Universidad Autónoma de Madrid
Escuela politécnica superior

Proyecto fin de carrera

UTILIZACIÓN DE MEDIDAS DE CALIDAD DE LA SEÑAL DE VOZ PARA COMPENSACIÓN DE VARIABILIDAD INTER SESIÓN EN RECONOCIMIENTO DE LOCUTOR

Ingeniería de Telecomunicación

Ana García Muro
Septiembre 2012

UTILIZACIÓN DE MEDIDAS DE CALIDAD DE LA SEÑAL DE VOZ PARA COMPENSACIÓN DE VARIABILIDAD INTER SESIÓN EN RECONOCIMIENTO DE LOCUTOR

AUTOR: Ana García Muro
TUTOR: Daniel Ramos Castro

ATVS - Grupo de Reconocimiento Biométrico
Dpto. de Ingeniería Informática
Escuela Politécnica Superior
Universidad Autónoma de Madrid
Septiembre 2012

Resumen

Resumen

En este proyecto final de carrera se estudian distintos métodos de agrupamiento sobre ficheros de voz en función de la calidad de éstos de cara a su aplicación en sistemas de reconocimiento de locutor.

Los diferentes algoritmos de agrupamiento se prueban sobre dos bases de datos distintas, una perteneciente a la evaluación bianual del NIST, muy utilizada en estudios del estado del arte y, la otra, una base de datos de origen forense perteneciente a la Guardia Civil formada por conversaciones reales con una gran variabilidad de la calidad de la señal de voz.

Al inicio del trabajo se realiza un análisis del estado del arte actual de los sistemas de reconocimiento de locutor. Posteriormente se realiza un estudio sobre distintas medidas de calidad en reconocimiento biométrico, centrándose especialmente en reconocimiento de audio, y un estudio de diferentes métodos de agrupamiento. A continuación se explica la metodología empleada en la fase de experimentación y se realizan diversos análisis sobre las dos bases de datos empleadas, utilizando diversos algoritmos de clasificación. De cada experimento se extraen diversas gráficas y valores estadísticos que permiten realizar un análisis detallado de las agrupaciones generadas y se realiza un estudio del rendimiento del sistema una vez agrupados los ficheros que forman cada base de datos empleando para ello curvas DET.

Por último se presentan las conclusiones del proyecto y las líneas de trabajo futuro.

Palabras Clave

Sistemas ă deăreconocimiento ădeă locutor,ăăcalidad,ă indicador ădeă degradación,ăagrupamiento, K-means, GMM, curva DET

Abstract

In this Master Thesis different clustering techniques are studied with a focus on the quality for their application to speaker-recognition systems.

The different proposed clustering methods are tested in two different databases, one of them belongs to biannual workshop NIST, very referenced in state-of-art studies. The other data base belongs to Guardia Civil and it contains real speech with a very huge variability.

At ă the ă beginningă of ă this ă work an introduction to speaker recognition systems state-of-art is made. Then, a study over several quality measures in biometrics recognition, mainly relatives to speaker recognition systems, and a study about several clustering methods are developed. Then,ă the experimentsă methodologyă is ă presented and, two database experiments are made using different clustering algorithms. For each experiment, different graphs and statistics values are evaluated. The statics values allow to made a clustering detailed analysis. Also a performance analysis, using DET curves, is made depending on clustered voice files.

Finally ă the ă project ăconclusions ă areă drawn ă and ă future ă lines ă of ăworkă areă presented.

Key words

Speaker Recognition System, quality,ăclustering, K-means, GMM, DET Curve

Agradecimientos

En primer lugar de esta lista de agradecimientos no puede estar otra persona que mi tutor, Daniel Ramos. Dani, muchas gracias por ofrecerme la oportunidad de realizar el proyecto final de carrera contigo cuando te lo pedí. Por todo lo que he aprendido durante este tiempo y, sobre todo, por la gran paciencia que has tenido conmigo. No creo que otro tutor hubiera hecho las cosas mejor de lo que las has hecho tu. Muchas gracias ya que sin ti no habría sido posible.

En segundo lugar, al Laboratorio de Criminalística de la Guardia Civil, por darme la oportunidad de trabajar con ellos para conseguir la acreditación de su laboratorio. Por la confianza depositada en el trabajo realizado durante este proyecto y permitirme realizar las pruebas necesarias sobre la base de datos AhumadaIV-BaezaI, punto clave para el inicio de este proyecto final de carrera.

A mis amigas de toda la vida, Marisa y Guiomar, por estar siempre apoyándome fuera cual fuera la circunstancia. Gracias por todos los consejos, las charlas, los cines, las cenas, los viernes viendo películas, los cotilleos y marujeos mientras tomamos algo. Gracias por las tardes de instituto comiendo pipas sentadas en un banco solo para divertirnos después de estudiar, por los ratos de biblioteca que se hacían más amenos al estar juntas. Gracias sobre todo por todo lo que nos queda por vivir.

A los nuevos amigos, esos que llegaron en primero de carrera o que se fueron sumando con los años posteriores. A los amigos que se crearon en primer año, Davor por estar desde 0 o desde el instante -1; Miguel por su fuerza de voluntad y tesón; Héctor por sus reflexiones que siempre consiguen sacarnos una sonrisa a todos; Kiri por ser el primero y uno de los más entusiasta; Esther porque nos encontramos cuando estábamos las dos perdidas; Álvaro por todas las borderías y ratos de pasillo y Patri a la que por fin te encontré y tantas cosas hemos vivido. A los que fueron llegando poco a poco después, Sarasúa por sus ganas de hacer cosas; Moni por ser siempre un apoyo y sus trucos para las galletas; Miriam por todos los consejos que me has dado; Sandra por todas las charlas profundas que acaban en risas; Eva por los ratos de biblioteca, no cambies nunca; José por apoyarme siempre y decirme que puedo. Nos quedan las cañas de los jueves, las cenas y fiestas por mil y un motivos y un montón de cosas más por hacer juntos.

A mis compañeros de trabajo en Deimos, sin ellos todo sería mucho más aburrido. A Mónica .ºn-rails"por evitar que me convirtiera en una planta decorando el despacho, a Nuria por su buen rollo y sus charlas conspiranóicas, a Nuria por su genial guía de Londres y hacernos siempre de chica del tiempo, a Jose María por ser el único teleco que me queda en el trabajo :P, a J.Arevalo por sus ganas contagiosas de mejorar y sus charlas de cine, a Laura por su entusiasmo con el gimnasio aunque luego no vayamos, a todos los que hacéis "móvilesz conseguís que todo sea mucho más divertido. Gracias a Esther, por apoyarme siempre, enseñarme un montón de cosas y ser mi guía cuando estaba perdida. Y por último, Sandra, Javi y Sylvia por ser tan buenos jefes.

A David, por este tiempo juntos y lo mucho que nos queda. Por las tardes en la Fnac, en el cine o paseando por Madrid. Por dejarme hacer todas las fotos que quiera aunque te aburras o te canses. Por estar siempre pendiente de mí. Por apoyarme, decirme que yo puedo y hacerme reír siempre. Por conocerme tan bien y saber lo que pienso solo con mirarme. Por todo lo que

nos queda por vivir. Muchas gracias, sin ti terminar este proyecto habría sido más difícil.

Y por último, el agradecimiento más especial de todos, a mi familia. Por apoyarme siempre, decirme que yo puedo, no dejar que me rinda y levantarme cuando me he caído. A mi madre, porque sin ella no sería la persona que soy hoy en día, darme ánimos en los momentos duros y decirme "que las lágrimas son muy caras para tirarlas por cualquier tontería". A mi padre por enseñarme un millón de cosas, siempre buscando una respuesta a las múltiples preguntas y hacerme descubrir mil mundos maravillosos en cada libro. Os quiero mucho a los dos y sé que sin vosotros no estaría aquí.

Índice general

Índice de figuras

Índice de cuadros

1 Introducción

Durante los últimos años los sistemas de verificación y de identificación de personas de forma automática han experimentado una gran mejora. Este gran avance ha propiciado su expansión y su aceptación en todo tipo de aplicaciones (seguridad, comercio electrónico, sistemas forenses, etc.).

Dentro de los sistemas automáticos, uno de los que más ha evolucionado ha sido el sistema de reconocimiento de locutor. Este sistema presenta varias ventajas frente a otros: la fácil adquisición de la muestra biométrica y la dificultad a la hora de falsear una identidad. Además, supone una forma más segura de identificación ya que no requiere el uso de tarjetas o claves que pueden ser olvidadas o perdidas.

Como una definición amplia del término, un sistema de reconocimiento de locutor es aquel que, comparando la medida de similitud frente a un modelo estadístico, pretende establecer la pertenencia o no pertenencia de un fichero de audio a un sujeto bajo estudio [1]. La medida de similitud empleada se denomina puntuación (score) y está directamente relacionada con la tasa de rendimiento del sistema.

Las principales ventajas del empleo de sistemas de verificación o autenticación de identidades por medio de reconocedores automáticos de locutor es la dificultad a la hora de falsear la identidad y la facilidad para obtener la muestra biométrica (obtención por medio de teléfonos o micrófonos, siendo una forma no intrusiva para el usuario). Sin embargo, a día de hoy todavía es necesario realizar un esfuerzo a la hora de diseñar e implementar sistemas más efectivos. Esto se debe, entre otra serie de factores, a que los rasgos biométricos se ven limitados por diferentes situaciones que modifican la muestra biométrica introduciendo variabilidad. La existencia de esta variabilidad afecta a la calidad de la muestra y, por tanto, al rendimiento del sistema. Dentro de los tipos de variabilidad que afectan a los archivos de audio se pueden encontrar las siguientes clases:

- Variabilidad inter-sesión: se debe a una interacción incorrecta con el sensor. Dentro de este tipo de variabilidad encontramos el uso de distintos micrófonos para registrar la voz del mismo usuario, entornos de grabación diferentes o, incluso, el cambio de edad del sujeto bajo estudio. Un ejemplo puede ser las condiciones presentes en la red telefónica (en grabaciones de tipo GSM).
- Variabilidad intra-sesión: a diferencia de la variabilidad inter-sesión, la variabilidad intra-

sesión hace referencia a la distorsión introducida por el usuario durante una misma sesión de captura. Un ejemplo de esta variabilidad puede ser las voces de otros usuarios en la grabación de audio (a menudo está presente en grabaciones de tipo ïnterview"de las que se hablará en próximos capítulos de este proyecto final de carrera).

Existen diversos métodos cuyo objetivo es compensar esta variabilidad y así obtener una información lo más precisa a la fuente. De estos métodos (como por ejemplo, compensación de canal o Factor Analysis [2] se hablará en los siguientes capítulos.

Este proyecto final de carrera se basa principalmente en variabilidad inter-sesión mostrándose un gran número de experimentos basados en diferentes métodos de obtención de muestras y cómo afecta este hecho a la calidad de la muestra biométrica (en el caso de un reconocimiento automático de locutor conviene recordar que se trata de ficheros de audio) y por tanto al rendimiento del sistema.

El instituto americano NIST(National Institute of Standards and Technology) realiza un workshop bianual sobre la calidad [3] en el que se muestran los últimos avances en la materia. Dicho workshop consiste en una serie de pruebas sobre un escenario común para todos los grupos participantes (existen diversas pruebas, pero siempre existe un núcleo común, denominado *core*). Durante años el grupo ATVS de la Universidad Autónoma de Madrid ha formado parte de esta lista de participantes .

Durante el desarrollo de este proyecto se ha empleado una base de datos proporcionada por el NIST (en concreto, la base de datos correspondiente a la evaluación del año 2008). Sobre la base de datos NIST2008 se hará un breve análisis, indicando número total de ficheros, tipos de ficheros y demás características propias. Además de la base de datos NIST2008 y con la finalidad de no limitar el estudio presente en este proyecto, también se ha utilizado una base de grabaciones de audio de la Guardia Civil. A esta base de datos se tuvo acceso gracias al acuerdo de colaboración existente entre dicho cuerpo de seguridad y el grupo ATVS.

1.1. Motivación del proyecto

La Guardia Civil y, más concretamente, el laboratorio forense de la Guardía Civil, poseen una serie de bases de datos con información biométrica de distintos usuarios. Una de estas bases de datos está formada por grabaciones de audio de distintos locutores en formato ïnterview". A partir de la necesidad de la Guardia Civil de obtener una serie de certificaciones sobre su base de datos se hace necesaria el contar con una herramienta que permita la clasificación de sus ficheros de audio.

De esta forma, con la Guardia Civil se ha desarrolla una colaboración en las primeras etapas de este proyecto final de carrera. Esta colaboración consistía en proporcionar un método capaz de clasificar los ficheros de audio en función de la calidad de las muestra biométricas recogidas en ellos. Para ello es necesario contar con las herramientas adecuadas para medir la calidad de los ficheros de audio así como sistemas de análisis de los datos extraídos.

Con una análisis inicial de los ficheros de audio se pueden llevar a cabo enfrentamientos más óptimos (por ejemplo, suprimiendo a priori una serie de datos que no van a aportar información suficiente debido a diferencias de sistema de grabación, procedencia del hablante, etc) entre diferentes elementos. A su vez también se persigue obtener una representación gráfica de las diferentes agrupaciones de ficheros de audio en función de su sistema de grabación.

Sobre las particularidades de la base de datos de la Guardia Civil, llamada AhumadaIV-BaezaI, se hablará en los capítulos siguientes en los que se aportará una descripción sobre que

clase de ficheros están presentes, que tipo de hablantes se pueden encontrar, así como otra serie de parámetros de interés.

1.2. Objetivos y enfoque

El objetivo de este proyecto es realizar un análisis de varios métodos de agrupación de ficheros de audio en base a su calidad y cómo afecta esta agrupación al rendimiento del sistema. Se pretende obtener la mayor cantidad de información posible de forma que sirva de base para futuros experimentos de agrupación y de compensación de calidad. Este objetivo se desglosa en los siguientes puntos:

1. Estudio del estado del arte. Recopilar las últimas tecnologías en Sistemas de reconocimiento de locutor (SRL), métodos utilizados para estimar la calidad de la señal de voz y sistemas de agrupación de ficheros.
2. Implementación de las agrupaciones. Tomando como base las medidas de calidad propuestas por Alberto Harriero en su proyecto final de carrera implementar métodos de clasificación de ficheros de audio que permitan obtener diferentes grupos en función de la calidad de éstos.
3. Implementar un sistema por el cual sea factible relacionar las agrupaciones obtenidas anteriormente con el rendimiento del sistema por medio de curvas DET.
4. Evaluar las agrupaciones propuestas basándose en medidas estadísticas como la entropía y la pureza de una agrupación.

Para cumplir estos objetivos se han realizado una serie de experimentos. Las características principales del planteamiento experimental son los siguientes:

- Distintas bases de datos. Para el desarrollo de este proyecto final de carrera se han utilizado dos bases de datos diferentes con la finalidad de poder realizar una comparativa entre los resultados obtenidos a partir éstas. La base de datos NIST2008 está proporcionada por el NIST [NIST SRE 2008]. Esta base de datos contiene locuciones tanto de tipo microfónico (sin transmitir por una red telefónica) así como locuciones telefónicas. La segunda base de datos forense empleada es la proporcionada por la Guardia Civil, en el contexto del acuerdo de colaboración que el cuerpo de seguridad mantiene con el grupo ATVS, llamada AhumadaIV-BaezaI. Esta base de datos también está formada por grabaciones microfónicas así como por habla telefónica, lo que facilitará la tarea posterior de análisis.
- Distintos sistemas. Se probará como de bueno es el agrupamiento empleando distintos métodos de clustering. El primero de los métodos empleados es el algoritmo K-means, sobre el que se realizan diversos experimentos al modificar el tipo de distancia empleada a la hora de calcular la pertenencia a un grupo. El segundo método empleado es el algoritmo GMM (*Gaussian Mixture Models*). Este algoritmo se basa en asemejar la distribución de probabilidad a una mezcla de gaussianas y clasificar los datos en función de la pertenencia a una u otra gaussiana.
- Distintos sistemas de análisis del rendimiento del sistema. El rendimiento del sistema se compone de varias fases: comenzando por una inspección visual sobre como de bueno ha sido el agrupamiento (fronteras definidas entre distintos grupos, grupos bien posicionados respecto a los datos iniciales, etc), realizando un análisis basado en la entropía y en la pureza de la agrupación, y, por último evaluando el rendimiento del sistema por medio de curvas DET.

1.3. Contribuciones del PFC

A continuación se citan las principales contribuciones que aporta este trabajo:

- **Análisis de la base de datos AhumadaIV-BaezaI**: Se presenta un análisis de la base de datos de la Guardia Civil obteniendo los indicadores de degradación. Estos indicadores de degradación serán presentados en la sección 3, así como el estudio pormenorizado de esta base de datos en la sección 4.
- **Análisis de agrupamiento**: En este proyecto se presentan distintas formas de agrupamiento de ficheros de audio basándose en los parámetros de calidad de éstos. Se presentan dos agrupaciones distintas, una por algoritmo K-means y otra mediante Modelo de mezcla de gaussianas (GMM).
- **Estudios experimentales**:
  - Estudio de agrupamiento por medio de K-means.
  - Estudio de agrupamiento por medio de GMM.
- **Evaluación del rendimiento**:
  - Datos de entropía y pureza de las agrupaciones.
  - Curvas DET de los agrupamientos realizados.

2

Sistemas de reconocimiento de locutor

2.1. Introducción a los sistemas biométricos

La biometría es la disciplina encargada del estudio de los métodos automáticos empleados para el análisis de rasgos personales para identificar o verificar la identidad de una persona. Para conseguir la identificación de una persona se debe recurrir a una serie de parámetros únicos de esa persona e inmutables en el tiempo, este tipo de parámetros son los rasgos biométricos.

2.1.1. Rasgos biométricos

Los rasgos biométricos son aquellas características que contienen información sobre la identidad de la persona bajo estudio. Según [4] se pueden encontrar los siguientes tipos:

- Rasgos físicos: aquellos que dependen del físico de cada persona. Dentro de los rasgos físicos podemos encontrar:
  - Huellas dactilares: patrón de crestas y valles en el dedo de un sujeto. Es un rasgo característico de cada persona aunque variable en el tiempo (edad del sujeto, heridas, etc).
  - Geometría de la mano: número de medidas tomadas de la mano del individuo. Estas medidas incluyen forma, tamaño de la mano, largo y ancho de los dedos, etc. Es un rasgo variable en el tiempo y, a su vez, no se pueden extraer características de forma fiable debido al uso de anillos o de enfermedades (por ejemplo, artritis).
  - Reconocimiento de iris: basado en la toma de imágenes del iris. Necesita un alto grado de cooperación con el sujeto bajo estudio debido a los sistemas de adquisición.
  - ADN: Código uni-dimensional único para cada persona (excepto gemelos idénticos, ya que éstos comparten ADN). Principalmente se emplea en análisis forenses.
  - Cara: método no intrusivo. Las imágenes faciales son, probablemente, la característica biométrica más empleada por el ser humano para realizar reconocimiento de personas.
  - Oreja: se basa en medidas de distancia entre diferentes puntos del pabellón auditivo respecto a una posición prefijada en el oído.

  - Voz: las características de la voz se basan en la forma y el tamaño del aparato vocálico.

- Rasgos de comportamiento: aquellos que el individuo adquiere a lo largo de su vida basados en su nivel de educación, su entorno, su comportamiento. Los rasgos de comportamiento más empleados en sistemas de biometría son:

  - Firma: rasgo biométrico ampliamente aceptado en transacciones comerciales, legales y gubernamentales. Cambia a lo largo del tiempo debido a la condición física del usuario así como a su estado anímico.
  - Dinámica del tecleo: este comportamiento biométrico no se espera que sea único para cada individuo, sin embargo, en aplicaciones de verificación de identidad puede ser lo suficientemente discriminativo como para ser considerado una buena alternativa en sistemas de verificación con pocos usuarios.
  - Forma de caminar: el movimiento de una persona al caminar es lo suficientemente significativo como para realizar verificaciones en conjuntos cerrados de población. Cambios de peso o heridas importantes (afectando a las articulaciones o, incluso, al cerebro) hacen variar la muestra biométrica dando como resultado errores en la identificación.
  - Voz: como rasgo de comportamiento, la voz presenta variaciones en función del nivel socio-lingüístico del hablante o de su estado de ánimo.

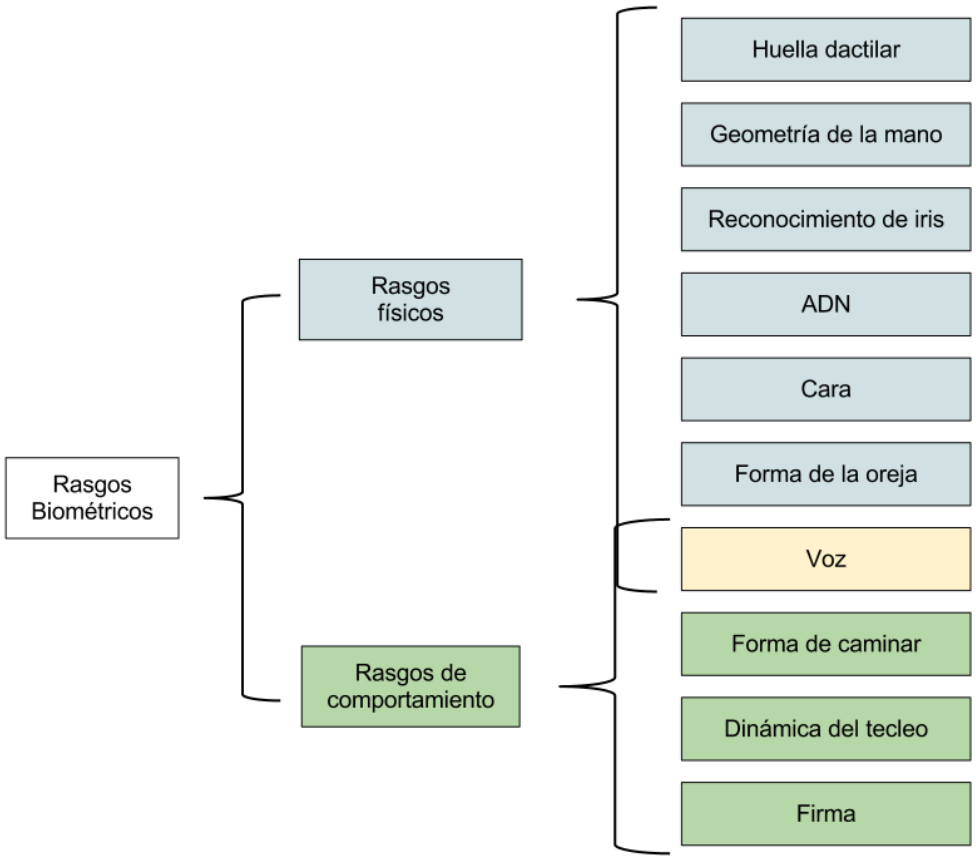


Figura 2.1: Tipos de rasgos biométricos

Como se puede ver en [4] existe una serie de requisitos que aseguran el correcto funcionamiento del sistema automático de reconocimiento biométrico.

- Universalidad: El rasgo biométrico debe ser universal, es decir, toda la población debe poseerlo.
- Distintividad: Debe ser lo suficientemente diferente de una persona a otra para ser capaz de distinguir entre ambas.
- Estabilidad: El rasgo debe ser estable. La estabilidad implica que su variabilidad temporal tiene que ser lo suficientemente pequeña como para que el cambio que se pueda producir no sea significativo.
- Evaluabilidad: Debe poder ser cuantificado de forma que se puedan realizar medidas sobre el rasgo.

2.1.2. Sistemas de reconocimiento

El sistema de reconocimiento biométrico funciona siguiendo un esquema parecido al de un clasificador de patrones. La señal de entrada del clasificador de patrones en este caso es un rasgo biométrico del sujeto a identificar. En la figura 2.2 se puede observar el esquema básico del sistema.

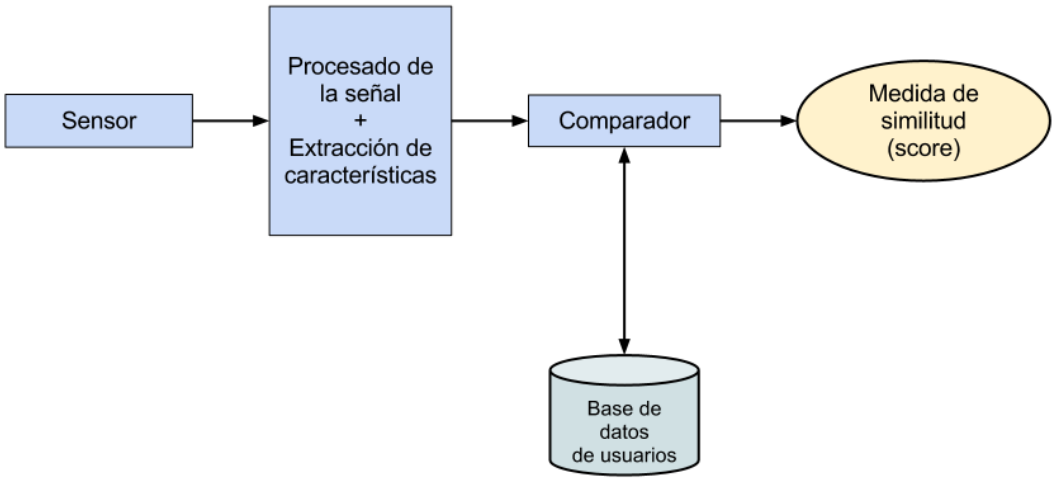


Figura 2.2: Funcionamiento básico de un sistema biométrico

Desde una vista de alto nivel se puede observar que el bloque inicial es un sensor encargado de obtener el rasgo biométrico del sujeto (voz por medio de un micrófono, huella dactilar con un lector de huellas, iris obtenido de un cámara especial, etc). Posteriormente esta señal se procesa (empleando técnicas de tratamiento digital de imágenes y de audio) para extraer características que contengan información de la identidad. Estos parámetros son los que contienen la información biométrica del usuario.

Los sistemas automáticos de reconocimiento biométrico pueden funcionar con dos modos principales, identificación y verificación, más un modo previo denominado registro.

- Registro: Modo de funcionamiento previo. En él se adquieren las muestras biométricas necesarias para el correcto funcionamiento del reconocedor. Éstas se obtienen por medio

de un sensor, se evalúa la calidad de la captura, se extraen los parámetros necesarios y se almacenan, junto con la identidad del sujeto, en una base de datos. El esquema de un reconocedor en modo registro se puede ver en la figura 2.3.

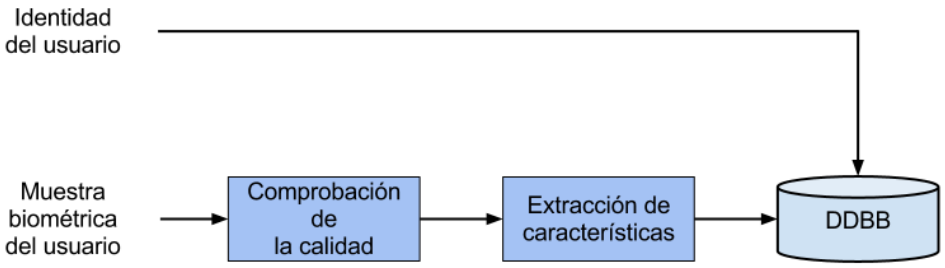


Figura 2.3: Sistema en modo registro

- Identificación: El objetivo de un sistema de identificación es determinar a partir de un grupo de voces conocidas cuál es la que coincide mejor con la muestra de voz suministrada. En ésta puede haber dos conjuntos distintos.
  - Conjunto cerrado: el individuo siempre pertenece a alguna de las clases incluidas en la base de datos.
  - Conjunto abierto: el usuario puede no pertenecer a alguna de las categorías definidas anteriormente.

  La figura 2.4 muestra el modo identificación.

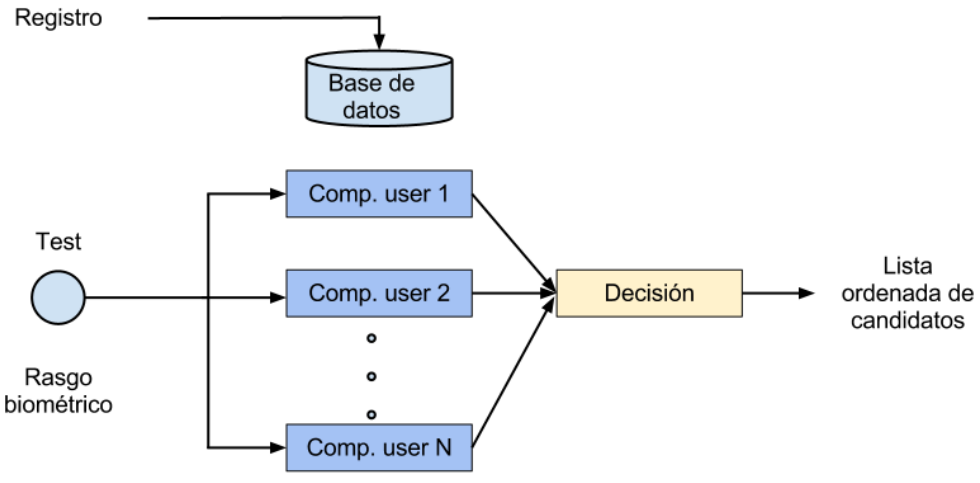


Figura 2.4: Sistema en modo identificación

- Verificación: El sistema comprueba que el usuario posee la identidad que dice poseer. Se realizan dos comparaciones: una de la muestra biométrica con el modelo correspondiente

a la identidad que el usuario dice poseer y otra comparación con un modelo representante de una población de interés. La figura 2.5 muestra un esquema de este sistema.

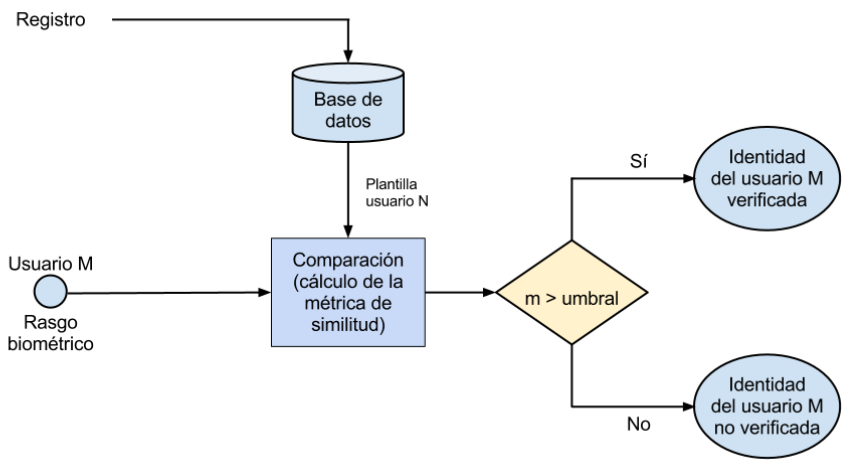


Figura 2.5: Sistema en modo verificación

Como se ha podido observar anteriormente, la voz como rasgo biométrico tiene características tanto físicas como de comportamiento. Por este motivo es muy versátil a la hora de servir de método de identificación/autenticación de personas.

2.2. La señal de voz

En este apartado se comentarán brevemente los aspectos más relevantes de la señal de voz de cara a su uso como rasgo biométrico.

2.2.1. Características

El empleo de la voz como sistema de reconocimiento tiene una gran importancia, incluso por encima de otros más fiables. Esto se debe a dos factores fundamentales:

- Aceptación: al ser la voz un acto natural, es un rasgo biométrico altamente aceptado ya que no supone desconfianza. Además su adquisición es un proceso fácil y natural que no necesita una alta colaboración ni es un método intrusivo para el sujeto bajo estudio.
- Gran cantidad de datos: gracias a la red telefónica se consigue acceder a estos rasgos desde cualquier parte del mundo.

Sin embargo, existen otros factores que penalizan el rendimiento de la voz como rasgo biométrico:

- Baja distintividad: la señal de voz es un rasgo dependiente del comportamiento por lo tanto factores externos e internos al sujeto pueden afectar a su calidad (ambiente ruidoso, estado físico y anímico del sujeto, etc).

- Baja estabilidad: la voz varía con el tiempo (con la edad del sujeto) lo que dificulta su uso en sistemas de reconocimiento de locutor. A su vez presenta variaciones inter-sesión (grabaciones obtenidas durante días diferentes) e intra-sesión (grabaciones obtenidas durante una misma sesión).

El ser humano es capaz de distinguir personas por su voz. En los sistemas de reconocimiento automático de locutor se ha implementado un sistema similar al humano, extrayendo una serie de parámetros característicos con los que realizar el reconocimiento. Como se observa en [5] estos niveles presentes en la señal de voz son los siguientes:

- Espectral: Representa el nivel más bajo de la señal de voz. La información se obtiene del espectro de la señal, enventanando ésta de tal forma que el tiempo de muestra sea tan pequeño que la voz se comporte como una señal estacionaria.
- Fonético: Basándose en la pronunciación de los diferentes fonemas es posible extraer parámetros que caractericen al sujeto.
- Prosódico: Es el estudio de los elementos relacionados con el tono y la entonación.
- Otros: léxico (palabras empleadas por el sujeto), semántico (significado), sociolingüísticos(características sociales de la persona), etc.

Los niveles citados anteriormente se muestran en formato de gráfica en la figura 2.6.

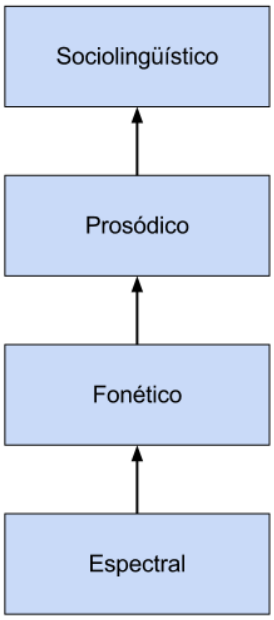


Figura 2.6: Niveles presentes en la señal de voz

En este proyecto final de carrera nos centraremos en las características de la voz en el nivel espectral, obteniendo de la señal enventanada los parámetros necesarios para su estudio.

2.2.2. Proceso de fomación

La señal de voz se forma en el tracto vocal de cada persona. El tracto vocal es diferente de un individuo a otro por lo que sirve para distinguir entre diferentes sujetos.

El tracto vocal esta formado por tres cavidades resonantes:

- Cavidad faríngea: situada detrás de la laringe.
- Cavidad oral: formada por la lengua, los dientes, los labios y el paladar.
- Cavidad nasal: situada entre el velo del paladar y los orificios nasales.

El tracto vocal y las cavidades que lo forman se puede observar en la figura 2.7.

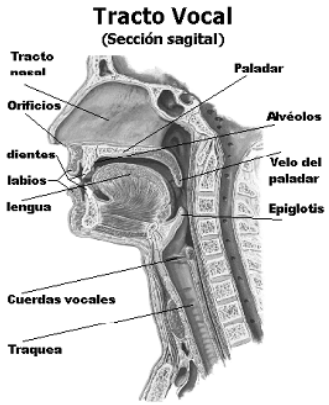


Figura 2.7: Tracto vocal con los principales órganos que lo forman

La señal hablada se produce al expelir el aire almacenado en los pulmones a través de la traquea y atravesar éste las cuerdas vocales. Esta señal puede presentar tramos periódicos o no periódicos, creándose distintos tipos de sonido:

- Tramo sonoro: de carácter periódico. En este caso las cuerdas vocales se hallan en tensión y vibran al atravesarlas el flujo de aire proveniente de los pulmones.
- Tramos sordos: son sonidos de carácter ruidoso. Se producen al estar las cuerdas vocales en relajación y el aire las atraviesa libremente. El sonido se produce por una turbulencia de aire.

En la producción de los sonidos hablados (tramo sonoro) la señal de excitación atraviesa las cavidades citadas anteriormente produciéndose determinadas frecuencias resonantes. A estas frecuencias se le denomina formantes (máximos relativos en la envolvente espectral de la señal de voz) y son una parte muy importante en el reconocimiento automático de locutores.

Los fonemas son la unidad mínima del lenguaje oral (sonidos que permiten diferenciar entre las palabras de una lengua determinada) y sus características dependen sobre todo de estos resonadores acústicos citados anteriormente y de la forma de las cavidades. Los fonemas se pueden organizar de la siguiente forma:

- Fonativo: tramo sonoro. Es generado cuando el flujo de aire se modula por las cuerdas vocales. Al ser una señal periódica, los armónicos (múltiplos enteros de la frecuencia fundamental) aparecen configurando la estructura fina del espectro.

- Fricativos: tramos sordos. Se generan por constricciones del tracto vocal. El lugar, la forma y el grado de constricción determina la forma de la excitación ruidosa.
- Consonantes oclusivas: formadas por un tiempo de cierre y, a continuación, un tiempo de explosión. Se pueden observar tanto tramos sonoros como sordos.

2.3. Reconocimiento de locutores

Una vez explicado el proceso de formación de la voz es hora de hablar como se emplean las características de esta señal para identificar personas. Esa es la tarea principal de un sistema de reconocimiento de locutor.

2.3.1. Tipos de reconocimiento

Como ya se ha comentado anteriormente los tipos de aplicaciones en un sistema de reconocimiento biométrico se pueden clasificar como verificación o identificación.

Sin embargo, como se muestra en [1], en el campo del reconocimiento de locutores se puede hacer otra clasificación en función del campo de aplicación:

- Dependiente de texto: En este sistema el usuario debe pronunciar una serie de palabras o de frases conocidas de antemano. Habitualmente las palabras consisten en un código PIN reforzando la seguridad. Este tipo de sistemas se aplican para control de acceso, por ejemplo en aplicaciones de banca telefónica (el usuario debe identificarse con un número PIN o con una serie de frases al azar pero ya entrenadas) o de acceso seguro a determinadas instalaciones.
- Independiente de texto: Es el método más habitual en los sistemas automáticos de reconocimiento de locutor. A diferencia de las aplicaciones dependientes del texto, el sistema desconoce lo qué dirá el usuario.

En el desarrollo de este proyecto nos basaremos en los sistemas independientes de locutor relacionados con las aplicaciones forenses debido a la colaboración con la Guardia Civil que el grupo ATVS lleva realizando desde hace años. En los siguientes apartados se detallarán los aspectos básicos de funcionamiento de un sistema de reconocimiento de locutor.

2.3.2. Funcionamiento básico

El funcionamiento básico de un sistema de reconocimiento automático de locutor se puede dividir en tres funciones básicas: adquisición de los datos, extracción de parámetros principales y obtención de resultados.

El proceso de adquisición de los datos se realiza por medio de micrófonos o de teléfonos, de tal forma que no será detallado en este proyecto, se partirá de la versión digitalizada de la señal.

Extracción de parámetros

La extracción de parámetros consiste en transformar la señal digitalizada de voz en un conjunto de vectores de características. Con este procedimiento la señal queda representada de forma

más compacta y menos redundante lo que favorece el resto de las etapas, como el modelado estadístico o la toma de decisión.

La mayor parte de los sistemas de reconocimiento de locutor se basan en una análisis homomórfico en el dominio cepstral. El dominio cepstral es un dominio temporal en el que la unidad básica es el cepstrum (coeficiente cepstral). Este conjunto de coeficientes representa la señal de voz , siendo los primeros coeficientes los relacionados con la envolvente de la señal, mientras que los más altos ofrecen una estimación de la estructura fina de la señal, de la cual se puede extraer información de la frecuencia fundamental.

Existen dos formas comunes de obtener los coeficientes cepstrales:

- MFCC: Coeficientes Cepstrales en Frecuencia Mel
- LPCC: Coeficientes de Predicción Lineal

A continuación se describen brevemente estos dos sistemas de obtención de los coeficientes cepstrales de una señal de voz:

MFCC: Coeficientes Cepstrales en Frecuencia Mel

En la figura 2.8 se puede observar una representación modular de una parametrización por medio de un banco de filtros, método empleado en el sistema MFCC.

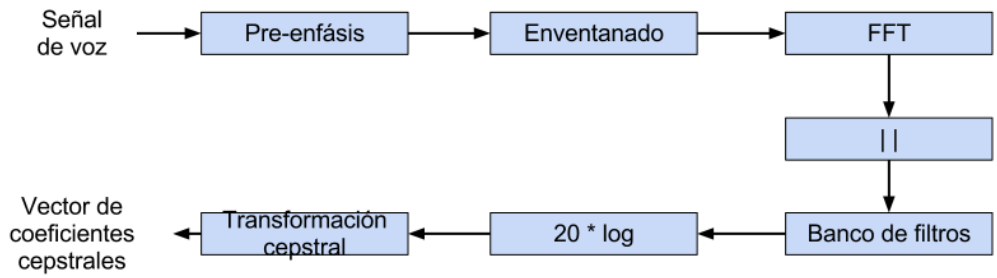


Figura 2.8: Parametrización cepstral por medio de un banco de filtros

Sobre la señal digitalizada, se aplica un filtro de preenfásis con la finalidad de aumentar las altas frecuencias del espectro. La señal preenfatizada se obtiene tras aplicar el siguiente filtro:

$$x_p(t) = x(t) - ax(t-1)$$

donde a es la amplitud del filtro y su valor oscila entre $0{,}95 < a < 0{,}98$. Esta etapa de preenfásis no siempre es necesaria.

La siguiente etapa consiste en el enventanado de la señal. El proceso de enventanado se basa en desplazar desde el principio hasta el final de la señal de voz una ventana cuya duración temporal es menor que la de la señal entera. Las ventanas generalmente tienen una duración entre 20 ó 30 milisegundos de esta forma se consigue que la señal de voz en el tramo enventanado se

comporte como una señal estacionaria. Es necesario recordar que un enventanado en el dominio temporal produce, en el dominio espectral, una convolución entre la transformada de la señal y la transformada de Fourier de la ventana. Para minimizar el efecto de la convolución se han de emplear ventanas con el lóbulo principal estrecho y lóbulos secundarios pequeños. Una solución de compromiso entre estos requisitos consiste en emplear una ventana tipo Hamming (ponderación tipo coseno alzado) en la que las muestras de los extremos se encuentran ponderadas quedando las muestras de la señal enventanada minimizadas en los extremos. Para solucionar este problema, las ventanas se solapan para que las muestras en los extremos de una ventana sean las centrales en las ventanas consecutivas. Generalmente el tiempo de solape suele ser de unos 10 milisegundos.

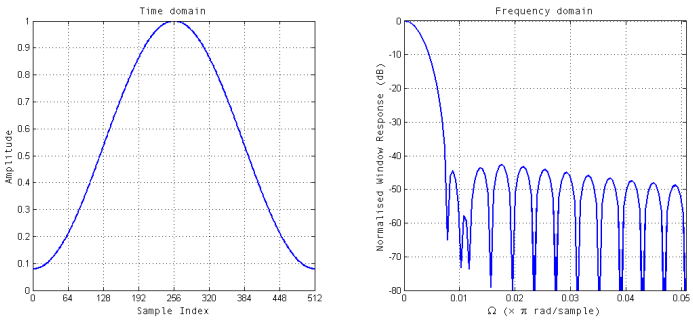


Figura 2.9: Ventana de Hamming en el dominio temporal y en el dominio frecuencial.

Una vez obtenida la señal enventanada se aplica una transformada rápida de Fourier (FFT). El número de puntos para calcular la FFT se fija en un número potencia de dos, habitualmente se emplean 512 puntos.

El espectro de la señal obtenido presenta muchas fluctuaciones y un gran nivel de detalle, sin embargo, desde el punto de vista de la extracción de parámetros es necesario la envolvente de la señal. Para obtener la envolvente del espectro se aplica un banco de filtros. Un banco de filtros es una serie de filtros paso banda que multiplican uno a uno el espectro para obtener un valor medio en la banda de frecuencias del filtro correspondiente. En este caso hablaremos del filtrado MEL. La escala MEL es una escala logarítmica similar a la escala de frecuencias del oído humano. Para construirla se equipara un tono de 1000Hz y a 40 dBs por encima del umbral de audición con un tono de 1000 mels. Sobre los 500Hz, los intervalos de frecuencia espaciados exponencialmente son percibidos como si estuvieran espaciados linealmente. Como consecuencia, cuatro octavas en escala lineal se comprimen a unas dos octavas en escala mel.

$$f_{MEL} = 1000 \cdot \frac{\log(1 + f_{LIN}/1000)}{\log 2}$$

El banco de filtros se crea por medio de filtros triangulares situados de forma que la frecuencia central de cada filtro siga la escala Mel. A su vez los bordes de los filtros coinciden con la frecuencia central de los filtros adyacentes. Gráficamente un banco de filtros Mel tiene la forma de la Figura 2.10.

En [5] se describe el proceso final de obtención de los MFCCs. Una vez filtrado el espectro de la señal, se obtiene la señal filtrada, denotada $Y(m), m = 1, ..., M$. Los coeficientes MFCC, C_n, se obtienen de la siguiente forma:

$$c_n = \sum_{m=1}^{M} [logY(m)] \, cos \left[\frac{\Pi n}{M} (n - \frac{1}{2}) \right],$$

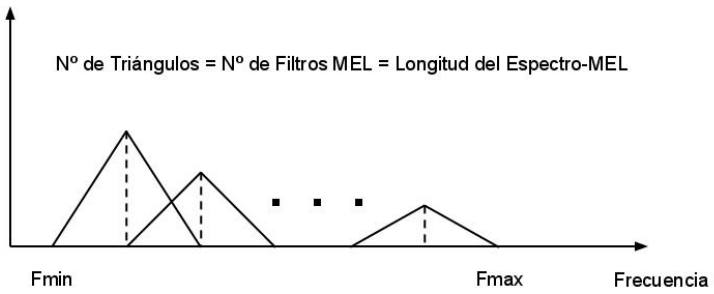


Figura 2.10: Banco de filtros Mel

donde n es el índice del coeficiente cepstral. El vector final de coeficientes MFCC se obtiene de los coeficientes más bajos, hasta el orden 12 ó 15.

LPC: Coeficientes de Predicción Lineal

Siguiendo con el estudio realizado por [5], la predicción lineal es una alternativa al cálculo de coeficientes MFCC, ya que ofrece una interpretación muy intuitiva tanto en el dominio frecuencial (los polos del espectro se corresponden con la estructura de resonancia) como en el dominio temporal (las muestras adyacentes están correladas).

En el dominio del tiempo, la ecuación de un sistema de predicción lineal es la siguiente:

$$\tilde{x}[n] = \sum_{k=1}^{p} a_k s[n-k]$$

donde $s[n]$ es la señal observada, a_k son los coeficientes de predicción y $\tilde{s}[n]$ es la señal predecida. El error de predicción (residual) se calcula como $e[n] = s[n] - \tilde{s}[n]$. Los coeficientes a_k se calculan minimizando la energía de la señal de error (algoritmo *Levinson-Durbin*).

El modelo espectral se define como:

$$H(z) = \frac{1}{1 - \sum_{k=1}^{p} a_k z^{-k}},$$

y consiste únicamente en picos y polos espectrales.

Los coeficientes de predicción lineal, a_k, raramente son empleados como vector de características, pero trasladando estos coeficientes al dominio cepstral se obtienen los LPCCs (*Linear Predictive Cepstral Coefficients*).

El esquema básico de un sistema LPC se muestra en la figura 2.11.

Al extraer la envolvente LPC se consigue obtener la estructura formántica de la señal hablada. Con una representación correlativa (en cascada) se consigue visualizar la frecuencia de la señal en una forma alternativa a los espectrogramas como se observa en la figura 2.12.

Una comparativa de las dos formas de obtención de parámetros (MFCCs, LPCCs) se encuentra en la figura 2.13. Esta figura está extraída de [5] y muestra como un espectro de la señal calculado por medio de FFT (con 512 puntos) puede ser reducido a, tan solo, 12 coeficientes MFCCs o 12 LPCCs.

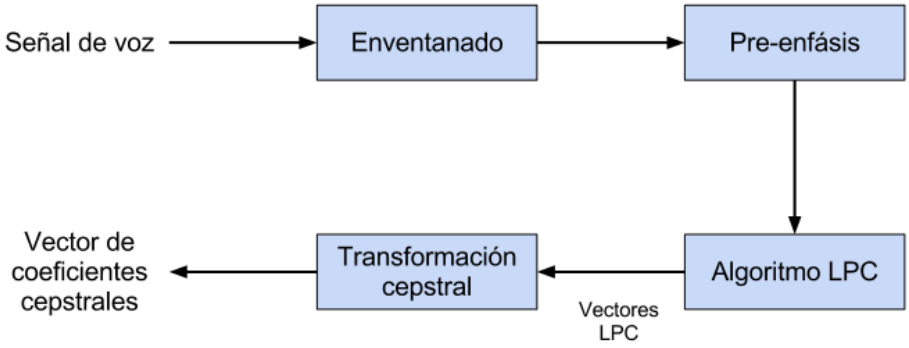


Figura 2.11: Parametrización LPC

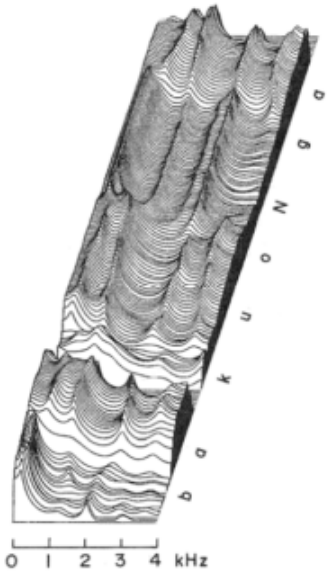


Figura 2.12: Estructura formántica de la señal de voz obtenida por LPC

2.3.3. Modelado y cálculo de medidas de similitud

Una vez obtenidos los vectores cepstrales que caracterizan la señal de voz se procede a realizar un entrenamiento de los modelos y, a continuación, calcular el score de similitud.

Según el trabajo desarrollado por [6] el sistema GMM se basa en un cálculo de la verosimilitud por medio de un modelado estadístico de los parámetros cepstrales. Dado un segmento de voz, Y, y un locutor hipotético S, la tarea de un reconocedor de audio es determinar si Y pertenece al locutor S. Se asume que Y contiene voz de un único locutor.

Para ello se realiza una comprobación sobre dos simples hipótesis:

H_0: Y pertenece al locutor S
H_1: Y no pertenece a S

La relación de verosimilitud (*Likelihood Ratio*) es dado por:

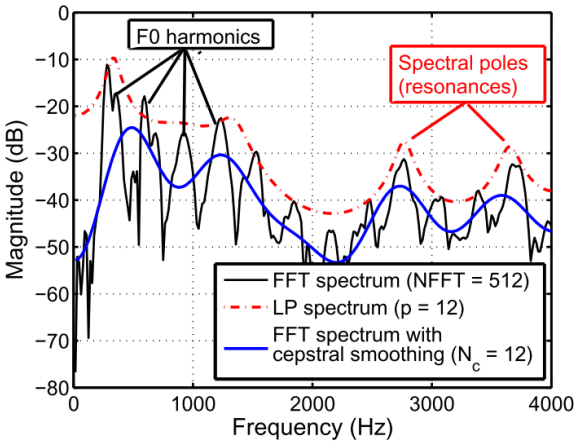


Figura 2.13: Comparativa entre MFFCs y LPCCs elaborada por [5]

$$\frac{p(Y|H_0)}{p(Y|H_1)} \begin{cases} \geq \theta & \text{aceptar} H_0 \\ < \theta & \text{rechazar} H_0 \end{cases}$$

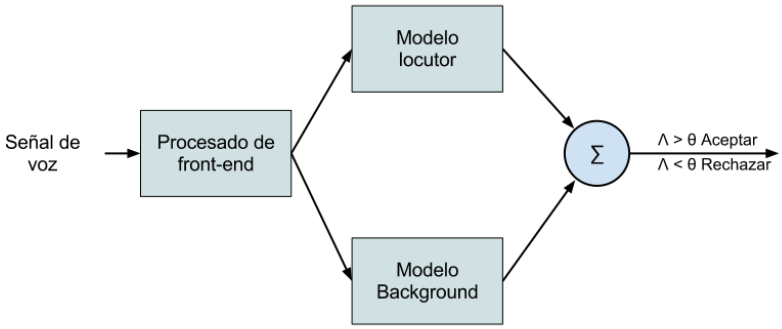


Figura 2.14: Sistema de verificación basado en análisis probabilístico

Continuando con el trabajo desarrollado por Reynolds en [6], matemáticamente H_0 es representado por un modelo denotado λ_{hyp} que caracteriza al locutor S en el espacio de características x. La hipótesis alternativa, H_1, es representada como $\lambda_{\overline{hyp}}$. Siendo X una secuencia de vectores características extraídos de la locución de test, $X = \{x_1, ..., x_T\}$ donde x_t es un vector características en el tiempo $t \in [1, 2, ..., T]$, de esta forma, el cálculo de verosimilitud queda definido como $p(X|\lambda_{hyp})/p((X|\lambda_{\overline{hyp}}))$.

Mientras que el modelo del locutor está bien definido, con el modelo de la hipótesis alternativa no sucede lo mismo. Este modelo, $h_{\overline{hyp}}$, deber representar el espacio formado por todas las posibles alternativas al locutor bajo estudio. Un método tradicional para estimar el modelo $h_{\overline{hyp}}$ consiste en utilizar un conjunto de otros locutores de background. A este conjunto de locutores se le denomina cohorte. Dado un conjunto de N modelos de locutor $\{\lambda_1, \lambda_2, ..., \lambda_N\}$, el modelo de la hipótesis H_1 es:

$$p(X|\lambda\overline{hyp}) = F(p(X|\lambda_1), ..., p(X|\lambda_N))$$

donde $F()$ es alguna función, como media o máximo, de los valores de similitud del conjunto de locutores de background.

GMM: Gaussian Mixture Models

Según [1] los Modelos de Mezclas Gaussianas consisten en el modelado del habla a partir de mezclas de distribuciones gaussianas basándose en que la distribución de los coeficientes (MFCC o LPCC) se aproxima a la de una mezcla de gaussianas. De esta forma se definen dos modelos: λ_t es el modelo del locutor a comprobar y λ_{UBM} es el modelo del habla universal (Background Universal Model). Este modelo UBM se entrena con bases de datos que representan la población bajo estudio mientras que el modelo del usuario se obtiene adaptando el UBM a los parámetros extraídos de las locuciones de entrenamiento.

La función de densidad de probabilidad de los modelos se obtiene por:

$$f\left(x|\lambda\right) = \sum_{i=1}^{M} w_i f_i(x)$$

donde x representa un vector de características, w_i es el peso de cada una de las M gaussianas $f_i(x)$ que a su vez se descompone en:

$$f_i(x) = \frac{1}{2\Pi^{\frac{d}{2}}|\sum_i|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_i)^T(\sum i)^{-1}(x-\mu_i)}$$

donde μ_i representa la media y $\sum i$ la matriz de covarianzas de las gausianas y d el número de dimensiones del modelo. Además se debe cumplir que $\sum_{i=1}^{M} wi = 1$. Por medio de estos parámetros se puede describir el modelo de la siguiente forma: $\lambda = (w_i, \overrightarrow{\mu_i}, \sum i)$.

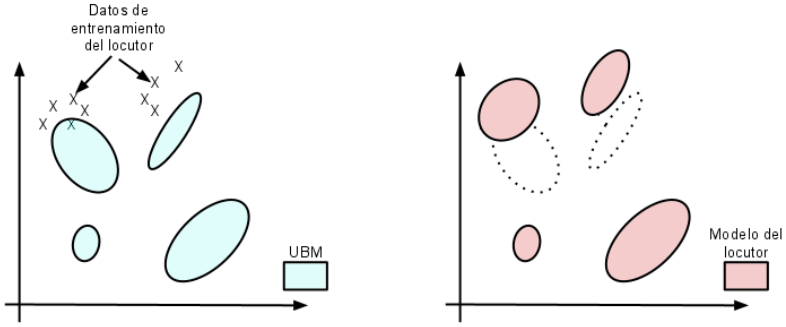


Figura 2.15: Adaptación de un locutor a partir de sus ficheros de training sobre UBM

Una vez obtenidos los modelos estadísticos representados anteriormente, λ_t y λ_{UBM} más una secuencia de observaciones extraída de un segmento de voz $O = O_1, O_2, ..., O_N$ se obtiene una medida de la similitud entre el vector O y λ_t. Este score se obtiene evaluando las funciones de densidad de probabilidad de los modelos para el vector de características O:

$$S(O, \lambda_t) = \log f(O|\lambda_t) - \log f(O|\lambda_{UBM})$$

donde $f(O|\lambda_t)$ y $f(O|\lambda_{UBM})$ son las funciones de densidad de probabilidad evaluadas para el vector O.

Compensación de variabilidad inter-sesión

Uno de los problemas presentes en el sistema automático de reconocimiento de locutores es la variabilidad introducida en diferentes sesiones de adquisición de datos (variabilidad inter-sesión).

El método empleado en este proyecto final de carrera para contrarrestar esta variabilidad es Join Factor Analisys.

Es una técnica de compensación desarrollada en los últimos años que ha demostrado reducir de forma significativa la influencia del canal en las locuciones [7]. Consiste en modelar las direcciones de máxima variabilidad interlocutor e intra-locutor de las características extraídas. A partir de esta información se trata de compensar aquellas variaciones relacionadas con la variabilidad intra-locutor y mantener las variaciones interlocutor.

La limitación principal de Join Factor Analisys es que su rendimiento depende de la disponibilidad de un corpus que contenga las mismas condiciones de la voz a reconocer. Lamentablemente, la existencia de un corpus así no es frecuente en aplicaciones reales, especialmente en aplicaciones forenses.

2.4. Evaluación de rendimiento

En los sistemas de reconocimiento automático de locutores que funcionan en modo verificación, en los que se toma una decisión de aceptación o rechazo, se presentan dos tipos de errores principales [8]:

- Falso rechazo: se rechaza a un usuario que posee realmente la identidad que dice poseer.
- Falsa aceptación: el sistema acepta una identidad falsa proveniente de un impostor.

Estos dos errores dependen del umbral escogido para realizar la tarea de decisión. Un umbral bajo provoca que el sistema tienda a aceptar un gran número de identidades, lo que produce muy pocos casos de falso rechazo y muchos de falsa aceptación. Sin embargo, un umbral alto produce el efecto contrario: muy pocas identidades son validadas con lo que el falso rechazo aumenta mientras que la falsa aceptación desciende. A partir de los valores de la probabilidad de estos dos errores (P_{FA} y P_{FR})se define el *punto de operación* del sistema que corresponde con un determinado valor del umbral de decisión.

2.4.1. Curvas DET

Según el trabajo realizado por [9] en el momento en el que existe un compromiso entre tipos de error (P_{FA} y P_{FR}) un simple número de rendimiento no es suficiente para representar las capacidas del sistema. Un sistema de reconocimiento de locutor tiene muchos puntos de funcionamiento y está mejor representado por medio de una curva de rendimiento.

Tradicionalmente se han empleado las curvas ROC (*Receiver Pperating Characteristic*. Como se describe en [1] la curva ROC muestra la probabilidad de FA respecto a la probabilidad de FR. En [9] la tasa de falsa alarma es situada en el eje horizontal miesntras que la tasa de acierto se muestra en el vertical. La familia de curvas generadas representa diferentes sistemas con FA-FR, donde los mejores sistemas son aquellos más cercanos al origen. Para cualquier sistema particular la curva ROC asociada cambia en función de la variación del umbral de selección escogido. Una curva ROC básica se muestra en la gráfica 2.16.

En este proyecto se ha hecho empleo de la curvas DET (*Detection Error Trade-off*, una variación de las curvas ROC. En la curva DET se representa cada tasa de error en un eje, usando una escala para cada eje que lo extiende y hace que se distingan mejor las diferentes pruebas realizadas. La curva generada es monótona decreciente y determina las características de operación del sistema.

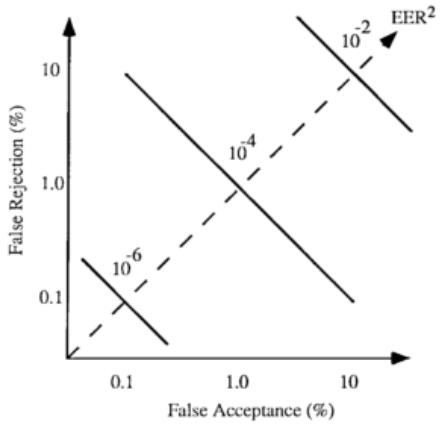


Figura 2.16: Ejemplo básico de curva ROC

En un sistema con los scores del locutor y de los impostores que se ajustan bien a distribuciones Gausianas con la misma varianza se obtiene una curva lineal con una pendiente igual a -1. Cuanto mejor sea el sistema más cerca estará la curva del origen. En la práctica, las distribuciones de puntuación de similitud no siguen una distribución exactamente gaussiana pero son bastante cercanas por lo que un análisis de rendimiento por medio de curvas DET se muestra adecuado para evaluar el rendimiento del sistema. Un ejemplo de curva DET para distintos conjuntos de scores se muestra en la figura 2.17

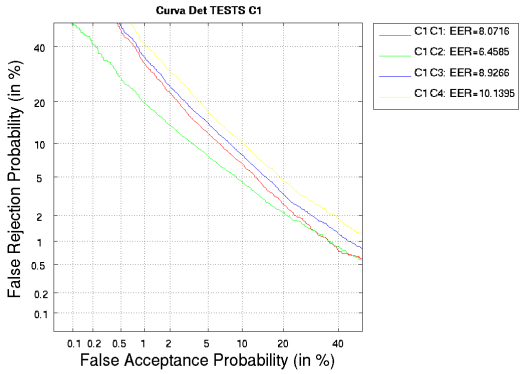


Figura 2.17: Ejemplo de curva DET para distintos conjuntos de scores

2.4.2. Evaluaciones NIST

Las evaluaciones NIST (National Institute of Standards and Technology) son un evento periódico (en la actualidad con carácter bianual) cuya finalidad es contribuir a la investigación de nuevas técnicas de reconocimiento automático de locutores en aplicaciones independientes del texto y, a su vez, calibrar las capacidades técnicas actuales.

Según la convocatoria oficial de para la evaluación de reconocimiento de locutores NIST Speaker Recognition Evaluation (NIST SRE) de 2008 [3], que se utilizará en este proyecto, consiste en 13 tests diferentes definidos por la duración, el tipo del entrenamiento y el tipo de los datos de test. La tarea principal de la prueba es averiguar cuando un determinado locutor

está hablando durante un segmento de conversación. De los 13 test presentados, uno de ellos es de obligada realización (*core* de la prueba) pudiéndose escoger el resto de test.

Las condiciones de entrenamiento son extractos de una conversación en los que no existe un sistema de eliminación de silencios. Los datos de entrenamiento son los siguientes:

- 10-sec: Un extracto de dos canales sobre una conversación telefónica que contiene aproximadamente 10 segundos de voz.
- short2: una conversación telefónica en dos canales de unos 5 minutos de duración, con el canal del locutor target pre-designado o empleando una grabación microfónica de aproximadamente tres minutos de duración en la que están presentes el locutor y un entrevistador. Para los segmentos de tipo ïnterview", la mayor parte del habla es generada por el locutor tarde y, para obtener coherencia durante la comparación se añade un segundo canal a ceros. Estos datos de entrenamiento son los usados durante este proyecto.
- 3conv: Tres conversaciones telefónicas de dos canales.
- 8conv: Ocho conversaciones telefónicas de dos canales.
- long: Una conversación telefónica en un único canal grabada por medio de un micrófono y con una duración de ocho minutos o más en la que intervienen el locutor y un entrevistador.
- 3summed: Tres canales telefónicos sumados formados por la suma muestra a muestra de sus dos canales.

Los segmentos de test son extractos de una conversación. Como en el caso de los segmentos de entrenamiento tampoco existe un sistema de eliminación de silencios. Los segmentos de tests son los siguientes:

- 10-sec: Un extracto de dos canales sobre una conversación telefónica que contiene aproximadamente 10 segundos de voz.
- short3: una conversación telefónica en dos canales de unos 5 minutos de duración, con el locutor de test obtenido de un canal microfónico o de una entrevista de tres minutos formada por el locutor y un entrevistador. Para los segmentos de tipo interview, la mayor parte del habla es generada por un locutor targe y, para mantener la consistencia a través de la condición de evaluación se añade un canal de ceros. Este es el tipo de datos a emplear.
- long: Una conversación telefónica en un único canal grabada por medio de un micrófono y con una duración de ocho minutos o más en la que intervienen el locutor y un entrevistador.
- summed: Una conversación telefónica en un canal formado por la suma de sus dos canales.

Las pruebas se organizan siguiendo esta tabla:

Una vez procesados los datos se calculan valores del rendimiento del sistema y se envía al instituto NIST junto con una breve descripción del sistema biométrico desarrollado.

		Test Segment Condition			
		10sec	**short3**	**long**	**summed**
Training Condition	**10 sec**	optional			
	short2	optional	required		optional
	3conv		optional		optional
	8conv	optional	optional		optional
	long		optional	optional	
	3summed		optional		optional

Figura 2.18: Tabla Nist

3

Calidad y sistemas biométricos

3.1. Introducción

En este capítulo se presentan los aspectos más importantes de la calidad en la señal de voz, así como un repaso a las técnicas de agrupamiento empleadas para la realización de este proyecto final de carrera.

3.2. Robustez en sistemas de reconocimiento automático

Ante los distintos factores que pueden afectar el rendimiento de un sistema biométrico uno de los más importantes es la calidad de las muestras biométricas que emplea el sistema. Con muestras de baja calidad se pueden obtener tasas de error muy superiores a la medida con lo que el objetivo principal, identificar usuarios, queda seriamente comprometido.

La toma de los umbrales de decisión es la última etapa en los sistemas de reconocimiento automático de locutor. La elección del valor numérico sigue siendo un tema abierto (por lo general, el umbral se fija empíricamente) y su fiabilidad no se puede garantizar, mientras el sistema está funcionando. Esta incertidumbre en la elección del umbral se debe a la variabilidad del score presente entre distintas pruebas.

Como se puede observar en [10] la variabilidad proviene de distintas fuentes. Por un lado, los sistemas de adquisición de datos pueden variar entre los distintos hablantes que forman parte del conjunto de estudio. Las diferencias también provienen del contenido fonético, la duración de la grabación, el ruido ambiental así como la calidad del entrenamiento del modelo del locutor. Por otro lado, la posible falta de correspondencia entre los datos de la grabación (obtenidos para el modelado del locutor) y los datos de test. Este es el principal problema que perdura en el reconocimiento de locutor.

De acuerdo con lo descrito [10] son dos los motivos principales por los que se crea esta falta de correspondencia:

- Variabilidad intra-sesión: variabilidad presente en la voz del locutor debido a emociones, estado de salud y la edad, así como a otra serie de condiciones medioambientales, cambios

en el canal de transmisión de la voz, en los métodos empleados para grabar la voz o condiciones acústicas (eco durante la grabación).

- Variabilidad inter-sesión: variación de voces entre distintos locutores. Esto es un problema a la hora de fijar un umbral de decisión en los sistemas independientes de locutor, ya que no es una variabilidad directamente medible. De esta forma no se puede fijar un umbral que proteja el sistema de verificación en contra de un impostor.

Por último, la naturaleza y la calidad de los segmentos de test (igual que con los segmentos de modelo de locutor) influyen en el valor de los scores tanto para pruebas de verificación con el sujeto correcto como para pruebas empleando impostores.

De cara a solventar estos problemas se han empleado una serie de técnicas que pretenden compensarlos. En esta sección se seguirá haciendo hincapié en los factores que producen esa pérdida de fiabilidad así como en los métodos más habituales para combatirla.

3.2.1. Factores degradantes

Existen una serie de factores que afectan a la calidad final del sistema. Las etapas más sensibles en el reconocedor están relacionadas con las muestras empleadas y con la extracción de características a partir de éstas.

Dentro de las muestras empleadas se pueden realizar dos grandes clasificaciones que serán vistas más en detalle de aquí en adelante. Una de ellas depende únicamente del sujeto bajo estudio mientras que la otra depende de la muestra ya obtenida.

Factores relacionados con los sujetos

Como se observa en el trabajo realizado por [11] los factores relacionados con los sujetos se basan en las características físicas y de comportamiento de un sujeto que pueden afectar a la muestra a emplear en las siguientes etapas del reconocimiento.

- Limitación en las características del sujeto: factores como la edad del sujeto (el reconocimiento facial en niños es más difícil de realizar que en adultos debido a que la estructura ósea todavía no está fijada). Igualmente una huella dactilar puede haber sido capturado de forma muy nítida, sin embargo, al contener ésta pocas minucias puede ser difícilmente distinguible.

- Cambios en las características: Las muestras biométricas pueden ser muy variables con el tiempo, operaciones de estética, las enfermedades o, simplemente, el estado físico y anímico del sujeto. En la señal de voz por ejemplo, un simple resfriado o una afonía pueden cambiar las características de la voz produciendo un cambio en las condiciones.

- Comportamiento del sujeto: La cooperación del sujeto, el conocimiento del sujeto de las técnicas de obtención de las muestras y el estado emocional son factores que afectan a la calidad de la muestra biométrica y que introducen variabilidad en las condiciones de ésta.

- Fraude: El fraude por evasión consiste en que el sujeto puede intentar modificar u ocultar un rasgo para así evitar ser identificado. Otro tipo de fraude es el denominado "spoofing.^en el que se intenta culpar a otro sujeto empleando una muestra biométrica del mismo.

Factores relacionados con la adquisición de datos

Esta serie de factores están ligados a la forma en la que se obtiene la muestra biométrica de los sujetos. Comprende tanto los dispositivos físicos de adquisición como los procesos por los cuales se obtiene o se transmite la muestra biométrica.

- Dispositivos de adquisición: Diferentes tipos de sistemas extraen los datos biométricos con diferente calidad y en diferentes condiciones. El empleo de un tipo u otro determinará en muchos casos la precisión del reconocimiento biométrico. Según el trabajo previo realizado por [12] uno de los sistemas más sensibles en cuanto al método de adquisición son los reconocedores por medio de huella dactilar. En ese trabajo se presentan las diferencias encontradas empleando sensores capacitivos, ópticos y térmicos. De la misma forma, en este proyecto se emplean muestras obtenidas por dos dispositivos de adquisición diferentes, por ejemplo, señal de voz de origen microfónico o de origen telefónico.
- Procesos de adquisición: Efectos del medio en el que se obtiene la muestra (humedad ambiental, ambiente ruidoso, situaciones de escasa iluminación, etc) determinan su calidad. También influyen el grado de supervisión del sujeto durante el proceso de adquisición o el número de capturas de la muestra realizadas.
- Canal de transmisión y almacenamiento: En sistemas de reconocimiento de locutores afectan muchísimo. En voz existe una gran diferencia entre GSM, voz de origen telefónico o cinta magnetofónica (hasta hace poco era el sistema empleado en grabaciones forenses).

Factores relacionados con el procesado y la extracción de características

La compresión de las muestras biométricas supone una pérdida de información que afecta a la calidad de dicha muestra. Por lo tanto existe un compromiso entre la cantidad de información almacenada en el sistema (sistemas de compresión con pérdidas) y la calidad de éstos.

En el caso de la voz, quizá uno de los ejemplos más característicos sean las conversaciones telefónicas. En éstas se recorta las frecuencias para adecuarlas al canal de transmisión, con lo que se produce una pérdida de calidad.

El empleo de diferentes métodos de extracción de características produce diferente información extraída de las muestras. Esto afectará al rendimiento del sistema.

3.2.2. Técnicas de compensación de variabilidad intersesión

La señal de voz está sujeta a muchas variaciones, relacionadas con el canal de transmisión así como el estado físico y anímico del sujeto. La mayoría de las veces estas variaciones son indeseables. Uno de los mayores problemas a los que se tienen que enfrentar los sistemas de reconocimiento es cómo contrarrestar esta variabilidad en la muestra biométrica.

Detector de Actividad de Voz (VAD)

Un detector de Actividad de Voz (VAD, *Voice Active Detector*) es un sistema por el cual se localizan los tramos de voz en una señal. El VAD es una parte importante ya que evita procesado de silencios (que no aportan información) y, además, reduce la complejidad de la señal a la hora de procesarla [5].

El detector de actividad más sencillo emplea la energía de la señal para determinar las zonas de actividad. En primer lugar se estima un umbral de actividad, es decir, por encima de ese

nivel de energía habrá actividad de voz y por debajo de ese nivel será silencio. Un método para estimar el umbral consiste en computar la energía de todos las muestras de la señal de voz.El umbral se sitúa a 30dB por debajo del valor máximo. Tradicionalmente un VAD descarta el 20-25 % de la señal si la fuente es una conversación telefónica.

Compensación de variabilidad a nivel de características cepstrales

La normalización de características sirve para eliminar la variabilidad intersesión de los coeficientes cepstrales.

Según el trabajo realizado por [5] en principio, es posible usar técnicas genéricas de eliminación de ruido en el dominio temporal antes de realizar la extracción de características. Por ese motivo, se normalizan las características una vez han sido extraídas y pertenecen al dominio cepstral.

Los métodos más básicos son:

- **Resta de la media cepstral** (CMS - *Cepstral Mean Subtraction*): En [13] se explica el procesado realizado en CMS. Al restar el vector media a los dos canales obtenidos se obtiene una media cero. Este procesado reduce el efecto del canal, bajo la hipótesis de que dicho canal es un elemento de variación lineal en el dominio cepstral (de esta forma su contribución principal es a la media de los vectores cepstrales).

- **Filtrado RASTA** (*RElative SpecTrAl* [14]): consiste en aplicar un filtro paso banda en el dominio log-spectral o en el dominio cepstral. El filtro suprime las frecuencias moduladas que se encuentran fuera de los rangos típicos en la señal de voz. Por ejemplo, el ruido convolucional puede ser visto como baja frecuencia, con lo que el filtrado RASTA lo eliminaría. La complejidad de esta técnica es mayor que la de CMS, haciendo que el espectro de la señal de voz obtenido tras el filtrado dependa de instantes pasados.

Otro método para normalizar las características extraídas de la señal consiste en modificar la distribución de probabilidad de las características. El método **Feature Warping** [15] se basa en conseguir que los parámetros sigan una distribución gaussiana de media nula y varianza unidad por medio de la transformación del histograma.

Por medio del trabajo realizado por [16] se observa que **Feature Mapping (FM)** presenta mejores resultados que Feature Warping. El método consiste en una transformación de las características obtenidas de canales con diferentes condiciones a un espacio de características independiente del canal, teniendo en cuenta la correlación entre dimensiones a la hora de realizar la normalización.

Por último, **Factor Analysis** es una técnica de compensación de variabilidad desarrollada en los últimos años [17]. Se basa en modelas las direcciones de máxima variabilidad (tanto interlocutor como intralocutor) de las características extraídas del habla del locutor. A partir de esta informa se compensa las variaciones intralocutor y se pretende mantener las interlocutor. Para mayor información sobre este tema, consúltese [2].

Normalización de score

Una vez obtenida la puntuación de verosimilitud, ésta se puede normalizar respecto a un conjunto de otros modelos locutor (cohorte). Se transforman los scores de diferentes locutores en un rango similar de forma que se pueda emplear un umbral común. La normalización del score también puede servir para corregir parámetros no compensados anteriormente. En [5] se realiza

siguiendo el siguiente desarrollo:

$$s' = \frac{s - \mu_I}{\sigma_I}$$

donde s' es el score normalizado, s es el score original, μ_I y σ_I es la media y la desviación estándar de la distribución cohorte (conjunto de todos los posibles locutores).

Uno de los métodos más empleados a la hora de realizar una normalización del score es el método **T-Norm**. T-norm intenta compensar las variaciones en los segmentos de test (como por ejemplo la duración del fichero y el contenido lingüístico)[5]. Este sistema de normalización consiste en que a la vez que se enfrenta el fichero de test al modelo bajo estudio, se enfrenta también a una cohorte de modelos de impostores (usuarios distintos al fichero de test). De esta distribución cohorte se obtiene la media y la varianza que se aplicará a los enfrentamientos obteniéndose un alineamiento de la distribución de probabilidad *non-target* dependiente del fichero a identificar:

$$s_{T_{norm}} = \frac{s_{raw} - \mu_{T_{norm}}}{\sigma_{T_{norm}}},$$

donde s_{raw} es el score sin normalizar, $\mu_{T_{norm}}$ y $\sigma_{T_{norm}}$ son los valores media y varianza de la distribución gaussiana aproximada y obtenida por el enfrentamiento del modelo de test y a la cohorte de impostores y $s_{T_{norm}}$ se define como el score normalizado.

En la normalización cero **Z-Norm** ([10]) el modelo del locutor es probado contra un conjunto de señales de voz producidas por algún impostor, obteniendo una distribución de score similar a un impostor. La media y la varianza se estiman en función a esta distribución. La ventaja de esta técnica es que los parámetros de normalización pueden ser obtenidos offline durante el entrenamiento del modelo del locutor. La desventaja de este método frente a T-Norm, es que Z-Norm requiere un método para evaluar la probabilidad de la muestra de test bajo la suposición de que el locutor real es un locutor diferente al sujeto bajo estudio (es un hablante "desconocido") [18].

$$s_{Z_{norm}} = \frac{s_{raw} - \mu_{Z_{norm}}}{\sigma_{Z_{norm}}},$$

donde s_{raw} es el score sin normalizar, $\mu_{Z_{norm}}$ y $\sigma_{Z_{norm}}$ son los parámetros de la distribución de los scores non target extraídos del enfrentamiento entre el modelo bajo estudio y la cohorte de ficheros de test y $s_{Z_{norm}}$ es el score normalizado.

3.3. Medidas de calidad

Como se puede observar en [19] la calidad se puede definir como el grado de bondad de un elemento dado un cierto criterio. Matemáticamente se puede considerar que tiene interpretación como una probabilidad, es decir

$$Q^{\varepsilon} = p(Y \text{ respecto a } \varepsilon)$$

donde ε es un criterio de bondad respecto a Y. La salida esperada de esta función es 0 cuando Y no establece el criterio impuesto por ε y 1 cuando Y satisface totalmente el criterio de ε.

Al trabajar con señales de voz se puede considerar Y como la energía de la señal y ε como un criterio calidad basado en SNR. De esta forma tendríamos

$$Q^{\varepsilon=SNR}(Y) = p(Y > ruido)$$

Para obtener una medida de calidad exitosa es muy importante cómo escoger el criterio de bondad. Por definición, cualquier factor que afecte al comportamiento de un elemento en un sistema de reconocimiento automático de locutor puede ser empleado como criterio de bondad. Existen dos tipos de criterios de bondad, aquellos que son dependientes de la identidad propuesta (necesitan información del entrenamiento del sujeto) o independientes de la identidad propuesta.

Por medio de una función de mapeo $Q(x)$ se expresa cada indicador de degradación (parámetros que indican la degradación de la voz [20]) perteneciente a un intervalo comprendido entre 0 y 1, donde 0 es el valor mínimo de calidad y 1 el máximo. De esta forma se consigue trabajar de forma heterogénea con diferentes tipos de calidad. Para cada una de las medidas de calidad propuestas más adelante se proporcionará la función de mapeo correspondiente.

3.3.1. La calidad en la señal de voz

Antecedentes

La estimación de la calidad de la voz es un tema ampliamente estudiado debido a la importancia de este factor en las redes de comunicaciones telefónicas. Desde los primeros métodos desarrollados (observar la calificación otorgada a una serie de locuciones por un un grupo de personas) se ha llevado a cabo una importante actualización en los métodos de estimación de calidad debido principalmente a la irrupción de las nuevas tecnologías de voz (comunicaciones móviles y voz sobre IP) que hacen necesario monitorizar la calidad de forma constante.

Este proyecto final se basa en las herramientas y estándares presentes en la actualidad que han aportado amplios conocimientos sobre estimación de calidad [21].

Continuando con el trabajo realizado por [21], la calidad de una muestra de voz viene determinada por los siguientes criterios básicos:

- Fidelidad: exactitud y precisión con la que una muestra es capturada y procesada en el sistema.
- Carácter: Altamente dependiente de los factores conductuales ya que se basa en la actitud del usuario.
- Utilidad: ser capaz de evaluar y predecir el rendimiento de un sistema.

En los siguientes subapartados se describen brevemente las medidas de calidad empleadas a lo largo del desarrollo de este proyecto final de carrera. Estas medidas han sido seleccionadas de acuerdo a criterios de utilidad fundamentalmente, como se puede comprobar en [20].

SNR: Relación Señal a Ruido

Como su propio nombre indica esta medida de calidad se basa en la relación entre la potencia de la señal de voz y la potencia del ruido. Formulando esta definición se obtiene:

$$SNR = 10 \cdot \log \frac{E_{voz}}{E_{silencio}},$$

donde E_{voz} y E_{ruido} representan la energía media de la señal y la energía media de los silencios. La forma más habitual de estimar los silencios presentes en la señal de voz es por medio de un Detector de Actividad de Voz.

La función de mapeo que permite convertir la SNR obtenida una representación comprendida entre 0 y 1 es la siguiente:

$$Q_{SNR}(x) = \frac{x}{60},$$

donde x se corresponde con el valor obtenido en el cálculo de la SNR de tal forma que $x \in [0, 60]$.

KLPC

La kurtosis es una medida estadística basada en medir el parecido de una distribución con una gaussiana empleando la energía de las colas y de la simetría. Para realizar el cálculo, basta con obtener el cuarto momento estadístico de la distribución.

Se denomina Kurtosis LPC porque se emplean los P coeficientes LPC:

$$k = \frac{1}{P} \cdot \sum_{p=1}^{P} \left(\frac{a_p - \frac{1}{P} \cdot \sum_{p=1}^{P} a_p}{\sigma} \right) - 3,$$

donde a_p son los valores de la distribución de los P coeficientes LPC y σ es la desviación típica.

Igual que sucedía en el caso de la medida de calidad SNR, es necesario implementar una función de mapeo que nos permita transformar los valores obtenidos, k, a un rango comprendido entre 0 y 1. De este modo la función de mapeo empleada es la siguiente:

$$Q_{KLPC}(x) = 1 - \frac{x - 3}{8},$$

donde x es el valor de la kurtosis de los coeficientes LPC obtenidos en las etapas previas del sistema biométrico.

KCEP

La Kurtosis cepstral es idéntica a la kurtosis LPC descrita en el apartado anterior, sin embargo en este caso se emplean los coeficientes MFCC.

La función de mapeo correspondiente es:

$$Q_{KCEP}(x) = \frac{x - 10}{6}$$

P563

El indicador de degradación P-563 es un método de evaluación de la calidad percibida. Esta percepción de la calidad varía de individuo a individuo siendo una medida subjetiva. Sin embargo, P-563 es el estándar de la ITU (*International Telecommunications Union*) empleado para medir la calidad de servicio de las redes telefónicas. Históricamente, la red telefónica ha

sido su ámbito de utilización básico, pero con el auge de los sistemas biométricos su uso se ha extendido a estas nuevas tecnologías. El cálculo básico comprende factores que afectan a las redes de comunicaciones modernas como ruido, ecos, jitter, pérdida de paquetes, etc.

Según la norma de la ITU para esta calidad (ITU-P563) se calculan un total de 51 parámetros sobre los que se estima el tipo de degradación presenta. Además se calcula una puntuación MOS (*Mean Opinion Score*).

Una vez obtenidos estos 51 parámetros, se analizan 8 de ellos para calcular el tipo de degradación más abundante. Cada uno de estos 8 parámetros guarda una gran relación con un factor de degradación determinado.

Como en los indicadores de degradación presentados anteriormente, P-563 también necesita ser mapeado para obtener el rango adecuado de representación. Para realizar la función de mapeo se emplea la escala MOS donde 1 se corresponde con la peor calidad posible y 5 con la mejor. De esta forma se obtiene:

$$Q_{P-563}(x) = \frac{x-1}{4}$$

UBML

La medida de calidad UBML fue propuesta por Harriero, Alberto [20] como contribución original de su proyecto final de carrera. Consiste en aproximar la similitud entre una locución y el modelo universal utilizado para generar el modelo estadístico del locutor.

Al emplear un sistema de reconocimiento basado en GMM su obtención es inmediata, sin suponer un cálculo computacional extra. Empleando la fórmula para la extracción del score de verosimilitud:

$$S(O|\lambda_t) = \log \frac{f(O|\lambda_t)}{f(O|\lambda_{UBM})},$$

de donde el numerador se corresponde con la similitud de la locución con el modelo estadístico del locutor. Mientras que el denominador $f(O|\lambda_{UBM})$ es la verosimilitud del habla de testeo frente a la función densidad de probabilidad del modelo universal. Ver la sección 4.5.

Este indicador de degradación aporta la idea de que si un modelo universal está entrenado con un tipo de habla, una locución de un tipo de habla diferente funcionará peor ya que el UBM no es representativo.

3.4. Clustering

En este aparto se muestran las técnicas para realizar el agrupamiento de ficheros de audio utilizando medidas de calidad. El objetivo de este proyecto es conseguir, a partir de los indicadores de degradación propuestos anteriormente, generar grupos automáticamente de forma que cada grupo esté formado por aquellos ficheros de audio obtenidos con el mismo sistema de adquisición.

La realización de este tipo de agrupaciones surge de la necesidad de la Guardia Civil de contar con un sistema de definición de patrones de cara a la acreditación de sus laboratorios. Una vez obtenido este sistema con la base de datos de la Guardia Civil se ha extrapolado su uso para emplearlo con otras bases de datos (en este caso NIST 2008) y, así, comprobar cómo de buenos son los resultados obtenidos en otras bases de datos.

3.4.1. Algoritmos de agrupamiento

El clustering (agrupamiento en castellano) [22] consiste en la tarea de asignar un conjunto de objetos en diferentes grupos (llamados clusters) de forma que los elementos que forman uno de los grupos tengan en común ciertas características que los distingan de los demás.

Existen diferentes tipos de agrupamiento de los datos en función de la metodología seguida para crear los grupos:

- Aglomerativos (*bottom-up*): la etapa inicial consiste en que cada elemento forma un cluster independiente y el algoritmo va formando clusters cada vez más grandes al ir agrupando los distintos elementos.
- Divisivos (*top-down*): Es el caso contrario a los algoritmos aglomerativos. Se parte de una situación inicial en la cual todos los elementos forman un mismo cluster para ir creando clusters más pequeños a medida que avanzan las iteraciones del algoritmo.

Los métodos citados arriba forman parte de los métodos jerárquicos que encuentran los nuevos clusters usando previamente unos ya establecidos (ya sea un único grupo o tantos grupos como elementos haya). En el sistema implementado en este proyecto final de carrera el clustering de las medidas de calidad se ha realizado mediante algoritmos divisivos en los que el número de clusters se iba variando hasta encontrar el número óptimo de éstos.

Otro factor muy importante de cara a la correcta asignación de los elementos a los grupos generados es la forma de medir cómo de similares son dos elementos. Para ello se emplean medidas de distancias. Estas distancias influirán en la forma que tengan los clusters (la frontera entre éstos)y, por lo tanto, en el rendimiento obtenido en el proceso. Las distancias utilizadas en la realización de este proyecto son las siguientes [23]:

- Distancia euclídea: es la distancia básica empleada. Muestra la distancia entre dos puntos en un plano euclídeo tal y cómo lo mediaríamos con una regla. Siendo $P1$ y $P2$ dos puntos en el espacio euclídeo con coordenadas $(x1, y1)$ y $(x2, y2)$ respectivamente, se obtiene que

$$d(P1, P2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y1)^2}$$

- Distancia Cityblock (también conocida como distancia Manhattan o Taxicab): Con este sistema la distancia se obtiene como la suma de las diferencias (absolutas) de las coordenadas de los puntos. Teniendo dos vectores p y q en un espacio vectorial real n-dimensional la distancia es la suma de las longitudes de las proyecciones del segmento de línea sobre el sistema de ejes coordenados:

$$d_1(p, q) = ||p - q||_1 = \sum_{i=1}^{n} |p_i - q_i|$$

- Distancia Cosine: Obtiene la distancia entre dos vectores al medir el coseno del ángulo que forman. El coseno de dos vectores se puede obtener usando el producto: $a \cdot b = \|a\| \, \|b\| \cos\theta$. De esta forma:

$$distancia = \cos\theta = \frac{A \cdot B}{\|A\| \, \|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}}$$

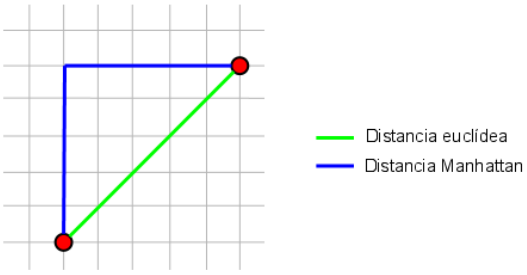


Figura 3.1: Comparativa entre las distancias euclídea y Manhattan.

- Distancia de Mahalanobis: Calcula la similitud entre dos variables aleatorias en un espacio multidimensional, teniendo en cuenta la correlación entre las variables. Sean $\overrightarrow{x}$ y $\overrightarrow{y}$ dos variables aleatorias con la misma distribución de probabilidad y con matriz de covarianza Σ:

$$d_m(\overrightarrow{x}, \overrightarrow{y}) = \sqrt{(\overrightarrow{x} - \overrightarrow{y})^T \Sigma^{-1} (\overrightarrow{x} - \overrightarrow{y})}$$

  Una comparativa entre las distancias euclídea y Manhattan puede ser vista en la figura 3.1

En los siguientes subapartados de esta parte de la memoria se describen dos métodos principales para conseguir el agrupamiento de las medidas de calidad. Los dos métodos siguientes pertenecen ambos a la métodos jerárquicos, más concretamente, a los métodos divisivos, ya que parten de un cluster inicial formado por todos los elementos presentes para llegar a un número de clusters definido en el comienzo. Para determinar el número óptimo de clusters se han realizado varios experimentos, los cuales serán detallados más adelante.

3.4.2. Algoritmo Kmeans

Se trata de un algoritmo de clustering cuyo objetivo es particionar n observaciones $(x_1, x_2, ..., x_n)$en k clusters $(k \leq n)$ de tal forma que cada observación pertenezca al cluster con la media más cercana. Cada cluster es parametrizado por un vector $m^{(k)}$, media del vector.

Según el trabajo desarrollado por [24] los datos son denotados por $\{x^{(n)}\}$ donde el parámetro n va desde 1 hasta el número total de puntos N. Cada vector x tiene I componentes x_i. Se asume que el espacio al que pertenece x es un espacio real en el que existe una métrica para definir distancias entre puntos, por ejemplo:

$$d(x, y) = \frac{1}{2} \sum_i (x_i - y_i)^2$$

El algoritmo K-means es iterativo ya que se realizan diversas ejecuciones hasta encontrar la convergencia de éste. Además es un algoritmo heurístico, es decir, no hay garantías de que converja al óptimo y, además, el resultado se ve influido por la elección de los clusters iniciales.

Para realizar el algoritmo se siguen los siguientes pasos:

1. Inicialización: Asignar al vector K-means $\{m^{(k)}\}$ valores aleatorios.

2. Fase de asignación: Cada punto es asignado a la media más cercana. Se denota la estimación al cluster $k^{(n)}$ de tal forma que el punto $x^{(n)}$ pertenezca al $\widehat{k}^{(n)}$.

$$\widehat{k}^{(n)} = \underset{i}{\operatorname{argmin}}\{d(m^{(k)}, x^{(n)})\}$$

   Como alternativa, una representación equivalente de esta asignación de puntos consiste en emplear un indicador de pertenencia, $r_k^{(n)}$. Este indicador toma el valor 1 en el caso de que media k sea la media más cercana al punto $x^{(n)}$, de otra forma $r_k^{(n)}$ es cero.

$$r_k^{(n)} = \begin{cases} 1 & \text{si } \widehat{k}^{(n)} = k \\ 0 & \text{si } \widehat{k}^{(n)} \neq k \end{cases}$$

3. Fase de actualización(actualización de centroides): Las medias se ajustan para que coincida con el valor medio de los puntos de cada cluster.

$$m^{(k)} = \frac{\sum_n r_k^{(n)} x^{(n)}}{R^{(k)}},$$

   donde $R^{(k)}$ es la suma de todos los indicadores de pertenencia.

$$R^{(k)} = \sum_n r_k^{(n)}$$

4. Iteración: se repiten los pasos de asignación y actualización hasta que las asignaciones no cambian.

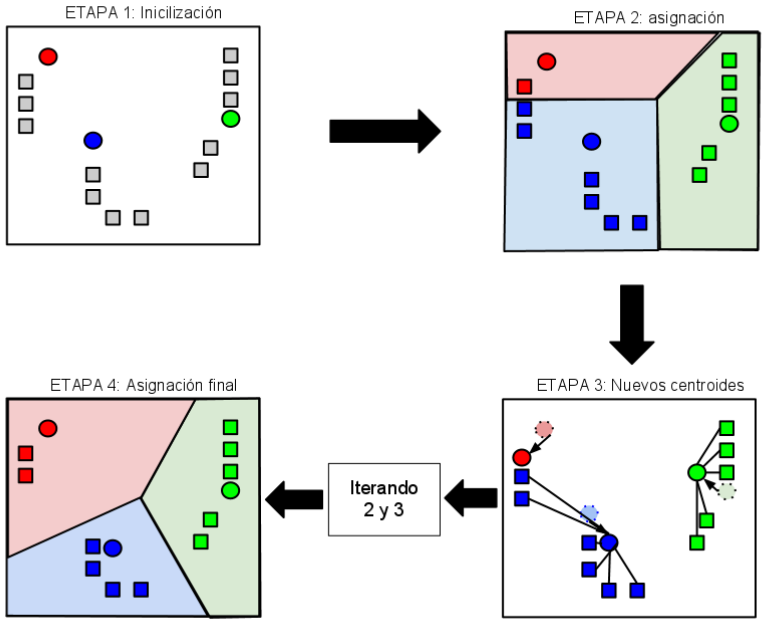


Figura 3.2: Etapas del algoritmo Kmeans

Al ser un algoritmo heurístico, se hace necesario emplear una condición de parada en el caso de que no converja. Esta condición de parada consiste en realizar un número determinado de iteraciones (paso 2 y 3).

3.4.3. Clasificador GMM

A diferencia del algoritmo K-means, el clasificador GMM se basa en un modelo de distribución de probabilidad.

El método empleado para realizar la clasificación de GMM es el algoritmo EM (*Expectation-Maximization Algorithm*). En este algoritmo, el conjunto de datos se modela siguiendo un número determinado de distribuciones gaussianas inicializadas aleatoriamente y cuyos parámetros se van optimizando para ajustarse mejor al conjunto de datos. El objetivo principal del algoritmo consiste en encontrar la máxima verosimilitud o el *maximum a posteriori (MAP)*

Según el trabajo realizado por [5] un modelo de mezcla de gaussianas, denotado por λ está caracterizado por su función densidad de probabilidad (*fdp*).

$$p(x|\lambda) = \sum_{k=1}^{K} P_k N(x|\mu_k, \Sigma_k),$$

donde K es el número de los componentes gaussianos, P_k es la probabilidad del componente k de la gaussiana y

$$N(x|\mu_k, \Sigma_k) = (2\Pi)^{-\frac{1}{2}} |\Sigma_k|^{-\frac{1}{2}} exp\left\{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)\right\},$$

es la función densidad de probabilidad con vector de medias, μ_k, y matriz de covarianzas Σ_k. Las probabilidades deben cumplir que $P_k \geq 0$ y $\sum_{k=1}^{K} P_k = 1$.

Entrenar un modelo GMM consiste en estimar los parámetros $\lambda = \{p_k, \mu_k, \Sigma_k\}_{k=1}^{k}$ a partir de una muestra de entrenamiento $X = \{x_1, ..., x_T\}$. Empleando un criterio de máxima verosimilitud (*Maximum Likelihood Estimation - ML Estimation*) se puede calcular la verosimilitud media de X con respecto al modelo definido λ (LL_{avg} *Log-Likelihood*):

$$LL_{avg}(X, \lambda) = \frac{1}{T} \sum_{t=1}^{T} \log \sum_{k=1}^{K} P_k N(x_t|\mu_k, \Sigma_k)$$

Cuanto mayor sea el valor obtenido en $LL_{avg}X|\lambda$, más probable es que los vectores desconocidos procedan del modelo. Por medio de algoritmos de maximización se consigue maximizar la verosimilitud con respecto a los datos proporcionados. En reconocimiento automático de locutores, uno de los algoritmos más empleados es el EM *Expectation-Maximization*.

Para emplear el algoritmo EM, primero se entrena un modelo de habla universal (*UBM*). Este nuevo modelo representa una distribución independiente de locutor de los vectores de características. De esta forma, los parámetros del modelo no se estiman desde cero, si no que se parte de un conocimiento previo ("datos de locutores universales"). Dada la muestra inicial, $X = \{x_1, ..., x_T\}$ y los datos del modelo de habla universal, $\lambda_{UBM} = \{P_k, \mu_k, \Sigma_k\}_{k=1}^{K}$, el vector de medias adaptadas, (μ_k') se obtiene por medio de *Maximum a posteriori (MAP)*:

$$\mu_k' = \alpha_k \tilde{x}_k + (1 - \alpha_k)\mu_k,$$

donde

$$\alpha_k = \frac{n_k}{n_k + r}$$

$$\tilde{x}_k = \frac{1}{n_k} \sum_{t=1}^{T} P(k|x_t) x_t$$

$$n_k = \sum_{t=1} T P(k|x_t)$$

$$P(k|x_t) = \frac{P_k N(x_t|\mu_k, \Sigma_k)}{\sum_{m=1}^{K} P_m N(x_t|\mu_m, \Sigma_m)}$$

En estas ecuaciones se puede observar, como el factor de relevancia r y, especialmente, α_k controlan cuanto contribuye el UBM al modelo.

En GMM cuando se emplea a la vez UBM, el reconocedor funciona siguiendo un modelo GMM-UBML. El score de similitud depende tanto del modelo de locutor, λ_{target}, como del modelo de hablante universal entrenado, λ_{UBM}:

$$LLR_{avg}\{(X, \lambda_{target}, \lambda_{UBM}\} = \frac{1}{T} \sum_{t=1}^{T} \{\log p(x_t|\lambda_{target}) - \log p(x_t|\lambda_{UBM})\}$$

Principalmente, el método GMM-UBM consiste en medir la *diferencia* del target y del modelo de habla universal. En la figura 3.3 se observa un ejemplo de GMM-UBM empleando un algoritmo *Maximum a Posteriori*. Esta figura ha sido extraída del trabajo de [5].

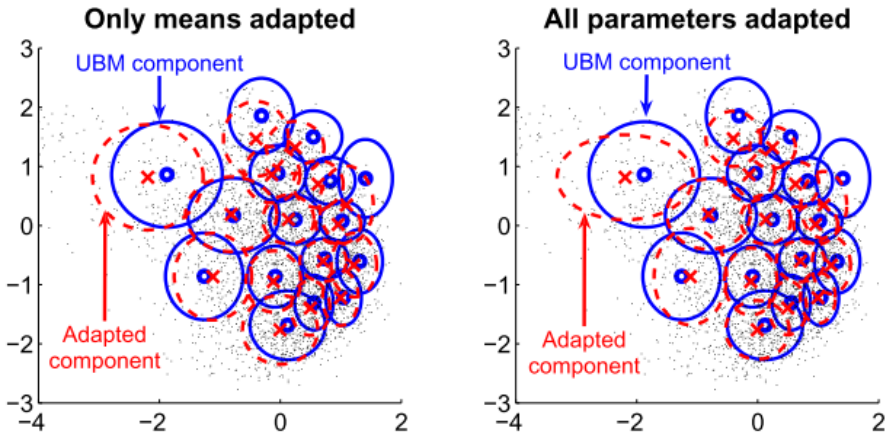
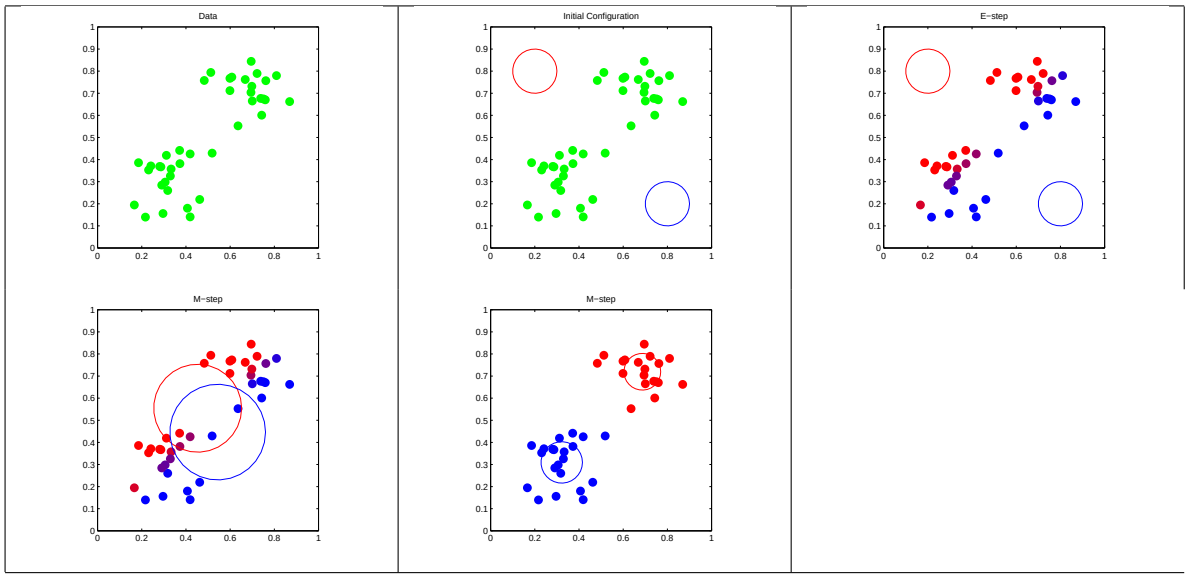


Figura 3.3: GMM-UBM con MAP. Fuente: [5]

De esta forma podemos resumir las etapas básicas de un clasificador:

1. Inicialización
2. Búsqueda de los centroides de las gaussianas
3. Cálculo de parámetros característicos
4. Actualización (algoritmo E-M)



Cuadro 3.1: Agrupación GMM desarrollada en Matlab

3.4.4. Rendimiento de un agrupamiento

En sistemas de reconocimiento biométrico y, en especial, sistemas con aplicaciones forenses, es de vital importancia contar con métodos que nos permitan evaluar cómo de buenas son las técnicas empleadas. Para el esquema de agrupamiento propuesto en este PFC consiste en calcular la entropía de los clusters y la pureza de éstos.

Entropía

La entropía mide el grado de desorden de un determinado agrupamiento. Se basa en indicar como de bueno es nuestro sistema de clasificación en base a la proporción de los distintos tipos de ficheros encontrados en cada cluster. Una entropía con valor cercano a 0 implica que la separación en clusters ha sido perfecta, es decir un cluster esta formado únicamente por un tipo determinado de ficheros.

Entropía de un Cluster

En [25] se presenta una forma de calcular la entropía de una agrupación. Sea $C_i, i = 1, ..., C$ el número de clusters generados cada uno de ellos formados por un número j de elementos.

Para cada cluster C_i se calcula el número de elementos que lo forman, siendo éstos de un tipo determinado $R(C_i)$, de esta forma $f_k(R(C_i))$ es el número de elementos del tipo k donde $k = 1, ..., K$:

$$f_{ik} = f_k(R(C_i))$$

El número de elementos que forman el cluster C_i viene dado por: $n_i = \sum_k f_{ik}$, y el número total de elementos a agrupar(es decir, número de ficheros de audio): $N = \sum_i n_i$

Una vez definidos estos parámetros se puede obtener fácilmente la probabilidad de que un cluster esté formado por un determinado tipo de elementos:

$$p_{ik} = \frac{f_{ik}}{n_i}$$

De esta forma se define la entropía de un cluster determinado, H_i^c:

$$H_i^c = -\sum_k p_{ik} \log_2 p_{ik}$$

Una vez obtenidas todas las entropías parciales, se calcula una entropía total de todos los clusters:

$$H^{total} = \frac{1}{N} \cdot \sum_i n_i H_i^C$$

Entropía del tipo de fichero

Similar a la entropía analizada anteriormente, se analiza la entropía del tipo de fichero. En ella se analiza cómo de bueno ha sido un clustering, pero desde la referencia del tipo de archivo, donde el tipo de archivo viene determinado por el origen de la grabación, es decir, la clase del sensor empleado para adquirir los datos.

De este modo una entropía de 0 indicará que todos los ficheros pertenecientes a un mismo tipo están recogidos dentro del mismo cluster. Mientras que una entropía mayor que 1 indica que los ficheros de un tipo determinado no se corresponden a un único cluster, si no que se encuentran mezclados en diferentes clusters.

Sea $T_m, i = 1, ..., T$ el tipo de ficheros encontrados en el sistema, cada uno de ellos formado por un conjunto de j de elementos.

Sea $C_i, i = 1, ..., C$ el número de agrupaciones que se crean en el sistema.

El objetivo es hallar la probabilidad de que los ficheros pertenecientes al tipo T_m estén agrupados bajo el mismo cluster C_i. Para ello se busca por cada tipo de fichero los clusters que lo forman: $f_{mi} = f_i(R(T_k))$.

El número de elementos que forman cada tipo de fichero T_m viene dado por: $n_m = \sum_i f_{mi}$, y el número total de elementos a agrupar(es decir, número de ficheros de audio): $T = \sum_m n_m$.

A continuación se puede obtener la probabilidad de que un mismo tipo de fichero esté agrupado bajo el mismo cluster:

$$p_{mi} = \frac{f_{mi}}{n_m}$$

De esta forma se define la entropía de un tipo de fichero, H_m^T:

$$H_m^T = -\sum_k p_{mi} \log_2 p_{mi}$$

Una vez obtenidas todas las entropías parciales, se calcula una entropía total de todos los tipos de fichero:

$$H^{total} = \frac{1}{T} \cdot \sum_m n_m H_m^T$$

Se puede observar que ambas entropías están inversamente relacionadas, a medida que se aumenta el número de agrupaciones en las que se dividen los datos, la entropia de un cluster disminuye (es más factible realizar una agrupación formada únicamente por un tipo determinado de ficheros), mientras que la entropía por tipo de fichero aumenta.

Pureza

Siguiendo con el trabajo realizado por [25] la pureza da una idea de rendimiento similar a la entropía. Cuanto mejor sea el clustering, la pureza obtendrá un valor mayor, cercano a 1, mientras que valores cercanos a 0 representan un clustering formado por la unión de varios tipos de ficheros distintos.

$$I^c = \frac{\text{máx}(num_{elementos})}{n_i},$$

donde $num_{elementos}$ representa el número de elementos de un tipo determinado que se encuentran en el cluster c y n_i es el número total de elementos presentes en el cluster c.

4

Agrupamiento de audio basado en medidas de calidad

4.1. Bases de datos y protocolo

Para el desarrollo de este proyecto se han empleado dos bases de datos formadas por ficheros de audio de distintos locutores. Una de ellas pertenece al instituto NIST y es empleada en las Evaluaciones bianuales de rendimiento de las que ya se ha hablado anteriormente. La otra base de datos se obtiene a raíz de la colaboración con la Guardia Civil. En las siguientes subsecciones se explicará en profundidad estas dos bases de datos.

4.1.1. Base de Datos NIST 2008

Las evaluaciones NIST son un método altamente fiable para incentivar y evaluar los diferentes sistemas de reconocimiento de locutores que se están desarrollando en el mundo.

La base de datos NIST SRE 2008 [3] está compuesta de dos tipos de habla: microfónica y telefónica, tanto para locuciones de test como para el entrenamiento de los modelos. Existen locuciones de tipo *interview* en las cuales existe un locutor principal y un entrevistador. Estas conversaciones se emplean tanto en la fase de entrenamiento de modelos como en la fase de test. Estas locuciones están formadas por audio de origen microfónico o telefónico. Adicionalmente se incluyen locuciones de testeo grabadas sobre canal microfónico.

Las conversaciones definidas como tipo *short* tienen una duración media de 5 minutos, con una media de 2.5 minutos de habla (una vez eliminados los silencios). Las locuciones de tipo interview contienen 3 minutos de habla registrada por un micrófono, de esos minutos la mayor parte corresponden al locutor y una pequeña parte al entrevistador.

Dentro del protocolo de evaluación NIST 2008 se definen cuatro condiciones: tel-tel, mic-mic, tel-mic y mic-tel.

Para el desarrollo de este proyecto final de carrera se han empleado las grabaciones correspondientes a tel-tel y mic-mic.

En la tabla 4.1 se observan los distintos ficheros de tipo *model* que forman la base de datos NIST2008.

Modelos					
Interview					
mic-02	mic-03	mic-07	mic-08	mic-12	Total
296	293	295	295	296	1475
Phonecall					
1788					1788
Total de todos los archivos					3263

Cuadro 4.1: Ficheros tipo *model* en NIST 2008

Tests							
Interview							
mic-02	mic-05	mic-09	mic-13				Total
596	596	596	596				2384
Phonecall							
mic-02	mic-04	mic-05	mic-06	mic-07	mic-09	N/A	Total
247	247	247	247	247	247	2213	3695
Total de todos los archivos							6079

Cuadro 4.2: Ficheros tipo *test* en NIST 2008

En la figura 4.1 se observa un gráfico que muestra los valores recogidos en la tabla 4.1. El conjunto de los ficheros *model* está formado prácticamente por el mismo número de elementos de habla *microfónica* como de habla *telefónica*. En conjunto existe una diferencia de 313 grabaciones de origen telefónico más que de origen microfónico. Esto representa un 9 % más de ficheros telefónicos, siempre para el caso de las locuciones tipo *model*.

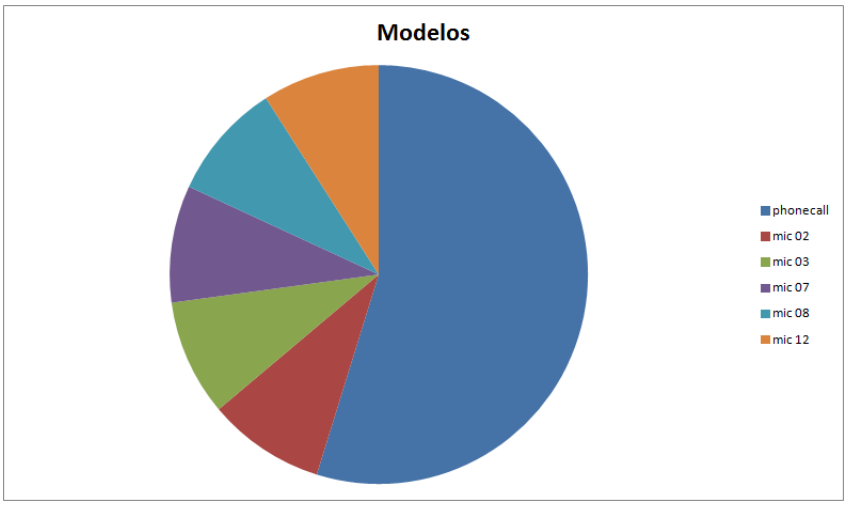
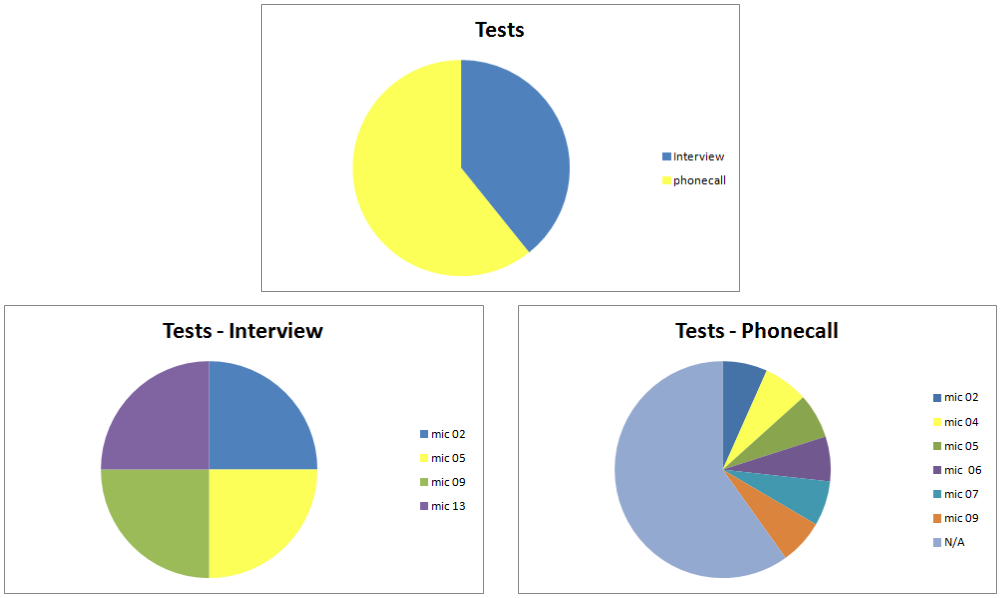


Figura 4.1: Ficheros modelos presentes en la base de datos NIST 2008

Se observa que los ficheros de tipo *test* (en el cuadro 4.2) están igualmente formados por un mayor número de locuciones telefónicas. En este caso, el porcentaje asciende hasta un 21 %. Los ficheros te tipo *test* presentan una organización ligeramente diferente a la presente en los ficheros *model*. En este caso, los ficheros *interview* están subdivididos a la vez en los distintos micrófonos empleados para su captación.

Por lo que se observa en las tablas anteriores, existe un mayor número de ficheros de origen



Cuadro 4.3: Distribución de ficheros *test* en NIST 2008

AhumadaIV-BaezaI						
Ahumada IV			Baeza I			Total
Modelos	Tests	Total	Modelos	Tests	Total	
91	442	533	91	91	182	715

Cuadro 4.4: Distribución de ficheros en AhumadaIV-BaezaI

telefónico en la base de datos. Este dato será importante a la hora de realizar los distintos grupos, de tal forma que en una agrupación perfecta existirá un grupo que contenga un mayor número de elementos que los demás.

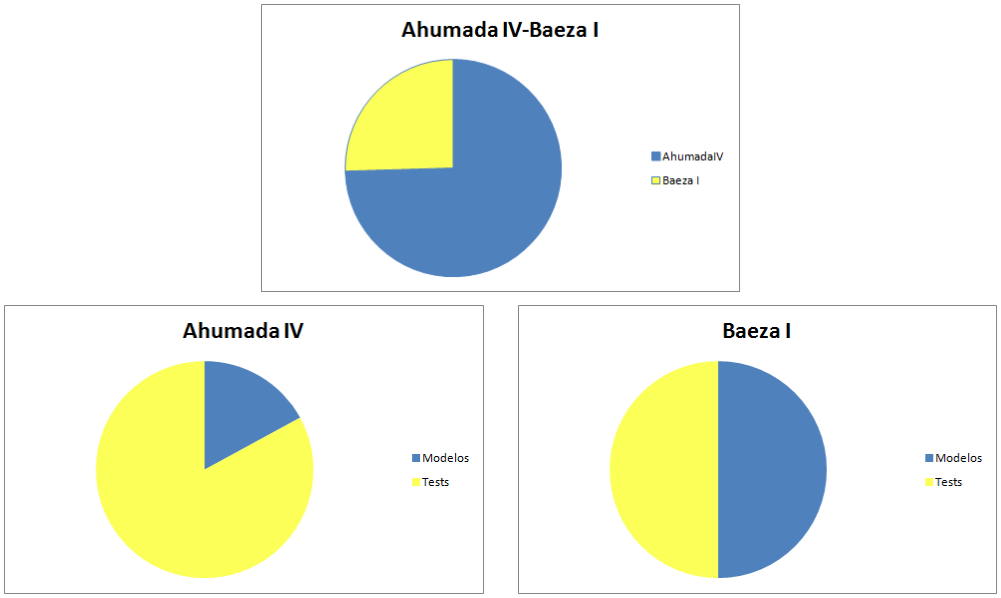
4.1.2. Base de Datos AHUMADAIV-BAEZA

La siguiente base de datos bajo estudio es AhumadaIV-Baeza I. Esta base de datos forma parte de los sistemas de reconocimiento biométrico de la Guardia Civil. En el marco de colaboración del grupo ATVS con este cuerpo de seguridad se tuvo acceso a una parte de los ficheros de audio pertenecientes a ella. Se trata de ficheros de audio anónimos realizados en distintos ambientes y con distintos tipos de sensor de adquisición (habla telefónica o habla microfónica).

A continuación, se da una breve descripción de cómo se organizan los ficheros en esta base de datos.

Como se observa en la tabla 4.4, la base de datos AhumadaIV-BaezaI está formada por la unión de dos bases de datos. De todos los ficheros existentes (un total de 715) 533 ficheros pertenecen a AhumadaIV y 182 ficheros a BaezaI. En la tabla 4.5 se muestra la distribución de los ficheros dentro de cada base de datos.

Los hablantes de la base de datos de la Guardia Civil son castellano parlantes, existiendo diferente acento en función de la región de origen de cada uno de los locutores. De esta forma, las dos bases de datos contienen grabaciones de hablantes andaluces, canarios, castellanos o gallegos. Además, la base de datos AhumadaIV está subdividida en ficheros grabados en interiores y en ficheros grabados en exteriores.



Cuadro 4.5: Gráfica de distribución de ficheros en AhumadaIV-BaezaI

Para poder obtener los valores de calidad de cada grabación necesarios para el desarrollo de la colaboración con la Guardia Civil y de este proyecto,se tuvieron que generar una serie de scripts que, a partir de los ficheros de audio, obtuviera los valores de cada indicador de degradación (ID). Estos scripts se comentarán en la sección 4.2.2..

4.2. Agrupamiento de ficheros

En esta sección de la memoria se presentan los resultados obtenidos en los diferentes experimentos realizados. Además se explicará brevemente el modo de obtención de los resultados así como la metodología empleada.

4.2.1. Metodología empleada

Aunque existen varios pasos comunes en la realización de los experimentos, la forma de obtener los datos de partida es diferente. Mientras que para la base de datos NIST 2008 ya existen una serie de indicadores de degradación calculados debido a trabajos previos [20], para la base de datos AhumadaIV-BaezaI hay que realizar una operativa adicional para obtener los valores de los indicadores de degradación.

Por lo tanto los pasos a seguir para la realización de los experimentos son los siguientes (clasificados por base de datos):

- **AhumadaIV-BaezaI**:
  1. Creación de scripts (bash) para la obtención de los indicadores de degradación.
  2. Creación de un método de lectura en Matlab de los resultados de estos scritps. Adaptación a un formato común ya existente (similar al trabajo realizado con la base de datos extraída de la evaluación Nist 2008).
  3. Implementación de un agrupamiento por K-means. Variación de medidas de distancias empleadas para realizar el agrupamiento.

4. Obtención de resultados: comparativa entre distintos métodos por medio de la entropía de cada agrupamiento.

- **Nist 2008:**

1. Implementación de un agrupamiento por K-means. Variación de medidas de distancias empleadas para realizar el agrupamiento.
2. Implementación de un agrupamiento por GMM.
3. Obtención de resultados: comparativa entre distintos métodos por medio de la entropía de cada agrupamiento.
4. Comparación del rendimiento del sistema por medio de Curvas DET.

4.2.2. Obtención de medidas de calidad

A continuación, se detalla la forma de obtener los valores de los indicadores de degradación necesarios para realizar los agrupamientos.

Se especificará el trabajo desarrollado para cada base de datos por separado, ya que el desarrollo, tal y como se ha comentado en la sección 4.2.1., es ligeramente diferente.

Ahumada IV - Baeza I

La obtención de las medidas de calidad se realiza a partir de los ficheros de audio en formato .wav pertenecientes a la base de datos de la Guardia Civil.

Para la obtención de los indicadores de degradación SNR, UBML y P.563 se generan 3 archivos de bash. Estos scripts están basados en el trabajo previo realizado por Alberto Harriero [20].

El indicador de degradación UBML es el más sencillo de obtener ya que es un dato obtenido directamente de la cabecera del fichero. Se crea un método que recorre todos los ficheros pertenecientes a la base de datos y el valor extraído se escribe en un fichero de texto que luego será cargado desde Matlab.

El indicador de degradación SNR se calcula empleando como herramientas auxiliares el *w*ave2feat y *C*alcularHablaNetaSPHINX, este último generado por el grupo ATVS. Procesando la salida generada por *C*alcularHablaNetaSPHINX se obtiene la SNR. Esta forma de cálculo es similar a la empleada con el indicador de degradación P.563.

Para calcular los valores de la kurtosis cepstral y la kurtosis sobre los parámetros LPC se emplea un código generado en Matlab.

En la tabla 4.6 se muestran los rangos de valores que toman los indicadores de degradación para esta base de datos.

NIST 2008

Para el empleo de la base de datos NIST 2008 estos valores fueron guardados en una estructura de datos de Matlab de forma que no fuera necesario procesar los archivos de audio cada vez que fuera necesario realizar una simulación.

En la tabla 4.7 se presenta el rango de los indicadores de degradación empleados en los siguientes experimentos de este proyecto final de carrera.

VALORES AHUMADA IV - BAEZA I		
Medida	min	max
SNR	9.1317	55.1934
UBML	-11.8264	-7.9164
KLPC	11.7295	16.1167
KCEP	4.0079	10.1611
P563	1	5

Cuadro 4.6: Valores máximo y mínimo de los indicadores de degradación para AhumadaIV-BaezaI

VALORES NIST 2008		
Medida	min	max
SNR	5.1034	90
UBML	-13.3470	-7.4029
KLPC	3.4085	15.5528
KCEP	2.6322	16.6098
P563	0	5

Cuadro 4.7: Valores máximo y mínimo de los indicadores de degradación para NIST 2008

4.2.3. Agrupamiento de medidas de calidad

Una vez obtenidas las medidas de calidad necesarias, se comienza su análisis para realizar las agrupaciones necesarias para cada simulación.

Sobre ambas bases de datos se realizan diferentes experimentos con la finalidad de observar cual ofrece mejor rendimiento (mejores valores de entropía). Las diferentes simulaciones se obtuvieron combinando los diferentes indicadores de degradación entre sí:

$$\sum_{i=2}^{n} C(n,i)$$

donde i se corresponde con el número de indicadores de degradación que forman parte del experimento, n el número total de indicadores de degradación bajo estudio y $C(n,i)$ es la combinación de n elementos escogiendo i.

En las simulaciones realizadas se toman 5 indicadores de degradación diferentes por lo que el número de experimentos realizados por simulación es

$$\sum_{i=2}^{5} C(5,i) = 26$$

De éstos, diez se obtienen empleando dos indicadores de degradación ($C(5,2)$), diez con tres indicadores de degradación ($C(5,3)$), cinco con cuatro indicadores simultáneos ($C(5,4)$) y, por último, un experimento involucrando los cinco indicadores de degradación empleados ($C(5,5)$).

A continuación se detallan los experimentos llevados a cabo en función del algoritmo de agrupación empleado.

ENTROPÍA AHUMADA IV BAEZA I - ANÁLISIS 2 CLUSTERS		
	Distancia Euclídea	Distancia Cityblock
snr - verosim	1.2874	**0.9041**
snr - verosim - kcep	1.2872	0.9924
snr - verosim - klpc - kcep	1.2872	0.9882
snr - verosim - klpc - kcep - p563	1.0491	1.0776

Cuadro 4.8: Análisis con dos grupos para el algoritmo K-means empleando AhumadaIV-BaezaI

K-means

Los estudios sobre K-means se basan en diferentes simulaciones empleando agrupaciones de 2, 3, 4 y 9 clusters. A continuación se indican los resultados más relevantes obtenidos para ambas bases de datos. En el anexo "Gráficas adicionales"se pueden consultar otra serie de gráficas, así como las tablas completas.

Ahumada IV - Baeza I

El primer bloque de experimentos realizado se basa en emplear el algoritmo K-means con dos únicas agrupaciones pero variando la distancia empleada (distancia euclídea o distancia cityblock).

En este apartado de la memoria se muestra un experimento por cada conjunto de indicadores de degradación empleados, así el primer experimento se corresponde con los mejores valores de k-means obtenidos empleando tan solo dos indicadores de degradación, el segundo experimento con el que mejor valor de entropía proporciona dentro del conjunto de tres indicadores de degradación, etc.

Como se puede observar en la tabla 4.8, el agrupamiento con una menor entropía es el correspondiente al obtenido por medio de dos indicadores de degradación (snr - verosim) con un valor de 0.9041. Consistente con la hipótesis de que el máximo valor de entropía del agrupamiento debe ser menor que el valor de la entropía de los distintos tipos presentes en la base de datos, en este caso cuatro tipos, lo que da lugar a una entropía de 2.

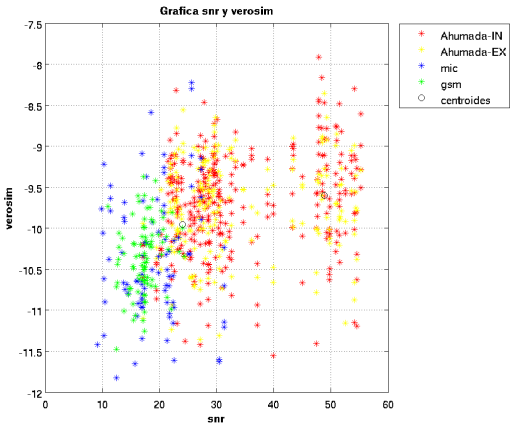
Y en la tabla 4.9 se muestran las gráficas correspondientes al experimento snr-verosim que es el que mejor entropía muestra (entropía más cercana a 0). Esta gráfica muestra en primer lugar los ficheros de audio sin agrupar para en la siguiente gráfica mostrar los mismos ficheros agrupados.

A continuación se puede observar en la tabla 4.10 un nuevo bloque de experimentos, los correspondientes a realizar tres agrupaciones. Los indicadores de degradación mejores prácticamente coinciden con los empleados con dos agrupaciones, con una única excepción, en vez de ser snr-verosim-kcep la mejor combinación de tres elementos, en este caso es snr-verosim-klcp. Esto se debe a la gran similitud existente entre la kurtosis cepstral y la kurtosis de los coeficientes de predicción lineal para la base de datos de la Guardia Civil. En este caso, la combinación que obtiene menor valor de entropía es la formada por los indicadores de degradación snr - verosim - klpc - kcep con un valor de 0.9087.

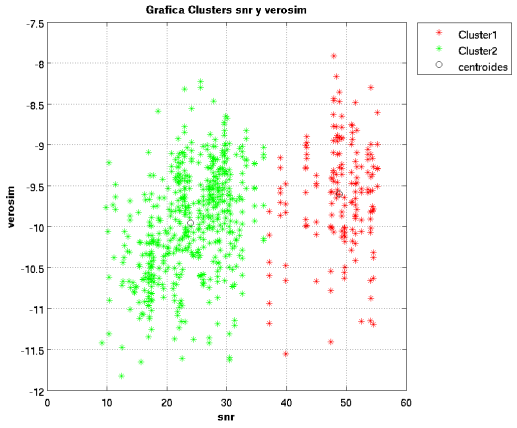
En la tabla 4.11 se observa la representación gráfica de los grupos realizados en el caso de dos únicos indicadores de degradación.

En la tabla 4.12 se presenta un nuevo bloque de experimentos, los correspondientes a realizar cuatro agrupaciones. Los indicadores de degradación coinciden con los presentados para tres agrupaciones. El valor mínimo de entropía (0.9051) se obtiene para snr - verosim - klpc - kcep.

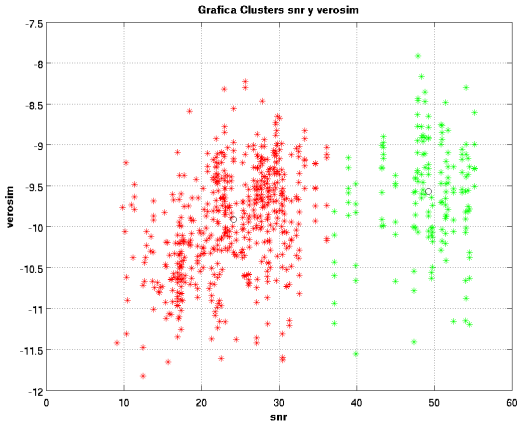
En la tabla 4.13 se observa la representación gráfica de los grupos realizados en el caso de



Representación gráfica snr - verosim



Distancia Euclídea

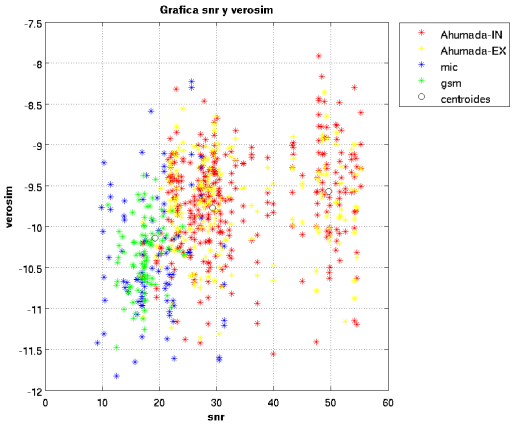


Distancia Cityblock

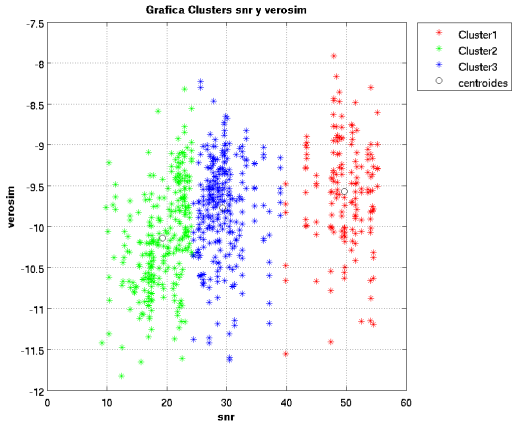
Cuadro 4.9: Snr-Verosim con distancia euclídea y cityblock para AhumadaIV-BaezaI con 2 agrupamientos

ENTROPÍA AHUMADA IV BAEZA I - ANÁLISIS 3 CLUSTERS		
	Distancia Euclídea	Distancia Cityblock
snr - verosim	1.0013	0.9793
snr - verosim - klpc	0.9841	0.9892
snr - verosim - klpc - kcep	1.1351	**0.9087**
snr - verosim - klpc - kcep - p563	1.2086	1.1868

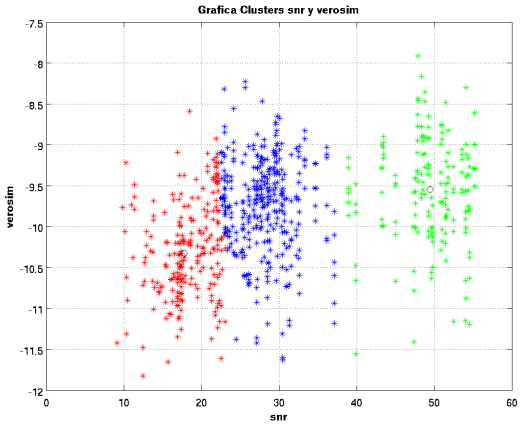
Cuadro 4.10: Análisis con tres grupos para el algoritmo K-means empleando AhumadaIV-BaezaI



Representación gráfica snr - verosim



Distancia Euclídea



Distancia Cityblock

Cuadro 4.11: Snr-Verosim con distancia euclídea y cityblock para AhumadaIV-BaezaI con tres agrupamientos

dos indicadores snr-verosim.

ENTROPÍA AHUMADA IV BAEZA I - ANÁLISIS 4 CLUSTERS		
	Distancia Euclídea	Distancia Cityblock
snr - verosim	1.0227	0.9898
snr - verosim - klpc	0.9815	1.0222
snr - verosim - klpc - kcep	0.9222	**0.9051**
snr - verosim - klpc - kcep - p563	1.0351	0.9731

Cuadro 4.12: Análisis con cuatro grupos para el algoritmo K-means empleando AhumadaIV-BaezaI

Una vez obtenidos todos los resultados para AhumadaIV-BaezaI empleando K-means se realiza un estudio detallado de los valores de entropía empleados para los diferentes agrupamientos realizados (tabla 4.14).

En la mayor parte de los experimentos se puede observar que la menor entropía se alcanza con la distancia cityblock, sin embargo no es un resultado concluyente, ya que existen pocas muestras en el sistema (tan solo 715 ficheros forman parte de la base de datos). El valor de cityblock como la mejor forma de realizar agrupamientos empleando K-means se observa tanto en los valores de las tablas 4.8, 4.10 y 4.12 como en la tabla 4.14.

En la tabla 4.9 se puede observar un buen agrupamiento si sólo se tuvieran en cuenta los dos grandes tipos de audio presentes en la base de datos de la Guardia Civil. Es decir, se puede distinguir una agrupación prácticamente formada por ficheros de la base de datos Ahumada IV y otra agrupación formada por ficheros de BaezaI.

En la tabla 4.13, se observa que se realiza un buen agrupamiento con los ficheros de tipo Baeza GSM ya que su distribución espacial esta más acotada que con los ficheros de tipo Ahumada (tanto ficheros interiores como exteriores).

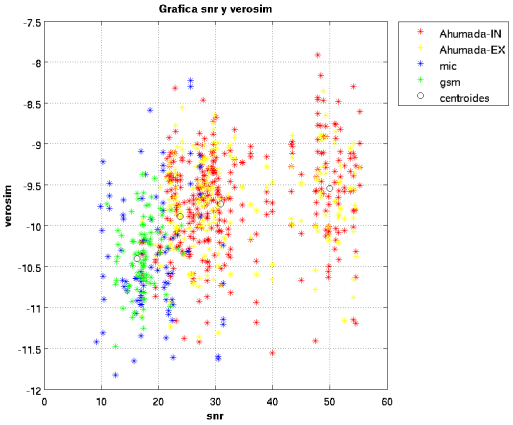
NIST 2008

Para mostrar de forma clara los resultados obtenidos sobre la base de datos NIST 2008 se muestra una tabla resumen en la que figuran los mejores resultados de entropía en función del agrupamiento realizado para cada una de las combinaciones explicadas anteriormente en la sección 4.2.3.

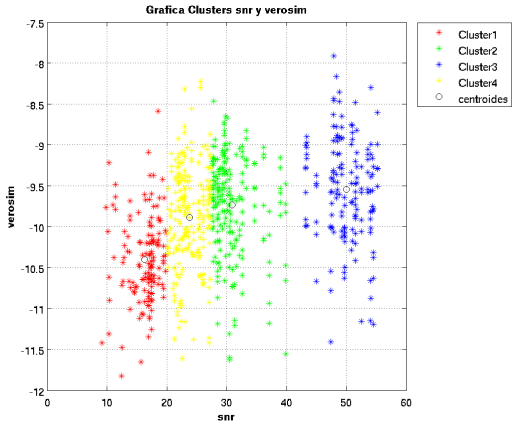
Como se puede observar en la tabla 4.15, el agrupamiento con dos clusters que tiene una menor entropía es el obtenido por medio de cuatro indicadores de degradación (snr - verosim - klpc - kcep) con un valor de 0.46415. Se trata de un valor adecuado ya que el valor máximo de entropía para esta simulación se fija como el $\log_2 2 = 1$.

En la tabla 4.16 se muestran las gráficas correspondientes al experimento snr - kcep. En estas gráficas podemos observar las diferencias obtenidas en función del método de cálculo de distancia empleado. La frontera generada por medio de la distancia Cityblock no es una línea recta como en el caso de la distancia euclídea o la distancia cosine. La existencia de una frontera curva da una idea más precisa sobre los agrupamientos ya que el algoritmo está realizando una mejor segmentación de los datos.

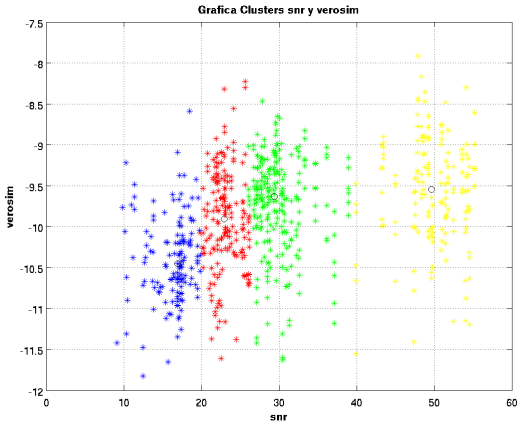
En la tabla 4.13 se muestra unos nuevos experimentos en los que se emplean tres agrupaciones. El agrupamiento con menor entropía es el correspondiente al obtenido por medio de los cinco indicadores de degradación (snr - verosim - klpc - kcep - p563) con un valor de 0.4581. En este caso, el límite teórico está fijado en $\log_2 3 = 1{,}5849$ por lo tanto es consiste con los datos presentados en la tabla 4.13.



Representación gráfica snr - verosim



Distancia Euclídea



Distancia Cityblock

Cuadro 4.13: Snr-Verosim con distancia euclídea y cityblock para AhumadaIV-BaezaI con cuatro agrupamientos

En la tabla 4.14 se muestra la agrupación generada para la combinación snr-kcep. En la figura generada con la distancia cosine se observa muy bien el procedimiento de realización de esta medida de similitud, situando el origen de los vectores empleados en el punto $(0, 0)$.

Para cuatro clusters, el agrupamiento con menor entropía es el correspondiente al obtenido por medio de los cinco indicadores de degradación (snr - verosim - klpc - kcep - p563) con un valor de 0.41661, según lo indicado en la tabla 4.15. A su vez, en la tabla 4.16 se muestran las imágenes correspondientes a este experimento.

En esta última ronda de experimentos, se realizan comprobaciones sólo sobre las distancia euclídea y la distancia cityblock. Empleando el sistema cosine, existen problemas con el código desarrollado, ya que el método kmeans encuentra clusters vacíos debido a la gran cantidad de grupos que se generan para estos experimentos. En la tabla 4.17 se muestran los resultados númericos donde la menor entropía, con valor 0.376, se obtiene a partir de la combinación de las cinco medidas de calidad. En la tabla 4.18, se puede observar al agrupación de los diferentes elementos en la base de datos NIST 2008. En este último caso, las fronteras con líneas irregulares son más notorios que en los casos anteriores. Esto es debido a la existencia de más grupos a la hora de realizar el agrupamiento, por lo que las fronteras en la distancia cityblock pierden tramos rectos para ganar líneas curvas.

Por último, como ya se hizo en el análisis de la base de datos AhumadaIV-BaezaI (sección

4.2.3) se muestra una comparativa con todas las posibles combinaciones generadas a partir de los cinco indicadores de degradación. Los resultados numéricos más importantes se muestran en la tabla 4.23, para la tabla completa consúltese el anexo. Esta tabla se ha generado empleando la distancia Cityblock ya que es la distancia con la que menor entropía se obtiene independientemente del número de grupos empleado.

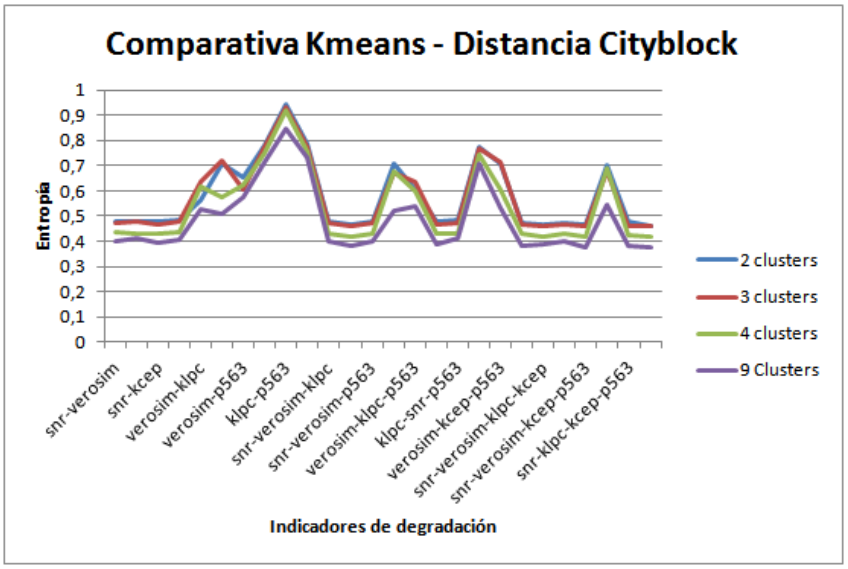


Figura 4.2: Comparativa Algoritmo K-means distancia CityBlock sobre NIST 2008

En la tabla 4.23, la de la comparativa entre clusters, es fácilmente observable como a medida que crece el número de clusters empleados en el sistema, el valor de la entropía disminuye. En el caso de emplear nueve grupos, es decir, el mismo número de grupos que de sensores diferentes en la base de datos, la entropía tiene 0,37352 como valor mínimo y 0,84425 como valor máximo. El valor máximo está muy alejado del valor máximo teórico, en el caso de nueve grupos este valor se fija en $\log_2 9 = 3{,}1699$, por lo que se puede concluir que el método de agrupación está funcionando como se espera.

También es importante destacar, que aunque varíe el número de clusters empleado en los agrupamientos, las combinaciones con menor entropía son siempre las mismas. Es decir, para una agrupamiento realizado empleando sólo dos indicadores de degradación, la combinación *snr - kcep* se mantiene como la que obtiene una entropía más cercana a cero, independientemente del número de clusters en los que agrupar los datos de origen.

La gráfica 4.2 muestra todos los valores de entropía hallados para las agrupaciones formadas por dos, tres, cuatro y nueve grupos y para todas las posibles combinaciones generadas en los experimentos. Se puede observar como a medida que aumenta el número de agrupaciones realizadas disminuye el valor de entropía.

Como conclusión final de esta ronda de experimentos se puede decir que se ha observado que la distancia Cityblock es la que mejor se comporta en cuanto a términos de entropía. Los valores de entropía más bajos se corresponden con los experimentos en los que se han generado nueve agrupaciones que, a su vez, es el número de tipos distintos de fichero presentes en la base de datos NIST 2008. La entropía más baja de todos los experimentos generados hasta esta parte del proyecto final de carrera tiene un valor de 0,376 obtenida para la distancia cityblock en el experimento que engloba todos los indicadores de degradación. Esta entropía es un valor obtenido muy bueno ya que está muy cercana al 0 y, a su vez, está muy alejada del valor teórico máximo para agrupaciones con nueve clusters ($log(9)/log(2) = 3{,}1699$). El caso peor para las agrupaciones generadas por K-means y distancia cityblock obtiene una entropía de 0,94242 (en

el experimento klpc-p563 para dos agrupamientos) pero aún así, este valor entra dentro del rango de valores teóricos para esa agrupación ($log(2)/log(2) = 1$).

GMM

En el análisis por GMM se han realizado experimentos sobre la base de datos NIST 2008.

Se desarrollaron dos experimentos diferenciados, por un lado la obtención de la entropía de los grupos y por otro lado la entropía de los distintos tipos de ficheros.

Entropía de un cluster

En este caso es similar al estudio presentado en el método K-means. Se analizan agrupaciones de dos, tres, cuatro y nueve clusters empleando el método GMM. De los posibles experimentos realizados se escoge la combinación de dos indicadores de degradación que mejor resultado ofrece (menor valor de entropía), la mejor de tres, la mejor de cuatro y, por último, la combinación de cinco.

En la tabla 4.24 se observa el análisis realizador para dos clusters. Las entropías son bastante cercanas a 0. En este caso el valor óptimo se encuentra en la combinación snr verosim klpc kcep con un valor de 0.43830.

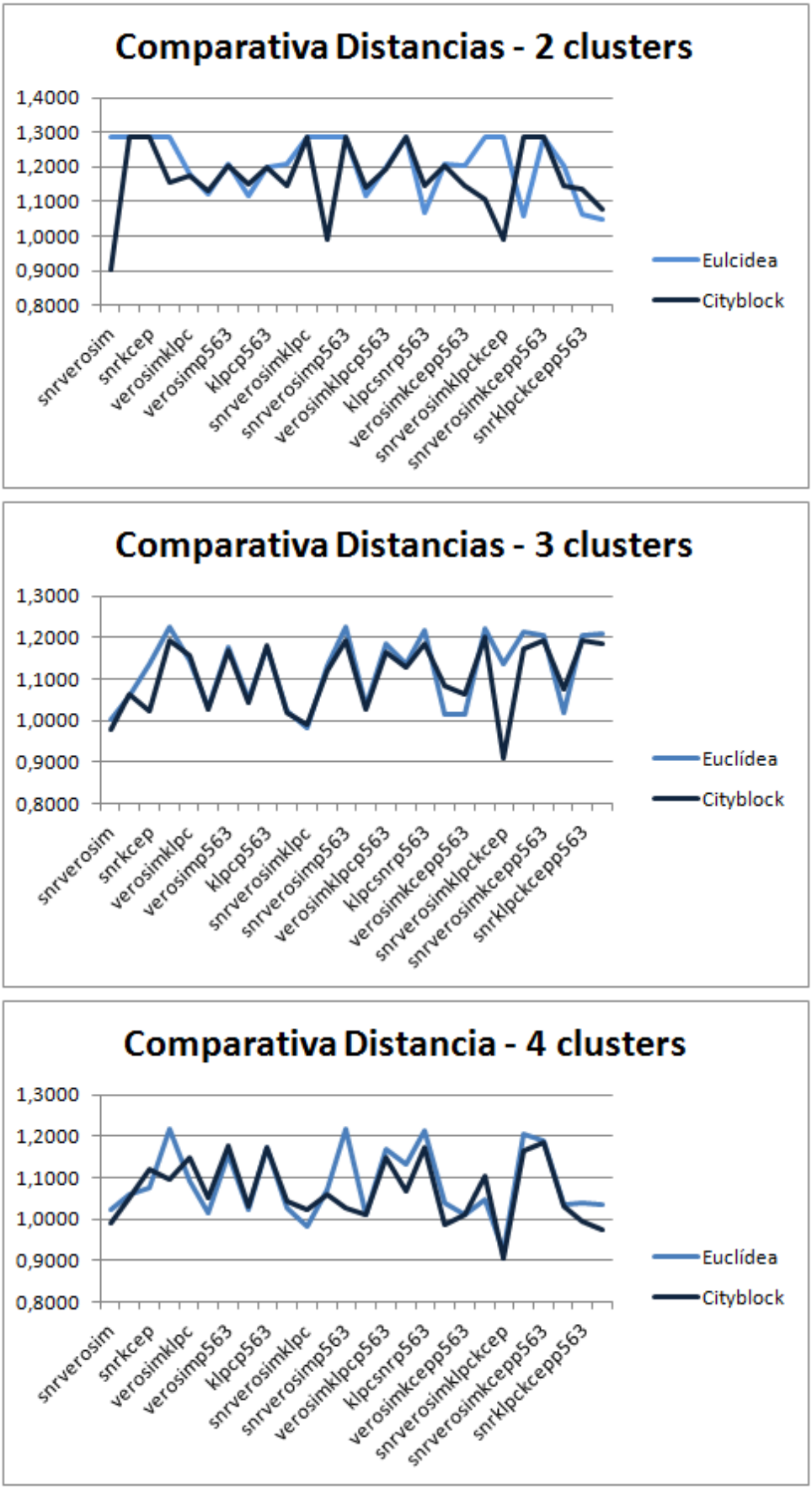
En la tabla 4.25 se muestra el agrupamiento para la combinación de dos indicadores de degradación (snr kcep) con una frontera claramente definida. Esta frontera presenta una gran similitud con la forma de una distribución gaussiana. Es fácilmente observable como las dos gaussianas se cruzan debido a la existencia de una pequeña agrupación de color verde separada del centroide por la gaussiana que da origen al otro grupo.

En la tabla 4.26 se realiza el mismo tipo de experimento, pero variando el número de grupos. De esta forma se obtiene una agrupación de tres elementos en el que, de nuevo, el grupo generado en el experimento con cuatro indicadores de degradación (snr verosim kcep y P.563) es el que mejor resultado ofrece (entropía de 0.41015).

En la tabla 4.27 se observa la agrupación generada para la combinación snr-kcep. De nuevo se observa que las fronteras se corresponden con los centroides generados y como la curva se asemeja a una distribución gaussiana. Existe dos centroides que no están dentro de su propio cluster, esto es debido a los distintos pesos que obtiene cada gaussiana y a los solapes que se producen entre ellas.

La tabla 4.28 muestra el mismo análisis para el caso de emplear cuatro clusters. En este caso el agrupamiento con menor entropía es el formado por snr verosim kcep con un valor de 0.40454. En la tabla 4.29 se muestran las gráficas correspondientes al agrupamiento snr - kcep con cuatro grupos.

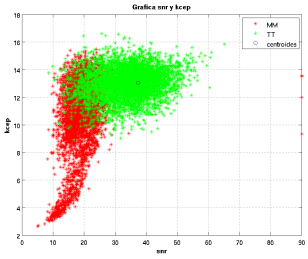
El último de los experimentos consiste en la agrupación con GMM empleando 9 clusters. En la tabla 4.30 se muestran los valores finales, donde el mejor agrupamiento vuelve a ser el obtenido empleando cuatro agrupaciones (snr verosim kcep p563) con una entropía de 0.40240. En la tabla 4.31 se ven las agrupaciones generadas.



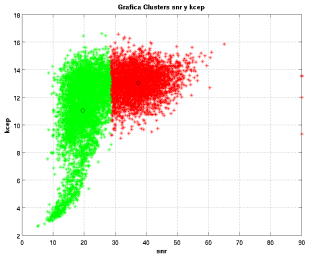
Cuadro 4.14: Comparativa: Entropía en función de los indicadores de degradación empleados.

ENTROPÍA NIST 2008 - ANÁLISIS 2 CLUSTERS			
	Distancia Euclídea	Distancia Cityblock	Distancia Cosine
snr - kcep	0.49555	0.47659	0.81423
snr - verosim - kcep	0.49308	0.46425	0.55038
snr - verosim - klpc - kcep	0.49337	**0.46415**	0.51225
snr - verosim - klpc - kcep - p563	0.49337	0.46267	0.51225

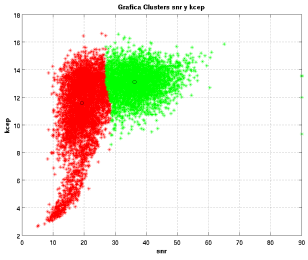
Cuadro 4.15: Análisis con dos grupos para el algoritmo K-means empleando NIST 2008



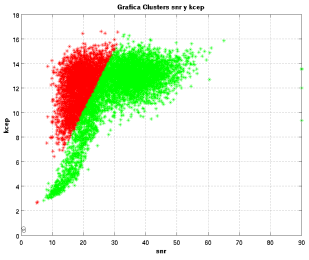
Representación gráfica snr - kcep



Distancia Euclídea



Distancia Cityblock

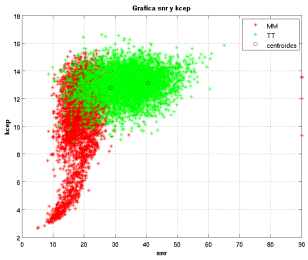


Distancia cosine

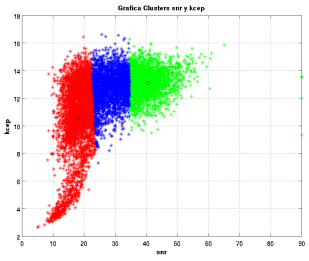
Cuadro 4.16: Snr-Kcep con distancia euclídea, cityblock y cosine para NIST 2008 con dos agrupamientos

ENTROPÍA NIST 2008 - ANÁLISIS 3 CLUSTERS			
	Distancia Euclídea	Distancia Cityblock	Distancia Cosine
snr - kcep	0.48273	0.46905	0.81216
snr - verosim - kcep	0.48128	0.46117	0.53321
snr - verosim - klpc - kcep	0.47937	0.45877	0.49227
snr - verosim - klpc - kcep - p563	0.48022	**0.4581**	0.4754

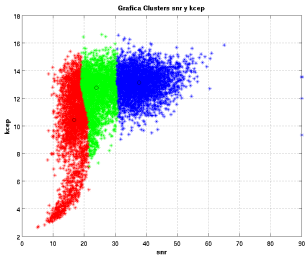
Cuadro 4.17: Análisis con 3 grupos para el algoritmo K-means empleando NIST 2008



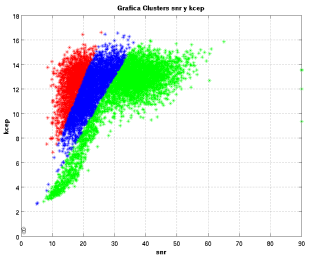
Representación gráfica snr - kcep



Distancia Euclídea



Distancia Cityblock

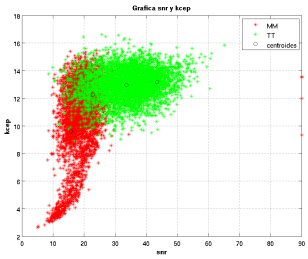


Distancia cosine

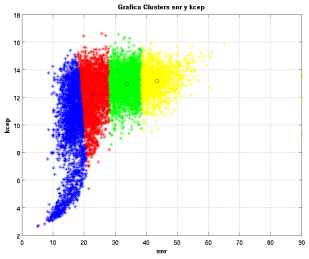
Cuadro 4.18: Snr-Kcep con distancia euclídea, cityblock y cosine para NIST 2008 con 3 agrupamientos

ENTROPÍA NIST 2008 - ANÁLISIS 4 CLUSTERS			
	Distancia Euclídea	Distancia Cityblock	Distancia Cosine
snr - kcep	0.43601	0.42709	0.79688
snr - verosim - kcep	0.43461	0.41927	0.47114
snr - verosim - klpc - kcep	0.43484	0.41685	0.45648
snr - verosim - klpc - kcep - p563	0.43474	**0.41661**	0.45397

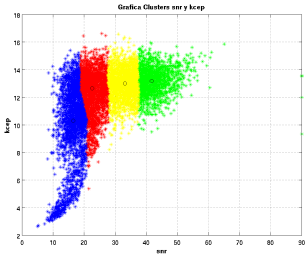
Cuadro 4.19: Análisis con cuatro grupos para el algoritmo K-means empleando NIST 2008



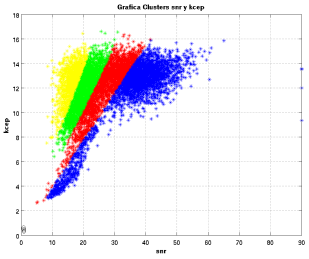
Representación gráfica snr - kcep



Distancia Euclídea



Distancia Cityblock

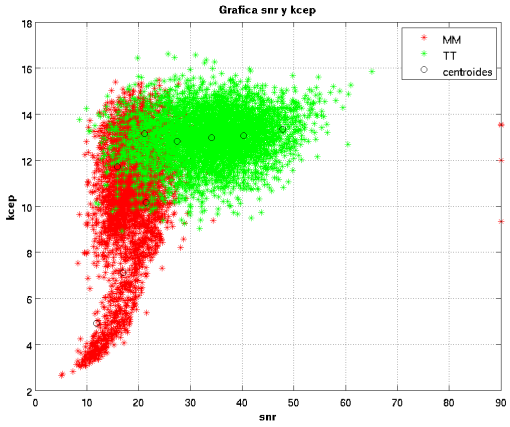


Distancia cosine

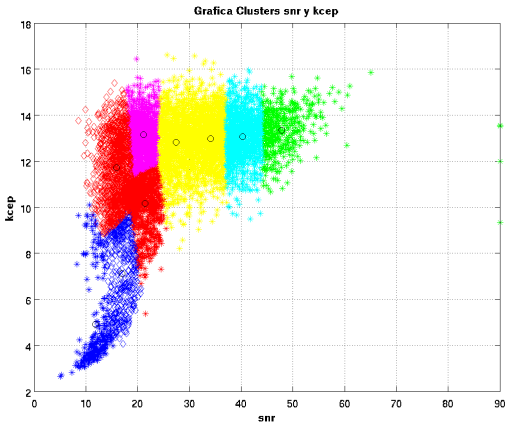
Cuadro 4.20: Snr-Kcep con distancia euclídea, cityblock y cosine para NIST 2008 con cuatro agrupamientos

ENTROPÍA NIST 2008 - ANÁLISIS 9 CLUSTERS		
	Distancia Euclídea	Distancia Cityblock
snr - kcep	0.3778	0.39119
snr - verosim - kcep	0.39113	0.37949
snr - verosim - klpc - kcep	0.38424	0.38778
snr - verosim - klpc - kcep - p563	0.38556	**0.376**

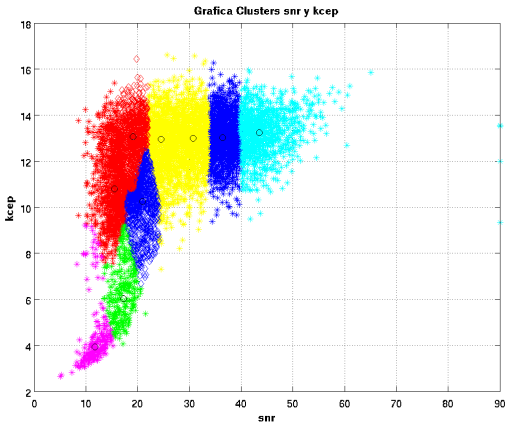
Cuadro 4.21: Análisis con nueve grupos para el algoritmo K-means empleando NIST 2008



Representación gráfica snr - kcep



Distancia Euclidea



Distancia Cityblock

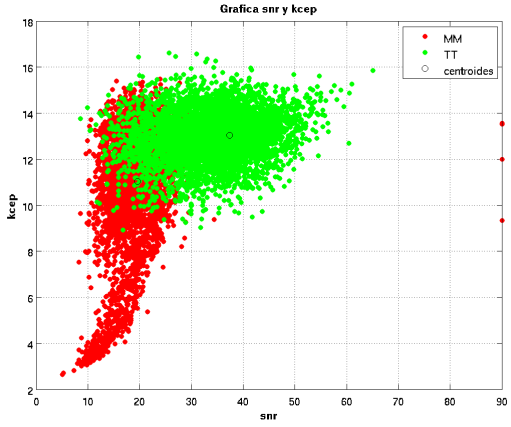
Cuadro 4.22: Snr-Kcep con distancia euclídea y cityblock para NIST 2008 con nueve agrupamientos

COMPARATIVA Kmeans - Distancia Cityblock				
	2 clusters	3 clusters	4 clusters	9 Clusters
snr-kcep	0,47659	0,46905	0,42709	0,39119
snr-verosim-kcep	0,46425	0,46117	0,41927	0,37949
snr-verosim-klpc-kcep	0,46415	0,45877	0,41685	0,38778
snr-verosim-klpc-kcep-p563	0,46267	0,4581	0,41661	0,376

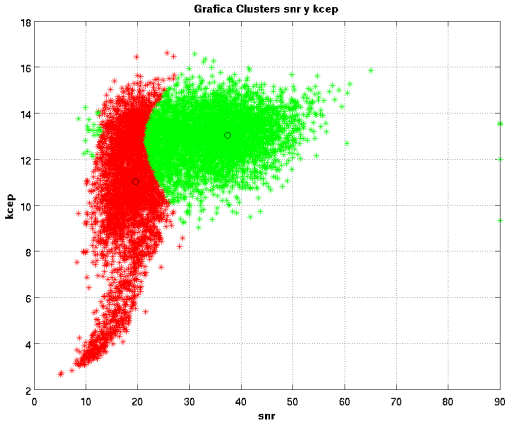
Cuadro 4.23: Entropías obtenidas por K-means (distancia Cityblock) en NIST 2008

ENTROPÍA NIST 2008 - ANÁLISIS 2 CLUSTERS	
snr - kcep	0.54935
snr - verosim - klpc	0.44592
snr - verosim - klpc - kcep	**0.43830**
snr - verosim - klpc - kcep - p563	0.52791

Cuadro 4.24: Análisis con dos grupos para el algoritmo GMM empleando NIST 2008



Gráfica inicial

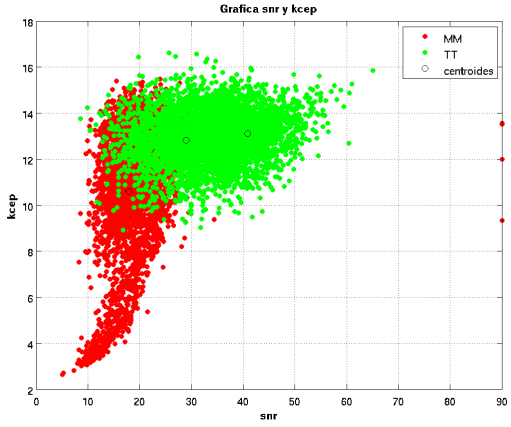


Gráfica dos clusters por GMM

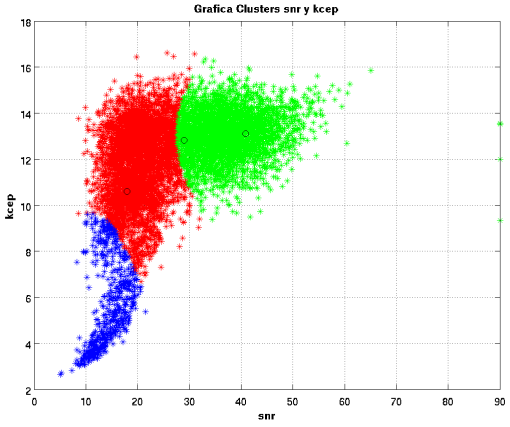
Cuadro 4.25: Snr-Kcep con algoritmo GMM sobre NIST 2008 con dos agrupamientos

ENTROPÍA NIST 2008 - ANÁLISIS 3 CLUSTERS	
snr - kcep	0.46588
snr - verosim - kcep	0.44420
snr - verosim - kcep - p563	0.41015
snr - verosim - klpc - kcep - p563	0.84004

Cuadro 4.26: Análisis con tres grupos para el algoritmo GMM empleando NIST 2008



Gráfica inicial

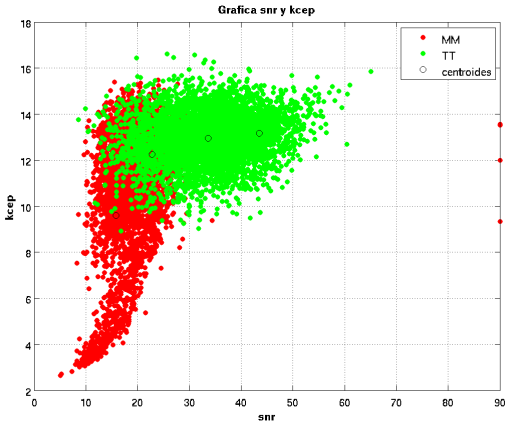


Gráfica tres clusters por GMM

Cuadro 4.27: Snr-Kcep con algoritmo GMM sobre NIST 2008 con 3 agrupamientos

ENTROPÍA NIST 2008 - ANÁLISIS 4 CLUSTERS	
snr - kcep	0.46013
snr - verosim - kcep	**0.40454**
snr - verosim - kcep - p563	0.40793
snr - verosim - klpc - kcep - p563	0.48414

Cuadro 4.28: Análisis con cuatro grupos para el algoritmo GMM empleando NIST 2008



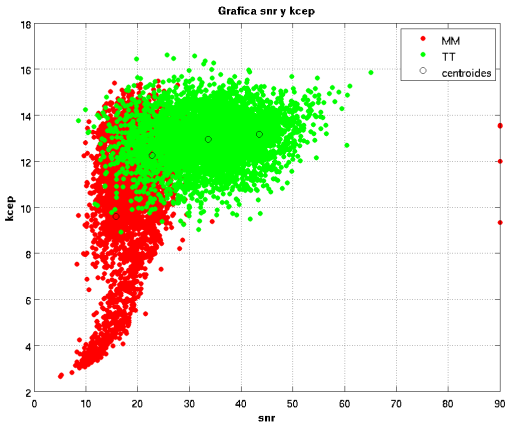
Gráfica inicial

Gráfica cuatro clusters por GMM

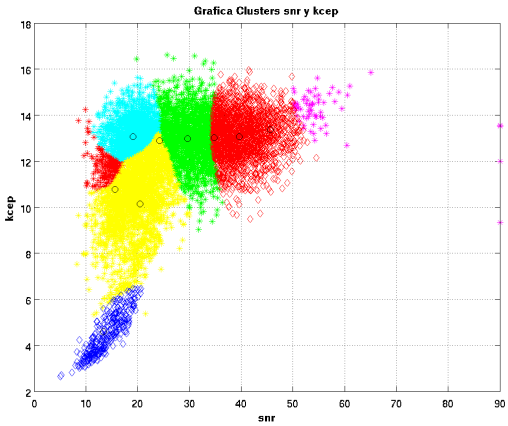
Cuadro 4.29: Snr-Kcep con algoritmo GMM sobre NIST 2008 con cuatro agrupamientos

ENTROPÍA NIST 2008 - ANÁLISIS 9 CLUSTERS	
snr - kcep	0.41605
snr - verosim - kcep	0.41785
snr - verosim - kcep - p563	**0.40240**
snr - verosim - klpc - kcep - p563	0.42768

Cuadro 4.30: Snr-Kcep con algoritmo GMM sobre NIST 2008 con nueve agrupamientos



Gráfica inicial



Gráfica nueve clusters por GMM

Cuadro 4.31: Snr-Kcep con algoritmo GMM sobre NIST 2008 con nueve agrupamientos

COMPARATIVA GMM				
	2 Clusters	3 Clusters	4 Clusters	9 Clusters
snrkcep	0,54935	0,46588	0,46013	0,41605
snrverosimklpc	0,44592	0,44420	0,53350	0,39742
snrverosimkcep	0,77354	0,78864	0,40454	0,41785
snrverosimklpckcep	0,43830	0,73285	0,43871	0,41331
snrverosimkcepp563	0,47971	0,54209	0,40793	0,40240
snrverosimklpckcepp563	0,52791	0,84004	0,48414	0,42765

Cuadro 4.32: Comparativa GMM para distinto número de agrupamiento

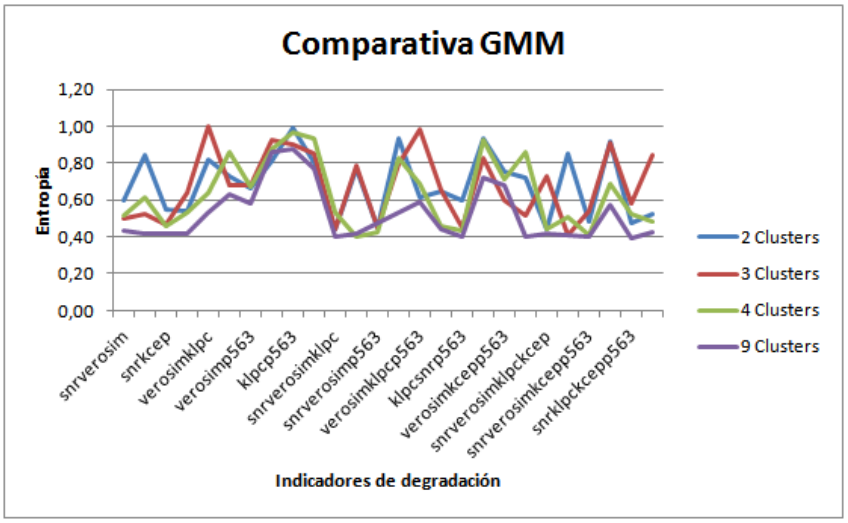


Figura 4.3: Comparativa Algoritmo GMM sobre NIST 2008

En la comparativa presente en la tabla 4.32 se muestran los valores de entropía más importantes variando el número de clusters empleados. En este caso se obtienen más combinaciones representativas debido a que durante los cambios de número de agrupamientos las mejores combinaciones de indicadores de degradación no se han repetido siempre como sucedía en el caso K-means. Es decir, en este caso en el caso de dos clusters, los indicadores de degradación (tres) son snr-verosim-klpc, mientras que para el resto de clusters es snr - verosim - kcep. Lo mismo sucede para el caso de cuatro indicadores de degradación. Esto puede indicar un peor agrupamiento, ya que los indicadores de degradación que mejor resultados de entropía obtienen son dependientes del número de agrupaciones que se quieran generar.

En la figura 4.3 se puede ver cómo a medida que aumenta el número de grupos formados disminuye el valor de la entropía, sin embargo esta variación no es tan significativa como la sucedida en el caso del algoritmo K-means. El valor de entropía mejor, aquel más cercano a 0, es el obtenido en el experimento snr - verosim - klpc para nueve agrupaciones. Este valor de entropía es 0.39742 muy alejado del valor máximo teórico ($log(9)/log(2) = 3{,}1699$). El caso peor se obtiene para la combinación de cinco indicadores de degradación en el caso de realizar tres agrupamientos distintos: 0,84004. Este valor más alto de entropía es un caso extraño, ya que rompe la tendencia de que al incrementar el número de clusters mejora (tiene un cluster más que la agrupación realizada con dos que tiene una entropía de 0,52791) y, a su vez, es el resultado del experimento con todos los indicadores de degradación que también debería obtener un valor de entropía más bajo que el resto. Este hecho da una idea de la importancia de la matriz de inicialización de los centroides del método GMM, ya que probablemente este valor obtenido haya sido debido a una mala inicialización.

Comparativa K-means y GMM

Después del trabajo realizado en la sección anterior, en este nuevo apartado se muestra una comparativa entre los métodos de agrupamiento empleados. La medida de similitud empleada en el agrupamiento K-means es la distancia Cityblock.

En la tabla 4.33 se muestran los valores de entropía encontrados para los diversos experimentos realizados, tanto para el algoritmo K-means (con distancia Cityblock) como para el algoritmo GMM. Se han escogido los indicadores de degradación que mejor funcionan en el caso de la agrupación por K-means. Los experimentos mostrados en dicha tabla son los siguientes:

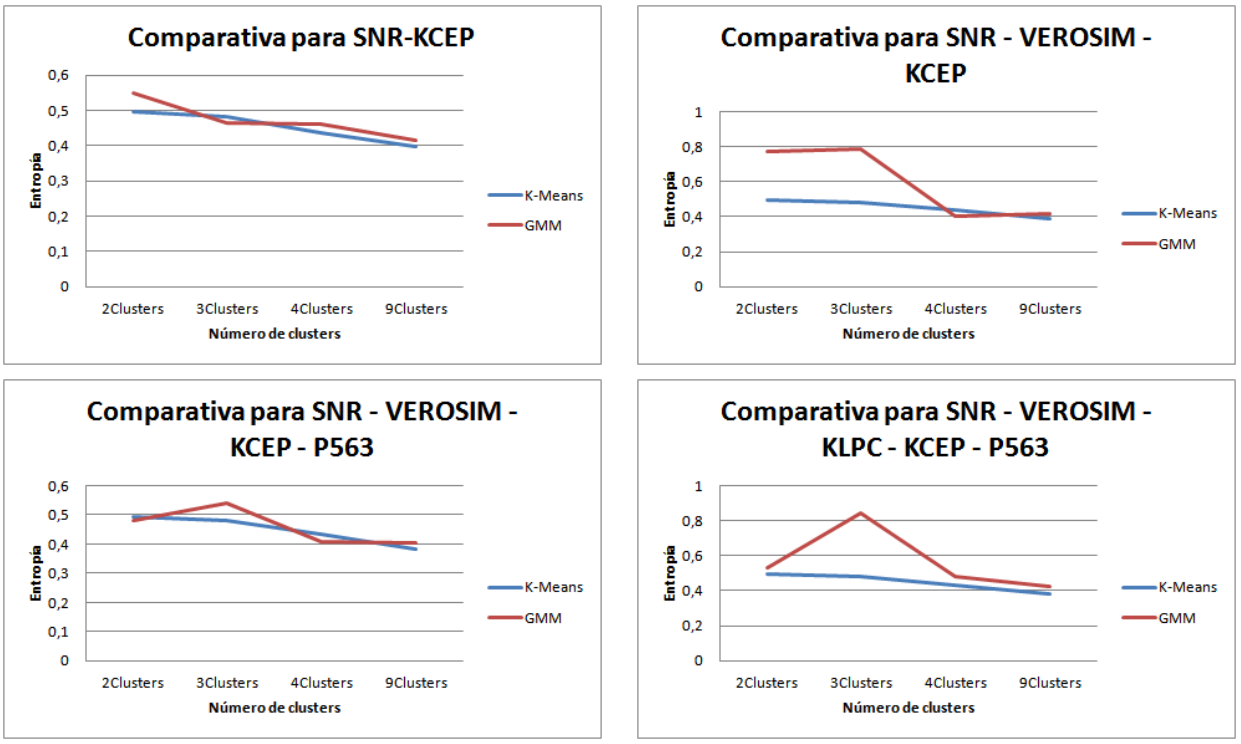
- snr - kcep
- snr - verosim - kcep
- snr - verosim - kcep - p563
- snr - verosim - klpc - kcep - p563

2 Clusters		3 Clusters		4 Clusters		9 Clusters	
K-means	GMM	K-means	GMM	K-means	GMM	K-means	GMM
0,4956	0,54935	0,4827	0,46588	0,436	0,46013	**0,3968**	0,41605
0,4946	0,77354	0,4813	0,78864	0,4363	0,40454	**0,3864**	0,41785
0,4931	0,47971	0,4812	0,54209	0,434	0,40793	**0,3835**	0,40240
0,4934	0,52791	0,4803	0,84004	0,432	0,48414	**0,3839**	0,42765

Cuadro 4.33: Comparativa K-means y GMM para distintas combinaciones

Como se puede observar las agrupaciones con nueve clusters son las que menor entropía proporcionan independientemente del sistema con el que se haya realizado la prueba. Para el caso K-means la mejor entropía es la obtenida con cuatro indicadores de degradación (0,3835) igual que en el caso GMM (0.40240). Esta tabla indica que las agrupaciones formadas por indicadores de degradación klpc - p563 no aportan buen resultado, ya que nunca obtienen un buen valor de entropía.

A continuación, en la tabla 4.34 se muestra una gráfica de la evolución de los valores de entropía tanto para K-means como para GMM. Los valores de esta tabla son los correspondientes con la tabla 4.33. Es importante destacar como la variación de la entropía en el caso de realizar los agrupamientos por medio de K-means es prácticamente lineal, a diferencia de lo que sucede con GMM. En ambos casos se puede comprobar como a medida que aumenta el número de agrupamientos la entropía disminuye.



Cuadro 4.34: Evolución de la entropía para K-means y GMM variando el número de agrupamientos.

4.3. Estudio de agrupamiento en función de los tipos de ficheros

Sobre la base de datos NIST 2008 se ha realizado un estudio paralelo al de agrupamiento descrito en el apartado anterior. En este nuevo estudio se pretende reflejar cómo de bueno es un agrupamiento en función de la distribución de los tipos de ficheros existentes en la base de datos. Es decir, se analiza si todos los elementos de un mismo tipo están agrupados en el mismo agrupamiento.

4.3.1. Algoritmo K-means

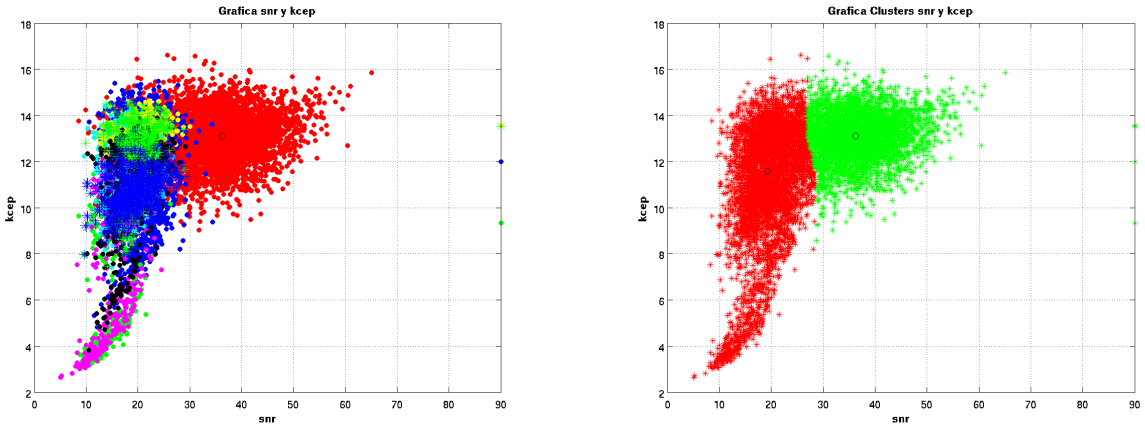
En el análisis siguientes se muestran agrupaciones que mejores resultados de entropía total proporcionan. Estas combinaciones son las siguientes:

- snr - kcep
- snr - verosim - kcep
- snr - verosim - klpc - kcep
- snr - verosim - klpc - kcep - p563

Los valores obtenidos para este experimento se encuentran en la tabla 4.38. Las gráficas generadas, en este caso para dos indicadores de degradación y para tres, se muestran en la tabla 4.35. Esta segmentación es muy buena, ya que existen dos tipos de ficheros (mic 05 y mic 07) con entropía 0 lo que quiere decir que todos los elementos de ese mismo tipo forman parte del mismo cluster (aunque hay que distinguir que si se mira la agrupación desde el punto del cluster,

Análisis Entropía dos Clusters									
phonecall	mic02	mic03	mic05	mic07	mic08	mic09	mic12	mic13	TOTAL
0,7762	0,2358	0,0132	0,0000	0,0000	0,1693	0,0071	0,0805	0,0520	0,4462
0,7581	0,2521	0,0132	0,0000	0,0000	0,1568	0,0071	0,0475	0,0520	0,4363
0,7598	0,2217	0,0132	0,0000	0,0000	0,1693	0,0071	0,1030	0,0520	0,4372
0,7514	0,2370	0,0132	0,0000	0,0000	0,2081	0,0071	0,1030	0,0520	0,4363

Cuadro 4.35: Entropía para cada tipo de fichero generados dos agrupamientos por K-means



Cuadro 4.36: Gráficas por tipo de fichero para dos agrupamientos por K-means

este agrupamiento puede estar formado por todos los elementos de tipo mic 05 (o mic 07) y a su vez contener algún fichero de otro tipo) independientemente de los indicadores de degradación empleados. Desde el punto de vista de la representación gráfica de los agrupamientos generados la agrupación también es muy buena ya que la mayoría de ficheros rojos presentes en la primera gráfica de la tabla 4.36 se agrupan juntos en la segunda gráfica de dicha tabla.

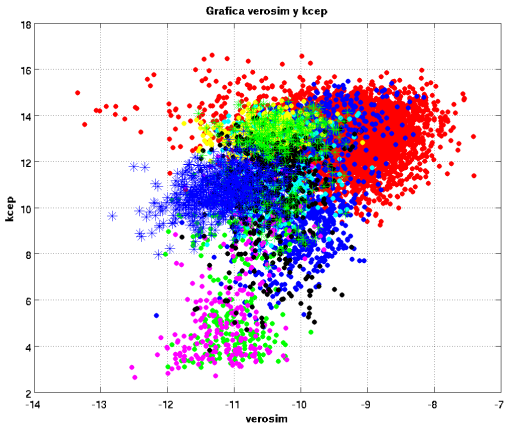
Para tres clusters las combinaciones que mejores resultados de entropía total presentan son las siguientes:

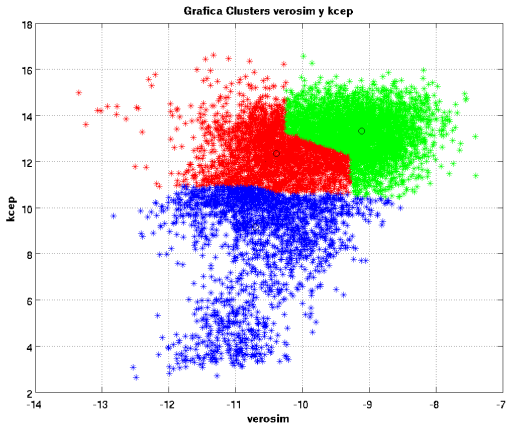
- verosim - kcep
- verosim - kcep - p563
- snr - verosim - kcep - p563
- snr - verosim - klpc - kcep - p563

En la tabla 4.37 están los valores obtenidos para el experimento descrito anteriormente. Se puede observar como los valores de la entropía aumentan mucho en todos los experimentos generados. El valor de entropía más bajo es el correspondiente con dos clusters 0,4782 y, sin embargo, sigue siendo un valor alto en comparación con los experimentos mostrados en la tabla 4.35. Las figuras correspondientes al agrupamiento verosim-kcep se muestran en la tabla 4.38.

Análisis Entropía 3 Clusters									
phonecall	mic02	mic03	mic05	mic07	mic08	mic09	mic12	mic13	TOTAL
0,9146	1,5043	0,2788	1,1562	0,6173	0,6319	1,0654	1,3618	0,9771	0,9689
0,4782	1,2439	0,9867	0,9191	1,1112	0,5167	1,1463	0,9212	0,2447	0,6912
1,1418	1,0749	0,1743	0,6659	0,1891	1,0006	0,8056	1,0112	0,7564	0,9584
1,1416	1,0771	0,1743	0,6591	0,2422	0,9883	0,7843	1,0114	0,7876	0,9607

Cuadro 4.37: Entropía para cada tipo de fichero generados tres agrupamientos por K-means





Cuadro 4.38: Gráficas por tipo de fichero para tres agrupamientos por K-means

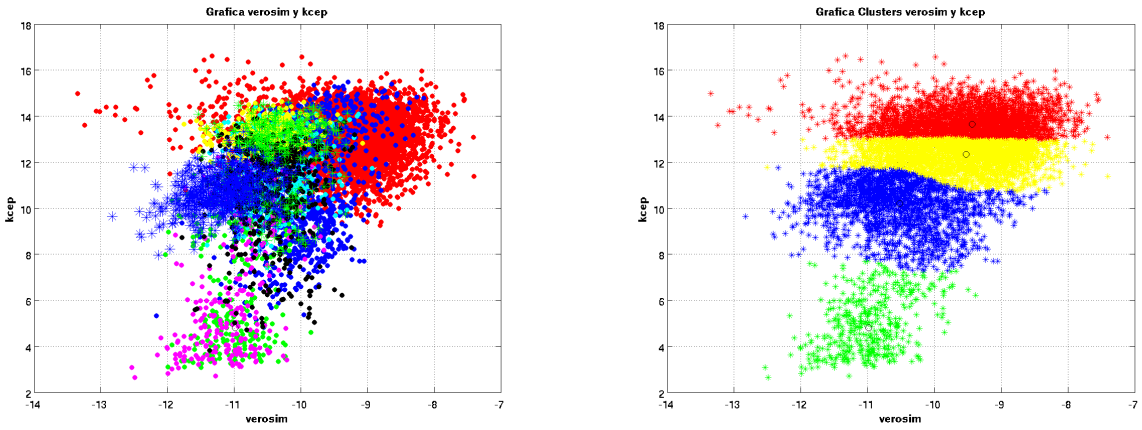
Para el estudio realizado sobre cuatro clusters, las mejores combinaciones de indicadores de degradación son las siguientes:

- verosim - kcep
- verosim - kcep - p563
- verosim - klpc - kcep - p563
- snr - verosim - klpc - kcep - p563

Análisis Entropía 4 Clusters									
phonecall	mic02	mic03	mic05	mic07	mic08	mic09	mic12	mic13	TOTAL
1,2010	1,6884	0,9908	1,5046	1,2971	1,0228	1,4624	1,5168	0,680	1,257
1,2008	1,6901	0,9931	1,4241	1,3033	1,0587	1,4355	1,5306	0,663	1,253
1,2614	1,5564	1,1490	1,3742	1,2610	0,8624	1,3871	1,5443	0,792	1,269
1,6906	1,1333	0,2303	0,8649	0,3059	0,9949	0,9470	1,0529	0,987	1,295

Cuadro 4.39: Entropía para cada tipo de fichero generados cuatro agrupamientos por K-means

En la tabla 4.39 se muestran los valores correspondientes al agrupamiento con cuatro clusters. Los valores de entropía aumentan mucho más y los tipos de fichero mic-05 y mic-07 ya no presentan entropía 0, lo que quiere decir que están divididos en varios clusters. Las gráficas correspondientes a este experimento se muestran en la tabla 4.40.



Cuadro 4.40: Gráficas por tipo de fichero para cuatro agrupamientos por K-means

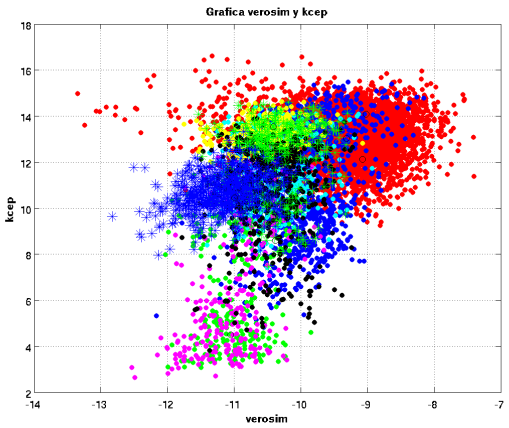
Para el último caso analizado realizando un clustering de nueve agrupaciones se obtienen los siguientes resultados (tabla 4.41 y 4.42):

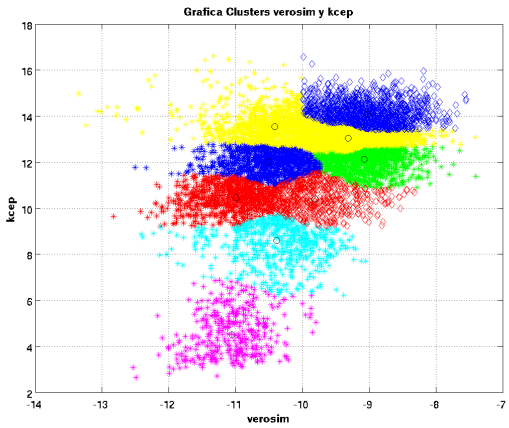
- verosim - kcep
- verosim - klpc - kcep
- verosim - klpc - kcep - p563
- snr - verosim - klpc - kcep - p563

Análisis Entropía 9 Clusters									
phonecall	mic02	mic03	mic05	mic07	mic08	mic09	mic12	mic13	TOTAL
2,3150	2,7346	1,6310	2,3740	1,7600	1,1312	2,4724	2,3746	1,3188	2,2008
2,4816	2,2138	1,0271	2,2211	1,8294	1,2975	2,0724	2,4433	1,2609	2,1803
2,4587	2,2816	1,0262	2,1573	1,8030	1,4226	2,0970	2,4638	1,3011	2,1815
2,6296	2,1758	1,4866	1,7421	1,5827	1,6215	2,1387	1,9872	1,8209	2,2665

Cuadro 4.41: Entropía para cada tipo de fichero generados nueve agrupamientos por K-means

Como se puede comprobar el valor de la entropía por tipo de fichero aumenta a medida que se realizan más agrupaciones. Las mejores agrupaciones están presentes en el caso de realizar solo dos agrupamientos, con tipos de fichero (mic05 y mic07) pertenecientes íntegramente al mismo cluster. Sin embargo, a medida que vamos segmentando más los datos, al algoritmo K-means le resulta más difícil realizar una buena agrupación.





Cuadro 4.42: Gráficas por tipo de fichero para nueve agrupamientos por K-means

4.3.2. Algoritmo GMM

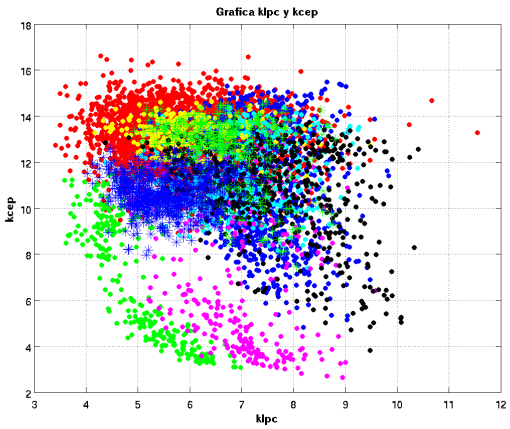
Para que la comparación sea más óptima también se han realizado un análisis similar por medio de GMM. Durante la siguiente sección se detallan los resultados obtenidos.

Empezando por dos agrupamientos se obtienen los siguientes resultados. Los indicadores de degradación que mejores resultados ofrecen son los siguientes:

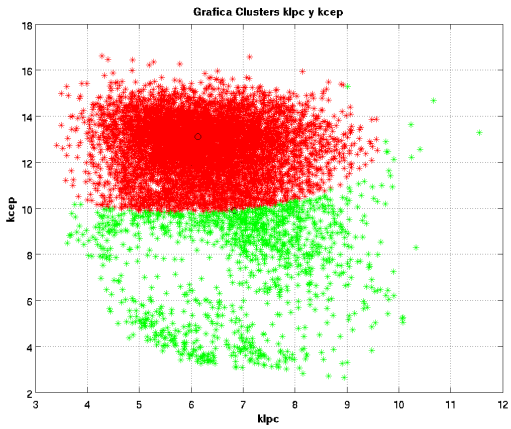
- klpc - kcep
- klpc - kcep - p563
- verosim - klpc - kcep - p563
- snr - verosim - klpc - kcep - p563

Análisis Entropía dos Clusters									
phonecall	mic02	mic03	mic05	mic07	mic08	mic09	mic12	mic13	TOTAL
0,0639	0,9821	0,5940	0,4834	0,7073	0,0000	0,9253	0,5064	0,5467	0,3433
0,0776	0,9780	0,6272	0,4231	0,7145	0,0000	0,9042	0,4635	0,4136	0,3371
0,0461	0,9492	0,8526	0,1838	0,8161	0,0000	0,8466	0,3637	0,1333	0,2940
0,7921	0,9480	0,9927	0,0294	0,8812	0,0131	0,7392	0,2867	0,0000	0,6659

Cuadro 4.43: Entropía para cada tipo de fichero generados dos agrupamientos por GMM



Gráfica inicial



Gráfica 2 clusters por Kmeans

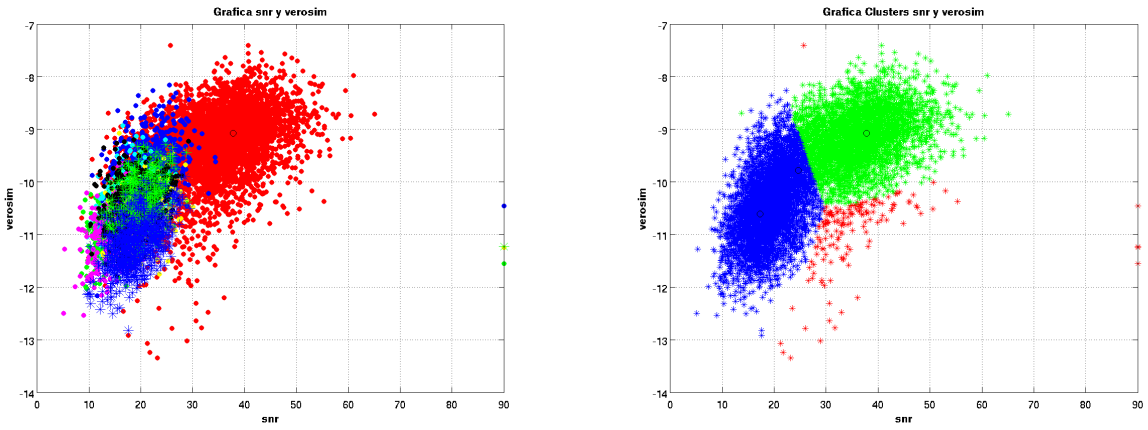
Cuadro 4.44: Gráficas por tipo de fichero para dos agrupamientos por GMM

A continuación realizamos una nueva agrupación empleando tres clusters. Los indicadores de degradación obtenidos son los siguientes:

- snr - verosim
- snr - verosim - klpc
- snr - verosim - klpc - p563
- snr - verosim - klpc - kcep - p563

Análisis Entropía 3 Clusters									
phonecall	mic02	mic03	mic05	mic07	mic08	mic09	mic12	mic13	TOTAL
0,8847	0,4672	0,0132	0,0000	0,0000	0,1278	0,0515	0,0926	0,0634	0,5297
0,8015	0,4700	0,0829	0,0073	0,0000	0,1432	0,0515	0,1819	0,0676	0,4957
0,6194	0,7522	0,0822	0,0528	0,3580	0,4023	0,2215	0,3451	0,0807	0,4825
1,1416	1,0447	0,0132	0,6497	0,0637	0,9410	0,3968	1,0234	0,2516	0,8831

Cuadro 4.45: Entropía para cada tipo de fichero generados tres agrupamientos por GMM



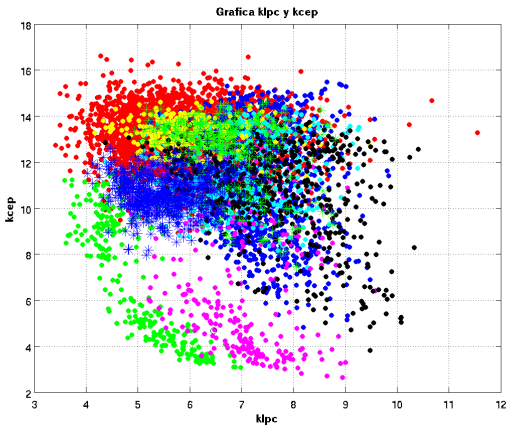
Cuadro 4.46: Gráficas por tipo de fichero para tres agrupamientos por GMM

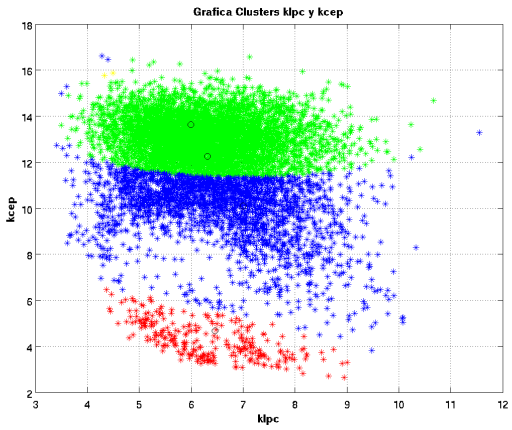
En el siguiente experimento presentado se realizan agrupaciones de 4 clusters:

- klpc - kcep
- klpc - kcep - p563
- verosim - klpc - kcep - p563
- snr - verosim - klpc - kcep - p563

Análisis Entropía cuatro Clusters									
phonecall	mic02	mic03	mic05	mic07	mic08	mic09	mic12	mic13	TOTAL
0,4274	0,9912	0,9988	0,9829	1,3941	0,0362	0,9714	0,7595	0,7182	0,6422
0,4380	1,0168	1,1634	0,8269	1,4454	0,0408	1,0063	0,6867	0,9937	0,6669
0,3682	1,0526	0,9998	0,9401	1,2616	0,8935	0,7864	0,9471	0,1638	0,6148
1,3236	1,6435	1,0108	0,2446	0,9857	0,8077	0,8033	0,7605	0,1372	1,1095

Cuadro 4.47: Entropía para cada tipo de fichero generados cuatro agrupamientos por GMM





Cuadro 4.48: Gráficas por tipo de fichero para cuatro agrupamientos por GMM

- snr - klpc
- snr - kcep - p563
- snr - klpc - kcep - p563
- snr - verosim - klpc - kcep - p563

[h]

Análisis Entropía nueve Clusters									
phonecall	mic02	mic03	mic05	mic07	mic08	mic09	mic12	mic13	TOTAL
1,2233	1,4859	1,5596	0,9885	1,1354	1,2761	1,1684	1,2019	1,2816	1,2514
0,7444	1,4118	1,2757	1,0747	1,6023	1,5400	1,0203	1,5441	0,5432	0,9836
1,4085	2,1243	1,1584	1,6758	1,8106	1,4532	2,0782	1,9355	1,2203	1,5658
1,1191	1,5918	1,0435	1,6146	1,5726	1,0684	1,6336	1,7628	0,5402	1,2425

Cuadro 4.49: Entropía para cada tipo de fichero generados nueve agrupamientos por GMM

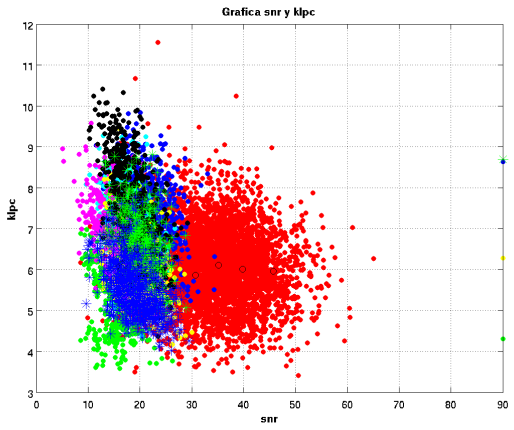
Por último, en el caso de los nueve grupos distintos los valores obtenidos son los que se presentan a continuación:

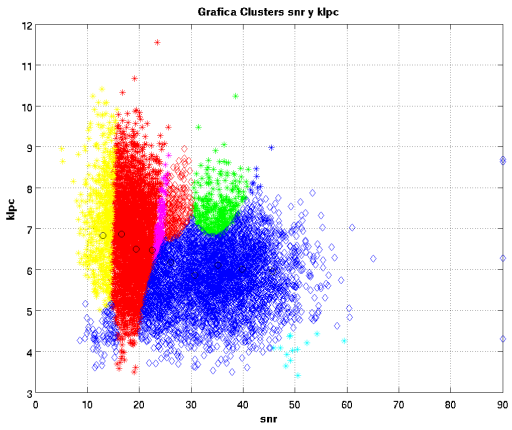
Como se ha podido observar en las tablas 4.43 a 4.50 la entropía por tipo de fichero de los agrupamientos generados por GMM aumenta a la vez que se aumenta el número de clusters.

4.3.3. Comparativa K-means-GMM y conclusiones

Por último, en la tabla 4.54 y 4.55 y figuras se puede observar el valor de la entropía de fichero tanto para GMM como Kmeans de forma simultánea. Se han escogido los indicadores de degradación que mejor han funcionado sobre pruebas anteriores:

- snr - kcep
- snr - verosim - kcep
- snr - verosim - kcep - p563
- snr - verosim - klpc - kcep - p563



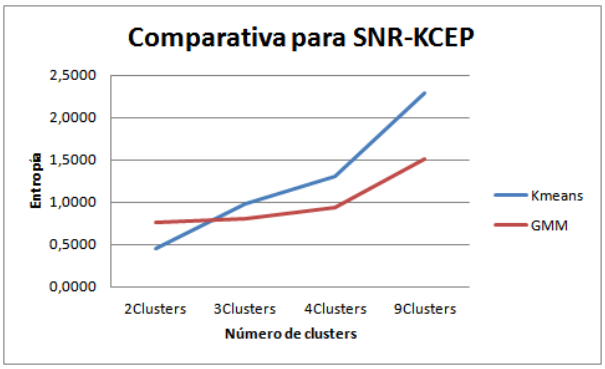


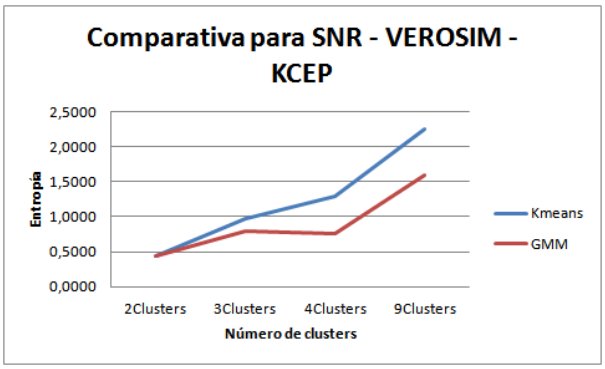
Cuadro 4.50: Gráficas por tipo de fichero para cuatro agrupamientos por GMM

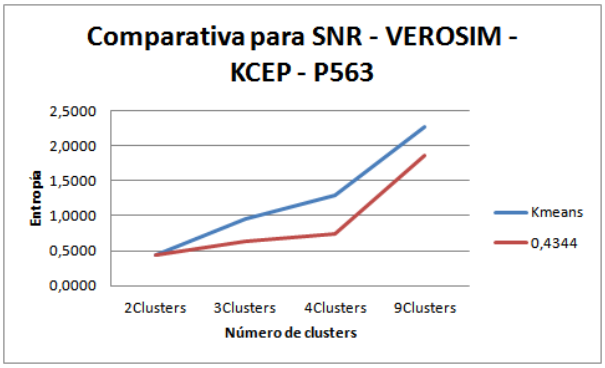
	2 Clusters		3 Clusters		4 Clusters		9 Clusters	
	K-means	GMM	K-means	GMM	K-means	GMM	K-means	GMM
snrkcep	0,4462	0,7634	0,9761	0,8041	1,305	0,9336	2,2917	1,5076
snrverosimkcep	0,4363	0,4455	0,9689	0,7989	1,297	0,7666	2,2536	1,5944
snrverosimkcepp563	0,4410	0,4457	0,9584	0,6414	1,294	0,7381	2,2733	1,8675
snrverosimklpckcepp563	0,4363	0,6659	0,9607	0,8831	1,295	1,1095	2,2665	1,2425

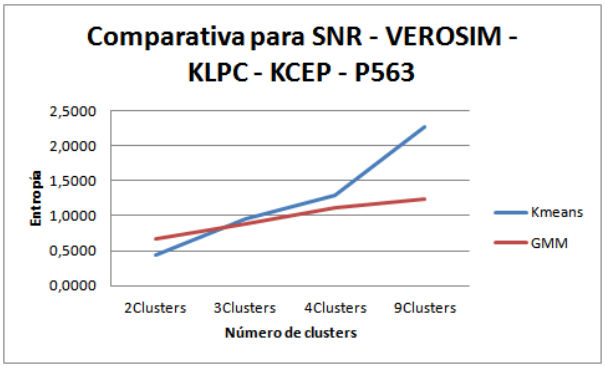
Cuadro 4.51: Entropía por tipo de fichero para K-means GMM variando el número de clusters

Como se ha podido observar durante esta sección la entropía por tipo de fichero aumenta a medida que se crean más agrupaciones. Esto se debe a que a medida que intentas hacer más agrupaciones los ficheros de audio se van clasificando en los diferentes grupos generados de tal forma que algunos ficheros de audio son separados de los de su mismo grupo. Para este tipo de análisis el algoritmo de clasificación que mejor ha funcionado ha sido el K-means, ya que en algunos casos (número de clusters bajo) se ha conseguido agrupar todos los ficheros de un mismo tipo en un mismo cluster. El comportamiento del algoritmo GMM es peor, los valores de entropía por tipo de fichero son más altos y no consigue en un experimento completo (variando los indicadores de degradación mientras el número de agrupamientos permanece constante) agrupar bien un tipo de ficheros.









Gráfica inicial

Gráfica 2 clusters por Kmeans

Cuadro 4.52: Gráficas de la comparativa entre K-means y GMM de la entropía por tipo de fichero

4.4. Selección del número óptimo de clusters

En esta sección se presenta una forma de obtener el número óptimo de agrupaciones (el número que ofrece menor entropía tanto el agrupamiento como el tipo de fichero) que se puede realizar sobre la base de datos NIST 2008. Como se ha visto en secciones anteriores, la entropía por cluster disminuye a medida que se crean más agrupaciones mientras que la entropía por tipo de fichero aumenta. En el punto de cruce de ambas entropías es donde se encontrará el punto óptimo de agrupamiento.

Para realizar este experimento se han seguido dos aproximaciones distintas, una con una sola iteración de experimentos y otra por medio de la media de varias simulaciones. Con estas distintas aproximaciones se pretende comprobar si los resultados son válidos.

4.4.1. Número óptimo de cluster con una única simulación

Para este experimento se realiza una ronda de experimentos tal y cómo se ha realizado en las secciones anteriores. El algoritmo de agrupación escogido ha sido el Gaussian Mixture Models (GMM). Se obtienen los valores del tipo de entropía del agrupamiento y, por otro lado, la entropía del tipo de fichero para agrupamientos en los que varía el número de grupos desde dos hasta doce.

En la tabla 4.56 se muestran los mejores resultados para distintas combinaciones de indicadores de degradación. Para dos combinaciones se ha escogido snr - kcep y el punto de cruce se obtiene para 8 clusters. Para tres indicadores se ha escogido snr - klpc - kcep y el número óptimo de agrupaciones son 6. También 6 grupos es el número óptimo para cuatro indicadores de

degradación (snr - verosim - klpc - kcep). Por último, en el caso de emplear todos los indicadores de degradación, es cuando se obtiene un menor número de agrupaciones, cinco.

En la tabla 4.56 se observan cuatro gráficas correspondientes cada una de ellas a las distintas combinaciones de indicadores de degradación. La línea azul (con tendencia decreciente) representa la entropía por agrupamiento, mientras que la línea roja representa la entropía por tipo de fichero (tendencia creciente).

Entropías sin realizar iteraciones			
	Cluster	Fichero	Num. Cluster óptimo
snr - kcep	1,5951	1,6628	8
snr - klpc - kcep	1,4467	1,5622	6
snr - verosim - klpc - kcep	1,3903	1,0994	6
snr - verosim - klcp - kcep - p563	1,4345	1,4137	5

Cuadro 4.53: Valores del entropía del agrupamiento y de tipo de fichero y número óptimo de agrupaciones con una iteración

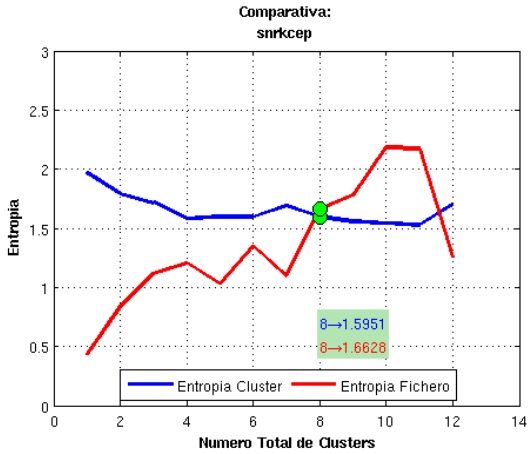
4.4.2. Número óptimo de cluster con diez simulaciones

Según el resultado obtenido en la subsección anterior, 4.4.1, el valor mínimo del cluster puede variar en función de la ejecución. Esto se debe a que las ejecuciones no siempre son iguales ya que la iniciación del vector gmm puede variar de una simulación a otra. Por ese motivo se realiza una simulación más compleja, en la cual, se ejecuta un total de diez iteraciones del mismo algoritmo con la finalidad de ver la variación de la entropía del agrupamiento y la entropía del tipo de fichero.

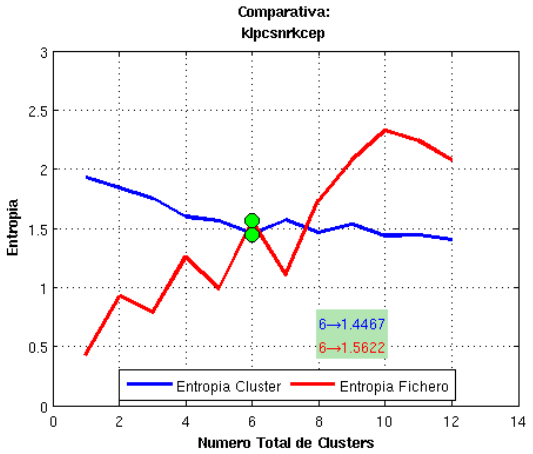
En la tabla 4.58 se muestran los valores de las entropías del cluster medias obtenidas después de realizar diez iteraciones. Para la combinación de dos indicadores de degradación (verosim - kcep) se obtiene que realizando siete agrupaciones los valores de la entropía, tanto del agrupamiento como del tipo de fichero. Para los experimentos con tres indicadores de degradación se obtiene el mejor resultado con verosim - klpc y kcep y cinco agrupamientos. En el caso de cuatro indicadores de degradación (snr - verosim - klpc - kcep) y de cinco indicadores de degradación el número óptimo de clusters a realizar es de seis agrupaciones.

4.4.3. Conclusiones

Como se ha podido observar en las tablas y gráficas anteriormente presentadas en esta sección. Los valores de número óptimo de clusters son coherentes con el número máximo de tipos de ficheros de audio presentes en el sistema (8 microfónicos y un telefónico) ya que por encima de 9 clusters se puede observar en las tablas 4.57 y 4.59 los valores obtenidos empeoran, sobre todo la entropía por tipo de fichero que muestra comportamientos no esperados.

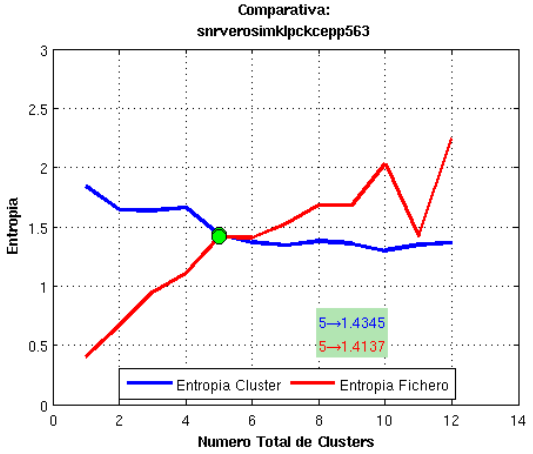


Comparativa snr - kcep



Comparativa klpc - snr - kcep

Comparativa snr - verosim - klpc - kcep

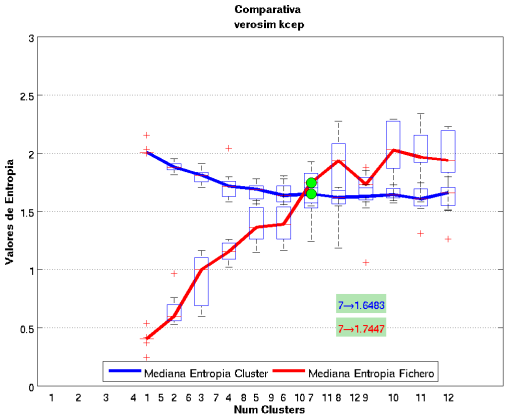


Comparativa snr - verosim - klpc - kcep - p563

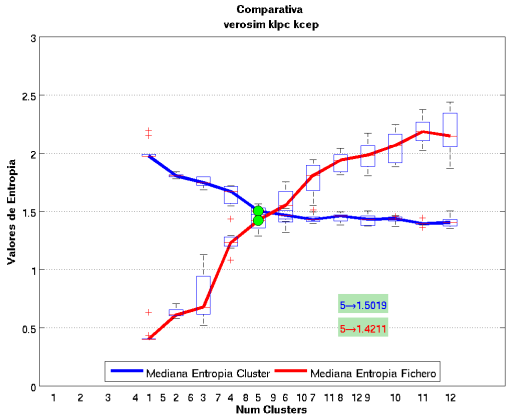
Cuadro 4.54: Gráficas comparativas para una iteración en la selección del número óptimo de clusters.

Entropías (medias) con iteraciones			
	Cluster	Fichero	Num. Cluster óptimo
Verosim - kcep	1,6483	1,7447	7
Verosim - klpc - kcep	1,5019	1,4211	5
Snr - verosim - klpc - kcep	1,4788	1,5021	6
Snr - verosim - klcp - kcep - p563	1,4828	1,4947	6

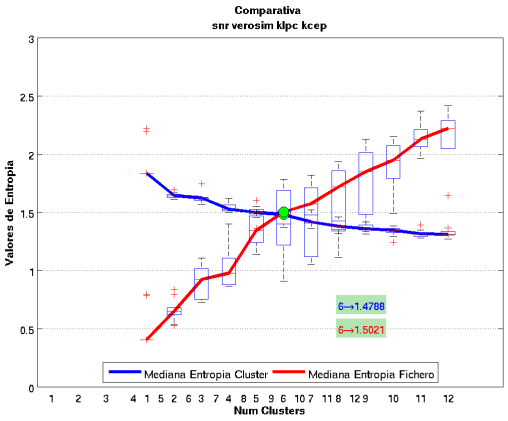
Cuadro 4.55: Valores del entropía del agrupamiento y de tipo de fichero y cluster óptimo con diez iteración



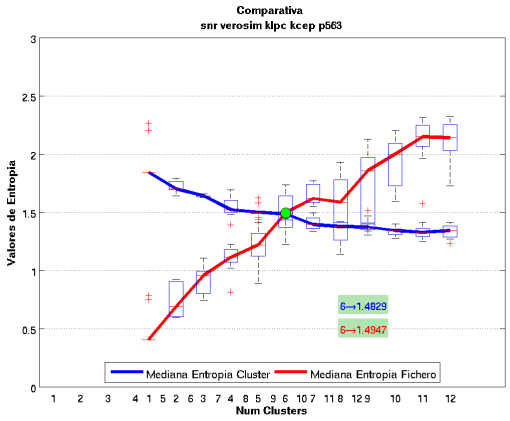
Comparativa verosim - kcep



Comparativa verosim - klpc - kcep



Comparativa snr - verosim - klpc - kcep



Comparativa snr - verosim - klpc - kcep - p563

Cuadro 4.56: Gráficas comparativas para diez iteraciones en la selección del número óptimo de clusters.

4.5. Medida del rendimiento por medio de curvas DET

En esta sección, 4.5, se presentan las gráficas de curvas DET correspondientes a los escenarios expuestos en los apartados anteriores. Se muestran en diferentes tablas por un lado las curvas DET obtenidas por medio de una agrupación por medio del algoritmo k-means y, en la siguiente gráfica, las obtenidas por medio de GMM.

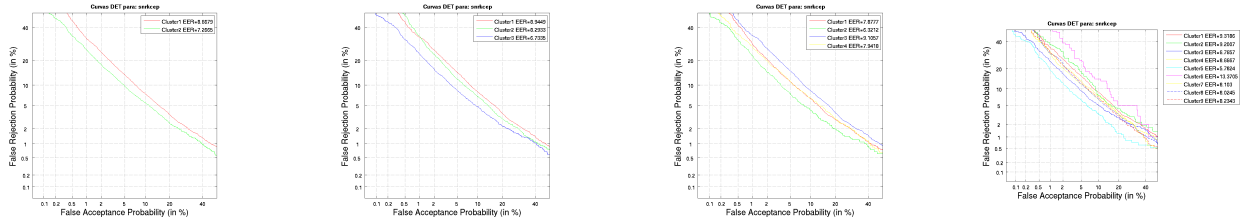
4.5.1. Análisis sobre K-means

Durante esta subsección se mostrarán las curvas DET obtenidas para los agrupamientos realizados con el algoritmo K-means y empleando la medida de distancia cityblock. Los valores para los que se muestran las curvas DET son los que mejores resultados (menor valor de entropía) han ofrecido a lo largo de este proyecto final de carrera. Estos valores son los siguientes:

- snr - kcep
- snr - verosim - kcep
- snr - verosim - kcep - P.563
- snr - verosim - klpc - kcep - P.563

Curvas DET para SNR-KCEP

2 Clusters | 3 Clusters | 4 Clusters | 9 Clusters



Cuadro 4.57: K-means: Curvas DET para snr-kcep variando el número de agrupaciones generadas

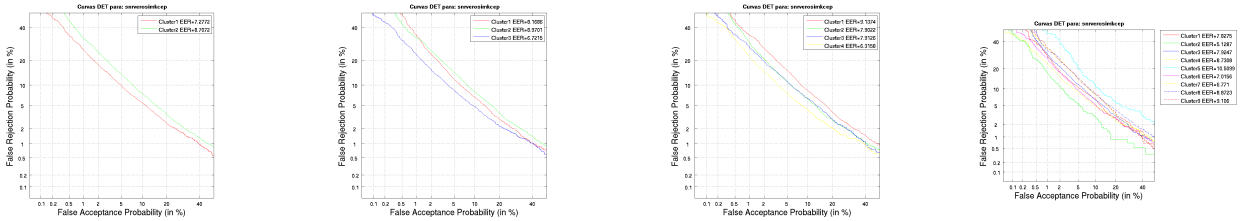
EER para snr-kcep				
ERR	2Clusters	3Clusters	4Clusters	9Clusters
C1	8,6679	8,9449	7,8777	9,3186
C2	7,2265	8,2933	6,3212	9,2007
C3		6,7335	9,1057	6,7656
C4			7,9418	8,667
C5				5,7824
C6				13,3705
C7				8,103
C8				8,0245
C9				8,2343

Cuadro 4.58: K-means: Valores EER para snr-kcep variando el número de agrupamientos

Como se ha observado en las tablas desde las 4.57 a la 4.64 las curvas DET son independientes a la agrupación, es decir, a medida que aumentan los grupos en el sistema no se nota un cambio

Curvas DET para SNR-VEROSIM-KCEP

2 Clusters 3 Clusters 4 Clusters 9 Clusters



Cuadro 4.59: K-means: Curvas DET para snr - verosim - kcep variando el número de agrupaciones generadas

EER para snr - verosim - kcep				
ERR	2Clusters	3Clusters	4Clusters	9Clusters
C1	7,2772	8,1686	9,1374	7,8275
C2	8,7072	8,9701	7,9022	5,1287
C3		6,7215	7,9126	7,9247
C4			6,3158	8,7308
C5				10,5099
C6				7,0156
C7				6,771
C8				8,8723
C9				9,106

Cuadro 4.60: K-means: Valores EER para snr - verosim -kcep variando el número de agrupamientos

brusco del valor del EER. Sin embargo, si que es destacable que en el caso de mostrar las curvas DET obtenidas de los experimentos con nueve agrupaciones, existen curvas escalonadas lo que indica que no existen suficientes valores en ese cluster como para realizar un análisis correcto. Esto indica que nueve agrupaciones son demasiadas para el caso K-means ya que existe un descenso del rendimiento del sistema.

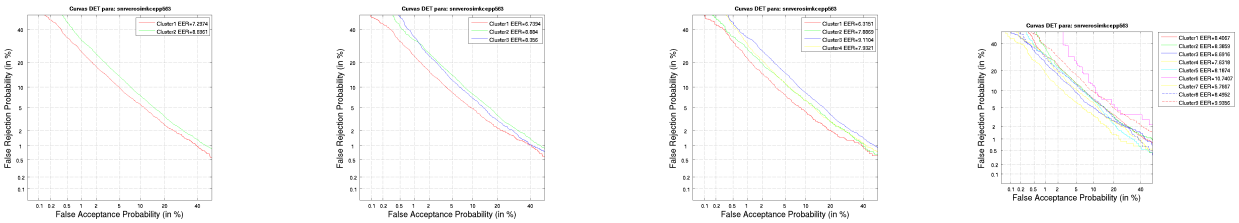
4.5.2. Análisis sobre GMM

A continuación se muestra el mismo conjunto de figuras que en la sección anterior pero para el algoritmo de clasificación GMM. De nuevo, las combinaciones sobre los que se muestran las curvas DET son los siguientes:

- snr - kcep
- snr - verosim - kcep
- snr - verosim - kcep - p563
- snr - verosim - klpc - kcep - p563

Curvas DET para SNR-VEROSIM-KCEP-P563

2 Clusters | 3 Clusters | 4 Clusters | 9 Clusters



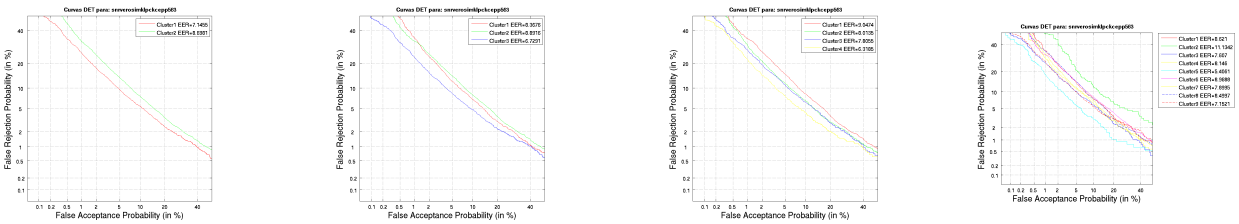
Cuadro 4.61: K-means: Curvas DET para snr- verosim - kcep - P.563 variando el número de agrupaciones generadas

EER para snr - verosim - kcep - P.563				
ERR	2Clusters	3Clusters	4Clusters	9Clusters
C1	7,2974	6,7394	6,3151	8,4067
C2	8,6961	8,884	7,8869	8,3859
C3		8,356	9,1104	6,6916
C4			7,9321	7,6318
C5				8,1874
C6				10,7407
C7				5,7667
C8				8,4952
C9				9,9356

Cuadro 4.62: K-means: Valores EER para snr - verosim -kcep - P.563 variando el número de agrupamientos

Curvas DET para SNR-VEROSIM-KLPC-KCEP-P563

2 Clusters | 3 Clusters | 4 Clusters | 9 Clusters



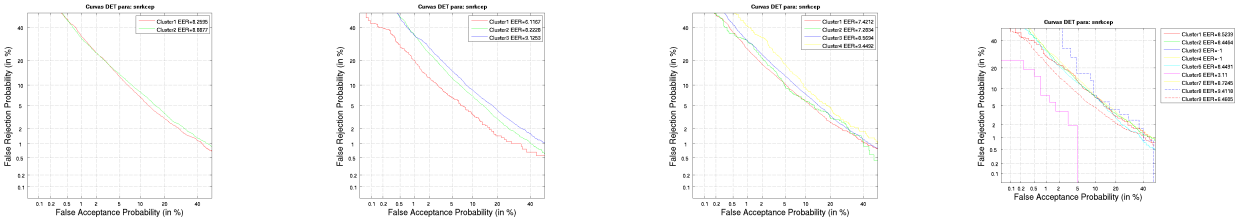
Cuadro 4.63: K-means: Curvas DET para snr - verosim - klpc - kcep - P.563 variando el número de agrupaciones generadas

EER para snr - verosim - klpc - kcep - P.563				
ERR	2Clusters	3Clusters	4Clusters	9Clusters
C1	7,14,55	8,3676	9,0474	8,621
C2	8,6981	8,8916	8,0135	11,1342
C3		6,7291	7,8055	7,607
C4			6,3185	8,146
C5				5,4061
C6				8,9688
C7				7,8995
C8				8,4997
C9				7,1521

Cuadro 4.64: K-means: Valores EER para snr - verosim -kcep - P.563 variando el número de agrupamientos

Curvas DET para SNR-KCEP

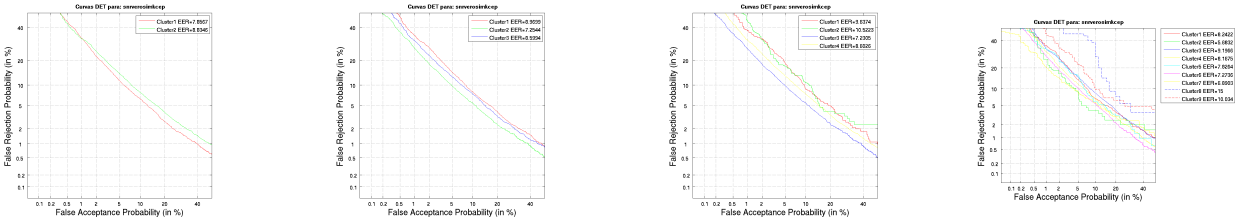
2 Clusters 3 Clusters 4 Clusters 9 Clusters



Cuadro 4.65: GMM: Curvas DET para snr - kcep variando el número de agrupaciones generadas

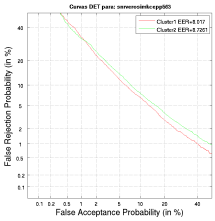
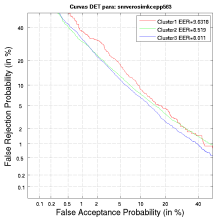
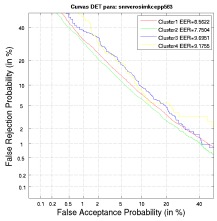
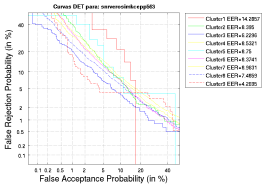
Curvas DET para SNR-VEROSIM-KCEP

2 Clusters 3 Clusters 4 Clusters 9 Clusters



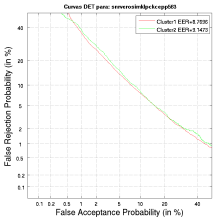
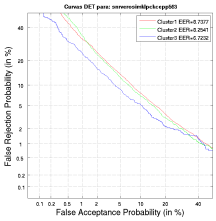
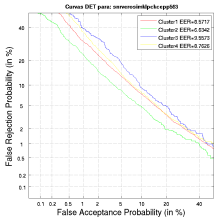
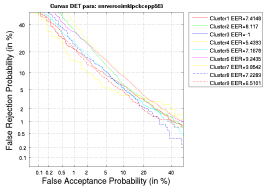
Cuadro 4.66: GMM: Curvas DET para snr - verosim - kcep variando el número de agrupaciones generadas

Curvas DET para SNR-VEROSIM-KCEP-P563

2 Clusters	3 Clusters	4 Clusters	9 Clusters
			

Cuadro 4.67: GMM: Curvas DET para snr - verosim - kcep - p563 variando el número de agrupaciones generadas

Curvas DET para SNR-VEROSIM-KLPC-KCEP-P563

2 Clusters	3 Clusters	4 Clusters	9 Clusters
			

Cuadro 4.68: GMM: Curvas DET para snr - verosim - klpc - kcep - p563 variando el número de agrupaciones generadas

Para el caso de la agrupación realizada por medio de GMM se observa un rendimiento ligeramente peor que el obtenido en el agrupamiento K-means (los valores de EER son un poco más altos que los anteriormente obtenidos). Como sucedía en el caso K-means, al realizar agrupaciones de 9 elementos se observa la existencia de algunos clusters con pocos elementos.

Este resultado en las agrupaciones de nueve elementos es coherente con los valores obtenidos en las curvas de corte de la entropía del agrupamiento con la entropía del tipo de fichero donde el número óptimo de agrupamientos se sitúa por debajo de nueve, en concreto, para todas combinaciones, oscila entre seis y siete agrupamientos.

4.5.3. Conclusiones

El objetivo principal de realizar curvas DET es comprobar cómo de bueno ha sido el agrupamiento realizado. Cuanto más bajo el EER mejor rendimiento presentará el sistema.

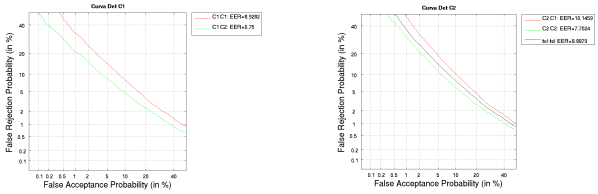
Las gráficas presentadas desde el cuadro 4.59 al 4.61 muestran la curva DET para los ficheros pertenecientes a un cluster determinado. A partir de la tabla 4.60, con cuatro agrupamientos realizados, se observa como algunos de los clusters (C3 en el caso de snr-kcep para K-means) no tienen suficientes valores como para generar una curva DET. Este hecho se muestra en los extremos de la gráfica, donde aparecen .^{es}calonesïndicando la falta de datos. Este hecho es más acusado a medida que hacemos más clusters, de forma que en las curvas pertenecientes a 9 agrupamientos existen clusters con pocos elementos. A medida que aumentamos el número de indicadores de degradación involucrados esta falta de datos se hace más acusada. Sobre todo en el caso de realizar el agrupamiento empleando el algoritmo GMM.

Las diferencias presentes en el EER se deben a las diferencias de entropía de cada cluster. De esta forma, los clusters con menor entropía presentan un valor de EER más cercano a 0, lo que relaciona el agrupamiento con el rendimiento del sistema. Cuanto mejor es un agrupamiento, la entropía parcial del cluster tiende a 0 a la vez que el valor de EER disminuye significativamente.

4.6. Comparativa de agrupamiento por medio de curvas DET en enfrentamientos tel-tel

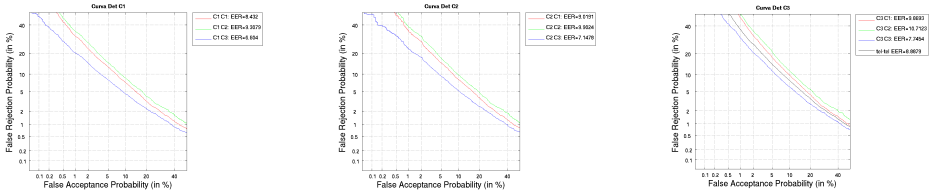
En esta última sección del capítulo de resultados se muestran diferentes curvas y gráficas que permiten realizar una comparativa entre las curvas DET de los scores originales obtenidos por enfrentamientos tel-tel con las curvas DET generadas a partir de los agrupamientos. La finalidad de este último experimento es comprobar cómo de bueno ha sido el agrupamiento, es decir, si

Comparativa Curvas DET dos agrupaciones



Cuadro 4.69: Comparativa Curvas DET de agrupamiento y Curva DET enfrentamiento tel-tel para dos agrupaciones

Comparativa Curvas DET tres agrupaciones



Cuadro 4.70: Comparativa Curvas DET de agrupamiento y Curva DET enfrentamiento tel-tel para tres agrupaciones

el valor de la EER de los scores originales tipo tel-tel es similar a los scores obtenidos de la agrupación con mayor número de ficheros telefónicos.

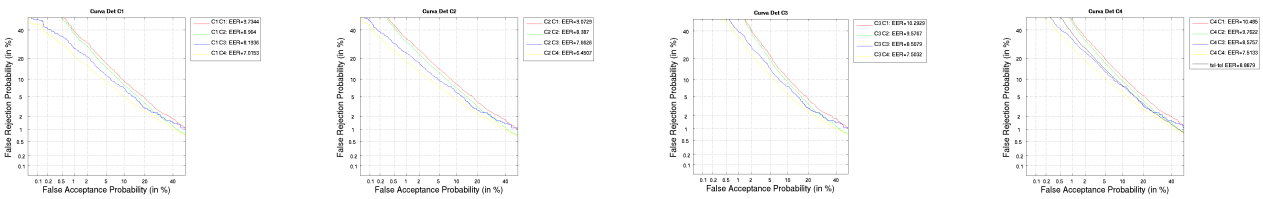
Para los scores generados por los agrupamientos se escoge el que presenta más ficheros de tipo tel-tel y, a su vez, es el que menor entropía parcial tiene.

Para todas las gráficas se ha tenido en cuenta el mismo algoritmo de agrupamiento GMM y la misma combinación de indicadores de degradación, cinco indicadores correspondientes con snr, verosim, klpc, kcep y P.563.

Las tablas siguientes, 4.69, 4.70 y 4.71, muestran estas gráficas.

En la tabla 4.72 se realiza el estudio para dos agrupamientos. El cluster con menor entropía parcial es cluster 2, que a su vez es el que tiene mayor número de elementos de tipo telefónicos. En este caso se comprueba que el valor del EER está por debajo del valor obtenido directamente de los scores tel-tel.

Comparativa Curvas DET tres agrupaciones



Cuadro 4.71: Comparativa Curvas DET de agrupamiento y Curva DET enfrentamiento tel-tel para tres agrupaciones

Entropía de dos agrupaciones		
Cluster 1	Cluster 2	Entropía Total
3.0672	1.0358	1.8407

Cuadro 4.72: Entropía para dos agrupaciones

Entropía de tres agrupaciones			
Cluster 1	Cluster 2	Cluster 3	Entropía Total
3.0760	1.8768	0.8512	1.6985

Cuadro 4.73: Entropía para tres agrupaciones

En la tabla 4.73 el cluster con menor entropía parcial es el tercero. De esta forma en las gráficas presentadas en la tabla 4.70 todas las curvas DET que pertenecen al C3 tienen un valor EER menor que el obtenido por los scores tel-tel.

Entropía de cuatro agrupaciones				
Cluster 1	Cluster 2	Cluster 3	Cluster 4	Entropía Total
2.0841	3.0830	0.5525	0.4927	1.6433

Cuadro 4.74: Entropía para cuatro agrupaciones

En la tabla 4.74 el cluster con menor entropía parcial es el cuarto. igual que sucede en el caso anterior, en las gráficas presentadas en la tabla 4.71 las curvas EER del C4 obtienen menor valor de EER que la obtenida por los scores tel-tel.

4.6.1. Conclusiones

Como se ha visto en las tablas y gráficas anteriores el agrupamiento realizado es bastante bueno. Los valores de entropía obtenidos son cercanos a 0 y no superan el valor teórico máximo ($\log_2 9$). A su vez en comparación con el valor EER obtenido de los scores tel-tel, las agrupaciones son buenas, ya que el cluster con la mayor parte de los ficheros de tipo telefónico (que además tiene menor entropía parcial) tiene un EER más parecido al original. Esto significa que la agrupación es buena, sobre todo en los ficheros de tipo telefónico.

5

Conclusiones y trabajo futuro

5.1. Conclusiones

En este trabajo final de carrera se han estudiado distintos métodos de agrupación de ficheros de audio en función de su calidad en sistemas de reconocimiento de locutor.

Para que el estudio realizado tuviera una base experimental sólida y ser capaces de extraer la mayor cantidad de información posible se han utilizado dos bases de datos bien diferencias. Por una parte, la base de datos procedente de la evaluación bianual NIST 2008 y por otra parte, una base de datos forenses formada por grabaciones de audio reales. El estudio ha sido realizado sobre los dos tipos de canal más habitual, canal telefónico y canal microfónico. El empleo de dos bases de datos muy diferentes entre sí, tanto por número total de ficheros como por calidad de éstos, ofrece dos sistemas de estudio bien diferenciados.

La base de datos de la Guardia Civil, permitió hacer un estudio del problema desde cero, implementando métodos para obtener las medidas de calidad de cada fichero de audio y, posteriormente, adaptando estos valores experimentales de tal forma que fueran compatibles con los presentes en los estudios de la base de datos NIST 2008.

Los estudios realizados sobre la base de datos de la Guardia Civil, AhumadaIV-BaezaI, han servido para demostrar qué es posible realizar una clasificación de ficheros de audio ateniéndose a criterios de calidad de la muestra obtenida. En esta base de datos es muy fácil observar como una medida de carácter subjetivo (P.563) no se muestra muy determinante a la hora de realizar agrupamientos, sin embargo indicadores de degradación como snr y verosim aportan, para todos los experimentos realizados y para todas las combinaciones, los mejores valores de entropía del agrupamiento.

Al contar con la base de datos de la Guardia Civil ha quedado demostrado que la clasificación de ficheros de audio puede ser empleada en múltiples aplicaciones en la vida real. En este caso, el trabajo presentado a lo largo de este proyecto final de carrera sirvió para obtener la acreditación de los laboratorios de acústica de la Guardia Civil.

En la base de datos NIST 2008 se llevaron a cabo más experimentos debido a que existen más ficheros de audio sobre los que realizar pruebas. Se calcularon tanto entropías por agrupamiento como entropías por tipo de fichero, lo que ha permitido obtener un número de óptimo de clusters tanto para agrupamientos con algoritmo K-means como empleando algoritmo GMM. Sobre la

base de datos NIST 2008 no existe una gran diferencia entre emplear K-means o GMM, aunque en los resultados presentes en este proyecto final de carrera, los resultados con mejor agrupamiento se obtienen con K-means, debido probablemente a los problemas de inicialización de las matrices de covarianzas del algoritmo GMM.

Los experimentos de variación del número de grupos generados tanto para el algoritmo K-means como para GMM indican que existe una serie de combinaciones de indicadores de degradación que presentan buen rendimiento independientemente del número del clusters. De esta forma, combinaciones como snr - kcep o snr - verosim - kcep, son las que obtienen siempre menor nivel de entropía tanto para la entropía del agrupamiento como para la entropía del tipo de fichero en el algoritmo K-means. Este es un dato muy importante ya que demuestra que estos dos indicadores de degradación son los aportan más información a al hora de realizar el agrupamiento. Sin embargo, medidas subjetivas como P.563 no parecen tener tanto peso a la hora de realizar el agrupamiento.

El análisis del número óptimo de clusters sobre la base de datos NIST 2008 indica que las agrupaciones con mejor rendimiento del sistema son aquellas que están formadas por un número de cluster entre cinco y siete, tanto para K-means como para GMM. Este número concuerda con el número de tipos de ficheros en el sistema, nueve en total (ocho de ellos de tipo microfónico, y uno de ellos de tipo telefónico). De estos tipos de ficheros, los tipos microfónicos son muy parecidos entre sí por lo que no es descartable que dos o tres de ellos sean tan similares que el algoritmo de clasificación no sea capaz de segmentarlos más.

Este número óptimo está relacionado con las curvas DET, ya que en la mayoría de las simulaciones para un experimento con nueve clusters existen agrupaciones con un número muy pequeño de ficheros, lo que conlleva una pérdida de información en la curva DET. Sin embargo, con agrupaciones de cuatro elementos todos los clusters tienen un número suficiente de elementos como para permitir que se obtenga un buen rendimiento del sistema.

A la hora de realizar una comparación entre los valores de EER de los scores obtenidos por enfrentamientos tel-tel y los valores obtenidos por la agrupación realizada durante este proyecto final de carrera, los resultados obtenidos son bastante buenos. El agrupamiento que presenta un mayor número de scores de tipo telefónico-telefónico tiene un EER más parecido al que se obtiene si se calcula directamente a partir de los scores tel-tel. Esto se debe, a que el cluster que presenta menor entropía parcial es, a su vez, el que presenta mayor número de scores tel-tel (existen muchos más ficheros de audio tipo telefónico que micrófonico), por lo tanto, el sistema propuesto en este proyecto final de carrera en general, clasifica muy bien los ficheros de audio de tipo telefónico.

A continuación, se citan las aportaciones que se han realizado en relación a los puntos planteados en los objetivos del proyecto:

1. Estado del arte: Se han documentado las últimas tecnologías utilizadas en sistemas de reconocimiento de locutor, algunos de ellos empleados en este proyecto final de carrera. Además se realiza un estudio de los indicadores de degradación empleados para evaluar el rendimiento del sistema. Posteriormente, estos indicadores de degradación son usados en los distintos experimentos realizados. Además se ha realizado un estudio de los algoritmos de agrupamiento K-means y GMM de cara a la utilización de ambos para realizar las agrupaciones de los ficheros de audio presentes en las dos bases de datos empleadas.

2. Implementación de las agrupaciones: se han realizado agrupamientos empleando los dos algoritmos descritos en el estado del arte sobre las dos bases de datos usadas en este proyecto por medio de los indicadores de degradación propuestos por [20]. Se ha comprobado que la base de datos de origen forense, AhumadaIV-BaezaI, ofrece peores resultados debido principalmente a los escasos elementos que lo forman así como la variabilidad presente

en la adquisición de la voz (ruidos ambientales, escasos silencios, habla simultanea de dos personas, etc.).

3. Implementación de un sistema de análisis por medio de curvas DET. Se analizan las agrupaciones obtenidas en la etapa anterior del proyecto siguiendo los criterios propuestos por [9].

4. Evaluación de la entropía de los agrupamientos: se realizan distintas medidas de entropía, por una parte la entropía obtenida directamente de los agrupamientos generados y por otra parte la entropía obtenida del tipo de fichero. Sobre estas dos entropías se han generado una serie de curvas que permiten obtener el número óptimo de clusters en el sistema (punto de cruce de ambas entropías). Relacionado con la entropía se ha llevado a cabo una análisis de cómo de bueno ha sido el agrupamiento, relacionando la entropía parcial de cada agrupamiento con los valores de los scores obtenidos de enfrentamientos tipo tel - tel.

5.2. Trabajo futuro

La información obtenida en este proyecto final de carrera, tanto desde el punto de vista del análisis del estado del arte, como la parte experimental, podrían servir para implementar métodos de compensación de variabilidad intersesión utilizando medidas de calidad en función de los agrupamientos generados.

Otra posible línea de desarrollo futuro podría ser la implementación de un discriminador de ficheros de la base de datos empleando los valores de calidad. De esta forma se podría excluir de los experimentos diferentes tipos de micrófonos que no vayan a realizar un buen aporte de cara al rendimiento del sistema. Un posible método de realizar esto sería obtener la mejor combinación de medidas de calidad desde el punto de vista de la entropía y luego descartar los ficheros que no pertenezcan al cluster adecuado.

Un estudio de otra serie de indicadores de degradación también sería útil, ya que aportaría más combinaciones sobre las que poder realizar agrupamientos y no lo limitaría tan solo a cinco combinaciones. Como ha quedado demostrado a lo largo de este PFC se obtienen mejores resultados tanto de entropía de agrupamiento como de entropía de fichero con la inclusión de más indicadores de degradación en el experimento, por lo que realizar un estudio exhaustivo de nuevos indicadores conseguiría reducir más los valores de entropía de fichero y por tanto de las curvas DET de rendimiento del sistema.

Para todo ello es necesario tener una base de datos sólida. En este proyecto ha quedado demostrado que existe una gran diferencia entre las dos bases de datos presentes, sobre todo debido a que existe un número muy diferente de ficheros de audio en ambas. Mientras que la base de datos NIST 2008 cuenta con 6236 modelos y 6079 test, AhumadaIV-BaezaI está formada por 715 ficheros (182 modelos y 533 tests). Por lo que sería importante concentrar esfuerzos futuros en crear una base de datos forense de referencia, en la que se incluyan más ficheros de audio y, además, se mejore la calidad de éstos.

Glosario de acrónimos

- **DET**: Detection Error Tradeoff
- **EER**: Equal Error Rate
- **GMM**: Gaussian Mixture Model
- **ID**: Indicador de degradación
- **KCEP**: Kurtosis sobre coeficientes cepstrales
- **KLPC**: Kurtosis sobre coeficientes LPC
- **LLR**: Log-Likelihood Rate
- **LPC**: Linear Predictive Coding
- **MFCC**: Mel-Frequencyă Cepstrală Coefficients
- **SNR**: Relación señal a ruido
- **SRL**: Sistema de reconocimiento de locutor
- **UBM**: Universal Background Model
- **UBML**: Universal Background Model Likelihood

Bibliografía

[1] J.P. Campbell Jr. Speaker recognition: A tutorial. *Proceedings of the IEEE*, 85(9):1437–1462, 1997.

[2] R.J. Vogt, C.J. Lustri, and S. Sridharan. Factor analysis modelling for speaker verification with short utterances. *Proc. of Odyssey, 2008*, 2008.

[3] NIST. The nist year 2008 speaker recognition evaluation plan 1. *2008 NIST Speaker Recognition Evaluation*, 2008:1–10, 2008.

[4] A.K. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *Circuits and Systems for Video Technology, IEEE Transactions on*, 14(1):4–20, 2004.

[5] T. Kinnunen and H. Li. An overview of text-independent speaker recognition: From features to supervectors. *Speech Communication*, 52(1):12–40, 2010.

[6] D.A. Reynolds, T.F. Quatieri, and R.B. Dunn. Speaker verification using adapted gaussian mixture models. *Digital signal processing*, 10(1-3):19–41, 2000.

[7] P. Kenny, P. Ouellet, N. Dehak, V. Gupta, and P. Dumouchel. A study of interspeaker variability in speaker verification. *Audio, Speech, and Language Processing, IEEE Transactions on*, 16(5):980–988, 2008.

[8] G. Doddington, W. Liggett, A. Martin, M. Przybocki, and D.A. Reynolds. Sheep, goats, lambs and wolves: A statistical analysis of speaker performance in the nist 1998 speaker recognition evaluation. In *Fifth International Conference on Spoken Language Processing*, 1998.

[9] A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki. The det curve in assessment of detection task performance. Technical report, DTIC Document, 1997.

[10] F. Bimbot, J.F. Bonastre, C. Fredouille, G. Gravier, I. Magrin-Chagnolleau, S. Meignier, T. Merlin, J. Ortega-García, D. Petrovska-Delacrétaz, and D.A. Reynolds. A tutorial on text-independent speaker verification. *EURASIP journal on applied signal processing*, 2004:430–451, 2004.

[11] A. Hicklin and R. Khanna. The role of data quality in biometric systems. *White Paper. Mitretek Systems*, 2006.

[12] F. Alonso-Fernandez, F. Roli, G.L. Marcialis, J. Fierrez, J. Ortega-Garcia, and J. Gonzalez-Rodriguez. Performance of fingerprint quality measures depending on sensor technology. *Journal of Electronic Imaging*, 17:011008, 2008.

[13] S. Furui. Cepstral analysis technique for automatic speaker verification. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 29(2):254–272, 1981.

[14] H. Hermansky and N. Morgan. Rasta processing of speech. *Speech and Audio Processing, IEEE Transactions on*, 2(4):578–589, 1994.

[15] J. Pelecanos and S. Sridharan. Feature warping for robust speaker verification. *Proc. of Odyssey, 2001*, 2001.

[16] D.A. Reynolds. Channel robust speaker verification via feature mapping. *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*, 2:II–53, 2003.

[17] P. Kenny and P. Dumouchel. Disentangling speaker and channel effects in speaker verification. *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*, 1:I–37, 2004.

[18] R. Vogt and S. Sridharan. Explicit modelling of session variability for speaker verification. *Computer Speech & Language*, 22(1):17–38, 2008.

[19] D. Garcia-Romero, J. Fierrez-Aguilar, J. Gonzalez-Rodriguez, and J. Ortega-Garcia. Using quality measures for multilevel speaker recognition. *Computer Speech & Language*, 20(2):192–209, 2006.

[20] Alberto Harriero. Fiabilidad en sistemas forenses de reconocimiento de locutor explotando la calidad de la señal de voz. *Proyecto Fin de Carrera*, 2010.

[21] V Grancharov and W.B. Hleijn. Speech quality assessment. *Springer Handbook of Speech-Processing*, 2007.

[22] R. Agrawal, J. Gehrke, D. Gunopulos, and P. Raghavan. Automatic subspace clustering of high dimensional data. *Data Mining and Knowledge Discovery*, 11(1):5–33, 2005.

[23] D. Burago, Y. Burago, S. Ivanov, and American Mathematical Society. *A course in metric geometry*. American Mathematical Society Providence, 2001.

[24] D. MacKay. An example inference task: Clustering. *Information Theory, Inference and Learning Algorithms*, pages 284–292, 2003.

[25] D.A. Van Leeuwen. Speaker linking in large data sets. *Proceedings of Odyssey*, 2010.

Presupuesto

1)	**Ejecución Material**	
	▪ Compra de ordenador personal (Software incluido)	1.5000 €
	▪ Material de oficina	200 €
	▪ Total de ejecución material	1.700 €
2)	**Gastos generales**	
	▪ sobre Ejecución Material	240 €
3)	**Beneficio Industrial**	
	▪ sobre Ejecución Material	90 €
4)	**Honorarios Proyecto**	
	▪ 1800 horas a 15 €/ hora	27000 €
5)	**Material fungible**	
	▪ Gastos de impresión	100 €
	▪ Encuadernación	200 €
6)	**Subtotal del presupuesto**	
	▪ Subtotal Presupuesto	29.330 €
7)	**I.V.A. aplicable**	
	▪ 18 % Subtotal Presupuesto	5.279,4 €
8)	**Total presupuesto**	
	▪ Total Presupuesto	34.609,4 €

Madrid, Septiembre 2012

El Ingeniero Jefe de Proyecto

Fdo.: Ana García Muro

Ingeniero Superior de Telecomunicación

Pliego de condiciones

Pliego de condiciones

Este documento contiene las condiciones legales que guiarán la realización, en este proyecto, de un *Utilización de medidas de calidad de la señal de voz para compensación de variabilidad inter sesión en reconocimiento de locutor*. En lo que sigue, se supondrá que el proyecto ha sido encargado por una empresa cliente a una empresa consultora con la finalidad de realizar dicho sistema. Dicha empresa ha debido desarrollar una línea de investigación con objeto de elaborar el proyecto. Esta línea de investigación, junto con el posterior desarrollo de los programas está amparada por las condiciones particulares del siguiente pliego.

Supuesto que la utilización industrial de los métodos recogidos en el presente proyecto ha sido decidida por parte de la empresa cliente o de otras, la obra a realizar se regulará por las siguientes:

Condiciones generales.

1. La modalidad de contratación será el concurso. La adjudicación se hará, por tanto, a la proposición más favorable sin atender exclusivamente al valor económico, dependiendo de las mayores garantías ofrecidas. La empresa que somete el proyecto a concurso se reserva el derecho a declararlo desierto.

2. El montaje y mecanización completa de los equipos que intervengan será realizado totalmente por la empresa licitadora.

3. En la oferta, se hará constar el precio total por el que se compromete a realizar la obra y el tanto por ciento de baja que supone este precio en relación con un importe límite si este se hubiera fijado.

4. La obra se realizará bajo la dirección técnica de un Ingeniero Superior de Telecomunicación, auxiliado por el número de Ingenieros Técnicos y Programadores que se estime preciso para el desarrollo de la misma.

5. Aparte del Ingeniero Director, el contratista tendrá derecho a contratar al resto del personal, pudiendo ceder esta prerrogativa a favor del Ingeniero Director, quien no estará obligado a aceptarla.

6. El contratista tiene derecho a sacar copias a su costa de los planos, pliego de condiciones y presupuestos. El Ingeniero autor del proyecto autorizará con su firma las copias solicitadas por el contratista después de confrontarlas.

7. Se abonará al contratista la obra que realmente ejecute con sujeción al proyecto que sirvió de base para la contratación, a las modificaciones autorizadas por la superioridad o a las órdenes que con arreglo a sus facultades le hayan comunicado por escrito al Ingeniero Director de obras siempre que dicha obra se haya ajustado a los preceptos de los pliegos de condiciones, con arreglo a los cuales, se harán las modificaciones y la valoración de las

diversas unidades sin que el importe total pueda exceder de los presupuestos aprobados. Por consiguiente, el número de unidades que se consignan en el proyecto o en el presupuesto, no podrá servirle de fundamento para entablar reclamaciones de ninguna clase, salvo en los casos de rescisión.

8. Tanto en las certificaciones de obras como en la liquidación final, se abonarán los trabajos realizados por el contratista a los precios de ejecución material que figuran en el presupuesto para cada unidad de la obra.

9. Si excepcionalmente se hubiera ejecutado algún trabajo que no se ajustase a las condiciones de la contrata pero que sin embargo es admisible a juicio del Ingeniero Director de obras, se dará conocimiento a la Dirección, proponiendo a la vez la rebaja de precios que el Ingeniero estime justa y si la Dirección resolviera aceptar la obra, quedará el contratista obligado a conformarse con la rebaja acordada.

10. Cuando se juzgue necesario emplear materiales o ejecutar obras que no figuren en el presupuesto de la contrata, se evaluará su importe a los precios asignados a otras obras o materiales análogos si los hubiere y cuando no, se discutirán entre el Ingeniero Director y el contratista, sometiéndolos a la aprobación de la Dirección. Los nuevos precios convenidos por uno u otro procedimiento, se sujetarán siempre al establecido en el punto anterior.

11. Cuando el contratista, con autorización del Ingeniero Director de obras, emplee materiales de calidad más elevada o de mayores dimensiones de lo estipulado en el proyecto, o sustituya una clase de fabricación por otra que tenga asignado mayor precio o ejecute con mayores dimensiones cualquier otra parte de las obras, o en general, introduzca en ellas cualquier modificación que sea beneficiosa a juicio del Ingeniero Director de obras, no tendrá derecho sin embargo, sino a lo que le correspondería si hubiera realizado la obra con estricta sujeción a lo proyectado y contratado.

12. Las cantidades calculadas para obras accesorias, aunque figuren por partida alzada en el presupuesto final (general), no serán abonadas sino a los precios de la contrata, según las condiciones de la misma y los proyectos particulares que para ellas se formen, o en su defecto, por lo que resulte de su medición final.

13. El contratista queda obligado a abonar al Ingeniero autor del proyecto y director de obras así como a los Ingenieros Técnicos, el importe de sus respectivos honorarios facultativos por formación del proyecto, dirección técnica y administración en su caso, con arreglo a las tarifas y honorarios vigentes.

14. Concluida la ejecución de la obra, será reconocida por el Ingeniero Director que a tal efecto designe la empresa.

15. La garantía definitiva será del 4

16. La forma de pago será por certificaciones mensuales de la obra ejecutada, de acuerdo con los precios del presupuesto, deducida la baja si la hubiera.

17. La fecha de comienzo de las obras será a partir de los 15 días naturales del replanteo oficial de las mismas y la definitiva, al año de haber ejecutado la provisional, procediéndose si no existe reclamación alguna, a la reclamación de la fianza.

18. Si el contratista al efectuar el replanteo, observase algún error en el proyecto, deberá comunicarlo en el plazo de quince días al Ingeniero Director de obras, pues transcurrido ese plazo será responsable de la exactitud del proyecto.

19. El contratista está obligado a designar una persona responsable que se entenderá con el Ingeniero Director de obras, o con el delegado que éste designe, para todo relacionado con ella. Al ser el Ingeniero Director de obras el que interpreta el proyecto, el contratista deberá consultarle cualquier duda que surja en su realización.

20. Durante la realización de la obra, se girarán visitas de inspección por personal facultativo de la empresa cliente, para hacer las comprobaciones que se crean oportunas. Es obligación del contratista, la conservación de la obra ya ejecutada hasta la recepción de la misma, por lo que el deterioro parcial o total de ella, aunque sea por agentes atmosféricos u otras causas, deberá ser reparado o reconstruido por su cuenta.

21. El contratista, deberá realizar la obra en el plazo mencionado a partir de la fecha del contrato, incurriendo en multa, por retraso de la ejecución siempre que éste no sea debido a causas de fuerza mayor. A la terminación de la obra, se hará una recepción provisional previo reconocimiento y examen por la dirección técnica, el depositario de efectos, el interventor y el jefe de servicio o un representante, estampando su conformidad el contratista.

22. Hecha la recepción provisional, se certificará al contratista el resto de la obra, reservándose la administración el importe de los gastos de conservación de la misma hasta su recepción definitiva y la fianza durante el tiempo señalado como plazo de garantía. La recepción definitiva se hará en las mismas condiciones que la provisional, extendiéndose el acta correspondiente. El Director Técnico propondrá a la Junta Económica la devolución de la fianza al contratista de acuerdo con las condiciones económicas legales establecidas.

23. Las tarifas para la determinación de honorarios, reguladas por orden de la Presidencia del Gobierno el 19 de Octubre de 1961, se aplicarán sobre el denominado en la actualidad "Presupuesto de Ejecución de Contrataz anteriormente llamado "Presupuesto de Ejecución Material"que hoy designa otro concepto.

Condiciones particulares.

La empresa consultora, que ha desarrollado el presente proyecto, lo entregará a la empresa cliente bajo las condiciones generales ya formuladas, debiendo añadirse las siguientes condiciones particulares:

1. La propiedad intelectual de los procesos descritos y analizados en el presente trabajo, pertenece por entero a la empresa consultora representada por el Ingeniero Director del Proyecto.

2. La empresa consultora se reserva el derecho a la utilización total o parcial de los resultados de la investigación realizada para desarrollar el siguiente proyecto, bien para su publicación o bien para su uso en trabajos o proyectos posteriores, para la misma empresa cliente o para otra.

3. Cualquier tipo de reproducción aparte de las reseñadas en las condiciones generales, bien sea para uso particular de la empresa cliente, o para cualquier otra aplicación, contará con autorización expresa y por escrito del Ingeniero Director del Proyecto, que actuará en representación de la empresa consultora.

4. En la autorización se ha de hacer constar la aplicación a que se destinan sus reproducciones así como su cantidad.

5. En todas las reproducciones se indicará su procedencia, explicitando el nombre del proyecto, nombre del Ingeniero Director y de la empresa consultora.

6. Si el proyecto pasa la etapa de desarrollo, cualquier modificación que se realice sobre él, deberá ser notificada al Ingeniero Director del Proyecto y a criterio de éste, la empresa consultora decidirá aceptar o no la modificación propuesta.

7. Si la modificación se acepta, la empresa consultora se hará responsable al mismo nivel que el proyecto inicial del que resulta el añadirla.

8. Si la modificación no es aceptada, por el contrario, la empresa consultora declinará toda responsabilidad que se derive de la aplicación o influencia de la misma.

9. Si la empresa cliente decide desarrollar industrialmente uno o varios productos en los que resulte parcial o totalmente aplicable el estudio de este proyecto, deberá comunicarlo a la empresa consultora.

10. La empresa consultora no se responsabiliza de los efectos laterales que se puedan producir en el momento en que se utilice la herramienta objeto del presente proyecto para la realización de otras aplicaciones.

11. La empresa consultora tendrá prioridad respecto a otras en la elaboración de los proyectos auxiliares que fuese necesario desarrollar para dicha aplicación industrial, siempre que no haga explícita renuncia a este hecho. En este caso, deberá autorizar expresamente los proyectos presentados por otros.

12. El Ingeniero Director del presente proyecto, será el responsable de la dirección de la aplicación industrial siempre que la empresa consultora lo estime oportuno. En caso contrario, la persona designada deberá contar con la autorización del mismo, quien delegará en él las responsabilidades que ostente.

Figuras generadas durante la experimentación

A continuación y, durante toda la extensión de este anexo, se mostrarán las figuras generadas durante los experimentos realizados.

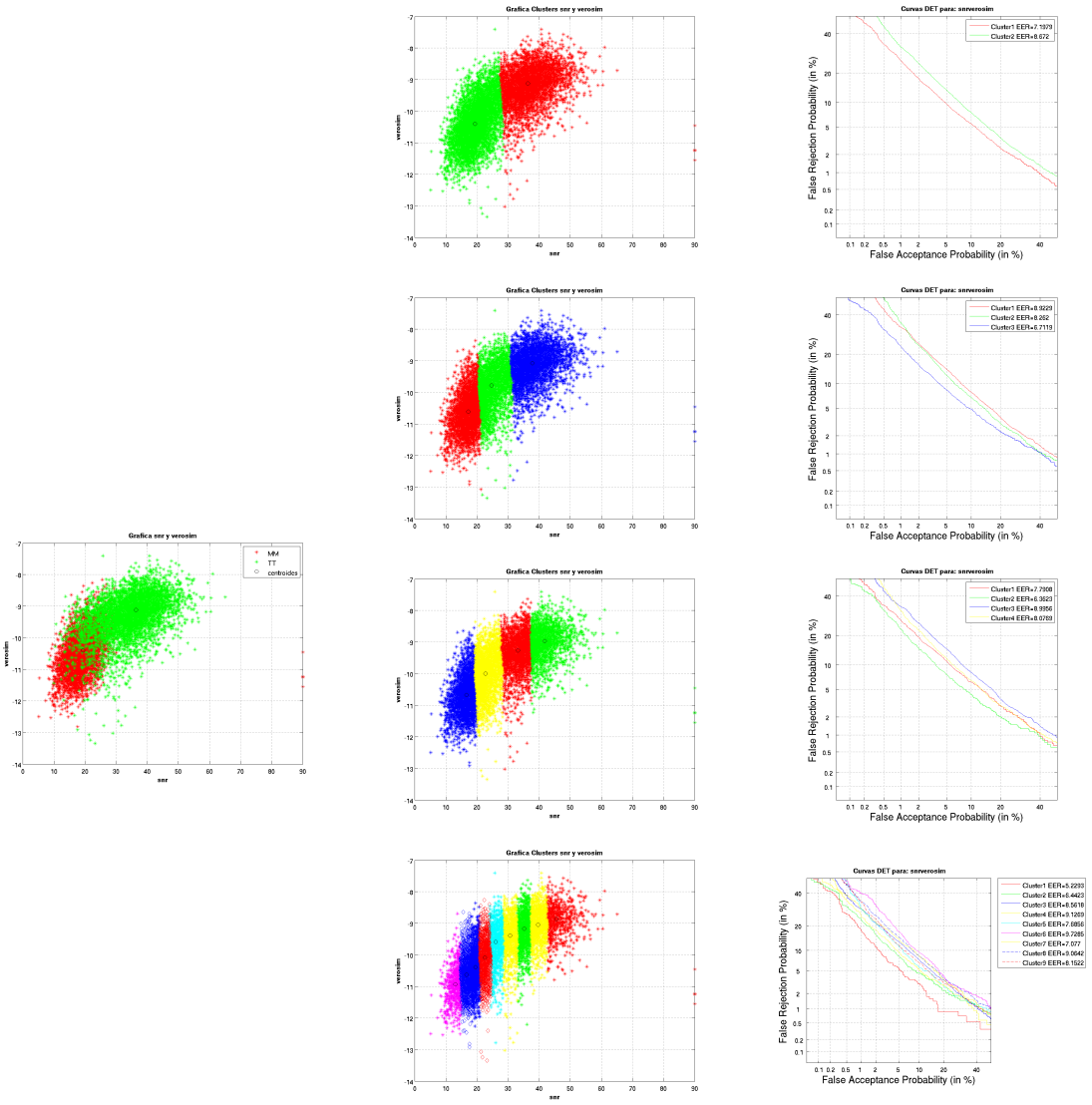
Estas figuras son las más representativas dentro de su tipo de experimento debido a que son las que mejor reflejan los agrupamientos de datos.

El motivo por el que no se muestran todas las figuras es por su extensión ya que existen numerosas gráficas realizadas durante este proyecto final de carrera.

En total se han generado más de mil gráficas correspondientes con los distintos experimentos:

- Distinto tipo de algoritmo de agrupamiento.
  - Medidas de distancias empleadas para el caso del algoritmo K-means
  - Diferentes matrices de covarianza empleadas en el caso de clasificación por GMM.
- Variación en el número de clusters.
- Empleo de distintas bases de datos.

2 indicadores de degradación: Snr verosim

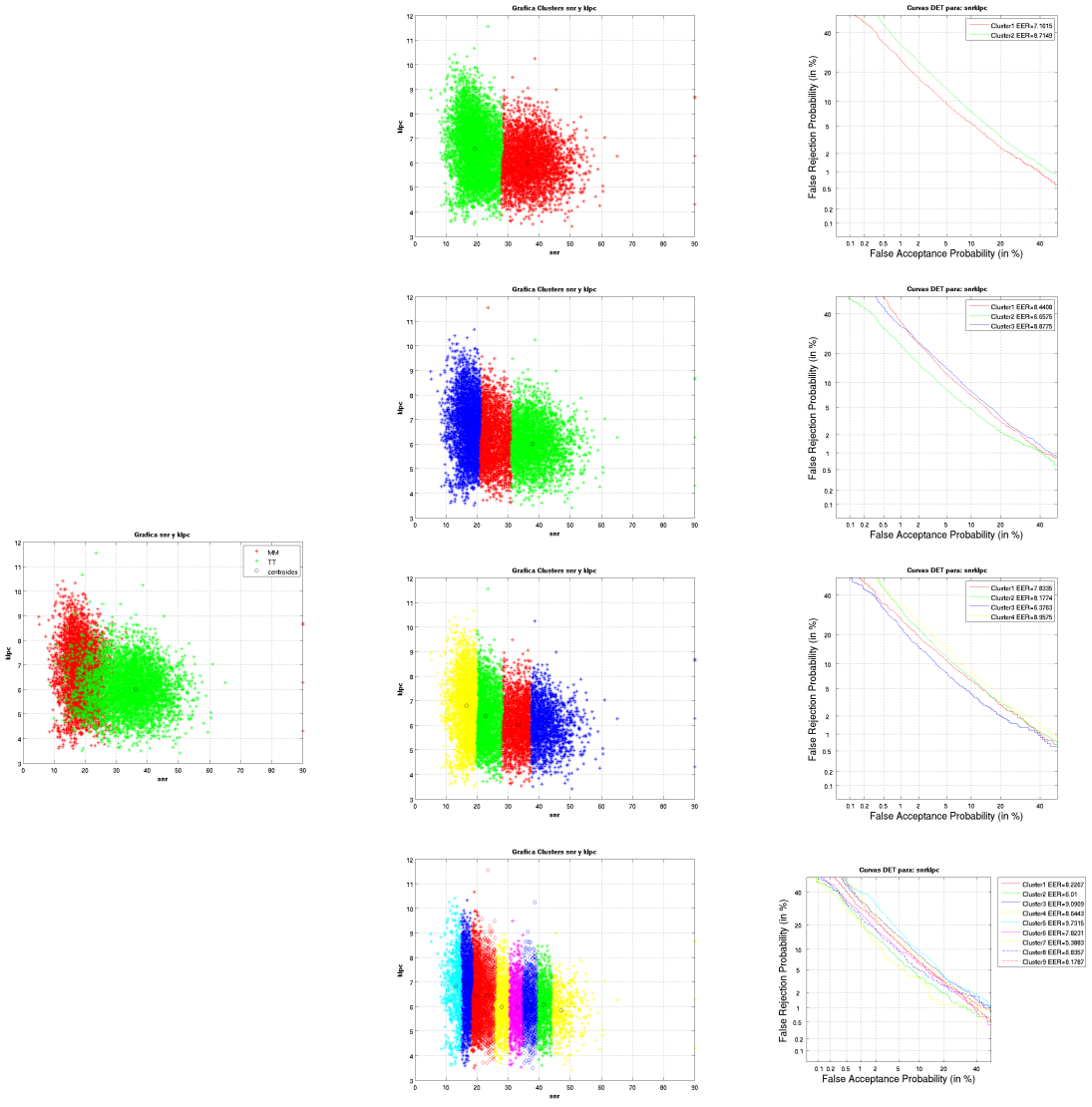


Cuadro 1: Ficheros agrupados y curvas DET para snr verosim con K-means-Cityblock

.0.1. NIST 2008 - K-means con distancia Citblock

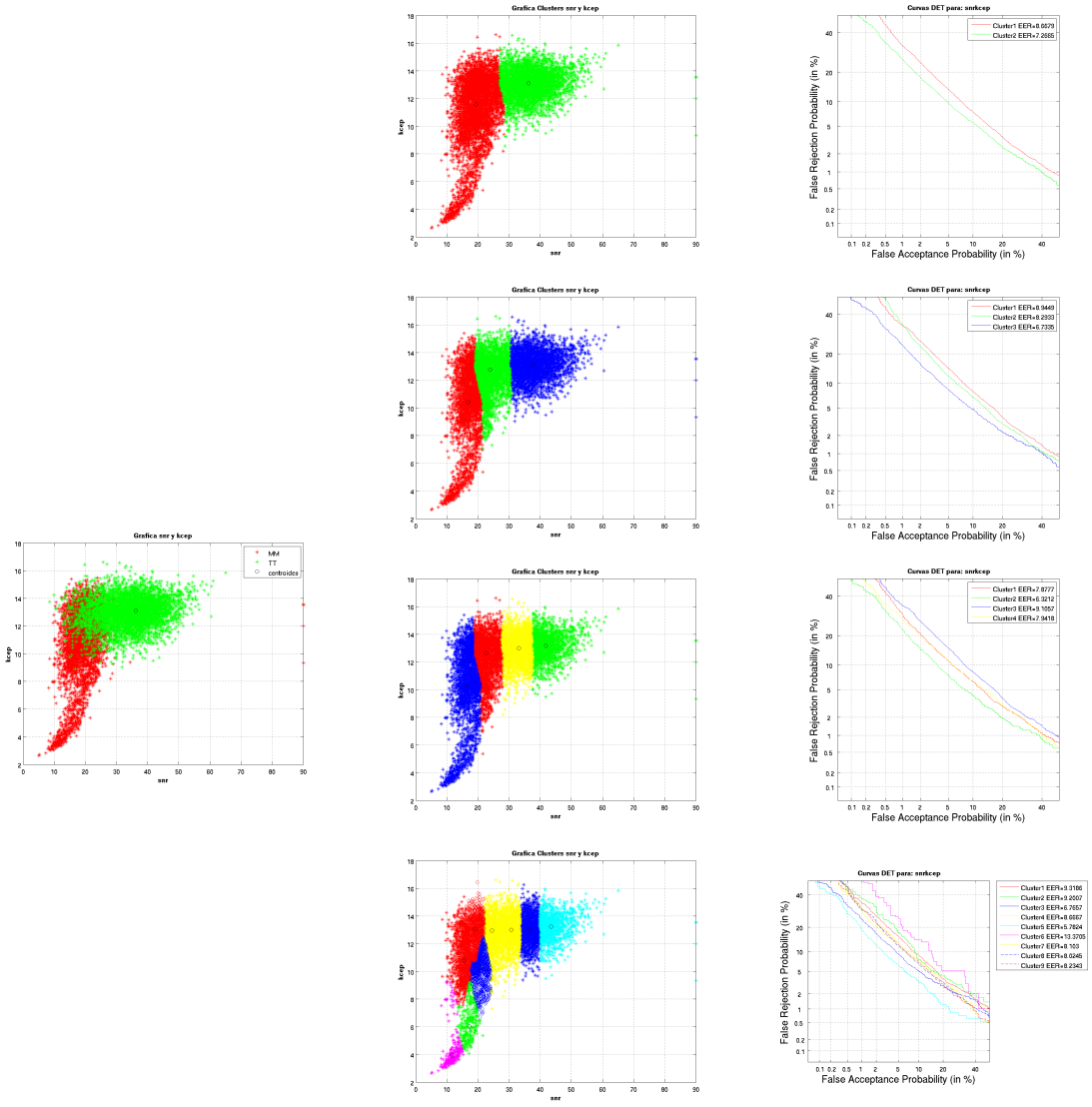
En los cuadros del 1 al 20 se muestran las agrupaciones K-means generadas para la base de datos NIST 2008. El tipo de indicador de degradación empleado se muestra en la cabecera de la sección de la tabla y, a continuación, se muestran las diferentes figuras asociadas. En la primera columna se presenta la gráfica inicial, sin haberse realizado sobre ella ninguna agrupación. La segunda columna muestra la gráfica ya agrupada en función del número de clusters y, por último, la tercera columna muestra la curva DET correspondiente con ese experimento.

2 indicadores de degradación: Snr klpc



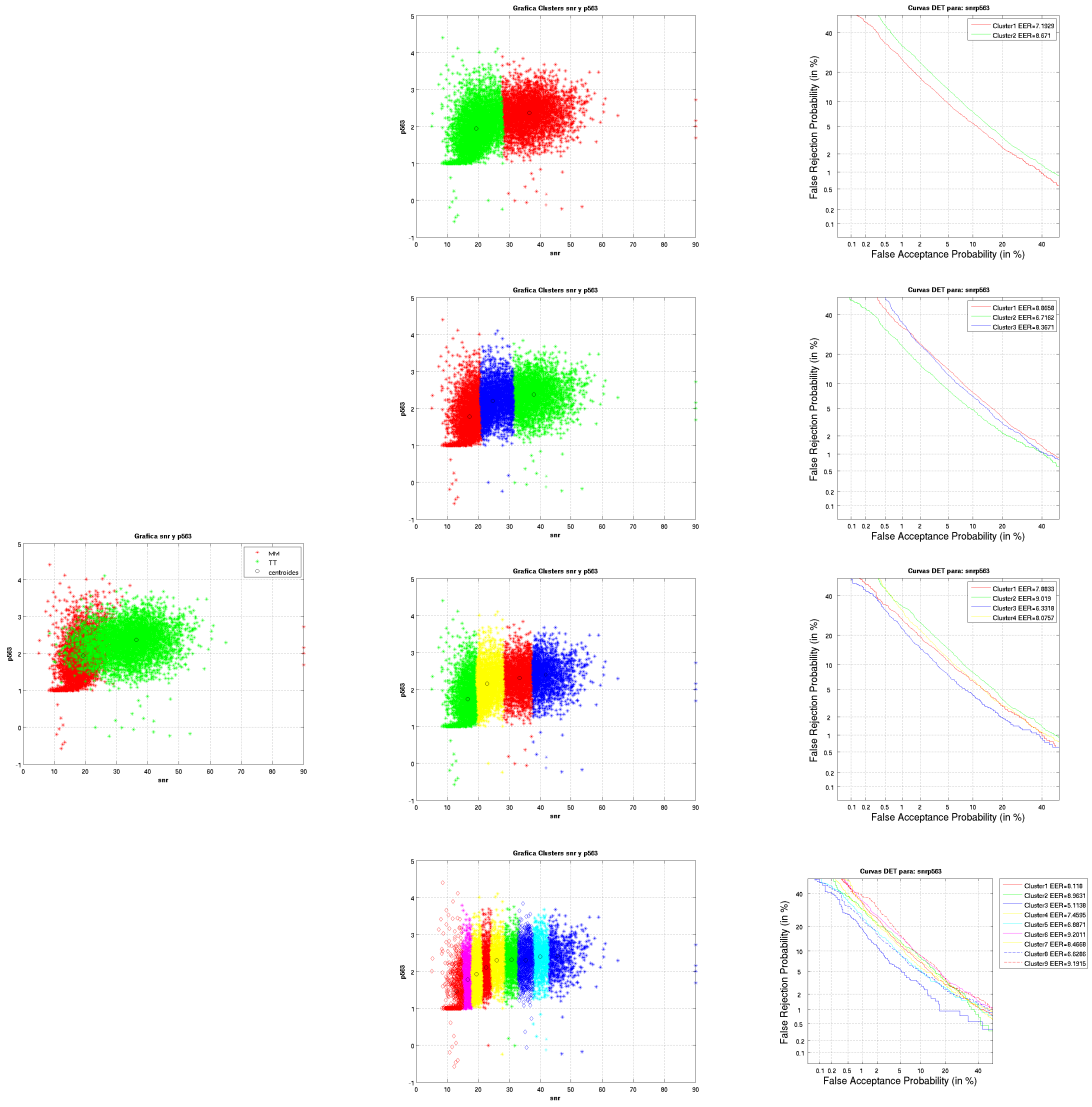
Cuadro 2: Ficheros agrupados y curvas DET para snr klpc con K-means-Cityblock

2 indicadores de degradación: Snr kcep



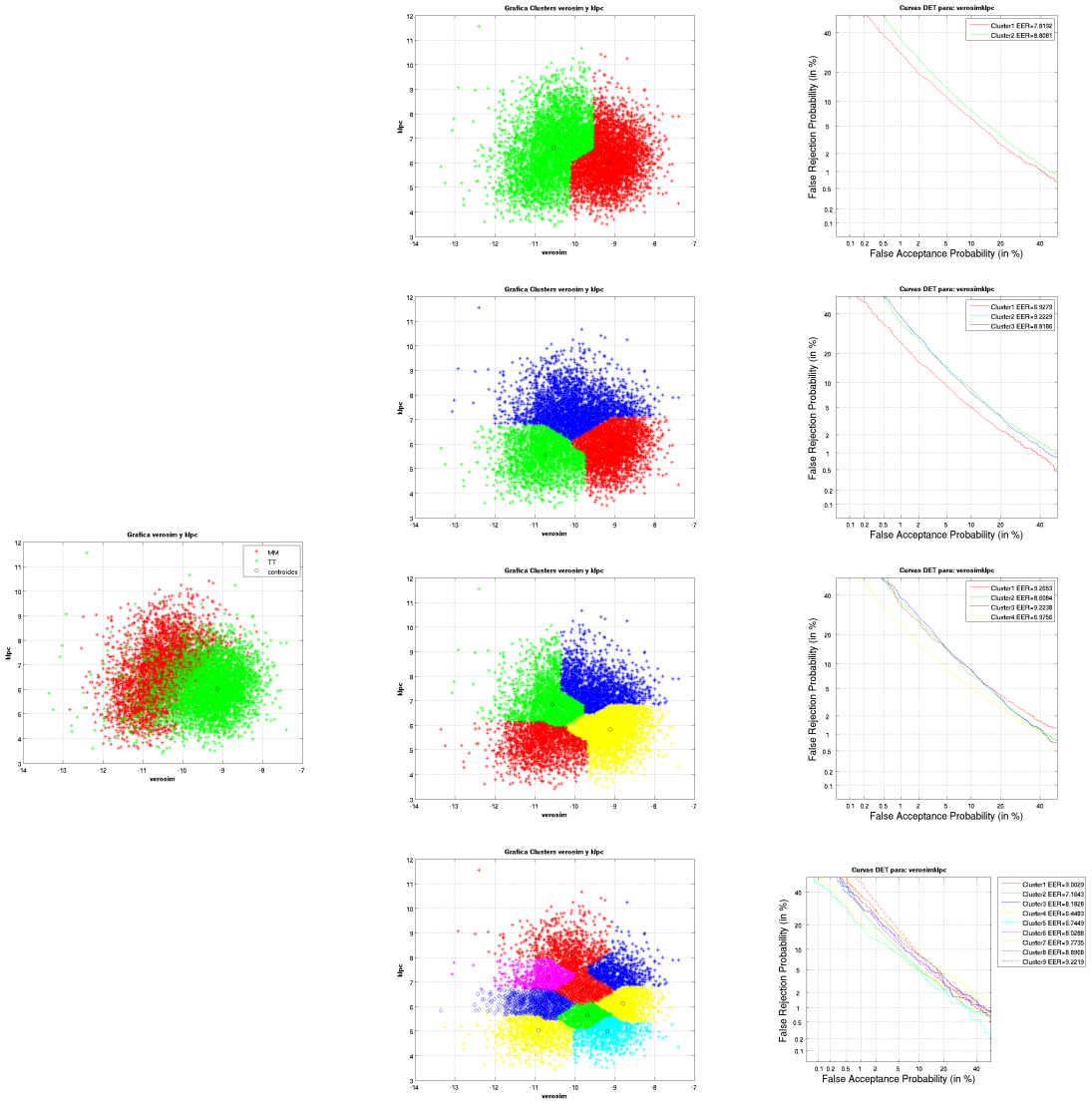
Cuadro 3: Ficheros agrupados y curvas DET para snr kcep con K-means-Cityblock

2 indicadores de degradación: Snr P.563



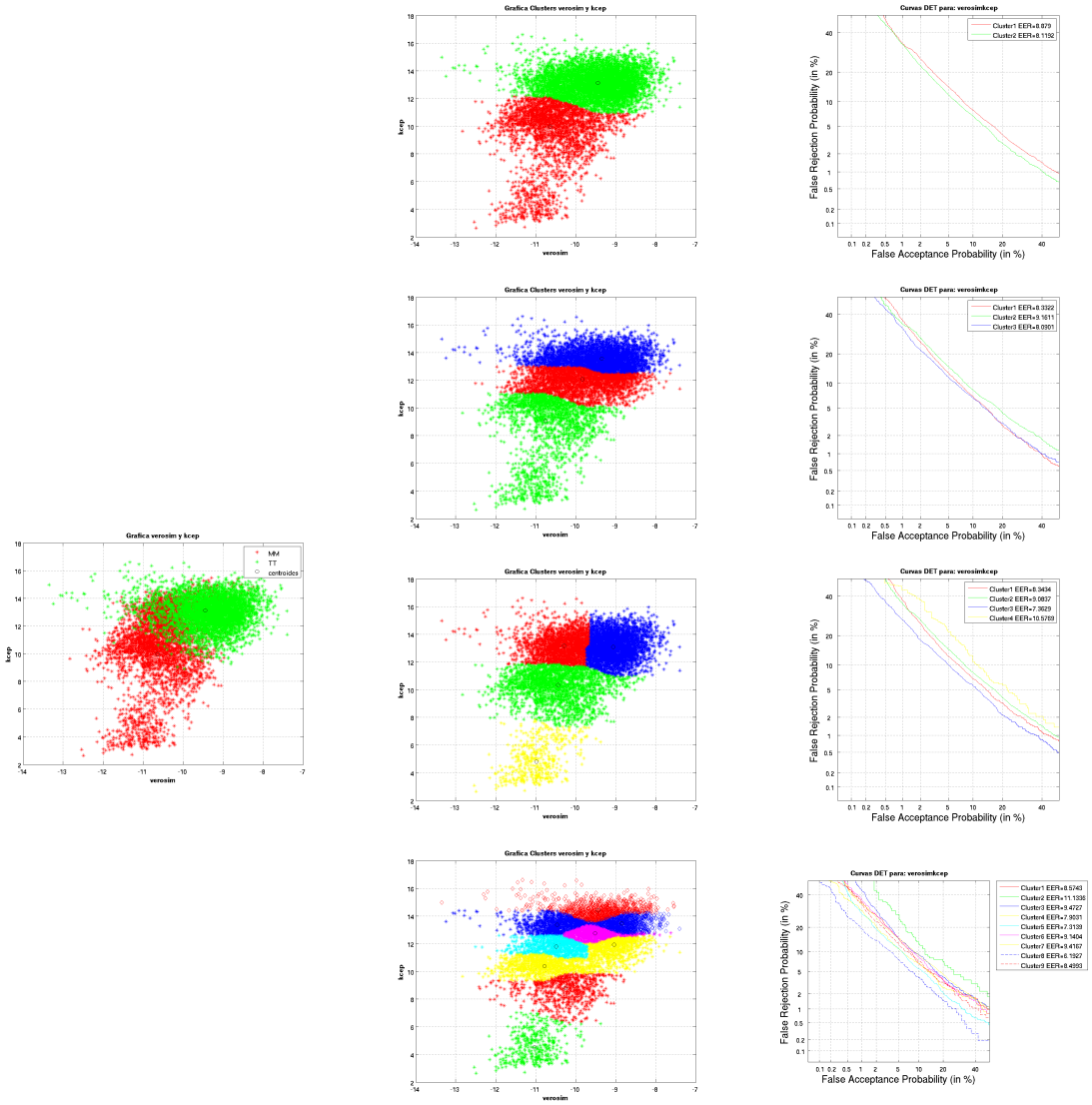
Cuadro 4: Ficheros agrupados y curvas DET para snr P.563 con K-means-Cityblock

2 indicadores de degradación: verosim klpc



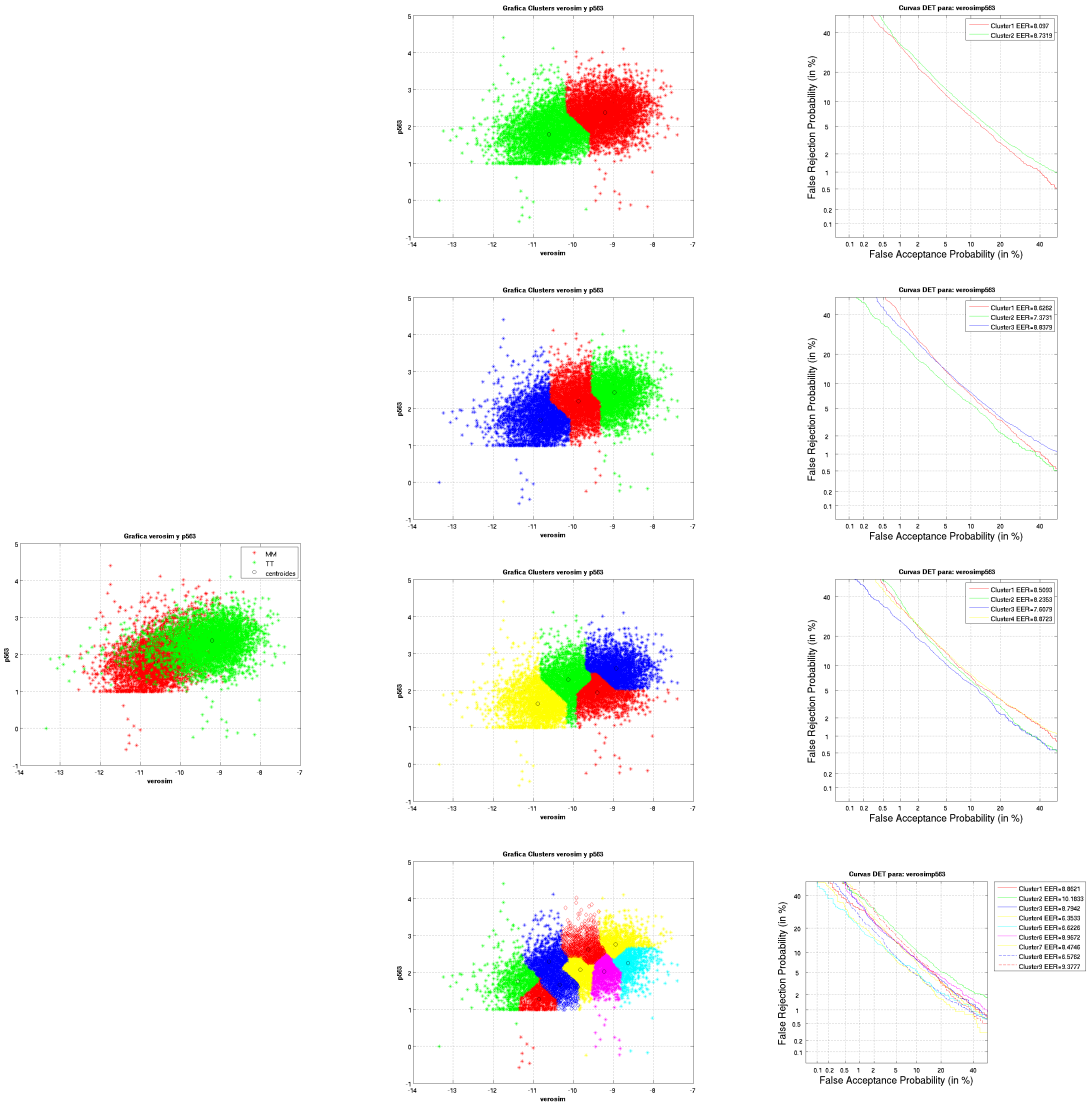
Cuadro 5: Ficheros agrupados y curvas DET para verosim klpc con K-means-Cityblock

2 indicadores de degradación: verosim kcep



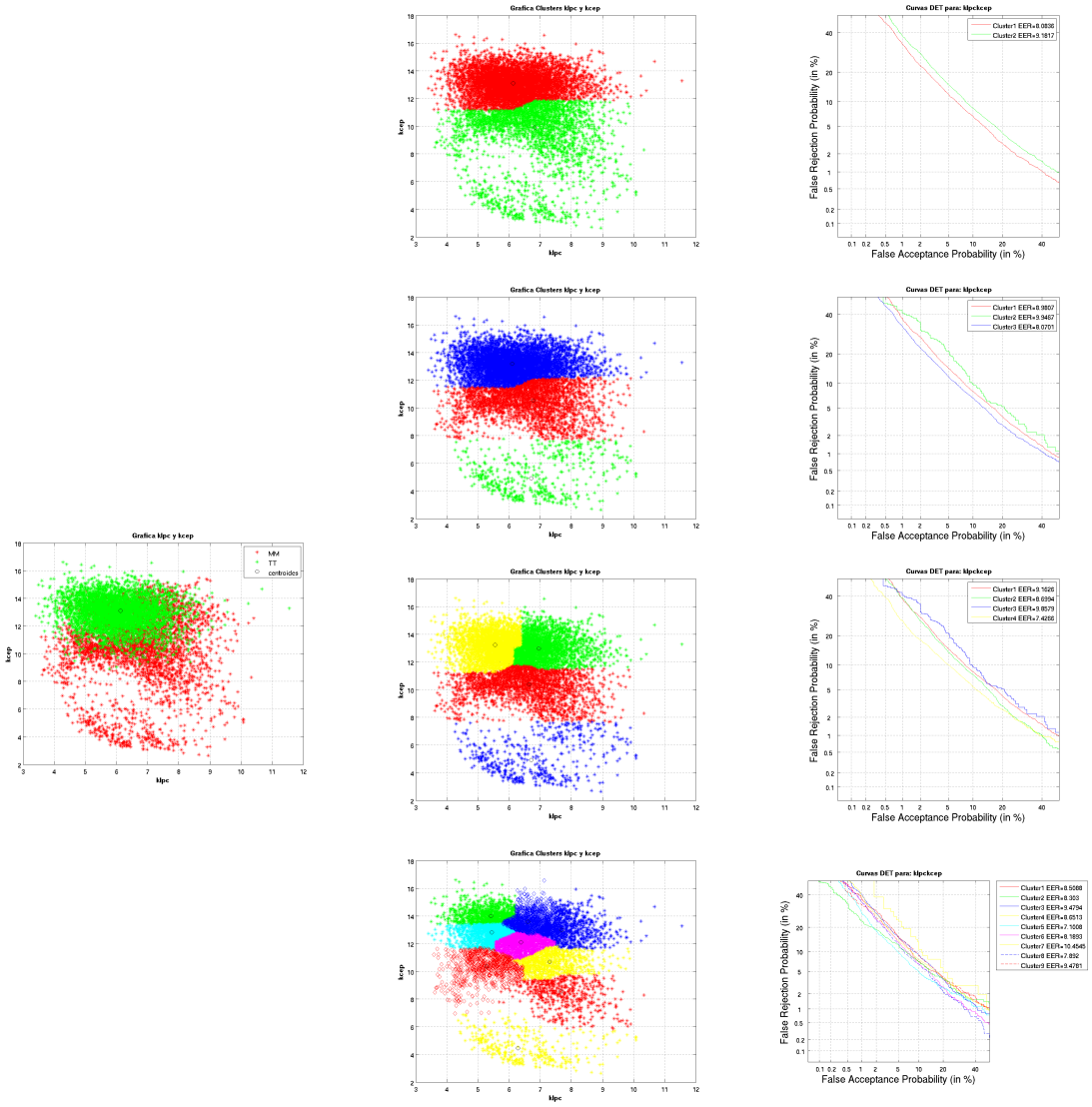
Cuadro 6: Ficheros agrupados y curvas DET para verosim kcep con K-means-Cityblock

2 indicadores de degradación: verosim P.563



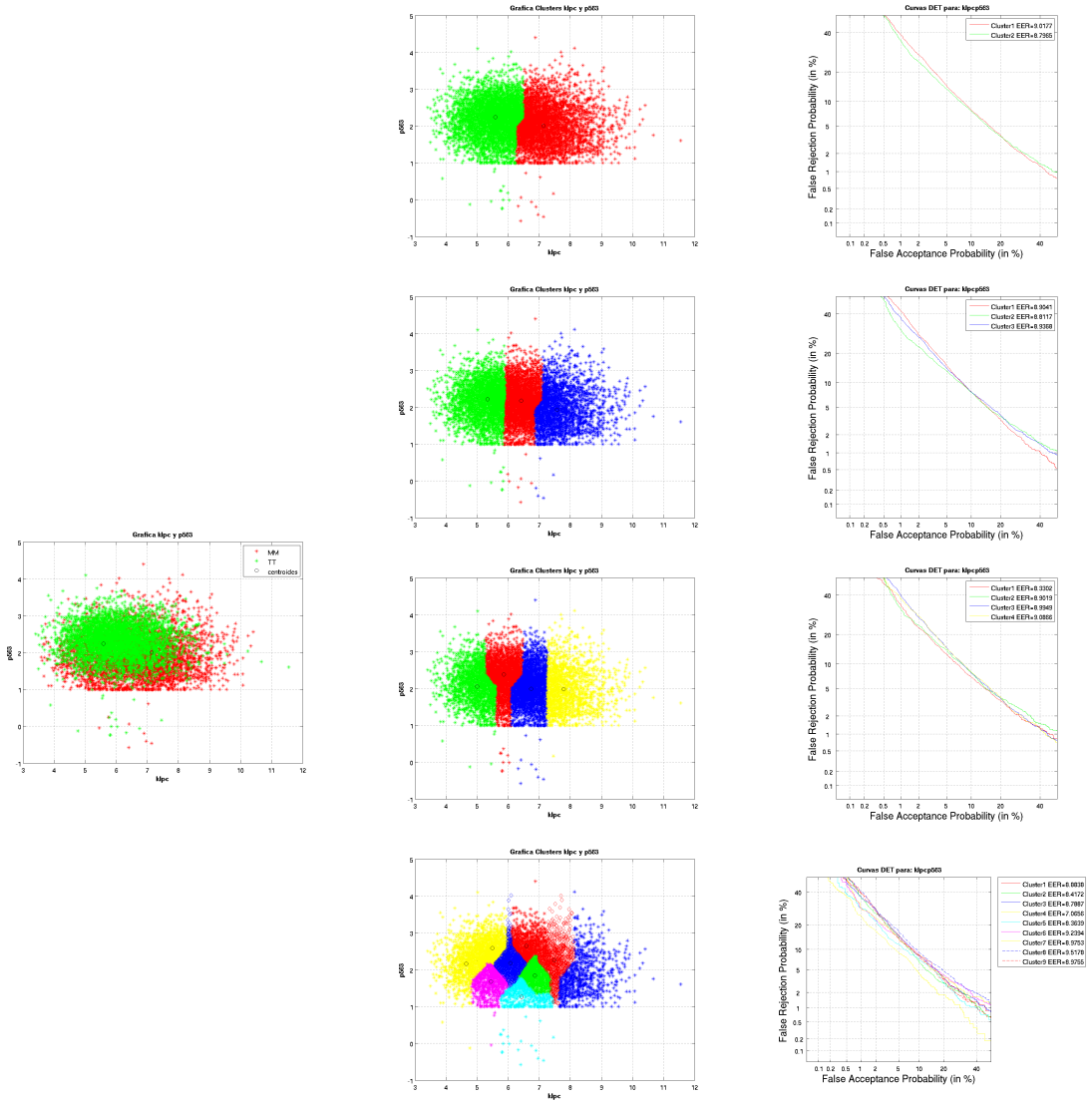
Cuadro 7: Ficheros agrupados y curvas DET para verosim P.563 con K-means-Cityblock

2 indicadores de degradación: klpc kcep



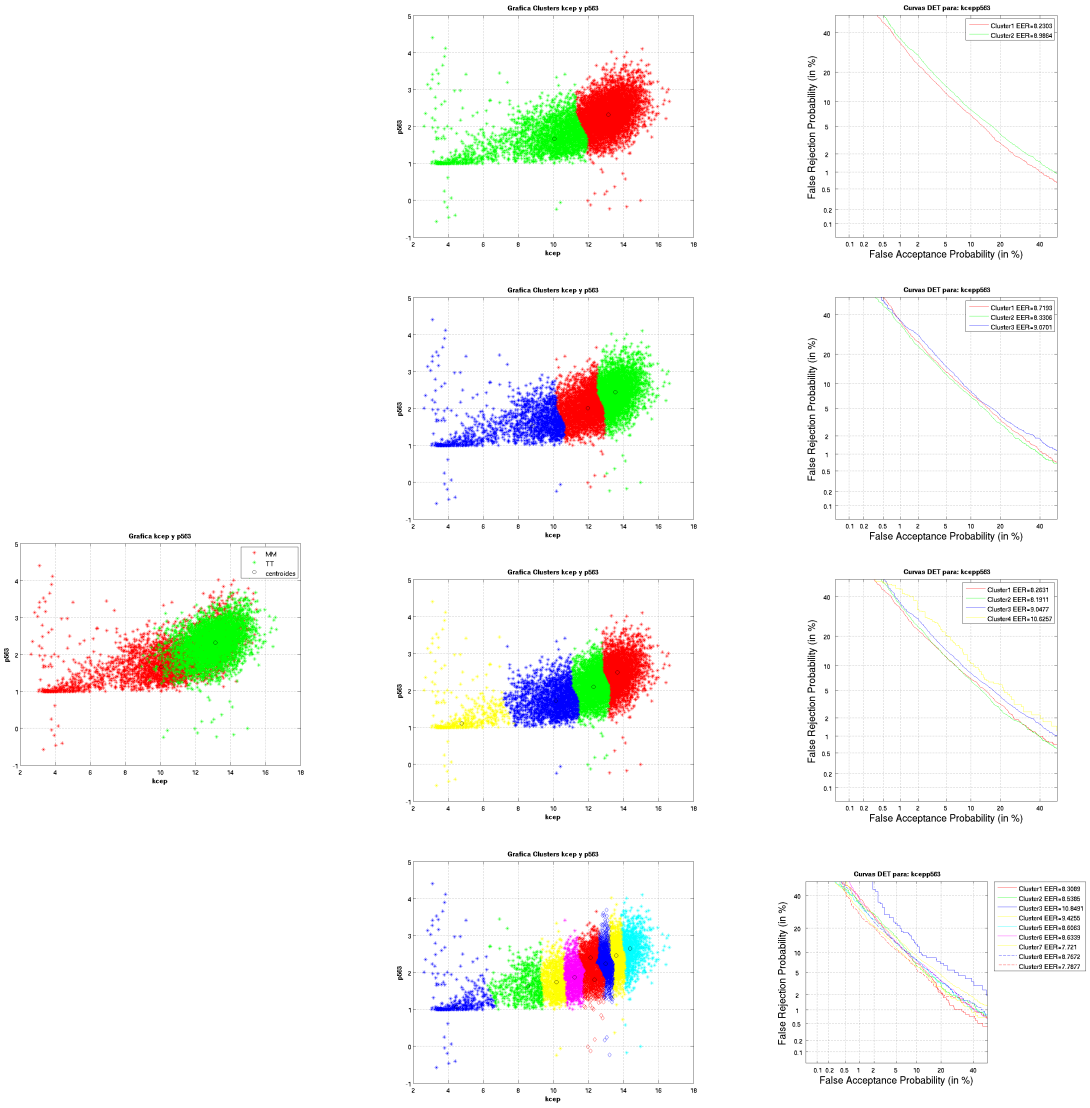
Cuadro 8: Ficheros agrupados y curvas DET para klpc kcep con K-means-Cityblock

2 indicadores de degradación: klpc P.563



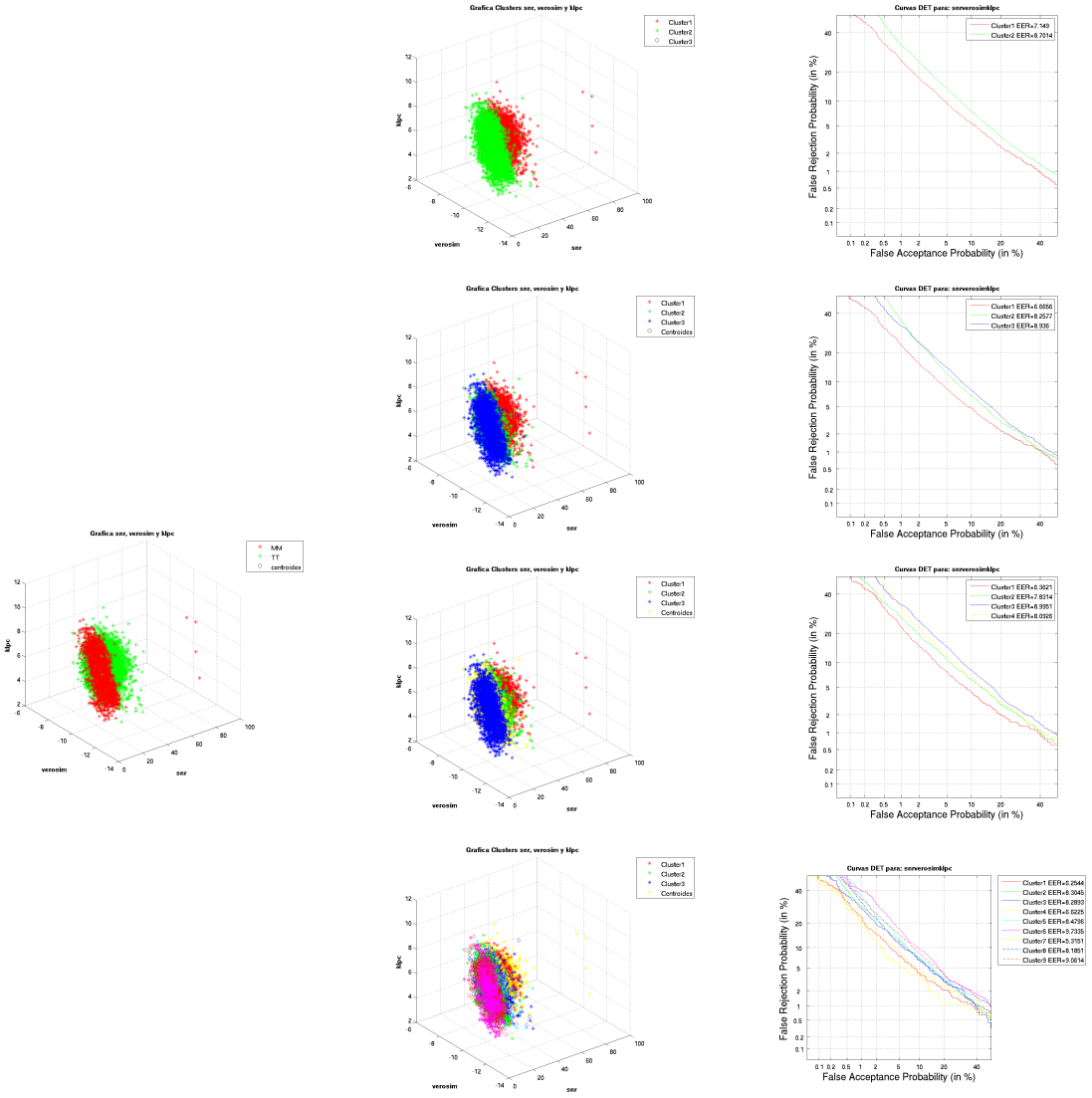
Cuadro 9: Ficheros agrupados y curvas DET para klpc P.563 con K-means-Cityblock

2 indicadores de degradación: kcep P.563



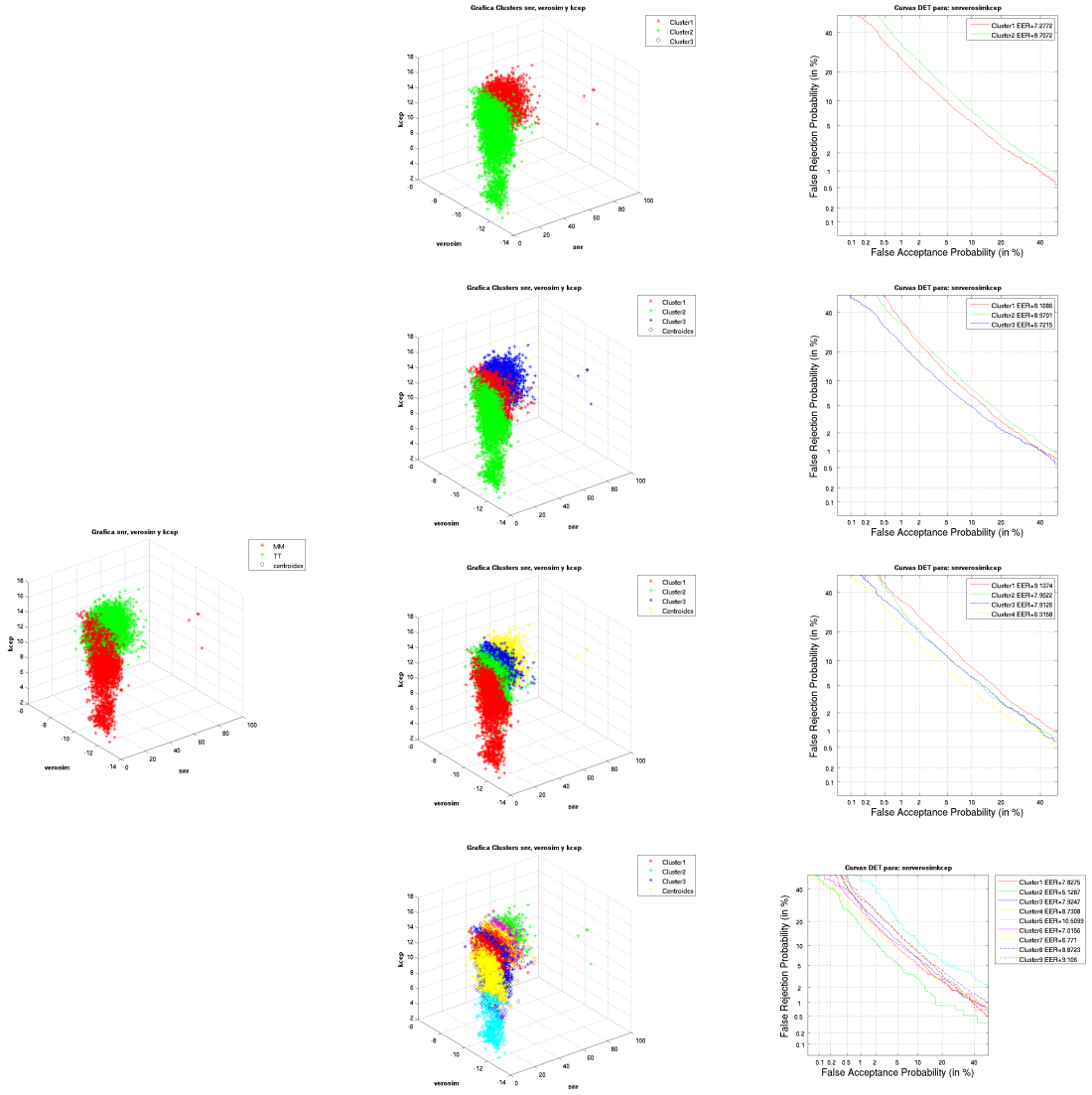
Cuadro 10: Ficheros agrupados y curvas DET para kcep P.563 con K-means-Cityblock

3 indicadores de degradación: snr verosim klpc



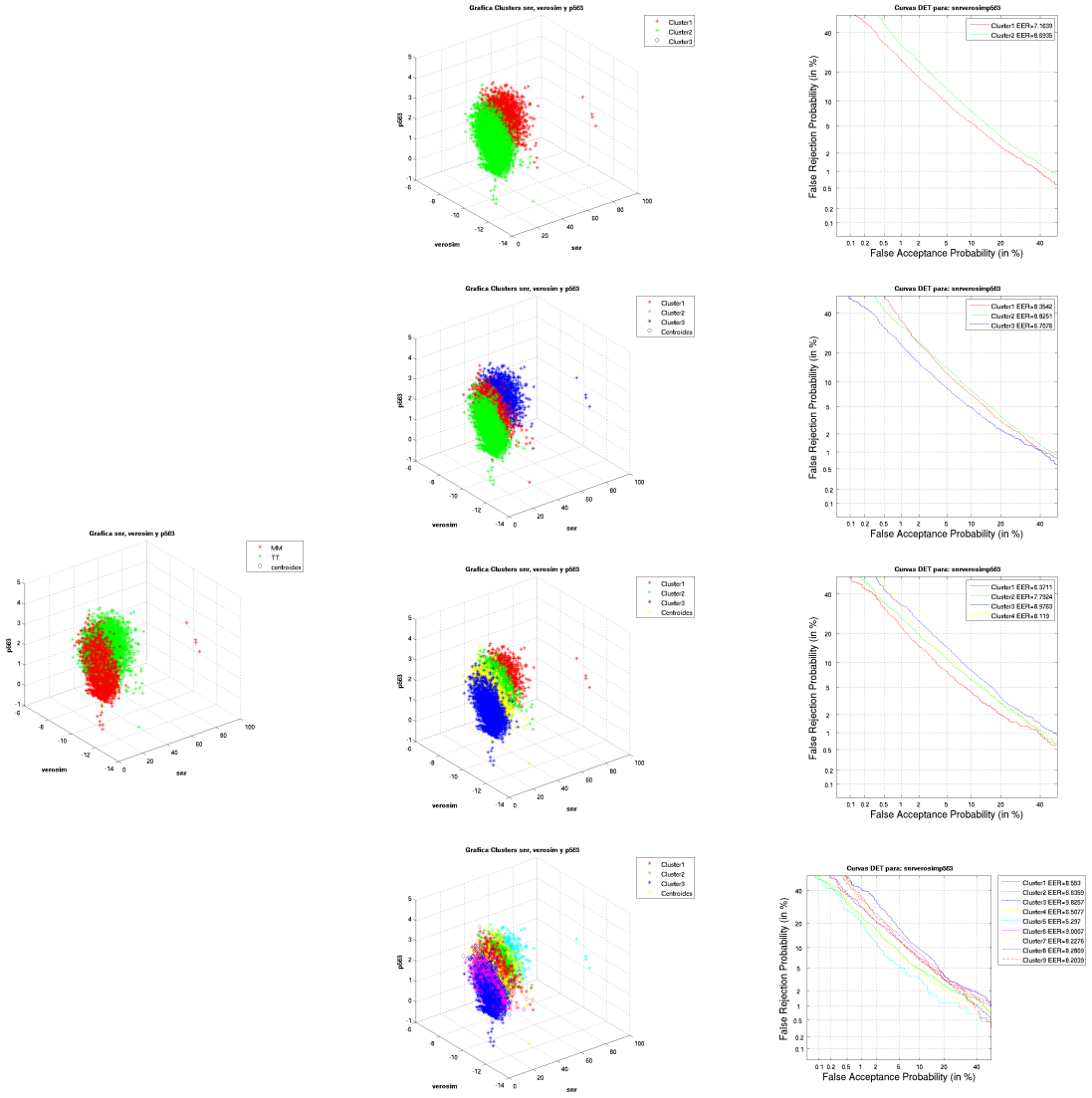
Cuadro 11: Ficheros agrupados y curvas DET para snr verosim klpc con K-means-Cityblock

3 indicadores de degradación: snr verosim kcep



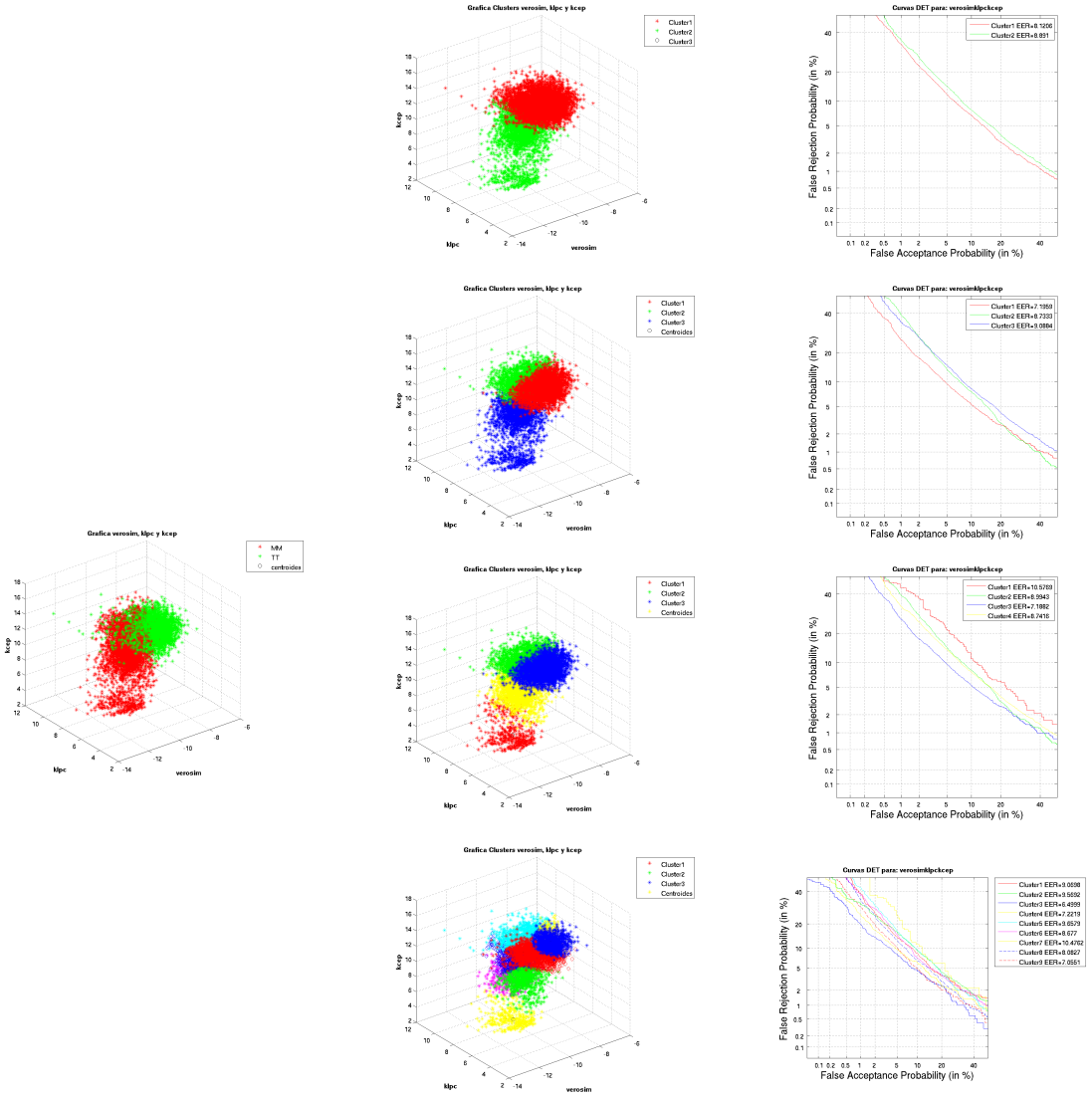
Cuadro 12: Ficheros agrupados y curvas DET para snr verosim kcep con K-means-Cityblock

3 indicadores de degradación: snr verosim P.563



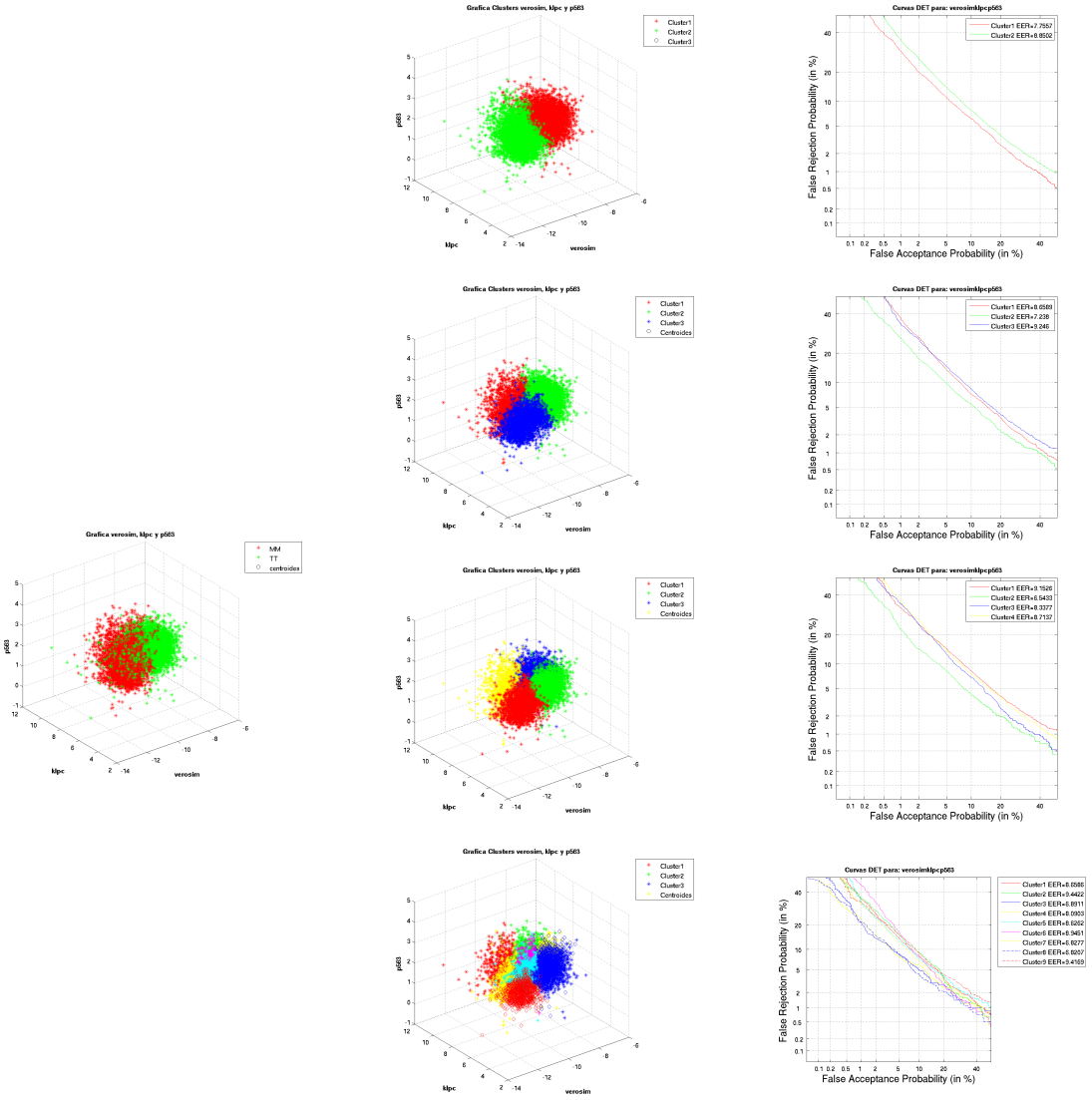
Cuadro 13: Ficheros agrupados y curvas DET para snr verosim P.563 con K-means-Cityblock

3 indicadores de degradación: verosim klpc kcep



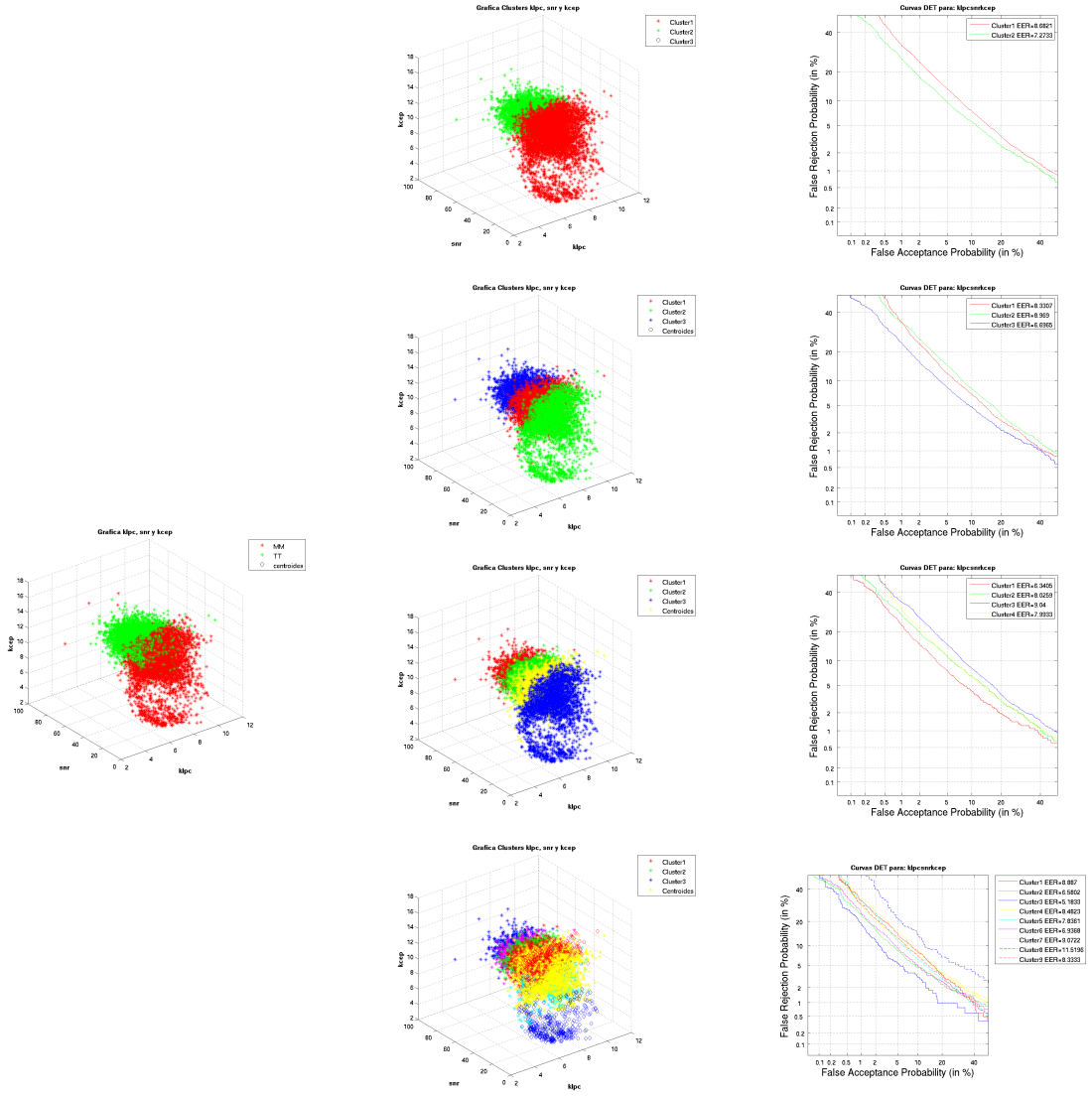
Cuadro 14: Ficheros agrupados y curvas DET para verosim klpc kcep con K-means-Cityblock

3 indicadores de degradación: verosim klpc P.563



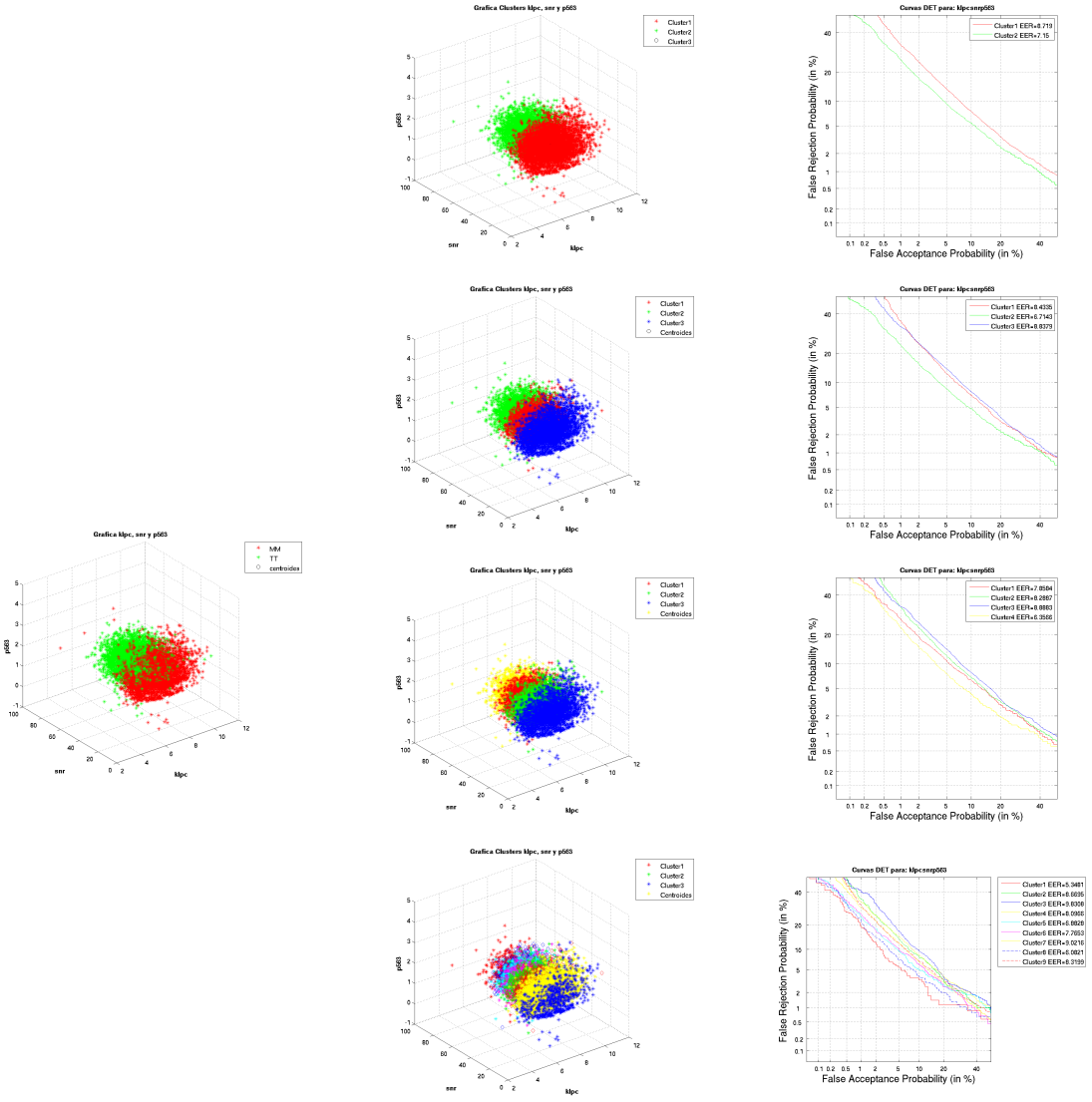
Cuadro 15: Ficheros agrupados y curvas DET para verosim klpc P.563 con K-means-Cityblock

3 indicadores de degradación: klpc snr kcep



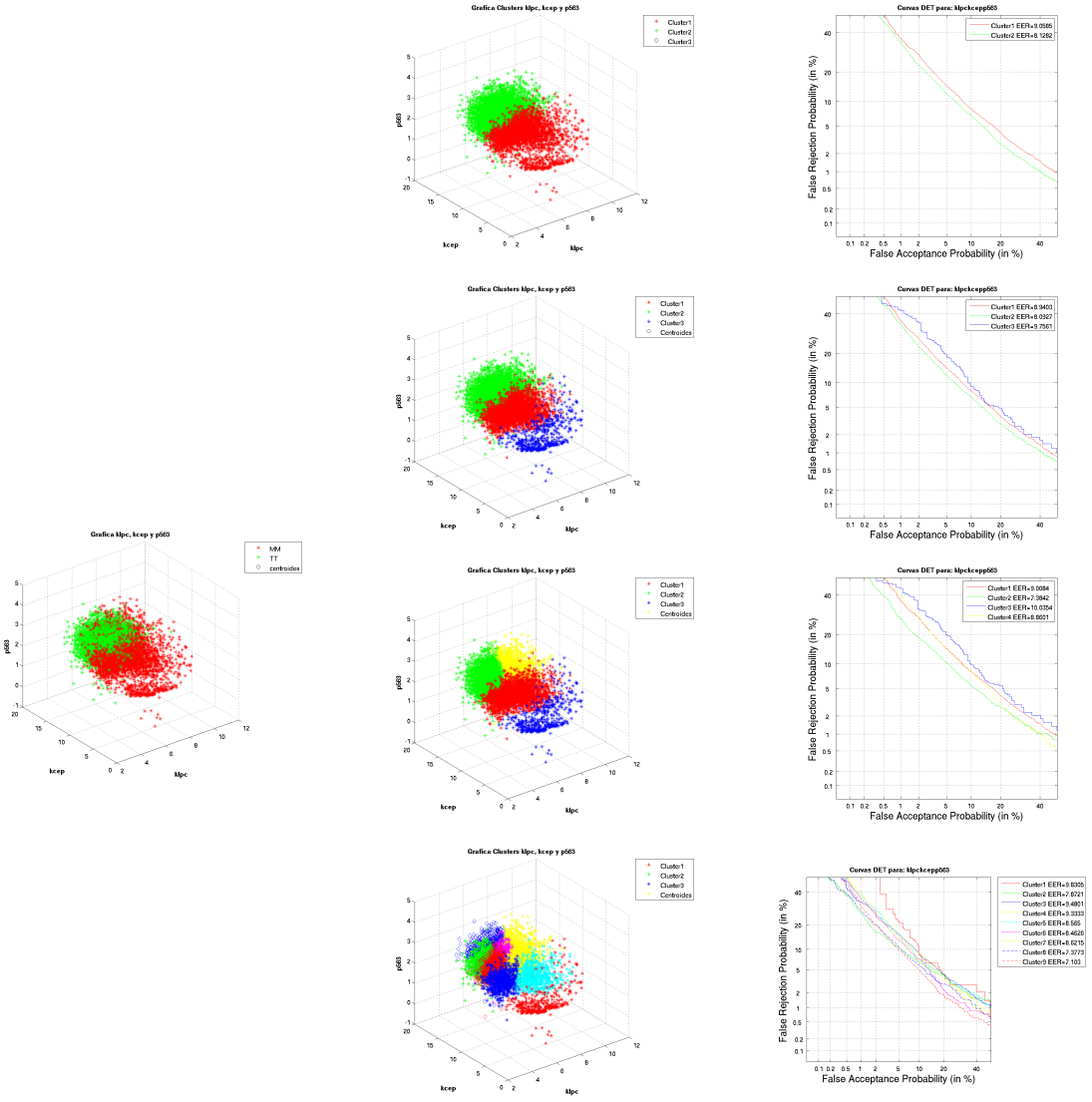
Cuadro 16: Ficheros agrupados y curvas DET para klpc snr kcep con K-means-Cityblock

3 indicadores de degradación: klpc snr P.563



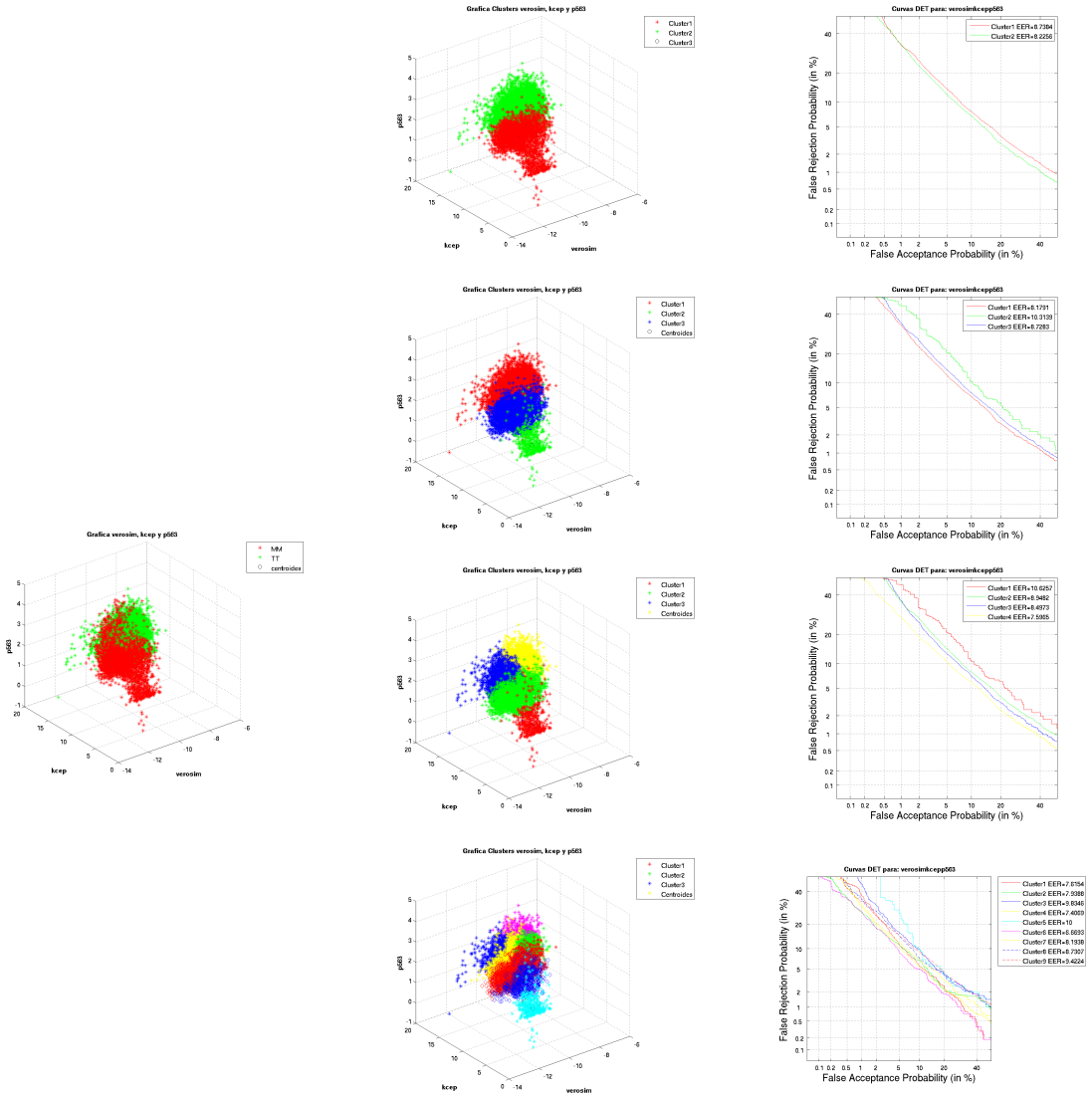
Cuadro 17: Ficheros agrupados y curvas DET para klpc snr P.563 con K-means-Cityblock

3 indicadores de degradación: klpc kcep P.563



Cuadro 18: Ficheros agrupados y curvas DET para klpc kcep P.563 con K-means-Cityblock

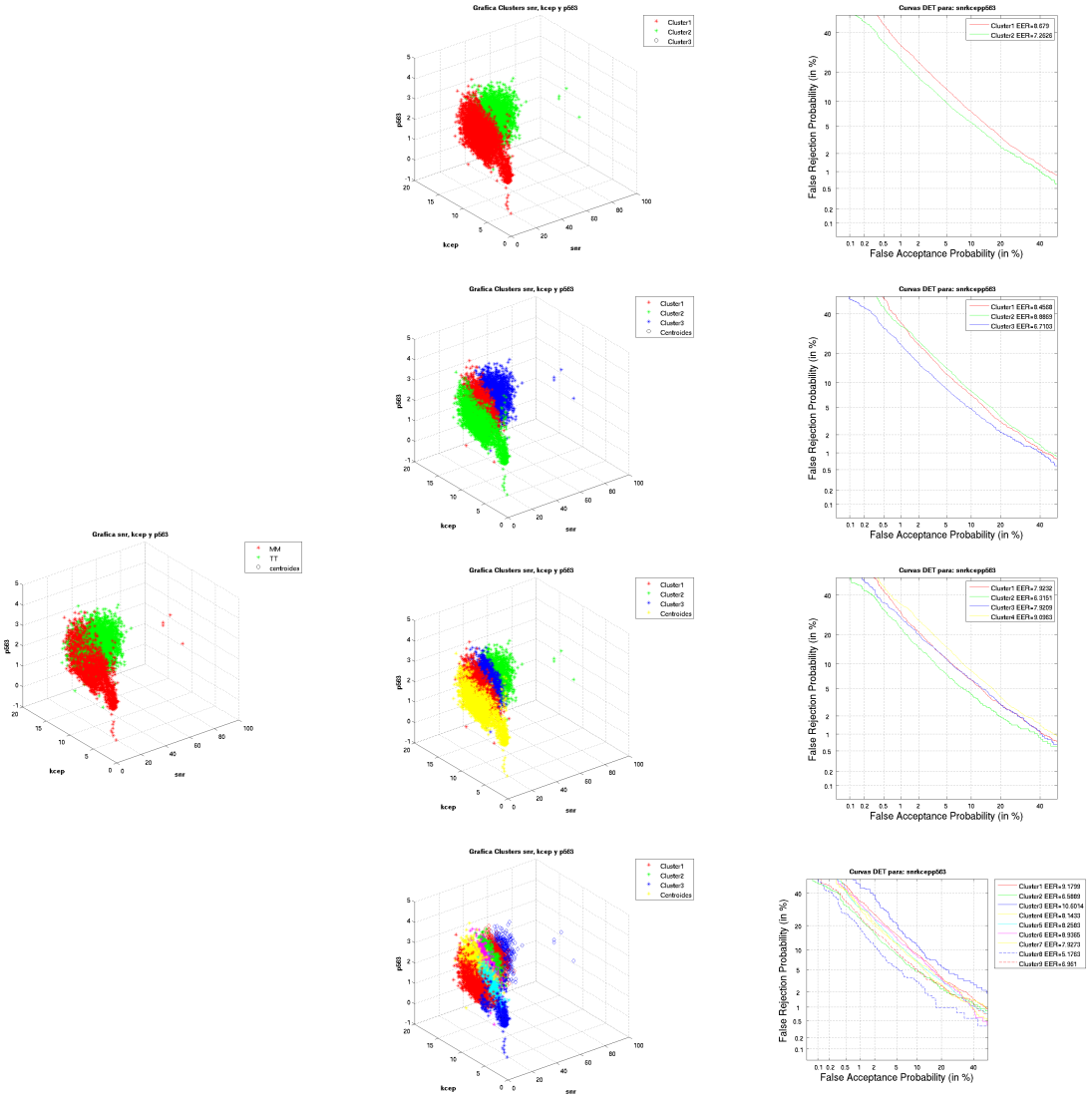
3 indicadores de degradación: verosim kcep P.563



Cuadro 19: Ficheros agrupados y curvas DET para verosim kcep P.563 con K-means-Cityblock

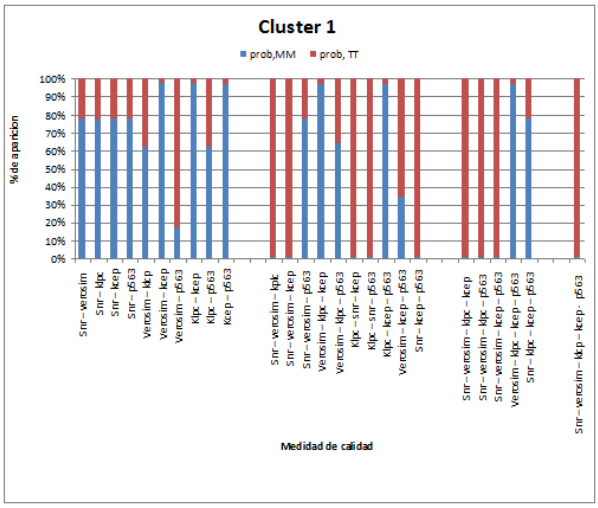
En los cuadros siguientes (Cuadro21 al Cuadro24) se muestran la distribución por agrupamiento de los distintos ficheros en función de su origen, telefónico o microfónico. Estas gráficas están relacionadas con las gráficas presentes en las tablas mostradas anteriormente (Cuadro 1 al Cuadro20).

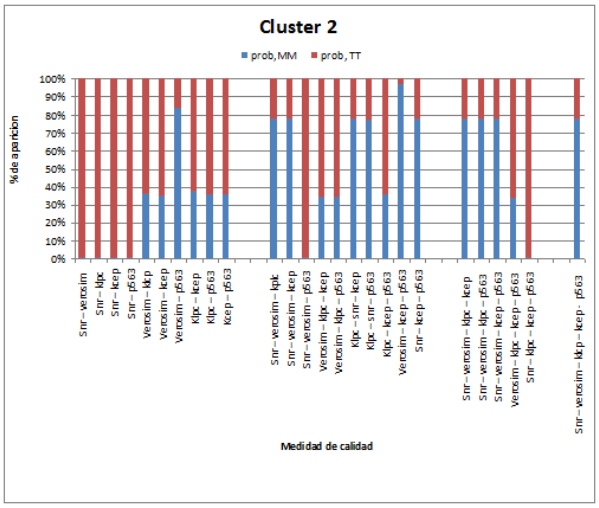
3 indicadores de degradación: snr kcep P.563



Cuadro 20: Ficheros agrupados y curvas DET para snr kcep P.563 con K-means-Cityblock

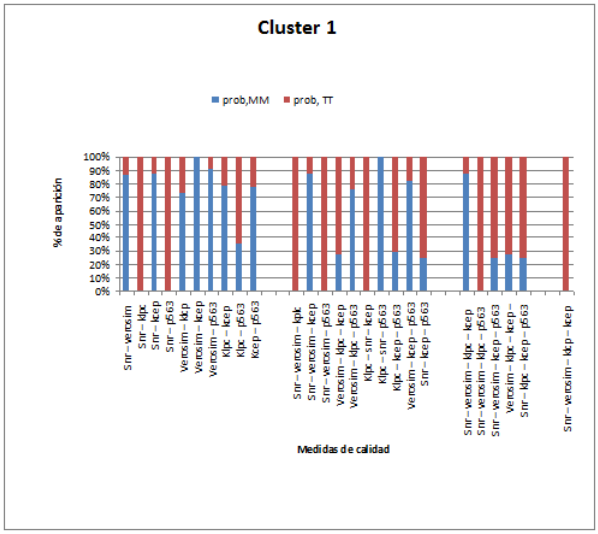
2 agrupamientos

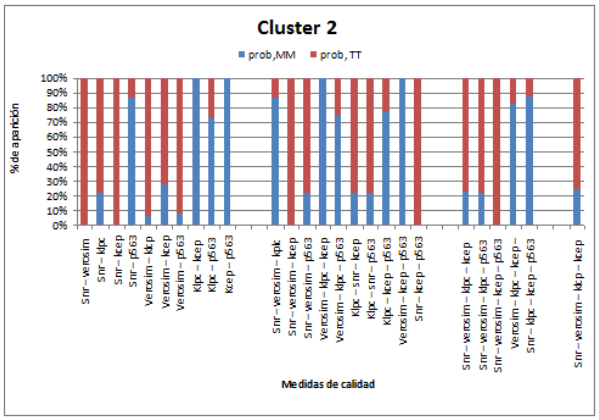


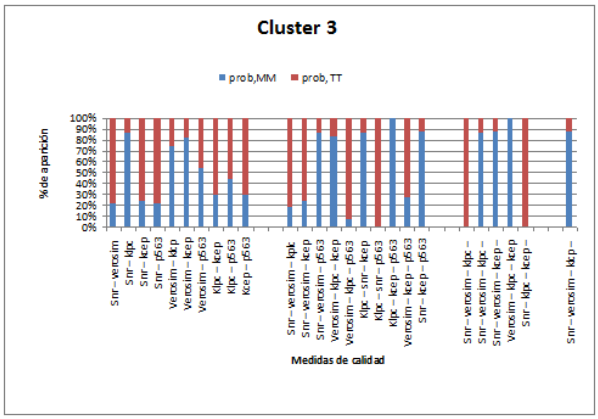


Cuadro 21: Origen del fichero por cluster para dos agrupaciones

3 agrupamientos

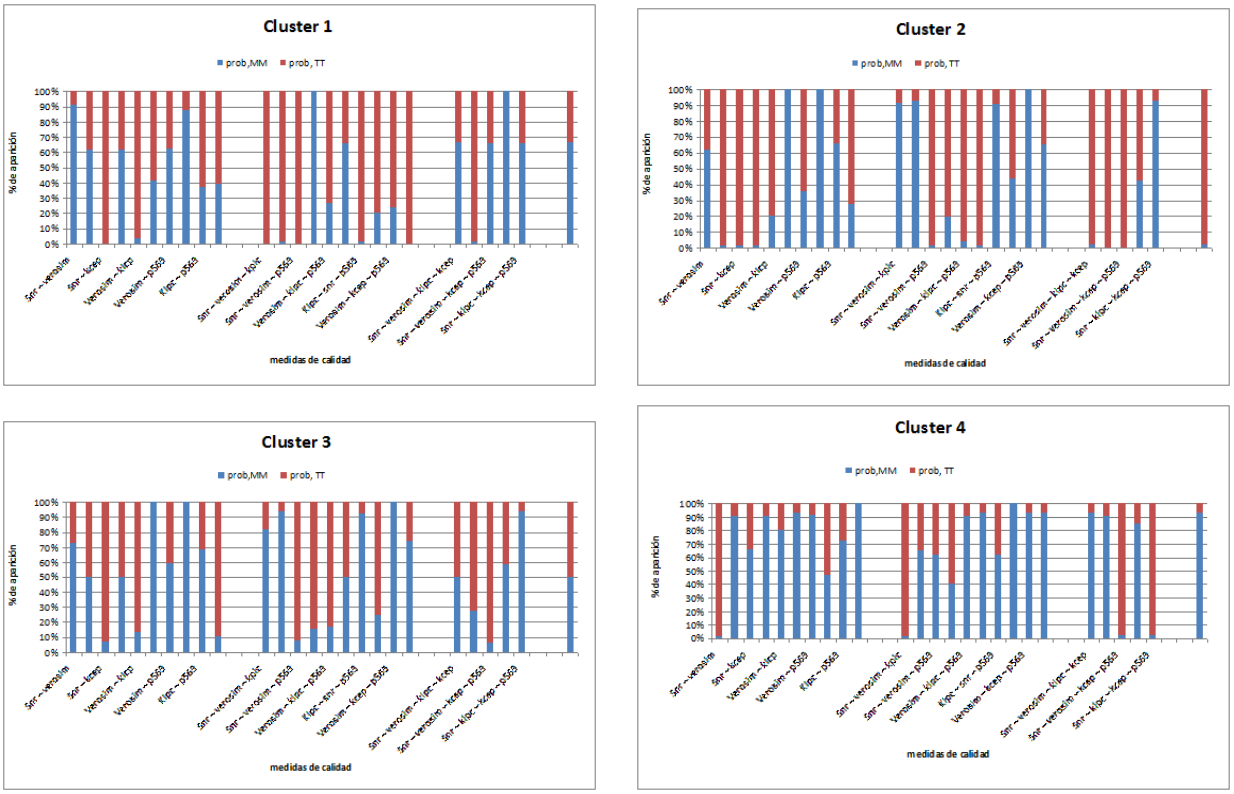






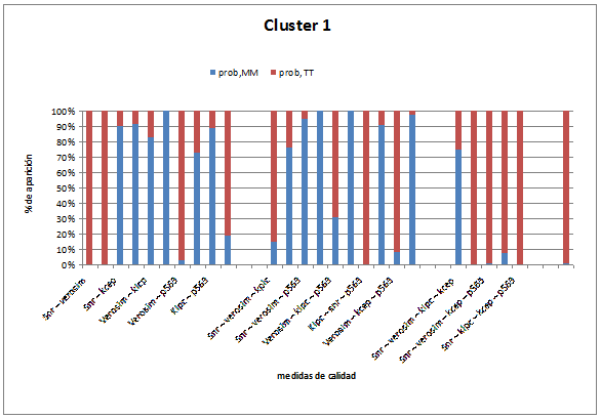
Cuadro 22: Origen del fichero por cluster para tres agrupaciones

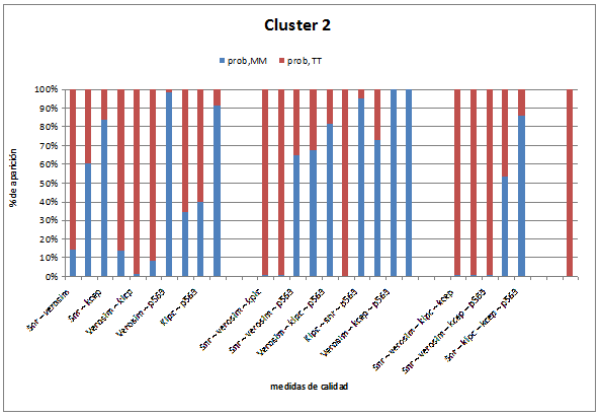
4 agrupamientos

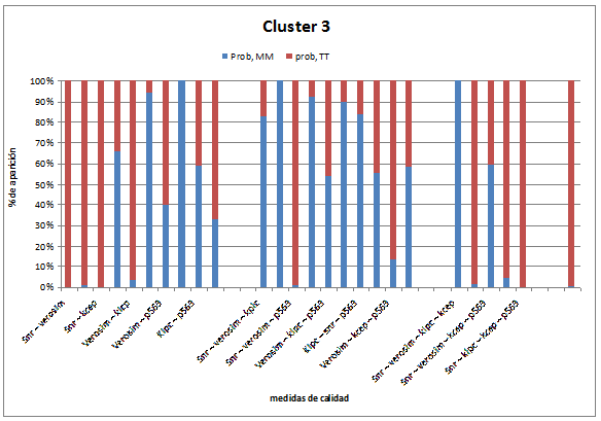


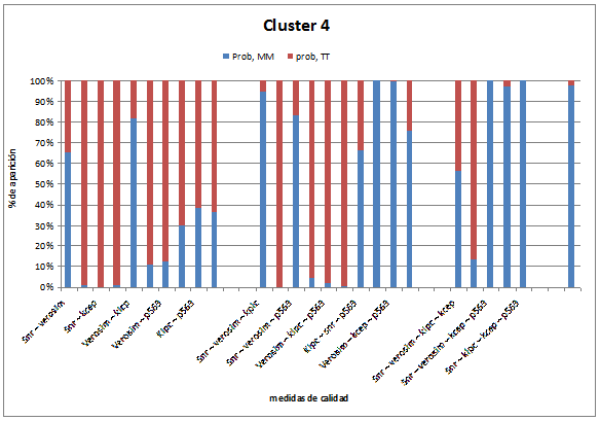
Cuadro 23: Origen del fichero por cluster para cuatro agrupaciones

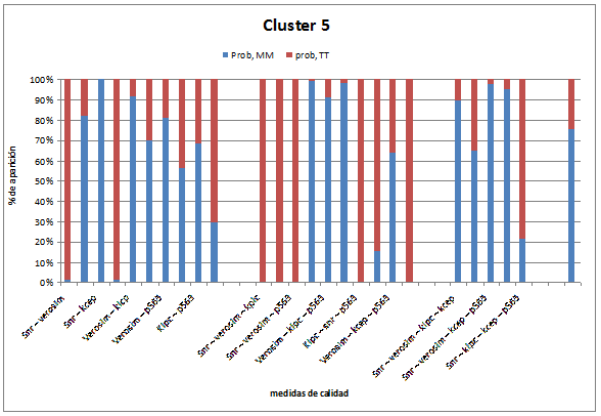
9 agrupamientos

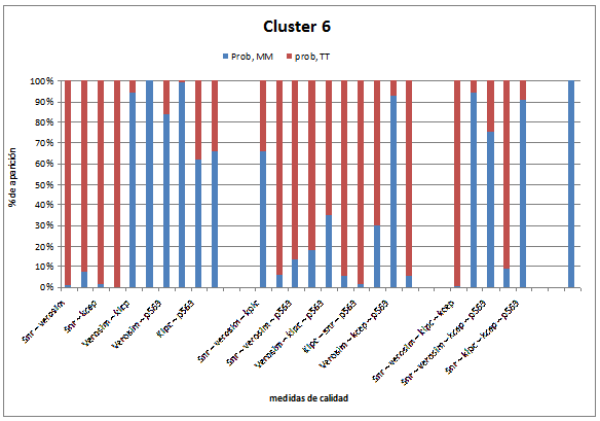


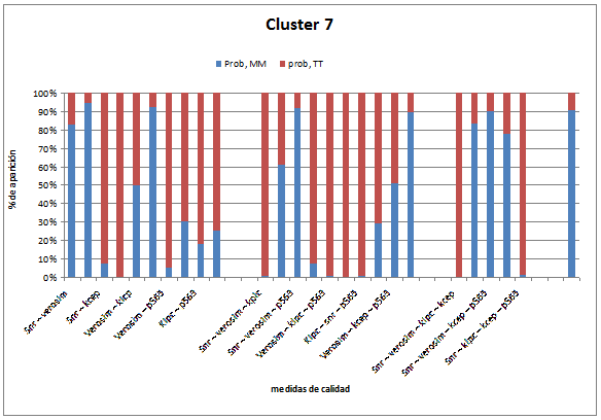


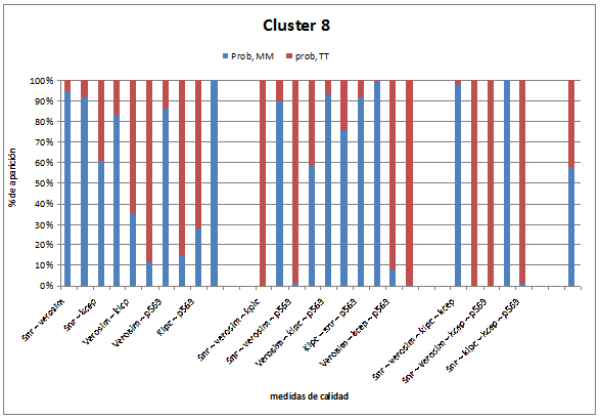


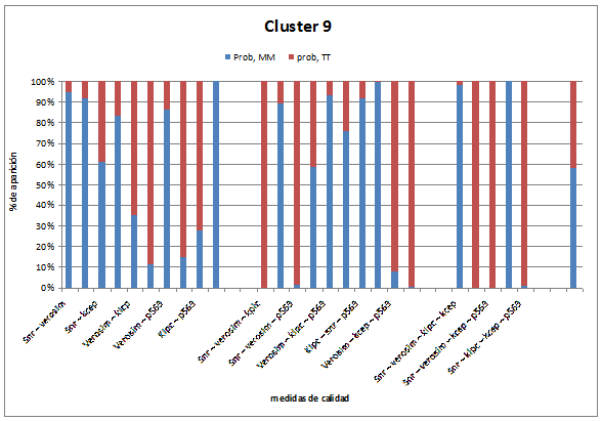












Cuadro 24: Origen del fichero por cluster para nueve agrupaciones

En la tabla 25 se muestra una comparativa para todas las combinaciones de indicadores de degradación y para distintos clusters de los valores de entropía obtenidos empleando la distancia Cityblock.

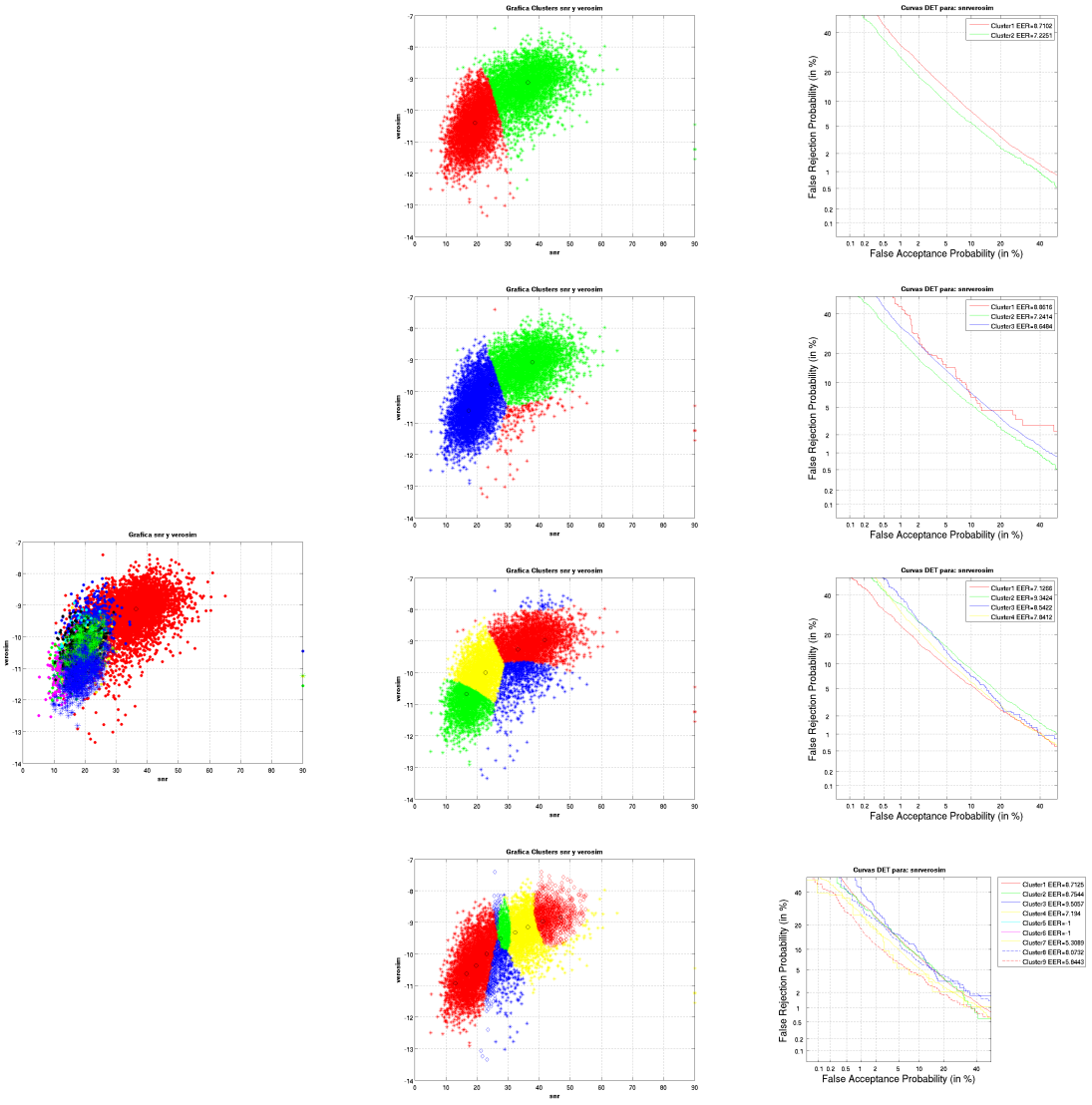
COMPARATIVA Kmeans - Distancia Cityblock				
	2 clusters	3 clusters	4 clusters	9 Clusters
snr-verosim	0,47873	0,47352	0,43334	0,40254
snr-klpc	0,48107	0,47795	0,43182	0,40886
snr-kcep	0,47659	0,46905	0,42709	0,39119
snr-p563	0,48455	0,47637	0,43647	0,40608
verosim-klpc	0,56512	0,63813	0,61421	0,5281
verosim-kcep	0,70932	0,71816	0,57341	0,51054
verosim-p563	0,65427	0,60532	0,62026	0,57523
klpc-kcep	0,78153	0,77474	0,74726	0,71446
klpc-p563	0,94242	0,93112	0,92074	0,84425
kcep-p563	0,78628	0,77388	0,75848	0,73029
snr-verosim-klpc	0,47688	0,46999	0,43248	0,39766
snr-verosim-kcep	0,46425	0,46117	0,41927	0,37949
snr-verosim-p563	0,47953	0,47045	0,43136	0,39814
verosim-klpc-kcep	0,70486	0,66999	0,67872	0,51916
verosim-klpc-p563	0,60528	0,63722	0,6019	0,54057
klpc-snr-kcep	0,47562	0,46408	0,42704	0,38489
klpc-snr-p563	0,4852	0,47427	0,43015	0,41106
klpc-kcep-p563	0,77343	0,76829	0,74307	0,7055
verosim-kcep-p563	0,70569	0,7124	0,60777	0,53155
snr-kcep-p563	0,47371	0,4641	0,42725	0,38339
snr-verosim-klpc-kcep	0,46415	0,45877	0,41685	0,38778
snr-verosim-klpc-p563	0,47501	0,46527	0,42918	0,39849
snr-verosim-kcep-p563	0,46898	0,45823	0,4171	0,37352
verosim-klpc-kcep-p563	0,70061	0,67992	0,69029	0,54442
snr-klpc-kcep-p563	0,47707	0,46272	0,4251	0,379
snr-verosim-klpc-kcep-p563	0,46267	0,4581	0,41661	0,376

Cuadro 25: Entropías obtenidas por K-means (distancia Cityblock) en NIST 2008

.0.2. NIST 2008 - GMM

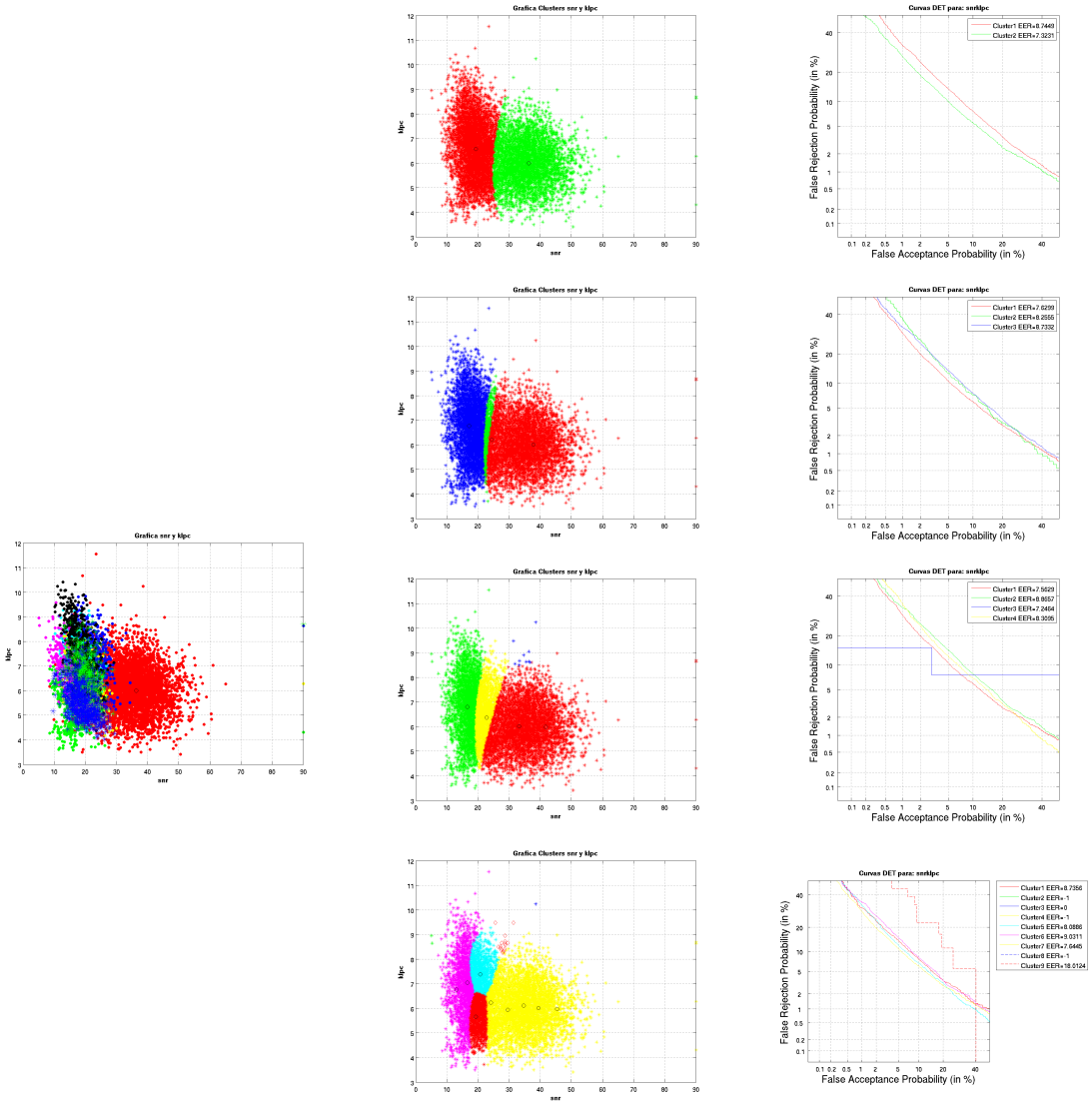
En los cuadros del X al X se muestran las agrupaciones GMM generadas para la base de datos NIST 2008. El tipo de indicador de degradación empleado se muestra en la cabecera de la sección de la tabla y, a continuación, se muestran las diferentes figuras asociadas. En la primera columna se presenta la gráfica inicial, sin haberse realizado sobre ella ninguna agrupación. La segunda columna muestra la gráfica ya agrupada en función del número de clusters y, por último, la tercera columna muestra la curva DET correspondiente con ese experimento.

2 indicadores de degradación: Snr verosim



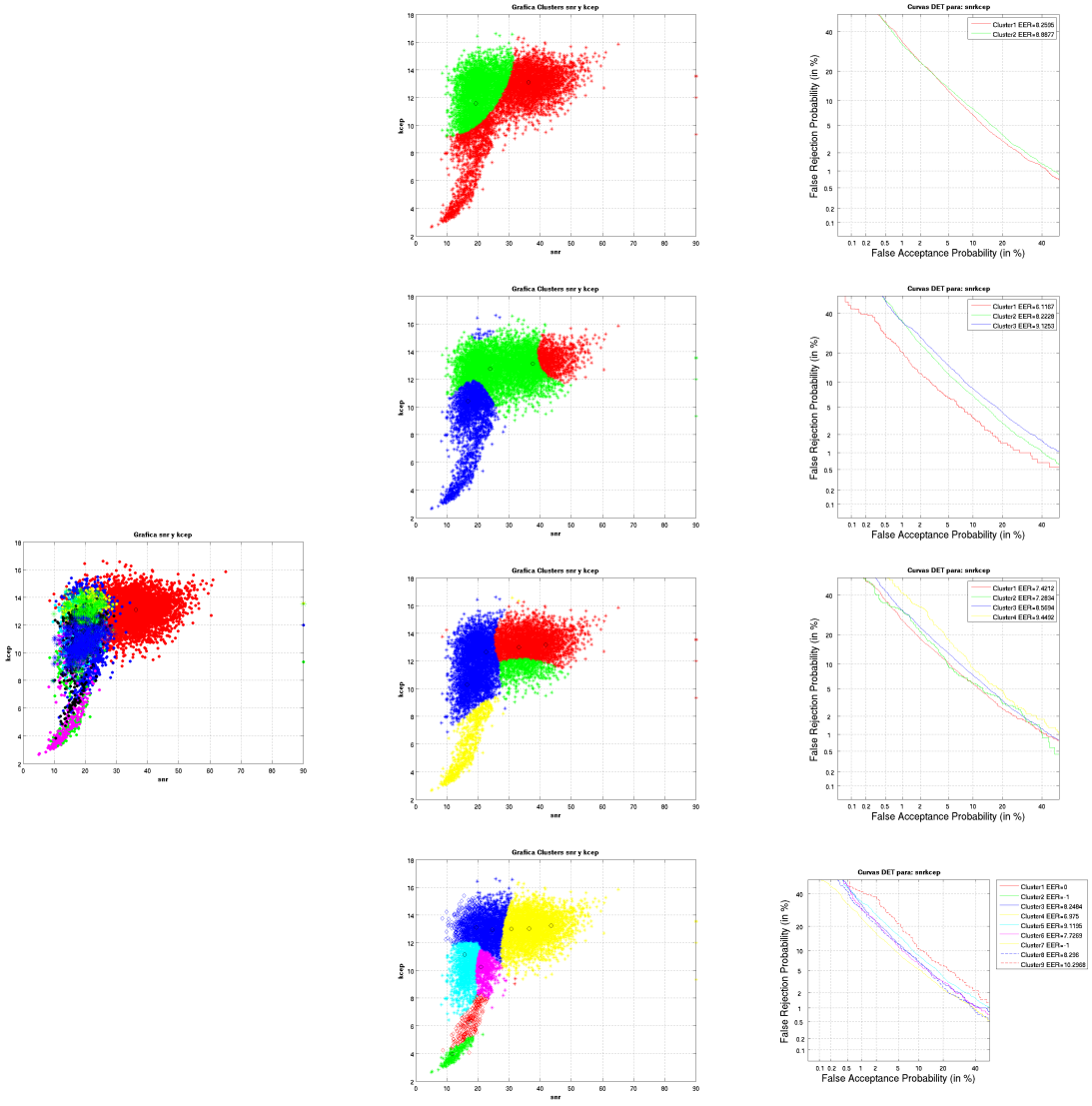
Cuadro 26: Ficheros agrupados y curvas DET para snr verosim con GMM

2 indicadores de degradación: Snr klpc



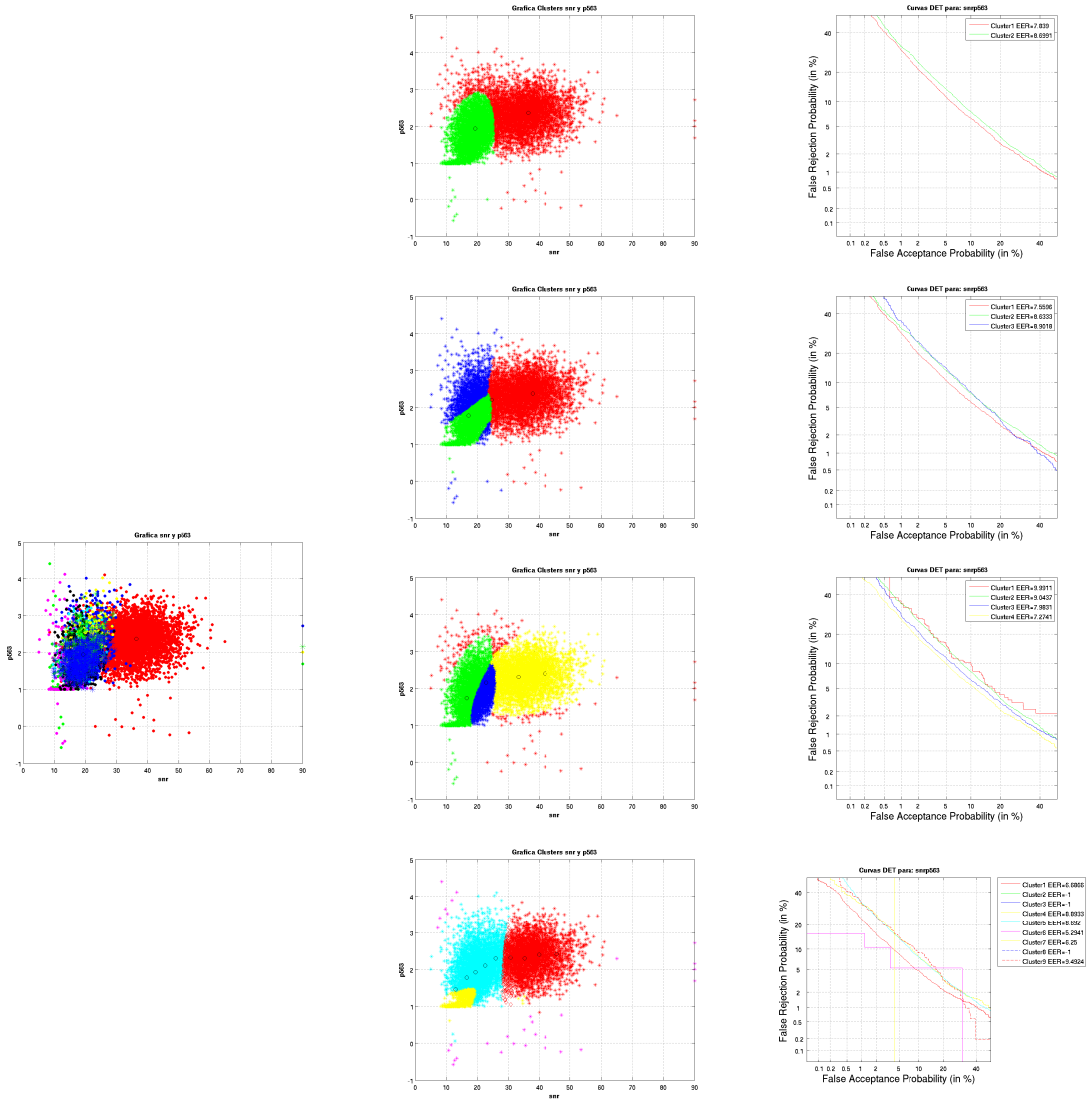
Cuadro 27: Ficheros agrupados y curvas DET para snr klpc con GMM

2 indicadores de degradación: Snr kcep



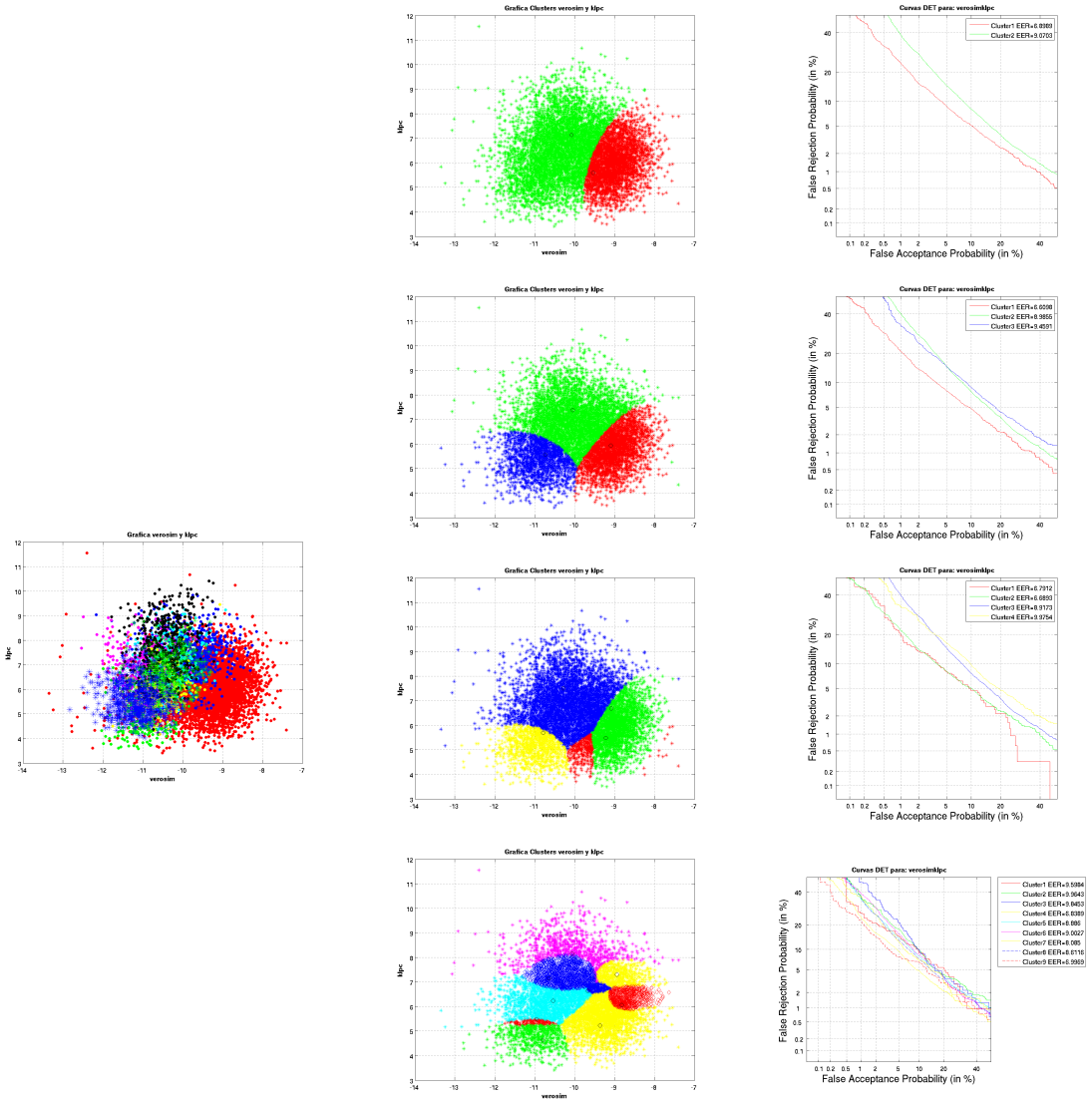
Cuadro 28: Ficheros agrupados y curvas DET para snr kcep con GMM

2 indicadores de degradación: Snr P.563



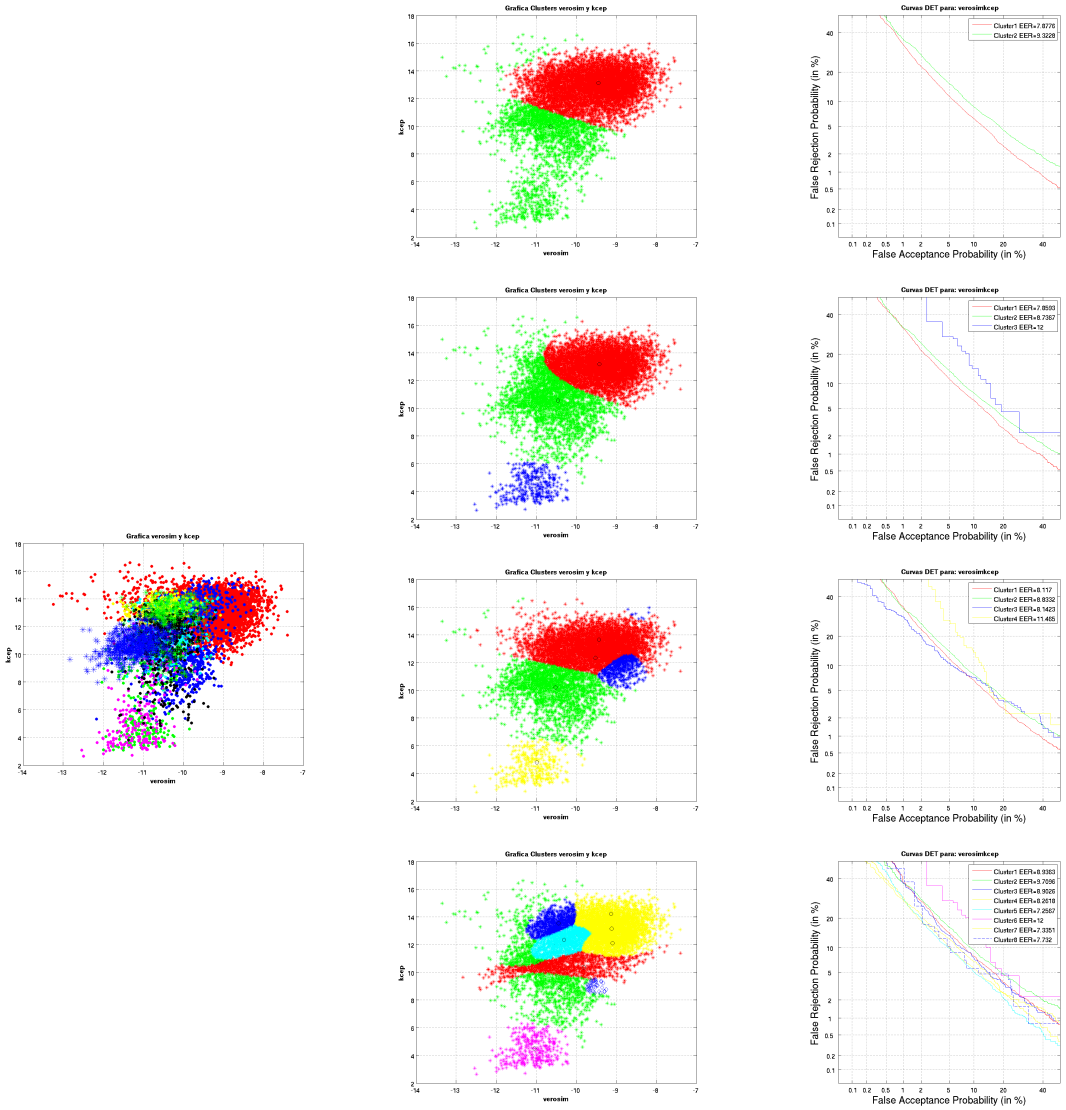
Cuadro 29: Ficheros agrupados y curvas DET para snr P.563 con GMM

2 indicadores de degradación: verosim klpc



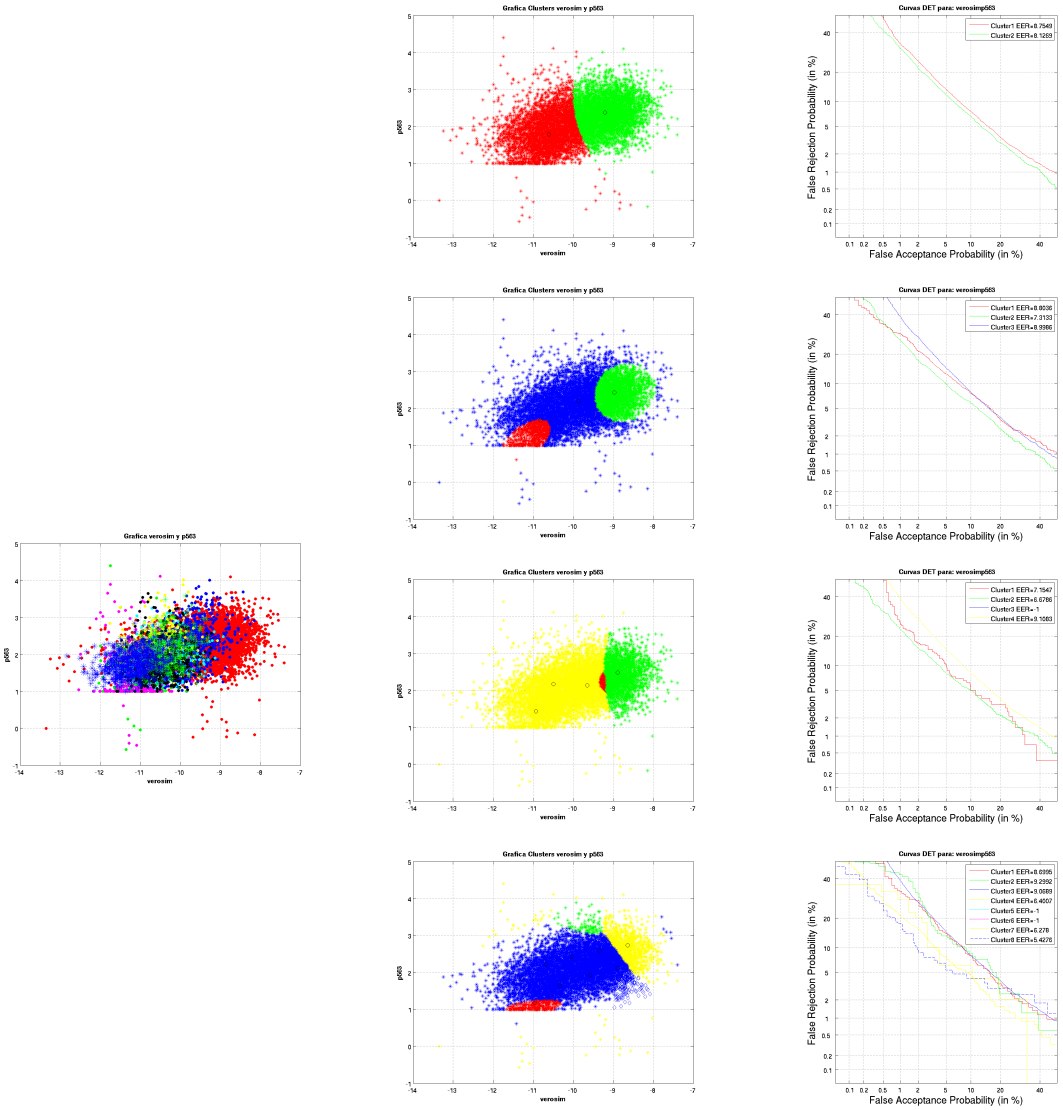
Cuadro 30: Ficheros agrupados y curvas DET para verosim klpc con GMM

2 indicadores de degradación: verosim kcep



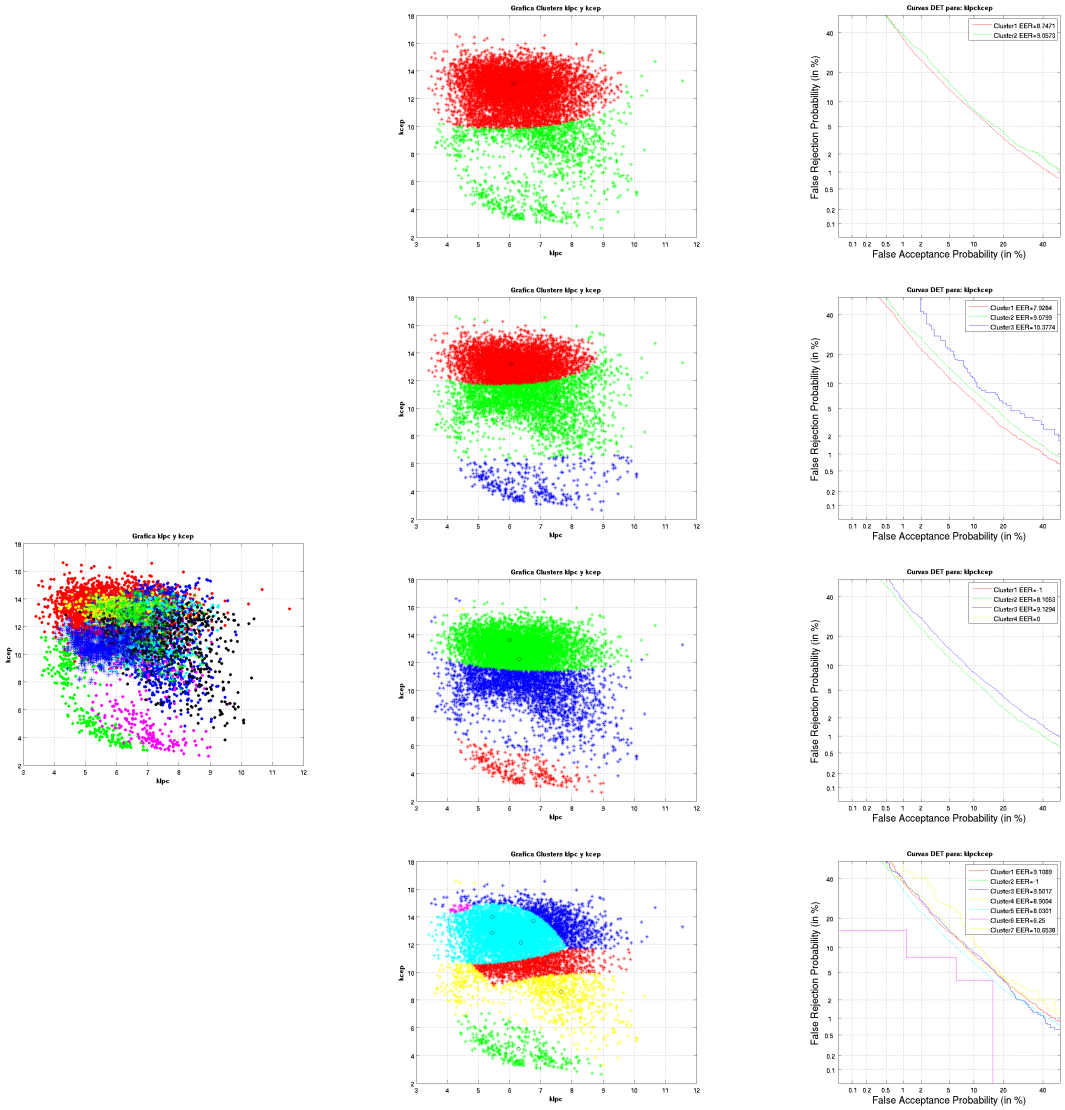
Cuadro 31: Ficheros agrupados y curvas DET para verosim kcep con GMM

2 indicadores de degradación: verosim P.563



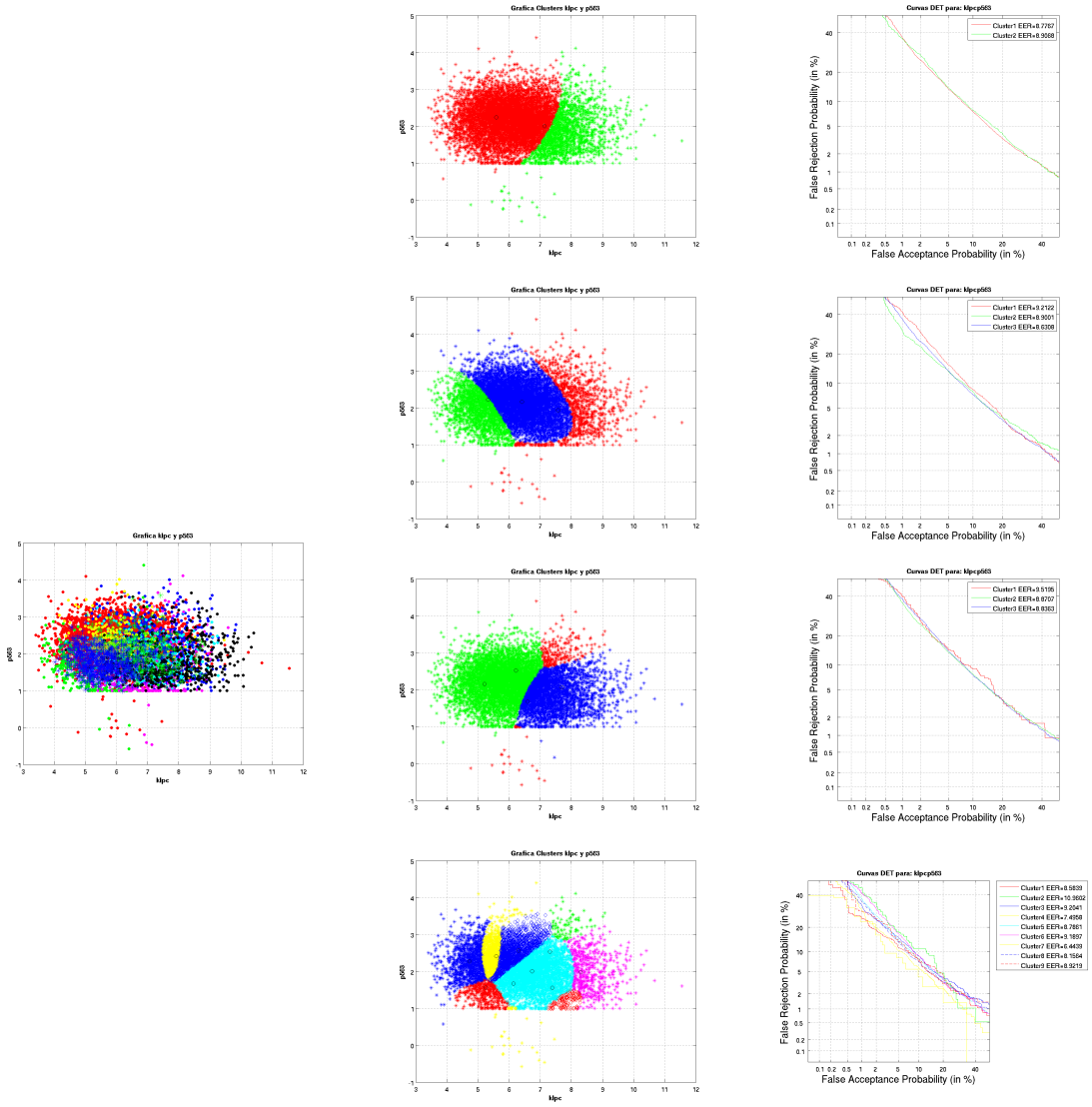
Cuadro 32: Ficheros agrupados y curvas DET para verosim P.563 con GMM

2 indicadores de degradación: klpc kcep



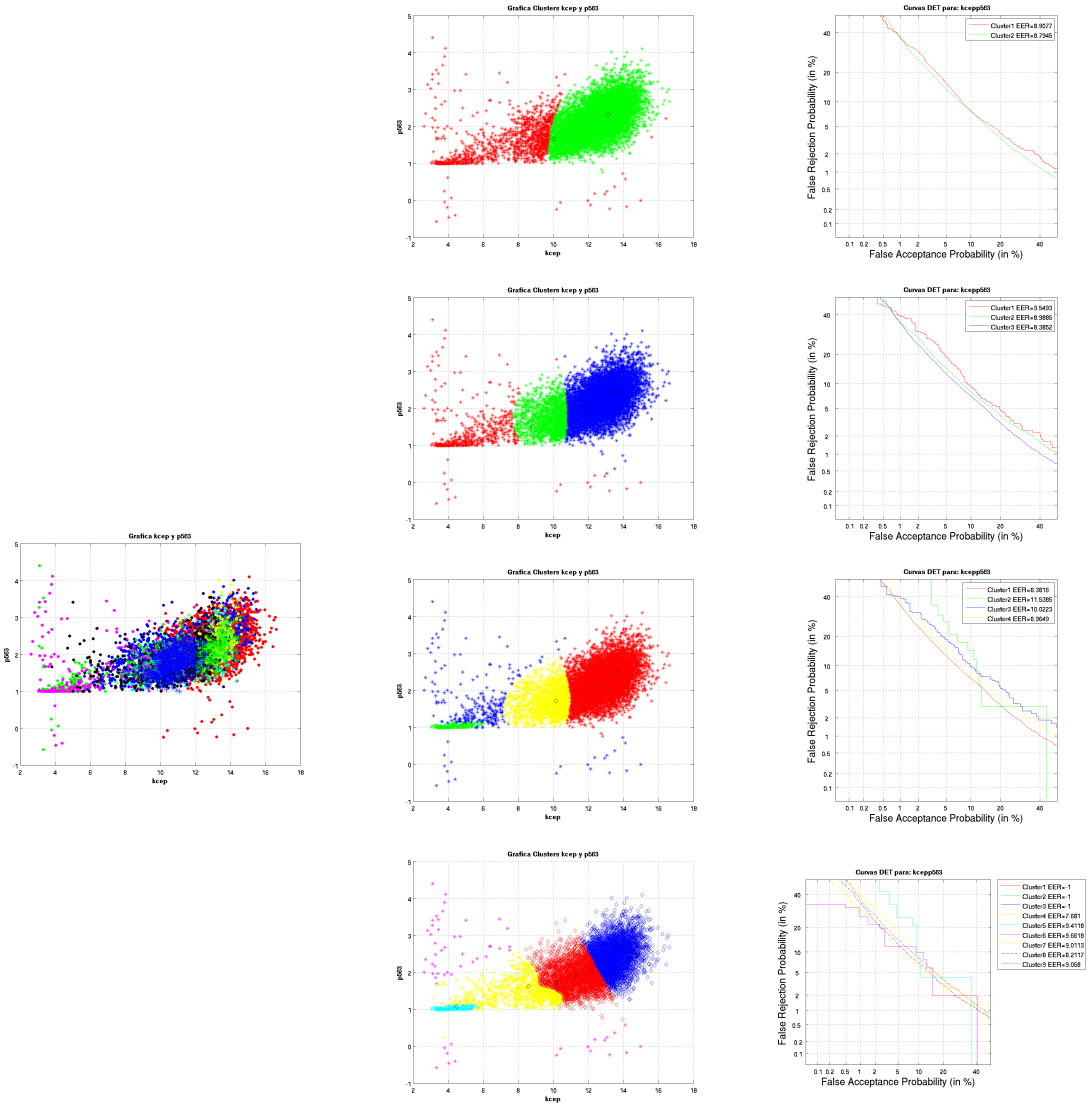
Cuadro 33: Ficheros agrupados y curvas DET para klpc kcep con GMM

2 indicadores de degradación: klpc P.563



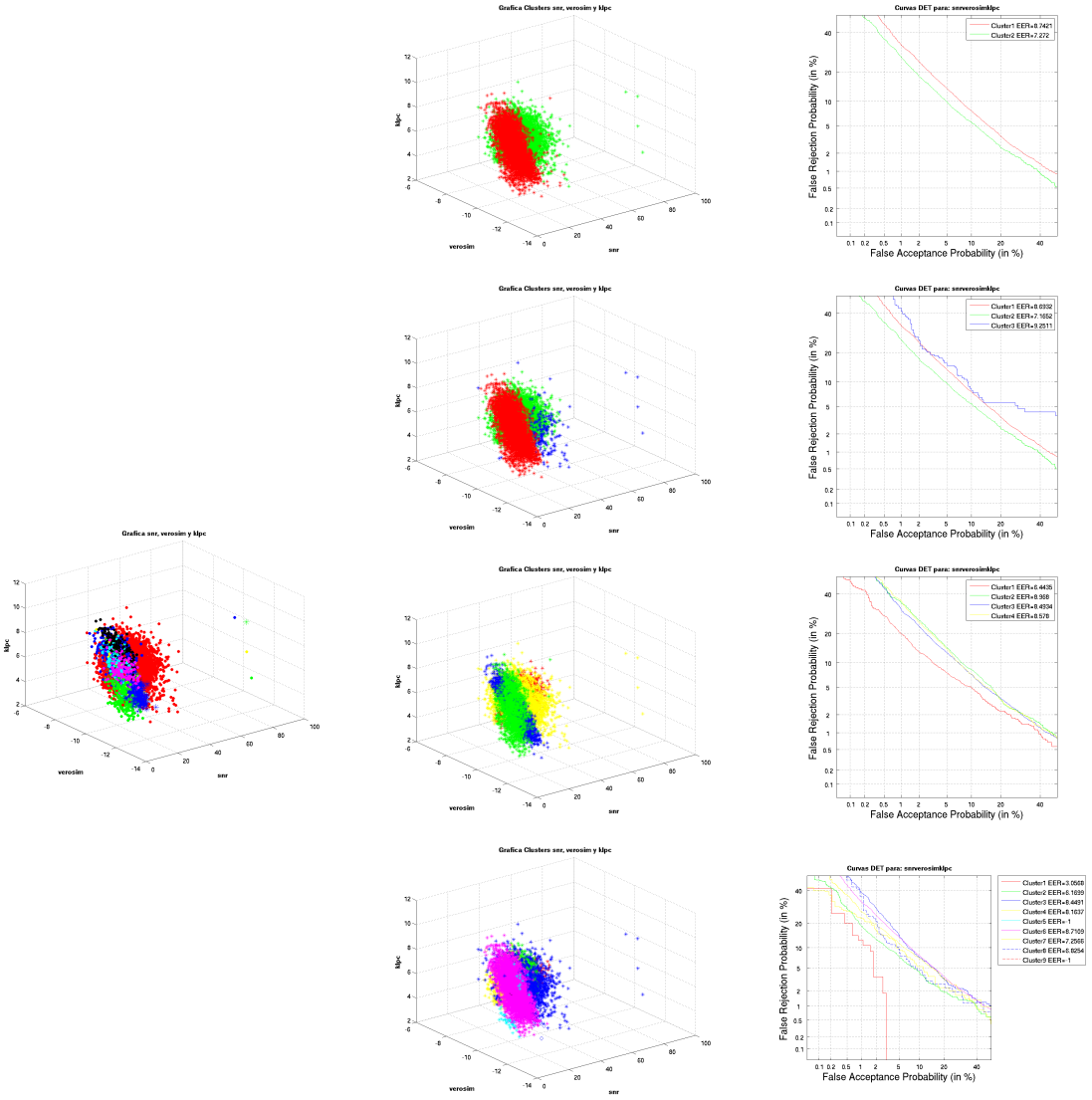
Cuadro 34: Ficheros agrupados y curvas DET para klpc P.563 con GMM

2 indicadores de degradación: kcep P.563



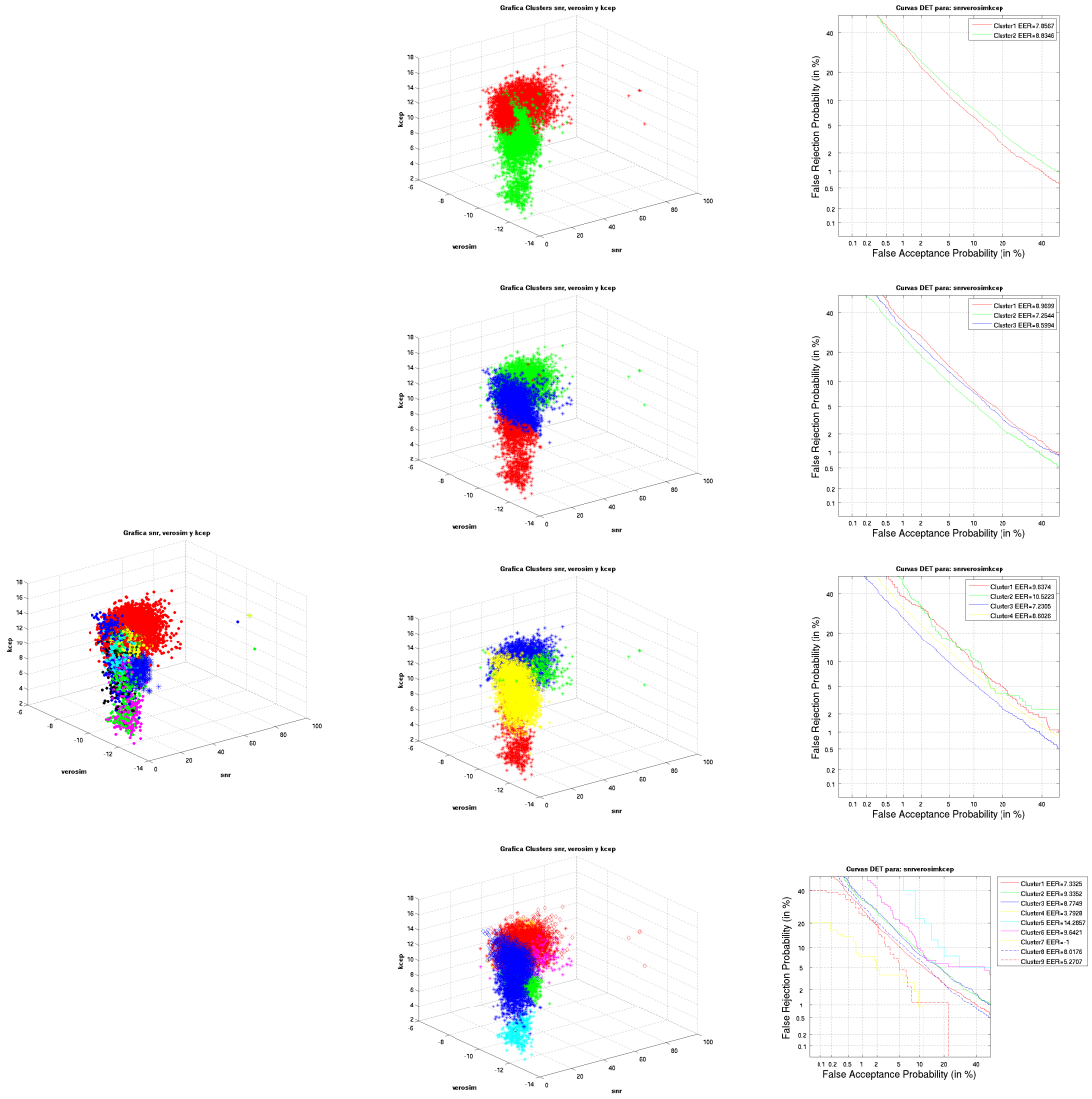
Cuadro 35: Ficheros agrupados y curvas DET para kcep P.563 con GMM

3 indicadores de degradación: snr verosim klpc



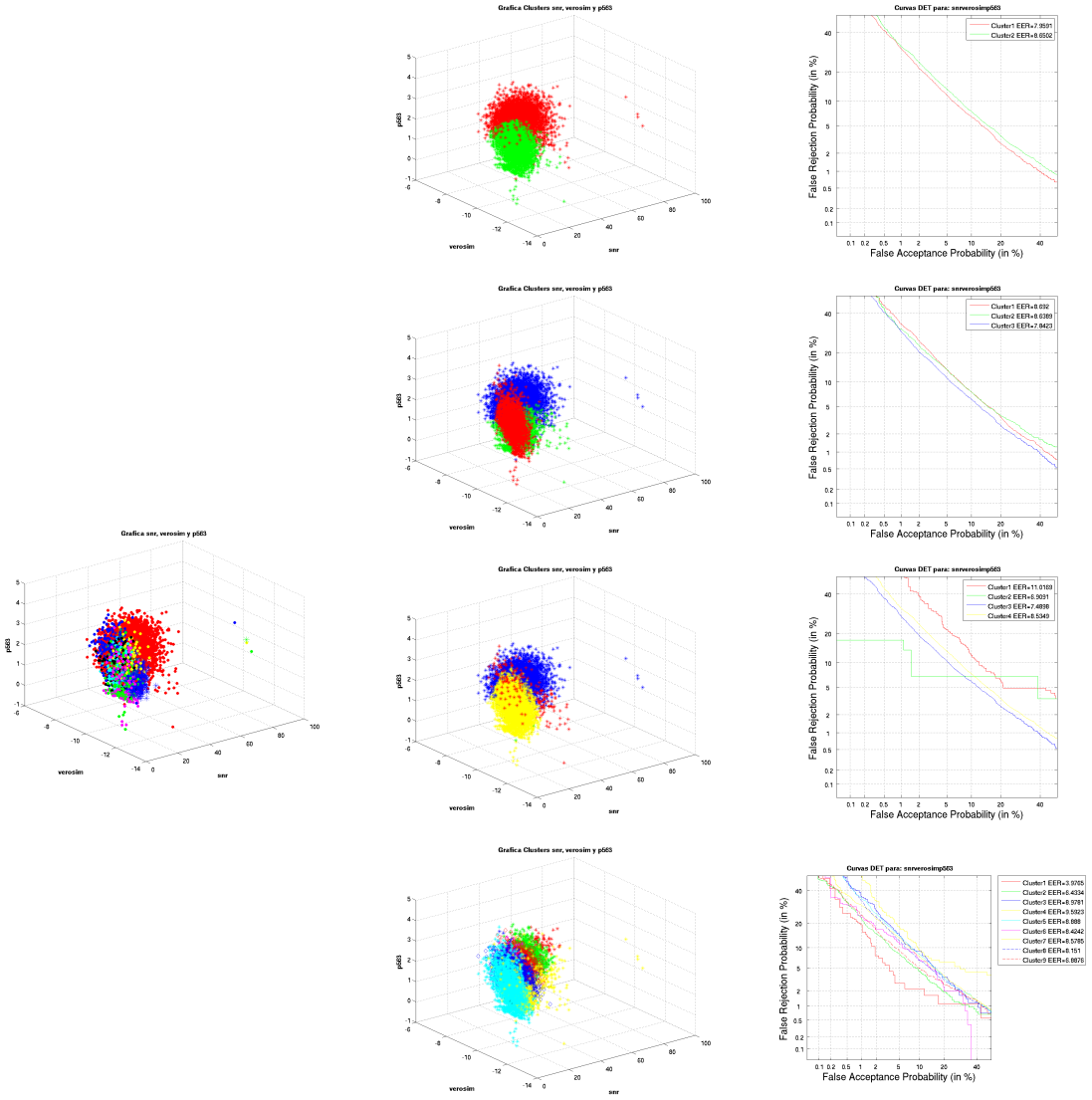
Cuadro 36: Ficheros agrupados y curvas DET para snr verosim klpc con GMM

3 indicadores de degradación: snr verosim kcep



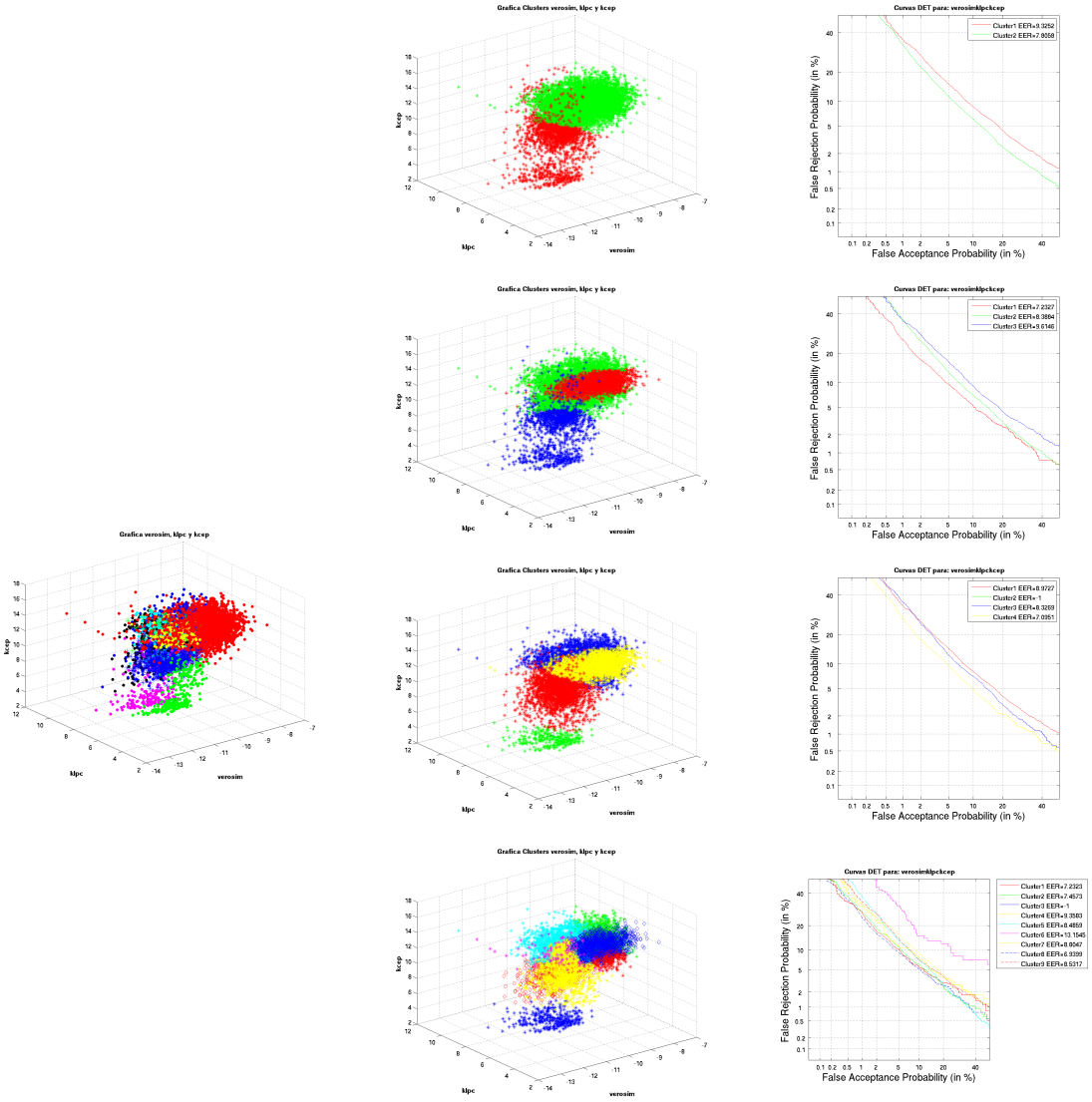
Cuadro 37: Ficheros agrupados y curvas DET para snr verosim kcep con GMM

3 indicadores de degradación: snr verosim P.563



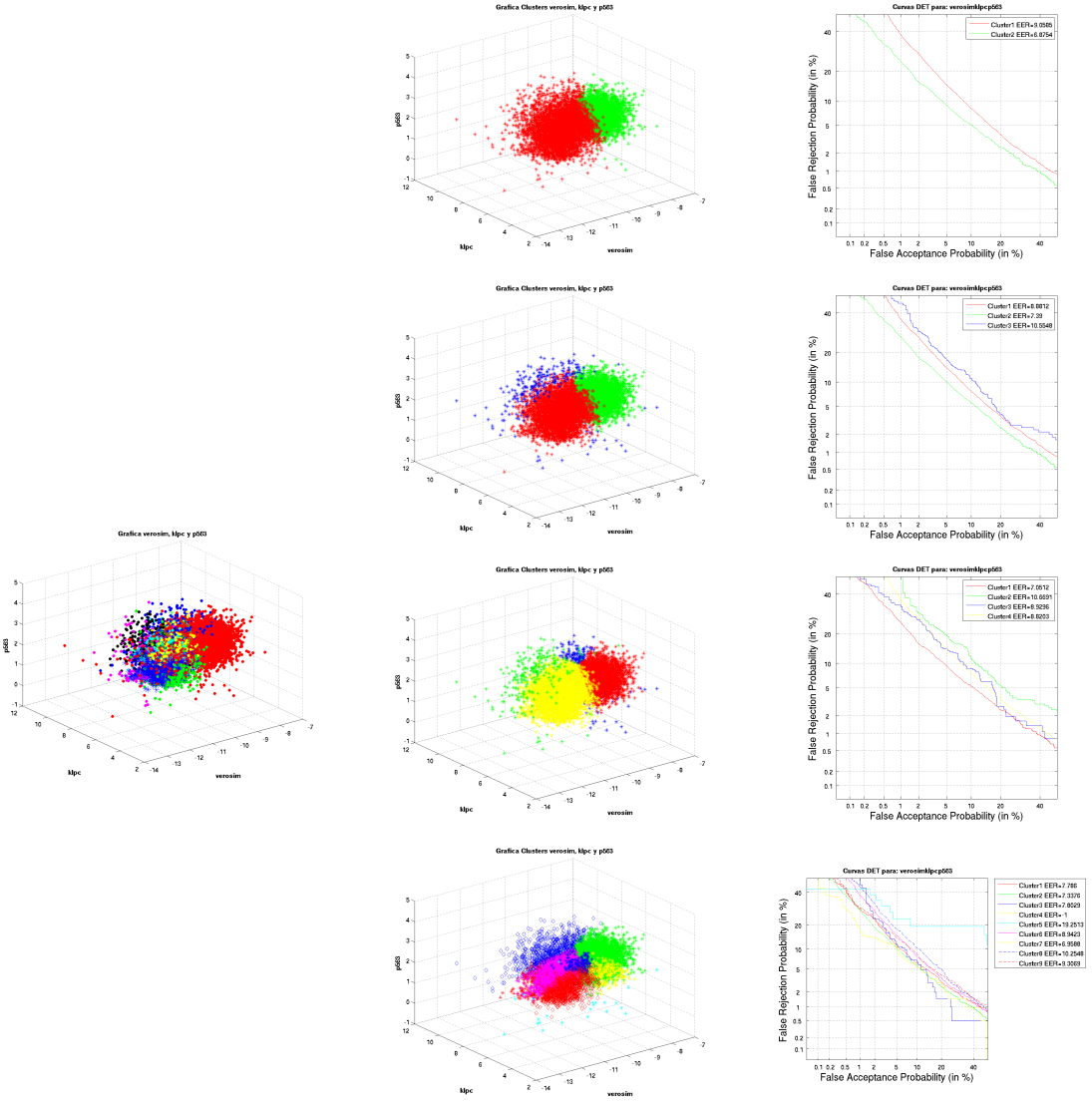
Cuadro 38: Ficheros agrupados y curvas DET para snr verosim P.563 con GMM

3 indicadores de degradación: verosim klpc kcep



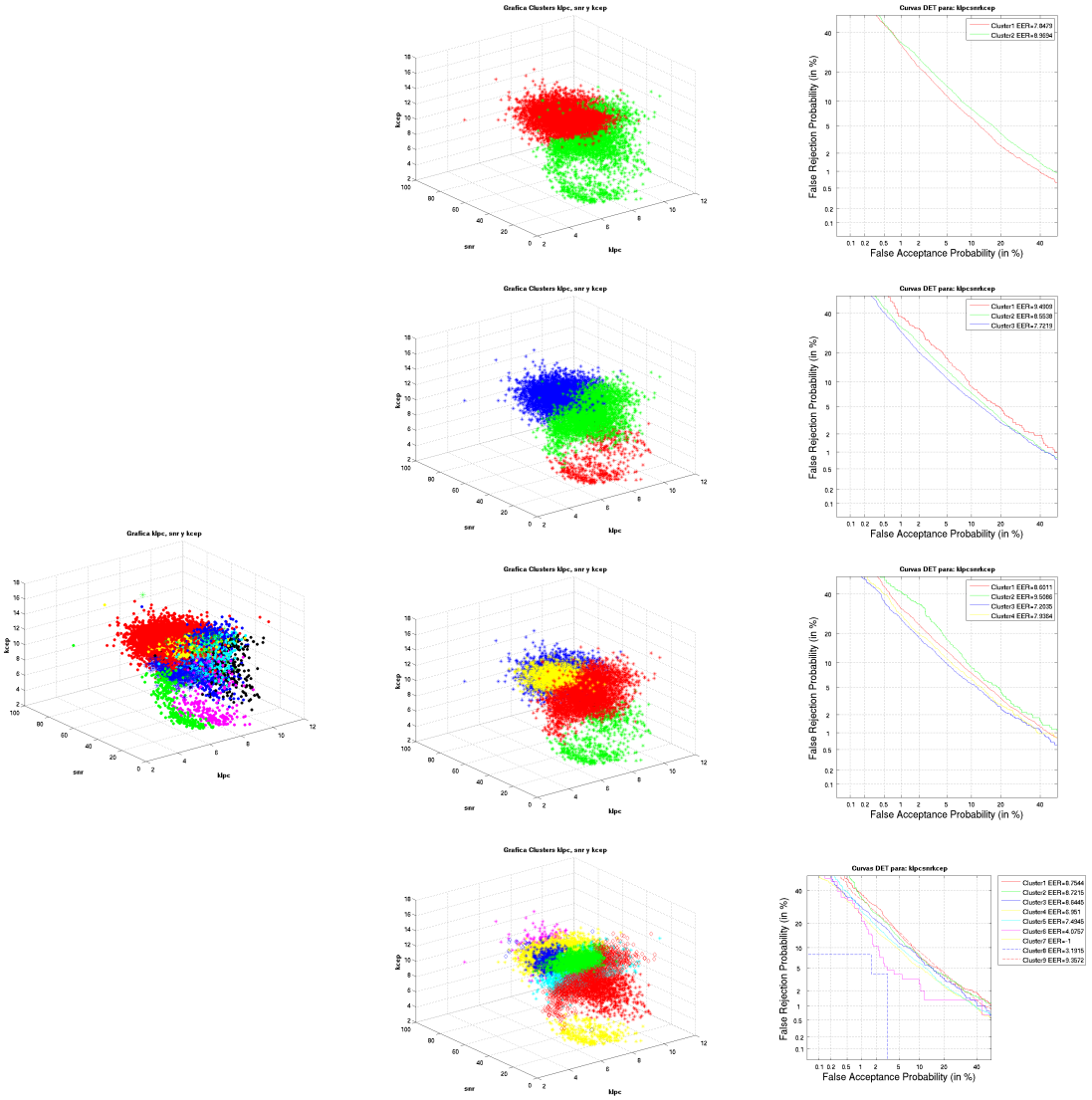
Cuadro 39: Ficheros agrupados y curvas DET para verosim klpc kcep con GMM

3 indicadores de degradación: verosim klpc P.563



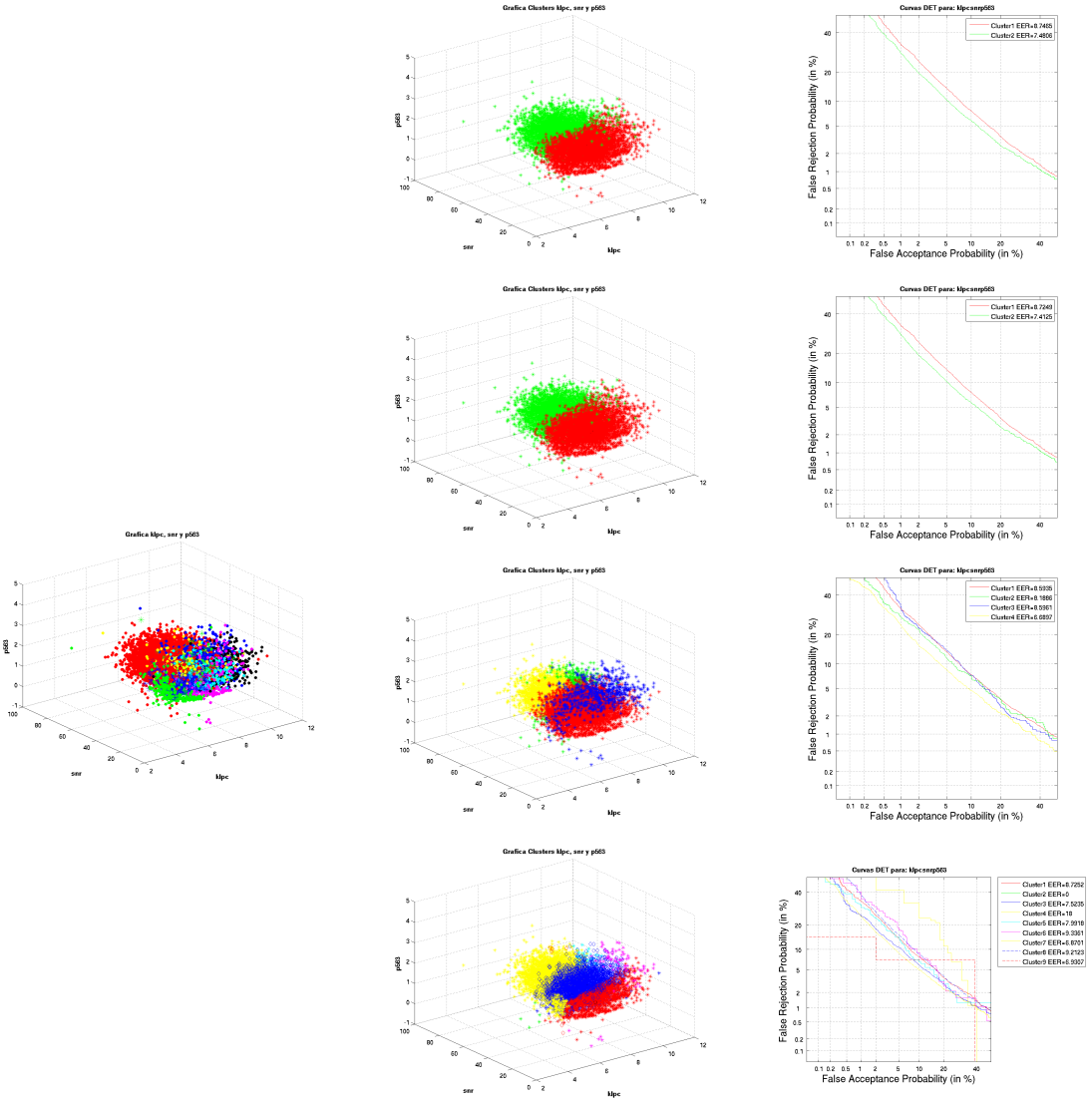
Cuadro 40: Ficheros agrupados y curvas DET para verosim klpc P.563 con GMM/

3 indicadores de degradación: klpc snr kcep



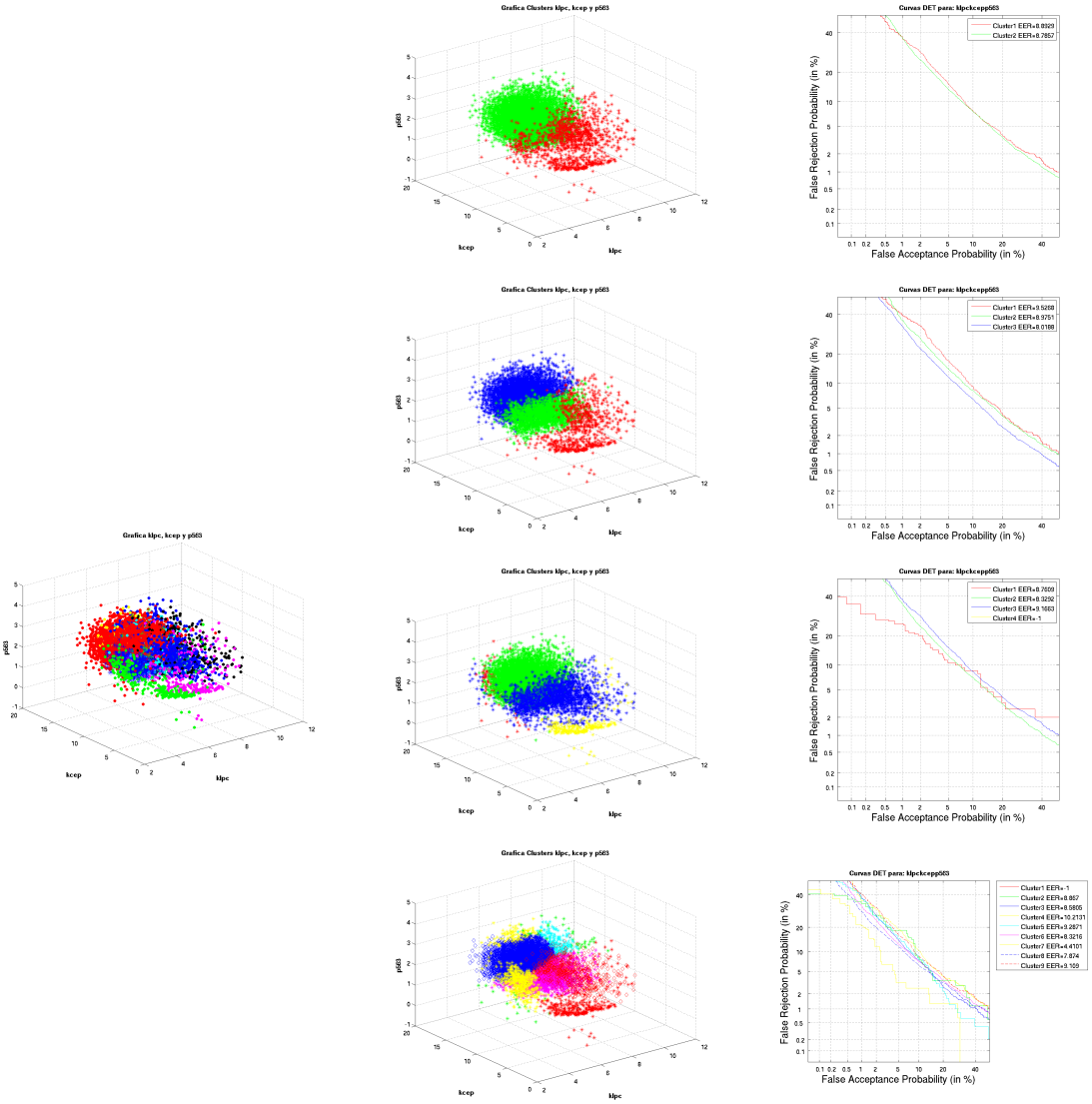
Cuadro 41: Ficheros agrupados y curvas DET para klpc snr kcep con GMM

3 indicadores de degradación: klpc snr P.563



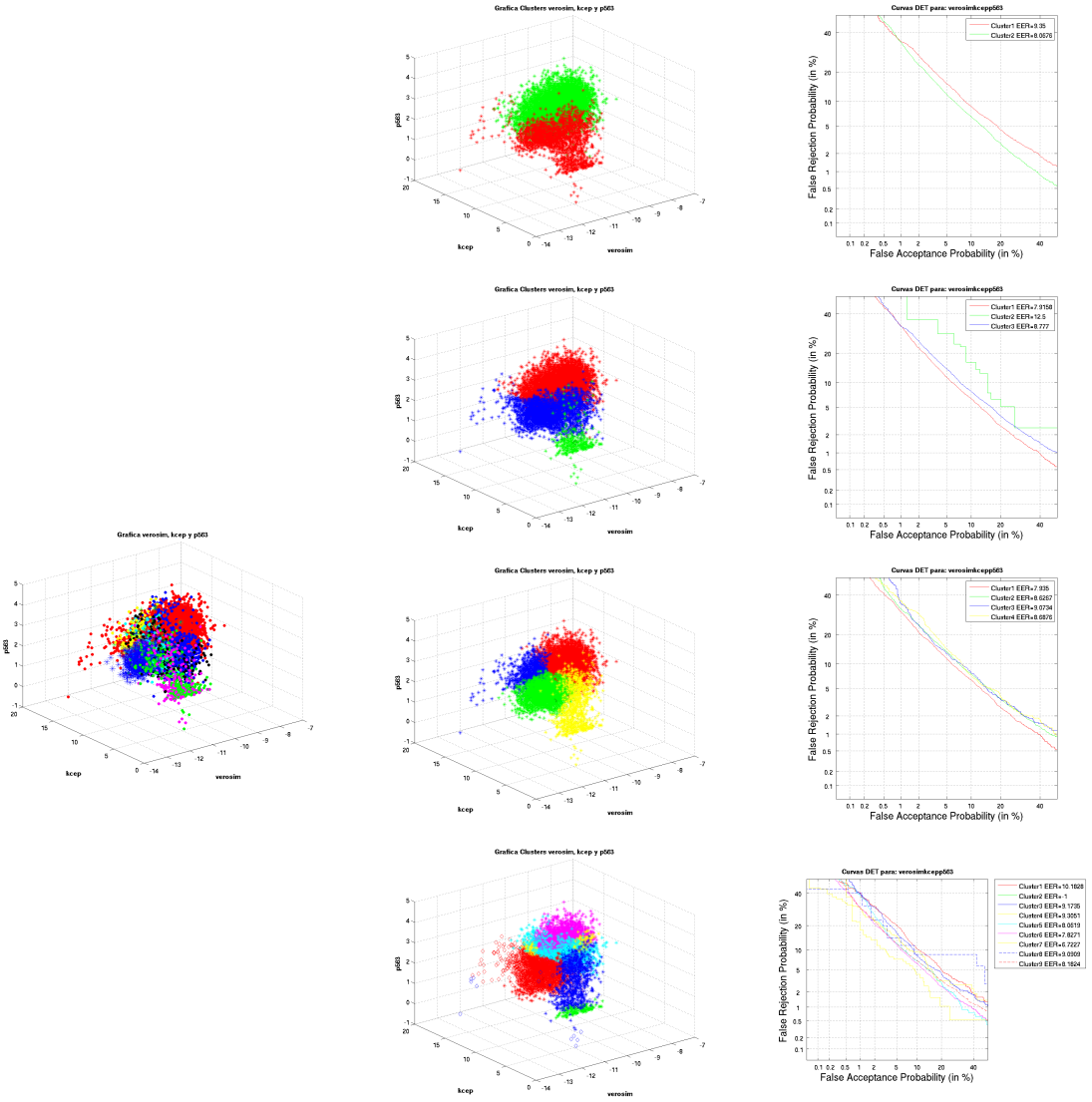
Cuadro 42: Ficheros agrupados y curvas DET para klpc snr P.563 con GMM

3 indicadores de degradación: klpc kcep P.563



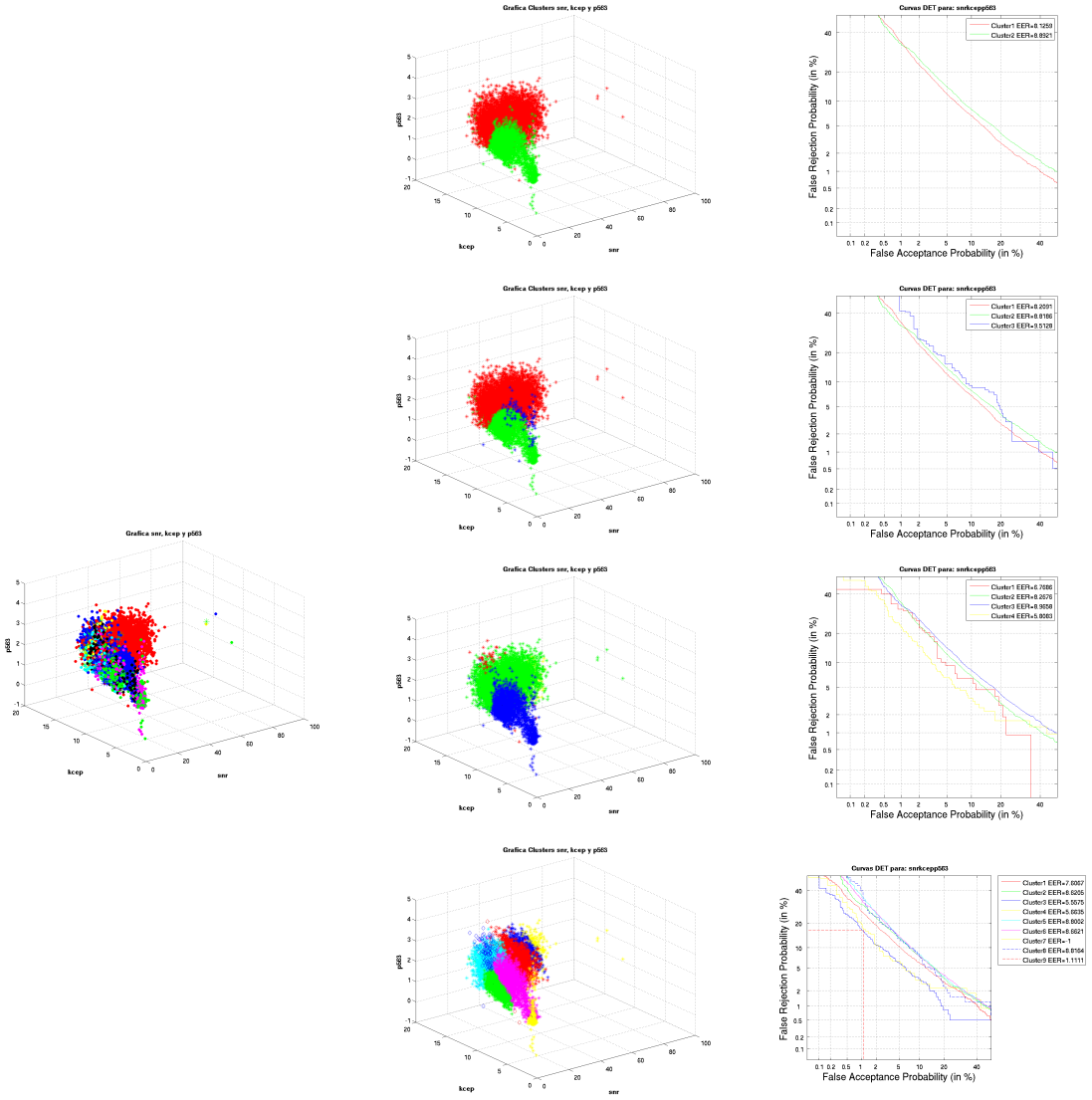
Cuadro 43: Ficheros agrupados y curvas DET para klpc kcep P.563 con GMM

3 indicadores de degradación: verosim kcep P.563



Cuadro 44: Ficheros agrupados y curvas DET para verosim kcep P.563 con GMM

3 indicadores de degradación: snr kcep P.563



Cuadro 45: Ficheros agrupados y curvas DET para snr kcep P.563 con GMM

COMPARATIVA GMM				
	2 Clusters	3 Clusters	4 Clusters	9 Clusters
snrverosim	0,59734	0,49568	0,51816	0,43637
snrklpc	0,84598	0,52571	0,61323	0,41741
snrkcep	0,54935	0,46588	0,46013	0,41605
snrp563	0,53818	0,63840	0,53104	0,41616
verosimklpc	0,82314	0,99836	0,63972	0,52972
verosimkcep	0,72584	0,67738	0,85800	0,62826
verosimp563	0,66283	0,68236	0,67416	0,58272
klpckcep	0,81420	0,92860	0,87653	0,85931
klpcp563	0,99319	0,90477	0,96623	0,87377
kcepp563	0,80201	0,85175	0,93339	0,76759
snrverosimklpc	0,44592	0,44420	0,53350	0,39742
snrverosimkcep	0,77354	0,78864	0,40454	0,41785
snrverosimp563	0,44847	0,44975	0,42406	0,47345
verosimklpckcep	0,93090	0,79350	0,82937	0,53085
verosimklpcp563	0,61721	0,98630	0,69140	0,59280
klpcsnrkcep	0,64663	0,65158	0,45672	0,43785
klpcsnrp563	0,60075	0,44775	0,43179	0,40391
klpckcepp563	0,93014	0,83079	0,92575	0,71970
verosimkcepp563	0,75188	0,59463	0,70972	0,67683
snrkcepp563	0,72193	0,51561	0,86199	0,39872
snrverosimklpckcep	0,43830	0,73285	0,43871	0,41331
snrverosimklpcp563	0,85184	0,41015	0,50795	0,40567
snrverosimkcepp563	0,47971	0,54209	0,40793	0,40240
verosimklpckcepp563	0,91837	0,90985	0,68944	0,56929
snrklpckcepp563	0,47695	0,58030	0,52386	0,39178
snrverosimklpckcepp563	0,52791	0,84004	0,48414	0,42765

Comparativa Kmeans - GMM - dos Clusters		
	Kmeans	GMM
Snr - verosim	0,4983	0,59734
Snr - klpc	0,5004	0,84598
Snr - kcep	0,4956	0,54935
Snr - p563	0,4988	0,53818
Verosim - klcp	0,9506	0,82314
Verosim - kcep	0,7725	0,72584
Verosim - p563	0,6519	0,66283
Klpc - kcep	0,8196	0,81420
Klpc - p563	0,95	0,99319
Kcep - p563	0,8007	0,80201
Snr - verosim - kplc	0,4991	0,44592
Snr - verosim - kcep	0,4946	0,77354
Snr - verosim - p563	0,4988	0,44847
Verosim - klpc - kcep	0,7716	0,93090
Verosim - klpc - p563	0,9427	0,61721
Klpc - snr - kcep	0,4941	0,64663
Klpc - snr - p563	0,5004	0,60075
Klpc - kcep - p563	0,8	0,93014
Verosim - kcep - p563	0,7702	0,75188
Snr - kcep - p563	0,4946	0,72193
Snr - verosim - klpc - kcep	0,4934	0,43830
Snr - verosim - klpc - p563	0,4991	0,85184
Snr - verosim - kcep - p563	0,4931	0,47971
Verosim - klpc - kcep - p563	0,767	0,91837
Snr - klpc - kcep - p563	0,4944	0,47695
Snr - verosim - klcp - kcep - p563	0,4934	0,52791

Cuadro 46: Comparativa K-means- GMM para dos agrupamientos

.0.3. Comparativa K-means - GMM para NIST 2008

Comparativa Kmeans - GMM - tres Clusters		
	Kmeans	GMM
Snr - verosim	0,4894	0,49568
Snr - klpc	0,489	0,52571
Snr - kcep	0,4827	0,46588
Snr - p563	0,4899	0,63840
Verosim - klcp	0,6329	0,99836
Verosim - kcep	0,7454	0,67738
Verosim - p563	0,6174	0,68236
Klpc - kcep	0,7741	0,92860
Klpc - p563	0,9376	0,90477
Kcep - p563	0,7816	0,85175
Snr - verosim - kplc	0,4879	0,44420
Snr - verosim - kcep	0,4813	0,78864
Snr - verosim - p563	0,4893	0,44975
Verosim - klpc - kcep	0,7363	0,79350
Verosim - klpc - p563	0,6282	0,98630
Klpc - snr - kcep	0,4795	0,65158
Klpc - snr - p563	0,4881	0,44775
Klpc - kcep - p563	0,7753	0,83079
Verosim - kcep - p563	0,7412	0,59463
Snr - kcep - p563	0,4827	0,51561
Snr - verosim - klpc - kcep	0,4795	0,73285
Snr - verosim - klpc - p563	0,488	0,41015
Snr - verosim - kcep - p563	0,4812	0,54209
Verosim - klpc - kcep - p563	0,7338	0,90985
Snr - klpc - kcep - p563	0,4813	0,58030
Snr - verosim - klcp - kcep - p563	0,4803	0,84004

Cuadro 47: Comparativa K-means- GMM para tres agrupamientos

Comparativa Kmeans - GMM - cuatro Clusters		
	Kmeans	GMM
Snr - verosim	0,4443	0,51816
Snr - klpc	0,4455	0,61323
Snr - kcep	0,436	0,46013
Snr - p563	0,4453	0,53104
Verosim - klcp	0,5566	0,63972
Verosim - kcep	0,7285	0,85800
Verosim - p563	0,6301	0,67416
Klpc - kcep	0,7489	0,87653
Klpc - p563	0,8919	0,96623
Kcep - p563	0,7609	0,93339
Snr - verosim - kplc	0,4427	0,53350
Snr - verosim - kcep	0,4363	0,40454
Snr - verosim - p563	0,4443	0,42406
Verosim - klpc - kcep	0,7544	0,82937
Verosim - klpc - p563	0,5598	0,69140
Klpc - snr - kcep	0,4361	0,45672
Klpc - snr - p563	0,4436	0,43179
Klpc - kcep - p563	0,7509	0,92575
Verosim - kcep - p563	0,7248	0,70972
Snr - kcep - p563	0,4364	0,86199
Snr - verosim - klpc - kcep	0,4328	0,43871
Snr - verosim - klpc - p563	0,4427	0,50795
Snr - verosim - kcep - p563	0,434	0,40793
Verosim - klpc - kcep - p563	0,706	0,68944
Snr - klpc - kcep - p563	0,4348	0,52386
Snr - verosim - klcp - kcep - p563	0,432	0,48414

Cuadro 48: Comparativa K-means- GMM para cuatro agrupamientos

Comparativa Kmeans - GMM - nueve Clusters		
	Kmeans	GMMs
Snr - verosim	0,4127	0,43637
Snr - klpc	0,4106	0,41741
Snr - kcep	0,3968	0,41605
Snr - p563	0,4133	0,41616
Verosim - klcp	0,5269	0,52972
Verosim - kcep	0,5179	0,62826
Verosim - p563	0,5754	0,58272
Klpc - kcep	0,7166	0,85931
Klpc - p563	0,8539	0,87377
Kcep - p563	0,7348	0,76759
Snr - verosim - kplc	0,4131	0,39742
Snr - verosim - kcep	0,3864	0,41785
Snr - verosim - p563	0,4117	0,47345
Verosim - klpc - kcep	0,5362	0,53085
Verosim - klpc - p563	0,5355	0,59280
Klpc - snr - kcep	0,3859	0,43785
Klpc - snr - p563	0,4143	0,40391
Klpc - kcep - p563	0,7127	0,71970
Verosim - kcep - p563	0,5227	0,67683
Snr - kcep - p563	0,3868	0,39872
Snr - verosim - klpc - kcep	0,3827	0,41331
Snr - verosim - klpc - p563	0,4113	0,40567
Snr - verosim - kcep - p563	0,3835	0,40240
Verosim - klpc - kcep - p563	0,5161	0,56929
Snr - klpc - kcep - p563	0,3927	0,39178
Snr - verosim - klcp - kcep - p563	0,3839	0,42765

Cuadro 49: Comparativa K-means- GMM para nueve agrupamientos

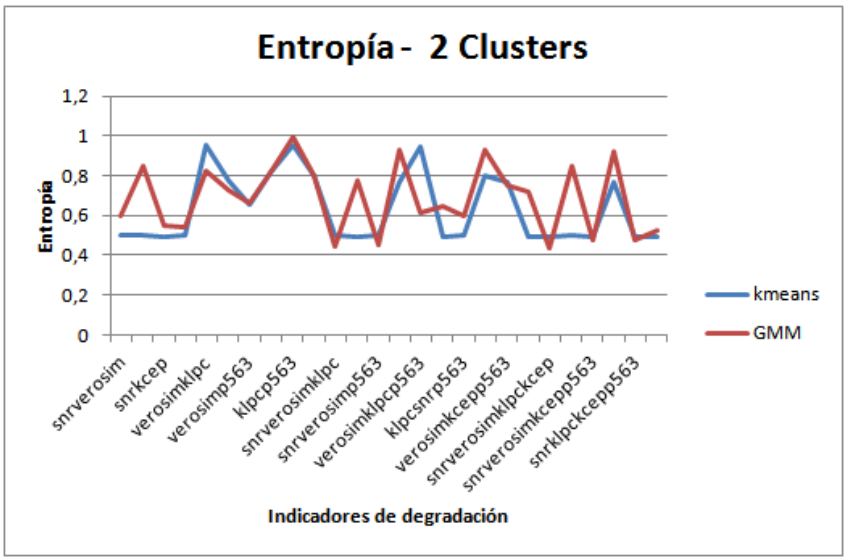


Figura 1: Entropía en función del tipo de algoritmo de agrupación empleado para dos grupos

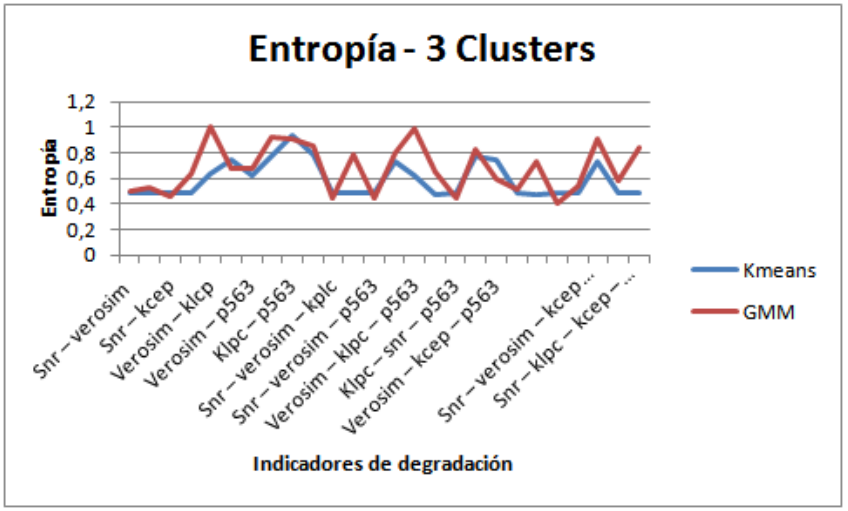


Figura 2: Entropía en función del tipo de algoritmo de agrupación empleado para tres grupos

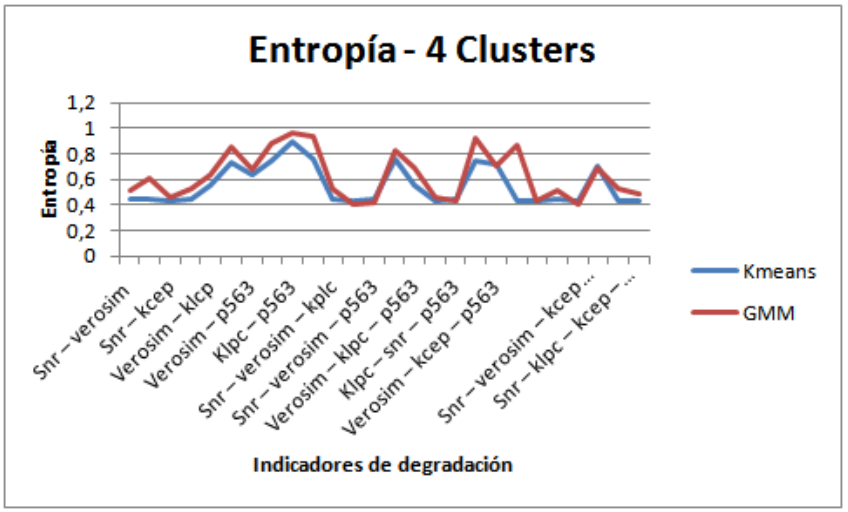


Figura 3: Entropía en función del tipo de algoritmo de agrupación empleado para cuatro grupos

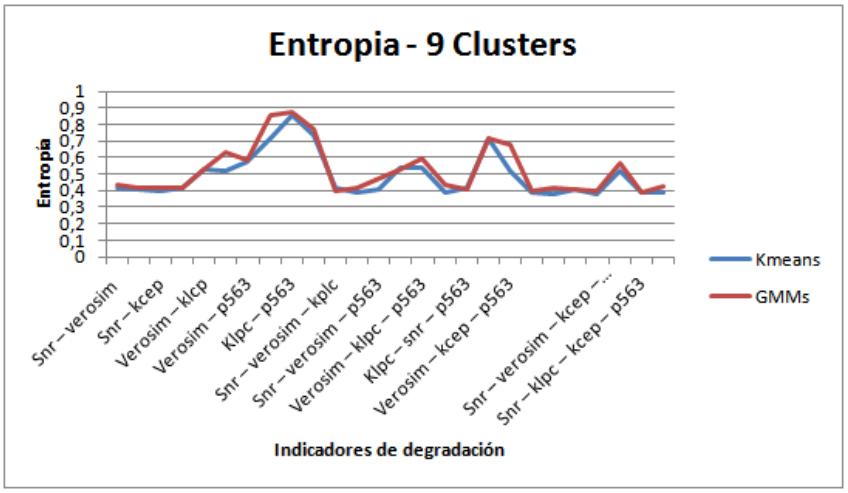


Figura 4: Entropía en función del tipo de algoritmo de agrupación empleado para nueve grupos