

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.

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

Study conducted at Departamento de Perícias Forenses da Universidade de Pernambuco - UPE, Recife, Pernambuco, Brazil.

Financial support: Nothing to declare.

Conflict of interests: Nonexistent.

Corresponding author:

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

Received on: November 3, 2022
Accepted on: March 31, 2023



This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

In ancient and contemporary history, there are several reports of people recognition through voice, the most famous being the Lindberg case in 1932. Since voice recognition is a fragile test, based on only one sense of a single person, currently the proposal is to identify speakers, using scientifically based protocols.

Studies are constantly evolving, and several methods have been used for the forensic identification of speakers, in most cases. In Brazil, voice identification methods were introduced for forensic purposes in the 1990s, involving experts from the states, the Federal Police and the Federal District¹. The interception of telephone communications for investigation and as evidence in the Brazilian Criminal Proceedings is an increasingly used procedure².

To assist and support the preparation of forensic evidence, Forensic Science is available, which is the set of all scientific knowledge and techniques that are used to unravel not only crimes, but also other legal issues. Concerning sciences, those directly involved with the forensic identification of speakers for legal purposes include Forensic Linguistics, Forensic Phonetics and Forensic Speech Therapy, whose professionals are dedicated to the complex task of identifying speakers through their voice and speech.

Forensic Linguistics is a branch of applied linguistics dedicated to the investigative context that points to elements that analyze communication in its several aspects³. Forensic Phonetics goes beyond the identification of speakers; it permeates many criminalistic mysteries. The main objective of Forensic Speech Therapy is to respond to legal demands related to human communication, acting in several analyses involving forensic comparison of voice, speech and language; graphotechnics; facial biometrics; transcription, textualization and analysis of audio, video and image content; and description of the communicative profile¹.

Recently, on October 22nd 2020, the Brazilian Federal Council of Speech Therapy recognized the field of Forensic Speech Therapy by resolution n. 584⁴.

For the Forensic Identification of Speakers, it is necessary to compare the standard sample with the sample under analysis⁵. It should be explained that the standard sample is the audio recording that contains the speech of the suspect, accused or defendant (of known identity), and the questioned sample is the audio recording that contains the speaker's speech, whose identity must be known⁶.

Three methods are used by specialists in the field of forensic speaker identification: the auditory-perceptual method, the acoustic method and the automatic method⁷.

The perceptual-auditory method highlights the parameters to be analyzed and presents a strong subjective aspect through a qualitative approach⁸.

The acoustic method uses the spectrogram to analyze the waves produced at the moment of vocal emission, allowing quantitative analysis⁹. The evaluation by acoustic parameter must be standardized, since this analysis provides a number¹⁰, which facilitates analysis, comparisons and storage of measurements. The spectrogram generated in this method is a three-dimensional graph that records the acoustic measurement of the sound wave. It contains information related to sound parameters, i.e., intensity, duration and frequency (time on the horizontal axis, frequency in Hertz on the vertical axis and intensity in Decibel by the color⁹).

In a simplified manner, the acoustic evaluation quantifies the sound signal, which leads to an objective analysis of voice. There is also the following distinction: while acoustics performs measurement of the sound signal, the auditory-perceptual evaluation offers a description of the vocal signal with only hearing as a basic instrument¹¹. The importance of the two proposed methods (perceptual-auditory and acoustic) in association, besides confirming that one is not better than the other but complement each other, was the conclusion of a recent study at the University of Pernambuco⁷.

The other method, the automatic, is performed by softwares that try to reduce subjective analyses as much as possible. The software is fed with information such as vocabulary, programmed and pronounced in many different manners. In some European countries, the use of automatic systems is accompanied by insights from a professional with knowledge in phonetics and even linguistics. For example, at the University of Gothenburg, the software used is ALIZE SpkDet, and the results obtained by the software are combined with traditional acoustic and auditory analysis¹².

Automatic systems are subject to the so-called incompatibility conditions, which occur when differences between voice samples may also appear due to differences in transmission channels, which is a relevant and worrying problem in this type of analysis method¹².

All legal and technological devices support the forensic identification/comparison of speakers, and

more studies are being conducted in this field, so that the binary comparison of voices may be used for legal purposes.

The general objective of this study was to verify the contributions of acoustic spectrographic analysis in the forensic identification of speakers in auditorily similar voices, and to propose an objective method of using the analyzed parameters. The specific objectives were to verify the usefulness of the acoustic parameters: formants of vowel “é”, mean fundamental frequency in Hz, formants F1, F2, F3 in speech, linear prediction curve (LPC) curve of vowel “é”, and area of the LP for distinguishing auditorily similar voices.

METHODS

The study was conducted at the state of Pernambuco and was approved by the Institutional Review Board of the State Hematology and Hemotherapy Foundation, Brazil, under report n. 4.303.659 and CAAE 38306620.3.0000.5195. The independent variables were place of birth, age, sibling and gender, and the dependent variables were the first four formants of vowel “é” (represented by “/ε/”); mean fundamental frequency, F1, F2, F3 in connected speech, LPC of vowel /ε/ and area of the LPC curve.

The study was conducted on 32 people, being 16 pairs, two brothers from each family. The following inclusion criteria were adopted: being brothers (due to genetics), being male (due to the proximity of vocal frequency), being aged between 18 and 60 years (since the voice does not undergo significant changes in this age group) and being native and residing in the state of Pernambuco (due to the accent and especially the pronunciation of vowel “e”, marked in the region). Exclusion criteria were: being twins, considering the existence of previous studies on twins, and/or having a viral, bacterial or inflammatory process in the upper airway on the day of collection, which would influence the voice and possibly the distinction of voice among peers, and/or not having signed the Informed Consent Form.

The investigator (S.C.W.C) recruited participants randomly, sending an invitation specifically designed for this purpose, on social networks and institutions

in the state of Pernambuco. After the participants were defined according to the previously described inclusion and exclusion criteria, data were collected by video, captured by the participant’s cell phone using the device software. The videos had the following recording script, previously explained to the participants: say the name, the date, show an identification document with photograph and date of birth; talk about the state of Pernambuco for 3 to 5 minutes. Afterwards, the videos were sent to the researcher. To perform the first methodological stage, listening to the voice samples, the videos were converted into audio in Wav format by the investigator, using the multimedia conversion software Format Factory®. Preparation of the material for the stage of listening and pairing of voice samples constituted the formation of two groups GimV (group of older brothers) and GimN (group of younger brothers). Then, the names of participants in group (GimV) were replaced by consecutive numbers from 1 to 16. In the group of younger siblings (GimN), the names were randomly replaced by numbers 17 to 32. After this procedure, two groups of voice samples were obtained, GimV with numbers from 1 to 16 and GimN with random numbers between 17 and 32.

To compose the samples of auditorily similar voice to be later investigated by the acoustic spectrographic analysis by the investigator in the second stage, the voice samples of the GimV and GimN groups were submitted to perceptual-auditory pairing, conducted by three speech therapists specialized in Voice by the Federal Council of Speech Therapy – CFFa. The speech therapists who performed the perceptual-auditory pairing were asked to listen to the GimV voices and to indicate the pair of the respective sibling in the GimN and record each pair using a pairing table (Chart 1). Acoustic analysis was performed on pairs of siblings considered to be auditorily similar in a correct manner, belonging to the same family, appointed as peers by at least two of the three speech therapists. Of the 16 pairs submitted to perceptual-auditory pairing performed by speech therapists, six were coincident and submitted to acoustic analysis. The result of the perceptual-auditory pairing is shown in Chart 1.

Chart 1. Perceptual-auditory analytical pairing performed by speech pathologists specialized in voice by the Federal Council of Speech Pathology

Examiner pair	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

Captions: C = Coincident; D = Divergent.

Source: Carmo et al. (2021).

In the second stage, the correctly paired samples were analyzed using acoustic spectrographic analysis, aiming to verify whether and which of the analyzed acoustic parameters would have sufficient statistical power to distinguish people from the same family with auditorily similar voices, and whether and which acoustic parameters were coincident in people born and residing in the State of Pernambuco. The acoustic spectrographic analyses were performed by the investigator (S.C.W.C) using the acoustic analysis software PRAAT®.

In this study, individual acoustic parameters were verified and later compared between the paired brothers, between the pairs and between the two groups (GimV and GimN). The acoustic parameters analyzed were the first four formants (F1, F2, F3, F4) of vowel /ε/, which were extracted after the first minute of speech; mean fundamental speech frequency in Hz; F1, F2 and F3 in connected speech, which were extracted in the first four minutes of speech; LPC curve by the PRAAT® software. The area of the LPC curve was also analyzed from the graphs of the individual LPC curves generated by the PRAAT® software, to propose an original analysis method in the present study. Calculation of the area generated by the comparative LPC graph of each pair studied was performed by an Informatics professional, who generated an algorithm

specifically for this purpose. The LPC curve of each audio separately generated in PRAAT® was submitted to analysis of its area to obtain measurements of the areas formed below the curves, which could be analyzed and submitted to intrapair comparison in the statistical analysis.

To achieve this area, an algorithm was used to generate graphs and calculate the integral (area under the curve). Initially, the image was converted from RGB to a monochrome version and the intermediate gray levels were removed, leaving only completely white or completely black pixels.

Then, a loop was made, first varying the “y” coordinate, in principle, from the first to the last line of the figure. Since the study was dealing with 3,600 x 2,400 resolution figures, this means varying “y” from 0 to 2,399; in each interaction of the “y” loop, another loop was performed, this time varying the “x” coordinate, in principle, from the first to the last column of the figure, i.e., varying “x” from 0 to 3,599. This is described as “in principle” because the pixel colors are evaluated during scanning, and initially all are white pixels. When the first black pixel was found, both loops ended, since it was known to be the upper left part of the graph, reminding that the coordinate point (0,0) is on the first line (uppermost) and first column (leftmost). From the point immediately before this pixel found,

i.e., the coordinates ($x_{\text{black}} - 1, y_{\text{black}}$), in which the coordinates ($x_{\text{black}}, y_{\text{black}}$) are those of that first black pixel found, the “y” coordinate was increased, recording the “y” values where variations are found from white to black, or vice versa. Since the column was being scanned immediately before the “y” axis of the graph, these variations are found in the markings on the “y” axis scale (0, 20, 40, and 60 dB/Hz, depending on the graph being analyzed). Thus, the T “y” Map table was generated, in which the mean “y” coordinate between the transition from white to black and the following transition from black to white was recorded, assuming that the scale value is exactly on the half of the marking stroke. This T “y” Map table allows to map the “y” coordinates expressed in pixels in the figure to their respective values in dB/Hz.

Following, an analogous table T “x” Map was created, this time varying the “x” coordinates from the point ($x_{\text{black}}, y_{\text{mark_min}}$), in which x_{black} is the “x” coordinate of the first black point found above, and $y_{\text{mark_min}}$ is the “y” coordinate of the mark with the lowest dB/Hz value on the “y” axis. Thus varying, the “x” coordinate of the first transition from black to white was recorded, x_{ini} , which characterizes the first column of the graph region; as well as the last transition from white to black, x_{end} , characterizing the last column of this region. The T “x” Map table, thus created, allowed mapping of “x” coordinates, with $x_{\text{ini}} \rightarrow 0$ dB, and $x_{\text{end}} \rightarrow 104$ dB. Finally, the “y” coordinate of ($x_{\text{ini}}, y_{\text{stroke_min}}$) was varied, increasing the “y” value, i.e., following downwards on the graph until finding a transition from white to black, which will occur on the coordinate y_{bottom} , where the “x” axis is located.

Similarly, the “y” coordinate was varied again, this time decreasing it (i.e., going upwards), until finding the y_{top} coordinate, where the upper frame of the graph is located. From there, the dx value was calculated,

defined as: $dx = x_{\text{f_im}} - x_{\text{ini}} / 104$, since 104 is the final value of the “x” axis in all graphs, and the initial value is zero. Then, an integral variable was initiated with value zero, and a loop was started varying the “x” coordinate, in principle, from x_{ini} to x_{end} , and at each iteration of this loop the “y” coordinate was varied, in principle, from y_{bottom} to y_{top} , that is, going upwards, passing through white pixels, then through black pixels (the graph line), and stopping one pixel before the transition from black to white, where the graph point is, at coordinate ($x_i, y_f(x_i)$).

Each time a point ($x_i, y_f(x_i)$) was found, the coordinates expressed in pixels were converted to coordinates expressed in graph units, using the T “x” Map and T “y” Map tables. The value $y_f(x_i)$ is added to the integral variable, zeroed at the beginning of the outermost loop, so that its value at the end of loops is multiplied by the dx value obtained above, providing the final value of the integral, i.e., the area under the curve.

For statistical analysis, the results of the analyzed acoustic parameters were extracted and inserted in a digital spreadsheet. Descriptive analyses were performed, using measures of central tendency, and inferential, using non-parametric comparison tests, since the data did not meet the normality criteria. The Wilcoxon test was used for paired analysis between siblings, and the Kruskal-Wallis test was used to compare groups of older and younger siblings and comparison between pairs of siblings, besides the post hoc Bonferroni test for multiple comparisons. The SPSS software version 21 was used at a significance level of 5% ($p < 0.05$).

RESULTS

Table 1 shows the comparison of measurements of formants of vowel /ε/ between the older and younger brothers of each pair.

Table 1. Comparison of each extracted acoustic measure referring to formants of vowel /ε/ between older and younger siblings of each pair

Pairs	F1 vowel /ε/	p value	F2 vowel /ε/	p value	F3 vowel /ε/	p value	F4 vowel /ε/	p value
Pair 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	
Pair 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	
Pair 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	
Pair 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	
Pair 10-25	498.0	0.233	1629.0	0.317	2338.0	0.317	3549.0	0.317
	335.0		1776.0		2675.0		3566.0	
Pair 14-19	529.0	0.327	1822.0	0.317	2499.0	0.317	3550.0	0.317
	471.0		1640.0		2372.0		4372.0	

Wilcoxon signed-rank test; significance $p < 0.05^*$

Captions: F1 = First formant; F2 = Second formant; F3 = Third formant; F4 = Fourth formant.

Source: Carmo et al. (2021)

The acoustic measurements extracted from vowel /ε/ for F1, F2, F3 and F4 did not show statistically significant differences, as shown in the results in Table 1.

Table 2 presents the comparison of formant measures, of the mean frequency in connected speech between older and younger siblings of each pair.

Table 2. Comparison of each extracted acoustic measure referring to speech formants, mean frequency of speech among older and younger siblings of the same pair

Variables	F1 speech	p value	F2 speech	p value	F3 speech	p value	Mean F0 speech	p value
Pair 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	
Pair 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	
Pair 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	
Pair 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	
Pair 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	
Pair 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	

Wilcoxon signed-rank test; significance $p < 0.05^*$

Captions: F0 = Fundamental frequency; F1 = First formant; F2 = Second formant; F3 = Third formant.

Source: Carmo et al. (2021)

The acoustic measurements presented in this table are not statistically significant.

In Table 3, the possibility of differences in measurements between pairs was considered, since these

subjects are not related, but only have a common birthplace. Thus, Table 3 shows the comparison of acoustic measurements between pairs.

Table 3. Comparison of general means of voice acoustic measures between the six pairs of older and younger siblings.

Variables	F1 vowel /ε/	F2 vowel /ε/	F3 vowel /ε/	F4 vowel /ε/	Mean Fq speech	F1 speech	F2 speech	F3 speech
PAIR 1-31	11.00	6.00	5.00	8.50	4.00	3.50	2.50	4.00
PAIR 3-21	3.50	6.00	10.00	4.50	7.00	4.50	3.50	9.50
PAIR 6-28	3.50	11.50	6.00	10.50	3.50	5.50	8.50	8.00
PAIR 9-32	7.00	9.00	9.00	3.50	11.50	6.00	8.50	4.00
PAIR 10-25	5.00	2.50	5.50	4.00	8.25	11.50	11.50	9.00
PAIR 14-19	9.00	4.00	3.50	8.00	4.75	8.00	4.50	4.50
p value	0.361	0.073	0.422	0.367	0.048*	0.265	0.381	0.572

Kruskal-Wallis test; significance $p < 0.05^*$

Captions: Fq = Frequency; F1 = First formant; F2 = Second formant; F3 = Third formant; F4 = Fourth formant.

Source: Carmo et al. (2021)

The frequency parameter between the six pairs (Table 3) revealed a statistically significant difference between peers, i.e., even knowing that this parameter has a population mean, interpair differences were found.

The Bonferroni's test for multiple comparisons was then performed to observe where these differences occurred, as shown in Chart 2, considering that such differences may contribute to the forensic identification of speakers in general.

Chart 2. Post-hoc test for multiple comparisons between general means of frequency measures of the six pairs of older and younger siblings

Dependent variable	(I) PAIR	(J) PAIR	Mean difference (I-J)	Standard error	Sig.	90% confidence interval	
						Lower limit	Upper limit
Mean speech frequency	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

Test: Bonferroni; significance $p < 0.05^*$

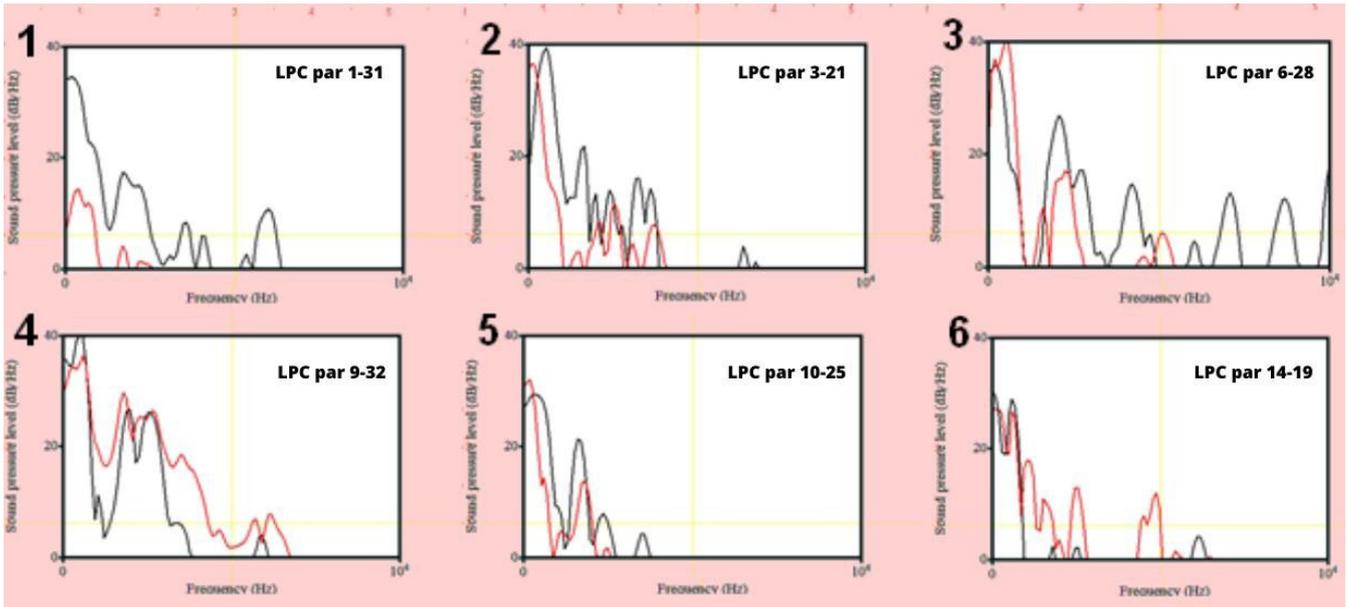
Captions: Sig. = statistical significance; Fq = Frequency; % = Percentage.

Source: Carmo et al. (2021)

With this analysis, no significance was found between the pairs in relation to frequency, i.e., even between all pairs there was not a frequency that could highlight a pair, or even a voice, as previously observed.

Figure 1 presents six images that represent the LPC curve between pairs, the siblings' audios in the graphs are represented by curves with different colors.

The following images demonstrate the differences between audios, since the two resulting curves are distinct, even though in some cases they superimpose or even intertwine.

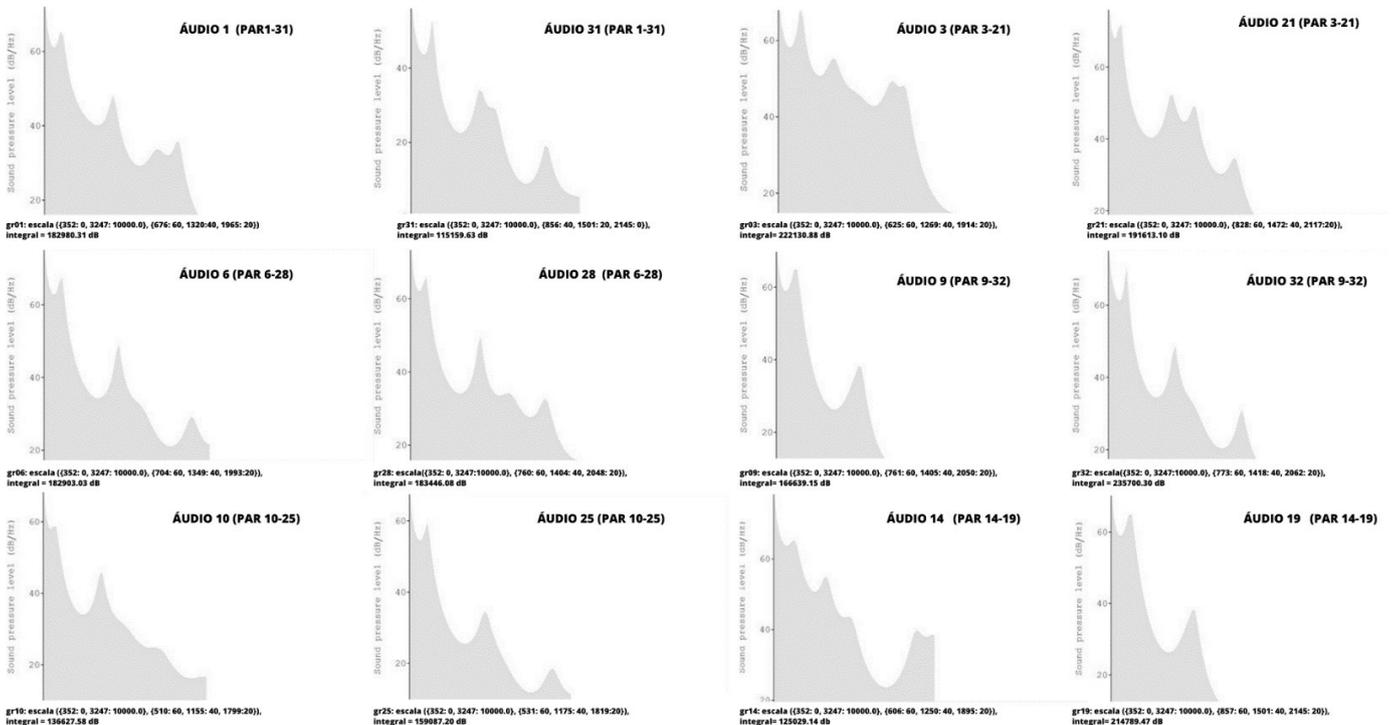


Captions: LPC = Linear Prediction Curve; LPC Par = LPC Pair.
 Source: Carmo et al. (2021)

Figure 1. Linear Prediction Curve of the same pair with different colors for each curve on the same screen

In the present study, the LPC was considered in vowel /ε/, whose results are presented in Figure 1. The analysis applied to a speech signal allows achieving the spectral envelope and the frequencies corresponding to the formants.

Figure 2 presents 12 images with measurements of the area of LPC graphs.



Captions: Áudio = Audio; Par = Pair.
 Source: Carmo et al. (2021)

Figure 2. Area measurements of Linear Prediction Curve graphs in different graphs

Table 4 compares the areas of LPC curves and shows that this measure is able to distinguish, as an

objective parameter, more than 50% of siblings with auditorily similar voices.

Table 4. Comparison of areas of Linear Prediction Curve measurements of the voice of siblings of each pair.

Variables	LPC area	p value
PAIR 10-25	136627.58	0.317
	159087.20	
PAIR 14-19	125029.14	0.0001*
	214789.47	
PAIR 6-28	182903.03	0.817
	183446.08	
PAIR 1-31	182980.31	0.0001*
	115159.63	
PAIR 3-21	222130.88	0.039*
	191613.10	
PAIR 9-32	166639.15	0.001*
	235700.30	

Wilcoxon test; significance $p < 0.05^*$
Caption: LPC = Linear Prediction Curve.
Source: Carmo et al. (2021)

DISCUSSION

As shown in the results of comparison of each extracted acoustic measurement, referring to the / ϵ / vowel formants between older and younger brothers of each pair, the measurements were not able to differentiate the brothers even in the high frequency formant, which is in line with the findings of studies described below.

A recent study¹³ revealed consistent patterns regarding the comparison of high- and low-frequency formants in pairs of twins and non-genetically related speakers, with high-frequency formants exhibiting greater speaker discriminatory power compared to low-frequency formants. It should be mentioned that this study was conducted on pairs of twins (genetically related) and on non-genetically related subjects.

Another study¹⁴ demonstrated that male and female speakers produced vowels with F1 and F2 values relatively close to the targets of native speakers of the state of Paraíba (PB), and the mean values for non-native male speakers were almost identical to the means of native speakers. Formantic measurements are the main acoustic correlates associated with the description of vowel segments¹⁵. In the present findings, the values of vowel / ϵ / formants were not sufficient to differentiate pairs of siblings with auditorily

similar voices. The absence of distinctive vowel characteristics indicates that this parameter should be used with caution in the forensic identification of speakers among siblings. That is, once again in this study, formants that are classified as highly individual¹¹ were not able to identify the auditorily similar voices in each pair, demonstrating limitations in the use of formants for the identification of speakers with auditorily similar voices.

Regarding the fundamental frequency, it was observed that the acoustic measurements referring to the means in connected speech between siblings of the same pair did not present statistical significance, corroborating a study¹⁶ that analyzed the mean fundamental frequency of speech of twins and its standard deviation in a reading task. The mentioned study investigated to which extent the similarity observed for the fundamental frequency was genetically influenced when comparing data from monozygotic twins (MZ) with data from heterozygotic twins (HZ). In that study, there were no differences between MZ twins and HZ twins in terms of mean fundamental frequency of speech (FFF) and its variation (standard deviation), although correlations were observed between measurements in the first group.

Therefore, as observed in the present study, the fundamental frequency, when used between siblings with auditorily similar voices, will probably not be efficient to distinguish such speakers.

The research also analyzed the LPC curve. When the exam to be performed is the identification of speakers, in which it is important to study the resonance poles of the vocal tracts, it is also necessary to study the response curve in Frequency, which is obtained by the LPC¹⁷. Whenever possible, the examiner should use linear prediction analysis (LPC), since this strategy is the most adequate for measuring sound formants¹¹.

The LPC graphs generated from the acoustic analysis of vowel /ε/ of the pairs of siblings, in the present study, corroborate the literature, showing different curves between siblings of the same pair (curves were traced with different colors for each sibling of the same pair for easy viewing). However, to allow their use as forensic evidence, it was decided to generate values that could be statistically analyzed to prove whether or not there were significant differences between siblings in pairs. Under this scientific view, the graphs were submitted to measurement of the area of the LPC curve generated from the audio of vowel /ε/ of each subject. This resource was used to provide a new method for forensic use based on an objective parameter herein represented by the measurement of area of the LPC curve.

After analyzing the graphs resulting from the measurements of areas of the LPC curves, values were generated, in which the measurements of pairs of siblings are statistically compared.

Comparing the areas of the LPC curves between pairs of siblings, it was observed that there were statistically significant differences in pairs 1-31, 3-21, 9-32, 14-19. In pairs 6-28 and 10-25, no statistically significant differences were observed. It is relevant to mention that, at study onset, in the perceptual-auditory pairing, the pair 6-28 was the only considered coincident by the three examiners specialized in voice. In general, this resource was able to differentiate the voice of older and younger brothers in the same pair, except when there is marked auditory similarity.

This resource demonstrates the importance of analyzing the area of the LPC curve in differentiating auditorily similar voices. The results of the LPC curves visually demonstrated that the curves must belong to different subjects. However, since this is a scientific research and aiming to exclude subjectivity in data interpretation, measurements of the LPC areas were

generated, in an unprecedented manner, which were submitted to statistical analysis. With the analysis of these measurements, it was possible to detect the distinction in most pairs, except for those in which the vocal similarity was high. Other studies on larger samples are needed to assess the sensitivity of this new method. This resource proved to be promising for the distinction of voices and should be combined with acoustic evaluations to complement and strengthen the delineation of cases, since this is an innovative measurement that can contribute to greater reliability in future forensic reports by bringing less subjectivity and providing reproducibility for the work of forensic experts.

This study reinforces how delicate is the forensic identification of speakers mainly with auditorily similar voices. It also points to acoustic analysis and its tools used in line with the desired forensic analysis; the more similar the compared voices, the more resources should be used.

This study is completed and simultaneously raises new hypotheses for studies in this field, which has been growing as recorded oral communication is increasingly used in the most diverse processes as an element of forensic evidence.

CONCLUSION

This study demonstrated that the formants of vowel “é” and connected speech, and the mean fundamental frequency in Hz were not enough to distinguish auditorily similar voices. It also showed that the unprecedented resource of measuring the area of the LPC curve was able to distinguish most of them, thus, representing an objective and reproducible parameter to be used in forensic evidence.

REFERENCES

1. Cazumbá LF, Sanches AP, Telles IFC. Introdução à fonoaudiologia forense. In: Rehder MI, Cazumbá L, Cazumbá M, editors. 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, Vasconcellos ZMC. Linguística: aspectos fonéticos. In: Rehder MI, Cazumbá L, Cazumbá M, editors. 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. Available at: https://www.fonoaudiologia.org.br/resolucoes/resolucoes_html/CFFa_N_584_20.htm. [Accessed 2021 fev 27].

5. Vieira RC. Identificação de falante: um estudo perceptivo da qualidade de voz [thesis]. 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, editors. 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 [dissertation]. 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, editors. 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, editor. 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, editors. 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.

Author contributions:

SC: design, ideas, formulation or evolution of overall research goals and objectives;

MIBCR: supervision and leadership responsibility in the planning and accomplishment of research activity, including external guidance of the main staff;

LNA, CV: formal analysis, application of statistical, mathematical, computational or other formal techniques to analyze or synthesize the study data;

CRVD, DV: research, accomplishment of research and investigation process, accomplishment of data collection/evidence;

EA: methodology, development or design of methodology, creation of models.