

# *Avaliação impressionística e fonético-descritiva de qualidades de voz:*

*convergências, divergências e contextos de aplicação forense*

Rafael Cury Scarpelli, Renata Regina Passetti, Sandra Madureira

Laboratório Integrado de Análise Acústica e Cognição  
Pontifícia Universidade Católica de São Paulo (PUC-SP)  
São Paulo, Brasil

rafacuryscarpelli@gmail.com, re.passetti@gmail.com, sandra.madureira.liaac@gmail.com

**Resumo:** Este trabalho tem como objetivo analisar os rótulos de qualidade de voz atribuídos por leigos e investigar como podem ser correlacionados a termos técnicos de descrição da qualidade de voz. A avaliação perceptiva dos áudios foi realizada em três etapas: (1) descrição das características das vozes de locutores por ouvintes leigos; (2) identificação das vozes por ouvintes treinados com base nos qualificadores fornecidos pelos ouvintes leigos; e (3) avaliação das vozes por especialistas no uso do *Vocal Profile Analysis*. Os resultados permitiram estabelecer correlações entre julgamentos impressionísticos e foneticamente embasados e apresentam desdobramentos para situações fonético-forenses envolvendo testemunhas auriculares.

*Palavras-chave-* Comparação de locutor; Fonética forense; Percepção de fala; Qualidade e dinâmica de voz; *Vocal Profile Analysis*.

## I. INTRODUÇÃO

A partir da fala de um indivíduo, os ouvintes são capazes de fazer uma série de inferências sobre a sua identidade, o seu estado emocional, sua faixa etária e seu local de origem, entre outras características [1] [2] [3]. Também podem atribuir às vozes ouvidas rótulos semânticos impressionísticos. Sobre esse tipo de rotulação de vozes se encontra na literatura o estudo de Watt e Burns [4] que demonstra haver convergências no emprego de descritores semânticos por leigos e especialistas em voz. É importante ressaltar, contudo, que essas relações entre descritores e características perceptivas de voz se baseiam em associações entre som e sentido e, conforme argumenta Tsur [5], a natureza de tais relações é *double-edged*, isto é, deriva do fato de que uma característica acústica, um ruído de fricativa, por exemplo, pode ser associado a efeitos de sentido diferenciados, quais sejam, à continuidade ou à aspereza.

Resultados de estudos sobre as características que qualificam as vozes de locutores trazem contribuições ao exercício da perícia vocal, campo de trabalho que se ocupa de questões relacionadas à comparação de locutores e que tem na fonética forense o apoio teórico [6] [7] [8].

Ao levantar as informações de um crime em um trabalho de perícia vocal, os investigadores podem se defrontar com relatos de ouvintes leigos, as testemunhas auriculares do crime, que fornecem um relato das circunstâncias do crime ou da voz do criminoso. Nessas situações, em alguns países, é comum o recurso à utilização de filas de reconhecimento (*line-ups*) pela perícia vocal. Filas de reconhecimento compreendem, além da amostra de fala do suspeito de haver cometido o crime, pelo menos cinco vozes distratoras, ou seja, de outros locutores não envolvidos no crime. Esse procedimento de verificação da voz de suspeitos tem como objetivo determinar se alguma das vozes é compatível com a memória e, conseqüentemente, com a descrição que a testemunha auricular fez da voz do criminoso [7]. Diante desse contexto de verificação de compatibilidade, as características de voz fornecidas por essas testemunhas podem ser de utilidade para a construção de uma fila de reconhecimento [4]. Entretanto, conforme apontam Kreiman et al [9] uma questão se coloca: qual seria o grau de confiabilidade dos descritores dados por testemunhas auriculares? Essa é uma questão de relevância para o contexto forense e nosso estudo busca explorá-la, confrontando termos atribuídos por leigos e por especialistas em qualidade de voz.

Desta forma, a descrição de vozes realizada por ouvintes leigos constitui o objeto de análise do presente trabalho. Nosso objetivo consiste em analisar os rótulos de qualidade de voz atribuídos por ouvintes leigos e investigar como podem ser relacionados a termos técnicos de descrição da qualidade de voz. Por meio de filas de reconhecimento contendo vozes-alvo e vozes distratoras, verificaremos o quão consistentes e reveladores esses rótulos são para a identificação de uma voz-alvo por um ouvinte treinado. Pretendemos, por meio de nosso estudo, contribuir para questões que envolvem testemunhas auriculares em contextos fonético-forenses reais.

Assim, como questão de pesquisa postulamos: descrições verbais de qualidade de voz por ouvintes leigos podem funcionar como bons indicadores para elaboração de um perfil vocal? Como hipótese de pesquisa, pressupomos que haverá pareamentos entre termos impressionísticos que remetem a sensações de altura melódica, taxa de elocução, extensão, contração muscular, tensão muscular ou irregularidades fonatórias.

## II. FUNDAMENTAÇÃO TEÓRICA

### A. Definições de voz e de qualidade de voz

O termo “voz” pode ser empregado para designar o processo fonatório de vibração das pregas vocais, o vozeamento. Na produção dos sons da fala, a corrente de ar pode ser modificada pela vibração das pregas vocais, originando os sons vozeados que se opõem aos não vozeados que são produzidos sem vibração das pregas vocais.

O termo “qualidade de voz”, por sua vez, é comumente utilizado em dois sentidos. Em sentido estrito, compreende os modos de fonação, ou seja, as diferentes formas de vibração das pregas vocais que são determinadas pelas configurações laríngeas. Em sentido amplo, compreende não apenas os modos de fonação, mas também aspectos de configuração e de tensão do trato vocal [10].

Descrever a qualidade de voz de um falante como produção única e exclusiva das pregas vocais é limitante, pois não permite contemplar as demais características que possibilitam aos ouvintes perceber as produções de fala como pertencentes a um determinado falante.

Neste trabalho adotamos a noção de qualidade de voz concebida em sentido amplo, segundo a definição proposta por Laver [11] que a descreve como resultado de uma combinação de ajustes (*settings*) de voz de curto e longo termo que caracterizam a fala de um indivíduo. Esses ajustes compreendem configurações e tensões no trato vocal e na parte laríngea. Nesse sentido, o perfil vocal de um indivíduo é concebido como fruto da combinação de componentes, os diversos ajustes que caracterizam a sua voz [12].

Tendo essa conceituação em mente, a utilização de um protocolo que avalie a qualidade de voz de maneira componencial faz-se necessária. Para a finalidade do presente trabalho, utilizamos o *Vocal Profile Analysis* [13], doravante referido como VPA, que se baseia no modelo de descrição fonética de qualidade de voz, desenvolvido por Laver [11].

## III. METODOLOGIA

A metodologia empregada na realização do estudo é de natureza experimental de carácter descritivo, com foco na análise fonético-perceptiva. A seguir expomos o material, o método e a descrição dos experimentos de percepção efetuados.

### B. *Material*

Todas as amostras de fala utilizadas neste trabalho foram extraídas de *podcasts* disponibilizados na plataforma *Youtube*<sup>1</sup>. Os locutores selecionados são profissionais de comunicação. Por utilizar gravações de acesso aberto, o estudo não foi submetido ao Comitê de Ética em Pesquisa.

Foram selecionados trechos de fala com duração de 30 segundos de oito vozes masculinas (VM1 à VM8) e oito vozes femininas (VF1 à VF8). Com a intenção de manter o ouvinte focado nas características de qualidade da voz e não no conteúdo dito, trechos que pudessem identificar os falantes foram evitados.

### C. *Método*

O estudo compreendeu a avaliação perceptiva de amostras da fala realizada em três etapas: (1) descrição das características das vozes de locutores por ouvintes leigos; (2) identificação das vozes por ouvintes treinados (peritos em voz e profissionais com experiência em análise de voz) com base nos qualificadores fornecidos pelos ouvintes leigos; e (3) avaliação das vozes por especialistas no uso do VPA.

Primeiramente, as amostras de fala de indivíduos dos sexos masculino e feminino foram descritas por ouvintes leigos, por meio de até três adjetivos que julgassem serem apropriados para caracterizar as vozes. Posteriormente, os três adjetivos mais frequentes para cada voz-alvo foram agrupados e associados a filas de reconhecimento com amostras de fala de quatro indivíduos (voz-alvo + três vozes distratoras) [14].

As filas de reconhecimento e o conjunto de adjetivos associado a cada uma delas foram enviados para ouvintes treinados, os quais foram instruídos a selecionar a voz-alvo, ou seja, aquela que acreditavam estar mais próxima da descrição a partir dos adjetivos fornecidos.

O sistema VPA foi preenchido, por três juízes com experiência no sistema, para cada uma das 16 vozes. Apresentam-se nas Figuras 1 e 2, respectivamente, o VPA preenchido com os ajustes das vozes femininas e o VPA com os ajustes das vozes masculinas.

---

<sup>1</sup> Plataforma de vídeos da internet - <https://www.youtube.com/>

	Neutral	Non-neutral	SETTING	moderate		
				1	2	3
<b>A. VOCAL TRACT FEATURES</b>						
1. Labial			Lip rounding	VF7		
			Lip spreading	VF6		
			Labiodentalization			
			Extensive range	VF1 VF3 VF5	VF2	
			Minimised range	VF8		
2. Mandibular			Close jaw	VF8		
			Open jaw			
			Protruded jaw			
			Extensive range	VF1 VF5	VF2 VF3	
			Minimised range			
3. Lingual tip/blade			tip/blade			
			Retracted	VF6 VF8		
4. Lingual body			Fronted tongue body	VF4 VF7		
			Backed tongue body	VF1 VF2		
			Raised tongue			
			Lowered tongue body	VF1 VF3 VF7		
			Extensive range	VF2		
			Minimised range		VF7	
5. Pharyngeal			Pharyngeal constriction			
			Pharyngeal expansion	VF1		
6. Velopharyngeal			Audible nasal escape			
			Nasal			
			Denasal			
7. Larynx height			Raised larynx	VF5 VF6 VF8		
			Lowered larynx	VF1		
<b>B. OVERALL MUSCULAR TENSION</b>						
8. Vocal tract tension			Tense vocal tract	VF2 VF7		
			Lax vocal tract			
9. Laryngeal tension			Tense larynx	VF2 VF4 VF5		
			Lax larynx			
<b>C. PHONATION FEATURES</b>						
			Present		Moderate	
	SETTING	Neutral	Non-neutral	1	2	3
10. Voicing type	Voice					
	Falsetto					
	Creak					
	Creaky		VF2i VF4i VF7i	VF6		
11. Laryngeal frication	Whisper					
	Whispery			VF8	VF5	
12. Laryngeal irregularity	Harsh			VF2 VF4	VF7	
	Tremor					

				moderate		
				1	2	3
<b>D. PROSODIC FEATURES</b>						
13. Pitch	Mean	High			VF5	
		Low	VF1 VF2 VF4 VF7			
	Range	Extensive range	VF3 VF4		VF5	VF6
		Minimised range				
	Variability	High	VF1 VF3		VF6	
Low		VF8				
14. Loudness	Mean	High	VF1 VF3 VF4 VF5 VF7		VF2	
		Low	VF8			
	Range	Extensive range	VF2 VF6 VF8			
		Minimised range				
	Variability	High	VF2 VF4 VF5i VF6			
Low						
<b>E. TEMPORAL ORGANIZATION</b>						
15. Continuity		Interrupted	VF1 VF2			
16. Rate		Fast	VF1 VF4 VF7	VF5	VF3	
		Slow	VF8			
<b>F. OTHER FEATURES</b>						
17. Respiratory support		Adequate				
		Inadequate				
18. Diplophonia		Absent				
		Present				

Figura 1. Protocolo VPA preenchido com os ajustes das vozes femininas. Obs: ajustes com “i” são intermitentes

	FIRST PASS		SETTING	SECOND PASS		
	Neutral	Non-neutral		moderate		
				1	2	3
<b>A. VOCAL TRACT FEATURES</b>						
1. Labial			Lip rounding			
			Lip spreading	VM6		
			Labiodentalization			
			Extensive range	VM1 VM4 VM7 VM8		
			Minimised range	VM3 VM6		
2. Mandibular			Close jaw	VM3		
			Open jaw	VM8		
			Protruded jaw			
			Extensive range	VM1 VM5 VM8		
3. Lingual tip/blade			Minimised range	VM3		
			Advanced tip/blade			
			Retracted tip/blade			
4. Lingual body			Fronted tongue body	VM2 VM4 VM8		
			Backed tongue body	VM3 VM7		
			Raised tongue body			
			Lowered tongue body	VM1 VM4 VM6 VM7		
			Extensive range			
			Minimised range	VM6		
5. Pharyngeal			Pharyngeal constriction	VM5		
			Pharyngeal expansion			
6. Velopharyngeal			Audible nasal escape			
			Nasal			
			Denasal			
7. Larynx height			Raised larynx	VM5 VM8		
			Lowered larynx		VM4	VM3
<b>B. OVERALL MUSCULAR TENSION</b>						
8. Vocal tract tension			Tense vocal tract			
			Lax vocal tract			
9. Laryngeal tension			Tense larynx	VM7	VM5	
			Lax larynx			
<b>C. PHONATION FEATURES</b>						
				Present	Moderate	
	SETTING		Neutral	Non-neutral	1	2
10. Voicing type	Voice					
	Falsetto					
	Creak					
	Creaky			VM2i	VM4 VM6	VM3
11. Laryngeal frication	Whisper					
	Whispery				VM2 VM6 VM7	VM1
12. Laryngeal irregularity	Harsh				VM2 VM5	
	Tremor					

			moderate		
			1	2	3
<b>D. PROSODIC FEATURES</b>					
13. Pitch	Mean	High	VM2 VM6 VM7	VM5	
		Low	VM4	VM3	
	Range	Extensive range	VM1 VM4 VM7	VM5	
		Minimised range	VM8		
	Variability	High	VM6 VM7	VM5	
		Low			
14. Loudness	Mean	High	VM4 VM5		
		Low	VM1 VM2		
	Range	Extensive range	VM1 VM5		
		Minimised range			
	Variability	High	VM5 VM6 VM8		
		Low	VM2		
<b>E. TEMPORAL ORGANIZATION</b>					
15. Continuity		Interrupted	VM5		VM3
16. Rate	Fast		VM1i VM4 VM6i	VM7	
	Slow		VM8	VM3	
<b>F. OTHER FEATURES</b>					
17. Respiratory support		Adequate			
		Inadequate			
18. Diplophonia		Absent			
		Present			

Figura 2. Protocolo VPA preenchido com os ajustes das vozes masculinas. Obs: ajustes com “i” são intermitentes

#### D. Experimento de percepção I: ouvintes leigos

Enviamos para ouvintes leigos um formulário *online* contendo o conjunto de vozes e pedimos para as que descrevessem com até três adjetivos, conforme demonstrado na Figura 3.

Vozes femininas

A seguir são apresentadas 8 (oito) vozes femininas. Lembre-se que é de extrema importância que se mantenha atento(a) apenas à impressão que as características de cada uma dessas vozes lhe provoca, e deixe de lado aspectos como sotaques, gírias, assuntos abordados etc.

Também se lembre de que não precisa se manter preso a termos “técnicos”. Pode usar, por exemplo: preguiçosa, fraca, forte, abafada, acelerada, suave, ágil, irritante, insegura, trepidante, esganiçada etc.

Primeira voz feminina

**ÁUDIO 1 FEMININO**

Insira aqui até 3 adjetivos para descrever a primeira voz feminina. \*

Sua resposta

Figura 3. Modelo da apresentação da voz no formulário

Nesta etapa participaram 30 indivíduos, 10 do sexo masculino e 20 do sexo feminino. A idade dos participantes do sexo masculino variou de 23 a 58 anos (média de 36,7; desvio-padrão de 13,8 anos). A idade dos participantes do sexo feminino variou de 23 a 78 anos (média de 42,2 desvio-padrão de 16,8 anos). Todos os participantes tinham curso superior concluído ou em andamento. Nossos critérios de inclusão nesta etapa foram por escolaridade e ausência de queixas auditivas<sup>2</sup>.

#### E. Experimento de percepção II: ouvintes treinados

Analisamos os dados provenientes do primeiro experimento de percepção e selecionamos os três adjetivos mais representativos de cada voz, conforme apresentado nos Quadros 1 e 2.

QUADRO 1. ADJETIVOS REPRESENTATIVOS DAS VOZES FEMININAS

	VF1	VF2	VF3	VF4	VF5	VF6	VF7	VF8
<b>1° adjetivo</b>	Forte	Forte	Rápida	Rouca	Aguda	Rouca	Grave	Calma
<b>2° adjetivo</b>	Grave	Rouca	Forte	Forte	Suave	Forte	Abafada	Suave
<b>3° adjetivo</b>	Firme	Rápida	Firme	Abafada	Rápida	Esganiçada	Forte	Insegura

QUADRO 2. ADJETIVOS REPRESENTATIVOS DAS VOZES MASCULINAS

	VM1	VM2	VM3	VM4	VM5	VM6	VM7	VM8
<b>1° adjetivo</b>	Rouca	Rouca	Insegura	Rápida	Esganiçada	Rápida	Rápida	Rápida
<b>2° adjetivo</b>	Rápida	Fraca	Grave	Forte	Irritante	Suave	Abafada	Segura
<b>3° adjetivo</b>	Abafada	Abafada	Forte	Grave	Aguda	Abafada	Jovem	Suave

Realizada a caracterização das vozes com os adjetivos mais representativos, enviamos um novo formulário, desta vez para ouvintes treinados, no qual constavam 16 filas de reconhecimento com cada uma das vozes na posição de alvo.

As amostras de fala (voz-alvo e distratoras) foram posicionadas em ordem aleatória nas filas de reconhecimento, a fim de que o ouvinte não fosse influenciado por fator algum ao escolher o alvo. Além disso, separamos as vozes masculinas e femininas. Em seguida, subdividimos as vozes de falantes do mesmo sexo em dois grupos de quatro vozes. Por fim, utilizamos o *script* “Create Lineup” [15] para a montagem das filas de reconhecimento. Por meio deste *script*, as vozes eram organizadas de forma aleatória e separadas por meio de um intervalo de silêncio de 5 segundos, o qual também era adicionado no início e no fim de cada arquivo.

Nesta etapa, a tarefa dos ouvintes treinados era escolher a voz descrita pelos adjetivos. As filas de reconhecimento foram enumeradas em L1, L2, L3 e assim por diante, conforme consta no Quadro 3.

<sup>2</sup> Embora a faixa etária não tenha constituído critério de inclusão ou exclusão neste estudo, tendo em vista o objetivo de abarcar uma população potencial de testemunhas auriculares, em futuros estudos pretendemos verificar se há diferenças de julgamentos entre faixas etárias.



QUADRO 3. RELAÇÃO DAS FILAS DE RECONHECIMENTO E SEUS RESPECTIVOS ALVOS (EM NEGRITO)

L1	L2	L3	L4	L5	L6	L7	L8
VM4	VM7	VM2	VM1	VM2	<b>VM7</b>	VM3	VM7
<b>VM2</b>	<b>VM1</b>	<b>VM3</b>	<b>VM6</b>	<b>VM4</b>	VM1	VM2	VM6
VM5	VM8	VM5	VM7	VM5	VM6	VM4	VM1
VM3	VM6	VM4	VM8	VM3	VM8	<b>VM5</b>	<b>VM8</b>
L9	L10	L11	L12	L13	L14	L15	L16
VF3	<b>VF1</b>	VF2	<b>VF4</b>	VF3	<b>VF5</b>	VF2	VF1
VF7	VF6	VF7	VF1	<b>VF7</b>	VF4	<b>VF6</b>	VF5
<b>VF2</b>	VF4	<b>VF3</b>	VF5	VF8	VF1	VF7	VF4
VF8	VF5	VF8	VF6	VF2	VF6	VF3	<b>VF6</b>

Participaram deste teste de percepção 13 ouvintes treinados, sendo quatro do sexo masculino e nove do sexo feminino. A idade dos participantes do sexo masculino variou de 36 a 66 anos (média de 48,77 e desvio-padrão de 8,52 anos), enquanto a das participantes do sexo feminino variou de 37 a 60 anos (média de 49,83 e desvio-padrão de 7,95 anos). Os participantes possuíam graduação em diversas áreas, sendo a maioria fonoaudiólogos, mas todos atuavam na área de perícia de voz. Nossos critérios de inclusão nesta etapa foram por escolaridade, ausência de queixas auditivas e pela experiência em análise de vozes.

Dos 13 participantes que preencheram o formulário com as filas de reconhecimento com vozes masculinas, apenas 12 preencheram as filas de reconhecimento com vozes femininas.

#### IV. RESULTADOS E DISCUSSÃO

Os resultados obtidos revelaram relação entre termos leigos e técnicos. Os resultados das relações foram categorizados em tabelas, assim como a porcentagem de acerto das filas de reconhecimento, que foram significativamente assertivas.

##### *F. Análise das vozes por ouvintes leigos*

Após identificarmos os adjetivos mais frequentes atribuídos pelos ouvintes leigos, reunimos de acordo com a carga semântica aqueles que caracterizavam da mesma maneira uma voz. Em seguida, elegemos um dos adjetivos de cada grupo, geralmente o avaliado com maior frequência, como o hiperônimo do grupo, conforme apresentado no Quadro 4.

QUADRO 4. RELAÇÃO ENTRE ADJETIVOS E SUA CARGA SEMÂNTICA

Adjetivos	Campo Semântico
Forte, Segura, Direta, Firme, Encorpada, Seca	Forte
Fraca, Insegura	Fraca
Ágil, Acelerada, Rápida, Agitada, Apressada	Ágil

Preguiçosa, Calma, Lenta, Cansada, Tranquila	Preguiçosa
Esganiçada, Estridente, Trepidante	Esganiçada
Rouca	Rouca
Abafada, Nasalada/Nasal	Abafada
Jovem, Jovial, Juvenil	Jovem
Velha	Velha
Confusa, Hesitante	Confusa
Irritante, Chata	Irritante
Oscilante, Inconstante	Oscilante
Suave, Agradável, Boa	Agradável
Boa dicção, Limpa, Clara	Limpa
Aguda, Fina	Aguda
Grave, Grossa	Grave

### G. Relação entre descritores leigos e termos técnicos

Após fazer esse agrupamento, buscamos a correspondência entre os adjetivos mais frequentemente atribuídos pelos ouvintes leigos e os termos técnicos, referentes aos *settings* que compõem o sistema VPA. Conforme consta no Quadro 5.

QUADRO 5. RELAÇÃO ENTRE ADJETIVOS E SUA CARGA SEMÂNTICA

Adjetivos	Settings do VPA
Abafada	Low Mean Pitch, Lowered Tongue Body & Backed Tongue Body
Forte	Extensive labial Range, Extensive Mandibular Range & High Mean Loudness
Ágil	Fast Rate of Temporal Organization
Preguiçosa	Slow Rate of Temporal Organization

Esganiçada	Tense Vocal Tract, High Mean Pitch, High Pitch Variability, High Loudness Variability, Harsh & Creaky Voice
Rouca	Creaky, Harsh e Whispery Voice
Jovem	Raised Larynx, High Mean Pitch & Whispery Voice
Confusa	Interrupted Temporal Organization
Irritante	Creaky Voice, Harsh Voice, Interrupted Temporal Organization, Tense Larynx, Pharyngeal constriction, e High Pitch Variability, High Loudness Variability, Extensive Pitch Range & Loudness Extensive Range
Limpa	Uninterrupted Temporal Organization
Grave	Lowered Larynx & Low Mean Pitch
Aguda	High Mean Pitch
Oscilante e Inconstante	Presença intermitente de um ajuste perceptível (Creaky Voice), High Pitch Variability & High Loudness Variability

Em suma, as associações não são exclusivamente mútuas. O perito criminal, ao exercer a perícia de voz deve se atentar a essas dificuldades, uma vez que diferentes descritores leigos podem se referir ao mesmo *setting*. O campo semântico “forte” foi associado a uma fala ininterrupta e com enunciação clara. O campo semântico “esganiçada”, por sua vez, faz referência a variações de *pitch* e *loudness* na voz e a ajustes de *creaky voice*, *harsh voice* e de tensão no trato vocal. Os demais descritores foram associados de maneira similar.

#### H. Análise das vozes por ouvintes treinados

As associações entre adjetivos e vozes resultantes do segundo experimento perceptivo foram, em sua maioria, assertivas. No entanto, algumas possuíram uma taxa de erro maior do que a de acerto. As porcentagens de escolha das vozes de cada fila de reconhecimento estão apresentadas nas Tabelas 1 e 2.

TABELA 1. PORCENTAGENS DE ESCOLHA DE CADA VOZ NAS FILAS DE RECONHECIMENTO DE VOZES FEMININAS

<b>Filas de Reconhecimento</b>	<b>Voz 1</b>	<b>Voz 2</b>	<b>Voz 3</b>	<b>Voz 4</b>	<b>Voz-alvo</b>
<b>1ª fila</b>	NA	58,30%	41,70%	NA	<b>Voz 3</b>
<b>2ª fila</b>	58,30%	NA	41,70%	NA	<b>Voz 1</b>
<b>3ª fila</b>	16,70%	NA	75,00%	8,30%	<b>Voz 3</b>
<b>4ª fila</b>	91,70%	NA	8,30%	NA	<b>Voz 1</b>
<b>5ª fila</b>	8,30%	66,70%	NA	25,00%	<b>Voz 2</b>

<b>6ª fila</b>	83,30%	NA	NA	16,70%	<b>Voz 1</b>
<b>7ª fila</b>	NA	100,00%	NA	NA	<b>Voz 2</b>
<b>8ª fila</b>	NA	58,30%	16,70%	25,00%	<b>Voz 4</b>

NA - Não avaliado

TABELA 2. PORCENTAGENS DE ESCOLHA DE CADA VOZ NAS FILAS DE RECONHECIMENTO DE VOZES MASCULINAS

<b>Filas de Reconhecimento</b>					
<b>VM</b>	<b>Voz 1</b>	<b>Voz 2</b>	<b>Voz 3</b>	<b>Voz 4</b>	<b>Voz-alvo</b>
<b>1ª fila</b>	NA	46,20%	30,80%	23,10%	<b>Voz 2</b>
<b>2ª fila</b>	15,40%	76,90%	NA	7,70%	<b>Voz 2</b>
<b>3ª fila</b>	NA	100%	NA	NA	<b>Voz 2</b>
<b>4ª fila</b>	23,10%	30,80%	23,10%	23,10%	<b>Voz 2</b>
<b>5ª fila</b>	NA	100%	NA	NA	<b>Voz 2</b>
<b>6ª fila</b>	38,50%	NA	46,20%	15,40%	<b>Voz 1</b>
<b>7ª fila</b>	NA	7,70%	NA	92,30%	<b>Voz 4</b>
<b>8ª fila</b>	7,70%	15,40%	7,70%	69,20%	<b>Voz 4</b>

NA - Não avaliado

Na 1ª Fila de Reconhecimento VF (L9 no Quadro 3), por exemplo, a Voz 2 (VF7) foi mais selecionada que a voz-alvo (VF2). A razão pode estar relacionada ao fato de que são vozes com características similares. No entanto, na 5ª Fila de Reconhecimento VF (L13 no Quadro 3), contendo as mesmas vozes, mas em posições diferentes, a VF7 (voz-alvo) foi a mais selecionada, mesmo com a presença da VF2. Os adjetivos da voz-alvo na 1ª Fila de Reconhecimento VF foram “forte, rouca e rápida”, na 5ª foram “grave, abafada e forte”. Nota-se que ambas as vozes foram consideradas “fortes”, mas a presença do modificador “abafada” causou uma maior taxa de acertos, ao contrário do modificador “rouca”.

Um dado interessante e que talvez tenha provocado essa estranheza é que a VF2 foi predominantemente caracterizada como uma voz rouca, enquanto a VF7 não. Além disso, a VF7 possui um ajuste de *harsh voice* de grau 2, junto a um ajuste de *creaky voice* intermitente, enquanto a VF2 possui um ajuste de *harsh voice* de grau 1, também com um ajuste de *creaky voice* intermitente. Assim, a identificação de ajustes na VF2 que não foram identificados na VF7, pode ter ocorrido pela predominância de outras qualidades de voz na VF7.

Por motivos similares, outras filas provocaram divergências nas respostas dos ouvintes treinados. Geralmente, os descritores mais frequentes atribuídos pelos ouvintes leigos eram genéricos e não contribuíam de forma categórica para a identificação das vozes por ouvintes treinados. Isso foi o que ocorreu na 1ª e na 2ª filas de reconhecimento VF, e nas 1ª, 4ª e 6ª filas de reconhecimento VM.

Vozes femininas com descrições como “forte, grave e rouca” tiveram as maiores taxas de erro. No caso das vozes masculinas, as vozes descritas como “joviais” ou “fracas” foram as que tiveram a maior taxa de erro. Isso se deu possivelmente por causa da falta de descritores que um ouvinte leigo encontra a sua disposição para descrever uma voz com muitos ajustes neutros de qualidade de voz.

A 7ª Fila de Reconhecimento VF, a 3ª Fila de Reconhecimento VM e a 5ª Fila de Reconhecimento VM, tiveram 100% de acerto. Os descritores das vozes-alvo dessas filas de reconhecimento eram, respectivamente, “calma, suave e insegura”, “insegura, grave e forte” e “rápida, forte e grave”. Em todas, os descritores englobam de forma muito precisa e correta as principais percepções que poderiam se obter dessas vozes.

Outras tiveram uma taxa de acerto alta, mas não alcançaram os 100%. Esse foi o caso das 3ª, 4ª, 5ª, 6ª e 8ª filas de reconhecimento VF, e das 2ª, 7ª e 8ª filas de reconhecimento VM. Nestes casos, embora os adjetivos tenham sido mais específicos para as características salientes das vozes, nem todos os ouvintes treinados foram capazes de identificar as vozes referidas por eles.

## V. CONCLUSÃO

O ouvinte leigo tende a descrever sua percepção de uma voz com base nas suas próprias experiências. Assim, o que caracterizaria uma voz como “rouca” para um indivíduo, não necessariamente a caracterizaria da mesma forma para outro, por exemplo. A função do treinamento de um ouvinte na identificação de ajustes de qualidade de voz é associar a sua percepção auditiva a conceitos previamente estabelecidos na literatura científica. A falta desse embasamento teórico e do conhecimento de conceitos técnicos de voz é justamente o que causa as divergências entre os descritores dados pelos participantes do primeiro teste perceptivo.

Além disso, observa-se que não há uma relação biunívoca entre os sentidos dos termos leigos e os descritores técnicos, ou seja, vários descritores podem equivaler a um mesmo ajuste de qualidade de voz, o que também está de acordo com o argumento de *double edgedness* [5]. Além disso, quando os ouvintes treinados tiveram que identificar as vozes-alvo a partir dos adjetivos atribuídos pelos ouvintes leigos, foram justamente essas equivalências entre a carga semântica dos descritores que enviesaram a correta identificação das vozes-alvo. Assim, a dificuldade em relacionar os ajustes de qualidade de voz aos descritores de ouvintes leigos está associada ao fato de que os descritores não são precisos o suficiente ou nem sempre são adequados às vozes de um ponto de vista técnico.

Em suma, um perito, ao considerar os depoimentos de testemunhas auriculares de um crime, deve manter a cautela ao parear as descrições dadas para a voz do criminoso às vozes dos suspeitos. A equivalência entre os termos utilizados por ouvintes não treinados e os descritores de qualidade de voz não é absoluta e, por consequência, pode causar conflitos na investigação. No entanto, há de se considerar o potencial descritivo dos termos leigos, que muitas vezes podem ser de auxílio para o especialista.

Desta forma, este trabalho contribui para a categorização dos tipos de descritores leigos e suas respectivas correspondências terminológicas técnicas. Os resultados desta pesquisa podem auxiliar no pareamento de descritores leigos com termos técnicos, uma tarefa complexa e minuciosa, porém necessária para a área forense.

## REFERÊNCIAS

- [1] Barbosa PA, Madureira S, Passetti RR, Brescancini CR. Fonética, que bicho é esse? *Cadernos de Linguística* 2021; 2(1):1-19.
- [2] Tanuri MAF, Madureira S, de Souza Fontes MA, Caltabiano MA. A avaliação perceptiva de atributos paralinguísticos e extralinguísticos na fala de uma comunidade de prática: um experimento sociofonético perceptivo. *Cadernos de Estudos Linguísticos* 2021, 63: e021014-e021014.
- [3] Madureira S. Fala e expressividade. *Verbetes LBASS* 2020. Disponível em: <[http://www.letas.ufmg.br/padrao\\_cms/index.php?web=lbass&lang=1&page=3646&menu=&tipo=1](http://www.letas.ufmg.br/padrao_cms/index.php?web=lbass&lang=1&page=3646&menu=&tipo=1)>
- [4] Watt, D.; Burns, J. *Verbal descriptions of voice quality differences among untrained listeners*, York Papers in Linguistics Series, v. 2, p. 1-28, 2012.

- [5] Tsur, R. What Makes Sound Patterns Expressive? The Poetic Mode of Speech Perception. Durham, NC: Duke University Press, 1992.
- [6] Rose P. *Forensic Speaker Identification*. New York: Taylor & Francis 2003.
- [7] Eriksson A. Tutorial sobre Fonética Forense. *ReVEL* 2014, 12(23): 297-322. Tradução Pablo Arantes e Suska Gutzeit. [www.revel.inf.br].
- [8] Barbosa PA, Carrenho JM, Cazumba LAF, Constantini AC, Machado AP, Passetti RR, Sanches AP, Silva BA. Protocolo de análise fonético-forense. In: Barbosa et al. (Orgs.). *Análise Fonético-Forense: em tarefa de Comparação de Locutor*. 1ed. Campinas: Millennium Editora, 2020, pp. 3-15.
- [9] Kreiman, D.; Vanlancker-Sidtis; R. Gerratt, B. Perception of Voice Quality. In: PISONI, D. B.; REMEZ, R. E. *The Handbook of Speech Perception*. [S.l.]: Blackwell Publishing Ltd, 2005. Cap. 14, p. 338-362
- [10] Esling, J., Moisik, S., Benner, A., & Crevier-Buchman, L. (2019). Voice and Voice Quality. In *Voice Quality: The Laryngeal Articulator Model* (Cambridge Studies in Linguistics, pp. 1-36). Cambridge: Cambridge University Press. doi:10.1017/9781108696555.001
- [11] Laver J. *The phonetic description of voice quality*. Cambridge: Cambridge University Press 1980.
- [12] Laver, J. *Voice quality and indexical information*. Department of Phonetics and Linguistics, University of Edinburgh, 1968.
- [13] Laver J, Mackenzie-Beck J. *Vocal Profile Analysis Scheme-VPAS*. Queen Margaret University College-QMUC, Speech Science Research Centre, Edinburgh 2007.
- [14] Broeders T, Van Amelsvoort AG. Lineup Construction for Forensic Earwitness Identification: a Practical Approach. *Conference: XIVth International Congress of Phonetic Sciences*, San Francisco CA 1999, v. Proceedings, pp. 1373-1376.
- [15] Barbosa P. Create Lineup. [*Script para o Praaf*] 2021.