

Perícia em vozes auditivamente semelhantes

Forensic analysis of auditorily similar voices

Sandra Carmo¹ Maria Inês Beltrati Cornacchioni Rehder² Larissa Nadjara Almeida³ Cristian Villegas⁴ Cirana Raquel Vasconcelos Dantas⁵ Daniela Vasconcelos⁶ Emanuel Andrade⁷ 

¹ Consultório Particular; Instituto de Desenvolvimento Educacional, Cursos de Pós-graduação em Linguagem e Fonoaudiologia Educacional, Recife, Pernambuco, Brasil.

² CEFAC - Saúde e Educação, Cursos de Aprimoramento em Identificação Humana Forense e Especialização em Perícia Fonoaudiológica, São Paulo, São Paulo, Brasil.

³ Centro Universitário de João Pessoa – UNIPE, João Pessoa, Paraíba, Brasil.

⁴ Universidade de São Paulo – USP, Piracicaba, São Paulo, Brasil.

⁵ Consultório Particular, Colégio Diocesano de Caruaru, Caruaru, Pernambuco, Brasil.

⁶ Hospital das Clínicas da Universidade Federal de Pernambuco – UFPE, Recife, Pernambuco, Brasil.

⁷ Universidade de Pernambuco – UPE, Departamento de Perícias Forenses, Recife, Pernambuco, Brasil.

Estudo realizado no Departamento de Perícias Forenses da Universidade de Pernambuco - UPE, Recife, Pernambuco, Brasil.

Fonte de financiamento: Nada a declarar

Conflito de interesses: Inexistente

Endereço para correspondência:
Sandra Cristina Wanderley Carmo
Alameda das hortênsias 101,
bl 01 apto 703, Imbiribeira
CEP: 51160-400 - Recife, Pernambuco,
Brasil
E-mail: sandracarmow@yahoo.com.br

Recebido em: 03/11/2022

Aceito em: 31/03/2023

RESUMO

Objetivo: verificar contribuições da análise espectrográfica acústica na identificação forense de falantes em vozes auditivamente semelhantes, considerando o comportamento distintivo dos parâmetros acústicos: formantes da vogal “é”, da fala encadeada, média da frequência fundamental em Hz, curva de predição linear da vogal “é” e área da curva de predição linear; propor um método objetivo da utilização dos parâmetros analisados.

Métodos: estudo quantitativo, qualitativo e descritivo, realizado em Pernambuco com 16 pares de irmãos do sexo masculino, entre 18-60 anos. Os sujeitos gravaram vídeos de onde extraíram-se os áudios que foram numerados e enviados a três avaliadores, em dois grupos: dos irmãos mais velhos e dos irmãos mais novos, para pareamento perceptivo-auditivo. Os pareamentos corretos, apontados por pelo menos dois avaliadores, foram submetidos à análise acústica. Os testes estatísticos foram Wilcoxon, Kruskal-Wallis, Bonferroni, com $p < 0,05$.

Resultados: os resultados das análises dos formantes e da média da frequência fundamental não foram suficientes para distinguir as vozes semelhantes. Ineditamente nas medidas das áreas geradas pelos gráficos da curva de predição linear, foi verificada significância estatística distintiva.

Conclusão: concluiu-se que entre os parâmetros estudados, as medidas das áreas da curva de predição linear apontaram, objetivamente, eficácia na distinção de falantes com vozes auditivamente semelhantes.

Descritores: Acústica; Voz; Fala

ABSTRACT

Purpose: to verify contributions of acoustic spectrographic analysis in the forensic identification of speakers with auditorily similar voices, considering the distinctive behavior of acoustic parameters: formants of vowel “é”, of connected speech, mean fundamental frequency in Hz, linear prediction curve of vowel “é” and linear prediction curve area; and to propose an objective method to use the analyzed parameters.

Methods: a quantitative, qualitative and descriptive study, conducted in Pernambuco on 16 pairs of male siblings, aged 18-60 years. The subjects recorded videos from which the audios were extracted, numbered and sent to three examiners, in two groups: older brothers and younger brothers, for perceptual-auditory pairing. The correct pairings, indicated by at least two examiners, were submitted to acoustic analysis. The statistical tests included Wilcoxon, Kruskal-Wallis and Bonferroni, with $p < 0.05$.

Results: the results of analyses of formants and the mean fundamental frequency were not enough to distinguish similar voices. Unprecedentedly, in the measurements of areas generated by the linear prediction curve graphs, a distinctive statistical significance was observed.

Conclusion: it was concluded that, among the parameters studied, the measurements of areas of the linear prediction curve objectively indicated effectiveness in distinguishing speakers with auditorily similar voices.

Descriptors: Acoustics; Voice; Speech



Este é um artigo publicado em acesso aberto (Open Access) sob a licença Creative Commons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

INTRODUÇÃO

Na história antiga e contemporânea, há vários relatos de reconhecimento de pessoas por meio da voz, o mais emblemático foi o caso Lindberg em 1932. Por ser o reconhecimento vocal uma prova frágil, uma vez que se baseia em somente um dos sentidos de uma só pessoa, atualmente a proposta é de identificação de falantes, utilizando-se protocolos cientificamente embasados.

Os estudos estão em constante evolução e vários são os métodos utilizados para a identificação forense de falantes, na maioria dos casos. No Brasil, os métodos de identificação de voz iniciaram-se na perícia oficial na década de 1990, envolvendo peritos dos estados, da Polícia Federal e do Distrito Federal¹. A interceptação de comunicações telefônicas como meio de investigação e de prova no Processo Penal Brasileiro é um procedimento cada vez mais utilizado².

Para auxiliar e fundamentar a elaboração da prova pericial, conta-se com a Ciência Forense que é o conjunto de todos os conhecimentos científicos e técnicas que são utilizados para desvendar não só crimes, como também outros assuntos legais. Em relação às ciências, àquelas que estão diretamente envolvidas com a identificação forense de falantes com intuito de esclarecer demandas judiciais, encontram-se a Linguística Forense, a Fonética Forense e a Fonoaudiologia Forense, onde seus profissionais se dedicam na tarefa complexa de identificar um falante por meio da sua voz e da fala.

A Linguística Forense é um ramo da linguística aplicada que se dedica ao contexto investigativo que nos aponta para elementos que analisam a comunicação em seus diversos aspectos³. A Fonética Forense vai além da identificação de falantes, ela permeia muitos misteres criminalísticos. A Fonoaudiologia Forense tem como principal objetivo responder às demandas judiciais relacionadas à comunicação humana, atuando em várias análises que envolvam comparação forense de voz, fala e linguagem; a grafotécnica; a biometria facial; a transcrição, textualização e análise de conteúdo de áudio, vídeo e imagens e a descrição do perfil comunicativo¹.

Recentemente, em 22 de outubro de 2020, o Conselho Federal de Fonoaudiologia reconheceu a área de Perícia Fonoaudiológica por meio da resolução 584⁴.

Para a perícia de Identificação Forense de Falantes, faz-se necessário a comparação da amostra padrão com a amostra questionada⁵. Deve-se esclarecer que

amostra padrão é o registro do áudio no qual consta a fala do suspeito, indiciado ou réu (de identidade conhecida) e amostra questionada é o registro do áudio no qual consta a fala do locutor, cuja identidade se deseja conhecer⁶.

Os métodos utilizados por especialistas na área da perícia de identificação de falantes são três: o método perceptivo-auditivo, método acústico e o método automático⁷.

O método perceptivo auditivo destaca auditivamente os parâmetros a serem analisados e apresenta um forte aspecto subjetivo por meio de uma abordagem qualitativa⁸.

O método acústico é aquele que utiliza o espectrograma para analisar as ondas produzidas no momento da emissão vocal, permitindo uma análise quantitativa⁹. A avaliação por parâmetro acústico é importante de ser padronizada, uma vez que essa análise fornece um número¹⁰, o que facilita análises, comparações e armazenamento de medidas. O espectrograma gerado neste método é um gráfico tridimensional que registra a mensuração acústica da onda sonora. Ele contém informações relativas aos parâmetros do som, ou seja, intensidade, duração e frequência (tempo no eixo horizontal, frequência em Hertz no eixo vertical e intensidade em Decibel por meio do grau de coloração⁹).

De modo simplificado a avaliação acústica quantifica o sinal sonoro, o que nos leva em direção a uma análise objetiva da voz. E ainda se tem a seguinte distinção: enquanto a acústica realiza a mensuração do sinal sonoro, a avaliação perceptivo-auditiva oferece uma descrição do sinal vocal tendo como instrumento básico apenas a audição¹¹. A importância dos dois métodos propostos (perceptivo-auditivo e acústico) associados, além de confirmar que um não é melhor do que o outro, mas se complementam foi a conclusão de um recente estudo na Universidade de Pernambuco⁷.

O outro método, o automático, é realizado por *softwares* que tentam reduzir ao máximo as análises subjetivas. Os *softwares* são alimentados com informações como o vocabulário, programado e pronunciado de várias formas diferentes. Em alguns países da Europa, o uso de sistemas automáticos é acompanhado de *insights* de um profissional com conhecimentos em fonética e até linguística. Por exemplo, na Universidade de Gotemburgo, o *software* utilizado é o ALIZE SpkDet, e os resultados obtidos pelo programa são combinados com análise acústica e auditiva tradicional¹².

Os sistemas automáticos estão sujeitos às chamadas condições de incompatibilidade, que acontecem quando as diferenças entre as amostras das vozes também podem aparecer devido a diferenças nos canais de transmissão, sendo esse um problema relevante e preocupante nesse tipo de método de análise¹².

Todos os dispositivos legais e tecnológicos trazem suporte para a perícia forense de identificação/comparação de falantes, e nesta área são cada vez mais realizadas pesquisas, para que a comparação binária de vozes tenha escopo perante a justiça.

O objetivo geral desta pesquisa foi verificar as contribuições da análise espectrográfica acústica na identificação forense de falantes em vozes auditivamente semelhantes e propor um método objetivo da utilização dos parâmetros analisados. Os objetivos específicos foram verificar a utilidade dos parâmetros acústicos: formantes da vogal “é”, média da frequência fundamental em Hz, formantes F1, F2, F3 na fala, curva de predição linear (LPC) da vogal “é” e área da curva LPC para a distinção de vozes auditivamente semelhantes.

MÉTODOS

O estudo foi realizado no estado de Pernambuco e tem a aprovação do Comitê de Ética em Pesquisa da Fundação de Hematologia e Hemoterapia do Estado, Brasil, sob o número de parecer 4.303.659 e CAAE 38306620.3.0000.5195. As variáveis independentes foram naturalidade, idade, irmandade e sexo e as variáveis dependentes: quatro primeiros formantes da vogal “é” (representada por “/ε/”); média de frequência fundamental, F₁, F2, F3 na fala encadeada, LPC da vogal /ε/ e área da curva LPC.

Participaram deste estudo, 32 pessoas, sendo 16 pares, dois irmãos de cada família. Foram critérios de inclusão: serem irmãos (devido à genética), serem do sexo masculino (devido à proximidade da frequência vocal), estarem na faixa etária entre 18 e 60 anos (devido a, nesta faixa etária a voz não passar por mudanças significativas) e serem naturais e residentes do estado de Pernambuco (devido ao sotaque e especialmente à pronúncia da vogal “e”, marcante na região). Foram critérios de exclusão: serem irmãos gêmeos, considerando a existência de estudos anteriores em gêmeos, e/ou se estivessem com algum processo viral, bacteriano ou inflamatório nas vias aéreas superiores no dia da coleta, o que influenciaria na voz e possivelmente na distinção da

voz entre os pares, e/ou não terem assinado o Termo de Consentimento Livre e Esclarecido (TCLE).

A pesquisadora (S.C.W.C) fez o recrutamento dos participantes, de forma aleatória, enviando um convite especificamente desenhado para este fim, por meio de redes sociais e instituições do estado de Pernambuco. Definidos os participantes, após utilizados os critérios de inclusão e exclusão previamente descritos, procedeu-se a coleta de dados por meio de vídeo, captado pelo celular do participante por *software* do próprio aparelho. Os vídeos possuíam o seguinte roteiro de gravação, previamente explicado, para os participantes: dizer o nome, a data, mostrar um documento de identificação com foto e data de nascimento; falar sobre o estado de Pernambuco durante 3 a 5 minutos. Depois, os vídeos foram enviados para a pesquisadora. Para que fosse possível a realização da primeira etapa metodológica, audição das amostras de vozes, os vídeos foram convertidos em áudio, formato Wav, pela pesquisadora com o auxílio do programa de conversão multimídia Format Factory®. O preparo do material para a etapa de audição e pareamento das amostras de voz constituiu na formação de dois grupos GimV (Grupo dos Irmãos mais velhos) e GimN (Grupo dos irmãos mais novos). Em seguida, os nomes dos participantes do grupo (GimV) foram substituídos por números consecutivos de 1 a 16. No grupo dos irmãos mais novos (GimN), os nomes foram substituídos por números de 17 a 32 de forma randomizada. Realizado este procedimento, obtiveram-se dois grupos de amostras de vozes, GimV com numeração de 1 a 16 e GimN com numeração randomizada entre 17 e 32.

Para compor as amostras de vozes auditivamente semelhantes e que posteriormente foram investigadas por meio da análise espectrográfica acústica pela pesquisadora na segunda etapa, as amostras de vozes dos grupos GimV e GimN foram submetidas ao pareamento perceptivo-auditivo, realizado por três avaliadores fonoaudiólogos especialistas em Voz pelo Conselho Federal de Fonoaudiologia – CFFa. Aos fonoaudiólogos que realizaram o pareamento perceptivo-auditivo, foi solicitado que ouvissem as vozes do GimV e que indicassem o par do respectivo irmão no GimN e realizassem o registro de cada par por meio de uma tabela de pareamento (Quadro 1). Foram conduzidos à análise acústica, os pares de irmãos considerados auditivamente semelhantes de forma correta, pertencentes à mesma família, apontados como pares por pelo menos dois, dos três fonoaudiólogos avaliadores. Dos 16 pares submetidos ao pareamento

perceptivo-auditivo realizado pelos fonoaudiólogos, seis foram coincidentes e seguiram para a análise

acústica. O resultado do pareamento perceptivo-auditivo encontra-se no Quadro 1.

Quadro 1. Pareamento analítico perceptivo – auditivo realizado por fonoaudiólogos especialistas em voz pelo Conselho Federal de Fonoaudiologia

Avaliador Par	1	2	3
1 – 31	D	C	C
2 – 27	D	D	C
3 – 21	D	C	C
4 – 20	D	D	C
5 – 29	D	D	C
6 – 28	C	C	C
7 – 30	D	D	D
8 – 24	D	D	D
9 – 32	C	C	D
10 -25	D	C	C
11 -26	D	C	D
12 -23	D	C	D
13 -17	D	C	D
14 -19	C	C	D
15 -18	D	D	D
16 -22	D	C	D

Legenda: C = Coincidente; D = Divergente

Fonte: Carmo et al. (2021).

Na segunda etapa, as amostras pareadas corretamente foram analisadas por meio da análise espectrográfica acústica, buscando verificar se, e quais dos parâmetros acústicos analisados, teriam robustez estatística suficiente para distinguir pessoas da mesma família com vozes auditivamente semelhantes, e se, e quais parâmetros acústicos, eram coincidentes em pessoas nascidas e residentes no Estado de Pernambuco. As análises espectrográficas acústicas foram realizadas pela pesquisadora (S.C.W.C), com o auxílio do programa de análise acústica PRAAT®.

Neste estudo foram verificados parâmetros acústicos individuais e posteriormente comparados entre os irmãos do par, entre os pares e entre os dois grupos (GimV e GimN). Os parâmetros acústicos analisados foram os quatro primeiros formantes (F1, F2, F3, F4) da vogal /ε/, que foram extraídos após o primeiro minuto de fala; média da frequência fundamental na fala em Hz; F1, F2 e F3 na fala encadeada, que foram extraídos nos quatro primeiros minutos de fala; curva LPC por meio do programa PRAAT®. Também foi feita a análise da área da curva LPC: realizada a partir

dos gráficos das curvas LPC individuais geradas no PRAAT® com o objetivo de fornecer um método de análise proposto de forma inédita no presente estudo. O cálculo da área gerada pelo gráfico LPC comparativo de cada par estudado foi realizado por um profissional da área de Informática, que gerou um algoritmo especificamente para este fim. A curva LPC de cada áudio gerada no PRAAT® separadamente foi submetida a uma análise de sua área para assim obter-se medidas das áreas formadas abaixo das curvas, capazes de serem analisadas e comparadas intrapar na análise estatística.

Para obtenção desta área utilizou-se um algoritmo para geração dos gráficos e cálculo da integral (área sob a curva). Inicialmente, a imagem foi convertida de RGB para uma versão monocromática, e foram removidos os níveis intermediários de cinza, deixando-se apenas pixels totalmente brancos ou totalmente pretos.

Em seguida, foi feito um *loop*, variando a coordenada “y”, em princípio, desde a primeira até a última linha da figura. Como estava-se trabalhando

com figuras de resolução 3.600x2.400, isso significa variar “y” de 0 a 2.399; em cada interação do *loop* em “y”, foi feito outro *loop*, desta vez variando a coordenada “x”, em princípio, desde a primeira até a última coluna da figura, ou seja, isso significa variar “x” de 0 a 3.599. A ressalva “em princípio” é feita porque, durante a varredura, as cores dos pixels são testadas, e inicialmente todos são pixels brancos. Ao ser encontrado o primeiro pixel preto, terminou-se ambos os *loops*, pois soube-se ter encontrado a parte superior esquerda do gráfico, lembrando que o ponto de coordenadas (0,0) é o da primeira linha (mais de cima) e da primeira coluna (mais à esquerda). A partir do ponto imediatamente anterior a esse pixel encontrado, ou seja, das coordenadas (xpreto – 1, ypreto), onde as coordenadas (xpreto, ypreto) são as desse primeiro pixel preto encontrado, seguiu-se incrementando a coordenada “y”, tomando nota dos valores de “y” em que são encontradas variações de branco para preto, ou vice-versa. Como estava-se percorrendo a coluna imediatamente anterior à do eixo “y” do gráfico, estas variações se encontram nas marcações da escala do eixo “y” (0, 20, 40, e 60 dB/Hz, a depender do gráfico sendo analisado). Foi gerada, assim, a tabela T “y” Map, em que se registrou a coordenada “y” média entre a transição do branco para o preto e a transição seguinte do preto para o branco, assumindo que o valor da escala esteja exatamente na metade do traço de marcação. Essa tabela T “y” Map nos permite mapear as coordenadas “y” expressas em pixels na figura para seus respectivos valores em dB/Hz.

Em seguida, foi criada uma tabela análoga, T “x” Map, desta vez variando as coordenadas “x” desde o ponto (xpreto, ymarca_min), onde xpreto é a coordenada “x” do primeiro ponto preto encontrado acima, e ymarca_min é a coordenada “y” da marca de menor valor em dB/Hz do eixo “y”. Assim variando, anotou-se a coordenada “x” da primeira transição de preto para branco, xini, que caracteriza a primeira coluna da região do gráfico; e da última transição de branco para preto, xfim, caracterizando a última coluna dessa região. A tabela T “x” Map, assim criada, permitiu o mapeamento das coordenadas “x”, com xini → 0 dB, e xfim → 104 dB. Por fim, variou-se a coordenada “y”, de (xini, ymarca_min), incrementando o valor de “y”, ou seja, seguindo para baixo no gráfico, até encontrar

uma transição do branco para o preto, que se dará na coordenada ybaixo, onde se encontra o eixo dos “x”.

De forma análoga, variou-se novamente a coordenada “y”, dessa vez decrementando-a (ou seja, seguindo para cima), até encontrar a coordenada ycima, onde se encontra a moldura superior do gráfico. A partir daí, foi calculado o valor dx, definido como: $dx = xfim - xini / 104$, já que 104 é o valor final do eixo “x” em todos os gráficos, e o valor inicial é zero. Em seguida, foi inicializada uma variável integral com o valor zero, e iniciou-se um *loop* variando a coordenada “x”, em princípio, de xini a xfim, e a cada interação desse *loop* variou-se a coordenada “y”, em princípio, de ybaixo até ycima, ou seja, seguindo para cima, passando por pixels brancos, depois por pixels pretos (a linha do gráfico), e parando um pixel antes da transição de preto para branco, onde está o ponto do gráfico, na coordenada (xi, yf(xi)).

A cada vez que foi encontrado um ponto (xi, yf(xi)), converteram-se as coordenadas expressas em pixels para coordenadas expressas nas unidades do gráfico, utilizando as tabelas T “x” Map e T “y” Map. O valor yf(xi) é somado à variável integral, zerada no início do *loop* mais externo, para que seu valor ao final dos *loops* seja multiplicado pelo valor dx obtido acima, fornecendo o valor final da integral, ou seja, da área sob a curva.

Para a análise estatística, os resultados dos parâmetros acústicos analisados foram extraídos e alocados em planilha digital. Foram realizadas análises descritivas, por meio de medidas de tendência central, e inferenciais, por meio de testes de comparação não paramétricos, visto que os dados não atenderam o critério de normalidade. Foi utilizado o teste de Wilcoxon, para análise pareada entre os irmãos, e o teste de Kruskal-Wallis para comparação dos grupos de irmãos mais velhos e mais jovens e comparação entre os pares de irmãos, além do teste post hoc de Bonferroni para observar comparações múltiplas. Foi utilizado o *software* SPSS, versão 21, e considerada significância de 5% ($p < 0,05$).

RESULTADOS

A Tabela 1 mostra a comparação das medidas dos formantes da vogal /ε/ entre os irmãos mais velhos e mais novos de cada par.

Tabela 1. Comparação de cada medida acústica extraída referente aos formantes da vogal /ε/ entre os irmãos mais velhos e mais novos de cada par

Pares	F1 vogal /ε/	Valor de p	F ₂ vogal /ε/	Valor de p	F3 vogal /ε/	Valor de p	F4 vogal /ε/	Valor de p
Par 1-31	517,0	0,317	1861,0	0,317	2579,0	0,317	3997,0	0,317
	561,0		1787,0		2445,0		3867,0	
Par 3-21	466,0	0,317	1739,0	0,317	2869,0	0,317	3522,0	0,317
	336,0		1926,0		2647,0		3866,0	
Par 6-28	388,0	0,317	2042,0	0,317	2657,0	0,317	4077,0	0,317
	455,0		2059,0		2374,0		4115,0	
Par 9-32	470,0	0,317	2003,0	0,317	2701,0	0,317	3418,0	0,317
	479,0		1875,0		2598,0		3733,0	
Par10-25	498,0	0,233	1629,0	0,317	2338,0	0,317	3549,0	0,317
	335,0		1776,0		2675,0		3566,0	
Par14-19	529,0	0,327	1822,0	0,317	2499,0	0,317	3550,0	0,317
	471,0		1640,0		2372,0		4372,0	

Teste dos postos sinalizados de Wilcoxon; significância $p < 0,05^*$

Legenda:

F1 = Primeiro Formante;

F2 = Segundo Formante;

F3 = Terceiro Formante;

F4 = Quarto Formante.

Fonte Carmo et al. (2021)

As medidas acústicas extraídas da vogal /ε/ para F1, F2, F3 e F4 não exibiram diferenças estatisticamente significantes, como foi mostrado nos resultados expostos na Tabela 1.

A Tabela 2 mostra a comparação das medidas dos formantes, da média da frequência na fala encadeada entre os irmãos mais velhos e irmãos mais novos de cada par.

Tabela 2. Comparação de cada medida acústica extraída referente aos formantes da fala, da média de frequência da fala entre os irmãos mais velhos e mais novos do mesmo par

Variáveis	F1 fala	Valor de p	F2 fala	Valor de p	F3 fala	Valor de p	Média FO fala	Valor de p
Par 1-31	612,0	0,317	1694,0	0,317	2921,0	0,317	117,0	0,317
	593,0		1629,0		2761,0		113,0	
Par 3-21	539,0	0,317	1712,0	0,317	3081,0	0,317	138,0	0,317
	686,0		195,0		2983,0		115,0	
Par 6-28	672,0	0,317	1803,0	0,317	2854,0	0,317	123,0	0,317
	662,0		1785,0		3987,0		104,0	
Par 9-32	565,0	0,317	1733,0	0,317	2823,0	0,317	152,0	0,317
	716,0		1845,0		2908,0		159,0	
Par 10-25	759,0	0,317	1904,0	0,317	2927,0	0,317	126,0	0,317
	750,0		2012,0		3163,0		135,0	
Par 14-19	714,0	0,317	1696,0	0,317	2948,0	0,317	126,0	0,317
	683,0		1700,0		2752,0		109,0	

Teste dos postos sinalizados de Wilcoxon; significância $p < 0,05^*$

Legenda:

FO = Frequência fundamental;

F1 = Primeiro Formante;

F2 = Segundo Formante;

F3 = Terceiro Formante.

Fonte: Carmo et al. (2021)

As medidas acústicas reunidas nessa tabela não possuem significância estatística.

Na Tabela 3 considerou-se a possibilidade de haver diferenças das medidas entre os pares, visto que estes

sujeitos não têm relação de parentesco, mas apenas a naturalidade em comum. Dessa forma, apresentou-se na Tabela 3, a comparação das medidas acústicas entre os pares.

Tabela 3. Comparação das médias gerais das medidas acústicas da voz entre os seis pares de irmãos mais velhos e mais novos

Variáveis	F1 vogal /ε/	F2 vogal /ε/	F3 vogal /ε/	F4 vogal /ε/	Média Fq Fala	F1 Fala	F2 Fala	F3 Fala
PAR 1-31	11,00	6,00	5,00	8,50	4,00	3,50	2,50	4,00
PAR 3-21	3,50	6,00	10,00	4,50	7,00	4,50	3,50	9,50
PAR 6-28	3,50	11,50	6,00	10,50	3,50	5,50	8,50	8,00
PAR 9-32	7,00	9,00	9,00	3,50	11,50	6,00	8,50	4,00
PAR 10-25	5,00	2,50	5,50	4,00	8,25	11,50	11,50	9,00
PAR 14-19	9,00	4,00	3,50	8,00	4,75	8,00	4,50	4,50
Valor de p	0,361	0,073	0,422	0,367	0,048*	0,265	0,381	0,572

Teste Kruskal-Wallis; significância $p < 0,05^*$

Legenda:

Fq = Frequência;

F1 = Primeiro Formante;

F2 = Segundo Formante;

F3 = Terceiro Formante;

F4 = Quarto Formante.

Fonte: Carmo et al. (2021)

No parâmetro de frequência entre os seis pares (Tabela 3), foi encontrada diferença estatisticamente significativa entre os pares, ou seja, mesmo sabendo-se que esse parâmetro possui média populacional, encontrou-se diferenças interpares.

Realizou-se então o teste de comparações múltiplas de Bonferroni, para observar onde ocorreram essas diferenças, como demonstram no Quadro 2. Considerando-se que tais diferenças podem contribuir para a perícia de identificação forense de falantes de forma geral.

Quadro 2. Teste post hoc, para comparações múltiplas entre médias gerais das medidas de frequência dos seis pares de irmãos mais velhos e mais novos

Variável dependente	(I) PAR	(J) PAR	Diferença média (I-J)	Erro Padrão	Sig.	Intervalo de Confiança 90%	
						Limite inferior	Limite superior
Fala Média da Fq	1-31	3-21	-11,5000	10,5079	1,000	-54,140	31,140
		6-28	1,5000	10,5079	1,000	-41,140	44,140
		9-32	-40,5000	10,5079	,126	-83,140	2,140
		10-25	-15,5000	10,5079	1,000	-58,140	27,140
		14-19	-2,5000	10,5079	1,000	-45,140	40,140
	3-21	1-31	11,5000	10,5079	1,000	-31,140	54,140
		6-28	13,0000	10,5079	1,000	-29,640	55,640
		9-32	-29,0000	10,5079	,493	-71,640	13,640
		10-25	-4,0000	10,5079	1,000	-46,640	38,640
		14-19	9,0000	10,5079	1,000	-33,640	51,640
	6-28	1-31	-1,5000	10,5079	1,000	-44,140	41,140
		3-21	-13,0000	10,5079	1,000	-55,640	29,640
		9-32	-42,0000	10,5079	,107	-84,640	,640
		10-25	-17,0000	10,5079	1,000	-59,640	25,640
		14-19	-4,0000	10,5079	1,000	-46,640	38,640
	9-32	1-31	40,5000	10,5079	,126	-2,140	83,140
		3-21	29,0000	10,5079	,493	-13,640	71,640
		6-28	42,0000	10,5079	,107	-,640	84,640
		10-25	25,0000	10,5079	,822	-17,640	67,640
		14-19	38,0000	10,5079	,167	-4,640	80,640
	10-25	1-31	15,5000	10,5079	1,000	-27,140	58,140
		3-21	4,0000	10,5079	1,000	-38,640	46,640
		6-28	17,0000	10,5079	1,000	-25,640	59,640
		9-32	-25,0000	10,5079	,822	-67,640	17,640
		14-19	13,0000	10,5079	1,000	-29,640	55,640
	14-19	1-31	2,5000	10,5079	1,000	-40,140	45,140
		3-21	-9,0000	10,5079	1,000	-51,640	33,640
		6-28	4,0000	10,5079	1,000	-38,640	46,640
		9-32	-38,0000	10,5079	,167	-80,640	4,640
		10-25	-13,0000	10,5079	1,000	-55,640	29,640

Teste: Bonferroni; significância $p < 0,05^*$

Legenda:

Sig. = Significância;

Fq = Frequência.

Fonte: Carmo et al. (2021)

Com essa análise não foi constatada nenhuma significância entre os pares em relação à frequência, ou seja, mesmo entre todos os pares não houve uma frequência que conseguisse destacar um par, ou mesmo uma voz, como foi visto anteriormente.

A Figura 1, apresenta seis imagens que representam a curva de LPC entre os pares, os áudios dos

irmãos nos gráficos são representados por curvas com cores diferentes.

As imagens a seguir demonstram as diferenças entre os áudios, pois as duas curvas resultantes são distintas, mesmo quando em alguns casos elas se sobrepõem ou mesmo se entrelaçam.

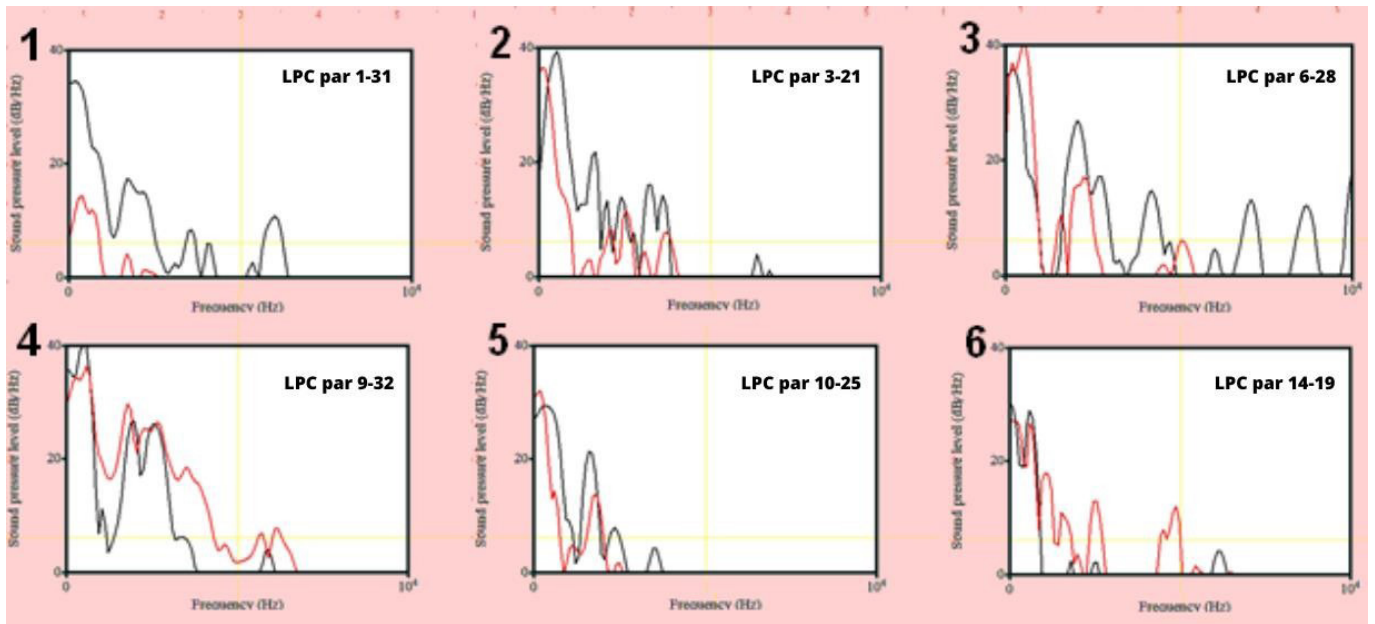


Figura 1. Curva de Predição Linear do mesmo par com cores diferentes para cada curva na mesma tela

Na presente pesquisa considerou-se o LPC no recorte da vogal /ε/, cujos resultados foram apresentados na Figura 1. A análise aplicada a um sinal de fala permite obter a envoltória espectral e as frequências correspondentes aos formantes.

A Figura 2 reúne 12 imagens com as medidas da área dos gráficos LPC.

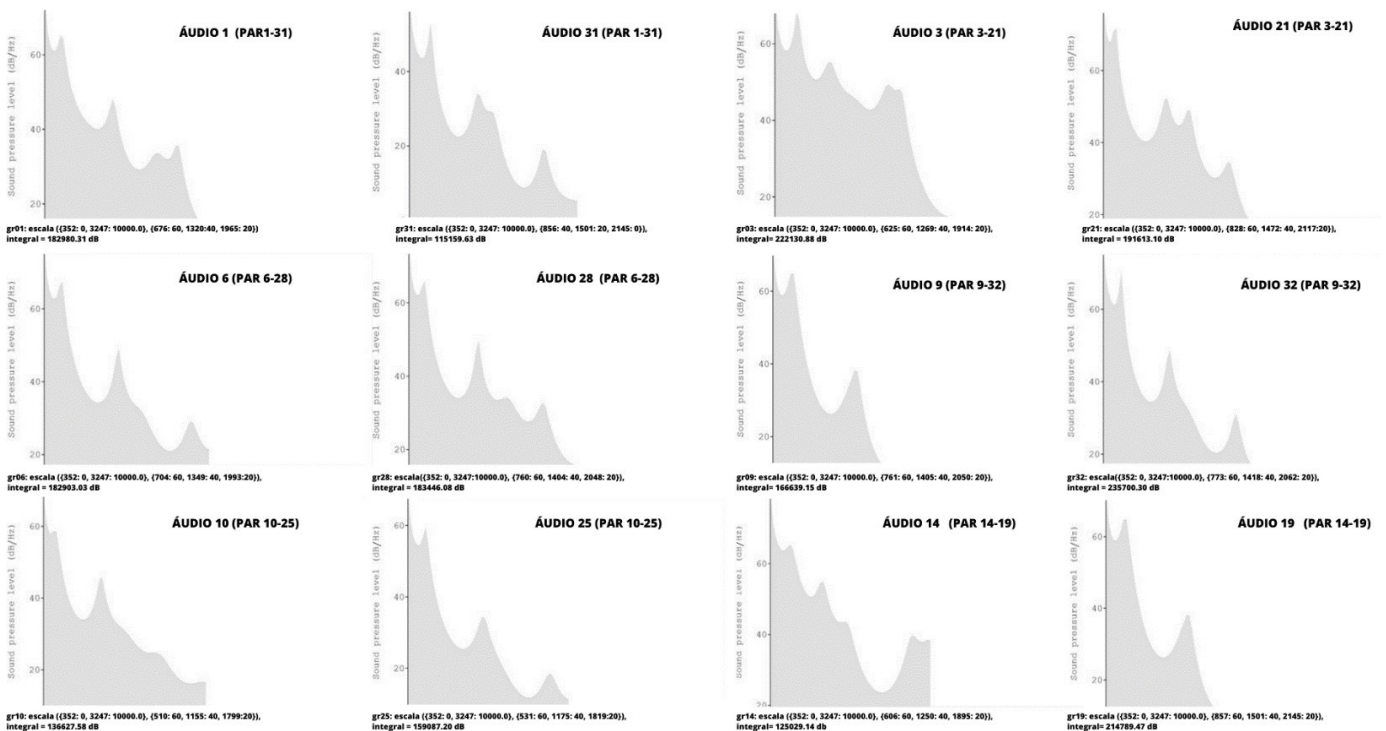


Figura 2. Medidas da área dos gráficos das Curvas de Predição Linear em gráficos distinto

A Tabela 4 compara as áreas das curvas de LPC e mostra que essa medida consegue distinguir como um

parâmetro objetivo mais de 50% dos irmãos com vozes auditivamente semelhantes.

Tabela 4. Comparação das medidas da área da Curva de Predição Linear da voz dos irmãos de cada par

Variáveis	Área LPC	Valor de p
PAR 10-25	136627,58 159087,20	0,317
PAR 14-19	125029,14 214789,47	0,0001*
PAR 6-28	182903,03 183446,08	0,817
PAR 1-31	182980,31 115159,63	0,0001*
PAR 3-21	222130,88 191613,10	0,039*
PAR 9-32	166639,15 235700,30	0,001*

Teste de Wilcoxon; significância $p < 0,05^*$

Legenda: LPC = Curva de Predição Linear.

Fonte: Carmo et al. (2021)

DISCUSSÃO

Como foi mostrado nos resultados referentes à comparação de cada medida acústica extraída, referente aos formantes da vogal /ε/ entre os irmãos mais velhos e mais novos de cada par, as medidas não foram capazes de diferenciar os irmãos nem mesmo no formante de alta frequência, dado que vem de encontro aos achados dos estudos abaixo citados.

Recente estudo¹³ revelou padrões consistentes em relação à comparação de formantes de alta e baixa frequência em pares de gêmeos e falantes não geneticamente relacionados, com os formantes de alta frequência exibindo um maior poder discriminatório do falante em comparação com formantes de baixa frequência. Vale a pena ressaltar que esse estudo foi realizado com pares de irmãos gêmeos (geneticamente relacionados) e com sujeitos não geneticamente relacionados.

Outro estudo¹⁴ demonstrou que os falantes, masculinos e femininos, produziram vogais com valores de F_1 e F_2 relativamente próximos aos alvos do falante nativo do estado da Paraíba (PB), os valores médios para falantes não nativos do sexo masculino mostraram-se quase idênticos às médias dos falantes nativos. Medidas formânticas são os principais correlatos acústicos associados à descrição de segmentos

vocálicos¹⁵. Nos achados da presente pesquisa, os valores dos formantes da vogal /ε/ não foram suficientes para diferenciar pares de irmãos com vozes auditivamente semelhantes. A ausência de características vocálicas distintivas aponta que este parâmetro deve ser utilizado com cautela em perícias de identificação forense de falantes entre irmãos. Ou seja, mais uma vez na pesquisa realizada neste estudo, os formantes que são classificados como altamente individuais¹¹, não foram capazes de identificar as vozes auditivamente semelhantes em cada par, demonstrando limitações na utilização dos formantes na identificação de falantes com vozes auditivamente semelhantes.

Em relação à frequência fundamental, observou-se que as medidas acústicas referentes às médias na fala encadeada entre os irmãos do mesmo par não apresentaram significância estatística, corroborando estudo¹⁶ que analisou a frequência fundamental média da fala de gêmeos e o seu desvio padrão numa tarefa de leitura. O estudo mencionado investigou até que ponto a semelhança observada para a frequência fundamental foi geneticamente influenciada ao confrontar dados de gêmeos monozigóticos (MZ) com dados de gêmeos heterozigóticos (HZ). No referido estudo não houve diferenças entre os gêmeos MZ e os gêmeos HZ em se tratando de frequência fundamental

da fala (FFF) média e sua variação (desvio padrão) apesar de terem sido observadas correlações das medidas no primeiro grupo.

Portanto, como verificado no presente estudo, a frequência fundamental, quando utilizada entre irmãos com vozes auditivamente semelhantes, provavelmente não será eficiente para distinguir tais falantes.

A pesquisa também ocorreu com a análise da curva LPC. Quando o exame a ser realizado é o de identificação de falantes, no qual é importante estudar os polos de ressonâncias dos tratos vocais, é necessário também o estudo da curva de resposta em Frequência, a qual é obtida pela LPC¹⁷. Sempre que possível o avaliador deve usar a análise de predição linear (LPC), por ser essa estratégia a mais indicada para a medida dos formantes do som¹¹.

Os gráficos LPC gerados a partir da análise acústica da vogal /ε/ dos pares de irmãos, na presente pesquisa, corroboram com a literatura estudada, mostrando curvas diferentes entre os irmãos do mesmo par (as curvas foram traçadas com cores diferentes para cada irmão do mesmo par para facilitar a visualização). Contudo, para que possam ser usados na prova pericial, optou-se por gerar valores que fossem passíveis de serem analisados estatisticamente para comprovar se havia ou não diferenças significativas entre os irmãos nos pares. Sob este olhar científico, os gráficos foram submetidos à medição da área da curva LPC gerada a partir do áudio da vogal /ε/ de cada sujeito. Este recurso foi utilizado com o objetivo de fornecer um novo método para utilização em perícias baseado em parâmetro objetivo representado aqui pela medida da área da curva LPC.

Após a análise dos gráficos resultantes das medidas das áreas das curvas LPC, foram gerados valores, nos quais as medidas dos pares de irmãos são comparadas estatisticamente.

Na comparação das áreas das curvas LPC entre os pares de irmãos, observou-se que houve diferenças estatisticamente significantes nos pares 1-31, 3-21, 9-32, 14-19. Nos pares 6-28 e 10-25, não foram observadas diferenças estatisticamente significantes. É relevante buscar, no início da presente pesquisa, no pareamento perceptivo-auditivo, que o par 6-28 foi o único que foi considerado coincidente pelos três avaliadores especialistas em voz. De modo geral, esse recurso foi capaz de diferenciar a voz dos irmãos mais velhos e mais novos no mesmo par, exceto quando a semelhança for muito grande auditivamente.

Esse recurso demonstra a importância da análise da área da curva LPC na distinção de vozes auditivamente semelhantes. Os resultados das curvas de LPC demonstraram visualmente que as curvas devem pertencer a sujeitos distintos. Contudo, por se tratar de uma pesquisa científica e visando excluir a subjetividade na interpretação dos dados, foram geradas, de forma inédita, medidas das áreas do LPC que foram submetidas à análise estatística. Com a análise dessas medidas, conseguiu-se detectar a distinção na maioria dos pares, exceto naqueles em que a semelhança vocal foi alta. Outros estudos com uma amostra maior são necessários para aferir a sensibilidade deste novo método. Esse recurso mostrou-se promissor para a distinção de vozes e deve ser aliado às avaliações acústicas com intuito de complementar e fortalecer o deslinde de casos, pois é uma medição inovadora e que pode contribuir com maior confiabilidade nos futuros laudos periciais por trazer menor subjetividade e conferir reprodutibilidade para a atuação do perito.

Este estudo reforça o quão delicada é uma identificação forense de falantes principalmente com vozes auditivamente semelhantes. Também aponta para análise acústica e suas ferramentas usadas de forma alinhada com a perícia desejada, quanto mais semelhantes as vozes comparadas, mais recursos necessitam ser utilizados.

Essa pesquisa finda e ao mesmo tempo consegue despertar novas hipóteses para estudos nesta área que cresce muito à medida que cada vez mais a comunicação oral gravada é amplamente utilizada nos mais diversos processos como elemento da prova pericial.

CONCLUSÃO

Este estudo demonstrou que os formantes da vogal “é” e da fala encadeada, e média da frequência fundamental em Hz não foram suficientes para distinguir as vozes auditivamente semelhantes e que o recurso inédito da medida da área da curva LPC foi a ferramenta que as distinguiu em sua maioria, representando assim, um parâmetro objetivo e reprodutível para ser utilizado na prova pericial.

REFERÊNCIAS

1. Cazumbá LF, Sanches AP, Telles IFC. Introdução à fonoaudiologia forense. In: Rehder MI, Cazumbá L, Cazumbá M, organizadores. Identificação de falantes: uma introdução à fonoaudiologia forense. Rio de Janeiro: Revinter; 2015. p.7-24.
2. Azzariti M. Diálogos de uma tortura: Discursos de um crime. Rio de Janeiro: Rei dos Livros; 2016.

3. Azzariti M, Gomes RV, Vasconcelos ZMC. Linguística: aspectos fonéticos. In: Rehder MI, Cazumbá L, Cazumbá M, organizadores. Identificação de falantes: uma introdução à fonoaudiologia forense. Rio de Janeiro: Revinter; 2015. p.119-37.
4. Conselho Federal de Fonoaudiologia. Resolução 584, 22 de outubro de 2020. Disponível em: https://www.fonoaudiologia.org.br/resolucoes/resolucoes_html/CFFa_N_584_20.htm. [acessado 27 fev 2021].
5. Vieira RC. Identificação de falante: um estudo perceptivo da qualidade de voz. [Tese] São Paulo (SP): Pontifícia Universidade Católica de São Paulo; 2018.
6. Gonçalves CS, Petry T. Comparação forense de locutores no âmbito da perícia oficial dos estados. In: Rehder MI, Cazumbá L, Cazumbá M, organizadores. Identificação de falantes: uma introdução à fonoaudiologia forense. Rio de Janeiro: Revinter; 2015. p.241-64.
7. Lucena LVO. Relação entre as análises acústica e perceptivo auditiva da voz na identificação forense de falantes: uma revisão sistemática [Dissertação]. Recife (PE): Universidade de Pernambuco; 2018.
8. Cazumbá LF, Rehder MI, Sanches AP. Investigação e análise perceptivo-auditiva. In: Cazumbá L, Cazumbá M, Rehder MI, organizadores. Identificação de falantes: uma introdução à fonoaudiologia forense. Rio de Janeiro: Revinter; 2015. p.89-101.
9. Karakoç MM, Varol A. Visual and auditory analysis methods for speaker recognition in digital forensic. In: International Conference on Computer Science and Engineering. Antalya. Anais. 2017:1189-1192. <https://doi.org/10.1109/UBMK.2017.8093505>.
10. Behlau M, Almeida AA, Amorim G, Balata P, Bastos S, Cassol AA et al. Reducing the GAP between science and clinic: lessons from academia and professional practice - part A: perceptual-auditory judgment of vocal quality, acoustic vocal signal analysis and voice self-assessment. *CoDAS*. 2022;34(5):e20210240. <https://doi.org/10.1590/2317-1782/20212021240en>. PMID:35920467.
11. Behlau M, Madazio G, Feijó D, Pontes P. Avaliação de Voz. In: Behlau M, organizador. O livro do especialista. v. 1. Rio de Janeiro: Revinter; 2001. p.85-245.
12. Eriksson A. Aural/Acoustic vs. Automatic methods in forensic phonetic case work. In: Neustein A, Patil HA, organizadores. Forensic speaker recognition: law enforcement and counter-terrorism. New York: Springer-Verlag; 2012. p.41-69.
13. Cavalcanti JC, Eriksson A, Barbosa PA. Acoustic analysis of vowel formant frequencies in genetically-related and nongenetically related speakers with implications for forensic speaker comparison. *Plos One*. 2021;16(2):1-31; <https://doi.org/10.1371/journal.pone.0246645>. PMID: 33600430.
14. Franks S, Barbosa R. A importância da duração da vogal final da palavra para a identificação de falantes não nativos de português por meio de máquinas de vetores de suporte. *RBLA*. 2014;14(3):689-714. <https://doi.org/10.1590/S1984-63982014000300009>.
15. França FP, Almeida AA, Lopes LW. Immediate effect of different exercises in the vocal space of women with and without vocal nodules. *CoDAS*. 2022;34(5):e 20210157. <https://doi.org/10.1590/2317-1782/20212021157pt>. PMID: 35894373.
16. Debruyne F, Decoster W, Gijssels AV, Vercammen J. Speaking fundamental frequency in monozygotic and dizygotic twins. *J Voice*. 2002;16(4):466-71. [https://doi.org/10.1016/s0892-1997\(02\)00121-2](https://doi.org/10.1016/s0892-1997(02)00121-2). PMID: 12512633.
17. Fernandes JR. Perícias em áudios e imagens forenses. Campinas: Millennium; 2014.

Contribuições dos autores:

SC: concepção, ideias, formulação ou evolução de metas e objetivos abrangentes de pesquisa;

MIBCR: supervisão e responsabilidade de liderança pelo planejamento e execução da atividade de pesquisa, incluindo orientação externa para a equipe principal;

LNA, CV: análise formal, aplicação de técnicas estatísticas, matemáticas, computacionais ou outras técnicas formais para analisar ou sintetizar dados de estudos;

CRVD, DV: pesquisa, condução do processo de pesquisa e investigação, realização dos experimentos ou coleta de dados/evidências;

EA: metodologia, desenvolvimento ou desenho de metodologia, criação de modelos.