7 EXPERIMENTOS

En este capítulo se procede a incluir los diferentes experimentos realizados en la elaboración de este proyecto fin de carrera. Se mostrarán los resultados obtenidos con las cuatro técnicas analizadas. En capítulos anteriores hemos visto que independientemente de la técnica de verificación de locutor con la que se desee trabajar, siempre existen una gran cantidad de parámetros y opciones comunes que hay que definir: número de componentes o mezclas gaussianas, dimensión de los vectores cepstrales, tipo de normalización de puntuaciones, modo de SAD, etc. Debido a la importancia fundamental e influencia de GMM-UBM en las restantes técnicas (el punto de partida siempre se basa en GMM-UBM), todos los experimentos necesarios para seleccionar estas opciones se llevarán a cabo dentro de esta primera técnica. Los parámetros escogidos con GMM-UBM serán los mismos que se empleen en el resto de técnicas, para así poder realizar una comparación completamente justa entre todas las técnicas, ya que todas serán evaluadas ante la misma base de datos y ante las mismas condiciones. Para medir el rendimiento de los diferentes sistemas haremos uso de la curva DET, de la EER y del minDCF. En líneas generales, valores relativamente bajos de la EER deberían de corresponderse con valores pequeños del minDCF. Sin embargo, la relación entre estas magnitudes no es lineal, y se darán casos para sistemas con rendimientos muy parecidos en los que uno presenta una minDCF más pequeña y una EER mayor. En dichas situaciones se dará mayor prioridad al minDCF, puesto que es el parámetro que utiliza el NIST para evaluar los sistemas y se ajusta más a una interpretación real.

7.1 GMM-UBM

En esta sección se llevarán a cabo los experimentos relacionados con la selección de los parámetros y opciones relativos a SAD, dimensión de los coeficientes cepstrales, número de mezclas gaussianas y tipo de normalización de puntuaciones. La elección de estos parámetros es determinante puesto que como se adelantó anteriormente, el resto de técnicas utilizarán las mismas opciones seleccionadas por GMM-UBM. Una vez definido estos parámetros, se obtendrá el resultado representativo de GMM-UBM, el cual se podrá comparar con el del resto de técnicas.

7.1.1 Weight SAD vs Mean SAD

En este apartado se ejecutarán pruebas a fin de averiguar cuál de las dos opciones explicadas en 3.3 es más óptima. Tal y como definimos en el apartado 6.2, la herramienta SPro permite obtener coeficientes cepstrales de 60 dimensiones. Se le pueden exigir otros valores, pero en este proyecto se ha decidido partir con 60 componentes, de las cuales:

- 0-18: coeficientes estáticos
- 19: energía de los coeficientes estáticos
- 20-38: coeficientes de velocidad
- 39: energía de la velocidad
- 40-58: coeficientes de aceleración
- 59: energía de la aceleración

Sin embargo, a la hora de procesar los coeficientes cepstrales para el entrenamiento o el test, no se tiene necesariamente que trabajar con las 60 componentes. En este proyecto se han barajado 8 opciones con ALIZE, expuestas en la siguiente tabla:

Número de componentes	Componentes utilizadas	Tipos de coeficientes
30	0-15, 20-27, 39-44	16+8Δ+5ΔΔ+ ΔE
32	0-15, 20-35	16+16Δ
33	0-15, 20-35, 39	16+16Δ+ ΔE
39	0-18, 20-39	19+19Δ+ ΔE
49	0-15, 20-35, 39-55	16+16Δ+16ΔΔ+ ΔE
50	0-18, 20-50	19+19Δ+11ΔΔ+ ΔE
58	0-18, 20-58	19+19Δ+19ΔΔ+ ΔE
60	0-59	19+19Δ+19ΔΔ+E+ ΔE+ ΔΔE

Tabla 2. *Opciones disponibles en ALIZE en este proyecto sobre el número de componentes cepstrales con las que trabajar dentro del conjunto de 60.*

En la tabla anterior se puede contemplar que un número no acompañado del operador delta hace referencia al número de coeficientes estáticos, mientras que Δ y ΔΔ se utiliza para indicar el número de coeficientes dinámicos. E, ΔE y ΔΔE identifican la componente de energía de los coeficientes de posición, velocidad y aceleración, respectivamente.

La opción de 60 es la que no descarta ninguna componente. Resulta llamativo que todas las opciones salvo la 32 cuentan con la componente ΔE (energía de los 19 coeficientes de velocidad). Ello se debe a que tradicionalmente siempre se le ha dado muchísima importancia a este parámetro, y se considera fundamental para la verificación (aporta información sobre la variabilidad de un locutor en promedio).

En el siguiente experimento, se han obtenido los valores de EER y minDCF para estas 8 posibilidades con la técnica de GMM-UBM tanto para Mean SAD como para Weight SAD. El número de gaussianas con el que se ha trabajado en esta prueba es de M=64 y no se ha llevado a cabo ningún tipo de normalización de puntuaciones. Con respecto al parámetro α_{SAD} descrito en el apartado 3.3, de momento y hasta que no se diga lo contrario se trabajará con el valor por defecto que equivale a cero. En posteriores apartados se analizará la importancia de dicho parámetro en la eficiencia del sistema, así como el del número de mezclas gaussianas y el tipo de normalización de puntuaciones.

	WEIGHT SAD		MEAN SAD	
#componentes	EER (%)	minDCF (×100)	EER (%)	minDCF (×100)
30	10.64	4.784	10.48	4.663
32	10.64	4.850	10.24	4.739
33	10.64	4.844	10.48	4.798
39	10.89	5.004	9.99	4.826
49	10.56	4.782	10.40	4.671
50	10.72	4.910	10.24	4.639
58	10.64	4.934	10.64	4.721
60	10.56	4.863	10.32	4.781

Tabla 3. *Resultados obtenidos en GMM-UBM, con M=64 y para los dos modos posibles de SAD.*

Para ilustrar gráficamente los resultados expuestos en la tabla 3, hemos incluido las dos gráficas siguientes:

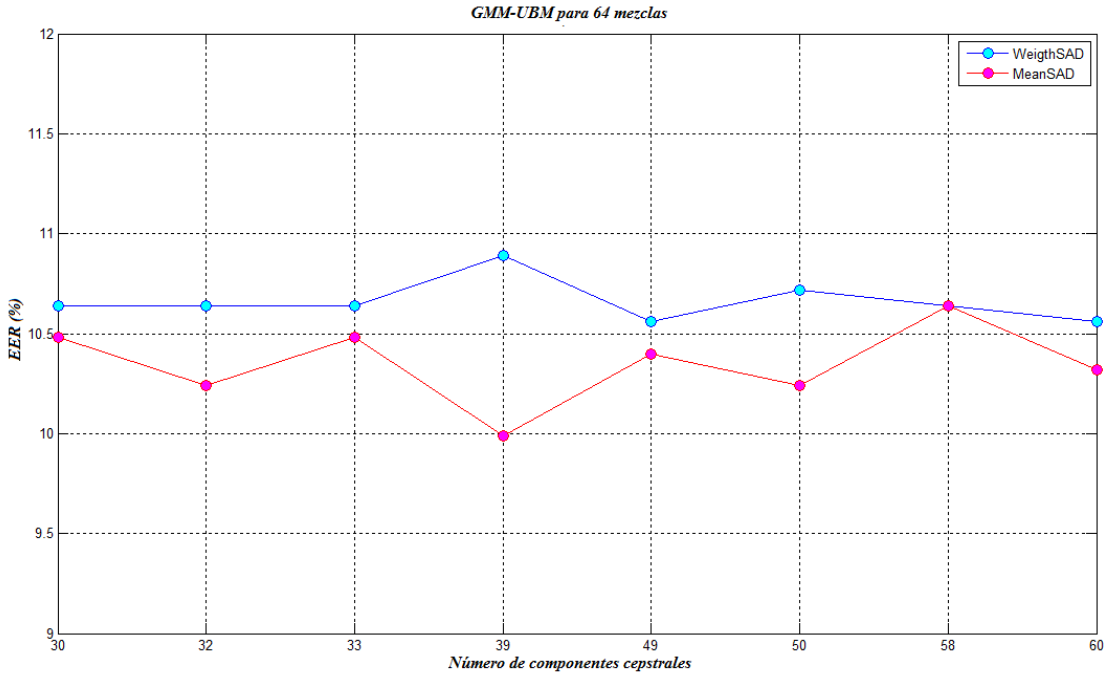

Figura 38. *Resultados con respecto a la EER obtenidos en GMM-UBM, con M=64 y para los dos modos posibles de SAD.*

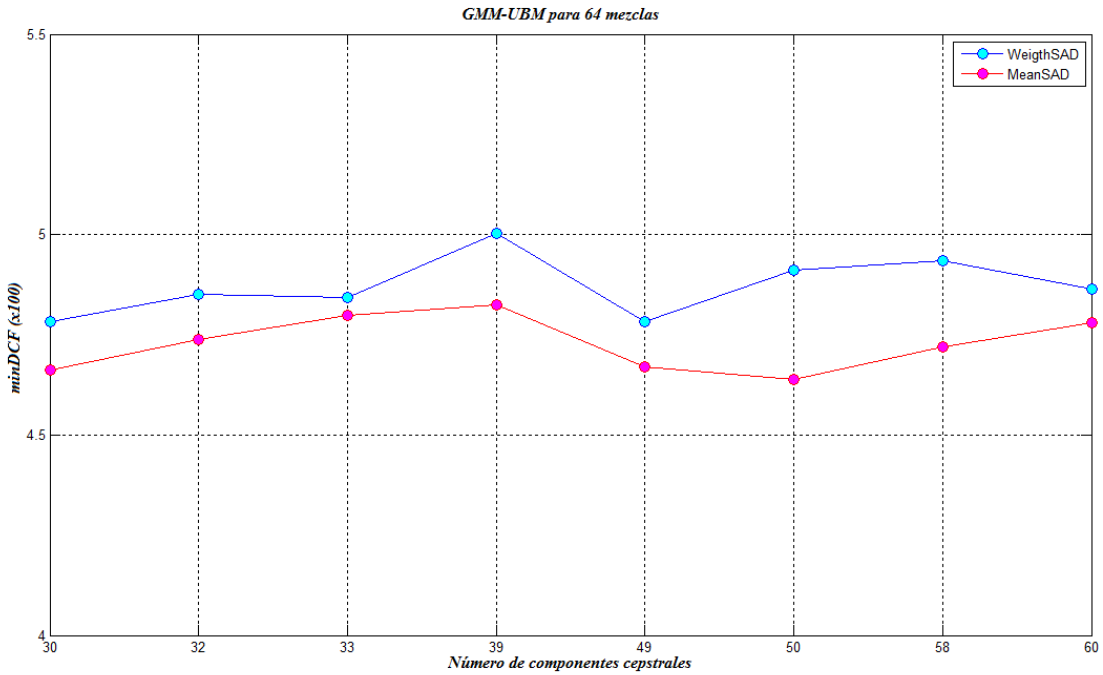

Figura 39. *Resultados con respecto al minDCF(×100) obtenidos en GMM-UBM, con M=64 y para los dos modos posibles de SAD.*

A la vista de los resultados anteriores, podemos observar como independientemente de la dimensión de los vectores cepstrales, el modo Mean siempre produce mejores resultados tanto a nivel de minDCF como a nivel de EER. Luego a partir de este momento y hasta el final del proyecto, siempre se trabajará con este modo en la etapa de detección de energía.

7.1.2 Elección del número de components cepstrales

Tras escoger el modo de SAD en el que se van a implementar todas las técnicas en este proyecto, el siguiente paso consistirá en escoger el número de componentes cepstrales que producen mejor rendimiento. Puesto que la dimensión óptima de los vectores de características depende notablemente de los datos específicos con los que se esté trabajando y no suele tener una influencia relativamente alta en el rendimiento final (como si ocurre por ejemplo con el modo de SAD que se emplee), el criterio de selección será únicamente el minDCF (priorizamos el criterio de evaluación del NIST).

Tal y como hemos comentado anteriormente, un parámetro clave y fundamental en este proyecto es la elección del número de mezclas gaussianas, puesto que en la sección de teoría hemos comprobado cómo las cuatro técnicas hacen uso del concepto de GMM. Sin embargo, aún no sabemos el valor óptimo para el número de mezclas. En este caso, debido a que la dimensión de los vectores cepstrales y el número de mezclas

sí mantienen un cierto grado de correlación, se ha decidido abordar este experimento con todas las opciones posibles sobre el número de mezclas.

Tradicionalmente, en GMM, el número de mezclas siempre se ha correspondido con potencias de 2. En este proyecto se han barajado las opciones de 16, 32, 64, 128, 256, 512, 1024 y 2048. Generalmente no se suele emplear nunca la opción de 4096 puesto que se considera que no produce mejores resultados que 2048, ya que 2048 componentes se considera una cantidad suficiente para representar a la población universal. Sin hacer uso de momento de ningún tipo de normalización y utilizando Mean SAD, los resultados obtenidos para la minDCF aparecen reflejados en la tabla siguiente:

		MEZCLAS							
		16	**32**	**64**	**128**	**256**	**512**	**1024**	**2048**
#COMPONENTES	**30**	5.864	5.102	4.663	4.391	4.043	3.988	3.834	3.823
	32	5.792	5.351	4.739	4.551	4.208	4.091	3.875	3.869
	33	5.776	5.241	4.798	4.485	4.197	4.026	3.837	3.782
	39	5.764	5.222	4.826	4.502	4.107	3.869	3.754	3.667
	49	5.353	5.133	4.671	4.333	4.061	3.834	3.723	3.713
	50	5.433	5.035	4.639	4.243	3.93	3.681	3.44	3.441
	58	5.480	5.031	4.721	4.428	4.039	3.736	3.66	3.560
	60	5.377	5.219	4.781	4.308	4.056	3.760	3.584	3.424

Tabla 4. *Valores obtenidos para el minDCF(×100) mediante GMM-UBM sin normalización de puntuaciones y en modo MeanSAD.*

En la figura 40 se puede contemplar una gráfica asociada a los resultados expuestos en la tabla 4. A la vista de ellos, se observa que a excepción de los casos en los que el número de mezclas es 16 o 32, el uso de vectores cepstrales con 50 componentes maximiza el rendimiento del sistema. Con lo cual, se concluye que el número de componentes cepstrales con el que se procesará los ficheros de entrenamiento y evaluación en este proyecto es de 50. Es decir, la opción elegida será la de tratar con vectores cepstrales D-dimensionales con $D = 50 = 19 + 19\Delta + 11\Delta\Delta + \Delta E$.

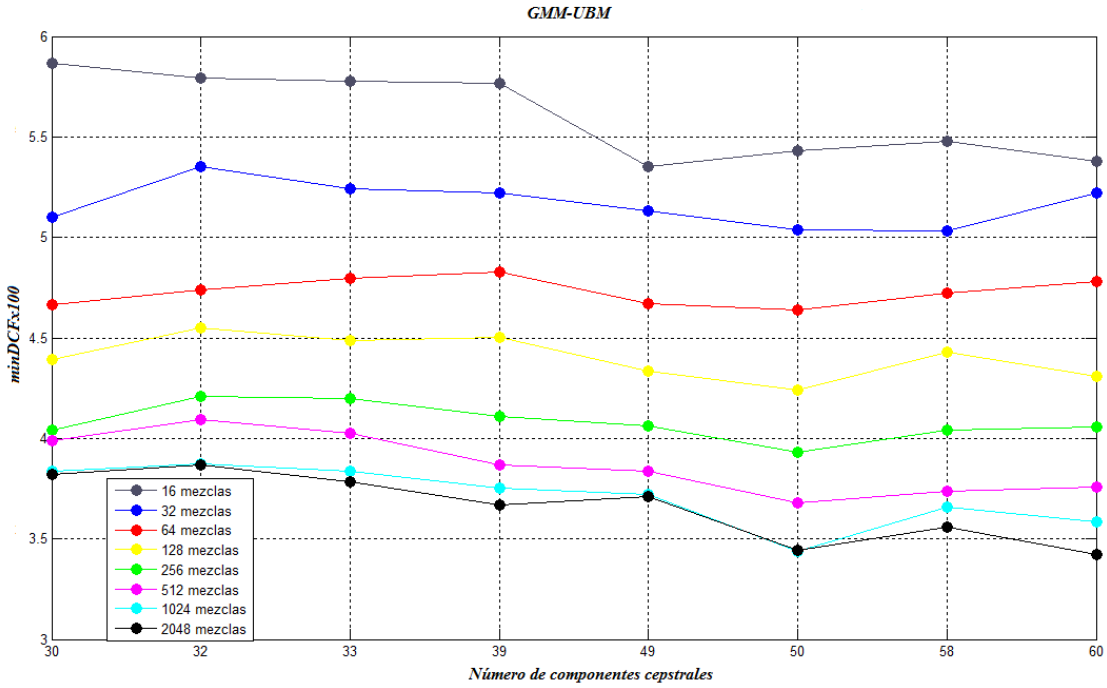

Figura 40. *Valores obtenidos para el minDCF(×100) mediante GMM-UBM sin normalización de puntuaciones y en modo MeanSAD.*

La gráfica anterior volverá a tenerse en cuenta en apartados posteriores donde se tendrá que seleccionar el número de mezclas gaussianas más óptimo.

De los resultados obtenidos con respecto a la elección de la dimensión D de los vectores cepstrales, resultan llamativos los siguientes aspectos a tener en cuenta:

- Analizando la mejora en el rendimiento al utilizar 33 componentes en lugar de 32 se contempla la importancia de introducir el término ΔE (variabilidad de interlocutor).

- Sin embargo, se observa que es preferible trabajar con 30 componentes en lugar de 32 o 33, con lo cual es recomendable tener en cuenta los coeficientes de aceleración también y no trabajar únicamente con los estáticos y de velocidad.

- Comparando 33 con 39 se extrae que es aconsejable trabajar con todos los coeficientes estáticos y de velocidad posibles.

- La mejora de 50 componentes sobre 49 refleja que en verificación de locutor, los coeficientes de velocidad son más importantes que los de aceleración.

- A pesar de que en puntos anteriores hemos remarcado que siempre resulta positivo incluir algunos coeficientes de aceleración, la degradación en los resultados al pasar de 50 a 58 refleja que un exceso en el uso de coeficientes de aceleración puede provocar más errores en el sistema.

7.1.3 Elección de la normalización de puntuaciones

En el apartado 3.5 se describió los cuatro tipos de normalización sobre puntuaciones con las que se suele trabajar en sistemas de Verificación Automática de Locutor (Tnorm, Znorm, TZnorm y ZTnorm). En esta sección se analizará la eficiencia de GMM-UBM en función del tipo de normalización y además se comparará con el caso en el que no se hace ninguna normalización (como se ha venido haciendo en todos los experimentos hasta ahora). Al igual que en el apartado anterior, puesto que aún no se ha fijado el número de mezclas gaussianas con la que se va a trabajar, se ha optado por mostrar los resultados para todas las posibilidades existentes. Los resultados de esta prueba experimental aparecen recogidos en la siguiente tabla:

	MEZCLAS							
	16	**32**	**64**	**128**	**256**	**512**	**1024**	**2048**
No-norm	5.433	5.035	4.639	4.243	3.930	3.681	3.440	3.441
T-norm	4.017	3.816	3.646	3.522	3.270	3.270	3.060	3.034
Z-norm	4.564	4.1	3.953	3.795	3.544	3.587	3.729	4.095
TZ-norm	4.709	4.261	3.947	3.701	3.451	3.346	3.174	3.172
ZT-norm	5.255	4.614	4.387	4.136	3.822	3.650	3.379	3.285

Tabla 5. *Valores obtenidos para el minDCF(×100) mediante GMM-UBM con D=50 y en modo MeanSAD.*

Analizándolo desde un punto de vista gráfico, disponemos de la siguiente figura:

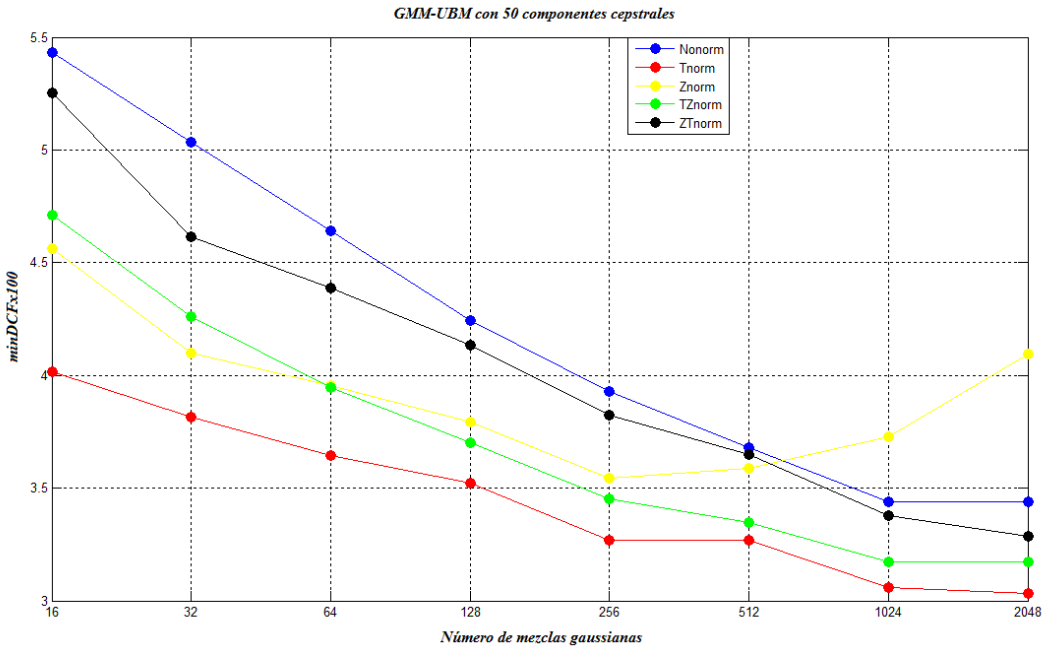

Figura 41. *Representación gráfica de los valores obtenidos para el minDCF(×100) mediante GMM-UBM con D=50 y en modo MeanSAD.*

Analizando la gráfica, se observa claramente que T-norm es la opción que proporciona mejor rendimiento (siguiendo el criterio del minDCF). Por lo cual será la normalización con la que se trabajará en lo que queda de proyecto. Resulta especialmente llamativo que cuando el número de mezclas gaussianas es muy alto (1024 o 2048), la normalización tipo Z genera un rendimiento inferior que cuando no se hace uso de ningún tipo de normalización sobre puntuaciones. Además, de los datos se concluye que a partir de 64 mezclas, TZ-norm genera mejores resultados que ZT-norm o Z-norm, de ahí que en el caso en el que se optase por desarrollar en serie ambas normalizaciones (Z y T) es preferible la normalización de test primero.

Atendiendo a la teoría, se podría pensar que cuando se hacen ambas tipos de normalizaciones se deberían de obtener mejores resultados que cuando se hace solamente una. Sin embargo, con TZ o ZT no se están combinando ambas normalizaciones realmente, sino que se desarrolla una y después la otra en serie. De cualquier manera, no existe una regla general para los tipos de normalizaciones, ya que el rendimiento de cada una de ellas depende en cierta medida de los datos con los que se esté trabajando. Finalmente, cabría comentar que esta gráfica, al igual que la del apartado anterior, será muy útil para la siguiente sección, en la que seleccionaremos el número de mezclas gaussianas que proporciona un resultado más eficiente.

7.1.4 Elección del número de mezclas gaussianas

Como ya se ha comentado, todas las técnicas en este proyecto parte o hacen uso de GMM, de ahí que la selección del número de mezclas gaussianas juegue un papel fundamental. Hasta el momento, se ha concluido que se trabajará en modo Mean SAD, con 50 componentes cepstrales y con la normalización de puntuaciones T-Norm. Con estas condiciones, los valores de EER y minDCF obtenidos variando el número de componentes gaussianas aparecen reflejados en la siguiente tabla:

Número de mezclas	EER (%)	minDCF × 100
16	10.97	4.017
32	10.15	3.816
64	9.91	3.646
128	9.75	3.522
256	8.85	3.270
512	7.88	3.270
1024	7.88	3.060
2048	7.72	3.034

Tabla 6. *Valores obtenidos para el minDCF y EER mediante GMM-UBM en modo MeanSAD, con 50 componentes cepstrales y con T-norm*

Analizando en detalle esta tabla junto con las figuras 40 y 41 se observa que conforme se incrementa el número de mezclas gaussianas, casi siempre se obtienen mejores resultados. Esto es así puesto que al utilizar un numero alto de componentes gaussianas, se está teniendo en cuenta mayor variabilidad para la población de locutores.

De hecho, atendiendo a las figuras 15 y 16 correspondientes a la sección 4.2 de GMM-UBM, donde se representaba la distribución de una componente cepstral de un locutor determinado mediante suma ponderada de gaussianas, se puede observar la diferencia en la precisión del modelo en función de si se hace con 4 o 32 mezclas. Al trabajar con 32, se están abarcando más posibilidades para modelar al usuario, o dicho de otro modo, se está trabajando con mayor precisión. El desajuste entre el modelo del usuario mediante 4 mezclas gaussianas y su distribución real es consecuencia de lo que en la literatura se conoce como *underfitting*.

Sin embargo, no siempre es posible trabajar con un número muy elevado de mezclas. Para poder llevar a cabo esta posibilidad, es necesario que el fichero que se utiliza para generar el modelo universal tenga un elevado número de grabaciones, como se ha desarrollado en este proyecto. De lo contrario, existiría un alto grado de incoherencia, puesto que se estaría representado el modelo UBM con muy pocos datos o con poca variabilidad, pero al mismo tiempo se estaría imponiendo que el modelo universal es muy variable puesto que se disponen de muchas componentes gaussianas. Éste efecto negativo se conoce como *overfitting*.

Hasta aquí, se podría razonar perfectamente de que el caso ideal es usar 2048 gaussianas. Sin embargo, hay dos factores que rechazan esta posibilidad. Por una parte, hay que tener en cuenta el alto coste computacional que supone trabajar por ejemplo con 2048 o 1024 componentes (teniendo en cuenta la cantidad de grabaciones enviados por el NIST, las simulaciones pueden llegar a tardar incluso más de 48 horas) De hecho, en relación a la figura 40, se puede contemplar cómo la línea correspondiente a 1024 mezclas y la de 2048 se solapan. Incluso de la tabla anterior se extrae que 512 y 1024 generan exactamente la misma EER (7.88%), mientras que la mejora en 2048 es muy reducida (7.72%). Esta hipótesis se puede justificar perfectamente con la figura 42, donde se ha representado la curva DET para las 8 componentes gaussianas.

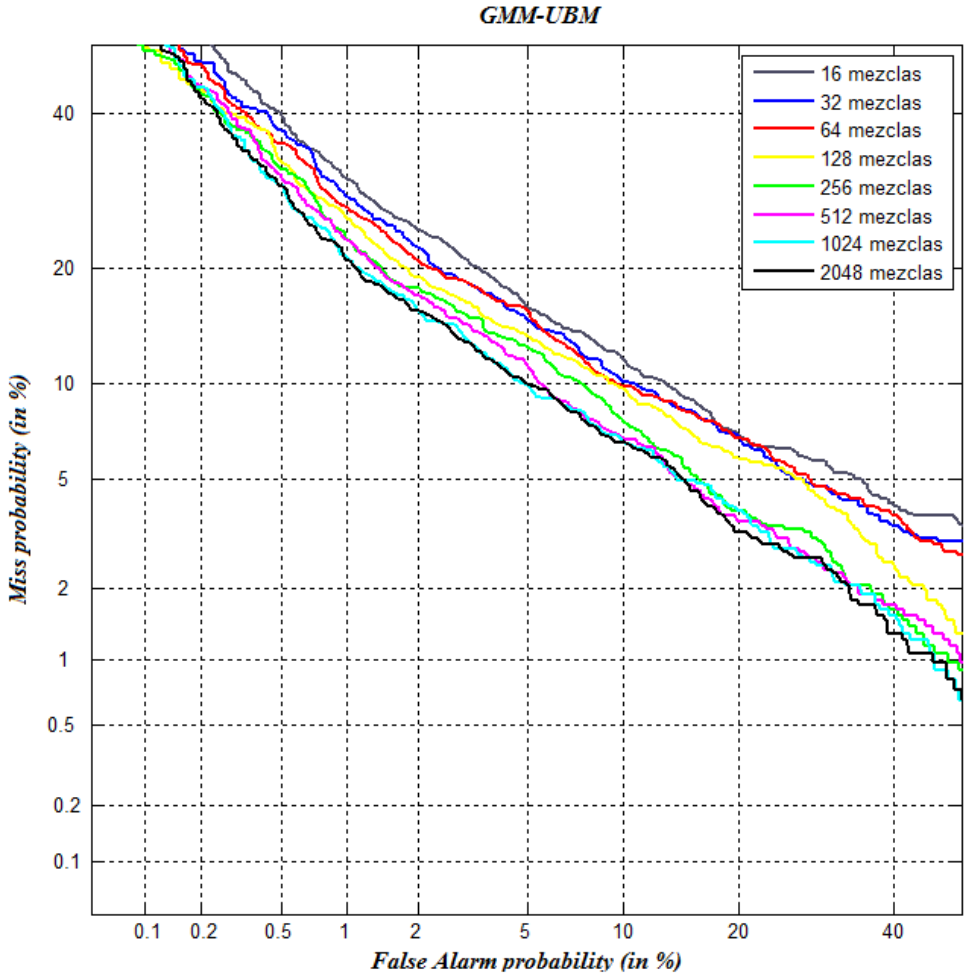

Figura 42. *Curvas DEC obtenidas mediante GMM-UBM en modo MeanSAD, con D=50 y con T-norm.*

La curva DET constituye una forma muy objetiva e imparcial de analizar la eficiencia de un sistema, puesto que nos permite ver con total detalle todos los puntos en los que puede trabajar el sistema (todos los combinaciones de P_{miss} y P_{fa}). Además, se vuelve a demostrar cómo conforme se aumenta el tamaño de las mezclas, las curvas descienden gradualmente, mejorando por lo tanto el rendimiento. Sin embargo, en este caso resulta complicado distinguir la mejora entre 512, 1024 y 2048, debido a la existencia de puntos en las que dichas curvas se solapan entre ellas. Aplicando un zoom a la figura anterior, se puede vislumbrar este hecho de manera más concisa en la siguiente figura:

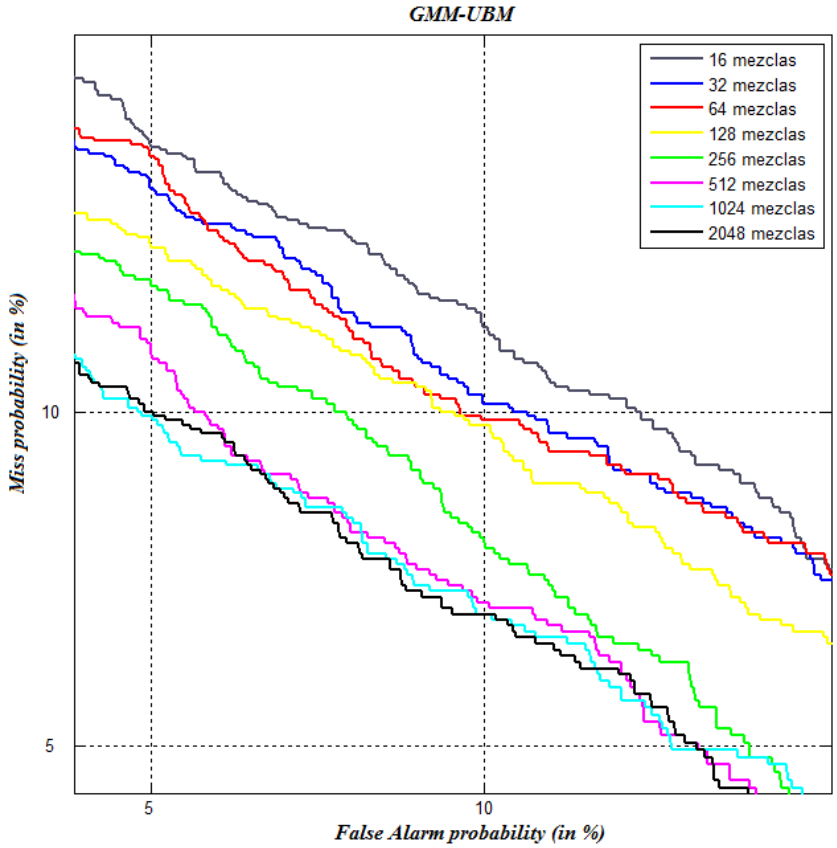

Figura 43. *Zoom aplicado a las curvas DEC obtenidas mediante GMM-UBM en modo MeanSAD, con D=50 y con T-norm.*

Con el zoom anterior, se puede contemplar perfectamente cómo queda claro la diferencia en el rendimiento para mezclas entre 16 y 256 componentes, y cómo a partir de 512 existe un solapamiento entre las curvas que se traduce en un estado de saturación en el rendimiento del sistema. Generalmente, este fenómeno de saturación se suele atribuir al hecho de que el fichero utilizado para generar el UBM no es muy extenso, o no contiene suficiente variabilidad en los datos para producir una mejora sustancial al trabajar con 1024 o 2048 mezclas. Con lo cual habría que incorporar más grabaciones de interlocutores a dicho fichero para evitar la saturación anterior.

Atendiendo a lo explicado en el párrafo anteior y al hecho de que las diferencias en la EER y la minDCF son mínimas a partir de 512 componentes gaussianas, junto con el aliciente del coste computacional y el tiempo de ejecución para las pruebas experimentales, el número de mezclas gaussianas con el que se trabajará será de M=512.

7.1.5 Elección del umbral para SAD

En el capítulo 3.3 correspondiente a SAD, se explicó la influencia del parámetro α_{SAD} a la hora de definir el umbral que se utiliza para descartar las tramas silenciosas en la etapa de detección de actividad de energía. Tal y como ya se referenció al inicio de este capítulo de simulaciones, todas los resultados mostrados hasta este momento se han llevado a cabo tomando la opción por defecto que se corresponde con α_{SAD}=0. En esta sección se realizarán pruebas para investigar la eficiencia de GMM-UBM en función del valor de este parámetro. Atendiendo a la ecuación 3.13, en la opción de MeanSAD (que es la opción que ha sido seleccionada para todos los experimentos) se observa como el parámetro α_{SAD} multiplica a la desviación típica, produciendo un desplazamiento relativo con respecto a la media para establecer el umbral. Debido a que no tiene sentido desviarse mucho de la media, las posibilidades que se han abarcado para este parámetro oscilan desde -0.5 hasta 0.5, llegando a los resultados representados en la tabla siguiente:

α_{SAD}	EER (%)	minDCF × 100
-0.5	7.72	3.150
-0.4	7.47	3.092
-0.3	7.72	3.044
-0.2	7.80	3.111
-0.1	7.88	3.091
0	7.88	3.270
0.1	8.29	3.271
0.2	9.26	3.521
0.3	9.99	3.697
0.4	9.91	3.862
0.5	10.15	4.536

Tabla 7. *Valores obtenidos para el minDCF y EER mediante GMM-UBM en modo MeanSAD, con M=512, D=50 y con T-norm.*

De la tabla 7 se concluye que el valor óptimo de α_{SAD} se da para -0.4 utilizando como criterio la EER y -0.3 para el minDCF. Así que se optará por trabajar con la segunda opción debido a que en este proyecto el minDCF tiene más prioridad que la EER. En las figuras 44 y 45 se han mostrado una representación gráfica de estos resultados, para intentar analizar de manera más clara la influencia de este parámetro desde un punto de vista teórico. En una primera etapa muy breve, observamos como las curvas correspondientes al minDCF o a la EER comienzan a descender, puesto que se están estableciendo valores del umbral excesivamente altos (el sistema de detección de energía está actuando de manera muy 'rigurosa' a la hora de descartar tramas). Sin embargo, rápidamente empiezan a aumentar, puesto que conforme se incrementa el parámetro, de acuerdo con la ecuación 3.13 disminuye el umbral de energía, lo cual permite que los tramos silenciosos se tengan en cuenta como información útil.

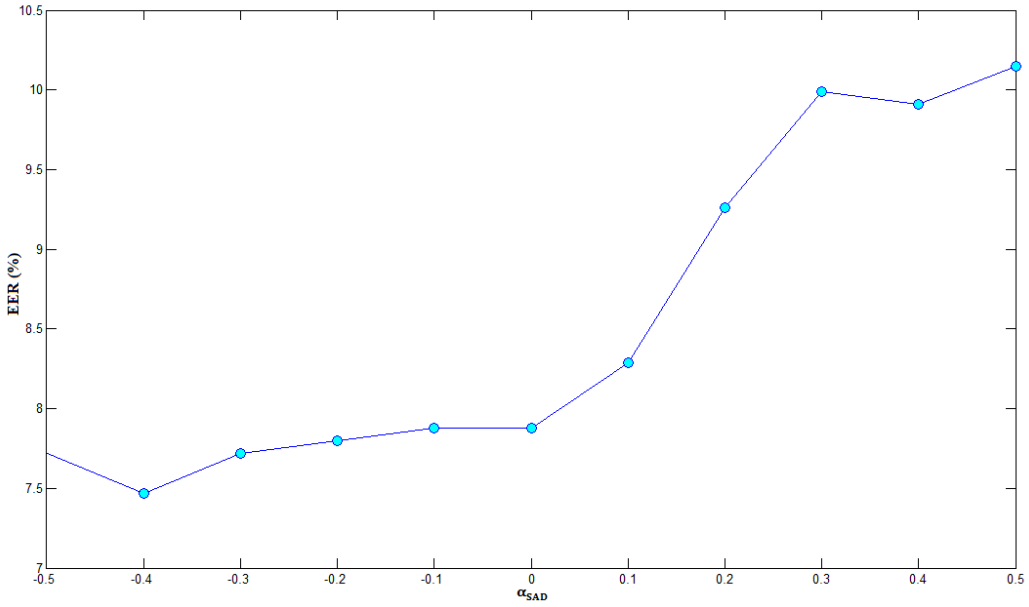

Figura 44. *Resultados para la EER mediante GMM-UBM en modo MeanSAD, con M=512, D=50 y con T-norm.*

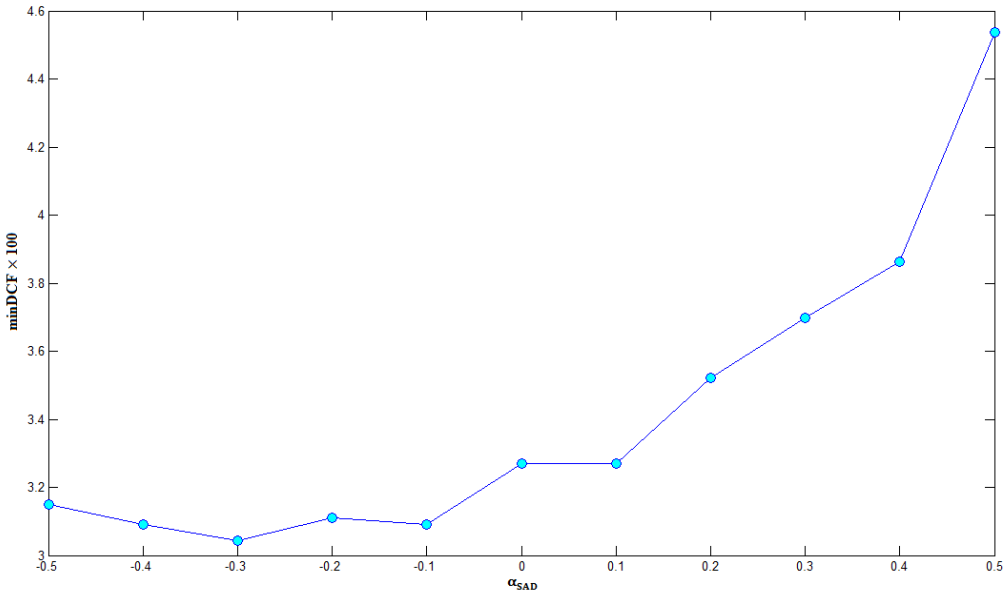

Figura 45. *Resultados para el minDCF mediante GMM-UBM en modo MeanSAD, con M=512, D=50 y con T-norm.*

Una vez llegado hasta este punto, tenemos todos los parámetros necesarios definidos para poder adentrarnos en otras técnicas y comparar en igualdad de condiciones. Los parámetros que han sido seleccionados como aquellos para los cuales GMM-UBM produce un rendimiento más eficiente se pueden resumir como:

- Mean SAD con $\alpha_{SAD} = -0.3$
- D=50 componentes cepstrales
- T-norm como técnica de normalización para las puntuaciones.
- M=512 mezclas para los modelos gaussianas.

Para el análisis experimental de las otras tres restantes técnicas, se utilizarán los mismos parámetros, ya que si estamos interesados en comparar unas técnicas con otras de una manera justa y objetiva, todas deberían de estar sometidas a las mismas condiciones. Recordar que todos los parámetros que se han analizado exhaustivamente en esta sección con GMM-UBM son comunes a todas las técnicas, aunque hemos decidido experimentarlos con GMM-UBM por ser ésta la primera técnica que se ha abordado en este proyecto y que se considera además como base fundamental para las otras tres. Se podría dar el caso de que los parámetros óptimos para GMM-UBM no lo sean para otra técnica, pero esa posibilidad no ha sido contemplada en este proyecto puesto que el objetivo final es comparar las técnicas con los mismos parámetros comunes de configuración. Además el hecho de que los experimentos para maximizar estos parámetros se hayan llevado a cabo con GMM-UBM cobra bastante sentido debido a su influencia en todas las técnicas restantes, puesto que todas ellas hacen uso del concepto de mezclas gaussianas y del modelo universal. Por ejemplo, en SVM se crean los supervectores concatenando las medias correspondientes a mezclas gaussianas. Si estamos interesados en comparar GMM-UBM y SVM de manera objetiva, no tiene sentido que en una técnica trabajemos con un número de mezclas determinado y en la otra con otro diferente o que se utilice en un caso normalización T y en otro Z. Es necesario usar una referencia común para todas ellas.

Tras haber finalizado la selección de todos los parámetros más óptimos y descartado el resto de posibilidades, el rendimiento final de GMM-UBM aparece reflejado en la siguiente figura, donde se ha representado la curva DET junto con los valores de EER y minDCF. Dicha figura representativa de GMM-UBM se comparará con el resto de técnicas para analizar los diferentes rendimientos.

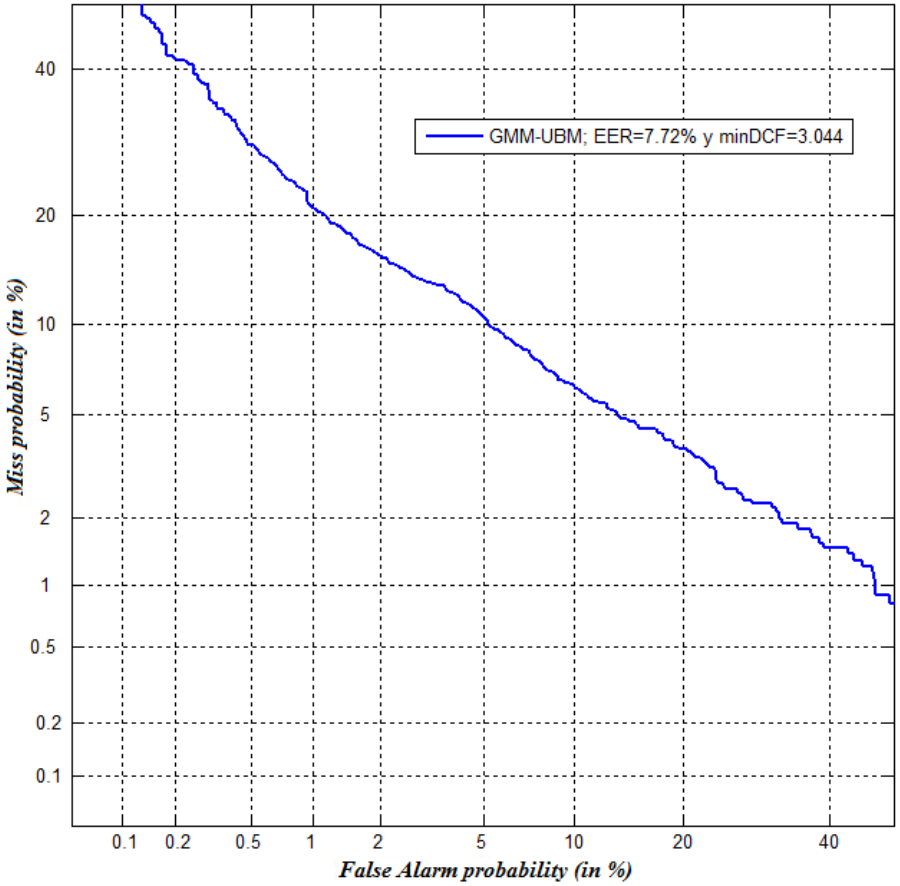

Figura 46. *Rendimiento de GMM-UBM en modo MeanSAD, con $\alpha_{SAD} = -0.3$, M=512, D=50 y con T-norm.*

7.1.6 Rendimiento de GMM-UBM en función de las condiciones de entrenamiento y evaluación

En este apartado nos desviaremos un poco de lo que hemos venido haciendo hasta ahora y antes de adentrarnos en el análisis del rendimiento de las restantes técnicas, continuaremos con GMM-UBM para comparar la eficiencia del sistema en función de las condiciones de entrenamiento y de test. Tal y como se afirmó en el apartado 6.1 sobre las evaluaciones del NIST de 2005, existen 20 combinaciones posibles sobre condiciones de training y de test. Durante todo el proyecto, a excepción de este apartado, se ha trabajado con la única condición obligatoria impuesta por el NIST de 1conv4w-1conv4w (2.5 minutos de grabación para el entrenamiento y test, respectivamente). Sin embargo, una de las condiciones posibles analizando la tabla 1 es trabajar únicamente con 10 segundos. En este apartado vamos a comparar los resultados entre 2.5 minutos y 10 segundos, donde se obtendrá una conclusión muy relevante y además general para cualquier sistema de reconocimiento biométrico.

Atendiendo a lo anteriormente comentado, se disponen de 4 posibilidades o combinaciones:

- 1conv4w-1conv4w: 2.5 minutos para entrenamiento y test (combinación estándar de este proyecto y que se ha usado en todas las simulaciones hasta este momento)

- 10sec4w-10sec4w: 10 segundos de grabaciones para entrenamiento y test

- 1conv4w-10sec4w: 2.5 minutos para entrenamiento y 10 segundos para evaluación

- 10sec4w-1conv4w: 10 segundos para training y 2.5 minutos para evaluación

Trabajando con los parámetros seleccionados en los apartados anteriores, con GMM-UBM se ha obtenido la siguiente figura para las 4 posibilidades existentes:

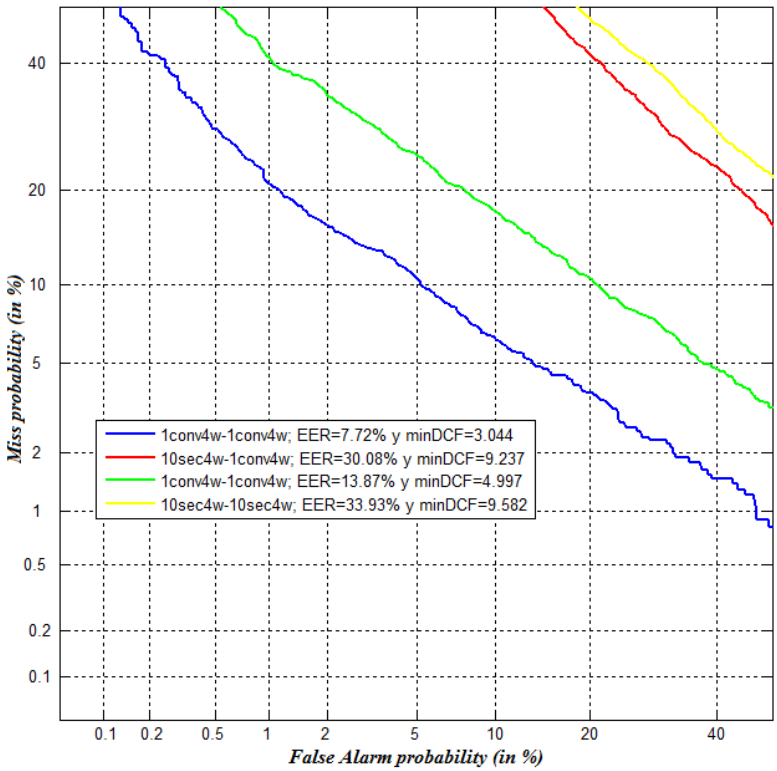

Figura 47. *Rendimiento de GMM-UBM para diferentes condiciones de training y test en modo MeanSAD, con* $\alpha_{SAD} = -0.3$*, M=512, D=50 y con T-norm.*

Naturalmente, la opción de 2.5 minutos para el entrenamiento y test genera mejores resultados que 10 segundos para ambas fases, puesto que al disponer de más tiempo de grabación se dispone de más información del locutor. Sin embargo, resulta llamativo cómo la línea roja correspondiente a la opción 10sec4w-1conv4w se encuentra muy próxima a la amarilla correspondiente a 10sec4w-10sec4w, con lo que se concluye que siempre que no se haya entrenado adecuadamente un locutor, el aumentar el tiempo para el test no produce una mejora sustancial en el sistema (observamos como para el test hemos pasado de 10 segundos a 2.5 minutos y sin embargo la EER se ha reducido tan solo desde un 32.85% hasta un 29.84%). Además, de la figura anterior se extrae que siempre y cuando tengamos dos ficheros de distinta longitud, es aconsejable utilizar el fichero largo para el entrenamiento y el corto para el test (véase las diferencias existentes entre las líneas verde y amarilla).

7.2 SVM

En este apartado se visualizarán los resultados obtenidos con SVM. Tal y como se añadió anteriormente, todos los parámetros relativos a tipo de normalización, número de mezclas gaussiana, etc. coincidirán con los seleccionados con GMM. Recordar que aunque SVM se trata de una técnica discriminativa, para obtener los supervectores previamente es necesario crear los modelos GMM de cada locutor, permitiendo así concatenar las medías para poder formar dichos supervectores. Los resultados de SVM, comparados con los calculados previamente en GMM-UBM se muestran en la siguiente figura:

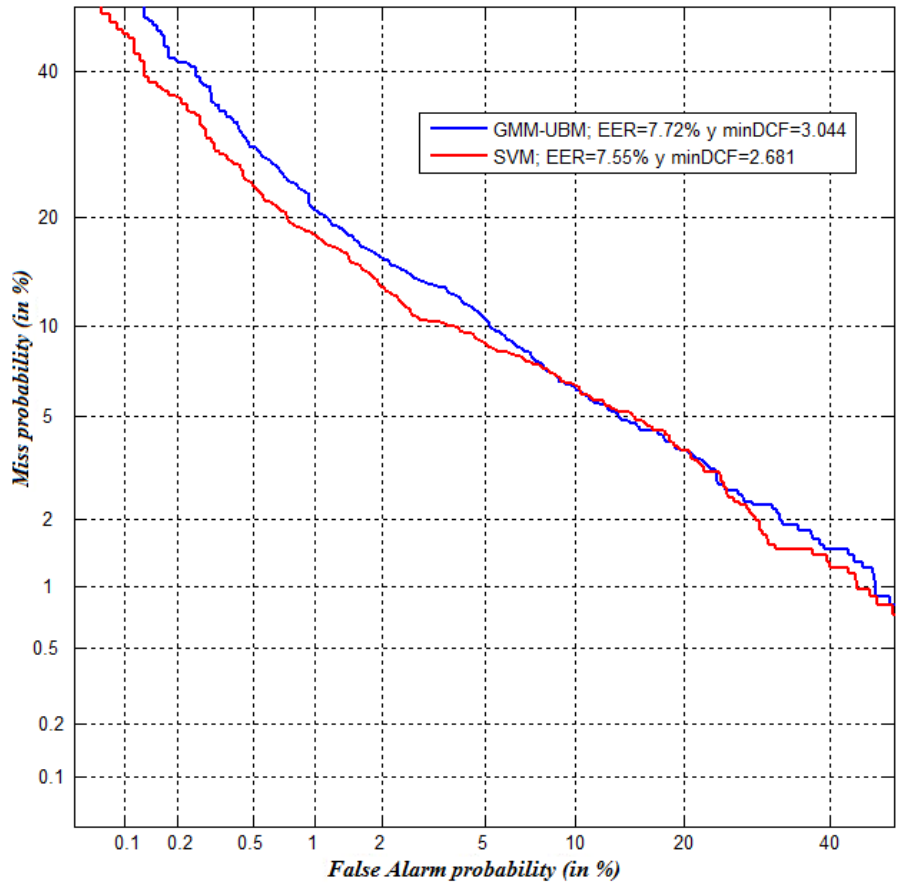

Figura 48. *Rendimiento de GMM-UBM y SVM en modo MeanSAD, con $\alpha_{SAD} = -0.3$, M=512, D=50 y con T-norm.*

Vemos como SVM mejora ligeramente sobre GMM-UBM en cuanto al minDCF y a la EER. Salvo en una pequeña zona en la que las dos técnicas se solapan, la curva DET correspondiente a SVM siempre se encuentra por debajo de la curva asociada a GMM-UBM. Aunque la mejora en el rendimiento tampoco es muy significativa.

7.3 NAP

En esta sección trataremos de analizar los resultados experimentales obtenidos con esta técnica y compararlos con las dos anteriores. En el apartado 4.4, donde se describió de manera precisa todo el funcionamiento de NAP, se hizo especial hincapié en remarcar el hecho de que el subespacio vectorial donde el canal afectaba de manera más notable y que estaba caracterizado por una base vectorial de K autovectores era de pocas dimensiones, con lo que $K \ll MD$. La primera prueba experimental con NAP consistirá en averiguar el valor de K que produce el resultado más óptimo. Generalmente, se suelen trabajar con valores de K inferiores a 100, con lo cual se ha probado con la posibilidad de $K = 1$ hasta $K = 100$. A continuación se muestra la tabla de resultados junto con sus correspondientes figuras:

K	EER (%)	minDCF ($\times$ 100)
1	6.66	2.342
10	5.52	1.813
20	5.12	1.717
30	4.87	1.572
40	4.87	1.596
50	4.71	1.605
60	4.71	1.572
70	4.96	1.587
80	4.87	1.541
90	4.79	1.572
100	4.87	1.572

Tabla 8. *Valores obtenidos para el minDCF y EER mediante NAP en modo MeanSAD, con $\alpha_{SAD} = -0.3$, M=512, D=50 y con T-norm.*

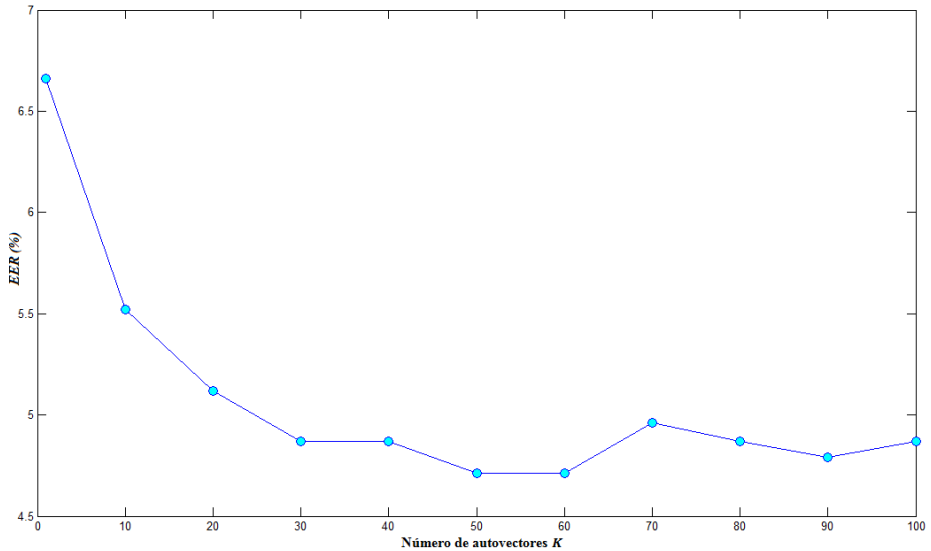

Figura 49. *Valores obtenidos para la EER mediante NAP en modo MeanSAD, con $\alpha_{SAD} = -0.3$, M=512, D=50 y con T-norm.*

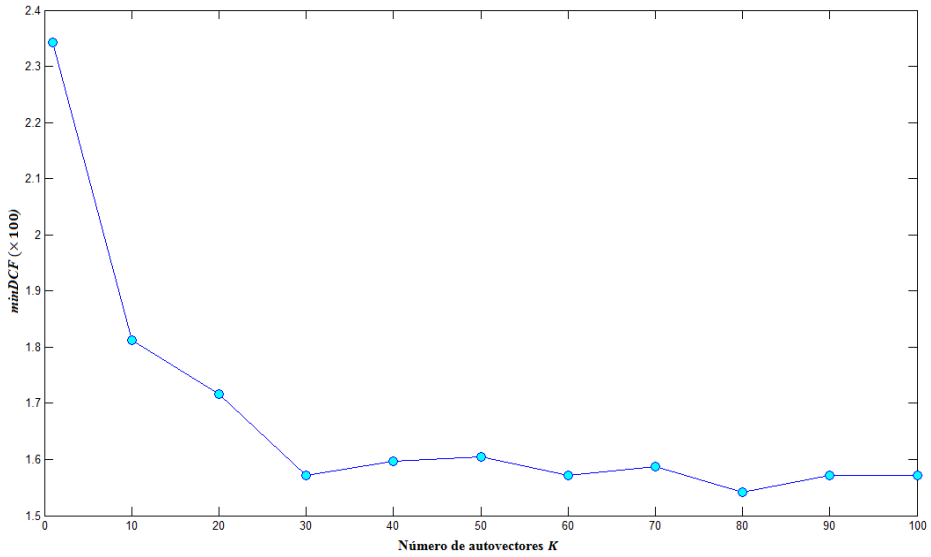

Figura 50. *Valores obtenidos para el minDCF mediante NAP en modo MeanSAD, con $\alpha_{SAD} = -0.3$, M=512, D=50 y con T-norm.*

Analizando los resultados, concluimos que de acuerdo con el criterio de la EER el número óptimo de autovectores se da para K=50 o K=60, mientras que en función del minDCF, el valor más eficiente se da para 80 autovectores. Priorizando con el criterio del minDCF, en NAP se optará por trabajar con esta última opción. De esta forma, ya estamos en completa disposición para comparar NAP con las dos técnicas anteriores. Dicha comparación se ha llevado a cabo en la figura 51.

A la luz de los resultados, se aprecia una notable mejora en NAP en relación con las dos técnicas anteriores (tanto a nivel de EER como de minDCF). Sin embargo, esta mejora no se produce en la parte derecha de la curva DET (alta probabilidad de falsa alarma y pequeña de rechazo). Aunque tal y como se explicó detalladamente en la introducción de este proyecto, a esta zona en verificación no se le presta especial importancia, ya que es muy difícil encontrar un sistema real en donde se exija o se espere una probabilidad de falsa alarma mayor que la de rechazo. Además, el NIST opera en el extremo superior izquierdo de la curva DET mientras que la EER se sitúa en la zona central. Aunque desde un punto de vista completamente objetivo, sí habría que señalar que a pesar de la clara mejora de NAP en la zona superior izquierda y la zona central de la curva DEC con respecto a las otras dos técnicas, si es cierto que su rendimiento disminuye notablemente conforme nos vamos desplazando hacia la derecha en la curva DEC, siendo superado incluso por GMM-UBM y SVM.

Si SVM y NAP son muy similares salvo en el detalle de la previa transformación de los supervectores, ¿por qué esa mejora sustancial con esta última técnica? No se debe olvidar el detalle fundamental de que NAP está trabajando con datos adicionales correspondientes a las grabaciones de locutores en diferentes condiciones que se utiliza para crear la matriz de proyección característica. Con lo cual a través de los resultados

experimentales se puede apreciar la importancia de incluir nuevos datos, los cuales además nos permiten mitigar el efecto negativo de las condiciones de canal.

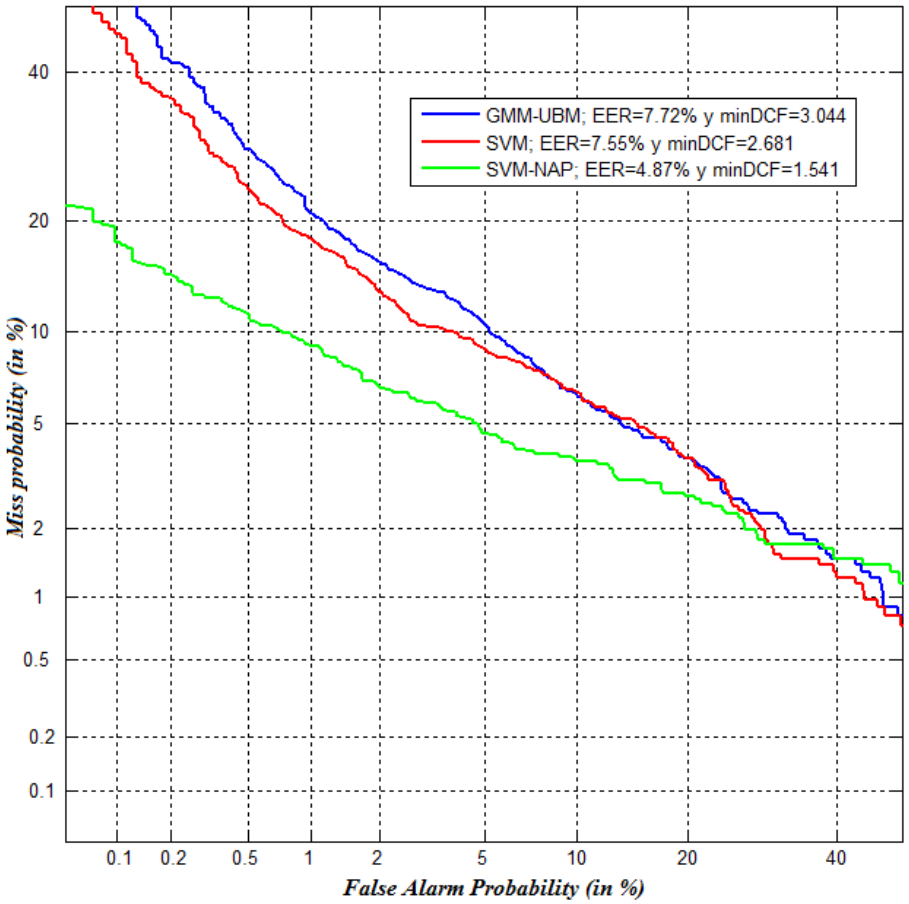

Figura 51. *Rendimiento de NAP (K=80), SVM y GMM-UBM en modo MeanSAD, con* $\alpha_{SAD} = -0.3$, *M=512, D=50 y con T-norm.*

7.4 FA

Al igual que en NAP se analizó la eficiencia del sistema en función del número de autovectores K, o lo que es lo mismo, el número de columnas de la matriz $\boldsymbol{S}$, para Factor Analysis se evaluará la bondad del sistema en función del número de columnas de la matriz de canal $\boldsymbol{U}$ que también ha sido identificado en este proyecto mediante la nomenclatura K. Si bien las matrices $\boldsymbol{S}$ y $\boldsymbol{U}$ no son las mimas, si es cierto que ambas representan al subespacio vectorial en el que el canal está afectando en mayor medida. En el apartado de teoría se explicó que en Factor Analyis se pueden puntuar tanto con GMM (FA-GMM-UBM) como con SVM (FA-SVM), por lo que se tendrán en cuenta ambas posibilidades en este apartado de simulaciones.

7.4.1 FA-GMM-UBM

Variando el rango de la matriz $\boldsymbol{U}$ o lo que es lo mismo, modificando el parámetro K, los resultados obtenidos con FA puntuando con la opción de GMM-UMB aparecen recogidos en la siguiente tabla:

K	EER (%)	minDCF (× 100)
1	7.07	2.824
10	5.12	2.139
20	4.31	1.838
30	4.63	1.741
40	4.47	1.70
50	4.06	1.645
60	3.90	1.654
70	4.06	1.660
80	4.06	1.645
90	4.06	1.604
100	3.98	1.652

Tabla 9. *Valores obtenidos para el minDCF y EER mediante FA-GMM-UBM en modo MeanSAD, con* $\alpha_{SAD} = -0.3$*, M=512, D=50 y con T-norm.*

7.4.2 FA-SVM

Los resultados obtenidos de las pruebas experimentales de Factor Analysis puntuando con SVM se muestran a continuación:

K	EER (%)	minDCF (× 100)
1	6.42	2.327
10	5.28	1.831
20	4.71	1.678
30	4.71	1.587
40	4.87	1.562
50	4.55	1.529
60	4.39	1.513
70	4.55	1.513
80	4.55	1.521
90	4.47	1.545
100	4.66	1.531

Tabla 10. *Valores obtenidos para el minDCF y EER mediante FA-SVM en modo MeanSAD, con* $\alpha_{SAD} = -0.3$*, M=512, D=50 y con T-norm.*

Agrupando los resultados obtenidos con las dos variantes permitidas en FA, se han representado las dos figuras siguientes, en donde se pueden comparar y diferenciar el rendimiento del sistema entre las dos técnicas de scoring para FA:

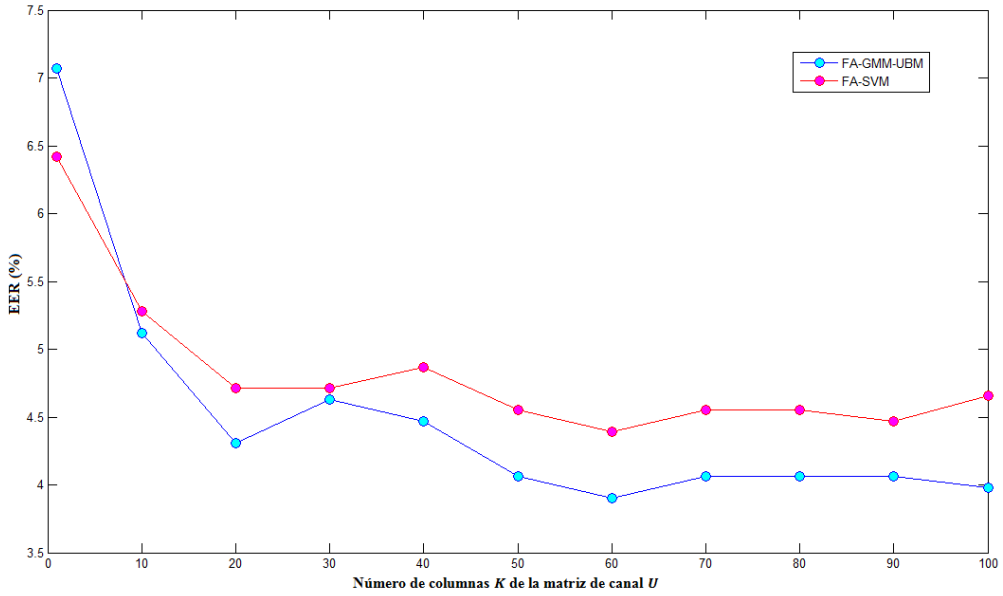

Figura 52. *Valores obtenidos para la EER mediante FA en modo MeanSAD, $\alpha_{SAD} = -0.3$, M=512, D=50 y con T-norm.*

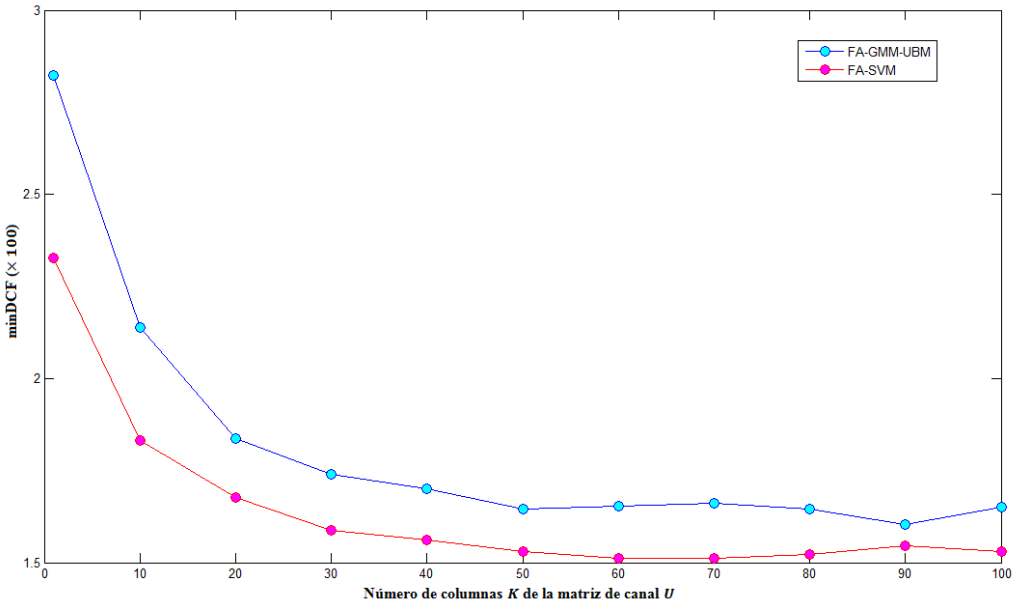

Figura 53. *Valores obtenidos para el minDCF mediante FA en modo MeanSAD, $\alpha_{SAD} = -0.3$, M=512, D=50 y con T-norm.*

Tras visualizar el resultado de las pruebas experimentales y atendiendo con mayor prioridad al criterio del NIST (minDCF), para FA-GMM-UBM se escogerá la opción de matriz de canal con 90 columnas. Para FA-SVM en principio podríamos seleccionar la opción de 60 o 70 columnas indistintamente, pues ambas generan el mismo valor de minDCF. Debido a ello, se ha hecho uso de la EER como juez en dicha elección, llevándonos a escoger la opción de K=60. Observando las gráficas 52 y 53 resulta especialmente llamativo como el score con GMM-UBM produce mejores resultados si atendemos a la EER, mientras que en el caso del minDCF es SVM la opción de puntuación que proporciona mayor eficiencia. Esto ratifica el argumento que se explicó al comienzo de este proyecto cuando se establecía la idea de que el minDCF y la EER no guardaban ningún tipo de relación líneal.

Así que ya podemos mostrar las curvas DEC de todas las técnicas descritas e implementadas en este proyecto, en donde se podrán apreciar las diferencias en cuanto a EER, minDCF y curvas DEC.

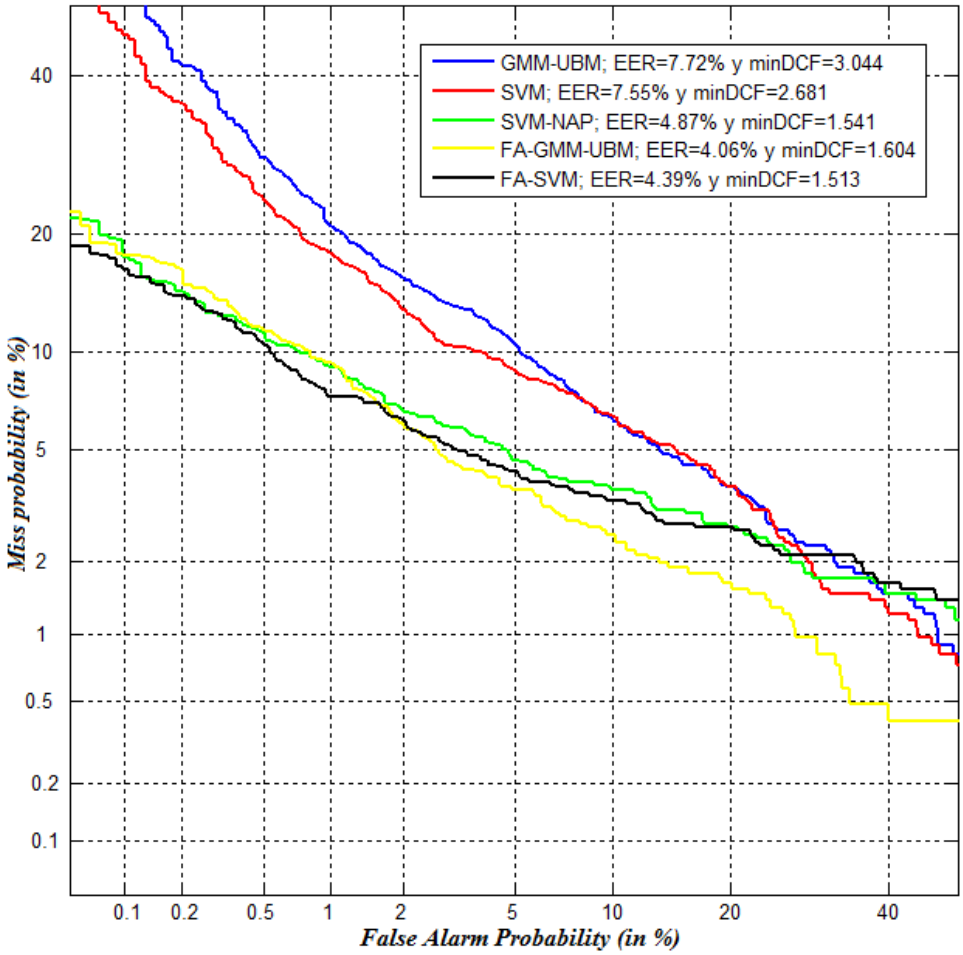

Figura 54. *Comparación de los rendimientos de GMM-UBM, SVM, NAP (K=80), FA-GMM-UBM (K=90) y FA-SVM (K=60) en modo MeanSAD,* $\alpha_{SAD} = -0.3$*, M=512, D=50 y con T-norm.*

A la luz de los resultados, se puede concluir afirmando que la técnica ganadora con respecto al NIST es FA-SVM, con un minDCF(×100) correspondiente a 1.513. De acuerdo con la EER, es FA-GMM-UBM la que produce un resultado más eficiente con una EER igual a 4.06%. Si analizamos en detalle las curvas DEC, podemos ver como SVM-NAP y FA-SVM, a pesar de la notable mejora con respecto a las dos primeras técnicas (GMM-UBM y SVM), sufren una degradación importante conforme nos desplazamos hacía la derecha en la gráfica. En cambio, FA-GMM-UBM se mantiene constante con excelentes resultados independientemente de la zona de la curva DEC en la que se encuentre. Así que a pesar de que en una hipotética evaluación del NIST FA-SVM habría sido considerada la técnica más óptima, siendo puramente objetivo habría que argumentar que la técnica que produce resultados más eficientes para cualquier punto de trabajo en la base de datos del NIST de 2005 sería FA-GMM-UBM.

7.5 Fusión de sistemas

La fusión de sistemas es un concepto ampliamente empleado y muy eficiente en verificación automática de locutor. La fusión entre distintas técnicas consiste en ponderar sus correspondientes puntuaciones, generando un nuevo fichero de scores al cual se le puede calcular de nuevo la EER, minDCF y su correspondiente curva DET. Se ha decidido abordar esta opción en este proyecto para ver sí resulta posible mejorar los resultados previamente generados. Generalmente la fusión se suele llevar a cabo entre técnicas o sistemas que presentan un rendimiento similar. Es por ello que atendiendo a la figura 54, y especialmente en la zona izquierda (minDCF) y central (EER) de la curva DET, los cinco resultados obtenidos en este proyecto se podrían clasificar en dos grupos distintos: por una parte tenemos a GMM-UBM junto a SVM como los dos sistemas con rendimiento más bajo y resultados muy similares entre ellos, y por otra parte a NAP y FA (con sus dos variantes) como el otro grupo de técnicas que presentan mejor rendimiento y con valores asociados a la EER y minDCF no muy dispersos entre ellas.

Para la fusión, generalmente se utilizan algoritmos que intentar obtener los pesos de ponderación adecuados de tal manera que el nuevo sistema fusionado resulte lo más óptimo posible. Sin embargo, para evitar mayores complicaciones, en este trabajo se ha tratado en todo momento con pesos equitativos. Así que para el primer sistema de fusión entre GMM-UBM (50%) y SVM (50%) se muestra a continuación su correspondiente curva DET, donde además se puede comparar los resultados de las dos técnicas sin fusionar.

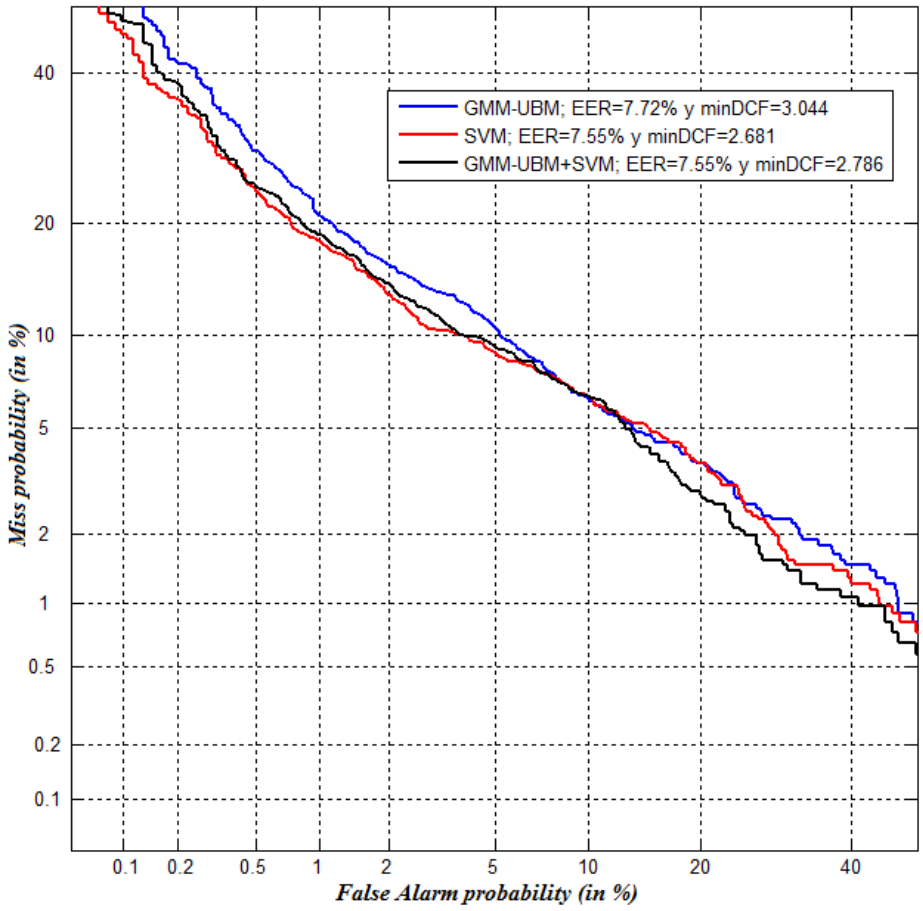

Figura 55. *Fusión entre GMM-UBM y SVM en modo MeanSAD, con* $\alpha_{SAD} = -0.3$*, M=512, D=50 y con T-norm.*

Lamentablemente, la fusión no produce mejores resultados. La EER resultante de la fusión es similar a SVM mientras que el minDCF resulta mejor comparándolo con GMM-UBM pero peor que con SVM. Únicamente se aprecia una notable mejoría en la parte derecha de la curva, en donde se observa como la curva correspondiente a la fusión (línea negra) siempre se encuentra por debajo que las otras dos curvas (azul y roja) en esa zona.

En la figura 56 se ha reflejado la curva DET correspondiente a la fusión entre NAP y FA (con las dos variantes de puntuación). De nuevo, se han establecido pesos equitativos (33% para cada una de las 3 técnicas). A diferencia del caso anterior, aquí sí que hay una mejora en el minDCF (1.482), resultando por lo tanto esta fusión en el sistema más eficiente del proyecto teniendo en cuenta la metodología de evaluación del NIST. Con respecto a la EER, la fusión mejora a SVM-NAP y FA-SVM pero se ve superado por FA-GMM-UBM. Además, el sistema fusionado conserva la ventaja principal de esta última, que como dijimos consistía en la constancia en el alto rendimiento a lo largo de toda la curva DET, característica de la que carece tanto SVM-NAP como FA-SVM.

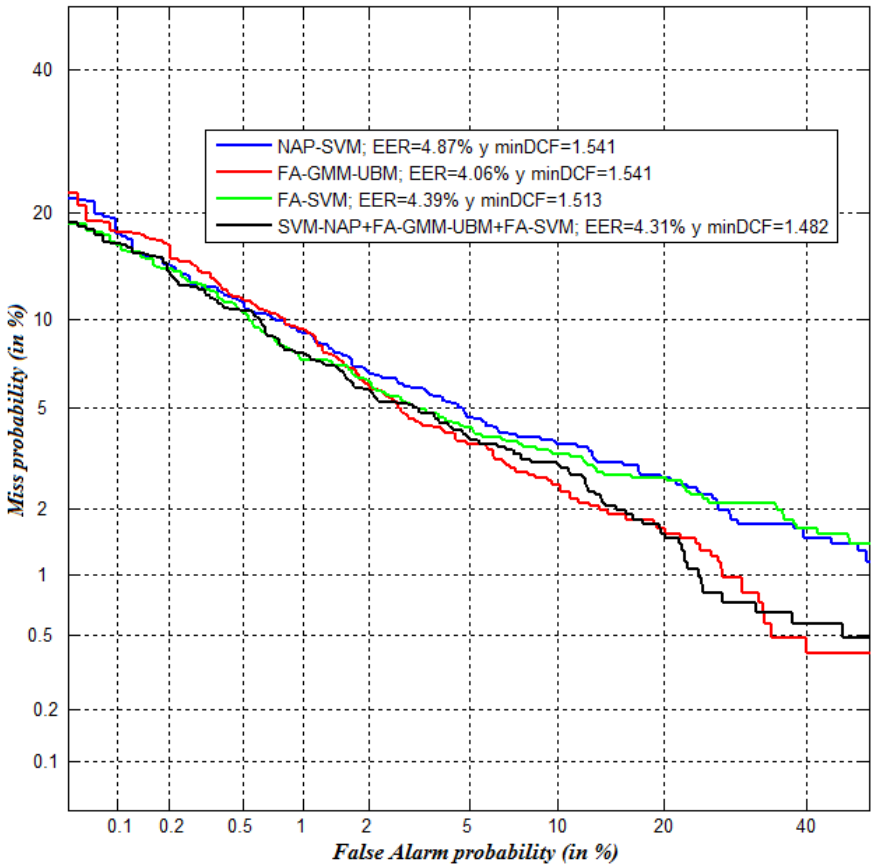

Figura 56. *Fusión entre NAP y FA en modo MeanSAD, con* $\alpha_{SAD} = -0.3$*, M=512, D=50 y con T-norm.*

A lo largo de este proyecto, además del minDCF, también se le ha ido dando especial relevancia a la EER. Así que una manera práctica de ilustrar el rendimiento de todas las técnicas junto con las dos fusiones teniendo en cuenta estos dos parámetros podría ser mediante una gráfica bidimensional con un eje representando el minDCF y el otro eje la EER. Esta opción se ha llevado a cabo en la figura 57, en donde se han representado todos los resultados obtenidos en este proyecto. De una manera completamente objetiva y otorgando la misma importancia tanto a la EER como al minDCF, el mejor sistema sería aquel cuyo punto se encuentra más cercano al origen en la figura siguiente. Analizando los resultados se concluye que atendiendo a este nuevo criterio, el mejor sistema es FA-GMM-UBM, seguido a continuación de la fusión entre FA y NAP.

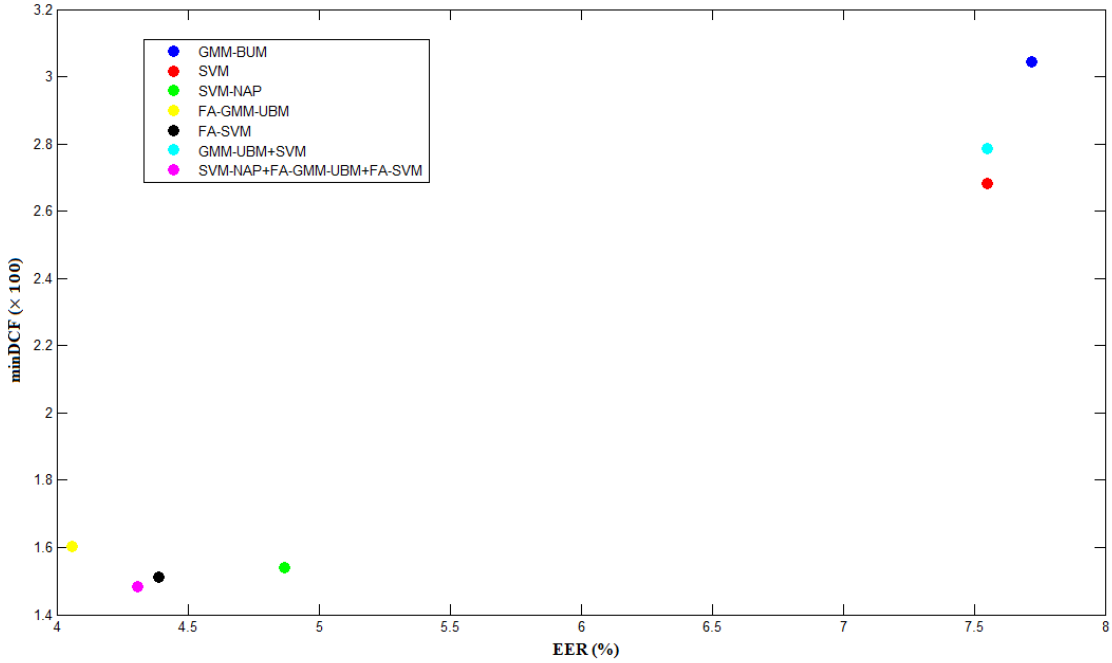

Figura 57. *Comparación de todas las técnicas atendiendo por igual al minDCF y EER en modo MeanSAD, con* $\alpha_{SAD} = -0.3$*, M=512, D=50 y con T-norm.*