



SABER, arte y técnica

Minerva. Saber, arte y técnica

AÑO 9 • VOL. 1 • JUNIO 2025

Dossier Ciencias Forenses Aplicadas

ISSN en línea 2545-6245

ISSN impreso 2591-3840

Análisis y medición de atributos acústicos de los formantes del habla del español rioplatense. Un abordaje para la confección de una base de datos de referencia para las pericias forenses de voz¹

VANESA VIÑA*

Instituto Universitario de la Policía Federal Argentina (IUPFA), Argentina
vwina@universidad-policial.edu.ar

MARÍA JORGELINA PACHAME**

Policía Federal Argentina (PFA), Argentina
maria.pachame@hospitalitaliano.org.ar

RECIBIDO: 8 de marzo de 2025

ACEPTADO: 21 de abril de 2025

Resumen

Desde el año 2005, se produjo un cambio de paradigma en la ciencia forense de identificación en referencia a la discernibilidad y unicidad de las muestras. Desde entonces, surge la necesidad de lograr niveles de confiabilidad similares a los obtenidos en la identificación con ADN por medio de ensayos estandarizados con base empírica y probabilística. Las investigaciones en Acústica Forense para la identificación a través de la voz no son ajenas a este cambio, lo que hace necesario

¹ Este trabajo se basa en la investigación realizada en la Secretaría de Investigación y Desarrollo del IUPFA aprobada por Resolución (CA) N° 046/2022. Equipo integrado por Vanesa Viña, María Jorgelina Pachame y las investigadoras estudiantes tesis: Giuliana Grosso, Mora Delgado Barros, Bestard, Aime, Baez Marín, Paola, Camila Alejandra Rodríguez. Una primera versión se presentó en el Congreso Internacional Policía Científica y Criminalística- 50 Años de Formación en Ciencias Forenses.

contar con una base de datos de voces institucional regional que permita ponderar estadísticamente los atributos acústicos de una población de referencia. De este modo se podrá informar en las pericias la verosimilitud del habla de personas vinculadas a una causa judicial respecto del total de población. El objetivo de este trabajo es crear dicha base de datos, evaluar estadísticamente los atributos acústicos de las vocales para personas femeninas y masculinas de entre 20 y 60 años de edad, al menos bajo dos configuraciones: microfónicas y con codificación la plataforma de mensajería Whatsapp y ponderar la variabilidad intra-hablante entre ambos canales. En principio, se trabajó con voces de español rioplatense, pero no se descarta la posibilidad de ampliar a otros idiomas o dialectos.

Palabras clave: base de datos; voz; tipicidad; identificación forense de hablantes

Analysis and Measurement of Acoustic Attributes of Formants in Rioplatense Spanish Speech. An Approach to the Creation of a Reference Database for Forensic Voice Examinations

Abstract

Since 2005, a paradigm shift has occurred in forensic identification science regarding the discernibility and uniqueness of samples. Since then, the need has arisen to achieve reliability levels comparable to those obtained in DNA identification through standardized tests based on empirical data and probabilistic models. Research in Forensic Acoustics for voice-based identification has not been exempt from this shift, making it necessary to establish a regional institutional database that allows the statistical assessment of the acoustic attributes of a reference population. This would enable forensic experts to determine the likelihood that a given speech sample belongs to an individual involved in legal proceedings, relative to the general population. The objective of this study is to create such a database, statistically evaluate the acoustic attributes of vowels in male and female speakers aged 20 to 60 years under at least two conditions—microphone recordings and audio encoded through the WhatsApp messaging platform—and to assess intra-speaker variability across both channels. Initially, the study focuses on Rioplatense Spanish, though the possibility of expanding to other languages or dialects is not ruled out.

Keywords: database; voice; typicality; statutory definition of the offense; forensic speaker identification

Introducción

La Policía Federal Argentina realiza sistemáticamente comparación forense de locutores desde el año 1997. Desde entonces, con la certificación de la Dirección Nacional de Protección de Datos Personales, la Fonoteca o Base de

Datos de Voces Judicializadas se nutre permanentemente de la voz de quienes delinquen o bien de quienes requieren salvaguardar su identidad.² No obstante, la variedad de estas voces y su restringido número (dado que, a diferencia de la imagen del rostro y la huella dactilar, no es un requisito obligatorio para acreditar identidad) hacen que solo sean utilizadas en caso de pedido judicial de cotejo. Paralelamente, desde el año 2009 el Instituto Universitario de la Policía Federal (IUPFA) cuenta con un laboratorio destinado al análisis forense de la voz. Este laboratorio es de uso estrictamente académico y, actualmente, solo posee una única computadora cuyo software ha quedado desactualizado.

Ahora bien, la realización de una pericia de comparación forense de locutores implica hacer uso de este registro biométrico del individuo. Todo registro biométrico es un rasgo que posee universalidad (lo posee la mayoría de las personas nacidas al término de un desarrollo embriológico sin alteraciones), individualización (deben ser suficientemente diferentes en ese rasgo), estabilidad (el rasgo debe permanecer invariable a lo largo de un período de tiempo aceptable) y evaluabilidad (es factible de medir, de cuantificar) (Maltoni, 2003). De este modo, la biometría de voz es calificada y aceptada en la comunidad científica y forense a nivel mundial, como una biometría de tipo dinámica o comportamental-conductual, lo que implica la factibilidad de evaluar y encontrar regularidad en la variación natural de los elementos del habla, de allí su “dinamismo”.

Tal como se entiende actualmente, la biometría designa una tecnología de identificación y autenticación que consiste en transformar una característica biológica, morfológica o de comportamiento en una marca numérica. Su objetivo es determinar la unicidad de una persona a partir de la medida de una parte inmutable o irrepitable de su cuerpo (Foessel y Garapon, 2007) o producciones del mismo, por ejemplo, su voz.

El reconocimiento, la identificación y la verificación del hablante se basan en “la modelización estadística o matemática de las características del tracto vocal de una persona”. Este proceso representa la fisiología de la persona que produce el habla humana, expresada en una señal acústica. “Una vez que un modelo es asociado a una persona, se calcula la verosimilitud de la emisión acústica incógnita como emitida por dicho modelo en contraposición con la de otros modelos de diferentes hablantes” (Univaso, 2016, p. 1). Esto significa que toda pericia de la voz trae aparejada un proceso scopométrico (de comparación o cotejo).

² Es dable destacar que el acceso al uso (ya sea para investigación judicial o académica) de las voces registradas en la Base de Datos de Voces Judicializadas que centraliza la PFA solo es posible bajo un requerimiento judicial. Esta base de datos además de contar con voces coloquiales, contiene voces con las variaciones lingüísticas y de entorno propias de delitos federales, tales como secuestros extorsivos.

La PFA utiliza sistemas de comparación combinados. Esto implica el uso de enfoques tanto espectrográfico-auditivos como auditivo-perceptuales, fonético-lingüísticos, y automáticos y semiautomáticos. Aquí se plantea un papel determinante del especialista, tanto en la selección de los elementos para la comparación como en la interpretación de resultados que arroja la aplicación de cálculo o análisis que se emplea.

Univaso y otros (2020, p. 117) explican que, en el año 2005,

Saks y Koehler propusieron un cambio de paradigma en la ciencia forense de identificación en referencia a la discernibilidad y unicidad de las muestras, en base a las evidencias de error en pruebas realizadas y casos reales. Dicho postulado concluye con la necesidad de lograr niveles de confiabilidad similares a los obtenidos en la identificación con ADN por medio de ensayos estandarizados con base empírica y probabilística.

Por otro lado, a partir del año 2010, se inició a nivel internacional, sobre todo en Europa, un movimiento a favor de optimizar la biometría de voz y sus alcances. Por ejemplo, Interpol (2017) creó un proyecto de investigación sobre tecnología de identificación de hablantes, de cuatro años de duración (de mayo de 2014 a abril de 2018), financiado por la Unión Europea y llevado a cabo por un consorcio de diecinueve socios, entre los que figuran usuarios finales y representantes del sector privado y del mundo académico conocido como SIIP (Speaker Identification Integrated Project).³ Este proyecto inició en 2014 en el marco de una reunión celebrada en la Secretaría General de Interpol, a la que asistieron representantes de los organismos encargados de la aplicación de la ley, el sector privado y el mundo académico (Policía Federal Argentina, representada por la Lic. Vanesa Viña).

Asimismo, el investigador Andrzej Drygajlo, del Instituto de Ciencias Forenses de Suiza, preside el proyecto europeo "Pautas metodológicas para el reconocimiento automático y semiautomático de hablantes para la evaluación e interpretación de casos". Este proyecto se ha llevado a cabo en el marco del Programa Monopoly de la Red Europea de Institutos de Ciencias Forenses 2011 (European Network of Forensic Science Institutes, ENFSI, por sus siglas en inglés), Mejora de las Metodologías Forenses en Europa (IFMAE, por sus siglas en inglés) en el contexto del Grupo de Trabajo de Análisis de Audio y Habla Forense (FSAAWG, por sus siglas en inglés) de la ENFSI (Drygajlo *et al.*, 2015).

A raíz de estos cambios de paradigma y contextos, en Argentina, de manera conjunta, el CONICET y el Ministerio de Seguridad de la Nación (Gendarmería

³ <https://www.interpol.int/es/content/download/12660/file/SIIP%20Project%20Factsheet-ES.pdf>

Nacional Argentina y Policía Federal Argentina) propusieron un Protocolo Único para la Comparación Forense de Voces. Allí se reglamenta que los resultados de los informes forenses deben ser comunicados por medio de la expresión verbal de la relación logarítmica de probabilidades obtenida de los métodos aplicados (Resolución del Ministerio de Seguridad 526-24).

En este contexto, las probabilidades que debe relacionar el perito serán, por un lado, la que permita evaluar la verosimilitud de que el habla del imputado sea el habla de los audios dúbidos o cuestionados (similitud), con respecto a la verosimilitud de que el habla del imputado sea la de cualquier otra persona de una población relevante (tipicidad). Otros investigadores resaltan la importancia del uso de bases de datos representativas de la población relevante que reflejen las del caso bajo investigación, así como el empleo de modelos estadísticos (Morrison, 2014).

Cada uno de los enfoques utilizados para la realización de cotejos de voces deberá aportar su índice de verosimilitud (también conocido como Likelihood Ratio o LR). El cálculo del logaritmo de esta relación se indica como LLR, que se irá combinando con los otros parámetros hasta dar el resultado numérico final del estudio pericial.⁴

De las estadísticas internas realizadas por el numerario de la Sección Acústica Forense de la PFA, la gran mayoría de los delitos peritados son llevados a cabo por personas de habla hispana de la región rioplatense (zona de la cuenca del Río de la Plata, una extensa zona de Argentina y en la totalidad de Uruguay, y regiones aledañas). Particularmente centrada en las siguientes aglomeraciones urbanas: Buenos Aires y su área metropolitana (definición donde actualmente está incluida La Plata), Montevideo, Rosario, Mar del Plata, Santa Fe-Paraná, Bahía Blanca y Neuquén, cada ciudad con su correspondiente conurbano o área metropolitana.

Queda planteada, entonces, la necesidad de contar con bases de datos de voces con parámetros característicos acústicos representativos de una población de referencia para cada caso a analizar pericialmente, así como también dejar abierta la posibilidad de generar sistemas de detección automática de voces en investigaciones futuras.

Aunque existen estudios cuyo objetivo ha sido desarrollar una normativa de los valores formánticos (resonancias naturales características de un tracto vocal) de las vocales de hablantes argentinos del Río de la Plata, con el fin de utilizarlos en estudios de diferentes patologías vocales, como elemento clínico objetivo de

⁴ Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) (2018). *Protocolo para las pericias forenses de voz en el ámbito judicial*. <https://www.conicet.gov.ar/wp-content/uploads/Protocolo-para-la-Pericias-Forenses-de-Voz.pdf>

la medida de desviación de esos valores (Aronson *et al.*, 2000), dichos estudios no plantean el análisis de variaciones de canales de grabación, necesario en el ámbito legal.

El objetivo general del trabajo es diseñar una base de datos de voces de referencia y calibración para ser utilizada en las pericias de comparación de locutores, a partir de medir y evaluar estadísticamente el valor de centro de frecuencia de los primeros tres o cuatro formantes del habla para encontrar su tipicidad en población masculina y femenina de entre 20 y 60 años en la región rioplatense y/o otras regiones.

Los objetivos específicos, por su parte, son tres: 1) cuantificar el centro frecuencial para cada uno de los tres formantes principales del habla de las vocales A y E que se presentan en la situación de coarticulación más usual (lingualveolar-dental). 2) Tipificar el valor de cada uno de los formantes y frecuencia fundamental del habla según edad, sexo sin ser voces disfónicas. 3) Evaluar y ponderar estadísticamente la influencia en la variabilidad intrahablante en la aplicación de distintos formatos microfónicos de grabación y audios de plataforma WhatsApp.

Metodología

Enfoque conceptual

El sistema fonador humano (Figura N°1) puede reducirse básicamente a la interrelación del sistema respiratorio, la laringe que alberga los pliegues o “cuerdas” vocales y las cavidades resonantes.

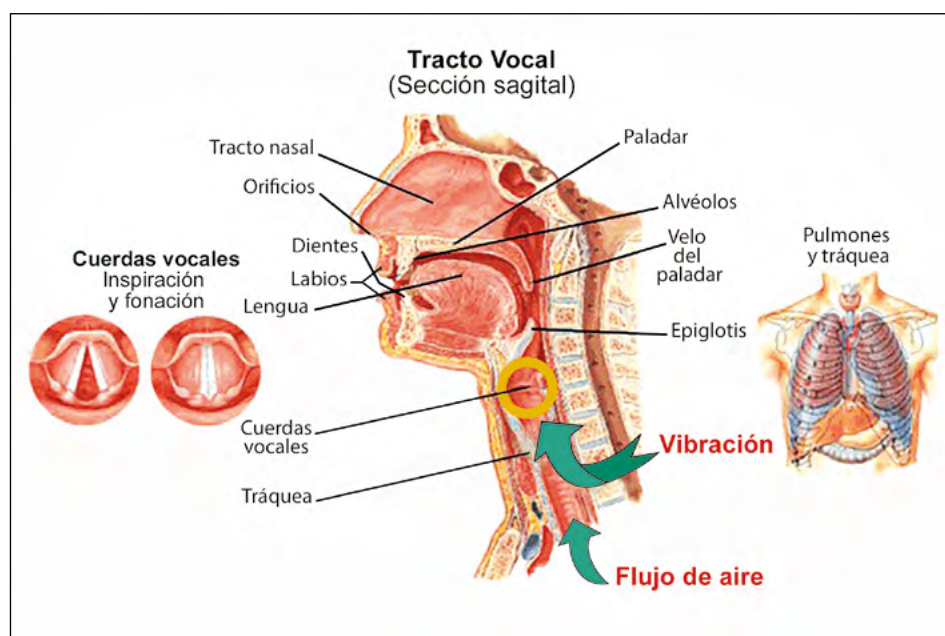


Figura N° 1. Tracto vocal humano Fuente: <https://www.pictoeduca.com/leccion/193/aparato-respiratorio/pag/1350>

La frecuencia promedio de cierre y apertura de los pliegues vocales (glotis), debido a la presión del flujo de aire traqueal, caracteriza la emisión de sonidos sonoros en las personas. A esta frecuencia se la denomina frecuencia fundamental o F0.

Aronson y otros (2000, p. 1) explican que

el espectro acústico del pulso glótico generado (espectro de la fuente glotal) será afectado por los filtros del tracto vocal, para producir sonidos vocálicos. Es decir, el tracto vocal es un filtro o un conjunto de filtros mecánicos que enfatizan algunas de las componentes del espectro glotal, especialmente, las que coinciden con sus propias frecuencias de resonancia. Esas componentes o zonas de componentes de mayor energía en el espectro resultante son los formantes del habla. Por lo tanto, la modulación del espectro de la fuente glotal y las resonancias del tracto vocal afectarán la distribución de los formantes.

Los espectros de los sonidos vocálicos están bien diferenciados entre sí y caracterizan cada vocal /i, e, a, o, u/ de las cuales /i, u/ se las clasifica como vocales cerradas y /e, o, a/ como abiertas. Si tomamos en cuenta el modo de articulación, las vocales /i/ y /u/ son altas; las /e/ y /o/ son medias; y la /a/ es baja. Según el lugar de articulación, las vocales /u/ y /o/ son anteriores, la /a/ es central y las vocales /i/ y /e/ son posteriores (Aronson *et al.*, 2000, p. 1).

Las dos primeras estructuras formánticas (F1 y F2) dependen de la posición de la mandíbula, y por la posición anteroposterior y la altura de la lengua dentro de la cavidad oral, lo que permite evaluar la inteligibilidad del habla, además de verse afectadas por la voluntad del hablante, por lo tanto, pueden imitarse. Mientras que las estructuras de los picos de resonancia superiores (F3 y F4), que dependen del largo del tracto vocal general, son menos vulnerables a la modificación voluntaria por parte del hablante, convirtiéndose en un indicio de identidad de quien las produce (Koval, 2006).

De este modo, tomar el valor del atributo de los formantes del habla permite tener acceso al registro biométrico del tracto vocal humano, independientemente del idioma en que estas vocales se emitan.

Ahora bien, el registro del valor de las estructuras formánticas puede estar influenciado por el canal y formato de grabación diferente al microfónico (Byrne y Foulkes, 2004; Hughes *et al.*, 2019; Koval *et al.*, 2010). En ese sentido se cree conveniente contar con una base de datos que permita hacer las compensaciones pertinentes en función de las estadísticas que se vayan a obtener.

Metodología y Diseño Experimental

Sujetos

En este estudio participó de manera anónima y voluntaria la comunidad educativa de entre 20 y 60 años de edad –tanto masculina como femenina– del Instituto Universitario de la Policía Federal Argentina (IUPFA) que acreditaran haber nacido en la región geográfica rioplatense.

Metodología

La investigación que da lugar a este trabajo es de tipo cuantitativa y estadística. Cuantitativa, ya que se quiere medir y evaluar numéricamente ciertos parámetros de producciones de la fuente glótica y del tracto vocal, como la frecuencia fundamental y el valor del centro frecuencial de los formantes del habla. Estadística, en virtud de que el análisis y orden de los parámetros a evaluar numéricamente se hará mediante el uso de esta rama de la matemática.

Registros

El material y las muestras a evaluar se recolectaron de la siguiente manera: se grabaron simultáneamente –mediante el micrófono de un smartphone y en audio de la plataforma de mensajería instantánea de Whatsapp– voces femeninas y masculinas de individuos de entre 20 y 60 años (rango etario que condensa la población delictiva más frecuente)⁵ de al menos un minuto de duración neta en habla continua y espontánea. Esto es que no implique lectura, sino más bien un diálogo distendido donde la naturalidad del repertorio vocal y del habla queden en evidencia.

La evaluación de los parámetros a investigar se realizó con uso del programa de PRAAT, software diseñado especialmente para hacer investigaciones en fonética, de libre distribución, código abierto, multiplataforma y, además, gratuito.⁶

A modo de control de los valores arrojados por el software de código abierto, se utilizó el programa exclusivo de uso forense SIS I de la empresa Speech Technology Center, con el que cuenta el Laboratorio de la Voz del IUPFA en una única computadora. Y se llevaron algunas muestras para control a la Sección Acústica Forense de la Dirección General de Policía Científica de PFA que cuenta

⁵ Estadística interna de Sección Acústica Forense de PFA, coincidente con la presentada por la Comisión por la Memoria en <https://www.comisionporlamemoria.org/datosabiertos/carceles/poblacion-detenido/historica/>

⁶ <https://www.fon.hum.uva.nl/praat/>

con tres computadoras con la versión SIS II STC-S521 v2. 4. 296 de dicho software. Esta versión data del año 2014 y, en la actualidad, ya no es posible su compra dado que la empresa STC comercializa la versión SIS III Software Forense de Análisis de Grabaciones.

Ante la falta de micrófonos profesionales en el IUPFA, las grabaciones se registraron en ambientes silenciosos mediante el uso del micrófono de los dispositivos smartphone particulares de las participantes en la presente investigación, con la aplicación “Grabadora de voz Fácil”, versión 2.8.2, de la compañía DIGIPOM, descargable desde la plataforma Google Play para Android. La aplicación se configuró para registrar en formato 44 KHz, códec PCM (sin pérdida de información) y extensión .wav, a un canal de grabación.⁷

Por otro lado, los audios fueron grabados mediante la aplicación WhatsApp en simultáneo con la grabación microfónica, realizada con otro dispositivo. Una vez grabados, eran enviados a través de esa misma aplicación al responsable de grupo de cada recolección, para que pudieran ser descargados y analizados.

En virtud del tiempo que requiere analizar cada parámetro acústico de cada una de las vocales en cada muestra de audio, se tomó la decisión de recortar el análisis a las dos vocales más pronunciadas por los y las hablantes, es decir: la vocal A y la vocal E en la posición más frecuente de coarticulación (linguo-alveolar/dental, lo cual incluye la articulación del grupo consonante + vocal para los fonemas l, s, d, t, r, rr). De modo que, sin desmedro de cumplir con el objetivo general, puedan obtenerse datos estadísticos amplios y abarcativos para caracterizar y tipificar a la población masculina y femenina de la región rioplatense a partir de una cantidad de fonemas plausible de analizar.

Asimismo, al hacer el análisis de aptitud técnico, se ha observado que con el canal microfónico es posible ponderar los cuatro primeros formantes del habla, no así con el recorte que digitalmente realiza el códec con que se registran los audios de la plataforma de mensajería de Whatsapp, en los que solo se visualiza hasta el tercer formante. Por esto, se decidió analizar estos tres primeros formantes únicamente para homogeneizar el estudio en todas las muestras en ambos formatos de canal de grabación.

En relación al análisis, se realizó con el programa gratuito y de código abierto Praat para obtener el fonograma, espectrograma de banda ancha y picos en

⁷ Las características de aptitud técnica del material tales como la relación SNR (Señal-Ruido) o la ausencia de *clipping* (o saturación de la señal) fueron analizadas con programas de código abierto y libre acceso como el Audacity® o el IZotope® en las PC de cada una de las participantes de la presente investigación, mientras que, por ejemplo, el análisis de tiempo de reverberación se hizo con el software de análisis forense SIS II de la Sección Acústica Forense de la Superintendencia de Policía Científica de la PFA.

LPC (Linear Predictive Code) de cada muestra, previamente calibrado con el software de exclusivo uso forense versión SIS II STC-S521 v2. 4. 296 del paquete IKARLab de la empresa Speech Technology Center (STC).

Resultados y análisis

Se han obtenido y analizado un total de 2070 fonemas E y 2068 fonemas A, para 15 locutores masculinos, incluyendo la diferencia según el formato del canal de grabación. La Figura N° 2 es de carácter ilustrativo, para que, de manera gráfica, se facilite la comprensión de lo que implica la búsqueda y medición de los centros frecuenciales de los picos de resonancia para una vocal.

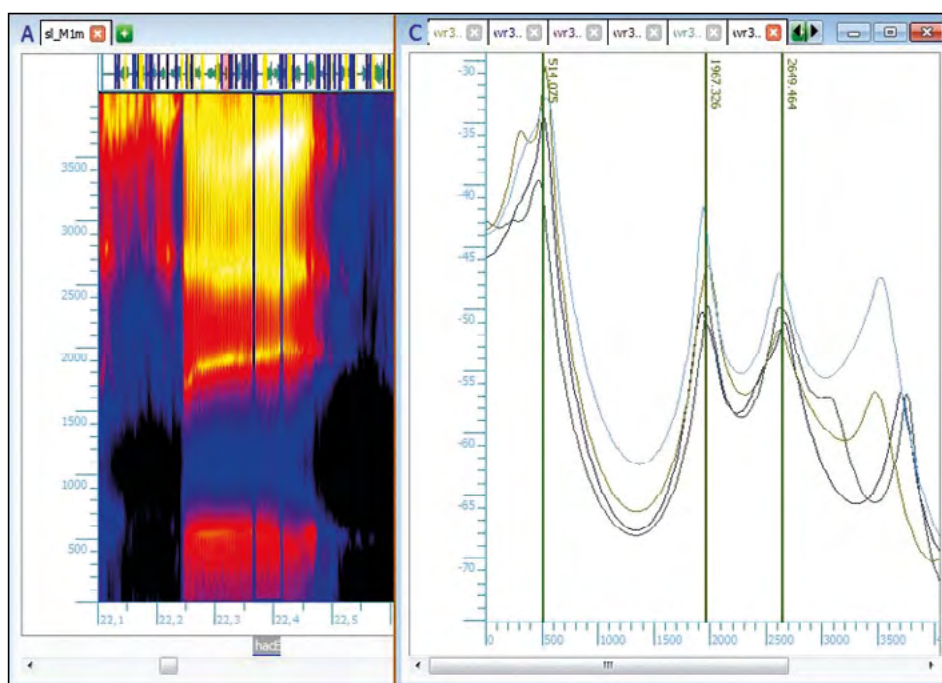


Figura N° 2. Análisis en espectrograma de banda ancha (izquierda) y LPC (derecha) en la variabilidad natural de la vocal /E/ masculina. Fuente: elaboración propia.

En relación con las frecuencias fundamentales (F0) de hablantes masculinos y femeninos (Tabla N° 1) se obtienen los resultados (Tabla N° 2) comparando entre canales de grabación microfónica (mic) y de mensajería Whatsapp (Wpp).

FEMENINO						
	FO MIC	FO MIN	FO MAX	FO WHATSAPP	FO MIN	FO MAX
promedio	185,2916667	124,25	370,7916667	185,4583333	126,375	371,4583333
DE	17,60553067	13,69941287	52,30677001	17,52756628	13,72972814	54,005619
MASCULINO						
	FO MIC	FO MIN	FO MAX	FO WHATSAPP	FO MIN	FO MAX
promedio	116,125	78,4375	233,4375	120,75	79	235,5
DE	18,22772613	13,80805924	61,09661065	27,31666158	9,872520786	61,11955497

Tabla N° 1. Estadística de valores de F0 en la población analizada.

	F0 femén Mic	F0 femén Wpp	F0 masc Mic	F0 masc Wpp
promedio	185,2916667	185,4583333	116,125	120,75
DE	17,60553067	17,52756628	18,22772613	27,31666158
variabilidad % entre canales	0,09		3,83	
promedio mic vs Wpp	185,375		118,4375	
de	0,12		3,27	

Tabla N° 2. Comparación de F0 entre canales de grabación.

Además, se evaluó el promedio de valores de F0 para las muestras femeninas y masculinas. Gráficamente se puede observar su estabilidad en función de la edad (Figura N° 3).

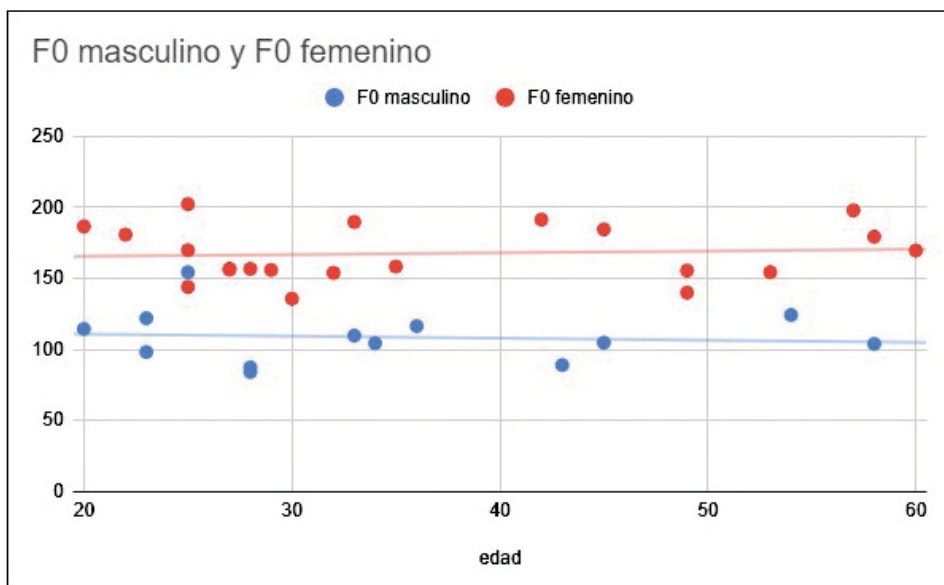


Figura N° 3. Gráfico de promedio general (mic y wpp) para la F0 en función de la edad. Fuente: elaboración propia.

En relación con la cuantificación del centro frecuencial para cada uno de los tres formantes principales del habla de las vocales A y E que se presentan en la situación de coarticulación más usual (linguoalveolar-dental), el resultado de análisis para vocal /A/ en el perfil de locutor masculino de habla español rioplatense se puede ver en la Tabla N° 3.

Masculinos			
VOCAL A TODAS LAS MUESTRAS	F1	F2	F3
PROMEDIOS GENERALES	645,1836546	1475,053031	2498,433394
DE	43,21690455	54,22958663	129,2120132
PROMEDIO POSICION LINGUO/DENTAL	F1	F2	F3
PROMEDIO	641,5036138	1492,504145	2513,863575
DE	42,99054948	55,92579507	139,1644919
PROMEDIOS MICROFÓNICOS	F1	F2	F3
PROMEDIO	643,3301322	1471,19031	2496,639269
DE	41,48152395	57,97778559	129,7720649
PROMEDIOS WHATSAPP	F1	F2	F3
PROMEDIO	647,037177	1478,915751	2500,22752
DE	46,17216775	51,81214546	132,8792029

Tabla 3. Vocal /A/ en el perfil de locutor masculino de habla español rioplatense.

A continuación, se presentan los resultados del análisis para vocal /E/ en el perfil de locutor masculino de habla español rioplatense (Tabla N°4).

Masculinos			
VOCAL E TODAS LAS MUESTRAS	F1	F2	F3
PROMEDIOS GENERALES	477,1199253	1844,452228	2566,806106
DE	33,31201619	75,67185657	113,5336805
PROMEDIO POSICION LINGUO/DENTAL	F1	F2	F3
PROMEDIO	473,8763608	1838,163199	2571,055593
DE	31,90445185	66,60410527	119,9960985
PROMEDIOS MICROFÓNICOS	F1	F2	F3
PROMEDIO	480,2732846	1845,455299	2567,272086
DE	32,98248687	78,4450972	114,5174832
PROMEDIOS WHATSAPP	F1	F2	F3
PROMEDIO	473,9665659	1843,449157	2566,340125
DE	34,41369775	75,35534938	116,2944645

Tabla N° 4. Vocal /E/ en el perfil de locutor masculino de habla español rioplatense.

En relación con voces femeninas, se han analizado el registro de 20 voluntarias, analizando un total de 1133 fonemas /E/ y un total de 1021 fonemas /A/. Los resultados obtenidos para los fonemas de la vocal /E/ se muestran en la Tabla N° 5. Y para el fonema de la vocal /A/ ver la Tabla N° 6.

Fem vocal E	F1	F2	F3
Promedio gral	525,11	2152,13	2923,15
DE	72,71	272,07	237,39
Promedio Wpp	524,84	2158,49	2971,38
DE	72,39	278,83	215,70
Promedio Mic	525,38	2145,78	2874,91
DE	73,10	265,23	248,24

Tabla N° 5. Vocal /E/ en el perfil de locutor femenino de habla español rioplatense.

Fem vocal A	F1	F2	F3
Promedio gral	724,84	1731,95	2784,24
DE	105,39	204,80	308,26
Promedio Wpp	713,66	1738,24	2826,62
DE	97,40	267,45	295,67
Promedio Mic	730,27	1725,32	2761,52
DE	113,39	209,59	334,29

Tabla N° 6. Vocal /A/ en el perfil de locutor femenino de habla español rioplatense.

Producciones de la investigación

Por un lado, cada vocal de análisis dio origen a un trabajo de tesina para graduarse en la Licenciatura en Criminalística en el Instituto Universitario de la Policía Federal Argentina (IUPFA) de cada una de las estudiantes, que participaron como integrantes del equipo de la presente investigación⁸ como se detalla a continuación:

⁸ Proyecto Análisis y medición de atributos acústicos de las formantes del habla del español rioplatense. Un abordaje para la confección de una base de datos de referencia para las pericias forenses de voz, en el marco de la Sexta convocatoria Interna de Proyectos de Investigación. (2022-2023). Res CA (68/2023).

- Análisis y medición de atributos acústicos de los formantes del habla del español rioplatense para la vocal /a/. Un abordaje para la confección de una base de datos de referencia para las pericias forenses de voz. (Báez Marín, Paola Desiré)
- Análisis y medición de atributos acústicos de los formantes del habla del español rioplatense para la vocal /e/. Un abordaje para la confección de una base de datos de referencia para las pericias forenses de voz. (Delgado Barros, Mora)
- Análisis y medición de atributos acústicos de los formantes del habla del español rioplatense para la vocal /i/. Un abordaje para la confección de una base de datos de referencia para las pericias forenses de voz. (Bestard, Aime)
- Análisis y medición de atributos acústicos de los formantes del habla del español rioplatense para la vocal /o/. Un abordaje para la confección de una base de datos de referencia para las pericias forenses de voz. (Rodríguez, Camila Alejandra)

Por otro lado, ha generado el análisis estadístico y la inclusión de registros en una base de datos que permita obtener el perfil de los atributos acústicos de la voz para hablantes masculinos de entre 20 y 60 años de la región de habla en español rioplatense.

Esto permitirá realizar el cálculo de la tipicidad. Es decir, la probabilidad de los rasgos de la voz de un material de pericia indubitado dados los rasgos de la voz de una población. Dicho de otro modo, qué tan similar es la voz de un individuo al resto de la población con la cual comparte características como edad, sexo y dialecto. Conocer la tipicidad de una voz es fundamental para redactar las conclusiones de una pericia forense de voz en la Sección Acústica Forense de la PFA. En ellas debe incluirse el índice de verosimilitud o LR según lo expresado en el “Protocolo Único para la Comparación Forense de Voces”, acorde a los Principios de Daubert para la prueba pericial.

Asimismo, se ha logrado crear un perfil de calibración (Figura N° 4) entre el programa de exclusivo uso forense SIS II STC-S521 v2. 4. 296 del paquete IKARLab de la empresa Speech Technology Center (STC) con el software de código abierto Praat®.

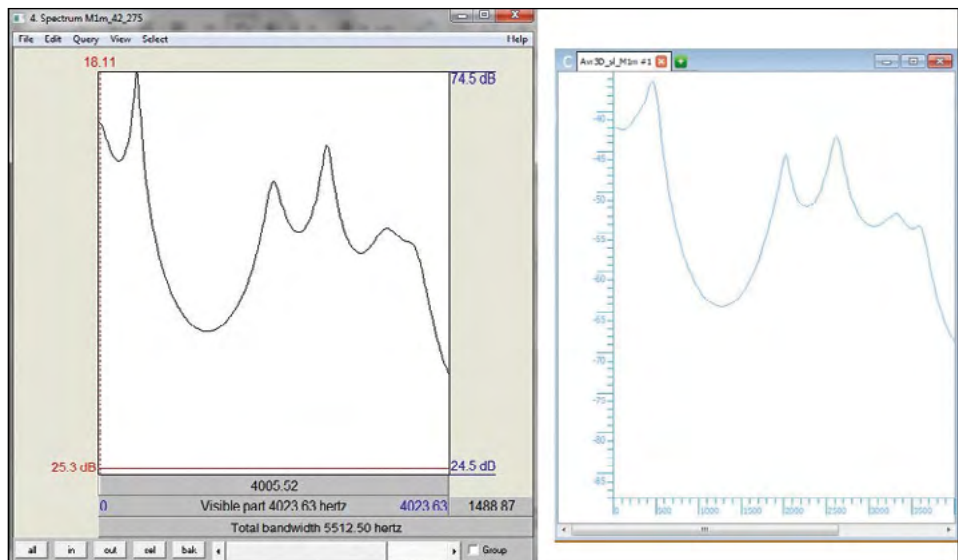


Figura N° 4. Calibración entre PRAAT (código abierto) y programa SIS II (uso exclusivo forense del paquete IKALab de la empresa STC). Fuente: elaboración propia.

Conclusiones y análisis de resultados

Se ha finalizado el diseño de una base de datos de voces de referencia y calibración para su utilización en pericias de comparación de locutores. Esta base se diseñó a partir de la medición y evaluación estadísticamente el valor de centro de frecuencia de los primeros tres o cuatro formantes del habla con el fin de encontrar su tipicidad en una población masculina y femenina de entre 20 y 60 años en la región rioplatense.

Para la población masculina, se ha logrado cuantificar e informar el centro frecuencial para cada uno de los tres formantes principales del habla de las vocales A y E que se presentan en la situación de coarticulación más usual (lingualveolar-dental), junto con su desviación estándar.

Sin embargo, para la población femenina registrada, solo pudo analizarse y cuantificar el centro frecuencial general de los tres formantes principales del habla de dichas vocales sin haber logrado indagar sobre la situación de coarticulación lingualveolar-dental.

Por otro lado, se tipificó el valor de cada uno de los formantes y frecuencia fundamental del habla según edad y sexo en voces sanas. En relación al atributo conocido como frecuencia fundamental o F0, en mujeres se observa un promedio de 185 Hz, y en varones 116 Hz para calidad microfónica y 185 Hz y 120 Hz para mujeres y varones respectivamente para calidad WhatsApp. Esto concuerda con toda la bibliografía consultada en el promedio estadístico de la frecuencia fundamental para habla masculina y femenina (F0 masculina entre 50 y 200 Hz y la femenina entre 150 y 350 Hz).

La diferencia observada en varones puede deberse a que el micrófono capta mejor las frecuencias graves dado que no presenta algoritmo de compresión, por lo cual el sistema no debe recalcular la frecuencia como ocurre en Whatsapp, sino que la toma directamente de la información acústica disponible.

La desviación estándar al tomar el tono mínimo y máximo aparenta ser mayor en mujeres que en varones, siendo de 174-173 Hz y 109-120 Hz respectivamente. Ahora bien, si tomamos los valores en Hz de las notas, observamos que la extensión tonal del habla en las mujeres es de 9.5 tonos y la de los varones es de 10.5 tonos, lo cual indica que, a pesar de que las mujeres pueden ser, de algún modo, “más musicales” al hablar, dado que varían en mayor medida de tono dentro de la frase melódica del habla, la extensión vocal en el habla (cantidad de tonos totales que producen) es muy similar entre ambos, siendo de hecho, en esta muestra, más amplia por un tono en los varones.

La diferencia entre los sistemas de análisis es mayor en la medición para mujeres, observando que la misma es de casi dos tonos diatónicos, a diferencia de esta misma medición para varones, para quienes es de un semitono a un tono. Esto probablemente radica en el hecho de que el programa de uso exclusivo forense de la empresa STC permite ser más preciso en la medición, dado que se pueden corregir las trayectorias de la frecuencia fundamental calculadas por el sistema. Corrección que no puede hacerse en el software gratuito y de código abierto Praat.

Finalmente, se ha logrado evaluar y ponderar estadísticamente la influencia en la variabilidad intrahablante –tanto en la población masculina como en la femenina– de la aplicación de distintos formatos y códecs de grabación, comparando el canal microfónico con el de audios de la plataforma de mensajería instantánea de Whatsapp. Puede concluirse que para todos los casos el porcentaje de variabilidad intrahablante entre canales es muy inferior al 5 %. Esto garantiza la homogeneidad de las muestras al realizar investigaciones cotejando voces registradas en ambos canales.

Este trabajo pone de relevancia lo minucioso que se debe ser al realizar una pericia de comparación de locutores, donde se evalúan los valores frecuenciales de los formantes del habla y el tiempo que esta tarea conlleva.

Es importante que esta investigación se amplíe, analizando mayor cantidad de voces, con el fin de obtener un corpus más representativo de la población general. Además, se recomienda extenderlo a otros dialectos o idiomas con los que convivimos, analizando todos los fonemas vocálicos y con mayor cantidad de situaciones de coarticulación que la linguoalveolar-dental. De este modo, la base de datos será más robusta.

Bibliografía

- Aronson, L.; Rufiner, H. L.; Furmanski, H. y Estienne, P. (2000). Características acústicas de las vocales del español rioplatense. *Fonoaudiológica*, 46(2), 12-20. https://sinc.unl.edu.ar/sinc-publications/2000/ARFE00/sinc_ARFE00.pdf
- Byrne, C. y Foulkes, P. (2004). The Mobile phone effect on Vowel Formants. *International Journal of Speech Language and the Law*, 11(1), 83-102. <https://doi.org/10.1558/ijsl.v11i1.83>
- Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) (2018). *Protocolo para las pericias forenses de voz en el ámbito judicial*. <https://www.conicet.gov.ar/wp-content/uploads/Protocolo-para-la-Pericias-Forenses-de-Voz.pdf>
- Drygajlo, A.; Jessen, M.; Gfroerer, S.; Wagner, I.; Vermeulen, J. y Niemi, T. (2015). *Methodological Guidelines for Best Practice in Forensic Semiautomatic and Automatic Speaker Recognition*. https://enfsi.eu/wp-content/uploads/2016/09/guidelines_fasr_and_fasar_0.pdf
- Foessel, M. y Garapon, A. (2007). *Biometría: las nuevas formas de la identidad*. Amorrortu.
- Hughes, V.; Harrison, P. T. y Foulkes, P. (2019). *Effects of formant settings and channel mismatch on semi-automatic systems in forensic voice comparison*. En *Proceedings of the International Congress of Phonetic Sciences (ICPhS)* (pp. 3129–3133). International Phonetic Association. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2019/papers/ICPhS_3129.pdf
- Interpol (julio de 2017). *Proyecto SIIP (Speaker Identification Integrated Project): Identificación de hablantes según normas estrictas de protección de la privacidad*. <https://www.interpol.int/es/content/download/12660/file/SIIP%20Project%20Factsheet-ES.pdf>
- Sartore, J. T. y Van Doren, R. (2006). *El veredicto Daubert obliga a los jueces a valorar las pruebas científicas*. *Pediatrics* (Ed. Esp.), 62(5), 747-749. <https://www.elsevier.es/es-revista-pediatrics-10-articulo-el-veredicto-daubert-obliga-jueces-13113406> (artículo en inglés: <https://doi.org/10.1542/peds.2006-0052>)
- Koval, S. (2006). *Formants matching como un método robusto para la identificación forense de hablantes*. En *Proceedings of the 11th International Conference on Speech and Computer (SPECOM)* (pp. 125-128). <https://eurasip.org/Proceedings/Ext/SPECOM2006/papers/021.pdf>
- Koval, S.; Barinov, A.; Pavel, I. y Stolbov, M. (2010). *Channel Compensation for Forensic Speaker Identification Using Inverse Processing*. En *AES 39th International Conference, Hillerød, Dinamarca*, 17–19. https://www.researchgate.net/publication/273443766_Channel_Compensation_for_Forensic_Speaker_Identification_Using_Inverse_Processing
- Maltoni, D.; Maio, D.; Jain, A. K. y Prabhakar, S. (2003). *Handbook of Fingerprint Recognition*. Springer. <https://doi.org/10.1007/978-1-84882-254-2>
- Morrison, G. S. (2014). *Distinguishing between forensic science and forensic pseudoscience: Testing of validity and reliability, and approaches to forensic voice comparison*. *Science & Justice*, 4(3), 245-256. <https://doi.org/10.1016/j.scijus.2013.07.004>
- Resolución del Ministerio de Seguridad 526-24. (14 de junio de 2024). <https://www.argentina.gob.ar/normativa/nacional/resoluci%C3%B3n-526-2024-400746/texto>
- Univaso, P. (2016). *Identificación de hablantes en Argentina: un tutorial*. <http://doi.org/10.13140/RG.2.1.4252.3768>
- Univaso, P.; Gurekian, J.; Martínez Soler, M. y Stalker, G. (2020). *FORENSIA: un sistema de identificación forense por voz*. En *Actas del 49° JAIIO - Simposio de Informática y Derecho (SID 2020)*. Sociedad Argentina de Informática (SADIO). <https://sedici.unlp.edu.ar/handle/10915/116758>
- Vázquez-Rojas, C. (2014). *Sobre la científicidad de la prueba científica en el proceso judicial*. *Anuario de Psicología Jurídica*, 24(1), 65-73. <https://doi.org/10.1016/j.apj.2014.09.001>

Cita sugerida: Viña, V. y Pachame, M. J. (2025). Análisis y medición de atributos acústicos de los formantes del habla del español rioplatense. Un abordaje para la confección de una base de datos de referencia para las pericias forenses de voz. *Minerva. Saber, arte y técnica*, 9(1). Instituto Universitario de la Policía Federal Argentina (IUPFA), pp. 26-43.

***VIÑA, VANESA**

Lic. en Física y Matemática por la Universidad Católica de Salta (UCASAL). Diplomada en Ciencias Forenses y Especialista en Lingüística. Durante siete años, se desempeñó en investigación policial en identificación de voces. Neurosicoeducadora Universidad de Buenos Aires (UBA). Diplomada en Ciencias Forenses, Instituto Universitario de la Policía Federal Argentina (IUPFA).

****PACHAME, MARÍA JORGELINA**

Licenciatura en Fonoaudiología en Universidad del Salvador (USAL). Coordinadora de la Subsección Voz de la Sección Fonoaudiología del Hospital Italiano de Buenos Aires. Staff del Sector Voz de la Sección Fonoaudiología del Hospital Italiano de Buenos Aires desde el año 2007. Personal de Planta en la Sección Acústica Forense de Policía Federal Argentina.

