

HACIA LA EVALUACIÓN PERCEPTUAL FORENSE

TRANSFORMACIÓN DE PUNTAJES DE SIMILITUD A RELACIONES DE VEROSIMILITUD

PARTE 1: HABLANTES MASCULINOS

TOWARDS FORENSIC PERCEPTUAL ASSESSMENT

TRANSFORMATION OF SIMILARITY SCORES TO LIKELIHOOD RATIOS

PART 1: MALE SPEAKERS

Molina N.¹
Suligoy S.²
Masessa E.³
Torres H.⁴
Univaso P.⁵
Gurlekian J.⁶

¹. Doctora en Fonoaudiología, Facultad de Medicina, Universidad de Buenos Aires.

². Licenciatura en Fonoaudiología, Universidad del Salvador, Gendarmería Nacional Argentina.

³. Licenciatura en Fonoaudiología, Universidad de Buenos Aires.

⁴. Laboratorio de Investigaciones Sensoriales, Instituto de Inmunología, Genética y Metabolismo, Universidad de Buenos Aires.

⁵. Doctor en Ingeniería, Universidad de Buenos Aires.

⁶. Ingeniero Electrónico. Laboratorio de Investigaciones Sensoriales. Programa Ciencia y Justicia CONICET. Argentina.

Correspondencia: nancyemolina@fibertel.com.ar

Resumen: Objetivo: obtener las distribuciones discriminantes de igual/diferentes hablantes para la tipificación del modelo de voces masculinas para el cálculo del cociente de verosimilitudes (LR). Desarrollo experimental: se inicia con el desarrollo de una base de datos de frases que considera la grabación de 2520 frases en distintas sesiones y grabaciones de habla espontánea de 40 hablantes nativos de la Argentina. Se crearon pares de frases provenientes del mismo hablante y de distintos hablantes combinando frases diferentes que provenían de diferentes sesiones. Se realizó la evaluación perceptual de los pares donde los oyentes debían indicar si las voces provenían de un mismo hablante o de distintos hablantes indicando el grado de confianza en la respuesta. Resultados: a partir de las respuestas obtenidas de la clasificación y el grado de confianza se obtuvieron las funciones de densidad de probabilidad para pares provenientes del mismo hablante y de diferentes hablantes. Se comparan las respuestas acumulativas de 1 a 5 oyentes para comprobar el efecto del número de evaluadores en las distribuciones obtenidas. Conclusiones: La evaluación perceptual en este trabajo permitió construir un modelo de las distribuciones de las respuestas a voces del mismo hablante (mismo origen) y de distintos hablantes (distinto origen) para una población de hablantes nativos masculinos. El número de evaluadores óptimo oscila entre 4 y 5 y se obtuvo mediante el análisis de errores. Con las distribuciones obtenidas será posible realizar la conversión del puntaje resultante de las comparaciones de la evidencia en los cocientes de verosimilitudes LR y LLR denominados en el ámbito forense como "la fuerza de la evidencia".

Palabras clave: ámbito forense, identificación de hablantes masculinos, evaluación perceptual auditiva, fuerza de la evidencia.

Abstract: Objective: to obtain the discriminant distributions of equal/different speakers for the typification of the model of male voices for the calculation of the likelihood ratio (LR). Experimental development: it starts with the development of a database of sentences considering the recording of 2520 sentences in different sessions and recordings of spontaneous speech of 40 native speakers from Argentina. Pairs of sentences coming from the same speaker and from different speakers were created by combining different sentences coming from different sessions. Perceptual evaluation of the pairs was carried out where listeners had to indicate whether the voices came from the same speaker or from different speakers indicating the degree of confidence in the response. Results: from the responses obtained from the classification and the degree of confidence, probability density functions were obtained for pairs coming from the same speaker and from different speakers. Cumulative responses from 1 to 5 listeners are compared to check the effect of the number of evaluators on the obtained distributions. Conclusions: The perceptual evaluation in this work allowed us to construct a model of the distributions of responses to voices from the same speaker (same origin) and from different speakers (different origin) for a population of male native speakers. The optimal number of evaluators ranges between 4 and 5 and was obtained by error analysis. With the obtained distributions it will be possible to convert the score resulting from the evidence comparisons into the likelihood ratios LR and LLR, known in the forensic field as "the strength of the evidence".

Keywords: forensics, male speaker identification, auditory perceptual evaluation, strength of evidence.

1. INTRODUCCIÓN

En el contexto forense, el grado de similitud entre una grabación desconocida de carácter ofensivo y la voz de un hablante sospechoso que es conocido, se denomina evidencia (2). Actualmente, en el reconocimiento forense del hablante, se utilizan métodos auditivos, semiautomáticos, automáticos y una combinación de estos para estimar esta similitud.

En la identificación de hablantes, se debe poder estimar la probabilidad de que el habla desconocida sea la del sospechoso (similitud), y estimar también la probabilidad de que el habla desconocida sea la de otra persona de la población relevante (tipicidad), empleando para ello resultados cuantitativos (15). Esto significa que no basta con establecer si dos muestras son similares o diferentes, sino que, en el caso de que se asemejen, es importante establecer cuanto se diferencia la muestra dudosa del resto de la población relevante. Para lograr este objetivo se aplica en la pericia un enfoque bayesiano, que expresa las probabilidades en forma de relaciones de verosimilitud (LR: likelihood ratio) (8, 4).

La evaluación perceptual auditiva es la herramienta más antigua y con mayor uso para determinar la similitud entre las voces (11). El “timbre” se considera como el principal atributo de la voz que puede identificar o diferenciar a dos hablantes. Ha sido definido por los estándares ANSI/ASA (American National Standards Institute, 2015) como un atributo multidimensional de la sensación auditiva que permite al oyente determinar si dos emisiones de habla son de diferente o el mismo origen, cuando no se considera la percepción de sonoridad y altura tonal (5). Los seres humanos tienen una capacidad notable para distinguir a los individuos por sus voces. Cada emisión vocal es distinta a otra cualquiera, aunque sea producida por la misma persona y de la forma más parecida posible a otra realizada con anterioridad. Incluso el habla registrada puede llegar a contener matices diferenciadores al ser reproducida o transmitida con distintos medios, en distintos instantes temporales o en distintos espacios acústicos.

Hay diferentes factores que dificultan las calificaciones perceptuales de la voz; el conocimiento de los mismos y sus efectos en la señal de habla a analizar, permiten valorizar la confiabilidad de los resultados obtenidos. Se detallan a continuación los factores más preponderantes.

El principal factor es la variación intra-hablante. Las características identificatorias que presenta la voz del hablante proceden de su anatomía y de sus hábitos articulatorios aprendidos: estos aspectos no son constantes y poseen un grado de variabilidad que determina la posibilidad de realizar una comparación correcta.

También hay otros factores no menos importantes, como las emociones (stress, cólera, tristeza, depresión, excitación, euforia y alegría), estados inducidos por agentes externos (intoxicación por drogas o alcohol), comportamientos intencionales (engaño, falsedades, enmascaramiento, insolencia, prevención), estados de salud (resfrío, gripe, fatiga) y envejecimiento del hablante (8).

Un factor externo al hablante es la calidad de las muestras. En el ámbito forense se suele trabajar con muestras de voz obtenidas a partir de grabaciones telefónicas captadas en el canal de transmisión durante una investigación autorizada por un juez. Puesto que el canal telefónico es un filtro pasa banda que sólo permite pasar las frecuencias entre los 300 Hz y los 3500 Hz; este canal reduce la información acústica. Además, las muestras de la voz interceptada, no sólo poseen esa calidad que podríamos denominar “de tipo telefónico”, sino que, en muchas ocasiones, conllevan ruidos externos procedentes de la calle o, incluso, el solapamiento de distintas voces, si la llamada se ha realizado desde un espacio público (4). Sin embargo, existen estudios que señalan que el reconocimiento auditivo es robusto a los cambios en las condiciones de grabación (1). Bregman en 1990, afirma que el sistema auditivo humano es capaz de adaptarse a los efectos del enmascaramiento al ruido y otras distorsiones (2).

Las características del oyente son tan importantes como las del hablante y su entorno, al momento de analizar las variables que afectan la identificación. Existen factores internos intrínsecamente inestables de los oyentes (por ejemplo, fallas en la memoria, la atención, fatiga y errores), la restricción de la memoria auditiva para el reconocimiento, el sesgo del oyente frente al conocimiento de los antecedentes médicos, la experiencia del evaluador (que disminuye o aumenta la confiabilidad), si tiene formación profesional (por ejemplo, en otorrinolaringología, fonoaudiología, profesores de canto, fonetistas y estudiantes de pregrado o postgrado), formación musical; si son oyentes nativos ya que se desempeñan mejor que los no nativos (1) y el entrenamiento en el juicio perceptivo auditivo. No debemos olvidar también que en la tarea pericial los expertos muchas veces disponen de un tiempo limitado para llevar a cabo un peritaje. Realizar un examen completo de una muestra de voz implica varias escuchas sucesivas y detalladas; además, en un contexto judicial, existen por lo general varias voces que el experto debe analizar (4).

Otro factor a tener en cuenta es la capacidad del oyente para identificar muestras de habla del mismo hablante no contemporáneas. La no-contemporaneidad se refiere a muestras de habla de un mismo hablante extraídas en distintos instantes de tiempo, las cuales serán posteriormente usadas en un proceso de identificación. El envejecimiento produce cambios fisiológicos del hablante que se manifiestan en la voz de varias formas: disminución de la frecuencia fundamental, cambios en el timbre o calidad tonal, junto a una creciente inestabilidad del tono y la intensidad del habla, incremento de la respiración, y disminución de la precisión articulatoria (19).

Con relación a los oyentes, un factor que se estudió es la influencia del género en los resultados de identificación correcta. En el trabajo realizado por Hollien, 2012 (14) no encontraron diferencias entre los resultados obtenidos por hombres y mujeres actuando como oyentes y empleando el método auditivo-perceptual.

Otro factor importante es la familiaridad del oyente con el hablante a identificar. Se considera que una voz es familiar cuando se la ha escuchado frecuentemente por un período de tiempo de al menos dos años (8). Los seres humanos a menudo son capaces de reconocer una voz familiar después de escucharlo durante unos segundos. Las máquinas superan a los humanos en expresiones largas, pero en expresiones muy breves, las máquinas aparentemente funcionan peor que los humanos; por lo tanto, es capacidad de los humanos poder distinguir entre voces desconocidas de corta duración (1). Los métodos perceptuales sobresalen cuando la evidencia tiene una duración efectiva menor a 8-10 segundos donde los métodos automáticos no son aplicables.

En esta investigación se utilizan frases de menos de 5 segundos y la separación entre estímulos del par a comparar es menor a 4 segundos para favorecer la memoria de corto plazo.

Los resultados obtenidos en este trabajo se completarán más adelante con una planilla de evaluación perceptual de atributos semejante a la descrita por Hollien, 2012 (14) y se incluirá la información de tipicidad.

Los métodos indirectos (15) para el cálculo de la relación LR consisten en dos transformaciones:

1. Convertir el conjunto multidimensional de los atributos perceptuales incluyendo la tipicidad en un puntaje de la evidencia.

2. Transformar el puntaje de la evidencia en la relación de verosimilitudes LR.

En este trabajo se tratará la segunda transformación. Para ello se estimarán las distribuciones que permitirán asociar el puntaje de la evidencia con el valor LR llamado fuerza de la evidencia. Las distribuciones que se presentarán corresponden a las respuestas promedio de los oyentes cuando comparan voces de diferentes hablantes masculinos y de voces provenientes del mismo hablante. Esta información permitirá a los peritos en evaluación perceptual presentar una respuesta que considere tanto la similitud como la tipicidad de las voces que representa la fuerza de la evidencia. La cuantificación de la fuerza de la evidencia permitirá comunicar el puntaje obtenido en la evaluación de la evidencia en función de las distribuciones obtenidas.

2. DESARROLLO EXPERIMENTAL

Se eligieron tres frases del párrafo “*Los Sentidos*” (13) que tienen una duración máxima de 4 segundos. Se presentan los textos, las transcripciones fonéticas con el alfabeto SAMPA (12) y el número de grupos entonativos.

Las frases son:

1. ¿Sabías que el cuerpo humano tiene sensores? ¡Si! ¡Como los robots!

[saBiah ke el kwerpo wmano posee sensores, si, komo loh roBots].

Posee tres grupos entonativos. Uno en modalidad interrogativa y dos en modalidad exclamativa.

2. Algunos nos mantienen en equilibrio, otros actúan cuando comemos algo picante.

[alGunoh noh mantjenen en ekiliBrjo, otros aktuan kwando komemos alGo pikante].

Posee dos grupos entonativos en modalidad enunciativa.

3. Un conocimiento que tenemos desde pequeños son los sonidos de nuestra lengua.

[un konosjmnto ke tenemoh DehDe pekeJos son los soniDoh De nwehtra leNgwa].

Posee un grupo entonativo enunciativo.

Estas frases fueron producidas por cada hablante en forma de entrevista dirigida (leídas y luego repetidas sin leerlas) en tres sesiones diferentes (cambiando el día y la hora). Los sujetos fueron 40 hablantes adultos masculinos del español de Buenos Aires. La grabación se realizó en formato wav sin compresión, en 44.100 Hz, 32bits, canal mono. Una vez obtenidas las grabaciones, se construyeron los pares a evaluar utilizando distintas frases obtenidas de distintas sesiones. Se prepararon 40 pares provenientes del mismo hablante y 800 pares con voces provenientes de diferentes hablantes. El total de 840 pares fue listado en forma aleatoria y dividido en 10 grupos de 84 pares cada uno para cada sesión de evaluación a los fines de no fatigar a los evaluadores. A partir de una base de 15 evaluadores en total, cada grupo fue evaluado por 5 evaluadores diferentes utilizando una interfase gráfica diseñada para este objetivo como la indicada en la Figura 1.

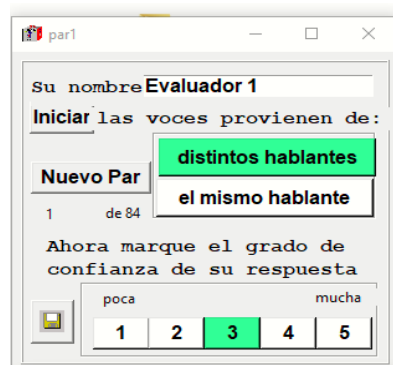


Figura 1. Interfase gráfica del usuario para responder sobre el origen de los pares y marcar el grado de confianza. Imagen propiedad autores.

La instrucción a los evaluadores fue “Usted oirá dos frases mediante auriculares, su tarea consiste en indicar si provienen del mismo o diferentes hablantes y luego marcar el grado de confianza con que ha dado su respuesta”. Las respuestas numéricas de los evaluadores al grado de confianza se multiplicaron por +1 cuando eligieron el mismo hablante y por -1 cuando eligieron distintos hablantes.

2.1 MODELADO PROBABILÍSTICO

Se empleó una variante del “método de puntajes” (*scoring method*) (9) conformado por tres etapas. En la primera etapa se determinó una medida de similitud (puntaje perceptual) entre los pares de emisiones, la cual puede tomar valores en el rango $[-5,+5]$, donde los valores negativos representan emisiones de diferentes hablantes (hipótesis H_1) y los positivos del mismo hablante (hipótesis H_0), representando su magnitud el grado de confianza de la evaluación. A partir de las estimaciones de las distribuciones de probabilidad de ambas hipótesis se determinaron los cocientes de verosimilitud (LR – *Likelihood Ratio*).

En una segunda etapa se calcularon las distribuciones de densidad de probabilidad univariadas, una representa las verosimilitudes entre emisiones que pertenecen a los mismos hablantes (MIS) y la otra a diferentes hablantes (DIF). Dichas distribuciones se estimaron con la fórmula de densidad de función núcleo (*KDE - Kernel Density Estimation*) desarrollada por Silverman, 1986 (17) y expresada por la siguiente ecuación:

$$K(\theta | x_i, \lambda) = \frac{1}{\lambda s \sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{(x_i - \theta)^2}{\lambda^2 s^2}\right) \quad (1)$$

Siendo D el conjunto de datos, correspondiente en nuestro caso a las medidas de similitud, conteniendo k elementos, $D = x_1, x_2, \dots, x_k$, donde s es la varianza de la muestra, θ el valor medio del elemento y λ el parámetro de suavización a ser elegido. Aitken y Taroni (2004) sugieren que este parámetro puede elegirse subjetivamente basándose en la experiencia del personal científico. En nuestro caso se consideró un valor de $\lambda=0.3$. Quedando la estimación de la distribución de densidad de probabilidades como:

$$f_j(\theta | D, \theta) = \frac{1}{k} \sum_{i=1}^k K(\theta | x_i, \lambda) \quad (2)$$

Finalmente, la distribución final se calibró considerando a la función de costo logarítmica C_{llr} (Brümmer, 2004) como función objetivo.

En la tercera etapa se determinó el cociente de verosimilitudes logarítmica (LLR) entre ambas hipótesis en función de la medida de similitud y se la aproximó por medio de una regresión lineal. Esta última ecuación permite calcular el aporte, expresado en LLR , de cada parámetro al sistema de identificación de hablantes final, quedando expresado como:

$$LLR = \log \frac{f(x_i | H_0)}{f(x_i | H_1)} \cong a \cdot x_i + b \quad (3)$$

Para representar cotejos en los que participan diferente cantidad de evaluadores, se aplicó esta metodología considerando como medida de similitud el promedio de los puntajes correspondientes a un mismo par de frases en grupos de 1, 2, 3, 4 y 5 evaluadores.

3. RESULTADOS

Como resultado de la primera etapa del método propuesto, en la Figura 2 se pueden ver los histogramas de frecuencia de los puntajes perceptuales promedios correspondientes a los mismos hablantes (MIS) y a diferentes hablantes (DIF) para un conjunto de 5 evaluadores. Los puntajes perceptuales promedio de cada par de frases se obtuvieron promediando los resultados obtenidos por cada uno de los 5 evaluadores. A partir de estos valores se calcularon los cocientes de verosimilitud (LR) y su logaritmo (LLR) correspondientes a cada puntaje perceptual como puede verse en la Tabla I. En la misma se muestra la equivalencia entre estos valores y una interpretación numérica que facilita la presentación de la fuerza de la evidencia al juzgado. En este caso se consideraron dos hipótesis competitivas: la de la fiscalía y la de la defensa. La fiscalía propone que ambas grabaciones fueron emitidas por la misma fuente y la defensa

por diferentes personas, pudiéndose presentar el resultado como una cantidad de veces más probable de alcanzar una de estas hipótesis con respecto a la otra.

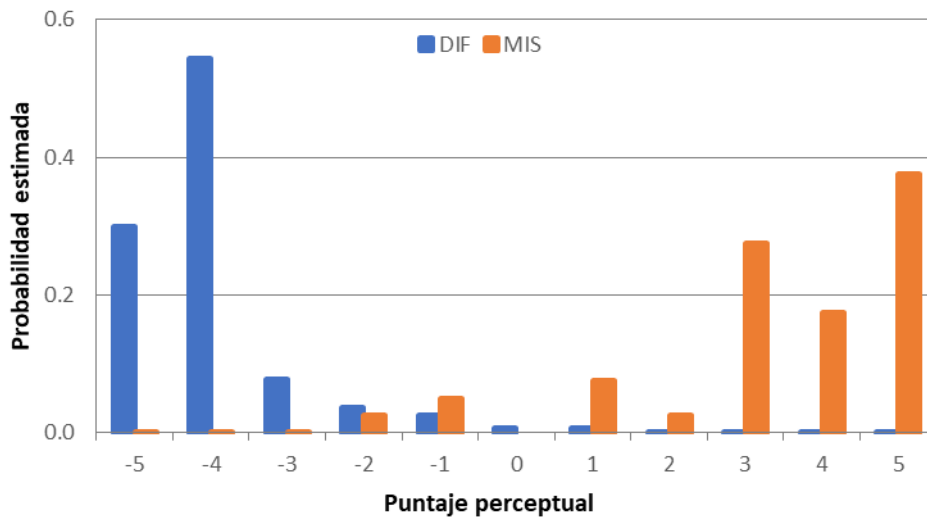


Figura 2. Histograma de las probabilidades obtenidas para los puntajes de -5 (diferente origen - DIF) hasta +5 (mismo origen - MIS) para los 840 pares por parte de una base de 5 evaluadores por par. Imagen propiedad autores.

Puntaje perceptual	LLR	LR	veces	
-5	-2.37	0.00	234	a favor de la defensa
-4	-1.81	0.02	65	
-3	-1.25	0.06	18	
-2	-0.69	0.20	5	
-1	-0.13	0.74	Indefinido	
0	0.43	2.69	3	a favor de la fiscalía
1	0.99	9.77	10	
2	1.55	35.48	35	
3	2.11	128.82	129	
4	2.67	467.74	468	
5	3.23	1698.24	1698	

Tabla 1 Cociente de verosimilitud (LR) y su logaritmo (LLR), para 5 evaluadores, en función del puntaje perceptual y su interpretación numérica. Propiedad autores.

En la Figura 3 se muestran las distribuciones del cociente de verosimilitud logarítmica (LLR) al comparar voces masculinas correspondientes al mismo hablante (MIS) y a diferentes hablantes (DIF). Las mismas se estimaron por medio de la fórmula de densidad de función núcleo (KDE) de acuerdo a lo expresado en (2).

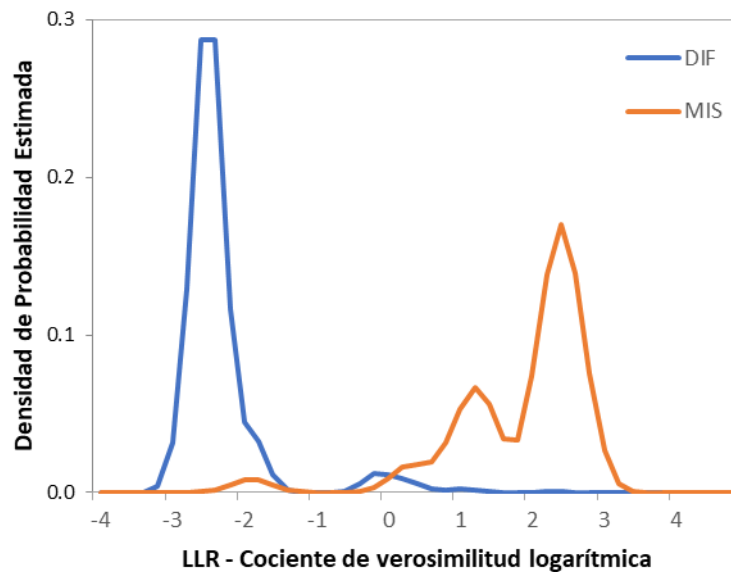


Figura 3 – Distribución del cociente de verosimilitud logarítmica (*LLR*) del mismo hablante (MIS) y de diferentes (DIF) hablantes.
Imagen propiedad autores.

3.1 EFECTO DEL NÚMERO DE EVALUADORES.

La evaluación del desempeño perceptual se realiza a partir de dos métricas: una directa como la proporción de igual error (*EER – Equal Error Rate*) referida a los errores por pérdidas y falsas alarmas y otra más sutil llamada la función de costo de la relación de verosimilitudes logarítmica (*Cllr –function cost of log-likelihood-ratios*). La métrica de *EER* es utilizada frecuentemente como indicador global, pero no evalúa la calibración, es decir la habilidad de los evaluadores de fijar umbrales de decisión correctos como ocurre con *Cllr*, que informa además sobre la pérdida de información. Cuanto más grande es este valor menor es la información que se obtiene de las comparaciones. En particular, el valor mínimo de *Cllr* nos habla de la mayor información que puede rescatarse de cada par comparado. (16).

$$Cllr = \frac{1}{2} \left[\sum_{t=mis} \frac{\log_2(1 + e^{-LRt})}{Nm_{is}} + \sum_{t=dis} \frac{\log_2(1 + e^{-LRt})}{Ndis} \right] \quad (4)$$

Donde *LRt* es la relación de verosimilitudes para el par *t*, *Nmis* es el conjunto de pares del mismo origen y *Ndis* es el conjunto de pares de distinto origen. Las sumatorias normalizadas representan las expectativas o los costos logarítmicos de los pares con el mismo origen y de los pares de distinto origen respectivamente. Para el desarrollo analítico de la ecuación (4) (20).

Estas dos métricas fueron calculadas para comparar el desempeño en función del número de evaluadores. En las figuras 4 y 5 se observan las respuestas acumuladas desde 1 a 5 evaluadores. Se observa que a pesar que el porcentaje de *EER* disminuye con la cantidad de evaluadores, el número óptimo para el rescate de información de pares iguales y diferentes se alcanza entre 4 y 5 evaluadores siguiendo el criterio del valor mínimo de *Cllr*.

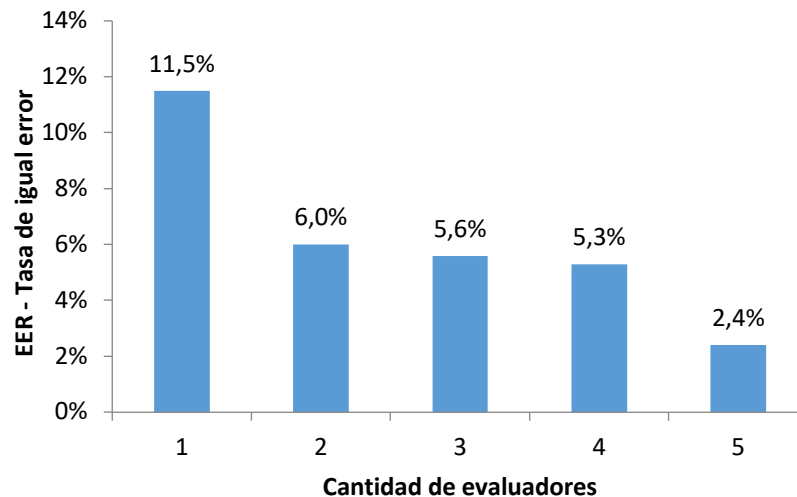


Figura 4. Efecto del número de evaluadores en la performance de la comparación de voces expresada como la tasa de igual de error (*EER*). Imagen propiedad autores.

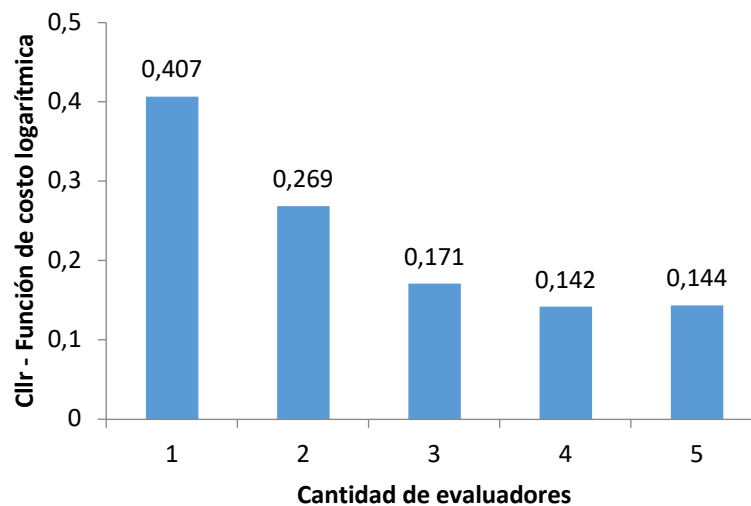


Figura 5. Efecto del número de evaluadores en la performance de la comparación de voces expresada como la función de costo logarítmica (*Cllr*). Imagen propiedad autores.

4. CONCLUSIONES Y TRABAJO FUTURO

En este trabajo se ha generado una base de datos apropiada para obtener las distribuciones de probabilidad de las respuestas a pares de sonidos del mismo y distinto origen.

Estas funciones permitirán convertir los puntajes resultantes de las pruebas de comparación de pares de voces dubitada e indubitada en valores *LR* y *LLR*. De esta forma los expertos en evaluación forense podrán informar sobre la fuerza de la evidencia en una escala equiparable a la de los sistemas semiautomáticos y automáticos.

Se ha verificado que aun cuando la tasa de igualación de errores por pérdidas y falsas alarmas disminuye con el número de evaluadores, el rescate de la mayor información calibrada de los pares se resuelve con cuatro evaluadores.

En un próximo trabajo y de acuerdo al método de cálculo indirecto, se desarrollará la transformación de los atributos multidimensionales de la voz a un puntaje unidimensional que deberá considerar el grado de similitud entre las muestras de voz dubitada e indubitada, incluyendo el grado de tipicidad de ambas muestras.

5. REFERENCIAS

1. Afshan A, Kreiman J y Alwan A. Speaker discrimination in humans and machines: Effects of speaking style variability - arXiv preprint rXiv:2008.03617, 2020 - arxiv.org
2. Aitken CG. Statistics and the evaluation of evidence for forensic scientists. John Wiley & Sons. 1995.
3. Aitken C G y Taroni F. Statistics and the evaluation of evidence for forensic scientists. John Wiley & Sons. 2004.
4. Amber, Kreiman J, Yeung G y Alwan A. Target and Non-target Speaker Discrimination by Humans and Machines. 10.1109/ICASSP.2019.8683362
5. Alexander A, Botti F, Dessimoz D y Drygajlo A. The effect of mismatched recording conditions on human and automatic speaker recognition in forensic applications. Forensic Sci Int. 2004 Dec 2;146 Suppl:S95-9. doi: 10.1016/j.forsciint.2004.09.078. PMID: 15639600
6. Alexander A, Dessimoz D, Botti F y Drygajlo A. Aural and automatic forensic speaker recognition in mismatched conditions. International Journal of Speech, Language and the Law. 12. 10.1558/sll.2005.12.2.214.
7. Amber, Kreiman J, Yeung G y Alwan A. Target and Non-target Speaker Discrimination by Humans and Machines. 10.1109/ICASSP.2019.8683362
8. Brümmer N. Application-independent evaluation of speaker detection. In Proc. Odyssey, Speaker and Language recognition workshop, ISCA. 2004, pp. 33-40.
9. Drygajlo, A., Jessen, M., Gfroerer, S., Wagner, I., Vermeulen, J., y Niemi, T. Methodological guidelines for best practice in forensic semiautomatic and automatic speaker recognition. Verlag für Polizeiwissenschaft. 2015. pp. 11-14.
10. Gil J y San Segundo E. La calidad de voz en fonética judicial. In Elena Garayzábal & Mercedes Reigosa (eds.), Lingüística forense. La lingüística en el ámbito legal y policial (pp. 154–199). Madrid: Euphonía Ediciones. 2014.
11. Gordillo LF Hitos de la evaluación perceptual auditiva de la voz: ¿hay evidencia?. Areté issn-l:1657-2513. 2018; 18 (2), 65-74. Obtenido de: <https://revistas.iberamericana.edu.co/index.php/arete/article/view/1413>
12. Gurlekian JA, Colantoni L y Torres HM. El alfabeto fonético SAMPA y el diseño de corpora fonéticamente balanceados. Revista Fonoaudiologica. Editorial: ASALFA, 2001. Tomo: 47, Número: 3, pp 58-69.
13. Gurlekian JA, Güemes M, Evin D y Torres H. Normalización del texto “Los sentidos” y su aplicación en la evaluación de habla continua. Revista Onomazein (en prensa 2021).
14. Hollien H. About forensic phonetics. Linguistica .2012. 52(1):27. DOI: 10.4312/linguistica.52.1.27-53
15. Morrison GS y Enzinger E. Score based procedures for the calculation of forensic likelihood ratios – Scores should take account of both similarity and typicality, Science & Justice, Volume 58, Issue 1, 2018, Pages 47-58, ISSN 1355- 306. <https://doi.org/10.1016/j.scijus.2017.06.005>. (<https://www.sciencedirect.com/science/article/pii/S1355030617300849>).
16. Park S L, Afshan A, Kreiman J, Yeung G. Target and Non-target Speaker Discrimination by Humans and Machines”. DOI:10.1109/ICASSP.2019.8683362
17. Silverman BW Density Estimation for Statistics and Data Analysis. London: Chapman & Hall; 1986.
18. Brümmer N. Application-independent evaluation of speaker detection. In Proc. Odyssey, Speaker and Language recognition workshop, ISCA 2004, pp. 33-40.
19. Univaso P. Identificación forense de hablantes en Argentina: un tutorial. ©ResearchGate 2016 DOI 10.13140/RG.2.1.4252.3768
20. Van Leeuwen DA y Brümmer N, “An Introduction to Application-Independent Evaluation of Speaker Recognition Systems,” in Speaker Classification I: Fundamentals, Features, and Methods. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 330–353.