



UNIVERSITAT POLITÈCNICA DE CATALUNYA  
Departament de Teoria del Senyal i Comunicacions

Proyecto Final de Carrera:  
**Reducción de Ruido en Comunicaciones Móviles**  
**Proyecto RERCOM**

Septiembre 2004

Autores: Antonio Marín Jiménez  
Marta del Pozo Ríos  
Director: Josep M<sup>a</sup> Salavedra Molí



**Índice:**

0.- Introducción.	0-1
1.- Modelado de la voz.	I-1
1.1.-Mecanismo de producción de la voz y su modelado.	I-1
1.1.1.-Esquema general de producción de la voz.	I-10
1.2.- El oído.	I-11
1.3.-Características acústicas y estadísticas de la voz y el oído.	I-12
2.- Técnicas básicas de eliminación de ruido en señales de voz.	II-1
2.1.- Técnicas basadas en el dominio de la frecuencia.	II-2
2.1.1.- Filtrado de Wiener.	II-2
2.1.2.- Filtrado de Wiener basado en la sustracción espectral.	II-4
2.1.3.- Filtro en peine adaptativo.	II-6
2.2.- Técnicas basadas en el dominio del tiempo.	II-7
2.2.1.- Filtrado de mediana.	II-7
3.- Modelo Autorregresivo.	III-1
3.1.- Función de transferencia.	III-2
3.2.- Modelo AR de segundo orden.	III-3
3.2.1.- Minimización del error cuadrático medio.	III-3
3.2.2.- Efectos del modelo AR de segundo orden.	III-7
3.3.- Estadísticas de orden superior (HOS).	III-10
3.3.1.-Definiciones Temporales y Frecuenciales.	III-14
3.3.1.1.-Momentos y Cumulantes.	III-14
3.3.1.2.- Propiedades de los Cumulantes.	III-17
3.3.2.-Espectros de Orden Superior.	III-19
3.3.3.-Estimación de los cumulantes y sus poliespectros.	III-22
3.3.3.1.-Estimadores Paramétricos.	III-23
3.3.3.1.1.-Método recursivo de tercer orden (TOR).	III-25
3.3.3.1.2.-Método de los momentos promedio de tercer orden (CTOR).	III-27
3.3.3.1.2.-Método AR Optimizado (OARM).	III-28
3.3.3.1.4.-Método de las ecuaciones de Yule-Walker de orden superior.	III-29
3.3.4.-Modelo autorregresivo de orden 2, orden 3, orden 4.	III-32
4.-Efectos del filtrado de Wiener con modelos AR.	IV-1

## Índice

---

4.1.-Reducción del ancho de banda de los formantes.	IV-1
4.1.1.-Efecto de picado en el método de correlaciones.	IV-2
4.1.2.-Efecto de picado en el método de cumulantes.	IV-14
4.2 Distorsión espectral.	IV-17
4.3 Desplazamiento de los formantes.	IV-19
4.4 Ruido Residual. Pérdida de reconocimiento del locutor.	IV-20
5.-VAD: Detector de actividad de Voz.	V-1
5.1.-Estimación del Ruido.	V-3
5.2.-VAD basado en la energía.	V-4
5.3.-VAD basado en la distancia espectral.	V-7
6.- Implementación programa de simulación RERCOM.	VI-1
6.1.- Segmentación.	VI-3
6.2.- Estimación Ruido.	VI-5
6.2.1.- Enventanado de trama.	VI-7
6.2.2.- Estimación densidad espectral de energía de ruido.	VI-8
6.2.3.- Promediado de espectros.	VI-9
6.3.- Prefiltro.	VI-11
6.3.1.- Enventanado de trama.	VI-11
6.3.2.- Estimación densidad espectral de energía de señal.	VI-13
6.3.3.- Construcción prefiltro.	VI-14
6.3.4.- prefiltrado.	VI-16
6.4.- Filtro.	VI-18
6.4.1.- Enventanado de trama.	VI-19
6.4.2.- Estimación densidad espectral de energía de señal.	VI-19
6.4.3.- Modelado AR de la voz.	VI-20
6.4.4.- VAD.	VI-23
6.4.5.- Estimación pitch.	VI-25
6.4.6.- Construcción filtro.	VI-28
6.4.7.- filtrado.	VI-32
6.5.- Postfiltrado.	VI-34
6.6.- Reconstrucción de la señal de voz.	VI-36
6.7.- Modificaciones del algoritmo básico de filtrado.	VI-40
6.7.1.- Filtrado de Wiener iterativo.	VI-40
6.7.2.- Modelado AR utilizando estadísticas de orden superior.	VI-42

## Índice

---

7.- Resultados pruebas de simulación.	VII-1
7.1.- Medidas de evaluación objetiva	VII-1
7.1.1.- SNR global.	VII-2
7.1.2.- SNR segmentada.	VII-2
7.1.3.- Distancia Itakura (LLR).	VII-3
7.1.4.- Distancia Itakura-Saito (IS).	VII-3
7.1.5.- Distancia Cepstrum (Ceps).	VII-4
7.1.6.- Distancia Log Area Ratio (LAR).	VII-5
7.1.7.- Cálculo de la amplificación aplicada a la señal test.	VII-6
7.2.- Elección optima de parámetros.	VII-7
7.2.1.- Elección de la longitud de trama.	VII-7
7.2.2.- Elección del desplazamiento de trama.	VII-9
7.2.3.- Activación de filtros.	VII-7
7.2.4.- Parámetros $\beta$ y $\delta$ .	VII-14
7.2.5.- Parámetro número de iteraciones	VII-19
7.2.6.- Parámetro activar filtro peine.	VII-26
7.2.7.- Parámetro añadir prefiltro.	VII-29
7.2.8.- Parámetro activar VAD.	VII-32
7.2.9.- Modificación del orden de predicción para el Modelado AR del filtro.	VII-39
7.2.10.- Parámetro activar postfiltro.	VII-42
7.3.- Comparativa con estándar Advance Front-End.	VII-45
7.3.1.-Comparativa de medidas objetivas.	VII-45
7.3.1.1.-Ruidos de banda ancha.	VII-47
7.3.1.1.1.-Ruido blanco.	VII-47
7.3.1.1.2.-Ruido rosa.	VII-52
7.3.1.2.-Ruidos de motor.	VII-57
7.3.1.2.1.-Ruido de motor de avión.	VII-57
7.3.1.2.2.-Ruido de un motor de explosión.	VII-62
7.3.1.2.3.-Ruido de motor de coche.	VII-67
7.3.1.2.4.- Ruido de motor de coche con fluctuaciones.	VII-72
7.3.1.3.-Ruidos de diferentes ambientes reales.	VII-77
7.3.1.3.1.-Ruido de fábrica.	VII-77
7.3.1.3.2.-Ruido de tráfico.	VII-82
7.3.1.3.3.-Ruido de tren.	VII-87

## Índice

---

7.4.-Comparativa de reconocimiento de voz con Aurora versión 008.	VII-92
8.- Implementación y pruebas del programa RERCOM_DSP.	VIII-1
8.1- Implementación RERCOM_DSP.	VIII-1
8.1.1.- Segmentación y enventanado.	VIII-2
8.1.2.- Algoritmo de doblefiltrado.	VIII-4
8.1.2.1.- Prefiltro.	VIII-5
8.1.2.2.- VAD.	VIII-5
8.1.2.3.- Estimación DEE ruido.	VIII-5
8.1.2.4.- Filtro.	VIII-6
8.1.2.- Reconstrucción señal.	VIII-7
8.2.- Comparativa RERCOM_DSP vs Advance Front-End.	VIII-8
8.2.1.- Comparativa de medidas objetivas.	VIII-8
8.2.1.1.- Ruidos de banda ancha.	VIII-8
8.2.1.1.1.- Ruido blanco.	VIII-8
8.2.1.1.2.- Ruido rosa.	VIII-9
8.2.1.2.- Ruidos de motor.	VIII-10
8.2.1.2.1.- Ruido de motor de avión.	VIII-10
8.2.1.2.2.- Ruido de un motor de explosión.	VIII-11
8.2.1.2.3.- Ruido de motor de coche.	VIII-11
8.2.1.2.4.- Ruido de motor de coche con fluctuaciones.	VIII-12
8.2.1.3.- Ruidos de diferentes ambientes reales.	VIII-13
8.2.1.3.1.- Ruido de fábrica.	VIII-13
8.2.1.3.2.- Ruido de tráfico.	VIII-14
8.2.1.3.3.- Ruido de tren.	VIII-14
8.2.2.- Comparativa de reconocimiento	VIII-16
9.- Conclusiones	IX-1
10.- Bibliografía	X-1

## Índice

---

### **- Tomo Anexos -**

Anexo A: Código RERCOM 1.18b	A-1
definiciones.h	A-6
main.c	A-8
senal.h	A-27
senal.c	A-28
parametros.h	A-36
parametros.c	A-38
estimacion.h	A-46
estimacion.c	A-47
modelado.h	A-50
modelado.c	A-51
vads.h	A-56
vads.c	A-57
filtrado.h	A-59
filtrado.c	A-60
utilidades.h	A-65
utilidades.c	A-67
Anexo B: Código DISTCALC 1.02b	B-1
definiciones.h	B-3
main.c	B-5
senal.h	B-16
senal.c	B-17
medidas.h	B-20
medidas.c	B-21
Anexo C: Código RERCOM_DSP 0.4	C-1
definiciones.h	C-2
main.c	C-5
senal.h	C-10
senal.c	C-11
wiener.h	C-18
wiener.c	C-20
funciones.h	C-27
funciones.c	C-29



## 0.- Introducción

Los avances tecnológicos de los últimos años han permitido recuperar y poner en práctica potentes técnicas ya introducidas a mediados del siglo XX, como las redes neuronales o las estadísticas de orden superior. A pesar de sus variopintas posibilidades, las limitaciones de cálculo de la época no contribuyeron a que estas herramientas matemáticas tan versátiles fueran más allá del plano teórico. El avance a pasos agigantados de la electrónica, especialmente en el diseño de circuitos integrados y, en menor medida, de los circuitos impresos, ha facilitado mediante tecnología VLSI (Very Large Scale Integration) disponer de máquinas cada día más potentes. Tanto para un estudio en laboratorio, gracias a microprocesadores de las últimas generaciones y memorias del orden de Gbytes, como a nivel de implementación práctica con el uso de DSP's (Digital Signal Processing) cada vez más rápidos, con menor consumo y más espacio de memoria, el nivel de cálculo disponible ha puesto en primer plano técnicas muy potentes, pero de una carga computacional elevada.

No es de extrañar que debido a estos avances, en las dos últimas décadas, gran parte de los esfuerzos invertidos en el Procesado Digital de la Señal hayan ido dirigidos hacia la mejora de esos algoritmos, rescatando en algunos casos potentes herramientas matemáticas que antes resultaban excesivamente lentas y complejas.

Sin embargo, desgraciadamente, aunque en la mayoría de aplicaciones los sistemas de procesado de voz presentan excelentes prestaciones en ambiente de laboratorio, se degradan drásticamente al considerar entornos reales de funcionamiento, especialmente en los ambientes más ruidosos (coche, avión,...) dada su poca robustez. Por este motivo se han destinado muchos recursos a la investigación de técnicas de tratamiento del habla en entornos adversos, originándose las denominadas Técnicas Robustas de Procesado de Voz.

Se entiende por **Speech Enhancement** el procesado o pre-procesado de señales de voz destinadas a ser directamente escuchadas por el oído humano, o bien como preparación previa a otros sistemas de tratamiento de la señal, como por ejemplo Codificación o Reconocimiento. El objetivo fundamental de estas técnicas consiste en mejorar uno o varios aspectos perceptuales de la voz, como son básicamente su calidad, su inteligibilidad y el grado de fatiga que produce en el oyente.

Centraremos nuestro estudio en aquellos casos en que el ruido, presente en el entorno del locutor, se superponga y degrade la señal de voz original, siempre desde la perspectiva de un Sistema de **Micrófono Simple**, es decir, disponiendo únicamente de la señal de voz ya contaminada con ruido. Estaremos entonces condicionados, como veremos, por la estacionariedad de la interferencia.

Nos encontramos, por tanto, ante una de las situaciones que presentan más dificultades en Speech Enhancement, pues no disponemos de una señal de referencia del ruido y la señal de voz original no ha sido tampoco previamente procesada. Este desconocimiento inevitable de ambas señales originará que todas las soluciones a las que se llegue sean subóptimas.

En el contexto descrito se ha implementado un algoritmo basado en el filtrado de Wiener iterativo, partiendo de la técnica propuesta por Lim y Oppenheim [Lim-79] e inspirado en proyectos anteriores como [Jove-93] y [Esta-95], donde se realizaron implementaciones de sistemas de Speech Enhancement con filtrado de Wiener iterativo que introducían el uso de las estadísticas de orden superior (HOS), permitiendo dar un paso más en la mejora de señal de voz contaminada con ruido (entornos de oficina, calles, vehículos a motor, comunicaciones telefónicas...), ya sea mediante un aumento de la calidad de la señal, una mayor inteligibilidad o en una menor fatiga del oyente.

Con la introducción de las estadísticas de orden superior en el cálculo del modelado AR se produce la deseada descorrelación entre voz y ruido blanco. Este efecto se basa en una propiedad aplicable a procesos Gaussianos, todos los cumulantes de orden superior a dos son idénticamente nulos. Si consideramos un p.d.f. Gaussiana o simétrica (buena aproximación de ambientes reales) y las características no-Gaussianas de la voz (principalmente en las tramas sonoras) es posible obtener un modelado AR del espectro de la voz más independiente del ruido.

A parte de estas mejoras, en este proyecto se han introducido otras técnicas, como el filtrado multietapa utilizando filtros de sustracción, peine y mediana. Esto nos proporciona la posibilidad de añadir al núcleo del filtrado de Wiener nuevas capacidades que le permitirán conseguir una mayor robustez frente a ambientes horriblemente adversos.

La introducción de un VAD en el sistema permitirá, además, una mejor estimación del ruido al poder realizar actualizaciones durante las tramas sin actividad de voz. El VAD también conseguirá hacer el sistema más robusto frente a ruidos no estacionarios.

Los más de 30 parámetros que el sistema simulador RERCOM permitirá modificar como argumentos lo convertirán en un sistema totalmente configurable, agilizando el proceso de evaluación de las diferentes variantes del algoritmo básico de Wiener.

La estructura del contenido de este Proyecto pretende introducir en un primer momento al lector en el marco del problema, para pasar posteriormente al análisis de sus posibles soluciones, como veremos a lo largo de los sucesivos capítulos y apartados que lo componen.

En el capítulo 1 se estudia la voz, no como señal aislada a procesar, sino intentando cubrir las etapas que tras ella se esconden, desde el modelo de producción de la voz hasta la subjetividad en el sistema auditivo, dando un breve repaso de las características y propiedades más relevantes de cara a un procesado digital para la mejora de la calidad y/o de la inteligibilidad del habla.

En el capítulo 2 se hace un repaso a los dos grupos básicos de técnicas de mejora de la voz, las basadas en el dominio de la frecuencia, como son el filtrado de Wiener, el filtrado de Wiener basado en la sustracción espectral y el filtro peine adaptativo, y las basadas en el dominio del tiempo como es el caso del filtrado de mediana. Analizar estas técnicas nos facilitará la comprensión del sistema de mejora de la voz que implementaremos a continuación.

En el capítulo 3 se procederá a un estudio teórico de los modelados autorregresivos aplicados sobre orden dos y mediante el uso de las estadísticas de orden superior, además de un resumen de sus propiedades. Son en total tres modelos básicos AR2, AR3 y AR4, según el orden de la estadística que utilicen. Al final de este capítulo estableceremos una comparativa entre los tres modelados.

También se recogen un conjunto de problemas derivados del uso del filtrado iterativo de Wiener y del modelo AR al aplicarlos sobre la señal de voz contaminada con ruido (capítulo 4). Por otra parte, se muestra un análisis del por qué se producen.

La teoría que hemos necesitado para la implementación del VAD híbrido de energía y distancia espectral aparece en el capítulo 5.

En el capítulo 6 realizamos una exhaustiva descripción mediante diagramas de bloques de la implementación de nuestro sistema de simulación que bautizaremos como RERCOM (REducción de Ruido en COmunicaciones Móviles).

El amplio conjunto de parámetros y coeficientes, que modifican el comportamiento del algoritmo básico, nos permite disponer de un gran número de variantes. La evaluación de éstas, así como las herramientas de medida empleadas para su comparación se pueden encontrar en el capítulo 7. También aquí se realizan las primeras pruebas para la elección óptima de un subgrupo de parámetros generales comunes a todas las variantes que posteriormente estudiaremos.

En una segunda parte del capítulo 7 realizaremos una comparativa de diferentes variantes de filtrado frente al estándar de la ETSI Advance Front-End. Esta comparativa será en primer lugar frente a medidas objetivas obtenidas al filtrar una pequeña base de datos con 9 ruidos diferentes con 3 SNRs distintas. Y en segundo lugar una comparativa de reconocimiento utilizando la base de datos TIdigits que utiliza el sistema de reconocimiento AURORA versión 008.

En el capítulo 8, explicamos las modificaciones necesarias para reducir 10 veces el tiempo de procesado del algoritmo de filtrado con los parámetros óptimos obtenidos en el capítulo 7. Así como unos breves resultados para comprobar el buen funcionamiento de estas modificaciones.

En los anexos se muestra el uso y código fuente de los tres programas implementados; simulador RERCOM versión 1.18b, programa de medidas DISTCALC versión 1.02b y la primera aproximación a DSP RERCOM\_DSP versión 0.4.

Todos los resultados obtenidos, documentación, base de datos, archivos filtrados, códigos fuente y referencias electrónicas se encuentran en el CD adjunto para una visión más global y rápida de los mismos. También se puede consultar esta información vía Internet a través de la página web del proyecto: [www.webpersonal.net/rercom](http://www.webpersonal.net/rercom).



## 1.-Modelado de la voz

Antes de buscar técnicas concretas que nos permitan subsanar los efectos de elementos perturbadores que actúan sobre la correcta recepción del mensaje, debemos plantearnos cuales son los fundamentos de la producción y audición del habla y sus características básicas, así como los principales grupos de técnicas para mejorar su calidad frente a dichos elementos perturbadores.

### 1.1.-Mecanismo de producción de la voz y su modelado

Si enfocamos la producción del habla desde un punto de vista fisiológico [Furu-89], diremos que el aire impulsado rítmicamente por los pulmones es modificado a su paso por las cuerdas vocales, produciendo una onda que es radiada por la boca.

Dicho proceso puede considerarse como un filtro, que estaría constituido por el tracto bucal, que actúa sobre la excitación del sistema, constituida por el aire proveniente de los pulmones, tal y como queda representado en la siguiente figura:

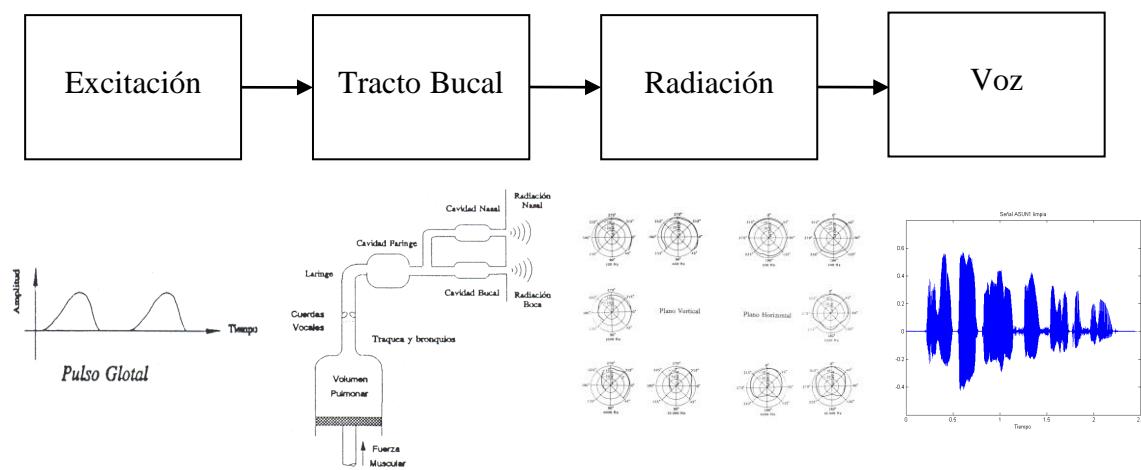


Fig. 1.1: Diagrama de bloques del proceso de producción del habla. [Jove-93]

La excitación del sistema es el llamado **Pulso Glotal**, pulso proveniente de los pulmones que pasa por la glotis con una cadencia de ~8 ms. Esta forma de onda es muy rica en armónicos espectrales, con una caída de unos 12dB/octava. En la figura 1.2(a) puede observarse el espectro de rayas de la señal pulsante excitadora, pues ha sido considerada periódica. Realmente dicha periodicidad se restringe a los regímenes

estacionarios de producción de los sonidos sonoros, por tanto dichas rayas no serían tan puras, sino que deberían tener cierta amplitud espectral.

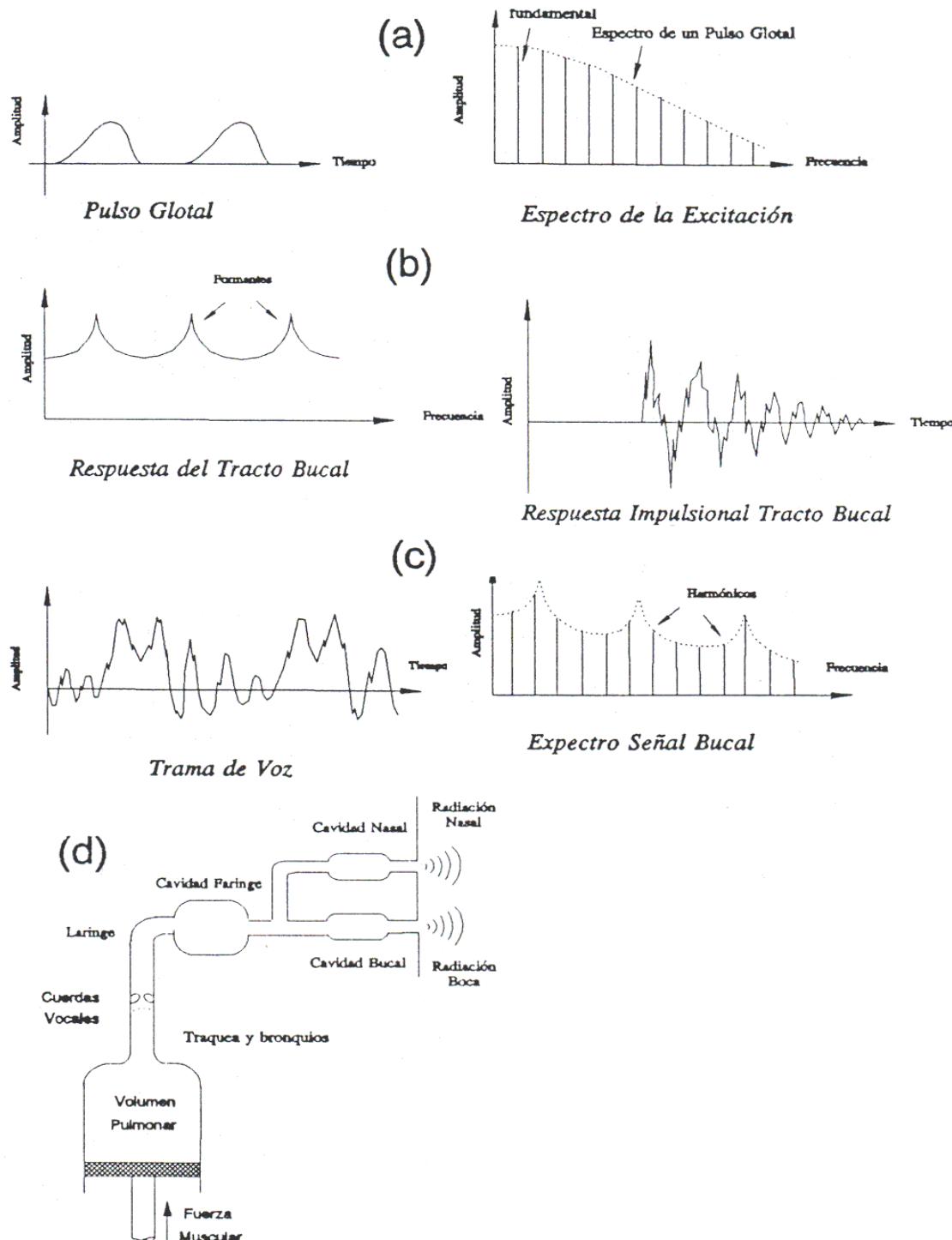


Fig. 1.2: Proceso de producción del habla:(a) Pulso Glotal y su espectro; (b) Respuesta del tracto bucal; (c) Trama de señal de voz; (d) Esquema del tracto bucal. [Jove-93]

El armónico fundamental se define como el **Pitch** de la persona. Dado el período de repetición aproximado del Pulso Glotal, el Pitch suele situarse en torno a los 80-300 Hz. El Pitch o período fundamental de la onda de excitación pseudoperiódica suele variar muy poco para una misma persona, ya que viene impuesto por las características de su tracto bucal. Tan sólo el paso del tiempo, sobre todo de niño a adulto, y posibles problemas físicos pueden dar cambios apreciables.

Esta señal de excitación pasa posteriormente por las cuerdas vocales y las cavidades bucal y nasal, modificándose su espectro.

El paso por la laringe, donde se encuentran las cuerdas vocales, actúa a modo de filtro paso banda en selección de los armónicos. Boca y nariz, cavidades resonantes, actúan también como filtros paso banda, dando a la señal final un espectro caracterizado por la presencia de unos **formantes** no fijos. Su posición depende de la articulación de las cuerdas vocales y de la forma de las cavidades vocal y nasal en el instante de fonación, variando de un fonema al siguiente y de una persona a otra. Dichos formantes se asocian a las frecuencias de resonancia de las diferentes cavidades paso banda del tracto bucal, es decir, a laringe, nariz y boca

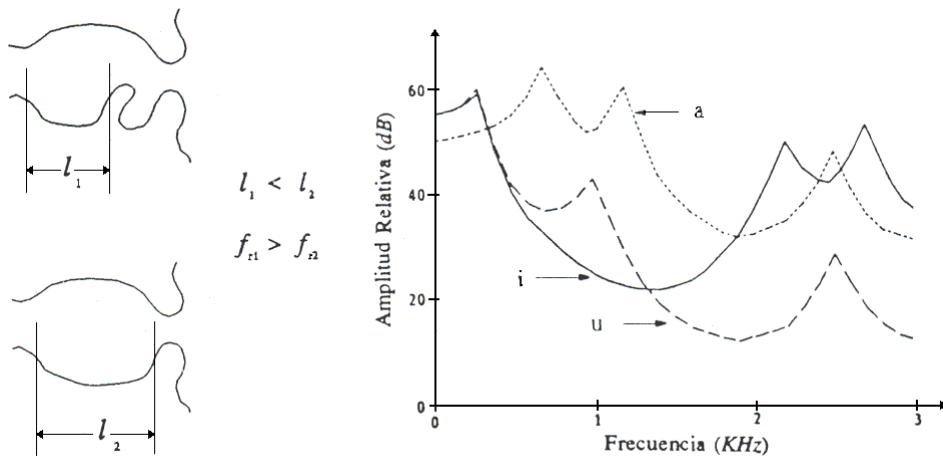


Fig.1.3: Resonancias de la boca en la boca según la posición de la lengua.  
Principales formantes de las vocales /a/, /i/, /u/. [Jove-93].

Esta modulación del aire procedente de los pulmones en las diversas cavidades atravesadas (laringe, nariz y boca) da lugar a una respuesta impulsional típica del tracto bucal, figura (1.2.b); en la cual se observan los formantes principales del espectro, polos que poseen una amplitud parecida. La suma de los filtros parciales paso banda de las cavidades dan lugar al filtro mostrado en la figura (1.2.b), un filtro de rizado de amplitud casi constante.

El número de formantes puede variar, la cavidad bucal puede verse pluralizada por la acción de la lengua, siendo la zona moduladora sobre la que podemos ejercer mayor control. La variación de la posición de la lengua da lugar a sonidos totalmente distintos, aún manteniendo el resto del tracto idénticamente gesticulado. Los fonemas /u/ y /i/ son modulados con una estructura resonante de la boca muy diferente, mostrada en la figura (1.3) Acortar la zona de la cavidad supone un aumento de la frecuencia de resonancia.

Fisiológicamente, vemos que los pulmones crean una onda pulsante de excitación hacia las cuerdas vocales, las cuales en su vibración modulan el pulso excitante que vuelve a ser modificado en las cavidades para su radiación. Este último paso no supone la variación de la forma del espectro del habla, no obstante enfatiza ciertas frecuencias. Si bien el filtrado intrínseco del tracto bucal supone un filtro paso banda que da lugar a distintos formantes, la radicación de la boca supone un filtrado paso alto de unos 6 dB/octava de ganancia.

Se observa en la figura (1.2.c) como la trama de voz generada posee un espectro de menor caída que el pulso Glotal de entrada al sistema debido a la ganancia de radiación a altas frecuencias. Ver diagramas de radiación de la figura (1.4).

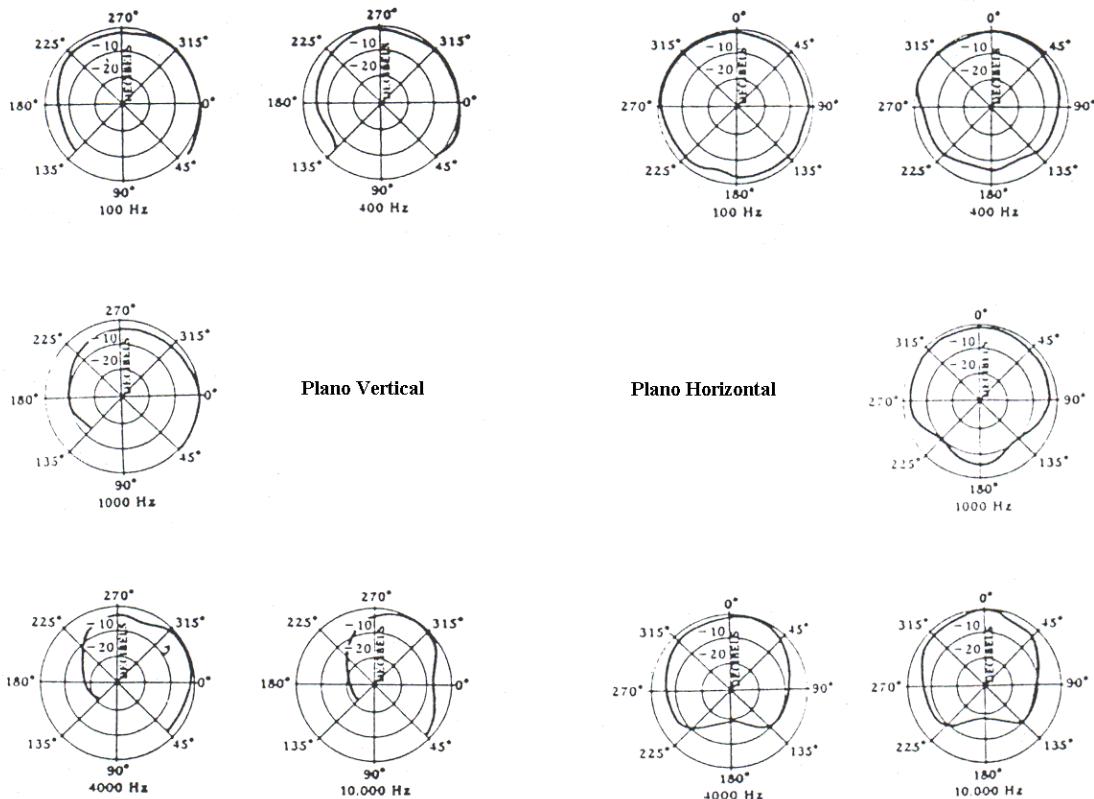


Fig.1.4: Diagramas de radiación de la boca. [Jove-93].

Dicha excitación, Pulso Glotal, toma una forma de onda de pulso de aire pseudoperiódico, modulado por el cierre de las cuerdas vocales, motivando la apariencia periódica por tramas de la voz y la concentración de su energía en frecuencias discretas múltiplos del Pitch. Todo este modelado se corresponde a unos dos tercios del habla, el resto está compuesto por sonidos aperiódicos, son los sonidos **sordos**, mientras que los anteriores eran **sonoros**.

En la figura (1.5) se muestran todos los fonemas del castellano diferenciando entre los sonidos vocálicos que son sonoros y los sonidos consonánticos que se dividen en sonoros y sordos. También se observa la clasificación de los distintos fonemas según la posición del tracto bucal, su abertura y el modo en que expulsamos el aire para la articulación de los mismos.

		Bilabial		Labiodental		Linguodental		Linguointerdental		Linguoalveolar		Linguopalatal		Linguovelar	
		sor.	son.	sor.	son.	sor.	son.	sor.	son.	sor.	son.	sor.	son.	sor.	son.
Oclusiva	p	b			t	d								k	g
Fricativa			f			θ		s			ʃ	x			
Africada										c					
Nasal		m							n		p				
Lateral									l		ʎ				
Vibrante simple									r						
Vibrante múltiple									ɾ						

VOCALES				EJEMPLOS:											
	Anterior	Central	Posterior	p	t	k	petaca	ʃ	ay	er	ŋ	caña	l	la	o
Cerrada	i		u	b	d	g	bodega	x	jota	ota	χ	lado	ɛ	la	o
Media	e		o	θ			zumo	c	chico	chico	χ	calle	ɛ	la	o
Abierta		a		f			fin	m	mamá	mamá	ɾ	pero	ɛ	la	o
				s			sol	n	no	no	ɾ	perro	ɛ	la	o

Fig.1.5: Fonemas del castellano.

Los sonidos sordos resultan del paso rápido de aire expulsado por los pulmones entre las cuerdas vocales sin provocar su vibración, pudiendo ser modelados por ruido blanco filtrado, dando lugar a las consonantes y los transitorios (p. ej.: paradas y enlace de diptongos).

Los dos grupos principales de consonantes se distinguen según la actuación de las cuerdas vocales obstruya parcial, pero permanentemente, el paso del aire a través suyo, dando lugar a los sonidos fricativos; o bien cuando las cuerdas vocales obstruyen inicialmente el tracto vocal de forma total con una posterior liberación brusca de la energía acumulada, nos referimos a las consonantes oclusivas.

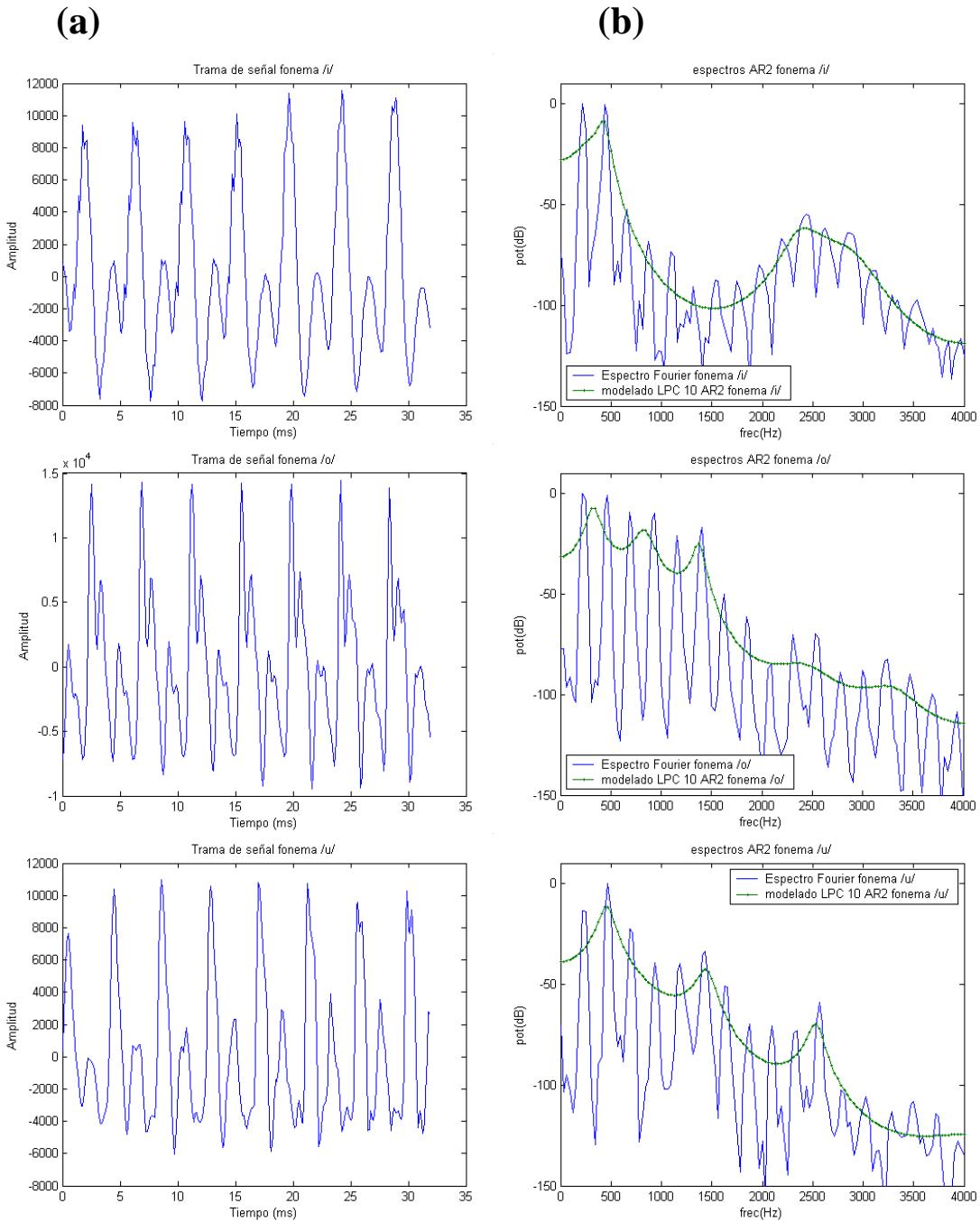


Fig.1.6: (a) Representación en el dominio temporal de los sonidos sonoros /i/, /o/ y /u/.  
 (b) Espectro y envolvente LPC de los sonidos sonoros, /i/, /o/ y /u/.  
 Pueden observarse los picos del espectro y su periodicidad.

En la generación de los fonemas consonánticos el sistema funciona de forma parecida, el tracto bucal sigue siendo un banco de filtros paso banda y la boca el sistema radiante, la diferencia radica en la señal entrante y en la pasividad de las cuerdas vocales. En esta generación de sonidos sordos puede asociarse la excitación a un ruido de banda ancha (como puede ser el Ruido Blanco Gaussiano).

Aún cuando las cuerdas vocales no actúan en el modelado de los sonidos sordos, los formantes tienen el mismo papel importante al actuar sobre una señal aleatoria a la cual dan forma espectral.

Un tercer tipo de sonidos, son los nasales. Estos tienen un tracto cuya función de transferencia se caracteriza por tener ceros y polos. Los ceros son debidos al acople de la cavidad nasal a la cavidad de la boca. No obstante en los cálculos en tiempo real la localización de los ceros resulta un factor limitador, siendo lo normal prescindir de ellos y subsanar su falta con el uso de en mayor número de polos.

Un modelado paralelo al expuesto se basa en **tubos o filtros FIR** [O'Sha-00]. Las cavidades llevan asociada una frecuencia de resonancia  $f_r$ , que depende de la longitud de la cavidad (equivalente al retardo en las líneas de transmisión, íntimamente ligado a la energía transmitida y reflejada). La excitación siguen siendo los pulsos que ascienden por la traquea hacia las cuerdas vocales, definiendo el Pitch característico de cada persona (puede ser un factor de identificación en reconocimiento de locutor).

Una longitud mayor supone un aumento del retardo de propagación de la energía, y en consecuencia una  $f_r$  menor. Esto repercute directamente en la posición de los diferentes formantes asociados a cada fonema, mientras que el pulso excitador y su periodicidad influyen en el valor comparativo del Pitch de un niño, una mujer y un hombre.

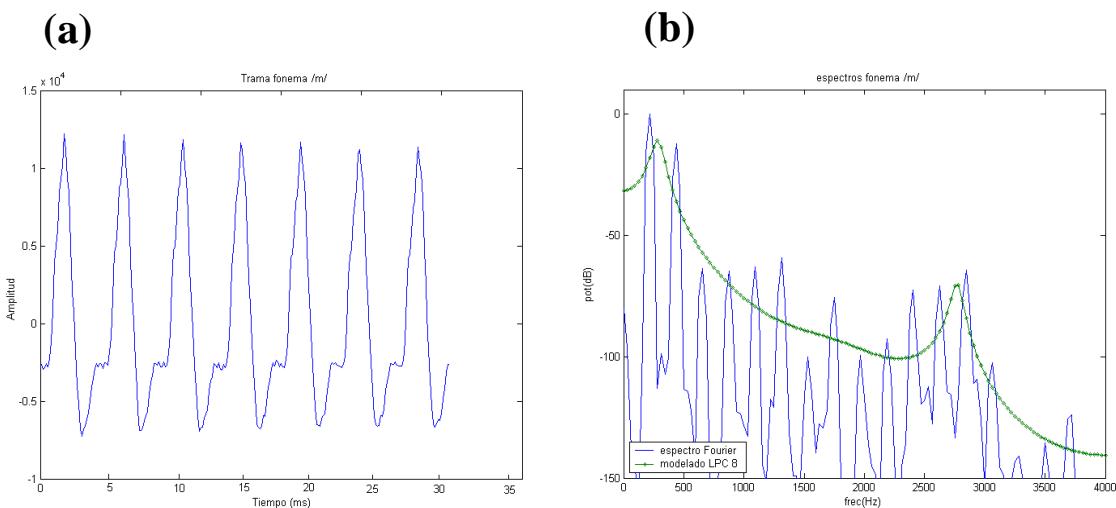


Fig.1.7a: (a) Representación en el dominio temporal de una consonante sonora, /m/.  
(b) Espectro y envolvente LPC de una consonante sonora, /m/.

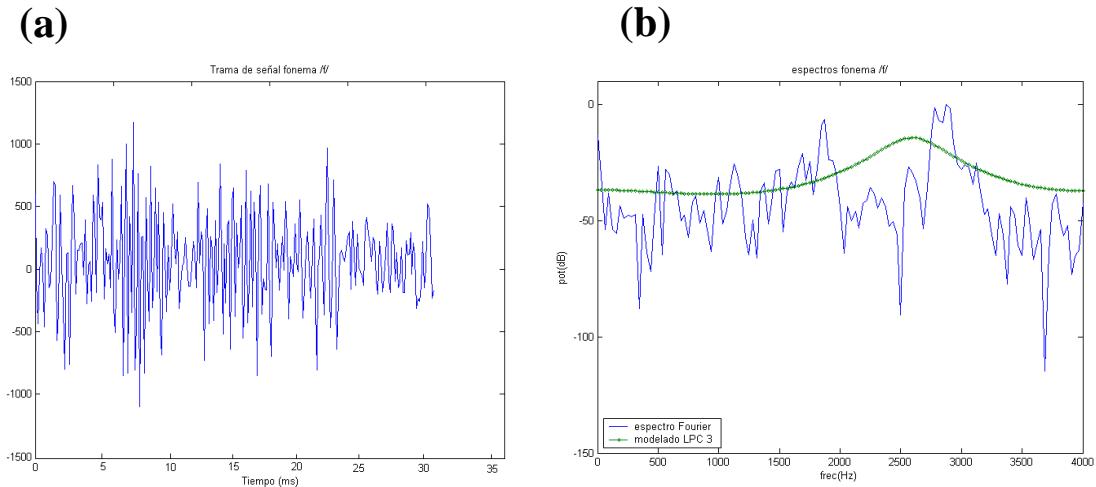


Fig.1.7b: (a) Representación en el dominio temporal de una consonante sorda, /f/. (b) Espectro y envolvente LPC de una consonante sorda, /f/.

Mediante la ecuación de presión y velocidad volumétrica del gas dentro del tracto bucal se puede modelar el paso del aire a través suyo. El tubo bucal se esquematiza como un conjunto de secciones equiespaciadas de área constante, en la práctica se limita el número a 8 ó 12 secciones. Por condiciones de continuidad sobre la presión y velocidad volumétrica entre secciones y condiciones de contorno en la Glotis y en los labios y orificios nasales se calculan los *Parcours* o coeficientes de reflexión entre las diferentes áreas fronterizas.

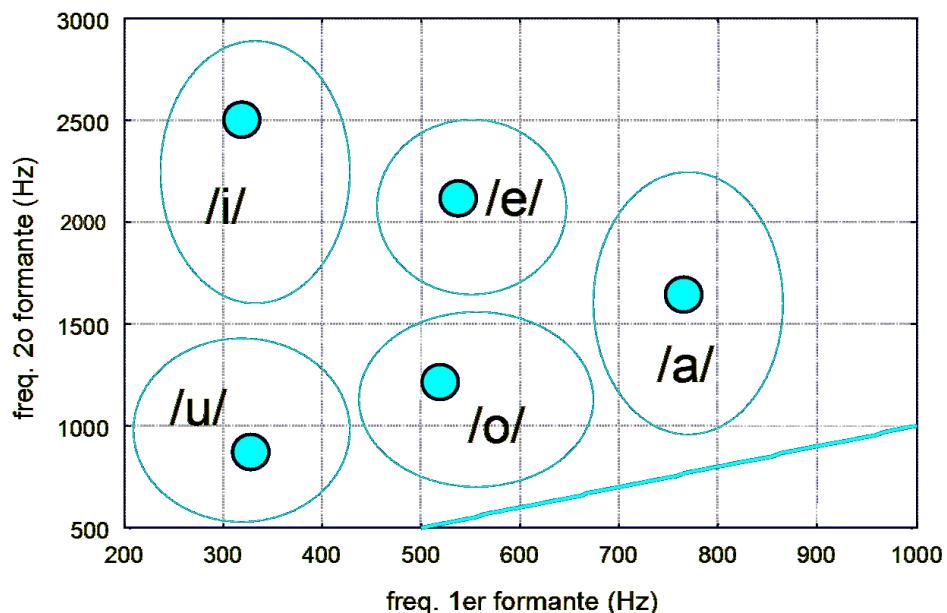


Fig.1.8: Gráfico frecuencial de los dos formantes principales de las vocales. [Torr-02]

Este modelado da lugar a una función de transferencia *Sólo polos*. No obstante la inclusión del acople de la cavidad nasal aporta los ceros a la función de transferencia,

complicando el estudio sin obtener resultados sustancialmente mejores. Un análisis que en general nos lleva a una función del tubo bucal del tipo:

$$H(z) = \frac{1}{P(z)} \quad (1.1)$$

Este modelado lineal de la producción de la voz, basado en las frecuencias de resonancia o *formantes* permite escribir la función de transferencia:

$$H(z) = \frac{g}{1 - \sum_{k=1}^p a_k \cdot z^{-k}} \quad (1.2)$$

Donde los coeficientes  $\{a_k\}$  definen el tubo generador y  $g$  es un término ajustable de ganancia.

Este modelado varía de un fonema a otro y por lo tanto debe irse actualizando trama a trama para conservar su validez. La función de transferencia en cada uno de estos períodos de tiempo puede considerarse estacionario. El valor del intervalo temporal varía entre los 10 y 30 ms.

La variación de la función de transferencia responde directamente a una variación en las cavidades resonantes y por tanto en la posición, número y ancho de banda de los *formantes*.

### **1.1.1.-Esquema general de producción de la Voz.**

Para una mejor comprensión del sistema de producción del habla intentamos agrupar en un solo modelo lineal simplificado todo el mecanismo fisiológico [O'Sha-89]. Para ello englobaremos en el esquema los dos tipos de excitación posibles y el modelado del tracto según se ha comentado (obviaremos los ceros existentes en los sonidos nasales).

Así el modelo distinguirá entre sonidos sonoros o *voiced* y sonidos sordos o *unvoiced*. Para simular la excitación sonora utilizaremos un tren de pulsos con una periodicidad y características espectrales adecuadas, lo más similares al mencionado *pulso Glotal*. Su período fundamental se corresponderá al *Pitch* deseado. La excitación de los sonidos sordos se consigue con la generación de un ruido de banda ancha.

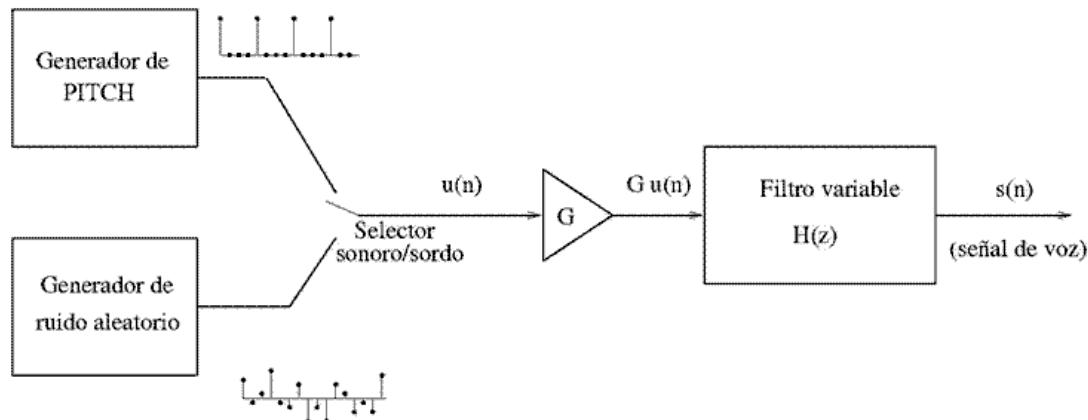


Fig.1.9: Diagrama de bloques del modelo de generación del habla.

Un conmutador accionado por el conocimiento del tipo de fonema a generar, sonido sordo o sonoro, seleccionará la entrada al sistema modulante del pulso. El bloque modulador viene caracterizado por la función de transferencia  $H(z)$ .

## 1.2.- El oído

A nosotros nos interesa analizar la voz en relación con el oído humano y su percepción de los sonidos [Furu-89] [O'Sha-89] [O'Sha-00]. Para ello haremos un breve repaso de este sentido: el oído.

Unos primeros rasgos que dan idea del potencial del oído son su capacidad de percepción en un margen de 20 a 20.000 Hz, en un rango de presiones de 20  $\mu\text{Pa}$  a 20 Pa, asimilando desplazamientos en el tímpano mínimos del orden de  $10^{-8}$  mm.

El estudio del oído se realiza de acuerdo con sus tres zonas bien diferenciadas:

- El **oído externo** formado por un colector de sonidos, el pabellón auditivo externo, y el conducto auditivo externo que actúa a modo de protección a sonidos elevados y es además una cavidad resonante que aumenta la sensibilidad auditiva en la zona de los ~3.000 Hz. El conducto de 0.7cm de diámetro y una longitud de 2.5cm (equivalentes a  $\lambda/4$  de la onda resonante) nos da una  $f_r$  donde se produce el énfasis frecuencial auditivo. Como característica adicional del oído externo debe mencionarse la difracción de las altas frecuencias que las corrugaciones de la oreja ejercen sobre los

sonidos captados.

- El **tímpano** está en el paso del oído externo al oído medio, un paso de presión sonora a vibración mecánica, en el que la cadena de huesecillos (martillo, yunque y estribo) protegen de sonidos elevados (reflejo acústico) y aumentan el nivel de presión sonora en la ventana oval respecto al nivel en el tímpano. Este aumento se basa en el efecto palanca de la cadena de huesos y en la diferencia de superficie entre tímpano y ventana oval. Esta ganancia de unos 30dB sirve para compensar los 35dB de perdida al pasar del aire del oído medio al ambiente líquido del oído interno.
- Las **ventanas redonda y oval** comunican con el oído interno compuesto por los canales semicirculares del sentido del equilibrio y el caracol (tubo de sección constante lleno de líquido).

El tubo que compone el caracol está dividido en dos rampas separadas por la cóclea. En ambas semisecciones se desplaza una onda longitudinal en sentido opuesto que da lugar a una onda transversal que pone en movimiento las células ciliadas. Al tocarse generan impulsos eléctricos que son recogidos por unas 30.000 terminaciones nerviosas.

Dos características son propias de este proceso, el tiempo refractario (tiempo para que una célula ciliada vuelva a su posición de equilibrio y esté en condiciones de volver a ser excitada, ~1 ms) y el umbral diferencial de frecuencia, unos 3 ciclos sobre 1.000 (poder de distinción entre dos frecuencias).

### **1.3.-Características acústicas y estadísticas de la voz y el oído**

Desde el punto de vista lingüístico, el habla puede verse como un conjunto de 5.000 a 10.000 palabras de uso común, o como un conjunto de 2.000 a 3.000 sílabas, o simplemente como 40 ó 50 fonemas, todo ello teniendo en cuenta las variaciones a veces sustanciales entre idiomas (el chino se compone tan sólo de 300 sílabas, o el inglés que no admite el uso de la división por sílabas). Existen otras posibles divisiones fonéticas muy utilizadas, puntos importantes en reconocimiento y síntesis de voz, pero más superfluo en nuestro caso. Para nosotros la división básica podría asociarse al fonema.

Existen dos características propias del oído interno que nos son útiles, la división en bandas y el enmascaramiento. Características útiles tanto en la compresión de información como en la comprensión de los test de audición.

La zona de células excitadas, es decir, la mayor o menor extensión de células excitadas partiendo de la ventana oval depende de la frecuencia percibida. A mayor frecuencia menor tramo de células en movimiento, dándose siempre la máxima excitación en la zona más lejana (ver figura 1.10).

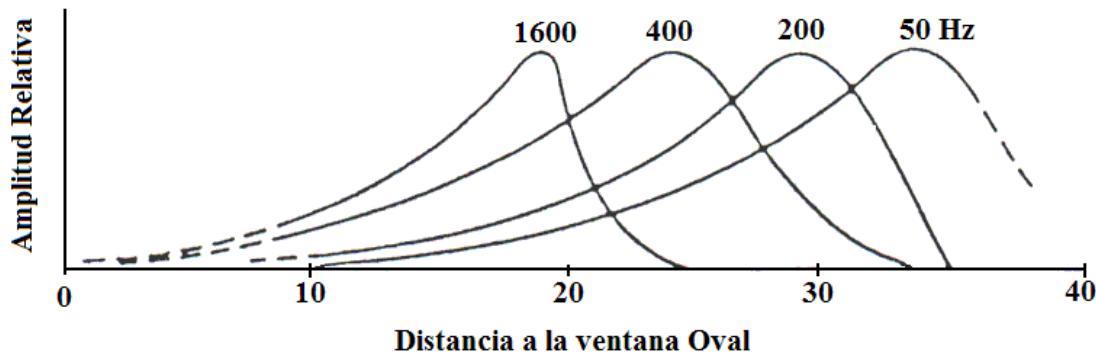


Fig.1.10: Desplazamiento de amplitudes en la membrana basilar en función de la distancia a la ventana oval para diferentes frecuencias. [Jove-93].

Este punto de máxima excitación está relacionado con la tonalidad escuchada. La sonoridad se asocia al número de células excitadas.

El enmascaramiento (figura 1.10) se basa en la división frecuencial del sonido, para una mayor comprensión se explica el caso de dos tonos. Debido a la caída de 150 dB tras el máximo, un tono de baja frecuencia podrá enmascarar parcial o totalmente un tono de alta frecuencia. La pendiente de subida es de unos 25dB por década, lo que posibilita cubrir un tono de alta frecuencia, insignificante tras su máximo (caso c, figura (1.11)).

En la figura 1.11 se observan esquemáticamente cuatro casos sobre la membrana basilar. En el caso (a) el tono B se superpone parcialmente al tono A, enmascarándolo ligeramente. En (b) el enmascaramiento es apreciable y en (c), aún siendo el tono B de frecuencia muy inferior, su elevado nivel encubre al tono A. (d) nos muestra la dificultad de enmascarar un tono de baja frecuencia B por un tono A de alta frecuencia.

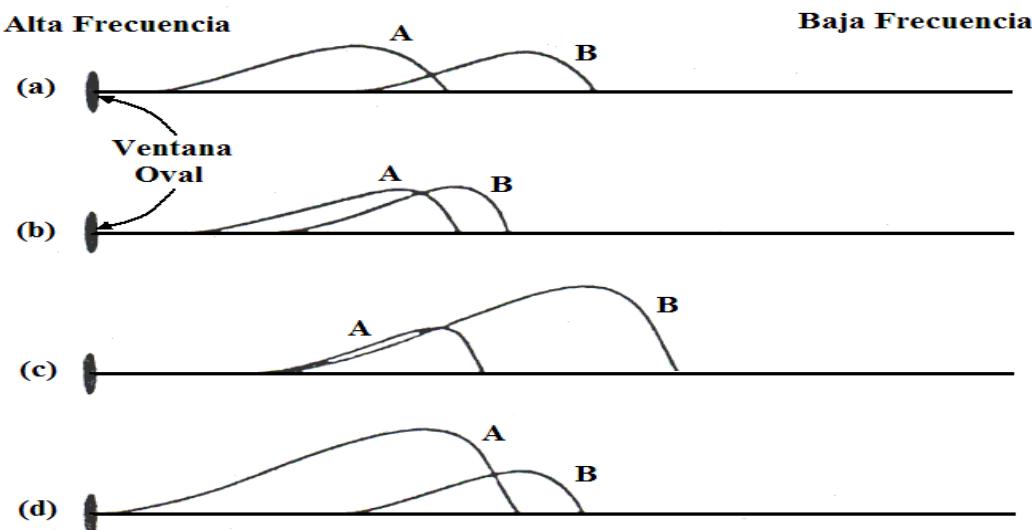


Fig.1.11: Respuesta simplificada de la membrana Basilar para dos tonos A y B. [Jove-93].

Se desprenden tres conclusiones:

- El ruido de banda estrecha provoca un enmascaramiento mayor que un tono de la misma intensidad y de frecuencia igual a la frecuencia central del ruido.
- A medida que el nivel de la señal enmascarante aumenta, también aumenta la banda de frecuencias sobre las cuales ejerce un efecto enmascarante.
- Las frecuencias superiores a la frecuencia central de la señal enmascarante son enmascaradas más fácilmente que las inferiores.

Al igual que existe un enmascaramiento simultáneo, existe el enmascaramiento temporal. Una señal de baja frecuencia y nivel alto de pronto desaparece y aparece otro tono de frecuencia mayor, si el retardo entre ambos tonos es inferior a 200 ms. se produce la falta de percepción acústica.

El otro punto a describir son las bandas críticas, el oído funciona como un analizador de espectros con un banco de 24 filtros paso banda con un ancho de banda de un tercio de octava.

Imaginemos ruido blanco de banda estrecha que está enmascarando un tono a la frecuencia central del ruido. Incrementamos el ancho de banda del ruido, incrementamos por tanto la potencia de ruido. Seguimos aumentando y cada vez aumenta más el enmascaramiento del ruido. Pero al llegar a un cierto ancho de banda un

aumento del mismo no repercute en un mayor enmascaramiento del tono. Los límites de esta banda es lo que llamamos banda crítica y el ancho de banda crítico es de  $\frac{1}{3}$  de octava. A nivel de membrana basilar esta banda crítica ocupa 1.3 mm. Esto no quiere decir que las frecuencias centrales sean fijas.

El concepto de banda crítica está íntimamente relacionado con los filtros paso banda, al igual que si superamos la banda crítica enmascaramos igual al tono deseado, si estamos en el límite de la banda crítica y disminuimos el ancho de banda perturbador, cada vez enmascaramos menos.

Se observa que el oído actúa subjetivamente, tanto en frecuencia como en nivel de presión acústica, la sensación percibida no tiene porque estar directamente relacionada con la tonalidad o sonoridad que apreciamos nosotros.

Esto lleva a definir las curvas isofónicas o de *Robinson-Dadson* que corresponden a la sensación auditiva creada por distintos tonos de diferentes frecuencias y diversos niveles (a cada curva se le asigna un número). Cada curva produce la misma sensación de percepción.

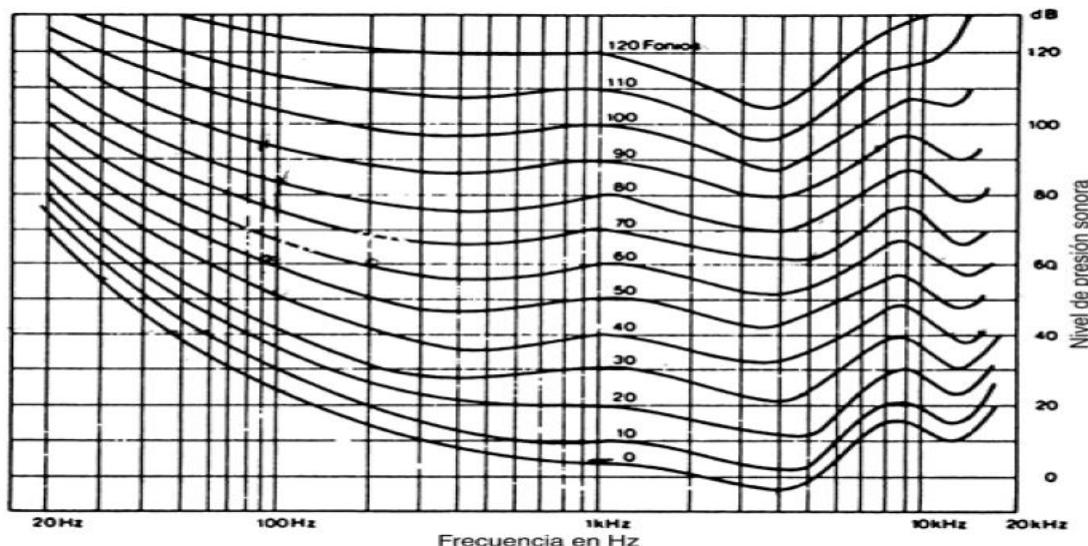


Fig.1.12: Curvas Isófonas.

Este número es el nivel del tono de 1 Khz. que se oye igual que los otros tonos. La curva indica el nivel de estos tonos.

A cada curva de nivel se le asocia un nivel de sonoridad, el cual coincide con el valor numérico del nivel de presión de un tono de 1 Khz. que tenga la misma sonoridad que la

señal, y su unidad se denomina *Phon*.

De las curvas isofónicas de la figura 1.12 debemos observar:

- Para niveles bajos de señal, el oído se comporta de forma no lineal, debemos aumentar mucho los niveles para tener la misma sonoridad tanto a bajas como a altas frecuencias.
- El umbral de percepción es de  $20 \mu\text{Pa}$  a  $1 \text{ KHz}$ , aunque el valor varía en función de la frecuencia, pudiendo ser superior o inferior.
- A medida que aumenta el nivel, la tendencia del comportamiento del oído es igualar la curva de percepción hacia la linealidad (curva más plana).

Esta caracterización del sistema productivo y auditivo del habla nos da idea de las zonas más delicadas y las más superfluas en el procesado de la voz. Es importante tener en cuenta la naturaleza sonora o sorda del sonido, su Pitch y el filtro modelador del tubo  $H(z)$ , siendo las demás características de generación menos importantes.

En relación al oído no debe olvidarse la importancia de la fase en los sonidos sordos y en transitorios, siendo menos influyente su papel en fonemas sonoros.

Más importante e indeseable resulta la distorsión que un procesado de señal pueda introducir, el oído es muy sensible a este efecto.

Finalmente daremos algunos valores de interés:

El Pitch medio asociable a las personas se puede suponer de  $110 \text{ Hz}$  en el hombre,  $220 \text{ Hz}$  en la mujer y  $300 \text{ Hz}$  en los niños.

La potencia asociable a la voz en media es de  $10 \mu\text{W}$ , estando concentrada la mayor parte de la energía en el margen de  $500$  a  $2.000 \text{ Hz}$ . La inteligibilidad de la voz tiene mayor peso en la zona de  $1$  a  $4 \text{ KHz}$ . Esta concentración de la energía por debajo de los  $4 \text{ KHz}$  permite limitar el ancho de banda en aquellas aplicaciones que no requieran elevada calidad (p. ej.: telefonía). Por encima de los  $4 \text{ KHz}$  la energía del espectro es pequeña y tan sólo contribuye a mejorar la calidad.

Por otra parte la gran redundancia de información que contiene la señal de voz permite comprender el mensaje con tan solo parte de él, debido a la lenta variación del espectro y su gran periodicidad que permite una descripción, como hemos visto, con pocos parámetros.

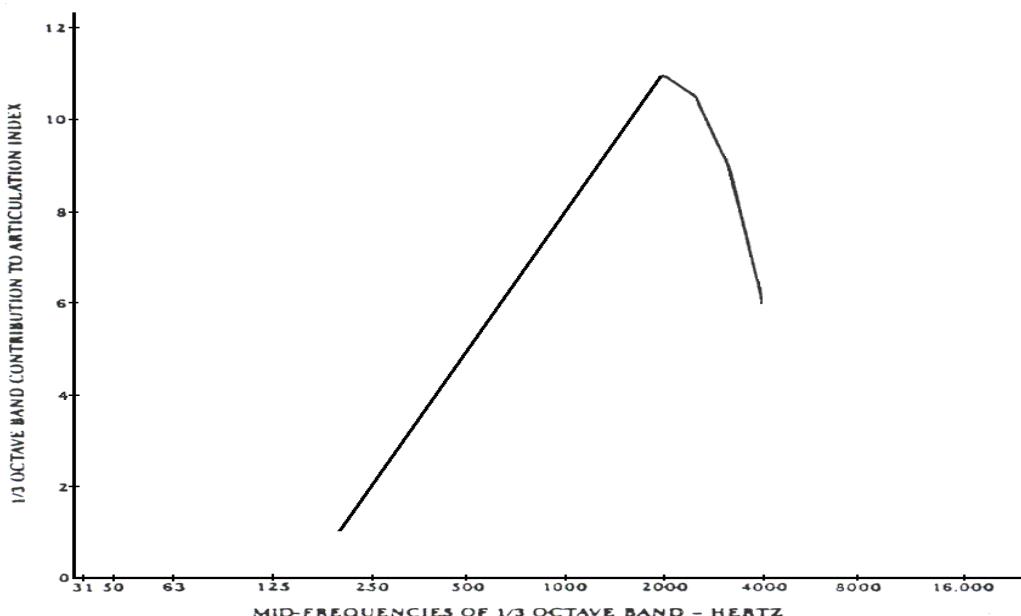


Fig.1.13: Contribución a la inteligibilidad por bandas frecuenciales de  $\frac{1}{3}$  de octava.

La radiación de la boca es omnidiireccional a baja frecuencia cuando  $d \ll \lambda$ , pero a alta frecuencia aumenta la directividad enfatizando en la dirección frontal la alta frecuencia a modo de filtro paso alto.

Un matiz propio de la voz, provocado por el espectro de la señal excitadora, los formantes y los transitorios de ataque y extinción, es el timbre, factor diferenciador de las personas.

Los algoritmos de mejora y codificación concentran su esfuerzo en no modificar los picos del espectro de amplitud; se basan en explotar la mayor sensibilidad del sistema auditivo a la presencia de energía que a su ausencia, dejando más de lado la fase o la energía en frecuencias débiles.

Las partes sonoras, zonas de gran amplitud temporal y concentración de la energía a bajas frecuencia, son más importantes que los fonemas sordos para garantizar la calidad del sistema. La mayoría de algoritmos tiende a mejorar las partes periódicas

correspondientes a voiced.

La representación de las amplitudes espectrales en las frecuencias de los armónicos en los tres primeros formantes es esencial.

Algunos test de audición demuestran que los sonidos sordos y débiles son menos importantes frente a los sonoros. No obstante la tarea más difícil reside en resolver la supresión del ruido en las zonas con menor redundancia, zonas de poco nivel energético y las zonas de transición.

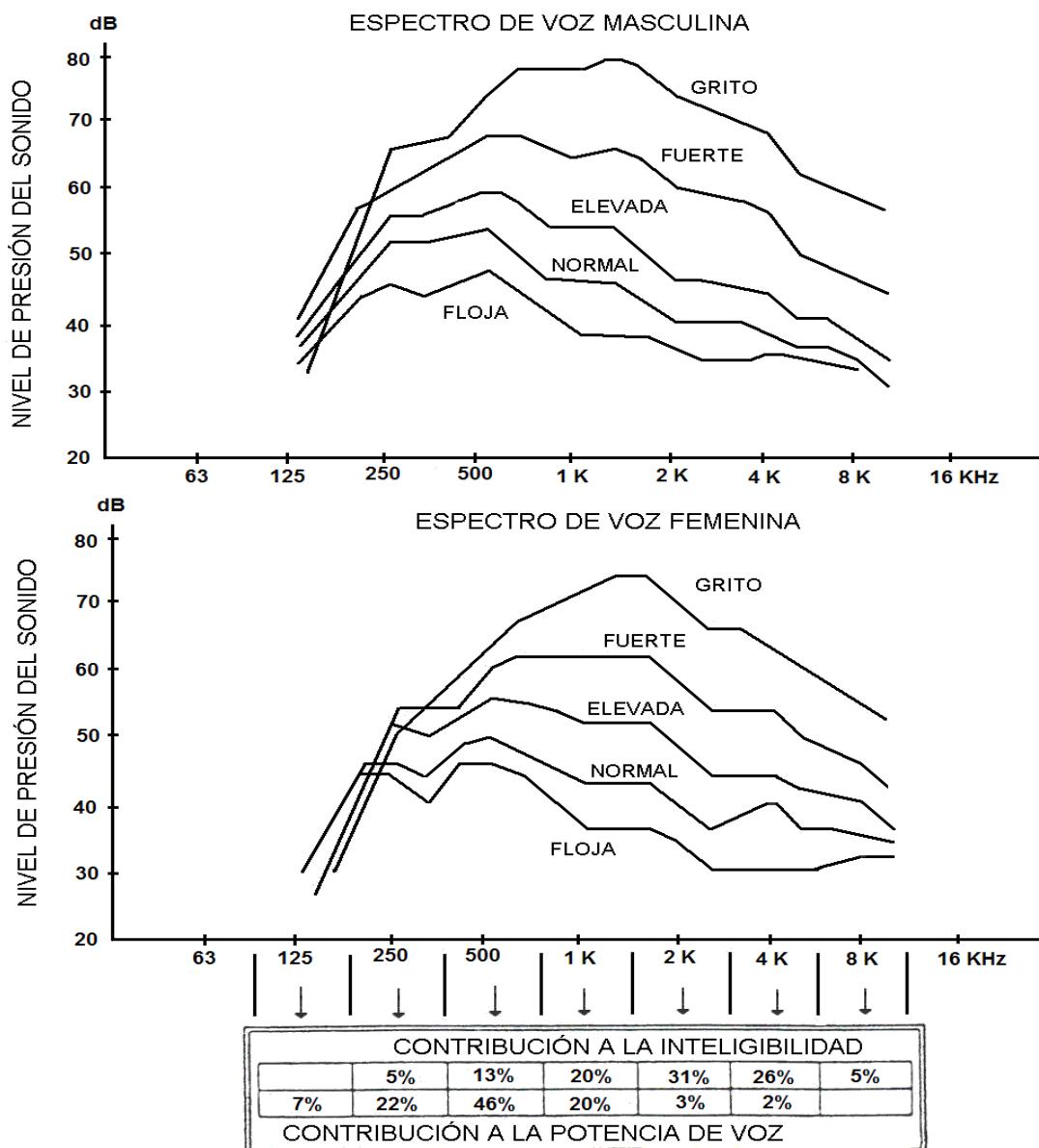


Fig.1.14: Nivel de presión sonora para la voz de hombre y mujer. Contribuciones por márgenes frecuenciales a la inteligibilidad y al nivel de presión acústica. [Jove-93].



## 2.- Técnicas básicas de eliminación de ruido en señales de voz

Una vez analizado el proceso de producción del habla y de la percepción de la voz por parte de nuestros oídos, ahora nos planteamos el desarrollo de un correcto procesado de la voz que nos permita mejorar su inteligibilidad en ambientes adversos [O'Sha-89], [Lim-79].

El efecto de los agentes perturbadores debe minimizarse para preservar los aspectos fundamentales del mensaje de cara a su comprensión. Ya que nuestro oído discrimina entre las diferentes características de la voz, otorgándole mayor importancia a algunas, éstas son las que debemos conservar lo más fielmente posible durante el procesado de la voz.

Respecto al rango de frecuencias de la señal de voz, nos debemos centrar, principalmente, en la zona hasta los 4 KHz en la que se concentra la energía y las frecuencias con principal contribución a la inteligibilidad. Por encima de los 4 KHz la fracción de energía presente es muy pequeña respecto al total. En el sistema que hemos implementado consideraremos una frecuencia de muestreo fija de 8 KHz, esto supone un filtrado con frecuencia de corte a 3.4 KHz, eliminando toda la señal y ruido por encima de esta frecuencia.

Otro aspecto importante para conseguir la inteligibilidad del mensaje recibido es la correcta captación del módulo del espectro por tramas de la señal vocal. En cambio, aunque el oído es sensible a la fase, ésta es relativamente poco importante en los sonidos sonoros y su importancia sólo es remarcable en transitorios y consonantes.

Un factor que debemos evitar en el procesado es la distorsión de la voz. Para entender los resultados obtenidos debemos poner especial cuidado en inspeccionar los formantes, su forma y su posición.

Además debemos tener en cuenta, tal y como vimos en el capítulo anterior, que el oído tiene la capacidad de enmascarar una señal con otra. Por ejemplo la introducción de un ruido de banda ancha puede esconder la presencia de otro ruido o señal de banda estrecha en función de su nivel.

A la hora de desarrollar una técnica concreta de procesado hay que tener en cuenta que ésta puede ser efectiva frente a un tipo de ruido o perturbación pero inútil frente al resto de perturbaciones. Nosotros nos enfrentaremos al caso de señal más ruido aditivo, para el cual existen, así mismo, múltiples sistemas de mejora, de los cuales explicaremos los más importantes a continuación.

## 2.1.- Técnicas basadas en el dominio de la frecuencia

En este punto trataremos las principales técnicas de procesado de la voz basadas en el dominio de la frecuencia, es decir, que se basan en características frecuenciales de la señal de entrada al sistema. Entre las cuales veremos el filtrado de Wiener, el filtrado de Wiener basado en la sustracción espectral y el filtro peine adaptativo.

### 2.1.1.- Filtrado de Wiener

Si tenemos un ruido aditivo  $d'(n)$  incorrelado con la voz que contamina a una señal de voz limpia  $s'(n)$  entonces la señal recibida es:

$$y'(n) = s'(n) + d'(n) \quad (2.1)$$

Con lo cual:

$$P'_y(\omega) = P'_s(\omega) + P'_d(\omega) \quad (2.2)$$

Donde  $P'_y$ ,  $P'_s$  y  $P'_d$  son las densidades espectrales de potencia de la señal contaminada, de la señal de voz limpia (señal original) y del ruido aditivo que contamina la señal respectivamente. Recordemos que en nuestro caso tan sólo disponemos directamente de  $y(n)$ .

La señal de voz es no estacionaria durante períodos de tiempo demasiado largos; además, aunque lo fuese, un número demasiado grande de muestras podría introducir un elevado retardo y se perdería otro factor deseable: el procesado en tiempo real.

Procederemos, por tanto, a un enventanado de la señal con una ventana  $w(n)$  de duración igual a la longitud de trama básica de análisis que elijamos de acuerdo con el

criterio de estacionariedad de la voz. Deberemos ir desplazando la ventana sobre la señal y tras el procesado recomponerla sin distorsionarla.

Si  $s(n)$ ,  $d(n)$  e  $y(n)$  son ahora las señales enventanadas:

$$\begin{aligned}s(n) &= s'(n) \cdot w(n) \\ d(n) &= d'(n) \cdot w(n) \\ y(n) &= y'(n) \cdot w(n)\end{aligned}\tag{2.3}$$

y  $P_s(\omega)$ ,  $P_d(\omega)$ ,  $P_y(\omega)$  sus respectivos espectros de potencia, entonces:

$$y(n) = s(n) + d(n)\tag{2.4}$$

La idea del filtrado de Wiener ([O'Sha-89] [Lim-79]) es minimizar el error cuadrático medio entre la señal filtrada y la señal original, lo que nos conduce a un filtro de la forma:

$$H(\omega) = \frac{P_s(\omega)}{P_s(\omega) + P_d(\omega)}\tag{2.5}$$

Se trata de un filtro no causal, obtenido en el supuesto que  $d(n)$  y  $s(n)$  sean procesos estacionarios incorrelados.

Podemos obtener una aproximación de las densidades espectrales de potencia haciendo un promediado de espectros de varias tramas de señal, construyendo un filtro tal como:

$$H(\omega) = \frac{E\{|S(\omega)|^2\}}{E\{|S(\omega)|^2\} + E\{|D(\omega)|^2\}}\tag{2.6}$$

Para la obtención de  $E\{|D(\omega)|^2\}$  promediaremos los espectros de las tramas de silencio, es decir, en ausencia de voz. Por otro lado, podemos obtener  $E\{|S(\omega)|^2\}$  realizando la sustracción espectral de la estimación del ruido a la señal contaminada,  $y(n)$ , que tenemos a la entrada del sistema:

$$E\{|S(\omega)|^2\} = E\{|Y(\omega)|^2\} - E\{|D(\omega)|^2\}\tag{2.7}$$

En este caso, por simplicidad, hemos aproximado las densidades espectrales de potencia por un promediado de los periodogramas de las tramas anteriores, esto en el

caso de la señal de voz no es demasiado efectivo, porque tal y como hemos visto, la señal de voz no es estacionaria y su espectro varía trama a trama. Así pues, sería más lógico pensar en un sistema que adapte el filtro para cada trama, al espectro concreto de cada una de ellas.

La búsqueda de  $\mathbf{P}_s(\omega)$  la realizaremos basándonos en el modelo de producción de la voz que vimos en el capítulo anterior y donde llegábamos a la siguiente ecuación:

$$H'(z) = \frac{g}{1 - \sum_{k=1}^p a_k \cdot z^{-k}} \quad (2.8)$$

De esta manera el problema de la estimación de  $\mathbf{P}_s(\omega)$  se traslada al del cálculo de los coeficientes  $a_k$ ,  $g$  y el valor de  $P$  (orden del modelo) que se adapte bien a nuestras necesidades. Todo esto conlleva un estudio detallado que se aborda en el capítulo 3.

La señal estimada a la salida del sistema será el resultado de pasar por el filtro de Wiener  $H(\omega)$  la señal contaminada con ruido  $y(n)$ :

$$\hat{S}(\omega) = Y(\omega) \cdot H(\omega) \quad (2.9)$$

Una expresión del filtro de Wiener más genérica, donde se han introducido los parámetros  $\delta$  y  $\beta$ , sería:

$$H(\omega) = \left( \frac{\mathbf{P}_s(\omega)}{\mathbf{P}_s(\omega) + \beta \cdot \mathbf{P}_d(\omega)} \right)^{\delta} \quad (2.10)$$

En nuestro sistema obtendremos diferentes resultados de ir variando  $\delta$  y  $\beta$ , tratando de obtener sus valores óptimos.

### 2.1.2.-Filtrado de Wiener basado en la sustracción espectral.

Este método consiste en realizar una estimación, tan fiable como sea posible, del espectro del ruido que contamina la señal. Para esto se utilizan o bien los intervalos de silencio, o bien debemos estar en condiciones de obtener por otro canal la señal de ruido. Siempre debemos estar seguros de utilizar sólo ruido para estimar su espectro.

Tal y como se explicó en el punto anterior, la señal de entrada al sistema, después de enventanar es:

$$y(n) = s(n) + d(n) \quad (2.11)$$

Donde  $y(n)$  es la señal contaminada con ruido enventanada,  $s(n)$  la señal original enventanada y  $d(n)$  el ruido aditivo enventanado. Y por tanto:

$$|Y(\omega)|^2 = |S(\omega)|^2 + |D(\omega)|^2 + S(\omega) \cdot D^*(\omega) + S^*(\omega) \cdot D(\omega) \quad (2.12)$$

$|Y(\omega)|^2$  se obtiene directamente y  $S(\omega) \cdot D^*(\omega)$ ,  $|D(\omega)|^2$  y  $S^*(\omega) \cdot D(\omega)$  se aproximan por  $E[S(\omega) \cdot D^*(\omega)]$ ,  $E[|D(\omega)|^2]$  y  $E[S^*(\omega) \cdot D(\omega)]$  respectivamente.

$D^*(\omega)$  y  $S^*(\omega)$  representan el complejo conjugado de  $D(\omega)$  y  $S(\omega)$ . Si  $d(n)$  es ruido incorrelado con la señal  $s(n)$  entonces:

$$E[S(\omega) \cdot D^*(\omega)] = E[S^*(\omega) \cdot D(\omega)] = 0 \quad (2.13)$$

Y queda:

$$|\hat{S}(\omega)|^2 = |Y(\omega)|^2 - E\{|D(\omega)|^2\} \quad (2.14)$$

Donde  $\hat{S}(\omega)$  es el espectro estimado y  $E\{|D(\omega)|^2\}$  se obtiene por conocimiento de las propiedades de  $d'(n)$  o por procesado de los intervalos de silencio de voz.

La  $|\hat{S}(\omega)|^2$  obtenida en (2.14) no nos garantiza que no sea negativa. En algunos algoritmos se toma el valor absoluto, mientras que en otros se da un valor positivo muy pequeño a  $|\hat{S}(\omega)|^2$  si  $|Y(\omega)|^2$  es inferior a  $E\{|D(\omega)|^2\}$ .

Tal y como vimos en el punto anterior el filtro de Wiener responde a la siguiente expresión:

$$H(\omega) = \frac{E\{|S(\omega)|^2\}}{E\{|S(\omega)|^2\} + E\{|D(\omega)|^2\}}$$

Si sustituimos  $E\{|S(\omega)|^2\}$  por la expresión obtenida para  $|\hat{S}(\omega)|^2$  tenemos:

$$H(\omega) = \frac{|Y(\omega)|^2 - E\{|D(\omega)|^2\}}{|Y(\omega)|^2 - E\{|D(\omega)|^2\} + E\{|D(\omega)|^2\}} = \frac{P_y(\omega) - P_d(\omega)}{P_y(\omega)} \quad (2.15)$$

Que corresponde a la expresión del filtro de Wiener basado en la sustracción espectral, al que también se le pueden añadir los parámetros  $\delta$  y  $\beta$ , obteniendo:

$$H(\omega) = \left( \frac{P_y(\omega) - \beta \cdot P_d(\omega)}{P_y(\omega)} \right)^{\delta} \quad (2.16)$$

### 2.1.3.- Filtro en peine adaptativo.

Generar un filtro peine que deje pasar la frecuencia fundamental y todos sus armónicos en banda, eliminando las componentes frecuenciales intercaladas, lleva a una mejora de la señal de voz.

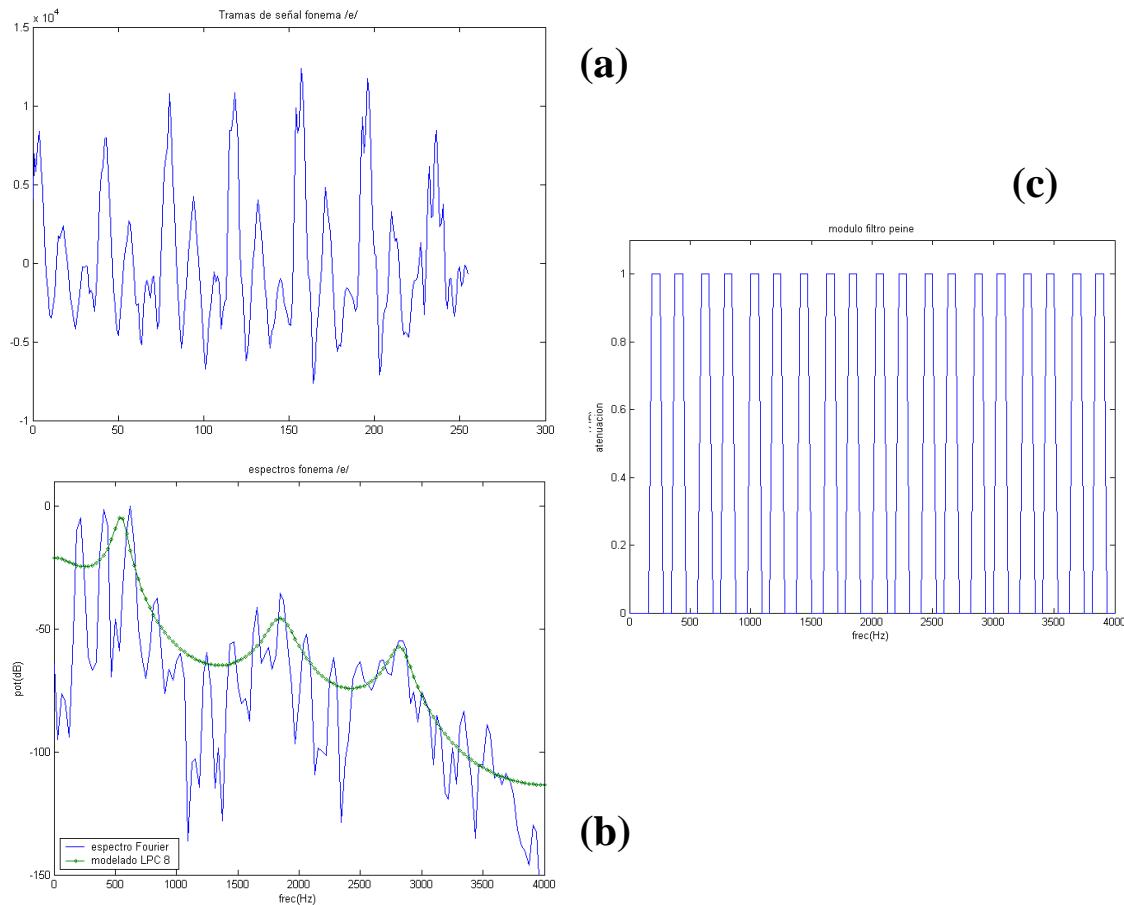


Fig.2.1: (a) Señal periódica en el tiempo. (b) Espectro de una señal periódica. (c) Respuesta frecuencial del filtro peine.

Esta técnica utiliza la propiedad de periodicidad de los sonidos sonoros (voiced): la energía de una señal periódica se concentra en los armónicos de la misma (figura 2.1(a)). Mientras la señal interferente suele poseer energía en un rango frecuencial más amplio.

Estos algoritmos carecen de eficacia cuando la estimación precisa del Pitch se ve dificultada al disponer de una señal fuertemente contaminada con ruido.

Aunque la voz sea sólo aproximadamente periódica por tramas el método es aplicable. Sin embargo, a la hora de buscar el valor  $T$  de separación entre los distintos armónicos, necesitamos que éste se adapte local y globalmente a la señal, esto lleva a modificar el algoritmo de respuesta de Shields que utiliza un solo espaciado  $T$ , frente al de Frazier que lo desglosa en los  $T_i$  adaptados a las variaciones locales del Pitch en las zonas sonoras de la voz, tal y como se observa en la figura 2.2.

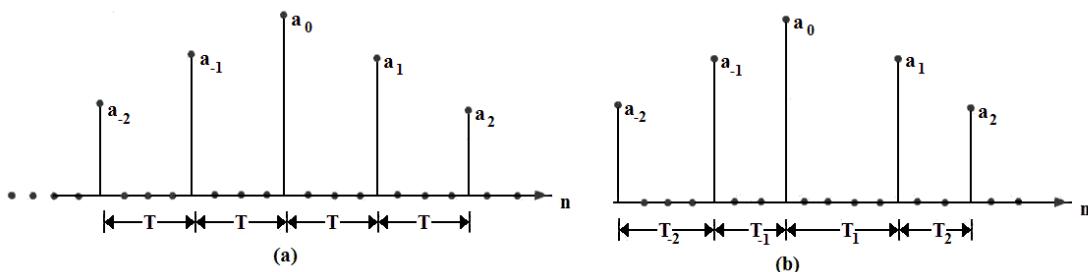


Fig.2.2: (a) Respuesta del filtro adaptativo de Shields.  
(b) Respuesta del filtro adaptativo de Frazier. [Jove-93]

Este método requiere de un procesado en paralelo que estime el Pitch para generar el filtro adaptativo. Este procesado puede ser necesario y muy útil en ambientes donde coexisten varias conversaciones.

## 2.2.- Técnicas basadas en el dominio del tiempo.

En este punto trataremos las principales técnicas de procesado de la voz basadas en el dominio del tiempo, es decir, que se basan en características temporales de la señal de entrada al sistema. Entre las cuales veremos el filtrado de mediana.

### 2.2.1.- Filtro de mediana.

La función del filtro de mediana en procesado de voz es eliminar el ruido residual que haya podido quedar de etapas de filtrado anteriores, así como la eliminación de ruido musical generado en el filtrado. También tiene como objetivo eliminar el ruido de tipo impulsional que contenga la señal.

El funcionamiento del filtro de mediana se muestra en la siguiente figura:

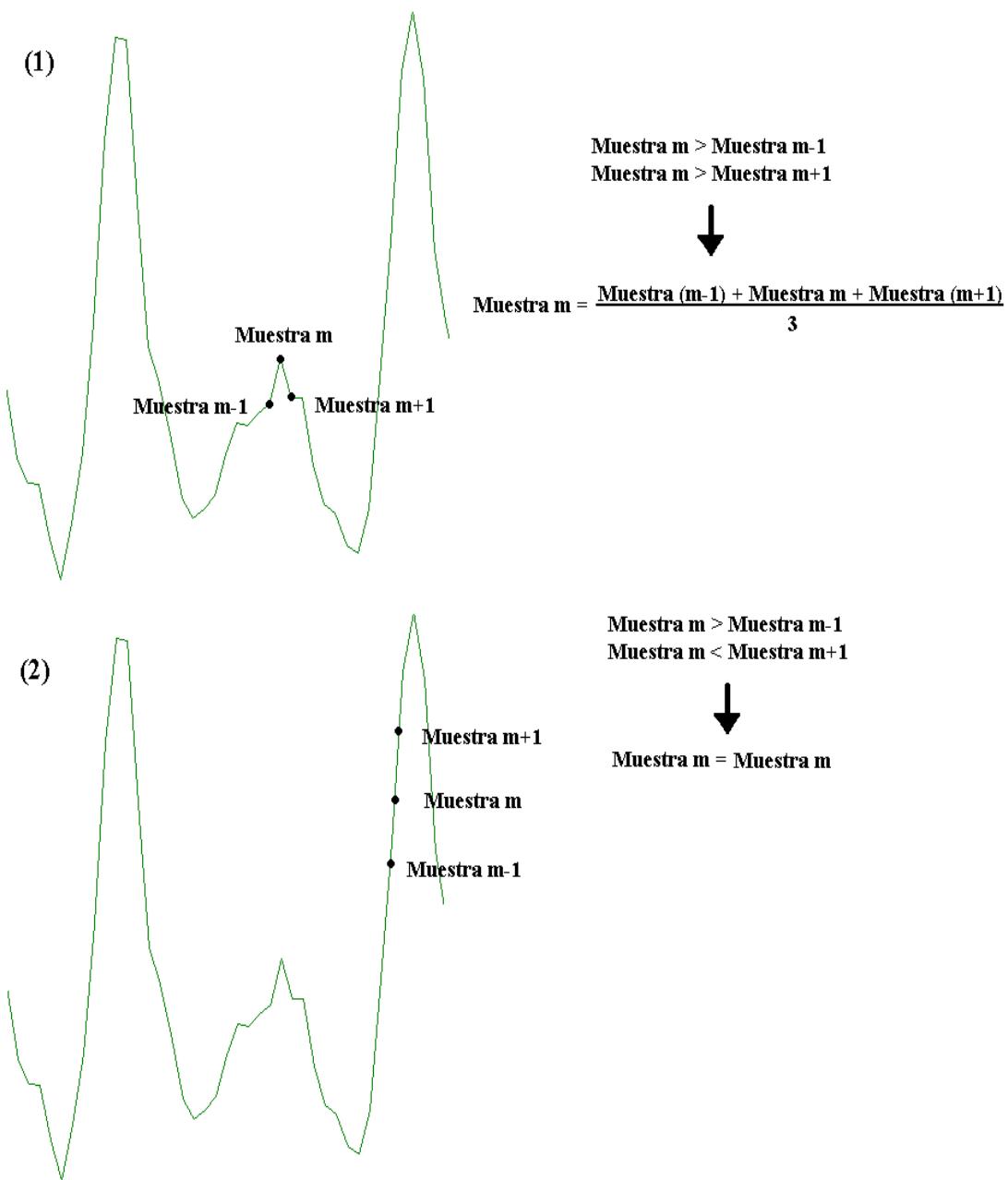


Fig.2.3: Funcionamiento del filtro de mediana.

El filtrado de mediana utiliza un **elemento estructurante**, que normalmente es de 3 muestras; **m-1, m, m+1**. Este elemento estructurante recorre la señal entera comparando los valores de las tres muestras que en cada momento señale este elemento, y ejecutando las siguientes directrices:

- Si el valor de las tres muestras, que no sea ni el máximo ni el mínimo de ellas, es decir, el valor mediano, se corresponde con la muestra m que indique el elemento estructurante, **no se hace nada**.
- Si el valor de las tres muestras que no sea ni el máximo ni el mínimo de ellas, es decir, el valor mediano, no corresponde con la muestra m que indique el elemento estructurante, se substituye ésta por la **media aritmética** de las tres muestras.



### 3.- Modelo Autorregresivo

Tal y como avanzábamos en el capítulo anterior, en la aplicación del filtrado de Wiener la estimación de la densidad espectral de potencia de la señal de voz se hará mediante modelos de predicción lineal o autorregresivos.

Hace unas 2 décadas, los sistemas basados en este modelo habían perdido terreno frente a la sustracción espectral, que aún teniendo prestaciones inferiores aportaban menor complejidad de cálculo y, en consecuencia, eran más adecuados para la implementación con los medios disponibles en la época.

Posteriormente, el método sólo polos recobró su interés gracias a las mejoras aportadas a sus algoritmos de resolución, solucionando algunos de los problemas que presentaban y aumentando sus prestaciones.

La posibilidad de realizar un mayor número de operaciones en un tiempo menor gracias al DSP reactivó hace muy poco la posibilidad de introducir los modelos paramétricos de orden superior (cumulantes de tercer y cuarto orden) en la resolución del modelo autorregresivo.

#### 3.1.- Función de transferencia

El modelo autorregresivo sólo polos (sin acoplos nasales), AR, utilizando la segmentación de la señal de voz por tramas de corta duración temporal (10-30 ms), para poder considerar que el tracto bucal no varía en el tiempo que lo procesamos, nos permite modelar la fonación con la siguiente función de transferencia:

$$H(z) = \frac{g}{1 - \sum_{k=1}^p a_k \cdot z^{-k}} \quad (3.1)$$

Donde **P** es el orden del modelo. Por lo tanto, la señal de voz en el dominio temporal satisface la ecuación:

$$s(n) = \sum_{k=1}^p a_k \cdot s(n-k) + u(n) + e(n) \quad (3.2)$$

Donde  $\mathbf{u}(n)$  es la excitación de entrada y  $\mathbf{e}(n)$  es el error que introducimos a la salida del sistema al considerar una excitación simplificada. Para generar los sonidos sonoros  $\mathbf{u}(n)$  es un tren de pulsos de período  $T=1/\text{Pitch}$  y ruido de banda ancha para sonidos sordos. En adelante consideraremos ambos en un único término  $\mathbf{w}(n)$ . La ecuación (3.2) se denomina modelo autorregresivo, AR, o también modelo de predicción lineal, ya que esta ecuación obtiene cada muestra  $s(n)$  como combinación lineal de sus anteriores  $P$  muestras con un error de predicción igual a  $e(n)$ .

A partir de ahora progresaremos hacia la determinación de los coeficientes  $\{\mathbf{a}_k\}$  y el factor de ganancia  $\mathbf{g}$ . Utilizaremos la notación vectorial para escribir los coeficientes  $\{\mathbf{a}_k\}$  del modelo sólo polos:

$$\underline{\mathbf{a}} = \begin{pmatrix} a_1 \\ \vdots \\ a_P \end{pmatrix} \quad (3.3)$$

Y por tanto la señal de voz puede escribirse como:

$$s(n) = \underline{\mathbf{a}}^T \cdot \underline{s}_P + w(n) \quad (3.4)$$

Donde  $\underline{s}_P$  es:

$$\underline{s}_P = \begin{pmatrix} s(n-1) \\ \vdots \\ s(N-P) \end{pmatrix} \quad (3.5)$$

En ausencia de ruido, el problema de estimación de los coeficientes  $\{\mathbf{a}_k\}$  se reduce a la resolución de un sistema de ecuaciones del tipo:

$$\underline{\underline{\mathbf{M}}} \cdot \underline{\mathbf{a}} = \underline{\mathbf{b}} \quad (3.6)$$

Donde  $\underline{\underline{M}}$  es una matriz de  $P \times P$  valores y  $\underline{b}$  un vector de longitud  $P$ . En función de cómo obtengamos la matriz  $\underline{\underline{M}}$  y el vector  $\underline{b}$  a partir de las muestras  $s(n)$ , sus elementos serán estadísticas de un tipo u otro, correlaciones, covarianzas o cumulantes.

Cuando usemos estadísticas de segundo orden hablaremos del método de correlaciones y del método de covarianzas, mientras que al utilizar estadísticas de orden superior podremos hablar del método de cumulantes de tercer orden y del método de cumulantes de cuarto orden.

### 3.2.- Modelo AR de segundo orden.

Si suponemos que no hay ruido que nos dificulte el cálculo de los coeficientes  $\{a_k\}$  y empleamos estadísticas de segundo orden, el sistema de ecuaciones (3.6) queda simplificado de la siguiente manera.

$$\underline{\underline{R}} \cdot \underline{a} = \underline{r} \quad (3.7)$$

Siendo  $\underline{\underline{R}}$  una matriz  $P \times P$  valores y  $\underline{r}$  un vector de longitud  $P$ , que contienen el mismo tipo de estadísticas, correlaciones.

El método de las correlaciones cuenta con ciertas ventajas por ser la matriz  $\underline{\underline{R}}$  de Toeplitz, matriz definida positiva y simétrica, para la cual existen algoritmos de resolución como el de Levinson muy eficientes, consiguiendo una mayor velocidad de resolución del modelo, además de garantizar su estabilidad.

#### 3.2.1.- Minimización del error cuadrático medio.

Una alternativa para el cálculo de los coeficientes  $\{a_k\}$  del modelo es intentar minimizar el error cuadrático medio (ECM) entre la señal que tenemos en realidad y la predicción lineal que hacemos de ella en el dominio temporal.

Supongamos que la excitación  $u(n)$  nos es totalmente desconocida, por lo que deberemos hacer una estimación del valor actual de la señal  $s(n)$  en función de las  $P$  muestras anteriores  $s(n-k)$ , con  $k=1..P$ . Por tanto:

$$\hat{s}(n) = \sum_{k=1}^p a_k \cdot s(n-k) \quad (3.8)$$

El error cometido en esta predicción vendrá dado por:

$$e_n = s(n) - \hat{s}(n) = s(n) - \sum_{k=1}^p a_k \cdot s(n-k) \quad (3.9)$$

Supondremos que  $s(n)$  es un proceso aleatorio, entonces podremos calcular el error cuadrático medio como:

$$E_{cm} = E\{e_n^2\} = E\left\{\left(s(n) - \sum_{k=1}^p a_k \cdot s(n-k)\right)^2\right\} \quad (3.10)$$

Procedemos a minimizar el valor del error cuadrático medio  $E_{cm}$ , según la condición de minimización:

$$\frac{\partial E_{cm}}{\partial a_k} = 0 \quad 1 \leq k \leq P \quad (3.11)$$

En consecuencia obtenemos el conjunto de ecuaciones.

$$E\left\{2 \cdot \left(s(n) - \sum_{k=1}^p a_k \cdot s(n-k)\right) \cdot s(n-i)\right\} = 0 \quad (3.12)$$

Desarrollando el producto teniendo en cuenta que los coeficientes  $\{a_k\}$  son de valor constante dentro de este estudio, la ecuación anterior queda como:

$$\sum_{k=1}^p a_k \cdot E\{s(n-k) \cdot s(n-i)\} = E\{s(n) \cdot s(n-i)\} \quad i = 1..P \quad (3.13)$$

Además, como el sistema es considerado estacionario, las esperanzas de (3.13) se pueden escribir como:

$$E\{s(n-k) \cdot s(n-i)\} = R(i-k) \quad (3.14)$$

Donde los  $R(i-k)$  representan las correlaciones del proceso  $s(n)$  para los distintos desplazamientos de la señal sobre si misma.

Aplicando (3.14), rescribimos el sistema de ecuaciones (3.13) como:

$$\sum_{k=1}^p a_k \cdot R(i-k) = R(i) \quad 1 \leq i \leq p \quad (3.15)$$

Donde:

$$R(i) = \sum_{n=-\infty}^{\infty} s(n) \cdot s(n+i) \quad (3.16)$$

Es la autocorrelación de la señal  $s(n)$ , que cumple la propiedad de simetría:

$$R(i) = R(-i) \quad (3.17)$$

Y por tanto, el conjunto de ecuaciones lineales (3.13) puede escribirse como:

$$\begin{aligned} a_1 \cdot R(0) + a_2 \cdot R(1) + \dots + a_p \cdot R(p-1) &= R(1) \\ a_1 \cdot R(1) + a_2 \cdot R(0) + \dots + a_p \cdot R(p-2) &= R(2) \\ \vdots &= \vdots \\ \vdots &= \vdots \\ \vdots &= \vdots \\ a_1 \cdot R(p-1) + a_2 \cdot R(p-2) + \dots + a_p \cdot R(0) &= R(p) \end{aligned} \quad (3.17)$$

Estas ecuaciones son conocidas como las ecuaciones de Yule-Walker y corresponden a la ecuación matricial que ya apuntábamos en (3.7):

$$\underline{\underline{R}} \cdot \underline{a} = \underline{r}$$

Donde tanto la matriz  $\underline{\underline{R}}$  como el vector  $\underline{r}$  están formados por elementos que corresponden a autocorrelaciones de la señal  $s(n)$ :

$$\underline{\underline{R}} = \begin{pmatrix} R(0) & R(1) & \dots & R(p-1) \\ R(1) & R(0) & \dots & R(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ R(p-1) & R(p-2) & \dots & R(0) \end{pmatrix} \quad \underline{r} = \begin{pmatrix} R(1) \\ R(2) \\ \vdots \\ \vdots \\ R(p) \end{pmatrix} \quad (3.18)$$

Este conjunto de ecuaciones son también conocidas como el principio de ortogonalidad:

Si trasladamos las ecuaciones a un espacio vectorial, donde la esperanza,  $E\{\cdot\}$ , representa el producto escalar de dos vectores en dicho espacio, exigir que cada una de las ecuaciones sea idénticamente nula supone imponer que el error  $e_n$  sea ortogonal a todos los datos o, para nosotros, muestras de la señal de voz,  $s(n-i)$ . Por tanto la mejor aproximación a  $s(n)$  que podemos predecir viene dada por  $\hat{s}(n)$ ; entonces  $\hat{s}(n)$  es la proyección de  $s(n)$  sobre el espacio vectorial de datos  $s(n-1), \dots, s(n-p)$ . Esto implica que el error  $e_n$  es mínimo cuando es ortogonal a  $\hat{s}(n)$  o, lo que es lo mismo, ortogonal a cada una de las muestras  $s(n-1), \dots, s(n-p)$ , ya que  $\hat{s}(n)$  es combinación lineal de las muestras de  $s(n)$  utilizadas como datos.

La figura 3.1 muestra la interpretación geométrica para un espacio predictor de dimensión 2 ( $P=2$ ), donde el plano de datos  $s(n-1), s(n-2)$  contiene a la estimación  $\hat{s}(n)$  y es ortogonal al error mínimo,  $e_{min}(n)$ .

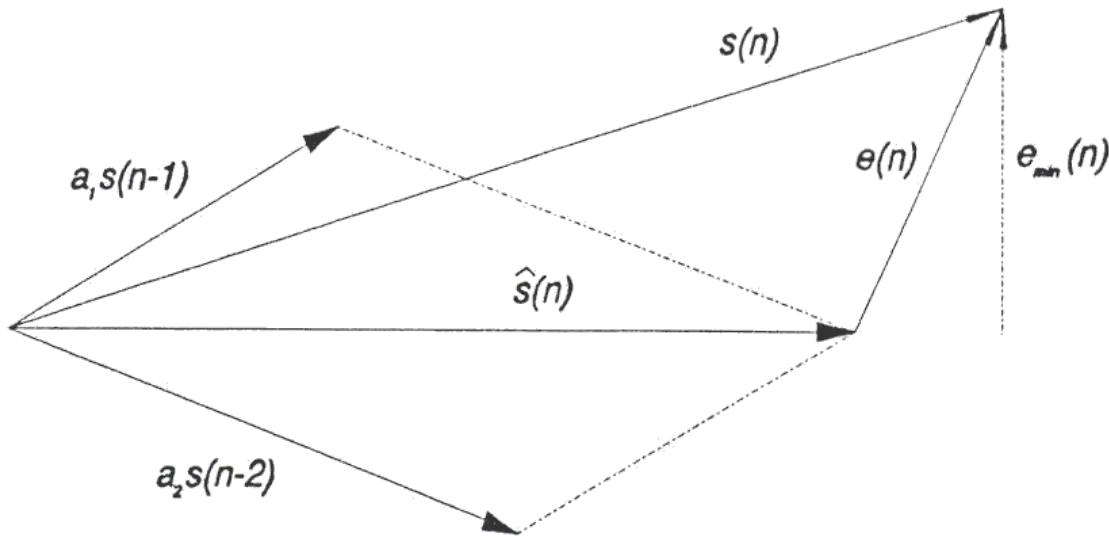


Fig.3.1: Interpretación geométrica para  $P=2$  del predictor lineal.

Existen dos limitaciones para la aplicación directa de este algoritmo. Por una parte no disponemos de la señal en un intervalo de tiempo infinito, y en segundo lugar hemos de mantener la condición de estacionariedad impuesta sobre la función de transferencia. Para poder usar este método dentro de las condiciones expuestas, tal y como comentamos en puntos anteriores, utilizaremos tramas de corta duración en el dominio

temporal, aproximadamente de unos 30 ms como máximo, para tener un sistema estacionario.

Recurriendo a la técnica de enventanado por tramas de la señal  $s(n)$  mediante una ventana  $w(n)$ , obtenemos una señal  $s'(n)$  de duración finita; y solucionamos así la primera limitación del algoritmo:

$$s'(n) = \begin{cases} s(n) \cdot w(n) & 0 \leq n \leq N-1 \\ 0 & \text{resto} \end{cases} \quad (3.19)$$

Ahora calcularemos la autocorrelación tan sólo en la porción de  $s(n)$  que hemos enventanado, es decir, sobre  $s'(n)$ :

$$R(i) = \sum_{n=0}^{N-1-i} s'(n) \cdot s'(n+i) \quad i \geq 0 \quad (3.20)$$

Posteriormente comentaremos los diferentes tipos de ventanas existentes y las ventanas que hemos usado en nuestra implementación, así como el solapamiento entre las tramas. No obstante, existe una condición general que las ventanas deben cumplir para evitar introducir constantes multiplicativas en el sistema:

$$\sum_i w_i(n - N \cdot i) = 1 \quad \text{para } \forall n \quad (3.21)$$

Donde  $w_i$  representa cada una de las infinitas ventanas a utilizar para segmentar por tramas la señal.

### 3.2.2.- Efectos del modelo AR de segundo orden

Debido a las limitaciones del sistema real en el que realizaremos el procesado, sólo dispondremos de la señal de voz contaminada con ruido a la entrada de éste, y el hecho de realizar la estimación de la densidad espectral de potencia de voz,  $P_s(\omega)$ , para realizar el filtrado de Wiener, mediante modelos autorregresivos de segundo orden nos lleva ha encontrarnos con una serie de problemas.

Los modelos de predicción lineal de segundo orden optimizan los coeficientes  $\{a_k\}$ , es decir, el modelo LPC sigue fielmente el espectro de la señal de entrada al estimador.

Los **akóptimos** siguen a la señal de entrada, pero ésta no es la señal limpia de voz que queremos modelar, sino que es señal más ruido:

$$\begin{aligned} s(n) &\rightarrow \text{AR } 2^{\circ} \text{ orden} \rightarrow a_{k\text{óptimo}} \\ y(n) = s(n) + d(n) &\rightarrow \text{AR } 2^{\circ} \text{ orden} \rightarrow a'_k \end{aligned} \quad (3.22)$$

Por tanto no estamos siguiendo a la señal de voz sino a  $y(n)$ . Esta limitación repercute principalmente y de forma más sensible en la posición de los formantes. Este efecto es más acusado al aumentar el ruido enmascarante, es decir, al disminuir la relación señal a ruido a la entrada del sistema,  $(S/N)_i$ .

En la figura (3.2) se observa como los **formantes se desplazan** hacia la posición donde están si nos guiamos por la señal  $y(n)$  (modelo LPC superior) en lugar de buscar la posición marcada en  $s(n)$  (línea continua). Este efecto más acusado en los formantes más lejanos del origen, ya que su  $(S/N)_i$  frecuencia a frecuencia es menor respecto a los primeros formantes.

Otro efecto que también se da al filtrar los sonidos sonoros es la reducción del ancho de banda de los formantes. El efecto se agrava al aumentar el número de iteraciones aplicadas, a mayor número de filtrados menor ancho de banda. Este progresivo aspecto más puntiagudo de los formantes se conoce como **picado de los formantes** (spectral peaking) [Masg-92b].

Este efecto es intrínseco al filtrado de Wiener iterativo ya que pondera cada frecuencia según su relación señal a ruido para otorgarle su valor en el filtro. Así pues, la zona central de los formantes posee mayor relación señal a ruido y el filtro le da más credibilidad dejándola pasar, mientras que en los valles del espectro es más riguroso y además de eliminar ruido elimina parte de la señal produciendo el picado de los formantes. Este efecto puede apreciarse también en la figura (3.2).

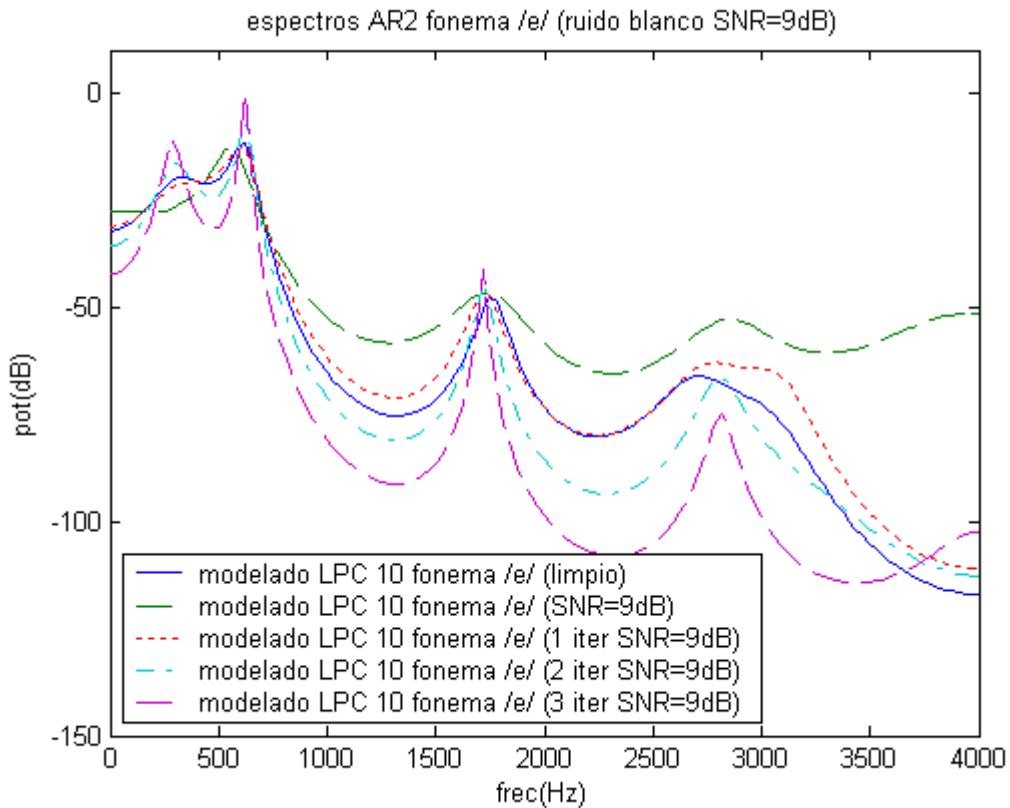


Fig.3.2: Efecto del picado espectral y desplazamiento de los formantes para una trama de señal con  $(S/N)=9\text{dB}$  durante 4 iteraciones de filtrado.  
Línea continua: señal limpia; línea discontinua superior: señal más ruido;  
las demás son de arriba a abajo las sucesivas iteraciones de la 1 a la 3.

El espectro que seguimos en la primera iteración es:

$$P_y(\omega) = P_s(\omega) + P_d(\omega) \quad (3.23)$$

Este error inicial proporcional al ruido de entrada se arrastra en las iteraciones posteriores produciendo el desplazamiento de los formantes alrededor de su posición real.

En el método de correlaciones al tener  $\mathbf{y}(n)$  únicamente:

$$\begin{aligned} R_{yy}(k) &= E\{(s(n) + d(n)) \cdot (s(n+k) + d(n+k))\} \\ R_{yy} &= E\{s(n) \cdot s(n+k)\} + E\{d(n) \cdot d(n+k)\} \end{aligned} \quad (3.24)$$

$$R_{yy}(k) = R_{ss}(k) + R_{dd}(k)$$

La ecuación anterior la hemos simplificado considerando que la señal de voz  $\mathbf{s}(n)$  y el ruido  $\mathbf{d}(n)$  están incorrelados, y por tanto:

$$E\{s(n) \cdot d(n+k)\} = E\{s(n+k) \cdot d(n)\} = 0 \quad (3.25)$$

Entonces el espectro estimado será:

$$\begin{aligned} P_y(\omega) &= F[R_{yy}(k)] \\ P_y(\omega) &= F[R_{ss}(k) + R_{dd}(k)] = P_s(\omega) + P_d(\omega) \end{aligned} \quad (3.26)$$

El tercer problema planteado, que resulta de interés también para los dos primeros, es la estimación de los coeficientes  $\{ak\}$  en presencia del ruido. Para ello se propone la introducción de las estadísticas de orden superior, cumulantes de tercer y cuarto orden, que permiten distinguir mejor entre señal y ruido.

### 3.3.- Estadísticas de orden superior (HOS)

La mayor parte de los métodos clásicos de estimación espectral, mencionados hasta ahora, se caracterizan porque durante la estimación espectral se procesa una señal cuyo espectro se interpreta como una superposición de componentes frecuenciales armónicas, estadísticamente incorreladas, y se obtiene su distribución de potencia a lo largo de sus componentes frecuenciales, perdiéndose las relaciones de fase existentes entre estas componentes. La información contenida en el módulo espectral de una señal se corresponde, esencialmente, con la información presente en su secuencia autocorrelación, y resulta ser suficiente para la caracterización estadística completa de una señal con distribución Gaussiana, cuya media sea conocida. Sin embargo, hay situaciones reales donde es necesario extraer información referente a la desviación que una determinada señal o proceso presenta respecto a la **Gaussianidad** o, incluso, disponer de la información contenida en su **fase espectral** y, entonces, el método clásico de autocorrelación resulta ser ciego en lo referente a las características anteriores [Niki-87], [Mend-91].

En cambio, esta información se encuentra presente en los espectros de orden superior (también conocidos como poliespectros), definidos a partir de las estadísticas de orden superior (cumulantes) de una señal. Casos particulares de estos espectros de orden superior son el espectro de tercer orden (Biespectro), que por definición se corresponde con la transformada de Fourier de las estadísticas de tercer orden, y el espectro de cuarto

orden (Triespectro), definido como la transformada de Fourier de los cumulantes de cuarto orden de una señal estacionaria. Nótese que el espectro de potencia clásico puede verse como el espectro de segundo orden dentro del entorno de las estadísticas de orden superior (HOS). La siguiente figura muestra la clasificación de los espectros de orden superior correspondientes a una señal discreta dada.

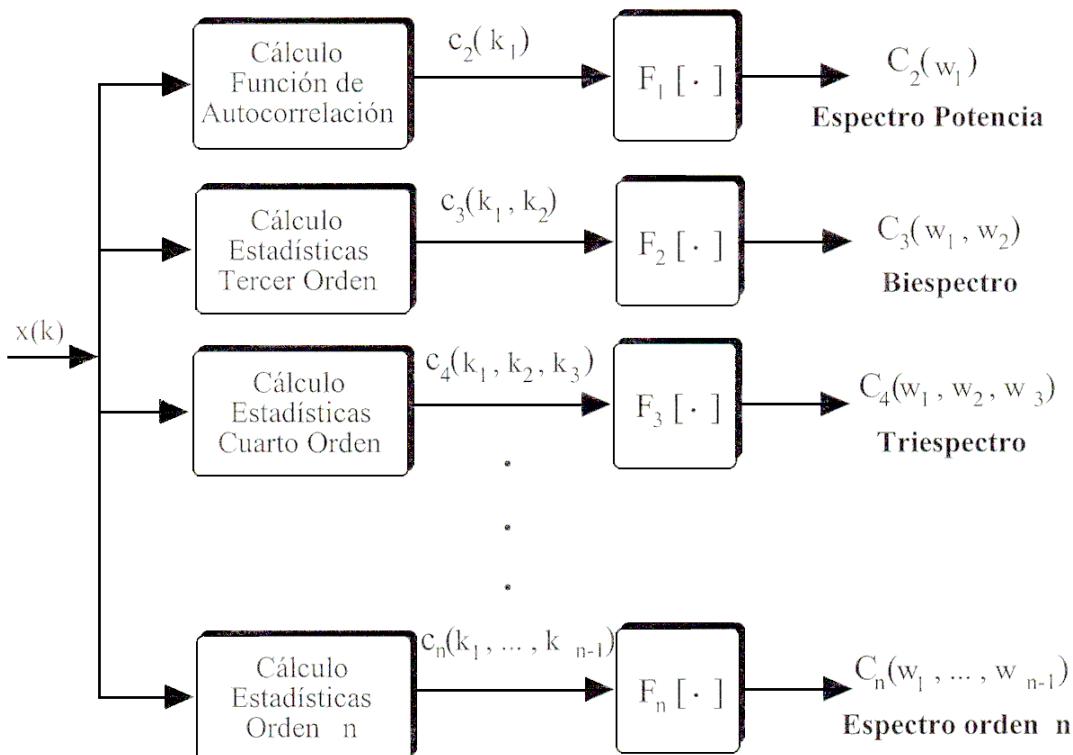


Fig.3.3: Esquema de clasificación de los espectros de orden superior para una señal discreta  $x(k)$ , donde  $F_n[\cdot]$  representa la transformada de Fourier de dimensión  $n$ . [Sala-95]

Las estadísticas y espectros de orden superior pueden expresarse en función de los momentos o de los cumulantes. Mientras que los momentos, y sus correspondientes espectros, resultan muy provechosos para el análisis de señales deterministas (periódicas o transitorios), los cumulantes y sus espectros son de gran importancia durante el análisis de procesos estocásticos [Niki-91]. El uso de los espectros de orden superior es adecuado en determinados campos del procesado de la señal, especialmente, en técnicas para:

- Suprimir el ruido aditivo Gaussiano, blanco o coloreado, cuyo espectro de potencia sea desconocido.

- Identificar sistemas de fase no mínima o reconstruir señales de fase no mínima.
- Extraer información relativa a las desviaciones respecto a las características de un proceso Gaussiano.
- Detectar y caracterizar propiedades no lineales de algunas señales determinadas o identificar sistemas no lineales [Niki-87].

La primera de estas posibles aplicaciones se refiere a la supresión o reducción de ruido Gaussiano que ha degradado una señal no Gaussiana. Las HOS pueden aplicarse en este contexto porque una de sus propiedades fundamentales es la de tener idénticamente nulos todos los cumulantes, y sus espectros de orden superior a dos solamente para el caso de señales con distribución Gaussiana. De esta manera, cuando se recibe una señal no Gaussiana degradada con ruido aditivo Gaussiano, se puede observar esta señal sin degradación al situarnos en el dominio de los cumulantes de orden superior, y obtener unas prestaciones muy superiores en relación al dominio clásico de los momentos de segundo orden o dominio de la función autocorrelación.

La segunda aplicación se fundamenta en la propiedad que todos los poliespectros de cumulantes o de momentos, conservan la información de fase de la señal tratada. Durante el modelado de series temporales, en procesado de la señal, suelen usarse las estadísticas de segundo orden porque resultan de aplicar el criterio de optimización de Mínimos Cuadráticos, que conduce a la estimación de Máxima Verosimilitud de los parámetros de Procesos Gaussianos y, además, conduce a sistemas de ecuaciones lineales donde aparece la función autocorrelación. Cuando el proceso no sea Gaussiano el criterio de Mínimos Cuadráticos ya no conduce a la solución de Máxima Verosimilitud. Si consideramos un proceso  $\mathbf{x}(k)$  real, estacionario y aleatorio de media nula, su función autocorrelación  $R(k_1)$  da una medida de lo correlada que está esta secuencia respecto a desplazamientos de ella misma:

$$R(k_1) = E\{x(k) \cdot x(k + k_1)\} \quad (3.27)$$

Además, tal y como vimos, la función autocorrelación presenta simetría par respecto al origen:

$$R(-k_1) = R(k_1) \quad (3.28)$$

Y, en consecuencia, su transformada de Fourier también es una función real y par en el dominio frecuencial, es decir, resulta una función de fase nula. La información contenida en la fase del proceso  $\mathbf{x}(\mathbf{k})$  se ha perdido al considerar  $\mathbf{R}(\mathbf{k}_1)$ . Así, la información contenida en la fase desaparece en el dominio de la autocorrelación, pudiéndose recuperar solamente para el caso de procesos de fase mínima. De esta manera, el uso de espectros de orden superior se impone para la reconstrucción de procesos de fase no mínima, o para la identificación de sistemas de fase no mínima, puesto que los poliespectros contienen las informaciones correspondientes al módulo y a la fase no mínima. Así, por ejemplo, dos procesos distintos, uno de fase mínima y otro espectralmente equivalente de fase no mínima, pueden presentar la misma función de autocorrelación y diferentes estadísticas de tercer orden, es decir, el mismo espectro de potencia y distintos biespectros. Nótese, que en el caso de procesos Gaussianos de fase no mínima ningún método permite recuperar la información de fase.

La mayor parte de señales o procesos reales no presentan una distribución Gaussiana y, por este motivo, presentan unos espectros de orden superior no nulos. Un proceso no Gaussiano puede descomponerse entre sus funciones spectrales de orden superior y cada una de éstas puede contener distinta información acerca de este proceso. Esto resulta muy útil en aplicaciones de clasificación de procesos donde distintas características de clasificación pueden extraerse a partir de los distintos dominios spectrales de orden superior.

Finalmente, el uso de los espectros de orden superior parece bastante lógico cuando se pretenda analizar alguna no linealidad de un sistema, operando con una entrada aleatoria. Durante los últimos años se han estudiado, de forma bastante extensa, las relaciones entre determinadas señales aleatorias estacionarias y su paso a través de determinados sistemas lineales. Sin embargo, la mayoría de estas propiedades se han establecido considerando criterios basados en el espectro de potencia o en la función de autocorrelación y, además, la mayoría de estas relaciones dejan de cumplirse cuando se trabaja con sistemas no lineales. Cada tipo de alinealidad se debe investigar como un caso especial y los Poliespectros pueden jugar un papel clave en vistas a detectar y caracterizar cada tipo de alinealidad de un sistema, a partir de la señal a la salida del sistema analizado. Bajo las premisas anteriores se han desarrollado distintos métodos para la detección y caracterización de alinealidades en series temporales mediante el uso

de espectros de orden superior [Niki-91].

### 3.3.1.-Definiciones Temporales y Frecuenciales

En este apartado se presentan las principales definiciones correspondientes a las estadísticas de orden superior y sus correspondientes espectros de orden superior. En principio no se distingue entre señales deterministas y estocásticas porque todas las definiciones que se presentan tienen validez para ambos tipos de señales, bajo la suposición de estacionariedad. Aunque en este trabajo las señales deterministas no son de interés. En el primer subapartado se discuten estas definiciones desde el dominio temporal, dejando la discusión correspondiente al dominio frecuencial para el segundo subapartado. En ambos casos se parte de la definición correspondiente a un orden **n** genérico para, seguidamente, situarse en los casos de segundo, tercero y cuarto orden, donde se sedimenta el trabajo realizado en el presente estudio.

#### 3.3.1.1.-Momentos y Cumulantes

Sea  $\mathbf{x}(\mathbf{k})$  un proceso discreto, real y estacionario cuyos momentos existen hasta un orden **n**, entonces se define su momento de orden **n** como:

$$m_n(k_1, k_2, \dots, k_{n-1}) = E\{x(k), x(k+k_1), \dots, x(k+k_{n-1})\} \quad (3.29)$$

Y  $E\{\cdot\}$  representa el operador esperanza estadística. Puede observarse que este momento de orden **n** depende, solamente, de los desplazamientos temporales  $k_1, k_2, \dots, k_{n-1}$  debido a la estacionariedad del proceso  $\mathbf{x}(\mathbf{k})$ . Cuando el proceso considerado no sea estacionario, también depende de la posición temporal  $m_n(k, k_1, k_2, \dots, k_{n-1})$ . Se aprecia, también, que el momento de segundo orden  $m_2(k_1)$  se corresponde claramente con la función autocorrelación clásica, mientras que los momentos de tercer y cuarto orden vienen representados, respectivamente, por  $m_3(k_1, k_2)$  y  $m_4(k_1, k_2, k_3)$ .

Para el caso de un proceso  $\mathbf{x}(\mathbf{k})$  aleatorio, estacionario y no Gaussiano los cumulantes de tercer y cuarto orden pueden expresarse en función de los momentos como sigue (**n=3,4**):

$$c_n(k_1, k_2, \dots, k_{n-1}) = m_n(k_1, k_2, \dots, k_{n-1}) - m_n^G(k_1, k_2, \dots, k_{n-1}) \quad (3.30)$$

Donde:

$$m_n^G(k_1, k_2, \dots, k_{n-1}) = E\{g(k), g(k+k_1), \dots, g(k+k_{n-1})\} \quad (3.31)$$

Representa el momento de orden **n** de un proceso Gaussiano **g(k)** equivalente a **x(k)**, de tal manera, que ambas presenten la misma media y la misma secuencia de autocorrelación, es decir, con idénticos momentos de primer y segundo orden. Evidentemente, esta definición concuerda con una propiedad de los cumulantes mencionada anteriormente, respecto a la nulidad de los cumulantes de tercer y cuarto orden, **c<sub>n</sub>(k<sub>1</sub>, k<sub>2</sub>, ..., k<sub>n-1</sub>)=0**, cuando el proceso **x(n)** presenta una distribución Gaussiana, puesto que se verifica:

$$m_n(k_1, k_2, \dots, k_{n-1}) = m_n^G(k_1, k_2, \dots, k_{n-1}) \quad (3.32)$$

A continuación se presentan las relaciones básicas entre cumulantes y momentos, resultantes de combinar las expresiones (3.29) y (3.30), particularizando para los cuatro primeros órdenes que serán los que utilizaremos en este proyecto.

a) Cumulantes de primer orden:

$$c_1 = m_1 = E\{x(k)\} \quad (3.33)$$

b) Cumulantes de segundo orden:

$$c_2(k_1) = m_2(k_1) - [m_1]^2 = R(k_1) - [m_1]^2 = R(-k_1) - [m_1]^2 = c_2(-k_1) \quad (3.34)$$

Donde **R(k<sub>1</sub>)** se corresponde con la secuencia de autocorrelación y, asimismo, los cumulantes de segundo orden **c<sub>2</sub>(k<sub>1</sub>)** con la secuencia de covarianza.

c) Cumulantes de tercer orden:

$$c_3(k_1, k_2) = m_3(k_1, k_2) - m_1[R(k_1) + R(k_2) - R(k_1 - k_2)] + 2 \cdot [m_1]^3 \quad (3.35)$$

d) Cumulantes de cuarto orden:

$$\begin{aligned} c_4(k_1, k_2, k_3) &= m_4(k_1, k_2, k_3) - R(k_1) \cdot R(k_3 - k_2) - R(k_2) \cdot R(k_3 - k_1) - R(k_3) \cdot R(k_2 - k_1) \\ &\quad - 6 \cdot [m_1]^4 - m_1[m_3(k_2 - k_1, k_3 - k_1) + m_3(k_2, k_3) + m_3(k_1, k_3) + m_3(k_1, k_2)] \\ &\quad + [m_1]^2 \cdot [R(k_1) + R(k_2) + R(k_3) + R(k_3 - k_1) + R(k_3 - k_2) + R(k_2 - k_1)] \end{aligned} \quad (3.36)$$

Si el proceso  $\mathbf{x}(\mathbf{k})$  presenta media nula,  $\mathbf{m1}=\mathbf{0}$ , entonces las expresiones anteriores se simplifican y los cumulantes de segundo y tercer orden se corresponden idénticamente con los momentos de segundo y tercer orden respectivamente:

$$c_2(k_1) = m_2(k_1) = R(k_1) \quad (3.37)$$

$$c_3(k_1, k_2) = m_3(k_1, k_2) \quad (3.38)$$

sin embargo, para generar los cumulantes de cuarto orden se precisa del conocimiento de los momentos de cuarto y segundo orden:

$$c_4(k_1, k_2, k_3) = m_4(k_1, k_2, k_3) - R(k_1) \cdot R(k_3 - k_2) - R(k_2) \cdot R(k_3 - k_1) - R(k_3) \cdot R(k_2 - k_1) \quad (3.39)$$

En nuestro sistema siempre lograremos que  $\mathbf{m1}=\mathbf{0}$ , calculando la media de la señal a la entrada y restándosela seguidamente, de esta manera los cálculos de los cumulantes se simplifican y el tiempo de procesado se reduce, al disminuir el cálculo computacional.

Cuando nos situamos en el origen, es decir, considerando desplazamientos temporales nulos,  $k_1=k_2=k_3=0$ , se presentan los conceptos de varianza, skewness y kurtosis correspondientes, respectivamente, a los dominios de segundo, tercero y cuarto orden:

$$\text{Varianza: } \gamma_2 = E\{x^2(k)\} = c_2(0) \quad (3.40)$$

$$\text{Skewness: } \gamma_3 = E\{x^3(k)\} = c_3(0,0) \quad (3.41)$$

$$\text{Kurtosis: } \gamma_4 = E\{x^4(k)\} - 3 \cdot [\gamma_2]^2 = c_4(0,0,0) \quad (3.42)$$

En el dominio de los cumulantes de orden superior, un valor nulo en el origen no implica que se anulen los cumulantes para cualquier punto del plano multidimensional. Así, por ejemplo, un valor nulo de la skewness no implica que los cumulantes de tercer orden sean idénticamente cero.

Aunque los cumulantes de cuarto orden implican un incremento considerable de la complejidad de cálculo, resultan especialmente necesarios cuando los cumulantes de tercer orden se anulan para el caso de procesos distribuidos simétricamente, tales como los procesos uniformes, procesos de Laplace, procesos Gaussianos y los procesos de Bernoulli-Gaussianos. Los cumulantes de tercer orden no se anulan para los procesos cuya función densidad de probabilidad no sea simétrica, como por ejemplo los procesos

exponenciales o los de Rayleigh, pero pueden tomar valores extremadamente pequeños en comparación a los valores que presentan sus cumulantes de cuarto orden y, entonces, también parece lógico usar éstos últimos.

### 3.3.1.2.- Propiedades de los Cumulantes

Los cumulantes pueden usarse como un operador, de la misma forma que tratamos el operador esperanza estadística. Las principales propiedades de los cumulantes, que sostienen esta afirmación, son las siguientes (su demostración puede hallarse en [Mend-91]):

1) Los cumulantes de señales o procesos escalados, no siendo estos factores de escala aleatorios, se corresponden con el producto de todos estos factores por los cumulantes del proceso sin escalar:

$$\text{cum}\{\lambda_0 \cdot x(k), \dots, \lambda_{n-1} \cdot x(k - k_{n-1})\} = \left( \prod_{i=0}^{n-1} \right) \cdot \text{cum}\{x(k), \dots, x(k - k_{n-1})\} \quad (3.43)$$

Donde  $\lambda_i$  son constantes y

$$\text{cum}\{x(k), \dots, x(k - k_{n-1})\} = c_n(k_1, \dots, k_{n-1}) \quad (3.44)$$

2) Los cumulantes son simétricos respecto la posición de sus argumentos ( $k_0=0$ ):

$$\text{cum}\{x(k - k_0), \dots, x(k - k_{n-1})\} = \text{cum}\{x(k - k_{i_0}), \dots, x(k - k_{i_{n-1}})\} \quad (3.45)$$

Donde **(i0,..., in-1)** es una permutación de **(0, 1,..., n-1)**. Esto significa que se pueden intercambiar los argumentos de los cumulantes sin modificar su valor. De este modo los cumulantes de cuarto orden verifican:

$$c_4(k_1, k_2, k_3) = c_4(k_3, k_1, k_2) = c_4(k_2, k_3, k_1) \quad (3.46)$$

3) Los cumulantes son aditivos respecto a sus argumentos, es decir, los cumulantes de una suma de argumentos se corresponde con la suma de cumulantes:

$$\begin{aligned} \text{cum}\{x(k) + y(k), x(k - k_1), \dots, x(k - k_{n-1})\} &= \text{cum}\{x(k), x(k - k_1), \dots, x(k - k_{n-1})\} \\ &+ \text{cum}\{y(k), x(k - k_1), \dots, x(k - k_{n-1})\} \end{aligned} \quad (3.47)$$

De ahí viene el nombre "cumulant" en inglés.

4) Los cumulantes son transparentes respecto la adición de constantes. Siendo  $\delta$  una constante, entonces se verifica:

$$\text{cum}\{\delta + x(k), x(k - k_1), \dots, x(k - k_{n-1})\} = \text{cum}\{x(k), x(k - k_1), \dots, x(k - k_{n-1})\} \quad (3.48)$$

5) Si dos procesos  $x(\mathbf{k})$  e  $y(\mathbf{k})$  son independientes, los cumulantes del proceso suma toma el valor de la suma de los cumulantes de cada proceso por separado:

$$\begin{aligned} \text{cum}\{x(k) + y(k), \dots, x(k - k_{n-1}), y(k - k_{n-1})\} &= \text{cum}\{x(k), \dots, x(k - k_{n-1})\} \\ &+ \text{cum}\{y(k), \dots, y(k - k_{n-1})\} \end{aligned} \quad (3.49)$$

Nótese que si los procesos  $x(\mathbf{k})$ ,  $y(\mathbf{k})$  no fueran independientes, según la propiedad 3 aparecerían  $2n$  términos en el lado derecho de esta última expresión.

6) Si un subconjunto de  $r$  argumentos ( $r \leq n$ ) son independientes del resto entonces se verifica:

$$\text{cum}\{x(k), y(k - k_1), \dots, z(k - k_{n-1})\} \quad (3.50)$$

7) Los cumulantes de orden  $n$  presentan  $n!$  regiones de simetría [Gian-90]:

$$\begin{aligned} c_n(k_1, k_2, \dots, k_{n-1}) &= c_n(k_2, k_1, \dots, k_{n-1}) = \dots = c_n(k_{n-1}, k_{n-2}, \dots, k_1) = \\ c_n(-k_1, k_2 - k_1, \dots, k_{n-1} - k_1) &= \dots = c_n(k_{n-1} - k_1, k_{n-2} - k_1, \dots, -k_1) = \\ c_n(k_1 - k_{n-1}, k_2 - k_{n-1}, \dots, -k_{n-1}) &= \dots = c_n(-k_{n-1}, k_{n-2} - k_1, \dots, k_1 - k_{n-1}) \end{aligned} \quad (3.51)$$

Pero, contrariamente a lo que sucede en el dominio de la función autocorrelación, no se verifica generalmente la simetría par para  $n \geq 3$ :

$$c_n(k_1, k_2, \dots, k_{n-1}) \neq c_n(-k_1, -k_2, \dots, -k_{n-1}) \quad (3.52)$$

Nótese que para el caso de segundo orden se reduce a  $c2(\mathbf{k1}) = c2(-\mathbf{k1})$  y se cumple la simetría par.

Sea  $v(\mathbf{k})$  un proceso Gaussiano independiente de  $x(\mathbf{k})$  (blanco o coloreado), que ha degradado el proceso  $x(\mathbf{k})$  originando el proceso  $y(\mathbf{k}) = x(\mathbf{k}) + v(\mathbf{k})$ , entonces para  $n \geq 3$  se verifica:

$$c_n^y(k_1, k_2, \dots, k_{n-1}) = c_n(k_1, k_2, \dots, k_{n-1}) \quad (3.53)$$

mientras que en el caso de segundo orden clásico se cumple:

$$c_n^y(k_1) = c_n(k_1) + c_n^v(k_1) \quad (3.54)$$

Esta última característica muestra la mayor robustez de las estadísticas de orden superior frente a la función autocorrelación clásica, incluso cuando este ruido sea coloreado. En consecuencia, los cumulantes pueden obtener información de procesos no Gaussianos sin afectarles la presencia de ruidos Gaussianos y, por ello, están viendo unas relaciones señal a ruido efectivas superiores.

### 3.3.2.-Espectros de Orden Superior

Los espectros de orden superior se obtienen al aplicar la Transformada de Fourier multidimensional  $F_n[\cdot]$  sobre las estadísticas de orden superior. Para un orden  $n$  genérico se define el espectro de momentos como:

$$M_n(\omega_1, \omega_2, \dots, \omega_{n-1}) = F_n[m_n(k_1, k_2, \dots, k_{n-1})] \quad (3.55)$$

Y, análogamente, se define el espectro de cumulantes:

$$C_n(\omega_1, \omega_2, \dots, \omega_{n-1}) = F_n[c_n(k_1, k_2, \dots, k_{n-1})] \quad (3.56)$$

Nótese que el espectro de cumulantes de orden  $n$  es también periódico con periodo  $2\pi$ :

$$C_n(\omega_1, \omega_2, \dots, \omega_{n-1}) = C_n(\omega_1 + 2\pi, \omega_2 + 2\pi, \dots, \omega_{n-1} + 2\pi) \quad (3.57)$$

Al trabajar con procesos estocásticos, como pueden ser la señal de voz o el ruido, el espectro de cumulantes presenta una serie de ventajas respecto al espectro de momentos:

- a) Para procesos Gaussianos todos los cumulantes de orden superior a dos se anulan y, por esta razón, el espectro de cumulantes puede medir la no Gaussianidad de un proceso concreto.
- b) Los cumulantes dan una medida bastante conveniente de la extensión de las

relaciones estadísticas que las series temporales presentaban en el caso de segundo orden.

- c) Para el caso de ruido blanco de media no nula, solamente su función covariancia se corresponde con la función impulso y, por consiguiente, presenta un espectro plano. Sus cumulantes de orden superior presentan la forma de una función impulso multidimensional y los poliespectros de este ruido son multidimensionalmente planos.
- d) Los cumulantes de dos procesos aleatorios estadísticamente independientes se corresponden con la suma de los cumulantes de cada proceso individual, a diferencia de los momentos de orden superior que no cumplen esta propiedad.

A partir de ahora sólo tendremos en cuenta los espectros de los cumulantes y particularizaremos para los casos del Espectro de Potencia, el Biespectro y el Triespectro.

### 1) Espectro de Potencia:

$$C_2(\omega_1) = \sum_{k_1=-\infty}^{\infty} c_2(k_1) \cdot e^{-j(\omega_1 \cdot k_1)} \quad (3.58)$$

Donde  $|\omega_1| \leq \pi$  y  $c_2(k_1)$  representa la secuencia de covariancia del proceso  $\mathbf{x}(k)$ . Esta expresión se conoce, también, como Teorema de Wiener-Khintchine.

### 2) Biespectro:

$$C_3(\omega_1, \omega_2) = \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} c_3(k_1, k_2) \cdot e^{-j(\omega_1 \cdot k_1 + \omega_2 \cdot k_2)} \quad (3.59)$$

$$|\omega_1| \leq \pi, |\omega_2| \leq \pi, |\omega_1 + \omega_2| \leq \pi$$

Donde  $c_3(k_1, k_2)$  representa la secuencia de cumulantes de tercer orden de  $\mathbf{x}(k)$ . Al combinar la expresión (3.35) con las propiedades de los momentos, se deducen unas relaciones de simetría muy importantes para estos cumulantes de tercer orden:

$$\begin{aligned} c_3(k_1, k_2) &= c_3(k_2, k_1) = c_3(-k_2, k_1 - k_2) = c_3(k_2 - k_1, -k_1) = \\ c_3(k_1 - k_2, -k_2) &= c_3(-k_1, k_2 - k_1) \end{aligned} \quad (3.60)$$

A partir de estas cinco ecuaciones aparece una división del plano  $k_1 k_2$  en seis

regiones donde se repite esta función y, en consecuencia, al conocer los cumulantes de tercer orden en cualquiera de estas seis regiones, representadas en la figura (3.4), se puede reconstruir la secuencia completa correspondiente a los cumulantes de tercer orden. Nótese que cada una de estas regiones contiene a su frontera. Así, por ejemplo, el sector 1 es una región infinita caracterizada por  $0 < \mathbf{k}_2 \leq \mathbf{k}_1$  (sector de  $45^\circ$  perteneciente al primer cuadrante). **Para procesos no estacionarios estas seis regiones de simetría desaparecen.** A partir de estas relaciones y de la definición del espectro de cumulantes de tercer orden se obtienen las siguientes relaciones en el dominio frecuencial bidimensional:

$$\begin{aligned} C_3(\omega_1, \omega_2) = C_3(\omega_2, \omega_1) = C_3(-\omega_2, -\omega_1) = C_3(-\omega_1 - \omega_2, \omega_2) = \\ C_3(\omega_1, -\omega_1 - \omega_2) = C_3(-\omega_1 - \omega_2, \omega_1) = C_3(\omega_2, -\omega_1 - \omega_2) \end{aligned} \quad (3.61)$$

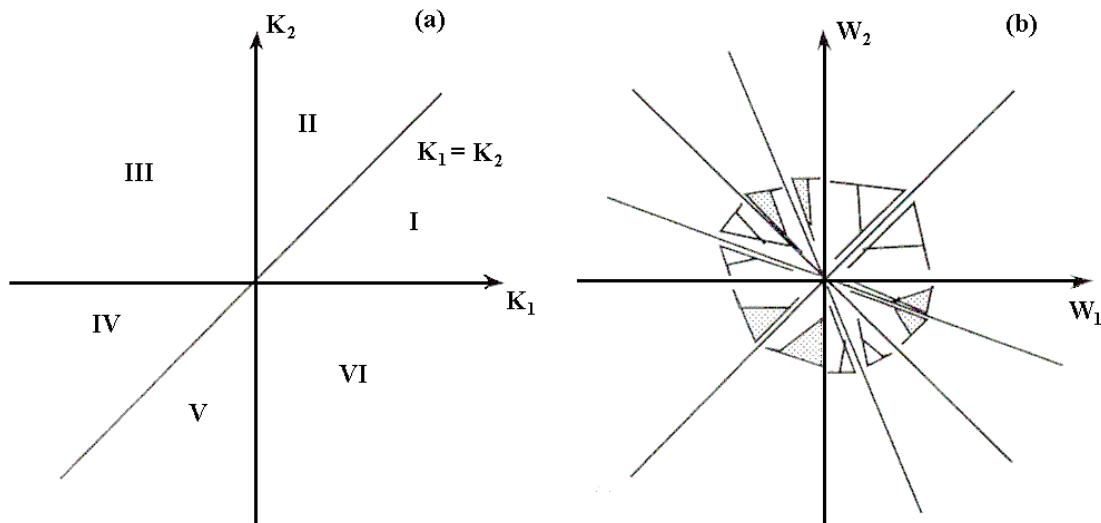


Fig.3.4: Regiones de simetría para: a) cumulantes de tercer orden; b) Biespectro

En la figura (3.4.b) se representan las 12 regiones de simetría del Biespectro cuando se consideran procesos estocásticos reales y, análogamente al dominio temporal, el conocimiento del Biespectro en la región triangular  $\omega_2 \geq 0, \omega_1 \geq \omega_2, \omega_1 + \omega_2 \leq \pi$  es suficiente para una total reconstrucción del Biespectro. Nótese que, en el dominio frecuencial las regiones de simetría presentan un área finita y en ellas, en general, el Biespectro toma valores complejos y, consecuentemente, no se destruye la información de fase.

### 3) Triespectro:

$$C_4(\omega_1, \omega_2, \omega_3) = \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} \sum_{k_3=-\infty}^{\infty} c_4(k_1, k_2, k_3) \cdot e^{-j(\omega_1 k_1 + \omega_2 k_2 + \omega_3 k_3)}$$

$$|\omega_1| \leq \pi, |\omega_2| \leq \pi, |\omega_3| \leq \pi, |\omega_1 + \omega_2 + \omega_3| \leq \pi \quad (3.62)$$

Donde  $c_4(k_1, k_2, k_3)$  representa la secuencia de cumulantes de cuarto orden. Al combinar la definición del Triespectro y la de los cumulantes de cuarto orden se pueden deducir 96 regiones de simetría [Pflu-92], cuando se evalúan procesos reales.

A partir de los espectros de cumulantes de orden superior, en el dominio frecuencial, se pueden recuperar las expresiones de sus respectivos cumulantes, en el dominio temporal, aplicando la Transformada de Fourier Inversa de orden  $n$ :

$$c_n(k_1, \dots, k_{n-1}) = \frac{1}{(2 \cdot \pi)^{n-1}} \cdot \int_{-\pi}^{\pi} \dots \int_{-\pi}^{\pi} C_n(\omega_1, \dots, \omega_{n-1}) \cdot e^{j(\omega_1 k_1 + \dots + \omega_{n-1} k_{n-1})} \cdot d\omega_1 \cdot \dots \cdot d\omega_{n-1}$$

$$(3.63)$$

### 3.3.3.-Estimación de los cumulantes y sus poliespectros

En una situación real de aplicación no se dispone de un conjunto infinito de valores de un proceso determinado sino, de un conjunto finito de  $N$  valores o muestras del proceso. Además, algunos procesos solo presentan estacionariedad durante intervalos relativamente cortos de tiempo, resultando necesaria la obtención de los cumulantes de una forma adaptativa a lo largo del tiempo, reduciéndose, aún más, el conjunto de valores disponibles del proceso  $\mathbf{x}(k)$ .

En estas condiciones se impone, pues, la necesidad de una estimación de los cumulantes, o sus poliespectros asociados, a partir de un conjunto finito de observaciones de este determinado proceso. En este apartado la discusión se centra en el caso de la estimación de tercer orden, pudiéndose deducir el caso de cuarto orden por simple extensión a partir del caso de tercer orden. En principio, existen dos estrategias básicas para estimar estos poliespectros:

- 1) métodos convencionales o tipo Fourier,
- 2) métodos paramétricos fundamentados en modelos ARMA, AR o MA.

Los métodos convencionales aplicados a la estimación de un proceso no Gaussiano

presentan las ventajas de facilidad de implementación y fidelidad de la estimación cuando se dispone de grandes registros de datos correspondientes al proceso a estimar. Sin embargo, su capacidad para discernir componentes armónicas en el dominio biespectral es limitada debido al "Principio de Incertidumbre" de la Transformada de Fourier. Además, para el caso de procesos paramétricos estos métodos convencionales obtienen una fidelidad biespectral bastante pobre. Nótese que la voz puede modelarse con bastante fidelidad por un modelado paramétrico Autorregresivo (AR).

Debido a esto, en nuestro sistema utilizaremos los métodos paramétricos, que pasamos a explicar a continuación.

### 3.3.3.1.-Estimadores Paramétricos

Tenemos un conjunto finito de datos  $\{x(1), x(2), \dots, x(N)\}$ , donde  $x(k)$  es un proceso AR real y de orden  $P$  descrito por:

$$\sum_{i=0}^P a_i \cdot x(k-i) = w(k) \quad \text{con } a_0 = 1 \quad (3.64)$$

$w(k)$  representa un proceso independiente e idénticamente distribuido de media nula, es decir  $E\{w(k)\}=0$ , y con momentos de segundo y tercer orden dados por:

$$\begin{aligned} E\{w(k), w(k+k_1)\} &= Q \cdot \delta(k_1) \\ E\{w(k), w(k+k_1), w(k+k_2)\} &= \beta \cdot \delta(k_1, k_2) \end{aligned} \quad (3.65)$$

y  $x(k')$  es independiente respecto  $w(k)$  para todo  $k' < k$ . Además,  $w(k)$  y  $x(k)$  no son procesos Gaussianos. Aplicando la definición de cumulantes de tercer orden, en este caso:

$$c_3(k_1, k_2) = E\{x(k), x(k+k_1), x(k+k_2)\} \quad (3.66)$$

Así, obtenemos la siguiente recursión de tercer orden:

$$c_3(-k_1, -k_2) + \sum_{i=1}^P a_i \cdot c_3(i-k_1, i-k_2) = \beta \cdot \delta(k_1, k_2) \quad k_1, k_2 \geq 0 \quad (3.67)$$

donde  $\delta(k_1, k_2)$  es la función impulso unidad bidimensional. A partir de (3.67),

tomando la recta  $\mathbf{k}_1 = \mathbf{k}_2$  aparecen  $2P+1$  valores de los cumulantes de tercer orden que satisfacen la ecuación [Ragh-85], [Ragh-86]:

$$\underline{\underline{R_c}} \cdot \underline{a} = \underline{\beta} \quad (3.68)$$

Donde:

$$\underline{\underline{R_c}} = \begin{pmatrix} c_3(0,0) & c_3(1,1) & \dots & c_3(p,p) \\ c_3(-1,-1) & c_3(0,0) & \dots & c_3(p-1,p-1) \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ c_3(-p,-p) & c_3(-p+1,-p+1) & \dots & c_3(0,0) \end{pmatrix} \quad (3.69)$$

$$\underline{a} = [1, a_1, a_2, \dots, a_p]^T \quad (3.70)$$

$$\underline{\beta} = [\beta, 0, 0, \dots, 0]^T \quad (3.71)$$

Esta matriz  $\underline{\underline{R_c}}$  es Toeplitz pero, en general, no es simétrica. Además si se escribe el vector de parámetros de la forma  $\underline{a} = [a_p, a_{p-1}, \dots, a_1, 1]^T$ , entonces  $\underline{\underline{R_c}}$  es una matriz de Hankel. Otra posible representación de (3.67) consiste en permitir valores de  $(\mathbf{k}_1, \mathbf{k}_2)$  pertenecientes al primer sector de la figura (3.4.a), siendo ahora una región triangular porque se dispone de un conjunto finito de datos pertenecientes al proceso  $\mathbf{x}(\mathbf{k})$ :

$$c_3(-k_1, -k_2) + \sum_{i=1}^p a_i \cdot c_3(i - k_1, i - k_2) = \beta \cdot \partial(k_1, k_2) \quad (3.72)$$

$$k_1 = 0, 1, \dots, L_1$$

$$k_2 = \begin{cases} 0, 1, \dots, k_1 & \text{si } k_1 < L_1 \\ 0, 1, \dots, L_2 & \text{si } k_1 = L_1 \end{cases}$$

Donde  $(k_1, k_2)$  se eligen de manera tal que se verifique:

$$L_2 \leq L_1$$

$$p = 1 + L_2 + \frac{(L_1 - 1) \cdot (L_1 + 2)}{2} \quad (3.73)$$

La matriz correspondiente a las ecuaciones anteriores no se corresponde con una

matriz de Toeplitz, pero representa un caso más general que (3.68) porque obtiene la información a partir de un área finita de los cumulantes y no a lo largo de una línea recta  $\mathbf{k}_1 = \mathbf{k}_2$ . La expresión representada en (3.68) obtiene el modelado AR a partir de **2P+1** valores de la secuencia de cumulantes o momentos de tercer orden, pertenecientes a la recta  $\mathbf{k}_1 = \mathbf{k}_2$  de la figura (3.4.a): uno correspondiente al origen,  $\mathbf{k}_1 = \mathbf{k}_2 = \mathbf{0}$ , y **P** puntos a cada lado del origen. Este modelo se ajusta al proceso en el sentido de ajuste perfecto entre la secuencia de momentos de tercer orden de la salida del filtro generador AR y las muestras disponibles del proceso en estos instantes correspondientes, es decir, si se encuentra un modelo AR de orden **P**, impulsado por ruido blanco no Gaussiano, cuya secuencia de momentos de tercer orden a su salida se iguale a las muestras dadas del proceso para los puntos  $\mathbf{k}_1 = \mathbf{k}_2 = \mathbf{0}, \pm\mathbf{1}, \dots, \pm\mathbf{p}$ , entonces, sus parámetros  $\underline{\mathbf{a}}$  verificarían (3.68) como condición necesaria. Si se considera la expresión (3.72) se precisan más de **2P+1** valores para ajustar un modelo AR según el criterio anterior. Si las muestras  $\mathbf{x}(\mathbf{k})$  disponibles han sido generadas a partir de valores reales de la secuencia de momentos de tercer orden correspondientes a un proceso de orden **P** que satisface todas las condiciones anteriormente especificadas, entonces, los parámetros  $\underline{\mathbf{a}}$  obtenidos en ambos casos son los mismos. En cualquier otra situación las dos soluciones anteriores pueden ser distintas.

A continuación se presentan algunos métodos para estimar el Biespectro de un proceso  $\mathbf{x}(\mathbf{k})$  mediante un modelado AR. Estos dos métodos se diferencian en la forma de estimar los cumulantes de tercer orden, a partir del conjunto finito de **N** muestras del proceso  $\mathbf{x}(\mathbf{k})$ , para resolver el sistema de ecuaciones de tercer orden (3.67). En todos estos métodos se consideran unas condiciones comunes de trabajo: un proceso  $\mathbf{v}(\mathbf{k})$ , uniformemente distribuido, independiente y no Gaussiano, excita un modelo  $\mathbf{h}(\mathbf{k})$  paramétrico AR y causal, cuya fase puede no ser mínima, y a su salida se obtiene un proceso  $\mathbf{x}(\mathbf{k})$  no Gaussiano, del cual se dispone un conjunto finito de **N** muestras. Normalmente el ruido  $\mathbf{v}(\mathbf{k})$  y su distribución se suponen desconocidos.

### **3.3.3.1.1.-Método recursivo de tercer orden (TOR)**

La matriz  $\underline{\underline{\mathbf{R}}}_c$  que aparece en la expresión (3.68) es Toeplitz pero, en general, no es simétrica. Si el filtro AR es estable, luego, esta ecuación (3.68) existe. Una condición

suficiente, aunque no necesaria, para la estabilidad del filtro AR impone que  $\underline{\underline{R}_c}$  sea una matriz Toeplitz, simétrica y definida positiva [Makh-75]. De este modo se puede asegurar la estabilidad de las representaciones AR, sólo para aquellos procesos cuya matriz de cumulantes de tercer orden verifique las tres condiciones previamente citadas.

Sin embargo, en el caso que nos ocupa se dispone de un conjunto limitado de datos, N. Así, se consideran las **P+1** ecuaciones compuestas cuando  $\mathbf{k}_1 = \mathbf{k}_2$  (3.68) y la aproximación de los cumulantes de tercer orden mediante  $\hat{c}_3(\mathbf{k}_1, \mathbf{k}_2)$ , que se obtienen realizando los siguientes pasos:

1) Segmentar el conjunto de N valores en K registros con M valores cada uno,  
 $N=K \cdot M$ .

2) Sustraer la media  $\mathbf{m}_1$  de cada registro, si la hubiera.

3) tomando  $\{x^i(k), k = 0, 1, \dots, M-1\}$  como el conjunto de valores correspondiente al segmento i-ésimo, se obtiene la secuencia de cumulantes o momentos de tercer orden:

$$r^i(k_1, k_2) = \frac{1}{M} \cdot \sum_{k=S_1}^{S_2} x^i(k) \cdot x^i(k+k_1) \cdot x^i(k+k_2) \quad (3.74)$$

Donde:

$$\begin{aligned} i &= 1, 2, \dots, K \\ S_1 &= \max(0, -k_1, -k_2) \\ S_2 &= \min(M-1, M-1-k_1, M-1-k_2) \end{aligned}$$

4) Promediar los valores obtenidos para cada uno de los K segmentos:

$$\hat{c}_3(k_1, k_2) = \frac{1}{K} \cdot \sum_{i=1}^K r^i(k_1, k_2) \quad (3.75)$$

Así, obtenemos el siguiente sistema:

$$\hat{\underline{\underline{R}}}_c \cdot \hat{\underline{a}} = \hat{\underline{\beta}} \quad (3.76)$$

$$\hat{\underline{R}}_c = \begin{pmatrix} \hat{c}_3(0,0) & \hat{c}_3(1,1) & \cdots & \hat{c}_3(p,p) \\ \hat{c}_3(-1,-1) & \hat{c}_3(0,0) & \cdots & \hat{c}_3(p-1,p-1) \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \hat{c}_3(-p,-p) & \hat{c}_3(-p+1,-p+1) & \cdots & \hat{c}_3(0,0) \end{pmatrix} \quad (3.77)$$

$$\hat{\underline{a}} = [1, \hat{a}_1, \hat{a}_2, \dots, \hat{a}_p]^T \quad (3.78)$$

$$\hat{\underline{\beta}} = [\hat{\beta}, 0, 0, \dots, 0]^T \quad (3.79)$$

Siendo  $\hat{\underline{a}}$  una estimación de los parámetros AR y la estimación del momento de tercer orden correspondiente al ruido blanco se representa por  $\hat{\underline{\beta}}$ . También se pueden aplicar las ecuaciones (3.72) para la estimación  $\hat{\underline{a}}$  de los parámetros AR. Debe remarcarse que la aplicación de este método presupone la ergodicidad del proceso  $\mathbf{x}(k)$ .

### 3.3.3.1.2.-Método de los momentos promedio de tercer orden (CTOR)

Si se forma la función de tercer orden:

$$\hat{q}^k(k_1, i) = x(k-i) \cdot x^2(k-k_1) \quad i, k_1 = 1, \dots, p \quad (3.80)$$

aplicando el operador Esperanza se verifica:

$$E\{\hat{q}^k(k_1, i)\} = c_3(i - k_1, i - k_1) \quad (3.81)$$

Si en lugar de disponer de la muestras del proceso  $\mathbf{x}(k)$  se dispusiera de muestras pertenecientes a su secuencia de cumulantes de tercer orden  $c_3(k_1, k_2)$ , entonces, el proceso AR de orden  $P$  se ajustaría al proceso dado mediante la resolución de las ecuaciones (3.67):

$$E\left\{\hat{q}^k(k_1, 0) + \sum_{i=0}^p \hat{a}_i \hat{q}^k(k_1, i)\right\} = 0 \quad k_1 = 1, \dots, P \quad (3.82)$$

Si la expresión interior del operador se nota como  $\hat{e}_3(k_1, k_2)$ :

$$E\{\hat{e}_3(k_1, k_2)\} = 0 \quad \begin{matrix} k_1 = 1, 2, \dots, P \\ k = P + 1, P + 2, \dots, N \end{matrix} \quad (3.83)$$

Donde  $\hat{e}_3(k_1, k_2)$  se refiere al proceso error de predicción de tercer orden. A partir de las  $N$  muestras disponibles del proceso  $\mathbf{x}(k)$  se pueden obtener  $N-P$  valores de  $\hat{e}_3(k_1, k_2)$  por cada valor de  $k_1$ . Para el caso que nos ocupa, el operador Esperanza se traduce en un cálculo de la media de la secuencia error, y se llega a un sistema de  $P$  ecuaciones lineales, cuya resolución conduce a los parámetros  $\hat{a}$  estimados:

$$\frac{1}{N-P} \cdot \sum_{k=P+1}^N \hat{e}_3(k, k_1) = 0 \quad k_1 = 1, \dots, P \quad (3.84)$$

Que admite la siguiente representación matricial:

$$\underline{\hat{Q}} \cdot \underline{\hat{a}} = \underline{\hat{b}} \quad (3.85)$$

Donde:

$$\underline{\hat{Q}} = \begin{pmatrix} \hat{q}_{11} & \cdot & \cdot & \cdot & \cdot & \hat{q}_{1P} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \hat{q}_{P1} & \cdot & \cdot & \cdot & \cdot & \hat{q}_{PP} \end{pmatrix} \quad (3.86)$$

$$\underline{\hat{a}} = [\hat{a}_1, \dots, \hat{a}_P]^T \quad (3.87)$$

$$\underline{\hat{b}} = [\hat{q}_{10}, \dots, \hat{q}_{P0}]^T \quad (3.88)$$

$$\hat{q}_{ij} \equiv \sum_{k=P+1}^N \hat{q}^k(i, j) \quad (3.89)$$

Los resultados obtenidos mediante este algoritmo son significativos si se verifica, como condición necesaria, la ergodicidad de tercer orden para el proceso  $\mathbf{x}(k)$ .

### 3.3.3.1.2.-Método AR Optimizado (OARM)

Este algoritmo se debe a An, Kim y Powers [An-88] y consiste en una extensión de la metodología TOR hacia un sistema sobredeterminado de ecuaciones. Se considera la ecuación recursiva de tercer orden (3.67) y se toman  $P+I$  posibles valores para:

$$k_1, k_2 = 0, 1, \dots, P \quad (3.90)$$

resultando la siguiente formulación matricial:

$$\underline{r} = \begin{pmatrix} c_3(0,0) & c_3(1,1) & \cdots & c_3(P,P) \\ c_3(0,-1) & c_3(1,0) & \cdots & c_3(P,P-1) \\ \cdots & \cdots & \cdots & \cdots \\ c_3(0,-P) & c_3(1,1-P) & \cdots & c_3(P,0) \\ c_3(-1,0) & c_3(0,1) & \cdots & c_3(P-1,P) \\ \cdots & \cdots & \cdots & \cdots \\ c_3(-1,-P) & c_3(0,1-P) & \cdots & c_3(P-1,0) \\ \cdots & \cdots & \cdots & \cdots \\ c_3(-P,0) & c_3(1-P,1) & \cdots & c_3(0,P) \\ \cdots & \cdots & \cdots & \cdots \\ c_3(-P,-P) & c_3(1-P,1-P) & \cdots & c_3(0,0) \end{pmatrix} \quad (3.91)$$

$$\underline{a} = [1, a_1, \dots, a_p]^T \quad (3.92)$$

$$\underline{b} = [\beta, 0, \dots, 0]^T \quad (3.93)$$

Nótese que  $\underline{r}$  es una matriz de  $(P+1)^2$  filas y  $P+1$  columnas que contiene todos los cumulantes de tercer orden. Al tratarse de un sistema sobre determinado, la solución se obtiene aplicando el criterio de Mínimo Error Cuadrático Medio:

$$\hat{\underline{a}} = (\underline{r}^T \cdot \underline{r})^{-1} \cdot \underline{r}^T \cdot \underline{b} \quad (3.94)$$

Mediante la resolución de la ecuación (3.94) se obtienen los parámetros AR. Este algoritmo origina una buena estimación especialmente para el caso de disponer de un conjunto pequeño de datos  $\{\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(N)\}$  y/o para entornos altamente ruidosos. Bajo estas condiciones de trabajo, el método TOR ofrece pobres prestaciones según se demuestra en [An-88].

### 3.3.3.1.4.-Método de las ecuaciones de Yule-Walker de orden superior

Giannakis demostró que esta metodología ofrece siempre una solución y ésta es única [Gian-90]. En principio, el desarrollo analítico correspondiente a este algoritmo se

deduce para cualquier orden  $n$  de las estadísticas de orden superior consideradas, no restringiéndose al caso más simple de tercer orden. Se consideran los cumulantes de orden  $n \geq 3$  con  $n-3$  grados de libertad fijados a cero y, por simplicidad, se hace uso de la siguiente nomenclatura:

$$d_n(k_1, k_2) = c_n(k_1, k_2, 0, \dots, 0) \quad (3.95)$$

La fórmula de Brillinger-Rosenblatt, particularizada para esta situación es:

$$d_n(k_1, k_2) = \gamma_n^v \cdot \sum_{k=0}^{\infty} h^{n-2}(k) \cdot h(k + k_1) \cdot h(k + k_2) \quad (3.96)$$

Por otra parte, el filtro AR satisface la recursión:

$$h(k_1, k_2) = -\sum_{i=1}^p a_i \cdot h(k + k_1 - i) + b(k + k_1) \quad (3.97)$$

y sustituyendo en (3.96);

$$d_n(k_1, k_2) + \sum_{i=1}^p a_i \cdot d_n(k_1 - i, k_2) = \gamma_n^v \cdot \sum_{k=0}^{\infty} h^{n-2}(k) \cdot b(k + k_1) \cdot h(k + k_2) \quad (3.98)$$

Como el modelo considerado es AR, entonces, el término  $b(k + k_1)$  se anula siempre para  $k_1 > 0$  y se obtienen las **ecuaciones de Yule-Walker** en el dominio de los cumulantes de orden superior:

$$\sum_{i=1}^p a_i \cdot d_n(k_1 - i, k_2) = -d_n(k_1, k_2) \quad \begin{array}{l} k_1 > 0 \\ k_2 \geq 0 \end{array} \quad (3.99)$$

Fijando  $k_2$  y tomando  $k_1 = 1, \dots, P$  aparece un sistema lineal de  $P$  ecuaciones:

$$Q_{k_2} \cdot \underline{a} = \underline{b}_{k_2} \quad (3.100)$$

Donde:

$$\underline{\underline{Q}}_{k_2} = \begin{pmatrix} d_n(0, k_2) & d_n(-1, k_2) & \dots & d_n(1 - P, k_2) \\ d_n(1, k_2) & d_n(0, k_2) & \dots & d_n(2 - P, k_2) \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ d_n(P - 1, k_2) & d_n(P - 2, k_2) & \dots & d_n(0, k_2) \end{pmatrix} \quad (3.101)$$

$$\underline{a} = [a_1, a_2, \dots, a_P]^T \quad (3.102)$$

$$\underline{b}_{k_2} = [d(1, k_2), d(2, k_2), \dots, d(p, k_2)]^T \quad (3.103)$$

En [Gian-89] se demuestra que esta matriz presenta rango completo  $P$  y origina una solución única cuando se toman  $P+1$  slices correspondientes a los valores  $\mathbf{k}_2 = -P, 1-P, \dots, -1, 0$ . En principio, parece que un valor de  $\mathbf{k}_2$  de la secuencia de cumulantes de orden  $n$  podría ser suficiente puesto que origina  $P$  ecuaciones para las  $P$  incógnitas  $\underline{a}$ , esta posibilidad se descarta posteriormente [Swam-89]. Así, se puede afirmar que ningún valor concreto de  $\mathbf{k}_2$  proporciona una matriz  $\mathbf{PxP}$  de Hankel que sea de rango completo  $P$ . Por esta razón se debe considerar el siguiente sistema:

$$\underline{Q} \cdot \underline{a} = \underline{b} \quad (3.104)$$

Donde:

$$\underline{\underline{Q}} = \begin{pmatrix} d_n(0,0) & d_n(-1,0) & \dots & d_n(1 - P, 0) \\ \dots & \dots & \dots & \dots \\ d_n(0,-P) & d_n(-1,-P) & \dots & d_n(1 - P, -P) \\ d_n(1,0) & d_n(0,0) & \dots & d_n(2 - P, 0) \\ \dots & \dots & \dots & \dots \\ d_n(1,-P) & d_n(0,-P) & \dots & d_n(2 - P, -P) \\ \dots & \dots & \dots & \dots \\ d_n(p - 1,0) & d_n(p - 2,0) & \dots & d_n(0,0) \\ \dots & \dots & \dots & \dots \\ d_n(P - 1,-P) & d_n(P - 2,-P) & \dots & d_n(0,-P) \end{pmatrix} \quad (3.105)$$

$$\underline{a} = [a_1, a_2, \dots, a_P]^T \quad (3.106)$$

$$\underline{b} = [d_n(1,0), \dots, d_n(1,-P), d_n(2,0), \dots, d_n(2,-P), \dots, d_n(p,0), \dots, d_n(p,-P)]^T \quad (3.107)$$

Obsérvese que esta matriz  $\underline{\mathbf{Q}}$  está formada por  $\mathbf{P} \cdot (\mathbf{P} + 1)$  filas y  $\mathbf{P}$  columnas y, aplicando las propiedades de simetría de los cumulantes de orden  $\mathbf{n}$ , precisa calcular los siguientes valores de los cumulantes:

$$d_n(0, \dots, 2P, 0, \dots, P) \quad \text{y} \quad d_n(2P, P) \quad (3.108)$$

En nuestro proyecto hemos utilizado este método porque nos proporciona una solución única, y presenta menor dificultad que los demás método explicados aquí.

### 3.3.4.-Modelo autorregresivo de orden 2, orden 3, orden 4

Cojamos una vocal, /e/ y apliquemos los modelos de segundo, tercero y cuarto orden bajo diferentes condiciones de ruido. Las figuras 3.5, 3.6 y 3.7 nos muestran el resultado del modelado.

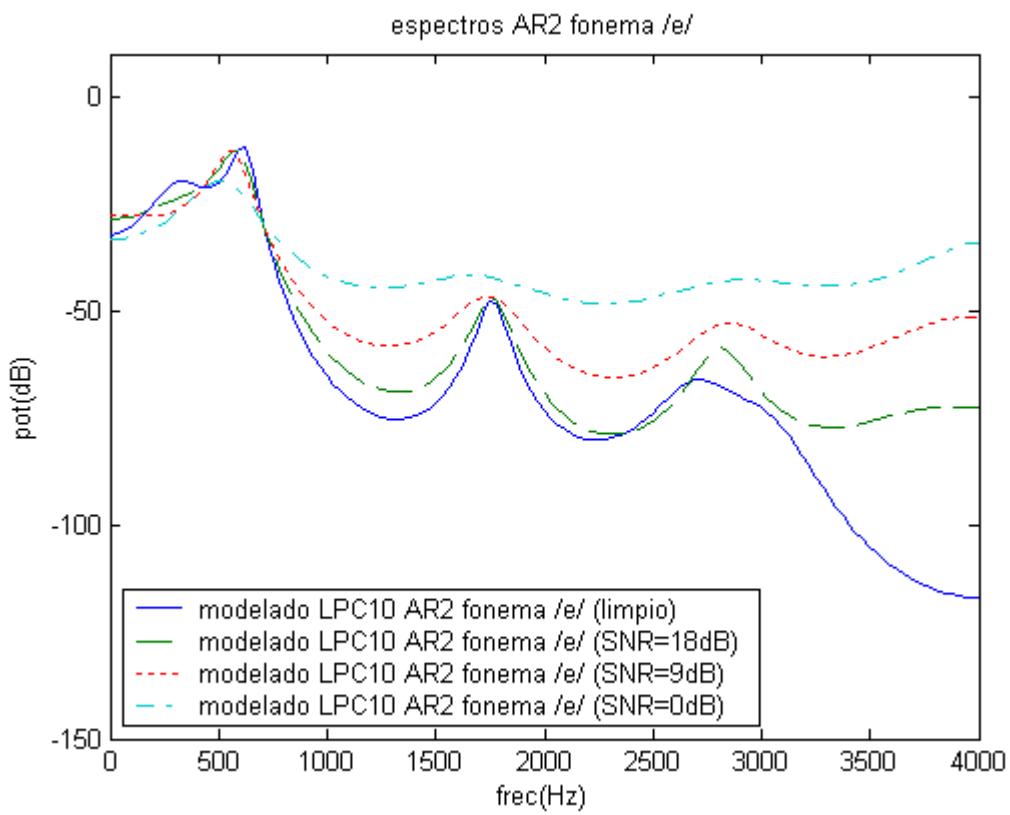


Fig.3.5: Modelado de una trama por el método de correlaciones.

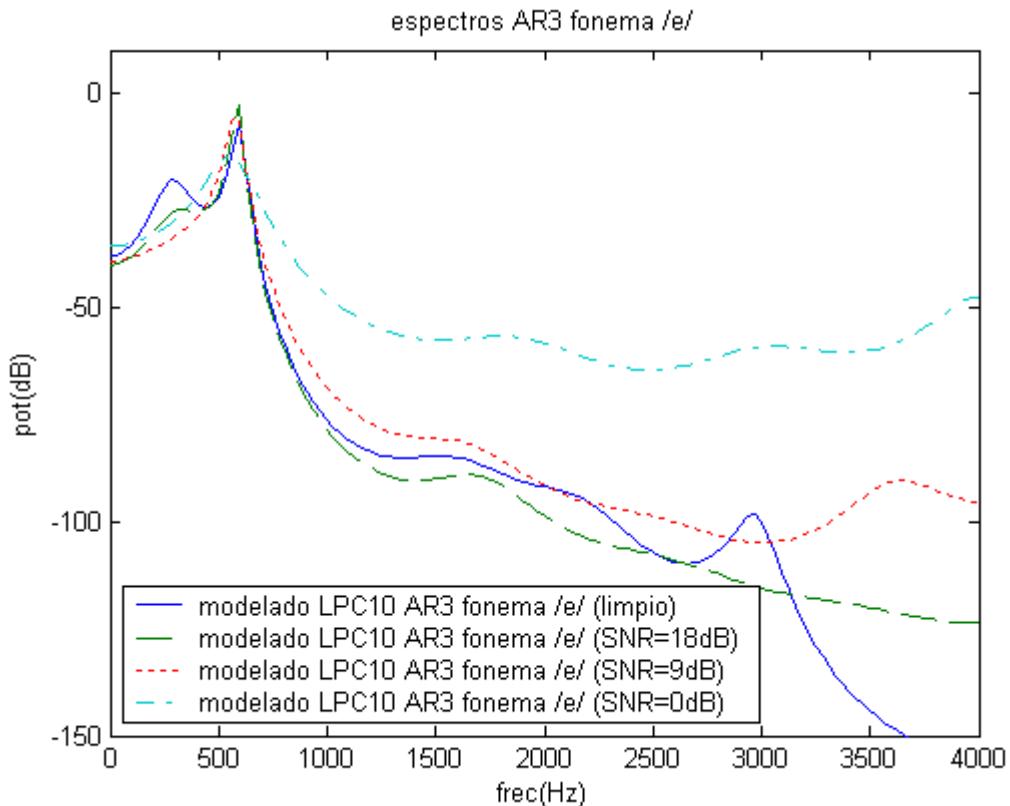


Fig.3.6: Modelado de una trama por el método de cumulantes de tercer orden.

Con trazo continuo se muestra la señal limpia, las líneas discontinuas son los casos de relación señal a ruido ( $S/N$ ) = 0, 9, 18 dB. El método de las correlaciones sigue exactamente la señal de entrada, sin determinar la presencia del ruido; así para 18 dB su estimación es casi exactamente el valor real para la señal limpia. Sin embargo, al aumentar el ruido presente en la señal de voz sintética, ( $S/N$ ) disminuyendo, el modelado LPC de segundo orden pierde paulatinamente de vista la vocal y, por tanto, deja de conocer con exactitud la posición de los formantes, incluso del segundo para  $S/N = 0$  dB. A 9dB, el método de correlaciones, ya no puede determinar la posición del tercer formante; no obstante, sigue notando la presencia de un último pico.

Si observamos la figura 3.6, vemos como la estimación por cumulantes no se aleja de la original hasta que no bajamos a  $SNR=0$  dB, en cuyo caso incluso tiene problemas de estabilidad (para la resolución de problemas en la estabilidad consultar [Marp-87]). Hasta para dicha SNR no perdemos de vista los dos primeros formantes, manteniendo un exactitud aceptable en posición y forma; aunque para el tercer formante ya no podemos apuntar con precisión su posición, pero sabemos que existe (con AR2 los perdíamos totalmente de vista).

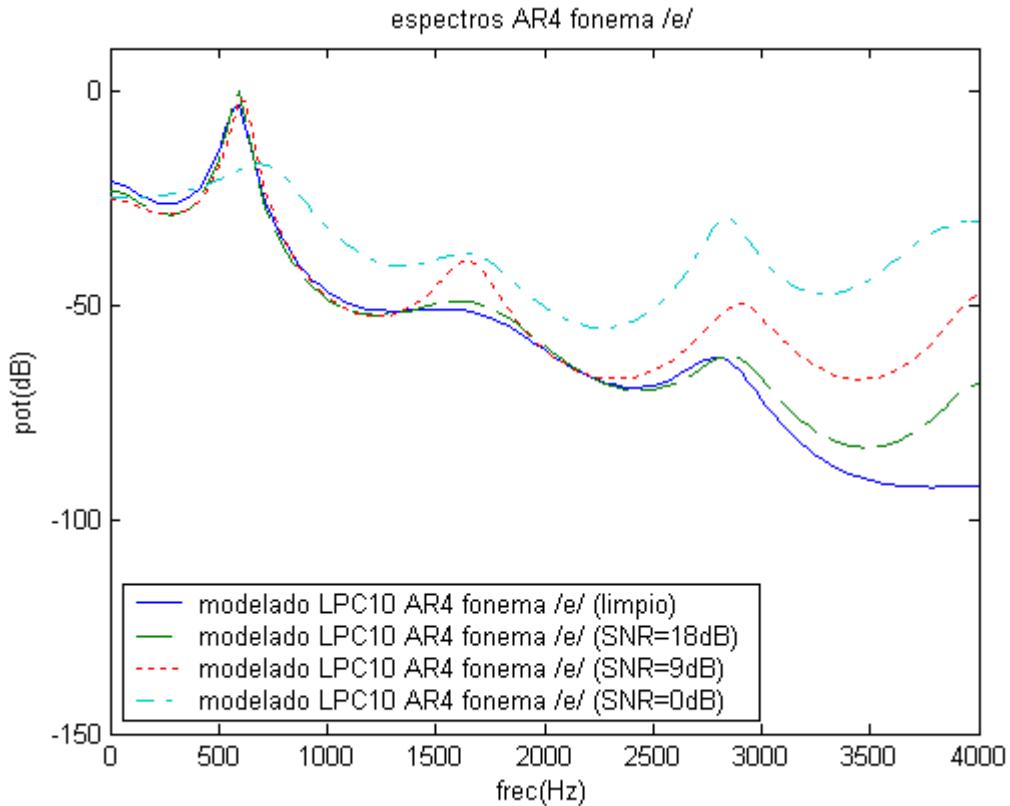


Fig.3.7: Modelado de una trama por el método de cumulantes de cuarto orden.

Finalmente, en la figura 3.7, podemos ver el método de cumulantes de cuarto orden. Al igual que para tercer orden, detecta la posición de todos los picos hasta relaciones señal a ruido de 9dB. Vemos que la trama seleccionada tiene un cuarto formante que no perdemos de vista. La diferencia más importante está en el paso a 0 dB. Ahora, aún perdiendo el nivel de amplitud de los formantes de alta frecuencia, seguimos posicionándolos con bastante fiabilidad.

Haciendo una comparación directa entre los tres ordenes, diremos que con correlaciones estamos siguiendo a la señal más ruido, mientras que los métodos de cumulantes intentan desacoplar señal de ruido. Así con el modelo *LP* mediante correlaciones, al aumentar el nivel de ruido, perdemos rápidamente la amplitud de los formantes, su posición puede resistir un poco más (obsérvese que al pasar de ruido blanco a un ruido pasó banda se castigaría muchísimo los formantes de la zona). Los cumulantes de tercer y cuarto orden nos dan la posición y amplitud conjuntamente hasta una SNR más baja; aquí cometemos ligeros errores en los valles, pero determinamos muy bien los picos. En el paso a 0 dB, en la zona de alta frecuencia, (menor relación señal a ruido en  $\omega$ ) sólo el método de cumulantes de cuarto orden puede dar

información del tercer y cuarto formante, frente a una pérdida de estabilidad para el método de tercer orden.

Frente a ruido gaussiano blanco ensuciando un sonido sonoro el sistema funciona aceptablemente con cumulantes. No obstante, hemos de recordar que éste es el caso que nos ha llevado a utilizar estadísticas de orden superior (señal de voz no-gaussiana frente a ruido gaussiano).



## 4.-Efectos del filtrado de Wiener con modelos AR

Como ya comentamos anteriormente al utilizar el filtro de Wiener (con estimación de los coeficientes por predicción lineal) para intentar extraer el ruido de la voz recibida surgen ciertos problemas. Estos se deben principalmente a la imposibilidad de disponer de más señal que la de entrada, ya contaminada con ruido.

Ahora examinaremos cada uno de los problemas intentando averiguar el porqué de dicho efecto indeseado. Algunos ya fueron presentados en [Jove-93] [Masg-92b].

### 4.1.-Reducción del ancho de banda de los formantes

El filtrado iterativo de Wiener, modelando el espectro de voz mediante la aplicación reiterada del algoritmo de estimación autorregresiva propuesto por Lim - Oppenheim, nos lleva a un espectro de la señal de salida caracterizado por la reducción del ancho de banda de los formantes. Mediante el método de correlaciones de segundo orden llegamos a un espectro como el de la figura (4.1).

El efecto se agudiza al disminuir la (S/N) disponible a la entrada. Para 0 dB el efecto es muy acusado, mientras que para 9 dB sólo se presenta de forma apreciable para el tercer formante.

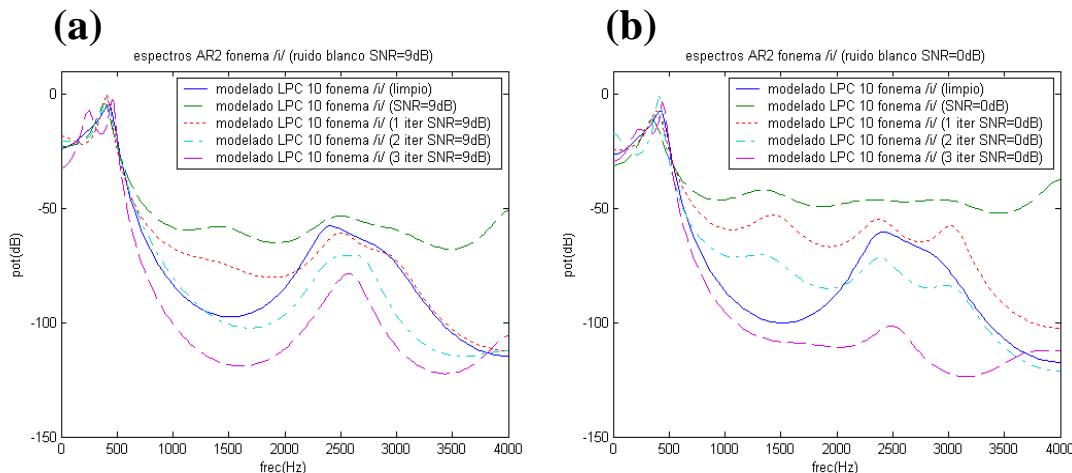


Fig.4.1: Envolvente LPC de una trama de voz sonora, vocal /i/. Evolución para iteraciones 1, 2, 3 y 4. (a) SNR=9 dB. (b) SNR=0 dB.

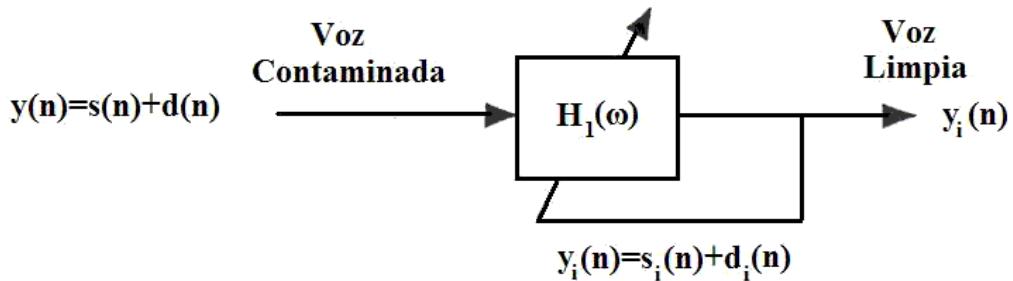
A continuación se expone una posible explicación para este fenómeno conocido como picado espectral o spectral peaking. Esta explicación fue presentada en [Masg-92b].

#### 4.1.1.-Efecto de picado en el método de correlaciones

Estamos usando el algoritmo de Lim – Oppenheim en un filtrado de Wiener iterativo. El filtro de Wiener no iterativo utilizado en el caso ideal, disponiendo de  $s(n)$  y  $d(n)$  por separado, tendría el aspecto siguiente:

$$H(\omega) = \frac{P_s(\omega)}{P_s(\omega) + P_d(\omega)} \quad (4.1)$$

El diagrama de bloques del filtrado de en el caso iterativo se muestra en la siguiente figura:



$$H_i(\omega) = \frac{P_{y_{i-1}}(\omega)}{P_{y_{i-1}}(\omega) + P_d(\omega)} \quad \text{donde:} \quad P_{y_{i-1}}(\omega) = \frac{g^2}{|1 + \sum a_k e^{-j\omega k}|}$$

Fig.4.2: Esquema del algoritmo iterativo de Wiener.

El filtro está en función de la estimación de  $P_s(\omega)$  que somos capaces de hacer, es decir, disponemos de:

$$y(n)=s(n)+d(n) \quad (4.2)$$

por lo que deberemos usar un filtro del tipo:

$$H(\omega) = \frac{P_y(\omega)}{P_y(\omega) + P_d(\omega)} \quad (4.3)$$

Así pues, haremos un estudio iterativo del filtrado, en el que intentaremos deducir la

relación entre el filtro en la iteración **i-ésima** y la **i+1-ésima**. En adelante, por comodidad, omitiremos la dependencia de los espectros de potencia con la frecuencia,  $\mathbf{P}(\mathbf{f}) = \mathbf{P}$ . Empezaremos por escribir el filtro inicial, tras la primera estimación de  $\mathbf{P}_s$ ,  $\hat{\mathbf{P}}_{s_1}$ . Para ello sólo disponemos de la señal con ruido,  $\mathbf{y}(\mathbf{n})$ :

$$H_1 = \frac{\hat{P}_{s_1}}{\hat{P}_{s_1} + \hat{P}_{d_1}} = \frac{P_y}{P_y + P_d} \quad (4.4)$$

Aplicando (4.2) en frecuencia:

$$P_y(\omega) = P_s(\omega) + P_d(\omega) \quad (4.5)$$

nos queda el filtro para la primera iteración:

$$H_1 = \frac{P_s + P_d}{P_s + P_d + P_d} = \frac{1}{1 + \frac{P_d}{P_s + P_d}} = \frac{1}{1 + H^c_{\text{ópt}}} \quad (4.6)$$

Donde se ha utilizado la definición del filtro óptimo de Wiener,  $H_{\text{ópt}}$  o de su complementario,  $H^c_{\text{ópt}}$ :

$$H_{\text{ópt}} = \frac{P_s}{P_s + P_d} \quad (4.7a)$$

$$H^c_{\text{ópt}} = 1 - H_{\text{ópt}} = \frac{P_d}{P_s + P_d} \quad (4.7b)$$

A simple vista se ve que se cumple siempre, dado que los espectros son positivos, las condiciones de rango:

$$\begin{aligned} 0 &\leq H_{\text{ópt}} \leq 1 \\ 0 &\leq H^c_{\text{ópt}} \leq 1 \end{aligned} \quad (4.8)$$

Para la segunda iteración obtendremos:

$$H_2 = \frac{\hat{P}_{s_2}}{\hat{P}_{s_2} + \hat{P}_{d_2}} = \frac{\hat{P}_{y_1}}{\hat{P}_{y_1} + P_d} \quad (4.9)$$

Arreglando la expresión:

$$H_2 = \frac{1}{1 + \frac{P_d}{\hat{P}_{y_1}}} = \frac{1}{1 + \frac{P_d}{\hat{P}_{s_1} + P_d}} \quad (4.10)$$

Hemos aplicado que la estimación a la salida del primer filtro se utiliza para crear el filtro de la segunda iteración:

$$\hat{P}_{y_1} = \hat{P}_{s_1} + P_d = (P_s + P_d) \cdot H_1^2 \quad (4.11)$$

Introduciendo (4.11) en (4.10) podemos expresar  $H_2$  como:

$$H_2 = \frac{1}{1 + \frac{P_d}{P_s + P_d} \cdot \frac{1}{H_1^2}} = \frac{1}{1 + H_{\text{opt}}^c \cdot (H_1^{-1})^2} \quad (4.12)$$

Para la tercera iteración con un proceso similar se llega a:

$$H_3 = \frac{1}{1 + H_{\text{opt}}^c \cdot (H_2^{-1})^2} \quad (4.13)$$

Y, en general, para la **i-ésima** iteración podemos hallar la fórmula de recurrencia siguiente:

$$H_i = \frac{1}{1 + H_{\text{opt}}^c \cdot (H_{i-1}^{-1})^2} \quad (4.14)$$

Teniendo en cuenta siempre que:

$$P_{y_i} = P_{y_{i-1}} \cdot H_1^2 \quad (4.15)$$

Ahora realizaremos un estudio de convergencia del filtrado iterativo de Wiener a partir de la ecuación de recurrencia a la que hemos llegado. Debemos recordar siempre que  $H_i$  representa a  $H_i(\omega)$  o  $H_i(f)$ . Para mayor comodidad en las siguientes expresiones reescribimos (4.14) como:

$$H_i = \frac{1}{D_i} \quad (4.16)$$

Donde:

$$D_i = 1 + H_{\text{opt}}^c \cdot D_{i-1}^2 \quad (4.17)$$

Puesto que:

$$D_{i-1}^2 = (H_{i-1}^{-1})^2 \quad (4.18)$$

Podemos escribir la ecuación de recurrencia más claramente tras realizar el siguiente cambio de variable:

$$d(i) = D_i(f)$$

$$r = H_{\text{opt}}^c \quad (4.19)$$

Entonces (4.17) se convierte en:

$$d(n) = 1 + r \cdot d^2(n-1) \quad (4.20)$$

Para que la recursión converja debe cumplirse:

$$d(n) = d(n-1) \quad \text{cuando } n \rightarrow \infty \quad (4.21)$$

con lo que nos queda la ecuación cuadrática:

$$d(\infty) = 1 + r \cdot d^2(\infty) \quad (4.22)$$

$$r \cdot d^2(\infty) - d(\infty) + 1 = 0$$

Resolviendo:

$$d(\infty) = \frac{1 \pm \sqrt{1 - 4 \cdot r}}{2 \cdot r} \quad (4.23a)$$

$$\begin{cases} r > 1/4 \Rightarrow d(\infty) \rightarrow \infty & \text{diverge} \\ r = 1/4 \Rightarrow d(\infty) = 2 \\ r < 1/4 \Rightarrow d(\infty) \text{ tiene dos soluciones} \end{cases} \quad (4.23b)$$

El caso **r>1/4** no tiene solución analítica, pero de un estudio numérico se desprende que  $d$  tiende a infinito cuando lo hace  $n$ , es decir, diverge. Anteriormente eliminamos la dependencia con la frecuencia en la ecuación de recurrencia y, por tanto, la solución es válida para una frecuencia concreta. Para tener la forma del filtro en todo el espectro de frecuencias debería particularizarse el resultado.

Analicemos el caso  $r < 1/4$ , para el cual:

$$\text{Si } r < 1/4 \Rightarrow 1 - 4 \cdot r > 0 \quad (4.24)$$

Pudiendo definir:

$$\begin{aligned} 1 - 4 \cdot r &= \frac{a^2}{b^2} \quad a < b \\ r &= \frac{b^2 - a^2}{4 \cdot b^2} \end{aligned} \quad (4.25)$$

Donde  $a$  y  $b$  son cualquier real que cumplan  $a < b$ , para que:

$$0 < \frac{b^2 - a^2}{4 \cdot b^2} = r < 1/4 \quad (4.26)$$

$$0 < r \quad r < 1/4$$

$$0 < b^2 - a^2 \quad 4 \cdot (b^2 - a^2) < 4 \cdot b^2$$

$$a^2 < b^2 \quad -4 \cdot a^2 < 0$$

Siempre se cumpla. Calculamos el valor de  $r(\infty)$ :

$$r(\infty) = \frac{\frac{1 \pm \sqrt{\frac{a^2}{b^2}}}{b^2 - a^2}}{2 \cdot b} = \begin{cases} r_A = \frac{2 \cdot b}{b - a} \\ r_B = \frac{2 \cdot b}{b + a} \end{cases} \quad (4.27)$$

En función del valor inicial utilizado en la recursión llegaremos a una u otra solución, o el sistema divergirá. A continuación se estudian las diferentes combinaciones:

a)

$$r(0) = \frac{2 \cdot b}{b - a} = r_A$$

$$r(1) = 1 + r \cdot r^2(0)$$

$$r(1) = 1 + \frac{b^2 - a^2}{4 \cdot b^2} \cdot \frac{4 \cdot b^2}{(b - a)^2} = \frac{2 \cdot b}{b - a} \quad (4.28)$$

$$\begin{aligned} r(1) &= r(0) \\ r(2) &= r(1) = r(0) \end{aligned}$$

.

.

$$r(\infty) = r(0)$$

b)

$$r_B < r(0) < \frac{2 \cdot b}{b - a}$$

$$r(0) = \frac{2 \cdot b}{b - a} - \xi = r_A - \xi \quad \xi < 0$$

$$r(1) = 1 + r \cdot r^2(0)$$

$$r(1) = r_A - \frac{\xi}{b} \cdot (b + a) \cdot \left[ 1 - \frac{\xi}{4 \cdot b^2} \cdot (b - a) \right] = r_A(0) - \xi'$$

$$\xi < \xi' \Rightarrow 1 < \frac{(b + a)}{b} \cdot \left[ 1 - \frac{\xi}{4 \cdot b^2} (b - a) \right]$$

$$\xi < \frac{4 \cdot b}{b - a} = \frac{4}{1 - \frac{a}{b}} = \frac{4}{1 - \sqrt{1 - 4 \cdot r}}$$

$$0 < r < 1/4 \Rightarrow \begin{cases} \xi < \infty \\ \xi < 4 \end{cases} \quad \text{que siempre es cierto}$$

$$r(1) < r(0) \quad (4.29)$$

$$r(n) < r(n-1)$$

c)

$$r_A > r(0) > \frac{2 \cdot b}{b + a}$$

$$r(0) = \frac{2 \cdot b}{b + a} + \xi = r_B + \xi \quad (4.30)$$

reducción al caso (b)

d)

$$r(0) > \frac{2 \cdot b}{b - a} = r_A$$

$$r(0) = \frac{2 \cdot b}{b - a} + \xi = r_A + \xi \quad \xi > 0$$

$$r(1) = 1 + r \cdot r^2(0)$$

$$r(1) = r_A + \frac{\xi}{b} \cdot (b + a) \cdot \left[ 1 + \frac{\xi}{4 \cdot b^2} \cdot (b - a) \right] = r_A(0) + \xi'$$

$$\xi < \xi' \Rightarrow 1 < \frac{(b + a)}{b} \cdot \left[ 1 + \frac{\xi}{4 \cdot b^2} (b - a) \right]$$

$$\xi > \frac{-4 \cdot b}{b - a} < 0$$

$$\xi > 0 \quad \text{que siempre es cierto} \quad (4.31)$$

$$r(1) > r(0)$$

e)

$$r(0) = \frac{2 \cdot b}{b + a} = r_B$$

$$r(1) = 1 + r \cdot r^2(0)$$

$$r(1) = 1 + \frac{b^2 - a^2}{4 \cdot b^2} \cdot \frac{4 \cdot b^2}{(b + a)^2} = \frac{2 \cdot b}{b + a} \quad (4.32)$$

$$\begin{aligned} r(1) &= r(0) \\ r(2) &= r(1) = r(0) \end{aligned}$$

.

.

$$r(\infty) = r(0)$$

f)

$$0 < r(0) < \frac{2 \cdot b}{b + a}$$

$$r(0) = \frac{2 \cdot b}{b + a} - \xi = r_B - \xi \quad \xi > 0$$

$$r(1) = 1 + r \cdot r^2(0)$$

$$r(1) = r_B - \frac{\xi}{b} \cdot (b - a) \cdot \left[ 1 - \frac{\xi}{4 \cdot b^2} \cdot (b + a) \right] = r_B - \xi'$$

$$\xi < \xi' \Rightarrow 1 > \frac{(b - a)}{b} \cdot \left[ 1 - \frac{\xi}{4 \cdot b^2} (b + a) \right]$$

$$\xi > \frac{-4 \cdot b}{b + a} < 0$$

$\xi > 0$  que siempre es cierto

(4.33)

$$r(1) > r(0)$$

En el caso (a) la recursión se queda clavada en el punto de partida,  $r_A$ . Si tomamos el valor inicial por debajo de  $r_A$  y superior a  $r_B$  convergerá siempre hacia  $r_B$  de manera monótona decreciente. Como se ve siempre se cumplirá que  $\xi$  sea inferior a 4, sino partiríamos de un valor inicial negativo, para nosotros no tiene sentido físico. El (c) es idéntico a (b) y la última alternativa nos lleva a  $r_B$  de manera monótona creciente. En (d) se muestra la divergencia si se toma  $r(0) > r_A$ , las sucesivas iteraciones se apartan de los  $r(\infty)$  calculados en (4.27). El caso (e) se clava en el valor inicial  $r_B$ . Podemos decir que para valores iniciales tomados en el intervalo  $(0, r_A)$  convergeremos a  $r_B$ , mientras que para  $(r_A, \infty)$  divergirá. Siendo la frontera,  $r_A$ , un punto de equilibrio inestable. La figura (4.2) ilustra las zonas de convergencia en función del valor inicial que tomemos, siempre para el caso  $r < 1/4$ . Se deduce que el ROC de  $d(0)$  es de la forma:

ROC :

$$d(0) < B(r)$$

$$d(0) < r_A = \frac{2 \cdot b}{b - a} = \frac{2}{1 - \frac{a}{b}} \quad (4.34)$$

$$d(0) < \frac{2}{1 - \sqrt{1 - 4 \cdot r}}$$

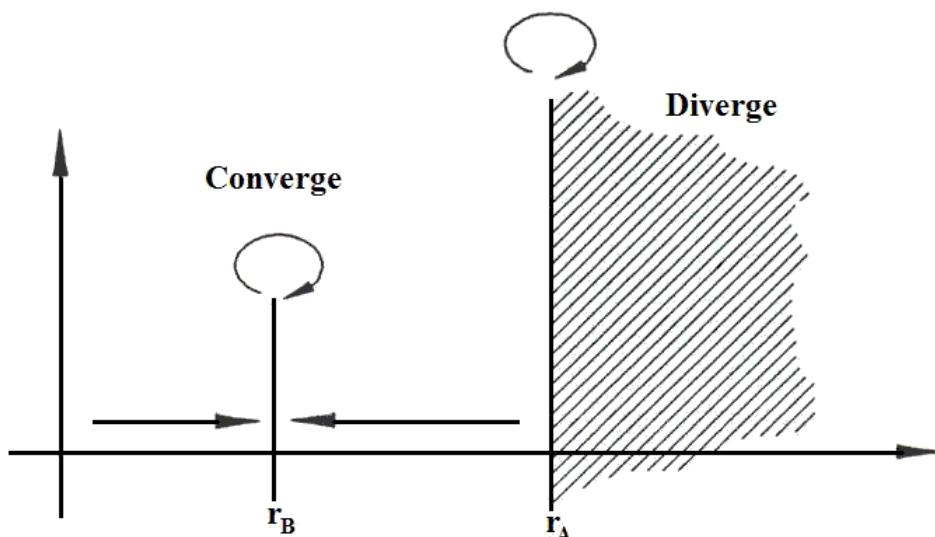


Fig.4.3: Zonas de convergencia para la elección del  $d(0)$ . [Jove-93].

Con  $\mathbf{B}(r)$  monótona decreciente con  $r$ , como se observa en la figura (4.3). Teniendo en cuenta que:

$$0 < r < 1/4$$

$$d(1) = H_1^{-1} \quad (4.35)$$

$$r = H^c_{\text{opt}}$$

Podemos estar seguros que siempre convergerá, ya que:

$$\begin{aligned} y(1) &= \frac{1}{H_1} = 1 + H^c_{\text{opt}} \\ y(1) &= 1 + r \cdot y(0) \end{aligned} \quad (4.36)$$

$$1 + H^c_{\text{opt}} = 1 + r \cdot y(0)$$

Uniendo (4.35) y (4.36) concluimos:

$$y(0) = 1 \in \text{ROC}[y(0) < 2] \quad (4.37)$$

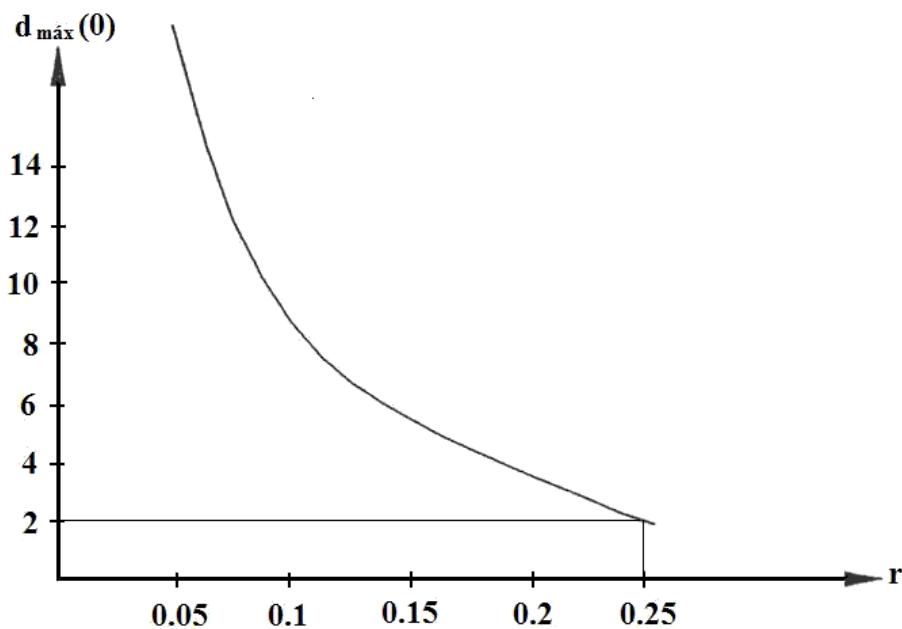


Fig.4.4: Valor máximo de  $d(0)$  para un  $r$  determinado. [Jove-93].

A partir de las siguientes igualdades entre la relación señal a ruido a la entrada, el filtro óptimo de Wiener y el filtro para la iteración infinita:

$$(S/N)_i = 10 \cdot \log \left( \frac{P_s}{P_d} \right)$$

$$H_{\text{opt}}(f) = \frac{P_s(f)}{P_s(f) + P_d(f)} \quad (4.38)$$

$$H_\infty(f) = \frac{2 \cdot (1 - H_{\text{opt}}(f))}{1 - \sqrt{1 - 4 \cdot (1 - H_{\text{opt}}(f))}}$$

Podemos calcular algunos valores de  $H_{\text{opt}}(f)$  en comparación al valor que debería tener,  $H_\infty(f)$ . Ver la tabla (4.1).

SNR	$P_s$ vs. $P_d$	$H_{\text{opt}}(f)$	$H_\infty(f)$
< 4,77 dB	$P_s < 3 \cdot P_d$	< 0,75	0
4,77 dB	$P_s = 3 \cdot P_d$	0,75	0,50
6,00 dB	$P_s = 4 \cdot P_d$	0,80	0,72
7,53 dB	$P_s = 5,6 \cdot P_d$	0,85	0,82
9,54 dB	$P_s = 9 \cdot P_d$	0,90	0,89
12,79 dB	$P_s = 19 \cdot P_d$	0,95	0,95
16,90 dB	$P_s = 99 \cdot P_d$	0,99	0,99
$\infty$	$P_d = 0$	1	1

Tabla 4.1: Valores del filtro en el límite, iteración infinita, en función de la relación señal a ruido.

Tenemos tres zonas bien definidas, tabla 4.2.

SNR	r	$H_{\text{opt}}(f)$	$H_\infty(f)$
< 4,77 dB	> 1/4	< 3/4	0
4,77 dB	= 1/4	= 3/4	0,5
> 4,77 dB	> 3/4	> 3/4	$0,5 \leq H_\infty(f) \leq 1$

Tabla 4.2: Zonas definidas sobre el filtro iterativo en el límite.

Si se representa sobre una gráfica  $H_{\text{opt}}(f)$  y  $H_{\infty}(f)$  se aprecia claramente la aparición de un pico en la respuesta frecuencial del filtro, figura (4.5).

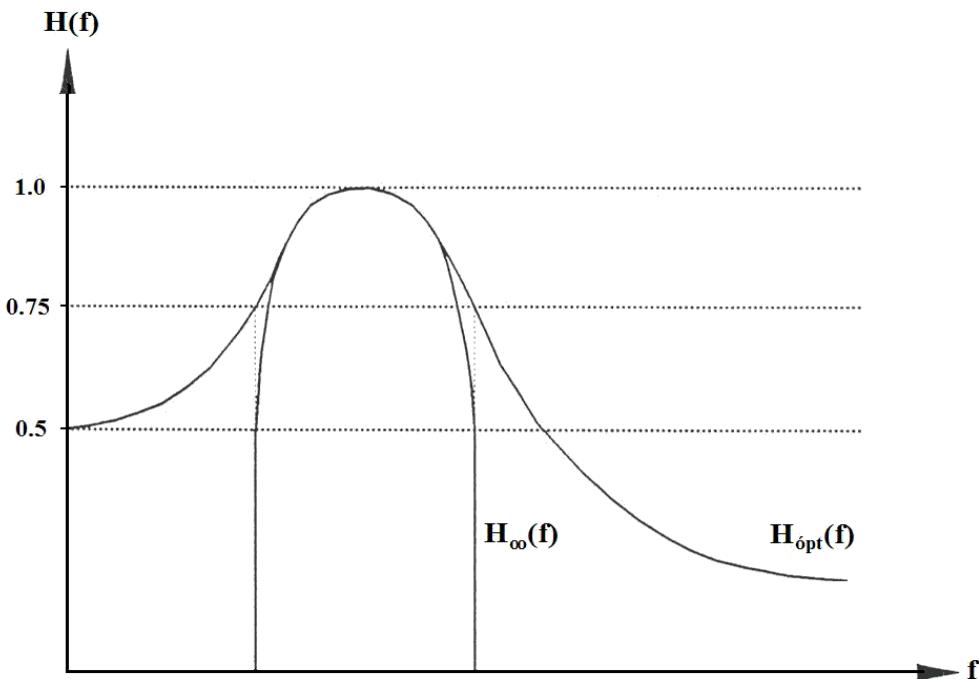


Fig. 4.5: Efecto de picado espectral debido a la presencia de ruido.  
Obsérvese el límite cuando el filtro óptimo vale 0,75. [Jove-93].

El efecto del picado espectral sobre los formantes que el filtro de Wiener produce al ser aplicado de manera reiterada puede tener su causa aquí.

Como ya se comentó al hablar de él (capítulo 3), este efecto es acusado por la diferencia de relación señal a ruido. En los alrededores del pico, donde existe una mayor (**S/N**) (en frecuencia) el picado no es significativo; mientras que en los valles se agudiza su profundidad paulatinamente, ocurriendo de manera brusca si estamos por debajo de los 4,77 dB.

Su evolución se presenta en la tabla (4.3) para poder apreciar lo lejos que estamos del valor óptimo del filtro en función de la iteración aplicada y el efecto degradador.

		$H_i$					
I	$H_{opt}$ dB	4,77	4,00	3,00	2,00	1,00	0,00
	$H_{opt}$	0,75	0,72	0,67	0,61	0,56	0,50
1		0,80	0,78	0,75	0,72	0,69	0,67
2		0,72	0,68	0,63	0,57	0,52	0,47
3		0,67	0,62	0,54	0,46	0,38	0,31
4		0,65	0,57	0,47	0,35	0,25	0,16
5		0,62	0,54	0,40	0,24	0,12	0,05
6		0,61	0,50	0,32	0,13	0,03	
7		0,60	0,47	0,23	0,04	0,00	
8		0,59	0,44	0,14	0,00		
9		0,58	0,40	0,06			
10		0,57	0,36	0,01	$10^{-8}$	$10^{-19}$	$10^{-33}$

Tabla 4.3: Evolución del filtro iterativo en una zona de “valle espectral”.

A medida que la relación señal a ruido disminuye el efecto de picado cae más rápido a cero, tendiendo a aislar los formantes. Para valores de ruido elevado, SNR<4,77 dB, el filtro iterativo y el filtro óptimo se aproximan en la segunda iteración.

#### 4.1.2.-Efecto de picado en el método de cumulantes

El método de los cumulantes nos ofrecía una cierta independencia al calcular el espectro de la señal a partir de ruido más señal. Para plasmar esto vamos a pensar que existe un factor de desacoplo señal-ruido  $\gamma$ . Mientras en el caso de las correlaciones al estimar  $P_s(\omega)$  obteníamos  $P_s(\omega) + P_d(\omega)$  pero un método ideal, aún en presencia del ruido debería darnos el valor de  $P_s(\omega)$  únicamente y no con ruido.

$$y(n) = s(n) + d(n) \Rightarrow \hat{P}_s(\omega) = \begin{cases} P_s(\omega) + P_d(\omega) & \text{método de correlaciones} \\ P_s(\omega) + \gamma \cdot P_d(\omega) & \text{método de cumulantes} \end{cases} \quad (4.39)$$

Los valores de  $\gamma$  nos darán una idea de lo robusto que es el algoritmo de predicción lineal frente al ruido, para mostrar la bondad de la estimación del espectro de la voz. Gamma estará comprendida entre:

$$0 \leq \gamma \leq 1 \quad (4.40)$$

- »  $\gamma = 1$       método de correlaciones o covarianzas.
- »  $0 < \gamma < 1$     casos intermedios (cumulantes).
- »  $\gamma = 0$         desacoplo ideal voz-ruido.

Introduciendo este parámetro modificamos la ecuación (4.28a), quedando como:

$$H_1 = \frac{1}{1 + \frac{P_d}{P_s + \gamma \cdot P_d}} = \frac{1}{1 + H^{c' \text{ ópt}}} \quad (4.41)$$

donde:

$$H^{c' \text{ ópt}} = \frac{P_d}{P_s + \gamma \cdot P_d} \quad (4.42)$$

El proceso para la obtención de los siguientes filtros es idéntico al desarrollado en el apartado anterior. En la segunda iteración:

$$H_2 = \frac{1}{1 + \frac{P_d}{P_s + \gamma \cdot P_d} \cdot \frac{1}{H_1^2}} = \frac{1}{1 + H^{c' \text{ ópt}} \cdot (H_1^{-1})^2} \quad (4.43)$$

Y, en general, para la **i-ésima** iteración podemos hallar una fórmula de recurrencia tal como:

$$H_i = \frac{1}{1 + H^{c' \text{ ópt}} \cdot (H_{i-1}^{-1})^2} \quad (4.44)$$

Aplicando los mismos cambios que en (4.16) y (4.19) a la ecuación recursiva:

$$D_i = 1 + H^{c' \text{ ópt}} \cdot D_{i-1}^2 \quad (4.45)$$

Obtenemos:

$$d(n) = 1 + r' \cdot d^2(n-1) \quad (4.46)$$

Con:

$$r' = H^{c' \text{ ópt}}$$

Si resolvemos la ecuación (4.46) para que converja la recurrencia, tal como en (4.20),

obtenemos el conjunto de soluciones de (4.23) para el número de iteraciones tendiendo a infinito:

$$r' > 1/4 \Rightarrow d(\infty) \rightarrow \infty \Rightarrow H_{\infty} = 0$$

$$r' = 1/4 \Rightarrow d(\infty) = 2 \Rightarrow H_{\infty} = 1/2$$

$$r' < 1/4 \Rightarrow d(\infty) \text{ converge} \Rightarrow 1/2 < H_{\infty} < 1$$

Si  $r \leq 1/4$  significa que  $H^{c'}_{\text{opt}} \leq 1/4$ :

$$\frac{P_d}{P_s + \gamma \cdot P_d} \leq \frac{1}{4} \quad (4.47)$$

O lo que es lo mismo, teniendo en cuenta la relación  $H^{c'}_{\text{opt}} = 1 - H'_{\text{opt}}$ :

$$\frac{P_s}{P_s + \gamma \cdot P_d} \leq 4 \quad (4.48)$$

De donde se debe cumplir:

$$P_s \geq (4 - \gamma) \cdot P_d \quad (4.49)$$

Rescribimos el filtro como:

$$H_{\text{opt}} = \frac{P_s}{P_s + P_d} > \frac{(4 - \gamma) \cdot P_d}{(5 - \gamma) \cdot P_d} = \frac{(4 - \gamma)}{(5 - \gamma)} \quad (4.50)$$

El valor de gamma es un parámetro que vendrá dado por el algoritmo de estimación utilizado. Podemos hallar que sucederá en los casos extremos: método de correlaciones,  $\gamma = 1$ , y el caso ideal, desacoplo total entre voz y ruido, caso donde el coeficiente vale  $\gamma = 0$ .

$$\begin{aligned} \gamma = 1 &\Rightarrow C = 3/4 \\ \gamma = 0 &\Rightarrow C = 4/5 \end{aligned} \quad (4.51)$$

Es decir el filtro converge a un valor inferior al del filtro óptimo si el filtro óptimo es superior a  $C$ . Si el filtro óptimo está por debajo de  $C$  entonces  $H_{\infty}(f_0)$  converge a cero.

$$\begin{aligned}
 H_{\infty}(\omega) &\leq H_{\text{opt}}(\omega) \quad \text{si} \quad H_{\text{opt}}(\omega) \geq C \\
 H_{\infty}(\omega) &= 0 \quad \text{si} \quad H_{\text{opt}}(\omega) \leq C \\
 C &= \frac{4-\gamma}{5-\gamma}
 \end{aligned} \tag{4.52}$$

La ilustración (4.6) muestra el efecto de picado. Donde, dado que el ruido de los cumulantes nos proporciona un mayor desacoplo voz-ruido respecto al método de correlaciones, el filtro tiende a ser más estrecho. No obstante la convergencia es más rápida para el algoritmo de cumulantes, por existir un mayor desacoplo voz-ruido, permite realizar un menor número de iteraciones.

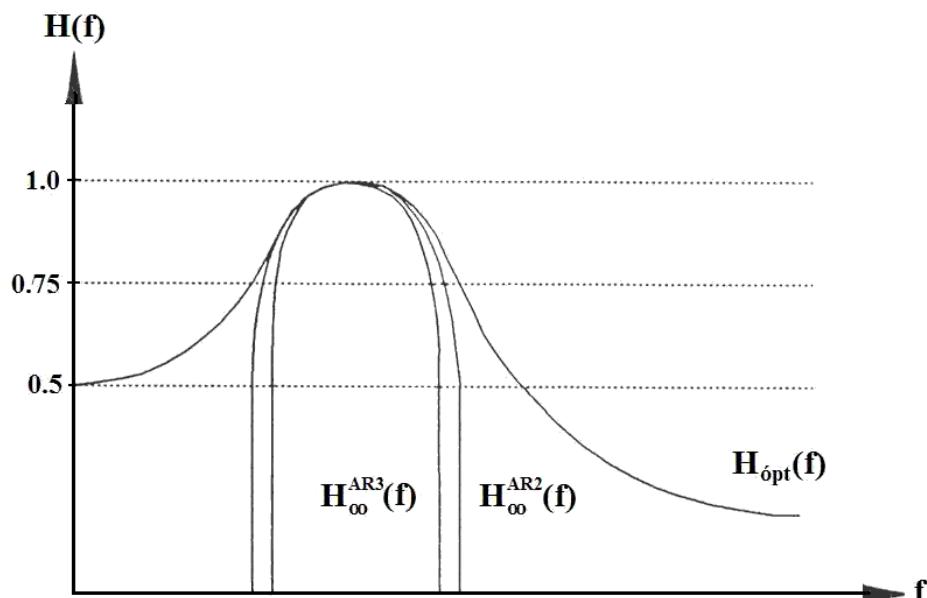


Fig.4.6: Variación del picado de la banda del filtro con el factor de acople, gamma. [Jove-93].

## 4.2 Distorsión espectral

Para implementar el filtro de Wiener:

$$H(\omega) = \frac{P_s(\omega)}{P_s(\omega) + P_d(\omega)} \tag{4.53}$$

Hemos de calcular el espectro de la voz,  $P_s(\omega)$  y el espectro del ruido,  $P_d(\omega)$ . Sin

embargo no disponemos de ninguno de los dos espectros, tan sólo podemos hacer estimaciones de ambos. Para el espectro del ruido,  $P_d(\omega)$ , debemos promediar las tramas disponibles en los períodos de silencio o las tramas que el VAD estime que son de ruido, para establecer la estimación de su espectro.

Igualmente no disponemos del espectro de la voz,  $P_s(\omega)$ , ya que ni tan siquiera tenemos  $s(n)$  aislada. Hemos aproximado su espectro a partir de la envolvente de predicción lineal de la señal sucia de entrada,  $y(n)$ . Nunca tendremos a nuestro alcance el espectro real, aún suponiendo un algoritmo AR que discriminara perfectamente ruido y señal. Esta aproximación del espectro por la envolvente no es muy limitadora, ya que para el oído es más importante la envolvente del espectro que el espectro en sí.

El efecto de picado en los formantes, que hemos explicado en el apartado 4.1, también impulsa el proceso iterativo hacia la distorsión espectral. Siendo el problema ligeramente más acusado para orden 3 que en orden 2 ó 4.

Estamos aplicando un filtro, que aun siendo la mejor aproximación que podemos obtener bajo estas limitaciones, no es el óptimo de Wiener. La aplicación del filtro mejora la señal de entrada, pero deberemos tomar precauciones. Si bien mejora la señal de entrada, introduciremos distorsión en la señal de voz. La aplicación iterativa del filtro nos da un efecto delante-atrás en la mejora global de la señal contaminada. Inicialmente mejoramos la calidad de la voz al reducir el ruido presente; pero poco a poco el ruido desaparece y surge la tendencia contraria, la merma en la inteligibilidad por distorsión espectral empeora la señal. La figura (4.7) muestra como la distancia espectral entre la señal de salida del sistema y la señal de voz limpia describe un valle. Primero disminuye hasta un mínimo en la iteración  $k_i$ , para aumentar cuando domina el efecto de distorsión ( $k > k_i$ ).

Aunque la curva que dibuja esta tendencia es general, su mínimo se desplaza en función del ruido de entrada. Cuanto mayor sea más tendremos que eliminar para llegar al mínimo, más iteraciones se empleará para que la señal pierda inteligibilidad. Si la  $SNR_i$  es más alta alcanzaremos la máxima mejora antes, además la distancia espectral mínima será menor cuan menor sea la  $SNR_i$ .

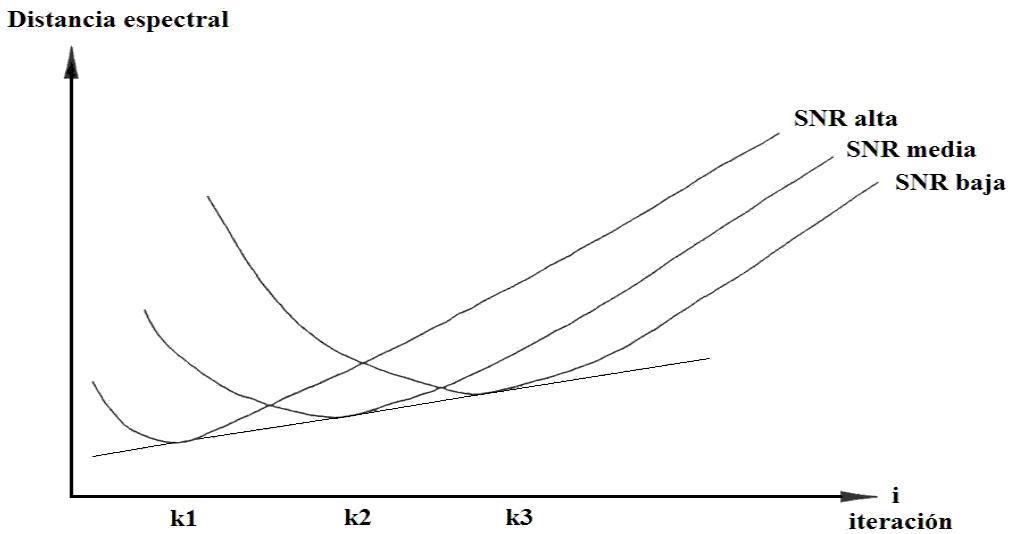


Fig.4.7: Evolución del Punto de mínima distancia espectral en función del número de iteraciones y  $\text{SNR}_i$ .

Podríamos asociar la curva de SNR alta con valores de **SNR > 18 dB**, siendo  $k_1 = 1$ . Para SNR media se incluyen valores de **6 a 15 dB** y un mínimo en la segunda iteración. Cuando tenemos a la entrada SNR baja, de **0 a 3 dB**,  $k_3$  se sitúa aproximadamente en la tercera iteración.

Los valores anteriores no son generales, y se refieren al caso clásico de correlaciones (orden 2). En función del modelo utilizado en la predicción lineal y del valor de los parámetros introducidos en el algoritmo el proceso puede acelerarse o retrasarse. Por ejemplo en el caso de cumulantes de orden tres, podemos llegar al mínimo en una o dos iteraciones incluso para  $\text{SNR}_i = 0$ , aumentando después la distorsión más rápidamente cuanto más rápido llegamos al mínimo. Existe un compromiso entre la velocidad hacia mínima distancia espectral y distorsión residual.

### 4.3 Desplazamiento de los formantes

La presencia del ruido nos dificulta a la hora de estimar el modelo autorregresivo. Esto influye sobre los formantes desde otro flanco, en la ubicación de los mismos. Sería lógico que para una misma trama en distintas iteraciones, o de una trama a las contiguas, la posición de los formantes fuera bastante fija (en señal sintética). No obstante, sufren un cambio de situación más o menos errático.

Por tener que hacer la estimación LP en presencia del ruido sufrimos un desplazamiento alrededor de la posición real. El grado de dispersión vendrá dado por el modelo de predicción lineal que utilicemos; a mayor robustez frente al ruido, es decir, mayor desacoplo voz-ruido en la estimación, menor será la incertidumbre en la posición del formante.

En la figura (4.8) se aprecia como, incluso dentro de una trama, se produce el desplazamiento al iterar. El primer formante se engancha con precisión, en el segundo erramos en la estimación inicial de su posición y, el tercero, no se engancha en ningún momento. El efecto de la  $\text{SNR}_i$ , ya sea para toda la trama o para una zona frecuencial concreta, limita la capacidad para determinar la posición de los formantes.

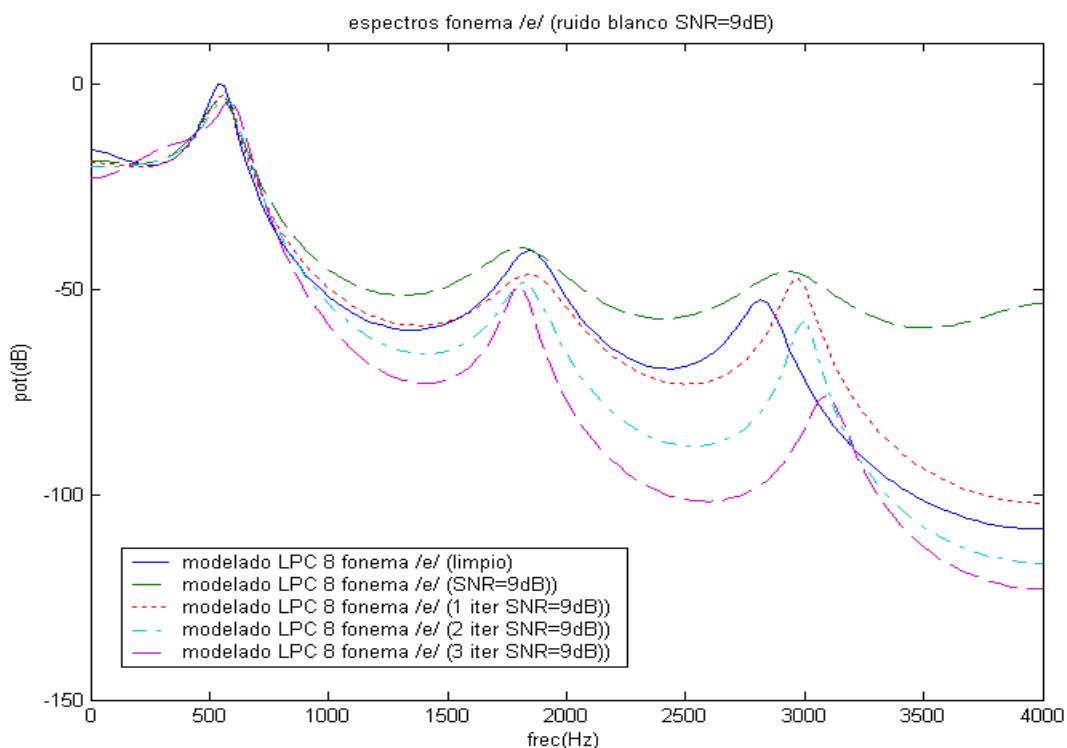


Fig.4.8: Desplazamiento de los formantes.

#### 4.4 Ruido 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 la identidad del locutor. Aún teniendo una señal de partida con una SNR prefijada, no podemos decir que cualquier trama de voz dentro de la frase posea esta misma relación señal a ruido. Cada segmento de señal tendrá su energía para la voz, valor variante, y un nivel para el ruido muy cercano a su energía media. Por

tanto, nos encontraremos con zonas donde domine la voz sobre el ruido, partes enmascaradas y trazos que sigan aproximadamente la relación de partida.

En cada iteración el ruido de fondo, muy molesto, disminuye hasta desaparecer; pero puede surgir un ruido musical, no por ello agradable, que desmerezca la mejora en la señal de voz. Este ruido es más acusado en las tramas o zonas del espectro donde existe una menor relación señal a ruido. La aparición de picos espurios a frecuencias altas del espectro de la voz, donde existe mayor relación ruido a señal, dan estas notas musicales.

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 del ruido de entrada (en función del orden del modelo). Al progresar en iteraciones aparecen espurios en la parte de menor SNR del espectro. En los siguientes filtrados este ruido musical tiende a desaparecer; pero aparece otro efecto no deseado. Nos referimos a la pérdida del poder de identificación de la persona que habla. Habremos eliminado el ruido musical; pero, por contra, se pierde el timbre de la voz, es difícil reconocer el locutor. Este efecto, a pesar de ser mucho menos molesto, puede ser muy indeseado en ciertas aplicaciones.

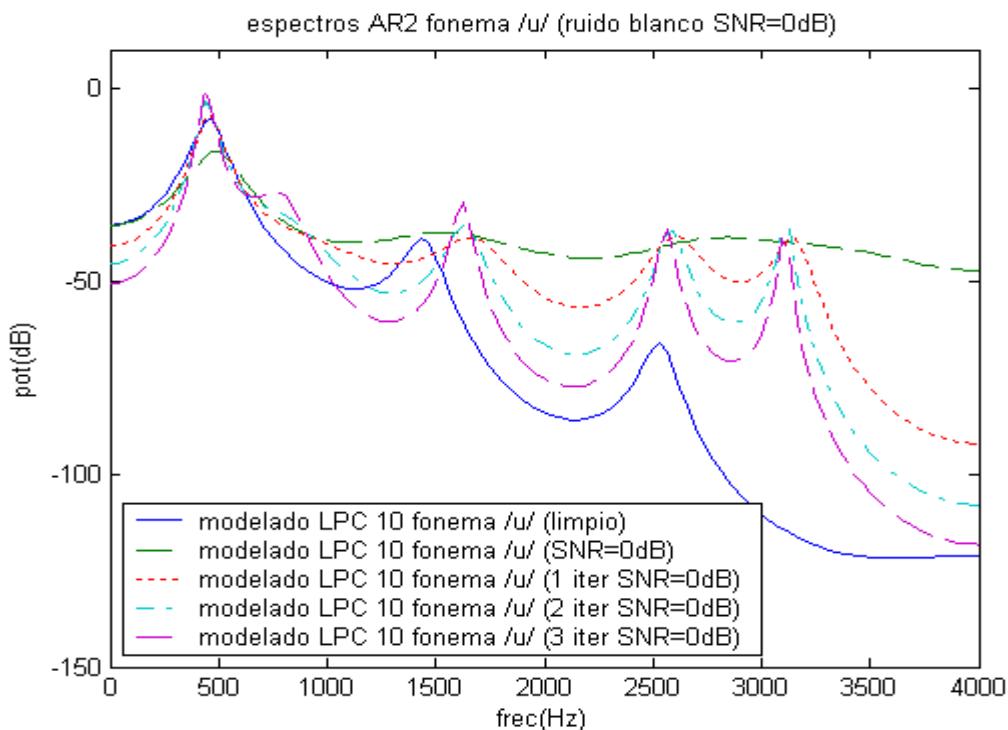


Fig.4.9: Evolución del filtrado de una trama de poca energía. Obsérvese la presencia de los espurios causantes de los tonos musicales.

En la figura (4.9) se observa la aparición de un espurio en la primera iteración para la parte alta del espectro (estamos filtrando una señal real a 0 dB). En las tramas de poca energía, como en la anterior, el ruido puede llegar a enmascarar totalmente la presencia de la señal y burlar nuestra estimación (la línea superior representa el modelo LP de la señal más ruido y la línea continua el de la señal original). Parece más evidente, según señal+ruido, la presencia de un cuarto pico que no aparece en la señal limpia. La estimación a partir de la señal más ruido nos vuelve a mostrar el efecto de desplazamiento de los formantes (Obsérvese el primer formante para la señal sucia y para las posteriores iteraciones de filtrado respecto a la original).

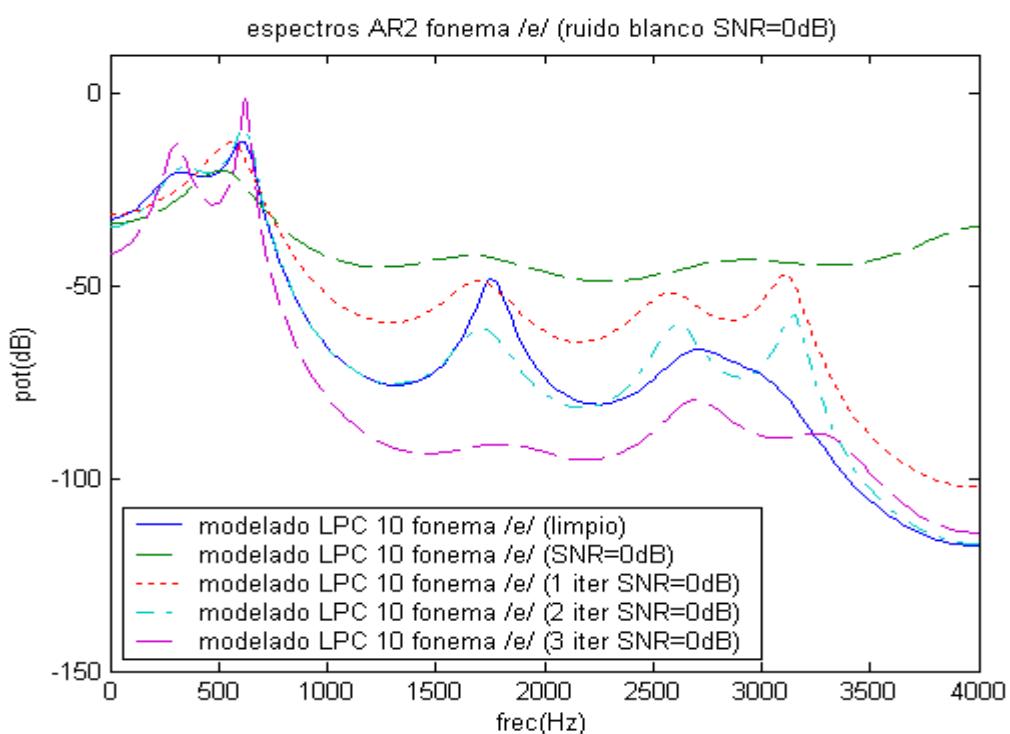


Fig.4.10: Filtrado de una trama de energía baja.

En la figura (4.10) los primeros formantes quedan bien determinados (estamos a 0 dB); por contra su dominio energético sobre el resto del espectro y sobre el ruido nos llevarán a una señal más ronca. Los dos primeros formantes muestran un efecto de picado respecto a la señal limpia. Mientras que el tercer formante, que tiene mayor energía que el cuarto, no desaparece, pero es absorbido paulatinamente por el primero, por el formante dominante. Esta reducción de los formantes que ocupan la parte superior del espectro lleva a una señal con dominio marcadamente más grave, una voz menos rica espectralmente.

Por último, veremos que pasa para un caso en el cual señal y ruido debaten por igual (figura (4.13), señal real a 0 dB). En una primera iteración eliminamos el ruido blanco, pero los formantes se desplazan, produciéndose un solapamiento entre el segundo y el tercero.

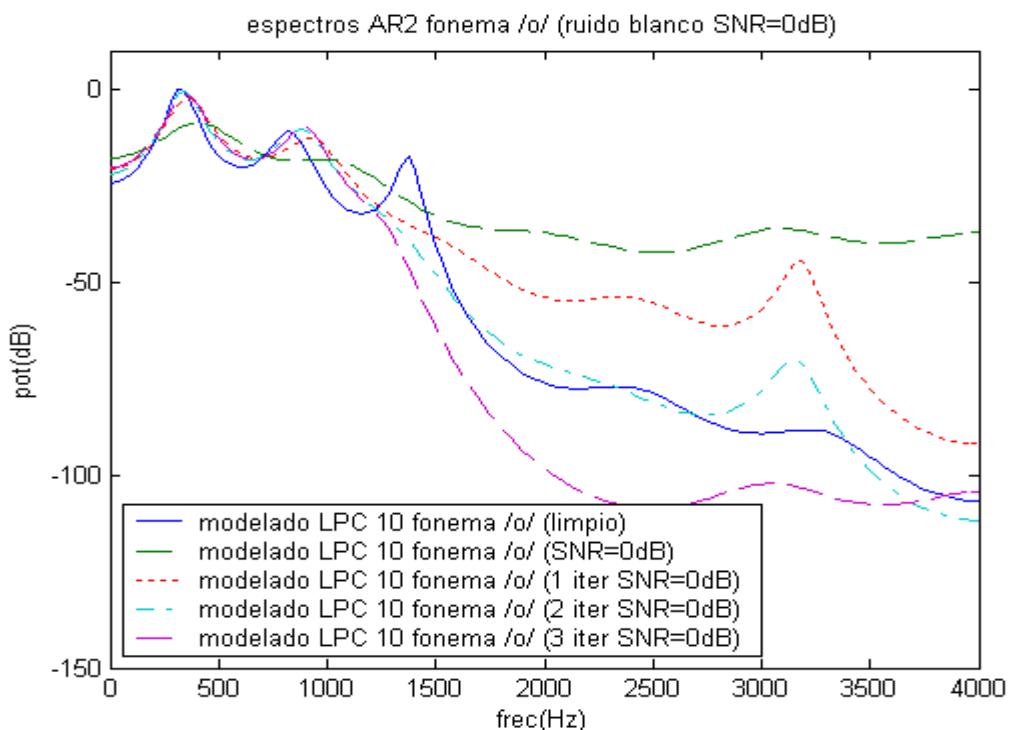


Fig.4.11: Iteraciones del filtrado de una trama de señal real a 0 dB con energía de ruido y señal equilibradas.

Ciertas variantes de la técnica básica de Lim-Oppenheim pretenden eliminar el ruido sin provocar la aparición de efectos no deseados, que conllevan distorsión espectral (p.ej.: disminución del ancho de banda de los formantes) y la aparición del ruido musical.

Hansen y Clements [Hans-91] sugirieron aplicar una serie de restricciones en el modelado para evitar la disminución del ancho de banda de los formantes, alejamiento de los formantes, distorsión espectral y otros efectos indeseados. La introducción de parámetros o condiciones en el algoritmo del modelo facilitan el control sobre la evolución de las estimaciones, esencialmente sobre la velocidad del filtrado y la varianza de las predicciones. Las restricciones aplicadas se muestran en los capítulos 6-7. Básicamente se ha utilizado la introducción de parámetros en el filtro de Wiener y el promediado (de coeficientes o de correlaciones) entre tramas contiguas o en una misma trama para las sucesivas iteraciones.



## 5.-VAD: Detector de actividad de Voz

En nuestro sistema realizaremos la estimación del ruido de dos maneras, al inicio de la transmisión antes del comienzo de la señal de voz, promediando las tramas de silencio que aparecen, realizando así una estimación inicial, y durante la conversación, en los instantes de silencio, en ausencia de señal de voz, que se detectan gracias al VAD. De esta manera la estimación de la potencia del ruido se va actualizando de manera continuada y es más fiable que si sólo hiciéramos una única estimación inicial.

El bloque VAD puede verse como un decisor: la trama i-ésima será de señal o de ruido en función de un determinado criterio que variará en función de su implementación. Por lo tanto, será necesario definir umbrales que separen las tramas en alta y baja energía.

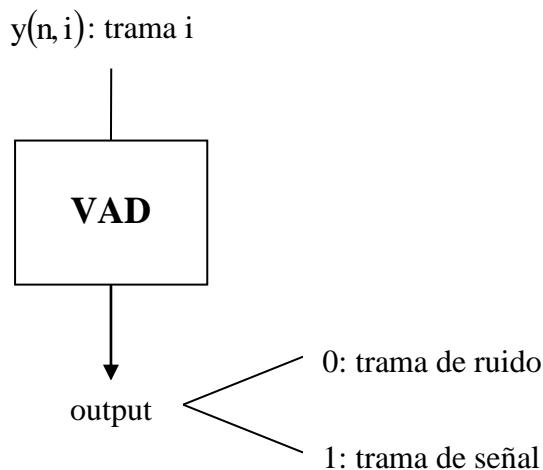


Fig.5.1: Esquema básico del VAD.

El subsistema VAD realiza una discriminación entre señal y silencio. Tal y como hemos visto, la señal de entrada al sistema está formada por voz+ruido y por eso los silencios se consideran tramas de ruido.

Si la trama es de señal, se aplicará el filtrado que corresponda utilizando como densidad espectral de potencia de ruido la estimada con todas las tramas anteriores de ruido. Si se considera trama de ruido se actualizará la estimación del ruido y se atenuará la trama.

Se consideran dos implementaciones básicas [Sala-02], diferentes para el VAD:

- Basado en Energía.
- Basado en Energía y Distancia Espectral.

Los parámetros básicos que utilizará un VAD "genérico", cualquiera que sea su implementación, son:

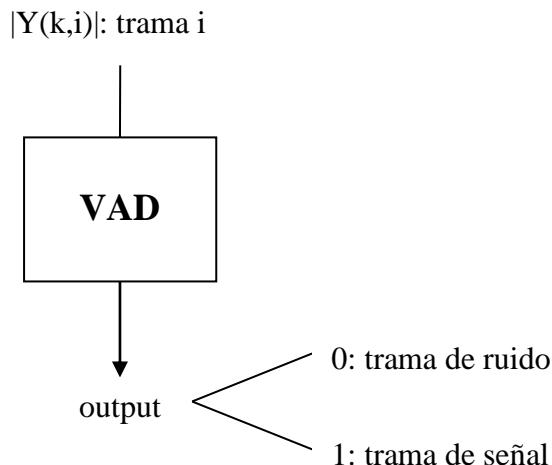


Fig.5.2: Parámetros básicos de un VAD.

- Módulo de  $\mathbf{Y}(k,i)$ :  $|Y(k,i)|$ . Se trabaja con el módulo porque la información de fase no aporta ninguna característica significativa a la señal de voz.
- Energía de la señal: Calculada a partir de sus muestras temporales.

$$E = \sum_{n=0}^{N-1} \frac{y^2(n)}{N} \quad (5.1)$$

donde N es el número de muestras de la trama.

El sistema VAD asume que las primeras tramas de cualquier señal son de ruido, puesto que antes de que un locutor comience a hablar ya se han digitalizado unos centenares de muestras que representan unos milisegundos.

Para tramas de muy alta energía, de un orden de magnitud de miles como en el caso de las vocales, el sistema VAD no tiene dificultades en detectarlas. Las decisiones comprometidas se producirán en las regiones de transición de silencio a señal y viceversa (cuando finaliza la pronunciación), sobretodo si en esas regiones aparecen consonantes sordas de espectro y energía similar al ruido.

La implementación del VAD basada en energía utiliza un umbral adaptativo de este parámetro, que se actualiza cada vez que se decide que una trama es de ruido, promediando su densidad espectral de potencia con el valor del umbral hasta ese momento. El objetivo del umbral es distinguir entre tramas de alta energía y tramas de baja energía. Entre estas últimas encontraremos tramas de ruido y consonantes sordas.

La segunda implementación del VAD, define un parámetro denominado **distancia espectral** cuyo objetivo es filtrar entre las tramas de baja energía. También será necesario definir un umbral para este parámetro.

$$d(P_y, P_d) = \sum_{k=0}^{N-1} |(\log P_y(k, i) - \log P_d(k, i-1))| \quad (5.2)$$

Donde:

$$P_y(k, i) = \frac{|Y(k, i)|^2}{N} \quad (5.3)$$

Es la densidad espectral de potencia para la trama de voz **i-ésima**. Y  $P_d(k, i-1)$  es la estimación de la densidad espectral de potencia de ruido calculada hasta ese momento, hasta la trama **i-ésima-1**, promediando las tramas de silencio al inicio de la transmisión y las tramas que el VAD ha considerado de silencio durante la misma.

Se trabaja con logaritmos para reducir el margen dinámico. Finalmente se hace un promedio para todas las muestras, de la diferencia entre las densidades espectrales de potencia de la trama actual y del ruido calculada hasta el momento.

### 5.1.-Estimación del Ruido.

La estimación del ruido sigue la siguiente expresión:

$$P_d(k, i) = (1 - \gamma) \cdot P_d(k, i-1) + \gamma \cdot P_y(k, i) \quad (5.4)$$

$$k : 0 \dots N-1$$

$$\gamma \in \mathbb{R} \text{ donde } 0 \leq \gamma \leq 1$$

Esta estimación se actualizará cada vez que el subsistema VAD detecte una trama que sea considerada de ruido. Como se observa, esta expresión se evalúa con densidades espectrales de potencia: módulos al cuadrado del espectro.

Por tanto, según se deduce de la expresión (5.3), se trata de una estimación acumulada de variación lenta, ya que la trama actual de ruido se ve afectada por el factor  $\gamma$ .

Para justificar esta última afirmación podemos interpretar  $\mathbf{P}_d(k,i)$  como la respuesta de un filtro IIR (respuesta impulsional infinita)  $\mathbf{H}(z)$  a una excitación  $\mathbf{P}_y(k,i)$ .

$$\mathbf{H}(z) = \frac{\gamma}{1 - (1-\gamma) \cdot e^{-z}} \quad y(n) = (1-\gamma) \cdot y(n-1) + \gamma \cdot x(n) \quad (5.5)$$

$\mathbf{H}(z)$  corresponde a un filtro paso bajo (FPB), lo cual se interpreta como una eliminación de las altas frecuencias del ruido.

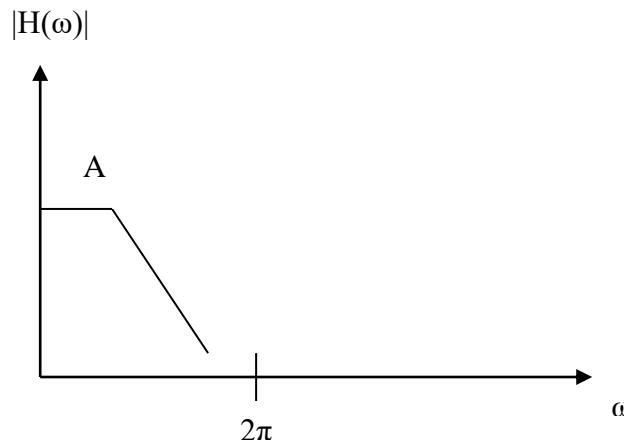


Fig.5.3: Filtro paso bajo en el dominio de la frecuencia.

Este comportamiento de la estimación del ruido contribuye a garantizar el supuesto de ruido aditivo y estacionario.

### 5.2.-VAD basado en la energía

El objetivo del VAD es decidir si la trama actualmente procesada es de alta o de baja energía. Esta implementación utiliza la energía como único parámetro para tomar esta decisión. El siguiente diagrama refleja esta idea:

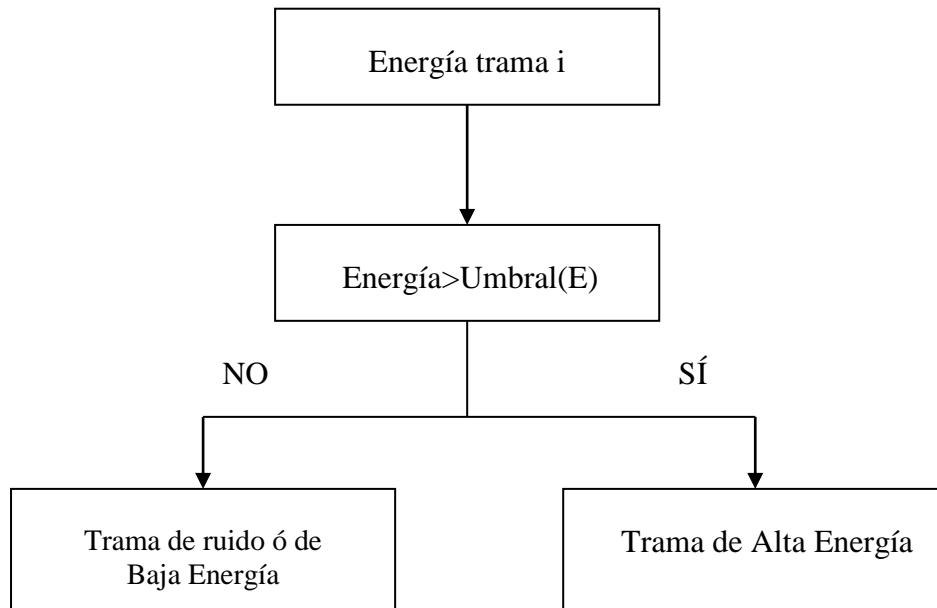


Fig.5.4: Esquema decisor de VAD basado en energía.

Se puede definir tanto a nivel temporal como a nivel frecuencial. Es un promedio del cuadrado de las muestras temporales o del cuadrado de las muestras en frecuencia, es decir, muestras de la FFT de la señal, según las siguientes expresiones:

$$E = \sum_{k=0}^{N-1} \frac{|Y(k,i)|^2}{N} \quad (5.6)$$

$$E = \sum_{k=0}^{N-1} \frac{y^2(n,i)}{N} \quad (5.7)$$

donde N, tal y como hemos visto, es el número de muestras de cada trama.

Al ser el ruido aditivo, lógicamente se produce un incremento de la energía. Esto puede apreciarse en la siguiente figura para una misma señal y  $\text{SNR}_2 < \text{SNR}_1$ .

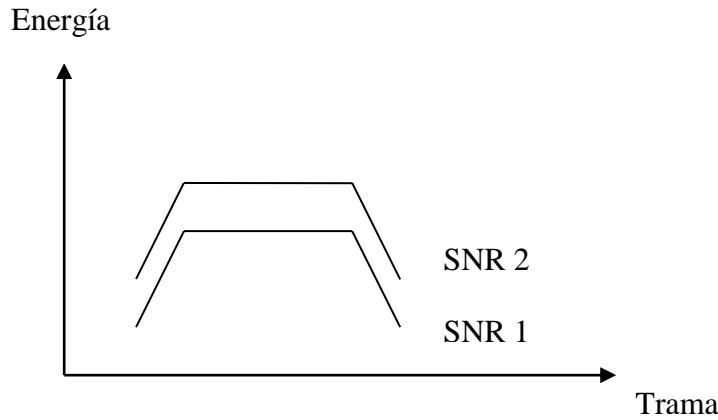


Fig.5.5: Relación entre Energía y SNR.

La implementación del VAD basada en energía utiliza un **umbral adaptativo**. Inicialmente el umbral de decisión es la energía del promedio de las densidades espectrales de energía de las tramas de silencio que siempre hay al inicio de la transmisión y por tanto, se realiza la comparación entre la energía de la trama actual y este valor inicial. A partir de este momento, cada vez que el bloque VAD detecta una trama de silencio (sólo contiene ruido), actualiza el valor del umbral promediando las densidades espectrales de energía de ruido calculado hasta ese momento y de la trama de ruido actual, de la siguiente manera:

$$P_d(k,i) = (1 - \gamma) \cdot P_d(k,i-1) + \gamma \cdot P_y(k,i) \quad (5.8)$$

$$k : 0 \dots N-1$$

$$\gamma \in \Re \text{ donde } 0 \leq \gamma \leq 1$$

Y a partir de  $P_d(k,i)$  calcula el nuevo umbral de energía:

$$E(\text{umbral}) = \sum_{k=0}^{N-1} P_d(k,i) \quad (5.9)$$

La ventaja de esta implementación de VAD respecto de otra es su simplicidad conceptual al utilizar sólo la energía, pero precisamente por ello es preciso afinar mucho el funcionamiento del algoritmo en sus condiciones iniciales y sus supuestos. De esa precisión depende que el sistema funcione bien para todos los casos presentados de ruido y relación señal a ruido.

A la pregunta de cómo distinguir el ruido de las consonantes sordas mediante la energía, se responde que para ambos casos el orden de magnitud es similar y que es necesario aplicar el algoritmo del VAD basado en distancia espectral, para poder afinar más en la decisión.

### 5.3.-VAD basado en la distancia espectral

El VAD basado en energía utiliza ésta como el único parámetro necesario para adaptarse a una variación lenta de energía de ruido. Para ruido con un ancho de banda más estrecho se produce una acentuación de la falsa clasificación entre tramas de ruido y tramas de baja energía de señal correspondientes básicamente a consonantes sordas. Es conveniente, por tanto, refinar el algoritmo basado en la energía, añadiendo un nuevo parámetro: la distancia espectral, con objeto de solucionar el problema planteado.

La siguiente figura nos muestra qué es lo que ocurre y cual es el papel de la distancia espectral:

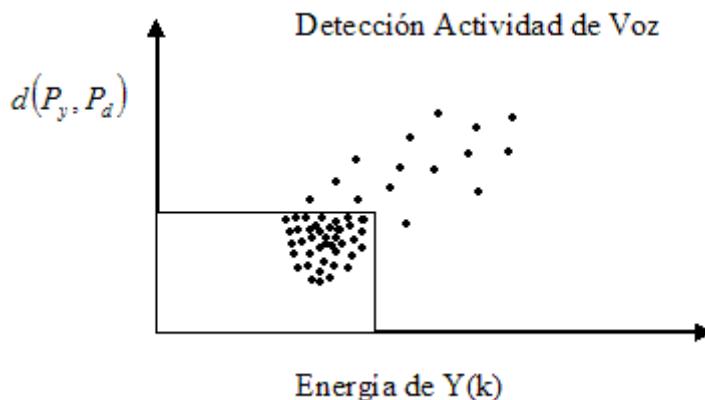


Fig.5.6: Área de decisión formada por ambos umbrales.

Como puede observarse, con ambos umbrales: energía y distancia espectral se define un área rectangular con un mejor ajuste a las tramas de ruido.

El umbral de energía determinará las tramas de baja energía, el umbral de distancia espectral intentará discriminar dentro de las tramas de baja energía, el ruido de las consonantes sordas. Por tanto, en el área rectangular están contenidas las energías de las tramas de ruido. Con este refinamiento estamos dotando de robustez al sistema.

La definición de distancia espectral, tal y como vimos, responde a la expresión:

$$d(P_y, P_d) = \sum_{k=0}^{N-1} |(\log P_y(k,i) - \log P_d(k,i-1))| \quad (5.10)$$

El algoritmo decisor en este caso será una extensión del utilizado en el VAD basado en energía, representado en la siguiente figura. La única modificación viene dada por la incorporación del parámetro distancia espectral.

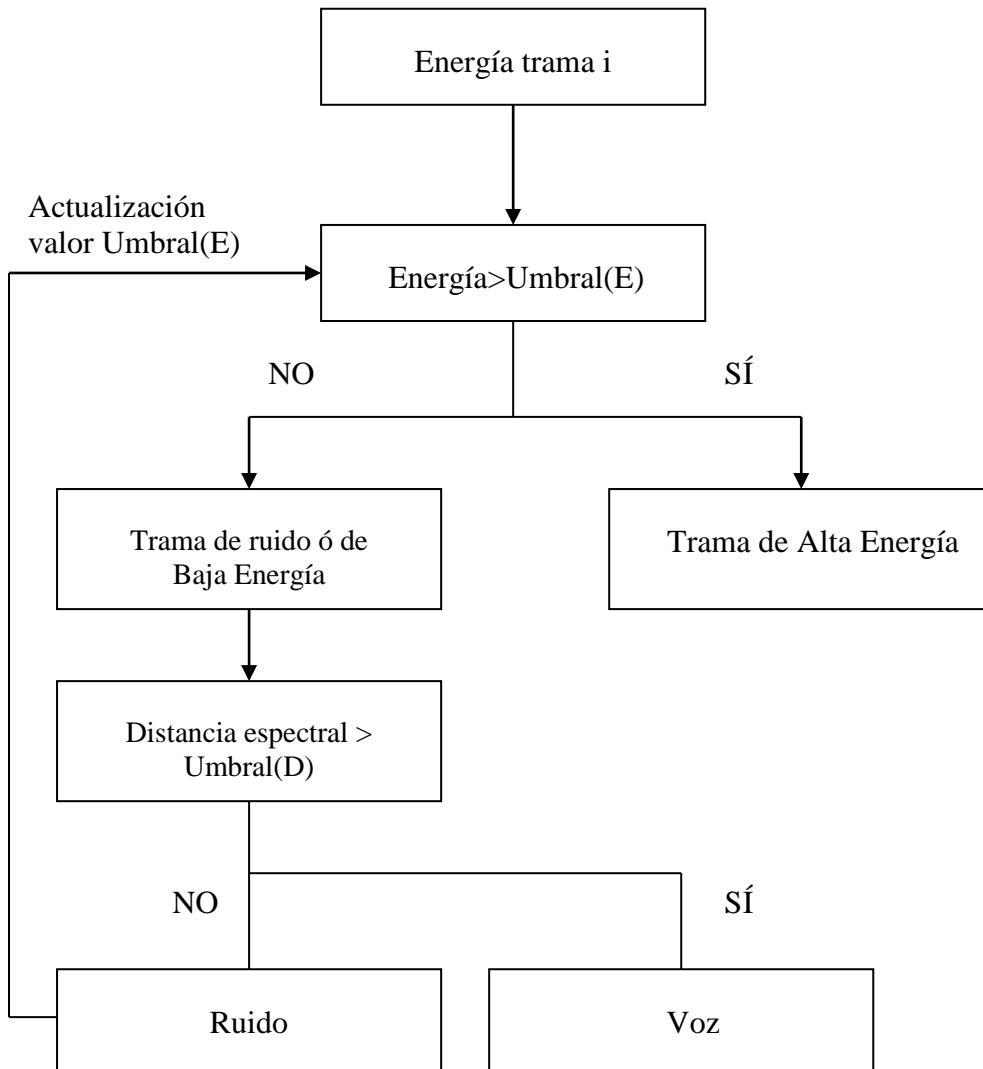


Fig.5.7: Algoritmo decisor VAD de distancia espectral.

Las tramas que en la primera implementación del VAD eran consideradas de baja energía y directamente consideradas como ruido, ahora son nuevamente procesadas.

Para establecer la decisión necesitamos un umbral para la distancia espectral: umbral( $D$ ).

Este **umbral** será **fijo** y se determinará a partir de las tramas iniciales de silencio al empezar la transmisión, calculando las distancias espetrales entre estas tramas y quedándonos con la distancia máxima como valor para el umbral, tal y como se detalla en la siguiente ecuación:

$$d(P_y, P_d) = \max_{i=1 \dots S} \left\{ \sum_{k=0}^{N-1} |\log P_d(k, i) - \log P_d(k, i-1)| \right\} \quad (5.11)$$

donde  $S$  es el número de tramas de silencio al inicio de la transmisión, que normalmente será de 10.

Así pues, realizaremos la comparación con este umbral de la distancia espectral entre cada trama de baja energía y la densidad espectral de potencia de ruido estimada hasta ese momento, si esta distancia es menor al umbral consideraremos que se trata de ruido, en caso contrario, consideraremos que se trata de un sonido consonántico.



## 6.- Implementación programa de simulación RERCOM.

En este capítulo trataremos de explicar las partes que componen el programa en C, que hemos bautizado como RERCOM (REducción Ruido en COmunicaciones Móviles). Este programa es capaz de eliminar “cualquier tipo” de ruido aditivo  $\mathbf{d(n)}$  mas o menos estacionario, que contamine una señal de voz  $\mathbf{y(n)}$ , obteniendo una estimación de voz realizada  $\mathbf{s_{est}(n)}$  lo más parecida posible a la señal de voz limpia original  $\mathbf{s(n)}$ .

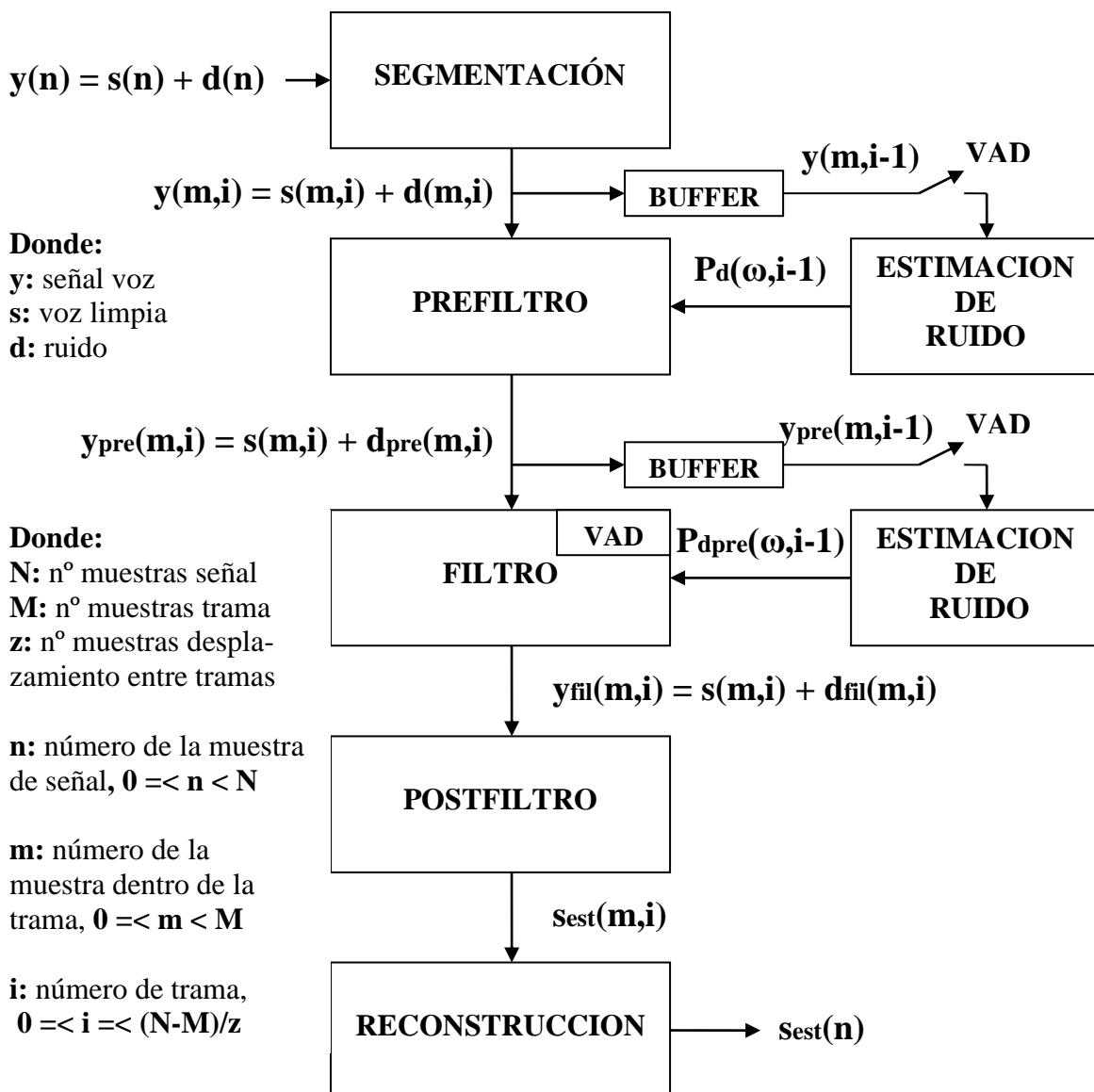


Fig. 6.1: esquema de filtrado utilizado en el programa RERCOM

Para conseguir esta meta se propone un esquema de 3 filtrados en cascada, cada uno de ellos con diferentes propiedades de eliminación de ruido aditivo. De esta manera, se ataca el ruido aditivo que contiene una señal de voz contaminada  $\mathbf{y(n)} = \mathbf{s(n)} + \mathbf{d(n)}$ ,

desde tres frentes diferentes. El esquema mostrado en la figura 6.1 permite obtener un gran rendimiento de eliminación de ruido con una distorsión mínima de la señal de voz limpia original.

### 6.1.- Segmentación.

El bloque que hemos llamado segmentación se encarga de subdividir la señal de voz ruidosa  $y(n)$  en segmentos de señal que llamaremos tramas  $y(m,i)$ . La longitud, en tiempo, de estas tramas depende de varios factores que expondremos seguidamente.

Por norma general, una señal de voz no cumple la propiedad de estacionariedad para intervalos de tiempo relativamente largos: la continua variación del tracto vocal y su excitación al generar los distintos fonemas que componen la señal de voz hace inviable un procesado directo de más de 30 ms de señal. Además, si queremos que nuestro programa trabaje en tiempo real, hemos de tener en cuenta que el retardo que introduce el sistema va ligado directamente a la longitud de trama utilizada. Así, tanto por motivos de estacionariedad, como por el retardo introducido nos interesará que la longitud de trama, sea lo más pequeña posible.

Por otro lado, para obtener unas estimaciones de densidad espectral de potencia consistentes, tanto en sesgo como en variancia, es necesario que el tiempo de trama sea lo mas grande posible, ya que el rendimiento de eliminación de ruido depende por completo de la calidad de estas estimaciones. Además, es necesario, por motivo de rendimiento computacional de la FFT, que el número de muestras por trama sea un múltiplo de  $2^n$  con n entero positivo.

A una frecuencia de muestreo típica de 8Khz, hemos determinado experimentalmente que una longitud de trama de 256 muestras, que corresponde a 32 ms de tiempo, es la decisión óptima al compromiso que se nos planteó, así como un desplazamiento, entre trama y trama, de 64 o 128 muestras, obteniendo así un solapamiento entre tramas del 50% o 75%.

Con un solapamiento de trama del 75%, podremos obtener a la salida del sistema una señal reconstruida obtenida del promediado de 4 tramas filtradas, este promediado nos ayudará a obtener unos mejores resultados, en rendimiento de eliminación de ruido a costa de aumentar el número de filtrados, y por tanto mayor coste computacional, por trama reconstruida.

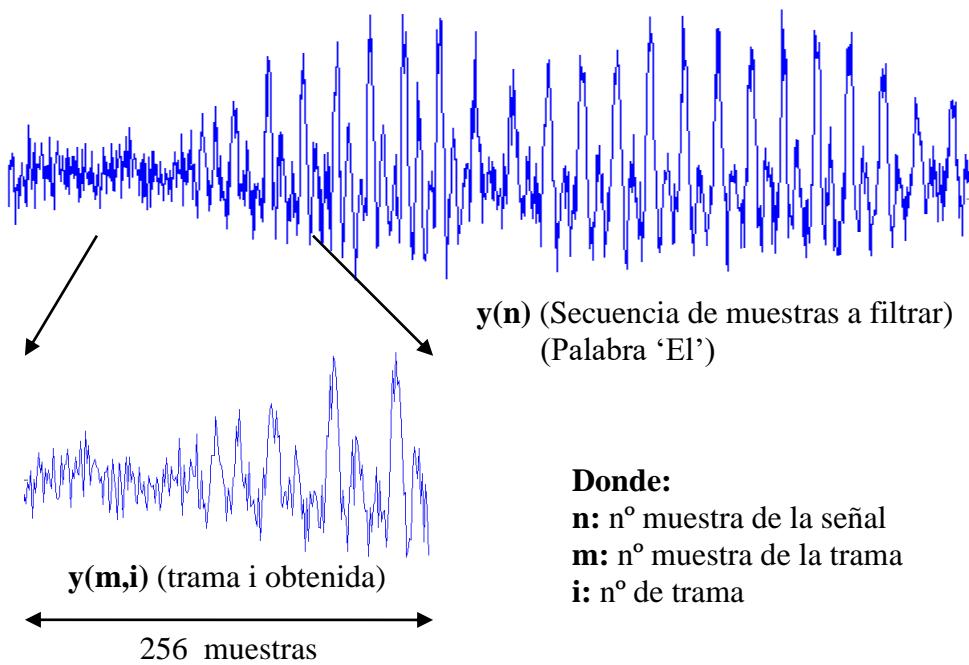


Fig. 6.2: Obtención de una trama (256 muestras) de señal a procesar. La trama corresponde al inicio del fonema /e/ del fichero ASUN1 + ruido blanco (SNR = 9 dB)

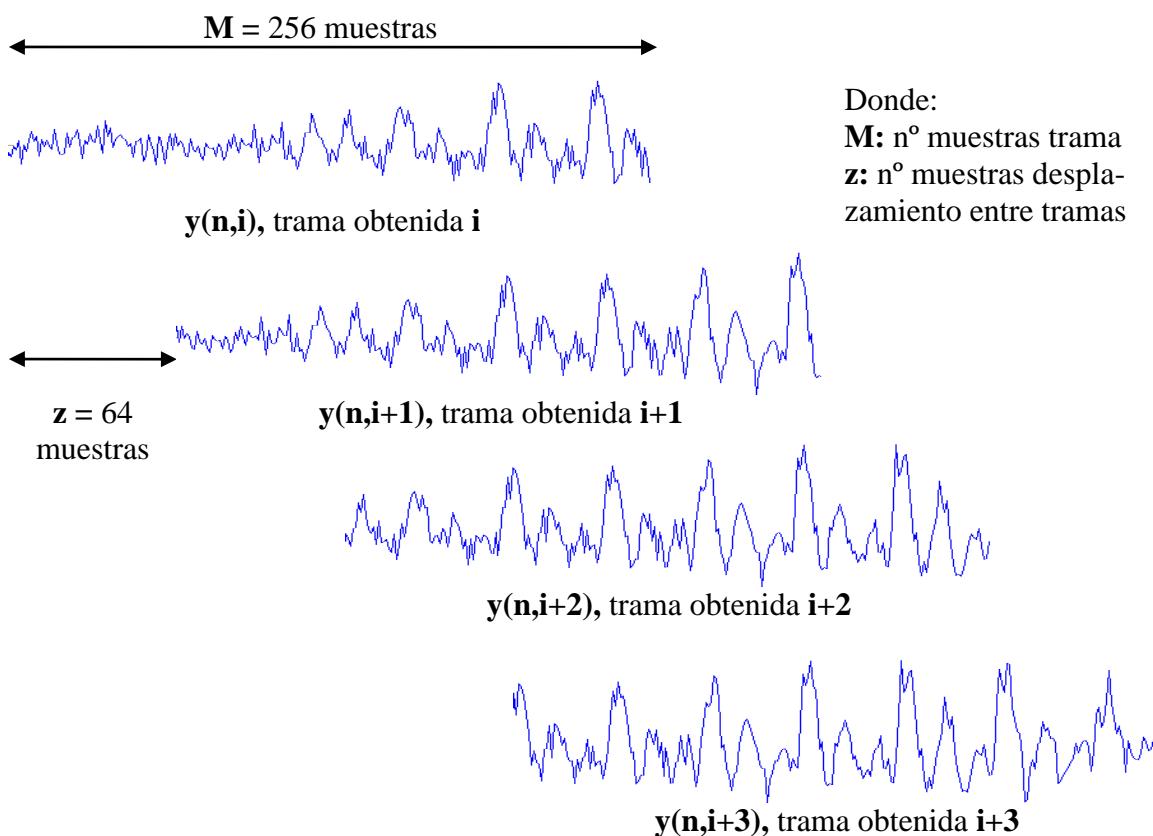


Fig. 6.3: Tramas de señal obtenidas utilizando un solapamiento del 75%. Las tramas corresponden al fonema /e/ del fichero ASUN1 con ruido blanco (SNR = 9 dB)

## 6.2.- Estimación Ruido

Para poder construir los filtros necesario para los bloques **Prefiltro** y **Filtro**, es necesaria una buena estimación de la densidad espectral de energía (DEE) del ruido aditivo que contiene la señal de voz. En nuestro caso, el bloque **Estimación de Ruido**, sólo tendrá disponible las tramas de voz;  $y(m,i) = s(m,i) + d(m,i)$ . Por lo tanto la mejor forma de poder estimar la DEE del ruido  $Pd(\omega)$  de la señal de voz ruidosa  $y(n)$ , es realizando esta estimación durante los intervalos de silencio, en donde  $y(n) = d(n)$ .

La estimación de DEE del ruido  $Pd(\omega)$  se realiza en dos fases: En una primera fase, se considera que las tramas de señal sólo contienen ruido, a partir de esta hipótesis, el sistema realiza por defecto una primera estimación de ruido siguiendo el esquema:

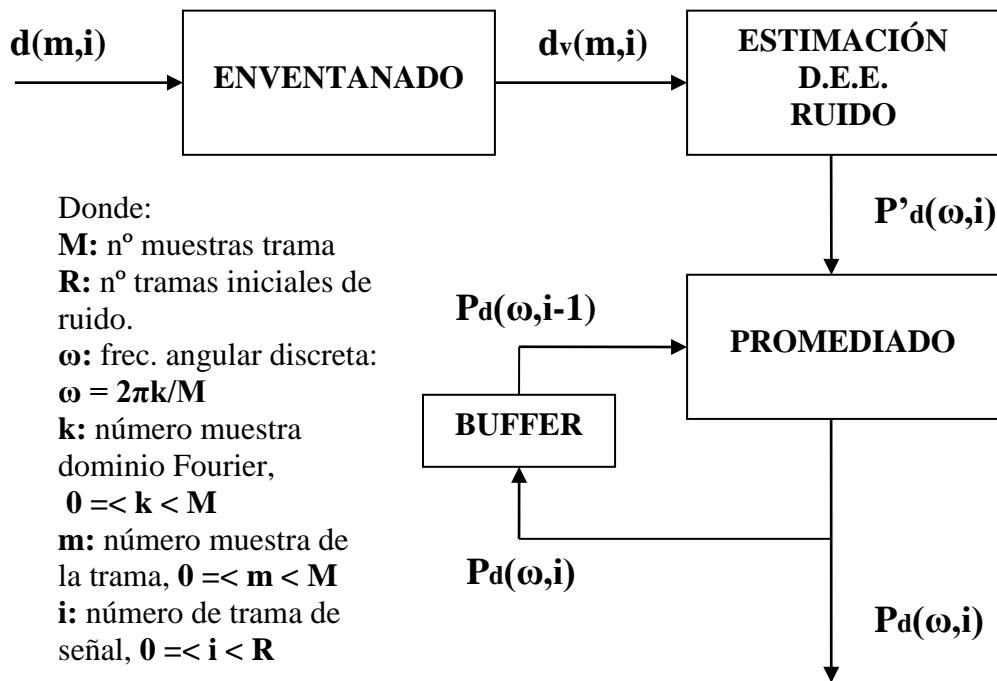


Fig. 6.4: Diagrama bloques estimación inicial de ruido

En una segunda fase, se realizan reestimaciones de DEEs de ruido que actualizarán la estimación de la primera fase, de esta forma el sistema será capaz de detectar y eliminar ruidos no estacionarios. Un VAD, situado en el bloque de filtrado, será el encargado de decidir si la trama actual **i** contiene voz,  $y(m,i) = s(m,i) + d(m,i)$ , o no la contiene,

$y(m,i) = d(m,i)$ . Por lo tanto, sólo en el caso de no detectar voz, el sistema actualizará la estimación de D.E.E de ruido  $Pd(\omega)$  que será utilizada en la trama de señal  $i+1$ .

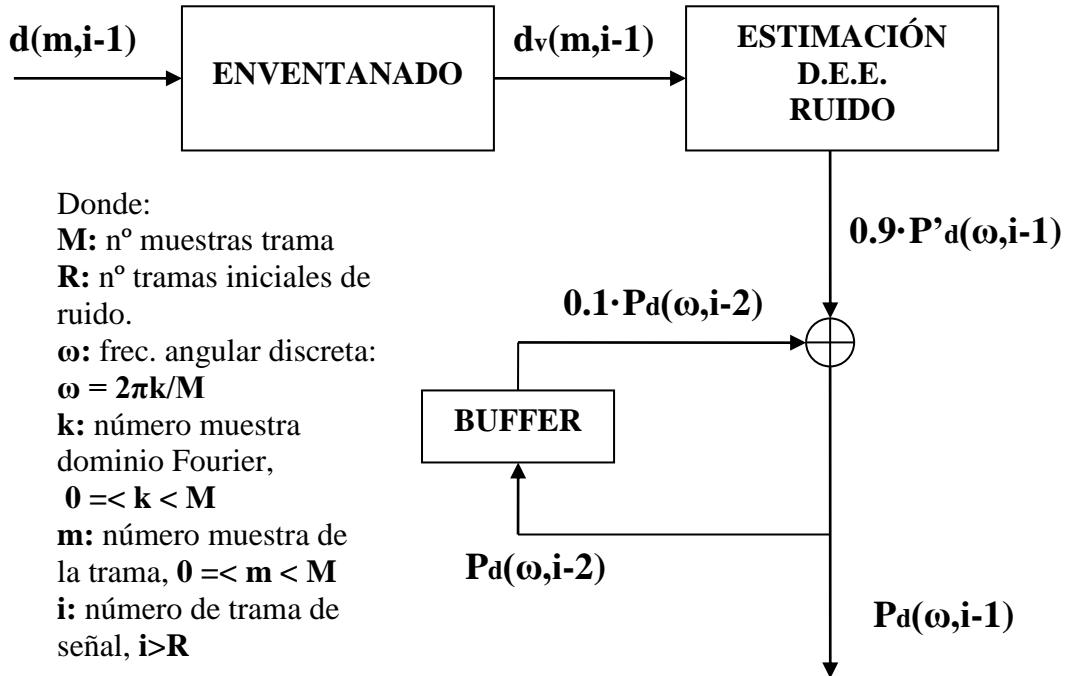


Fig. 6.5: Diagrama bloques para la actualización de la estimación ruido

### 6.2.1.- Enventanado de trama.

El enventanado de la trama de entrada  $d(m,i)$  se realiza mediante la ventana de Hanning:

$$d_v(m, i) = d(m, i) \cdot v_{\text{hann}}(m) \quad (6.1)$$

Con:

$$v_{\text{hann}}(m) = 0.5 \cdot \left[ 1 - \cos\left(2 \cdot \pi \cdot \frac{m}{M-1}\right) \right] \quad (6.2)$$

Donde: **m:** número muestra de la trama,  $0 \leq m < M$   
**M:** n° muestras de la trama

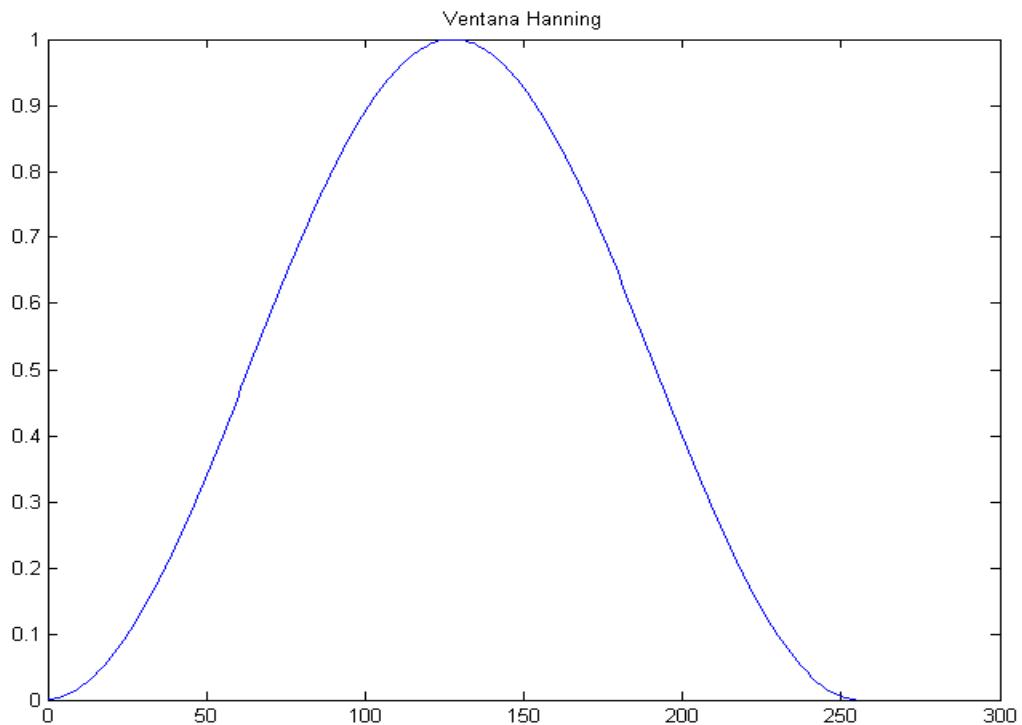


Fig. 6.6: Ventana de Hanning

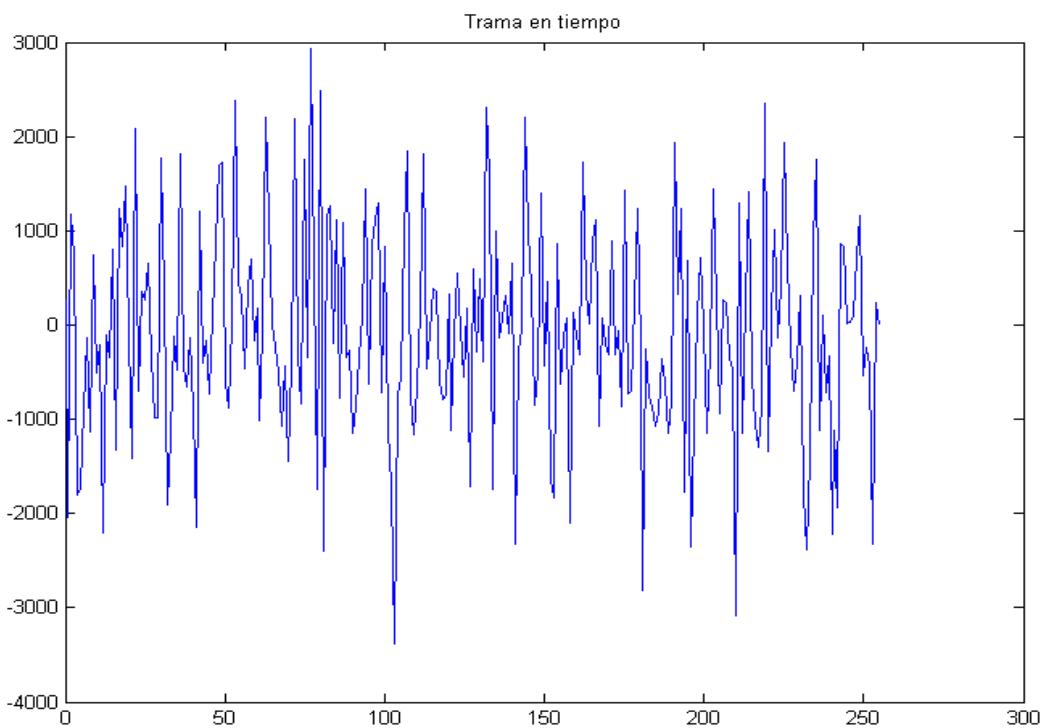


Fig. 6.7: Trama ruido blanco sin enventanar

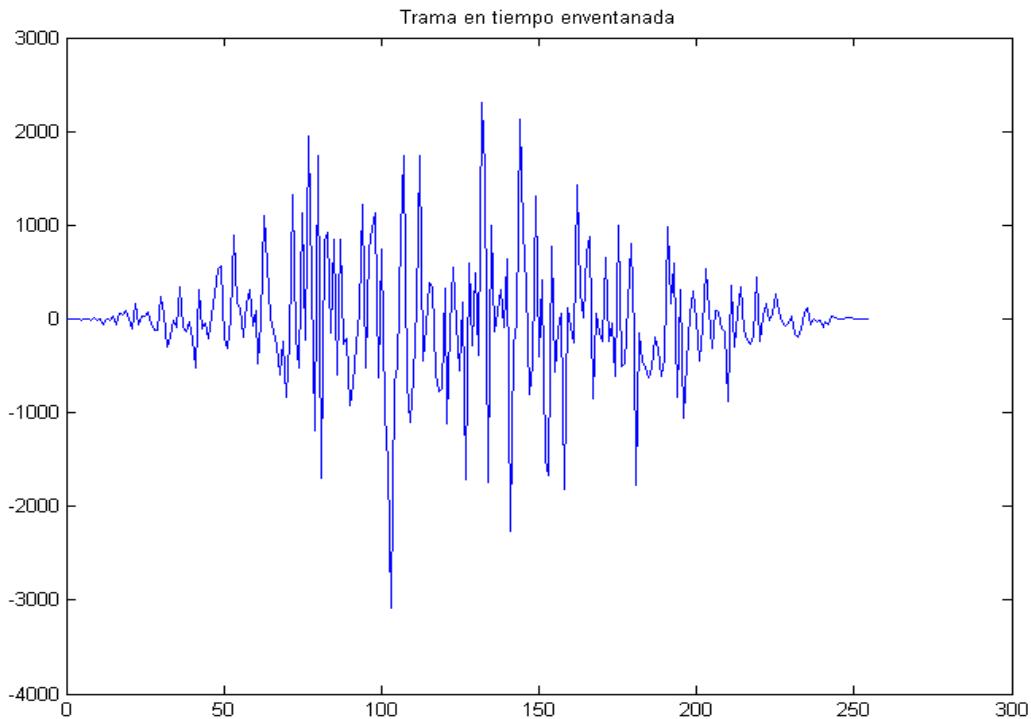


Fig. 6.8: Trama ruido blanco enventanada

Con la aplicación de esta ventana en cada una de las tramas de ruido  $\mathbf{d}(m,i)$ , se pretende conseguir un suavizado de los extremos de la trama y, de esta manera, mejorar la estimación de la DEE de ruido  $\mathbf{P}'\mathbf{d}(\omega)$ .

#### 6.2.2.- Estimación densidad espectral de energía de ruido.

La estimación de la DEE de ruido sin promediado  $\mathbf{P}'\mathbf{d}(\omega)$  se realiza de la siguiente forma:

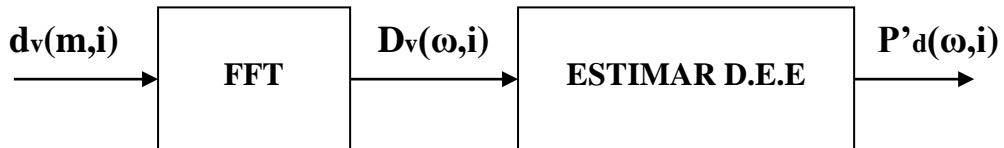


Fig. 6.9: Diagrama de bloques estimación DEE ruido

El bloque **FFT** realiza la transformada rápida de Fourier [Rodr-03] obteniendo la trama de ruido en el dominio transformado  $\mathbf{Dv}(\omega,i)$ .

El bloque **Estimar DEE** realiza la siguiente operación:

$$P'_d(\omega, i) = \frac{|D_v(\omega, i)|^2}{M} \quad (6.3)$$

Donde: **M**: n° muestras trama

**ω**: frec. angular discreta:  $\omega = 2\pi k/M$

**k**: número muestra dominio Fourier,  $0 \leq k < M$

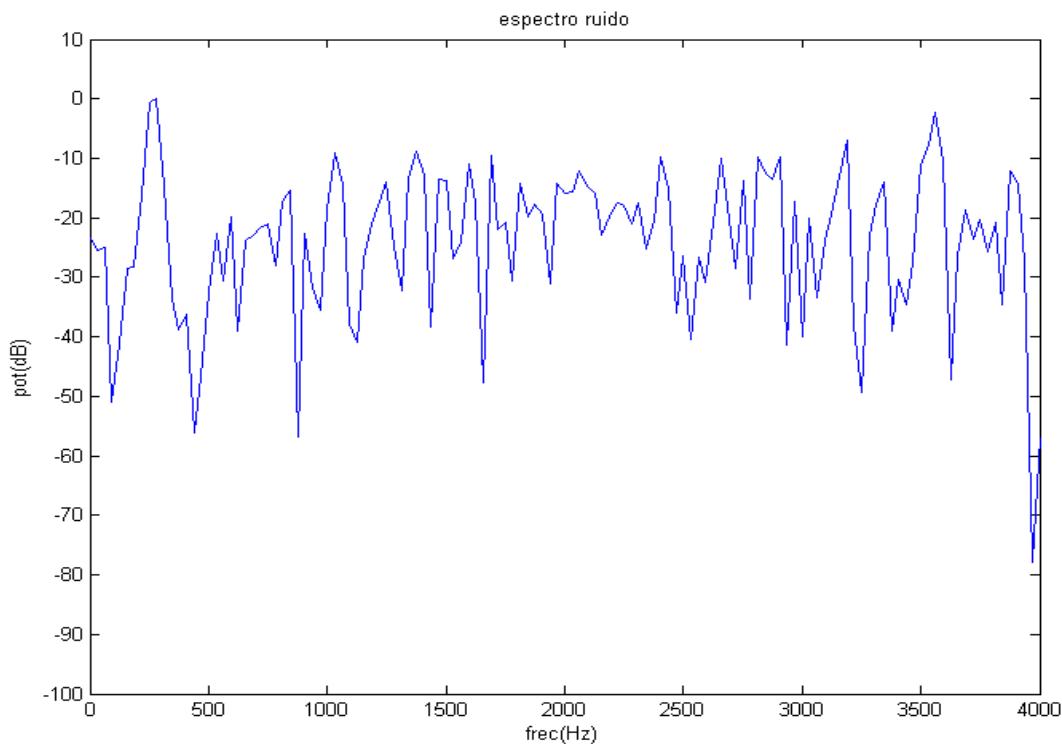


Fig. 6.10: Estimación de la DEE de ruido normalizada a 0dB de la trama 0

### 6.2.3.- Promediado de espectros.

Una vez calculada la DEE de ruido, se procede al promediado de esta  $P'_d(w, i)$  por el espectro promediado en una trama anterior  $P_d(w, i-1)$  según la ecuación:

$$P_d(\omega, i) = \gamma \cdot P'_d(\omega, i) + (1 - \gamma) \cdot P_d(\omega, i - 1) \quad (6.4)$$

Con:

$$\gamma = \frac{1}{i+1} \quad (6.5)$$

Donde: **i**: número de trama de ruido,  $0 \leq i < R$

**R**: n° tramas iniciales de ruido.

Con el promediado de DEEs de ruido se consigue disminuir el sesgo y variancia de la estimación, haciendo esta más consistente.

El programa RERCOM parte de la hipótesis de que las primeras 10 tramas de señal son ruido **R=10**, es decir, un tiempo de **104 ms**, para una longitud de trama **M=256** muestras y un desplazamiento de trama **z=64**, suficiente para asegurarnos que no falsearemos la DEE de ruido con tramas de voz.

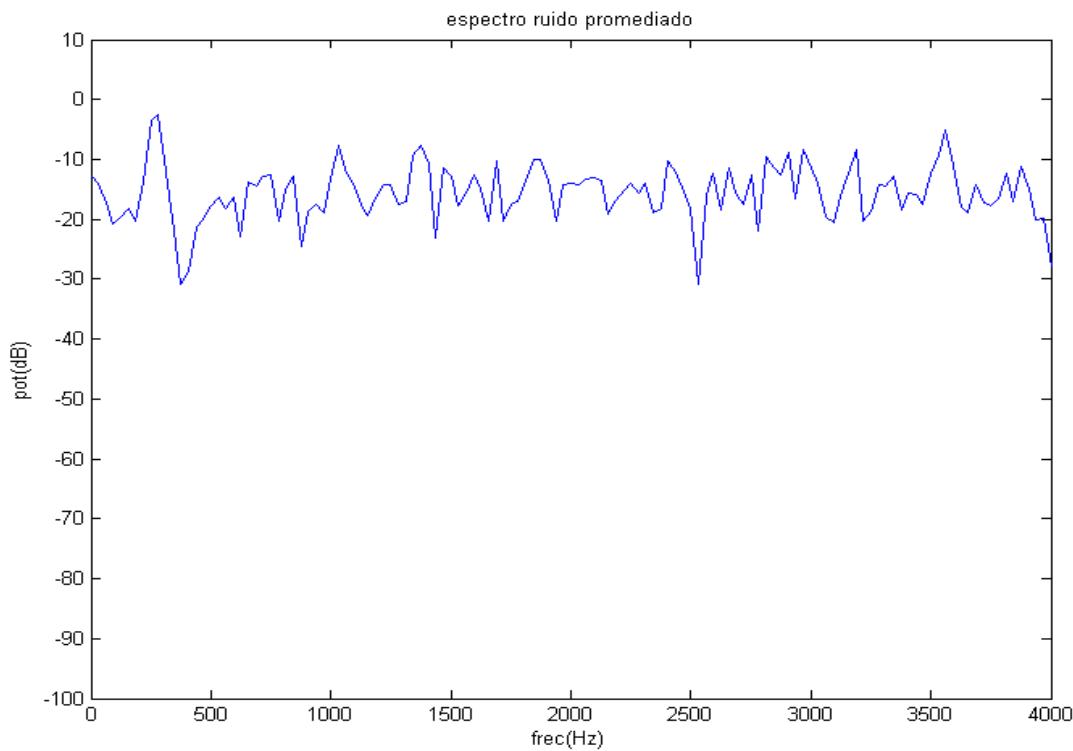
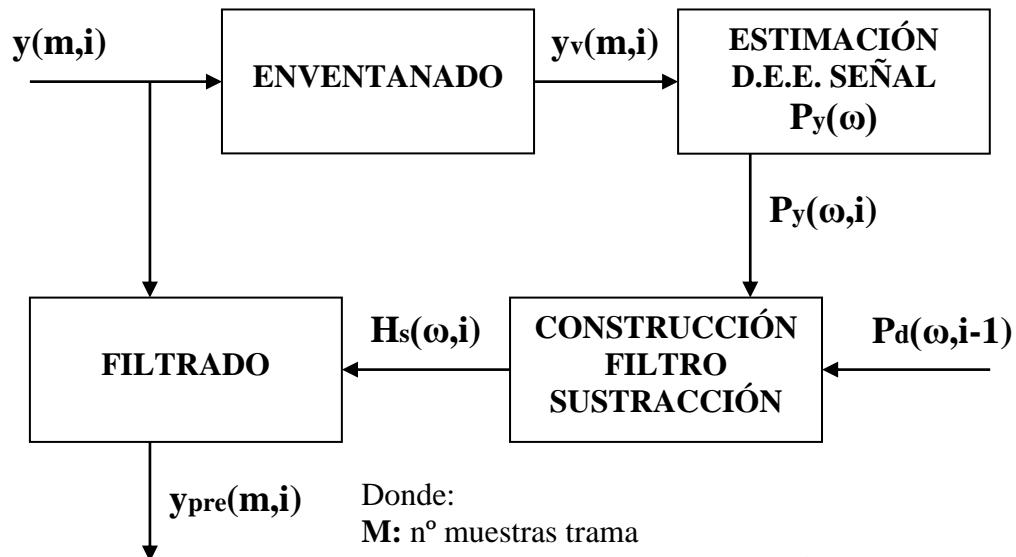


Fig. 6.11: Estimación de la DEE de ruido normalizada  
a 0dB después del promediado inicial de 10 tramas

### 6.3.- Prefiltro

El objetivo del prefiltro es realizar una primera reducción del ruido contenido en la trama de voz, de esta forma se prepara la trama de señal para que en el bloque de filtrado se obtenga un mejor rendimiento.



Donde:

- $M$ : n° muestras trama
- $\omega$ : frec. angular discreta:  $\omega = 2\pi k/M$
- $k$ : número muestra dominio Fourier,  $0 \leq k < M$
- $m$ : número muestra de la trama,  $0 \leq m < M$
- $i$ : número de trama de señal

Fig. 6.12: Diagrama de bloques del prefiltro

#### 6.3.1.- Enventanado de trama.

El enventanado de la trama de entrada  $y(m,i)$  se realiza de la siguiente forma:

$$y_v(m,i) = y(m,i) \cdot v_{hann}(m) \quad (6.6)$$

Con:

$$v_{hann}(m) = 0.5 \cdot \left[ 1 - \cos\left(2 \cdot \pi \cdot \frac{m}{M-1}\right) \right] \quad (6.7)$$

Donde:  $m$ : número muestra de la trama,  $0 \leq m < M$   
 $M$ : n° muestras de la trama

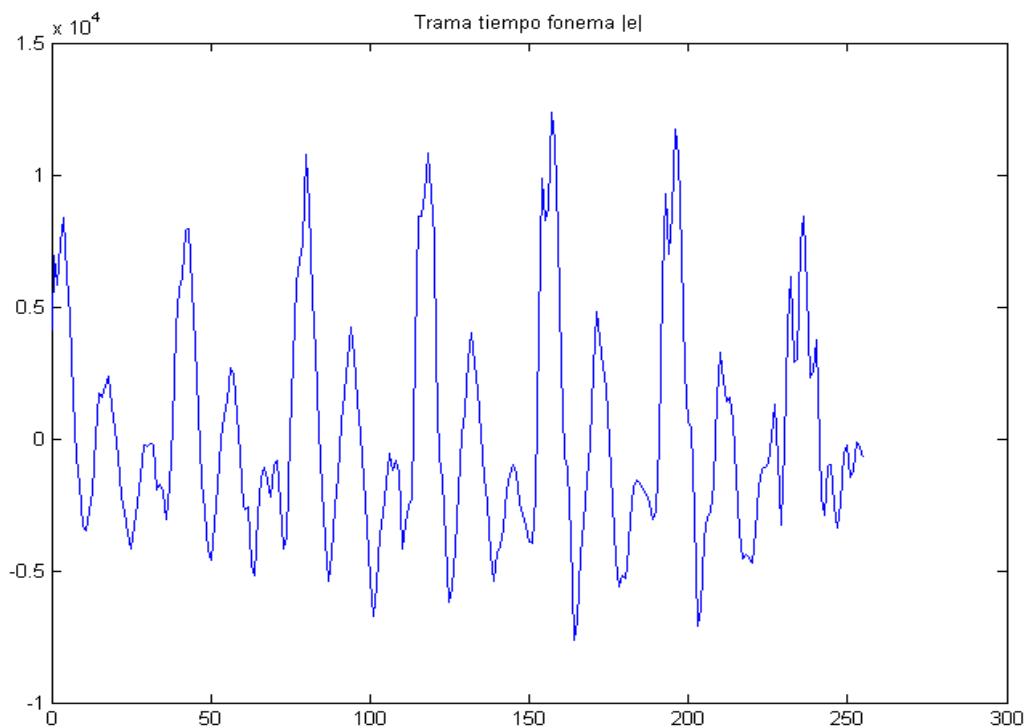


Fig. 6.13: Trama señal sin enventanar, correspondiente al fonema /e/

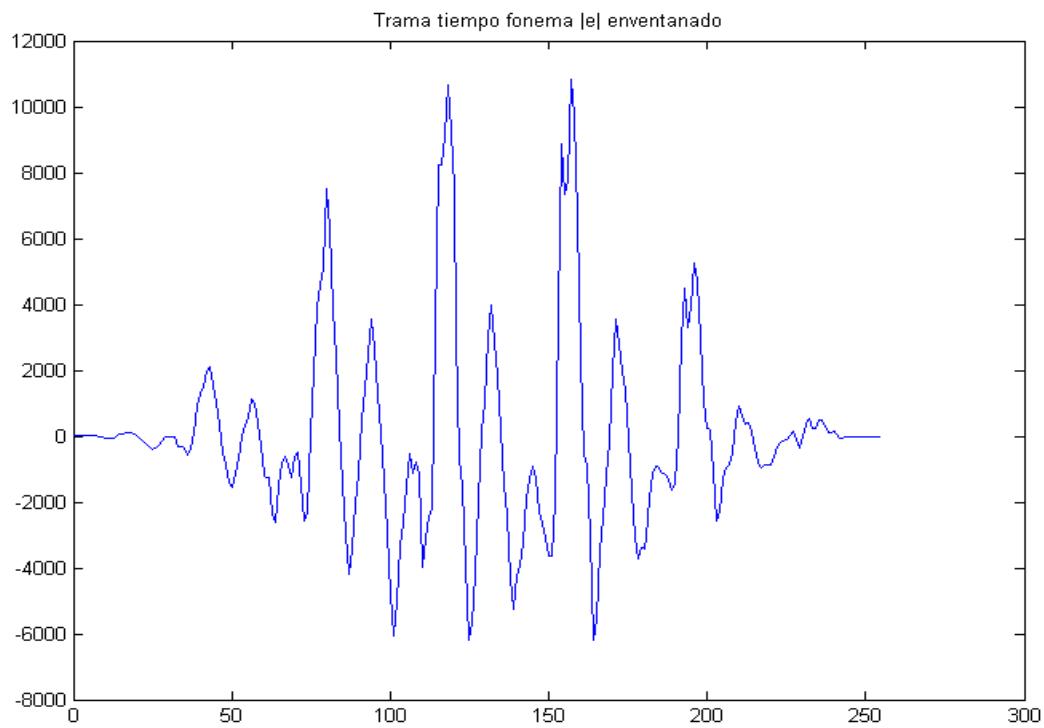


Fig. 6.14: Trama señal correspondiente al fonema /e/ enventanada con Hanning.

Con la aplicación de esta ventana en cada una de las tramas de señal  $y(m,i)$ , se pretende conseguir un suavizado de los extremos de la trama y, de esta manera, mejorar la estimación de la DEE de señal  $\mathbf{Py}(\omega,i)$ .

### 6.3.2.- Estimación densidad espectral de energía de señal.

La estimación de la DEE de señal, voz + ruido,  $\mathbf{P}_y(\omega, i)$  se realiza de la siguiente forma:

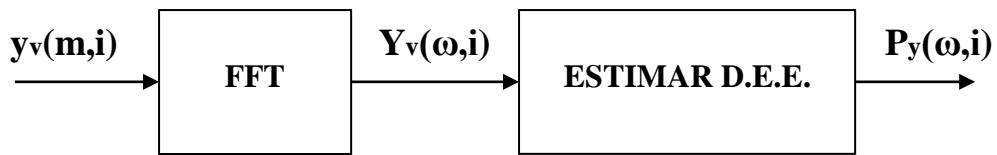


Fig. 6.15: Diagrama de bloques estimación DEE señal.

El bloque **FFT** realiza la transformada rápida de Fourier [Rodr-03] obteniendo la trama de señal en el dominio transformado  $\mathbf{Y}_v(\omega, i)$ .

El bloque **Estimar DEE** realiza la siguiente operación:

$$P_y(\omega, i) = \frac{|Y_v(\omega, i)|^2}{M} \quad (6.8)$$

Donde:  
**M**: n° muestras trama.

**ω**: frec. angular discreta:  $\omega = 2\pi k/M$ .

**k**: número muestra dominio Fourier,  $0 \leq k < M$ .

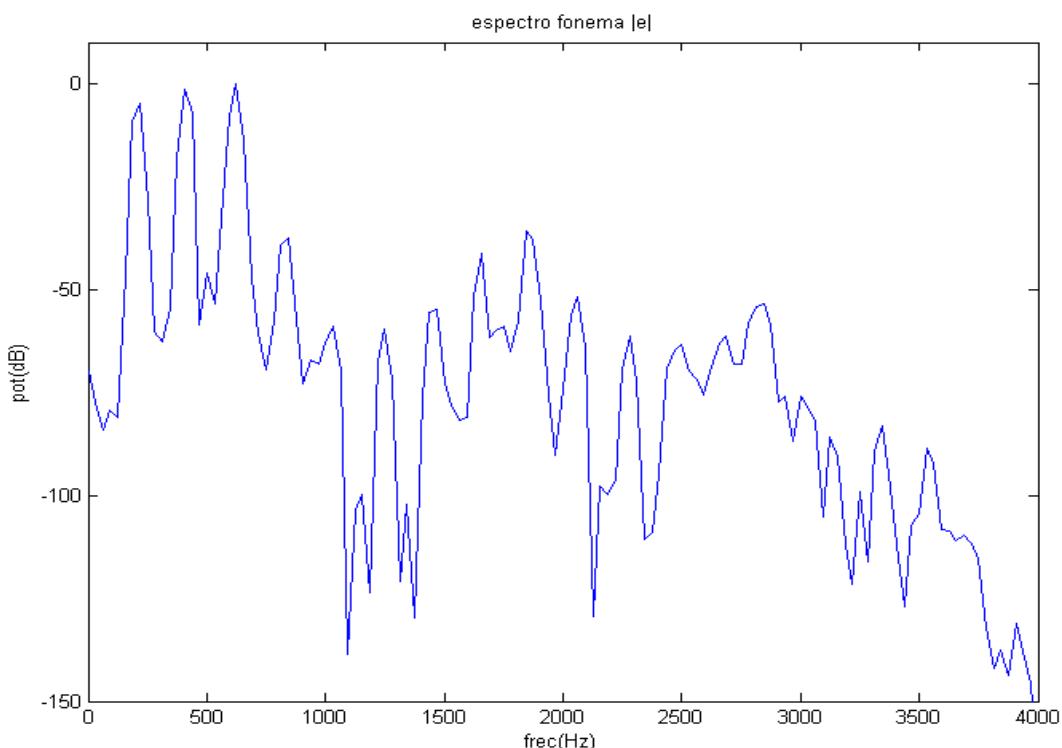


Fig. 6.16: Estimación de la DEE normalizada a 0dB del fonema /e/.

utilizando una trama de 256 muestras enventanada con Hanning

### 6.3.3.- Construcción prefiltro.

El modulo del filtro de Wiener basado en sustracción espectral que se utiliza en el prefiltrado responde a la siguiente expresión [Edua-00] y [Sovk-96]:

$$H_s(\omega, i) = \left[ \frac{P_y(\omega, i) - \beta \cdot P_d(\omega, i-1)}{P_y(\omega, i)} \right]^\delta \quad \text{con} \quad Aten \leq H_s(\omega, i) \leq 1 \quad (6.9)$$

$$\text{Donde: } Aten = \frac{\text{Nivel\_ruido}}{\text{Pot}\{P_d(\omega, i-1)\}} \quad \text{con } 0.5 \leq Aten \leq 1 \quad (6.10)$$

$$\text{Donde: } \text{Pot}\{P_d(\omega, i-1)\} = \frac{1}{M} \cdot \sum_{k=0}^{M-1} P_d(\omega, i-1) \quad (6.11)$$

Donde:  
**Py( $\omega, i$ ):** DEE de señal (voz + ruido), trama actual.

**Pd( $\omega, i-1$ ):** DEE de ruido actualizada en trama anterior.

**$\omega$ :** frec. angular discreta:  $\omega = 2\pi k/M$ .

**k:** número muestra dominio Fourier,  $0 \leq k < M$ .

**M:** n° muestras trama.

**$\beta$ :** parámetro de sobreestimación de ruido.

**$\delta$ :** parámetro de filtrado no lineal.

**Aten:** atenuación máxima del prefiltro para cada frecuencia.

**Nivel\_ruido:** potencia de ruido residual deseada (2500).

Observamos en esta expresión que la atenuación máxima de cada componente frecuencial discreta está limitada a una división entre 2, es decir, atenuación máxima de 3dB para cualquier trama de señal de entrada sea ésta de voz + ruido o ruido solamente.

Por lo que respecta a la fase, se asume que el filtro es no causal, introduciendo de este modo un retardo mínimo entre la entrada y la salida del filtro equivalente a una trama, es decir, 256 muestras a 8Khz que en tiempo se traducen en 32 ms. Asumiendo este retardo consideraremos que la fase de la trama de señal prefiltrada será la misma que la de la trama de señal de entrada al prefiltro.

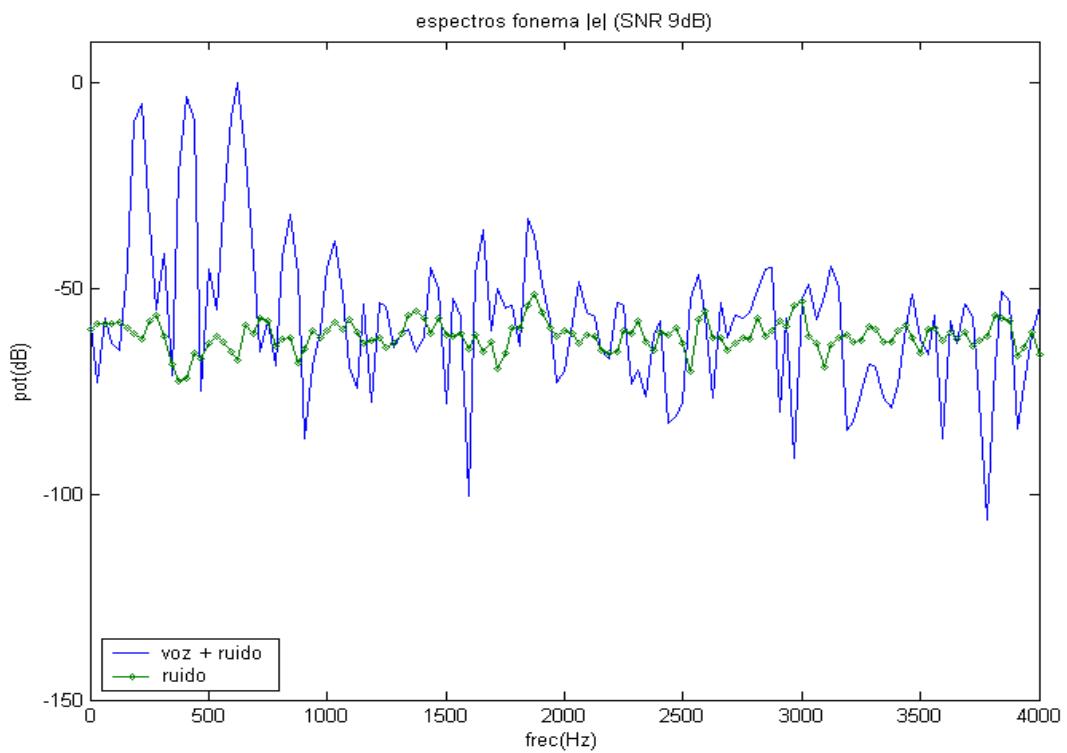


Fig 6.17: DEEs correspondientes a  $\mathbf{P}_y(\omega, i)$  (voz + ruido) y  $\mathbf{P}_d(\omega, i-1)$  (ruido), del fonema /e/ bajo condiciones de SNR=9dB (ruido blanco).

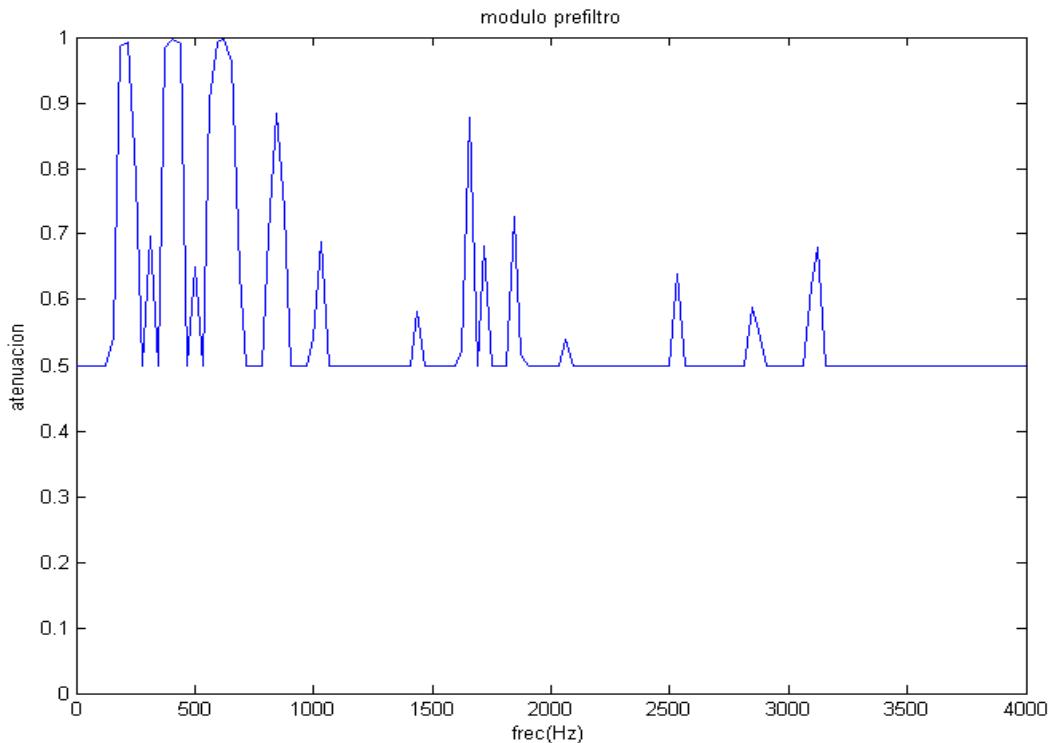


Fig 6.18: Modulo prefiltro construido para el fonema /e/. Utilizando los parámetros  $\delta=2$  y  $\beta=1.2$ .

### 6.3.4.- prefiltroado.

El prefiltroado de la señal se realiza en el dominio de la frecuencia y responde al siguiente diagrama de bloques:

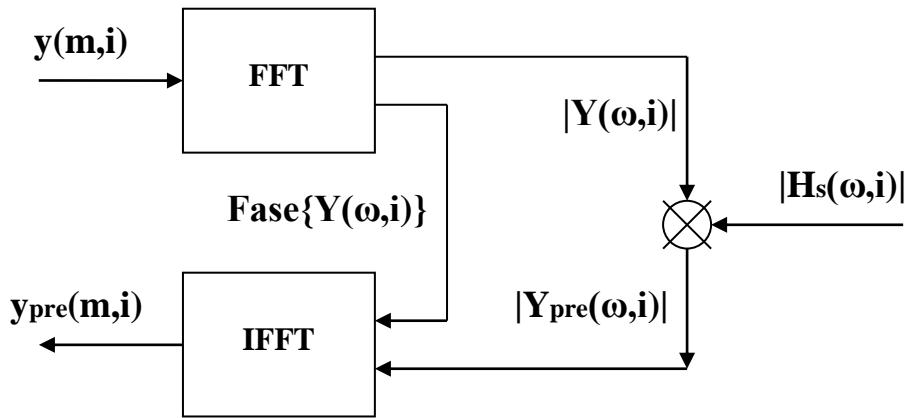


Fig. 6.19: Diagrama de bloques del prefiltroado.

Como se puede observar en la figura 6.19, para realizar los cambios de dominio se utilizan la FFT e IFFT complejas. En el caso de la FFT, trasformada rápida de Fourier, se obtiene una trama de señal transformada compleja  $\mathbf{Y}(\omega,i)$  a partir de una trama de señal real  $\mathbf{y}(m,i)$ , en cambio, en la IFFT solo nos interesará la parte real de la trama de señal  $\mathbf{y}_{pre}(m,i)$ , obtenida a partir de una trama compleja en el dominio transformado  $\mathbf{Y}_{pre}(\omega,i)$ .

El modulo en el dominio frecuencial de la trama de señal prefiltrada  $|\mathbf{Y}_{pre}(\omega,i)|$  se obtiene del producto:

$$|\mathbf{Y}_{pre}(\omega,i)| = |\mathbf{Y}(\omega,i)| \cdot |\mathbf{H}_s(\omega,i)| \quad (6.12)$$

Donde:  $\mathbf{Y}(\omega,i)$ : Transformada de Fourier  $y(m,i)$

$\mathbf{H}_s(\omega,i)$ : Filtro sustracción trama  $i$ .

$\omega$ : freq. angular discreta:  $\omega = 2\pi k/M$ .

$k$ : número muestra dominio Fourier,  $0 \leq k < M$ .

$M$ : n° muestras trama.

La fase, en cambio, se considera que es la misma que la señal de entrada:

$$\text{fase}\{\mathbf{Y}_{pre}(\omega,i)\} = \text{fase}\{\mathbf{Y}(\omega,i)\} \quad (6.13)$$

Donde:  $\mathbf{Y}(\omega,i)$ : Transformada de Fourier  $y(m,i)$

**$\omega$** : frec. angular discreta:  $\omega = 2\pi k/M$ .  
**k**: número muestra dominio Fourier,  $0 \leq k < M$ .  
**M**: n° muestras trama.

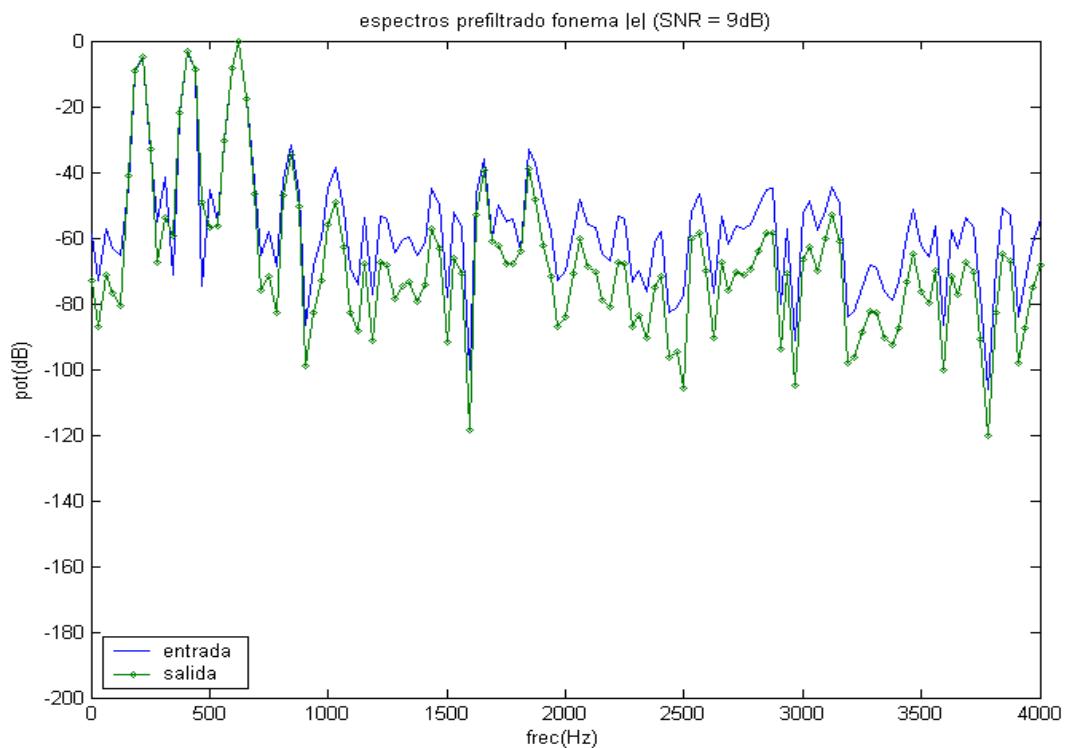


Fig. 6.20: Espectros normalizados de entrada y salida del prefiltro para el fonema /e/ bajo condiciones de ruido blanco SNR = 9 dB.

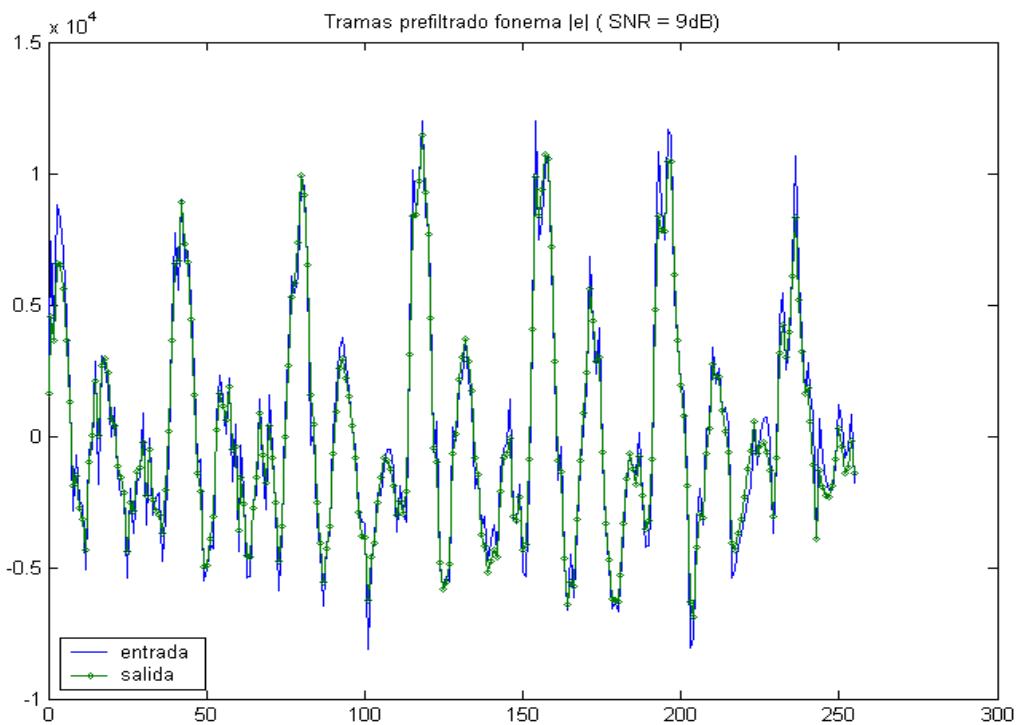


Fig. 6.21: Tramas en tiempo de entrada y salida del prefiltro para el fonema /e/ bajo condiciones de ruido blanco SNR = 9 dB.

#### 6.4.- Filtro

El filtro realiza una supresión adaptativa de ruido, en función del ruido presente en la trama de señal prefiltrada  $y_{pre}(m,i)$ , así como la detección de tramas de voz y silencio, por medio de un VAD, para poder realizar la reestimación del ruido presente en la señal  $d(n)$ .

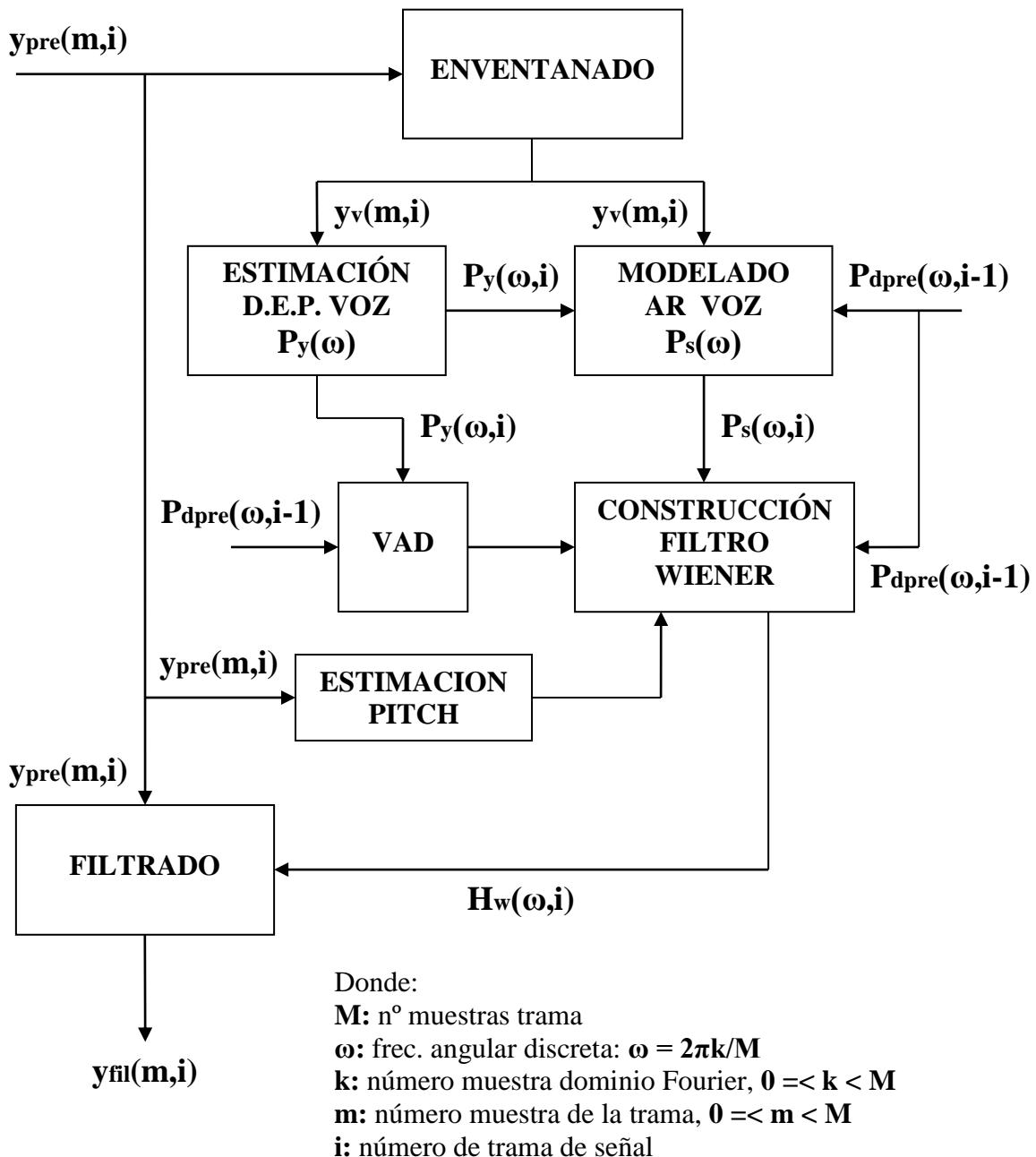


Fig. 6.22: Diagrama de bloques del filtro.

#### 6.4.1.- Enventanado de trama.

El enventanado de la trama de entrada  $y_{\text{pre}}(m,i)$  se realiza de la misma forma que en el punto 6.3.1:

$$y_v(m,i) = y_{\text{pre}}(m,i) \cdot v_{\text{hann}}(m) \quad (6.14)$$

Con:

$$v_{\text{hann}}(m) = 0.5 \cdot \left[ 1 - \cos\left(2 \cdot \pi \cdot \frac{m}{M-1}\right) \right] \quad (6.15)$$

Donde: **m**: número muestra de la trama,  $0 \leq m < M$

**M**: n° muestras de la trama

#### 6.4.2.- Estimación densidad espectral de energía de señal.

La estimación de la DEE de señal, voz + ruido,  $P_y(\omega,i)$  se realiza de manera idéntica al punto 6.3.2:

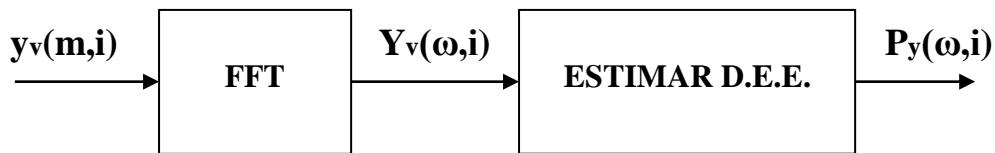


Fig. 6.23: Diagrama de bloques estimación DEE señal.

El bloque FFT realiza la transformada rápida de Fourier obteniendo la trama de señal en el dominio transformado  $Y_v(\omega,i)$ .

El bloque Periodograma realiza la siguiente operación:

$$P_y(\omega,i) = \frac{|Y_v(\omega,i)|^2}{M} \quad (6.16)$$

Donde: **M**: n° muestras trama.

**ω**: frec. angular discreta:  $\omega = 2\pi k/M$ .

**k**: número muestra dominio Fourier,  $0 \leq k < M$ .

#### 6.4.3.- Modelado AR de la voz.

La DEE de voz se calcula según la siguiente figura:

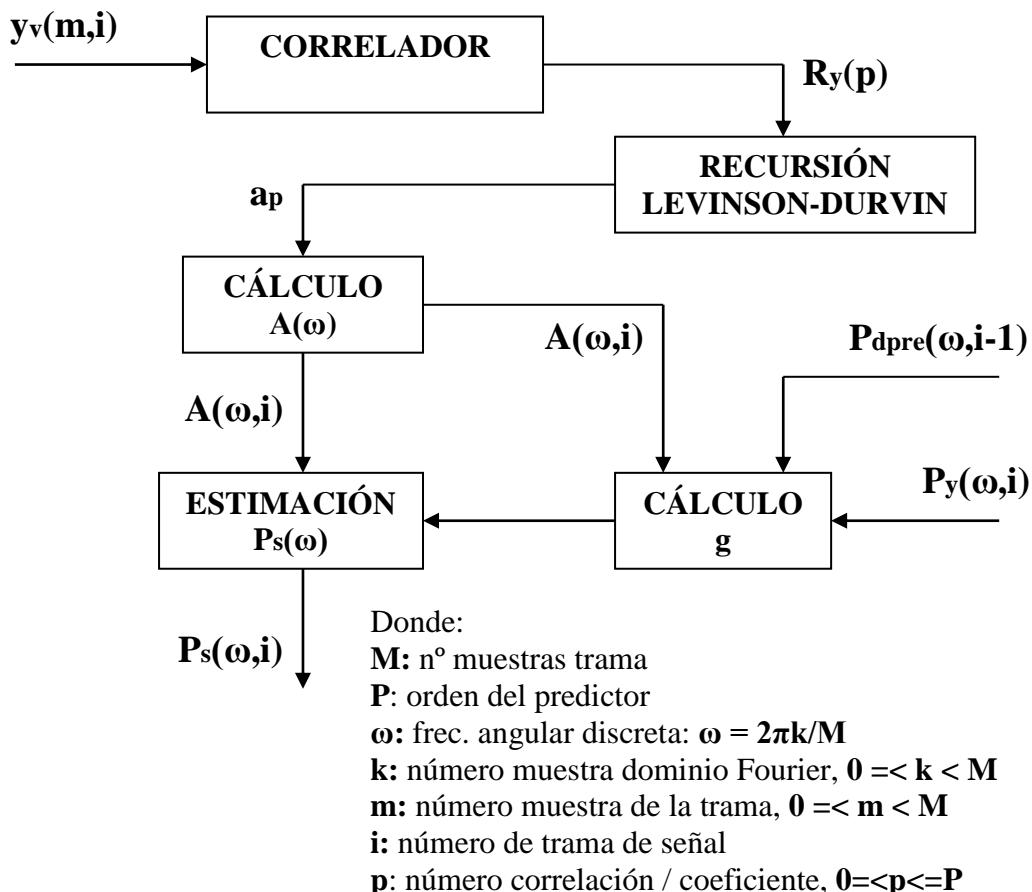


Fig. 6.24: Esquema para la estimación espectro LPC

El bloque **correlador** realiza la siguiente operación:

$$R(p) = \sum_{m=0}^{M-p-1} y_v(m,i) \cdot y_v(m+p,i) \quad (6.17)$$

Donde:  
**M:** n° muestras trama  
**P:** orden del predictor  
**m:** número muestra de la trama,  $0 \leq m < M$   
**i:** número de trama de señal  
**p:** número correlación / coeficiente,  $0 \leq p \leq P$

El bloque recursión de Levinson-Durbin realiza la siguiente secuencia de operaciones, para resolver de forma eficiente las ecuaciones de Yule-Walker para obtener los coeficientes de predicción **a<sub>p</sub>**:

$$\begin{aligned}
 E_0 &= R_y(0) \\
 \text{para } p &= 1, \dots, P \\
 r_p &= -\frac{R_y(p) + \sum_{k=1}^{p-1} a_k^{p-1} R_y(p-k)}{E_{p-1}} \\
 \text{para } r &= 1, p-1 \\
 a_k^p &= a_k^{p-1} + r_p a_{p-k}^{p-1} \\
 \text{end} \\
 a_p^p &= r_p \\
 E_p &= E_{p-1}(1 - r_p^2) \\
 \text{end}
 \end{aligned} \tag{6.18}$$

Donde:  
**P:** orden del predictor  
**p:** número coeficiente  
**r:** coeficiente de reflexión  
**E:** Energía error residual de predicción  
**a:** coeficiente predicción AR  
**Ry:** autocorrelación señal entrada (voz + ruido)

El cálculo del denominador de la DEE de voz **Ps(ω)**, es decir, **A(ω)** se realiza mediante la siguiente ecuación:

$$A(\omega, i) = 1 + \sum_{p=1}^P a_p e^{-j\omega p} \tag{6.19}$$

Donde:  
**M:** n° muestras trama  
**P:** orden del predictor  
**ω:** freq. angular discreta:  $\omega = 2\pi k/M$   
**k:** número muestra dominio Fourier,  $0 \leq k < M$   
**p:** número coeficiente  
**i:** número de trama de señal

El cálculo de la constante de ganancia del espectro de modelado AR **g** se realiza mediante la siguiente ecuación:

$$g^2 = \left[ \left( \frac{1}{M} \cdot \sum_{k=0}^{M-1} P_y(\omega, i) \right) - \left( \frac{1}{M} \cdot \sum_{k=0}^{M-1} P_{\text{pre}}(\omega, i-1) \right) \right] \cdot \frac{1}{\frac{1}{M} \cdot \sum_{k=0}^{M-1} \frac{1}{|A(\omega)|^2}} \quad (6.20)$$

Donde:  
**M**: n° muestras trama

**ω**: frec. angular discreta:  $\omega = 2\pi k/M$

**k**: número muestra dominio Fourier

**i**: número de trama de señal

Finalmente la DEE de modelado AR correspondiente a la voz  $P_s(\omega)$ , se calcula de la forma:

$$P_s(\omega, i) = \frac{g^2}{|A(\omega, i)|^2} \quad (6.21)$$

Donde:  
**M**: n° muestras trama

**ω**: frec. angular discreta:  $\omega = 2\pi k/M$

**k**: número muestra dominio Fourier

**i**: número de trama de señal

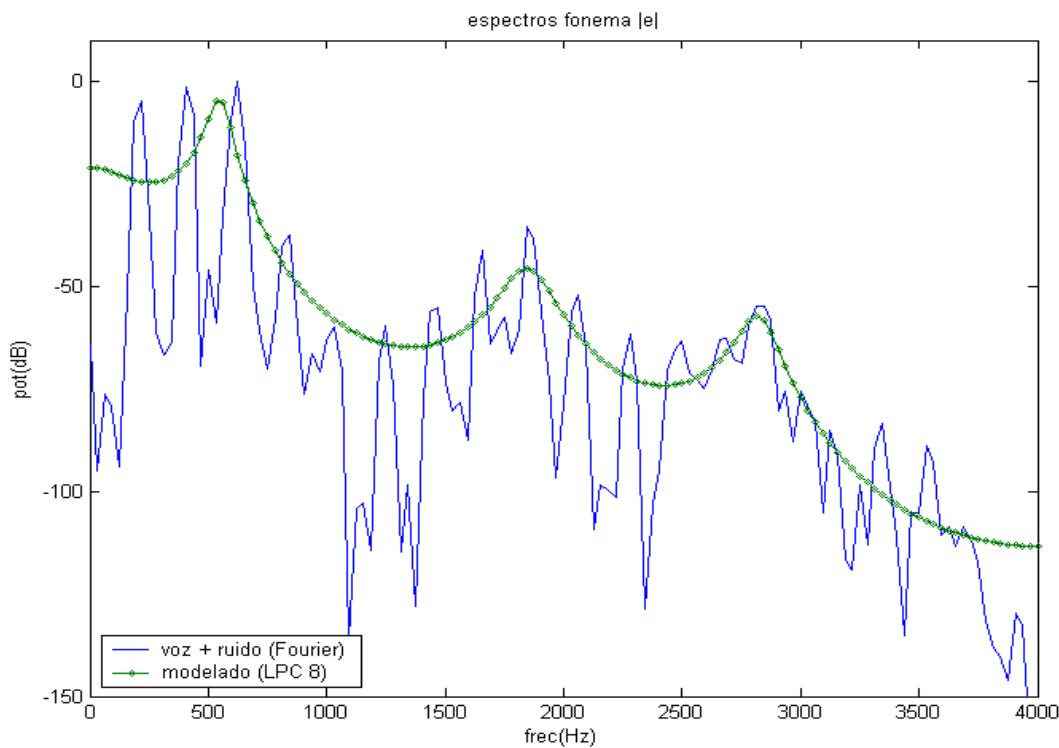


Fig 6.25: DEEs correspondientes a  $P_y(\omega, i)$  (voz + ruido) y  $P_s(\omega, i)$  (voz), bajo condiciones de señal limpia.

#### 6.4.4.- VAD.

El VAD, o detector de actividad de voz, detecta la actividad de voz a partir del esquema siguiente:

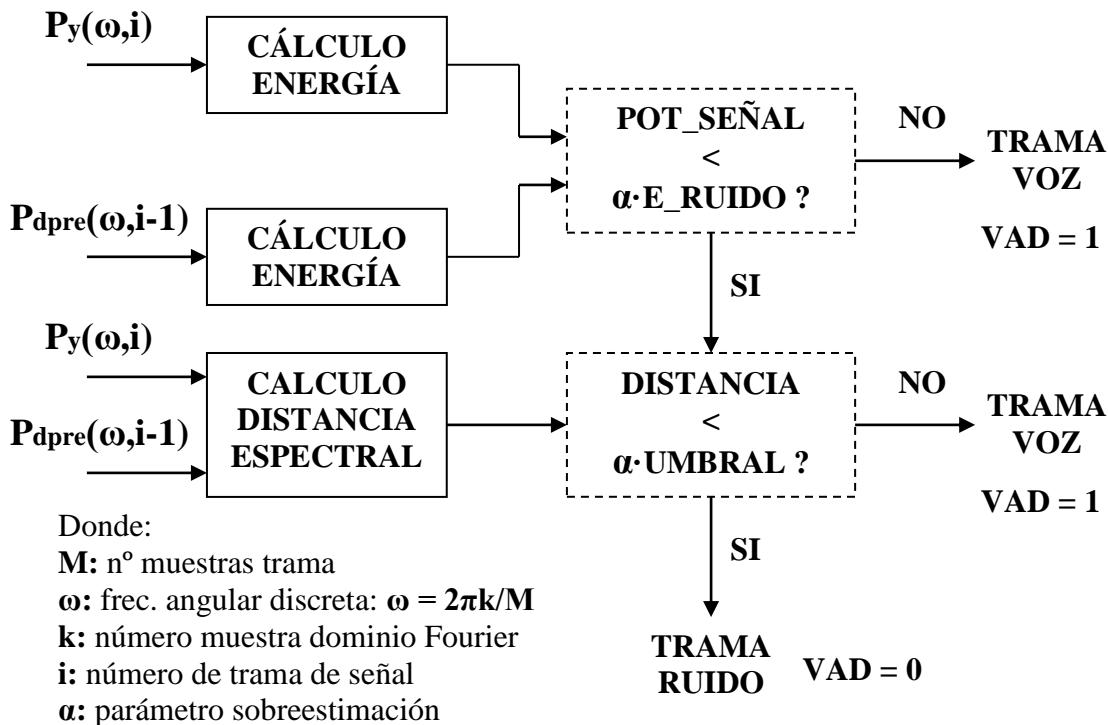


Fig. 6.26: Diagrama de decisión del detector de voz.

El cálculo de las energías de la trama de señal  $y(m,i)$  y ruido  $d_{pre}(m,i-1)$  se realiza a partir de sus DEEs mediante las siguientes ecuaciones:

$$E_y = \frac{1}{M} \cdot \sum_{k=0}^{M-1} P_y(\omega, i-1) \quad (6.22)$$

$$E_{d_{pre}} = \frac{1}{M} \cdot \sum_{k=0}^{M-1} P_{d_{pre}}(\omega, i-1) \quad (6.23)$$

Y el cálculo de la distancia espectral se realiza también a partir de sus respectivas DEEs mediante la ecuación:

$$d(P_y(\omega, i), P_{d_{pre}}(\omega, i-1)) = \sum_{k=0}^{M-1} |\log(P_y^{\text{norm}}(\omega, i)) - \log(P_{d_{pre}}(\omega, i-1))| \quad (6.24)$$

$$\text{Con: } P_y^{\text{norm}}(\omega, i) = P_y(\omega, i) \cdot \frac{E_{\text{dpre}}}{E_y} \quad (6.25)$$

Donde: **M**: n° muestras trama

**ω**: freq. angular discreta:  $\omega = 2\pi k/M$

**k**: número muestra dominio Fourier

**i**: número de trama de señal

Según se observa en el diagrama de la fig. 6.26 para tomar la decisión de que una cierta trama sea de voz + ruido o sólo ruido, se tienen que realizar dos test:

- En el primer test se debe superar la energía que impone por la estimación de ruido  $P_{\text{dpre}}(\omega, i-1)$  con la sobreestimación de esta energía por el parámetro **a**, para poder determinar que la trama actual es voz + ruido,  $y_{\text{pre}}(m, i) = s(m, i) + d_{\text{pre}}(m, i)$ .
- Si no se supera el primer test, se realiza un segundo test, en el que finalmente se decidirá que la trama es sólo ruido  $y_{\text{pre}}(m, i) = d_{\text{pre}}(m, i)$ , si la distancia espectral no supera el **umbral** de distancia, calculado en la estimación inicial de ruido, sobreestimado por el parámetro **a**, en caso contrario se decidirá finalmente que la trama es voz + ruido.

$$\text{umbral} = \max \{d(P_y(\omega, i), P_{\text{dpre}}(\omega, i-1))\} \quad \text{con } 0 \leq i < R \quad (6.26)$$

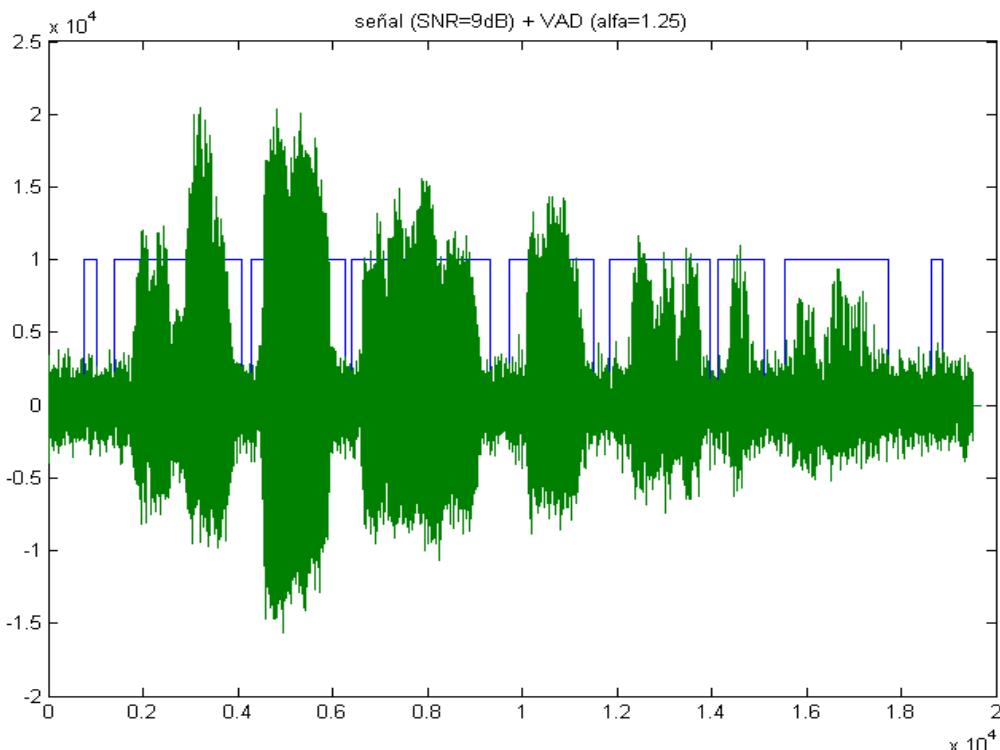


Fig. 6.27: Detección de actividad de voz por el VAD con  $\alpha=1.25$  utilizando la señal ASUN1 + ruido blanco (SNR=9dB).

#### 6.4.5.- Estimación pitch.

La estimación del pitch, es decir, del armónico principal de la voz, se realiza según el siguiente esquema:

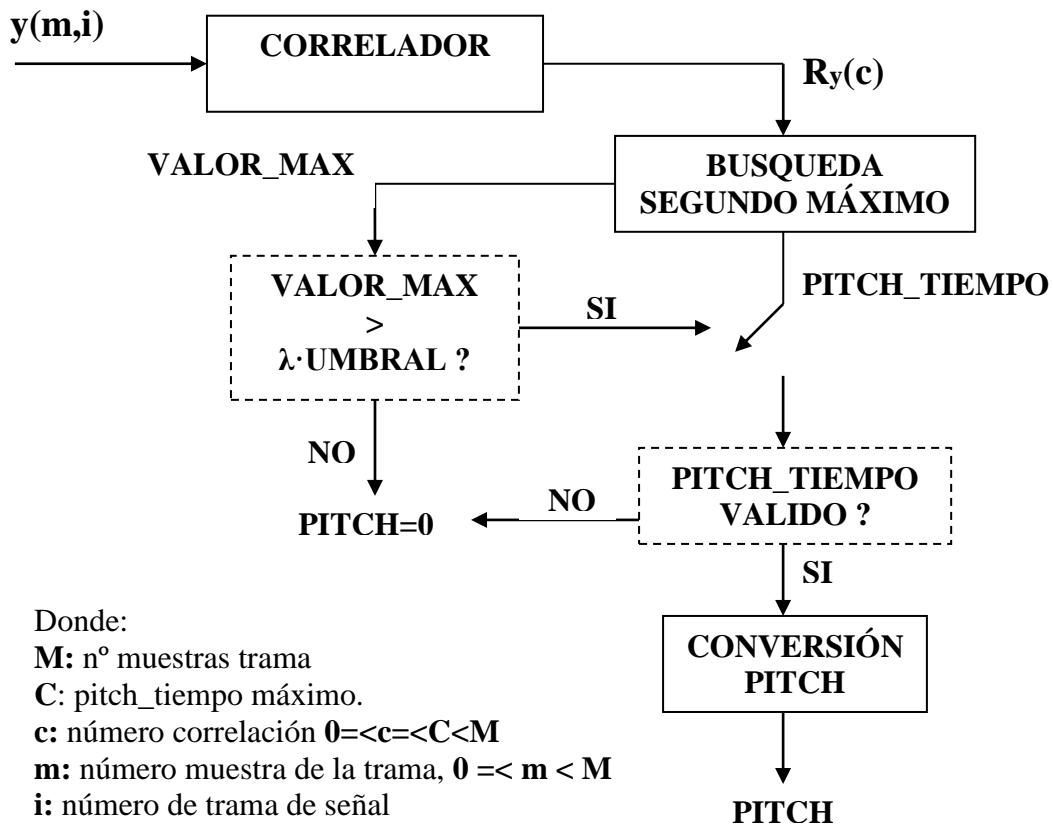


Fig. 6.28: Diagrama de bloques para la búsqueda del pitch de voz.

El bloque cálculo correlaciones realiza la siguiente operación:

$$R_y(c) = \sum_{m=0}^{M-1} y_{\text{pre}}(m,i) \cdot y_{\text{pre}}(m+c,i) \quad (6.27)$$

Donde:  
**M:** n° muestras trama  
**C:** pitch\_tiempo máximo.  
**c:** número correlación  $0 \leq c \leq C < M$   
**m:** número muestra de la trama,  $0 \leq m < M$   
**i:** número de trama de señal

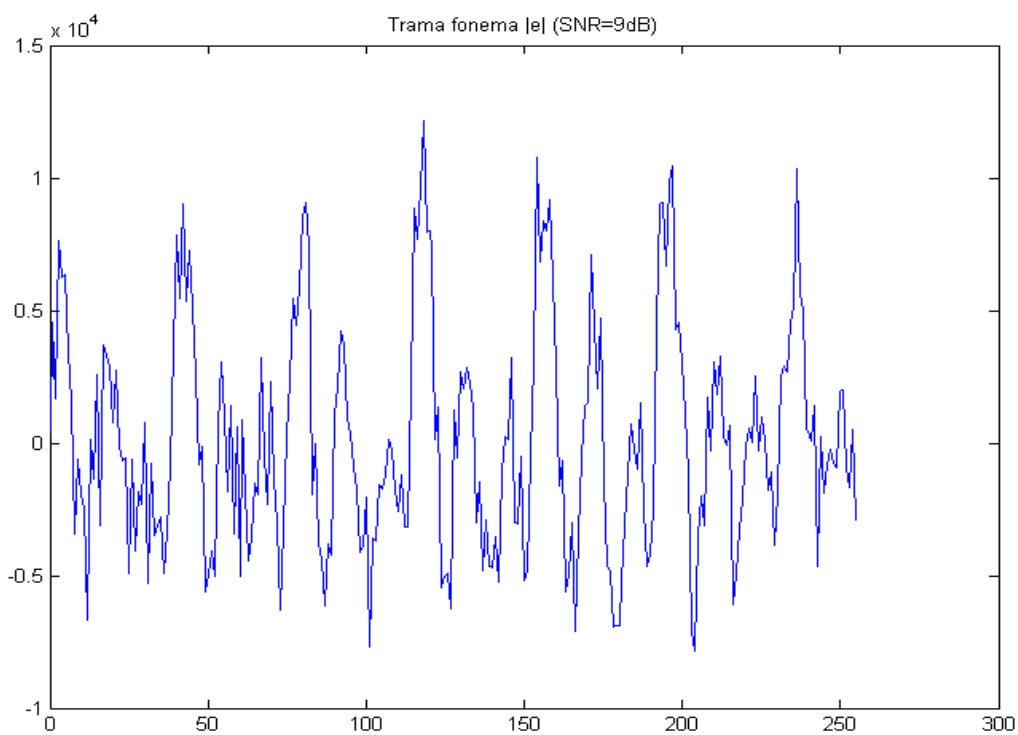


Fig. 6.29: Trama de voz correspondiente al fonema /e/ contaminada con ruido blanco (SNR = 9dB) .

La búsqueda del segundo máximo de la función de correlación de la trama  $y(m,i)$  se realiza porque el primer máximo siempre corresponderá a  $Ry(0)$ . Seguimos las siguientes pautas:

- Primero se busca el primer mínimo de la función autocorrelación, que para una trama periódica corresponde al mínimo de mayor valor absoluto.
- A partir de la posición de este mínimo, se busca el máximo de mayor valor que, para una trama periódica corresponderá al segundo máximo que nos dará el valor de periodo de la señal.

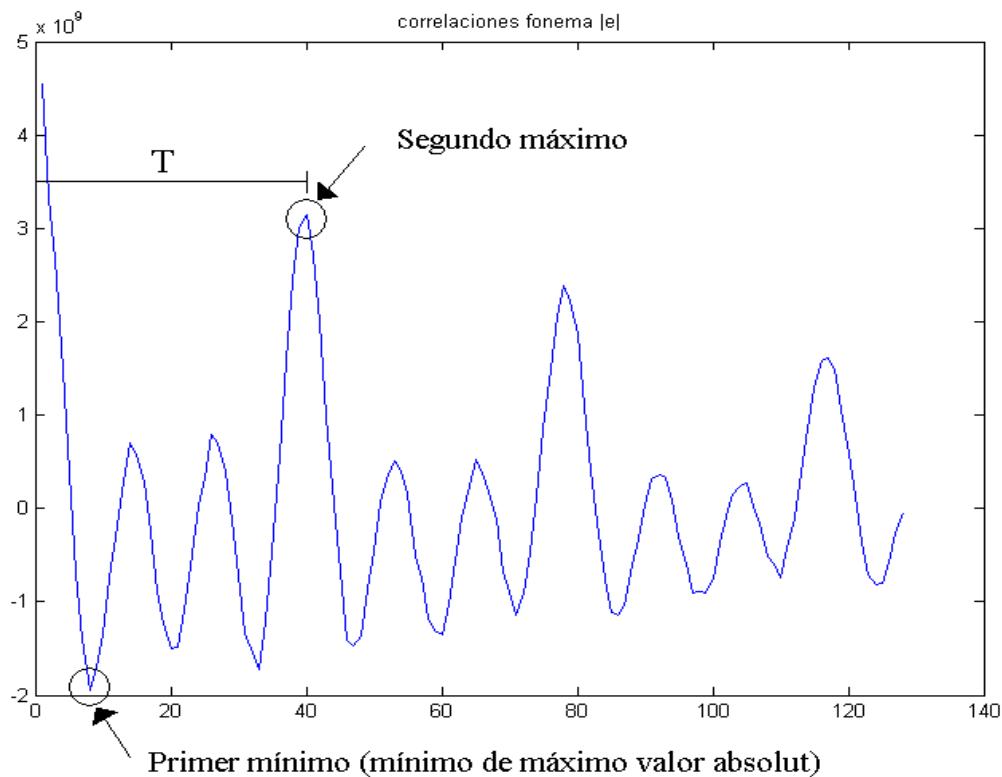


Fig.6.30: Posición del primer mínimo y del segundo máximo en las correlaciones calculadas para el fonema /e/ (SNR=9dB).

Para asegurarnos de que hemos obtenido un periodo de trama válido es necesario que el valor del segundo máximo sea superior a un umbral, que se calcula de la siguiente manera:

$$\text{Umbral} = R_y(0) \cdot \frac{M - T}{M} \quad (6.28)$$

Donde:  
**M**: n° muestras trama  
**T**: pitch\_tiempo/periodo calculado

Para considerar que un periodo/pitch\_tiempo es válido, debe estar en un rango entre 20 muestras (400 Hz) y 100 muestras (80Hz), entonces  $20 \leq T \leq 100$  muestras.

Una vez obtenido un periodo/pitch\_tiempo válido se procede a realizar la conversión de periodo temporal a frecuencia obteniendo el pitch de la señal según la ecuación:

$$\text{Pitch} = \frac{F_m}{T} \quad (6.29)$$

Donde:  
**Fm**: frecuencia de muestreo señal  $y(m,i)$ , **Fm = 8 KHz**

**T:** pitch\_tiempo/periodo calculado

El valor de pitch calculado, considerando la limitación impuesta en el cálculo del período, estará acotado en un rango de frecuencias de  $80 \leq \text{Pitch} \leq 400 \text{ Hz}$ .

#### 6.4.6.- Construcción filtro.

El módulo de la respuesta frecuencial utilizada en la segunda etapa de filtrado es una mezcla entre filtro de Wiener de modelado AR y filtro peine, este módulo se construye a partir de las siguientes ecuaciones:

$$H_f(\omega, i) = H_w(\omega, i) \cdot H_p(\omega, i) \quad \text{con} \quad Aten \leq H_f(\omega, i) \leq 1 \quad (6.30)$$

$$\text{Donde: } Aten = \frac{\text{Nivel\_ruido}}{\text{Pot}\{P_d(\omega, i-1)\}} \quad \text{con} \quad 0.1 \leq Aten \leq 1 \quad (6.31)$$

$$H_w(\omega, i) = \left[ \frac{P_s(\omega, i)}{P_s(\omega, i) + \beta \cdot P_d(\omega, i-1)} \right]^\delta \quad \text{si} \quad \text{VAD} = 1 \quad (\text{trama de voz}) \quad (6.32)$$

$$H_w(\omega, i) = Aten \quad \text{para toda } k \quad \text{si} \quad \text{VAD} = 0 \quad (\text{trama de silencio/ruido}) \quad (6.33)$$

$$H_p(\omega, i) = 1 \quad \text{para} \quad k = h-1, h, h+1, 2 \cdot h-1, 2 \cdot h, 2 \cdot h+1, \dots \quad (6.34)$$

$$\text{Con } h = \left\lfloor \text{pitch} \cdot \frac{M}{F_m} + 0.5 \right\rfloor \quad (6.35)$$

$$H_p(\omega, i) = 0 \quad \text{para cualquier otra } k$$

$$\text{Si} \quad \text{Pitch}=0 \quad \text{entonces} \quad H_p(\omega, i)=1 \quad \text{para toda } k$$

Donde:**Ps( $\omega, i$ ):** DEE de modelado de voz, trama actual.

**Pd( $\omega, i-1$ ):** DEE de ruido actualizada en trama anterior

**$\omega$ :** freq. angular discreta:  $\omega = 2\pi k/M$ .

**k:** número muestra dominio Fourier,  $0 \leq k < M$ .

**M:** n° muestras trama.

**$\beta$ :** parámetro de sobreestimación de ruido

**$\delta$ :** parámetro de filtrado no lineal

**Aten:** atenuación máxima del filtro para cada frecuencia

**Nivel\_ruido:** potencia de ruido residual deseada

Observamos en esta expresión que la atenuación máxima de cada componente frecuencial discreta esta limitada a una división entre 10, es decir, atenuación máxima de 20dB para cualquier trama de señal de entrada sea esta de voz + ruido o ruido solamente.

Por lo que respecta a la fase, se asume que el filtro es no causal, introduciendo de este modo un retardo mínimo entre la entrada y la salida del filtro equivalente a una trama, es decir, 256 muestras a 8Khz que en tiempo se traducen en 32 ms. Asumiendo este retardo consideraremos que la fase de la trama de señal filtrada será la misma que la de la trama de señal de entrada al filtro.

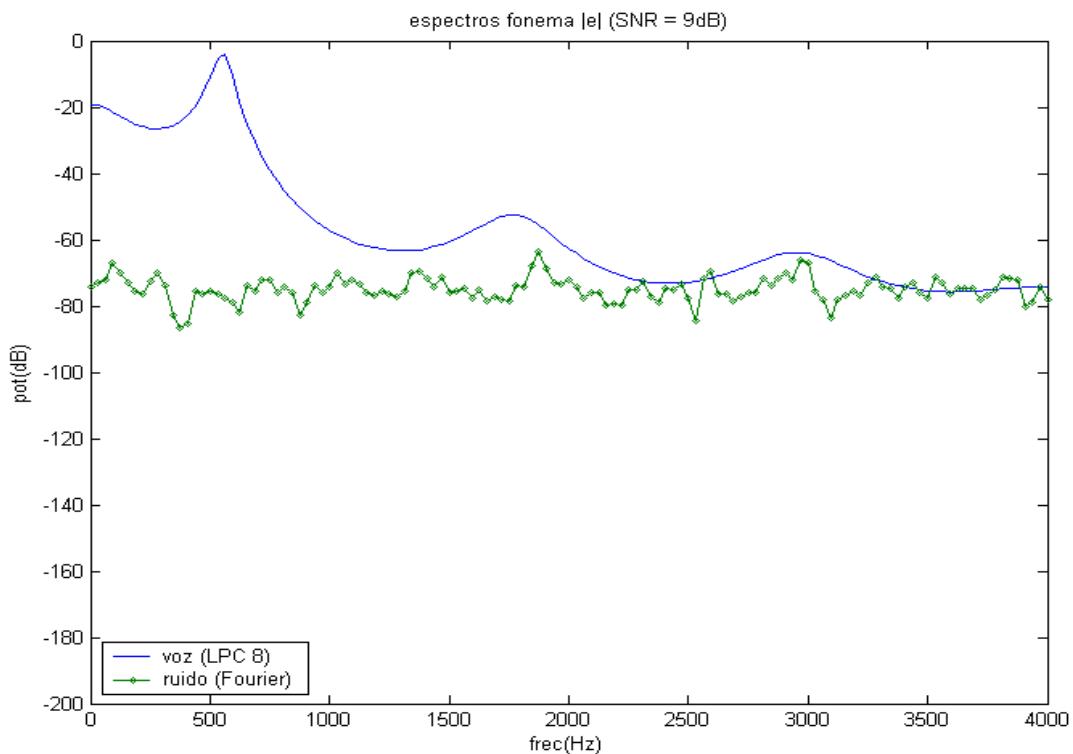


Fig 6.31: DEEs correspondientes a  $\mathbf{Ps}(\omega, i)$  (voz) y  $\mathbf{Pd}(\omega, i-1)$  (ruido), para El fonema /e/ bajo condiciones de SNR=9dB (ruido blanco).

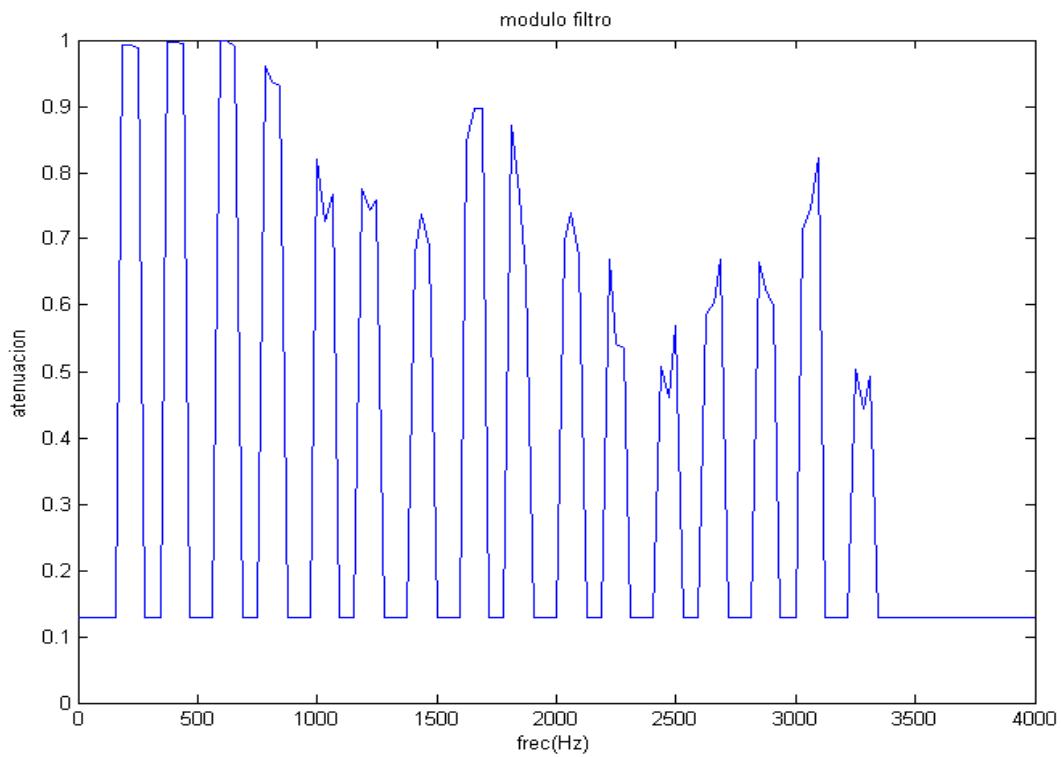
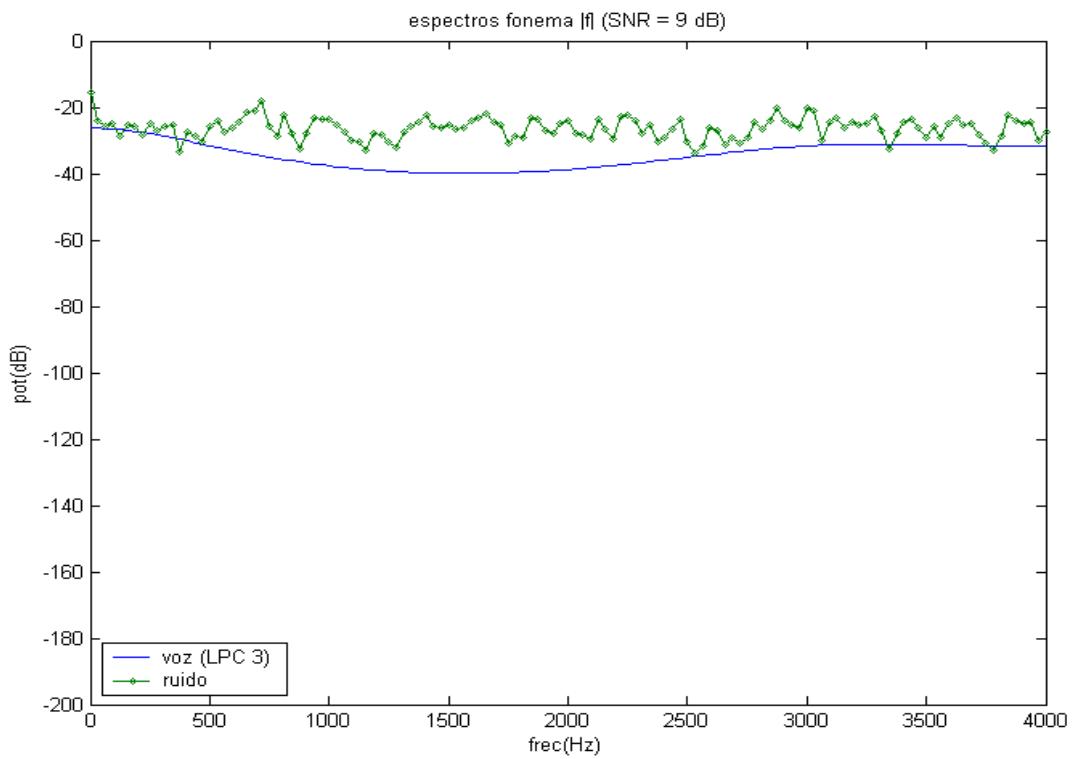


Fig 6.32: Modulo filtro construido para el fonema /e/.

Fig. 6.33: DEEs correspondientes a  $Ps(\omega, i)$  (voz) y  $Pd(\omega, i-1)$  (ruido), para El fonema /f/ bajo condiciones de SNR=9dB (ruido blanco).

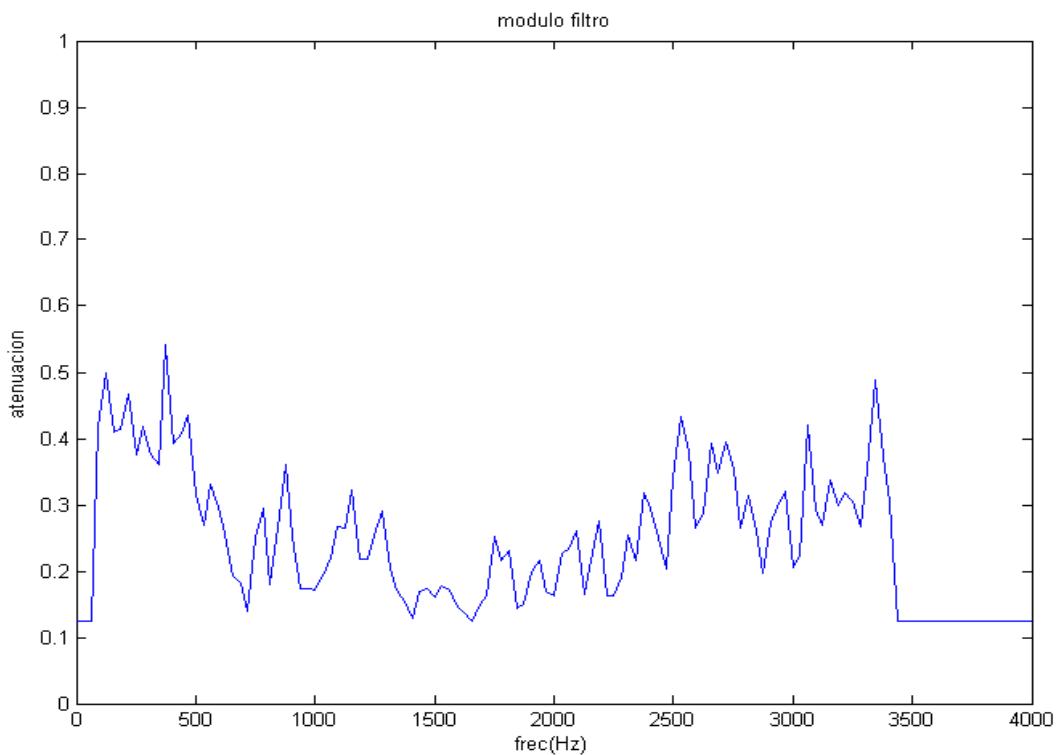


Fig 6.34: Modulo filtro construido para el fonema /f/.

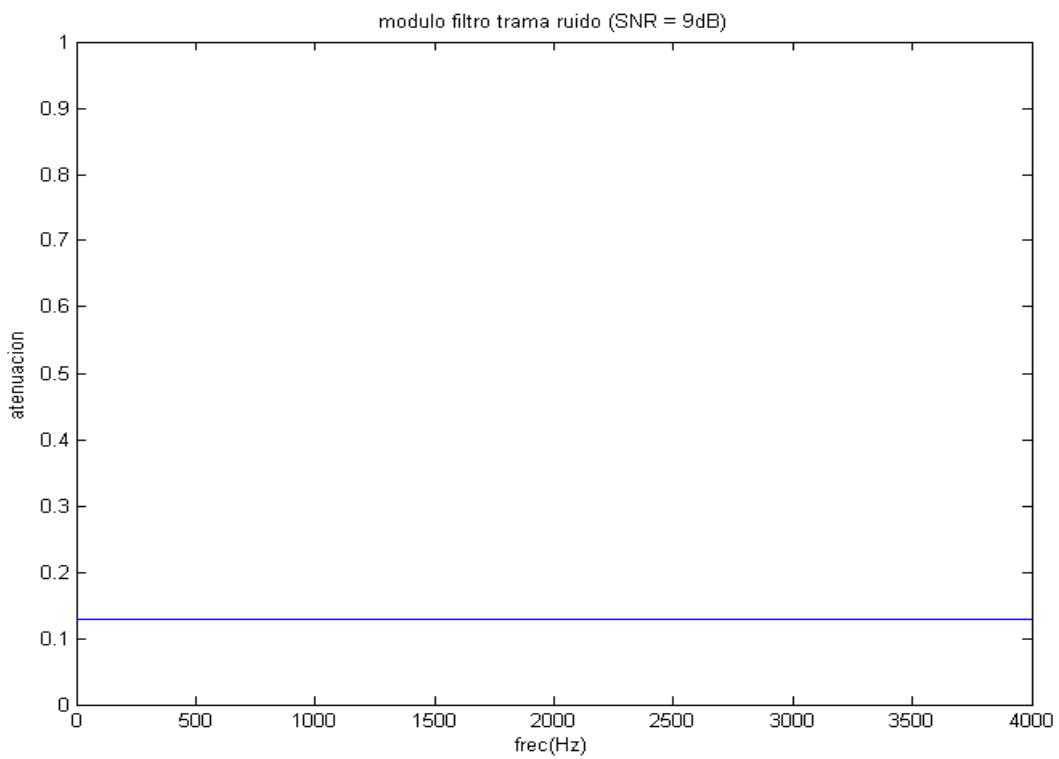


Fig. 6.35: Modulo filtro construido para una trama de ruido (VAD=0).

#### 6.4.7.- filtrado.

El filtrado de la trama de señal  $y_{pre}(m,i)$  se realiza en el dominio de la frecuencia y responde al siguiente diagrama de bloques:

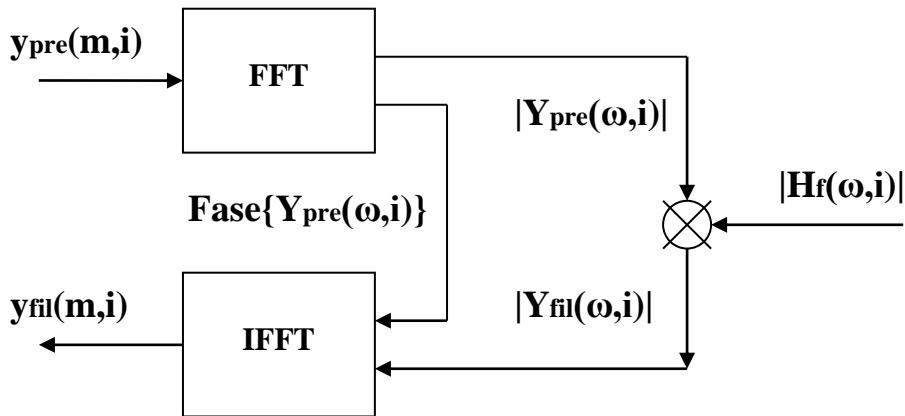


Fig. 6.36: Diagrama de bloques del filtrado.

Como se puede observar en la figura 6.36, para realizar los cambios de dominio se utilizan la FFT e IFFT complejas. En el caso de la FFT, transformada rápida de Fourier, se obtiene una trama de señal transformada compleja  $\mathbf{Y}_{pre}(\omega,i)$  a partir de una trama de señal real  $y_{pre}(m,i)$ , en cambio, en la IFFT solo nos interesará la parte real de la trama de señal  $y_{fil}(m,i)$ , obtenida a partir de una trama compleja en el dominio transformado  $\mathbf{Y}_{fil}(\omega,i)$ .

El modulo en el dominio frecuencial de la trama de señal prefiltrada  $|\mathbf{Y}_{fil}(\omega,i)|$  se obtiene del producto:

$$|\mathbf{Y}_{fil}(\omega,i)| = |\mathbf{Y}_{pre}(\omega,i)| \cdot |H_f(\omega,i)| \quad (6.36)$$

Donde:  $\mathbf{Y}_{pre}(\omega,i)$ : Transformada de Fourier  $y_{pre}(m,i)$

$H_f(\omega,i)$ : Filtro construido, trama  $i$ .

$\omega$ : freq. angular discreta:  $\omega = 2\pi k/M$ .

$k$ : número muestra dominio Fourier,  $0 \leq k < M$ .

$M$ : n° muestras trama.

La fase, en cambio, se considera que es la misma que la señal de entrada:

$$\text{fase}\{\mathbf{Y}_{fil}(\omega,i)\} = \text{fase}\{\mathbf{Y}_{pre}(\omega,i)\} \quad (6.37)$$

Donde:  $\mathbf{Y}(\omega,i)$ : Transformada de Fourier  $y(m,i)$

$\omega$ : freq. angular discreta:  $\omega = 2\pi k/M$ .

**k:** número muestra dominio Fourier,  $0 \leq k < M$ .  
**M:** n° muestras trama.

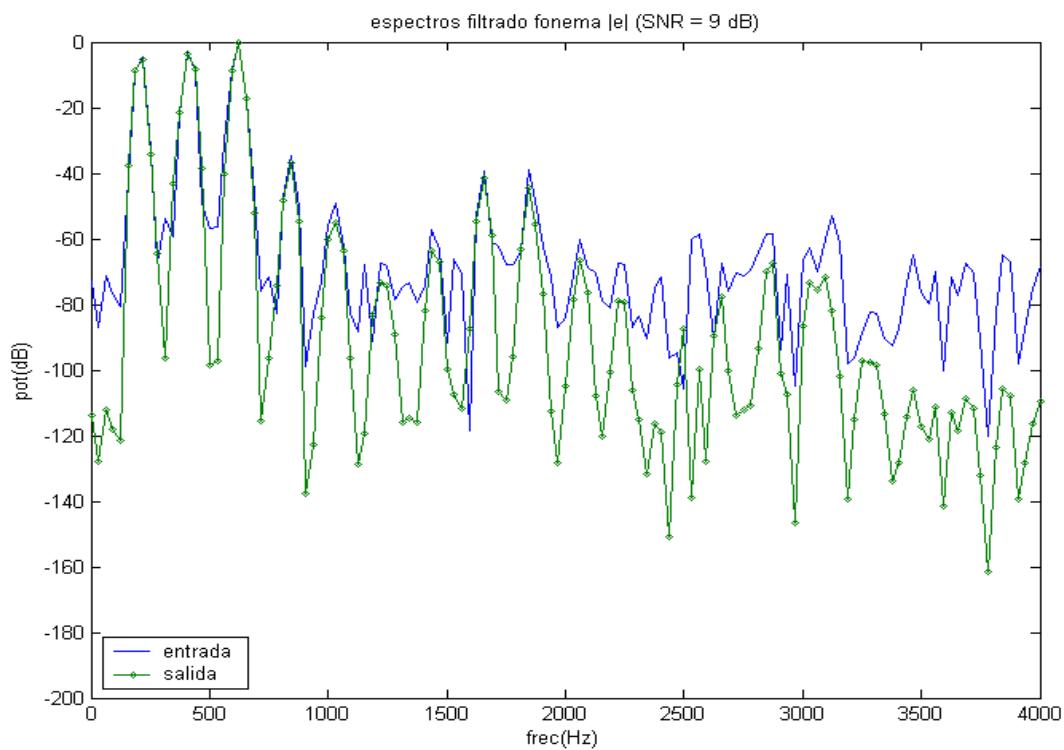


Fig. 6.37: Espectros de entrada y salida del filtro para el fonema /e/ bajo condiciones de ruido blanco SNR = 9 dB.

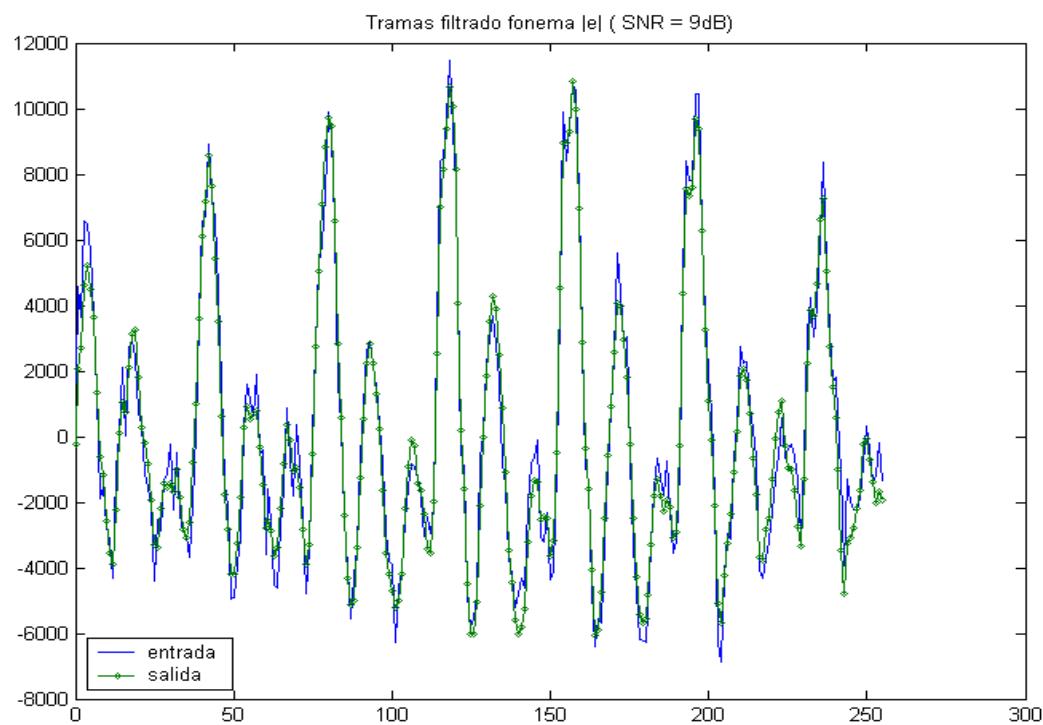
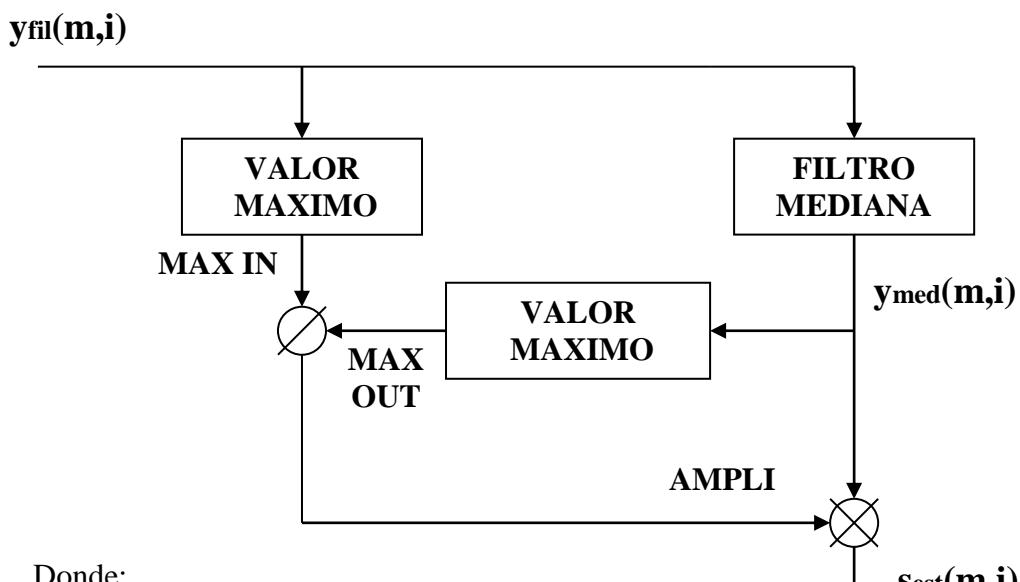


Fig. 6.38: Tramas en tiempo de entrada y salida del filtro para el fonema /e/ bajo condiciones de ruido blanco SNR = 9 dB.

### 6.5.- Postfiltrado.

La función del postfiltrado es eliminar el ruido residual que haya quedado de las etapas de filtrado anteriores, así como la eliminación de ruido musical generado en el filtrado. También tiene como objetivo eliminar el ruido de tipo impulsional que contenga la trama  $y_{fil}(m,i)$ .

Este postfiltrado no es mas que un simple filtro de mediana, pero con unas modificaciones que evitan el efecto de saturación de picos y conserva los márgenes dinámicos de la trama  $y_{fil}(m,i)$ . El esquema de este postfiltrado es el siguiente:



Donde:

**M:** n° muestras trama

**m:** número muestra de la trama,  $0 \leq m < M$

**i:** número de trama de señal

Fig.6.39: Diagrama de bloques del filtro de mediana.

El filtrado de mediana utiliza un elemento estructurante de 3 muestras;  $m-1, m, m+1$  donde  $1 \leq m \leq M-1$  con **M=número de muestras por trama**. Este elemento estructurante recorre la trama entera comparando los valores de las tres muestras que en cada momento señale este elemento, y ejecutando las siguientes directrices:

- Si el valor de las tres muestras que no sea ni el máximo ni el mínimo de ellas, es decir, el valor mediano, corresponde con la muestra **m** que indique el elemento estructurante, **no se hace nada**.

- Si el valor de las tres muestras que no sea ni el máximo ni el mínimo de ellas, es decir, el valor mediano, **no** corresponde con la muestra **m** que indique el elemento estructurante, se substituye esta por la **media aritmética** de las tres muestras.

Los márgenes dinámicos de la trama postfiltrada se recuperan multiplicando esta por una amplificación que responde a la ecuación:

$$\text{Ampli} = \frac{\max \{y_{\text{fil}}(m,i)\}}{\max \{y_{\text{med}}(m,i)\}} \quad (6.38)$$

Donde:  
**M**: nº muestras trama

**m**: número muestra de la trama,  $0 \leq m < M$

**i**: número de trama de señal

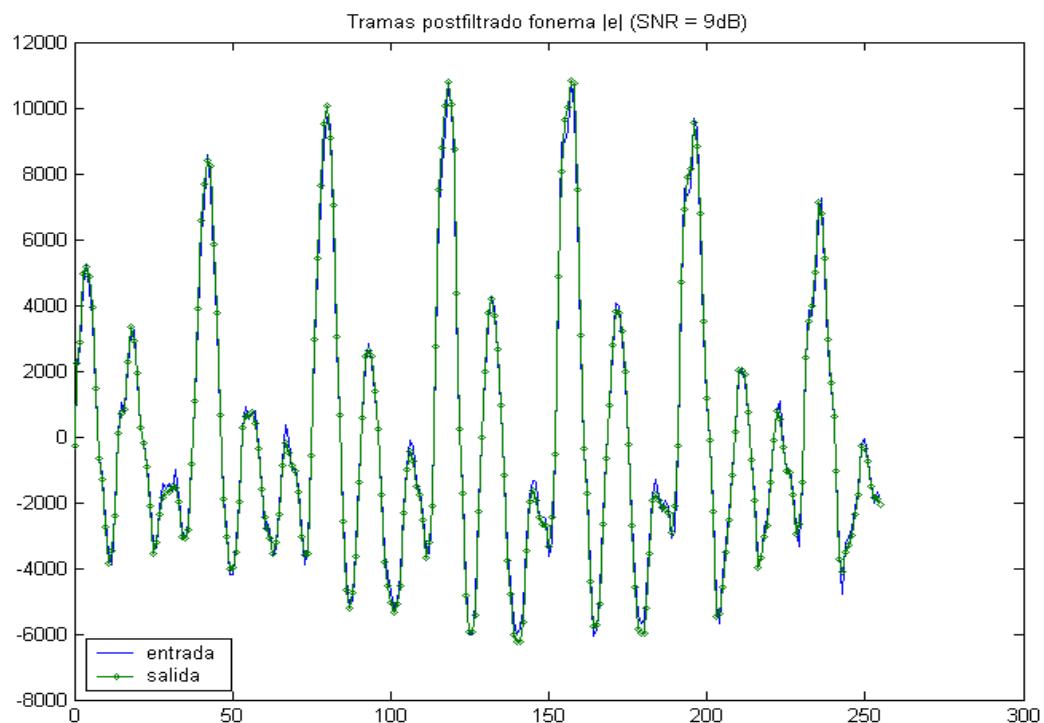


Fig. 6.40: Tramas en tiempo de entrada y salida del postfiltro para el fonema /e/ bajo condiciones de ruido blanco SNR = 9 dB.

### 6.6.- Reconstrucción de la señal de voz

Las tramas filtradas tienen un desplazamiento de  $z$  muestras, entonces si la longitud de trama son  $M$  muestras, tendremos un solapamiento de trama de  $\frac{M-z}{M} \cdot 100\%$ . Esto viene traducido en que la reconstrucción de la señal filtrada  $s_{est}(n)$  se realiza a partir del promediado de  $\frac{M}{z}$  tramas filtradas.

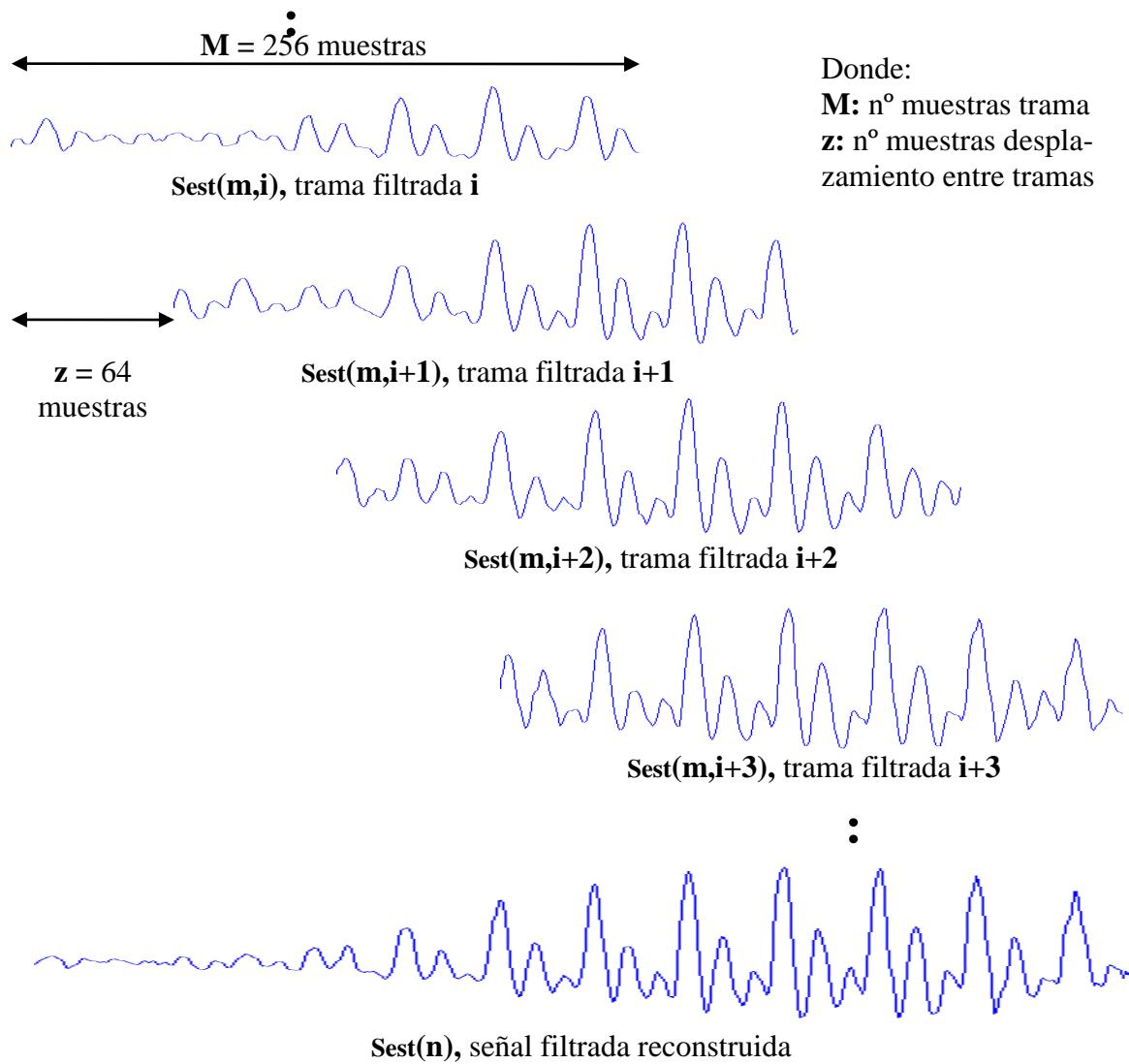


Fig. 6.41: Reconstrucción de parte del fonema /e/ a partir de diferentes tramas filtradas

En el caso de  $M=256$  y  $z=64$  muestras se deduce que la estimación de señal de voz  $s_{est}(n)$  se realiza de promediar 4 filtrados de trama. Este promediado es especialmente eficaz para la no eliminación de los sonidos sordos de principio y final de

palabra, ya que el simple hecho de que se detecte voz en 1 o 2 tramas de las 4 que se promediarán, es suficiente para poder escucharlo.

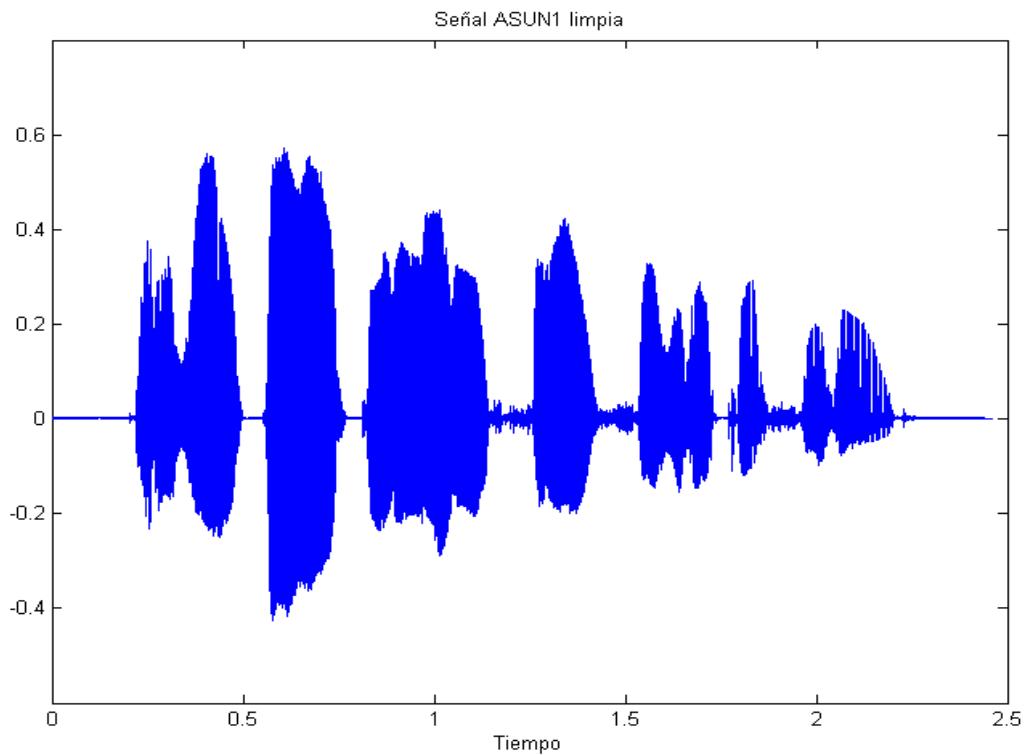


Fig.6.42: Señal ASUN1 limpia.

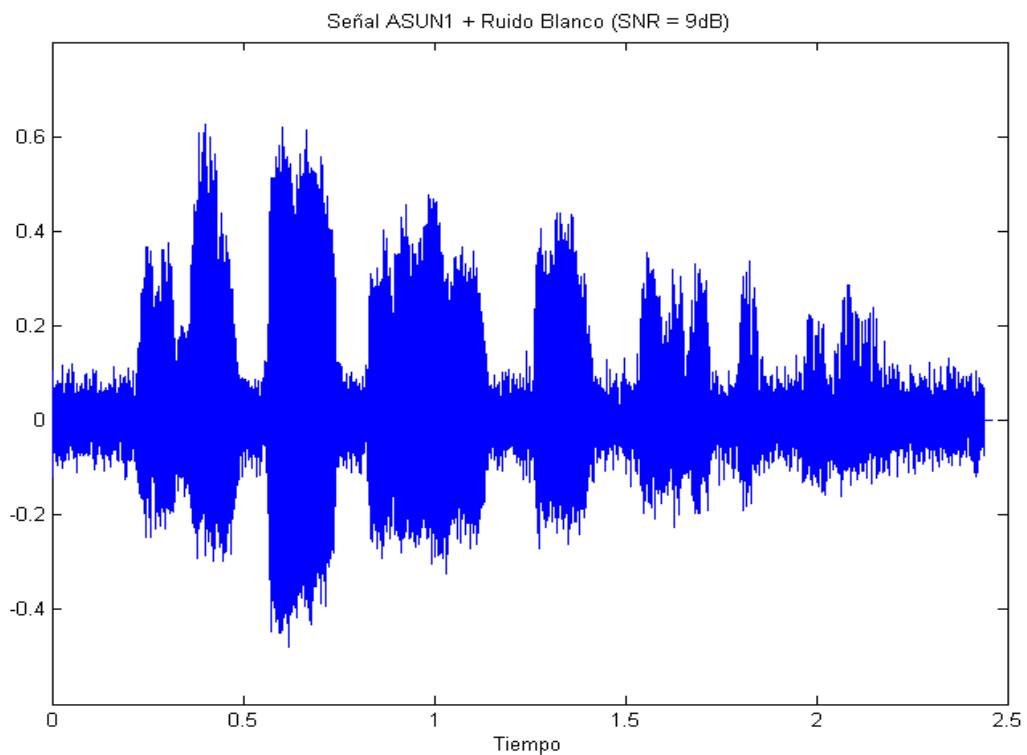


Fig.6.43: Señal ASUN1 con ruido blanco aditivo (SNR = 9 dB)

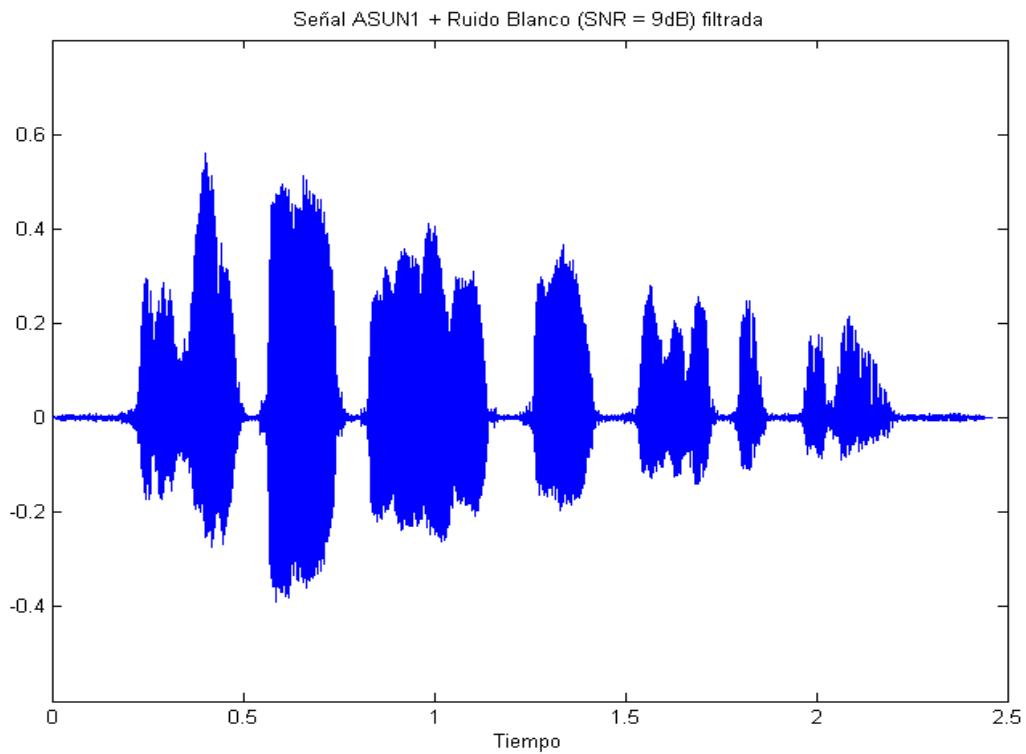


Fig.6.44: Señal ASUN1 + ruido blanco (SNR = 9dB)  
filtrada con RERCOM 3 etapas + peine.

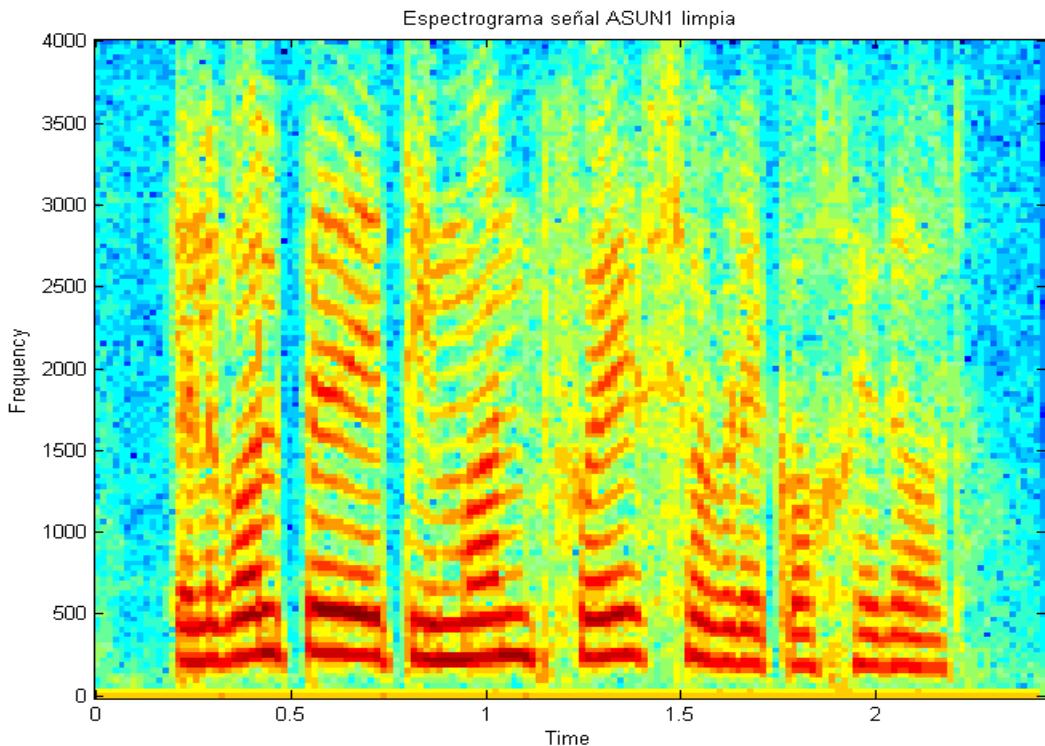


Fig.6.45: Espectrograma correspondiente a la señal ASUN1 limpia

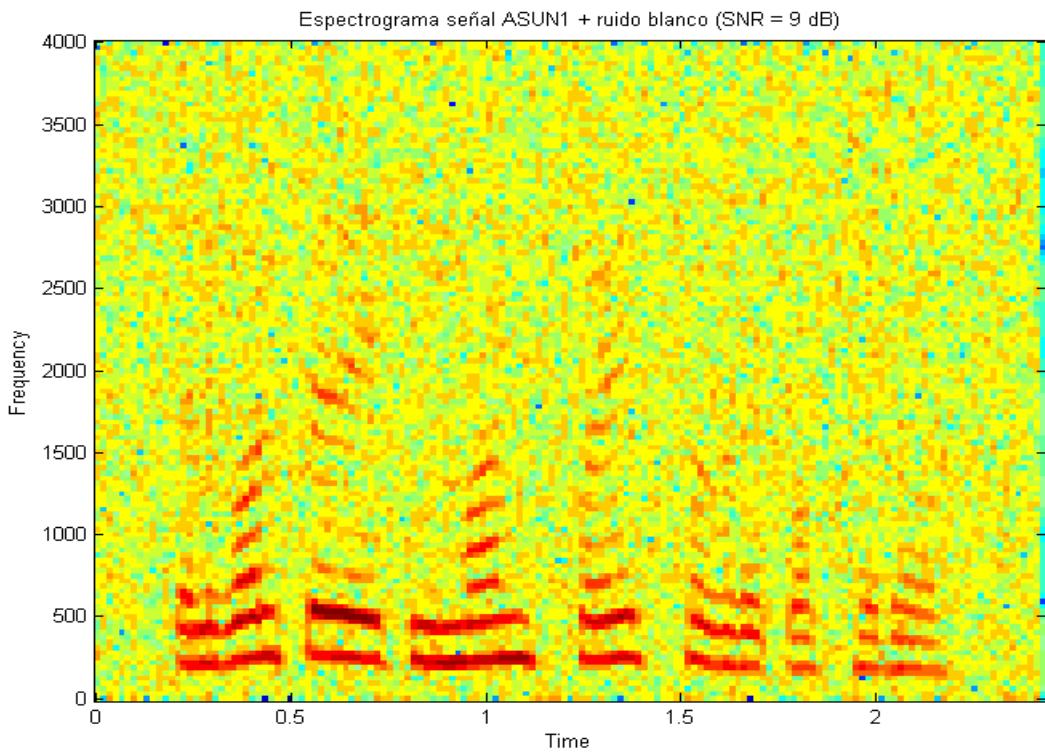


Fig.6.46: Espectrograma correspondiente a la señal ASUN1 con ruido blanco aditivo (SNR = 9 dB)

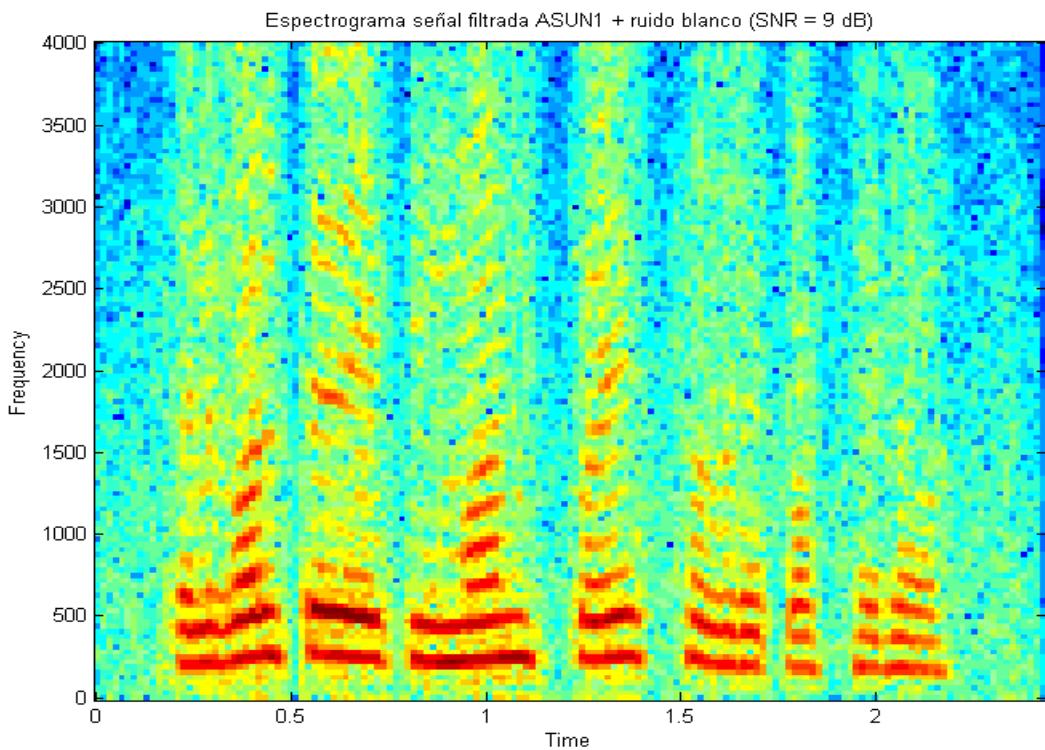


Fig.6.47: Espectrograma correspondiente a la señal ASUN1 + ruido blanco (SNR = 9dB) filtrada con RERCOM 3 etapas + peine.

## 6.7.- Modificaciones del algoritmo básico de filtrado

En este apartado explicaremos las diferentes mejoras que el simulador RERCOM permite realizar sobre el esquema básico de filtrado. Estas modificaciones avanzadas son la introducción de un esquema de filtrado iterativo y la introducción de las estadísticas de orden superior en el calculo de los coeficientes necesarios en el modelado AR de la voz.

### 6.7.1- Filtrado de Wiener iterativo

El filtrado de Wiener iterativo es un esquema que permite al sistema realizar una estimación del modelado AR de la voz a partir de una trama previamente filtrada un número arbitrario de veces, de esta manera se pueden obtener, teóricamente, una mejor estimación de la voz, ya que se utiliza una trama con un nivel de ruido reducido.

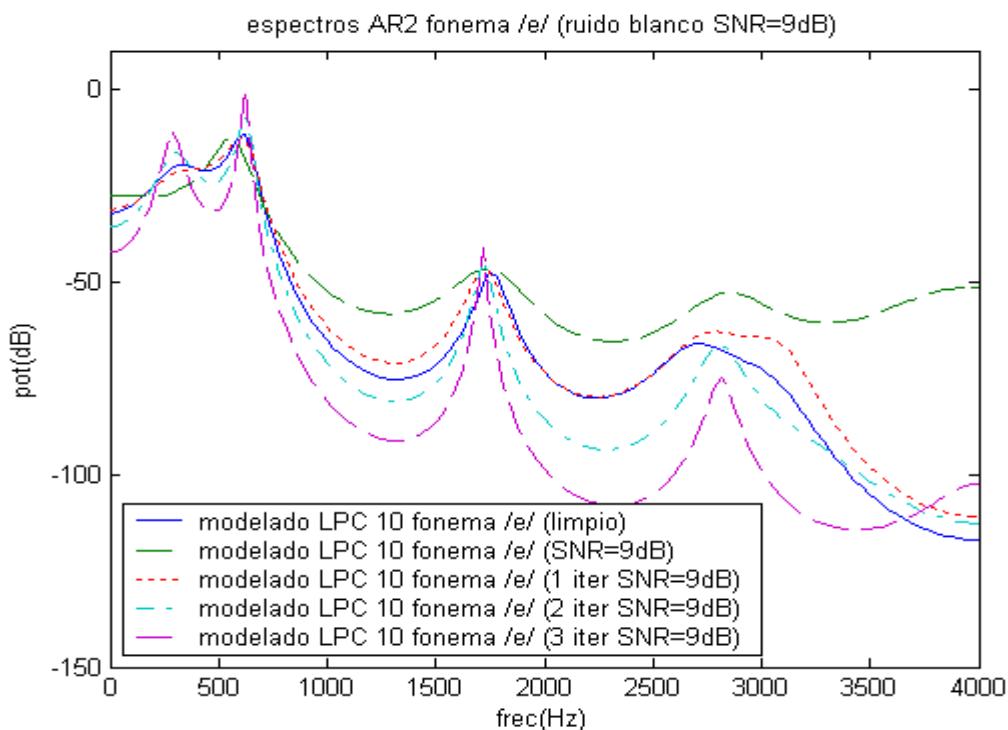


Fig. 6.48: Espectro modelado LPC10 AR2 obtenido después de varias iteraciones

Al introducir esta característica de filtrado, el esquema de la figura 6.22 se modifica según la figura 6.49. En esta figura, podemos observar que el realizar una nueva iteración de filtrado nos permite actualizar la estimación de modelado AR de voz a partir de una trama más “ limpia”, es decir, con un nivel menor de ruido, ya que la trama

ha sido previamente filtrada utilizando un filtro de Wiener calculado en la iteración anterior. Con esta técnica, a base de iterar deberíamos aproximarnos al filtro óptimo.

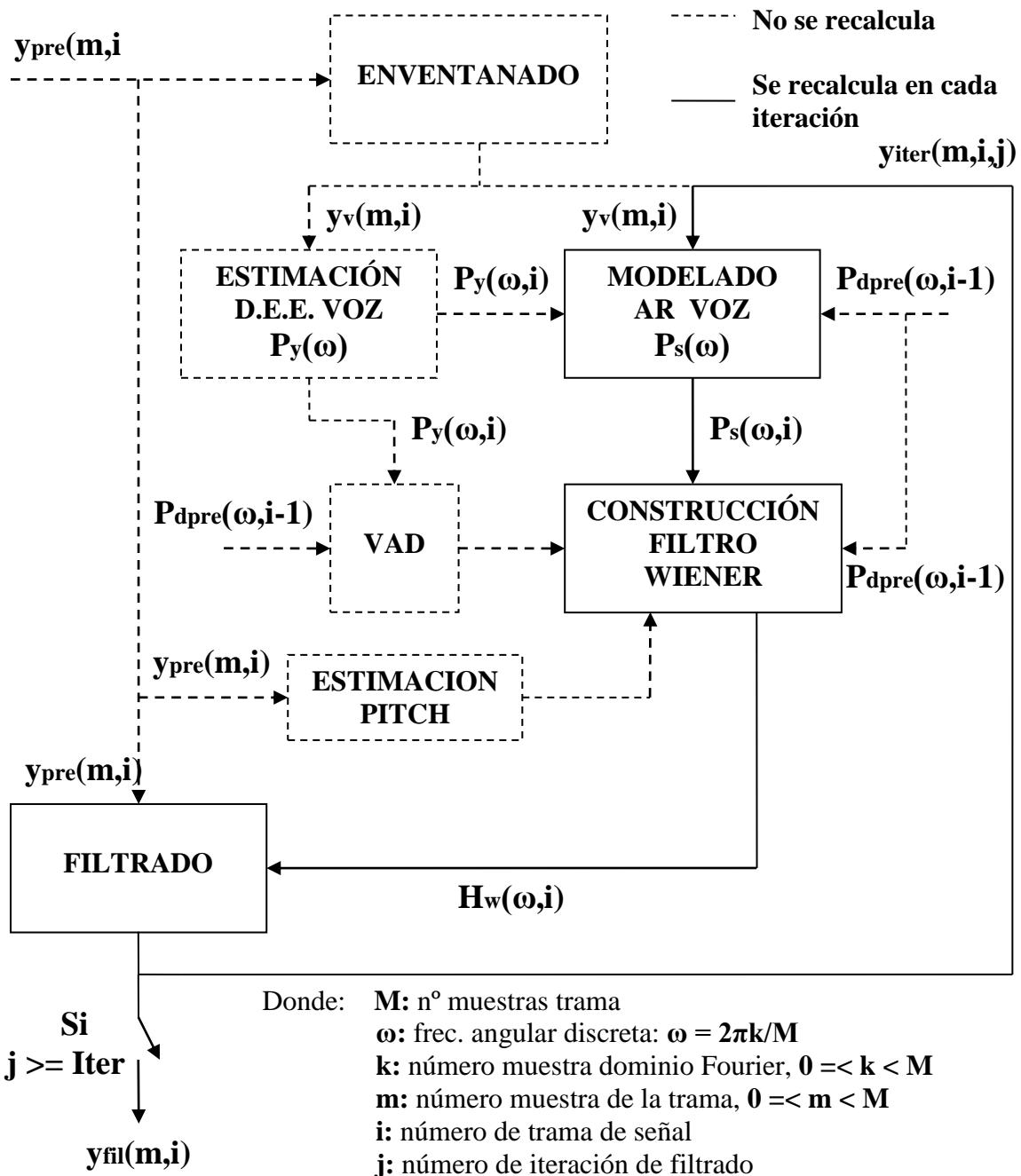


Fig. 6.49: Diagrama de bloques del filtro en una nueva iteración.

En el esquema observado en la figura 6.49 cabe destacar que en la primera iteración de filtrado es necesario realizar todos los cálculos necesarios para construir un primer filtro de Wiener, como se muestra en la figura 6.22. A partir de aquí, con cada nueva iteración sólo se actualizan las estimaciones de los bloques que aparecen en línea continua en la figura 6.48, en cambio, se conservan las estimaciones de los bloques en

línea discontinua. De esta manera el coste computacional de una nueva iteración es inferior. Además nos aseguramos que el valor del parámetro **g** de la ecuación (6.20) no acabe siendo nulo o negativo al aumentar en exceso el numero de iteraciones.

### 6.7.2.- Modelado AR utilizando estadísticas de orden superior

En el capítulo 3 vimos que las estadísticas de orden superior son más robustas que las estadísticas de segundo orden para suprimir ruido aditivo Gaussiano de una señal de voz. Es decir, que los cumulantes obtenidos a partir de una trama de voz libre de ruido o de la misma trama contaminada con ruido aditivo Gaussiano son teóricamente iguales si el número de muestras de la trama es suficientemente elevado. Esta característica propia de los cumulantes nos permite obtener unos coeficientes **ap**, necesarios para la estimación del modelado AR, mejores que los obtenidos utilizando correlaciones, bajo condiciones de ruido blanco.

Los diferentes ordenes de cumulantes nos permiten obtener una mejor estimación de modelado AR, pero con una penalización en coste computacional, ya que los cálculos necesarios para la obtención de los coeficientes **ap** son muy superiores. A continuación mostraremos en que consisten estos cálculos y como afectan en el algoritmo básico de filtrado.

El sistema RERCOM es capaz de simular un filtrado de Wiener utilizando cumulantes de orden 3 y 4 para la estimación de coeficiente **ap**. La diferencia básica entre estos dos tipos de estadísticas es que mientras que los cumulantes de orden 3 conservan la información de fase, los cumulantes de orden 4 la pierden al igual que las correlaciones.

El procedimiento de cálculo de los coeficientes **ap** utilizando cumulantes es el indicado por la figura 6.49, el cual substituye a los bloques **correlador** y **recursión Levinson-Durbin** de la figura 6.24.

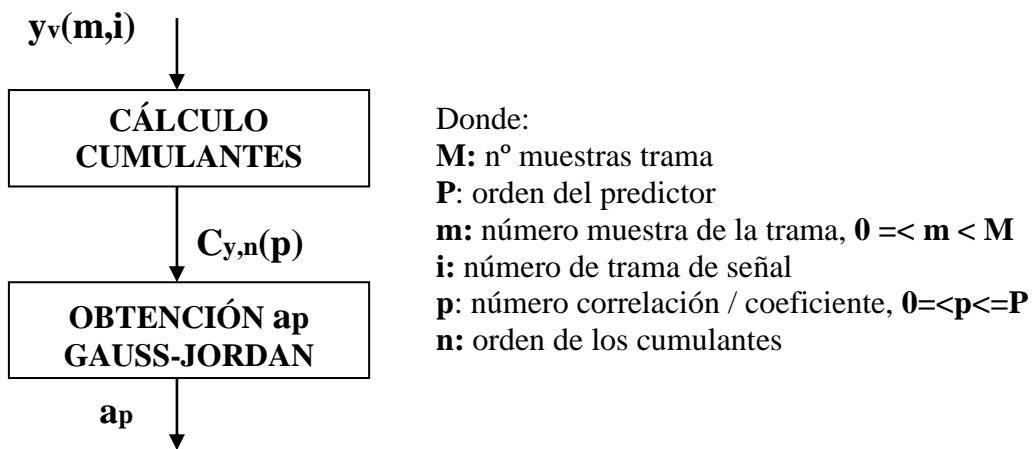


Fig. 6.50: Esquema para la estimación coeficientes **ap** a partir de cumulantes  
El bloque **Cálculo Cumulantes** realiza las siguientes operaciones [Sala-95]:

$$C_y^3(p_1, p_2) = \sum_{m=0}^{M-p-1} y_v(m, i) \cdot y_v(m + p_1, i) \cdot y_v(m + p_2, i) \quad \text{para orden 3} \quad (6.39)$$

Aplicando la propiedad de simetría:  $C_y^3(p_1, p_2) = C_y^3(p_2, p_1)$

$$C_y^4(p_1, p_2, 0) = \sum_{m=0}^{M-p-1} y_v(m, i) \cdot y_v(m + p_1, i) \cdot y_v(m + p_2, i) \cdot y_v(m, i) \quad \text{para orden 4} \quad (6.40)$$

Aplicando la propiedad de simetría:  $C_y^3(p_1, p_2, 0) = C_y^3(p_2, p_1, 0)$

Donde: **M:** n° muestras trama

**P:** orden del predictor

**m:** número muestra de la trama,  $0 \leq m < M$

**i:** número de trama de señal

**p:**  $\text{MAX}\{p_1, p_2\}$

con  $0 \leq p_1, p_2 \leq P$  variables desplazamiento de los cumulantes

El bloque **Obtención ap Gauss-Jordan** soluciona el sistema de ecuaciones de Yule-Walker para estadísticas de orden superior, obteniendo los coeficientes **ap**.

$$\underline{a}_p = (\underline{\underline{Q}}^T \cdot \underline{\underline{Q}})^{-1} \cdot \underline{\underline{Q}}^T \cdot \underline{b} \quad (6.41)$$

Donde: **n:** orden de los cumulantes

**P:** orden del predictor

$$d_3(p_1, p_2) = C_y^3(p_1, p_2) \quad \text{Para orden 3}$$

$$d_4(p_1, p_2) = C_y^4(p_1, p_2, 0) \quad \text{Para orden 4}$$

$$\underline{\underline{Q}} = \begin{pmatrix} d_n(0,0) & d_n(-1,0) & \cdots & d_n(1-P,0) \\ \cdots & \cdots & \cdots & \cdots \\ d_n(0,-P) & d_n(-1,-P) & \cdots & d_n(1-P,-P) \\ d_n(1,0) & d_n(0,0) & \cdots & d_n(2-P,0) \\ \cdots & \cdots & \cdots & \cdots \\ d_n(1,-P) & d_n(0,-P) & \cdots & d_n(2-P,-P) \\ \cdots & \cdots & \cdots & \cdots \\ d_n(p-1,0) & d_n(p-2,0) & \cdots & d_n(0,0) \\ \cdots & \cdots & \cdots & \cdots \\ d_n(P-1,-P) & d_n(P-2,-P) & \cdots & d_n(0,-P) \end{pmatrix} \quad (6.42)$$

$$\underline{\underline{a}}_p = [a_1, a_2, \dots, a_p]^T \quad (6.43)$$

$$\underline{\underline{b}} = [d_n(1,0), \dots, d_n(1,-p), d_n(2,0), \dots, d_n(2,-p), \dots, d_n(p,0), \dots, d_n(p,-p)]^T \quad (6.44)$$

Aplicando las propiedades de cumulantes:

$$\begin{aligned} d_n(p_1, p_2) &= d_n(p_2, p_1) = d_n(-p_1, p_2 - p_1) = \\ &= d_n(p_2 - p_1, -p_1) = d_n(p_1 - p_2, -p_2) = d_n(-p_2, p_1 - p_2) \end{aligned}$$

La inversión de la matriz  $(\underline{\underline{Q}}^T \cdot \underline{\underline{Q}})^{-1}$  se realiza mediante el método iterativo de Gauss – Jordan [Nume-92].

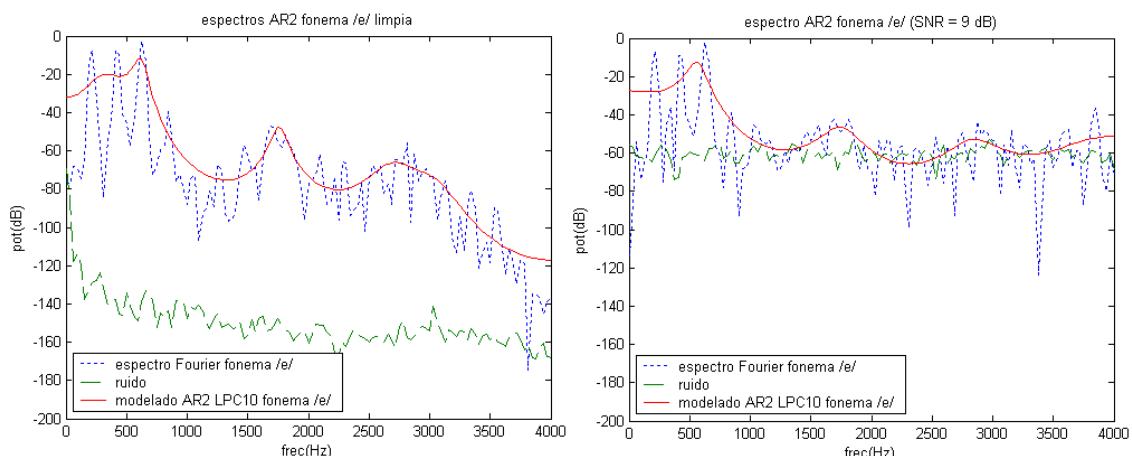


Fig. 6.51: Espectros modelado AR2 (orden 10) del fonema /e/ bajo condiciones de señal limpia y ruido blanco con SNR = 9dB.

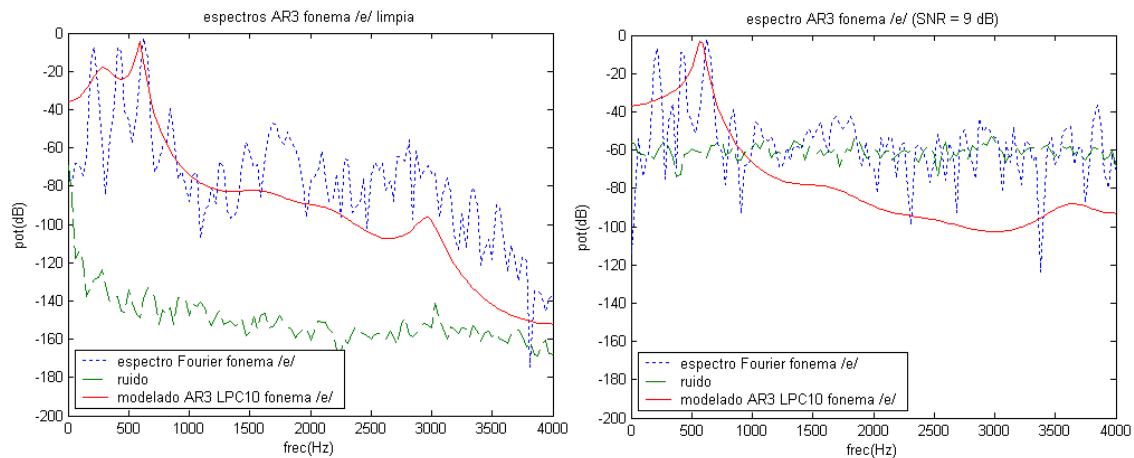


Fig. 6.52: Espectros modelado AR3 (orden 10) del fonema /e/ bajo condiciones de señal limpia y ruido blanco con SNR = 9dB.

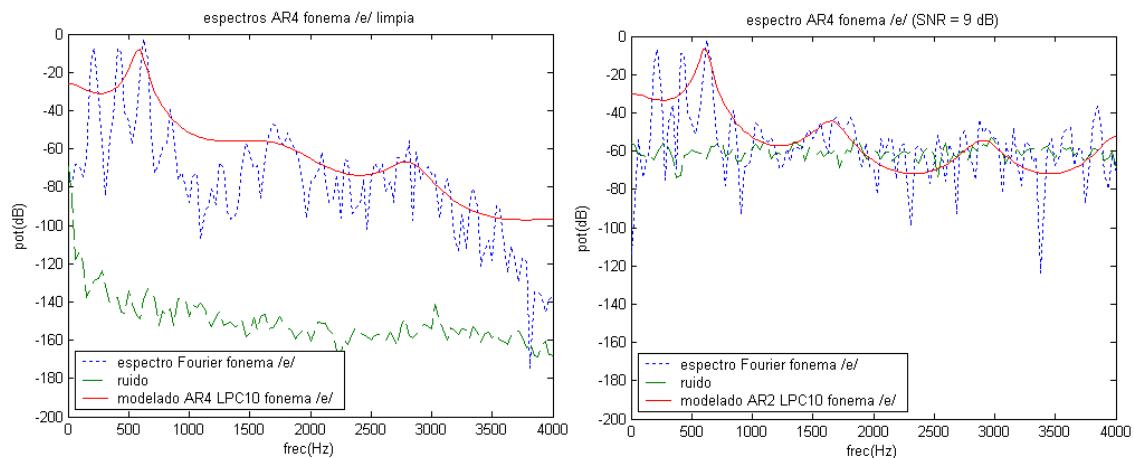


Fig. 6.53: Espectros modelado AR4 (orden 10) del fonema /e/ bajo condiciones de señal limpia y ruido blanco con SNR = 9dB.



## 7.- Resultados pruebas de simulación.

En este capítulo realizaremos una evaluación exhaustiva del sistema de reducción de ruido RERCOM. Esta evaluación consta de dos partes.

En una primera parte se realizará una optimización de los parámetros del sistema a partir de una serie de medidas objetivas basadas en SNR y distancia espectral, realizadas por un programa en C de elaboración propia que hemos bautizado como DISTCALC. Una vez obtenidos los parámetros óptimos, según las medidas obtenidas, procederemos a la comparación del sistema RERCOM con el estándar AdvFrontEnd de la ETSI.

En una segunda parte, evaluaremos el aumento de rendimiento que el sistema RERCOM introduce en un sistema de reconocimiento de voz, utilizando la base de datos SPEECHDATCAR del proyecto AURORA versión 008, comparando una vez más los resultados con el estándar AdvFrontEnd.

### 7.1.- Medidas de evaluación objetiva.

El programa DISTCALC permite realizar dos medidas de SNR y cuatro de distancia espectral, basadas en modelado LPC. Sus cinco parámetros de entrada permiten especificar:

- Longitud de trama en muestras;
  - o 128 muestras por defecto.
- Desplazamiento de trama en muestras;
  - o 128 muestras por defecto.
- Orden de predictor AR;
  - o 10 por defecto.
- Igualación de márgenes dinámicos y eliminación de componente de continua (DC) de la señal test y la señal de referencia;
  - o 0 -> desactivado.
  - o 1 -> activado (por defecto).
- Eliminación de silencios en las medidas de distancia espectral;
  - o 0 -> desactivado.
  - o 1 -> activado (por defecto).

### 7.1.1.- SNR global (SNR).

El cálculo de la SNR se realiza mediante la siguiente ecuación [Jove-93]:

$$\text{SNR} = 10 \cdot \log_{10} \left( \frac{\sum_{n=0}^{N-1} r^2(n)}{\sum_{n=0}^{N-1} (r(n) - t(n))^2} \right) \quad (7.1)$$

Donde:  
**N**: longitud total de señal.

**n**: número muestra de señal.

**r(n)**: señal de referencia; señal original limpia.

**t(n)**: señal de test; señal filtrada.

### 7.1.2.- SNR segmentada (SNRs).

La correlación de la SNR con la calidad subjetiva percibida es bastante pobre, ya que no refleja con exactitud los errores puntuales graves de filtrado. La SNR segmentada refleja mejor esta heterogeneidad de filtrado, siendo más sensible a estos errores.

La SNR segmentada responde a la ecuación [Hans-98]:

$$\text{SNRs} = \frac{10}{I} \sum_{i=0}^{I-1} \text{SNRtrama}(i) \quad (7.2)$$

$$\text{Donde: } \text{SNRtrama}(i) = \log_{10} \left( \frac{\sum_{n=N \cdot i}^{N \cdot i + N - 1} r^2(n)}{\sum_{n=N \cdot i}^{N \cdot i + N - 1} (r(n) - t(n))^2} \right) \quad (7.3)$$

Con:  $-10 \text{dB} \leq \text{SNRtrama}(i) \leq 35 \text{dB}$

Donde:  
**M**: número de muestras de cada trama de señal .

**N**: número de muestras de señal.

**I**: número de tramas de señal;  $I = N/M$ .

**i**: número de trama de señal.

**n**: número muestra de señal.

**r(n)**: señal de referencia; señal original limpia.

**t(n)**: señal de test; señal filtrada.

### 7.1.3.- Distancia Itakura (LLR).

La distancia LLR (Log-Likelihood Ratio) para una determinada trama **i**, responde a la siguiente ecuación [Hans-98]:

$$d_{LLR}(i) = \log \left( \frac{\underline{t} \cdot \underline{\underline{R}_r} \cdot \underline{t}^T}{\underline{r} \cdot \underline{\underline{R}_r} \cdot \underline{r}^T} \right) \quad (7.4)$$

Donde:  
**P**: orden predictor AR.

$\underline{t}$ : vector coeficientes AR trama señal test:  $\underline{t} = [1 \ a_1 \ .. \ a_P]$

$\underline{r}$ : vector coeficientes AR trama señal referencia:  $\underline{r} = [1 \ a_1 \ .. \ a_P]$

$\underline{\underline{R}_r}$ : matriz de correlaciones trama señal referencia:

$$\underline{\underline{R}_r} = \begin{bmatrix} R_r(0) & R_r(1) & R_r(2) & .. & R_r(P) \\ R_r(1) & R_r(0) & R_r(1) & .. & R_r(P-1) \\ R_r(2) & R_r(1) & R_r(0) & .. & R_r(P-2) \\ .. & .. & .. & .. & .. \\ R_r(P) & R_r(P-1) & R_r(P-2) & .. & R_r(0) \end{bmatrix}$$

El promediado de distancias LLR calculadas en las diferentes tramas se realizan mediante la ecuación:

$$d_{LLR}^{promed} = \frac{1}{I'} \sum_{i=0}^{I'-1} d_{LLR}(i) \quad (7.5)$$

Donde:  
**M**: número de muestras de cada trama de señal .

**N**: número de muestras de señal.

**I'**: número de tramas de señal a promediar;  $I' = (N/M) - \text{tramas\_silencio}$ .

**i**: número de trama de señal a promediar.

### 7.1.4.- Distancia Itakura-Saito (IS).

La distancia IS (Itakura-Saito) correspondiente a la trama **i**, responde a la siguiente ecuación [Hans-98]:

$$d_{IS}(i) = \left[ \frac{\sigma_r^2}{\sigma_t^2} \right] \cdot \left[ \frac{\underline{t} \cdot \underline{\underline{R}_r} \cdot \underline{t}^T}{\underline{r} \cdot \underline{\underline{R}_r} \cdot \underline{r}^T} \right] + \log \left( \frac{\sigma_t^2}{\sigma_r^2} \right) - 1 \quad (7.6)$$

Donde:  
**P**: orden predictor AR

$\underline{t}$ : vector coeficientes AR trama señal test:  $\underline{t} = [1 \ a_1 \ \dots \ a_P]$

$\underline{r}$ : vector coeficientes AR trama señal referencia:  $\underline{r} = [1 \ a_1 \ \dots \ a_P]$

$\underline{\underline{R}_r}$ : matriz de correlaciones trama señal referencia:

$$\underline{\underline{R}_r} = \begin{bmatrix} R_r(0) & R_r(1) & R_r(2) & \dots & R_r(P) \\ R_r(1) & R_r(0) & R_r(1) & \dots & R_r(P-1) \\ R_r(2) & R_r(1) & R_r(0) & \dots & R_r(P-2) \\ \dots & \dots & \dots & \dots & \dots \\ R_r(P) & R_r(P-1) & R_r(P-2) & \dots & R_r(0) \end{bmatrix}$$

$\sigma_t^2$ : valor de ganancia AR trama señal de test:

$$\sigma_t^2 = [R_t(0) \ R_t(1) \ \dots \ R_t(P)] \cdot \underline{t}^T$$

$\sigma_r^2$ : valor de ganancia AR trama señal de referencia:

$$\sigma_r^2 = [R_r(0) \ R_r(1) \ \dots \ R_r(P)] \cdot \underline{r}^T$$

El promediado de distancias IS calculadas en las diferentes tramas se realizan mediante la ecuación:

$$d_{IS}^{promed} = \frac{1}{I'} \sum_{i=0}^{I'-1} d_{IS}(i) \quad (7.7)$$

Donde:  
**M**: número de muestras de cada trama de señal .

**N**: número de muestras de señal.

**I'**: número de tramas de señal a promediar; **I'=(N/M) – tramas\_silencio**.

**i**: número de trama de señal a promediar.

En esta ecuación se observa que el valor de la distancia depende de los valores de ganancia de AR de la señal de test y señal de referencia, así, al igual que las medidas de SNR y SNRs, será necesario igualar los márgenes dinámicos y el valor continua de la señal test a los de la señal de referencia.

### 7.1.5.- Distancia Cepstrum (Ceps).

La distancia cepstrum para la trama **i** responde a la ecuación [Jove-93]:

$$d_{ceps}(i) = \frac{1}{K} \cdot \sum_{k=0}^{K-1} \left| \log \left( \frac{\sigma_t^2}{|A_t(\omega)|^2} \right) - \log \left( \frac{\sigma_r^2}{|A_r(\omega)|^2} \right) \right| \quad (7.8)$$

$$\text{Donde: } A(\omega) = 1 + \sum_{p=1}^P a_p e^{-j\omega p} \quad (7.9)$$

Donde:  
**K**: número muestras por trama en dominio Fourier.

**k**: número muestra dominio Fourier.

**P**: orden predictor AR.

**ω**: frec. angular discreta:  $\omega = 2\pi k/M$

$\sigma_t^2$ : valor de ganancia AR trama señal de test:

$$\sigma_t^2 = [R_t(0) \ R_t(1) \ \dots \ R_t(P)] \cdot \underline{t}^T$$

$\sigma_r^2$ : valor de ganancia AR trama señal de referencia:

$$\sigma_r^2 = [R_r(0) \ R_r(1) \ \dots \ R_r(P)] \cdot \underline{r}^T$$

El promediado de distancias cepstrum calculadas en las diferentes tramas se realizan mediante la ecuación:

$$d_{\text{ceps}}^{\text{promed}} = \frac{1}{I'} \sum_{i=0}^{I'-1} d_{\text{ceps}}(i) \quad (7.10)$$

Donde:  
**M**: número de muestras de cada trama de señal .

**N**: número de muestras de señal.

**I**: número de tramas de señal a promediar;  $I'=(N/M) - \text{tramas\_silencio}$ .

**i**: número de trama de señal a promediar.

En esta ecuación se observa que el valor de la distancia depende de los valores de ganancia de AR de la señal de test y señal de referencia, así, al igual que las medidas de SNR, SNRs y IS será necesario igualar los márgenes dinámicos y el valor continua de la señal test a los de la señal de referencia.

#### 7.1.7.- Distancia Log Area Ratio (LAR).

La medida LAR se basa también en la diferencia entre los coeficientes AR2 de la señal de test y la señal de referencia, pero en este caso los coeficientes utilizados son los P coeficientes de reflexión de la señal de referencia y la señal de test. La distancia LAR de una cierta trama responde a la ecuación [Hans-98]:

$$d_{\text{LAR}}(i) = \sqrt{\frac{1}{P} \cdot \sum_{p=0}^{P-1} \left( \log \left( \frac{1+t^r(p)}{1-t^r(p)} \right) - \log \left( \frac{1+r^r(p)}{1-r^r(p)} \right) \right)^2} \quad (7.11)$$

Donde:**P**: orden predictor AR.

$\underline{t}$ : vector coeficientes reflexión AR trama señal test:

$$\underline{t}^r = [rx_0 \quad rx_1 \quad \dots \quad rx_{P-1}]$$

$\underline{r}$ : vector coeficientes reflexión AR trama señal referencia.

$rx_i$ : coeficiente reflexión obtenido de la recursión Levinson-Durbin:

$$\underline{r}^r = [rx_0 \quad rx_1 \quad \dots \quad rx_{P-1}]$$

El promediado de distancias LAR calculadas en las diferentes tramas se realizan mediante la ecuación:

$$d_{LAR}^{promed} = \frac{1}{I'} \sum_{i=0}^{I'-1} d_{LAR}(i) \quad (7.12)$$

Donde:**M**: número de muestras de cada trama de señal.

**N**: número de muestras de señal.

**I**: número de tramas de señal a promediar;  $I'=(N/M) - \text{tramas\_silencio}$ .

**i**: número de trama de señal a promediar.

### 7.1.7.- Cálculo de la amplificación aplicada a la señal de test.

Como se ha comentado anteriormente, para evitar falsear las medidas de SNR's y distancias espectrales, es necesario igualar los márgenes dinámicos de las señales de test y de referencia. La medida AMPLI proporciona la información en dB de la amplificación uniforme que ha sido necesaria aplicar a toda la señal de test  $t(n)$ , para igualar su margen dinámico a la señal de referencia  $r(n)$ .

$$\text{Ampli} = 10 \cdot \log_{10} \left( \frac{\text{Max}_r}{\text{Max}_t} \right) \quad (7.13)$$

Donde:**Maxr**: Valor máximo señal referencia.

**Maxt**: Valor máximo señal test.

## 7.2.-Elección óptima de parámetros

En este punto realizaremos un estudio de los distintos parámetros de entrada a nuestro sistema, con el objetivo de obtener los valores óptimos de estos parámetros que nos permiten obtener las mejores estadísticas. Para ello, hemos obtenido tablas comparativas para distintos ficheros con diferentes niveles de ruido blanco. En cada tabla aparecen los valores de la señal limpia (REFEREN), los valores de la señal antes de filtrar (ORIGINAL) y los valores de la señal original después de filtrar en función del parámetro que estamos analizando en cada caso. En las siguientes tablas hemos resaltado los mejores resultados de cada parámetro.

### 7.2.1.-Elección de la longitud de trama

En esta serie de experimentos intentaremos descubrir cual es la longitud de trama óptima para el algoritmo de filtrado. Para ello utilizaremos una sola etapa de filtrado basado en filtro de wiener con modelado AR. Consideraremos un solapamiento de trama del 0%, es decir, el parámetro desplazamiento de trama es igual a la longitud de trama.

Parámetros utilizados:

*Longitud de trama: \*\*\* parámetro a optimizar \*\*\**

*Desplazamiento de trama: \*\*\* igual a longitud de trama \*\*\**

*Multiplicador de muestras de trama: 1 (desactivado)*

*Tipo de ventana (estimación espectros): 2 (Hanning)*

*Tipo de prefiltrado: 0 (desactivado)*

*Parámetro beta del prefiltrado: 1.00*

*Parámetro delta del prefiltrado: 1.00*

*Atenuación máxima de prefiltrado: 0.50*

*Tipo de filtro: 1 (wiener basado en modelado AR)*

*Orden del predictor sonoras(modelado): 10*

*Orden del predictor sordas(modelado): 10*

*Orden de los cumulantes (modelado): 1 (AR2 – Levinson Durbin)*

*Sobreumbral pitch: 0.00 (desactivado)*

*Numero de iteraciones del filtro: 1*

*Factor intertrama (ak): 1.00*

*Factor iteración trama previa: 1*

*Parámetro beta del filtro: 1.00*

*Parámetro delta del filtro: 1.00*

*Atenuación máxima de filtrado: 0.01*

*Nivel de ruido (silencio) en filtrado: 2500.00*

*Tipo de postfiltrado: 0 (desactivado)*

*Nivel activación postfiltrado: 0.50*

*Orden del filtro de mediana: 1*

*Parámetro sobreestimación ruido (VAD): 0.00 (desactivado)*

*Numero de tramas de ruido iniciales: 10*

*Parámetro promediado espectro ruido: 0.10*

*Reestimar ruido en tramas silencio: 0 (desactivado)*

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMPLI(dB)
REFEREN	154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL	18,079	8,169	1,018	1,483	2,118	2,407	0,000
64	1,716	-0,069	1,146	3,511	1,574	2,476	-2,391
128	4,266	2,119	<b>0,752</b>	<b>2,119</b>	<b>1,456</b>	<b>2,138</b>	-2,405
256	<b>7,957</b>	<b>3,740</b>	0,941	5,368	2,374	2,173	<b>0,231</b>
512	2,574	0,634	1,029	7,760	3,737	2,373	1,711

Tabla 7.1: Resultados utilizando el fichero ASUN1 con ruido blanco (SNR=18dB).

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMPLI(dB)
REFEREN	154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL	9,061	1,522	1,745	2,718	3,500	3,192	0,000
64	1,343	-0,741	1,872	4,889	2,042	3,203	-2,390
128	4,005	1,273	<b>1,337</b>	<b>2,848</b>	<b>1,917</b>	2,953	-2,404
256	<b>7,302</b>	<b>2,371</b>	1,512	5,386	2,255	<b>2,834</b>	<b>0,113</b>
512	2,480	-0,052	1,667	7,421	3,093	3,035	1,663

Tabla 7.2: Resultados utilizando el fichero ASUN1 con ruido blanco (SNR=9dB).

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMPLI(dB)
REFEREN	154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL	0,061	-3,918	2,523	4,266	5,146	3,787	0,000
64	0,812	-1,640	2,612	6,145	2,599	3,843	-2,390
128	3,007	-0,327	<b>2,093</b>	<b>3,888</b>	<b>2,495</b>	3,606	-2,404
256	<b>5,208</b>	<b>0,468</b>	2,266	5,802	2,595	<b>3,499</b>	<b>0,284</b>
512	1,904	-0,968	2,387	7,501	2,950	3,589	1,451

Tabla 7.3: Resultados utilizando el fichero ASUN1 con ruido blanco (SNR=0dB).

	SNR(dB)	SNRs (dB)	LLR	IS	CEPS	LAR	AMPLI(dB)
REFEREN	154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL	18,048	9,165	2,220	3,075	3,818	3,961	0,000
64	2,447	1,382	2,312	3,929	2,586	3,807	-2,019
128	15,578	9,464	1,940	<b>2,280</b>	2,650	3,593	<b>-0,051</b>
256	<b>17,807</b>	<b>10,980</b>	<b>1,811</b>	2,290	<b>2,512</b>	<b>3,444</b>	-0,155
512	1,954	0,727	2,113	6,747	2,835	3,646	2,786

Tabla 7.4: Resultados utilizando el fichero ESCA con ruido blanco (SNR=18dB).

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMPLI(dB)
REFEREN	154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL	9,031	2,262	3,195	4,548	5,390	4,659	0,000
64	1,781	0,080	3,302	5,738	3,287	4,482	-2,021
128	11,879	5,504	2,824	<b>3,710</b>	3,588	4,292	<b>-0,033</b>
256	<b>13,171</b>	<b>6,927</b>	<b>2,615</b>	3,744	3,305	<b>4,135</b>	-0,084
512	2,183	0,644	2,886	6,599	<b>3,143</b>	4,204	2,646

Tabla 7.5: Resultados utilizando el fichero ESCA con ruido blanco (SNR=9dB).

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMPLI(dB)
REFEREN	154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL	0,023	-3,563	4,120	6,179	7,146	5,148	0,000
64	1,098	-1,081	4,304	6,898	4,068	5,019	-2,021
128	6,237	1,117	3,616	<b>4,881</b>	4,563	4,872	<b>0,103</b>
256	<b>7,139</b>	<b>2,532</b>	<b>3,369</b>	4,995	<b>4,005</b>	4,701	0,199
512	2,538	0,615	3,599	6,377	3,598	<b>4,669</b>	2,215

Tabla 7.6: Resultados utilizando el fichero ESCA con ruido blanco (SNR=0dB).

Como era de esperar, con un solapamiento de trama del 0%, es decir, sin solapamiento, se obtienen unos rendimientos de reducción de ruido muy pobres, pero nos muestra de manera muy clara cual es la longitud de trama óptima a utilizar en el filtrado.

En las tablas de resultados anteriores observamos que una longitud de trama de 512 muestras resulta excesiva por motivos de no estacionariedad de la señal, ya que 512 muestras corresponden a 64 ms de señal.

Por otro lado una longitud de trama de 128 o 64 muestras (16 y 8 ms) cumplen la condición de estacionariedad, pero permiten obtener espectros de muy poca resolución para una frecuencia de muestreo de 8 Khz (62,5 Hz por muestra para 128 muestras y 125 Hz por muestra para 64 muestras).

Como era de suponer el parámetro de **longitud de trama óptimo es 256 muestras**, que corresponden a 32 ms (a frecuencia de muestreo de 8Khz), tiempo máximo en el cual la estadística de la voz se mantiene estacionaria, con una resolución espectral de 31,125 Hz por muestra.

### 7.2.2.-Elección del desplazamiento de trama

En esta serie de experimentos intentaremos descubrir cual es el desplazamiento de trama óptimo para el algoritmo de filtrado. Consideraremos los mismos parámetros del

apartado 7.2.1., a excepción del parámetro longitud de trama que lo mantendremos fijado a 256 muestras.

Parámetros utilizados:

*Longitud de trama: 256*

*Desplazamiento de trama: \*\*\* parámetro a optimizar \*\*\**

*Multiplicador de muestras de trama: 1 (desactivado)*

*Tipo de ventana (estimación espectros): 2 (Hanning)*

*Tipo de prefiltrado: 0 (desactivado)*

*Parámetro beta del prefiltrado: 1.00*

*Parámetro delta del prefiltrado: 1.00*

*Atenuación máxima de prefiltrado: 0.50*

*Tipo de filtro: 1 (wiener basado en modelado AR)*

*Orden del predictor sonoras(modelado): 10*

*Orden del predictor sordas(modelado): 10*

*Orden de los cumulantes (modelado): 1 (AR2 – Levinson Durbin)*

*Sobreumbral pitch: 0.00 (desactivado)*

*Numero de iteraciones del filtro: 1*

*Factor intertrama (ak): 1.00*

*Factor iteración trama previa: 1*

*Parámetro beta del filtro: 1.00*

*Parámetro delta del filtro: 1.00*

*Atenuación máxima de filtrado: 0.01*

*Nivel de ruido (silencio) en filtrado: 2500.00*

*Tipo de postfiltro: 0 (desactivado)*

*Nivel activación postfiltro: 0.50*

*Orden del filtro de mediana: 1*

*Parámetro sobreestimación ruido (VAD): 0.00 (desactivado)*

*Numero de tramas de ruido iniciales: 10*

*Parámetro promediado espectro ruido: 0.10*

*Reestimar ruido en tramas silencio: 0 (desactivado)*

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMPLI(dB)
REFEREN	154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL	18,079	8,169	1,018	1,483	2,118	2,407	0,000
256 (0%)	7,957	3,740	0,941	5,368	2,374	2,173	0,231
128 (50%)	21,127	10,419	<b>0,668</b>	<b>0,956</b>	<b>1,302</b>	<b>1,879</b>	<b>-0,021</b>
64 (75%)	21,339	10,568	0,695	0,968	1,335	1,926	-0,052
32 (87,5%)	<b>21,540</b>	<b>10,743</b>	0,717	1,042	1,346	1,992	-0,028

Tabla 7.7: Resultados utilizando el fichero ASUN1 con ruido blanco (SNR=18dB).

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMPLI(dB)
REFEREN	154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL	9,061	1,522	1,745	2,718	3,500	3,192	0,000
256 (0%)	7,302	2,371	1,512	5,386	2,255	2,834	<b>0,113</b>
128 (50%)	14,053	5,347	<b>1,282</b>	2,060	<b>2,020</b>	<b>2,699</b>	-0,195
64 (75%)	14,095	5,379	1,304	<b>2,035</b>	2,057	2,743	-0,186
32 (87,5%)	<b>14,121</b>	<b>5,496</b>	1,317	2,227	2,058	2,797	-0,168

Tabla 7.8: Resultados utilizando el fichero ASUN1 con ruido blanco (SNR=9dB).

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMPLI(dB)
REFEREN	154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL	0,061	-3,918	2,523	4,266	5,146	3,787	0,000
256 (0%)	5,208	0,468	2,266	5,802	<b>2,595</b>	3,499	0,284
128 (50%)	<b>7,362</b>	0,932	<b>1,997</b>	2,875	2,847	<b>3,381</b>	<b>0,185</b>
64 (75%)	7,167	0,773	2,045	<b>2,857</b>	2,924	3,427	0,314
32 (87,5%)	7,186	<b>1,054</b>	2,006	3,092	2,820	3,474	0,277

Tabla 7.9: Resultados utilizando el fichero ASUN1 con ruido blanco (SNR=0dB).

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMPLI(dB)
REFEREN	154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL	18,048	9,165	2,220	3,075	3,818	3,961	0,000
256 (0%)	17,807	10,980	<b>1,811</b>	2,290	<b>2,512</b>	<b>3,444</b>	-0,155
128 (50%)	18,935	11,532	1,824	<b>2,099</b>	2,601	3,453	-0,087
64 (75%)	19,273	11,826	1,849	2,126	2,649	3,485	-0,052
32 (87,5%)	<b>19,513</b>	<b>12,130</b>	1,840	2,126	2,636	3,532	<b>-0,023</b>

Tabla 7.10: Resultados utilizando el fichero ESCA con ruido blanco (SNR=18dB).

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMPLI(dB)
REFEREN	154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL	9,031	2,262	3,195	4,548	5,390	4,659	0,000
256 (0%)	13,171	6,927	<b>2,615</b>	3,744	<b>3,305</b>	<b>4,135</b>	-0,084
128 (50%)	13,621	6,729	2,637	<b>3,319</b>	3,496	4,145	-0,045
64 (75%)	13,724	6,728	2,675	3,327	3,563	4,186	<b>-0,005</b>
32 (87,5%)	<b>13,768</b>	<b>6,960</b>	2,627	3,373	3,515	4,238	0,028

Tabla 7.11: Resultados utilizando el fichero ESCA con ruido blanco (SNR=9dB).

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMPLI(dB)
REFEREN	154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL	0,023	-3,563	4,120	6,179	7,146	5,148	0,000
256 (0%)	7,139	2,532	3,369	4,995	<b>4,005</b>	<b>4,701</b>	0,199
128 (50%)	<b>7,179</b>	1,960	3,390	<b>4,318</b>	4,404	4,702	<b>0,140</b>
64 (75%)	7,123	1,750	3,464	4,414	4,502	4,752	0,174
32 (87,5%)	7,084	<b>2,011</b>	<b>3,348</b>	4,394	4,368	4,806	0,260

Tabla 7.12: Resultados utilizando el fichero ESCA con ruido blanco (SNR=0dB).

En las tablas anteriores observamos que los mejores valores de SNR y SNRs se obtienen con un desplazamiento de trama igual a 32 muestras correspondiente a un solapamiento del 87,5%, en cambio los mejores resultados de distancias espectrales se obtienen con un desplazamiento de trama de 128 muestras, correspondiente a un solapamiento entre tramas del 50%. Hemos de tener en cuenta que la elección de este parámetro influye mucho en el tiempo de procesado, ya que el número de tramas a procesar aumenta, así, con desplazamiento de trama de 128 muestras (solapamiento del 50%) se han de procesar la mitad de tramas que para un desplazamiento de trama de 64 muestras (solapamiento del 75%).

En este caso, descartaremos los parámetros de desplazamiento de trama; 256 muestras por motivos de continuidad de la forma de onda de la señal de voz y 32 muestras por motivos de aumento excesivo de tiempo necesario de procesado a costa de un inapreciable incremento de SNR y SNRs.

Consideraremos entonces como valor óptimo un **desplazamiento de trama 128 muestras correspondiente a un solapamiento del 50%**.

### 7.2.3.- Activación de filtros

En esta prueba mantendremos los parámetros del apartado 7.2.2 utilizando un desplazamiento de trama de 128 muestras, es decir, un solapamiento entre tramas de 50%.

En este experimento activaremos un filtro paso banda con atenuación variable en función de la potencia de ruido de la señal. Este filtro paso banda atenúa las frecuencias donde la energía de la señal de voz es muy baja, correspondiente a la banda frecuencial entre 0 y 50 Hz y la banda frecuencial entre 3500 y 4000 Hz, para una frecuencia de muestreo de 8Khz.

Parámetros utilizados:

*Longitud de trama: 256*

*Desplazamiento de trama: 128*

*Multiplicador de muestras de trama: 1 (desactivado)*

*Tipo de ventana (estimación espectros): 2 (Hanning)*

*Tipo de prefiltro: 0 (desactivado)*

*Parámetro beta del prefiltro: 1.00*

*Parámetro delta del prefiltro:* 1.00  
*Atenuación máxima de prefiltrado:* 0.50  
*Tipo de filtro:* 1 (wiener basado en modelado AR)  
*Orden del predictor sonoras(modelado):* 10  
*Orden del predictor sordas(modelado):* 10  
*Orden de los cumulantes (modelado):* 1 (AR2 – Levinson Durbin)  
*Sobreumbral pitch:* 0.00 (desactivado)  
*Numero de iteraciones del filtro:* 1  
*Factor intertrama (ak):* 1.00  
*Factor iteración trama previa:* 1  
*Parámetro beta del filtro:* 1.00  
*Parámetro delta del filtro:* 1.00  
*Atenuación máxima de filtrado:* 0.01  
*Nivel de ruido (silencio) en filtrado:* 2500.00  
*Tipo de postfiltro:* 0 (desactivado)  
*Nivel activación postfiltro:* 0.50  
*Orden del filtro de mediana:* 1  
*Parámetro sobreestimación ruido (VAD):* 0.00 (desactivado)  
*Numero de tramas de ruido iniciales:* 10  
*Parámetro promediado espectro ruido:* 0.10  
*Reestimar ruido en tramas silencio:* 0 (desactivado)

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL		18,079	8,169	1,018	1,483	2,118	2,407	0,000
SIN FILT		21,127	10,419	0,668	0,956	1,302	1,879	-0,021
CON FILT		<b>21,285</b>	<b>10,555</b>	<b>0,495</b>	<b>0,795</b>	<b>1,091</b>	<b>1,435</b>	<b>-0,019</b>

Tabla 7.13: Resultados utilizando el fichero ASUN1 con ruido blanco (SNR=18dB).

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL		9,061	1,522	1,745	2,718	3,500	3,192	0,000
SIN FILT		14,053	5,347	1,282	2,060	2,020	2,699	-0,195
CON FILT		<b>14,348</b>	<b>5,565</b>	<b>0,947</b>	<b>1,722</b>	<b>1,641</b>	<b>2,118</b>	<b>-0,184</b>

Tabla 7.14: Resultados utilizando el fichero ASUN1 con ruido blanco (SNR=9dB).

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL		0,061	-3,918	2,523	4,266	5,146	3,787	0,000
SIN FILT		7,362	0,932	1,997	2,875	2,847	3,381	<b>0,185</b>
CON FILT		<b>7,628</b>	<b>1,116</b>	<b>1,563</b>	<b>2,420</b>	<b>2,400</b>	<b>2,832</b>	0,300

Tabla 7.15: Resultados utilizando el fichero ASUN1 con ruido blanco (SNR=0dB).

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL		18,048	9,165	2,220	3,075	3,818	3,961	0,000
SIN FILT		<b>18,935</b>	11,532	1,824	2,099	2,601	3,453	<b>-0,087</b>
CON FILT		18,932	<b>11,579</b>	<b>1,631</b>	<b>1,902</b>	<b>2,396</b>	<b>2,989</b>	-0,115

Tabla 7.16: Resultados utilizando el fichero ESCA con ruido blanco (SNR=18dB).

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL		9,031	2,262	3,195	4,548	5,390	4,659	0,000
SIN FILT		13,621	6,729	2,637	3,319	3,496	4,145	<b>-0,045</b>
CON FILT		<b>13,856</b>	<b>6,992</b>	<b>2,285</b>	<b>2,969</b>	<b>3,107</b>	<b>3,507</b>	-0,126

Tabla 7.17: Resultados utilizando el fichero ESCA con ruido blanco (SNR=9dB).

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL		0,023	-3,563	4,120	6,179	7,146	5,148	0,000
SIN FILT		7,179	1,960	3,390	4,318	4,404	4,702	<b>0,140</b>
CON FILT		<b>7,565</b>	<b>2,245</b>	<b>2,956</b>	<b>3,872</b>	<b>3,947</b>	<b>4,038</b>	<b>0,140</b>

Tabla 7.18: Resultados utilizando el fichero ESCA con ruido blanco (SNR=0dB).

En las tablas anteriores se observa de forma inequívoca que la **activación del filtro paso banda** mejora todos los resultados de SNR y distancia espectral, aunque en las pruebas de audición las señales filtradas con aplicación de filtro paso banda pierden un poco de riqueza a altas frecuencias.

#### 7.2.4.-Parámetros $\beta$ y $\delta$

En esta prueba trataremos de encontrar cuales son los parámetros óptimos beta y delta utilizados en la construcción del filtro de Wiener.

Como ya se explicó en el capítulo anterior, el parámetro  $\beta$  es utilizado para realizar una sobreestimación de la densidad espectral de energía  $Pd(\omega)$  de la estimación de ruido. El parámetro  $\delta$  se utiliza para dar al filtro un carácter no lineal, atenuando de forma más notable las bandas de poca energía, valores de atenuación del filtro pequeños. Por el contrario, el valor de atenuación del filtro se mantiene prácticamente igual en las zonas de alta energía, valores de atenuación del filtro próximos a 1.

Utilizaremos los parámetros utilizados en el apartado 7.2.3. activando el filtro paso banda.

Parámetros utilizados:

*Longitud de trama: 256*

*Desplazamiento de trama: 128*

*Multiplicador de muestras de trama: 1 (desactivado)*

*Tipo de ventana (estimación espectros): 2 (Hanning)*

*Tipo de prefiltrado: 0 (desactivado)*

*Parámetro beta del prefiltrado: 1.00*

*Parámetro delta del prefiltrado: 1.00*

*Atenuación máxima de prefiltrado: 0.50*

*Tipo de filtro: 1 (wiener basado en modelado AR)*

*Orden del predictor sonoras(modelado): 10*

*Orden del predictor sordas(modelado): 10*

*Orden de los cumulantes (modelado): 1 (AR2 – Levinson Durbin)*

*Sobreumbral pitch: 0.00 (desactivado)*

*Numero de iteraciones del filtro: 1*

*Factor intertrama (ak): 1.00*

*Factor iteración trama previa: 1*

*Parámetro beta del filtro: \*\*\* parámetro a optimizar \*\*\**

*Parámetro delta del filtro: \*\*\* parámetro a optimizar \*\*\**

*Atenuación máxima de filtrado: 0.01*

*Nivel de ruido (silencio) en filtrado: 2500.00*

*Tipo de postfiltro: 0 (desactivado)*

*Nivel activación postfiltro: 0.50*

*Orden del filtro de mediana: 1*

*Parámetro sobreestimación ruido (VAD): 0.00 (desactivado)*

*Numero de tramas de ruido iniciales: 10*

*Parámetro promediado espectro ruido: 0.10*

*Reestimar ruido en tramas silencio: 0 (desactivado)*

$\delta=1$	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL		18,079	8,169	1,018	1,483	2,118	2,407	0,000
$\beta=1$		21,285	10,555	0,495	<b>0,795</b>	1,091	1,435	-0,019
$\beta=1.2$		21,343	10,619	0,485	0,816	1,064	1,418	-0,016
$\beta=1.5$		<b>21,374</b>	<b>10,663</b>	0,477	0,858	1,039	<b>1,408</b>	-0,012
$\beta=2$		21,334	10,656	0,474	0,937	1,025	1,420	-0,004

Tabla 7.19: Resultados utilizando el fichero ASUN1 con ruido blanco (SNR=18dB).

$\delta=2$	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL		18,079	8,169	1,018	1,483	2,118	2,407	0,000
$\beta=1$		21,297	10,641	<b>0,472</b>	1,043	<b>1,016</b>	1,441	-0,005
$\beta=1.2$		21,173	10,554	0,487	1,145	1,037	1,492	<b>0,001</b>
$\beta=1.5$		20,958	10,405	0,514	1,320	1,082	1,572	0,010
$\beta=2$		20,584	10,147	0,558	1,612	1,160	1,691	0,026

Tabla 7.20: Resultados utilizando el fichero ASUN1 con ruido blanco (SNR=18dB).

$\delta=1$	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL		9,061	1,522	1,745	2,718	3,500	3,192	0,000
$\beta=1$		14,348	5,565	0,947	<b>1,722</b>	1,641	2,118	-0,184
$\beta=1.2$		14,497	5,704	0,930	1,769	1,581	2,073	-0,158
$\beta=1.5$		14,641	5,841	0,914	1,849	1,517	2,023	-0,121
$\beta=2$		14,739	5,950	0,902	1,967	1,453	1,978	-0,064

Tabla 7.21: Resultados utilizando el fichero ASUN1 con ruido blanco (SNR=9dB).

$\delta=2$	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL		9,061	1,522	1,745	2,718	3,500	3,192	0,000
$\beta=1$		<b>14,810</b>	<b>6,058</b>	<b>0,887</b>	2,223	1,372	<b>1,942</b>	-0,048
$\beta=1.2$		14,752	6,036	0,893	2,361	<b>1,365</b>	1,963	<b>-0,003</b>
$\beta=1.5$		14,593	5,946	0,912	2,614	1,389	2,028	0,061
$\beta=2$		14,245	5,739	0,960	3,234	1,464	2,171	0,160

Tabla 7.22: Resultados utilizando el fichero ASUN1 con ruido blanco (SNR=9dB).

$\delta=1$	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL		0,061	-3,918	2,523	4,266	5,146	3,787	0,000
$\beta=1$		7,628	1,116	1,563	2,420	2,400	2,832	<b>0,300</b>
$\beta=1.2$		7,799	1,289	1,537	2,415	2,301	2,787	0,397
$\beta=1.5$		7,945	1,457	1,510	<b>2,402</b>	2,184	2,732	0,523
$\beta=2$		8,015	1,596	1,484	2,419	2,048	2,665	0,695

Tabla 7.23: Resultados utilizando el fichero ASUN1 con ruido blanco (SNR=0dB).

$\delta=2$	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL		0,061	-3,918	2,523	4,266	5,146	3,787	0,000
$\beta=1$		<b>8,087</b>	<b>1,842</b>	<b>1,417</b>	2,998	1,777	2,546	0,803
$\beta=1.2$		7,928	1,799	1,421	3,254	1,730	<b>2,526</b>	0,927
$\beta=1.5$		7,652	1,693	1,440	3,762	<b>1,706</b>	2,532	1,075
$\beta=2$		7,195	1,482	1,495	4,687	1,738	2,600	1,262

Tabla 7.24: Resultados utilizando el fichero ASUN1 con ruido blanco (SNR=0dB).

$\delta=1$	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL		18,048	9,165	2,220	3,075	3,818	3,961	0,000
$\beta=1$		18,932	11,579	1,631	1,902	2,396	2,989	-0,115
$\beta=1.2$		18,974	11,667	1,607	1,881	2,345	2,973	-0,110
$\beta=1.5$		19,007	11,750	1,583	<b>1,864</b>	2,286	<b>2,961</b>	-0,104
$\beta=2$		<b>19,013</b>	11,811	1,562	1,870	2,223	2,959	-0,094

Tabla 7.25: Resultados utilizando el fichero ESCA con ruido blanco (SNR=18dB).

$\delta=2$	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL		18,048	9,165	2,220	3,075	3,818	3,961	0,000
$\beta=1$		18,992	<b>11,837</b>	<b>1,540</b>	1,899	2,164	2,970	-0,093
$\beta=1.2$		18,954	11,825	1,548	1,953	<b>2,151</b>	2,991	-0,084
$\beta=1.5$		18,880	11,780	1,570	2,056	2,156	3,022	-0,071
$\beta=2$		18,703	11,640	1,617	2,259	2,187	3,068	<b>-0,061</b>

Tabla 7.26: Resultados utilizando el fichero ESCA con ruido blanco (SNR=18dB).

$\delta=1$	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL		9,031	2,262	3,195	4,548	5,390	4,659	0,000
$\beta=1$		13,856	6,992	2,285	2,969	3,107	3,507	-0,126
$\beta=1.2$		14,024	7,202	2,252	<b>2,962</b>	3,026	3,469	-0,102
$\beta=1.5$		14,201	7,429	2,217	2,965	2,931	3,424	-0,068
$\beta=2$		14,364	7,657	2,181	2,984	2,821	3,369	-0,016

Tabla 7.27: Resultados utilizando el fichero ESCA con ruido blanco (SNR=9dB).

$\delta=2$	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL		9,031	2,262	3,195	4,548	5,390	4,659	0,000
$\beta=1$		14,495	7,925	2,122	3,050	2,654	3,302	<b>-0,003</b>
$\beta=1.2$		<b>14,510</b>	<b>7,969</b>	<b>2,112</b>	3,107	2,599	3,281	0,040
$\beta=1.5$		14,462	7,963	<b>2,112</b>	3,256	2,551	<b>3,276</b>	0,101
$\beta=2$		14,283	7,859	2,138	3,618	<b>2,522</b>	3,307	0,165

Tabla 7.28: Resultados utilizando el fichero ESCA con ruido blanco (SNR=9dB).

$\delta=1$	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL		0,023	-3,563	4,120	6,179	7,146	5,148	0,000
$\beta=1$		7,565	2,245	2,956	3,872	3,947	4,038	<b>0,140</b>
$\beta=1.2$		7,809	2,472	2,918	3,828	3,833	4,005	0,200
$\beta=1.5$		8,056	2,716	2,873	3,789	3,697	3,964	0,282
$\beta=2$		8,271	2,966	2,824	<b>3,750</b>	3,534	3,913	0,404

Tabla 7.29: Resultados utilizando el fichero ESCA con ruido blanco (SNR=0dB).

$\delta=2$	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL		0,023	-3,563	4,120	6,179	7,146	5,148	0,000
$\beta=1$		<b>8,553</b>	3,539	2,675	3,907	3,144	3,801	0,406
$\beta=1.2$		8,504	<b>3,581</b>	<b>2,659</b>	3,972	3,047	3,763	0,464
$\beta=1.5$		8,333	3,530	2,660	4,175	2,957	3,732	0,552
$\beta=2$		7,963	3,331	2,692	4,640	<b>2,893</b>	<b>3,725</b>	0,699

Tabla 7.30: Resultados utilizando el fichero ESCA con ruido blanco (SNR=0dB).

En las pruebas, si nuestro sistema sólo consistiera en una etapa de filtrado de Wiener podríamos decantarnos a valores de  $\delta=2$  y  $\beta=1$ , valores que en general obtienen los mejores resultados de SNR y distancia espectral en condiciones de mucho ruido, pero las pruebas de audición revelan que la señal se vuelve demasiado paso bajo.

Teniendo en cuenta que el sistema posee un prefiltro que elimina algo de ruido, los valores  **$\delta=1$  y  $\beta=1.2$  podrían ser los indicados**, ya que con ellos se obtiene una ligera mejora de resultados SNR y distancia espectral sin apenas distorsionar la señal en las pruebas de audición.

### 7.2.5.-Parámetro número de iteraciones

En este experimento intentaremos determinar el rendimiento en reducción de ruido de la técnica de filtrado de Wiener iterativo.

Como se ha explicado ya en anteriores capítulos el filtrado iterativo intenta aumentar el rendimiento de filtrado reconstruyéndolo una y otra vez el filtro, tantas veces como iteraciones, a partir de una mejor estimación del espectro de voz  $\mathbf{Ps}(\omega)$ , esta mejor estimación se obtiene a partir de la trama filtrada con un filtro construido en una iteración anterior. De esta manera, cada vez se estima el espectro de voz de una trama de señal mas “ limpia”, ya que se ha filtrado con un filtro mejor construido.

También analizaremos las ventajas de utilizar estadísticas de orden superior en la estimación del espectro de voz en lugar de las estadísticas de segundo orden tradicionales o correlaciones.

En las siguientes pruebas utilizaremos los mismos parámetros utilizados en el punto 7.2.4., con  $\delta=1$  y  $\beta=1.2$ .

Parámetros utilizados:

*Longitud de trama: 256*

*Desplazamiento de trama: 128*

*Multiplicador de muestras de trama: 1 (desactivado)*

*Tipo de ventana (estimación espectros): 2 (Hanning)*

*Tipo de prefiltro: 0 (desactivado)*

*Parámetro beta del prefiltro: 1.00*

*Parámetro delta del prefiltro: 1.00*

*Atenuación máxima de prefiltrado: 0.50*

*Tipo de filtro: 1 (wiener basado en modelado AR)*

*Orden del predictor sonoras(modelado): 10*

*Orden del predictor sordas(modelado): 10*

*Orden de los cumulantes (modelado): \*\*\* parámetro a optimizar \*\*\**

*Sobreumbral pitch: 0.00 (desactivado)*

*Numero de iteraciones del filtro: \*\*\* parámetro a optimizar \*\*\**

*Factor intertrama (ak): 1.00*

*Factor iteración trama previa: 1*

*Parámetro beta del filtro: 1.20*

*Parámetro delta del filtro: 1.00*

*Atenuación máxima de filtrado: 0.01*

*Nivel de ruido (silencio) en filtrado: 2500.00*

*Tipo de postfiltro: 0 (desactivado)*

*Nivel activación postfiltro: 0.50*

*Orden del filtro de mediana: 1*

*Parámetros sobreestimación ruido (VAD): 0.00 (desactivado)*

*Numero de tramas de ruido iniciales: 10*

*Parámetros promediado espectro ruido: 0.10*

*Reestimar ruido en tramas silencio: 0 (desactivado)*

AR2	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL		18,079	8,169	1,018	1,483	2,118	2,407	0,000
iter=1		21,343	10,619	0,485	<b>0,816</b>	1,064	<b>1,418</b>	<b>-0,016</b>
iter=2		<b>21,377</b>	<b>10,735</b>	<b>0,476</b>	0,981	<b>1,000</b>	1,445	-0,030
iter=3		21,204	10,619	0,543	1,407	1,088	1,585	-0,028
iter=4		21,071	10,528	0,585	1,687	1,141	1,661	-0,020

Tabla 7.31a: Resultados con modelado AR2 utilizando el fichero ASUN1 con ruido blanco (SNR=18dB).

AR3	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL		18,079	8,169	1,018	1,483	2,118	2,407	0,000
iter=1		<b>19,862</b>	<b>9,918</b>	<b>0,660</b>	<b>2,374</b>	<b>1,271</b>	<b>1,975</b>	<b>-0,030</b>
iter=2		18,641	9,284	0,752	3,453	1,427	2,239	-0,039
iter=3		18,236	9,122	0,789	3,863	1,477	2,340	-0,045
iter=4		17,785	8,890	0,804	4,082	1,510	2,405	-0,022

Tabla 7.31b: Resultados con modelado AR3 utilizando el fichero ASUN1 con ruido blanco (SNR=18dB).

AR4	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL		18,079	8,169	1,018	1,483	2,118	2,407	0,000
iter=1		<b>21,134</b>	<b>10,498</b>	<b>0,521</b>	<b>1,002</b>	<b>1,069</b>	<b>1,509</b>	-0,013
iter=2		20,832	10,374	0,606	1,763	1,121	1,743	-0,010
iter=3		20,680	10,300	0,635	2,090	1,158	1,809	-0,009
iter=4		20,005	10,069	0,642	2,208	1,174	1,829	<b>-0,008</b>

Tabla 7.31c: Resultados con modelado AR4 utilizando el fichero ASUN1 con ruido blanco (SNR=18dB).

AR2	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL		9,061	1,522	1,745	2,718	3,500	3,192	0,000
iter=1		14,497	5,704	0,930	<b>1,769</b>	1,581	2,073	-0,158
iter=2		<b>15,040</b>	6,108	<b>0,837</b>	2,087	<b>1,376</b>	<b>1,998</b>	-0,102
iter=3		14,984	<b>6,113</b>	0,992	4,132	1,595	2,376	-0,049
iter=4		14,787	6,022	1,100	5,513	1,754	2,606	<b>-0,011</b>

Tabla 7.32a: Resultados con modelado AR2 utilizando el fichero ASUN1 con ruido blanco (SNR=9dB).

AR3	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL		9,061	1,522	1,745	2,718	3,500	3,192	0,000
iter=1		<b>14,358</b>	<b>5,844</b>	<b>0,985</b>	<b>3,807</b>	<b>1,600</b>	<b>2,339</b>	-0,072
iter=2		14,028	5,764	1,244	6,243	1,823	2,815	<b>0,055</b>
iter=3		13,595	5,574	1,273	6,628	1,944	2,953	0,090
iter=4		13,131	5,352	1,307	6,730	1,966	3,001	0,129

Tabla 7.32b: Resultados con modelado AR3 utilizando el fichero ASUN1 con ruido blanco (SNR=9dB).

AR4	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL		9,061	1,522	1,745	2,718	3,500	3,192	0,000
iter=1		14,306	5,611	<b>0,909</b>	<b>1,832</b>	1,516	<b>1,999</b>	<b>-0,167</b>
iter=2		<b>14,489</b>	<b>5,762</b>	1,033	3,376	<b>1,489</b>	2,241	-0,185
iter=3		14,172	5,601	1,181	5,086	1,621	2,533	-0,168
iter=4		14,112	5,564	1,216	5,675	1,697	2,670	-0,148

Tabla 7.32c: Resultados con modelado AR4 utilizando el fichero ASUN1 con ruido blanco (SNR=9dB).

AR2	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL		0,061	-3,918	2,523	4,266	5,146	3,787	0,000
iter=1		7,799	1,289	1,537	<b>2,415</b>	2,301	2,787	<b>0,397</b>
iter=2		8,282	1,460	<b>1,323</b>	2,617	<b>1,828</b>	<b>2,633</b>	0,541
iter=3		8,279	1,501	1,585	5,571	2,002	3,033	0,542
iter=4		<b>8,289</b>	<b>1,543</b>	1,738	6,898	2,197	3,320	0,483

Tabla 7.33a: Resultados con modelado AR2 utilizando el fichero ASUN1 con ruido blanco (SNR=0dB).

AR3	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL		0,061	-3,918	2,523	4,266	5,146	3,787	0,000
iter=1		8,210	1,593	<b>1,429</b>	<b>2,675</b>	<b>1,957</b>	<b>2,571</b>	<b>0,417</b>
iter=2		8,650	1,870	1,607	5,962	2,071	3,068	0,535
iter=3		8,594	1,879	1,689	6,631	2,165	3,320	0,727
iter=4		<b>8,696</b>	<b>2,050</b>	1,732	6,897	2,185	3,322	0,627

Tabla 7.33b: Resultados con modelado AR3 utilizando el fichero ASUN1 con ruido blanco (SNR=0dB).

AR4	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL		0,061	-3,918	2,523	4,266	5,146	3,787	0,000
iter=1		7,789	1,324	1,531	<b>2,454</b>	2,225	2,783	<b>0,165</b>
iter=2		8,248	1,503	<b>1,428</b>	3,069	<b>1,892</b>	<b>2,698</b>	0,320
iter=3		<b>8,344</b>	<b>1,546</b>	1,671	5,147	1,916	3,060	0,381
iter=4		8,318	1,545	1,771	5,979	1,998	3,200	0,398

Tabla 7.33c: Resultados con modelado AR4 utilizando el fichero ASUN1 con ruido blanco (SNR=0dB).

AR2	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL		18,048	9,165	2,220	3,075	3,818	3,961	0,000
iter=1		18,974	11,667	1,607	1,881	2,345	2,973	-0,110
iter=2		<b>19,068</b>	<b>11,808</b>	<b>1,529</b>	<b>1,847</b>	<b>2,138</b>	<b>2,969</b>	-0,106
iter=3		19,005	11,761	1,594	2,091	2,164	3,025	<b>-0,102</b>
iter=4		18,958	11,720	1,634	2,251	2,193	3,053	-0,098

Tabla 7.34a: Resultados con modelado AR2 utilizando el fichero ESCA con ruido blanco (SNR=18dB).

AR3	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL		18,048	9,165	2,220	3,075	3,818	3,961	0,000
iter=1		<b>18,491</b>	<b>11,390</b>	<b>1,712</b>	<b>2,647</b>	<b>2,246</b>	<b>3,124</b>	<b>-0,107</b>
iter=2		18,106	11,089	1,778	2,967	2,300	3,180	-0,113
iter=3		17,693	10,927	1,795	3,090	2,311	3,191	-0,115
iter=4		17,926	10,964	1,800	3,134	2,316	3,197	-0,117

Tabla 7.34b: Resultados con modelado AR3 utilizando el fichero ESCA con ruido blanco (SNR=18dB).

AR4	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL		18,048	9,165	2,220	3,075	3,818	3,961	0,000
iter=1		<b>18,888</b>	11,610	<b>1,576</b>	<b>1,904</b>	2,227	<b>2,956</b>	-0,116
iter=2		18,885	<b>11,615</b>	1,603	2,164	2,168	3,001	<b>-0,114</b>
iter=3		18,102	11,548	1,612	2,261	<b>2,166</b>	3,018	-0,114
iter=4		18,782	11,547	1,622	2,296	2,168	3,034	-0,114

Tabla 7.34c: Resultados con modelado AR4 utilizando el fichero ESCA con ruido blanco (SNR=18dB).

AR2	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL		9,031	2,262	3,195	4,548	5,390	4,659	0,000
iter=1		14,024	7,202	2,252	2,962	3,026	3,469	-0,102
iter=2		<b>14,691</b>	<b>7,666</b>	<b>2,045</b>	<b>2,873</b>	2,603	<b>3,242</b>	-0,036
iter=3		14,686	7,643	2,110	3,916	<b>2,545</b>	3,342	<b>0,008</b>
iter=4		14,558	7,538	2,195	4,617	2,588	3,431	0,043

Tabla 7.35a: Resultados con modelado AR2 utilizando el fichero ESCA con ruido blanco (SNR=9dB).

AR3	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL		9,031	2,262	3,195	4,548	5,390	4,659	0,000
iter=1		<b>14,143</b>	<b>7,362</b>	<b>2,111</b>	<b>4,048</b>	<b>2,610</b>	<b>3,369</b>	-0,099
iter=2		13,692	7,174	2,225	5,092	2,619	3,451	-0,067
iter=3		13,736	7,242	2,270	5,308	2,632	3,490	-0,022
iter=4		13,551	7,222	2,285	5,371	2,645	3,498	<b>0,015</b>

Tabla 7.35b: Resultados con modelado AR3 utilizando el fichero ESCA con ruido blanco (SNR=9dB).

AR4	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL		9,031	2,262	3,195	4,548	5,390	4,659	0,000
iter=1		13,822	7,038	2,175	<b>2,965</b>	2,892	3,352	-0,161
iter=2		<b>14,334</b>	<b>7,378</b>	<b>2,025</b>	3,221	2,519	<b>3,290</b>	-0,134
iter=3		14,228	7,343	2,065	3,751	<b>2,483</b>	3,317	-0,129
iter=4		14,135	7,282	2,090	4,065	2,488	3,330	<b>-0,127</b>

Tabla 7.35c: Resultados con modelado AR4 utilizando el fichero ESCA con ruido blanco (SNR=9dB).

AR2	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL		0,023	-3,563	4,120	6,179	7,146	5,148	0,000
iter=1		7,809	2,472	2,918	3,828	3,833	4,005	0,200
iter=2		8,835	3,120	<b>2,539</b>	<b>3,537</b>	3,162	<b>3,715</b>	0,195
iter=3		<b>8,987</b>	<b>3,274</b>	2,627	4,985	<b>2,969</b>	3,781	0,188
iter=4		8,925	3,231	2,755	6,058	2,995	3,917	<b>0,172</b>

Tabla 7.36a: Resultados con modelado AR2 utilizando el fichero ESCA con ruido blanco (SNR=0dB).

AR3	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL		0,023	-3,563	4,120	6,179	7,146	5,148	0,000
iter=1		8,367	2,758	2,657	<b>3,804</b>	3,305	3,727	0,213
iter=2		8,548	3,076	<b>2,543</b>	5,504	<b>2,953</b>	<b>3,720</b>	<b>0,189</b>
iter=3		8,526	3,212	2,624	6,111	2,965	3,751	0,245
iter=4		<b>8,662</b>	<b>3,312</b>	2,716	6,300	3,001	3,783	0,311

Tabla 7.36b: Resultados con modelado AR3 utilizando el fichero ESCA con ruido blanco (SNR=0dB).

AR4	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL		0,023	-3,563	4,120	6,179	7,146	5,148	0,000
iter=1		7,577	2,327	2,954	3,894	3,774	4,039	0,032
iter=2		8,390	2,805	2,528	<b>3,634</b>	3,095	<b>3,750</b>	<b>-0,023</b>
iter=3		<b>8,540</b>	<b>2,933</b>	<b>2,526</b>	4,313	2,881	3,800	0,033
iter=4		8,298	2,797	2,578	4,883	<b>2,878</b>	3,885	0,119

Tabla 7.36c: Resultados con modelado AR4 utilizando el fichero ESCA con ruido blanco (SNR=0dB).

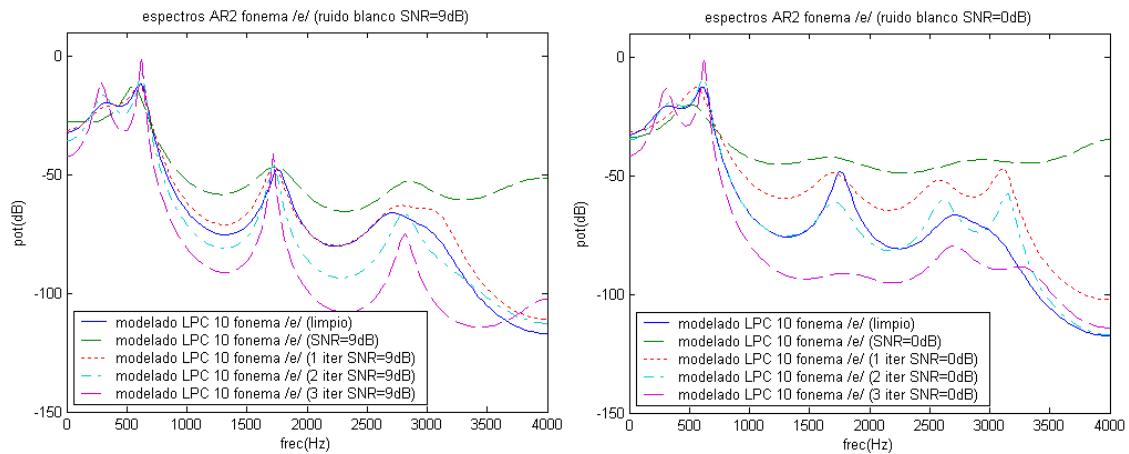


Fig. 7.1: Espectros de modelado AR2 LPC10 obtenidos en diferentes iteraciones bajo condiciones de ruido blanco con  $\text{SNR} = 9$  y  $0$  dB

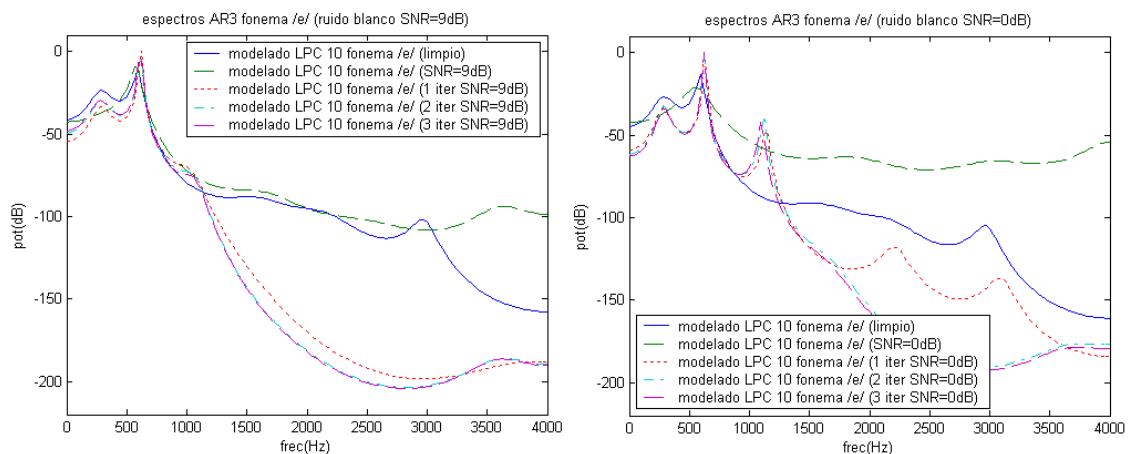


Fig. 7.2: Espectros de modelado AR2 LPC10 obtenidos en diferentes iteraciones bajo condiciones de ruido blanco con  $\text{SNR} = 9$  y  $0$  dB

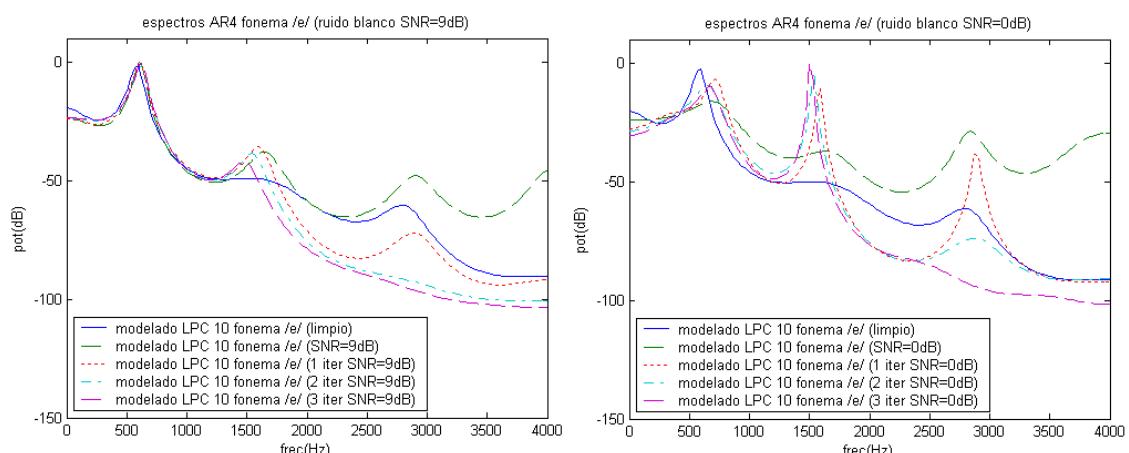


Fig. 7.3: Espectros de modelado AR2 LPC10 obtenidos en diferentes iteraciones bajo condiciones de ruido blanco con  $\text{SNR} = 9$  y  $0$  dB

Después de analizar las medidas objetivas de las diferentes pruebas podemos obtener dos conclusiones. Por un lado, la utilización de estadísticas de orden superior frente a estadísticas de segundo orden sólo mejora los resultados de medidas objetivas en condiciones de mucho ruido,  $SNR = 0$  dB. Estos resultados pueden venir condicionados al hecho de la utilización de una longitud de trama de 256 muestras, con este reducido número de muestras, con el que el sistema se ve obligado a trabajar por motivos de estacionariedad, no es suficiente para obtener unas buenas estadísticas de orden superior, y por lo tanto de realizar un mejor modelado que las estadísticas de segundo orden.

Por otro lado, observamos que la técnica de filtrado iterativo, en condiciones de ruido de  $SNR = 0$  y 9 dB, mejora ligeramente los resultados de medidas objetivas, pero a costa de introducir un desagradable ruido musical que por consecuencia de la generación espontánea de espurios en las zonas de baja energía del modelado espectral de la voz.

#### **7.2.6.-Parámetro activar filtro peine**

En este experimento evaluaremos el aumento de rendimiento de reducción de ruido que se introduce en el sistema al complementar el filtro de Wiener, basado en modelado AR, con un filtro peine adaptativo como se explica en el capítulo anterior.

En las pruebas se utilizaran los mismos parámetros utilizados en el punto 7.2.5., utilizando una sola iteración de filtrado. Por otro lado, en el punto anterior comienza a ser evidente, por las pruebas de audición, que un desplazamiento de 128 muestras resulta insuficiente, ya que no elimina las discontinuidades creadas al realizar un filtrado segmentado. Estas discontinuidades son probablemente producidas por el hecho de que en nuestro sistema sólo se ha realizado enventanado de la señal al realizar estimaciones de su espectro y no para filtrar la señal, dando de esta manera mayor libertad (rompiendo con la limitación de solapamiento al 50%), mayor precisión (ya que un solapamiento mayor del 50% mejora la adaptación a transiciones) y mayor promediado de tramas al sistema.

Por los motivos anteriores realizaremos las pruebas con un sobreumbral de detección de pitch igual a 0.5 y comparando los resultados utilizando un desplazamiento de trama igual a 128 y 64 muestras.

Parámetros utilizados:

*Longitud de trama: 256*

*Desplazamiento de trama: 128 y 64*

*Multiplicador de muestras de trama: 1 (desactivado)*

*Tipo de ventana (estimación espectros): 2 (Hanning)*

*Tipo de prefiltro: 0 (desactivado)*

*Parámetro beta del prefiltro: 1.00*

*Parámetro delta del prefiltro: 1.00*

*Atenuación máxima de prefiltrado: 0.50*

*Tipo de filtro: 1 (wiener basado en modelado AR)*

*Orden del predictor sonoras(modelado): 10*

*Orden del predictor sordas(modelado): 10*

*Orden de los cumulantes (modelado): 1 (AR2 – Levinson Durbin)*

*Sobreumbral pitch: \*\*\* parámetro a optimizar \*\*\**

*Numero de iteraciones del filtro: 1*

*Factor intertrama (ak): 1.00*

*Factor iteración trama previa: 1*

*Parámetro beta del filtro: 1.20*

*Parámetro delta del filtro: 1.00*

*Atenuación máxima de filtrado: 0.01*

*Nivel de ruido (silencio) en filtrado: 2500.00*

*Tipo de postfiltro: 0 (desactivado)*

*Nivel activación postfiltro: 0.50*

*Orden del filtro de mediana: 1*

*Parámetro sobreestimación ruido (VAD): 0.00 (desactivado)*

*Numero de tramas de ruido iniciales: 10*

*Parámetro promediado espectro ruido: 0.10*

*Reestimar ruido en tramas silencio: 0 (desactivado)*

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL		18,079	8,169	1,018	1,483	2,118	2,407	0,000
NO PEI128		21,343	10,619	<b>0,485</b>	<b>0,816</b>	1,064	<b>1,418</b>	-0,016
NO PEI_64		<b>21,716</b>	<b>10,869</b>	0,522	0,836	1,120	1,492	-0,018
PEINE128		15,911	7,756	0,495	0,882	<b>1,051</b>	1,493	<b>0,010</b>
PEINE_64		18,904	9,385	0,517	0,876	1,109	1,551	0,072

Tabla 7.37: Resultados utilizando el fichero ASUN1 con ruido blanco (SNR=18dB).

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL		9,061	1,522	1,745	2,718	3,500	3,192	0,000
NO PEI128		14,497	5,704	0,930	1,769	1,581	2,073	-0,158
NO PEI_64		14,584	5,768	0,958	1,732	1,637	2,121	-0,139
PEINE128		13,485	5,343	<b>0,876</b>	<b>1,673</b>	<b>1,444</b>	<b>1,937</b>	0,098
PEINE_64		<b>14,983</b>	<b>6,164</b>	0,896	1,703	1,488	1,968	<b>0,027</b>

Tabla 7.38: Resultados utilizando el fichero ASUN1 con ruido blanco (SNR=9dB).

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL		0,061	-3,918	2,523	4,266	5,146	3,787	0,000
NO PEI128		7,799	1,289	1,537	2,415	2,301	2,787	<b>0,397</b>
NO PEI_64		7,681	1,202	1,594	2,451	2,380	2,821	0,439
PEINE128		8,917	1,911	<b>1,430</b>	2,255	<b>2,109</b>	<b>2,590</b>	0,620
PEINE_64		<b>9,116</b>	<b>1,966</b>	1,489	<b>2,214</b>	2,199	2,623	0,853

Tabla 7.39: Resultados utilizando el fichero ASUN1 con ruido blanco (SNR=0dB).

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL		18,048	9,165	2,220	3,075	3,818	3,961	0,000
NO PEI128		18,974	11,667	1,607	1,881	2,345	2,973	-0,110
NO PEI_64		<b>19,409</b>	<b>12,049</b>	1,647	1,925	2,413	3,037	-0,068
PEINE128		15,223	9,245	<b>1,555</b>	<b>1,817</b>	<b>2,248</b>	<b>2,932</b>	<b>-0,032</b>
PEINE_64		18,170	11,230	1,591	1,858	2,344	2,998	0,216

Tabla 7.40: Resultados utilizando el fichero ESCA con ruido blanco (SNR=18dB).

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL		9,031	2,262	3,195	4,548	5,390	4,659	0,000
NO PEI128		14,024	7,202	2,252	2,962	3,026	3,469	-0,102
NO PEI_64		14,217	7,282	2,300	2,961	3,113	3,529	<b>-0,048</b>
PEINE_128		12,790	6,672	<b>2,128</b>	2,795	<b>2,792</b>	<b>3,292</b>	-0,090
PEINE_64		<b>14,947</b>	<b>7,911</b>	2,156	<b>2,747</b>	2,937	3,345	0,309

Tabla 7.41: Resultados utilizando el fichero ESCA con ruido blanco (SNR=9dB).

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL		0,023	-3,563	4,120	6,179	7,146	5,148	0,000
NO PEI128		7,809	2,472	2,918	3,828	3,833	4,005	0,200
NO PEI_64		7,795	2,321	2,993	3,937	3,929	4,052	<b>0,191</b>
PEINE128		8,991	3,148	<b>2,776</b>	<b>3,636</b>	<b>3,630</b>	<b>3,844</b>	0,371
PEINE_64		<b>9,425</b>	<b>3,181</b>	2,849	3,704	3,807	3,901	0,688

Tabla 7.42: Resultados utilizando el fichero ESCA con ruido blanco (SNR=0dB).

Observando los resultados de medidas, observamos que los mejores resultados de SNR y SNRs en condiciones de mucho ruido se consiguen con la activación del filtro peine y utilizando un desplazamiento de trama de 64 muestras, por el contrario, los mejores resultados de distancia espectral se consiguen con una longitud de trama de 128 muestras. En situaciones de poco ruido, si nos guiamos por los resultados objetivos, es mejor no activar el filtro peine.

Las pruebas de audición nos confirman lo que planteábamos al principio del apartado, en caso de activar el filtro peine, es necesario utilizar un desplazamiento de trama de 64 muestras (solapamiento del 75%) para poder eliminar el molesto ruido de discontinuidades de señal. Por otro lado, la activación del filtro peine mejora mucho la calidad de audición de los fonemas sonoros en condiciones de mucho ruido, pero a costa de añadirles un ligero zumbido.

### 7.2.7.-Parámetro añadir prefiltro

En este experimento intentaremos mejorar los resultados en condiciones de mucho ruido con la activación de un prefiltrado basado en sustracción espectral. Este método es muy efectivo para reducir ruido, pero, desgraciadamente, también es un gran generador de ruido musical. Este ruido musical es muy desagradable para el oído humano, así que atenuar sus efectos hemos impuesto una atenuación máxima de prefiltrado de 3dB en amplitud (6 dB en energía).

Para realizar las pruebas utilizaremos los mismos parámetros del apartado 7.2.6., con el filtro peine activado y utilizando un desplazamiento de trama de 64 muestras.

Parámetros utilizados:

*Longitud de trama:* 256

*Desplazamiento de trama:* 64

*Multiplicador de muestras de trama:* 1 (desactivado)

*Tipo de ventana (estimación espectros):* 2 (Hanning)

*Tipo de prefiltro:* \*\*\* parámetro a optimizar \*\*\*

*Parámetro beta del prefiltro:* \*\*\* parámetro a optimizar \*\*\*

*Parámetro delta del prefiltro:* \*\*\* parámetro a optimizar \*\*\*

*Atenuación máxima de prefiltrado:* 0.50

*Tipo de filtro:* 1 (wiener basado en modelado AR)

*Orden del predictor sonoras(modelado):* 10

*Orden del predictor sordas(modelado):* 10

*Orden de los cumulantes (modelado):* 1 (AR2 – Levinson Durbin)

*Sobreumbral pitch:* 0.50 (activado)

*Numero de iteraciones del filtro:* 1

*Factor intertrama (ak):* 1.00

*Factor iteración trama previa:* 1

*Parámetro beta del filtro:* 1.20

*Parámetro delta del filtro:* 1.00

*Atenuación máxima de filtrado:* 0.01

*Nivel de ruido (silencio) en filtrado:* 2500.00

*Tipo de postfiltro:* 0 (desactivado)

*Nivel activación postfiltro:* 0.50

*Orden del filtro de mediana:* 1

*Parámetro sobreestimación ruido (VAD):* 0.00 (desactivado)

*Numero de tramas de ruido iniciales:* 10

*Parámetro promediado espectro ruido:* 0.10

*Reestimar ruido en tramas silencio:* 0 (desactivado)

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL		18,079	8,169	1,018	1,483	2,118	2,407	0,000
NO PRE		18,904	9,385	0,517	0,876	1,109	1,551	<b>0,072</b>
PREβ1δ1		19,703	9,827	<b>0,491</b>	<b>0,819</b>	<b>1,065</b>	<b>1,521</b>	0,083
PREβ1.2δ1		<b>19,707</b>	<b>9,840</b>	0,497	0,860	1,069	1,539	0,085
PREβ1.2δ2		19,571	9,767	0,528	1,079	1,103	1,609	0,101

Tabla 7.43: Resultados utilizando el fichero ASUN1 con ruido blanco (SNR=18dB).

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL		9,061	1,522	1,745	2,718	3,500	3,192	0,000
NO PRE		14,983	6,164	0,896	1,703	1,488	<b>1,968</b>	<b>0,027</b>
PREβ1δ1		15,376	6,314	<b>0,833</b>	<b>1,447</b>	1,467	1,969	0,052
PREβ1.2δ1		15,437	6,396	0,837	1,542	1,457	1,978	0,067
PREβ1.2δ2		<b>15,475</b>	<b>6,579</b>	0,864	1,988	<b>1,446</b>	2,020	0,134

Tabla 7.44: Resultados utilizando el fichero ASUN1 con ruido blanco (SNR=9dB).

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL		0,061	-3,918	2,523	4,266	5,146	3,787	0,000
NO PRE		9,116	1,966	1,489	2,214	2,199	2,623	0,853
PREβ1δ1		9,792	1,986	1,403	<b>1,920</b>	2,240	2,574	<b>0,520</b>
PREβ1.2δ1		9,932	2,145	1,399	1,948	2,179	2,558	0,573
PREβ1.2δ2		<b>10,090</b>	<b>2,603</b>	<b>1,387</b>	2,196	<b>1,973</b>	<b>2,512</b>	0,792

Tabla 7.45: Resultados utilizando el fichero ASUN1 con ruido blanco (SNR=0dB).

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL		18,048	9,165	2,220	3,075	3,818	3,961	0,000
NO PRE		18,170	11,230	1,591	1,858	2,344	2,998	0,216
PREβ1δ1		18,594	11,476	1,592	<b>1,838</b>	2,275	<b>2,972</b>	<b>0,169</b>
PREβ1.2δ1		<b>18,614</b>	11,534	1,586	1,843	2,262	2,984	<b>0,169</b>
PREβ1.2δ2		18,524	<b>11,595</b>	<b>1,583</b>	1,907	<b>2,236</b>	3,018	0,172

Tabla 7.46: Resultados utilizando el fichero ESCA con ruido blanco (SNR=18dB).

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL		9,031	2,262	3,195	4,548	5,390	4,659	0,000
NO PRE		14,947	7,911	2,156	2,747	2,937	3,345	0,309
PREβ1δ1		15,079	7,574	2,110	<b>2,573</b>	2,872	3,252	<b>0,299</b>
PREβ1.2δ1		15,128	7,726	2,099	2,598	2,828	<b>3,248</b>	0,301
PREβ1.2δ2		<b>15,186</b>	<b>8,236</b>	<b>2,072</b>	2,800	<b>2,707</b>	<b>3,248</b>	0,310

Tabla 7.47: Resultados utilizando el fichero ESCA con ruido blanco (SNR=9dB).

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL		0,023	-3,563	4,120	6,179	7,146	5,148	0,000
NO PRE		9,425	3,181	2,849	3,704	3,807	3,901	0,688
PREβ1δ1		9,895	2,929	2,707	3,341	3,848	3,781	<b>0,528</b>
PREβ1.2δ1		10,127	3,188	2,680	<b>3,308</b>	3,745	3,747	0,534
PREβ1.2δ2		<b>10,659</b>	<b>4,115</b>	<b>2,640</b>	3,422	<b>3,452</b>	<b>3,673</b>	0,588

Tabla 7.48: Resultados utilizando el fichero ESCA con ruido blanco (SNR=0dB).

Los resultados de medidas objetivas indican claramente que los mejores resultados se obtienen con la activación del prefiltro en cualquier situación.

Las pruebas de audición nos confirman lo mismo, la introducción de un prefiltrado hace que la voz suene más clara, pero, como ya temíamos, el prefiltrado nos introduce también un desagradable ruido musical.

Los parámetros óptimos para el prefiltrado son  **$\beta=1.2$  y  $\delta=2$** , con los que se obtienen los mejores resultados con mínimo ruido musical.

### 7.2.8.-Parámetro activar VAD.

Con la activación del VAD pretendemos conseguir dos metas; por un lado actualizar la estimación inicial de ruido en los espacios de silencio, es decir en las tramas que sean sólo ruido. Y por otro lado minimizar el ruido musical introducido en el filtrado con la substitución del filtrado de Wiener por un atenuador adaptativo en las tramas donde el VAD nos indique que no detecta voz.

En las pruebas utilizaremos los mismos parámetros que en el apartado 7.2.7., activando el prefiltrado con  $\beta=1.2$  y  $\delta=2$ .

Parámetros utilizados:

*Longitud de trama: 256*

*Desplazamiento de trama: 64*

*Multiplicador de muestras de trama: 1 (desactivado)*

*Tipo de ventana (estimación espectros): 2 (Hanning)*

*Tipo de prefiltrado: 2 (wiener basado en modelado AR)*

*Parámetro beta del prefiltrado: 1.20*

*Parámetro delta del prefiltrado: 2*

*Atenuación máxima de prefiltrado: 0.50*

*Tipo de filtro: 1 (wiener basado en modelado AR)*

*Orden del predictor sonoras(modelado): 10*

*Orden del predictor sordas(modelado): 10*

*Orden de los cumulantes (modelado): 1 (AR2 – Levinson Durbin)*

*Sobreumbral pitch: 0.50 (activado)*

*Numero de iteraciones del filtro: 1*

*Factor intertrama (ak): 1.00*

*Factor iteración trama previa: 1*

*Parámetro beta del filtro: 1.20*

*Parámetro delta del filtro: 1.00*

*Atenuación máxima de filtrado: 0.01*

*Nivel de ruido (silencio) en filtrado: 2500.00*

*Tipo de postfiltrado: 0 (desactivado)*

*Nivel activación postfiltrado: 0.50*

*Orden del filtro de mediana: 1*

*Parámetro sobreestimación ruido (VAD): \*\*\* parámetro a optimizar \*\*\**

*Numero de tramas de ruido iniciales: 10*

*Parámetro promediado espectro ruido: 0.10*

*Reestimar ruido en tramas silencio: \*\*\* parámetro a optimizar \*\*\**

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL		18,079	8,169	1,018	1,483	2,118	2,407	0,000
SIN VAD		19,571	9,767	0,528	1,079	1,103	1,609	0,101
VAD=1.1		19,617	9,818	0,535	1,097	1,101	1,613	0,096
VAD=1.25		19,588	9,807	0,530	1,114	1,102	1,615	0,096
VAD=1.5		19,317	9,579	0,573	1,597	1,152	1,663	0,106

Tabla 7.49: Resultados utilizando el fichero ASUN1 con ruido blanco (SNR=18dB).

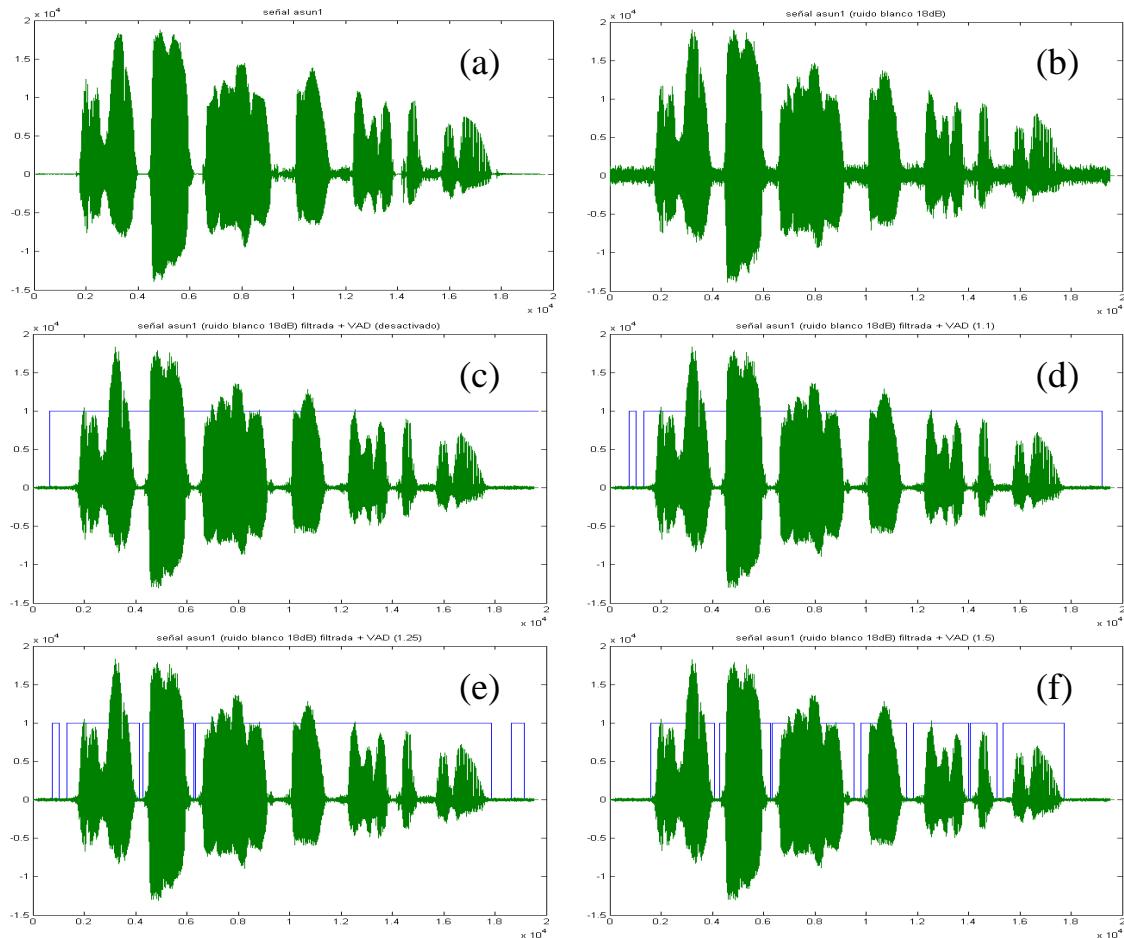


Fig.7.4: (a) Señal ASUN1 limpia. (b) Señal ASUN1 con ruido blanco (SNR=18dB). (c) Señal ASUN1 con ruido blanco (SNR=18dB) filtrada sin activar el VAD. (d) Señal ASUN1 con ruido blanco (SNR=18dB) filtrada activando el VAD (sobreumbral=1.1). (e) Señal ASUN1 con ruido blanco (SNR=18dB) filtrada activando el VAD (sobreumbral=1.25). (f) Señal ASUN1 con ruido blanco (SNR=18dB) filtrada activando el VAD (sobreumbral=1.5).

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN	🔊	154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL	🔊	9,061	1,522	1,745	2,718	3,500	3,192	0,000
SIN VAD	🔊	15,475	6,579	0,864	1,988	1,446	2,020	0,134
VAD=1.1	🔊	15,654	6,734	0,869	2,109	1,446	2,011	0,147
VAD=1.25	🔊	15,629	6,764	0,863	2,271	1,439	2,001	0,152
VAD=1.5	🔊	15,484	6,689	0,869	2,470	1,432	2,012	0,180

Tabla 7.50: Resultados utilizando el fichero ASUN1 con ruido blanco (SNR=9dB).

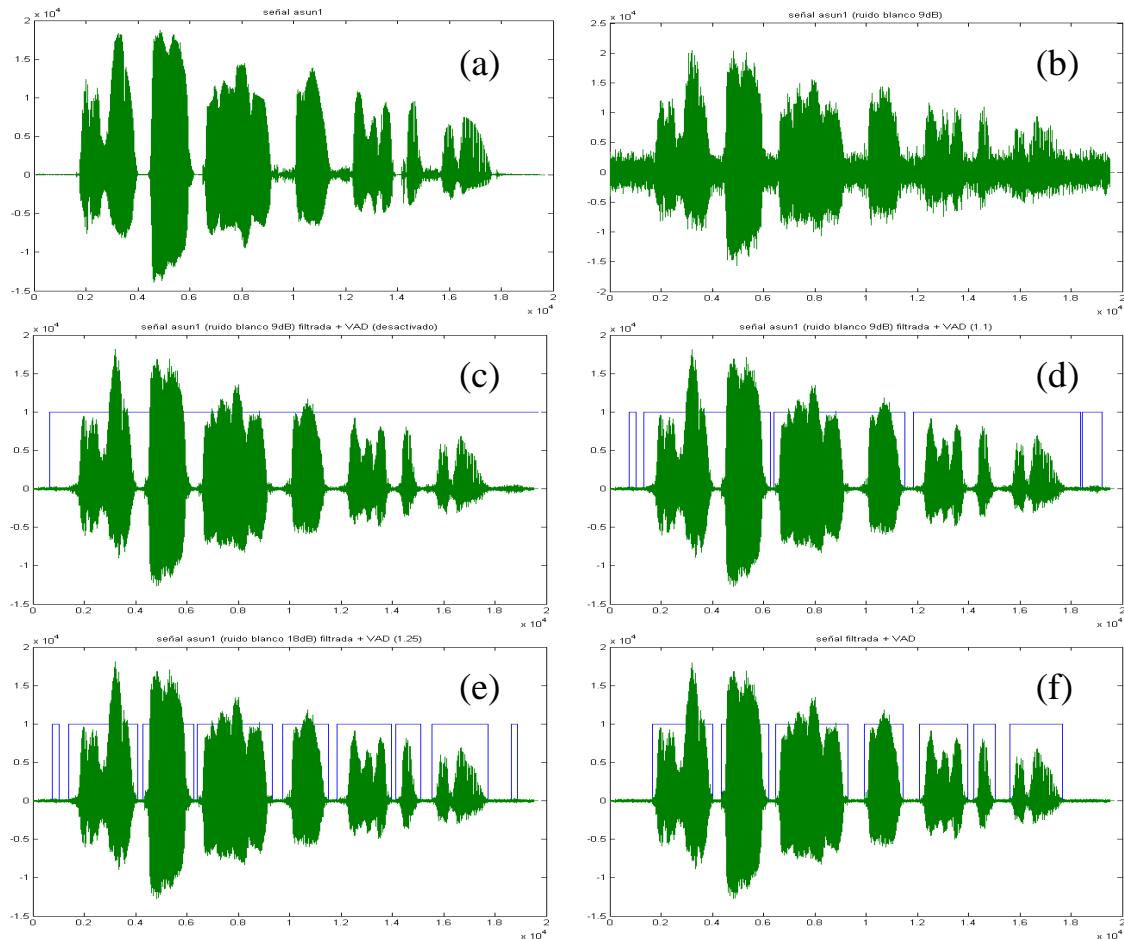


Fig.7.5: (a) Señal ASUN1 limpia. (b) Señal ASUN1 con ruido blanco (SNR=9dB). (c) Señal ASUN1 con ruido blanco (SNR=9dB) filtrada sin activar el VAD. (d) Señal ASUN1 con ruido blanco (SNR=9dB) filtrada activando el VAD (sobreumbral=1.1). (e) Señal ASUN1 con ruido blanco (SNR=9dB) filtrada activando el VAD (sobreumbral=1.25). (f) Señal ASUN1 con ruido blanco (SNR=9dB) filtrada activando el VAD (sobreumbral=1.5).

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN	🔊	154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL	🔊	0,061	-3,918	2,523	4,266	5,146	3,787	0,000
SIN VAD	🔊	10,090	2,603	1,387	2,196	1,973	2,512	0,792
VAD=1.1	🔊	10,480	2,996	1,355	2,310	1,894	2,436	0,729
VAD=1.25	🔊	10,557	3,220	1,387	2,978	1,804	2,464	0,715
VAD=1.5	🔊	9,974	3,051	1,495	3,655	1,763	2,522	0,802

Tabla 7.51: Resultados utilizando el fichero ASUN1 con ruido blanco (SNR=0dB).

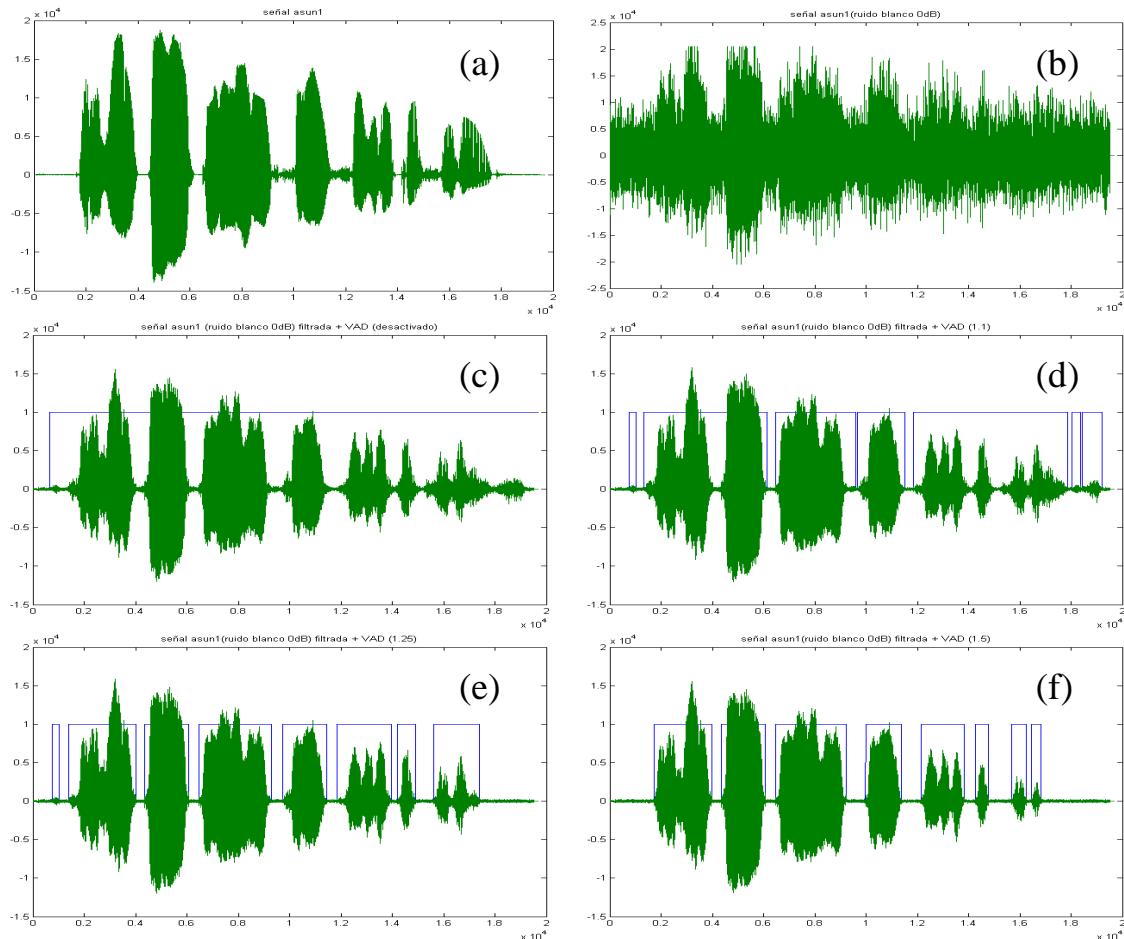


Fig.7.6: (a) Señal ASUN1 limpia. (b) Señal ASUN1 con ruido blanco (SNR=0dB). (c) Señal ASUN1 con ruido blanco (SNR=0dB) filtrada sin activar el VAD. (d) Señal ASUN1 con ruido blanco (SNR=0dB) filtrada activando el VAD (sobreumbral=1.1). (e) Señal ASUN1 con ruido blanco (SNR=0dB) filtrada activando el VAD (sobreumbral=1.25). (f) Señal ASUN1 con ruido blanco (SNR=0dB) filtrada activando el VAD (sobreumbral=1.5).

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN	🔊	154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL	🔊	18,048	9,165	2,220	3,075	3,818	3,961	0,000
SIN VAD	🔊	18,524	11,595	1,583	1,907	2,236	3,018	0,172
VAD=1.1	🔊	18,579	11,652	1,579	1,910	2,232	3,022	0,164
VAD=1.25	🔊	18,489	11,636	1,570	1,928	2,209	3,019	0,162
VAD=1.5	🔊	18,225	11,488	1,564	2,030	2,188	3,004	0,168

Tabla 7.52: Resultados utilizando el fichero ESCA con ruido blanco (SNR=18dB).

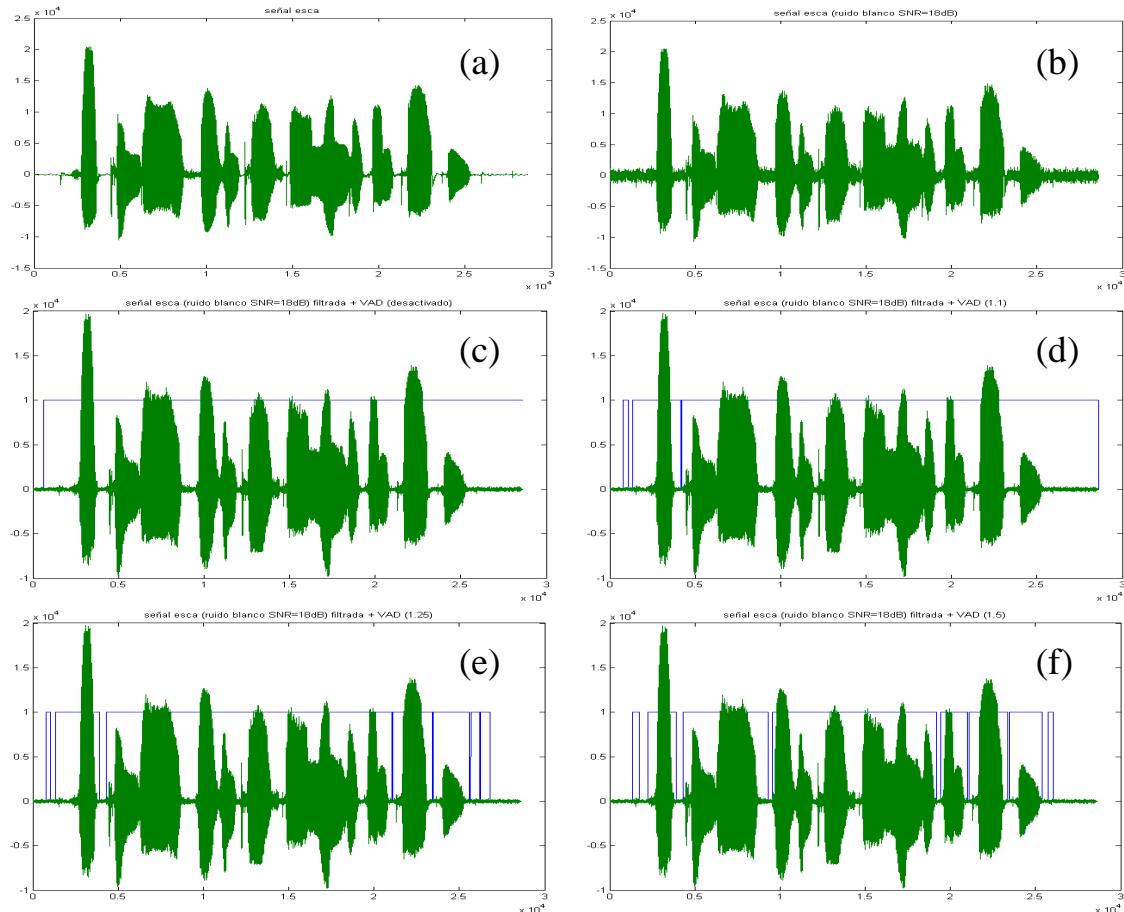


Fig.7.7: (a) Señal ESCA limpia. (b) Señal ESCA con ruido blanco (SNR=18dB). (c) Señal ESCA con ruido blanco (SNR=18dB) filtrada sin activar el VAD. (d) Señal ESCA con ruido blanco (SNR=18dB) filtrada activando el VAD (sobreumbral=1.1). (e) Señal ESCA con ruido blanco (SNR=18dB) filtrada activando el VAD (sobreumbral=1.25). (f) Señal ESCA con ruido blanco (SNR=18dB) filtrada activando el VAD (sobreumbral=1.5).

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN	🔊	154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL	🔊	9,031	2,262	3,195	4,548	5,390	4,659	0,000
SIN VAD	🔊	15,186	8,236	2,072	2,800	2,707	3,248	0,310
VAD=1.1	🔊	15,240	8,359	2,039	2,815	2,672	3,270	0,305
VAD=1.25	🔊	15,241	8,640	2,036	2,893	2,612	3,260	0,307
VAD=1.5	🔊	15,135	8,637	2,064	3,067	2,583	3,270	0,322

Tabla 7.53: Resultados utilizando el fichero ESCA con ruido blanco (SNR=9dB).

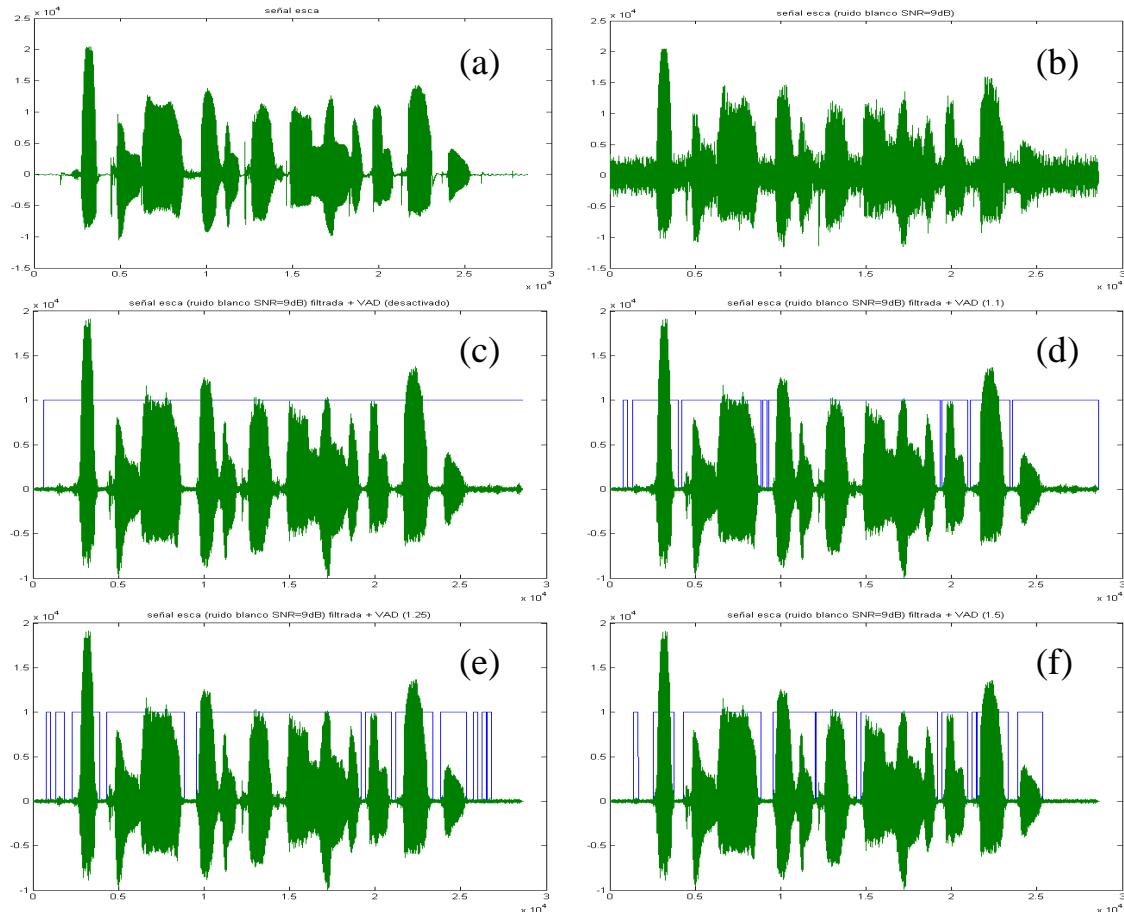


Fig.7.8: (a) Señal ESCA limpia. (b) Señal ESCA con ruido blanco (SNR=9dB). (c) Señal ESCA con ruido blanco (SNR=9dB) filtrada sin activar el VAD. (d) Señal ESCA con ruido blanco (SNR=9dB) filtrada activando el VAD (sobreumbral=1.1). (e) Señal ESCA con ruido blanco (SNR=9dB) filtrada activando el VAD (sobreumbral=1.25). (f) Señal ESCA con ruido blanco (SNR=9dB) filtrada activando el VAD (sobreumbral=1.5).

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN	🔊	154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL	🔊	0,023	-3,563	4,120	6,179	7,146	5,148	0,000
SIN VAD	🔊	10,659	4,115	2,640	3,422	3,452	3,673	0,588
VAD=1.1	🔊	10,838	4,389	2,621	3,443	3,400	3,672	0,561
VAD=1.25	🔊	10,920	5,222	2,625	3,627	3,143	3,689	0,569
VAD=1.5	🔊	10,827	5,292	2,656	3,840	3,054	3,720	0,569

Tabla 7.54: Resultados utilizando el fichero ESCA con ruido blanco (SNR=0dB).

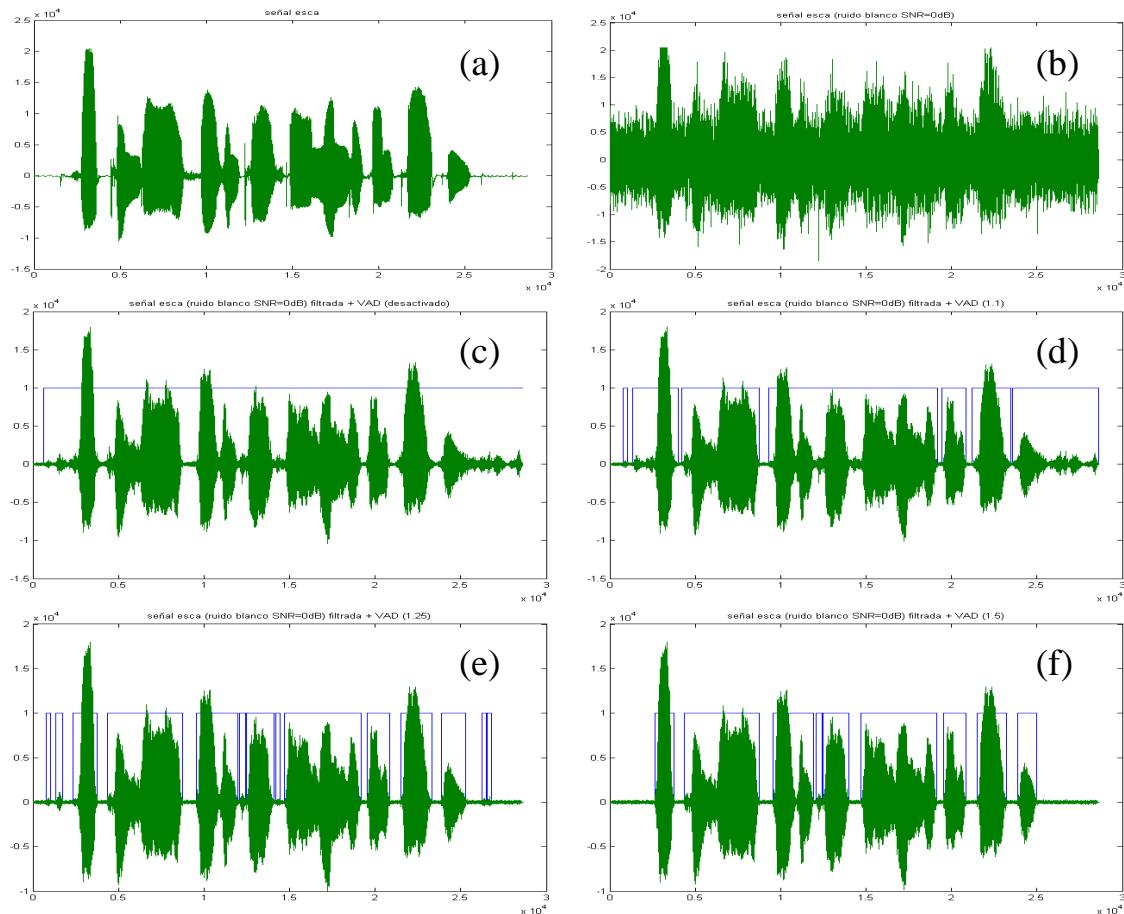


Fig.7.9: (a) Señal ESCA limpia. (b) Señal ESCA con ruido blanco (SNR=0dB). (c) Señal ESCA con ruido blanco (SNR=0dB) filtrada sin activar el VAD. (d) Señal ESCA con ruido blanco (SNR=0dB) filtrada activando el VAD (sobreumbral=1.1). (e) Señal ESCA con ruido blanco (SNR=0dB) filtrada activando el VAD (sobreumbral=1.25). (f) Señal ESCA con ruido blanco (SNR=0dB) filtrada activando el VAD (sobreumbral=1.5).

Las medidas objetivas realizadas en el experimento nos muestran unos resultados dispares, ya que en función de la situación de ruido es mejor utilizar un umbral de VAD más agresivo. A partir de los resultados objetivo obtenidos, el aumento de rendimiento del VAD es muy pobre.

A partir de las pruebas de audición, hemos observado claramente que en las tramas de silencio, en donde el VAD no ha detectado voz, desaparece el ruido musical generado, obteniendo un sonido más agradable en las pausas de voz.

Las graficas obtenidas en el experimento nos confirman lo que ya habíamos comentado, los silencios filtrados con un atenuador proporcionan un ruido residual de fondo homogéneo, eliminando por completo el ruido musical en estas tramas.

Utilizando un sobreumbral de detección de 1,5 o 1,25, se obtienen la mayor homogeneidad de silencio y mayor facilidad de adaptación del sistema a ruidos no estacionarios, eliminando por completo el ruido musical en las tramas de silencio al minimizar las falsas alarmas, el problema de utilizar estos valores es el hecho de que el VAD tiene mayor facilidad de confundir los fonemas sordos de principio de palabra con ruido, sobretodo en condiciones de mucho ruido, atenuándolos al mismo nivel que el ruido.

Un **sobreumbral de detección de VAD igual a 1,1** nos proporciona una peor adaptación a ruido no estacionarios y un número, no despreciable, de falsas alarmas en tramas claramente de silencio, pero asegura la conservación de fonemas sordos al inicio y fin de palabra.

#### **7.2.9.- Modificación del orden de predicción para el Modelado AR del filtro.**

Hasta el momento, en todos los experimentos hemos utilizado de forma invariable un orden de predicción de 10 para el cálculo de espectros con modelado AR. Este orden de predicción viene motivado por el hecho de que un espectro de modelado AR, a diferencia de un espectro de modelado ARMA, carece de ceros, por lo que se opta por aumentar el orden de modelado AR para compensar esta carencia.

Los estudios realizados en anteriores capítulos nos demuestran que los fonemas sonoros se ajustan mejor a un modelo AR de orden 8, en cambio, los fonemas sordos se modelan mejor con un orden 3.

Nuestro sistema es capaz de diferenciar entre fonemas sonoros y sordos gracias a la detección de pitch, otorgándole la capacidad de modificar el orden de predicción AR adecuándolo al tipo de fonema, por otro lado la construcción híbrida del filtro, utilizando la información de un modelado AR y de un filtro peine, hace innecesaria la utilización de un orden 10 para modelado AR, siendo suficiente con un orden 8.

En este experimento observaremos el efecto que produce la modificación adaptativa del orden de predicción de modelado AR en función del tipo de fonema a filtrar. En las pruebas utilizaremos los mismos parámetros del apartado 7.2.8, con un sobreumbral de estimación de VAD igual a 1,1.

Parámetros utilizados:

*Longitud de trama: 256*

*Desplazamiento de trama: 64*

*Multiplicador de muestras de trama: 1 (desactivado)*

*Tipo de ventana (estimación espectros): 2 (Hanning)*

*Tipo de prefiltro: 2 (wiener basado en modelado AR)*

*Parámetro beta del prefiltro: 1.20*

*Parámetro delta del prefiltro: 2*

*Atenuación máxima de prefiltrado: 0.50*

*Tipo de filtro: 1 (wiener basado en modelado AR)*

*Orden del predictor sonoras(modelado): \*\*\* parámetro a optimizar \*\*\**

*Orden del predictor sordas(modelado): \*\*\* parámetro a optimizar \*\*\**

*Orden de los cumulantes (modelado): 1 (AR2 – Levinson Durbin)*

*Sobreumbral pitch: 0.50 (activado)*

*Numero de iteraciones del filtro: 1*

*Factor intertrama (ak): 1.00*

*Factor iteración trama previa: 1*

*Parámetro beta del filtro: 1.20*

*Parámetro delta del filtro: 1.00*

*Atenuación máxima de filtrado: 0.01*

*Nivel de ruido (silencio) en filtrado: 2500.00*

*Tipo de postfiltro: 0 (desactivado)*

*Nivel activación postfiltro: 0.50*

*Orden del filtro de mediana: 1*

*Parámetro sobreestimación ruido (VAD): 1.10 (activado)*

*Numero de tramas de ruido iniciales: 10*

*Parámetro promediado espectro ruido: 0.10*

*Reestimar ruido en tramas silencio: 1 (activado)*

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL		18,079	8,169	1,018	1,483	2,118	2,407	0,000
ORD10-10		<b>19,617</b>	<b>9,818</b>	<b>0,535</b>	1,097	1,101	1,613	0,096
ORD8-3		19,607	9,800	0,539	<b>1,095</b>	<b>1,092</b>	<b>1,603</b>	<b>0,094</b>

Tabla 7.55: Resultados utilizando el fichero ASUN1 con ruido blanco (SNR=18dB).

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL		9,061	1,522	1,745	2,718	3,500	3,192	0,000
ORD10-10		<b>15,654</b>	<b>6,734</b>	0,869	2,109	1,446	2,011	0,147
ORD8-3		15,643	6,732	<b>0,866</b>	<b>2,072</b>	<b>1,436</b>	<b>1,987</b>	<b>0,143</b>

Tabla 7.56: Resultados utilizando el fichero ASUN1 con ruido blanco (SNR=9dB).

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL		0,061	-3,918	2,523	4,266	5,146	3,787	0,000
ORD10-10		10,480	2,996	1,355	2,310	1,894	2,436	<b>0,729</b>
ORD8-3		<b>10,515</b>	<b>3,054</b>	<b>1,326</b>	<b>2,211</b>	<b>1,888</b>	<b>2,381</b>	0,738

Tabla 7.57: Resultados utilizando el fichero ASUN1 con ruido blanco (SNR=0dB).

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL		18,048	9,165	2,220	3,075	3,818	3,961	0,000
ORD10-10		18,579	11,652	1,579	1,910	2,232	3,022	<b>0,164</b>
ORD8-3		<b>18,586</b>	<b>11,660</b>	<b>1,577</b>	<b>1,902</b>	<b>2,229</b>	<b>3,007</b>	<b>0,164</b>

Tabla 7.58: Resultados utilizando el fichero ESCA con ruido blanco (SNR=18dB).

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL		9,031	2,262	3,195	4,548	5,390	4,659	0,000
ORD10-10		15,240	8,359	<b>2,039</b>	2,815	<b>2,672</b>	3,270	0,305
ORD8-3		<b>15,256</b>	<b>8,412</b>	2,040	<b>2,783</b>	<b>2,672</b>	<b>3,248</b>	<b>0,305</b>

Tabla 7.59: Resultados utilizando el fichero ESCA con ruido blanco (SNR=9dB).

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL		0,023	-3,563	4,120	6,179	7,146	5,148	0,000
ORD10-10		10,838	4,389	2,621	3,443	3,400	3,672	0,561
ORD8-3		10,898	4,463	2,603	3,419	3,403	3,643	0,555

Tabla 7.60: Resultados utilizando el fichero ESCA con ruido blanco (SNR=0dB).

Tanto los resultados de medidas objetivas como las pruebas de audición indican que el cambio de orden realizado ayuda ligeramente a aumentar el rendimiento de reducción de ruido, con la ventaja añadida de que al utilizar un menor número de coeficientes de predicción AR el coste computacional es menor.

### 7.2.10.-Parámetro activar postfiltro

Con la activación del postfiltro, bloque basado en filtrado de mediana, pretendemos intentar atenuar el ruido residual y los efectos indeseados que han generados los dos filtrados previos.

En las pruebas utilizaremos un elemento estructurante de tamaño en muestras igual a 3 (orden 1), 5 (orden 2) y 7 (orden 3), utilizando los mismos parámetros del apartado 7.2.9, con ordenes del predictor iguales a 8-3.

Parámetros utilizados:

*Longitud de trama: 256*

*Desplazamiento de trama: 64*

*Multiplicador de muestras de trama: 1 (desactivado)*

*Tipo de ventana (estimación espectros): 2 (Hanning)*

*Tipo de prefiltro: 2 (wiener basado en modelado AR)*

*Parámetro beta del prefiltro: 1.20*

*Parámetro delta del prefiltro: 2*

*Atenuación máxima de prefiltrado: 0.50*

*Tipo de filtro: 1 (wiener basado en modelado AR)*

*Orden del predictor sonoras (modelado): 8*

*Orden del predictor sordas (modelado): 3*

*Orden de los cumulantes (modelado): 1 (AR2 – Levinson Durbin)*

*Sobreumbral pitch: 0.50 (activado)*

*Numero de iteraciones del filtro: 1*

*Factor intertrama (ak): 1.00*

*Factor iteración trama previa: 1*

*Parámetro beta del filtro: 1.20*

*Parámetro delta del filtro: 1.00*

*Atenuación máxima de filtrado: 0.01*

*Nivel de ruido (silencio) en filtrado: 2500.00*

*Tipo de postfiltro: \*\*\* parámetro a optimizar \*\*\**

*Nivel activación postfiltro: 0.50*

*Orden del filtro de mediana: \*\*\* parámetro a optimizar \*\*\**

*Parámetro sobreestimación ruido (VAD): 1.10 (activado)*

*Numero de tramas de ruido iniciales: 10*

*Parámetro promediado espectro ruido: 0.10*

*Reestimar ruido en tramas silencio: 1 (activado)*

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL		18,079	8,169	1,018	1,483	2,118	2,407	0,000
SIN POST		<b>19,607</b>	<b>9,800</b>	<b>0,539</b>	<b>1,095</b>	<b>1,092</b>	<b>1,603</b>	0,094
POSTF=1		18,483	9,282	0,632	1,771	1,171	1,814	<b>-0,031</b>
POSTF=2		14,065	7,339	0,965	2,436	1,532	2,234	0,051
POSTF=3		10,162	5,272	1,151	2,939	1,701	2,420	0,177

Tabla 7.61: Resultados utilizando el fichero ASUN1 con ruido blanco (SNR=18dB).

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL		9,061	1,522	1,745	2,718	3,500	3,192	0,000
SIN POST		<b>15,643</b>	<b>6,732</b>	<b>0,866</b>	<b>2,072</b>	1,436	<b>1,987</b>	0,143
POSTF=1		15,472	6,727	<b>0,845</b>	2,579	<b>1,395</b>	2,015	<b>0,065</b>
POSTF=2		12,495	5,470	1,075	2,822	1,604	2,307	0,354
POSTF=3		10,255	4,620	1,262	3,247	1,755	2,501	0,114

Tabla 7.62: Resultados utilizando el fichero ASUN1 con ruido blanco (SNR=9dB).

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,035	34,708	0,000	0,000	0,000	0,000	0,000
ORIGINAL		0,061	-3,918	2,523	4,266	5,146	3,787	0,000
SIN POST		<b>10,515</b>	3,054	1,326	<b>2,211</b>	1,888	2,381	0,738
POSTF=1		10,499	<b>3,150</b>	<b>1,172</b>	2,403	<b>1,680</b>	<b>2,226</b>	<b>0,667</b>
POSTF=2		8,518	2,201	1,311	2,753	1,757	2,414	1,142
POSTF=3		8,584	2,490	1,370	3,199	1,784	2,541	0,696

Tabla 7.63: Resultados utilizando el fichero ASUN1 con ruido blanco (SNR=0dB).

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL		18,048	9,165	2,220	3,075	3,818	3,961	0,000
SIN POST		<b>18,586</b>	<b>11,660</b>	1,577	<b>1,902</b>	2,229	3,007	<b>0,164</b>
POSTF=1		17,820	11,250	<b>1,355</b>	2,036	<b>2,116</b>	<b>2,737</b>	0,225
POSTF=2		13,862	9,454	1,559	2,582	2,267	2,903	0,278
POSTF=3		9,853	7,588	1,568	2,610	2,379	2,894	0,403

Tabla 7.64: Resultados utilizando el fichero ESCA con ruido blanco (SNR=18dB).

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL		9,031	2,262	3,195	4,548	5,390	4,659	0,000
SIN POST		<b>15,256</b>	<b>8,412</b>	2,040	2,783	2,672	3,248	0,305
POSTF=1		15,066	8,396	<b>1,678</b>	<b>2,639</b>	<b>2,442</b>	<b>2,919</b>	<b>0,256</b>
POSTF=2		12,687	7,328	1,737	2,929	2,515	3,031	0,385
POSTF=3		9,451	5,979	1,724	2,911	2,605	3,044	0,535

Tabla 7.65: Resultados utilizando el fichero ESCA con ruido blanco (SNR=9dB).

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFEREN		154,448	35,000	0,000	0,000	0,000	0,000	0,000
ORIGINAL		0,023	-3,563	4,120	6,179	7,146	5,148	0,000
SIN POST		10,898	4,463	2,603	3,419	3,403	3,643	0,555
POSTF=1		<b>11,025</b>	<b>4,695</b>	2,056	<b>3,068</b>	3,035	<b>3,133</b>	<b>0,515</b>
POSTF=2		9,943	4,342	1,976	3,181	<b>2,927</b>	3,152	0,677
POSTF=3		7,993	3,575	<b>1,959</b>	3,144	2,975	3,218	0,921

Tabla 7.66: Resultados utilizando el fichero ESCA con ruido blanco (SNR=0dB).

Observamos que la activación del postfiltrado reduce el ruido residual y el ruido musical generado por las dos etapas de filtrado anteriores, pero sólo funciona bien en condiciones de mucho ruido, ya que resulta ser un filtro muy agresivo para señales de voz con poco ruido.

### 7.3.-Comparativa con estándar Advance Front-End

En este apartado evaluaremos los resultados obtenidos en nuestra implementación del sistema en comparación con los resultados obtenidos por el estándar de la ETSI, Advance Front-End.

#### 7.3.1.-Comparativa de medidas objetivas

Tal y como hemos visto, el algoritmo de orden 2 se basa en obtener una buena estimación de la densidad espectral de potencia correspondiente a la señal de voz original,  $P_s(\omega)$ . En nuestro sistema de 3 etapas, utilizaremos esta estimación cuando, posteriormente, realicemos un filtrado de Wiener.

Testaremos la calidad de nuestro sistema frente a fuentes perturbadoras consistentes no sólo en ruido aditivo Gaussiano blanco (AWGN) o rosa, sino también con ruidos reales como el que producen diversos tipos de motores o ambientes normales de trabajo.

Los resultados expuestos corresponden a un fichero de voz facilitado por la sociedad "European Speech Communication Association", **ESCA** y a un fichero, **ASUN1**, que forma parte de la base de datos del Departamento de Teoría de Señal y comunicación de la UPC. Evidentemente, los resultados varían para cada fichero de voz, aunque el comportamiento de sus prestaciones suele ser similar para todos los ficheros testeados.

Para simular el margen de distintos niveles de ruido, abarcando desde los entornos más ruidosos hasta los más silenciosos, se ha degradado un mismo fichero de señal de voz con distintos niveles de ruido desde una SNR global (SNR) de -6 dB hasta unos 18 dB. Así, en las siguientes tablas se muestra la evolución de las distintas medidas, temporales y espectrales, en función de la SNR de entrada al sistema y del tipo de ruido perturbador.

En las tablas (7.67) y (7.68) hemos obtenido los valores de todos los parámetros de evaluación para los dos algoritmos, tanto el estándar AURORA, como para nuestro sistema, RERCOM (Reducción de Ruido en COmunicaciones Móviles), cuando se filtran los ficheros ASUN1 y ESCA limpios, es decir sin ningún ruido añadido. Dentro del algoritmo RERCOM, hemos evaluado los resultados para el caso en que sólo se

realice una etapa del procesado, para el caso en que se realicen dos etapas y para el caso en que se realice todo el procesado completo, es decir, las tres etapas.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFERENCIA		154,035	34,708	0,000	0,000	0,000	0,000	0,000
RERCOM2e		<b>58,611</b>	<b>32,357</b>	<b>0,000</b>	<b>0,000</b>	<b>0,013</b>	<b>0,017</b>	<b>-0,000</b>
RERCOM2e+p		<b>58,611</b>	<b>32,357</b>	<b>0,000</b>	<b>0,000</b>	<b>0,013</b>	<b>0,017</b>	<b>-0,000</b>
RERCOM3e+p		19,493	16,275	0,387	0,784	0,810	1,437	0,041
ADVFRONT		25,567	20,520	0,064	0,358	0,295	0,407	0,346

Tabla.7.67 : Evaluación del algoritmo AR2 para el fichero de entrada ASUN1 limpio.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
REFERENCIA		154,448	35,000	0,000	0,000	0,000	0,000	0,000
RERCOM2e		<b>35,852</b>	<b>27,362</b>	<b>0,057</b>	<b>0,080</b>	<b>0,267</b>	<b>0,381</b>	<b>0,002</b>
RERCOM2e+p		<b>35,852</b>	<b>27,362</b>	<b>0,057</b>	<b>0,080</b>	<b>0,267</b>	<b>0,381</b>	<b>0,002</b>
RERCOM3e+p		23,889	21,846	0,688	1,222	1,199	1,737	0,005
ADVFRONT		23,740	19,453	0,363	1,782	0,908	1,135	0,292

Tabla.7.68 : Evaluación del algoritmo AR2 para el fichero de entrada ESCA limpio.

En las tablas anteriores hemos resaltado los mejores valores de cada medida.

Observamos que tanto la primera como la segunda etapa de filtrado del sistema RERCOM son muy poco agresivas, obteniendo a la salida una señal idéntica a la señal de entrada. En cambio, la tercera etapa es algo más agresiva modificando de manera perceptible la señal limpia de entrada.

El estándar AURORA observamos que utiliza un filtrado más o menos igual de agresivo que las tres etapas del sistema RERCOM.

### 7.3.1.1.-Ruidos de banda ancha

En este apartado nos centramos en las perturbaciones provocadas por ruidos de banda ancha, el ruido blanco y el ruido rosa. Realizamos la evaluación para SNR de 18 dB, 9 dB y 0 dB, para los ficheros ASUN1 y ESCA.

#### 7.3.1.1.1.-Ruido blanco

El ruido blanco es un tipo de ruido de banda ancha que posee un espectro plano, esto se traduce en un ruido que tiene la misma energía en todas las frecuencias de la banda de interés, en nuestro caso de 0 a 4 khz.

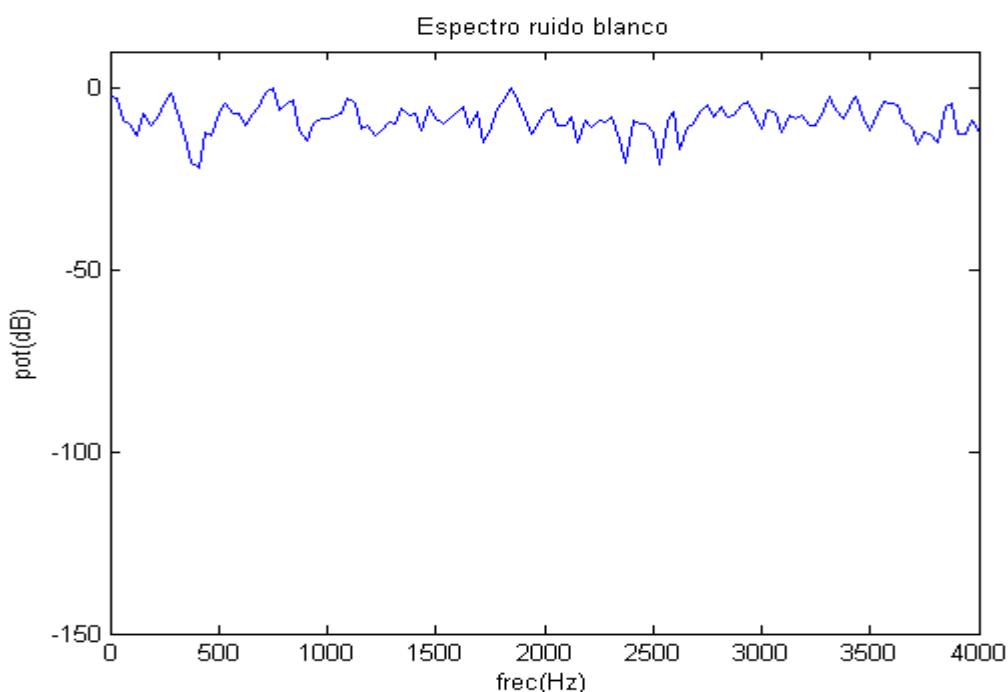


Fig. 7.10: Densidad espectral de energía del ruido blanco

Los resultados obtenidos para el fichero de voz ASUN1 son:

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	🔊	18,079	8,169	1,018	1,483	2,118	2,407	0,000
RERCOM2e	🔊	<b>21,938</b>	<b>11,162</b>	<b>0,504</b>	<b>0,954</b>	<b>1,048</b>	<b>1,491</b>	0,051
RERCOM2e+p	🔊	19,607	9,800	0,539	1,095	1,092	1,603	0,094
RERCOM3e+p	🔊	18,483	9,282	0,632	1,771	1,171	1,814	<b>-0,031</b>
ADVFRONT	🔊	15,506	9,152	0,526	2,202	1,261	1,643	1,419

Tabla.7.69 : Evaluación del algoritmo AR2, con SNR=18 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		9,061	1,522	1,745	2,718	3,500	3,192	0,000
RERCOM2e		<b>16,174</b>	<b>6,906</b>	0,850	<b>1,802</b>	1,450	<b>1,916</b>	0,085
RERCOM2e+p		15,643	6,732	0,866	2,072	1,436	1,987	0,143
RERCOM3e+p		15,472	6,727	<b>0,845</b>	2,579	<b>1,395</b>	2,015	<b>0,065</b>
ADVFRONT		12,939	5,364	1,031	2,262	1,653	2,304	2,388

Tabla.7.70 : Evaluación del algoritmo AR2, con SNR=9 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		0,061	-3,918	2,523	4,266	5,146	3,787	0,000
RERCOM2e		9,427	2,287	1,398	<b>2,176</b>	2,072	2,535	<b>0,647</b>
RERCOM2e+p		<b>10,515</b>	3,054	1,326	2,211	1,888	2,381	0,738
RERCOM3e+p		10,499	<b>3,150</b>	<b>1,172</b>	2,403	<b>1,680</b>	<b>2,226</b>	0,667
ADVFRONT		7,380	1,073	1,843	2,480	2,441	3,118	3,576

Tabla.7.71 : Evaluación del algoritmo AR2, con SNR=0 dB.

Los resultados obtenidos para el fichero de voz ESCA son:

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		18,048	9,165	2,220	3,075	3,818	3,961	0,000
RERCOM2e		<b>19,767</b>	<b>12,581</b>	1,571	<b>1,884</b>	2,221	<b>2,996</b>	<b>0,061</b>
RERCOM2e+p		18,586	11,660	1,577	1,902	2,229	3,007	0,164
RERCOM3e+p		17,820	11,250	1,355	2,036	2,116	2,737	0,225
ADVFRONT		15,744	10,803	<b>1,381</b>	2,650	<b>1,847</b>	2,893	2,067

Tabla.7.72 : Evaluación del algoritmo AR2, con SNR=18 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		9,031	2,262	3,195	4,548	5,390	4,659	0,000
RERCOM2e		<b>15,569</b>	<b>8,475</b>	2,126	2,822	2,797	3,336	<b>0,160</b>
RERCOM2e+p		15,256	8,412	2,040	2,783	2,672	3,248	0,305
RERCOM3e+p		15,066	8,396	<b>1,678</b>	<b>2,639</b>	<b>2,442</b>	<b>2,919</b>	0,256
ADVFRONT		12,573	7,071	2,106	2,891	2,824	3,620	3,068

Tabla.7.73 : Evaluación del algoritmo AR2, con SNR=9 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	🔊	0,023	-3,563	4,120	6,179	7,146	5,148	0,000
RERCOM2e	🔊	9,735	3,588	2,750	3,622	3,659	3,857	<b>0,427</b>
RERCOM2e+p	🔊	10,898	4,463	2,603	3,419	3,403	3,643	0,555
RERCOM3e+p	🔊	<b>11,025</b>	<b>4,695</b>	<b>2,056</b>	<b>3,068</b>	<b>3,035</b>	<b>3,133</b>	0,515
ADVFRONT	🔊	6,986	1,320	3,022	3,707	4,140	4,439	4,308

Tabla.7.74 : Evaluación del algoritmo AR2, con SNR=0 dB.

En las tablas anteriores observamos que el sistema RERCOM obtiene mejores resultados que el sistema ADVANCE frente a ruido blanco, siendo la variante de tres etapas la que mejor se comporta frente a niveles de SNR muy bajos, 0 dB. Las variante de dos etapas sin peine es la que mejor se comporta con niveles bajos de ruido blanco, 18 dB. Y finalmente la variante de dos etapas con peine se comporta mejor en entornos de SNR = 9 dB.

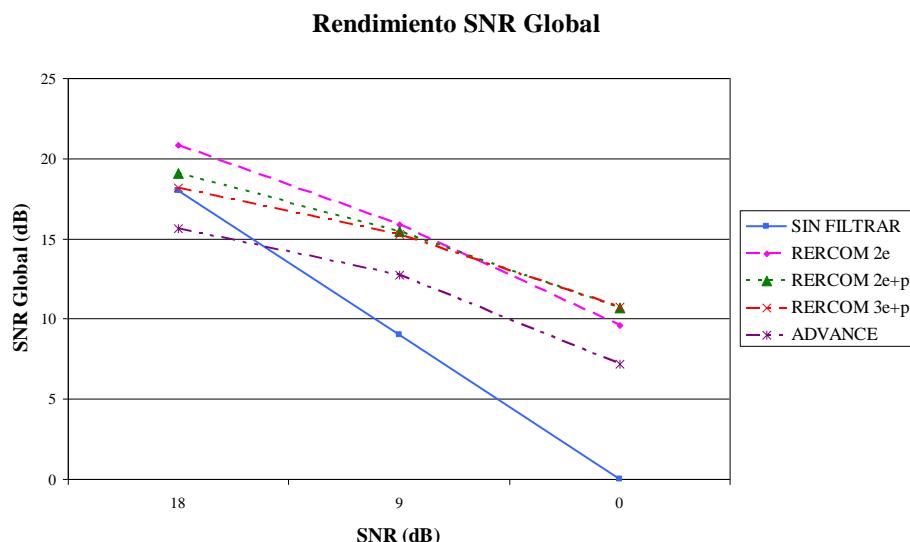


Gráfico 7.1: Promedio de rendimiento de SNR Global para ruido blanco

En el gráfico 7.1 podemos observar que frente a ruido blanco, el sistema RERCOM, en todas sus variantes, supera en una media de unos 3 dB a los resultados de SNR global obtenidos por el sistema ADVANCE. Por otro lado, cabe destacar que el sistema RERCOM, en condiciones de mucho ruido, llega a mejorar el nivel de SNR global de la señal de voz en más de 10 dB.

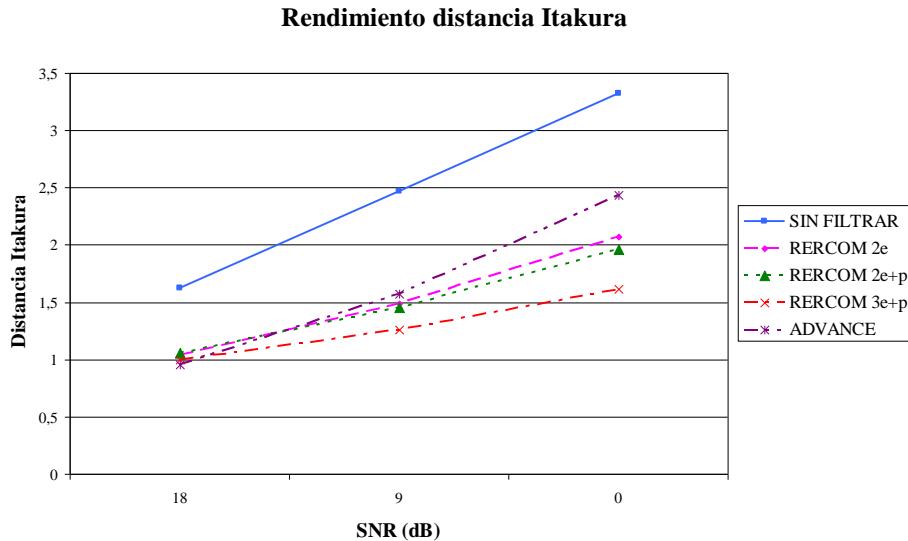


Gráfico 7.2: Promedio de rendimiento de distancia Itakura para ruido blanco

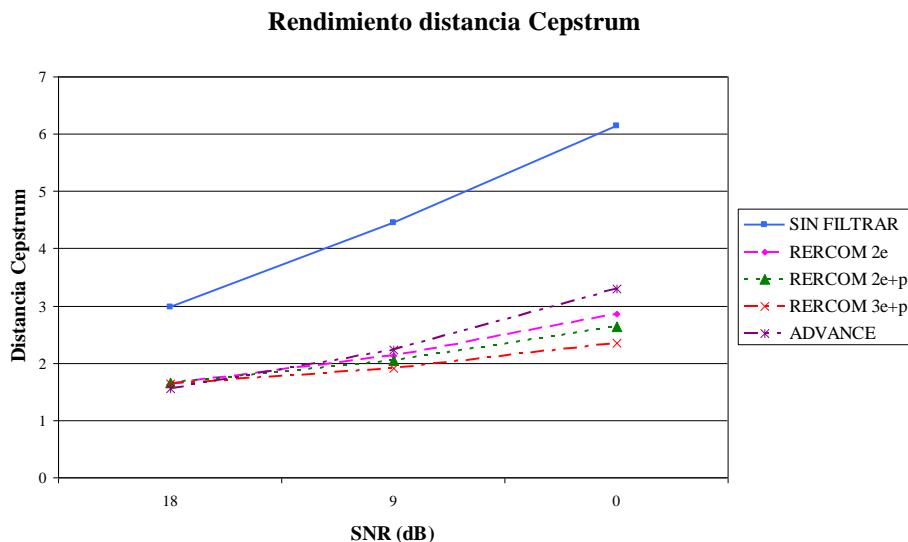


Gráfico 7.3: Promedio de rendimiento de distancia Cepstrum para ruido blanco

Los gráficos 7.2 y 7.3 indican que el sistema RERCOM en su variante de 3etapas + peine obtiene las menores distancias Itakura y Cepstrum, llegando a reducir estas medidas en aproximadamente un 50%. Frente a este tipo de ruido el sistema ADVANCE obtiene los peores resultados, sobretodo en niveles bajos de SNR.

Los resultados obtenidos con ruido blanco sugieren la creación de un algoritmo adaptativo que active o desactive etapas de filtrado en función del nivel de ruido existente en la señal de voz.

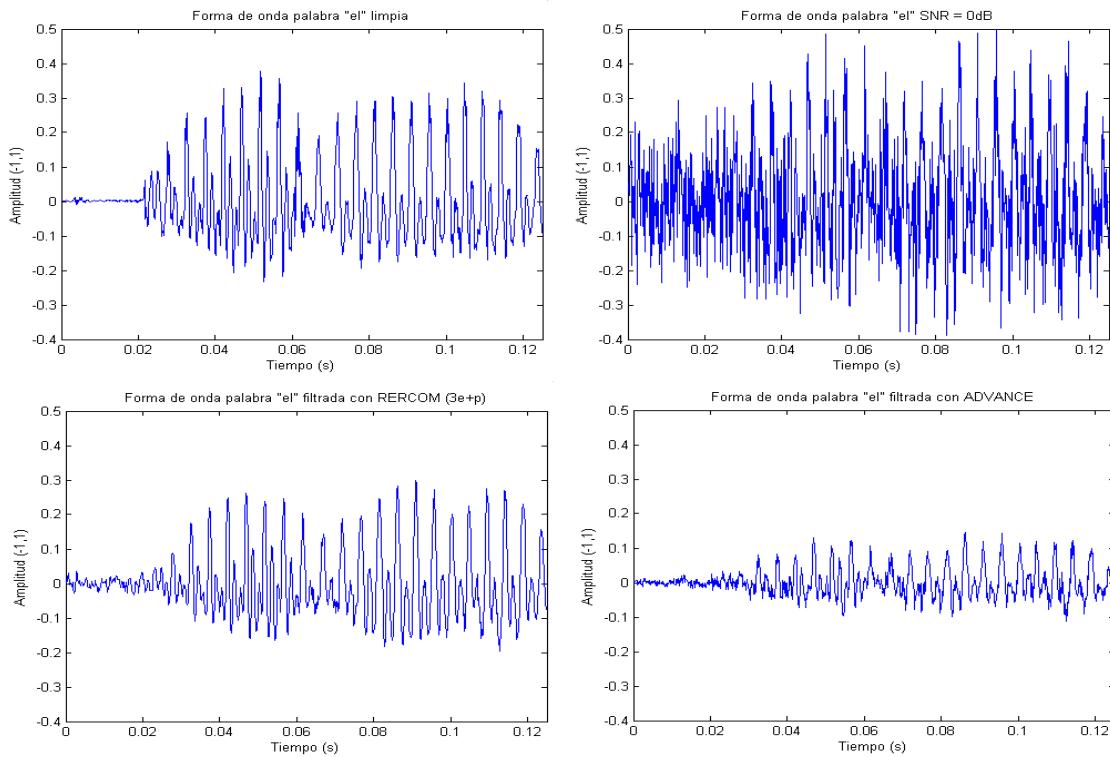


Fig. 7.11: Formas de onda de la palabra “el” del fichero ASUN1; comparativa RERCOM vs ADVANCE frente a ruido blanco con SNR = 0 dB

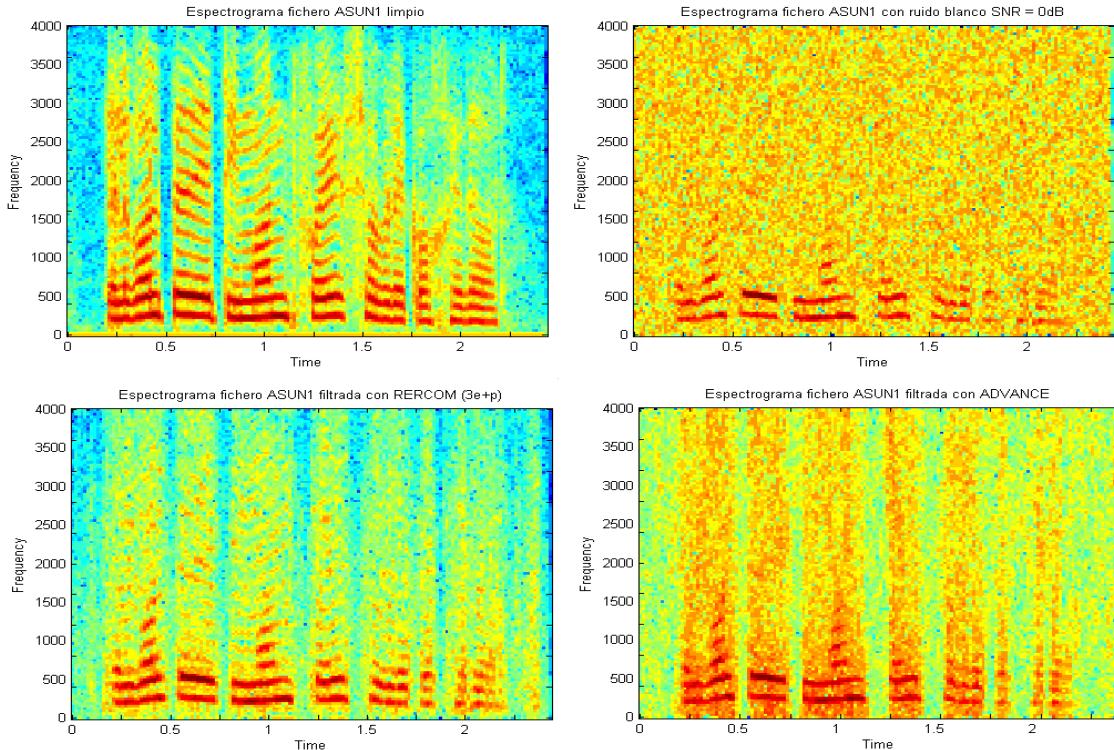


Fig. 7.12: Espectrogramas fichero ASUN1; comparativa RERCOM vs ADVANCE frente a ruido blanco con SNR = 0 dB

### 7.3.1.1.2.-Ruido rosa

El ruido rosa es un tipo de ruido de banda ancha que posee componentes frecuenciales en toda la banda de interés, pero, a diferencia del ruido blanco, su espectro no es plano, sino que tiene un ligero rizado con tendencia paso bajo.

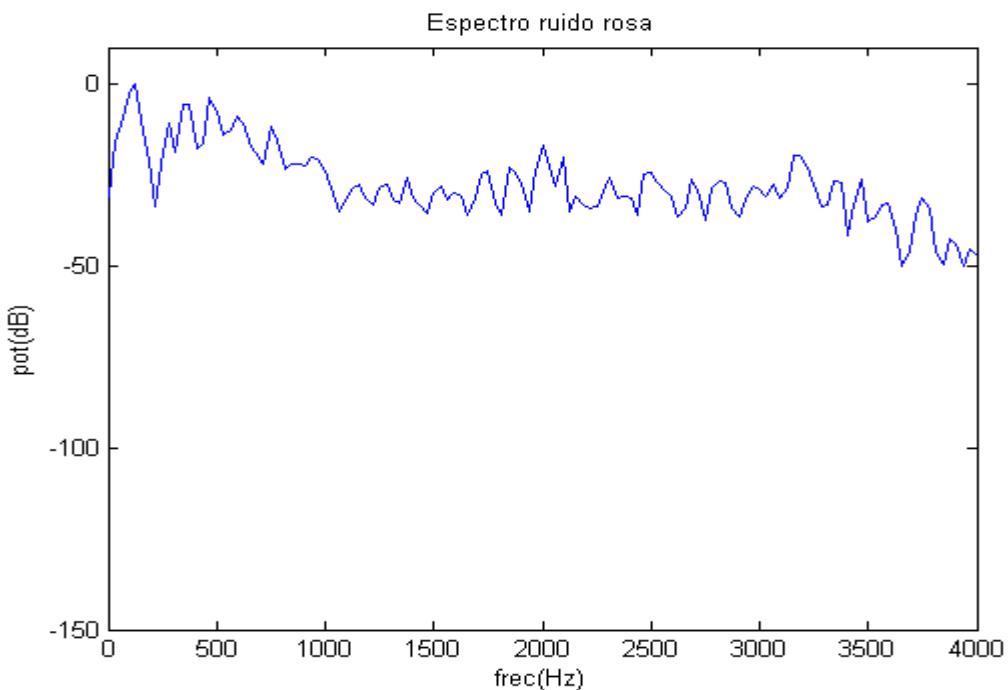


Fig. 7.13: Densidad espectral de energía del ruido rosa

Los resultados obtenidos para el fichero de voz ASUN1 son:

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	🔊	18,104	8,214	0,683	1,133	1,721	1,795	0,000
RERCOM2e	🔊	20,995	10,439	0,379	0,891	0,903	1,204	0,052
RERCOM2e+p	🔊	19,327	9,486	0,413	0,982	0,963	1,331	0,162
RERCOM3e+p	🔊	17,849	8,694	0,585	1,661	1,145	1,805	0,118
ADVFRONT	🔊	14,409	8,234	0,411	2,139	1,171	1,300	1,781

Tabla.7.75 : Evaluación del algoritmo AR2, con SNR=18 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		8,899	1,465	1,309	2,308	3,065	2,586	0,000
RERCOM2e		14,250	5,563	<b>0,679</b>	<b>1,848</b>	<b>1,310</b>	<b>1,743</b>	<b>0,027</b>
RERCOM2e+p		<b>14,549</b>	<b>5,795</b>	0,698	2,127	1,326	1,805	0,276
RERCOM3e+p		14,173	5,597	0,807	2,531	1,436	2,069	0,316
ADVFRONT		9,908	3,448	0,852	2,259	1,454	1,944	2,504

Tabla.7.76 : Evaluación del algoritmo AR2, con SNR=9 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		0,199	-3,828	1,939	3,776	4,648	3,142	0,000
RERCOM2e		7,718	1,210	1,247	2,606	1,760	2,388	<b>0,078</b>
RERCOM2e+p		<b>8,681</b>	<b>1,772</b>	1,183	<b>2,644</b>	1,730	<b>2,262</b>	0,713
RERCOM3e+p		8,582	1,729	<b>1,128</b>	2,888	<b>1,719</b>	2,290	0,770
ADVFRONT		4,991	-0,194	1,625	2,828	2,057	2,779	3,576

Tabla.7.77 : Evaluación del algoritmo AR2, con SNR=0 dB.

Los resultados obtenidos para el fichero de voz ESCA son:

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		17,982	9,171	1,764	2,641	3,365	3,234	0,000
RERCOM2e		<b>18,994</b>	<b>11,789</b>	1,219	<b>1,696</b>	1,903	2,463	<b>0,073</b>
RERCOM2e+p		17,963	11,063	1,237	1,749	1,924	2,490	0,152
RERCOM3e+p		17,459	10,686	1,227	2,052	1,952	2,515	0,159
ADVFRONT		15,811	9,927	<b>1,097</b>	2,550	<b>1,567</b>	<b>2,380</b>	2,033

Tabla.7.78 : Evaluación del algoritmo AR2, con SNR=18 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		9,081	2,358	2,498	4,062	4,897	3,889	0,000
RERCOM2e		14,090	7,287	1,614	<b>2,415</b>	2,401	2,774	<b>0,027</b>
RERCOM2e+p		<b>14,258</b>	<b>7,545</b>	1,579	2,472	2,341	2,727	0,241
RERCOM3e+p		14,076	7,408	<b>1,460</b>	2,541	<b>2,254</b>	<b>2,674</b>	0,258
ADVFRONT		11,716	6,059	1,715	2,629	2,380	3,043	2,874

Tabla.7.79 : Evaluación del algoritmo AR2, con SNR=9 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		-0,020	-3,545	3,254	5,738	6,685	4,366	0,000
RERCOM2e		7,478	1,969	2,157	3,106	3,160	3,385	<b>-0,093</b>
RERCOM2e+p		8,573	2,674	2,010	2,985	3,030	3,173	0,400
RERCOM3e+p		<b>8,651</b>	<b>2,704</b>	<b>1,753</b>	<b>2,888</b>	<b>2,782</b>	<b>2,888</b>	0,474
ADVFRONT		5,283	0,534	2,545	3,395	3,372	3,844	3,420

Tabla 7.80: Evaluación del algoritmo AR2, con SNR=0 dB.

A partir de los resultados obtenidos con ruido rosa podemos extraer las mismas conclusiones que con ruido blanco; el sistema RERCOM es mas robusto, frente a ruidos de banda ancha que el sistema ADVANCE. Además, el número de etapas óptimo a utilizar va en función del nivel de ruido que contamina la señal de voz.

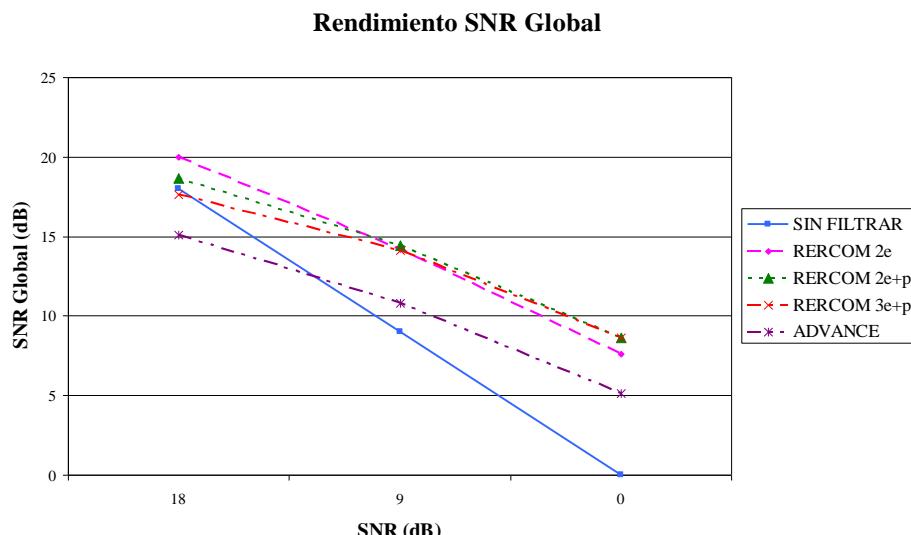


Gráfico 7.4: Promedio de rendimiento de SNR Global para ruido rosa

En el gráfico 7.4 podemos observar que frente a ruido rosa, el sistema RERCOM, en todas sus variantes, supera en una media de unos 3-4 dB a los resultados de SNR global obtenidos por el sistema ADVANCE. También observamos que este tipo de ruido es más difícil de suprimir que el ruido blanco, obteniendo un peor rendimiento de SNR global.

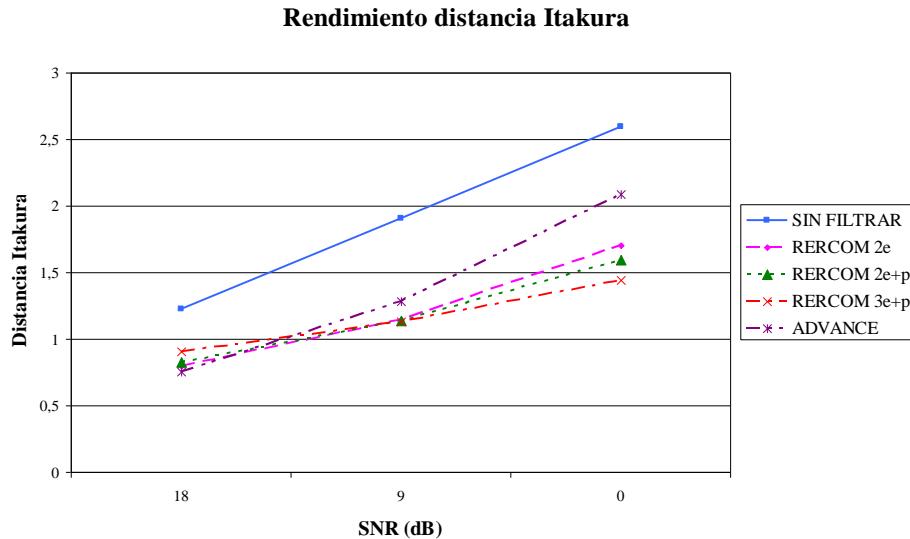


Gráfico 7.5: Promedio de rendimiento de distancia Itakura para ruido rosa

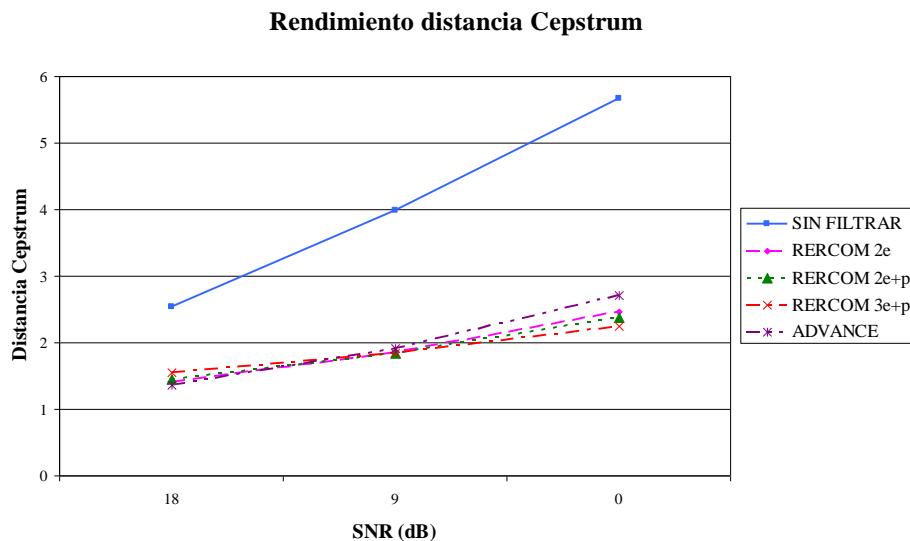


Gráfico 7.6: Promedio de rendimiento de distancia Cepstrum para ruido rosa

Los gráficos 7.5 y 7.6 indican que, al igual que para ruido blanco, el sistema RERCOM en su variante de 3etapas + peine obtiene las menores distancias Itakura y Cepstrum, llegando a reducir la distancia itakura en casi un 50% y la distancia Cepstrum en algo más. Frente a este tipo de ruido el sistema ADVANCE, frente a niveles bajos de SNR, obtiene los peores resultados.

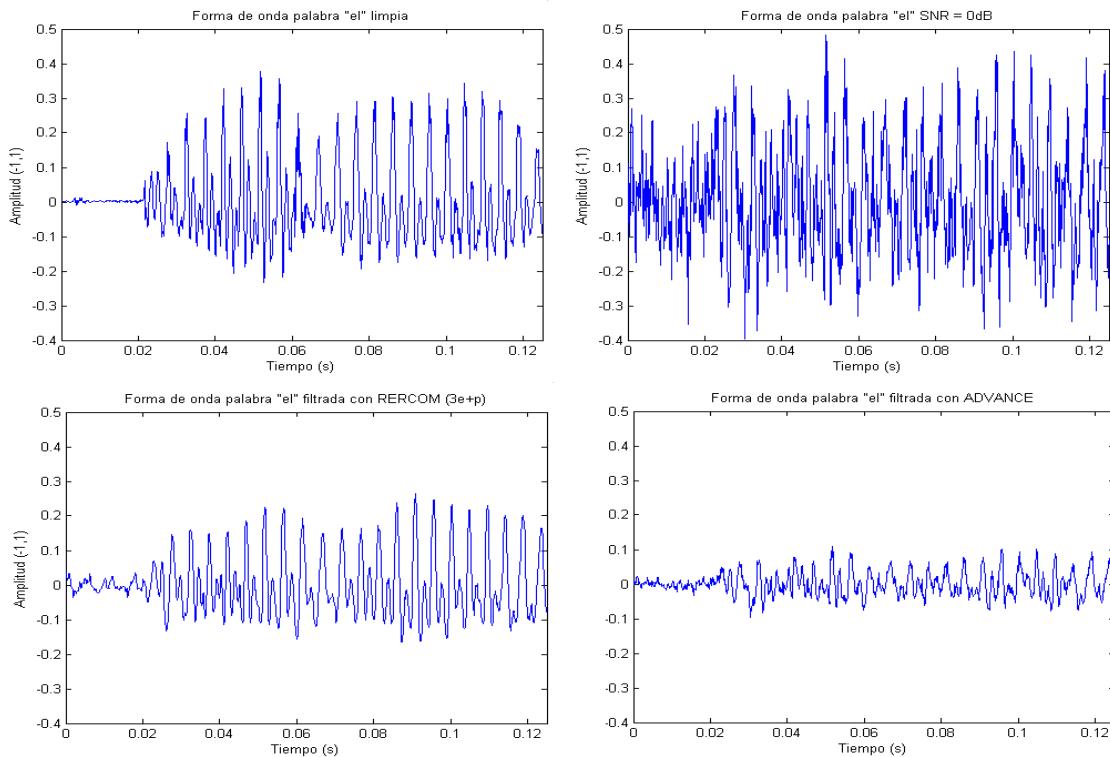


Fig. 7.14: Formas de onda de la palabra “el” del fichero ASUN1; comparativa RERCOM vs ADVANCE frente a ruido rosa con SNR = 0 dB

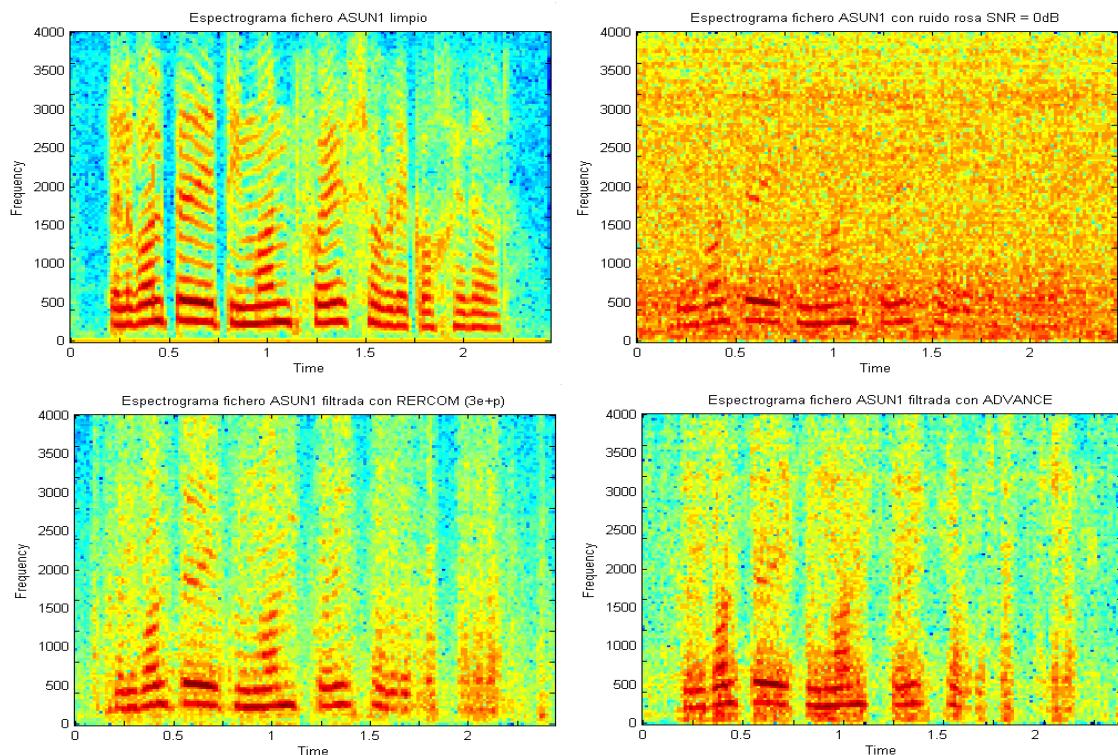


Fig. 7.15: Espectrogramas fichero ASUN1; comparativa RERCOM vs ADVANCE frente a ruido rosa con SNR = 0 dB

### 7.3.1.2.-Ruidos de motor

En este apartado nos centramos en las perturbaciones provocadas por ruidos de banda estrecha, como son los ruidos generados por motores, trataremos 4 casos, el ruido de motor de avión f16, el ruido de un motor genérico y dos ruidos provocados por el motor de un coche, uno de ellos con ruido musical. Aunque parezca incoherente respecto al apartado anterior, realizaremos la evaluación para SNR de 9 dB, 0 dB y -6 dB, ya que los tipos de ruido que analizamos en este apartado son más fáciles de eliminar que los de banda ancha, por lo que un nivel de SNR = 18 dB resulta irrelevante a la hora de obtener conclusiones. También utilizaremos los ficheros de referencia ASUN1 y ESCA.

#### 7.3.1.2.1.-Ruido de motor de avión.

El ruido del motor de un avión f16 es un tipo de ruido que tiene componentes importantes a las frecuencias de 200 Hz, 600 Hz y 2,6 KHz, pero conservando un espectro bastante plano, por lo que resulta un ruido difícil de eliminar.

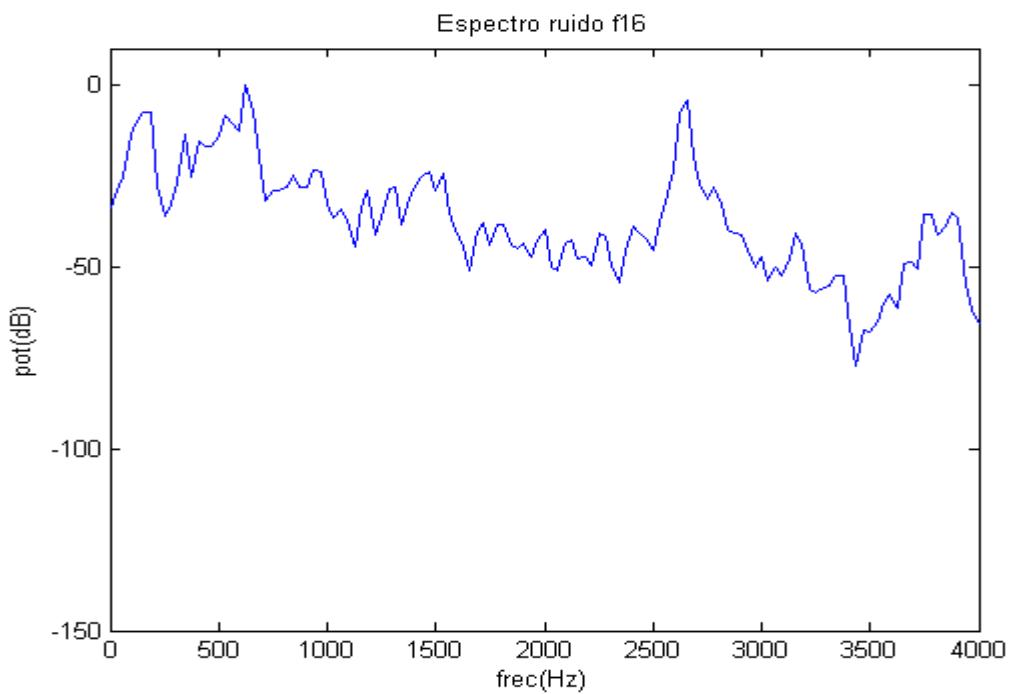


Fig. 7.16: Densidad espectral de energía del ruido de f16

Los resultados obtenidos para el fichero de voz ASUN1 son:

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		9,159	1,664	1,037	1,867	2,576	2,609	0,000
RERCOM2e		14,492	5,751	<b>0,630</b>	<b>1,738</b>	<b>1,208</b>	1,830	<b>-0,007</b>
RERCOM2e+p		<b>14,860</b>	<b>6,057</b>	0,662	1,999	1,259	<b>1,824</b>	0,200
RERCOM3e+p		14,480	5,801	0,799	2,467	1,386	2,099	0,103
ADVFRONT		9,535	3,451	0,689	2,393	1,362	1,973	1,963

Tabla.7.81: Evaluación del algoritmo AR2, con SNR=9 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		0,258	-3,724	1,589	3,284	4,142	3,200	0,000
RERCOM2e		8,163	1,399	1,038	<b>2,407</b>	<b>1,575</b>	2,482	<b>-0,126</b>
RERCOM2e+p		<b>9,296</b>	<b>2,218</b>	<b>1,036</b>	2,607	1,626	<b>2,259</b>	0,247
RERCOM3e+p		9,267	2,177	1,076	2,879	1,615	2,304	0,360
ADVFRONT		4,773	-0,117	1,205	2,669	1,733	2,748	2,443

Tabla.7.82 : Evaluación del algoritmo AR2, con SNR=0 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		-4,542	-6,358	1,865	4,193	5,111	3,401	0,000
RERCOM2e		5,141	-0,568	1,320	<b>3,095</b>	<b>1,804</b>	2,835	-0,320
RERCOM2e+p		5,963	0,003	1,338	3,134	1,917	2,529	<b>0,116</b>
RERCOM3e+p		<b>6,079</b>	<b>0,074</b>	<b>1,259</b>	3,325	1,808	<b>2,374</b>	0,196
ADVFRONT		2,337	-1,785	1,590	3,836	2,050	3,136	2,790

Tabla.7.83: Evaluación del algoritmo AR2, con SNR=-6 dB.

Los resultados obtenidos para el fichero de voz ESCA son:

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		9,069	2,438	2,291	3,629	4,436	3,949	0,000
RERCOM2e		13,450	6,551	1,576	2,329	2,253	3,088	-0,141
RERCOM2e+p		14,005	7,156	1,534	<b>2,283</b>	<b>2,231</b>	<b>2,907</b>	<b>0,116</b>
RERCOM3e+p		<b>14,034</b>	<b>7,183</b>	<b>1,492</b>	2,590	2,242	2,819	0,092
ADVFRONT		11,144	5,694	1,591	2,708	2,145	3,224	2,349

Tabla.7.84: Evaluación del algoritmo AR2, con SNR=9 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	🔊	-0,030	-3,456	3,050	5,271	6,172	4,412	0,000
RERCOM2e	🔊	6,783	1,537	2,300	3,426	2,899	3,639	-0,387
RERCOM2e+p	🔊	7,965	2,341	2,165	3,404	2,905	3,418	0,137
RERCOM3e+p	🔊	8,116	2,487	1,915	3,478	2,728	3,107	0,030
ADVFRONT	🔊	4,983	0,434	2,442	3,751	3,027	3,965	2,639

Tabla.7.85: Evaluación del algoritmo AR2, con SNR=0 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	🔊	-5,928	-6,808	3,429	6,346	7,396	4,577	0,000
RERCOM2e	🔊	2,827	-1,413	2,877	4,305	3,524	4,187	-0,095
RERCOM2e+p	🔊	3,526	-0,934	2,803	4,140	3,784	3,889	0,575
RERCOM3e+p	🔊	3,747	-0,754	2,280	3,643	3,368	3,352	0,535
ADVFRONT	🔊	1,841	-2,053	3,185	5,682	3,717	4,408	2,865

Tabla.7.86 : Evaluación del algoritmo AR2, con SNR=-6 dB.

Para este tipo de ruido los mejores resultados de medidas objetivas se obtienen con el sistema RERCOM de 2 etapas con peine, siendo el de 3 etapas el que mejor se comporta en las peores condiciones de ruido. El sistema ADVANCE obtiene los peores resultados, sobretodo en las medidas de SNR y SNRs provocado principalmente a la excesiva atenuación de la señal de voz que este sistema introduce.

### Rendimiento SNR Global

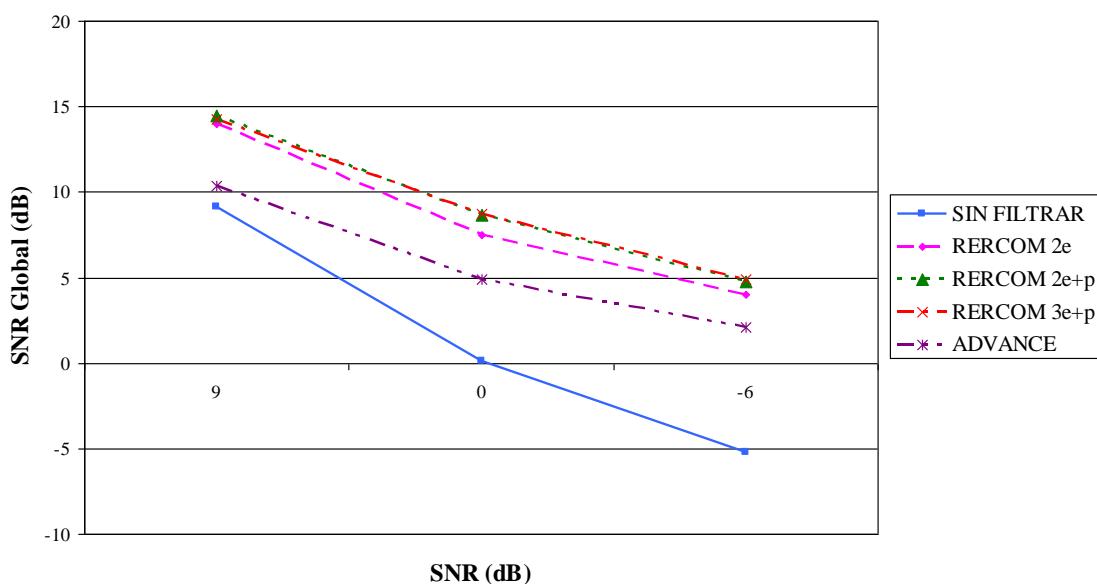


Gráfico 7.7: Promedio de rendimiento de SNR Global para ruido f16

En el gráfico 7.7 podemos observar que frente a ruido de f16, el sistema RERCOM, en todas sus variantes, supera en una media de unos 3-4 dB a los resultados de SNR global obtenidos por el sistema ADVANCE. Por otro lado, cabe destacar que el sistema RERCOM, en condiciones de mucho ruido, llega a mejorar el nivel de SNR global de la señal de voz en algo menos de 10 dB.

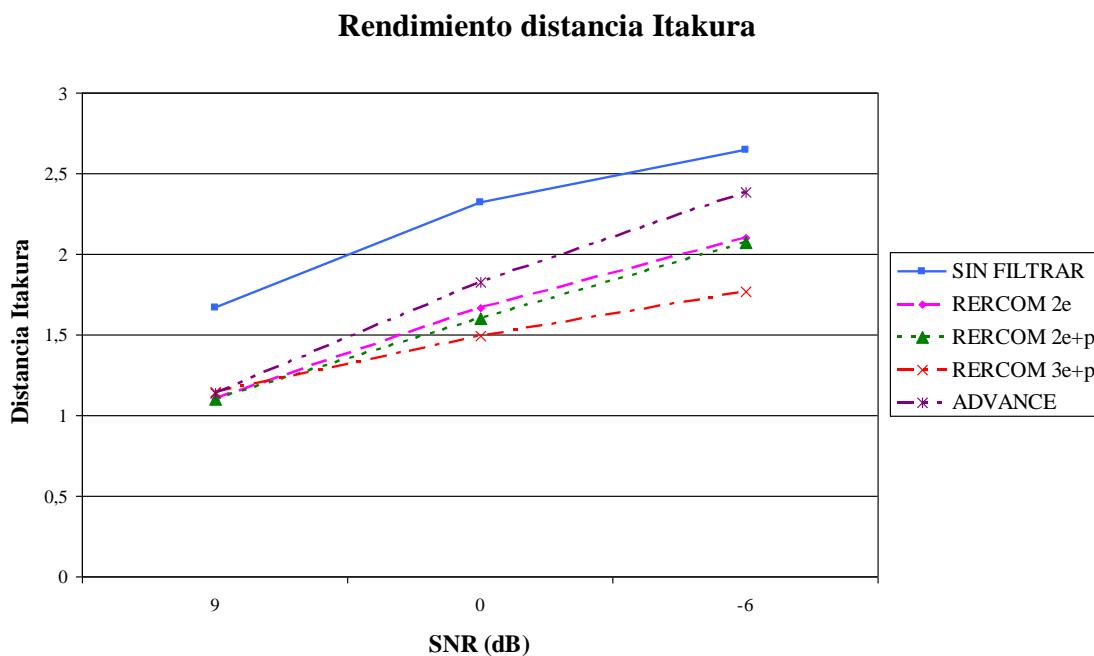


Gráfico 7.8: Promedio de rendimiento de distancia Itakura para ruido f16

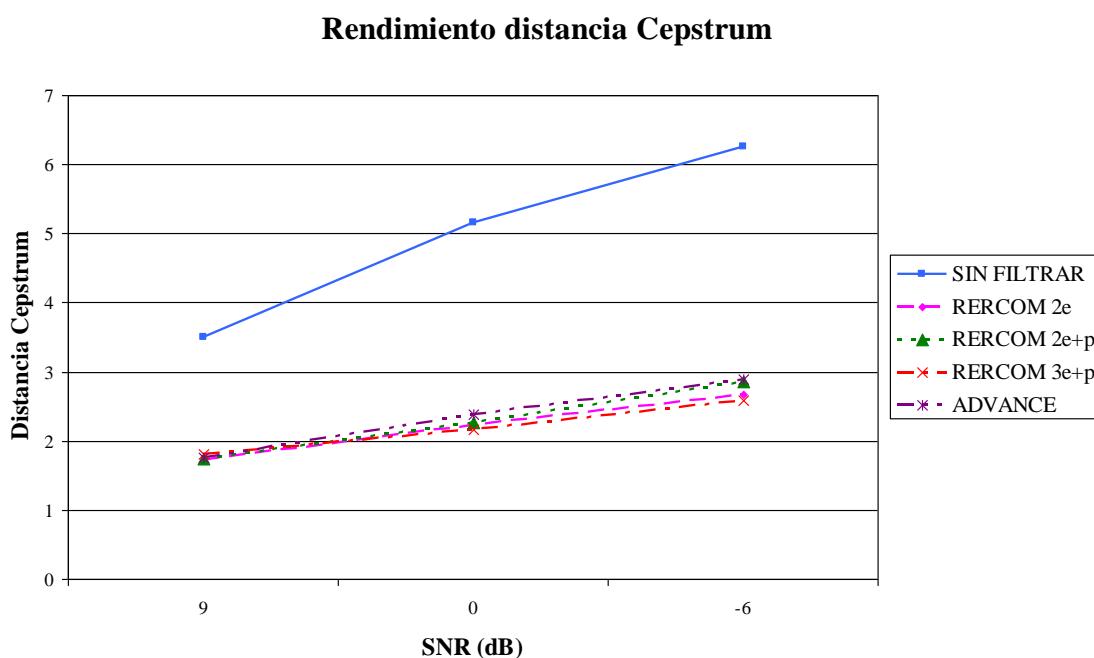


Gráfico 7.9: Promedio de rendimiento de distancia Cepstrum para ruido f16

Los gráficos 7.8 y 7.9 indican que el sistema RERCOM en su variante de 3 etapas + peine obtiene las menores distancias Itakura y Cepstrum, llegando a reducir la medida Itakura en un 30%. Para la medida Cepstrum todos los sistemas analizados se comportan de forma muy similar, obteniendo todos ellos una reducción del 50%.

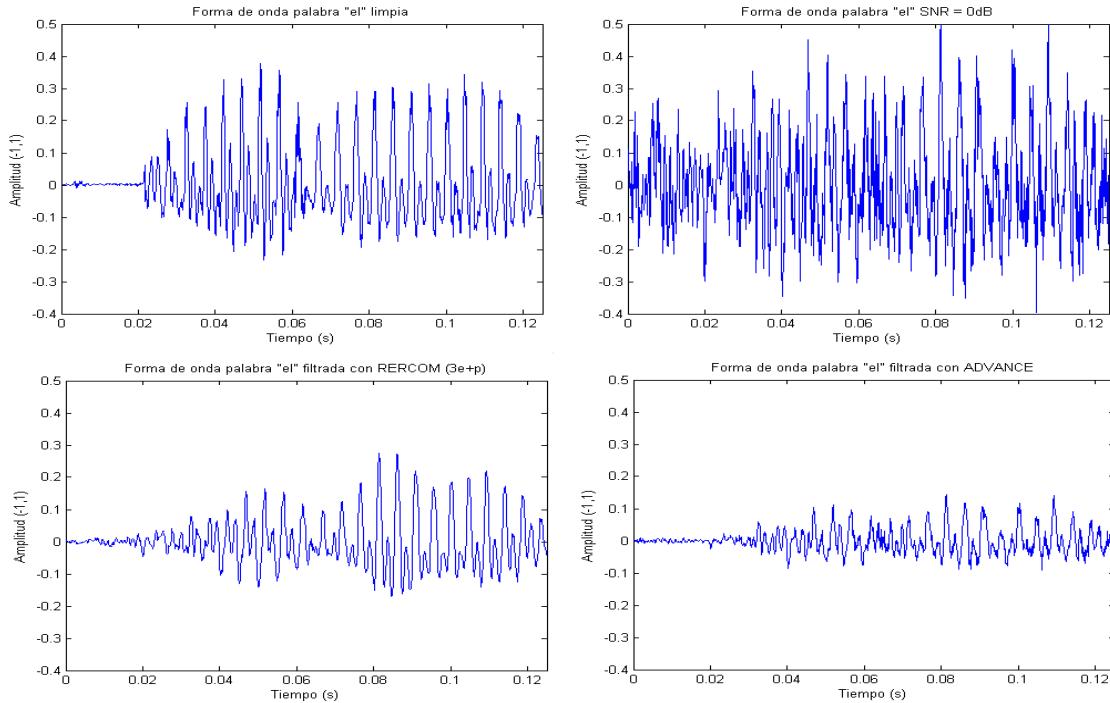


Fig. 7.17: Formas de onda de la palabra “el” del fichero ASUN1; comparativa RERCOM vs ADVANCE frente a ruido avión f16 con SNR = 0 dB

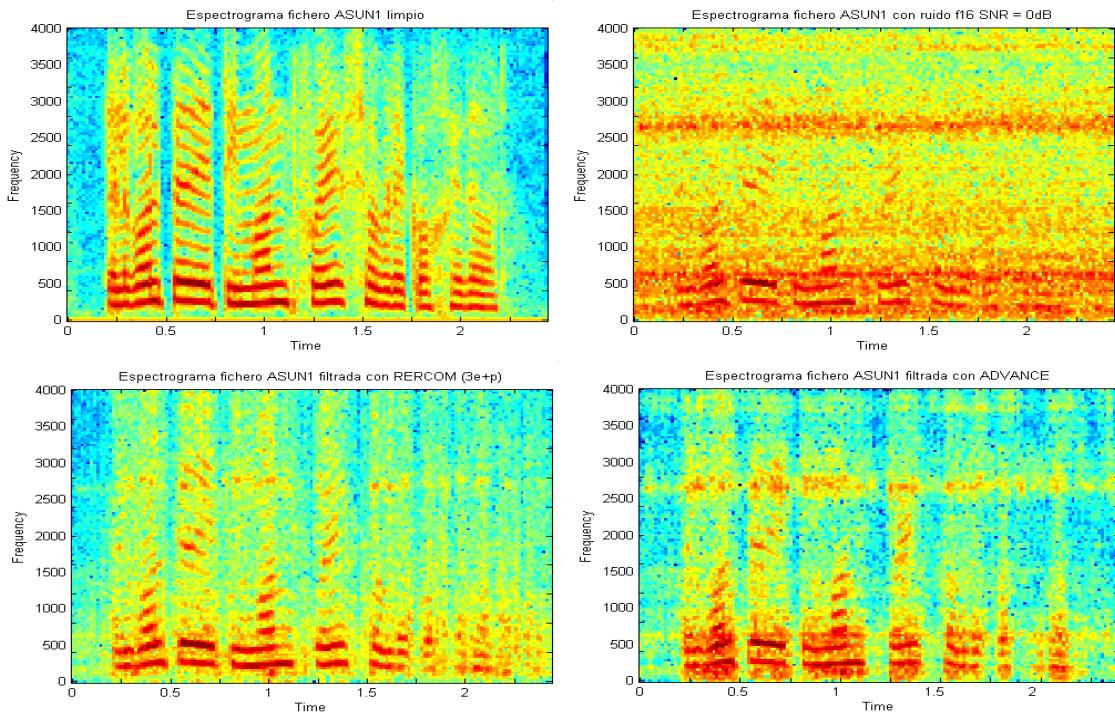


Fig. 7.18: Espectrogramas fichero ASUN1; comparativa RERCOM vs ADVANCE frente a ruido avión f16 rosa con SNR = 0 dB

### 7.3.1.2.2.-Ruido de un motor de explosión.

El ruido de un motor de tipo impulsional lo podríamos identificar con el típico ruido que produciría un generador (motor explosión) de electricidad. Este ruido tiene componentes importantes a las frecuencias de 100-200 Hz y 800 Hz, con tendencia paso bajo.

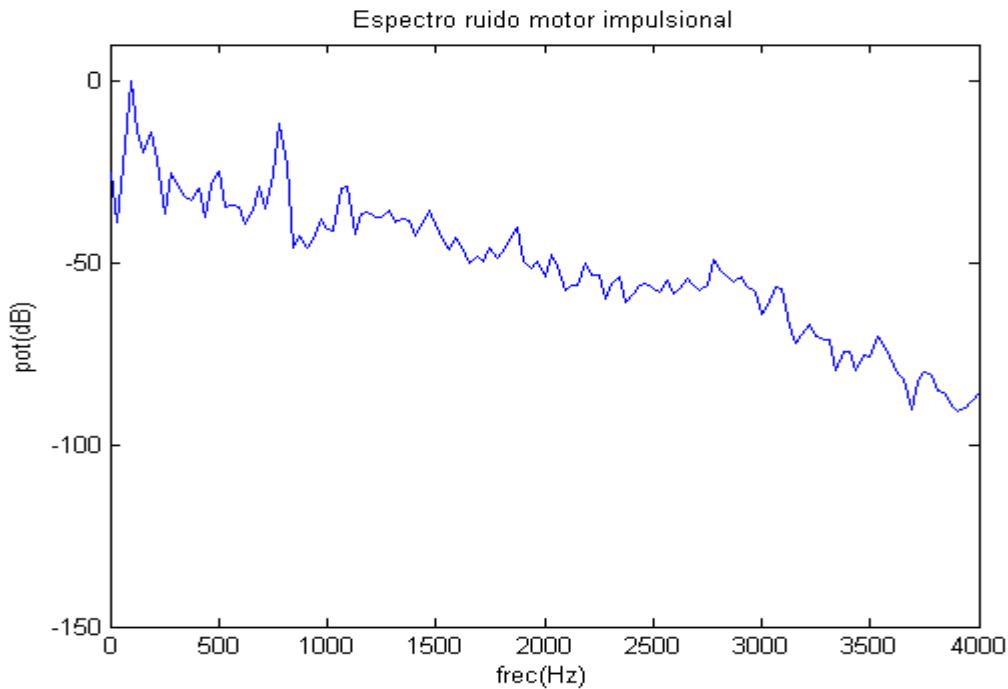


Fig. 7.19: Densidad espectral de energía del ruido de motor impulsional

Los resultados obtenidos para el fichero de voz ASUN1 son:

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	🔊	8,871	1,524	0,574	1,163	1,776	1,534	0,000
RERCOM2e	🔊	14,596	5,748	0,555	2,212	1,109	1,509	0,073
RERCOM2e+p	🔊	15,075	6,075	0,600	2,540	1,202	1,667	0,357
RERCOM3e+p	🔊	14,108	5,475	0,762	2,530	1,347	2,065	0,332
ADVFRONT	🔊	10,018	3,568	0,434	2,186	1,110	1,393	2,403

Tabla.7.87 : Evaluación del algoritmo AR2, con SNR=9 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		-0,029	-3,866	1,001	2,383	3,194	2,058	0,000
RERCOM2e		8,457	1,185	0,828	2,006	1,336	1,877	<b>0,142</b>
RERCOM2e+p		<b>9,420</b>	<b>1,793</b>	0,863	2,361	1,422	2,015	0,625
RERCOM3e+p		9,216	1,667	0,928	2,306	1,450	2,255	0,545
ADVFRONT		4,824	-0,207	<b>0,795</b>	<b>2,000</b>	<b>1,334</b>	<b>1,844</b>	3,376

Tabla.7.88: Evaluación del algoritmo AR2, con SNR=0 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		-5,728	-6,920	1,264	3,406	4,303	2,272	0,000
RERCOM2e		4,477	-1,589	1,033	<b>2,023</b>	1,642	<b>2,127</b>	<b>0,335</b>
RERCOM2e+p		<b>4,793</b>	<b>-1,344</b>	<b>1,026</b>	2,136	1,672	2,181	0,754
RERCOM3e+p		4,690	-1,407	1,075	2,191	1,689	2,395	0,743
ADVFRONT		2,094	-2,015	1,156	2,541	<b>1,571</b>	2,172	4,138

Tabla.7.89: Evaluación del algoritmo AR2, con SNR=-6 dB.

Los resultados obtenidos para el fichero de voz ESCA son:

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		8,854	2,305	1,247	2,612	3,390	2,495	0,000
RERCOM2e		13,186	6,421	1,016	<b>2,012</b>	1,725	2,228	<b>-0,160</b>
RERCOM2e+p		<b>14,213</b>	<b>7,238</b>	1,021	2,159	1,754	2,277	0,274
RERCOM3e+p		13,961	7,018	1,207	2,371	1,980	2,542	0,261
ADVFRONT		10,761	5,679	<b>0,795</b>	2,105	<b>1,399</b>	<b>1,961</b>	2,452

Tabla.7.90 : Evaluación del algoritmo AR2, con SNR=9 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		0,153	-3,347	1,799	4,134	5,020	2,929	0,000
RERCOM2e		6,949	1,381	1,353	<b>2,349</b>	2,275	2,497	<b>-0,344</b>
RERCOM2e+p		<b>8,275</b>	<b>2,308</b>	<b>1,309</b>	2,378	2,304	<b>2,475</b>	0,498
RERCOM3e+p		8,181	2,228	1,420	2,636	2,444	2,668	0,504
ADVFRONT		4,815	0,797	1,346	2,376	<b>2,078</b>	2,517	2,694

Tabla.7.91: Evaluación del algoritmo AR2, con SNR=0 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	🔊	-5,946	-6,840	2,126	5,338	6,289	3,101	0,000
RERCOM2e	🔊	3,106	-1,555	1,676	<b>2,722</b>	2,877	2,758	<b>-0,117</b>
RERCOM2e+p	🔊	<b>3,408</b>	<b>-1,379</b>	<b>1,652</b>	2,729	2,923	<b>2,715</b>	0,368
RERCOM3e+p	🔊	3,370	-1,383	1,680	2,958	2,947	2,870	0,271
ADVFRONT	🔊	1,891	-1,665	1,984	3,978	<b>2,587</b>	2,995	2,804

Tabla 7.92: Evaluación del algoritmo AR2, con SNR=-6 dB.

En las tablas anteriores podemos observar que el sistema RERCOM obtiene los mejores resultados de SNR segmentada y SNR segmentada, pero en este caso el sistema ADVANCE consigue minimizar, en la mayoría de casos, las distancia Itakura y Cepstrum.

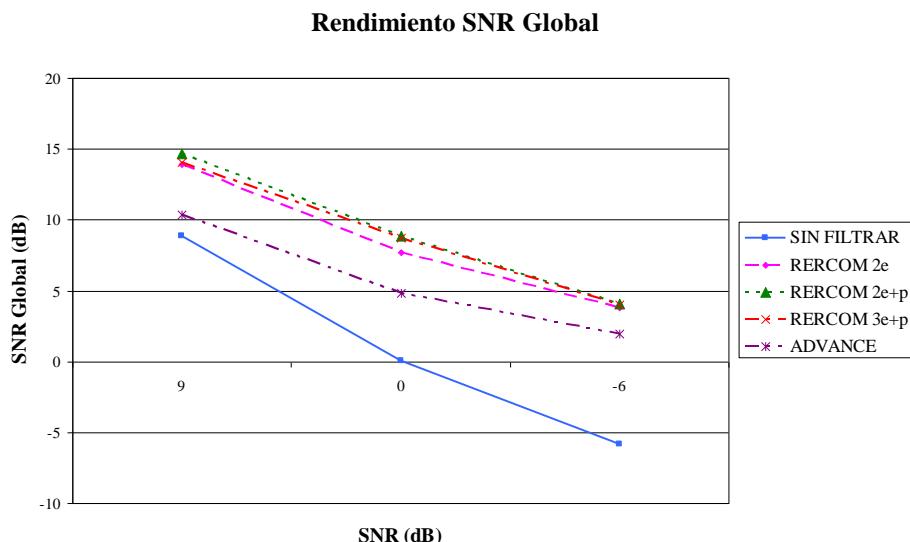


Gráfico 7.10: Promedio de rendimiento de SNR Global para ruido mower

En el gráfico 7.10 podemos observar que frente a ruido de motor de tipo impulsional, el sistema RERCOM, en todas sus variantes, supera en una media de unos 3 dB a los resultados de SNR global obtenidos por el sistema ADVANCE. Pero en este caso la variante de 2 etapas + peine es la que obtiene los mejores resultados.

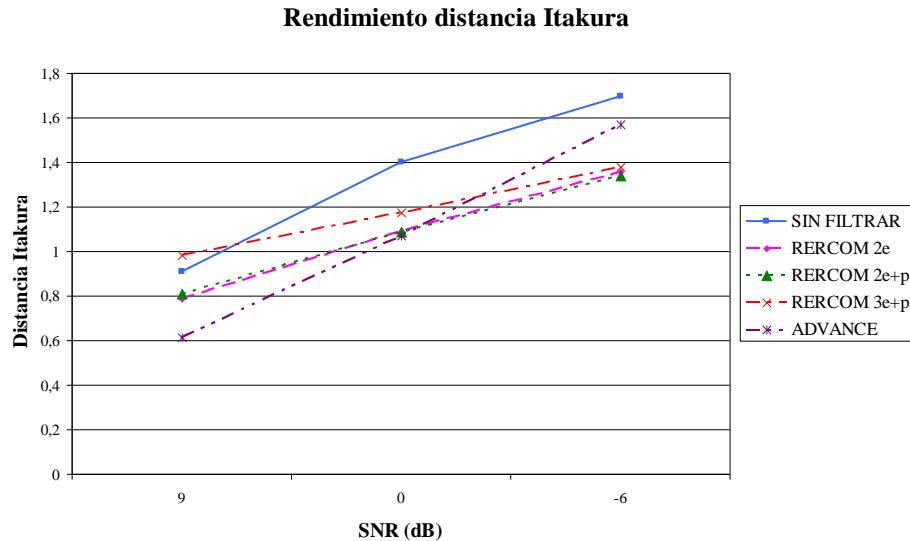


Gráfico 7.11: Promedio de rendimiento de distancia Itakura para ruido mower

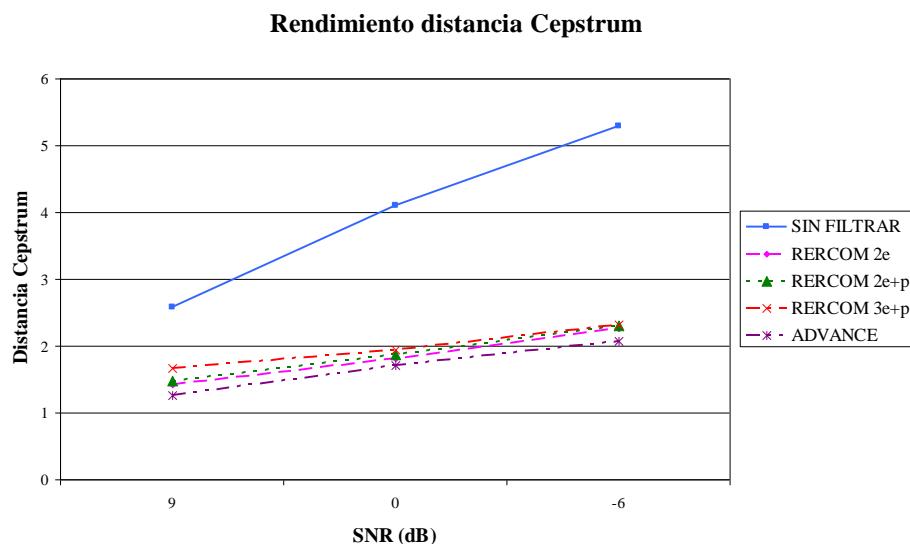


Gráfico 7.12: Promedio de rendimiento de distancia Cepstrum para ruido mower

Los gráficos 7.11 y 7.12 indican que el sistema ADVANCE, a pesar de obtener peores prestaciones en SNR global, consigue los menores valores de distancia Itakura y Cepstrum.

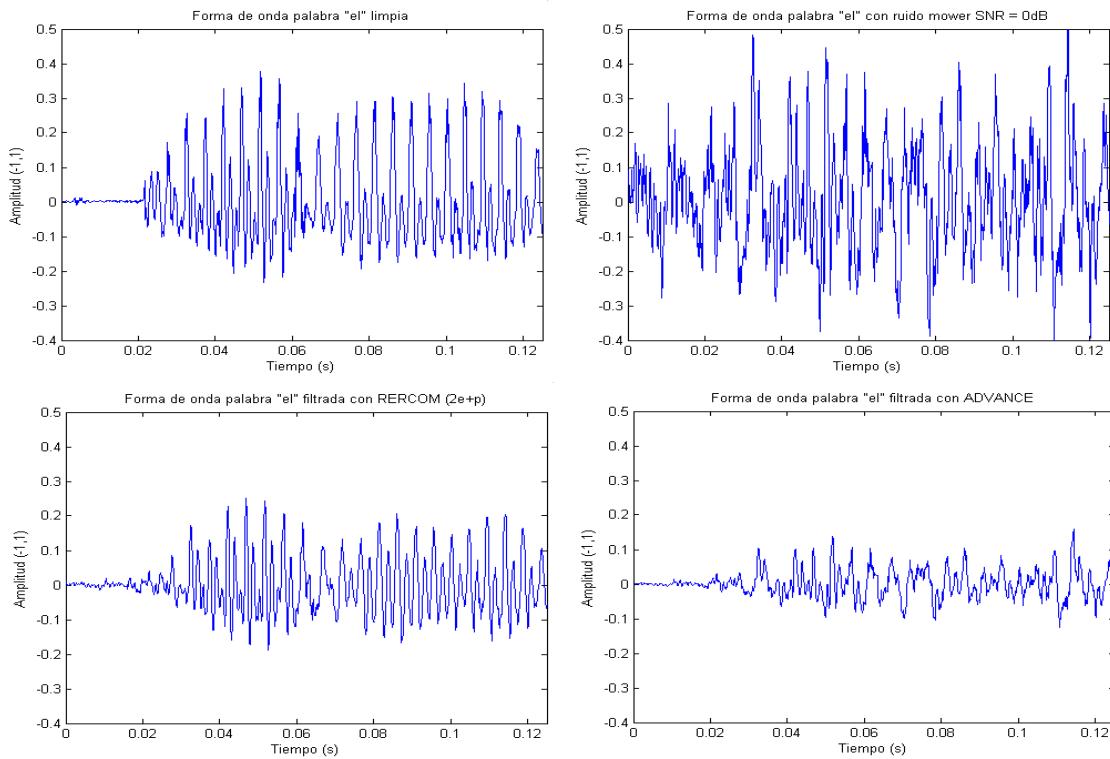


Fig. 7.20: Formas de onda de la palabra “el” del fichero ASUN1; comparativa RERCOM vs ADVANCE frente a ruido motor explosión con SNR = 0 dB

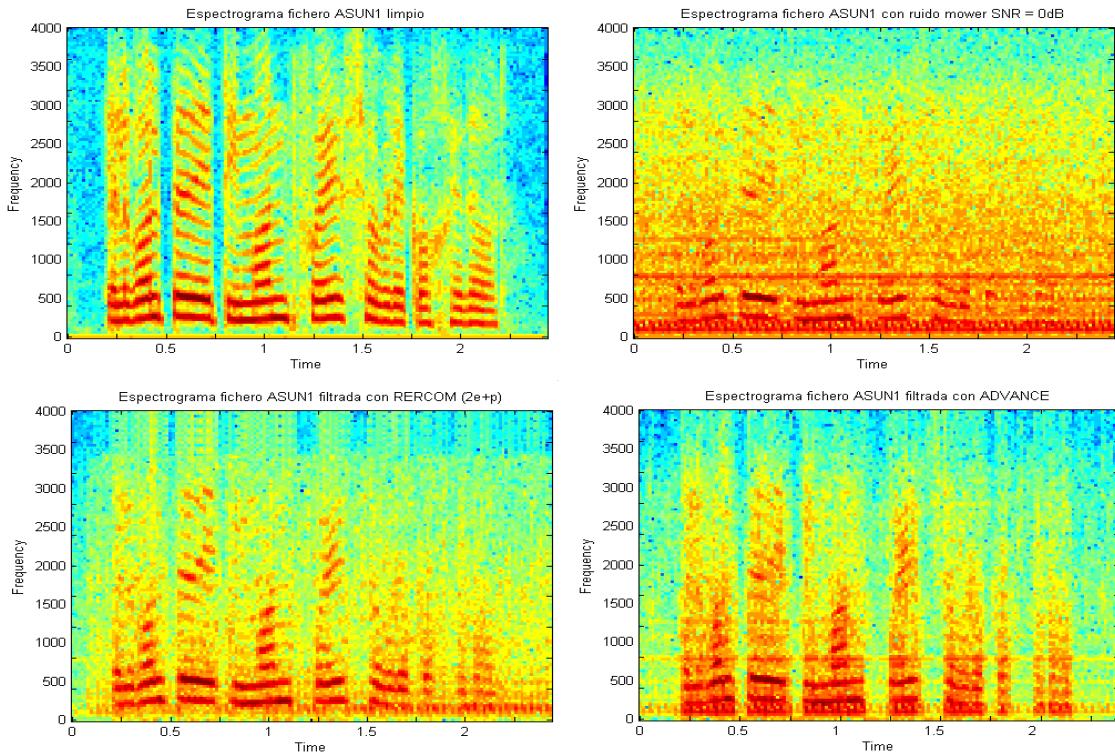


Fig. 7.21: Espectrogramas fichero ASUN1; comparativa RERCOM vs ADVANCE frente a ruido motor explosión con SNR = 0 dB

### 7.3.1.2.3.-Ruido de motor de coche.

El motor de coche produce un ruido con una pronunciada tendencia paso bajo, con sólo componentes importantes alrededor de los 100 Hz. Este tipo de ruido resulta muy sencillo de eliminar, ya que toda su energía se concentra en una banda muy estrecha, dejando la mayoría del espectro de voz intacto.

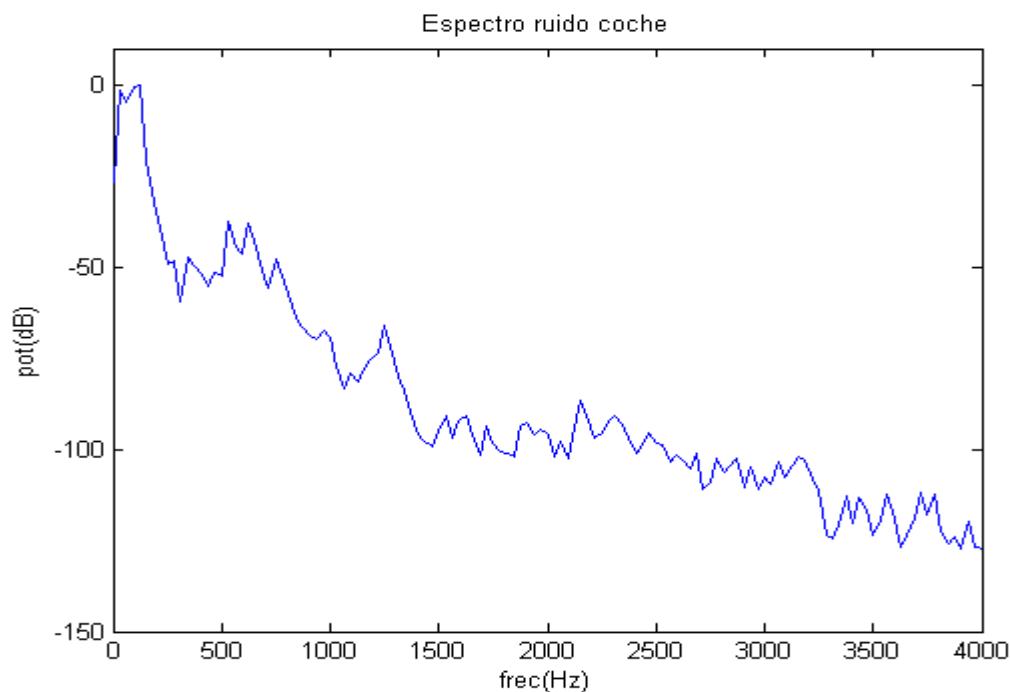


Fig. 7.22: Densidad espectral de energía del ruido de coche

Los resultados obtenidos para el fichero de voz ASUN1 son:

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	🔊	9,548	2,629	0,165	0,212	0,428	1,091	0,000
RERCOM2e	🔊	<b>19,243</b>	<b>9,394</b>	0,415	2,382	1,056	1,476	0,096
RERCOM2e+p	🔊	18,716	8,969	0,496	2,807	1,177	1,641	0,179
RERCOM3e+p	🔊	17,337	8,108	0,608	2,204	1,292	2,037	<b>0,081</b>
ADVFRONT	🔊	10,856	4,213	<b>0,179</b>	<b>0,643</b>	<b>0,598</b>	<b>1,067</b>	0,768

Tabla.7.93: Evaluación del algoritmo AR2, con SNR=9 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		0,347	-3,076	0,379	0,572	0,978	1,834	0,000
RERCOM2e		12,549	4,883	0,532	2,333	<b>1,176</b>	<b>1,680</b>	<b>-0,009</b>
RERCOM2e+p		<b>13,684</b>	<b>5,501</b>	0,580	2,460	1,203	1,719	0,265
RERCOM3e+p		13,042	4,949	0,663	<b>2,030</b>	1,329	2,114	0,139
ADVFRONT		4,517	-0,378	<b>0,388</b>	2,894	1,309	1,717	0,482

Tabla.7.94: Evaluación del algoritmo AR2, con SNR=0 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		-6,053	-6,496	0,619	1,054	1,612	2,426	0,000
RERCOM2e		6,773	1,248	0,693	2,605	<b>1,337</b>	<b>2,023</b>	<b>-0,056</b>
RERCOM2e+p		<b>7,241</b>	<b>1,482</b>	0,712	2,663	1,352	2,040	0,091
RERCOM3e+p		6,970	1,300	0,817	<b>2,544</b>	1,443	2,390	0,145
ADVFRONT		0,949	-2,790	<b>0,642</b>	4,678	1,874	2,420	0,186

Tabla.7.95: Evaluación del algoritmo AR2, con SNR=-6 dB.

Los resultados obtenidos para el fichero de voz ESCA son:

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		9,156	3,164	0,347	0,661	1,089	1,306	0,000
RERCOM2e		16,127	9,241	0,832	2,969	1,531	2,146	<b>-0,077</b>
RERCOM2e+p		<b>16,599</b>	<b>9,726</b>	0,880	2,982	1,555	2,181	0,108
RERCOM3e+p		16,210	9,240	1,130	2,931	1,845	2,534	0,228
ADVFRONT		10,949	5,499	<b>0,429</b>	<b>2,267</b>	<b>1,021</b>	<b>1,398</b>	1,002

Tabla.7.96: Evaluación del algoritmo AR2, con SNR=9 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		0,156	-2,767	0,660	1,493	2,136	2,005	0,000
RERCOM2e		10,503	4,780	1,148	3,807	1,823	2,507	-0,128
RERCOM2e+p		<b>11,366</b>	<b>5,764</b>	1,171	3,827	1,828	2,483	0,064
RERCOM3e+p		11,205	5,668	1,272	3,304	1,978	2,670	0,125
ADVFRONT		3,850	0,078	<b>0,633</b>	<b>2,277</b>	<b>1,309</b>	<b>1,897</b>	1,233

Tabla.7.97 : Evaluación del algoritmo AR2, con SNR=0 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		-5,545	-6,104	0,963	2,257	3,006	2,461	0,000
RERCOM2e		5,493	1,329	1,305	3,783	2,183	2,763	-0,151
RERCOM2e+p		<b>5,962</b>	<b>1,721</b>	1,355	4,041	2,182	2,734	<b>-0,107</b>
RERCOM3e+p		5,847	1,701	1,409	3,609	2,280	2,856	<b>-0,107</b>
ADVFRONT		0,731	-2,106	<b>0,838</b>	<b>3,574</b>	<b>1,744</b>	<b>2,416</b>	0,381

Tabla 7.98 : Evaluación del algoritmo AR2, con SNR=-6 dB.

En las tablas anteriores podemos observar que el sistema RERCOM obtiene los mejores resultados de SNR segmentada y SNR segmentada muy por encima del sistema ADVANCE. Pero, en el caso de medidas de distancias espectrales, el sistema ADVANCE obtiene los mejores valores.

Rendimiento SNR Global

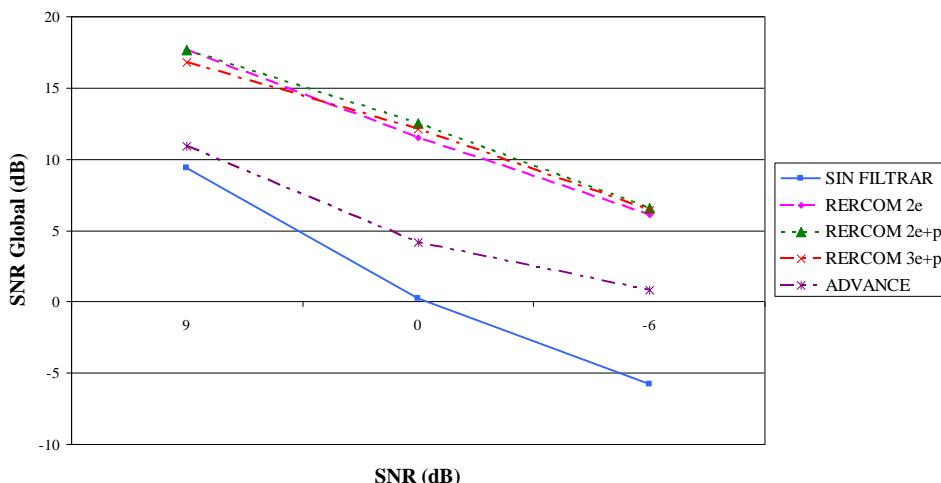


Gráfico 7.13: Promedio de rendimiento de SNR Global para ruido coche

En el gráfico 7.13 podemos observar que frente a ruido de coche, el sistema RERCOM, en todas sus variantes, supera en una media de unos 6 dB a los resultados de SNR global obtenidos por el sistema ADVANCE. La variante del sistema RERCOM de 2 etapas + peine es la que obtiene los mejores resultados mejorando en unos 12 dB el nivel de SNR de la señal de voz contaminada con ruido.

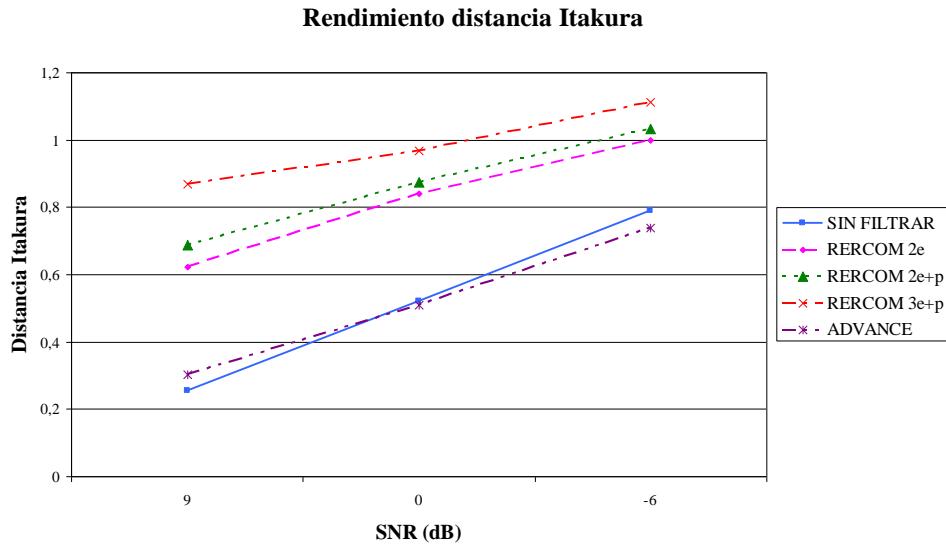


Gráfico 7.14: Promedio de rendimiento de distancia Itakura para ruido de coche

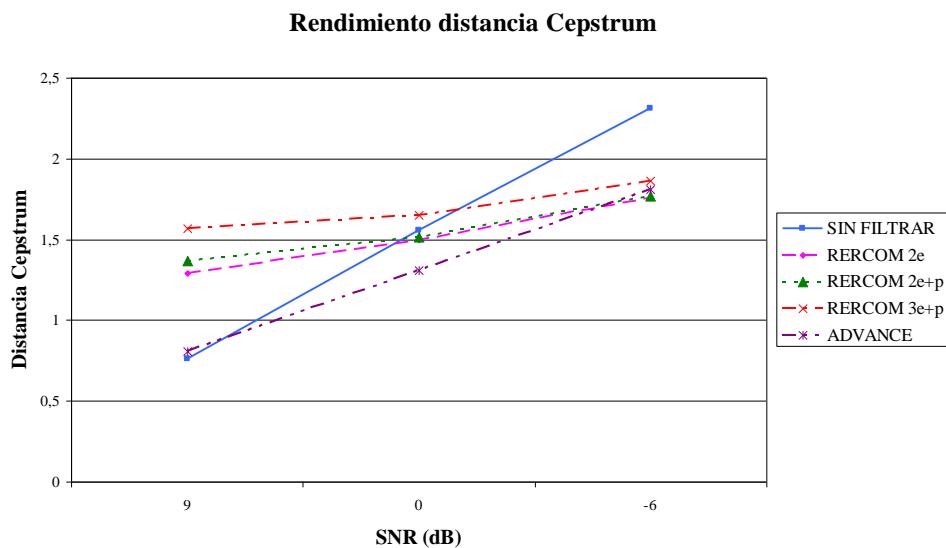


Gráfico 7.15: Promedio de rendimiento de distancia Cepstrum para ruido coche

Los gráficos 7.14 y 7.15 nos muestran unos resultados desconcertantes, a pesar de la evidente mejora de la señal filtrada respecto a su original sin filtrar, los resultados de distancias indican que en algunos casos es mejor dejar la señal de voz contaminada de ruido. En los otros casos, el sistema ADVANCE, sobretodo en distancia Itakura, es el que los mejora ligeramente.

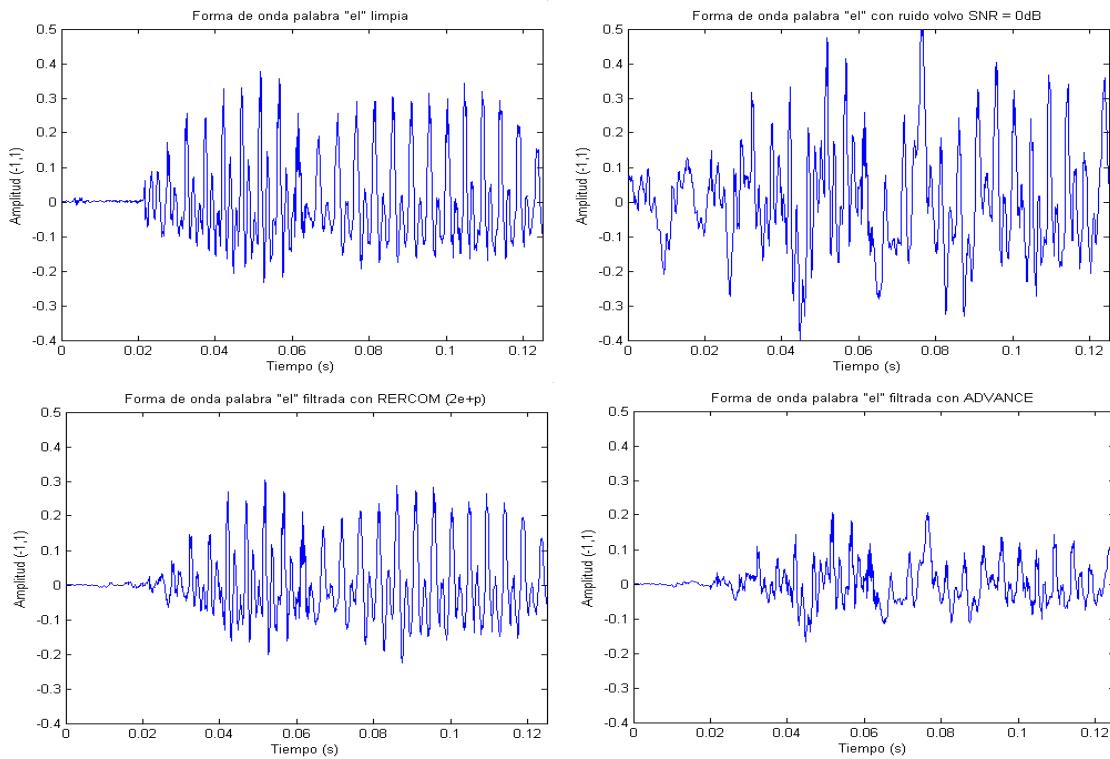


Fig. 7.23: Formas de onda de la palabra “el” del fichero ASUN1; comparativa RERCOM vs ADVANCE frente a ruido volvo con SNR = 0 dB

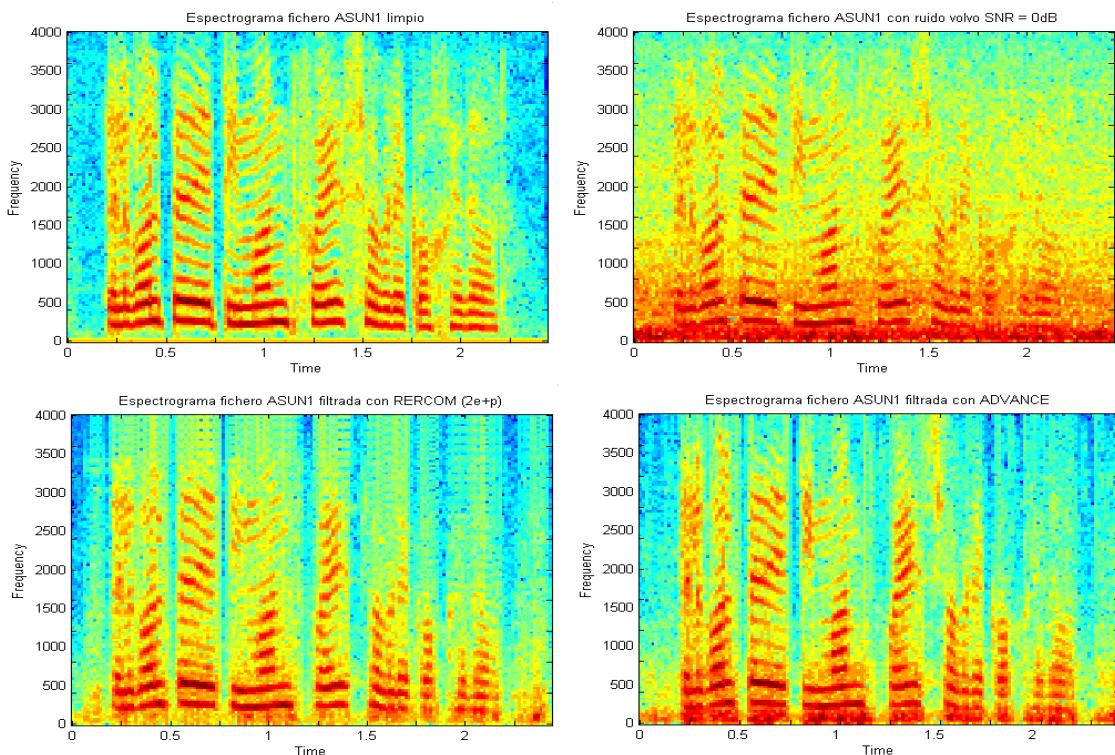


Fig. 7.24: Espectrogramas fichero ASUN1; comparativa RERCOM vs ADVANCE frente a ruido volvo con SNR = 0 dB

### 7.3.1.2.4.-Ruido de motor de coche con fluctuaciones.

Este tipo de ruido es producido por el mismo coche del apartado anterior, pero con la particularidad de que en esta ocasión el coche circula por un arcén en malas condiciones, generando una especie de fluctuación en el nivel de ruido. Este ruido, al igual que el anterior, sólo tiene una componente importante a la frecuencia de 100 Hz.

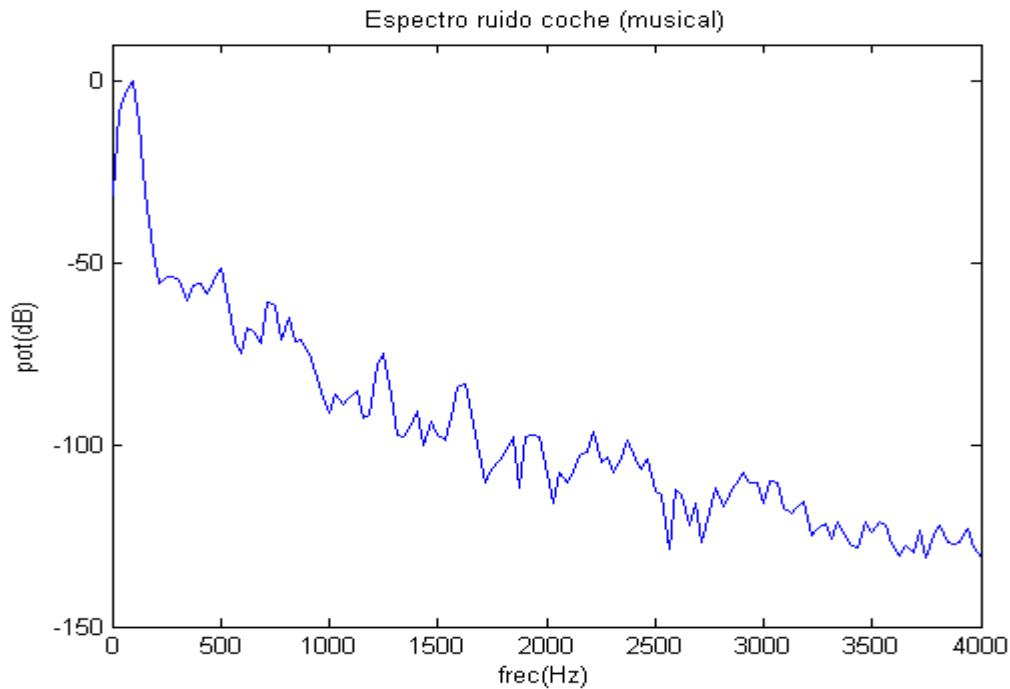


Fig. 7.25: Densidad espectral de energía del ruido de coche con fluctuaciones

Los resultados obtenidos para el fichero de voz ASUN1 son:

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	🔊	9,182	2,396	0,207	0,306	0,582	1,156	0,000
RERCOM2e	🔊	<b>18,759</b>	<b>9,114</b>	0,300	1,553	0,923	<b>1,136</b>	<b>0,040</b>
RERCOM2e+p	🔊	17,473	8,305	0,353	1,756	1,040	1,311	0,169
RERCOM3e+p	🔊	16,257	7,458	0,557	2,070	1,288	1,897	0,190
ADVFRONT	🔊	9,896	4,233	<b>0,238</b>	<b>1,326</b>	<b>0,820</b>	1,170	1,362

Tabla.7.99: Evaluación del algoritmo AR2, con SNR=9 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		-0,018	-3,316	0,462	0,809	1,300	1,885	0,000
RERCOM2e		11,813	4,416	0,555	3,247	1,536	<b>1,755</b>	<b>-0,002</b>
RERCOM2e+p		<b>12,208</b>	<b>4,695</b>	0,593	3,371	1,545	1,830	0,238
RERCOM3e+p		11,792	4,339	0,733	<b>3,051</b>	1,677	2,214	0,321
ADVFRONT		4,809	0,140	<b>0,498</b>	3,759	<b>1,501</b>	1,930	2,849

Tabla.7.100: Evaluación del algoritmo AR2, con SNR=0 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		-5,719	-6,356	0,682	1,361	2,012	2,386	0,000
RERCOM2e		7,890	1,124	<b>0,668</b>	3,009	1,402	<b>1,891</b>	<b>-0,070</b>
RERCOM2e+p		<b>8,136</b>	<b>1,334</b>	0,705	2,946	<b>1,380</b>	1,952	0,135
RERCOM3e+p		8,048	1,281	0,810	<b>2,758</b>	1,455	2,293	0,151
ADVFRONT		1,526	-2,253	0,697	4,840	1,843	2,334	0,687

Tabla.7.101: Evaluación del algoritmo AR2, con SNR=-6 dB.

Los resultados obtenidos para el fichero de voz ESCA son:

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		9,139	3,081	0,413	0,951	1,483	1,535	0,000
RERCOM2e		<b>16,753</b>	<b>9,722</b>	0,690	<b>1,634</b>	1,289	1,874	<b>0,013</b>
RERCOM2e+p		16,453	9,598	0,709	1,698	1,326	1,912	0,201
RERCOM3e+p		16,053	9,129	1,034	2,232	1,705	2,390	0,196
ADVFRONT		10,799	5,480	<b>0,457</b>	2,666	<b>1,227</b>	<b>1,533</b>	1,934

Tabla.7.102: Evaluación del algoritmo AR2, con SNR=9 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		-0,162	-2,986	0,848	2,035	2,751	2,252	0,000
RERCOM2e		10,814	4,619	0,921	2,361	1,599	2,228	<b>-0,040</b>
RERCOM2e+p		<b>11,648</b>	<b>5,290</b>	0,937	<b>2,350</b>	1,609	2,227	0,232
RERCOM3e+p		11,625	5,191	1,173	2,649	1,867	2,561	0,307
ADVFRONT		4,663	0,586	<b>0,688</b>	2,740	<b>1,381</b>	<b>1,992</b>	2,049

Tabla.7.103: Evaluación del algoritmo AR2, con SNR=0 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		-5,951	-6,500	1,165	2,893	3,697	2,572	0,000
RERCOM2e		4,991	0,476	1,161	2,786	2,121	2,480	<b>-0,110</b>
RERCOM2e+p		<b>5,161</b>	<b>0,760</b>	1,166	<b>2,722</b>	2,139	2,462	0,235
RERCOM3e+p		5,127	0,711	1,319	3,011	2,292	2,697	0,158
ADVFRONT		0,412	-2,484	<b>0,979</b>	2,812	<b>1,824</b>	<b>2,432</b>	1,542

Tabla 7.104: Evaluación del algoritmo AR2, con SNR=-6 dB.

En las tablas anteriores podemos observar que el sistema RERCOM obtiene los mejores resultados de SNR segmentada y SNR segmentada muy por encima del sistema ADVANCE. Pero, en el caso de medidas de distancias espectrales, el sistema ADVANCE obtiene los mejores valores.

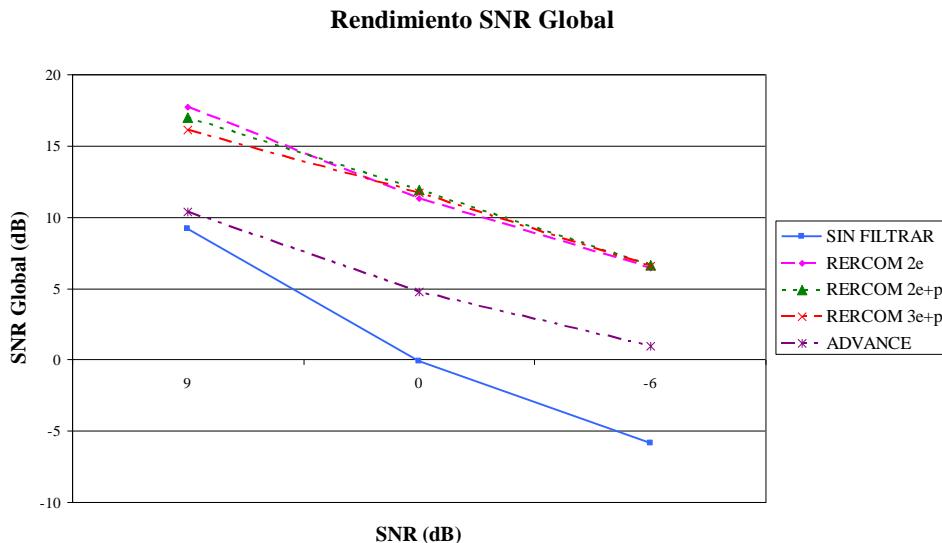


Gráfico 7.16: Promedio de rendimiento de SNR Global para ruido coche2

En el gráfico 7.16 podemos observar que frente a ruido de coche, el sistema RERCOM, en todas sus variantes, supera en una media de unos 6 dB a los resultados de SNR global obtenidos por el sistema ADVANCE. La variante del sistema RERCOM de 2 etapas + peine es la que obtiene los mejores resultados mejorando en unos 12 dB el nivel de SNR de la señal de voz contaminada con ruido.

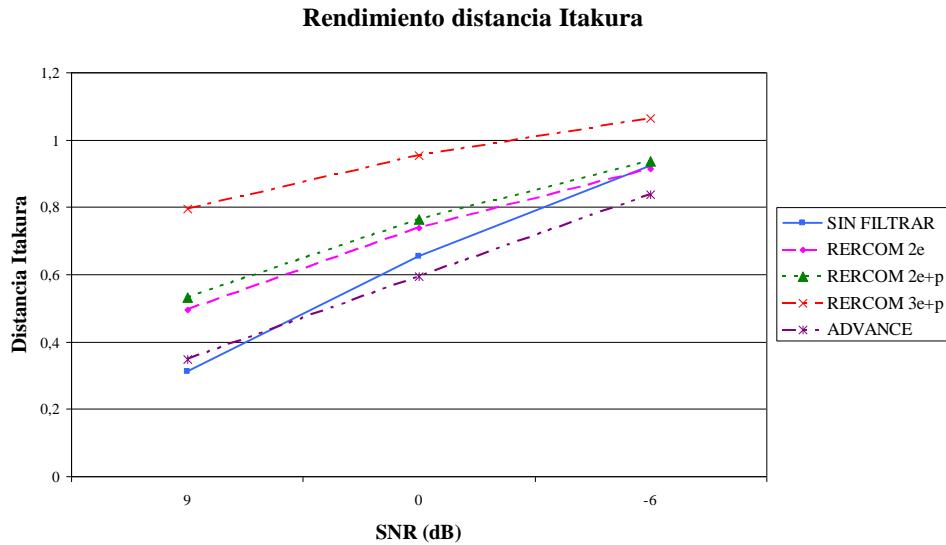


Gráfico 7.17: Promedio de rendimiento de distancia Itakura para ruido de coche2

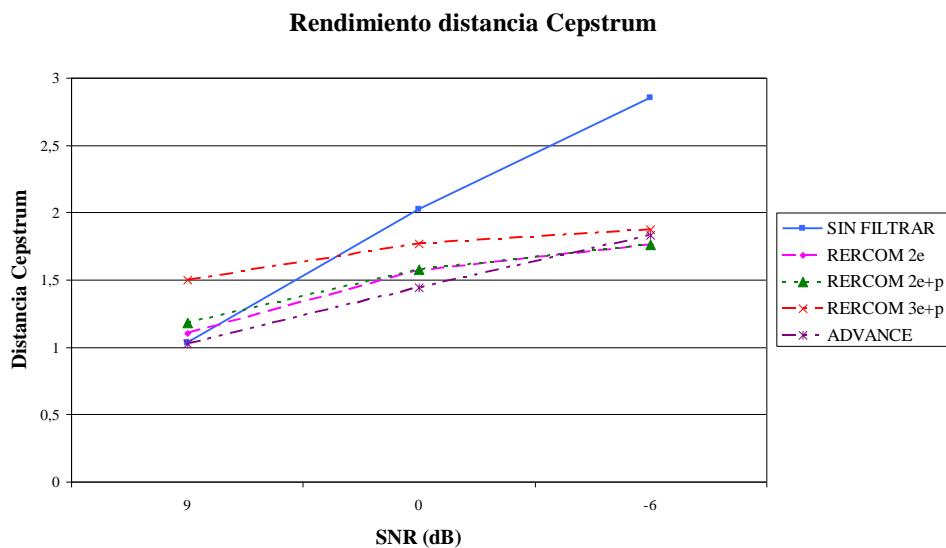


Gráfico 7.18: Promedio de rendimiento de distancia Cepstrum para ruido coche2

En este caso los gráficos 7.17 y 7.18 nos muestran los mismos resultados desconcertantes del apartado anterior, a pesar de la evidente mejora de la señal filtrada respecto a su original sin filtrar, los resultados de distancias indican que en algunos casos es mejor dejar la señal de voz contaminada de ruido. En los otros casos, el sistema ADVANCE, sobretodo en distancia Itakura, es el que los mejora ligeramente.

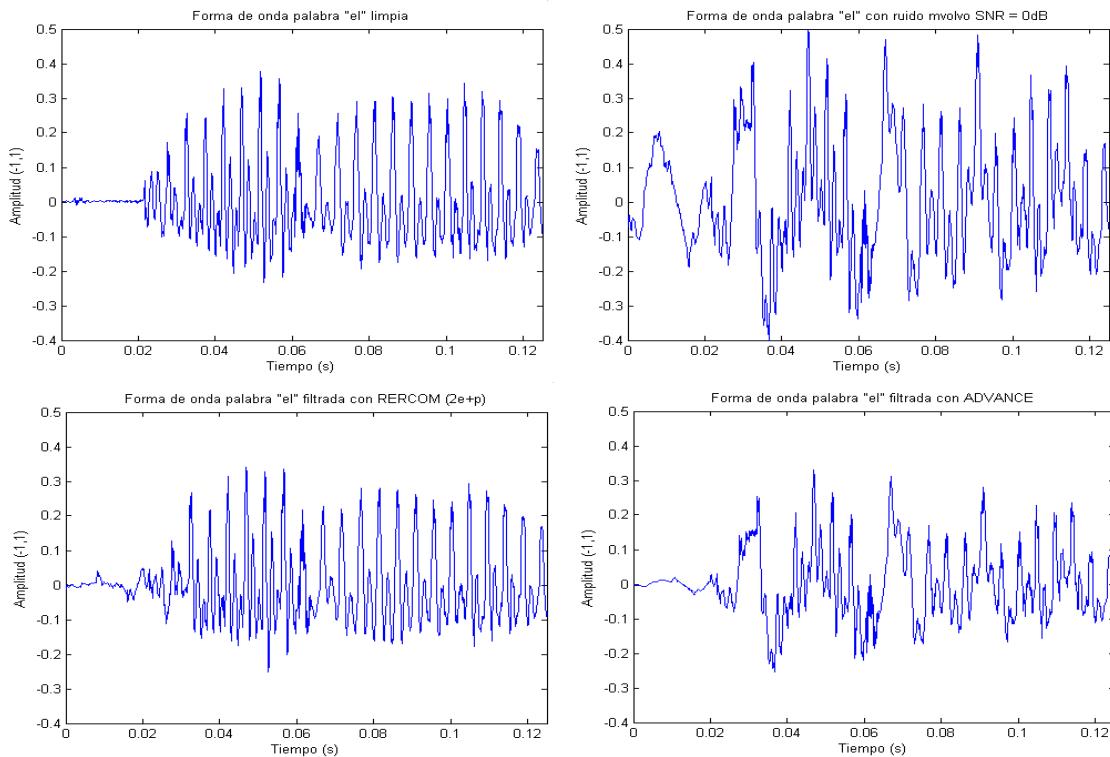


Fig. 7.26: Formas de onda de la palabra “el” del fichero ASUN1; comparativa RERCOM vs ADVANCE frente a ruido mvolve con SNR = 0 dB

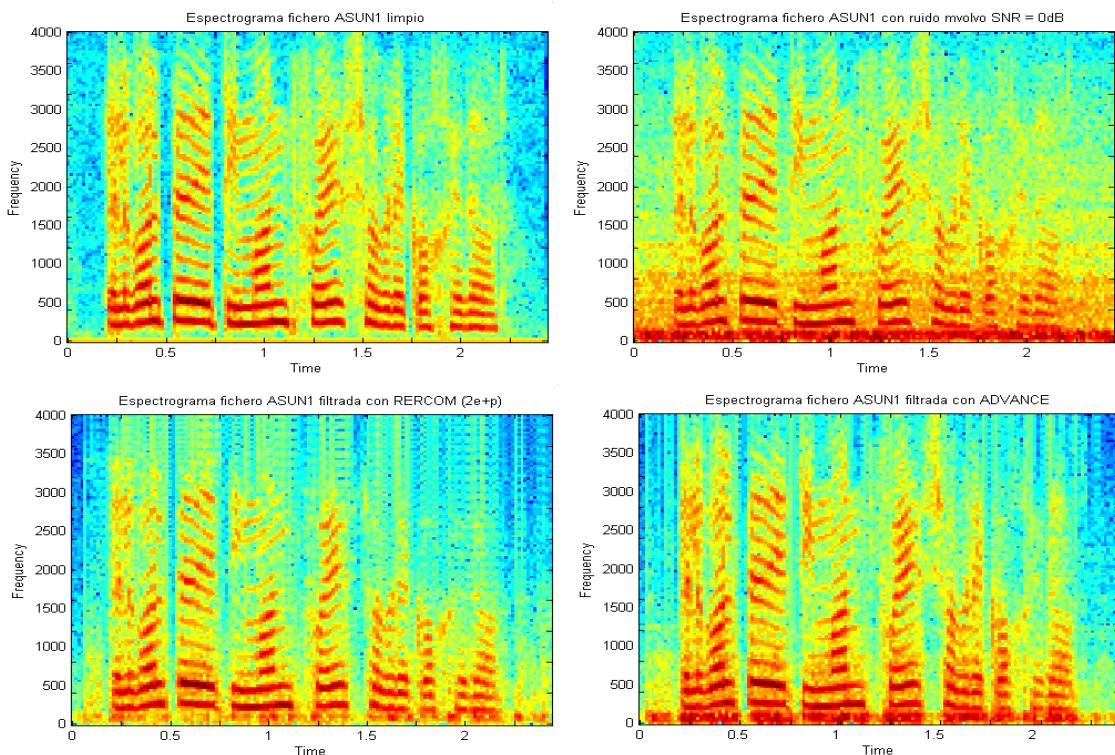


Fig. 7.27: Espectrogramas fichero ASUN1; comparativa RERCOM vs ADVANCE frente a ruido mvolve con SNR = 0 dB

### 7.3.1.3.-Ruidos de diferentes ambientes reales

En este apartado nos centramos en las perturbaciones provocadas por ruidos reales que dependen del ambiente en que se produce la comunicación y afectan negativamente a ésta. Trataremos el ruido que existe en el entorno de una fábrica, el ruido que produce el tráfico y el ruido que existirá en un tren o sus proximidades. Realizamos la evaluación para SNR de 9 dB, 0 dB y -6 dB, para los ficheros ASUN1 y ESCA.

#### 7.3.1.3.1.-Ruido de fábrica

Este es el tipo de ruido que generarían las diversas máquinas de una fábrica. En su espectro característico no se observan componentes frecuenciales destacadas y su espectro tiene la tendencia a disminuir su energía con el aumento de la frecuencia. Además, este tipo de ruido es poco estacionario, ya que se puede oír el vaivén de los golpes producidos por diferentes mecanismos.

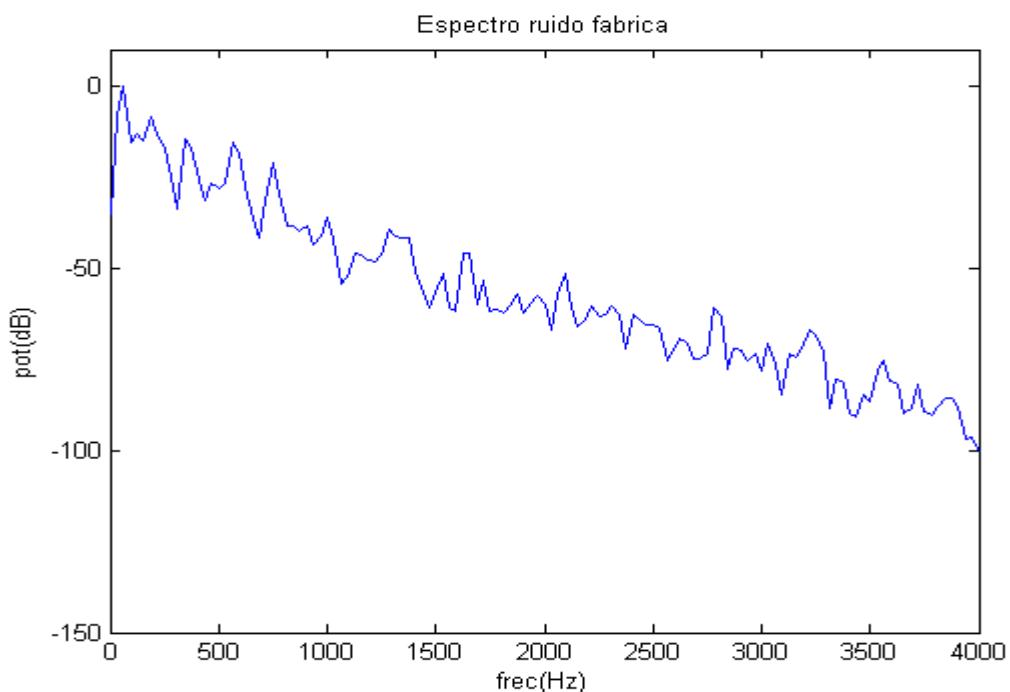


Fig. 7.28: Densidad espectral de energía del ruido de una fábrica

Los resultados obtenidos para el fichero de voz ASUN1 son:

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		9,049	1,826	0,625	1,214	1,824	1,679	0,000
RERCOM2e		15,124	6,107	0,506	1,517	<b>1,028</b>	1,452	<b>0,128</b>
RERCOM2e+p		<b>15,073</b>	<b>6,113</b>	0,561	1,907	1,134	1,563	0,266
RERCOM3e+p		14,533	5,690	0,696	1,972	1,233	1,971	0,198
ADVFRONT		10,731	3,666	<b>0,502</b>	<b>1,278</b>	1,038	<b>1,444</b>	2,708

Tabla.7.105: Evaluación del algoritmo AR2, con SNR=9 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		-0,052	-3,658	1,094	2,449	3,250	2,265	0,000
RERCOM2e		7,861	0,802	0,943	2,002	<b>1,484</b>	<b>2,006</b>	<b>0,304</b>
RERCOM2e+p		<b>8,465</b>	<b>1,260</b>	<b>0,941</b>	2,429	1,582	2,044	0,536
RERCOM3e+p		8,389	1,185	0,983	2,336	1,530	2,251	0,481
ADVFRONT		4,650	-0,809	1,056	<b>1,892</b>	1,692	2,181	4,368

Tabla.7.106: Evaluación del algoritmo AR2, con SNR=0 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		-5,552	-6,527	1,332	3,416	4,303	2,495	0,000
RERCOM2e		4,106	-1,892	1,223	2,634	1,969	2,292	<b>0,567</b>
RERCOM2e+p		<b>4,292</b>	<b>-1,751</b>	1,220	2,914	2,043	<b>2,284</b>	0,781
RERCOM3e+p		4,273	-1,757	<b>1,178</b>	2,809	<b>1,894</b>	2,420	0,670
ADVFRONT		1,382	-3,034	1,396	<b>2,560</b>	2,167	2,582	4,646

Tabla.7.107: Evaluación del algoritmo AR2, con SNR=-6 dB.

Los resultados obtenidos para el fichero de voz ESCA son:

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		9,046	2,401	1,352	2,676	3,443	2,614	0,000
RERCOM2e		14,257	7,336	1,014	2,043	1,722	2,193	<b>0,133</b>
RERCOM2e+p		<b>14,355</b>	<b>7,511</b>	1,032	2,199	1,727	2,231	0,321
RERCOM3e+p		14,173	7,317	1,214	2,443	1,926	2,517	0,333
ADVFRONT		10,274	5,117	<b>0,982</b>	<b>1,910</b>	<b>1,615</b>	<b>2,161</b>	2,933

Tabla.7.108: Evaluación del algoritmo AR2, con SNR=9 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	🔊	-0,055	-3,438	1,945	4,236	5,115	3,074	0,000
RERCOM2e	🔊	6,796	1,503	1,497	2,598	2,290	2,561	0,165
RERCOM2e+p	🔊	7,390	1,833	1,445	2,621	2,299	2,497	0,628
RERCOM3e+p	🔊	7,417	1,783	1,536	2,884	2,379	2,728	0,656
ADVFRONT	🔊	3,986	-0,292	1,721	2,824	2,594	2,884	4,067

Tabla.7.109: Evaluación del algoritmo AR2, con SNR=0 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	🔊	-5,555	-6,593	2,243	5,308	6,245	3,245	0,000
RERCOM2e	🔊	3,360	-1,277	1,893	3,247	2,957	2,845	0,386
RERCOM2e+p	🔊	3,493	-1,237	1,807	3,120	3,067	2,734	1,131
RERCOM3e+p	🔊	3,433	-1,282	1,796	3,172	3,030	2,876	1,168
ADVFRONT	🔊	1,135	-2,647	2,177	3,744	3,317	3,213	4,805

Tabla.7.110: Evaluación del algoritmo AR2, con SNR=-6 dB.

A partir de las tablas anteriores obtenemos las mismas conclusiones que hasta ahora nos han proporcionado los ruidos anteriores, el sistema RERCOM obtiene los mejores resultados de SNR segmentada y SNR segmentada. En el caso de medidas de distancias espectrales, el sistema RERCOM obtiene los mejores valores.

#### Rendimiento SNR Global

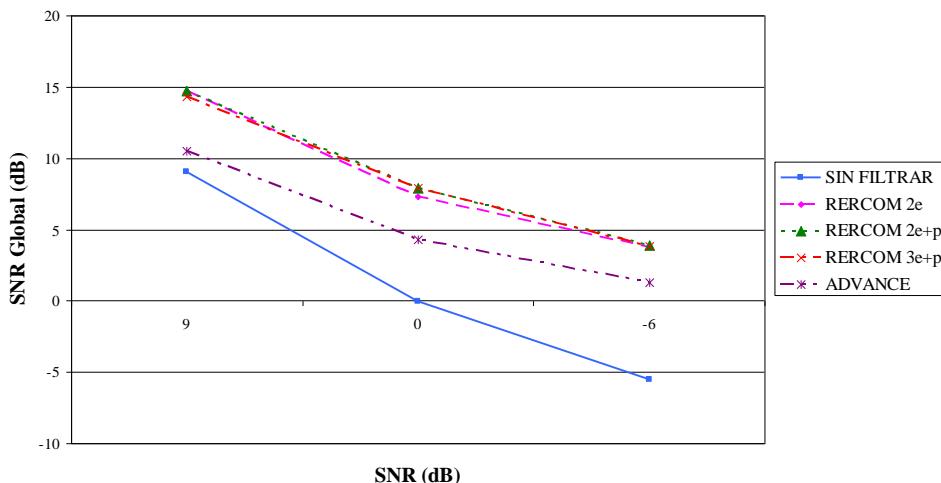


Gráfico 7.19: Promedio de rendimiento de SNR Global para ruido de fábrica

En el gráfico 7.19 podemos observar que frente a ruido de una fábrica, el sistema RERCOM, en todas sus variantes, supera en una media de unos 4 dB a los resultados de SNR global obtenidos por el sistema ADVANCE, mejorando la señal original en un

máximo de 10 dB (SNR = -6 dB). Frente a este tipo de ruido las diferentes variante del sistema RERCOM obtienen resultados muy similares.

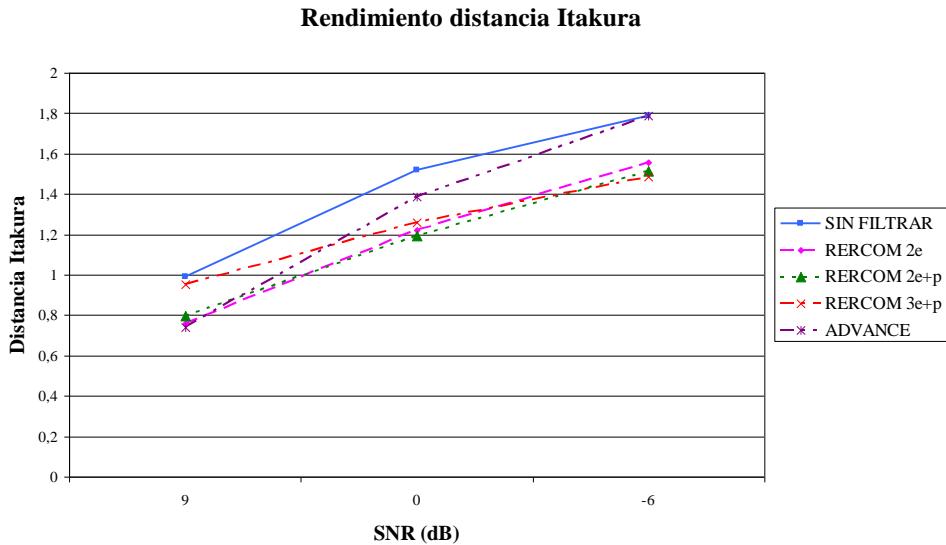


Gráfico 7.20: Promedio de rendimiento de distancia Itakura para ruido de fábrica

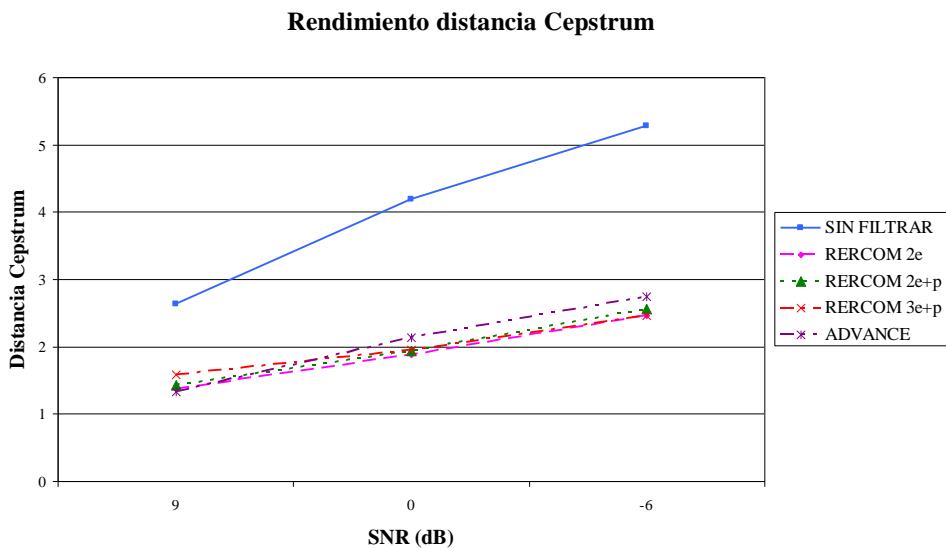


Gráfico 7.21: Promedio de rendimiento de distancia Cepstrum para ruido fábrica

En este caso los gráficos 7.20 y 7.21 observamos que el mejor filtrado, RERCOM 2 etapas + peine, sólo es capaz de reducir un 15% la distancia Itakura, sin embargo en distancia cepstrum, todos los sistemas se comportan de forma similar reduciendo en un 50% la distancia.

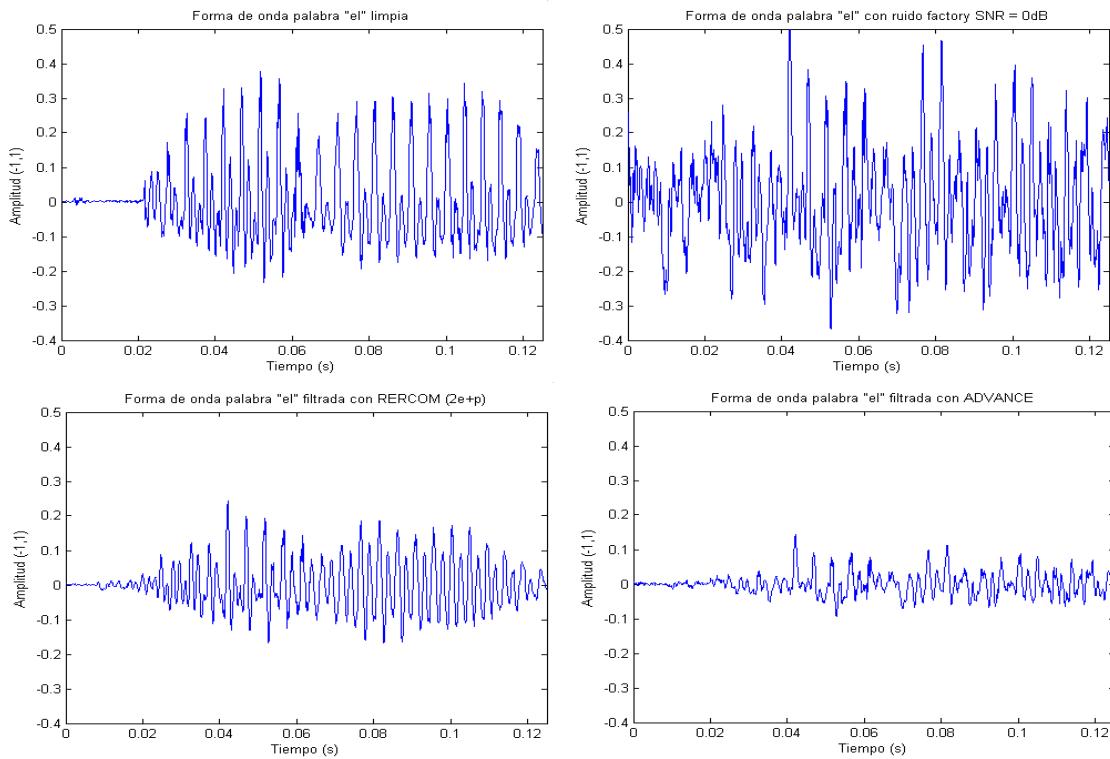


Fig. 7.29: Formas de onda de la palabra “el” del fichero ASUN1; comparativa RERCOM vs ADVANCE frente a ruido fábrica con SNR = 0 dB

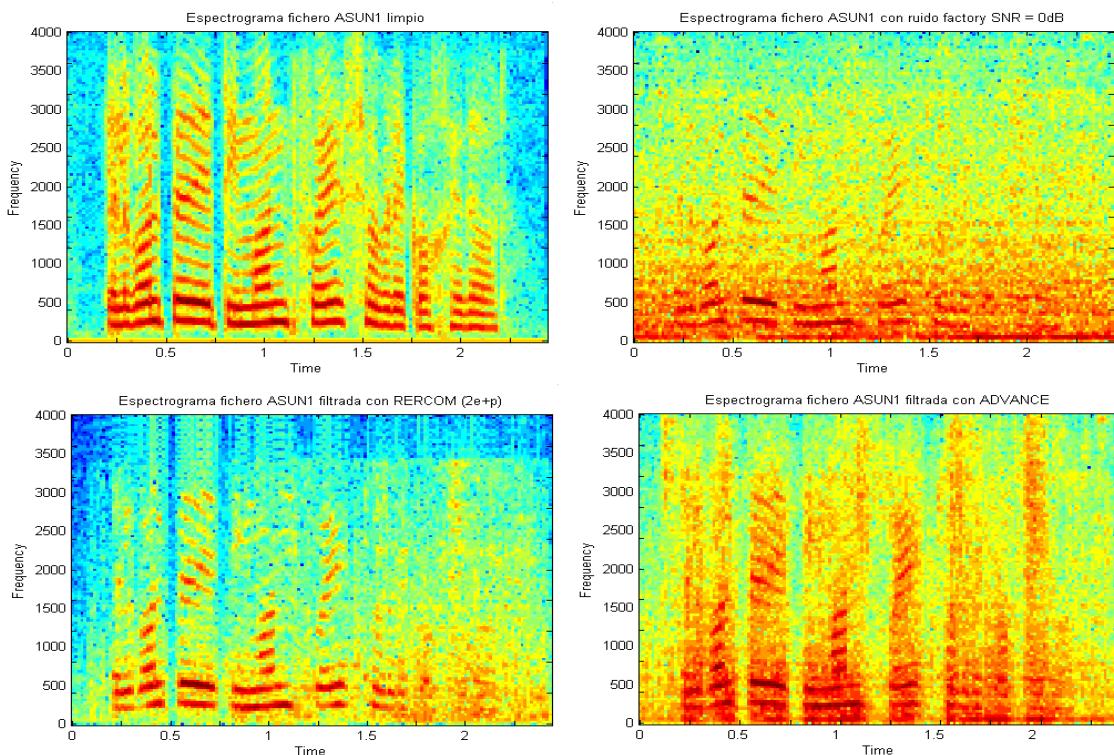


Fig. 7.30: Espectrogramas fichero ASUN1; comparativa RERCOM vs ADVANCE frente a ruido fábrica con SNR = 0 dB

### 7.3.1.3.2.-Ruido de tráfico

Este es el tipo de ruido de fondo que podría escucharse en una oficina o en cualquier otro lugar que estuviese cerca de una avenida con tráfico de coches. Este ruido tiene la importante componente situada a 125 Hz característica del ruido producido por coches, pero a diferencia del ruido de coche, la envolvente del espectro no decae monótonamente, sino que genera un pico redondeado de unos 35 dB de energía menos que el primero situado a unos 1,1 Khz, a partir de este punto la envolvente decae monótonamente.

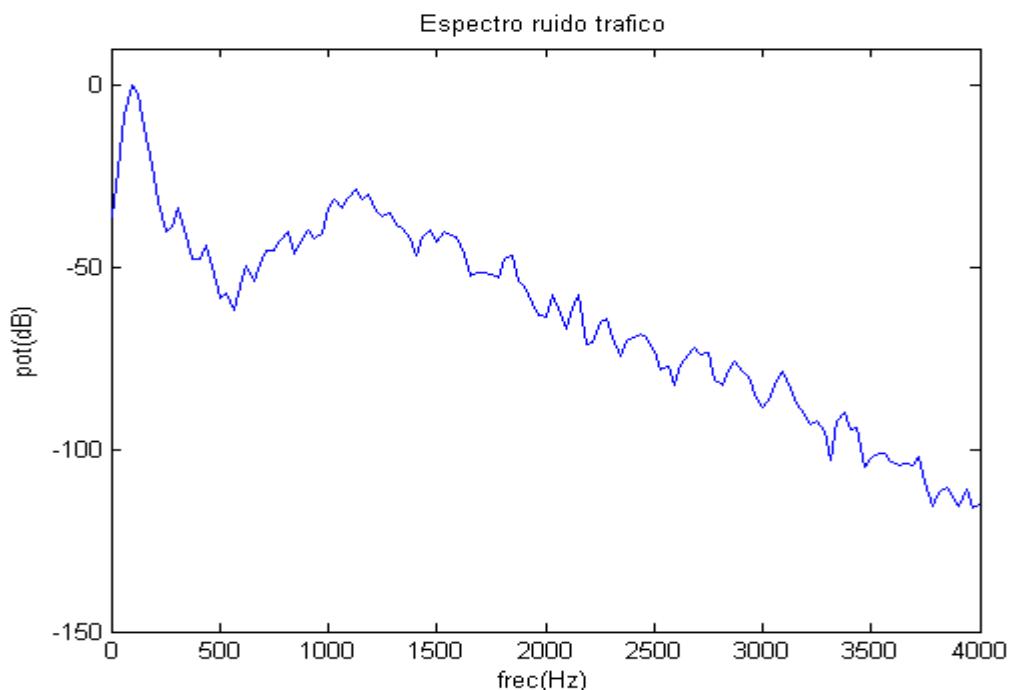


Fig. 7.31: Densidad espectral de energía del ruido de tráfico

Los resultados obtenidos para el fichero de voz ASUN1 son:

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	🔊	9,027	1,899	0,498	0,857	1,381	1,533	0,000
RERCOM2e	🔊	16,170	6,801	0,571	2,307	1,144	1,589	0,100
RERCOM2e+p	🔊	16,736	7,184	0,604	2,612	1,238	1,727	0,326
RERCOM3e+p	🔊	15,493	6,518	0,735	2,540	1,362	2,089	0,252
ADVFRONT	🔊	10,634	4,086	0,402	2,095	1,033	1,343	1,943

Tabla.7.111: Evaluación del algoritmo AR2, con SNR=9 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		-0,374	-3,857	0,977	1,937	2,686	2,155	0,000
RERCOM2e		9,461	2,179	0,856	<b>2,202</b>	1,362	1,941	<b>0,157</b>
RERCOM2e+p		<b>10,232</b>	<b>2,840</b>	0,889	2,434	1,422	2,009	0,442
RERCOM3e+p		9,834	2,545	0,927	2,246	1,455	2,236	0,439
ADVFRONT		4,330	-0,347	<b>0,715</b>	2,454	<b>1,298</b>	<b>1,907</b>	2,689

Tabla.7.112: Evaluación del algoritmo AR2, con SNR=0 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		-5,674	-6,727	1,310	2,783	3,626	2,417	0,000
RERCOM2e		5,148	-0,441	1,020	2,281	1,583	2,170	<b>0,133</b>
RERCOM2e+p		<b>5,307</b>	<b>-0,229</b>	<b>1,009</b>	2,339	1,591	<b>2,162</b>	0,532
RERCOM3e+p		5,067	-0,375	1,035	<b>2,210</b>	1,641	2,401	0,562
ADVFRONT		1,505	-2,122	1,071	3,176	<b>1,523</b>	2,259	3,003

Tabla.7.113: Evaluación del algoritmo AR2, con SNR=-6 dB.

Los resultados obtenidos para el fichero de voz ESCA son:

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		9,362	2,996	1,019	1,993	2,709	2,325	0,000
RERCOM2e		14,365	7,215	<b>0,968</b>	<b>1,993</b>	1,633	2,223	<b>-0,013</b>
RERCOM2e+p		<b>15,160</b>	<b>8,056</b>	0,998	2,164	1,657	2,278	0,148
RERCOM3e+p		14,931	7,820	1,187	2,476	1,908	2,557	0,137
ADVFRONT		11,196	5,674	0,693	2,016	<b>1,275</b>	<b>1,839</b>	2,366

Tabla.7.114: Evaluación del algoritmo AR2, con SNR=9 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		0,261	-2,970	1,607	3,443	4,294	2,849	0,000
RERCOM2e		7,612	2,114	1,312	2,358	2,240	2,593	<b>-0,093</b>
RERCOM2e+p		<b>8,660</b>	<b>3,210</b>	1,285	2,500	2,207	2,591	0,266
RERCOM3e+p		8,551	3,129	1,402	2,661	2,395	2,784	0,252
ADVFRONT		4,574	0,355	<b>1,208</b>	<b>2,304</b>	<b>1,944</b>	<b>2,431</b>	2,579

Tabla.7.115 : Evaluación del algoritmo AR2, con SNR=0 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	🔊	-5,639	-6,409	1,982	4,560	5,472	3,048	0,000
RERCOM2e	🔊	2,962	-0,841	1,559	2,582	2,797	2,830	-0,121
RERCOM2e+p	🔊	3,063	-0,496	1,543	2,616	2,782	2,811	0,298
RERCOM3e+p	🔊	2,970	-0,525	1,625	2,878	2,902	2,983	0,271
ADVFRONT	🔊	1,586	-1,778	1,678	3,663	2,401	2,834	2,287

Tabla.7.116 : Evaluación del algoritmo AR2, con SNR=-6 dB.

En las tablas anteriores podemos observar que el sistema RERCOM obtiene los mejores resultados de SNR segmentada y SNR segmentada por encima del sistema ADVANCE. Pero, en el caso de medidas de distancias espectrales, el sistema ADVANCE obtiene los mejores valores.

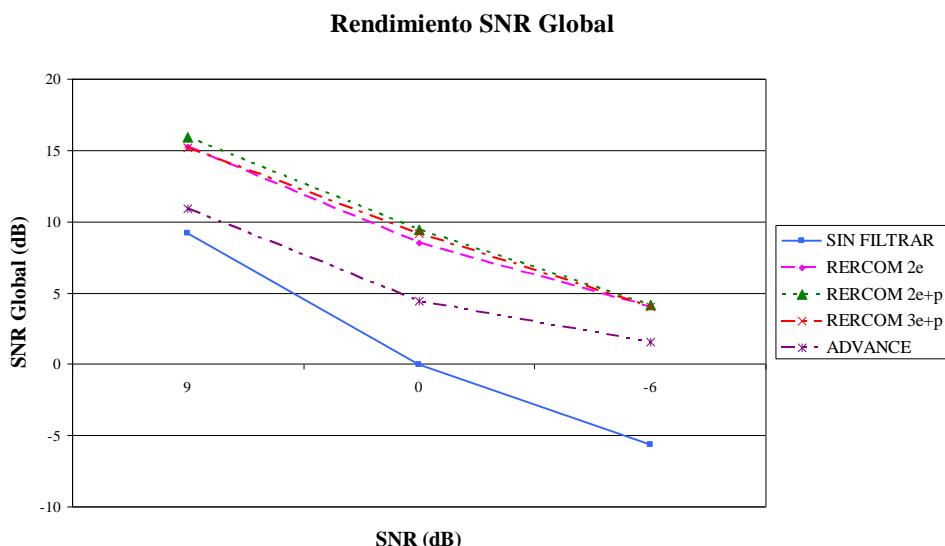


Gráfico 7.22: Promedio de rendimiento de SNR Global para ruido de tráfico.

En el gráfico 7.22 observamos que frente a ruido de tráfico el sistema RERCOM, en todas sus variantes, supera en una media de unos 5 dB a los resultados de SNR global obtenidos por el sistema ADVANCE, mejorando la señal original en un máximo de 10 dB (SNR = -6 dB). Por lo que respecta a mediadas de distancia el sistema ADVANCE, al igual que con los ruidos de coche obtiene los mejores resultados.

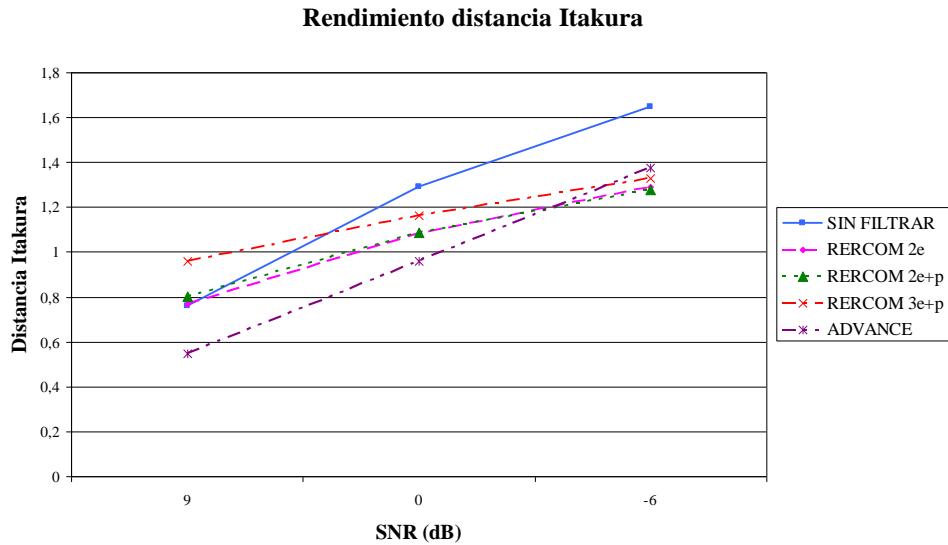


Gráfico 7.23: Promedio de rendimiento de distancia Itakura para ruido de tráfico.

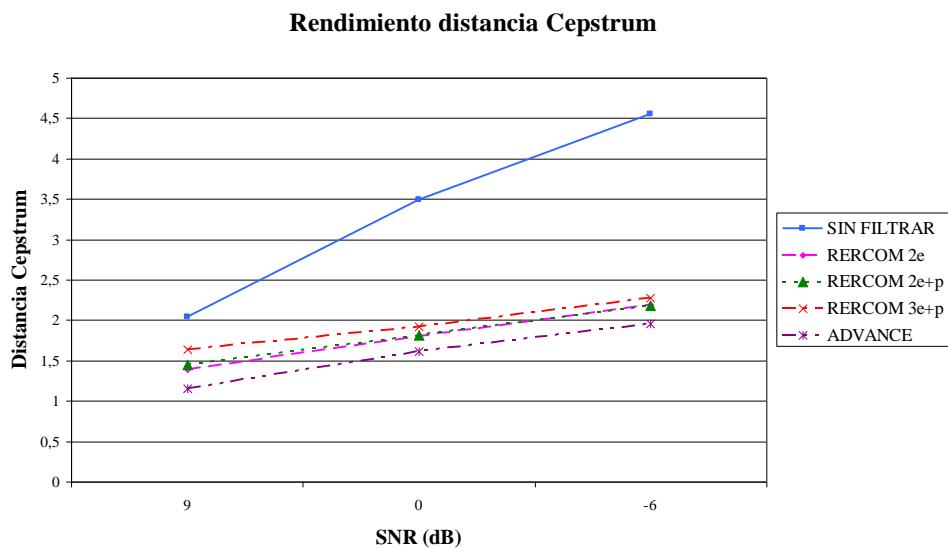


Gráfico 7.24: Promedio de rendimiento de distancia Cepstrum para ruido de tráfico.

Los gráficos 7.23 y 7.24 muestran que los mejores resultados de distancias espectrales, al igual que ocurría con los ruidos de coche, los obtiene el sistema ADVANCE. Por otro lado, el sistema RERCOM 3e + p es el que obtiene los peores resultados incluso llegando a empeorar la distancia Itakura frente a una SNR superior a 9 dB. Frente a este ruido, el mejor filtrado reduce un 20% la distancia Itakura y un 50% la distancia Cepstrum.

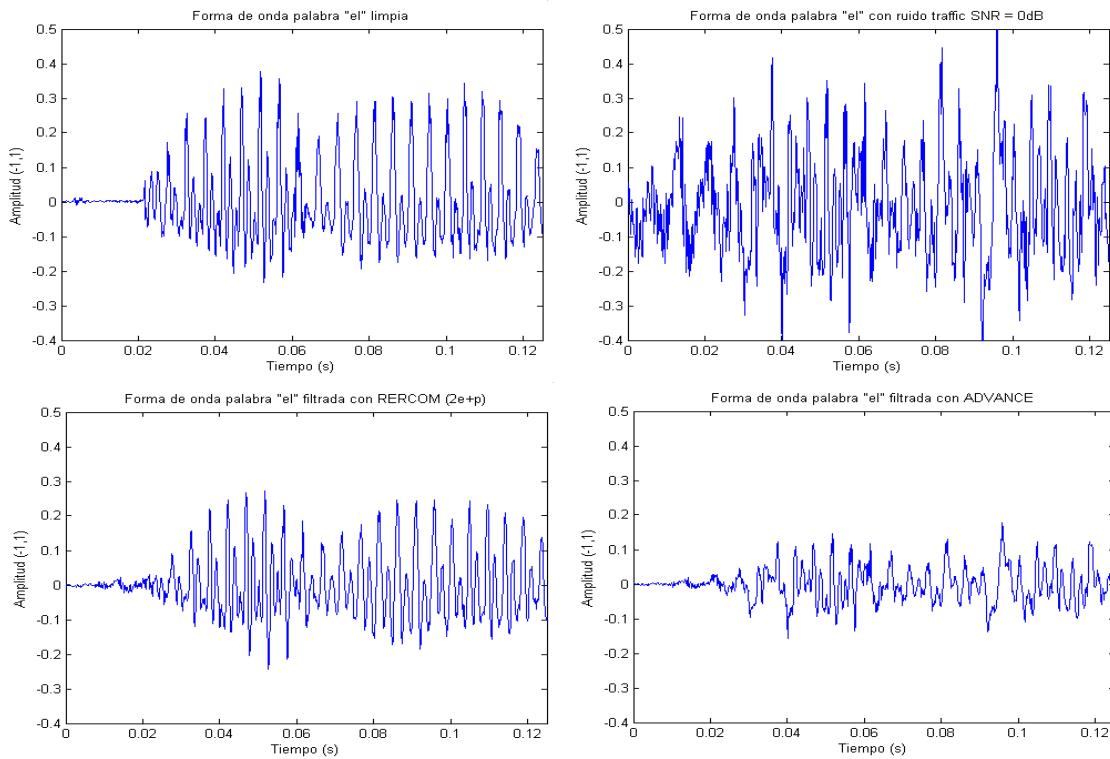


Fig. 7.32: Formas de onda de la palabra “el” del fichero ASUN1; comparativa RERCOM vs ADVANCE frente a ruido tráfico con SNR = 0 dB

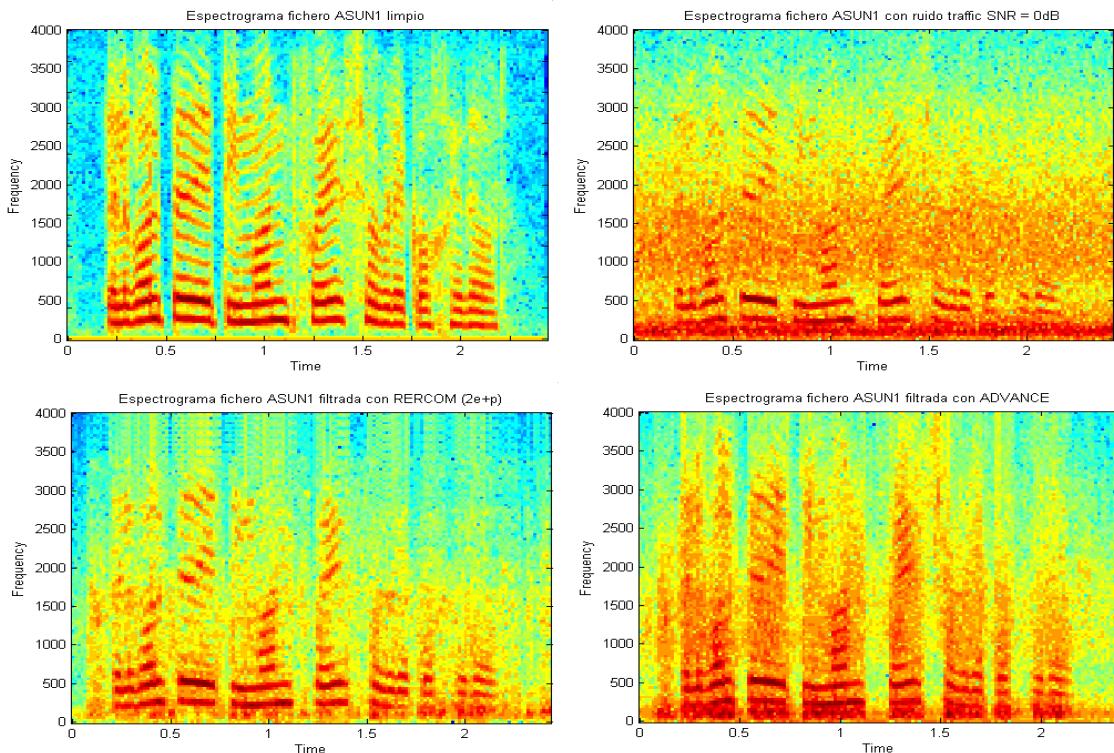


Fig. 7.33: Espectrogramas fichero ASUN1; comparativa RERCOM vs ADVANCE frente a ruido tráfico con SNR = 0 dB

### 7.3.1.3.3.-Ruido de tren

Este es el tipo de ruido que podría escucharse en el interior de un tren en movimiento. Observamos en este ruido tres componentes frecuenciales importantes, situadas a 50, 450 y 800 Hz con niveles de energía similares. A partir del último pico el espectro decae 50 dB de manera bastante brusca, en un rango de unos 500 Hz. A partir de este punto el espectro se mantiene bastante plano, con un ligero decaimiento con el aumento de la frecuencia..

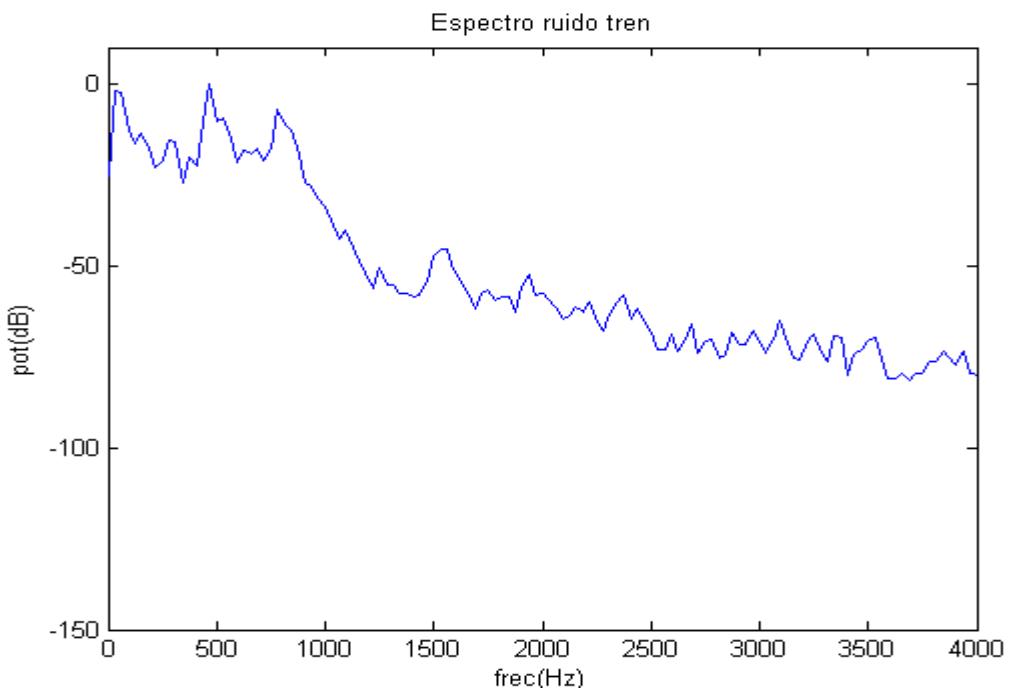


Fig. 7.34: Densidad espectral de energía del ruido de tren

Los resultados obtenidos para el fichero de voz ASUN1 son:

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	🔊	9,116	1,707	0,520	0,926	1,464	1,556	0,000
RERCOM2e	🔊	15,219	6,080	0,448	1,101	0,897	1,306	0,038
RERCOM2e+p	🔊	15,352	6,238	0,486	1,274	0,994	1,451	0,248
RERCOM3e+p	🔊	14,773	5,815	0,651	1,612	1,198	1,967	0,118
ADVFRONT	🔊	9,783	3,303	0,401	1,391	0,929	1,345	1,852

Tabla.7.117: Evaluación del algoritmo AR2, con SNR=9 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		-0,085	-3,929	0,968	2,008	2,764	2,167	0,000
RERCOM2e		7,660	0,516	0,797	<b>1,615</b>	<b>1,347</b>	<b>1,841</b>	<b>0,002</b>
RERCOM2e+p		<b>8,383</b>	<b>1,001</b>	<b>0,779</b>	1,744	1,399	1,852	0,320
RERCOM3e+p		8,299	0,914	0,896	1,758	1,505	2,183	0,289
ADVFRONT		4,211	-0,822	0,814	2,139	1,381	1,966	2,293

Tabla.7.118 : Evaluación del algoritmo AR2, con SNR=0 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		-5,585	-6,804	1,236	2,906	3,759	2,431	0,000
RERCOM2e		3,524	-2,265	1,050	<b>1,933</b>	<b>1,678</b>	2,167	<b>-0,033</b>
RERCOM2e+p		<b>3,951</b>	<b>-1,958</b>	<b>1,011</b>	2,145	1,710	<b>2,108</b>	0,069
RERCOM3e+p		3,839	-2,013	1,111	2,128	1,841	2,355	0,198
ADVFRONT		1,540	-2,541	1,163	3,980	1,724	2,407	2,397

Tabla.7.119: Evaluación del algoritmo AR2, con SNR=-6 dB.

Los resultados obtenidos para el fichero de voz ESCA son:

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		8,748	2,194	1,162	2,304	3,036	2,595	0,000
RERCOM2e		13,878	6,547	0,892	<b>1,551</b>	1,618	2,224	-0,147
RERCOM2e+p		<b>14,448</b>	<b>7,057</b>	0,927	1,569	1,631	2,243	<b>0,113</b>
RERCOM3e+p		14,361	6,878	1,213	2,095	2,001	2,610	0,187
ADVFRONT		10,505	4,868	<b>0,787</b>	1,605	<b>1,455</b>	<b>2,105</b>	2,287

Tabla.7.120: Evaluación del algoritmo AR2, con SNR=9 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL		-0,254	-3,657	1,709	3,751	4,606	3,006	0,000
RERCOM2e		7,499	1,393	<b>1,297</b>	2,182	<b>2,270</b>	2,582	-0,325
RERCOM2e+p		<b>8,592</b>	<b>2,083</b>	1,331	2,261	2,326	<b>2,566</b>	<b>0,189</b>
RERCOM3e+p		8,541	2,045	1,484	2,571	2,594	2,805	0,336
ADVFRONT		3,927	-0,500	1,356	<b>2,171</b>	2,339	2,723	2,781

Tabla.7.121: Evaluación del algoritmo AR2, con SNR=0 dB.

	PLAY	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	🔊	-5,854	-6,859	2,003	4,809	5,728	3,151	0,000
RERCOM2e	🔊	3,431	-1,432	1,613	2,565	2,847	2,760	-0,347
RERCOM2e+p	🔊	3,501	-1,332	1,634	2,611	3,017	2,693	0,433
RERCOM3e+p	🔊	3,371	-1,388	1,758	2,890	3,249	2,940	0,480
ADVFRONT	🔊	0,939	-2,505	1,762	2,965	2,829	3,012	2,887

Tabla 7.122: Evaluación del algoritmo AR2, con SNR=-6 dB.

A partir de las tablas observamos que en este caso las tres variantes del sistema RERCOM se comportan de manera similar, siendo la variante RERCOM 3 etapas + peine la que peores resultados de distancias espectrales obtiene. En este caso el sistema ADVANCE obtiene peor rendimiento que las variantes del sistema RERCOM en SNR global y segmentada, pero en distancia espectrales sus resultados son similares a las variantes RERCOM de 2 etapas y 2 etapas + peine.

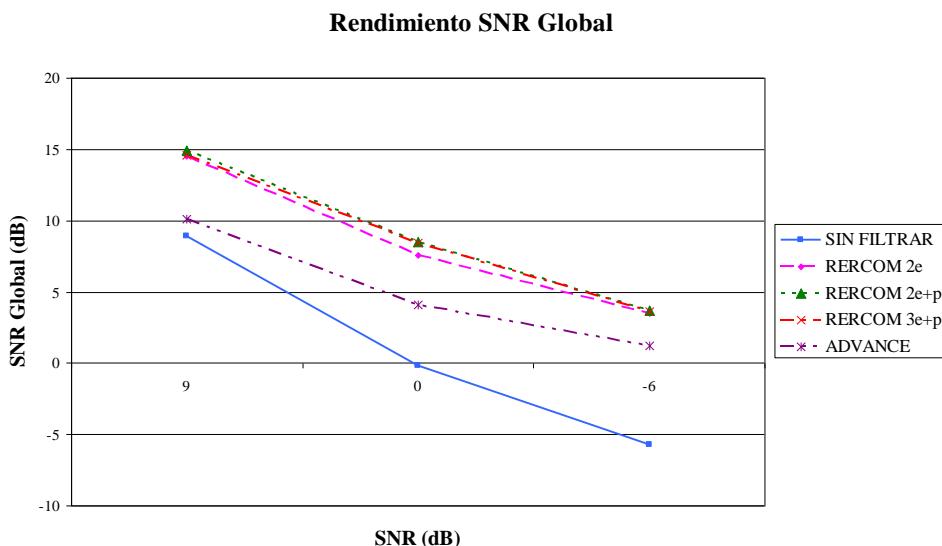


Gráfico 7.25: Promedio de rendimiento de SNR Global para ruido de tren.

En el gráfico 7.25 observamos que frente a ruido de tráfico todos las variantes del sistema RERCOM se comportan de forma similar, superando en una media de unos 4 dB a los resultados de SNR global obtenidos por el sistema ADVANCE, mejorando la señal original entre mínimo de 6 dB (SNR = 9 dB) y un máximo de 10 dB (SNR = -6 dB).

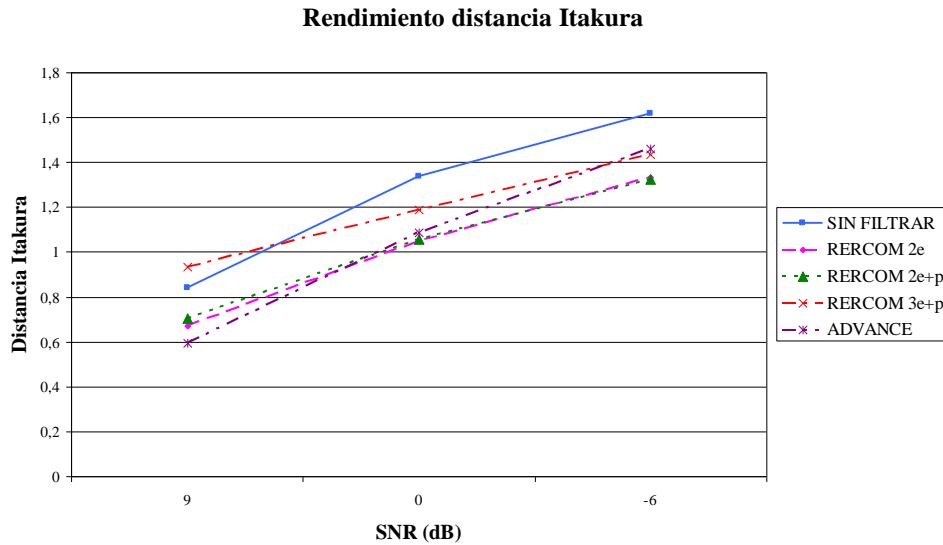


Gráfico 7.26: Promedio de rendimiento de distancia Itakura para ruido de tren.

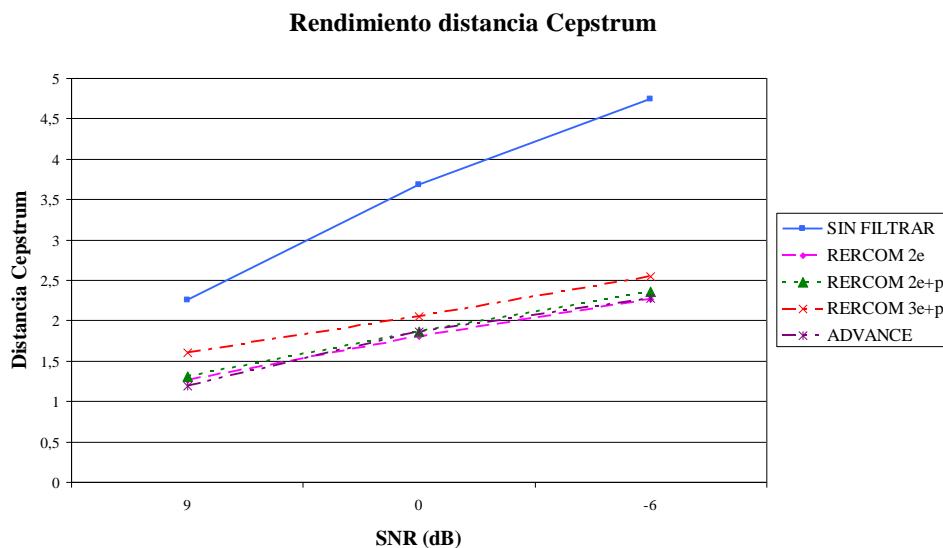


Gráfico 7.27: Promedio de rendimiento de distancia Cepstrum para ruido de tren.

Los gráficos 7.26 y 7.27 muestran que en este caso el sistema que obtiene unos resultados ligeramente peores es la variante RERCOM 3 etapas + peine. Los demás sistemas se comportan de manera muy similar, mejorando en un 10-15 % la distancia Itakura y un 50% la distancia Cepstrum.

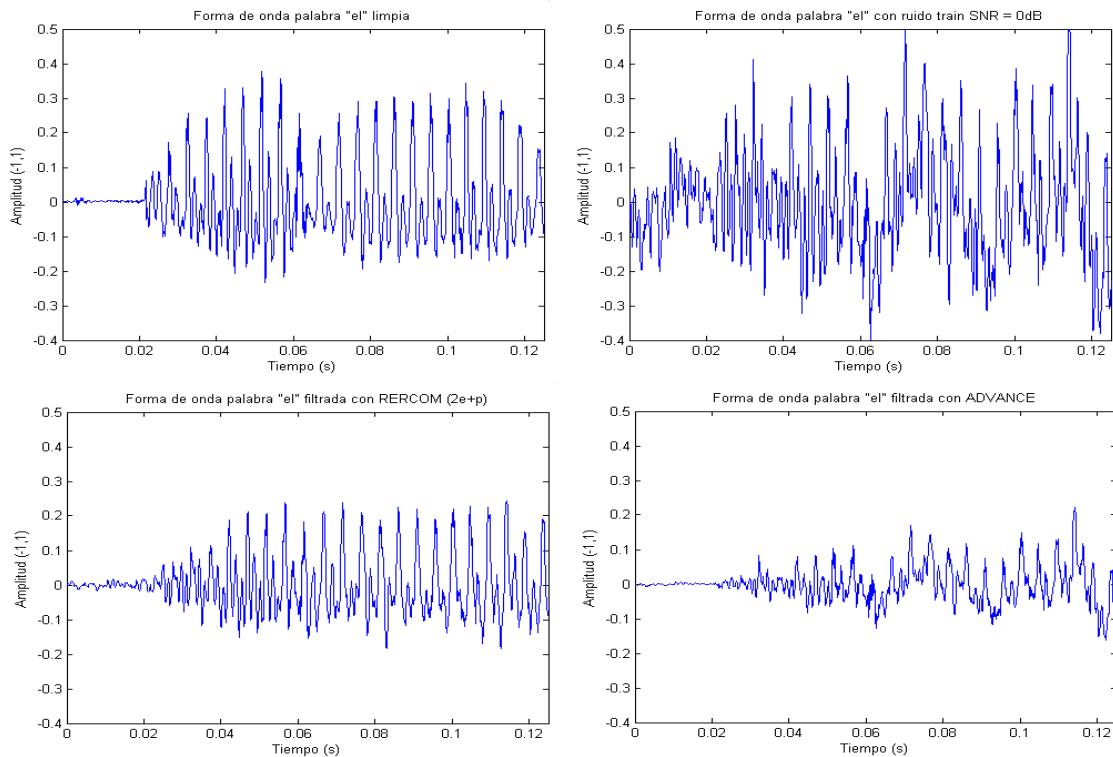


Fig. 7.35: Formas de onda de la palabra “el” del fichero ASUN1; comparativa RERCOM vs ADVANCE frente a ruido de tren con SNR = 0 dB

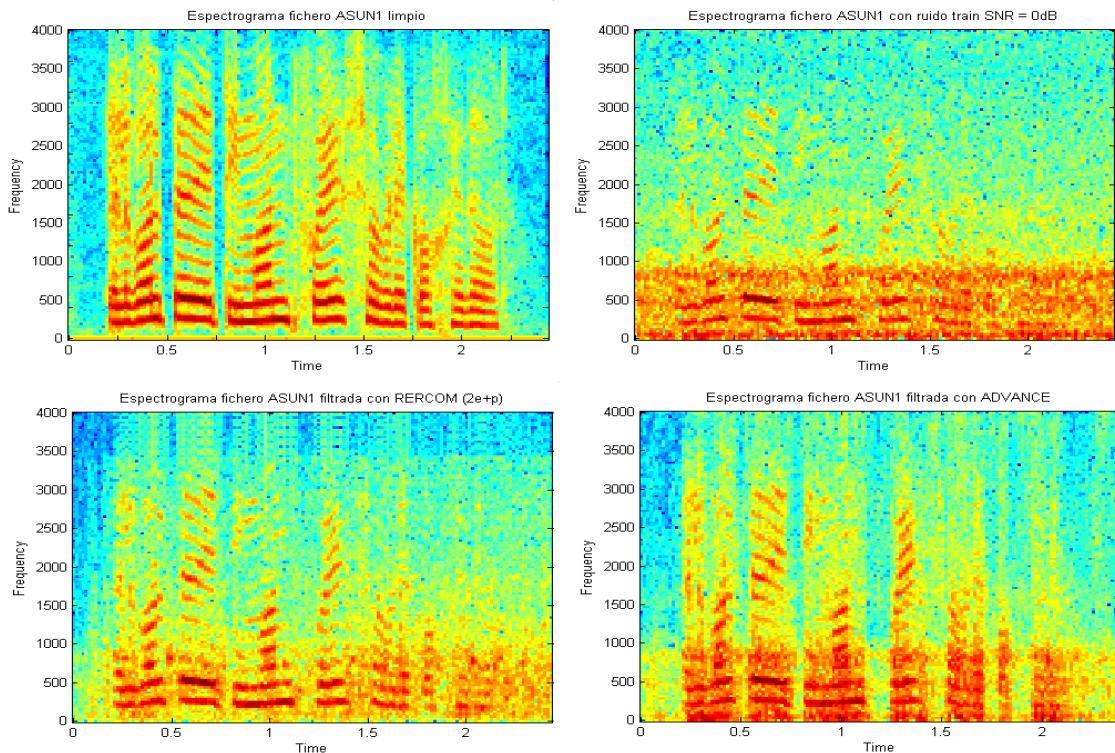


Fig. 7.36: Espectrogramas fichero ASUN1; comparativa RERCOM vs ADVANCE frente a ruido de tren con SNR = 0 dB

#### 7.4.-Comparativa de reconocimiento de voz con Aurora versión 008

La prueba de reconocimiento de voz se ha realizado siguiendo las indicaciones del documento [Pear-98], utilizando un entrenamiento “multi-condition Training” del sistema de reconocimiento.

Para realizar este entrenamiento se han filtrado los 8440 archivos entrenamiento de TIdigits con los programas Rercom y AdvFrontEnd, estos archivos de entrenamiento, creados por 110 locutores diferentes (55 hombres y 55 mujeres), se distribuyen en diferentes conjuntos en función del tipo de ruido y de su SNR según la siguiente tabla:

	Ruido 1	Ruido 2	Ruido 3	Ruido 4
Limpio	422	422	422	422
20dB	422	422	422	422
15dB	422	422	422	422
10dB	422	422	422	422
5dB	422	422	422	422

Fig 7.123 : Tabla de distribución de ficheros de entrenamiento “multi-condition” en SNR y tipo de ruido.

Donde los ruidos añadidos artificialmente son:

- Ruido 1: Sala de exhibición (Exhibition hall).
- Ruido 2: Voces (Babble noise (big room, office, people chatting)).
- Ruido 3: Metro (suburban train).
- Ruido 4: Coche (Car moving).

Una vez entrenado el sistema de reconocimiento se obtienen los resultados de rendimiento de reconocimiento para 7 niveles de SNR diferentes, para cada nivel de reconocimiento es necesario analizar 4000 archivos (1000 por cada ruido), según se indica en la siguiente tabla:

	Ruido 1	Ruido 2	Ruido 3	Ruido 4
Limpio	1000	1000	1000	1000
20dB	1000	1000	1000	1000
15dB	1000	1000	1000	1000
10dB	1000	1000	1000	1000
5dB	1000	1000	1000	1000
0dB	1000	1000	1000	1000
-5dB	1000	1000	1000	1000

Fig 7.124 : Tabla de distribución de ficheros de test en SNR y tipo de ruido.

Los rendimientos de reconocimiento están expresados en tanto por ciento de precisión de palabra según la ecuación:

$$\text{Precisión\_palabra} = \frac{n^{\circ}\text{palabras\_acertadas} - 0.5 \cdot n^{\circ}\text{palabras\_insertadas}}{n^{\circ}\text{palabras\_totales}} \cdot 100$$

Como se ha mencionado antes, el test se realiza para 7 condiciones de SNR diferentes, desde señal limpia hasta -5 dB. Según el documento [], el rendimiento general se obtiene a partir del promediado de los resultados obtenidos (precisión de palabra) entre 0 dB a 20 dB, según la tabla:

SNR	Limpio	20 dB	15 dB	10 dB	5 dB	0 dB	-5 dB
PESO	0,00	0,20	0,20	0,20	0,20	0,20	0,00

Fig 7.125 : Tabla de pesos utilizado en el promedio “Media1”.

Los resultados obtenido con este tipo de promedio están anotados como “Media1”. La etiqueta ‘Media2’, en cambio, indica que en los resultados obtenidos se ha añadido al promedio los rendimientos obtenidos en el test de ruido con SNR=-5dB, con un peso igual al que indica la siguiente tabla: nosotros hemos querido ir más allá,

SNR	Limpio	20 dB	15 dB	10 dB	5 dB	0 dB	-5 dB
PESO	0,00	1/6	1/6	1/6	1/6	1/6	1/6

Fig 7.126 : Tabla de pesos utilizado en el promedio “Media2”.

Por otro lado, los sistemas analizados corresponden a:

- **BASELINE**: Las pruebas se han realizado utilizando el Front-End de reducción de ruido de referencia.
- **RERCOM2e**: Las pruebas se han realizado utilizando el programa Rercom utilizando un esquema de prefiltro + filtro utilizando los parámetros optimizados para ruido blanco.
- **RERCOM2e+p**: Las pruebas se han realizado utilizando el programa Rercom utilizando un esquema de prefiltro + filtro con peine utilizando los parámetros optimizados para ruido blanco.
- **RERCOM3e+p**: Las pruebas se han realizado utilizando el programa Rercom utilizando un esquema de prefiltro + filtro con peine + postfiltro utilizando los parámetros optimizados para ruido blanco.
- **ADVFRONT**: Las pruebas se han realizado utilizando el programa Advance Front-End modificado. Se ha adaptado el programa para extraiga, en un fichero, la señal de test filtrada por el modulo de reducción de ruido de este programa.

Los resultados obtenidos en las pruebas de reconocimiento bajo las condiciones anteriores son los siguientes:

RUIDO 1	Limpia	20 dB	15dB	10 dB	5 dB	0 dB	-5 dB	Media1	Media2
BASELINE	98,65	97,6	96,16	92,85	83,11	47,31	18,61	83,406	72,607
RERCOM2e	98,89	97,73	96,81	94,44	88,67	71,88	41,69	89,906	81,870
RERCOM2e+p	<b>98,93</b>	97,66	<b>96,96</b>	94,14	<b>89,35</b>	74,18	42,92	<b>90,458</b>	82,535
RERCOM3e+p	98,62	97,11	95,92	93,68	88,67	<b>75,16</b>	<b>45,78</b>	90,108	<b>82,720</b>
ADVFRONT	98,71	<b>98,03</b>	<b>96,96</b>	<b>94,57</b>	88,33	70,86	34,85	89,750	80,600

Tabla 7.127: Porcentajes de precisión de reconocimiento de palabra para el ruido tipo1 (sala/hall).

Rendimiento reconocimiento ruido sala

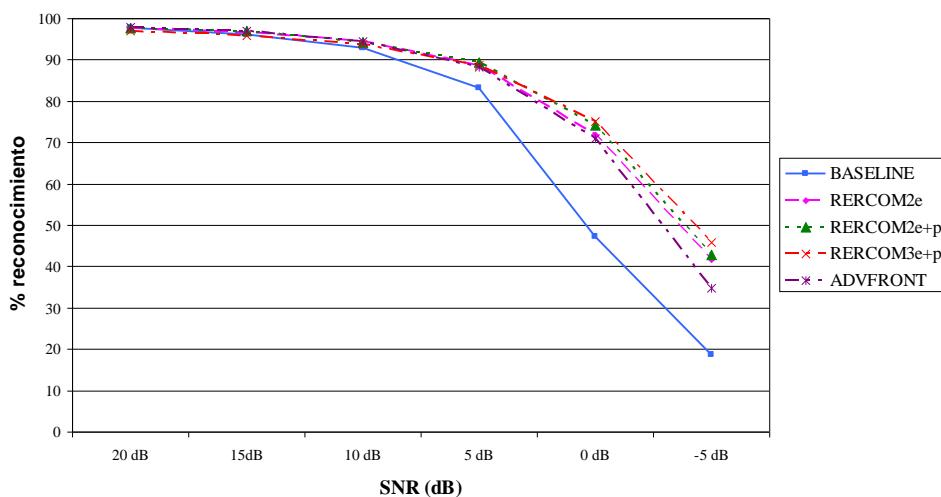


Gráfico 7.28: Rendimiento de reconocimiento para ruido de Sala

Los resultados de esta tabla nos indican que el tipo de esquema de reducción de ruido que obtiene un mejor rendimiento global frente al ruido tipo 1 (sala de exhibición) es la estructura **RERCOM2e+p**, obteniendo el mejor resultado “Media1”, por otro lado, el esquema **RERCOM3e+p**, obtiene los mejores resultados en condiciones de mucho ruido (0, -5dB), obteniendo el mejor rendimiento “Media2”.

RUIDO 2	Limpia	20 dB	15dB	10 dB	5 dB	0 dB	-5 dB	Media1	Media2
BASELINE	98,49	96,67	93,80	86,40	70,25	49,27	<b>31,68</b>	79,278	71,345
RERCOM2e	98,58	96,77	<b>94,50</b>	<b>87,61</b>	71,49	46,01	24,30	79,276	70,113
RERCOM2e+p	<b>98,70</b>	96,67	94,35	86,82	71,64	45,80	25,06	79,056	70,057
RERCOM3e+p	98,43	96,19	93,17	84,64	68,26	44,71	23,58	77,394	68,425
ADVFRONT	98,37	<b>96,81</b>	93,78	87,58	<b>74,24</b>	<b>53,72</b>	31,50	<b>81,226</b>	<b>72,938</b>

Tabla 7.128: Porcentajes de precisión de reconocimiento de palabra para el ruido tipo 2 (voz/babble).

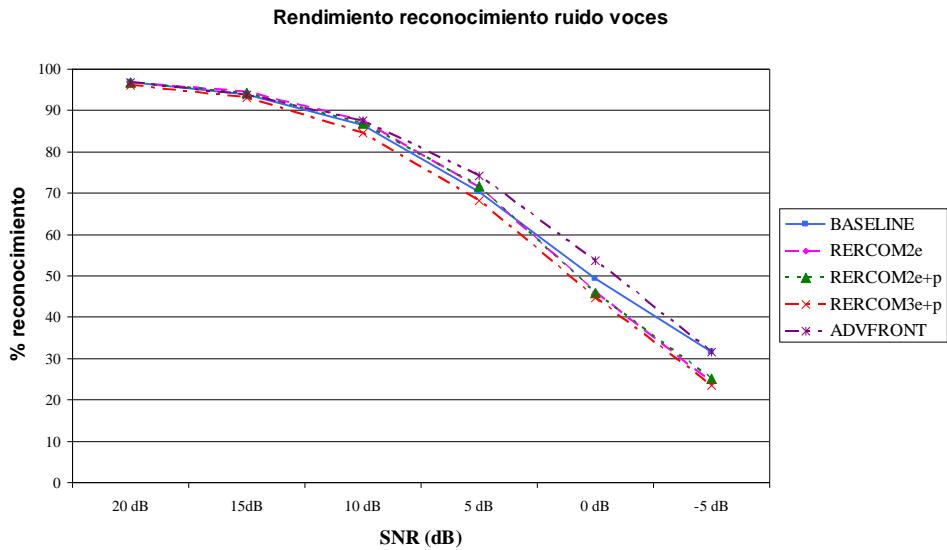


Gráfico 7.29: Rendimiento de reconocimiento para ruido de Voces

Los resultados obtenidos con este tipo de ruido nos indican que el sistema RERCOM es poco robusto frente a este tipo de ruido, obteniendo valores similares al front-end de referencia en condiciones de poco ruido (20 → 5 dB). Frente a este tipo de ruido, **ADVFRONT** obtiene los mejores resultados, tanto en “Media1” como “Media2”.

RUIDO 3	Limpia	20 dB	15dB	10 dB	5 dB	0 dB	-5 dB	Media1	Media2
BASELINE	<b>98,48</b>	98,03	97,26	94,78	88,22	66,51	30,63	88,960	79,238
RERCOM2e	98,36	97,97	<b>97,64</b>	95,82	91,17	78,17	51,45	92,154	85,370
RERCOM2e+p	98,45	97,76	97,55	96,00	<b>92,66</b>	80,79	56,13	<b>92,952</b>	86,815
RERCOM3e+p	98,24	97,64	96,75	95,88	92,19	<b>81,78</b>	<b>59,08</b>	92,848	<b>87,220</b>
ADVFRONT	98,09	<b>98,18</b>	<b>97,64</b>	<b>96,12</b>	91,08	76,95	42,98	91,994	83,825

Tabla 7.129: Porcentajes de precisión de reconocimiento de palabra para el ruido tipo 3 (metro/suburban train).

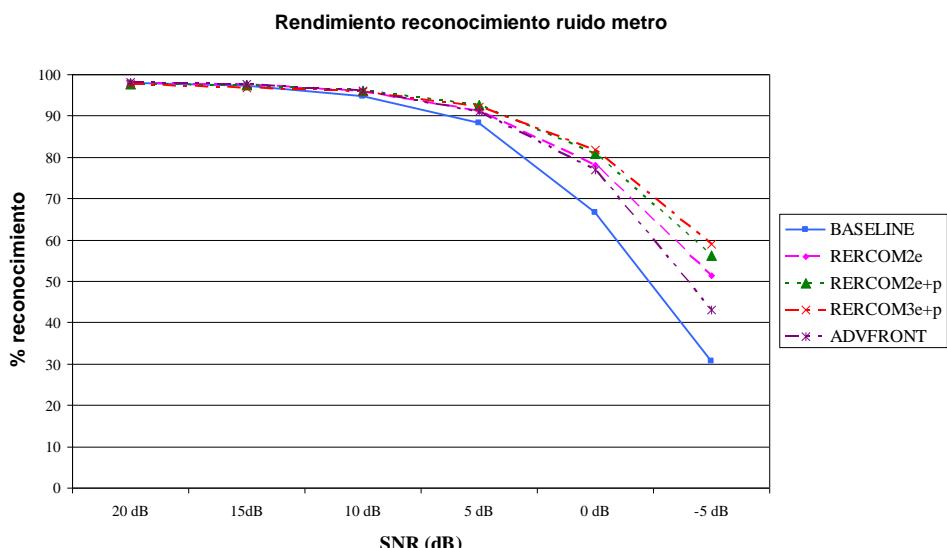


Gráfico 7.30: Rendimiento de reconocimiento para ruido de Metro

Los resultados de la tabla para el ruido tipo 3 nos indican lo mismo que para el ruido tipo 1, el esquema de reducción de ruido que obtiene un mejor rendimiento global es la estructura **RERCOM2e+p**, obteniendo el mejor resultado “Media1”, por otro lado, el esquema **RERCOM3e+p**, obtiene los mejores resultados en condiciones de mucho ruido (0, -5dB), obteniendo el mejor rendimiento “Media2”.

RUIDO 4	Limpia	20 dB	15dB	10 dB	5 dB	0 dB	-5 dB	Media1	Media2
BASELINE	98,64	98,43	98,12	97,59	94,72	79,67	47,24	93,706	85,962
RERCOM2e	98,89	98,55	98,33	97,84	97,10	93,37	81,92	97,038	94,518
RERCOM2e+p	98,95	98,64	98,40	97,87	96,95	93,18	81,33	97,008	94,395
RERCOM3e+p	98,67	98,24	97,81	96,85	95,80	91,82	79,82	96,104	93,390
ADVFRONT	98,86	98,86	98,58	97,96	96,36	90,99	70,84	96,550	92,265

Tabla 7.130: Porcentajes de precisión de reconocimiento de palabra para el ruido tipo 4 (coche/car).

Rendimiento reconocimiento ruido coche

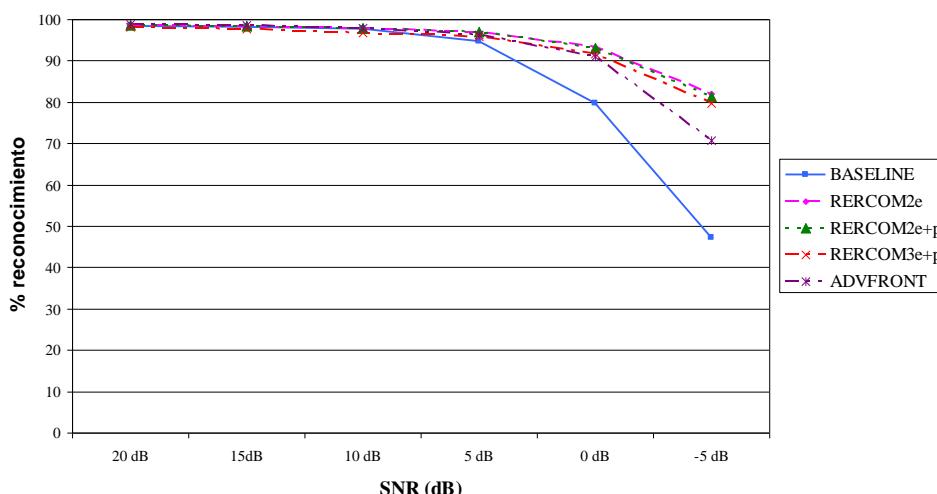


Gráfico 7.31: Rendimiento de reconocimiento para ruido de Coche

La tabla para el ruido tipo 4 nos muestra una variación en sus resultados respecto a los 3 ruidos anteriores; los mejores rendimientos, tanto en “Media1” como en “Media2”, los obtiene la variante de **2 etapas del sistema RERCOM**.

	RUIDO 1	RUIDO 2	RUIDO 3	RUIDO 4	MEDIA
BASELINE	83,406	79,278	88,960	93,706	86,3375
RERCOM2e	89,906	79,276	92,154	97,038	89,5935
RERCOM2e+p	90,458	79,056	92,952	97,008	89,8685
RERCOM3e+p	90,108	77,394	92,848	96,104	89,1135
ADVFRONT	89,750	81,226	91,994	96,550	89,8800

Tabla 7.131: Porcentajes de precisión de reconocimiento de palabra finales utilizando el promedio “Media1”

En la tabla anterior observamos que a pesar de que el esquema RERCOM2e+p obtiene los mejores resultados en 3 de los cuatro tipos de ruido, **el mejor rendimiento global lo obtiene ADVFRONT, superando en 0,0115% al esquema RERCOM2e+p.** Este hecho es provocado por el bajo rendimiento, menor que el esquema BASELINE, que se obtiene con el programa RERCOM frente al ruido tipo 2 (Voces), un tipo de ruido que no se ha tenido en cuenta en el diseño del programa.

	RUIDO 1	RUIDO 2	RUIDO 3	RUIDO 4	MEDIA
BASELINE	72,607	71,345	79,238	85,962	77,2879
RERCOM2e	81,870	<b>70,113</b>	85,370	<b>94,518</b>	82,9679
RERCOM2e+p	82,535	70,057	86,815	94,395	<b>83,4504</b>
RERCOM3e+p	<b>82,720</b>	68,425	<b>87,220</b>	93,390	82,9388
ADVFRONT	80,600	72,938	83,825	92,265	82,4071

Tabla 7.132: Porcentajes de precisión de reconocimiento de palabra finales utilizando el promedio “Media2”

En la tabla observamos que teniendo en cuenta los resultados obtenidos a muy baja SNR (-5dB), la disminución de rendimiento que experimenta el esquema ADVFRONT supera la penalización que añadía el ruido tipo 2 en el esquema **RERCOM2e+p, obteniendo los mejores resultados de precisión de reconocimiento, superando en algo más del 1% al esquema ADVFRONT.**



## 8.- Implementación y pruebas del programa RERCOM\_DSP.

### 8.1- Implementación RERCOM\_DSP.

En este apartado trataremos de explicar las partes que componen el programa en C, que hemos bautizado como RERCOM\_DSP. Este programa es una versión optimizada del simulador que realiza un filtrado similar al programa de simulación RERCOM. La diferencia básica entre ambos es la supresión de transformadas de fourier innecesarias, y la eliminación de algunas funciones prescindibles. De esta manera hemos obtenido un programa sin parámetros de entrada y sin un control de la evolución de la forma de onda en el dominio temporal, con solamente las capacidades de filtrado que hemos considerado óptimas a partir de las pruebas que hemos realizado en el capítulo 7 con del programa RERCOM de simulación, ponderando su relación de aumento de rendimiento / aumento de potencia de calculo. El resultado es un programa 10 veces más rápido que su versión simulador, convirtiéndolo en un respetable competidor del Standard de la ETSI AdvanceFrontEnd.

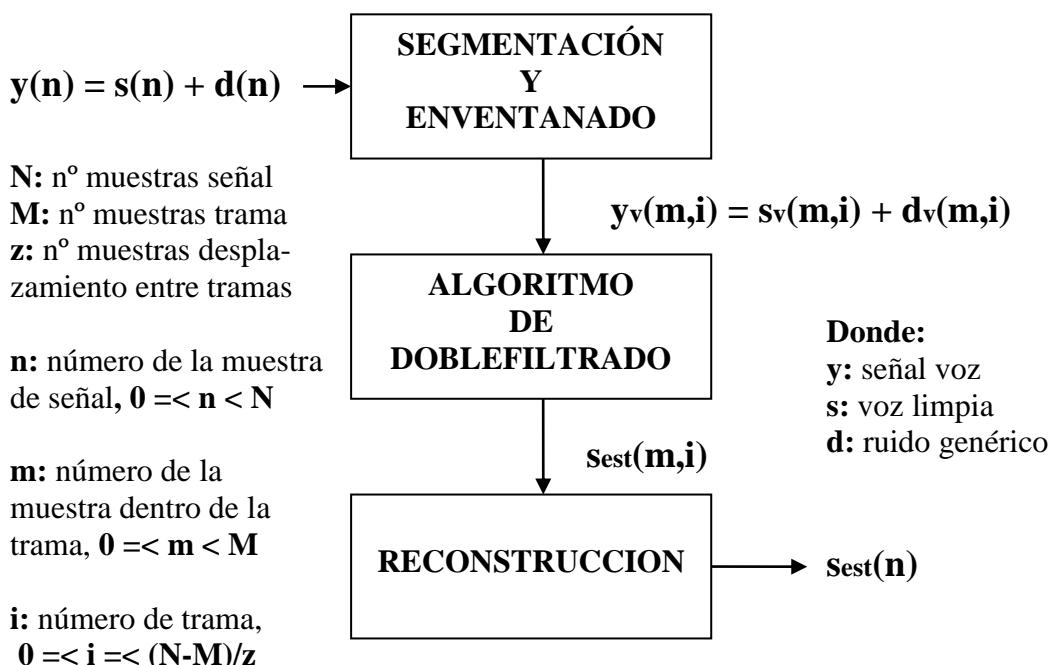


Fig. 8.1: esquema de filtrado utilizado en el programa RERCOM\_DSP

La figura 8.1, nos muestra un diagrama de bloques que utiliza el programa RERCOM\_DSP, este sistema realiza el filtrado en una sola etapa a partir de un algoritmo optimizado basado en una sustracción espectral y un filtrado de wiener en cascada.

### 8.1.1.- Segmentación y enventanado.

El bloque que hemos llamado segmentación y enventanado es la unión de los bloques segmentación del apartado 6.1.1. y enventanado del apartado 6.1.3.1.. En nuestro algoritmo se ha impuesto un desplazamiento de trama del 50% y se ha eliminado la posibilidad de realizar un alargamiento virtual del numero de muestras de una trama a partir de su extensión periódica.

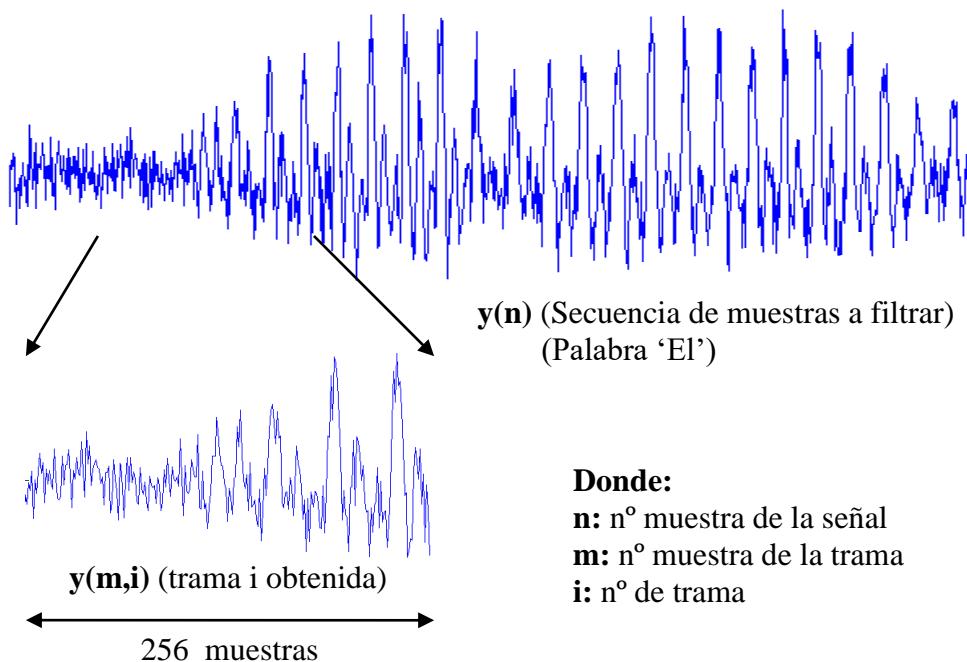


Fig. 6.49: Obtención de una trama (256 muestras) de señal a procesar. La trama corresponde al inicio del fonema /e/ del fichero ASUN1 + ruido blanco (SNR = 9 dB)

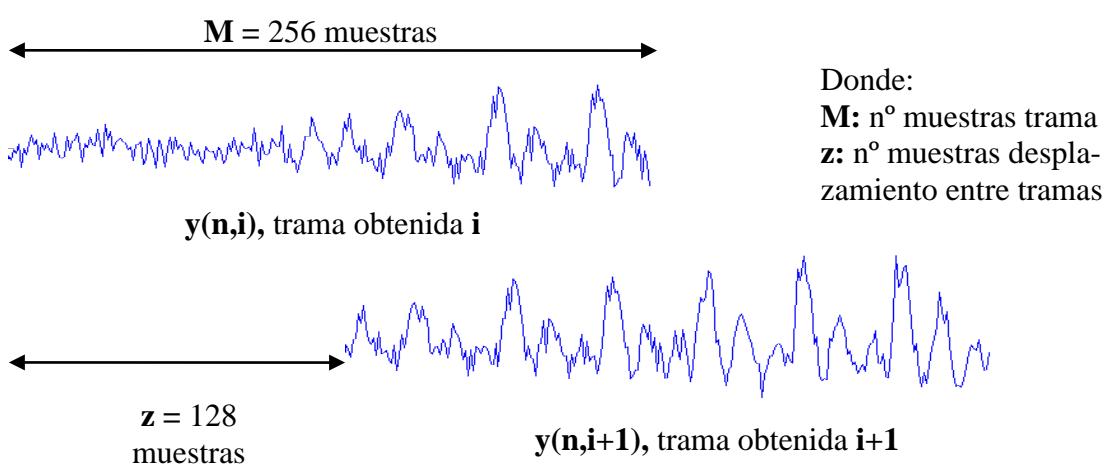


Fig. 8.2: Tramas de señal obtenidas utilizando un solapamiento del 50%. Las tramas corresponden al fonema /e/ del fichero ASUN1 con ruido blanco (SNR = 9 dB)

El enventanado de la trama de entrada  $y(m,i)$  se realiza mediante las ecuaciones (6.6) y (6.7)

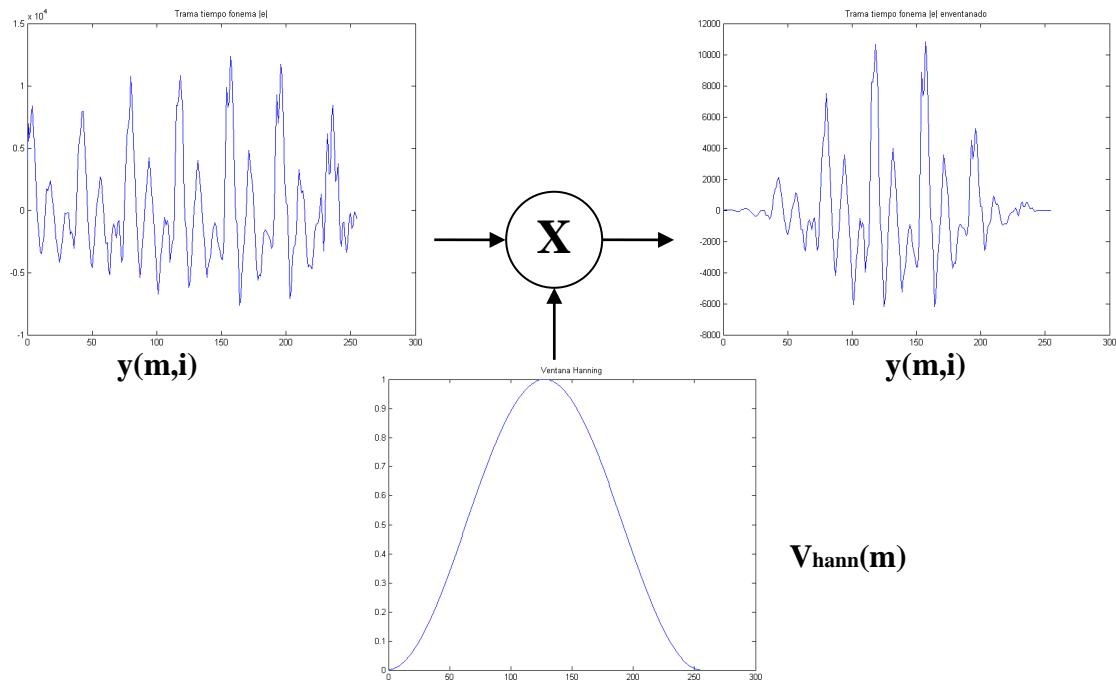


Fig. 8.3: Proceso de enventanado de una trama de señal obtenida, la trama corresponde al fonema /e/ del fichero ASUN1.

### 8.1.2.- Algoritmo de doblefiltrado

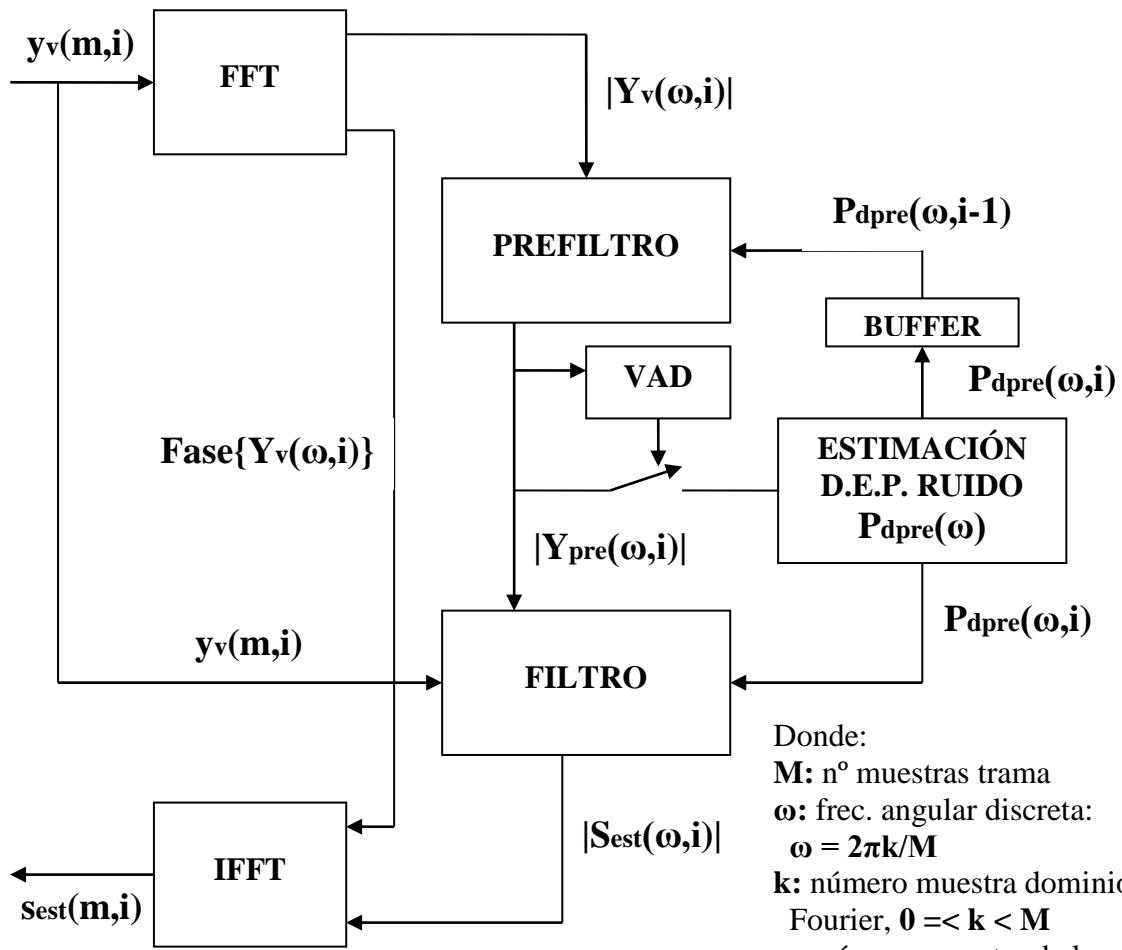
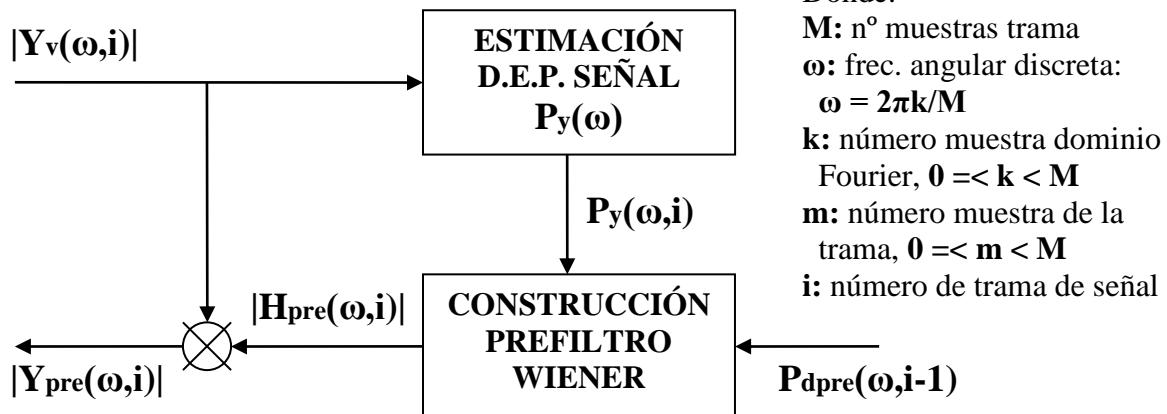


Fig 8.4: Diagrama de bloques del algoritmo de doblefiltrado

La figura 8.4 muestra el diagrama de bloques del algoritmo de doble filtrado, observamos que este algoritmo esconde un filtrado de dos etapas, una primera etapa, prefiltro, basada en un filtrado de sustracción espectral y una segunda etapa, filtro, basada en un filtrado de Wiener con modelado AR y filtrado peine.

Este algoritmo es menos robusto que realizar el filtrado en dos etapas, ya que el modelado AR y la detección de pitch se realizan a partir de la trama temporal original, no obstante permite reducir el número de transformadas de Fourier a sólo dos, siendo necesario una menor potencia de cálculo.

### 8.1.2.1.- Prefiltro



Donde:  
**M:** n° muestras trama  
**ω:** frec. angular discreta:  
 $\omega = 2\pi k/M$   
**k:** número muestra dominio Fourier,  $0 \leq k < M$   
**m:** número muestra de la trama,  $0 \leq m < M$   
**i:** número de trama de señal

Fig. 8.5: Diagrama de bloques del módulo prefiltro

La figura 8.5 nos muestra las operaciones que realiza el prefiltro, observamos que es un esquema similar al prefiltro del punto 6.3, pero con un buen número de simplificaciones, ya que el filtrado se realiza a partir de una trama enventanada y transformada al dominio de Fourier.

El bloque “Estimación DEE señal” es un bloque similar al del punto 6.3.2, pero suprimiendo el módulo FFT. El bloque “Construcción Prefiltro Wiener” es idéntico al del punto 6.3.3.

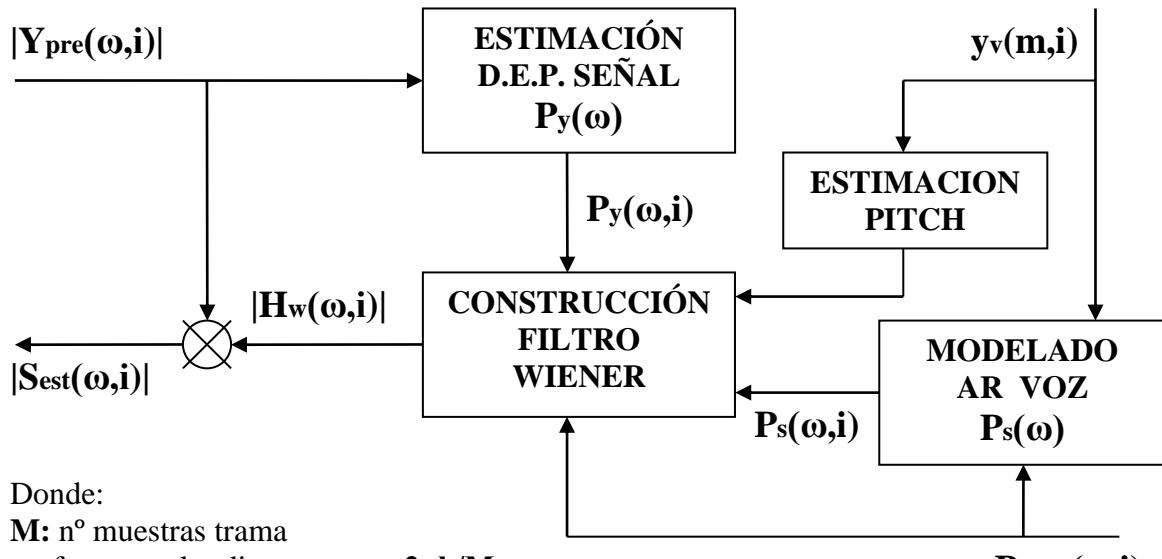
### 8.1.2.2.- VAD

El modulo VAD es un detector de actividad de voz idéntico al del punto 6.4.4, se utiliza para indicar al sistema que debe realizar la actualización de la estimación de la DEE de ruido necesaria para la construcción de los filtros.

### 8.1.2.3.- Estimación DEE ruido.

El bloque “Estimación DEE ruido” es similar a la estimación que se realiza en el punto 6.2, pero a partir de una trama enventanada y en el dominio de Fourier, pudiendo eliminar los módulos “enventanado” y “FFT” del punto 6.2.

### 8.1.2.4.- Filtro



Donde:

**M:** n° muestras trama

**ω:** freq. angular discreta:  $\omega = 2\pi k/M$

**k:** número muestra dominio Fourier,  $0 \leq k < M$

**m:** número muestra de la trama,  $0 \leq m < M$

**i:** número de trama de señal

Fig. 8.6: Diagrama de bloques del módulo filtro

La figura 8.6 muestra el diagrama de bloques del filtro, como en los anteriores bloques se parte de una trama enventanada y en el dominio de fourier, con lo cual se obtienen ventajas con la reducción del número de cálculos a efectuar.

El módulo “Estimación DEE señal” es la versión simplificada del bloque del punto 6.4.2. El modulo “Estimación pitch” es idéntico al del punto 6.4.5. El bloque “Modelado AR Voz” es también idéntico al del punto 6.4.3. Y finalmente el bloque “Construcción Filtro Wiener” es lo mismo que el bloque del punto 6.4.6.

### 8.1.3.- Reconstrucción señal

En la implementación de este sistema se ha utilizado un desplazamiento de trama del 50%, por lo que para una longitud de trama de 256 muestras, el desplazamiento entre tramas será de 128 muestras. La figura 8.7 muestra la forma en que se realiza esta reconstrucción.

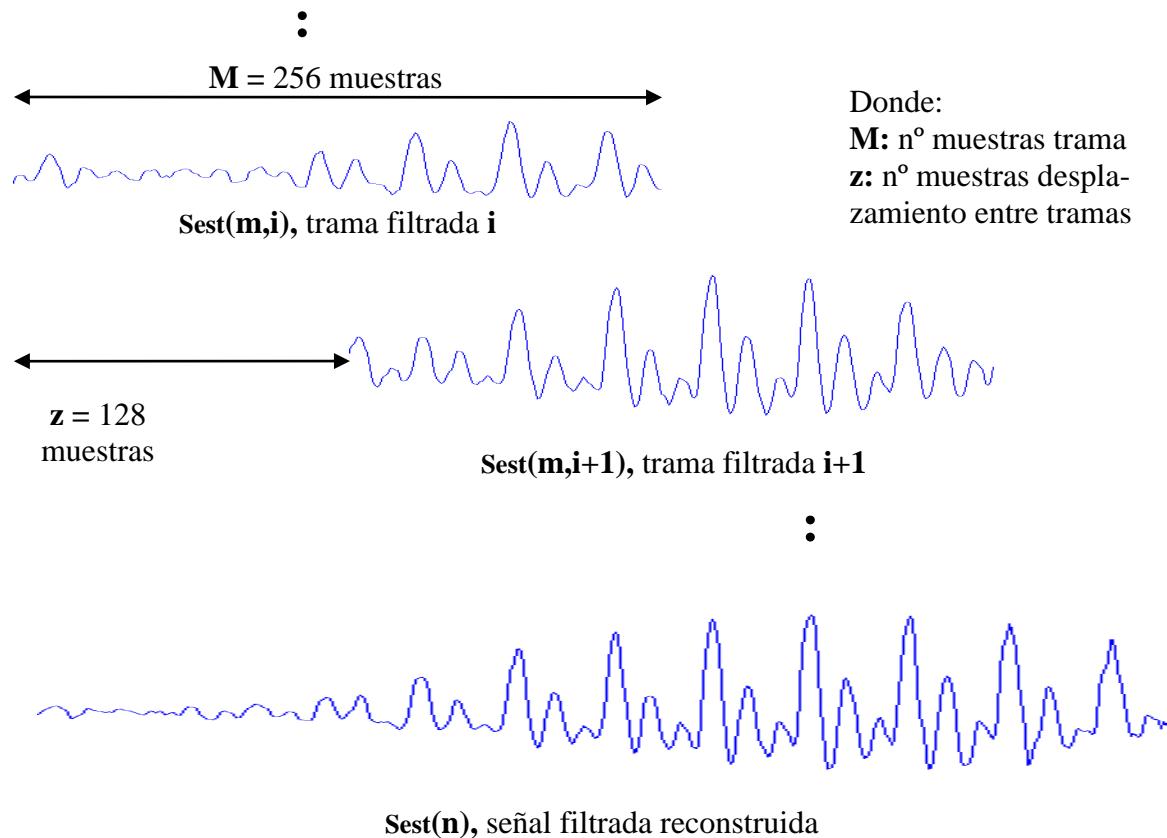


Fig. 8.7: Reconstrucción de parte del fonema /e/ a partir de diferentes tramas filtradas

## 8.2 - Comparativa RERCOM\_DSP vs Advance Front-End

En este capítulo, al igual que los apartados 7.3 y 7.4 realizaremos una comparativa del programa optimizado RERCOM\_DSP, creado a partir de los resultados obtenidos con el programa simulador RERCOM, con el estándar de la ETSI, Advance Front-End.

### 8.2.1.-Comparativa de medidas objetivas

Al igual que en el apartado 7.3, testearemos la calidad del sistema RERCOM\_DSP frente a fuentes perturbadoras consistentes no sólo en ruido aditivo Gaussiano blanco (AWGN) o rosa, sino también con ruidos reales como el que producen diversos tipos de motores o ambientes normales de trabajo.

En esta comparativa nos centraremos en ambiente con ruido ( $\text{SNR} = 9 \text{ dB}$ ) y mucho ruido ( $\text{SNR} = 0 \text{ dB}$ ), descartando los ambientes con  $\text{SNR}$  igual  $18 \text{ dB}$  ya que los consideramos poco relevantes. Y descartando también los ambientes con  $\text{SNR} = -6 \text{ dB}$ , puesto que es un nivel de ruido poco habitual, ya que supondría que el nivel de ruido es mucho mas alto que el de voz.

#### 8.2.1.1.-Ruidos de banda ancha

En este apartado nos centramos en las perturbaciones provocadas por ruidos de banda ancha, el ruido blanco y el ruido rosa. Realizamos la evaluación para  $\text{SNR}$  de  $9 \text{ dB}$  y  $0 \text{ dB}$ , para los ficheros ASUN1 y ESCA.

##### 8.2.1.1.1.-Ruido blanco

Los resultados obtenidos para el fichero de voz ASUN1 son:

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	9,061	1,522	1,745	2,718	3,500	3,192	0,000
RERCOM	16,177	6,983	0,808	2,152	1,417	1,838	-0,223
ADVANCE	12,939	5,364	1,031	2,262	1,653	2,304	2,388

Tabla 8.1: Comparativa RERCOM vs ADVANCE, con  $\text{SNR}=9 \text{ dB}$ .

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	0,061	-3,918	2,523	4,266	5,146	3,787	0,000
RERCOM	9,801	2,731	1,386	2,720	2,047	2,570	0,316
ADVANCE	7,380	1,073	1,843	2,480	2,441	3,118	3,576

Tabla 8.2: Comparativa RERCOM vs ADVANCE, con  $\text{SNR}=0 \text{ dB}$ .

Los resultados obtenidos para el fichero de voz ESCA son:

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	9,031	2,262	3,195	4,548	5,390	4,659	0,000
RERCOM	15,568	8,826	2,135	3,033	2,708	3,301	-0,123
ADVANCE	12,573	7,071	2,106	2,891	2,824	3,620	3,068

Tabla 8.3: Comparativa RERCOM vs ADVANCE, con SNR=9 dB.

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	0,023	-3,563	4,120	6,179	7,146	5,148	0,000
RERCOM	9,730	4,387	2,799	4,092	3,472	3,913	0,158
ADVANCE	6,986	1,320	3,022	3,707	4,140	4,439	4,308

Tabla 8.4: Comparativa RERCOM vs ADVANCE, con SNR=0 dB.

### 8.2.1.1.2.-Ruido rosa

Los resultados obtenidos para el fichero de voz ASUN1 son:

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	8,899	1,465	1,309	2,308	3,065	2,586	0,000
RERCOM	14,925	5,966	0,666	1,906	1,301	1,792	-0,088
ADVANCE	9,908	3,448	0,852	2,259	1,454	1,944	2,504

Tabla 8.5: Comparativa RERCOM vs ADVANCE, con SNR=9 dB.

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	0,199	-3,828	1,939	3,776	4,648	3,142	0,000
RERCOM	7,694	1,458	1,281	3,273	1,786	2,437	-0,302
ADVANCE	4,991	-0,194	1,625	2,828	2,057	2,779	3,576

Tabla 8.6: Comparativa RERCOM vs ADVANCE, con SNR=0 dB.

Los resultados obtenidos para el fichero de voz ESCA son:

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	9,081	2,358	2,498	4,062	4,897	3,889	0,000
RERCOM	14,376	7,602	1,579	2,511	2,355	2,814	-0,204
ADVANCE	11,716	6,059	1,715	2,629	2,380	3,043	2,874

Tabla 8.7: Comparativa RERCOM vs ADVANCE, con SNR=9 dB.

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	-0,020	-3,545	3,254	5,738	6,685	4,366	0,000
RERCOM	7,503	2,792	2,176	3,483	3,018	3,441	-0,384
ADVANCE	5,283	0,534	2,545	3,395	3,372	3,844	3,420

Tabla 9.8: Comparativa RERCOM vs ADVANCE, con SNR=0 dB.

### 8.2.1.2.-Ruidos de motor

En este apartado nos centramos en las perturbaciones provocadas por ruidos de banda estrecha, como son los ruidos generados por motores, trataremos 4 casos, el ruido de motor de avión f16, el ruido de un motor genérico y dos ruidos provocados por el motor de un coche. Realizamos la evaluación para SNR de 9 dB y 0 dB, para los ficheros ASUN1 y ESCA.

#### 8.2.1.2.1.-Ruido de motor de avión

Los resultados obtenidos para el fichero de voz ASUN1 son:

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	9,159	1,664	1,037	1,867	2,576	2,609	0,000
RERCOM	15,405	6,385	0,580	1,937	1,163	1,740	-0,212
ADVANCE	9,535	3,451	0,689	2,393	1,362	1,973	1,963

Tabla 8.9: Comparativa RERCOM vs ADVANCE, con SNR=9 dB.

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	0,258	-3,724	1,589	3,284	4,142	3,200	0,000
RERCOM	8,285	1,810	0,971	2,833	1,709	2,273	-0,567
ADVANCE	4,773	-0,117	1,205	2,669	1,733	2,748	2,443

Tabla 8.10: Comparativa RERCOM vs ADVANCE, con SNR=0 dB.

Los resultados obtenidos para el fichero de voz ESCA son:

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	9,069	2,438	2,291	3,629	4,436	3,949	0,000
RERCOM	13,806	7,240	1,383	2,311	2,044	2,763	-0,401
ADVANCE	11,144	5,694	1,591	2,708	2,145	3,224	2,349

Tabla 8.11: Comparativa RERCOM vs ADVANCE, con SNR=9 dB.

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	-0,030	-3,456	3,050	5,271	6,172	4,412	0,000
RERCOM	6,940	2,390	2,132	3,733	2,674	3,479	-0,777
ADVANCE	4,983	0,434	2,442	3,751	3,027	3,965	2,639

Tabla 8.12: Comparativa RERCOM vs ADVANCE, con SNR=0 dB.

### 8.2.1.2.2.-Ruido de un motor genérico

Los resultados obtenidos para el fichero de voz ASUN1 son:

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	8,871	1,524	0,574	1,163	1,776	1,534	0,000
RERCOM	15,494	6,515	0,630	3,032	1,242	1,463	-0,158
ADVANCE	10,018	3,568	0,434	2,186	1,110	1,393	2,403

Tabla 8.13: Comparativa RERCOM vs ADVANCE, con SNR=9 dB.

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	-0,029	-3,866	1,001	2,383	3,194	2,058	0,000
RERCOM	8,956	1,812	0,851	2,739	1,458	1,804	-0,285
ADVANCE	4,824	-0,207	0,795	2,000	1,334	1,844	3,376

Tabla 8.14: Comparativa RERCOM vs ADVANCE, con SNR=0 dB.

Los resultados obtenidos para el fichero de voz ESCA son:

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	8,854	2,305	1,247	2,612	3,390	2,495	0,000
RERCOM	14,172	7,518	0,786	1,759	1,443	1,890	-0,150
ADVANCE	10,761	5,679	0,795	2,105	1,399	1,961	2,452

Tabla 8.15: Comparativa RERCOM vs ADVANCE, con SNR=9 dB.

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	0,153	-3,347	1,799	4,134	5,020	2,929	0,000
RERCOM	7,194	2,192	1,321	2,566	2,109	2,462	-0,702
ADVANCE	4,815	0,797	1,346	2,376	2,078	2,517	2,694

Tabla 8.16: Comparativa RERCOM vs ADVANCE, con SNR=0 dB.

### 8.2.1.2.3.-Ruido de motor de coche

Los resultados obtenidos para el fichero de voz ASUN1 son:

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	9,182	2,396	0,207	0,306	0,582	1,156	0,000
RERCOM	18,598	9,268	0,862	6,128	1,959	2,046	-0,107
ADVANCE	9,896	4,233	0,238	1,326	0,820	1,170	1,362

Tabla 8.17: Comparativa RERCOM vs ADVANCE, con SNR=9 dB.

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	-0,018	-3,316	0,462	0,809	1,300	1,885	0,000
RERCOM	10,840	4,145	0,924	6,734	2,369	2,185	-0,226
ADVANCE	4,809	0,140	0,498	3,759	1,501	1,930	2,849

Tabla 8.18: Comparativa RERCOM vs ADVANCE, con SNR=0 dB.

Los resultados obtenidos para el fichero de voz ESCA son:

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	9,139	3,081	0,413	0,951	1,483	1,535	0,000
RERCOM	16,873	10,072	0,594	4,555	1,561	1,793	-0,182
ADVANCE	10,799	5,480	0,457	2,666	1,227	1,533	1,934

Tabla 8.19: Comparativa RERCOM vs ADVANCE, con SNR=9 dB.

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	-0,162	-2,986	0,848	2,035	2,751	2,252	0,000
RERCOM	10,529	4,948	0,655	3,596	1,509	1,764	-0,210
ADVANCE	4,663	0,586	0,688	2,740	1,381	1,992	2,049

Tabla 8.20 : Comparativa RERCOM vs ADVANCE, con SNR=0 dB.

#### 8.2.1.2.4.-Ruido de motor de coche (musical)

Los resultados obtenidos para el fichero de voz ASUN1 son:

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	9,548	2,629	0,165	0,212	0,428	1,091	0,000
RERCOM	20,208	10,200	0,947	6,243	1,989	2,126	-0,074
ADVANCE	10,856	4,213	0,179	0,643	0,598	1,067	0,768

Tabla 8.21: Comparativa RERCOM vs ADVANCE, con SNR=9 dB.

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	0,347	-3,076	0,379	0,572	0,978	1,834	0,000
RERCOM	12,988	5,285	1,033	6,321	2,461	2,405	-0,117
ADVANCE	4,517	-0,378	0,388	2,894	1,309	1,717	0,482

Tabla 8.22: Comparativa RERCOM vs ADVANCE, con SNR=0 dB.

Los resultados obtenidos para el fichero de voz ESCA son:

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	9,156	3,164	0,347	0,661	1,089	1,306	0,000
RERCOM	17,267	10,792	1,013	6,483	2,087	3,073	-0,160
ADVANCE	10,949	5,499	0,429	2,267	1,021	1,398	1,002

Tabla 8.23: Comparativa RERCOM vs ADVANCE, con SNR=9 dB.

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	0,156	-2,767	0,660	1,493	2,136	2,005	0,000
RERCOM	10,743	5,675	1,116	7,890	2,624	2,944	-0,319
ADVANCE	3,850	0,078	0,633	2,277	1,309	1,897	1,233

Tabla 8.24: Comparativa RERCOM vs ADVANCE, con SNR=0 dB.

### 8.2.1.3.-Ruidos de diferentes ambientes reales

En este apartado nos centramos en las perturbaciones provocadas por ruidos reales que dependen del ambiente en que se produce la comunicación y afectan negativamente a ésta. Trataremos el ruido que existe en el entorno de una fábrica, el ruido que produce el tráfico y el ruido que existirá en un tren o sus proximidades. Realizamos la evaluación para SNR de 9 dB, 0 dB y -6 dB, para los ficheros ASUN1 y ESCA.

#### 8.2.1.3.1.-Ruido de fábrica

Los resultados obtenidos para el fichero de voz ASUN1 son:

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	9,049	1,826	0,625	1,214	1,824	1,679	0,000
RERCOM	16,016	6,798	0,610	3,039	1,251	1,485	-0,086
ADVANCE	10,731	3,666	0,502	1,278	1,038	1,444	2,708

Tabla 8.25: Comparativa RERCOM vs ADVANCE, con SNR=9 dB.

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	-0,052	-3,658	1,094	2,449	3,250	2,265	0,000
RERCOM	8,354	1,729	0,908	2,771	1,511	1,937	0,082
ADVANCE	4,650	-0,809	1,056	1,892	1,692	2,181	4,368

Tabla 8.26: Comparativa RERCOM vs ADVANCE, con SNR=0 dB.

Los resultados obtenidos para el fichero de voz ESCA son:

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	9,046	2,401	1,352	2,676	3,443	2,614	0,000
RERCOM	14,556	7,916	0,790	2,042	1,436	1,840	-0,122
ADVANCE	10,274	5,117	0,982	1,910	1,615	2,161	2,933

Tabla 8.27: Comparativa RERCOM vs ADVANCE, con SNR=9 dB.

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	-0,055	-3,438	1,945	4,236	5,115	3,074	0,000
RERCOM	7,326	2,468	1,368	2,723	2,149	2,408	-0,063
ADVANCE	3,986	-0,292	1,721	2,824	2,594	2,884	4,067

Tabla 8.28: Comparativa RERCOM vs ADVANCE, con SNR=0 dB.

### 8.2.1.3.2.-Ruido de tráfico

Los resultados obtenidos para el fichero de voz ASUN1 son:

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	9,027	1,899	0,498	0,857	1,381	1,533	0,000
RERCOM	16,831	7,452	0,959	5,142	1,506	1,774	0,038
ADVANCE	10,634	4,086	0,402	2,095	1,033	1,343	1,943

Tabla 8.29: Comparativa RERCOM vs ADVANCE, con SNR=9 dB.

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	-0,374	-3,857	0,977	1,937	2,686	2,155	0,000
RERCOM	8,806	2,245	0,994	3,760	1,633	1,929	-0,124
ADVANCE	4,330	-0,347	0,715	2,454	1,298	1,907	2,689

Tabla 8.30: Comparativa RERCOM vs ADVANCE, con SNR=0 dB.

Los resultados obtenidos para el fichero de voz ESCA son:

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	9,362	2,996	1,019	1,993	2,709	2,325	0,000
RERCOM	15,292	8,234	0,800	2,606	1,419	1,892	-0,195
ADVANCE	11,196	5,674	0,693	2,016	1,275	1,839	2,366

Tabla 8.31: Comparativa RERCOM vs ADVANCE, con SNR=9 dB.

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	0,261	-2,970	1,607	3,443	4,294	2,849	0,000
RERCOM	8,135	2,916	1,168	2,650	1,891	2,318	-0,223
ADVANCE	4,574	0,355	1,208	2,304	1,944	2,431	2,579

Tabla 8.32: Comparativa RERCOM vs ADVANCE, con SNR=0 dB.

### 8.2.1.3.3.-Ruido de tren

Los resultados obtenidos para el fichero de voz ASUN1 son:

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	9,116	1,707	0,520	0,926	1,464	1,556	0,000
RERCOM	15,609	6,243	0,557	1,601	0,978	1,333	-0,119
ADVANCE	9,783	3,303	0,401	1,391	0,929	1,345	1,852

Tabla 8.33: Comparativa RERCOM vs ADVANCE, con SNR=9 dB.

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	-0,085	-3,929	0,968	2,008	2,764	2,167	0,000
RERCOM	7,842	0,861	0,861	2,217	1,421	1,751	-0,320
ADVANCE	4,211	-0,822	0,814	2,139	1,381	1,966	2,293

Tabla 8.34: Comparativa RERCOM vs ADVANCE, con SNR=0 dB.

Los resultados obtenidos para el fichero de voz ESCA son:

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	8,748	2,194	1,162	2,304	3,036	2,595	0,000
RERCOM	14,344	6,987	0,719	1,600	1,365	1,809	-0,272
ADVANCE	10,505	4,868	0,787	1,605	1,455	2,105	2,287

Tabla 8.35: Comparativa RERCOM vs ADVANCE, con SNR=9 dB.

	SNR(dB)	SNRs(dB)	LLR	IS	CEPS	LAR	AMP(dB)
ORIGINAL	-0,254	-3,657	1,709	3,751	4,606	3,006	0,000
RERCOM	7,476	1,648	1,246	2,292	2,204	2,338	-0,465
ADVANCE	3,927	-0,500	1,356	2,171	2,339	2,723	2,781

Tabla 8.36: Comparativa RERCOM vs ADVANCE, con SNR=0 dB.

### 8.2.2.-Comparativa de reconocimiento

La prueba de reconocimiento de voz se ha realizado un entrenamiento “multi-condition Training” del sistema de reconocimiento según [Pear-98].

Para realizar este entrenamiento se han filtrado (FrontEnd) los 8440 archivos de entrenamiento de TIdigits, estos archivos de entrenamiento, creados por 110 locutores diferentes (55 hombres y 55 mujeres), se distribuyen en diferentes conjuntos en función del tipo de ruido y de su SNR según la siguiente tabla:

<b>SNR</b>	<b>RUIDO1</b>	<b>RUIDO2</b>	<b>RUIDO3</b>	<b>RUIDO4</b>
Limpia	422	422	422	422
20 dB	422	422	422	422
15 dB	422	422	422	422
10 dB	422	422	422	422
5 dB	422	422	422	422

Donde los ruidos añadidos artificialmente son:

- Ruido 1: Sala de exhibición (Exhibition hall).
- Ruido 2: Voces (Babble noise (big room, office, people chatting)).
- Ruido 3: Metro (suburban train).
- Ruido 4: Coche (Car moving).

Una vez entrenado el sistema de reconocimiento se obtienen los resultados de rendimiento de reconocimiento filtrando (FrontEnd) los 28028 archivos de test distribuidos en 7 niveles de SNR diferentes, para cada nivel de reconocimiento es necesario analizar 4004 archivos (1001 por cada ruido), según se indica en la siguiente tabla:

<b>SNR</b>	<b>RUIDO1</b>	<b>RUIDO2</b>	<b>RUIDO3</b>	<b>RUIDO4</b>
Limpia	1001	1001	1001	1001
20 dB	1001	1001	1001	1001
15 dB	1001	1001	1001	1001
10 dB	1001	1001	1001	1001
5 dB	1001	1001	1001	1001
0 dB	1001	1001	1001	1001
-5 dB	1001	1001	1001	1001

Los rendimientos de reconocimiento están expresados en tanto por ciento de precisión de palabra según la ecuación:

$$\text{Precision\_palabra} = 100 * (\text{nº palabras acertadas} - 0.5 * \text{nº palabras insertadas}) / \text{nº palabras totales}$$

Resultados utilizando el FrontEnd de referencia (BASELINE)

<b>SNR</b>	<b>RUIDO1</b>	<b>RUIDO2</b>	<b>RUIDO3</b>	<b>RUIDO4</b>	<b>PROMEDIO</b>
Limpia	98.65	98.49	98.48	98.64	98.5650
20 dB	97.60	96.67	98.03	98.43	97.6825
15 dB	96.16	93.80	97.26	98.12	96.3350
10 dB	92.85	86.40	94.78	97.59	92.9050
5 dB	83.11	70.25	88.22	94.72	84.0750
0 dB	47.31	49.27	66.51	79.67	60.6900
-5 dB	18.61	31.68	30.63	47.24	32.0400

**Rendimiento promedio 20 dB a 0 dB: 86.3375 %**

**Rendimiento promedio 20 dB a -5 dB: 77.2879 %**

Resultados utilizando como FrontEnd el programa RERCOM DSP v0.4

<b>SNR</b>	<b>RUIDO1</b>	<b>RUIDO2</b>	<b>RUIDO3</b>	<b>RUIDO4</b>	<b>PROMEDIO</b>
Limpia	98.98	98.78	98.47	99.03	98.8150
20 dB	98.00	96.86	98.06	98.70	97.9050
15 dB	97.05	94.32	97.58	98.49	96.8600
10 dB	94.69	87.64	96.21	98.03	94.1425
5 dB	89.44	71.98	91.95	97.16	87.6325
0 dB	72.83	46.64	79.57	93.49	73.1325
-5 dB	40.25	24.58	52.40	79.23	49.1150

**Rendimiento promedio 20 dB a 0 dB: 89.9345 %**

**Rendimiento promedio 20 dB a -5 dB: 83.1312 %**

Resultados utilizando como FrontEnd el programa AdvanceFrontEnd

<b>SNR</b>	<b>RUIDO1</b>	<b>RUIDO2</b>	<b>RUIDO3</b>	<b>RUIDO4</b>	<b>PROMEDIO</b>
Limpia	98,71	98,37	98,09	98,86	98,5075
20 dB	98,03	96,81	98,18	98,86	97,9700
15 dB	96,96	93,78	97,64	98,58	96,7400
10 dB	94,57	87,58	96,12	97,96	94,0575
5 dB	88,33	74,24	91,08	96,36	87,5025
0 dB	70,86	53,72	76,95	90,99	73,1300
-5 dB	34,85	31,50	42,98	70,84	45,0425

**Rendimiento promedio 20 dB a 0 dB: 89,8800 %**

**Rendimiento promedio 20 dB a -5 dB: 82,4071 %**



## 9.- Conclusiones

Este proyecto de Final de Carrera se nos planteo como una reescritura en C de un programa escrito en Fortran, llamado NETPOLS, que realizaba un filtrado de Wiener iterativo, con la posibilidad de utilizar estadísticas de orden superior (HOS).

Iniciamos la programación de un primer algoritmo basandonos en la implementación realizada en anteriores proyectos [Jove-93] y [Esta-95]. De esta forma conseguimos realizar un primer programa de filtrado de Wiener, totalmente compatible con cualquier sistema operativo, que realizaba el filtrado iterativo propuesto por Lim y Oppenheim [Lim-79]. Además, este programa permitía realizar modelados AR utilizando correlaciones y estadísticas de orden superior, como se propone en la tesis [Sala-95].

Un nuevo problema se nos presentó al iniciar la evaluación de rendimiento de este primer algoritmo, carecíamos de un programa funcional que realizara medidas objetivas. Por ello realizamos un nuevo programa en C, que bautizamos como DISTCALC, basandonos en el tipo de evaluación que propone el documento [Hans-98]. Este programa realiza las medidas objetivas SNR global y segmentada, Itakura, Itakura-Saito, Cepstrum y Log Area-Ratio.

Al evaluar el aumento de rendimiento que introducían las técnicas de filtrado iterativo y estadísticas de orden superior, ver apartado 7.2.5., observamos que introducían una gran distorsión en la señal de voz, además de otros efectos desagradables como el ruido musical, y que apenas mejoraban el filtrado básico de una iteración utilizando correlaciones.

Con el ánimo de mejorar los mediocres resultados que obtenía esta primera versión del programa RERCOM, introducimos mejoras que no se habían implementado en anteriores proyectos como:

- Posibilidad de modificar de forma arbitraria la longitud y desplazamiento de trama.
- Una etapa de prefiltrado basado en sustracción espectral.
- Introducción de un VAD híbrido basado en energía y distancia espectral, que permite la reestimación del ruido de una señal de voz en tramas de silencio.
- Mejora del filtrado de Wiener con la introducción de un filtrado en peine adaptativo.

- Introducción de un postfiltrado basado en mediana para eliminar ruido musical.
- Introducción de un sistema de detección de tramas sonoras/sordas, que permite modificar el número de coeficientes AR a utilizar.
- Introducción de un sistema que alarga artificialmente la longitud de trama de las tramas sonoras o periódicas. Este alargamiento de las tramas permite mejorar las estimaciones de cumulantes de orden superior.

Estas principales mejoras, junto a otras de menor importancia, y la posibilidad de activarlas y desactivarlas por medio de argumentos nos permitieron conseguir un programa de simulación que obtenía muy buenos resultados frente a ruido blanco. Este programa lo bautizamos como simulador RERCOM.

Las pruebas realizadas con este programa nos permitieron llegar a las siguientes conclusiones, ver apartado 7.2:

- La longitud de trama óptima es 256 muestras para 8Khz, lo que supone un tiempo de trama de 32 ms.
- Un desplazamiento de trama razonable de 128 o 64 muestras, solapamiento del 50% y 75%.
- La activación de unos filtros que atenúen las bandas de frecuencia donde, a priori, hay poca energía de voz, de 0-50Hz y de 3,5-4 KHz. para una frecuencia de muestreo de 8 KHz., mejoran sensiblemente todas las medidas.
- La utilización de una  $\beta=1,2 \delta=1$ , en la expresión del filtro de Wiener genérico, obtiene una ligera mejora de medidas sin apenas distorsionar la señal de voz.
- La utilización de un filtro peine adaptativo mejora la calidad de audición de las tramas de voz frente a ruidos de banda ancha como son el ruido blanco o el ruido rosa.
- La utilización de un prefiltro basado en sustracción espectral, utilizando los parámetros  $\beta=1,2 \delta=2$ , mejoran sensiblemente los resultados de medidas objetivas y calidad de audición a costa de introducir un casi imperceptible ruido musical.
- La activación de un VAD permite mejorar enormemente la calidad de audición de las tramas de silencio, así como la actualización y adaptación del sistema frente a ruidos no estacionarios.
- La modificación del orden de predicción en la estimación de modelado AR del filtro de Wiener en función de si la trama es sonora o sorda, utilizando ordenes 8 y 3,

permite obtener un rendimiento similar a utilizar un orden 10 y 10, con un menor tiempo de procesado.

- La activación del postfiltro reduce el ruido residual y musical generado en etapas previas, pero sólo en condiciones de mucho ruido, ya que es un filtro muy agresivo.

Después de comprobar el excelente rendimiento del programa frente a condiciones muy adversas de SNR, se nos facilitó el código fuente y la documentación del programa Advance Front-End, para que pudiéramos comparar sus resultados con los del programa simulador RERCOM, por lo que decidimos crear una pequeña base de datos con diversos ruidos, además del ruido blanco, para que la comparación fuera más realista.

Los resultados de esta comparación fueron muy contundentes, en todas las pruebas el programa RERCOM, con los parámetros optimizados para ruido blanco, obtenía los mejores resultados de SNR global y segmentada. Y en la mayoría de casos, sobretodo a SNR bajas, el programa RERCOM obtenía las menores distancias espectrales, ver apartado 7.3.

Para confirmar estos excelentes resultados, realizamos pruebas de reconocimiento de voz utilizando la base de datos TIdigits que utiliza el sistema de reconocimiento de voz AURORA versión 008 (estándar para el sistema GSM). Los resultados obtenidos nos confirmaron que el programa RERCOM mejoraba los resultados obtenidos por Advance Front-End en tres de los cuatro ruidos de la base de datos, sobretodo en condiciones de SNR bajas.

A partir de las pruebas realizadas, comprobamos que el sistema RERCOM ofrece prestaciones claramente superiores al estandar de la ETSI Advance-FrontEnd [Etsi-03] para la mayoría de tipos de ruido. Pero también comprobamos que el sistema RERCOM estaba poco optimizado, ya que su tiempo de procesado era unas 10 veces mayor que el sistema ADVANCE. Para disminuir el tiempo de procesado del programa RERCOM realizamos una nueva versión optimizada del programa, que bautizamos como RERCOM\_DSP. Esta nueva versión era una primera aproximación para la implementación en DSP del sistema RERCOM. Este programa obtiene una prestaciones ligeramente inferiores, pero con un tiempo de procesado 10 veces menor a su antecesor.

A partir de este proyecto quedan abiertas nuevas líneas de futuro como son una mejor implementación de un VAD que no dependa tan estrechamente de la estimación inicial de ruido, ya que el sistema parte de la suposición de que las primeras tramas, unos 100 ms, son ruido. Y la implementación física del algoritmo en DSP. Líneas que actualmente ya están siendo estudiadas por otros proyectos.



## 10.- Bibliografía

- [An-88] C.K.An, S.B.Kim, E.J.Powers. “Optimized Parametric Bispectrum Estimation”. Proc. ICASSP, pp. 2392-2395. Nueva York, USA. 11-14 de Abril, 1988.
- [Andr-86] H.C.Andrews, B.R.Hunt. “Digital Image Restoration”. Prentice Hall. 1986.
- [Boll-79] S.F.Boll. “Suppression of Acoustic Noise in Speech Using Spectral Subtraction”. IEEE Trans. on ASSP, Vol. ASSP-27, N° 2, pp. 113-120. Abril 1979.
- [Edua-00] Eduardo Lleida Solano. “Cancelación de Ruido Aditivo: Sustracción Espectral” [En línea]. URL <[http://www.gtc.cps.unizar.es/~eduardo/investigacion/voz/susespec/sustrace\\_spectral.html](http://www.gtc.cps.unizar.es/~eduardo/investigacion/voz/susespec/sustrace_spectral.html)>. Communication Technology Group, Universidad de Zaragoza. Octubre 2000.
- [Edua-04] Eduardo Lleida Solano, Salvador Olmos Gassó. “Tratamiento digital de la señal” [En línea]. URL <[http://www.gtc.cps.unizar.es/~eduardo/docencia/tds/Temario.html#estim\\_n\\_oparam](http://www.gtc.cps.unizar.es/~eduardo/docencia/tds/Temario.html#estim_n_oparam)>. Communication Technology Group, Universidad de Zaragoza. 2004.
- [Esta-95] J.J.Estarellas. “Técnicas de Speech Enhancement para Entornos Altamente Ruidosos”. Proyecto Final de Carrera, ETSETB, UPC. Barcelona. 1995.
- [Etsi-03] ETSI Standard. ”ETSI ES 202 212 V1.1.1” [En línea]. URL <<http://www.etsi.org>>. European Telecommunications Standards Institute, Francia. Noviembre 2003.
- [Fono-91] J.A.R.Fonollosa, J.Vidal, E.Masgrau. “Adaptative System Identification Based on Higher-Order Statistics “. Proc. ICASSP, pp. 3437-3440. Toronto, Ontario, Canada. 14-17 de Mayo. 1991.
- [Furu-89] S.Furui. “Digital Speech Processing, Synthesis and Recognition”. Marcel Dekker, Cop. Nueva York. 1989.
- [Gian-89] G.B.Giannakis, J.M.Mendel. “Identification of Nonminimum Phase Systems Using HOS”. IEEE Trans. on ASSP, Vol. 37, N° 3, pp. 360-377. Marzo 1989.
- [Gian-90] G.B.Giannakis. “On the Identifiability of Non-Gaussian ARMA Models Using Cumulants”. IEEE Trans. on Automatic Control, Vol. 35, N° 1, pp. 1284-1296. Julio 1990.
- [Hans-91] J.H.L.Hansen, M.A.Clements, “Constrained Iterative Speech Enhancement with Application to Speech Recognition”. IEEE Trans. on ASSP, Vol. ASSP-39, N° 4, pp. 795-805. Abril 1991.

- [Hans-98] J.H.L. Hansen, B. Pellom, "An effective Quality Evaluation Protocol for Speech Enhancement Algorithms" ICSLP-98: Inter. Conf. on Spoken Language Processing, vol. 7, pp. 1819-2822, Sydney, Australia, Dec. 1998.
- [Jin-03] Kyung Jin Byun, Sangbae Jeong, Hoi Rin Kim, and Minsoo Hahn. "Noise Whitening-Based Pitch Detection for Speech Highly Corrupted by Colored Noise". ETRI Journal, Volumen 25, Numero 1, pp. 49-51. Febrero 2003
- [Jove-93] F.X.Jové. "Técnicas Robustas de Procesado de Señal de Voz Usando Estadísticas de Orden Superior". Proyecto Final de Carrera, ETSETB, UPC. Barcelona. 1993.
- [Lim-78] J.S.Lim, A.V.Oppenheim. "All-Pole Modeling of Degraded Speech". IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-26, N° 3, pp. 197-210. Junio 1978.
- [Lim-79] J.S.Lim, A.V.Oppenheim. "Enhancement and Bandwidth Compression of Noisy Speech". Proc. of The IEEE, Vol. 67, N° 12, pp. 1586-1604. Diciembre 1979.
- [Mark-04] Mark S. Csele. "WAV File Format Description" [En línea]. URL <<http://technology.niagarac.on.ca/courses/comp630>>. Niagara Collage, Canada. 2004.
- [Makh-75] J.H.Makhoul. "Linear Prediction: A Tutorial Review". Proc. of The IEEE, Vol. 63, pp. 561-580. Abril 1975.
- [Marpe-87] S.L.Marple Jr. "Digital Spectral Analysis with Applications". Prentice Hall. Englewoods Cliffs. 1987.
- [Masg-92a] E.Masgrau, J.A.R.Fonollosa, A.Arduanuy. "Enhancement of Speech by Using Higher-Order Spectral Modelling". Proc. of EUSIPCO, pp. 307-310. Bruselas, Bélgica. 24-27 de agosto, 1992.
- [Masg-92b] E.Masgrau, J.M.Salavedra, A.Moreno, A.Arduanuy. "Speech Enahncement by Adaptative Wiener Filtering Based on Cumulant AR Modelling". Proc. of ESCA Workshop on Speech Processing in Adverse Conditions, pp 143-146. Cannes, Francia. 10-13 de noviembre, 1992.
- [Mend-91] J.R.Mendel. "Tutorial on Higher-Order-Statistics (Spectra) in Signal Processing and System Theory: Theoretical Results and Some Applications". Proc. of The IEEE, Vol. 79, N° 3. Marzo 1991.
- [Niki-87] C.L.Nikias, M.R.Raghubeer. "Bispectrum Estimation: A Digital Signal Processing Framework". Proc. of The IEEE, Vol. 75, N° 7, pp. 869-891. Julio 1987.
- [Niki-91] C.L.Nikias, A.P.Petropulu. "Higher-Order Spectra Analysis". IEEE, Cop. Piscataway, NJ. 1991.

- [Nume-92] Numerical Recipes. “Numerical Recipes in C: The Art of Scientific Computing”. [En línea]. URL <<http://www.nr.com>>. Cambridge University Press. Programs. 1992.
- [Oppe-97] A.V.Oppenheim, A.S.Willsky, I.T.Young. “Señales y sistemas”. Hispanoamericana, Cop. 1997.
- [O'Sha-89] D.O'Shaughnessy. “Enhancing Speech Degraded by Additive Noise or Interfering Speakers”. IEEE Communications Magazine. Febrero 1989.
- [O'Sha-00] D.O'Shaughnessy. “Speech Communication: Human and Machine”. IEEE Press, Cop. Nueva York. 2000.
- [Pali-91] K.K.Paliwal, M.M.Sondhi, “Recognition of Noisy Speech Using Cumulants-Based Linear Prediction Analysis”. Proc. of ICASSP, Vol. 1, pp. 429-432. Toronto. Mayo 1991.
- [Papo-91] A.Papoulis. “Probability, Random Variables, and Stochastic Processes”. McGraw Hill. Nueva York. 1991.
- [Pear-98] David Pearce. “Experimental Framework for the Performance Evaluation of Distributed Speech Recognition Front-ends”. Motorota. Septiembre 1998.
- [Pflu-92] L.A.Pflug, G.E.Ioup, J.W.Ioup, R.L.Field. “Properties of Higher-Order Correlations and Spectra for Bandlimited, deterministic Transients”. J. Acoust. Soc. Am., Vol. 91, Nº 2, pp. 975-988. Febrero 1992.
- [Proa-96] J.G.Proakis, D.G.Manolakis. “Digital Signal Processing Principles, Algorithms, and Applications”. Prentice Hall Internacional, Inc. New Jersey. 1996.
- [Ragh-85] M.R.Raghubeer, C.L.Nikias. “Bispectrum Estimation: A Parametric Approach”. IEEE Trans. on ASSP, Vol. 33, Nº 4, pp. 1213-1230. Octubre 1985.
- [Ragh-86] M.R.Raghubeer, C.L.Nikias. “Bispectrum Estimation via AR Modeling”. Signal Processing, Vol. 10, Nº 1, pp. 35-48. Enero 1986.
- [Rodr-03] Rodrigo Huerta Cortés. “Transformada Rápida de Fourier” [En línea]. URL <<http://www.elo.utfsm.cl/~elo385/Documentos/Transformada%20R%e1pi%20de%20Fourier.pdf>>. Universidad Técnica Federico Santa María, Dpto. Electrónica, Valparaíso. Abril 2003.
- [Sala-93a] J.M.Salavedra, E.Masgrau, A.Moreno, X.Jové. “Comparation of Different Order Cumulants in a Speech Enhancement System by Adaptative Wiener Filtering”. Signal Processing Workshop on Higher-Order Statistics, IEEE. South Lake Tahoe, USA. 7-9 de Junio, 1993.

- [Sala-93b] J.M.Salavedra, X.Jové, E.Masgrau, A.Moreno. “Sistema de Mejora de Voz Usando Estimación AR de Orden Superior en Ambientes Reales”. URSI. Valencia. 21-24 de Septiembre, 1993.
- [Sala-94] J.M.Salavedra, E.Masgrau, A.Moreno, J.Estarellas. “Some Robust Speech Enhancement Techniques Using Higher-Order AR Estimation”. Proc. EUSIPCO, pp. 1194-1197. Edimburgo, Escocia. 13-16 de septiembre, 1994.
- [Sala-95] J.M.Salavedra. “Técnicas de Speech Enhancement Considerando Estadísticas de Orden Superior”. Tesis Doctoral, UPC, Departamento de Teoría de Señal y Comunicaciones. Barcelona. Junio 1995.
- [Sala-02] J.M.Salavedra. “La Detección de Actividad Oral en Ambientes Ruidosos”. Actas del II Congreso de la Sociedad Española de Acustica Forense, pp 197-204. Abril 2003.
- [Sovk-96] Pavel Sovka, Petr Pollak, Jan Kybic. “Extended Spectral Subtraction”. Czech Technical University, Faculty of Electrical Engineering, Czech Republic. 1996.
- [Swam-89] A.Swami, J.M.Mendel. “AR Identifiability Using Cumulants”. Proc. Workshop on HO Spectral Analysis, pp. 13-18. Vail, CO, USA. Junio 1989.
- [Torr-02] Angel de la Torre Vega.”Procesamiento de señales de voz” [En línea]. URL <<http://ceres.ugr.es/~atv/Documents/Docencia/voz.ppt>>. Dpto. Electrónica y Tecn. Computadores – UGR. 2002.
- [Vida-93] J.Vidal, E.Masgrau, A.Moreno, J.A.R.Fonollosa. “Speech Analysis Using Higher Order Statistics”. Eds. Martin Cooke, Steve Beet, Malcolm Crawford: “Visual Representations of Speech Signals”. Ed. Wiley Professional Computing, Chapter 38, pp. 347-354. 1993.