

HACIA LA EVALUACIÓN PERCEPTUAL FORENSE
TRANSFORMACIÓN DE PUNTAJES DE SIMILITUD A RELACIONES DE VEROSIMILITUD
PARTE 2: HABLANTES FEMENINOS
TOWARDS FORENSIC PERCEPTUAL ASSESSMENT
TRANSFORMATION OF SIMILARITY SCORES TO LIKELIHOOD RATIOS
PART 2: FEMALE SPEAKERS

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

¹- Licenciatura en Fonoaudiología, Universidad de Buenos Aires.

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

³- Doctora en Fonoaudiología, Facultad de Medicina, 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: evamasessa@hotmail.com

Resumen: Objetivo: obtener las distribuciones probabilísticas de las respuestas a la evaluación de pares de voces femeninas provenientes de la misma y diferentes hablantes que permiten desarrollar el método indirecto para el cálculo del cociente de verosimilitudes (LR). Desarrollo experimental: se inicia con la creación de una base de datos que considera el diseño de 2520 frases, las cuales serán emitidas por 40 hablantes femeninos nativos de Argentina, en distintas sesiones. Se crearon estímulos con pares de frases provenientes de la misma y de distintas hablantes combinando frases diferentes grabadas en distintas sesiones. Se realizó la evaluación auditiva de los pares donde los oyentes debían indicar si las dos frases habían sido generadas por un mismo hablante o por distintas hablantes, indicando el grado de confianza en la respuesta. Resultados: a partir de las respuestas de los evaluadores, se estimaron las funciones de densidad de probabilidad para pares provenientes del mismo hablante y de diferentes hablantes. Al comparar las respuestas acumulativas de 1 a 5 oyentes para comprobar el efecto del número de evaluadores en las distribuciones obtenidas, se verifica una disminución en las mediciones de error efectuadas. Conclusiones: La evaluación perceptual de voces permite representar las funciones distribución de probabilidad de las respuestas a pares de voces de un mismo y diferentes hablantes provenientes de una población de hablantes nativos femeninos. Con las distribuciones estimadas será posible realizar la conversión del puntaje resultante de las comparaciones de la evidencia a los cocientes de verosimilitudes, denominado en el ámbito forense como “la fuerza de la evidencia”.

Palabras clave: percepción auditiva, forense, método de puntajes, distribuciones para voces femeninas.

Abstract: Objective: to obtain the probabilistic distributions of the responses to the evaluation of pairs of female voices coming from the same and different speakers that allow the development of the indirect method for the calculation of the likelihood ratio (LR). Experimental development: it starts with the development of a database that considers the design of 2520 sentences, which will be emitted by 40 female native speakers from Argentina, in different sessions. Pairs of sentences from the same and different speakers were created by combining different sentences recorded in different sessions. The auditory evaluation of the pairs was carried out where the listeners had to indicate whether the two sentences had been generated by the same speaker or by different speakers, indicating the degree of confidence in the response. Results: from the evaluators' responses, probability density functions were obtained for pairs coming from the same speaker and from different speakers. When comparing the cumulative responses from 1 to 5 listeners to check the effect of the number of evaluators on the distributions obtained, a decrease in the error measurements was verified. Conclusions: The perceptual evaluation of voices makes it possible to represent the probability distribution functions of the responses to pairs of voices from the same and different speakers from a population of female native speakers. With the distributions obtained it

will be possible to convert the score resulting from the evidence comparisons into likelihood ratios, known in the forensic field as "the strength of the evidence".

Keywords: auditory perception, forensics, scoring method, female distributions.

1. INTRODUCCIÓN

En un trabajo anterior se presentó la primera parte de la conversión del puntaje de similitud a la relación de verosimilitudes para hablantes masculinos. En esta segunda parte tratamos la misma conversión para hablantes femeninos.

El marco de referencia donde se insertan ambos desarrollos es el método de puntajes para el cálculo de la fuerza de la evidencia. Por este motivo se repasarán brevemente los conceptos básicos asociados desde el punto de vista de los peritos forenses con formación en foniatría.

En el ámbito forense, la comparación por voz consiste en cotejar una grabación de un sujeto desconocido (evidencia) con una o varias grabaciones de un sujeto conocido con el fin de determinar si ambas grabaciones fueron producidas o no por el mismo hablante. El resultado de esta comparación se traduce en un puntaje de similitud o disimilitud. Con el objetivo de uniformar las unidades de comparación se ha recomendado presentar este resultado en una escala que represente la fuerza de la evidencia tal como ocurre en las pruebas de ADN. Con este fin se utiliza el método de puntajes p método indirecto en dos etapas que convierte los atributos perceptuales en puntajes y luego los puntajes en relaciones de verosimilitud. En este trabajo presentaremos la metodología y los resultados que permite realizar la conversión de puntajes obtenidos por cualquier método en relaciones de verosimilitud (del inglés likelihood ratio: LR) para hablantes femeninos.

1.1. Métodos perceptuales

El método de evaluación perceptual es uno de los análisis utilizados en la tarea de comparación forense de hablantes junto con el análisis espectrográfico manual y el análisis automático.

Los métodos perceptuales de evaluación vocal tienen la ventaja de no necesitar el uso de instrumentación de alto costo, pero si, requieren que el profesional que realiza dicha tarea posea, en primera instancia, un buen nivel de audición y por otro lado un oído entrenado (oído musical), una agudeza y discriminación auditiva, una sensibilidad entrenada y una sólida experiencia en las variantes dialectales de la región.

El oído musical es una destreza necesaria para poder discriminar pequeñas variaciones de sonoridad, altura tonal y timbre, poder percibir las y analizarlas con el máximo de precisión (11). Además, el evaluador debe poseer la habilidad de reconocer cambios en la resonancia (oral, nasal, faríngea), variaciones en la prosodia, alteraciones en la coordinación fonorespiratoria y en fonética/fonología (variaciones en el punto y modo de articulación, sustitución y/u omisión de fonemas). Es preciso realizar una escucha óptima de cada uno de los atributos nombrados precedentemente. Entiéndase por escucha óptima a la habilidad de usar el oído para enfocar voluntariamente y con atención un sonido entre los demás sonidos del ambiente. La función primaria del oído es escuchar, de esta manera podemos, no solamente captar la información, sino también filtrar la información irrelevante de manera de organizar y jerarquizar dicha información (11).

A la hora de enfrentarnos a la tarea de comparación de voces mediante el método auditivo- perceptual, muchas veces, el perito se ve expuesto a distintos desafíos como por ejemplo el comparar muestras que son atemporales, encontrarse en presencia de simuladores o recibir muestras de voces las cuales denotan una situación de estrés. El caso de los simuladores es un factor importante a tener en cuenta. La mayoría de las personas son hábiles para determinar si una voz ha sido modificada por algún tipo de aparato electrónico o por algún cambio provocado por el mismo hablante (por ej. nasalizar, susurrar). Pero al momento de comparar dos pares de voces resulta, muchas veces imposible, determinar si ambas emisiones fueron producidas por la misma persona en el caso que una de ellas presente algún tipo de distorsión

como la mencionada precedentemente. Ante dicha situación se requerirá de un experto a la hora de realizar la toma de muestra indubitada de la voz del sospechoso, el cual utilizará recursos apropiados para desenmascarar esa voz (6).

Emoción y estrés son otros factores a los que se enfrenta el perito al momento de realizar su labor. Generalmente en las muestras recibidas (muestras dubitadas), según el delito que se trate, el sospechoso se encuentra en un estado de emoción fuerte (ira, excitación, estrés) el cual va a determinar ciertos cambios en su voz como por ejemplo aumento de la velocidad del discurso, elevación de la frecuencia fundamental y de la intensidad (6). Aquí también, al igual que en el caso de los simuladores, se necesitará de la experticia del entrevistador para poder obtener una muestra indubitada con el mismo estado emocional que la dubitada.

Cabe destacar que aquellas voces que presentan rasgos muy notables (características distintivas), son más fáciles de reconocer por sobre el resto de las voces de una población (6), por ejemplo, rasgos prosódicos particulares, trastornos fonéticos- fonológicos, alteración en la calidad vocal (ronquera, aspereza, soplo, astenia, tensión, inestabilidad), entre otros.

1.2. Cociente de verosimilitudes

Con el objetivo de mejorar y estandarizar la presentación de los informes periciales la Red Europea de Institutos de Ciencias Forenses (ENFSI, del inglés European Network of Forensic Science Institutes), ha publicado en el año 2015 una guía para la elaboración de informes evaluativos en ciencias forenses (2). En dicha directriz se plantea el uso del cociente de probabilidades en los informes evaluativos de todas las disciplinas forenses con el fin de establecer un lenguaje común y facilitar la comprensión de los mismos (9).

El likelihood-ratio indica cuántas veces es más probable que la evidencia haya sido producida por el sospechoso que por cualquier otro hablante. Este cociente de probabilidad estima la relación entre dos probabilidades P las cuales son mutuamente excluyentes. Por lo tanto, podemos suponer la probabilidad que tiene la evidencia E si se cumple la hipótesis de la fiscalía $H0$ por sobre la probabilidad que se cumpla la hipótesis de la defensa $H1$. El LR puede expresarse en la siguiente ecuación:

$$LR = \frac{P(E/H0)}{P(E/H1)} \quad (1)$$

Donde LR es el cociente de probabilidad, P es la probabilidad, E es la evidencia, $H0$ es la hipótesis de la fiscalía y $H1$ es la hipótesis de la defensa (9).

La hipótesis de la fiscalía o de la parte acusatoria ($H0$) afirma que la evidencia proviene del hablante sospechoso en tanto que la hipótesis de la defensa ($H1$) sostiene que la evidencia no proviene del hablante sospechoso, por lo cual, podría corresponder a cualquier hablante de la población.

Por lo expuesto anteriormente, al hacer referencia al cociente de probabilidades, debemos saber que dicho término lleva implícito tres conceptos: la similitud, la tipicidad y la población de referencia. La similitud se relaciona con la probabilidad que la evidencia provenga del sospechoso, cuán semejantes o menos diferentes son entre sí. La tipicidad expresa la probabilidad que la evidencia pueda ser encontrada en cualquier hablante de la población. La relación entre similitud/tipicidad es lo que se denomina cociente de verosimilitudes (9).

1.3. Ventajas del método perceptual

Los métodos perceptuales sobresalen cuando la evidencia tiene una duración efectiva menor a 8-10 segundos donde los métodos automáticos son descartados. En estos trabajos se utilizan frases de menos de 5 segundos y la separación inter par es menor a 4 segundos para favorecer la memoria de corto plazo. La ventaja de los métodos perceptuales se potencia en confiabilidad cuando se considera no solo el grado de similitud/disimilitud alcanzada en la

comparación que realiza el perito, sino también cuando se considera la tipicidad, es decir cuán parecida es la voz dubitada o desconocida con las voces típicas de la población. Dicho de una manera intuitiva: un alto puntaje obtenido en la comparación será menos contundente si esa voz se parece a las voces típicas de la población.

Como se mencionó, el método de puntajes o indirecto (8) para el cálculo de la relación LR consiste en dos pasos:

1. Convertir los atributos perceptuales en un puntaje de la evidencia
2. Estimar el puntaje de la evidencia en la relación de verosimilitudes LR

En esta etapa se obtiene el puntaje promedio de similitud/disimilitud comparando una serie de atributos perceptuales relevantes y afectando las respuestas por factores de tipicidad para cada uno de los parámetros (5).

En este trabajo se tratará la segunda etapa del método de puntajes. Para ello se estimarán las distribuciones que permitirán asociar el puntaje de la evidencia con el valor LR llamado fuerza de la evidencia.

2. DESARROLLO EXPERIMENTAL

2.1. Participantes

En el presente estudio participaron en forma voluntaria y anónima 40 hablantes de género femenino del español de la Argentina que residen en la provincia de Buenos Aires, entre 25 y 60 años de edad.

2.2. Grabación de las voces

Las grabaciones fueron realizadas mediante una aplicación genérica de los teléfonos celulares llamada “grabadora de sonidos”, en formato wav sin compresión en 44.100 Hz, 32bits, canal mono. Se utilizó el micrófono incorporado colocando el teléfono a una distancia de 10 cm de la boca.

Se eligieron tres frases del párrafo “Los Sentidos” (4) que tienen una duración máxima de 4 segundos. En la Tabla 1 se presentan los textos, las transcripciones fonéticas con el alfabeto SAMPA (3) y el número de grupos entonativos.

Estas frases fueron grabadas en dos sesiones diferentes, en distintos días y en distinto horario, en forma de entrevista dirigida (leídas y luego repetidas sin leerlas).

FRASE	GRUPOS ENTONATIVOS	MODALIDAD
¿Sabías que el cuerpo humano tiene sensores? ¡Si! ¡Como los robots! [saBiah ke el kwerpo wmano posee sensores, si, komo loh roBots]	3	1 interrogativa 2 exclamativas
Algunos nos mantienen en equilibrio, otros actúan cuando comemos algo picante. [alGunoh noh mantjenen en ekiliBrjo, otros aktuan kwando komemos alGo pikante].	2	enunciativa
Un conocimiento que tenemos desde pequeños son los sonidos de nuestra lengua. [un konosmjento ke tenemoh DehDe pekeJos son los soniDoh De nwehtra leNgwa].	1	enunciativa

Tabla 1: Frases utilizadas con transcripción grafémica y fonética con el alfabeto SAMPA, estructura de grupos entonativos y modalidades. Propiedad autores.

2.3. Evaluación perceptual de los pares de voces

Una vez obtenidas las grabaciones de los 40 locutores, se construyeron los estímulos a evaluar utilizando distintas frases obtenidas de distintas sesiones en su creación. 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 (ver Parte 1). Los evaluadores fueron sujetos con un buen nivel de audición y un oído entrenado (oído musical), fonoaudiólogas y músicos, con la destreza necesaria para analizar cada atributo perceptual con el máximo de precisión. Las instrucciones a los evaluadores fueron las mismas que para los hablantes masculinos: “Usted oirá dos frases mediante auriculares, su tarea consiste en indicar si provienen del mismo o de diferentes hablantes y luego marcar el grado de confianza con que ha dado su respuesta”. Las respuestas se obtuvieron multiplicando el grado de confianza de 1 a 5 por +1 si la elección era “mismo hablante” y por -1 si la elección era “hablantes diferentes”. Se obtuvieron de este modo las respuestas en una escala de puntajes discreta de -5 a +5.

3. RESULTADOS

En la Figura 1 se presentan las frecuencias del grado de similitud agrupados en intervalos discretos de -5 a +5. El histograma obtenido representa las probabilidades de las respuestas a los estímulos formados por pares de voces provenientes de hablantes femeninos evaluados como de origen semejante u orígenes diferentes por los oyentes.

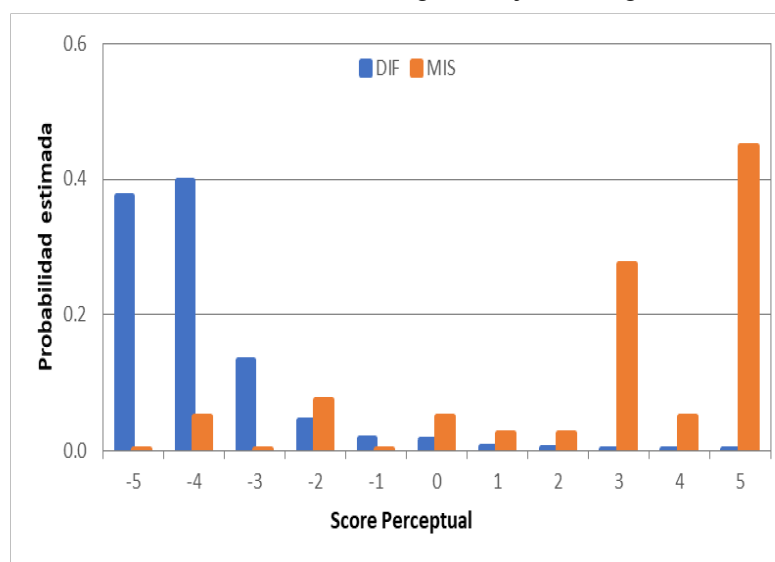


Figura 1. Histograma de las probabilidades obtenidas para los puntajes de -5 (diferente origen) hasta +5 (mismo origen) para los 840 pares de estímulos por parte de una base de 5 evaluadores por par. Imagen propiedad autores.

Puntaje	LLR	LR	Veces	
-5	-1.70	0.02	50	a favor de la defensa
-4	-1.23	0.06	17	
-3	-0.76	0.17	6	
-2	-0.29	0.51	2	
-1	0.18	1.51	2	a favor de la fiscalía
0	0.65	4.47	4	
1	1.12	13.18	13	
2	1.59	38.90	39	
3	2.06	114.82	115	
4	2.53	338.84	339	
5	3.00	1000.00	1000	

Tabla 2 Cálculos del LLR y LR. Propiedad autores.

En la Figura 2 se muestran las distribuciones del cociente de verosimilitud logarítmica (LLR) al comparar voces femeninas 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) presentada en la Parte 1.

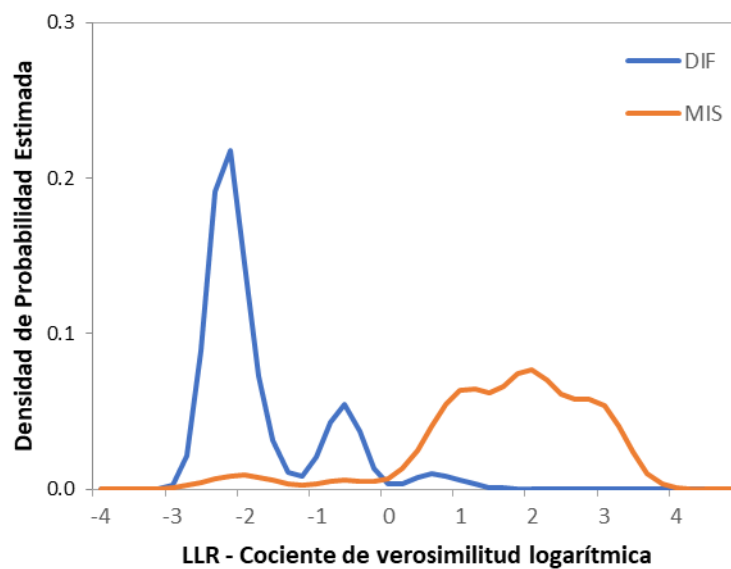


Figura 2. Funciones densidad de probabilidad estimada. Imagen propiedad autores.

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*) (1). 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. Estas dos métricas fueron calculadas para comparar el desempeño en función del número de evaluadores. En las figuras 3 y 4 se observan las medidas acumuladas desde 1 a 5 evaluadores para los estímulos de voces femeninas comparadas con las mediciones para voces masculinas obtenidas en la parte 1 (7).

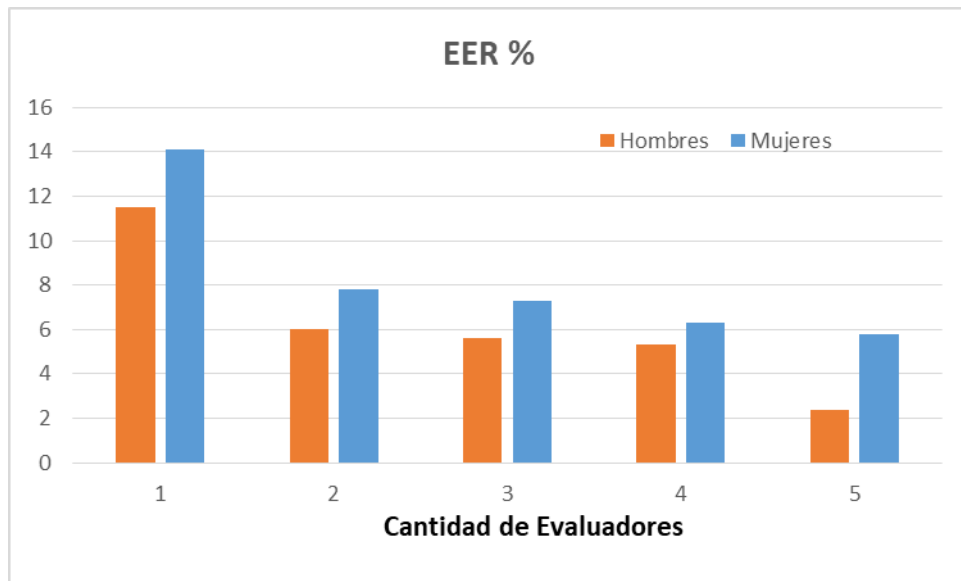


Figura 3. Porcentajes de la proporción de errores iguales (EER) en función del número de evaluadores de voces masculinas y femeninas. Imagen propiedad autores.

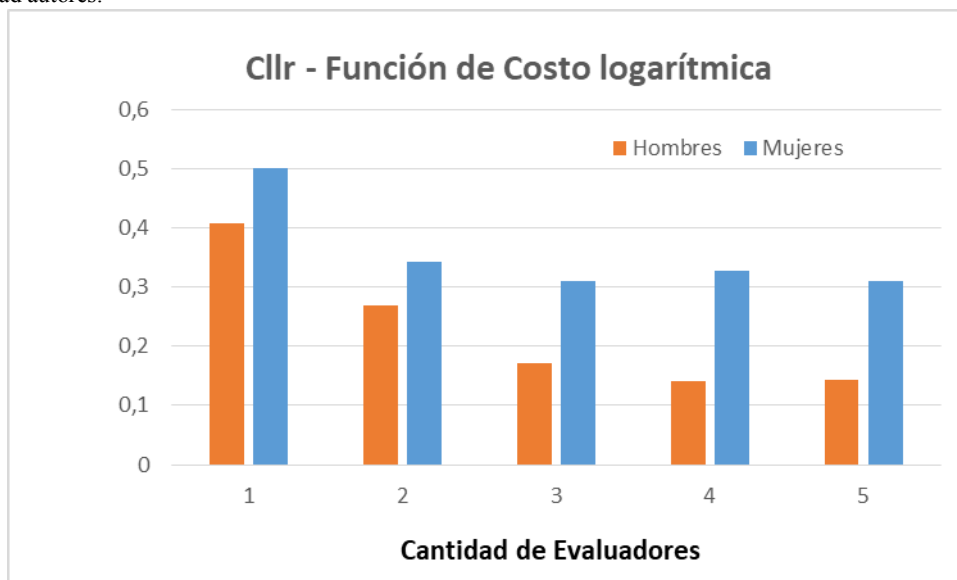


Figura 4. Función de Costo logarítmico mínimo (Cllrmin) en función del número de evaluadores de voces masculinas y femeninas. Imagen propiedad autores.

Se observa que la proporción de igual error (EER) disminuye con el número de evaluadores sin alcanzar un mínimo con 5 evaluadores. La función de costo logarítmica (Cllr) también disminuye y se estabiliza alrededor de 3 a 5 evaluadores.

Para obtener una equivalencia entre los puntajes perceptuales obtenidos y el cociente de verosimilitud logarítmico (LLR) se aproximó cada equivalencia según una ecuación lineal de acuerdo a (2) según pueden verse en la Figura 5.

$$LLR = \log \frac{f(x_i | H_0)}{f(x_i | H_1)} \cong a \cdot x_i + b \quad (2)$$

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.

<i>Evaluadores</i>	<i>a</i>	<i>b</i>
1	0.24	0.27

2	0.36	0.50
3	0.50	0.40
4	0.53	0.54
5	0.56	0.43

Tabla 3. Coeficientes de la ecuación lineal (3) de equivalencia LLR vs *score* en función de la cantidad de evaluadores. Propiedad autores.

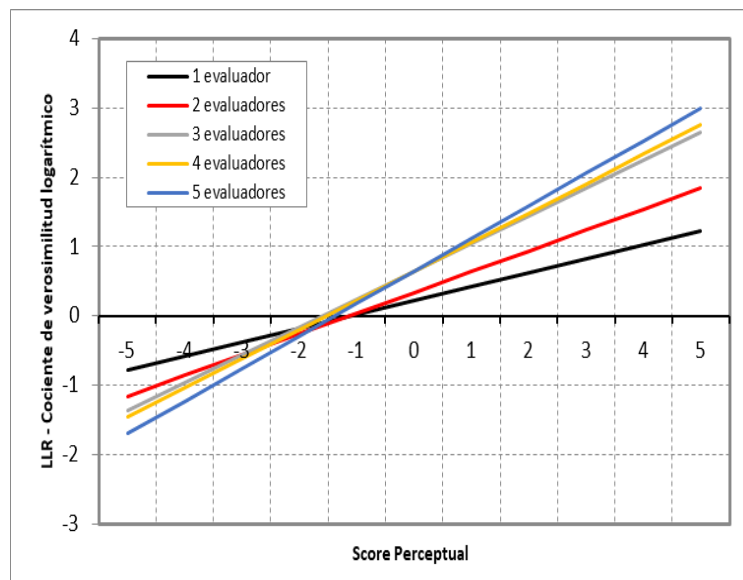


Figura 5. Familia de rectas LLR según el número de evaluadores. Imagen propiedad autores.

4. CONCLUSIONES Y TRABAJO FUTURO.

En este trabajo se ha creado una base de datos para obtener las distribuciones probabilísticas de las respuestas de comparación perceptual de pares del mismo origen y pares de diferentes orígenes para hablantes femeninos. Se verifica que a mayor número de evaluadores disminuye el EER y se estabiliza en un mínimo la función de costo logarítmica Cllr alrededor de 4-5 jurados evaluadores. La evaluación de pares de voces femeninas presentó más errores que los medidos para pares de voces masculinas.

Estas funciones permitirán convertir los puntajes de las pruebas forenses estándar de comparación de pares en valores LR y LLR de acuerdo al método de estimación indirecto.

Se espera mejorar las pruebas forenses de comparación de voces al considerar no sólo el grado de similitud global y de los atributos sino también el grado de tipicidad de las muestras a ser comparadas.

Agradecimientos: Este trabajo se realizó sin ningún subsidio a proyectos de investigación. Surgió como resultado del interés y trabajo demostrado por los tres primeros autores en el curso de “Identificación Forense por Voz” dictado en la UTN por los tres últimos en forma remota en el 2020. Agradecemos a todos los participantes: evaluadores, colaboradores y a los propios autores que lo hicieron posible.

5. REFERENCIAS

1. Brümmer N. Application-independent evaluation of speaker detection. In Proc. Odyssey, Speaker and Language recognition workshop, ISCA. 2004, pp. 33-40.
2. ENFSI Guideline for evaluative reporting in forensic science. <http://enfsi.eu>
3. 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.
4. 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).
5. Gurlekian JA, Suligoy S, Univaso P, Torres H, Masessa E, Molina N. Similitud y tipicidad en la evaluación forense de atributos perceptuales. En desarrollo. (2021).
6. Hollien, H. Forensic Voice Identification. Academic Press. 2002.
7. Molina N, Suligoy S, Masessa E, Torres H, Univaso P, Gurlekian JA. Hacia la evaluación perceptual forense. Transformación de puntajes de similitud a relaciones de verosimilitud. Parte 1: hablantes masculinos. En desarrollo. (2021).
8. 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>).
9. Rosas Aguilar C, Sommerhoff Hyde J, Sáez Elgueta C, Saavedra Cea S. Comparación de voz bajo el cociente de probabilidad en el caso de Luis Tralca. RLA, Revista de Lingüística teórica y aplicada. <http://dx.doi.org/10.4067/S0718-48832014000100002>. 2011
10. Silverman BW Density Estimation for Statistics and Data Analysis. London: Chapman & Hall; 1986.
11. Trallero Flix C. El oído musical. Obtenido de: <http://diposit.ub.edu/dspace/bitstream/2445/11525/1/EL%20OIDO%20MUSICAL.pdf>. 2008.