

Universidad Central “Marta Abreu” de Las Villas

Facultad de Ingeniería Eléctrica

Centro de Estudios de Electrónica y Tecnologías de la

Información CEETI



TRABAJO DE DIPLOMA

**Análisis comparativo de técnicas de procesamiento
de señales en prótesis auditivas.**

Autor: Julio Alejandro Martínez Piñeiro

Tutor: Dr.C.T. Julián Cárdenas Barrera.

Santa Clara

2011

"Año 53 de la Revolución"

Universidad Central “Marta Abreu” de Las Villas

Facultad de Ingeniería Eléctrica

**Centro de Estudios de Electrónica y Tecnologías de la
Información CEETI**



TRABAJO DE DIPLOMA

Análisis comparativo de técnicas de procesamiento de señales en prótesis auditivas.

Autor: Julio Alejandro Martínez Piñeiro

E-mail: jmpineiro@uclv.edu.cu

Tutor: Dr.C.T. Julián Cárdenas Barrera.

E-mail: julian@uclv.edu.cu

Cotutor: Roberto Jiménez Yera

E-mail: robertojy@cpe.vcl.sld.cu

Consultante: Armando Alvaré Jaramillo.

E-mail: armandoaj@uclv.edu.cu

Santa Clara_2010

"Año 53 de la Revolución"



Hago constar que el presente trabajo de diploma fue realizado en la Universidad Central “Marta Abreu” de Las Villas como parte de la culminación de estudios de la especialidad de Ingeniería en Automática, autorizando a que el mismo sea utilizado por la Institución, para los fines que estime conveniente, tanto de forma parcial como total y que además no podrá ser presentado en eventos, ni publicados sin autorización de la Universidad.

Firma del Autor

Los abajo firmantes certificamos que el presente trabajo ha sido realizado según acuerdo de la dirección de nuestro centro y el mismo cumple con los requisitos que debe tener un trabajo de esta envergadura referido a la temática señalada.

Firma del Tutor

Firma del Jefe de Departamento
donde se defiende el trabajo

Firma del Responsable de
Información Científico-Técnica

PENSAMIENTO

"En cuestiones de cultura y de saber, sólo se pierde lo que se guarda; sólo se gana lo que se da."

Antonio Machado.

DEDICATORIA

A mis padres...

AGRADECIMIENTOS

A mis padres, por las innumerables muestras de sacrificio, apoyo y amor que me brindaron.

A mi tutor Julián por su confianza en mí y por su sabia conducción.

A mi tutor Roberto por su dedicación y esfuerzo personal en el desarrollo de este proyecto.

A Armando por su confianza en mí.

A todos mis amigos, amigas, y compañeros de aula.

A toda mi familia por estar siempre pendientes de mí.

A mis vecinos por la ayuda ofrecida en los momentos más difíciles.

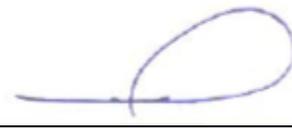
A todos mis profesores y maestros.

TAREA TÉCNICA

1. Estudio de las técnicas de procesamiento actuales empleadas en prótesis auditivas y métodos que permitan su comparación.
 - a. Análisis de la literatura científica sobre el tema.
 - b. Análisis de algoritmos desarrollados en la bibliografía.
2. Estudio de los algoritmos empleados en prótesis desarrolladas en el CEETI.
3. Diseño de experimento comparativo entre los algoritmos seleccionados.
 - a. La comparación incluye: Calidad de la reducción de ruido ante diferentes tipos de ruidos, Empleo de criterios objetivos que permitan evaluar la calidad de las señales sintetizadas por las prótesis desarrolladas en las tesis mencionadas. Experimento subjetivo utilizando criterios MOS.
 - b. Extraer conclusiones sobre qué métodos son los más apropiados a incluir en una prótesis a desarrollar.
4. Escritura del informe final.



Firma del Autor



Firma del Tutor

RESUMEN

Las hipoacusias o pérdidas parciales de la capacidad auditiva constituyen problemas de salud que afectan a cientos de miles de personas en el mundo. El tratamiento de estas patologías depende de la correcta selección y funcionamiento de las prótesis auditivas. La característica no uniforme de resolución de frecuencia del oído humano, sugieren implementaciones eficientes que traten de simular este comportamiento, y faciliten la aplicación de técnicas no lineales de amplificación en correspondencia con la codificación de intensidad que realiza el sistema auditivo, mientras se evita sobrepasar el umbral de audición del paciente. Por sus características no estacionarias el ruido de la vida real constituye una barrera para el desempeño de las prótesis, los diseños tienen que ser capaces de reducir este fenómeno y garantizar al mismo tiempo, calidad e inteligibilidad en la señal de habla. En correspondencia con lo anteriormente dicho, este trabajo evalúa el desempeño de dos algoritmos de reducción de ruido: el de Sustracción Espectral Geométrico y el basado en Wavelet Packets; y dos algoritmos de compensación de pérdidas auditivas: AGC utilizando un Banco de Filtro Gammatone y AGC utilizando la Transformada Wavelet Packets. Para las comparaciones de las técnicas elegidas se emplearon criterios objetivos y subjetivos, ampliamente usados y sugeridos en la literatura científica consultada. Como resultado final se concluye que el algoritmo de Sustracción Espectral Geométrico es una alternativa adecuada para su implementación en prótesis auditivas reales; mientras que el AGC basado en Wavelet Packets realiza una buena amplificación de las pérdidas auditivas en todas las bandas de frecuencias.

TABLA DE CONTENIDOS

PENSAMIENTO.....	i
DEDICATORIA.....	ii
AGRADECIMIENTOS.....	iii
TAREA TÉCNICA	iv
RESUMEN	v
INTRODUCCIÓN	1
CAPÍTULO 1. REVISIÓN BIBLIOGRÁFICA.....	4
1.1 Tipos de sorderas.	4
1.2 Prótesis auditivas.	6
1.2.1 Compensación de pérdidas auditivas.	6
1.2.2 Elementos de las prótesis auditivas digitales.....	7
1.2.2.1 Bancos de Filtros.	7
1.2.2.2 Reducción de ruido.....	9
1.2.2.3 AGC.	11
1.2.3 Tendencias actuales.....	11
1.3 Técnicas para la comparación de prótesis auditivas digitales.	12
1.4 Conclusiones parciales del capítulo.....	14
CAPÍTULO 2. MATERIALES Y MÉTODOS.....	15

2.1	Algoritmos objeto de análisis.....	15
2.1.1	Método 1.....	16
2.1.1.1	Algoritmo Geométrico de Reducción de Ruido (AG).	17
2.1.1.2	Banco de Filtros Gammatone.	20
2.1.1.3	AGC.	22
2.1.2	Método 2.....	24
2.1.2.1	Algoritmo de reducción de ruido basado en Wavelet Packets.	24
2.1.2.2	Algoritmo de compensación de pérdidas auditivas mediante Wavelet Packets.....	28
2.2	Características de las Bases de Datos empleadas.	29
2.2.1	AHUMADA.	29
2.2.2	AURORA y sitio Web HearingLossSamples.....	30
2.3	Evaluación objetiva a prótesis digitales auditivas.	30
2.3.1	Evaluación empleando mediciones objetivas.....	31
2.3.1.1	Evaluación perceptual de la calidad de voz (PESQ).	31
2.3.1.2	Razón de verosimilitud logarítmica (LLR).....	32
2.3.1.3	Distancia del espectro de pendiente ponderada WSS.	32
2.3.1.4	Relación señal a ruido por segmentos (RSRseg).....	32
2.3.1.5	Medidas objetivas compuestas.	33
2.3.2	Diseño de los experimentos.....	33
2.3.2.1	Evaluación de los algoritmos de reducción de ruido.	33
2.3.2.2	Evaluación de los algoritmos de compensación de pérdidas auditivas.....	34
2.4	Evaluación subjetiva a prótesis digitales auditivas.	35
2.4.1	Experimento subjetivo de reducción de ruido.....	35
2.4.1.1	Metodología empleada en las pruebas.	36

2.4.1.2	Preparación de las secuencias de pruebas.....	37
2.4.1.3	Presentación de audio.....	37
2.4.2	Experimento subjetivo de compensación de pérdidas auditivas.	38
2.5	Conclusiones parciales del capítulo.	39
CAPÍTULO 3. RESULTADOS Y DISCUSIÓN		40
3.1	Comparación objetiva de las prótesis auditivas digitales.	40
3.1.1	Desempeño de los algoritmos de reducción de ruido.	40
3.1.2	Desempeño de los algoritmos de compensación de pérdidas auditivas.	49
3.2	Comparación subjetiva de las prótesis auditivas digitales.	53
3.2.1	Desempeño de los algoritmos de reducción de ruido.	53
3.2.2	Desempeño de los algoritmos de compensación de pérdidas auditivas.	57
3.3	Análisis comparativo de las medidas de evaluación utilizadas.....	60
3.4	Conclusiones parciales del capítulo.	62
CONCLUSIONES Y RECOMENDACIONES		64
	Conclusiones	64
	Recomendaciones.....	65
REFERENCIAS BIBLIOGRÁFICAS.....		66
ANEXOS		70
Anexo I	Tablas de los cálculos de los parámetros objetivos de reducción de ruido. ...	70
Anexo II	Espectrogramas y gráficos de potencia de compensación de pérdidas auditivas.....	77

INTRODUCCIÓN

Para el ser humano, el sistema auditivo es uno de los sistemas sensoriales más importantes debido a la gran cantidad de información ambiental que procesa. Además, es uno de los sistemas involucrados en la adquisición, desarrollo y mantenimiento del lenguaje hablado, por lo que su funcionamiento tiene una gran influencia en su desempeño profesional, intelectual y social.

La interacción de este sistema con el mundo exterior permite reconocer palabras, notas musicales, emociones, etc. Una buena escucha asegura cómodamente la integración de la persona a la sociedad, no siendo así para individuos que presentan problemas auditivos, los cuales tienen que imponerse ante disímiles barreras para lograr comunicarse.

La mayoría de las personas con alguna deficiencia en el sistema auditivo, experimentan ciertos sentimientos de aislamiento con respecto a la sociedad en que viven. El desarrollo personal de las mismas puede ser limitado a una minoría, ya sea por causa del rechazo por parte de las personas que no presentan este problema o por complejos desencadenados propiamente por los pacientes con problemas auditivos.

Las pérdidas auditivas más comunes en los pacientes son las hipoacusias, las cuales constituyen afecciones del sistema auditivo de manera parcial, y afectan tanto a niños como a adultos, sin importar el sexo o nivel social. Las principales causas de estas pérdidas auditivas son la vejez, los antecedentes familiares de hipoacusia, el ambiente urbano ruidoso o la utilización subtóxica de medicamentos y sustancias lesivas para el oído, enfermedades vasculares que someten a hipoxia al aparato auditivo, enfermedades metabólicas como la diabetes mellitus, hiperlipoproteinemia u obesidad y el alzheimer, enfermedad que muestra degeneración en la cóclea, con un patrón histológico distinto al que se produce en el sistema visual y olfatorio[1].

La primera referencia escrita sobre la utilización de un aparato para la corrección auditiva fue escrita por el ensayista y filósofo inglés Francis Bacon en el año 1664, titulado *Sylva Sylvarum* y fue editado en Frankfurt am Main por la editorial Schonwetter. En el se describe: "... existe en España un aparato para usar en el oído, capaz de ayudar a aquellos que oyen poco." En la actualidad ha existido un desarrollo vertiginoso de los dispositivos encaminados a tratar las deficiencias del oído y las prótesis auditivas digitales han ocupado un espacio importante en el diseño de equipos electrónicos con resultados muy alentadores para el tratamiento de las hipoacusias en todas sus modalidades.

De manera general, las prótesis auditivas actuales emplean un conjunto de bancos de filtros con diferentes ganancias que se ajustan en correspondencia con el audiograma específico de cada paciente. La selección de la ganancia depende, además, de la intensidad del sonido a amplificar, pues su percepción en el discapacitado posee comportamientos no lineales [2]. La posibilidad de implementaciones que compensen eficazmente los déficits en la audición existe [3], y de hecho, muchos de los diseños de bancos de filtros; técnicas de amplificación selectivas en frecuencia y adaptación no lineal a estrechos márgenes dinámicos se diseñan a partir de modelos del funcionamiento del oído humano.

Los Bancos de Filtros, utilizados en múltiples aplicaciones dentro del procesamiento digital de señales y particularmente en el procesamiento de voz, requieren de implementaciones rápidas y de bajo consumo de potencia en aplicaciones de prótesis auditivas [4]. El control automático de ganancia (AGC – de sus siglas en inglés –) debe cumplir con las características de amplificación de intensidad prescritas y tratar de no sobrepasar estos niveles [5].

El umbral de escucha, determinado mediante técnicas de diagnósticos como la audiometría en una persona con hipoacusia, se ve aumentado en cuanto a intensidad para diferentes bandas de frecuencia. Generalmente, estas bandas son contiguas, lo que deriva dificultades que el AGC debe solucionar, pues la amplificación de bandas cercanas podría provocar enmascaramientos en bandas vecinas. Los desarrollos recientes en la tecnología de micrófonos direccionales y en la comunicación inalámbrica, han mostrado mejoras en el reconocimiento del habla cuando el usuario se expone a entornos ruidosos [6]. Además, los

algoritmos de reducción de ruido han avanzado considerablemente con el objetivo de lograr comodidad y alivio entre los portadores de prótesis auditivas [7]

En nuestro país la adquisición de prótesis auditivas resulta cada día más compleja debido a las restricciones de la política internacional y los altos precios de estos dispositivos. Por estas razones adquieren enorme importancia la contribución a la mejora del desempeño de las prótesis auditivas mediante métodos comparativos.

Basado en lo antes expuesto el objetivo general de este trabajo es:

- Determinar, a partir del análisis comparativo de técnicas de procesamiento de voz, los elementos más adecuados a incluir en una prótesis auditiva real.

A partir de este objetivo general se derivan los siguientes objetivos específicos:

1. Estudiar las técnicas de procesamiento empleadas en las prótesis auditivas que serán objeto de estudio.
2. Analizar las métricas empleadas para determinar la efectividad de los resultados obtenidos en las propuestas desarrolladas por los autores de las tesis seleccionadas.
3. Diseñar el experimento comparativo entre los algoritmos desarrollados.
4. Extraer conclusiones sobre qué métodos son los más apropiados a incluir en una prótesis a desarrollar.

Organización del informe.

Este trabajo se estructura a partir de un informe que cuenta con: resumen; introducción, la que concluye con el planteamiento de un objetivo general y cuatro específicos, seguido de un capitulo compuesto por una fundamentación teórica, con epígrafes que refieren los antecedentes teóricos de la investigación; un segundo capítulo que se refiere a la metodología y materiales empleados, en el cual se incluyen los diseños de los experimentos; y un tercer capítulo donde se presentan y discuten los resultados alcanzados, terminando con su validación. Se concluye con los resultados obtenidos y se recomiendan aspectos para la continuación de los estudios relacionados con el tema tratado.

CAPÍTULO 1. REVISIÓN BIBLIOGRÁFICA

En este capítulo se presenta el marco teórico alrededor del tema de las prótesis auditivas digitales. Primeramente, se tratan los tipos de sorderas y las características de estas que dan paso a la utilización de prótesis auditivas (epígrafe 1.1). En el epígrafe 1.2, se describen los factores que influyen en la compensación de las pérdidas auditivas, los principales elementos de las prótesis digitales auditivas y las tendencias actuales en el diseño de las mismas. En el epígrafe 1.3 se abordan las técnicas de medición más utilizadas en la validación de las prótesis auditivas, y por último, en el epígrafe 1.4, se comentan las conclusiones parciales del capítulo.

1.1 Tipos de sorderas.

Los problemas de audición afectan a un número importante de personas a lo largo de todo el planeta y aproximadamente una de cada siete personas presentan una pérdida auditiva en mayor o menor grado. Se le denomina sordera o hipoacusia al déficit funcional que ocurre cuando un sujeto pierde capacidad auditiva.

Para estudiar una sordera es necesario caracterizarla según la cantidad de pérdida de audición (clasificación cuantitativa), con respecto al lugar donde se asienta la lesión que produce el déficit (clasificación topográfica), de acuerdo con la etiología de la hipoacusia (clasificación etiológica) y por último en relación con el lenguaje (clasificación locutiva)[8].

Clasificación cuantitativa:

1. Hipoacusia leve: Pérdida no superior a 40 dB para las frecuencias centrales.
2. Hipoacusia moderada: Pérdida comprendida entre 40 y 70 dB.

3. Hipoacusia grave: Pérdida comprendida entre 70 y 90 dB.

4. Hipoacusia profunda: Pérdida superior a 90 dB.

Clasificación topográfica:

1. Hipoacusias de transmisión: Aparecen por alteraciones del oído externo y medio, y se llaman también hipoacusias de conducción. En general, estas hipoacusias son recuperables médica o quirúrgicamente.

2. Hipoacusias de percepción o sensorineurales: Ocurren, bien por lesión del órgano de CORTI (hipoacusias cocleares), o bien por lesiones de las vías acústicas (hipoacusias retrocleares o neuropatías).

- Hipoacusias cocleares: No son curables por procedimientos médicos o quirúrgicos.
- Hipoacusias retrocleares o neuropatías: Las más frecuentes son debidas a enfermedades del sistema nervioso, como el neurinoma del nervio acústico: proceso tumoral benigno pero que necesita de tratamiento quirúrgico.

3. Hipoacusias mixtas: Ocasionalmente por varias causas y que afectan al mismo tiempo al oído externo-medio y al interno.

4. Hipoacusias centrales: Ocasionalmente por trastornos a nivel cerebral y que causan una falta de decodificación del mensaje sonoro. Se salen del ámbito de la Otorrinolaringología para entrar de lleno en la neurología o neuropsiquiatría.

Clasificación etiológica:

Esta clasificación está basada en el origen de la sordera y en el momento en que debuta:

1. Hipoacusias hereditarias o genéticas: Son sorderas transmitidas por una alteración genética y pueden ser: a) precoces (desde el momento del nacimiento), o b) tardías (cuando se desarrollan a lo largo de la vida).

2. Sorderas adquiridas: Originadas por enfermedades sobrevenidas en algún momento de la vida y pueden ser: a) prenatales (cuando un agente patógeno ha actuado en período embrionario o fetal), b) perinatales (cuando la enfermedad ha ocurrido en el momento del parto) o c) posnatales (cuando el daño se establece a lo largo de la vida).

Clasificación locutiva:

1. Sordera Prelocutiva: Si la sordera ocurre antes de la adquisición del lenguaje.
2. Sordera Postlocutiva: Si la sordera ocurre después de la adquisición del lenguaje.

Atendiendo a la clasificación topológica, el diagnóstico de prótesis auditivas resulta un medio de solución en caso de que las hipoacusias sean cocleares (sensorineurales), y en algunos casos de transmisión, pues estas últimas pueden ser resueltas quirúrgicamente [9].

1.2 Prótesis auditivas.**1.2.1 Compensación de pérdidas auditivas.**

Las pérdidas auditivas sensorineurales de tipo coclear se encuentran entre las más comunes y, a pesar de que no tienen cura, pueden compensarse a través de prótesis auditivas. Desafortunadamente, a diferencia de otros tipos de hipoacusias, éstas no pueden ser tratadas por simple amplificación debido a múltiples razones:

1. No todas las regiones del espectro audible sufren igual atenuación en el oído interno; por lo general, las componentes de altas frecuencias sufren mayor atenuación que las de baja frecuencia. Bajo estas premisas la compensación del déficit auditivo debe realizarse a través de amplificación selectiva en frecuencia, ofreciendo diferentes ganancias en diferentes regiones de su espectro. Generalmente las prótesis implementan descomposición en sub bandas de frecuencias empleando bancos de filtros cuyas ganancias se modifican a partir de conocer las pérdidas del individuo que la usará.
2. El umbral de audición se incrementa mucho más que el umbral de confortabilidad. Para compensar este efecto, se debe amplificar más los sonidos suaves que los intensos, generalmente siguiendo una función no lineal.
3. Pobre resolución espectral. Este fenómeno se manifiesta por la incapacidad del individuo para reconocer sonidos de frecuencias cercanas. A medida que este problema se hace más agudo los sonidos de baja frecuencia con relativamente alta intensidad, enmascaran totalmente a los sonidos de alta frecuencia. La compensación de este fenómeno necesita garantizar que ningún sonido sobrepase niveles que puedan provocar enmascaramiento. El ruido ambiente ejerce una influencia extremadamente negativa,

provocando en muchos casos enmascaramiento total del sonido útil. Las prótesis minimizan estas situaciones empleando diversas técnicas, entre las que se encuentran los algoritmos de reducción de ruido.

4. Pobre resolución temporal: los sonidos débiles de corta duración, pueden ser enmascarados por sonidos de mayor amplitud que le preceden o suceden. Este es un comportamiento que provoca disminución en la inteligibilidad del lenguaje. Su compensación es difícil, especialmente en ambientes ruidosos, dado que pueden amplificarse sonidos débiles que no se corresponden con información del habla, limitando la comprensión del mensaje.

Las prótesis auditivas constituyen hoy en día la única solución a las hipoacusias sensorineurales. Su función reside fundamentalmente en intentar remediar los problemas anteriormente mencionados, o sea: compensar elevados umbrales de audición dependientes de la frecuencia, así como un margen dinámico auditivo con un comportamiento generalmente no lineal, minimizar los efectos que provocan el enmascaramiento espectral, la pobre resolución espectral, la pobre resolución temporal, la pobre integración temporal y la percepción distorsionada de envolvente [2] [5].

1.2.2 Elementos de las prótesis auditivas digitales.

1.2.2.1 Bancos de Filtros.

La descomposición de señales en bandas de frecuencia, ha demostrado eficacia para lograr altos niveles de reconocimiento del habla e inteligibilidad, además, permite la inserción de ganancias prescritas para compensar los umbrales de escucha. Incluso, la aplicación independiente de técnicas de procesamiento de voz a bandas de frecuencias específicas ha conducido a importantes resultados sobre todo en lo referente a la reducción de ruido.

Los bancos de filtros que se utilizan en este tipo de aplicación pueden ser clasificados en dos categorías: banco de filtro Uniforme y banco de filtro No Uniforme. La desventaja principal de los bancos de filtros Uniformes es la alta complejidad computacional que se necesita para cubrir los requisitos de resolución de frecuencia no-uniforme del sistema auditivo humano. Por otra parte, dentro de los bancos de filtros No Uniforme, pueden ser clasificados como los más apropiados los Bancos de Filtro de Bandas-Octava, de Bandas Críticas y de 1/3 Banda-Octava [4]. En el procesamiento de voz y específicamente en el

diseño de prótesis digitales auditivas, los bancos de filtros representan la etapa que más energía consume.

Un banco de filtro de 7 bandas-octava diseñado en [10], usa técnicas de interpolación en filtros FIR (IFIR). Por otro lado en [11] se propone un banco de filtro de 8 bandas-octava con IFIR y técnica de enmascaramiento de respuesta de frecuencia (FRM -de sus siglas en inglés-) que reduce la complejidad computacional considerablemente. En general, el banco de filtro 8 bandas-octava puede hacer buena compensación para la típica pérdida auditiva postlocutiva, cuya característica es plana. Pero muchas son las personas actualmente con pérdidas auditivas debido a la exposición a altos niveles de ruido, las cuales presentan generalmente déficits auditivo en varios rangos estrechos de frecuencia. Por esta razón es necesaria una mayor cantidad de bandas de frecuencias para tratar como corresponde el déficit auditivo de estas personas.

Un banco de filtro de 16 bandas críticas se diseña en [12]. Las bandas críticas toman en cuenta el funcionamiento del sistema auditivo para proveer al diseño de una compensación de pérdida auditiva exacta.

Aunque los bancos de filtros de bandas-críticas son buenos para prótesis digitales, el banco de filtro deseado para esta aplicación debe ser diseñado de forma que, cada ganancia de inserción obtenida para la prótesis, pueda ser aplicada acorde con las fórmulas de prescripción. La fórmula de prescripción conocida para prótesis digitales, NAL-NL1 [13] prescribe, por ejemplo, la inserción de ganancia para frecuencias 1/3-octava definida por el estándar ANSI S1.11 [14].

Existen muchos bancos de filtros ANSI S1.11 implementados con filtro de respuesta infinita al impulso (IIR) [15]. El banco de filtro IIR puede ser un buen diseño con baja complejidad computacional, sin embargo, los filtros (FIR) son preferidos [12] no sólo por su respuesta de fase lineal sino también por su estabilidad y por su estructura. El error de redondeo de los filtro FIR es fácil de analizar y controlar y permite el diseño de un hardware más simple. Aunque el oído humano normal no es consciente del atraso de fase, el diseñar bancos de filtros de compensación auditiva con fase lineal exacta tiene algunas ventajas, pues la propiedad de fase-lineal puede hacer que el sonido procesado sea más similar al sonido que normalmente una persona sana está acostumbrado a escuchar. Esto

mejora considerablemente la inteligibilidad del habla, además, mantener la fase, ayuda a la localización del sonido [16] para el caso de usuarios con prótesis bilaterales, además de ser importante para los algoritmos de cancelación de eco o de realimentación [12].

Otra técnica propuesta en la bibliografía es el método de Wavelet-Packets (WP), que es una generalización del proceso de descomposición de la Transformada Wavelet y ofrece un mayor rango de probabilidades para el análisis de la señal. En el análisis con wavelet la señal se distribuye en coeficientes de aproximación y detalle. Los coeficientes de aproximación son transformados para un segundo nivel de descomposición, en nuevos coeficientes de aproximación y detalle. La compresión con base wavelet es un tipo de compresión paramétrica donde los parámetros son los CW, obtenidos mediante el cálculo de TWD o mediante la utilización de WP. El proceso para la compensación de pérdidas auditivas es variante en el tiempo y dependiente de la frecuencia, se fundamenta en la aplicación de una razón de compresión a cada CW en cada banda de frecuencia dependiendo del nivel de intensidad propio del coeficiente [17].

El banco de filtro Gammatone [18] se basa en el concepto de describir el funcionamiento de la cóclea, teniendo como premisa que la codificación en frecuencia del oído humano no es uniforme y se describe usualmente en base a las llamadas bandas críticas. El filtro Gammatone lleva a cabo un análisis espectral y convierte la onda acústica en una representación multi- canal del movimiento de la membrana de basilar. El filtro fue introducido por Johannesma en 1972 y el nombre se refiere al hecho de que la respuesta de impulso pertenece la familia de funciones gamma de estadísticas.

1.2.2.2 Reducción de ruido.

El habla es un mecanismo de comunicación altamente redundante; los seres humanos normo-oyentes, en presencia de ruido, y aun cuando parte del mensaje se enmascara por este, son capaces de entender el habla. Las personas con pérdidas auditivas, sin embargo, pierden gran parte de información, pues algunos sonidos no los escuchan y otros sufren distorsiones más o menos severas [19]. El ruido constituye un grave problema para las personas con déficits auditivos, incluso a niveles perfectamente inocuos bajo condiciones de audición normal. Las prótesis auditivas pueden incluso acrecentar los efectos nocivos

del ruido. Por esta razón, el empleo de algoritmos de reducción de ruido avanzados que permitan aumentar la inteligibilidad del habla resulta esencial.

Varios algoritmos de reducción de ruido para prótesis digitales auditivas han sido propuestos durante los últimos años, la mayoría de estas técnicas realizan la limpieza del habla contaminada, luego de efectuar alguna transformación sobre la señal temporal utilizando como por ejemplo: la Transformada Discreta de Fourier (DFT, -del Inglés Discrete Fourier Transform-), Transformada de Karhunen-Loeve, la Transformada Discreta en Cosenos (DCT,- del Inglés Discrete Cosine Transform-). Las técnicas clásicas se han centrado principalmente sobre:

1. La sustracción espectral [20] donde se le sustrae, al espectro del habla contaminada, el espectro del ruido estimado, logrando así un excelente desempeño del sistema.
2. El filtrado óptimo y adaptativo [21], específicamente el filtrado de Wiener, es una técnica relacionada con las estrategias de filtrado óptimo que intenta minimizar el error medio cuadrático de la señal limpia.
3. El procesamiento basado en modelos estadísticos, donde se intenta modelar el habla limpia sobre la señal de habla contaminada, así como modelar las características del ruido presente en la señal contaminada [22], a través de las características estadísticas de este.

Cualquiera que sea el método escogido para estimar el ruido y la técnica elegida para suprimirlo, el objetivo del procesamiento es mejorar la calidad del habla y preservarla, o incluso incrementar su inteligibilidad. La evaluación del éxito de los distintos métodos de reducción de ruido no es un problema sencillo, debido a que la calidad del habla es un fenómeno complejo de percepción cuya descripción en términos puramente matemáticos no resulta fácil ni eficaz [23] [7].

A la hora de implementar una prótesis digital, es importante tener presente que la etapa de reducción de ruido tenga un buen desempeño, pues de eso dependen las posteriores etapas en el funcionamiento del equipo, como es el caso del bloque del control automático de ganancia (AGC), que de no asegurar la reducción de ruido se podrían amplificar sonidos indeseables y molestos para el usuario.

1.2.2.3 AGC.

Los pacientes con pérdidas auditivas de tipo sensorineural, generalmente experimentan déficits auditivos frente a las altas frecuencias, lo que obliga hacer una amplificación específica sobre tales bandas; al mismo tiempo existen otros tipos de sorderas sensorineural en las que el déficit auditivo está presente en algunas bandas de alta frecuencia y en otras de bajas frecuencias en las que también se requiere cierta especificidad en cuanto al proceso de amplificación [8]. Por lo tanto, para las pérdidas auditivas sensorineural u otro tipo de sordera en las que las extensiones dinámicas están seriamente restringidas, el procesamiento lineal de amplificación tiene limitaciones o es totalmente ineficiente. El manejo automático de la amplificación en las prótesis digitales auditivas resulta complejo, pues es necesario lograr ajustes rápidos en la ganancia de la señal del habla mientras se asegura la amplificación en componentes de frecuencia específicas. Este tema debe ser tratado con mucha delicadeza, pues en caso de incurrir en amplitudes de alta intensidad que superen el umbral del dolor, ya sea por niveles no ajustados de intensidad o por amplificaciones en bandas indebidas, el paciente sería el mayor perjudicado. Investigaciones realizadas muestran cómo es posible implementar un AGC con excelente desempeño frente a ambientes ruidosos [9]

1.2.3 Tendencias actuales.

El uso del PDS en prótesis auditivas ha abierto las puertas a muchas clases de algoritmos y actualmente, figuran entre los avances más significativos la utilización de micrófonos direccionales. Se han desarrollado muchas alternativas para el direccionamiento, incluyendo dirección adaptable y asimétrica automática adaptable, así como descomposición multi-banda automática por parte de los micrófonos [6].

Por otra parte, la utilización de tecnología inalámbrica en prótesis auditivas, logra comodidad para los usuarios, pero la introducción de técnicas inalámbricas como el Bluetooth, métodos modernos de transmisión magnética y tecnología de onda de radio FM, expanden aún más las tendencias actuales de desarrollo, con notables logros en conexiones a dispositivos externos como teléfonos [24] y reproductores mp3.

Los algoritmos de reducción de ruido han experimentado también cambios importantes producto del surgimiento de los PSD, ahora es posible utilizar algoritmos más complejos

computacionalmente, y a través de la combinación de estos algoritmos con los micrófonos direccionales se está logrando significativos avances.

Las opciones de amplificación o control automático de ganancia se incrementan cada vez más con mejoras en la cancelación de realimentación y en la comunicación inalámbrica, ejemplo de esto lo constituyen los significativos avances obtenidos en la conducción ósea para prótesis trans-craneal [25].

Los micrófonos direccionales y algoritmos de reducción de ruido han mejorado durante los últimos años, apoyados por la reintroducción de la tecnología inalámbrica. Con la evolución de la microelectrónica, el futuro de prótesis auditivas parece muy prometedor. Con el desarrollo en el procesamiento de voz, producto del aumento de la memoria de los dispositivos, los algoritmos se irán sofisticando, intentado mejorar, cada vez más, la inteligibilidad del habla, así como la percepción de los sonidos como lo hace una persona normo-oyente [26]. Además, la conectividad con otros dispositivos, permite ampliar el campo de desarrollo y mejorar las posibilidades del usuario.

1.3 Técnicas para la comparación de prótesis auditivas digitales.

Durante las últimas tres décadas, han sido propuestas técnicas para la comparación de prótesis auditivas digitales encaminadas a hacer más eficaz el rendimiento de los dispositivos de comunicación modernos en ambientes ruidosos. Sin embargo, todavía sigue sin estar claro qué algoritmo de mejora del habla se desempeña bien en diferentes situaciones de escucha en el mundo real, en las que el nivel de ruido de fondo y sus características están constantemente cambiando. Una comparación confiable y justa entre los algoritmos ha sido difícil de alcanzar por varias razones, entre ellas la falta de base de datos de lenguaje común para la evaluación de nuevos algoritmos, las diferencias entre los tipos de ruido y las disímiles metodologías de evaluación [7].

Las prótesis auditivas digitales requieren en su etapa de desarrollo de experimentos que se validen o refuten el funcionamiento de las mismas. Esta validación se realiza mediante la medición de diferentes parámetros, ya sean objetivos o subjetivos, análisis de los espectrogramas de las señales, inspección visual de formas de ondas de señales, criterios MOS para la evaluar calidad perceptualmente, etc.

Las mediciones subjetivas son la forma más precisa y efectiva de evaluar el comportamiento de algoritmos de reducción de ruido. Sin embargo la ejecución de estas mediciones es altamente costosa y consumidora de tiempo, debido a la gran cantidad de pruebas que deben llevarse a cabo para lograr un resultado válido. Algunas de estos experimentos subjetivos son el análisis de la calidad de los audios sintetizados, la evaluación de la inteligibilidad, la valoración de que sonidos son más afectados por determinados tipos de dolencias, etc. Otros parámetros que se pueden evaluar mediante pruebas a los oyentes son la señal de voz independiente a la distorsión de la señal (SIG), el ruido de fondo independiente a la señal (BAK), el efecto general (calidad general) de la señal (OVRL), un test diagnóstico (DRT) donde el paciente debe elegir una palabra entre dos posibilidades, etc.

Debido al tiempo y costo de estas medidas subjetivas, mencionadas anteriormente, se ha incrementado el esfuerzo en los últimos años para desarrollar mediciones objetivas que logren de una manera efectiva sustituirlas y cuyo resultado mantenga una alta correlación con el obtenido en las mediciones subjetivas (Referencias). De esta forma se puede hacer menos costosa la validación de algoritmos de reducción de ruido u otros relacionados con procesamiento de voz [23].

Las mediciones objetivas se fundamentan básicamente en el cálculo de las diferencias entre la señal procesada y la original (limpia), comparándolas tanto en el dominio del tiempo como de la frecuencia. De esta forma se pueden obtener medidas como: evaluación perceptual de la calidad de voz (PESQ, por sus siglas en inglés), razón de verosimilitud logarítmica (LLR, por sus siglas en inglés), distancia espectral de pendiente ponderada (WSS por sus siglas en inglés), relación señal a ruido por segmentos (RSRseg), medidas compuestas, etc. [23].

Otras técnicas de comparación de prótesis auditivas digitales son los espectrogramas y oscilogramas de las señales. El espectrograma es una representación gráfica que permite analizar el comportamiento de las componentes espectrales a lo largo del tiempo. Este experimento se realiza con varios fines, entre los cuales se encuentran observar la afectación y conservación de la estructura de los formantes.

1.4 Conclusiones parciales del capítulo.

En este capítulo se definió el marco conceptual de la investigación, mencionando los principales tipos de afecciones en el sistema auditivo con el objetivo de conocer las deficiencias a solucionar mediante las prótesis auditivas digitales.

Se concluyó que las etapas de reducción de ruido y compensación de pérdidas auditivas son de suma importancia en el correcto funcionamiento de las prótesis auditivas digitales; por lo que el mal trabajo de una de ellas, afecta seriamente la calidad general de la señal procesada.

Se analizaron las principales técnicas de comparación de las prótesis auditivas digitales, concluyéndose que es de suma importancia la medición de parámetros objetivos y subjetivos en la validación de estas, valorándose los más susceptibles a ser utilizados para el presente trabajo.

CAPÍTULO 2. MATERIALES Y MÉTODOS

En el presente capítulo se describirán los algoritmos que serán objeto de análisis y comparación, conjuntamente al diseño del experimento que evalúa criterios de carácter objetivo y subjetivo, los cuales permitirán determinar la efectividad de cada uno de ellos.

El capítulo se desglosa de la siguiente manera: en el epígrafe 2.1 se explicarán las principales características de los métodos a evaluar, tanto de reducción de ruido como de compensación de pérdidas auditivas; las bases de datos utilizadas en los experimentos se describen brevemente en el epígrafe 2.2, justificándose las grabaciones de audios escogidas; en el epígrafe 2.3 se propone el diseño de los experimentos, los cuales, mediante el análisis de sus resultados permitirán determinar los elementos susceptibles a mejorar en cada algoritmo y/o determinar los más apropiados para su implementación en prótesis auditivas reales; y por último, termina con las conclusiones parciales del capítulo (epígrafe 2.4).

2.1 Algoritmos objeto de análisis.

De forma general, una prótesis auditiva está compuesta por los transductores tanto de entrada como de salida, filtros encargados de seleccionar qué partes del espectro de frecuencia serán procesados posteriormente y percibidos por el individuo. Dicho procesamiento incluye esencialmente un control automático de ganancias y reducción de ruido lo cual garantiza al individuo una correcta percepción del sonido tratado.

Algunos investigadores optan por desarrollar prótesis auditivas con un solo canal y optimizan los algoritmos para lograr una correcta cancelación del ruido de fondo. Muchos autores han desarrollado procedimientos para la solución de este problema que vinculan técnicas convencionales como la sustracción espectral, filtrado de Wiener, métodos basados

en modelos estadísticos; como modelos ocultos de Markov y métodos basados en wavelet. Sin embargo las técnicas que emplean un micrófono solamente en prótesis auditivas se diseñan para mejorar medidas subjetivas referidas al campo del confort del sonido [27].

Tomando como referencia la tendencia actual para los diseños de prótesis digitales auditivas, así como métodos y/o técnicas desarrolladas para el mejoramiento de las mismas, se tomaron como base para la comparación dos prótesis auditivas digitales desarrolladas en el departamento del CEETI de la Universidad Central Marta Abreu de Las Villas, las cuales serán explicadas a continuación.

2.1.1 Método 1.

El primer método fue realizado por el Ingeniero en Biomedicina Roberto Jiménez Yera (situar referencia a la tesis), y consta de 4 etapas principales: Reducción de Ruido, Banco de Filtros, Control Automático de Ganancia (AGC) y Compresión.

Con la perspectiva de seleccionar un algoritmo, apropiado para su diseño de prótesis, en la etapa de Reducción de Ruido, el autor analiza y compara tres métodos diferentes: Algoritmo de Sustracción Espectral Multi-Banda (MBSS -de sus siglas en inglés-), Filtrado de Wiener utilizando Detección de Pausas en el Habla (FW) y Algoritmo Geométrico de Reducción de Ruido (AG).

En la etapa de Banco de Filtros, con el objetivo de abordar varias técnicas populares de filtrado dentro esta rama de investigación y realizar una comparación en cuanto a desempeño de los mismos dentro de una prótesis auditiva, el autor implementó tres bancos de filtros diferentes: Banco de Filtros Uniforme, Banco de Filtros Gammatone y Banco de Filtros de 1/3 Bandas-Octava.

En el bloque de control automático de ganancia (AGC) se efectúa el reconocimiento de intensidad de la señal y se controlan automáticamente estos niveles teniendo en cuenta la prescripción del déficit auditivo que se desea tratar. El enfoque de implementación utilizado en esta etapa permite el ajuste de estos parámetros de intensidad o déficit auditivo fácilmente. Finalmente, en la etapa de compresión, se realiza la sumatoria de todas las bandas procesadas y se obtiene la señal procesada.

Como resultados de esta comparación el autor obtiene que el Algoritmo Geométrico de Reducción de Ruido (AG), entre los métodos reducción de ruidos propuestos es el que mejor desempeño tiene y análogamente concluye, de acuerdo a las medidas de evaluación, que el banco de filtro Gammatone puede ser utilizado para la implementación de la prótesis. El AGC por otra parte mostró buen desempeño frente a deficiencias auditivas reales.

2.1.1.1 Algoritmo Geométrico de Reducción de Ruido (AG).

El enfoque de sustracción espectral basado en principios geométricos aborda dos defectos muy importantes de la sustracción espectral: el ruido musical y las suposiciones de existencia de términos cruzados [23]. El enfoque se basa en la representación del espectro del habla con ruido en el plano complejo como la suma del vector espectral ruidoso y el vector espectral señal limpia.

En el enfoque de sustracción espectral utilizando algoritmos geométricos (AG), se calcula la transformada de Fourier de tiempo corto a la señal del habla con ruido (véase Ecuación (1)), donde $X(wk)$ y $D(wk)$ son el espectro del habla limpia y del ruido respectivamente para $wk=2\pi k/N$ y $k=0,1,2,\dots,N-1$, donde N es la longitud en muestras de la trama.

$$Y(wk) = X(wk) + D(wk) \quad (1)$$

Luego de multiplicar $Y(wk)$ en la Ecuación anterior por su conjugada $Y'(wk)$ y aproximando los términos resultantes que no pueden ser obtenidos directamente, se puede calcular la estimación del espectro del habla limpia [9].

En la Ecuación (1) se muestra como $Y(wk)$ para la frecuencia wk es obtenida mediante la sumatoria de dos valores complejos espectrales a la frecuencia wk . Sin embargo, $Y(wk)$ puede ser representada en el plano complejo como la suma de dos números complejos, $X(wk)$ y $D(wk)$. En la Figura 2.1 se muestra la representación de $Y(wk)$ como el vector adición de $X(wk)$ y $D(wk)$.

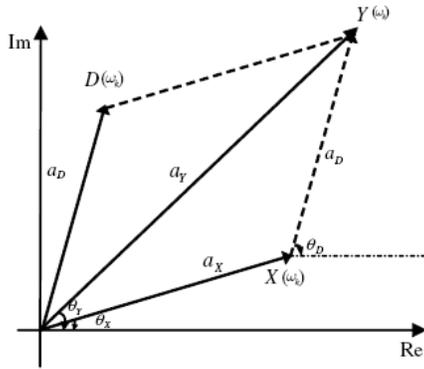


Figura 2.1. Representación del espectro del habla contaminada $Y(\omega_k)$ en el plano complejo como la suma del espectro de la señal limpia $X(\omega_k)$ y el espectro del ruido $D(\omega_k)$.

$H(\omega_k)$ es comúnmente usada como función de ganancia del algoritmo de sustracción espectral de potencia y es obtenida en [9], después de ser asumido que los términos cruzados son ceros o equivalentemente, que la diferencia de fase $[\theta_x(k) - \theta_D(k)]$ es igual a $\pm\pi/2$. Teniendo esto en cuenta, primeramente se escribe $y(n)=x(n)+d(n)$ en forma polar:

$$a_Y e^{j\theta_Y} = a_X e^{j\theta_X} + a_D e^{j\theta_D} \quad (2)$$

Donde $\{a_Y, a_X, a_D\}$ son las magnitudes y $\{\theta_Y, \theta_X, \theta_D\}$ son las fases del espectro del habla ruidoso, habla limpia y ruido respectivamente.

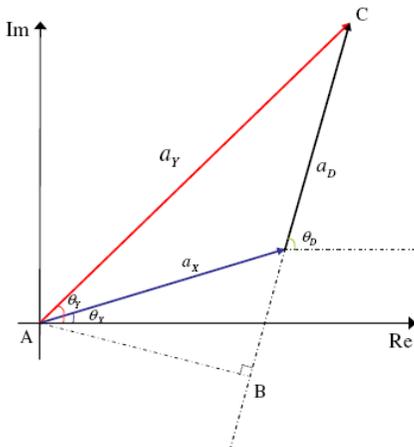


Figura 2.2. Triángulo que muestra la relación entre las fases de habla con ruido, ruido y habla limpia espectral.

Para el triángulo mostrado en la Figura 2.2, usando la ley de los senos se puede obtener

una nueva función de ganancia:

$$H_{GA} = \frac{a_X}{a_Y} = \sqrt{\frac{1-C_{YD}^2}{1-C_{XD}^2}} \quad (3)$$

La función (3) depende del cálculo de las diferencias de fase entre el ruido y la señal de ruido. Uno de las posibilidades para obtener y utilizar las diferencias de fases es a través de los principios trigonométricos. Dicho lo anterior se puede calcular CYD y CXD [23].

$$C_{YD} = \frac{\gamma+1-\xi}{2\sqrt{\xi}} \quad (4)$$

$$C_{XD} = \frac{\gamma-1-\xi}{2\sqrt{\xi}} \quad (5)$$

Donde las variables γ y ξ son definidas como:

$$\gamma \triangleq \frac{a_Y^2}{a_D^2} \quad (6)$$

$$\xi \triangleq \frac{a_X^2}{a_D^2} \quad (7)$$

Note que los términos γ y ξ son versiones instantáneas de la SNR a posteriori y a priori respectivamente [28]. Luego, mediante la sustitución de ecuaciones se obtiene la siguiente función de supresión en función de los términos γ y ξ :

$$H_{GA}(\xi, \gamma) = \sqrt{\frac{1 - \frac{(\gamma+1-\xi)^2}{4\gamma}}{1 - \frac{(\gamma-1-\xi)^2}{4\xi}}} \quad (8)$$

La ganancia obtenida en la Ecuación (8) es ideal, en la práctica esta necesita ser estimada por la observación del ruido. La implementación de la función de ganancia requiere estimaciones de γ y ξ , los cuales son valores instantáneos y de corto plazo.

Combinando las dos estimaciones de ξ se puede entonces obtener información espectral pasada y presente mediante la Ecuación siguiente:

$$\xi(\lambda, k) = \alpha \left[\frac{a_Y(\lambda-1, k)}{a_D(\lambda-1, k)} \right]^2 + (1 - \alpha) * (\sqrt{\gamma(\lambda, k)} - 1)^2 \quad (9)$$

Donde α es la constante de suavizado, y $a_D(\lambda, k)$ es un estimado de la magnitud espectral del ruido. La Ecuación (9) es un promedio ponderado de las mediciones instantáneas de la

SNR pasada y presente y la constante de suavizado controla el peso de la información pasada y futura.

Para $\gamma(\lambda, k)$, se usa la siguiente estimación instantánea:

$$\gamma(\lambda, k) = \left(\frac{a_\gamma(\lambda, k)}{a_D(\lambda, k)} \right)^2 \quad (10)$$

Donde $a_D(\lambda, k)$ es una estimación del espectro de ruido obtenido usando un algoritmo de estimación de ruido. Se considera el suavizado y el limitado de los valores de $\gamma(\lambda, k)$ para reducir las fluctuaciones rápidas relacionadas con el cálculo anterior de $\gamma(\lambda, k)$ y también limitar la sobre-supresión producto a valores grandes de $\gamma(\lambda, k)$. Entonces se calcula el suavizado de $\gamma(\lambda, k)$ y se limita H_GA (ξ, γ) a ser siempre menor o igual que 1 [9].

En modo de resumen y para aumentar la comprensión del AG se muestra su diagrama en bloques.

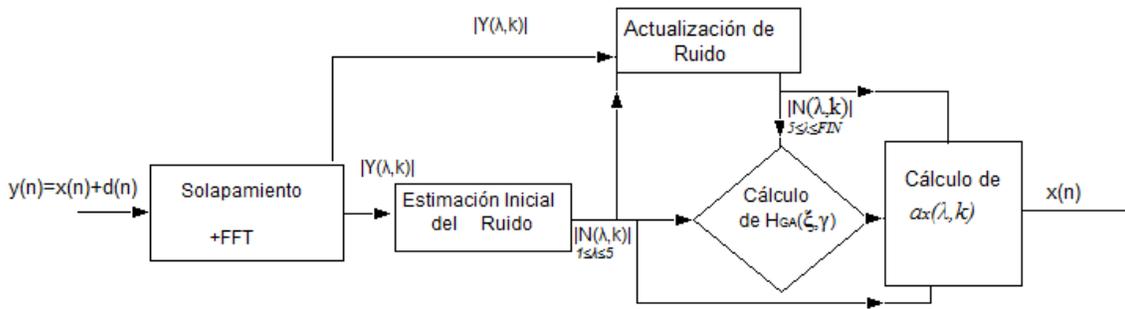


Figura 2.3. Diagrama en bloque del AG.

2.1.1.2 Banco de Filtros Gammatone.

El banco de filtro Gammatone [18] constituye un modelo de procesamiento de voz que se basa en simular el funcionamiento de la cóclea. El filtro Gammatone lleva a cabo un análisis espectral y convierte la onda acústica en una representación multi-canal del movimiento de la membrana basilar.

La respuesta de impulso de la función del filtro Gammatone suministra una compatibilidad excelente con la respuesta de impulso obtenida mediante la técnica de revcor (Correlación entre los datos de una fibra auditiva principal y la forma de onda recibida por el oído) [29].

El filtro Gammatone es definido en el dominio de tiempo por la respuesta de impulso siguiente:

$$Gt(t) = at^{(n-1)}exp(-2pbt) \cos(2pf_0t + \phi) \quad (t > 0) \quad (11)$$

Dónde:

n : orden del filtro.

f_0 : frecuencia central

b : Determina la longitud de duración de la respuesta al impulso y por lo tanto del ancho de banda del filtro Gammatone

a : constante real de normalización

p : constante π

Una característica de magnitud del filtro GT de orden 4 es que es muy similar a la función roex, comúnmente usada para representar la característica de magnitud del filtro auditivo humano. En segundo lugar, el estudio realizado por indica que los anchos de banda del filtro auditivo tienen una distancia fija en la membrana basilar. En este estudio, específicamente, se han resumido los datos humanos sobre el ancho de banda rectangular equivalente (ERB) al filtro auditivo humano como [9]:

$$ERB = 24.7(4.37fc/100 + 1) \quad (12)$$

Las ecuaciones (11) y (12) juntas definen al banco de filtro Gammatone si se incluye la suposición común de que las frecuencias centrales del filtro están distribuidas a través de la membrana basilar en proporción a su ancho de banda. Cuando fc/b es grande como en el caso auditivo, el ancho de banda del filtro es proporcional a b y la constante de proporcionalidad solo depende del orden del filtro, n . [18]. Finalmente en cuanto a la fase el GT es un filtro de fase mínima y, aunque la fase característica del filtro auditivo no se conoce, resulta razonable suponer que estará cerca a la fase mínima.

Un análisis en el dominio de la frecuencia del filtro GT revela que un n -enésimo orden del filtro GT puede ser aproximado por una cascada de n filtro idénticos de primer orden y el filtro de primero orden GT digital tiene una recursividad muy eficiente. El banco de filtro auditivo Gammatone resultante suministra una buena compensación entre la exactitud con

la que simula el filtrado coclear y la carga computacional para aplicaciones donde el nivel del estímulo es moderado.

El Banco de Filtro Gammatone implementado, consta de 17 bandas de frecuencia (ajustables) y su ancho de banda está dado por la Ecuación (12).

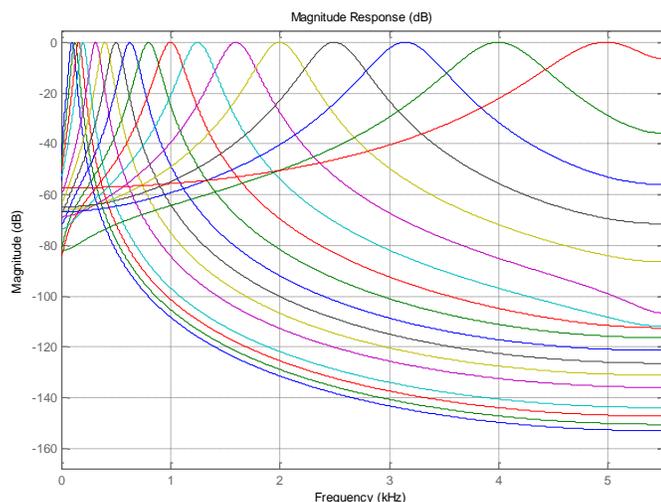


Figura 2.4. Respuesta de impulso del banco de filtro Gammatone.

2.1.1.3 AGC.

En procesamiento digital de voz resulta crucial la etapa en la que se realiza la calibración de intensidad de la señal del habla para su posterior procesamiento. Un elemento esencial en este tipo de transformaciones es establecer una correspondencia entre la intensidad del sonido en decibeles y el volumen realmente percibido por el oído. Con el objetivo de lograr esta conversión entre valores de intensidad se hace una calibración de la señal a modificar [30].

En la plataforma de desarrollo de software utilizada (MatLab) [30], los sonidos que son adquiridos mediante la lectura de archivos .wav, son normalizados para que sus amplitudes queden dentro de rango entre 1 y -1; lo que no representa los verdaderos valores de reproducción del sonido. Para lograr una representación real de los niveles de sonido, su amplitud debe ser escalada obteniendo el sonido a los niveles de dB SPL deseados [30].

Las conversiones de este tipo requieren de la elección de un nivel de intensidad de referencia (en dB SPL), correspondiente al nivel de intensidad sonora presente en el momento de la grabación. [30].

La transformación entre los niveles de intensidad se hace según la ecuación:

$$st = \frac{s}{\sqrt{\frac{\sum_{l=1}^L s^2(l)}{L}}} N, \quad (13)$$

Donde: s es la señal original, st la señal luego de su transformación, L la longitud de s y N el equivalente lineal del valor en dB SPL de referencia seleccionado. Se calcula según la relación $N_{dB SPL} = 20 \log N$.

Los pacientes con sordera experimentan pérdidas auditivas en una o varias bandas de frecuencia; para que estas bandas sean escuchadas es necesario hacer un procedimiento de amplificación a la señal del habla con el objetivo de acondicionar la misma para asegurar la comprensión de estas por parte del paciente. Este proceso de amplificación o AGC no es más que la multiplicación de la trama o segmento de la señal que se está analizando por una ganancia determinada, la cual está en correspondencia con el umbral auditivo del paciente y debe cumplir como requisito esencial, no sobrepasar este umbral. El algoritmo de AGC que utiliza el autor es una arista del estudio realizado en [31]. Donde se definen primeramente, los umbrales de intensidad mínimo ($UmBmin$) y máximo ($UmBmax$) de escucha, los mismos son 0db y 120db respectivamente, el objetivo de definir estos parámetros es calcular un coeficiente de ganancia normalizado para la trama que se analiza y asegurar de esta manera que la intensidad de la trama después de ser amplificada no sobrepase el umbral del dolor y fluctúe dentro del rango de escucha humana.

El cálculo de este coeficiente se realiza mediante la Ecuación siguiente:

$$\text{Coef (db)} = (UmBmax - UmBd) / (UmBmax - UmBmin) \quad (14)$$

Donde $UmBd$ se define como la intensidad de referencia, es decir, nivel al cual deben ser llevados cada intensidad presente en la trama que se procesa, o lo que es lo mismo, el umbral de deficiencia auditiva del paciente en ese segmento de señal. Para lograr que todas las muestras en la trama tengan el mismo nivel de intensidad es preciso analizar una por una. Supongamos que se está realizando el proceso de amplificación a la muestra n , de

intensidad ($InT(n)$) y después de determinar su amplitud en db SPL se desea calcular el nuevo valor intensidad ($InT''(n)$) para sustituir a la ($InT(n)$) y de esta forma lograr que la muestra n cumpla con el requisito de ser aproximadamente igual en amplitud a UmB . El cálculo de ($InT''(n)$) se lleva a cabo mediante la fórmula siguiente:

$$InT''(n) = UmBd + (InT(n) - UmBmin) * Coef \quad (15)$$

Obtenido entonces la intensidad deseada para la muestra n es preciso realizar el proceso inverso a la conversión en db SPL para crear una nueva señal, esta nueva señal tendrá de diferencia con la señal original solamente los niveles de intensidad modificados pero la información de las dos será exactamente la misma [9].

2.1.2 Método 2.

El segundo método fue desarrollado por la Ingeniera en Biomedicina Beatriz Peón Pérez y son algoritmos para prótesis auditivas digitales utilizando la Transformada Wavelets Packets.

La aplicación de la Transformada Wavelet a una señal es equivalente a la implementación de un banco de filtros no uniforme que usa una ventana de tamaño variable para diferentes componentes de frecuencia. Permite obtener intervalos largos de tiempo para bajas frecuencias y cortos para frecuencias altas; aumentando la resolución temporal cuando los cambios en el tiempo son más lentos, de esta forma se logra una mejor representación de la señal en bandas de frecuencia.

La descomposición es análoga a un filtrado sucesivo cuya frecuencia de corte es siempre la mitad de la mayor frecuencia de la señal. Esta forma de descomposición en bandas de frecuencia se asemeja en gran medida a la descomposición del sonido realizada por la cóclea del sistema auditivo humano. Esto sugiere que la aplicación de la Transformada Wavelet y WP en este tipo de procesamiento sea adecuada. En la implementación de los algoritmos de reducción de ruido y de compensación de la pérdida auditiva la autora emplea la wavelet db8 [27].

2.1.2.1 Algoritmo de reducción de ruido basado en Wavelet Packets.

Las técnicas de reducción de ruido basadas en la TW se sustentan en las diferencias espectrales del ruido y la señal del habla. Los componentes de ruido se encuentran

distribuidos en coeficientes de valor absoluto cercano a cero en las descomposiciones mediante TW, lo que posibilita una mejor eliminación del mismo sin afectar la información útil de la señal de voz.

Sin embargo, métodos de aplicación de umbral como el introducido por Donoho, 1995[27] presenta algunas dificultades cuando son aplicados a señales corrompidas por ruido, sobre todo si es ruido real.

El algoritmo elegido por el autor se basa en los descritos por Ghanbari y cols. 2005, [32]; haciendo modificaciones en algunos aspectos de su funcionamiento con el objetivo de mejorar su desempeño. En la Figura 2.5 se muestra un diagrama en bloques que representa las etapas de dicho algoritmo.

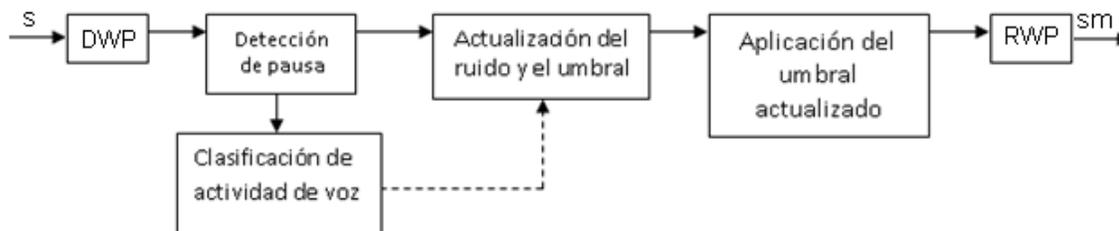


Figura 2.5. Diagrama en bloques del algoritmo de reducción de ruido basado en WP. Donde s: señal ruidosa, sm: señal modificada mediante filtrado con base wavelet, DWP: Descomposición mediante WP, RWP: Reconstrucción mediante WP.

Como se puede apreciar en los bloques de la Figura 2.5. inicialmente se calculan los coeficientes wavelets de la trama de la señal a partir de la descomposición mediante WP, (DWP). Luego se realiza un proceso de detección de pausa basado en la energía de los coeficientes. Si la trama de la señal es detectada como pausa se actualizan los valores del patrón de ruido y el umbral y energía de ruido en cada banda de frecuencia. Posteriormente se obtiene la relación señal a ruido, también por banda entre la trama de la señal no detectada como silencio y el ruido estimado. A partir de esta relación y del umbral actualizado se calcula un nuevo umbral. El bloque nombrado como Clasificador de actividad de voz se encarga de etiquetar los segmentos como sordos “unvoiced” o sonoros “voiced”. Para prevenir degradación en las regiones sordas, primero se detectan los segmentos sordos de los sonoros. En este sistema la detección sorda-sonora se hace basada

en la transformada Wavelet Packets. A partir de este bloque se modifican los valores de umbral obtenidos luego de la detección de pausa.

Posterior al cálculo del umbral, este se le aplica a los coeficientes de señal para obtener los coeficientes de la señal filtrada. El último bloque del diagrama pertenece al proceso de recomposición de la señal mediante WP, (RWP) [27].

Para la estimación del ruido se realiza una descomposición en tres niveles lo que provocaría un total de 8 bandas de frecuencia. Cuando una trama de la señal es detectada como pausa o en ausencia de voz se toma el patrón de ruido como la trama en sí y los coeficientes wavelet del ruido (CWR) como los coeficientes que resultan de la descomposición de la trama, en todas las bandas de frecuencia. Para todas las tramas de la señal se actualiza la energía estimada del ruido (EEruido) en cada banda de frecuencia, a partir de los CWR de la siguiente forma:

$$EEruido_{k,n} = \alpha EEruido_{k,n-1} (1-\alpha) \sum_{i=1}^{L_k} CWR_k^2(i) , \quad (16)$$

Se denota cada banda de frecuencia con el índice k y cada trama de la señal con n, α es un factor establecido por optimización, $0 < \alpha < 1$. La cantidad de coeficientes de la k-ésima banda de frecuencia se identifica por L_k . Para la primera trama de la señal la energía del ruido se calcula como se describe en [27].

Para la obtención de umbrales con el objetivo de eliminar ruido en señales de voz, se utiliza el umbral básico para uso en algoritmos basados en Wavelet Packets como se muestra en la ecuación siguiente:

$$U = \hat{\sigma} \sqrt{2 \log(L \log_2 L)} , \quad (17)$$

Siendo L la longitud de la señal ruidosa y $\hat{\sigma}$ la desviación estándar del ruido

Cuando una trama de la señal es detectada como silencio se actualiza el umbral.

Luego se realiza el cálculo de un umbral según la ecuación (18) que posibilita el ajuste a los cambios en las características inherentes a la señal y el ruido.

$$U_{k,n} = \begin{cases} U_{k,n} \left(1 + e^{-\frac{RSRb_{k,n}}{t}}\right) & RSRb \geq 0 \\ 2U_{k,n}RSRb & RSRb < 0 \end{cases}, \quad (18)$$

El factor t , utilizado en la expresión anterior, se calcula de manera experimental.

La aplicación del umbral adaptativo $U_{k,n}$ a los coeficientes wavelet de la señal (CWS) se hace en cada trama y para todas las bandas de frecuencia; a partir de una modificación del umbral duro convencional. Cuando el valor absoluto del coeficiente está por debajo del valor del umbral, este no se elimina completamente sino que se reduce de manera sustancial, donde el factor d se obtiene de manera experimental. En el algoritmo la autora utiliza un valor de d igual a 4, este hace una eliminación lo suficientemente fuerte de los coeficientes.

$$CWSM_{k,n} = \begin{cases} CWS_{k,n} & |CWS_{k,n}| \geq U_{k,n} \\ signo(CWS_{k,n}) \frac{|CWS_{k,n}|^d}{U_{k,n}^{d-1}} & |CWS_{k,n}| < U_{k,n} \end{cases}, \quad (19)$$

De esta forma se obtienen los coeficientes wavelets de la señal modificada de la k -ésima banda de frecuencia y la n -ésima trama de la señal. A partir de estos coeficientes se obtiene la señal filtrada mediante la recomposición por WP.

La detección de pausas permite establecer las diferencias entre silencio y voz en presencia de un ambiente ruidoso. Para la clasificación de un segmento de señal del habla como silencio se utiliza una comparación entre la energía de los coeficientes de la señal modificada y la energía del ruido estimado. La detección de la actividad de voz parte de asumir que la primera trama de la señal es silencio por lo tanto en esta trama los valores de E_{ruido} , $RSRb$ y umbral se actualizan. Para las tramas posteriores de la señal la detección se hace según la secuencia de pasos descrita en [27].

Posterior a la detección de una trama como pausa se procede a reemplazar los valores de ruido, umbral y E_{ruido} . Estos valores son utilizados en las tramas siguientes para realizar el mismo proceso de detección de segmentos de silencio. Luego de la obtención del umbral este se aplica a los coeficientes de la trama. Con el objetivo de obtener la señal filtrada se reconstruye la señal a partir de los coeficientes wavelet modificados.

La clasificación de las tramas de la señal en sordas o sonoras se hace basada en la Transformada Wavelet Packets. En este método para cada banda de frecuencia se calcula la energía promedio. Acumulando la energía promedio de las bandas por debajo de 2kHz se obtiene EL (energía de baja frecuencia) y similarmente, mediante la acumulación la energía promedio de las bandas por encima de 2kHz tenemos EH (energía de alta frecuencia) [32].

Luego de la clasificación se modifica el valor de umbral que le será aplicado a los coeficientes, incrementando el mismo para las bandas altas de los segmentos etiquetados como sonoros y para las bandas bajas de frecuencia de los segmentos considerados como sordos. Las bandas por debajo de 2kHz son las nombradas como bandas bajas y las que se encuentran por encima de este valor como bandas altas. [32].

2.1.2.2 Algoritmo de compensación de pérdidas auditivas mediante Wavelet Packets.

Primeramente se realiza la calibración de la señal de igual forma que en el epígrafe 2.1.1.3, llevando los niveles de intensidad a dB SPL.

El análisis de señales mediante WP posibilita una forma de descomposición en frecuencia útil en este tipo de aplicaciones. A partir del umbral de audición de discapacidad y del rango dinámico audible es posible lograr una modificación acertada de los coeficientes proporcionados en la descomposición mediante WP con el objetivo de compensar la pérdida auditiva. Además la compresión de la amplitud mediante WP brinda un ajuste rápido de la ganancia de la señal del habla para asegurar la amplificación de los componentes de bajo nivel de intensidad sin que los de mayor amplitud sobrepasen el umbral del dolor [27].

Con el propósito de alcanzar una modificación enfocada a la compensación de este tipo de pérdidas auditivas, la autora propone un algoritmo que utiliza técnicas de compresión de amplitud con WP basado en el diseño utilizado por Trenas, M. A, y cols, 1999[31].

El objetivo de esta modificación de la señal es aumentar la intensidad por encima del umbral de audición del individuo, posibilitando una compensación para su hipoacusia.

La modificación realizada a los CW de la señal en cada banda de frecuencia se corresponde con el umbral de audición propio del individuo, además de la razón existente entre el rango dinámico de discapacidad del sujeto y el rango dinámico normal de audición. Luego de

obtener la equivalencia en dB de los coeficientes de la señal ($CWdb_{k,n}$) se obtienen los coeficientes modificados también en dB según la ecuación (20)[27].

$$CWMdb_{k,n} = Ud + (CWdb_{k,n} - Un) * Rrd, \quad (20)$$

La relación entre los coeficientes de la señal y su valor en dB es de la forma:

$$CWdb = 20 \log Cw, \quad (21)$$

Los valores de umbral de discapacidad, umbral normal y umbral del dolor se denotan como Udk , Unk y $Udolork$ de la k -ésima banda de frecuencia, respectivamente y la razón del rango dinámico (Rrd) se calcula para cada banda como:

$$Rrd = \frac{Udolor - Ud}{Udolor - Un'} \quad (22)$$

En la implementación de los algoritmos de reducción de ruido y de compensación de la pérdida auditiva la autora emplea la wavelet db8.

2.2 Características de las Bases de Datos empleadas.

Para obtener resultados válidos y repetibles en los experimentos que se proponen realizar para la evaluación de las prótesis, se seleccionaron grabaciones de audio provenientes de bases de datos internacionales de dominio público. De estas bases de datos se eligieron un subconjunto de alocuciones que garantizasen variabilidad en la información fonética y presencia de información ruidosa proveniente de escenarios clásicos, como restaurantes, mercados, estaciones de transporte, automóviles, etc. Esta selección permite analizar el comportamiento de la prótesis ante situaciones reales. Se utilizaron alocuciones provenientes de bases de datos en inglés y español, grabadas en contextos de mínimo ruido.

El material de habla utilizado para las pruebas de evaluación de los algoritmos fueron las bases de datos AHUMADA, AURORA y el sitio Web *HearingLossSamples*. A continuación se explicarán las características principales de algunas de ellas.

2.2.1 AHUMADA.

La base de datos AHUMADA fue elaborada en Madrid en 1998. Contiene grabaciones de alocuciones en español de 103 hombres en dos versiones que difieren por las características del canal: una es por teléfono y la otra con micrófono de alta calidad. Es una base de datos

multisesión, con 3 sesiones de grabación por cada versión. El intervalo intercesión es de aproximadamente 20 días. AHUMADA está diseñada principalmente para experimentos de identificación de hablantes y presenta un texto equilibrado fonológicamente, lo que significa que todos los fonemas del español se encuentran repetidos con su frecuencia de aparición relativa a lo largo del texto.

La frecuencia de muestreo es 16.000 Hz para los ficheros que se corresponden con las sesiones microfónicas y 8000 Hz para los que se corresponden con las sesiones donde ha estado implicado el teléfono. El formato de todos los ficheros es binario de 16 bits. Cada fichero posee una cabecera de 20 bytes propias del software utilizado en la segmentación y que son ajenas a las muestras de voz del locutor.

2.2.2 AURORA y sitio Web HearingLossSamples.

Los ruidos para contaminar las señales fueron tomados de la base de datos AURORA [33] con una frecuencia de muestreo de 8000Hz, un tamaño de muestra de 16 bits, con un formato de audio PCM. Algunos de los ruidos con que cuenta esta base de datos provienen de escenarios muy comunes en la vida cotidiana: multi-locutor (babble), coches (car), tren (train), sala de exposiciones, restaurantes, calle, aeropuerto, etc.

En el sitio web HearingLossSamples[34] se obtuvieron las señales para la validación del algoritmo de compensación de pérdidas auditivas. Estas señales se adquirieron con una frecuencia de muestreo de 44KHz, con un solo canal (mono) y una velocidad de transmisión de 64 kbps.

2.3 Evaluación objetiva a prótesis digitales auditivas.

En la actualidad las mediciones subjetivas son la forma más precisa y efectiva de evaluar el comportamiento de algoritmos de reducción de ruido. Sin embargo, no siempre es posible realizar este tipo de mediciones. Por este motivo se ha incrementado el esfuerzo en los últimos años para desarrollar mediciones objetivas que logren sustituirlas de una manera efectiva y cuyo resultado mantenga una alta correlación con el obtenido en las mediciones subjetivas [23]. De esta forma se puede hacer menos costosa la validación de algoritmos de reducción de ruido u otros relacionados con procesamiento de voz.

2.3.1 Evaluación empleando mediciones objetivas.

Las medidas objetivas que evalúan la calidad del habla intentan predecir la calificación que daría un grupo de oyentes en base a medir alguna distancia entre la señal limpia y la obtenida luego del procesamiento. Aunque esta distancia debería ser perceptualmente significativa, las medidas analíticas clásicas se han centrado principalmente sobre diferencias simples en las representaciones temporales o espectrales de ambas señales. Para este trabajo se escogieron las medidas: evaluación perceptual de la calidad de voz (PESQ, por sus siglas en inglés), razón de verosimilitud logarítmica (LLR, por sus siglas en inglés), distancia espectral de pendiente ponderada (WSS por sus siglas en inglés), relación señal a ruido por segmentos (RSRseg) y medidas compuestas. Estos parámetros se eligieron por ser los más utilizados por los investigadores en los experimentos de reducción de ruido.

2.3.1.1 Evaluación perceptual de la calidad de voz (PESQ).

Es la medida objetiva más compleja computacionalmente, se calcula mediante una combinación lineal del valor del disturbio promedio D_{prom} y del valor del disturbio promedio asimétrico A_{prom} de la siguiente forma:

$$PESQ = a_0 + a_1 D_{prom} + a_2 A_{prom}, \quad (23)$$

Donde los parámetros a_0 , a_1 y a_2 fueron calculados mediante optimización mediante un análisis múltiple de regresión lineal por (Hu y Loizou, 2008) para distintos tipos de análisis como distorsión de la señal y calidad general de la voz. Para distorsión de la señal se utilizaron los valores siguientes: $a_0 = 4.754$, $a_1 = -0.186$, $a_2 = -0.008$; para calidad general de la voz estos valores se sustituyen por: 4.906 , -0.148 y -0.0021 respectivamente; mientras que para distorsión de ruido se utilizan los siguientes $a_0 = 5.611$, $a_1 = -0.070$, $a_2 = -0.068$. Los valores de D_{prom} y A_{prom} fueron tratados como variables independientes en un análisis múltiple de regresión lineal. Para la medición de PESQ las señales original y procesada son ecualizadas a un nivel estándar de audición y alineadas temporalmente para evitar efectos propios del retardo introducido por el procesamiento. El cálculo de esta medida produce un escala entre 1.0 y 4.5 en la que los valores altos indican mejor calidad de voz.

2.3.1.2 Razón de verosimilitud logarítmica (LLR).

Es una medida basada en el análisis de coeficiente de predicción lineal (LPC por sus siglas en inglés) de las señales. Esta medida se define como:

$$d_{LLR}(\vec{a}_p, \vec{a}_c) = \log \left(\frac{\vec{a}_p R_c \vec{a}_p^T}{\vec{a}_c R_c \vec{a}_c^T} \right), \quad (24)$$

Donde \vec{a}_c es el vector LPC de la trama de la señal original del habla, \vec{a}_p es el vector LPC de la señal procesada y R_c la matriz de autocorrelación de la señal original.

Esta medida valora las diferencias de la envolvente espectral entre la señal original y la procesada o corrompida. Los valores de LLR por segmentos están limitados al rango entre 0 y 2, mientras más cercano a cero esté el valor de LLR menos diferencias existen entre la señal procesada y la original.

2.3.1.3 Distancia del espectro de pendiente ponderada WSS.

La medición de la distancia WSS calcula las diferencias de los pesos entre las pendientes espectrales en cada banda de frecuencia. La pendiente espectral es obtenida como la diferencia entre magnitudes espectrales adyacentes, en decibeles. Esta medida se define como:

$$d_{WSS} = \frac{1}{M} \sum_{m=0}^{M-1} \frac{\sum_{j=1}^K W_{WSS}(j,m) (S_c(j,m) - S_p(j,m))^2}{\sum_{j=1}^K W_{WSS}(j,m)}, \quad (25)$$

Donde $W_{WSS}(j, m)$ es el peso calculado por Ma y cols, 2009, K el número de bandas, M el número de segmentos de la señal y $S_c(j, m)$ y $S_p(j, m)$ son las pendientes espectrales de la banda de frecuencia j de la señal limpia y procesada respectivamente.

2.3.1.4 Relación señal a ruido por segmentos (RSRseg).

Esta medida se centra sobre la diferencia entre la forma de onda de la señal limpia y de la estimada. Se calcula promediando las SNR obtenidas para cada tramo de análisis, donde la SNR de cada segmento puede tomar valores entre -10 y 35 dB. Se supone que fuera de este

rango la variación de la SNR no está acompañada de una variación en la percepción del habla. La medida está dada entonces por:

$$RSRseg = \frac{10}{M} \sum_{m=0}^{M-1} \log_{10} \frac{\sum_{n=Nm}^{Nm+N-1} x^2(n)}{\sum_{n=Nm}^{Nm+N-1} (x(n) - \hat{x}(n))^2}, \quad (26)$$

Donde $x(n)$ es la señal limpia, $\hat{x}(n)$ es la señal procesada, N el tamaño de la trama y M es el número de tramas en la señal.

2.3.1.5 Medidas objetivas compuestas.

Existen también medidas objetivas compuestas por combinaciones lineales de algunas de las explicadas anteriormente. Fueron desarrolladas por Hu y Loizou, 2006 con el objetivo de mejorar los coeficientes de correlación entre la valoración obtenida por estas medidas sobre la calidad de voz y la evaluación resultante de las medidas subjetivas. Mediante regresión lineal se forman las siguientes medidas compuestas:

1. Csig: Representa en una escala de cinco puntos la distorsión de la señal del habla (SIG -de sus siglas en inglés-) y está formada por la combinación lineal de las medidas objetivas, LLR, PESQ y WSS.
2. Cbak: Representa en una escala de cinco puntos la molestia del ruido de fondo (BAK -de sus siglas en inglés-) y está formada por la combinación lineal de las medidas objetivas, segSNR, PESQ y WSS.
3. Covl: Calidad del conjunto usando el criterio MOS, (OVRL - de sus siglas en inglés-) y está formada por la combinación lineal de las medidas objetivas, PESQ, LLP y WSS.

Estas medidas compuestas presentan un moderado incremento de los coeficientes de correlación con respecto a las medidas convencionales [23].

2.3.2 Diseño de los experimentos.

2.3.2.1 Evaluación de los algoritmos de reducción de ruido.

En este trabajo, se evalúa el desempeño de varias medidas objetivas en términos de la predicción de la calidad de la voz con el objetivo de comparar dos algoritmos de reducción de ruido: Algoritmo basado en Wavelet Packets y algoritmo basado en el Método de

Sustracción Espectral Geométrico, frente a ruidos típicos y simulados, empleando medidas objetivas y compuestas, tales como LLR, PESQ, SNRseg, WSS, Covl, Cbak y Csig (explicadas en el sub-epígrafes anteriores). Estas mediciones permiten ofrecer criterios de calidad globales que en mayor o menor grado reflejan la calidad perceptual de la señal filtrada.

Se tomaron ocho señales para la comparación de los algoritmos procedentes de la base de datos AHUMADA, muestreadas a 8000Hz y 16bpm (bits por muestra). Las alocuciones fueron pronunciadas por voces masculinas.

Las señales fueron contaminadas artificialmente con tres tipos de ruido: murmullo (babble), ruido presente en autos (car) y ruido de tren (train), a diferentes relaciones señal a ruido. Se tomó en consideración los siguientes niveles de SNR 0dB, 5dB, 10dB, 15dB, de forma similar a otros estudios reportados en la bibliografía consultada. Los ruidos fueron tomados de la base de datos AURORA [38] con una frecuencia de muestreo de 8000Hz [7] [23].

Con vistas a ofrecer evidencias gráficas se propone el empleo de espectrogramas y oscilogramas de las ocho señales, con el objetivo de observar las pérdidas por procesamiento inferidas a las grabaciones por cada método.

2.3.2.2 Evaluación de los algoritmos de compensación de pérdidas auditivas.

Para la comparación de los algoritmos de compensación se utilizaron señales de voz que tenían aplicada una simulación de pérdidas auditivas; de manera tal que tienen una variación específica de la potencia por bandas de frecuencia. Estas señales (seis) se obtuvieron en el sitio web *HearingLossSamples* [34] con una frecuencia de muestreo de 44KHz. Se procedió a realizarle un sub-muestreo a 8000Hz; ya que el análisis de la señal se hace alrededor de las frecuencias entre 500 y 8000Hz, pues son las frecuencias utilizadas en los audiogramas. Mediante este proceso no ocurren pérdidas importantes de la información de la señal pues los componentes de la voz están por debajo de los 4kHz de frecuencia y las bandas que más interesa analizar se encuentran por debajo de 8KHz.

Para la comprobación del desempeño de estos algoritmos se utilizaron señales con pérdidas inducidas en las frecuencias medias y altas. El objetivo de la compensación consiste en

modificar la potencia por bandas de frecuencia de la señal con pérdida simulada de forma que quede lo más cercana posible a los niveles de audición normal.

El análisis de los resultados se realizó de mediante espectrogramas de las señales amplificadas y gráficas de intensidad contra frecuencia.

2.4 Evaluación subjetiva a prótesis digitales auditivas.

Con el objetivo de mejorar la calidad del habla y preservar, o incluso incrementar, su inteligibilidad se han desarrollado muchos algoritmos de reducción de ruido, sin embargo la evaluación del éxito de los distintos algoritmos en esta tarea no es un problema sencillo. A diferencia de otros campos donde una medida computacional simple (como el error cuadrático medio) puede ser suficiente, la inteligibilidad y la calidad del habla son fenómenos complejos de percepción cuya descripción en términos puramente matemáticos no resulta eficaz. Esto ha conducido a la aceptación general de que las pruebas subjetivas son el mejor instrumento para evaluar en última instancia el desempeño de las distintas tecnologías de procesamiento de voz.

2.4.1 Experimento subjetivo de reducción de ruido.

Este experimento que se define a continuación se ha diseñado siguiendo los requisitos del estándar ITU-T (P.835), para incitar a los oyentes a integrar los efectos de la señal y de la distorsión de fondo en sus evaluaciones de la calidad en general. Se informa sobre la comparación y evaluación subjetiva de dos métodos de reducción de ruido: el de Sustracción Espectral Geométrico y el basado en Wavelet Packets.

Para reducir la duración de las evaluaciones subjetivas, sólo un subconjunto de frases de la base de datos AHUMADA (explicada en el epígrafe 2.2.) fue procesado por los 2 algoritmos para la evaluación oficial subjetiva. Un total de 4 oraciones (véase Tabla 2.1) fueron contaminadas con dos ambientes de ruido de fondo (murmullo y coche) en dos niveles de SNR (5 dB y 10 dB), y luego fueron procesadas. Los ruidos se tomaron de la base de datos AURORA. Estas oraciones fueron realizadas por cuatro parlantes masculinos y se utilizaron por poseer frases fonéticamente equilibradas con un contexto de palabras con relativa baja previsibilidad.

Tabla 2.1. Oraciones utilizadas en la comparación subjetiva.

Señal #1	He comprado en la tienda tres camisetas y un vestido rojo.
Señal #2	Detrás de la escuela veíamos poco cine antiguo.
Señal #3	Mi estufa caliente bien el salón de unos diez metros.
Señal #4	No salten sobre la picota en lugar de estudiar.

2.4.1.1 Metodología empleada en las pruebas.

Las pruebas subjetivas fueron diseñadas de acuerdo a la recomendación ITU-T P.835, la cual fue planteada para reducir la incertidumbre del oyente en una prueba subjetiva, de manera que pueda distinguir la señal de voz, el ruido de fondo, o ambos, los que deben constituir la base de sus calificaciones de la calidad en general. Este método instruye al oyente a participar de forma sucesiva y estimar la señal de voz mejorada en:

1. La señal de voz independiente utilizando una escala de cinco puntos para medir la distorsión de la señal (SIG) (Tabla2.2).
2. El ruido de fondo independiente utilizando una escala de cinco puntos para medir el ruido intrusivo. (Tabla2.3)
3. El efecto general usando la escala de la *Mean Opinion Score* (OVRL) [1 = mala, 2 = pobre, 3=aceptable, 4 = buena, 5 = excelente.]

Tabla 2.2. Escala de la señal distorsionada.

Puntuación	Descripción
5	Muy natural, no degradada.
4	Bastante natural, pequeña degradación.
3	Algo natural, algo degradada.
2	Bastante no natural, bastante degradada.

1	Muy no natural, muy degradada
---	-------------------------------

Tabla 2.3. Escala del ruido de fondo.

Puntuación	Descripción
5	No perceptible.
4	Algo perceptible
3	Perceptible pero no intrusivo
2	Bastante visible, algo intrusivo
1	Muy visible, muy intrusivo

2.4.1.2 Preparación de las secuencias de pruebas.

Un total de 16 oyentes hablantes hispanos fueron reclutados para las pruebas de escucha, los cuales se dividieron en cuatro grupos. Los oyentes estaban entre las edades de 20 y 50 años y nunca habían participado en una prueba de audición.

Las pruebas se dividieron en dos sesiones. A cada oyente se le reprodujeron tríadas de muestras con intervalos de diez minutos entre cada una de las tríadas, durante cuatro días de pruebas. El estándar P.835 permite el uso de tríadas formado por tres muestras diferentes o la misma muestra repetida tres veces; en este experimento se aplicó la primera opción. De acuerdo con la norma P.835, para la mitad de las pruebas en el experimento el orden de la escala de calificación fue SIG, BAK, y OVRL, y para la otra mitad de las pruebas, el orden fue BAK, SIG, y OVRL. Las condiciones fueron asignadas de tal manera que los principales factores involucrados en el experimento (algoritmo, el tipo de ruido, SNR) no mantuvieran un orden en específico.

2.4.1.3 Presentación de audio.

El material procesado del habla fue presentado a los oyentes sentados por separados y se proyectaron visualmente las frases que escuchaban. En cada estación de escucha las

escalas de calificación fueron presentadas en un monitor de PC y las clasificaciones se registraron con un teclado de PC.

Se utilizaron para las pruebas los auriculares Nippon América CD 900, los cuales tienen una gama de frecuencia de reproducción de 4Hz a 24kHz y una sensibilidad de 100 dB/mW. Los sujetos fueron instruidos para usar los auriculares en su oído preferido para escuchar. El otro oído estaba destapado con un constante nivel de ruido ambiente que se mantuvo alrededor de 25dB.

Al final de las pruebas los oyentes emitieron un total de 1536 evaluaciones; 96 criterios por cada oyente y 32 muestras de audio modificadas.

2.4.2 Experimento subjetivo de compensación de pérdidas auditivas.

En este experimento se informa sobre la comparación y evaluación subjetiva de dos métodos de compensación de pérdidas auditivas:

- Basado en Wavelet Packets.
- Utilizando Banco de Filtro Gammatone.

Para el experimento de compensación pérdidas auditivas se reclutaron 32 oyentes con dominio del idioma inglés, los cuales fueron divididos en cuatro grupos de ocho para cada prueba. Las muestras fueron presentadas a los oyentes en grupos de cuatro, con un descanso corto de diez minutos entre cada uno de estos cuartetos. Los encuestados se encontraban entre las edades de 20 y 45 años.

Se utilizó como material de muestra una alocución en inglés a la cual se le aplicó seis simulaciones de pérdidas auditivas, las cuales fueron obtenidas en el sitio web *HearingLossSamples* [34]. Estas señales tenían pérdidas en las medias y altas frecuencias fundamentalmente.

Se utilizó para la calificación de las señales criterios MOS (Mean Opinion Score) (OVRL) [1 = mala, 2 = pobre, 3=aceptable, 4 = buena, 5 = excelente.], que nos brinda una calidad general de la señal amplificada.

La presentación de audio se desarrolló de manera idéntica al primer experimento subjetivo (sub-epígrafe 2.4.1.3.).

Al final de las pruebas los oyentes emitieron un total de 384 evaluaciones; 12 criterios por cada oyente y 12 muestras de audio modificadas.

2.5 Conclusiones parciales del capítulo.

Se han descrito los algoritmos de reducción de ruido y de control automático de ganancia de los algoritmos a comparar: Método basado en Wavelet Packets y Método que utiliza Sustracción Espectral Geométrica (reducción de ruido) y Bancos de Filtros Gammatone (compensación de pérdidas auditivas).

Se preparó el diseño de los experimentos, que incluyen pruebas objetivas y subjetivas, acorde a requerimientos y normas actuales con excelente prestigio. Las medidas objetivas y subjetivas servirán de base para el análisis comparativo de los desempeños de los algoritmos de reducción de ruido y compensación de pérdidas auditivas elegidos. Se emplearon alocuciones de bases de datos internacionales de dominio público en español, sin embargo, algunas de las medidas propuestas son evaluadas en el marco del idioma Inglés.

Se utilizó el software Matlab versión 7.9b para la implementación de los algoritmos. Este software brinda múltiples facilidades en el procesamiento digital de señales y es altamente utilizado en la actualidad en estas aplicaciones.

CAPÍTULO 3. RESULTADOS Y DISCUSIÓN

En este capítulo se comparan los algoritmos de reducción de ruido y compensación de pérdidas auditivas explicados en el capítulo anterior. En el epígrafe 3.1 se discuten los resultados de los parámetros objetivos medidos y la información brindada por los oscilogramas y espectrogramas de las señales modificadas por los métodos de comparados. Los experimentos subjetivos se analizan en el epígrafe 2.2. En el epígrafe 2.3 se muestra y discute la correlación existente entre los parámetros medidos (objetivos y subjetivos) para observar la fiabilidad de los experimentos. Por último se emiten las conclusiones parciales del capítulo.

3.1 Comparación objetiva de las prótesis auditivas digitales.

3.1.1 Desempeño de los algoritmos de reducción de ruido.

Se condujo un experimento como el descrito en el sub-epígrafe 2.3.2.1. que contempla el desempeño de los algoritmos de reducción de ruido basados en Wavelet Packets (WP) y el método de Sustracción Espectral Geométrico (AG), frente a ruidos típicos y simulados, empleando medidas objetivas y compuestas. Las señales fueron contaminadas artificialmente con tres tipos de ruido: murmullo (babble), ruido presente en autos (car) y ruido de tren (train). Se tomaron en consideración los siguientes niveles de SNR: 0dB ,5dB, 10dB y 15dB.

En las tablas 3.1, 3.2 y 3.3 se muestran los valores resultantes del cálculo de las medidas objetivas y las medidas compuestas explicadas anteriormente para un caso de estudio elegido. (Señal 1). La señal de voz es contaminada con ruido “babble”, “car” y “train” a 0, 5, 10 y 15dB.

Tabla 3.1. Medidas objetivas y compuestas para el ruido babble (señal 1).

Métodos	PESQ	LLR*	SNRseg	WSS*	Csig	Cbak	Covl
AG y SNR=0dB	2,206504	0,980981	-1,440178	84,917156	2,6498	2,0036	2,2736
AG y SNR=5dB	2,452803	0,856148	0,793882	72,955377	3,0345	2,3458	2,6195
AG y SNR=10dB	2,775708	0,716715	3,592808	61,282830	3,4777	2,7582	3,0325
AG y SNR=15dB	3,066664	0,608594	6,193825	49,122500	3,8739	3,1462	3,4072
WP y SNR=0dB	1,988204	1,231476	-2,316876	57,947657	2,5032	2,0328	2,1584
WP y SNR=5dB	2,258812	1,080468	0,351054	48,481871	2,9069	2,3965	2,5198
WP y SNR=10dB	2,574630	0,980258	3,243411	41,505432	3,2633	2,7785	2,8741
WP y SNR=15dB	2,896530	0,759544	6,709760	34,960177	3,7434	3,1965	3,2921

*Los parámetros señalados presentan mejores resultados en la medida que tengan valores menores.

Tabla 3.2. Medidas objetivas y compuestas para el ruido Car (señal 1).

Métodos	PESQ	LLR	SNRseg	WSS	Csig	Cbak	Covl
AG y SNR=0dB	1,378115	1,133440	-0,418647	67,336181	2,1517	1,7950	1,6517
AG y SNR=5dB	1,505196	0,908130	1,522646	58,728880	2,5376	2,0383	1,9296
AG y SNR=10dB	1,658042	0,718673	3,745043	47,693473	2,9240	2,3286	2,2269
AG y SNR=15dB	1,794199	0,564962	6,028366	41,382158	3,2211	2,5817	2,4594
WP y SNR=0dB	1,131606	1,443273	-0,153891	58,046383	1,7678	1,7589	1,3597
WP y SNR=5dB	1,382419	1,184558	2,136320	48,754961	2,2689	2,0881	1,7591
WP y SNR=10dB	1,571317	0,983056	4,873533	38,949748	2,6784	2,4195	2,0829
WP y SNR=15dB	1,727601	0,762947	8,088554	30,404008	3,0760	2,7565	2,3813

Tabla 3.3. Medidas objetivas y compuestas para el ruido Train (señal 1).

Métodos	PESQ	LLR	SNRseg	WSS	Csig	Cbak	Covl
AG y SNR=0dB	1,334583	1,237348	-0,418647	67,336181	2,0991	1,8267	1,6261
AG y SNR=5dB	1,488509	0,999547	1,522646	58,728880	2,5106	2,0948	1,9293
AG y SNR=10dB	1,617470	0,833901	3,745043	47,693473	2,8202	2,3566	2,1657
AG y SNR=15dB	1,755129	0,677590	6,028366	41,382158	3,1241	2,6210	2,4033
WP y SNR=0dB	1,062760	1,937330	-0,153891	58,046383	1,2602	1,6064	1,0842
WP y SNR=5dB	1,285224	1,578495	2,136320	48,754961	1,8348	1,9494	1,5024
WP y SNR=10dB	1,487893	1,276491	4,873533	38,949748	2,3493	2,3143	1,8835
WP y SNR=15dB	1,627096	0,992136	8,088554	30,404008	2,7969	2,6757	2,1965

Al analizar los valores de las medidas contempladas en las Tablas anteriores se deduce que el algoritmo de Sustracción Espectral Geométrico (AG) y el algoritmo de reducción de ruido mediante Wavelet Packets (WP) presentan resultados muy similares en cuanto a todas las medidas calculadas, tanto a las referidas a la reducción de ruido como a la conservación de la calidad de voz. Sin embargo, se puede observar que el AG presenta ligeramente mejores valores que el WP.

Con el fin de lograr una comparación más fundamentada con respecto a estos métodos de reducción de ruido y dejar creadas las condiciones para posteriormente cotejarlas con los resultados de las medidas subjetivas, se le realizaron las pruebas anteriores a 7 señales más; y para ofrecer evidencia gráfica de los resultados obtenidos fueron promediadas cada una de las medidas objetivas y compuestas para cada tipo de ruido con los cuales se contaminaron las ocho señales analizadas. En las Figuras 3.1, 3.2, 3.3 y 3.4 se muestran los resultados referidos.

Observación: Para observar los resultados de las señales analizadas puede verse el Anexo 1.

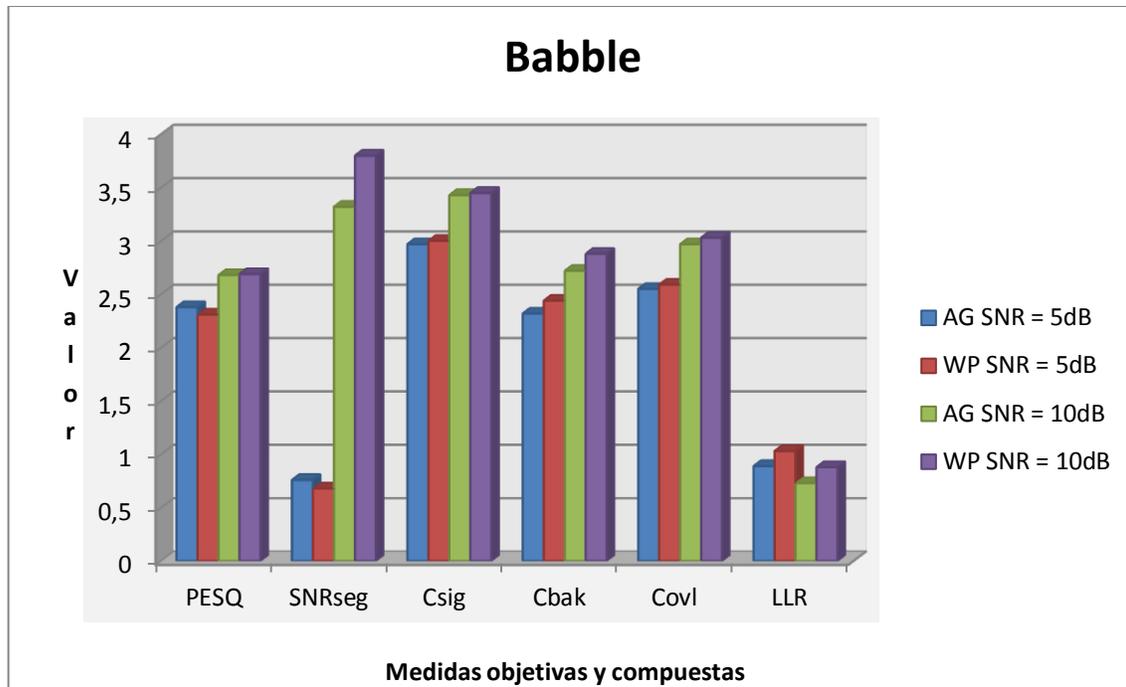


Figura 3.1. Comportamiento de los algoritmos ante ruido Babble.

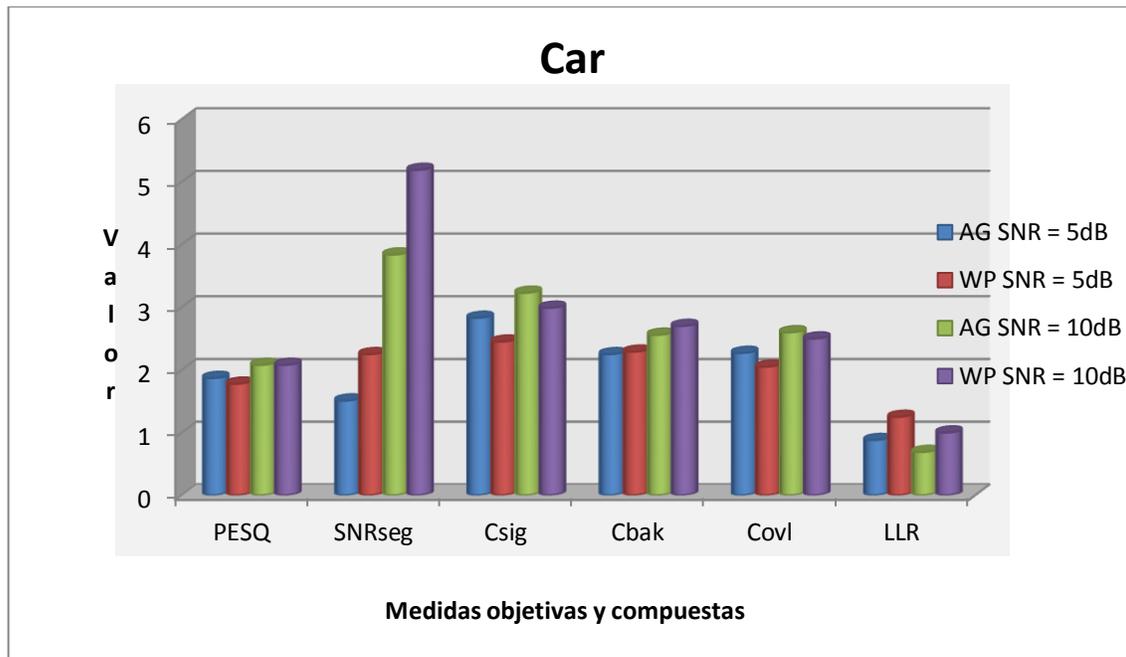


Figura 3.2. Comportamiento de los algoritmos ante ruido Car.

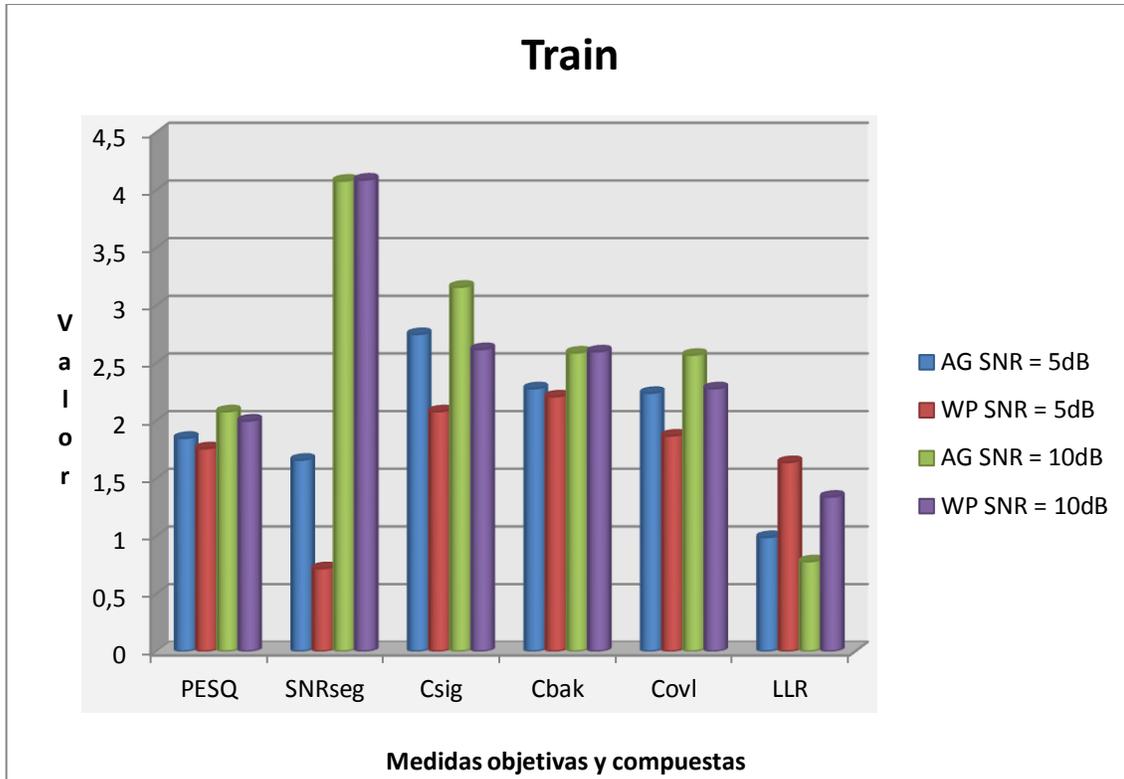


Figura 3.3: Comportamiento de los algoritmos ante ruido Train.

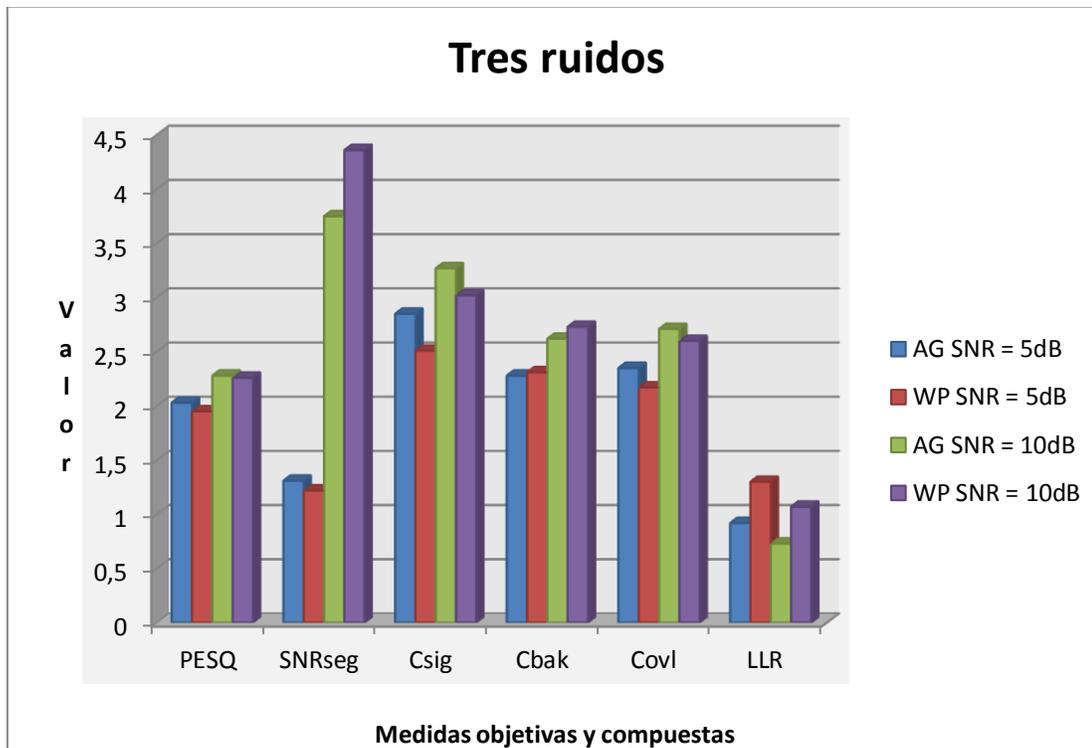


Figura 3.4. Comportamiento de los algoritmos ante los tres ruidos.

En las figuras anteriores se puede observar el comportamiento de ambos algoritmos de reducción de ruido ante los diferentes tipos de ruido y SNR. Para el ruido babble (murmullo) el algoritmo basado en WP obtuvo mejores resultados en cuanto a las tres medidas compuestas evaluadas (C_{sig} , C_{bak} y C_{ovl}) y las diferencias entre las formas de las ondas limpias y las estimadas (RSR_{seg}), para las dos SNR medidas. En cuanto a la calidad de voz (PESQ) ambos métodos se comportaron de manera similar, sin diferencias apreciables y el AG presentó mejores números en cuanto a la diferencia entre las representaciones espectrales del habla limpia y la procesada (LLR). Para el ruido car (autos), los algoritmos se comportaron de manera similar que para el ruido babble, sin embargo cabe destacar que tanto el AG como el WP empeoraron sus resultados en todas las medidas calculadas, excepto en RSR_{seg} ; por lo que ante este ruido, la calidad de la señal procesada empeoró en ambos casos. Para el ruido train (tren) el AG tuvo mejores resultados en todos los aspectos evaluados. En este ruido todos los valores se comportaron similares a los ruidos anteriores, con un empeoramiento considerable del algoritmo de WP en cuanto a LLR, donde sus valores se acercan a los más malos de esta medida.

En la Figura 3.4 podemos ver el comportamiento promedio de ambos algoritmos ante los tres tipos de ruido. El AG de forma general tuvo mejores resultados en cuanto a la razón de verosimilitud logarítmica (LLR) y en las medidas compuestas C_{sig} , C_{bak} y C_{ovl} , referentes a la distorsión de la voz, el ruido de fondo y la calidad general de la señal respectivamente, obteniendo resultados aceptables en todos sus valores. La medida RSR_{seg} para una SNR = 5dB mostró un mejor comportamiento del AG y para SNR = 10dB, el algoritmo de WP fue mejor, mostrando buenos resultados. En cuanto al PESQ, que calcula la calidad de voz, ambos métodos se comportaron parejos, con valores de 2 aproximadamente.

De acuerdo a los parámetros objetivos y compuestos analizados, el algoritmo de Sustracción Espectral Geométrico mantiene mejores resultados en la mayoría de las medidas calculadas, tanto las referidas a la reducción de ruido como a la conservación de la calidad de voz, aunque en algunos casos es imperceptible la diferencia con respecto al algoritmo basado en Wavelet Packets.

Con vistas a ampliar el criterio de evaluación de los resultados obtenidos, las Figuras 3.5 y 3.6 muestran las formas de onda de la primera señal utilizada para la comparación de los

algoritmos: señal limpia en la parte superior, la señal contaminada con ruido car y una SNR de 0dB en el centro y la señal filtrada por el método de reducción de ruido basado en WP y AG propuestos respectivamente. Se eligió una $SNR = 0dB$, porque es el caso más crítico para el que se evaluaron los algoritmos.

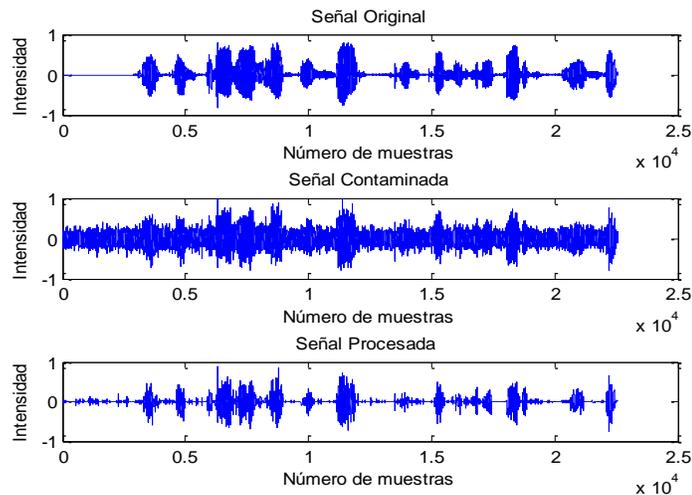


Figura 3.5. Formas de onda de la señal limpia, contaminada (Ruido Car y $SNR=0dB$) y filtrada por el algoritmo basado en WP.

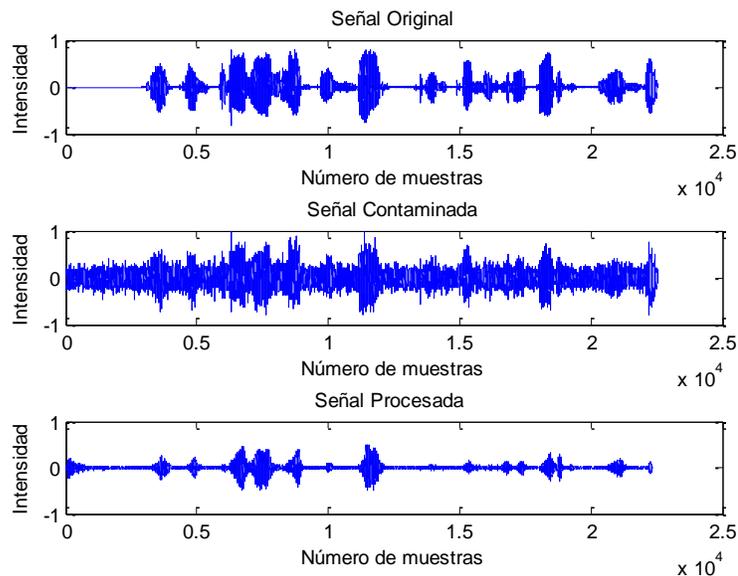


Figura 3.6. Formas de onda de la señal limpia, contaminada (Ruido Car y $SNR=0dB$) y filtrada por el algoritmo de AG.

Como se puede observar en las figuras anteriores el algoritmo basado en WP realiza una reducción del ruido de fondo, sin embargo, no lo elimina sustancialmente. Se puede ver que en los momentos de silencio de la señal ninguno de los algoritmos son capaces de eliminar totalmente el ruido presente, con un peor comportamiento para el algoritmo de WP, donde se puede apreciar una mayor intensidad del ruido en las tramas de silencio. Sin embargo, los oscilogramas de las señales no dan una medida exacta de la calidad, porque aunque difieran mucho en el dominio del tiempo, no necesariamente se podrá escuchar muy diferente la señal.

Con el objetivo de poder definir mejor las pérdidas por procesamiento de ambos métodos de reducción de ruido, las Figuras 3.7 y 3.8 muestran espectrogramas que expresan el comportamiento de estos algoritmos ante las SNR extremas ($SNR=0dB$ y $SNR=15dB$) utilizadas en el experimento con ruido car y la señal 1.

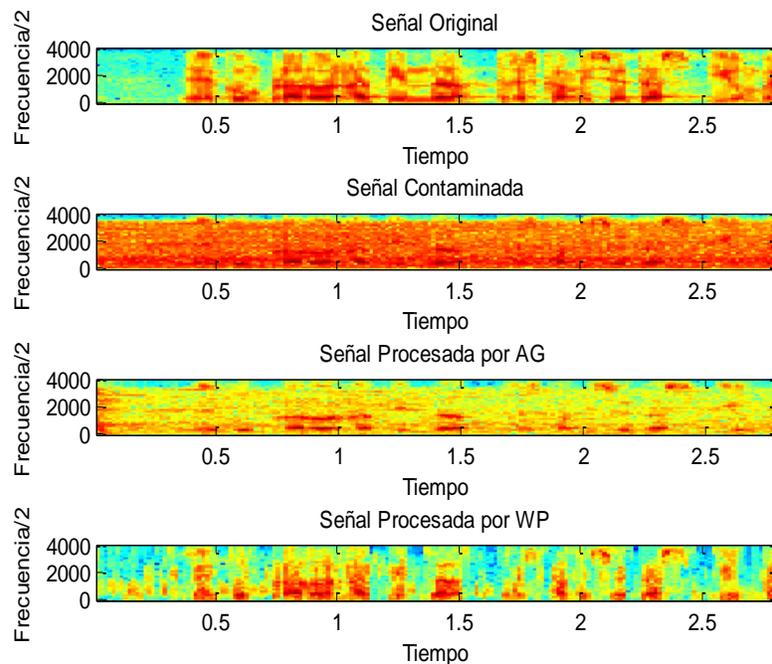


Figura 3.7. Espectrogramas de la señal limpia, señal contaminada (Ruido Car y $SNR=0dB$), filtrada por el algoritmo de AG y filtrada por el algoritmo basado en WP respectivamente.

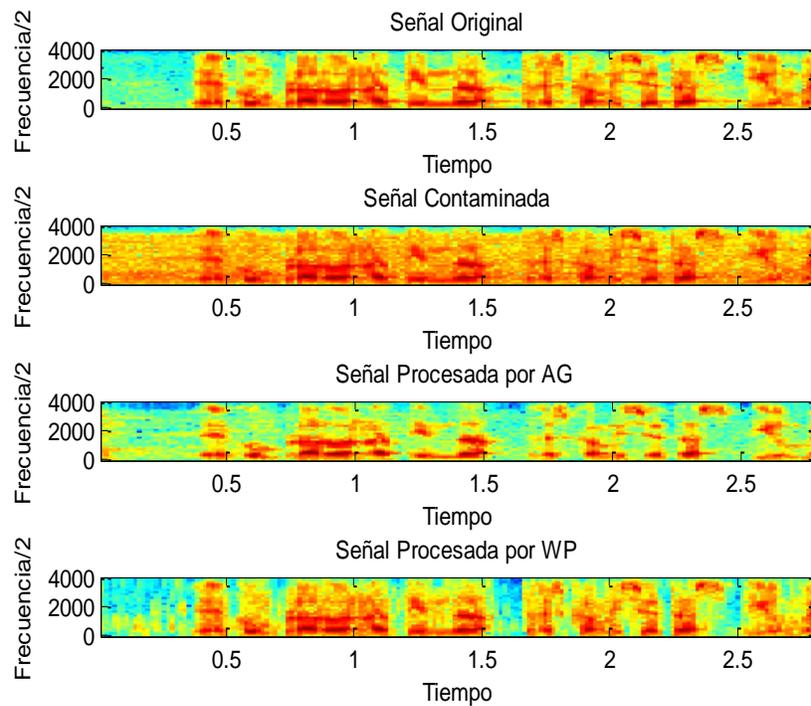


Figura 3.8. Espectrogramas de la señal limpia, señal contaminada (Ruido Car y SNR=15dB), filtrada por el algoritmo de AG y filtrada por el algoritmo basado en WP respectivamente.

En la Figura 3.7 (SNR = 0dB) se observa que los algoritmos no sólo eliminan el ruido sino además, información de voz potencialmente importante, corroborando la afectación sugerida por las medidas objetivas a la calidad e inteligibilidad de habla. Para SNR = 15dB (Figura 3.8) los espectrogramas corroboran los resultados sugeridos por las medidas empleadas, conservando gran parte de la información de la estructura de formantes presentes en la señal de habla. Sin embargo se puede observar que en ambos métodos hay eliminación de potencia de señal en segmentos con información útil, a pesar de que las medidas objetivas indican que no redundan en grandes detrimentos de la calidad de la señal procesada.

Se ratifica que desde el punto de vista perceptual, el algoritmo de reducción de ruido basado en WP no realiza una reducción sustancial del ruido de fondo, arrojando peores resultados que el algoritmo Geométrico de Sustracción Espectral.

De acuerdo a los parámetros objetivos, los oscilogramas y los espectrogramas analizados, el algoritmo de Sustracción Espectral Geométrico tiene mejores resultados que el algoritmo basado en Wavelet Packets, sin embargo para la total comprobación del desempeño de ambos se realizaron evaluaciones mediante medidas subjetivas según normas de la ITU-T P. 835 [35].

3.1.2 Desempeño de los algoritmos de compensación de pérdidas auditivas.

En este experimento se informa sobre la comparación objetiva de dos algoritmos de compensación de pérdidas auditivas: AGC basado en Wavelet Packets y AGC utilizando un Banco de Filtros Gammatone. Se utilizó una alocución en inglés a la cual se le aplicaron seis simulaciones de pérdidas auditivas reales en las frecuencias medias y altas.

Para el caso del algoritmo basado en WP se tomó como umbral de discapacidad los valores de potencia por banda de las señales con la simulación de pérdidas. Luego de la aplicación del algoritmo de compensación se calcularon los valores de potencia por banda de las señales modificadas [27].

El otro algoritmo a comparar emplea el filtro Gammatone de 17 bandas (Schneider & R.L, 1997), que posibilita la simulación de las pérdidas del audiograma a las frecuencias presentadas, y en el resto de las bandas se sitúan las pérdidas obtenidas mediante interpolación de las mismas. De esta forma se logra utilizar las 11 bandas restantes para cubrir el rango de audición y permitir, de manera general, mejor compensación de las pérdidas auditivas [9].

La primera pérdida auditiva (Señal 1) a amplificar es una presbiacusia (patología auditiva a causa de la edad) en las altas frecuencias, llegando hasta los niveles de 50dB de intensidad. En la Figura 3.9 se muestran las representaciones de las potencias por banda de la señal original, la señal con la simulación de pérdidas (pérdida en las frecuencias altas en este caso) y la señal modificada mediante ambos algoritmos del primer caso de estudio elegido.

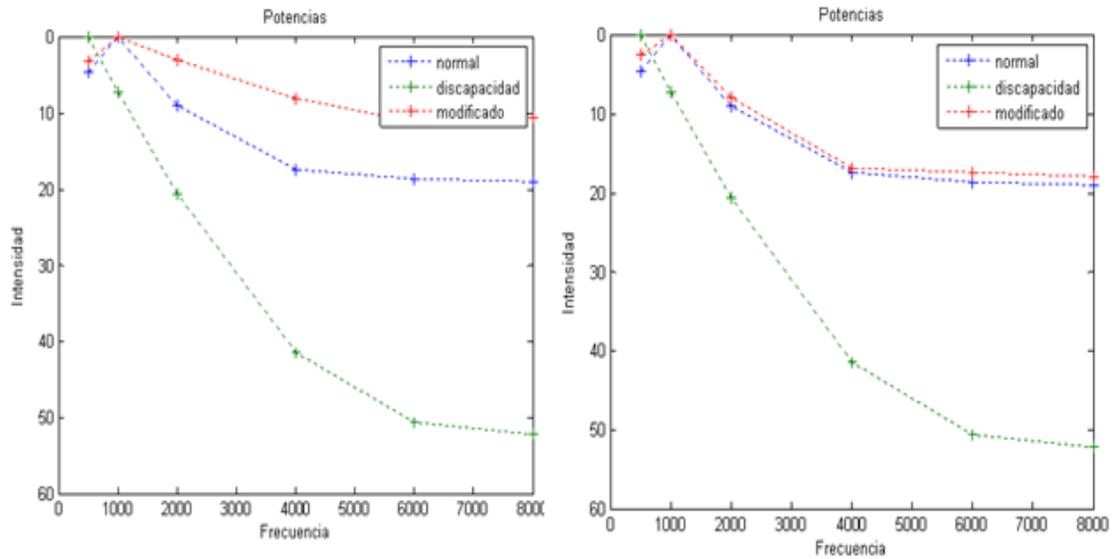


Figura 3.9. Representación de la potencia en dB SPL de la señal original (azul), señal con pérdida simulada en las frecuencias altas (verde), señal modificada por el algoritmo basado en WP (rojo, izquierda) y señal utilizando Gammatone (rojo, derecha).

El objetivo de la compensación de pérdidas auditivas reside en amplificar la potencia por bandas de frecuencia de la señal con pérdida simulada, de forma que quede lo más cercana posible a los niveles de audición normal. Partiendo de esta premisa se puede concluir que el método de AGC utilizando Gammatone realiza una amplificación casi perfecta de la pérdida auditiva, llevando la señal modificada a niveles casi iguales de la señal original. El método de AGC basado en WP también realiza una excelente compensación de pérdidas auditivas, porque aunque amplifica aproximadamente 10dB por encima de la señal original, esto prácticamente es imperceptible para el sistema auditivo humano.

Para obtener un mejor criterio de análisis, en la Figura 3.10 se presenta el espectrograma de la señal original en la parte superior, amplificada por Gammatone en el centro y basado en Wavelet Packets en la parte inferior.

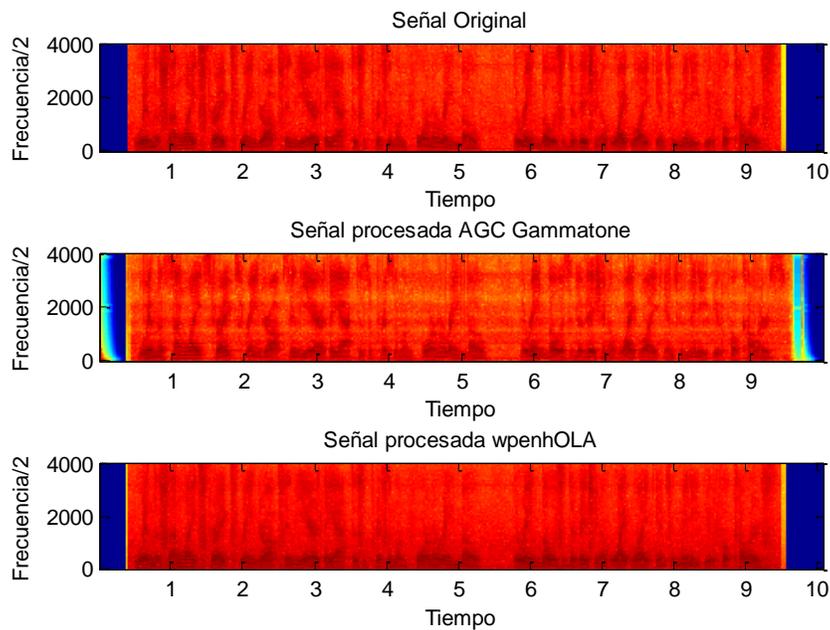


Figura 3.10. Espectrogramas de la señal original, procesada por Gammatone y procesada por Wavelet Packets.

Al igual que los resultados que arrojó la gráfica vista anteriormente (Figura 3.9) el método que utiliza Gammatone tiene superiores efectos sobre la señal con pérdida simulada que la basada en WP, ya que como se puede observar, logra una mejor amplificación en todas las bandas de la señal.

De igual forma a la señal analizada precedentemente se comparó otra señal (Señal 2) con pérdidas inducidas en las frecuencias medias y altas. En la Figura 3.11 se realiza una representación de las potencias por banda de la señal original, la señal con la simulación de pérdidas y la señal modificada mediante ambos algoritmos.

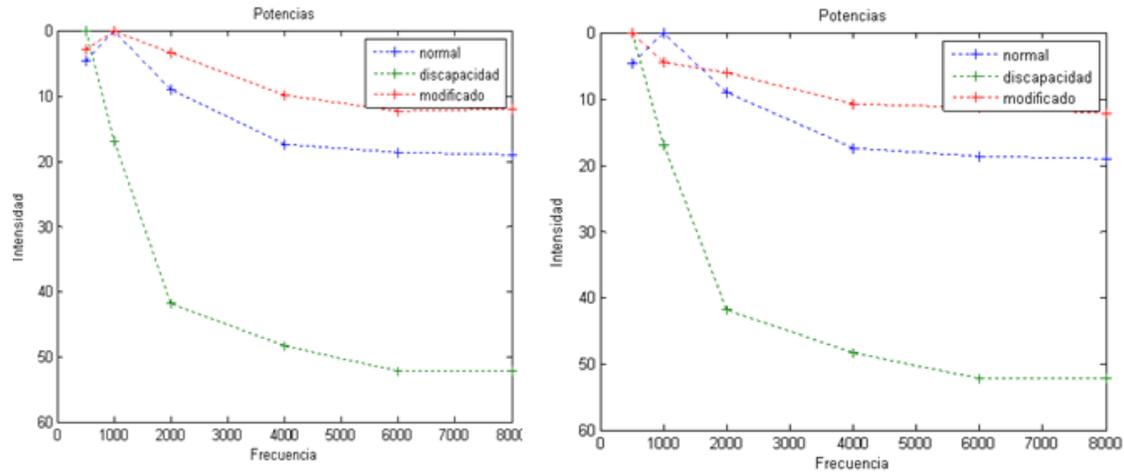


Figura 3.11. Representación de la potencia en dB SPL de la señal original (azul), señal con pérdida simulada en las frecuencias altas (verde), señal modificada por el algoritmo basado en WP (rojo, izquierda) y señal utilizando Gammatone (rojo, derecha).

En este caso ambos algoritmos realizan una buena amplificación, tanto en las medias como en las altas frecuencias. De forma general hay un aumento de la intensidad de la señal modificada con respecto a la normal, sin embargo esto es casi imperceptible para el oído humano, y en ambos casos se logra corregir la pérdida satisfactoriamente.

El tercer grupo de señales (Señal 3) se muestra a continuación, y se realizó el mismo análisis comparativo. En este caso la deficiencia auditiva es una hipoacusia sensorineural leve. Al igual que en las señales anteriores se realiza una representación de las potencias por banda de la señal original, la señal con la simulación de pérdidas y la señal modificada mediante ambos algoritmos en la Figura 3.12.

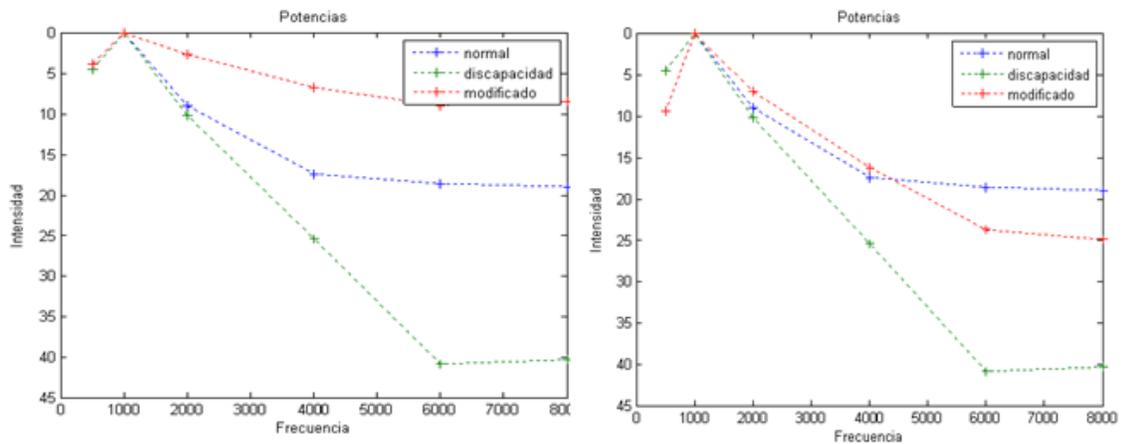


Figura 3.12. Representación de la potencia en dB SPL de la señal original (azul), señal con pérdida simulada en las frecuencias altas (verde), señal modificada por el algoritmo basado en WP (rojo, izquierda) y señal utilizando Gammatone (rojo, derecha).

Como se puede ver en la figura anterior, el AGC que utiliza un banco de Filtros Gammatone no llega en las altas frecuencias a la amplificación deseada, quedándose por debajo de la intensidad de la señal original. El AGC basado en WP mantiene buenos resultados de amplificación en todas las bandas de frecuencia.

De forma general, basado en las 6 pérdidas analizadas (Véase Anexo2), se concluye que el AGC basado en WP tuvo un mejor comportamiento que el AGC que utiliza un Banco de Filtros Gammatone, ya que tuvo mejores resultados de amplificación en todas las pérdidas inducidas, tanto en las medias como en las altas frecuencias. Sus niveles de intensidad estaban en los límites permisibles, ya que aunque en algunas bandas la potencia está hasta 10dB por encima del nivel esperado, esto no constituye una deficiencia significativa del algoritmo. Un aumento de 10dB es prácticamente imperceptible para el sistema auditivo humano.

El AGC que utiliza Gammatone también tuvo buenos resultados de compensación, sin embargo, en algunas pérdidas en las altas frecuencias, no llegó a los niveles deseados de amplificación.

3.2 Comparación subjetiva de las prótesis auditivas digitales.

3.2.1 Desempeño de los algoritmos de reducción de ruido.

Se condujo un experimento como el descrito en el sub-epígrafe 2.4.1 siguiendo los requisitos del estándar UIT-T (P.835). En este experimento se muestran los resultados del desempeño de los algoritmos de reducción de ruido basados en Wavelet Packets (WP) y el método de Sustracción Espectral Geométrico (AG), frente a ruidos típicos y simulados, empleando medidas subjetivas. Las señales fueron contaminadas artificialmente con dos tipos de ruido: murmullo (babble) y ruido presente en autos (car). Se tomó en consideración dos niveles de SNR: 5 y 10dB.

En las figuras 3.13, 3.14 y 3.15 se muestran las puntuaciones medias de las escalas de SIG, BAK y OVRL derivadas de los dos algoritmos de reducción de ruido evaluados, con el objetivo de comparar el desempeño de los mismos ante ruidos y SNR diferentes.

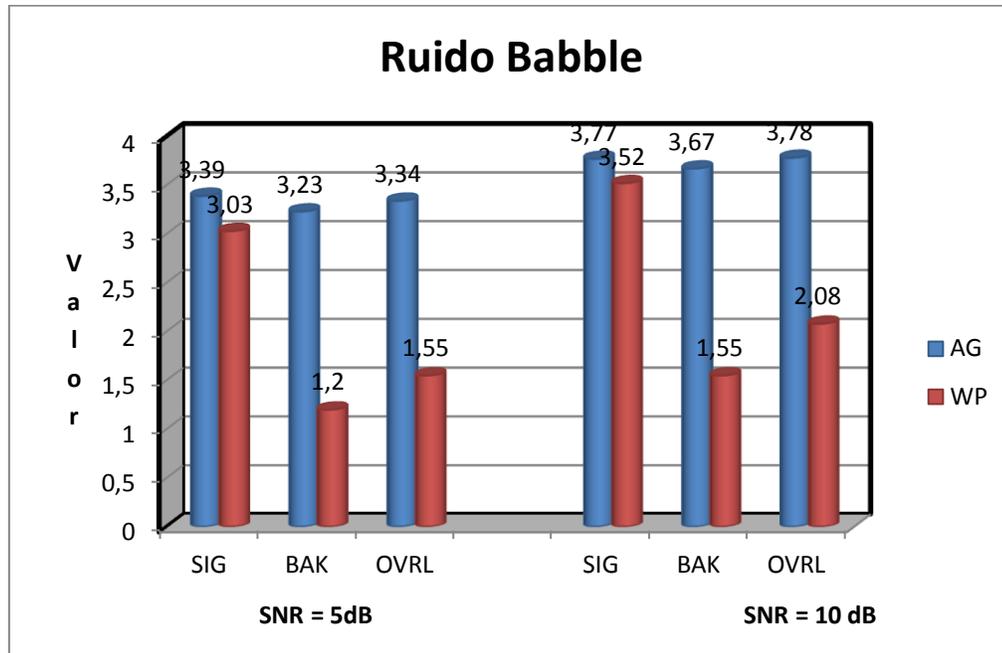


Figura 3.13. Puntuaciones medias de las medidas evaluadas para ruido babble.

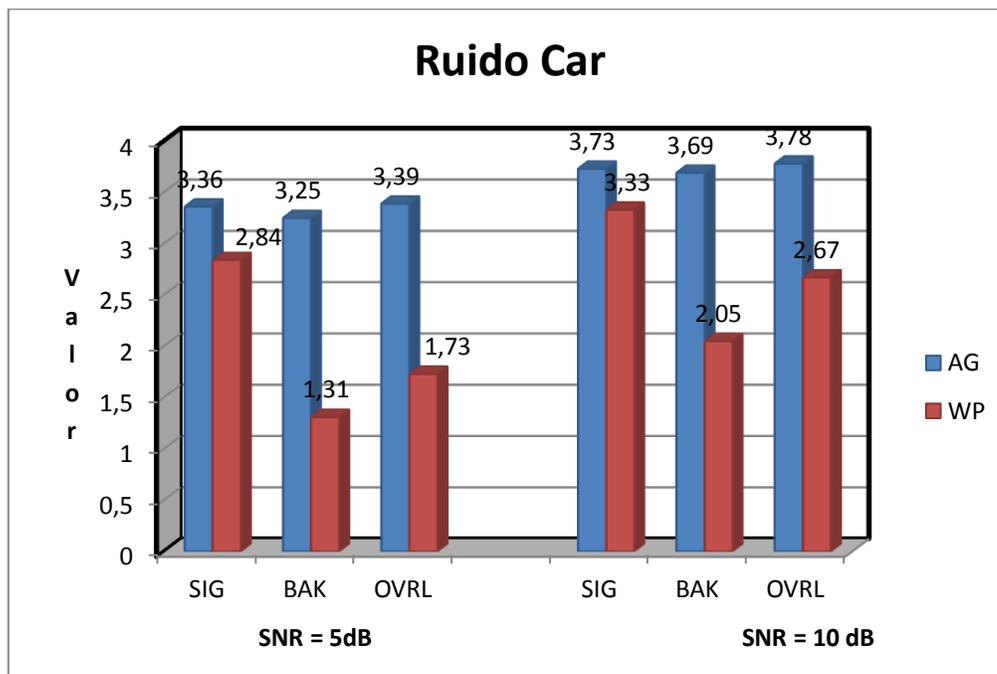


Figura 3.14. Puntuaciones medias de las medidas evaluadas para ruido car.

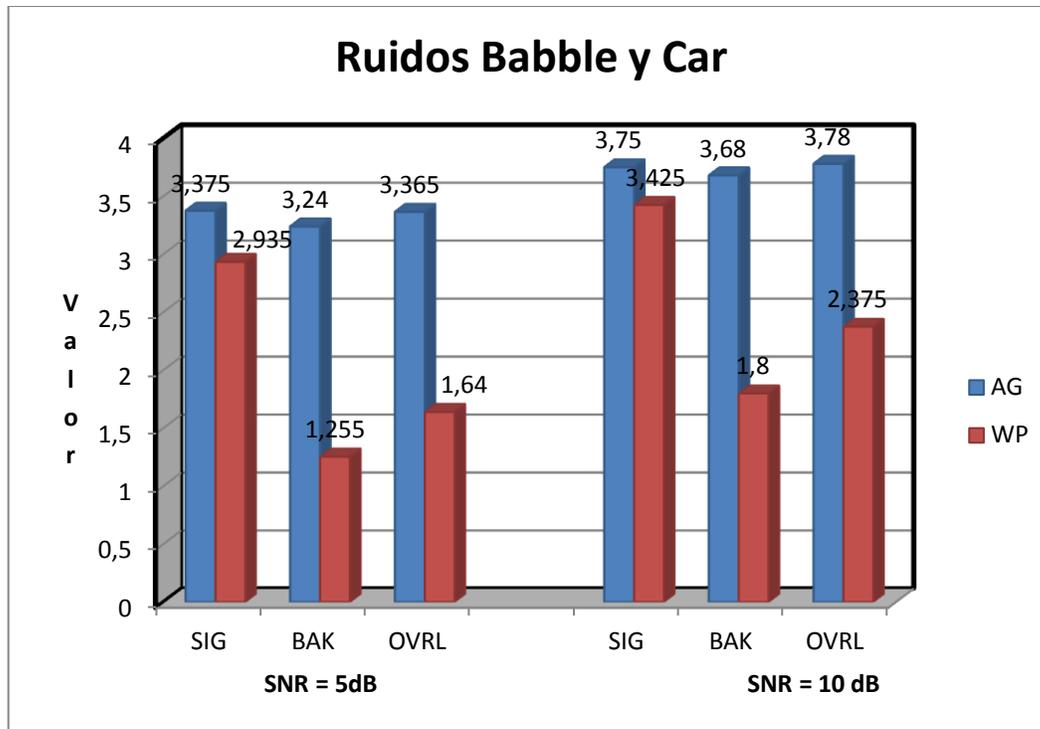


Figura 3.15: Puntuaciones medias de las medidas evaluadas para ambos ruidos (babble y car).

Como se puede apreciar en las figuras anteriores el Algoritmo de Sustracción Espectral Geométrico se comportó mucho mejor que el Algoritmo basado en Wavelet Packets. En dos de las medidas evaluadas, las referentes a la distorsión del ruido de fondo (BAK) y la calidad general de la señal (OVRL), el AG obtuvo una significativa diferencia de calidad con respecto al algoritmo de WP para todos los tipos de ruido y SNR. En cuanto a la distorsión de la voz (SIG), ambos métodos de reducción de ruido se comportaron similarmente, sin embargo el AG también obtuvo puntuaciones más altas.

Las bajas calificaciones del Algoritmo basado en WP con respecto a la medida BAK y OVRL están en correspondencia con los resultados de los oscilogramas y espectrogramas analizados en el epígrafe anterior, en los cuales se podía apreciar que este algoritmo no hacía una reducción sustancial del ruido de fondo. Esto está en contraposición con las medidas objetivas y compuestas calculadas, donde se le otorgan mejores puntuaciones a este algoritmo, que las reflejadas por las medidas subjetivas.

De gran interés es descubrir la contribución individual de la distorsión de la voz y del ruido de fondo en el juicio de los oyentes para determinar la calidad en general de la señal. Nuestros datos anteriores (Figuras 3.13, 3.14 y 3.15) nos llevaron a creer que los oyentes fueron más afectados por la distorsión del ruido de fondo a la hora de emitir sus criterios de la calidad general de la señal. A fin de fundamentar esto, se realizó un análisis de regresión lineal múltiple con las calificaciones obtenidas por las tres medidas subjetivas evaluadas. Tratamos la puntuación global de calidad (OVRL) como la variable dependiente y las puntuaciones del habla (SIG) y el ruido de fondo (BAK) como variables independientes. El análisis de regresión reveló la siguiente relación entre las tres escalas de calificación:

$$Rovl = 0,2887 + 0,2446 \times Rsig + 0,6726 \times Rbak \quad (27)$$

Donde $Rovl$ es la predicción general del promedio de calificación del OVRL, $Rsig$ el promedio de SIG y $Rbak$ el promedio de BAK. El coeficiente de determinación resultante es $r^2=0.6814$, por lo que las predicciones de las variables anteriores son aceptables (0=mala, 1=excelente). La ecuación anterior confirma que los oyentes en realidad prestaron atención tanto a la calidad de la voz como a los efectos del ruido al presentar sus valoraciones sobre la calidad en general de la señal; sin embargo se ratifica que los oyentes parecen poner más énfasis en el ruido de fondo que en la distorsión del habla cuando emiten sus juicios de la calidad en general.

Con el objetivo de observar la dispersión de los valores respecto a la media (valor promedio) y la medida de la asimetría de los datos alrededor de la misma se muestra la Tabla 3.4.

Tabla 3.4: Valor medio, desviación estándar y asimetría de las muestras.

Condiciones	Valor medio			Desviación Estándar			Asimetría		
	SIG	BAK	OVRL	SIG	BAK	OVRL	SIG	BAK	OVRL
AG y SNR=5dB babble	3,391	3,234	3,344	0,581	0,611	0,597	0,188	-0,164	0,622
AG y SNR=10dB babble	3,766	3,672	3,781	0,496	0,473	0,453	-0,432	-0,732	-0,832

WP y SNR=5dB babble	3,031	1,203	1,547	0,712	0,406	0,5021	-0,044	1,476	-0,188
WP y SNR=10dB babble	3,516	1,547	2,078	0,591	0,502	0,543	-0,759	-0,188	0,061
AG y SNR=5dB car	3,359	3,25	3,391	0,484	0,471	0,492	0,586	0,687	0,448
AG y SNR=10dB car	3,734	3,688	3,781	0,512	0,467	0,417	-0,319	-0,809	-1,361
WP y SNR=5dB car	2,844	1,31	1,734	0,569	0,467	0,623	-0,015	0,809	0,245
WP y SNR=10dB car	3,328	2,047	2,672	0,619	0,575	0,644	-0,338	0,005	-0,662
Total	3.371	2.494	2.791	0.571	0.497	0.534	-0.142	0.039	-0.273

En la tabla anterior se aprecia que la desviación estándar de las tres medidas subjetivas evaluadas se comportan de forma excelente, ya que no existe una dispersión apreciable de los valores emitidos por los oyentes con respecto a la media; teniendo en cuenta que el rango de calificaciones se encuentra del 1 al 5. La medida de la asimetría de los datos alrededor del valor medio de las muestras presenta un satisfactorio resultado, pues la distribución normal perfecta es 0 y el promedio total de las tres medidas calculadas están muy cercanas a este valor. Se debe destacar que las medidas SIG y OVRL tienden más a la izquierda del valor medio, y BAK a la derecha, sin embargo no son diferencias significativas. De forma general se puede apreciar que existieron semejanzas entre las puntuaciones obtenidas en el experimento subjetivo.

3.2.2 Desempeño de los algoritmos de compensación de pérdidas auditivas.

En este experimento se informa sobre la comparación subjetiva de dos algoritmos de compensación de pérdidas auditivas: AGC basado en Wavelet Packets y AGC utilizando un Banco de Filtros Gammatone (Véase sub-epígrafe 2.4.2.). Los oyentes emitieron sus criterios basando sus juicios en la calidad global de la señal amplificada mediante criterios

MOS (Mean Opinion Score) (OVRL) [1=mala, 2=pobre, 3=aceptable, 4=buena, 5=excelente].

En la Tabla 3.5 se muestran los valores medios, la medida de la asimetría de los datos alrededor de los mismos y la desviación estándar para las calificaciones emitidas por los encuestados.

Tabla 3.5: Valor medio, desviación estándar y asimetría de las muestras.

Condiciones	Valor medio	Desviación Estándar	Asimetría
Señal 1 AGCWP	4,375	0,49186938	0,5164
Señal 1 AGCGamma	4,875	0,33601075	-2,2678
Señal 2 AGCWP	3,5625	0,50401613	-0,2520
Señal 2 AGCGamma	3,46875	0,50700735	0,1252
Señal 3 AGCWP	4,6875	0,53506105	-1,4342
Señal 3 AGCGamma	3,84375	0,51489899	-0,2462
Señal 4 AGCWP	4,5	0,50800051	0
Señal 4 AGCGamma	3,625	0,49186938	-0,5164
Señal 5 AGCWP	4,28125	0,45680341	0,9731
Señal 5 AGCGamma	3,71875	0,52267149	-0,2623
Señal 6 AGC WP	4,4375	0,50401613	0,2520
Señal 6 AGC Gamma	3,15625	0,44788932	0,7160
Total	4.04427083	0.4850095	-0.1996883

Como se puede observar en la Tabla 3.5 la desviación estándar de la medida subjetiva evaluada obtiene valores aceptables de dispersión, por lo que no existieron demasiadas

diferencias entre la media y la totalidad de puntuaciones analizadas. La medida de la asimetría de los datos alrededor del valor medio de las muestras presenta buenos resultados, en el orden de -0.2 (tiende más a la izquierda de la media), pudiéndose ver que el método que utiliza Bancos de Filtros Gammatone (AGC Gamma) presentó menor simetría y promedio de puntuaciones (media) que el método basado en Wavelet Packets.

Para poder comparar las puntuaciones otorgadas a la calidad de las señales procesadas por el AGC Gamma y el AGC WP de una forma más visible se muestra la Figura 3.16.

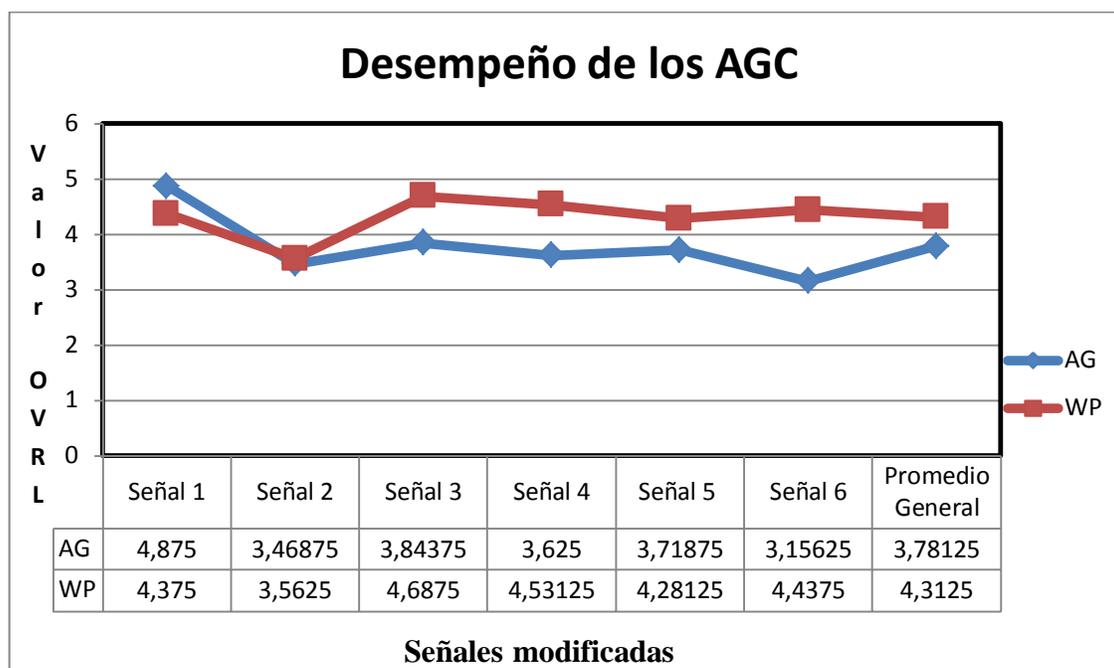


Figura 3.16: Desempeño de los AGC ante las seis señales de acuerdo al valor de las calificaciones obtenidas.

En correspondencia con los resultados obtenidos en las pruebas objetivas, el AGC basado en Wavelet Packets (WP) tuvo un mejor desempeño que el AGC utilizando Banco de Filtros Gammatone, observándose claramente en las puntuaciones obtenidas sobre la calidad general de las señales modificadas. De acuerdo a las calificaciones emitidas por los oyentes basados en los criterios MOS (OVRL), los valores de ambos algoritmos logran amplificar satisfactoriamente las señales en todas las frecuencias que se analizaron. El AGC basado en WP tuvo un valor medio superior a 4,4, lo que en la escala MOS se interpreta como una señal amplificada en el rango de calidad: buena – excelente.

3.3 Análisis comparativo de las medidas de evaluación utilizadas.

Este análisis comparativo de las medidas utilizadas en la evaluación se realiza condicionado por la diversidad de criterios en la bibliografía consultada con respecto a las correlaciones y las desviaciones estándar del error entre las medidas analíticas (prescindiendo de la opinión de los oyentes) y las pruebas subjetivas. Numerosos métodos de mediciones objetivas han sido propuestos por los investigadores, con sus ventajas y limitaciones. La abundancia de medidas objetivas ha conducido al mismo tiempo al uso arbitrario de una diversidad de ellas.

Para realizar esta comparación se calculan dos figuras de mérito para cada medida objetiva y compuesta. La primera es el coeficiente de correlación (correlación de Pearson) entre las medidas de calidad subjetivas (Sd) y las objetivas (Od) [23].

$$\rho = \frac{\sum_d (Sd - \overline{Sd})(Od - \overline{Od})}{[\sum_d (Sd - \overline{Sd})^2]^{1/2} [\sum_d (Od - \overline{Od})^2]^{1/2}} \quad (28)$$

Donde \overline{Sd} y \overline{Od} son los valores medios de Sd y Od respectivamente.

La segunda figura de mérito es una estimación de la desviación estándar del error cuando la medida objetiva se utiliza en lugar de la medida subjetiva, y está dada por:

$$\hat{\sigma}_e = \hat{\sigma}_s \sqrt{1 - \rho^2} \quad (29)$$

Donde $\hat{\sigma}_s$ es la desviación estándar de Sd y $\hat{\sigma}_e$ es la desviación estándar calculada del error. Un valor menor de $\hat{\sigma}_e$ indica que la medida objetiva es mejor para predecir la calidad subjetiva [23].

En la Tabla 3.6 se muestra el coeficiente de correlación (ρ) de cada una de las medidas objetivas y compuestas con las tres medidas subjetivas evaluadas y en la Tabla 3.7 la desviación estándar del error para cada correlación de las medidas objetivas y compuestas.

Tabla 3.6: Coeficiente de correlación estimado.

	PESQ	LLR	SNRseg	WSS	Csig	Cbak	Covl
SIG	0,4	-0,707	0,268	0,073	0,575		
BAK	0,061	-0,695	0,026	0,532		-0,101	
OVRL	0,087	-0,69	0,154	0,45			0,138

Tabla 3.7: Desviación estándar del error para la correlación de las medidas objetivas y compuestas.

	PESQ	LLR	SNRseg	WSS	Csig	Cbak	Covl
SIG	0,348	0,268	0,366	0,379	0,311		
BAK	1,038	0,753	1,04	0,881		1,034	
OVRL	0,883	0,641	0,875	0,791			0,878

En las Tablas 3.6 y 3.7 se puede observar que de todas las medidas objetivas y compuestas convencionales la medida Csig fue la que mejores resultados obtuvo en términos de predicción de la calidad de voz (SIG), seguida de PESQ, SNRseg, WSS y LLR. En términos de distorsión del ruido de fondo (BAK), la medida WSS resultó superior, y a continuación PESQ, SNRseg, Cbak y LLR. En términos de la predicción de la calidad global de la señal la mejor fue la medida WSS, seguida de SNRseg, Covl, PESQ y LLR.

De manera general se ve claramente que la mayoría de las medidas objetivas incluidas, no alcanzan una buena correlación con las valoraciones de los oyentes. En consecuencia, no resulta apropiado evaluar el desempeño de las distintas técnicas de reducción de ruido con un solo método: objetivo o subjetivo. La divergencia entre las mediciones subjetivas y objetivas de calidad, sugieren la exploración de nuevas técnicas que contemplen mejor los aspectos perceptuales del proceso de audición. A la vez, la falta de correlación con medidas tan utilizadas como LLR y PESQ enfatiza en la necesidad de establecer protocolos universales y realistas de evaluación, a fin de escoger técnicas de referencia para nuevos

desarrollos. Asimismo, es necesario considerarlos posibles efectos de la variabilidad del juicio de los oyentes, y el número y experiencia de éstos en pruebas de este tipo.

Todos estos aspectos, sumado a que en general no se ha utilizado un conjunto común de señales para la evaluación, ha dificultado la existencia de una base de referencia verdaderamente comparativa del desempeño de las distintas técnicas de reducción de ruido disponibles. Por otra parte, la eficacia de los sistemas de procesamiento de voz depende de las características del lenguaje en el cual se aplican y la mayoría de las técnicas propuestas son evaluadas en el marco del idioma Inglés, extendiéndose muy pocas veces esta evaluación a múltiples idiomas. Ejemplo de esto, es que no se dispone actualmente de una base de referencia, construida sobre el idioma Español, que permita evaluar algoritmos de reducción de ruido. Muchas medidas objetivas de evaluación encontradas en la bibliografía no están validadas para distintos idiomas, y generalmente sus resultados sólo han sido correlacionados con las calificaciones de sujetos normo-oyentes.

Son necesarios mayores estudios para establecer cuáles medidas objetivas y en qué grado, se ajustan a la aceptación de los oyentes. Sin embargo, estas pruebas iniciales nos permiten comprender la complejidad de los procedimientos de evaluación del desempeño de algoritmos de reducción de ruido en el habla, y representan un primer paso hacia la elaboración de un protocolo confiable de evaluación destinado a pacientes hipoacúsicos.

El experimento subjetivo de compensación de pérdidas auditivas se comportó de acuerdo a las informaciones que nos mostraron los gráficos de potencias y espectrogramas de las señales, por lo que se puede evaluar de excelente la correspondencia entre ellos.

3.4 Conclusiones parciales del capítulo.

En este capítulo se analizó el desempeño de los algoritmos descritos en el capítulo 2. Los resultados obtenidos muestran que el algoritmo de Sustracción Espectral Geométrico es la técnica de reducción de ruido con mejor desempeño entre las evaluadas. La superioridad de este algoritmo se refleja en los valores de las medidas objetivas y compuestas calculadas y en la mayor aceptación mostrada por los oyentes en la totalidad de los criterios y condiciones consideradas. El algoritmo basado en Wavelet Packets a pesar de haber presentado valores aceptables de las medidas objetivas y compuestas calculadas, no hace

una suficiente sustracción del ruido de fondo desde el punto de vista perceptual, lo que se demostró en los juicios emitidos por los oyentes.

En el caso de los algoritmos para la compensación de pérdidas auditivas personalizadas, los dos algoritmos comparados mostraron un buen desempeño amplificando las pérdidas simuladas que se le aplicaron. Sin embargo, el AGC basado en Wavelet Packets, tuvo mejores resultados en todas las bandas de frecuencias analizadas, obteniendo valores muy cercanos a la señal original.

CONCLUSIONES Y RECOMENDACIONES

Conclusiones

- Mediante análisis objetivos y subjetivos de las técnicas de reducción de ruido seleccionadas se determinó que el algoritmo de Sustracción Espectral Geométrico presenta una gran aplicabilidad en el tratamiento de señales con diferentes ruidos y SNR, mientras que el algoritmo basado en Wavelet Packets no realiza una reducción sustancial del ruido de fondo.
- Se demostró que los AGC utilizando Filtros Gammatone y los basados en la transformada Wavelet Packets, realizan una buena amplificación de las pérdidas auditivas en las medias y altas frecuencias, obteniendo este último algoritmo resultados excelentes.
- Se demostró que la mayoría de las medidas objetivas que evalúan la calidad en el proceso de reducción de ruido, no son adecuadas para predecir los criterios subjetivos emitidos por los oyentes con respecto a la de calidad de voz, distorsión del ruido de fondo y calidad global de la señal.
- Mediante análisis estadísticos se demostró que los oyentes se vieron más influidos en la distorsión del ruido de fondo que en la calidad de la voz al emitir sus criterios de la calidad general de la señal.
- Se concluyó que no existe un protocolo confiable de evaluación de métodos destinados a pacientes hablantes hispanos con deficiencias auditivas, por lo que este trabajo representa un primer paso para su elaboración.

Recomendaciones

- Utilizar el algoritmo de compensación de pérdidas auditivas basado en la Transformada Wavelet Packets en el procesamiento digital de señales de una prótesis auditiva digital real.
- Utilizar el algoritmo de Sustracción Espectral Geométrica para la reducción de ruido en una prótesis auditiva digital real.
- Implementar una base de referencia construida sobre el idioma español que permita evaluar algoritmos de reducción de ruido.
- Validar los algoritmos analizados con pacientes con deficiencias auditivas.

REFERENCIAS BIBLIOGRÁFICAS

- [1] Efren Ruiz, "Contaminación acústica: efectos sobre parámetros físicos y fisiológicos," Universidad de La Laguna, Tesis grado Dr C..
- [2] Edwards, B., 2007. Beyond Amplification: Signal Processing Techniques for Improving Speech Intelligibility in Noise with Hearing Aids. In *Seminars In Hearing*, 2007.
- [3] MarkeTrakV. & S., K., 2005. Customer satisfaction with hearing instruments in the digital age. *Hear J.*, 58, pp.30-42.
- [4] Kuo, Y.-T., Lin, T.-J., Li, Y.-T. & Liu, C.-W., 2010. Design and Implementation of Low-Power ANSI S1.11 Filter Bank for Digital Hearing Aids. *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS*.
- [5] Nikoleta, G., 2009. Compression Techniques For Digital Hearing Aids. Ms.Tesis.
- [6] Preves, D. & Banerjee, S., 2008. Hearing aid instrumentation signal processing and electroacoustic testing. Thieme Medical. *Audiology treatment*, pp.1-35.
- [7] Loizou, P., 2005. Speech enhancement based on perceptually motivated Bayesian estimators of the speech magnitude spectrum. *IEEE Trans.Speech Audio Process*, 13(5), pp.857-69.
- [8] Trinidad, G., 2003. [Online] Available at: HYPERLINK "<http://www.fineprint.com>" <http://www.fineprint.com> [Accessed 18 November 2009].
- [9] Jiménez Yera, Roberto, 2010. Desarrollo de una prótesis auditiva digital. Tesis de pregrado. Universidad Central de Las Villas.
- [10] T.Lunner & J.Hellgren, 1991. A digital filterbank hearing aid—Design, implementation and evaluation. *Proc. ICASSP Conf*, pp.3661-64.

- [11] Y.Lian & Y.Wei, 2005. A computationally efficient nonuniform FIR digital filter bank for hearing aids. *IEEE Tran.Circuits Syst*, 52(12), pp.2754-62.
- [12] K.S.Chong, Gwee, B.H. & J.S.Chang, 2006. A 16-channel low-power nonuniform spaced filter bank core for digital hearing aid. *IEEE Tran.Circuits Syst*, 53(9), pp.853-57.
- [13] Byrne, D. et al., 2001. NAL-NL1 procedure for fitting nonlinear hearing aids: Characteristics and comparisons with other procedures. *J.Amer.Acad. of Audiology*, 12(1), pp.37-54.
- [14] Standard ANSI, S1.11-2004. Specification for Octave-Band and Fractional-Octave-Band Analog and Digital Filters.
- [15] Lozano, A. & A.Carlosena, 2003. DSP-based implementation of an ANSI S1.11 acoustic analyzer. *IEEE Trans. Instrum. Meas*, 52(4), pp.1213-19.
- [16] Bogaert, T.V.d., Wouters, J., Klasen, T.J. & Moone, M., 2005. Distortion of interaural time cues by directional noise reduction system in modern digital hearing aids. *Proc. IEEE Workshop on Appl. Of Signal Process. to Audio and Acoust.*
- [17] Heinrich Löllman and Peter Vary, "Low Delay Noise Reduction and Deverberetion for Hearing Aids," *EURASIP Journal of Advances in signal Processing*, April 2009.
- [18] Patterson, R.D. & Holdsworth, J., 1993. A FUNCTIONAL MODEL OF NEURAL ACTIVITY PATTERNS AND AUDITORY IMAGES. *Advances in Speech, Hearing and Language Processing*, 3(Parte B), pp.547-63.
- [19] C.Loizou, P. & Dorman, M., 1999. On the number of channel needed to understanding speech. Ph.D. Universidad de Dallas,Texas.
- [20] D.Kamath, S. & C.Loizou, P., 2002. A MULTI-BAND SPECTRAL SUBTRACTION METHOD FOR ENHANCING SPEECH CORRUPTED BY COLORED NOISE.
- [21] Goyé, A., Wable, J., Guilmin, G. & Frachet, B., 2000. Réduction de bruit et traitements paramétriques de la parole en large bande destinés à améliorer la compréhension des sujets malentendants.
- [22] Martin, R., 2001. Noise power spectral density estimation based on optimal smoothing and minimum statistics. *IEEE Trans. Speech Audio Process*, 9(5), p.504–512.

- [23] Hu, Y. & Loizou, P.C., 2008. Evaluation of Objective Measures for Speech Enhancement.
- [24] Levitt, H., 2007. Historically, the paths of hearing aids and telephones have often intertwined. *The Hearing Journal*, 60(11), pp.20-24.
- [25] Kerckhoff, J., Listenberger, J. & Valente, M., 2008. Advances in Hearing Aid Technology. *Contemporary Issues in Communication Science and Disorders* , 35, pp.102-12.
- [26] Launer, S., 2008. The core of future hearing instrument innovation. *Advance for Audiologists*, 60(11), pp.20-24.
- [27] Peón Pérez, Beatriz, 2010. Algoritmos para prótesis auditivas utilizando la Transformada Wavelet Packets. Tesis de Pregrado. Universidad Central de Las Villas.
- [28] Loizou, P., 2005. Speech enhancement based on perceptually motivated Bayesian estimators of the speech magnitude spectrum. *IEEE Trans.Speech Audio Process*, 13(5), pp.857-69.
- [29] Carney, L.H. & Yin, C.T., 1988. Temporal coding of resonances by low-frequency auditory nerve fibers: Single fibre responses and a population model. *J.Neurophysiology*, 60, pp.1653-77.
- [30] Joseph Timoney, Thomas Lyshaght, and Marc Shoenweisner, "Implementing loudness model in Matlab," *Proc of the 7th Int. Conference on Digital Audio Effects DAFx'04*, pp. 177-180, October 2004.
- [31] Trenas, M.A., Rutledge, J.C. & Whitmal, N.A., 1999. Wavelet-Based Speech Enhancement for Hearing Aids. Vienna, Austria: European Medical & Biological Engineering Conf. (EMBEC'99) University of Malaga
- [32] Y. Ghambari, S. Y. Mortazavi, and M. R. Karami, "A New Speech Enhancement System Based on the Adaptive Thresholding of Wavelet Packets," *13th ICEE2005*, vol. 1, May 2005.

- [33] H. Hirsch and D. Pearce, "The aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions.," ISCA ITRW ASR2000, September 2000.
- [34] Scott Bradley. (2007, Dec.) HEARING LOSS SAMPLER. [Online]. <http://facstaff.uww.edu/bradleys/radio/hlsimulation/>
- [35] Yi Hu and Philipos C. Loizou, "Subjective comparison and evaluation of speech enhancement algorithms," *Speech Communication*, vol. 49, pp. 588-601, 2007.

ANEXOS

Anexo I Tablas de los cálculos de los parámetros objetivos de reducción de ruido.

Tabla 1. Medidas objetivas y compuestas para el ruido babble (señal 2).

Métodos	PESQ	LLR	SNRseg	WSS	Csig	Cbak	Covl
AG y SNR=5dB	2,183479	0,981122	1,334311	74,338869	2,7310	2,2414	2,3290
AG y SNR=10dB	2,607961	0,807231	4,001381	57,849652	3,3143	2,7277	2,8752
WP y SNR=5dB	2,178820	1,049366	1,251991	51,405305	2,8644	2,3945	2,4508
WP y SNR=10dB	2,575473	0,931812	4,360200	42,612837	3,3037	2,8415	2,8919

Tabla 2. Medidas objetivas y compuestas para el ruido Car (señal 2).

Métodos	PESQ	LLR	SNRseg	WSS	Csig	Cbak	Covl
AG y SNR=5dB	1,565891	0,957518	1,742002	57,436299	2,5350	2,0902	1,9622
AG y SNR=10dB	1,753584	0,715890	4,084872	46,491018	2,9953	2,4041	2,3137
WP y SNR=5dB	1,438224	1,293692	2,397989	50,070436	2,1784	2,1221	1,7389
WP y SNR=10dB	1,758696	1,058411	5,512403	38,869705	2,7146	2,5498	2,1958

Tabla 3. Medidas objetivas y compuestas para el ruido Train (señal 2).

Métodos	PESQ	LLR	SNRseg	WSS	Csig	Cbak	Covl
AG y SNR=5dB	1,517738	1,128546	2,039901	46,969165	2,4242	2,1592	1,9092
AG y SNR=10dB	1,769539	0,851277	4,633749	39,711784	2,9267	2,4938	2,3046
WP y SNR=5dB	1,401606	1,762824	1,148015	43,514110	1,7326	2,0717	1,5151
WP y SNR=10dB	1,683493	1,429213	4,733411	33,820642	2,3331	2,5002	1,9807

Tabla 4. Medidas objetivas y compuestas para el ruido babble (señal 3).

Métodos	PESQ	LLR	SNRseg	WSS	Csig	Cbak	Covl
AG y SNR=5dB	2,358871	0,925104	-0,139134	70,955269	2,9249	2,2561	2,5226
AG y SNR=10dB	2,610984	0,803292	2,261646	59,793694	3,3027	2,6060	2,8660
WP y SNR=5dB	2,332687	0,976196	0,310984	42,526354	3,1124	2,4709	2,6743
WP y SNR=10dB	2,756729	0,851985	3,317263	35,366282	3,5603	2,9131	3,1294

Tabla 5. Medidas objetivas y compuestas para el ruido Car (señal 3).

Métodos	PESQ	LLR	SNRseg	WSS	Csig	Cbak	Covl
AG y SNR=5dB	2,192072	0,849315	0,959940	47,190672	3,1162	2,4120	2,5934
AG y SNR=10dB	2,368495	0,640822	3,274511	42,249535	3,4816	2,6767	2,8768
WP y SNR=5dB	2,135140	1,213203	1,904075	42,346550	2,7510	2,4781	2,3952
WP y SNR=10dB	2,535416	0,888287	5,122249	31,027819	3,4286	2,9514	2,9630

Tabla 6. Medidas objetivas y compuestas para el ruido Train (señal 3).

Métodos	PESQ	LLR	SNRseg	WSS	Csig	Cbak	Covl
AG y SNR=5dB	2,139375	0,960323	1,267632	48,437884	2,9589	2,3974	2,4854
AG y SNR=10dB	2,415025	0,726113	3,541916	41,953364	3,4245	2,7178	2,8727
WP y SNR=5dB	2,041837	1,512411	0,516804	37,595159	2,4296	2,3794	2,2002
WP y SNR=10dB	2,383815	1,163438	3,886533	28,958888	3,0726	2,8156	2,7146

Tabla 7. Medidas objetivas y compuestas para el ruido babble (señal 4).

Métodos	PESQ	LLR	SNRseg	WSS	Csig	Cbak	Covl
AG y SNR=5dB	2,488875	0,872818	-0,247321	72,036750	3,0473	2,3038	2,6464
AG y SNR=10dB	2,783339	0,776337	2,403758	63,076123	3,4048	2,6743	2,9956
WP y SNR=5dB	2,372690	1,155015	0,252792	47,566934	2,9071	2,4511	2,5797
WP y SNR=10dB	2,764556	1,003830	2,948096	40,347498	3,3640	2,8588	3,0231

Tabla 8. Medidas objetivas y compuestas para el ruido Car (señal 4).

Métodos	PESQ	LLR	SNRseg	WSS	Csig	Cbak	Covl
AG y SNR=5dB	1,850196	1,051731	0,440690	56,280264	2,6199	2,1522	2,1510
AG y SNR=10dB	2,002945	0,836598	2,903446	46,771515	3,0190	2,4469	2,4506
WP y SNR=5dB	1,757904	1,294496	1,580139	49,135844	2,3788	2,2299	2,0024
WP y SNR=10dB	2,101610	1,118224	4,699039	38,420067	2,8638	2,6657	2,4443

Tabla 9. Medidas objetivas y compuestas para el ruido Train (señal 4).

Métodos	PESQ	LLR	SNRseg	WSS	Csig	Cbak	Covl
AG y SNR=5dB	1,799117	1,174657	1,200238	49,702817	2,5218	2,2217	2,0929
AG y SNR=10dB	2,010614	0,976986	3,714143	44,749133	2,8973	2,5158	2,3991
WP y SNR=5dB	1,970042	1,720243	0,209931	42,854062	2,1251	2,2889	1,9991
WP y SNR=10dB	1,992857	1,417100	3,726034	33,150536	2,5381	2,5893	2,2406

Tabla 10. Medidas objetivas y compuestas para el ruido babble (señal 5).

Métodos	PESQ	LLR	SNRseg	WSS	Csig	Cbak	Covl
AG y SNR=5dB	2,384026	0,853977	1,053290	66,446852	3,0538	2,3748	2,6108
AG y SNR=10dB	2,611374	0,704942	3,815485	54,672523	3,4502	2,7399	2,9525
WP y SNR=5dB	2,352658	0,916959	1,010671	46,244017	3,1519	2,4985	2,6947
WP y SNR=10dB	2,661511	0,791738	4,003254	36,589478	3,5539	2,9023	3,0750

Tabla 11. Medidas objetivas y compuestas para el ruido Car (señal 5).

Métodos	PESQ	LLR	SNRseg	WSS	Csig	Cbak	Covl
AG y SNR=5dB	1,716051	0,923942	1,515191	49,561107	2,7310	2,2028	2,1554
AG y SNR=10dB	1,942544	0,744937	4,008202	43,504260	3,1063	2,5105	2,4718
WP y SNR=5dB	1,623209	1,151912	2,360151	46,591029	2,4672	2,2324	1,9848
WP y SNR=10dB	1,858948	0,926522	5,480011	33,948987	2,9550	2,6302	2,3784

Tabla 12. Medidas objetivas y compuestas para el ruido Train (señal 5).

Métodos	PESQ	LLR	SNRseg	WSS	Csig	Cbak	Covl
AG y SNR=5dB	1,800681	0,983361	1,711161	48,858879	2,7272	2,2605	2,1981
AG y SNR=10dB	1,950810	0,848566	4,368361	40,337880	3,0331	2,5593	2,4476
WP y SNR=5dB	1,634759	1,529099	0,785205	42,551543	2,1224	2,1670	1,8292
WP y SNR=10dB	1,846939	1,163966	4,160789	32,082687	2,7202	2,5544	2,2603

Tabla 13. Medidas objetivas y compuestas para el ruido babble (señal 6).

Métodos	PESQ	LLR	SNRseg	WSS	Csig	Cbak	Covl
AG y SNR=5dB	2,467771	1,001616	1,159121	72,840634	2,8948	2,3767	2,5578
AG y SNR=10dB	2,825876	0,824109	3,592757	57,526975	3,4313	2,8084	3,0442
WP y SNR=5dB	2,447910	1,061081	0,379445	47,737093	3,0476	2,4938	2,6871
WP y SNR=10dB	2,801780	0,942846	3,798866	39,524438	3,4566	2,9359	3,0900

Tabla 14. Medidas objetivas y compuestas para el ruido Car (señal 6).

Métodos	PESQ	LLR	SNRseg	WSS	Csig	Cbak	Covl
AG y SNR=5dB	1,651463	0,802753	1,825881	52,797329	2,7876	2,1688	2,1428
AG y SNR=10dB	1,900973	0,646229	4,216784	44,451008	3,1743	2,4972	2,4823
WP y SNR=5dB	1,640599	1,393953	2,432273	44,800536	2,2447	2,2578	1,8874
WP y SNR=10dB	1,851825	1,097060	5,073848	37,589322	2,7425	2,5757	2,2599

Tabla 15. Medidas objetivas y compuestas para el ruido Train (señal 6).

Métodos	PESQ	LLR	SNRseg	WSS	Csig	Cbak	Covl
AG y SNR=5dB	1,643481	1,007784	1,772202	50,305995	2,5943	2,1791	2,0489
AG y SNR=10dB	1,814459	0,732720	4,290476	43,943693	3,0377	2,4640	2,3719
WP y SNR=5dB	1,626512	1,845979	0,045725	42,194622	1,7945	2,1190	1,6628
WP y SNR=10dB	1,794197	1,512735	3,527810	33,865577	2,3135	2,4768	2,0267

Tabla 16. Medidas objetivas y compuestas para el ruido babble (señal 7).

Métodos	PESQ	LLR	SNRseg	WSS	Csig	Cbak	Covl
AG y SNR=5dB	2,426673	0,993387	1,514123	78,210401	2,8302	2,3419	2,4914
AG y SNR=10dB	2,713845	0,759858	4,029418	62,650699	3,3837	2,7465	2,9510
WP y SNR=5dB	2,466947	1,205567	0,715303	52,825902	2,8646	2,4885	2,5929
WP y SNR=10dB	2,752774	1,049223	3,974694	44,901552	3,2692	2,8859	2,9585

Tabla 17. Medidas objetivas y compuestas para el ruido Car (señal 7).

Métodos	PESQ	LLR	SNRseg	WSS	Csig	Cbak	Covl
AG y SNR=5dB	2,319981	0,862160	2,501223	50,484107	3,1504	2,5471	2,6668
AG y SNR=10dB	2,590482	0,670960	4,983824	44,532477	3,5639	2,8745	3,0241
WP y SNR=5dB	2,234739	1,509512	2,820627	49,616169	2,4407	2,5326	2,2728
WP y SNR=10dB	2,602836	1,239504	5,554894	40,579761	3,0218	2,9441	2,7706

Tabla 18. Medidas objetivas y compuestas para el ruido Train (señal 7).

Métodos	PESQ	LLR	SNRseg	WSS	Csig	Cbak	Covl
AG y SNR=5dB	2,331498	0,935096	2,671574	48,768999	3,0978	2,5754	2,6507
AG y SNR=10dB	2,652227	0,698901	5,177678	42,363976	3,5918	2,9314	3,0747
WP y SNR=5dB	2,142034	2,029276	0,355837	44,333000	1,8975	2,3700	1,9690
WP y SNR=10dB	2,459212	1,835496	3,986637	36,030828	2,3629	2,8084	2,3817

Tabla 19. Medidas objetivas y compuestas para el ruido babble (señal 8).

Métodos	PESQ	LLR	SNRseg	WSS	Csig	Cbak	Covl
AG y SNR=5dB	2,301132	0,626912	0,642714	66,741035	3,2348	2,3072	2,6582
AG y SNR=10dB	2,521828	0,475363	2,901450	50,727893	3,6680	2,6671	3,0256
WP y SNR=5dB	2,085414	0,765448	1,173752	48,499498	3,1264	2,3653	2,5414
WP y SNR=10dB	2,622785	0,477888	4,741903	37,397670	3,8462	2,9246	3,1989

Tabla 20. Medidas objetivas y compuestas para el ruido Car (señal 8).

Métodos	PESQ	LLR	SNRseg	WSS	Csig	Cbak	Covl
AG y SNR=5dB	2,126217	0,686727	1,593303	51,895675	3,2014	2,3874	2,5907
AG y SNR=10dB	2,407091	0,538604	3,487805	41,794311	3,6141	2,7118	2,9634
WP y SNR=5dB	2,000111	0,942645	2,349489	49,627228	2,8824	2,3907	2,3741
WP y SNR=10dB	2,391495	0,698255	5,208153	37,748053	3,4768	2,8410	2,8974

Tabla 21. Medidas objetivas y compuestas para el ruido Train (señal 8).

Métodos	PESQ	LLR	SNRseg	WSS	Csig	Cbak	Covl
AG y SNR=5dB	2,113606	0,700254	1,072067	49,203584	3,2041	2,3674	2,5925
AG y SNR=10dB	2,370974	0,586760	3,150822	42,111876	3,5399	2,6710	2,9074
WP y SNR=5dB	2,005875	1,160894	0,553004	45,237335	2,7008	2,3110	2,2977
WP y SNR=10dB	2,331203	0,881895	3,806543	34,582939	3,2800	2,7460	2,7770

Anexo II Espectrogramas y gráficos de potencia de compensación de pérdidas auditivas.

Señal 4

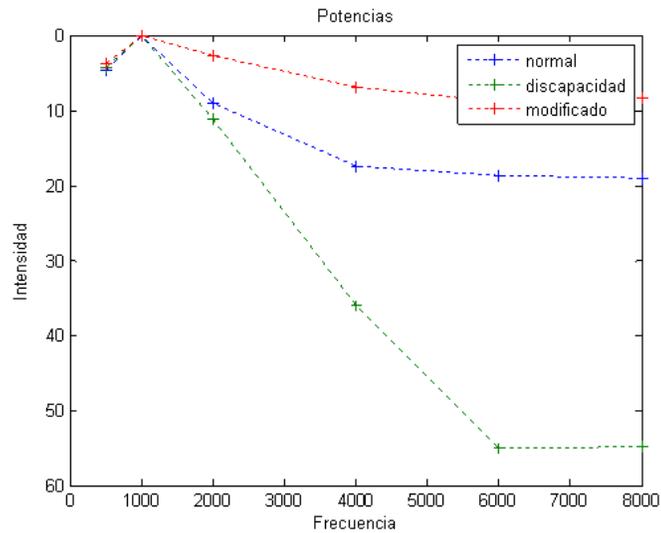


Figura 1. Representación de la potencia en dB SPL de la señal original (azul), señal con pérdida simulada (verde) y de la señal modificada por el algoritmo basado en WP (rojo).

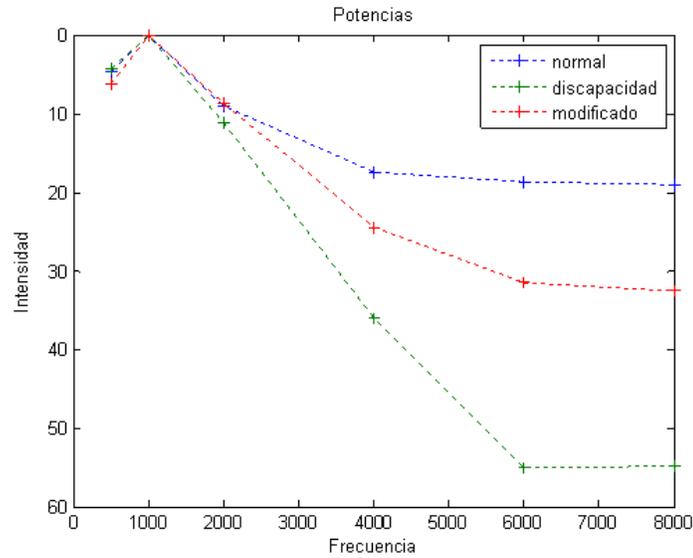


Figura 2. Representación de la potencia en dB SPL de la señal original (azul), señal con pérdida simulada (verde) y de la señal utilizando Gammatone (rojo).

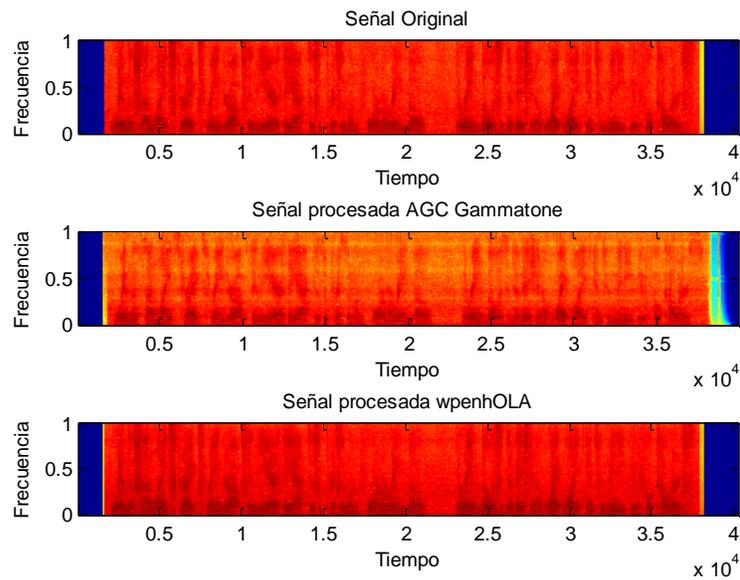


Figura 3. Espectrogramas de la señal original, procesada por Gammatone y procesada por Wavelet Packets.

Señal 5

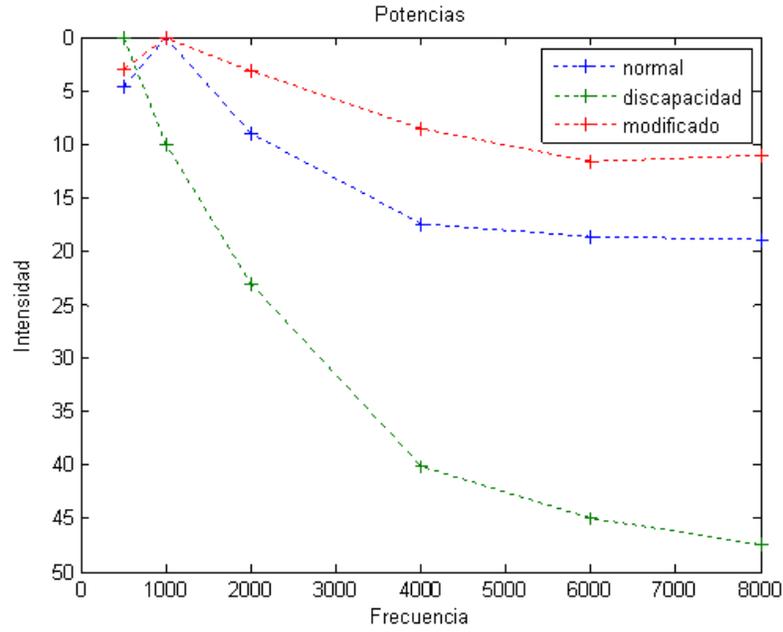


Figura 4. Representación de la potencia en dB SPL de la señal original (azul), señal con pérdida simulada (verde) y de la señal modificada por el algoritmo basado en WP (rojo).

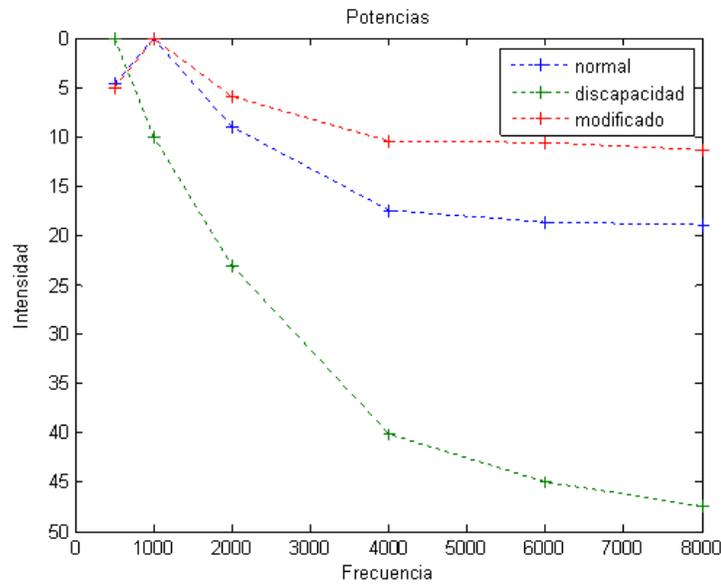


Figura 5. Representación de la potencia en dB SPL de la señal original (azul), señal con pérdida simulada (verde) y de la señal utilizando Gammatone (rojo).

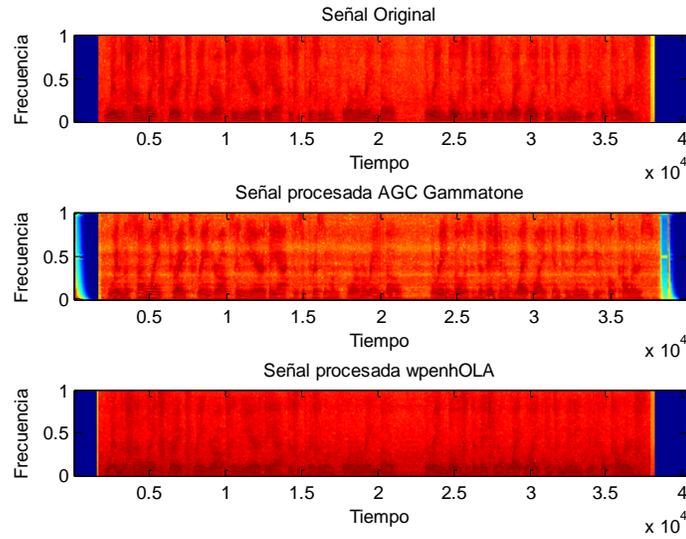


Figura 6. Espectrogramas de la señal original, procesada por Gammatone y procesada por Wavelet Packets.

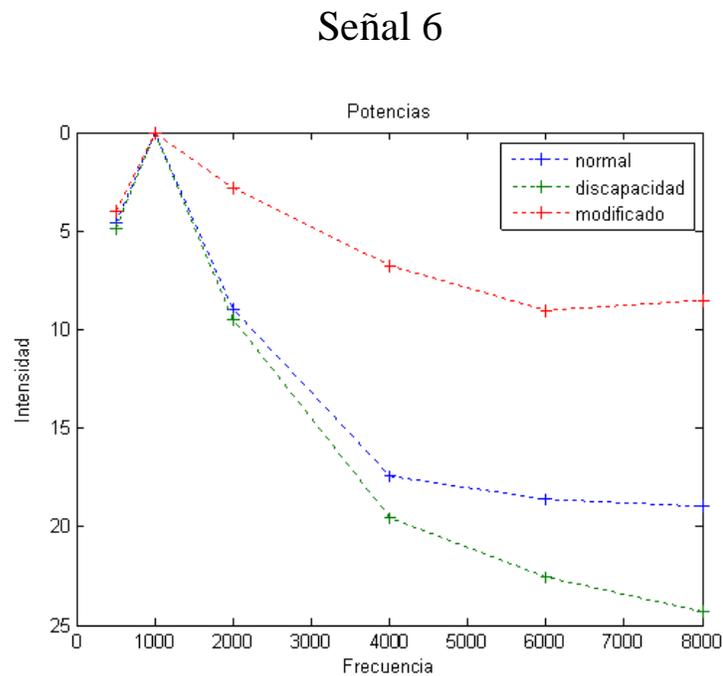


Figura 7. Representación de la potencia en dB SPL de la señal original (azul), señal con pérdida simulada (verde) y de la señal modificada por el algoritmo basado en WP (rojo).

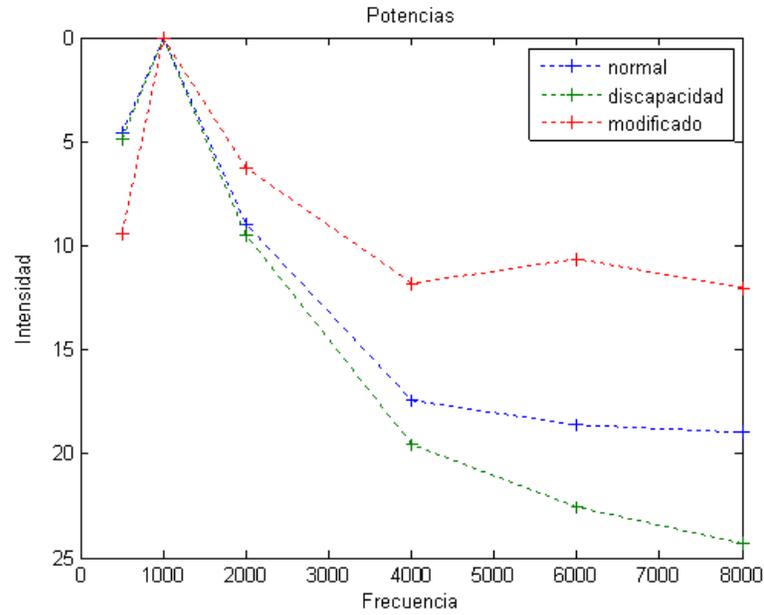


Figura 8. Representación de la potencia en dB SPL de la señal original (azul), señal con pérdida simulada (verde) y de la señal utilizando Gammatone (rojo).

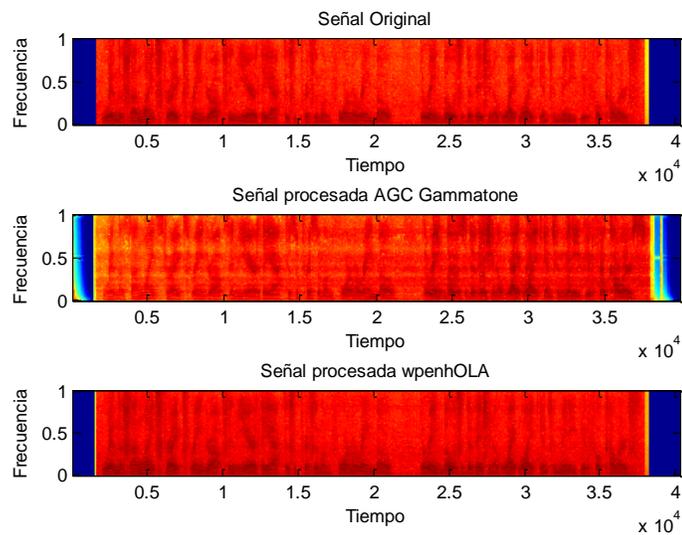


Figura 9. Espectrogramas de la señal original, procesada por Gammatone y procesada por Wavelet Packets.