas e quantitativas, considerando a projeção dessas variáveis no espaço vetorial dentro das dimensões com maior ganho inercial, ou seja, aquelas que compreendem as variáveis significativas. Dentro de uma mesma dimensão, as variáveis apresentam correlações, sejam elas positivas, ou negativas (Husson; Le; Pagè, 2020).

Resultados

Dentre os adjetivos mais selecionados no primeiro teste perceptivo para descrever as vozes do *corpus* de pesquisa deste trabalho, podemos destacar “grave”, “entusiasmada” e “fraca”.

As taxas de acerto, referentes à escolha das vozes-alvo corretamente com base no conjunto de adjetivos que as descrevia, do segundo teste perceptivo variaram entre 6% e 66%, com algumas vozes apresentando alarmes falsos significativos (vozes distratoras selecionadas com mais frequência do que a voz-alvo). Esses dados mostram uma diversidade nas percepções e ajustes para diferentes vozes.

Dentre os ajustes mais frequentes nos perfis vocais dos falantes, advindos da etapa de avaliação das vozes por meio do protocolo VPA, destacam-se *creaky voice*, nasalidade, estiramento de lábios e ponta de língua avançada.

Os resultados das três etapas da pesquisa, para cada uma das vozes-alvo, estão resumidos na lista abaixo:

- **V1:** Adjetivos: grave, relaxada, segura, forte. Taxa de acerto: 56%. Ajustes: corpo de língua abaixado, nasal, *creaky voice*, média alta de *loudness*.
- **V2:** Adjetivos: relaxada, límpida, desanimada, lenta. Taxa de acerto: 22%. Voz distratora (alarme falso): V7 (34%). Ajustes: nasal, corpo de língua avançado, *creaky voice*, alta variabilidade de *loudness*.
- **V3:** Adjetivos: tensa, desanimada, lenta, fraca. Taxa de acerto: 59%. Ajustes: extensão de lábios minimizada, ponta de língua avançada, corpo de língua avançado, nasal intermitente, trato vocal relaxado, laringe relaxada, *creaky voice* intermitente, média alta de *pitch*, fala interrompida.

- **V4:** Adjetivos: relaxada, entusiasmada, segura, forte. Taxa de acerto: 6%. Voz distratora (alarme falso): V5 (50%). Ajustes: estiramento de lábios, corpo de língua abaixado, nasal, *creaky voice* (grau 2).
- **V5:** Adjetivos: límpida, entusiasmada, segura, forte. Taxa de acerto: 44%. Ajustes: estiramento de lábios, extensão ampla de lábios, extensão ampla de língua, nasal, trato vocal tenso, *creaky voice* (grau 2), extensão ampla de *pitch*, variabilidade ampla de *pitch*, média alta de *loudness*.
- **V6:** Adjetivos: grave, tensa, límpida, insegura. Taxa de acerto: 22%. Voz distratora (alarme falso): V10 (53%). Ajustes: lábios arredondados, corpo de língua recuado, laringe abaixada, extensão ampla de *loudness* (grau 2), variabilidade alta de *loudness*, fala interrompida.
- **V7:** Adjetivos: grave, relaxada, suave, segura. Taxa de acerto: 38%. Ajustes: extensão minimizada de lábios, extensão minimizada de mandíbula, corpo de língua levantado, *creaky voice* intermitente, *whispery*.
- **V8:** Adjetivos: aguda, relaxada, esganiçada, segura. Taxa de acerto: 44%. Ajustes: extensão ampla de mandíbula, corpo de língua recuado, corpo de língua abaixado, nasal, *creaky voice* (grau 2), variabilidade alta de *pitch*.
- **V9:** Adjetivos: aguda, límpida, rápida, fraca. Taxa de acerto: 66%. Ajustes: estiramento de lábios, extensão minimizada de lábios, ponta de língua avançada, corpo de língua avançado, denasal, *creaky voice*, média alta de *pitch*, média alta de *loudness*, variabilidade alta de *loudness*, taxa de elocução rápida.
- **V10:** Adjetivos: grave, tensa, insegura, forte. Taxa de acerto: 50%. Ajustes: extensão ampla de lábios, extensão ampla de mandíbula, corpo de língua abaixado, *creaky voice* intermitente, extensão ampla de *pitch*, fala interrompida.

Na tabela 1 apresentamos a relação dos resultados do segundo teste perceptivo com filas de reconhecimento de voz. Na parte superior da tabela, em negrito, temos a porcentagem de seleção da voz-alvo (taxa de acerto); de E1 a E5, temos a ordem apresentada na fila de reconhecimento; e na parte inferior da tabela, temos os estímulos apresentados em cada fila, com destaque para a voz-alvo em negrito.

Tabela 1. Resultados do segundo teste perceptivo (taxas de acerto e alarmes falsos)

	Fila 1	Fila 2	Fila 3	Fila 4	Fila 5	Fila 6	Fila 7	Fila 8	Fila 9	Fila 10
E1	13%	22%	16%	19%	9%	16%	22%	16%	16%	9%
E2	9%	28%	59%	50%	44%	9%	38%	9%	6%	9%
E3	0%	0%	9%	6%	6%	22%	13%	3%	66%	9%
E4	56%	34%	9%	22%	22%	53%	13%	28%	9%	50%
E5	22%	16%	6%	3%	19%	0%	16%	44%	3%	22%
	Fila 1	Fila 2	Fila 3	Fila 4	Fila 5	Fila 6	Fila 7	Fila 8	Fila 9	Fila 10
E1	V4	V2	V10	V7	V4	V3	V2	V5	V3	V2
E2	V2	V3	V3	V5	V5	V5	V7	V2	V7	V3
E3	V9	V5	V6	V4	V9	V6	V10	V10	V9	V1
E4	V1	V7	V2	V1	V1	V10	V3	V4	V6	V10
E5	V7	V8	V9	V8	V6	V8	V8	V8	V1	V4

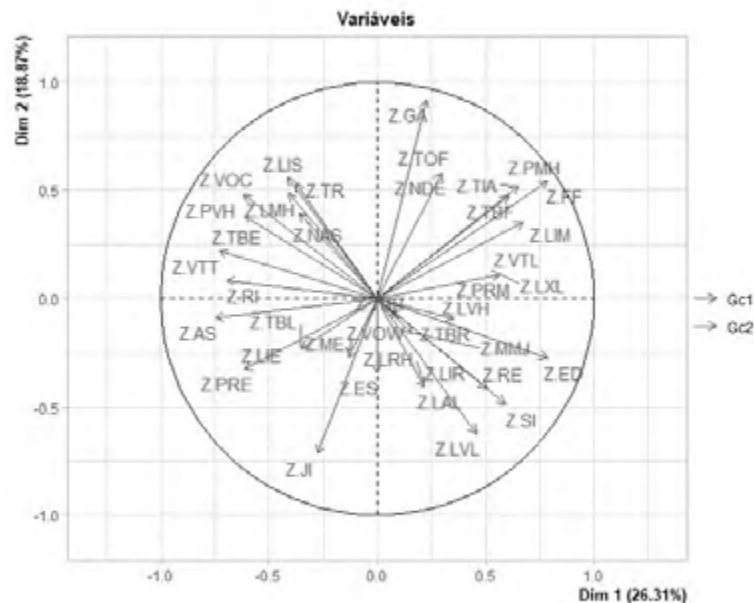
Fonte: Elaboração própria

Resultado da análise multifatorial para todas as variáveis

Para a análise da correlação entre descritores impressionísticos e técnicos da qualidade de voz, foram considerados dois grupos de dados: as médias das pontuações obtidas para cada par de adjetivos no primeiro teste perceptivo (Gc1) e os graus atribuídos a cada ajuste nos estímulos avaliados como não neutros no VPA (Gc2). Ambos os conjuntos de dados foram tratados como variáveis quantitativas na análise, utilizando o ambiente e a linguagem R. Essa abordagem permitiu investigar a estrutura subjacente dos dados e identificar as relações entre as variáveis de tipos diferentes.

Na Figura 2, temos o gráfico de vetores para todas as variáveis nas dimensões 1 e 2. Por meio deste gráfico, podemos observar a relação de oposição ou semelhança entre as variáveis. Por exemplo, a variável GA (Grave/Aguda) está em oposição à variável JI (Jovem/Idosa), ou seja, quando o valor de uma sobe, o da outra decresce. A mesma relação pode ser apontada em relação às variáveis FF (Forte/Fraca) e PRE (Extensão ampla de *pitch*), RI (Rouca/Límpida) e ED (Entusiasmada/Desanimada), LVL (Fala interrompida) e LIS (Lábios estirados). Em termos de relação de semelhança, podemos observar as variáveis que estão linearmente dependentes. É o caso de FF (Forte/Fraca), PMH (Média de *pitch* alto), LIM (Extensão de lábios minimizada), TIA (Ponta de língua avançada) e TBF (Corpo de língua avançado); LMH (Média alta de *loudness*), LIS (Lábios estirados) e TR (Tensa/Relaxada), dentre outras.

Figura 2. Gráfico de vetores para todas as variáveis nas dimensões 1 e 2



Fonte: Elaboração própria

As variáveis significativas com maior valor de correlação, em módulo, são: AS (Abafada/ Clara), ED (Entusiasmada/Desanimada), ES (Esganiçada/Suave), FF (Forte/Fraca), GA (Grave/aguda), JI (Jovem/Idosa), TR (Tensa/Relaxada), RI (Rouca/Límpida), LIM (Extensão de lábios minimizada), LVH (Alta variabilidade de *loudness*), PMH (Média alta de *pitch*), TBE (Extensão ampla de corpo de língua), TIA (Ponta de língua avançada) e VTT (Trato vocal tenso).

Com base nessas variáveis, foi realizada uma nova análise multifatorial para explorar as relações subjacentes entre elas. As variáveis presentes na Tabela 2 foram utilizadas nesta nova etapa. Os dados da Tabela 2 mostram as correlações e valores de p para as variáveis investigadas dentro das três dimensões com maior ganho inercial. Na análise multivariada, o termo “dimensões” refere-se ao número de variáveis ou características em um conjunto de dados, que podem ser visualizadas como eixos em um espaço geométrico. As variáveis dentro de uma mesma dimensão estão correlacionadas de acordo com o valor apresentado.

Tabela 2. Variáveis projetadas em cada dimensão e respectivos índices de correlação e valor de significância

Dimensão	Variável	Correlação	Valor de p
Dim 1	Z,ED (Entusiasmada/Desanimada)	0,788	0,0068
Dim 1	Z,FF (Forte/Fraca)	0,785	0,0071
Dim 1	Z,LIM (Extensão minimizada de lábios)	0,6717	0,0334
Dim 1	Z,TIA (Ponta de língua avançada)	0,6491	0,0423
Dim 1	Z,PMH (Média alta de <i>pitch</i>)	0,6491	0,0423
Dim 1	Z,RI (Rouca/Límpida)	-0,6957	0,0255
Dim 1	Z,VTT (Trato vocal tenso)	-0,7275	0,0171
Dim 1	Z,TBE (Extensão ampla de corpo de língua)	-0,7275	0,0171
Dim 1	Z,AS (Abafada/Clara)	-0,7441	0,0136
Dim 2	Z,GA (Grave/Aguda)	0,9156	0,0002
Dim 2	Z,JI (Jovem/Idosa)	-0,7137	0,0204
Dim 3	Z,LVH (Alta variabilidade de <i>pitch</i>)	0,6676	0,0349
Dim 3	Z,TR (Tensa/Relaxada)	-0,6441	0,0444
Dim 3	Z,ES (Esganiçada/Suave)	-0,7091	0,0217

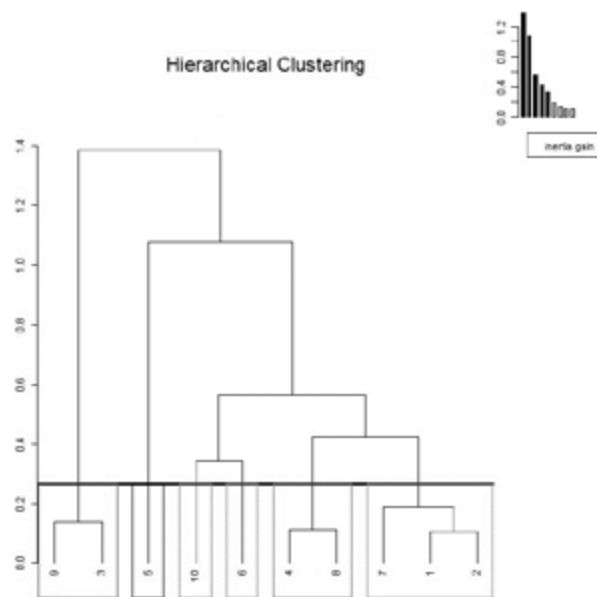
Fonte: Elaboração própria

As variáveis presentes na Tabela 2 geraram a distribuição dos estímulos nos *clusters* compreendidos no dendrograma da Figura 3. Considerando a extensão no eixo vertical, quanto menor a extensão da chave, maior a similaridade entre os estímulos dentro de um *cluster*. Dessa maneira, os estímulos 1 e 2 são os que apresentam maior grau de similaridade, seguidos dos estímulos 4 e 8.

Considerado o corte em 0,25 no dendrograma da Figura 3, configuraram-se os seguintes *clusters*: *Cluster 1* – V5; *Cluster 2* – V10; *Cluster 3* – V1, V2 e V7; *Cluster 4* – V4 e V8; *Cluster 5* – V6; *Cluster 6* – V3 e V9. Os dados do dendrograma indicam similaridade entre os estímulos agrupados dentro de um mesmo *cluster*.

A seguir, apresentamos o dendrograma com a distribuição de todos os estímulos. Na parte superior direita, o dendrograma verifica-se pelas extensões das barras que as dimensões 1, 2 e 3 são as que apresentaram maior ganho inercial.

Figura 3. Dendrograma para variáveis seleccionadas



Fonte: Elaboração própria

Discussão

Os estímulos V1 e V5, avaliados no VPA com média alta de *loudness*, foram associados ao adjetivo “forte” pelos participantes. O mapa vetorial (Figura 2) mostrou que as variáveis PRE (Extensão ampla de *pitch*), LIE (Extensão ampla de lábios) e MEJ (Extensão ampla de mandíbula) também se relacionam com a percepção de “forte”.

No entanto, ao se observar os resultados obtidos no segundo teste perceptivo, nota-se que o estímulo V9, apesar de ter sido avaliado no VPA com uma média alta de *loudness*, foi percebido pelos ouvintes treinados como “fraco”. Provavelmente esta percepção deve-se à interferência de outros ajustes de qualidade de voz no perfil vocal deste falante, como TIA (Ponta de língua avançada), TBF (Corpo de língua avançado), PMH (Média de *pitch* alta) e LIM (extensão de lábios minimizada). Além disso, esses ajustes também são compartilhados pelo estímulo V3, o que pode explicar a maior taxa de alarmes falsos (16%) para essa voz na fila em que V9 era a voz-alvo.

A análise multifatorial evidencia essas observações, mostrando uma correlação entre o aumento da variável FF (Forte/Fraca) e os ajustes TIA, TBF, PMH e LIM. O mapa vetorial para todas as variáveis nas dimensões 1 e 2 (Figura 2) destaca essa relação, indicando que a combinação desses ajustes influencia como os ouvintes percebem uma voz como “forte” ou “fraca”.

A fala interrompida, caracterizada por uma alta incidência de pausas silenciosas (ajuste "*Interrupted*") de organização temporal referente à avaliação de continuidade na fala (Laver; Mackenzie-Beck, 2007) foi relacionada à percepção de "tensão" e "insegurança" nos estímulos V3, V6 e V10. Em contraste, a fala "relaxada" foi associada à ausência de interrupções. O gráfico vetorial para todas as variáveis nas dimensões 1 e 2 (Figura 2) evidencia essa correlação, mostrando uma relação de oposição entre as variáveis LVL (fala interrompida) e TR (tensa/relaxada), e uma relação linear entre SI (segura/insegura) e LVL. Esse resultado demonstra que os adjetivos "tensa" e "relaxada" possivelmente foram interpretados pelos ouvintes como o estado mental de tensão ou relaxamento, respectivamente, e não como uma característica fisiológica relacionada ao estado de rigidez ou relaxamento de partes dos subsistemas laríngeo e supralaríngeo (trato vocal), como avaliado no VPA.

A presença do ajuste não neutro de lábios estirados no perfil vocal do VPA dos estímulos V4 e V5 foi associada ao adjetivo "entusiasmada" pelos ouvintes leigos, provavelmente devido à sensação de "sorriso" nas vozes desses falantes. O estímulo V5 foi mais frequentemente descrito com esse adjetivo, porque combinava diferentes ajustes de qualidade de voz: extensão ampla de língua, trato vocal tenso, extensão ampla de *pitch* e média alta de *loudness*. O gráfico vetorial da análise multifatorial para todas as variáveis nas dimensões 1 e 2 (Figura 2) confirma essa relação componencial. O gráfico mostra que as variáveis LIS (estiramento de lábios), LMH (média alta de *loudness*), TBE (Extensão ampla de corpo de língua), VTT (trato vocal tenso) e PVH (alta variabilidade de *pitch*) diminuem quando a variável ED (entusiasmada/desanimada) aumenta, o que indica a seleção do adjetivo "entusiasmada" para esses ajustes de qualidade de voz.

Por fim, o dendrograma para variáveis selecionadas (Figura 3) explica alguns dos alarmes falsos que ocorreram no segundo teste perceptivo, com ouvintes treinados. Os estímulos V2 e V7 foram agrupados dentro de um mesmo *cluster* e foram confundidos na segunda fila de reconhecimento. Os estímulos V6 e V10, que foram confundidos na sexta fila, foram agrupados de forma muito próxima, dentro de um mesmo *cluster*, se considerarmos um corte em 0,4 em vez de 0,25 como foi realizado. Estímulos com distribuição próxima dentro do dendrograma, tais como os estímulos V3 e V9 na terceira fila de reconhecimento e V4 e V8 na oitava fila de reconhecimento, produziram alarmes falsos.

Conclusão

A análise multifatorial revelou correlações entre adjetivos utilizados por ouvintes sem treinamento em fonética e ajustes do VPA relacionados à modificação, extensão e tensão do trato vocal, além da dinâmica vocal. Relações estabelecidas entre adjetivos impressionísticos, ajustes do VPA e taxas de acerto no teste perceptivo com ouvintes treinados evidenciam o potencial de alguns adjetivos para descrição de vozes.

Os adjetivos que tiveram melhor desempenho na descrição dos estímulos foram: o adjetivo “fraca”, correlacionado com corpo de língua avançado, ponta de língua avançada, extensão minimizada de lábios e média de *pitch* alta; o adjetivo “entusiasmada”, associado à extensão ampla de corpo de língua, variabilidade alta de *pitch*, trato vocal tenso, lábios estirados e média alta de *loudness*; o adjetivo “forte”, relacionado à extensão ampla de lábios e mandíbula; os adjetivos “tensa” e “insegura”, correlacionados à fala interrompida.

O adjetivo “límpida” possuiu pouco potencial descritor no segundo teste perceptivo, e foi correlacionada à ausência de ajustes que provocassem sensação de “rouquidão” na voz, como “*Creaky*”, “*Harsh*” e “*Whispery*”.

Essas relações confirmam a hipótese de que há pareamentos entre termos impressionísticos e técnicos na descrição de qualidade de voz. Este trabalho preenche uma lacuna na pesquisa sobre associações entre ajustes de qualidade de voz e descrições de vozes com base em descritores semânticos, lançando luz sobre as atribuições feitas por ouvintes sem treinamento em fonética.

Este estudo experimental, ao comparar descrições de vozes feitas por leigos com aquelas realizadas por especialistas no VPA, apresenta pareamentos existentes entre termos impressionísticos e fonético-acústicos de descrição da qualidade de voz no português brasileiro. Essas correlações fornecem subsídios para a condução de análises envolvendo o depoimento de testemunhas auriculares e para a composição de laudos periciais em Fonética Forense. Como continuidade da pesquisa, pretende-se aprofundar a investigação da relação entre descritores impressionísticos e fonéticos da qualidade de voz, incluindo a análise acústica de ajustes de qualidade de voz. Isso enriquecerá a discussão sobre descritores impressionísticos e terá desdobramentos para áreas como a Fonética Forense.

Agradecimentos

O primeiro autor agradece ao CNPq e ao Programa de Pós-Graduação em Linguística Aplicada e Estudos da Linguagem da PUCSP (PPG-LAEL) pela Bolsa de Mestrado concedida (Processo CNPq nº 130094/2022-6). A segunda autora agradece à CAPES e ao Programa de Pós-Graduação em Linguística da UFSCar (PPGL/UFSCar) pela Bolsa de Pós-Doutorado concedida (Processo CAPES nº 88887.804443/2023-00). A segunda autora agradece, ainda, o apoio financeiro concedido pelo PPGL/UFSCar, proveniente de verba CAPES/PROAP, para sua participação no evento que gerou a presente publicação.

Referências

BARBOSA, P. *Create Lineup*, 2021.

BECK, J. Perceptual analysis of voice quality: the place of vocal profile analysis. In: HARDCASTLE, W. J.; BECK, J. (org.). *A Figure of Speech: A Festschrift for John Laver*. 1. ed. [S. l.]: Routledge, 2005. p. 285-322.

BECK, J. *Vocal Profile Analysis Scheme: A User's Manual*. Edinburgh: Queen Margaret University College-QMUC, Speech Science Research Centre, 2007.

BROEDERS, T.; VAN AMELVOORT, A. G. Lineup Construction for Forensic Earwitness Identification: A Practical Approach. 1999, San Francisco, CA. *Proceedings of the XIVth International Congress of Phonetic Sciences*. San Francisco, CA: [s. n.], 1999. p. 1373-1376.

BURNS, J. *Voice Quality: Labelling by Untrained Listeners*. [S. l.: s. n.], 2010.

ERIKSSON, A. Aural/Acoustic vs. Automatic Methods in Forensic Phonetic Case Work. In: NEUSTEIN, A.; PATIL, H. (org.). *Forensic Speaker Recognition*. [S. l.]: s.n., 2012. p. 41-69.

ERIKSSON, A. Tutorial sobre fonética forense. *Revel*, [s. l.], v. 12, n. 23, 2014.

FRASER, H. The reliability of voice recognition by 'ear witnesses'. *Language and Law / Linguagem e Direito*, [S. l.], v. 6, p. 1-9, 2019.

HUSSON, F.; LE, S.; PAGÈS, J. *Exploratory multivariate analysis by example using R*. 2. ed. Boca Raton: Chapman and Hall/CRC, 2020.

JESSEN, M. Forensic phonetics. *Language and Linguistics Compass*, [s. l.], v. 2, n. 4, p. 671-711, 2008.

JONG, G. D. *et al.* Voice lineups: A practical guide. 2015. *Proceedings of the International Congress of Phonetic Sciences*. [S. l.]: s.n., 2015.

KÖSTER, O.; JESSEN, M.; KHAIRI, F.; ECKERT, H. Auditory-perceptual identification of voice quality by expert and non-expert listeners. In: INTERNATIONAL CONGRESS OF PHONETIC SCIENCES, 16., 2007, Saarbrücken. *Proceedings [...]*. Saarbrücken: ICPhS, 2007. p. 1845-1848. Disponível em: <http://www.icphs2007.de>. Acesso em: 2 out. 2024.

KREIMAN, J.; SIDTIS, D. Identifying unfamiliar voices in forensic contexts. In: *Foundations of voice studies: An interdisciplinary approach to voice production and perception*. Oxford: Wiley Blackwell, 2011. Capítulo 7, p. 237-259.

LAVER, J. *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press, 1980.

LAVER, J. Voice Quality and Indexical Information. *British Journal of Disorders of Communication*, [s. l.], v. 3, n. 1, p. 43-54, 1968.

LAVER, J.; WIRS, S.; MACKENZIE-BECK, J.; HILLER, S. M. A perceptual protocol for the analysis of vocal profiles. In: LAVER, J. (ed.). *The gift of speech: papers in the analysis of speech and voice*. Edinburgh: Edinburgh University Press, 1991. p. 265-280.

LAVER, J.; MACKENZIE-BECK, J. *Vocal Profile Analysis Scheme-VPAS*. Queen Margaret University College-QMUC, Speech Science Research Centre. Edinburgh. 2007.

NOLAN, F.; GRABE, E. Preparing a voice lineup. *Forensic Linguistics*, [S. l.], v. 3, p. 74-94, 1996.

PASSETTI, R. O EFEITO DO TELEFONE CELULAR NO SINAL DA FALA: Uma análise fonético-acústica com implicações para a verificação de locutor em português brasileiro. Dissertação de Mestrado – Universidade Estadual de Campinas (Unicamp), Instituto de Estudos da Linguagem Campinas, SP, 2015.

SCARPELLI, R; PASSETTI, R; MADUREIRA, S. Avaliação impressionística e fonético-descritiva de qualidades de voz: convergências, divergências e contextos de aplicação forense. In: CONGRESSO BRASILEIRO DE PROSÓDIA, 2., 2022, Florianópolis. Anais [...]. Florianópolis: Lbass, 2023. v. 1, p. 120-133. Disponível em: http://www.periodicos.letras.ufmg.br/index.php/anais_coloquio/index.

SCHILLER, O. N.; KOSTER, O. *The ability of expert witnesses to identify voices: a comparison between trained and untrained listeners*. [S. l.]: University of Birmingham Press, 1998.

Anexo – Protocolo Vocal Profile Analysis (Laver; Mackenzie-Beck, 2007)

		FIRST PASS		SECOND PASS						
		Neutral	Non-neutral	SETTING	moderate			extreme		
					1	2	3	4	5	6
A. VOCAL TRACT FEATURES										
1. Labial			Lip rounding							
			Lip spreading							
			Labiodentalization							
			Extensive range							
			Minimised range							
2. Mandibular			Close jaw							
			Open jaw							
			Protruded jaw							
			Extensive range							
			Minimised range							
3. Lingual tip/blade			Advanced tip/blade							
			Retracted tip/blade							
4. Lingual body			Fronted tongue body							
			Backed tongue body							
			Raised tongue body							
			Lowered tongue body							
			Extensive range							
5. Pharyngeal			Minimised range							
			Pharyngeal constriction							
6. Velopharyngeal			Pharyngeal expansion							
			Audible nasal escape							
7. Larynx height			Nasal							
			Denasal							
			Raised larynx							
			Lowered larynx							
B. OVERALL MUSCULAR TENSION										
8. Vocal tract tension			Tense vocal tract							
			Lax vocal tract							
9. Laryngeal tension			Tense larynx							
			Lax larynx							
C. PHONATION FEATURES										
	SETTING	Present		Scalar degree						
		Neutral	Non-neutral	Moderate			Extreme			
				1	2	3	4	5	6	
10. Voicing type	Voice									
	Falsetto									
	Creak									
	Creaky									
11. Laryngeal friction	Whisper									
	Whispery									
12. Laryngeal irregularity	Harsh									
	Tremor									
				moderate			extreme			
				1	2	3	4	5	6	
D. PROSODIC FEATURES										
13. Pitch	Mean	High								
		Low								
	Range	Extensive range								
		Minimised range								
Variability	High									
	Low									
14. Loudness	Mean	High								
		Low								
	Range	Extensive range								
		Minimised range								
Variability	High									
	Low									
E. TEMPORAL ORGANIZATION										
15. Continuity	Interrupted									
16. Rate	Fast									
	Slow									
F. OTHER FEATURES										
17. Respiratory support	Adequate									
	Inadequate									
18. Diplophonia	Absent									
	Present									