

FORENSIA: un sistema de identificación forense por voz

Pedro Univaso¹, Jorge Gurlekian^{1,2,3}, Miguel Martínez Soler¹ y Germán Stalker³

¹ BlackVOX, Buenos Aires, Argentina
punivaso@blackvox.com.ar, miguelmsoler@gmail.com
<http://www.blackvox.com.ar>

² Laboratorio Investigaciones Sensoriales, INIGEM, CONICET-UBA, Buenos Aires, Argentina
anagraf99@hotmail.com

³ Programa Ciencia y Justicia, CONICET, Argentina
gstalker@conicet.gov.ar

Resumen. Se presentan los antecedentes históricos que llevaron al desarrollo informático impulsado por el Ministerio de Ciencia, Tecnología e Innovación Productiva y plasmado por el grupo de ingenieros de BlackVOX que crearon el sistema FORENSIA para su empleo en las pericias forenses de voz para la justicia y las fuerzas de seguridad. Los desarrollos informáticos asociados a las técnicas de procesamiento digital del habla, la modelación de las características distintivas del hablante, los métodos de clasificación de patrones y la construcción de bases de datos de voces, han permitido a los ingenieros informáticos y electrónicos realizar investigaciones que respondieron a una extensa controversia relacionada a la identificación de personas por su voz. Los aportes de los métodos probabilísticos automatizados superaron la desconfianza creada por la subjetividad de las evaluaciones auditivas y visuales que habían sido empleadas en las últimas décadas. La creación del Programa Ciencia y Justicia del CONICET está logrando la concientización en el Poder Judicial para lograr el perfeccionamiento de los procesos al utilizar los métodos científicos en las áreas forenses que exceden la medicina. La accesibilidad, el soporte local y el empleo de bases de datos nacionales han hecho que FORENSIA se esté empleando en la Asesoría Pericial de La Plata y en la Gendarmería Nacional Argentina. La utilización de la herramienta se complementa con un protocolo de buenas prácticas, elaborado en conjunto con los principales profesionales de las fuerzas de seguridad y judiciales del país, indispensable para el correcto accionar de los jueces y auxiliares de la justicia.

Palabras clave: Identificación forense de hablantes, comparación de voces, protocolo pericial.

1 Breve historia del reconocimiento forense de hablantes

La historia del reconocimiento de hablantes en el ámbito forense se remonta al siglo XVII en la corte de Carlos I de Inglaterra [1]. La primera investigación científica fue realizada en 1937 en el caso Lindbergh [2], donde la voz del inculcado, Bruno R. Hauptmann, fue positivamente identificado por parte del médico John F. Condon, intermediario en la negociación del rescate.

La identificación de hablantes comenzó empleando técnicas de reconocimiento por escucha, haciendo uso de la discriminación de voces por parte de los humanos. La identificación de hablantes comienza a ser practicada de forma sistemática por cientí-

ficos forenses a finales de los 40s en la Unión de Repúblicas Socialistas Soviéticas y en los 50s en los Estados Unidos de Norteamérica.

Durante los años 60, y a la luz de la confiabilidad de los sistemas de huellas dactilares, desarrollados a fines del siglo XIX, se propuso un sistema similar denominado: huella de voz [3], basado en el análisis visual de espectrogramas, el cual no alcanzó los resultados esperados dada la variabilidad intrínseca del habla. La controversia sobre la conveniencia de su uso duró más de 10 años y concluyó con el informe de Stevens et al. [4] de 1968 que determinó que el método auditivo era más preciso que la inspección visual.

A partir de los 70s, los métodos basados en humanos, contaron con el aporte de los sistemas automáticos de reconocimiento de hablantes desarrollados para aplicaciones biométricas. A pesar de que dichos sistemas evolucionaron en aplicaciones comerciales, no se pudieron crear sistemas confiables a ser empleados en el ámbito forense, existiendo una permanente controversia en su uso por parte de los expertos forenses y los científicos del reconocimiento de hablantes.

El comité de expertos convocado por la Academia Nacional de Ciencias de los Estados Unidos de América durante el año 1976 a pedido del FBI [5] concluyó que para lograr un sistema de identificación completo basado en la percepción auditiva-visual y con el empleo de métodos automáticos se deben continuar con las investigaciones y desarrollos en el área. Y sugirió que se expliciten las limitaciones de las técnicas empleadas ante cada presentación de una evidencia de identificación. Después de la publicación de este reporte el FBI dejó de brindar servicios de identificación de hablantes con el propósito de ser empleados como testimonio en las cortes judiciales norteamericanas.

En el 2003 Bonastre et al. [6] presentaron un informe alertando sobre la imposibilidad de identificar unívocamente a una persona por su voz con los avances científicos alcanzados hasta el momento, especialmente en el ámbito forense donde el entorno y los factores que afectan la performance pueden variar tremendamente con respecto al ámbito comercial.

En 2005 Saks y Koehler [7] 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. En esa misma época se comenzó a emplear una nueva metodología de evaluación de los sistemas automáticos de reconocimiento de hablantes en el ámbito forense, denominada función de costo logarítmica (Cllr) [8], la cual introduce medidas de validez y confiabilidad.

En el contexto de la identificación de hablantes en el ámbito forense, el investigador debe poder evaluar la verosimilitud de que el habla del delincuente sea la del sospechoso (similitud), con respecto a la verosimilitud de que el habla del delincuente sea la de otra persona de la población relevante (tipicidad). La Red Europea de Institutos de Ciencias Forenses (European Network of Forensic Science Institutes, ENFSI) publicó en 2015 unas directrices para estandarizar y mejorar los informes periciales (de tipo evaluativo) en el conjunto de las disciplinas forenses [9]. El reconocimiento

de hablantes figura como una de las áreas en la que son aplicables dichas directrices. El punto 2.4 de las directrices del ENFSI establece que la evaluación forense, independientemente de la disciplina concreta de la que se trate, se basará en la asignación de un ratio o relación de verosimilitud (LR de Likelihood Ratio). Esta relación mide la fuerza de apoyo que los resultados proporcionan para discriminar entre las proposiciones de interés. Generalmente estas proposiciones son la hipótesis de que las muestras de habla proceden del mismo hablante (H0 o hipótesis del fiscal) y la hipótesis de que proceden de distinto hablante (H1 o hipótesis de la defensa). Por otro lado, las LR “están científicamente aceptadas y proporcionan una forma lógica de lidiar con el razonamiento inferencial” [9:6].

En el año 2011 Dehak et al. [11] propusieron un nuevo marco de referencia para los sistemas automáticos de reconocimiento de hablantes basado en el análisis factorial como extractor de parámetros. El análisis factorial es empleado para definir un nuevo espacio de baja dimensionalidad denominado espacio de variabilidad total. En este nuevo espacio, una emisión de habla es representada por un nuevo vector denominado de factor total o i-vector. La compensación entre sesiones se realiza en este nuevo espacio de dimensiones reducidas, para lo cual se han propuesto diferentes técnicas de normalización de los i-vectores, siendo el análisis discriminativo lineal probabilístico (PLDA) [12] el que ha logrado mejores resultados. Los sistemas automáticos basados en i-vectors/PLDA son considerados en la actualidad el estado del arte.

Los diferentes enfoques empleados para la comparación o identificación forense de hablantes actualmente son los siguientes:

- Análisis espectrográfico-auditivo
- Enfoque auditivo-perceptual
- Enfoque fonético-acústico
- Métodos automáticos
- Métodos semi-automáticos
- Métodos combinados

Los métodos combinados que incluyen las principales metodologías que sustentan la identificación forense del habla (espectrográfica-auditiva, auditiva-perceptual, fonético-acústica, semi-automática y automática) son considerados por la comunidad forense de vanguardia, la alternativa metodológica de mayor fiabilidad. Cuando cualquier procedimiento científico aborda un objeto de estudio con carácter variable suele utilizar el mayor número de perspectivas de estudio posible para que las conclusiones alcanzadas posean un alto grado de objetividad. Por esta razón, se propone como el mejor acercamiento al problema la utilización de los métodos combinados, que permitan utilizar un grupo de procedimientos complementarios pero de distinta o similar naturaleza.

2 El ámbito de las pericias forenses por voz

Dentro de las principales asociaciones internacionales que estudian la identificación de hablantes en el ámbito forense se encuentra la Asociación Internacional de Acústica y Fonética Forense (IAFPA – International Association for Forensic Phonetics and Acoustics), y la Asociación Internacional de Lingüística Forense (IAFL - International Association of Forensic Linguistic), las cuales congregan a sendas reuniones anuales, y están representadas en la Revista Internacional de Habla, Lenguaje y Ley (International Journal of Speech, Language and the Law). En las mismas se congregan principalmente fonetistas, y en menor medida ingenieros de tecnologías del habla. Otra agrupación que representa a 58 laboratorios en 33 países es el Consejo Europeo de Redes de Instituciones de la Ciencia Forense (ENFSI - Board of the European Network of Forensic Science Institutions). En este grupo los ingenieros y computadores científicos están al menos tan representados como los fonetistas y lingüistas.

En el 13º Simposio de Ciencias Forenses, que tuvo lugar en Lyon, Francia, en octubre de 2001 [10] se presentó el resultado de una encuesta realizada a los gobiernos y países miembros de la INTERPOL, para conocer el estado de actividad en el campo de la lingüística forense. Se recibieron 30 cuestionarios de 21 de los 190 países miembros, de los cuales 5 laboratorios indicaron que no estaban activos en esta área. Los países que respondieron, algunos con más de una respuesta, fueron: Alemania, Austria, Belarrús, Bélgica, Brasil, Canadá, Eslovaquia, Eslovenia, España, Estados Unidos de Norteamérica (USA), Finlandia, Francia, Holanda, Israel, Italia, Lituania, Noruega, la República Checa, el Reino Unido, Suecia, y Suiza. Los resultados mostraron que los laboratorios llegan a procesar desde unos pocos casos de identificación de hablantes por año hasta varios cientos.

En la Argentina los casos de identificación de hablantes por medio de la voz solicitados por la justicia son procesados por la Policía Federal Argentina, la Gendarmería Nacional Argentina y por la Dirección General de la Asesoría Pericial del Poder Judicial en la provincia de Buenos Aires. La Policía Federal Argentina, a través del Gabinete de Identificación de la Voz, perteneciente a la División Scopometría de la Superintendencia de Policía Científica, emplea una metodología semi-automática de medición de formantes con la capacidad de detectar automáticamente los segmentos vocálicos. El sistema fue desarrollado por el Dr. Sergey Koval, reconocido como uno de los expertos mejor calificados en el mundo en el análisis forense de audio, y cuenta con más de 32 años de experiencia y más de 1.000 peritajes de audio forense ejecutados por su autoría y bajo su dirección. Es uno de los creadores de las metodologías de identificación de hablantes utilizadas por los órganos de seguridad interna de la Federación Rusa. La Policía Federal Argentina que durante el año 2001 realizó un total de 50 identificaciones de voz está encarando a partir del año 2011 un proyecto para la creación de un Banco Nacional de Voces Delictivas a incorporar al servicio que actualmente brinda. En la Gendarmería Nacional Argentina, la División Sónica, perteneciente al Departamento de Estudios Especiales, posee un grupo de trabajo en el cual los expertos cuentan con la misma tecnología rusa empleada por la policía. En la actualidad están adquiriendo el sistema automático nacional FORENSIA para complementar la medición de formantes con la metodología i-vector/PLDA, el actual estado

del arte en identificación forense. En cambio la Asesoría Pericial de la provincia de Buenos Aires emplea la metodología de identificación perceptual en conjunto con el sistema automático FORENSIA. El análisis perceptual es llevado a cabo por un equipo de fonoaudiólogos expertos para la realización de las pericias de voz que se le encomiendan. El mismo está basado en la metodología audio-perceptual desarrollada por el Dr. Harry Hollien. La cual un panel de fonoaudiólogos valoriza subjetivamente una serie de rasgos distintivos predeterminados: entonación, articulación, calidad de la voz, prosodia, intensidad, dialecto y otras características como las disfluencias y los desórdenes del habla.

La identificación de hablantes por los métodos perceptual y de formantes, empleadas por la policía, la gendarmería y la justicia, comparan las voces del sospechoso y la evidencia sin contrastarlas con otras bases universales de voces. Como se vio anteriormente, esta metodología está siendo cuestionada por los científicos de diversas ramas de las ciencias forenses, no sólo de la identificación de voces, sino también de la identificación de huellas dactilares, escritura, marcas de herramientas, de proyectiles de armas de fuego, cabellos, marcas dentales, etc., debido a los errores que se han encontrado en diversos casos judiciales. La ciencia forense tradicional de individualización o identificación se basa en una premisa central, denominada unicidad discernible, que es la de suponer que dos marcas indistinguibles deben ser producidas por el mismo objeto o persona, excluyendo cualquier otro objeto o persona en el mundo que la hubiera podido producir. Esta premisa le brinda importantes beneficios prácticos a la ciencia forense, la cual se excusa de desarrollar mediciones de atributos objetivos, de recolectar costosas bases de datos universales de frecuencia de variación de dichos atributos, analizando la independencia de los atributos o calculando la probabilidad de que diferentes objetos puedan poseer los mismos atributos observables. Esta nueva visión de la ciencia de la identificación forense es denominada por Saks et al. [7] como el “nuevo cambio de paradigma”. Con la incorporación de sistemas automáticos, como FORENSIA de BlackVOX, las instituciones de la Argentina están comenzando a apoyar este nuevo paradigma. En el país se realizan trabajos de investigación relacionados con el reconocimiento de hablantes para el ámbito forense en la Carrera de Fonoaudiología de la Facultad de Medicina de la Universidad de Buenos Aires y en el Laboratorio de Investigaciones Sensoriales, LIS (INIGEM, UBA-CONICET). El LIS desarrolló un Proyecto de Investigación y Desarrollo financiado por el Ministerio de Ciencia, Tecnología e Innovación Productiva, en el cual la institución adoptante fue la Gendarmería Nacional Argentina. El proyecto orientó a la institución adoptante acerca de los protocolos aceptados para el diseño de bases de datos de hablantes para la construcción de una base de referencia de Argentina en forma directa y por medio del canal telefónico y proveerla de entrenamiento, recursos humanos capacitados y desarrollos de software de aplicación forense.

Como en el resto de las ciencias forenses en general, es cada vez más común los procedimientos de aseguramiento de la calidad de los laboratorios que realizan estos tipos de análisis. Existen varios organismos encargados de certificar laboratorios, entre los que se encuentra el ENFSI en Europa, UKAS en Gran Bretaña, el Consejo de Acreditación en Holanda, grupos en USA como el SWGDOC (Scientific Working Group for Document Examination), o el SWGMAT (Scientific Working Group for

Materials Analysis), auspiciados por el FBI, y el SAG (Scientific Advisory Groups) en Australia y Nueva Zelanda. Uno de los primeros laboratorios que preparó la certificación de su metodología de identificación de hablantes durante el año 2001 fue el IRCGN de Francia. La IAFP inició, también ese año, un procedimiento de acreditación práctica de fonetistas forenses.

Las bases de datos de uso forense reportadas en el 13° Simposio de Ciencias Forenses [10] fueron las siguientes: DRUG (Alemania), KISTE y TELDAT (Suiza), ETH (Suiza), AHUMADA/GAUDI conformada por 450 hablantes (España), corpus de diferentes lenguas (Austria), llamadas telefónicas anónimas (República Checa), base de datos de 250 hombres y 150 mujeres (Francia), grabaciones de eventos (Israel), CGN (Holanda), LOCOPOL (Policía Científica de España), IDEM base de datos de formantes (Italia).

A pesar de los importantes avances tecnológicos en la materia y en contra de la visión simplista y superadora que nos transmiten algunas películas cinematográficas y series televisivas policíacas, podemos concluir que aunque se ha producido un gran avance en la identificación de hablantes durante estos últimos años, especialmente en algunas aplicaciones comerciales, es en el ámbito forense donde mayor resistencia encuentra en comparación con otras técnicas biométricas de mayor confiabilidad como las de ADN y huellas dactilares. El desarrollo de FORENSIA intenta mejorar la confiabilidad de las pericias de voz permitiendo combinar los diferentes métodos utilizados por los laboratorios forenses del país con un sistema de identificación automático (e.g. métodos perceptual y automático, métodos formantes y automático).

3 FORENSIA: un sistema automático de identificación forense de hablantes

FORENSIA es un sistema diseñado por BlackVOX en Argentina para cotejar muestras de voces, con el objeto de determinar identidades en el contexto de una investigación judicial. BlackVOX es una empresa de base tecnológica, incubada en el Laboratorio de Investigaciones Sensoriales (INIGEM, CONICET-UBA), que brinda herramientas tecnológicas para aplicaciones de audición y habla.

El motor de identificación automática de hablantes está desarrollado con la última tecnología basada en aproximaciones por i-vectors. El sistema realiza la comparación entre la evidencia (emisión dubitada) y la plana de voz (emisión indubitada). El resultado obtenido indica la proximidad del par de voces comparadas, que se evalúan en relación a una base de datos universal (UBM) que incluye voces de diferentes regiones de Argentina. El sistema incluye una etapa re-calibrable (PLDA) que permite compensar los desajustes entre las dos grabaciones a ser comparadas: llamadas desde teléfono fijo o celular, entrevistas por grabación directa, diferencias en los micrófonos o del ambiente de grabación, estado de las personas, paso del tiempo, etc. Para la recalibración BlackVOX selecciona datos discriminados por hablante. En la Fig. 1 podemos ver un esquema funcional de FORENSIA 2.0 y en la Tabla 1. las características técnicas del mismo [19].

Un problema al que debe prestarse atención al emplear un sistema automático de reconocimiento de hablantes es el error de calibración. A pesar que un sistema presente una tasa de error muy baja, el resultado realmente producido está sujeto a diferentes niveles de variación. Estas variaciones pueden producirse las combinaciones de canales en los que fueron grabadas las voces dubitadas e indubitadas utilizado, el cual generalmente se determina por medio de un procedimiento de calibración con datos grabados, en esos canales, de hablantes conocidos. FORENSIA 2.0 incluye calibración para los siguientes canales: teléfono de línea, celular, micrófono electrostático (o a condensador) –descrito como micrófono-, micrófono dinámico (o de bobina móvil) -descrito como Hi micrófono-, Whatsapp y YouTube. Las bases de datos empleadas para el diseño, entrenamiento y calibración de FORENSIA quedaron conformadas por más de 1.200 hablantes de 6 países (Argentina Australia, Colombia, España, Norteamérica y Uruguay) y 360.000 cotejos de voces correspondientes a más de 170 horas de grabación.

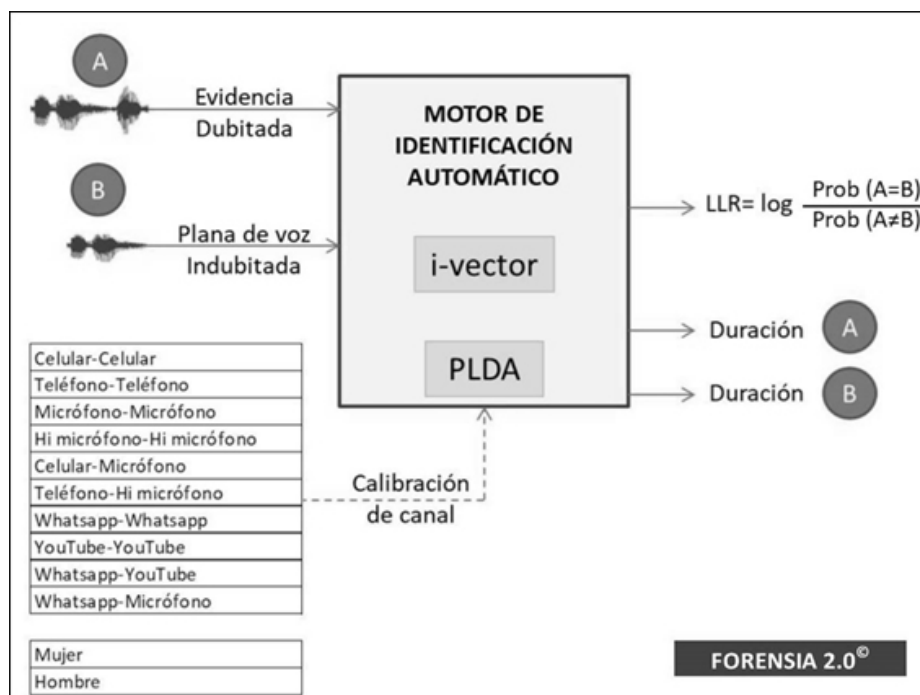


Fig. 1. Esquema funcional de FORENSIA.

Tabla 1. Características técnicas de FORENSIA.

Etapa	Descripción
Pre-procesamiento	Tramado TPDF (dithering) + filtros Wiener y RASTA-LDA
Parámetros acústicos	39 parámetros = 12 MFCC + Energía + Δ + Δ^2
Detección de actividad de habla (VAD)	Basado en energía + algoritmo heurístico de restricción de duraciones
Post-procesamiento de parámetros	Normalización cepstral media (CMN)
Modelo de referencia universal (UBM)	Modelo de mezclas gaussianas (GMM): 1024 mezclas independientes del género
Modelado de i-vectors	400 dim
Post-procesamiento de i-vectors	Normalización esférica de atributos perjudiciales (sph-Norm) + análisis de discriminación lineal (LDA)
Compensación y puntuación	Análisis discriminativo lineal probabilístico (PLDA): 300 dim p/hablante y 100 dim p/canal
Calibración	Regresión logística: duración y Cllr
Post-calibración	Hombres/mujeres: micrófono dinámico, micrófono electrostático, celular, teléfono, Whatsapp y Youtube.

Los resultados de cada comparación se expresan en términos de cocientes de verosimilitudes (LR) y en escala logarítmica (LLR), similares a los utilizados en técnicas de análisis de ADN (Fig. 2). El cociente de verosimilitudes es la relación entre la probabilidad de que dos muestras pertenezcan al mismo hablante y la probabilidad de que las mismas muestras pertenezcan a hablantes distintos. Un valor de LR alto significa una alta probabilidad de que ambas muestras pertenezcan al mismo hablante, un valor de LR cercano a 1 delimita una región de incertidumbre en la identificación, y un valor de LR cercano a 0 significa una alta probabilidad de que las muestras pertenezcan a distintos hablantes.

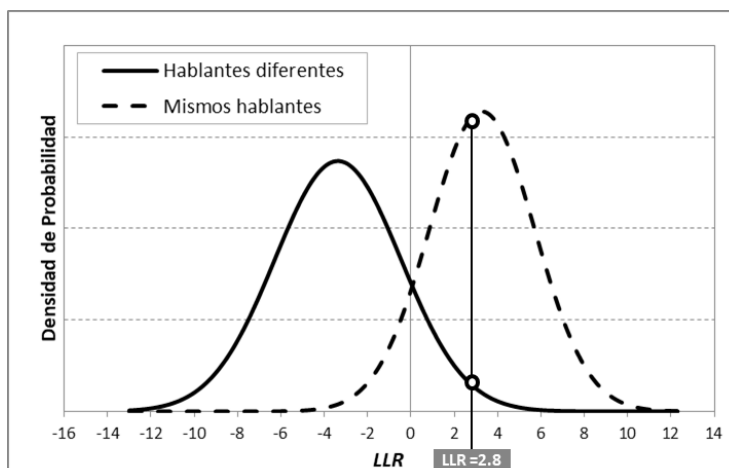


Fig. 2. Ejemplo gráfico del resultado de un cotejo de voces expresado en LLR.

3.1 Validación de FORENSIA con bases de datos internacionales

La validación de los sistemas de identificación de hablantes en Argentina es un tema de interés del Programa Nacional Ciencia y Justicia del CONICET. Para ello propone emplear bases de datos que representen los casos forenses actuales y una metodología de validación de todas las tecnologías de identificación de hablantes (e.g. auditivo-perceptuales, automáticas, semi-automáticas, etc). FORENSIA fue evaluado en base a la guía metodológica para sistemas de identificación de hablantes para casos de evaluación e interpretación elaborada por la European Network of Forensic Sciences Institutes - ENFSI [15].

Para ello se emplearon las siguientes bases de datos: 1) SITW 2016 de SRI International [16], la que incluye 648 pares de emisiones empleando canales micrófono-teléfono y 110 pares de emisiones empleando micrófono-micrófono, 2) forensic_eval_01 de Morrison [17] conformada por 1.250 pares de emisiones empleando canales de teléfono-teléfono. Las grabaciones fueron previamente editadas manualmente eliminando segmentos de habla superpuestos, ruidos y otras anomalías que se incluyen en las grabaciones originales (perturbaciones técnicas locales y sonidos no usuales). No se realizó ningún proceso de mejoramiento del audio (e.g. filtrado).

Métricas de evaluación. Las métricas de evaluación proveen un dato numérico que describe la performance del sistema de identificación en términos de precisión, confiabilidad, poder de discriminación y calibración. Nota: H_0 es la hipótesis de diferencia (hipótesis de que las emisiones a comparar han sido emitidas por diferentes personas) y H_1 la hipótesis de semejanza (hipótesis de que las emisiones a comparar han sido emitidas por la misma persona).

- **Cllr:** la función de costo logarítmica (Cllr) mide la efectiva cantidad de información que el sistema de identificación le brinda al usuario, en forma independiente de la aplicación empleada.
- **Cllr min:** es el valor mínimo que puede obtenerse del sistema de identificación.
- **Cllr cal:** es la pérdida de calibración del sistema de identificación. $Cllr = Cllr\ min + Cllr\ cal$
- **EER%:** la tasa de igual error (EER%) es la tasa de error en la que la probabilidad de falsas alarmas es igual a la probabilidad de casos perdidos.
- **PME H_0 :** probabilidad de evidencia engañosa en favor de la hipótesis H_1 . Es la probabilidad de todos los LLR que son menores de 0, sabiendo que la hipótesis H_0 es verdadera.
- **PME H_1 :** probabilidad de evidencia engañosa en favor de la hipótesis H_0 . Es la probabilidad de todos los LLR que son mayores de 0, sabiendo que la hipótesis H_1 es verdadera.

- **95% CI:** intervalo de credibilidad de que en el 95% de los casos una medición de LLR se encuentra acotada por dicho intervalo. Corresponde a la confiabilidad de las mediciones del sistema [18]

Validación. El criterio de validación presenta condiciones relacionadas con las métricas de evaluación que deben cumplirse como una condición necesaria para que el sistema de identificación sea considerado válido. Los resultados de la evaluación para cada par de canales pueden verse en las Tablas 2, 3 y 4.

Tabla 2. Evaluación y validación de FORENSIA para el canal micrófono-teléfono

Métricas de evaluación	Resultado	Criterio de validación	Cumplimiento del criterio de validación	Tipo de parámetro
Cllr	0.43	0.65	SI	Precisión
Cllr min	0.43	0.46	SI	Discriminación
Cllr cal	0.00	0.05	SI	Calibración
EER%	10%	13%	SI	Discriminación
PME H1	0.10	0.30	SI	Precisión
PME H0	0.10	0.15	SI	Precisión
95% CI	±1.38	±1.50	SI	Confiabilidad

Tabla 3. Evaluación y validación de FORENSIA para el canal micrófono-micrófono

Métricas de evaluación	Resultado	Criterio de validación	Cumplimiento del criterio de validación	Tipo de parámetro
Cllr	0.18	0.65	SI	Precisión
Cllr min	0.17	0.46	SI	Discriminación
Cllr cal	0.01	0.05	SI	Calibración
EER%	6%	13%	SI	Discriminación
PME H1	0.06	0.30	SI	Precisión
PME H0	0.06	0.15	SI	Precisión
95% CI	±1.38	±1.50	SI	Confiabilidad

Tabla 4. Evaluación y validación de FORENSIA para el canal teléfono-teléfono

Métricas de evaluación	Resultado	Criterio de validación	Cumplimiento del criterio de validación	Tipo de parámetro
------------------------	-----------	------------------------	---	-------------------

Cllr	0.46	0.65	SI	Precisión
Cllr min	0.46	0.46	SI	Discriminación
Cllr cal	0.00	0.05	SI	Calibración
EER%	8%	13%	SI	Discriminación
PME H1	0.08	0.30	SI	Precisión
PME H0	0.08	0.15	SI	Precisión
95% CI	±1.38	±1.50	SI	Confiabilidad

4 Protocolo para pericias forenses de voz en Argentina

El experto forense Daniel Corach de la Argentina ha desarrollado una guía [14] para la recepción de muestras de ADN donde expresa que “Los auxiliares de justicia y peritos actuantes habitualmente desconocen que existen pautas de procedimiento que deben cumplirse para garantizar que los resultados de las pericias forenses tengan valor de prueba al momento del juicio y que no sean declaradas nulas por errores en los procedimientos. Al recibir un indicio de interés pericial, deberán cumplirse una serie de requisitos indispensables que permitirán otorgar valor de prueba a los resultados de los estudios derivados de su análisis”. El protocolo elaborado durante el año 2018 por el Programa Nacional de Ciencia y Justicia del CONICET está basado en la guía mencionada, en los fundamentos de la identificación forense de hablantes mediante la voz [21] y en las recomendaciones de la Red Europea ENFSI [15]. Para su elaboración fueron consultados profesionales periciales de las fuerzas de seguridad y de la justicia [20]. Se presenta a continuación una síntesis del mismo.

4.1 Recepción de las grabaciones dubitadas

Regla 1: Para que una muestra de voz tenga valor como evidencia en la instancia de juicio deberá estar acompañada por un documento denominado cadena de custodia, recomendándose adicionar una función criptográfica hash (e.g. MD 5, SHA-1, SHA-2) del archivo de audio,

Regla 2: En caso de contar con peritos de parte, el solicitante deberá citar a los peritos en función de la disponibilidad de fechas del laboratorio actuante.

Regla 3: Bajo estas circunstancias, se deberán labrar actas de apertura que serán firmadas por los peritos actuantes juntamente con el integrante del laboratorio que coordina la apertura.

Regla 4: En el cuaderno de registro de entrada se deberá tomar nota de: nombre y documento de la persona que entrega el material, código del caso, si corresponde a una grabación dubitada, tipo de soporte, fecha y lugar de grabación, formato de audio, canales, nombre e identificación de los archivos, condiciones ambientales durante la toma de la grabación, observaciones.

Regla 5: Se deberá realizar una copia digital del audio recibido como dubitado con misma calidad del original para realizar el pre-procesamiento y cotejo posterior manteniendo el original sin ningún tipo de modificación.

4.2 Preprocesamiento de las grabaciones dubitadas

Regla 6: Conversión del formato de audio: en caso que la grabación esté en formato analógico es necesario digitalizarla previamente.

Regla 7: Separación del habla del hablante relevante y eliminación de silencios

Regla 8: Se deben extraer y conservar los ruidos y distorsiones locales de la señal de audio que puedan interferir el análisis, como así también los segmentos de audio que incluyan aspectos extra-lingüísticos.

Regla 9: En el ámbito legal el mejoramiento del audio (e.g. filtrado de ruidos) puede ser considerado una manipulación de la evidencia, con lo cual debe realizarse con sumo cuidado y solamente si el método de identificación de hablantes lo requiere.

Regla 10: Se considera una duración efectiva mínima de entre 3 y 7 segundos y una duración efectiva óptima de entre 15 y 30 segundos.

4.3 Condiciones para el rechazo de las grabaciones dubitadas

Regla 11: Se deberá rechazar la muestra de habla dubitada si no cumple con los requisitos de: duración efectiva mínima, nivel de ruido y/o reverberación aceptable, contemporaneidad de las muestras, diferencia de idioma o dialecto salvable, que la muestra no haya sido recogida cuando el hablante estaba severamente influido por las drogas, el alcohol, una disfonía, enfermedad, un estado emocional que pudiera alterar el resultado de la identificación, o estuviera enmascarando su voz.

Regla 12: En todos los casos se deberá incluir en el informe final un detalle de las causas del rechazo de la grabación dubitada o evidencia y consecuentemente de la pericia.

4.4 Creación de las grabaciones indubitadas

Regla 13: En esta etapa deben grabarse las voces de los sospechosos en condiciones controladas en una cámara acústica.

Regla 14: Las grabaciones indubitadas deberán contener habla espontánea del hablante en situación de diálogo con el auxiliar o perito.

Regla 15: La grabación deberá cumplir con los requisitos mínimos: Relación señal a ruido menor a 40 dB, formato riff (archivos WAV), nivel de cuantización de 16 bits y frecuencia de muestreo de 16KHz, duración efectiva mínima de 60 segundos.

4.5 Análisis y cotejo de las muestras

Regla 16: se deberá verificar que cada segmento de habla continua no haya sido editado internamente utilizando para ello programas de detección de ediciones.

Regla 17: Para las grabaciones indubitadas se deberá verificar si cumplen con las condiciones de calidad y variedad indicadas en el punto 2 de este protocolo.

Regla 18: Deberá garantizar la trazabilidad de las muestras de voz, debiendo quedar ésta documentada en el cuaderno de laboratorio correspondiente al grupo de análisis.

Regla 19: El resultado se expresará en forma cuantitativa como la relación logarítmica de probabilidades.

4.6 Informe Pericial

Regla 20: El auxiliar que realiza el informe pericial debe tener su firma registrada, conocer con precisión los términos legales que empleará y poseer capacidad de comunicación verbal y escrita de los resultados cuantitativos obtenidos para informar con claridad a la corte.

Regla 21: El informe final a ser presentado al juez o al jurado debe contener toda la información recabada en el análisis previo y la fuerza de la evidencia, que estará representada por la medida de similitud o por el cociente de verosimilitudes obtenido.

Regla 22: Se debe acompañar este resultado con todas las aclaraciones del caso y eventualmente con gráficos que permitan una fácil interpretación de los resultados. En la Fig. 3 puede verse un ejemplo del informe que genera FORENSIA.

Regla 23: El informe deberá acomodarse a los requerimientos de cada sistema legal o a las regulaciones propias del laboratorio forense, pero debería contener al menos los siguientes ítems: lista de todos los materiales recibidos, información sobre el equipamiento y metodología empleados, datos descriptivos de los archivos y audios recibidos y si fue necesario convertir o mejorar el audio, información administrativa referida a la recepción y envío de material e informes, naturaleza del caso requerido por la autoridad competente, cualquier hipótesis (proposición) referida al caso, intercambio de información entre el laboratorio y la autoridad competente, si el laboratorio realizó grabaciones del sospechoso, estrategia de análisis y una descripción detallada de la metodología empleada, observaciones y análisis de resultados, conclusión final y método de devolución del material recibido.

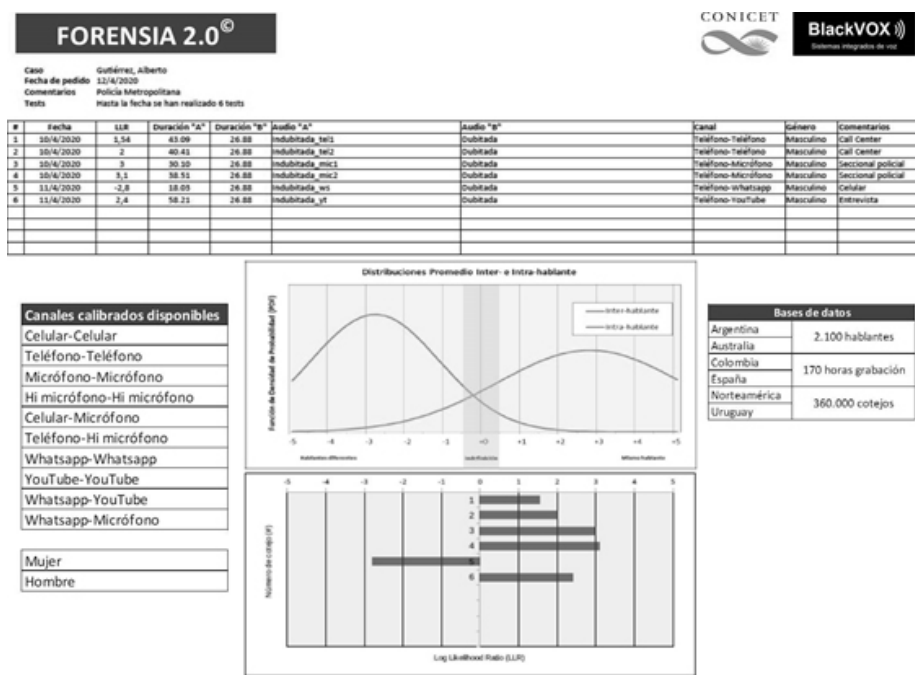


Fig. 3. Informe de una pericia de voz emitido por FORENSIA.

Referencias

1. Furui, S.: Acoustic and Speech Engineering. Kindai Kagaku-sha Publishing Company, Tokyo (1992).
2. National Research Council: On the theory and practice of voice identification. National Academy of Science, Washington, 3-13 (1979).
3. Kersta, L.G.: Voiceprint identification. Nature 196(4861), 1253-1257 (1962).
4. Stevens, K. N., Williams, C. E., Carbonell, J. R. and Woods, B.: Speaker authentication and identification: a comparison of spectrographic and auditory presentations of speech material. Journal of the Acoustical Society of America 44, 1596-1607 (1968).
5. Koenig, B., Federal Bureau of Investigation and US Dept of Justice: Speaker Identification-Part 2-Results of the National Academy of Sciences' study, FBI Law Enforcement Bulletin 49(2), 20-22 (1980).
6. Bonastre, J. F., Bimbot, F., Boë, L. J., Campbell, J. P., Reynolds, D. A. and Magrin-Chagnolleau, I.: Person Authentication by Voice: A Need for Caution. Proceedings of. Eurospeech, 1-4 (2003).
7. Saks, M. J. and Koehler, J. J.: The coming paradigm shift in forensic identification science. Science 309(5736), 892-895 (2005).
8. Brümmer, N. and Du Preez, J.: Application-independent evaluation of speaker detection. Computer Speech & Language, 20(2), 230-275 (2006).
9. ENFSI, European Network of Forensic Science Institutes: ENFSI Guideline for Evaluative Reporting in Forensic Science, http://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf [19/02/2019] (2015).

10. Broeders, A. P. A.: Forensic speech and audio analysis forensic linguistics. 13th INTERPOL Forensic Science Symposium, Lyon (Francia), 26 (2001).
11. Dehak, N., Kenny, P., Dehak, R., Dumouchel, P. and Ouellet, P.: Front-end factor analysis for speaker verification. Audio, Speech, and Language Processing, IEEE Transactions on 19(4), 788-798 (2011).
12. Prince, S. J. and Elder, J. H.: Probabilistic linear discriminant analysis for inferences about identity. Computer Vision, ICCV 2007. IEEE 11th International Conference, 1-8 (2007).
13. Potter, R. K.: "Visible Patterns of Sound", in Science, 1945.
 1. Corach D.: Pautas y Recomendaciones para la Recepción, Preservación y Devolución de Muestras Biológicas en el Marco de una Investigación Judicial. Informe técnico. Programa Ciencia y Justicia. CONICET (2018).
 2. Drygajlo A, Jessen M, Gfroerer S, Wagner I, Vermeulen J and Niemi T.: Methodological Guidelines for Best Practice in Forensic Semiautomatic and Automatic Speaker Recognition. European Network of Forensic Science Institutes (2015).
14. <http://www.speech.sri.com/projects/sitw/>
15. Morrison, G. S. and Enzinger, E.: Multi-laboratory evaluation of forensic voice comparison systems under conditions reflecting those of a real forensic case (forensic_eval_01)-Introduction. Speech Communication 85, 119-126 (2016).
16. Morrison, G. S.: Measuring the validity and reliability of forensic likelihood-ratio systems. Science & Justice 51(3), 91-98 (2011).
17. Martínez Soler, M., Univaso, P. and Gurlekian, J. A: FORENSIA - Technical Specifications. Research Gate, DOI: 10.13140/RG.2.2.36718.92488 (2018).
18. Gurlekian, J. A., Univaso, P. and Martínez Soler, M.: Protocolo para pericias forenses de voz. Research Gate, DOI: 10.13140/RG.2.2.22133.88801 (2018).
19. Rose P.: Forensic Speaker Identification, London, Taylor & Francis (2002).