UNIVERSIDAD CARLOS III DE MADRID

ESCUELA POLITÉCNICA SUPERIOR

DEPARTAMENTO DE INFORMÁTICA

Doctorado en Ciencia y Tecnología Informática

TESIS DOCTORAL

Influencia de los segmentos del discurso en la discriminación del locutor.

Autor:

Luis Antonio Puente Rodríguez

Directores:

Belén Ruíz Mezcua

Ángel García Crespo

Año 2013

"Dessine-moi un mouton."

Antoine de Saint-Exupéry

"… y una escuadra es un triángulo."

Jose Palop Gómez

A todos y todas:

A los que habéis estado ahí ayudándome a ponerme en pie tras cada caída, a los que habéis tenido una palabra de aliento en cada momento desánimo, a los que habéis prodigado paciencia en cada arrebato, a los que me habéis predicado templanza ante la imprudencia. A los que sí y a los que no. A los que habéis puesto piedras en mi camino, a los que habéis mirado hacia otro lado cuando he gritado "ayuda", a los habéis huido cuando os llamaba, a los que no llamé pero lamentablemente vinisteis.

A todos, sí, a todos, porque gracias a todos soy lo que soy.

Índice

Índice de figuras

Índice de tablas

Resumen

La autenticación de la identidad de las personas es hoy en día una tarea crucial, ya que una amplia variedad de sistemas precisan de un método fiable, bien para determinar o bien para confirmar la identidad de los individuos.

Entre los métodos de autenticación, el "reconocimiento biométrico" ha recibido una considerable atención en los últimos años debido principalmente a dos motivos: el fuerte crecimiento de la demanda de aplicaciones de seguridad, tanto comerciales como militares y el rápido desarrollo de la tecnología que las soporta. Su finalidad es la determinación de la identidad de las personas basándose en uno o más rasgos físicos o de comportamiento, elementos, que a diferencia de los utilizados por otras técnicas, siempre acompañan al individuo.

En este área, la utilización de la voz humana como rasgo presenta un conjunto de características que la hacen especialmente practica y la convierten en la mejor opción, cuando no la única, en un amplio conjunto de aplicaciones.

El esquema general del proceso de reconocimiento define dos grandes etapas: la extracción de la información relevante de las muestras de voz capturadas, y la comparación de dicha información con otra de las mismas características previamente almacenada; comparación, esta última, para lo cual se suele hacer uso de técnicas de clasificación provenientes del área de la inteligencia artificial.

Dado el estado actual de los algoritmos de clasificación, parece difícil pensar que los sistemas de reconocimiento biométrico puedan mejorar sustancialmente sus tasas a partir de la mejora de los mismos; es necesario, por tanto mejorar la calidad de la información que se les suministra.

En este trabajo, el autor presenta un nuevo enfoque que permite la mejora de las tasas del reconocimiento del locutor mediante la selección de la dicha información, proponiendo, asimismo, un sencillo algoritmo que realiza este filtrado. Sus resultados no sólo son aplicables al diseño de nuevos sistemas, sino que resultan útiles a la hora de mejorar las prestaciones de los que se encuentran en funcionamiento.

Abstract

The authentication of people identity is nowadays a crucial task, since a wide variety of systems requires a reliable method either to determine or to confirm the identity of individuals.

Among all the authentication methods, the "biometric recognition" has received considerable attention in the recent years mainly due to two reasons: the strong growth in demand for security applications them, both commercial and military, and the rapid development of technology supporting it. Its purpose is to determine the identity of the person based on one or more physical or behavioural traits, elements that unlike those used by other techniques, always go with the individual.

In this area, the use of the human voice as a trait has a set of characteristics that make it especially practical and it becomes the best choice, if not the only available one, for a wide range of applications.

The general scheme of the recognition process is defined in two mayor stages: extracting the relevant information from the captured voice samples, and matching that information to another one previously stored of the same trait; matching, the latter, for which usually makes use of classification techniques inherit from the artificial intelligence area.

Considering the current state of classification algorithms, it seems hard to believe that biometric recognition systems can substantially improve their rates just by improving them, it is therefore necessary to pay attention to improve the quality of information supplied.

In this document, the author presents a new approach which allows the improvement of speaker recognition rates by the selection of such information, proposing, likewise, a simple algorithm that performs this filtering. Their results are not only applicable to the design of new systems, but also are useful in improving the performance of those which are in operation.

1 Introducción

En el contexto actual de inseguridad creado por el final de la bipolaridad internacional, la globalización, la diseminación de la violencia y los ataques del 11 de septiembre, identificar se ha convertido en una necesidad, lo que ha intensificado de forma dramática el interés en los medios y tecnologías de la identificación; ya que, conocer quién es quién con certeza ha pasado a ser una importante misión de las fuerzas del orden. Actualmente más y más gobiernos adoptan nuevas tecnologías de identificación para la vigilancia y la monitorización de los movimientos de las personas en y dentro de sus fronteras (Ceyhan, 2008) (Adler, 2003).

Así, el reconocimiento de las personas representa, hoy en día, una tarea vital en la protección de espacios, sean estos físicos o virtuales. En las últimas décadas, múltiples grupos de investigación han trabajado en el desarrollo de sistemas que puedan realizar una identificación (positiva o negativa) de los individuos bajo control, basándose en su físico, en su psicología o en su comportamiento, elementos que han venido a denominarse "Rasgos Biométricos" (Rubio-Ayuso and Hernández-Rioja, 2005).

Una amplia variedad de sistemas precisan de un método de reconocimiento fiable de las personas bien para determinar o bien para confirmar la identidad de los individuos que solicitan un servicio. El propósito de estos métodos es asegurar que los servicios provistos son únicamente accesibles a usuarios legítimos y a nadie más, servicios que pueden incluir el acceso a espacios restringidos, ya sean físicos (léase edificios, salas,…) o espacios virtuales (zonas para servicios específicos) o incluso equipamiento (ordenadores, laptops, móviles,…). Sin un método robusto tales espacios quedarían a merced de impostores que harían un uso indeseado y mal intencionado de los mismos (Jain et al., 2004b).

1.1 Técnicas de reconocimiento

Tradicionalmente la confianza en la identificación automática de las personas ha venido representada por dos alternativas "el conocimiento" o "la posesión".

El primero de ellos se resume en la frase "Soy quién soy porque sé algo que sólo yo sé", concepto que se ha implementado habitualmente en forma de claves aunque existen otras variantes. Un ejemplo lo encontramos en la frontera alemana, en virtud de los acuerdos establecidos con las excolonias, se precisaba determinar en la frontera la nacionalidad de inmigrantes que estaban desprovistos de identificación. La solución adoptada fue requerir de ellos información sobre su país de origen

(presidente, capital, etc.) para confirmar su procedencia. Barreras basadas en esta idea pueden ser violentadas con relativa facilidad, ya que la información puede ser filtrada hacia una persona no autorizada, deducida cuando su complejidad es baja, u olvidada cuando es alta. En este último caso existe una situación que lo agrava ya que el usuario tiende mantenerla anotada, anulando, en consecuencia, la protección que aportaba su complejidad (Gómez et al., 2007).

El segundo concepto ("Posesión") se resumen en la frase: "Soy quién soy porque poseo algo que sólo yo poseo", que en la actualidad viene implementada mediante pasaportes, tarjetas de identificación de banda magnética o de chip, pero que en la literatura romántica y de aventuras hemos visto representado por "medios medallones", salvo conductos, cartas de presentación, sellos reales, espadas, joyas, etc. Estas barreras pueden ser también fácilmente rebasadas al poder ser estas posesiones sujetos de sustracción o pérdida, en aquellos casos en los que no sea posible la duplicación(Puente et al., 2008b).

Alternativamente a "soy lo que sé" o "soy lo que tengo", los modelos de reconocimiento biométrico se basan en el concepto de "soy lo que soy" (o más apropiadamente "soy lo que parezco ser"). Se utilizan rasgos personales e intransferibles que acompañan siempre a la persona y son relativamente difíciles de falsificar. Con ello la biometría pretende facilitar y robustecer el proceso de reconocimiento de los individuos.

Lo facilitan, porque los rasgos biométricos básicamente no son susceptibles de hurto ni de olvido, van con la persona donde quiera que ella vaya (aunque el sujeto no quiera), con lo que casi siempre están disponibles. Lo robustecen ya que por su propia naturaleza, los rasgos biométricos tienen, en general, el carácter de información pública, lo que a la larga supone en sí mismo una fortaleza, ya que la presunción de que una información secreta no es conocida o el objeto no es poseído por terceros es la principal debilidad de los sistemas de seguridad tradicionales (Ruiz, et al., 2008).

Múltiples tipos de rasgos biométricos son objeto de estudio por parte de los distintos grupos de investigación(Rubio-Ayuso and Hernández-Rioja, 2005); algunos popularizados por sus espectaculares apariciones en productos cinematográficos como las huellas digitales, la retina, la geometría de la cara, el ADN, la firma manuscrita; otros no tan popularizados pero bien conocidos como la voz, el iris, la caligrafía, la estructura del pabellón auditivo, o muy poco populares como el sistema vascular de la palma de la mano, la geometría de la mano, la huella de la palma, la forma de caminar, el ritmo de tecleo.

Prácticamente cualquier aspecto del cuerpo o del comportamiento humano que sea registrable puede convertirse en un rasgo biométrico con mejores o peores

resultados. En (Singh et al., 2012b) y artículos anteriores se afirma que señales eléctricas como el electrocardiograma o el electroencefalograma presentan características que permiten distinguir entre individuos, por lo que pueden ser propuestas para su utilización en procesos de reconocimiento, y todavía está por ver que alguien pueda simular el comportamiento neuronal o cardíaco.

Todas estas modalidades ofrecen coyunturalmente ventajas y desventajas. Mientras que rasgos como Iris, Retina, Sistema Vascular de la mano y otros ofrecen una confianza muy alta en el reconocimiento, la captura de la muestra biométrica exige, en la práctica, la colaboración del donante y en muchos casos unas estrictas directrices en el protocolo de su toma. En el punto opuesto rasgos como el modo de caminar o los patrones de tecleo, son relativamente fáciles de capturar, pero el grado de confianza que ofrecen es sensiblemente menor.

A tenor de esta coyuntura se están poniendo de moda los sistemas multimodales en los que se opera simultáneamente con varios rasgos, así por un lado se previene la ausencia de muestras de calidad de alguno de ellos, y por otra se refuerza la confianza en la decisión tomada sobre la identidad del individuo a partir de todas las modalidades disponibles.

1.2 La voz como rasgo biométrico

En el caso particular de la voz existen tres circunstancias que la convierten en una modalidad biométrica de mucha utilidad. En primer lugar, se trata de una señal que se produce de forma natural y el procedimiento de su captura no resulta incómodo para el usuario, que no considera la donación de una muestra como un proceso intrusivo ni separado de su actividad habitual (Ruiz-Mezcua, 1998). En segundo lugar, en muchas aplicaciones la voz puede ser la principal, a veces la única, modalidad disponible. Por último, la proliferación de las redes celulares ha convertido al teléfono móvil en un sensor familiar y cuya utilización simplifica notablemente la operativa de captura y transmisión de una muestra de voz. En aplicaciones basadas en la telefonía, no hay necesidad de transductores de señal especiales; y para las aplicaciones no telefónicas, tanto las tarjetas de sonido como los micrófonos son de bajo coste y de alta disponibilidad. Debido a la ubiquidad de la red telefónica y a la capacidad actual de los ordenadores, el coste de un sistema de reconocimiento de locutor queda vinculado tan sólo al desarrollo del software (Bimbot et al., 2004).

Adicionalmente, el área del procesado del habla tiene una larga y rica tradición de investigación, desarrollo y evaluación. En los últimos tiempos su tecnología ha hecho aparición en múltiples productos comerciales. Por ello es probablemente el

método más natural y económico para resolver los problemas del uso no autorizado de ordenadores y de sistemas de comunicación (Campbell Jr, 1997).

Los sistemas de reconocimiento de locutor más exitosos en resultados y que han tenido un mayor número de implementaciones prácticas, utilizan el análisis espectral continuo de la señal vocal, en concreto la caracterización de la voz mediante los coeficientes "Mel-Cepstrum" y sus derivados (Ganchev, 2011), que tienen la habilidad de analizar de forma separada el tracto vocal y la actividad de las cuerdas vocales.

El conjunto de los coeficientes obtenidos a partir de un pequeño fragmento de audio se denomina vector de características. El conjunto de vectores de características obtenidos de los múltiples fragmentos de una o de un conjunto de locuciones representativas, define una región en un espacio multidimensional donde vectores de futuros fragmentos del mismo locutor tienen alta probabilidad de localizarse. Esta región suele resultar definida con mayor precisión cuando se obtiene a partir de la comparación entre la voz del locutor y la voz del resto del mundo, o de forma más práctica de un conjunto de personas suficientemente representativo del resto del mundo.

Establecidas esas regiones, se construye una expresión matemática que calcule cuando un vector está dentro o fuera de ellas; expresión que constituye el denominado "modelo del locutor". En la práctica este modelo corresponde a los valores de los parámetros de una ecuación definida para el tipo de reconocedor que utilizado.

En este contexto, verificar la identidad de un locutor se resume en determinar si un porcentaje suficiente de vectores de características de una nueva locución se encuentra dentro de la región definida para el usuario, o dentro de la definida para el mundo, lo que corresponde a la funcionalidad de un clasificador binario.

Este tipo de técnicas basadas en el análisis espectral de la voz, obvian la estructura de la comunicación y realizan su función con independencia de que la emisión corresponda o no a un mensaje que busque transmitir una información. Pero también existen en la literatura otras aproximaciones en las se considera la existencia propia de un contenido en el mensaje. Históricamente se ha centrado la atención en caracterizar el espectro analizando fragmentos muy pequeños (virtualmente instantes) dejando a un lado la caracterización de la evolución temporal de la voz, cuando probablemente el oído humano basa su capacidad de reconocimiento en patrones de larga duración (sílabas, palabras o frases), y en las características asociadas a estos patrones (entonación, énfasis, ritmo,...). Con toda seguridad este tipo de diferencias idolectales definen de forma más precisa al locutor. [Doddington 2001].

En éste documento se resume el trabajo de investigación realizado para aportar una técnica más refinada en el reconocimiento del locutor, basada en la asunción que no toda la información contenida en el discurso humano tiene la misma capacidad de discriminación de las personas.

1.3 Trabajos previos

Las bases de la presente tesis se desarrollaron durante la participación del doctorando en los proyectos de investigación PIBES y SEGUR@.

PIBES fue un proyecto competitivo del Plan Nacional, financiado por el Ministerio de Ciencia y Tecnología, con el título "Perfeccionamiento de la Identificación Biométrica y Evaluación de su Seguridad (PIBES)" (Referencia: TEC2006-12365-C02-01). Con una duración de 36 meses (del 01/10/2006 al 30/09/2009) en el que participaron la Universidad Carlos III de Madrid y la Universidad Politécnica de Cataluña.

"SEGUR@: Seguridad y Confianza en la Sociedad de la Información", fue un proyecto CENIT en el que participaron SECUWARE, ATOS ORIGIN SAE, ERICSSON, S21SEC, ISDEFE, SAFELAYER, Universidad Politécnica de Madrid, ALCATEL, CSIC, y Telefónica ID, S.A., y donde el equipo de investigación de la Universidad Carlos III de Madrid fue contratado para el desarrollo de los temas de biometría y seguridad biométrica. Tuvo una duración de 42 meses (del 15/12/2007 al 15/12/2010)

En estos trabajos se llevaron a cabo las evaluaciones del estado del arte, la propuesta de algoritmos y la implantación de soluciones adecuadas al entorno de explotación elegido, en las áreas del reconocimiento del locutor y de sistemas polibiométricos, en concreto los de fusión multimodal. Los resultados de estos trabajos fueron presentados entre otros en los siguientes artículos:

> "Biometric Authentication Devices and Semantic Web Services: An Approach for Multi Modal Fusion Framework". L.Puente, J.M.Gómez, B.Ruiz, M.J.Poza BIODEVICES 2008: International Conference on Biomedical Electronics and Devices, (Madeira, Portugal, Jan-2008).
>
> "Identity Authentication Services". J. Poza, L. Puente, B. Ruiz and J.M. Gómez Proceedings of the 2008 International Conference on Security and Management (SAM'08) |pp 655-659, (Las Vegas, Nevada, USA Jul. 2008), ISBN #: 1-60132-085-X.

"Study of Different Fusion Techniques for Multimodal Biometric Authentication". L. Puente Rodríguez , A. García Crespo, M. J. Poza Lara .B. Ruiz Mezcua. 4th IEEE International Conference on Wireless & Mobile Computing, Networking & Communication (WiMob 2008). pag. 666-671 (Avignon. France, Oct-2008).

SVM Speaker Verification System Based on a Low-Cost FPGA. Rafael Ramos, Mariano López, Enrique Cantó and Luis Puente-Rodríguez, 19th International Conference on Field Programmable Logic and Applications (FPL?09), Prague, Czech Republic, pp. 582-586, June 2009. ISBN: 978-1-4244-3892-1.

1.4 Motivación

En los trabajos previos descritos en la sección anterior se pudo comprobar que las tasas de error se encuentran en la horquilla del 5 al 12% para un entorno ambiente controlado y alcanzan el 25% cuando éste se degrada (Sanchez, 2010) (Sanchez, 2009) (Puente et al., 2008a), apreciación que es refrendada por la literatura (Stolcke et al., 2007). Estas tasas parecen apuntar a que es necesario seguir trabajando en las prestaciones de estos sistemas.

No parece posible obtener grandes mejoras de los resultados de los reconocedores de locutor en base a la utilización de clasificadores más perfeccionados, sin previamente lograr una identificación más precisa de los mecanismos fisiológicos de la fonación, la determinación de cómo estos quedan representados en la locución y la obtención una algorítmica eficaz con la que extraer esa información de la señal portadora de la voz.

En los últimos años la utilización de las herramientas de reconocimiento de habla en la investigación del reconocimiento del locutor ha recibido un gran impulso. En (Stolcke et al., 2007) se señala que el conocimiento del contenido del discurso es necesario para mejorar el proceso de extracción de características y por tanto, que la salida ofrecida por un reconocedor del habla puede ser utilizada a dicho fin. La información generada por este tipo de reconocedor puede ser utilizada de múltiples formas. Obviamente esto incluye reconocimiento de palabras para permitir el uso de técnicas de creación de modelos del locutor dependientes del texto cuando el discurso no está prefijado (Stolcke et al., 2007), la extracción de características supra-segmentales y la eliminación de los elementos que no aportan información sobre el locutor.

La aportación principal del trabajo, que hasta el momento no se ha visto planteada en ninguna publicación, es la determinación de aquellos segmentos más determinantes en el reconocimiento y cuál es la metodología adecuada para su selección durante el proceso de entrenamiento en el supuesto de que su selección sea dependiente del locutor y/o de la aplicación objetivo.

1.5 Hipótesis

Con este estudio se pretende verificar que es posible plantear mejoras en los sistemas de reconocimiento del habla actuales, sacando partido al hecho de que la capacidad discriminadora de la voz no es uniforme a lo largo de todo el discurso, característica que es expuesta en la hipótesis principal de la presente tesis:

Hipótesis

> **"En el discurso existen segmentos que son más característicos del locutor que otros que no lo son tanto".**

Siendo cierta esta afirmación el trabajo quedaría incompleto si no ofrecieran caminos para sacar partido a esta propiedad y por ello se han planteado las siguientes hipótesis secundarias:

Hipótesis Secundaria 1

> **"Los segmentos del discurso no presentan prestaciones uniformes en la caracterización del locutor".**

Hipótesis Secundaria 2

> **"Es posible determinar qué segmentos del discurso ofrecen mejores prestaciones".**

Hipótesis Secundaria 3

> **"Es posible diseñar un modelo de procesamiento del habla para el reconocimiento del locutor que pueda sacar provecho de la identificación de los segmentos característicos".**

1.6 Estructura del documento

La presente memoria se estructura de la siguiente forma: En la sección 2 se realiza la revisión del estado del arte presentando los conceptos básicos de la

biometría, de los mecanismos de la producción de la voz, de la percepción de los sonidos y finalmente de las tecnologías al uso en el área. En la sección 3 se realiza la presentación del problema y la propuesta de solución. En la 4 se definen los criterios metodológicos definidos para la investigación asociada a la tesis. En la sección 5 se revisan la evolución de la investigación a través del desarrollo de la experimentación y los resultados obtenidos en ellas. En la sección 6 se presentan las conclusiones de la investigación realizada. Finalmente la sección 6.3 plantea los trabajos que se pueden desarrollar como continuación de éste.

2 Estado del arte

En esta sección se presentará el estado del arte de las áreas de conocimiento necesarias para abordar el problema de la eficacia de la identificación biométrica en toda su magnitud.

2.1 Biometría

2.1.1 Historia

La palabra biometría proviene de los términos griegos *bios* que significa vida y *metron* que significa medida, lo que se traduce por *medida de las cosas vivas*. Por extensión se ha convertido en la caracterización de los rasgos de los seres vivos, y por el uso dentro del área específica del reconocimiento de las personas (Tapiador and Singüenza, 2005), en la utilización de estas caracterizaciones para distinguirlas a pesar de las similitudes entre ellas y de las variaciones que pudieran aparecer a lo largo del tiempo aún perteneciendo a un único individuo (Wayman et al., 2004).

Desde que el hombre es hombre distingue entre las personas próximas (conocidas) de aquellas que no lo son. La cualidad de animal social que nos define, exigió de la evolución la capacidad identificar mediante los sentidos a los miembros del propio clan, clan que le que proveía de seguridad (Poza et al., 2008a).

La historia de la humanidad recoge como las sociedades han venido utilizando rasgos tales como la cara o la voz para el reconocimiento mutuo. Ya Quintilio (35-90 D.C.) apostillaba que *"la voz de un locutor es tan fácilmente distinguible por el oído como la cara lo es por la vista"*.

La biometría se inició como ciencia con Alphonse Bertillon (1883-1914), jefe de la división criminal de la policía de París, quien fue el primero en utilizar el una sistemática en la identificación de las personas que se basaba en tres grupos de características físicas (Rhodes, 1956) (Dantcheva et al., 2011).

- Medidas antropométricas (altura, longitud del brazo forma y tamaño de la cabeza, peso, altura…)
- Descripción morfológica (color de los ojos, anomalías en los dedos,…).
- Marcas peculiares observables en el cuerpo (marcas, lunares, cicatrices, tatuajes…)

Aunque este sistema era útil en la persecución de criminales, tenía una elevado porcentaje de error ya que las características elegidas no poseían grandes propiedades diferenciadoras (varias personas podían presentar el mismo conjunto de medidas) y tampoco eran permanentes (para el mismo individuo el conjunto de medidas podía cambiar con el tiempo)(Jain et al., 2004a)

Esté avance en el concepto de la criminalística fue perfeccionado con el descubrimiento, a finales del mismo siglo, de la capacidad de discriminación que poseen las huellas digitales. A partir de ahí los departamentos legales abordaron la labor de recopilar impresiones de las de los criminales en una amplia base de datos. Es aquí cuando la biometría se transforma en la ciencia que permite sistematizar el reconocimiento de los individuos a partir de sus características personales (Poza et al., 2008b).

Finalmente, en la década de los sesenta, el desarrollo del proceso digital de la señal llevo inmediatamente a utilizarlo en la identificación de las personas. La voz y la huella digital fueron las primeras modalidades en ser estudiadas. Los setenta vieron la aplicación a la geometría de la mano, la utilización de grandes volúmenes de datos y el interés de los gobiernos por las aplicaciones potenciales de estas tecnologías; los ochenta popularizaron el estudio de los rasgos faciales mientras los noventa añadieron a este conjunto el reconocimiento de los patrones de iris (Wayman et al., 2004).

Desde Bertillon la tecnología biométrica ha mejorado sus técnicas de reconocimiento aportando nuevos rasgos como la cara, las huellas digitales, la geometría de la mano, el iris, la retina, la firma manuscrita, la voz, todas ellas bien conocidas pero también otras menos populares como la forma del pabellón auditivo, la forma de caminar, el olor corporal, electrocardiograma, el electroencefalograma (Ruiz, et al., 2008).

2.1.2 Biometría y seguridad

Es costumbre asumir que el impulso actual a las tecnologías de la seguridad es consecuencia de los hechos ocurridos el 11 de septiembre, pero realmente encuentran sus raíces a comienzos de los años ochenta, época en que los Estados Unidos iniciaron la recuperación de los dispositivos utilizados en la Guerra de Vietnam y se produce su redespliegue en la frontera mexicana, con el fin de interceptar a los traficantes durante la “Guerra de las drogas” (Poza et al., 2008c). El desarrollo continuó en los noventa con el agudizamiento de los problemas de inmigración que llevo al endurecimiento de los controles fronterizos. Bajo estas primeras directrices, la biometría fue introducida como técnica para la identificación de inmigrantes indocumentados que trababan de entrar en EE.UU. cruzando de la frontera del sur. En

consecuencia biometría, sensores, detectores de movimiento cámaras de altas prestaciones y de visión nocturna fueron instalados para detectar a las personas que cruzaban ilegalmente la frontera (Ceyhan, 2008).

2.1.3 Rasgos biométricos

Básicamente hay tres clases de rasgos biométricos, aquellos vinculados con el aspecto físico, aquellos vinculados a aspectos biológicos y aquellos vinculados a como se muestra la psicología personal (comportamiento). Pero también existen rasgos que se muestran al exterior como el resultado de la combinación de ellos (Puente et al., 2008a).

Desde el punto de vista de proceso se requieren diferentes técnicas según sea el rasgo a utilizar. Algunos requieren que el sistema procese una única señal unidimensional (voz), múltiples señales unidimensionales simultáneas (escritura manuscrita), una única imagen bidimensional (huella digital), una secuencia temporal de imágenes bidimensionales (la forma de caminar)(Wayman et al., 2004) (Puente et al., 2008a) (Carrero et al., 2010).

Para que un rasgo biométrico sea de utilidad práctica debería cumplir cinco requisitos (Wayman et al., 2004):

- Estable: que indica que el rasgo no cambia o cambia poco en una persona a lo largo del tiempo.
- Diferenciador: que sugiere que el rasgo presenta grandes o al menos suficientes variaciones de un individuo a otro.
- Disponible: que presupone que toda la población, o al menos la parte de la población que será observada por el sistema, puede exponer ese rasgo.
- Accesible: que requiere la existencia sensores electrónicos que puedan capturar una muestra de ese rasgo de la forma más sencilla posible.
- Aceptable: que implica que los donantes de muestras biométricas no pongan objeciones a la donación.

Podríamos añadir una más a la lista de Wayman:

- Computable: que indicaría que es preciso que la muestra biométrica pueda ser convertida en valores, que estos puedan ser almacenados y finalmente procesados por un ordenador en un plazo y espacio aceptables.

Toda una rama de esta área de conocimiento ha sido desarrollada para definir medidas cuantitativas de estas características y establecer los criterios de evaluación de los sistemas biométricos.

La estabilidad es medida por el "false non-match rate" o "false rejection rate" o "Error de Tipo I", tres términos que representan el mismo concepto y cuya utilización depende principalmente del uso al que se destina el sistema de reconocimiento. Cuantifica la probabilidad de que una muestra dada no cumpla con el modelo creado para definir al individuo del cual procede (Wayman et al., 2004).

La "diferenciabilidad" se mide mediante el "false match rate" o "false acceptation rate" o "Error de Tipo II", al igual que antes y por la misma razón tres términos para el mismo concepto. Recoge la probabilidad de que una muestra de un rasgo de una persona sea conforme con el modelo creado para otro individuo (Wayman et al., 2004).

Para la disponibilidad se utiliza el "Failure to enrol rate", la probabilidad de que un usuario no pueda proveer una muestra utilizable durante la fase de construcción de su modelo (Wayman et al., 2004)(Reynolds, 2002)(Furui, 1996).

La accesibilidad se valora mediante la determinación de la existencia o no de los sensores adecuados para el rasgo elegido en el modo de funcionamiento establecido. Mientras que no suelen existir problemas para la captura de una muestra en escenarios colaborativos (el sujeto hace donación de la misma voluntariamente), la accesibilidad se ve comprometida en escenarios que no lo son. Disponer de cámaras de vigilancia en la frontera puede resultar poco efectivo para el reconocimiento facial, cuando la cara está oculta tras una barba o los ojos cubiertos por la visera de una gorra.

La aceptabilidad se mide encuestando a los futuros usuarios sobre su disponibilidad a proporcionar la muestra necesaria y la incomodidad o daño que les produce. (Wayman et al., 2004).

Finalmente la complejidad computacional puede medirse utilizando por un lado la cantidad de memoria utilizada, básicamente, la cantidad de memoria necesaria para almacenar todos los modelos de todos los posibles usuarios; por otro, el tiempo de computación para la generación de los modelos, y el tiempo de respuesta del sistema desde que se ofrece el donante hasta que se obtiene una decisión ("throughput rate") (Wayman et al., 2004).

2.1.4 Taxonomía

Las aplicaciones de la biometría son tan numerosas como las tecnologías que se utilizan. Se diseñan aplicaciones que detectan la presencia de personas conocidas, otras se diseñan para emitir una alarma ante la presencia de las que no lo son; unas verifican que la identidad de una persona es la que se presupone, unas comprueban que la persona es quien dice no ser, mientras que otras comprueban si el individuo se encuentra o no registrado en el sistema. Algunas aplicaciones recogen una o varias muestras para compararla con una base de datos de millones de modelos previamente construidos, mientras que otras buscan en una base de datos reducida a unos pocos modelos. Unas comparan una muestra con múltiples modelos, otras comparan una muestra con un modelo y otras comparan múltiples muestras con múltiples modelos(Phillips et al., 2000). Todo multiplicado por las múltiples alternativas de entorno de funcionamiento: exterior o interior, supervisado o no, gente entrenada o no, ambiente controlado o ambiente ruidoso, equipo de captura de altas prestaciones o comercial, etc. Todo lo anterior hace muy complejo establecer criterios de clasificación de estos sistemas que resulten útiles.

Un criterio habitual considera que existen tres tipos de sistemas biométricos: de verificación, de identificación y de búsqueda.

En un sistema de verificación se utiliza una muestra biométrica de una persona la cual, al mismo tiempo, proporciona su identidad de una forma u otra o bien existe una suposición sobre la misma (identidad proclamada). El sistema busca en la base de datos de modelos el correspondiente a esa identidad y lo compara con la muestra; como resultado se obtiene una decisión sobre la veracidad o no de la proclamación (Phillips et al., 2000). Este tipo de aplicaciones se encuentran el acceso a espacios reservados, en los puntos de venta para comprobar la identidad del comprador, etc.

En un sistema de identificación se utiliza una muestra biométrica de una persona sin que sea posible presuponer su identidad. El sistema la compara con una lista de modelos biométricos construidos para definir a sujetos ya registrados; en base a esa comparación se estima la identidad del donante. Aplicaciones de este tipo de sistemas se encuentran en los controles de fronteras o en apoyo de las técnicas forenses y en todos aquellos sistemas en los que no se puede, no se quiere o no es convincente la colaboración del donante (Phillips et al., 2000) (Ruiz-Mezcua, 1998). Conceptualmente una identificación se puede considerar una secuencia de verificaciones sucesivas, en las que la muestra es la misma mientras que el modelo cambia.

En un sistema de búsqueda se conoce a priori la identidad pero se desconoce a la persona en sí. Múltiples muestras de múltiples donantes son comparadas con el

modelo correspondiente hasta el momento en el que coinciden. Su utilización se orienta a sistemas de seguimiento, vigilancia y prevención. Al igual que la identificación una búsqueda puede considerarse una secuencia de verificaciones sucesivas, pero aquí el modelo es único mientras que quién cambia es la muestra.

2.1.5 Técnica biométrica

Las tecnologías biométricas son un conjunto de métodos automatizados de reconocimiento de la identidad de las personas basándose en características fisiológicas o de comportamiento. Aunque el ámbito de la tesis es el reconocimiento individual de las personas, las técnicas biométricas permiten su utilización para la identificación del grupo (social, étnico, …) al que pertenece el individuo (Wayman et al., 2004).

El reconocimiento biométrico habitualmente es realizado por una aplicación software de cierto grado de sofisticación que controla el interfaz de usuario, administra la base de datos e interactúa con el algoritmo de reconocimiento específico (Adler, 2003) (Ruiz et al., 2009). Estos algoritmos suponen la ejecución un proceso compuesto por los siguientes pasos (ver Figura 1 esquema inferior):

- Captura de una muestra biométrica.
- Procesado de la muestra
- Comparación del resultado con uno o varios modelos previamente almacenados
- Obtención del resultado de la comparación que se expresa con un valor denominado puntuación.

Por convenio se asume que a mayor puntuación mayor probabilidad de que el sujeto asociado al modelo (usuario) y el que ha proporcionado la muestra (donante) sean la misma persona.

La necesidad explícita de modelos con los que comparar exige acompañar el proceso descrito antes de otro previo dedicado a la construcción de esos modelos que básicamente consta de las siguientes etapas (ver Figura 1 esquema superior):

- Captura de una o varias muestras biométricas de un único usuario.
- Opcionalmente captura de una o varias muestras de otros individuos (mundo).
- Procesado de las muestras

- Definición de un modelo biométrico adecuado a la caracterización del usuario.

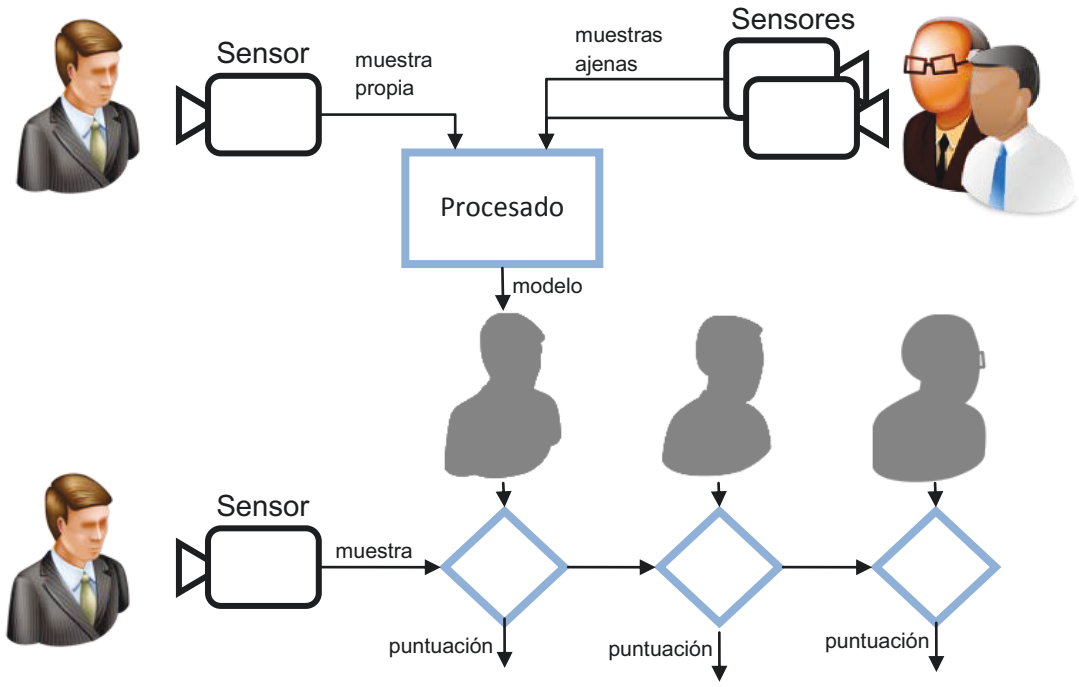


Figura 1: Diagrama de bloques de un sistema de reconocimiento biométrico.

Para distinguir ambos procesos, a este último es denominado entrenamiento, reservando el término reconocimiento para etiquetar propiamente al primero.

Al finalizar el entrenamiento, el modelo obtenido debe describir al usuario, lo que quiere decir que debe definir cómo son sus muestras biométricas en contraposición a cómo son las muestras de otros individuos.

2.1.6 Distancias

Existen dos conceptos básicos que determinan el comportamiento de los sistemas biométricos, y por tanto su diseño.

Uno es la *distancia inter-clases* que representa cuán distintas son las muestras biométricas de dos sujetos diferentes, el otro es la *distancia intra-clase* que representa la diferencia entre muestras distintas del mismo sujeto.

A la hora de diseñar, se buscan rasgos que proporcionen muestras para las cuales la distancia intra-clase sea lo menor posible (que la variabilidad de las

características sea pequeña) y que la distancia inter-clases sea alta (que las diferencias entre sujetos sean apreciables).

Cuanto mayor es la distancia intra-clase, tanto más genéricos serán los modelos que representan a los distintos sujetos. Cuanto menor es la distancia inter-clase mayor es la probabilidad de que una muestra verifique el patrón de un sujeto cercano.

Obtener un equilibrio entre ambas y un umbral discriminador que optimice su funcionamiento es uno de los objetivos de cualquier diseño.

2.1.7 Función de coste

En (Phillips et al., 2000) se apunta que las prestaciones de un sistema biométrico dependen de la población objetivo, de la utilización específica que se quiera dar al sistema y del equipamiento utilizado; lo que impide predecir el comportamiento de un sistema a partir del de otro, siendo además, que las características que definen sus prestaciones están alta y complejamente correladas entre sí. Para un sistema cuyo objetivo primordial sea impedir el acceso, probablemente se estará interesado en reducir el porcentaje de impostores que logran acceder (tasa de falsa aceptación), pero ello será siempre a costa de utilizar un proceso más complejo (reducción de la accesibilidad, de la aceptabilidad y de los requerimientos de computación) y de incrementar el número de veces que un usuario genuino es rechazado (tasas de falso rechazo). Complementariamente si se pretende un sistema de fácil utilización será siempre a costa de incrementar las tasas de fallos.

En un sistema de verificación que compara una única muestra con un único modelo, antes de tomar la decisión (aceptación o rechazo de la identidad proclamada) se obtiene un valor denominado puntuación que el algoritmo que lo genera vincula directamente a la confianza en la decisión. Como se ha dicho, a mayor puntuación mayor probabilidad de que la muestra sea **genuina** (corresponda a la identidad) y a menor puntuación mayor probabilidad de que la muestra sea **impostada** (proporcionada por un sujeto que no es el usuario proclamado).

Estas puntuaciones tienen un comportamiento aleatorio representable por distribuciones gaussianas cuando los rangos de variación de las puntuaciones no están acotados (Figura 2).

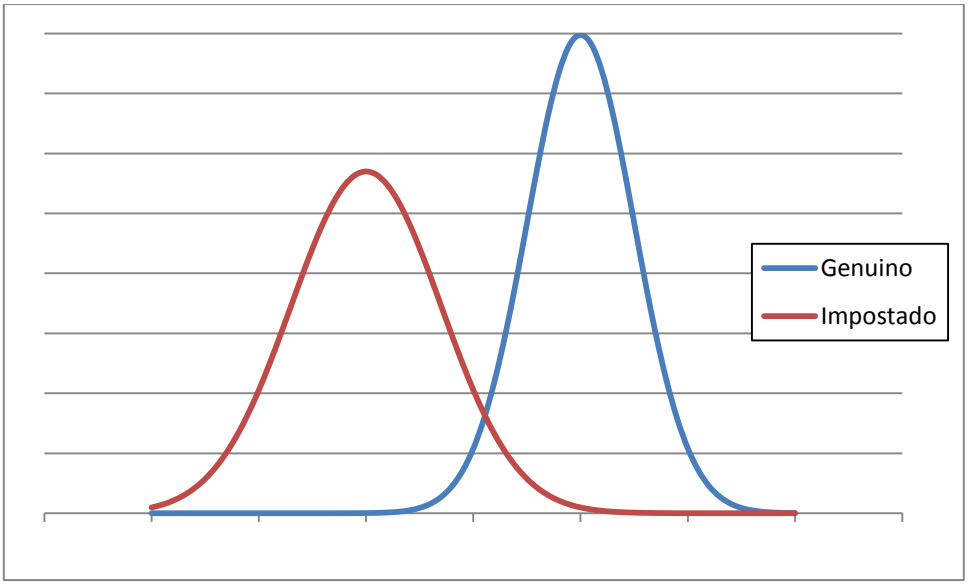


Figura 2: Distribución estadística de puntuaciones gaussianas.

Más comúnmente se obtienen resultados con puntuaciones acotadas, (entre 0 y 1 ó 0% y 100%), donde las puntuaciones de las muestras impostadas se agrupan en las cercanías del mínimo y las genuinas en las cercanías del máximo con comportamientos que pueden ser descritos con funciones de distribución Beta (Puente et al., 2011a).

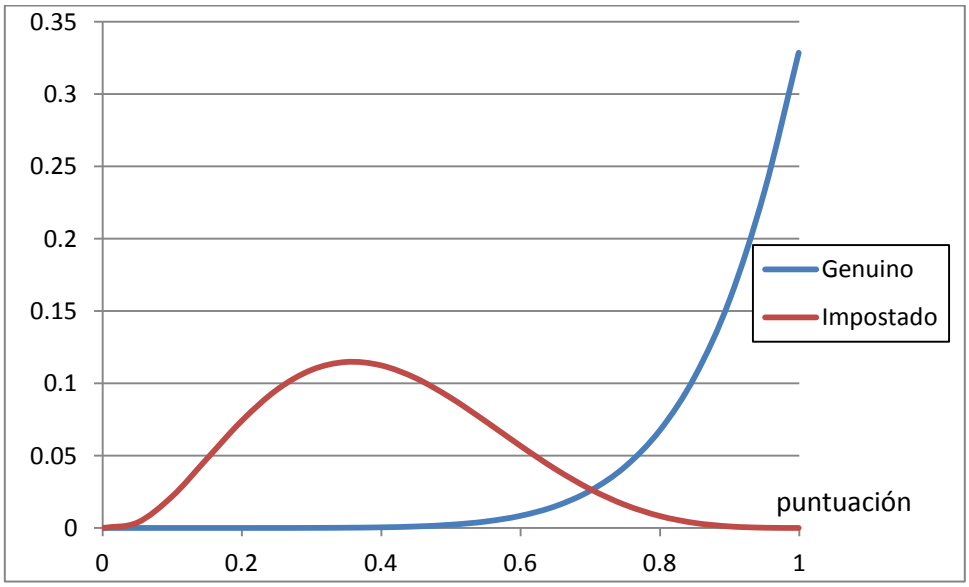


Figura 3: Distribución Beta de las puntuaciones de las muestras.

Para aceptar al donante como genuino, el sistema debe determinar si la puntuación es lo suficientemente alta o no, y por tratarse de una decisión binaria, si no es aceptado como genuino, tendrá que ser rechazado por impostor.

Determinar si la puntuación es lo suficientemente alta implica establecer un umbral de decisión. Si u es ese umbral, f_i la función de densidad de probabilidad de la puntuación de muestras impostadas y f_g la puntuación de las genuinas, la tasa de falso rechazo (FRR) vendrá definida por la Ecuación 1 y la de falsa aceptación por la Ecuación 2, pudiéndose ver su representación gráfica en la Figura 4.

$$FRR(u) = \int_{-\infty}^{u} f_g(p)\ dp = F_g(u) \qquad \text{Ecuación 1}$$

$$FAR(u) = \int_{u}^{\infty} f_i(p)\ dp = 1 - \int_{-\infty}^{u} f_i(p)\ dp = 1 - F_i(u) \qquad \text{Ecuación 2}$$

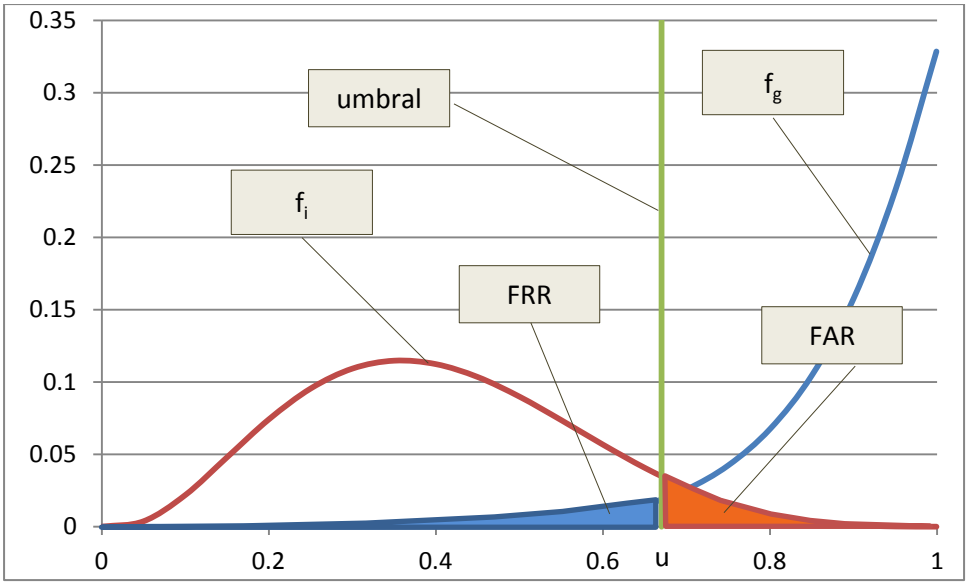


Figura 4: Umbral, falsa aceptación y falso rechazo.

Es posible representar la FAR y la FRR en función del antedicho umbral obteniéndose gráficas como las de la Figura 5.

Si este umbral se reduce es más probable que una muestra impostada pase por genuina (falsa aceptación), y si el umbral se aumenta, aumenta con él la probabilidad de que una muestra genuina sea rechazada (falso rechazo)

En la gráfica de la Figura 5 se puede apreciar la existencia de un punto característico denominado "cross point" que corresponde al lugar donde la falsa aceptación y el falso rechazo toman el mismo valor, valor que se denomina EER (Equal Error Rate) y que se suele utilizar como estimador de la calidad del sistema.

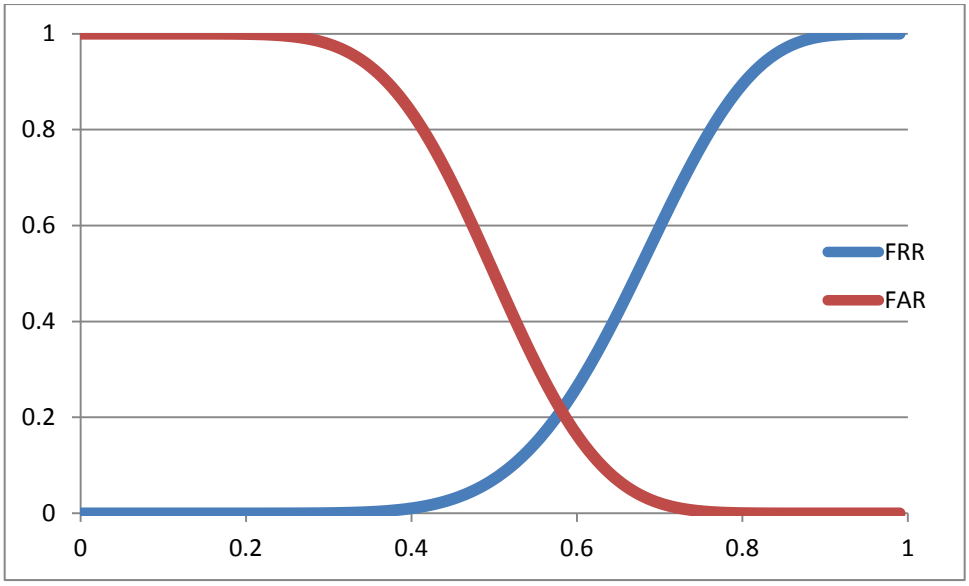


Figura 5: Tasas FAR y FRR en función del umbral.

2.1.8 Curva ROC

En escenarios como el presentado hasta el momento, en el que es necesario tomar una decisión dicotómica a partir de la categorización de ejemplos en una variable continua, es habitual analizar la calidad del procedimiento de decisión a través de la denominada curva ROC (Streiner, 2013).

El término ROC corresponde a las siglas de "Receiver Operating Characteristics" que se utiliza por razones históricas, ya que se concibió para el análisis de la detección de ecos radar sobre un fondo ruidoso, análisis que fue desarrollado durante la II Guerra Mundial en el área de la teoría de detección de la señal (Kay, 1998).

Las curvas ROC reflejan cuantitativamente aspectos sobre la calidad del conjunto de datos en análisis. Estas figuras representan en una gráfica la relación existente entre la tasa de aceptaciones correctas (*CAR = 1-FRR*) y el de las falsas aceptaciones (*FAR*), dado que ambas caracterizan la confianza en las decisiones tomadas con los datos experimentales (Maswadeh and Snyder, 2012).

Un parámetro importante de la curva ROC es el del área bajo ella (AUC, area under the curve). El AUC es una medida de la separación entre los dos grupos de datos a distinguir, lo que implica una medida de lo buena que resulta la variable para clasificarlos y en consecuencia cómo de robusto es el procedimiento ante variaciones en la estimación del umbral de decisión de la población general a partir de las muestras de ejemplo (Maswadeh and Snyder, 2012). La Figura 6 muestra un ejemplo de curva ROC y su AUC.

Dado que el FAR y el CAR son funciones del umbral

$$\begin{aligned} FAR &= far(th) \\ FRR &= frr(th) \\ CAR &= 1 - FRR = 1 - frr(th) \end{aligned}$$ Ecuación 3

puede expresarse esta última como función de la primera, lo que constituye la ROC, y establecer el AUC como la integral de esa curva:

$$ROC: CAR = 1 - frr(th) = 1 - frr(far^{-1}(FAR))$$
$$AUC = \int_0^1 1 - frr\left(far^{-1}(FAR)\right) d(FAR)$$ Ecuación 4

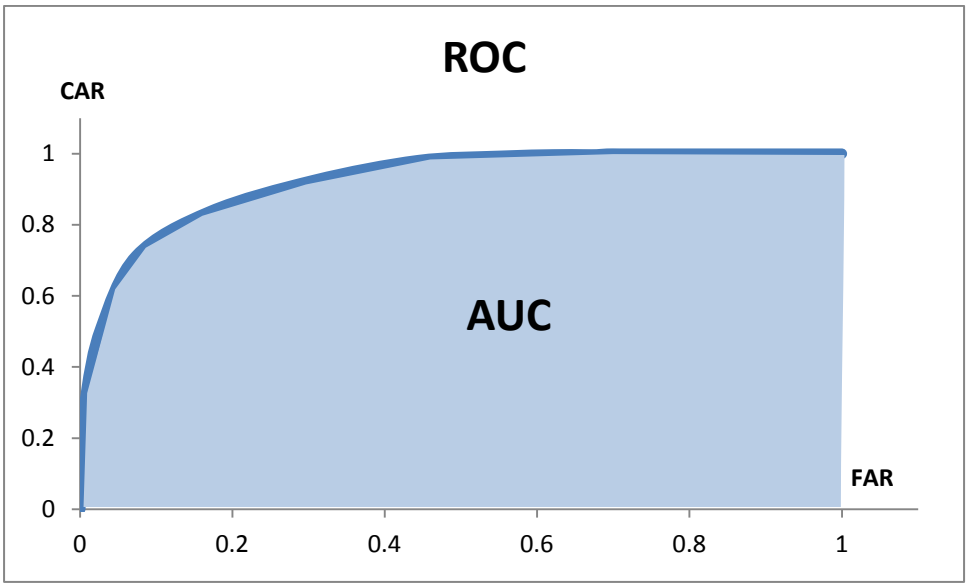


Figura 6: Curva ROC y AUC.

2.2 Percepción del sonido

Tras analizar el comportamiento de los sistemas de clasificación se procederá a revisar el funcionamiento de órganos humanos capaces de generar locuciones y reconocerlas. En primer lugar, y para analizar como los humanos reconocen los sonidos, se analiza el comportamiento del oído (órgano sensor de los mismos) desde un punto de vista puramente biomecánico, sin entrar en la consideración de las funciones neurológicas que permiten al cerebro entender el mensaje que contiene.

El oído humano es un complejo sistema para la percepción de los sonidos. Desde el punto de vista biológico está compuesto de múltiples elementos, agrupados en tres secciones: oído externo, oído medio y oído interno.

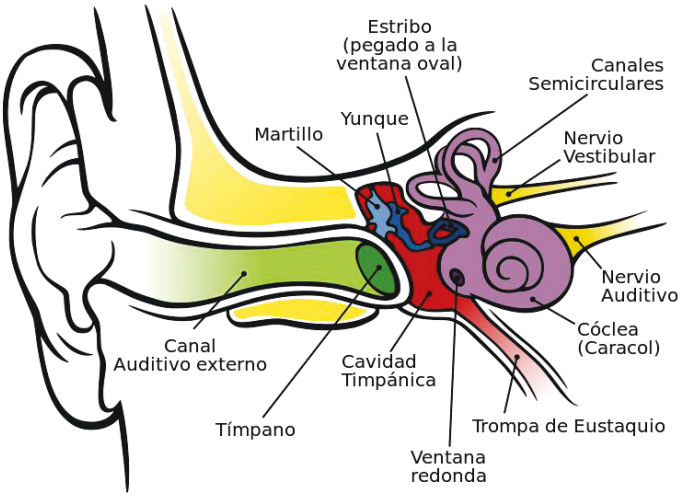


Figura 7: Esquema del sistema auditivo humano.[1]

El mecanismo de percepción del sonido comienza en el oído externo con la captación de las ondas de presión, que son dirigidas por el canal auditivo hasta el tímpano, que vibra por la acción de los cambios temporales de la presión del aire sobre él.

A lo largo del oído medio, esta vibración es transmitida por tres pequeños huesos: martillo, yunque y estribo, que realizan la adaptación de las impedancias del oído externo con las del oído interno, convirtiendo el movimiento de baja presión y alta amplitud en el tímpano, en otro de alta presión y baja amplitud en el apoyo del estribo sobre la cóclea denominado ventana oval.

En el oído interno las vibraciones de la ventana oval son convertidas en señales eléctricas que el nervio auditivo envía al cerebro.

La cóclea es una cavidad espiral situada en el oído interno de hasta 35mm de longitud, dividida longitudinalmente en tres cámaras llenas de fluido y delimitadas por la membrana basilar y la membrana vestibular (Figura 8).

Las características mecánicas de la membrana basilar varían progresivamente, siendo más estrecha y rígida en su inicio (base) y más ancha y elástica en su extremo final (ápex o ápice).

[1] Fuente: Perception Space—The Final Frontier, A PLoS Biology Vol. 3, No. 4, e137 doi:10.1371/journal.pbio.0030137 ([1]/[2]).

Sobre la membrana basilar se encuentra el *órgano de Corti*, formado por un conjunto de células ciliadas (aproximadamente 24000), que son sensibles a las vibraciones de la membrana basilar y están conectadas al nervio coclear.

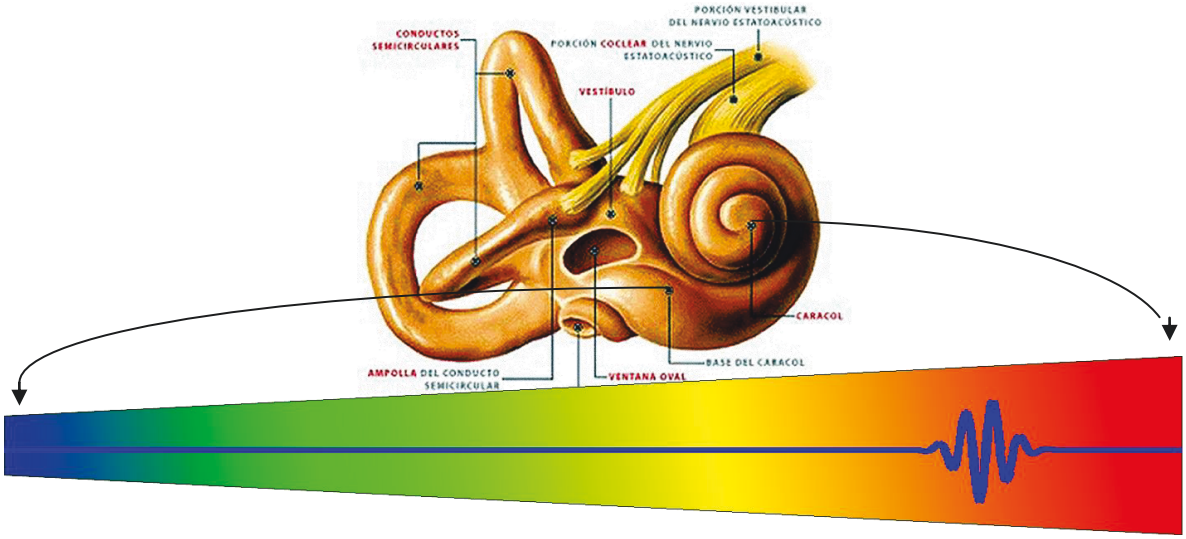


Figura 8: Respuesta en frecuencia de la membrana basilar dentro de la cóclea.[2]

Cuando una vibración hace oscilar la ventana oval, esta vibración se transmite al fluido interior de la cóclea y por éste a la membrana basilar, que es más sensible a esta vibración en el punto en el que la frecuencia de resonancia es más cercana a la frecuencia de la vibración. Ya que la frecuencia de resonancia depende las características de anchura y rigidez y estas dependen de la distancia a la base, se puede considerar que la membrana basilar realiza la conversión de frecuencia en espacio. Las membranas ciliadas convierten en señal eléctrica en el nervio coclear, la intensidad de la vibración a la que se ven sometidas, con lo que finalmente el cerebro recibe una representación espectral del sonido escuchado.

[2] Adaptada de www.sonoraudifonos.com

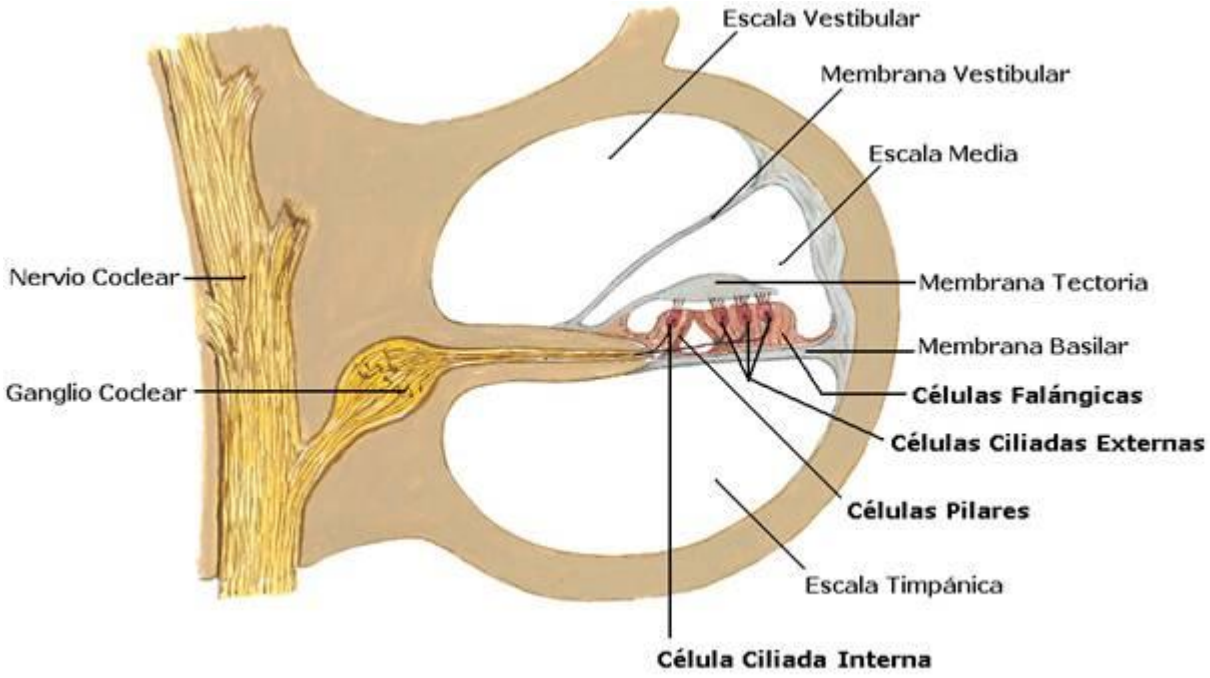


Figura 9: Sección de la cóclea con detalle del órgano de Corti.[3]

2.3 Producción de la voz

Es también esencial comprender cuales son los mecanismos con los que se produce la locución; por ello en la presente sección se revisa el comportamiento del aparato fonador, también desde el punto de vista biomecánico, sin entrar en la discusión del proceso cerebral encaminado a incorporar información en la señal acústica.

Desde el principio de los tiempos los seres vivos han sentido la necesidad de intercambiar información. Para cubrir esta necesidad, fue necesario desarrollar medios de señalización que tuviera suficiente capacidad de transmisión y que permitiera comunicar conceptos complejos. La señalización acústica ha demostrado ser la que presenta mejor cumplimiento de los requisitos en el hábitat natural, mejor que otras como las señales asociados a los otros cuatro sentidos (Crocker, 1997).

[3] Fuente: www.med.ufro.cl/Recursos/neuroanatomia/archivos/fono_oido_archivos/

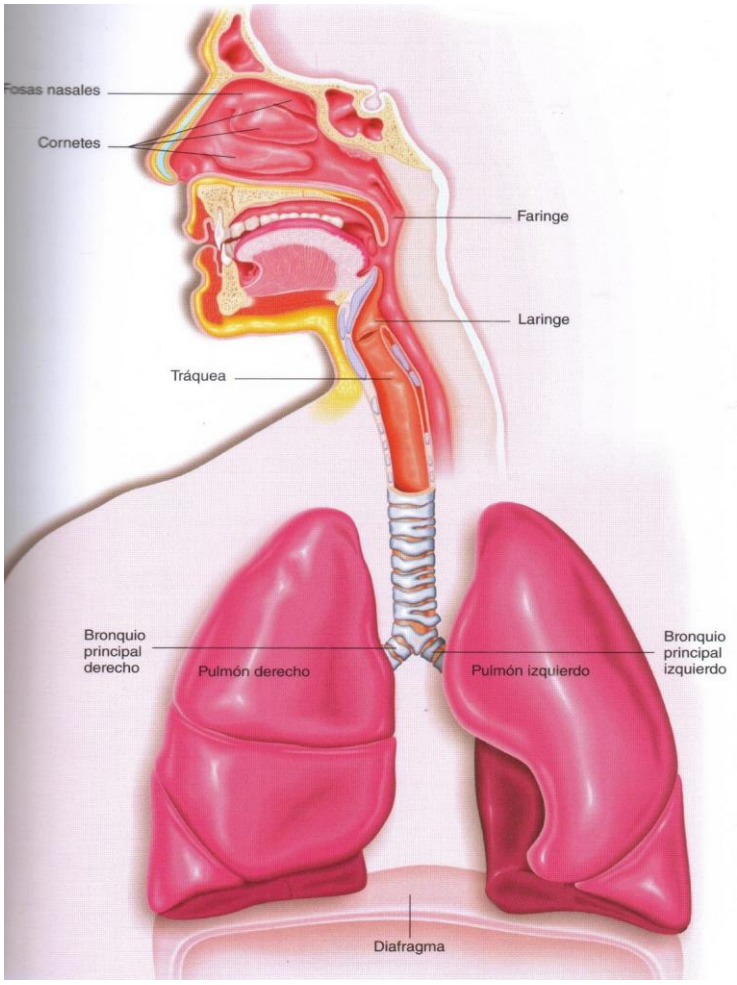


Figura 10: Sistema fonador.[4]

La evolución ha permitido que algunas especies de insectos emitan sonido haciendo entrechocar sus alas rígidas diseñadas inicialmente para proteger sus alas de vuelo mientras se encuentran en reposo; algunas aves hacen entrechocar sus picos creados para la alimentación y la caza o pesca; parece que la naturaleza ha reutilizado sistemas destinados a servir principalmente a otras funciones más críticas para la supervivencia. De la misma forma en el caso del hombre, la voz fue una adaptación evolutiva muy posterior a otras imprescindibles para la vida. Así, la laringe, a la que relacionamos de forma automática con la voz, tiene como función principal la protección de las vías respiratorias, de hecho muchos animales poseen pliegues vocales y no todos ellos emiten sonidos (Torres et al., 2007).

[4] Fuente: http://3.bp.blogspot.com/_B0jGw56qEd0/TUGN9RJW7hI/AAAAAAAAAGk/fdefH46SqUY/s1600/Sistema+Respiratorio+.JPG

2.3.1 Anatomía del tracto vocal

El aparato fonador humano está integrado por estructuras musculares de diferentes regiones y por elementos del aparato respiratorio y del aparato digestivo. El sistema respiratorio en combinación con las cuerdas vocales y las cavidades nasal y bucal pueden generar una gran diversidad de sonidos con patrones de energía diferenciados en el rango de frecuencias audibles por el oído humano. Estos patrones constituyen las señales del habla. Las convenciones culturales han establecido que estos sonidos sean emitidos en una secuencia precisa a la cual se le ha asignado un significado concreto constituyendo así el lenguaje hablado (Crocker, 1997).

En esencia los sonidos de la voz corresponden a los efectos producidos por una corriente de aire creada por la acción de los pulmones en sus movimientos de respiración pero principalmente durante la expiración. Dicha corriente circula por las vías respiratorias y es alterada por estas para producir ondas de presión, que finalmente serán emitidas al exterior por la boca y/o la nariz.

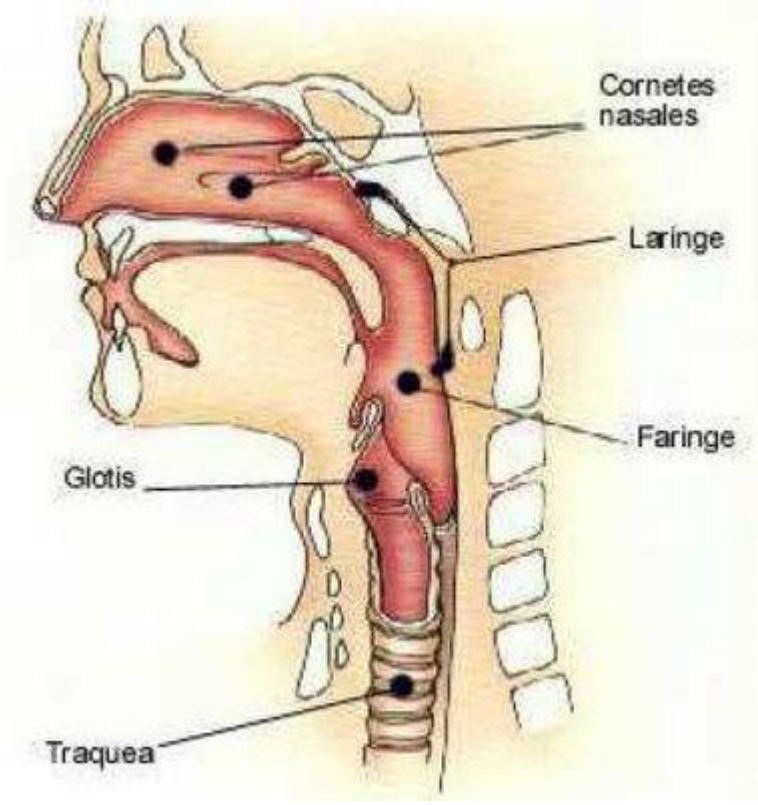


Figura 11: Esquema de las cavidades del tracto vocal en relación a la glotis.

El aparato fonador suele dividirse en tres partes fundamentales

- Cavidades infraglóticas

  En las cavidades infraglóticas se encuentran los órganos de la respiración: pulmones, diafragma, bronquios y tráquea.

- Cavidad laríngea

  En la laringe se encuentran las cuerdas vocales, dos pliegues laríngeos que son el principal generador de las ya citadas ondas de presión.

- Cavidades supraglóticas

  Por encima de la glotis se encuentran la faringe, la cavidad nasal y la cavidad bucal. Dentro de la cavidad bucal encontramos el paladar con sus partes dura y blanda, la lengua (ápice, dorso y raíz), los dientes y los labios. En conjunto modulan las ondas de presión de los sonidos vocálicos y participan activamente en la generación de los que no lo son.

Pulmones

Los pulmones son los órganos de la respiración por excelencia, su función básica es la de oxigenar la sangre. Son elásticos, esponjosos y ligeros. Cada pulmón está envuelto en su pleura. La pleura es un saco de doble pared, la interna unida al pulmón y la externa adherida a la pared torácica y a la cara craneal del diafragma. Gracias a esta unión los pulmones siguen al diafragma y a las costillas durante los movimientos respiratorios (Torres et al., 2007).

Clásicamente se describen tres tipos básicos de respiración: diafragmática, clavicular e intercostal.

Las respiraciones clavicular e intercostal no son muy útiles para la fonación ya que utilizan músculos accesorios de la respiración que al contraerse crean tensiones que la dificultan. Entre estos músculos encontramos los pectorales, algunos músculos del cuello y pequeños músculos profundos del tórax, que controlan las dimensiones de la caja torácica. Durante la inspiración se produce la elevación de las costillas, una proyección hacia delante de las superiores y una proyección lateral de las inferiores que incrementa el volumen torácico que deriva en un aumento de la capacidad de los pulmones, a la que acompaña de reducción de la presión intraalveolar y, como consecuencia, la aparición de una corriente de aire inspiratoria que la compensa. Durante la expiración las costillas vuelven a su posición de reposo y se produce el efecto inverso con una corriente de aire espiratoria (Torres et al., 2007).

La respiración diafragmática (o abdominal) es la que se produce en la frontera entre el tórax y el abdomen. En esa zona se dispone de mayor control voluntario los movimientos respiratorios principalmente gobernados por el diafragma.

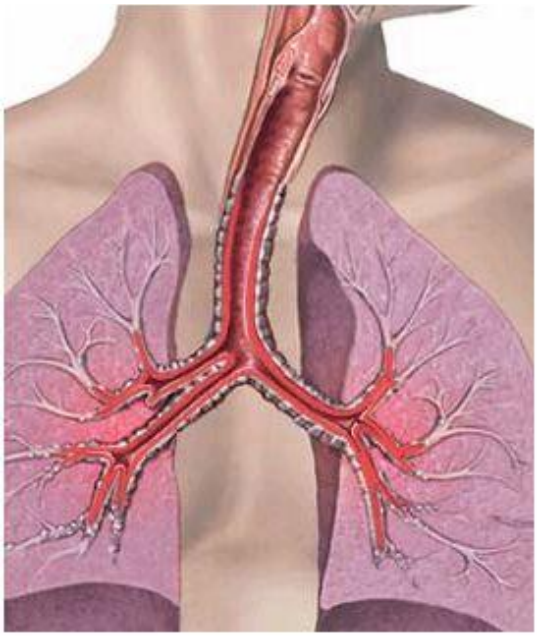


Figura 12: Tráquea bronquios y pulmones.[5]

Diafragma

El diafragma es el principal musculo de la inspiración. Se sitúa como una capa que separa la cavidad torácica de la abdominal, constituyendo el suelo de aquella y el techo de esta. Está formado por dos cúpulas convexas en dirección craneal. La cara superior se une a la pleura y el pericardio, y la inferior al peritoneo que envuelve las vísceras abdominales.

Durante la inspiración se contrae realizando un movimiento amplio de descenso que arrastra la pleura y con ella el pulmón aumentando verticalmente su volumen. La espiración normal es un proceso pasivo, en el que la fuerza de recuperación, debida a la elasticidad de las estructuras, produce un aumento de la presión y la consiguiente salida del aire. Por el contrario en la expiración forzada, la utilizada en la fonación, participan activamente múltiples músculos abdominales.

[5] Fuente: www.ferato.com/wiki/images/8/85/20071204_mgb_pulmones_.jpg

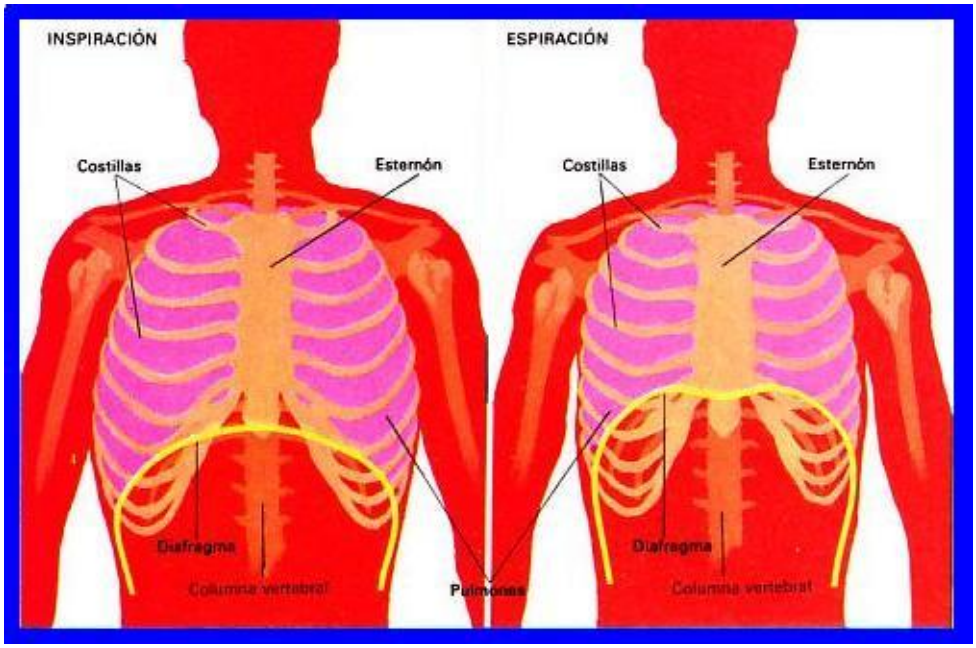


Figura 13: Inspiración, expiración, diafragma.[6]

Durante la inspiración el diafragma se contrae mientras la musculatura abdominal se relaja. De forma opuesta en la expiración activa el diafragma se relaja mientras la musculatura abdominal se contrae. Esta acción coordinada constituye *el soporte de la voz.*

Tráquea

La tráquea es un órgano respiratorio que conecta cranealmente con la laringe. Su función es proporcionar una vía de paso al aire durante la respiración. Es un tubo cilíndrico hueco y semi-rígido aplanado hacia atrás, de unos 12 cm de longitud y 2 cm de diámetro, formado por veinte anillos cartilaginosos en forma de herradura (Figura 11).

Bronquios

Los bronquios son cada uno de los dos conductos tubulares fibrocartilaginosos en que se divide la tráquea para entrar en los respectivos pulmones. Los bronquios son estructuras tubulares con ramificaciones progresivas arboriformes de diámetro decreciente, cuyas paredes están formadas por cartílagos y capas muscular, elástica y mucosa. Están encargados de conducir el aire al interior de los pulmones. Los bronquios primarios se dividen en bronquios secundarios y estos en bronquios terciarios.

[6] Fuente: esunmomento.es/contenido.php?recordID=413

Cuando las sucesivas divisiones alcanzan un tamaño muy pequeño, no tienen cartílago en la pared y se denominan bronquiolos. Los bronquiolos continúan dividiéndose y disminuyendo su calibre. Finalmente los alvéolos son expansiones sacciformes de la pared de los bronquiolos y que permiten el intercambio gaseoso con la sangre. Ver Figura 12.

Laringe

La laringe tiene la función vital de proteger las vías respiratorias y adicionalmente la de producir sonidos bajo la acción del aire durante la expiración. Se sitúa en la parte media y anterior del cuello. Comunica a través de la faringe con la cavidad bucal y las fosas nasales.

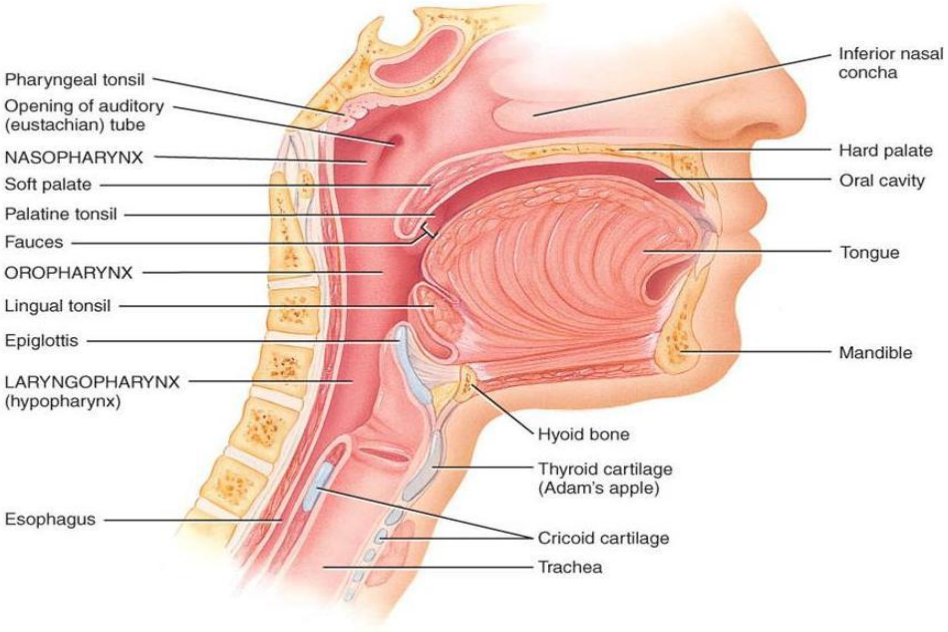


Figura 14: Sección sagital de la zona laríngea.[7]

La laringe está formada por un pequeño esqueleto de piezas cartilaginosas que se articulan entre sí. A ambos lados de la superficie interna se encuentran dos grupos de pliegues superpuestos: los superiores (pliegues vesticulares) y los inferiores (pliegues vocales). Los pliegues vesticulares se forman por la presencia del ligamento vesticular al cual recubren.

De la misma forma que los pliegues vesticulares, los pliegues vocales (cuerdas vocales) recubren el músculo y ligamento vocal y son responsables, con su vibración, de los sonidos vocálicos. Los pliegues vocales son muy elásticos y ello les permite

[7] www.uaz.edu.mx/histo/TortorAna/ch23/23_04.jpg

generar una gran variedad de sonidos. La posición y tensión de las cuerdas vocales determina el sonido producido, situación que está controlada por unos pequeños músculos que se insertan en los cartílagos laríngeos(Torres et al., 2007).

Estos músculos son:

- Cricotiroideo: que alarga tensa y aduce los pliegues vocales.
- Cricoaritenoideo posterior: único músculo abductor de los pliegues vocales.
- Cricoaritenoideo lateral: aductor de los pliegues vocales
- Vocal: Constituye la mayor parte del pliegue vocal. Es responsable de las variaciones locales de tensión durante la fonación.
- Tiroaritenoideo: Algunas de sus fibras se extienden hasta la epiglotis. Es aductor de los pliegues vocales
- Aritenoideo transverso: Es aductor de los pliegues vocales
- Aritenoideo oblicuo: Es aductor de los pliegues vocales. También algunas de sus fibras llegan hasta la epíglotis

En la laringe se distinguen tres regiones:

- Región supraglótica: es el espacio situado por encima de los pliegues o cuerdas vocales.
- Región media: donde se encuentra la glotis y los ventrículos laríngeos.
- Región infraglótica: espacio situado debajo de los pliegues vocales.

Glotis

La glotis es el espacio triangular que queda entre ambos pliegues vocales, cuando están separados. Durante la emisión de sonidos vocálicos son estos pliegues (cuerdas) los que con su vibración ondas producen las ondas de sonoras.

Faringe

La faringe es la estructura tubular que conecta la nariz y la boca con la laringe y el esófago. Por la faringe circula tanto el aire con dirección a los pulmones como los alimentos con dirección al estómago

Cavidad nasal

La cavidad nasal son dos espacios que se sitúan bajo el cráneo y sobre la cavidad bucal de la que la separa el paladar. Actúa como caja de resonancia en la emisión de ciertos sonidos, cuando el velo palatino está relajado y el aire expirado circula por ella.

Lengua

La lengua es un órgano musculoso situado en el suelo de la boca. En ella se distinguen las siguientes partes: la raíz, es la base de la lengua, que se sitúa por delante de la epiglotis; vértice o punta, la parte final de la lengua, el cuerpo, que se sitúa entre la raíz y el vértice; el dorso la parte superior del cuerpo de la lengua que en reposo contacta con el paladar, los bordes, laterales de la lengua que en reposo contactan con las encías y los dientes y la cara inferior que suele descansar sobre el suelo de la boca (Torres et al., 2007).

Paladar

El paladar es el cielo de la boca en los humanos y otros mamíferos. Separa las cavidades nasal y bucal. Se distinguen dos partes, el paladar duro en la parte frontal y el paladar blando o velo en la parte posterior

Velo

Es el tejido blando que forma la parte posterior del paladar. Se articula en la emisión de los sonidos nasales y en conjunción con la lengua para la producción de los sonidos velares.

Dientes:

Son estructuras duras que se asientan en los alveolos de los maxilares su función principal es la masticación, pero también actúan como obstáculos pasivos a la circulación del aire dando características especificas a cierto tipo de sonidos.

Alveolos

Los alveolos son compartimentos que presentan las mandíbulas donde están alojados los dientes

Labios

Los labios limitan la fisura oral. Lateralmente se unen formado la comisura labial. Para el control del gesto de los labios existen 22 músculos, entre ellos los más importantes son el músculo orbicular y el bucinador. Interactúan con otros órganos de

la boca, colaboran con la mandíbula en definir la apertura de la salida del sonido y pueden pronunciarse hacia el exterior modificando con ello la longitud del canal resonador (Torres et al., 2007).

2.3.2 La fonación

Los sonidos orales son producidos por la modulación del aire que circula a través de los estrechamientos que dinámicamente se forman en la laringe y las cavidades supraglóticas. De este modo se obtiene, a partir de un flujo continuo, un tren de ondas de presión que constituye el sonido.

En el periodo previo a la producción del sonido, el diafragma se contrae y los pulmones recogen aire por las vías respiratorias. Inmediatamente antes de la fonación (periodo prefonatorio) los pliegues vocales mantienen cerrada la glotis impidiendo la circulación del aire. A continuación la musculatura abdominal se contrae mientras el diafragma se relaja y aumenta la presión intraalveolar, presión que se transmite por la tráquea hasta la región subglótica. Cuando esta presión supera a la fuerza de cierre de la glotis, las cuerdas vocales son obligadas a separarse y el aire sale con fuerza, produciéndose un descenso brusco de la presión subglótica. La reducción de esta presión junto con la elasticidad de los pliegues vocales produce que la glotis vuelva a cerrarse y el ciclo comienza de nuevo (Torres et al., 2007). Este proceso se repite en rápida secuencia durante toda la fonación, produciendo un tren de ondas de presión que constituyen el componente fundamental de la voz vocálica. La frecuencia de este tren (frecuencia fundamental) está directamente relacionada con el grosor, tensión y longitud de las cuerdas vocales (en adultos 18 mm para los hombres y 10mm para las mujeres) produciendo una frecuencia fundamental de unos 125Hz para hombres y unos 200Hz para mujeres. Por otro lado la intensidad de la voz dependerá principalmente de la presión del aire espirado.

La presión existente en los pulmones es la causa de que el aire se mueva por las vías aéreas. Durante una emisión de una frase, la presión en los pulmones es relativamente constante, y el flujo de aire es controlado creando uno o más estrechamientos en la laringe o por encima de ella. El mantenimiento de la presión constante de los pulmones se consigue por la contracción de los músculos de la pared torácica, el diafragma y los músculos abdominales; asistidos por la fuerza de restauración que es consecuencia de la dilatación del espacio pleural producida cuando los pulmones fueron inflados (Stevens, 2007).

Una fuente de sonido es la vibración de las cuerdas vocales, que produce alteraciones cuasi-periódicas en el flujo de aire que cruza la glotis. Otro tipo de modulación es la causada por la aparición de obstáculos en el camino de la expiración, estos obstáculos producen turbulencias que constituyen un sonido de tipo "ruido".

Pueden formarse fuentes de sonido transitorias cuando tras el cierre momentáneo de la vía aérea y de producirse un aumento de la presión delante de él, sigue una rápida liberación que provoca un cambio abrupto en dicha presión.

Para la mayoría de los sonidos del habla, las vías aéreas anteriores a la laringe no tienen una influencia significativa en el sonido emitido. Sin embargo las resonancias en este espacio subglotal pueden ser observadas en el habla de algunas personas en sonidos con la glotis relativamente abierta.

La laringe, cuya función principal es la de proteger las vías respiratorias, es capaz de producir sonidos utilizando el flujo de aire que circula por la glotis durante la respiración. Las cuerdas vocales pueden ser configuradas principalmente de dos formas:

a) Por la contracción de los grupos musculares anexos a varios cartílagos de la laringe. Un tipo de ajuste que cambia la rigidez de las cuerdas vocales modificando la frecuencia de vibración.

b) Por la modificación de la distancia entre ambas cuerdas, distancia que puede variar desde que estén juntas hasta una configuración en la que se encuentren desplegadas.

La vibración de las cuerdas vocales sucede sólo en unos rangos concretos de rigidez y separación.

En presencia de presión subglotal estable, el flujo de aire en el tracto vocal está controlado por la configuración de la laringe y/o de estructuras supra-laríngeas que forman uno o más estrechamientos de la vía aérea.

2.3.3 Resonadores

Las cavidades situadas por encima de la glotis (la faringe, la cavidad nasal y la cavidad bucal) pueden ser configuradas para que actúen como cajas de resonancia.

La faringe es un resonador relativamente dinámico, aunque tal dinamismo está limitado a los desplazamientos que la musculatura produce sobre la laringe, o a los cambios en su terminación que produce el movimiento de la lengua y el velo palatino.

La cavidad nasal es un resonador estático, no es posible cambiar sus características aunque, como ya se ha visto, es posible hacer que participe o no gracias a la acción del velo palatino.

La boca es un resonador móvil, su configuración puede ser alterada en múltiples aspectos. Se puede modificar la apertura mandibular que afecta tanto al

volumen como a la adaptación con el exterior; se puede modificar la forma de los labios alargando o cerrando en mayor o menor medida el tracto; la lengua puede modificar la forma y volumen de la cavidad, finalmente la interacción entre labios, dientes y lengua, alveolos y paladar, pueden no sólo modificar los sonidos procedentes de la glotis sino ser ellos mismo fuentes de otros sonidos. La habilidad de la boca para cambiar su configuración permite la generación la gran variedad de sonidos distintos que constituyen los componentes del habla.

Las configuraciones que pueden adoptar las vías supraglóticas tienen dos funciones en la producción de la voz. Una es formar estrechamientos en algunos puntos. La segunda es configurar las vías aéreas para tener frecuencias naturales particulares, de tal modo que constituyan un filtro que modifique el sonido. (Stevens, 2007)

Hay cinco estructuras anatómicas principales que pueden ser utilizadas para cambiar la configuración del tracto vocal (Stevens, 2007).

1. Al ser contraídos los músculos que rodean la faringe producen el estrechamiento de la zona.
2. El cuerpo de la lengua puede ser desplazado de forma vertical y horizontal.
3. El vértice lingual puede configurarse como extensión del cuerpo de la lengua y pude ser desplazado hacia arriba para formar un estrechamiento extendido desde los dientes hasta el paladar duro.
4. Los labios pueden ser configurados y desplazados para formar un estrechamiento o sobresalir para incrementar la longitud efectiva de la vía aérea.
5. Finalmente el paladar blando subido o bajado para variar el área de la sección transversal del paso del la faringe alta a la cavidad nasal llegando a cerrar este paso completamente .

2.3.4 Articulación de sonidos

El fonema es la unidad menor distinguible del sistema fonológico de una lengua, sin embargo cada fonema puede ser realizado de formas muy diversas dependiendo del resto de fonemas de su entorno, de la variedad lingüística del locutor o de sus propias condiciones. Cada una de estas realizaciones es conocida como alófono.

Articular un fonema es el acto de posicionar de una manera concreta los órganos articulatorios (lengua, paladar, dientes, labios) para producir un sonido específico alterando el flujo proveniente de la laringe.

En la producción de los sonidos alguno de esos órganos, denominado articulador activo (la punta de la lengua, labio inferior, dorso de la lengua…) se ubica respecto a un articulador pasivo (labio superior, incisivos superiores, alveolos, paladar duro, velo palatino) para interponer obstáculo al flujo de aire o formar un estrechamiento en el conducto vocal (constricción).

Desde el punto de vista de la fonética, las consonantes se producen bloqueando el flujo de aire procedente de los pulmones lo que causa una diferencia de presión a ambos lados del obstáculo, mientras que las vocales se producen sin dicha obstrucción, con el tracto vocal abierto y sin dicha diferencia de presión (Hualde, 2005).

Al estudiar los sonidos del habla es costumbre establecer la antedicha distinción entre vocales y consonantes; pero de esta gran división, impuesta por la tradición lingüística, no se pueden inferir criterios articulatorios o acústicos que caractericen a las unas respecto de las otras, ya que desde la vocal más pura (/a/) hasta la consonante más pura (/p/), existe toda una gradación de sonidos que no presenta fin de continuidad alguno (Martínez Celdrán, 1996). Así, por ejemplo, desde la perspectiva fonológica[8] se clasifica a las "aproximantes" (ver 2.3.5) como consonantes mientras que la fonética[9] las considera próximas a las vocales ya que no existe ningún obstáculo que interrumpa el flujo de aire.

En este trabajo, más interesado en el sonido en sí que en la función que desempeñan en la comunicación, se analizan los fonemas como resultado de los mecanismos que los producen. A tal fin se consideran los tres atributos que definen de configuración los órganos articulatorios y con ello el sonido producido. Estos atributos son: la forma en que se produce la articulación, el punto en que se produce la articulación y la actividad de las cuerdas vocales durante la articulación.

2.3.5 Forma en que se articulan

La forma de la articulación hace referencia a cómo se produce la constricción. En razón de ello se pueden obtener sonidos: vocálicos, aproximantes, fricativos, oclusivos, nasales, africados, laterales y vibrantes (Hualde, 2005).

[8] Estudio de los sonidos del habla en relación a su función en la comunicación.
[9] Estudio acerca de la acústica y fisiología de los sonidos del habla.

Sonidos vocálicos:

No se produce ninguna obstrucción, pero la disminución en el calibre de la vía aérea, definido por el hueco entre la lengua y el paladar, junto con el punto donde se produce esta disminución son las causas de los distintos fonemas de este tipo. Ejemplos de este tipo de sonido son: /a/, /e/, /i/, /o/, /u/.

Sonidos aproximantes:

Son sonidos que son articulados por una aproximación entre dos órganos que ni interrumpe totalmente la corriente de aire, ni la apertura es suficiente estrecha como para producir fricción, ni la apertura es tan amplia como en los vocálicos. Ejemplos: /l/ /ll/.

Sonidos fricativos:

En estos sonidos el flujo de aire no es completamente interrumpido, sino que escapa por una vía estrecha entre los articuladores, produciendo turbulencia o fricción. Ejemplos de este tipo de sonido son: /f/, /z/, /j/, /s/.

Sonidos oclusivos:

Estos son sonidos en los que durante la articulación el flujo de aire proveniente de los pulmones es completamente bloqueado para a continuación ser liberado de forma explosiva, originando un ruido característico. Ejemplos de este tipo de sonido son: /p/, /b/, /t/, /d/, /k/, /g/, /n/, /m/.

Sonidos nasales:

Al igual que los oclusivos, se produce un bloqueo completo de la vía oral, pero se permite que el aire fluya por la vía nasal, lo que se consigue bajando el velo palatino. La cavidad bucal actúa como cavidad de la resonancia, pero el aire no sale por la boca al ser bloqueado por la lengua. Ejemplos de este tipo de sonido son: /m/, /n/, /ñ/.

Sonidos africados:

La articulación de estos sonidos consta de dos fases, comienza con una oclusión y terminar con una liberación fricativa. Ejemplos de este tipo de sonido son: /ch/, /ñ/.

Sonidos laterales:

Estos sonidos se producen cuando los articuladores se mantienen en contacto en el eje central, obligando al aire a fluir por uno o ambos lados. Ejemplos de este tipo de sonido son: /l/, /ll/.

Sonidos vibrantes:

Las vibrantes se caracterizan por presentar periodos seguidos de periodos vocálicos, en rápida sucesión. Este ciclo oclusivo-vocálico puede producirse una sola vez (vibrante simple) o varias veces seguidas (vibrante múltiple). Ejemplos de este tipo de sonido son: /r/, /rr/.

2.3.6 Punto de articulación

Este criterio clasifica los sonidos en función de los articuladores implicados, así tenemos sonidos: bilabiales, labiodentales, interdentales, dentales, alveolares, palatales, prepalatales, velares y glotales (Hualde, 2005).

Sonidos bilabiales:

La articulación se produce por la aproximación o el contacto del labio superior y el inferior. Ejemplos de este tipo de sonido son: /p/, /b/, /m/

Sonidos labiodentales:

En la producción de este tipo de sonidos el labio inferior es el articulador activo que entra en contacto con los incisivos superiores (articulador pasivo). Ejemplo de este tipo de sonido es: /f/

Sonidos interdentales:

En los sonidos interdentales la punta de la lengua se coloca entre los incisivos superiores e inferiores. Ejemplos de este tipo de sonido es la /z/

Sonidos dentales:

En estos casos el articulador pasivo es la base de los incisivos superiores donde se apoya la punta de la lengua. Ejemplos de este tipo de sonido son: /t/, /d/.

Sonidos alveolares:

Para los sonidos alveolares, la lengua se apoya sobre las inserciones de los incisivos superiores. Ejemplos de este tipo de sonido son: /s/, /l/, /r/, /rr/, /n/

Sonidos palatales:

Se producen por la aproximación del dorso de la lengua hacia el paladar duro. Ejemplos de este tipo de sonido son: /ch/, /y/, /ll/, /ñ/

Sonidos prepalatales:

Para estos sonidos la constricción se forma por el contacto de la parte anterior del dorso de la lengua con el área comprendida entre los alveolos y el paladar duro. Ejemplo de este tipo de sonidos es la realización francesa de la /j/

Sonidos velares:

Se articulan por la aproximación de la parte posterior del dorso de la lengua al velo palatino. Ejemplos de este tipo de sonido son: /k/, /g/, /j/

Sonidos glotales:

A diferencia de los demás, en estos sonidos la obstrucción se produce en la glotis. Ejemplo de este tipo de sonido es la realización francesa de la /r/

2.3.7 Actividad de las cuerdas vocales

Finalmente el tercer atributo distingue los sonidos en función del estado vibratorio de las curdas vocales. Se distinguen así:

Sonidos sonoros:

Aquellos en que las cuerdas vocales están vibrando durante la articulación. Ejemplos de este tipo de sonidos son: /b/, /r/, /n/…

Sonidos sordos:

Aquellos en el que las cuerdas vocales están en reposo. Ejemplos de este tipo de sonidos son: /p/, /t/, /z/, /s/…

2.3.8 Resumen

En resumen, son precisos tres grupos de órganos funcionando de forma coordinada para que se pueda producir el lenguaje hablado. Estos grupos son:

- **El sistema respiratorio:** que suministra el flujo de aire y la energía necesaria para la emisión de sonidos
- **La laringe:** que proporciona, cuando se precisa, una alteración cuasi-periódica del flujo de aire.
- **El sistema articulatorio:** constituye un sistema que provoca turbulencias en el flujo de aire y/o lo modifica utilizando las cualidades resonantes de las cavidades supraglóticas.

2.4 Modelos de la producción de la voz

Un modelo lineal de producción del habla fue desarrollado por Fant a finales de la década de los 50. En él (Fant, 1970) se afirma que la señal de voz es la respuesta del sistema de filtros del tracto vocal a una o varias fuentes de sonido. Al considerar los filtros independientes de las fuentes deduce que la señal de voz puede ser descrita exclusivamente en términos de las características de la/s fuente/s y del filtro. El concepto de esta teoría de la producción del habla ("fuente y filtro") está representada en las Figura 15 y Figura 16.

La Figura 15 presenta el modelo de Fant de producción de sonidos cuando se emplea la glotis como única fuente. En ella se muestra el detalle del acoplamiento de las cavidades nasales al resto del tracto vocal en el paso de la cavidad faríngea a la cavidad bucal.

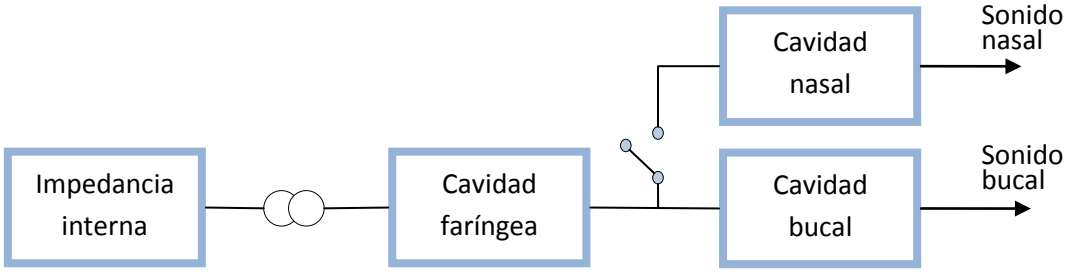


Figura 15: Modelo de producción del habla.[10]

De forma complementaria, la Figura 16 representa la producción de un sonido no nasal considerando variable la posición en la que se encuentra fuente del sonido. Para ello contempla dos bloques de filtros, "cavidades posteriores" para la sección del aparato fonador que se encuentra después de la fuente de sonido y "cavidades anteriores" para las que se encuentran antes que ella. Cubre así los casos de los fonemas sonoros, donde "cavidades posteriores" representa a todas las cavidades supraglóticas, hasta los oclusivos donde la faringe es parte las "cavidades anteriores". En consecuencia resalta las diversas funciones de los segmentos vocales, y apunta a que es posible dividir el tracto vocal en relación a la fuente del sonido.

La teoría acústica ha ampliado este modelo y demostrado la validez del concepto, donde las características de transmisión del tracto vocal pueden ser aproximadas por una cascada de resonadores y anti-resonadores cuyos anchos de

[10] Adaptado de: Acoustic Theory of Speech Production(Fant, 1970)

banda y frecuencias centrales son controladas independientemente (Childers and Wong, 1994).

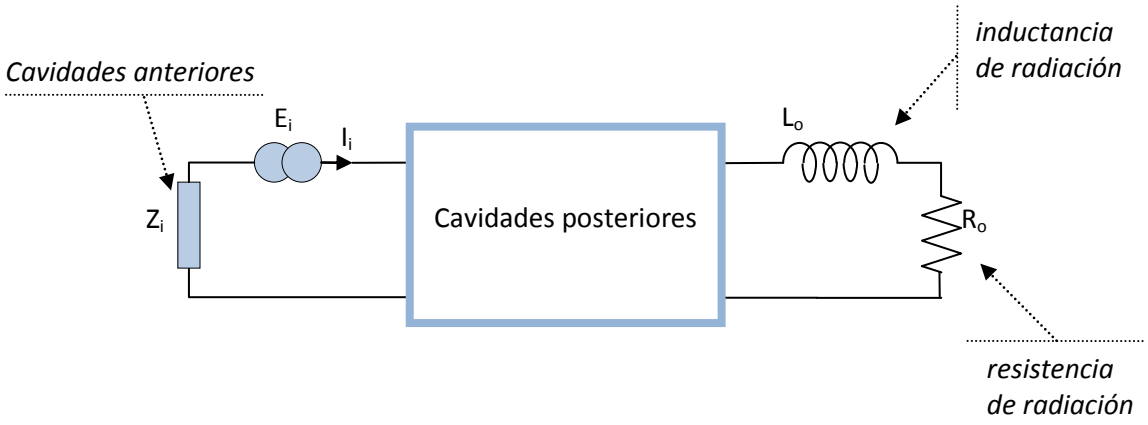


Figura 16: Modelo del proceso de filtrado.[11]

Esta teoría de *"filtro y fuente"* establece que las fuentes y el filtro del tracto vocal son linealmente separables y no interactúan. En consecuencia: el tracto vocal puede ser representado como un filtro invariante en el corto plazo, la fuente glotal (en los sonidos sonoros) por un generador del tren de pulsos cuasi-periódicos con un periodo y amplitud controlable; y el resto de fuentes es representado por un generador de ruido aleatorio. En base a ello, la Figura 17 presenta una versión más detallada del esquema de producción de voz en tiempo discreto.

2.4.1 Función de transferencia

La función de filtrado (función de transferencia), está definida como el cociente entre la presión del sonido en un punto específico próximo a los labios del locutor (P_L) y la presión del sonido en la fuente (P_G), presiones dependientes de la frecuencia. Una versión equivalente de cálculo es el cociente de las respectivas velocidades volumétricas[12] (Ecuación 5).

$$T(f) = \frac{P_L(f)}{P_G(f)} = \frac{U_L(f)}{U_G(f)} \qquad \textbf{Ecuación 5}$$

[11] Adaptado de: Acoustic Theory of Speech Production(Fant, 1970)

[12] Dado un flujo donde la velocidad media de la partículas en un punto determinado es *v* y el área de la sección transversal en ese punto es *S*, la velocidad volumétrica será: $U = v \cdot S$.

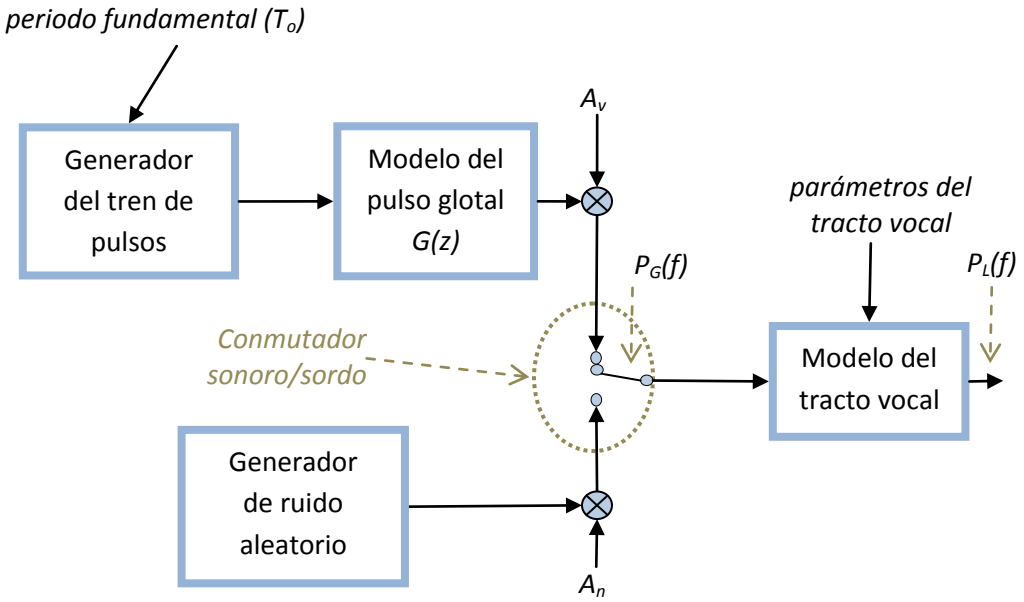


Figura 17: Modelo de la producción de la voz en tiempo discreto.[13]

Si se considera que la señal es la presión en la fuente $S = P_G$

$$P_L(f) = S(f) \cdot T(f) \qquad \text{Ecuación 6}$$

La propiedad básica de las cuerdas vocales, y por tanto de P_L en sonidos sonoros, es su periodicidad expresada por la duración *(T_0)* de un periodo completo y su inversa, la *frecuencia fundamental* de la voz *(F_0).* Dicha duración a menudo varía de un periodo al siguiente; tales variaciones son en parte sistemáticas en un determinado el patrón de entonación, y en parte no intencionadas aunque importantes para la naturalidad del habla humana.

Como se ha dicho con anterioridad, los sonidos vocálicos se producen normalmente teniendo como fuente la glotis y con las vías aéreas abiertas, de modo que se cualquier estrechamiento no es suficiente para causar una diferencia de presión a ambos lados del mismo.

Para este caso el tracto puede aproximarse, en primera instancia, a un tubo de longitud *l* (Figura 18), la distancia desde la glotis hasta los labios (Stevens, 2007), en el que la configuración viene definida por el área de la sección transversal como función de su distancia a la glotis (función de área). Asumiendo que la función de área es

[13] Adaptado de: Acoustic Theory of Speech Production(Fant, 1970)

uniforme, ignorando las pérdidas acústicas y siendo c la velocidad de propagación del sonido[14], la función de transferencia es (Stevens, 2007):

$$T(f) = \frac{1}{\cos(2\pi f l/c)} \qquad \text{Ecuación 7}$$

Esta función (Ecuación 7) sólo posee polos y estos se encuentran en las frecuencias dadas por la Ecuación 8 en una sucesión infinita. Estos polos, denominados *formantes,* corresponden a las frecuencias naturales del tubo cuando está cerrado a nivel de la glotis.

$$F_n = \frac{(2n-1)c}{4l} \qquad \text{Ecuación 8}$$

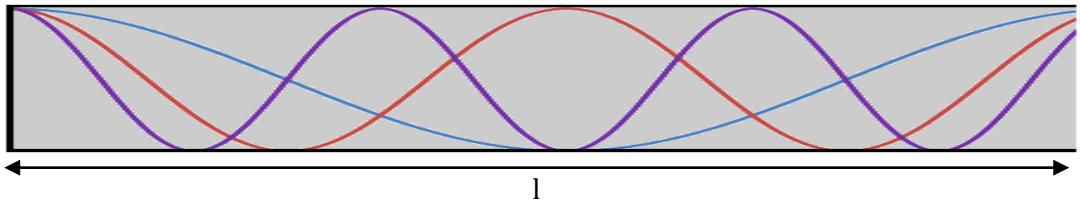


Figura 18: Formación de sonidos en un tubo.

Si se incluyen los efectos de las perdidas y el valor finito de la impedancia de radiación, los polos de la función de transferencia se ven afectados, de modo que reducen ligeramente de la frecuencia de las formantes y se establece un ancho de banda de valor finito (Stevens, 2007).

La función de transferencia, cuando la función de área no es uniforme, puede ser aproximada por la correspondiente a un tubo de sección variable y éste por una cadena de tubos cada uno de ellos de sección constante. Los estudios teóricos muestran que si existe un estrechamiento que encuentre próximo al punto donde se hace máxima la energía cinética las frecuencias disminuyen. De forma análoga los un estrechamiento en un punto cercano al máximo de energía potencial de un nodo hará que las frecuencias aumenten. Por el contrario, para ambos caos, las frecuencias aumentan y disminuyen respectivamente si se produce un ensanchamiento en lugar de un estrechamiento (Stevens, 2007).

[14] Velocidad del sonido c = 343 m/s (en el aire a una temperatura de 20 °C)

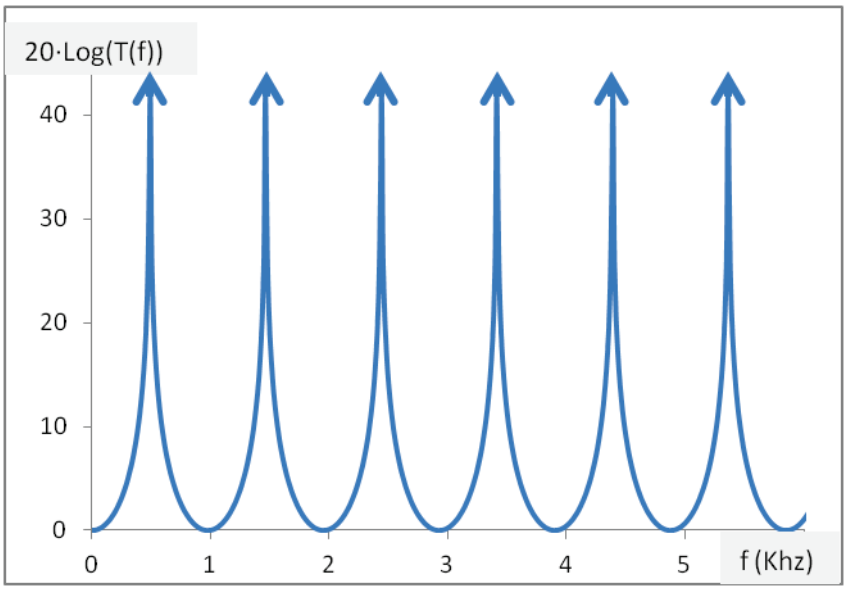


Figura 19: Función de transferencia de un tubo.

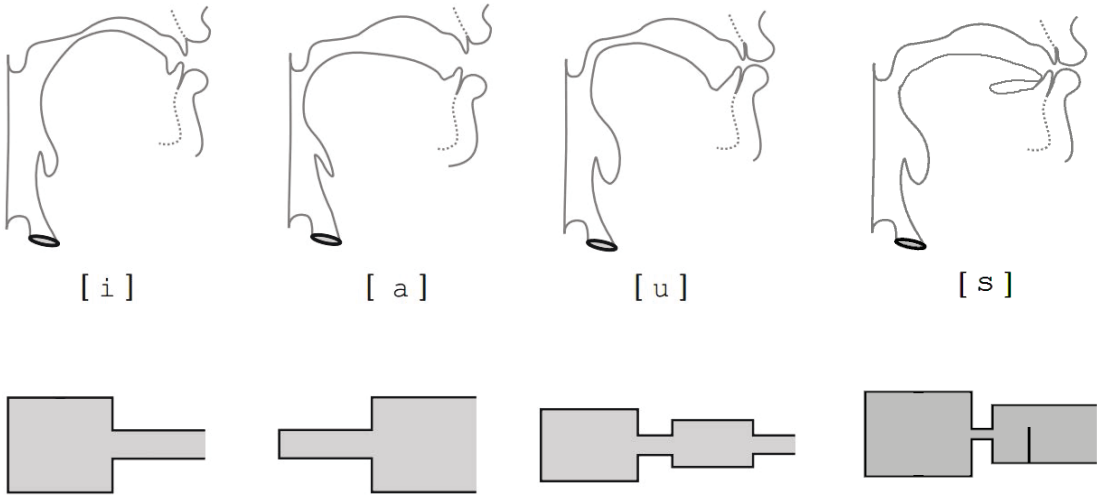


Figura 20: Esquema de la configuración diversos fonemas.[15]

Los sonidos más puramente consonánticos se producen cuando aparece una constricción relativamente estrecha en algún punto de la vía aérea con la incorporación de fuentes de ruido y/o glóticas. La Figura 20 muestra entre otras la configuración y el modelo del fonema fricativo /s/; para esta configuración el flujo de aire es enviado por la constricción hacia los incisivos inferiores y la fuente principal

[15] ocw.mit.edu

de las turbulencias se localiza en las proximidades de estos, con lo que puede ser modelado como un tubo con una constricción y con un obstáculo donde se ubica la fuente de sonido. Para dimensiones típicas de 1 cm desde la constricción al obstáculo y una cavidad posterior de 1 cm de longitud y de 1 cm^2 de área de la sección transversal, la función de transferencia presenta un pico alrededor de los 4400Hz y un cero cercano a los 0 Hz.

2.5 Voz y reconocimiento

La extrema complejidad del sistema fonador es la causa de que la voz sea a su vez extremadamente variante; no sólo por tratarse de una señal (por definición variable en el tiempo), sino porque no presenta patrones temporales claramente estables, ni tampoco presenta un comportamiento muy similar entre distintos individuos. Esta variabilidad está causada por dos tipos de actuantes, unos voluntarios y otros intrínsecos.

Los voluntarios corresponden a alteraciones intencionadas de la configuración del aparato fonador que resultan en la producción de sonidos diversos. El cambio de la presión del aire en las cavidades produce directamente un cambio en el volumen de la voz, la alteración en la tensión de las cuerdas vocales causa inmediatamente el cambio del tono, el cambio en la longitud de la laringe deriva en un cambio en las frecuencias naturales del filtro vocal y en consecuencia en los armónicos que acompañan al fundamental, el cierre, con la lengua, de la cavidad bucal está asociado a la aparición los bien conocidos de sonidos nasales, la aparición de distintos obstáculos y distintas constricciones ocasionan la producción de los distintos fonemas, etc. Es este gran repertorio de posibilidades el que permite alojar en la voz tan gran contenido de información que ha convertido el lenguaje hablado en el principal medio de comunicación natural (Doddington et al., 2000).

Los actuantes intrínsecos no están relacionados con la voluntad del locutor, pero establecen parámetros de la voz que varían poco en el corto plazo, algunos debidos a las características anatómicas de la persona, pero también debidos a las características psicológicas, ambientales, de la salud, culturales y otros tipos de coyunturas en las que se puede ver inmerso el locutor.(Enos et al., 2007) (Reynolds, 2002).

En consecuencia la señal de voz reúne un alto volumen de información de diversos aspectos, donde lo fundamental, desde el punto de vista evolutivo, está relacionado con el mensaje que el locutor pretende transmitir al oyente, mensaje que está formado por secuencias de sonidos que forman palabras, que a su vez se agrupan en secuencias que constituyen frases. De forma colateral también transporta

información relativa al emisor en sí mismo y a sus circunstancias. Información, esta última, que puede ser aprovechada para el reconocimiento del sujeto emisor.

2.6 Tecnología

Los humanos pueden reconocer a otras personas por la voz sin necesidad de ver a quien habla. Más allá de la capacidad del hombre de transmitir y recibir lenguaje hablado, la naturaleza ha dotado al ser humano de la capacidad de reconocer a las personas por su voz. La voz transporta junto con la información semántica, datos relativos a la identidad del emisor, elementos que son personales y discriminativos. En el modo de hablar de la persona se resumen características físicas, psíquicas, idolectales[16], sociolectales[17], dialectales[18], emocionales…, parámetros que no tan sólo le distinguen como individuo, sino también como perteneciente a cierto grupo o por estar sometido a determinadas circunstancias (Ruiz-Mezcua, 1998) (Zetterholm, 2007) (Dellwo et al., 2007); información que puede ser utilizada por los oyentes o equipos especializados para describir, clasificar y/o identificar a las personas.

El reconocimiento de locutor, considerado como el proceso automático por el que se determina quién está hablando basándose en la información contenida en la señal acústica producida (Furui, 1996), es un ejemplo del reconocimiento biométrico de las personas.

Por razones de coevolución el sistema auditivo y el fonador se han desarrollado paralelamente y por lo tanto los mecanismos de producción del habla están íntimamente vinculados a los de su recepción, no es posible entender cómo se reconoce sin entender cómo se produce, pero a su vez es cierto lo contrario, ya que el habla se ha desarrollado para construir las secuencias de sonido que son inteligibles por el cerebro cuando han sido tratadas por el oído. Pese a la complejidad que supone el reconocimiento que las personas hacen de sus congéneres gracias a su voz es tan rápido y eficaz que la ciencia intenta entender ese proceso para poder reproducirlo en sistemas artificiales.

Parece razonable decir que la investigación del habla está actualmente lejos de proponer un modelo detallado que explique cómo los oyentes acometen la percepción de la voz. Aunque se han propuesto múltiples e influyentes teorías y éstas han sido sometidas a experimentación, no hay todavía acuerdo sobre los mecanismos que hacen que la percepción del habla sea tan singularmente rápida y robusta.

[16] Relativo a la forma de hablar característica de cada persona.
[17] Relativo a la variedad lingüística usada por una clase social.
[18] Relativo a la variedad de la lengua determinada por la zona geográfica.

A partir de la propia observación queda claro que los humanos basan su capacidad de reconocer a otros tan sólo por su voz, en la información que a varios niveles se aloja en ella. Empezando por el nivel más bajo que representa simplemente el sonido de la voz y terminado por el más alto (idiolectal) que define la particularización cada uno hace de las palabras disponibles en el idioma, nivel que está relacionado con los hábitos y estilos aprendidos

2.6.1 Reconocimiento del locutor

Bajo el concepto de Reconocimiento (de locutor en el presente caso) se reúnen dos tipos de proceso *verificación* e *identificación* (Campbell Jr, 1997).

En la *verificación automática de locutor* (ASV) la indeterminación está acotada; se dispone de una muestra de la voz y de una identidad pretendida (c*laimed identity*), proporcionada voluntariamente por el locutor o se trata de una suposición realizada por el operador del reconocedor. En este caso la misión del sistema es establecer si es cierto o no que la muestra corresponde a dicha identidad; para ello, el sistema debe disponer de una descripción de la voz (*modelo*) de la persona a la que corresponde la identidad (*usuario*); la decisión se toma a partir de la comparación de la muestra con dicho modelo (Carrero et al., 2008a).

En la *identificación automática de locutor* (ASI) no se cuenta con una presunción de la identidad del *donante,* y es preciso que el sistema establezca a quién, si alguien, corresponde la muestra entre el conjunto de personas para las que se dispone de modelo (*usuarios*). El sistema ya no emite una respuesta booleana, sino que principalmente ofrece la identidad encontrada o, alternativamente, la identidad junto con la *confianza* que se tiene en la decisión, o un conjunto de identidades y sus confianzas.

Los anteriores procesos comparten en gran medida sus técnicas, algoritmos y criterios. Básicamente se considera el proceso de identificación como un conjunto de n procesos de verificación en los cuales se plantea la comparación con todos y cada uno de los modelos disponibles.

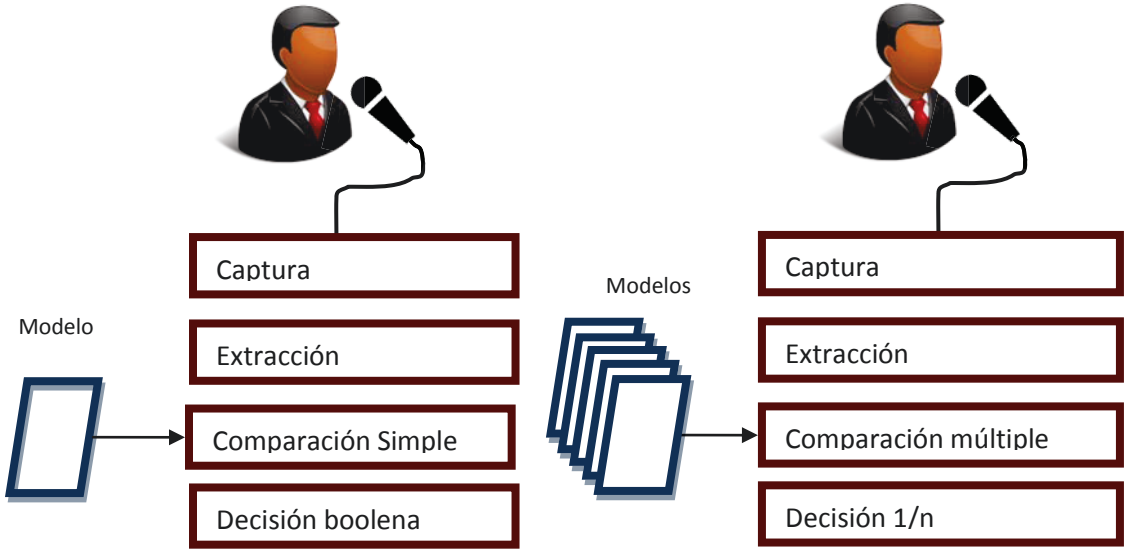


Figura 21: Procesos de verificación (izq.) e identificación (dcha.)

2.6.2 Dependencia e independencia del texto

La mayoría de las aplicaciones desarrolladas precisan de la cooperación del usuario con el sistema, ya que se presupone que está interesado en que el sistema les reconozca, enumerando una serie de dígitos, enunciando claves o repitiendo una frase que le ha sido mostrada, y que ha sido construida a partir de un vocabulario reducido (*text-dependent* o *text-constrained*) (Poza et al., 2008d).

En esos escenarios estas restricciones pueden ser razonables y gracias a ellas se obtiene una mejora sustancial de las prestaciones del sistema. Sin embargo existen situaciones donde este tipo de restricción puede complicar el proceso o resultar de todo punto inviable, en estos casos es necesario un sistema de verificación capaz de operar en situaciones de ausencia de cooperación y que basan su tarea en la independencia del texto (Bimbot et al., 2004)(Ruiz-Mezcua, 1998).

2.6.3 Short range

Del mismo modo la mayoría de los sistemas comerciales, utilizan las características espectrales de la voz obtenidas habitualmente a partir de piezas subsegmentales del discurso, pequeños fragmentos del discurso de duraciones del orden de pocas decenas de milésimas de segundo (*short-range*).

Este enfoque resulta altamente exitoso en condiciones acústicas limpias, pero sufre una degradación significativa en sus prestaciones cuando se presentan alteraciones en el canal (Shriberg et al., 2005).

2.6.4 Long range

Dado que las muestras espectrales no son construidas ni consideradas en atención a su orden temporal, el enfoque short-range no resulta eficaz cuando se persiguen características estilísticas personales, tales como hábitos léxicos, hábitos prosódicos, etc. Para solventarlo es posible cambiar el planteamiento y estudiar fragmentos mayores (*long-range*) donde muestran sus efectos las estas características suprasegmentales.

Los beneficios de la utilización este enfoque en reconocimiento automático de locutor son al menos tres.

1. Tales características pueden mejorar las prestaciones de las obtenidas de los patrones espectrales. Se ha encontrado que añadir características long-range, puede ofrecer una gran mejora de prestaciones cuando se dispone de grandes cantidades de datos de entrenamiento.

2. A diferencia de las características basadas en tramas, las long-range reflejan el comportamiento voluntario y esto potencialmente puede resultar útil no tan sólo en reconocimiento de locutores, sino también en el reconocimiento de las características del discurso tales como el estilo.

3. Es fundamental para la investigación del comportamiento para entender las características de su manifestación en el habla.

2.6.5 Técnicas suprasegmentales

En lingüística y especialmente en fonética y fonología, se considera como *segmento* cualquier unidad discreta que pueda ser identificada emisiva o auditivamente en flujo de producción del habla. El segmento es un concepto muy cercano al de fonema, y se consideran discretos porque pueden ser identificados individualmente (como las vocales o las consonantes) ya que ocurren en distintos instantes de la locución. Algunos efectos pueden producirse a lo largo de varios segmentos y en dependencia con ellos. Estos efectos son conocidos como *suprasegmentales.*

De ahí que las técnicas suprasegmentales analicen el comportamiento de la voz sobre componentes que se mantienen a lo largo de más de un segmento, como son la entonación, el ritmo, la duración.

La prosodia, encargada del estudio de este tipo de características está recibiendo una creciente atención dado que este tipo de características aportan información adicional muy valiosa sobre la identidad del locutor.

2.6.6 Niveles de información

Es posible definir (Reynolds, 2006) (Schultz, 2007) una jerarquía de criterios que el oído humano aplica en la tarea del reconocimiento de las personas a través de su voz. En esa jerarquía en el nivel más alto se utiliza la semántica, dicción, características idiolécticas, pronunciación,... Este nivel distingue el estatus socio-económico, la educación y lugar de nacimiento. En un segundo nivel se recogen características como la prosodia, ritmo, velocidad, entonación, modulación, peculiaridades que vienen determinadas por el carácter y la influencia de las personas cercanas al locutor. En el nivel lingüístico más bajo, se consideran los aspectos puramente acústicos del sonido vinculados a la estructura anatómica del tracto vocal.

Al contrario que las de más alto nivel, las características de bajo nivel son relativamente sencillas de extraer de forma automática. A consecuencia de ello la mayoría de los sistemas ASR concentran su esfuerzo en los componentes espectrales del habla.

Un importante número de sistemas de reconocimiento modelan al locutor a partir de las características espectrales de fragmentos subsegmentales, construyendo *"Modelos de Mezclas Gaussianas"* o *"Modelos de Vectores Soporte",* en la suposición de que estos fragmentos son independientes entre sí (Reynolds, 2002). Este tipo de modelado habitualmente falla al detectar información específica que abarque más de una trama, y por tanto se adaptan a la identificación de locutor sólo utilizando la variabilidad de las características de bajo nivel. Para soslayar este inconveniente la investigación en reconocimiento de locutor ha aplicado esfuerzos a incorporar las características procedentes de la idiosincrasia del individuo (Doddington, 2001).

2.7 Clasificadores

Los clasificadores son máquinas (algoritmos matemáticos implementados en un ordenador) diseñados para extraer patrones a partir de un conjunto de datos.

El problema genérico de clasificación podría enunciarse del siguiente modo:

> *Dado un conjunto de ejemplos x_i, muestras etiquetadas de un espacio de d dimensiones, con una función densidad de probabilidad p(x), donde las etiquetas (y_i) indican la clase a la que pertenece cada muestra y dada la relación entre la muestra y su clase que verifique*

una función de densidad de probabilidad condicionada p(y/x) desconocida; determinar la función $\hat{y}_i = f(x_i)$ que estime de forma óptima la clase de cada muestra.

Cada clasificador emplea una familia específica de funciones diferenciadas por el valor de sus parámetros (w), de donde podemos reformular *f(x)* como *f(x,w),* lo que convierte el objetivo de la clasificación en determinar w que optimiza las discrepancias entre las clases estimadas por el algoritmo y las clases reales conocidas.

Para analizar la calidad de un clasificador se establece una función de coste $C\big(y, f(x,w)\big)$ que calcula un valor asociado a una determinada decisión. A partir de ella se puede evaluar el riesgo que implica la máquina:

$$R(f_w) = \int_{X\times Y} C\big(y, f(x,w)\big) p(x,y)\, dx\, dy \qquad \textbf{Ecuación 9}^{19}$$

y concluir que la optimización buscada se resume es en realidad en minimizar el riesgo.

Como el conjunto de ejemplos es finito y se desconoce la función densidad de probabilidad conjunta, se opta por definir el riesgo empírico:

$$R_{emp}(f_w) = \frac{1}{n}\sum_{i=0}^{n} C\big(y_i, f(x_i, w)\big) \qquad \textbf{Ecuación 10}$$

Un primer enfoque de la solución buscaría minimizar este riesgo (ERM empirical risk minimization). Sin embargo este criterio no resulta óptimo ya que no asegura la minimización del riesgo real de la máquina debido al sesgo que introduce disponer de un número limitado de ejemplos. Por otra parte puede producirse un efecto de sobre entrenamiento bajo el cual la máquina es capaz de clasificar perfectamente los ejemplos pero no así nuevos datos (Solera Ureña, 2011).

Alternativamente se ha propuesto un límite superior del riesgo real denominado riesgo estructural, que se formula:

$$R\big(f_w\big) \le R_{est}\big(f_w\big) = R_{emp}\big(f_w\big) + \Phi(h) \qquad \textbf{Ecuación 11}$$

[19]El dominio de la ecuación es un espacio (n+1) dimensional siendo n la dimensión del vector x, por lo que la integral corresponde a una de la misma complejidad.

donde *h* es una medida de la complejidad de la familia de funciones y Φ define el riesgo asociado a dicha familia, lo que impone una limitación a la complejidad del tipo de funciones que pueden ser utilizadas por el reconocedor. En (Vapnik, 1999) se exponen técnicas para su evaluación.

2.7.1 Support Vector Machine (SVM)

Una máquina de vectores soporte es un clasificador binario que estima el mejor hiperplano separador para dos clases de vectores pertenecientes a un espacio n-dimensional, aplicando el criterio de margen de separación máxima y minimizando el riesgo estructural.

Este algoritmo generaliza el método presentado en (Vapnik and Lerner, 1963) para la resolución de problemas de clasificación de vectores linealmente separables, extendida posteriormente a la resolución del mismo problema cuando los vectores no lo son, mediante la utilización de algoritmos no lineales (Vapnik, 1999)(Burges, 1998).

La formulación de las SVM se basa en el concepto del hiperplano separador óptimo, que puede expresarse en función de los vectores ejemplo propuestos para el entrenamiento. También incorpora aspectos derivados del aprendizaje estadístico que le dotan al algoritmo de una capacidad de generalización superior a la de otros métodos de aprendizaje automático, aportando excelentes resultados al aplicarse a gran variedad de problemas prácticos (Solera Ureña, 2011)(Gómez et al., 2009).

En el problema de la separación lineal considérese un conjunto de *n* vectores (x_i) de un espacio de dimensión *d* y sus respectivas etiquetas (y_i).

$$X = \{x_i \in R^d \,|\, 1 \leq i \leq n\}$$
$$Y = \{y_i \in \{-1, +1\}\}$$

Ecuación 12

Si son linealmente separables existe un hiperplano separador definido por su vector director (w) y su sesgo (b) tal que:

$$w^T x + b = 0$$

Ecuación 13

separa todos ejemplos positivos (con etiqueta +1) de los negativos.

$$y_i = \begin{cases} +1 & f(x_i) = (w^T x_i + b) > 0 \\ -1 & en\ otro\ caso \end{cases}$$

Ecuación 14

de lo que se deduce que:

$$y_i \cdot (w^T x_i + b) \geq M > 0 \quad \forall i < n$$ Ecuación 15

La distancia de cualquier vector al separador viene dada por d y su valor mínimo por el margen δ en la Figura 22.

$$d = \frac{|f(x_i)|}{|w|} \geq \frac{M}{|w|} = \delta$$ Ecuación 16

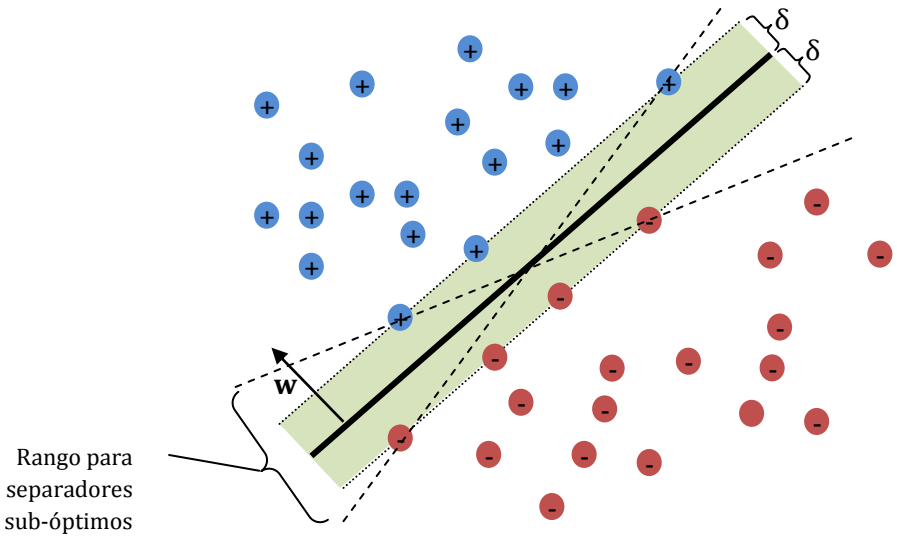


Figura 22: Hiperplano separador óptimo y margen máximo.

En consecuencia el hiperplano separador óptimo será aquel capaz de separar correctamente los ejemplos de cada clase con el mayor margen (δ) posible, o lo que es lo mismo minimizando $\|w\|$

$$min_{w,b} \frac{1}{2}\|w\|$$
$$con: y_i \cdot (w^T x_i + b) > 0 \quad \forall i < n$$ Ecuación 17

lo que lo hace consistente con el criterio de minimización del riesgo estructural, ya que la maximización del margen elimina las posibles soluciones sub-óptimas con lo que se reduce la complejidad estructural del clasificador.

El concepto en que se basa la aplicación del clasificador lineal cuando los ejemplos no son linealmente separables es la transformación virtual del espacio de vectores en un espacio de Hilbert de mayor dimensión (pudiera ser infinita) donde los ejemplos sean linealmente separables (Vapnik, 1999)(Burges, 1998).

$$x_i' = \varphi(x_i)$$ Ecuación 18

Ya que la formulación del SVM puede expresarse únicamente en función del producto escalar de los vectores transformados

$$K(x_i, x_j) = \varphi^T(x_i) \cdot \varphi(x_j)$$ Ecuación 19

El truco es utilizar funciones kernel que evitan tener que transformar explícitamente cada uno de los vectores, extendiendo la eficacia de la solución y mejorando su eficiencia. (Solera Ureña, 2011).

Las implementaciones de las máquinas de vectores soporte suelen implementar las funciones kernel más clásicas: lineal, polinómica, radial, sigmoide…

$$\begin{aligned} Kernel\ lineal \quad & K(x_i, x_j) = x_i \cdot x_j \\ Kernel\ polinómico \quad & K(x_i, x_j) = (x_i \cdot x_j + \beta)^d \\ Kernel\ radial \quad & K(x_i, x_j) = e^{-\frac{\|x_i - x_j\|}{\sigma^2}} \\ Kernel\ sigmoidal \quad & K(x_i, x_j) = tanh(v(x_i \cdot x_j) + c) \end{aligned}$$ Ecuación 20

Este tipo de clasificadores se han convertido en los últimos tiempos en tema más estudiado en el área del aprendizaje automático (Zhang et al., 2013) ya que con sus características SVM ha demostrado poder alcanzar las mejores prestaciones en la resolución de problemas de clasificación y regresión (Jiang et al., 2012) (Kulkarni and Gadhe, 2013), siendo una herramienta de discriminación exitosa en el campo del reconocimiento del locutor (Cumani and Laface, 2012).

2.7.2 Gaussian Mixture Models

El modelo de mezclas gaussianas (Gaussian Mixture Model, GMM) es un tipo de modelado estocástico que representa al sujeto mediante la función densidad de probabilidad de sus vectores, estimada a partir de la de los ejemplos de entrenamiento; dado que la forma esta función de densidad es desconocida se aproxima por la suma ponderada de funciones gaussianas en un espacio de n dimensiones, dada la capacidad que tienen estas mezclas, de representar densidades arbitrarias(Reynolds and Rose, 1995)(Carrero et al., 2008b).

De tal modo que:

$$fdp_u(x) = \sum_{i=1}^{K} w_i \cdot N(x, \mu_i, \Sigma_i) \qquad \text{Ecuación 21}$$

donde K es el número de gaussianas del modelo, w_i es el peso de la gaussiana en la mezcla, μ_i es el vector de medias y $\sum_i$ la matriz de covarianzas. En consecuencia el modelo M vendrá definido por los parámetros necesarios para definir la mezcla, pesos, vectores de medias y matrices de covarianzas:

$$M = \{(w_i, \mu_i, \Sigma_i) \mid i = 1 \dots K\} \qquad \text{Ecuación 22}$$

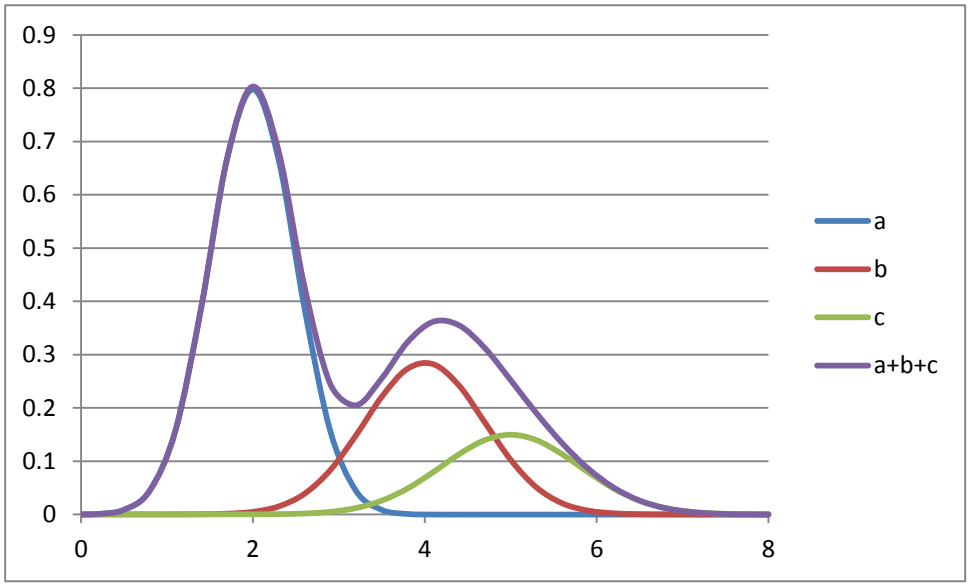


Figura 23: Ejemplo de suma de gaussianas en un espacio unidimensional.

El uso de los GMM para modelar locutores surge de la convicción en que las componentes gaussianas representan recogen algunos patrones espectrales que de forma genérica caracterizan el tracto vocal del locutor(Reynolds and Rose, 1995)

La verificación de locutor puede plantearse como un test entre dos hipótesis, la primera (H_u) que postula que la muestra (S) de sonido ha sido emitida por el usuario, y la segunda (H_w) que postula que el sonido ha sido emitido por cualquier otro. Test para decidir entre estas dos hipótesis calcula el ratio de verosimilitud (likelihood ratio, LR):

$$\mathrm{LR} = \frac{\mathrm{fdp}(S|H_u)}{\mathrm{fdp}(S|H_w)} = \frac{\mathrm{fdp_u}(S)}{\mathrm{fdp_w}(S)} \begin{cases} \geq \mathrm{th} \implies H_u \text{ es cierta} \\ < th \implies H_u \text{ es falsa} \end{cases} \qquad \text{Ecuación 23}$$

donde $fdp(S|H_x)$ es la función densidad de probabilidad para la hipótesis H_x para la muestra *S*. En consecuencia esta técnica obliga a evaluar/estimar la función densidad de probabilidad para el usuario, para las personas que no son el usuario (mundo) y el umbral (*th*) de decisión optimo(Reynolds et al., 2000).

2.8 Extracción

La elección de los parámetros que caracterizan una señal biométrica, es la primera decisión del proceso de diseño de un sistema de reconocimiento, afecta directa y sensiblemente a las prestaciones de éste. Tanto para el caso del reconocimiento del locutor como para el caso del reconocimiento del habla los parámetros (características) que son a día de hoy más populares, debido a la calidad de los resultados que obtienen los sistemas basados en ellos, son los denominados Mel-Frequency Cepstral Coefficients (MFCC) (Beigi, 2011) (Skowronski and Harris, 2002) (Campbell Jr, 1997).

En definitiva la extracción de características consiste en transformar la muestra original en un conjunto de valores (vector de características) que exponga numéricamente los aspectos más relevantes del donante.

2.8.1 Coeficientes cepstrales.

Cepstrum es una de esas transformaciones. Su definición original corresponde a la "*Transformada inversa de Fourier del logaritmo del módulo de la Transformada de Fourier de la señal*".

Históricamente cepstrum tiene su base en el problema general de desconvolucionar dos señales. Originalmente fue propuesta en (Bogert et al., 1963) para detectar la recepción del eco de una señal radar, al observar que el efecto del retraso del eco se aprecia una ondulación en el logaritmo del espectro. La "frecuencia" de esta ondulación es fácilmente establecida calculando el espectro del logaritmo del espectro de la señal, ahí la ondulación aparece como un pico. Ya que la medida de esta "frecuencia" se expresa en unidades de tiempo y para evitar la confusión que puede producir, Bogert buscó un nuevo término para denominar el nuevo dominio obtenido al transformar el de la frecuencia (frequency), y alterando el orden fonético del vocablo en inglés propuso "quefrency" (quefrencia); lo mismo para el espectro (spectrum, representación de la señal en el dominio de la frecuencia) al que en el dominio de la quefrencia llamó cepstrum.

Aunque quefrency y cepstrum son los que han tenido mayor éxito, Bogert realizo el mismo ejercicio para el resto de conceptos comunes en el análisis espectral ahora análisis cepstral (Furui, 2001) (Childers et al., 1977).

- Frequency → Quefrency
- Spectrum → Cepstrum
- Filtering → Liftering
- Phase → Saphe
- Amplitude → Gamnitude
- Harmonic → Rahmonic
- Periode → Repiod

El computo de los coeficientes cepstrales se desarrolla a partir del concepto del *filtrado homomórfico,* el cual consiste en una transformación no lineal utilizada para convertir una señal obtenida como resultado de la convolución de otras dos, en la suma de dos obtenidas de una transformación de estas (Oppenheim and Schafer, 2004).

$$s(nT) = e(nT) * h(nT) \Rightarrow \hat{s}(nT) = \hat{e}(nT) + \hat{h}(nT) \qquad \text{Ecuación 24}$$

Según esto, se puede expresar para una secuencia temporal como*:*

> *"El cuadrado de la transformada Z inversa del logaritmo del cuadrado del módulo de la transformada Z de la secuencia".*

$$\hat{s}(n) = \left(\mathrm{Z}^{-1}(\log(|\mathrm{S}(z)|^2))\right) = \left[\frac{1}{2\pi \mathrm{j}} \oint \log|\mathrm{X}(z)|^2 \, \mathrm{dz}\right]^2 \qquad \text{Ecuación 25}$$

En consecuencia aplicando la transformada Z a la secuencia de la Ecuación 24

$$s(n) = e(n) * h(n) \Rightarrow |S(z)|^2 = |E(z)|^2 \cdot |H(z)|^2 \qquad \text{Ecuación 26}$$

Donde se pueden aplicar logaritmos y obtener

$$\begin{gathered} log|S(z)|^2 = log|E(z)|^2 + log|H(z)|^2 \\ \Rightarrow \hat{S}(z) = \hat{E}(z) + \hat{H}(z) \end{gathered} \qquad \text{Ecuación 27}$$

Y finalmente aplicando la transformada inversa

$$\hat{s}(nT) = \hat{e}(nT) + \hat{h}(nT) \qquad \text{Ecuación 28}$$

Con lo cual, si los dos sumandos están separados es posible aislarlos aplicando un simple "liftering".

A tenor del modelo de producción del habla *"fuente y filtro"*, la voz se compone de una señal de excitación *e(t)* linealmente convolucionada con la respuesta impulsional del filtro que supone el tracto vocal *h(t)*. El proceso de desconvolución cepstral proporciona, pues, una vía para obtener la caracterización de ambos componentes de forma independiente.

2.8.2 Percepción espectral

Muchos experimentos psicológicos han demostrado que la percepción auditiva de los componentes de la frecuencia en el discurso no sigue una escala lineal, por tanto no parece adecuado utilizar una de tal tipo a la hora de reproducir los mecanismos humanos del reconocimiento de la voz. Otras escalas como la musical no incorporan los elementos subjetivos propios de la comunicación humana. (Chakroborty et al., 2007)

En 1937 (Stevens et al., 1937) publica los resultados de experimentos sobre la apreciación de las frecuencias y las diferencias subjetivas entre ellas. La sensación que produce el fundamental y la que produce la diferencia entre dos frecuencias es una función rectilínea del ancho de la membrana basilar en el punto de resonancia; de tal modo que cuando se reduce a la mitad la frecuencia de un tono se reduce también a la mitad la distancia entre el punto de resonancia y el ápice.

Las medidas del tamaño subjetivo de intervalos musicales como las octavas, muestran que los intervalos crecen al mismo tiempo que la frecuencia del punto central del intervalo, excepto para las dos primeras octavas audibles. Proponen una escala basada en estos resultados cuyas unidades subjetivas denominan mels y en correspondencia la escala "escala mel".

No hay una única la escala mel, la más popular es la propuesta en (O'shaughnessy, 1987) establece la siguiente relación entre la frecuencia en Hz y la frecuencia subjetiva en mels:

$$M(f) = 2595 \cdot log_{10}\left(1 + {}^{f}/_{700}\right) \qquad \text{Ecuación 29}$$

Otros autores proponen ecuaciones alternativas:

(Fant, 1967)

$$M(f) = \frac{1000}{log(2)} \cdot log_{10}\left(1 + {}^{f}/_{1000}\right) \qquad \text{Ecuación 30}$$

(Lindsay and Norman, 1977)

$$M(f) = 2410 \cdot log_{10}(0.0016 \cdot f + 1)$$ Ecuación 31

(Zwicker, 1961)

$$M(f) = 100 \cdot \left(13 \cdot atan(0.00076 \cdot f) + 3.5 \cdot \text{atan}\left(\left({}^{f}/_{7500}\right)^{2}\right)\right)$$ Ecuación 32

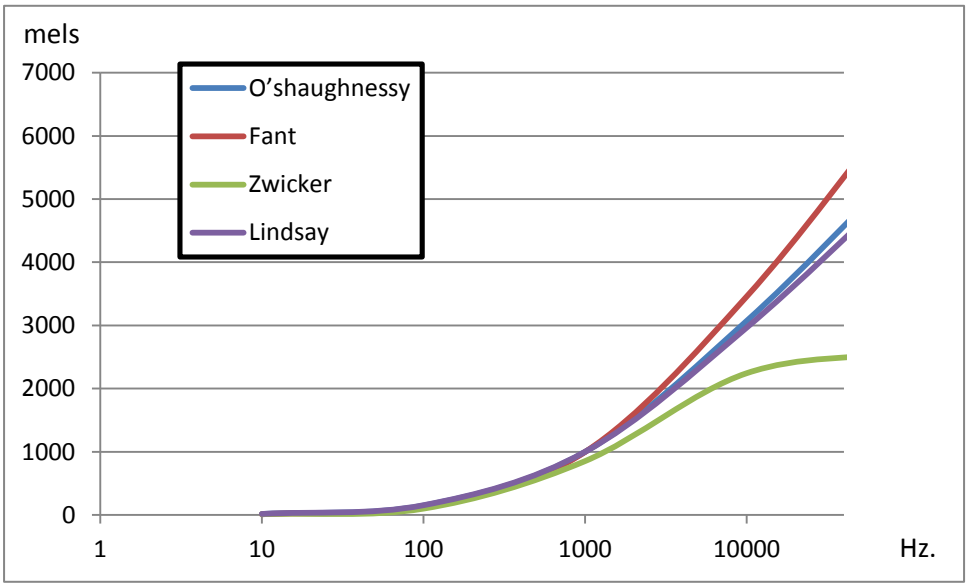


Figura 24: Funciones de conversión escala mel.

Otro hecho fundamental es que también la experimentación sobre humanos demuestra que las frecuencias de un sonido complejo dentro cierto ancho de banda entorno a una determinada frecuencia no puede ser identificadas individualmente, pero sí cuando un componente de ese sonido cae fuera de esa banda. De forma nominal el ancho de banda (ancho de banda crítico) se encuentra entre el 10 y el 20% de la frecuencia central. (Picone, 1993).

$$BW_{critical} = 25 + 75 \cdot \left[1 + 1.4 \cdot \left({}^{f}/_{1000}\right)^{2}\right]^{0.69}$$ Ecuación 33

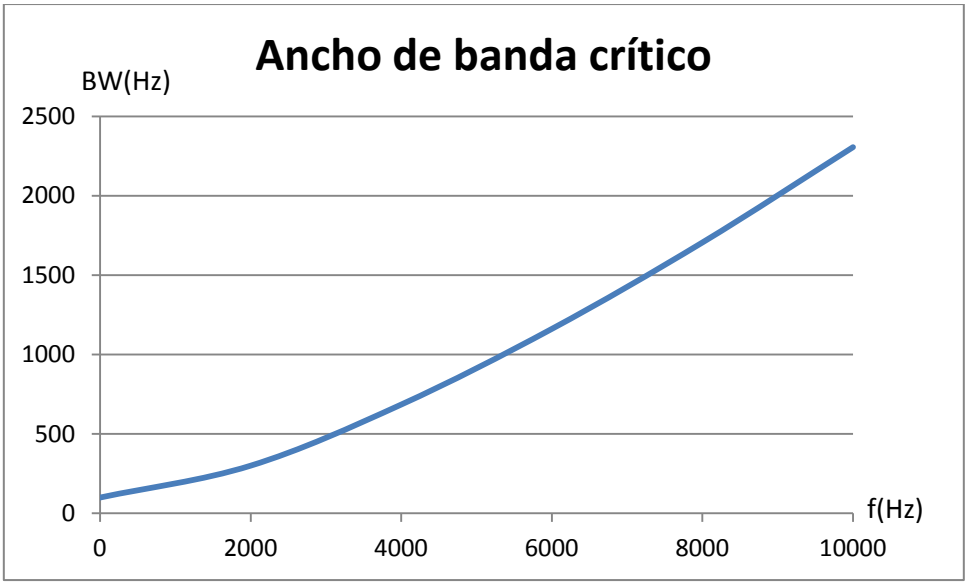


Figura 25: Ancho de banda critico como función de la frecuencia central.

En consecuencia el modelado del sistema auditivo debe ir en la línea de considerar bandas de frecuencias distribuidas no linealmente. Asumiendo que el oído es un buen reconocedor de la voz este enfoque resulta de gran utilidad ya que se aproxima al modo como se realiza la transducción del sonido. Lo que está por demostrar es que el oído pueda ser el mejor reconocedor posible (Chakroborty et al., 2007).

2.8.3 Mel Frequency Cepstral Coefficients (MFCC)

La idea original de los MFCC, es decir, aplicar la DCT (Transformada discreta del coseno) a la compresión logarítmica de la salida de un banco de filtros con un espaciado no lineal de las frecuencias centrales (Skowronski and Harris, 2003) (Oppenheim and Schafer, 2004) y su uso en el procesado del habla corresponde a Bridle y Brown (Bridle and Brown, 1974), pero fue el trabajo de Davis y Mermelstein (Davis and Mermelstein, 1980) el que popularizó este concepto ente los investigadores. A partir de ahí ha sido ampliamente utilizado y con multitud de variantes y mejoras sobre la propuesta inicial. Estas propuestas difieren en el número de filtros, en la forma de estos, en su espaciamiento, en su ancho, en el rango de frecuencias cubierto, o en el número de coeficientes calculados (Ganchev, 2011).

El algoritmo puede ser resumido del siguiente modo: la señal pasa a través de un banco de filtros triangulares uniformemente espaciados en la escala mel, la energía obtenida a la salida de cada filtro es comprimida logarítmicamente y convertida con la DTC en coeficientes cepstrales. Corresponde a revisión del algoritmo de Bogert, donde se considera la componente real de la señal y no su módulo, se tienen en cuenta los

componentes subjetivos de la percepción del habla (escala de mel) y se utiliza la transformada del coseno como alternativa a la transformada Z inversa.

Esta transformación aporta un conjunto de características robustas, elemento imprescindible en un sistema de reconocimiento robusto. Los MFCC modelan la forma en el que el sistema auditivo humano percibe los sonidos y en razón de ello ha sido utilizada como la forma estándar de caracterizar la voz pese a que la estructura de su banco de filtros recoge las características del tracto vocal de forma más efectiva en la zona de las bajas frecuencias (Chakroborty et al., 2007) (Ganchev, 2011).

En base a las consideraciones anteriores Davis y Mermelstein proponen utilizar un banco que es simplemente un conjunto de filtros paso banda dispuestos linealmente a lo largo de la escala mel. Los anchos de banda de son elegidos para corresponder al ancho de banda crítico de la frecuencia central de la banda implementada (Ramos Lara et al., 2012).

El banco de filtros triangulares es, con mucho, el más utilizado en reconocimiento de locutor (Khalifa et al., 2013). La función de transferencia de ese conjunto de filtros viene dada por:

$$H_i(f)\begin{cases} 0 & f < f_{i-1} \\ \dfrac{2\cdot(f-f_{i-1})}{(f_{i+1}-f_{i-1})\cdot(f_i-f_{i-1})} & f_{i-1} \le f \le f_i \\ \dfrac{2\cdot(f_{i+1}-f)}{(f_{i+1}-f_{i-1})\cdot(f_{i+1}-f_i)} & f_i \le f \le f_{i+1} \end{cases}$$

Ecuación 34

Donde, si Q es el número de bancos, f_{min} la frecuencia mínima y f_{max} la frecuencia máxima.

$$\begin{gathered} m_{min} = M(f_{min}) \\ m_{max} = M(f_{max}) \\ \Delta m = \frac{m_{max} - m_{min}}{Q} \\ m_i = i\cdot\Delta m, \quad 0 \le i \le Q \\ f_i = M^{-1}(m_i) \end{gathered}$$

Ecuación 35

A modo de ejemplo, para una frecuencia mínima de 0Hz y una máxima de 8KHz (banda típica telefónica) y proponiendo un banco de 10 filtros, al aplicar la Ecuación 34 se obtienen las gráficas del a Figura 26:

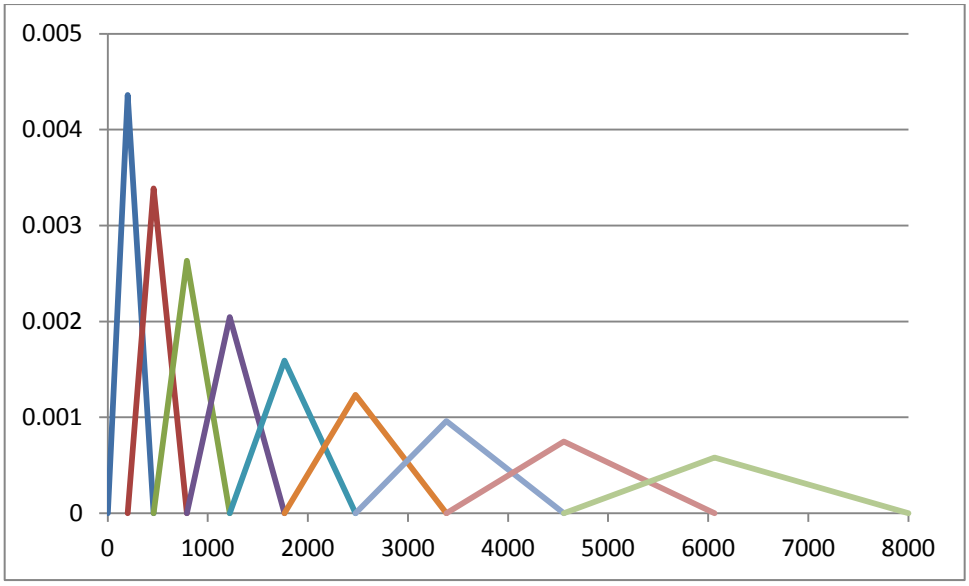


Figura 26: Ejemplo de banco de filtros mel.

2.8.4 Estructura de la extracción de características

Para la obtención de los MFCC lo descrito hasta ahora se completa con un conjunto de pasos previos, repetidos a lo largo de la literatura del estado del arte, para el preprocesado de la señal: preénfasis, fragmentado y enventanado (Ganchev et al., 2005)(Mezghani and O'Shaughnessy, 2005)(Togneri and Pullella, 2011) (Singh et al., 2012a) (Ramaiah and Rao, 2013) (Zergat and Amrouche, 2013)(Kulkarni and Gadhe, 2013).

La secuencia completa de pasos puede verse en el esquema de la Figura 27

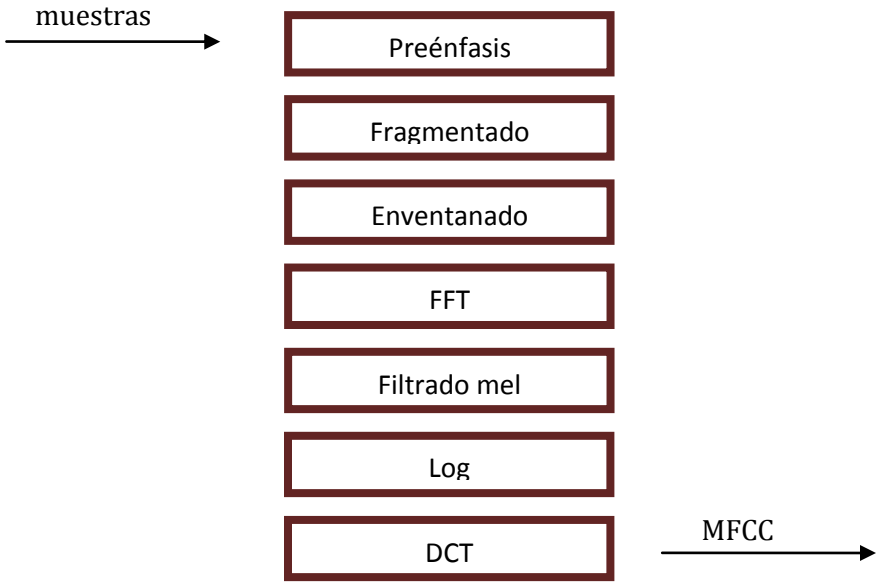


Figura 27: Etapas de la extracción de características.

Preénfasis

Las características del tracto vocal definen el sonido emitido, dichas características quedan evidenciadas en el dominio de la frecuencia por la ubicación de las formantes. Pese a tener información relevante las formantes de alta frecuencia tienen menor amplitud que las de baja (Khalifa et al., 2013). Para compensar esta diferencia se aplica un filtrado del tipo:

$$H(z) = 1 - \alpha \cdot z^{-1} \; con \; 0 \leq \alpha \leq 1 \qquad \text{Ecuación 36}$$

Que en el dominio del tiempo se resume en aplicar a la señal de entrada:

$$x_n = s_n - \alpha \cdot s_{n-1} \qquad \text{Ecuación 37}$$

Fragmentado

Fragmentado es el proceso de trocear la señal de audio en el tiempo. Se utilizan tramas (x) de duración $20ms < d_f < 40ms$, lo que equivale que para una frecuencia de muestreo f_s se deberán obtener tramas de longitud $N = d_f \cdot f_s$[20].

Dos tramas consecutivas estarán separadas un número de muestras $l_h < N$ ($d_h < d_f$) lo que implica que estarán solapadas N-l_h muestras. Los valores típicos están en el entorno de $d_h = 10ms$ (Ramos-Lara et al., 2009).

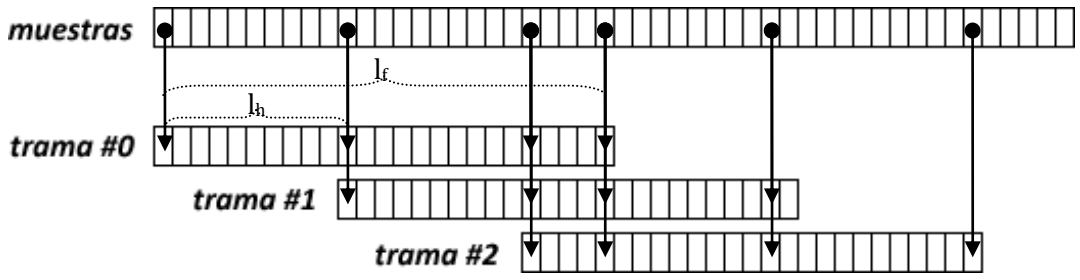


Figura 28: Ejemplo de fragmentación.

Enventanado

El enventanado es preciso para minimizar los efectos de las discontinuidades al comienzo y final de la trama (Kulkarni and Gadhe, 2013). Para ello se utiliza la ventana de Hamming definida como un vector de la misma longitud que la trama (l_f) cuyos componentes toman valores según la Ecuación 38:

[20] d_f: frame duration, f_s: sample frequency, l_h: hop length, d_h hop duration.

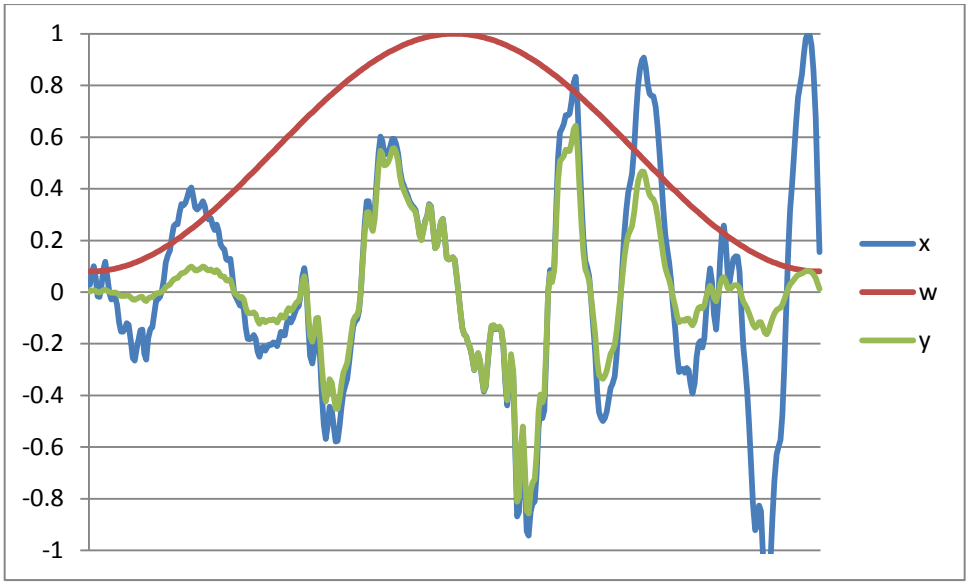


Figura 29: Ejemplo de enventanado.

$$w_n = 0.54 - 0.46\,cos\left(\frac{2\pi n}{l_f - 1}\right) \forall 1 \leq i \leq N$$ Ecuación 38

Así realizar el *enventanado* se reduce a multiplicar componente a componente la trama por la ventana de Haming.

$$y_n = x_n \cdot w_n \quad \forall 1 \leq n \leq N$$ Ecuación 39

Transformada rápida de Fourier (FFT)

La transformada rápida de Fourier es utilizada para convertir la señal en el dominio del tiempo al dominio de la frecuencia, obteniendo así su espectro. En esta aplicación permite convertir la convolución del pulso glotal y la respuesta *impulsional* del tracto vocal en su producto.

$$|Y_k|^2 = \sum_{n=1}^{N} y_n \cdot e^{-\frac{j2\pi nk}{N}} \forall\, 1 \leq k \leq N$$ Ecuación 40

Filtrado mel

Como se ha visto en la sección anterior el filtrado de mel consiste en el cálculo de la potencia obtenida a la salida de cada uno de los Q filtros triangulares del banco.

$$R_i = \sum_{k=1}^{N/2} |Y_k|^2 \cdot M_i(k) \quad 1 \leq i \leq Q$$ Ecuación 41

Log

Simplemente se trata de la aplicación de logaritmos a los resultados del filtrado mel, lo que permite convertir el producto de la excitación y filtro en una suma.

Transformada discreta del coseno(DCT)

Esta transformación convierte el logaritmo del espectro al dominio de la quefrencia. El resultado es el buscado conjunto de los M valores MFCC's que constituye el vector de características.

$$C_m = \sqrt{\frac{2}{Q}} \sum_{i=1}^{Q} \log(R_i) \cdot cos\left[\left(i - \frac{1}{2}\right)\frac{m \cdot \pi}{Q}\right] \quad 1 \leq m \leq M$$ Ecuación 42

2.8.5 Energía y velocidades

Se ha demostrado que la energía contenida en la trama aporta información relevante vinculada a la identidad del locutor (Zetterholm, 2007). La forma habitual de computarla es mediante su logaritmo que reduce la sensibilidad a la presencia de ruido.

$$C_o = log\left(\frac{1}{N}\sum_{n=1}^{N} x_n^2\right)$$ Ecuación 43

Así mismo se ha comprobado que se incrementan las prestaciones del sistema si se agregan las derivadas temporales de los coeficientes estáticos obtenidos (Wong and Sridharan, 2001). Estos componentes conocidos como velocidades o deltas pueden obtenerse mediante(Ramos-Lara et al., 2009):

$$\Delta_m = \frac{\sum_{h=1}^{H} h \cdot (C_{m+h} - C_{m-h})}{2\sum_{h=1}^{H} h^2} \quad 0 \leq m \leq M$$ Ecuación 44

2.8.6 Aceleraciones

En el área del reconocimiento del habla se hace un uso sistemático de las aceleraciones (segundas derivadas de los coeficientes estáticos) para completar el conjunto del vector de características. Sin embargo en los trabajos previos a la realización de la tesis (1.3) se pudo comprobar que su utilización no aporta ninguna mejora estructural en el reconocimiento del locutor. Razón por la cual no van a ser consideradas dentro del presente estudio.

2.9 Arquitectura del reconocedor

Como cualquier otro sistema de reconocimiento el de locutor también involucra dos fases distintas denominadas entrenamiento o enrolamiento y reconocimiento o test. El entrenamiento es el proceso en el que, a partir de muestras de voz, se construyen modelos representativos de un conjunto concreto de locutores denominados usuarios. Reconocimiento es el proceso mediante el cual nuevas muestras de voz pertenecientes o no a los usuarios son comparadas con los modelos obtenidos en el entrenamiento lo que permite determinar si la voz corresponde al usuario pretendido (verificación) o a qué usuario, si alguno, pertenece la voz (identificación) (Tiwari, 2010).

Prácticamente todo el estado del arte utiliza durante el entrenamiento, de una forma u otra, un conjunto de locuciones de referencia ("mundo") para aumentar la robustez y la eficiencia computacional del reconocedor (Prasad et al., 2012a). Las locuciones referencia (ejemplos ajenos) son utilizadas como ejemplos negativos para los clasificadores, o para crear un modelo del mundo (virtualmente un modelo del conjunto de todos los locutores) con el que comparar los resultados de los test.

2.9.1 Entrenamiento

Para un caso típico como es el del análisis subsegmental basado en características espectrales de la voz, es necesario disponer de múltiples ejemplos de fragmentos del habla que fehacientemente haya sido producida por el usuario. La necesidad de múltiples muestras deriva de la necesidad de tener representada toda la variabilidad posible de la voz, que, como ya se ha comentado, aparece tanto en el corto como en el medio como en el largo plazo; por ello se suelen realizar múltiples tomas en diversos días y, dentro de las posibilidades, en las todas las circunstancias en las que tuvieran que realizarse los futuros reconocimientos. Estos ejemplos se denominan **propios.**

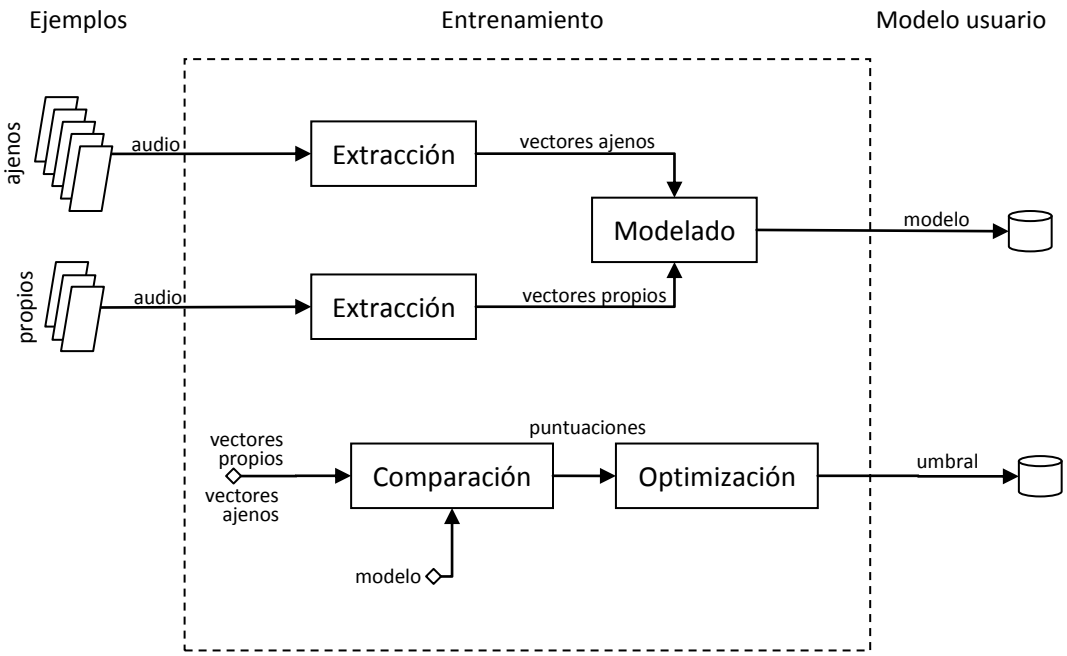


Figura 30: Esquema del proceso de entrenamiento.

También como se ha dicho, es necesario disponer de muestras de audio de otras personas, ejemplos negativos, ejemplos ajenos.

Por cada ejemplo se obtiene una secuencia de tramas de audio, tal y como se ha descrito en la sección 2.8.4, y dado que cada trama da lugar a único vector, a la finalización se dispondrá para cada ejemplo de un número de ellas definido por la Ecuación 45.

$$N_t \approx (d - d_f)/d_h \qquad \textbf{Ecuación 45}$$

donde N_t es número de vectores, *d:* duración del ejemplo, d_f duración de la trama, d_h duración del intervalo.

Para evitar los efectos del sobre entrenamiento, de caja conjunto de vectores (propios y ajenos) se escoge aleatoriamente un subconjunto representativo. Estos dos subconjuntos son los que en definitiva serán utilizados para construir el modelo.

Igualmente, durante el reconocimiento también se dispone del conjunto de vectores construido a partir de la locución del donante, el clasificador ofrecerá un resultado por cada vector, por ello, se ha de fijar la función que transforme ese conjunto de verosimilitudes en una decisión binaria (aceptar o rechazar). La opción es operar las verosimilitudes obteniendo un resultado que valore los ejemplos y calcular un umbral para el resultado de esa operación que minimice la función de coste que se

considere adecuada a la utilización (ver 2.1.7) o como se realizará a lo largo de esta tesis, iguale las tasas de fallo, EER.

2.9.2 Reconocimiento

Teniendo en mente la equivalencia descrita en 2.6.1 entre identificación y verificación la siguiente descripción se centrará sólo en esta última.

En la verificación se dispone de una captura de una muestra de voz de un donante cuya identidad se supone, de esta muestra se obtienen todos los vectores posibles atendiendo a la Ecuación 44. Se utiliza el mismo módulo de extracción que en el entrenamiento.

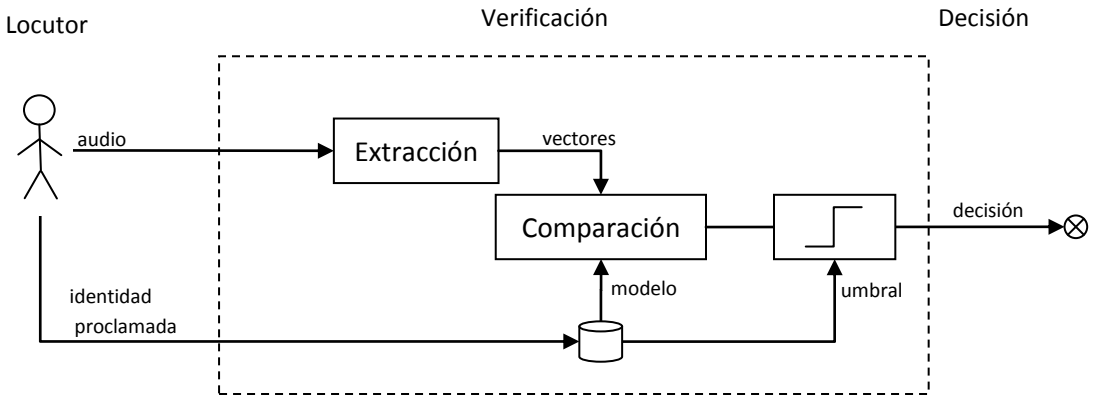


Figura 31: Esquema del proceso de verificación.

Se compara cada vector con el modelo existente para la identidad pretendida y se obtiene para cada uno de ellos un resultado (clase, verosimilitud).

Se opera el conjunto de resultados y se compara con el umbral asociado al modelo obteniendo una decisión sobre la veracidad o no de la identidad.

2.10 Evaluación

En un sistema ideal todos los individuos de la población son poseedores del rasgo que el sistema utiliza, cada patrón biométrico difiere sensiblemente de los del resto de la población, dicho patrón no depende de de las condiciones en las fueron recolectados los datos que se utilizaron en su generación y finalmente el sistema es resistente a las medidas que pueda adoptar cualquier impostor (Phillips et al., 2000); bajo este criterio, la evaluación biométrica cuantifica en qué grado los sistemas verifican estas características.

La medición de las prestaciones de sistemas de verificación difieren sustancialmente de aquellos dedicados a la identificación de las personas, ya que en esta última a medida de rendimiento corresponde al porcentaje de veces en el que la respuesta correcta se encuentra entre las que ofrecen valores de confianza más altos, mientras que para sistemas de verificación son las tasas de falsa aceptación y falso rechazo las que definen las prestaciones del sistema (Phillips et al., 2000).

Un protocolo de evaluación establece cómo se deben seleccionar los datos, cómo se debe ejecutar el test, y cómo se han de medir las prestaciones, lo que supone establecer un marco de trabajo en el cual se definan las métricas y la sistemática en la realización de las mediciones y los criterios de su análisis (Marcel, 2013).

Típicamente se requiere que un organismo independiente realice el diseño de la evaluación, recopile los datos de test y analice los resultados.

El grupo del habla del National Institute of Standards and Technology (NIST) ha venido coordinando las evaluaciones de sistemas e reconocimiento de locutor independientes del texto desde 1996 (Przybocki and Martin, 2004)(Doddington et al., 2000). Proponiendo planes específicos de evaluación de sistemas, conjuntos de datos de evaluación, medidas estándar de error y el foro de participantes, donde discutir abiertamente los éxitos y fracasos de los algoritmos. Las series del NIST para la evaluación del reconocimiento del locutor (SRE) ha permitido difundir el progreso de las prestaciones del estos sistemas.

De tal modo que el NIST dispone de un amplio repertorio de audios para diversos escenarios. Asociado a estos audios ellos existe para cada uno una identificación del locutor que lo ha producido, información que es desconocida por los investigadores, de tal modo que es posible realizar un test ciego cuyos resultados pueden ser remitidos al NIST quien evalúa las prestaciones del sistema en prueba.

2.11 Alternativas

En resumen, el proceso de reconocimiento se puede considerar como una secuencia de cinco etapas Captura, Preprocesado, Extracción, Clasificación, Decisión.

Captura: Dónde se obtienen muestras de voz y se selecciona el material acústico que debe ser utilizado

Preprocesado: Donde se acondiciona el material acústico, para la etapa posterior, realzando las características de interés y atenuando los efectos indeseados del ambiente o los inherentes al procesado digital de la señal.

Extracción: Donde se evalúan los parámetros (características) que mejor caracterizan la señal acústica a los fines del reconocimiento.

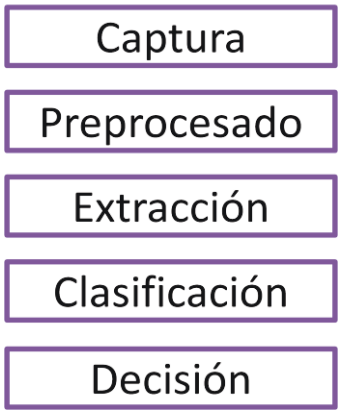


Figura 32: Etapas del reconocimiento.

Clasificación: Donde se compara la caracterización del material acústico con los patrones establecidos para la identidad proclamada, obteniendo una puntuación asocia da a dicha comparación.

Decisión: Donde se determina, en función de la puntación obtenida en el etapa anterior, si la locución es genuina o es impostada.

Bajo esta perspectiva, lo expuesto en las secciones anteriores representa lo más clásico del reconocimiento del locutor, existiendo múltiples alternativas en cada una de las citadas etapas.

A nivel de decisión puede realizarse a partir la normalización previa (Ruiz et al., 2009) (Puente et al., 2009) (Puente et al., 2011b) de las puntuaciones siguiendo algoritmos como:

- Zero normalization

  Z-norm Es una de las normalizaciones más populares en el área del procesado del habla, donde la puntuación normalizada (s') se obtiene a partir de la puntuación original según Ecuación 46 (Reynolds, 1997)

  $$s' = \frac{s - \mu}{\sigma} \qquad \text{Ecuación 46}$$

  siendo μ y σ la media y desviación estándar de las distribuciones de la puntuaciones del mundo (suponiendo que son gaussianas)(Wallace et al., 2012).

- Test normalization

  T-Norm es otro tipo de normalización igualmente popular obtenida utilizando la misma fórmula que en el caso de Z-norm, la diferencia radica en que la estimación de media y desviación se realiza durante la etapa de verificación enfrentando a locución contra modelos de locutores del mundo (Auckenthaler et al., 2000)

- Zero & Test Normalization

  Las técnicas anteriores pueden ser combinadas ejecutando T-norm después de la Z-norm (Machlica, 2012).

- Double Linear normalization

  Propuesta por el doctorando en (Ruiz et al., 2009), pretende una distribución homogénea de las confianzas vinculadas a las puntuaciones en los rangos que se encuentran por encima y por debajo del umbral de decisión, aplicando una transformación lineal distinta para cada uno de ellos.

A nivel de clasificación la literatura referencia la utilización también otras máquinas entre las que destaca:

- Redes Neuronales (ANN)

  Una ANN es una máquina diseñada para modelar la forma en que el cerebro, considerado como un computador paralelo altamente complejo y no lineal, realiza una tarea concreta (Haykin, 1994), empleando de forma masiva interconexiones en componentes computacionales muy simples denominados "neuronas". Sus principales ventajas son su capacidad discriminatoria, una arquitectura flexible que permite la incorporación de información contextual y la escasa vinculación con las hipótesis sobre la distribución estadística de los vectores de datos (Bimbot et al., 2004).

Para la etapa de extracción se han propuesto otros tipos de caracterización de la señal acústica, como entre otros:

- LPCC: linear prediction cepstral coefficients

  LPCC es una técnica popular para la extracción de características. Toma como base la suposición hecha en "Linear Prediction

Coefficients" (LPC) según la cual una muestra puede ser estimada a partir la suma ponderada de las *p* muestras anteriores que, en el dominio de la frecuencia, corresponde a un filtro "sólo polos" que se puede reconocer por procedimientos de análisis cepstral. (Hanilçi and Ertas, 2013)(Dehak et al., 2011).

- LSF: line spectral frequencies

  LSF propuesta por (Itakura, 1975) es un modo alternativo de presentar los coeficientes de predicción lineal (LPC), para el cual el polinomio divisor de la representación espectral del filtro puede ser descompuesto en otros dos polinomios uno simétrico y otro antisimétrico.

- PLP: perceptual linear prediction

  El análisis PLP fue originalmente concebido para suprimir las caracterizaciones dependientes del locutor en los procesos de reconocimiento del habla, pero como colateralmente se observó que proporciona un mecanismo eficaz en el reconocimiento del locutor (zohra Chelali, 2011). Fue propuesto en (Hermansky, 1990), proporciona una mayor aproximación al modo en que el oído percibe la voz al considerar que a ciertas bandas de frecuencias es más sensible que a otras. Critical-band spectral resolution, equal loudness curve y intensity-loudness power law son los conceptos psicofísicos en que se apoya (Jamaati et al., 2008)

Y también algunos criterios de normalización de los valores obtenidos a través de algoritmos como:

- Mean subtraction

  Se aplica con objeto de eliminar los efectos inducidos en el sonido por los canales lineales. Este método sustrae el valor medio de los coeficientes a lo largo de toda la locución (Ramos-Lara et al., 2009).

- Feature warping

  La idea central del feature warping es construir una representación más robusta de la distribución de los coeficientes.

Esto se logra acondicionando el flujo de características a una distribución objetivo (Pelecanos and Sridharan, 2001).

- Short-term Gaussianization

  Gaussianization fue propuesta en primera instancia para estimar distribuciones de alta dimensionalidad y explotar la independencia intrínseca de los datos para reducir la carga que supone esta alta dimensionalidad (Xiang et al., 2002).

- Relative spectral analysis (RASTA)

  Este tipo de procesado del habla fue originalmente propuesto por Hermansky y Morgan para reducir los efectos del ruido convolucinado y aditivo, realzando la voz en ambientes ruidosos. Suprime los componentes espectrales que varían de forma más lenta o más rápida de lo que lo hace típicamente la voz (Hermansky and Morgan, 1994).

- Feature mapping (FM)

  Feature mapping es una técnica de normalización que identifica un conjunto de transformaciones no lineales para la transformación de un espacio de características dependiente del contexto en uno independiente, compensando con ello, los efectos adversos de la variación de los canales (Reynolds, 2003) (Mason et al., 2005).

Para esta etapa los científicos han redescubierto la ventajas que puede proporcionar caracterizar un fragmento de voz, no por un conjunto de vectores sino por uno único de mayor dimensionalidad, los llamados supervectores mayor (Campbell et al., 2006)(Prasad et al., 2012b). Habitualmente supervector indica la combinación de múltiples vectores de pequeña dimensionalidad en uno, donde el quid de la cuestión se encuentra en realizar la caracterización cuando la duración del fragmento es de por sí variable. En esta línea existen múltiples enfoques, como los GMM Supervector en que se utiliza máquinas de vectores soporte para clasificar supervectores construidos a partir de las medias de los modelos GMM (Campbell et al., 2006)

A nivel de preprocesado se utilizan distintas técnicas de filtrado adaptativo, en general para paliar los efectos del ruido.

Pero a nivel de captura poco se ha propuesto más allá de los importantes avances en el campo de los VAD (Voice Activity Detector), que determinan en qué

instantes el audio contiene voz y en cuáles no; de este modo es posible eliminar del proceso estos últimos que carecen de información alguna sobre él locutor, elemento de gran complejidad cuando el ambiente es ruidoso.

3 Planteamiento

Siguiendo el modelo de la Figura 33, mejorar las prestaciones del sistema implicaría actuar en la mejora de alguna de estas etapas.

Mejorar la etapa de clasificación y decisión implica diseñar algoritmos más eficaces, aunque en opinión de este doctorando y dado el estado de la técnica parece difícil pensar en sustanciales mejoras en el reconocedor basándose sólo en estos nuevos algoritmos, sin mejorar previamente la información con la que son alimentados y que es producida en las etapas previas.

Mejorar en la etapa de extracción implica encontrar caracterizaciones con mayor capacidad discriminadora, atendiendo a que describan de forma más precisa el bien tracto vocal y/o los mecanismos psicomotores involucrados en la producción del habla.

Mejorar en la etapa de preprocesado implicaría seleccionar nuevos algoritmos que realcen las manifestaciones de las características propias de la voz de cada locutor.

Estos últimos avances requerirían conocer más en profundidad cómo los elementos característicos de la persona se manifiestan en la voz y cómo el oído las distingue para poder así realizar una emulación más eficaz.

Finalmente mejorar la captura, apunta a seleccionar más adecuadamente el material acústico que será utilizado en el proceso, ignorando en él aquel que por contener información poco relevante añade dispersión en la caracterización de las antedichas manifestaciones.

En el estado del arte presentado en la sección 2 se ha mencionado que los factores idiolécticos, sociolécticos, dialécticos (factores culturales) afectan a la fonación, y por tanto introducen en la señal vocal características que permiten diferenciar a los distintos grupos humanos que pueden ser estudiados en base a estos factores. También afectan a la voz las circunstancias del medioambiente, de la salud o del estado anímico (factores circunstanciales), alteraciones que pueden ser analizadas en procesos diagnósticos. Finalmente, desde el punto de vista del reconocimiento, la voz está definida por las características acústicas del aparato fonador y sesgos vinculados al temperamento (factores individuales).

En el contexto del reconocimiento del locutor con independencia del texto, la distancia intra-clase se ve perjudicada, ya que las variaciones a muy corto plazo,

imprescindibles en la función comunicadora de la voz, se traduce el aumento de la dispersión intrínseca de sus características; pero también existen piezas de la locución que incluyen poca o ninguna información referente a su emisor: el breve silencio que precede a una oclusión, el sonido debido a las turbulencias producidas por una interposición dental o labio dental, etc., que afectan perjudicialmente a las distancias inter-clases.

De los tres grupos de características descritos, las propiedades de interés en el los sistemas de reconocimiento de locutor, son aquellas que son más invariantes en el tiempo. Aquellas que son de variación más o menos lenta afectan negativamente a los resultados ya que aumentan la distancia intra-clase. También existen causas y por tanto propiedades comunes a las voces humanas por el mero hecho de serlas, que disminuyen la distancia inter-clases. Sería interesante eliminar de la señal vocal los efectos debidos tanto a las variaciones de los actuantes como a las características comunes y con ello mejor las tasas de reconocimiento.

Eliminar los efectos producidos por causas generadoras comunes o por las variantes puede no estar a nuestro alcance en este momento y no es objeto de la presente memoria, pero tal vez sí eliminar aquellos segmentos del discurso más afectados por ellas o mejor aún, seleccionar para su tratamiento sólo aquellos segmentos que se encuentran muy afectados por las características individuales. Para ello sería necesario determinar dónde se encuentran las tramas más propensas a producir falsas aceptaciones o falsos rechazos, de forma que sea posible sesgar el proceso de reconocimiento, lo que implicaría no considerar aquellos fragmentos de voz de baja confianza (cercanos al umbral de decisión).

Es en esta dirección donde este trabajo quiere incidir. No pretende por lo tanto ni proponer un nuevo modelo de proceso de reconocimiento, ni una nueva algorítmica de clasificación, ni plantear la definición de nuevas características la voz, ni tan siquiera cómo realzarlas; simplemente busca confirmar que la información discriminatoria no aparece uniformemente distribuida a lo largo del discurso y que es posible predecir los instantes en que su densidad aumenta o disminuye; momentos que se mantienen bajo control en entornos de texto conocido pero que en la actualidad están fuera de él en los independientes del texto. Permitiría así, capturando sólo estos instantes, reducir la dispersión de la información que se debe procesar, tal y como se esquematiza en la Figura 33.

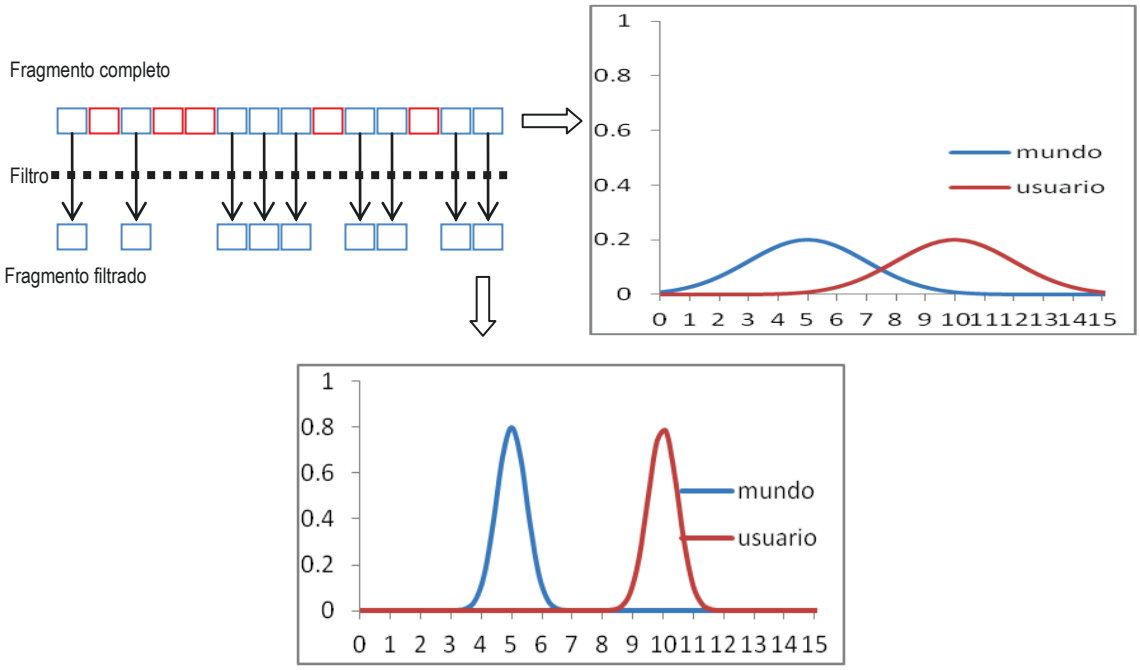


Figura 33: Efecto de la selección del material acústico.

3.1 Definición del problema

Siendo el escenario el de reconocimiento del locutor en independencia del texto, siendo V el conjunto de todos los vectores de características, siendo V_p el conjunto de vectores propios del usuario y siendo V_a el conjunto de vectores ajenos al usuario

$$V_p = \{v_p \in V \mid v_p \in usuario\}$$
$$V_a = \{v_a \in V \mid v_p \in mundo - usuario\}$$

Ecuación 47

se busca diseñar un diezmado cuyo criterio seleccione qué vectores ($V' \subset V$) deben ser utilizados y cuáles no; de tal modo que, si el diezmado es el adecuado, los centroides de los conjuntos de vectores de cada usuario se encuentren más alejados aumentando la distancia inter-clases, y los vectores propios sean menos dispersos reduciendo la distancia intra-clase.

De este modo en una representación del problema en el espacio de vectores, tal y como se formula para las SVM, se persigue eliminar del proceso de comparación aquellos vectores que probablemente se vayan a encontrar próximos a la hipersuperficie separadora, aquellos que aportan mayor nivel de incertidumbre, aumentando el margen de separación entre las clases;

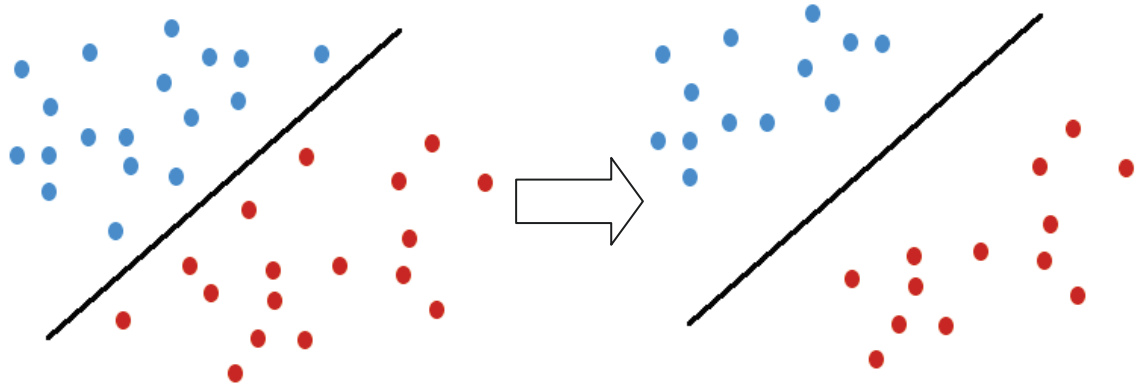


Figura 34: Supresión de los vectores cercanos a la superficie separadora.

o en la formulación estadística, como la realizada para los clasificadores GMM, los que se encuentran en los solapes de las fdp del usuario y del mundo.

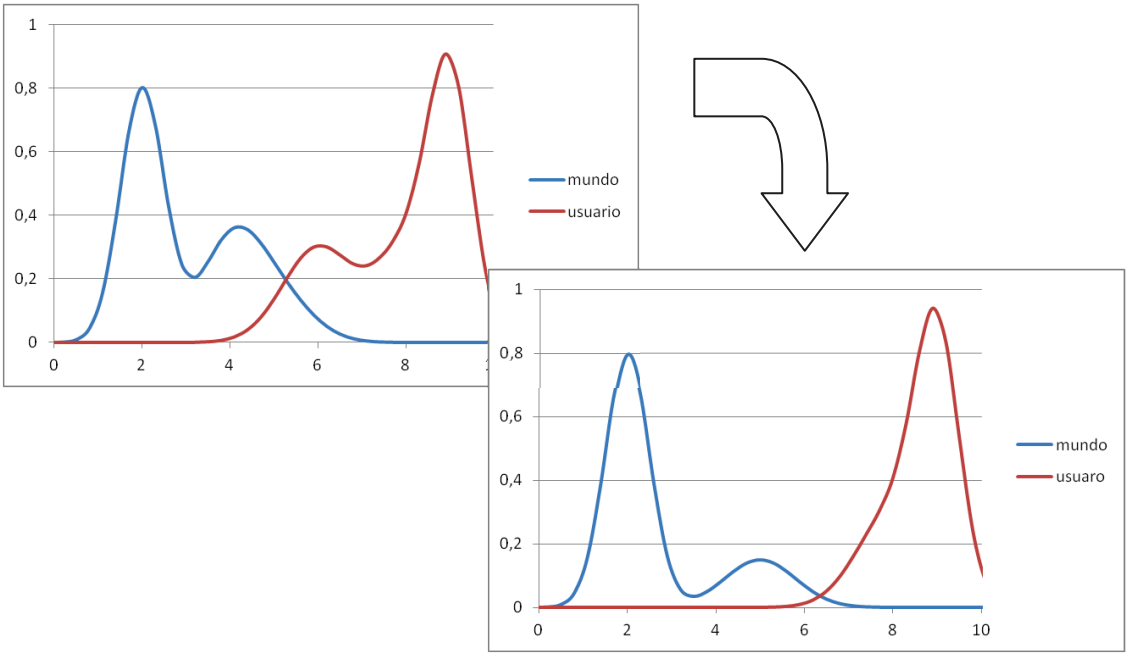


Figura 35: Supresión de los vectores en el solape de las funciones densidad.

Los objetivos de la investigación se traducen en establecer la función de diezmado que defina los vectores seleccionados de modo que mejoren las tasas de reconocimiento,

$$V' = \Pi(V) \qquad \text{Ecuación 48}$$

comprobando que:

> *"Dado un conjunto de audios adecuadamente etiquetado y para los cuales el proceso estándar de reconocimiento*[21] *tiene tasas conocidas (línea base), el proceso de reconocimiento modificado con la función de diezmado, aplicado a las mismas muestras obtiene mejores resultados."*

3.2 Formulación de la propuesta de solución

Como se ha dicho en secciones anteriores, los motores de clasificación implementan la "función de puntuación" (T) cuyo resultado está directamente relacionado con la probabilidad de que el vector dado pertenezca al usuario.

$$T(v) \sim p(v \in V_p) \qquad \text{Ecuación 49}$$

Así el propio clasificador ofrece un criterio de ordenación de los vectores en función de su comportamiento como discriminadores de la identidad del usuario. En consecuencia es de esperar que los segmentos del discurso con mayor capacidad discriminativa sean aquellos que acumulan gran cantidad de vectores de alta puntuación y pocos de los de baja.

Siendo cierto lo anterior, será posible establecer a partir de ello un criterio de valoración del carácter discriminador de los segmentos, criterio que puede expresarse como el valor medio de las puntuaciones de sus vectores.

$$s = \{v_i \in V_p, i = 1 \dots N\}$$
$$\mathrm{T_s}(s) = \frac{1}{N}\sum_{i=1}^{N} T(v_i) \qquad \text{Ecuación 50}$$

Y dado que es posible asociar cada segmento del discurso a un tipo de componente de la locución (unidad de análisis), también lo es valorar la calidad de cada uno según la puntuación media de los vectores de los segmentos con los que está asociado, de tal modo que si se define una función de etiquetación (E) que determine que unidad corresponde a un segmento,

$$u = E(s) \qquad \text{Ecuación 51}$$

[21] Por proceso estándar debe entenderse el descrito en la sección 4.1 Resultados previos.

y se define el conjunto de vectores de la unidad (q) como todos los vectores que pertenecen a segmentos asociados con ella,

$$S_u = \{s \mid E(s) = u\}$$
$$q_u = \{v \in s \mid \forall s \in S_u\}$$

Ecuación 52

es posible definir una valoración para cada unidad (T_u)

$$T_u(u) = \frac{1}{|q_u|}\sum_{q_u} T(v_i)$$

Ecuación 53

La ecuación anterior ofrece un método de ordenación del conjunto de unidades, y de su clasificación en aquellas que tienen un buen comportamiento discriminador (mejores, U^+) y aquellas que no lo tienen bueno (peores, U^-). Con ello construir una agrupación de segmentos distinguiendo los "mejores" de los "peores" para cada usuario según su capacidad discriminadora a partir de la definición de los límites (l^-, l^+) de las puntuaciones de los segmentos para pertenecer a cada grupo.

$$U^- = \{u \mid T_u(u) > l^-\}$$
$$U^+ = \{u \mid T_u(u) < l^+\}$$

Ecuación 54

A partir de aquí es posible establecer dos algoritmos de diezmado de los fragmentos, uno por exclusión de los peores segmentos (EPS) y otro por la selección de los mejores (SMS). De modo que considerando que un fragmento (f) es una secuencia de segmentos puede obtenerse un nuevo fragmento ya diezmado con:

$$f = \{s\}$$
$$EPS: f^- = \{s \mid s \in f \wedge E(s) \notin U^-\}$$
$$SMS: f^+ = \{s \mid s \in f \wedge E(s) \in U^+\}$$

Ecuación 55

3.3 Propuesta del algoritmo

Resumiendo la formulación matemática presentada en la sección anterior, es preciso filtrar el audio eliminando aquellos segmentos correspondientes a las unidades de análisis con menor densidad de información discriminadora; lo que se traduce en eliminar del proceso de clasificación los vectores asociados a ellas.

La determinación de cuáles son estas unidades se realiza en la fase de entrenamiento, para ello se ha de obtener la puntuación de cada uno de los vectores obtenidos audio de entrenamiento, enfrentándolos al modelo de usuario, y promediar todos aquellos asociados a cada una de las unidades de análisis identificadas en el discurso (Ecuación 53).

Establecida la valoración de cada unidad, es posible ordenar el conjunto de unidades, queda por establecer que fracción de este conjunto (factor de filtrado) es el elegido.

A partir de aquí es necesario estimar el umbral de decisión utilizando los fragmentos de audio de entrenamiento una vez excluidos los vectores correspondientes a la unidades no seleccionadas, optimizando la función de coste.

Finalmente los modelos específicos de cada locutor quedan definidos por el modelo propiamente dicho creado por la máquina de clasificación, el listado de las unidades seleccionadas y el umbral de decisión.

3.4 Diseño del proceso

Incorporar la solución propuesta en la sección anterior a un sistema de reconocimiento de locutores implica una leve adaptación tanto del proceso de entrenamiento como el de reconocimiento, esquematizados anteriormente en Figura 30 y Figura 31 a los modelos que aquí se amplían en Figura 36 y Figura 37 respectivamente, en los que se modifica, como se ha propuesto previamente, tan sólo la etapa de captura.

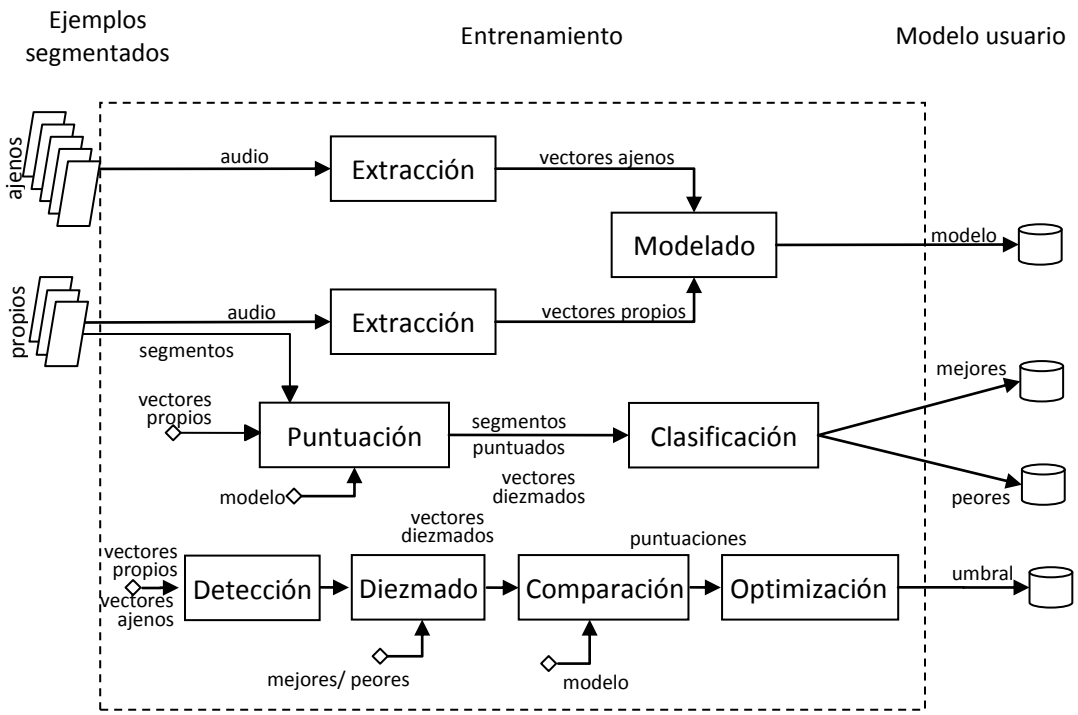


Figura 36: Esquema del nuevo proceso de entrenamiento.

La primera modificación se refiere a los requisitos de los audios de entrada, estos deben estar etiquetados señalando el inicio y final de cada uno de los segmentos. Al igual que en proceso estándar, a partir de los ejemplos propios y ajenos se obtiene el modelo del usuario; modelo que será utilizado para reprocesar los vectores propios puntuándolos y con ello obtener la valoración de los segmentos que los contienen, valoraciones que permitirán identificar las peores unidades o alternativamente aquellas considerados de mejor comportamiento.

Finalmente, como es obvio, la determinación del umbral de decisión debe realizarse a partir de los vectores (propios y ajenos) sólo después de haber sido diezmados.

Igualmente el proceso de verificación se ve modificado, y al mismo se añade la correspondiente etapa de diezmado.

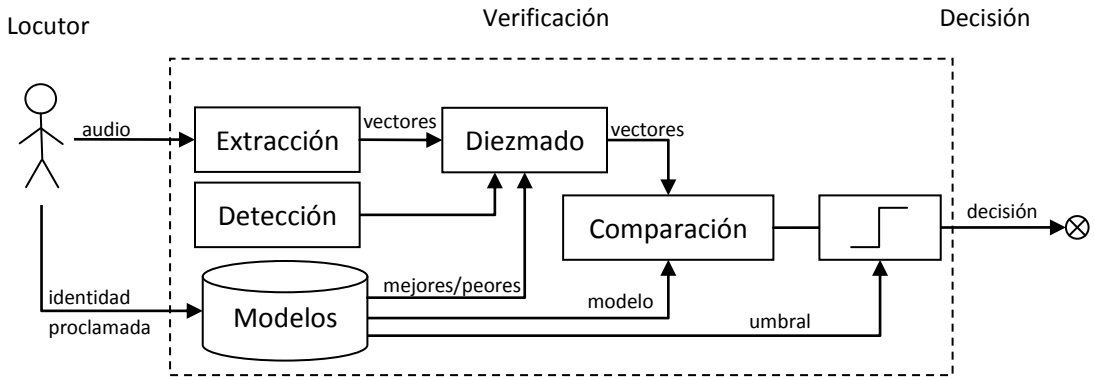


Figura 37: Esquema del nuevo proceso de verificación.

4 Metodología de la investigación

4.1 Resultados previos

Durante el desarrollo del proyecto PIBES (ver 1.3 Trabajos previos) se comprobó la eficacia del proceso de reconocimiento habitual en la literatura del estado de la técnica. Dicho proceso recoge las conclusiones de las necesidades de computación relatadas hasta aquí (Ramos-Lara et al., 2013).

La extracción de características, una vez obtenido el audio se puede resumir de la siguiente forma:

- El audio (en formato pcm) con frecuencia de muestreo conocida (f_s) es fragmentado en tramas de duración predeterminada (d_f).
- La longitud de las tramas es de $N = d_f \cdot f_s$ muestras
- El intervalo entre dos tramas consecutivas está también predeterminado (d_h), y por tanto el número de muestras del salto es $l_h = d_h \cdot f_s < N$.
- Cada trama se preenfatiza con un coeficiente predeterminado (α) (ver Ecuación 37)
- Se *enventana* (Ecuación 39).
- Se obtiene la Transformada de Fourier (Ecuación 40).
- Se aplican los filtros de mel con Q de bancos y unas frecuencias mínima y máxima f_{min} y f_{max} (Ecuación 34).
- Se obtiene la transformada del coseno del logaritmo, calculando M coeficientes cepstrales (Ecuación 40).
- Se calcula el logaritmo de la energía media de la trama (Ecuación 43).
- Se calculan las velocidades de las *M+1* características (Ecuación 44)
- Se reúnen energía, coeficientes y deltas en un solo vector.
- Se normalizan los coeficientes con el algoritmo "mean substraction"
- Para cada usuario

- Se seleccionan (*aleatoriamente*) N_p vectores propios entre todos los obtenidos de los audios seleccionados para entrenamiento
- Se seleccionan (*aleatoriamente*) N_a vectores ajenos entre todos los obtenidos de los archivos seleccionados para entrenamiento.
- Se entrena un clasificador con ambos conjuntos de vectores, y se obtiene el modelo de voz del usuario.

- Para cada fragmento de audio elegido para el test:
  - Se toman todos los vectores producidos.
  - Se utiliza el clasificador para comparar cada uno con el modelo.
  - Se cuenta el n_a número de vectores aceptados.
  - Se cuenta el n_t número total de vectores.
  - Se obtiene la puntuación del fragmento $p = \frac{n_a}{n_t}$.
- Con el conjunto de puntuaciones y el conocimiento de la clase de cada fragmento se determina el EER.

En coincidencia con el estado del arte, durante los trabajos previos se estableció que la mejor parametrización de la extracción de características corresponde a:

f_s	Frecuencia de muestreo	16 KHz
d_f	Duración de la trama	25 ms
d_h	Duración del salto	10 ms
α	Coeficiente de preénfasis	0.97
Q	Número de filtros del banco	26
f_{min}	Frecuencia mínima	0
f_{max}	Frecuencia máxima	8KHz
M	Número de coeficientes cepstrales	12
N_p	Número de vectores propios	4.000
N_a	Número de vectores ajenos	4.000

Tabla 1: Parametrización del extractor de características.

Las variaciones probadas sobre esta configuración cuando mostraban mejoras en los resultados, estos eran coyunturales y no estructurales.

4.2 Proceso experimental

De lo propuesto en la hasta aquí se sigue que los pasos a dar para la demostración de la hipótesis deben ser los siguientes

- Seleccionar los audios a utilizar en la experimentación (base de datos, véase 4.2.1)
- Etiquetar fragmentos de audio con el nombre del locutor (véase 4.2.2).
- Determinar las tasas correspondientes la línea base (véase 4.2.3).
- Proponer la función de diezmado (véase 4.2.4).
- Determinar las tasas tras incorporar dicha función al proceso (véase 4.2.5).
- Evaluar los resultados (véase 4.2.6).

4.2.1 Selección de la base de datos

Para la selección de la base de datos a utilizar durante el desarrollo de los experimentos vinculados a esta tesis se estableció el siguiente conjunto de requisitos.

1. El audio debe ser limpio, carente de ruido significativo.
2. Las locuciones deben ser naturales evitando la lectura de textos o la enunciación de palabras separadas.
3. Las locuciones deben ser independientes, sin que se produzcan varias de forma simultánea.
4. El número de locutores debe ser amplio y con cierto nivel de equilibrio de audios por género.
5. Los locutores deben ser perfectamente identificables.
6. Un número razonable de locutores debe tener un alto grado de participación.
7. Las locuciones deben haberse realizado en castellano.
8. Los audios deben corresponder a muestras obtenidas de un locutor en un amplio periodo de tiempo.

9. El conjunto de locuciones debe subdividirse en el subconjunto de locuciones de entrenamiento y en el de locuciones de test.

4.2.2 Etiquetación del audio

Debe definirse de forma clara cuáles son los criterios de etiquetación del audio a la hora de identificar claramente qué locutor es el propietario de cada fragmento y donde comienza y acaba cada una de los segmentos con indicación de la unidad a la que corresponden.

4.2.3 Establecimiento de la línea base

A fin de poder reconocer las beneficios que la propuesta aporta, deberá definirse una plataforma base que soporte un proceso predefinido estándar. Plataforma que aplicada a los fragmentos de la base de datos proporcionará unos resultados en forma de EER y AUC.

4.2.4 Función de diezmado

Deberá proponerse como serán identificadas las unidades de análisis y como los segmentos asociados a ellas se eliminan en el proceso de reconocimiento.

4.2.5 Aplicación del diezmado

Tras modificar la plataforma de reconocimiento incorporando la función de diezmado (Figura 37), se aplicará a los fragmentos de la base de datos de forma equivalente a lo realizado para el establecimiento de la línea base.

4.2.6 Evaluación de resultados

Finalmente se compararán los resultados obtenidos en la línea base con aquellos alcanzado tras la aplicación del diezmado, examinando las diferencias en términos de EER y AUC.

4.3 Base de datos

La investigación de las bases de datos de habla a las que el autor ha podido acceder ofreció el siguiente resultado.

BANCA:

BANCA es un proyecto Europeo cuyo objetivo es desarrollar e implementar un control de acceso mejorado que combina sistemas clásicos de seguridad con un esquemas de reconocimiento multimodal (cara y voz) (Bailly-Bailliére et al., 2003),

proyecto para el que fue necesario recopilar una base de datos que contemplara diversos escenarios realistas, con variedad de prestaciones en los sensores (2 cámaras y 2 micrófonos). 52 sujetos, 26 hombres y 26 mujeres, en cuatro idiomas distintos (inglés, francés, italiano y español).

Por cada idioma los 52 sujetos grabaron simultáneamente con los dos micrófonos en 4 sesiones distintas en cada uno de los tres ambientes acústicos predefinidos (controlado, degradado y adverso). La señal de los micrófonos fue capturada con 12 y 16 bits a una frecuencia de 32KHz. En cada sesión el donante lee algunos dígitos, un nombre propio, una dirección y una fecha.

BioSec

BioSec baseline corpus fue recopilada dentro del "FP6 EU BioSec Integrated Project". Se trata de una base de datos multimodal que incluye huellas dactilares capturadas con tres sensores distintos, imágenes frontales del rostro capturadas con una webcam, muestras de iris, voz capturada en proximidad con unos microcascos y en distancia con un micrófono de webcam(Fierrez et al., 2007). Contiene muestras multimodales reales de 200 personas capturadas en dos sesiones.

Con respecto al rasgo de la voz, locuciones de un número de ocho dígitos específico para el usuario y tres locuciones de tres supuestos impostores locutando el mismo número, en todos los casos tanto en inglés como en español, lo que hace que el total de 11.200 grabaciones.

MicroAES[22]

La base de datos "ATLAS Spanish Microphone Database" (MICROAES) ha sido recolectada en español por Applied Technologies on Language and Speech, S.L. Se trata de una extensa base de datos de habla española; una de las más completas en este idioma con un elevado número de muestras por sujeto.

Incluye grabaciones microfónicas de 300 locutores diferentes de seis áreas dialectales. Recoge la lectura de 450 párrafos en la que se ha cuidado la ocurrencia de alófonos. Constituye un total de 30h de grabación.

La señal hay sido muestreadas a 16 KHz y con 16 bits. Incluye la transcripción ortográfica y marcas que dividen el audio en fragmentos de menos de diez segundos.

XM2VTS

[22] http://www.elda.org

XM2VTS database (Bengio et al., 2001) contiene video y voz sincronizadas de 295 personas grabadas en cuatros sesiones distribuidas a lo largo de cuatro meses, en cada sesión se realizaron dos grabaciones en las que se recitaba una frase. De los 295 donantes fueron divididos en 200 usuarios, 25 como mundo y 70 impostores. Las locuciones de esta base de datos fueron grabadas en inglés.

General

Las bases orientadas al reconocimiento del locutor no suelen aportar una transcripción del discurso y mucho menos su etiquetado fonético. Si bien sería posible realizar una transcripción manual, las locuciones se suelen limitar a la lectura de pequeñas frases, a la enumeración de dígitos o comandos, que por un lado están lejos de ser habla natural y por otro limitan mucho la variedad de segmentos.

Las bases de datos pensadas para la investigación en el área del reconocimiento del habla si disponen de esta transcripción, pero sin embargo no existe control alguno sobre los locutores, e identificarlos auditivamente resultaría una labor muy poco eficaz.

Ante la coyuntura se optó por construir un corpus ad hoc al objetivo. Se buscó un programa de televisión ya que ello permitiría identificar visualmente a los locutores en el momento en que estuvieran hablando. El programa debería ser de debate, ya que este tipo de programas aporta una cantidad importante de locutores distintos, todos ellos haciendo uso de una locución natural y nunca forzada. Sin embargo este tipo de programas suele estar caracterizado por la controversia, la réplica no organizada y la simultaneidad de los discursos. Por ello se seleccionó entre las opciones uno que mantiene cierto control en las intervenciones y donde una vez eliminados las colisiones y los pocos instantes en los que aparece ruido de fondo, todavía se dispone de un volumen de datos apreciable.

La fuente de información fueron los archivos “podcast” (sesiones) publicados en la página web de la cadena, archivos seleccionados aleatoriamente entre los emitidos en los años 2010, 2011 y 2012.

4.4 Corpus

Se recopilaron doce archivos (sesiones) casi todos de entorno a 90 min de duración (ver Tabla 2), lo que constituye un total de casi 18h de audio bruto.

Sesión	Duración
s100922	1h 28"11'
s101104	1h 33"50'

s101223	1h 25"41'
s111102	0h 49"15'
s111123	1h 29"57'
s111130	1h 30"08'
s111221	1h 30"23'
s120118	1h 30"49'
s120125	1h 31"03'
s120201	1h 30"59'
s120208	1h 30"32'
s120215	1h 32"55'

Tabla 2: Sesiones y duraciones.

Tomando como criterio de las pausas realizadas por los diversos locutores se trocearon dichos archivos en fragmentos, eliminando aquellos que no correspondían a un audio suficientemente limpio. Se obtuvieron 12673 fragmentos.

Se procedió a etiquetar los fragmentos con un identificador del locutor que los había producido. Como resultado se identificaron 99 locutores distintos con muy diversa participación tanto en número de fragmentos como en duración total de sus intervenciones a cada uno de los cuales se asigno un código de identificación formado por dos o tres letras.

La distribución del número de fragmentos por sesión y locutor se muestra en la Tabla 3.

LOCUTOR	s100922	s101104	s101223	s111102	s111123	s111130	s111221	s120118	s120125	s120201	s120208	s120215	TOTAL
AA		167											167
AB	1												1
AC		6											6
AD			3										3
AEp		2											2
AEr	39		168		116					121		50	494
AEx						182				182			364
AGa					34								34
AGo			3										3
AGu									4				4
AGz												266	266
AH						154							154
AI			3										3
AL			190	95			55	88		167			595

LOCUTOR	s100922	s101104	s101223	s111102	s111123	s111130	s111221	s120118	s120125	s120201	s120208	s120215	TOTAL
AMa					24								24
AMo			2										2
AN		18											18
AO		192											192
APa										5			5
APr		2		15						5	2	16	40
ARa		2											2
ARz									9				9
BG										11			11
BP					6								6
CC		299											299
CCh										17			17
CG				110									110
CH						52		54			51		157
CL											151		151
CMe	1	4	6									10	21
CMo								9					9
DM									6				6
EE		407					71						478
EEk				145		197			52			104	498
EG		5											5
EP				87		177				152			416
FA										10			10
FB									27				27
FC										42			42
FM		225	204										429
FO			200				58	85			141	106	590
IA			3										3
IC					95			51			147		293
IF	1	2	9									2	14
IL					89								89
IM				105									105
J2		14	5										19
J3						8							8
J4									112				112
JAn									40				40
JAr				4									4
JBa			4										4

LOCUTOR	s100922	s101104	s101223	s111102	s111123	s111130	s111221	s120118	s120125	s120201	s120208	s120215	TOTAL
JBo		4											4
JCa											98		98
JCr				106									106
JE			6										6
JF	67												67
JG	71		179		105					143		86	584
JJ									7				7
JL		8											8
JM		283			196			42			100		621
JMa		8											8
JMo					14								14
JRo			3		9							4	16
JRz	4	20		5	23	19							71
JS	92											301	393
JSa										2			2
JTa									3				3
JTo								57					57
JV												75	75
JW										6			6
JZ	71			112			76		57				316
LA		196			146		59	58			128		587
LG									13				13
LM		5											5
M2	3	11				14							28
MA				7	289								296
MB										2			2
MC	40	183	100	85	131		41	57	46	83	79	79	924
MCh			2										2
MN										58			58
MP									23				23
MR		7		25	26					21	6	4	89
MS			194									88	282
NF										3			3
NR	54					175	81						310
PL		4	5										9
RJ				5		284							289
RN		167											167
SF				37									37

LOCUTOR	s100922	s101104	s101223	s111102	s111123	s111130	s111221	s120118	s120125	s120201	s120208	s120215	TOTAL
SR												6	6
SS			3									4	7
TG			8								1		9
TJ	1										2		3
UH			9	2	23						10		44
UM				3									3
VG		5											5
VL				70		172				175			417
VP					163				69				232
TOTAL	**445**	**2246**	**1309**	**1018**	**1489**	**1434**	**441**	**501**	**468**	**1205**	**916**	**1201**	**12673**

Tabla 3: Número de fragmentos por locutor y sesión.

La relación de los locutores y la duración en segundos de los fragmentos obtenidos en cada sesión se muestran en la Tabla 4

LOCUTOR	s100922	s101104	s101223	s111102	s111123	s111130	s111221	s120118	s120125	s120201	s120208	s120215	TOTAL
AA		235											235
AB	25												25
AC		8											8
AD			14										14
AEp		9											9
AEr	366		415		277					367		196	1620
AEx						377				451			828
AGa					98								98
AGo			11										11
AGu									20				20
AGz												826	826
AH						342							342
AI			10										10
AL			446	203			324	388		474			1836
AMa					53								53
AMo			7										7
AN		34											34
AO		385											385

LOCUTOR	s100922	s101104	s101223	s111102	s111123	s111130	s111221	s120118	s120125	s120201	s120208	s120215	TOTAL
APa										14			14
APr		6		34						18	7	62	127
ARa		3											3
ARz									81				81
BG										25			25
BP					20								20
CC		556											556
CCh										41			41
CG				215									215
CH						182		277			241		700
CL											606		606
CMe	4	4	12									15	35
CMo								42					42
DM									35				35
EE		494					353						847
EEk				248		461			315			335	1359
EG		12											12
EP				221		464				510			1195
FA										31			31
FB									189				189
FC										131			131
FM		584	630										1213
FO			412				321	401			413	339	1886
IA			10										10
IC					329			280			504		1113
IF	3	12	15									9	39
IL					291								291
IM				289									289
J2		18	9										26
J3						16							16
J4									457				457
JAn									194				194
JAr				5									5
JBa			9										9
JBo		14											14
JCa											352		352
JCr				242									242
JE			11										11

LOCUTOR	s100922	s101104	s101223	s111102	s111123	s111130	s111221	s120118	s120125	s120201	s120208	s120215	TOTAL
JF	573												573
JG	555		523		336					454		312	2179
JJ									24				24
JL		10											10
JM		479			477			206			288		1449
JMa		10											10
JMo					32								32
JRo			22		23							12	57
JRz	61	25		7	40	23							156
JS	728											763	1491
JSa										7			7
JTa									17				17
JTo								277					277
JV												272	272
JW										41			41
JZ	690			230			445		315				1680
LA		382			358		311	269			375		1695
LG									74				74
LM		5											5
M2	19	21				36							75
MA				14	558								572
MB										10			10
MC	301	342	263	230	383		209	281	221	284	290	286	3091
MCh			8										8
MN										201			201
MP									154				154
MR		15		31	49					59	26	20	201
MS			463									288	752
NF										13			13
NR	454					493	386						1333
PL		12	9										20
RJ				10		618							628
RN		317											317
SF				125									125
SR												15	15
SS			10									8	18
TG			14								3		16
TJ	4										4		8

LOCUTOR	s100922	s101104	s101223	s111102	s111123	s111130	s111221	s120118	s120125	s120201	s120208	s120215	TOTAL
UH			29	9	38						18		94
UM				6									6
VG		11											11
VL				133		373				505			1011
VP					370				346				716
TOTAL	3781	4003	3351	2253	3731	3385	2348	2420	2443	3636	3126	3759	38238

Tabla 4: Intervención total por locutor y sesión.

La Figura 38 resume gráficamente los datos anteriores mostrando el número de locutores agrupados en bandas de tiempos de participación total. Se aprecia la existencia de un considerable número de participantes esporádicos con menos de 200 segundos, la participación destacada de la moderadora del programa con un total superior a los 3000 segundos, pero principalmente la existencia de un grupo, los contertulios habituales del programa en cantidad aceptable, con una abundante participación (por encima de los mil segundos).

Como resultado se obtuvo un total de 12.673 fragmentos de diversa longitud. Con valores *mínimo = 0,1s; máximo= 30,37s; media= 3,01s; desviación estándar = 2,36s; mediana = 2,44s;* y *moda = 1,5s* La Figura 39 muestra la estimación de la función densidad probabilidad de la duración de dichos fragmentos.

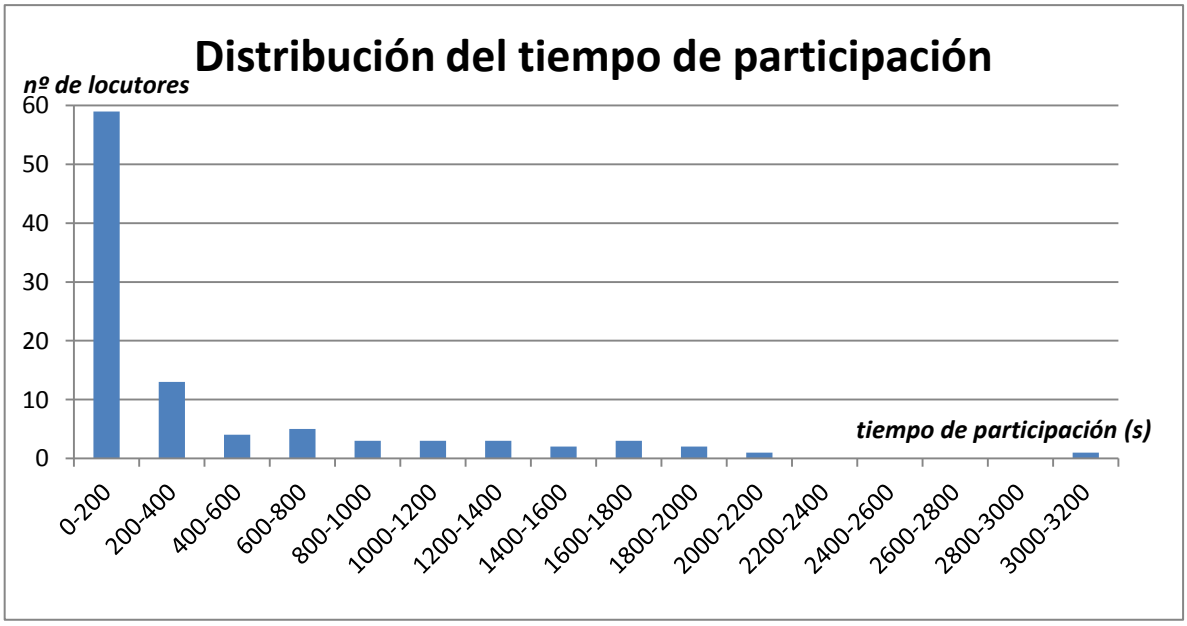


Figura 38: Distribución de la participación en la base de datos.

En esta etiquetación desecharon los fragmentos con sonido de baja calidad, aquellos que mezclan voz con sintonía, videos, etc. o que correspondían a varias locuciones simultáneas; con ello se redujo el audio útil a 10h 37min.

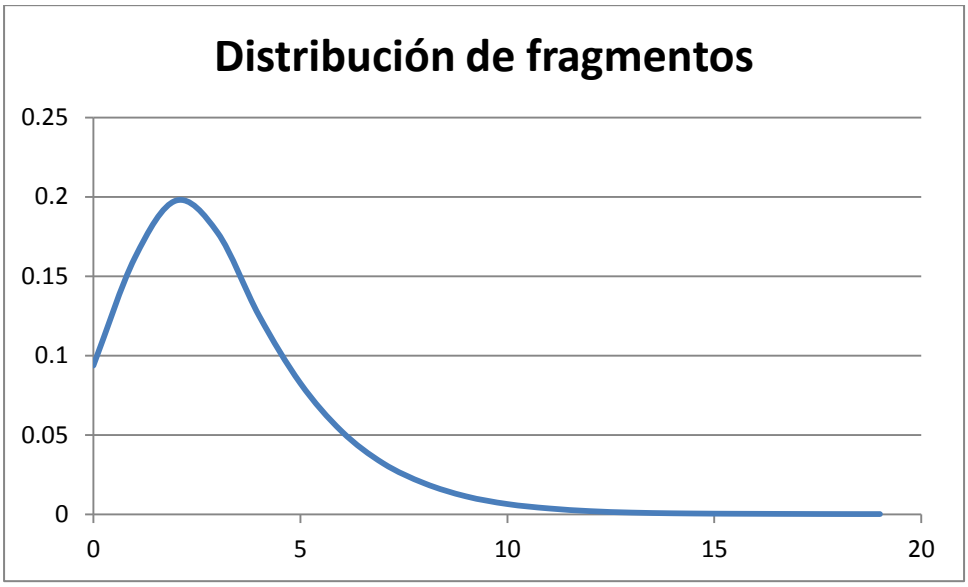


Figura 39: Distribución de la participación en la base de datos.

En los archivos recogidos se etiquetaron fragmentos de discurso con los identificadores de los locutores. Como criterio de fragmentado se utilizaron las pausas realizadas por el locutor, que en la mayoría de los casos corresponden a las dedicadas a la inspiración y en un número menor las utilizadas para enfatizar la expresión.

4.5 Estructuración del corpus

Del conjunto de locutores fue necesario seleccionar cuales formarían el conjunto de usuarios y cuales pasarían a formar parte del conjunto "mundo". También resultaba necesario establecer el criterio por el cual se dictamine que fragmentos serán utilizados en el entrenamiento del sistema.

Por ello se consideró que los usuarios deberían tener participación en varias de las sesiones recopiladas, y en cada una de ellas aportar un volumen de audio y número de intervenciones suficientes para el entrenamiento y el reconocimiento. Como consecuencia se seleccionaron veinte usuarios identificados con las siguientes claves: *MC, JG, FO, AL, LA, JZ, AEr, JS, JM, EEk, NR, FM, EP, IC, VL, EE, AEx, MS, VP y CH.*

Para los usuarios se designaron como de entrenamiento los fragmentos de la primera sesión en la que intervienen. Para el entrenamiento del mundo se utilizaron los fragmentos correspondientes a las dos primeras sesiones. El detalle de esta designación se muestra en Tabla 5, que corresponde a la relación de los locutores

seleccionados como usuarios con indicación de su identificación, de la duración total de su fragmento, la sesión utilizada como entrenamiento, la duración total de los fragmentos utilizados para entrenamiento y los utilizados para test expresada en forma de duración (segundos) y en forma de número de fragmentos.

			Duración				Fragmentos		
Locutor	Sesión		Total	Entrenamiento	Test		Total	Entrenamiento	Test
MC	s100922		3091	301	2790		924	40	884
JG	s100922		2179	555	1625		584	71	513
FO	s101223		1886	412	1474		590	200	390
AL	s101223		1836	446	1390		595	190	405
LA	s101104		1695	382	1313		587	196	391
JZ	s100922		1680	690	990		316	71	245
AEr	s100922		1620	366	1254		494	39	455
JS	s100922		1491	728	763		393	92	301
JM	s101104		1449	479	971		621	283	338
EEk	s111102		1359	248	1111		498	145	353
NR	s100922		1333	454	879		310	54	256
FM	s101104		1213	584	630		429	225	204
EP	s111102		1195	221	974		416	87	329
IC	s111123		1113	329	784		293	95	198
VL	s111102		1011	133	878		417	70	347
EE	s101104		847	494	353		478	407	71
AEx	s111130		828	377	451		364	182	182
MS	s101223		752	463	288		282	194	88
VP	s111123		716	370	346		232	163	69
CH	s111130		700	182	518		157	52	105
MUNDO	s100922 s100104		7784	2753	5032		2691	1213	1478

Tabla 5: Usuarios seleccionados y sus participaciones.

4.6 Escenarios

Se plantearon dos escenarios en los cuales desarrollar la experimentación.

En el primero, durante el reconocimiento no se tendría en consideración el sexo del locutor, en razón de ello, los modelos fueron construidos en base a esta

suposición, de tal modo que para el entrenamiento de cada locutor se utilizaron sus vectores propios y como vectores ajenos una selección de los pertenecientes al mundo sin distinción de su género, este escenario ha sido denominado escenario "genérico".

Para el segundo escenario se planteó la diferenciación de sexo a la hora de la construcción de los modelos, por ello, los vectores propios de cada usuario son contrapuestos a una selección de los vectores del mundo correspondientes a locutores del mismo sexo que el usuario. Este escenario ha sido denominado "específico".

4.7 Selección de unidades de análisis

La siguiente decisión a fue la definición de las unidades de análisis que deberían ser utilizadas en el procedimiento de verificación de la hipótesis, para lo cual se estudiaron las unidades lingüísticas (palabras, silabas, fonemas), intentando predecir cuáles serían la capacidades de cada una de ellas a la hora de ser utilizadas al objetivo de la tesis.

La opción más natural es la que implica la utilización de las palabras individuales como unidad. Este enfoque presenta algunos puntos fuertes:

- Es relativamente sencilla la preparación del corpus: etiquetar los archivos de audio con las palabras que se pronuncian en cada momento, 35 hh[23] por hora de audio bruto.
- La pronunciación de palabras completas incorpora características prosódicas, las cuales tienen un carácter altamente discriminador.

Lamentablemente tras realizar un primer intento en esta dirección se pudo apreciar que el número de repeticiones de cada vocablo era realmente bajo y que, como es evidente, este número mantiene una relación inversa con la longitud del vocablo.

En consecuencia, no existe suficiente información que procesar, en las palabras largas, ya que el número de repeticiones es bajo y en las palabras cortas porque lo es su longitud. Esto deja un repertorio muy limitado unidades útiles para la experimentación; con el añadido de que en las construcciones cortas como los monosílabos tienden a desaparecer los efectos prosódicos.

También es cierto, según se pudo observar, que donde las características prosódicas se hacen más significativas, no en la fonación de palabras individuales, sino en secuencias de estas de uso habitual como: "es decir", "yo creo"…

[23] hh: Horas humano.

Atendiendo al hecho de que la etiquetación de los audios se realizaría de forma manual, las silabas y fonemas presentan un problema práctico: la imposibilidad de marcar el comienzo y fin de cada una de ellas en un abundante número de casos. En el habla natural los fonemas se presentan ligados en secuencias, incluso abarcado grupos de palabras contiguas gramaticalmente separadas, en los que no aparece fin de continuidad alguno, ya que en el habla natural se concatenan, se anticipan, se mezclan e incluso desaparecen (ver Figura 40), hecho que hace imposible establecer los instantes de inicio u final, básicamente porque estos instantes no suelen existir.

Existen otras potenciales unidades de análisis, muy utilizadas en el área del reconocimiento del habla: bifonemas y trifonemas.

Tanto el bifonema (encadenamiento de dos fonemas consecutivos en una locución) como el trifonema (sucesión de tres fonemas consecutivos en una locución), considerados como unidad de análisis, aunque se requieren un importante esfuerzo para su identificación (140 hh/h), contienen información suprasegmental, en concreto la intersegmental vinculada a la coarticulación, la forma en que evoluciona el tracto vocal para pasar de la emisión de un fonema a la emisión del siguiente.

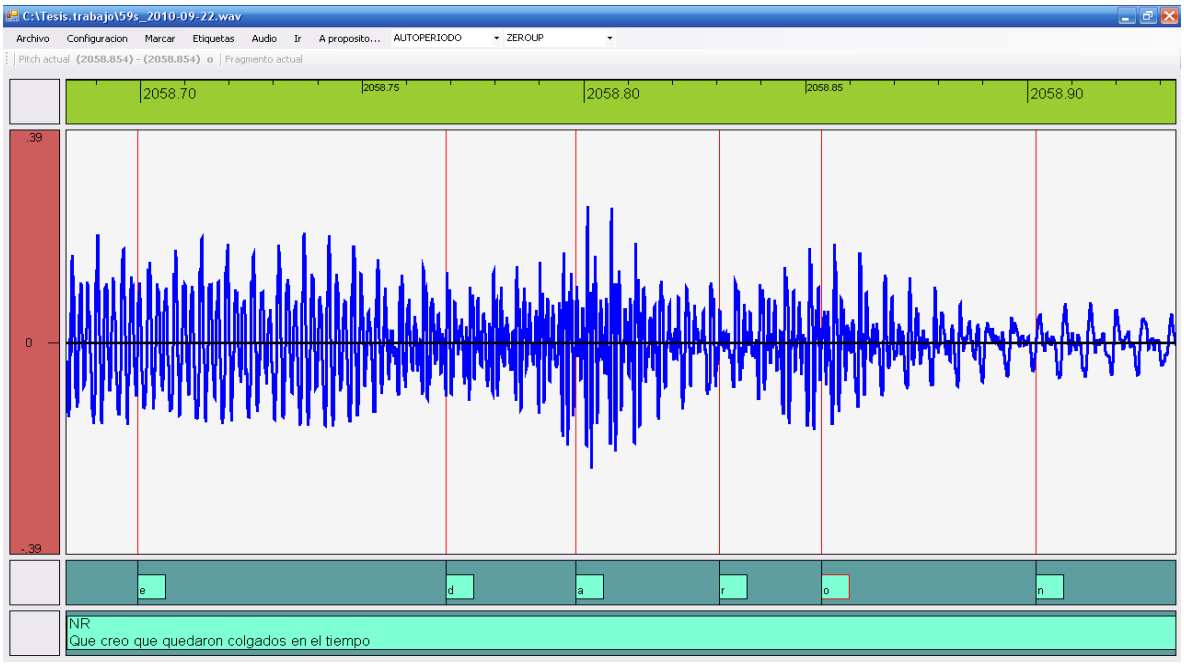


Figura 40: Ejemplo de la evolución de la forma de onda.

Por otro lado ambos presentan como ventaja que no sea necesario identificar con precisión los fonemas que lo componen sino simplemente marcar aproximadamente su centro en la seguridad de que la información buscada se encuentra entre dos de esas marcas.

Aunque el número de bifonemas en castellano es relativamente alto, permanece suficientemente limitado como para asegurar que un buen número de ellos tendrá suficiente número de repeticiones presentes en cada sesión, y aquellos que no, corresponderán a bifonemas no lo suficientemente comunes como para resultar interesantes en este trabajo.

Una estimación razonable de trifonemas en castellano es de 5000, esto hace que sea difícil obtener un subconjunto significativo de ellos para los que se disponga de suficiente información y que finalmente resulten significativos.

En segundo lugar en el trifonema existen dos grupos coarticulatorios distintos, y por lo tanto o se tratan de forma individual, es decir como bifonemas, o se tratan de forma conjunta con lo que estarán gobernados por una función de distribución que tendrá una mayor dispersión y por tanto perderá gran parte de su especificidad.

En razón de lo dicho más arriba, se tomó la decisión de proponer los bifonemas por ser una unidad prometedora que mantiene alta la relación *[volumen de audio/número de unidades]*. De este modo el criterio de detección de los segmentos se resume en la identificación del bifonema pronunciado en el, cosa que puede realizarse de forma automática si se dispone de un reconocedor fonético, núcleo central de un ASR (Automatic Speech Recognizer), con un modelo específico para el idioma objetivo.

4.8 Etiquetado del corpus

Para la etiquetación de los fragmentos con los códigos de locutor, fue de gran ayuda disponer de los videos que permitieron una identificación visual de la persona. Como apoyo se desarrolló una herramienta especifica al caso (Apéndice III - d. Transcriptor), al apreciar que los "open source" que funcionalmente cubrían las necesidades requeridas tienen una sensible caída de prestaciones cuando trabajan con archivos del volumen de los utilizados.

La carencia, para este trabajo, del modelo fonético para castellano, obligó a realizar el etiquetado fonético de los audios de forma manual, localizando cada uno de los centros de los fonemas e identificando el bifonema comprendido entre dos consecutivos, tarea que resulta necesaria para la preparación del corpus, como se muestra en la Figura 41. En apoyo se readapto un desarrollo previo (Apéndice III - c. PitchMarker) creado por el doctorando para la detección del *pitch* en los procesos de modelado para la síntesis del habla.

Se observó que resultaba mucho más productivo realizar el etiquetado fonético si era conocido el texto de la locución, por ello se procedió a realizar la transcripción

de los fragmentos utilizando la misma herramienta que para el etiquetado de locutores.

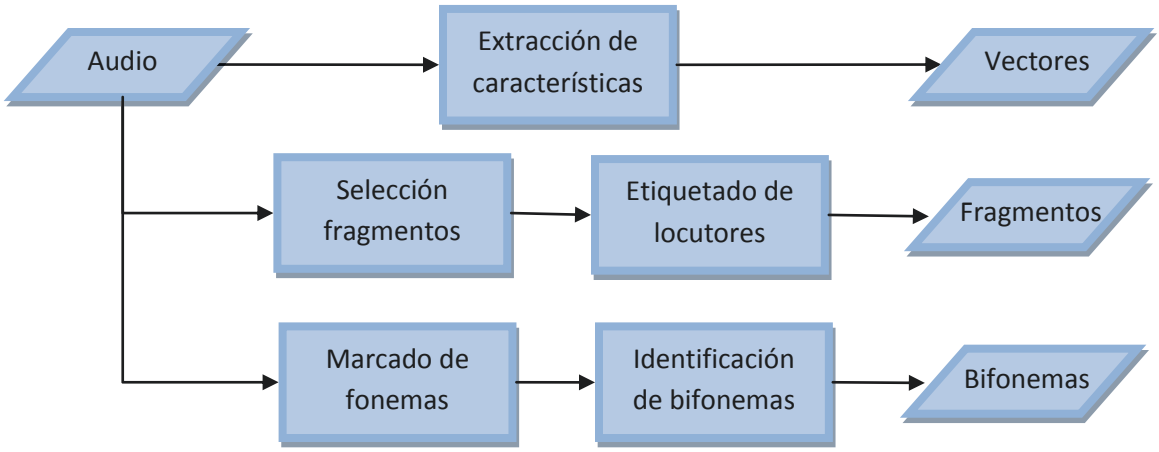


Figura 41: Proceso ampliado de preparación del corpus.

Dado que el objetivo es realizar el reconocimiento de locutor en base al habla natural y dada la amplia variedad de realizaciones de fonemas, no sólo debidos a los alófonos propios del idioma o de su variedad lingüística, si no a también a las producidas por alteraciones prosódicas coyunturales (perdidas de fonemas, sustitución, etc.) fue necesario establecer una disciplina de trabajo a la hora de realizar el marcado de fonemas, de modo que su sistemática diera robustez al corpus, y mantuviera el proceso constantemente orientado al objetivo final.

Las pautas establecidas para esta disciplina se pueden resumir en:

- Utilizar un conjunto de unidades reducido, ignorando alófonos y concentrándose única y exclusivamente en los fonemas.
- Utilizar un conjunto de etiquetas sencillo: mono-carácter y dentro lo más cercano posible a los símbolos escritos correspondientes, y en evitación de los problemas de codificación mantenerse dentro del conjunto de caracteres ASCII; todo lo cual simplificó drásticamente el marcado (véase Apéndice I Etiquetación de los fonemas).
- Marcar los fonemas que se han producido y no aquellos que debían haberse producido, del mismo modo que en el ejemplo:

si la transcripción del fragmento dice:

es decir yo creo que es posible

en vez de etiquetar:

e s d e z i R y o k R e o k e e s p o s i b l e

etiquetar:

s z i R y k r o k e j p o s b l e

Realizando esta labor para los archivos de la base de datos, en los fragmentos de entrenamiento se identificaron 375 bifonemas distintos, haciendo un total de 128.838 repeticiones (segmentos), en el Apéndice II Catalogación de las unidades se presenta el detalle de su distribución.

4.9 Evaluación

La evaluación de los resultados se realizó en cada uno de los dos escenarios propuestos en base a la comparación de las tasas de error mediante el EER definido en 2.1.7, y al comportamiento de la ROC mediante su AUC definidos en 2.1.8.

5 Desarrollo experimental

5.1 Línea Base

La línea base corresponde a la determinación de las prestaciones del procedimiento estándar en los escenarios planteados; permite establecer el marco de comparación de los resultados experimentales que se obtengan a partir de la aplicación de la propuesta de solución realizada en la sección 3.2, comparación que deberá demostrar la validez de la hipótesis que fundamenta esta tesis.

El proceso experimental a ejecutar para el establecimiento de la línea base se desglosa en:

- Entrenamiento, para todos los fragmentos seleccionados como de entrenamiento:
  - Realizar la extracción de características.
  - Seleccionar 4000 vectores de cada usuario (propios).
  - Seleccionar 4000 vectores del resto de locutores (ajenos).
  - Construir el modelo de cada locutor utilizando propios + ajenos.
- Reconocimiento, para cada uno de los fragmentos:
  - Compararlo con cada uno de los modelos de usuario.
  - Obtener su puntuación, para lo cual se contabiliza el número total de vectores rechazados del fragmento (n_r, $T(v)<0$) y el número de vectores aceptados (n_a, $T(v)\geq 0$) en la comparación, obteniéndose el valor buscado del cociente:

$$T_s(v) = \frac{n_a}{n_a + n_r} \qquad \text{Ecuación 56}$$

  - Etiquetar la puntuación como genuina o impostada, dependiendo de si el modelo y el fragmento corresponden al mismo locutor o no.
- Evaluación: para cada uno de los usuarios con las puntuaciones etiquetadas obtenidas de la comparación de los fragmentos con su modelo:

- Determinar el cross point y calcular el EER asociado.
- Determinar la ROC y calcular su AUC.

Para el establecimiento de la línea base, a partir de la definición ya realizada de la selección de usuarios, de fragmentos y de las utilizaciones de los mismos (véase 4.4 Corpus), se procedió a considerar dos escenarios etiquetados LBg[24] y LBe[25].

5.1.1 Escenario genérico (LBg)

En el escenario genérico, se considera que no existe conocimiento sobre el sexo del locutor, bajo tales condiciones, los modelos de los distintos usuarios se construyen con independencia del mismo, de forma coherente el test se realiza confrontado todos los modelos frente a todos los fragmentos.

Dentro de este escenario el umbral de decisión se establece para cada usuario de forma independiente de modo que optimicen los resultados de los test de todos los fragmentos con su modelo personal bajo el criterio del EER.

LBg				
usr.	g	umbral	eer	auc
AEx	h	0.73	7.7%	96.9%
AL	h	0.46	7.5%	96.6%
AEr	h	0.47	21.2%	83.2%
CH	m	0.70	4.6%	98.7%
EEk	h	0.70	16.5%	92.3%
EE	m	0.49	2.0%	98.5%
EP	m	0.66	9.9%	96.3%
FO	h	0.47	8.0%	96.5%
FM	h	0.45	9.0%	93.0%
IC	h	0.71	12.6%	95.0%
JG	h	0.48	7.7%	96.3%
JM	h	0.43	11.4%	93.5%
JS	h	0.50	4.2%	99.1%
JZ	h	0.48	4.3%	98.7%
LA	h	0.41	12.5%	93.3%
MS	m	0.48	13.3%	94.5%
MC	m	0.29	10.9%	92.7%
NR	m	0.43	4.2%	98.5%

[24] LBg: Línea Base escenario Genérico
[25] LBe: Línea Base escenario Específico.

LBg				
usr.	g	umbral	eer	auc
VL	m	0.68	17.9%	90.5%
VP	m	0.76	4.7%	98.9%
Promedio			**9.5%**	**95.2%**
Desv. Estándar			**5.1%**	**3.8%**
Desv. Estándar/Promedio			**53.7%**	**4.0%**

Tabla 6: Resultados de la evaluación individualizada en el escenario LBg.

5.1.2 Escenario específico (LBe)

En contraposición con el escenario LBg en este se tiene presente en todo momento el género del locutor, ello implica que el entrenamiento se realiza utilizando un mundo con fragmentos correspondientes a locutores del mismo género y coherentemente los test de evaluación se realizan enfrentando modelos sólo a tramas del genero correspondiente. Bajo los dos criterios de evaluación antes descritos (global e individualizado) se obtuvieron los resultados presentados en la Tabla 7.

LBe				
usr.	g	umbral	eer	auc
AEx	h	0,74	8.6%	96.3%
AL	h	0.48	8.4%	96.0%
AEr	h	0.48	17.5%	86.7%
CH	m	0.76	6.5%	97.7%
EEk	h	0.71	18.7%	90.0%
EE	m	0.64	1.7%	98.6%
EP	m	0.73	20.0%	88.2%
FO	h	0.48	11.4%	94.8%
FM	h	0.45	9.6%	92.5%
IC	h	0.71	10.9%	95.7%
JG	h	0.48	8.8%	95.9%
JM	h	0.43	13.1%	92.2%
JS	h	0.50	4.4%	98.7%
JZ	h	0.50	4.6%	98.9%
LA	h	0.41	14.0%	92.6%
MS	m	0.53	13.9%	94.9%
MC	m	0.37	11.3%	92.5%
NR	m	0.49	6.6%	97.7%
VL	m	0.73	18.4%	89.9%
VP	m	0.81	6.9%	97.8%

LBe				
usr.	g	umbral	eer	auc
Promedio			10.8%	94.4%
Desv. Estándar			5.2%	3.7%
Desv. Estándar/Promedio			47.8%	3.9%

Tabla 7: Resultados de la evaluación individualizada en el escenario LBe.

5.1.3 Resumen

Los resultados obtenidos pueden resumirse promediando los correspondientes a cada uno de los usuarios tal y como se presenta en Tabla 8, acompañados de ejemplos del comportamiento de la falsa aceptación, falso rechazo en las Figura 42 y de la ROC en la Figura 43, el resto de las gráficas pueden ser consultadas en el OJO

Escenario	EER	AUC
LBg	9,5%	95,2%
LBe	10,8%	94,4%

Tabla 8: Resultados de la línea base.

5.2 Categorización de unidades de análisis.

La categorización de las unidades de análisis supuso para cada usuario el ordenamiento de las mismas según su capacidad discriminatoria; donde, para el presente caso, se ha utilizado la algorítmica presentada en la sección 3.2, que implica, para cada unidad:

1. Determinar qué segmentos del audio del audio le corresponden y qué vectores componen dichos segmentos.
2. Obtener la puntuación de cada uno de esos vectores tras la comparación con el modelo del usuario.
3. Promediar todas las puntuaciones de los vectores correspondientes a la unidad.

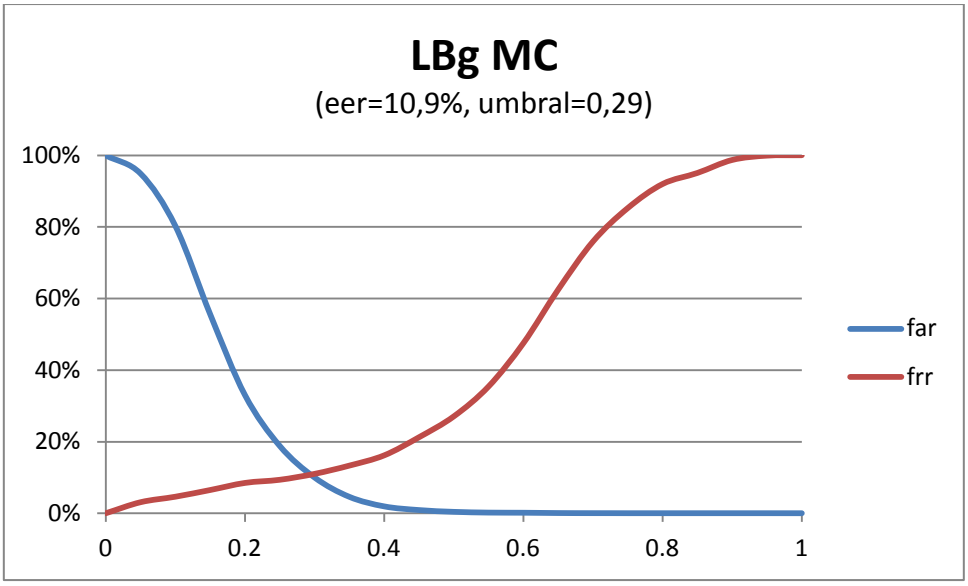


Figura 42: Ejemplo de curvas FAR-FRR, sujeto MC escenario genérico.

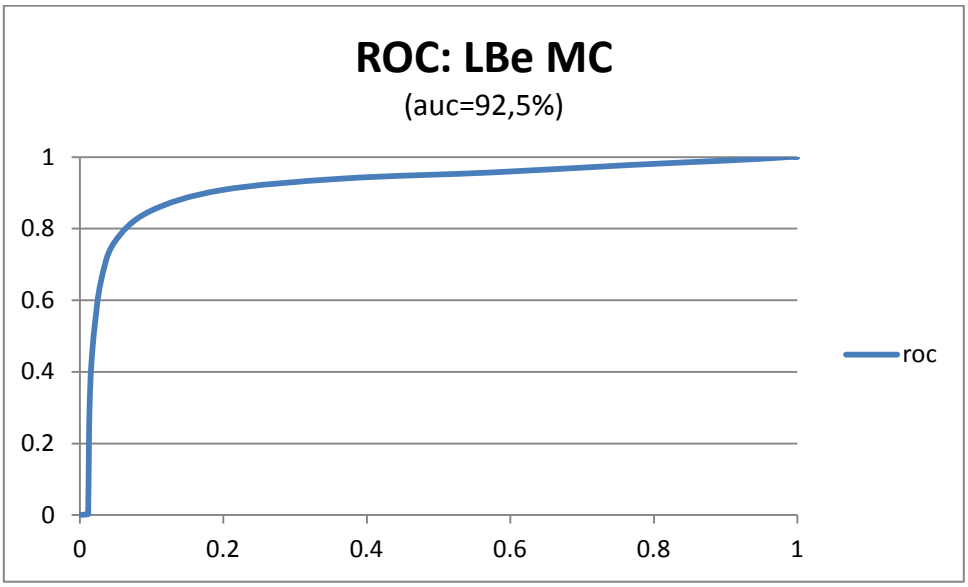


Figura 43: Ejemplo del comportamiento de la ROC, sujeto MC escenario específico.

Utilizando sólo los fragmentos de entrenamiento de cada usuario, se calculó la valoración de cada una de las unidades presentes siguiendo la Ecuación 53, eliminando de la lista aquellas con número escaso de ejemplos.

Aplicando esta categorización a todos los bifonemas identificados en los fragmentos de entrenamiento de cada usuario se obtienen los resultados presentados en el Apéndice II Catalogación de las unidades.

Si sobre dichos resultados se ignoran aquellas unidades con pocas repeticiones (menos de 100 vectores) observando sólo aquellas de alcanzan las mayores

valoraciones y las que alcanzan las menores se obtiene la información de las siguientes tablas:

AEx		AL		AE		CH		EEk		EE		EP	
[e-p]	1.17	[t-a]	0.84	[e-e]	1.01	[l-p]	1.39	[t-e]	1.26	[l-p]	1.33	[t-a]	1.34
[e-t]	1.13	[a-n]	0.84	[t-e]	0.83	[a-n]	1.24	[o-k]	1.14	[o-p]	1.24	[a-n]	1.30
[a-t]	1.11	[e-m]	0.83	[n-d]	0.81	[e-k]	1.23	[n-t]	1.07	[d-a]	1.22	[e-n]	1.19
[l-p]	1.07	[e-n]	0.80	[d-e]	0.76	[p-e]	1.21	[e-n]	1.00	[l-a]	1.20	[s-o]	1.12
[a-p]	1.06	[a-d]	0.79	[s->]	0.70	[u-n]	1.21	[t-a]	0.98	[e-f]	1.17	[p-a]	1.10
[p-e]	1.05	[l-a]	0.79	[<-e]	0.70	[s-o]	1.20	[s-e]	0.98	[r-o]	1.16	[a-k]	1.09
[o-k]	1.05	[a-m]	0.79	[n-t]	0.70	[t-a]	1.20	[i-e]	0.95	[r-a]	1.14	[s-u]	1.08
[e-n]	1.05	[d-a]	0.77	[e-s]	0.67	[e-p]	1.19	[t-o]	0.93	[p-o]	1.12	[u-n]	1.07
[n-t]	1.04	[z-i]	0.75	[o-d]	0.64	[k-a]	1.17	[k-e]	0.92	[R-a]	1.10	[s-e]	1.07
[o-p]	1.04	[i-e]	0.74	[o-k]	0.64	[t-e]	1.16	[o-n]	0.92	[y-a]	1.10	[s-i]	1.06
...		...		...		...		...		...		...	
[s->]	0.66	[e-R]	0.44	[e-l]	0.37	[a-d]	0.81	[e-R]	0.87	[s-k]	0.68	[a-s]	0.88
[a-o]	0.65	[d-o]	0.44	[e->]	0.37	[R-a]	0.80	[t-i]	0.86	[n-k]	0.68	[l-a]	0.87
[e-R]	0.64	[e-k]	0.43	[l-o]	0.36	[a-l]	0.80	[R-e]	0.86	[m-a]	0.66	[o-s]	0.84
[e-d]	0.63	[s-o]	0.43	[R-e]	0.35	[e-R]	0.79	[s-i]	0.84	[e->]	0.66	[n-a]	0.83
[<-n]	0.62	[a->]	0.43	[e-z]	0.31	[a-R]	0.77	[k-i]	0.81	[a-R]	0.64	[e-s]	0.82
[a-d]	0.60	[t-o]	0.39	[s-t]	0.29	[d-o]	0.74	[l-a]	0.78	[c-o]	0.64	[d-e]	0.81
[o-l]	0.59	[k-o]	0.35	[k-o]	0.25	[R-e]	0.74	[e-s]	0.78	[m-p]	0.63	[a-d]	0.81
[n-o]	0.58	[a-R]	0.34	[p-a]	0.24	[i-e]	0.73	[a-s]	0.77	[i-e]	0.62	[a-l]	0.78
[a-r]	0.57	[o-R]	0.28	[l-a]	0.22	[l-e]	0.71	[e-l]	0.77	[a->]	0.61	[i-s]	0.76
[n-i]	0.56	[o-k]	0.28	[a-l]	0.10	[a->]	0.69	[n-a]	0.77	[n-e]	0.53	[d-a]	0.73

Tabla 9: Bifonemas extremos, locutores AEx-EP.

En ellas se presentan los fonemas extremos en puntuación para cada uno le los usuarios seleccionados.

Puede observarse que es común que entre los mejores se encuentre bifonemas que incluyan fonemas como “m” o “n” en combinación con vocales abiertas como “a” o ”e”; mientras que entre los peores se localizan muchas veces los bifonemas compuestos por consonantes como “R, “r”, “t”, ”k”, ”s” en combinación con vocales cerradas como “o” y “u”.

FO		FM		IC		JG		JM		JS		JZ	
[u-n]	1.34	[o-o]	0.84	[e-p]	1.08	[e-n]	1.03	[s-m]	0.92	[<-n]	1.47	[s->]	1.07
[<-m]	1.05	[R-n]	0.78	[l-p]	1.07	[a-n]	1.00	[e-m]	0.86	[<-m]	1.41	[<-m]	0.95

[n-a]	1.02	[n-p]	0.77	[a-p]	1.02	[e-a]	1.00	[e-e]	0.85	[i-m]	1.28	[s-k]	0.92
[m-e]	1.01	[i-a]	0.75	[u-t]	0.99	[o-n]	0.94	[b-i]	0.84	[n-u]	1.14	[o-p]	0.89
[a-m]	0.99	[b-i]	0.75	[a-t]	0.98	[<-m]	0.94	[e-d]	0.83	[R->]	1.10	[<-e]	0.87
[e-m]	0.89	[k-a]	0.74	[e-n]	0.97	[i-m]	0.89	[k-e]	0.80	[e-n]	1.07	[d-a]	0.85
[o-n]	0.86	[a-p]	0.74	[n-t]	0.92	[u-n]	0.88	[i-d]	0.80	[o-m]	1.04	[e-e]	0.84
[i-e]	0.85	[n-a]	0.73	[z-e]	0.91	[i-a]	0.87	[i-p]	0.78	[s->]	1.03	[t-e]	0.84
[n-e]	0.82	[a-t]	0.72	[u-n]	0.90	[o-m]	0.85	[<-e]	0.76	[i-n]	1.00	[e->]	0.83
[z-e]	0.78	[a-~]	0.72	[a-n]	0.89	[r-a]	0.85	[o->]	0.75	[u-n]	0.92	[<-a]	0.81
...		...		...		...		...		...		...	
[s-a]	0.43	[m-i]	0.37	[i-o]	0.63	[o-b]	0.48	[k-i]	0.44	[e-l]	0.45	[o-s]	0.34
[a-d]	0.41	[e-g]	0.37	[s->]	0.63	[a-m]	0.48	[m-p]	0.44	[a-b]	0.44	[a-m]	0.34
[z-i]	0.40	[g-R]	0.35	[e-e]	0.63	[p-e]	0.48	[o-R]	0.44	[e-d]	0.42	[k-o]	0.34
[i-o]	0.39	[e-R]	0.34	[<-a]	0.61	[o-e]	0.46	[k-u]	0.44	[d-a]	0.41	[s-u]	0.34
[i-s]	0.37	[a-d]	0.33	[a-l]	0.61	[a-l]	0.43	[t-o]	0.41	[s-a]	0.38	[p-e]	0.33
[d-e]	0.35	[a-R]	0.30	[p-o]	0.59	[z-e]	0.43	[i-s]	0.40	[k-a]	0.38	[d-i]	0.32
[a-l]	0.33	[e-r]	0.24	[n-o]	0.58	[n-o]	0.40	[<-i]	0.38	[n-e]	0.32	[n-o]	0.27
[s-t]	0.32	[n->]	0.21	[o-l]	0.54	[m-o]	0.39	[r-o]	0.32	[n-a]	0.32	[i-k]	0.22
[a-k]	0.31	[e->]	0.19	[o-r]	0.53	[a-d]	0.28	[a-o]	0.24	[e-r]	0.32	[e-l]	0.20
[a-b]	0.28	[<-d]	0.04	[o-R]	0.53	[e-d]	0.26	[b-a]	0.23	[i-a]	0.28	[a-b]	0.18

Tabla 10: Bifonemas extremos, locutores FO-JZ.

LA		MS		MC		NR		VL		VP			
[a-p]	1.07	[a-d]	1.13	[i-t]	1.67	[a-p]	1.18	[o-p]	1.56	[l-p]	1.36		
[e-f]	1.02	[n-t]	1.00	[s-d]	1.59	[a-t]	1.16	[e-n]	1.19	[o-p]	1.26		
[a-t]	0.91	[a-n]	0.99	[i-d]	1.58	[n-t]	1.14	[o->]	1.15	[a-n]	1.24		
[r-i]	0.89	[a-l]	0.98	[m-p]	1.56	[i-d]	1.02	[o-n]	1.15	[d-a]	1.16		
[o->]	0.87	[o->]	0.98	[s-i]	1.56	[d-a]	1.02	[d-e]	1.12	[a-m]	1.14		
[o-k]	0.86	[r-a]	0.90	[e-t]	1.55	[o-k]	1.00	[a-k]	1.09	[l-a]	1.14		
[e-p]	0.86	[t-e]	0.89	[l-i]	1.55	[o-s]	0.95	[l-o]	1.05	[p-o]	1.13		
[a->]	0.84	[s->]	0.88	[e-m]	1.55	[o->]	0.95	[s-t]	0.97	[R-o]	1.11		
[p-a]	0.84	[a->]	0.88	[l-p]	1.53	[i-n]	0.94	[t-a]	0.97	[r-o]	1.11		
[n-t]	0.84	[k-e]	0.87	[i-m]	1.53	[a->]	0.86	[n-e]	0.97	[e-n]	1.11		
...		...		...		...		...		...			
[e-a]	0.53	[o-s]	0.63	[<-i]	0.54	[a-l]	0.54	[e-s]	0.95	[a-s]	0.76		
[o-R]	0.52	[d-e]	0.63	[d-o]	0.52	[e-l]	0.54	[p-e]	0.95	[s-u]	0.76		
[a-b]	0.52	[s-a]	0.62	[r-a]	0.52	[u-n]	0.52	[e->]	0.92	[a-l]	0.76		
[i->]	0.52	[a-s]	0.61	[k-o]	0.52	[k-o]	0.50	[k-a]	0.87	[s-t]	0.75		
[o-i]	0.50	[R-e]	0.57	[o-R]	0.52	[s-e]	0.48	[a-s]	0.87	[m-a]	0.73		

[s-i]	0.49	[s-o]	0.57	[a-o]	0.50	[<-a]	0.47	[k-e]	0.83	[s-k]	0.72
[a-r]	0.49	[m-e]	0.56	[s->]	0.50	[m-e]	0.46	[s-i]	0.82	[e->]	0.67
[a-i]	0.48	[i-a]	0.55	[a-y]	0.47	[i-a]	0.46	[e-l]	0.80	[n-e]	0.67
[y-o]	0.45	[e->]	0.54	[R->]	0.45	[R-e]	0.45	[o-R]	0.78	[i-e]	0.65
[e-d]	0.42	[e-l]	0.45	[u-r]	0.32	[i-k]	0.41	[s->]	0.77	[a->]	0.62

Tabla 11: Bifonemas extremos, locutores LA-VP.

5.3 Exclusión de los peores

Con los mismos criterios experimentales utilizados para el establecimiento de la línea base se procedió a evaluar el algoritmo propuesto de "Exclusión de los Peores" que identifica las unidades con peor valoración para los segmentos que les corresponden durante la locución sean ignorados en la fase de reconocimiento. Se utilizó como 50% como factor de filtrado (porcentaje de las unidades categorizadas que se consideran, l^+ en la Ecuación 54)

De nuevo todos los fragmentos fueron contrastados con los modelos, tan sólo excluyendo previamente de aquellos todos los vectores correspondientes a unidades que se encontraban entre las elegidas.

5.3.1 Escenario EPSg

Con este algoritmo en el escenario genérico se obtuvieron los siguientes resultados por usuario:

EPSg				
usr.	g	umbral	eer	auc
AEx	h	0,68	6,6%	97,7%
AL	h	0,57	3,9%	98,1%
AEr	h	0,60	9,5%	95,3%
CH	m	0,76	4,0%	99,2%
EEk	h	0,73	7,9%	97,6%
EE	m	0,61	1,6%	98,4%
EP	m	0,87	5,1%	98,2%
FO	h	0,61	6,1%	97,9%
FM	h	0,47	6,2%	97,0%
IC	h	0,66	6,2%	98,4%
JG	h	0,53	1,9%	98,6%
JM	h	0,56	9,2%	94,8%
JS	h	0,63	3,7%	98,0%
JZ	h	0,55	2,2%	99,2%

EPSg				
usr.	g	umbral	eer	auc
LA	h	0,63	7,5%	97,2%
MS	m	0,71	8,3%	97,2%
MC	m	0,27	8,8%	94,3%
NR	m	0,61	2,7%	99,1%
VL	m	0,85	8,5%	96,7%
VP	m	0,72	3,6%	98,6%
Promedio			**5,6%**	**97,6%**
Desv, Estándar			**2,6%**	**1,4%**
Desv, Estándar/Promedio			**46,1%**	**1,5%**

Tabla 12: Resultados de la evaluación individualizada en el escenario EPSg.

5.3.2 Escenario EPSe

Mientras que en el escenario específico los resultados son los listados en la siguiente tabla para cada uno de los usuarios:

EPSe				
usr.	g	umbral	eer	auc
AEx	h	0,68	5,9%	98,4%
AL	h	0,59	6,3%	97,7%
AEr	h	0,61	9,8%	97,5%
CH	m	0,86	4,1%	99,3%
EEk	h	0,70	11,6%	95,0%
EE	m	0,67	1,2%	98,1%
EP	m	0,86	11,3%	95,3%
FO	h	0,61	7,6%	97,4%
FM	h	0,46	9,1%	96,1%
IC	h	0,79	8,3%	97,2%
JG	h	0,52	3,4%	99,6%
JM	h	0,59	10,2%	95,8%
JS	h	0,67	3,8%	99,2%
JZ	h	0,58	2,1%	99,2%
LA	h	0,65	10,5%	96,7%
MS	m	0,68	7,2%	98,0%
MC	m	0,32	9,2%	95,4%

NR	m	0,63	5,1%	99,0%
VL	m	0,86	11,8%	95,0%
VP	m	0,32	4,4%	99,4%
Promedio			**7,2%**	**97,5%**
Desv, Estándar			**3,3%**	**1,6%**
Desv, Estándar/Promedio			**45,6%**	**1,6%**

Tabla 13: Resultados de la evaluación individualizada en el escenario EPSe.

5.4 Comparativa

Finalmente en Tabla 14 se presentan los promedios de los resultados obtenidos en comparación con los obtenidos para la línea base de cada escenario.

Escenario	EER	AUC
EPSg	5,6%	97,6%
LBg	9,5%	95,2%
EPSe	7,2%	97,5%
LBe	10,8%	94,4%

Tabla 14: Resultados promedio EPS.

En resumen puede observarse en Figura 44 y Figura 45 la consistencia de la mejora en todos los casos tanto medido en forma de EER como en AUC. Es de notar que es más sustancial en aquellos casos en los que la línea base presenta peor comportamiento.

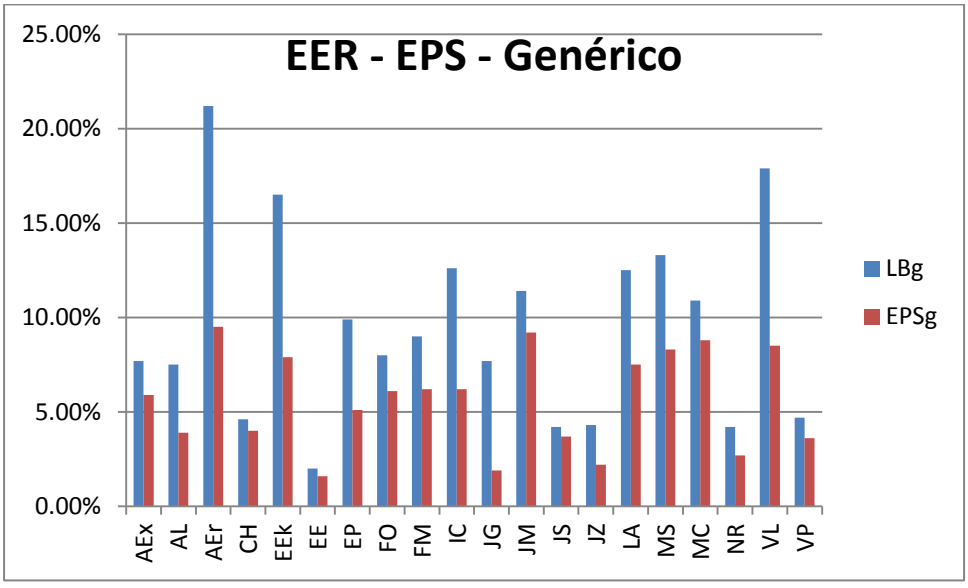


Figura 44: Comparativa de los EER de LB y EPS en el escenario genérico.

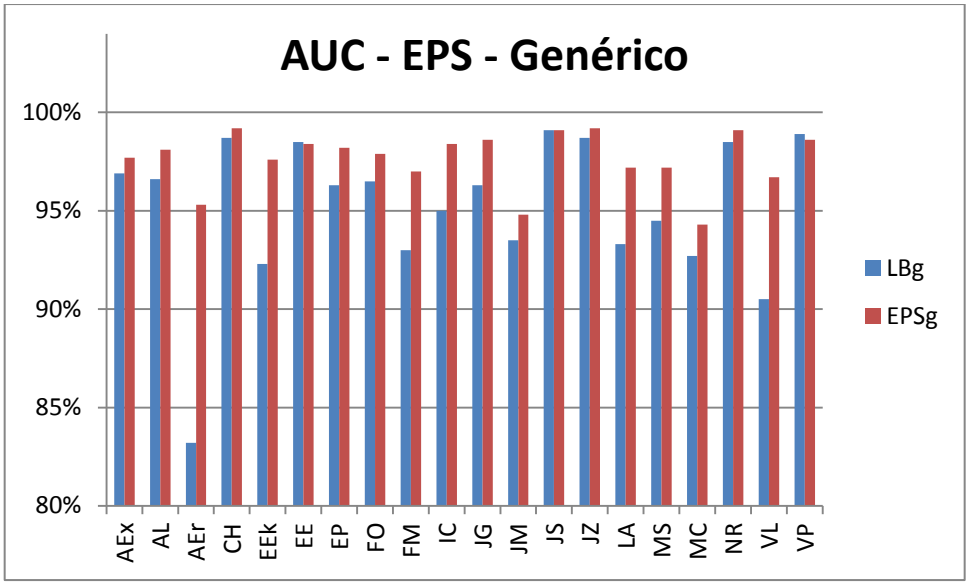


Figura 45: Comparativa de los AUC de LB y EPS en el escenario genérico.

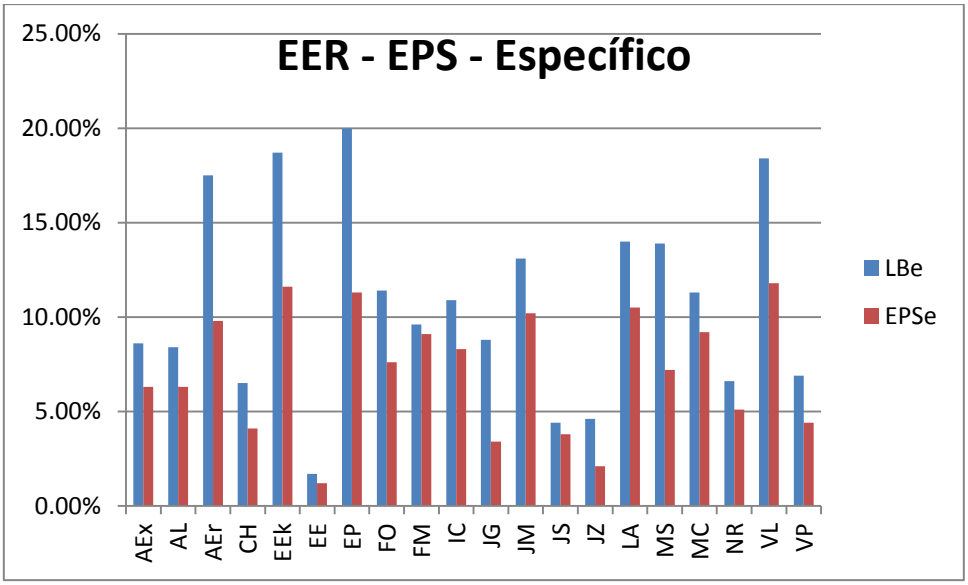


Figura 46: Comparativa de los EER de LB y EPS en el escenario específico.

Un comportamiento similar es que reflejan los resultados en el escenario específico, tal y como muestran Figura 46 y Figura 47.

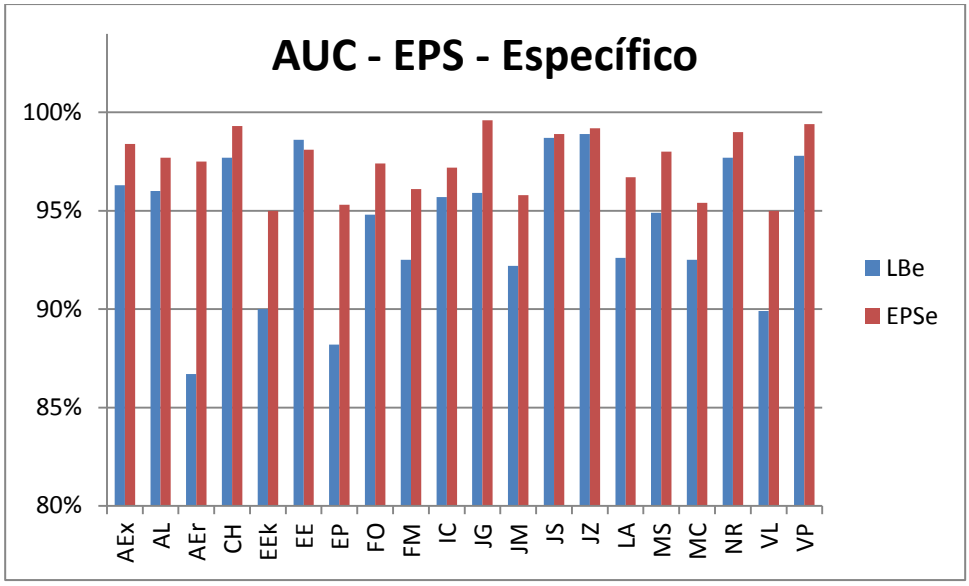


Figura 47: Comparativa de los AUC de LB y EPS en el escenario específico.

5.5 Selección de los mejores

Haciendo uso de los mismos criterios expuestos en la sección anterior se procedió a evaluar el algoritmo propuesto de "Selección de los Mejores" según el cual sólo se utilizan los segmentos correspondientes a las unidades con mejor comportamiento. De la misma forma que para EPS, se estableció el factor de filtrado en el 50% (l en la Ecuación 54), lo que significa que las unidades elegidas son el 50% mejor valorado entre las unidades identificadas en el entrenamiento.

Utilizar sólo un corto repertorio de bifonemas lleva en algunos casos a disponer de un fragmento con muy pocos ejemplos, en estos casos el fragmento no fue considerado.

5.5.1 Escenario SMSg

Los resultados de este algoritmo aplicado en el escenario genérico corresponden a los listados en la siguiente tabla:

SMSg				
usr.	g	umbral	eer	auc
AEx	h	0.70	6.6%	98.5%
AL	h	0.57	3.5%	97.9%
AEr	h	0.64	4.3%	99.5%
CH	m	0.83	4.5%	98.8%

SMSg				
usr.	g	umbral	eer	auc
EEk	h	0.76	7.0%	97.6%
EE	m	0.58	1.6%	98.0%
EP	m	0.88	4.7%	99.0%
FO	h	0.62	4.7%	98.1%
FM	h	0.48	7.8%	97.3%
IC	h	0.83	7.8%	97.1%
JG	h	0.60	2.1%	99.3%
JM	h	0.57	9.0%	95.3%
JS	h	0.68	4.2%	98.2%
JZ	h	0.56	2.3%	98.2%
LA	h	0.63	6.1%	97.3%
MS	m	0.76	8.2%	96.8%
MC	m	0.30	8.5%	95.8%
NR	m	0.56	2.1%	98.6%
VL	m	0.89	7.0%	97.4%
VP	m	0.81	3.5%	98.5%
Promedio			**5.3%**	**97.9%**
Desv. estándar			**2.4%**	**1.1%**
Desv. Estándar/Promedio			**44.6%**	**1.1%**

Tabla 15: Resultados de la evaluación individualizada en el escenario SMSg.

5.5.2 Escenario SMSe

Los resultados obtenidos en el escenario específico se presentan en la siguiente tabla:

SMSe				
usr.	g	umbral	eer	auc
AEx	h	0.67	5.6%	98.5%
AL	h	0.59	5.3%	98.1%
AEr	h	0.62	9.6%	96.0%
CH	m	0.88	4.0%	99.0%
EEk	h	0.75	8.8%	96.5%
EE	m	0.65	1.1%	98.7%
EP	m	0.88	8.5%	96.6%
FO	h	0.64	6.5%	96.9%
FM	h	0.49	8.3%	95.9%
IC	h	0.79	7.5%	97.9%
JG	h	0.59	1.9%	99.0%

SMSe				
usr.	g	umbral	eer	auc
JM	h	0.61	10.1%	95.6%
JS	h	0.61	4.1%	99.8%
JZ	h	0.58	2.0%	97.5%
LA	h	0.65	10.7%	95.2%
MS	m	0.69	7.3%	96.9%
MC	m	0.35	8.5%	95.0%
NR	m	0.70	4.0%	98.9%
VL	m	0.89	9.3%	96.5%
VP	m	0.82	4.1%	98.9%
Promedio			**6.4%**	**97.3%**
Desv. Estándar			**2.9%**	**1.4%**
Desv. Estándar/Promedio			**45.7%**	**1.4%**

Tabla 16: Resultados de la evaluación individualizada en el escenario SMSe.

5.5.3 Comparativa

En la Tabla 18 se presentan los resultados promedio en comparación con los obtenidos para la línea base:

Escenario	EER	AUC
SMSg	5,3%	97,9%
LBg	9,5%	95,2%
SMSe	6,4%	97,3%
LBe	10,8%	94,4%

Tabla 17: Resultados EPS.

Las Figura 48 a la Figura 51 resumen dichos resultado donde al igual que con el algoritmo EPS se observa la consistencia de la mejora en EER y AUC.

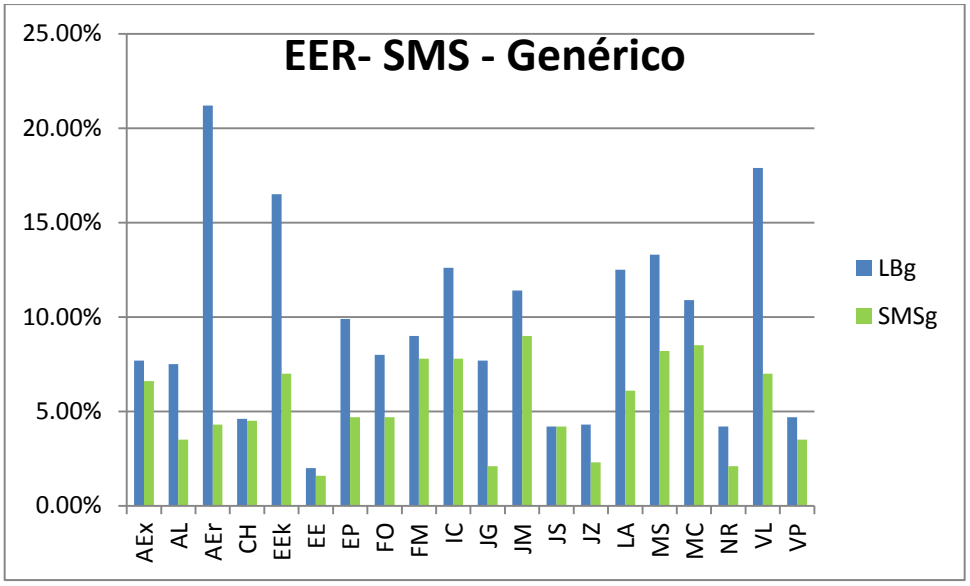


Figura 48: Comparativa de los EER de LB y SMS en el escenario genérico.

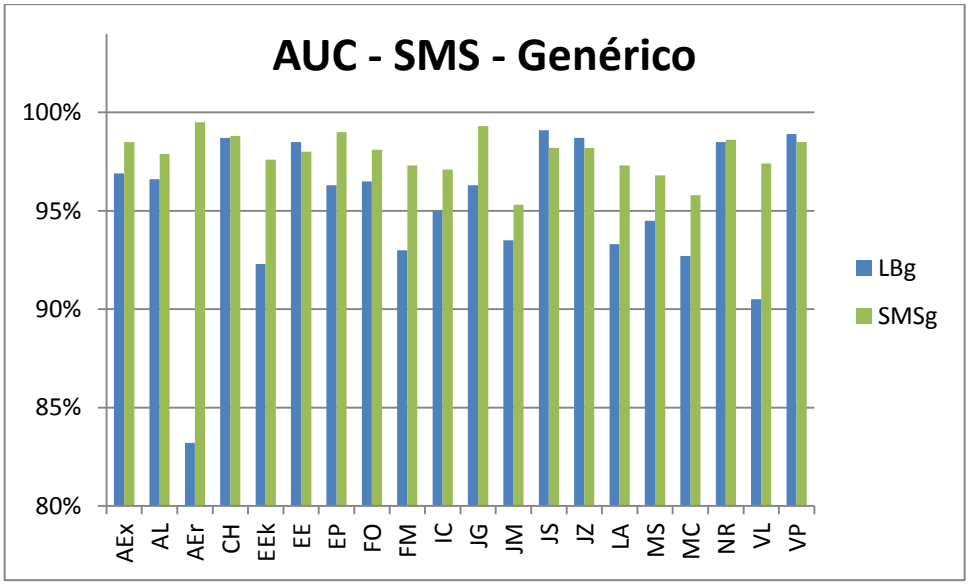


Figura 49: Comparativa de los AUC de LB y SMS en el escenario genérico.

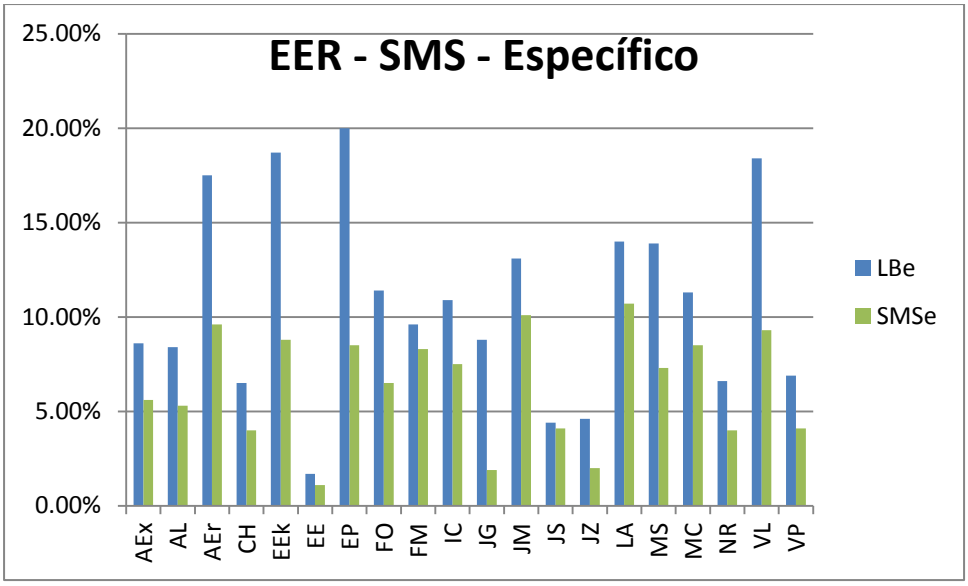


Figura 50: Comparativa de los EER de LB y SMS en el escenario específico.

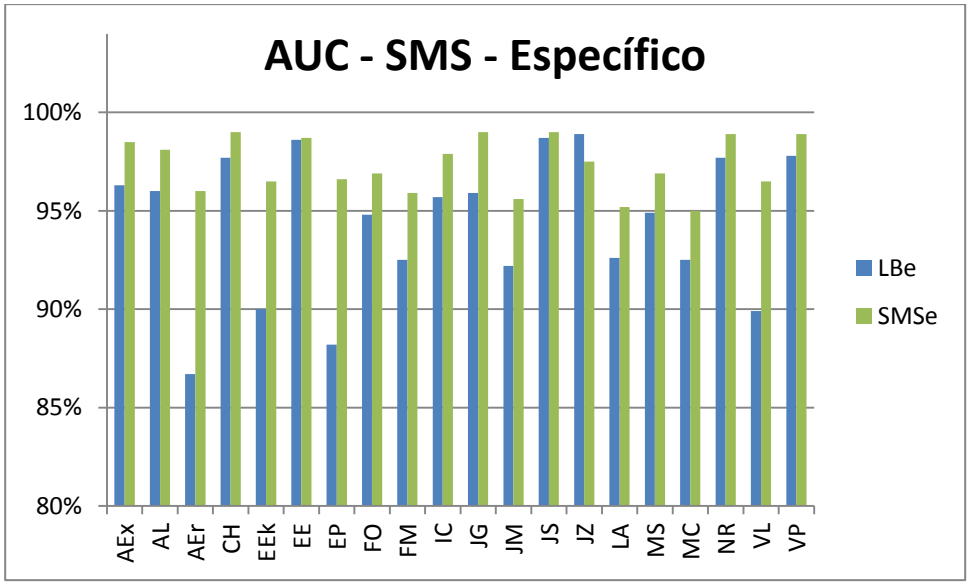


Figura 51: Comparativa de los AUC de LB y SMS en el escenario específico.

5.6 Análisis de los resultados

Pese ha haber utilizado un factor de filtrado del 50% existen diferencias entre EPS y SMS debido a aquellas unidades para las que en el fase de entrenamiento no existía información suficiente y que sin embargo aparecen en la fase de test. Dichas unidades son seleccionadas cuando se aplica el algoritmo de "exclusión de los peores" y sin embargo son excluidas cuando el que se aplica es el de "selección de los mejores".

Algoritmo	EER	AUC
LBg	9,5%	95,2%
EPSg	5,6%	97,6%
SMSg	5,2%	97,9%

Algoritmo	EER	AUC
LBe	10,8%	94,4%
EPSe	7,2%	97,5%
SMSe	6,4%	97,3%

Tabla 18: Comparativa de los resultados promedio.

Los resultados generales de la siguiente tabla demuestran la sustancial mejora que supone la utilización de la selección de segmentos tanto en términos de EER como en términos de ROC en sus valores promedio.

Que se traducen en que EPS produce una mejora en EER de 3,6 puntos en el escenario especifico (3,9 en el genérico) que corresponde a un 33% (41%) sobre la línea base, y con SMS 4,4 puntos (4,3) correspondiente al 50% (70%).

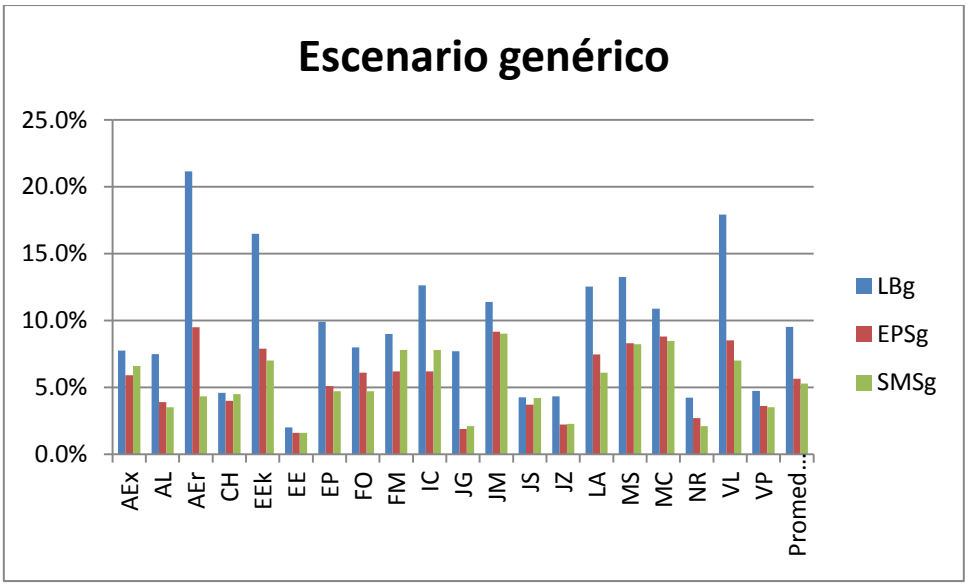


Figura 52: Comparativa de los resultados individuales en el escenario genérico.

Las Figura 52 y Figura 53 resumen los resultados obtenidos que de forma más de tallada se pueden encontrar más arriba; presentan una comparativa de los resultados con los tres algoritmos (línea base, exclusión de los peores y selección de los mejores) para cada usuario. En ellas se puede constatar la consistencia de la mejora aportada.

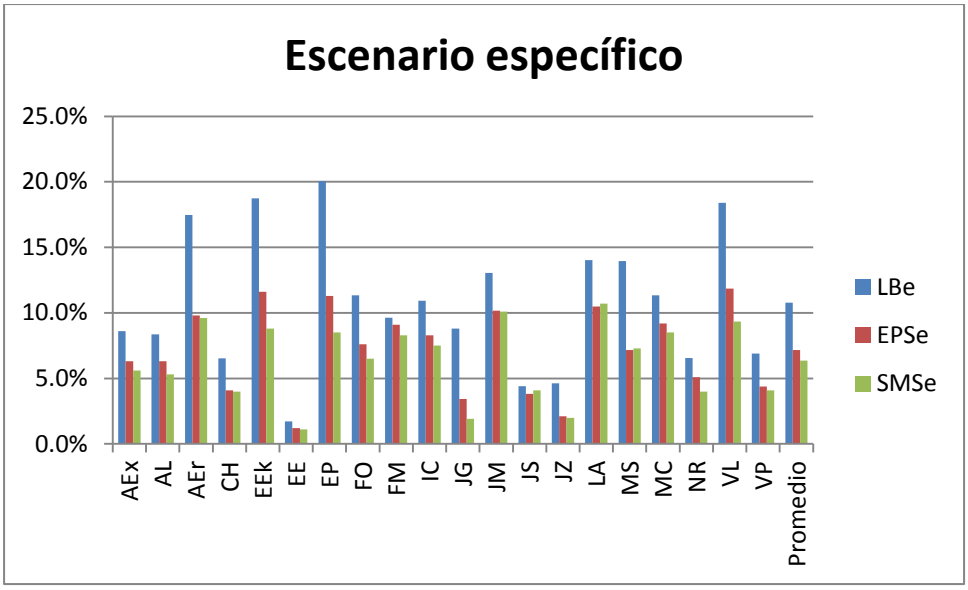


Figura 53: Comparativa de los resultados individuales en el escenario genérico.

Como se ha mencionado los dos algoritmos presentan mayores mejoras para aquellos locutores en los que el algoritmo de la línea base funciona peor, aspecto que puede observarse en el comportamiento de la desviación estándar de los resultados individuales.

Algoritmo	EER	AUC
LBg	5,1%	3,8%
EPSg	2,6%	1,4%
SMSg	2,4%	1,8%

Algoritmo	EER	AUC
LBe	5,2%	3,7%
EPSe	3,4%	1,6%
SMSe	2,9%	1,4%

Tabla 19: Comparativa de las desviaciones estándar de los resultados.

La dispersiones se reducen sensiblemente lo que puede ser interpretado como una medida de la consistencia de la técnica utilizada, ya que no sólo los resultados son mejores sino que también estos son más fácilmente predecibles.

5.7 Estudio del factor de filtrado

Como se ha citado anteriormente para el desarrollo experimental relatado en las secciones previas, se utilizó como límite en la selección de segmentos, "factor de filtrado", el 50% de las unidades con información representativa, queda como final de este trabajo estudiar cuál es la influencia de ese valor en las tasas resultado.

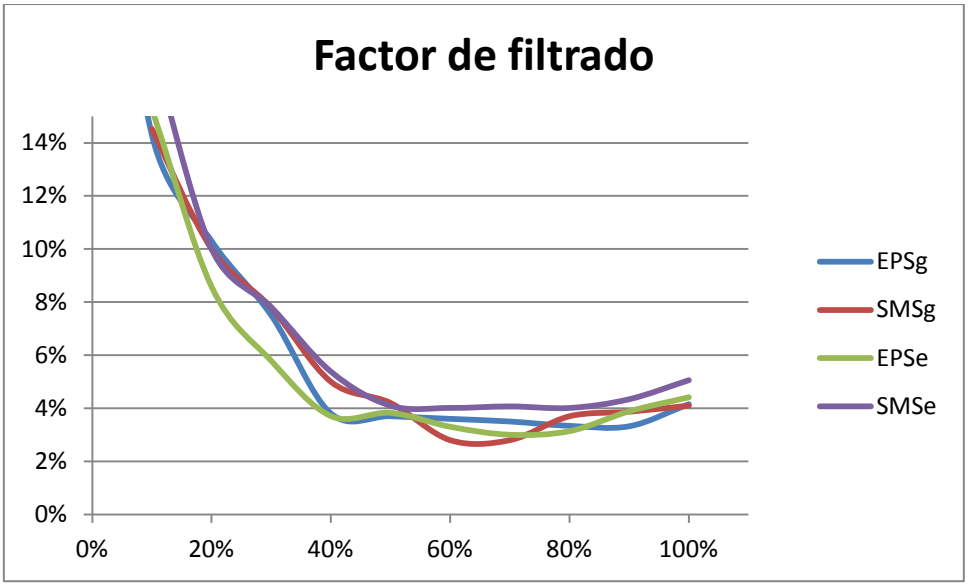


Figura 54: Ejemplo del comportamiento del EER en función factor de filtrado (MC).

Para realizar esta parte del estudio se repitieron los experimentos para once valores distintos del facto de filtrado (0-100%, Δ10%). La Figura 54 muestra un ejemplo del comportamiento de los EER en función de él. Como era de esperar, el 100% en EPS corresponde a la línea base, mientras que 0% en SMS implica la eliminación de toda la información que produce un resultado no computable.

Los resultados pueden analizarse visualmente a partir de las gráficas, apreciándose que para todos los usuarios presenta un comportamiento bastante similar. De dicha inspección visual se puede concluir:

- Que los valores del *factor de filtrado* que minimizan los EER no necesariamente se encuentran en el entorno del 50%.
- Que los resultados aparentemente anómalos presentados en las gráficas de la sección anterior, por ejemplo el del sujeto JS cuyo SMSg empeora la línea base, se explica no con la ausencia de segmentos característicos sino con la ubicación del mínimo, como se observa en la Figura 54 dónde el mínimo del SMG se encuentra en torno al 70% mientras que el 50% supera ligeramente la línea base.

6 Conclusiones y trabajos futuros

6.1 Conclusiones

En el presente trabajo se ha expuesto el estado del arte del reconocimiento biométrico, en el que se encuentra la necesidad de seguir progresando en la mejora de las prestaciones de la modalidad de reconocimiento de locutor, ya que sus tasas de acierto se encuentran lejos de las de otras modalidades actualmente en boga como el reconocimiento del iris o de la huella digital.

Se ha comentado que es posible mejorar en cualquiera de las cinco etapas que componen su proceso (captura, preprocesado, extracción, clasificación y decisión) y que este trabajo se enfoca exclusivamente a la selección del material acústico durante la etapa de captura, optando por analizar, categorizar y optimizar la información que tiene la voz para distinguir a los diferentes locutores.

Para ello ha sido necesario ahondar en el estudio y los procesos involucrados en el reconocimiento a través de la voz como son su producción, su caracterización y su discriminación.

La presente tesis propone en esta línea apuntar en dirección de realizar un prefiltrado de unidades concretas del discurso (Figura 33: Efecto de la selección del material acústico.), apoyándose en el hecho de que la información discriminadora no está uniformemente distribuida a lo largo de él y que se encuentra más representada en segmentos específicos de cada locutor, hecho que se comprueba a partir de los resultados de la experimentación (sección 5).

Con la línea base se han ejemplificado las prestaciones típicas de un proceso estándar de reconocimiento, cuya definición puede encontrarse en la sección 5.1 y que consta de la extracción de características del audio, la construcción de modelos de locutor a partir de los fragmentos de entrenamiento, la evaluación de los fragmentos de test y del análisis del comportamiento del reconocedor para cada modelo mediante el cálculo del EER y el AUC. Todo ello en escenarios genérico y específico (sección 5.6).

El presente trabajo propone como pieza de la algorítmica del reconocimiento del locutor en independencia del texto, la selección del material acústico, u para ello la clasificación de los segmentos de dicho material en las que se han denominado unidades de análisis.

Para la selección de las unidades de análisis adecuadas al fin del trabajo, se han evaluado diferentes alternativas como palabras, fonemas, bifonemas... Descartando

otras por su baja viabilidad práctica, se ha considerando tan sólo el comportamiento de los bifonemas.

Con este objetivo y bajo los criterios ya descritos (véase 4.2.2 y Apéndice I Etiquetación de los fonemas), se han marcado los fragmentos de todo el audio seleccionado, formando las piezas de locución naturales de cada uno de los locutores, se han transcrito estos fragmentos y se han etiquetado fonéticamente sus componentes segmentales. A partir de esta información ha sido posible la identificación y localización de los bifonemas involucrados; resultado éste que constituye una de las aportaciones del presente trabajo: un corpus de más de 10h netas de audio etiquetado.

Se ha verificado que su análisis y utilización para la definición del criterio de prefiltrado, ofrece un aumento de las prestaciones tanto en tasas como en robustez. Esta selección se ha llevado a cabo heurísticamente comprobando sobre el sistema de referencia definido para la línea base el comportamiento de las diferentes unidades e infiriendo, a partir de él, que estas unidades contienen información del locutor que presenta una buena calidad en la discriminación y un nivel de procesamiento optimizado. Lo que constituye otra de las contribuciones de la tesis: los bifonemas son unidades acústicas de discriminación del locutor óptimas en la relación de EER/cómputo y entrenamiento.

Asimismo dentro de los bifonemas se ha visto como algunos presentan un comportamiento diferenciado respecto de otros ante la tarea de la discriminación y caracterización del locutor, siendo esta una contribución principal de la presente investigación y soporte de la hipótesis de esta tesis.

Colateralmente se ha observado como tiende a ser más representativos aquellos sonidos en los que su producción obliga a exponer mayor parte del aparato fonador (fuentes y filtro), hecho predecible ya que como se ha afirmado previamente es tipo de reconocimiento discrimina dichos aparatos en la suposición de que son distintos los de dos individuos cualquiera de la población humana.

Resumiendo los resultados obtenidos se puede establecer que se ha obtenido una sustancial mejora sobre la línea base representativa del proceso estándar, donde el peor resultado lo representa la exclusión de los peores segmentos en un escenario especifico con una mejora del 33%, y el mejor se presenta la selección de los mejores segmentos en un escenario genérico que mejora la línea base en un 45%.

Escenario	EER	AUC	Mejora
LBg	9,5%	95,2%	
EPSg	5,6%	97,6%	41%
SMSg	5,2%	97,9%	45%

Tabla 20: Comparativa de los resultados del escenario genérico.

Escenario	EER	AUC	Mejora
LBe	10,8%	94,4%	
EPSe	7,2%	97,5%	33%
SMSe	6,4%	97,3%	41%

Tabla 21: Comparativa de los resultados del escenario específico.

A la vista de estos datos ha de realizarse una consideración, como era de suponer desde un principio, y simplemente por el hecho de atender al concepto base de la presente tesis, el algoritmo de selección de los mejores ofrece también mejores resultados. Sin embargo el hecho de que en torno al veinte por ciento de las tramas no presentaron suficiente información como para ser consideradas en la experimentación realizada en esta tesis, obligará a evaluar para una implantación y aplicación concreta si es aceptable o no este procedimiento..

En conjunto estos resultados dan pie a plantear una evolución en este tipo de sistemas que incorporen como apoyo un sistema de reconocimiento de fonemas/bifonemas, tecnología ya existente que funciona en el núcleo de la mayoría de los sistemas de reconocimiento del habla (Figura 55).

Con independencia de la mejora que supone esta aproximación general al problema, los resultados de esta tesis pueden ser aprovechados para la construcción de claves orales en sistemas de reconocimiento de locutor dependientes del texto, o bajo texto aleatorio propuesto por la máquina, en la que la elección de estas claves debería estar sometida a los criterios de mejor discriminación individual.

Como resultado del trabajo de investigación expuesto hasta aquí podemos afirmar que para cualquier locutor existen bifonemas dentro del discurso que son más discriminadores de su emisor que otros; que la adecuada utilización de esta información permite mejorar las tasas de reconocimiento de locutores. Dicho lo cual se puede considerar comprobada la hipótesis secundaria #1:

“Los segmentos del discurso no presentan prestaciones uniformes en la caracterización del locutor”.

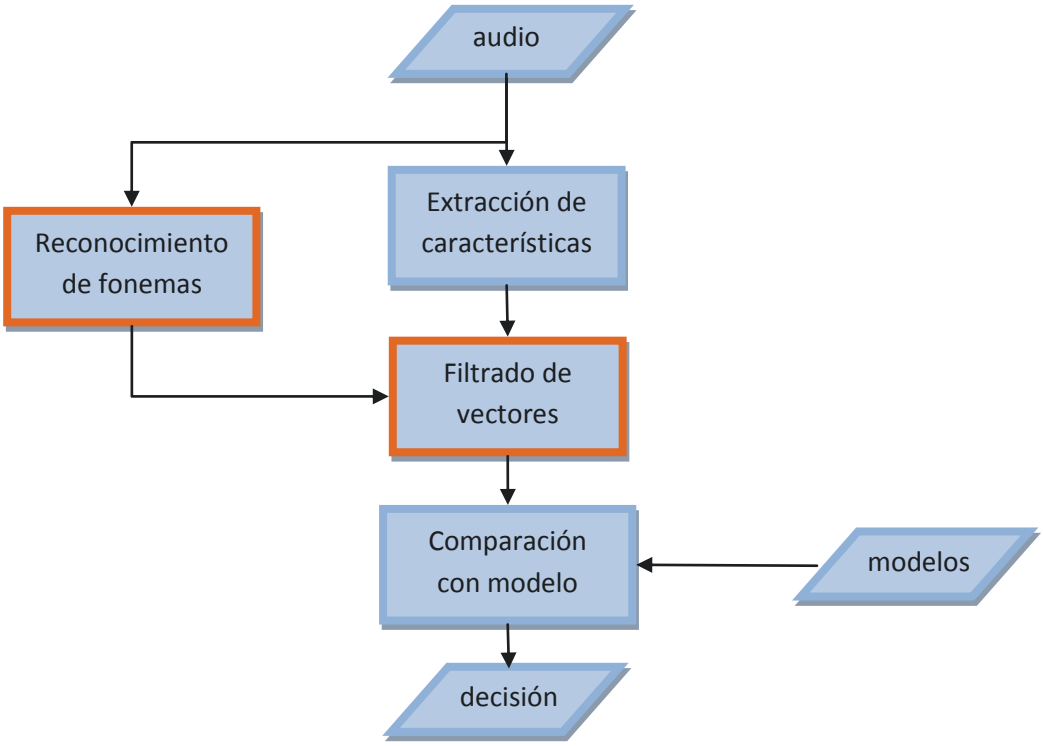


Figura 55: Esquema evolucionado de reconocimiento de locutor.

Se ha presentado en este trabajo un algoritmo simple para ordenar esos bifonemas de acuerdo a su capacidad de discriminación, que no implica modificaciones en la línea del proceso de la fase de entrenamiento. Tan sólo añade al final de la misma la categorización de las unidades, y tampoco altera sustancialmente la fase de reconocimiento ya que sólo inserta una etapa de filtrado del material acústico. El algoritmo está basado en computar las puntuaciones de los vectores que asociados al audio. De lo cual se puede afirmar que la hipótesis secundaria #2 ha quedado verificada:

> **"Es posible determinar qué segmentos del discurso ofrecen mejores prestaciones".**

Se ha propuesto una arquitectura que saca provecho de las conclusiones anteriores, arquitectura que incluye un reconocedor del habla dentro del esquema tradicional. Por tanto puede considerarse que la hipótesis secundaría #3 está confirmada:

> **"Es posible diseñar un modelo de procesamiento del habla para el reconocimiento del locutor que pueda sacar provecho de la identificación de los segmentos característicos".**

Con lo que finalmente podemos afirmar que:

> **"En el discurso existen segmentos que son más característicos del locutor que otros que no lo son tanto".**

que constituye la hipótesis principal de la presente tesis.

Dado el escaso impacto que la propuesta realizada tiene sobre la arquitectura de las dos fases del reconocimiento es razonable suponer que es aplicable con resultados similares a casi todos, si no a todos, los modelos de proceso.

Por tanto y resumiendo las principales contribuciones de la tesis:

- Haber constatado la no uniformidad del discurso en cuanto la densidad de información relativa al locutor.
- Haber formulado un algoritmo de valoración de los segmentos del discurso respecto de su capacidad para representar al locutor.
- Haber propuesto los bifonemas como segmentos adecuados para el antedicho algoritmo, donde la falta de uniformidad es detectable, y haber constatado que dicho algoritmo revela sus diferencias.
- Haber extendido el proceso estándar de reconocimiento para sacar provecho de esta falta de uniformidad, a partir de máquinas presentes ya en el estado de la técnica.

6.2 Contribuciones científicas

Durante la realización de la investigación que se ha presentado en esta memoria, el doctorando ha venido publicando los resultados de su trabajo directa o colateralmente relacionados con el planteamiento expuesto aquí:

6.2.1 Capítulos de libros

2011: Biometrical Fusion – Input Statistical Distribution. Luis Puente, Maria-Jesús Poza, Belén Ruíz, Diego Carrero. Advanced Biometric Technologies, Edited by Girija Chetty and Jucheng Yang p. cm. ISBN 978-953-307-487-0, pp 87-110. Jul 2011.

6.2.2 Congresos

2007 "SIRE: A Semantic Biometric Identity Browsing and Searching System using Multi-modal Fusion". J.M. Gómez, B. Ruiz, M.J. Poza, L. Puente, RIAO'07 (Recherche d'Information Assistée par Ordinateur: Large-Scale

Semantic Access to Content (Text, Image, Video and Sound), Pittsburgh, Pennsylvania, USA. Jun-2007.

2008 "Biometric Authentication Devices and Semantic Web Services: An Approach for Multi Modal Fusion Framework". L.Puente, J.M.Gómez, B.Ruiz, M.J.Poza BIODEVICES 2008: International Conference on Biomedical Electronics and Devices, (Madeira, Portugal, Jan-2008).

2008 "A Knowledge-Based Distributed Biometric Authentication Entity Mechanism". M.J. Poza, L. Puente, D. Carrero and J.M. Gómez: Proceedings of the 2008 International Conference on Security & Management, pp. 660-662. (Las Vegas, Nevada, USA Jul. 2008), ISBN: 1-60132-085-X.

2008 "Identity Authentication Services". J. Poza, L. Puente, B. Ruiz and J.M. Gómez Proceedings of the 2008 International Conference on Security and Management (SAM'08) |pp 655-659, (Las Vegas, Nevada, USA Jul. 2008), ISBN #: 1-60132-085-X.

2008 "Biometric Accreditation Entities: An Approach for Web Accreditation Services". B. Ruiz, L. Puente, M. J. Poza, D. Carrero. SIGMAP-2008. International Conference on Signal Processing and Multimedia Applications. pp 216-220 (Porto. Portugal. July 26-29, 2008)

2008 An Approach for Web Accreditation Services. Poza M.J.,Puente L., Mencke M. and Gómez J.M. Proceedings of Intl Conf. Semantic Web and Web Services (SWWS'08), pp 333-336.Las Vegas (USA), July 14-17 2008. ISBN:1-60132-089-2

2008 "Verificación de Locutores Mediante Modelos de Mezclas Gaussianas (GMM)" D.Carrero, L.Puente, B.Ruiz, M.J.Poza. IV Jornadas de Reconocimiento de Personas pags 103-112. (Valladolid. España. Septiembre 11-12,2008)

2008 "Study of Different Fusion Techniques for Multimodal Biometic Authentication". L. Puente Rodríguez , A. García Crespo, M. J. Poza Lara .B. Ruiz Mezcua. 4th IEEE International Conference on Wireless & Mobile Computing, NetWorking & Comunication (WiMob 2008). pags 666-671 (Avignon. France, Oct-2008).

2008 "Prestaciones de la Verificación de Locutores en Aplicaciones Forenses". D. Carrero, L. Puente, M.J. Poza y B. Ruíz Mezcua: Actas del I INGEFOR. Oct. 2008.

2008 “Evaluación ALBAYZÍN-08 de Sistemas de Verificación de la Lengua: Sistema del Grupo SOFTLAB de la UC3M”. M.J. Poza, B. Ruiz, L. Puente y D. Carrero. Actas de las V Jornadas sobre Tecnologías del Habla, pp. 112-114. Nov. 2008. ISBN: 978-84-9860-169-5.

2009 A Pervasive Biometric Identification Services Platform using Support Vector Machines. Juan Miguel Gómez, Maria Jesús Poza, Belén Ruiz Mezcua, Luis Puente. UBICC, the Ubiquitous Computing and Communication Journal, Volume 4 Number 2, January 2009. ISSN Online 1992-8424. ISSN Print 1994-4608

2009: SVM Speaker Verification System Based on a Low-Cost FPGA. Rafael Ramos, Mariano López, Enrique Cantó and Luis Puente-Rodríguez, 19th International Conference on Field Programmable Logic and Applications (FPL?09), Prague, Czech Republic, pp. 582-586, June 2009. ISBN: 978-1-4244-3892-1.

2009: Implementación mediante FPGA de un sistema SVM de verificación de locutor. Rafael Ramos, Mariano López, Enrique Cantó y Luis Puente-Rodríguez, “”, Jornadas de Computación Reconfigurable y Aplicaciones (JCRA’09), Alcalá de Henares, pp. 99-108, Sep. 2009. ISBN: 978-84-8138-832-9.

2009: Multimodal Biometrics: Topics in Score Fusion. Luis Puente, M. Jesús Poza, Juan Miguel Gómez, and Diego Carrero Computational Intelligence in Security for Information Systems. Proceedings of CISIS’09 2nd International Workshop Burgos, Spain, September 2009, pp. 155-162. ISBN 978-3-642-04090-0.

2010: Prestaciones de la Normalización del Rostro en el Reconocimiento Facial. D. Carrero, B. Ruíz, L. Puente y M.J. Poza. Actas de las V Jornadas de Reconocimiento Biométrico de Personas. Huesca (España), Septiembre 2010, Sep. 2010.

2010: Score Normalization for Multimodal Recognition Systems. Luis Puente, Maria-Jesús Poza, Belén Ruíz, Angel Garcia-Crespo. Journal of Information Assurance and Security, volume 5, 2010, pp. 409-417. ISBN: xxx.

2011: Real-time Subtitle Synchronization in Live Television Programs. Mercedes de Castro, Diego Carrero, Luis Puente, Belen Ruiz. Universidad Carlos III de Madrid. 2011 IEEE International Simposium on Broadband Multimedia Systems and Broadcasting

Ramos Lara, R.R., López García, M., Canto Navarro, E.F., and Puente Rodriguez, L. (2012). Implementación mediante FPGA de un sistema SVM de verificación de locutor.

2013 “SAVIA: Accesibilidad en el sistema hospitalario”, Luis Puente, Belén Ruíz, Alejandro Lozano, Rafael Gálvez, Roberto Peña, International Simposium of Artificial Inteligence and Assistive Technology. VI Congreso Español de Informática

6.2.3 Revistas

2009 A simple score normalization technique for multimodal biometric authentication. Belen Ruiz, Maria-Jesus Poza, Luis Puente, Angel Garcia. International Journal of Biometrics (IJBM), Vol. 1, No. 4, 2009, pp. 374-392. Inderscience Publishers, Geneva, SWITZERLAND. ISSN: 1755-8301 EISSN: 1755-831X doi 10.1504/IJBM. 2009. 027302

2013 Ramos-Lara, R., López-García, M., Cantó-Navarro, E., and Puente-Rodriguez, L. (2013). Real-time speaker verification system implemented on reconfigurable hardware. Journal of Signal Processing Systems 71, 89–103.

6.3 Trabajos futuros

Dado lo expuesto hasta aquí, es razonable preguntarse ante el trabajo aquí expuesto que dado que se ha realizado en base al etiquetado manual de las unidades de análisis, serían trasladables los resultados a una conceptualización más cercana a una aplicación práctica, en la que esta labor manual ya no es posible, y que por tanto fuera la anteriormente propuesta máquina de reconocimiento fonético, quién realizará en tiempo real la selección del material acústico. En este marco es claro que las identificaciones realizadas probablemente no corresponderán a las que una persona podría hacer, eludiendo con toda seguridad las definiciones ortodoxas. Ello sólo supone que las unidades de análisis serían redefinidas sin por ello perder ni un ápice de validez, añadiendo sin embargo cierto nivel de homogeneidad del cuál un operador humano está carente. Resulta por tanto interesante realizar el mismo análisis hecho aquí a partir del etiquetado automático en tiempo real emitido por un reconocedor del habla adaptado al caso, una vez se disponga del modelo fonético adecuado.

Se ha reflejado en la sección 5.7, que uno de los elementos de optimización es la selección correcta del factor de filtrado, queda por estudiar la algorítmica para determinar el facto adecuado a cada locutor que formara parte del modelo individual.

Considerando que se ha utilizado exclusivamente el análisis espectral de la señal de voz, reconociendo el tracto vocal del hablante atendiendo a los sonidos que produce, es lógico proponer considerar esos sonidos no con relación a la articulación del tracto vocal sino a los sonidos producidos por el locutor en su manera personal de emitir elementos suprasegmentales del habla.

Al igual que para grafología la firma, por razones psicosomáticas es mucho más representativa de la persona que otro tipo de escritura manual, parece lógico suponer que por las mismas razones el propio nombre o alguna otra secuencia de palabras con alta carga emotiva para el sujeto aporten mayor información discriminadora que un texto aséptico. Analizar este contenido psíquico aportaría un volumen extra de información al reconocimiento del locutor, pero también analizar las variaciones de estas caracterizaciones podrían ser una aportación útil en el estudio anímico de las personas.

Existen protocolos de diagnóstico basados en el habla, para la detección temprana de enfermedades que provocan alteraciones en ella. La determinación de los grupos fonéticos más afectados por cada una de las dolencias podría ser la base para el desarrollo de algoritmos simplificados de mayor capacidad de resolución.

Se ha citado, a la hora de categorizar las unidades, la aparente tendencia que tienen los fonemas nasales y vocálicos abiertos a situarse entre las mejor valoradas al contrario que ocurre con los más puramente sonidos consonánticos y vocales cerradas. Es posible preguntarse si existe o no un patrón general independiente de locutor que defina el conjunto de unidades óptimo o subóptimo utilizable.

Los criterios de etiquetado expuestos en la sección 4.8 apuntaban a la necesidad de etiquetar los fonemas tal y como se han pronunciado con independencia de cuál era la intención del locutor o la ortodoxia de la pronunciación. Sin embargo elementos característicos de la persona pudieran estar vinculados con la relación existente entre la voluntad del locutor y la realización de la fonación, y elementos culturales podrían estarlo con la relación entre la ortodoxia y la voluntad. Analizar esta relación sobre aquellas unidades de alta categoría presentaría aportaciones en la caracterización tanto de las personas como de los grupos sociales.

7 Referencias

Adler, A. (2003). Sample images can be independently restored from face recognition templates. In Canadian Conference on Electrical and Computer Engineering, 2003. IEEE CCECE 2003, pp. 1163–1166 vol.2.

Auckenthaler, R., Carey, M., and Lloyd-Thomas, H. (2000). Score normalization for text-independent speaker verification systems. Digital Signal Processing *10*, 42–54.

Bailly-Bailliére, E., Bengio, S., Bimbot, F., Hamouz, M., Kittler, J., Mariéthoz, J., Matas, J., Messer, K., Popovici, V., Porée, F., et al. (2003). The BANCA database and evaluation protocol. In Audio-and Video-Based Biometric Person Authentication, pp. 1057–1057.

Beigi, H. (2011). Speaker Recognition. In Biometrics, J. Yang, ed. (InTech),.

Bengio, S., Mariethoz, J., and Marcel, S. (2001). Evaluation of biometric technology on XM2VTS.

Bimbot, F., Bonastre, J.F., Fredouille, C., Gravier, G., Magrin-Chagnolleau, I., Meignier, S., Merlin, T., Ortega-García, J., Petrovska-Delacrétaz, D., and Reynolds, D.A. (2004). A tutorial on text-independent speaker verification. EURASIP Journal on Applied Signal Processing *2004*, 430–451.

Bogert, B.P., Healy, M.J., and Tukey, J.W. (1963). The quefrency alanysis of time series for echoes: Cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking. In Symposium on Time Series Analysis, pp. 209–243.

Bridle, J.S., and Brown, M.D. (1974). An experimental automatic word recognition system. JSRU Report *1003*, 5.

Burges, C.J. (1998). A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery *2*, 121–167.

Campbell, W.M., Sturim, D.E., Reynolds, D.A., and Solomonoff, A. (2006). SVM based speaker verification using a GMM supervector kernel and NAP variability compensation. pp. I–I.

Campbell Jr, J.P. (1997). Speaker recognition: A tutorial. Proceedings of the IEEE *85*, 1437–1462.

Carrero, D., Puente, L., Ruiz-Mezcua, B., and Poza, M.J. (2008a). Prestaciones de la Verificación de Locutores en Aplicaciones Forenses.

Carrero, D., Puente, L., Ruiz-Mezcua, B., and Poza, M.J. (2008b). Verificación de Locutores Mediante Modelos de Mezclas Gaussianas (GMM). 103–112.

Carrero, D., Ruíz, B., Puente, L., and Poza, M.J. (2010). Prestaciones de la Normalización del Rostro en el Reconocimiento Facial. Actas de Las V Jornadas de Renocimiento Biometrico de Las Personas.

Ceyhan, A. (2008). Technologization of security: Management of uncertainty and risk in the age of biometrics. Surveillance & Society *5*, 102–123.

Chakroborty, S., Roy, A., and Saha, G. (2007). Improved closed set text-independent speaker identification by combining MFCC with evidence from flipped filter banks. International Journal of Signal Processing *4*, 114–122.

Zohra Chelali, F. (2011). Amar. Djeradi, Rachida. Djeradi," Speaker Identification System based on PLP Coefficients and Artificial Neural Network. In Proceedings of the World Congress on Engineering, pp. 6–8.

Childers, D.G., and Wong, C.-F. (1994). Measuring and modeling vocal source-tract interaction. IEEE Transactions on Biomedical Engineering *41*, 663 –671.

Childers, D.G., Skinner, D.P., and Kemerait, R.C. (1977). The cepstrum: A guide to processing. Proceedings of the IEEE *65*, 1428–1443.

Crocker, M.J. (1997). Encyclopedia of acoustics (John Wiley).

Cumani, S., and Laface, P. (2012). Analysis of Large-Scale SVM Training Algorithms for Language and Speaker Recognition. IEEE Transactions on Audio, Speech, and Language Processing *20*, 1585–1596.

Dantcheva, A., Velardo, C., D'Angelo, A., and Dugelay, J.L. (2011). Bag of soft biometrics for person identification. Multimedia Tools and Applications *51*, 739–777.

Davis, S., and Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. Acoustics, Speech and Signal Processing, IEEE Transactions on *28*, 357–366.

Dehak, N., Kenny, P.J., Dehak, R., Dumouchel, P., and Ouellet, P. (2011). Front-end factor analysis for speaker verification. Audio, Speech, and Language Processing, IEEE Transactions on *19*, 788–798.

Dellwo, V., Huckvale, M., and Ashby, M. (2007). How is individuality expressed in voice? An introduction to speech production and description for speaker classification. Speaker Classification I 1–20.

Doddington, G. (2001). Speaker recognition based on idiolectal differences between speakers. In Seventh European Conference on Speech Communication and Technology,.

Doddington, G.R., Przybocki, M.A., Martin, A.F., and Reynolds, D.A. (2000). The NIST speaker recognition evaluation – Overview, methodology, systems, results, perspective. Speech Communication *31*, 225–254.

Enos, F., Shriberg, E., Graciarena, M., Hirschberg, J., and Stolcke, A. (2007). Detecting deception using critical segments. In Eighth Annual Conference of the International Speech Communication Association,.

Fant, C.G.M. (1967). Analysis and synthesis of speech processes (North-Holland Publishing Comp.).

Fant, G. (1970). Acoustic Theory of Speech Production (Walter de Gruyter).

Fierrez, J., Ortega-Garcia, J., Torre Toledano, D., and Gonzalez-Rodriguez, J. (2007). Biosec baseline corpus: A multimodal biometric database. Pattern Recognition *40*, 1389–1392.

Furui, S. (1996). An overview of speaker recognition technology. KLUWER INTERNATIONAL SERIES IN ENGINEERING AND COMPUTER SCIENCE 31–56.

Furui, S. (2001). Digital speech processing, synthesis, and recognition (CRC Press).

Ganchev, T. (2011). Contemporary Methods for Speech Parameterization (Springer).

Ganchev, T., Fakotakis, N., and Kokkinakis, G. (2005). Comparative evaluation of various MFCC implementations on the speaker verification task. In Proceedings of the SPECOM, pp. 191–194.

Gómez, J.M., Mezcua, B.R., Poza, M.J., and Puente, L. (2007). SIRE: A Semantic Biometric Identity Browsing and Searching System using Multi-modal Fusion.

Gómez, J.M., Poza, M.J., Ruiz,, B., and Puente, L. (2009). A Pervasive Biometric Identificacion Services Platform using Support Vector Machines". UBICC, the Ubiquituos Computing and Comunication Journal *4*.

Hanilçi, C., and Ertas, F. (2013). Optimizing acoustic features for source cell-phone recognition using speech signals. In Proceedings of the First ACM Workshop on Information Hiding and Multimedia Security, pp. 141–148.

Haykin, S. (1994). Neural networks: a comprehensive foundation (Prentice Hall PTR).

Hermansky, H. (1990). Perceptual linear predictive (PLP) analysis of speech. The Journal of the Acoustical Society of America *87*, 1738.

Hermansky, H., and Morgan, N. (1994). RASTA processing of speech. Speech and Audio Processing, IEEE Transactions on *2*, 578–589.

Hualde, J.I. (2005). The sounds of Spanish (Cambridge University Press).

Itakura, F. (1975). Line Spectrum Representation of Linear Predictor Coefficients of Speech Signals. The Journal of the Acoustical Society of America 57.

Jain, A., Dass, S., and Nandakumar, K. (2004a). Soft Biometric Traits for Personal Recognition Systems. In Biometric Authentication, D. Zhang, and A. Jain, eds. (Springer Berlin / Heidelberg), pp. 1–40.

Jain, A.K., Ross, A., and Prabhakar, S. (2004b). An introduction to biometric recognition. Circuits and Systems for Video Technology, IEEE Transactions on *14*, 4–20.

Jamaati, M., Marvi, H., and Lankarany, M. (2008). Vowels recognition using mellin transform and plp-based feature extraction. Journal of the Acoustical Society of America *123*, 3177.

Jiang, J., Wu, Z., Xu, M., Jia, J., and Cai, L. (2012). Comparison of adaptation methods for GMM-SVM based speech emotion recognition. In 2012 IEEE Spoken Language Technology Workshop (SLT), pp. 269–273.

Kay, S.M. (1998). Fundamentals of Statistical signal processing, Volume 2: Detection theory (Prentice Hall PTR).

Khalifa, O.O., El-Darymli, K.K., Abdullah, A.-H., and Daoud, J.I. (2013). Statistical Modeling for Speech Recognition.

Kulkarni, P.N., and Gadhe, D.L. (2013). Comparison Between SVM & Other Classifiers For SER. International Journal of Engineering *2*.

Lindsay, P.H., and Norman, D.A. (1977). ${$Human Information Processing$}$.

Machlica, L. (2012). High Dimensional Spaces and Modelling in the task of Speaker Recognition.

Marcel, S. (2013). BEAT – biometrics evaluation and testing. Biometric Technology Today *2013*, 5–7.

Martínez Celdrán, E.E. (1996). El sonido en la comunicación humana: introducción a la fonética.

Mason, M.W., Vogt, R.J., Baker, B.J., and Sridharan, S. (2005). Data-driven clustering for blind feature mapping in speaker verification.

Maswadeh, W.M., and Snyder, A.P. (2012). Multivariable and Multigroup Receiver Operating Characteristics Curve Analyses for Qualitative and Quantitative Analysis (DTIC Document).

Mezghani, A., and O'Shaughnessy, D. (2005). Speaker verification using a new representation based on a combination of MFCC and formants. In Canadian Conference on Electrical and Computer Engineering, 2005, pp. 1461–1464.

O'shaughnessy, D. (1987). Speech communication: human and machine (Universities press).

Oppenheim, A.V., and Schafer, R.W. (2004). From frequency to quefrency: a history of the cepstrum. IEEE Signal Processing Magazine *21*, 95–106.

Pelecanos, J., and Sridharan, S. (2001). Feature warping for robust speaker verification.

Phillips, P.J., Martin, A., Wilson, C.L., and Przybocki, M. (2000). An introduction evaluating biometric systems. Computer *33*, 56–63.

Picone, J.W. (1993). Signal modeling techniques in speech recognition. Proceedings of the IEEE *81*, 1215 –1247.

Poza, M.J., Puente, L.A., Carrero, D., and Gómez, J.M. (2008a). A Knowledge-Based Distributed Biometric Authentication Entity Mechanism. In Security and Management, pp. 660–662.

Poza, M.J., Puente, L.A., Ruíz, B., and Gómez, J.M. (2008b). Identity Authentication Services. In Security and Management, pp. 655–659.

Poza, M.J., Puente, L.A., Mencke, M., and Gómez, J.M. (2008c). An Approach for Web Accreditation Services. In SWWS, pp. 333–335.

Poza, M.J., Ruiz, B., Puente, L., and Carrero, D. (2008d). Evaluación Albayzín-08 de Sistemas de Verificación de la Lengua: Sistema del Grupo Softlab de La UC3M. Actas Dela V Jornadas Sober Tecnologías Del Habla 112–114.

Prasad, K., Lotia, P., and Khan, M.R. (2012a). A Review on Text-Independent Speaker Identification Using Gaussian Supervector SVM. International Journal of U-and E-Service, Science and Technology *5*, 71–82.

Prasad, K., Lotia, P., and Khan, M.R. (2012b). A Review on Text-Independent Speaker Identification Using Gaussian Supervector SVM. *5*, 71–82.

Przybocki, M., and Martin, A.F. (2004). NIST speaker recognition evaluation chronicles. In ODYSSEY04-The Speaker and Language Recognition Workshop,.

Puente, L., Crespo, A.G., Lara, M., and Mezcua, B.R. (2008a). Study of different fusion techniques for multimodal biometric authentication. In Networking and Communications, 2008. WIMOB'08. IEEE International Conference on Wireless and Mobile Computing,, pp. 666–671.

Puente, L., Poza, M.J., Gómez, J.M., and Carrero, D. (2009). Multimodal Biometrics: Topics in Score Fusion. In Computational Intelligence in Security for Information Systems, (Springer), pp. 155–162.

Puente, L., Poza, M.J., Ruíz, B., and Carrero, D. (2011a). Biometrical Fusion–Input Statistical Distribution.

Puente, L., Poza, M.J., Ruiz, B., and Crespo, A.G. (2011b). Score Normalization for Multimodal Recognition Systems. Jouernal of Information Assurance and Security 87–110.

Puente, L.A., Poza, M.J., Gómez, J.M., and Ruíz-Mezcua, B. (2008b). Biometric Authentication Devices and Semantic Web Services-An Approach for Multi Modal Fusion Framework. In BIODEVICES (1), pp. 95–100.

Ramaiah, V.S., and Rao, R.R. (2013). Modeling Speaker Specific F Modeling Speaker Specific Features for Automatic Text eatures for Automatic Text eatures for Automatic Text Independent Speaker Independent Speaker Tracking System u Tracking System u Tracking System using Support Vector Machines (SVMs).

Ramos Lara, R.R., López García, M., Canto Navarro, E.F., and Puente Rodriguez, L. (2012). Implementación mediante FPGA de un sistema SVM de verificación de locutor.

Ramos-Lara, R., López-García, M., Cantó-Navarro, E., and Puente-Rodriguez, L. (2009). SVM speaker verification system based on a low-cost FPGA. In Field Programmable Logic and Applications, 2009. FPL 2009. International Conference on, pp. 582–586.

Ramos-Lara, R., López-García, M., Cantó-Navarro, E., and Puente-Rodriguez, L. (2013). Real-time speaker verification system implemented on reconfigurable hardware. Journal of Signal Processing Systems *71*, 89–103.

Reynolds, D. (2006). Tutorial on SuperSID. JHU 2002 Workshop.In: JHU 2002 Workshop. Retrieved December,2006 (2002).

Reynolds, D.A. (1997). Comparison of background normalization methods for text-independent speaker verification. In Eurospeech,.

Reynolds, D.A. (2002). An overview of automatic speaker recognition technology. In Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on, pp. IV–4072.

Reynolds, D.A. (2003). Channel robust speaker verification via feature mapping. In Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on, pp. II–53.

Reynolds, D.A., and Rose, R.C. (1995). Robust text-independent speaker identification using Gaussian mixture speaker models. IEEE Transactions on Speech and Audio Processing *3*, 72–83.

Reynolds, D.A., Quatieri, T.F., and Dunn, R.B. (2000). Speaker verification using adapted Gaussian mixture models. Digital Signal Processing *10*, 19–41.

Rhodes, H.T.F. (1956). Alphonse Bertillon, father of scientific detection (Abelard-Schuman).

Rubio-Ayuso, A., and Hernández-Rioja, I. (2005). Libro blanco de las tecnologías del habla (Antonio Rubio Ayuso).

Ruiz, B., Poza, M.J., Puente, L., and Garcia, A. (2009). A simple score normalisation technique for multimodal biometric authentication. International Journal of Biometrics *1*, 374–392.

Ruiz,, B., Poza, M.J., and Carrero, D. (2008). Biometric Accreditation Entities: An Approach for Web Accreditation Services. Proceeding of SIGMAP-2008.

Ruiz-Mezcua, B. (1998). Modelado estadístico y conexionista para reconocimiento de locutores con aprendizaje de la variabilidad temporal del habla. Unpublished Doctoral Dissertation, Escuela Técnica Superior de Ingenieros de Telecomunicación–Universidad Politécnica de Madrid, Spain.

Sanchez, R. (2009). PIBES: Algoritmos biométricos y metodología de evaluación.

Sanchez, R. (2010). CENIT: SEGUR@ - Seguridad y confianza en la sociedad de la información.

Schultz, T. (2007). Speaker characteristics. Speaker Classification I 47–74.

Shriberg, E., Ferrer, L., Kajarekar, S., Venkataraman, A., and Stolcke, A. (2005). Modeling prosodic feature sequences for speaker recognition. Speech Communication *46*, 455–472.

Singh, L., Chetty, G., and Singh, S. (2012a). A NOVEL ALGORITHM USING MFCC AND ERB GAMMATONE FILTERS IN SPEECH RECOGNITION.

Singh, Y.N., Singh, S.K., and Ray, A.K. (2012b). Bioelectrical Signals as Emerging Biometrics: Issues and Challenges.

Skowronski, M.D., and Harris, J.G. (2002). Increased mfcc filter bandwidth for noise-robust phoneme recognition. In 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp. I–801–I–804.

Skowronski, M.D., and Harris, J.G. (2003). Improving the filter bank of a classic speech feature extraction algorithm. In Circuits and Systems, 2003. ISCAS'03. Proceedings of the 2003 International Symposium on, pp. IV–281.

Solera Ureña, R. (2011). Máquinas de vectores soporte para reconocimiento robusto de habla.

Stevens, K.N. (2007). Models of speech production. Encyclopedia of Acoustics, Volume Four 1565–1578.

Stevens, S.S., Volkmann, J., and Newman, E.B. (1937). A Scale for the Measurement of the Psychological Magnitude Pitch. The Journal of the Acoustical Society of America *8*, 185–190.

Stolcke, A., Shriberg, E., Ferrer, L., Kajarekar, S., Sonmez, K., and Tur, G. (2007). Speech recognition as feature extraction for speaker recognition. In Signal Processing Applications for Public Security and Forensics, 2007. SAFE'07. IEEE Workshop on, pp. 1–5.

Streiner, D.L. (2013). A Guide for the Statistically Perplexed: Selected Readings for Clinical Researchers (University of Toronto Press).

Tapiador, M., and Singüenza, J.A. (2005). Tecnologías biométricas aplicadas a la seguridad. México, DF: Alfaomega.

Tiwari, V. (2010). MFCC and its applications in speaker recognition. International Journal on Emerging Technologies *1*, 19–22.

Togneri, R., and Pullella, D. (2011). An Overview of Speaker Identification: Accuracy and Robustness Issues. IEEE Circuits and Systems Magazine *11*, 23–61.

Torres, B., Antonio, J., and Casado Morente, J.C. (2007). Anatomía funcional de la voz. Medicina Del Canto. Capítulo *1*.

Vapnik, V. (1999). The nature of statistical learning theory (springer).

Vapnik, V., and Lerner, A.J. (1963). Generalized portrait method for pattern recognition. Automation and Remote Control *24*, 774–780.

Wallace, R., McLaren, M., McCool, C., and Marcel, S. (2012). Cross-pollination of normalization techniques from speaker to face authentication using gaussian mixture models. Information Forensics and Security, IEEE Transactions on *7*, 553–562.

Wayman, J.L., Jain, A.K., Maltoni, D., and Maio, D. (2004). Biometric systems: Technology, design and performance evaluation (Springer).

Wong, E., and Sridharan, S. (2001). Comparison of linear prediction cepstrum coefficients and mel-frequency cepstrum coefficients for language identification. In Intelligent Multimedia, Video and Speech Processing, 2001. Proceedings of 2001 International Symposium on, pp. 95–98.

Xiang, B., Chaudhari, U.V., Navratil, J., Ramaswamy, G.N., and Gopinath, R.A. (2002). Short-time Gaussianization for robust speaker verification. In 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp. I–681–I–684.

Zergat, K.Y., and Amrouche, A. (2013). Robust Support Vector Machines for Speaker Verification Task.

Zetterholm, E. (2007). Detection of speaker characteristics using voice imitation. Speaker Classification II 192–205.

Zhang, X., Liu, X., and Wang, Z.J. (2013). Evaluation of a set of new ORF kernel functions of SVM for speech recognition. Engineering Applications of Artificial Intelligence.

Zwicker, E. (1961). Subdivision of the audible frequency range into critical bands (Frequenzgruppen). The Journal of the Acoustical Society of America *33*, 248–248.

Apéndice I. Etiquetación de los fonemas

Durante el proceso de construcción del corpus se ha tomado un criterio de etiquetación de fonemas simple, adecuado a los fines de esta tesis en base a las siguientes normas:

- El etiquetado corresponde al fonema pronunciado, ignorando cual en correcta fonación debería haber sido emitido.
- Sólo se consideran los fonemas y no sus alófonos.
- Los símbolos que representen cada fonema serán lo más próximos a la grafía que les corresponde y siempre dentro del código ASCII.

Finalmente se ha utilizado los símbolos que se representan en la siguiente tabla:

símbolo	utilización
a	a
b	b, v
c	ch
d	d
e	e
f	f
g	g(a), gu(e), gu(i) g(o), g(u), g(ü)
i	i, y{vocálica}
j	j, g(e), g(i), h{aspirada}
k	k, qu, c(a), c(o), c(u),c(consonante)
l	l
L	ll
m	m

símbolo	utilización
n	n
~	ñ
o	o
p	p
r	r {simple}
R	rr{múltiple}
s	s
t	t
u	u
x	x
y	y {consonante}
z	z, c(e), c(i)

Acompañando a estos símbolos se han utilizado también

<: Para indicar el comienzo de una emisión, en otras palabras un silencio delante de un fonema.

>: Para indicar el final de una emisión.

$: Para marcar una vocalización ininteligible.

Apéndice II. Catalogación de las unidades

A continuación se presenta el listado de la catalogación de las unidades de cada uno de los usuarios.

a. Usuario AEx

[e-p]	1.17/171	[R->]	0.96/140	[g-a]	0.84/44	[a-j]	0.79/126
[R-k]	1.15/87	[a->]	0.96/582	[e-~]	0.84/36	[d-a]	0.78/209
[u-z]	1.14/15	[i-t]	0.96/111	[i-b]	0.84/28	[p-R]	0.78/175
[a-f]	1.13/31	[n-k]	0.96/64	[o-e]	0.83/138	[p-i]	0.78/63
[e-t]	1.13/144	[j-u]	0.95/44	[j-a]	0.83/82	[i-j]	0.78/26
[t-u]	1.13/53	[d-e]	0.95/413	[l-d]	0.83/29	[l-a]	0.78/317
[a-t]	1.11/168	[e-y]	0.94/23	[n->]	0.83/131	[s-m]	0.77/50
[l-p]	1.07/107	[o-z]	0.94/56	[m-p]	0.83/95	[e-s]	0.77/1068
[o-t]	1.07/49	[t-o]	0.93/447	[s-a]	0.83/403	[r-e]	0.77/88
[a-p]	1.06/285	[z-e]	0.93/270	[<-e]	0.82/337	[j-e]	0.77/175
[l-u]	1.06/49	[m-u]	0.93/46	[n-m]	0.82/40	[d-i]	0.77/163
[u-p]	1.06/25	[n-f]	0.93/34	[j-o]	0.82/66	[<-i]	0.77/302
[p-e]	1.05/214	[a-k]	0.93/314	[R-a]	0.82/115	[u-r]	0.77/12
[o-k]	1.05/294	[f-u]	0.93/48	[a-s]	0.82/513	[x-i]	0.76/34
[e-n]	1.05/800	[a-n]	0.93/296	[k-a]	0.82/254	[o-j]	0.76/74
[n-t]	1.04/578	[n-z]	0.92/125	[n-e]	0.82/180	[e-j]	0.76/85
[o-p]	1.04/128	[s-d]	0.92/41	[s-e]	0.81/649	[l-s]	0.76/123
[l-m]	1.03/21	[e-z]	0.92/191	[e-m]	0.81/285	[i-s]	0.76/338
[z-a]	1.03/46	[y-a]	0.91/43	[<-d]	0.81/143	[e-e]	0.75/234
[a-z]	1.02/171	[n-d]	0.91/156	[k-o]	0.81/439	[R-s]	0.75/31
[f-a]	1.01/19	[i-u]	0.91/19	[o-n]	0.81/446	[k-u]	0.75/171
[s-u]	1.01/174	[p-u]	0.91/105	[p-a]	0.81/321	[b-i]	0.75/99
[t-e]	1.01/510	[l-o]	0.89/217	[R-e]	0.80/221	[n-a]	0.74/243
[e-f]	1.00/57	[R-t]	0.88/126	[t-i]	0.80/307	[~-o]	0.73/95
[s-p]	1.00/101	[z-i]	0.88/401	[i-e]	0.80/283	[y-o]	0.73/48
[u-t]	0.99/111	[a-a]	0.88/98	[i-m]	0.80/147	[a-R]	0.73/230
[t-R]	0.99/96	[u-s]	0.88/53	[o-i]	0.80/55	[m-i]	0.73/110
[k-t]	0.99/30	[i-n]	0.87/168	[s-k]	0.80/125	[b-s]	0.73/49
[i-k]	0.99/264	[a-m]	0.86/212	[g-u]	0.79/60	[u-c]	0.72/19
[e-k]	0.99/188	[g-i]	0.86/15	[m-e]	0.79/257	[b-a]	0.72/101
[i-z]	0.98/140	[r-a]	0.86/233	[a-e]	0.79/128	[a-u]	0.71/53
[u-n]	0.98/303	[m-o]	0.86/223	[e-a]	0.79/234	[e-r]	0.71/215
[c-o]	0.98/88	[i-r]	0.86/14	[t-a]	0.79/514	[o-m]	0.71/270
[o->]	0.98/562	[n-l]	0.86/35	[u-e]	0.79/120	[d-o]	0.71/319
[<-b]	0.98/30	[s-n]	0.86/13	[n-u]	0.79/19	[u-l]	0.71/41
[l->]	0.98/145	[k-i]	0.85/130	[e-i]	0.79/109	[o-s]	0.71/777
[R-z]	0.97/19	[k-e]	0.85/579	[s-o]	0.79/348	[e-o]	0.70/85

[i-i]	0.70/79	[a-o]	0.65/128	[R-j]	0.54/30
[<-m]	0.70/99	[r-o]	0.65/99	[b-u]	0.51/30
[e->]	0.70/392	[e-R]	0.64/249	[e-b]	0.49/78
[<-a]	0.70/201	[o-b]	0.63/83	[R-o]	0.48/74
[i-a]	0.70/226	[e-d]	0.63/109	[o-u]	0.44/28
[a-b]	0.69/196	[i-g]	0.62/27		
[b-l]	0.69/53	[n-s]	0.62/52		
[e-l]	0.69/407	[R-m]	0.62/31		
[o-d]	0.69/119	[<-n]	0.62/135		
[i-o]	0.68/183	[<-o]	0.61/42		
[u-b]	0.68/29	[e-u]	0.61/63		
[o-R]	0.68/236	[R-u]	0.60/34		
[n-g]	0.68/30	[o-o]	0.60/66		
[b-o]	0.67/42	[o-r]	0.60/82		
[i-d]	0.67/167	[a-d]	0.60/353		
[m-a]	0.67/339	[j-i]	0.59/17		
[o-a]	0.67/137	[o-l]	0.59/203		
[p-o]	0.67/267	[n-o]	0.58/334		
[a-l]	0.67/441	[a-r]	0.57/166		
[l-i]	0.67/111	[n-i]	0.56/105		
[s->]	0.66/418	[g-o]	0.54/98		

b. Usuario AL

[u-i]	1.07/21	[a-b]	0.82/58	[a-o]	0.75/27	[e-i]	0.70/30
[f-e]	1.02/12	[n-z]	0.81/50	[a-p]	0.75/51	[m-o]	0.68/156
[s-l]	1.02/11	[l->]	0.81/37	[i-e]	0.74/151	[i-i]	0.68/30
[p-l]	1.00/16	[e-n]	0.80/393	[a-t]	0.74/87	[l-o]	0.68/130
[l-m]	0.99/22	[a-d]	0.79/135	[n-a]	0.74/123	[a-z]	0.67/109
[R-s]	0.99/11	[u-a]	0.79/36	[a-i]	0.73/73	[n-k]	0.66/21
[a-g]	0.93/26	[i-d]	0.79/67	[<-u]	0.73/11	[o-z]	0.66/39
[m-e]	0.92/70	[l-a]	0.79/231	[l-s]	0.73/18	[r-a]	0.66/54
[i-m]	0.91/33	[a-m]	0.79/202	[t-i]	0.73/121	[e-a]	0.66/93
[<-i]	0.90/51	[R-a]	0.78/55	[o-e]	0.73/63	[R-o]	0.65/39
[i-z]	0.89/53	[g-a]	0.78/86	[i-n]	0.73/96	[e-l]	0.65/296
[R-k]	0.89/84	[e-r]	0.78/55	[<-l]	0.72/32	[s-t]	0.64/161
[u-e]	0.87/23	[e-z]	0.78/57	[e-p]	0.72/44	[n-p]	0.64/24
[d->]	0.86/13	[d-a]	0.77/114	[n->]	0.72/102	[s-k]	0.63/115
[o-a]	0.85/38	[a-~]	0.77/14	[e-d]	0.72/58	[n-f]	0.63/33
[n-j]	0.84/26	[b-a]	0.77/51	[n-e]	0.71/181	[o-n]	0.63/242
[R-l]	0.84/25	[m-i]	0.77/67	[t-R]	0.71/58	[o-t]	0.62/62
[u-p]	0.84/20	[e-f]	0.77/51	[u-n]	0.71/127	[l-e]	0.61/93
[t-a]	0.84/200	[<-n]	0.77/56	[s-e]	0.71/291	[i-g]	0.61/38
[a-n]	0.84/236	[R-d]	0.77/27	[o-m]	0.71/56	[m-a]	0.61/183
[r-o]	0.83/36	[f-i]	0.76/49	[c-a]	0.70/12	[o-b]	0.61/46
[e-m]	0.83/157	[b-e]	0.76/68	[s-d]	0.70/28	[e-s]	0.60/568
[z-e]	0.83/78	[z-i]	0.75/164	[a-e]	0.70/60	[d-e]	0.60/244

[b-o]	0.60/44	[n-i]	0.52/62	[p-R]	0.44/33	[i->]	0.30/15
[a-s]	0.60/292	[<-e]	0.52/143	[e-k]	0.43/149	[o-R]	0.28/139
[o-s]	0.60/436	[i-t]	0.51/77	[o-d]	0.43/52	[o-k]	0.28/160
[i-s]	0.60/117	[<-m]	0.51/26	[i-l]	0.43/51	[e-e]	0.27/82
[i-R]	0.59/42	[k-u]	0.51/56	[s-o]	0.43/145	[z-a]	0.27/30
[<-s]	0.59/31	[u-d]	0.51/36	[a->]	0.43/192	[l-t]	0.26/29
[i-p]	0.58/40	[i-a]	0.51/107	[l-k]	0.42/30	[b-l]	0.25/19
[s->]	0.58/199	[e->]	0.50/183	[o-l]	0.42/99	[m-u]	0.24/23
[<-p]	0.57/39	[p-a]	0.50/93	[u-l]	0.42/22	[l-i]	0.23/60
[n-o]	0.56/184	[s-i]	0.50/198	[t-u]	0.41/32	[r-e]	0.21/54
[o->]	0.56/185	[R-t]	0.49/30	[p-u]	0.41/22	[g-o]	0.21/67
[n-d]	0.56/118	[p-o]	0.49/212	[b-i]	0.40/41	[a-j]	0.20/39
[R->]	0.56/62	[s-u]	0.49/102	[e-g]	0.40/64	[y-o]	0.18/47
[<-t]	0.56/39	[q-o]	0.48/57	[t-o]	0.39/146	[a-a]	0.13/81
[s-p]	0.55/105	[a-l]	0.48/236	[o-i]	0.39/43	[m-p]	0.12/11
[a-r]	0.55/51	[d-i]	0.48/63	[R-e]	0.38/85	[e-c]	0.12/24
[e-t]	0.55/49	[a-k]	0.48/96	[n-s]	0.38/85	[<-b]	0.08/26
[j-e]	0.55/70	[g-u]	0.48/47	[o-p]	0.38/97	[<-y]	0.05/40
[k-e]	0.54/385	[n-t]	0.47/161	[e-u]	0.37/39	[i-k]	0.03/57
[i-o]	0.54/104	[t-e]	0.47/218	[k-o]	0.35/198	[R-i]	0.01/22
[a-q]	0.54/79	[r-i]	0.47/27	[a-R]	0.34/127	[l-d]	-0.16/47
[o-o]	0.54/20	[k-a]	0.46/54	[f-R]	0.34/13	[o-r]	-0.17/58
[f-u]	0.54/45	[<-a]	0.45/114	[k-i]	0.34/53		
[a-u]	0.53/28	[<-k]	0.45/21	[i-b]	0.33/34		
[p-e]	0.53/129	[s-m]	0.45/32	[<-d]	0.32/40		
[s-a]	0.53/180	[e-R]	0.44/122	[j-o]	0.32/59		
[g-i]	0.52/19	[d-o]	0.44/117	[e-b]	0.32/28		

c. Usuario AEr

[n-p]	1.47/21	[e-i]	0.79/19	[n-k]	0.66/64	[i-a]	0.59/148
[<-n]	1.22/30	[R->]	0.77/32	[o-f]	0.64/17	[i-o]	0.58/98
[<-p]	1.10/33	[d-e]	0.76/350	[o-d]	0.64/103	[o->]	0.58/140
[R-m]	1.09/18	[u-c]	0.75/19	[o-k]	0.64/181	[z-i]	0.58/146
[u-m]	1.03/19	[s-p]	0.73/73	[p-e]	0.64/49	[d-a]	0.58/133
[s-n]	1.03/27	[m-b]	0.73/24	[a-o]	0.64/67	[i-n]	0.58/74
[e-d]	1.02/70	[e-o]	0.72/39	[e-a]	0.63/98	[s-d]	0.57/34
[e-e]	1.01/173	[e-f]	0.72/36	[o-s]	0.62/416	[l-f]	0.57/27
[n-m]	1.00/25	[o-m]	0.72/61	[R-s]	0.62/27	[u-n]	0.57/76
[l-i]	1.00/27	[R-k]	0.72/20	[i-b]	0.62/31	[t-i]	0.57/96
[n-b]	0.97/11	[s->]	0.70/232	[<-i]	0.62/75	[n->]	0.56/47
[n-u]	0.93/15	[<-e]	0.70/171	[m-a]	0.61/95	[r-e]	0.56/61
[m-p]	0.89/20	[n-t]	0.70/134	[u-s]	0.61/11	[n-i]	0.56/61
[l-s]	0.86/11	[o-j]	0.69/26	[i-e]	0.60/91	[s-a]	0.55/104
[m-u]	0.85/47	[i-f]	0.69/13	[a-s]	0.60/219	[e-b]	0.55/38
[z-o]	0.83/12	[l-p]	0.68/24	[m-o]	0.60/73	[u-a]	0.55/20
[t-e]	0.83/145	[e-s]	0.67/533	[n-e]	0.60/106	[n-l]	0.55/28
[n-d]	0.81/119	[f-i]	0.66/15	[m-i]	0.59/78	[i-r]	0.54/28

[e-u]	0.54/38	[o-o]	0.48/17	[b-a]	0.41/78	[o-a]	0.25/80
[o-t]	0.54/48	[e-r]	0.47/64	[a-j]	0.40/66	[p-a]	0.24/125
[o-n]	0.54/227	[i-z]	0.47/27	[i-t]	0.39/35	[i-k]	0.24/72
[u-d]	0.54/33	[u-e]	0.47/39	[l-d]	0.39/22	[k-R]	0.23/25
[d-i]	0.54/96	[R-l]	0.47/33	[<-o]	0.39/12	[f-o]	0.23/50
[e-n]	0.54/340	[a-z]	0.47/83	[s-i]	0.39/188	[e-k]	0.23/92
[b-l]	0.54/37	[e-m]	0.47/106	[a->]	0.39/229	[l-a]	0.22/208
[i-m]	0.54/85	[b-o]	0.46/39	[a-d]	0.39/132	[p-l]	0.21/11
[i-s]	0.53/110	[e-t]	0.46/113	[n-a]	0.39/74	[R-d]	0.21/31
[j-e]	0.53/54	[n-o]	0.46/180	[a-~]	0.38/46	[t-R]	0.20/98
[i-d]	0.53/42	[<-k]	0.46/51	[e-l]	0.37/180	[f-e]	0.19/12
[i-j]	0.53/35	[a-e]	0.46/70	[e->]	0.37/131	[k-u]	0.18/57
[r-o]	0.53/34	[d-o]	0.46/102	[l-o]	0.36/158	[e-L]	0.18/22
[a-i]	0.53/21	[<-d]	0.46/57	[R-e]	0.35/137	[e-p]	0.17/40
[s-e]	0.53/242	[t-o]	0.46/215	[R-i]	0.35/56	[a-r]	0.16/82
[R-a]	0.52/64	[z-k]	0.45/13	[f-u]	0.35/30	[o-e]	0.16/75
[r-i]	0.52/48	[s-o]	0.45/152	[o-l]	0.35/81	[L-o]	0.15/23
[a-n]	0.52/219	[b-R]	0.45/15	[<-y]	0.34/18	[o-b]	0.15/74
[u-i]	0.52/24	[p-i]	0.45/14	[o-r]	0.34/37	[p-o]	0.15/63
[l-m]	0.52/11	[s-u]	0.45/46	[n-z]	0.33/36	[y-o]	0.13/46
[p-R]	0.51/69	[l->]	0.44/40	[g-a]	0.33/14	[R-o]	0.12/57
[k-e]	0.51/239	[a-p]	0.44/79	[g-o]	0.32/36	[a-l]	0.10/137
[a-t]	0.50/125	[o-R]	0.43/76	[b-i]	0.32/72	[a-k]	0.10/84
[k-a]	0.50/122	[j-o]	0.43/47	[a-b]	0.31/74	[i-l]	0.10/37
[<-l]	0.50/40	[m-e]	0.43/124	[e-z]	0.31/101	[e-g]	0.09/31
[e-j]	0.50/18	[b-e]	0.43/60	[s-t]	0.29/141	[s-b]	0.06/14
[<-a]	0.50/65	[z-e]	0.43/83	[s-m]	0.29/29	[a-y]	-0.02/20
[t-a]	0.49/173	[i-R]	0.43/38	[c-o]	0.27/20		
[o-p]	0.48/58	[t-u]	0.42/21	[p-u]	0.27/33		
[R-t]	0.48/44	[a-m]	0.42/100	[e-R]	0.27/94		
[s-k]	0.48/119	[r-a]	0.42/77	[n-s]	0.26/32		
[i-p]	0.48/13	[l-e]	0.41/61	[k-o]	0.25/218		

d. Usuario CH

[u-p]	1.46/22	[R-k]	1.27/65	[b-e]	1.18/41	[u-t]	1.13/18
[o-p]	1.46/27	[o-o]	1.25/15	[m->]	1.18/39	[n->]	1.13/31
[l-t]	1.41/12	[a-n]	1.24/220	[k-a]	1.17/185	[g-u]	1.12/22
[p-a]	1.40/78	[e-k]	1.23/170	[o-k]	1.17/96	[R-i]	1.12/14
[l-p]	1.39/115	[s-d]	1.23/57	[t-e]	1.16/186	[s-m]	1.11/47
[s-l]	1.38/25	[p-e]	1.21/166	[R-d]	1.16/14	[a-t]	1.11/149
[f-e]	1.33/28	[u-n]	1.21/106	[e-t]	1.16/109	[f-i]	1.11/44
[n-d]	1.31/68	[<-m]	1.21/45	[l-m]	1.14/18	[e-n]	1.11/298
[j-a]	1.30/35	[s-o]	1.20/147	[a-m]	1.14/209	[k-e]	1.11/309
[i-n]	1.29/66	[t-a]	1.20/195	[j-o]	1.14/26	[m-e]	1.10/109
[c-o]	1.29/47	[e-p]	1.19/112	[i-k]	1.14/69	[z-a]	1.10/13
[n-p]	1.27/31	[p-o]	1.18/99	[a-p]	1.14/111	[i-b]	1.10/32

[f-o]	1.10/37	[l-d]	0.98/39	[s->]	0.88/109	[e-c]	0.78/27
[e-m]	1.09/97	[r-a]	0.98/112	[e-u]	0.88/12	[a-j]	0.78/22
[o-t]	1.09/25	[a-i]	0.98/41	[k-u]	0.88/35	[p-R]	0.78/80
[c-a]	1.09/21	[i-t]	0.98/55	[e-s]	0.88/512	[n-g]	0.78/20
[z-e]	1.09/77	[m-a]	0.97/103	[b-l]	0.87/14	[~-a]	0.77/41
[e->]	1.09/18	[s-n]	0.96/21	[o-s]	0.87/294	[a-R]	0.77/109
[R-m]	1.08/21	[n-l]	0.96/44	[b-s]	0.87/15	[u-c]	0.76/37
[a-r]	1.08/40	[s-t]	0.96/165	[k-i]	0.87/24	[o-b]	0.76/27
[s-a]	1.06/189	[n-o]	0.96/167	[R-u]	0.87/14	[u-r]	0.75/20
[l-k]	1.06/36	[e-b]	0.95/52	[R-o]	0.87/36	[d-o]	0.74/136
[s-p]	1.06/58	[b-i]	0.95/53	[i-m]	0.86/50	[a-b]	0.74/90
[l-u]	1.05/33	[n-u]	0.95/24	[a-z]	0.86/106	[R-e]	0.74/141
[l->]	1.05/18	[l-o]	0.95/84	[e-l]	0.86/308	[n-s]	0.74/39
[n-i]	1.05/29	[o-R]	0.95/74	[t-i]	0.86/171	[u-z]	0.73/24
[m-p]	1.05/56	[n-a]	0.95/176	[n-k]	0.86/66	[i-e]	0.73/141
[m-u]	1.05/49	[l-i]	0.94/58	[<-i]	0.86/51	[i-s]	0.73/73
[b-o]	1.04/29	[r-o]	0.94/31	[e-f]	0.86/58	[e-r]	0.72/73
[e-a]	1.04/57	[s-k]	0.94/42	[e-o]	0.86/29	[l-s]	0.72/14
[t-R]	1.04/80	[n-e]	0.93/154	[r-e]	0.86/15	[m-b]	0.71/11
[k-t]	1.04/39	[o-j]	0.93/15	[e-i]	0.85/35	[l-e]	0.71/136
[<-p]	1.04/16	[j-e]	0.92/71	[s-i]	0.84/172	[<-b]	0.71/14
[n-t]	1.04/166	[i-a]	0.92/162	[o->]	0.84/209	[o-r]	0.70/32
[i-o]	1.03/75	[o-m]	0.92/80	[o-a]	0.84/30	[i-R]	0.70/47
[a-k]	1.03/169	[g-a]	0.92/25	[o-d]	0.83/72	[a-e]	0.70/67
[l-a]	1.03/318	[o-l]	0.92/53	[e-d]	0.83/79	[o-u]	0.69/14
[d-a]	1.03/151	[u-e]	0.92/56	[<-e]	0.83/88	[i-l]	0.69/24
[t-o]	1.03/176	[a-a]	0.91/69	[d-i]	0.82/48	[a->]	0.69/123
[m-o]	1.02/78	[d-e]	0.91/310	[a-d]	0.81/198	[~-o]	0.68/34
[b-a]	1.02/85	[a-s]	0.91/366	[i-j]	0.81/29	[<-d]	0.63/29
[k-o]	1.01/235	[j-i]	0.91/62	[o-e]	0.81/49	[r-i]	0.62/39
[a-u]	1.00/20	[e-j]	0.91/87	[R-z]	0.81/18	[e-~]	0.62/26
[<-a]	1.00/34	[R-t]	0.90/59	[<-k]	0.80/27	[e-g]	0.61/41
[i-d]	1.00/149	[z-i]	0.90/186	[R-a]	0.80/108	[i-g]	0.60/12
[e-e]	1.00/139	[<-n]	0.90/24	[a-l]	0.80/256		
[p-u]	0.99/48	[R-s]	0.89/28	[e-R]	0.79/100		
[o-n]	0.99/274	[i-z]	0.89/42	[n-m]	0.79/17		
[s-e]	0.99/313	[g-o]	0.89/21	[e-z]	0.79/62		
[t-u]	0.99/24	[i-i]	0.88/27	[o-z]	0.79/20		

e. Usuario EEk

[m-p]	1.56/15	[e-p]	1.28/35	[a-t]	1.21/30	[i->]	1.15/45
[n-b]	1.50/21	[i-p]	1.27/16	[p-u]	1.20/20	[s-m]	1.14/25
[R-p]	1.41/14	[f-i]	1.26/12	[i-t]	1.19/60	[o-k]	1.14/123
[<-n]	1.35/12	[t-e]	1.26/103	[<-p]	1.18/13	[o-t]	1.14/18
[a-p]	1.29/53	[e-t]	1.25/56	[a->]	1.16/82	[b-o]	1.12/16

[o-r]	1.11/18	[e->]	0.95/64	[s-i]	0.84/117	[<-s]	0.74/30
[i-r]	1.10/12	[s-o]	0.95/74	[a-d]	0.84/54	[o-R]	0.74/88
[e-o]	1.08/19	[<-e]	0.94/54	[d-e]	0.84/99	[j-o]	0.73/34
[n-t]	1.07/120	[l->]	0.94/14	[e-m]	0.84/74	[a-a]	0.72/30
[R-n]	1.07/22	[o-b]	0.93/15	[u-m]	0.84/25	[n-e]	0.72/56
[i-z]	1.06/41	[t-o]	0.93/112	[d-i]	0.84/57	[s-t]	0.71/77
[R-b]	1.06/11	[k-e]	0.92/199	[k-a]	0.84/48	[g-a]	0.71/52
[z-e]	1.06/69	[n-o]	0.92/64	[u-n]	0.83/76	[n-k]	0.71/14
[m-b]	1.06/16	[m-e]	0.92/73	[o-z]	0.83/23	[t-R]	0.71/12
[R->]	1.06/40	[o-n]	0.92/104	[e-z]	0.83/60	[i-a]	0.69/79
[c-o]	1.05/16	[z-i]	0.91/142	[o-e]	0.83/34	[a-l]	0.68/75
[n-n]	1.05/15	[s-p]	0.91/21	[i-n]	0.83/78	[k-R]	0.68/41
[f-e]	1.05/45	[i-s]	0.91/57	[l-i]	0.83/54	[u-d]	0.67/16
[<-d]	1.05/17	[k-u]	0.91/24	[l-u]	0.82/13	[s-k]	0.67/31
[k-t]	1.02/19	[i-l]	0.91/22	[n-z]	0.82/21	[R-l]	0.67/24
[s-u]	1.02/30	[p-e]	0.91/25	[u-e]	0.81/33	[u-l]	0.66/35
[d-o]	1.01/69	[o->]	0.91/135	[a-m]	0.81/63	[l-e]	0.66/19
[l-k]	1.01/32	[a-n]	0.90/156	[<-y]	0.81/26	[r-a]	0.63/35
[b-e]	1.00/57	[b-a]	0.89/40	[k-i]	0.81/106	[a-j]	0.61/14
[i-o]	1.00/85	[n-d]	0.89/87	[s->]	0.80/57	[n-u]	0.61/32
[m-a]	1.00/33	[<-i]	0.89/52	[k-o]	0.79/83	[u-a]	0.61/11
[e-n]	1.00/193	[o-p]	0.89/50	[e-r]	0.78/56	[a-g]	0.59/31
[d-u]	1.00/40	[R-t]	0.88/31	[l-a]	0.78/115	[o-l]	0.58/67
[e-f]	1.00/50	[R-i]	0.88/36	[o-s]	0.78/96	[e-d]	0.58/62
[<-u]	1.00/14	[i-b]	0.88/12	[e-s]	0.78/235	[s-a]	0.56/25
[a-b]	0.99/18	[i-k]	0.88/61	[a-R]	0.78/88	[n-s]	0.54/14
[e-a]	0.99/55	[a-e]	0.87/43	[b-i]	0.78/21	[l-t]	0.54/19
[o-d]	0.99/32	[e-R]	0.87/122	[o-m]	0.77/52	[g-R]	0.52/16
[e-k]	0.99/82	[m-o]	0.87/66	[m-i]	0.77/50	[o-a]	0.52/25
[t-a]	0.98/101	[<-l]	0.86/22	[a-s]	0.77/126	[u-s]	0.51/14
[l-p]	0.98/21	[t-i]	0.86/138	[e-l]	0.77/154	[k-l]	0.50/27
[s-e]	0.98/137	[R-k]	0.86/39	[n-a]	0.77/102	[e-u]	0.49/33
[e-g]	0.97/34	[a-r]	0.86/27	[n-i]	0.77/25	[e-e]	0.47/17
[y-o]	0.97/44	[R-e]	0.86/100	[a-y]	0.76/13	[r-i]	0.42/17
[a-k]	0.97/63	[a-i]	0.86/48	[e-b]	0.76/42	[e-j]	0.36/19
[n->]	0.96/21	[p-a]	0.86/67	[R-a]	0.76/24		
[i-R]	0.96/15	[l-o]	0.86/80	[d-a]	0.76/51		
[i-d]	0.96/24	[<-k]	0.85/15	[p-o]	0.76/96		
[e-i]	0.96/29	[i-m]	0.85/17	[a-z]	0.75/80		
[i-e]	0.95/141	[r-e]	0.85/62	[<-a]	0.75/42		

f. Usuario EE

[y-i]	1.84/11	[i-f]	1.26/20	[l-a]	1.20/560	[f-a]	1.14/33
[b-o]	1.42/60	[a-o]	1.24/91	[e-f]	1.17/112	[i-y]	1.12/23
[z-k]	1.38/11	[o-p]	1.24/234	[r-o]	1.16/174	[p-o]	1.12/379
[n-f]	1.33/23	[j-u]	1.23/11	[p-r]	1.16/18	[<-p]	1.11/81
[l-p]	1.33/143	[s-g]	1.23/14	[i-L]	1.16/25	[z->]	1.11/14
[a-c]	1.28/19	[d-a]	1.22/307	[r-a]	1.14/258	[g-R]	1.11/20

[x-i]	1.10/18	[R->]	0.94/164	[b-e]	0.82/186	[l-s]	0.73/40
[R-a]	1.10/227	[a-r]	0.93/186	[t-i]	0.82/291	[k-s]	0.73/12
[y-a]	1.10/114	[s-a]	0.93/425	[e-d]	0.82/210	[m-e]	0.72/295
[r-i]	1.10/97	[r-p]	0.92/19	[o-z]	0.82/167	[p-l]	0.72/86
[j-a]	1.10/24	[f-o]	0.92/54	[t-u]	0.82/94	[s-p]	0.72/136
[l-o]	1.09/356	[l-e]	0.92/339	[c-e]	0.82/15	[t-R]	0.72/147
[a-f]	1.09/50	[u-t]	0.92/71	[o->]	0.82/979	[e-R]	0.72/291
[<-d]	1.07/40	[o-i]	0.91/77	[k-u]	0.81/193	[s-y]	0.70/54
[i-r]	1.06/24	[k-e]	0.91/878	[p-u]	0.81/144	[a-j]	0.70/143
[z-e]	1.06/176	[e-g]	0.91/111	[o-s]	0.81/813	[i-l]	0.70/94
[n-s]	1.04/137	[i-d]	0.91/316	[R-k]	0.81/124	[i-o]	0.70/203
[z-a]	1.04/63	[e-r]	0.91/246	[i-p]	0.81/43	[s-t]	0.69/594
[d-e]	1.03/612	[p-R]	0.90/136	[e-t]	0.81/180	[a-s]	0.69/710
[n-l]	1.03/51	[u-a]	0.90/135	[R-e]	0.80/292	[d->]	0.69/23
[s->]	1.02/352	[p-p]	0.90/11	[n-t]	0.80/667	[s-k]	0.68/206
[o-u]	1.01/38	[a-t]	0.90/226	[R-f]	0.80/11	[n-k]	0.68/110
[b-a]	1.01/224	[l-i]	0.90/263	[i-s]	0.80/365	[a-~]	0.67/77
[e-o]	1.01/157	[<-k]	0.90/164	[<-a]	0.80/215	[m-a]	0.66/355
[l-d]	1.01/64	[e-p]	0.89/284	[R-t]	0.80/122	[e->]	0.66/539
[e-z]	1.01/215	[i-b]	0.89/40	[l-m]	0.80/61	[a-g]	0.65/37
[e-e]	1.01/89	[<-s]	0.89/169	[e-b]	0.79/115	[a-R]	0.64/381
[o-t]	1.00/197	[j-i]	0.88/77	[a-b]	0.79/260	[c-o]	0.64/111
[a-y]	1.00/50	[a-i]	0.88/164	[e-x]	0.79/27	[o-c]	0.63/13
[t-a]	1.00/691	[u-e]	0.88/136	[p-e]	0.78/260	[n-b]	0.63/36
[e-k]	0.99/368	[u-s]	0.88/126	[o-e]	0.78/154	[m-p]	0.63/139
[<-l]	0.99/91	[g-e]	0.88/13	[f-e]	0.78/75	[i-e]	0.62/349
[o-a]	0.99/176	[i-t]	0.87/226	[d-R]	0.78/15	[p-z]	0.62/11
[s-b]	0.99/26	[d-o]	0.87/447	[o-d]	0.78/239	[n-p]	0.61/48
[a-e]	0.98/136	[a-p]	0.87/217	[a-d]	0.78/450	[l-k]	0.61/39
[R-o]	0.98/191	[i-R]	0.86/60	[e-s]	0.78/1282	[s-s]	0.61/21
[e-y]	0.97/45	[a-z]	0.85/195	[s-e]	0.77/456	[a->]	0.61/590
[o-k]	0.97/616	[i-a]	0.85/452	[l->]	0.77/127	[R-p]	0.59/21
[k-a]	0.97/341	[e-a]	0.85/202	[o-l]	0.77/335	[l-t]	0.56/64
[y-o]	0.96/260	[s-f]	0.85/12	[r-t]	0.76/14	[i->]	0.56/56
[i-z]	0.96/148	[t-e]	0.85/600	[s-u]	0.76/210	[s-d]	0.55/45
[u->]	0.95/37	[j-o]	0.85/82	[n-z]	0.75/88	[n-e]	0.53/287
[o-y]	0.95/39	[a-k]	0.84/349	[u-d]	0.75/44	[c-a]	0.52/30
[i-k]	0.95/265	[z-i]	0.84/355	[<-e]	0.75/375	[k-l]	0.48/75
[r-e]	0.95/107	[k-t]	0.84/67	[n->]	0.75/104	[~-a]	0.34/77
[u-p]	0.95/61	[g-a]	0.83/54	[a-a]	0.75/76		
[p-i]	0.94/24	[r-k]	0.83/28	[a-l]	0.74/629		
[j-e]	0.94/118	[o-r]	0.83/130	[e-l]	0.74/541		
[<-R]	0.94/17	[s-o]	0.83/311	[n-y]	0.73/93		
[p-a]	0.94/329	[t-o]	0.82/606	[n-a]	0.73/250		

g. Usuario EP

[i-n]	1.68/13	[s-l]	1.42/12	[a-i]	1.35/21	[o-p]	1.33/32
[k-u]	1.54/16	[u-m]	1.36/14	[t-a]	1.34/69	[p-e]	1.33/15

[a-n]	1.30/68	[r-a]	1.03/23	[R-e]	0.91/78	[n-o]	0.76/32
[e-n]	1.19/102	[t-o]	1.02/57	[i-a]	0.90/62	[i-s]	0.76/58
[a-t]	1.18/12	[i->]	1.01/12	[e-R]	0.90/21	[g-a]	0.75/13
[e-m]	1.18/19	[a-r]	1.01/30	[<-e]	0.89/21	[k-i]	0.75/14
[m-a]	1.14/36	[i-d]	1.01/40	[n-s]	0.89/34	[a->]	0.75/17
[i-m]	1.13/31	[a-z]	1.01/67	[e-e]	0.89/21	[e-p]	0.74/33
[s-d]	1.12/23	[a-e]	1.00/44	[a-s]	0.88/90	[d-a]	0.73/52
[s-o]	1.12/64	[e->]	1.00/29	[<-s]	0.88/21	[a-g]	0.73/12
[n-t]	1.12/47	[a-a]	1.00/27	[l-a]	0.87/50	[o-d]	0.73/31
[a-m]	1.10/34	[r-i]	0.99/11	[o-e]	0.87/24	[t-e]	0.72/39
[p-a]	1.10/69	[n-d]	0.99/48	[n-i]	0.87/22	[i-k]	0.72/26
[a-k]	1.09/75	[e-a]	0.98/34	[i-l]	0.87/19	[r-e]	0.70/12
[R-n]	1.09/17	[m-b]	0.98/13	[j-i]	0.86/23	[<-i]	0.68/15
[z-e]	1.09/40	[e-z]	0.97/64	[o-s]	0.84/104	[s-k]	0.67/28
[s-u]	1.08/66	[k-l]	0.97/15	[n-a]	0.83/61	[u-s]	0.67/15
[a-o]	1.08/11	[e-r]	0.96/14	[b-i]	0.83/28	[o-z]	0.63/37
[l-o]	1.08/21	[l-e]	0.95/17	[n-l]	0.83/16	[o-a]	0.62/30
[i-o]	1.08/38	[d-o]	0.94/68	[e-s]	0.82/154	[n-e]	0.59/26
[s-p]	1.07/26	[o-k]	0.94/13	[d-e]	0.81/83	[i-t]	0.58/16
[u-n]	1.07/60	[R->]	0.94/11	[m-u]	0.81/12	[e-j]	0.58/31
[s-e]	1.07/85	[m-e]	0.93/43	[a-d]	0.81/103	[m-i]	0.45/21
[a-p]	1.06/22	[l-u]	0.93/22	[k-o]	0.81/45	[<-m]	0.41/24
[e-u]	1.06/31	[o-b]	0.93/14	[i-e]	0.79/42	[g-R]	0.39/14
[e-l]	1.06/42	[s-t]	0.93/49	[k-a]	0.79/44	[o-m]	0.34/26
[o->]	1.06/42	[e-g]	0.92/28	[e-t]	0.79/19	[i-R]	0.29/13
[s-i]	1.06/86	[s->]	0.92/34	[r-o]	0.78/31	[o-R]	0.24/15
[g-o]	1.05/14	[l-i]	0.92/18	[e-k]	0.78/42		
[b-R]	1.05/16	[o-n]	0.92/71	[a-l]	0.78/73		
[t-i]	1.04/20	[p-o]	0.91/35	[n-z]	0.78/16		
[m-p]	1.04/29	[k-e]	0.91/60	[a-R]	0.77/36		
[s-a]	1.03/66	[z-i]	0.91/117	[<-a]	0.77/11		

h. Usuario FO

[n-p]	1.47/21	[u-R]	0.98/19	[<-b]	0.88/25	[e-g]	0.80/35
[u-n]	1.34/122	[r-i]	0.98/60	[o-n]	0.86/173	[R-l]	0.78/30
[q-o]	1.21/55	[m-o]	0.97/79	[e-x]	0.86/12	[z-e]	0.78/102
[u-d]	1.21/15	[a-q]	0.96/90	[i-e]	0.85/103	[l-k]	0.78/29
[d->]	1.20/28	[m-p]	0.95/48	[e-o]	0.84/65	[R-k]	0.78/58
[<-p]	1.15/29	[n-i]	0.94/21	[e-a]	0.84/63	[o-k]	0.77/216
[e-e]	1.08/43	[o-z]	0.94/26	[i-j]	0.82/23	[e-i]	0.77/23
[o-m]	1.06/71	[<-n]	0.93/35	[n-e]	0.82/119	[n-l]	0.76/30
[<-m]	1.05/153	[s-k]	0.93/14	[l-e]	0.82/45	[c-o]	0.76/11
[n-a]	1.02/125	[R-n]	0.91/26	[f-u]	0.82/15	[s->]	0.75/231
[m-e]	1.01/150	[i-m]	0.91/28	[e-L]	0.81/11	[e-u]	0.75/55
[k-t]	1.00/24	[m-u]	0.90/19	[s-m]	0.81/71	[o-a]	0.75/31
[l-m]	0.99/25	[o-p]	0.90/70	[o-e]	0.81/49	[a-z]	0.74/95
[a-m]	0.99/118	[e-m]	0.89/161	[<-k]	0.81/32	[j-o]	0.73/33
[i-c]	0.98/11	[y-o]	0.89/29	[l-t]	0.80/21	[d-i]	0.73/81

[i-b]	0.73/14	[R-e]	0.59/119	[k-o]	0.49/128	[b-a]	0.36/40
[b-e]	0.73/34	[j-u]	0.59/23	[<-e]	0.49/370	[f-o]	0.36/34
[i-n]	0.72/68	[p-R]	0.59/38	[e-d]	0.49/59	[d-e]	0.35/281
[<-a]	0.72/64	[i-a]	0.58/139	[e-n]	0.48/361	[g-i]	0.35/14
[o-u]	0.72/25	[a-i]	0.58/33	[a-j]	0.48/61	[f-a]	0.35/14
[a->]	0.71/190	[b-l]	0.58/38	[l-i]	0.48/38	[s-o]	0.34/85
[e-s]	0.71/460	[e-l]	0.57/230	[r-e]	0.48/68	[a-l]	0.33/115
[z-a]	0.70/43	[a-t]	0.57/26	[l-p]	0.48/50	[s-d]	0.33/39
[<-l]	0.70/145	[e-r]	0.57/95	[<-i]	0.47/101	[a-o]	0.33/49
[r-o]	0.69/75	[o-t]	0.56/47	[l-a]	0.47/267	[<-y]	0.32/29
[l-d]	0.67/14	[<-s]	0.56/64	[t-a]	0.47/169	[s-t]	0.32/120
[o-d]	0.67/63	[s-p]	0.56/84	[n-k]	0.46/45	[k-a]	0.31/73
[o->]	0.67/329	[o-s]	0.56/209	[p-a]	0.46/187	[a-k]	0.31/104
[o-R]	0.66/61	[<-o]	0.56/14	[t-i]	0.45/86	[u-r]	0.30/27
[t-R]	0.66/71	[l-o]	0.56/138	[k-R]	0.45/28	[k-u]	0.28/49
[o-l]	0.65/70	[s-u]	0.55/51	[u-b]	0.44/38	[a-b]	0.28/124
[n-o]	0.65/101	[a-r]	0.55/80	[e-b]	0.43/14	[o-r]	0.26/52
[u-a]	0.65/31	[e-z]	0.55/110	[m-a]	0.43/106	[p-l]	0.26/17
[<-u]	0.64/66	[r-a]	0.54/100	[s-a]	0.43/112	[b-u]	0.25/13
[i-t]	0.64/23	[e-j]	0.53/27	[i-d]	0.42/69	[j-e]	0.24/94
[p-e]	0.63/108	[n-t]	0.53/201	[i-k]	0.42/48	[b-o]	0.24/30
[n->]	0.63/114	[a-g]	0.53/18	[d-a]	0.41/79	[a-R]	0.21/73
[e->]	0.63/243	[t-o]	0.52/108	[a-d]	0.41/226	[a-e]	0.18/78
[R-o]	0.63/37	[a-n]	0.52/132	[c-a]	0.40/40	[R->]	0.16/79
[i-l]	0.62/19	[p-o]	0.52/88	[z-i]	0.40/171	[q-a]	0.16/46
[R-a]	0.61/44	[a-s]	0.52/167	[l->]	0.39/47	[i-z]	0.15/53
[t-e]	0.61/174	[R-m]	0.51/27	[i-o]	0.39/109	[f-i]	0.15/20
[a-p]	0.61/81	[s-i]	0.51/166	[k-l]	0.39/11	[u-e]	0.12/21
[k-i]	0.61/40	[n-d]	0.51/93	[o-b]	0.39/51	[i-p]	-0.01/23
[n-z]	0.61/38	[e-f]	0.51/79	[n-s]	0.38/69	[f-e]	-0.11/32
[d-o]	0.60/149	[m-i]	0.50/49	[b-i]	0.37/69		
[e-t]	0.60/71	[t-u]	0.50/30	[e-R]	0.37/98		
[s-e]	0.60/190	[e-p]	0.50/97	[i-s]	0.37/104		
[e-k]	0.60/79	[p-u]	0.50/54	[i-R]	0.37/22		
[k-e]	0.60/288	[o-o]	0.50/42	[g-o]	0.37/39		

i. Usuario FM

[m-t]	1.19/14	[r-d]	0.83/15	[r-t]	0.77/15	[u-k]	0.73/12
[x-t]	1.10/18	[i-p]	0.83/93	[R-p]	0.76/80	[a-t]	0.72/313
[e-~]	0.92/19	[R-m]	0.81/90	[u-f]	0.76/17	[j-i]	0.72/63
[s-s]	0.85/20	[l-m]	0.80/32	[k-z]	0.76/17	[a-~]	0.72/122
[l-y]	0.85/12	[r-k]	0.79/11	[i-a]	0.75/575	[m-a]	0.71/582
[o-o]	0.84/127	[R-n]	0.78/156	[b-i]	0.75/319	[l-s]	0.69/85
[d-d]	0.84/38	[t-r]	0.78/19	[k-a]	0.74/578	[i-z]	0.69/182
[r-n]	0.84/13	[y-i]	0.77/15	[a-p]	0.74/335	[s-n]	0.68/162
[p-p]	0.83/14	[n-p]	0.77/158	[n-a]	0.73/721	[i-n]	0.68/341

[r-i]	0.68/225	[R-t]	0.57/206	[l-k]	0.50/116	[e-t]	0.41/404
[e-a]	0.67/441	[u-r]	0.57/61	[j-a]	0.50/101	[m-u]	0.41/142
[i-e]	0.67/517	[l-t]	0.57/81	[n-z]	0.50/207	[e-c]	0.41/36
[u-m]	0.67/40	[a-o]	0.57/159	[<-t]	0.50/30	[z-e]	0.41/375
[t-a]	0.66/975	[s-y]	0.56/43	[r-a]	0.50/278	[e-s]	0.41/2780
[y-e]	0.66/75	[b-e]	0.56/189	[k-e]	0.49/1316	[d-a]	0.41/497
[x-i]	0.66/36	[R-z]	0.56/43	[s-k]	0.49/526	[a-s]	0.40/1237
[l-p]	0.66/150	[g-l]	0.56/41	[m-p]	0.49/224	[~-o]	0.40/84
[z-k]	0.65/43	[g-a]	0.56/216	[a-m]	0.49/526	[R-e]	0.40/578
[o-t]	0.65/283	[i-r]	0.55/54	[p-a]	0.49/608	[e-i]	0.40/117
[a-L]	0.65/25	[l-i]	0.55/227	[o->]	0.49/791	[s-i]	0.40/850
[a-a]	0.65/246	[u-p]	0.55/44	[n-y]	0.48/53	[s-a]	0.39/639
[g-i]	0.65/23	[n-e]	0.55/547	[l-b]	0.47/32	[u-e]	0.38/250
[k-l]	0.64/60	[f-o]	0.55/83	[n-o]	0.47/1050	[R-s]	0.38/78
[b-a]	0.64/283	[t-i]	0.54/444	[i-u]	0.47/108	[o-s]	0.37/1475
[o-k]	0.64/642	[i-t]	0.54/235	[p-i]	0.47/33	[y-o]	0.37/231
[e-x]	0.64/107	[r-R]	0.54/11	[d-o]	0.47/644	[n-d]	0.37/425
[n-t]	0.63/775	[e-e]	0.54/400	[e-z]	0.47/351	[e-m]	0.37/592
[a-x]	0.63/20	[f-u]	0.54/58	[a->]	0.47/755	[m-i]	0.37/342
[o-n]	0.63/953	[t-e]	0.53/765	[m-o]	0.46/445	[e-g]	0.37/150
[R-i]	0.63/159	[p-e]	0.53/442	[a-k]	0.46/435	[L-i]	0.36/22
[s-g]	0.63/51	[a-g]	0.53/158	[l-r]	0.46/14	[u-a]	0.35/88
[e-p]	0.62/451	[a-e]	0.53/473	[L-a]	0.46/46	[g-R]	0.35/101
[a-z]	0.62/492	[o-p]	0.53/330	[s-t]	0.46/643	[e-R]	0.34/524
[m-e]	0.62/355	[z-i]	0.53/1040	[s-m]	0.46/131	[a-d]	0.33/759
[u-z]	0.62/61	[n-k]	0.53/169	[o-a]	0.46/255	[p-r]	0.31/11
[y-a]	0.61/152	[a-r]	0.53/339	[k-t]	0.45/71	[a-R]	0.30/515
[z-a]	0.61/126	[u-n]	0.53/522	[o-e]	0.45/558	[<-k]	0.30/88
[e-n]	0.61/1471	[b-o]	0.52/132	[e-l]	0.45/807	[<-p]	0.26/81
[n-m]	0.61/47	[o-z]	0.52/197	[u-t]	0.45/85	[e-r]	0.24/337
[l-a]	0.61/989	[a-n]	0.52/907	[g-u]	0.45/92	[n->]	0.21/151
[i-k]	0.60/399	[i-L]	0.52/41	[x-p]	0.45/30	[e->]	0.19/282
[R-a]	0.60/383	[R-b]	0.52/19	[i-f]	0.44/29	[i->]	0.19/69
[<-b]	0.60/17	[o-i]	0.52/147	[e-o]	0.43/197	[d->]	0.08/17
[~-a]	0.60/65	[i-b]	0.51/149	[p-u]	0.43/354	[R->]	0.05/82
[y-k]	0.60/24	[R-k]	0.51/138	[d-e]	0.43/1177	[<-d]	0.04/101
[a-i]	0.60/365	[a-y]	0.51/129	[<-i]	0.43/237		
[s-z]	0.59/46	[r-e]	0.51/189	[<-z]	0.42/31		
[e-k]	0.58/491	[i-o]	0.51/426	[o-m]	0.42/490		
[a-f]	0.58/79	[i-i]	0.51/65	[j-e]	0.42/226		
[e-b]	0.58/274	[e-d]	0.50/412	[s-p]	0.42/358		

j. Usuario IC

[a-f]	1.13/31	[R-k]	1.02/85	[f-a]	1.01/19	[a-t]	0.98/127
[e-p]	1.08/107	[r-k]	1.02/85	[l-m]	1.01/21	[e-t]	0.98/96
[u-i]	1.07/21	[f-e]	1.02/6	[p-l]	1.00/8	[c-o]	0.98/88
[l-p]	1.07/107	[s-l]	1.02/11	[u-t]	0.99/111	[R-z]	0.97/19
[l-u]	1.06/49	[a-p]	1.02/168	[k-t]	0.99/30	[e-n]	0.97/596

[i-z]	0.96/96	[t-a]	0.81/357	[e-r]	0.73/135	[b-o]	0.64/43
[e-y]	0.94/23	[u-e]	0.80/71	[k-e]	0.73/482	[d-o]	0.64/218
[l->]	0.94/91	[g-a]	0.80/65	[<-l]	0.72/16	[i-a]	0.63/166
[a-g]	0.93/26	[o-e]	0.80/100	[a-b]	0.72/127	[i-o]	0.63/143
[n-t]	0.92/369	[t-o]	0.80/296	[u-c]	0.72/19	[s->]	0.63/308
[y-a]	0.91/43	[u-a]	0.79/18	[s-k]	0.72/120	[e-e]	0.63/158
[i-u]	0.91/19	[n-u]	0.79/19	[i-s]	0.71/227	[o-b]	0.62/64
[z-e]	0.91/174	[<-i]	0.79/176	[e-s]	0.71/818	[o-i]	0.62/49
[u-n]	0.90/215	[m-o]	0.79/189	[o-m]	0.71/163	[r-m]	0.62/31
[e-f]	0.89/54	[L-a]	0.78/274	[o-a]	0.71/87	[R-m]	0.62/31
[a-n]	0.89/266	[l-a]	0.78/274	[i-d]	0.71/117	[i-g]	0.62/32
[n-z]	0.89/87	[n->]	0.78/116	[j-e]	0.71/122	[l-e]	0.61/46
[t-R]	0.88/77	[n-f]	0.78/33	[<-d]	0.71/91	[<-o]	0.61/42
[e-z]	0.88/124	[t-i]	0.78/214	[e-o]	0.70/42	[o-d]	0.61/85
[n-k]	0.88/42	[p-i]	0.78/63	[c-a]	0.70/6	[<-a]	0.61/157
[a-z]	0.88/140	[i-e]	0.78/217	[m-u]	0.70/34	[a-l]	0.61/338
[u-s]	0.88/53	[s-e]	0.78/470	[k-i]	0.70/91	[u-l]	0.61/31
[o->]	0.88/373	[d-a]	0.78/161	[a-r]	0.70/140	[R-u]	0.60/34
[t-u]	0.86/42	[o-k]	0.78/227	[a-R]	0.70/140	[j-i]	0.59/17
[p-e]	0.86/171	[i-t]	0.77/94	[R-o]	0.69/67	[p-o]	0.59/239
[n-l]	0.86/35	[s-p]	0.77/103	[r-o]	0.69/67	[<-s]	0.59/15
[r-a]	0.85/144	[e-i]	0.77/69	[i-i]	0.69/54	[o-o]	0.58/43
[R-a]	0.85/144	[r-d]	0.77/13	[k-u]	0.69/113	[i-p]	0.58/40
[t-e]	0.85/364	[R-d]	0.77/13	[d-i]	0.69/113	[j-o]	0.58/62
[n-j]	0.84/26	[o-j]	0.76/74	[r-e]	0.69/153	[n-o]	0.58/259
[R-l]	0.84/25	[a-e]	0.76/94	[R-e]	0.69/153	[<-p]	0.57/19
[z-i]	0.84/282	[n-e]	0.76/180	[s-o]	0.68/246	[b-l]	0.57/36
[R->]	0.84/101	[n-d]	0.76/137	[n-g]	0.68/30	[i-b]	0.56/31
[j-a]	0.83/82	[f-i]	0.76/24	[e-l]	0.67/351	[<-b]	0.56/28
[s-d]	0.83/34	[b-e]	0.76/34	[e-L]	0.67/351	[<-t]	0.56/39
[a->]	0.83/387	[e-j]	0.76/85	[a-o]	0.67/77	[n-i]	0.54/83
[a-m]	0.83/207	[m-p]	0.76/53	[o-s]	0.67/606	[o-l]	0.54/151
[a-k]	0.82/205	[k-a]	0.75/154	[k-o]	0.67/318	[o-r]	0.53/187
[o-z]	0.82/47	[e-a]	0.75/163	[<-m]	0.66/62	[o-R]	0.53/187
[n-m]	0.82/40	[o-p]	0.75/112	[<-n]	0.66/95	[l-i]	0.51/85
[m-e]	0.82/163	[o-n]	0.75/344	[i-r]	0.66/28	[u-d]	0.51/36
[p-u]	0.82/63	[m-i]	0.75/88	[i-R]	0.66/28	[s-i]	0.50/99
[o-t]	0.82/55	[p-a]	0.74/207	[e-d]	0.66/83	[n-s]	0.47/68
[i-m]	0.82/90	[e-k]	0.74/168	[g-u]	0.65/53	[r-i]	0.47/13
[e-m]	0.82/221	[a-s]	0.74/402	[a-u]	0.65/40	[R-i]	0.47/13
[i-n]	0.82/132	[n-a]	0.74/183	[a-d]	0.65/244	[y-o]	0.46/47
[d-e]	0.82/328	[s-a]	0.74/291	[a-j]	0.65/82	[<-k]	0.45/10
[i-k]	0.82/160	[b-a]	0.74/76	[m-a]	0.65/261	[o-u]	0.44/28
[s-u]	0.82/138	[a-i]	0.73/36	[b-i]	0.65/70	[e-b]	0.44/53
[l-o]	0.81/173	[<-e]	0.73/240	[s-m]	0.65/41	[i-l]	0.43/25
[R-s]	0.81/21	[p-R]	0.73/104	[s-t]	0.64/80	[l-k]	0.42/15
[R-t]	0.81/78	[z-a]	0.73/38	[n-p]	0.64/12	[g-o]	0.40/82
[r-t]	0.81/78	[e-R]	0.73/135	[e->]	0.64/287	[e-g]	0.40/32

[i->]	0.30/15	[l-d]	0.22/38	[<-y]	0.05/40
[l-t]	0.26/14	[e-c]	0.12/12		

k. Usuario JG

[n-m]	1.56/28	[o-o]	0.86/15	[m-a]	0.69/149	[a->]	0.56/227
[d-k]	1.40/11	[<-p]	0.85/71	[l-o]	0.68/168	[t-e]	0.56/279
[r-m]	1.36/18	[a-p]	0.85/76	[i->]	0.68/43	[r-t]	0.54/15
[R-k]	1.35/24	[o-m]	0.85/118	[r-i]	0.68/43	[e-L]	0.54/11
[n-n]	1.31/15	[r-a]	0.85/102	[k-e]	0.68/482	[k-u]	0.54/44
[R-m]	1.24/47	[s->]	0.84/212	[n-u]	0.67/49	[e-j]	0.54/35
[s-m]	1.20/42	[a-s]	0.83/451	[o-k]	0.67/309	[u-s]	0.54/16
[<-t]	1.19/11	[n-e]	0.83/156	[R->]	0.67/50	[o-u]	0.54/13
[<-r]	1.14/16	[R-f]	0.83/13	[l-e]	0.66/104	[i-t]	0.54/68
[R-p]	1.13/31	[o-s]	0.83/611	[e->]	0.66/187	[t-o]	0.53/304
[l-u]	1.13/15	[s-k]	0.82/102	[n-g]	0.66/18	[d-i]	0.53/61
[~-o]	1.10/12	[e-m]	0.82/124	[<-b]	0.66/21	[n-f]	0.53/11
[n-R]	1.06/13	[a-e]	0.82/100	[n-d]	0.65/155	[<-o]	0.52/28
[<-k]	1.06/35	[n-l]	0.81/43	[a-z]	0.64/199	[s-a]	0.52/287
[x-p]	1.03/18	[u-i]	0.81/53	[r-e]	0.64/69	[k-t]	0.52/39
[e-n]	1.03/445	[n->]	0.80/66	[a-b]	0.64/90	[y-o]	0.52/130
[a-n]	1.00/256	[R-u]	0.80/13	[b-i]	0.64/99	[s-o]	0.51/286
[<-n]	1.00/44	[<-e]	0.79/257	[j-e]	0.63/54	[c-o]	0.51/51
[e-a]	1.00/100	[m-p]	0.79/90	[m-e]	0.63/213	[e-f]	0.51/83
[y-a]	0.99/19	[i-n]	0.78/133	[u-m]	0.63/19	[s-e]	0.50/272
[<-l]	0.96/59	[r-d]	0.78/16	[i-o]	0.62/133	[n-z]	0.50/49
[n-s]	0.95/52	[R-n]	0.78/22	[i-s]	0.62/116	[s-s]	0.50/15
[o-n]	0.94/220	[n-k]	0.78/89	[e-z]	0.62/241	[l->]	0.50/79
[<-m]	0.94/113	[e-l]	0.78/235	[R-z]	0.62/25	[d-a]	0.50/118
[p-p]	0.92/16	[R-i]	0.76/54	[e-s]	0.62/733	[a-i]	0.50/80
[i-p]	0.90/33	[p-a]	0.76/211	[s-t]	0.61/249	[a-r]	0.49/112
[s-b]	0.90/19	[l-p]	0.75/35	[l-i]	0.61/111	[a-f]	0.49/26
[n-i]	0.90/53	[m-i]	0.75/59	[L-a]	0.60/19	[l-a]	0.49/244
[i-m]	0.89/195	[a-t]	0.75/146	[i-e]	0.60/158	[R-e]	0.49/207
[r-n]	0.89/13	[i-i]	0.74/39	[t-a]	0.59/299	[<-s]	0.49/56
[e-e]	0.88/68	[p-i]	0.74/22	[d-e]	0.59/290	[m-u]	0.49/149
[b-l]	0.88/32	[s-p]	0.73/106	[z-i]	0.58/303	[a-k]	0.49/131
[a-j]	0.88/17	[o->]	0.73/265	[i-k]	0.58/124	[o-l]	0.48/99
[n-p]	0.88/53	[e-p]	0.73/192	[e-y]	0.57/38	[n-a]	0.48/149
[u-n]	0.88/221	[k-i]	0.73/103	[g-o]	0.57/49	[o-b]	0.48/110
[l-t]	0.88/35	[a-~]	0.73/17	[s-u]	0.57/119	[a-m]	0.48/149
[i-u]	0.88/36	[<-i]	0.72/90	[e-t]	0.57/108	[p-e]	0.48/169
[i-a]	0.87/173	[n-t]	0.71/330	[e-x]	0.57/14	[f-o]	0.46/32
[<-d]	0.87/46	[k-a]	0.71/239	[e-i]	0.57/69	[o-r]	0.46/52
[R-s]	0.87/30	[o-j]	0.70/49	[f-i]	0.56/39	[u-a]	0.46/25
[R-t]	0.86/70	[e-k]	0.70/280	[u-t]	0.56/16	[o-e]	0.46/106
[o-p]	0.86/98	[g-e]	0.69/17	[k-o]	0.56/254	[p-u]	0.45/93

[a-g]	0.45/34	[R-o]	0.40/52	[g-u]	0.37/40	[e-d]	0.26/118
[z-a]	0.44/73	[n-o]	0.40/147	[o-a]	0.36/71	[o-y]	0.21/32
[a-l]	0.43/350	[m-o]	0.39/135	[e-o]	0.35/28	[m-r]	0.18/18
[s-d]	0.43/44	[o-i]	0.39/48	[<-y]	0.34/24		
[z-e]	0.43/191	[s-l]	0.38/43	[l-m]	0.34/20		
[l-g]	0.42/56	[e-c]	0.38/21	[l-d]	0.32/30		
[R-a]	0.42/88	[t-y]	0.38/12	[g-a]	0.29/28		
[y-u]	0.41/11	[a-y]	0.37/80	[a-d]	0.28/111		

l. Usuario JM

[u->]	1.33/13	[~-a]	0.73/45	[o-n]	0.66/590	[k-o]	0.58/464
[n-m]	1.26/16	[n-y]	0.73/52	[s-o]	0.66/468	[k-a]	0.58/390
[l-r]	1.18/16	[z-e]	0.73/212	[e-y]	0.65/43	[a-a]	0.58/146
[m-m]	1.08/14	[r-i]	0.73/101	[t-a]	0.65/596	[R->]	0.58/76
[R-n]	1.05/54	[u-n]	0.72/294	[d-u]	0.65/32	[t-R]	0.57/136
[u-m]	1.02/28	[d-e]	0.72/530	[e->]	0.65/433	[<-b]	0.57/34
[<-r]	0.96/23	[g-a]	0.72/92	[t-i]	0.65/405	[z-k]	0.57/23
[g-R]	0.94/30	[a-i]	0.71/166	[<-m]	0.65/25	[o-i]	0.57/98
[k-z]	0.93/27	[<-p]	0.71/57	[z-i]	0.65/574	[r-n]	0.57/13
[s-m]	0.92/112	[u-d]	0.71/67	[s->]	0.65/389	[n-t]	0.57/572
[e-f]	0.92/70	[r-e]	0.71/60	[a-s]	0.65/777	[i-n]	0.57/208
[o-y]	0.91/24	[e-a]	0.70/214	[e-x]	0.64/36	[<-a]	0.56/130
[j-a]	0.88/63	[a-e]	0.70/162	[k-R]	0.64/119	[n-l]	0.56/88
[i-b]	0.87/58	[p-e]	0.70/265	[l-e]	0.64/285	[a-l]	0.56/449
[e-m]	0.86/265	[p-i]	0.70/21	[R-a]	0.64/168	[e-p]	0.56/231
[L-a]	0.86/19	[b-e]	0.69/151	[i-o]	0.63/234	[b-o]	0.56/67
[e-e]	0.85/255	[s-d]	0.69/96	[i-a]	0.63/286	[m-u]	0.56/72
[x-p]	0.84/34	[o-g]	0.69/21	[e-n]	0.62/1063	[n-o]	0.55/495
[b-i]	0.84/118	[e-z]	0.69/188	[j-e]	0.62/143	[r-t]	0.55/23
[e-b]	0.83/96	[o-k]	0.69/634	[s-e]	0.62/584	[u-t]	0.55/88
[e-d]	0.83/232	[i-e]	0.69/321	[d-a]	0.62/184	[a-b]	0.55/196
[z->]	0.81/25	[<-n]	0.69/147	[o-e]	0.61/264	[m-o]	0.54/237
[k-e]	0.80/885	[x-u]	0.68/18	[e-i]	0.61/87	[p-a]	0.54/325
[i-d]	0.80/223	[l->]	0.68/58	[e-t]	0.61/244	[e-s]	0.54/1366
[f-e]	0.79/53	[s-a]	0.68/338	[o-m]	0.61/220	[a-u]	0.54/71
[f-a]	0.79/17	[i-g]	0.68/67	[i-m]	0.61/157	[R-d]	0.54/38
[<-t]	0.78/21	[m-e]	0.68/214	[n-p]	0.61/94	[s-i]	0.54/445
[l-z]	0.78/26	[u-z]	0.68/56	[a-t]	0.61/181	[a-R]	0.54/277
[o-z]	0.78/82	[R-i]	0.68/114	[i-z]	0.61/256	[o-p]	0.54/169
[i-p]	0.78/124	[m-a]	0.68/345	[e-k]	0.60/343	[l-d]	0.54/56
[c-e]	0.77/13	[i->]	0.67/44	[l-p]	0.60/146	[f-i]	0.54/49
[<-e]	0.76/380	[<-u]	0.67/18	[u-c]	0.59/58	[n-e]	0.54/409
[o->]	0.75/603	[t-u]	0.67/116	[i-R]	0.59/82	[n-s]	0.53/125
[d-i]	0.75/240	[s-y]	0.66/23	[s-l]	0.59/93	[R-t]	0.53/156
[n-a]	0.75/352	[a-m]	0.66/363	[n-d]	0.59/180	[a-d]	0.52/328
[l-i]	0.74/157	[b-l]	0.66/45	[a-n]	0.58/448	[l-a]	0.52/519
[i-t]	0.74/118	[e-g]	0.66/130	[z-a]	0.58/44	[b-u]	0.52/51
[a-p]	0.73/122	[t-e]	0.66/412	[R-e]	0.58/442	[r->]	0.51/23

[m-i]	0.51/205	[u-i]	0.47/26	[k-i]	0.44/110	[t-o]	0.41/602
[s-g]	0.50/12	[o-a]	0.47/162	[m-p]	0.44/183	[d-d]	0.41/13
[R-m]	0.50/30	[a-k]	0.47/357	[o-R]	0.44/276	[i-s]	0.40/366
[e-c]	0.50/75	[r-a]	0.46/145	[e-o]	0.44/71	[<-i]	0.38/188
[o-s]	0.49/718	[R-p]	0.46/21	[g-e]	0.44/15	[u-s]	0.36/69
[e-R]	0.49/284	[L-o]	0.46/32	[k-u]	0.44/184	[a-g]	0.35/25
[R-u]	0.49/20	[y-a]	0.46/84	[l-b]	0.43/15	[i-r]	0.32/53
[e-L]	0.48/21	[o-u]	0.46/26	[<-d]	0.43/52	[r-o]	0.32/102
[u-e]	0.48/107	[g-u]	0.46/49	[u-a]	0.43/61	[a-o]	0.24/101
[e-r]	0.48/155	[k-t]	0.45/48	[p-l]	0.42/79	[b-a]	0.23/129
[j-i]	0.47/44	[i-L]	0.45/15	[u-p]	0.42/14		

m. Usuario JS

[r-n]	1.49/11	[n-m]	0.95/11	[l-o]	0.79/139	[a-s]	0.69/451
[r-m]	1.48/11	[a-f]	0.94/19	[<-l]	0.79/57	[e->]	0.69/366
[<-n]	1.47/135	[p-u]	0.92/86	[c-o]	0.78/50	[l-p]	0.69/124
[<-m]	1.41/117	[o-e]	0.92/73	[m-o]	0.78/204	[e-b]	0.67/66
[R-t]	1.36/42	[u-n]	0.92/282	[l-t]	0.77/36	[r-o]	0.67/112
[e-u]	1.36/27	[e-m]	0.92/234	[<-i]	0.77/207	[i-p]	0.67/75
[l-m]	1.35/46	[n-t]	0.91/564	[o-k]	0.77/161	[<-e]	0.67/388
[u-m]	1.29/45	[s-k]	0.91/166	[i-c]	0.76/34	[e-o]	0.66/49
[i-m]	1.28/110	[u->]	0.90/20	[<-s]	0.76/141	[e-x]	0.66/35
[n-p]	1.27/19	[s-u]	0.89/97	[i-u]	0.76/21	[a-m]	0.65/230
[u-k]	1.21/11	[o-t]	0.89/98	[m-e]	0.75/297	[a-n]	0.65/264
[g-u]	1.20/65	[a-p]	0.89/137	[i-k]	0.75/86	[t-a]	0.65/352
[<-p]	1.19/51	[o-s]	0.89/766	[j-i]	0.75/21	[o-i]	0.64/64
[o-o]	1.16/24	[o->]	0.89/461	[u-t]	0.74/31	[p-i]	0.64/35
[o-p]	1.16/92	[b-d]	0.88/16	[y-o]	0.74/62	[t-e]	0.64/577
[f-u]	1.15/27	[u-d]	0.88/22	[p-R]	0.74/213	[~-o]	0.63/42
[m-u]	1.15/90	[R-o]	0.88/94	[a-u]	0.73/85	[k-e]	0.63/394
[<-f]	1.15/21	[a-t]	0.88/172	[s-b]	0.73/22	[d-e]	0.63/396
[r-k]	1.15/22	[p-l]	0.88/19	[d-i]	0.73/148	[i-t]	0.63/140
[n-u]	1.14/118	[m-p]	0.87/81	[o-n]	0.73/368	[e-g]	0.62/30
[R-m]	1.14/52	[k-l]	0.86/20	[n-o]	0.72/262	[l-i]	0.62/75
[<-b]	1.12/45	[R-k]	0.86/61	[n-s]	0.72/80	[b-i]	0.62/144
[R->]	1.10/111	[i-s]	0.86/145	[u-a]	0.72/48	[l-g]	0.62/57
[t-p]	1.09/16	[o-c]	0.85/21	[r-R]	0.72/20	[i-d]	0.61/210
[n-b]	1.07/22	[t-o]	0.84/368	[s-i]	0.72/453	[a-L]	0.61/11
[e-n]	1.07/622	[l-s]	0.84/40	[R-s]	0.72/32	[n-i]	0.61/129
[m-b]	1.05/26	[s-m]	0.84/92	[o-u]	0.72/16	[b-o]	0.60/69
[o-m]	1.04/171	[u-e]	0.84/120	[R-z]	0.71/16	[R-i]	0.60/83
[s->]	1.03/351	[<-u]	0.83/80	[e-s]	0.71/788	[s-o]	0.60/186
[e-f]	1.01/47	[d->]	0.83/47	[u-s]	0.71/60	[~-a]	0.59/37
[i-n]	1.00/197	[a->]	0.83/272	[<-d]	0.70/158	[l-u]	0.59/30
[l-k]	0.99/28	[a-~]	0.81/80	[i->]	0.70/104	[a-o]	0.59/87
[n-f]	0.97/26	[p-o]	0.81/218	[n->]	0.70/73	[o-a]	0.58/47
[l->]	0.96/97	[s-d]	0.81/70	[<-k]	0.69/104	[s-e]	0.58/353
[u-i]	0.95/51	[m-i]	0.80/125	[s-n]	0.69/53	[n-g]	0.57/16

[n-d]	0.57/152	[e-a]	0.54/137	[R-a]	0.47/154	[R-n]	0.37/52
[p-e]	0.57/96	[r-e]	0.54/68	[b-e]	0.46/124	[g-a]	0.35/24
[i-g]	0.57/30	[i-e]	0.53/190	[i-b]	0.46/80	[b-a]	0.34/96
[s-p]	0.56/182	[b-R]	0.53/33	[e-l]	0.45/261	[n-e]	0.32/127
[a-i]	0.56/84	[l-a]	0.53/256	[a-b]	0.44/199	[n-a]	0.32/222
[e-y]	0.56/14	[a-z]	0.53/123	[i-r]	0.44/22	[e-r]	0.32/143
[d-o]	0.56/231	[r-i]	0.52/62	[b-s]	0.43/37	[e-i]	0.31/42
[z-i]	0.56/347	[g-o]	0.51/58	[e-d]	0.42/118	[s-g]	0.31/27
[n-z]	0.55/54	[m-a]	0.51/232	[d-a]	0.41/200	[a-a]	0.30/19
[l-e]	0.55/102	[z-a]	0.51/41	[i-l]	0.40/86	[i-a]	0.28/138
[o-b]	0.55/73	[t-i]	0.50/204	[r-a]	0.40/70		
[t-R]	0.55/155	[f-i]	0.50/30	[s-a]	0.38/260		
[e-t]	0.55/144	[e-p]	0.49/129	[e-e]	0.38/48		
[p-a]	0.55/186	[a-d]	0.48/222	[k-a]	0.38/250		
[R-e]	0.54/313	[i-R]	0.48/57	[s-l]	0.37/37		

n. Usuario JZ

[<-p]	1.15/59	[e->]	0.83/355	[m-b]	0.68/22	[R-l]	0.54/33
[R->]	1.11/62	[<-a]	0.81/131	[e-k]	0.67/197	[i-z]	0.54/33
[u-b]	1.09/27	[e-x]	0.80/22	[z-d]	0.67/13	[a-R]	0.54/131
[d->]	1.08/14	[R-d]	0.80/18	[s-t]	0.67/216	[m-p]	0.54/64
[l-t]	1.07/50	[b-l]	0.80/23	[i-t]	0.67/105	[n-a]	0.53/146
[s->]	1.07/223	[s-m]	0.79/44	[g-a]	0.66/46	[m-i]	0.53/120
[e-i]	1.04/40	[i-a]	0.78/94	[i-e]	0.66/187	[o-k]	0.53/178
[j-u]	1.03/11	[a-~]	0.78/31	[k-a]	0.64/125	[<-i]	0.53/223
[s-p]	1.02/31	[a-k]	0.77/72	[p-a]	0.64/127	[o-u]	0.53/29
[x-p]	1.02/12	[i-R]	0.77/25	[n-i]	0.63/114	[k-t]	0.52/33
[<-d]	1.00/62	[R-y]	0.76/14	[e-m]	0.62/131	[z-i]	0.51/198
[l-u]	0.96/27	[a->]	0.76/419	[o->]	0.62/307	[i-o]	0.51/81
[<-m]	0.95/111	[n-t]	0.75/324	[e-s]	0.61/568	[o-l]	0.51/109
[~-o]	0.94/21	[u-k]	0.74/20	[e-n]	0.61/395	[a-d]	0.50/153
[r-i]	0.94/36	[a-e]	0.74/48	[u-s]	0.61/17	[n-g]	0.50/39
[l-p]	0.94/84	[z-e]	0.72/110	[e-d]	0.60/142	[n-k]	0.50/27
[s-k]	0.92/135	[g-u]	0.72/23	[e-p]	0.59/123	[R-a]	0.50/38
[R-t]	0.92/35	[i-n]	0.72/155	[e-a]	0.59/197	[i->]	0.50/115
[R-k]	0.90/55	[a-i]	0.71/80	[t-i]	0.59/141	[j-i]	0.49/34
[o-p]	0.89/138	[a-p]	0.71/54	[n-s]	0.58/39	[o-n]	0.49/315
[l-i]	0.89/47	[o-t]	0.71/97	[a-o]	0.58/71	[i-s]	0.49/160
[<-k]	0.88/54	[s-n]	0.71/25	[R-s]	0.58/51	[d-o]	0.49/123
[n->]	0.88/70	[b-i]	0.71/113	[j-a]	0.58/19	[a-f]	0.49/39
[<-e]	0.87/385	[a-t]	0.71/47	[l->]	0.58/47	[a-j]	0.49/73
[~-a]	0.86/15	[p-u]	0.70/51	[u-n]	0.57/232	[l-s]	0.48/28
[p-l]	0.85/20	[o-i]	0.70/52	[t-a]	0.57/299	[l-a]	0.48/193
[d-a]	0.85/124	[l-e]	0.70/104	[d-e]	0.56/383	[o-m]	0.48/81
[a-a]	0.84/69	[i-p]	0.69/37	[m-a]	0.56/116	[m-o]	0.48/124
[<-s]	0.84/24	[n-e]	0.69/167	[a-g]	0.55/14	[a-s]	0.47/301
[e-e]	0.84/372	[k-e]	0.69/383	[n-d]	0.55/91	[n-n]	0.47/19
[t-e]	0.84/287	[<-u]	0.69/48	[a-r]	0.55/63	[s-d]	0.47/54

[a-z]	0.46/105	[i-b]	0.39/29	[u-z]	0.33/24	[t-u]	0.24/16
[e-g]	0.45/12	[e-f]	0.39/30	[p-e]	0.33/123	[o-r]	0.23/39
[i-m]	0.45/69	[m-e]	0.38/116	[d-i]	0.32/100	[n-b]	0.23/31
[a-n]	0.44/233	[f-u]	0.38/40	[e-t]	0.32/48	[o-b]	0.23/78
[p-o]	0.43/162	[s-a]	0.38/139	[f-e]	0.31/27	[i-k]	0.22/183
[l-o]	0.43/153	[r-a]	0.38/108	[n-p]	0.30/30	[b-e]	0.21/30
[e-b]	0.43/64	[i-r]	0.38/37	[e-u]	0.30/59	[e-l]	0.20/335
[m-u]	0.42/28	[i-d]	0.38/228	[R-i]	0.29/48	[b-a]	0.19/76
[j-e]	0.42/90	[t-o]	0.37/163	[t-R]	0.28/63	[a-b]	0.18/110
[R-e]	0.42/172	[R-u]	0.36/22	[b-o]	0.28/48	[k-R]	0.17/34
[a-l]	0.42/178	[s-y]	0.36/28	[n-o]	0.27/248	[l-R]	0.09/21
[p-i]	0.42/38	[o-a]	0.36/97	[k-u]	0.27/50	[e-o]	0.08/60
[s-i]	0.42/278	[r-e]	0.35/40	[n-l]	0.27/43	[<-y]	0.06/32
[s-e]	0.42/248	[e-R]	0.35/166	[s-l]	0.27/22	[e-c]	-0.05/13
[e-r]	0.42/82	[o-s]	0.34/469	[R-o]	0.26/37		
[e-z]	0.41/135	[a-m]	0.34/203	[o-o]	0.25/43		
[n-z]	0.41/67	[<-n]	0.34/83	[s-b]	0.25/22		
[o-R]	0.41/105	[k-o]	0.34/309	[R-n]	0.25/66		
[y-a]	0.39/24	[s-u]	0.34/147	[l-g]	0.25/34		

o. Usuario LA

		[r-i]	0.89/109	[e-i]	0.81/88	[o-p]	0.74/194
[y-l]	1.46/13	[u-R]	0.89/17	[p-e]	0.80/208	[i-n]	0.74/238
[l-u]	1.27/31	[r-t]	0.89/13	[e-n]	0.80/771	[a-l]	0.74/285
[o-g]	1.24/21	[f-R]	0.89/13	[i-o]	0.79/211	[a-m]	0.74/312
[n-n]	1.22/31	[m-b]	0.89/18	[n->]	0.79/187	[u-n]	0.74/273
[u-t]	1.16/31	[b-l]	0.88/32	[u-i]	0.79/25	[o-n]	0.74/615
[o-f]	1.16/12	[z-o]	0.87/23	[b-e]	0.79/142	[o-d]	0.73/111
[n-u]	1.11/33	[g-o]	0.87/58	[u-p]	0.78/30	[c-a]	0.73/27
[n-y]	1.11/71	[o->]	0.87/475	[i-p]	0.78/52	[a-e]	0.72/121
[l-n]	1.09/12	[o-k]	0.86/508	[o-b]	0.78/116	[r-a]	0.72/97
[a-p]	1.07/228	[e-p]	0.86/236	[n-o]	0.77/539	[l-i]	0.72/107
[o-j]	1.07/25	[y-a]	0.86/56	[n-k]	0.77/103	[s-n]	0.72/57
[y-e]	1.05/57	[l-t]	0.84/57	[a-k]	0.77/453	[o-y]	0.72/41
[<-b]	1.02/16	[a->]	0.84/304	[m-a]	0.77/183	[i-g]	0.71/29
[l-p]	1.02/79	[R-n]	0.84/24	[f-a]	0.77/17	[o-a]	0.71/119
[e-f]	1.02/113	[p-a]	0.84/309	[m-o]	0.77/201	[<-d]	0.71/49
[<-f]	0.98/35	[n-t]	0.84/476	[x-p]	0.77/13	[e-b]	0.71/117
[m-u]	0.98/63	[e-t]	0.83/204	[z-a]	0.76/87	[o-l]	0.71/78
[l-g]	0.98/40	[b-u]	0.83/33	[g-i]	0.76/28	[e-o]	0.70/109
[R-p]	0.94/23	[a-j]	0.83/41	[p-R]	0.75/174	[e-u]	0.70/44
[l-k]	0.94/41	[k-u]	0.83/88	[n-p]	0.75/76	[z-e]	0.70/204
[i-j]	0.92/25	[k-a]	0.82/225	[R->]	0.75/89	[i-a]	0.70/308
[l->]	0.91/43	[g-a]	0.82/62	[a-n]	0.75/394	[d-a]	0.70/203
[a-t]	0.91/203	[R-f]	0.82/20	[R-z]	0.75/17	[b-a]	0.70/210
[g-R]	0.90/25	[L-a]	0.82/13	[o-m]	0.75/218	[m-i]	0.70/146
[R-u]	0.89/12	[o-t]	0.82/98	[t-e]	0.74/473	[a-o]	0.69/125
[a-y]	0.89/48	[a-z]	0.81/157	[b-o]	0.74/35	[d-o]	0.69/208

[<-y]	0.68/32	[a-g]	0.64/37	[g-u]	0.60/49	[i->]	0.52/130
[a-a]	0.68/66	[i-b]	0.64/118	[k-o]	0.60/383	[i-m]	0.51/86
[n-e]	0.68/312	[n-a]	0.64/303	[l-b]	0.59/12	[s-b]	0.51/46
[e-m]	0.68/301	[r-e]	0.64/108	[u-a]	0.59/43	[e-y]	0.51/27
[m-p]	0.68/139	[e-s]	0.64/1084	[t-o]	0.59/428	[j-e]	0.50/66
[t-a]	0.67/679	[R-o]	0.63/134	[R-s]	0.59/21	[o-i]	0.50/107
[a-u]	0.67/52	[e-r]	0.63/172	[e-R]	0.58/244	[s-i]	0.49/402
[z-i]	0.67/425	[i-d]	0.63/165	[l-m]	0.58/22	[a-r]	0.49/142
[b-R]	0.67/18	[m-e]	0.63/216	[e->]	0.57/281	[a-i]	0.48/135
[d-e]	0.66/395	[s-k]	0.63/232	[i-t]	0.56/114	[n-i]	0.47/90
[R-e]	0.66/298	[i-r]	0.62/21	[e-k]	0.56/288	[<-l]	0.46/17
[l-e]	0.66/237	[s-a]	0.62/344	[~-a]	0.55/17	[k-i]	0.45/90
[R-l]	0.66/58	[d-i]	0.62/151	[<-u]	0.55/39	[y-o]	0.45/133
[p-p]	0.65/17	[t-i]	0.62/200	[R-i]	0.55/76	[i-l]	0.45/95
[p-o]	0.65/268	[<-o]	0.62/65	[i-c]	0.54/34	[i-y]	0.43/36
[n-d]	0.65/98	[o-r]	0.61/52	[~-o]	0.54/43	[e-d]	0.42/196
[b-i]	0.65/185	[p-l]	0.61/72	[R-a]	0.54/177	[j-i]	0.40/18
[k-e]	0.65/686	[R-m]	0.61/22	[a-~]	0.53/72	[i-R]	0.38/94
[i-s]	0.65/216	[l-a]	0.61/337	[<-m]	0.53/37	[l-d]	0.32/61
[u-r]	0.65/27	[<-i]	0.61/125	[e-a]	0.53/174		
[a-d]	0.65/263	[k-R]	0.60/115	[e-g]	0.53/59		
[R-t]	0.64/148	[e-e]	0.60/300	[o-R]	0.52/179		
[i-e]	0.64/284	[a-s]	0.60/615	[a-b]	0.52/200		

p. Usuario MS

[n-p]	1.74/13	[a-z]	1.02/76	[z-e]	0.94/70	[g-u]	0.87/19
[l-p]	1.58/24	[o-R]	1.02/53	[R-k]	0.93/61	[k-e]	0.87/204
[u-t]	1.53/15	[a-t]	1.01/41	[R-d]	0.93/17	[e-m]	0.87/107
[u-p]	1.44/12	[e-u]	1.01/28	[o-r]	0.93/34	[g-o]	0.86/15
[<-p]	1.35/28	[n-k]	1.01/35	[y-o]	0.93/27	[f-u]	0.86/18
[<-m]	1.33/17	[r-i]	1.00/41	[u-n]	0.93/47	[i-e]	0.86/100
[o-p]	1.29/47	[n-t]	1.00/154	[i-n]	0.92/89	[a-u]	0.86/17
[u-d]	1.25/17	[a-n]	0.99/190	[l-m]	0.92/13	[t-o]	0.86/93
[o-t]	1.21/58	[s-d]	0.99/34	[a-j]	0.92/17	[s-t]	0.85/112
[m-p]	1.18/30	[i-t]	0.99/51	[o-m]	0.91/51	[o-o]	0.85/15
[<-n]	1.17/11	[b-a]	0.99/35	[e-t]	0.91/46	[j-e]	0.84/49
[a-p]	1.17/78	[R-a]	0.98/25	[a-q]	0.90/34	[R-i]	0.84/24
[a-d]	1.13/253	[f-o]	0.98/16	[r-a]	0.90/101	[e-s]	0.84/347
[s-m]	1.10/54	[a-l]	0.98/103	[m-i]	0.89/44	[R-o]	0.84/24
[a-i]	1.07/46	[o->]	0.98/130	[l-s]	0.89/16	[u-a]	0.83/41
[e-b]	1.06/17	[a-k]	0.97/91	[t-e]	0.89/134	[o-a]	0.83/35
[d-i]	1.05/48	[e-k]	0.97/58	[p-o]	0.89/99	[a-a]	0.83/61
[k-u]	1.04/22	[i-k]	0.96/47	[t-u]	0.89/48	[<-a]	0.83/56
[a-o]	1.04/28	[o-d]	0.96/71	[s->]	0.88/151	[i-j]	0.83/28
[o-z]	1.02/24	[m-a]	0.95/90	[i-s]	0.88/88	[p-R]	0.83/47
[n-d]	1.02/43	[j-a]	0.95/29	[a->]	0.88/146	[a-r]	0.82/95

[n-a]	0.82/75	[k-a]	0.72/140	[s-k]	0.62/73	[n-s]	0.50/71
[d-a]	0.82/92	[r-o]	0.72/44	[a-m]	0.61/74	[p-u]	0.49/56
[n-o]	0.82/97	[R-n]	0.72/25	[p-a]	0.61/77	[l-f]	0.49/14
[d-o]	0.82/171	[e-o]	0.71/28	[e-j]	0.61/25	[n-j]	0.48/11
[R-s]	0.81/17	[<-i]	0.71/13	[a-s]	0.61/197	[x-i]	0.47/27
[o-n]	0.81/142	[t-R]	0.70/56	[l-t]	0.60/14	[c-o]	0.47/11
[i-p]	0.80/30	[o-k]	0.70/65	[l-i]	0.60/52	[o-e]	0.46/37
[a-R]	0.80/110	[o-b]	0.69/39	[<-k]	0.60/49	[u-l]	0.46/27
[i-z]	0.80/31	[s-i]	0.68/224	[e-a]	0.59/96	[e-l]	0.45/178
[t-i]	0.80/59	[z-a]	0.68/63	[p-e]	0.59/90	[i->]	0.45/17
[a-b]	0.79/75	[s-e]	0.68/147	[R->]	0.58/38	[<-s]	0.44/35
[e-p]	0.79/74	[i-m]	0.68/27	[l-o]	0.57/76	[e-x]	0.42/14
[a-e]	0.79/55	[b-i]	0.67/67	[R-e]	0.57/128	[e-g]	0.42/42
[s-p]	0.79/45	[g-a]	0.66/33	[r-e]	0.57/81	[<-d]	0.40/44
[u-e]	0.78/19	[k-t]	0.66/27	[n-e]	0.57/57	[m-u]	0.39/18
[n-z]	0.78/26	[R-l]	0.66/20	[s-o]	0.57/101	[i-l]	0.38/33
[e-r]	0.77/66	[<-e]	0.66/64	[m-e]	0.56/103	[q-o]	0.38/23
[l-a]	0.77/180	[t-a]	0.66/160	[b-l]	0.55/21	[l-e]	0.37/58
[n->]	0.76/61	[k-o]	0.65/118	[i-a]	0.55/132	[i-r]	0.35/33
[e-z]	0.76/70	[n-i]	0.65/31	[e->]	0.54/122	[o-i]	0.30/17
[k-R]	0.76/34	[i-b]	0.65/19	[d-u]	0.54/15	[e-e]	0.27/80
[e-R]	0.76/102	[<-l]	0.64/33	[s-u]	0.54/41	[a-g]	0.25/13
[z-i]	0.75/114	[i-R]	0.64/13	[i-o]	0.53/63	[e-d]	0.17/42
[b-e]	0.75/21	[p-l]	0.64/22	[k-i]	0.52/28	[n-l]	0.11/16
[e-n]	0.75/205	[o-s]	0.63/358	[u-r]	0.52/24		
[i-d]	0.74/43	[d-e]	0.63/232	[f-e]	0.51/12		
[m-o]	0.73/67	[s-a]	0.62/152	[<-o]	0.50/18		

q. Usuario MC

[R-m]	1.97/51	[e-m]	1.55/346	[i-e]	1.44/348	[s-b]	1.37/28
[l-b]	1.86/16	[l-s]	1.54/20	[a-e]	1.44/279	[n-s]	1.36/180
[e-y]	1.78/42	[l-p]	1.53/239	[a-m]	1.44/519	[p-i]	1.36/62
[i-r]	1.75/52	[i-m]	1.53/164	[e-n]	1.43/1228	[<-e]	1.36/700
[n-y]	1.71/45	[s-m]	1.52/89	[e-p]	1.43/322	[k-e]	1.35/747
[k-z]	1.70/32	[a-p]	1.49/477	[t-e]	1.42/795	[n-k]	1.35/138
[n-m]	1.69/12	[n-i]	1.49/215	[i-i]	1.42/48	[s-n]	1.34/47
[i-t]	1.67/190	[e-e]	1.49/611	[z-i]	1.41/616	[i-y]	1.33/15
[p-p]	1.64/13	[i-p]	1.48/38	[m-i]	1.40/277	[l-m]	1.33/65
[l-k]	1.63/34	[t-i]	1.48/307	[e-i]	1.40/164	[m-e]	1.32/306
[<-n]	1.61/64	[R-s]	1.47/56	[n-t]	1.39/634	[r-e]	1.31/140
[s-d]	1.59/215	[b-i]	1.47/242	[m-b]	1.39/68	[<-m]	1.31/75
[i-d]	1.58/293	[r-n]	1.47/36	[s-g]	1.39/14	[l-n]	1.30/18
[r-i]	1.57/74	[z-o]	1.47/39	[o-p]	1.39/262	[a-L]	1.30/38
[m-p]	1.56/119	[l-t]	1.47/74	[d-e]	1.39/1193	[n-d]	1.29/269
[s-i]	1.56/699	[l-e]	1.46/364	[R-i]	1.38/107	[n-e]	1.29/566
[e-t]	1.55/278	[L-o]	1.46/38	[d-i]	1.38/185	[u-l]	1.29/64
[l-i]	1.55/157	[l-d]	1.45/42	[a-a]	1.37/787	[a-s]	1.29/1401

[R-t]	1.29/72	[u-t]	1.12/75	[l-u]	1.01/50	[f-u]	0.87/57
[e-s]	1.28/1982	[R-p]	1.11/66	[p-l]	1.01/47	[R-n]	0.87/59
[d-u]	1.27/35	[e-r]	1.11/339	[s-a]	1.01/856	[o-f]	0.87/20
[o-m]	1.27/283	[~-o]	1.11/34	[p-o]	1.00/298	[<-b]	0.87/36
[o-c]	1.27/43	[m->]	1.10/51	[o-j]	1.00/24	[u-a]	0.86/57
[u-b]	1.27/50	[y-a]	1.10/76	[o-x]	1.00/13	[u-~]	0.86/15
[u-n]	1.27/436	[e-o]	1.10/144	[p-a]	1.00/526	[o-d]	0.86/418
[o-s]	1.26/1324	[o-k]	1.10/338	[j-o]	1.00/129	[u-s]	0.85/147
[e->]	1.25/475	[u-z]	1.09/18	[i-b]	0.99/86	[<-d]	0.85/107
[L-s]	1.25/14	[f-e]	1.08/58	[t-a]	0.98/752	[s-u]	0.85/315
[l-a]	1.25/811	[n-a]	1.08/446	[s-p]	0.98/290	[b-R]	0.84/62
[e-l]	1.25/1095	[e-k]	1.08/556	[n-n]	0.98/14	[e-R]	0.84/338
[m-o]	1.25/297	[d-a]	1.08/262	[k-a]	0.98/410	[k-t]	0.84/74
[a-t]	1.25/388	[<-p]	1.08/65	[z-a]	0.97/112	[s-j]	0.82/19
[e-f]	1.25/100	[R-e]	1.08/592	[a-u]	0.97/103	[a-y]	0.82/100
[a-i]	1.24/145	[u-m]	1.08/17	[i-k]	0.97/289	[<-u]	0.82/77
[b-e]	1.24/119	[x-u]	1.08/11	[g-a]	0.96/125	[k-o]	0.80/507
[f-i]	1.24/86	[c-o]	1.08/113	[o->]	0.96/729	[i-g]	0.78/45
[i-s]	1.24/476	[g-u]	1.07/122	[g-e]	0.96/45	[<-i]	0.78/161
[s-e]	1.23/1114	[r-a]	1.07/528	[p-e]	0.96/169	[R->]	0.77/144
[o-u]	1.23/58	[o-o]	1.07/287	[m-u]	0.95/46	[y-o]	0.77/43
[<-s]	1.23/80	[n-l]	1.06/94	[f-a]	0.95/63	[e-h]	0.77/12
[o-e]	1.22/289	[e-a]	1.06/281	[a->]	0.95/880	[o-g]	0.77/39
[i-a]	1.22/529	[o-r]	1.06/101	[R-l]	0.95/67	[a-f]	0.76/111
[o-i]	1.21/141	[a-l]	1.05/545	[n-u]	0.95/63	[a-j]	0.76/197
[t-u]	1.20/58	[o-y]	1.05/32	[R-o]	0.95/209	[s-k]	0.76/258
[i-n]	1.20/285	[e-u]	1.05/144	[y-e]	0.95/70	[o-R]	0.75/232
[e-d]	1.19/441	[a-z]	1.05/279	[<-k]	0.94/92	[j-u]	0.75/13
[o-n]	1.19/665	[e-z]	1.05/319	[p-u]	0.94/163	[n-b]	0.75/62
[R-d]	1.18/63	[a-r]	1.04/319	[e-g]	0.94/138	[g-o]	0.75/129
[b-l]	1.17/114	[i->]	1.04/85	[n-o]	0.93/475	[o-a]	0.74/189
[n-z]	1.17/138	[k-i]	1.04/85	[p-R]	0.93/306	[s->]	0.73/420
[a-k]	1.17/416	[s-l]	1.03/73	[n-f]	0.93/21	[i-f]	0.73/33
[L-e]	1.17/44	[u-e]	1.03/204	[e-L]	0.92/17	[a-c]	0.72/11
[u-i]	1.15/41	[t-R]	1.03/180	[a-g]	0.92/131	[e-c]	0.72/90
[a-n]	1.15/706	[<-l]	1.03/45	[l-o]	0.92/367	[i-z]	0.71/169
[s-o]	1.14/504	[i-L]	1.03/44	[n-p]	0.92/29	[i-c]	0.71/25
[o-l]	1.14/146	[s-t]	1.03/621	[n->]	0.91/220	[k-u]	0.70/62
[u-c]	1.14/44	[e-b]	1.03/180	[a-b]	0.91/470	[c-a]	0.68/42
[e-x]	1.13/34	[R-a]	1.02/206	[a-~]	0.89/87	[R-k]	0.67/47
[z-e]	1.13/248	[f-o]	1.02/47	[d-o]	0.89/494	[R-z]	0.67/17
[j-e]	1.13/178	[l-g]	1.02/115	[o-b]	0.89/215	[j-a]	0.67/38
[m-a]	1.13/437	[r-s]	1.02/16	[a-R]	0.88/535	[~-a]	0.61/77
[b-o]	1.13/75	[i-R]	1.01/90	[r-o]	0.88/181	[u-r]	0.57/111
[i-o]	1.13/334	[o-t]	1.01/152	[l-z]	0.88/16	[l->]	0.56/108
[b-a]	1.13/489	[a-d]	1.01/559	[a-o]	0.87/111	[<-R]	0.56/15
[k-l]	1.12/49	[<-a]	1.01/310	[o-z]	0.87/79	[k-R]	0.52/60
[i-l]	1.12/114	[g-i]	1.01/21	[t-o]	0.87/573	[i-j]	0.51/51

[<-o]	0.47/32	[m-p]	1.08/119	[m-o]	0.92/297	[l-m]	0.80/65
[R-u]	0.46/36	[a-s]	1.07/1401	[i-c]	0.92/25	[b-i]	0.80/242
[e-j]	0.42/42	[<-s]	1.07/80	[a-m]	0.92/519	[b-e]	0.80/119
[b-u]	0.36/23	[i-p]	1.07/38	[n-d]	0.91/269	[e-f]	0.80/100
[s-y]	0.35/33	[e-n]	1.07/1228	[e-m]	0.91/346	[s-y]	0.80/33
[r->]	0.25/20	[y-e]	1.07/70	[z-a]	0.91/112	[n-u]	0.79/63
[<-y]	0.23/22	[n-k]	1.06/138	[u-n]	0.91/436	[s-o]	0.79/504
[<-j]	0.08/41	[n-s]	1.06/180	[i-o]	0.91/334	[e-k]	0.79/556
[y-l]	-0.51/23	[o-s]	1.06/1324	[a-L]	0.90/38	[s-b]	0.79/28
[n-n]	1.70/14	[r-e]	1.05/140	[j-e]	0.90/178	[a-z]	0.78/279
[<-n]	1.50/64	[o-p]	1.05/262	[a-u]	0.89/103	[o-k]	0.78/338
[k-z]	1.48/32	[l-b]	1.05/16	[o-c]	0.89/43	[s-a]	0.78/856
[l-k]	1.46/34	[o-x]	1.05/13	[e-z]	0.89/319	[a-d]	0.78/559
[i-r]	1.45/52	[k-e]	1.05/747	[s-t]	0.89/621	[a-i]	0.78/145
[e-y]	1.45/42	[R-m]	1.05/51	[s-e]	0.89/1114	[u-s]	0.78/147
[R-t]	1.33/72	[o-m]	1.04/283	[l-u]	0.89/50	[o->]	0.78/729
[i-d]	1.32/293	[<-m]	1.04/75	[e-d]	0.88/441	[a-r]	0.77/319
[l-s]	1.28/20	[i-n]	1.04/285	[e-i]	0.88/164	[n-m]	0.77/12
[d-e]	1.27/1193	[m-i]	1.04/277	[i-g]	0.88/45	[a-k]	0.77/416
[t-e]	1.26/795	[o-e]	1.03/289	[<-k]	0.87/92	[n-p]	0.76/29
[n-i]	1.26/215	[e-s]	1.02/1982	[i-y]	0.87/15	[z-e]	0.76/248
[l-d]	1.25/42	[z-i]	1.02/616	[o-i]	0.87/141	[u-t]	0.76/75
[p-p]	1.24/13	[l-e]	1.02/364	[R-o]	0.86/209	[a->]	0.76/880
[e-t]	1.23/278	[d-i]	1.02/185	[o-y]	0.86/32	[o-l]	0.76/146
[R-s]	1.21/56	[p-i]	1.01/62	[p-l]	0.85/47	[b-o]	0.76/75
[n-y]	1.20/45	[g-i]	1.01/21	[n->]	0.85/220	[l-a]	0.75/811
[u-m]	1.20/17	[i-s]	1.00/476	[<-e]	0.85/700	[s-p]	0.75/290
[a-p]	1.20/477	[f-i]	0.99/86	[s-n]	0.85/47	[f-o]	0.75/47
[i-m]	1.19/164	[R-d]	0.99/63	[o-n]	0.85/665	[a-b]	0.75/470
[l-i]	1.18/157	[r-n]	0.98/36	[j-a]	0.84/38	[a-g]	0.75/131
[s-m]	1.18/89	[l-n]	0.98/18	[i->]	0.84/85	[o-b]	0.75/215
[<-p]	1.17/65	[e-x]	0.98/34	[m-u]	0.84/46	[p-u]	0.75/163
[e-e]	1.17/611	[r-i]	0.98/74	[d-u]	0.84/35	[R-e]	0.74/592
[z-o]	1.17/39	[u-i]	0.98/41	[i-a]	0.84/529	[R-l]	0.74/67
[s-d]	1.17/215	[~-o]	0.97/34	[m->]	0.84/51	[i-b]	0.74/86
[l-t]	1.16/74	[<-b]	0.97/36	[s-g]	0.83/14	[L-o]	0.73/38
[L-s]	1.16/14	[a-a]	0.96/787	[r-s]	0.83/16	[n-a]	0.73/446
[s-i]	1.15/699	[i-R]	0.96/90	[m-b]	0.83/68	[e-g]	0.73/138
[n-t]	1.15/634	[a-t]	0.96/388	[e->]	0.83/475	[b-l]	0.73/114
[e-p]	1.15/322	[R-i]	0.95/107	[n-l]	0.82/94	[o-t]	0.73/152
[u-l]	1.13/64	[t-u]	0.95/58	[n-z]	0.81/138	[y-a]	0.72/76
[i-e]	1.13/348	[o-j]	0.94/24	[e-o]	0.81/144	[i-k]	0.72/289
[l-z]	1.13/16	[i-l]	0.94/114	[u-c]	0.80/44	[k-R]	0.72/60
[i-t]	1.10/190	[e-l]	0.94/1095	[e-u]	0.80/144	[i-L]	0.72/44
[l-p]	1.10/239	[s-j]	0.93/19	[a-n]	0.80/706	[R-z]	0.71/17
[u-z]	1.10/18	[u-b]	0.93/50	[c-o]	0.80/113	[k-a]	0.71/410
[t-i]	1.09/307	[a-e]	0.93/279	[p-e]	0.80/169	[t-a]	0.71/752
[n-e]	1.08/566	[m-e]	0.93/306	[a-l]	0.80/545	[e-r]	0.71/339

[k-i]	0.71/85	[f-a]	0.66/63	[c-a]	0.59/42	[a-y]	0.47/100
[d-a]	0.71/262	[R-p]	0.66/66	[e-a]	0.59/281	[u-a]	0.46/57
[f-e]	0.71/58	[R-a]	0.66/206	[e-R]	0.59/338	[e-L]	0.45/17
[a-R]	0.71/535	[k-t]	0.65/74	[o-a]	0.58/189	[k-u]	0.45/62
[l-g]	0.70/115	[a-~]	0.65/87	[o-r]	0.57/101	[R->]	0.45/144
[p-R]	0.70/306	[g-u]	0.65/122	[g-a]	0.56/125	[k-l]	0.43/49
[s-u]	0.70/315	[a-f]	0.65/111	[j-o]	0.56/129	[r->]	0.41/20
[i-i]	0.70/48	[o-u]	0.65/58	[o-z]	0.55/79	[e-c]	0.39/90
[i-z]	0.70/169	[<-d]	0.64/107	[<-i]	0.54/161	[o-g]	0.39/39
[e-h]	0.69/12	[<-l]	0.64/45	[b-R]	0.54/62	[<-o]	0.35/32
[e-b]	0.69/180	[o-o]	0.64/287	[d-o]	0.52/494	[u-r]	0.32/111
[p-a]	0.69/526	[x-u]	0.64/11	[r-a]	0.52/528	[n-b]	0.30/62
[R-n]	0.69/59	[<-a]	0.64/310	[i-j]	0.52/51	[<-u]	0.29/77
[l->]	0.69/108	[p-o]	0.63/298	[k-o]	0.52/507	[y-l]	0.24/23
[t-o]	0.69/573	[f-u]	0.63/57	[o-R]	0.52/232	[y-o]	0.20/43
[t-R]	0.68/180	[R-k]	0.63/47	[o-f]	0.51/20	[<-y]	0.17/22
[n-o]	0.68/475	[R-u]	0.62/36	[a-o]	0.50/111	[j-u]	0.16/13
[u-e]	0.67/204	[m-a]	0.61/437	[L-e]	0.50/44	[b-u]	0.15/23
[o-d]	0.66/418	[u-~]	0.61/15	[a-c]	0.50/11	[e-j]	0.09/42
[r-o]	0.66/181	[g-e]	0.61/45	[s->]	0.50/420	[<-R]	0.04/15
[b-a]	0.66/489	[s-k]	0.60/258	[n-f]	0.50/21	[<-j]	-0.27/41
[s-l]	0.66/73	[g-o]	0.60/129	[i-f]	0.49/33		
[a-j]	0.66/197	[l-o]	0.59/367	[~-a]	0.48/77		

r. Usuario NR

[e-i]	1.45/12	[i-d]	1.02/108	[o-R]	0.88/91	[l-g]	0.81/48
[a-j]	1.42/21	[d-a]	1.02/114	[s-g]	0.87/20	[R-s]	0.80/43
[l-t]	1.37/15	[l-m]	1.02/26	[L-a]	0.87/12	[<-l]	0.80/39
[o-p]	1.34/57	[n->]	1.02/68	[R-m]	0.87/38	[b-o]	0.80/28
[R-t]	1.32/36	[o-k]	1.00/179	[i-z]	0.86/36	[e-j]	0.79/47
[i-g]	1.31/11	[s-z]	0.99/13	[a->]	0.86/206	[i-r]	0.78/27
[n-g]	1.31/14	[m-p]	0.99/81	[s-d]	0.85/52	[<-d]	0.78/30
[o-t]	1.27/70	[l-s]	0.98/14	[z-a]	0.85/53	[e-f]	0.78/55
[a-p]	1.18/126	[s-p]	0.96/47	[e-s]	0.85/533	[R-i]	0.77/40
[n-z]	1.17/17	[n-d]	0.96/87	[e-t]	0.84/134	[s-u]	0.77/81
[e-p]	1.16/49	[<-p]	0.96/27	[R-d]	0.84/23	[s->]	0.77/202
[a-t]	1.16/180	[k-i]	0.95/19	[R->]	0.84/17	[<-s]	0.76/17
[R-z]	1.15/15	[m-b]	0.95/32	[m-i]	0.84/37	[s-a]	0.76/197
[n-t]	1.14/221	[o-s]	0.95/292	[r-R]	0.84/14	[m-a]	0.76/142
[d-t]	1.11/13	[s-k]	0.95/65	[p-R]	0.84/69	[l-k]	0.76/30
[n-p]	1.08/24	[o->]	0.95/296	[t-a]	0.83/215	[j-o]	0.76/36
[r-n]	1.07/23	[u-r]	0.94/12	[<-u]	0.83/14	[d-s]	0.75/16
[R-n]	1.05/60	[i-n]	0.94/120	[t-u]	0.83/39	[<-e]	0.75/198
[l-p]	1.04/30	[o-j]	0.94/16	[n-b]	0.82/16	[a-s]	0.75/350
[n-y]	1.04/15	[s-m]	0.94/55	[p-a]	0.81/107	[k-z]	0.75/16
[R-k]	1.03/43	[n-k]	0.90/52	[z-i]	0.81/142	[n-a]	0.75/180
[k-t]	1.03/16	[u-a]	0.89/35	[a-k]	0.81/141	[i-o]	0.74/112

[k-a]	0.74/190	[i-R]	0.66/50	[a-n]	0.58/209	[a-u]	0.51/32
[s-o]	0.74/118	[i-f]	0.66/29	[n-n]	0.58/16	[k-o]	0.50/239
[l-o]	0.74/122	[n-s]	0.66/28	[L-o]	0.58/12	[<-k]	0.49/78
[i-t]	0.74/21	[t-y]	0.65/20	[e-g]	0.56/16	[s-e]	0.48/246
[R-a]	0.73/37	[f-u]	0.65/54	[m-u]	0.56/19	[a-m]	0.47/92
[s-t]	0.73/172	[p-u]	0.65/45	[i-c]	0.56/30	[<-a]	0.47/158
[e-k]	0.73/151	[z-e]	0.64/121	[o-n]	0.56/278	[e-d]	0.46/79
[d-o]	0.73/183	[i-p]	0.64/47	[n-o]	0.55/194	[m-e]	0.46/115
[f-o]	0.72/66	[b-s]	0.63/19	[u-t]	0.55/20	[i-a]	0.46/145
[r-a]	0.72/100	[g-a]	0.63/41	[t-R]	0.55/66	[a-r]	0.46/56
[a-b]	0.72/132	[a-L]	0.63/16	[e-o]	0.55/51	[R-e]	0.45/198
[u-z]	0.71/26	[e-a]	0.62/102	[m-o]	0.54/126	[g-u]	0.45/49
[l->]	0.71/44	[o-m]	0.62/132	[s-s]	0.54/19	[o-z]	0.44/22
[i-i]	0.70/22	[l-a]	0.62/265	[a-l]	0.54/250	[o-e]	0.44/14
[b-a]	0.70/85	[r-d]	0.61/11	[i-s]	0.54/88	[<-m]	0.44/17
[t-o]	0.70/174	[a-a]	0.61/97	[a-e]	0.54/47	[e-c]	0.43/34
[o-b]	0.70/78	[<-n]	0.61/26	[e-l]	0.54/205	[k-R]	0.42/32
[o-d]	0.70/86	[a-y]	0.60/22	[a-g]	0.54/40	[c-o]	0.42/53
[p-e]	0.69/68	[a-z]	0.60/84	[b-l]	0.53/35	[u-l]	0.41/13
[a-f]	0.69/16	[r-o]	0.60/69	[j-a]	0.53/33	[k-s]	0.41/14
[t-i]	0.69/154	[a-d]	0.60/254	[i-b]	0.53/43	[i-k]	0.41/149
[e-b]	0.69/52	[e-n]	0.59/433	[g-o]	0.53/39	[s-l]	0.39/23
[s-i]	0.68/219	[l-e]	0.59/80	[n-l]	0.53/38	[m-r]	0.18/21
[p-o]	0.68/119	[i->]	0.59/52	[e-e]	0.52/65		
[o-i]	0.68/30	[y-o]	0.59/29	[u-n]	0.52/144		
[e-z]	0.67/77	[k-u]	0.59/102	[<-y]	0.52/17		
[e->]	0.67/151	[k-e]	0.59/380	[R-o]	0.52/65		
[t-e]	0.67/223	[d-e]	0.58/253	[k-l]	0.51/14		

s. Usuario VL

[R-m]	1.82/12	[m-e]	1.20/22	[d-e]	1.12/87	[l-p]	1.06/14
[e-k]	1.57/13	[i-R]	1.20/31	[u-m]	1.12/13	[n-k]	1.06/36
[o-p]	1.56/74	[e-n]	1.19/135	[m-i]	1.11/14	[e-R]	1.05/56
[i-m]	1.44/15	[m-u]	1.19/12	[e-e]	1.11/31	[l-o]	1.05/79
[o-l]	1.43/16	[f-e]	1.18/19	[b-a]	1.10/22	[a-b]	1.05/15
[s-d]	1.42/21	[u-r]	1.17/27	[l-k]	1.10/16	[a-n]	1.04/40
[u-p]	1.41/15	[o-a]	1.17/12	[r-a]	1.10/28	[a-l]	1.04/18
[d-o]	1.34/45	[p-o]	1.16/58	[o-o]	1.09/13	[e-z]	1.03/47
[e-b]	1.33/25	[u-e]	1.16/17	[r-o]	1.09/54	[s-k]	1.03/31
[l-i]	1.33/13	[o->]	1.15/67	[b-e]	1.09/31	[e-u]	1.03/29
[e-p]	1.32/30	[o-n]	1.15/66	[i-e]	1.09/50	[R->]	1.03/12
[t-o]	1.30/26	[i-z]	1.15/21	[a-k]	1.09/77	[n-o]	1.03/39
[t-e]	1.26/51	[a-d]	1.14/45	[a-r]	1.07/46	[<-k]	1.01/17
[z-e]	1.26/31	[n-t]	1.13/36	[u-s]	1.07/14	[<-i]	1.01/13
[s-u]	1.23/47	[m-b]	1.13/13	[l-e]	1.07/32	[o-k]	1.01/40
[a-p]	1.22/36	[e-o]	1.13/25	[a-t]	1.07/11	[z-i]	1.01/48
[<-m]	1.20/26	[r-e]	1.12/49	[s-p]	1.06/22	[i-n]	1.01/42

[a->]	1.01/38	[l-a]	0.95/59	[e-g]	0.89/50	[b-i]	0.81/27
[g-u]	1.01/16	[u-a]	0.95/13	[R-o]	0.89/17	[e-l]	0.80/97
[t-u]	1.00/12	[e-s]	0.95/222	[k-u]	0.89/27	[o-R]	0.78/64
[i-s]	1.00/35	[p-e]	0.95/62	[p-u]	0.89/18	[R-l]	0.78/23
[e-r]	1.00/42	[R-e]	0.95/54	[<-s]	0.88/23	[b-R]	0.77/18
[d-u]	1.00/17	[r-i]	0.94/15	[k-a]	0.87/63	[s->]	0.77/71
[b-o]	0.99/21	[d-i]	0.93/36	[a-s]	0.87/90	[<-p]	0.76/14
[n-a]	0.98/20	[e-d]	0.93/37	[n-d]	0.86/44	[<-l]	0.75/54
[p-i]	0.98/16	[<-e]	0.93/22	[d-a]	0.86/13	[m-a]	0.70/27
[g-a]	0.98/12	[p-a]	0.92/58	[i-k]	0.85/18	[i-d]	0.67/32
[o-r]	0.97/27	[a-e]	0.92/56	[k-l]	0.85/19	[R-p]	0.66/25
[k-i]	0.97/51	[e->]	0.92/79	[n-p]	0.85/11	[s-l]	0.65/11
[s-t]	0.97/81	[e-f]	0.92/30	[i-a]	0.85/19	[i-l]	0.58/13
[t-a]	0.97/73	[k-o]	0.92/52	[s-a]	0.84/34	[e-a]	0.51/27
[n-e]	0.97/61	[o-m]	0.91/34	[k-e]	0.83/167	[R-b]	0.38/12
[s-o]	0.96/48	[j-i]	0.91/18	[R-i]	0.83/25	[g-R]	0.26/12
[R-k]	0.96/48	[n-s]	0.90/36	[<-a]	0.83/13		
[o-s]	0.96/107	[u-n]	0.90/34	[f-o]	0.83/30		
[i-o]	0.96/16	[a-o]	0.90/25	[s-i]	0.82/76		
[s-e]	0.95/102	[a-R]	0.89/39	[p-R]	0.82/24		

t. Usuario VP

[l-p]	1.36/129	[<-p]	1.10/48	[f-o]	0.99/45	[e-t]	0.94/144
[n-d]	1.31/34	[R-a]	1.09/185	[s->]	0.99/230	[R-e]	0.94/61
[i-n]	1.29/33	[r-a]	1.09/185	[i-k]	0.99/167	[r-e]	0.94/61
[o-p]	1.26/130	[e-m]	1.09/48	[e-o]	0.99/93	[R->]	0.94/164
[o-o]	1.25/7	[u-p]	1.08/41	[<-l]	0.99/45	[j-e]	0.93/94
[a-o]	1.24/45	[l-o]	1.07/220	[i-b]	0.98/36	[i-l]	0.93/24
[a-n]	1.24/110	[z-e]	1.07/126	[a-t]	0.98/187	[f-e]	0.93/51
[j-a]	1.22/29	[e-k]	1.06/269	[e-p]	0.97/198	[s-d]	0.93/51
[u-n]	1.21/53	[n-i]	1.05/14	[s-a]	0.97/307	[t-e]	0.92/393
[<-m]	1.21/22	[z-a]	1.05/38	[n-s]	0.97/88	[o-m]	0.92/40
[d-a]	1.16/229	[m-u]	1.05/24	[o-a]	0.97/103	[o-i]	0.91/38
[R-d]	1.16/7	[t-a]	1.04/443	[k-e]	0.96/593	[k-t]	0.91/53
[a-m]	1.14/104	[k-a]	1.04/263	[y-o]	0.96/130	[e-j]	0.91/43
[l-a]	1.14/439	[p-a]	1.02/203	[u-t]	0.96/44	[l-i]	0.90/160
[R-k]	1.14/46	[m-o]	1.02/39	[e-z]	0.96/138	[a-k]	0.90/259
[p-o]	1.13/239	[b-a]	1.01/154	[n-o]	0.96/83	[u-a]	0.90/67
[g-u]	1.12/11	[o-t]	1.01/111	[a-r]	0.96/113	[a-i]	0.90/102
[R-o]	1.11/105	[k-o]	1.01/117	[a-R]	0.96/113	[<-n]	0.90/12
[r-o]	1.11/105	[a-u]	1.00/10	[a-p]	0.96/164	[i-t]	0.89/140
[s-m]	1.11/23	[e-e]	1.00/114	[b-i]	0.95/26	[e-a]	0.89/129
[e-n]	1.11/149	[o-k]	1.00/356	[p-e]	0.95/213	[R-s]	0.89/14
[x-i]	1.10/18	[n-l]	1.00/47	[s-o]	0.95/229	[u-e]	0.89/96
[r-i]	1.10/55	[o-n]	0.99/137	[i-z]	0.94/95	[a-e]	0.89/101
[R-i]	1.10/55	[d-e]	0.99/461	[i-d]	0.94/232	[g-o]	0.89/10

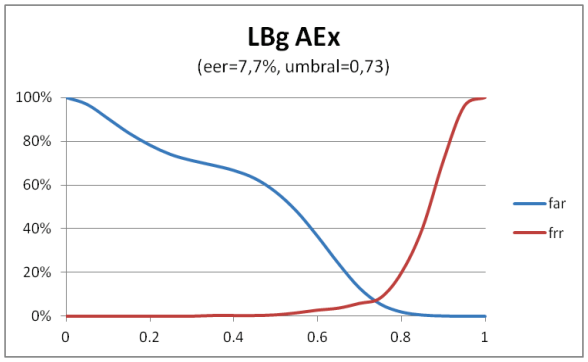
[b-e]	0.89/113	[p-u]	0.86/96	[i-R]	0.82/35	[u-r]	0.75/10
[<-s]	0.89/84	[a-z]	0.86/150	[i-r]	0.82/35	[n-z]	0.75/44
[<-k]	0.88/95	[t-u]	0.85/59	[o->]	0.82/594	[s-t]	0.75/379
[<-d]	0.88/34	[p-R]	0.85/49	[n-a]	0.82/213	[m-p]	0.75/97
[e-u]	0.88/6	[n-t]	0.85/416	[o-z]	0.82/93	[u-d]	0.75/44
[l-m]	0.87/39	[e-b]	0.84/83	[i-j]	0.81/14	[n-k]	0.75/88
[o-R]	0.87/102	[s-i]	0.84/86	[i-p]	0.81/43	[m-a]	0.73/229
[o-r]	0.87/102	[d-o]	0.84/291	[e-s]	0.81/897	[l-s]	0.72/27
[b-l]	0.87/7	[t-i]	0.84/231	[o-d]	0.79/155	[s-k]	0.72/124
[n-p]	0.87/39	[n->]	0.83/67	[i-s]	0.79/219	[p-l]	0.72/43
[i-a]	0.87/307	[t-R]	0.83/113	[a-d]	0.79/324	[a-j]	0.71/82
[k-i]	0.87/12	[c-o]	0.83/79	[i-o]	0.79/139	[l-t]	0.70/38
[t-o]	0.87/391	[e-g]	0.83/76	[e-x]	0.79/27	[e->]	0.67/278
[e-R]	0.86/159	[<-a]	0.83/124	[o-e]	0.79/101	[n-e]	0.67/220
[e-r]	0.86/159	[o-s]	0.83/553	[e-l]	0.78/424	[i-e]	0.65/245
[i-m]	0.86/25	[m-e]	0.83/202	[a-b]	0.78/175	[a-g]	0.65/37
[s-e]	0.86/384	[a-a]	0.83/72	[<-e]	0.76/231	[a->]	0.62/356
[<-i]	0.86/25	[d-i]	0.82/24	[a-s]	0.76/538	[i->]	0.56/56
[l-e]	0.86/237	[s-p]	0.82/97	[s-u]	0.76/105		
[g-a]	0.86/39	[k-u]	0.82/114	[o-b]	0.76/13		
[z-i]	0.86/270	[e-d]	0.82/144	[a-l]	0.76/442		

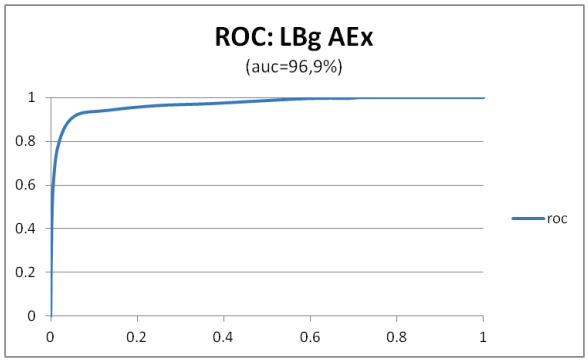
.

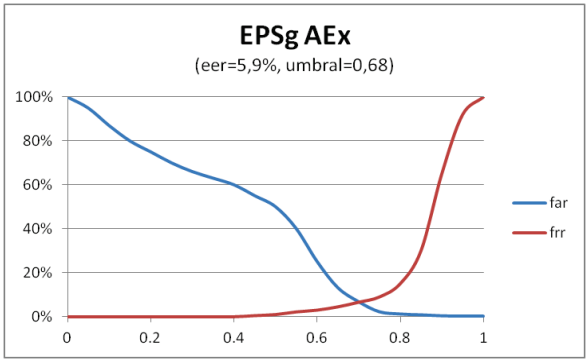
Apéndice III. Resultados individualizados

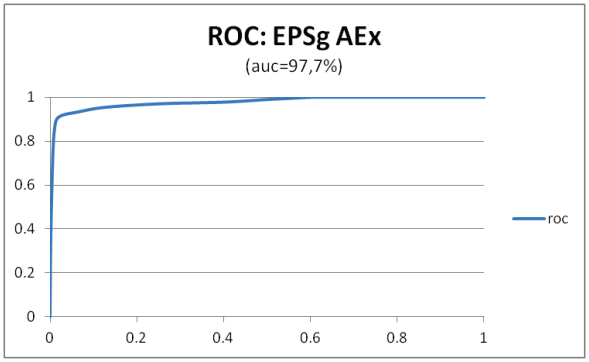
En las siguientes secciones que componen este apéndice se presentan las graficas obtenidas en los test para cada uno de los usuarios.

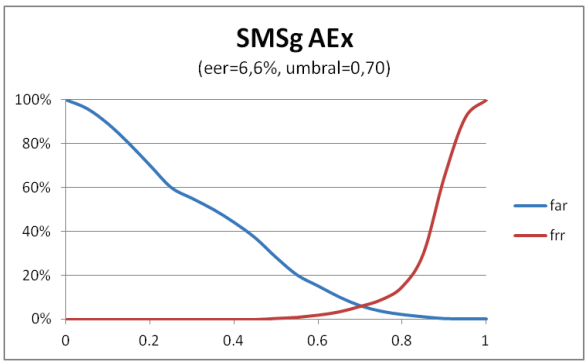
a. Usuario AEx

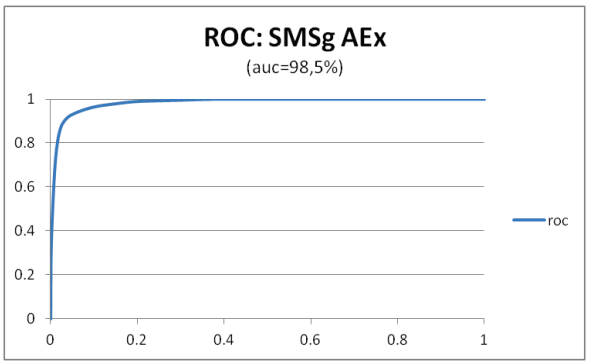


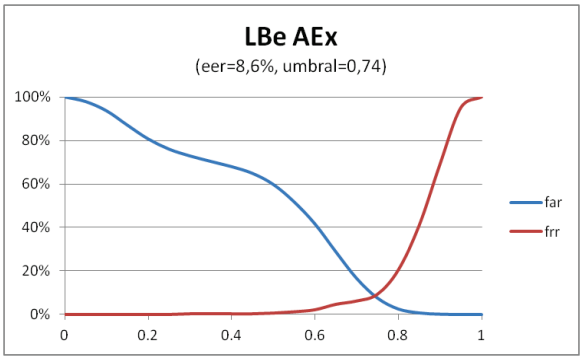


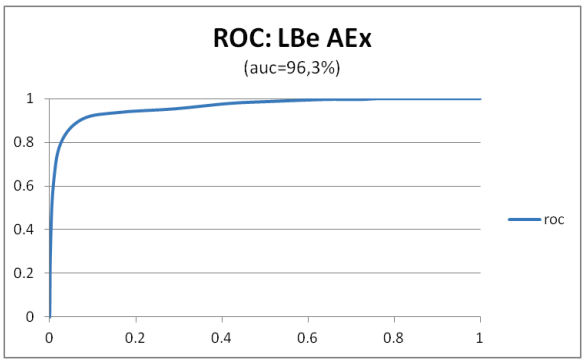


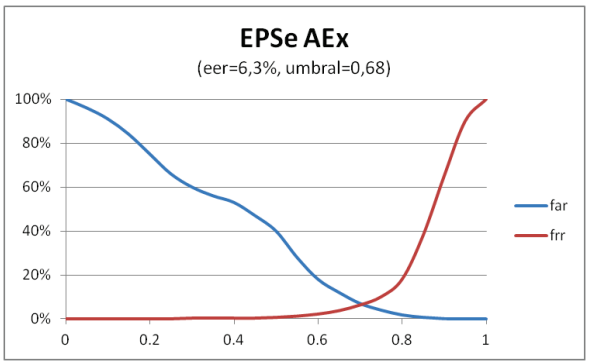


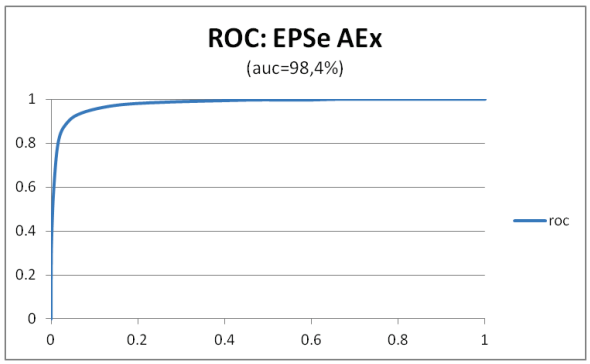


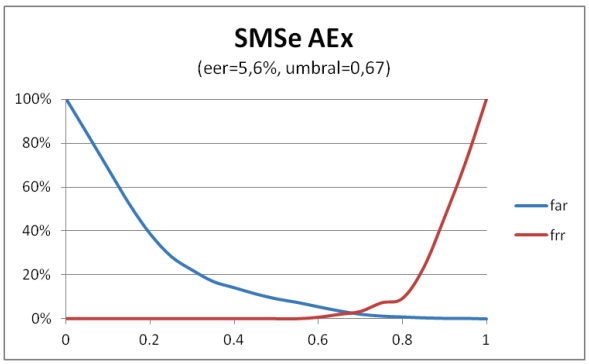


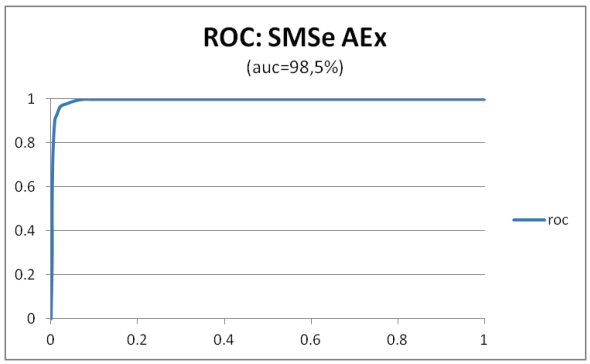

LBe AEx
(eer=8,6%, umbral=0,74)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr


ROC: LBe AEx
(auc=96,3%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

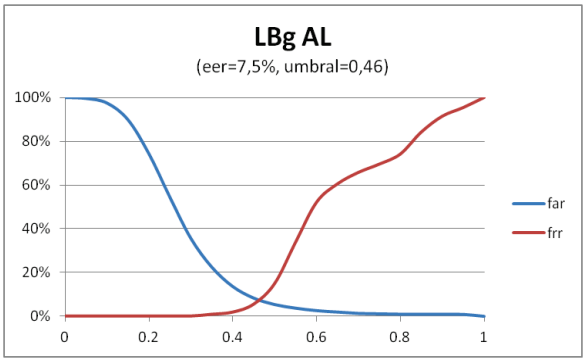

EPSe AEx
(eer=6,3%, umbral=0,68)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

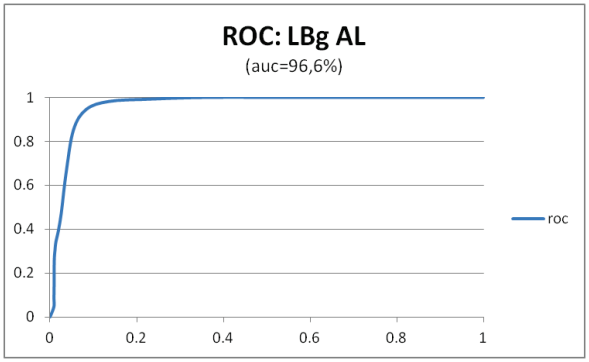

ROC: EPSe AEx
(auc=98,4%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

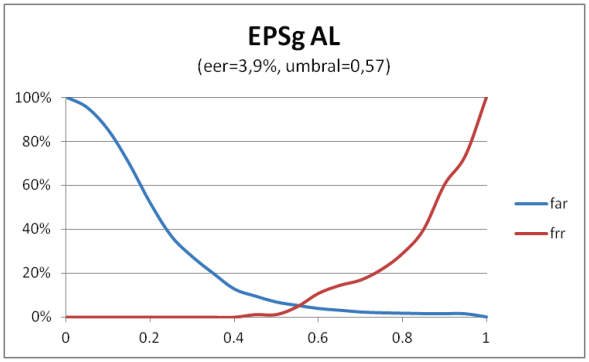

SMSe AEx
(eer=5,6%, umbral=0,67)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

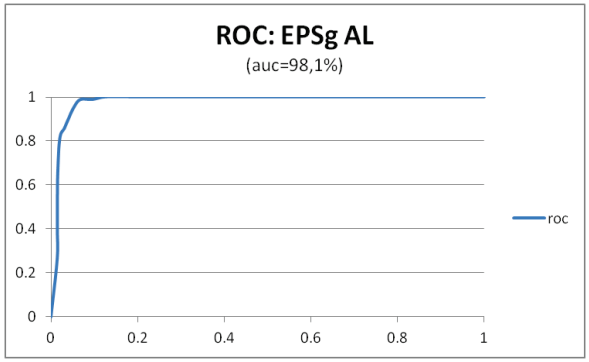

ROC: SMSe AEx
(auc=98,5%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

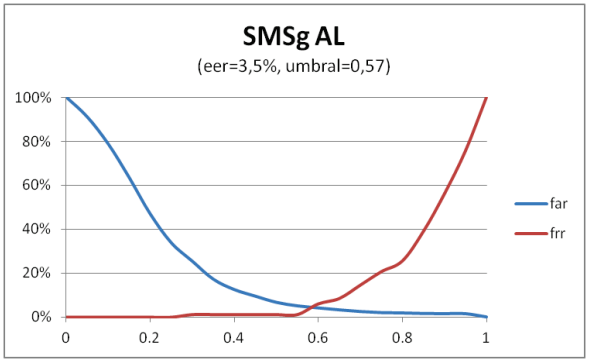
b. Usuario AL

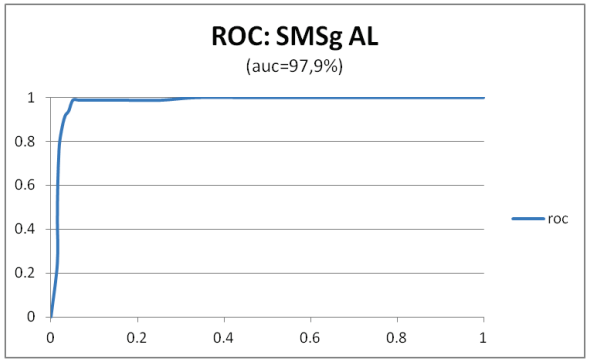

LBg AL
(eer=7,5%, umbral=0,46)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

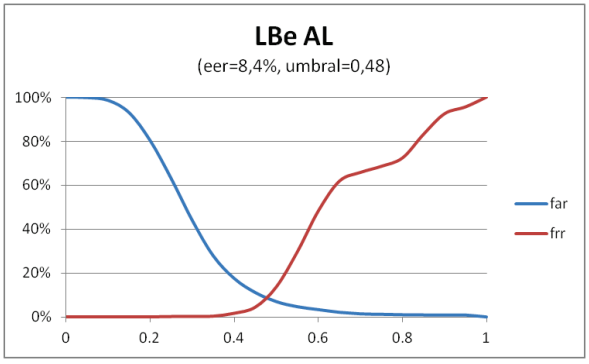

ROC: LBg AL
(auc=96,6%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

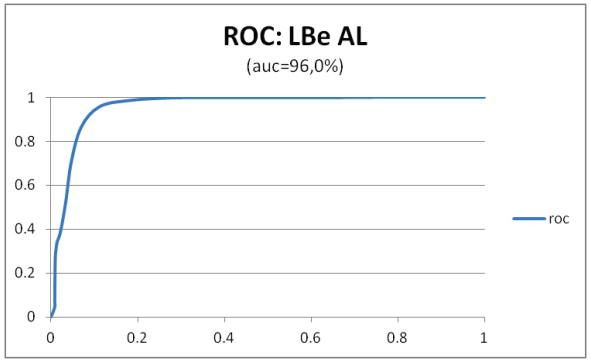

EPSg AL
(eer=3,9%, umbral=0,57)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

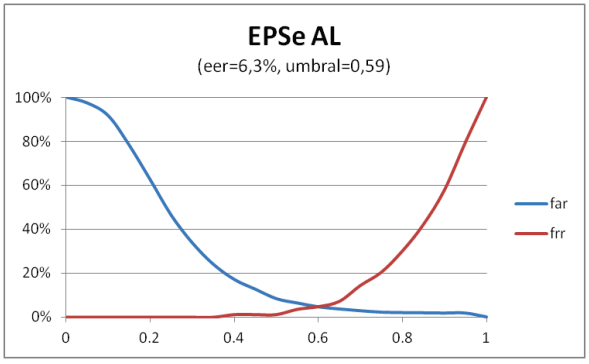

ROC: EPSg AL
(auc=98,1%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

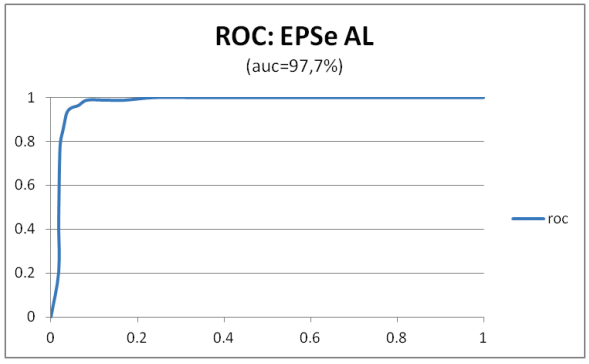

SMSg AL
(eer=3,5%, umbral=0,57)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

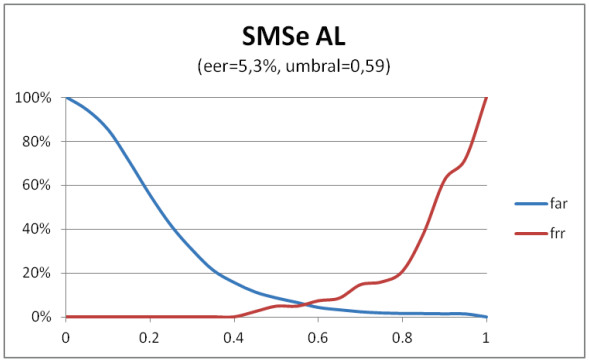

ROC: SMSg AL
(auc=97,9%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

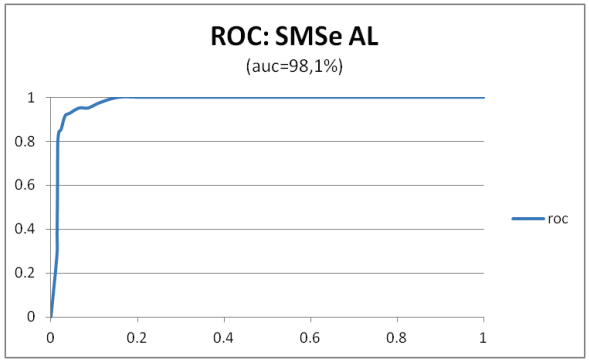

LBe AL
(eer=8,4%, umbral=0,48)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr


ROC: LBe AL
(auc=96,0%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

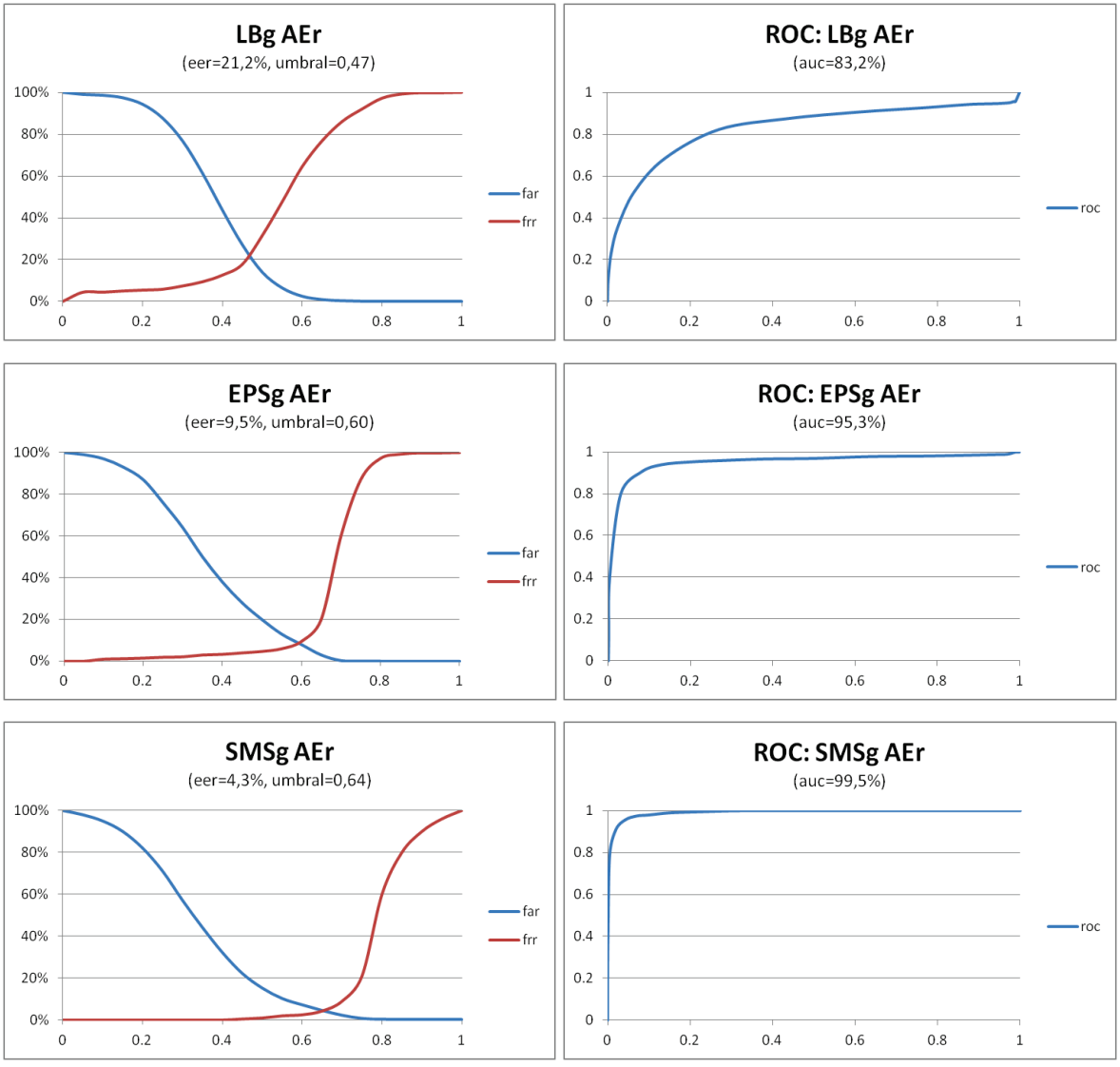

EPSe AL
(eer=6,3%, umbral=0,59)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

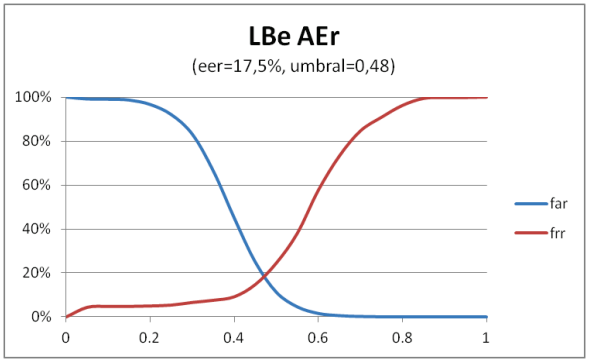

ROC: EPSe AL
(auc=97,7%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

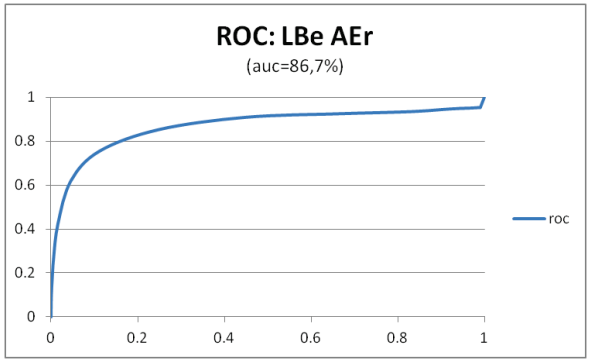

SMSe AL
(eer=5,3%, umbral=0,59)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

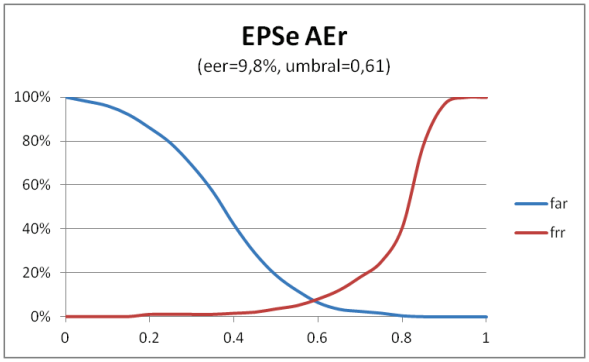

ROC: SMSe AL
(auc=98,1%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

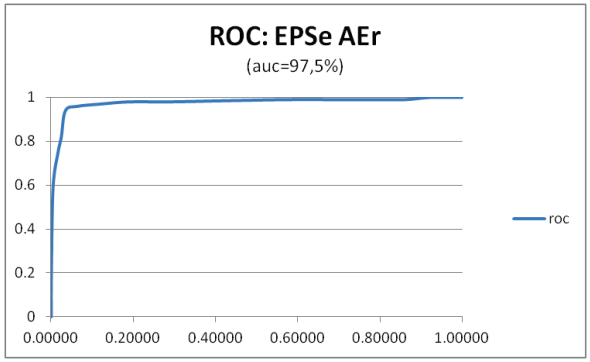
c. Usuario AEr

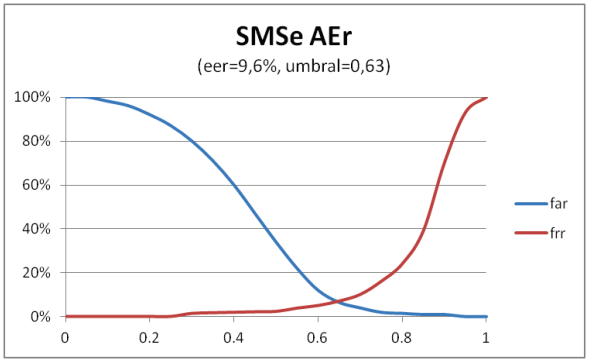


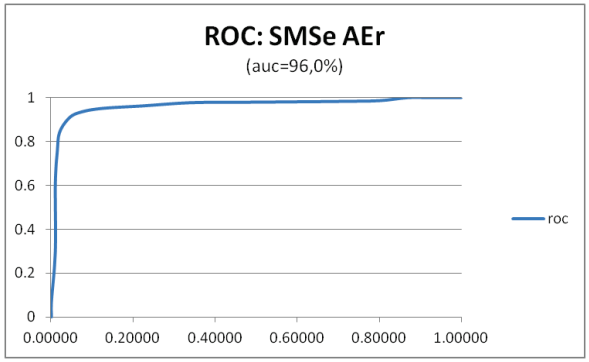

LBe AEr
(eer=17,5%, umbral=0,48)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr


ROC: LBe AEr
(auc=86,7%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

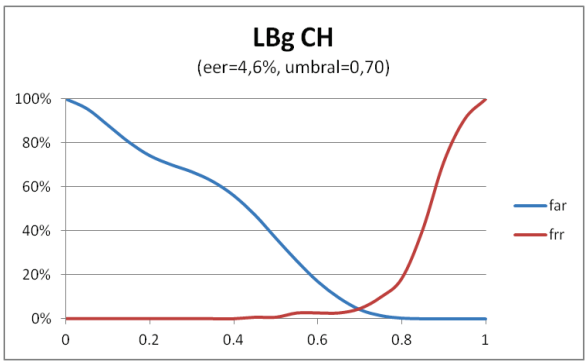

EPSe AEr
(eer=9,8%, umbral=0,61)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

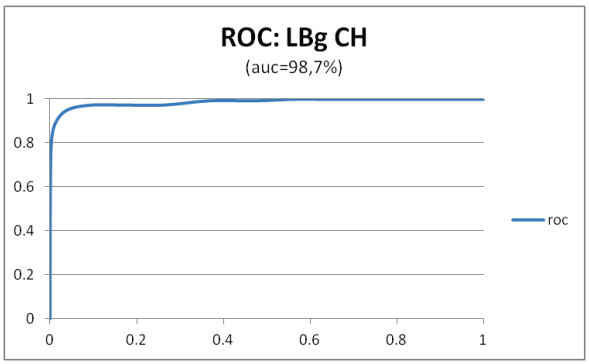

ROC: EPSe AEr
(auc=97,5%)
1
0.8
0.6
0.4
0.2
0
0.00000
0.20000
0.40000
0.60000
0.80000
1.00000
roc

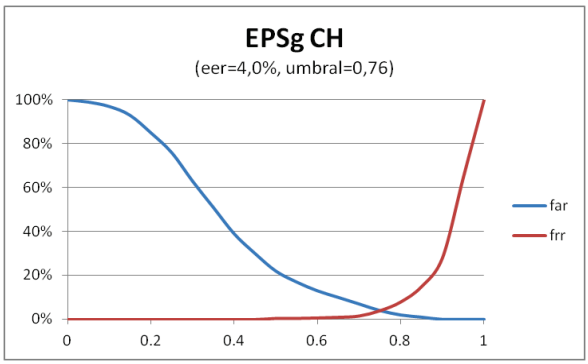

SMSe AEr
(eer=9,6%, umbral=0,63)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

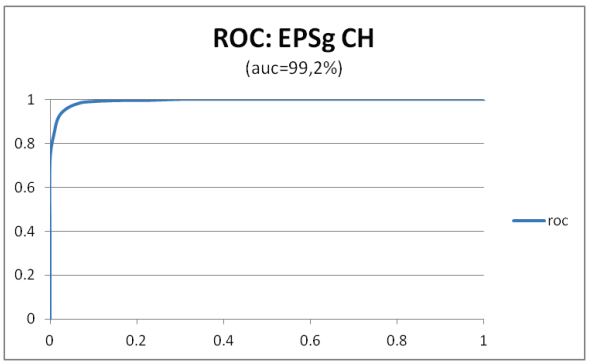

ROC: SMSe AEr
(auc=96,0%)
1
0.8
0.6
0.4
0.2
0
0.00000
0.20000
0.40000
0.60000
0.80000
1.00000
roc

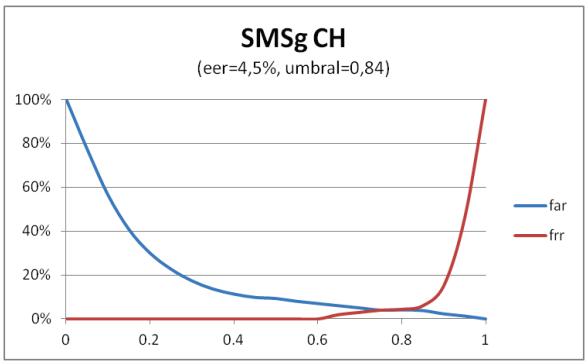
d. Usuario CH

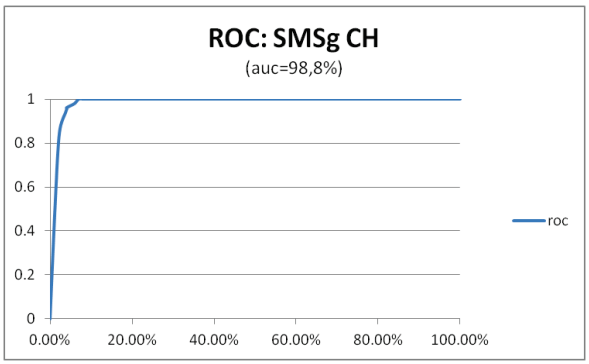

LBg CH
(eer=4,6%, umbral=0,70)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

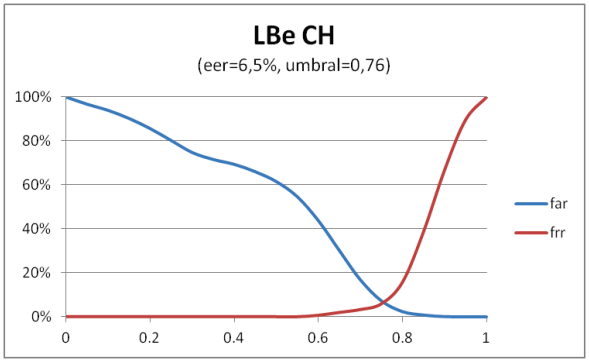

ROC: LBg CH
(auc=98,7%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

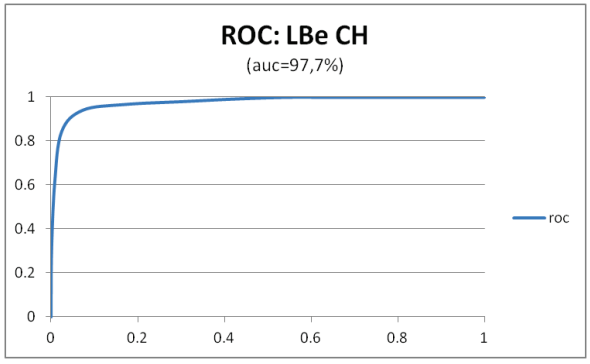

EPSg CH
(eer=4,0%, umbral=0,76)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

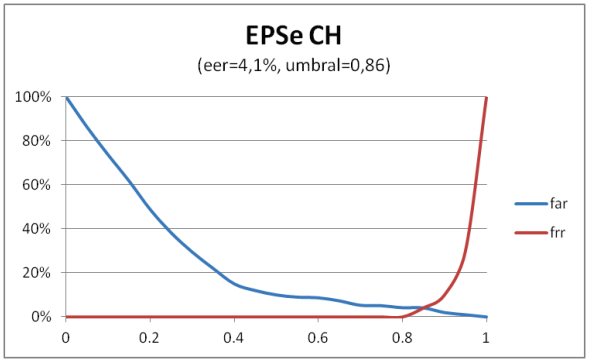

ROC: EPSg CH
(auc=99,2%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

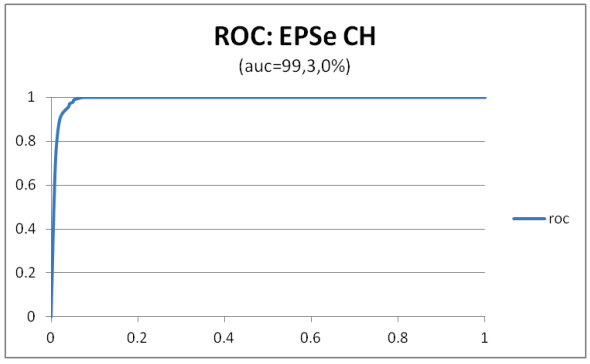

SMSg CH
(eer=4,5%, umbral=0,84)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

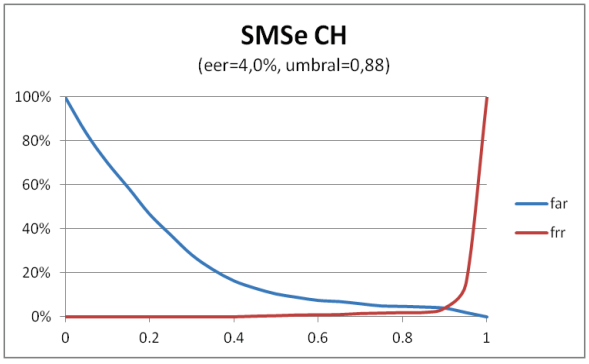

ROC: SMSg CH
(auc=98,8%)
1
0.8
0.6
0.4
0.2
0
0.00%
20.00%
40.00%
60.00%
80.00%
100.00%
roc

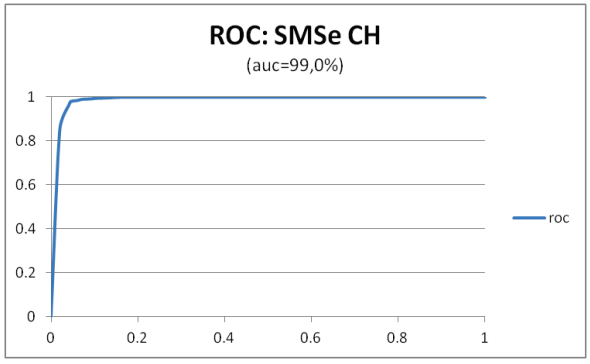

LBe CH
(eer=6,5%, umbral=0,76)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr


ROC: LBe CH
(auc=97,7%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

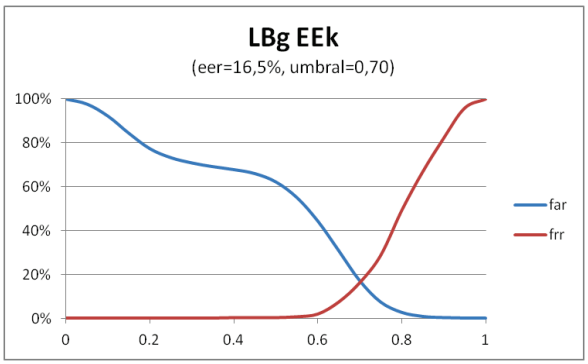

EPSe CH
(eer=4,1%, umbral=0,86)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

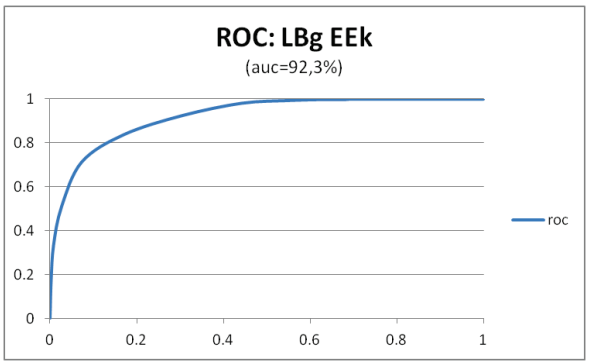

ROC: EPSe CH
(auc=99,3,0%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

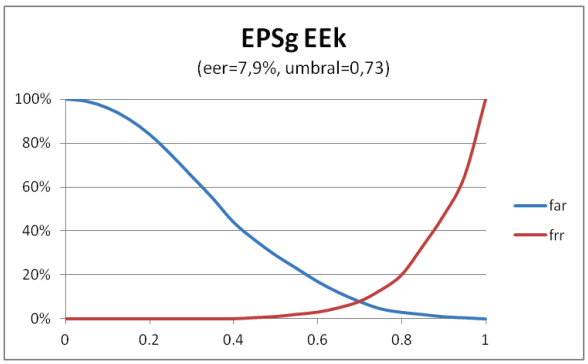

SMSe CH
(eer=4,0%, umbral=0,88)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

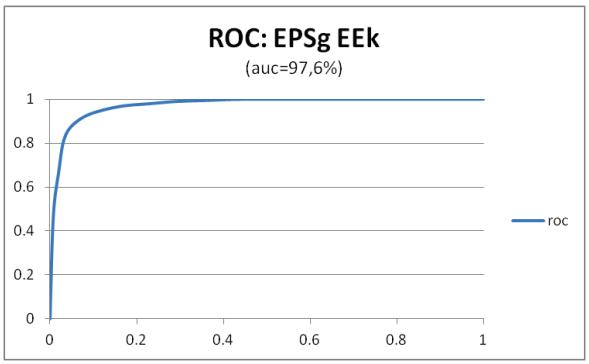

ROC: SMSe CH
(auc=99,0%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

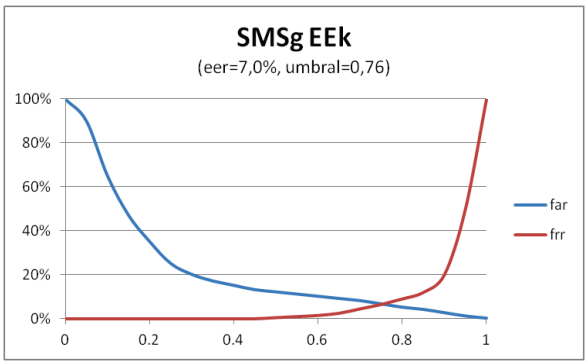
e. Usuario EEk

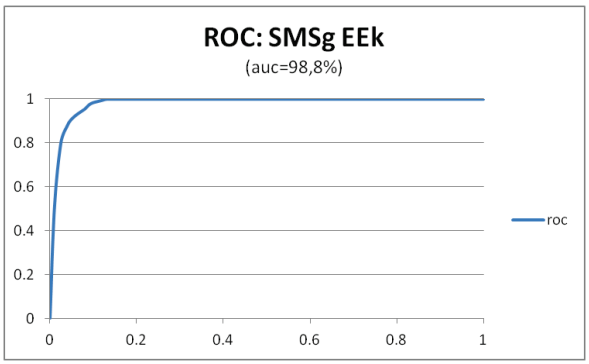


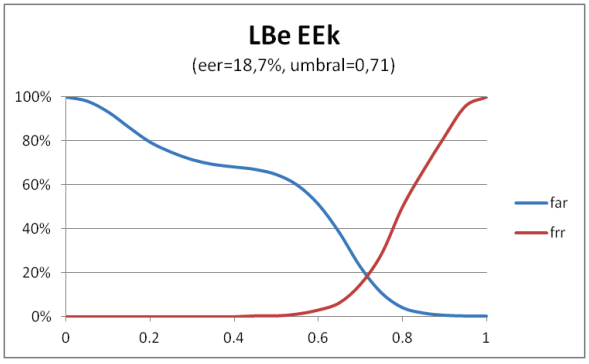


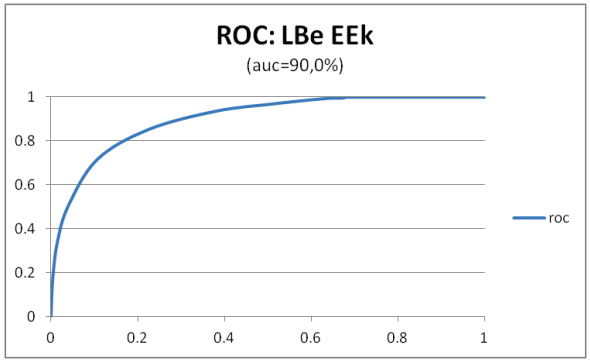


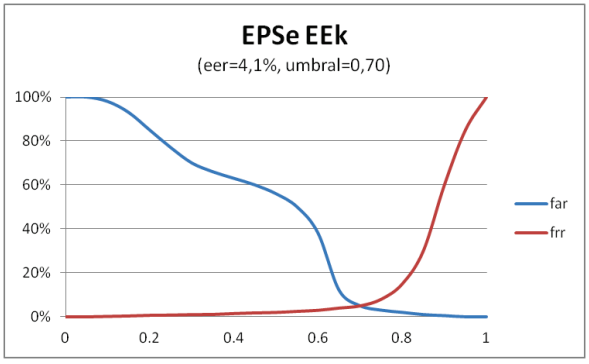


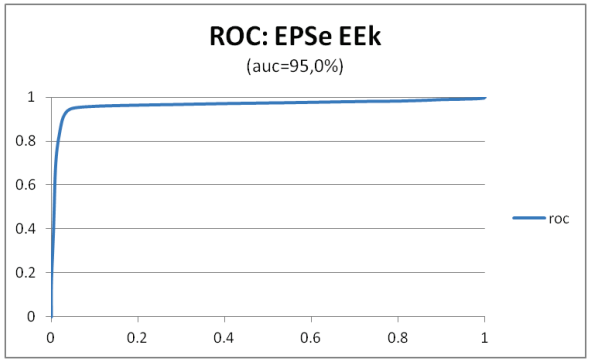


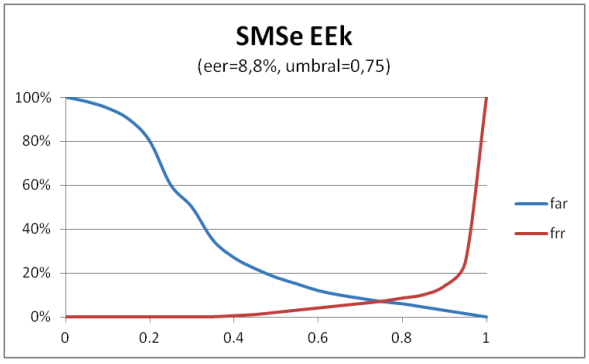


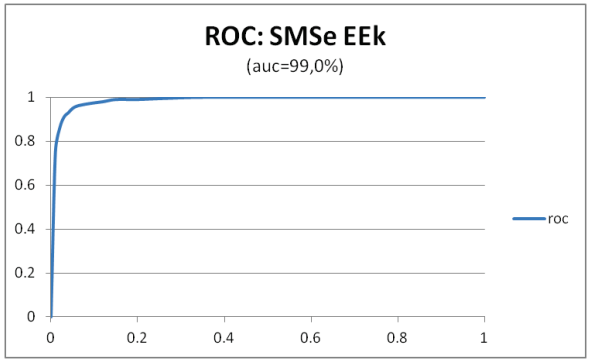

LBe EEk
(eer=18,7%, umbral=0,71)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr


ROC: LBe EEk
(auc=90,0%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

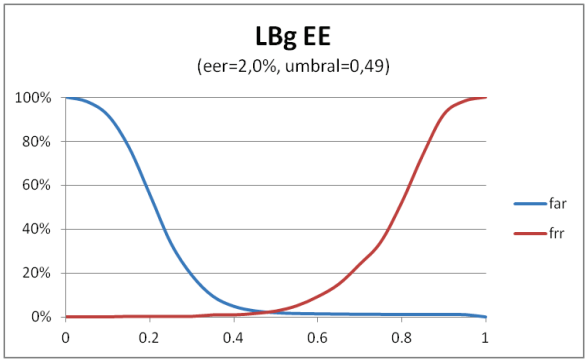

EPSe EEk
(eer=4,1%, umbral=0,70)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

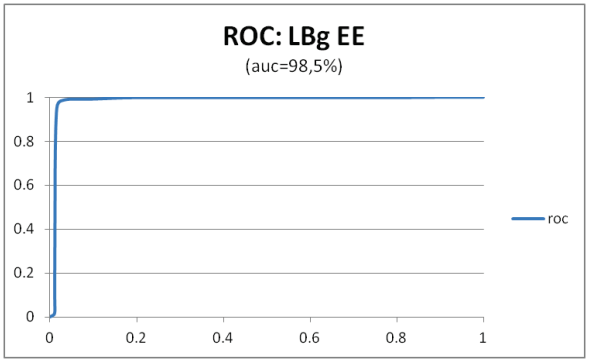

ROC: EPSe EEk
(auc=95,0%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

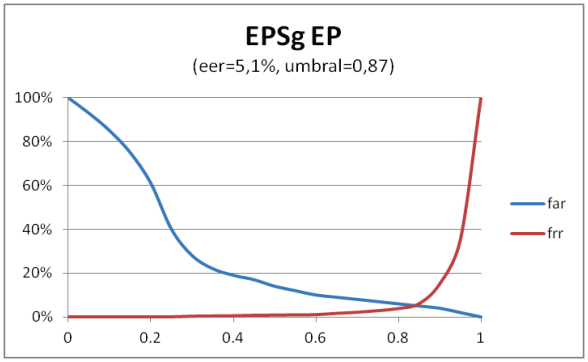

SMSe EEk
(eer=8,8%, umbral=0,75)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

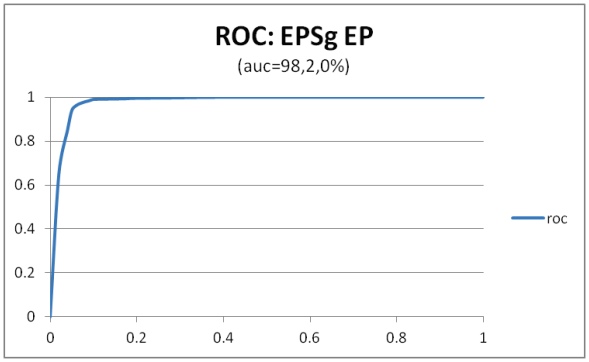

ROC: SMSe EEk
(auc=99,0%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

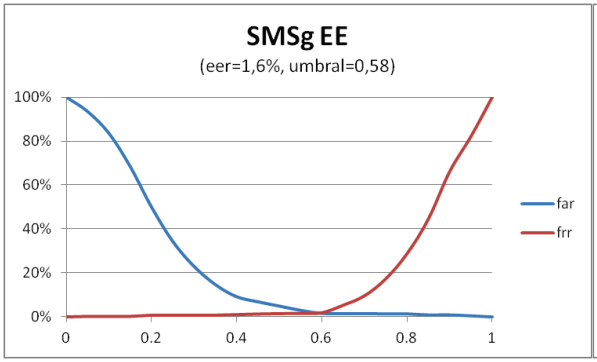
f. Usuario EE

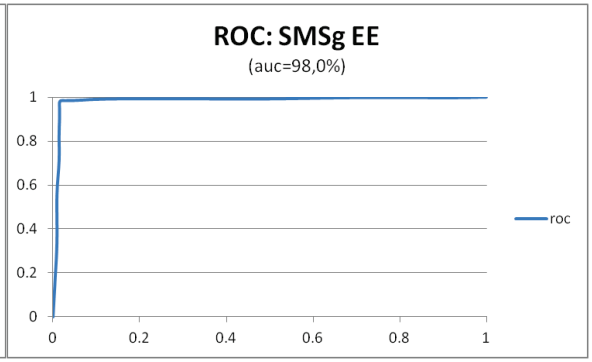

LBg EE
(eer=2,0%, umbral=0,49)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

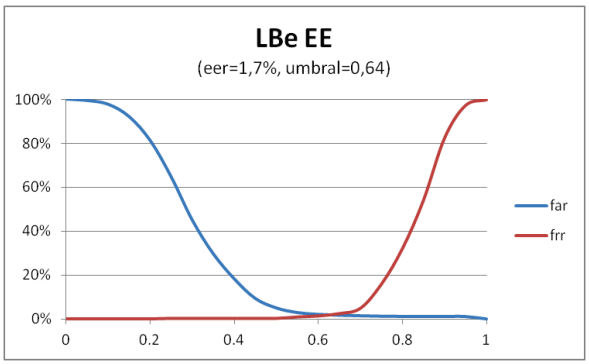

ROC: LBg EE
(auc=98,5%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

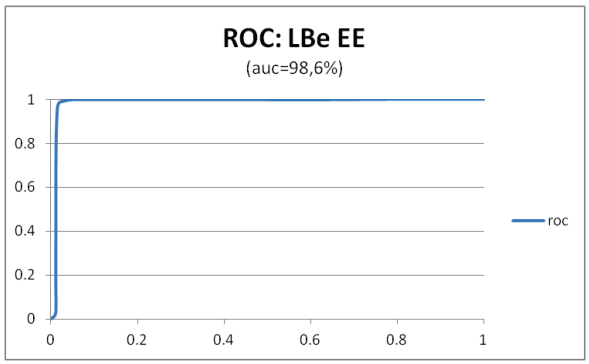

EPSg EP
(eer=5,1%, umbral=0,87)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

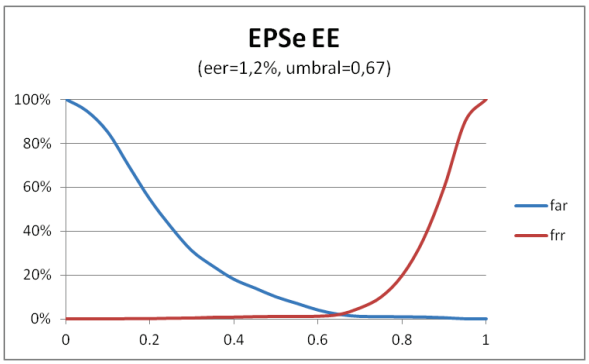

ROC: EPSg EP
(auc=98,2,0%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

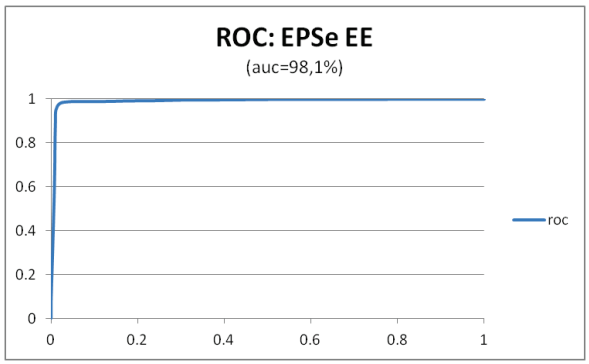

SMSg EE
(eer=1,6%, umbral=0,58)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

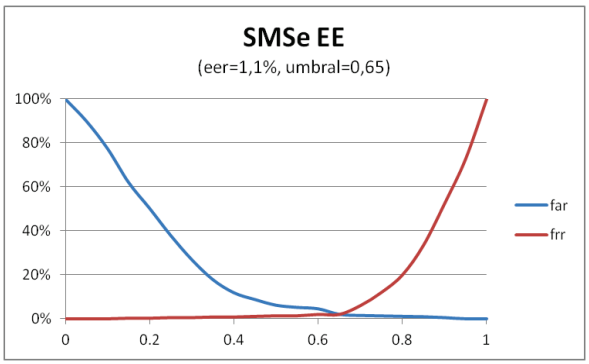

ROC: SMSg EE
(auc=98,0%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

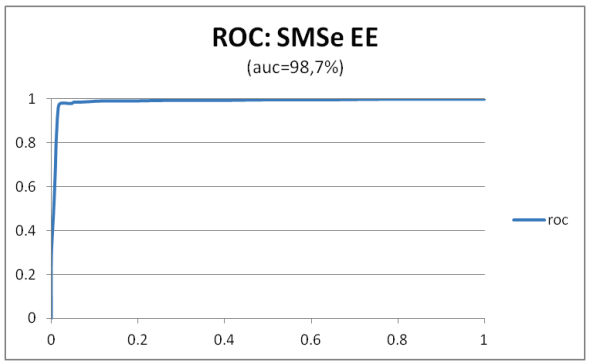

LBe EE
(eer=1,7%, umbral=0,64)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr


ROC: LBe EE
(auc=98,6%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

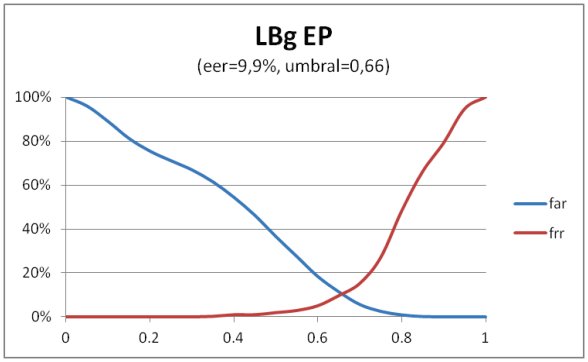

EPSe EE
(eer=1,2%, umbral=0,67)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

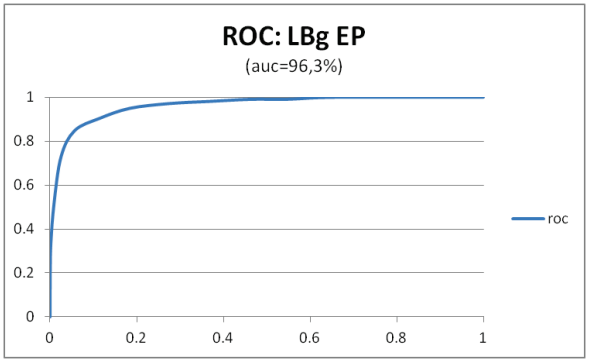

ROC: EPSe EE
(auc=98,1%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

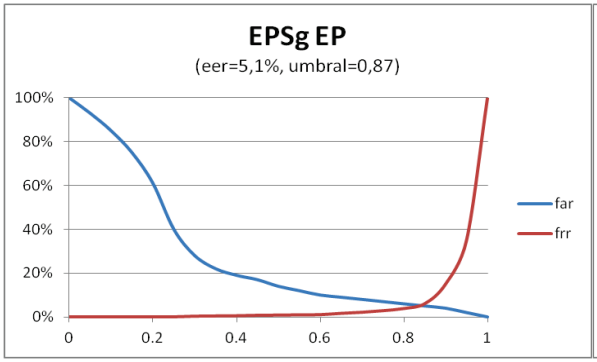

SMSe EE
(eer=1,1%, umbral=0,65)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

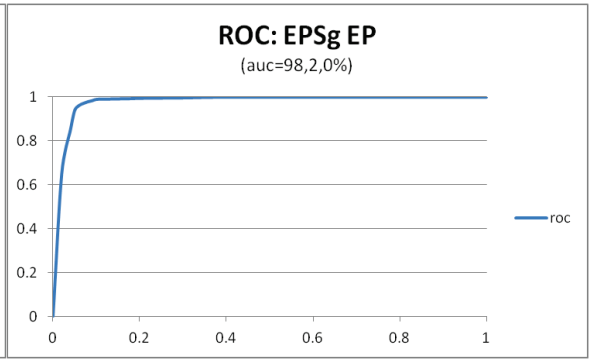

ROC: SMSe EE
(auc=98,7%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

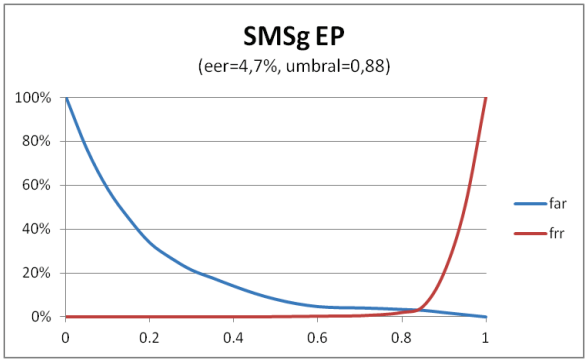
g. Usuario EP

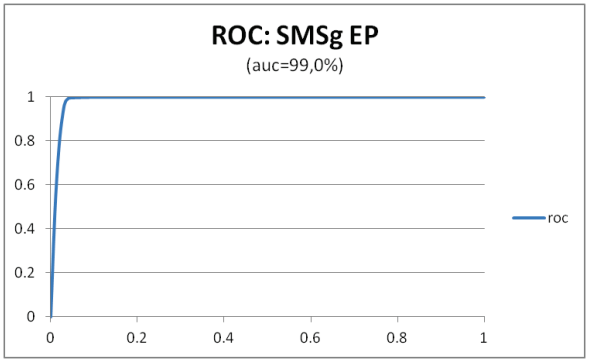

LBg EP
(eer=9,9%, umbral=0,66)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

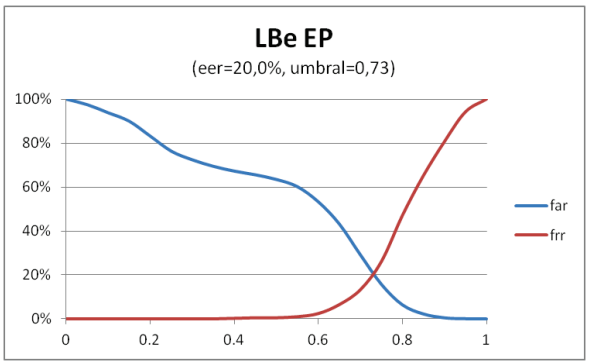

ROC: LBg EP
(auc=96,3%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

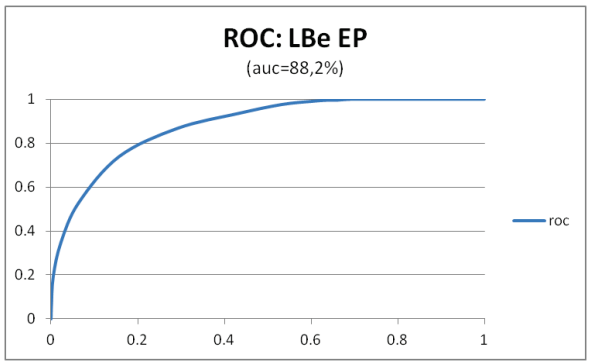

EPSg EP
(eer=5,1%, umbral=0,87)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

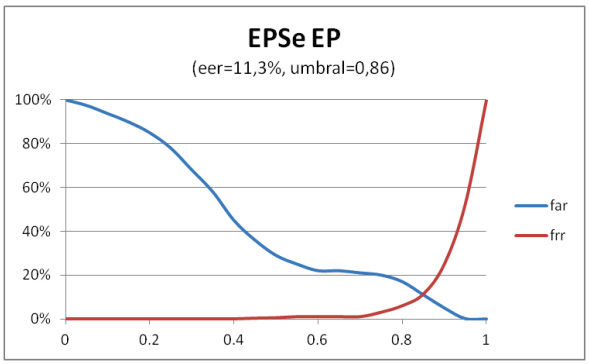

ROC: EPSg EP
(auc=98,2,0%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

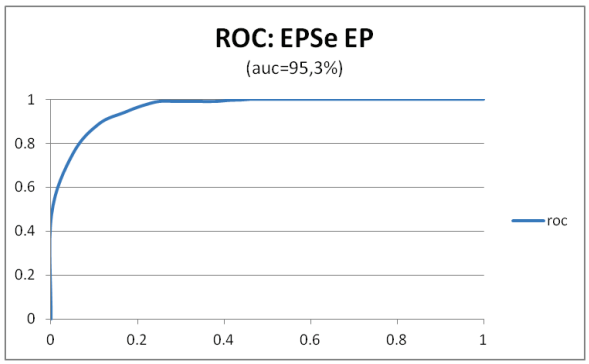

SMSg EP
(eer=4,7%, umbral=0,88)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

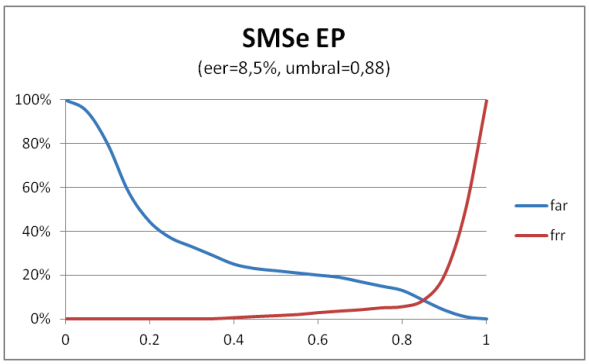

ROC: SMSg EP
(auc=99,0%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

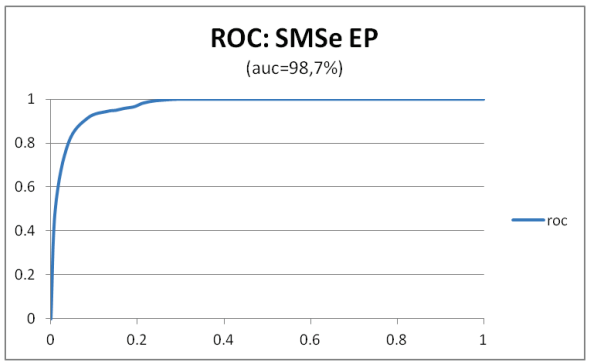

LBe EP
(eer=20,0%, umbral=0,73)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr


ROC: LBe EP
(auc=88,2%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

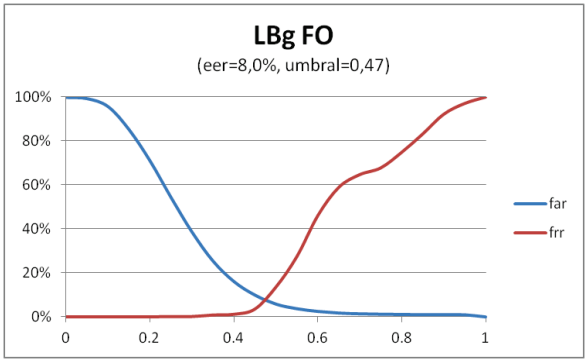

EPSe EP
(eer=11,3%, umbral=0,86)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

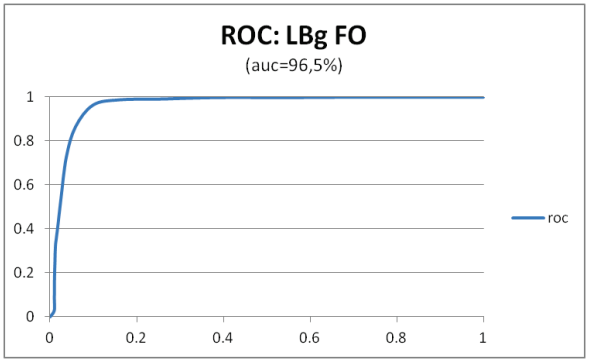

ROC: EPSe EP
(auc=95,3%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

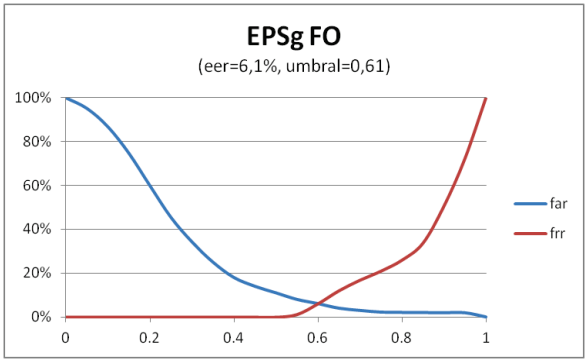

SMSe EP
(eer=8,5%, umbral=0,88)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

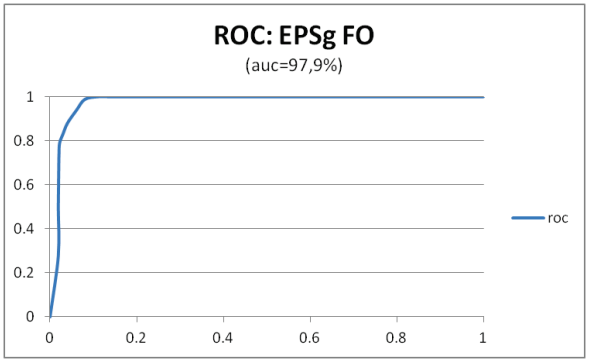

ROC: SMSe EP
(auc=98,7%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

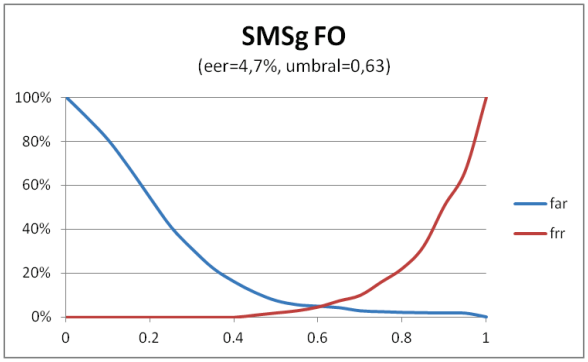
h. Usuario FO

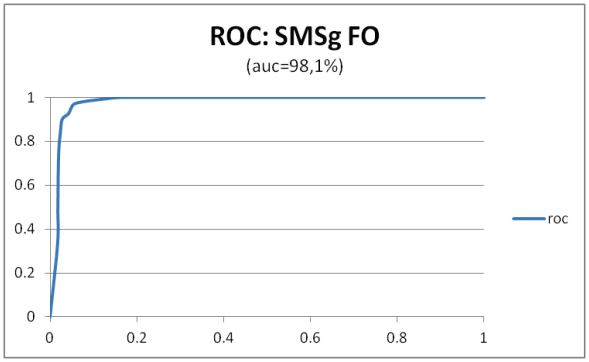

LBg FO
(eer=8,0%, umbral=0,47)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

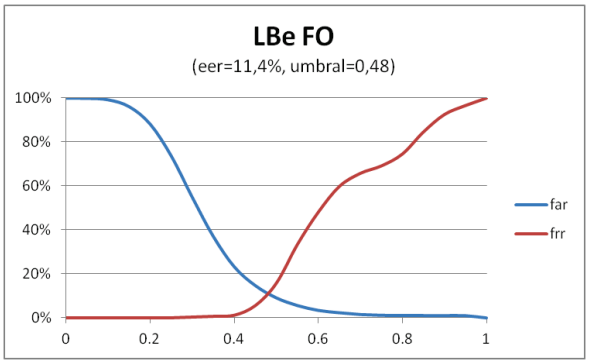

ROC: LBg FO
(auc=96,5%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

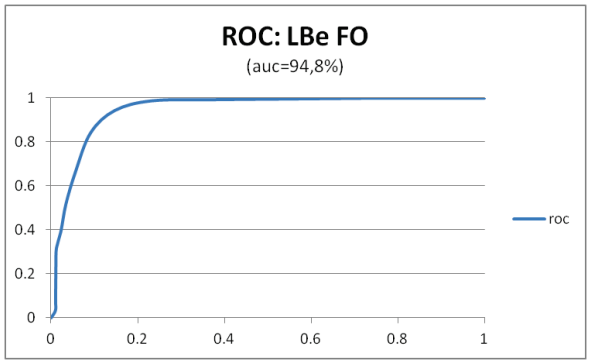

EPSg FO
(eer=6,1%, umbral=0,61)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

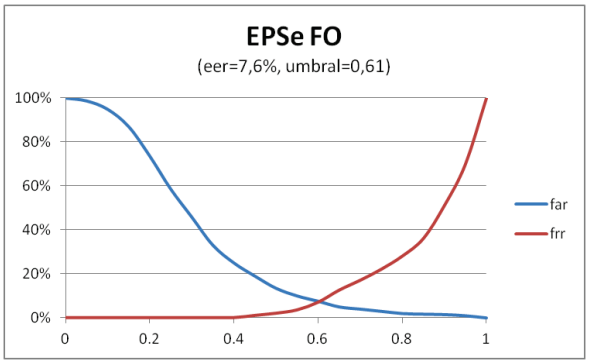

ROC: EPSg FO
(auc=97,9%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

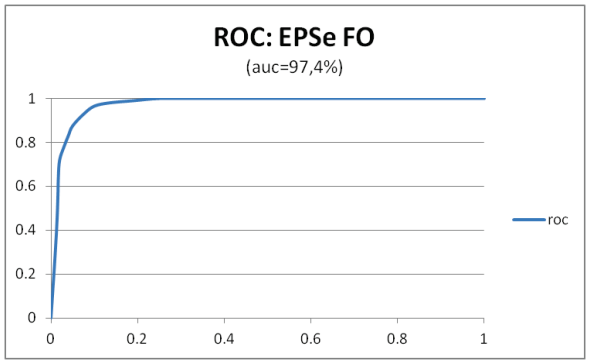

SMSg FO
(eer=4,7%, umbral=0,63)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

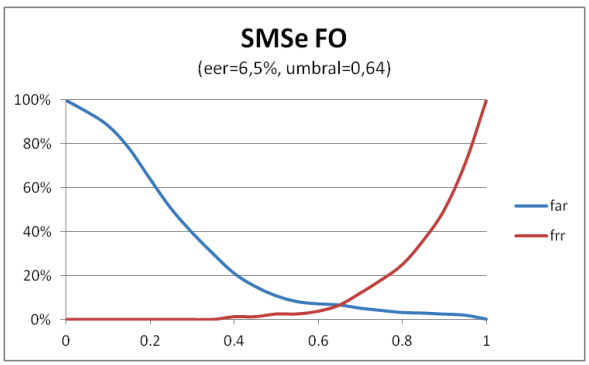

ROC: SMSg FO
(auc=98,1%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

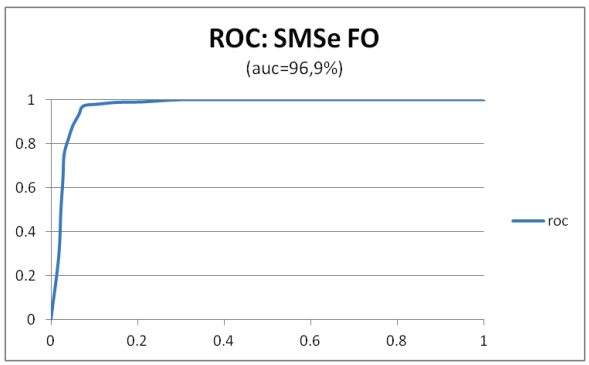

LBe FO
(eer=11,4%, umbral=0,48)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr


ROC: LBe FO
(auc=94,8%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

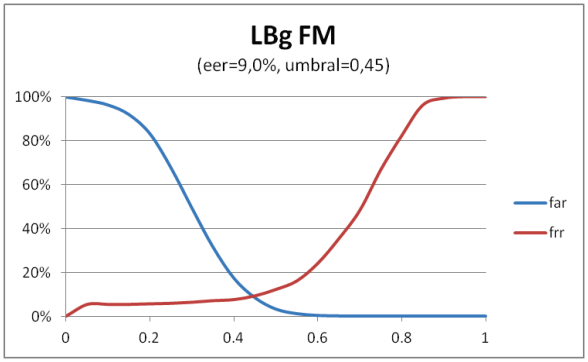

EPSe FO
(eer=7,6%, umbral=0,61)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

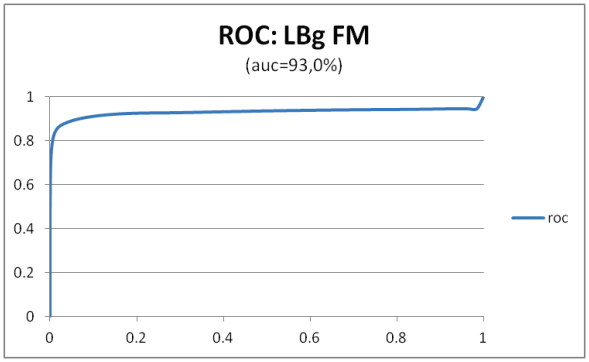

ROC: EPSe FO
(auc=97,4%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

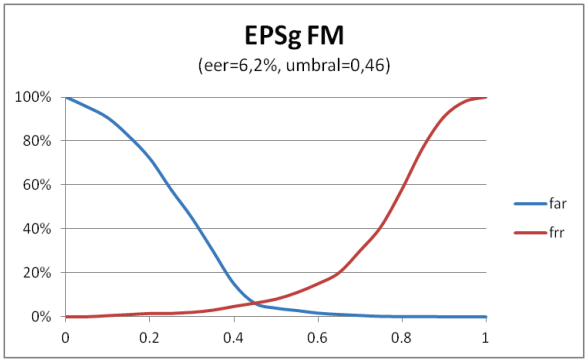

SMSe FO
(eer=6,5%, umbral=0,64)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

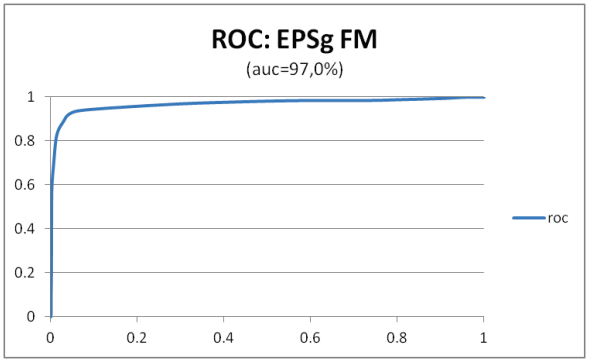

ROC: SMSe FO
(auc=96,9%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

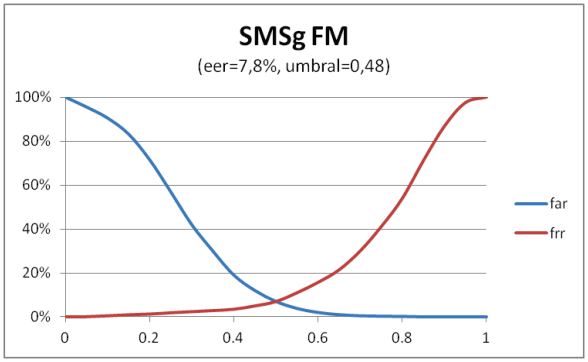
i. Usuario FM

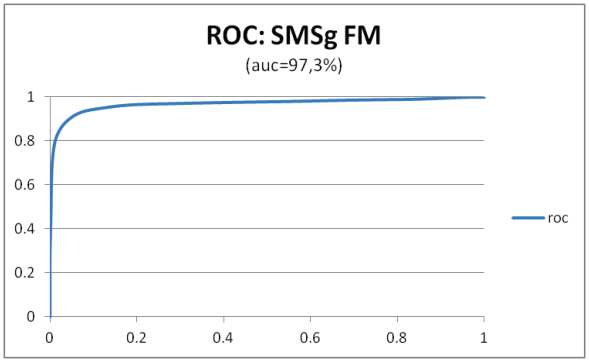

LBg FM
(eer=9,0%, umbral=0,45)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

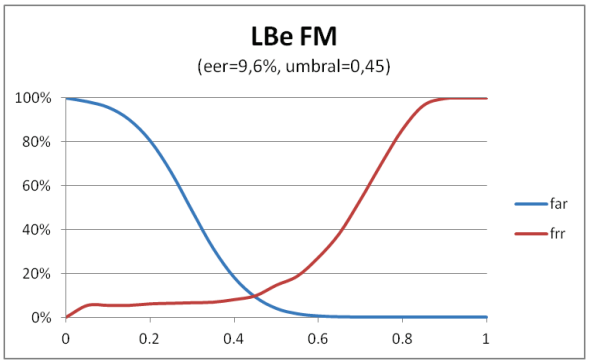

ROC: LBg FM
(auc=93,0%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

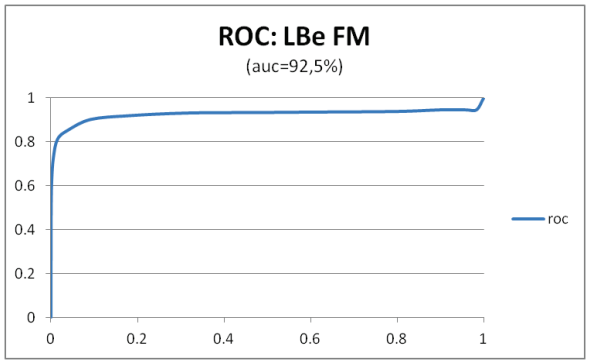

EPSg FM
(eer=6,2%, umbral=0,46)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

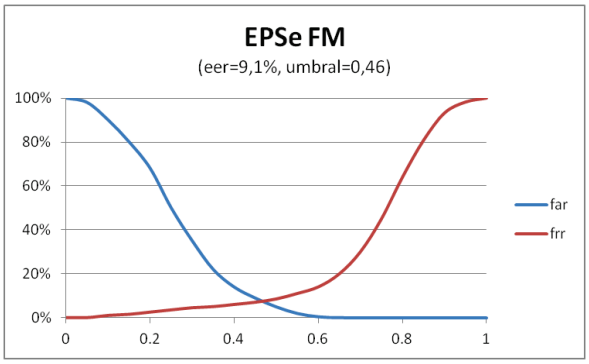

ROC: EPSg FM
(auc=97,0%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

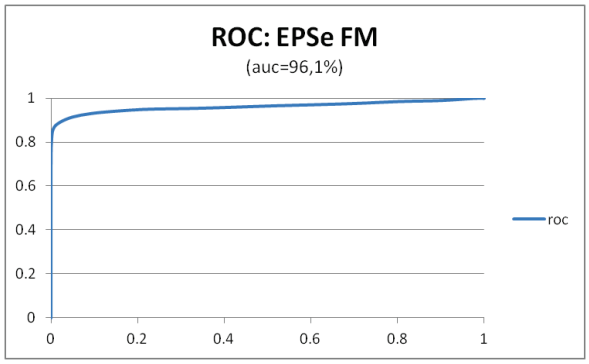

SMSg FM
(eer=7,8%, umbral=0,48)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

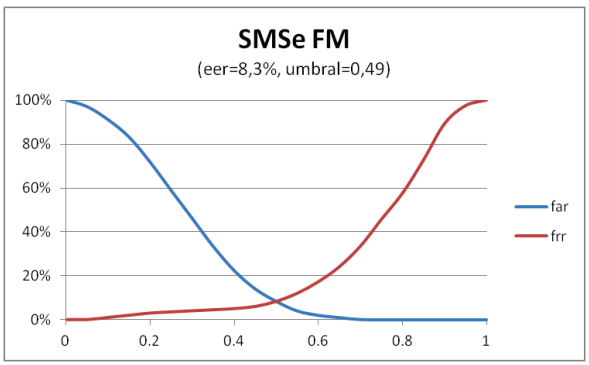

ROC: SMSg FM
(auc=97,3%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

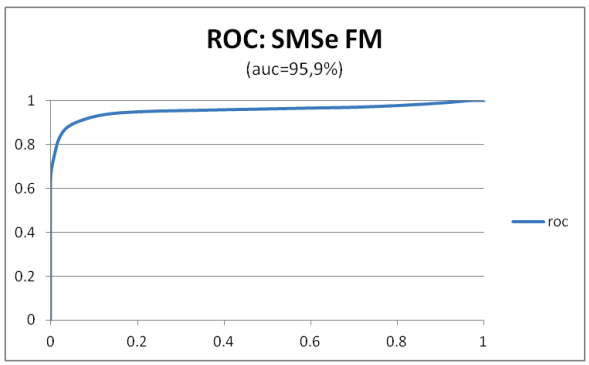

LBe FM
(eer=9,6%, umbral=0,45)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr


ROC: LBe FM
(auc=92,5%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

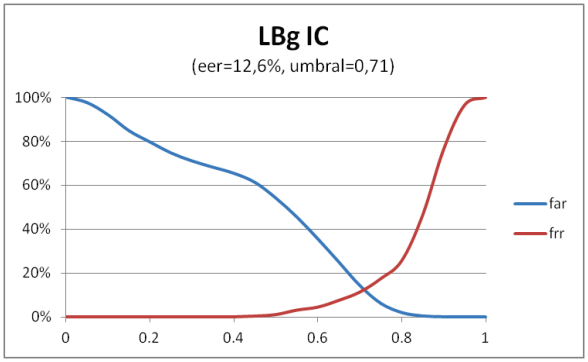

EPSe FM
(eer=9,1%, umbral=0,46)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

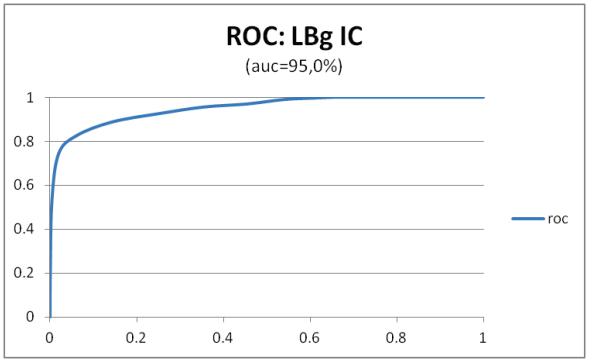

ROC: EPSe FM
(auc=96,1%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

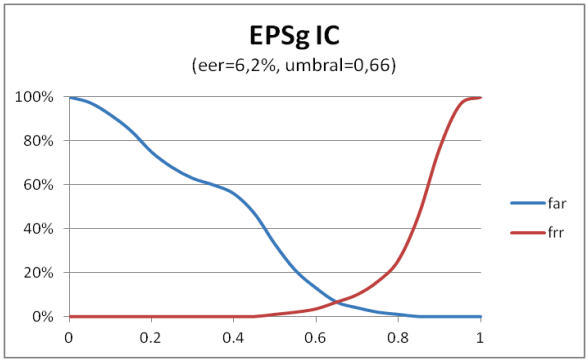

SMSe FM
(eer=8,3%, umbral=0,49)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

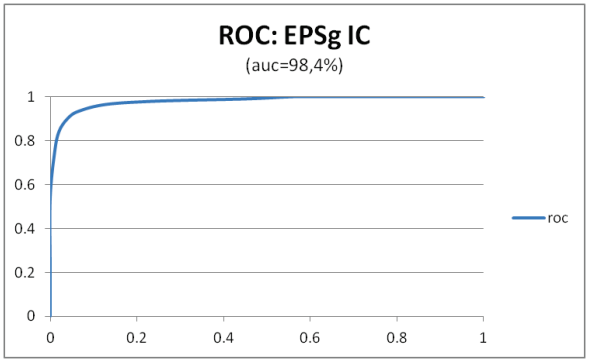

ROC: SMSe FM
(auc=95,9%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

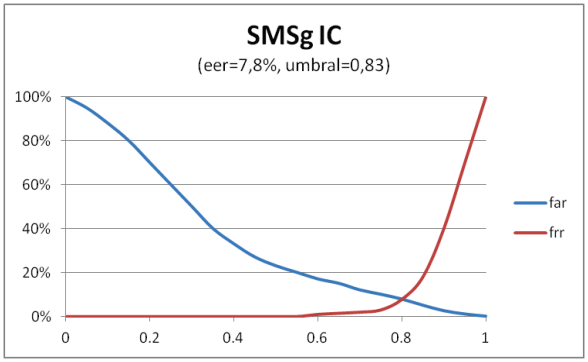
j. Usuario IC

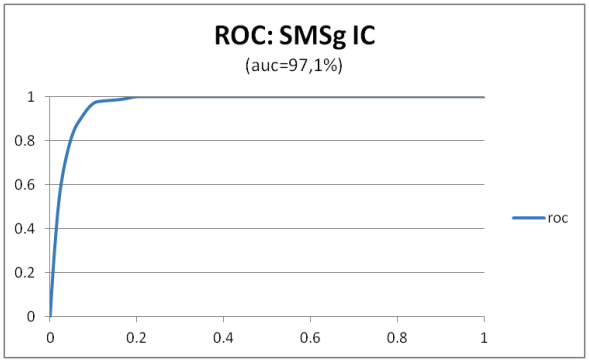

LBg IC
(eer=12,6%, umbral=0,71)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

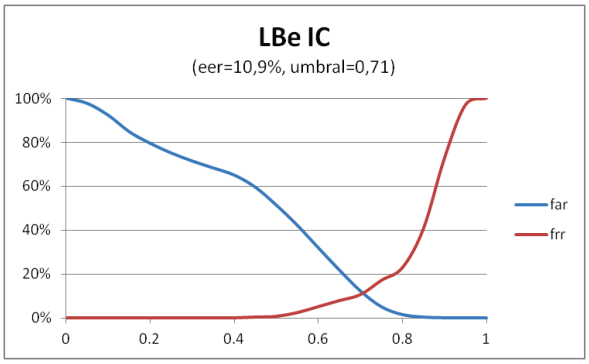

ROC: LBg IC
(auc=95,0%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

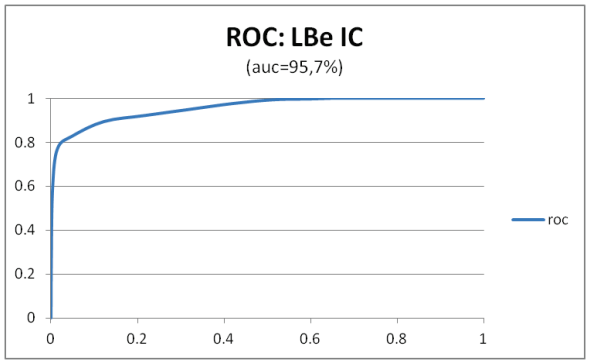

EPSg IC
(eer=6,2%, umbral=0,66)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

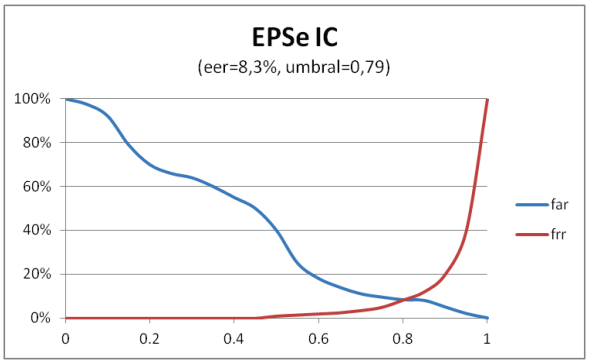

ROC: EPSg IC
(auc=98,4%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

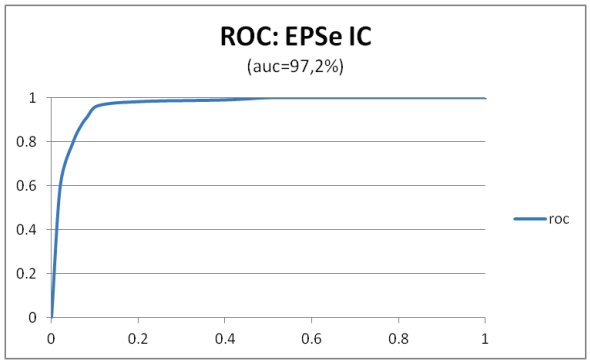

SMSg IC
(eer=7,8%, umbral=0,83)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

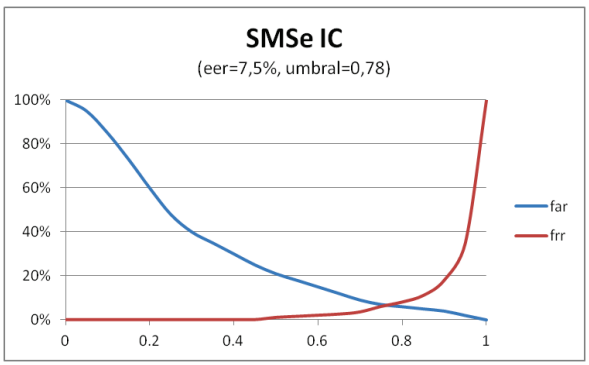

ROC: SMSg IC
(auc=97,1%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

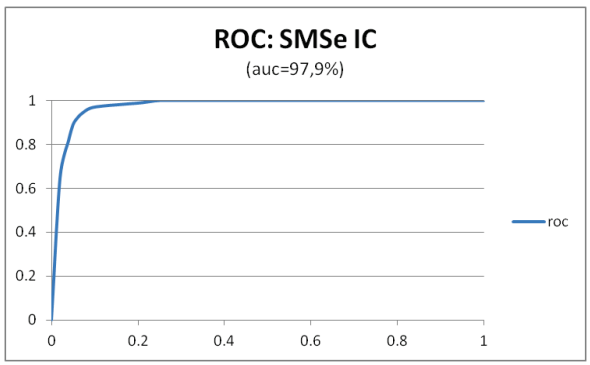

LBe IC
(eer=10,9%, umbral=0,71)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr


ROC: LBe IC
(auc=95,7%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

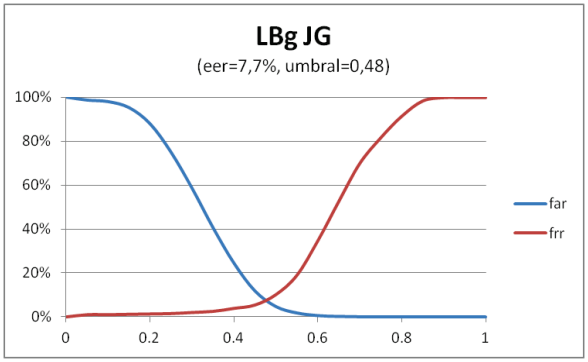

EPSe IC
(eer=8,3%, umbral=0,79)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

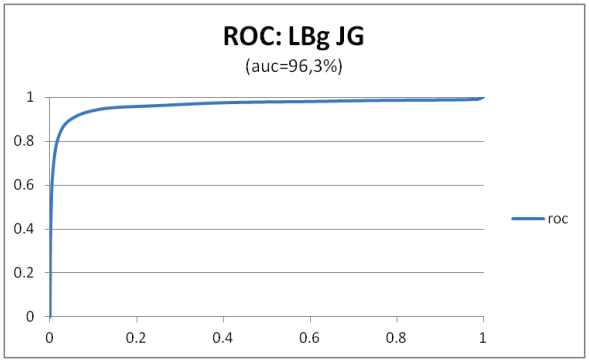

ROC: EPSe IC
(auc=97,2%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

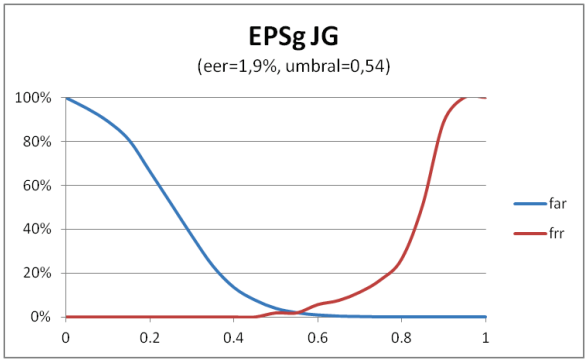

SMSe IC
(eer=7,5%, umbral=0,78)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

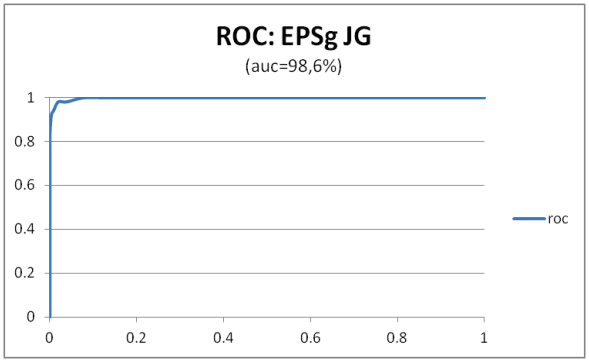

ROC: SMSe IC
(auc=97,9%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

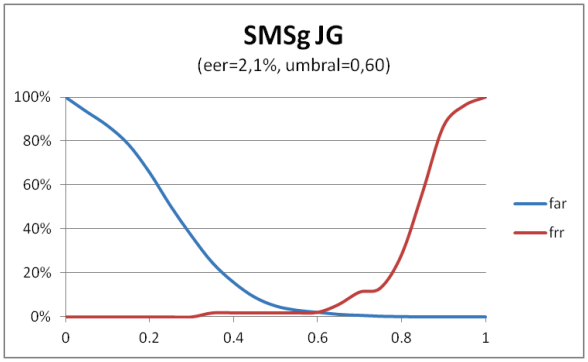
k. Usuario JG

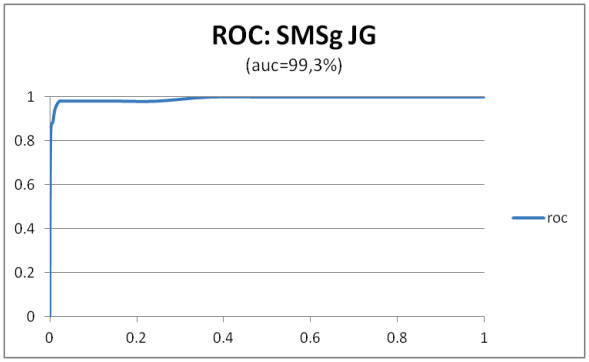

LBg JG
(eer=7,7%, umbral=0,48)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

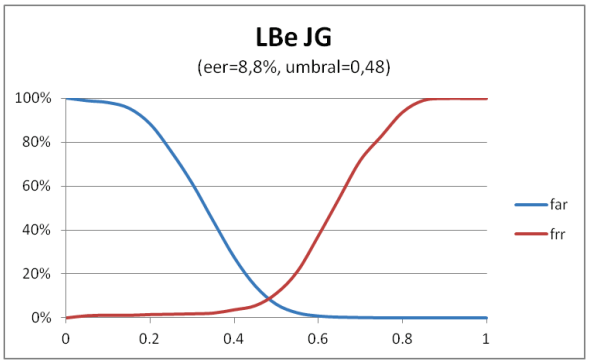

ROC: LBg JG
(auc=96,3%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

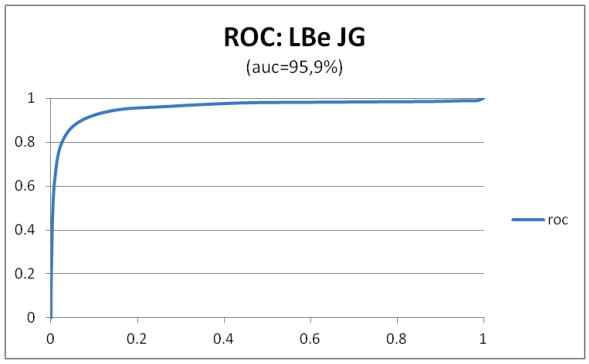

EPSg JG
(eer=1,9%, umbral=0,54)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

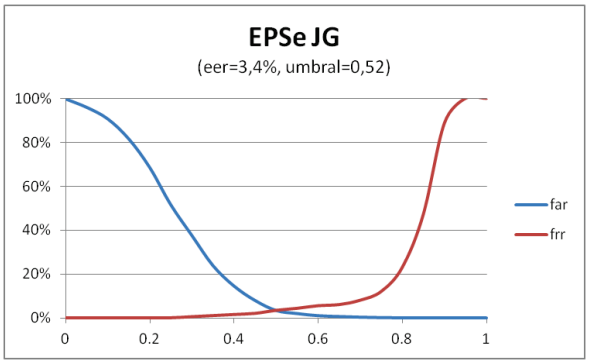

ROC: EPSg JG
(auc=98,6%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

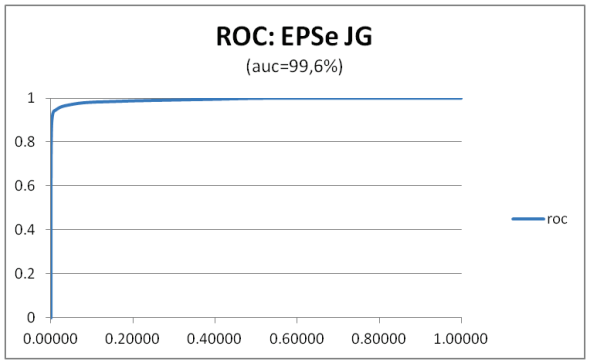

SMSg JG
(eer=2,1%, umbral=0,60)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

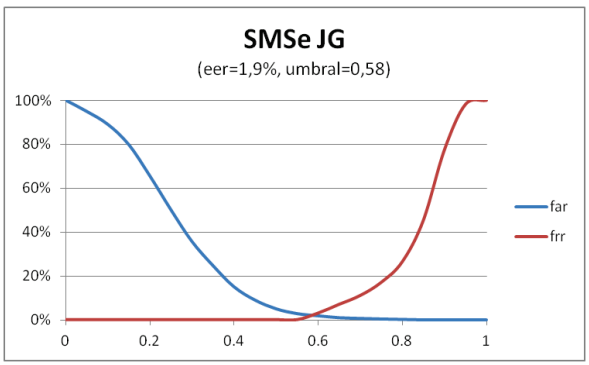

ROC: SMSg JG
(auc=99,3%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

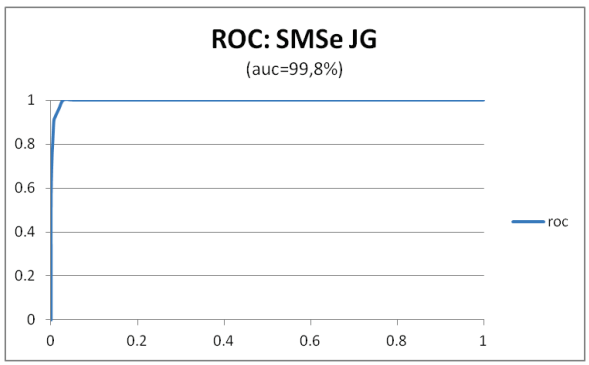

LBe JG
(eer=8,8%, umbral=0,48)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr


ROC: LBe JG
(auc=95,9%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

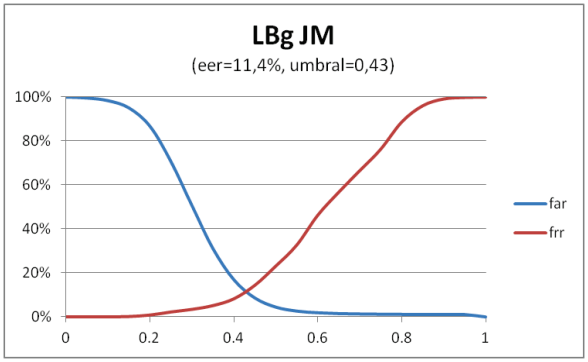

EPSe JG
(eer=3,4%, umbral=0,52)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

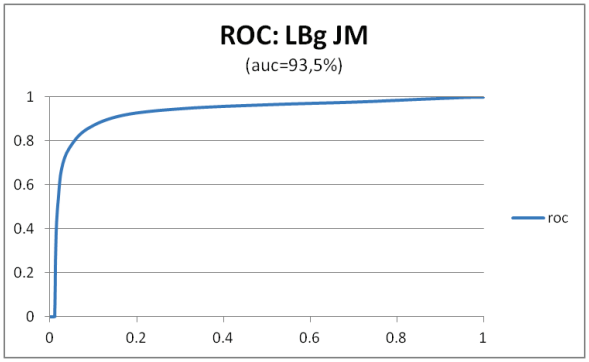

ROC: EPSe JG
(auc=99,6%)
1
0.8
0.6
0.4
0.2
0
0.00000
0.20000
0.40000
0.60000
0.80000
1.00000
roc

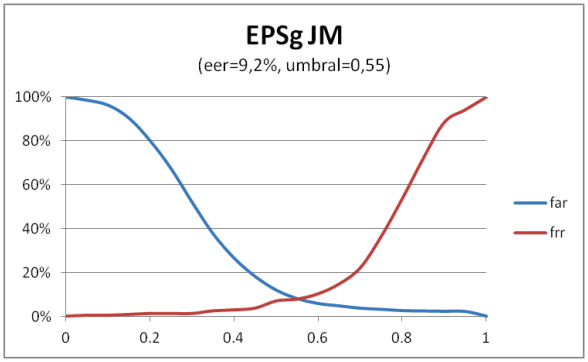

SMSe JG
(eer=1,9%, umbral=0,58)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

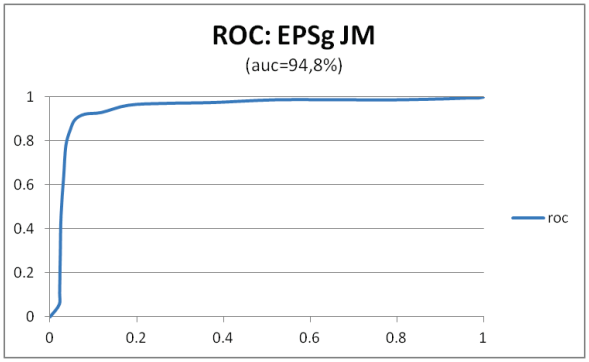

ROC: SMSe JG
(auc=99,8%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

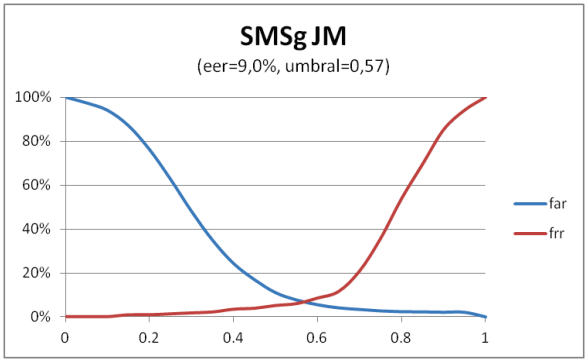
I. Usuario JM

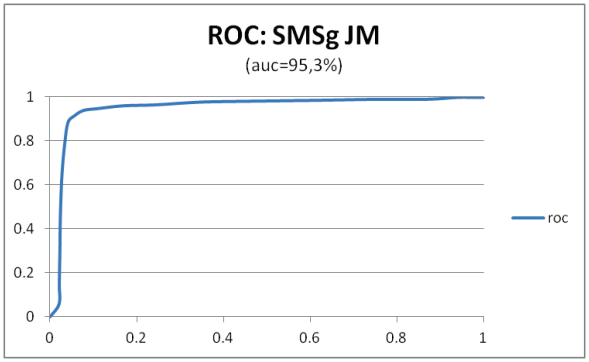

LBg JM
(eer=11,4%, umbral=0,43)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

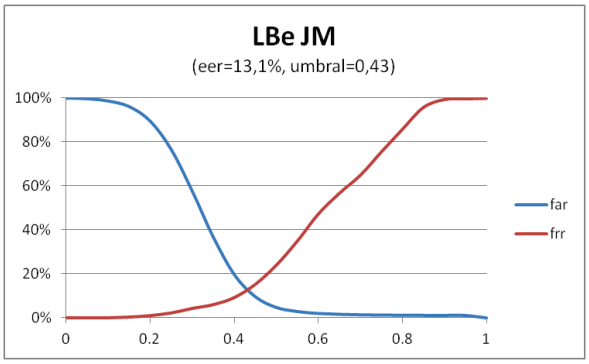

ROC: LBg JM
(auc=93,5%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

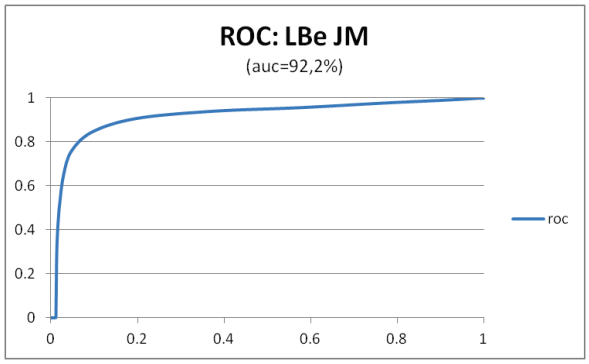

EPSg JM
(eer=9,2%, umbral=0,55)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

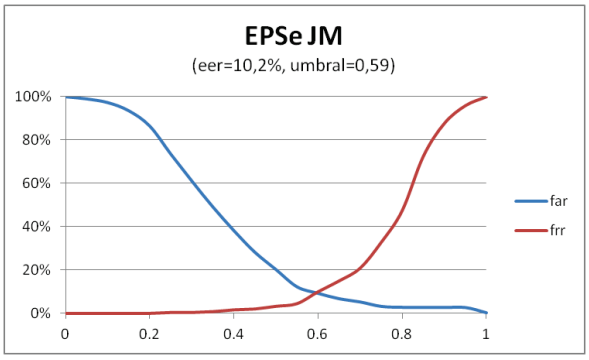

ROC: EPSg JM
(auc=94,8%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

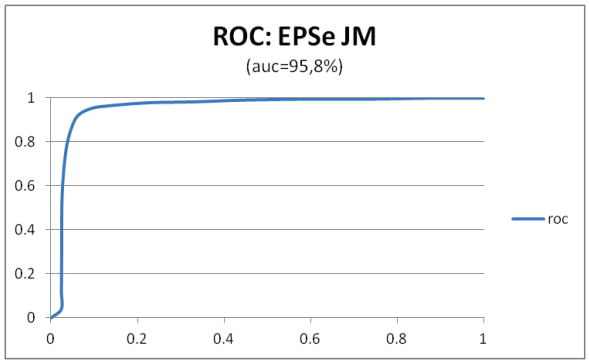

SMSg JM
(eer=9,0%, umbral=0,57)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

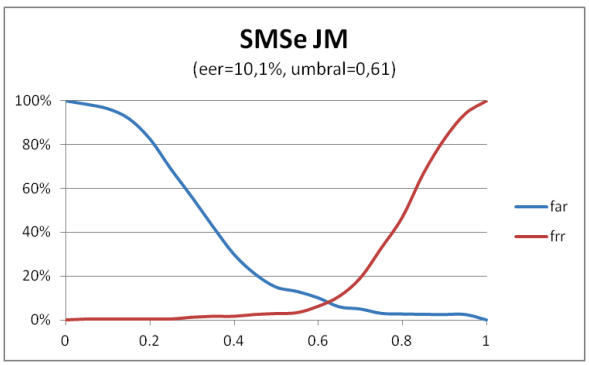

ROC: SMSg JM
(auc=95,3%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

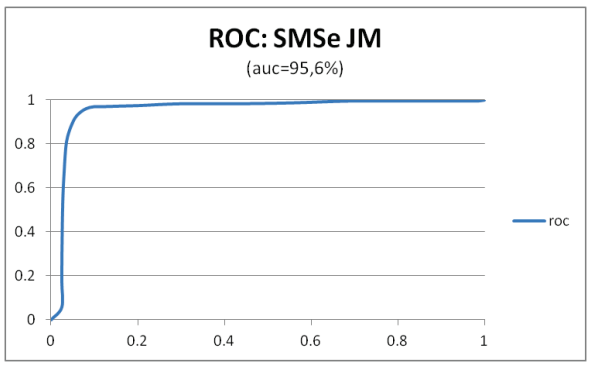

LBe JM
(eer=13,1%, umbral=0,43)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr


ROC: LBe JM
(auc=92,2%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

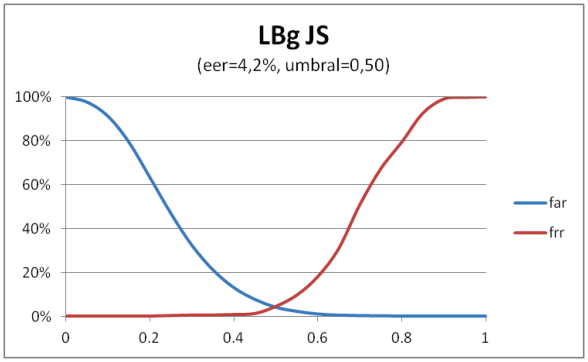

EPSe JM
(eer=10,2%, umbral=0,59)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

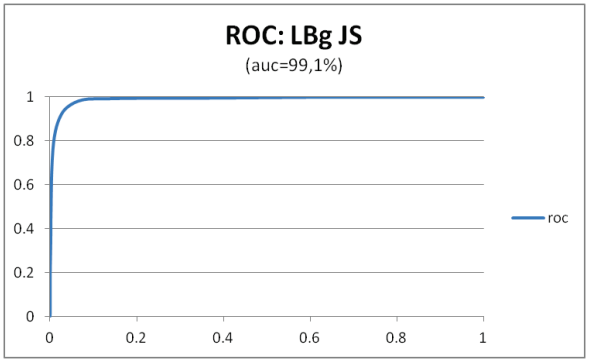

ROC: EPSe JM
(auc=95,8%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

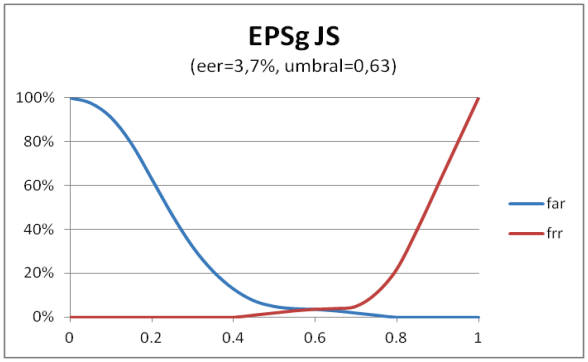

SMSe JM
(eer=10,1%, umbral=0,61)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

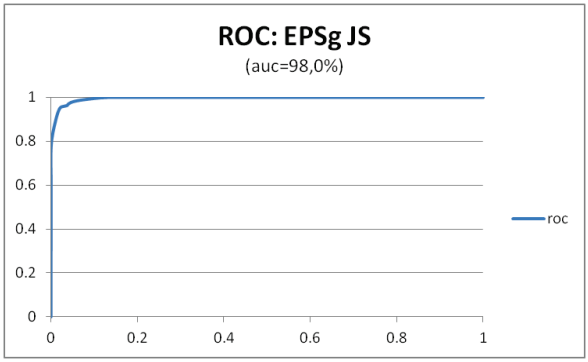

ROC: SMSe JM
(auc=95,6%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

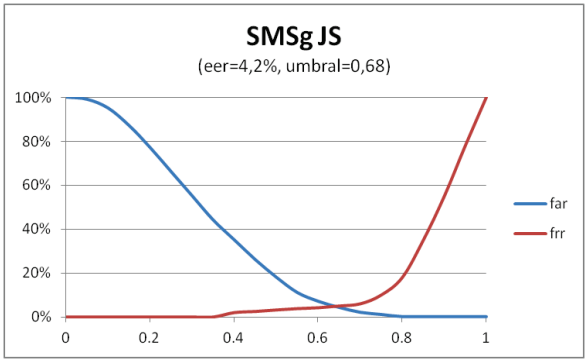
m. Usuario JS

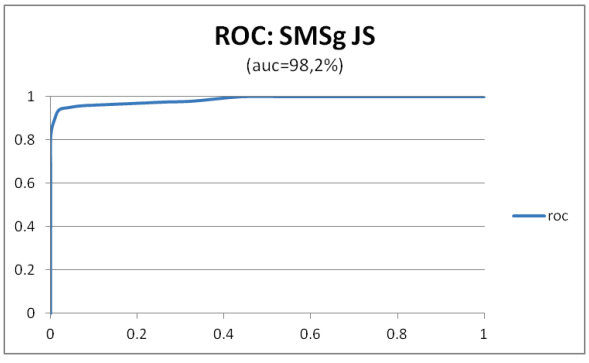

LBg JS
(eer=4,2%, umbral=0,50)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

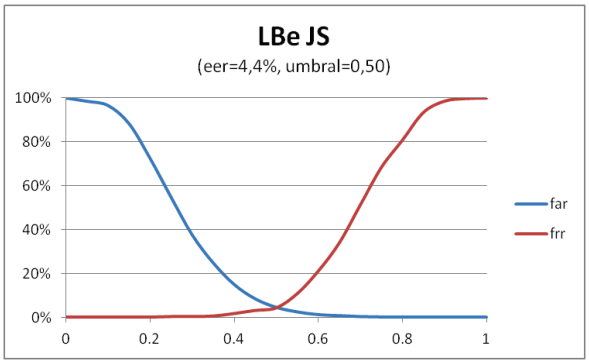

ROC: LBg JS
(auc=99,1%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

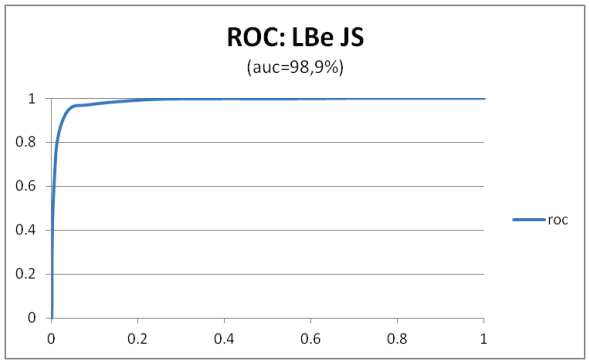

EPSg JS
(eer=3,7%, umbral=0,63)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

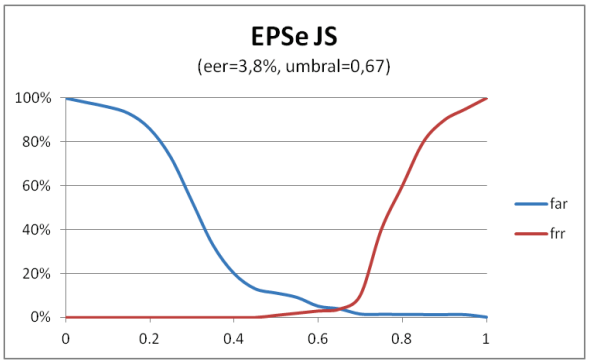

ROC: EPSg JS
(auc=98,0%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

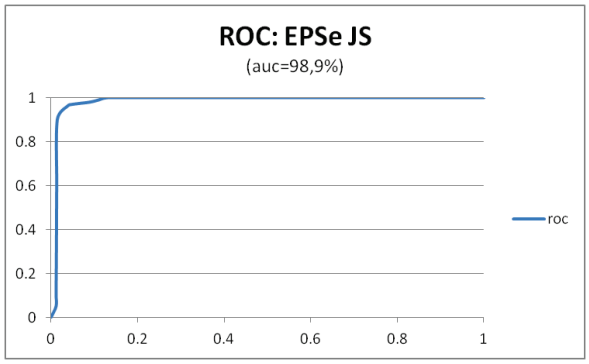

SMSg JS
(eer=4,2%, umbral=0,68)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

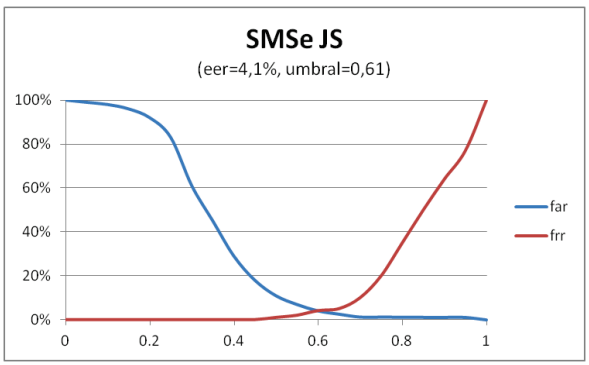

ROC: SMSg JS
(auc=98,2%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

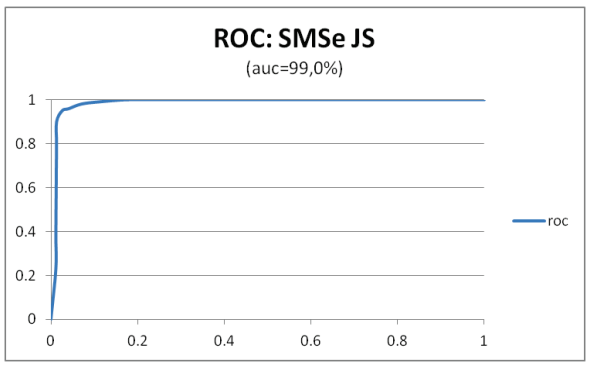

LBe JS
(eer=4,4%, umbral=0,50)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr


ROC: LBe JS
(auc=98,9%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

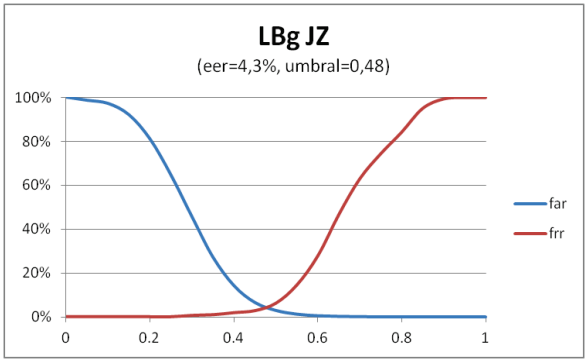

EPSe JS
(eer=3,8%, umbral=0,67)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

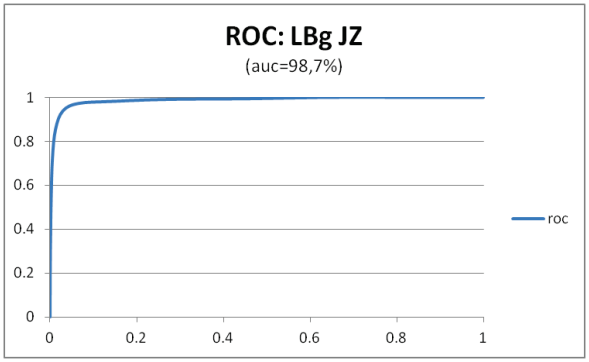

ROC: EPSe JS
(auc=98,9%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

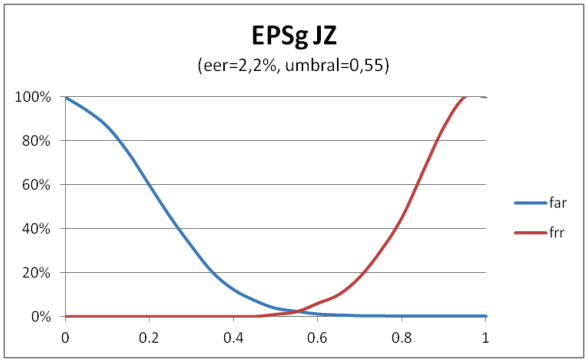

SMSe JS
(eer=4,1%, umbral=0,61)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

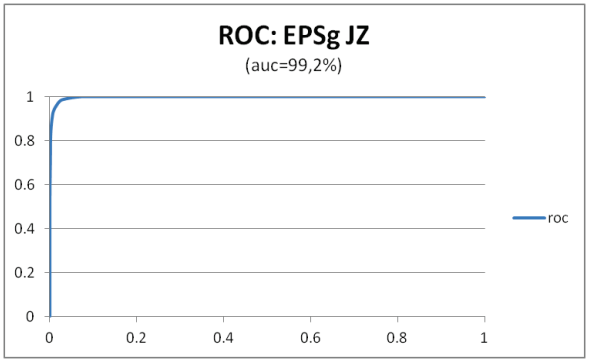

ROC: SMSe JS
(auc=99,0%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

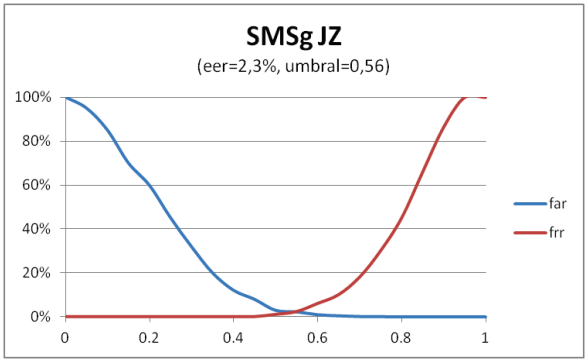
n. Usuario JZ

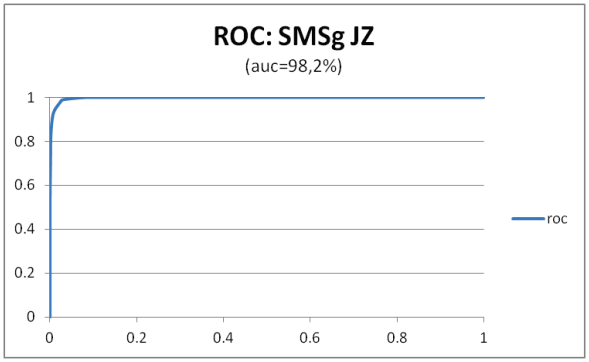

LBg JZ
(eer=4,3%, umbral=0,48)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

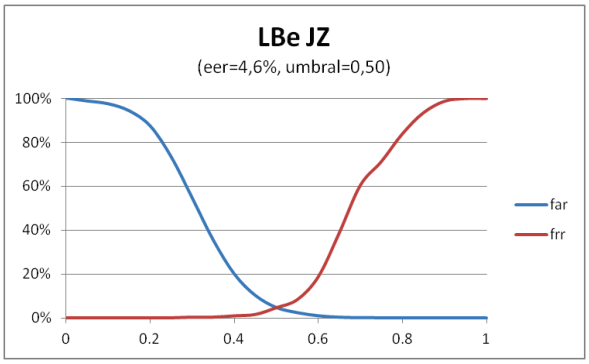

ROC: LBg JZ
(auc=98,7%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

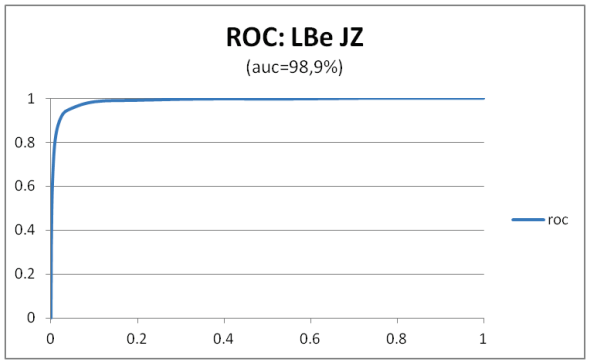

EPSg JZ
(eer=2,2%, umbral=0,55)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

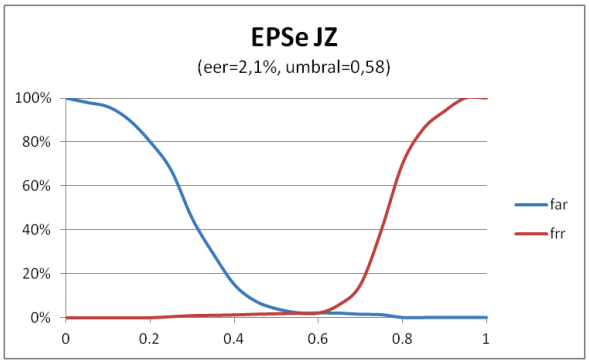

ROC: EPSg JZ
(auc=99,2%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

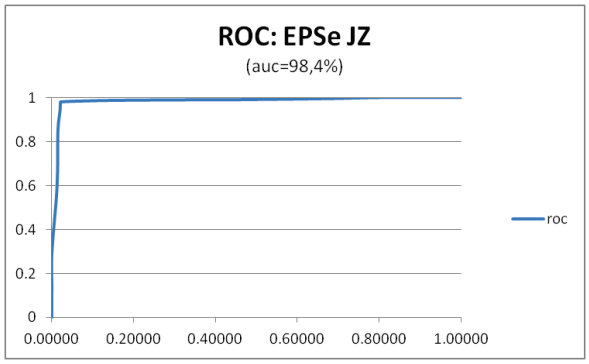

SMSg JZ
(eer=2,3%, umbral=0,56)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

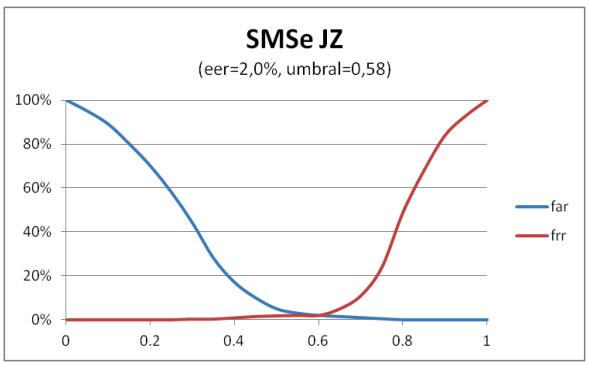

ROC: SMSg JZ
(auc=98,2%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

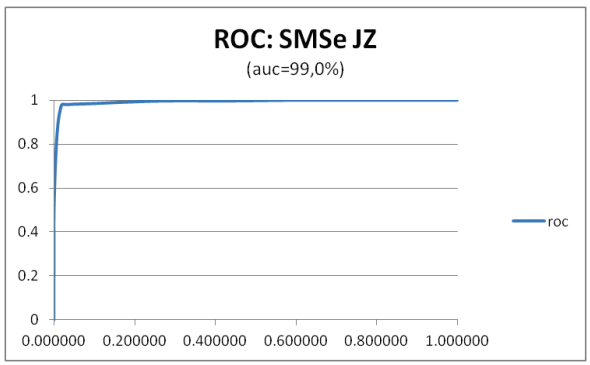

LBe JZ
(eer=4,6%, umbral=0,50)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr


ROC: LBe JZ
(auc=98,9%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

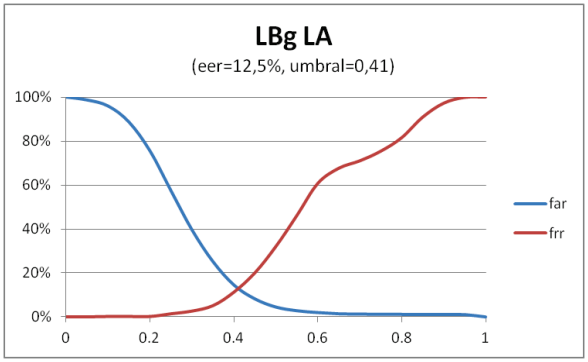

EPSe JZ
(eer=2,1%, umbral=0,58)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

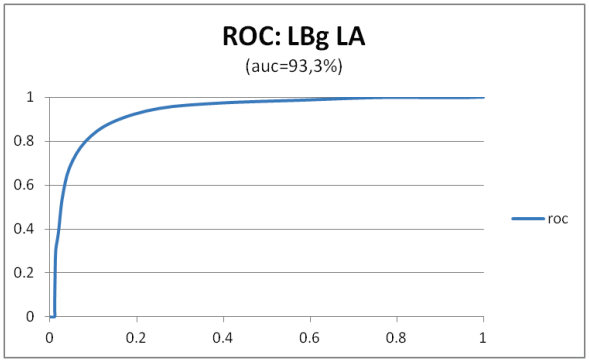

ROC: EPSe JZ
(auc=98,4%)
1
0.8
0.6
0.4
0.2
0
0.00000
0.20000
0.40000
0.60000
0.80000
1.00000
roc

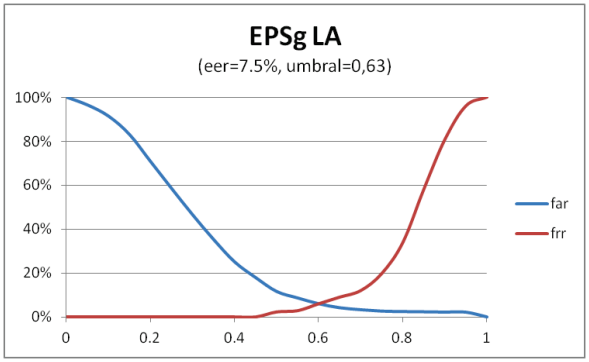

SMSe JZ
(eer=2,0%, umbral=0,58)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

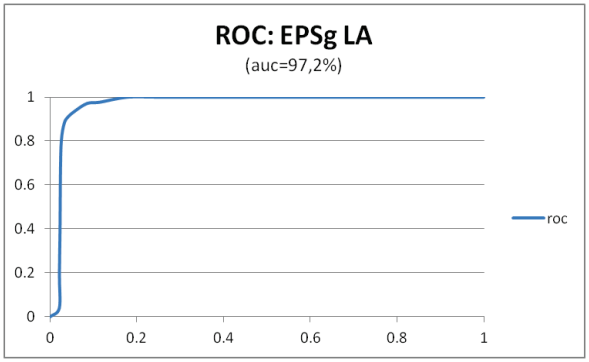

ROC: SMSe JZ
(auc=99,0%)
1
0.8
0.6
0.4
0.2
0
0.000000
0.200000
0.400000
0.600000
0.800000
1.000000
roc

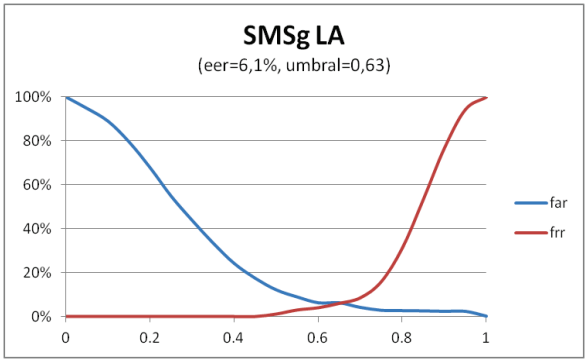
o. Usuario LA

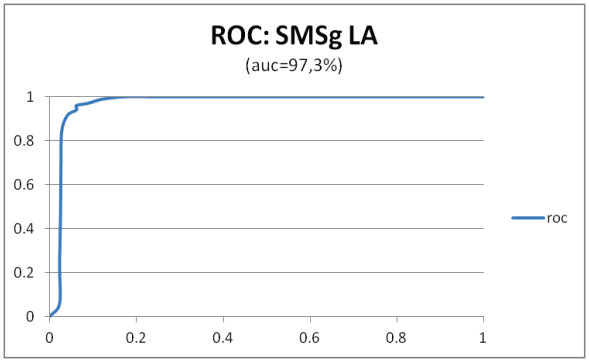

LBg LA
(eer=12,5%, umbral=0,41)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

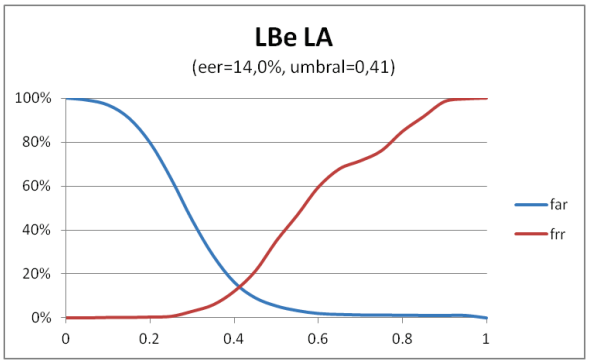

ROC: LBg LA
(auc=93,3%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

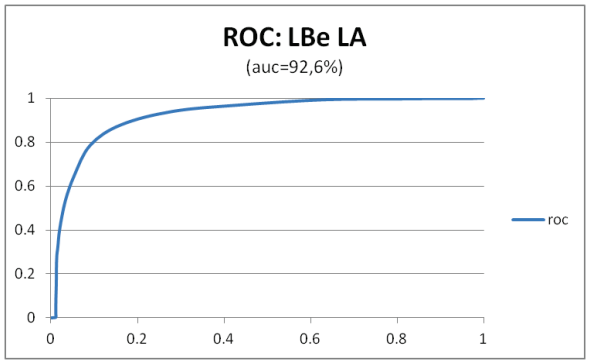

EPSg LA
(eer=7.5%, umbral=0,63)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

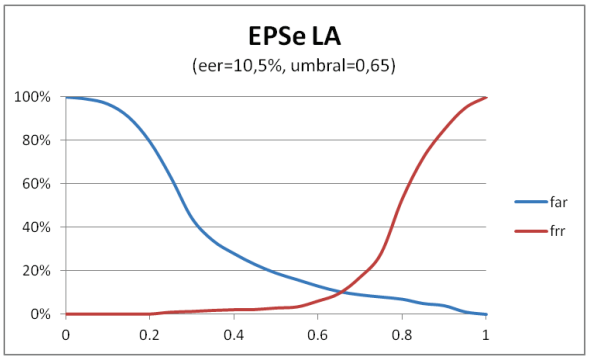

ROC: EPSg LA
(auc=97,2%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

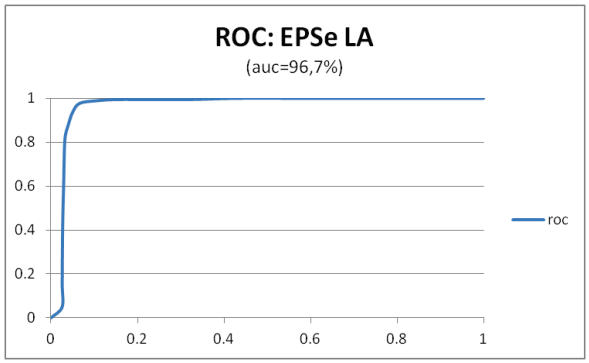

SMSg LA
(eer=6,1%, umbral=0,63)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

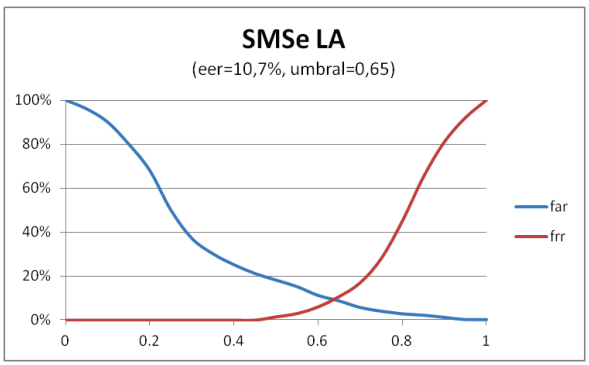

ROC: SMSg LA
(auc=97,3%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

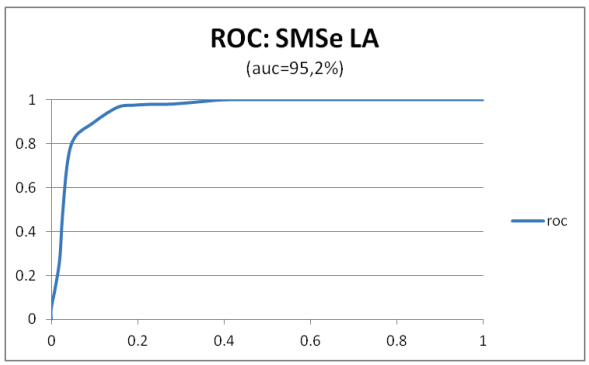

LBe LA
(eer=14,0%, umbral=0,41)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr


ROC: LBe LA
(auc=92,6%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

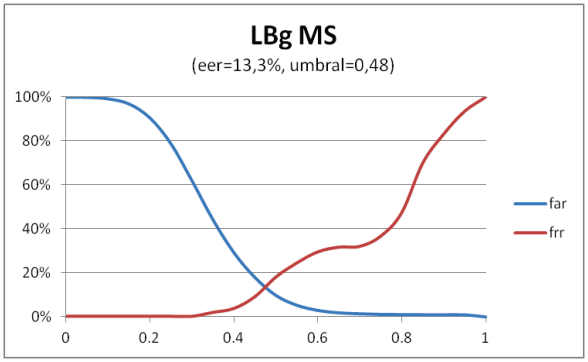

EPSe LA
(eer=10,5%, umbral=0,65)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

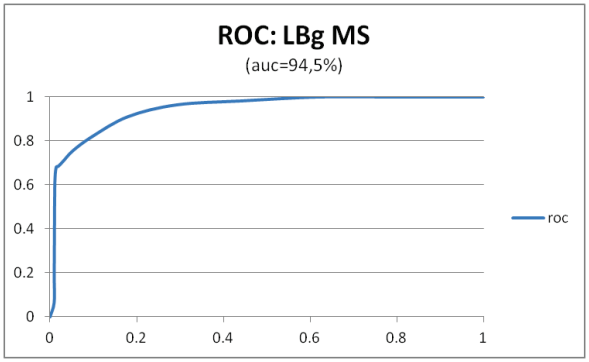

ROC: EPSe LA
(auc=96,7%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

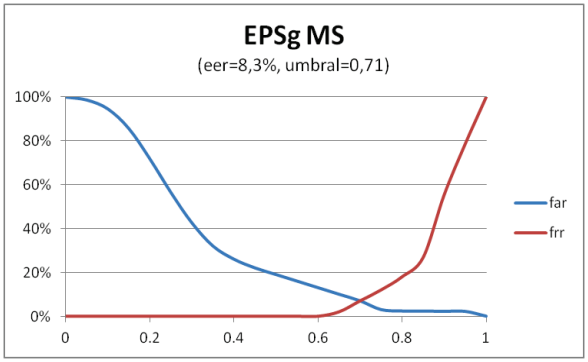

SMSe LA
(eer=10,7%, umbral=0,65)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

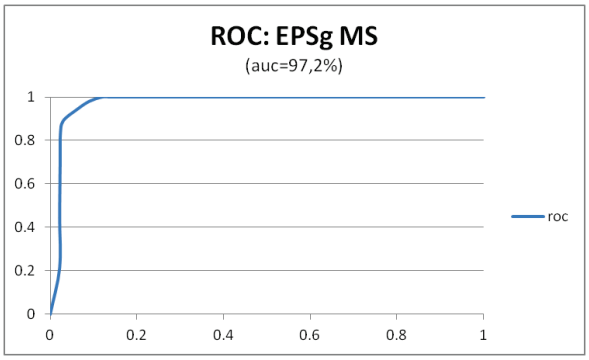

ROC: SMSe LA
(auc=95,2%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

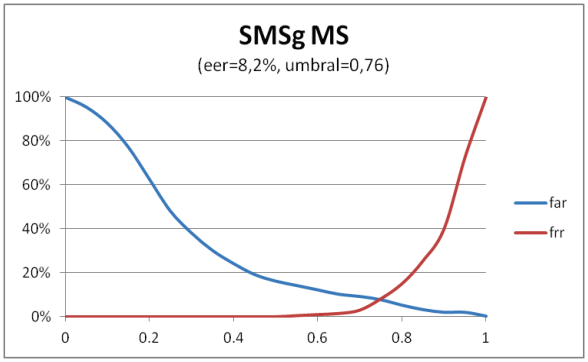
p. Usuario MS

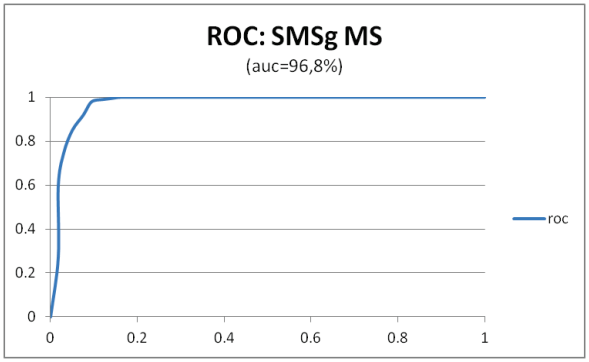

LBg MS
(eer=13,3%, umbral=0,48)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

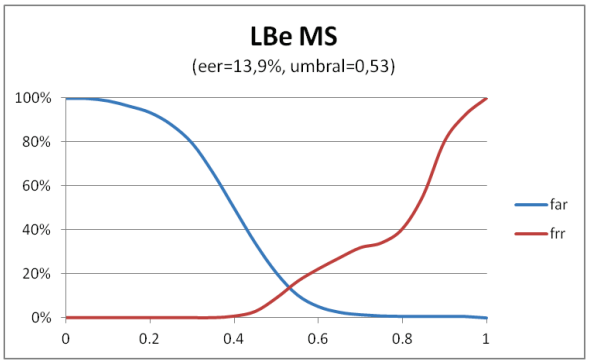

ROC: LBg MS
(auc=94,5%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

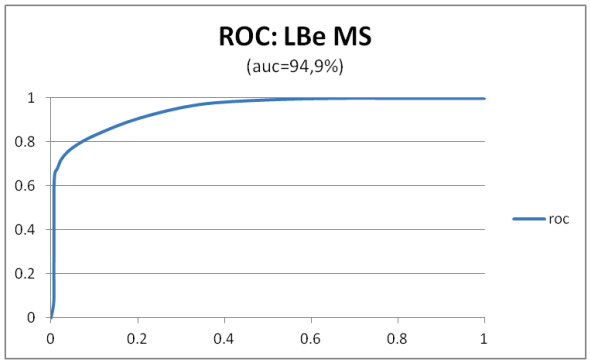

EPSg MS
(eer=8,3%, umbral=0,71)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

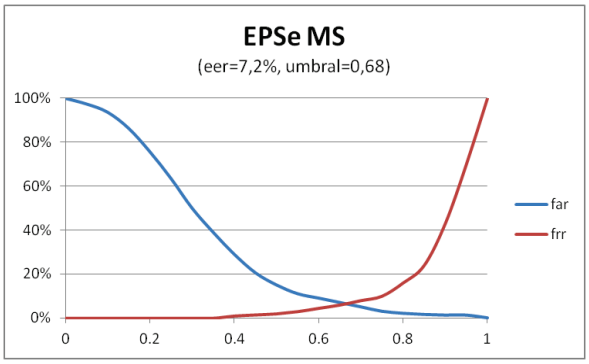

ROC: EPSg MS
(auc=97,2%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

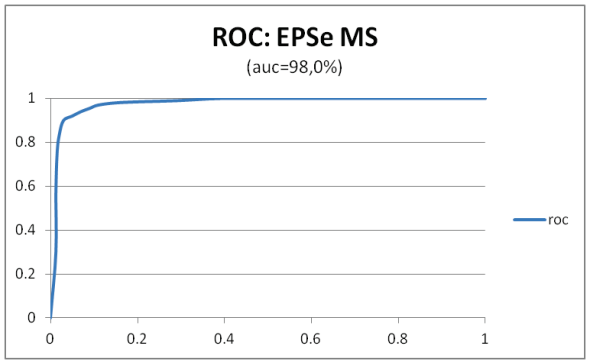

SMSg MS
(eer=8,2%, umbral=0,76)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

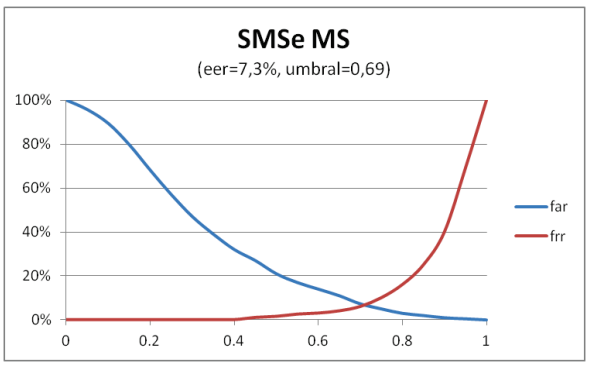

ROC: SMSg MS
(auc=96,8%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

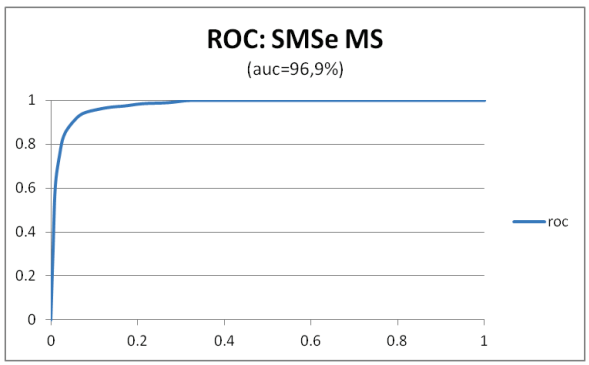

LBe MS
(eer=13,9%, umbral=0,53)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr


ROC: LBe MS
(auc=94,9%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

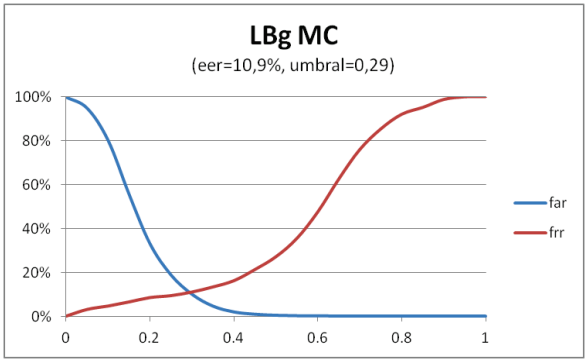

EPSe MS
(eer=7,2%, umbral=0,68)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

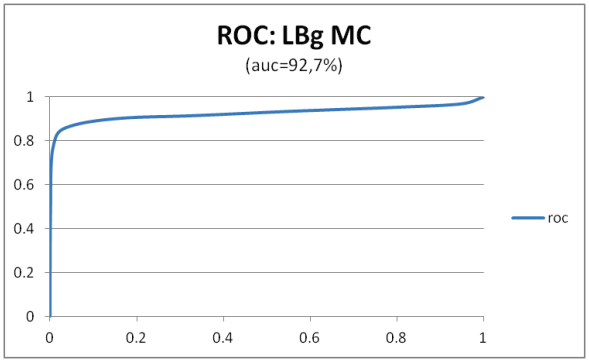

ROC: EPSe MS
(auc=98,0%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

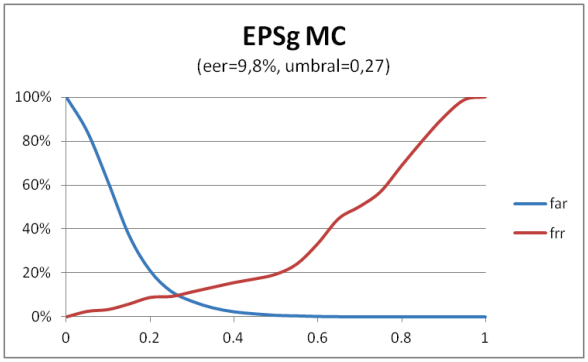

SMSe MS
(eer=7,3%, umbral=0,69)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

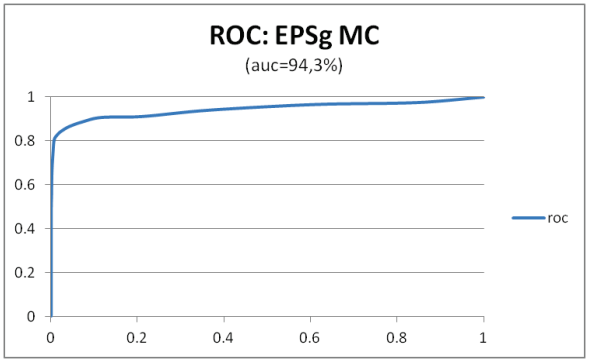

ROC: SMSe MS
(auc=96,9%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

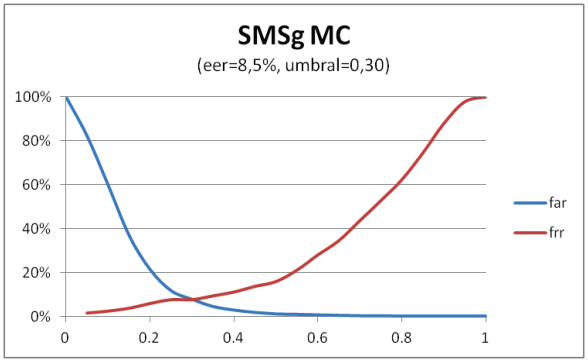
q. Usuario MC

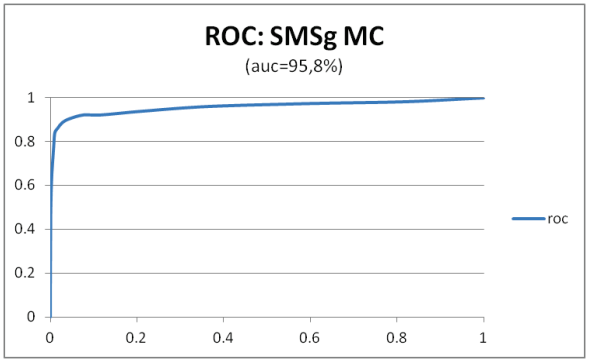

LBg MC
(eer=10,9%, umbral=0,29)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

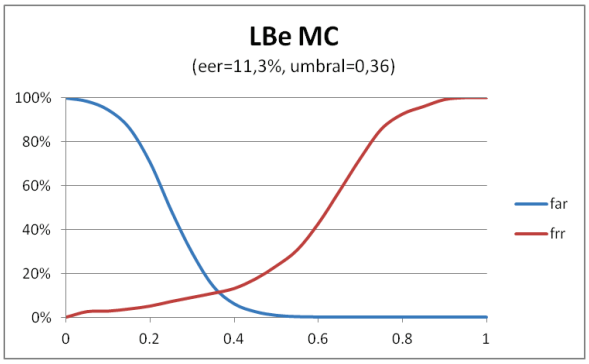

ROC: LBg MC
(auc=92,7%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

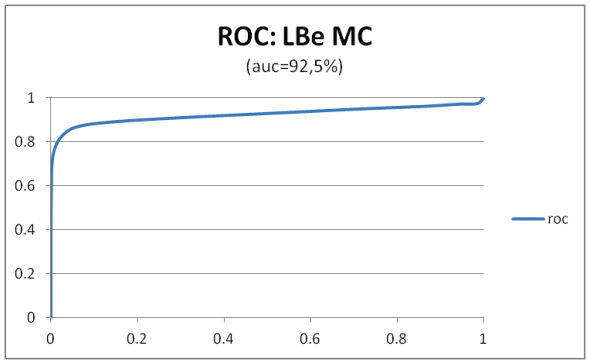

EPSg MC
(eer=9,8%, umbral=0,27)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

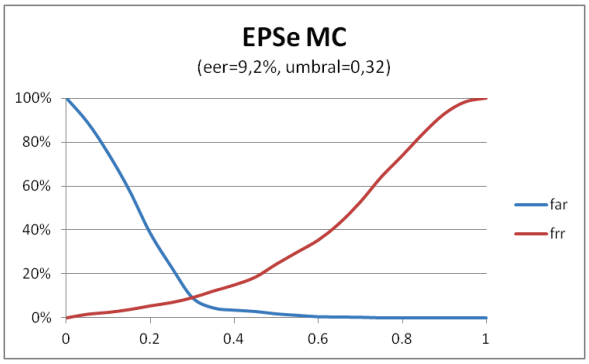

ROC: EPSg MC
(auc=94,3%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

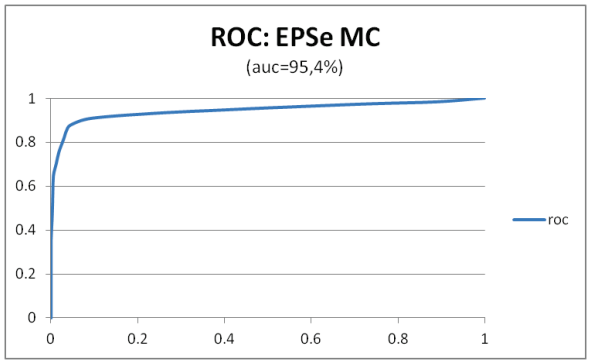

SMSg MC
(eer=8,5%, umbral=0,30)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

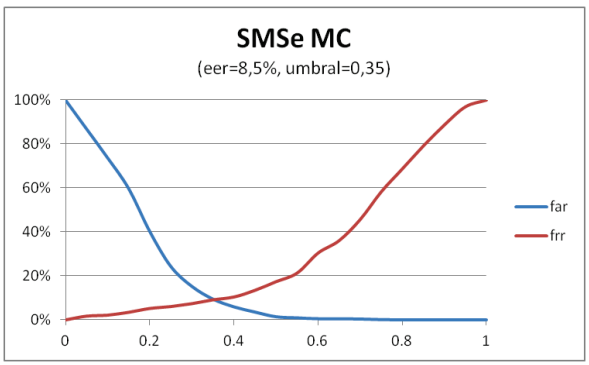

ROC: SMSg MC
(auc=95,8%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

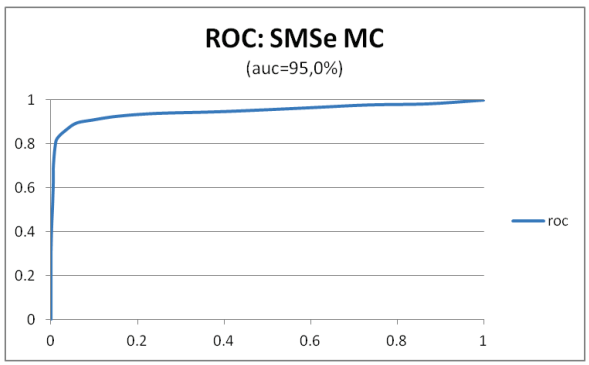

LBe MC
(eer=11,3%, umbral=0,36)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr


ROC: LBe MC
(auc=92,5%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

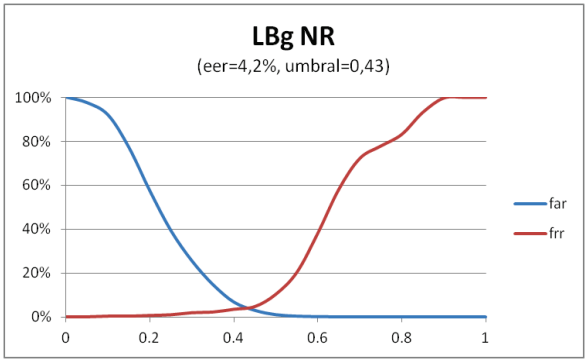

EPSe MC
(eer=9,2%, umbral=0,32)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

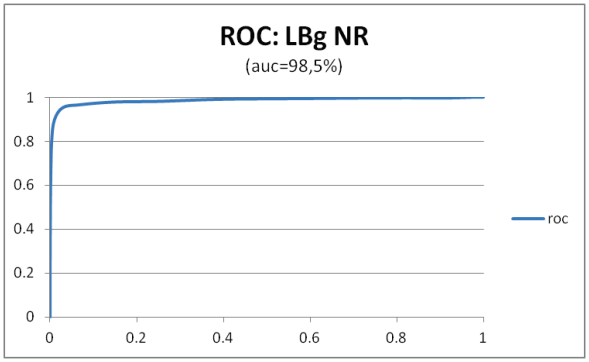

ROC: EPSe MC
(auc=95,4%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

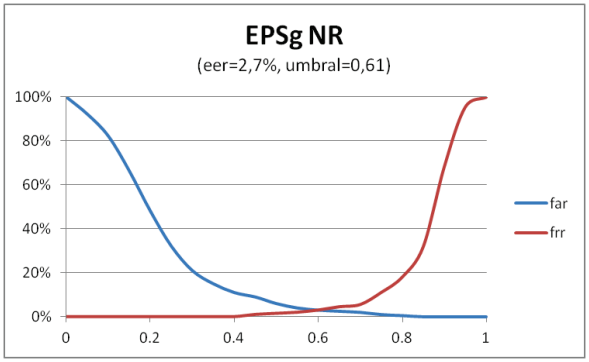

SMSe MC
(eer=8,5%, umbral=0,35)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

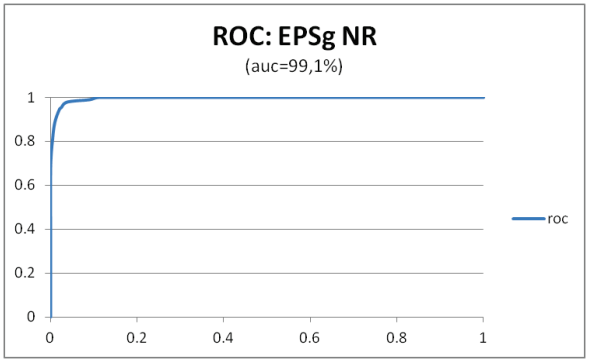

ROC: SMSe MC
(auc=95,0%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

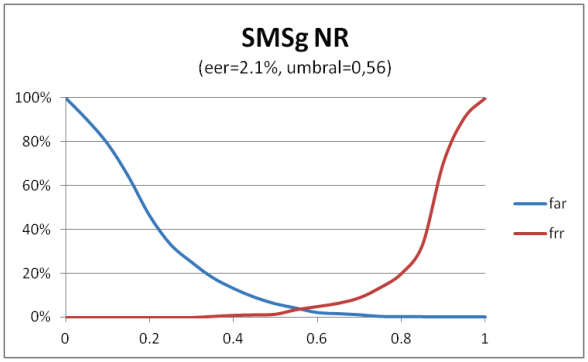
r. Usuario NR

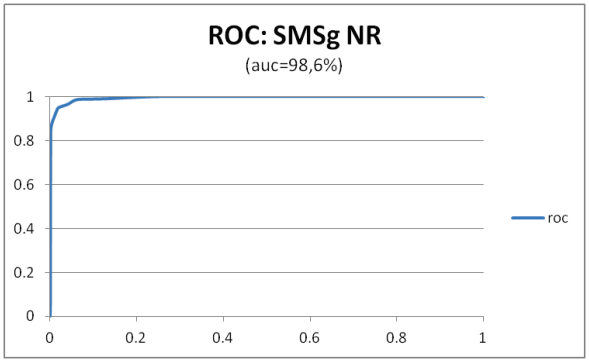

LBg NR
(eer=4,2%, umbral=0,43)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

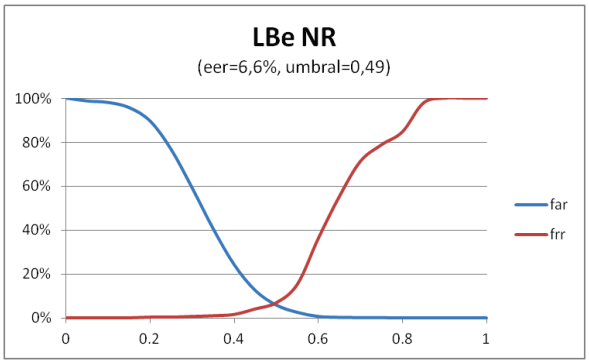

ROC: LBg NR
(auc=98,5%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

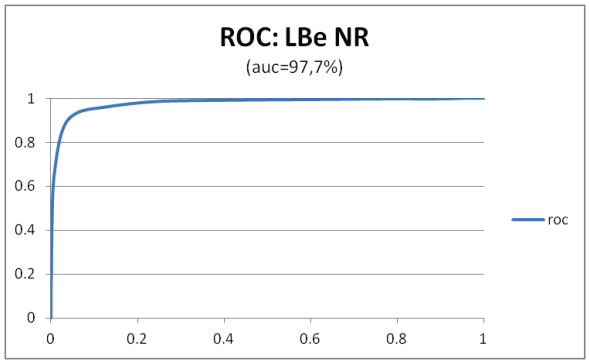

EPSg NR
(eer=2,7%, umbral=0,61)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

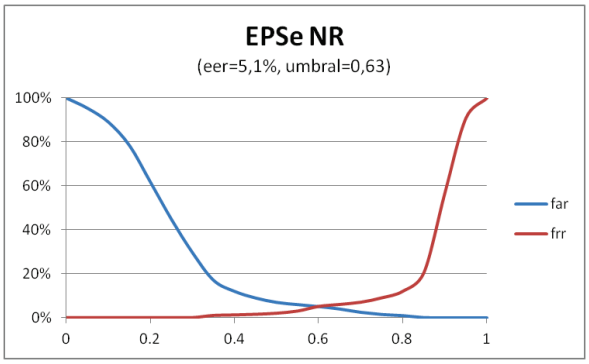

ROC: EPSg NR
(auc=99,1%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

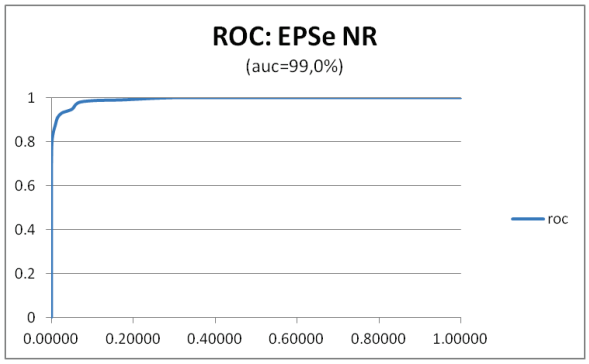

SMSg NR
(eer=2.1%, umbral=0,56)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

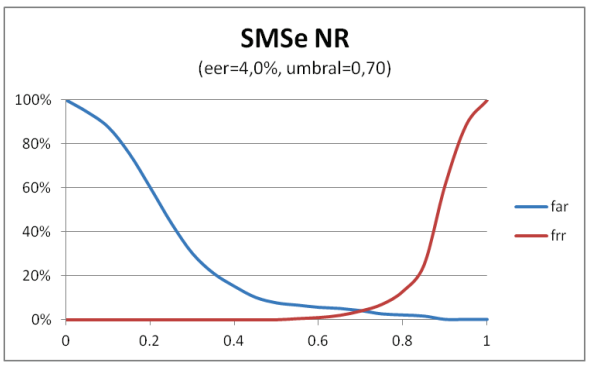

ROC: SMSg NR
(auc=98,6%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

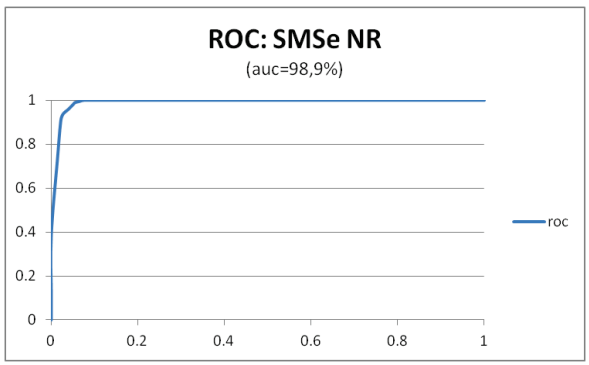

LBe NR
(eer=6,6%, umbral=0,49)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr


ROC: LBe NR
(auc=97,7%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

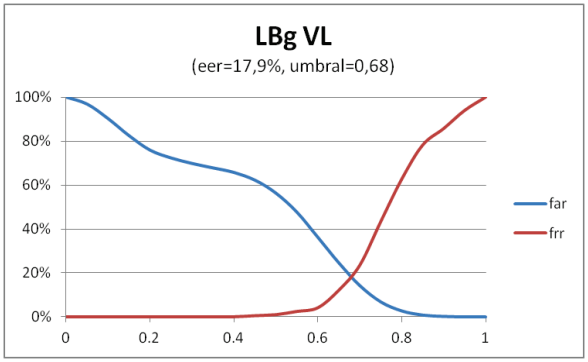

EPSe NR
(eer=5,1%, umbral=0,63)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

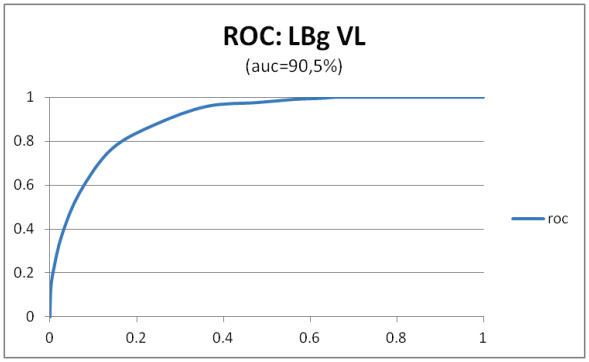

ROC: EPSe NR
(auc=99,0%)
1
0.8
0.6
0.4
0.2
0
0.00000
0.20000
0.40000
0.60000
0.80000
1.00000
roc

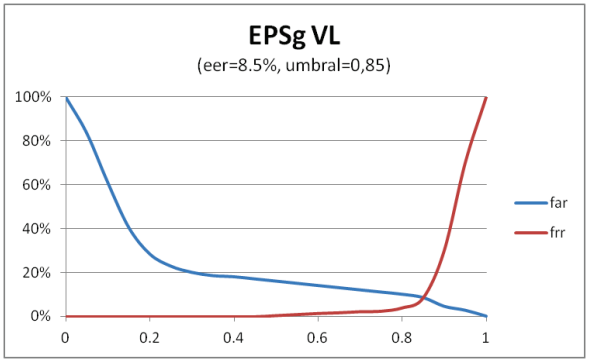

SMSe NR
(eer=4,0%, umbral=0,70)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

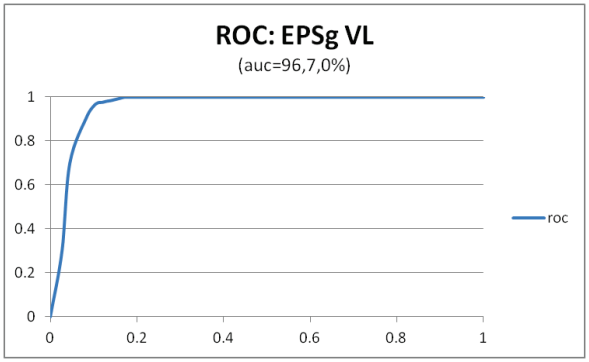

ROC: SMSe NR
(auc=98,9%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

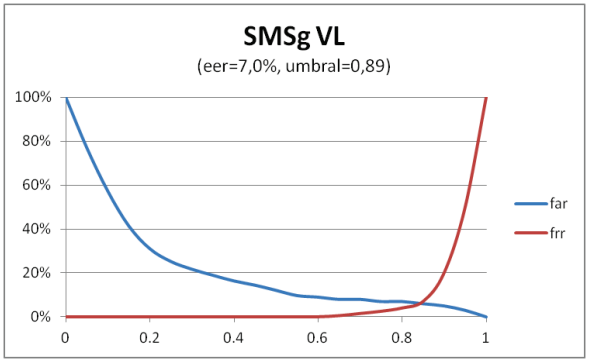
s. Usuario VL

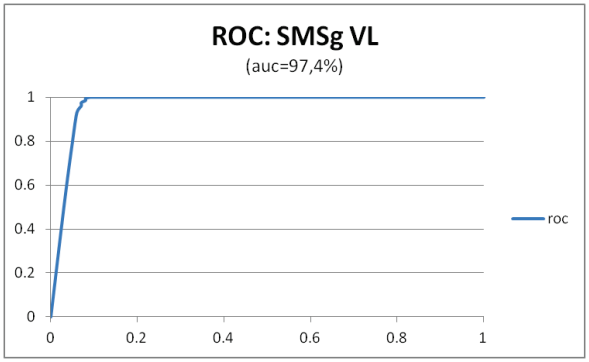

LBg VL
(eer=17,9%, umbral=0,68)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

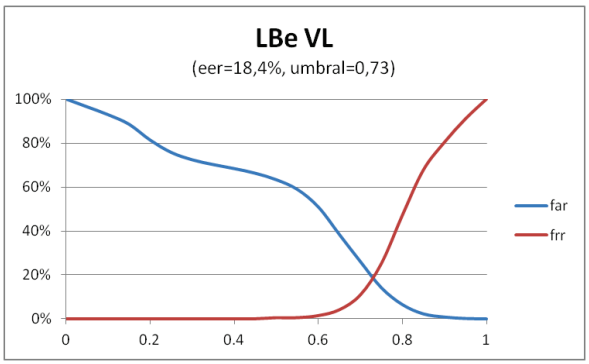

ROC: LBg VL
(auc=90,5%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

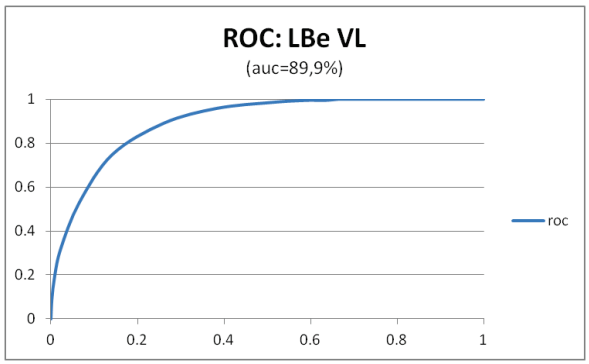

EPSg VL
(eer=8.5%, umbral=0,85)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

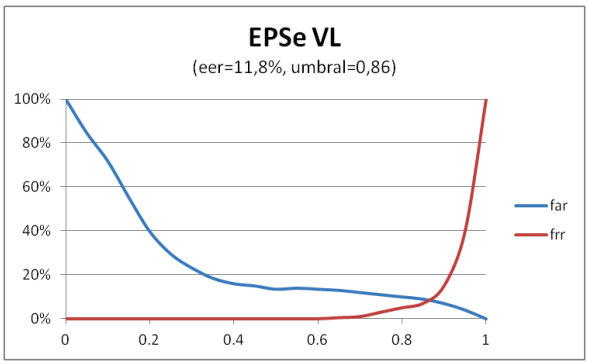

ROC: EPSg VL
(auc=96,7,0%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

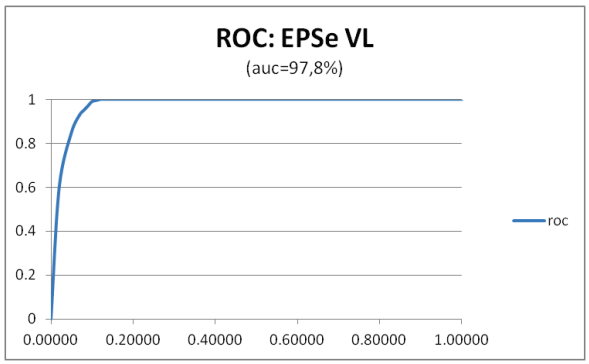

SMSg VL
(eer=7,0%, umbral=0,89)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

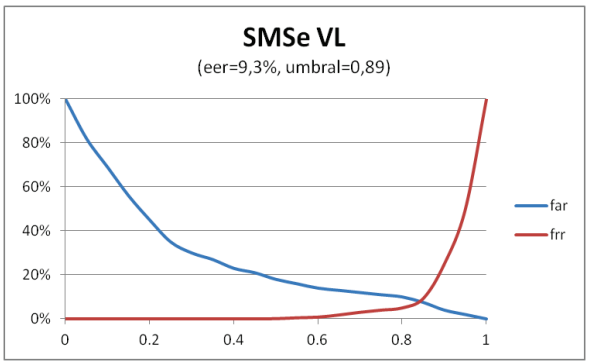

ROC: SMSg VL
(auc=97,4%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

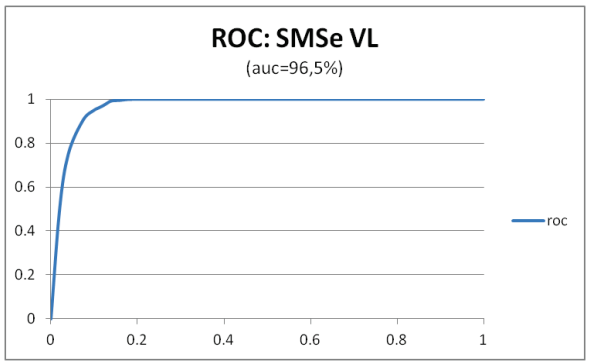

LBe VL
(eer=18,4%, umbral=0,73)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr


ROC: LBe VL
(auc=89,9%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

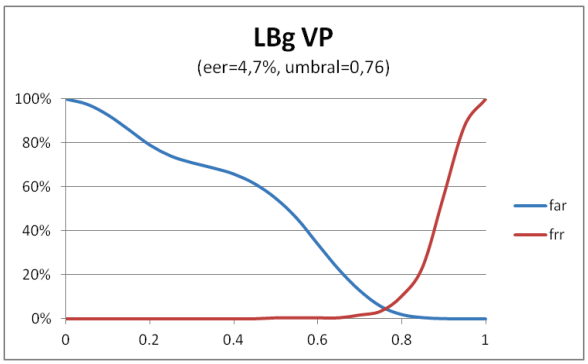

EPSe VL
(eer=11,8%, umbral=0,86)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

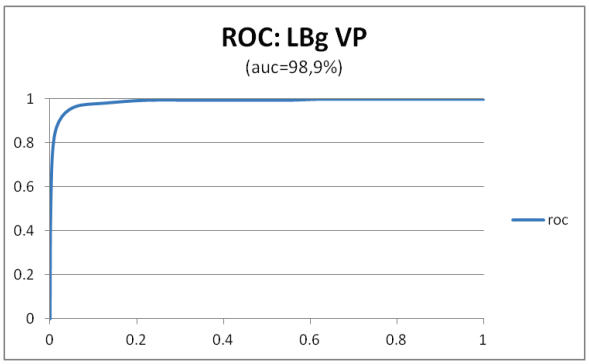

ROC: EPSe VL
(auc=97,8%)
1
0.8
0.6
0.4
0.2
0
0.00000
0.20000
0.40000
0.60000
0.80000
1.00000
roc

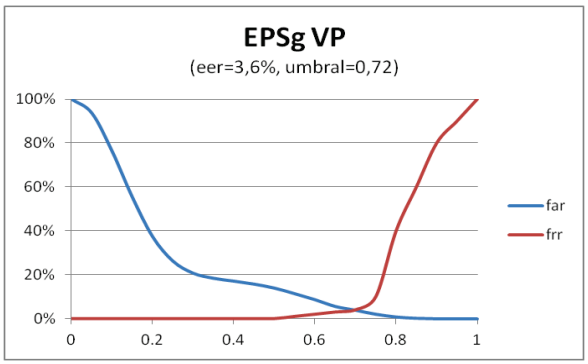

SMSe VL
(eer=9,3%, umbral=0,89)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

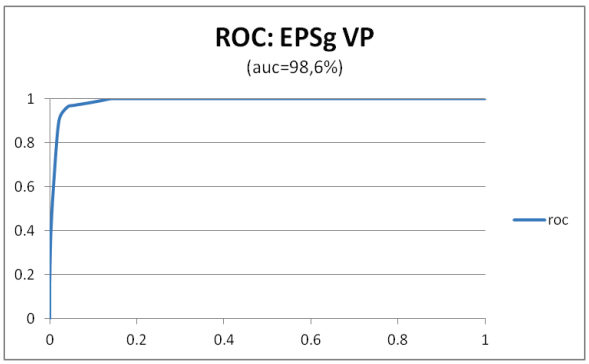

ROC: SMSe VL
(auc=96,5%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

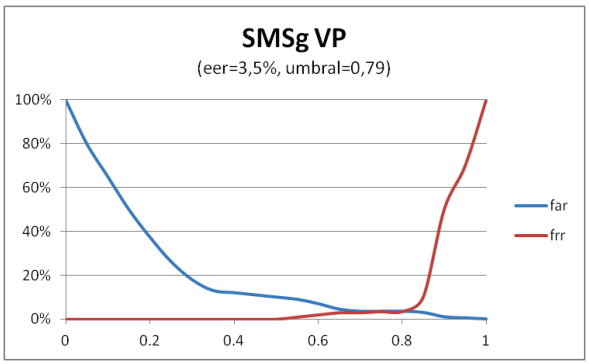
t. Usuario VP

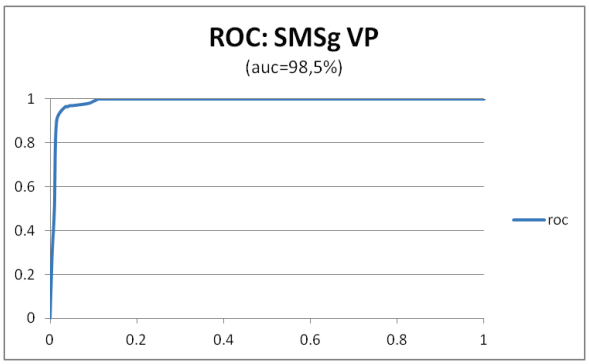

LBg VP
(eer=4,7%, umbral=0,76)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

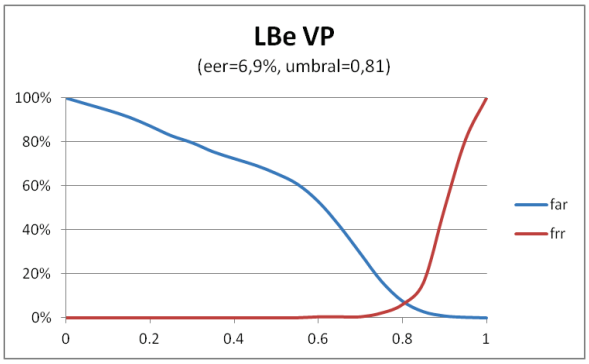

ROC: LBg VP
(auc=98,9%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

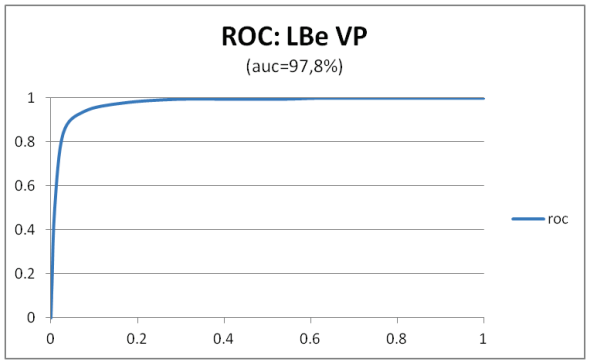

EPSg VP
(eer=3,6%, umbral=0,72)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

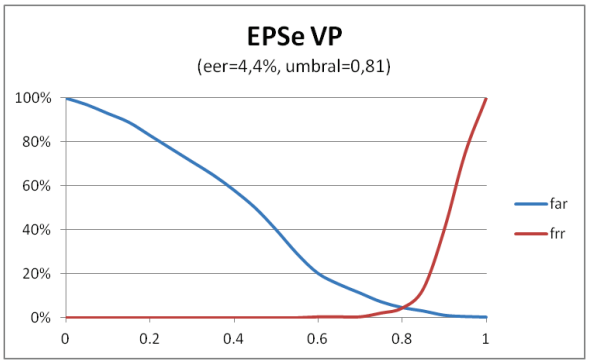

ROC: EPSg VP
(auc=98,6%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

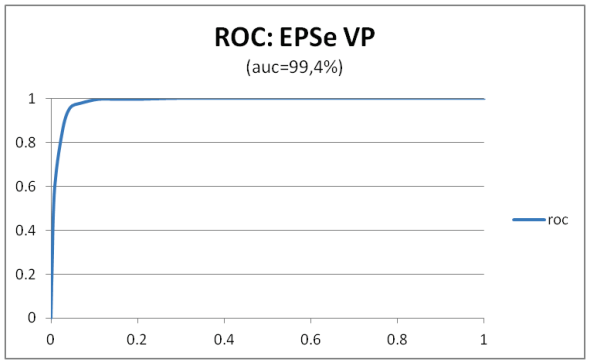

SMSg VP
(eer=3,5%, umbral=0,79)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

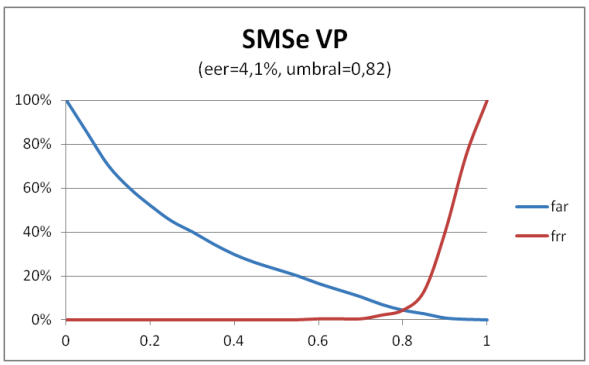

ROC: SMSg VP
(auc=98,5%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

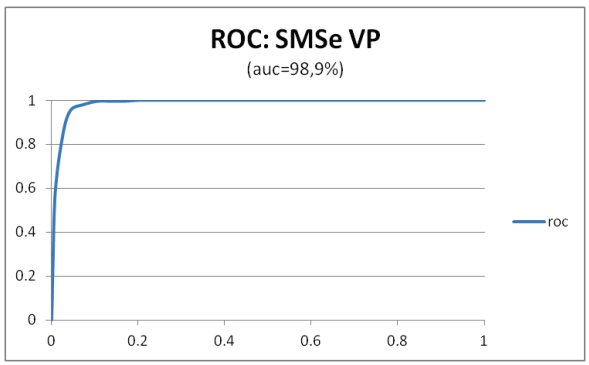

LBe VP
(eer=6,9%, umbral=0,81)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr


ROC: LBe VP
(auc=97,8%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc


EPSe VP
(eer=4,4%, umbral=0,81)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

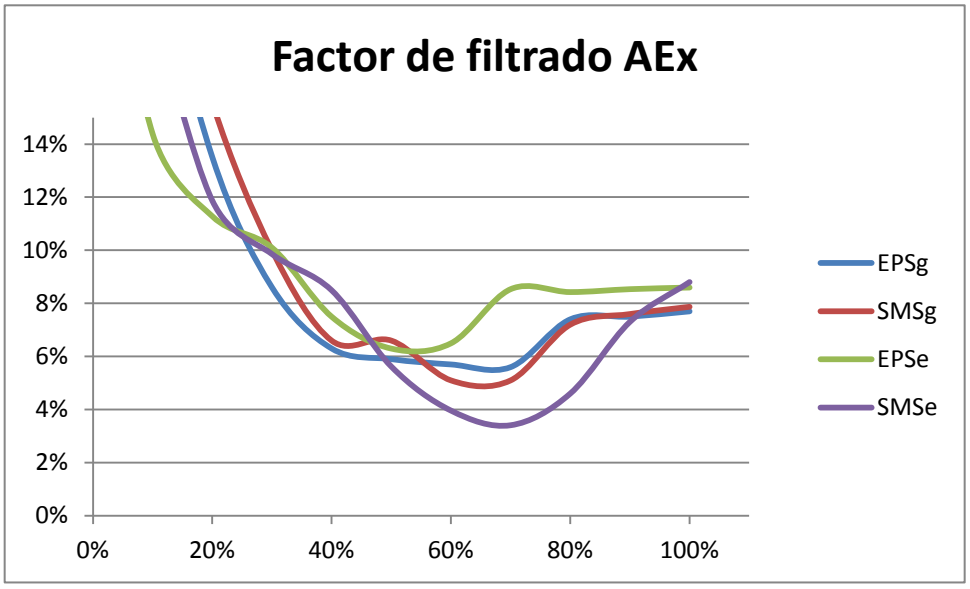

ROC: EPSe VP
(auc=99,4%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc


SMSe VP
(eer=4,1%, umbral=0,82)
100%
80%
60%
40%
20%
0%
0
0.2
0.4
0.6
0.8
1
far
frr

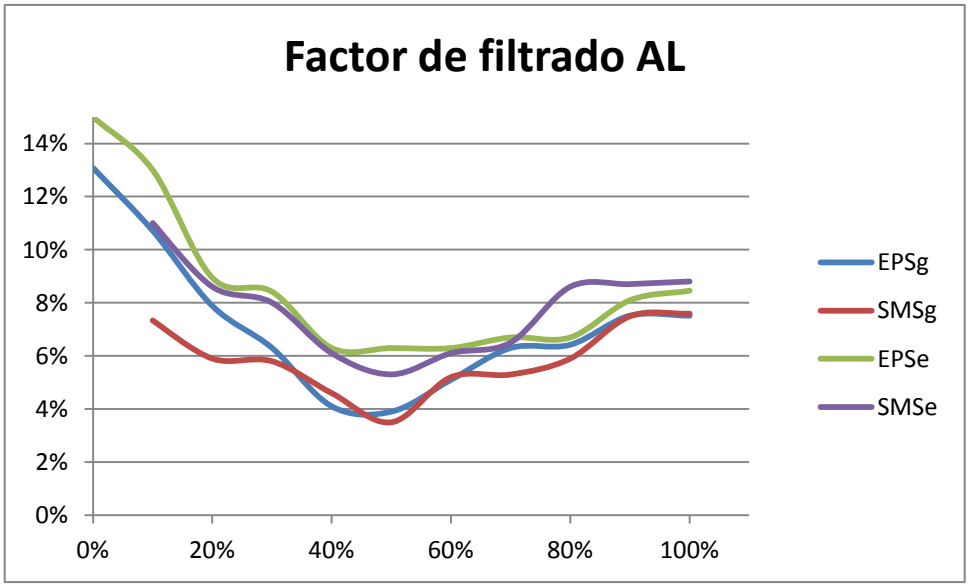

ROC: SMSe VP
(auc=98,9%)
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
roc

Apéndice IV. Factores de filtrado individuales

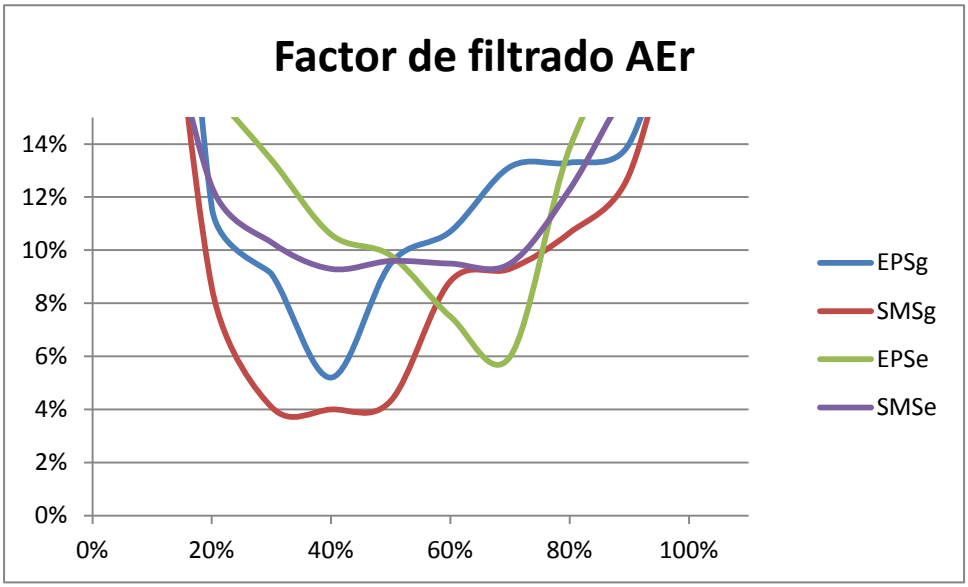
a. Usuario AEx



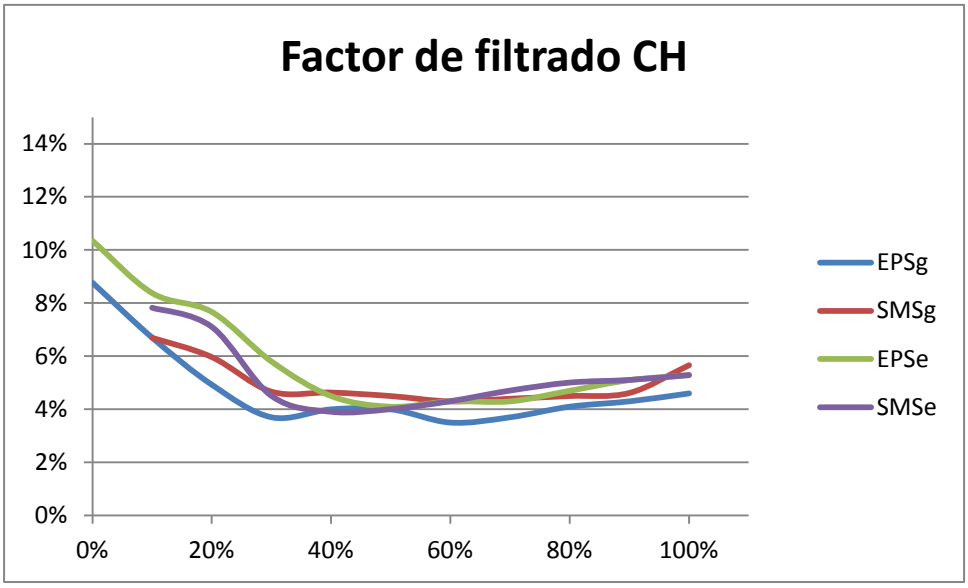
b. Usuario AL



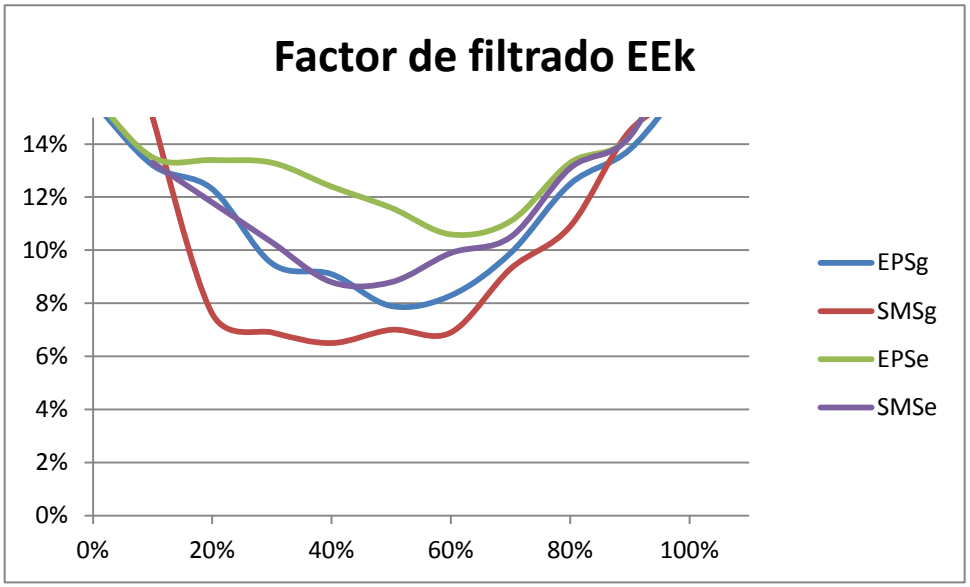
c. Usuario AEr



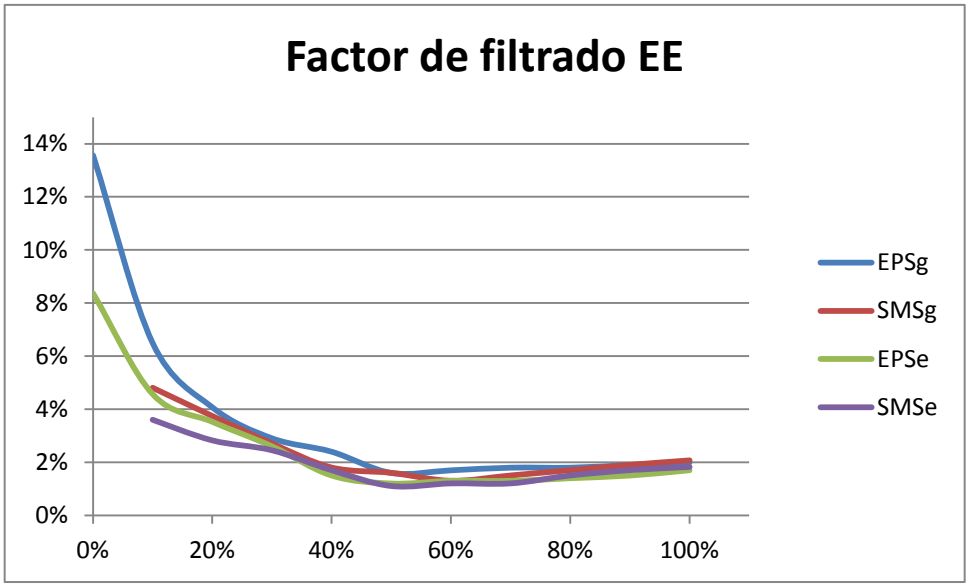
d. Usuario CH



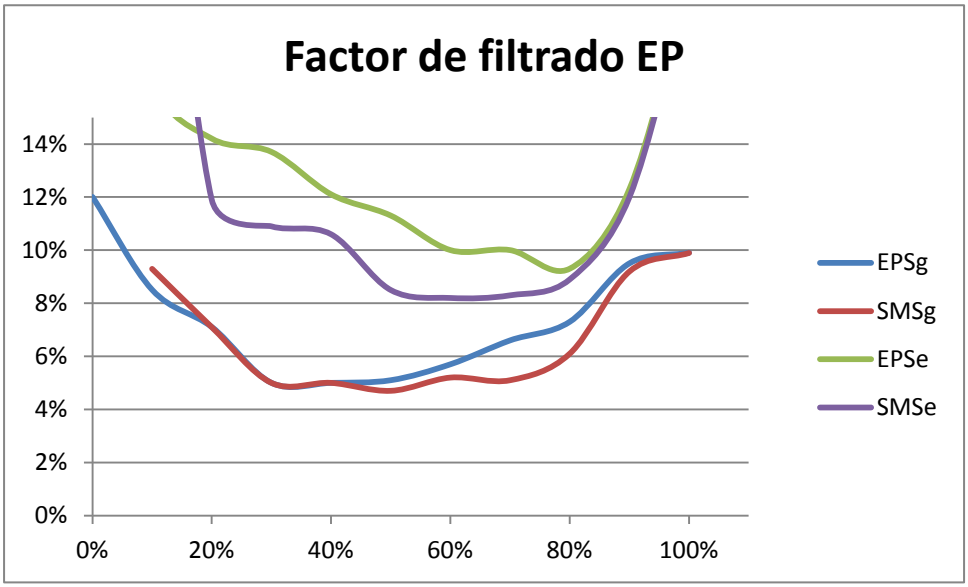
e. Usuario EEk



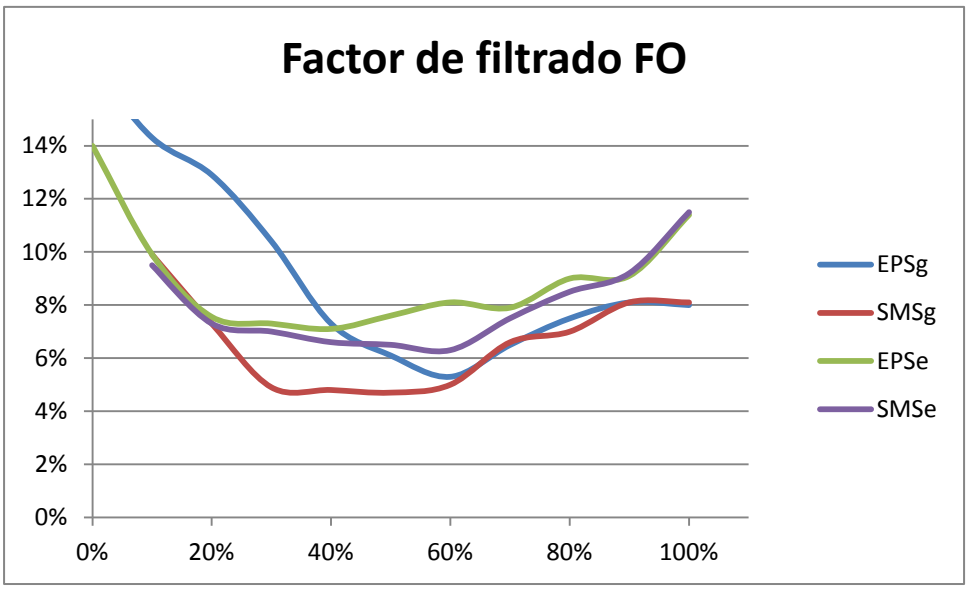
f. Usuario EE



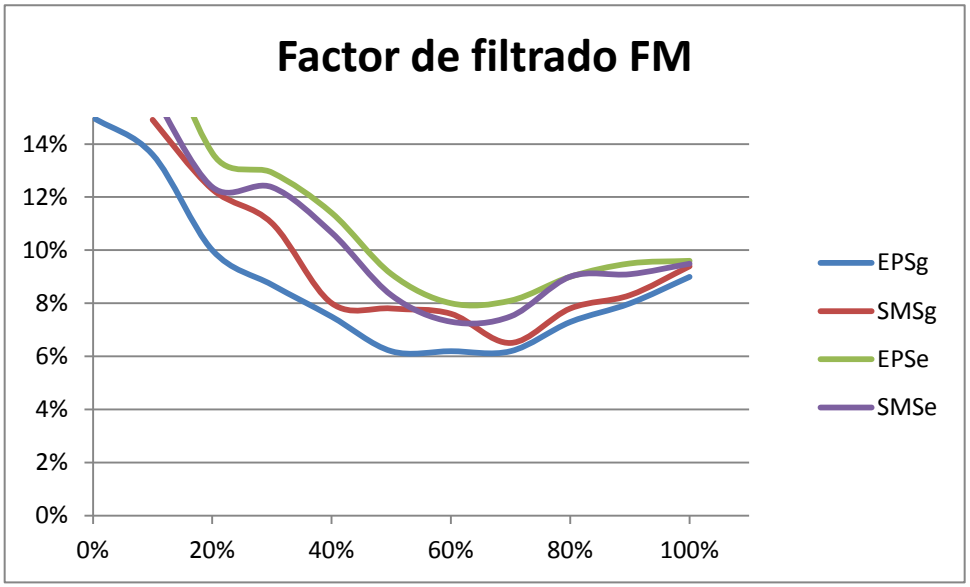
g. Usuario EP



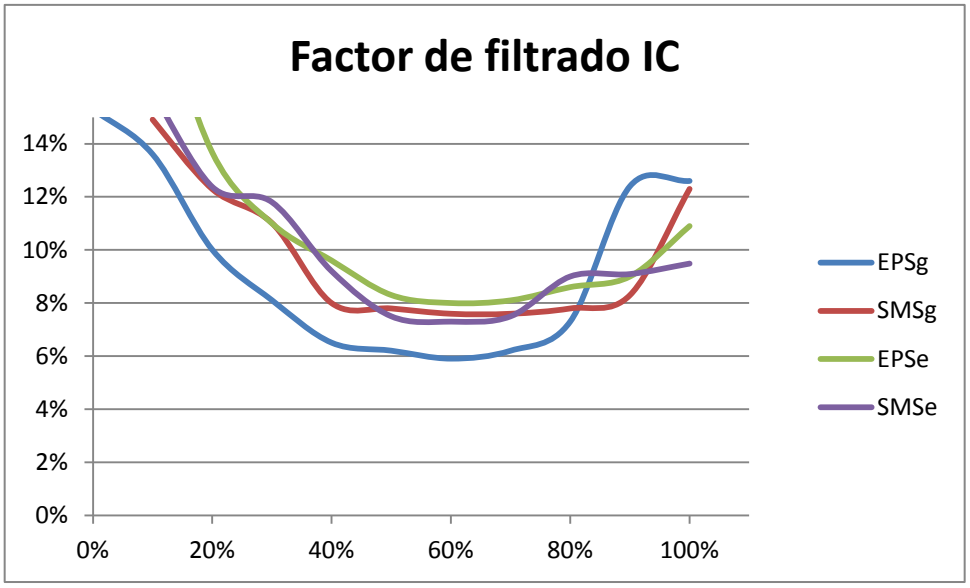
h. Usuario FO



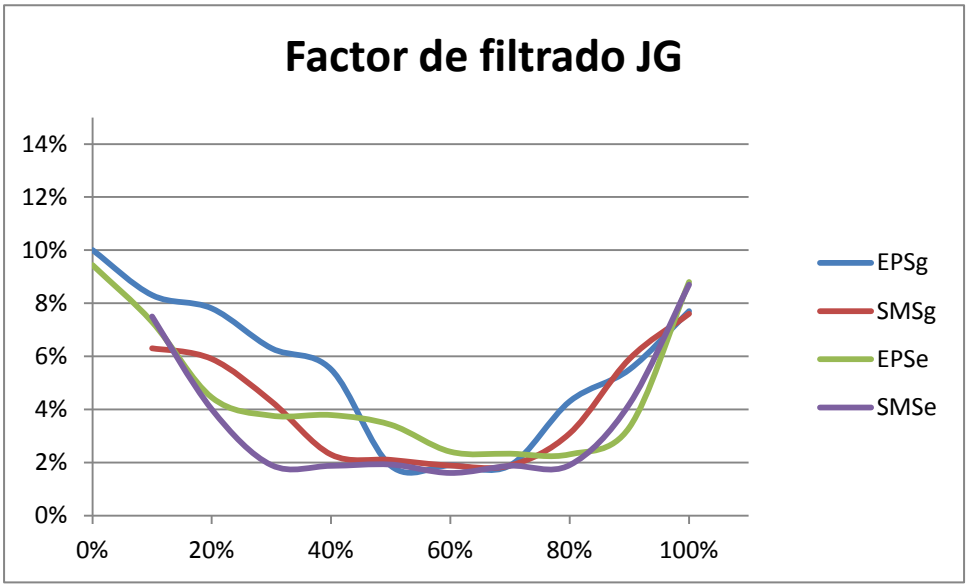
i. Usuario FM



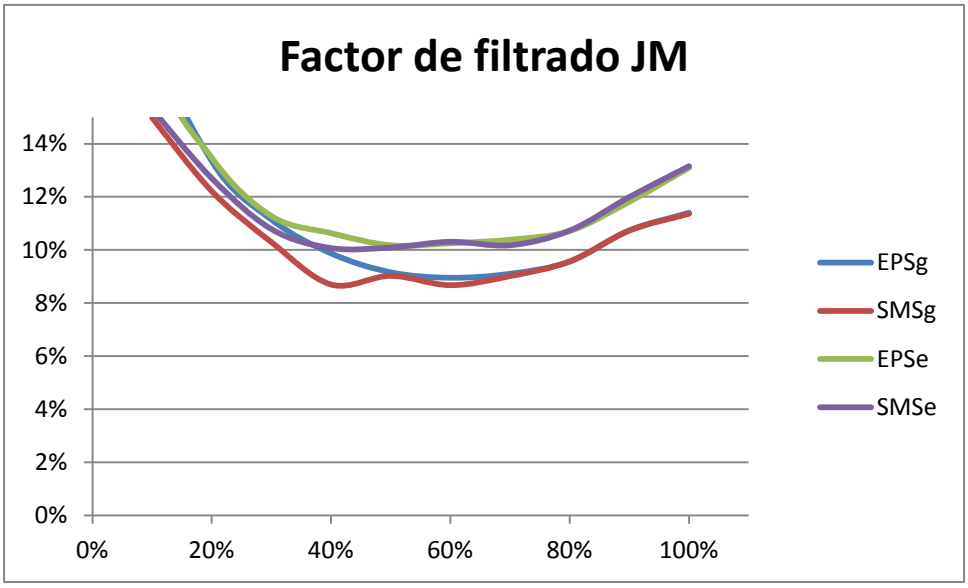
j. Usuario IC



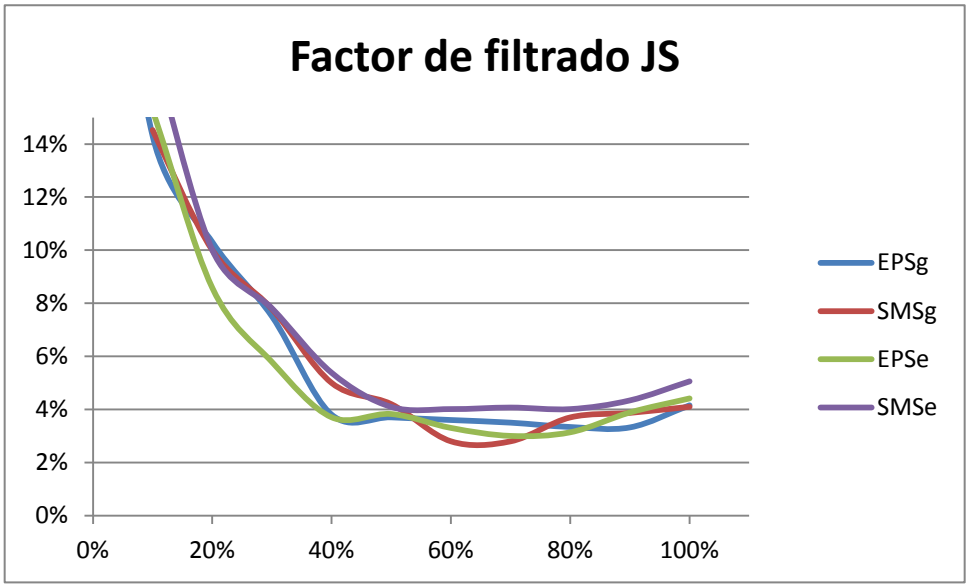
k. Usuario JG



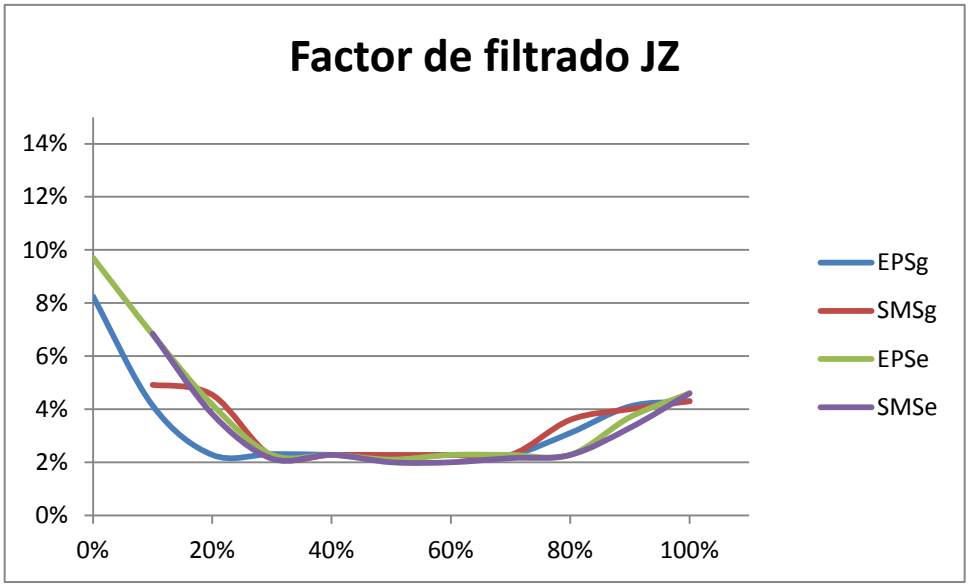
l. Usuario JM



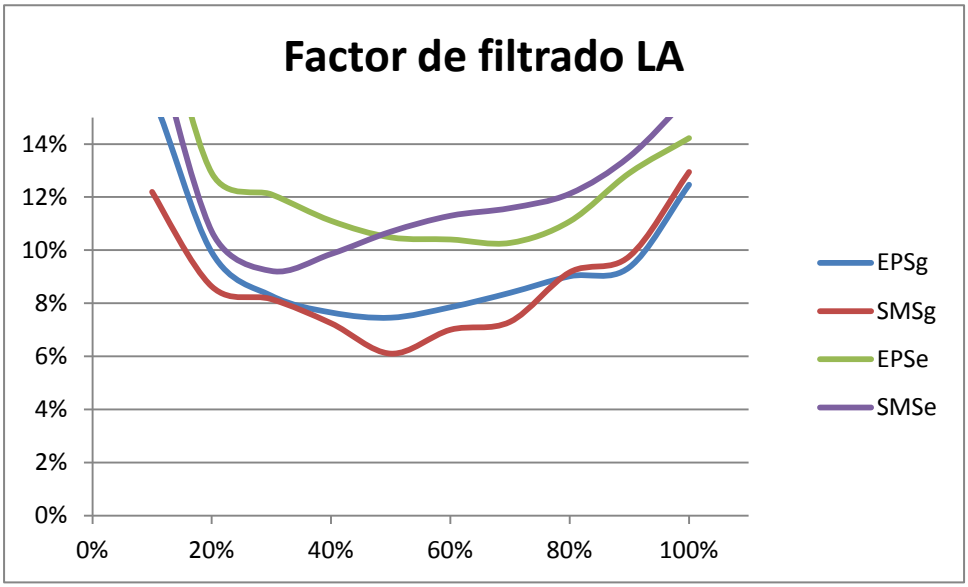
m. Usuario JS



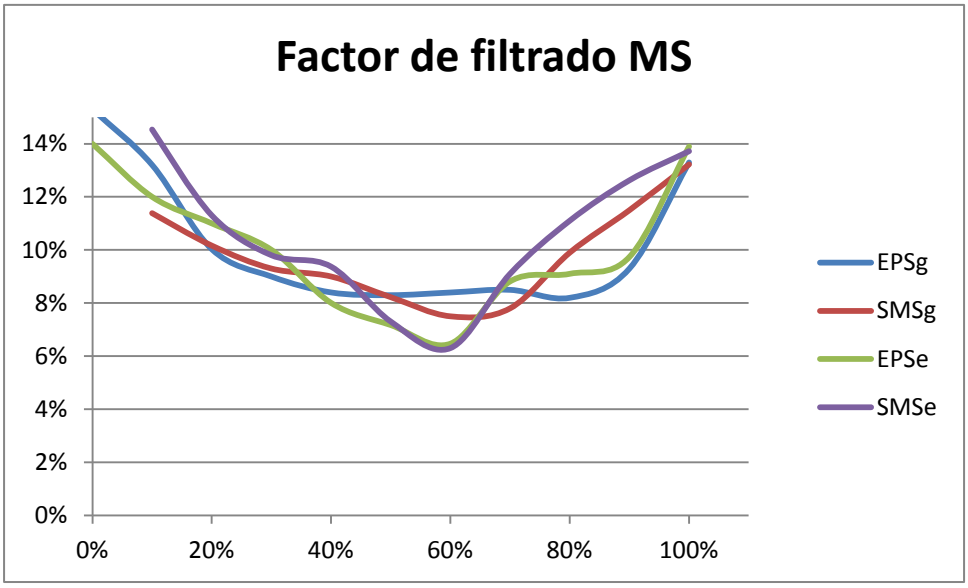
n. Usuario JZ



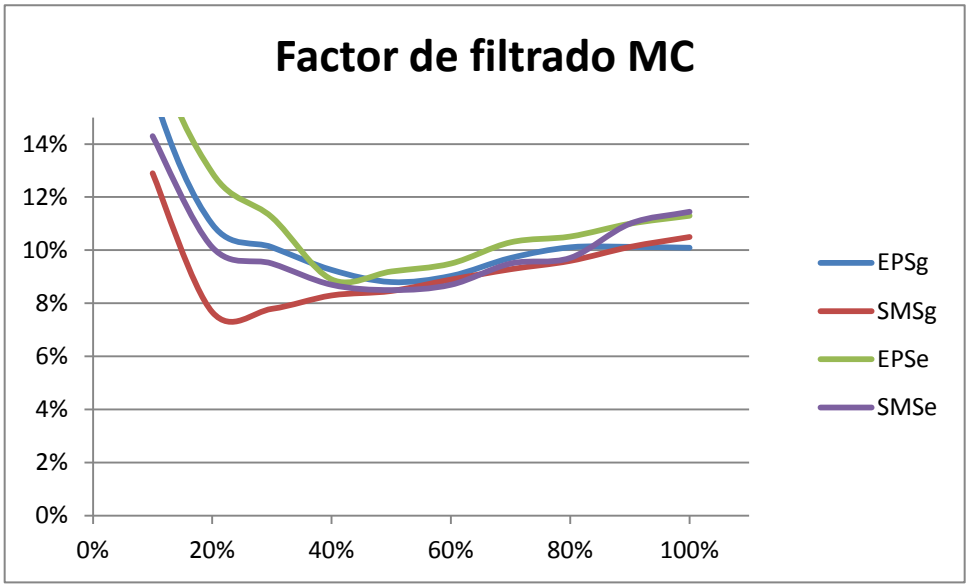
o. Usuario LA



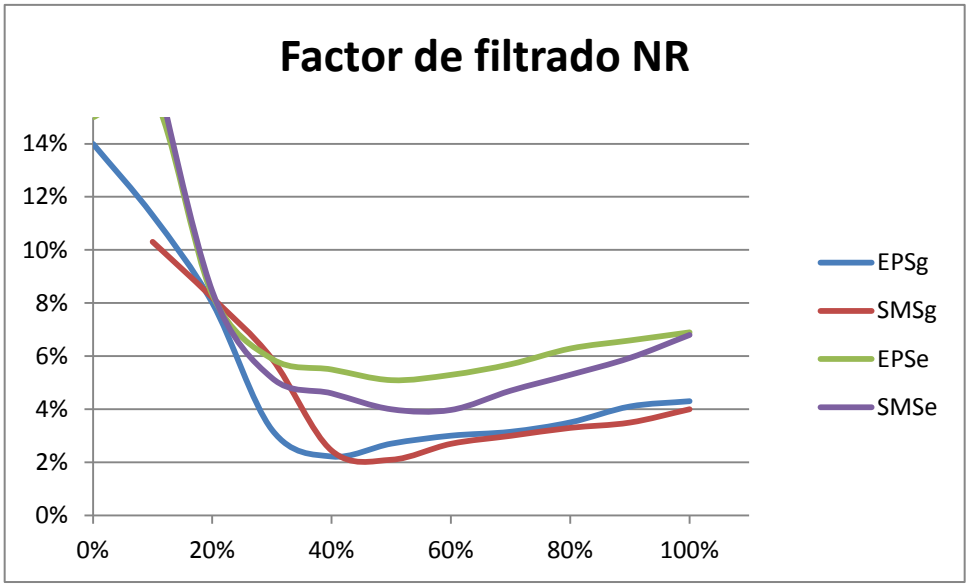
p. Usuario MS



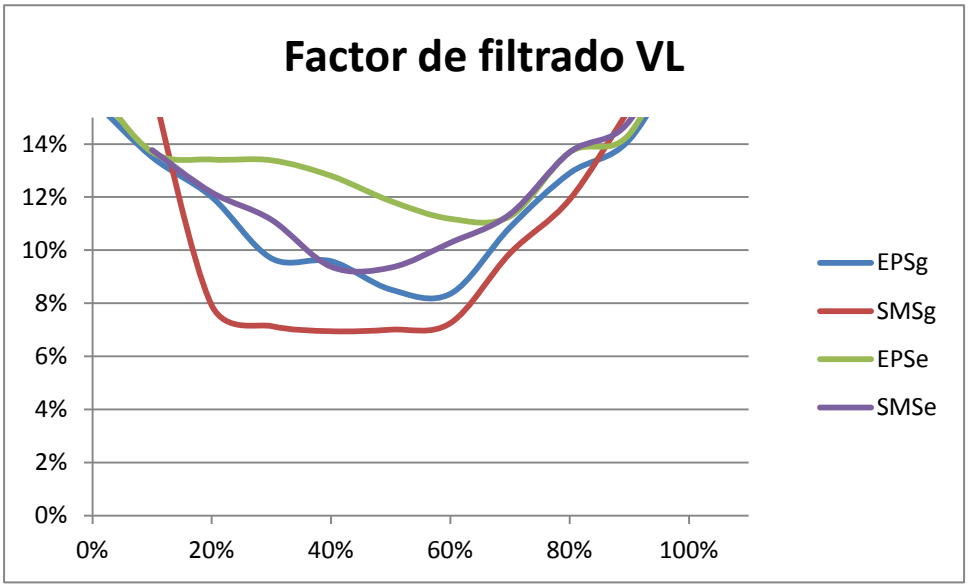
q. Usuario MC



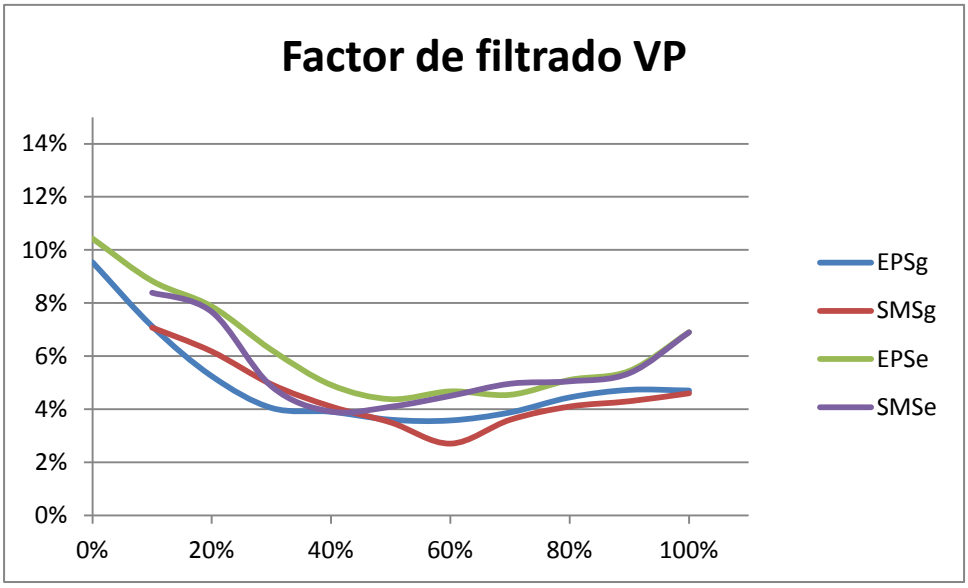
r. Usuario NR



s. Usuario VL



t. Usuario VP



Apéndice V. Herramientas

Para la realización de los trabajos de experimentación asociados a la presente tesis ha sido necesario utilizar un conjunto de herramientas en algunos casos desarrollos ad-hoc cuando las disponibles no cubrían adecuadamente las necesidades particulares de este trabajo.

a. Total Video Converter

Total Video Converter es un programa para extraer y convertir videos a formatos compatibles con numerosos dispositivos electrónicos.

El diseño y las funciones de este programa son sencillos y fáciles de navegar. Se puede acceder a las principales funciones desde la ventana principal. El tiempo de conversión variará según la tarea que realice, dependiendo del tamaño y del formato del archivo, pero maneja la mayoría de los formatos populares de audio, como mp4, 3gp, xvid, divx, mpeg4, avi y amr.

Esta herramienta fue utilizada para procesar los videos, videos que estaban disponibles en formatos avi y mpeg4. Con ella de los videos se extrajeron los audios en formato *wave*, que es con el que trabaja el extractor de características.

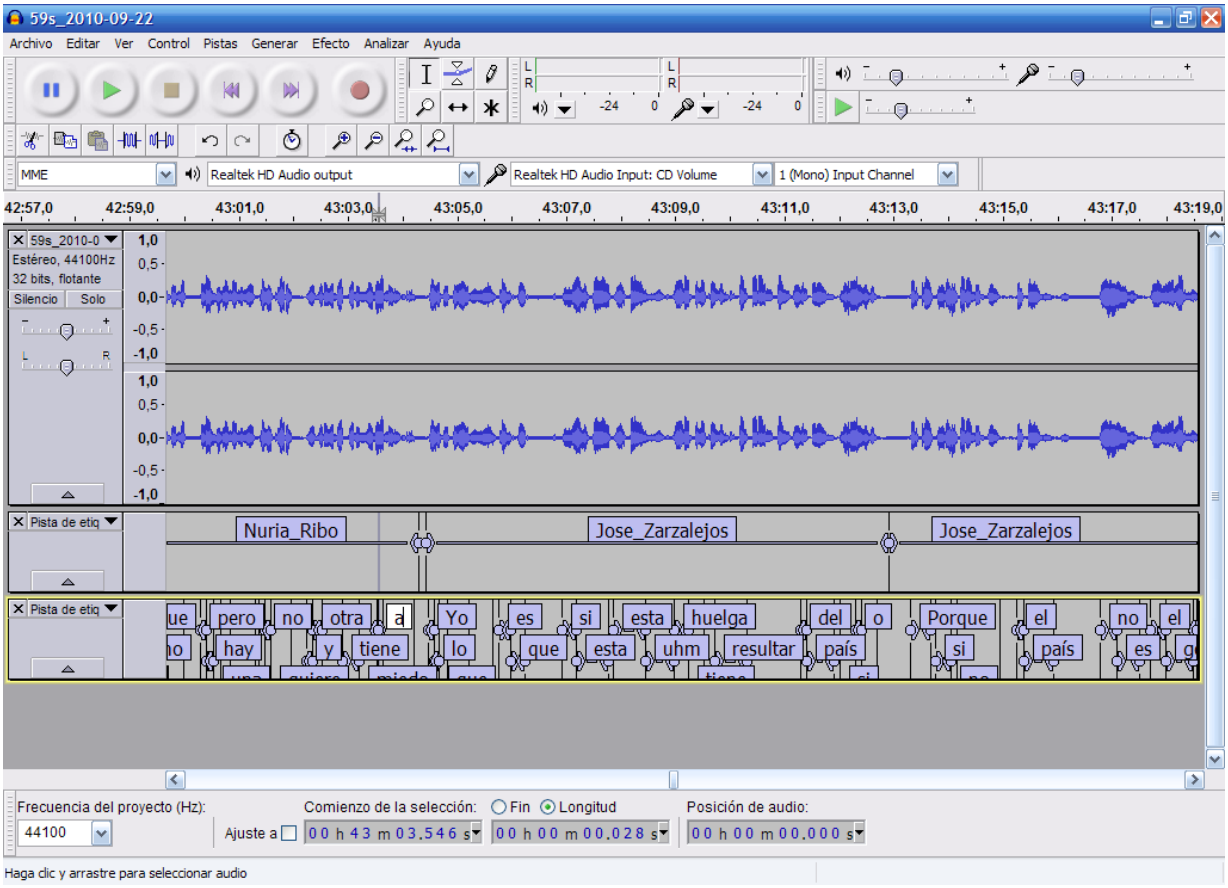
b. Audacity®

Audacity®[26] es un editor de grabación y edición de sonido libre, de código abierto y multiplataforma. Es una aplicación libre, escrita por un equipo de desarrolladores voluntarios disponible para múltiples sistemas operativos (Windows, Mac OS, Linux, Unix). El proyecto se puede encontrar tanto en SourceCode.net como en Google Code.

Este editor de audio fue utilizado por poseer la capacidad de asociar y editar de forma sencilla etiquetas a fragmentos de de un archivo de audio. De este modo permitió la etiquetación manual de todos los archivos de la base de datos.

Aunque en el momento de redacción de este documento la versión operativa es 2.0.3, para el desarrollo de esta tesis se utilizó una versión modificada por el doctorando de la 1.3.13, a la cual se le añadieron ciertas funcionalidades al interfaz de usuario que facilitaron las operaciones de etiquetado.

[26] http://audacity.sourceforge.net/



c. PitchMarker

El PitchMarker es una herramienta desarrollada por el doctorando, cuya finalidad original era el marchado de inicios los ciclos de la frecuencia fundamental, uno de los pasos iniciales en para la construcción de voces sintetizadas. Para los objetos de esta tesis fue adaptada para permitir el marcado de los centros de los fonemas.

d. Transcriptor

El transcriptor es una herramienta también desarrollada por el proyectando,, utilizada para realizar la transcripción de las locuciones, disponer de esta información simplifica en gran medida la identificación de los fonemas, ya que es habitualmente sencillo localizarlos cuando es conocido cuales son. No se hizo uso de herramientas existentes dada su mal comportamiento con archivos de audio grandes.

e. Extractor de características

El extractor de características fue desarrollado por el doctorando. Corresponde a una aplicación de consola cuyos argumentos de entrada son el archivo de audio a procesar, el archivo de vectores de características y un archivo con la parametrización.

El extractor de características realiza tanto el preprocesado como la obtención de los vectores de características.

Se opto por el desarrollo específico, en lugar de utilizar otros preexistentes como HLIST de HTK muy popular entre los investigadores de esta área de conocimiento, ya que estas herramientas son incapaces de trabajar con archivos de audio del volumen de los utilizados en esta tesis.

La aplicación fue desarrollada en C#, lenguaje utilizado por la alta productividad que permite, y por ser compatible con módulos y librerías ya desarrolladas para .NET.

Los archivos de audio que acepta deben encontrarse en formato wav. Los ficheros de vectores producidos se encuentran en un formato ad hoc denominado plt y que es manejado por una librería de utilidades denominada BinMatrix.

f. BinMatrix

BinMatrix es una librería de gestión, lectura y escritura de los archivos de formato plt, diseñada por el doctorando.

Se opto por el diseño del formato plt, como alternativa a los utilizados por otras herramientas como HTK, porque estos últimos almacenan la información en formato texto lo que presenta dos inconvenientes, el primero es la ineficiencia en el espacio ocupado por los archivos, crítico en el presente caso debido al tamaño de los archivos a manejar; el segundo es su ineficiencia en tiempo de proceso ya que en cada pasada del algoritmo es necesario convertir todos los textos a valores numéricos. Alternativamente plt es un formato crudo, que representa una matriz de valores reales en formato double, precedida de una cabecera con la descripción del contenido. Este formato crudo permite la utilización directa de las rutinas de bajo nivel de C++, sin necesidad de conversión de datos.

La librería fue desarrollada en C++ para entorno .NET.

g. Etiquetas

Etiquetas es una librería desarrollada en C# para la gestión de los archivos de etiquetas que produce Audacity®. Esta librería permite la integración directa de la información de dichos archivos en las aplicaciones que conforman la experimentación que soporta esta tesis.

h. SVM^{light}

SVM^{light} es una implementación de una máquina de vectores soporte (Support Vector Machine) en C. Cuyas prestaciones principal es la utilización de un rápido algoritmo de optimización.

Desde el punto de vista de interés para el desarrollo de la tesis, el motor se implementa en forma de dos aplicaciones una dedicada a la creación de un modelo que se ajuste a los vectores ejemplo proporcionados (SVM_Learn), y una segunda que clasifica un conjunto de vectores datos según un modelo (SVM_Classify). Soporta los kernel estándar entre ellos "Radial" utilizado en esta experimentación. Para la realización de esta tesis se utilizo una variación de SVM_Classify, que se ha denominado SVM_Scorer, el cual incluye una pequeña modificación para que el clasificador presente las puntuaciones de cada vector en lugar de la indicación de su clase.

No existe ningún criterio especial para haber escogido SVM^{light} frente a otros como LIBSVM, ya ante las necesidades de cálculo presentan prestaciones equivalentes.

i. SVMlpr

Es una librería desarrollada por el doctorando en C#, para encapsular el manejo de los archivos de vectores que SVM^{light} utiliza. Consta de dos clases SVMWriter y SVMReader que manejan la creación y lectura respectivamente de este tipo de archivos

j. Clasificador

Es una herramienta desarrollada por el doctorando en C#, realiza la clasificación de segmentos de audio haciendo, para ello, uso de SVM^{light}, quien realiza la clasificación de los vectores individualmente.

k. Secuenciador y automatizador de experimentos

Varias aplicaciones desarrolladas ad hoc por el doctorando, para cada tipo de experimento. Cada una secuencia y automatiza la realización de los experimentos: generación de modelos, evaluación de umbrales de decisión, verificación múltiple de los fragmentos de test, evaluación de los resultados, estimación de las tasas, y registro de la información obtenida.

l. Analizadores de resultados

Para el análisis de los resultados y su formateo se han utilizado hojas de cálculo Excel a las que se .importaron los resultados de las herramientas en formato XML