El elevado coste computacional del algoritmo que realiza el filtrado morfológico hace que éste sea inapropiado para multitud de aplicaciones relacionadas con el reconocimiento automático de voz, que muchas veces deben trabajar en tiempo real como las que se utilizan en la tecnología móvil. Por ello se motiva la reducción del tiempo de ejecución tratando de mantener resultados competitivos para hacer el algoritmo válido en todo tipo de contextos. Así se propondrá en primer lugar una forma de reducir las iteraciones basada en una estimación de la SNR y de las condiciones últimas óptimas obtenidas en el Capítulo 3. A continuación se propondrá una alternativa al costoso proceso de la propagación y finalmente se explicará una forma de definir las condiciones finales basada en el mismo principio que el algoritmo original pero llevado a cabo de una manera mucho más eficaz. Todas las modificaciones y las posibles combinaciones entre ellas serán evaluadas a través de experimentos que medirán tanto el tiempo de ejecución como los porcentajes de reconocimiento para compararlos con el algoritmo original.

Capítulo 6

REDUCCIÓN DE LA CARGA COMPUTACIONAL

6.1 ESTIMACIÓN DE SNR Y DEFINCIÓN DE CONDICIONES FINALES

El filtrado morfológico es un proceso iterativo, que busca regiones en el espectrograma dominadas por señal vocal mediante procedimientos de crecimiento morfológico. El proceso es iterativo porque se basa en la estimación de la PDF del ruido que esta presente junto con la señal vocal, que va determinando de manera cada vez más precisa en cada iteración. En cada iteración se estima dicha PDF a partir de los puntos del espectrograma que no han sido segmentados hasta ese momento, definiéndose a partir de dicha PDF un nuevo límite de propagación que discrimine las zonas dominadas por ruido de las zonas donde existe señal vocal. El proceso itera hasta que toda la zona de señal ha sido extraída, de manera que la estimación de la PDF alcanza un nivel de precisión adecuado y ya no varía. Esta PDF de precisión define un último límite de propagación que establece el área que se va a segmentar finalmente.

El proceso de optimización del apartado 4.3 establece, para cada SNR, los últimos límites de propagación que son óptimos desde el punto de vista del comportamiento del reconocedor. De esta forma, sabemos experimentalmente el límite de propagación medio que debemos ajustar para obtener resultados óptimos, de manera que si lo ajustamos manualmente al comienzo del proceso sólo necesitamos una iteración para llevar a cabo la segmentación completa. El único problema es que estos límites de propagación son distintos para cada SNR en cuestión, de forma que necesitamos ese conocimiento previo sobre la señal para poder realizar el ajuste manual.

Por ello, se desarrolla en este capítulo una manera de estimar la relación señal a ruido basada también en el filtrado morfológico, que dará lugar a un algoritmo independiente de la señal de entrada que hace que el proceso vaya directamente a la última de la iteraciones en las condiciones óptimas obtenidas experimentalmente, reduciendo considerablemente la carga computacional.

6.1.1 Estimación de la SNR

Cuando el filtrado morfológico se aplica a un espectrograma de una señal vocal inmersa en ruido, el resultado de la segmentación depende obviamente de la cantidad de ruido presente en la muestra. A más ruido presente más se enmascaran las características deterministas de la señal y, por consiguiente, el algoritmo detecta una menor área de señal. Por lo tanto, existe una relación directa entre la cantidad de área segmentada por el algoritmo y la relación señal a ruido de la señal. Como muestra de este hecho se muestra en la siguiente figura cómo el área segmentada de una señal sintética va aumentando a medida que aumenta su potencia y por lo tanto la relación señal a ruido, ya que el ruido se mantiene a un nivel constante.

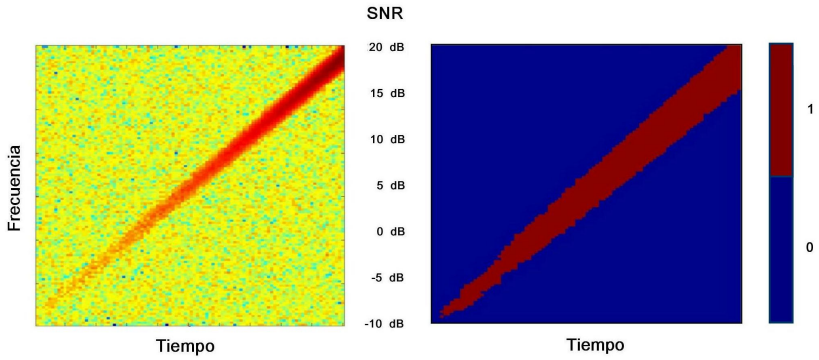


Figura 6.1 Dependencia del área segmentada con la relación señal a ruido

De esta forma, si medimos el área que se segmenta en la primera de las iteraciones podremos estimar a partir dicha medición la relación señal a ruido que existe en el archivo concreto. Esto llevará a un algoritmo que sólo tiene que realizar dos iteraciones para realizar la segmentación completa en lugar de las 4 de media que necesita el algoritmo original. La primera de las iteraciones servirá para determinar la SNR de la muestra, con lo que se define el límite final de propagación a partir de los datos experimentales obtenidos en el apartado 4.3; La segunda iteración parte de las condiciones finales para el Área de trabajo y la Región de confianza de ruido por lo que se obtiene toda el área final de segmentación en una sola propagación.

La medida que realizamos no será la SNR real, pero nos servirá para discriminar entre las diferentes SNR existentes y poder así definir los distintos límites de propagación finales. La medida consistirá en calcular la relación existente entre la media de los píxeles del espectrograma que son segmentados en la primera iteración y la media de los que no. Detalladamente la medida tiene la siguiente forma:

$$\overset{\Delta}{snr} = 20\log\frac{\frac{1}{N_1}\sum_{i=1}^{N_1} S_i}{\frac{1}{N_2}\sum_{i=1}^{N_2} R_i}$$

S_i son las amplitudes de los píxeles del espectrograma que han sido clasificados como señal determinista y N_1 es el número total de estos píxeles. R_i serán las amplitudes de los píxeles no clasificados y N_2 el número total de este tipo de píxeles.

Para determinar la utilidad de esa medida se realiza un experimento que obtiene el valor de la estimación para 125 muestras de cada uno de los grupos de SNR bajo estudio. La tabla que tenemos a continuación muestra los resultados de dicho experimento, la primera columna muestra el valor máximo para la estimación, la segunda el mínimo y la tercera la media de los valores obtenidos para cada uno de los grupos de archivos pertenecientes a una determinada SNR.

	Máximo	Mínimo	Media
Limpio	40.23	23.93	29.56
20dB	23.02	19.32	20.59
15dB	19.60	16.08	17.37
10dB	16.01	13.18	14.26
5dB	13.37	10.21	11.51
0dB	11.95	8.17	9.29
-5dB	9.91	6.37	7.59

Tabla 6.1 Resultados de la medida de SNR

Se observa que los valores no se corresponden con los valores reales de SNR, pero se comprueba que la medida es dependiente de la relación señal a ruido de manera que los valores obtenidos tienen una media distinta para cada uno de los grupos de SNR. Así, esta medida servirá para discriminar entre los diferentes grupos de SNR ya que si por ejemplo un archivo nos da como resultado un valor entre 16.08 y 19.60 lo clasificaremos como perteneciente al grupo de 15dB. Existe un cierto solape entre los rangos que obtenemos de las medidas para las distintas SNR, sobre todo en la zona donde ésta no es muy alta, por lo que en la clasificación se cometerá cierto error. Eligiendo los siguientes rangos para discriminar entre los diferentes grupos de SNR el error de clasificación resulta en un mínimo de 2.85% en media.

			Errores de clasificación
$23.50 \leq \infty$	$\longrightarrow$	Limpio	0%
$19.00 \leq 23.50$	$\longrightarrow$	20dB	0%
$16.05 \leq 19.00$	$\longrightarrow$	15dB	1.6%
$13.20 \leq 16.05$	$\longrightarrow$	10dB	0.8%
$10.30 \leq 13.20$	$\longrightarrow$	5dB	2.4%
$8.30 \leq 10.30$	$\longrightarrow$	0dB	5.6%
$-\infty \leq 8.30$	$\longrightarrow$	-5dB	9.6%

6.1.2 Definición de condiciones finales

Una vez realizada la estimación de la SNR tenemos el archivo etiquetado como perteneciente a una SNR determinada, por lo que a partir de tener esa información podemos ajustar el límite de propagación óptimo para cada SNR que se derivó en el apartado 4.3. De esta manera ahorraremos todo el proceso iterativo de búsqueda de este límite y los requerimientos computacionales del proceso se reducirán considerablemente.

A continuación se presentan dos formas de establecer dichos límites finales con sus respectivos resultados en exactitud de reconocimiento de palabra y el ahorro computacional que suponen. El primero de ellos sólo es válido para los rangos de SNR que normalmente tenemos bajo estudio y sirve para realizar la comprobación de la validez del método. El segundo de ellos extiende la definición de los límites a toda posible SNR por lo que es mucho más adecuado para utilizarlo en condiciones reales.

6.1.2.1 Definición de límites discreta

El proceso de estimación de la SNR explicado en el apartado anterior da como resultado unos rangos para el valor de la estimación que nos indican la SNR a la que pertenece el archivo, con un cierto grado de error por supuesto. Así, al realizar la estimación de la SNR obtendremos un valor que pertenecerá a uno de los 7 rangos definidos, que se corresponden con una de las 7 SNR reales bajo estudio para las que se ha derivado un límite de propagación óptimo. Dependiendo entonces del rango en el que se encuentre la estimación definiremos un límite de propagación concreto y se llevará a cabo la propagación en las condiciones últimas óptimas particulares para esa SNR. La función que se utiliza entonces para relacionar la estimación de la SNR con el ratio que debemos ajustar es la siguiente:

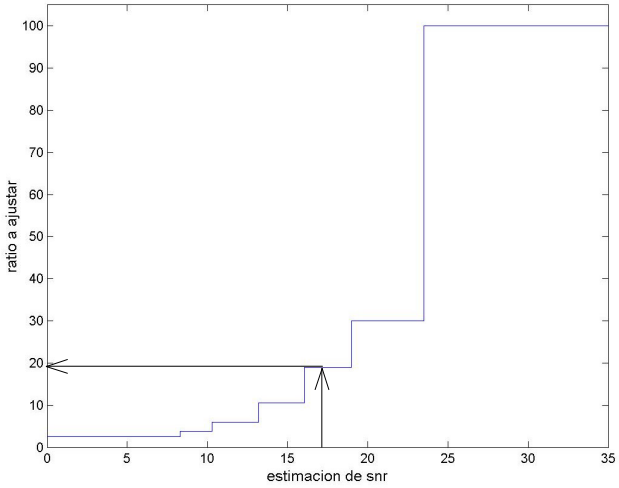


Figura 6.2 Función de relación entre la estimación de la SNR y el ratio a ajustar

Si por ejemplo el valor de la estimación nos da 17.4, los rangos establecidos en el apartado 6.1.1 nos determinan que la muestra debe ser clasificada como perteneciente al grupo de 15dB por lo que se ajustará ratio = 19 ya que es ratio óptimo para esta SNR que se derivó en el apartado de optimización 4.3. Para comprobar el comportamiento de este algoritmo modificado que estima la SNR a partir del área segmentada en la primera iteración y salta directamente a la última mediante un ajuste manual del último límite de propagación, se lleva a cabo el experimento estándar filtrando los 7 grupos de SNR cada uno de 1001 archivos. Durante el proceso se registra el tiempo de computación para ver la reducción existente respecto al algoritmo original. Aunque el comportamiento es dependiente del procesador, la comparación entre este algoritmo y el original es completamente válida ya que para medir los tiempos del original se utilizó el mismo procesador. Los archivos filtrados se envían al reconocedor entrenado con el entrenamiento óptimo (Pe =10%)

	limpio	**20dB**	**15dB**	**10dB**	**5dB**	**0dB**	**-5dB**
ORIGINAL	95.3	94.8	93.3	88.6	78.1	57.3	29.5
Estimación SNR	91.3	86.4	84.9	73.1	58.2	36.3	13.0

Tabla 6.2 Comparación de resultados en reconocimiento de palabra entre el algoritmo original y el modificado estimando la SNR

El algoritmo modificado consigue una reducción del tiempo de procesamiento del 35.7% ya que solo necesita ejecutar dos iteraciones del algoritmo, una para estimar la SNR al comienzo del proceso y la siguiente, más costosa, para realizar la propagación de toda el área a partir del límite de propagación ajustado de manera manual. Los resultados en reconocimiento automático muestran una reducción de la exactitud respecto al algoritmo original del 13.25% de media, siendo más acusada dicha reducción hacia valores menores de la SNR debido al mayor número de errores de clasificación. Además no hay que olvidar que los límites ajustados manualmente son definidos mediante ratios medios para todos los archivos, lo que también degrada el comportamiento como se explica en el Capítulo 4.

6.1.2.2 Definición de límites continua

El método anterior sólo es válido para los 7 grupos de SNR concretos que definen en el protocolo de pruebas Aurora 2 por lo que no es válido para aplicarlo a un archivo con cualquier SNR, lo que resta generalidad al algoritmo. Además se define un mismo ratio medio para todos los archivos clasificados como pertenecientes a uno de los grupos de SNR que contempla y esto reduce el comportamiento del algoritmo como se demuestra en el apartado 4.3. El algoritmo original, tras el proceso iterativo, ajusta al final un límite de propagación particular para cada archivo que depende de las características y de la potencia de ruido concreta que exista en el mismo. Así el algoritmo original actúa de manera particular con cada archivo y es completamente independiente de la SNR por lo que funciona igualmente bien para muestras fuera de los grupos típicos de prueba, como por ejemplo un archivo que tenga SNR 7.5dB.

Para extender la modificación del algoritmo que se esta llevando a cabo en este apartado y poder aplicarlo a una archivo de cualquier SNR se crea una nueva función para definir el ratio a utilizar a partir del valor de la estimación. Tomando los valores medios de cada rango definido en el apartado 6.1.1 y los ratios óptimos que debemos ajustar para cada SNR tenemos una serie de puntos que sirven para construir una función que relaciona de manera continua las estimaciones con los ratios. Utilizando un polinomio de quinto grado la función de relación resultante es la siguiente:

$$y = 9.15 \cdot 10^{-5} x^5 - 7.33 \cdot 10^{-3} x^4 + 0.227 x^3 - 3.25 x^2 + 23.3 x - 57.3$$

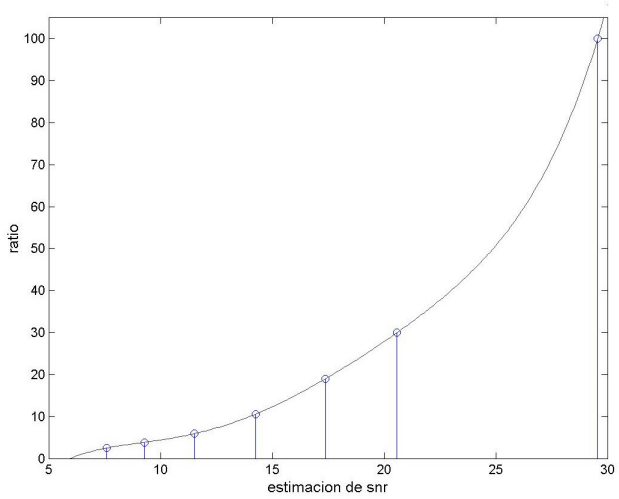


Figura 6.3 Función de relación continua entre la estimación de la SNR y el ratio a ajustar

Mediante esta nueva forma de ajustar el límite, el algoritmo modificado es completamente válido para cualquier SNR ya que simplemente trataremos de determinar la relación señal a ruido mediante nuestro estimador y a continuación utilizaremos la función de quinto grado para establecer una ratio en el proceso de segmentación de ese archivo concreto. El algoritmo resultante determina un ratio diferente para cada archivo dependiendo del valor concreto de la estimación por lo que en cierta manera se parece más al algoritmo original en la especificación de límites individualizada. Llevando a cabo el mismo experimento que en el apartado anterior, los resultados que se obtienen son los siguientes

	limpio	**20dB**	**15dB**	**10dB**	**5dB**	**0dB**	**-5dB**
ORIGINAL	95.3	94.8	93.3	88.6	78.1	57.3	29.5
Estimación SNR	91.1	87.4	82.9	70.6	57.2	36.1	13.7

Tabla 6.3 Comparación de resultados en reconocimiento de palabra entre el algoritmo original y el modificado estimando la SNR con la determinación de límites continua

La reducción del tiempo de proceso es similar al caso anterior y los resultados en exactitud del reconocimiento automático son un poco mejores aunque del mismo orden que los obtenidos con el establecimiento de ratios discreto. De esa forma la reducción del coste computacional sigue

siendo de un 37.5% y la reducción de la exactitud en el reconocimiento es de un 12% disminuyéndose un poco la tendencia al empeoramiento hacia SNR bajas. De esta forma comprobamos que la definición del ratio de manera particular para cada archivo no tiene gran trascendencia para el comportamiento del reconocedor; aunque si que es una gran ventaja desde el punto de vista de generalidad del algoritmo, ya que puede trabajar con todo tipo de niveles de ruido y no solo con los grupos definidos en el paquete Aurora 2.

6.1.3 Conclusiones

En este apartado se ha propuesto un primer método de reducir la carga computacional del algoritmo, basado en los resultados de optimización de parámetros llevada a cabo en el apartado 4.3 y de una estimación efectiva de la SNR de los archivos.

Una vez realizada la estimación de la SNR se proponen dos formas de establecer los límites de propagación, una discreta y la otra continua. El método continuo es superior en todos los aspectos al discreto ya que el porcentaje de reducción de carga computacional es el mismo, pero la reducción en exactitud de reconocimiento es ligeramente menor. Además el método continuo es válido para cualquier valor de SNR mientras que el discreto estaba limitado a los grupos definidos en Aurora 2. La reducción del coste computacional es del 37.5% y la reducción de exactitud del 12% lo que todavía da un margen de mejora considerable sobre la línea base. A continuación presentamos un gráfico donde se muestra la exactitud y el coste computacional relativo al algoritmo original de ambos métodos presentados en este apartado.

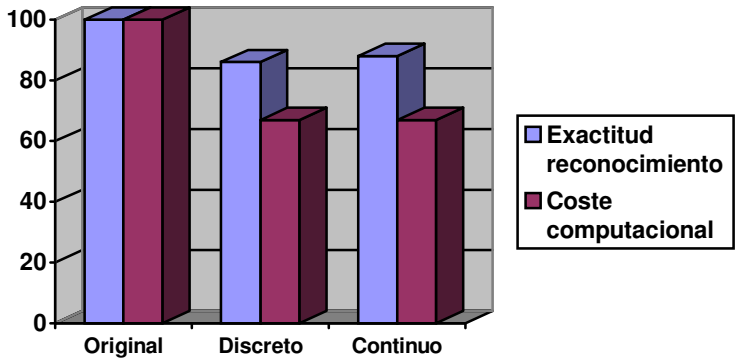


Figura 6.4 Comparación relativa de exactitud del reconocimiento y coste computacional del método de establecimiento de límites discreto y continuo frente al algoritmo original

6.2 ALTERNATIVA A LA PROPAGACIÓN

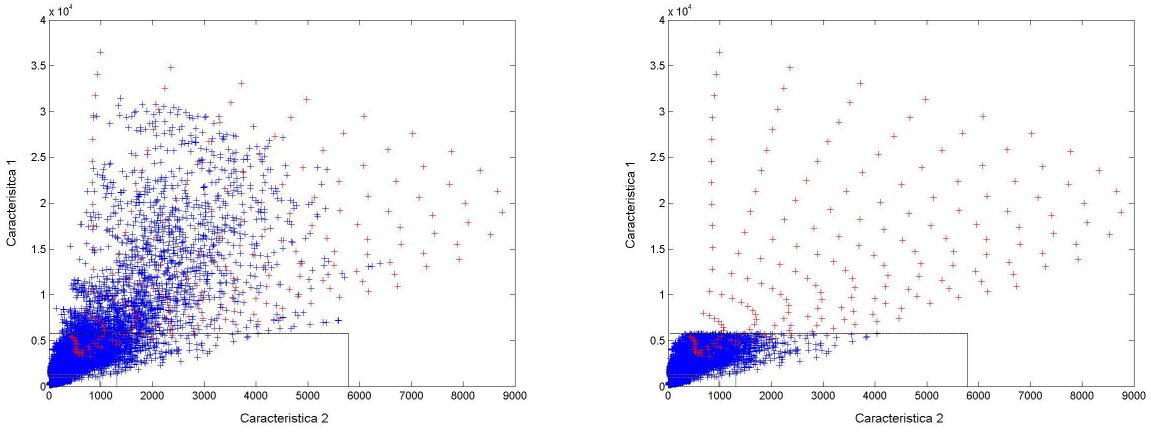
El algoritmo original lleva a cabo una serie de iteraciones en las que va segmentando área que clasifica como señal determinista a través de un algoritmo de región creciente, de forma que cuando se ha extraído toda la señal determinista la estimación de la PDF de ruido alcanza una precisión adecuada al basarse sólo en píxeles de ruido y el algoritmo detiene la extracción de señal. En cada iteración del proceso el algoritmo tiene que realizar el proceso de búsqueda y propagación de semillas que es la parte más costosa desde el punto de vista computacional.

El proceso de propagación en cada iteración comienza con la computación del Grid teórico. Una vez computado se procede a comprobar, para cada punto del Grid, si existe algún punto del Espacio Característico que se encuentre próximo, lo que supone evaluar la posición de todos los puntos del Espacio Característico. Si se encuentra algún punto, el píxel al que corresponde en el espectrograma es segmentado y a continuación se hace una comprobación de la posición en el Espacio Característico de los ocho píxeles vecinos al píxel segmentado en el espectrograma. Si alguno de estos cumple con la condición de estar en el Área de trabajo es a su vez segmentado y se realiza la misma comprobación de las vecindades. Una vez que no quedan píxeles vecinos que se encuentren en el Área de trabajo, el algoritmo cambia de punto de comprobación en el Grid teórico y se vuelven a realizar todas las comprobaciones descritas anteriormente. De esta forma, todo el proceso se realizará 225 veces, tantas como puntos forman el Grid teórico. Todo este proceso hace que el algoritmo tenga una carga computacional bastante elevada, ya que aparte de ser un proceso iterativo, cada una de las iteraciones supone un gran procesamiento.

El apartado anterior se concentró en intentar disminuir el coste computacional mediante una reducción de las iteraciones, conservando el proceso de propagación. En este apartado se propone una alternativa a la propagación de mucho menor coste computacional y que no disminuye de forma excesiva la exactitud en el reconocimiento de palabra. Así, se comprobarán las prestaciones de un algoritmo con la alternativa a la propagación pero que sigue siendo iterativo y a continuación se procederá a la eliminación de las iteraciones mediante una combinación con las modificaciones llevadas a cabo en el apartado anterior.

6.2.1 Motivación y descripción

El proceso de propagación, una vez extrae un píxel del espectrograma, elimina también el punto correspondiente a ese píxel en el Espacio Característico, de manera que a medida que se va propagando la región de señal determinista en el espectrograma, van desapareciendo los puntos correspondientes en el Espacio Característico. La extracción de estos puntos se lleva a cabo evaluando una doble condición sobre los mismos: estar cerca de algún punto del Grid teórico y encontrarse dentro del Área de trabajo. Si se observa el estado del Espacio Característico tras dar por finalizada una iteración se comprueba que han desaparecido prácticamente todos los puntos pertenecientes al Área de trabajo por lo que la condición de estar cerca de algún punto del Grid teórico es secundaria. La figura 6.5 muestra el estado del Espacio Característico antes y después de la propagación, donde se puede comprobar como queda distribuido el Espacio Característico.



a b

Figura 6.5 Estado del Espacio Característico antes y después de la propagación

En el ejemplo de la figura comprobamos cómo en este caso todos los puntos dentro del Área de trabajo han sido extraídos tras realizar la propagación de forma que las comprobaciones respecto a cercanías a puntos del Grid teórico y de vecindad en el espectrograma son redundantes, pudiendo ser eliminadas del proceso sin afectar al comportamiento final. Esto es así porque prácticamente todos los puntos dentro del Espacio Característico se encuentran en las cercanías de

algún punto del Grid teórico; y ser vecinos en el espectrograma supone compartir las mismas propiedades estadísticas y como consecuencia estar situados en la misma zona del Espacio Característico (dentro o fuera del Área de trabajo). Comprobando simplemente si los puntos del Espacio Característico están o no dentro del Área de trabajo podemos llevar a cabo prácticamente la misma extracción de señal en cada iteración de forma que llegaremos a los mismos resultados, pero de una manera notablemente más rápida. Los resultados no serán completamente exactos porque a veces, después de una iteración, si que quedan puntos dentro del Área de trabajo que no propaga el algoritmo original; pero el porcentaje de estos puntos es mínimo, de manera que no afectan al comportamiento general del algoritmo en cuanto a la definición de las correctas áreas de segmentación.

6.2.2 Experimentos y resultados

Para comprobar las prestaciones de esta nueva forma de extraer la señal sin utilizar el costoso proceso de la propagación se desarrolla una nueva versión del algoritmo. Esta modificación será un proceso iterativo en el que se va extrayendo los píxeles pertenecientes a área de señal determinista en base a una simple comprobación de si los puntos en el Espacio Característico están dentro o fuera del Área de trabajo. Al igual que el algoritmo original, sucesivas iteraciones van extrayendo más señal determinista de forma que las estimaciones de la PDF y con ello la determinación del límite de propagación se van haciendo cada vez más precisas, llegando finalmente a un nivel de precisión donde el algoritmo detiene la extracción de píxeles del espectrograma dando por concluida la segmentación.

El experimento que se realiza es el estándar de filtrado de 7 grupos de 1001 archivos pertenecientes a diferentes categorías de SNR. El tiempo que se tarda en realizar dicho proceso es registrado para poder comparar los requerimientos computacionales con los necesitados por el algoritmo original. Para obtener los resultados que alcanza este algoritmo en reconocimiento automático se envían las muestras filtradas al reconocedor entrenado con el algoritmo original optimizado (Pe = 10%). Los resultados se muestran a continuación en comparación con los alcanzados con el algoritmo original.

	limpio	20dB	15dB	10dB	5dB	0dB	-5dB
ORIGINAL	95.3	94.8	93.3	88.6	78.1	57.3	29.5
No propagación	94.8	93.8	91.1	83.6	72.4	49.1	23.8

Tabla 6.4 Comparación de resultados en reconocimiento de palabra entre el algoritmo original y el modificado evitando la propagación

En el proceso de segmentación de las muestras el algoritmo consigue una reducción del tiempo de procesamiento de nada menos que el 72.2%, lo que pone de manifiesto lo costoso que resulta el proceso de propagación ya que se comprueba que prácticamente las 3 cuartas partes del tiempo de procesamiento se dedica a las tareas de propagación de píxeles. Evitando ese proceso tenemos un algoritmo extraordinariamente más rápido y que no pierde demasiada competitividad en resultados de exactitud de reconocimiento vocal, situándose en media un 4% por debajo de los resultados del algoritmo original. Para SNR altas los resultados obtenidos son prácticamente iguales sin perder prestaciones, mientras que a medida que decrece la SNR el comportamiento ve empeorando llegando a degradaciones en las prestaciones del 8.2%.

Para relaciones señal a ruido altas, normalmente todos los píxeles dentro del Área de trabajo tienen características deterministas, ya que el ruido no tiene mucha amplitud y como consecuencia los puntos pertenecientes a ruido se encuentran en zonas de valores bajos en el Espacio Característico. De esta manera extraer todos los puntos dentro del Área de trabajo resulta acertado como finalmente demuestran los resultados experimentales. Cuando la SNR es más baja la potencia de ruido es mayor de manera que pueden existir puntos dentro del Área de trabajo que se correspondan con píxeles en zonas de ruido. Estos píxeles que son normalmente evitados por el algoritmo original mediante la comprobación de vecindad en el espectrograma, son ahora erróneamente segmentados. A pesar de ello, el nuevo algoritmo sigue dando unos resultados realmente buenos en reconocimiento vocal a la vez que reduce de manera drástica el tiempo de computación necesario.

6.2.3 Reducción de las iteraciones

El algoritmo presentado en este capítulo reduce el coste computacional evitando el largo proceso de la propagación en cada una de las iteraciones a través de una alternativa que elimina comprobaciones en la mayoría de los casos redundantes. Por el contrario, los algoritmos que se estudiaron en el apartado 6.1 buscaban la manera de reducir el número de iteraciones mediante una estimación de la SNR y una posterior definición de las condiciones de los límites de propagación en las condiciones óptimas de la última iteración. Así, el próximo paso en la reducción del coste computacional es la combinación de ambos esquemas de forma que se eviten tanto la propagación como la necesidad de todas las iteraciones.

La modificación del algoritmo para evitar las iteraciones de 6.1 se basa en la obtención anterior de las condiciones óptimas finales para los límites de propagación. Por ello el primer paso para poder reducir las iteraciones con el mismo esquema es obtener dichas condiciones óptimas para el algoritmo sin propagación.

6.2.3.1 Obtención de límites óptimos

El nuevo esquema sin propagación da unos resultados en cuanto a la extracción de señal que se va produciendo en cada iteración prácticamente iguales a los que se obtienen con el proceso de propagación original, de manera que es de esperar que las condiciones finales óptimas para el nuevo esquema no varíen demasiado respecto a las obtenidas con el proceso propagativo en el apartado 4.3. Debido a ello no se realiza el proceso de optimización completo sino que se hace el mismo experimento de optimización fina de los límites de 4.3.2.2 esperando que si los límites varían lo hagan sólo ligeramente. A continuación se presentan los resultados obtenidos con el reconocedor óptimo, para los diferentes grupos de SNR y los diferentes ratios utilizados en el ajuste manual:

Limpio	80	84	88	92	96	100	104	108	112	116
	95.02	95.41	95.89	96.14	96.59	96.93	96.41	95.99	95.58	94.33
20dB	20	22	24	26	28	30	32	34	36	38
	92.01	92.61	93.12	93.97	94.56	94.83	94.13	93.29	92.37	91.43
15dB	16	17	18	19	20	21	22	23	24	25
	88.46	89.47	89.75	89.60	88.31	87.36	86.12	85.89	85.27	84.73
10dB	9.5	10	10.5	11	11.5	12	12.5	13	13.5	14
	77.16	78.14	78.02	77.74	71.23	68.12	65.29	62.43	59.80	55.23
5dB	4	4.5	5	5.5	6	6.5	7	7.5	8	8.5
	54.29	57.33	60.94	63.67	64.65	65.05	63.81	60.27	58.40	56.31
0dB	3.2	3.4	3.6	3.8	4	4.2	4.4	4.6	4.8	5
	37.21	38.04	38.78	39.33	38.47	38.19	37.69	36.12	35.04	34.15
-5dB	1.8	1.9	2	2.1	2.2	2.3	2.4	2.5	2.6	2.7
	15.98	16.70	16.81	16.96	17.76	18.96	20.49	21.33	21.19	20.84

Tabla 6.5 Resultados de optimización de ratios para el algoritmo sin propagación

Cada fila corresponde a un grupo de SNR donde la parte de arriba en azul es el ratio utilizado y la parte de abajo el resultado obtenido para la exactitud en reconocimiento de palabra. Las cifras en rojo indican los máximos de cada conjunto de resultados de forma que, como esperábamos, los límites varían ligeramente, pero manteniéndose en el mismo rango. Por lo tanto se reafirma la decisión de utilizar P_e = 10% en el algoritmo del apartado 6.2.2. La tabla siguiente muestra los ratios a utilizar en cada SNR:

SNR	Limpio	20dB	15dB	10dB	5dB	0dB	-5dB
ratio	100	30	18	10	6.5	3.8	2.5

Tabla 6.6 Ratios a utilizar para el algoritmo sin propagación

6.2.3.2 Experimentos y resultados

El experimento que se lleva a cabo en este apartado tiene exactamente las mismas características que el del apartado 6.1.2.2. El algoritmo que se va a utilizar llevará a cabo una iteración para realizar la estimación de la SNR necesaria para poder establecer los límites finales que discriminan el Área de trabajo de la Región de confianza de ruido. Una vez hecha la estimación de la SNR se establecerá el límite apropiado utilizando la definición de límites continua, ya que frente a la discreta, se obtienen mejores resultados para la exactitud del reconocimiento de palabra y además resulta en un algoritmo válido para cualquier valor de SNR. Debido a que los ratios a definir son ligeramente diferentes la función que relaciona la estimación de SNR con los mismos también tendrá los coeficientes un poco distintos:

$$y = 2.02\cdot10^{-5}x^5 - 1.73\cdot10^{-3}x^4 + 6.33\cdot10^{-2}x^3 - 1.02x^2 + 8.28x - 23.9$$

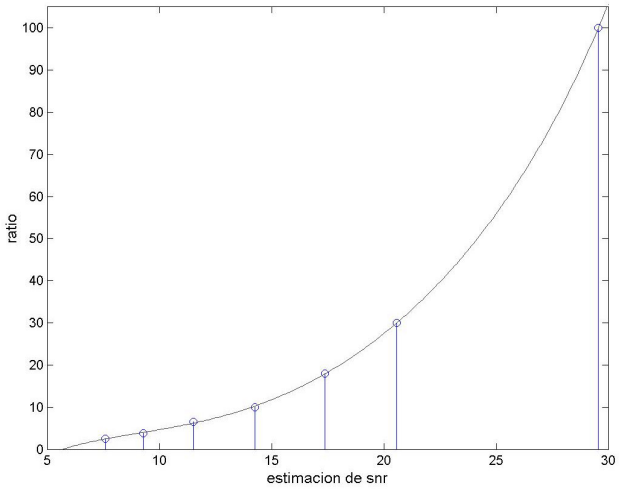


Figura 6.6 Función de relación continua entre la estimación de la SNR y el ratio a ajustar

Respecto al algoritmo en 6.1.2.2 la diferencia radica en que una vez se ha establecido con el último límite de propagación el Área de trabajo y la Región de confianza de ruido definitivas, en lugar de realizar la propagación de todas las semillas que se encuentren se segmentan directamente todos los puntos que se encuentren en el Área de trabajo por lo que los requerimientos computacionales serán bastante menores. Para evaluar las prestaciones en cuanto a exactitud de reconocimiento de voz llevamos a cabo, como siempre, la segmentación de los 7 grupos de 1001

archivos con diferente SNR y obtenemos los siguientes resultados del reconocedor entrenado de la manera óptima.

	limpio	20dB	15dB	10dB	5dB	0dB	-5dB
ORIGINAL	95.3	94.8	93.3	88.6	78.1	57.3	29.5
Estimación SNR y no propagación	90.9	88.4	84.7	70.8	56.1	30.8	12.2

Tabla 6.7 Comparación de resultados en reconocimiento de palabra entre el algoritmo original y el modificado estimando la SNR con la determinación de límites continua y sin propagación

En este caso, como cabría esperarse, es donde se consigue una mayor reducción del tiempo de computación de manera que el tiempo empleado para tratar las muestras es un 80.5% menor que el empleado por el algoritmo original. Pero los resultados en exactitud en reconocimiento de palabra caen en media un 14.3% respecto a los conseguidos con el algoritmo original. Esta degradación del comportamiento es más acusada cuanto menor es la SNR ya que en este caso se suman los problemas de los errores en la estimación de la SNR y los errores en la segmentación de píxeles de ruido que caen en el Área de trabajo para las SNR bajas.

6.2.4 Conclusiones

En este apartado se han estudiado dos nuevos algoritmos que tratan de reducir el coste computacional del filtrado morfológico. En ambos la novedad consiste en utilizar una alternativa al algoritmo de región creciente que lleva a cabo la propagación. Esta nueva forma viene motivada por una observación del estado en el que queda el Espacio Característico después de una iteración, en el que normalmente no queda ningún punto en el Área de trabajo. Así, la alternativa consiste en, una vez definido el límite que separa la Región de confianza de ruido del Área de trabajo, segmentar simplemente todos píxeles que se correspondan con puntos en dicha área. El algoritmo que utiliza

este método conservando el proceso iterativo consigue una reducción drástica del tiempo de ejecución, de manera que tarda en realizar el proceso de segmentación el 28.8% del tiempo que necesita el algoritmo original, por lo que se pone de manifiesto el gran coste que tiene el proceso de propagación. En cuanto a la exactitud del reconocimiento de palabra este nuevo enfoque sigue siendo muy competitivo situándose tan sólo un 4% en media por debajo del algoritmo original, lo que hace de esta modificación del algoritmo una opción mucho mejor cuando el tiempo de procesamiento sea un factor crítico.

El segundo de los algoritmos presentados en este apartado trata de combinar los dos enfoques de reducción del coste computacional tratados hasta ahora. Por una parte elimina el proceso de propagación como el anterior y por otra, en lugar de llevar a cabo las iteraciones del algoritmo para determinar el umbral definitivo entre el Área de trabajo y la Región de confianza de ruido, se lleva a cabo una estimación de la SNR y a partir de ella se define un límite que ha sido obtenido mediante un proceso de optimización. De esta forma se evita la necesidad de llevar a cabo un proceso iterativo de búsqueda de manera que se reduce aún más el tiempo de ejecución, llegando al 19.5% del tiempo empleado por el algoritmo original. La contrapartida es que los resultados para la exactitud en reconocimiento se degradan un 14.3% respecto al original, de manera que la utilización de este algoritmo puede no ser tan conveniente si se compara con otras alternativas de reducción del coste computacional manteniendo resultados competitivos, como es el caso del primer algoritmo de este apartado. A continuación presentamos un gráfico donde se muestra la exactitud y el coste computacional relativo al algoritmo original de ambos métodos presentados en este apartado.

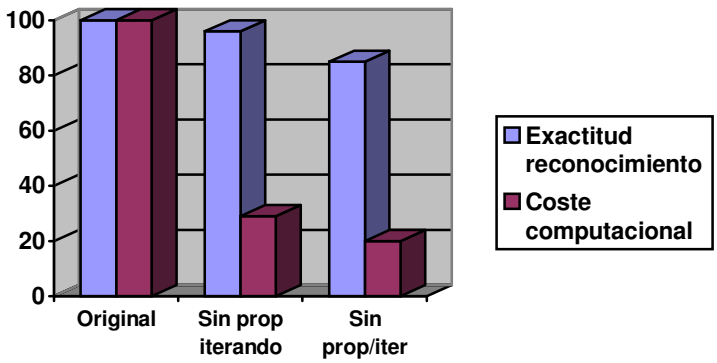
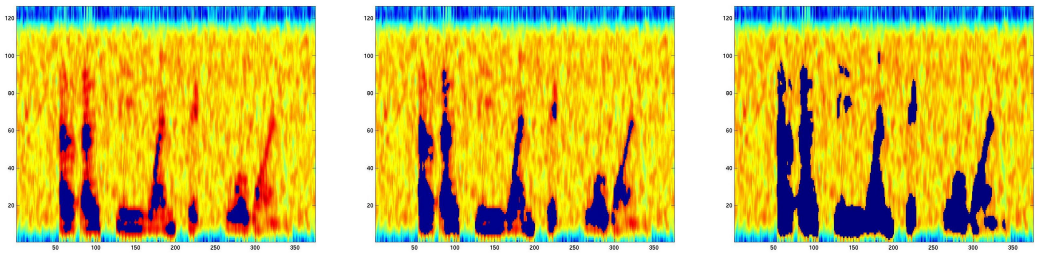


Figura 6.7 Comparación relativa de exactitud del reconocimiento y coste computacional frente al algoritmo original del método que evita la propagación conservando las iteraciones y el método que elimina tanto la propagación como el proceso iterativo

6.3 ESTIMACIÓN DIRECTA DEL LÍMITE

En los apartados anteriores, la definición de las condiciones finales para evitar el proceso iterativo se ha basado en el anterior proceso de optimización de los ratios medios para determinar los límites de propagación que se llevó a cabo en al apartado 4.3. Además dichos ratios obtenidos experimentalmente son dependientes de la SNR de la muestra, por lo que para poder aplicarlos es necesaria una anterior estimación de la SNR basada en el área segmentada por el filtrado morfológico en la primera de las iteraciones. Esto hace que el comportamiento de los algoritmos resultantes empeore de manera apreciable respecto al algoritmo general en exactitud del reconocimiento de palabra; ya que la definición de un ratio para obtener el límite de propagación es un método aproximado que no da los mejores resultados y además hay que sumar la degradación producida por los errores cometidos en la estimación de la SNR Asimismo, los algoritmos tienen que ejecutar una iteración para poder estimar la SNR y con ello definir las condiciones de la última iteración por lo que finalmente el proceso consta de dos iteraciones y la reducción del tiempo de computación no es tan notable, si se compara por ejemplo con la reducción conseguida al eliminar el proceso propagativo.

En el algoritmo original, el proceso iterativo se utiliza para encontrar de manera precisa el último límite de propagación que define un umbral efectivo entre el Área de trabajo donde se lleva a cabo la propagación de píxeles y la Zona de confianza de ruido donde suponemos que todos los píxeles son WGN. Al comienzo del proceso, el algoritmo realiza una estimación de la PDF de ruido a partir de todos los píxeles del espectrograma y a partir de ella define el límite de propagación. Este primer límite no es muy exacto ya que la estimación de la PDF de ruido se ha basado tanto en píxeles de ruido como en todos los píxeles que contienen señal determinista; pero sirve para realizar una primera extracción de señal determinista que hace que el espectrograma tenga una mayor proporción de píxeles de ruido. De esta forma, iteración tras iteración el proceso va extrayendo más y más zona de señal determinista por lo que las posteriores estimaciones de la PDF de ruido van siendo más precisas, al basarse en datos de mayor calidad para esta tarea (existe una mayor proporción de píxeles de ruido).



a b c

Figura 6.8 Proceso iterativo de extracción de señal determinista. (a) primera iteración, (b) tercera iteración y (c) sexta y última iteración.

La estimación de la última PDF de ruido y con ello el último límite de propagación se basa en un espectrograma donde teóricamente todos los píxeles pertenecen a zonas de ruido exclusivamente, por lo que se alcanza la exactitud adecuada y el límite definido determina el umbral real entre punto de ruido y de señal determinista en el Espacio Característico. Por lo tanto, lo que el algoritmo necesita para definir el último límite de propagación de forma exacta es una estimación de la PDF basada sólo en píxeles de ruido, de manera que si esto se realiza al comienzo del proceso no será necesario todo el proceso iterativo de incremento de la exactitud. Esto motiva la modificación del algoritmo que se lleva a cabo en este apartado, donde se lleva a cabo una estimación del último límite de propagación a partir de una zona exclusivamente ocupada por ruido, eliminando el proceso iterativo y reduciendo así los requerimientos computacionales del algoritmo. A continuación se presenta la manera de realizar dicha estimación y posteriormente se pone a prueba dicho método para comprobar las prestaciones del mismo.

6.3.1 Estimación del límite a partir de zona ruidosa

Para llevar a cabo una estimación eficiente de la PDF de ruido y con ello del verdadero umbral en el Espacio Característico entre puntos de señal y de ruido, sólo necesitamos una zona del espectrograma que este dominada exclusivamente por ruido, de forma que nos aporte información veraz sobre la potencia de ruido para poder derivar una PDF acorde al ruido existente en ese

espectrograma concreto. Tomando como referencia la técnica clásica de Substracción Espectral donde la potencia de ruido es estimada a partir del intervalo inicial del espectrograma donde no existe señal determinista, el algoritmo que se desarrollará en este apartado estima el límite de propagación a partir de las cinco primeras columnas del espectrograma.

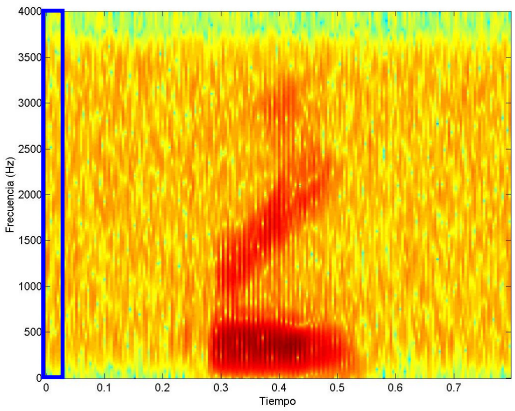


Figura 6.9 Espectrograma de señal vocal inmersa en ruido con el intervalo utilizado para estimar la PDF de ruido señalado en azul.

De esta manera, esta técnica presupone la existencia de una zona libre de señal vocal al comienzo del archivo lo que en principio resta cierta generalidad al proceso. Pero realmente, para realizar la estimación de ruido simplemente tomamos 5 columnas del espectrograma que, según los parámetros con los que ha sido generado, equivalen a 40ms. Este intervalo de tiempo mínimo siempre existirá al comienzo de una grabación por lo que el proceso siempre tendrá esa zona libre de señal vocal para poder realizar la estimación. De todas formas, si en algún caso no se dieran estas circunstancias, siempre podemos añadir al comienzo del algoritmo un sencillo detector de actividad vocal que no conlleva mucho tiempo de procesamiento y que nos indique las zonas más probables de estar libres de señal vocal y como consecuencia aptas como fuente de datos de nuestra estimación.

Para comprobar si el intervalo elegido para realizar la estimación es suficiente para llevar a cabo una determinación efectiva del límite de propagación se realiza un experimento comparativo entre el algoritmo original y el nuevo. El experimento consiste en tomar 125 muestras de tres grupos diferentes de SNR y tratarlas con el algoritmo original registrando los límites de propagación en los que se detiene, y a continuación tomar las mismas muestras y realizar la

estimación de límites a través del nuevo método simplificado para ver si los límites computados son los mismos. A continuación se muestran gráficamente los resultados obtenidos para los ratios con ambos métodos; la columna de la izquierda muestra el comportamiento del algoritmo original optimizado con P_e = 10% y la columna de la izquierda los límites computados a través de la estimación a partir del intervalo ruidoso inicial.

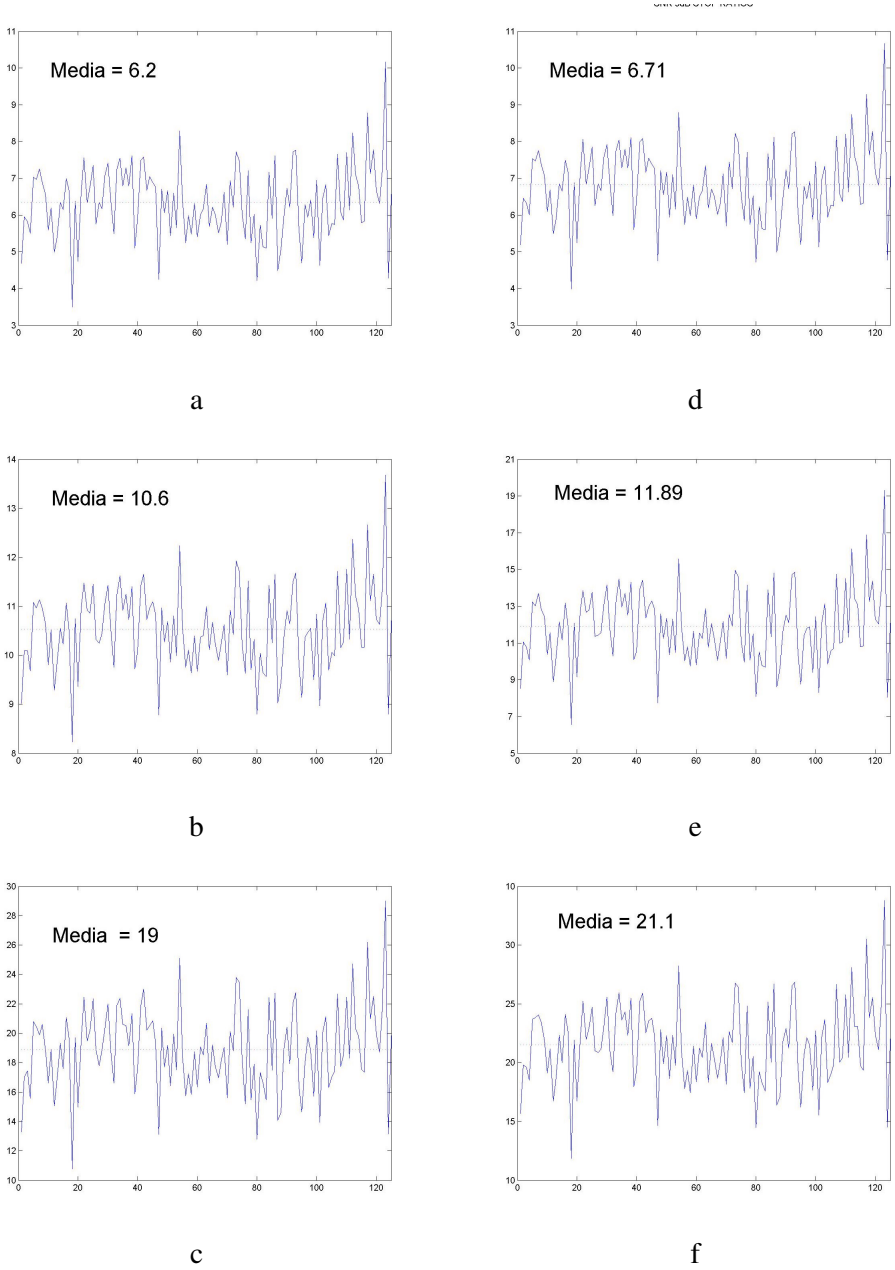


Figura 6.10 Comparación de los ratios definidos por el algoritmo original para 5dB (a), 10dB (b) y 15dB (c) y los ratios estimados a partir del intervalo inicial para 5dB (d), 10dB (e) y 15dB (f)

Como se puede comprobar en la figura anterior los ratios computados mediante los dos métodos son los mismos pero con una cierta desviación para todos ellos. El patrón de variación es el mismo lo que supone que el método simplificado lleva a cabo exactamente la tarea exactamente de la misma forma que el algoritmo original, pero de una forma muchísimo más eficiente, ya que no es necesario realizar ninguna iteración del algoritmo, sino que se puede llegar a las condiciones finales de manera directa. Como hemos dicho anteriormente los límites computados por la estimación tienen un offset respecto a los óptimos computados por el algoritmo con el parámetro P_e = 10%. Esto se debe a la diferencia existente entre los datos de partida del algoritmo original donde se toma todo el espectrograma no segmentado (en el que existirá cierta parte de señal determinista) y del nuevo enfoque que lleva a cabo una estimación más precisa debido a que sólo se basa en píxeles de ruido. Por lo tanto, para llegar a estos mismos límites óptimos con el proceso de estimación habrá que realizar un reajuste de la Pe, que presumiblemente debe ser menor, para hacer desaparecer el offset existente. A continuación se muestran los resultados de las medias computadas con la estimación para cuatro valores diferentes de P_e en comparación con las medias óptimas.

Original	**10%**		**9%**		**8%**		**7%**	
media	media	error	media	error	media	error	media	error
6.2	6.71	0.51	6.47	0.27	6.19	0.01	5.83	0.37
10.6	11.89	1.29	11.46	0.86	10.96	0.36	10.34	0.26
19	21.1	2.1	20.29	1.29	19.39	0.39	18.29	0.71

Tabla 6.8 Comparación de las medias de los límites para 4 valores diferentes de Pe en la estimación a partir del intervalo inicial y las medias óptimas computadas por el algoritmo original

Haciendo decrecer el valor de P_e del 10% al 8% cometemos el mínimo error respecto a los valores medios de los límites computados por el algoritmo original, de forma que definiendo esa nueva probabilidad reduciremos el offset al mínimo y los límites de propagación serán los más cercanos posible a los óptimos. Este decrecimiento en el valor de la probabilidad de error es lógico ya que la PDF computada a partir de píxeles de ruido exclusivamente es más precisa que la computada por el algoritmo original que siempre tendrá entre sus datos de origen píxeles de señal determinista, por lo que el error cometido a la hora establecer los límites es menor.

6.3.2 Experimentos y resultados

Para comprobar las prestaciones en cuanto a tiempo de computación y exactitud en reconocimiento de palabra se desarrollan dos nuevos algoritmos con los cuales se lleva a cabo dos experimentos. Ambos algoritmos, tras llevar a cabo la computación del Espacio Característico, realizan la estimación de la PDF de ruido a partir de los datos contenidos en las cinco primeras columnas del espectrograma los que dará lugar a una directamente a una PDF con la precisión adecuada, evitando el proceso iterativo de incremento de la exactitud de dicha estimación. A partir de esa estimación se definirá el límite de propagación con el parámetro P_e ajustado al 8% lo que dará lugar a los mismos límites computados por el algoritmo original, según se ha demostrado en el apartado anterior. Una vez aquí, el primero de los algoritmos propuestos llevará a cabo la segmentación de píxeles a través del proceso de propagación, lo que tendrá un coste de computación elevado pero presumiblemente conseguirá mejores resultados para el reconocimiento de palabra, especialmente para SNR bajas. El segundo de los algoritmos propuestos realizará la segmentación mediante la alternativa a la propagación descrita en el apartado 6.2.1 lo que dará lugar a un algoritmo mucho más rápido.

El experimento que se realiza con ambos algoritmos es el estándar de filtrado de 7 grupos de 1001 archivos pertenecientes a diferentes categorías de SNR. El tiempo que se tarda en realizar dicho proceso es registrado para poder comparar los requerimientos computacionales con los necesitados por el algoritmo original. Para obtener los resultados que alcanza este algoritmo en reconocimiento automático se envían las muestras filtradas al reconocedor entrenado con el algoritmo original optimizado (Pe = 10%). Los resultados se muestran a continuación en comparación con los alcanzados con el algoritmo original.

	limpio	20dB	15dB	10dB	5dB	0dB	-5dB
ORIGINAL	95.3	94.8	93.3	88.6	78.1	57.3	29.5
Estimación límite y propagación	89.9	94.8	93.2	89.0	77.9	56.9	29.3
Estimación límite y no propagación	90.2	94.7	94.1	86.5	74.0	53.2	25.4

Tabla 6.9 Comparación de resultados en reconocimiento de palabra entre el algoritmo original y los que estiman el límite directamente con y sin propagación posterior

La reducción del tiempo de computación respecto al algoritmo original es del 77.9% para el caso que lleva a cabo la propagación, lo que es un gran ahorro computacional pero que parece mínimo si se compara con el espectacular ahorro del 95% que se consigue con el algoritmo que utiliza la alternativa a la propagación. Además los resultados para exactitud del reconocimiento de palabra, exceptuando el caso limpio, son muy buenos en ambas versiones del algoritmo lo que pone de manifiesto la efectividad de esta técnica para definir unos límites de propagación adecuados que hagan segmentar las áreas óptimas. Los resultados son bajos en el caso limpio porque la determinación efectiva del límite se basa en la exactitud de estimación de la PDF de ruido, y para este caso los datos de origen para realizar la estimación no tienen las características del WGN (ya que no existe ruido al comienzo del archivo) lo que hace que se realice una estimación de una PDF que no determina con exactitud el límite óptimo de propagación. A pesar de ello los resultados son solamente un 0.85% en media por debajo del algoritmo original para el caso de la propagación y de un 2.7% para el caso sin propagación lo que hace que este último tenga el mejor balance entre coste computacional y exactitud en el reconocimiento de palabra.

6.3.3 Conclusiones

En este apartado se ha propuesto una nueva forma de estimar las condiciones últimas del algoritmo de manera que se elimina la necesidad del proceso iterativo de búsqueda de las mismas. La definición de las condiciones últimas en el algoritmo original se da cuando la estimación de la PDF de ruido solo tiene como datos de origen píxeles de ruido, estado al que se llega tras la extracción de la mayor parte de la señal tras llevar a cabo las iteraciones. La modificación llevada a cabo en este apartado utiliza directamente como datos de origen píxeles de ruido que se encuentran al comienzo de los archivos, de manera que se consigue directamente la precisión adecuada para la PDF y con ello la determinación efectiva del óptimo límite de propagación. Este procedimiento hace desaparecer la necesidad de las iteraciones, lo que reduce considerablemente el tiempo de proceso. Asimismo, el método mantiene los resultados en exactitud del reconocimiento de palabra ya que define los mismos límites de propagación que el algoritmo original, basándose en el mismo principio pero de una manera más eficiente.

El primero de los algoritmos puesto a prueba en este apartado define los límites de la nueva forma más eficiente y a continuación realiza el proceso de propagación clásico, consiguiendo así una reducción del coste computacional del 77.9% mientras que los resultados para el reconocimiento de palabra se mantienen al nivel del original con una diferencia relativa de tan sólo el 0.85% entre ambos métodos. El segundo de los algoritmos elimina las iteraciones mediante la nueva técnica de estimación de las condiciones finales y además elimina la necesidad de propagación mediante la utilización de la alternativa propuesta en el apartado 6.2.1. De esta forma el nuevo algoritmo es tremendamente más rápido que el original reduciendo los requerimientos computacionales en un 95%, a la vez que sigue siendo competitivo en cuanto exactitud del reconocimiento de palabra ya que se sitúa sólo un 2.7% en media por debajo de los resultados del original lo que hace de este algoritmo una técnica rápida y eficaz que debe servir como alternativa al algoritmo original.

El inconveniente que tiene esta técnica es la asunción de que al comienzo del archivo de voz existen al menos 40ms libres de actividad vocal para poder hacer a partir de esos datos la estimación de la PDF de ruido. Normalmente siempre tenemos este tipo de intervalos, y en ellos se basan técnicas como la substracción espectral con estimación de ruido clásica, pero en caso de no tenerlos sería necesario la inclusión en el algoritmo de un detector de actividad vocal basado en la

onda temporal o el espectrograma que no consumiera demasiados recursos para no degradar las buenas características del algoritmo. Como otra alternativa más eficiente también se podría aplicar un método de estimación de ruido basada en quantiles como la utilizada en la substracción espectral que se describe en el apartado 3.2.2. Además los resultados para exactitud en el reconocimiento de palabra no son demasiado buenos para el caso limpio, ya que la estimación de la PDF no se basa en muestras de ruido de forma que el límite determinado se aleja del óptimo dando como resultado una degradación de las áreas segmentadas respecto a las óptimas. De todas formas el caso limpio es un caso especial de referencia que no debe ser tomado en cuenta a la hora de evaluar la capacidad del algoritmo de limpiar muestras de voz inmersas en ruido. . A continuación presentamos un gráfico donde se muestra la exactitud y el coste computacional relativo al algoritmo original de ambos métodos presentados en este apartado.

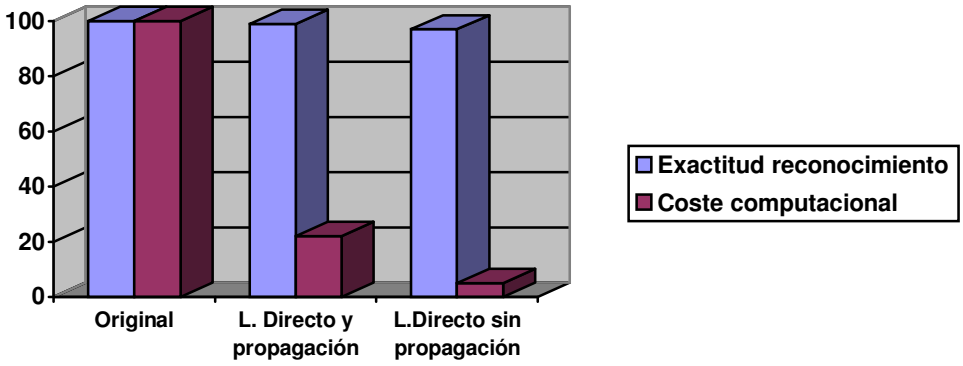
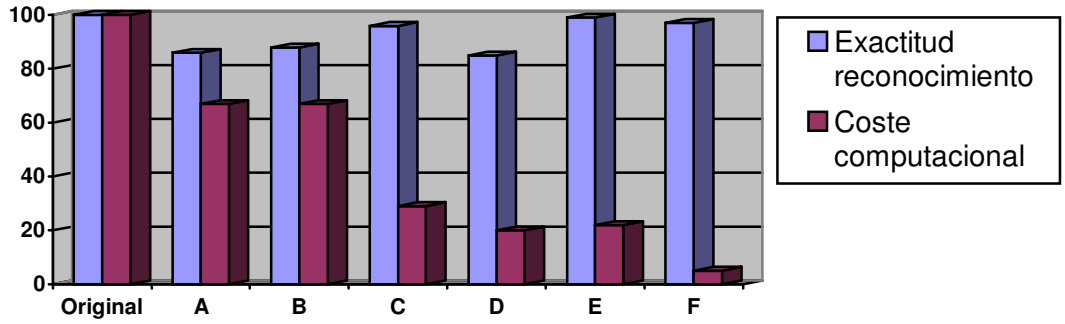


Figura 6.11 Comparación relativa de exactitud del reconocimiento y coste computacional frente al algoritmo original de los métodos que evitan las iteraciones mediante estimación directa del límite, con proceso de propagación y con la alternativa de 6.2.1.

6.4 CONCLUSIONES

A lo largo de este capítulo se han abordado diversas maneras de reducir el elevado coste computacional del filtrado morfológico manteniendo resultados en exactitud de reconocimiento de voz competitivos. Los métodos que se han utilizado se concentran por una parte en eliminar la

necesidad de llevar a cabo el proceso iterativo necesario para estimar con la precisión adecuada la PDF de ruido; y por otra parte en evitar el costoso proceso desde el punto de vista computacional que supone la búsqueda y propagación de semillas a lo largo tanto del Espacio Característico como del Espectrograma. A continuación se muestra un gráfico comparativo que expone los resultados de la evaluación de los diferentes métodos propuestos desde el punto de vista de la reducción de requerimientos computacionales y de la exactitud que mantienen en el reconocimiento de palabra frente al algoritmo original.



A	Estimación de la SNR y definición de condiciones finales de forma discreta
B	Estimación de la SNR y definición de condiciones finales de forma continua
C	Alternativa a la propagación conservando proceso iterativo
D	Alternativa a la propagación y combinación con B para evitar iteraciones
E	Estimación del límite a partir de intervalo inicial con propagación original
F	Estimación del límite a partir de intervalo inicial con alternativa a la propagación

Figura 6.12 Comparación de las prestaciones en reconocimiento automático de palabra y ahorro de coste computacional entre los distintos algoritmos propuestos

Los primeros acercamientos de *A* y *B* tratan de reducir el tiempo de proceso mediante la eliminación de las iteraciones al definir las condiciones finales del algoritmo directamente. Estas condiciones finales se obtienen de un anterior proceso de optimización de los ratios medios a utilizar para conseguir los mejores resultados en exactitud de reconocimiento de palabra llevado a

cabo en al apartado 4.3. Estos límites son dependientes de la SNR de forma que el algoritmo tiene que realizar primero una estimación de dicha SNR, para poder después definir los límites. *A* define los límites exclusivamente para los rangos de SNR definidos en Aurora 2 mediante una función de relación discreta, mientras que *B* construye a partir de la misma información una función continua válida para toda SNR. El algoritmo *C*, por el contrario, conserva el proceso iterativo y se centra en reducir en cada iteración el tiempo necesario para realizar la segmentación de los píxeles de señal determinista. Esto lo lleva a cabo mediante una alternativa a la propagación que consiste en segmentar directamente todos los puntos que caen en el Área de trabajo en una determinada iteración, evitando así el costoso proceso de búsqueda y propagación de semillas. *D* combina los acercamiento de *B* y *C* de forma que elimina las iteraciones mediante la estimación de la SNR y la definición a partir de ella de las condiciones últimas y una vez aquí realiza la segmentación mediante la alternativa a la propagación propuesta en *C*. En *E* y *F* se plantea una nueva forma mucho más eficiente de eliminar las iteraciones mediante una definición de las condiciones finales basada en el mismo proceso que lleva a cabo el algoritmo original pero llevado a cabo de manera más eficaz. Aquí se realiza una estimación de la PDF con la precisión final tomando como datos de origen las cinco primeras columnas del espectrograma ya que están completamente libres de señal determinista. *E* define el límite de propagación a partir de esta PDF y a continuación lleva a cabo el proceso de propagación original, mientras que *F* utiliza la alternativa a la propagación propuesta en *C*.

Los dos primeros enfoque deben realizar dos iteraciones del algoritmo, una para llevar a cabo la estimación de la SNR y la segunda para realizar la segmentación final. Esto hace que se reduzca el tiempo de proceso pero no de una forma tan notable como los otros acercamientos. Además, los errores en la estimación de la SNR y las bajas prestaciones del método de definir un ratio medio para determinar los límites de propagación hacen que la exactitud en reconocimiento vocal empeore considerablemente, por lo que estos algoritmos no serán candidatos a sustituir al original en comparación con los otros métodos. El acercamiento de *C* es bastante bueno, ya que la eliminación del proceso de propagación hace decrecer considerablemente el tiempo de ejecución mientras que no degrada excesivamente el comportamiento en el reconocedor. Al combinar en *D* los dos métodos conseguimos reducir aún más el tiempo de procesamiento pero aparecen los problemas que tiene *B* de manera que la exactitud en reconocimiento de palabra decrece de manera inaceptable. Por el contrario, los algoritmos *E* y F mantienen los resultados del reconocedor al nivel del algoritmo original mientras que reducen de manera espectacular el tiempo de ejecución (F llega a ser 20 veces más rápido que el algoritmo original). Si bien estos últimos algoritmos podrían

necesitar la inclusión de un detector de actividad vocal, las prestaciones que ofrecen los señalan como serios candidatos a sustituir al original cuando el tiempo de ejecución sea un factor limitante.

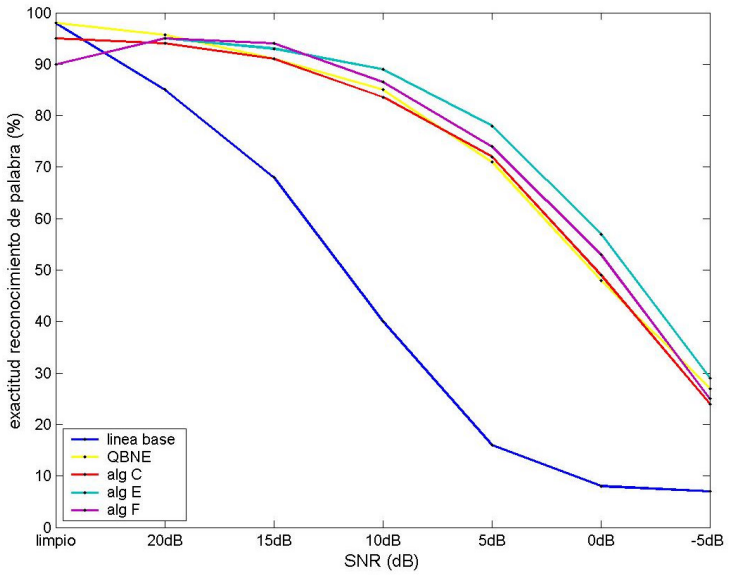


Figura 6.13 Comparación de resultados en exactitud de reconocimiento de palabra de los algoritmos C, E y F con la línea base y el perfil que se obtiene con la substracción espectral.

En la figura 6.13 observamos que cualquiera de los tres algoritmos alternativos tiene un comportamiento muy superior a la línea base, situándose en los niveles obtenidos con la substracción espectral. *C* consigue un porcentaje medio de mejora sobre la línea base del 26.6% mientras que sólo es un 1% peor que los resultados que alcanza la substracción espectral. *E* es en media un 29.9% superior a la línea base y 2.2% mejor que la substracción espectral, mientras que *F* supera a la línea base en un 27.9% y a la substracción espectral en un 0.25%. Por ello, podemos concluir que los algoritmos propuestos consiguen de manera efectiva el objetivo de reducir el coste computacional manteniendo resultados competitivos en exactitud de reconocimiento de palabra.