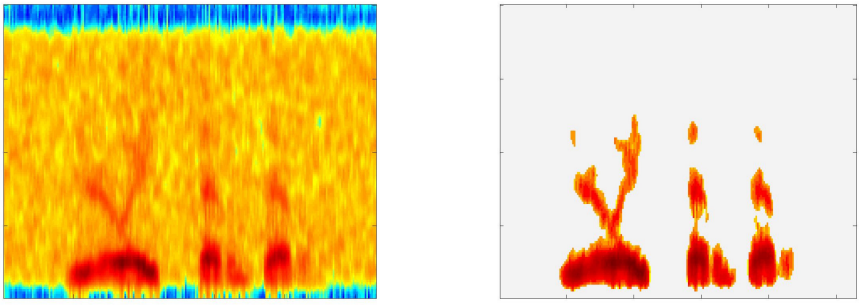
Las dos técnicas de mejora de la señal vocal para el reconocimiento que se estudian en este proyecto se basan en principios diferentes y ambas consiguen muy buenos resultados en exactitud del reconocimiento aplicadas de manera independiente. Persiguiendo un aprovechamiento de las prestaciones de ambas técnicas en un mismo algoritmo, para obtener aún mejores resultados en el reconocimiento automático, en este capítulo se estudia la manera de combinarlas. Así, se seguirá el paso lógico en el desarrollo del filtrado morfológico al intentar incluir las variaciones de amplitud espectral 'limpias' que ofrece la substracción espectral, en lugar de las ruidosas originales utilizadas en el algoritmo del capítulo anterior. Para realizar dicha combinación se propondrán dos configuraciones distintas, llamadas configuración serie y paralelo, que serán evaluadas a través de experimentos con el reconocedor de voz. Finalmente se llevará a cabo un estudio de los resultados obtenidos para valorar los beneficios de la posible combinación y que hará comprender mejor los principios de funcionamiento de ambas técnicas.

Capítulo 5

COMBINACIÓN FILTRADO MORFOLÓGICO - SUBSTRACCIÓN ESPECTRAL

5.1 INTRODUCCIÓN

El proceso del filtrado morfológico se basa en las características estadísticas del espectrograma de una señal de voz inmersa en ruido para detectar las zonas en las que existe señal determinista. Las características estadísticas generadas de manera local para cada píxel del espectrograma forman un nuevo espacio paralelo al espectrograma, en el que los puntos correspondientes a píxeles en zonas donde sólo existe ruido ocupan una posición diferente a los puntos correspondientes a píxeles en zonas donde además existe señal vocal. Así, el proceso consigue definir un umbral en dicho Espacio de Características que al aplicarlo al espectrograma permite distinguir las áreas donde sólo existe ruido y donde además tenemos las características deterministas de la señal vocal, procediendo finalmente a la extracción de esta última zona. De esta forma el filtrado morfológico simplemente detecta las áreas del espectrograma ocupadas por 'señal + ruido' dejándolas inalteradas, mientras que atenúa fuertemente las zonas clasificadas como 'ruido'.



a b

Figura 5.1 Espectrograma antes (a) y después (b) del filtrado morfológico

Así, las amplitudes espectrales presentes en el espectrograma con el que se genera la onda temporal resultado del filtrado morfológico están degradadas por el ruido que contamina el archivo, ya que la limpieza solo afecta a las zonas circundantes a lo que se considera la señal vocal. Esta limpieza se revela como notablemente beneficiosa desde el punto de vista de la mejora de la señal vocal para el reconocimiento automático de palabra, como demuestran los resultados obtenidos en el Capítulo 4. Utilizando las variaciones espectrales ruidosas se consiguen mejoras sobre la línea base para todas las SNR llegando a porcentajes de mejora de nada menos que el 62% para 5dB; e incluso cuando las variaciones de las amplitudes están fuertemente degradadas por el ruido como es el caso de -5dB se alcanza una mejora del 22.5%. Una vez demostrada entonces la utilidad de atenuar el ruido presente alrededor de la señal, el siguiente paso lógico en el desarrollo de una algoritmo que mejore aún más la señal vocal es tratar de eliminar o compensar de alguna forma el ruido presente en las zonas que contienen señal vocal.

La substracción Espectral explicada en profundidad en el apartado 3.2 lleva a cabo un tratamiento del espectrograma que da como resultado una limpieza de ruido en todo el espectrograma. Esta técnica lleva a cabo una estimación del ruido que esta corrompiendo la señal a través de un de las múltiples técnicas que han sido desarrolladas para tal efecto y a continuación realiza una compensación del espectrograma completo acorde a la estimación de la potencia de ruido que ha realizado anteriormente.

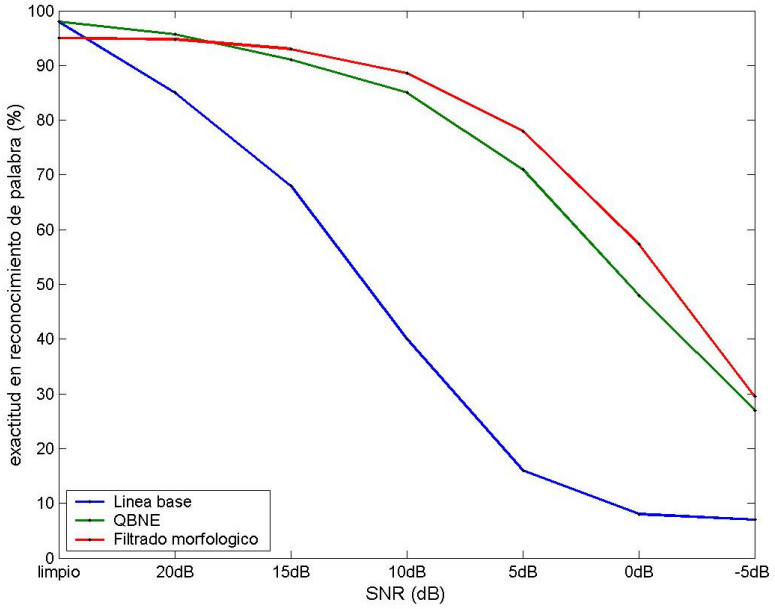


Figura 5.2 Comparación de comportamientos del filtrado morfológico y la substracción espectral basada en quantiles (QBNE) frente a la línea base

De esta forma las amplitudes espectrales de las zonas que contienen señal vocal son también tratadas con este proceso, por lo que en teoría la substracción espectral elimina el ruido presente en estas zonas. Mediante este tratamiento general, la substracción espectral consigue resultados en exactitud del reconocimiento de voz realmente buenos como se demuestra en el apartado 3.2 y como muestra el gráfico comparativo de la figura 5.2

Así, ambas técnicas demuestran tener un comportamiento eficaz en la mejora de la señal vocal para el reconocimiento automático, el filtrado morfológico mejorando la señal mediante una limpieza del ruido que rodea a las zonas de la señal vocal y la substracción espectral compensando el ruido presente en el espectrograma completo, dando como resultado comportamientos mucho más robustos frente al ruido. Por ello, el objetivo de este capítulo será tratar de combinar ambas técnicas para aprovechar las ventajas de los dos procedimientos, de manera que el filtrado morfológico se encargue de hacer la limpieza de las zonas que solo contienen ruido y la substracción espectral limpie las áreas de señal.

5.2 COMBINACIÓN DE AMBAS TÉCNICAS

A la hora de utilizar los dos procedimientos para tratar el espectrograma pueden proponerse varias configuraciones. El orden de aplicación de las técnicas es la primera de las variables a tener en cuenta, pero que se resuelve lógicamente a favor de aplicar primero la substracción espectral ya que esta técnica necesita el ruido alrededor de la señal para hacer la estimación de lo que debe substraer. De esta forma, la primera de las configuraciones propuestas llevará a cabo la aplicación uno tras otro de ambos procesos, comprobando el comportamiento que tiene el filtrado morfológico al aplicarlo a espectrogramas tratados mediante substracción espectral. A continuación se propondrá un segundo esquema el que se aplicarán ambas técnicas a un mismo espectrograma origen para combinar luego los resultados de una manera similar al procedimiento de inclusión de la amplitud desarrollado en el apartado 4.2. Ambas configuraciones serán evaluadas a través del comportamiento obtenido en reconocimiento automático de palabra.

5.2.1 Configuración serie

El acercamiento inmediato y más sencillo de la combinación de ambas técnicas es la aplicación de los procesos uno tras otro. El filtrado morfológico realiza una limpieza del ruido que corrompe el archivo alrededor de la zona de señal determinista, y esto es precisamente lo que necesita la técnica de substracción espectral para realizar la estimación de la potencia ruido que debe substraer por lo que lógicamente la substracción espectral deberá ser aplicada en primer lugar. De esta forma, tras la limpieza del espectrograma que realiza la substracción espectral la señal resultante tendrá teóricamente una relación señal a ruido mucho más elevada, por lo que el filtrado morfológico podrá detectar de manera más eficiente las áreas correspondientes a señal determinista. Así, el filtrado morfológico actuará siempre en condiciones mucho más cercanas a las condiciones limpias y teóricamente los resultados en exactitud del reconocimiento de palabra se acercarán también a los niveles conseguidos al tratar señales limpias. De esta forma, la aportación que tendrá el filtrado morfológico tras la substracción espectral será la de realizar una mejor atenuación de las áreas ocupadas por el ruido haciendo aumentar aún más la relación señal a ruido.

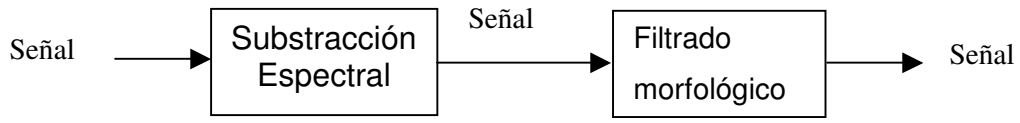


Figura 5.3 Configuración serie

Para comprobar las prestaciones de esta configuración realizamos el experimento clásico según las recomendaciones del paquete Aurora 2. Se lleva a cabo el procesamiento de 7 grupos de 1001 archivos con diferente relación señal a ruido y se obtienen los porcentajes de exactitud de reconocimiento de palabra que lleva a cabo el reconocedor entrenado de la manera óptima determinada en el Capítulo 4. La tabla que tenemos a continuación presenta dichos resultados en comparación con los obtenidos mediante ambas técnicas por separado.

	limpio	20dB	15dB	10dB	5dB	0dB	-5dB
Substracción espectral	98.5	95.7	91.9	85.2	71.0	48.6	27.4
Filtrado morfológico	95.3	94.8	93.3	88.6	78.1	57.3	29.5
Configuración serie	95.3	87.9	84.6	76.1	63.7	41.7	11.3

Tabla 5.1 Comparación de resultados en exactitud de reconocimiento automático de palabra entre Substracción espectral, Filtrado morfológico y la combinación de ambas técnicas según la configuración serie

Como puede comprobarse, aparte del resultado para el caso limpio, donde la técnica de substracción espectral deja las muestras sin procesar y como resultado tenemos el mismo resultado que con el filtrado morfológico independientemente, en lugar de una mejora lo que existe es una degradación acusada de los resultados. Esto se debe a que cuando un espectrograma de una señal vocal inmersa en WGN es procesada por la substracción espectral, las características del espectrograma resultado de la substracción no tiene las características ni de una señal completamente limpia, ni de una señal inmersa en WGN. Por lo tanto, al cambiar las características estadísticas del espectrograma, el Espacio Característico computado por el Filtrado morfológico tiene unas propiedades totalmente diferentes, de manera que los puntos generados en el Espacio Característico de las zonas que contienen señal y las zonas que contienen exclusivamente ruido no ocupan un lugar distinto en dicho espacio y como consecuencia no se puede definir un umbral que discrimine ambas zonas. Por ello, el filtrado morfológico no puede determinar las áreas correctas de señal vocal y no lleva a cabo una segmentación adecuada de la señal, lo que afecta notablemente al comportamiento con el reconocedor. Esto motiva la proposición de la siguiente configuración que conserva intactas las prestaciones tanto de la substracción espectral como del filtrado morfológico.

5.2.2 Configuración paralelo

Anteriormente concluimos que la substracción espectral necesita el espectrograma original ruidoso para poder realizar su tarea y en el apartado anterior se ha demostrado que el filtrado

morfológico también necesita el espectrograma original ya que se basa en las características estadísticas propias del mismo. Por ello se propone la configuración paralelo que estudiamos en esta apartado. La substracción espectral procesará el espectrograma original para dar como resultado un espectrograma mejorado donde se ha estimado el ruido que corrompe la señal y se ha realizado un proceso de extracción dando como resultado un espectrograma teóricamente limpio. El filtrado morfológico por su parte, procesará el espectrograma original para determinar las áreas que contienen señal determinista y las áreas que contienen exclusivamente ruido y a continuación se tomarán las variaciones espectrales resultado del proceso de substracción espectral para realizar el proceso de recuperación de amplitudes descrito en el apartado 4.2. El resultado de este procedimiento es el lógico a seguir en el desarrollo del algoritmo de filtrado morfológico; esto es, realizar todo el proceso de manera análoga pero a la hora de incluir las variaciones espectrales utilizar unos valores que han sido tratados para eliminar el ruido presente, por lo que al final tendremos en teoría un espectrograma totalmente limpio.

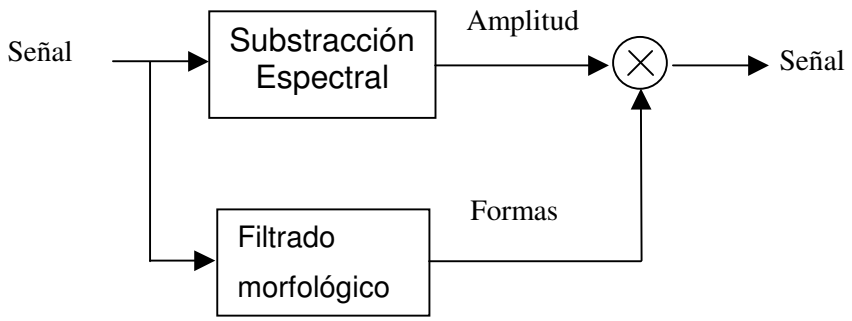


Figura 5.4 Configuración paralelo

Al igual que en el apartado anterior se comprueban las prestaciones de esta configuración procesando los 7 grupos de 1001 archivos con diferente relación señal a ruido y se obteniendo los porcentajes de exactitud de reconocimiento de palabra que lleva a cabo el reconocedor entrenado de la manera óptima determinada en el Capítulo 4. La tabla que tenemos a continuación presenta dichos resultados en comparación con los obtenidos mediante ambas técnicas por separado.

	limpio	20dB	15dB	10dB	5dB	0dB	-5dB
Substracción espectral	98.5	95.7	91.9	85.2	71.0	48.6	27.4
Filtrado morfológico	95.3	94.8	93.3	88.6	78.1	57.3	29.5
Configuración serie	95.3	95.2	93.9	87.9	77.9	59.8	30.8

Tabla 5.2 Comparación de resultados en exactitud de reconocimiento automático de palabra entre Substracción espectral, Filtrado morfológico y la combinación de ambas técnicas según la configuración paralelo

Al contrario que la configuración serie, la configuración paralelo sí mejora los resultados si bien los porcentajes de mejora no son demasiado amplios. En primer lugar observamos que para todo el rango de SNR bajo estudio (exceptuando el limpio en el que la Substracción espectral no realiza ningún proceso) la combinación de ambas técnicas da mejores resultados que la Substracción Espectral independientemente de forma que se demuestra la eficacia y utilidad que tienen el filtrado morfológico atenuando fuertemente las áreas de ruido que rodean a la zona de señal vocal en el espectrograma. La substracción espectral tiene que jugar con los parámetros α y β para dar un balance óptimo entre el ruido que substrae de la señal y la cantidad que deja como nivel base. El nivel base que se deja tiene como objetivo rellenar los valles producidos en las zonas de baja amplitud tras la substracción y la rectificación que hay que realizar para evitar los valores negativos. Estos esfuerzos se llevan a cabo para intentar eliminar el llamado 'musical noise' que queda tras realizar el proceso; y es aquí donde la combinación con el filtrado morfológico aporta sus ventajas, ya que elimina todo el ruido presente en esas zonas criticas de baja amplitud, erradicando todo rastro del 'musical noise'.

Respecto a los resultados obtenidos con el filtrado morfológico de manera independiente, la combinación con la substracción espectral aporta cierta mejora pero no la que podría esperarse y no para todas las relaciones señal a ruido. Esto es debido a que las amplitudes espectrales que estamos incluyendo no están realmente limpias, debido a la estimación que hace la substracción espectral del ruido presente. Como se describe detalladamente en el apartado 3.2 la substracción espectral obtiene un valor medio de la potencia de ruido y eso es lo que resta a las magnitudes del espectrograma ruidoso. Ese valor medio nada tiene que ver con los valores instantáneos que realmente contaminan el espectrograma, de forma que lo que se hace con las amplitudes espectrales es restarles un valor medio de manera general, que en el caso de WGN y si la estimación se hace de manera correcta, sería el mismo valor para todas las amplitudes del espectrograma. Por lo tanto, las amplitudes espectrales que se incluyen procesadas con la substracción espectral tienen las mismas variaciones instantáneas pero con un cierto offset negativo frente a las originales sin tratar. A pesar de ello, en ciertos casos parece que ayuda de alguna manera devolviendo la altura natural de los formantes frente al nivel de potencia cero y creando patrones de transición más suaves entre las zonas segmentadas que conservan las amplitudes y las que no segmentadas que son atenuadas a niveles mínimos.

El suave incremento en la calidad de la señal que aporta la substracción espectral depende obviamente de los parámetros que se utilizan α y β que se utilizan al aplicarla. Los parámetros que han sido utilizados en el experimento anterior son los que dan resultados óptimos cuando se aplica la técnica de substracción espectral de manera independiente tal y como se describe en el apartado 3.2.4. Estos parámetros se centran en intentar mejorar las características de la zona que rodea a la señal vocal para minimizar el 'musical noise' residual y en este caso esa zona no tiene en absoluto importancia ya que es limpiada de manera eficaz por el filtrado morfológico. Por ello los parámetros α y β utilizados probablemente no sean los óptimos para utilizar la substracción espectral de manera combinada con el filtrado morfológico. Para comprobarlo realizamos un experimento variando los parámetros α y β de la substracción espectral para ver como varía el comportamiento en cuanto a exactitud del reconocimiento de palabra. El experimento tendrá exactamente la misma configuración que el llevado a cabo anteriormente en este capítulo pero centrándonos sólo en el grupo de SNR 10dB y ajustando los parámetros de la substracción espectral a los valores indicados en la tabla de resultados siguiente.

		Reconocimiento %
3.25	0.01875	87.86
3.25	0.025	88.07
3.0	0.0125	87.77
3.0	0.025	88.07
3.0	0.0375	88.17
3.5	0.0125	87.77
3.5	0.025	88.07
3.5	0.0375	88.17
3.5	0.05	88.41
4.0	0.05	88.24

Tabla 5.3 Variación del comportamiento al variar los parámetros α y β de la substracción espectral

Se comprueba que efectivamente modificando los parámetros de la substracción espectral conseguimos un comportamiento mejorado del proceso global, de manera que los α y β utilizados no son los óptimos para la configuración combinada con el filtrado morfológico. A primera vista se observa que cambios en el parámetro α que es el que controla la sobre-estimación del ruido no afectan al comportamiento, mientras que variaciones de β que controla el nivel base que se deja en el espectrograma hacen que se modifiquen los resultados de forma que éstos son mejores a media que se incrementa el valor del parámetro.

5.3 CONCLUSIONES

Intentando mejorar aún más los resultados en reconocimiento automático de palabra frente a los obtenidos en el Capítulo 4 y motivados a seguir el paso lógico para hacer del filtrado morfológico una técnica más eficaz para la mejora de las señales vocales, se estudia en este capítulo la manera de incluir variaciones espectrales de mejores características que las ruidosas utilizadas en

el Capítulo 4. Para ello se utiliza la otra técnica de mejora de la señal vocal estudiada en este proyecto ya que ésta, a diferencia del filtrado morfológico, intenta eliminar el ruido de todo el espectrograma incluyendo las áreas ocupadas por la señal determinista. Por lo tanto la substracción espectral aporta las magnitudes espectrales libres de ruido que necesita el filtrado morfológico para intentar mejorar su comportamiento, por lo que se decide combinar ambas técnicas.

En el primer esquema de combinación, llamado configuración serie, los procesos actúan sobre el espectrograma ruidoso uno tras otro, de manera que uno de los algoritmos se basa en el resultado del algoritmo anterior. En este esquema la substracción espectral tiene que actuar primero de manera forzosa ya que esta técnica se basa en el ruido que elimina el filtrado morfológico de manera que si utilizásemos el orden inverso la substracción espectral no podría hacer nada ya que estimaría ruido nulo y no podría limpiar la zona de señal. De todas formas los resultados obtenidos en el único orden correcto no son buenos ya que la substracción espectral modifica las características estadísticas del espectrograma de manera que el filtrado morfológico ya no es capaz de determinar las áreas correctas ocupadas por la señal determinista.

El segundo esquema, llamado configuración paralelo, tiene como entrada a los dos procesos el espectrograma original ruidoso de forma que la substracción espectral obtiene un espectrograma mejorado donde las amplitudes están teóricamente libres de ruido y el filtrado morfológico define las áreas que hay que conservar y las que hay que atenuar. Frente a la substracción espectral de manera independiente este esquema obtiene mejores resultados lo que demuestra la eficacia y utilidad que tienen el filtrado morfológico atenuando fuertemente las áreas de ruido que rodean a la zona de señal vocal en el espectrograma, de manera que hace desaparecer todo el ruido residual que no elimina la substracción espectral.

Frente al filtrado morfológico usado de manera independiente no se observa una clara mejora para todas las SNR, debido a que si se estudia el proceso de estimación de ruido y compensación del mismo que lleva a cabo la substracción espectral para obtener las amplitudes espectrales, comprobamos que éstas realmente siguen siendo ruidosas. Esto nos lleva a dar un nuevo enfoque a la razón por la que funciona la substracción espectral. Esta técnica no obtiene la mejora en los resultados del reconocimiento vocal extrayendo el ruido la zona de señal, ya que en esa zona lo único que realiza es la resta de un nivel medio que nada tiene que ver con los valores instantáneos, sino que el mejor comportamiento es debido al alisamiento que consigue en la zona de ruido circundante, tal y como hace de una manera mucho más eficiente el filtrado morfológico.

A pesar de ello, la utilización conjunta a través del esquema paralelo, parece que en ciertos casos ofrece una cierta mejora. Presumiblemente el utilizar las amplitudes tratadas con la substracción espectral ayuda de alguna manera devolviendo la altura natural de los formantes frente al nivel de potencia cero y creando patrones de transición más suaves entre las zonas segmentadas que conservan las amplitudes y las que no segmentadas que son atenuadas a niveles mínimos. Esta ventaja sería mayor para ruidos que tuvieran alguna tendencia clara en frecuencia, como es el caso del ruido producido en un coche, de forma que tras la substracción espectral se devolvería la tendencia en frecuencia original de la señal vocal.

Finalmente estudiando los parámetros utilizados en la substracción espectral se concluye que éstos están optimizados para tratar el ruido circundante, lo que resulta inútil en este esquema ya ese ruido finalmente es eliminado por el filtrado morfológico. Variando estos parámetros en un experimento centrado en la relación señal a ruido 10dB vemos que efectivamente se obtienen mejores comportamientos, por lo que finalmente concluimos que la combinación entre el filtrado morfológico y la substracción espectral se muestra en principio beneficiosa, pero necesita una optimización de parámetros conjunta para llegar a un algoritmo realmente eficiente.