

La importancia de las bases de datos en la ciencia forense: el caso de la comparación forense de voz y la población de referencia en el cálculo del LR

López-Escobedo Fernanda¹✉, Martínez Sánchez Julio César¹, Huerta-Pacheco N. Sofía^{1,2}

¹Escuela Nacional de Ciencias Forenses. Investigación Científica S/N, C.U., Coyoacán, 04510, Ciudad de México, México.

²SECIHTI. Av. de los Insurgentes Sur 1582, Crédito Constructor, Benito Juárez, 03940, Ciudad de México, México.

✉ flopeze@unam.mx

Datos del artículo

Cita

López-Escobedo F, Martínez Sánchez JC, Huerta-Pacheco NS. La importancia de las bases de datos en la ciencia forense: el caso de la comparación forense de voz y la población de referencia en el cálculo del LR. ReCiF. 2026;(1):27-40.

Editor

Vicente Torres Zúñiga

Revisión por pares

Uno

Recibido

1/octubre/2025

Aceptado

14/enero/2026

Publicado

30/abril/2026

Creative Commons

Atribución: CC-BY-NC-SA
4.0 Internacional

Resumen

Las ciencias forenses atraviesan un cambio de paradigma, impulsado por la necesidad de transitar de métodos basados en el juicio subjetivo hacia enfoques sustentados en mediciones cuantitativas y modelos estadísticos. En este contexto, el enfoque bayesiano se presenta como el marco lógico para la interpretación de la evidencia, cuyo objetivo es contrastar la hipótesis de la defensa (diferente origen de las muestras) y la del Ministerio Público (mismo origen) mediante la razón de verosimilitud (RV).

Para su construcción, la RV requiere información sobre la similitud y la tipicidad de las características que se están comparando entre las muestras. En particular, la estimación de la tipicidad requiere conocer el comportamiento de dichas características en la población de la cual proviene la muestra de origen desconocido. Es en este punto donde las bases de datos adquieren un papel central.

En este trabajo se destaca la importancia de las bases de datos en el contexto de la comparación forense de voz y se muestra cómo el Corpus de Lengua Oral del Español de México (CLOE), recopilado en la Escuela Nacional de Ciencias Forenses de la UNAM, constituye un avance en este terreno, al reunir grabaciones que reflejan la diversidad sociodemográfica y lingüística de hablantes de la Ciudad de México. A partir de un caso práctico se muestra que una selección adecuada de la población mejora la capacidad discriminativa de la RV, mientras que elecciones inadecuadas distorsionan las conclusiones. En consecuencia, el desarrollo de corpus orales estandarizados y representativos es una condición indispensable para consolidar la comparación forense de voz.

Palabras clave: Comparación Forense de Voz; Paradigma Bayesiano; Razón de Verosimilitud; Bases de datos; Evaluación de la evidencia

Abstract

Forensic science is undergoing a paradigm shift in the evaluation of forensic evidence, driven by the need to transition from subjective methods to approaches based on quantitative characteristics and statistical models. Within this context, the Bayesian paradigm provides a logical framework for interpreting evidence. Its purpose is to contrast, through the use of the likelihood ratio (LR), two competing hypotheses: the prosecution's hypothesis, which states that the sample of questioned origin comes from the same source as the sample of known origin, and the defense's hypothesis, which states that the sample of questioned origin does not come from the same source as the sample of known origin.

The likelihood ratio requires information about both the similarity and the typicality of the characteristics being compared between the samples. In particular, estimating typicality requires knowledge of how these characteristics are distributed in the population to which the offender's recording belongs. At this stage, databases play a central role.

This work highlights the importance of databases in forensic voice comparison. The Corpus de Lengua Oral del Español de México (CLOE), compiled at the National School of Forensic Sciences at UNAM, represents significant progress in this field, as it collects recordings that reflect the sociodemographic and linguistic diversity of speakers from Mexico City. A practical case demonstrates that the likelihood ratio (LR) performs better when the reference population is appropriately selected, whereas inadequate choices distort the results and compromise the LR's performance. Therefore, the development of standardized and representative speech corpora is essential for advancing and consolidating forensic voice comparison.

Key words: Forensic voice comparison; Bayesian paradigm; Likelihood Ratio; Databases; Evaluation of forensic evidence

Introducción

Las ciencias forenses atraviesan un cambio de paradigma impulsado por el uso de información cuantitativa y modelos estadísticos. En este artículo se examina el papel de las bases de datos en la comparación forense de voz, un campo en el que el juicio subjetivo de los expertos resulta insuficiente. Para ello, se expone el proceso de construcción de la Razón de Verosimilitud (RV), herramienta que permite a jueces y tribunales tomar decisiones mejor fundamentadas al valorar la evidencia disponible en un proceso judicial.

A lo largo del texto se explica cómo se construye la RV y de qué manera contribuye a determinar si un audio de origen desconocido puede atribuirse a un sospechoso. La RV se formula como un cociente de dos probabilidades, lo que pone en primer plano la necesidad de contar con bases de datos de calidad, ya que en particular el denominador depende de la representatividad de la población de referencia. Si los datos son insuficientes o presentan deficiencias, aumenta el riesgo de obtener resultados inciertos que comprometan la solidez de la interpretación pericial.

Por lo tanto, la discusión trasciende el plano técnico y se conecta con los consensos alcanzados en la comunidad científica sobre la necesidad de comunicar los resultados de manera clara y responsable.

El objetivo es que la presentación de la RV refleje, con precisión, el peso de la evidencia en contextos reales (1,2).

Antecedentes

Las bases de datos son útiles en el ámbito forense porque permiten asignar peso a la evidencia recolectada en una investigación. Un ejemplo se encuentra en la genética, en donde los datos permiten calcular la frecuencia de perfiles de ADN y, con ello, establecer la probabilidad de coincidencia entre una muestra dubitada y una indubitada (3–6). No obstante, existen otras disciplinas como la dactiloscopia, la balística y el análisis de voz, en las que las bases de datos también resultan fundamentales. Esto es especialmente cierto si consideramos a estas disciplinas como áreas que utilizan métodos de comparación de características; en genética la comparación de alelos; en dactiloscopia, la de crestas y minucias; en balística, la de estrías; y en el análisis de voz, la de parámetros acústicos.

Este tipo de métodos de comparación de características se puede clasificar en objetivos y subjetivos, según el documento *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*, presentado al presidente de los Estados Unidos, Barack Obama, en 2016 (7). Según dicho reporte:

“By objective feature-comparison methods, we mean methods consisting of procedures that are each defined with enough standardized and quantifiable detail that they can be performed by either an automated system or human examiners exercising little or no judgment. By subjective methods, we mean methods including key procedures that involve significant human judgment—for example, about which features to select within a pattern or how to determine whether the features are sufficiently similar to be called a probable match (7: p.5)”¹

En este mismo documento (7: p.88) se señala que, en el caso de dactiloscopia, el recurso más importante para impulsar el desarrollo de métodos objetivos es la generación de bases de datos que incluyan impresiones de distinta calidad e integridad. Esto se debe a que, mientras que las impresiones conocidas suelen contar con una buena calidad llegando a tener hasta impresiones de los diez dedos capturadas en un entorno controlado, las impresiones latentes encontradas en un lugar de la investigación suelen ser incompletas y de calidad variable, dependiendo de factores como la superficie y la mecánica de contacto.

Es así como, en los métodos de comparación de características surge un cuestionamiento fundamental: una vez que se ha establecido algún tipo de correspondencia entre el material encontrado en el lugar de la investigación y el material asociado con un sospechoso, ¿cuál es el tipo de base de datos más relevante que debería utilizarse para evaluar el peso de la evidencia? Por ejemplo, ¿sería preferible una base de datos con observaciones realizadas sobre material dejado en diferentes casos?, ¿o una base

¹Los métodos objetivos de comparación de características, son aquellos que consisten en procedimientos estandarizados y cuantificables con suficiente detalle como para que puedan ser realizados por un sistema automatizado o por expertos humanos ejerciendo poco o ningún juicio. Por métodos subjetivos, nos referimos a aquellos que incluyen procedimientos clave que implican un juicio humano significativo; por ejemplo, sobre qué características seleccionar dentro de un patrón o cómo determinar si las características son lo suficientemente similares como para considerarse una posible coincidencia.

de datos de observaciones realizadas sobre sospechosos de delitos de esa naturaleza? o bien, ¿existiría algún otro tipo de estudio más apropiado? (8).

Este tipo de interrogantes son precisamente las que se plantean en el ámbito de la comparación forense de voz, donde la pertinencia y representatividad de las bases de datos resultan determinantes para la adecuada valoración de la evidencia. Morrison (9) advierte que la falta de bases de datos adecuadas constituye el principal obstáculo para la realización de peritajes en comparación forense de voz. Ante la pregunta de si se han realizado estudios orientados a establecer frecuencias de referencia (como ocurre en el caso de genética) de las características acústicas que se comparan en distintas condiciones, la respuesta es negativa. El autor destaca que son pocas, o prácticamente inexistentes, las bases de datos que cumplen con los criterios generales necesarios para llevar a cabo una comparación forense de voz.

Ante esta situación, conviene reflexionar: ¿por qué las bases de datos son esenciales para la comparación forense de voz?, y ¿qué impacto tienen en el peso que se les asignan a las evidencias?

Enfoque cuantitativo en el análisis de voz

La comparación de voz, al igual que otras disciplinas forenses, atraviesa lo que varios autores han denominado un “cambio de paradigma” en la práctica pericial e.g. (10,11). Este cambio consiste en la transición de los métodos basados principalmente en la percepción y el juicio subjetivo de los expertos, hacia enfoques fundamentados en mediciones cuantitativas y modelos estadísticos. El objetivo de este cambio es que las conclusiones periciales sean reproducibles, permitan minimizar el sesgo de quien investiga y adopten una metodología que pueda ser validada por otras personas.

En el caso específico del análisis de voz, este proceso se refleja en la creciente adopción de sistemas de comparación forense que integran datos cuantitativos, técnicas estadísticas y métricas de evaluación. Por ejemplo, en el artículo publicado por Morrison (12), se menciona que actualmente nos encontramos en medio de una transformación en la manera en que se evalúa y presenta la evidencia en las ciencias forenses. Este cambio implica una transición hacia el paradigma bayesiano, particularmente en aquellas disciplinas que utilizan métodos de comparación de características, como el análisis forense de voz.

Evet (13) sostiene que el enfoque bayesiano proporciona un marco lógico para la interpretación de la evidencia, basado en un concepto probabilístico aplicado a la inferencia forense. Para cualquier científico debe ser un principio rector no especular acerca de una hipótesis sin considerar, al menos, otra hipótesis posible; este mismo criterio es igualmente válido en la ciencia forense. En este sentido, el juez no debe limitarse a la hipótesis planteada por la fiscalía o Ministerio Público (MP), debe también considerar la hipótesis de la defensa. La decisión entre una u otra debe apoyarse en la evidencia disponible y una vía para lograrlo es cuantificar comparativamente la probabilidad de cada hipótesis.

La razón de verosimilitud (likelihood ratio en inglés), constituye una herramienta central del teorema de Bayes² y ofrece un criterio formal para contrastar dos explicaciones alternativas. Por un lado, mide la probabilidad de obtener las características observadas en la muestra de origen conocido y en la

² El Teorema de Bayes es una aplicación directa de la probabilidad condicional, que permite conocer la condición en probabilidades, es decir, si conocemos la probabilidad de un evento A dado otro evento B; se podrá calcular la probabilidad de B dado A. (24)

muestra de origen cuestionado bajo la hipótesis de que ambas proceden del mismo origen (hipótesis del fiscal o MP); por el otro, cuantifica la probabilidad de obtener dichas características bajo la hipótesis de que provienen de diferente origen (hipótesis de la defensa). Esto se expresa de la siguiente manera:

$$RV = \frac{p(E|H_{mo})}{p(E|H_{do})}$$

Donde RV es la razón de verosimilitud; E representa la evidencia (por ejemplo, las características acústicas medidas y analizadas en las grabaciones); H_{mo} denota la hipótesis del fiscal o MP, que sostiene que las muestras de voz tienen el mismo origen; y H_{do} se refiere a la hipótesis de la defensa, que sostiene que las muestras de voz tienen diferente origen.

El resultado de este cociente de probabilidades respalda una u otra hipótesis. Valores mayores a 1 apoyan la hipótesis del fiscal o MP; mientras que valores menores a 1 apoyan la hipótesis de la defensa. La magnitud de la desviación de la razón de verosimilitud (RV) respecto al valor 1 determina el peso que la evidencia proporciona a las hipótesis que se están contrastando.

Por ejemplo, Evett (13) propone una escala verbal en la que los valores de RV entre 1 y 10 indican un apoyo limitado a la hipótesis del fiscal o MP; entre 10 y 100, un apoyo moderado; entre 100 y 1000, un apoyo moderadamente fuerte; 1000 a 10000, un fuerte apoyo; y mayores a 10000, un apoyo muy fuerte a dicha hipótesis (H_{mo}).

Morrison (14), por su parte, señala que resulta más conveniente convertir estos valores a logaritmos con base 10, de modo que el peso de la evidencia se exprese en una escala simétrica respecto de cero, y propone la siguiente:

<i>RV</i>	1/1000	1/100	1/10	1	10	100	1000
$\log_{10}(RV)$	-3	-2	-1	0	1	2	3

Tabla 1. Conversión de razones de verosimilitud (RV) a razones de verosimilitud logarítmicas en base 10. (14)

Ahora bien, el cálculo del numerador (es decir, de la hipótesis que postula el mismo origen) se basa en la probabilidad de encontrar determinadas características acústicas si la muestra de origen cuestionado y la muestra de origen conocido fueron producidas por el mismo hablante, lo que se suele llamar similitud. En contraste, el cálculo del denominador se basa en la probabilidad de hallar las características de la muestra cuestionada y la muestra de origen conocido, considerando que esta última pudo haber sido producida por cualquier otro hablante de la población (tipicidad). Es precisamente en este punto donde surge la necesidad de conocer la probabilidad de que determinadas características acústicas, se presenten en la población a la cual pertenece el hablante de la muestra cuestionada.

En comparación forense de voz esta población se denomina población de referencia o población relevante. Morrison y otros (1) la describen como aquella de la cual procede el hablante de la muestra cuestionada, en caso de que no fuera el sospechoso. Señalan que, dependiendo de las circunstancias de cada caso, la población de referencia puede abarcar un grupo grande de hablantes (por ejemplo, hablantes nativos de la Ciudad de México), hasta un grupo reducido (como los integrantes de una banda de secuestradores), o incluso un solo hablante distinto del sospechoso (por ejemplo, su hermano gemelo).

De acuerdo con Morrison y otros (1), el experto que realiza una comparación forense de voz debe especificar las características de la población de referencia que utilizó en el cálculo de la razón de verosimilitud, y esta población debe reflejar de manera adecuada las condiciones de las grabaciones del hablante cuestionado y del hablante conocido. Contar con una población que refleje las condiciones del caso constituye un elemento esencial en el cálculo de la razón de verosimilitud. De ahí la importancia de contar con bases de datos que sean representativas de las diferentes condiciones que puedan presentarse en un caso.

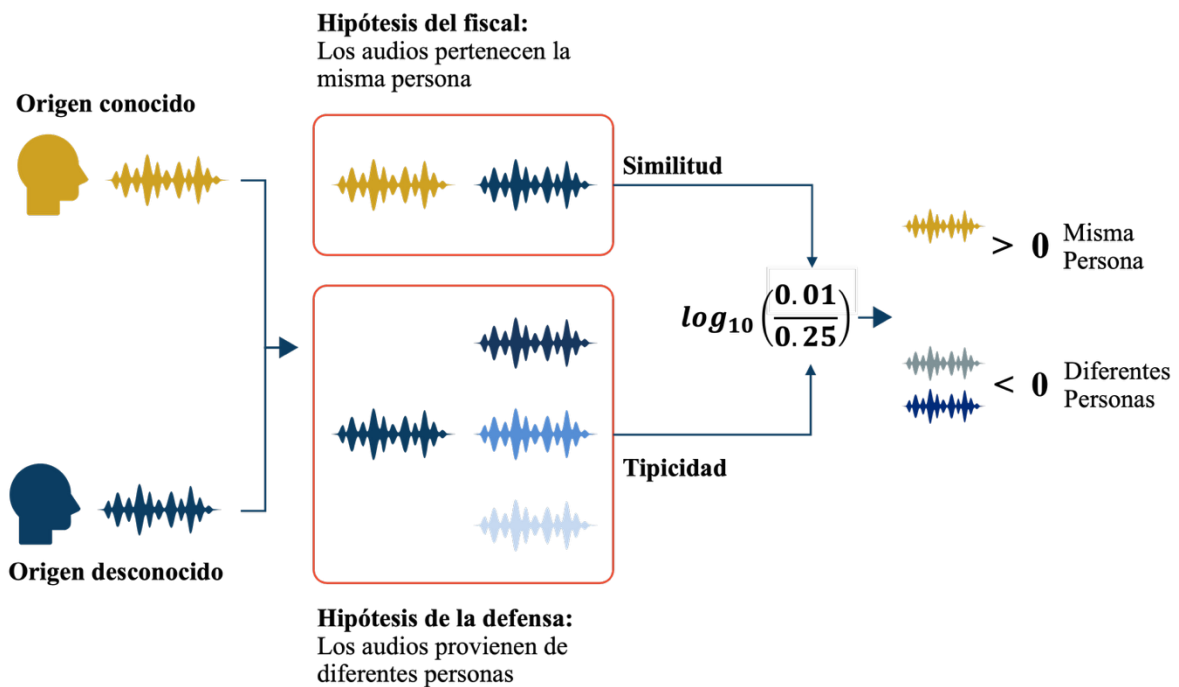


Figura 1. Ejemplo simulado de la representación del proceso de comparación forense de voz.

Ahora bien, ¿de qué forma pueden comunicarse o presentarse los resultados de una comparación forense de voz para que sean claros y útiles? En respuesta a ello, la comunidad científica especializada en este campo ha alcanzado consensos importantes respecto a la validación de los resultados en condiciones de casos reales (1). Para ello, se han propuesto métricas y representaciones gráficas como el Tippett plot o el costo del logaritmo de la razón de verosimilitud (Cllr), conocido en inglés como log-likelihood-ratio cost.

El *Tippett plot* muestra la distribución de los valores del logaritmo de la razón de verosimilitud que se obtienen al comparar muestras del mismo hablante y muestras de diferente hablante, a partir de un conjunto de datos de validación que reproducen las condiciones del caso. En este gráfico, el lado negativo del eje de las x 's representa la proporción de casos en los que se obtienen determinados valores de $\log_{10}(RV)$ cuando las muestras provienen de hablantes diferentes (diferente origen); mientras que el lado positivo representa la proporción de casos en los que se obtienen valores de $\log_{10}(RV)$ cuando las muestras corresponden al mismo hablante (mismo origen). Así, en el eje x se grafican los valores de $\log_{10}(RV)$ y en el eje y la probabilidad acumulada de dichos valores.

Una buena discriminación entre las hipótesis que se ponen en contraste (mismo origen versus distinto origen) se observa cuando las dos curvas muestran una mayor separación y presentan una

pendiente más suave, como se aprecia en la Figura 2: simulación realizada en R project (15), la cual se generó a partir de las descripciones metodológicas presentadas por Morrison (16), con el objetivo de ilustrar de manera idealizada la separación entre las hipótesis de mismo origen y distinto origen.

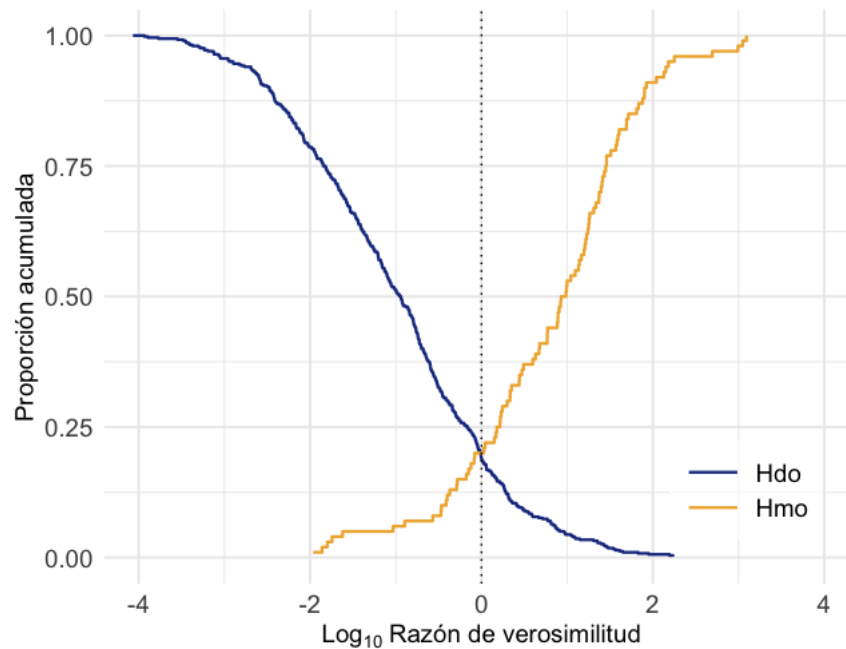


Figura 2. Ejemplo de Tippet plot con datos simulados

Además de la representación gráfica que ilustra el grado de discriminación del modelo empleado para contrastar las hipótesis a partir de un conjunto de datos de validación, se ha propuesto utilizar otras métricas complementarias. Entre ellas está el Cllr, que permite evaluar la calidad de los resultados obtenidos al comparar pares de grabaciones de un mismo hablante y de hablantes diferentes. En este sentido, cuando se comparan dos grabaciones del mismo hablante, se esperan valores de $\log_{10}(RV)$ superiores a 1; por el contrario, al comparar grabaciones de diferentes hablantes, se esperan valores menores a 1.

Para la aplicación de métricas como el Cllr, también se llevan a cabo procedimientos generales que utilizan un conjunto de pares de muestras de prueba del mismo y diferente origen, que son analizadas a través de un modelo estadístico. Estas muestras se evalúan en función del desempeño del modelo. El Cllr refleja el valor medio de los costos de ambos conjuntos de comparaciones, de modo que valores más bajos indican un mejor desempeño del modelo. En sistemas de comparación forense de voz bien calibrados, los valores de Cllr se sitúan en el rango de 0 hasta aproximadamente 1 (1).

Bases de datos para la comparación forense de voz

Desde el punto de vista estadístico, una base de datos se entiende como la codificación de los atributos de un conjunto de unidades de análisis. Tradicionalmente, se organiza en filas y columnas, donde las filas corresponden a los casos y las columnas a las variables que los describen. Para cumplir su función, una base de datos debe responder a una necesidad específica de información, reunir un número suficiente de registros que reflejen la diversidad de las unidades de estudio y mantener coherencia interna (17). En el caso de una base de datos de vehículos, por ejemplo, es recomendable

definir con claridad el objetivo de la recopilación, garantizar que se incluyan registros suficientes para mostrar sus distintas características y asegurar que todo el contenido se centre exclusivamente en ellos.

Además, existen aspectos fundamentales para asegurar la calidad de una base de datos. Uno de ellos es la reducción al mínimo de valores atípicos o faltantes, ya que su presencia puede distorsionar los resultados y afectar la fiabilidad de los análisis. También es importante documentar el proceso de construcción de la base de datos y especificar los métodos y criterios utilizados en su construcción (18). Todas estas prácticas buscan que los hallazgos, derivados del análisis estadístico, sean válidos para la generación de nuevo conocimiento.

En la comparación forense de voz, las bases de datos adquieren una forma distinta, aunque el principio de mantener una documentación adecuada es igualmente importante. En este caso, en lugar de registros tabulares, lo que se almacena son grabaciones de audio, generalmente en formato .wav, que contienen voces de distintas personas. Estas bases de datos de voz suelen denominarse corpus orales, y su particularidad nos lleva a preguntarnos: ¿cuáles son las condiciones que debe cumplir un corpus oral para ser utilizado en el ámbito forense?

Un primer aspecto se refiere a la representatividad del corpus. Es indispensable considerar las características sociodemográficas de los individuos que serán grabados, por lo que es fundamental que las grabaciones representen de manera adecuada a la población que se pretende estudiar. Esto implica incluir voces de hombres y mujeres, abarcar distintos rangos de edad y contemplar variables adicionales como el origen geográfico, nivel educativo u otros factores que permitan una caracterización más completa de los individuos (19).

El corpus oral también debe ser representativo del tipo de habla que se analiza. Una grabación puede provenir de habla espontánea, como ocurre en una conversación entre dos personas, de la lectura de un texto o de una interacción telefónica. La literatura ha mostrado que estas diferencias generan variaciones en las frecuencias acústicas, lo que repercute en los análisis posteriores. Por ejemplo, Nakamura y otros (20) demostraron que la variabilidad de los parámetros acústicos de los fonemas producidos por un hablante es mayor en el habla espontánea que en el habla leída. Por este motivo, los corpus deben incluir grabaciones obtenidas de distintos tipos de habla y, en el ámbito forense, es fundamental considerar que las grabaciones cuestionadas suelen provenir de llamadas telefónicas, lo cual también debe estar representado en el corpus oral.

Ahora bien, en lo que respecta a la calidad de las grabaciones, aunque en el ámbito forense estas suelen ser de baja calidad, en la construcción de un corpus oral se busca obtener registros de buena calidad que sirvan como referencia. Una grabación de buena calidad puede adaptarse posteriormente a condiciones técnicas menos favorables, pero no es posible realizar el proceso inverso. Por esta razón, se busca que las grabaciones estén libres de ruidos externos o interferencias que disminuyan su claridad. Idealmente, la recolección debe llevarse a cabo en ambientes controlados, como cabinas de audios insonorizadas, y utilizando equipos de grabación que minimicen el ruido medioambiental (21).

Además, el corpus debe contar con un número suficiente de grabaciones que permitan realizar los análisis de manera adecuada. Esto implica asegurar un balance entre las distintas clases representadas en el corpus y garantizar que dichas muestras sean representativas de la población de hablantes a la que se refieren; en este caso, del español de México.

Finalmente, el corpus oral debe cumplir con los aspectos éticos y legales relacionados con el uso de datos personales, por lo que cada grabación debe estar acompañada de su respectivo consentimiento informado.

Caso aplicado: comparación forense de voz

Para comprender con mayor claridad el papel que desempeña el corpus oral en los resultados de una comparación forense de voz, es necesario partir de un hecho clave: la escasez de corpus orales de acceso público. La falta de estos recursos limita la posibilidad de llevar a cabo confrontas sistemáticas, ya que no existe una base que integre de manera suficiente la diversidad acústica del país. En este contexto, el Corpus de Lengua Oral del Español de México³, desarrollado en la ENaCiF, constituye una alternativa relevante, pues reúne grabaciones de hablantes mexicanos que abarcan distintos rangos de edad y condiciones socioeconómicas, con el propósito de ofrecer un banco de voces representativo para aplicaciones forenses.

A partir de este marco, podemos ilustrar su importancia con un caso concreto con datos reales. Para ello, se trabajó con ocho participantes voluntarios — cuatro hombres y cuatro mujeres —, cada uno con cuatro grabaciones: dos en habla espontánea y dos en lectura, lo que sumó 32 audios. Con estas grabaciones se construyeron pares de grabaciones, donde un audio se trató como voz dubitada y el otro como indubitada, que fueron procesados y analizados a través del modelo Gaussian Mixture Model–Universal Background Model (GMM-UBM) (16) en R project (15) y Python (22). El GMM-UBM es un método que analiza un conjunto de vectores numéricos, los cuales describen las características espectrales de la voz del hablante en distintos rangos de frecuencia en intervalos cortos de tiempo. A partir de estos vectores, el sistema construye un modelo de fondo que representa las características acústicas de una población. Cuando se tiene un hablante específico, el modelo se adapta para capturar su forma particular de hablar. Así, cuando se analiza una voz de origen desconocido, el sistema evalúa qué tan probable es que las características acústicas provengan del modelo adaptado (23).

Para analizar el papel del corpus oral (base de datos) como población de referencia utilizamos el CLOE y evaluamos los dos escenarios. En el primero, empleamos exclusivamente voces femeninas tanto en los pares cuestionado–no cuestionado como en la población de referencia. En el segundo, mantuvimos las mismas voces dubitadas e indubitadas, pero utilizamos un conjunto mixto —hombres y mujeres— como población de referencia. Esta comparación permitió contrastar un escenario, donde la población de referencia coincide con las características de la voz dubitada y otro donde esa coincidencia no se cumple.

La estimación de similitud y tipicidad se realizó mediante el modelo GMM-UBM (16), una técnica utilizada en el reconocimiento automático de hablantes y aplicado también en la comparación forense de voz y a otras ramas de la ciencia forense. Este modelo calcula la razón de verosimilitud (RV) a partir de la generación de vectores de datos que obtiene de toda la señal sonora.

En este caso, la voz cuestionada corresponde a una mujer, por lo que lo más adecuado es que la población de referencia esté conformada por mujeres con características sociodemográficas similares a

³ El CLOE México se recopiló en el marco del proyecto PAPIIT IA401419 por el Seminario de Lingüística Forense de la Escuela Nacional de Ciencias Forenses, UNAM. Para mayor información y consultar sus estadísticas https://www.enacif.unam.mx/selifo/index.php/cloe_mexico/

las de la grabación cuyo origen se encuentra en duda. En estas condiciones, los resultados tienden a ser coherentes, como se aprecia del lado izquierdo en la Figura 3.a, la distribución acumulada de las razones de verosimilitud ($\log_{10}(RV)$) evidencia una separación clara entre las hipótesis: bajo la hipótesis del fiscal, las probabilidades se concentran en valores positivos, mientras que bajo la hipótesis de la defensa se acumulan en valores negativos. Además, la voz cuestionada versus voz conocida, da como resultado $\log_{10}(RV)$ de 2.5, lo que indica una evidencia moderadamente fuerte a favor de que la voz cuestionada pertenezca al hablante femenina de la voz conocida.

En contraste, si no se utiliza una población de referencia adecuada e, incluso, se consideran voces masculinas, se generan distribuciones atípicas que distorsionan la evaluación y reducen la solidez de las conclusiones. En la parte derecha de la Figura 3.b, se observa cómo las curvas de Tippett plot presentan un solapamiento más evidente. La línea correspondiente a la hipótesis del fiscal se desplaza hacia valores negativos, mientras que la curva asociada a la hipótesis la defensa se mantiene en la zona negativos. Esta superposición reduce la capacidad de discriminar entre hipótesis y debilita la interpretación probabilística del resultado, lo cual se corrobora con el valor de $\log_{10}(RV)$ de -0.63, que indican un apoyo limitado a la hipótesis del fiscal.

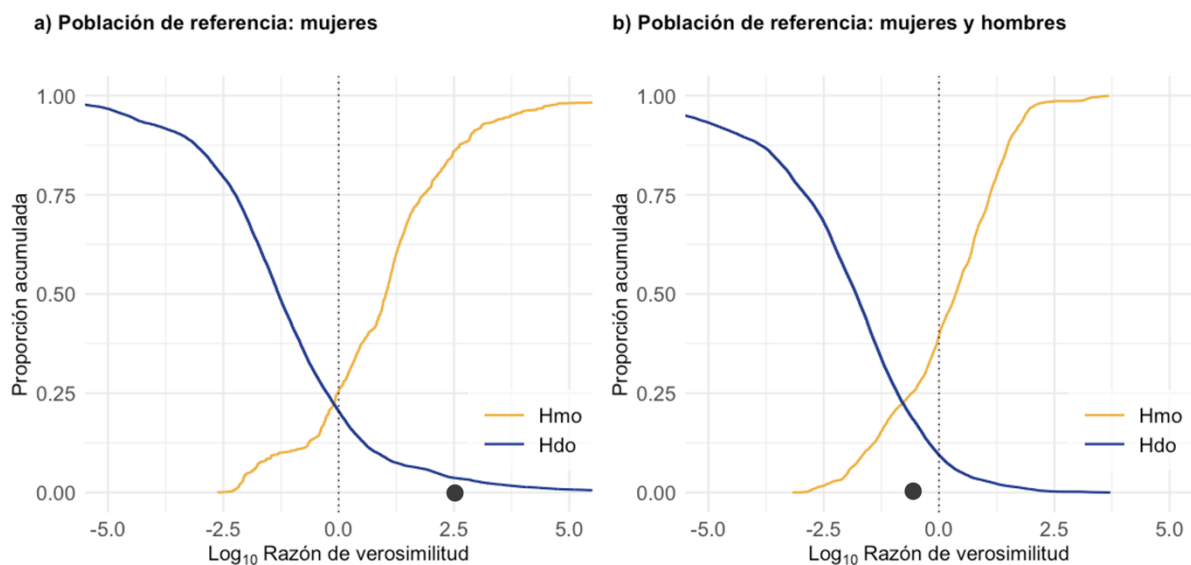


Figura 3. Comparación de datos reales de una voz cuestionada femenina versus voz conocida femenina con dos poblaciones referencia: a) Tippett plot con una población de referencia conformada por mujeres; y b) Tippett plot con una población de referencia conformada por mujeres y hombres.

Esto significa que una mala selección de la población de referencia en el caso analizado provocaría que, al comparar las muestras, la probabilidad de apoyar la hipótesis de la fiscalía o del Ministerio Público (que ambas voces tienen el mismo origen) sea menor únicamente por el hecho de haber utilizado una población inadecuada en el cálculo de la tipicidad. Es importante señalar que una buena selección de la población de referencia sólo es posible si se cuenta con un corpus oral representativo de la población, de modo que la elección pueda realizarse de manera adecuada.

Discusión y conclusiones

A lo largo de este documento se ha expuesto la importancia que tienen las bases de datos en el ámbito forense y, en particular, en la comparación forense de voz. Esta relevancia se ha evidenciado a partir del paradigma bayesiano, adoptado por diversas ciencias forenses y, de manera específica, por la comparación forense de voz. Dicho paradigma ofrece un marco lógico para la interpretación de la evidencia, basado en un enfoque probabilístico aplicado a la inferencia forense.

Este cambio ha permitido la transición de métodos sustentados principalmente en la percepción y el juicio subjetivo de los expertos, hacia enfoques fundamentados en mediciones cuantitativas y modelos estadísticos. De este modo, las conclusiones periciales resultan reproducibles, permiten minimizar el sesgo de quien investiga y se apoyan en una metodología susceptible de ser validada por otras personas.

No obstante, como se señaló previamente en el texto, para que este proceso sea posible resulta indispensable conocer la tipicidad de las características que se están midiendo y a partir de las cuales se ponen en contraste las dos hipótesis consideradas en el marco de la razón de verosimilitud: la hipótesis de la fiscalía, que sostiene que las muestras tienen el mismo origen, y la hipótesis de la defensa, que sostiene que las muestras no tienen el mismo origen.

Para determinar esta tipicidad es necesario contar con bases de datos adecuadas que, en el caso del análisis forense de voz, consisten en grabaciones orales que representen la diversidad acústica de la lengua en cuestión. En nuestro caso, el interés se centra en el español de México.

Mediante un ejemplo práctico, en el que se recurrió al Corpus de Lengua Oral del Español de México (CLOE), recolectado en la ENaCiF, se demostró que, si bien la razón de verosimilitud estimada a través del modelo GMM-UBM (16) constituye un marco sólido para emitir resultados objetivos y probabilísticos, su fiabilidad puede verse comprometida por la falta de representatividad de la base de datos empleada en el cálculo.

En México nos enfrentamos al problema de la carencia de corpus orales que representen de manera adecuada la diversidad acústica de la población mexicana que sean de acceso público y que puedan ser utilizados en evaluaciones periciales. La gran diversidad lingüística, sociocultural y demográfica del país hace indispensable contar con bases de datos capaces de reflejar esa heterogeneidad. Sin este esfuerzo, los modelos estadísticos corren el riesgo de sustentarse en muestras reducidas o sesgadas que no corresponden a las condiciones reales de los hablantes.

El desafío es considerable, pues la comparación forense de voz converge en un principio esencial: sin bases de datos pertinentes, la razón de verosimilitud pierde fiabilidad como herramienta interpretativa. Incluso los modelos estadísticos más sofisticados dependen de la validez empírica de los datos de referencia. Por ello, la construcción de corpus orales representativos y estandarizados constituye una condición indispensable para garantizar que el análisis forense de voz proporcione evidencia sólida en el ámbito judicial.

En este sentido, el futuro de la comparación forense de voz no solo depende del perfeccionamiento de los modelos estadísticos, sino, de la consolidación de bases de datos que reflejen la diversidad lingüística y contextual de cada comunidad. Solo a través de este esfuerzo colectivo será posible fortalecer la solidez científica de la disciplina y garantizar que sus aportaciones tengan un impacto real en la búsqueda de justicia.

Agradecimientos

Este trabajo ha sido realizado gracias al proyecto IH-2025-I-106 apoyado por la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI) en el año 2025. Además, este trabajo forma parte de una investigación realizada desde un enfoque estadístico gracias al Programa UNAM-PAPIIT IA101726. Finalmente, la tercera autora agradece el respaldo del proyecto SECIHTI 7144.

Disponibilidad de los datos

Los datos y los códigos empleados en los análisis de este estudio están disponibles a solicitud dirigida a la autora de correspondencia.

Uso de herramientas de Inteligencia Artificial

En la elaboración del presente artículo se ha empleado una herramienta basada en Grandes Modelos del Lenguaje (LLM) ChatGPT (OpenAI) como apoyo en diversas fases de la investigación y redacción. Su uso ha sido de carácter estrictamente complementario, con el objetivo de optimizar la eficiencia y la calidad del documento, y en ningún momento ha sustituido el trabajo intelectual, el juicio crítico ni la responsabilidad académica de los autores. Por lo que la autoría y responsabilidad final sobre la totalidad del contenido recae, de manera íntegra, en los autores de este trabajo.

Las funciones para las que se utilizó este tipo de herramienta fueron: el apoyo en la redacción y mejora del estilo del texto, la reformulación de ciertos pasajes para lograr una mayor claridad expositiva y la verificación de coherencia terminológica a lo largo del documento.

Referencias

1. Morrison GS, Enzinger E, Hughes V, Jessen M, Meuwly D, Neumann C, et al. Consensus on validation of forensic voice comparison. *Science & Justice*. 2021;61(3):299–309.
2. Bali AS, Edmond G, Ballantyne KN, Kemp RI, Martire KA. Communicating forensic science opinion: An examination of expert reporting practices. *Science & Justice* [Internet]. 2020;60(3):216–24. Disponible en: <https://www.sciencedirect.com/science/article/pii/S1355030619302102>
3. Balding DJ, Nichols RA. DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. *Forensic Sci Int* [Internet]. 1994;64(2):125–40. Disponible en: <https://www.sciencedirect.com/science/article/pii/0379073894902224>
4. Budowle B, Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM. Population Data on the Thirteen CODIS Core Short Tandem Repeat Loci in African Americans, U.S. Caucasians, Hispanics, Bahamians, Jamaicans, and Trinidadians. *Journal of Forensic Sciences* 1972–2005 [Internet]. el 1 de noviembre de 1999;44(6):1277–86. Disponible en: <https://doi.org/10.1520/JFS14601J>

5. Budowle B, Shea B, Niezgoda S, Chakraborty R. CODIS STR Loci Data from 41 Sample Populations. *Journal of Forensic Sciences* 1972-2005 [Internet]. el 1 de mayo de 2001;46(3):453–89. Disponible en: <https://doi.org/10.1520/JFS14996J>
6. Steele CD, Balding DJ. Choice of population database for forensic DNA profile analysis. *Science & Justice* [Internet]. 2014;54(6):487–93. Disponible en: <https://www.sciencedirect.com/science/article/pii/S1355030614001336>
7. President's Council of Advisors on Science, (PCAST) T. Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods [Internet]. 2016 sep. Disponible en: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf
8. Champod C, Evett IW, Jackson G. Establishing the most appropriate databases for addressing source level propositions. *Science & Justice* [Internet]. 2004;44(3):153–64. Disponible en: <https://www.sciencedirect.com/science/article/pii/S1355030604717086>
9. Morrison G. Response to the Forensic Science Questions Posed by the President's Council of Advisors on Science and Technology. 2015.
10. Koehler JJ, Mnookin JL, Saks MJ. The scientific reinvention of forensic science. *Proceedings of the National Academy of Sciences* [Internet]. el 10 de octubre de 2023;120(41):e2301840120. Disponible en: <https://doi.org/10.1073/pnas.2301840120>
11. Morrison GS. Advancing a paradigm shift in evaluation of forensic evidence: The rise of forensic data science. *Forensic Sci Int* [Internet]. 2022;5:100270. Disponible en: <https://www.sciencedirect.com/science/article/pii/S2589871X22000559>
12. Morrison GS, Enzinger E. Introduction to forensic voice comparison. 2019;
13. Evett IW. Towards a uniform framework for reporting opinions in forensic science casework. *Science & Justice*. 1998;3(38):198–202.
14. Morrison GS. Measuring the validity and reliability of forensic likelihood-ratio systems. *Science & Justice*. 2011;51(3):91–8.
15. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria; 2025. Disponible en: <https://www.R-project.org/>
16. Morrison GS. A comparison of procedures for the calculation of forensic likelihood ratios from acoustic-phonetic data: Multivariate kernel density (MVKD) versus Gaussian mixture model-universal background model (GMM-UBM). *Speech Commun*. 2011;53(2):242–56.
17. Wang J, Liu Y, Li P, Lin Z, Sindakis S, Aggarwal S. Overview of Data Quality: Examining the Dimensions, Antecedents, and Impacts of Data Quality. *Journal of the Knowledge Economy* [Internet]. 2024;15(1):1159–78. Disponible en: <https://doi.org/10.1007/s13132-022-01096-6>
18. Enders CK. Applied missing data analysis. Guilford Publications; 2022.
19. Hughes V, Foulkes P. What is the relevant population? Considerations for the computation of likelihood ratios in forensic voice comparison. En: *Proceedings of Interspeech 2017*. 2017.

20. Nakamura M, Iwano K, Furui S. Differences between acoustic characteristics of spontaneous and read speech and their effects on speech recognition performance. *Comput Speech Lang* [Internet]. 2008;22(2):171–84. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0885230807000459>
21. Bottalico P, Codino J, Cantor-Cutiva LC, Marks K, Nudelman CJ, Skeffington J, et al. Reproducibility of Voice Parameters: The Effect of Room Acoustics and Microphones. *Journal of Voice* [Internet]. 2020;34(3):320–34. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0892199718304338>
22. Van Rossum G, Drake FL. Python 3 reference manual [Internet]. CreateSpace; 2009. Disponible en: <https://dl.acm.org/doi/book/10.5555/1593511>
23. Reynolds DA, Quatieri TF, Dunn RB. Speaker Verification Using Adapted Gaussian Mixture Models. *Digit Signal Process* [Internet]. 2000;10(1):19–41. Disponible en: <https://www.sciencedirect.com/science/article/pii/S1051200499903615>
24. Canavos GC, Medal EGU. Probabilidad y estadística. McGraw Hill México; 1987.