

**INFLUENCIA DEL CANAL DE GRABACIÓN EN SISTEMAS  
FORENSES DE RECONOCIMIENTO AUTOMÁTICO DE  
LOCUTOR**

*INFLUENCE OF THE RECORDING CHANNEL ON OPERATOR'S  
AUTOMATIC RECOGNITION FORENSIC SYSTEMS*

**JOSÉ ANTONIO DÍAZ PÉREZ**

Tutores:

**FERMÍN DEL RÍO CANO**

Servicio de Criminalística de la Guardia Civil

**CARMEN VALERO GARCÉS**

Universidad Alcalá de Henares

**MÁSTER UNIVERSITARIO EN  
CIENCIAS POLICIALES**

**(Trabajo fin de Máster – 12 ECTS)**

**2020-2021**

## **RESUMEN DEL PROYECTO**

El enfoque inicial de este Proyecto es el análisis de la evidencia por los peritos forenses con la finalidad de identificar si una muestra proviene de una determinada fuente. “La Criminalística es la ciencia de la individualización” (Kirk, 1963), no obstante la muestra científica requiere especial atención en su interpretación, debiendo desechar la idea inicial de una certeza absoluta incluso en un proceso de identificación absolutamente objetivo. Lo que si podemos dar como cierto es la incertidumbre respecto a muchos de los procesos asociados a la observación científica pero esta apreciación no impide hacer una valoración de las diferencias cuantitativas y proporcionar una medida de incertidumbre asociada a esta valoración.

Partiendo de esta idea y volviendo al objeto de análisis del proyecto, los resultados de un sistema reconocimiento automático de locutor varían ateniéndose a las características del audio utilizado, en este caso el tipo de archivo de audio (.mp3, .wav, ogg), pudiendo afectar a la fiabilidad y por lo tanto a las graves repercusiones que podría alcanzar en el ámbito forense donde se exige la máxima robustez de resultados.

Se procurará el estudio de la incidencia de las variables que se verán afectadas por tipo de archivo, en la presentación final de los resultados de un sistema de reconocimiento automático de locutor, identificando su influencia en el trabajo forense.

Para el análisis establecido, se ha contado con software específico de reconocimiento automático de locutor, desarrollado por la comercial AGNITIO S.L. De los ensayos practicados se puede deducir el motivo de la elección de las actividades expresadas que mejoran los sistemas forenses y los importantes avances conseguidos gracias a muchas de ellas.

## **LISTADO DE PALABRAS CLAVE**

Biometría, reconocimiento automático de locutor, sistemas de ámbito forense, presunción de inocencia, parametrización y procesado de la voz, sistemas de identificación, sistemas de verificación, teoría bayesiana de valoración, relación de verosimilitud, puntuaciones, modelos, test, poblaciones, normalizaciones.

## Agradecimientos

*Me gustaría dedicar este Proyecto a mi familia por su paciencia y comprensión, que es la que más me ha apoyado pesar de haberles dejado apartados dándome ánimos en los momentos difíciles de su confección.*

*Muchas gracias a mis profesores y compañeros de Master durante este año inolvidable de Universidad.*

*Por último, quisiera darle las gracias especialmente a Fermín del Río, por su tiempo y dedicación, por supervisarme y guiarme en la realización de este proyecto a lo largo de estos meses.*

*¡¡Muchas gracias a todos!!*

## **PREFACIO**

La investigación ha demostrado que un modelo probabilístico basado en la forma de probabilidades del teorema de Bayes y la razón de verosimilitud parece ser una herramienta adecuada para ayudar a los expertos en ciencias forenses a interpretar la evidencia [1]. Este marco puede aplicarse al reconocimiento forense de hablantes mediante sistemas automáticos de reconocimiento de locutor [2]. Si bien, los sistemas de reconocimiento de hablantes de última generación muestran suficientes habilidades de discriminación entre hablantes, ha sido preciso un paso adelante para permitir su uso en un entorno forense de la vida real. El proceso de adaptación no fue sencillo, especialmente cuando falta una cantidad suficiente de datos del sospechoso o condiciones desajustadas (canal, ruido, emociones, lenguaje, impostación de voz, etc.) entre el locutor sospechoso y las grabaciones dubitadas.

Teniendo estos hechos en mente, las investigaciones se han centrado en adaptar los sistemas de reconocimiento de locutor, para que sea compatible con el marco de interpretación bayesiano, y en desarrollar contribuciones para una estimación robusta de las razones de verosimilitud [3]. El doble objetivo es: en primer lugar, proporcionar una razón de verosimilitud significativa (LR) para cada par de habla cuestionada y sospechosa; evitando o minimizando así la gran proporción de casos no reportados presentes en el reconocimiento de locutor forense debido a condiciones de no coincidencia o calidad limitada de los datos, y en segundo lugar, garantizar la presunción de inocencia, manteniendo potenciales locutores no sospechosos (non target) con puntuaciones de LR inferiores a uno, lo que corresponde a no respaldar la hipótesis de enjuiciamiento ( $H_p$ ).

## **ORIGEN DEL PROYECTO**

La idea inicial parte de la cadena de investigaciones concernientes al estado del arte de la tecnología aplicada a los sistemas forenses de reconocimiento de locutor desarrollada por la empresa AGNITIO SL y utilizada en el Servicio de Criminalística de la Guardia Civil.

La búsqueda de identificadores que individualicen la biometría de las personas ha sido una constante a lo largo del tiempo, encontrando en el reconocimiento automático de locutor uno

de los retos más complicados y a la vez más esperanzadores. Se trata de centrar esfuerzos en la búsqueda de nuevos algoritmos de tratamiento de voces, más eficientes. No obstante la documentación y los estudios relativos a la influencia de las características del audio en los sistemas, no son muy profundos. Con este estudio se pretende acercarse a ver cómo afecta a los resultados y su robustez, la configuración de los sistemas y formato de los indicios para su uso en un entorno forense.

## **OBJETIVOS DEL PROYECTO**

Este trabajo se intenta alcanzar tres objetivos principales. Primeramente, poner de manifiesto algunas de las variables o elementos más influyentes en este tipo de sistemas de reconocimiento automático de locutor. En segundo lugar, se tratará de ofrecer una respuesta a cómo afecta cada variable y sobre qué funciones del sistema, esta dependencia es más influyente. En último término se procurará condensar toda esta información en forma de conclusiones asequibles al usuario final del sistema que le permitan la alternativa de configuración del mismo.

## Índice de contenidos

<b>1.</b>	<b><i>SISTEMAS DE RECONOCIMIENTO AUTOMATICO DE LOCUTOR</i></b>	<b>8</b>
1.1.	<b>INTRODUCCIÓN</b>	<b>8</b>
1.2.	<b>FUNCIONAMIENTO BÁSICO</b>	<b>9</b>
1.2.1.	Parametrización de la voz y extracción de características	9
1.2.2.	Diagrama de funcionamiento de un sistema	12
1.3.	<b>CLASIFICACION DE LOS SISTEMAS</b>	<b>13</b>
1.4.	<b>SISTEMAS FORENSES</b>	<b>15</b>
1.5.	<b>LIKELYHOOD RATIOS (LR) Y ESTIMACIÓN BAYESIANA</b>	<b>16</b>
1.6.	<b>PROBLEMAS EN LA ESTIMACIÓN DEL LR</b>	<b>19</b>
1.7.	<b>EVALUACIÓN DE SISTEMAS FORENSES</b>	<b>20</b>
<b>2.</b>	<b><i>DESCRIPCIÓN DEL SISTEMA BATVOX</i></b>	<b>26</b>
2.1.	<b>CONCEPTOS DE RECONOCIMIENTO DE LOCUTOR</b>	<b>28</b>
2.1.1.	Entrenamiento	28
2.1.2.	Tareas de reconocimiento	29
2.1.3.	Herramienta EvalBatvox	36
<b>3.</b>	<b><i>ANALISIS DE VARIABLES</i></b>	<b>39</b>
3.1.	<b>CASO 1: Valores de referencia - Test: Ahumada IV (SITEL) VS Modelo: Ahumada IV (SITEL)</b>	<b>41</b>
3.1.1.	Resultados y curvas comparativas	41
3.2.	<b>CASO 2: Test: Ahumada IV (AAC) VS Modelo: Ahumada IV (SITEL)</b>	<b>43</b>
3.2.1.	Resultados y curvas comparativas	43
3.3.	<b>CASO 3: Formato MP3 - Test: Ahumada IV (MP3) VS Modelo: Ahumada IV (SITEL)</b>	<b>45</b>
3.3.1.	Resultados y curvas comparativas	45
3.4.	<b>CASO 4: Test: Ahumada IV (OGG) VS Modelo: Ahumada IV (SITEL)</b>	<b>48</b>
3.4.1.	Resultados y curvas comparativas	48

<b>3.5.</b>	<b>COMP 1: Test: Ahumada IV (SITEL) y Test: Ahumada IV (AAC) VS</b>	
	<b>Modelo: Ahumada IV (SITEL)</b> -----	<b>50</b>
3.5.1.	Resultados y curvas comparativas-----	50
<b>3.6.</b>	<b>COMP 2: Test: Ahumada IV (SITEL) y Test: Ahumada IV (MP3) VS</b>	
	<b>Modelo: Ahumada IV (SITEL)</b> -----	<b>52</b>
3.6.1.	Resultados y curvas comparativas-----	52
<b>3.7.</b>	<b>COMP 3: Test: Ahumada IV (SITEL) y Test: Ahumada IV (OGG) VS</b>	
	<b>Modelo: Ahumada IV (SITEL)</b> -----	<b>54</b>
3.7.1.	Resultados y curvas comparativas-----	54
<b>3.8.</b>	<b>COMP 4: Test: Ahumada IV (SITEL), Test: Ahumada IV (AAC), Test:</b>	
	<b>Ahumada IV (MP3) y Test: Ahumada IV (OGG) VS Modelo: Ahumada IV (SITEL)</b>	
	<b>55</b>	
3.8.1.	Resultados y curvas comparativas-----	56
<b>4.</b>	<b>CONCLUSIONES:</b> -----	<b>57</b>
<b>5.</b>	<b>TRABAJO DE FUTURO</b> -----	<b>58</b>
<b>6.</b>	<b>REFERENCIAS</b> -----	<b>59</b>

# **1. SISTEMAS DE RECONOCIMIENTO AUTOMATICO DE LOCUTOR**

## **1.1. INTRODUCCIÓN**

Las críticas hacia las formas clásicas de evaluación de la evidencia han puesto en duda su carácter científico [4] motivado por: condenas erróneas, el análisis de ADN como ejemplo científico de disciplina forense y por cambios en las leyes iniciados por reglas Daubert sobre los requisitos necesarios para que una evidencia sea admitida a juicio. Además, se contemplan otras necesidades que los métodos clásicos no ofrecían, como puede ser la transparencia en las técnicas para descartar prácticas no científicas, la disponibilidad de los datos para evaluación, comparación y mejora de los métodos empleados, finalmente se estima en gran valor la adopción de métodos comunes en la presentación de resultados.

En el sentido expuesto, se ha alcanzado la conclusión de seguir la corriente usada en la identificación por ADN: “el cálculo de LR”. Con esta metodología se ha aportado un apoyo probabilístico al peso de la evidencia, alejando las opiniones influenciadas por la experiencia del perito.

Los sistemas de reconocimiento automático de locutor son programas informáticos que comparan muestras de voz de unos individuos (test) con las características de la voz de otros individuos (modelos), dando como resultado puntuaciones de similitud (scores) para cada pareja test-modelo, ofreciendo de esta forma una visión clara sobre las posibilidades de que estos audios pertenezcan a la misma persona.

De una forma simple, la aplicación de estos sistemas al ámbito policial es lo que se denomina “entorno forense”, de este modo trataríamos de hacer un modelo de voz de un sospechoso o voz indubitada y enfrenar este modelo a un test (voz dubitada) para conocer el grado de similitud entre ambas. La necesaria precisión para diferentes condiciones acústicas exige modelos más complejos. Asimismo, la entidad del resultado de estas pruebas en un entorno como el descrito, ha necesitado de nuevas y más complejas fórmulas para cuantificar la similitud de los audios, adoptando como fórmula los valores de LR's (relaciones de verosimilitud). Con este objetivo, estos sistemas deben poder extraer con precisión las características de la voz que definan de forma unívoca al locutor.

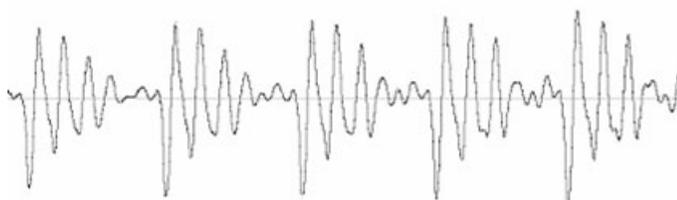
## **1.2. FUNCIONAMIENTO BÁSICO**

### **1.2.1. Parametrización de la voz y extracción de características**

La señal de voz puede contemplarse a largo plazo (en el orden de segundos) siendo en este caso no estacionaria toda vez que sus características temporales están sometidas a constantes variaciones, aunque las transiciones son progresivas; o a corto plazo (decenas milisegundos) donde la señal se muestra cuasi-estacionaria. De esta forma la señal hablada presenta tramos sonoros (carácter periódico) y tramos sordos (carácter ruidoso).

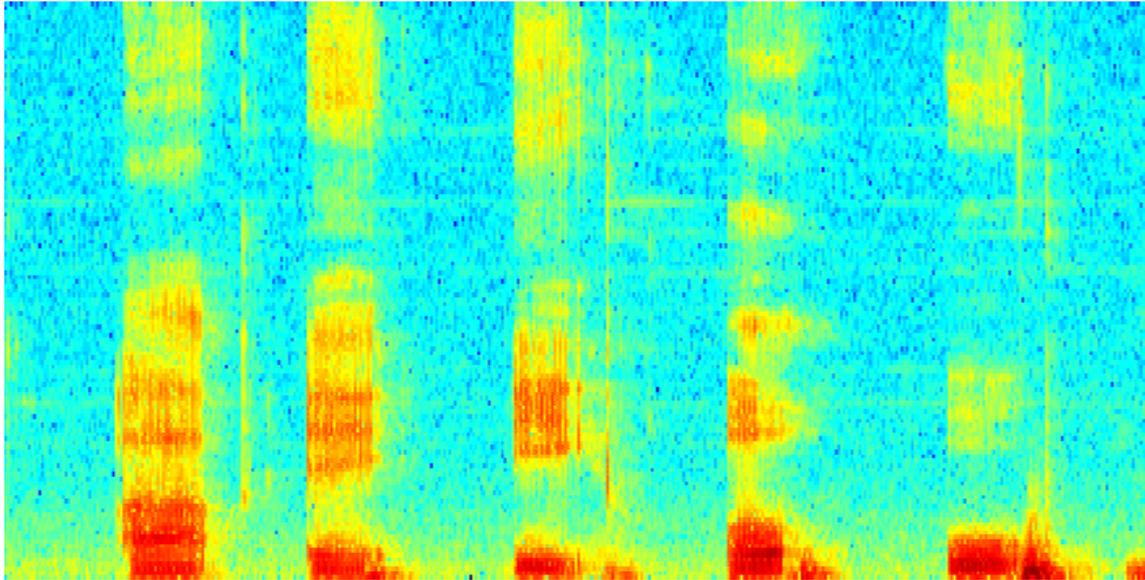


Locución de 5 segundos



Locución de una vocal de 80 ms

En la generación de sonido, intervienen las cuerdas vocales y el tracto vocal (faringe, cavidad oral y cavidad nasal). Los sistemas de reconocimiento automático usan el último básicamente, la laringe excita las cavidades del tracto vocal dando lugar a frecuencias de resonancia únicas llamadas “formantes”, que constituyen el rasgo diferenciador del habla humana.



Cada señal de voz dispone de dos componentes, la excitación, que contiene la estructura fina del espectro y la que se ocupa de la articulación de los sonidos, que es la que va a determinar la envolvente espectral característica, donde se pueden detectar los formantes.

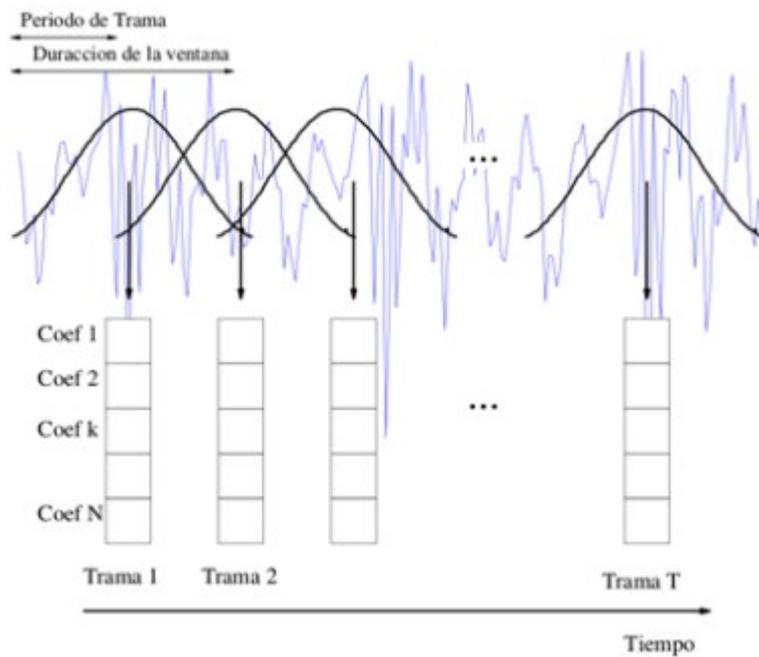
Al hecho de reducir la información disponible de la señal a procesar y a su vez, expresar esta información en dominios donde ésta se muestre suficientemente robusta e independiente se conoce como “parametrización”.

En la parametrización se trata de extraer parámetros a partir de muestras de la señal y transformar la señal en un vector de características, cuya representación sea más practicable a la hora del modelado estadístico y el cálculo de puntuaciones y es precisamente en la parametrización de la envolvente espectral, donde se fundamentarán los sistemas automáticos para conseguir su eficiencia.

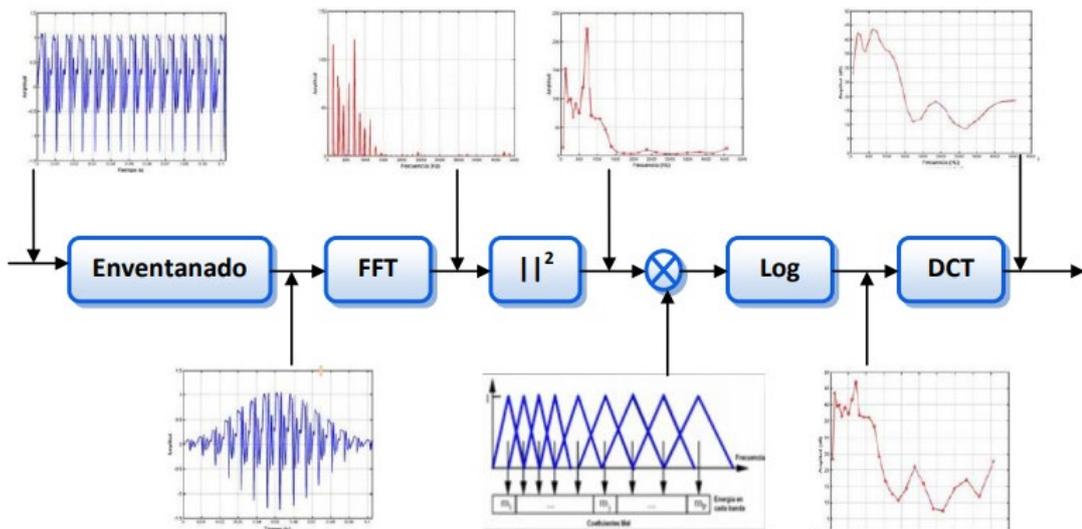
La señal de voz varía con el tiempo, cada pocos milisegundos existe un cambio en la señal de voz, por ello es preferible conocer las propiedades de la señal de voz en intervalos pequeños de tiempo que en toda la señal. En este sentido, para facilitar el análisis y tratamiento de voz se aplicará a la señal de voz una segmentación temporal, que permitirá realizar un análisis localizado mediante el uso de tramas consecutivas al que se denomina enventanado de la señal. Es precisamente, a partir de cada ventana, usualmente tipo Hamming, [5] de donde extraemos los parámetros de interés. Se trata en definitiva de trasponerlo a un modelo matemático que permita la representación de las características de las diferentes voces. No obstante, la finalidad de estos procesos no es la consecución de un modelo, sino la de conseguir el reconocimiento

de la voz, es decir, se trata de determinar si un audio de procedencia desconocida pertenece a uno de los modelos de locutores ya identificados.

En el siguiente esquema se eligen segmentos temporales de la locución de 20 ms, donde se establece un solapamiento del 50% (10 ms) a través de ventanas tipo Hamming. A partir de cada ventana se separan los parámetros de interés.



Enventanado de la señal de voz para extracción de características



Extracción de características MFCC

### 1.2.2. Diagrama de funcionamiento de un sistema

En este punto, se pretende analizar como los sistemas automáticos de locutor trabajan con las características parametrizadas.

El principal objetivo de los sistemas de reconocimiento automático de locutor es conseguir formulas y técnicas seguras para que, sin supervisión humana, puedan pronunciarse acerca de la identidad del locutor.

Para conseguir este propósito, los sistemas automáticos de reconocimiento de locutor, suelen trabajar en tres fases [6]:

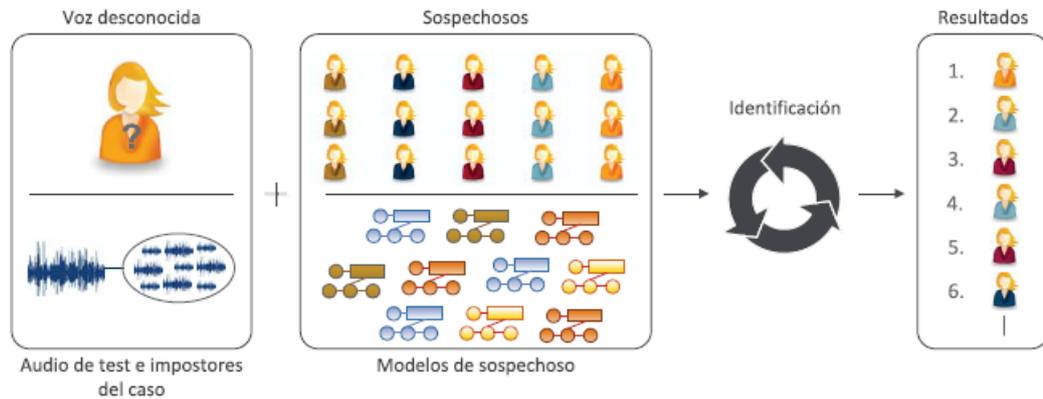
- **Fase de entrenamiento:**

Se trata de un proceso en el que se extraen las características de la voz de una persona y a partir de toda esta información, se genera un modelo biométrico.



- **Fase de prueba o reconocimiento:**

Conjunto de comparaciones de nuevas muestras de voz desconocidas (test) y modelos de sospechoso que, tras un proceso de normalización, el sistema tomará decisiones acerca de la identidad del hablante.



Fase de prueba o reconocimiento

- **Fase de actualización:**

El sistema deberá ser capaz de incorporar nuevos locutores, introducir y dar de baja a usuarios, actualizar o mejorar modelos y referencias de usuarios presentes en el sistema.

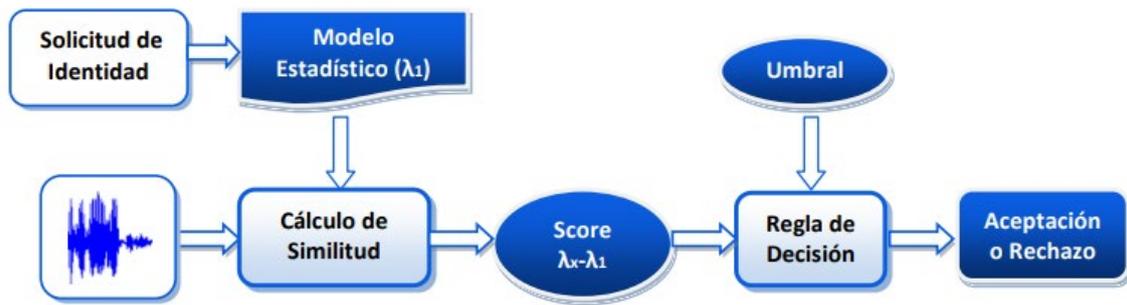
### 1.3. CLASIFICACION DE LOS SISTEMAS

Tomando como criterio la dependencia del sistema del texto pronunciado, obtenemos los sistemas dependientes y los sistemas independientes. En el primero, la locución de entrenamiento y la de prueba son idénticas, en los segundos la locución a pronunciar en la fase de entrenamiento y la de reconocimiento no coinciden.

Por otro lado, tomando como base la tarea específica a realizar por el sistema, nos encontramos con sistemas de identificación o de verificación de locutores.

#### Verificación

Se reciben dos entradas de audio. Una corresponde a la voz a verificar y la otra es una solicitud de identidad. De este modo las decisiones del sistema son la aceptación o rechazo de que ambas locuciones pertenezcan a la misma persona, debiendo contar para ello con umbrales de decisión que permitan al sistema un valor probabilístico.



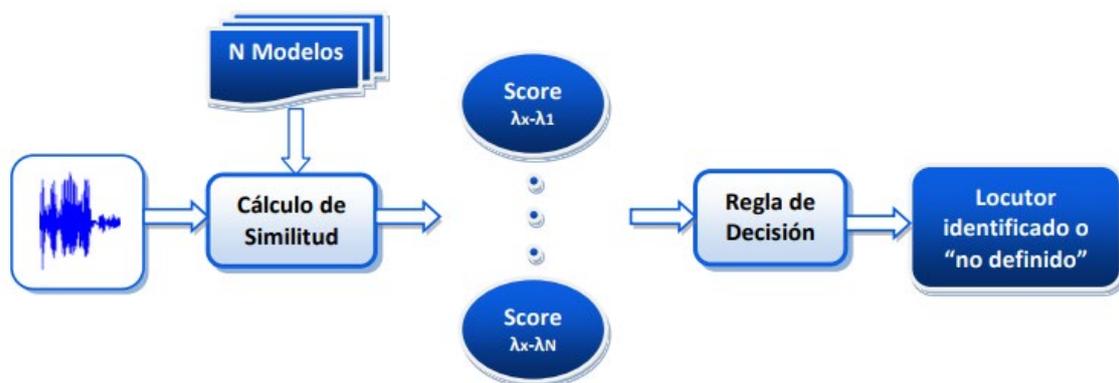
Esquema en modo verificación

### Identificación

El objetivo de los sistemas de identificación de locutores es clasificar una señal de voz a verificar, dentro de un conjunto de modelos almacenados. Aquí, se calculan todas las probabilidades para cada modelo y la mayor de todas se toma como perteneciente al modelo que se está analizando.

Casuística:

- Identificación en conjunto cerrado: Como resultado se produce la asignación de identidad a uno de los locutores modelados por el sistema. Existen, por tanto, N posibles salidas.
- Identificación en conjunto abierto: el sistema debe considerar la posibilidad de que la locución de entrada no pertenezca a ninguno de los usuarios de la base de datos.



Esquema en modo identificación

En todo el proceso el sistema ha de poder garantizar que las características de todos los modelos son similares, al objeto de impedir que el sistema puntúe más alto a modelos en el mismo canal,

aunque no pertenezcan a la misma persona. En este sentido, en caso de no poder garantizar este extremo se llevará a cabo una “normalización” que compense las puntuaciones para permitir su comparación.

#### **1.4. SISTEMAS FORENSES**

La finalidad última de un sistema forense está en asistir a un tribunal, con resultados objetivos y útiles para el Juez o para el jurado [7].

El análisis forense de la evidencia, persigue ofrecer al tribunal una estimación de la fuerza de la propia evidencia. No obstante, el resultado de una tarea de identificación ofrece una opinión, nunca una información, es decir indica la posibilidad de encontrar una persona de características similares [8].

La premisa fundamental de cualquier sistema es garantizar la presunción de inocencia, de este modo, debe poder minimizar los locutores sospechosos no autores (non-targets), falsamente acusados, aunque esto redunde en una peor discriminación entre sospechosos autores “targets” y “non-targets”.

El rendimiento, en estos sistemas, se mide mediante la relación FA/FR, cuya representación más adecuada es la curva DET (Detection Error Tradeoff) ofreciendo una muestra de la prestación del sistema casi lineal, al objeto de visualizar la productividad del mismo en todos los puntos de trabajo (umbrales).

En otro sentido, bajo un enfoque Bayesiano (firmemente consolidado como marco teórico para cualquier disciplina forense y de uso habitual en áreas tipo ADN), los sistemas forenses muestran los resultados en forma de “Relaciones de Verosimilitud” o LR (Likelihood Ratios), siendo evaluados a través de curvas Tippett, que indican cómo de fuerte es el apoyo a la hipótesis de partida para target (el sospechoso es el autor de la evidencia) y cómo de fuerte es la atenuación de la misma hipótesis para non-target (el sospechoso no coincide con el autor de la evidencia).

En resumen, se dispone de una muestra de voz desconocida y unas muestras de referencia de un sospechoso, obtenidas bajo unos parámetros controlados con las que se ha generado su modelo, siendo el objetivo comparar las muestras. Así, se calcula la similitud entre ambas muestras dando lugar a un score o puntuación que representará la evidencia forense. Para que

esta puntuación pueda mostrar su valoración, ha de transformarse en un LR que simboliza el peso de la evidencia.

### **1.5. LIKELYHOOD RATIOS (LR) Y ESTIMACIÓN BAYESIANA**

La metodología Bayesiana de estimación de la evidencia, permite que la valoración a priori (previa a la presentación de la evidencia) a favor de la culpabilidad o la inocencia, sea transformada en una valoración a posteriori (una vez dada la evidencia), basada en datos. De acuerdo con este enfoque, los sistemas forenses ofrecen sus resultados en forma de relaciones de verosimilitud (LR,s), ello implica una manifestación del **grado de apoyo hacia una hipótesis frente a la otra**, labor que ayudará al tribunal en la toma de decisión [9].

En un caso de atribución de fuentes se considerarán dos hipótesis:

- $H_p$ : proposición en la que el sospechoso es culpable, hipótesis inicial (la voz dubitada pertenece al sospechoso)
- $H_d$ : proposición en la que el sospechoso es inocente.
- Ev: evidencia, resultado de comparación de las muestras (score) realizado por el perito (sistema)
- I: información del caso, ajena a la observación del experto (sexo, raza, nacionalidad, lengua, etc) que influye en él.

A partir de aquí, obtenemos la apuesta a priori, es decir la relación de la probabilidad de que la voz dubitada pertenezca al sospechoso, partiendo las circunstancias del caso frente a la probabilidad de que no pertenezca. Esta apuesta puede determinar si la hipótesis inicial es correcta o no, sin llevar a cabo ningún estudio por el perito.

$$O(H_p / I) = P(H_p/I)/P(H_d/I)$$

Igualmente, se pueden deducir la apuesta a posteriori como:

$$O(H_p / Ev, I) = P(H_p/Ev, I)/P(H_d/Ev, I)$$

La relación entre las dos apuestas es una cte. de ponderación, representada por el resultado de la evidencia, denominada Relación de Verosimilitud o LR.

$$O(H_p / Ev, I) = \mathbf{LR} * O(H_p / I)$$

Donde el LR aparece, de esta forma, representado mediante el cociente entre dos probabilidades.

1. Probabilidad de, dada la observación, la  $H_p$  sea cierta.
2. Probabilidad de, tras la observación, la  $H_p$  sea incierta y sea cierta la hipótesis contraria,  $H_d$ .

$$LR = P(Ev / H_p, I) / P(Ev / H_d, I)$$

Para determinar que las probabilidades del numerador y del denominador sean objetivas, debemos contar con datos de una población de referencia.

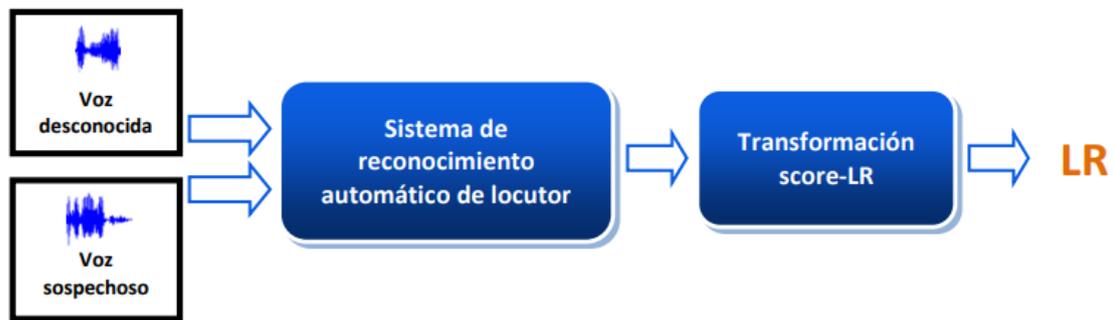
La existencia de bases de datos de locuciones es un requisito imprescindible para modelar la distribución de las muestras, sean estas de la misma fuente o entre fuentes, de este modo el LR se verá como un cociente de funciones de densidad de probabilidades en lugar de una cociente de probabilidades.

Siendo el numerador (within-source distribution) el grado de verosimilitud de la evidencia, cuando el locutor se supone el autor de la misma, representa una semejanza entre el material cuestionado y el sospechoso. Por otro lado, el denominador (between-source distribution) mide la verosimilitud de la evidencia, dado que el locutor se trata de otro sujeto perteneciente a una población de referencia, es decir, una medida de la tipicidad o rareza del sospechoso en una población relevante de individuos.

De esta forma tenemos que, si lo más probable es que la voz dubitada pertenezca al sospechoso,  $LR > 1$  y al contrario,  $LR < 1$  caso que la probabilidad de que no pertenezca al locutor es mayor. Por lo tanto, esta constante lo que hace es reforzar o debilitar la probabilidad a priori de que sea cierta la hipótesis de partida y el grado de apoyo o atenuación viene definido por la magnitud del LR.

- ✓ Ej.  $LR=3$ ; indica que el resultado del perito, es tres veces más probable si  $H_i$  fuera cierta que si fuera falsa. La apuesta a priori a favor de  $H_i$  se multiplica por 3.
- ✓  $LR=1/3$ , el resultado del perito es tres veces más probable si  $H_i$  fuese falso que si fuese verdadero. La apuesta a priori a favor de  $H_i$  se divide por 3.

El objetivo del cálculo



Computación de LR,s a partir de scores

En el marco bayesiano distinguimos muchas ventajas para el contexto forense. Primero, permite hacer una estimación e informar un valor significativo al tribunal [10]. En segundo lugar, el papel científico está claramente definido, dejando al tribunal la tarea de utilizar sentencias o costos previos en el proceso de la decisión [11]. En tercer lugar, las probabilidades se pueden interpretar como grados de creencia [12] lo que permite la incorporación de opiniones subjetivas como probabilidades.

La interpretación bayesiana de la evidencia forense mediante el uso de técnicas discriminatorias tiene como objetivo, el cálculo de la razón de verosimilitud (LR) como un grado de apoyo de una hipótesis frente a su opuesta.

Los sistemas automáticos estiman el LR a partir de puntuaciones de similitud [13].

- ✓ Ej. Suponemos que la evidencia (Ev) es una información extraída de una escucha telefónica y el material sospechoso, una grabación del sospecho en ambiente bajo control. En los sistemas automáticos, la Ev es una puntuación de similitud entre la escucha y el material sospechoso.

Al objeto de hacer más comprensible estos valores de LR, se utiliza como criterio la tabla de equivalencias verbales utilizada por el Servicio de Ciencias Forenses Británico (escala Evett):

<i>Equivalencia Verbal</i>	<i>LR</i>
<i>Respalda muy fuertemente la hipótesis del fiscal frente a la de la defensa</i>	<i>Más de 10000</i>
<i>Respalda fuertemente la hipótesis del fiscal frente a la de la defensa</i>	<i>1000 a 10000</i>
<i>Respalda de forma moderadamente fuerte la hipótesis del fiscal frente a la de la defensa</i>	<i>100 a 1000</i>
<i>Respalda de forma moderada la hipótesis del fiscal frente a la de la defensa</i>	<i>10 a 100</i>
<i>Respalda levemente la hipótesis del fiscal frente a la de la defensa</i>	<i>1 a 10</i>
<i>Respalda levemente la hipótesis de la defensa frente a la del fiscal</i>	<i>1 a 0,1</i>
<i>Respalda de forma moderada la hipótesis de la defensa frente a la del fiscal</i>	<i>0,1 a 0,01</i>
<i>Respalda de forma moderadamente fuerte la hipótesis de la defensa frente a la del fiscal</i>	<i>0,01 a 0,001</i>
<i>Respalda fuertemente la hipótesis de la defensa frente a la del fiscal</i>	<i>0,001 a 0,0001</i>
<i>Respalda muy fuertemente la hipótesis de la defensa frente a la del fiscal</i>	<i>Menos de 0,0001</i>

## **1.6. PROBLEMAS EN LA ESTIMACIÓN DEL LR**

Ya en casos reales, el reconocimiento forense del locutor presenta algunas peculiaridades que hacen que aparezcan errores en la estimación de las puntuaciones de LR, debiendo adoptar técnicas especiales para evitarlos.

Generalmente, en condiciones forenses, la calidad y cantidad de datos de voz que el experto forense puede manejar está lejos de ser la óptima, causando fuertes desajustes entre locuciones dubitadas, del sospecho y población de referencia así como la insuficiencia de datos para las estimaciones de distribución [14] [15].

Por otra parte, en aplicaciones forenses es conveniente, para garantizar la presunción de inocencia, que los locutores de la población de referencia (non-target) no obtengan puntuaciones de  $LR > 1$ , aunque ello conlleve una peor discriminación entre target y non-target.

El problema de la estimación de la variabilidad entre fuentes (between-source) está relacionado con la selección y el número de modelos disponibles de la población de referencia. Dado que la distribución de variabilidad entre fuentes representa la distribución de la probabilidad de coincidencia aleatoria de la evidencia dentro de la población de referencia, debe presentar las mismas características que la locución dubitada con respecto al canal, idioma, etc. Por lo tanto, hay muchos problemas cuando no se dispone de una adaptada población de referencia.

Otros problemas surgen por la estimación inadecuada de la distribución de variabilidad dentro de la fuente del sospechoso. A menudo, el perito forense cuenta sólo con una grabación dubitada

y una indubitada. Normalmente la tarea consiste en evaluar si la voz proviene de la misma persona.

## 1.7. EVALUACIÓN DE SISTEMAS FORENSES

En la evaluación de estos sistemas de forma científica se llevan a cabo comparaciones sobre muestras conocidas, que podemos clasificar en dos tipos:

- **Comparaciones target:** en las que la muestra indubitada y la dubitada tiene el mismo origen, evalúan casos en los que  $H_p$  es cierta, por lo tanto, deberían generar un  $LR > 1$ .
- **Comparaciones non-target:** las muestras tienen orígenes distintos. En este caso la  $H_d$  es cierta, por lo tanto, el  $LR < 1$ .

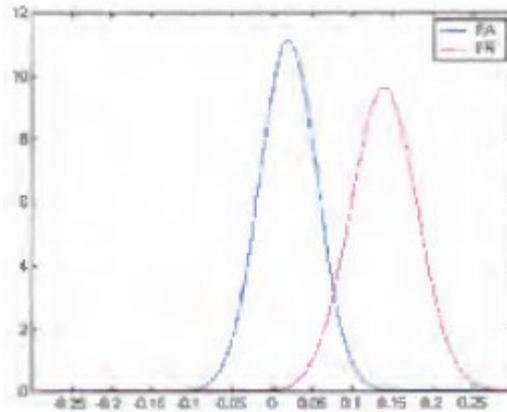
Por otro lado, la adopción del LR como indicador válido en la toma de decisiones puede dar lugar a:

- **Falsas aceptaciones (FA):** aceptar como target una comparación non-target (impostores como generadores de la voz dubitada).
- **Falsos rechazos (FR):** aceptar como non-target una comparación target (aceptar como impostores un enfrentamiento target).

Un acercamiento a la evaluación del sistema se puede obtener a partir de la distribución de puntuaciones de los tipos de errores en una misma gráfica.

Estos valores (FA, FR) dependerán del umbral establecido para separar resultados target ( $H_p$  es cierta) de los clasificados como non-target ( $H_d$  es cierta). Llamamos “discriminación” a la separación entre comparaciones target y non-target. Una menor discriminación viene determinada por una reducción en la tasa de FA, siendo FR constante o una reducción en la tasa de FR para un FA constante.

El procedimiento que se ha seguido, ha sido establecer el cálculo de la media y la varianza de cada lista de puntuaciones obteniendo una distribución en forma de gaussiana [16] como la de la figura.



Distribución de puntuaciones de FA y FR

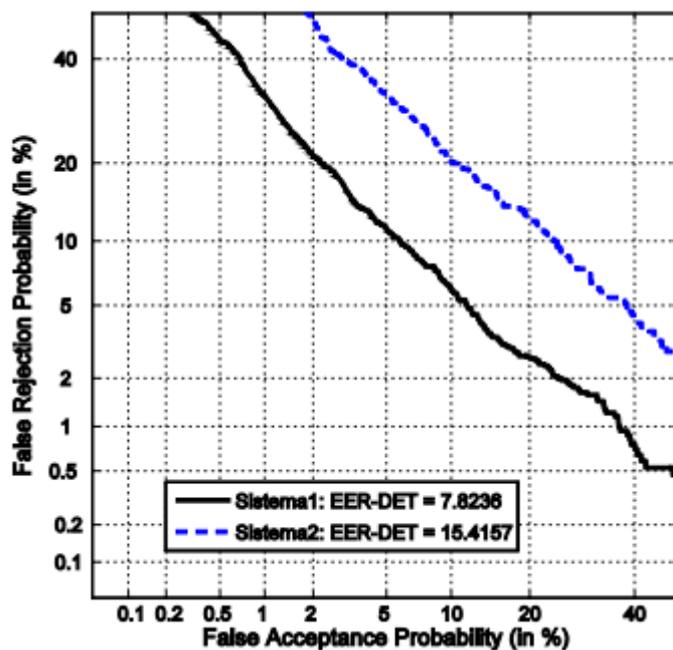
El punto en el que coinciden las dos representaciones se denomina EER (Equal Error Rate), donde  $FA=FR$  y determina el comportamiento resumido del sistema, aunque en una única posición.

### **DET (Detección Error Trade-off)**

Otro método de representación, es a través de las curvas DET (Detection Error Trade off), donde las distribuciones de verosimilitud gaussianas se vean como líneas rectas procurando una representación de la capacidad discriminativa de un sistema mediante pruebas controladas, en las que se conoce los resultados de antemano. Su representación en una única gráfica resulta muy intuitiva y útil en la comparación del rendimiento del sistema en cualquier punto.

Su cálculo se efectúa representando las probabilidades de falsa aceptación y falso rechazo en los ejes “x” e “y”, respectivamente, de esta forma, cuanto más cerca están estas curvas al origen de coordenadas, mejor es el poder discriminatorio del sistema. El punto EER se corresponde con la intersección de la curva DET del sistema con la bisectriz de los ejes, donde la probabilidad de FA y FR se igualan (definición del valor EER).

Se significa que, para dos curvas con un mismo EER, la de mayor interés forense será aquella que tenga una tasa de FA menor (evitar la culpabilidad de un inocente).



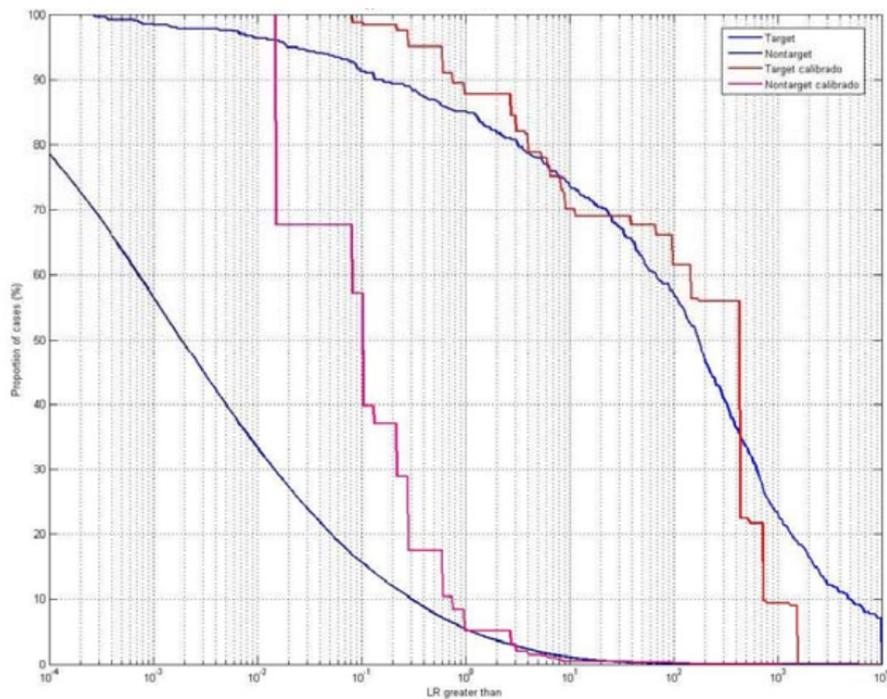
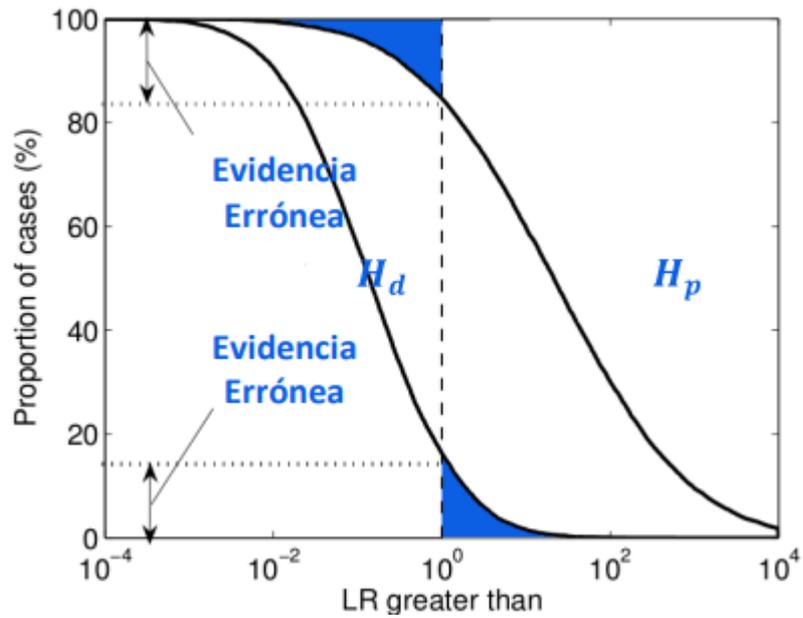
Ejemplo de curva DET

Dentro de la enorme aportación de este tipo de curvas para medir el poder discriminatorio del sistema, también muestra desventajas a la hora de medir la calidad de la calibración, mostrando su carencia a la hora de distinguir entre el rendimiento de sistemas con el mismo poder de discriminación. En este sentido, se recurre a otras técnicas que permitan evaluar el rendimiento.

### Curva Tippett

Representa la proporción de casos en los que cabe encontrar determinados valores de LR,s, cuando las voces que se comparan proceden de la misma persona, la voz pertenece al sospechoso ( $LR \gg 1$ ) o cuando proceden de individuos distintos, no pertenece al sospechoso ( $0 < LR < 1$ ). Representan las distribuciones acumulativas en dos curvas diferentes de los log-LR generados en las comparaciones de las que se conoce a priori que son target (misma fuente) y non-target (distinta fuente).

En las curvas Tippett no aparece representado ningún porcentaje de error porque en sistemas forenses estos errores, debido a que no hay ningún pronunciamiento sobre la identidad de los individuos, no existen.



Curvas Tippett

En las versiones escalonadas o calibradas se ha usado el 50% de datos target y non-target.

Están formadas por dos curvas, una de impostores (la más próxima al origen) y otra de usuarios (la más alejada del origen). Significar, que cuanto mayor sea la distancia entre curvas, mejor será la capacidad de discriminación y el rendimiento del sistema. En el eje de ordenadas se

representa el “% de casos” y en el eje de abscisas los valores de “LR mayores que”. De este modo, lo ideal será obtener curvas de locutores target cercanas al 100% de los casos y curvas de locutores non-target lo más cercanas posible al 0% de los casos.

Adquiere especial relevancia el punto  $LR=1$  (punto de inflexión entre  $H_p$  o  $H_d$ ), donde es primordial disponer de la mayor cantidad posible de locutores target y teniendo en cuenta la necesidad de mantener la presunción de inocencia, es vital minimizar la cantidad de locutores non-target, con  $LR>1$  (se ha de procurar que sea así, aún a costa de disminuir la cantidad de locutores target con puntuaciones  $LR>1$ ).

### **Medida del rendimiento de un sistema. Función de coste $C_{lr}$**

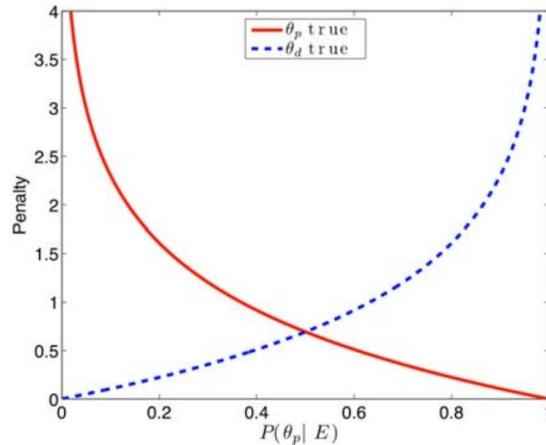
Como medida de rendimiento se ha introducido la función de coste  $C_{lr}$ , que mide el rendimiento de un conjunto de LR's obtenido de una prueba empírica [17]. Esta función se fundamenta en las “reglas de puntuación estrictamente propias” que proponen que cada valor de LR, en la comparación entre dos muestras (dubitada e indubitada), será penalizado dependiendo de la hipótesis ocurrida realmente, es decir, si un  $LR \gg 1$  y la hipótesis que ha ocurrido en realidad es que ambos materiales pertenecen a la misma fuente, la regla de puntuación dará una puntuación baja equivalente a la penalización por el error cometido, puesto que el LR es el que se esperaba. Al contrario, si lo ocurrido en realidad es que ambas muestras pertenecen a fuentes distintas, la penalización debida al  $LR>1$  será alta, con el significado de no comportarse como debiera.

Las reglas de puntuación se aplican sobre las probabilidades a posteriori, para ello es preciso conocer las probabilidades a priori y el LR.

La función de coste  $C_{lr}$ , está basada en la regla de puntuación logarítmica, donde la penalización se asigna de la forma:

- $\log_2(P(H_p/E,I))$  si  $H_p$  es cierta
- $\log_2(P(H_d/E,I))$  si  $H_d$  es cierta

Donde E: evidencia, I: información del caso que desconoce el perito.



En este contexto, se representan las siguientes medidas de rendimiento:

Precisión: es la medida global del rendimiento del sistema. Cuanto más alta sea esta medida, peor se supone el rendimiento del conjunto de LR. Su interpretación está relacionada con el coste medio de las decisiones que se tomarían utilizando el conjunto de LR (si los LR dan lugar a decisiones correctas  $C_{lr}$  disminuirá, y aumentará en caso contrario. Ese coste es promediado para todos los posibles valores de los costes de decisión y de la probabilidad a priori, y por lo tanto independiente de los mismos, por ello se considera un valor adecuado para medir el rendimiento en ciencia forense, donde el perito no se ocupa de las prioridades a priori.

Poder de discriminación: medido por el valor mínimo del  $C_{lr}$  ( $\min C_{lr}$ ). Este valor resume el poder de discriminación de un conjunto de LR, y será más alto cuanto mejor sea éste. Es una medida similar a la aportada por las curvas DET y el EER, que también miden el poder de discriminación.

Calibración: es la diferencia entre el  $C_{lr}$  y  $\min C_{lr}$ . Es una medida de lo correcta o incorrecta que es la interpretación de un conjunto de LR. Se representa por gráficos de barras.

De esta forma la función de coste  $C_{lr}$  resulta de la suma de la media de las reglas de puntuación logarítmicas para todos los LR en los que se cumple  $H_p$  más la media de las reglas logarítmicas para los LR del experimento en los que se cumple  $H_d$ . Expresado matemáticamente:

$$C_{lr} = \frac{1}{2N_{ss}} \sum_{i \in ss} \log_2 \left( 1 + \frac{1}{LR_i^{ss}} \right) + \frac{1}{2N_{ds}} \sum_{j \in ds} \log_2 (1 + LR_j^{ds})$$

Siendo:

ss: misma fuente (same source).

ds: fuentes diferentes (different source).

Cuanto mayor sea  $C_{lr}$  peor será el rendimiento medio del conjunto de puntuaciones de LR.

## **2. DESCRIPCIÓN DEL SISTEMA BATVOX**

BATVOX es un sistema forense desarrollado inicialmente por AGNITIO SL., que permite el reconocimiento automático de una voz dubitada utilizando la metodología bayesiana de valoración de la evidencia. Para ello, se realiza una comparación entre un segmento de grabación de origen desconocido con el modelo de voz generado por BATVOX, a partir de la toma indubitada de la voz a la persona sospecha.

Al objeto de obtener más fiabilidad sobre la fuerza de la evidencia, el sistema puede hacer la misma comparación con una población de referencia.

BATVOX organiza la información en dos grupos: casos y poblaciones de referencia.

Un caso almacena y organiza audios (de entrenamiento, de test, de impostor), modelos de sospechoso y tareas de reconocimiento (identificaciones y LR's).

Al objeto de obtener resultados fiables es necesario un proceso de normalización de los audios, a través del uso de impostores y/o poblaciones de referencia.

- Los audios de impostor se asignan y almacenan dentro de cada caso, junto a los audios de test y los audios de sospechoso.
- Las poblaciones de referencia se almacenan de forma independiente para ser utilizadas en todos los casos. De esta forma, los modelos añadidos al sistema, para su uso como población de referencia, no se asignan a ningún caso en particular.

Las poblaciones de referencia se pueden utilizar para realizar identificaciones, pero su uso no es obligatorio si se utilizan impostores para normalizar el proceso. No obstante, su uso es obligatorio para los cálculos de LR.

Las principales funcionalidades del sistema:

- **Gestión de casos**

El perito es capaz de organizar audios y modelos de voz en forma de casos en los cuales se ejecutan tareas de reconocimiento.

- **Tareas de reconocimiento de locutor**

El perito sólo necesita seleccionar el habla dubitada, locución indubitada del sospechoso y la población de referencia, tomada a partir de la información adicional, de esta forma el sistema emite un valor de LR para cada test de voz dubitada.

### **Tipo de Ficheros:**

Los audios que BATVOX acepta han de ser .wav con frecuencia de muestreo  $\geq 8$  kHz y una resolución  $\geq 16$  bits y no inferior, ya que los resultados se verían comprometidos.

Deberá prestarse atención a los archivos procedentes de compresiones con pérdidas como: .mp3, .mp4, .ogg, .wma, donde BATVOX provoca pérdidas de rendimiento, en este caso habrá que prestar atención a la tasa binaria, por debajo de la cual, el rendimiento decrece (para .mp3 la tasa binaria es 32 kbps) y no es recomendable su uso.

### **Calidad del audio:**

BATVOX chequea varios parámetros:

La longitud neta de un audio: es la relativa al mismo, sin considerar las zonas de silencio. Esta es una característica que influye en la fiabilidad.

- En el entrenamiento de un modelo, BATVOX permite audios cuya longitud sea  $>30$  segundos. Por debajo de este valor los modelos se convierten en poco fiables y los resultados se vuelven impredecibles.

La longitud neta óptima para entrenar un modelo es de 60 segundos.

- En relación a tareas de reconocimiento, BATVOX precisa de audios de test (voz dubitada) con una longitud  $> 10$  segundos.

La relación señal a ruido es otro factor a tener en cuenta.

- Los audios usados en BATVOX deberán tener al menos, 10 dB de relación señal a ruido (SNR)

## **2.1. CONCEPTOS DE RECONOCIMIENTO DE LOCUTOR**

La tecnología de reconocimiento de locutor se asienta en dos fases principales: entrenamiento y tareas de reconocimiento.

### **2.1.1. Entrenamiento**

#### **Conceptos relacionados:**

Sospechoso: Persona de identidad conocida. El objetivo es comparar su voz con otras desconocidas.

Audio de entrenamiento: Fichero de audio que contiene la voz del sospechoso, para crear un modelo de su voz que posteriormente se utilizará en un cálculo biométrico.

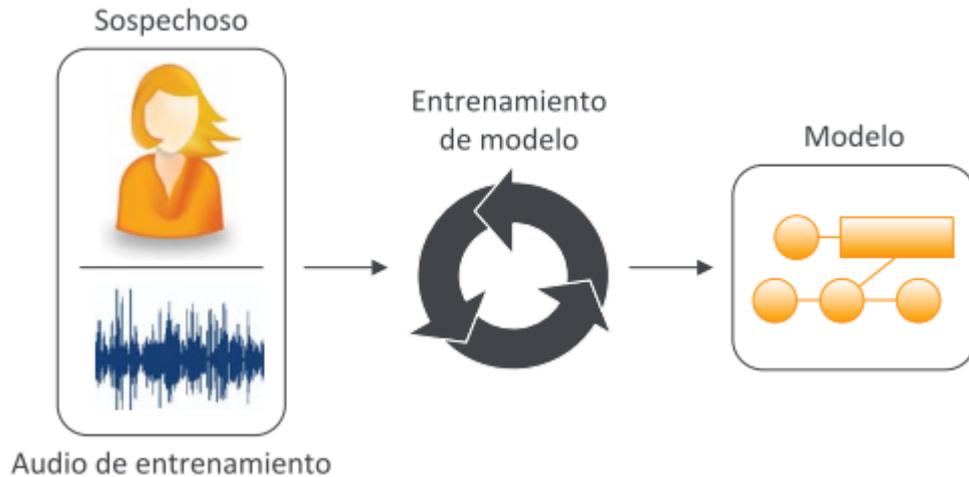
Modelo: Representa las características de la voz del locutor y es generado a partir de uno o varios ficheros de audio.

Entrenamiento de modelo: Proceso en el que se extraen las características de la voz de una persona y a partir de éstas se genera un modelo.

Existen unas condiciones mínimas que hacen a un modelo más o menos fiable.

- La longitud neta de los audios con los que se entrena un modelo influye en la fiabilidad de éste. La longitud optima recomendada en el sistema es de 60" de audio para entrenar un modelo, por debajo de esta longitud el rendimiento empeora y para longitudes muy superiores (2 ó 3 minutos) las mejoras dejan de apreciarse.
- Asimismo, la relación señal a ruido influye de manera destacada en la calidad de modelo, en este sentido, el mínimo SNR para el sistema es 10 dB, siendo optimo cuando  $SNR > 15$  dB.

Llevando estos conceptos a una representación gráfica:



### **2.1.2. Tareas de reconocimiento**

Dependiendo del objetivo perseguido, clasificamos las tareas en dos grandes grupos:

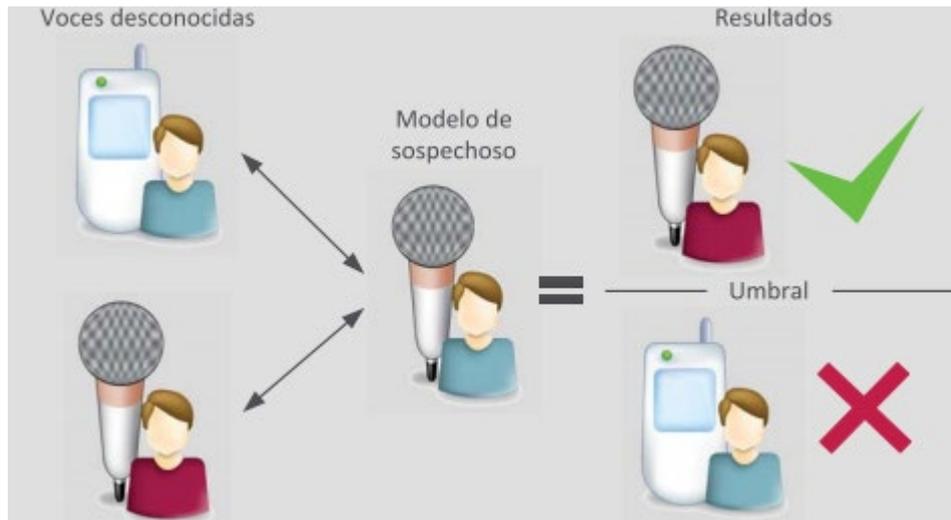
- Tareas de identificación
- Tareas de verificación

#### **Conceptos relacionados:**

Audio de test: Audio que contiene voz de identidad desconocida y que se utiliza para compararse con modelos de sospechoso con el objetivo de encontrar coincidencias para su identificación.

Población de referencia: Grupo de modelos, con las mismas características acústicas que las del modelo sospechoso, utilizados para llevar a cabo los procesos de normalización al objeto de realizar tareas de reconocimiento fiables. Cada modelo de sospechoso, objeto de reconocimiento, tendrá asociado una población de referencia.

La influencia del canal de grabación puede provocar resultados anómalos, es decir, dar puntuaciones altas con identidades distintas y bajas cuando proceden de la misma persona.



El uso de audios con distintas características provoca desalineamientos. La población de referencia ayuda a alinear las puntuaciones para determinar si el audio dubitado pertenece al sospechoso.

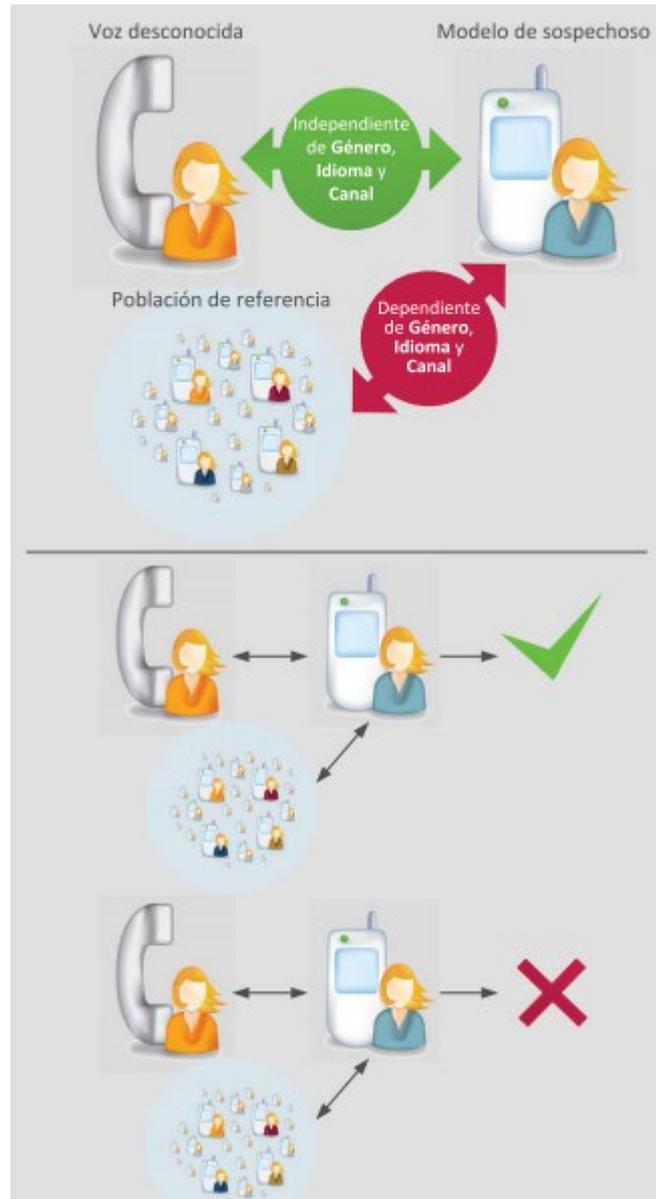
Una correcta normalización exige que el grupo de modelos usados como población de referencia, cuente con las mismas características que el modelo de sospechoso.

Audios de impostor: Audios de los que se tiene la certeza de que no pertenecen al sospechoso y que han sido grabados bajo las mismas condiciones que los audios de test. Al igual que las poblaciones de referencia, se utilizan para efectuar procesos de normalización para obtener resultados fiables en las tareas de reconocimiento.

Estos audios deben contar con más de 10" de habla y más de 10 dB de SNR. Por otro lado, es recomendable que el número de audios a utilizar para el cálculo  $> 35$  impostores.

Normalización: Proceso que se aplica en los cálculos de reconocimiento para corregir desalineamientos en las puntuaciones debido a la naturaleza de los audios. La normalización permite comparar resultados obtenidos con audios de muy distintas características (ej. canal de grabación e idioma).

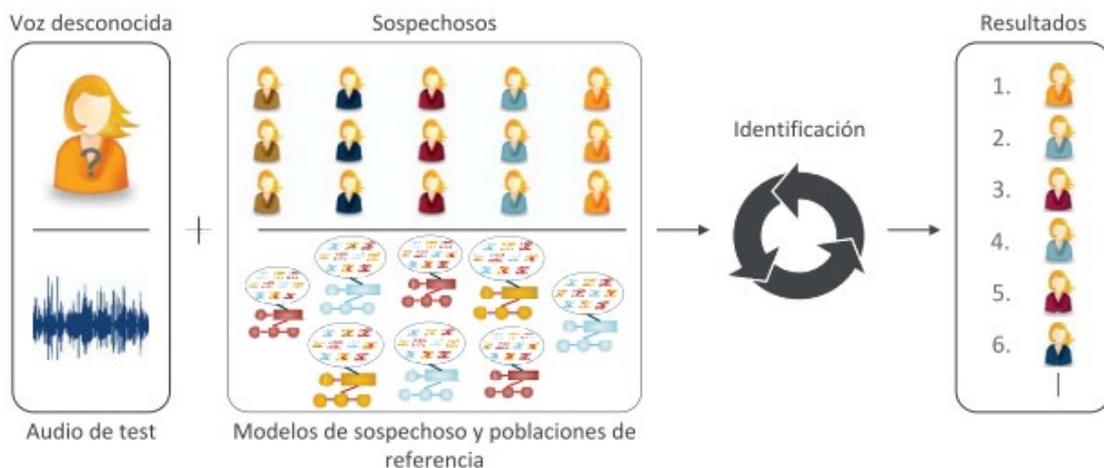
Este proceso consiste en comparar no solo el audio de test con el modelo de sospechoso sino también con todos los modelos de la población de referencia y utilizar estas puntuaciones para realizar la normalización haciendo que la comparación sea independiente del género, idioma y canal.



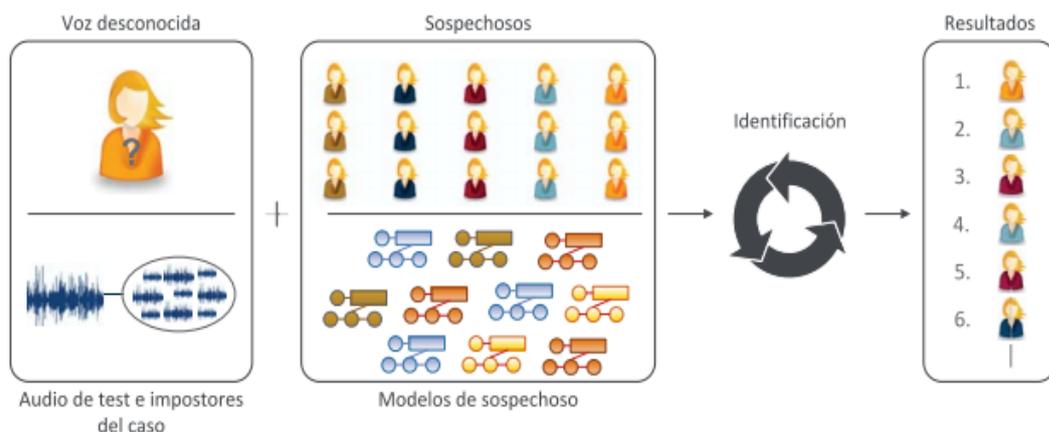
Se significa que BATVOX, solo permite la ejecución de identificaciones normalizadas. El proceso de normalización se puede llevar a cabo utilizando poblaciones de referencia, impostores o ambas. Como norma, en el proceso de normalización se utilizarán siempre poblaciones de referencia y como proceso de normalización adicional a éste, para mejorar los resultados o cuando no hay disponibles poblaciones de referencia, se utilizará impostores. Las poblaciones de referencia y los audios de impostor deberán estar correctamente adaptados a los modelos de sospecho y a los audios de test, respectivamente.

Identificación: Conjunto de comparaciones entre modelos y audio de test llevados a cabo por el sistema con el objetivo de clasificar una señal de voz, cuyo origen no conocemos, como perteneciente a uno entre N posibles locutores. El resultado de la identificación resulta una lista priorizada de modelos basada en la similitud con la voz desconocida. No obstante, en un sistema automático hay variables, como el canal que, para modelos de diferentes características (ej. grabación microfónica y telefónica) los resultados pueden aparecer falseados (ej. test en canal telefónico puntuará más con modelos telefónicos, aunque pertenezcan a personas diferentes) [18].

La normalización, en los casos en los que no tengamos la certeza de unas características similares en todos los modelos, se encargará de compensar las puntuaciones, permitiendo así poder compararlas.



Identificación normalizada utilizando poblaciones de referencia



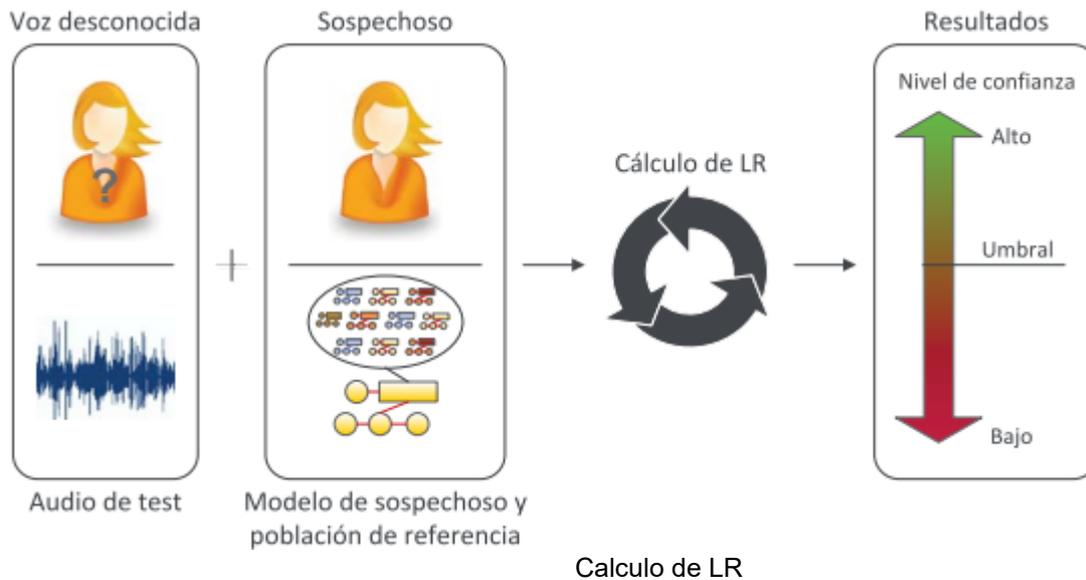
### Identificación normalizada utilizando impostores

**Verificación:** Comparación entre un modelo de sospechoso y un audio de test (1:1), devolviendo un resultado positivo o negativo. La verificación no está considerada como una buena práctica en el entorno forense ya que implica la toma de decisiones categóricas.

**Calculo de LR:** Una relación de verosimilitud o LR, es un valor probabilístico obtenido por aproximación Bayesiana. Por un lado, la probabilidad de que la voz desconocida o audio de test pertenezca al sospechoso y por otro estará la probabilidad de que la voz pertenezca a un tercero. Los cálculos de LR están considerados como el estado del arte de los sistemas de reconocimiento de locutor en el entorno forense.

Para obtener resultados fiables, se lleva a cabo un proceso de normalización en la ejecución de cálculos de LR. Hay que tener en cuenta que no se podrán realizar cálculos de LR seleccionando sólo impostores sin seleccionar población de referencia, toda vez que ésta es utilizada para calcular la probabilidad de que la voz desconocida pertenezca al resto del mundo. Así, si solo hay poblaciones de referencia, será posible lanzar identificaciones y cálculos de LR. Por otro lado, si solo hay audios de impostor, solo podrán lanzarse identificaciones ya que los cálculos de LR requieren de poblaciones de referencia.

**Umbral:** valor usado para determinar el nivel de confianza que se da a que el audio de test pertenezca al modelo de sospechoso. El umbral de LR es 1.



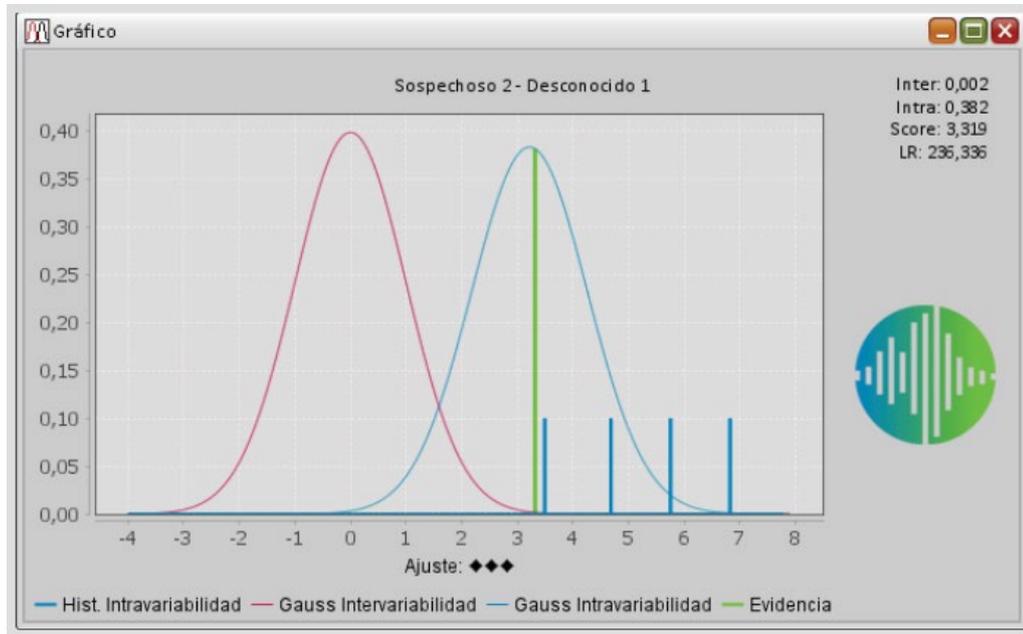
#### Interpretación de una gráfica de LR:

La distribución correspondiente a la probabilidad de que el audio de test pertenezca al sospechoso, se denomina intravariabilidad “within-source” (curva azul), estima la puntuación del sospechoso contra sí mismo. La distribución correspondiente a la probabilidad de que el audio de test pertenezca a terceros, se denomina intervariabilidad “between-source” (curva roja) y representa la puntuación del audio de test cuando se enfrenta a una población de referencia (modelos que no son la misma persona que el audio de test). De este modo, si la puntuación del audio de test frente al modelo de sospechoso (evidencia) cae en la zona de la curva de intervariabilidad, lo más probable es que no pertenezcan a la misma persona y al contrario se cae en la zona de la curva de intravariabilidad, la probabilidad de que pertenezca a la misma persona aumenta.

El cociente entre la altura de una curva y la otra en un punto donde haya puntuado el audio de test es precisamente el valor del LR.

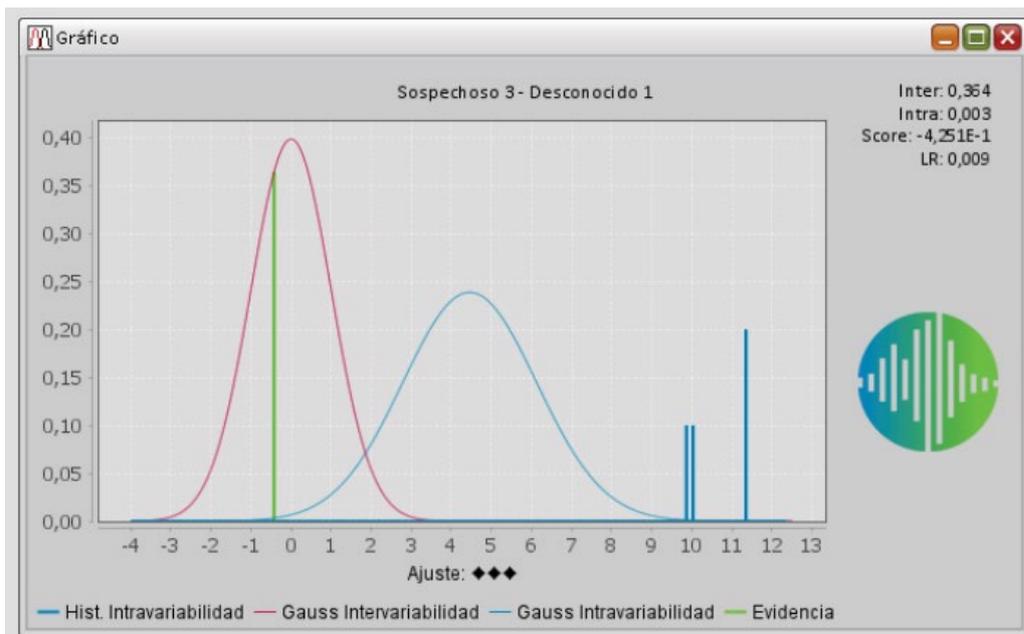
Para un resultado de 1, significa que la probabilidad de que el audio de test pertenezca al sospechoso es la misma que la probabilidad de que pertenezca a un tercero, no pudiendo inclinar la balanza hacia ninguna hipótesis.

Ej. El LR indica que es 236.336 veces más probable que el audio de test y el sospechoso sean la misma persona que el que pertenezca a un tercero.



Gráfica de LR en BATVOX

Ej. Indica que es 111 veces más probable que los audios no pertenezcan a la misma persona que el que si lo sean.



### **2.1.3. Herramienta EvalBatvox**

EvalBatvox tiene una estructura de carpetas basada en tres elementos principales posicionados al mismo nivel cada uno de ellos.

**Locutores:** Dentro de esta carpeta, hay tantas carpetas como locutores queramos evaluar. A su vez, dentro de la carpeta “LocutorN” tenemos dos subcarpetas “Entrenar” y “Target”. Todos los archivos de audio dentro de “Entrenar” se usarán para construir el modelo de locutor N (únicamente uno, independientemente del número de audios de que se disponga). Por otra parte, todos los audios dentro de “Objetivo” serán usados como test de audio.

**Población de referencia:** Dentro de esta carpeta, tenemos tantas carpetas como locutores queramos incluir como población de referencia. En este sentido, para cada locutor Y habrá una carpeta “PoblaciónY”. Todos los archivos de audio serán usados para generar un modelo. Todos los modelos creados se usarán para construir T-norm, como parte de S-norm.

**Impostores:** Dentro de la carpeta “Impostores”, tendremos tantas carpetas como hablantes tenemos en la carpeta “Locutores”. Así, cada “LocutorN” en la carpeta “Locutores” tiene la correspondiente carpeta (con el mismo nombre) en la carpeta “Impostores”. Todos los archivos de audio dentro de “LocutorN” se usará para construir z-norm.

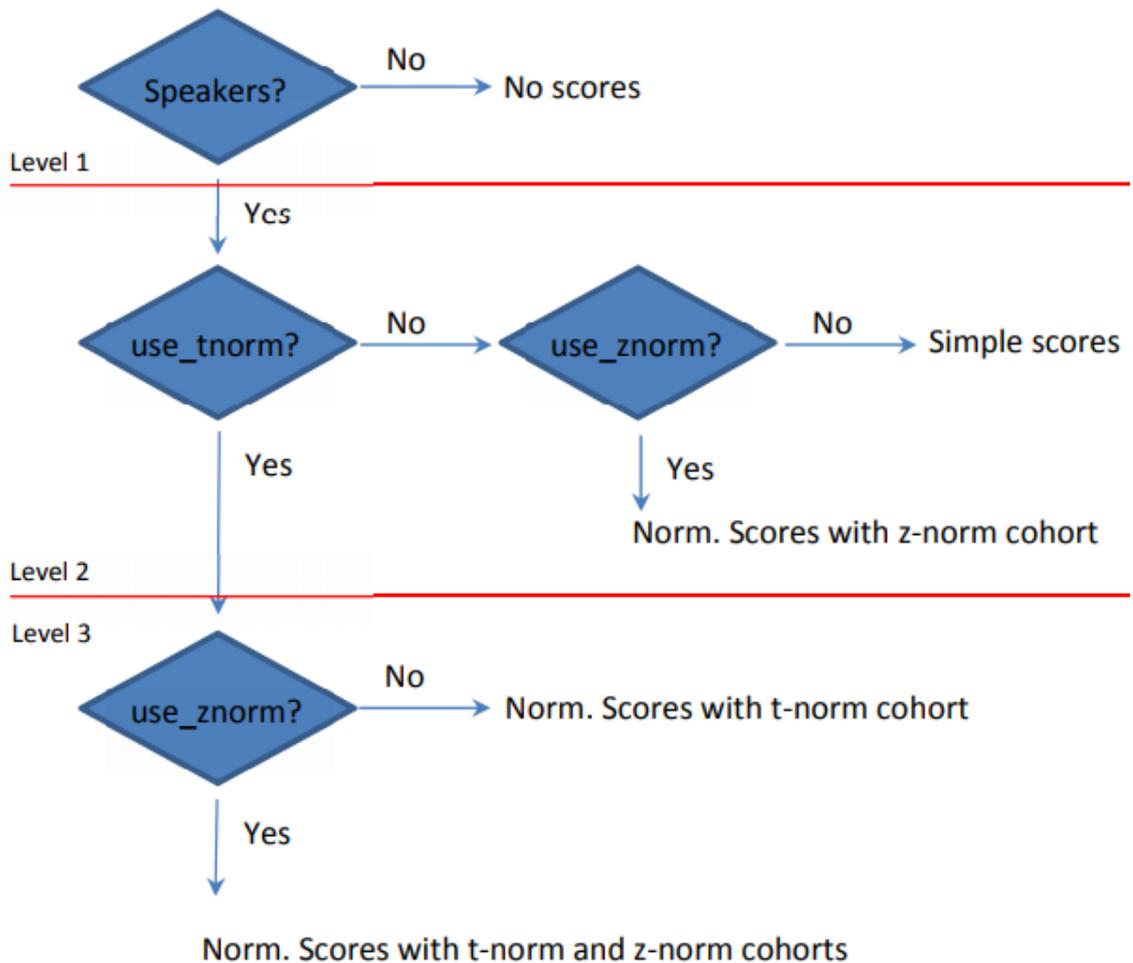
Puntuación normalizada de entrada para Eval

Aunque la estructura de carpeta podría usarse para determinar el tipo de puntuación normalizada que se requeriría, algunas entradas se usan para aumentar la flexibilidad de BATVOX EVAL.

En este sentido, se han definido las entradas “use\_tnorm” y “use\_znorm” que determinan si se utiliza el grupo t-norm y/o z-norm, independientemente de que exista o no la carpeta “Población de referencia” y/o “Impostores”.

Algunas otras entradas son obligatorias como el número de impostores (“numero\_impostores”) u opcionales como el número de modelos (“numero\_modelos”).

El flujo de trabajo para un locutor N de la carpeta “locutores” es:



### Construcción de la cohorte z-norm.

Cuando “use\_znorm” esta activado, se construirá una cohorte z-norm con el número de elementos especificados por el parámetro de entrada “número\_impostores”. Suponiendo que estamos usando el modelo del locutor N y un test del locutor T para calcular la puntuación simple, existen dos opciones para generar la cohorte z-norm:

- I. Existe la carpeta “Impostores/LocutorN/”. Partiendo de poder extraer R impostores de esa carpeta, se darán tres opciones:
  - i.  $R > \text{“número\_impostores”}$ : entonces eliminamos  $R - \text{“número\_impostores”}$  impostores para construir la cohorte z-norm. Se

lleva a cabo para proporcionar resultados consistentes cuando se repite el experimento.

- ii.  $R < \text{“número\_impostores”}$ . Necesitaremos más impostores para crear la cohorte de z-norm, utilizando para ello archivos adicionales de la carpeta “Target”, con las limitaciones siguientes:
  - No seleccionar archivos del locutor N o T.
  - Maximizar el número de locutores.
  - Robustez del experimento cuando se repite.

Si habiendo utilizado todos los archivos incluidos en la subcarpeta “Target”, no se alcanza el número de impostores, la cohorte z-norm se construirá con el mayor número posible de archivos.

- iii.  $R = \text{“numero\_impostores”}$ , no será preciso ningún ajuste de la cohorte z-norm.

- II. La carpeta “Impostores/LocutorN/” no existe. Se usará para ese propósito los archivos target de otros locutores siguiendo el procedimiento anterior.

La selección de impostores será siempre la misma para mantener la coherencia entre las ejecuciones con los mismos datos. Cualquier conjunto de archivos de audio que sirvan como impostores se ordenará por voz neta y si tuvieran la misma cantidad de voz, se ordenará por nombre del audio.

1. Después de ordenados, si la cohorte de impostores necesita más archivos de audio, cogerá archivos de audio de locutor.

## **Escenarios de validación de audio**

Audios de entrenamiento de locutor: se entrenará cualquier modelo a menos que no cumpla la condición:

- Calidad del modelo (suma neta de voz y SNR media de todos los audios) debe reunir los umbrales de entrenamiento (30” de voz neta y 10 dB para SNR).  
Si uno de los audios de entrenamiento no tiene más de 3” de voz neta (límite de la distorsión), se descartará.

Audios de sospechoso: cada audio debe superar los umbrales de prueba habituales (7" de voz neta y 10 dB para SNR).

Audios de población de referencia: como el entrenamiento de locutor, si un archivo de audio no pasa el validador de 3", se descarta el mismo.

Audios de impostor: cada audio será validado contra los umbrales de validación del test.

### **3. ANALISIS DE VARIABLES**

Se pretende estudiar el impacto que puede llegar a tener la elección del tipo de archivo a utilizar como test en algunas de las situaciones que podemos encontrar en un análisis forense. Las pruebas de han realizado con EVALBATVOX 4.0, partiendo de la situación de referencia con modelos y test en canal SITEL con habla espontánea. Posteriormente se ha ido modificando el canal de los test a los formatos: AAC, MP3 y OGG.

En todas las pruebas los archivos utilizados, tanto modelos como test, son los mismos en cuanto a su contenido acústico; la única variable que cambia es el canal de codificación de los test.

Las tareas, para cada uno de los casos, consisten en el enfrentamiento de 91 modelos de 120 segundos contra 442 archivos de test de 20 segundos, de los cuales 39780 son enfrentamientos non-target y 442 son target.

#### **Curvas DET y EER**

Como regla, cuanto más cerca esta la curva al origen, mejor es el poder de discriminación del sistema.

Ofreciendo una medida resumen del rendimiento tenemos el valor EER (Equal Error Rate), que corresponde al punto de la curva DET en el que se iguala la probabilidad de error de falsa aceptación y falso rechazo (este punto correspondería con el cruce de la curva con la bisectriz de los ejes de coordenadas). Cuanto menor sea el valor EER, mejor será el rendimiento.

#### **Curvas Tippett**

Se compone de dos curvas:

- Curva target (cierta la hipótesis del fiscal): es la proporción de LR mayores que el valor en el eje x en el conjunto de LR calculado cuando la hipótesis del fiscal es cierta.
- Curva non target (cierta la hipótesis del defensor): representa la proporción de LR mayores que el valor en el eje x en el conjunto de LR calculado, cuando la hipótesis del defensor es cierta.

### $C_{lr}$ y $\min C_{lr}$

El valor  $C_{lr}$ , se corresponde con la tasa de error real del sistema tal y como esta calibrado.

El  $\min C_{lr}$  representa la probabilidad de error del sistema de referencia, supuesta una calibración perfecta. Cuanto menor sea su valor, mayor será su poder de discriminación.

Representa las medidas de rendimiento de:

- Precisión: Este valor viene a valorar el coste medio de las decisiones tomadas bajo un conjunto de LR. Si de los LR se obtienen decisiones correctas,  $C_{lr}$  disminuirá y aumentará en caso contrario. Cuanto más alta sea esta medida peor es el rendimiento del conjunto de LR. El valor del  $C_{lr}$  debería figurar entre los valores de 0 y 1, para valores  $>1$  no conviene utilizar el conjunto de LR. El rendimiento de los LR es tanto mejor cuanto más próximo a 0 sea el valor  $C_{lr}$ .
- Discriminación: medido por el valor mínimo de  $C_{lr}$ . Representado como pérdidas de discriminación. El poder discriminatorio de un conjunto de LR será más alto cuanto mejor sea este valor.
- Calibración: es la diferencia entre  $C_{lr}$  y  $\min C_{lr}$ . Cuanto menor sea esta diferencia mejor es la calibración del sistema.

La disposición de los test y modelos objeto del enfrentamiento en cada uno de los casos a estudio se puede resumir como:

	CASO 1	CASO 2	CASO 3	CASO 4
TEST	SITEL	AAC	MP3	OGG
MODELO	SITEL	SITEL	SITEL	SITEL
POBLACIÓN	SITEL	SITEL	SITEL	SITEL

### 3.1. CASO 1: Valores de referencia - Test: Ahumada IV (SITEL) VS

#### Modelo: Ahumada IV (SITEL)

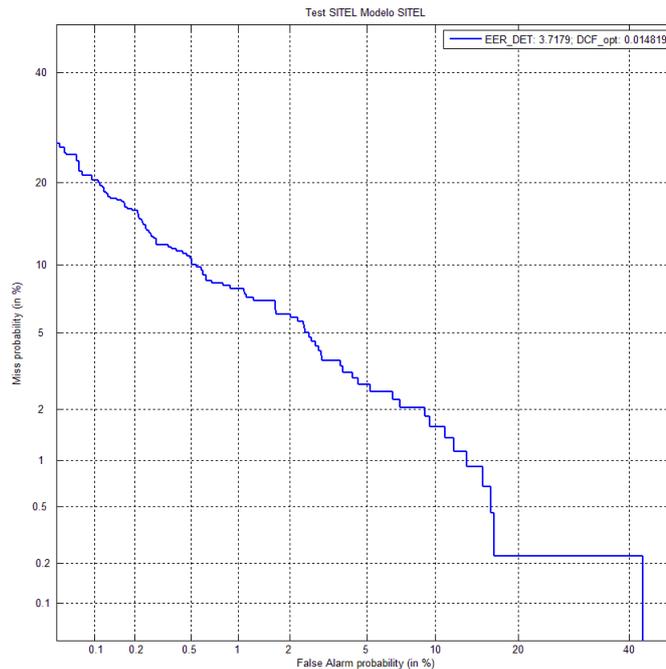
En esta prueba se han utilizado test y modelos dentro de una población masculina, tanto de la población de referencia como de los de locutor, procedentes de Ahumada IV (SITEL). La prueba es el resultado del enfrentamiento de 91 modelos de 120 segundos contra 442 archivos de test de 20 segundos, de los cuales 39780 son enfrentamientos non-target y 442 son target.

Se seleccionan los mejores 35 competidores de 91 en la población de referencia.

#### 3.1.1. Resultados y curvas comparativas

##### Curva DET

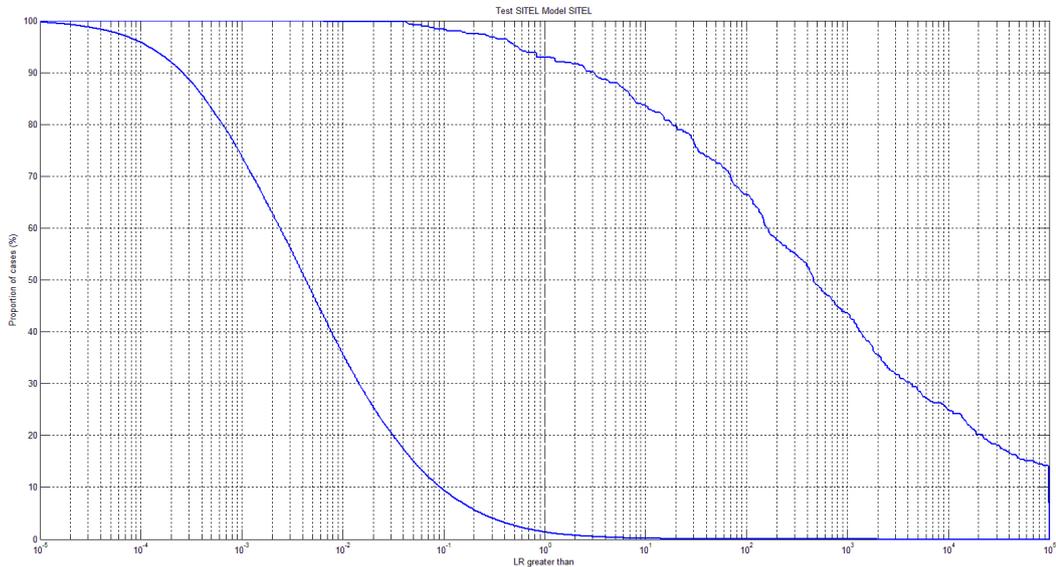
La medida de rendimiento obtenida EER es de 3.72%.



##### Curva Tippett

El sistema ofrece tasas de evidencia errónea de un 1,4% para valores de  $LR > 1$  cuando se refuerza la hipótesis del fiscal (que las voces cotejadas provengan del mismo individuo), siendo cierta la hipótesis contraria y del 7% para valores de  $LR > 1$  cuando se refuerza la hipótesis del

defensor (las voces cotejadas provengan de distintos individuos) siendo cierta la hipótesis contraria.

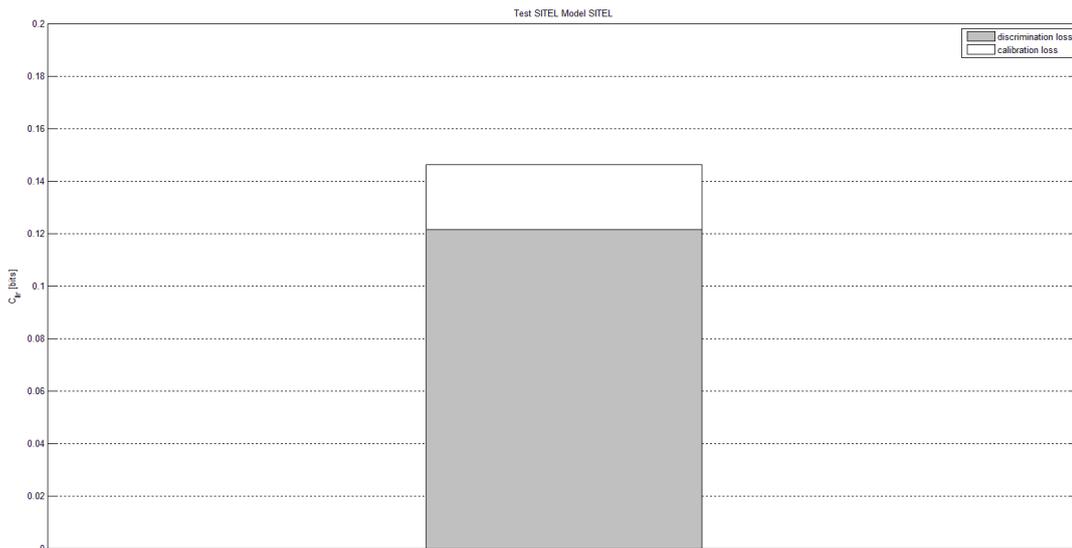


### Curva APE, $C_{lr}$ y $\min C_{lr}$

Para el cálculo de la curva APE se utilizan los mismos datos que para el cálculo de la Tippett.

Optamos por un diagrama de barras que representa la medida de la función de coste  $C_{lr}$  que cuantifica la probabilidad de error de las probabilidades a posteriori a favor de la hipótesis del Fiscal, dadas las posibilidades a priori a favor de ésta, así como las pérdidas de discriminación y calibración del sistema.

En este sentido, se observa un rendimiento en discriminación de 0.121, y unas pérdidas de calibración muy bajas,  $(0.146-0.121=0.025)$ . Por lo tanto, el sistema se estima fiable para su utilización en este escenario.



### **3.2. CASO 2: Test: Ahumada IV (AAC) VS Modelo: Ahumada IV (SITEL)**

En esta prueba se han utilizado hablantes masculinos, test Ahumada IV (AAC) de 20 segundos y modelos Ahumada IV (SITEL) de 120 segundos, tanto de la población de referencia como los de locutor.

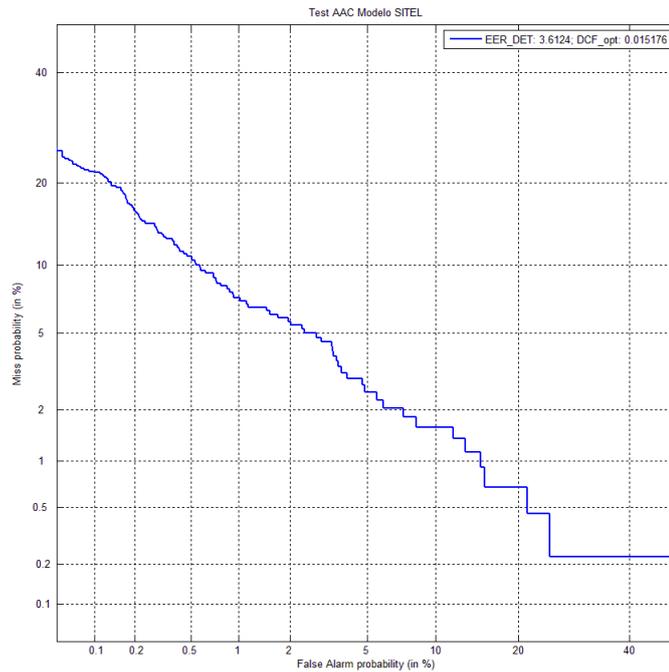
La prueba es el resultado del enfrentamiento de 91 modelos contra 442 archivos de test, de los cuales 39780 son non-target y 442 son target.

Se seleccionan los 35 mejores competidores de 91 en la población de referencia.

#### **3.2.1. Resultados y curvas comparativas**

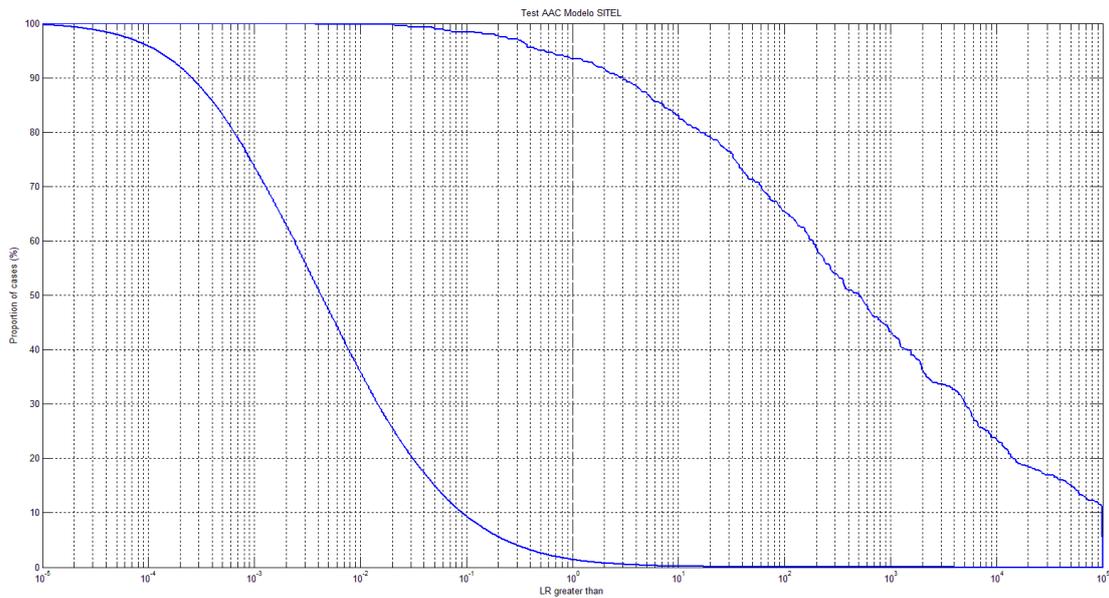
##### **Curva DET**

En este caso tenemos un parámetro EER de 3.61%.



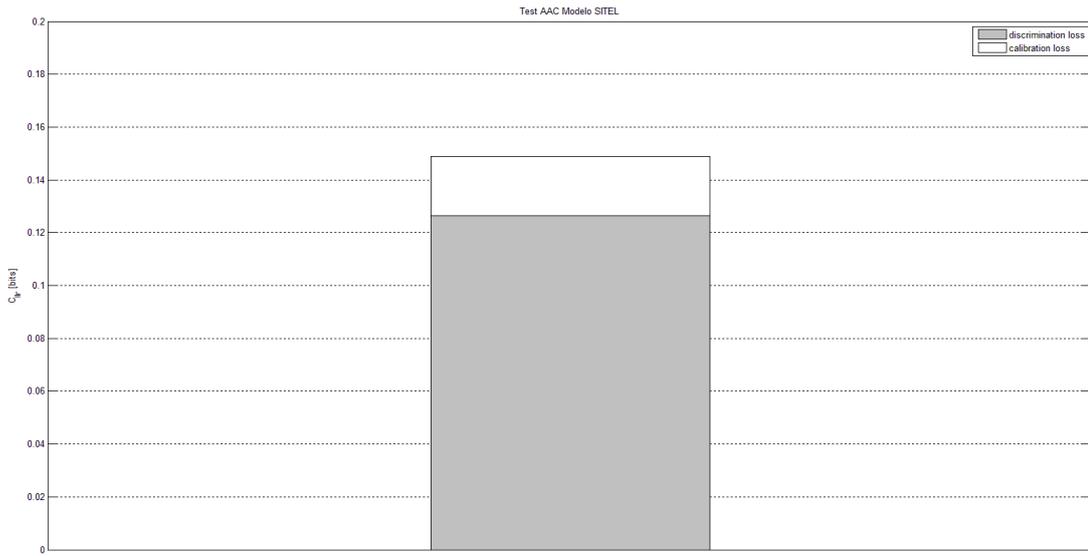
### Curva Tippett

El sistema ofrece tasas de evidencia errónea de un 1,4% para valores de  $LR > 1$  cuando se refuerza la hipótesis del fiscal (que las voces cotejadas provengan del mismo individuo), siendo cierta la hipótesis contraria y del 6.6% para valores de  $LR > 1$  cuando se refuerza la hipótesis del defensor (las voces cotejadas provengan de distintos individuos) siendo cierta la hipótesis contraria.



### Curva APE, C<sub>llr</sub> y minC<sub>llr</sub>

En la gráfica se pueden observar los valores de C<sub>llr</sub>, pérdida de discriminación y pérdidas de calibración. De nuevo, se observa que las dos para el conjunto de LR son buenas, con pérdidas muy bajas, obteniendo unas pérdidas de discriminación 0.126, y unas pérdidas de calibración de (0.149-0.126=0.023).



### 3.3. CASO 3: Formato MP3 - Test: Ahumada IV (MP3) VS Modelo: Ahumada IV (SITEL)

En esta prueba se han utilizado hablantes masculinos con test procedentes de Ahumada IV (MP3) de 20 segundos y modelos, tanto de población de referencia como de locutor, provenientes de Ahumada IV (SITEL), 120 segundos.

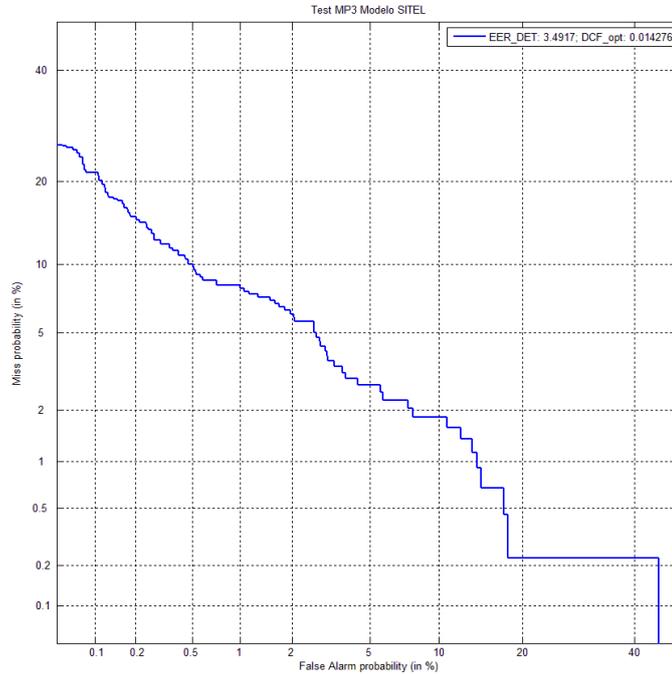
La prueba es el resultado del enfrentamiento de 91 modelos contra 442 archivos de test, de los cuales 442 enfrentamientos son target y 39780 son non-target.

Se realiza una selección de los 35 mejores competidores de 91 en la población de referencia.

#### 3.3.1. Resultados y curvas comparativas

## Curva DET

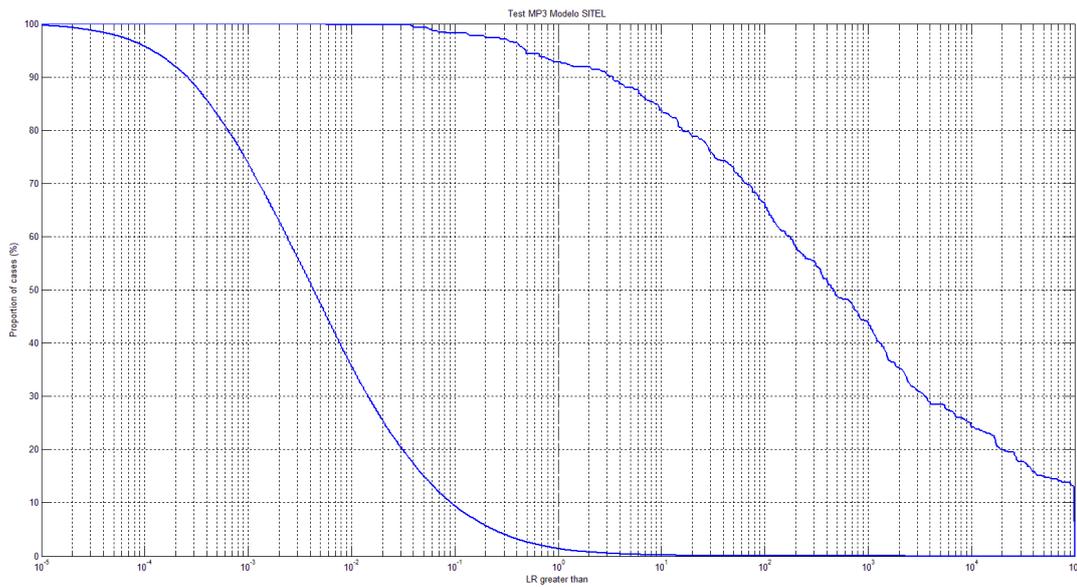
En este caso el parámetro de EER obtenido es de 3.49%



De esta curva no se desprende una conclusión clara del rendimiento del sistema, para ello debemos recurrir a valores de  $\min C_{llr}$ .

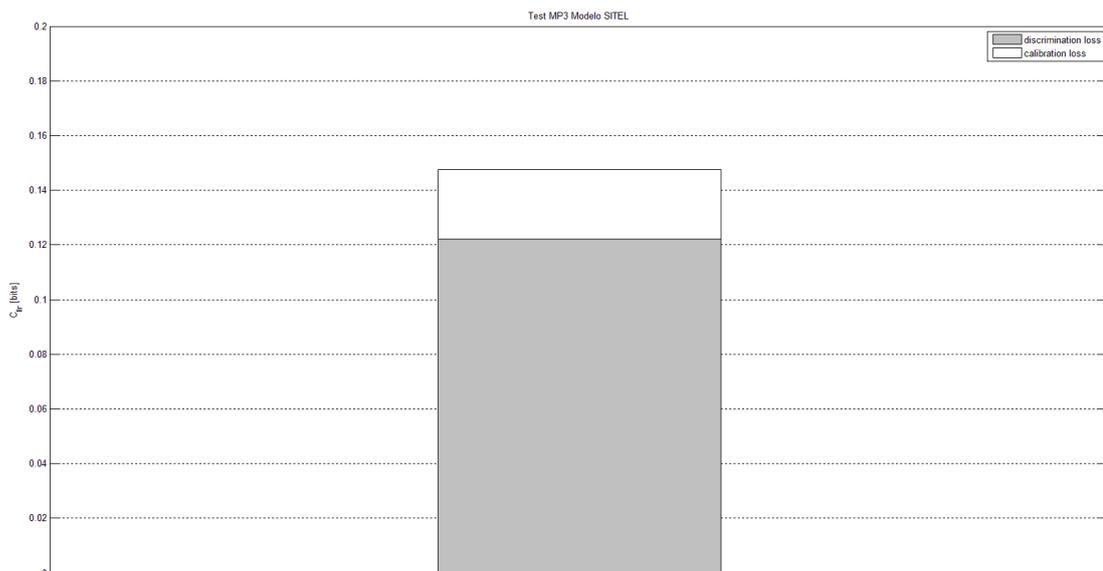
## Curva Tippett

El sistema ofrece tasas de evidencia errónea de un 1,4% para valores de  $LR > 1$  cuando se refuerza la hipótesis del fiscal (que las voces cotejadas provengan del mismo individuo), siendo cierta la hipótesis contraria y del 7.2% para valores de  $LR > 1$  cuando se refuerza la hipótesis del defensor (las voces cotejadas provengan de distintos individuos) siendo cierta la hipótesis contraria.



### Curva APE, $C_{lr}$ y $\min C_{lr}$

En la gráfica se pueden observar los valores de  $C_{lr}$ , pérdida de discriminación y pérdidas de calibración. Se observa que en el sistema la pérdida de discriminación 0.122 y unas pérdidas de calibración ( $0.149-0.122=0.027$ ) muy bajas y, por lo tanto, en este escenario el sistema se puede estimar como fiable.



### 3.4. CASO 4: Test: Ahumada IV (OGG) VS Modelo: Ahumada IV (SITEL)

En esta prueba se han utilizado hablantes masculinos con test procedentes de Ahumada IV (OGG), 20 segundos y los modelos, tanto de población de referencia como los de locutor, provenientes de Ahumada IV (SITEL), 120 segundos.

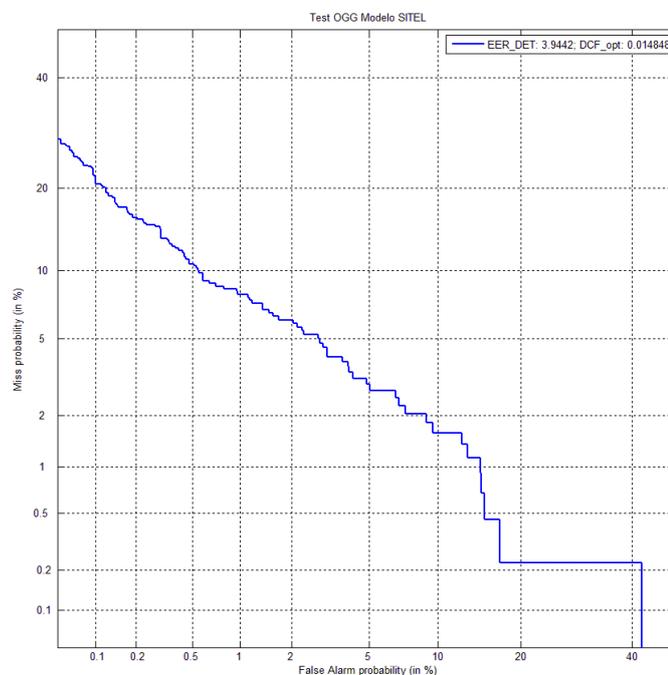
La prueba es el resultado del enfrentamiento de 91 modelos contra 442 archivos de test, de los cuales 442 enfrentamientos son target y 39780 son non-target.

Se seleccionan los 35 mejores competidores de 91 en la población de referencia.

#### 3.4.1. Resultados y curvas comparativas

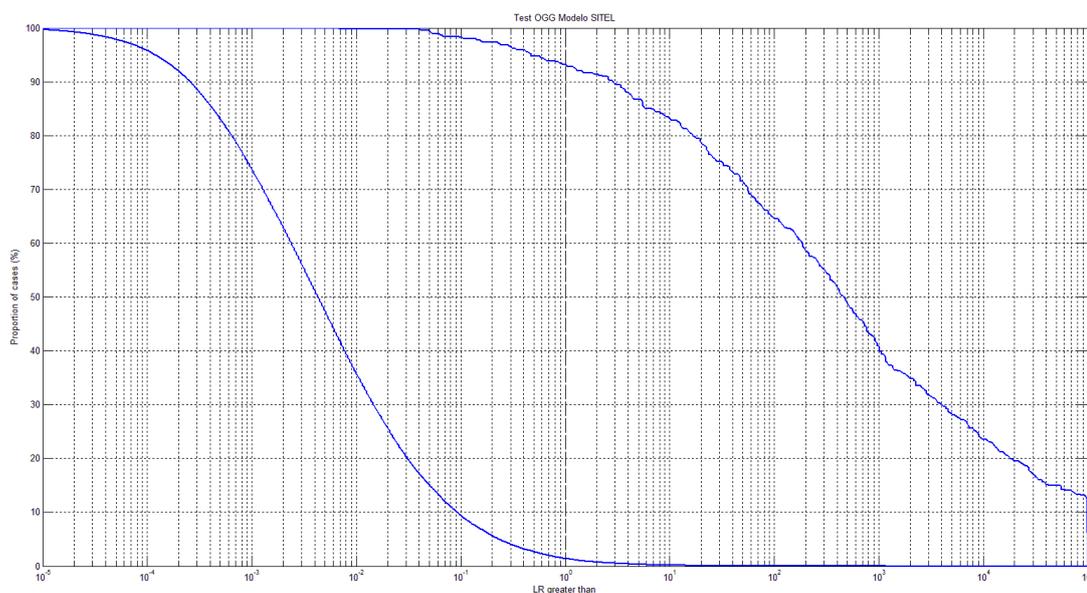
##### **Curva DET**

En este caso el parámetro de EER obtenido es nuevamente de 3.94%



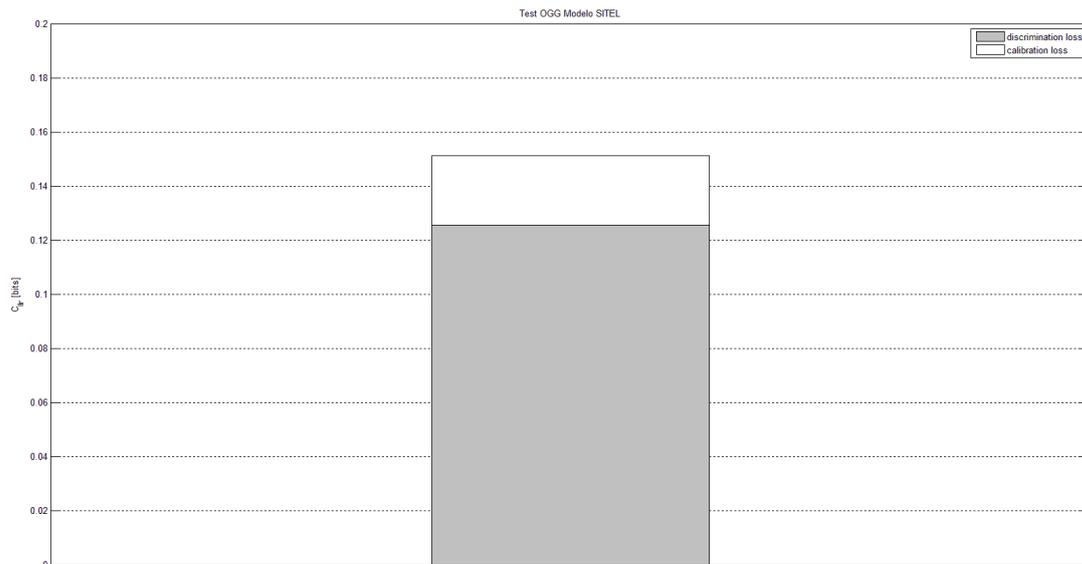
##### **Curva Tippett**

El sistema ofrece tasas de evidencia errónea de un 1,4% para valores de  $LR > 1$  cuando se refuerza la hipótesis del fiscal (que las voces cotejadas provengan del mismo individuo), siendo cierta la hipótesis contraria y del 6.8% para valores de  $LR > 1$  cuando se refuerza la hipótesis del defensor (las voces cotejadas provengan de distintos individuos) siendo cierta la hipótesis contraria.



### Curva APE, $C_{lr}$ y $\min C_{lr}$

En la gráfica se pueden observar los valores de  $C_{lr}$ , pérdida de discriminación y pérdidas de calibración. Como se visualizó en gráficas anteriores se observan resultados con una pérdida muy baja tanto en discriminación (0.126) como en calibración ( $0.151 - 0.126 = 0.025$ ) y, por lo tanto, en este escenario el sistema se puede estimar como fiable.



Una vez realizadas las curvas características de cada CASO, vamos a ver los efectos producidos al llevar a cabo comparaciones cuando enfrentamos test en los tipos de canales que vamos variando, mantenido fijo los modelos de una población (SITEL).

En esta ocasión se unifica en una gráfica el caso de referencia (SITEL) con cada uno de los formatos objeto de estudio AAC, MP3, OGG.

### **3.5. COMP 1: Test: Ahumada IV (SITEL) y Test: Ahumada IV (AAC) VS Modelo: Ahumada IV (SITEL)**

En esta prueba se han utilizado hablantes masculinos con test procedentes de Ahumada IV (SITEL y AAC), 20 segundos y los modelos, tanto de población de referencia como los de locutor, provenientes de Ahumada IV (SITEL), 120 segundos.

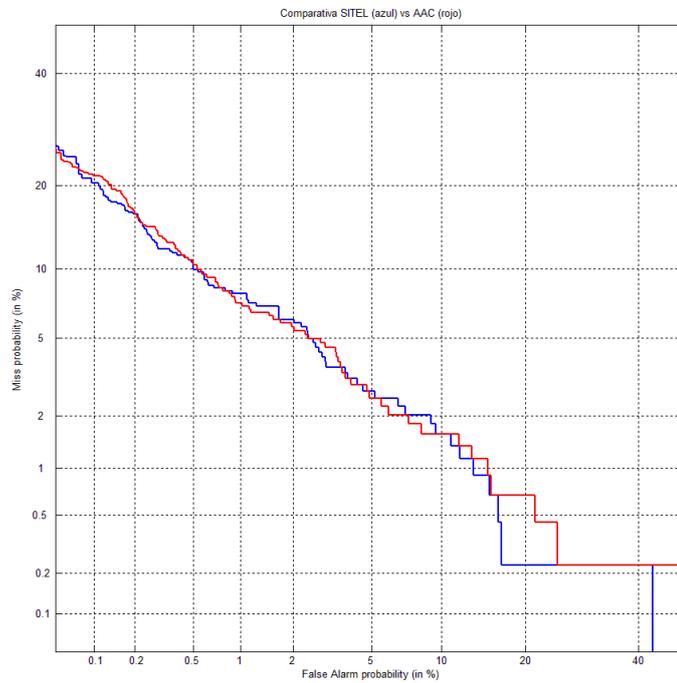
La prueba es el resultado del enfrentamiento de 91 modelos contra 442 archivos de test, de los cuales 442 enfrentamientos son target y 39780 son non-target.

Se seleccionan los 35 mejores competidores de 91 en la población de referencia.

#### **3.5.1. Resultados y curvas comparativas**

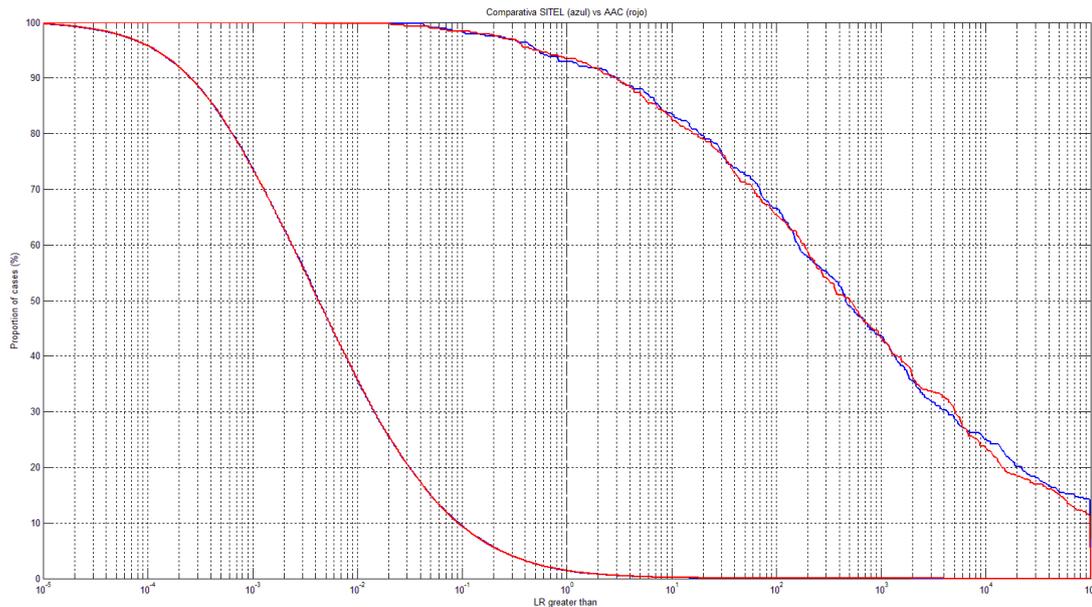
##### **Curva DET**

En este caso el parámetro de EER obtenido es nuevamente de 3.72% y 3.61% para las curvas de SITEL y AAC, respectivamente.



### Curva Tippett

El sistema ofrece nuevamente tasas de evidencia errónea de un 1.4% que apoya la hipótesis del fiscal y 7% y 6.60% que apoya la hipótesis del defensor, evaluadas para valores de LRs > 1, para las distribuciones correspondientes a los formatos SITEL y AAC.



### **3.6. COMP 2: Test: Ahumada IV (SITEL) y Test: Ahumada IV (MP3) VS Modelo: Ahumada IV (SITEL)**

En esta prueba se han utilizado hablantes masculinos con test procedentes de Ahumada IV (SITEL y MP3), 20 segundos y los modelos, tanto de población de referencia como los de locutor, provenientes de Ahumada IV (SITEL), 120 segundos.

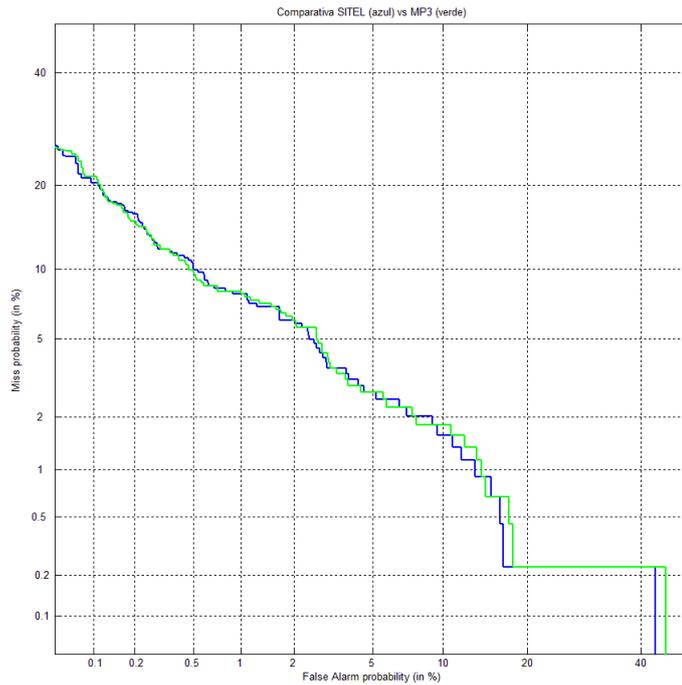
La prueba es el resultado del enfrentamiento de 91 modelos contra 442 archivos de test, de los cuales 442 enfrentamientos son target y 39780 son non-target.

Se seleccionan los 35 mejores competidores de 91 en la población de referencia.

#### **3.6.1. Resultados y curvas comparativas**

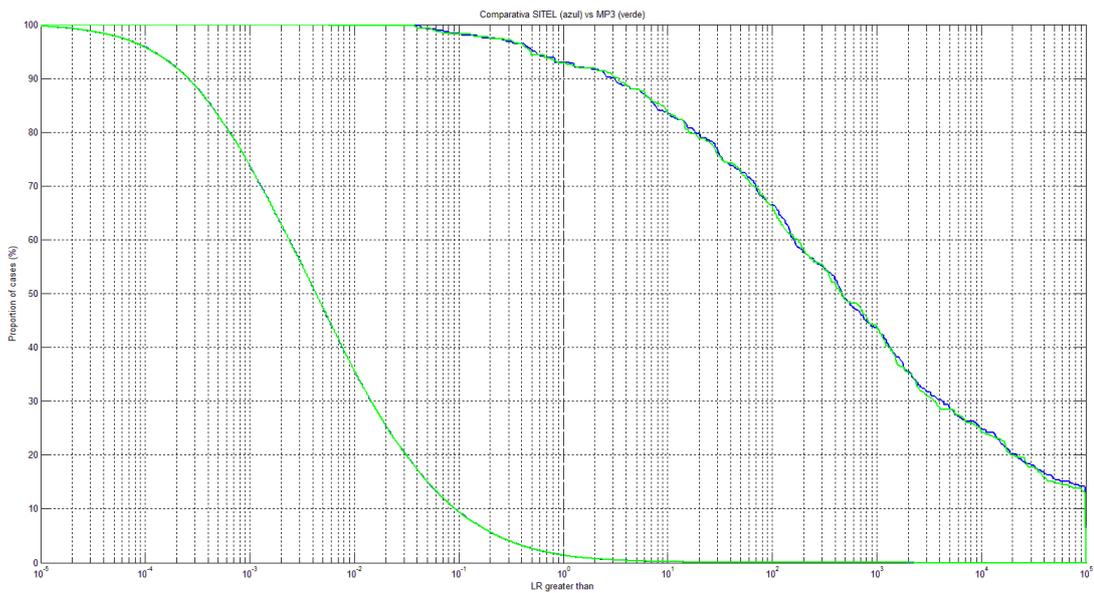
##### **Curva DET**

En este caso el parámetro de EER obtenido es de 3.72% y 3.49% para las curvas de SITEL y MP3, respectivamente.



## Curva Tippett

El sistema ofrece nuevamente tasas de evidencia errónea de un 1.4% que apoya la hipótesis del fiscal y de 7% y 7.2% que apoya la hipótesis del defensor, evaluadas para valores de LRs >1, para las distribuciones correspondientes a los formatos SITEL y MP3.



### **3.7. COMP 3: Test: Ahumada IV (SITEL) y Test: Ahumada IV (OGG)**

#### **VS Modelo: Ahumada IV (SITEL)**

En esta prueba se han utilizado hablantes masculinos con test procedentes de Ahumada IV (SITEL y OGG), 20 segundos y los modelos, tanto de población de referencia como los de locutor, provenientes de Ahumada IV (SITEL), 120 segundos.

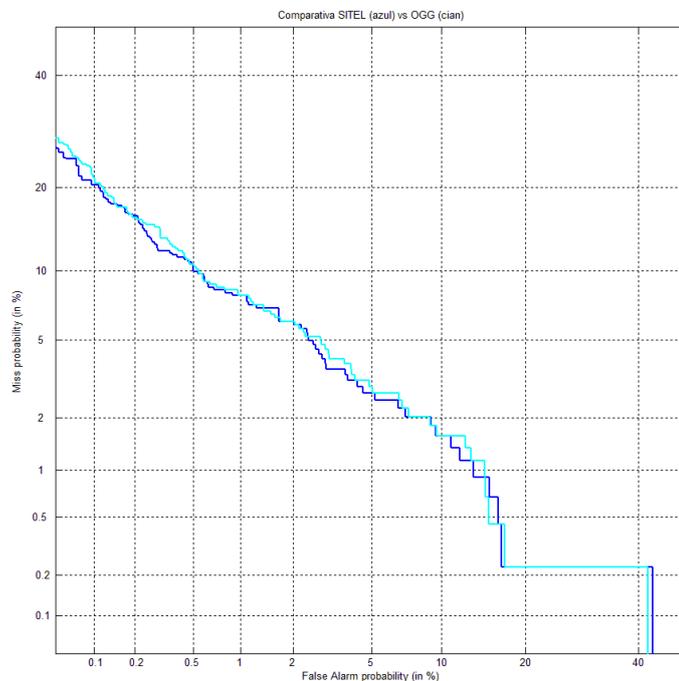
La prueba es el resultado del enfrentamiento de 91 modelos contra 442 archivos de test, de los cuales 442 enfrentamientos son target y 39780 son non-target.

Se seleccionan los 35 mejores competidores de 91 en la población de referencia.

#### **3.7.1. Resultados y curvas comparativas**

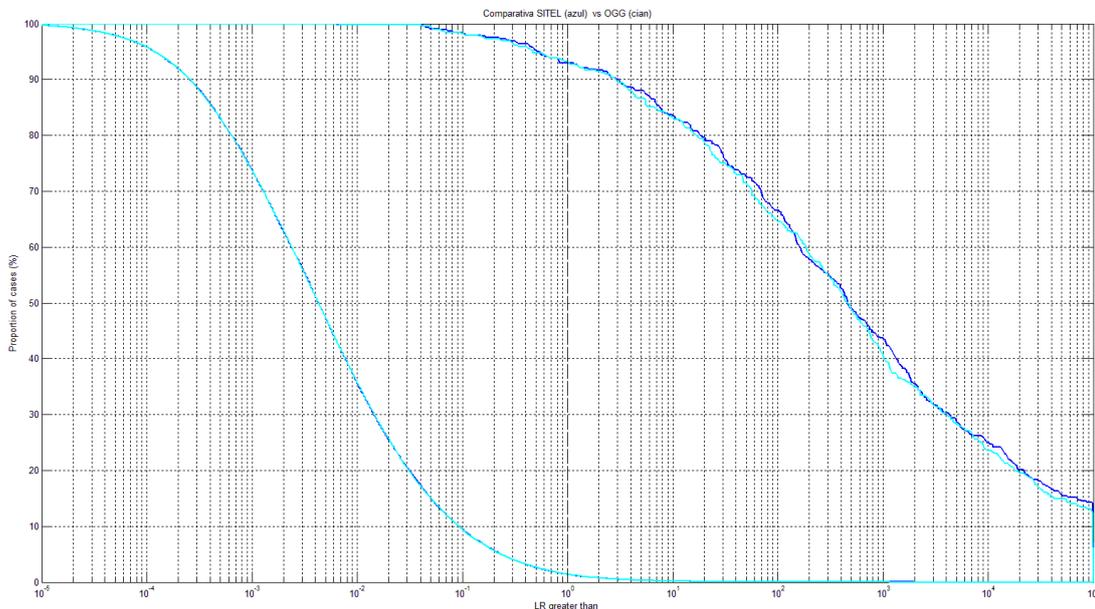
##### **Curva DET**

En este caso el parámetro de EER obtenido es de 3.72% y 3.94% para las curvas de SITEL y OGG, respectivamente.



##### **Curva Tippett**

El sistema ofrece nuevamente tasas de evidencia errónea de un 1.4% que apoya la hipótesis del fiscal y 7% y 6.8% que apoya la hipótesis del defensor, evaluadas para valores de LRs >1, para las distribuciones correspondientes a los formatos SITEL y OGG.



### **3.8. COMP 4: Test: Ahumada IV (SITEL), Test: Ahumada IV (AAC), Test: Ahumada IV (MP3) y Test: Ahumada IV (OGG) VS Modelo: Ahumada IV (SITEL)**

Por último, con esta comparativa procuramos reflejar en una sola gráfica los resultados obtenidos. Para ello utilizamos todos los test de que disponemos, hablantes masculinos con procedentes de Ahumada IV (SITEL, AAC, MP3 y OGG), 20 segundos y los modelos, tanto de población de referencia como los de locutor, provenientes de Ahumada IV (SITEL), 120 segundos.

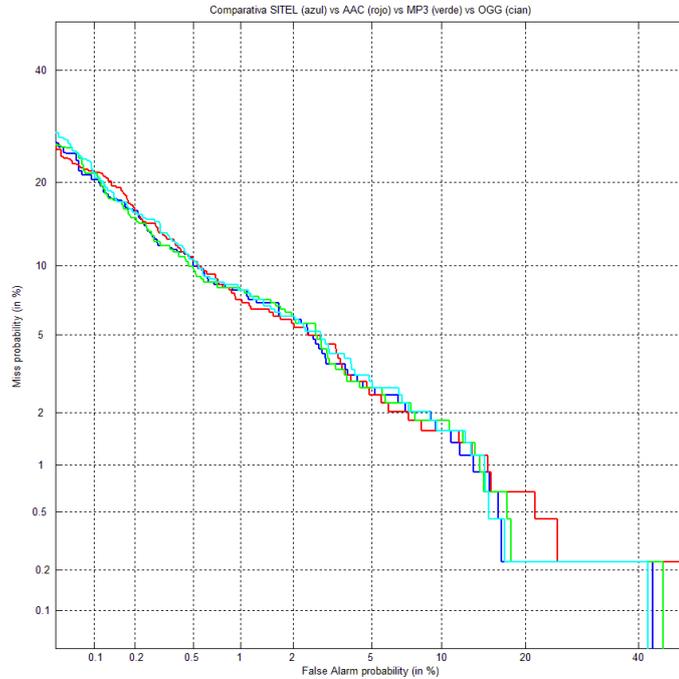
La prueba es el resultado del enfrentamiento de 91 modelos contra 442 archivos de test, de los cuales 442 enfrentamientos son target y 39780 son non-target.

Se seleccionan los 35 mejores competidores de 91 en la población de referencia.

### 3.8.1. Resultados y curvas comparativas

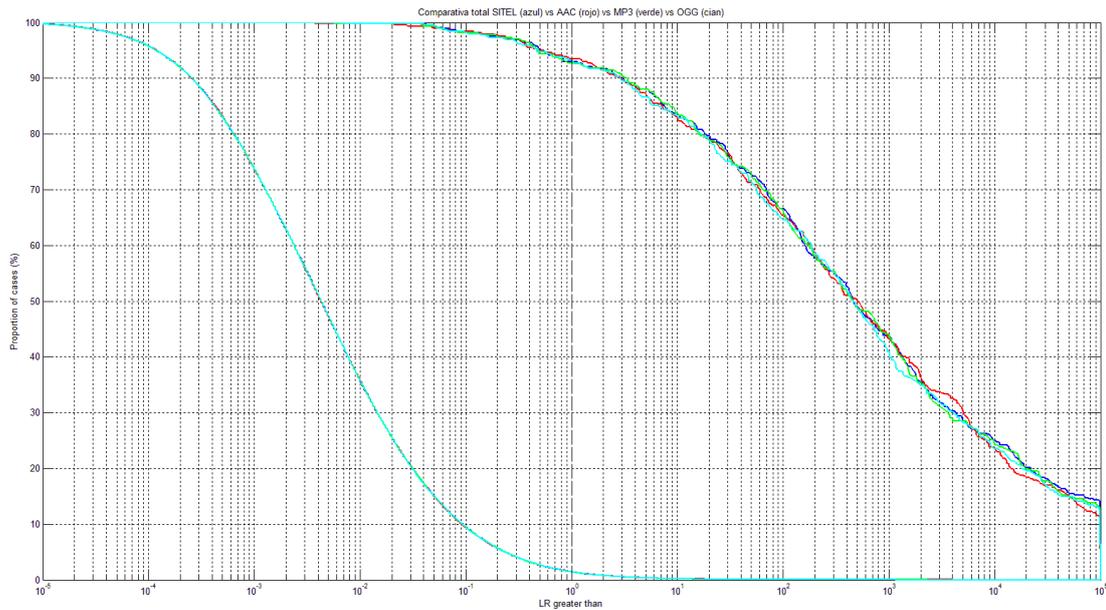
#### Curva DET

En este caso el parámetro de EER son los valores especificados para cada formato de audio.



#### Curva Tippett

Refleja unas tasas de evidencia errónea de un 1.4% que apoya la hipótesis del fiscal y próximas al 7%, 6.6%, 7.2% y 6.8% que apoya la hipótesis del defensor, evaluadas para valores de LR<sub>s</sub> >1, para las distribuciones correspondientes a los formatos SITEL, AAC, MP3, y OGG, respectivamente.



#### 4. CONCLUSIONES:

Trasladando los valores obtenidos de cada una de las gráficas, para una visión más genérica, intuitiva y resumida, tenemos:

**TABLA COMPARATIVA CON VALORES CARACTERÍSTICOS**

		<b>CASO 1 (SITES)</b>	<b>CASO 2 (AAC)</b>	<b>CASO 3 (MP3)</b>	<b>CASO 4 (OGG)</b>
<b>DET</b>	<b>EER</b>	3.72	3.61	3.49	3.94
<b>TIPPET</b>	<b>F<sub>A</sub></b>	1.40%	1.40%	1.40%	1.40%
	<b>F<sub>R</sub></b>	7.00%	6.60%	7.20%	6.80%
<b>C<sub>llr</sub></b>	<b>C<sub>llr</sub> CAL</b>	0.146	0.149	0.149	0.151
	<b>C<sub>llr</sub> DIS</b>	0.121	0.126	0.122	0.126

En el enfrentamiento alternado de los diferentes tipos de archivos de voz el sistema ofrece unos valores de rendimiento y calibración sistema cuyas diferencias son apenas. Si bien, se estima un mejor comportamiento en EER, de los archivos MP3, frente a una leve diferencia para este parámetro en los correspondientes a SITES.

En lo referente a la Falsa aceptación, no hay distinción a la hora de elección del formato del archivo de audio pero el formato AAC se comporta ligeramente mejor que los demás a la hora de discriminar el Falso rechazo.

Por último, el sistema presenta una mejor calibración con el formato de SITEL frente al OGG, siendo la diferencia poco notable. Asimismo, el sistema presenta una mejor discriminación con el formato de archivo SITEL frente a los AAC y MP3.

No obstante lo anterior, se puede afirmar que disponiendo de una población coincidente en canal para cada tipo de modelo, aunque en formatos diferentes, los resultados obtenidos serán muy similares y cuyas diferencias apenas pueden ser ponderables, como era de esperar.

No se observa que la modificación del canal de los test se traduzca en un cambio en el rendimiento del sistema, en relación al obtenido en el caso de referencia (SITEL). Ello implica la viabilidad de empleo de cualquiera de los formatos de archivo de audio sin que suponga aportar diferencias apreciables en el rendimiento.

## **5. TRABAJO DE FUTURO**

En relación a una posible ampliación sobre este estudio, se estima que pudiera ser productivo y por lo tanto ampliar el estudio al uso de otras variables, manteniendo el canal de test y procediendo a realizar las modificaciones en el modelo.

Asimismo, se indica como factible y de gran interés llevar a cabo el estudio de las correspondencias de canal entre test, modelo y población de referencia.

Finalmente, se propone ver las posibilidades en el caso de manteniendo el canal, variar el contenido acústico de los test.

## 6. REFERENCIAS

- [1] Evett, I.W., «Towards a uniform framework for reporting opinions in forensic science casework,» *Science and Justice*, vol. 38, n° 3, pp. 198-202, 1998.
- [2] Champod, C.; Meuwly, D., «The inference of identity in forensic speaker recognition,» *Speech Communication*, vol. 31, pp. 193-203, 2000.
- [3] J. Gonzalez Rodriguez, D. Garcia Romero, M. Garcia Gomar, D. Ramos Castro y J. Ortega Garcia, «On Robut Estimation of Likelihood Ratios: The ATVS-UPM System at 2003 NFI/TNO Forensic Evaluation,» de *The Speaker and Language Recognition Workshop*, Toledo, 2004.
- [4] F. Taroni y C. Aitken, «Forensic Science at Trial,» *Jurimetrics Journal* 37, pp. 327 - 337, 1997.
- [5] J. Ortega Garcia y J. Gonzalez Rodriguez, «Procesado de voz: Análisis y Parametrización de señal,» *Departamento de Ingeniería Audiovisual y Comunicaciones*, 1997.
- [6] J. Gonzalez Rodriguez y J. Ortega Garcia, «Procesado de Voz: Reconocimiento de mensaje y locutor,» *Departamento de Ingeniería Audiovisual y Comunicaciones de EUIT de Telecomunicación, Universidad Politécnica de Madrid*, 1997.
- [7] J. Gonzalez Rodríguez, J. Fierrez Aguilar y J. Ortega García, «Forensic Identification Reportin using Automatic Speaker Recognition Systems,» de *IEEE International Conference on Acoustics, Speech and Signal Precessing*, Hong Kong (China), 2003.
- [8] C. Champod, «Overview and Meaning of Identification,» de *Encyclopedia of Forensic Sciences*, Academic Press, 2000, pp. 1077-1084.
- [9] C. Champod y D. Meuwly, «The Inference of Identity in Forensic Speaker Recognition,» *Speech Communication*, vol. 31, pp. 193-203, 2000.
- [10] C. Champod y D. Meuwly, «The Inference of Identity in Forensic Speaker Recognition,» *Speech Communication*, vol 31, pp. 193 -203, 2000.
- [11] F. Taroni, S. Bozza y C. Aitken, «Decision analisis in forensic science,» *Journal of Forensic Sciences*, vol 50, p. 2005, 894 - 905.
- [12] F. Taroni, C. Aitken y P. Garbolino, «De Finetti's subjectivism, the assessment of probabilities and the evaluation of evidence,» *Science and Justice*, vol 41, pp. 145-150, 2001.

- [13] O. Ribaux, S. J. Walsh y P. Margot, «The contribution of forensic science to crime analysis and investigation: Forensic Intelligence,» *Forensic Science International*, vol. 153, nº 2-3, pp. 171-181, 2006.
- [14] A. Alexander, F. Botti y A. Drygailo, «Handling Mismatch in Corpus-Based Forensic Speaker Recognition,» de *The Speaker and Language Recognition Workshop*, Toledo, pp. 63-68, 2004.
- [15] F. Botti, A. Alexander y A. Drygailo, «An Interpretation Framework for the Evaluation of Evidence in Forensic Automatic Speaker Recognition with Limited Suspect Data,» de *The Speaker and Language Recognition Workshop*, Toledo, pp. 63-68, in Proceedings of Odyssey 2004.
- [16] D. Reynolds, T. Quatieri y R. Dunn, «Speaker Verification Using Adapted Gaussian Mixture Model,» *Digital Signal Processing*, vol. 10, pp. 19-41, 2000.
- [17] N. Brümmer y J. d. Preez, «Application-Independent Evaluation of Speaker Detection,» *Computerr Speech and Language*, vol. 20, nº 2-3, pp. 230-275, 2006.
- [18] . B. Gonzalez Sigüenza, «Batvox: Sistema automático de reconocimiento de locutor,» ISSN 1575-5533, 2008.



**INSTITUTO UNIVERSITARIO DE INVESTIGACIÓN EN CIENCIAS POLICIALES**

**Facultad de Derecho. Universidad de Alcalá**

**Libreros, 27, planta baja. 28801 Alcalá de Henares (Madrid)**

**<https://iuicp.uah.es>**