Case Report

Forensic Speaker Verification: Case Report at Brazilian Justice

Thyago J. Machado[1,*], Jozue Vieira Filho[2] and Mario A. de Oliveira[3]

[1] São Paulo State University (UNESP), Campus of Ilha Solteira, São Paulo, 15385-000, Brazil; tmachado@forenselab.com
[2] Telecommunications and Aeronautical Engineering, São Paulo State University (UNESP), São João da Boa Vista, SP 13876-750, Brazil; jozue.vieira@unesp.br
[3] Department of Electrical and Automation Engineering, Mato Grosso Federal Institute of Technology, Cuiabá 78005-200, Brazil; mario.oliveira@cba.ifmt.edu.br
* Correspondence: tmachado@forenselab.com; Tel.: +55-65-98112-2338

Abstract: This case report investigates 5 real cases which followed legal channels and were judged by Mato Grosso Court in Brazil. Audio systems served as elements of key evidence on those lawsuits. The goal here is to analyze the cases by using a methodology based on the forensic speaker verification by using the Ordinary Least Squares (OLS) algorithm and to compare results with analyses obtained on real cases. The comparative analysis is assessed for time elapsed for obtaining results, as well as results quality. In Brazil, the lawsuit duration is very important, since the Penal Code foresees prescription after a given time, and it may lead to impunity. Results show that the analysis, by using OLS, generates immediate, effective results when compared to those obtained with traditional methodologies on the studied Brazilian lawsuits.

Keywords: forensic speaker comparison, voice processing, ordinary least squares, OLS.

1. Introduction

Brazilian law has a device which pleads over prescription before definitive sentence and this is calculated based on the maximum penalty [1]. Article 109 from the civil Code says, "The prescription, before final sentence (...) is ruled by the maximum of freedom private penalty communicated to the crime" [2].

The same article 109 establishes the time needed to be elapsed before the final sentence, so that prescription occurs. Besides, this same article specifies that defendants under 21 years old, when the felony was committed, or over 70, at the conviction day, must have the prescription period reduced by half, which is defined, by the Penal Code, as "age prescription" [3]. This demands an even bigger effort forensic team, in order to obtain faster results. Table 1 shows a comparative chart between regular and age prescriptions times.

Table 1. Correlation between penalties regarding their prescriptions.

Penalty	Regular Prescription	Age Prescription
0 – 12 months	3 years	18 months
>1 year < 2 years	4 years	2 years
>2 years < 4 years	8 years	4 years
>4 years < 8 years	12 years	6 years
>8 years < 12 years	16 years	8 years
>12 years	20 years	10 years

From the moment there is a conviction by Brazilian justice's maximum instance (Federal Supreme Court) interrupts the prescription countdown based on the maximum penalty possible and the countdown based on the penalty in fact applied to the convicted

one. This is typified on article 110 of the Brazilian Penal Code, i.e.: "Prescription after final sentence is regulated by the penalty applied" [2].

As an example, consider a given lawsuit for approximately 3 to 5 years, when the individual committed a crime susceptible to a maximum penalty of 2 years. In this case, a prescription will occur in 4 years. However, if the conviction is the minimum penalty of 6 months, the prescription reduces to 3 years and the individual considered guilty will remain unpunished because the prescription occurred. Therefore, the time for the conclusion of works should be the smallest possible to allow convictions not to be prescribed and to make the convicted ones serve their times according to what justice establishes [2].

Mato Grosso State, located on center-western region in Brazil, has a pent-up demand for forensic analyses on the Audio sector. Nowadays, the average time to begin a forensic analysis on a given audio file is about one year. After the beginning of the works, there are 3 to 6 months for technical analysis and issuing of a final report. Naturally, all this time is due to the techniques used by the specialists, which contribute negatively to the efficiency of Brazilian justice. Thus, a quicker, more effective methodology might reduce significantly the times mentioned, and reduce the chances of prescription of crimes, thus reducing impunity.

Within the automatic speaker identification, several attempts have been proposed and applied to different groups of people. In [4], a method that assesses the performance of an acoustic-phonetic system using empirical testing was addressed. Results show that the performance obtained on auditory-acoustic-phonetic-spectrographic is inferior to the one obtained with automatic systems, i.e., objective analysis via empiric tests has bigger credibility than the subjective analysis via graphs. In [5], the authors propose a novel method for checking the homogeneity of an audio recording which includes a morphological examination of the sample and the step of deep analysis. They pointed out that due to the uniqueness of the case, only a limited sample was available for examination. Within the neural networks-based methods, recent advances have been proposed based on Convolutional Neural network [6-7]. Although these were good results, they did not take into account the time processing cost, as the method needs hours of learning. It is necessary for the announcer to record the maximum possible amount of different vocabulary in a row, with the aim of allowing the neural networks to achieve an enhanced success rate. A method of the speaker-discriminatory potential of vowel formant mean frequencies in comparisons of identical twin pairs and non-genetically related speakers was presented in [8]. They employed Praat software to extract F1-F4 formant to estimate automatically extracted from the middle points of each labelled vowel. Although the good results, the authors pointed out that even the identical twins displayed a higher phonetic similarity, they were not found phonetically identical. Vowel formants are also employed in [9], whereas twin pair speaker identification was also addressed in [10]. Readers are encouraged to consult the references [11-12] for further details about methods do identify the speaker.

Notwithstanding, acoustic-phonetic methods have been the subject of discussion which includes, for example, voice quality which is generally understood by forensic practitioners and how it contrasts with the practices of voice therapists [13]. It is also worthy to bear in mind that there is an absence of a voice database in Brazil, considering the Brazilian authorities such as the Federal Police and the Criminal Institutes of the Brazilian States. In the same way, several methods found in the literature (English database), for speaker identification, might not work accurately for Brazilian Portuguese. The Brazilian Portuguese language has a complex vocabulary, having a symmetrical and balanced phonetic system with the final notes more clearly than in European Portuguese [15]. Likewise, speaker recognition in Brazil still presents decision-making based on the subjective analysis of results using unreliable techniques [15]. To overcome that, a methodology based on analyzing formants via OLS algorithm was proposed in [16]. The authors stated that the obtained results are legitimate for Brazilian Portuguese and these strongly reinforce the potentiality of the methodology employed since the several voic-

es dataset is in English, which could provide distance from Brazilian Portuguese. Although the good results, it was missing to test their methodology in real cases of Brazilian justice.

Based on all the above, this report analyzes the performance of a technique proposed on forensic speaker verification based on the OLS algorithm [16], for assertive efficiency as well as for response time. Five cases, which have already prosecuted and sentenced by the Mato Grosso court in Brazil, were employed to evaluate the method. All cases involve voice verification and identification of those involved, to vouch for the truthfulness of the recording attached to the lawsuits. The obtained results demonstrated that the method presented an accuracy of 96% to identify automatically the true speaker and that may reduce a task of months for days can significantly affect the reduction of procedural costs.

2. Materials and Methods

2.1. Cases Information

Firstly, it is important to highlight that the Case Reports used on this paper are in public domain, because it rests absent the request of procedural secrecy.

Case 1: Domestic violence defined by Maria da Penha law (Brazilian law for women's protection) where, recurrently, the wife was victim of slander and psychological violence and could not stand living on this setting anymore, requiring protective measures against the husband, as well as divorce. By using a cell phone, the woman managed to record two episodes when the aggressor would have insulted her psychologically. Under oath, ex-husband denies the voice contained on the record is his, demanding, thus, an audio verification.

Case 2: Commercial trade signed through a phone call, confirming a debt of K millions of Reais (Brazilian currency) which was supposed to be paid in 5 equal parts (5 x K/5). On the lawsuit records, the payment of a single part has been registered, which would have been the first one, and the creditor charges, in justice, the other 4 remaining parts. The alleged debtor, by his turn, claimed in writing that he does not recognize the demanded debt and denies having made any arrangement or having confessed any debt, be it in writing or by phone call. With the audio file in his power, the creditor included the recording to the lawsuit as evidence to the agreement and the judge requested inspection to confirm the truthfulness of the facts.

Case 3: Signing of a monthly fashion magazine via corporative credit card payment made via phone call, with voice-confirmed data, in the value of X Reais. The company does not recognize the transaction claimed by the client, but considering it has only 21 employees, all males, decided to verify the recordings on the period of signing. With all audio files, the company requested a forensic analysis to verify if one of their employees made the transaction.

Case 4: A police officer approached a motorcyclist driving in high speed and claimed to have received an offer of bribe to release him immediately without making the breathalyzer test. The policeman recorded the conversation during the approach and, as soon as he asked for the man's documents, he informed the motorcyclist that he was being recorded. The motorcyclist denied the audio's authorship and a forensic analysis was requested.

Case 5: On a purchase request via WhatsApp of a private company, an exchange of voice messages between a company female employee and a client triggered a lawsuit for sexual harassment of the employee against the female client. By claiming compromising vexation situation and embarrassment, the employee demanded in justice a compensation for moral damages. Client vehemently denied having sent such kind of content to the employee and an evidence specification has been demanded.

2.2. Methodology applied

Voiceover identification through voice audio files has, for base, the comparison between the disputed audio and an audio file containing the same sentence recorded on the disputed audio file, which is called reference audio. The methodology used at the implementation of OLS algorithm demands some complementary steps, be it for analysis or final verification to obtain results, just like presented on Figure 1 [16]. Thus, audio files are initially windowed and transformed for the frequency domain for Linear Predictive Coding (LPC) parameters generation. After that, algorithm plied to obtain formants. Then, the model obtained compares statistically the significance degree between reference audio and disputed audio file, of each formant found.

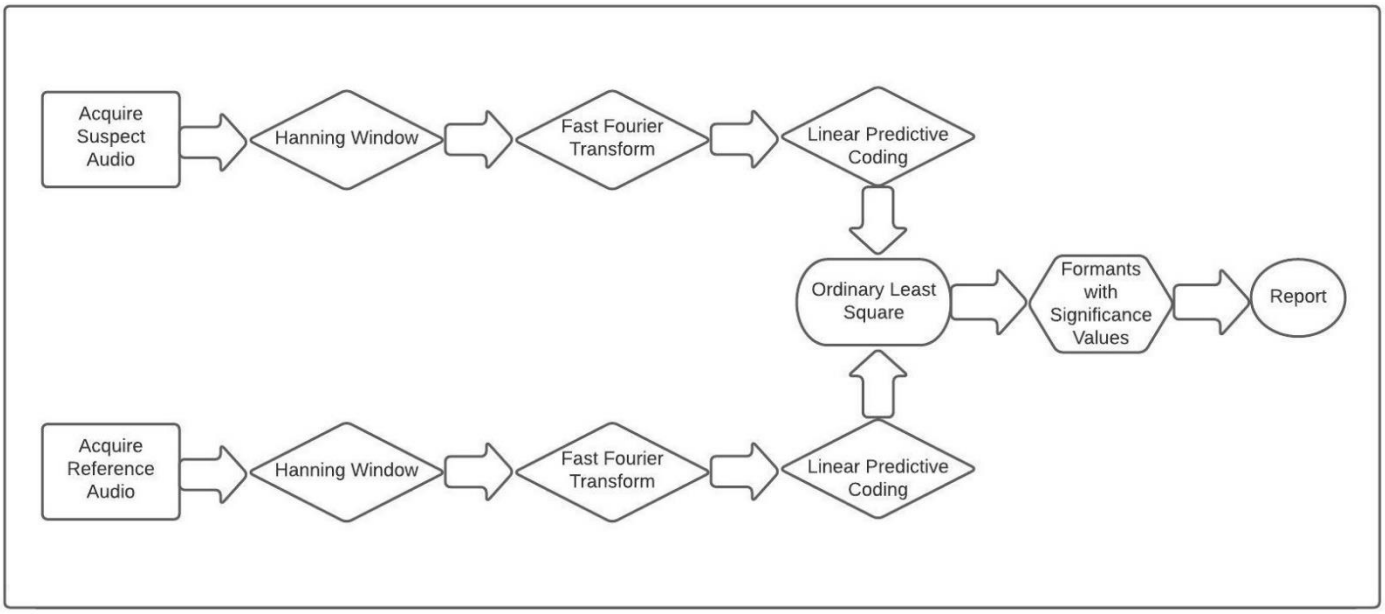

Figure 1. Steps for forensic speaker comparison based on OLS algorithm.

The first step of the methodology, for both audios reference (x[n]) and suspect (y[n]), consists in segmenting the analyzed audio in smaller pars through the following equation:

$$t_{jan} = \left(\frac{0{,}45}{pitch_{floor}}\right) * 1000 \quad (1)$$

where, t_{jan} is the window interval in seconds and $pitch_{floor}$ is the smallest frequency expected in the audio file. After that, audio files are passed through Hanning window, with the goal to smooth the border effects and to limit data [17]:

$$w(n) = 0.5 * \left[1 - \cos\left(\frac{2\pi n}{N}\right)\right] = \sin^2\left(\frac{\pi n}{N}\right) \quad (2)$$

where w(n) is the Hann windowing function, N is the window's width, and n is each one of the values throughout the interval $0 \leq n \leq N$.

The audio file signals frequency spectrums X[k] are computed through Discreet Fourier transform of N points (FFT algorithm), as follows:

$$X[k] = \sum_{n=0}^{N-1} x[n]\, e^{\frac{-j2\pi kn}{N}} \quad (3)$$

where k = 0, …, N-1.

Posteriorly, LPC algorithm is applied to get the roots of the coefficients for both reference (p) and suspect (l). For the suspect, the following equation is used (similar for the reference):

$$l = \sum_{n=1}^{N} s_p(n)\, x^{N-n} \quad (4)$$

where, sp are the coefficients of the polynomial (estimated model), n is the coefficient, and N the maximum number of the polynomial's coefficients.

The OLS algorithm is computed by:

$$p_{médio} = \frac{(\sum_{i=0}^{n} p_i)}{n} \quad (5)$$

$$l_{médio} = \frac{(\sum_{i=0}^{n} l_i)}{n} \quad (6)$$

$$\hat{a} = p_{médio} - \hat{\beta} * l_{médio} \quad (7)$$

$$\hat{\beta} = \frac{\sum_{i=0}^{n}(l_i - l) * (p_i - p_{médio})}{\sum_{i=0}^{n}(l_i - l_{médio})^2} = \frac{Cov(l,p)}{Var(l)} = r_{LP} * \frac{S_l}{S_p} \quad (8)$$

where r_{lp} is the coefficient of sampling correlation, and s_l and s_p are the not corrected sample standard deviation of l and p, $\hat{a}$ is the constant regressor term, $\hat{\beta}$ is the scalar regressor term of linear model, and the determination coefficient R-square is given by:

$$R^2 = r_{lp}^2 \quad (9)$$

What is assessed on test-F is if the model tested is capable to adjust itself to the data significantly better than the arbitrary model. The test-F is calculated from the following equation:

$$F = \frac{\frac{RSS_1 - RSS_2}{m_2 - m_1}}{\frac{RSS_2}{n - m_2}} \quad (10)$$

where RSS represents the residual squares of the reference model (m1) and of the suspect model (m_2), and n are each one of the data points compared among the models. Significance values are calculated considering an α = 90%, and the p-Value is represented according to Table 2.

Table 2. Intervals for determination of p-Value on test-F.

Symbol	p-Value
'NS'	P > 0.1
'*'	0.05 < p ≤ 0.1
'**'	0.01 < p ≤ 0.05
'***'	p ≤ 0.01

3. Results

3.1. Comprehensive analysis

Audio processing was conducted using an application developed in Matlab 2018b [16]. First, data were loaded by a guest user interface (GUI), where the user could select the desired audio file. The software is able to support any audio file format, such as m4a, ogg, wav, mp3, and mp4.

After loading the audio files, the algorithm performs the windowing with adjustable for both duration and overlap. The length window is equal to the number of formants compared to that which has fewer elements with the variables of windowing in the percentage of overlap between the windows of 90%. Posteriorly, the signal is decomposed by applying the Fourier transform. The spectrum is plotted by extracting the formant frequencies of each Fourier spectrum from each audio window. The extracted formants are rearranged and, followed by applying a move average filter, which has a default length of 11. However, it can be customized by the user. Then, the algorithm separates the maximum number of formants possible for the processed audio. The maximum number of formants varies according to the characteristics of each audio file (suspect and reference). However, the data must be formatted to present the same number of formants (columns) and samples (lines) before we compare them. The numbers of LPC coefficients to be extracted must be greater than the number of bins in the FFT spectrum. Afterwards, the algorithm removes the data windows the LPC did not return values and also does the conversion of the values of the roots of the LPC to real values, and determination of the phase associated with each root. To sum up, the separation of the formants is carried out according to the criteria of frequency> 90 Hz and bandwidth <400 Hz, removing the frequencies that are not considered formants (removes the zeros).

For application of the methodology previously described, 18 formants with white noise sensibility up to 1% of total audio time have been selected, by confronting the same sentences of the contested audio file with the reference audio. White noise is a random sign with equal intensity in different frequencies (spectral density of constant power). It is important to highlight that the reference audio is obtained through several voiceover recordings, "investigated or under suspicion" in the presence of an expert of the forensic area. Results of the several cases analyzed are summarized on Tables 3, 4, 5, 6 and 7, having as basic parameters of assessment the formants.

It is important to mention that the confirmation of the suspect, for the proposed methodology, is given by means of significance analysis (p-Values) for the formants (Table 2). The methodology tries to obtain and to analyze 20 formants. For example, if there is the obtaining of "***" for all formants, this implies in a perfect correlation degree, confirming the suspicion of the audio file. For the cases where there is no perfect correlation, the methodology searches, for audio authorship, the biggest number of "***". If there is more than one "NS" value, the suspicion is rejected.

3.1.1. Results for Case 1

Table 3 presents results of application of the methodology proposed for Case 1. Upon analyzing the results, it was verified that formant F18 was not obtained, possibly due to the cell phone suffocation during the argument. However, the suspect was considered positive due to the fact of obtaining "***" for several formants (Table 2). The small p-Value resulted in a higher significance. In this approach, the highest significance was represented by '***'.

Table 3 – The results obtained after applying the proposed method to the Case 1.

Suspect	Time(s)	Pitch	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18
A1	18,3569	***	***	***	**	***	**	***	***	***	**	**	***	***	***	***	**	***	***	

3.1.2. Results for Case 2

Table 4 shows the result for Case 2. Based on obtained results, one notes that, despite the conversation being long, the remarkable characteristics of the communicating voice were validated with standard audio, confirming that the voice in question belongs to the suspect (Case 2). Phone conversations tend to have a poorer quality due to technical limitations of the channel [18].That is why the results based on the proposed methodology did not present uniform significance for all formants.

Table 4 – The results obtained after applying the proposed method to the Case 2.

Suspect	Time(s)	Pitch	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18
B1	31,0560	***	**	***	***	NS	**	**	*	**	***	**	**	**	***	***	***	**	*	**

3.1.3. Results for Case 3

Table 5 presents the results obtained for Case 3. It is important to mention that, unlike the other cases, which had only one suspect, here it was collected audio files of the 21 male employees of the company. All suspects totally denied about the authorship of the audios. Thus, all were considered suspects and audios collecting were performed for the 21 communicating voices.

Table 5 – The results obtained after applying the proposed method to the Case 3.

Suspect	Time(s)	Pitch	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18
C1	3,2182	NS	*	*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	
C2	6,7436	*	NS	*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS				
C3	3,0043	**	NS	NS	*	*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	
C4	3,3206	**	NS	***	NS	NS	NS	NS	NS	*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
C5	3,2415	***	**	**	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS			
C6	3,2573	**	***	***	NS	NS	*	***	*	*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
C7	4,7832	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	
C8	3,2551	***	NS	*	NS	NS	NS	NS	*	**	**	NS	*	***	*	NS	NS	NS	NS	*
C9	3,4461	***	**	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS					
C10	3,4249	NS	NS	NS	*	NS	NS	NS	NS	*	NS	NS	NS	NS	NS	NS	NS			
C11	3,4270	NS	*	*	NS	***	***	NS	NS	NS	NS	NS	NS	NS	***	***	NS	**	NS	***
C12	3,4268	NS	***	NS	*	**	NS	NS	NS	NS	**	**	*	**	NS	*	NS	*	*	NS
C13	2,4598	NS	NS	*	***	NS	NS	NS	NS	**	*	*	*	NS	NS	NS	NS	NS	NS	NS
C14	3,0185	NS	NS	NS	*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
C15	3,3577	***	*	NS	NS	NS	NS	**	NS	NS	*	NS	NS	NS	NS	NS	NS	NS	NS	NS
C16	3,3351	*	**	NS	***	NS	NS	**	***	**	**	NS	***	**	*	NS	NS	NS	NS	NS
C17	3,3365	**	*	*	NS	**	*	*	*	*	*	**	**	*	*	*	*	*	*	*
C18	4,8113	*	***	NS	***	NS	**	*	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
C19	3,3365	NS	***	***	NS	**	***	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
C20	3,4141	***	***	**	***	***	***	***	**	***	***	***	***	**	***	***	***	***	***	**
C21	3,4102	NS	NS	*	*	NS	NS	NS	NS	NS	NS	*	NS	NS	NS	NS	NS	NS	NS	NS

Based on the results obtained, one notices that suspects C17 and C20 presented higher level of significances by considering all formants. Therefore, they could be classi-

fied as authors of the audios. However, analyzing thoroughly the formants, one realizes that suspect C20 presents a bigger amount of p-Values "***" if compared to C17. Besides, C17 presented an "NS". Thus, it would be possible to rule out C17 and to affirm the audio belongs to suspect C20. Upon verifying that the software delivers objective results, it is possible to classify as a false positive case the audio belonging to C17, because, despite the significance being relatively low, it did not fit on the automatic rejection filters. Considering 25 tests performed and one case of false positive, the result is an accuracy of 96% for the proposed method, which is considered enough for forensic analyses [19]. The suspicion of the possible methodology flaw are that the formants confronted have similarity, even low, and the technique shows that, despite the similarity is small, it was observed, but not enough to validate that C17 is the alleged author after reading the information.

3.1.4. Results for Case 4

Table 6 presents the results obtained for Case 4. The fact of the recording device being too close to the suspect communicating voice, without suffocation, generated great results with excellent in all formants verified, the high number of "***" ensured that the speaker was positively identified.

Table 6 – The results obtained after applying the proposed method to the Case 4.

Suspect	Time(s)	Pitch	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18
D1	64,3721	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***	***

3.1.5. Results for Case 5

Table 7 presents results obtained from application of the methodology proposed for Case 5. It is important to mention that WhatsApp has a codifier, which reduces the audio files quality, as well as images and videos, aiming to make it easy the exchange of information between the ends (Opus [20]) Compression results resulted significantly on not obtaining formants F15 to F18. However, the remaining formants were enough to attest the positive voice of the suspect, as observed on Table 7.

Table 7 – The results obtained after applying the proposed method to the Case 5.

Suspect	Time(s)	Pitch	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18
E1	5,1623	***	***	**	***	**	***	**	**	**	NS	***	*	**	**	**				

It is important to mention that, during production of a forensic report by the practitioner, those analyses contribute significantly for enhancement of the forensic practitioner's job, for time and technical detail, structure and format the expert reports to enhance their appropriate impact on the trier of fact [21].

3.2. Sentences Post Jury

Voiceover comparison was fundamental for disputes' solution, because it was key evidence to validate the facts presented for one of the parties. It is important to mention that in all lawsuits of these cases, the different judges handed down the sentences based on the forensic work applied to the audio analyses, which were preponderant for truth elucidation. Up next, a brief summary is presented of conclusion, taken from judicial sentences, of each case analyzed here.

Case1: After reconnaissance work of the voiceover, ex-husband was convicted for having performed psychological violence and got a penalty where it has a protective measure to keep away for at least 500 m from any family member of the ex-wife. Child

visit was made with the presence of a social worker, besides having been convicted for slander. However, this was converted into payment of care package to philanthropic institutions and the divorce is still under course.

Case 2: After tried, the credit owner won and the real estate are scheduled for going into auction to pay the updated debtor's debt. This was possible due to the fact of reconnaissance that the voice recording would be a debt confession. The crime of larceny was prescribed due to the overly extended time to compose the forensic evidence and to judge, there was not occurred the prison, only the debt payment.

Case 3: The employee accountable for the call was fired from the company and had to pay a compensation of 3 minimum wages to the magazine for false testimony. He was not arrested, because the crime prescribed. In this specific case, only to perform the forensic work, since the arrival of the request to the conclusion of the lawsuit, it took approximately one year and a half. Enough time for the felon to be unpunished.

Case 4: The motorcyclist was arrested and had his driver's license retained, however he paid a bailout and was released. The lawsuit on the corruption crime was prescribed due to the laxness on the lawsuit following of legal channels. Forensic contributed in some part for the significant for the excessive delay.

Case 5: The invasive flirting was considered a penal contravention of offensive heckle to modesty, defined by article 61 of Penal Contraventions Law (Law nº 3688/41). In this case, the penalty imposed was of a fine for the penal lawsuit and the same value to the civil lawsuit for moral damage.

4. Discussion

Table 8 presents the time comparison for the two methods (usual and proposed) highlighting the times of forensic work and of prescription. It is important to mention that the usual method applied is to listen to, countless times with the forensic practitioners' ear, trained to identify any point of speech that can be characteristic on the standard audio, for further localization and confrontation with the disputed audio, demanding more time.

Table 8 – Analogy of complete working time and prescription on the current applied and tested methods.

Case #	Complete Work Time		Prescription	
	Usual Method	Tested Method	Usual Method	Tested Method
1	3 months	1 day	No	No
2	4 months	1 day	Yes	No
3	6 months	3 days	Yes	No
4	3 months	1 day	Yes	No
5	3 months	1 day	No	No

It is important to mention that, to begin a voiceover comparison work, regardless of the methodology to be applied, the time for material collecting to be confronted with the disputed audio last about one hour for each communicant. This is due to the fact of being necessary to record the voiceover several times. However, if we consider all the voiceover identification methodology, when applying the methodology here proposed, and time difference is expressive, can be reduced up to 99% regarding the usual methods. That way, the waiting line, which today is up to one year waiting for the next forensic practitioners to begin his job, could be drastically reduced. This is possible if we take into account that the forensic practitioners will be busy, on average, 1 day for each case with few comparisons, i.e., more time do dedicate to a new job, with less processing time and voiceovers filter.

Considering, for example, if Case 3 had been analyzed using the voiceover comparison proposed on this report, the conviction on the case might not have been prescribed.

Even Case 3 was one week away from its deadline, with the methodology proposed here, the criminal would have his sentence to be served, inhibiting this gap of impunity that exists in Brazilian Penal Code. It is also important to highlight that the application of a faster forensic technique could lead the suspect to an earlier conviction, which is as important as the concern about the prescription.

Table 9 shows the comparison of efficiency on assertive result for Case 3. Analyzing the results presented on Table 9, one notes that only one of the communicants for Case 3 presented a false positive. In short, the efficiency observed was of 96% for the tested method, i.e., of 25 tests, only 1 of these presented false positive which can be considered robust enough for applications in the forensic area [19].

Table 9: Comparison of the efficiency on assertive result for Case 3: considering the usual and tested methodologies.

Communicating Voice	Method	
	Usual Method	Tested Method
Case #1 – A1	Correct	Correct
Case #2 – B1	Correct	Correct
Case #3 – C1	Correct	Correct
Case #3 – C2	Correct	Correct
Case #3 – C3	Correct	Correct
Case #3 – C4	Correct	Correct
Case #3 – C5	Correct	Correct
Case #3 – C6	Correct	Correct
Case #3 – C7	Correct	Correct
Case #3 – C8	Correct	Correct
Case #3 – C9	Correct	Correct
Case #3 – C10	Correct	Correct
Case #3 – C11	Correct	Correct
Case #3 – C12	Correct	Correct
Case #3 – C13	Correct	Correct
Case #3 – C14	Correct	Correct
Case #3 – C15	Correct	Correct
Case #3 – C16	Correct	Correct
Case #3 – C17	Correct	Wrong
Case #3 – C18	Correct	Correct
Case #3 – C19	Correct	Correct
Case #3 – C20	Correct	Correct
Case #3 – C21	Correct	Correct
Case #4 – D1	Correct	Correct
Case #5 – E1	Correct	Correct

Analyzing specifically Case 3, C17 (false positive) the forensic practitioner could easily distinguish through hearing of the audio that the author of the felony was communicant C17 instead of C20. This is because, despite both audios being of males, the disputed audio was more high-pitched in comparison to the reference audio (deep).Thus, despite the methodology not having identified clearly the author of the audio, the forensic practitioner could easily identify the real communicant.

It is also important to mention that the method tested consists on a true filter so that the responsible for the comparison have a smaller scope of audios to work with, thus speeding up the result. For example, for Case 3, which contained at first 21 comparisons of voiceovers to be performed in only 3 days, using the methodology presented here, this value was reduced to 2 possible individuals (suspects #17 and #20).

Formant-based acoustic-phonetic systems have been criticized to be vastly inferior to MFCC-based human-supervised-automatic systems [22]. However, those tests

were performed with only 4 formants, different from the model evaluated here, wherein the comparisons made, the obtained formants were between 14 to 18 (about 4 times higher), bringing more detail consequently refining the result as a filter for the expert to narrow the scope of analysis. It is important to point out that there is a good deal of metrics to be analyzed so that a single software, by itself, produces a reliable result, valid in the courts and that these were subject to reservations by practitioners how can one decide whether the system is good enough for its results to be used in court [23]. However, the proposed method showed 96% effective and may be effectively used as a filter by practitioners to get an effectiveness of 100%. It also reduces drastically the chance of the process being prescribed and losing its legal validity in the Laws of Brazil.

It is important to observe that the reduction of analysis time might mean economy to publish spending, because a lawsuit has costs with judges, prosecutors and forensic practitioners and, the longer the normal course through legal channels, until final sentence, the bigger will be the procedural expenses. Thus, a method that is able to reduce a task of months for days can significantly affect the reduction of procedural costs.

Author Contributions: For T.J.M., J.V.F and M.A.d.O. conceived and designed experiments; T.J.M. performed the experiments; T.J.M., M.A.d.O. and J.V.F. wrote the paper.

Funding: This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - Finance Code 001.

Conflicts of Interest: The authors declare no conflicts of interest.

References

[1] Andreucci, R. A. Manual de direito penal. Saraiva Educação SA, **2018**.
[2] BRASIL, Decreto-Lei 2.848, de 07 de dezembro de 1940, Código Penal. Diário Oficial da União, Rio de Janeiro, **1940**.
[3] Boschi, J. A. P. Das penas e seus critérios de aplicação, Livraria do Advogado Editora, **2018**.
[4] Enzinger, E.; Morrison, G.S. Empirical test of the performance of an acoustic-phonetic approach to forensic voice comparison under conditions similar to those of a real case. *Forens. Sci. Intern.* **2017**, 277, 30-40.
[5] Soni, M.; Sagarwal, N.; Abdullahi, Z.H. Application of Python Audio Analysis Library for Performing Deep Analysis to Test Signal Homogeneity on an Audio Sample: A Case Study. *J Forensic Res. Crime Stud.* **2020,** 5,1-7.
[6] Dhakal, P.; Damacharla, P.; Javaid, A.Y.; Devabhaktuni, V. A Near Real-Time Automatic Speaker Recognition Architecture for Voice-Based User Interface. *Mach. Learn. Knowl. Extr*. **2019,** 1, 504–520.
[7] Nassif, A.B.; Shahin, I.; Hamsa, S.; Nemmour, N.; Hirose, K. CASA-based speaker identification using cascaded GMM-CNN classifier in noisy and emotional talking conditions, *App. Soft Comp*. **2021**, 103, 107141.
[8] Cavalcanti, J.C.; Eriksson A.; Barbosa, P.A. Acoustic analysis of vowel formant frequencies in genetically-related and non-genetically related speakers with implications for forensic speaker comparison. *PLoS ONE* **2021**, 16(2).
[9] Cenceschi, S.; Meluzzi, C.; Trivilini A. The Variability of Vowels' Formants in Forensic Speech, *IEEE Instr. & Meas. Mag*. **2021**, 24, 1, 38-41.
[10] Revathi, A.; Sasikaladevi, N.; Geetha, K. Forensic investigation for twin identification from speech: perceptual and gamma-tone features and models. *Mult. Tools Appl.* **2021**.
[11] Hanifa, R. M.; Isa, K.; Mohamad, S. A review on speaker recognition: Technology and challenges, *Computers & Electrical Engineering*, **2021**, 90, 107005.
[12] Returi, K.D.; Radhika, Y.; Mohan; V.M. A Simple Method for Speaker Recognition and Speaker Verification. *In Intelligent System Design. Advances in Intelligent Systems and Computing*, Springer, Singapore; Satapathy S.; Bhateja V.; Janakiramaiah B.; Chen, Y.W. , Singapore, 2021, vol 1171. https://doi.org/10.1007/978-981-15-5400-1_64.
[13] San Segundo, E. International survey on voice quality: Forensic practitioners versus voice therapists. *Est. de Fonética Exper.* **2021**, XXIX.
[14] Teyssier, P. História da língua portuguesa, Lisboa; Sá da Costa Editora: Lisbon, Portugal, **1982**.
[15] Braid, A.C.M. Fonética Forense, 2nd ed.; Millennium: Campinas, Brazil, **2003**.
[16] Machado, T. J.; Vieira Filho, J.; De Oliveira, M.A. Forensic Speaker Verification Using Ordinary Least Squares. *Sensors* **2019**, 19, 20 4385.
[17] Esch, T., & Vary, P. Efficient musical noise suppression for speech enhancement system. *In 2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, April. (**2009**) 4409-4412.
[18] Iovan, C.; Olteanu-Raimond, A.M.; Couronné, T.; Smoreda, Z. Moving and calling: mobile phone data quality measurements and spatiotemporal uncertainty in human mobility studies. *In Geographic information science at the heart of Europe*. Springer, Cham, **2013**, 247-265.

[19] Morrison, G.T. Measuring the validity and reliability of forensic likelihood-ratio systems. *Sci. & Just.* **2011**, 51(3) 91-98.
[20] Filip, K.; Baggili, I.; Breitinger, F. WhatsApp network forensics: Decrypting and understanding the WhatsApp call signaling messages. *Dig. Invest.* **2015**, 15, 110-118.
[21] Goodman-Delahunty, J.; Dhami, M.K. A forensic examination of court reports. *Aust. Psych.* **2013**, 48 (1), 32-40.
[22] Morrison, G. S., Enzinger, E.; Hughes, V.; Jessen, M.; Meuwly, D.; Neumann, C.; Planting, S.; Thompson, W.C.; David van der Vloed, D.; Ypma, R.J.F; Zhang, C. Consensus on validation of forensic voice comparison, *Scienc. & Just.* **2021**.
[23] Zhang, C.; Morrison, G.S.; Enzinger, E.; Ochoa, F. Effects of telephone transmission on the performance of formant-trajectory-based forensic voice comparison - female voices. *Sp. Comm.* **2013**, 55, 796-813.