UNIVERSITAT POLITECNICA DE CATALUNYA
DEPARTAMENT DE TEORIA DEL SENYAL I COMUNICACIONS

Tesi Doctoral

TECNICAS DE SPEECH ENHANCEMENT CONSIDERANDO ESTADISTICAS DE ORDEN SUPERIOR

AUTOR : Josep Mª Salavedra Molí

DIRECTOR : Enrique Masgrau Gómez

TUTORA : Asunción Moreno Bilbao

Barcelona, Juny 1995

IV.5. Efectos nocivos asociados con el Filtrado iterativo de Wiener.

Como ya apuntamos en los apartados precedentes, el uso del Filtrado de Wiener estimando paramétricamente la voz original de una forma iterativa comporta una serie de efectos distorsionadores sobre la señal de voz original. Recuérdase que el principal propósito asociado con el uso del algoritmo iterativo de Wiener consiste en obtener una más fiel estimación espectral de la voz original pero, al mismo tiempo, ello conlleva la aparición de una cierta distorsión. El origen de tales efectos radica en la disponibilidad de la señal de voz ruidosa como única posibilidad donde se pueden estimar las características de la señal de voz durante la etapa de diseño del Filtro de Wiener. En este apartado se describen los principales efectos indeseados que presentan los algoritmos AR2, AR3 y AR4. Asimismo, se intenta justificar su origen mediante un estudio teórico relacionado con la convergencia de cada algoritmo iterativo. Previamente se realiza la evaluación de los algoritmos AR2 y AR3 mediante señal sintética para evitar las limitaciones impuestas por la estacionariedad de la voz durante la estimación de los cumulantes.

IV.5.1. Evaluación de los algoritmos AR2 y AR3 mediante señal sintética.

Debido a la no estacionariedad de la señal de voz, se debe procesar cada fichero de voz segmentado en varias tramas. Para cada trama debe obtenerse una estimación de los cumulantes o de la función autocorrelación. En el Capítulo III se ha visto como la disposición de una trama más larga origina mejores estimaciones, con una varianza y un sesgo menores. Sin embargo, el intervalo relàtivamente corto donde se puede aplicar la estacionariedad de la voz permite disponer de un conjunto de muestras de voz bastante pequeño en la mayoría de aplicaciones. Como las estadísticas de orden superior presentan una mayor varianza, durante la estimación de los cumulantes a partir de un conjunto finito de datos, puede suceder que la disposición de un conjunto pequeño de datos limite la eficacia propia del algoritmo bajo consideración.

Para evitar el efecto debido a las limitaciones asociadas con esta estacionariedad de la voz, se ha considerado la generación de sonidos sintéticos de una longitud predeterminada, normalmente 1024 muestras. De esta manera se suprime el efecto limitador asociado al

procesado por tramas y se puede evaluar el algoritmo de una forma más aislada. Para ello se han generado dos señales sintéticas mediante el modelo de producción de la voz representado en la Fig.II.7. Se ha tomado la envolvente de una vocal /a/ para diseñar el filtro lineal del tracto vocal y se han considerado una excitación consistente en un tren de deltas y otra formada por ruido exponencial para generar, respectivamente, un sonido sonoro y otro sordo. Nótese, además, que la señal sintética utilizada para esta evaluación se corresponde exactamente con un proceso paramétrico AR, mientras la señal de voz real se puede aproximar sólamente por un proceso AR.

Mientras los sonidos sonoros conllevan todas las características propias de la periodicidad asociada con la voz, los sonidos sordos las pierden y, además, presentan rasgos en cierta manera similares a las características particulares del ruido. Por esta razón se realiza este estudio sintético simulando voz sonora, diferenciando entre un pitch masculino y otro femenino, y voz sorda. Para generar los sonidos sonoros el filtro lineal del tracto vocal se ha excitado mediante un tren de deltas de periodo 64 y otro de periodo 32, equivalentes respectivamente a unos valores de pitch de 125Hz y 250Hz. A continuación se han degradado estas señales sintéticas mediante ruido AWGN, considerando los valores usados previamente para SNR_G entre 0dB y 24dB.

IV.5.1.1. Sonido sonoro (pitch=125 Hz).

Cuando comparamos los procedimientos normales de segundo y tercer orden podemos observar la presencia de las características que a continuación se enumeran.

Para la distancia Cepstrum, el procedimiento de tercer orden presenta una rapidísima supresión de ruido durante la primera iteración, frente al de segundo orden que obtiene la máxima mejora en la segunda. Para ambos métodos, a partir de la iteración posterior a esta óptima, la degradación aumenta rápidamente. Para SNR_G altas, se observa como todas las medidas de distancia empeoran al filtrar segun Wiener. En la primera iteración se obtiene un valor de distancia Cepstrum ligeramente peor al inicial correspondiente a la voz sin procesar. Inmediatamente las medidas correspondientes a iteraciones posteriores empeoran rápidamente en relación al valor obtenido en la primera iteración. Esta degradación, originada al superar la iteración óptima es mucho más acusada cuando el nivel de ruido es mayor.

En la Fig.IV.24, se puede observar como el método de cumulantes mejora la señal ruidosa de entrada cuando el nivel de ruido es alto ($SNR_G \leq 9dB$), y a partir de este valor (SNR_G superiores) no consigue mejorar ni en la primera iteración. En cambio, el método de correlaciones resiste hasta los 15 dB, a partir de aquí degrada la señal de entrada (como en el caso anterior). En vista de estos resultados puede parecer que cuando la SNR es alta es mejor no tocar nada, porque en lugar de eliminar el poco ruido que hay, distorsionaremos además la señal. Si bien objetivamente esto sería cierto desde este punto de vista, no olvidemos que en caso de que el oido u otro sistema sea el destinatario de la señal limpia, se pueden mezclar otros factores que aquí no tenemos en cuenta, entre ellos la subjetividad de la percepción.

Para la distancia de Itakura, ambos métodos presentan fuertes mejoras para $SNR_G \leq 9dB$ (mejoran casi 2.5dB para $SNR_G=0dB$). Pero cuando nos situamos en una $SNR_G=12dB$ la mejora ya tan sólo es de 0.3dB. A partir de este punto la degradación crece y ya no se obtiene ninguna mejora, sino que la señal empeora. La degradación en términos de distancia Itakura o Cosh es bastante menor. Así, la primera iteración mejora para SNR_G algo superiores al caso de distancia Cepstrum y, además, al sobrepasar la iteración óptima no se obtiene una rápida degradación sino unos valores estabilizados. Este comportamiento de las distancias espectrales conduce a situar la distorsión en las zonas correspondientes a los valles

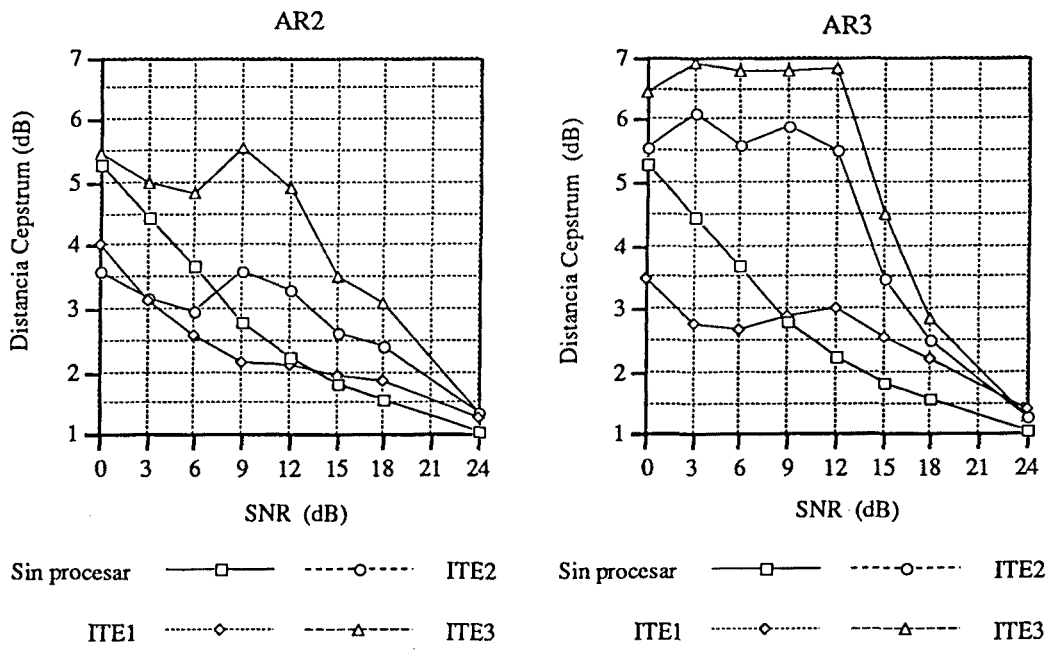


Figura IV.24 : Evolución de la distancia Cepstrum en función del nivel de ruido durante las tres primeras iteraciones procesadas con los algoritmos AR2 y AR3.

espectrales mientras la distorsión de los formantes es escasa. También se observa como esta distorsión localizada en los valles espectrales aumenta con el nivel de ruido.

En resumen, podemos decir que en este caso, el método de tercer orden es más rápido que el de segundo orden en cuanto a la mejora frente al número de iteraciones. Sin embargo, ambos procedimientos introducen una cierta degradación a partir de los 12-15 dB para esta señal sintética.

IV.5.1.2. Sonido sonoro (pitch=250 Hz).

Si bien en el caso anterior podríamos asimilar el valor del pitch utilizado (125 Hz) al caso de una voz masculina, este otro valor tendería a ser una voz femenina (más aguda). Cuando utilizamos un pitch distinto para generar la señal, el comportamiento del algoritmo (Fig.IV.25) es aproximadamente el mismo que en el caso anterior. La única diferencia existente en un primer examen de los resultados es una leve mejora de los valores en todas las

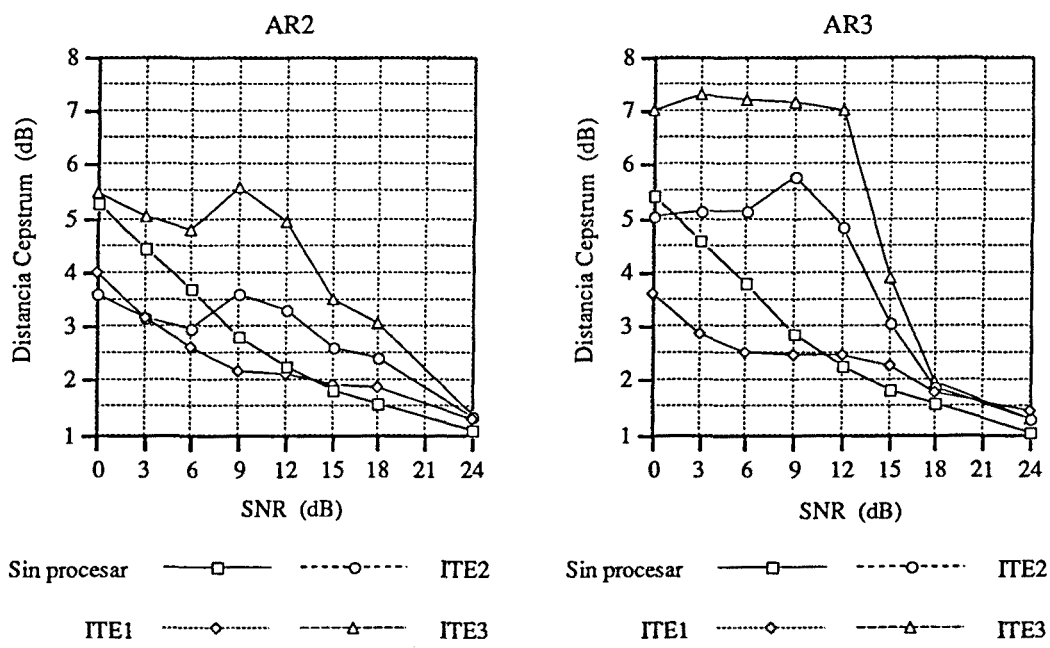


Figura IV.25 : Evolución de la distancia Cepstrum en función del nivel de ruido durante las tres primeras iteraciones procesadas con los algoritmos AR2 y AR3.

medidas, es decir, la ganancia es ligeramente superior y además se extiende hasta SNR_G algo más altas, empezando a introducir distorsión a partir de los 15-18 dB, esto es, tenemos un margen de unos 3dB superior sin que domine la distorsión.

Hay que hacer notar, sin embargo, que en estos resultados interviene un factor que Makoul comenta en [Makh-75]. Cuando tenemos un modelado AR, y la señal original no tiene un espectro continuo sino discreto, el modelo sólo tiene que pasar por los puntos donde tenemos espectro, quedando libre en el resto de puntos. Así, si en nuestro caso disminuimos el periodo del tren de deltas a la mitad, estamos separando las líneas espectrales (deltas) al doble de distancia (el valor del pitch será el doble). Esto significa que quizás ahora la envolvente LP tiene más libertad para adaptarse a los puntos por los que tendría que pasar.

IV.5.1.3. Sonido sordo.

Podemos sintetizar la señal utilizando en este caso la misma envolvente de la /a/ y una fuente de ruido exponencial, tomándose como aproximación de un sonido sordo. Se eligió esta fuente de ruido y no una gaussiana porque, en este último caso, nos encontraríamos con que el método de cumulantes no sabría distinguir entre voz y ruido (recordemos que los cumulantes de tercer orden de una distribución gaussiana se anulan).

El comportamiento que se observa ahora (Fig.IV.26) es distinto del caso de una excitación con trenes de deltas. Para la distancia Cepstrum, la mejora correspondiente a la primera iteración se extiende hasta SNR_G=15dB (algoritmo AR2) o hasta SNR_G=6dB (algoritmo AR3), mientras que para SNR_G=0dB existe una mejora de 1.6dB tras dos iteraciones (AR2) o 1.4dB en la primera iteración (AR3). El mínimo de distancia, resultante para el algoritmo AR2, se sitúa en la segunda iteración para las SNR bajas ($SNR_G \leq$3dB) y en la primera para las SNR_G a partir de 6dB. Con el uso del método de cumulantes (AR3) la iteración óptima siempre se corresponde con la primera y la mejora obtenida es notable hasta los 6dB de SNR_G (0.85dB), observándose que a partir de los 9dB empieza una degradación (0.5dB) que cada vez es más acusada. Si prolongamos las iteraciones hasta la tercera o cuarta, la distorsión acumulada es entonces ya muy acusada. Es decir, la distorsión ocasionada al superar la iteración óptima aumenta con el nivel de ruido y con la distancia entre la iteración considerada y la óptima. Además, se puede afirmar que este nivel de degradación supera la

obtenida para las tramas sonoras, puesto que ahora no se distingue tan claramente la voz y el ruido.

Para la distancia de Itakura, el comportamiento es distinto. Ahora, a diferencia del caso de una excitación sonora, esta distancia mejora, pero no de forma tan acusada (en los casos anteriores llegaba a tener un valor de casi 0dB para las SNR_G en el margen de 0dB-9dB), obteniendo una ganancia de 0.4dB a 1dB para el mismo margen de SNR_G. La distancia Cosh sigue un comportamiento parecido.

Hay, sin embargo, una diferencia en el comportamiento de la distancia Cepstrum. Cuando utilizamos señal real, o señal sintética sonora, es habitual que al disminuir el nivel de ruido la distancia Cepstrum empiece a degradarse antes que las distancias Itakura y Cosh. Sin embargo, al considerar ruido exponencial como excitación, no se produce este hecho sino que todos los valores de distancia empeoran aproximadamente al mismo tiempo, siendo la distancia Cepstrum la que resiste un poco más. Es difícil explicar el por qué sucede, aunque podríamos pensar que ahora no es posible aislar tan claramente los formantes, con lo cual las distancia de Itakura y Cosh no resulten tan favorecidas. No es intención tampoco de este trabajo entrar en análisis espectrales muy detallados que se dejan para futuros estudios.

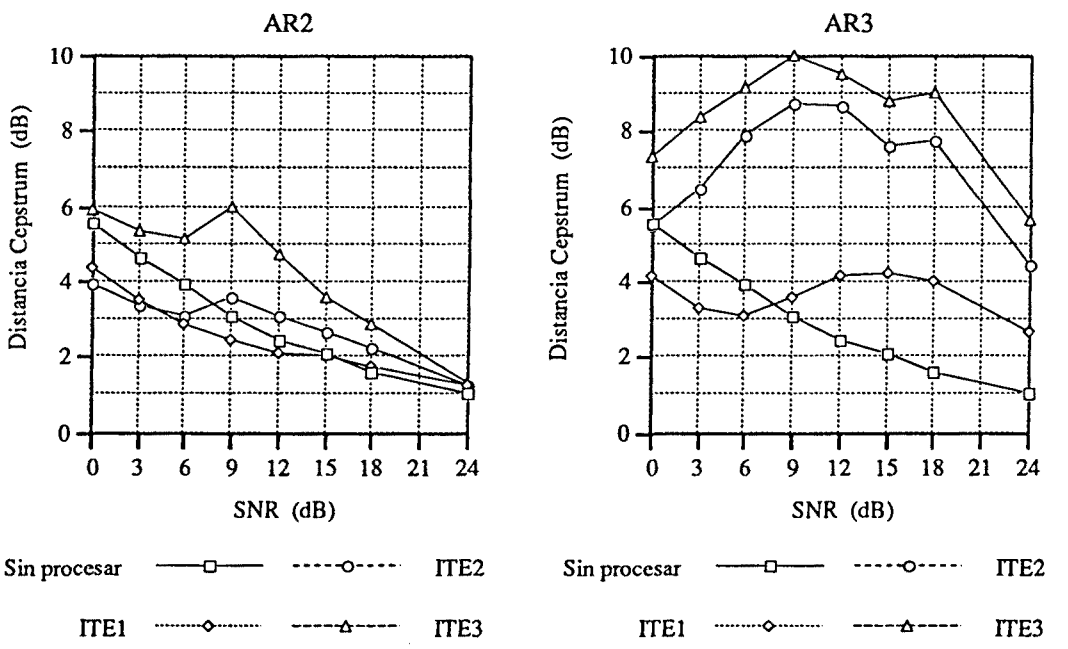


Figura IV.26 : Evolución de la distancia Cepstrum en función del nivel de ruido durante las tres primeras iteraciones procesadas con los algoritmos AR2 y AR3.

Para concluir, se puede afirmar que el algoritmo AR3 suprime el ruido de una forma más rápida y más notoria que el algoritmo AR2, cuando los niveles de ruido son apreciables, es decir, en aquellos entornos donde un sistema de realce de la voz puede ser más necesario. Para ambientes poco ruidosos, el algoritmo AR2 presenta unas prestaciones ligeramente superiores, ocasionando una menor distorsión. Sin embargo, este ligero incremento de distorsión no se aprecia durante las pruebas de audición.

Por otra parte, ambos algoritmos presentan mejores prestaciones para sonidos sonoros que cuando se aplican a sonidos sordos. Además al aumentar el Pitch asociado con los sonidos sonoros, las prestaciones mejoran ligeramente. En cambio, al trabajar con voz real, donde las funciones estadísticas se estiman a partir de conjuntos de datos más limitados en número de muestras y, además, la voz no se corresponde directamente con un proceso paramétrico AR, se aprecia un peor comportamiento para las tramas sonoras femeninas. El problema radica en que algunas tramas sonoras con un valor de Pitch alto (voz femenina) presentan una skewness muy pequeña (elevada componente simétrica de la voz) y, en consecuencia, la estimación AR resulta bastante pobre al obtenerse a partir de estos valores pequeños de los cumulantes de tercer orden.

IV.5.2. La Distorsión por Picado Espectral.

Para diseñar el filtro de Wiener:

$$W(w) = \frac{P_s(w)}{P_s(w) + P_r(w)} \qquad \text{(IV.37)}$$

tenemos que calcular el espectro de la voz original $\mathbf{P_s(w)}$ y el espectro del ruido $\mathbf{P_r(w)}$. Sin embargo, no disponemos de ninguno de los dos espectros, tan sólo podemos hacer estimaciones de ámbos.

Para el espectro del ruido $\mathbf{P_r(w)}$, debemos promediar las tramas disponibles en los periodos de silencio para establecer la estimación de su espectro, ya que en ausencia de actividad vocal la señal contiene sólamente ruido de fondo. Pero una vez transcurridos esos breves intérvalos de silencio, ya no podemos acceder al ruido como señal de referencia. Si bien de este modo se consigue una buena aproximación del espectro de ruido, tanto mejor

cuantas más tramas podamos promediar, esta estimación perderá credibilidad a medida que transcurra el tiempo sin poder hacer una actualización de dicho espectro. Por tanto podemos decir que la validez del espectro $P_r(w)$ se apoya en la estacionariedad del ruido. En los ensayos con ruido blanco gaussiano tal estacionariedad es realmente cierta, pero al enfrentar el sistema a ruidos reales que no sean estrictamente estacionarios, cuyo espectro de ruido irá modificándose paulatinamente de una trama a otra, la predicción irá en la misma medida perdiendo su validez. Para solucionarlo, deberemos actualizar el espectro cuanto antes sea posible. En este trabajo se utilizan frases cortas en las que puede aplicarse perfectamente la estacionariedad del ruido a lo largo de toda su duración, pero para una implementación práctica real debería usarse un detector voz-silencio.

Igualmente tampoco disponemos del espectro de la voz $P_s(w)$ ya que ni tan siquiera tenemos la señal de voz original $s(n)$ aislada. Hemos aproximado su espectro a partir de la envolvente de predicción lineal de la señal sucia de entrada $x(n)$. Nunca tendremos a nuestro alcance al espectro real, aún suponiendo un algoritmo AR que discriminara perfectamente ruido y señal. La aproximación del espectro por la envolvente no representa una limitación muy importante, puesto que tal como se comentó en el Capítulo II, para el oído humano es más importante la envolvente del espectro que el espectro en sí.

En resumen, estamos aplicando un filtro a la señal ruidosa de entrada que, aún siendo la mejor aproximación que podemos obtener bajo todas estas limitaciones, no es el óptimo de Wiener. La aplicación del filtro mejora la señal ruidosa, pero introduce también cierta distorsión. Inicialmente mejoramos la calidad de la voz al reducir el ruido de fondo presente, pero poco a poco el ruido desaparece y surge la tendencia contraria: la degradación en la inteligibilidad por distorsión espectral empeora la calidad de la señal.

La Fig.IV.30 muestra cómo la distancia espectral entre la señal de salida del algoritmo y la señal de voz limpia describe un valle que presenta un mínimo. A medida que vamos iterando, la distancia primero disminuye hasta ese mínimo, iteración k_i, para incrementarse posteriormente cuando el efecto distorsión es el dominante. En los tres algoritmos evaluados previamente se observa este comportamiento: el efecto supresión de ruido domina claramente durante las primeras iteraciones y, a partir de la iteración óptima, el efecto distorsión empieza a ser el efecto dominante, aunque exista todavía bastante nivel de ruido remanente. Esta situación se da para cada una de las tres distancias espectrales consideradas y para cada algoritmo particular, aunque los valores donde se alcanza la iteración óptima depende de cada caso específico. Para un algoritmo dado y bajo la consideración de una determinada medida de distancia espectral se obtiene una iteración óptima mayor a medida que se considera un mayor nivel de ruido. Así, para un nivel de ruido superior, el valor de la iteración óptima

aumenta y, además, el valor mínimo de distancia espectral que se puede alcanzar también es mayor. Esto se debe, en parte, a la superior presencia de ruido y, por otra, a la mayor distorsión ocasionada cuando el nivel de ruido aumenta, tal como veremos seguidamente.

Si asociamos valores de $SNR_G \geq 16dB$ para SNR alta, entre 8dB y 16dB para SNR media y de $SNR_G \leq 8dB$ para SNR baja, entonces, según lo discutido previamente, los mínimos respectivos se sitúan en la primera iteración para SNR alta ($k_1=1$), en la tercera iteración para SNR media ($k_2=3$) y alrededor de la cuarta o quinta iteración para SNR baja ($k_3=4$-5). Utilizando estadísticas de tercer orden reducimos el número necesario de iteraciones, aunque paralelamente también aumenta la distorsión más rápidamente cuanto más rápido llegamos al mínimo. Existe por tanto un compromiso entre la velocidad hacia mínima distancia espectral y distorsión espectral residual.

El Filtrado Iterativo de Wiener, que modela el espectro de voz mediante la aplicación reiterada del algoritmo de estimación AR propuesto por Lim y Oppenheim, nos conduce a un espectro de la señal de salida caracterizado por la reducción progresiva del ancho de banda de los formantes que la definen. Tal fenómeno ha sido bautizado como "spectral peaking" o picado espectral, y la posible justificación que le asociamos fue presentada en [Masg-92b].

A continuación veremos como el efecto de picado espectral en la zona de los formantes es uno de los efectos nocivos más importantes entre los que aparecen durante la ejecución del algoritmo iterativo de Wiener. Este efecto es más notorio al aumentar la iteración considerada y depende de cada algoritmo iterativo específico, siendo el problema más acusado para el

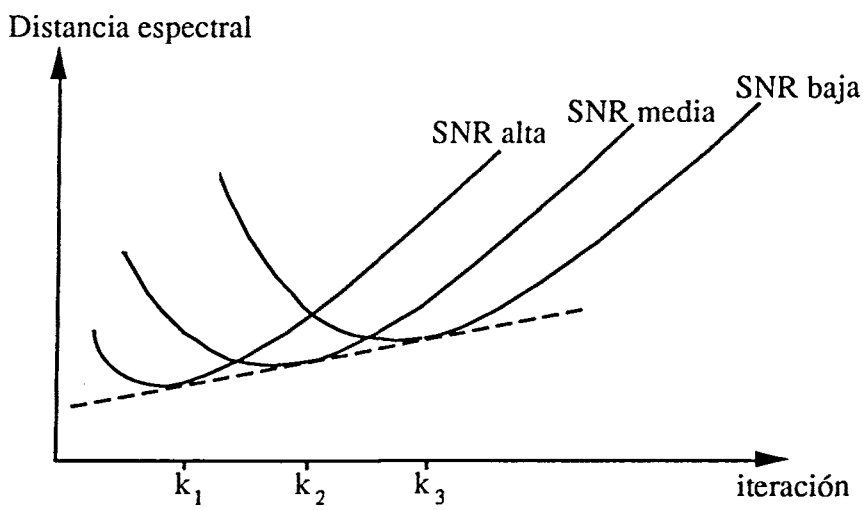


Figura IV.30 : Evolución del punto de mínima distancia espectral en función del número de iteraciones y para distintos niveles de ruido.

algoritmo AR3 que para los algoritmos AR2 o AR4, debido a la mayor agresividad de filtrado asociado con el método de tercer orden.

IV.5.2.1. Efecto de Picado Espectral para el algoritmo AR2.

Vamos a analizar el efecto distorsión por picado espectral originado por el algoritmo clásico AR2. En la Fig.IV.31 se han representado los distintos espectros LPC correspondientes a la señal de voz original (trazo continuo), a la señal de voz ruidosa y a las

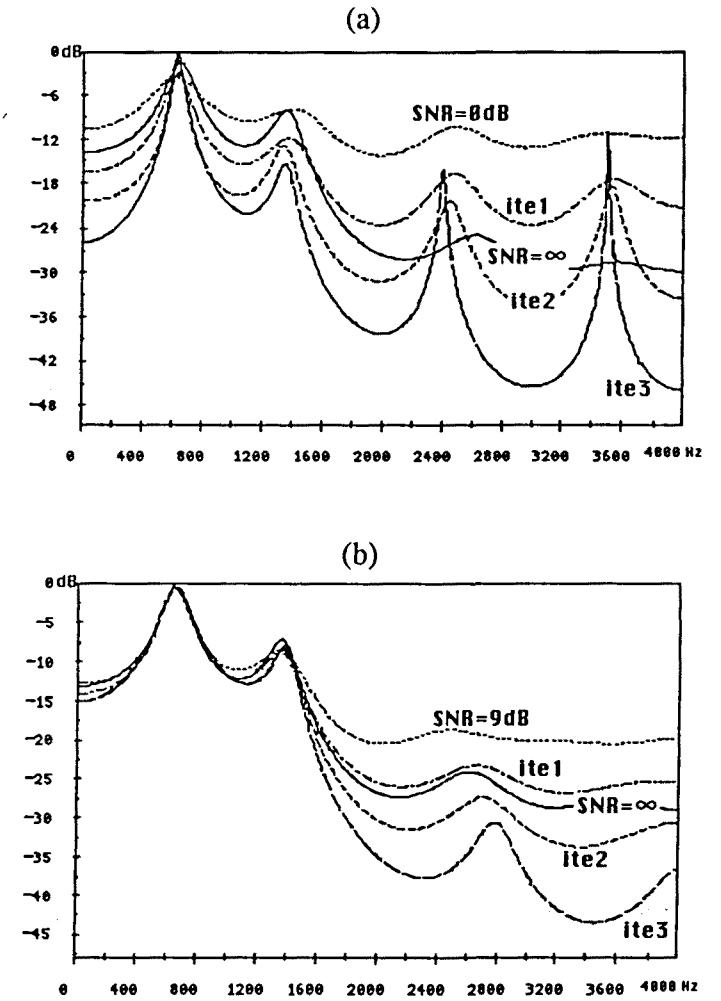


Figura IV.31 : Efecto de picado espectral en la envolvente LPC de una trama de voz sonora, vocal /a/. Evolución desde la estimación inicial hasta la estimación correspondiente a la tercera iteración (ITE3) para los niveles de ruido : (a) SNR_G=0dB ; (b) SNR_G=9dB

tres señales de voz realzada resultantes a la salida del filtro de Wiener durante las tres primeras iteraciones. Se ha considerado una trama de voz sonora, correspondiente al sonido vocálico /a/, y se ha degradado mediante dos niveles distintos de ruido AWGN para que resulten los valores SNR_G=0dB y SNR_G=9dB.

A primera vista, se observa como el efecto se agudiza al disminuir la relación señal a ruido disponible a la entrada. Para SNR_G=0dB el efecto es muy acusado, mientras que para SNR_G=9dB sólamente se presenta de forma apreciable para el tercer formante (zonas de baja energía). Para el caso de ambientes muy ruidosos, se aprecia como la sucesiva aplicación de la estimación AR sobre la señal de voz resultante de cada iteración provoca una continua erosión en la zona correspondiente a los valles espectrales, enfatizando la importancia relativa de los formantes. Obsérvese como la localización de los formantes se mantiene bastante bastante intacta, pero su ancho de banda se va reduciendo con el transcurrir de las sucesivas iteraciones, originando el fenómeno de distorsión conocido como picado espectral. En la zona de baja frecuencia, el ruido resulta enmascarado por la relativamente alta energía de la señal de voz y, en consecuencia, origina un picado espectral bastante suave. En la zona de alta frecuencia, la energía del ruido supera la energía de la voz y la enmascara. Entonces, el algoritmo iterativo permite reducir el nivel de ruido existente en esta zona, pero a cambio de la aparición de un efecto picado espectral bastante notorio.

Seguidamente intentamos justificar este comportamiento mediante un estudio teórico. Para poderlo razonar, debemos tener en mente el esquema general asociado con este filtrado iterativo de Wiener, representado al inicio del presente capítulo en las Fig.IV.1 y Fig.IV.2. En el caso ideal, suponiendo que las señales de voz original y ruido estan disponibles, el filtro óptimo de Wiener se diseñaría segun la expresión (IV.37). Sin embargo, sólamente se dispone de la señal de voz ruidosa y, en consecuencia, la información espectral de la voz original $\mathbf{s(n)}$ debe recuperarse a partir de la señal de voz ruidosa $\mathbf{x(n)}$, durante la primera iteración, o a partir de las señales de voz realzada $\mathbf{y_i(n)}$ para las restantes iteraciones.

A partir de nuestro modelo aditivo presentado en (II.1) y manteniendo las hipótesis de incorrelación entre los procesos de voz y ruido, se verifica que la función densidad espectral de potencia para la señal de voz ruidosa $\mathbf{P_x(w)}$ se corresponde con la suma de las densidades espectrales de la voz $\mathbf{P_x(w)}$ y del ruido $\mathbf{P_r(w)}$. Sin embargo, la no estacionariedad de la voz convierte esta igualdad en una aproximación cuando el intérvalo de tiempo es la duración de una trama. De esta manera vamos a considerar:

$$P_x(w) \;@\; P_s(w) + P_r(w) \qquad \text{(IV.38)}$$

Para la ejecución de este estudio teórico se hace uso de algunas definiciones y algunas condiciones, resumidas a continuación:

1) para obtener una mayor claridad, se omite la dependencia frecuencial en todas las expresiones, salvo cuando se crea necesario;

2) el filtro óptimo de Wiener W_{opt} se expresa como:

$$W_{opt} = \frac{P_s}{P_s + P_r} \qquad \text{(IV.39)}$$

3) se define el filtro complementario del óptimo de Wiener W^c_{opt} segun la relación:

$$W^c_{opt} = 1 - W_{opt} = \frac{P_r}{P_s + P_r} \qquad \text{(IV.40)}$$

4) se define el filtro D_i como el inverso del filtro de Wiener diseñado para la iteración i-ésima:

$$D_i = \frac{1}{W_i} \qquad \text{(IV.41)}$$

5) segun las definiciones anteriores, el filtro óptimo de Wiener y su complementario toman un valor acotado, para cualquier frecuencia del espectro:

$$0 \leq W_{opt} \leq 1 \qquad \text{(IV.42.a)}$$

$$0 \leq W^c_{opt} \leq 1 \qquad \text{(IV.42.b)}$$

$$W_{opt} + W^c_{opt} = 1 \qquad \text{(IV.42.c)}$$

El objetivo básico consiste en estudiar la convergencia de los sucesivos filtros suministrados por cada iteración del algoritmo iterativo AR2. En principio deberíamos esperar que el filtro de Wiener W_i diseñado en la iteración i-ésima converja hacia el filtro óptimo de Wiener W_{opt} para un valor de **i** suficientemente elevado. Así, se estamos buscando una ecuación de recurrencia que relacione el filtro W_i perteneciente a la iteración i-ésima con el filtro diseñado en la iteración anterior W_{i-1}.

Durante la primera iteración el filtro de Wiener se estima a partir de la señal de voz ruidosa **x(n)**, cumpliéndose la siguiente relación:

$$W_1 = \frac{P_x}{P_x + P_r} = \frac{1}{1 + \frac{P_r}{P_s + P_r}} = \frac{1}{1 + W_{opt}^c} \qquad \text{(IV.43)}$$

En la segunda iteración, se diseña el filtro de Wiener a partir de la señal de voz realzada $\mathbf{y_1(n)}$ disponible a la salida del filtro de Wiener $\mathbf{W_1}$ perteneciente a la primera iteración:

$$W_2 = \frac{P_{y_1}}{P_{y_1} + P_r} = \frac{1}{1 + \frac{P_r}{P_{y_1}}} \qquad \text{(IV.44)}$$

Esta señal de voz realzada $\mathbf{y_1(n)}$ resulta de filtrar la señal de voz ruidosa $\mathbf{x(n)}$ mediante el filtro de Wiener diseñado durante la primera iteración y, entonces, en el dominio frecuencial se verifica:

$$P_{y_1} = P_x \cdot W_1^2 = (P_s + P_r) \cdot W_1^2 \qquad \text{(IV.45)}$$

sustituyendo (IV.45) en (IV.44) resulta:

$$W_2 = \frac{1}{1 + \frac{W_{opt}^c}{W_1^2}} = \frac{1}{1 + W_{opt}^c \cdot D_1^2} \qquad \text{(IV.46)}$$

Obsérvese como esta expresión relaciona los filtros de Wiener diseñados durante las dos primeras iteraciones. Durante la tercera iteración, el filtro de Wiener se estima a partir de la señal de voz realzada $\mathbf{y_2(n)}$ resultante tras procesar la segunda iteración:

$$W_3 = \frac{P_{y_2}}{P_{y_2} + P_r} = \frac{1}{1 + \frac{P_r}{P_{y_2}}} \qquad \text{(IV.47)}$$

Análogamente, la señal de voz realzada $\mathbf{y_2(n)}$ resulta de filtrar la señal de voz ruidosa $\mathbf{x(n)}$ mediante el filtro de Wiener diseñado durante la segunda iteración:

$$P_{y_2} = P_x \cdot W_2^2 = (P_s + P_r) \cdot W_2^2 \qquad \text{(IV.48)}$$

sustituyendo (IV.48) en (IV.47):

$$W_3 = \frac{1}{1 + \frac{W_{opt}^c}{W_2^2}} = \frac{1}{1 + W_{opt}^c \cdot D_2^2} \qquad \text{(IV.49)}$$

De esta manera, para la iteración i-ésima, se puede obtener una relación de recurrencia en **i** que relacione los filtros de Wiener diseñados en dicha iteración y en la iteración precedente:

$$W_i = \frac{1}{1 + W^c_{opt} \cdot D^2_{i-1}} \qquad \text{(IV.50)}$$

o en términos del filtro de Wiener inverso:

$$D_i = 1 + W^c_{opt} \cdot D^2_{i-1} \qquad \text{(IV.51)}$$

Esta ecuación de recurrencia se puede constatar más claramente al aplicar el siguiente cambio de variable:

$$d(i) = D_i(f) \qquad \text{(IV.52.a)}$$

$$r = W^c_{opt}(f) \qquad \text{(IV.52.b)}$$

resultando la siguiente ecuación de recurrencia:

$$d(i) = 1 + r \cdot d^2(i-1) \qquad \text{(IV.53)}$$

y para que sea convergente debe cumplirse la siguiente condición:

$$d(i) = d(i-1) \qquad \text{cuando} \qquad i \longrightarrow \infty \qquad \text{(IV.54)}$$

y, de esta manera, en el límite resulta una ecuación cuadrática para **d(∞)**:

$$d(\infty) = 1 + r \cdot d^2(\infty) \qquad \text{(IV.55.a)}$$

$$r \cdot d^2(\infty) - d(\infty) + 1 = 0 \qquad \text{(IV.55.b)}$$

resolviéndola se obtiene la solución en función de los valores que tome **r**:

$$d(\infty) = \frac{1 \pm \sqrt{1 - 4r}}{2r} \qquad \text{(IV.56)}$$

donde deben distinguirse tres márgenes de valores para **r**, recordando (IV.52.b) y (IV.52.b):

$$\text{si} \quad 0.25 < r < 1 \quad \Rightarrow \quad d(\infty) \longrightarrow \infty \quad \text{diverge} \qquad \text{(IV.57.a)}$$

$$\text{si} \quad r = 0.25 \quad \Rightarrow \quad d(\infty) = 2 \qquad \text{(IV.57.b)}$$

$$\text{si} \quad 0 < r < 0.25 \quad \Rightarrow \quad d(\infty) \text{ tiene dos soluciones} \qquad \text{(IV.57.c)}$$

Para el caso **r>0.25** no existe solución analítica real, pero a partir de un estudio numérico sobre la ecuación de recurrencia (IV.53) se desprende que la variable **d(∞)** tiende a infinito con **i**, es decir, la solución diverge. Para **r<0.25** aparecen dos soluciones reales y en nuestro caso se debe elegir la pequeña:

$$d(\infty) = \frac{1 - \sqrt{1 - 4r}}{2r} \qquad \text{(IV.58)}$$

es decir, al disminuir los valores de **r** desde 0.25 hasta 0, la solución **d(∞)** recorre el camino de valores entre 2 y 1 y para un valor de r suficientemente pequeño dicha solución tiende al valor unidad.

Deshaciendo los cambios de variable realizados en (IV.52), se pueden interpretar estos resultados en el dominio frecuencial. Recuérdase que el análisis anterior adquiere validez para cada frecuencia del espectro. Evidentemente, para cada relación entre la energía de la señal de voz original y la energía del ruido, se obtiene un valor distinto del filtro óptimo de Wiener y, en consecuencia, un valor diferente del filtro de convergencia $W_\infty(f)$ del filtro de Wiener W_I. Por esta razón, parece adecuado considerar la relación señal a ruido para cada componente frecuencial del espectro, definida como:

SNR	$W_{opt}(f)$	$W_\infty(f)$
< 4.77 dB ($P_S<3P_r$)	< 0.75	0
4.77 dB ($P_S=3P_r$)	0.75	0.5
6.00 dB ($P_S=4P_r$)	0.80	0.72
7.53 dB ($P_S=5.6P_r$)	0.85	0.82
9.54 dB ($P_S=9P_r$)	0.90	0.89
12.79 dB ($P_S=19P_r$)	0.95	0.95
16.90 dB ($P_S=49P_r$)	0.98	0.98
∞ dB ($P_r=0$)	1	1

***Tabla IV.10** : Valores del filtro óptimo de Wiener y del filtro de convergencia (iteración infinita) en función de la SNR asociada con una frecuencia determinada.*

$$SNR(f) = 10 \cdot \log \frac{P_s(f)}{P_r(f)} = 10 \cdot \log \frac{W_{opt}(f)}{W^c_{opt}(f)} \qquad \text{(IV.59)}$$

En la Tabla IV.10 se han considerado distintos valores de **SNR(f)** para obtener una comparación entre los filtros óptimo de Wiener y el de convergencia. Además, ambos filtros se han representado en la Fig.IV.52. Como primera conclusión importante, se puede afirmar que el filtro óptimo de Wiener y el filtro de convergencia, originado por el algoritmo AR2, no coinciden. Ambos estan relacionados por la siguiente igualdad:

$$W_\infty(f) = \frac{2 \cdot \left(1 - W_{opt}(f) \right)}{1 - \sqrt{1 - 4\left(1 - W_{opt}(f) \right)}} \qquad \text{si} \qquad W_{opt}(f) > 0.75 \qquad \text{(IV.60)}$$

y sólamente para valores del filtro óptimo cercanos a la unidad, ambos filtros toman valores muy similares. Para la elaboración de la Tabla IV.10 se han considerado las expresiones (IV.59), (IV.42.c) y (IV.60).

Se aprecia claramente la aparición de un pico en la respuesta frecuencial del filtro de convergencia. En las frecuencias donde el ruido es considerable (SNR<4.77dB) el filtro de Wiener iterativo tiende al valor nulo a medida que aumente el número de iteraciones procesadas. Es decir, tiende a suprimir la información contenida en estas frecuencias. Por otra parte las frecuencias donde la SNR sea más alta (SNR>4.77dB), el filtro de convergencia y el

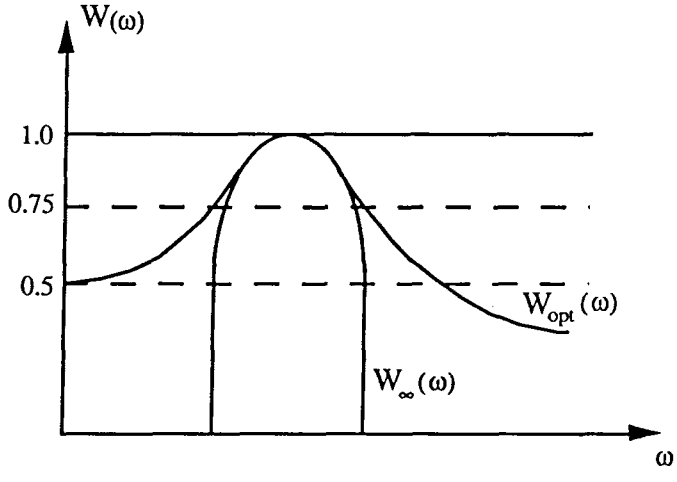


Figura IV.32 : Efecto de picado espectral debido a la presencia de ruido para el algoritmo AR2.

óptimo de Wiener tienden a coincidir, no produciéndose distorsión. Obviamente si consideramos ruido plano en frecuencia, esta zona de buen comportamiento se corresponde con los formantes de la voz, mientras la zona de peores valores de SNR se corresponde con los valles espectrales.

En resumen, se pueden distinguir tres zonas distintas, Tabla IV.11, para el filtro de convergencia $\mathbf{W}_\infty$ obtenido al aplicar este algoritmo iterativo AR2. En las zonas correspondientes a los formantes, el filtrado iterativo de Wiener tiende tiende hacia el filtro óptimo. En cambio, en los valles espectrales el filtro iterativo de Wiener $\mathbf{W}_i$ tiende a atenuar progresivamente la información contenida en éllos. Esto se traduce en una progresiva reducción del ancho de banda de los formantes y de la importancia relativa de los valles espectrales, dando lugar al efecto conocido como picado espectral. Nótese que esta distorsión adquiere una mayor importancia a medida que aumenta el nivel de ruido existente y que en el caso de ambientes silenciosos la distorsión ocasionada es prácticamente nula. De esta manera se ha podido justificar el comportamiento apreciado en la Fig. IV.31 y, además, los resultados obtenidos para señal sintética.

A medida que la relación señal a ruido disminuye, el efecto de picado espectral es más pronunciado, tendiendo a aislar los formantes. Esto puede apreciarse claramente en la Fig.IV.31.a, donde son los formantes con menor SNR de entrada, el tercero y sobretodo el cuarto, los que sufren de manera más acusada el fenómeno de picado espectral. También se aprecia en dicha figura como el ancho de banda de los formantes se va reduciendo a medida que se van ejecutando las distintas iteraciones, adoptando una forma de pico bastante acentuada en la tercera iteración, especialmente cuando el nivel de ruido es elevado. Se puede apreciar como este fenómeno de picado espectral es menos relevante cuando se aumenta la SNR para una determinada frecuencia. En entornos altamente ruidosos, la distorsión aumenta

SNR	$W_{opt}(f)$	$W_\infty(f)$
< 4.77 dB $(P_s<3P_r)$	< 0.75	0
4.77 dB $(P_s=3P_r)$	0.75	0.5
> 4.77 dB $(P_s<3P_r)$	> 0.75	$0.5 < W_\infty(f) \leq 1$

Tabla IV.11 : Zonas de comportamiento definido sobre el filtro de convergencia originado por el algoritmo AR2.

con el número de iteraciones consideradas. En este supuesto se debe evitar la ejecución de un número de iteraciones superior a dos, puesto que la distorsión por picado espectral aumenta rápidamente a partir de la segunda iteración. De esta manera, en la Tabla IV.12 se puede apreciar la rapidez de convergencia del filtro de Wiener W_i hacia su filtro de convergencia W_∞, segun distintos niveles de ruido para cada frecuencia particular. También se puede apreciar la velocidad de divergencia entre el filtro óptimo ideal W_{opt} y el filtro de convergencia W_∞.

		$W_i(f)$								
	SNR (f)	12	9	6	4.77	4	3	2	1	0
i	$W_{opt}(f)$	.94	.89	.80	.75	.72	.67	.61	.56	.50
1		.94	.90	.83	.80	.78	.75	.72	.69	.67
2		.94	.88	.78	.72	.68	.63	.57	.52	.47
3		.94	.87	.75	.67	.62	.54	.46	.38	.31
4		.94	.87	.74	.65	.57	.47	.35	.25	.16
5		.94	.87	.73	.62	.54	.40	.24	.12	.05
6		.94	.87	.73	.61	.50	.32	.13	.03	.00
7		.94	.87	.73	.60	47	.23	.04	.00	
8		.94	.87	.72	.59	.44	.14	.00		
9		.94	.87	.72	.58	.40	.06			
10		.94	.87	.72	.57	.36	.01	10^{-8}	10^{-19}	10^{-33}
∞		.94	.87	.72	.50	0	0	0	0	0

Tabla IV.12 : Evolución de la convergencia del filtro iterativo de Wiener para el algoritmo AR2.

En esta Tabla IV.12 puede apreciarse claramente la evolución del efecto degradador durante las sucesivas iteraciones. A medida que la relación señal a ruido disminuye, el efecto de picado cae más rápido a cero, para aquellas frecuencias donde $\mathbf{W_{opt}<0.75}$, tendiendo a aislar los formantes. Para estos valores de ruido elevado, el filtro iterativo y el filtro óptimo tienden a una buena aproximación durante la segunda iteración. Para la elaboración de dicha tabla se han aplicado las expresiones (IV.59), (IV.43), y (IV.50).

IV.5.2.2. Efecto de Picado Espectral para el algoritmo de cumulantes.

En los apartados IV.3 y IV.4 hemos visto como los algoritmos basados en los cumulantes de orden superior ofrecen un mayor desacoplo entre las señales de voz y ruido. Este cierto desacoplo se materializa en una mayor reducción de ruido y, además, de una forma más rápida, siendo necesario procesar un menor número de iteraciones. Durante la presentación de estos algoritmos AR3 y AR4, se ha reseñado como en condiciones ideales de estacionariedad este desacoplo sería total, pero como la voz sólamente puede considerarse estacionaria durante pequeños intérvalos de tiempo (trama), entonces, la propiedad anterior se verifica de forma parcial. Debido a esta limitación se plantea el estudio de distorsión por picado espectral mediante la introducción de un factor de desacoplo ∂, definido como:

$$\hat{P}_s(w) = P_x(w) = P_s(w) + \partial \cdot P_r(w) \qquad ,, \qquad 0 \leq \partial \leq 1 \qquad \text{(IV.61)}$$

Nótese como este factor de desacoplo ∂ dota de mayor generalidad al estudio de convergencia realizado previamente para el algoritmo AR2. Para el algoritmo clásico de segundo orden AR2, la estimación de $\mathbf{P_s(w)}$ en presencia de ruido conduce realmente al cálculo:

$$\hat{P}_s(w) = P_x(w) = P_s(w) + P_r(w) \qquad \text{(IV.62)}$$

correspondiéndose con el supuesto de máximo acoplamiento voz-ruido (∂=1). Pero el método de cumulantes intenta ver la presencia de una cantidad de ruido menor y, en consecuencia, el factor de desacoplo toma valores inferiores (∂<1). Unicamente en el supuesto de un desacoplo ideal voz-ruido, se obtiene un valor nulo para dicho factor (∂=0). Es decir, el método ideal debería ser totalmente ciego en relación a la presencia de ruido y obtener $\mathbf{P_s(w)}$ a partir de la observación de la señal de voz ruidosa $\mathbf{x(n)}$. Así, los valores del factor de desacoplo ∂

proporcionan una idea acerca de la robustez frente a la presencia de ruido para el algoritmo de predicción lineal AR y, en consecuencia, sirve para mostrar la bondad de la estimación espectral de la señal de voz. Las características del factor de desacoplo ∂ se pueden sintetizar a continuación:

$\partial = 1$ método de correlaciones o covarianzas (AR2) (IV.63.a)

$0 < \partial < 1$ método de cumulantes (AR3 y AR4) (IV.63.b)

$\partial = 0$ desacoplo ideal voz-ruido (IV.63.c)

La expresión (IV.61) se puede ver como una versión generalizada de la expresión (IV.38). Para realizar el estudio de convergencia correspondiente al algoritmo de cumulantes se usan las mismas definiciones usadas para el estudio anterior del algoritmo AR2. En relación al estudio de convergencia de segundo orden, este factor de desacoplo ∂ afecta las expresiones de los filtros W_i de tal manera que se precisa definir el filtro cuasicomplementario W^{cc}_{opt} del filtro óptimo de Wiener como:

$$W^{cc}_{opt} = \frac{P_r}{P_s + \partial \, . \, P_r} \tag{IV.64}$$

Obsérvese que continua verificando las condiciones (IV.42):

$$0 \leq W^{cc}_{opt} \leq 1 \tag{IV.65}$$

Para la primera iteración, la estimación AR de la voz se realiza a partir de la señal de voz ruidosa $\mathbf{x(n)}$, aunque en el caso de los cumulantes se aprecia un desacoplo ∂ en relación al ruido presente (IV.61), resultando el siguiente diseño para el filtro de Wiener correspondiente a la primera iteración:

$$W_1 = \frac{P_x}{P_x + P_r} = \frac{1}{1 + \dfrac{P_r}{P_s + \partial . P_r}} = \frac{1}{1 + W^{cc}_{opt}} \tag{IV.66}$$

En la segunda iteración, se diseña el filtro de Wiener a partir de la señal de voz realzada $\mathbf{y_1(n)}$ disponible a la salida del filtro de Wiener $\mathbf{W_1}$ perteneciente a la primera iteración:

$$W_2 = \frac{P_{y_1}}{P_{y_1} + P_r} = \frac{1}{1 + \dfrac{P_r}{P_{y_1}}} \tag{IV.67}$$

Esta señal de voz realzada $\mathbf{y_1(n)}$ resulta de filtrar la señal de voz ruidosa $\mathbf{x(n)}$ mediante el filtro de Wiener diseñado durante la primera iteración y, entonces, en el dominio frecuencial se verifica:

$$P_{y_1} = P_x \cdot W_1^2 = (P_s + \partial . P_r) \cdot W_1^2 \qquad \text{(IV.68)}$$

sustituyendo (IV.68) en (IV.67) resulta una relación entre los filtros de Wiener correspondientes a las dos primeras iteraciones:

$$W_2 = \frac{1}{1 + \dfrac{W_{opt}^{cc}}{W_1^2}} = \frac{1}{1 + W_{opt}^{cc} \cdot D_1^2} \qquad \text{(IV.69)}$$

De forma parecida al estudio anterior (AR2), se obtiene una relación de recurrencia para el filtro de Wiener correspondiente a la iteración i-ésima:

$$W_i = \frac{1}{1 + W_{opt}^{cc} \cdot D_{i-1}^2} \qquad \text{(IV.70)}$$

y vista en términos del filtro de Wiener inverso $\mathbf{D_i}$:

$$D_i = 1 + W_{opt}^{cc} \cdot D_{i-1}^2 \qquad \text{(IV.71)}$$

Aplicando el siguiente cambio de variables se puede constatar más claramente esta ecuación de recurrencia:

$$d(i) = D_i(f) \qquad \text{(IV.72.a)}$$

$$r' = W_{opt}^{cc}(f) \qquad \text{(IV.72.b)}$$

resultando la siguiente ecuación de recurrencia:

$$d(i) = 1 + r' \cdot d^2(i-1) \qquad \text{(IV.73)}$$

que en el límite origina la siguiente ecuación de segundo grado para $\mathbf{d(\infty)}$:

$$r' \cdot d^2(\infty) - d(\infty) + 1 = 0 \qquad \text{(IV.74)}$$

cuya solución viene dada por:

$$d(\infty) = \frac{1 - \sqrt{1 - 4r'}}{2r'} \qquad \text{(IV.75)}$$

Análogamente, segun los valores de **r'** la solución **d(∞)** adopta valores distintos:

$$\text{si} \quad 0.25 < r' \leq 1 \quad \Rightarrow \quad d(\infty) \longrightarrow \infty \quad \Rightarrow \quad W_\infty = 0 \qquad \text{(IV.76.a)}$$

$$\text{si} \quad r' = 0.25 \quad \Rightarrow \quad d(\infty) = 2 \quad \Rightarrow \quad W_\infty = 0.5 \qquad \text{(IV.76.b)}$$

$$\text{si} \quad 0 \leq r' < 0.25 \quad \Rightarrow \quad d(\infty) \text{ converge} \quad \Rightarrow \quad 0.5 < W_\infty \leq 1 \qquad \text{(IV.76.c)}$$

Vamos a analizar más detalladamente la expresión (IV.76.c). Deshaciendo el cambio de variable, la condición **r'<0.25** se traduce en:

$$W_{opt}^{cc} = \frac{P_r}{P_s + \partial \, . \, P_r} < 0.25 \qquad \text{(IV.77)}$$

Despejando P_s se obtiene una relación entre las densidades espectrales de la señal de voz y del ruido:

$$P_s > (4 - \partial) \, . \, P_r \qquad \text{(IV.78)}$$

Sustituyendo en la expresión del filtro óptimo de Wiener, se obtiene la condición anterior en términos de W_{opt}:

$$W_{opt} = \frac{P_s}{P_s + P_r} > \frac{(4 - \partial).P_r}{(4 - \partial).P_r + P_r} = \frac{(4 - \partial)}{(5 - \partial)} = C(\partial) \qquad \text{(IV.79)}$$

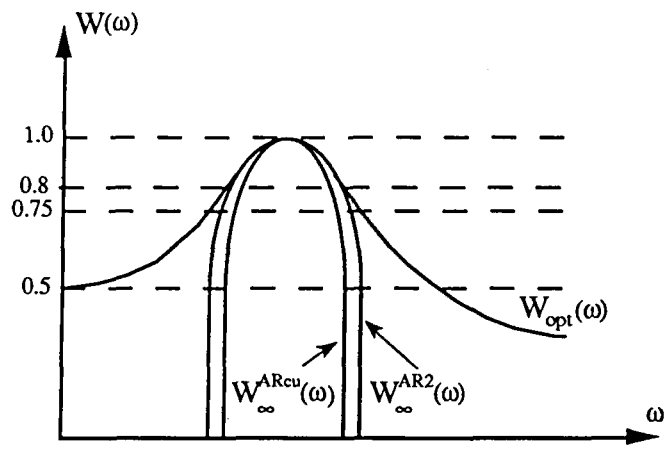


Figura IV.33 : Efecto de picado espectral debido a la presencia de ruido para los algoritmos AR2 y el de cumulantes ARcu.

El valor del factor de desacoplo ∂ es una característica propia de cada algoritmo de estimación AR particular, segun su eficiencia al diferenciar entre señal de voz y ruido. Para interpretar estos resultados, vamos a ver que sucede en los dos casos extremos, el método de correlaciones y el supuesto teórico ideal con total desacoplo voz-ruido (∂=0):

$$\begin{aligned} \partial = 1 \quad &=> \quad C(\partial) = 0.75 \\ \partial = 0 \quad &=> \quad C(\partial) = 0.8 \end{aligned} \qquad \text{(IV.80)}$$

Es decir, el filtro de Wiener tiende a un valor nulo cuando el filtro óptimo es inferior a $C(\partial)$. De esta manera, el filtro de Wiener, tras unas cuantas iteraciones, tiende a suprimir la información correspondiente a aquellas frecuencias donde el filtro óptimo de Wiener presenta un valor inferior a $C(\partial)$. Además, el valor de la constante $C(\partial)$ aumenta cuando un determinado algoritmo se caracteriza por un mayor desacoplo voz-ruido, es decir, cuando el factor de desacoplo tiende a cero. Por esta razón, se puede afirmar que los algoritmos de cumulantes, AR3 y AR4, presentan un mayor efecto de picado espectral en la región de los formantes en comparación al algoritmo clásico AR2 basado en la función autocorrelación, siempre que se considere un mismo número de iteraciones para ámbos. Los algoritmos de cumulantes se caracterizan por un valor $C(\partial)>0.75$ y, en consecuencia, el ancho de banda asociado con los formantes tiende a reducirse, en relación al caso $C(\partial)=0.75$ (AR2), tal como se muestra en la Fig.IV.33.

Por otra parte, el filtro de Wiener W_i converge a valores no nulos cuya cota superior viene dada por el filtro óptimo de Wiener, para aquellas frecuencias donde $W_{opt} \geq C(\partial)$:

$$W_\infty(f) = 0 \quad \text{si} \quad W_{opt} < C(\partial) \qquad \text{(IV.81.a)}$$

$$W_\infty(f) = 0.5 \quad \text{si} \quad W_{opt} = C(\partial) \qquad \text{(IV.81.b)}$$

$$0.5 < W_\infty(f) \leq W_{opt}(f) \quad \text{si} \quad W_{opt} > C(\partial) \qquad \text{(IV.81.c)}$$

Como conclusión se puede afirmar que los algoritmos cuya estimación AR se realiza a partir de los cumulantes de orden superior proporcionan un mayor desacoplo voz-ruido. A medida que este desacoplo ∂ tiende a cero nos estamos acercando más a las condiciones del algoritmo ideal caracterizado por el total desacoplo voz-ruido pero, a su vez, $C(\partial)$ aumenta desde 0.75 hasta un valor cercano a 0.8 y, consecuentemente, el efecto de picado espectral aumenta. No debe perderse de vista que la mayor distorsión asociada con los algoritmos de cumulantes, AR3 y AR4, resulta de la ejecución de un mismo número de iteraciones. Sin embargo, se ha visto que la convergencia asociada con estos algoritmos de cumulantes es bastante más rápida que la propia del algoritmo AR2 y, en consecuencia, se precisa procesar

un menor número de iteraciones y ello conlleva una menor distorsión final para los algoritmos de cumulantes.

En la Tabla IV.13 se han representado valores comparativos correspondientes al filtro óptimo de Wiener y al filtro de convergencia resultante para el supuesto de desacoplo ideal (∂=0). Para su elaboración se han considerado las expresiones (IV.59), (IV.63) y la solución (IV.75) deshaciendo el cambio de variable (IV.72):

$$W_{\infty}(f) = \frac{2 \cdot W_{opt}^{cc}(f)}{1 - \sqrt{1 - 4.W_{opt}^{cc}(f)}} \qquad \text{si} \qquad W_{opt}(f) > C(\partial) \qquad \text{(IV.82)}$$

Obsérvese que el caso ideal se verifica:

$$W_{opt}^{cc}(f) = 10^{-\frac{SNR(f)}{10}} \qquad \text{(IV.83)}$$

En comparación al algoritmo AR2 (Tabla IV.10), los filtros de convergencia y óptimo de Wiener toman valores aproximadamente idénticos a partir de una SNR más alta. Además, el filtro de convergencia para el caso ideal de total desacoplo voz-ruido es bastante más estrecho. Se puede concluir que el filtro de convergencia asociado con el algoritmo de cumulantes produce una mayor distorsión por picado espectral de los formantes. Esta

SNR	$W_{opt}(f)$	$W_{\infty}(f)$
< 6.00 dB (P_S<4P_r)	< 0.80	0
6.00 dB (P_S=4P_r)	0.80	0.5
7.53 dB (P_S=5.6P_r)	0.85	0.77
9.54 dB (P_S=9P_r)	0.90	0.87
12.79 dB (P_S=19P_r)	0.95	0.94
16.90 dB (P_S=49P_r)	0.98	0.98
∞ dB (P_r=0)	1	1

***Tabla IV.13** : Valores del filtro óptimo de Wiener y del filtro de convergencia (∂=0) en función de la SNR asociada con una frecuencia determinada.*

característica impone la consideración de un menor número de iteraciones en relación al algoritmo AR2. Es decir, si el algoritmo de cumulantes precisa procesar el mismo número de iteraciones para suprimir el ruido de un entorno específico, entonces, se debe elegir el algoritmo de segundo orden AR2 pues conduce a una menor distorsión.

En la Tabla IV.14 se ha realizado un pequeño estudio para evaluar la velocidad de convergencia del supuesto ideal (∂=0). Para su elaboración se han considerado las expresiones (IV.59), (IV.83), (IV.66) y (IV.70). Nótese que en el caso ideal se obtiene el filtro ideal en la primera iteración $\mathbf{W_1=W_{opt}}$. Así, en este supuesto no sería necesario continuar iterando puesto que en la primera iteración se dispone de $\mathbf{P_s(w)}$, ya que el ruido presente en $\mathbf{x(n)}$ resulta invisible durante el proceso de estimación AR. Sin embargo, el propósito de dicha tabla consiste en adquirir un conocimiento acerca de la velocidad de convergencia a partir de la evolución de $\mathbf{W_i}$ a lo largo de las diez primeras iteraciones, hacia el filtro de

		$W_i(f)$								
	SNR (f)	12	9	6	4.77	4	3	2	1	0
i	W_{opt} (f)	.94	.89	.80	.75	.72	.67	.61	.56	.50
1		.94	.89	.80	.75	.72	.67	.61	.56	.50
2		.93	.86	.72	.63	.56	.47	.37	.28	.20
3		.93	.86	.67	.54	.44	.31	.18	.09	.04
4		.93	.85	.64	.47	.33	.16	.05	.01	.00
5		.93	.85	.62	.40	.21	.05	.00	.00	
6		.93	.85	.61	.32	.10	.00			
7		.93	.85	.59	.24	.03				
8		.93	.85	.58	.15	.00				
9		.93	.85	.58	.06				10^{-64}	10^{-91}
10		.93	.85	.57	.01	10^{-10}	10^{-34}	10^{-72}		
∞		.93	.85	.50	0	0	0	0	0	0

Tabla IV.14 : Evolución de la convergencia del filtro iterativo de Wiener para ∂=0.

convergencia W_∞. Se aprecia una mayor velocidad de convergencia de W_i hacia W_∞ cuando se considera el algoritmo de cumulantes, especialmente para aquellas frecuencias donde el filtro tiende a atenuar ($W_{opt}<C$). Por esta razón, cuando se haga uso del algoritmo de cumulantes, se debe procurar la ejecución del número de iteraciones estrictamente necesario y no sobrepasarlo porque, en caso contrario, la distorsión aumenta considerablemente.

IV.5.2.3. Desplazamiento de los Formantes.

La presencia del ruido durante la estimación del modelo autorregresivo de la voz influye también sobre los formantes desde otro flanco, todavía no comentado: en la ubicación de los mismos. En nuestra situación, sería lógico esperar que la localización de los formantes fuera bastante fija al relacionar tramas consecutivas o, incluso, distintas iteraciones pertenecientes a la misma trama. No obstante, se puede observar como sufren un cambio progresivo de localización, más o menos errático.

Al hacer la estimación AR en presencia de ruido interferente sufrimos un desplazamiento de los formantes alrededor de la posición real. El grado de dispersión de tal desplazamiento vendrá dado, en primer lugar, por el modelo de predicción lineal que utilicemos: a mayor robustez frente al ruido, es decir, mayor desacoplo voz-ruido, menor será

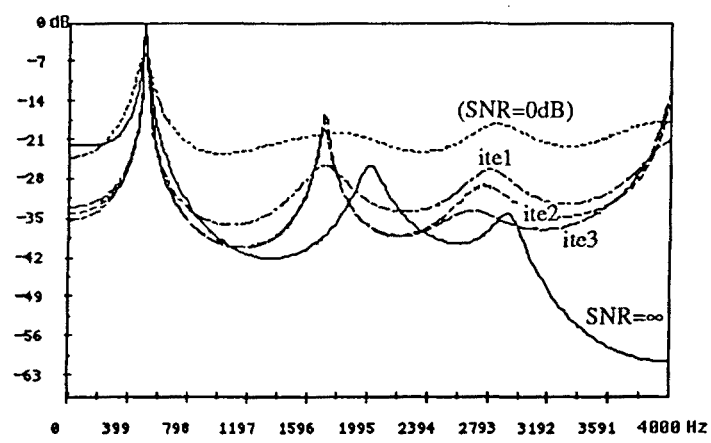


Figura IV.34 : Efecto de desplazamiento y mal posicionamiento de los formantes dentro de una misma trama sonora (/e/ real), a lo largo de las 3 primeras iteraciones del algoritmo. (método AR2 sobre un segmento de 512 muestras de señal).

la incertidumbre en la posición del formante. En segundo lugar, y como parece obvio, dependerá también de la cantidad de ruido presente en la señal de entrada. Si hay mucho ruido enmascarando la señal, más difícil será que el sistema sitúe correctamente los formantes que la forman, pudiendo incluso perder a alguno de vista cuando su SNR sea excesivamente baja. También influye negativamente la varianza y el sesgo de las estimaciones de las correspondientes funciones estadísticas, función autocorrelación y cumulantes, sobre las cuáles se realiza la extracción de las características espectrales de la voz.

En la Fig.IV.34 se aprecia como, incluso dentro de una misma trama, se produce un desplazamiento de los formantes al iterar. El primer formante se engancha con precisión a su auténtica localización; para el segundo formante erramos en la estimación inicial de su posición y se estabiliza en una localización errónea; y el tercero como presenta una SNR más baja que el resto no engancha en ningún momento y sufre un desplazamiento iteración a iteración. Nótese que este efecto es muy notorio para el supuesto representado en la Fig.IV.34 porque, tal como se ha visto en el Apartado IV.2, el algoritmo AR2 muestra una total incapacidad para afrontar niveles de ruido tan elevados. Para los algoritmos cuyas prestaciones son buenas al atacar un determinado nivel de ruido, el efecto es mucho menor pero existe y debe tenerse en cuenta. En nuestro caso los algoritmos de cumulantes ven una menor presencia de ruido que el algoritmo AR2 y ello se traduce en una mejor localización de los formantes. Pero, en algunas tramas, la estimación de los cumulantes presenta una mayor varianza que la estimación de la función autocorrelación, y ello comporta una mayor varianza en la localización de los formantes. Este efecto sólamente se puede corregir aumentando el número de muestras por trama de voz ruidosa.

IV.5.2.4. Ruido Musical Residual. Pérdida de Reconocimiento del Locutor.

Otros dos efectos adicionales que aparecen al filtrar con modelos AR son el ruido musical y la pérdida de reconocimiento del locutor.

Aún teniendo una señal de partida con una SNR media prefijada, no podemos decir que cualquier trama de voz dentro de dicha señal posea esa misma relación señal a ruido. Los distintos niveles de entonación que hace cualquier locutor al hablar, dentro de una misma sentencia (frases afirmativas, exclamativas, interrogativas, etc...) o incluso dentro de una

misma palabra (las consonantes sordas se pronuncian con menor energía que los sonidos sonoros), hacen que nos encontremos con zonas donde domina la voz sobre el ruido, con partes que quedan totalmente enmascaradas y trazos donde se tiene aproximadamente la relación SNR de partida. Es decir, cada trama de señal presenta un valor variante en la energía de la voz y, al mismo tiempo, un nivel de energía de ruido bastante próximo a su energía media.

En cada iteración el ruido de fondo, muy molesto, disminuye hasta casi desaparecer; pero puede surgir a la vez un nuevo tipo de ruido, el ruido musical, que desmerezca la mejora obtenida en la señal de voz. Este tipo de ruido, más acusado en las tramas o zonas del espectro donde existe una menor relación señal a ruido, se denomina musical por las notas musicales que produce (no por ello agradables) producto de la aparición de picos espurios a frecuencias altas del espectro de la voz.

La evolución del filtrado, iteración a iteración, sigue unos pasos generales para los métodos básicos de correlaciones y cumulantes de tercer y cuarto orden. Tras la primera iteración se logra eliminar un porcentaje elevado de ruido de entrada (en función del algoritmo considerado). Al progresar en iteraciones aparecen espurios en la parte de menor SNR del espectro, es decir la zona de alta frecuencia, y estos espurios de una trama a la siguiente varían en forma y posición (si es que aparecen), ofreciendo al escucharlo la sensación de oir brevísimas notas disonantes y cortantes que no corresponden en absoluto al espectro de voz original. Estos espurios de alta frecuencia pueden ser producto tanto de una

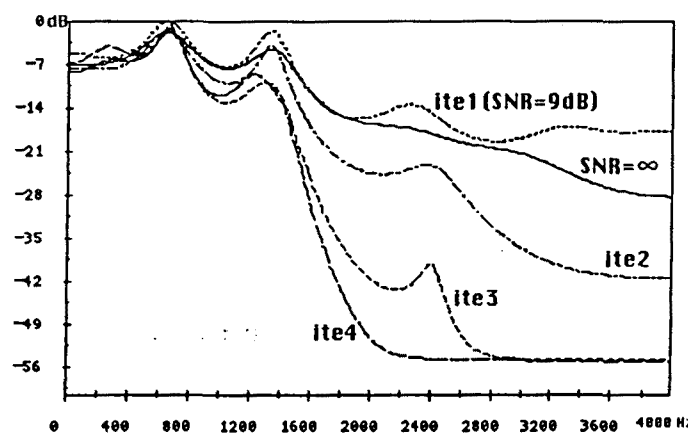


Figura IV.35 : Pérdida de reconocimiento del locutor por sobreestimación del espectro (método AR2). Los dos primeros formantes, de mayor energía, se constituyen como dominantes a partir de la segunda iteración.

deficiente estimación inicial del espectro (el ruido engaña al sistema estimador), como producto de un efecto exagerado de picado sobre los formantes de la voz que se encuentran en esa zona y que son más vulnerables debido a su menor SNR.

El otro efecto del que vamos a hablar en este apartado es la pérdida del poder de identificación de la persona que habla, que a pesar de ser mucho menos molesto que el ruido musical, puede ser bastante indeseado en ciertas aplicaciones. Al seguir iterando, el ruido musical tiende a desaparecer, pero también, se pierde el timbre de la voz y resulta dificultoso el reconocimiento de un locutor determinado.

El proceso de filtrado para una trama de señal donde la voz tenga un nivel de energía similar al del ruido (SNR_G=0dB), produce que el primer o primeros formante (los de mayor energía) queden muy bien determinados; por contra, su dominio energético sobre el resto del espectro y sobre el ruido nos llevarán a una señal más ronca. A partir de una tercera iteración se van perdiendo los formantes de menor energía (los de frecuencia más elevada), lo que conduce a una señal con dominio marcadamente más grave, una voz menos rica espectralmente. Se produce también este efecto para SNR's mayores, producto de una sobreestimación del espectro, cuando el sistema es demasiado agresivo, tal como podemos observar en la Fig.IV.35. Este efecto, juntamente con los restantes efectos negativos citados previamente, son los que producen la pérdida de identificación del locutor. Obsérvese como las dos primeras iteraciones consiguen ceñirse bastante correctamente a la señal original. Sin embargo, al seguir iterando, la zona alta del espectro tiende a desaparecer e, incluso, el segundo formante tiende a atenuarse y a ser absorvido por el primero. Esta supresión de los formantes situados en la parte alta del espectro conduce a una señal marcadamente más grave.

IV.5.2.5. Restricciones para controlar la Distorsión Espectral.

En este apartado vamos a exponer de una forma superficial las restricciones propuestas por Hansen y Clements para reducir estos efectos distorsivos cuando se consideran estadísticas de segundo orden. Ciertas variantes de la técnica básica de Lim y Oppenheim (Fig.II.16) pretenden eliminar el ruido sin provocar la aparición de efectos indeseados, como la distorsión por picado espectral y la aparición del ruido musical. Hansen y Clements en [Hansl-91] sugirieron aplicar una serie de restricciones al modelado AR para evitar la

reducción del ancho de banda de los formantes, el desplazamiento de sus localizaciones y otros efectos indeseados.

Tal como se ha expuesto en el Capítulo II, al procesar señal de voz se dispone de cierta información de antemano, ya que el tracto vocal es un sistema mecánico y como tal tiene una cierta inercia que imposibilita la existencia de cambios bruscos. La aplicación de ligaduras al sistema trata de aprovechar estos conocimientos previos acerca de la señal de voz, y de esta manera, acotar la estimación dentro de unos márgenes permitidos: cuantas más restricciones se impongan más restringido será el márgen de posibles soluciones. Las restricciones impuestas en [Hansl-91] se centran en las propiedades espectrales de los sonidos sonoros. Este conjunto de limitaciones pretenden acotar la señal en el dominio frecuencial, su envolvente LPC. Así, por ejemplo, los polos no pueden estar en cualquier posición ni tener un ancho de banda cualquiera para los formantes.

Su procedimiento consiste en efectuar inicialmente una transformación de coeficientes, pasando de los coeficientes del predictor LPC a los coeficientes LSP. Ello se realiza simplemente por una razón básica: los coeficientes LSP proporcionan una mejor interpretación sobre donde se actua y una menor complejidad computacional en relación al cálculo y uso de las raíces (polos del modelo) del polinomio del predictor lineal **A(z)**.

Cuando tenemos estos coeficientes, entonces se actúa sobre ellos, a través de lo que llaman parámetros posición y parámetros diferencia. Los parámetros posición indican aproximadamente, la ubicación de las raíces de **A(z)** (frecuencias de los formantes). Los parámetros diferencia indican de alguna forma, o estan relacionados con el ancho de banda de éstos. Si aplicamos restricciones sobre los parámetros posición o los parámetros diferencia, estamos controlando, según exponen Hansen y Clements, la posición y el ancho de banda de los formantes del modelo de la señal de voz.

Algunas de las restricciones o ligaduras aplicadas son:

a) Alisado o promediado de los parámetros en diversas tramas consecutivas o en diversas iteraciones dentro de una misma trama [Arsl-94].

b) Vigilancia de los valores de los parámetros para que no entren en una "zona prohibida", evitando efectos distorsivos tales como el picado espectral o movimientos erráticos de los polos.

c) Promediado de la autocorrelación entre diversas tramas o dentro de la misma trama. Recuérdase que ambos autores limitan su estudio a la consideración de las estadísticas de segundo orden.

Además es posible utilizar otras estrategias combinadas con éstas como representa la posible segmentación de trama mediante longitudes variables, en función de cuán rápidamente varía el espectro: tramos de señal de variación lenta en el modelo pueden admitir tramas de mayor longitud temporal mientras que tramos con cambios rápidos requieren fragmentos más pequeños de señal para tener un modelado correcto.

Los resultados, que presentan los citados autores en su estudio, superan los del método clásico de Lim y Oppenheim. El principal problema que aparece, como muestran ellos mismos, viene dado por el fuerte incremento de coste computacional, pasando a una carga de cálculo entre el doble y el triple de la carga correspondiente a la técnica clásica, libre de restricciones. El procedimiento que ofrece los mejores resultados también es el que más número de iteraciones necesita (7).

La introducción de parámetros o condiciones en el algoritmo de estimación AR facilitan un mejor control sobre las estimaciones finales obtenidas y, de esta manera, reducir los efectos indeseados. En este trabajo se consideran, en el capítulo siguiente, algunas alternativas a estos algoritmos AR de orden superior: el Filtrado de Wiener Generalizado y la Ponderación Intertrama.

IV.6. El Algoritmo Híbrido AR3H.

A la vista de los resultados obtenidos durante la evaluación comparativa de los dos algoritmos de filtrado iterativo de Wiener, correlaciones y cumulantes, se pensó en aprovechar las características favorables asociadas a ambos métodos, uniendo la estimación de segundo orden y la de tercer orden en un solo algoritmo. La evaluación de las prestaciones correspondientes a dicho algoritmo fueron presentadas en [Masg-92b] y [Sala-93c].

El algoritmo AR3 es la técnica que obtiene una mayor reducción de ruido durante la primera iteración, o en todo caso durante las dos primeras iteraciones cuando el nivel de ruido es alto. Sin embargo, a partir de la segunda o tercera iteración el efecto reducción de ruido se equipara al efecto distorsión espectral. También se ha demostrado su mayor distorsión por picado espectral cuando se consideran algunas iteraciones del algoritmo iterativo, produciéndose especialmente a partir de la segunda iteración una clara divergencia entre el filtro de Wiener $\mathbf{W_i}$ y su filtro óptimo $\mathbf{W_{opt}}$, especialmente para aquellas frecuencias pertenecientes a los valles espectrales.

Estas características de los algoritmos AR2 y AR3 condujo a un primer algoritmo modificado que pretende aglutinar en un mismo algoritmo las ventajas propias de los algoritmos de segundo y tercer orden: el algoritmo híbrido AR3H. Este algoritmo pretende combinar la velocidad de convergencia característica del algoritmo AR3 con la menor distorsión propia del algoritmo AR2. Para ello se procesan las primeras I_3 iteraciones de cada trama aplicando el método basado en estadísticas de tercer orden (AR3) y en las restantes iteraciones se hace uso del algoritmo AR2. Este algoritmo pretende reducir la mayor parte de ruido durante las primeras iteraciones sin ocasionar una distorsión apreciable y, en consecuencia, se limita el parámetro I_3 a los valores 1 ó 2 segun el nivel de ruido a combatir. Para las restantes iteraciones, se ataca el bajo nivel de ruido, sobrante después de actuar el algoritmo AR3, mediante el algoritmo AR2 y, de este modo, se ocasiona una menor distorsión en relación al supuesto de continuar iterando con el algoritmo AR3. Nótese que el filtro de Wiener $\mathbf{W_i}$ tiende al mismo filtro de convergencia $\mathbf{W_\infty}$ correspondiente al algoritmo AR2 y, además, la velocidad de convergencia inicial se corresponde con la asociada al algoritmo AR3.

En la Fig.IV.36 se ha representado la reducción de ruido obtenida para el algoritmo híbrido AR3H durante las 4 primeras iteraciones. En la Fig.IV.36.a se ha considerado el algoritmo AR3 únicamente durante la primera iteración ($\mathbf{I_3=1}$), mientras en la Fig.IV.36.b se ha evaluado un factor $\mathbf{I_3=2}$.

Las prestaciones del algoritmo AR3H con $I_3=2$ sólamente superan las del algoritmo AR3H con $I_3=1$ para niveles altos de ruido ($SNR_G \leq 6dB$), aunque la distorsión ocasionada siempre es mayor para el supuesto con las dos primeras iteraciones de tercer orden ($I_3=2$).

El comportamiento del método para la distancia Cepstral en la Fig.IV.36 muestra como la distancia decrece fuertemente en la primera iteración, en la segunda mejora levemente y en la tercera apenas presenta variación, el sistema se satura, sobre todo a SNR altas ($SNR_G>12dB$).

En la Fig.IV.37 se han comparado los algoritmos AR2, AR3, AR4 y AR3H con $I_3=1$ para una iteración específica, en función de los niveles de ruido. La segunda iteración correspondiente a estos tres algoritmos se ha representado en la Fig.IV.37.a, mientras los valores de distancia Cepstrum para la cuarta iteración se encuentran en la Fig.IV.37.b. Ambos casos presentan un comportamiento similar: para SNR_G bajas el algoritmo AR3H alcanza prestaciones similares al mejor algoritmo en esta zona, el algoritmo AR3; y para SNR_G altas introduce una menor distorsión originando valores mejores que AR3 y peores a los de AR2. Para las distancias Cosh e Itakura se observa el mismo comportamiento.

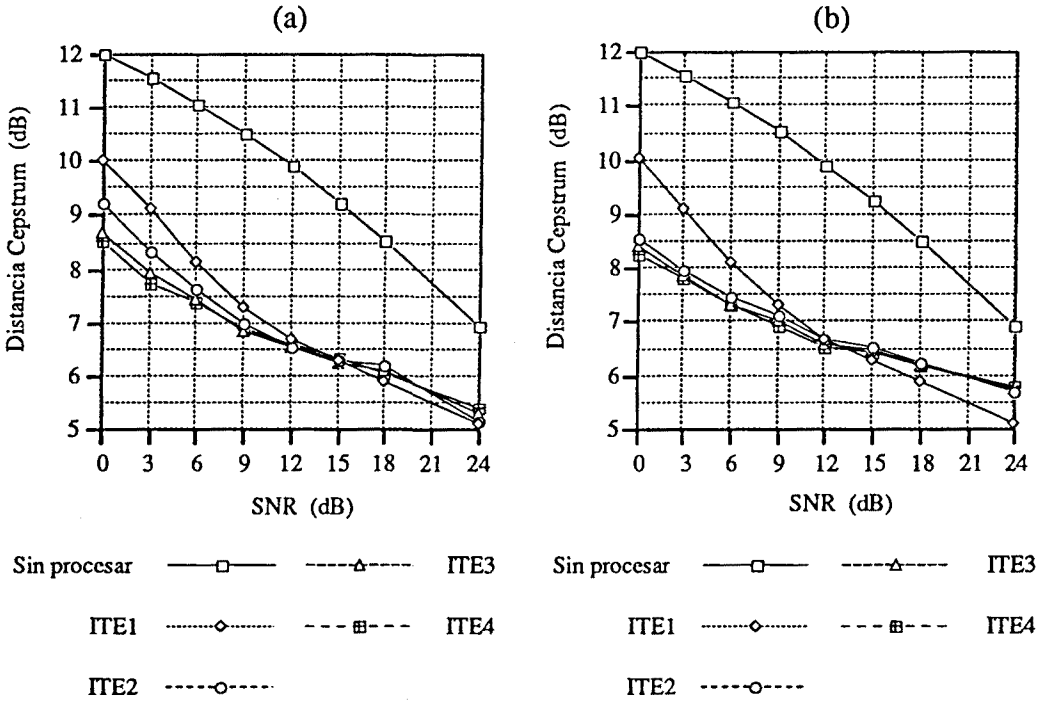


Figura IV.36 : Comportamiento del algoritmo AR3H durante las cuatro primeras iteraciones para los supuestos: a) $I_3=1$; b) $I_3=2$.

Se puede afirmar que se ha considerado un algoritmo híbrido de los algoritmos AR2 y AR3 para aprovechar las mejores características de ambos algoritmos. Aunque se ha presentado un algoritmo particular que combina el uso de los algoritmos básicos de segundo y tercer orden, la idea relacionada con la consideración de distintos algoritmos de estimación AR durante las distintas iteraciones de una misma trama, o para distintas tramas, queda como una linea de futuro del presente trabajo.

En realidad hemos visto que al combinar algoritmos distintos se pueden aprovechar las características más favorables de cada algoritmo, si se combinan de forma adecuada. En función de la aplicación concreta y del nivel de ruido existente, se puede elaborar un algoritmo a la medida. Incluso se puede pensar en un sistema adaptativo que combine los algoritmos y sus parámetros de forma adaptativa, a partir de ciertas medidas extraidas de la señal de voz ruidosa. En este caso también se debe considerar el destinatario final y el objetivo pretendido: un sistema de reconocimiento o un sistema de realce de la voz para mejorar la calidad o la inteligibilidad o para aminorar la fatiga del oyente.

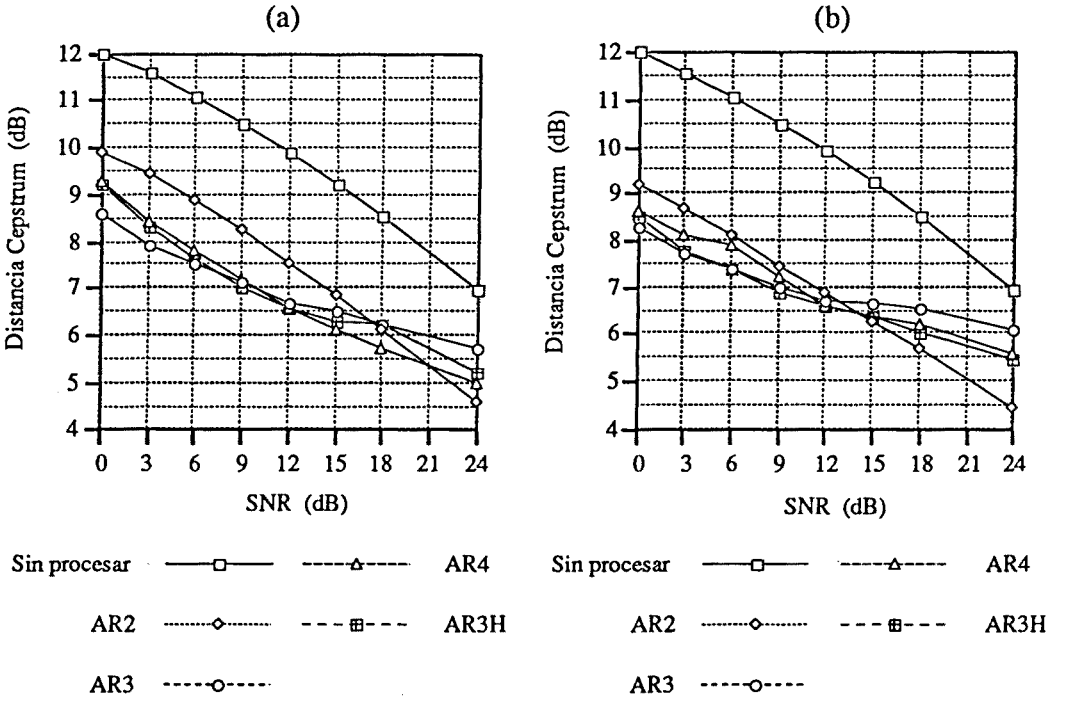


Figura IV.37 : Comparación de los algoritmos AR2, AR3, AR4 y AR3H con I_3=1 en función de distintos niveles de ruido tras procesar : a) 2 iteraciones ; b) 4 iteraciones.

En las pruebas realizadas con ruidos reales, tales como ruido de motor, de reactor o ruido paso bajo, este algoritmo híbrido no aporta prestaciones superiores puesto que obtiene el valor mínimo de distancia en la primera iteración y al añadir más iteraciones de segundo o tercer orden no aporta mejores notables. Estos resultados también muestran como el uso del algoritmo híbrido frena la degradación que sufre la señal al sobreiterar, especialmente para el algoritmo AR3.

IV.7. El Algoritmo Híbrido de Tercer y Cuarto Orden (AR34).

En la mayor parte de algoritmos evaluados se han obtenido, en presencia de un cierto nivel de ruido, superiores prestaciones en aquéllos que hacen uso de los cumulantes de tercer orden durante el proceso de estimación paramétrica AR, resultando unas medidas de distancia espectral bastante inferiores en relación a las obtenidas por aquellos algoritmos que hacen uso de los cumulantes de cuarto orden o de la función autocorrelación (segundo orden).

Estas medidas han sido obtenidas al evaluar globalmente un fichero de voz cuya duración suele oscilar entre 3seg. y 5seg. Es decir, representan una medida global sobre las prestaciones obtenidas para cada trama de voz, pudiendo existir alguna trama aislada donde el algoritmo fracace, sin que este hecho tenga grandes repercuciones en la medida de distancia objetiva global. Sin embargo, se ha comprobado [Vida-93] la existencia de tramas sonoras cuya skewness toma un valor demasiado pequeño. En estas condiciones los cumulantes de tercer orden estimados toman valores pequeños y, entonces, la estimación AR, resultante a partir de éstos, resulta de una calidad bastante deficiente.

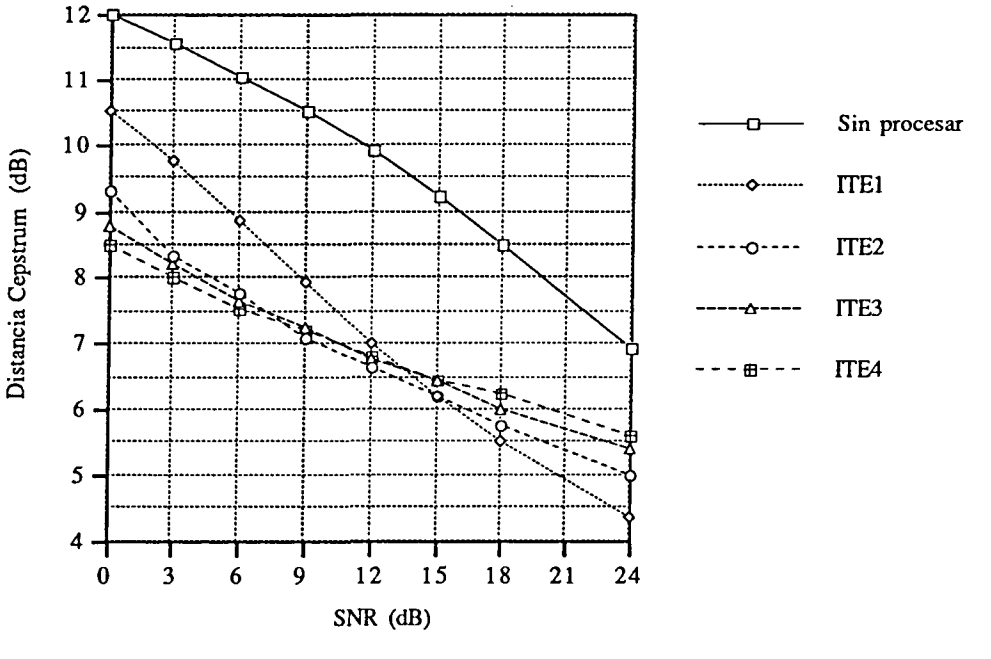


Figura IV.38 : Comportamiento del algoritmo AR34 en función de distintos niveles de ruido durante las cuatro primeras iteraciones.

Para corregir este comportamiento erróneo se pensó en extender el sistema de ecuaciones (III.99) de **p+1** slices de cumulantes de tercer orden mediante la sobredeterminación de otros **p+1** slices de cumulantes de cuarto orden. Así, en el caso de una trama cuyos cumulantes de tercer orden tomen valores demasiado pequeños se dispone también de los cumulantes de cuarto orden, donde no es posible la aparición de un valor tan pequeño para la kurtosis. y, en consecuencia, el método de mínimos cuadráticos, usado para resolver este sistema sobredeterminado, hace uso automáticamente de los slices de cuarto orden durante la estimación AR, desechando la consideración de los slices de cumulantes de cuarto orden por ser poco fiables.

La evaluación del comportamiento obtenido por este algoritmo AR34 se ha representado en la Fig.IV.38 para distintos niveles de ruido. Los resultados obtenidos muestran unas prestaciones muy similares al algoritmo AR4, aunque se puede situar en un nivel de calidad intermedio entre los algoritmos AR3 y AR4: los valores de distancia Cepstrum empeoran en relación al algoritmo AR3 aunque éstos son ligeramente mejores a los obtenidos por el algoritmo AR4. Así, se soluciona el problema debido a las tramas de baja skewnes pero, entonces, el comportamiento general tiende más hacia el algoritmo AR4, cuyas prestaciones son inferiores en relación al algoritmo AR3. Una posible razón que justifique esta superior tendencia hacia el algoritmo AR4 puede venir dada por la superior energía en los valores de los cumulantes de cuarto orden en relación a los valores correspondientes a los cumulantes de tercer orden.

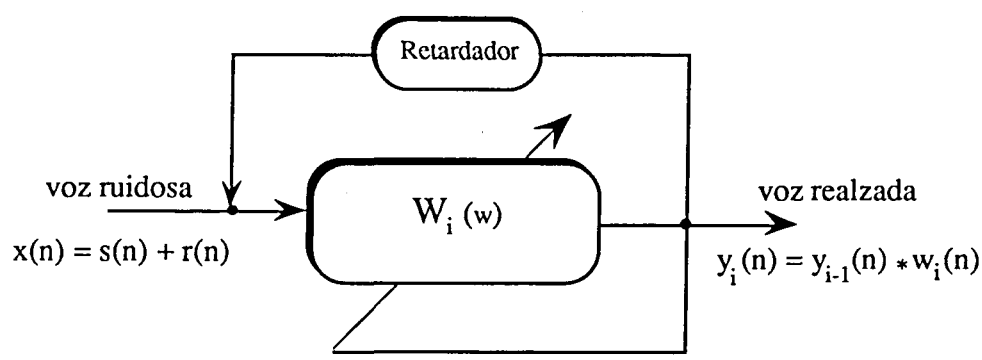
IV.8. El Filtrado de Wiener Realimentado (ARre).

Vamos a mostrar brevemente que sucede teóricamente cuando el filtro de Wiener se utiliza en modo realimentado, es decir, cuando la salida del filtro de Wiener se utiliza, no solamente para obtener la estimación AR de la siguiente iteración, sinó también como señal de entrada al filtro. Por consiguiente, no se filtra siempre la señal de voz ruidosa $\mathbf{x(n)=y_0(n)}$ en todas las iteraciones sinó que se filtra la señal de salida de la iteración precedente $\mathbf{y_i(n)}$. En la Fig.IV.39 se ha representado el esquema correspondiente a este Filtrado de Wiener realimentado, donde el retardador separa la señal de voz realzada correspondiente a la iteración i-ésima $\mathbf{y_i(n)}$ de la señal obtenida en la iteración precedente $\mathbf{y_{i-1}(n)}$, que a su vez es la entrada al filtro para la iteración i-ésima.

Seguidamente se realiza su estudio de convergencia de forma análoga a los algoritmos anteriores. En la primera iteración el filtro de Wiener se diseña como:

$$W_1 = \frac{P_x}{P_x + P_r} = \frac{1}{1 + \dfrac{P_r}{P_s + \partial . P_r}} = \frac{1}{1 + W_{opt}^{cc}} \qquad \text{(IV.94)}$$

En la segunda iteración, se diseña el filtro de Wiener a partir de la señal de voz realzada $\mathbf{y_1(n)}$ disponible a la salida del filtro de Wiener $\mathbf{W_1}$ perteneciente a la primera iteración:



$$W_i(w) = \frac{P_{y_{i-1}}(w)}{P_{y_{i-1}}(w) + P_r(w)} \qquad P_{y_{i-1}}(w) = \frac{g^2}{\left| 1 - \sum_{k=1}^{P} a_k \cdot e^{-jwk} \right|^2}$$

Figura IV.39 : Esquema del filtrado realimentado de Wiener con estimación AR.

$$W_2 = \frac{P_{y_1}}{P_{y_1} + P_r} = \frac{1}{1 + \frac{P_r}{P_{y_1}}} \qquad \text{(IV.95)}$$

Esta señal de voz realzada $\mathbf{y_1(n)}$ resulta de filtrar la señal de voz ruidosa $\mathbf{x(n)}$ mediante el filtro de Wiener diseñado durante la primera iteración y, entonces, en el dominio frecuencial se verifica:

$$P_{y_1} = P_x \cdot W_1^2 = (P_s + \partial . P_r) \cdot W_1^2 \qquad \text{(IV.96)}$$

sustituyendo (IV.96) en (IV.95) resulta una relación entre los filtros de Wiener correspondientes a las dos primeras iteraciones:

$$W_2 = \frac{1}{1 + \frac{W_{opt}^{cc}}{W_1^2}} = \frac{1}{1 + W_{opt}^{cc} \cdot D_1^2} \qquad \text{(IV.97)}$$

En la tercera iteración se diseña el filtro de Wiener a partir de la señal de voz realzada $\mathbf{y_2(n)}$ disponible a la salida del filtro de Wiener $\mathbf{W_2}$:

$$W_3 = \frac{P_{y_2}}{P_{y_2} + P_r} = \frac{1}{1 + \frac{P_r}{P_{y_2}}} \qquad \text{(IV.98)}$$

donde la señal de voz realzada $\mathbf{y_2(n)}$ resulta de filtrar la señal de voz realzada $\mathbf{y_1(n)}$, obtenida en la iteración anterior, mediante el filtro de Wiener diseñado durante la segunda iteración y, entonces, en el dominio frecuencial se verifica:

$$P_{y_2} = P_{y_1} \cdot W_2^2 = (P_s + \partial . P_r) \cdot W_1^2 . W_2^2 \qquad \text{(IV.99)}$$

sustituyendo (IV.99) en (IV.98) resulta una relación entre los filtros de Wiener correspondientes a las tres primeras iteraciones:

$$W_3 = \frac{1}{1 + \frac{W_{opt}^{cc}}{W_1^2 . W_2^2}} = \frac{1}{1 + W_{opt}^{cc} \cdot D_1^2 . D_2^2} \qquad \text{(IV.100)}$$

En general para la iteración i-ésima se obtiene la siguiente estimación para el filtro de Wiener:

$$W_i = \frac{1}{1 + W_{opt}^{cc} \cdot \prod_{j=1}^{i-1} D_j^2} \qquad \text{(IV.101)}$$

observándose una ecuación de recurrencia en términos del filtro de Wiener inverso $\mathbf{D_i}$:

$$D_i = 1 + W_{opt}^{cc} \cdot \prod_{j=1}^{i-1} D_j^2 \qquad \text{(IV.102)}$$

Aplicando los cambios de variable propuestos en (IV.82) se obtiene la siguiente ecuación de recurrencia:

$$d(i) = 1 + r' \cdot \prod_{j=1}^{i-1} d^2(j) \qquad \text{(IV.103)}$$

para un número de iteraciones **i** tendiendo a infinito y cuando la solución **d(∞)** converge se verifica la relación:

$$d(\infty) = 1 + r' \cdot \prod_{j=1}^{\infty} d^2(j) \qquad ,, \qquad 0 \le r' \le 1 \qquad \text{(IV.104)}$$

Esta serie sólamente puede converger si **d(∞)** toma el valor cero o la unidad:

$$\text{si } d(\infty) = 0 \Rightarrow \quad 0 = 1 + r' \cdot \prod_{j=1}^{\infty} d^2(j) \quad \Rightarrow \quad d(\infty) \text{ diverge} \qquad \text{(IV.105.a)}$$

$$\text{si } d(\infty) = 1 \Rightarrow \quad 0 = r' \cdot \prod_{j=1}^{\infty} d^2(j) \quad \Rightarrow \quad d(\infty) \text{ converge si } r' = 0 \qquad \text{(IV.105.b)}$$

Este resultado nos indica que el filtro de Wiener realimentado converge al filtro óptimo de Wiener sólamente para aquellas frecuencias donde SNR=∞, es decir, cuando el filtro óptimo vale la unidad. Para las restantes frecuencias el filtro de Wiener realimentado tiende a cero

Análogamente, segun los valores de **r'** la solución **d(∞)** adopta valores distintos:

$$W_\infty = \begin{cases} 0 & \text{si} \quad W_{opt} < 1 \\ 1 & \text{si} \quad W_{opt} = 1 \end{cases} \qquad \text{(IV.106)}$$

Es decir, se preserva la señal sólo en aquellos puntos donde no haya ruido, y en el resto la señal de salida se va anulando paulatinamente con el discurrir de las iteraciones. De esta manera se obtiene una fuerte degradación de la señal de voz, con un efecto de picado espectral mucho más acentuado y un total aislamiento de los formantes. A priori era

predicible esta superior distorsión puesto que a la distorsión ocasionada por el filtro se le añade de forma acumulativa la distorsión originada en la señal filtrada en cada iteración. Nótese, también, que el algoritmo de Wiener realimentado converge al mismo filtro para los casos de estadísticas de segundo orden (AR2re) y tercer orden (AR3re). En resumen se puede afirmar que esta configuración acentúa el efecto de picado espectral de los formantes y tiende hacia un filtro en peine que sólamente deja pasar los picos donde la relación señal a ruido sea muy elevada (SNR —> ∞).

CAPITULO V

El Algoritmo Iterativo de Wiener Generalizado.

Vimos en el Capítulo II cómo lo que llamamos 'generalización' del filtro de Wiener clásico utilizado en el algoritmo básico de Lim-Oppenheim podía aportarnos una serie de mejoras sin apenas representar un incremento en el tiempo de cálculo invertido en la estimación del filtro. Recordemos que tal generalización se limitaba a la inclusión de dos nuevos parámetros en la expresión del filtro, uno exponencial, a, y otro multiplicativo, b:

$$W_i(w) = \left(\frac{P_{y_{i-1}}(w)}{P_{y_{i-1}}(w) + \beta . P_r(w)} \right)^{\alpha} \qquad (V.1)$$

El objetivo y funciones de ambos parámetros fueron explicados en el segundo Capítulo (subapartado II.3.4.1.). En principio el parámetro b representa una sobreestimación de ruido cuando toma valores superiores a la unidad o una subestimación de ruido en caso contrario (b<1). Así, tomando valores b>1 el filtro de Wiener cree que hay un ruido de nivel superior y, entonces, intenta suprimir más ruido del que realmente está presente en un entorno determinado. Es decir, se dota de mayor agresividad al filtro de Wiener cuando se consideran sobreestimaciones de ruido y, en el caso opuesto, este filtrado se muestra más conservador ante las subestimaciones de ruido. Por otra parte, valores del parámetro a superiores a la

unidad dotan de mayor agresividad al filtrado puesto que acentúa la importancia relativa de las frecuencias con una mayor densidad espectral de potencia frente a las zonas frecuenciales de menor energía. A su vez valores $\alpha<1$ equivalen a un realce de las zonas espectrales de menor energía ya que se verifica $|W(w)|\leq 1$ para cualquier frecuencia del espectro. Es decir, el filtro resultante será menos selectivo en frecuencia debido a este suavizado. De esta forma, la consideración de estos dos parámetros permite un mayor control del filtrado iterativo de Wiener ante ruidos de distintos niveles y características. Un detallado estudio sobre los algoritmos AR3 y AR4 generalizados fue publicado en [Sala-94a].

El análisis de este algoritmo nos permitirá averiguar la respuesta a algunas cuestiones previas que pueden ser de interés para el lector. ¿Es conveniente considerar sobreestimaciones de ruido para alcanzar una reducción de ruido más rápida? La consideración de valores $\alpha>1$ combinado con el uso de estadísticas de orden superior, menos sensibles a la presencia de ruido ¿Pueden originar un filtro demasiado agresivo? ¿Resultaría más adecuada, en este supuesto, compensar esta capacidad de las estadísticas de orden superior mediante valores $\alpha<1$? ¿Cuán importante es el efecto correspondiente a la distorsión ocasionada? ¿Se puede eliminar totalmente el ruido residual dotando de mayor agresividad al filtro de Wiener? ¿Se puede alcanzar una convergencia más rápida y, en consecuencia, reducir el número de iteraciones a procesar? ¿A cuál parámetro, α o β, se muestra más sensible el Filtro de Wiener Generalizado? Estas y otras cuestiones intentan encontrar una respuesta a lo largo del presente capítulo.

V.1. El Método AR3 Generalizado.

A continuación se evalúa el método de filtrado iterativo de Wiener, descrito en el capítulo anterior, cuando se ha generalizado mediante la introducción de dos parámetros a y b. Se estudia su comportamiento cuando se utilizan estadísticas de tercer orden durante la estimación espectral de la señal de voz. Este algoritmo se nota como AR3 puesto que se fundamenta en un modelado AR de la señal de voz obtenido a partir de los cumulantes de tercer orden. Los resultados que se discuten han sido obtenidos a partir de la aplicación de las ecuaciones de Yule-Walker de tercer orden (III.96) para el proceso de cálculo de los coeficientes a_k.

Parece lógico suponer que el filtro iterativo de Wiener puede comportarse de forma desigual ante distintos niveles de ruido, para unos mismos valores prefijados de los parámetros a y b. Por esta razón se ha estudiado su comportamiento en tres ambientes diferenciados:

- entornos muy ruidosos (SNR=0dB),
- entornos con un nivel intermedio de ruido (SNR=9dB),
- entornos poco ruidosos (SNR=18dB).

Todos los resultados, que se exponen a continuación corresponden a ruido blanco Gaussiano de media nula, que ha degradado aditivamente la señal de voz. Los resultados obtenidos al aplicar otros tipos de ruido, tales como ruido Gaussiano coloreado o ruido de motor de coche, presentan una tendencia de comportamiento similar, aunque ésta es menos notoria en sentido cuantitativo debido, básicamente, a sus propias características espectrales.

La distancia Itakura privilegia las zonas correspondientes a los picos espectrales, pudiendo darse el supuesto que un buen comportamiento en estas zonas enmascare totalmente un comportamiento deficiente en las zonas correspondientes a los valles espectrales. En cambio, la distancia Cepstrum trata de forma más homogenea lo que sucede en las distintas zonas del espectro. Por esta razón se ha tomado la distancia Cepstrum como guía para comentar el comportamiento de este algoritmo en los distintos entornos considerados. No obstante, cuando se crea significativo también se mencionarán los valores correspondientes a otras medidas de distancias espectrales (Itakura, Cosh) o temporales (SNR Global, SNR Segmentada), así como las conclusiones derivadas de las pruebas de audición realizadas. Los resultados que se presentan corresponden a un locutor en inglés (Fichero ESCA) que ha sido

seleccionado por ser uno de los ficheros que mejor aproxima sus valores cuantitativos de distancia al comportamiento observado durante las pruebas de audición.

Este algoritmo iterativo de Wiener de tercer orden (AR3) ha sido evaluado en la zona del plano a-b dada por:

$$0.5 \leq a \leq 1.4$$
$$0.2 \leq b \leq 2.0 \qquad \text{(V.2)}$$

considerándose incrementos de 0.1 para el parámetro **a** e incrementos de 0.2 cuando se trate del parámetro **b**. Nótese que cada representación se compone de un total de 100 medidas correspondientes a filtros independientes de características distintas y cuya señal de entrada es común a todos éllos. Cada combinación de dos valores de estos dos parámetros se interpreta como un punto **(a, b)**, perteneciente al plano bidimensional **a-b**. Asimismo, la distancia Cepstrum asociada a este punto **(a, b)** se indica como C_i **(a, b)**, siendo **i** el número de la iteración considerada, según la notación introducida en el capítulo anterior. También se han denominado C_{MAX} y C_{MIN} los valores máximo y mínimo, respectivamente, de distancia Cepstrum dentro de la región (V.2), para tener una cierta orientación respecto los diferentes comportamientos ofrecidos por este filtrado parametrizado, según cada par **(a, b)** considerado.

Para cada caso concreto, estas medidas de distancia Cepstrum se han materializado gráficamente mediante una representación tridimensional, que ofrece una interpretación visual más directa, y una representación plana mediante curvas de nivel, que aporta una visión más detallada de la variación de esta medida a través de la zona (V.2) del plano **a-b**. Para dotar a la representación mediante curvas de nivel de una mejor interpretación se han situado flechas indicando la dirección hacia dónde la distancia Cepstrum tiende a disminuir. Es decir, se han indicado las regiones donde se obtiene una mayor reducción de ruido. Además, mediante el símbolo M se ha situado el punto **(a, b)**, de la representación por curvas de nivel, donde se obtiene un valor mínimo absoluto de distancia Cepstrum dentro de la región (V.2).

V.1.1. Ambientes Altamente Ruidosos.

Se ha considerado una relación señal a ruido global de 0dB para simular un entorno muy ruidoso. Evidentemente se trata de un caso muy desfavorable puesto que la potencia del ruido se equipara a la correspondiente a la señal de voz. En consecuencia, las estadísticas de tercer orden logran, en estas condiciones extremas, una clara superioridad respecto a las clásicas de segundo orden. El valor de distancia Cepstrum inicial, asociada con esta SNR=0dB, equivale a C_0=12.02dB.

Después de procesar la primera iteración, Fig.V.1, se observa una reducción bastante

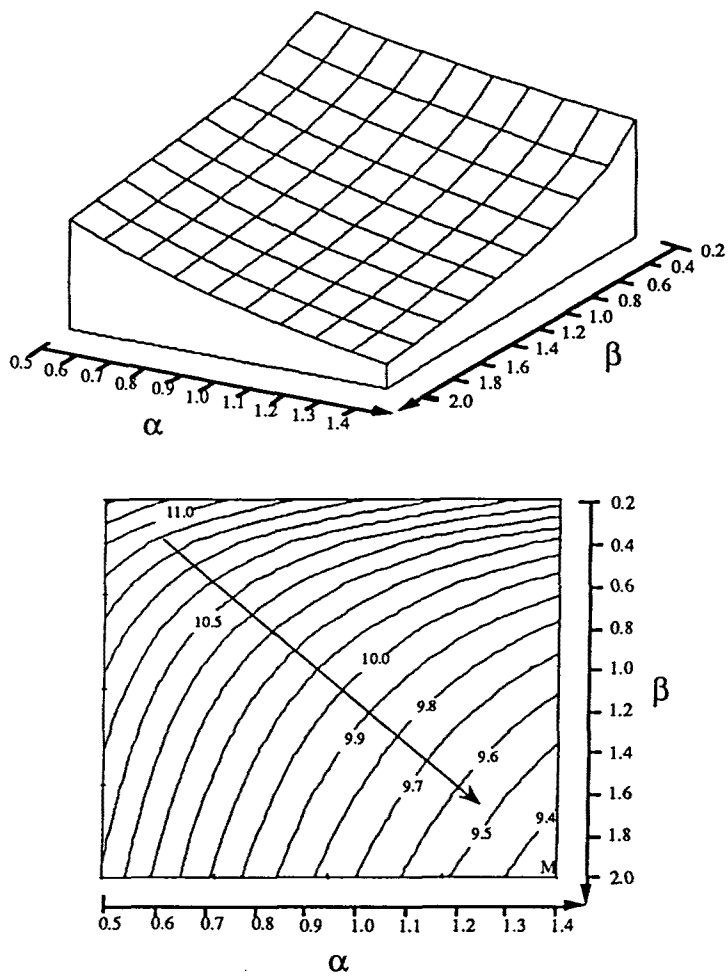


Figura V.1 : Distancia Cepstrum (dB) después de procesar la primera iteración mediante Filtro de Wiener Generalizado a SNR=0dB para ESCA+AGWN.

uniforme del nivel de ruido al aumentar los valores de a y b, desde su valor de máxima distancia $C_{MAX}= C_1$ (0.5, 0.2)= 11.22dB hasta su valor mínimo $C_{MIN}= C_1$ (1.4, 2)= 9.31dB. Durante esta primera iteración se aprecia como el efecto de supresión de ruido es el principal efecto y domina claramente sobre la distorsión introducida, debido principalmente al enorme nivel del ruido presente. Los valores de (a, b) elevados dotan de mayor agresividad al filtro supresor de ruido y, por esta razón, conducen a una mayor reducción de ruido (2.7dB).

En el lado opuesto, valores pequeños de los parámetros a y b originan un filtro muy conservador y conducen a una reducción de ruido poco significativa (inferior a 1dB). Obsérvese que la diferencia entre los valores C_{MIN} y C_{MAX} es bastante notoria (1.9dB). No debe olvidarse que valores elevados de estos parámetros llevan asociada una mayor distorsión, aunque ésta empieza a ser significativa a partir de la segunda iteración, tal como se

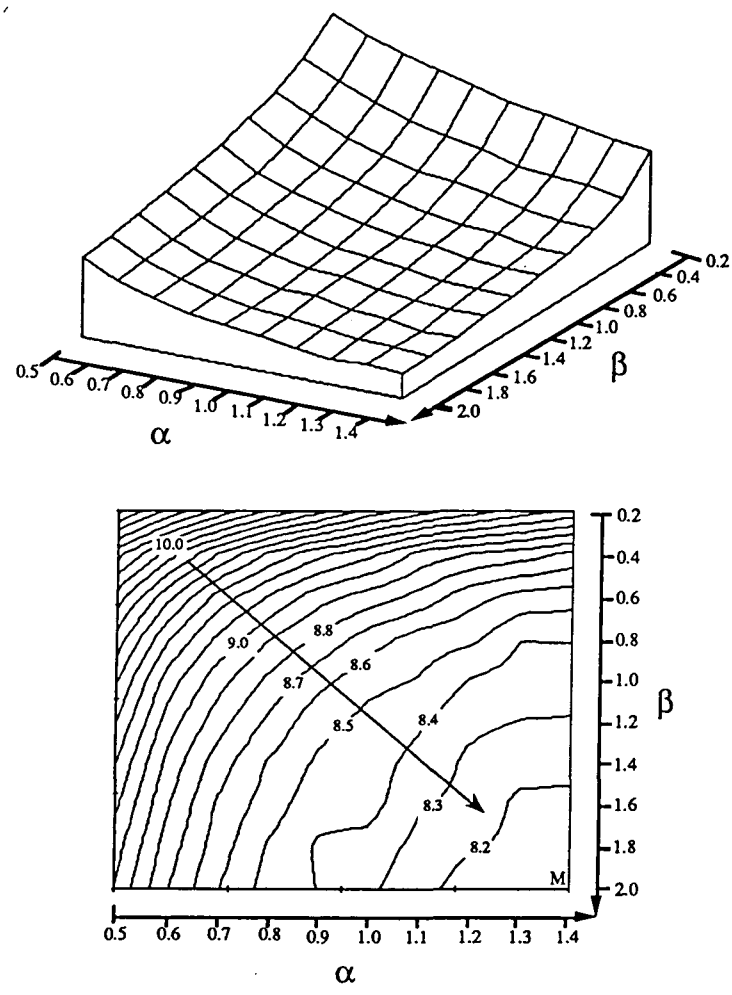


Figura V.2 : Distancia Cepstrum (dB) después de procesar la segunda iteración

ha comentado préviamente en el capítulo anterior. Nótese, además, como mediante estos dos parámetros se controla la agresividad del filtro: un valor elevado de **a** combinado con un valor pequeño de **b** origina un comportamiento equivalente al obtenido mediante un valor pequeño de **a** y una **b** alta. Por esta razón, las curvas de nivel tienden a tomar forma diagonal-circular sobre la región (V.2).

Tras la segunda iteración del algoritmo de tercer orden AR3, Fig.V.2, el nivel de distancia decrece desde el valor máximo $C_{MAX}= C_2$ (0.5, 0.2)= 10.62dB hasta el mínimo $C_{MIN}= C_2$ (1.4, 2.0)= 8.13dB. Aunque la posición de estos puntos máximo y mínimo no ha variado, la diferencia entre estos dos valores se ha incrementado a 2.5dB debido a que el efecto supresión de ruido continua predominando y ya se han acumulado dos filtrados de tendencias muy diferentes. Sin embargo, para valores elevados de **a** y **b** aparece una cierta saturación en el nivel de la distancia Cepstrum debido a que los efectos de supresión de ruido y distorsión ocasionada empiezan a equilibrarse. Este hecho se constata en una mayor separación entre las curvas de nivel correspondientes a la zona $20a + 12b \geq 39$.

Al mismo tiempo, en la zona de valores pequeños para ambos parámetros **a** y **b** se produce un cierto acercamiento entre curvas de nivel, originando una zona de pendiente elevada, puesto que valores un poquito más agresivos marcan claras diferencias, en relación a sus vecinos, tras acumularse las mejoras de procesado correspondientes a cada una de las dos iteraciones realizadas. En esta zona el nivel de ruido remanente continua siendo significativo y, en consecuencia, el efecto supresor de ruido domina claramente respecto la distorsión producida, que en esta zona es menos importante en relación a la esquina opuesta.

La confirmación de este efecto saturación entre efectos contrarios, reducción de ruido y distorsión, se aprecia en la tercera iteración (Fig.V.3): los resultados correspondientes a la zona más agresiva empiezan a deteriorarse. Se puede decir que la mayor parte de ruido ha sido suprimido tras las dos primeras iteraciones y en la tercera iteración esta reducción de ruido empieza a ser superada, en términos de distancia Cepstrum, por el efecto distorsión. Debe recordarse que al efecto distorsión inherente del algoritmo iterativo de Wiener, acrecentado en esta situación mediante valores agresivos de **a** y **b**, se le acumula el mayor efecto distorsionador ocasionado por los cumulantes de tercer orden. De esta manera el valor mínimo $C_{MIN}= C_3$ (1.2, 1.0)= 8.02dB se alcanza en el interior de la región (V.2) y no en su extremo más agresivo C_3 (1.4, 2.0)= 8.49dB, tal como sucedía en las dos iteraciones previas. Este mismo efecto se aprecia también en las medidas de distancia Itakura y Cosh, pudiéndose deducir que esta distorsión afecta de forma similar tanto a las regiones de los formantes como a los valles espectrales del espectro de la voz.

El valor máximo C_{MAX}= C_3 (0.5, 0.2)= 10.13dB sigue localizado en el mismo lugar y sigue disminuyendo, iteración a iteración, de forma uniforme pero lenta. En esta zona conservadora el nivel de ruido a suprimir es todavía considerable y por esta razón se marcan diferencias apreciables ante variaciones pequeñas de los parámetros a y b, originando una gran acumulación de curvas de nivel en esta esquina superior izquierda. La diferencia entre los valores C_{MAX} y C_{MIN} se ha reducido a 2.1dB debido a dos razones principales: la existencia de zonas donde la distorsión supera al efecto reductor de ruido y la paulatina, aunque lenta, reducción de ruido en cada iteración correspondiente a la zona más lenta.

En la cuarta iteración, Fig.V.4, se observa como la distorsión supera la reducción de ruido en el interior de la zona agresiva, que se traslada un poco hacia al interior. Este

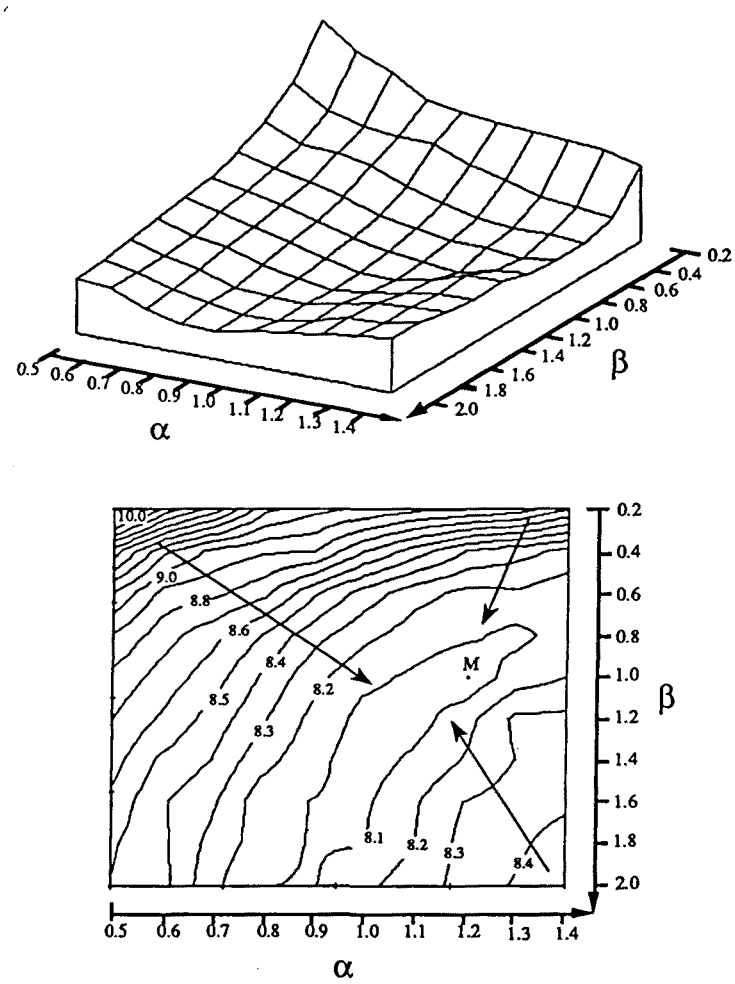


Figura V.3 : Distancia Cepstrum (dB) después de procesar la tercera iteración mediante Filtro de Wiener Generalizado a SNR=0dB para ESCA+AGWN.

deterioro origina un incremento máximo de unos 0.2dB en la distancia Cepstrum durante las iteraciones cuarta y quinta: C_4 (1.4, 2.0)= 8.67dB y C_5 (1.4, 2.0)= 8.81dB. Obviamente, este número de iteraciones se muestra claramente abusivo para valores elevados de a y b.

La zona de saturación, o de equilibrio entre ambos efectos opuestos, se traslada un poco hacia la izquierda donde los valores (a, b) se muestran menos agresivos, localizándose en forma de valle diagonal alrededor de las rectas $8a + b = 8.4$ para $b \geq 1.2$ y $6a + 5b = 11.4$ si $0.6 \leq b \leq 1.2$. El valor mínimo se alcanza en $C_{MIN} = C_4$ (0.9, 1.2)= 8.13dB, habiendo empeorado 0.1dB con respecto al valor mínimo de la iteración precedente. Esto se debe a que la distorsión inherente a cada iteración del algoritmo afecta a los pares (a, b) menos agresivos que precisan de un mayor número de iteraciones para alcanzar su valor mínimo en la iteración óptima. Estas características también se cumplen en la quinta iteración: La zona de equilibrio

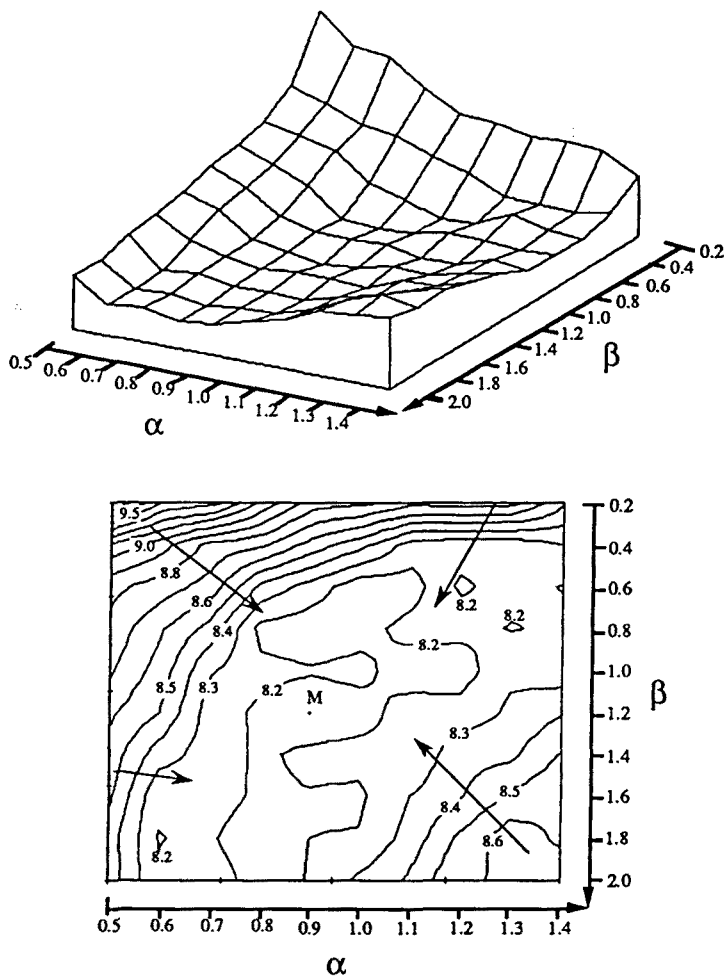


Figura V.4 : Distancia Cepstrum (dB) después de procesar la cuarta iteración mediante Filtro de Wiener Generàlizado a SNR=0dB para ESCA+AGWN

se ensancha un poco más y se traslada ligeramente hacia la izquierda y, también, su valor mínimo empeora suavemente $C_{MIN}= C_5$ (0.8, 1.8)= 8.15dB.

La zona de valores pequeños, puntos (α, β) pertenecientes a la región $\alpha \leq 0.7$ y $\beta \leq 0.4$, va mejorando lentamente y, además, esta mejora se va reduciendo con el transcurrir de las sucesivas iteraciones. En consecuencia, los valores máximos todavía se encuentran por encima de los 9dB tras la cuarta iteración $C_{MAX}= C_4$ (0.5, 0.2)= 9.71dB y la quinta iteración $C_{MAX}= C_5$ (0.5, 0.2)= 9.45dB. Así, el nivel de ruido presente en la señal de voz realzada es todavía considerable, después de procesar cinco iteraciones del algoritmo AR3, y se precisa de un número superior a las doce iteraciones para llegar a su valor de distancia mínima, tal como se muestra en la Tabla V.1. Aparte del efecto distorsionador ocasionado por este número elevado de iteraciones, aparece otro factor negativo a considerar: su coste de cálculo les hace claramente inviables ante posibles aplicaciones reales.

Para concluir este estudio del algoritmo AR3 generalizado correspondiente a ambientes muy ruidosos, en la Tabla V.1 se ha anotado la mínima distancia Cepstrum obtenida para cada par (α, β) y entre paréntesis se ha indicado el número óptimo de iteraciones necesario para alcanzarlo. Además estos valores mínimos correspondientes a la iteración óptima se han representado gráficamente en la Fig.V.5. Una primera conclusión, que se puede extraer en la zona de valores pequeños $2\alpha + \beta \leq 2.4$, viene dada por el elevado número de iteraciones necesario que ocasiona un insuperable coste de cálculo. Además, en la mayor parte de estos puntos, la distancia mínima alcanzada es demasiado elevada, superando incluso los 9dB en algunos casos $C_{MAX}=C_{13}$ (0.6, 0.2)= 9.11dB. Aunque el valor mínimo absoluto se alcanza en esta región $C_{MIN}= C_{17}$ (0.6, 1.2)= 7.95dB, se deben procesar 17 iteraciones. Los puntos pertenecientes a esta región conservadora progresan muy despacio, apenas 0.1dB por iteración después de la quinta o sexta iteración, hasta llegar a su valor mínimo correspondiente a la iteración óptima y, entonces, al seguir iterando su distancia Cepstrum se mantiene bastante invariante, a diferencia de lo que sucedía para valores (α, β) agresivos. Este comportamiento es el responsable del salto brusco que experimenta la iteración óptima, representada en la Tabla V.1, entre esta región y la correspondiente a valores intermedios de α y β, donde se precisan de 5 o 7 iteraciones, a lo sumo, para alcanzar la iteración óptima.

En la región opuesta, valores elevados de ambos parámetros, $5\alpha + 2\beta > 9$, originan una gran supresión de ruido (3.85dB) durante las dos primeras iteraciones y, entonces, su notoria agresividad ocasiona una significativa distorsión a partir de la tercera iteración y, en consecuencia, los valores de distancia Cepstrum se deterioran cuando se procesan más de dos iteraciones. Esta zona es la que requiere un menor tiempo de cálculo, sin embargo, los test de

audición muestran un nivel de distorsión ligeramente elevado y el ruido musical remanente todavía es bastante perceptible.

Entre las dos zonas anteriores, en un valle diagonal alrededor de la recta $2\alpha + \beta = 3.4$, se encuentra la región que mejor negocia un compromiso entre las ventajas e inconvenientes anteriormente citados: se alcanzan valores óptimos (la distancia Cepstrum se reduce 4dB) tras un tiempo de cálculo razonable (3 iteraciones) sin una pérdida de inteligibilidad significativa. Después de procesar tres iteraciones se alcanzan valores inferiores a 8.1dB, destacando su valor mínimo C_3 (1.2, 1.0)= 8.02dB y los valores C_3 (1.0, 1.2)= 8.07dB, C_3 (1.0, 1.4)= 8.04dB que implican un menor coste de cálculo debido a que el factor exponencial α del filtro vale la unidad. El punto (1.0, 1.0) correspondiente al filtrado clásico de Wiener pertenece a esta zona C_3 (1.0, 1.0)= 8.12dB pero, en cambio, el punto (0.5, 1.0) correspondiente al filtrado por

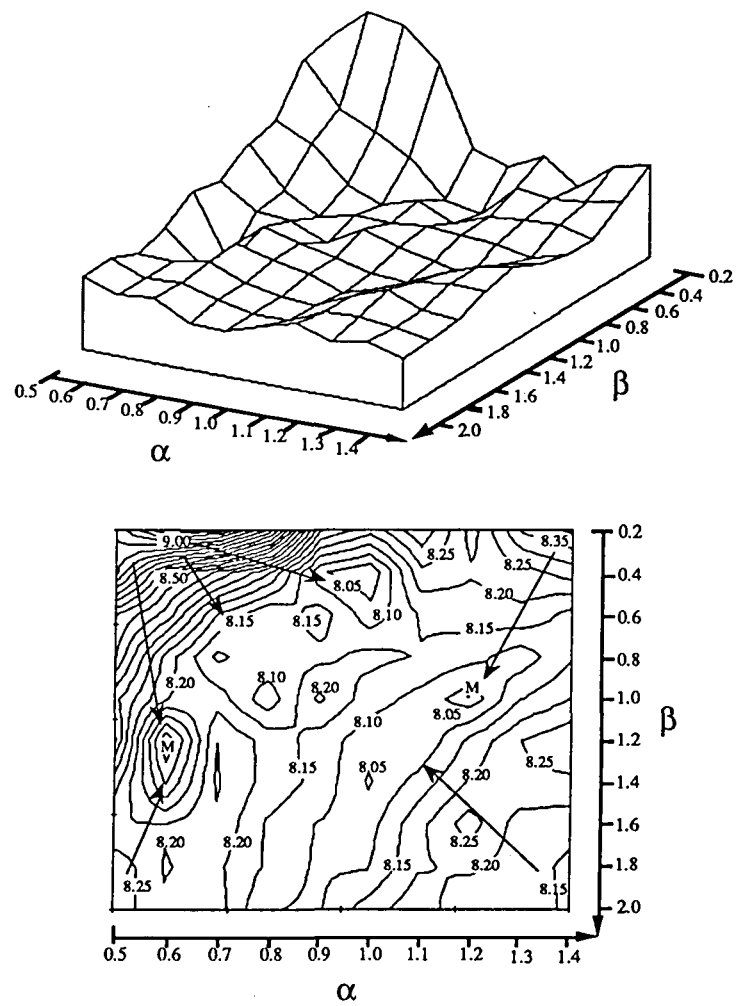


Figura V.5 : Distancia Cepstrum (dB) después de procesar la iteración óptima mediante Filtro de Wiener Generalizado a SNR=0dB para ESCA+AGWN.

espectro de potencia, también denominado como No Wiener durante la discusión del algoritmo Híbrido de tercer orden (apartado IV.6), pertenece a la zona lenta que conlleva un coste de cálculo excesivo C_{16} (0.5, 1.0)= 8.36dB.

Si analizamos los valores obtenidos en esta zona de equilibrio por las distancias espectrales Itakura y Cosh, se observa que siguen mejorando algunas iteraciones después de superar la iteración óptima en distancia Cepstrum. Como estas medidas prestan más atención a la zona del espectro correspondiente a los formantes, resulta hasta cierto punto lógico que acusen en menor medida los efectos de picado espectral, especialmente dañinos para las zonas correspondientes a los valles espectrales. Las medidas temporales, por su parte, confirman la bondad de la zona de equilibrio en comparación a la zona más agresiva, resultando valores de SNR global y SNR segmentada superiores en 2dB. No obstante, las mejores medidas temporales se obtienen en la zona lenta debido a la menor distorsión ocasionada, alcanzando

3		A			L		F		A		
	0dB	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4
	0.2	8.91 (14)	9.11 (13)	9.06 (18)	8.83 (8)	8.30 (19)	8.22 (14)	8.34 (7)	8.18 (17)	8.36 (7)	8.41 (5)
B	0.4	8.86 (19)	8.55 (16)	8.44 (19)	8.31 (17)	8.06 (14)	8.01 (13)	8.24 (6)	8.21 (12)	8.23 (4)	8.27 (3)
	0.6	8.55 (12)	8.43 (19)	8.19 (16)	8.10 (14)	8.17 (7)	8.07 (12)	8.16 (4)	8.17 (3)	8.19 (3)	8.13 (3)
E	0.8	8.50 (13)	8.23 (15)	8.09 (13)	8.14 (13)	8.13 (4)	8.17 (4)	8.15 (3)	8.12 (3)	8.07 (3)	8.14 (3)
	1.0	8.36 (16)	8.20 (14)	8.18 (13)	8.07 (16)	8.21 (4)	8.12 (3)	8.08 (3)	8.02 (3)	8.15 (3)	8.20 (3)
T	1.2	8.40 (12)	7.95 (17)	8.25 (4)	8.18 (3)	8.13 (4)	8.07 (3)	8.05 (3)	8.16 (3)	8.25 (3)	8.29 (2)
	1.4	8.27 (12)	8.04 (11)	8.25 (4)	8.18 (4)	8.13 (3)	8.04 (3)	8.12 (3)	8.20 (3)	8.24 (2)	8.24 (2)
A	1.6	8.20 (12)	8.20 (5)	8.24 (4)	8.14 (3)	8.10 (3)	8.08 (3)	8.18 (3)	8.27 (2)	8.18 (2)	8.17 (2)
	1.8	8.30 (7)	8.20 (4)	8.21 (4)	8.14 (4)	8.09 (3)	8.10 (3)	8.19 (3)	8.33 (2)	8.18 (2)	8.14 (2)
	2.0	8.28 (7)	8.20 (4)	8.23 (3)	8.10 (3)	8.09 (3)	8.17 (3)	8.23 (2)	8.16 (2)	8.20 (2)	8.13 (2)

Tabla V.1 : Distancia Cepstrum (dB) después de procesar la iteración óptima (indicada entre paréntesis) mediante Filtro de Wiener Generalizado a SNR=0dB para ESCA+AGWN.

incrementos, alrededor de 1dB, en relación a los valores de la zona de equilibrio. Los máximos valores obtenidos corresponden al punto (0.7, 0.6), obteniéndose una SNR global de 9.5dB y una SNR segmentada de 6.5dB a costa de procesar 17 iteraciones.

Como conclusión, para el algoritmo AR3 en ambientes altamente ruidosos (SNR=0dB) se ha determinado una región de compromiso entre los efectos de reducción de ruido y pérdida de inteligibilidad. Esta región de equilibrio se localiza en un valle diagonal alrededor de la recta $2a + b = 3.4$ y, en concreto, la elección del punto (1.0, 1.2) comporta, además, una reducción del coste de procesado. Aunque estas medidas objetivas muestran que esta zona de equilibrio alcanza su comportamiento óptimo después de procesar tres iteraciones, los tests de audición agradecen, en general, el procesado de una iteración adicional. El incremento de distorsión ocasionada por esta cuarta iteración resulta poco perceptible para el oido humano y, en cambio, se percibe un menor ruido musical residual.

Tal como se vio en el Capítulo IV, el uso de las estadísticas clásicas de segundo orden en ambientes altamente ruidosos conducía a la baja operatividad del algoritmo AR2: incapaz de suprimir el ruido presente en la señal de voz. Este hecho se traducía en la necesidad de procesar un elevado número de iteraciones y, además, el ruido musical residual era todavia considerable. En cambio, el algoritmo AR3 permite afrontar estos niveles elevados de ruido en pocas iteraciones (cuatro a lo sumo) y el ruido musical tras este procesado es considerablemente inferior.

V.1.2. Ambientes con un Nivel Intermedio de Ruido.

Se ha considerado una SNR global de 9dB para simular un entorno con un nivel intermedio de ruido, a mitad de camino entre un entorno muy ruidoso y un ambiente poco ruidoso. Las conclusiones obtenidas en ambientes altamente ruidosos permiten vislumbrar un buen comportamiento del algoritmo AR3 generalizado para atacar estos niveles de ruido, donde el algoritmo clásico de segundo orden obtenía pobres prestaciones. El valor de distancia Cepstrum inicial correspondiente a esta **SNR= 9dB** viene dada por C_0= 10.51dB.

Las distancias obtenidas tras estimar el filtro de Wiener a partir de la señal de voz ruidosa disponible y filtrar ésta, Fig.V.6, muestran como los valores de distancia Cepstrum

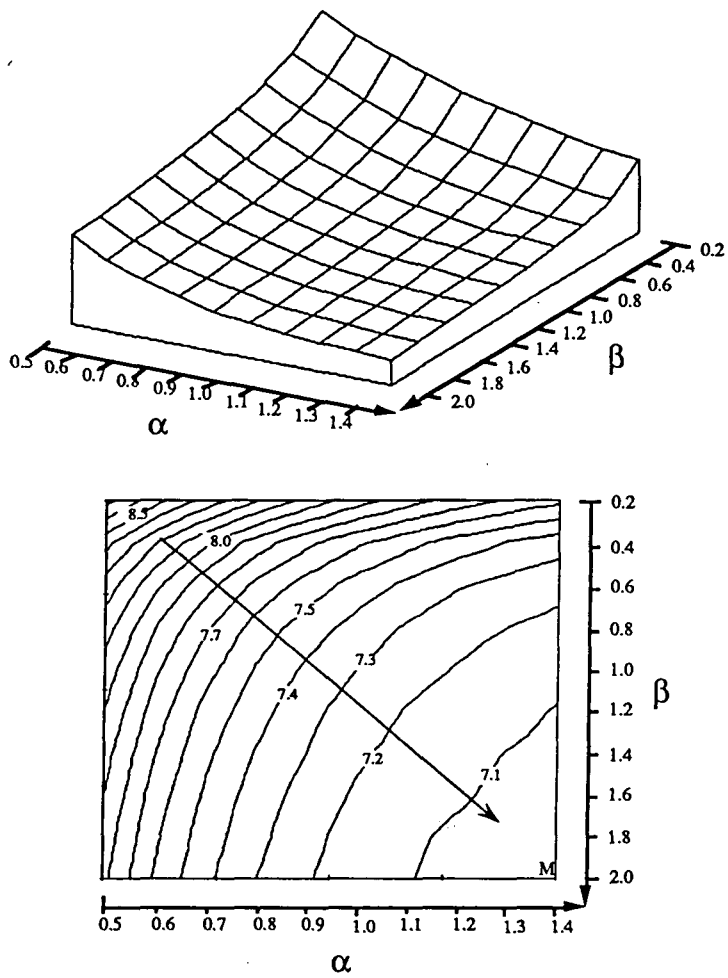


Figura V.6 : Distancia Cepstrum (dB) después de procesar la primera iteración mediante Filtro de Wiener Generalizado a SNR=9dB para ESCA+AGWN.

decrecen de forma continuada desde su valor máximo $C_{MAX}= C_1 (0.5, 0.2)= 8.73$dB hasta su valor mínimo $C_{MIN}= C_1 (1.4, 2.0)= 7.02$dB. La diferencia entre estos dos valores, entre la región más y menos agresiva, es de 1.7dB y resulta bastante similar a la obtenida para SNR=0dB. Pero en este entorno el algoritmo logra una mayor reducción de ruido (3.5dB) tras procesar una sola iteración. Este hecho demuestra la gran capacidad operativa de las estadísticas de tercer orden ante estos niveles de ruido. Obsérvese, además, como la reducción de ruido en la zona más conservadora también es considerable (1.8dB), mientras que para niveles superiores de ruido estos cumulantes de tercer orden resultaban demasiado sensibles a la presencia de éste y se obtenían peores ganancias (inferiores a 1dB). Estos motivos permiten adivinar la aparición de la zona de saturación entre reducción de ruido y pérdida de inteligibilidad en la región más agresiva $14a + 5b \geq 22.6$.

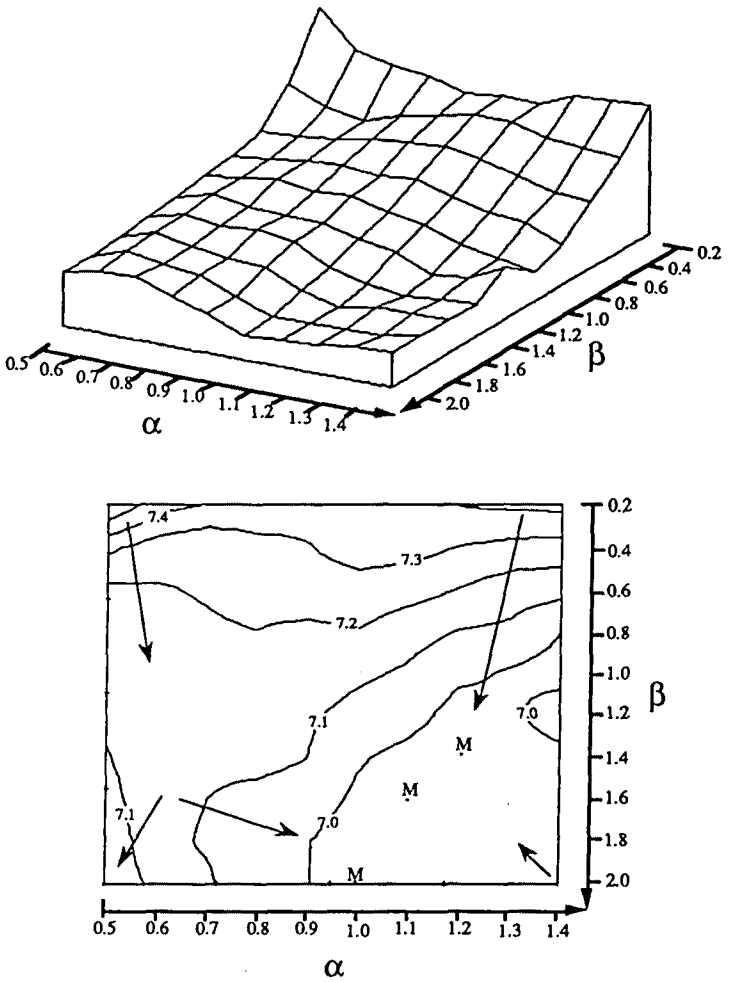


Figura V.7 : Distancia Cepstrum (dB) después de procesar la segunda iteración mediante Filtro de Wiener Generalizado a SNR=9dB para ESCA+AGWN.

Al procesar la segunda iteración, Fig.V.7, los valores de distancia oscilan entre su máximo valor $C_{MAX}= C_2(0.5, 0.2)= 7.59$dB y su valor mínimo $C_{MIN}= C_2(1.0, 2.0)= 6.93$dB. La diferencia entre estos valores extremos es bastante baja (0.65dB), debido a que la mayor parte de ruido se ha suprimido después de procesar sólamente las dos primeras iteraciones, de una forma bastante independiente con respecto a los valores de α y β. Se pueden destacar tres zonas diferenciadas: una planicie en la zona más agresiva donde se alcanzan las mejores distancias; un altiplano situado unos 0.2dB por encima y con pendiente decreciente hacia valores de β menores; y una pequeña región en la esquina de valores más conservadores donde se aprecia una mayor sensibilidad frente a variaciones de los valores de α y β.

La zona de mejor comportamiento se extiende a lo largo de la región donde se vislumbraba una cierta saturación en la iteración precedente. Las diferencias al variar los valores (α, β) son insignificantes y los mejores valores se localizan alrededor de la recta $5\alpha + 2\beta = 9$, destacando algunos mínimos locales $C_2(1.1, 1.6)= 6.94$dB y $C_2(1.2, 1.4)= 6.94$dB. Los puntos más agresivos no han empeorado $C_2(1.4, 2.0)= 6.98$dB, confirmándose la conclusión obtenida durante el estudio del algoritmo iterativo de Wiener referente a la baja pérdida de inteligibilidad durante la ejecución de las dos primeras iteraciones.

En la tercera iteración, Fig.V.8, los valores oscilan entre $C_{MAX}=C_3(1.0, 0.2)= 7.62$dB y $C_{MIN}= C_3(1.2, 0.8)= 6.84$dB. Resulta curioso observar como la zona de peor comportamiento pertenece a valores pequeños de β, como era lógico esperar, pero combinados con valores altos de α. Esto se debe a que los valores de α altos y β pequeña han sufrido un deterioro en relación a la iteración anterior, situándose alrededor de los 7.5dB durante unas dos o tres iteraciones para luego descender hasta niveles de 7.3dB, mientras que los valores de α y β pequeños han continuado mejorando para empezar a empeorar en la cuarta iteración debido a su mayor lentitud. Nótese, sin embargo, que estas diferencias en los valores de la distancia Cepstrum son muy pequeñas puesto que la mayor parte de ruido ha sido eliminada durante el procesado correspondiente a las dos primeras iteraciones.

La región que presenta un mejor comportamiento se situa alrededor del valle diagonal $2\alpha + \beta = 3.2$ *con* $\beta \geq 0.6$ *donde, además de* C_{MIN}, *se alcanza otro mínimo local en* $C_3(1.0, 1.2)= 6.88$dB. Se alcanzan valores mínimos ligeramente mejores a los proporcionados por la iteración anterior en la zona más agresiva, pero con un nivel de distorsión menor que evita su deterioro, incluso, tras procesar más iteraciones después de la iteración óptima. La región de valores más agresivos empeora un poco (0.2dB) debido al efecto distorsión que supera holgadamente la supresión de la pequeña cantidad de ruido que todavía estaba presente. A partir de la cuarta iteración, el efecto distorsión predomina claramente en la mayor parte de puntos (α, β).

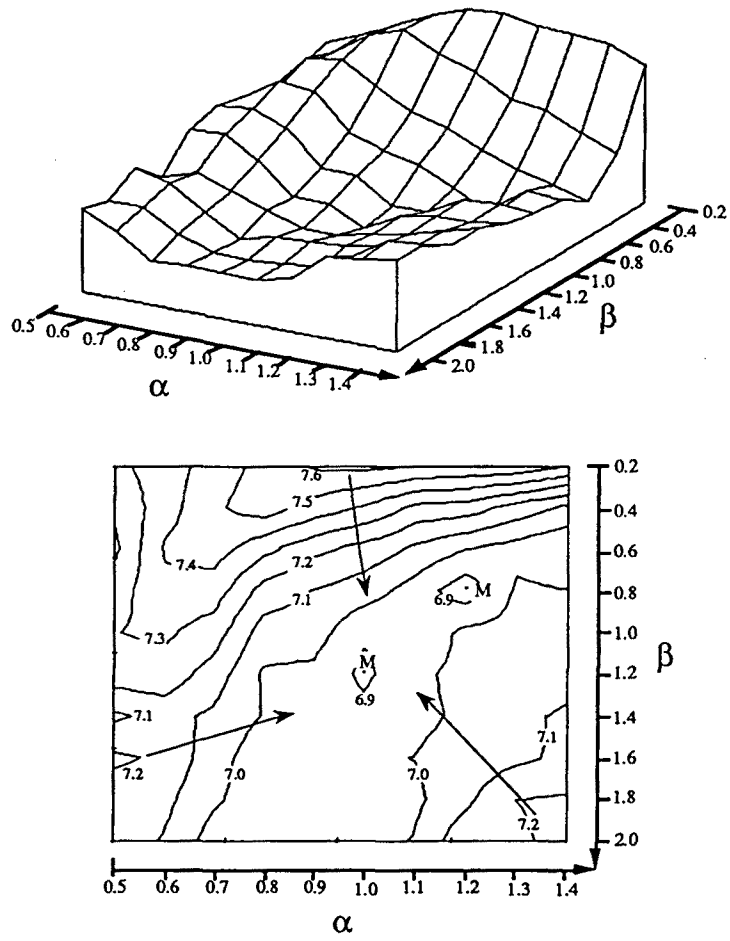


Figura V.8 : Distancia Cepstrum (dB) después de procesar la tercera iteración mediante Filtro de Wiener Generalizado a SNR=9dB para ESCA+AGWN.

Finalmente, en la Tabla V.2 y la Fig.V.9 se han representado los mejores valores obtenidos en la iteración óptima de cada par de valores (a, b). Los valores oscilan entre C_{MAX}= C_2 (0.7, 0.2)= 7.40dB y C_{MIN}= C_3 (1.2, 0.8)= 6.84dB. La diferencia entre éstos es bastante pequeña (0.55dB) e indica claramente como la supresión de ruido ha sido muy eficaz tras un pequeño número de iteraciones para este nivel de ruido. Este comportamiento es bastante independiente de los valores (a, b) considerados, perdiéndose un poco el efecto diagonal observado para SNR=0dB, y sólamente valores de b pequeños ($b \leq 0.4$) conducen a mínimos ligeramente peores (0.4dB). La zona de mejor comportamiento $2a + b = 3.2$ se localiza en el mismo lugar obtenido para SNR=0dB, siendo necesario el procesado de una iteración menos (3 iteraciones).

La distancia Itakura sigue mejorando, en general, hasta la cuarta iteración pero el efecto de picado espectral afecta de forma apreciable las zonas correspondientes a valles espectrales y, entonces, la distancia Cepstrum empeora a partir de la tercera iteración. Así, mientras la medida de distancia Cepstrum mejora apenas unos 0.2dB, al procesar la tercera iteración, la distancia Itakura suele mejorar más de 0.5dB. Por este motivo, la consideración de tres iteraciones parece un buen compromiso entre reducción de ruido y pérdida de inteligibilidad. Durante estas tres iteraciones la distancia Cepstrum decrece 3.7dB mientras la Itakura experimenta una mejora más notoria desde los 8.28dB de la voz ruidosa inicial hasta los 3.54dB, para los puntos mínimos señalizados en la Fig.V.9. Evidentemente, la distancia Itakura da mayor importancia a la zonas espectrales correspondientes a los formantes y en éstas la distorsión es menos apreciable. Al procesar tres iteraciones, la SNR global se sitúa en torno a los 14.7dB, mientras la SNR segmentada mejora desde 8.1dB hasta los 11dB. La distancia Cosh se reduce desde 9.92dB hasta 5.73dB.

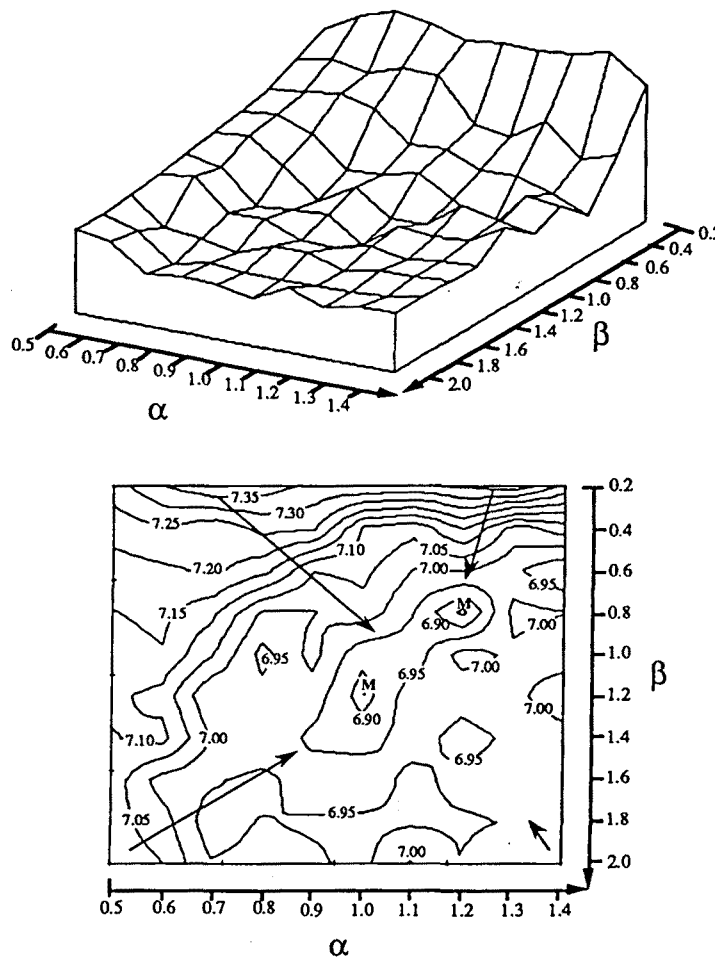


Figura V.9 : Distancia Cepstrum (dB) después de procesar la iteración óptima mediante Filtro de Wiener Generalizado a SNR=9dB para ESCA+AGWN.

En resumen, los cumulantes de tercer orden muestran un comportamiento muy superior en relación al método clásico basado en la función autocorrelación para estos niveles intermedios de ruido: la presencia de ruido musical residual es prácticamente nula y la pérdida de inteligibilidad bastante insignificante. En este caso, las diferentes configuraciones del filtro de Wiener generalizado son poco relevantes frente a la eficacia inherente a los cumulantes de tercer orden, en comparación a lo que ocurre en ambientes más ruidosos.

3				A		L		F	A		
	9dB	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4
B	0.2	7.28 (3)	7.35 (3)	7.40 (2)	7.38 (2)	7.33 (2)	7.35 (2)	7.34 (2)	7.36 (11)	7.39 (6)	7.29 (6)
	0.4	7.22 (3)	7.25 (2)	7.22 (2)	7.26 (2)	7.30 (4)	7.18 (5)	7.07 (4)	7.15 (4)	7.03 (4)	7.08 (3)
	0.6	7.18 (3)	7.19 (2)	7.21 (2)	7.17 (12)	7.05 (5)	7.06 (4)	7.00 (4)	7.00 (3)	6.97 (3)	6.90 (3)
E	0.8	7.15 (2)	7.16 (2)	7.11 (11)	6.99 (4)	7.00 (4)	7.03 (4)	6.95 (3)	6.84 (3)	7.02 (3)	7.00 (3)
	1.0	7.13 (2)	7.14 (12)	7.09 (15)	6.93 (5)	7.02 (4)	6.91 (3)	6.92 (3)	7.02 (2)	6.99 (2)	6.96 (2)
T	1.2	7.11 (2)	7.08 (12)	6.95 (4)	6.97 (4)	6.98 (3)	6.88 (3)	6.98 (3)	6.96 (2)	6.98 (2)	7.05 (3)
	1.4	7.10 (2)	7.12 (6)	7.00 (4)	7.00 (3)	6.94 (3)	6.93 (3)	6.97 (2)	6.94 (2)	6.96 (2)	6.96 (2)
A	1.6	7.08 (2)	6.98 (9)	6.96 (5)	6.93 (3)	6.98 (3)	6.98 (2)	6.94 (2)	6.96 (2)	6.97 (2)	6.97 (2)
	1.8	7.07 (2)	7.03 (5)	6.92 (4)	6.95 (3)	6.94 (3)	6.94 (3)	6.94 (2)	6.94 (2)	6.95 (2)	6.98 (2)
	2.0	7.08 (2)	7.05 (4)	6.95 (4)	6.97 (3)	6.96 (3)	6.93 (2)	6.94 (2)	6.95 (2)	6.98 (2)	6.98 (2)

Tabla V.2 : Distancia Cepstrum (dB) después de procesar la iteración óptima (indicada entre paréntesis) mediante Filtro de Wiener Generalizado a SNR=9dB para ESCA+AGWN.

V.1.3. Ambientes poco Ruidosos.

Se ha evaluado una SNR global correspondiente a 18dB para simular entornos poco ruidosos. Los resultados anteriores muestran un comportamiento muy superior para los cumulantes de tercer orden frente a la función autocorrelación, cuando se deben afrontar unos niveles de ruido medios o altos. Sin embargo, en el capítulo anterior se vió como el algoritmo AR2 ofrecía mejores prestaciones que AR3, al considerar SNR superiores a los 20dB, debido a la mayor agresividad de los cumulantes de tercer orden: el algoritmo AR2 eliminaba una cantidad de ruido similar pero a cambio de pagar un precio en distorsión algo menor. En este apartado se intenta estudiar si es posible corregir este defecto mediante el control que puedan ejercer los parámetros (α,β). El valor de distancia Cepstrum inicial correspondiente a esta señal de voz ligeramente ruidosa (**SNR=18dB**) viene cuantificada como C_0=8.52dB.

Tal como era previsible, al procesar la primera iteración (Fig.V.10) aparece una zona de saturación bastante marcada en la mayor parte de la región (V.2) bajo estudio. Se debe principalmente, a la baja cantidad de ruido a eliminar y, entonces, el efecto distorsión causada es bastante selectivo: para los posibles valores (α,β) que logran eliminar cantidades de ruido similares debemos escoger los causantes de una menor distorsión. Así, la zona de mejor comportamiento se sitúa alrededor de la recta $7\alpha + \beta = 5.5$, donde se alcanza el valor mínimo $C_{MIN}= C_1(0.6, 1.0)= 5.77$dB.

A su derecha valores (α,β) más agresivos conducen a valores peores, aunque las diferencias son pequeñas $C_1(1.4, 2.0)= 6.06$dB. En la esquina opuesta, para valores pequeños de (α,β) aún no se ha llegado a la saturación y queda algo de ruido a eliminar en la segunda iteración $C_{MAX}= C_1(0.5, 0.2)= 6.12$dB. Obsérvese que la diferencia entre estos valores extremos C_{MAX} y C_{MIN} es muy pequeña (0.35dB) debido a que los cumulantes de tercer orden logran eliminar el ruido durante un único filtrado, y de una forma bastante insensible a los valores (α,β) considerados. Si tenemos en mente lo que ocurría para entornos más ruidosos, SNR_G=0dB o SNR_G=9dB, parece como si la iteración óptima se alcanzara para una fracción de iteración menor a la unidad. Es decir, la agresividad del algoritmo AR3 parece excesiva para afrontar estos niveles bajos de ruido y se precisa que el filtro parametrizado la atenúe mediante valores (α,β) relativamente bajos.

Al procesar la iteración óptima, Fig.V.11, para la gran mayoría de valores (α, β) se alcanza la iteración óptima durante la primera. Sólamente los valores más pequeños de α y β mejoran su distancia Cepstrum tras la segunda iteración (ver Tabla V.3), marcando un nuevo valor mínimo $C_{MIN}= C_2(0.5, 0.2)= 5.76$dB, prácticamente idéntico al obtenido en la primera

iteración. El valor máximo corresponde a la zona más agresiva $C_{MAX}= C_1 (1.4, 2.0)= 6.06dB$, puesto que la zona más lenta mejora durante la segunda iteración.

La diferencia entre estos valores extremos es bastante insignificante y muestra claramente como los cumulantes de tercer orden logran eliminar este nivel de ruido tras una única iteración y no precisan sustentarse en una determinada configuración (a, b) del filtro de Wiener generalizado, tal como muestran los resultados indicados en la Tabla V.3. En consecuencia, la mayoría de valores (a, b) originan un buen comportamiento y, en esta situación, lo más conveniente consiste en tomar los valores más conservadores que logran afrontar este nivel de ruido, valle diagonal alrededor de $7a + b = 5.5$. Debe remarcarse que valores (a, b) más agresivos a los de este valle no introducen distorsión perceptible al oido humano y, sólamente, se detecta cuantitativamente en la medida de distancia Cepstrum.

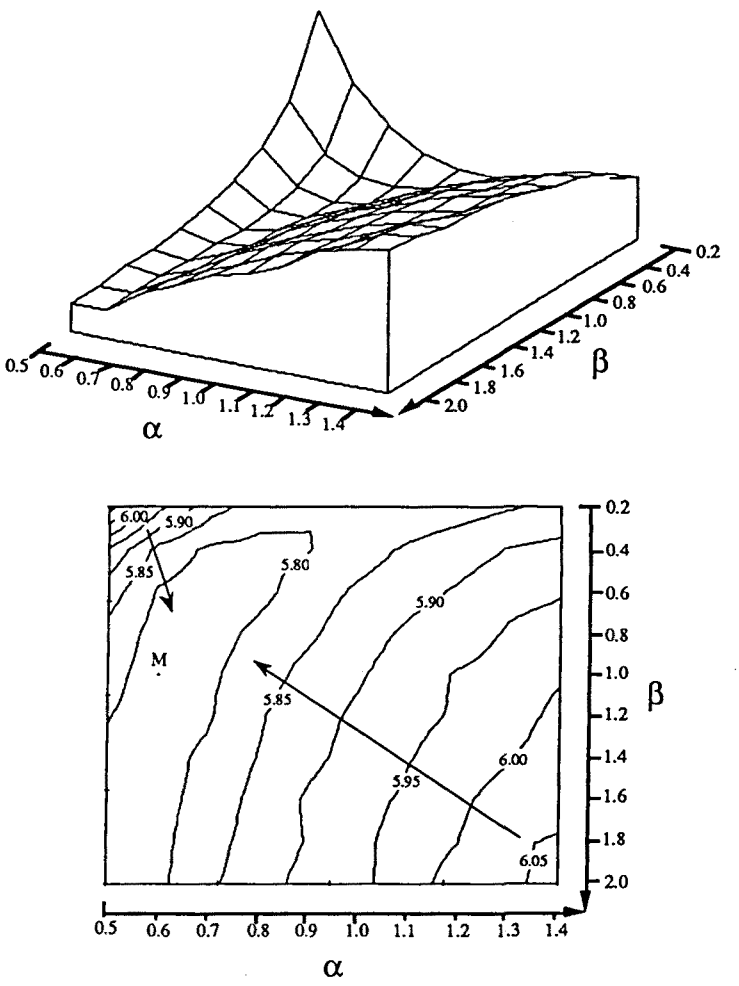


Figura V.10 : Distancia Cepstrum (dB) después de procesar la primera iteración mediante Filtro de Wiener Generalizado a SNR=18dB para ESCA+AGWN.

La distancia Itakura mejora de forma generalizada en la segunda iteración, decreciendo desde los 6.33dB iniciales hasta los 3dB tras la primera iteración y hasta los 2.7dB al final de la segunda iteración. Aunque la distancia Cepstrum indica un ligero incremento de pérdida de inteligibilidad durante la segunda iteración, C_2 (1.2, 2.0)= 6.37dB, ésta no es percibida por el oido humano, puesto que durante las dos primeras iteraciones este algoritmo AR3 no introduce distorsión significativa. Para las medidas temporales esta mejora se traduce en una SNR global de 22dB y una SNR segmentada que mejora desde 13.4dB hasta 17.3dB. Asimismo, la distancia Cosh decrece de 7.89dB hasta los 5dB.

Concluyendo, se puede afirmar que para este nivel de ruido los cumulantes de tercer orden obtienen resultados de distancia Cepstrum ligeramente peores a los obtenidos mediante las estadísticas de segundo orden (AR2). Sin embargo, las pruebas de audición muestran calidades parecidas. Además, el algoritmo de tercer orden presenta una ventaja adicional: mientras el algoritmo AR2 precisa 3 ó 4 iteraciones para suprimir el ruido existente, el

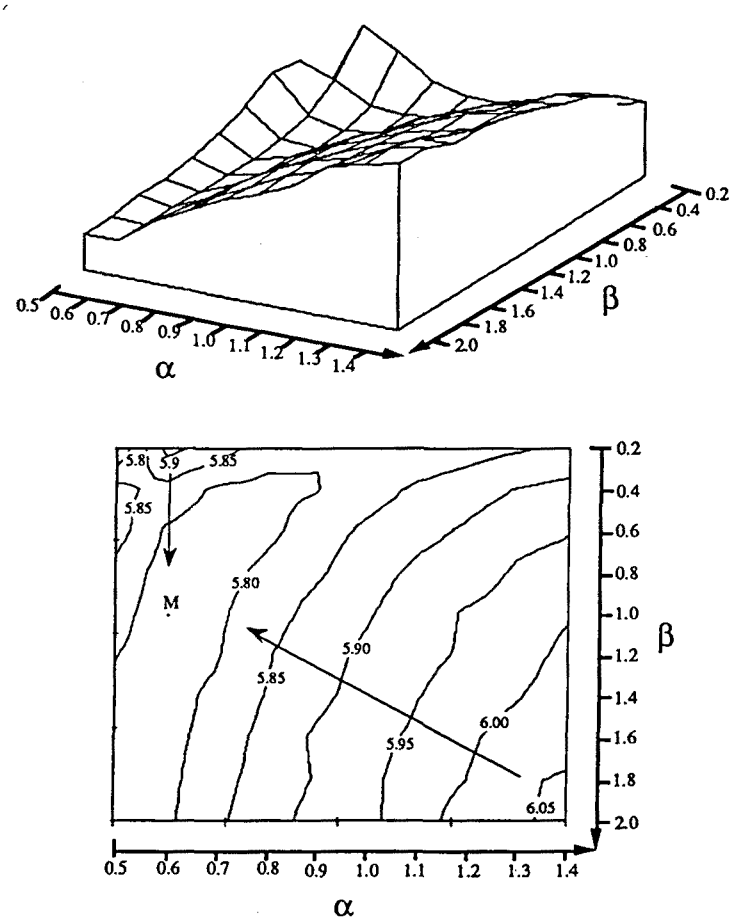


Figura V.11 : Distancia Cepstrum (dB) después de procesar la iteración óptima mediante Filtro de Wiener Generalizado a SNR=18dB para ESCA+AGWN.

algoritmo AR3 lo consigue en una sola iteración y, en consecuencia, comporta un ahorro considerable de tiempo de cálculo.

3		A			L		F		A		
	18dB	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4
	0.2	5.76 (2)	5.92 (2)	5.87 (1)	5.82 (1)	5.80 (1)	5.80 (1)	5.81 (1)	5.83 (1)	5.85 (1)	5.86 (1)
B	0.4	5.86 (1)	5.83 (1)	5.79 (1)	5.79 (1)	5.80 (1)	5.83 (1)	5.86 (1)	5.88 (1)	5.90 (1)	5.91 (1)
	0.6	5.87 (1)	5.79 (1)	5.77 (1)	5.78 (1)	5.83 (1)	5.86 (1)	5.89 (1)	5.91 (1)	5.94 (1)	5.95 (1)
E	0.8	5.84 (1)	5.78 (1)	5.78 (1)	5.81 (1)	5.84 (1)	5.88 (1)	5.91 (1)	5.93 (1)	5.96 (1)	5.97 (1)
	1.0	5.82 (1)	5.77 (1)	5.79 (1)	5.83 (1)	5.87 (1)	5.89 (1)	5.93 (1)	5.95 (1)	5.97 (1)	5.99 (1)
T	1.2	5.80 (1)	5.78 (1)	5.79 (1)	5.84 (1)	5.88 (1)	5.91 (1)	5.94 (1)	5.95 (1)	5.98 (1)	6.01 (1)
	1.4	5.79 (1)	5.78 (1)	5.81 (1)	5.85 (1)	5.89 (1)	5.92 (2)	5.95 (1)	5.97 (1)	6.00 (1)	6.03 (1)
A	1.6	5.79 (1)	5.78 (1)	5.82 (1)	5.86 (1)	5.91 (1)	5.94 (1)	5.96 (1)	5.99 (1)	6.02 (1)	6.04 (1)
	1.8	5.78 (1)	5.78 (1)	5.83 (1)	5.87 (1)	5.90 (1)	5.94 (1)	5.97 (1)	6.00 (1)	6.04 (1)	6.06 (1)
	2.0	5.78 (1)	5.79 (1)	5.84 (1)	5.88 (1)	5.91 (1)	5.94 (1)	5.97 (1)	6.03 (1)	6.04 (1)	6.06 (1)

***Tabla V.3** : Distancia Cepstrum (dB) después de procesar la iteración óptima (indicada entre paréntesis) mediante Filtro de Wiener Generalizado a SNR=18dB para ESCA+AGWN.*

V.2. Método AR4 Generalizado.

En este apartado se estudia el comportamiento del filtrado iterativo de Wiener generalizado, de forma análoga al apartado anterior, pero usando las estadísticas de orden cuarto durante la estimación espectral de la señal de voz. Este método se nota como AR4 ya que modela la señal de voz de forma paramétrica AR y halla los coeficientes ak, pertenecientes a este modelo, a partir de los cumulantes de cuarto orden, mediante la resolución de las ecuaciones de Yule-Walker de cuarto orden representadas en (III.96). También, se evaluan tres entornos diferenciados correspondientes a un nivel alto, medio o bajo de ruido. Los resultados presentados corresponden a ruido aditivo Gaussiano blanco de media nula. La región del plano a-b a estudiar se ha modificado ligeramente:

$$\begin{aligned} 0.5 \leq a \leq 1.5 \\ 0.2 \leq b \leq 1.8 \end{aligned} \qquad \text{(V.3)}$$

con los mismos incrementos anteriormente definidos: 0.1 para el parámetro a y 0.2 para b. Nótese que cada representación, correspondiente a una determinada iteración del filtrado AR4, se compone de 99 medidas asociadas a distintas configuraciones del filtro de Wiener estimado. La nomenclatura utilizada para interpretar cada uno de los gráficos se corresponde idénticamente a la definida durante el apartado anterior.

V.2.1. Ambientes Altamente Ruidosos.

Este tipo de entornos con la presencia de mucho ruido se han simulado mediante una SNR global de 0dB. Recordemos que la distancia Cepstrum inicial asociada a esta **SNR= 0dB** es C_0= 12.02dB.

Después de filtrar la señal de voz ruidosa disponible, Fig.V.12, se aprecia un descenso muy uniforme en los valores de distancia Cepstrum entre su valor máximo C_{MAX}= C_1 (0.5, 0.2)= 11.48dB y su valor mínimo C_{MIN}= C_1 (1.5, 1.8)= 9.81dB. El comportamiento es muy similar al caso anterior de tercer orden con un marcado efecto diagonal-circular, pero existen dos pequeñas diferencias: el descenso es más uniforme porque la zona más agresiva todavía

no ha empezado a saturar y, en segundo lugar, los cumulantes de cuarto orden se muestran más sensibles al ruido, originando una menor supresión de ruido: C_{MIN} y C_{MAX} empeoran 0.3dB y 0.5dB respectivamente. El efecto supresión de ruido domina claramente respecto los niveles de distorsión ocasionada: la máxima reducción de ruido se corresponde con una reducción de 2.2dB en la distancia Cepstrum, mientras que los valores más pequeños de α y β apenas reducen significativamente el nivel de ruido (0.54dB).

En la segunda iteración, Fig.V.13, empieza a saturar la zona más agresiva, aunque la región ocupada $8\alpha + 5\beta \geq 17$ es menor en relación al algoritmo AR3. Este efecto origina un descenso menos uniforme, en relación a la iteración anterior, entre su valor máximo $C_{MAX}= C_2\,(0.5, 0.2)= 11.28$dB y su valor mínimo $C_{MIN}= C_2\,(1.5, 1.8)= 8.59$dB. Estos valores también empeoran unos 0.5dB en comparación a los obtenidos mediante los cumulantes de

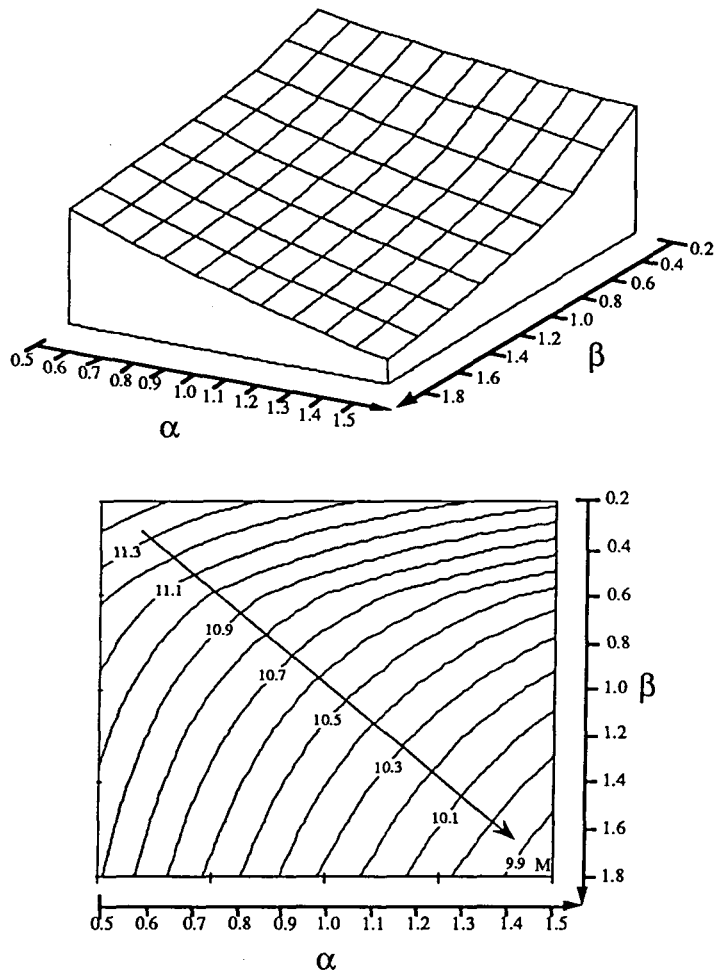


Figura V.12 : Distancia Cepstrum (dB) después de procesar la primera iteración mediante Filtro de Wiener Generalizado a SNR=0dB para ESCA+AGWN.

tercer orden, aunque su comportamiento ante la variación de los parámetros (a, b) es bastante similar: domina el efecto supresión de ruido y las localizaciones de los valores máximo y mínimo coinciden. Para estos cumulantes de cuarto orden se aprecia, no obstante, una menor capacidad para afrontar este elevado nivel de ruido y, en consecuencia, deben apoyarse en la agresividad suministrada por los valores grandes de los parámetros (a, b). La sensibilidad respecto los valores (a, b) es bastante similar para los algoritmos AR4 y AR3, puesto que las diferencias entre C_{MAX} y C_{MIN} son parecidas (2.7dB y 2.5dB respectivamente). Debe remarcarse su extremada lentitud en la región más conservadora, donde tras dos iteraciones la reducción de ruido es menor a 1dB.

En la tercera iteración, Fig.V.14, se confirma la presencia de la zona de saturación entre ambos efectos. Esta zona de equilibrio $10a + 7b \geq 20.6$ se ha extendido un poco hacia el interior de la región (V.3). El efecto distorsión empieza a sobrepasar el efecto reducción de

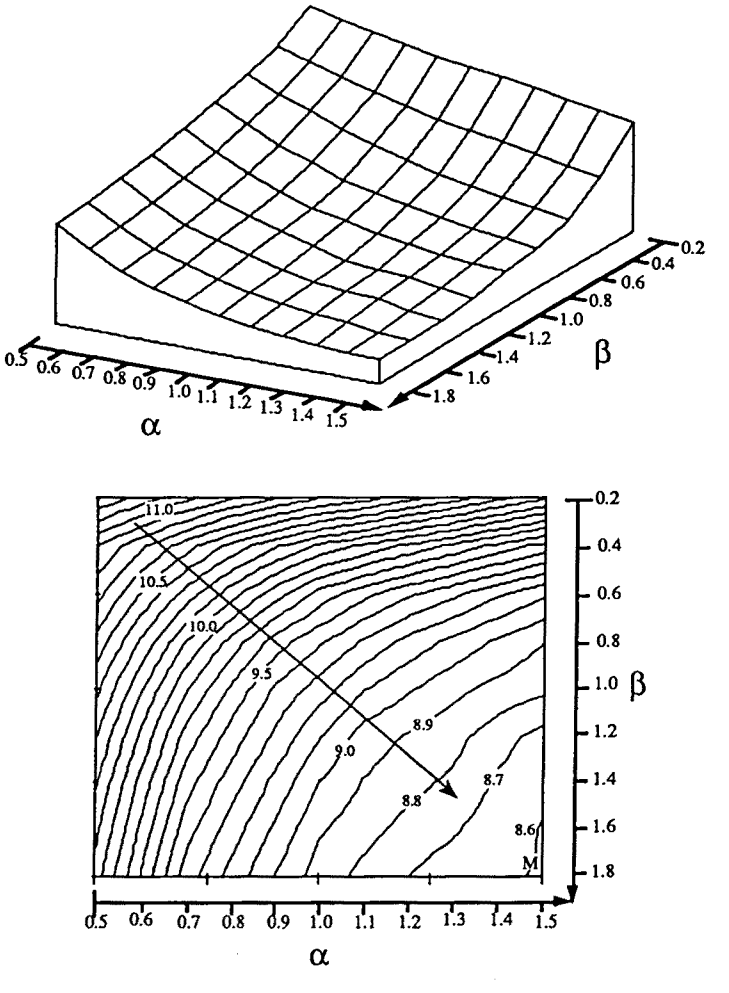


Figura V.13 : Distancia Cepstrum (dB) después de procesar la segunda iteración mediante Filtro de Wiener Generalizado a SNR=0dB para ESCA+AGWN.

ruido en la zona de valores a y b mayores, empeorando suavemente algún valor C_3 (1.5, 1.8)= 8.61dB. En consecuencia, la región de equilibrio se sitúa un poquito hacia el interior, alrededor de la recta 5a + 2b = 10.1, donde se localiza el valor mínimo C_{MIN}= C_3 (1.3, 1.8)= 8.26dB. La zona más conservadora apenas mejora: C_{MAX}= C_3 (0.5, 0.2)= 11.17dB. A grandes rasgos, estos resultados se pueden comparar con los obtenidos mediante el algoritmo de tercer orden tras la segunda iteración. Este comportamiento de los cumulantes de cuarto orden parece conducir a un claro inconveniente referente al coste de cálculo: el algoritmo AR4 suprime ruido de una forma más lenta que AR3 y, además, el cálculo de los cumulantes de cuarto orden es más complejo.

En la cuarta iteración, Fig.V.15, la zona de equilibrio se traslada hacia el interior a lo

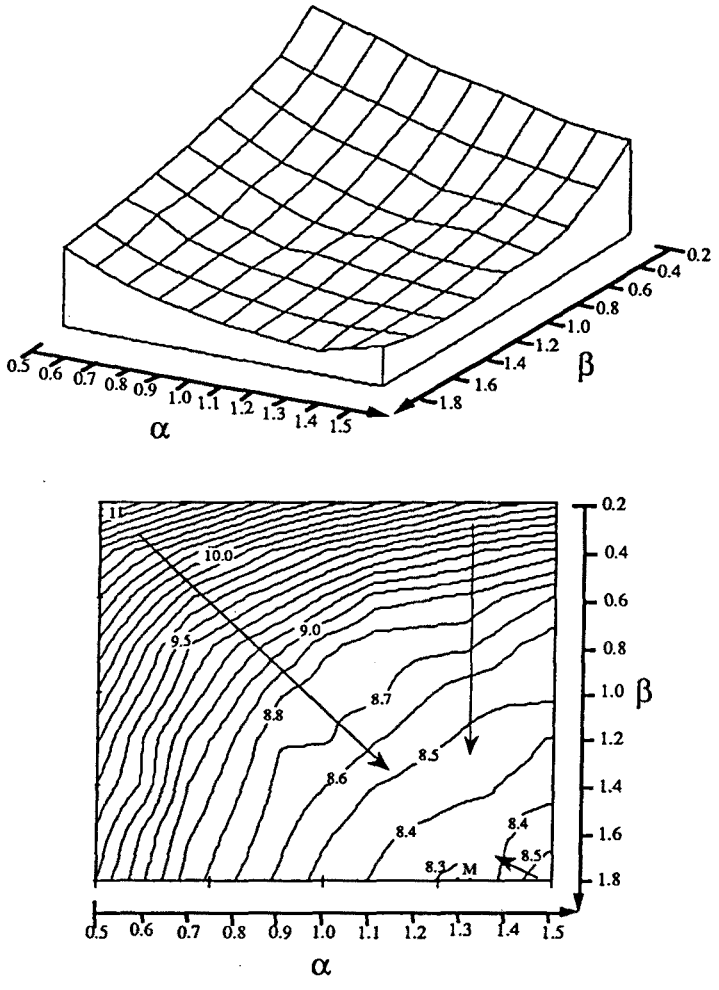


Figura V.14 : Distancia Cepstrum (dB) después de procesar la tercera iteración mediante Filtro de Wiener Generalizado a SNR=0dB para ESCA+AGWN.

largo de un valle diagonal más ancho, alrededor de la recta 12a + 5b = 21. El valor mínimo se alcanza en esta zona C_{MIN}= C_4 (1.0, 1.8)= 8.29dB, aunque ya no mejora respecto la iteración anterior porque se acumula la distorsión inherentemente ocasionada por este algoritmo iterativo tras cuatro iteraciones. Dentro de esta zona de mejor comportamiento se aprecia una ligera mejora de prestaciones al aumentar b y, por esta razón, se alcanza el valor mínimo cuando se considera el mayor (b=1.8) valor de b. Parece como si las sobreestimaciones de ruido ayudaran a suprimir cantidades de ruido algo superiores, sin ocasionar un incremento de distorsión como sucede al aumentar el valor de a.

La zona más conservadora continua sin conseguir mejoras apreciables y su valor máximo no se reduce apenas C_{MAX}= C_4 (0.5, 0.2)= 11.1dB, debido a que la cantidad de ruido suprimida se compensa con la pérdida de inteligibilidad. En la esquina contraria, el efecto

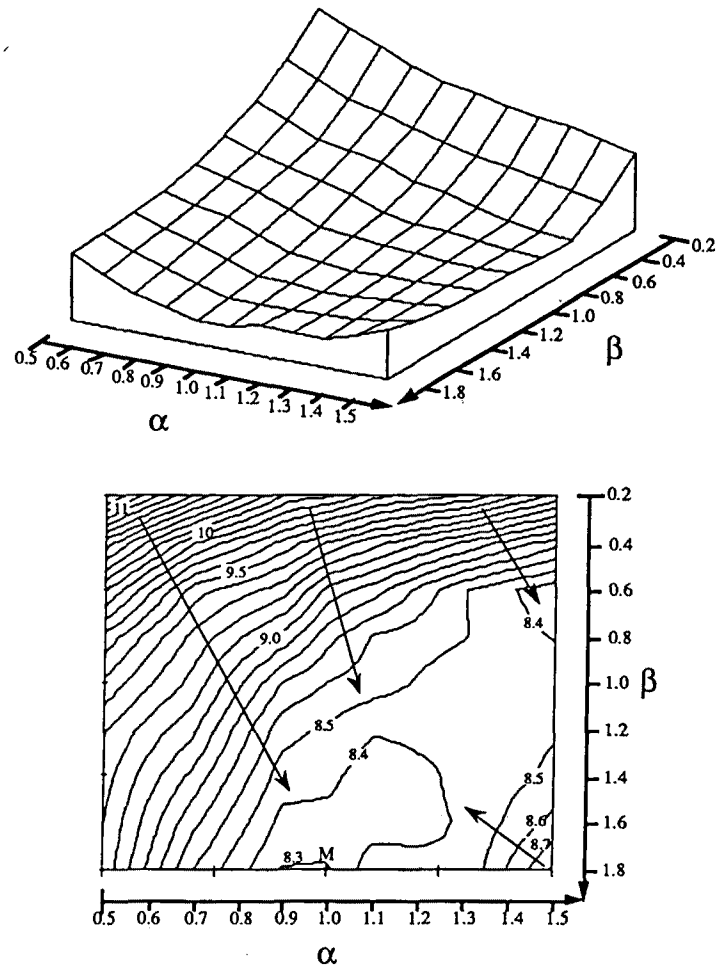


Figura V.15 : Distancia Cepstrum (dB) después de procesar la cuarta iteración mediante Filtro de Wiener Generalizado a SNR=0dB para ESCA+AGWN.

distorsión domina claramente y los valores de distancia empeoran ligeramente C_4 (1.5, 1.8)= 8.81dB.

Tras la quinta iteración, Fig.V.16, la zona de equilibrio se ensancha un poco hacia la izquierda, localizándose alrededor de 3a + b = 4.5 con b≥0.6. Sus valores mínimos C_{MIN}= C_5 (0.9, 1.6)= C_5 (0.9, 1.8)= 8.29dB se mantienen al mismo nivel obtenido en las dos iteraciones anteriores: se compensan los efectos de una menor agresividad debida al filtro (valores a y b menores) con la superior distorsión acumulada debida a un mayor número de iteraciones. La zona de valores a y b pequeños continua estancada a los valores obtenidos por las iteraciones precedentes C_{MAX}= C_5 (0.5, 0.2)= 11.01dB, progresando de forma lentísima. La zona más conservadora sigue empeorando suavemente (0.15dB) C_5 (1.5, 1.8)= 8.96dB ya que, una vez superada la iteración óptima, el efecto distorsión es dominante.

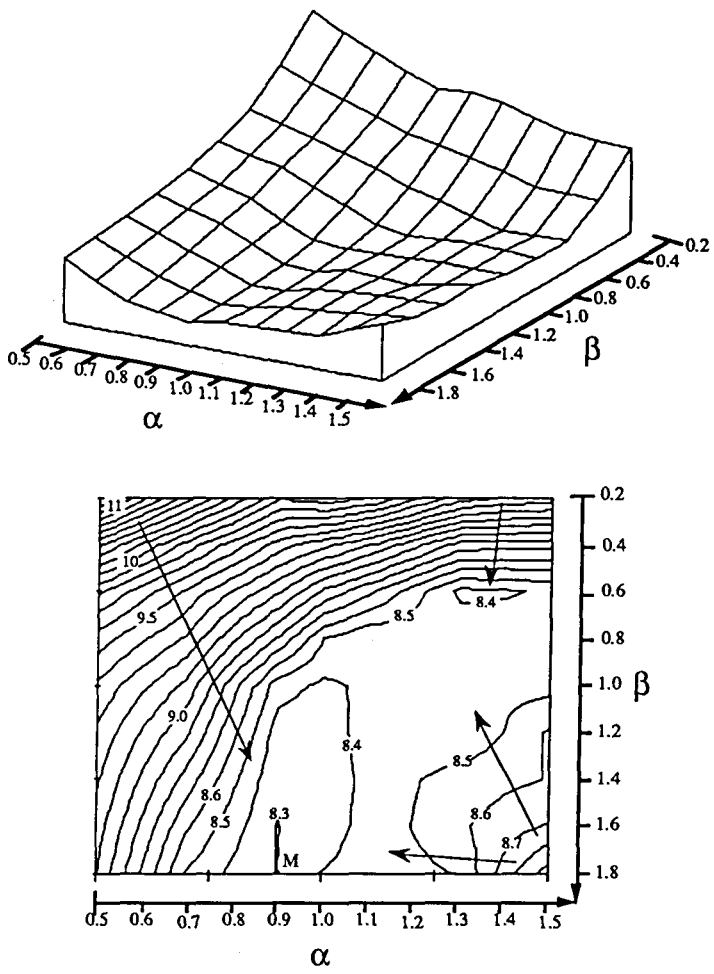


***Figura V.16** : Distancia Cepstrum (dB) después de procesar la quinta iteración mediante Filtro de Wiener Generalizado a SNR=0dB para ESCA+AGWN.*

Estamos en el punto límite entre eliminación de ruido y distorsión ocasionada, donde iterar más no conduce a una reducción de las distancias espectrales, sinó sólamente alguna pequeñísima mejora cuando se consideran valores (a, b) pequeños, aunque afectando muy seriamente la inteligibilidad del mensaje. En relación a AR3 se aprecia una variación más uniforme hacia la zona de valores mínimos y no aparecen zonas de máximos y mínimos locales. Nótese como esta quinta iteración origina peores resultados a los correspondientes a AR3 tras tres iteraciones.

En la Fig.V.17 se han representado las distancias Cepstrum correspondientes a la iteración óptima, mientras en la Tabla V.4 se han indicado estos valores conjuntamente con la iteración óptima donde se alcanzan. En comparación al algoritmo AR3, los mejores valores se obtienen para unos valores (a, b) menores, a costa de procesar un número bastante superior de

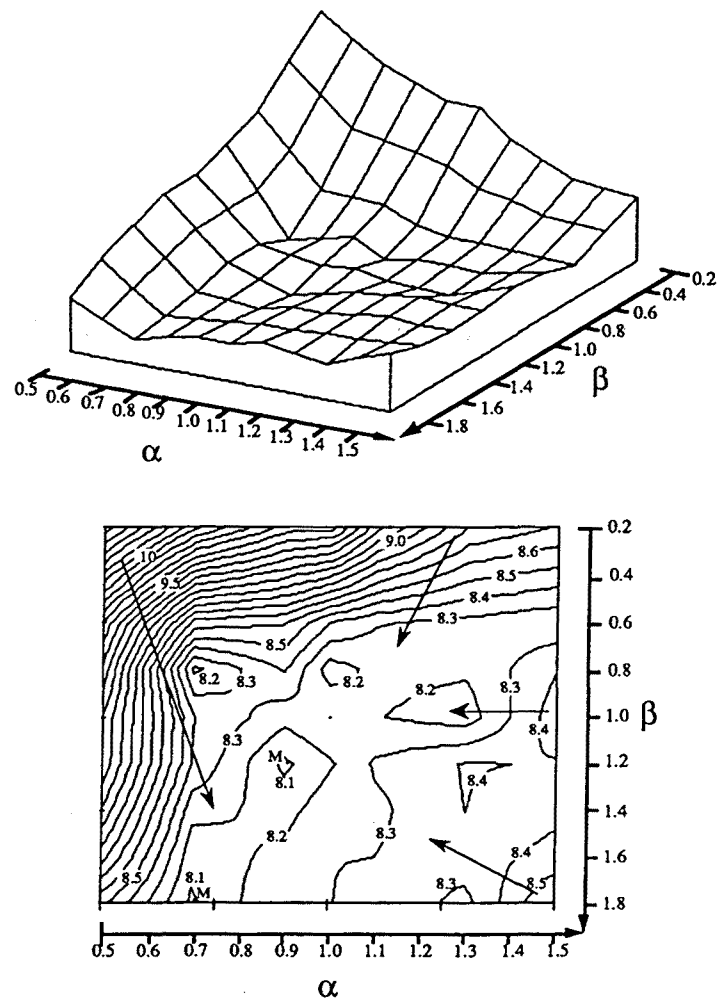


Figura V.17 : Distancia Cepstrum (dB) después de procesar la iteración óptima mediante Filtro de Wiener Generalizado a SNR=0dB para ESCA+AGWN.

de iteraciones: los valores mínimos se alcanzan tras procesar ocho iteraciones en los puntos $C_{MIN}=C_8$ (0.9, 1.2)=8.08dB y C_8 (0.7, 1.8)=8.09dB. Además, estos valores son ligeramente peores y comportan una distorsión bastante superior.

Considerando que el filtro de Wiener básico obtiene su mínimo tras seis iteraciones C_6 (1.0, 1.0)=8.30dB se pueden aceptar, como solución de compromiso, aquellos pares (α, β) que obtengan valores inferiores a los 8.3dB después de procesar sólamente 3 ó 4 iteraciones. Perteneciente a este caso el punto (1.3, 1.8) alcanza una distancia Cepstrum de 8.26dB tras tres iteraciones. Las pruebas de audición muestran una apreciable distorsión al sobrepasar la cuarta iteración. Si se procesan sólamente cuatro iteraciones, entonces, el ruido musical residual es todavía considerable.

La zona de valores más conservadores elimina cantidades de ruido demasiado pequeñas. Después de un número elevado de iteraciones, alcanza valores de distancia Cepstrum peores (1dB) en relación a los obtenidos por los cumulantes de tercer orden en idénticas condiciones. Obviamente el valor máximo corresponde a esta región $C_{MAX}=C_{12}$ (0.5, 0.2)=10.6dB.

4	A				L		F		A		
	0dB	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5
	0.2	10.6 (12)	10.3 (10)	10.0 (8)	9.83 (10)	9.65 (11)	9.63 (14)	9.22 (19)	9.01 (16)	8.79 (13)	8.74 (12)	8.69 (12)
B	0.4	10.2 (10)	9.75 (11)	9.34 (12)	9.23 (11)	9.13 (12)	9.01 (14)	8.77 (15)	8.65 (12)	8.53 (10)	8.49 (10)	8.46 (10)
	0.6	9.69 (7)	9.17 (10)	8.66 (15)	8.63 (13)	8.61 (12)	8.40 (14)	8.31 (10)	8.29 (8)	8.27 (6)	8.25 (7)	8.23 (8)
E	0.8	9.48 (6)	8.83 (11)	8.18 (15)	8.29 (12)	8.41 (10)	8.15 (8)	8.23 (6)	8.22 (8)	8.21 (11)	8.30 (7)	8.39 (4)
	1.0	9.29 (16)	8.85 (14)	8.40 (12)	8.32 (10)	8.24 (8)	8.30 (6)	8.21 (13)	8.18 (12)	8.15 (12)	8.30 (8)	8.46 (4)
T	1.2	9.28 (5)	8.83 (8)	8.39 (11)	8.23 (9)	8.08 (8)	8.18 (6)	8.32 (7)	8.36 (6)	8.41 (5)	8.40 (4)	8.39 (3)
	1.4	9.09 (8)	8.66 (9)	8.23 (10)	8.20 (8)	8.16 (7)	8.26 (5)	8.27 (5)	8.34 (4)	8.40 (4)	8.35 (3)	8.37 (3)
A	1.6	8.84 (13)	8.49 (11)	8.13 (9)	8.18 (7)	8.23 (5)	8.29 (4)	8.30 (4)	8.33 (3)	8.35 (3)	8.39 (3)	8.44 (3)
	1.8	8.53 (17)	8.31 (12)	8.09 (8)	8.19 (6)	8.29 (4)	8.29 (4)	8.40 (3)	8.33 (3)	8.26 (3)	8.44 (3)	8.61 (3)

***Tabla V.4** : Distancia Cepstrum (dB) después de procesar la iteración óptima (indicada entre paréntesis) mediante Filtro de Wiener Generalizado a SNR=0dB para ESCA+AGWN.*

En la región opuesta, valores altos de (a, b) conducen a valores de distancia un poco peores que los de AR3 y, además, precisan procesar una o dos iteraciones adicionales. Desde el punto de vista de coste de cálculo deberíamos situarnos en esta zona, pero la calidad apreciada en las pruebas de audición es peor a la obtenida por el algoritmo AR3, aunque superior a la correspondiente a las técnicas clásicas AR2.

Como conclusión, se aprecia una mayor sensibilidad al ruido cuando se usan los cumulantes de cuarto orden en lugar de los de tercer orden. Recuérdase que ante estos niveles elevados de ruido, los cumulantes de tercer orden lograban afrontarlo a costa de degradar un poco la inteligibilidad de la señal de voz y dejar algo de ruido musical tras unas cuatro iteraciones de procesado. Obviamente, el mayor conservadurismo que muestran los cumulantes de cuarto orden les hace inadecuados ante estos niveles elevados de ruido. El problema reside en el elevado número de iteraciones a procesar y, además, cada una de ellas conlleva implícitamente un mayor gasto computacional. Incluso en el supuesto que se pudiera aceptar este incremento de tiempo de procesado, los valores mínimos locales obtenidos son peores y la calidad mostrada por las pruebas de audición es peor. Por este motivo las mejores prestaciones del algoritmo AR4 se obtienen cuando el filtro toma valores agresivos de (a, b). Para valores conservadores de (a, b), se observa una total incapacidad en vistas a suprimir el ruido presente en la voz, de forma parecida a las prestaciones ofrecidas por las estadísticas de segundo orden clásicas. En resumen, el algoritmo AR4 conlleva una carga computacional bastante mayor y la calidad obtenida es peor con relación al anterior algoritmo AR3. Así, ante estos niveles de ruido es preferible el uso del algoritmo AR3.

V.2.2. Ambientes con un Nivel Intermedio de Ruido.

Los entornos con un nivel intermedio de ruido han sido simulados mediante la consideración de una SNR global de 9dB. En este tipo de entornos el algoritmo AR3 ofrece unas buenas prestaciones mientras el algoritmo clásico AR2 ofrece una calidad bastante pobre. En el presente apartado se trata de averiguar si el algoritmo AR4 puede afrontar el ruido presente en la señal de voz cuando su nivel no es tan elevado. La distancia Cepstrum inicial asociada a esta **SNR= 9dB** es C_0= 10.51dB.

En la primera iteración, Fig.V.18, se aprecia una variación bastante uniforme entre su punto más lento $C_{MAX}=C_1$ (0.5, 0.2)= 9.47dB y sus valores (a, b) más agresivos donde se localiza el mínimo $C_{MIN}=C_1$ (1.5, 1.8)= 7.33dB. La diferencia entre ambos valores es elevada (2.15dB) e indica claramente la ayuda ofrecida por los parámetros (a, b) a los cumulantes de cuarto orden en vistas a eliminar el ruido existente. En la zona de valores (a, b) altos, se aprecia una mayor separación entre curvas de nivel debido a que empieza a aparecer la zona de saturación. Según lo comentado anteriormente, este algoritmo de cuarto orden parece estar capacitado para suprimir este nivel de ruido. Durante este primer filtrado el efecto reducción de ruido domina claramente, dejando el efecto distorsión escondido en un segundo plano. Las reducciones de distancia Cepstrum van desde 1dB en la zona conservadora hasta los 3.2dB en la zona más agresiva. Aunque estas prestaciones resultan bastante buenas, son todavía unos 0.5dB peores a las del algoritmo AR3.

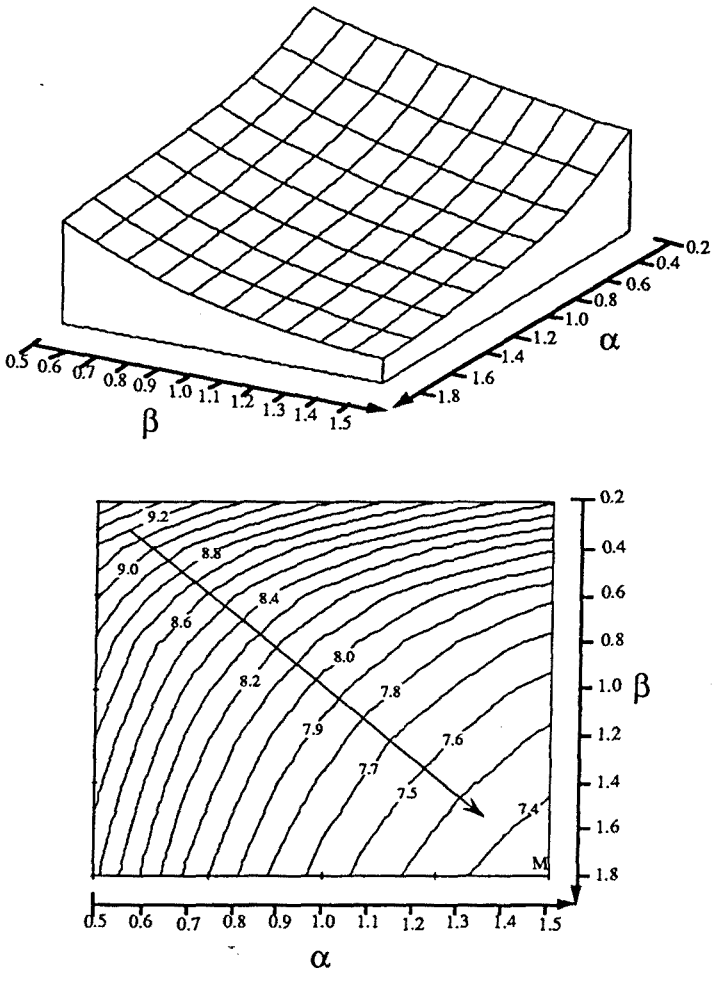


Figura V.18 : Distancia Cepstrum (dB) después de procesar la primera iteración mediante Filtro de Wiener Generalizado a SNR=9dB para ESCA+AGWN.

Al procesar la segunda iteración, Fig.V.19, aparece una zona de equilibrio muy extensa para $0.6 \leq a \leq 1.5$ y $1.0 \leq b \leq 1.8$, donde se alcanzan valores similares (0.15dB peores) a los obtenidos mediante el algoritmo AR3. En esta región se alcanzan dos valores mínimos: $C_{MIN}= C_2$ (0.7, 1.8)= 7.07dB y C_2(1.3, 1.8)= 7.08dB. Dentro de esta zona de equilibrio pueden distinguirse dos características: la zona de valores más agresivos no presenta un dominio del efecto distorsión sobre la reducción de ruido y, en consecuencia, la inteligibilidad no se deteriora tanto y, en segundo lugar, se obtiene un comportamiento ligeramente superior al incrementar b, alcanzando los mejores resultados cuando el parámetro b toma su valor máximo (b=1.8). Obsérvese como esta zona de mayor eficacia coincide con la consideración de una sobreestimación del ruido presente. También se aprecia como los cumulantes de cuarto orden tratan más cuidadosamente la señal de voz, ocasionando una menor distorsión cuando

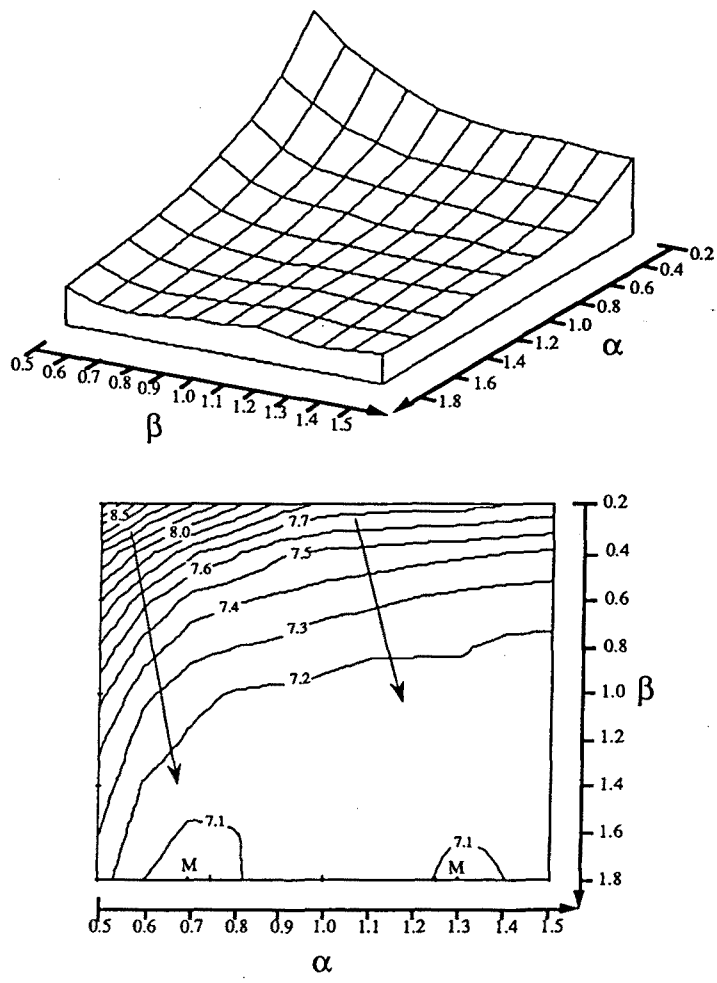


Figura V.19 : Distancia Cepstrum (dB) después de procesar la segunda iteración mediante Filtro de Wiener Generalizado a SNR=9dB para ESCA+AGWN.

el filtro es más agresivo C_2 (1.5, 1.8)= 7.16dB. Esto provoca que el efecto diagonal-circular presente una mayor sensibilidad con respecto a los valores de b.

En la zona más conservadora el filtro ralentiza la reducción de ruido propia de los cumulantes de cuarto orden. El valor máximo sigue localizado en la misma posición C_{MAX}= C_2 (0.5, 0.2)= 8.74dB. Aunque se ha reducido unos 0.7dB durante esta segunda iteración, el nivel de ruido presente en esta zona es todavía importante. Se puede afirmar que los cumulantes de cuarto orden se muestran un poco maniatados ante valores (a, b) conservadores en la configuración del filtro de Wiener.

En la tercera iteración, Fig.V.20, la zona de equilibrio se corresponde, básicamente, con la zona donde se realiza una sobreestimación del ruido (b>1). La dependencia respecto el

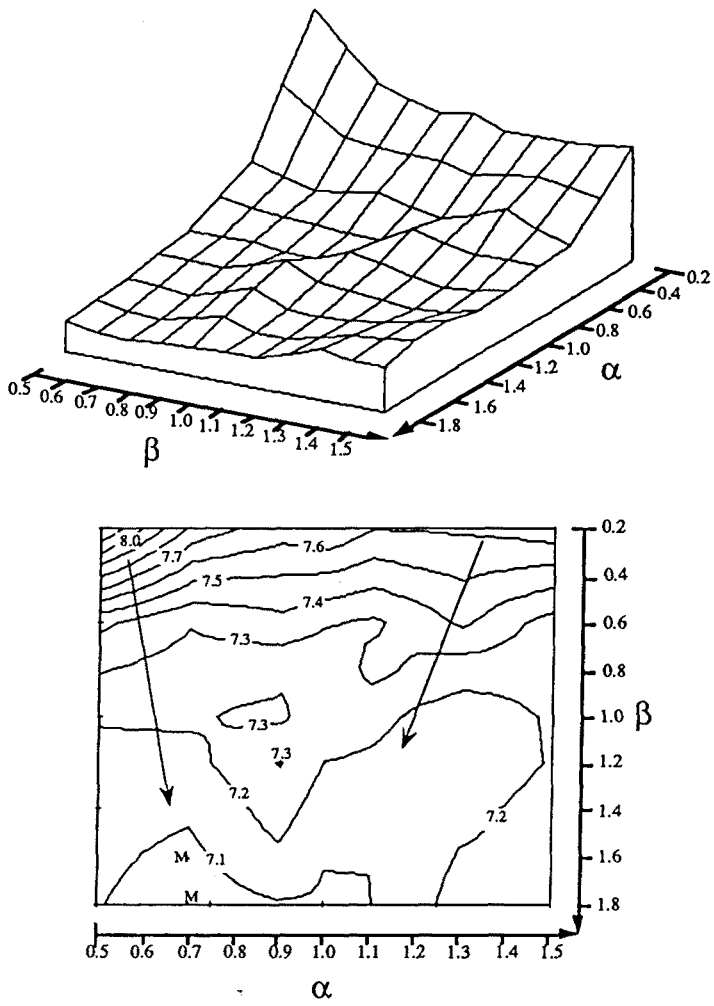


***Figura V.20** : Distancia Cepstrum (dB) después de procesar la tercera iteración mediante Filtro de Wiener Generalizado a SNR=9dB para ESCA+AGWN.*

parámetro a es muy pequeña. Los mínimos se alcanzan en $C_{MIN}= C_3 (0.7, 1.8)= 7.07dB$ y en $C_3 (0.7, 1.6)= 7.09dB$, siendo similares a los de la iteración anterior y ligeramente peores (0.2dB) a los obtenidos mediante AR3. La zona más agresiva apenas sufre distorsión tras superar la iteración óptima $C_3 (1.5, 1.8)= 7.29dB$. Mientras en la zona conservadora hay a un ruido pendiente de eliminación $C_{MAX}= C_3 (0.5, 0.2)= 8.22dB$.

Para concluir el análisis de comportamiento del algoritmo AR4 en entornos medianamente ruidosos se han representado los resultados correspondientes a la iteración óptima en la Fig.V.21 y en la Tabla V.5. Las distancias oscilan entre su valor máximo $C_{MAX}= C_6 (0.5, 0.2)= 7.75dB$ y sus mínimos $C_{MIN}= C_6 (0.7, 1.8)= C_6 (0.7, 1.6) = 6.98dB$, aunque también se alcanzan otros mínimos similares $C_6 (0.9, 1.0)= C_5 (1.0, 1.2)= 6.99dB$. Estos valores indican una eficaz supresión de ruido en la zona de mejor comportamiento,

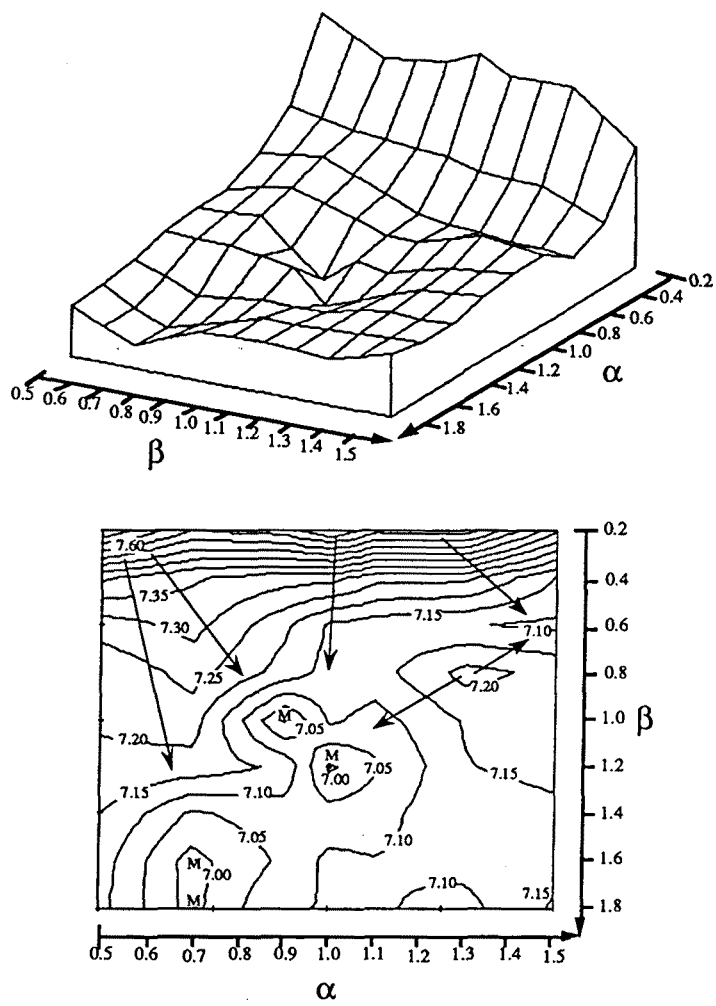


Figura V.21 : Distancia Cepstrum (dB) después de procesar la iteración óptima mediante Filtro de Wiener Generalizado a SNR=9dB para ESCA+AGWN.

donde se obtiene una calidad parecida a la del algoritmo AR3. Sin embargo, el principal inconveniente viene dado por el cuantioso número de iteraciones a procesar. Teniendo en mente esta complejidad de cálculo, se podría considerar una elección inteligente el punto C_2(1.3, 1.8)= 7.08dB, u otros de su vecindad que precisen 3 iteraciones a lo sumo. En comparación a AR3, se obtiene una menor sensibilidad al parámetro exponencial a. Además, la zona más conservadora (b≤0.6) consigue una menor reducción de ruido con el agravante de procesar un número doble de iteraciones.

La mejora en distancia Cepstrum tras 2 ó 3 iteraciones se sitúa alrededor de los 3.5dB para la zona computacionalmente aceptable. Asimismo, la distancia Itakura se reduce desde 8.28dB hasta 3.45dB y la Cosh desde 9.92 hasta 5.81dB. En relación a las medidas temporales, la SNR global sube hasta los 14dB y la SNR segmentada se traslada desde 8.1dB hasta 10.3dB. Las pruebas de audición muestran una buena inteligibilidad y buena calidad, con un nivel de distorsión ligerísimamente inferior al de AR3, aunque el ruido residual es algo superior.

4	A	L	F	A							
BETA	9dB	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5
	0.2	7.75 (6)	7.67 (5)	7.60 (4)	7.61 (5)	7.63 (5)	7.70 (3)	7.62 (3)	7.62 (3)	7.63 (3)	7.52 (8)	7.41 (15)
	0.4	7.41 (5)	7.38 (5)	7.36 (4)	7.34 (5)	7.32 (7)	7.32 (6)	7.27 (6)	7.27 (6)	7.26 (7)	7.21 (9)	7.15 (10)
	0.6	7.28 (4)	7.30 (4)	7.32 (3)	7.26 (6)	7.21 (10)	7.14 (9)	7.12 (10)	7.11 (10)	7.10 (11)	7.10 (9)	7.09 (7)
	0.8	7.21 (5)	7.24 (4)	7.27 (3)	7.22 (5)	7.18 (6)	7.13 (6)	7.13 (6)	7.17 (4)	7.22 (2)	7.20 (2)	7.19 (2)
	1.0	7.20 (4)	7.21 (3)	7.22 (3)	7.11 (4)	6.99 (6)	7.11 (5)	7.08 (7)	7.12 (4)	7.15 (3)	7.17 (3)	7.18 (2)
	1.2	7.19 (3)	7.18 (3)	7,18 (3)	7.16 (3)	7.14 (2)	6.99 (5)	7.04 (5)	7.09 (4)	7.15 (3)	7.16 (3)	7.17 (2)
	1.4	7.15 (3)	7.10 (3)	7.04 (4)	7.06 (4)	7.07 (4)	7.07 (4)	7.08 (4)	7.11 (3)	7.13 (2)	7.12 (2)	7.13 (2)
	1.6	7.13 (3)	7.05 (4)	6.98 (6)	7.02 (5)	7.06 (4)	7.11 (3)	7.10 (3)	7.11 (3)	7.11 (2)	7.12 (2)	7.13 (2)
	1.8	7.12 (3)	7.05 (4)	6.98 (6)	7.04 (4)	7.10 (3)	7.10 (3)	7.11 (3)	7.09 (3)	7.08 (2)	7.12 (2)	7.16 (2)

***Tabla V.5** : Distancia Cepstrum (dB) después de procesar la iteración óptima (indicada entre paréntesis) mediante Filtro de Wiener Generalizado a SNR=9dB para ESCA+AGWN.*

Como conclusión se puede afirmar que los cumulantes de cuarto orden son capaces de combatir estos niveles intermedios de ruido y conseguir una calidad final bastante buena. En comparación a los cumulantes de tercer orden, precisan procesar un número de iteraciones superior y precisan de una mayor ayuda de valores agresivos de a y b, especialmente del parámetro b de sobreestimación de ruido. En realidad el principal inconveniente viene dado por su mayor complejidad de cálculo: precisan alguna iteración más que AR3 y, además, cada iteración representa un mayor tiempo de cálculo debido a los cumulantes de cuarto orden.

V.2.3. Ambientes poco Ruidosos.

Se han simulado degradando la señal de voz original mediante un nivel de ruido tal que la SNR global resultante sea 18dB. El algoritmo AR3 ofrece un exceso de agresividad durante la primera iteración cuando el nivel de ruido bajo. La mayor lentitud y menor poder distorsionador asociada al algoritmo AR4 pueden hacerle apropiado para procesar este nivel de ruido. La duda estriba en el gasto computacional que precisan para conseguirlo. La distancia Cepstrum inicial correspondiente a esta **SNR= 18dB** viene dada por C_0= 8.52dB.

Al procesar la primera iteración, Fig.V.22, se obtiene una zona de equilibrio muy amplia $8a + 5b \geq 13.5$. El efecto distorsión en esta zona es muy insignificante, incluso para valores agresivos C_1 (1.5, 1.8)= 5.47dB. Recuérdase que el algoritmo AR3 empeoraba hasta unos 0.3dB cuando (a, b) tomaban valores altos. Además, este algoritmo AR4 llega a valores más óptimos (unos 0.4dB) $C_{MIN}= C_1$ (1.0, 1.8)= C_1 (1.1, 1.6)= 5.41dB. Obviamente el efecto supresión de ruido domina claramente durante esta primera iteración. La anchura de la zona de equilibrio indica claramente que se ha alcanzado la iteración óptima para la mayoría de puntos (a, b). Los puntos restantes, correspondientes a valores (a, b) menores, muestran peor comportamiento a medida que decrecen los valores de ambos parámetros hasta alcanzar su valor máximo $C_{MAX}= C_1$ (0.5, 0.2)= 6.72dB. El efecto diagonal-circular, característico de la sensibilidad a ambos parámetros, es bastante pronunciado.

Si se procesa la segunda iteración, los valores de distancia empeoran debido a la distorsión ocasionada C_2 (1.5, 1.8)= 6.02dB y al bajísimo nivel de ruido que aún tenia la señal de voz tras la primera iteración. Los mejores valores se obtienen para a=0.5 y b>0.6 y su valor mínimo $C_{MIN}= C_2$ (0.5, 1.2)= 5.34dB es bastante similar al de la iteración previa.

Aunque este valor mínimo es ligerísimamente mejor (0.07dB) no merece la pena procesar una iteración adicional. Esto permite afirmar que la mayoría de pares (a, b) conducen a una total supresión de ruido durante la primera iteración y, entonces, la consideración de una segunda iteración sólamente conlleva asociada una mayor distorsión.

En la Fig.V.23 y la Tabla V.6 se han representado las distancias correspondientes a sus valores mínimos en la iteración óptima asociada a cada par (a, b). Sus valores mínimos se corresponden con los indicados préviamente y se pierde el efecto diagonal-circular debido al salto de 1 a 2 iteraciones, apareciendo una marcada sensibilidad respecto a los valores del parámetro b. Los valores más conservadores siguen mejorando hasta la segunda o tercera iteración, situándose el valor máximo en $C_{MAX}= C_2 (1.5, 0.2)= 5.7$dB. La diferencia entre C_{MAX} y C_{MIN} es muy pequeña (0.3dB) y se constata que cualquier combinación de valores (a,

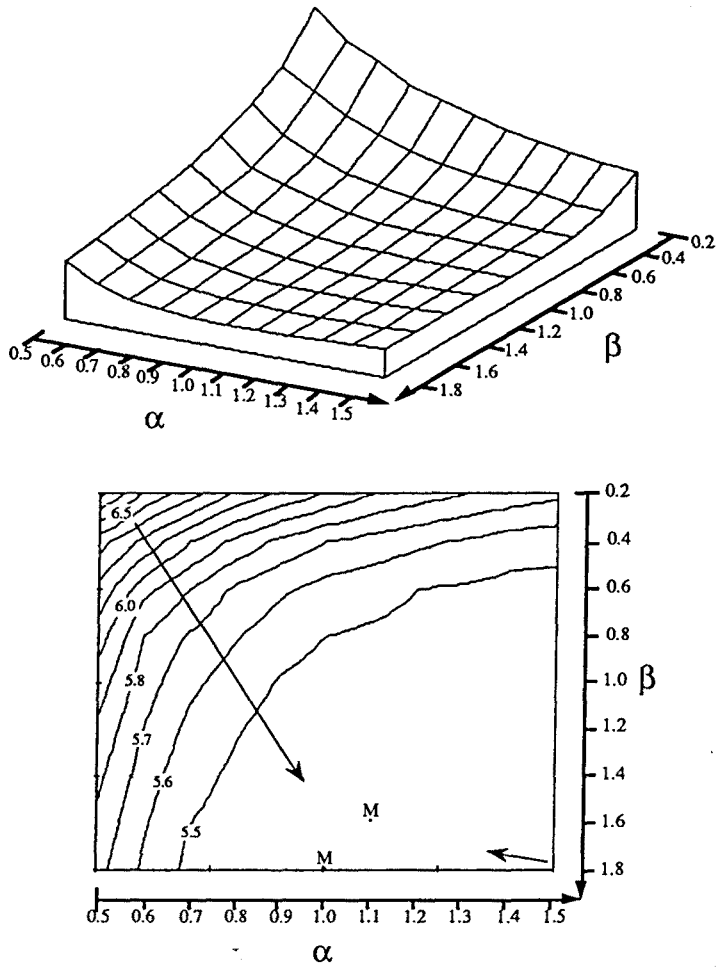


Figura V.22 : Distancia Cepstrum (dB) después de procesar la primera iteración mediante Filtro de Wiener Generalizado a SNR=18dB para ESCA+AGWN.

b) ofrece buenas prestaciones. Desde el punto de vista de cálculo, interesa procesar únicamente la primera iteración y, entonces, los mejores valores corresponden a los puntos (1, 1.8) y (1.1, 1.6).

La distancia Cepstrum en esta primera iteración mejora 3.1dB. En referencia a las otras medidas de distancia espectral, la distancia Itakura decrece desde 6.33dB hasta 2.7dB en la primera iteración y hasta los 2.2dB tras la segunda, debido a que la distorsión afecta, en menor cuantía, a los formantes de la voz. La distancia Cosh se comporta de forma parecida: decrece desde 7.89dB a 4.9dB durante la primera iteración y en la segunda llega hasta los 4.6dB. La SNR global alcanza los 21.8dB mientras la SNR segmentada aumenta de 13.4dB hasta los 17dB.

Se puede concluir afirmando que el algoritmo AR4 obtiene unas prestaciones

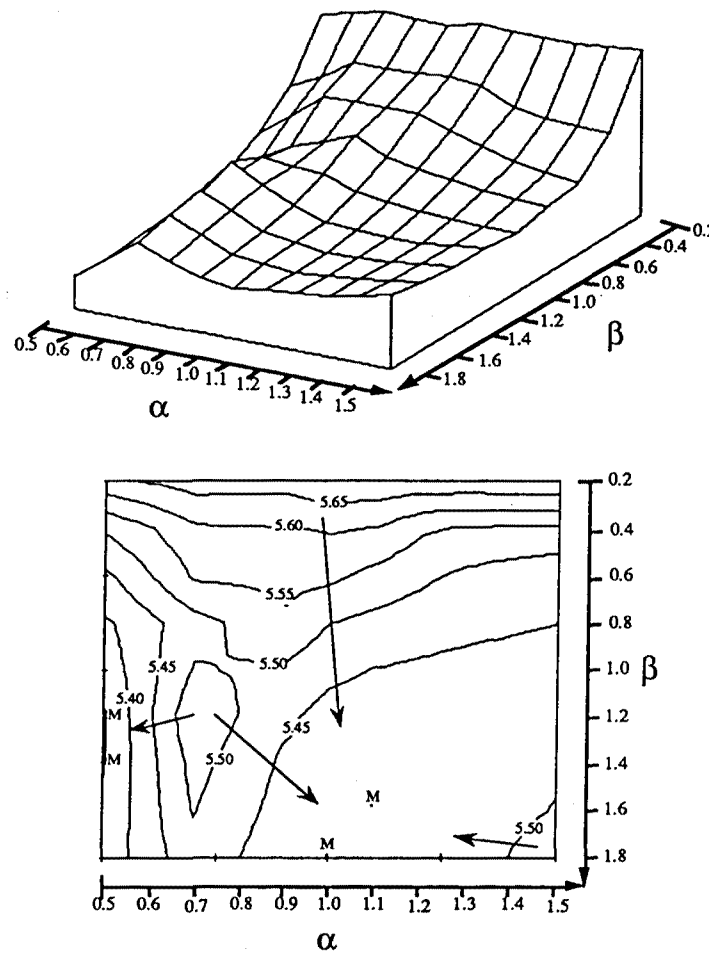


Figura V.23 : Distancia Cepstrum (dB) después de procesar la iteración óptima mediante Filtro de Wiener Generalizado a SNR=18dB para ESCA+AGWN.

ligeramente superiores a las del algoritmo AR3 en presencia de poco ruido. La sobreestimación de ruido ayuda a obtener resultados un poco mejores. La distorsión ocasionada es menor que en el caso AR3, aunque las pruebas de audición no marcan diferencias apreciables. Este algoritmo AR4 presupone un incremento del gasto operacional, aunque éste es poco relevante cuando se procesa únicamente la primera iteración. En comparación al algoritmo AR2, se obtienen prestaciones similares pero con un menor coste cuando se usa el algoritmo AR4 puesto que ahorra unas 2 ó 3 iteraciones de procesado.

4		A				L		F		A		
	18dB	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5
	0.2	5.64 (4)	5.66 (3)	5.67 (2)	5.67 (2)	5.67 (2)	5.69 (2)	5.68 (2)	5.68 (2)	5.68 (2)	5.69 (2)	5.70 (2)
B	0.4	5.55 (3)	5.60 (3)	5.62 (2)	5.62 (2)	5.63 (2)	5.64 (2)	5.68 (2)	5.68 (2)	5.62 (1)	5.61 (1)	5.60 (1)
	0.6	5.44 (2)	5.50 (2)	5.56 (2)	5.57 (2)	5.56 (2)	5.56 (1)	5.53 (1)	5.50 (1)	5.49 (1)	5.48 (1)	5.47 (1)
E	0.8	5.39 (2)	5.44 (2)	5.48 (2)	5.53 (2)	5.54 (1)	5.50 (1)	5.48 (1)	5.47 (1)	5.46 (1)	5.46 (1)	5.45 (1)
	1.0	5.37 (2)	5.44 (2)	5.50 (2)	5.52 (2)	5.49 (1)	5.46 (1)	5.45 (1)	5.44 (1)	5.44 (1)	5.43 (1)	5.43 (1)
T	1.2	5.35 (2)	5.45 (2)	5.54 (2)	5.51 (1)	5.46 (1)	5.43 (1)	5.43 (1)	5.42 (1)	5.42 (1)	5.43 (1)	5.44 (1)
	1.4	5.35 (2)	5.45 (2)	5.55 (2)	5.48 (1)	5.45 (1)	5.43 (1)	5.42 (1)	5.42 (1)	5.42 (1)	5.44 (1)	5.45 (1)
A	1.6	5.36 (2)	5.46 (2)	5.51 (1)	5.46 (1)	5.43 (1)	5.42 (1)	5.41 (1)	5.42 (1)	5.43 (1)	5.44 (1)	5.46 (1)
	1.8	5.37 (2)	5.46 (2)	5.48 (1)	5.45 (1)	5.42 (1)	5.41 (1)	5.42 (1)	5.43 (1)	5.44 (1)	5.45 (1)	5.47 (1)

***Tabla V.6** : Distancia Cepstrum (dB) después de procesar la iteración óptima (indicada entre paréntesis) mediante Filtro de Wiener Generalizado a SNR=18dB para ESCA+AGWN.*

V.3. Estudio de Convergencia del algoritmo de Wiener Generalizado.

En este apartado se realiza un estudio teórico sobre la convergencia del Algoritmo Iterativo de Wiener Parametrizado. Mediante este desarrollo teórico se pretende justificar el comportamiento apreciado en los apartados precedentes del presente capítulo. Para dar generalidad a este desarrollo se considera el factor de desacoplo definido en (IV.41) y (IV.43), así como la notación y definiciones utilizadas en los estudios de convergencia expuestos anteriormente. De esta manera se obtiene conjuntamente la tendencia de convergencia asociada con los algoritmos parametrizados AR2, AR3 y AR4 a partir de la consideración de distintos valores del factor de desacoplo ∂.

Tal como se ha presentado en (V.1) se han añadido los parámetros a y ß a la expresión del filtro de Wiener $\mathbf{W_1}$, para obtener un mejor control sobre las características de éste. Durante la primera iteración, la estimación AR de la voz se realiza a partir de la señal de voz ruidose **x(n)** resultando el siguiente diseño para el filtro de Wiener generalizado:

$$W_1^{1/\alpha} = \frac{P_x}{P_x + ß.P_r} = \frac{1}{1 + \dfrac{ß.P_r}{P_s + \partial.P_r}} = \frac{1}{1 + ß.W_{opt}^{cc}} \qquad (V.4)$$

En la segunda iteración, se diseña el filtro de Wiener a partir de la señal de voz realzada $\mathbf{y_1(n)}$ disponible a la salida del filtro de Wiener $\mathbf{W_1}$ perteneciente a la primera iteración:

$$W_2^{1/\alpha} = \frac{P_{y_1}}{P_{y_1} + ß.P_r} = \frac{1}{1 + \dfrac{ß.P_r}{P_{y_1}}} \qquad (V.5)$$

Dicha señal de voz realzada $\mathbf{y_1(n)}$ resulta de filtrar la señal de voz ruidosa **x(n)** mediante el filtro de Wiener diseñado durante la primera iteración y, entonces, en el dominio frecuencial se verifica:

$$P_{y_1} = P_x \,.\, W_1^2 = (P_s + \partial.P_r)\,.\,W_1^2 \qquad (V.6)$$

sustituyendo (V.6) en (V.5) resulta una relación entre los filtros de Wiener correspondientes a las dos primeras iteraciones:

$$W_3^{1/\alpha} = \frac{1}{1 + \dfrac{ß\,.\,W_{opt}^{cc}}{W_1^2}} = \frac{1}{1 + ß\,.\,W_{opt}^{cc}\,.\,D_1^2} \qquad (V.7)$$

De esta manera resulta la ecuación de recurrencia para el filtro de Wiener Generalizado correspondiente a la iteración i-ésima:

$$W_i^{1/\alpha} = \frac{1}{1 + \beta \cdot W_{opt}^{cc} \cdot D_{i-1}^2} \qquad (V.8)$$

que vista en términos del filtro de Wiener inverso $\mathbf{D_i}$ resulta:

$$D_i = \left[1 + \beta \cdot W_{opt}^{cc} \cdot D_{i-1}^2\right]^{\alpha} \qquad (V.9)$$

Evidentemente la aparición del parámetro a complica considerablemente su resolución. Por simplicidad nos limitaremos a reseñar su comportamiento en relación al algoritmo iterativo de Wiener (a=ß=1). Además, se analiza el caso particular del Filtrado de Wiener Pausado (a=0.5) para evaluar detalladamente sus prestaciones y por extensión trazar las características de este filtro ante variaciones del parámetro a. A continuación se discute como varían las características de cada filtro ante variaciones de uno de estos dos parámetros a y ß.

Fijando el parámetro ß=1 se puede estudiar la tendencia del filtro de Wiener ante distintos valores del parámetro a, pudiéndose analizar el comportamiento de convergencia de $\mathbf{W_i}$ hacia $\mathbf{W_\infty}$, o la de $\mathbf{D_i}$ hacia $\mathbf{D_\infty}$. Para ello se debe constatar que la expresión afectada por el parámetro a es siempre superior a la unidad para cualquier frecuencia donde exista presencia de ruido:

$$1 + \beta \cdot W_{opt}^{cc} \cdot D_{i-1}^2 > 1 \qquad (V.10)$$

Entonces, al considerar valores superiores a la unidad (a>1) se acelera la velocidad de convergencia del algoritmo, tendiendo $\mathbf{D_i}$ de forma más rápida hacia $\mathbf{D_\infty}$. De manera opuesta los valores inferiores a la unidad (a<1) dotan de menor agresividad al algoritmo y, en consecuencia, reducen su velocidad de convergencia.

Si se fija el parámetro a a la unidad, entonces se puede analizar el comportamiento de la convergencia de este algoritmo ante sobreestimaciones (ß>1) y subestimaciones (ß<1) de ruido. En este supuesto se considera el siguiente cambio de variables:

$$d(i) = D_i(f) \qquad (V.11.a)$$

$$r'' = \beta . W_{opt}^{cc}(f) \qquad (V.11.b)$$

se obtiene una solución similar a la obtenida anteriormente en (IV.56):

$$\text{si} \quad 0.25 < r'' \leq 1 \quad \Rightarrow \quad d(\infty) \longrightarrow \infty \quad \Rightarrow \quad W_\infty = 0 \qquad \text{(V.12.a)}$$

$$\text{si} \quad r'' = 0.25 \quad \Rightarrow \quad d(\infty) = 2 \quad \Rightarrow \quad W_\infty = 0.5 \qquad \text{(V.12.b)}$$

$$\text{si} \quad 0 \leq r'' < 0.25 \quad \Rightarrow \quad d(\infty) \text{ converge} \quad \Rightarrow \quad 0.5 < W_\infty \leq 1 \qquad \text{(V.12.c)}$$

Vamos a analizar más detalladamente la expresión (V.12.c). Deshaciendo el cambio de variable, la condición **r"<0.25** se traduce en:

$$W_{opt}^{cc} = \frac{P_r}{P_s + \partial \, . \, P_r} < \frac{1}{4ß} \qquad \text{(V.13)}$$

Despejando P_S se obtiene una relación entre las densidades espectrales de la señal de voz y del ruido:

$$P_s > (4ß - \partial) \, . \, P_r \qquad \text{(V.14)}$$

Sustituyendo en la expresión del filtro óptimo de Wiener, se obtiene la condición anterior (V.13) en términos de W_{opt}:

$$W_{opt} = \frac{P_s}{P_s + P_r} > \frac{(4ß - \partial).P_r}{(4ß - \partial).P_r + P_r} = \frac{(4ß - \partial)}{(4ß + 1 - \partial)} = C(\partial, ß) \qquad \text{(V.15)}$$

De este modo el filtro de Wiener W_i converge a valores no nulos cuya cota superior viene dada por el filtro óptimo de Wiener, para aquellas frecuencias donde $W_{opt} \geq C(\partial, ß)$:

$$W_\infty(f) = 0 \quad \text{si} \quad W_{opt} < C(\partial, ß) \qquad \text{(V.16.a)}$$

$$W_\infty(f) = 0.5 \quad \text{si} \quad W_{opt} = C(\partial, ß) \qquad \text{(V.16.b)}$$

$$0.5 < W_\infty(f) \leq W_{opt}(f) \quad \text{si} \quad W_{opt} > C(\partial, ß) \qquad \text{(V.16.c)}$$

$C(\partial, ß)$	$ß = 2$	$ß = 1$	$ß = 0.25$
$\partial = 1$	0.875	0.75	0
$\partial = 0$	0.888	0.8	0.5

Tabla V.7 : *Valores de la constante $C(\partial, ß)$ según el factor de desacoplo ∂ y el parámetro $ß$.*

En la ecuación de recurrencia (V.9) se puede observar como la velocidad de convergencia aumenta cuando se consideran sobreestimaciones de ruido. Desafortunadamente la Tabla V.7 muestra como el valor de la constante $C(\partial, ß)$ aumenta con el parámetro ß y, en consecuencia, se incrementa la distorsión por picado espectral. De forma análoga, las subestimaciones de ruido conducen a una menor distorsión, aunque la velocidad de convergencia se ralentiza. Obsérvese como para el caso de acoplo voz-ruido (∂=1) y valores del parámetro ß bastante pequeños (ß≤0.25), se puede obtener una convergencia bastante buena del filtro $\mathbf{W}_i$ hacia el filtro óptimo $\mathbf{W}_{opt}$. En este supuesto la dificultad viene impuesta por su elevada complejidad de coste de cálculo, puesto que se precisa un número de iteraciones demasiado elevado. De este modo se han justificado analíticamente los comportamientos observados con anterioridad para los algoritmos generalizados AR3 y AR4.

V.3.1. Filtrado de Wiener Pausado. Estudio de Convergencia.

El análisis anterior derivó el estudio analítico hacia otras implementaciones distintas al Filtrado de Wiener clásico propuesto por Lim y Oppenheim [Lim-79]. Se pretende evaluar el algoritmo iterativo de Wiener según un valor concreto del parámetro a distinto a la unidad. Se ha considerado un algoritmo que introduzca una menor distorsión y cuya convergencia sea hacia el filtro óptimo de Wiener. Para ello se ha particularizado el parámetro a al valor a=0.5, correspondiente al Filtrado por Espectro de Potencia considerado previamente en (II.28). En este algoritmo se propone una forma distinta de realizar las sucesivas aproximaciones hacia el filtro óptimo W_{opt}: se usa la raiz cuadrada del módulo del filtro de Wiener (a=0.5) para todas las iteraciones, a excepción de la última iteración donde se ejecuta un filtrado de Wiener estimado con a=1.

Según el análisis desarrollado previamente y los estudios de evaluación discutidos en los apartados V.1 y V.2 se espera, a priori, una menor velocidad de convergencia combinada con un efecto distorsión menos importante. En la Fig.V.24 se puede apreciar un nivel de distorsión bastante menor para el algoritmo de Wiener Pausado, en relación al algoritmo de Wiener clásico. Se han representado los espectros LPC para las señales de voz original, voz ruidosa y señales de voz realzada resultantes en las tres primeras iteraciones para cada algoritmo (AR2 y AR2 Pausado) y para los dos niveles de ruido considerados: ambientes muy ruidosos y entornos con un nivel intermedio de ruido. Es decir, para el algoritmo

iterativo de Wiener Pausado se han ejecutado las dos primeras iteraciones con un valor a=0.5 durante el diseño del filtro y la tercera, y última, iteración con el filtro de Wiener clásico (a=1).

Independientemente del nivel de ruido introducido, cuando se ejecuta el algoritmo de Wiener pausado (a=0.5) se observa como la degradación en los valles espectrales es mucho menor y, además, el estrechamiento del ancho de banda de los formantes es bastante menos notorio con el discurrir de las sucesivas iteraciones. Nótese que se ha representado el caso correspondiente a las estadísticas de segundo orden y, en consecuencia, estas diferencias

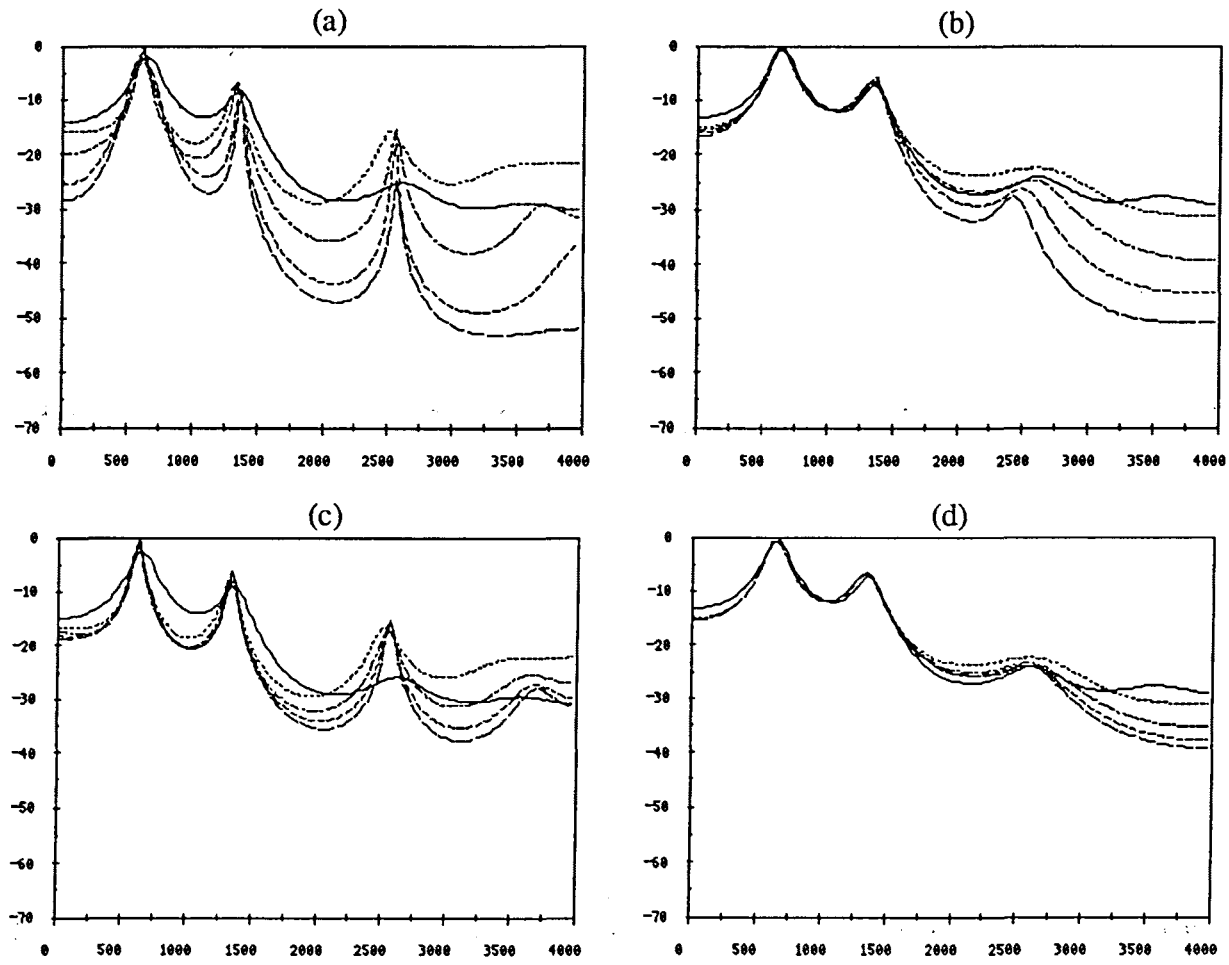


Figura V.24 : Espectro LPC usando estadísticas de segundo orden para un sonido sintético en ausencia de ruido (trazo continuo) y en presencia de ruido (trazo discontinuo) para las tres primeras iteraciones de los casos siguientes: a) a=ß=1 y SNR=0dB; b) a=ß=1 y SNR=9dB; c) a=0.5, ß=1 y SNR=0dB; d) a=0.5, ß=1 y SNR=9dB

serían bastante más apreciables si se ejecutara alguno de los algoritmos basados en las estadísticas de orden superior.

Seguidamente se realiza el estudio de convergencia correspondiente a este algoritmo iterativo de Wiener Pausado. Durante la primera iteración se estima el filtro de Wiener según la expresión siguiente:

$$W_1^2 = \frac{P_x}{P_x + P_r} = \frac{1}{1 + \dfrac{P_r}{P_s + \partial . P_r}} = \frac{1}{1 + W_{opt}^{cc}} \qquad (V.17)$$

En la segunda iteración, se diseña el filtro de Wiener a partir de la señal de voz realzada $\mathbf{y_1(n)}$ disponible a la salida del filtro de Wiener $\mathbf{W_1}$ perteneciente a la primera iteración:

$$W_2^2 = \frac{P_{y_1}}{P_{y_1} + P_r} = \frac{1}{1 + \dfrac{P_r}{P_{y_1}}} = \frac{1}{1 + \dfrac{W_{opt}^{cc}}{W_1^2}} \qquad (V.18)$$

sustituyendo (V.17) en (V.18):

$$W_2^2 = \frac{1}{1 + W_{opt}^{cc} \cdot \left(1 + W_{opt}^{cc}\right)} = \frac{1}{1 + W_{opt}^{cc} + \left(W_{opt}^{cc}\right)^2} \qquad (V.19)$$

En la tercera iteración se diseña el filtro de Wiener a partir de la señal de voz realzada resultante en la segunda iteración.

$$W_3^2 = \frac{P_{y_2}}{P_{y_2} + P_r} = \frac{1}{1 + \dfrac{P_r}{P_{y_2}}} = \frac{1}{1 + \dfrac{W_{opt}^{cc}}{W_2^2}} \qquad (V.20)$$

sustituyendo (V.19) en (V.20):

$$W_3^2 = \frac{1}{1 + W_{opt}^{cc} \cdot \left(1 + W_{opt}^{cc} + \left(W_{opt}^{cc}\right)^2\right)} = \frac{1}{\sum_{j=0}^{3} \left(W_{opt}^{cc}\right)^j} \qquad (V.21)$$

De esta manera resulta la siguiente ecuación para el filtro de Wiener Pausado correspondiente a la iteración i-ésima:

$$W_i^2 = \frac{1}{\sum_{j=0}^{i} \left(W_{opt}^{cc}\right)^j} \qquad (V.22)$$

donde se pude aplicar la formula de la suma de una serie geométrica cuya razón es inferior a la unidad:

$$W_i^2 = \frac{1 - W_{opt}^{cc}}{1 - \left(W_{opt}^{cc}\right)^{i+1}} \qquad (V.23)$$

Para obtener una total convergencia se supone que se procesa un número infinito de iteraciones resultando la siguiente expresión para el filtro de convergencia:

$$W_\infty^2 = 1 - W_{opt}^{cc} \Rightarrow W_\infty = \sqrt{1 - W_{opt}^{cc}} = \sqrt{W_{opt}} \cdot \left[\frac{(P_s + P_r).(P_s - P_r.(1-\partial))}{P_s \,.\, (P_s + \partial.P_r)}\right]^{\frac{1}{2}} \qquad (V.24)$$

Si particularizamos para el caso de las estadísticas de segundo orden (∂=1) esta expresión se reduce a la siguiente:

$$W_\infty = \sqrt{W_{opt}} \qquad (V.25)$$

es decir, el filtrado de Wiener bajo la consideración del parámetro a=0.5 tiende hacia la raiz cuadrada del filtro óptimo de Wiener. Sin embargo, este algoritmo de Wiener Pausado realiza la última iteración tomando a=1. De este modo se supone la ejecución de un número de iteraciones lo suficientemente elevado para que se pueda suponer que el filtro ha convergido hacia la raiz cuadrada del filtro óptimo de Wiener:

$$W_{\infty-1} = \sqrt{W_{opt}} \qquad (V.26)$$

y entonces se ejecuta la última iteración diseñando el filtro de Wiener clásico (a=1) a partir de la señal de voz realzada $\mathbf{y_{\infty-1}(n)}$ obtenida en la iteración anterior:

$$W_\infty = \frac{P_{y_{\infty-1}}}{P_{y_{\infty-1}} + P_r} = \frac{1}{1 + \frac{W_{opt}^c}{W_{\infty-1}^2}} = \frac{W_{opt}}{W_{opt} + W_{opt}^c} = W_{opt} \qquad (V.27)$$

De este modo se demuestra como este algoritmo iterativo de Wiener Pausado presenta un filtro de Wiener $\mathbf{W_i}$ que converge exactamente hacia el filtro óptimo de Wiener, bajo las hipótesis anteriores. Obsérvese que según este análisis teórico existe la posibilidad de llegar al filtro óptimo de Wiener, sin precisar la disponibilidad de la señal de voz original ($\mathbf{P_s(w)}$) de forma aislada en relación al ruido.

Si se considera el supuesto ideal de total desacoplo voz-ruido (∂=0) la expresión (V.24) se reduce a la siguiente:

$$W_{\infty-1} = \sqrt{W_{opt}} \cdot \left[1 - \left(\frac{P_r}{P_s} \right)^2 \right]^{\frac{1}{2}} \qquad (V.28)$$

que conduce en la iteración infinita al filtro:

$$W_{\infty} = W_{opt} \cdot \left[1 - \left(\frac{P_r}{P_s} \right)^2 \right] = 1 - \frac{P_r}{P_s} \qquad (V.29)$$

En este caso particular el filtro iterativo tiende al óptimo solamente para aquellas frecuencias donde no hay presencia de ruido. A medida que exista un nivel de ruido superior, en relación a la energía de la señal, el filtro de convergencia alcanza un valor inferior en relación al valor del filtro óptimo a esta frecuencia determinada. El filtro puede llegar a cancelar la información contenida en una determinada frecuencia si en ésta se igualan los niveles de señal y ruido. Si se considera una energía de ruido superior a la de la señal de voz, entonces, la expresión (V.23) deja de ser válida y se debe reconsiderar la expresión (V.22) observándose como el filtro también tiende a cancelar estas frecuencias.

Como conclusión a este estudio analítico sobre la convergencia del algoritmo del Algoritmo de Wiener Pausado se puede afirmar que la distorsión es muy pequeña cuando se aplican las estadísticas de segundo orden, puesto que al procesar un número muy elevado de iteraciones este filtro iterativo converge hacia el filtro óptimo de Wiener. A pesar de presentar un nivel de distorsión bastante menor en relación al algoritmo iterativo de Wiener (a=1), cuando se consideran las estadísticas de orden superior la distorsión por picado espectral aumenta considerablemente en relación al supuesto de segundo orden (∂=1).

Las pruebas realizadas con ruido blanco y ruido de motor muestran, por una parte, como el algoritmo es más lento en su velocidad de convergencia y, al mismo tiempo, menos agresivo en los valles espectrales. Por otra parte, la menor velocidad de convergencia conduce a la necesidad de ejecutar un número de iteraciones bastante superior, haciéndolo inviable para una aplicación real. La degradación que se introduce al sobrepasar la iteración óptima es ínfima en comparación al filtrado iterativo de Wiener. En resumen, el filtrado Pausado de Wiener ayuda a controlar la distorsión espectral derivada del efecto picado de los formantes, mejorando la inteligibilidad a costa de una menor calidad. Debe recordarse, también, que el algoritmo de tercer orden AR3 resulta demasiado agresivo en algunas aplicaciones, como por

ejemplo ante niveles bajos de ruido y, entonces, este algoritmo de Wiener Pausado puede ayudar a suprimir el mismo ruido pero ocasionando una menor distorsión.

V.4. Métodos de Promediado de Coeficientes AR.

Una variante de filtrado similar a la que vamos a presentar en este apartado fue planteada originariamente por Hansen y Clements en [Hans1-88] dentro de un conjunto de propuestas cuya función era la de actuar como restricciones sobre la estimación de los coeficientes $\mathbf{a_k}$, de manera que significaran un cierto control sobre la posición y el ancho de banda de los formantes de la voz. El objetivo principal era, sin duda, reducir los efectos propios del filtrado de Wiener con modelado AR (Capítulo IV.5), alguno de los cuáles ya hemos visto anteriormente. Sin embargo, como podremos comprobar, la variante que proponemos supone además una importante mejora respecto del método de estimación básico, sea éste de orden dos o superior. Distintas versiones correspondientes a este algoritmo basado en el promediado de coeficientes han sido publicados en [Sala-94b] y [Sala-94c].

V.4.1. La Ponderación Intertrama (IF).

Considerando el mismo esquema de procesado que hemos tratado hasta ahora, se observa en las pruebas realizadas que la mayor parte de eliminación de ruido interferente se produce en las dos primeras iteraciones del algoritmo, si bien en el caso de cumulantes de tercer orden es ya notable tras sólo una iteración. Esto se debe a que en la segunda iteración estimamos el filtro a partir de una señal más libre de ruido. En ambientes muy ruidosos, incluso se ha visto como durante tres o cuatro iteraciones se reduciendo apreciablemente el nivel de ruido presente en la señal de voz.

Por otro lado, sabiendo que el tracto vocal es un sistema mecánico con inercia que no puede variar bruscamente, estamos en condiciones de suponer que dicho tracto presenta suficiente estacionariedad entre dos tramas solapadas consecutivas como para poder afirmar que los valores de los parámetros que definen dichas tramas deben moverse dentro de un margen de deriva acotado. Esto nos sugiere que, situándonos en una trama **n** de la señal de voz ruidosa, disponemos "a priori" de cierta información de esos parámetros gracias a que hemos procesado ya su trama inmediatamente anterior. Se vislumbra la posibilidad de realizar un alisado o promediado entre tramas consecutivas de alguna característica propia de la voz,

con el doble objetivo de ayudar o acelerar el algoritmo de filtrado y de suavizar los efectos de posibles distorsiones, producidas por cambios bruscos de posición o forma de los formantes.

Intentar directamente promediar la posición de los polos del modelo AR nos llevaría al cálculo de las raíces de un polinomio de grado 10, algo que sin duda requiere un elevado gasto en tiempo de cálculo. Con la intención de evitar estos inconvenientes, se pensó en promediar directamente los coeficientes a_k. Dicha ponderación entre dos tramas consecutivas, o ponderación intertrama, puede verse como el resultado de alisar sus envolventes o espectros LPC respectivos. En efecto, sean dos estimaciones de los coeficientes, $\mathbf{a'_k}$ y $\mathbf{a''_k}$, tales que en una región determinada cumplen la siguiente relación:

$$\sum_{k=0}^{p} a'_k \cdot z^k \leq \sum_{k=0}^{p} a''_k \cdot z^k \qquad \text{(V.30)}$$

entonces se desprende que los coeficientes ponderados mediante un factor μ:

$$\bar{a}_k = \mu \cdot a'_k + (1 - \mu) \cdot a''_k \qquad \text{,,} \qquad 0 \leq \mu \leq 1 \qquad \text{(V.31)}$$

cumplen las desigualdades:

$$\sum_{k=0}^{p} a'_k \cdot z^k \leq \sum_{k=0}^{p} \bar{a}_k \cdot z^k \leq \sum_{k=0}^{p} a''_k \cdot z^k \qquad \text{(V.32)}$$

Es decir, sus envolventes LPC cumplen, en un cierto margen frecuencial, la desigualdad:

$$\frac{1}{\sum_{k=0}^{p} a'_k \cdot z^k} \geq \frac{1}{\sum_{k=0}^{p} \bar{a}_k \cdot z^k} \geq \frac{1}{\sum_{k=0}^{p} a''_k \cdot z^k} \qquad \text{(V.33)}$$

En la primera iteración de cada trama de voz ruidosa es donde aparecen las mayores dificultades en vistas a estimar el filtro de Wiener, puesto que se debe estimar el modelado AR de la voz original a partir de la señal de voz ruidosa en lugar de la señal de voz realzada tal como sucede en las restantes iteraciones. Sin embargo, bajo la hipótesis de una cierta estacionariedad del tracto vocal durante dos tramas consecutivas, solapadas al 50 por ciento, se dispone de un modelado AR obtenido durante la trama anterior a partir de unas condiciones de menor presencia de ruido. De este modo la estimación AR realizada en la primera iteración de cada trama se puede ayudar mediante alguna de las estimaciones AR realizadas durante las distintas iteraciones de la trama precedente. Mediante este promediado de coeficientes se acelera la velocidad de convergencia y, en consecuencia, la reducción de ruido durante la primera iteración debe aumentar. Por esta razón no se considera alisado de coeficientes a partir de la segunda iteración de cada trama, considerándose que dicha

estimación se realiza sobre una señal de voz realzada cuyo contenido de ruido es bastante menor.

En esta metodología se tratan dos tipos de coeficientes AR según su procedencia y deben ser diferenciados:

- los procedentes de una estimación directa sobre una trama de voz ruidosa, o voz realzada, los vamos a notar en minúscula a_k;

- los procedentes de una ponderación de coeficientes pertenecientes a distintas tramas consecutivas se notan en mayúscula A_k.

Entonces, diseñamos el filtro de Wiener correspondiente a la primera iteración de cada trama con unos nuevos coeficientes A_k, obtenidos aplicando la siguiente combinación lineal de estimaciones AR:

$$A_k(n, 1) = \mathbf{IF} \cdot a_k(n, 1) + (1 - \mathbf{IF}) \cdot a_k(n-1, \mathbf{PFI})$$
$$0 \le k \le p \quad ,, \quad 1 \le PFI \le Max\_Iter \quad ,, \quad 0 \le IF \le 1 \qquad \text{(V.34)}$$

donde **n** es la trama actual, **PFI** (Previous Frame Iteration) es la Iteración de la Trama Anterior cuyos coeficientes a_k consideramos para ayudar a la primera estimación de la trama en curso, **IF** (Interframe Factor) es el Factor Intertrama, y los a_k.son los coeficientes estimados directamente de la trama de voz contaminada con ruido. En esta nueva notación de los coeficientes, $a_k(n,i)$, la primera componente es el número de trama y la segunda, el número de iteración de dicha trama. El Factor Intertrama nos indica el tanto por uno de la estimación AR obtenida en la trama actual que se pone en la mezcla con la estimación AR procedente de la iteración PFI de la trama anterior. El valor **Max_Iter** se ha limitado a 5 debido a las características de distorsión propias del algoritmo iterativo de Wiener, descritas en los Capítulos IV.5 y V.3.

El filtro de Wiener correspondiente a la segunda y demás iteraciones se estima a partir de una señal más limpia de ruido, debido en parte a la ayuda que supone el promediado intertrama efectuado durante la primera iteración. En consecuencia no se considera alisado de coeficientes (**IF=1**) durante estas iteraciones:

$$A_k(n, iter) = \mathbf{IF} \cdot a_k(n, iter) \quad ,, \quad 2 \le iter \le Max\_Iter \qquad \text{(V.35)}$$

siendo **iter** el número de iteración de la trama en curso. Hay que hacer notar que estos $a_k(n, iter)$, para **iter**≥2, no son los mismos que obtiene el algoritmo clásico (**IF=1**), pues

ahora evolucionan a partir de una primera iteración ponderada. En la Fig.V.25 se ha representado esquemáticamente esta metodología de alisado de coeficientes. .

A modo de resumen, disponemos de dos parámetros de control sobre la combinación lineal de coeficientes: el factor intertrama **IF** y la iteración de la trama anterior **PFI** seleccionada durante la ponderación. Al principio de cada zona con actividad de voz será necesario inicializar el parámetro **IF=1**, puesto que en este caso la trama anterior (silencio o sólo ruido) no aporta información alguna. Cuando **IF=1** no tiene sentido considerar el segundo parámetro, **PFI**, ya que entonces obtenemos un modelado AR procedente únicamente de la trama actual. Pero si se decide considerar información de la trama anterior **(IF<1)**, entonces deberemos estudiar y decidir qué número de iteración, **PFI**, de la trama anterior puede resultar más favorable para un buen modelado. Este factor **PFI** dependerá básicamente del algoritmo utilizado y del nivel y características del ruido presente en la señal de voz. En este supuesto no resultará muy ventajosa la elección de un valor **PFI** procedente de una iteración donde se ha alcanzado la saturación entre los efectos de supresión de ruido y distorsión espectral, pues seguramente aportará una cierta distorsión que se acumula a la originada por la no estacionariedad del tracto vocal. Así, la elección de **PFI** debe corresponder a un compromiso entre menor presencia de ruido y baja distorsión.

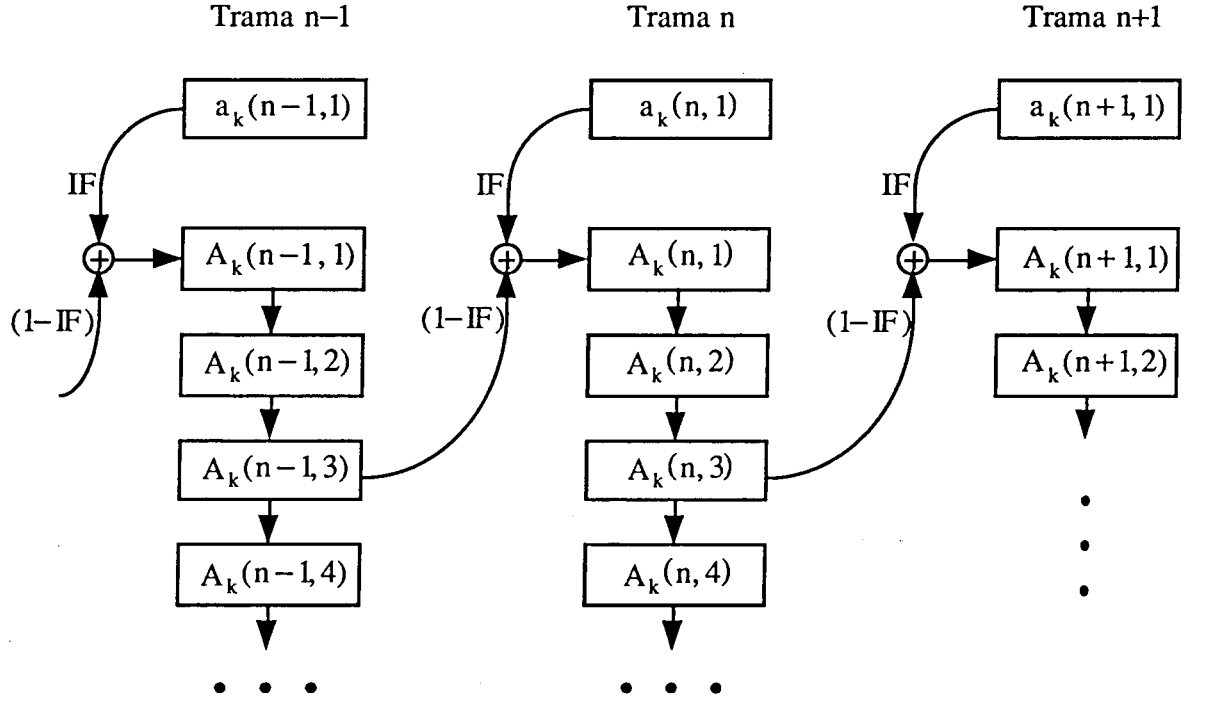


Figura V.25 : Metodología de ponderación de coeficientes AR entre dos tramas consecutivas de voz particularizado para un valor PFI=3.

V.4.1.1. Algoritmo AR2 con Promediado Intertrama (AR2_IF).

Vamos a ver ahora los resultados que se obtienen al aplicar la Ponderación Intertrama a los coeficientes cuando su estimación se realiza con estadísticas de segundo orden, es decir, mediante el método de las correlaciones AR2. El objetivo de un promediado entre los coeficientes $a_k(n, 1)$ de una trama de señal con los $a_k(n-1, PFI)$ de su trama inmediatamente anterior es doble:

- en primer lugar, para intentar suavizar de algún modo los cambios bruscos entre estimaciones consecutivas que podían dar lugar a espurios o distorsiones no deseados en el proceso de filtrado;

- en segundo lugar , para ayudar al sistema a obtener una mejor y más rápida estimación de los coeficientes de la trama en curso, dada la información de que disponemos 'a priori' al conocer ya las estimaciones AR de la trama anterior. Recordemos que este supuesto se sustenta en la estacionariedad de la voz; el tracto vocal no puede variar bruscamente de una trama a la siguiente (solapadas al 50%).

Tal como se ha comentado previamente entran en juego dos nuevos parámetros: el Factor Intertrama **IF** y la iteración de la trama anterior con cuyos coeficientes promediamos, **PFI**. Se han considerado valores de **PFI** entre 1 y 5, y cada caso ha sido analizado para un barrido de **IF** en todo su margen posible, o sea, entre 0 y 1. Las medidas se han efectuado sobre la frase ESCA degradada con ruido AGWN, para niveles de SNR_G de 0 y 9dB.

V.4.1.1.1. Ambientes altamente ruidosos.

En la Fig.V.26 podemos ver los resultados obtenidos para un nivel de ruido elevado (SNR_G=0dB). Estos resultados se han representado en términos de distancia Cepstrum e Itakura, cuando hemos procesado únicamente la primera iteración del algoritmo. Se observa que éstos están dentro de la lógica que cabía esperar de un promediado de este tipo. Para **PFI=1** aparecen unos resultados bastante lógicos: no hay ninguna mejoría respecto del caso

sin ponderar (**IF=1**). Estamos promediando con los coeficientes AR de la misma iteración pero de la trama anterior, es decir, con unos a_k con el mismo grado de fiabilidad que los de la trama en curso, y eso no ayuda en nada a la progresión del sistema (salvo en una cierta reducción en la varianza de la estimación del método).

Sin embargo, a medida que aumenta **PFI** empiezan a notarse los efectos positivos del promediado intertrama y los resultados son cada vez mejores. Considerando que el algoritmo básico alcanza su óptimo en la cuarta iteración, parece normal que los mejores resultados se obtengan al ayudar a la primera iteración de cada trama con los coeficientes estimados en la cuarta iteración de la trama anterior (**PFI=4**). Aunque en realidad con **PFI=5** las medidas sean levemente mejores, desestimaremos esta opción por la carga de distorsión que conlleva, derivada de la saturación que se produce con AR2 a partir de la cuarta iteración. Este efecto se aprecia fácilmente en ambas gráficas; de **PFI=4** a **PFI=5** no hay apenas evolución.

Independientemente del **PFI** considerado se observa también que los mínimos de cada curva se producen para valores bajos del factor **IF**, entre 0.1 y 0.4, o sea, que el sistema parece funcionar mejor mediante estimaciones que contengan más información (70-80%) de

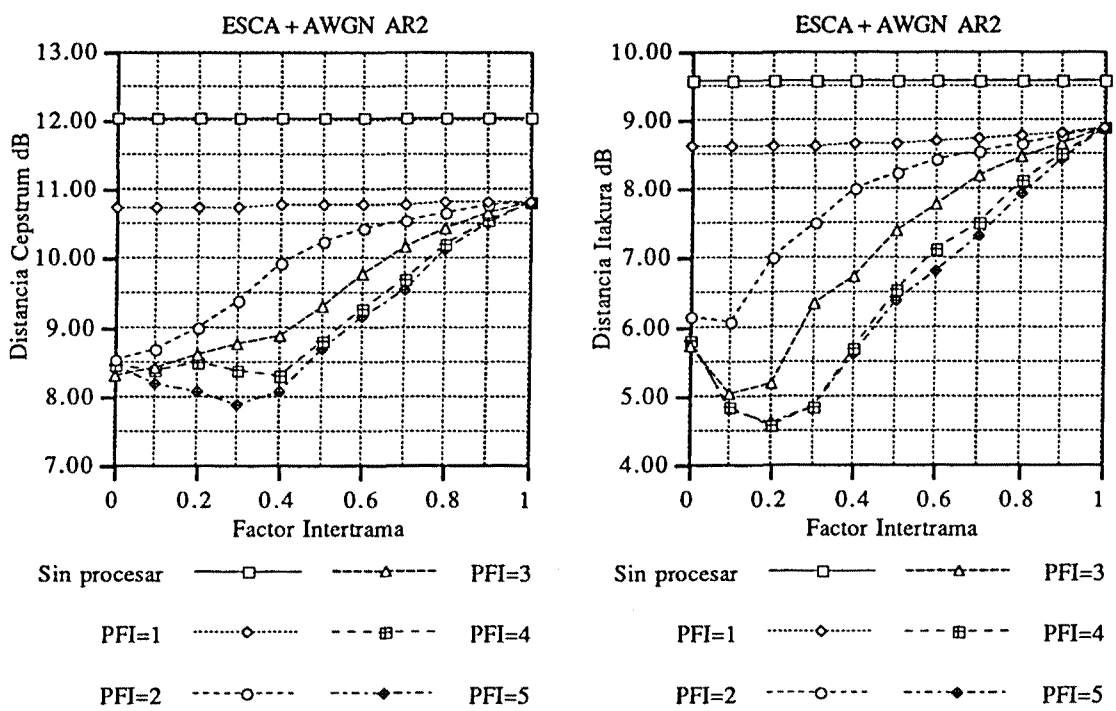


Figura V.26 : Distancias Cepstrum e Itakura durante la primera iteración del algoritmo AR2_IF para SNR_G=0dB.

la trama anterior, pero de mayor calidad, que de la trama actual (20-30%), de peor calidad.

No hay que olvidar, sin embargo, la carga distorsionante que la ponderación intertrama conlleva. Para tener una idea de dicha distorsión hemos efectuado las mismas medidas sobre señal original sin ruido una vez ésta ha sido procesada del mismo modo. Un filtrado ideal no tocaría la señal y consecuentemente las distancias espectrales serían idénticamente nulas. Por tanto, sabremos en qué grado nuestro filtrado distorsiona la señal en la medida de que las distancias espectrales se alejen más o menos de cero.

Los resultados obtenidos se presentan en la Fig.V.27. Se aprecia como las curvas de distorsión aumentan progresivamente a medida que disminuye el Factor Intertrama. Por tanto, los valores de **IF** correspondientes a mínimas distancias espectrales son también los que introducen mayor distorsión. Hay que tener en cuenta, además, que esta distorsión aumentará en las sucesivas iteraciones de filtrado. En consecuencia, todos los valores de **IF** inferiores a 0.5 quedan totalmente descartados y habrá que centrar nuestro estudio en el margen de **IF=0.5** a **IF=1.0**.

Se observa también que la distorsión aumenta con **PFI**, y aunque parece saturarse a partir de **PFI=3**, los tests de audición demuestran que el grado de inteligibilidad continúa sufriendo una lenta erosión. Habrá que seleccionar, entonces, una pareja de valores **IF** y **PFI** que signifiquen un buen compromiso entre la mejora que puedan aportar (y por tanto

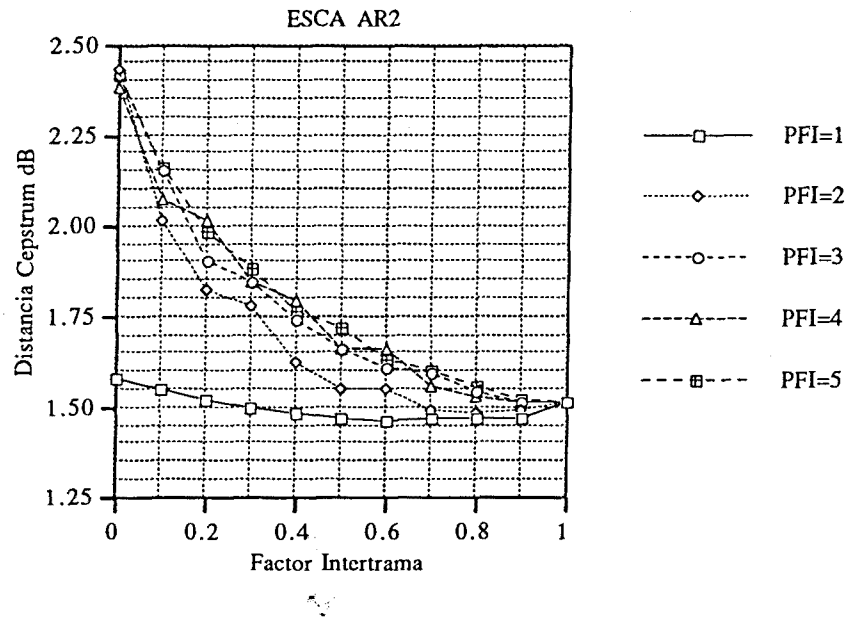


Figura V.27 : Distorsión introducida con el método AR2_IF durante la primera iteración.

velocidad de convergencia del algoritmo) y distorsión introducida por el método AR2 con ponderación intertrama de coeficientes cuando se consideren esos valores.

En el Capítulo IV.2 se ha evaluado el comportamiento del algoritmo AR2. Se aprecia como la supresión de ruido es bastante uniforme durante las 3 ó 4 primeras iteraciones, cuando se consideran ambientes altamente ruidosos. De este modo los resultados facilitados por el algoritmo AR2_IF concuerdan plenamente con el comportamiento asociado con el algoritmo AR2. Si optamos por **PFI=4** y estudiamos lo que sucede con los diferentes valores de **IF** obtenemos los resultados representados en la Fig.V.28.

La Fig.V.28.b representa de forma más nítida la evolución de la reducción de ruido a lo largo de las sucesivas iteraciones para algunos de los valores viables del factor **IF**. También se puede ver como cortes verticales provenientes de la Fig.V.28.a en aquellos valores **IF** cuyo nivel de distorsión es aceptable. Puesto que las medidas decrecen con **IF**, nos decidimos por un valor de este parámetro entre 0.6 y 0.7, en la zona límite a partir de la cual el grado de distorsión empieza a ser excesivo.

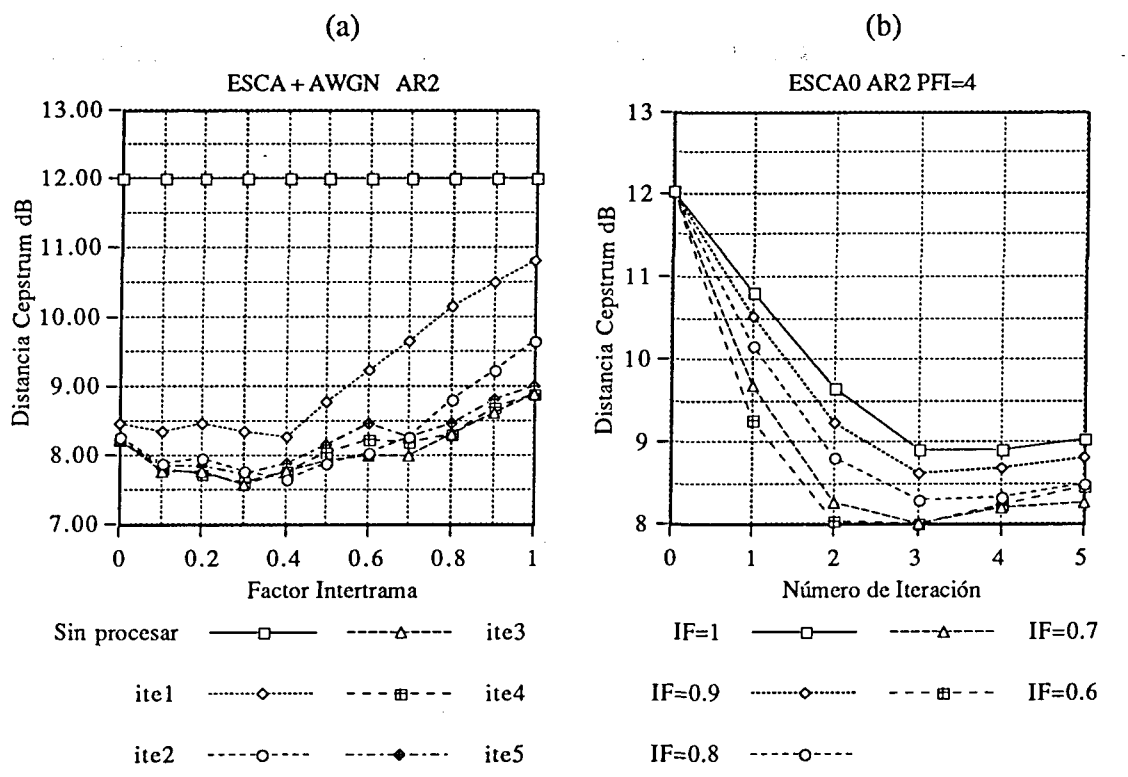


Figura V.28 : Comportamiento del algoritmo AR2_IF para PFI=4 y SNR_G=0dB.

0dB	SNR_{Global}	$SNR_{Seg.}$	ITAKURA	COSH	CEPSTRUM
Original	0.02	0.76	9.57	11.66	12.02
1 iter.	7.89	4.74	7.08	8.98	9.25
2 iter.	9.09	6.21	5.82	7.99	8.06
3 iter.	9.21	6.49	5.17	7.53	8.02
4 iter.	9.12	6.52	5.24	7.61	8.22
5 iter.	9.06	6.51	5.30	7.76	8.46

Tabla V.8** : Evolución del algoritmo AR2_IF a SNR_G=0dB para los valores **IF=0.6** y **PFI=4.

De este modo se obtiene un valor en distancia Cepstrum 1.8dB menor que el obtenido sin ponderar (**IF=1**), obteniéndose una reducción de 2.8dB en términos de distancia Cepstrum tras procesar únicamente la primera iteración. Las distancias Cosh e Itakura, a su vez, se reducen en términos similares, mientras que las medidas temporales (SNR) mantienen el mismo nivel. Además, como puede verse claramente en la Tabla V.8, la iteración óptima se consigue cuando se han procesado tan sólo 3 iteraciones de filtrado, mientras que el algoritmo básico necesitaba 4. Nótese como todos los valores del algoritmo AR2_IF, representados en la Fig.V.28, conducen a mejores reducciones de ruido en comparación al algoritmo AR2.

V.4.1.1.2. Ambientes con un nivel intermedio de ruido.

Si analizamos lo que sucede para un nivel intermedio de ruido, $\mathbf{SNR_G=9dB}$, obtenemos unos resultados ciertamente similares a lo que sucede con 0dB. A diferencia del caso anterior ($\mathbf{SNR_G=0dB}$), al existir menos ruido a eliminar no se precisa tanta ayuda de la trama anterior, y por eso los valores mínimos se sitúan en torno a valores **IF** algo más elevados, entre 0.4 y 0.6, tal como se muestra en la Fig.V.29 . Recuérdase, no obstante, que la distorsión ocasionada (Fig.V.27) obliga a considerar valores $\mathbf{0.6 \leq IF \leq 1.0}$.

Por la misma razón será suficiente considerar el valor de **PFI=3** para realizar la ponderación; los mínimos que se obtienen (3ª iteración) son similares e incluso inferiores a los que resultan de **PFI=4**. Además al considerar una iteración menos estamos introduciendo menor distorsión. Sin embargo, dentro del margen de baja distorsión para el parámetro **IF** se obtienen mejores valores al considerar la cuarta iteración de la trama precedente **(PFI=4)**.

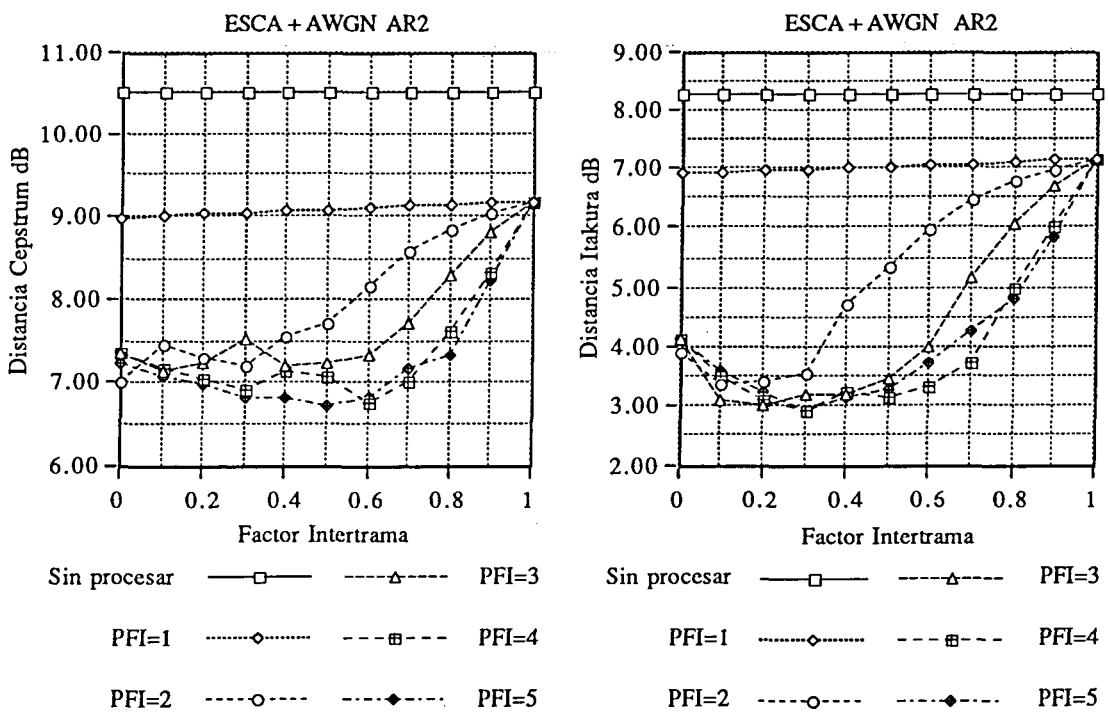


Figura V.29 : Distancias Cepstrum e Itakura tras procesar la primera iteración del algoritmo AR2_IF para un nivel de ruido SNR_G=9dB.

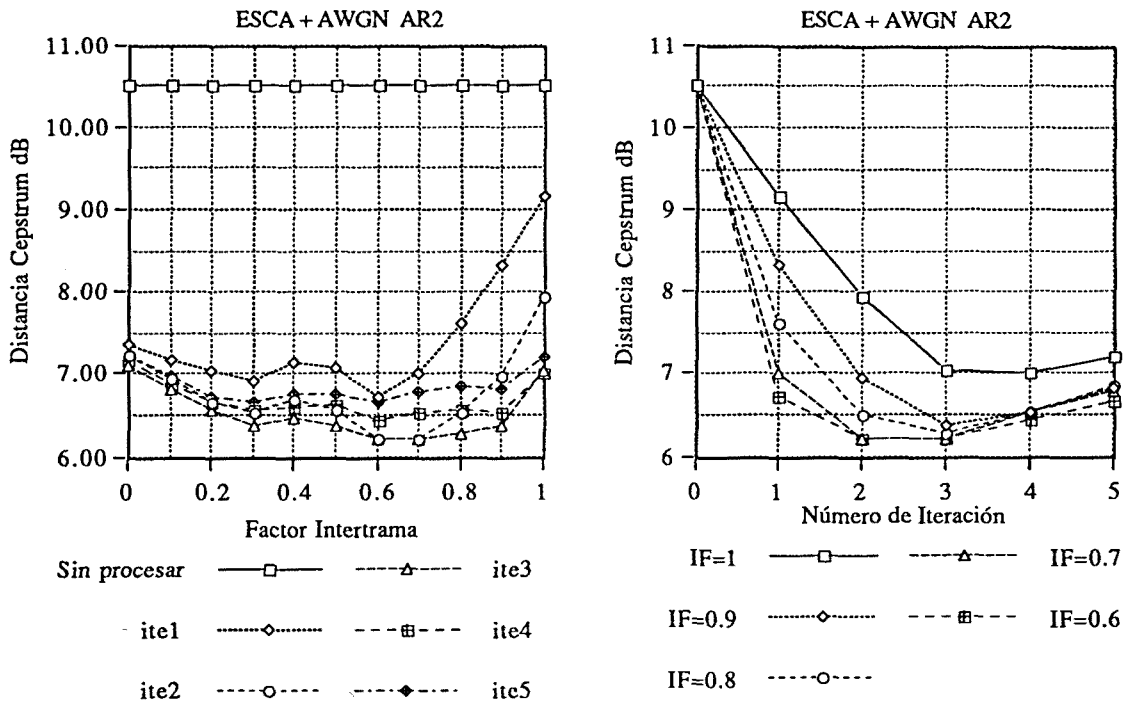


Figura V.30 : *Evolución del Algoritmo AR2_IF a SNR_G=9dB para **PFI=4.***

Vemos en la Fig.V.30 como los valores de **IF** más beneficiados están nuevamente situados entre 0.6 y 0.7 . Y como sucedía para 0dB, además de ganar una iteración respecto al caso sin ponderar (**IF=1**). En relación al algoritmo AR2, se obtienen ganancias más notables al usar esta ponderación intertrama durante la primera iteración, reduciéndose la distancia Cepstrum unos 3.5dB tras procesar únicamente la primera iteración.

9dB	SNR_{Global}	$SNR_{Seg.}$	ITAKURA	COSH	CEPSTRUM
Original	9.02	8.07	8.28	9.92	10.51
1 iter.	15.01	10.63	5.19	7.22	7.72
2 iter.	15.94	11.66	3.75	6.03	6.55
3 iter.	15.98	11.77	3.04	5.36	6.28
4 iter.	15.74	11.68	3.13	5.49	6.55
5 iter.	15.59	11.57	3.23	5.61	6.75

Tabla V.9 : *Medidas de distancia obtenidas por el algoritmo AR2_IF para **IF=0.7** y **PFI=3**.*

V.4.1.2. Algoritmo AR3 con Ponderación Intertrama (AR3_IF).

En el apartado precedente se han evaluado las prestaciones del algoritmo AR2_IF y se ha concluido que éstas resultan bastante superiores en relación al algoritmo iterativo AR2: la ponderación intertrama consigue una mejora en la calidad de la voz realzada y, asimismo, se reduce el número de iteraciones necesarias para llegar al resultado óptimo. Anteriormente, se han introducido las estadísticas de orden superior pensando en unas ventajas similares a las ofrecidas por la ponderación intertrama.

En este apartado se prevee acumular las mejoras debidas a ambas estrategias: el uso de estadísticas de tercer orden considerando ponderación intertrama. Puesto que el método AR3 básico es ya de por sí bastante más agresivo que el AR2, la utilización de dicho promediado en casos en que el ruido presente en la señal no sea suficientemente elevado dará lugar a señales distorsionadas, como consecuencia de un excesivo sobrefiltrado. Por este motivo nuestro estudio irá dirigido a las situaciones más adversas, alrededor de $\mathbf{SNR_G=0dB}$, donde otras técnicas de Speech Enhancement ofrecen unas pobres prestaciones. No obstante, analizaremos también el caso $\mathbf{SNR_G=9dB}$, señal de calidad media, y comprobaremos lo que sucede. Recuérdase que la técnica de ponderación intertrama adquiere mayor significado en aquellos supuestos donde se precisa procesar algunas iteraciones para alcanzar la iteración óptima. Por este motivo la consideración del algoritmo AR3_IF en la zona $\mathbf{SNR_G=9dB}$ puede situarnos justo en la frontera donde la ponderación intertrama pierde su razón de ser utilizada.

V.4.1.2.1. Ambientes altamente ruidosos.

Se evalúan valores de **PFI** comprendidos entre 1 y 5, y valores de **IF** entre 0 y 1, para ruido AGWN a 0dB, pero en este caso veremos lo que sucede con la frase ASUN1 además del fichero ESCA. Téngase en cuenta que estas dos señales corresponden a frases distintas, en idiomas diferentes, pronunciadas por locutores diferentes. A pesar de obtener pequeñas diferencias al evaluar locutores y contenidos distintos se mostrará como el comportamiento general es bastante similar.

Empecemos analizando el comportamiento de la distancia Cepstrum, que da una visión más general de todo el espectro. En la Fig.V.31 podemos ver lo que sucede con ESCA y ASUN1 durante la primera iteración. Para **PFI=1** sucede aproximadamente lo mismo que sucedía con AR2: ponderar con unos coeficientes $\mathbf{a_k}$ procedentes de la trama anterior, estimados sobre niveles de ruido similares a los de la trama actual, no aporta mejoras significativas y la curva resultante al barrer el parámetro **IF** es prácticamente constante.

Sin embargo, al subir en **PFI** ya aparecen las diferencias y la mejora aumenta de forma casi brusca por poco que el parámetro **IF** se reduzca a partir de la unidad: una reducción de distancia Cepstrum de unos 2dB para ESCA y unos 0.8dB para ASUN1. Al contrario de lo que sucedía con AR2, donde una mejora progresiva se apreciaba a medida que **IF** disminuía (el mínimo se situaba alrededor de **IF=0.3**), aquí la zona más beneficiada se sitúa entre **IF=0.6** e **IF=0.9**. La explicación puede encontrarse en la Fig.V.32, donde queda representada la distorsión introducida por el algoritmo cuando procesamos señal de voz original carente de ruido. Se observa como los niveles de distorsión, producto de la agresividad del método, están incluso por encima del doble de los que vimos que ocasionaba AR2. No es de extrañar, por

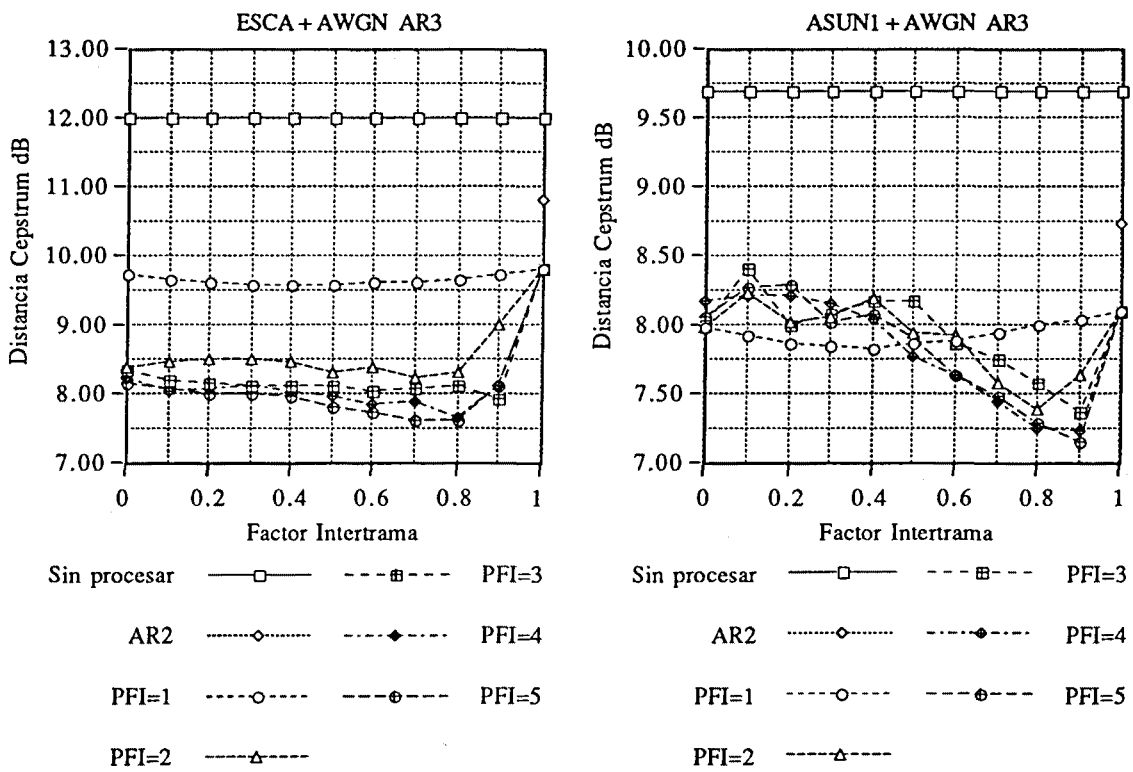


***Figura V.31** : Comportamiento del Algoritmo AR3_IF tras procesar la primera iteración para un nivel de ruido SNR_G=0dB.*

tanto, que los valores de **IF** menores a 0.7 se autodescarten como valores eficientes para el filtrado, a causa de la excesiva agresividad y consecuente distorsión ocasionada en un algoritmo ya de por sí agresivo.

Esta agresividad es también la explicación de lo que sucede con la frase ASUN1. A pesar de contener un nivel de ruido equivalente al de ESCA, por diversas razones (fisiología del locutor, distribución de la voz) parte de un nivel de distancia Cepstrum inferior (9.68dB frente a los 12.02dB del fichero ESCA); parece como si el mismo ruido lograra enmascarar menos la señal ASUN1 que la ESCA. Por este motivo, al filtrar, presenta de forma más acusada los efectos que hemos comentado en el párrafo anterior. La distancia Cepstrum, para **IF<0.6** y **PFI>1**, adopta incluso valores peores que los que se obtienen sin promediar **(IF=1)**.

Centrándonos en el intervalo **0.7≤IF≤0.9** podemos ver el comportamiento de la distorsión durante las sucesivas iteraciones. En la Fig.V.33 se ha representado la distorsión ocasionada para el peor caso **IF=0.7** según los distintos valores de **PFI** considerados. Si **PFI=1** lógicamente apenas se aprecian diferencias en relación al algoritmo AR3 **(IF=1)**. Para **PFI>1** la distorsión crece en cada iteración, hasta empezar a saturarse para **PFI=4** y **PFI=5**. Recordemos que el algoritmo AR3 básico obtiene una mejor estimación en la tercera iteración (ver Tabla IV.4 y Tabla V.10).

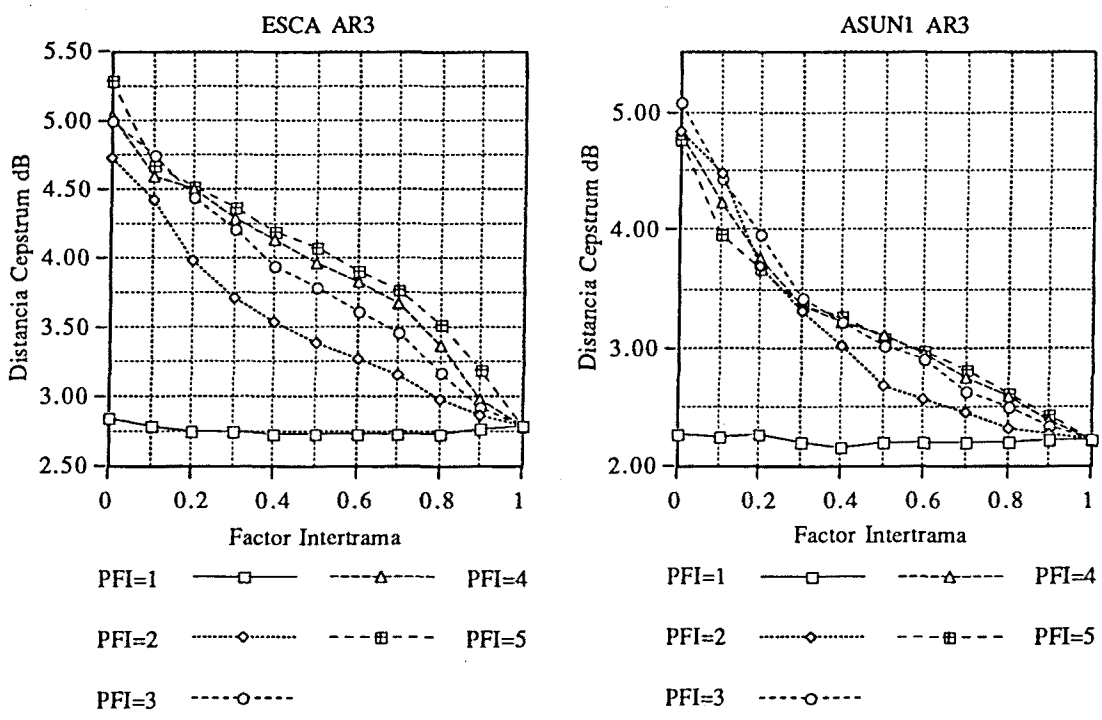


Figura V.32 : Distorsión introducida por el algoritmo AR3_IF durante la primera iteración.

.Entonces, parece lógico que funcione también mejor si ayudamos ahora a la primera iteración de cada trama con la mejor estimación de la trama anterior, es decir, la tercera (**PFI=3**). A pesar de que en la Fig.V.34 se evalua el valor **PFI=3**, un valor **PFI=2** también puede considerarse una buena elección, pues segun la Fig.V.33 este supuesto conduce a menores niveles de distorsión y para valores de **IF** adecuados conduce a mayores reducciones

0dB	SNR_{Global}	$SNR_{Seg.}$	ITAKURA	COSH	CEPSTRUM
Original	0.060	0.894	7.372	10.048	9.676
1 iter.	7.618	4.303	6.432	8.667	8.105
2 iter.	7.133	4.901	4.992	7.727	8.030
3 iter.	7.183	5.108	3.771	6.985	7.847
4 iter.	7.255	4.972	3.693	7.152	7.879
5 iter.	7.401	5.052	3.671	7.433	8.108

***Tabla V.10** : Comportamiento del Algoritmo AR3 básico para el fichero ASUN1 degradado mediante ruido Gaussiano a SNR_G=0dB.*

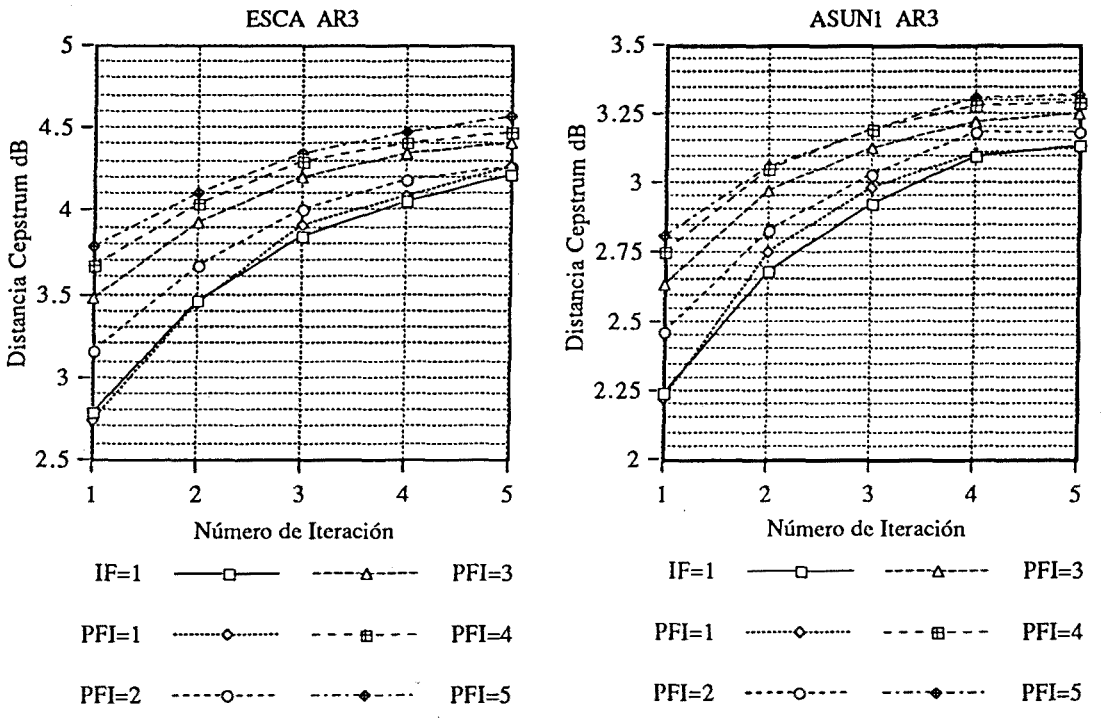


***Figura V.33:** Evolución de la distorsión introducida por el algoritmo AR3_IF a lo largo de las sucesivas iteraciones para una ponderación IF=0.7* .

de ruido para el Fichero ASUN1 (ver Fig.V.31). Nótese como en la Fig.V.33 las diferencias de distorsión, al variar los valores de **PFI**, son mucho más significativos en la primera iteración y, al incrementar la iteración considerada, luego tienden a agruparse todos los valores, puesto que la distorsión debida a la ponderación intertrama se va enmascarando con la distorsión propia del algoritmo iterativo de Wiener.

En la Fig.V.34 se presentan los resultados obtenidos por el algoritmo AR3_IF cuando se toma **PFI=3**. Tal como sucedía para el algoritmo AR2_IF se consigue el ahorro de una iteración en el proceso de filtrado, aunque entonces pasábamos de 4 a 3 y ahora serán suficientes 2 iteraciones para alcanzar una estimación todavía mejor ($C_3(1, 1.2)$=7.9dB) en relación a la suministrada por la tercera iteración del algoritmo AR3 (8.15dB en distancia Cepstrum). Observando los gráficos podríamos pensar en quedarnos directamente con la primera iteración, vistos los buenos resultados obtenidos en distancia Cepstrum, pero las pruebas de audición muestran como todavía no se han eliminado todos los espurios de la señal y, además, el resto de medidas espectrales continúan progresando.

El comportamiento de la distancia de Itakura al procesar la primera iteración de filtrado se ha representado en la Fig.V.35. La distorsión de los valles espectrales afecta poco a esta

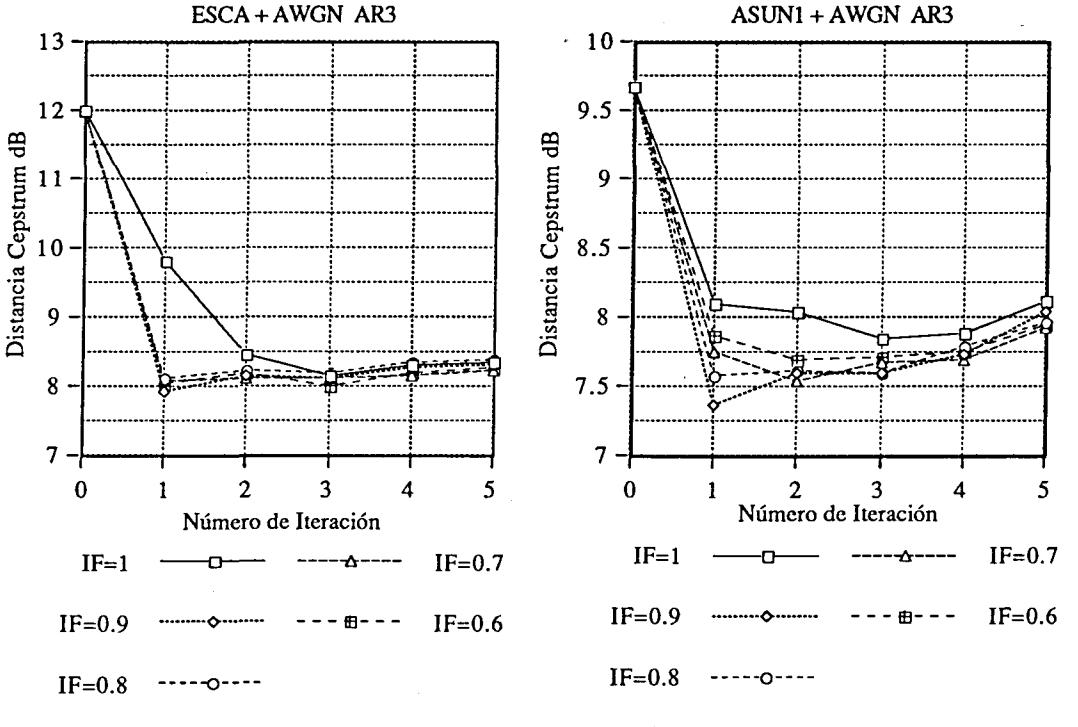


Figura V.34 : Evolución del algoritmo AR3_IF al procesar las 5 primeras iteraciones para PFI=3 y a un nivel de ruido dado por SNR_G=0dB.

medida, por lo que no es raro que sus mínimos se sitúen en la zona más baja (y agresiva) de **IF**. Por idéntico motivo tampoco se ve afectada por una subida de **PFI**, debido a su mayor insensibilidad al efecto de picado espectral discutido anteriormente.

Para **PFI=2** la mejora es muy notoria, 3.5dB y 2.5dB respecto de **IF=1** para ESCA y ASUN1 respectivamente, y para **PFI=3** baja hasta la saturación, de modo que **PFI=4** y **PFI=5** mejoran apenas unas décimas. Se constata, entonces, la gran ayuda que supone el promediado intertrama para un buen y rápido posicionamiento de los formantes de la voz.

En el método básico AR3 la primera iteración del algoritmo se limita a extraer ruido, para posicionar y dar forma a los formantes en una segunda o tercera iteración. Mediante el factor intertrama logramos ahora, en una sola iteración, extraer el ruido y posicionar con bastante precisión a los formantes del espectro, de modo que en una segunda vuelta el algoritmo podrá centrarse en afinar en la estimación y eliminar los espurios remanentes.

En la Fig.V.36 puede verse cómo en la primera iteración logramos casi la máxima mejora en distancia Itakura. Reducimos 3.8dB de golpe en la frase ASUN1 y hasta 4.7dB en

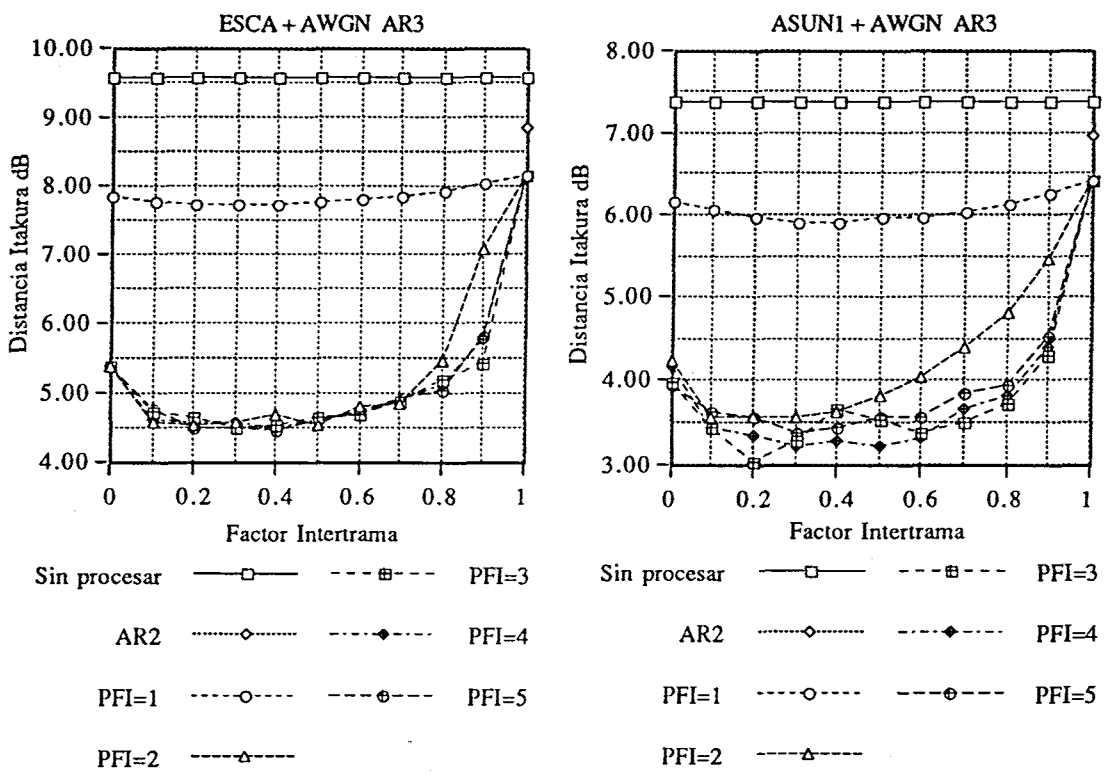


Figura V.35 : Evolución del algoritmo AR3_IF segun la distancia Itakura tras procesar la primera iteración para SNR_G=0dB.

la ESCA. La segunda iteración, más que empeorar, mantiene ese resultado; la mejora se produce sobretodo en los valles espectrales. Se aprecia un comportamiento muy similar para los valores **PFI=2** y **PFI=3**.

La distancia Cosh, como hemos visto otras veces, presenta un comportamiento muy parecido a la Itakura. Respecto a las medidas temporales hay que decir que presentan una evolución paralela a la resultante en el supuesto de no ponderar **(IF=1)**, pero con unas décimas de mejora.

A partir de los tests de audición realizados para este nivel de ruido, SNR_G=0dB, se desprenden las siguientes apreciaciones:

- Para **PFI=1** no se observa en absoluto ninguna mejora respecto del caso sin ponderar **(IF=1)**. Aunque para las zonas de baja energía (sobretodo los fonemas sordos) sí existe una degradación a medida que **IF** baja, para los fonemas de mayor energía no parece notarse ninguna diferencia en relación a **IF=1**, incluso para el caso extremo **IF=0** (que equivale a utilizar los $\mathbf{a_k}$ de la trama solapada anterior). Se pone aquí de manifiesto la estacionariedad de la voz entre cada dos tramas consecutivas, tomada como hipótesis de partida al inicio de la ponderación intertrama.

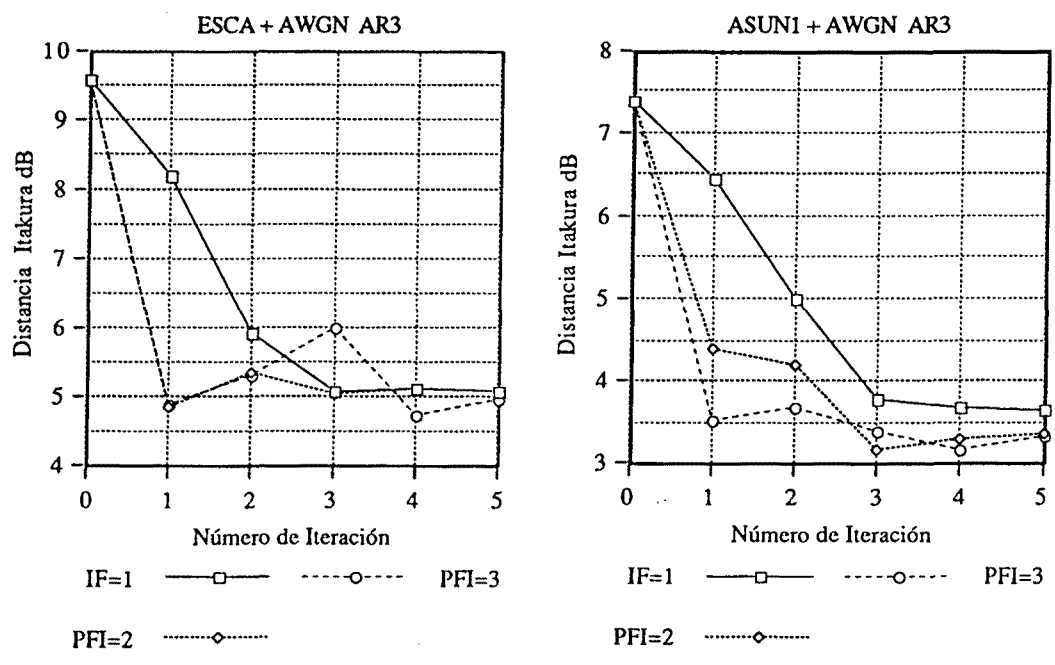


Figura V.36 : Evolución del algoritmo AR3_IF segun la distancia Itakura durante las 5 primeras iteraciones para IF=0.7 y SNR_G=0dB.

- A partir de **PFI=2** en adelante la distorsión sí es determinante para valores **IF** pequeños. Se observa que si bajamos de 0.6, como concluíamos de las medidas objetivas, la inteligibilidad se ve seriamente perjudicada. En cambio, para el margen útil **(0.6≤IF≤1.0)** el comportamiento es distinto. En la primera iteración se elimina gran parte del ruido de fondo, pero aparecen espurios musicales. En una segunda pasada el ruido que quedaba ha desaparecido y el nivel de los espurios es menor. A la salida del tercer filtrado éstos son prácticamente eliminados, y aunque empieza a notarse un dominio de los graves sobre los agudos, producto de la distorsión y el picado espectral del método, podemos asegurar que la señal es de mejor calidad que la del algoritmo AR3 sin promediado intertrama.

- Con **PFI=3** obtenemos un resultado óptimo. El comportamiento es muy similar a **PFI=2**, pero siempre con un grado de inteligibilidad ligeramente superior, la señal es más nítida, más clara.

- Para **PFI=4** y **PFI=5** la evolución de la señal filtrada no difiere en exceso de los casos anteriores, aunque siempre viéndose afectada por el hecho de estar utilizando unos coeficientes a_k correspondientes a una estimación con mayor contenido de distorsión. Como además estos casos suponen mayor tiempo de cálculo (1 ó 2 iteraciones más), serán ambos descartados.

En conclusión, utilizaremos los valores **PFI=3** o **PFI=2** en función de que haya más o menos ruido, respectivamente, en la señal a procesar, o lo que es lo mismo, seleccionaremos el número de **PFI** correspondiente al número de iteración donde se obtenga un mínimo en distancia Cepstrum cuando se utiliza el método AR3 básico **(IF=1)**. El valor de **IF** se situará entre 0.6 y 0.9, también en función de la agresividad o velocidad que queramos dar al algoritmo de filtrado (0.6 más agresivo, 0.9 menos agresivo).

V.4.1.2.2. Niveles intermedios de Ruido.

El algoritmo AR3 ante un nivel intermedio de ruido (SNR_G=9dB) alcanza una gran reducción de ruido durante la primera iteración y, en las iteraciones siguientes, la mejora obtenida en la mayor parte de medidas de distancia es bastante insignificante. En estas condiciones no se puede esperar una gran mejora mediante la aplicación de la ponderación intertrama. Los resultados correspondientes al algoritmo AR3_IF confirman plenamente estas previsiones.

En la Fig.V.37 se ha representado la distancia Cepstrum tras procesar la primera iteración con el algoritmo AR3_IF. Las mejoras obtenidas son muy poco importantes, apenas unos 0.5dB en distancia Cepstrum. En estas condiciones parece suficiente la consideración de la segunda iteración de la trama precedente (**PFI=2**) y un Factor Intertrama de valor entre 0.8 y 0.9 .

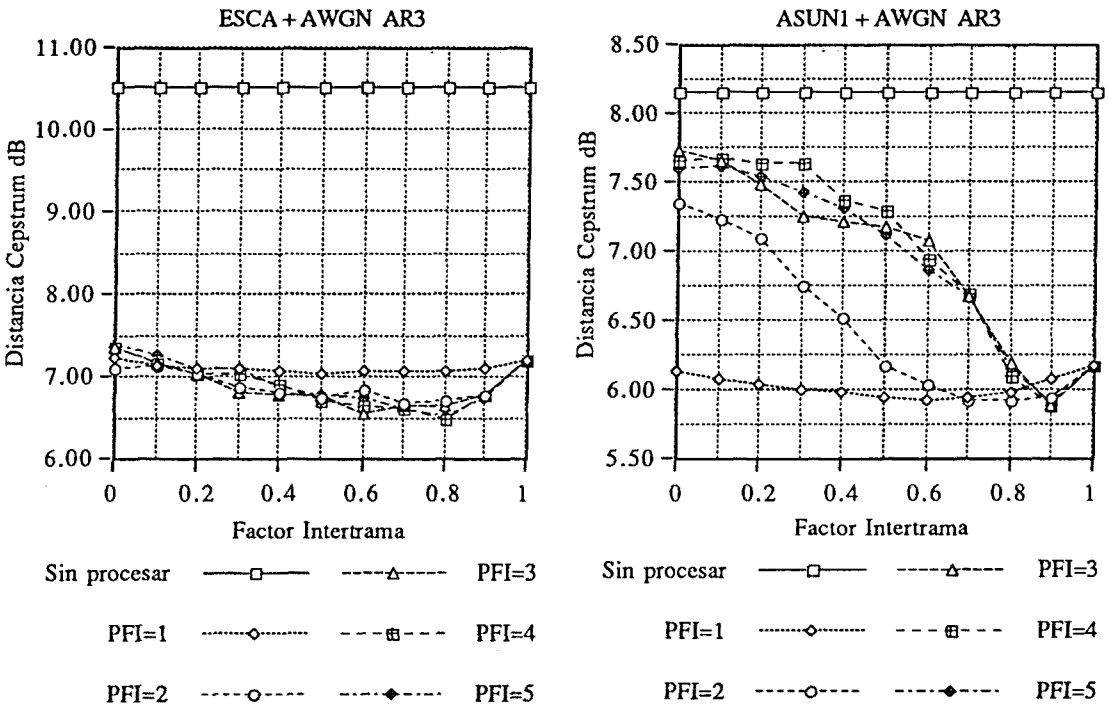


Figura V.37 : Evolución del Algoritmo AR3_IF tras procesar la primera iteración para SNR_G=9dB.

En la Fig.V.38 se ha representado la distancia Itakura para las mismas condiciones anteriores. Para los valores de distancia Itakura se pueden obtener mejoras superiores a 1dB, e incluso el valor **PFI=3** se muestra como claramente superior. Evidentemente, este pequeño engaño se debe a su menor dedicación hacia las zonas de los valles espectrales donde realmente se produce la distorsión.

De los comentarios anteriores parece lógico considerar el valor **PFI=2** como única posibilidad cuando el algoritmo AR3_IF se enfrenta a estos niveles intermedios de ruido. En la Fig.V.39 se ha representado la evolución de la reducción de ruido durante las 5 primeras iteraciones cuando en la primera iteración se ha considerado un promediado intertrama mediante la estimación AR de la segunda iteración de la trama precedente **(PFI=2)**. Al considerar distintos valores para el parámetro **IF** se obtienen prestaciones muy similares, aunque parecen ligeramente mejores cuando se considera un factor **IF=0.7**. A pesar de que los distintos valores del factor **IF** conducen a mejores resultados que los correspondientes al algoritmo AR3 **(IF=1)**, esta mejora es bastante pequeña en relación a la obtenida para niveles de ruido superiores, donde la ponderación intertrama de los coeficientes $\mathbf{a_k}$ se presenta como una estrategia mucho más atractiva. Así, para este nivel de ruido, se puede afirmar que estamos justo en el límite de utilidad del promediado intertrama cuando se utilizan las

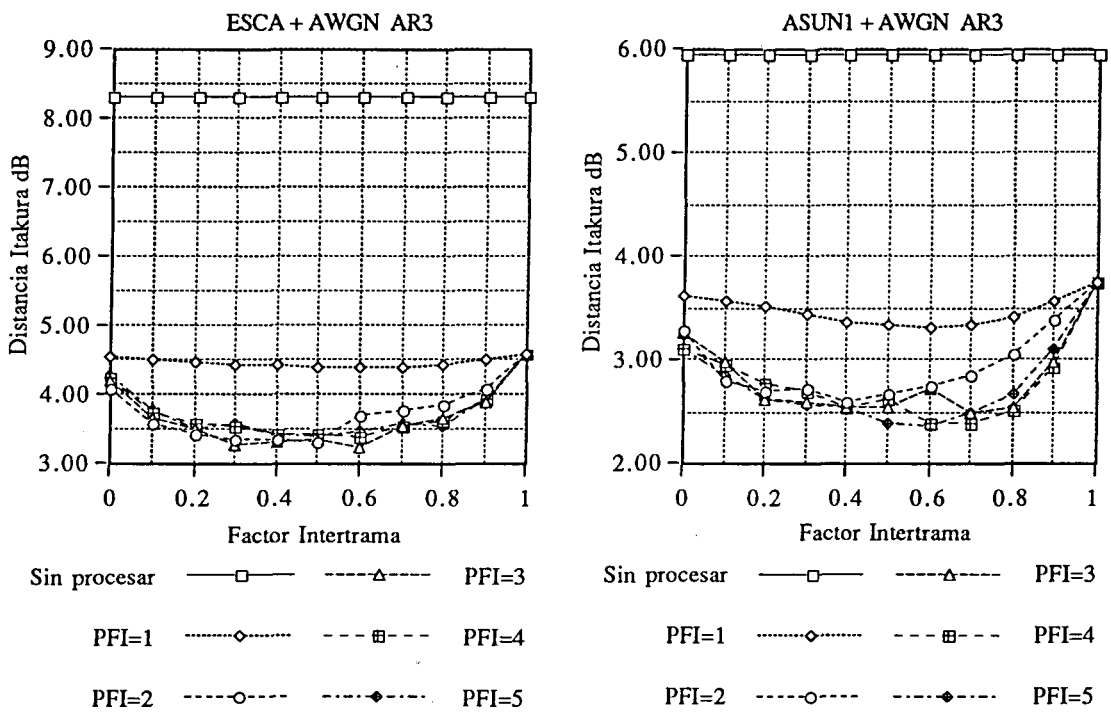


Figura V.38 : Evolución del Algoritmo AR3_IF segun la distancia Itakura tras procesar la primera iteración para un nivel de ruido SNR_G=9dB.

estadísticas de tercer orden. De hecho en el Apartado IV.3 se vió como el algoritmo AR3 es capaz de afrontar por si mismo estos niveles intermedios de ruido, a diferencia de los algoritmos AR2 y AR4.

Como conclusión, la estrategia de Ponderación Intertrama **IF** sirve de gran ayuda en ambientes muy ruidosos. Ante niveles de ruido intermedios o bajos su necesidad decrece a medida que un determinado algoritmo se muestra capaz de combatir el ruido presente en la señal de voz ruidosa **x(n)**. En presencia de poco ruido, se impone el uso de la Ponderación Intertrama sólamente para aquellos algoritmos cuya velocidad de convergencia sea baja, como por ejemplo el caso del algoritmo AR2. Cuando se dota a un determinado algoritmo de la posibilidad de una ponderación intertrama, la velocidad de convergencia se acelera llegándose a una iteración óptima menor y, en consecuencia, se alcanzan valores de distancia espectral mejores debido, en parte, a la menor distorsión acumulada. Al mismo tiempo también se disminuye la complejidad de cálculo del algoritmo iterativo.

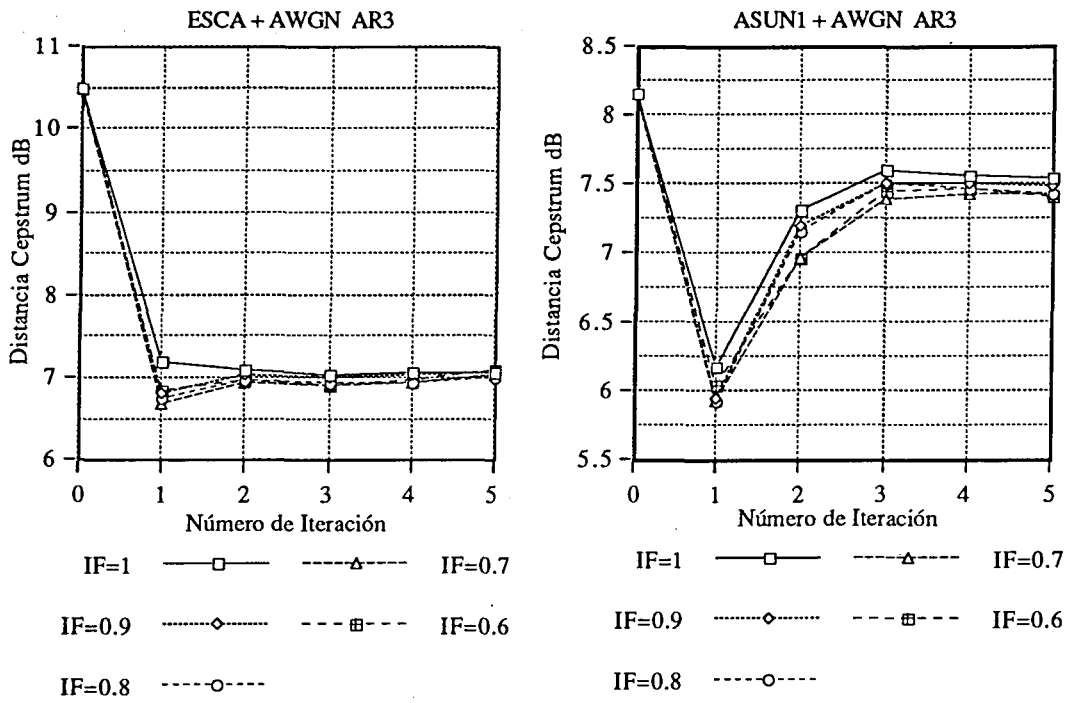


***Figura V.39** : Evolución del Algoritmo AR3_IF durante las 5 primeras iteraciones para **PFI**=2 y considerando un nivel de ruido SNR_G=9dB.*

CAPITULO VI

El Método de Preprocesado por Autocorrelaciones.

En este capítulo se propone una técnica basada en el cálculo de la parte causal de la función autocorrelación de la señal de voz como preprocesado a la estimación AR mediante las estadísticas de segundo orden. Este método consiste en aplicar el algoritmo AR2 sobre la parte causal de la función autocorrelación en lugar de aplicarlo directamente sobre la señal de voz ruidosa **x(n)**. Tal como se demuestra posteriormente la función autocorrelación causal de la señal de voz presenta los mismos polos que la propia señal de voz y, además, presenta la ventaja de ser menos sensible al ruido. Bajo estas hipótesis parece lógico esperar una mayor robustez en la estimación AR ejecutada en el dominio de la función autocorrelación causal. Esta técnica ha sido denotada por algunos autores [Nade-94], [Hern-92] como OSALPC (One-Sided Autocorrelation Linear Predictive Coding) y ha sido aplicada al reconocimiento robusto del habla y se ha visto su estrecha relación con las ecuaciones de Yule-Walker y las ecuaciones de Yule-Walker Sobredeterminadas [Hern-93]. De esta manera vamos a denotar el algoritmo resultante como OSA_AR2. Los resultados obtenidos mediante esta estrategia han sido publicados en [Sala-94d] y [Sala-94e].

VI.1. Predicción Lineal de la Parte Causal de la Autocorrelación.

En el Apartado VI.1.1. se presentan las propiedades de la parte causal de la secuencia de autocorrelación, el espectro analítico y la envolvente espectral. Se muestra como existe una correspondencia biunívoca entre el espectro y su envolvente y, en consecuencia, la estimación de la envolvente del espectro se corresponde con una única estimación del espectro y no representa ninguna pérdida de información. Seguidamente en los Apartados VI.1.2. y VI.1.3. se demuestra que, al considerar un Modelado Inverso de la función Autocorrelación Causal (MIAC), los polos de la Transformada Z de la parte causal de la autocorrelación de x(n) son los mismos que presenta la Transformada Z de la propia señal de voz x(n). Esta técnica MIAC representa una técnica bastante simple para estimar la envolvente espectral. Las prestaciones de esta técnica se pueden mejorar al considerar un sistema de ecuaciones sobredeterminado, dando lugar a la técnica denominada OSALPC. Debe remarcarse que la técnica MIAC se ha considerado en el presente trabajo únicamente como paso previo para el desarrollo teórico de la técnica OSALPC.

VI.1.1. La Parte Causal de la Función Autocorrelación.

Sea **r(m)** la secuencia de autocorrelación de una señal real **x(n)**, entonces se define su parte causal $\mathbf{r^+(m)}$ como:

$$r^+(m) = \begin{cases} r(m) & \text{si } m > 0 \\ \dfrac{r(0)}{2} & \text{si } m = 0 \\ 0 & \text{si } m < 0 \end{cases} \qquad \text{(VI.1)}$$

y se verifica:

$$r^+(m) + r^+(-m) = r(m) \quad ,, \quad -\infty \leq m \leq \infty \qquad \text{(VI.2)}$$

La transformada Z y la transformada de Fourier de $\mathbf{r^+(m)}$, introducidas en análisis espectral por Cadzow [Cadz-80], se denotan respectivamente como $\mathbf{R^+(z)}$ y $\mathbf{S^+(w)}$, verificando la siguiente relación:

$$S^+(w) = R^+(z)\,|_{z=e^{jw}} = R^+(e^{jw}) \qquad \text{(VI.3)}$$

mientras que las Transformadas Z y de Fourier de la función autocorrelación $\mathbf{r(m)}$ se denotan como $\mathbf{R(z)}$ y $\mathbf{S(w)}$ respectivamente. Análogamente se satisface:

$$S(w) = R(z)\,|_{z=e^{jw}} = R(e^{jw}) \qquad \text{(VI.4)}$$

Como $\mathbf{r^+(m)}$ es una secuencia real y causal y $\mathbf{r(m)}$ es dos veces la parte par de $\mathbf{r^+(m)}$, se satisface la siguiente relación entre $\mathbf{S^+(w)}$ y $\mathbf{S(w)}$ [Oppe-75]:

$$S^+(w) = \frac{1}{2} \cdot \left[\, S(w) + j.S_H(w) \,\right] \qquad \text{(VI.5)}$$

donde $\mathbf{S_H(w)}$ representa la transformada de Hilbert de $\mathbf{S(w)}$ y responde a la siguiente expresión:

$$S_H(w) = \frac{1}{2\pi} \cdot \lim_{\varepsilon \to 0} \left[\int_{w+\varepsilon}^{\pi} S(\phi) \cdot \cot\frac{\phi - w}{2} \cdot d\phi \; + \int_{-\pi}^{w+\varepsilon} S(\phi) \cdot \cot\frac{\phi - w}{2} \cdot d\phi \right] \qquad \text{(VI.6)}$$

debido a la analogía entre la expresión (VI.5) y la definición de la señal analítica utilizada en modulación de amplitud, se denomina Espectro Analítico a $\mathbf{S^+(w)}$ y Envolvente espectral a su módulo:

$$E(w) = |\, S^+(w) \,| \qquad \text{(VI.7)}$$

De esta forma se observa una correspondencia biunívoca entre la envolvente espectral y el espectro. En consecuencia, la envolvente espectral no representa ninguna pérdida de información en relación al espectro. Dado un espectro $\mathbf{S(w)}$, entonces, su envolvente asociada $\mathbf{E(w)}$ viene dada por las ecuaciones (VI.5), (VI.6) y (VI.7). Por otro lado, dada una envolvente espectral $\mathbf{E(w)}$, el espectro asociado $\mathbf{S(w)}$ viene definido unívocamente por la expresión:

$$S(w) = 2 \,.\, \mathrm{Re}\,[S^+(w)] = 2 \,.\, E(w) \,.\, \cos \phi(w) \qquad \text{(VI.8)}$$

donde $\mathbf{\phi(w)}$ es la curva de fase mínima asociada al módulo $\mathbf{E(w)}$. Ello es debido a que $\mathbf{R^+(z)}$ no tiene ceros ni polos fuera de la circunferencia de radio unidad, como se demuestra seguidamente.

Si el espectro S(w) está acotado, la parte causal de la secuencia autocorrelación $r^+(m)$ es una secuencia estable y, en consecuencia, $R^+(z)$ presenta todos sus polos, en el supuesto de tenerlos, en el interior de la circunferencia de radio unidad. Además, si el espectro no se anula para cualquier frecuencia se puede demostrar fácilmente que todos los ceros de $R^+(z)$, caso de tenerlos, estan también localizados en el interior de la circunferencia de radio unidad. Como r(m) es dos veces la parte par de $r^+(m)$, luego se verifica:

$$\mathrm{Re}\,[\,R^+(z)\,] = R(z)/2 \qquad \text{(VI.9)}$$

y al ser el espectro S(w) positivo se satisface:

$$S(w) = R(z)\,|_{z=e^{jw}} = 2\,\mathrm{Re}\,[\,R^+(z)\,|_{z=e^{jw}}\,] > 0 \qquad \text{(VI.10)}$$

para todo w. Consecuentemente no existen ceros de $R^+(z)$ en la circunferencia de radio unidad. Para demostrar que no existen ceros fuera de dicha circunferencia será suficiente demostrar:

$$\mathrm{Re}\,[\,R^+(z)\,|_{z=r.e^{jw}}\,] = \mathrm{Re}\left[\,\sum_{m=0}^{\infty} r^+(m).\rho^{-m}.e^{-jwm}\right] > 0 \qquad \text{(VI.11)}$$

para todo w y $1 < r < \infty$.

Para verificar (VI.11) se construye la secuencia:

$$r'(m) = r(m)\,.\,a^{|m|} \qquad \text{(VI.12)}$$

con $0<a<1$. La transformada de Fourier de $r'(m)$ es positiva para cualquier frecuencia ya que se corresponde, salvo un factor de escalado, a la convolución de S(w) con la transformada de Fourier de $a^{|m|}$. Esta última transformada es positiva por ser $a^{|m|}$ la autocorrelación de un proceso AR paso bajo de orden uno. Como la parte causal de $r'(m)$ es $r^+(m).a^{|m|}$, a partir de (VI.9) resulta:

$$\mathrm{Re}\left[\,\sum_{m=0}^{\infty} r^+(m).a^m.e^{-jwm}\right] = \frac{1}{2}\sum_{m=-\infty}^{\infty} r'(m).e^{-jwm} > 0 \qquad \text{(VI.13)}$$

quedando demostrada la expresión (VI.11) para todo w y $1<r<\infty$, considerando $r=a^{-1}$. En el infinito, el teorema del valor inicial nos garantiza la inexistencia de ceros, pues como $r^+(m)=0$ para $m<0$, entonces se cumple:

$$\lim_{z\to\infty} R^+(z) = r^+(0) = \frac{r(0)}{2} > 0 \qquad \text{(VI.14)}$$

De esta manera, si el espectro está acotado y toma valores no nulos para cualquier frecuencia, entonces $R^+(z)$ tiene sus polos y ceros en el interior de la circunferencia unidad, es decir, $r^+(m)$ es una secuencia de fase mínima. Si el espectro está acotado pero para algunas frecuencias se anula, entonces $R^+(z)$ tiene sus polos y ceros en el interior de la circunferencia unidad a menos que exista simetría par del espectro respecto a alguno de sus ceros. Ello se debe a que la función $\cot(\ (\phi+w)/2\ )$, que aparece en la relación de Hilbert (VI.6), es una función par en ϕ. Así, para que se cumpla $S_H(w_0)=0$, siendo $S(w_0)=0$, debe verificarse que $S(\phi+w_0)$ sea una función par en ϕ, es decir, $S(w)$ ha de presentar simetría par con respecto a w_0. La secuencia $r^+(m)$ es una secuencia de fase mínima en la mayoría de los casos, a menos que el espectro presente un cero en $w=0$ y esta situación suele ser bastante inusual. Si el espectro presenta simetría par respecto a algunos de sus ceros, entonces $R^+(z)$ tiene todos sus polos y ceros en el interior de la circunferencia unidad, a excepción de los ceros que verifiquen tal condición, cuya localización se situa encima de dicha circunferencia.

En cualquier caso, $R^+(z)$ nunca presenta polos ni ceros en el exterior de la circunferencia de radio unidad y consecuentemente existe una relación biunívoca entre el espectro y la envolvente espectral. Considerando que la envolvente no representa ninguna pérdida de información con respecto al espectro, entonces aparece como una buena candidata en vistas a ser utilizada para la estimación espectral [Amen-88], [Nade-89].

A partir de la expresión (VI.8) se deduce que la envolvente espectral $E(w)$ es más suave que su espectro asociado $S(w)$, ya que el término $\cos(\ \phi(w)\ )$ introduce variaciones en $S(w)$, no existentes en $E(w)$. Este carácter de envolvente, juntamente con el alto rango dinámico del espectro de voz, origina que $E(w)$ enfatice las bandas de frecuencia de mayor potencia, las cuáles son asimismo las más robustas a un ruido de banda ancha. En consecuencia, $E(w)$ es más robusto a este tipo de ruido que $S(w)$.

Teniendo en cuenta que el cuadrado de la envolvente espectral $E^2(w)$ es precisamente el espectro de $r^+(m)$, según (VI.7), el párrafo anterior equivale a afirmar que el espectro de $r^+(m)$, es decir $E^2(w)$, presenta una mayor robustez frente al ruido de banda ancha que el espectro de la propia señal $x(n)$, $S(w)$.

Por otra parte, al considerar un modelado AR para la señal de voz, los polos de la transformada Z de la parte causal de su función autocorrelación $R^+(z)$, son los mismos que presenta la transformada Z de la propia señal de voz $X(z)$, tal como se muestra en el apartado VI.1.2.

Ambos factores sugieren que los parámetros AR de la señal de voz pueden ser estimados de forma más fiable aplicando las técnicas de predicción lineal clásicas, basadas en

estadísticas de segundo orden, sobre $\mathbf{r^+(m)}$, en lugar de sobre la propia señal $\mathbf{x(n)}$, cuando la señal de voz está contaminada con ruido de banda ancha. Esta es la base de la técnica OSA_AR2 presentada en el Apartado VI.2. Nótese, para finalizar, que esta envolvente espectral se corresponde estadísticamente con una función de cuarto orden.

VI.1.2. El Modelado Inverso de la Autocorrelación Causal (MIAC).

A continuación se describe el método MIAC como una posibilidad para estimar los coeficientes AR de la señal de voz, de una manera muy simple y eficiente, realizando un modelado todo-polos de $\mathbf{R^+(z)}$, es decir, de $\mathbf{E^2(z)}$. A pesar de que este método no destaca de forma especial durante el tratamiento robusto de la voz, permite introducir de una manera simple la técnica OSALPC, cuya aplicación al tratamiento de la voz ruidosa es muy interesante debido, básicamente, a su simplicidad, su bajo coste computacional y las altas reducciones de ruido que alcanza.

Sea $\mathbf{x(n)}$ un proceso real autorregresivo de orden $\mathbf{p}$, cuyo espectro viene dado por la expresión:

$$S(w) = \frac{g^2}{\left| A(e^{jw}) \right|^2} \qquad \text{(VI.15)}$$

donde

$$A(z) = 1 + \sum_{k=1}^{p} a_k \cdot z^{-k} \qquad \text{(VI.16)}$$

entonces la transformada Z de su función autocorrelación $\mathbf{R(z)}$ se puede expresar como:

$$S(w) = \frac{g^2}{A(z) \cdot A(z^{-1})} \qquad \text{(VI.17)}$$

Como $\mathbf{A(z)}$ es el denominador de la función de transferencia del filtro del modelo $\mathbf{H(z)}$ para la voz, que se supone causal y estable, los ceros de $\mathbf{A(z)}$ están siempre localizados en el interior de la circunferencia de radio unidad. Por otro lado, resulta fácil comprobar como $\mathbf{R(z)}$ puede escribirse en función de la transformada Z de la parte causal de la secuencia autocorrelación $\mathbf{R^+(z)}$ según:

$$R(z) = R^{+}(z) + R^{+}(z^{-1}) \qquad \text{(VI.18)}$$

De esta manera, la expresión (VI.17) puede escribirse de la forma siguiente:

$$R(z) = \frac{C(z)}{A(z)} + \frac{C(z^{-1})}{A(z^{-1})} \qquad \text{(VI.19)}$$

donde el primer término se corresponde con $\mathbf{R^{+}(z)}$, por el hecho de tener los polos en el interior de la circunferencia unidad, según se ha visto en el apartado anterior:

$$R^{+}(z) = \frac{C(z)}{A(z)} \qquad \text{(VI.20)}$$

llegándose de esta forma a la conclusión deseada: la transformada Z de la parte causal de la autocorrelación $\mathbf{R^{+}(z)}$ tiene los mismos polos que el filtro $\mathbf{H(z)}$ y consecuentemente los mismos polos que la señal $\mathbf{x(n)}$ [Ginn-83a], [Ginn-83b].

En lo referente a los ceros de $\mathbf{R^{+}(z)}$, $\mathbf{C(z)}$ debe ser un polinomio en $\mathbf{z^{-1}}$ por ser $\mathbf{r^{+}(m)}$ una secuencia causal. Por otro lado, a partir de (VI.16) y (VI.20) se obtiene que $\mathbf{C(\infty)}$ es igual a $\mathbf{R^{+}(\infty)}$, y al aplicar el teorema del valor inicial coincide con $\mathbf{r^{+}(0)}$. Así resulta que el término independiente de $\mathbf{C(z)}$ es $\mathbf{r^{+}(0)}$ y, en consecuencia, toma un valor no nulo. Tomando en consideración que este término independiente de $\mathbf{C(z)}$ no en distinto de cero se obtiene:

$$g^2 = C(z) \cdot A(z^{-1}) + C(z^{-1}) \cdot A(z) \qquad \text{(VI.21)}$$

al mezclar las expresiones (VI.17) y (VI.19), de donde se deduce fácilmente que $\mathbf{C(z)}$ es un polinomio del mismo orden que $\mathbf{A(z)}$.

Además, como en la práctica el espectro (VI.15) suele ser distinto de cero para cualquier frecuencia, $\mathbf{r^{+}(m)}$ es una secuencia de fase mínima y, en consecuencia, los $\mathbf{p}$ ceros de $\mathbf{A(z)}$, coincidentes con los polos de la señal de voz, y los $\mathbf{p}$ ceros de C(z) están siempre en el interior de la circunferencia de radio unidad.

Finalmente, podemos expresar $\mathbf{R^{+}(z)}$ de la manera siguiente:

$$R^{+}(z) = \frac{C(z)}{A(z)} = \frac{c_0 + \sum_{k=1}^{p} c_k . z^{-k}}{1 + \sum_{k=1}^{p} a_k . z^{-k}} \qquad \text{(VI.22)}$$

y las expresiones para el espectro analítico y el cuadrado de la envolvente espectral se pueden expresar como:

$$S^{+}(w) = \frac{C(e^{jw})}{A(e^{jw})} \qquad \text{(VI.23)}$$

y

$$E^{2}(w) = \frac{\left|C(e^{jw})\right|^{2}}{\left|A(e^{jw})\right|^{2}} \qquad \text{(VI.24)}$$

respectivamente. A pesar de que aparezcan **2p+1** parámetros en la expresión (VI.22), sólo **p+1** de éllos son independientes, ya que **R(z)** queda especificada por g^2 y los **p** coeficientes de **A(z)** y ésta tiene una correspondencia biunívoca con $R^{+}(z)$. Esta relación de dependencia viene dada por (VI.21) y entonces puede calcularse **C(z)** a partir de g^2 y **A(z)**.

En el dominio temporal, la expresión (VI.21) adopta la forma:

$$r^{+}(m) = -\sum_{k=1}^{p} a_k \cdot r^{+}(m-k) + \sum_{k=0}^{p} c_k \cdot \delta(m-k) \qquad \text{(VI.25)}$$

donde **d(m)** representa la función impulso unidad. Esta expresión se transforma en una identidad con ambos términos nulos para **m<0** y, en consecuencia, sólamente se considera para **m≥0**.

En este capítulo se propone modelar $R^{+}(z)$ como una función todo-polos, aprovechando el hecho de que un modelo todo-polos permite aproximar cualquier modelo racional, utilizando un número de polos suficientemente elevado. Esto equivale a incrementar el valor de **p** y suponer que todos los ceros de **C(z)** están situados en el origen:

$$C(z) = c_0 = r^{+}(0) \qquad \text{(VI.26)}$$

En este supuesto las expresiones (VI.22) y (VI.25) se transforman en las siguientes:

$$R^{+}(z) = \frac{C(z)}{A(z)} = \frac{r^{+}(0)}{1 + \sum_{k=1}^{p} a_k . z^{-k}} \qquad \text{(VI.27)}$$

$$r^{+}(m) = -\sum_{k=1}^{p} a_k \cdot r^{+}(m-k) + r^{+}(0) \cdot \delta(m) \qquad \text{(VI.28)}$$

Esta expresión (VI.28) se convierte en una identidad para **m<0**, puesto que se anulan todos sus términos, y para **m=0** los dos términos de la igualdad valen $r^{+}(0)$. Por esta razón, sólo es necesario considerar esta expresión (VI.28) para valores **m>0**.

Una posible forma de estimar los coeficientes a_k a partir de la señal de voz consiste en evaluar (VI.28) para $1 \le m \le p$ y resolver el sistema de ecuaciones resultante, utilizando un estimador adecuado para obtener los valores de $r(m)$ y, luego, aplicar la relación (VI.1) entre $r^+(m)$ y $r(m)$. Este método de estimación de los coeficientes a_k se denota como MIAC, Modelado Inverso de la Autocorrelación Causal, debido al modelo $R^+(z)$ en el que se basa. De este modo, se origina un sistema de ecuaciones cuya matriz es triangular:

$$\begin{pmatrix} r(0)/2 & 0 & \cdots & 0 \\ r(1) & r(0)/2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ r(p-1) & r(p-2) & \cdots & r(0)/2 \end{pmatrix} \cdot \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} = - \begin{pmatrix} r(1) \\ r(2) \\ \vdots \\ r(p) \end{pmatrix} \qquad \text{(VI.29)}$$

Al tratarse de un sistema triangular su resolución es bastante simple. Su coste de cálculo es bastante menor, incluso, que el asociado al algoritmo de Levinson-Durbin utilizado para resolver las ecuaciones de Yule-Walker, el más eficiente y popular de los métodos de predicción lineal.

VI.1.3. Predicción Lineal de la Parte Causal de la Autocorrelación (OSALPC).

El método anterior (MIAC) es muy simple pero sus prestaciones en aplicaciones de tratamiento robusto de la voz no suelen ser muy destacadas. Sin embargo permite introducir la técnica OSALPC (One-Sided Autocorrelation Linear Predictive Coding) de una manera bastante simple. Para mejorar la estimación de los coeficientes a_k del modelo todo-polos de $R^+(z)$, expresión (VI.27), se suele considerar un sistema sobredeterminado con un número de M ecuaciones, donde $M>p$, evaluando (VI.28) para valores $m>0$.

Una justificación del uso de un sistema de ecuaciones sobredeterminado es que en la ecuaciones (VI.29) sólamente intervienen los valores estimados para la autocorrelación entre 0 y p, $\{r(m),\ 0 \le m \le p\}$. De este modo los coeficientes a_k obtenidos dependen totalmente de la calidad de las estimaciones de la secuencia autocorrelación para este márgen de valores de m y se sabe que éstos presentan un cierto error de estimación. Además, hay que tener en cuenta el efecto de bordes. Estos errores de estimación pueden compensarse mediante la utilización de más ecuaciones, en relación al mínimo número necesario (p ecuaciones), con lo

cual se hace intervenir en la obtención de los coeficientes a_k un mayor conjunto de valores estimados para la autocorrelación Por otro lado, si se pretende realizar una estimación fiable de los coeficientes a_k en presencia de ruido, la utilización de valores de autocorrelación alejados del origen puede ser favorable, pues, éstos son más robustos frente a un ruido de espectro plano que aquellos valores más cercanos al origen.

Siendo **M**, donde **(M>p)**, el mayor índice para el cual la autocorrelación puede ser estimada con fiabilidad, entonces, se puede construir el siguiente sistema sobredeterminado con **M+1** ecuaciones y **p** incógnitas, resultante de evaluar (VI.28) para **1≤m≤M** y añadir la ecuación **r(0)/2= e(0)** :

$$\begin{pmatrix} r(0)/2 & 0 & 0 & \cdots & 0 \\ r(1) & r(0)/2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r(p) & r(p-1) & r(p-2) & \cdots & r(0)/2 \\ r(p+1) & r(p) & r(p-1) & \cdots & r(1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r(M) & r(M-1) & r(M-2) & \cdots & r(M-p) \end{pmatrix} \cdot \begin{pmatrix} 1 \\ a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} = \begin{pmatrix} \varepsilon(0) \\ \varepsilon(1) \\ \vdots \\ \varepsilon(p) \\ \varepsilon(p+1) \\ \vdots \\ \vdots \\ \varepsilon(M) \end{pmatrix} \qquad \text{(VI.30)}$$

donde **e(m)** representa al error asociado con la estimación de los valores de la secuencia autocorrelación. Este sistema de ecuaciones puede resolverse minimizando el error cuadrático y se puede utilizar como un método fiable para estimar los coeficientes a_k del modelo todo-polos de $R^+(z)$, dado en (VI.27), utilizando un estimador adecuado para obtener los valores de la autocorrelación.

Este nuevo método de estimación de los coeficientes AR se denota como OSALPC. Este sistema (VI.30) puede resolverse de forma eficiente [Frie-79] debido al carácter Toeplitz de su matriz de coeficientes. Sin embargo, el método de autocorrelación de predicción lineal es más eficiente gracias a la posibilidad de aplicar el Algoritmo de Levinson-Durbin.

Esta técnica puede interpretarse como la predicción lineal de orden **p** de los valores de la secuencia autocorrelación **r(m)**, sustituyendo **r(0)** por **r(0)/2**, suponiendo nulos los valores de la secuencia autocorrelación anteriores y posteriores al intérvalo de predicción. Sin embargo, también puede interpretarse como la predicción lineal de orden **p** de los valores de la secuencia $r^+(m)$ con **1≤m≤M**, donde se suponen nulos los valores de dicha secuencia posteriores al intérvalo de predicción.

En ambas interpretaciones se supone un enventanado de la secuencia correspondiente hasta un índice **M** suficientemente grande. Sin embargo, sólo en la primera interpretación se hace necesaria la suposición de valores nulos cercanos al origen, cuyos efectos en la estimación pueden ser significativos. En consecuencia, la segunda interpretación de esta técnica OSALPC, como predicción lineal de la parte causal de la secuencia autocorrelación, es más realista. De ahí viene el nombre OSALPC para identificar esta técnica, pues en español se podría traducir como predicción lineal de la parte causal de la autocorrelación.

Su implementación práctica es bastante simple. Una vez estimados los valores de la parte causal de la secuencia de autocorrelación $\mathbf{r^+(m)}$ para $\mathbf{0 \le m \le M}$, se aplica sobre dicha secuencia el método de autocorrelación de predicción lineal. Para ello se calculan los coeficientes $\mathbf{rr(m)}$ de las ecuaciones de Yule-Walker aplicando un estimador sesgado de la autocorrelación sobre la parte causal de la secuencia de autocorrelación:

$$rr(m) = \sum_{n=0}^{M-m} r^+(n+m) \cdot r^+(n) \qquad \text{(VI.31)}$$

y finalmente se resuelven dichas ecuaciones utilizando el algoritmo de Levinson-Durbin.

De este modo, una vez estimada la parte causal de la secuencia de autocorrelación, $\mathbf{r^+(m)}$ para $\mathbf{0 \le m \le M}$, el coste de cálculo de esta técnica es el mismo asociado con el Método de Autocorrelación de predicción lineal aplicado sobre una trama de longitud $\mathbf{M+1}$ muestras. El mayor esfuerzo de cálculo pertenece al cálculo de las autocorrelaciones.

Interpretando la técnica OSALPC como predicción lineal de la parte causal de la secuencia autocorrelación, esta técnica se corresponde con un modelado todo-polos del espectro de $\mathbf{r^+(m)}$, es decir, el cuadrado de la envolvente espectral:

$$E^2(w) = \frac{\left(r^+(0) \right)^2}{\left| A\left(e^{jw}\right) \right|^2} \qquad \text{(VI.32)}$$

donde el polinomio $\mathbf{A(z)}$ del denominador viene dado por la expresión (VI.16) y el numerador se obtiene directamente a partir del modelo de $\mathbf{R^+(z)}$ dado en (VI.27).

Desde el punto de vista de modelado paramétrico de procesos también puede llegarse al mismo resultado. En la expresión (VI.31) puede observarse como los coeficientes de las ecuaciones de Yule-Walker son una estimación de la autocorrelación de la parte causal de la secuencia de autocorrelación. Entonces, la técnica OSALPC supone un modelado AR de la parte causal de la secuencia de autocorrelación y, en consecuencia, su espectro $\mathbf{E^2(w)}$ es todo-polos.

VI.2. El Algoritmo OSA_AR2.

Tal como hemos visto en algunos de los apartados anteriores, el uso de estadísticas de segundo orden durante la estimación de los coeficientes AR para el Filtrado Iterativo de Wiener, presenta dos deficiencias fundamentales frente a las estadísticas de tercer orden o cumulantes:

1) La menor velocidad de convergencia del método, puesto que el desacoplo entre señal y ruido que nos proporcionan las HOS dotan de mayor agresividad al modelo de orden superior.

2) Cierta cantidad de ruido musical o distorsión residual que permanece en la señal procesada a partir de la tercera iteración, acompañado de una pérdida de inteligibilidad progresiva.

Por contra, la mayor agresividad de la estimación con cumulantes se traduce también en un mayor efecto de picado espectral de los formantes de la señal de voz, para cada iteración en particular. Aquí, se propone un método alternativo de segundo orden que puede solucionar o compensar en parte las deficiencias arriba mencionadas y sin que éste represente una carga computacional añadida importante. El esquema general de filtrado es el mismo que hasta ahora. La diferencia estriba en que, en lugar de efectuar la estimación de coeficientes a partir de la señal de voz ruidosa, se realiza a partir de la autocorrelación causal de dicha señal

Esta idea se basa en el hecho de que ambas señales tienen los mismos polos en el denominador, tal como se ha mostrado en (VI.20), y por tanto las dos son igualmente válidas para estimar los coeficientes $\mathbf{a_k}$. Ello se debe a que la función de autocorrelación causal preserva los polos de la señal original [Ginn-83b]. La principal característica de dicha función reside en su menor sensibilidad al ruido y, en consecuencia, proporciona un modelado más rápido, acelerándose la velocidad de convergencia propia del Algoritmo AR2. Este algoritmo se ha denominado OSA_AR2 y su esquema general ha sido representado en la Fig.VI.1, donde se muestra como se obtiene el modelado AR en el dominio de la función autocorrelación causal, durante la etapa de diseño del filtro de Wiener $\mathbf{W_i}$.

En la Fig.VI.2 se comparan las estimaciones espectrales correspondientes al método de autocorrelación de predicción lineal aplicado sobre la señal de voz (algoritmo AR2) y la técnica OSALPC (algoritmo OSA_AR2) para un orden **p=10** del modelado AR. Se suponen

condiciones supuestamente libres de ruido (línea continua) y en presencia de ruido aditivo gaussiano blanco (líneas discontinuas) considerándose los tres niveles de ruido utilizados en capítulos anteriores. A primera vista, vemos en la Fig.VI.2.a y la Fig.VI.2.c cómo el método clásico es tal vez demasiado sensible al ruido cuando se consideran niveles medios y bajos de SNR, entre 0dB y 9 dB. En cambio, el algoritmo OSA_AR2 (Figs. VI.2.b y VI.2.d) consigue mejores resultados para estos niveles de ruido.

En ambientes altamente ruidosos, ambos métodos estiman bastante bien la zona alrededor del primer formante de la voz. Sin embargo en el resto de frecuencias hay una clara diferencia de comportamiento entre ambas técnicas en relación a su robustez frente al ruido. El espectro correspondiente a la técnica clásica (AR2) es muy sensible a la presencia de ruido: se produce una espectacular reducción del margen dinámico y la estructura de los formantes, a partir del segundo formante, queda totalmente alterada e incluso puede aparecer algun formante nuevo. Sin embargo, el cuadrado de la envolvente espectral (OSA_AR2) es mucho más insensible al ruido: se mantiene el margen dinámico y sólo cambian ligeramente la frecuencia central y el ancho de banda asociados a los formantes siguientes al primer formante. Así, se muestra claramente como la técnica OSALPC es mucho más robusta al ruido que la técnica clásica y, en consecuencia, se pueden esperar unas prestaciones mejores. Nótese, además, que en el caso de SNR_G=9dB y SNR_G=18dB, el método OSA_AR2 es muy poco sensible al ruido existente, mientras el algoritmo clásico AR2 muestra una gran sensibilidad a la presencia de éste, especialmente en la zona alta del espectro.

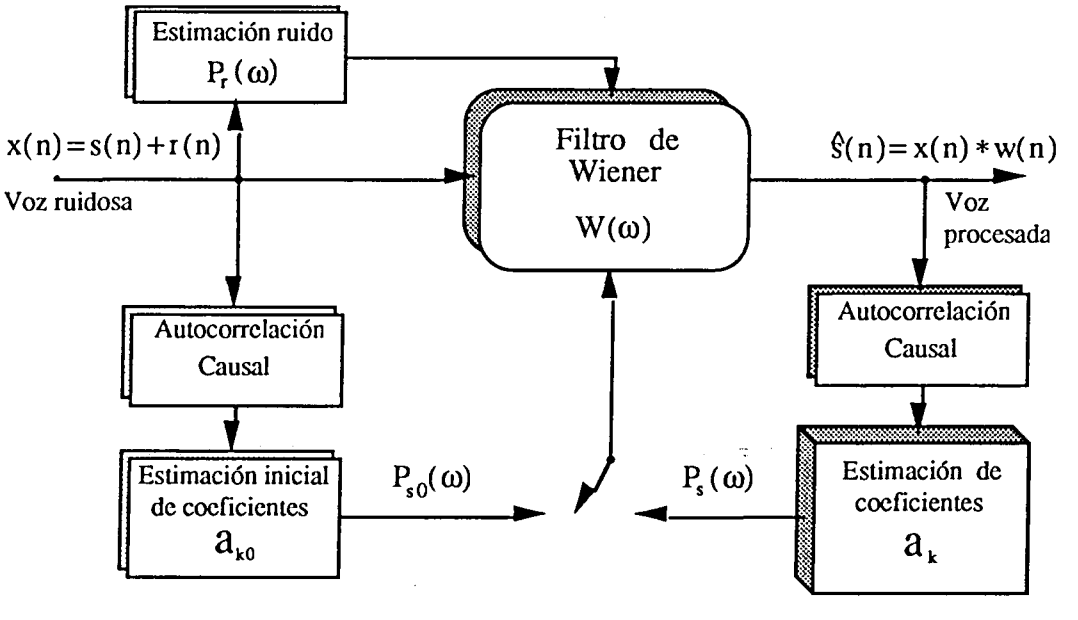


Figura VI.1 : Esquema general del Filtrado Iterativo de Wiener con estimación de coeficientes en el dominio de la Autocorrelación Causal.

También puede observarse en dicha figura como la envolvente espectral enfatiza fuertemente las bandas frecuenciales de mayor energía, hecho ya comentado en el Apartado VI.1.1. Por otro lado, si se comparan las dos estimaciones espectrales para el caso de ausencia de ruido, se observa como en la técnica OSALPC pueden aparecer formantes espurios con respecto a la técnica de predicción clásica. Este hecho puede explicarse teniendo en cuenta que el modelado espectral asociado a la técnica de predicción lineal clásica, consistente con el modelo lineal de producción de la voz (Fig.I.1), equivale a un modelado AR de la señal de voz. Considerando la señal de voz como un proceso autorregresivo, el cuadrado de la envolvente espectral de la señal de voz es una función con polos y ceros, tal como se ha visto anteriormente en (VI.24). Sin embargo la técnica OSALPC se ha derivado a partir de la simplificación (VI.26), que consiste en suponer un modelo todo-polos para el cuadrado de la envolvente espectral (VI.32).

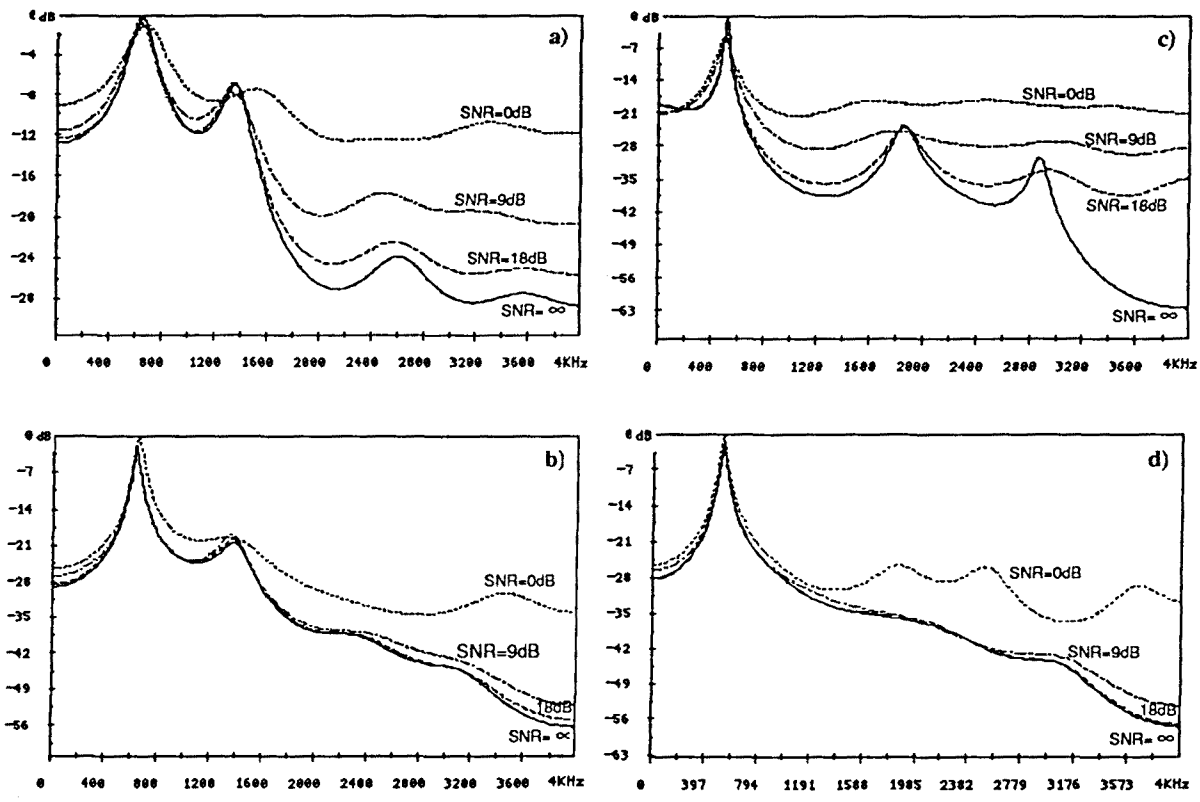


Figura VI.2 : Espectros LPC correspondientes a los supuestos siguientes : a) método clásico de 2º orden (AR2) sobre vocal /a/ sintética; b) método OSA_AR2 sobre vocal /a/ sintética; c) método AR2 sobre vocal /e/ real; d) método OSA_AR2 sobre vocal /e/ real.

Considerando que el procesado de la señal de voz ruidosa **x(n)** se lleva a cabo mediante tramas de longitud tal que nos garantice su estacionariedad, 256 muestras por trama, cabe ahora determinar qué longitud **L** de muestras de $\mathbf{r^+(m)}$ debe tomarse como entrada al algoritmo de estimación de los coeficientes AR.

En la Fig.VI.3 se muestra un estudio realizado para valores de **L** comprendidos entre 20 y 240 muestras, que nos ha permitido observar como para una longitud menor que el número de muestras equivalente al pitch del locutor ($\mathbf{L_p}$) las medidas empeoran, como consecuencia de una estimación demasiado pobre del filtro. Se debe esto a que la mayor parte de la energía de $\mathbf{r^+(m)}$ se concentra en las $\mathbf{L_p}$ primeras muestras de la secuencia, y por tanto, para $\mathbf{L<L_p}$ estamos perdiendo información. Por otra parte, longitudes mucho mayores de $\mathbf{L_p}$ no aportan apenas información adicional, aunque sí mayor tiempo de cálculo y cierta cantidad de ruido, puesto que los valores de la Autocorrelación Causal pierden fidelidad hacia los bordes, consecuencia inevitable de trabajar con tramas de duración finita. Como conclusión, se consideran valores de **L** comprendidos entre 60 y 120 muestras, como un buen compromiso entre el tanto por ciento de energía $\mathbf{r^+(m)}$ que se captura y el tiempo de cálculo invertido en la operación.

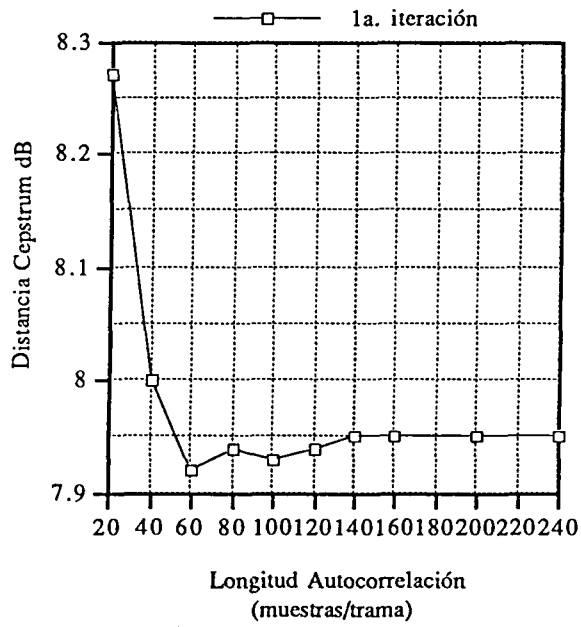


***Figura VI.3** : Estudio de las medidas espectrales obtenidas durante la primera iteración del algoritmo OSA_AR2, para distintas longitudes de la autocorrelación de la señal.*

A continuación vamos a evaluar el comportamiento de esta variante del método clásico, analizando primero el caso básico para pasar seguidamente a la inclusión del promediado intertrama de los coeficientes $\mathbf{a_k}$. En el párrafo anterior discutimos la conveniencia de considerar una u otra longitud de la autocorrelación de la señal en función del pitch del locutor por lo cual, dado que realizamos nuestro estudio sobre las señales ESCA y ASUN1, ámbas pertenecientes a locutores femeninos, se ha tomado un valor de **L=80** muestras en todas las pruebas presentadas a continuación.

Tal como hicimos con el resto de técnicas, analizamos el comportamiento general del método OSA_AR2 en el margen 0dB-24dB de SNR de la señal de entrada, para ver su capacidad ante señales de distinta calidad. Si nos fijamos en la representación de la distancia Cepstrum, Fig.VI.4, a simple vista se observa como su evolución no tiene nada que ver con la del método AR2 clásico, sinó que nos recuerda mucho más a las estadísticas de orden superior, HOS. Mientras AR2 obtenía una mejora escalonada, iteración a iteración, hasta llegar al mínimo, OSA_AR2 lo consigue tras casi una sola iteración en la mayoría de los casos. Incluso en la zona donde el nivel de ruido es muy importante (**SNR_G=0dB**) se alcanza la iteración óptima durante la segunda iteración, resultando una velocidad de convergencia superior a la asociada con el algoritmo AR3. Para **$SNR_G \geq$9dB** la iteración óptima resultante se corresponde con la primera, es decir, no se precisa ejecutar el algoritmo iterativo de

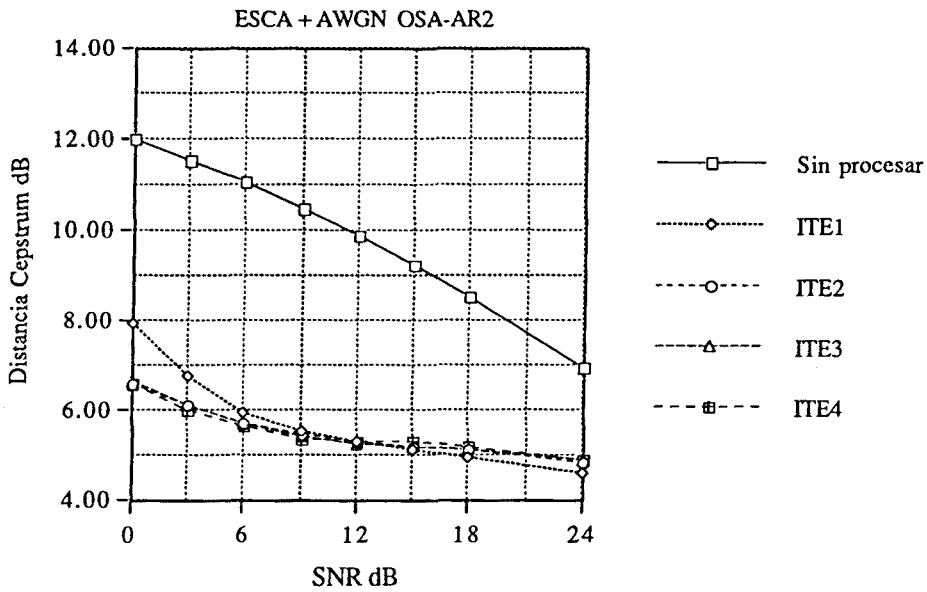


Figura VI.4 : Comportamiento del Algoritmo OSA_AR2 durante las 4 primeras iteraciones ante la presencia de distintos niveles de ruido.

Wiener: una única estimación por trama para el filtro de Wiener resulta suficiente en estas condiciones de ruido. Además se obtienen valores mínimos de distancia Cepstrum inalcanzables para los restantes algoritmos vistos anteriormente. Logramos bajar la medida de distancia Cepstrum desde 12dB hasta una medida absoluta de 6.6dB para SNR_G=0dB y hasta los 6.1dB para SNR_G=3dB, mientras que el método AR3 básico no iba más allá de 8.1dB y 7.7dB respectivamente.

A partir de SNR_G=12dB es suficiente con una sola pasada para alcanzar el mínimo, y se aprecia que la curva de mejora toma un aspecto aplanado, casi constante, que denota que se ha llegado al límite de extracción de ruido. Los resultados, aún así, siguen siendo superiores a los obtenidos con el resto de métodos, exceptuando el caso de señal poco ruidosa, SNR_G=24dB, donde el algoritmo AR2 consigue una medida de distancia más baja. Hay que hacer notar, por otro lado, que a pesar de seguir iterando tras haber llegado al mínimo en el primer filtrado, las medidas de distancia Cepstrum se mantienen aproximadamente en el mismo nivel. Sale a relucir aquí la naturaleza de segundo orden del método, que al igual que el algoritmo AR2 clásico, tras haber extraído todo el ruido que puede, degrada sólo muy lentamente la señal.

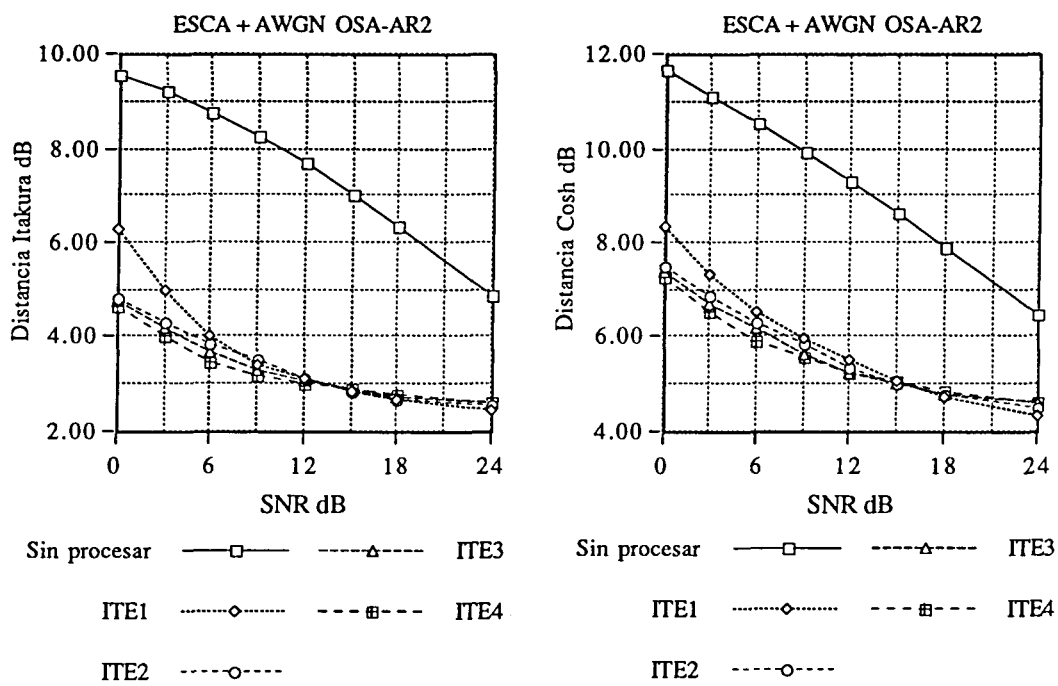


Figura VI.5 : Distancias Itakura y Cosh obtenidas durante las cuatro primeras iteraciones del Algoritmo OSA_AR2 para distintos niveles de ruido

El comportamiento de las distancias Cosh e Itakura, Fig.VI.5, es muy parecido al de la distancia Cepstrum, si bien no es tan contundente y en el margen 0dB-12dB, aunque se satura su mejora en la segunda iteración, continúa bajando unas décimas hasta la cuarta. Logramos de nuevo en este margen superar al resto de métodos, bajando para SNR_G=0dB hasta 7.2dB y 4.6dB en ambas distancias en la cuarta iteración, mientras que el algoritmo AR3 no conseguía bajar de 7.6dB y 4.9dB respectivamente en la quinta iteración. Aunque las diferencias no sean tan amplias como en distancia Cepstrum, siguen siendo significativas de la superior agresividad del método OSA_AR2. A partir de SNR_G=15dB vuelve a ser suficiente una iteración para alcanzar el mínimo, y al igual que antes continuar filtrando no empeora apenas la medida. Nuevamente para las SNR_G más altas el algoritmo OSA_AR2 se ve superado por otros métodos, sobretodo por AR2, como consecuencia de una excesiva agresividad en ese margen de bajos niveles de ruido degradante.

Si analizamos conjuntamente las mejoras globales introducidas en los tres tipos de distancia, Fig.VI.6, un hecho inusual nos llama especialmente la atención: mientras que con los otros métodos de extracción de ruido la distancia Cepstrum era la que experimentaba mejoras más bajas, con OSA_AR2 se sitúa en todo el margen por encima de la distancia Cosh, e incluso por encima de la Itakura para $SNR_G \leq 9$dB. La explicación es sencilla si nos atenemos al modo de funcionamiento de cada método. En los casos AR2 y AR3 el modelado de la señal se realiza de modo que los formantes de la voz (frecuencias de mayor energía del

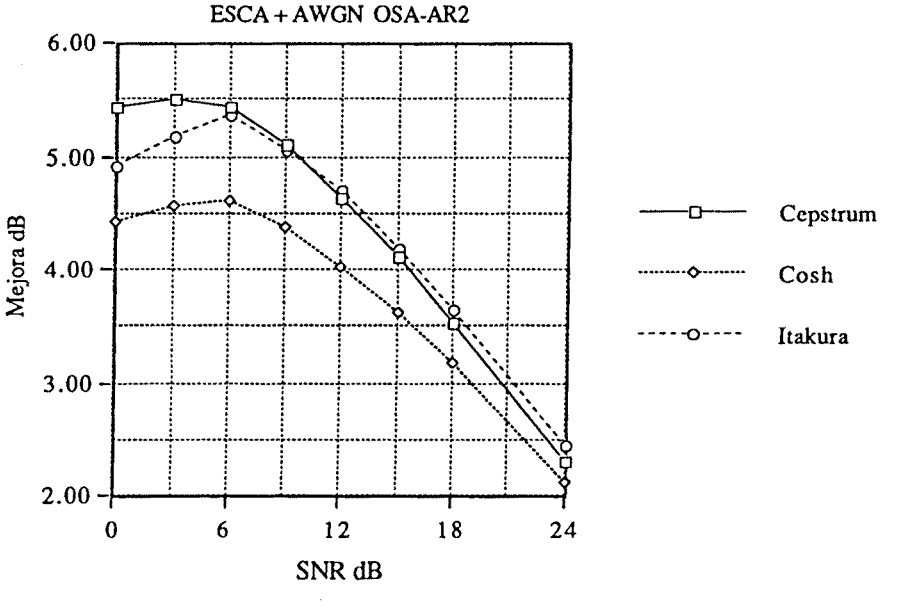


Figura VI.6 : Evolución de las mejoras globales introducidas por el método OSA_AR2 ante distintos niveles de ruido .

espectro) se ven favorecidos por la atenuación de los valles espectrales y la zona de alta frecuencia (frecuencias de menor energía del espectro). Recordemos que por este motivo el sobrefiltrado de la señal deriva en el efecto de picado espectral, discutido en los Capítulos VI.5. y V.3., más acusado en AR3 que en AR2. Puesto que la distancia Cepstrum evalúa todo el espectro por igual y las distancias Cosh e Itakura centran más su estudio en la forma y posición de los formantes, parece lógico que estas últimas obtengan en esos casos mejores resultados que la primera. En el caso OSA_AR2, sin embargo, la autocorrelación causal previa de la señal enfatiza las frecuencias de menor energía de la señal, por lo que el modelado posterior se hace a partir de una señal más equilibrada; los resultados son más equilibrados y la distancia Cepstrum se ve por ello favorecida. En cualquier caso, si analizamos la gráfica de mejoras veremos como la agresividad del método OSA_AR2 es especialmente efectiva en los casos de calidad media y baja de la señal, quedando definitivamente descartado para señales de calidad superior.

Un análisis de la SNR_G correspondiente a la señal de voz realzada obtenida al final de cada iteración, Fig.VI.7, confirma lo comentado anteriormente. Con mejoras globales por encima de las de AR3, la forma prácticamente recta de las curvas de resultados a lo largo de todo el margen de SNR de entrada y su posición casi constante a lo largo de las sucesivas iteraciones, son claros indicativos de la agresividad y robustez del método frente al ruido

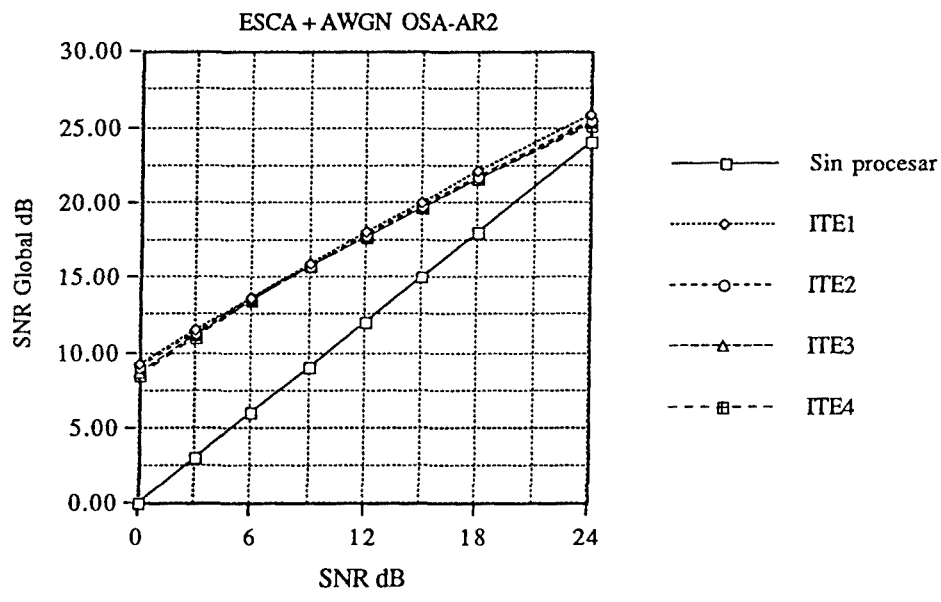


Figura VI.7 : Evolución de la medida de distancia temporal SNR_G para el Método OSA_AR2 ante distintos niveles de ruido.

perturbador, así como de la baja varianza que presenta y de la escasa degradación de la señal cuando se ejecuta un número de iteraciones superior a la óptima.

Las pruebas de audición desvelan, sin embargo, que en realidad los tests subjetivos resultantes de escuchar la voz procesada por este método no ofrecen una mejora tan espectacular como parecen indicar las medidas objetivas que hemos analizado. La convergencia es ahora más rápida y el ruido musical remanente, aunque no desaparece del todo, se reduce considerablemente. No obstante, también apreciamos una pérdida importante de inteligibilidad debida al fuerte recorte de la señal que se produce como consecuencia de la excesiva agresividad que el método parece tener gracias a esa autocorrelación causal previa.

A continuación presentamos las medidas resultantes de los casos SNR_G=0dB, 9dB y 18dB de entrada, representativos de entornos cuya calidad de la señal de voz es, respectivamente, baja, media y alta.

0dB	SNR_{Global}	$SNR_{Seg.}$	ITAKURA	COSH	CEPSTRUM
Original	0.022	0.765	9.575	11.665	12.020
1 iter.	9.177	5.986	6.282	8.335	7.936
2 iter.	9.058	6.144	4.794	7.467	6.577
3 iter.	8.768	6.046	4.758	7.359	6.650
4 iter.	8.579	5.974	4.642	7.249	6.599
5 iter.	8.451	5.938	4.675	7.278	6.608

Tabla VI.1 : Evaluación del Algoritmo OSA_AR2 en ambientes muy ruidosos (SNR_G=0dB).

9dB	SNR_{Global}	$SNR_{Seg.}$	ITAKURA	COSH	CEPSTRUM
Original	9.021	8.073	8.276	9.923	10.510
1 iter.	15.838	11.596	3.415	5.974	5.566
2 iter.	15.741	11.683	3.3.503	5.857	5.480
3 iter.	15.682	11.660	3.306	5.644	5.444
4 iter.	15.658	11.624	3.191	5.552	5.385
5 iter.	15.646	11.601	3.157	5.533	5.388

***Tabla VI.2** : Evaluación del Algoritmo OSA_AR2 en ambientes ruidosos (SNR_G=9dB).*

18dB	SNR_{Global}	$SNR_{Seg.}$	ITAKURA	COSH	CEPSTRUM
Original	18.019	13.408	6.328	7.893	8.518
1 iter.	22.094	17.274	2.691	4.725	4.976
2 iter.	21.782	17.009	2.665	4.760	5.116
3 iter.	21.593	16.881	2.709	4.803	5.135
4 iter.	21.492	16.819	2.767	4.849	5.180
5 iter.	21.437	16.790	2.783	4.865	5.183

***Tabla VI.3** : Evaluación del Algoritmo OSA_AR2 en ambientes poco ruidosos (SNR_G=18dB).*

A modo de conclusión, este método OSA_AR2 nos proporciona, respecto al algoritmo clásico de Lim-Oppenheim, un aumento de calidad de la señal en detrimento de una fuerte pérdida de inteligibilidad. Ello lo descarta lógicamente para aquellas aplicaciones en que la inteligibilidad sea un objetivo importante o necesario, pero será de gran utilidad en otras aplicaciones que no requieran esa faceta. Un campo importante de investigación dentro del Procesado de la Señal que cumple esos requisitos es, por ejemplo, el de Reconocimiento del Habla. Sabemos que existen distintos sistemas y algoritmos de Reconocimiento que basan sus lógicas de decisión únicamente en la distancia Cepstrum existente entre las señales, independientemente del grado de inteligibilidad del mensaje. La aplicación del método

OSA_AR2 en estos casos puede resultar innovadora respecto de la estimación clásica AR2 de coeficientes que se viene utilizando de manera casi tradicional. De hecho, ya se han realizado pruebas con una estimación inicial (sin filtrado) de coeficientes a_k efectuada con este nuevo método, y los resultados de tasas de reconocimiento obtenidos, son claramente mejores que los alcanzados mediante los algoritmos AR2 y AR3 [Sala-95a], [Sala-95b]. Recordemos que para una SNR_G=0dB, zona donde más espectacularmente se degradan las tasas de éxito de los distintos sistemas de reconocimiento, la distancia Cepstrum disminuía de forma significativa hasta la segunda iteración (la mejora a partir de ahí era mínima), por lo que una estimación de los coeficientes a partir de la señal de voz realzada saliente del primer o segundo filtrado puede dar mejores resultados que esa estimación inicial obtenida a partir de la voz ruidosa.

VI.3. El Algoritmo OSA_AR2 con Ponderación Intertrama (OSA_AR2_IF).

En este apartado se trata de mejorar las prestaciones del algoritmo OSA_AR2, en la orientación de aplicaciones sensibles a la calidad de la señal y no a su inteligibilidad. Para ello se ha efectuado el estudio del algoritmo OSA_AR2 con la incorporación del promediado intertrama de coeficientes. Ello aumenta todavía más, si cabe, la agresividad del algoritmo OSA_AR2 básico. Para este algoritmo la ponderación intertrama de los parámetros AR tiene sentido sólamente para niveles altos de ruido (SNR_G<9dB), segun los resultados obtenidos en el apartado anterior.

El estudio del método OSA_AR2_IF se ha llevado a cabo siguiendo las mismas pautas que con los métodos AR2 y AR3, es decir, evaluando valores del factor **IF** entre 0 y 1 y valores de **PFI** comprendidos entre 1 y 5, aunque valores altos de **PFI** carecen de sentido en este caso particular debido a la alta velocidad de convergencia del algoritmo OSA_AR2. Considerando que estamos de nuevo ante un sistema ya de por sí muy agresivo, se empieza analizando el caso correspondiente al entorno más ruidoso (SNR_G=0dB), precisamente donde parece que la ponderación intertrama puede aportar mas beneficios. Posteriormente se evaluan sus prestaciones para un nivel de ruido en la banda opuesta (SNR_G=9dB). Se presentan los resultados obtenidos para los ficheros ESCA y ASUN1.

VI.3.1. Ambientes altamente ruidosos.

Empezamos analizando el comportamiento de la distancia Cepstrum para ambos ficheros de voz. En la Fig.VI.8 se han representado los resultados obtenidos durante la primera iteración de filtrado. De entrada hay que resaltar que partimos de unos valores de distancia realmente muy bajos correspondientes al supuesto de no promediado intertrama (**IF=1**). Véase a qué valores de distancia Cepstrum están situados los algoritmos AR2 e incluso AR3 tras procesar la primera iteración. No resulta extraño que las mejoras relativas (que no absolutas) introducidas por el factor **IF** para este método sean inferiores a las alcanzadas cuando se aplica dicho factor **IF** a los algoritmos AR2 y AR3, bastante más sensibles al ruido.

En concordancia con los estudios anteriores, la consideración del valor **PFI=1** conduce a mejoras escasas o nulas, su curva es casi constante, ya que significa promediar con coeficientes de la trama anterior obtenidos en presencia de un nivel de ruido similar al de la trama actual.

Ambos ficheros de señal alcanzan la curva de mínima distancia Cepstrum para el valor **PFI=2.** La consideración de un valor superior para **PFI** no aporta apenas mejoras, si bien tampoco empeora. Se pone de manifiesto la naturaleza del método OSA_AR2 básico, que alcanza su mínimo en la segunda pasada del filtro y no distorsiona al ir más allá. Lógicamente, el único valor esperado como óptimo para el parámetro **PFI** era, a priori, el valor correspondiente a la segunda iteración, segun el comportamiento observado durante la evaluación del algoritmo OSA_AR2.

En relación a la distorsión ocasionada, los valores útiles del parámetro **IF** son los comprendidos entre 0.5 y 1 . Aunque estamos ahora ante un método que supuestamente introduce menos distorsión 'iterativa' que el algoritmo AR3, en realidad ésta puede llegar

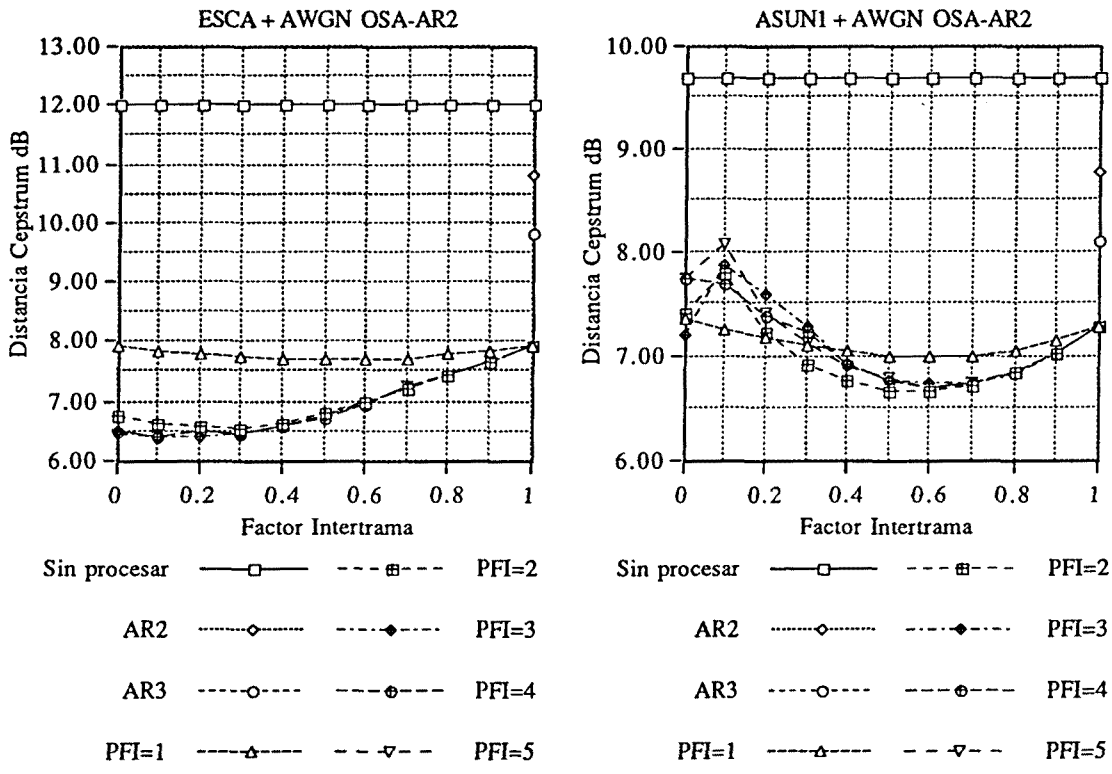


Figura VI.8 : Comportamiento del Algoritmo OSA_AR2_IF tras procesar la primera iteración para un nivel de ruido alto (SNR_G=0dB)

también a ser importante, como queda demostrado con la señal ASUN1 para **IF<0.5**, tal como corroboran las pruebas de audición realizadas. Se ha observado con distintas señales que mientras para **IF≥0.5** el método presenta un comportamiento regular , para **IF<0.5** puede dar lugar a comportamientos extraños, por lo que este margen será definitivamente descartado.

Concretando, en el límite inferior del margen útil, la señal ESCA para unos valores **PFI=2** e **IF=0.5** logra disminuir los 7.94dB, obtenidos mediante el algoritmo OSA_AR2 **(IF=1)**, hasta los 6.62dB tras procesar únicamente la primera iteración, resultando valores de distancia Cepstrum realmente inimaginables con otros métodos. La frase ASUN1, por su parte, también con **PFI=2** e **IF=0.5**, baja hasta los 6.66dB, mientras que con **IF=1.0** se situaba en 7.28dB. Nótese como se alcanzan reducciones de distancia Cepstrum superiores a los 5dB tras procesar una sola iteración del algoritmo OSA_AR2_IF.

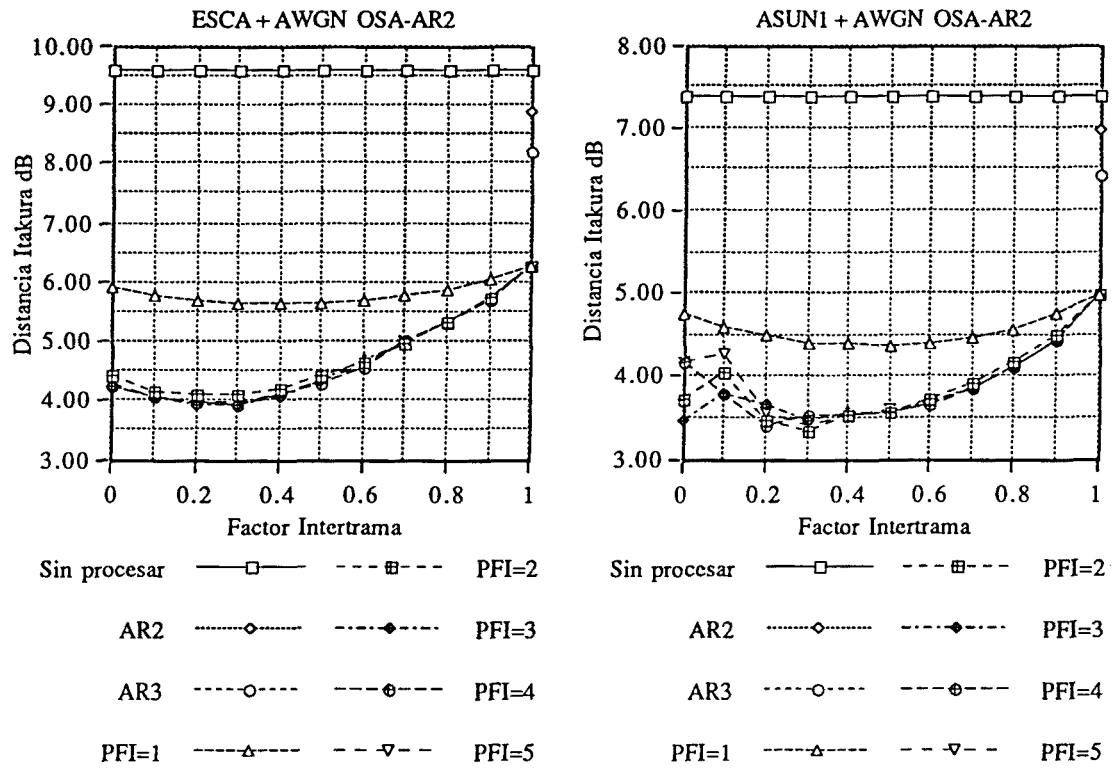


Figura VI.9 : Distancias Cosh e Itakura tras la primera iteración para el método OSA_AR2_IF en presencia de un nivel de ruido SNR_G=0dB.

Analizando la evolución de la distancia Itakura, Fig.VI.9, se aprecian resultados similares a los de la distancia Cepstrum, aunque la distorsión parece afectar en menor medida. Pero, al igual que sucedía antes, a partir de **PFI=2** las curvas no mejoran, aunque tampoco empeoran. Nuevamente estamos muy por debajo de los valores de distancia resultantes de la aplicación de los algoritmos AR2 y AR3. Pasamos así de los 6.28dB para el algoritmo OSA_AR2, para la señal ESCA, a los 4.20dB si utilizamos **PFI=2** con **IF=0.5** procesando únicamente la primera iteración. Del mismo modo se reduce este valor de distancia para el fichero ASUN1, pues de los 4.99dB para **(IF=1)** pasamos a 3.57dB al considerar también **PFI=2** con **IF=0.5**.

0dB	SNR_{Global}	$SNR_{Seg.}$	ITAKURA	COSH	CEPSTRUM
Original	0.022	0.765	9.575	11.665	12.020
1 iter.	9.142	5.963	4.190	7.225	6.620
2 iter.	9.009	6.163	4.453	7.164	6.426
3 iter.	8.766	6.113	4.579	7.111	6.512
4 iter.	8.599	6.059	4.406	6.993	6.412
5 iter.	8.469	6.017	4.419	7.018	6.387

TablaVI.4 : Evaluación del Algoritmo OSA_AR2_IF tomando unos valores IF=0.5 y PFI=2 en un ambiente muy ruidoso (SNR_G=0dB) para el locutor ESCA.

0dB	SNR_{Global}	$SNR_{Seg.}$	ITAKURA	COSH	CEPSTRUM
Orig.	0.056	0.894	7.372	10.048	9.677
1 iter.	8.366	4.946	3.519	6.930	6.648
2 iter.	8.001	4.872	3.450	6.699	7.072
3 iter.	7.550	4.587	3.500	6.529	6.933
4 iter.	7.282	4.373	3.456	6.556	6.965
5 iter.	7.149	4.259	3.499	6.647	6.964

Tabla VI.5 : Evaluación del Algoritmo OSA_AR2_IF tomando unos valores IF=0.5 y PFI=2 en un ambiente muy ruidoso (SNR_G=0dB) para el locutor ASUN1.

En las Tablas VI.4 y VI.5 puede apreciarse la evolución de todas las medidas correspondientes a este nivel de ruido SNR_G=0dB bajo la consideración de los valores **PFI=2** e **IF=0.5** . Se observa como la distancia Cosh mejora de forma similar a las distancias Cepstrum e Itakura, mientras las medidas temporales se mantienen a valores muy similares a los resultantes para el supuesto de no usar la ponderación intertrama **(IF=1)**.

No hay que olvidar, sin embargo, el deficiente grado de inteligibilidad proporcionado por el algoritmo OSA_AR2 básico, por lo que ahora continua siéndolo. Recordemos que por este motivo enfocamos los avances conseguidos con este método hacia aplicaciones en que prime la calidad de la señal (o incluso la medida objetiva que es la distancia espectral) independientemente de su inteligibilidad. Además de sus atractivos valores de distancia espectral, merece destacar su elevada velocidad de convergencia, pudiéndose destacar que el algoritmo OSA_AR2_IF prácticamente no necesita hacer uso del carácter iterativo del algoritmo y, en consecuencia, implica un importante ahorro de complejidad de cálculo.

VI.3.2. Ambientes con un nivel intermedio de ruido.

Al considerar un nivel de ruido inferior SNR_G=9dB para la señal de entrada representa situarnos ya en el límite de utilidad del algoritmo de promediado, puesto que la segunda iteración correspondiente al algoritmo OSA_AR2 apenas mejora 0.1dB en términos de distancia Cepstrum. El promediado intertrama pierde, entonces, su efectividad (no tiene sentido promediar con una señal igual o peor que la actual). La mejora en este caso se reduce a unas pocas décimas de decibelio para la distancia Cepstrum correspondiente a la primera iteración del algoritmo OSA_AR2_IF al tomar los valores **IF=0.6** y **PFI=2**, tal como se muestra en las Tablas VI.6 y VI.7 .

Como conclusión, se puede afirmar que no resulta aconsejable el uso de la Ponderación Intertrama para señales de voz ruidosa cuya relación señal a ruido global supere los 9dB, pues, el método OSA_AR2 básico **(IF=1)** obtiene unos resultados más óptimos bajo estas condiciones.

9dB	SNR_{Global}	$SNR_{Seg.}$	ITAKURA	COSH	CEPSTRUM
Original	9.021	8.073	8.276	9.923	10.510
1 iter.	15.822	11.581	3.020	5.768	5.396
2 iter.	15.749	11.702	3.255	5.637	5.335
3 iter.	15.687	11.679	3.095	5.484	5.308
4 iter.	15.663	11.642	2.995	5.378	5.239
5 iter.	15.652	11.616	2.965	5.350	5.240

Tabla VI.6 : Evaluación del Algoritmo OSA_AR2_IF tomando unos valores ***IF=0.5*** *y* ***PFI=2*** *en un ambiente ruidoso (SNR_G=9dB) para el locutor ESCA.*

9dB	SNR_{Global}	$SNR_{Seg.}$	ITAKURA	COSH	CEPSTRUM
Original	9.050	8.108	5.960	8.317	8.155
1 iter.	14.730	10.222	2.235	5.513	5.419
2 iter.	14.711	10.389	2.470	5.663	5.868
3 iter.	14.678	10.372	2.503	5.598	5.851
4 iter.	14.657	10.348	2.431	5.508	5.763
5 iter.	14.644	10.334	2.379	5.519	5.672

Tabla VI.7 : Evaluación del Algoritmo OSA_AR2_IF tomando unos valores ***IF=0.6*** *y* ***PFI=2*** *en un ambiente ruidoso (SNR_G=9dB) para el locutor ASUN1.*