

3
2ej



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE INGENIERIA

"ELIMINACION DE RUIDO PARA EL REALCE DE VOZ MEDIANTE LA FUNCION DE COHERENCIA"

T E S I S

QUE PARA OBTENER EL TITULO DE:

INGENIERO EN TELECOMUNICACIONES

P R E S E N T A :

EDUARDO SANCHEZ SOTO

DIRECTOR DE TESIS: DR. BOHUMIL PSENICKA



MEXICO, D. F.

1999

TESIS CON
MATE...

27/2/96



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Contenido

I	Introducción (Descripción del Problema)	1
II	Realce de Voz	3
II.1	Historia del Procesamiento de Voz	3
II.2	Clasificación de los Métodos de Realce de Voz	5
II.3	Realce de Voz antes de su Degradación	8
II.3.1	Sistemas Basados en el mejoramiento de SNR	9
II.3.2	Compresión de Amplitud	9
II.3.3	Filtrado y Compresión de Amplitud	12
II.4	Realce de Voz Degradada por Reflexiones	12
II.4.1	Reducción de Eco Lejano	13
II.4.2	Reducción de Eco Cercano	13
II.4.3	Reducción de Eco en comunicaciones de Larga Distancia	16
II.5	Realce de Voz por el cambio en la escala	17
II.5.1	Segmentación de la Forma de Onda	17
II.5.2	Segmentación Espectral	18
III	Filtrado Adaptable	19
III.1	Los Filtros	19
III.1.1	Filtros Analógico y Digitales	19
III.2	Filtrado adaptable	23
III.2.1	Características de un Sistema Adaptable	25
III.2.2	Sistemas de Lazo Cerrado	26
III.2.3	Forma de los Filtros Adaptables	27
III.2.4	Clasificación de los Algoritmos Adaptables	28
III.2.4.1	Algoritmos Tipo Gradiente	28
III.2.4.2	Método de Mínimos Cuadrados	29
IV	Eliminación de Ruido	30
IV.1	Control Activo de Ruido	30
IV.1.1	Aplicaciones del Control Activo de Ruido	32
IV.1.2	Sistemas de Control Activo de Ruido	33
IV.1.2.1	Sistemas de Banda Ancha	34
IV.1.2.2	Sistemas de Banda Angosta	34
IV.1.2.3	Sistemas con mas de un Canal	35
IV.2	Realce de Voz Degradada por ruido Aditivo	35
IV.2.1	Clasificación de los Métodos de Realce de Voz Degradada por Ruido Aditivo	36
IV.2.2	Técnicas de Amplitud Espectral	37

IV.2.2.1 Resta Espectral	38
IV.2.3 Filtrado de Wiener	39
IV.2.4 Filtro Comb	40
IV.2.5 Realce de voz Mediante la Función de Coherencia	41
IV.2.5.1 La Función de Coherencia	41
IV.2.5.1 Reducción de Ruido	42
V Resultados y Conclusiones	46
V.1 Material de Prueba	46
V.2 Pruebas con el Programa	46
V.3 Conclusiones	54

APÉNDICES

A Características Fundamentales de la Voz	55
A.1 Potencia de la Voz	56
A.2 Frecuencia de la Voz	56
A.3 Frecuencia Fundamental de la Voz	57
A.4 Características fundamentales de los sonidos en la Voz	58
A.4.1 Vocales	58
A.4.2 Consonantes	59
B Señales Aleatorias y Densidad Espectral de Potencia	60
B.1 Proceso Aleatorio	60
B.2 Proceso Aleatorio Estacionario	61
B.3 Densidad Espectral de Potencia	61
C Procesador Digital de Señales (DSP) TMS320C54x	63
C.1 Descripción General	63
C.2 Características	63
C.3 Arquitectura	65
C.3.1 Estructura de Buses	65
C.3.2 Organización de la Memoria Interna	67
C.3.2.1 Memoria ROM	67
C.3.2.1 Memoria DARAM y SARAM	67
C.3.3 Unidad Central de Proceso (CPU)	69
C.3.3.1 Unidad Aritmética Lógica (ALU)	69
C.3.3.2 Acumuladores	70
C.3.3.3 Operador de Desplazamientos	71
C.3.3.4 Multiplicador / Sumador	71
C.3.3.5 Unidad de Comparación, Selección y Escritura (CSSU)	72
C.4 Direccionamiento de Datos	73

C.5	Direccionamiento en la Memoria de Programa	73
C.6	Operación del Pipeline	73
C.7	Periféricos	74
D	Listados	75

Capítulo I

Introducción

La esencia y propósito de la voz es la comunicación. El hombre se ha comunicado usando alguna forma de expresión oral por más de 50 000 años; durante estos años en los que el habla se ha convertido en un proceso muy natural, complejo y muy rápido, una simple frase puede contener una cantidad enorme de información y puede expresar ideas tan complejas que van más allá de la simple interpretación literal de las palabras. Una frase puede mostrarnos, entre otras cosas, la edad aproximada, el humor, la educación y el sexo de la persona que habla.

Los efectos que genera el ruido en la voz van desde la simple molestia y fatiga, la dificultad en el entendimiento y hasta la falta total de la comunicación. Por lo cual la ciencia en general y en especial la ingeniería se han interesado en el problema de la eliminación de la interferencia y del ruido en una gran variedad de situaciones en las que es necesario el mejoramiento en la calidad de la voz degradada. Ejemplo de tales situaciones son los sistemas de compresión, reconocimiento y detección de voz.

Las fuentes de ruido son tan variadas como ambientes existen. El problema del ruido es cada vez más evidente, ya que el número de industrias y el gran crecimiento que han sufrido las poblaciones expone a la gente a fuentes de ruido cada vez más variadas. Si pensamos en los cambios que han sufrido los materiales con los que se construye en la actualidad y en la forma en la que se ha tratado de reducir los costos de estos nos daremos cuenta que se tiende a aligerarlos aumentándose así el ruido ambiental. El ruido es una característica de cualquier ambiente, pero es especialmente un problema en las oficinas, vehículos y en las calles de las grandes ciudades. En un ambiente de oficina típico el ruido está formado por las percusiones de máquinas de escribir, los ventiladores de las computadoras, las campanas de los teléfonos, así como el generado por las conversaciones de otras personas y por todo lo que existe en una oficina. En un automóvil se ha demostrado que la relación señal a ruido (SNR) de voz grabada dentro de un auto moviéndose a una velocidad de 90 km/hr con las ventanas cerradas y sin tener encendidos los ventiladores se ve afectada al menos por 5 dB [Lec89][Roe87]; si agregamos a todo esto una cantidad extra de ruido por el ambiente donde se transita, o el tráfico en una calle muy concurrida nos podremos dar cuenta que la calidad de la voz puede ser afectada realmente.

Adicionalmente al ruido ambiental la voz está sujeta a una gran variedad de distorsiones. La reverberación producida por el rebote de las señales en las paredes y el techo. Y más aún, en un sistema de comunicaciones en general la

adquisición de la señal de voz se ve afectada por el tipo de micrófonos usados, su orientación y posición, así como por los medios de transmisión y la forma en la que se lleve a cabo. Es diferente la distorsión y el ruido que sufre una señal en una red telefónica normal a la que se sufre dentro de una red de datos con las nuevas tecnologías como es voz sobre IP (Internet Protocol).

Debido a las grandes cantidades de información que se transmiten por los sistemas de comunicaciones la tecnología y la investigación se han encaminado a mejorar las técnicas de transmisión, de cancelación de ruido e interferencia y de codificación de las señales para mejorar el uso de los canales. Las técnicas adaptables han obtenido grandes resultados y han sido de gran aplicación ya que los medios de transmisión son cambiantes y sobre todo debido a que las señales que se transmiten son aleatorias.

El presente trabajo trata acerca del Realce de Voz, Las Técnicas Adaptables del Filtrado y el Control Activo del Ruido (exclusivamente del ruido en el aire); en especial la que utiliza la función de coherencia como parámetro de decisión y su aplicación a la cancelación de ruido como tema central.

En una primera fase se programó el algoritmo de cancelación de ruido con la función de coherencia en MATLAB y se hicieron algunas pruebas de desempeño para probar su eficacia. Posteriormente debido a que el proceso se lleva a cabo sobre grandes cantidades de información y que se requiere de un proceso en tiempo real se implantó el algoritmo sobre un DSP (Procesador Digital de Señales) de la familia TMS320 de Texas. Finalmente se evalúan los resultados y las condiciones en las que se llevaron a cabo los experimentos y se concluye a partir de ellos.

Capítulo II

Realce de Voz

En las últimas décadas se ha puesto mucho interés en el problema de mejorar la calidad de la voz degradada por ruido aditivo ambiental y la eliminación de eco en las comunicaciones. Se ha intentado resolver este problema de muchas maneras, cada una de estas ha intentado utilizar alguna característica de la voz y/o del ruido que pueda ser de utilidad para llevar a cabo el realce. El grado de éxito de estos métodos depende del objetivo final y de las suposiciones hechas dentro del algoritmo. Dependiendo de cual sea la aplicación final; mejoramiento de la calidad general de la voz, aumento de la inteligibilidad¹ o reducción de la fatiga al escuchar, cada método tiene ciertas ventajas y por lo general no da buenos resultados en otra área de aplicación. En el presente capítulo se tratarán las partes que forman el área del realce de voz sin tomar en cuenta la eliminación de ruido que será estudiada en el siguiente capítulo.

II.1 HISTORIA DEL PROCESAMIENTO DE VOZ

Los primeros pasos que se dieron en el área de procesamiento de voz se remontan al año de 1791 cuando Wolfgang von Kempelen describe en su libro [Kem1791] un modelo mecánico del aparato vocal humano. Posteriormente Charles Wheatstone, 40 años más tarde, construye otro modelo basado en el de Kempelen con capacidad de producir varios sonidos, vocales, consonantes y hasta pequeñas frases.

Durante el siglo XX se pueden distinguir dos vertientes en el procesamiento de voz. Mientras varios hombres de ciencia, entre ellos Graham Bell, siguen el modelo mecánico de producción de voz de Kempelen, otro grupo de investigadores (Helmholz, Miller y Koenig) proponen otro modelo basado en la síntesis de sonidos mediante la superposición armónica de varias ondas senoidales con diferentes amplitudes. De tal manera que en los últimos años estas han sido las dos aproximaciones que han dominado el campo del procesamiento de voz. Por una parte, la aproximación física con el modelado del tracto vocal, y por otra parte la aproximación matemática con el modelado senoidal.

Podríamos hablar de la parte moderna del procesamiento de voz cuando hablamos de los trabajos de Homer Dudley un ingeniero eléctrico de los Laboratorios Bell que en los años 30 [Dud36] construye por primera vez un modelo

¹ Inteligibilidad es la calidad de una palabra, de un discurso, etc. de ser comprensible, es decir, que se pueda entender claramente.

del tracto vocal mediante circuitos eléctricos. Para modular el mensaje su modelo utilizaba un banco de filtros con diez canales de respuesta ajustable. Posteriormente con la ayuda de Riesz y Watkins, Dudley construye el VODER (VOICE DEMONSTRATION RECORDER) y el VOCODER (VOICE CODER).

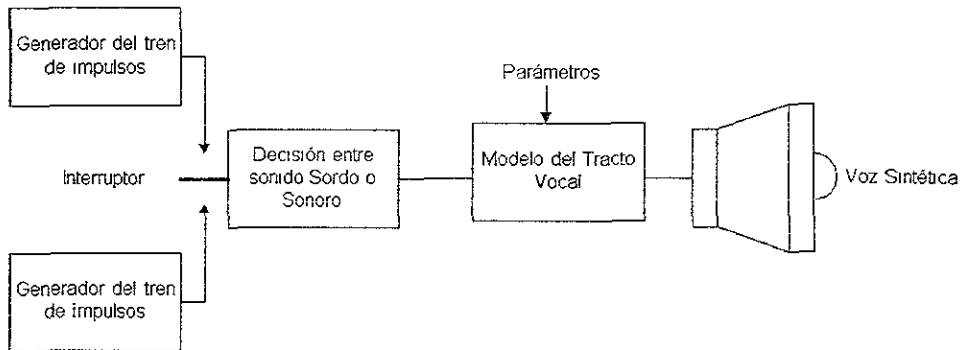


Fig. II.1.1 Esquema de un Vocoder.

El Voder es un sistema en el que un operador puede manipular una serie de palancas y pedales para controlar las fuentes de sonido y el banco de filtros y de esta manera producir voz. El Vocoder, por otra parte, obtiene los parámetros de control a partir de una señal de voz grabada mediante la utilización de un micrófono. El Vocoder obtiene automáticamente la frecuencia fundamental y la ganancia para los filtros directamente de la señal grabada.

Con la utilización del Vocoder, Dudley dio el primer paso en el área de compresión de voz cuando demostró que una señal de voz se puede representar por medio de una serie de parámetros que cambian lentamente en el tiempo, y que con ellos se podía resintetizar la voz mas tarde. De la misma manera cuando utilizó un banco de filtros para modelar el tracto vocal Dudley entró en el área del análisis espectral y se dio cuenta de que la forma en la que se modificaban las componentes en frecuencia podría ser de gran utilidad para el reconocimiento de voz que es otra parte de gran auge del procesamiento de voz en la actualidad.

En los años 60 la atención de los investigadores se vio atraída por la representación en frecuencia del tracto vocal independientemente de los parámetros de la fuente. Atal [Ata71] e Itakura [Ita68] desarrollaron de manera independiente un método de análisis espectral que ahora conocemos como Predicción Lineal. El método se basa en la idea de que una señal de voz en un tiempo t puede ser predecida con gran precisión por una combinación lineal de los valores anteriores de la misma señal.

$$s_i \approx \hat{s}_i = \sum_{j=1}^p \alpha_j s_{i-j} \quad (II.1)$$

donde p es llamado el orden y donde la parte importante del método es la obtención de los coeficientes $\{\alpha_j\}$.

La predicción Lineal tiene muchas interpretaciones. Por ejemplo, en el dominio de la frecuencia podemos decir que los coeficientes calculados $\{\alpha_j\}$ pueden formar un modelo matemático que contiene polos solamente $G/A(e^{j\omega})$, donde $A(z) = 1 - \sum_{j=1}^p \alpha_j z^{-j}$, representa a la envolvente de la respuesta en frecuencia del tracto vocal de la señal. La representación que se obtiene de la predicción lineal se ha utilizado muy eficientemente en el área de codificación.

A partir de los años 80 los modelos de estudio han incluido a las propiedades estadísticas de la señal como un parámetro muy importante. Con la ayuda de herramientas tales como: Modelos de Markov (HMM), la Cuantificación Vectorial y la Teoría de Bayes empresas muy importantes entre las que destacan IBM, los Laboratorios Bell y NEC junto con universidades como CMU, MIT y los Laboratorios Lincoln han puesto mucho énfasis principalmente en el reconocimiento, la codificación y el realce de voz.

II.2 CLASIFICACIÓN DE LOS MÉTODOS DE REALCE DE VOZ

En general el área del realce en el procesamiento de voz intenta mejorar la calidad de la señal de voz, de tal manera que no solamente se incluye la eliminación de ruido. Como ya se ha dicho, también incluye una gran variedad de aplicaciones tales como el control de eco y el procesamiento de la voz antes de que sea degradada entre muchas otras.

Existen muchas maneras de clasificar los métodos de realce de voz. Podemos clasificar a los algoritmos dependiendo de las características de las señales que se utilizan como parte esencial en el método; podríamos clasificarlos según el número de entradas utilizadas o por la aplicación para la que se diseñaron. Siendo esta última una forma muy ilustrativa, ya que nos permite explorar los diferentes ámbitos en los que se utiliza el realce de voz y nos permite también observar la forma en la que se relacionan los algoritmos será la que se utilice para clasificar las diferentes partes que componen esta área del procesamiento de señales.

Una de las parte esenciales de estudio en el realce de voz es la eliminación de ruido. Como la eliminación de ruido es parte muy importante en el presente trabajo se discutirá en detalle en el siguiente capítulo.

Otra parte que tornó gran importancia a partir de la Segunda Guerra mundial es el procesamiento de la señal antes de que esta sea degradada por ruido aditivo. Típicamente tiene aplicación en situaciones donde la persona que escucha se encuentra en un ambiente lleno de ruido y el hablante en un lugar sin interferencias. Este es el caso específico de un piloto que se encuentra dentro de la cabina de un avión o de un helicóptero y tiene que entender las ordenes que le son dichas desde la base en tierra.

Dado que se dispone de la señal sin degradación, los algoritmos de procesamiento de señales desarrollados en este caso en particular son mucho más simples conceptualmente y tienen mucho menos carga operacional que en los casos en los que se dispone solamente de la señal cuando ya esta mezclada con el ruido. Debido a todo esto, la eficiencia de estos algoritmos es mucho mayor; la mejoría en la inteligibilidad de las palabras es sustancial.

Una manera muy fácil y hasta cierto punto obvia de mejorar la calidad de la señal recibida es aumentar el volumen de la señal antes de trasmitirla. Sin embargo, la amplificación de la señal tiene un límite. La señal puede ser amplificada solamente hasta un nivel en el que la persona que escucha no experimente molestias o dolor.

Cuando la intensidad del ruido es constante el método anterior no tiene buenos resultados. No se puede exponer a una persona a niveles de sonido muy altos durante mucho tiempo, si así se hiciera el hombre experimentaría dolor, fatiga y finalmente no comprendería lo que se dice. Para resolver este problema se han propuesto, por una parte, métodos en los que se hace énfasis a aspectos de percepción del ser humano y se enfoca la atención a los sonidos sordos (ver Apéndice A). De tal manera que se amplifican estos sonidos hasta los mismos niveles del resto de la señal. Por otra parte se tiene la compresión de amplitud; se ha probado que el recorte de los picos en una señal da muy buenos resultados al obtenerse una señal mas inteligible, pero las mediciones cuantitativas no son muy alentadoras.

Finalmente, sabemos que la respuesta espectral de la voz se reduce en 6 dB por octava conforme aumenta la frecuencia y que la mayoría de los sonidos que son importantes en la inteligibilidad tienen sus principales componentes en altas frecuencias, de tal manera que una buena forma de aumentar la calidad de la voz es amplificando las bandas de frecuencia que son importantes en la percepción humana.

Cuando estamos dentro de un lugar cerrado el fenómeno de reflexión que se presenta en las paredes, techo y en todos los muebles que estén dentro del cuarto ocasionan efectos muy desagradables en las comunicaciones, tal es el grado de molestia que en ocasiones no existe comunicación. Este tipo de degradación en la voz se refiere especialmente al caso de lugares cerrados en los que el sonido no

solamente viaja de la fuente hasta el receptor, sino que existen múltiples trayectorias por las reflexiones mencionadas. La figura II.1 ilustra este fenómeno.

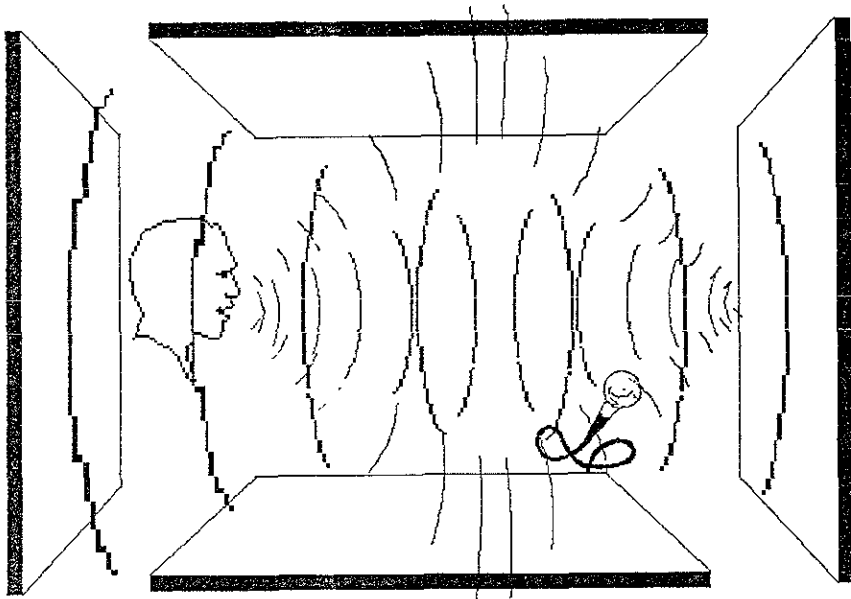


Figura II.1 Fenómeno de reflexión en lugares cerrados que ocasiona ecos en la comunicación.

El problema del realce de voz degradada por eco ha sido estudiado a partir de los años 60 y muy especialmente por los Laboratorios Bell. Durante años el eco ha sido uno de los mayores problemas en las comunicaciones telefónicas, especialmente las comunicaciones telefónicas de larga distancia.

El eco es un fenómeno que tiene muy variadas implicaciones a nivel de comunicación y que puede ser medido de varias maneras. Principalmente el eco en un lugar cerrado puede ser dividido en dos componentes bien diferenciadas. Primero, el eco cercano, que es la componente formada por las primeras reflexiones y que tiene mucha intensidad y en segundo lugar el eco lejano que se forma de las subsecuentes reflexiones y que en intensidad es menor pero no en cantidad. Las primeras aproximaciones para resolver el problema de las reflexiones no toman en cuenta esta clasificación.

Una forma mediante la que se ha atacado este problema es con modelados matemáticos, la respuesta al impulso, del lugar donde se tienen las reflexiones. Posteriormente, con la respuesta al impulso se puede modelar la respuesta total del sistema como la convolución entre la voz y el modelo del cuarto, de tal manera que por medios matemáticos y de transformaciones es posible separar a la voz y

al eco en dos componentes bien definidas. Como se puede intuir es necesario obtener la respuesta al impulso en cada caso en especial, lo que en general limita a esta clase de algoritmos.

Otra aproximación ha sido la síntesis de voz con la ayuda de la predicción lineal. Primeramente se hace una estimación de los parámetros con los que se puede modelar la voz tomando en cuenta la presencia de las reflexiones y posteriormente se sintetiza la voz a partir de los parámetros calculados.

Como ya se ha mencionado, es posible hacer una diferencia entre el eco cercano y lejano. La energía del eco lejano es mucho menor que la energía en la componente de la señal principal y la energía en el eco cercano y se puede representar fácilmente como una degradación en el dominio del tiempo. Para eliminar la influencia de la componente del eco lejano se puede utilizar la supresión de señales de amplitud muy pequeña. Si observamos que las componentes del eco llegan en diferentes tiempos a lugares diferentes podemos pensar que utilizando dos entradas podremos distinguir entre estas componentes debido al retardo. En el dominio de la frecuencia, el retraso provoca la eliminación de diferentes componentes en la frecuencia en cada receptor, de tal manera que si combinamos las dos entradas en cada banda de frecuencias podremos limpiar la señal. Este método solamente elimina las componentes del eco cercano debido a la energía que puede ser medida fácilmente.

Para eliminar el eco en su totalidad se puede utilizar combinaciones de los métodos que se han descrito, de cualquier manera aún existe muchos problemas que no se han podido resolver. La calidad de la señal resultante es mucho mejor si tomamos en cuenta solamente mediciones cuantitativas, la inteligibilidad de la señal aún no ha podido ser mejorada.

Si el fin del realce de voz por una parte es que la persona que escucha pueda entender bien, aumentar o disminuir la velocidad con la que se transmite la señal puede ser considerada como una parte del área del realce de voz en algunos casos (personas con algún tipo de deficiencia, personas que están aprendiendo una nueva lengua o la comunicación con un buzo que esta utilizando una mezcla con helio para bucear a grandes profundidades) a pesar de que la señal no ha sido degradada por ruido.

II.3 REALCE DE VOZ ANTES DE SU DEGRADACIÓN

Como ya se ha mencionado el realce de voz anterior a la degradación por ruido aditivo es una parte muy importante del procesamiento de voz. En la resolución de este problema se han utilizado muchas técnicas con muy diferentes puntos de vista.

El interés final es aumentar la inteligibilidad, de tal manera que las mediciones y los algoritmos que se utilizan son cualitativos y se basan en su mayoría en aspectos de la percepción del ser humano y en mediciones experimentales.

II.3.1 Sistemas Basados en el mejoramiento de la Relación Señal a Ruido (SNR).

Para aumentar la relación Señal a Ruido (SNR) sin distorsionar la señal original lo mas sencillo es amplificarla antes de ser enviada sin que sea reducida su calidad. Si el ruido es intermitente el proceso puede ser selectivo, es decir, se puede amplificar la señal de manera proporcional a la cantidad de ruido existente en el ambiente a través de un sistema de control automático de ganancia (AGC).

II.3.2 Compresión de Amplitud

Distorsionar una señal puede provocar resultados alarmantes si se hace una medición cuantitativa, por ejemplo la SNR, pero perceptualmente se ha observado que puede dar buenos resultados. Existen muchas formas de distorsionar una señal, en amplitud, en fase o en frecuencia. Para fines del realce de voz la distorsión en amplitud ha demostrado una gran eficiencia.

La distorsión en amplitud es la deformación que experimenta una señal al pasar a través de un circuito no lineal. Existen muchas formas de distorsión en amplitud, casi tantas como curvas nos podamos imaginar, pero para este caso en particular se pueden mencionar el recorte de picos simétrico y asimétrico, el recorte central y la rectificación lineal (ver fig. II.3.2).

A partir de estudios hechos por Licklider [Lic46] se ha podido observar que la perdida de inteligibilidad depende de la distorsión a la que sea sometida la voz. En un ambiente sin ruido la distorsión en amplitud en forma de recorte de picos, ya sea simétrico o no simétrico, no reduce la inteligibilidad, de hecho es casi nula. A pesar de haber llegado a niveles de distorsión en los que se conservó solo un décimo de la amplitud el 96% de las palabras en las mediciones fueron reconocidas. En cambio con la distorsión lineal la voz fue casi irreconocible y con el recorte central la voz se escuchaba como ruido atmosférico. La inteligibilidad en el recorte central se reduce en gran medida debido a que elimina las consonantes débiles y mantiene las vocales que son menos importantes para la comprensión.

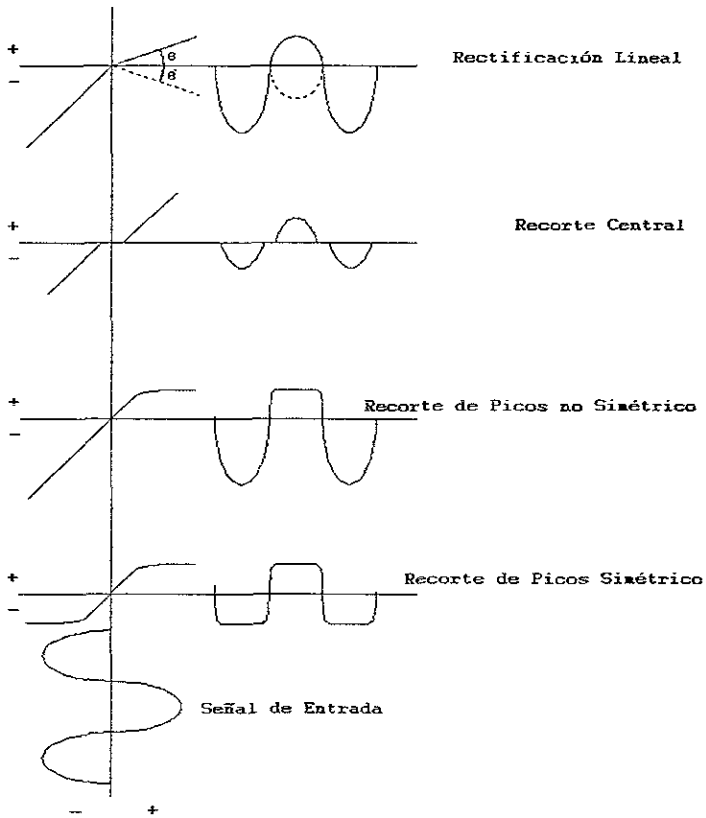


Fig. II.3.2 Diferentes Tipos de Distorsión en Amplitud

El resultado de los experimentos al agregar ruido mostraron que si en ambos extremos el ruido estaba presente la distorsión de amplitud por recorte de picos deterioraba aún mas a la señal, pero cuando solo el ruido estaba presenta del lado del receptor la inteligibilidad no disminuyo a pesar de tener un recorte muy grande.

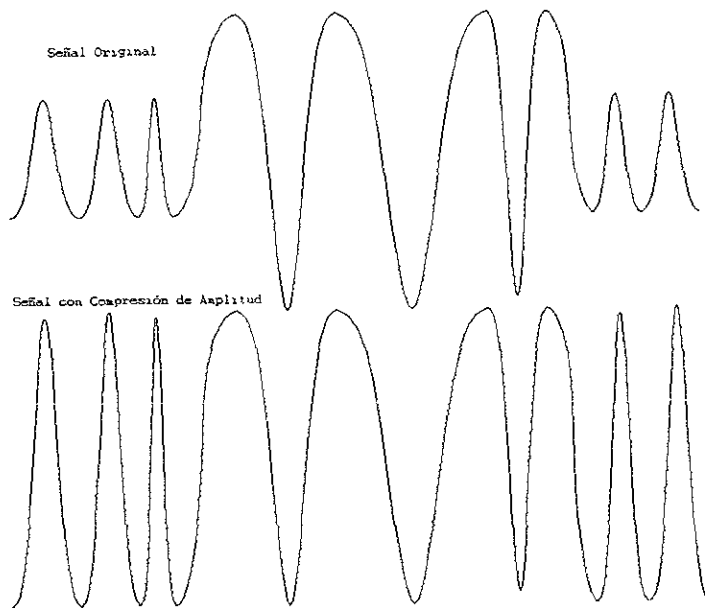


Fig. II.3.3 Compresión de Amplitud en una señal

Del análisis de la voz y de estudios de fonación (ver apéndice A) sabemos que la potencia de las consonantes, parte esencial en la comprensión, es del orden de 12 dB por debajo de la potencia de las vocales. Cuando se recortan los picos en una señal de voz, se están recortando componentes de las vocales y al amplificar de nuevo la señal se amplifican tanto vocales como consonantes de tal manera que se les da énfasis a las últimas.

En un ambiente con ruido la potencia de transmisión necesaria para obtener un grado de inteligibilidad puede ser muy alta, si enviamos la misma señal con la misma potencia y con recorte de picos la eficiencia disminuye, pero si queremos obtener un grado mas alto de eficiencia en la transmisión es necesaria más potencia en la primera señal que en la que tiene recorte de picos. Esto quiere decir, que en un sistema de comunicaciones que tiene limitaciones en la amplitud y potencia de las señales que puede manejar se obtienen mucho mejores resultados, hablando de inteligibilidad, en las señales con un recorte de picos. Este efecto ha sido muy utilizado en las estaciones de radio debido al ahorro que se obtiene en la potencia transmitida. Por ejemplo, en un transmisor de voz con modulación AM² se pueden ganar 14 dB en potencia de transmisión, sin pérdida en inteligibilidad, si la señal de voz se pasa primero por un recorte de picos de 24

² Kryter, K. D., J. C. R. Licklider, y S. S. Stevens. "Premodulation Clipping in AM Voice communication." *Journal of Acoustical Society of America* 19(January 1947):125-131.

dB³ y posteriormente una amplificación de los mismos 24 dB para obtener un porcentaje de modulación del 100%.

II.3.3 Filtrado y Compresión de Amplitud

La compresión de amplitud que se ha explicado en la sección anterior tiene ciertas desventajas que tienen como principal inconveniente la distorsión armónicas causadas por el recorte de la señal. Las frecuencias generadas se encuentran en su mayoría por arriba de los 5kHz, pero también se obtienen frecuencias que afectan a los formantes de los fonemas (ver Apéndice A). Se ha demostrado que el primer formante no es tan importante como el segundo y los formantes de orden alto en la inteligibilidad [Tho68], de tal manera que si podemos eliminamos el primer formante y aumentar la potencia de las consonantes sin generar armónicas, en este orden, podremos lograr una gran calidad en la señal para que pueda ser fácilmente comprendida.

Para eliminar el primer formante se utiliza un filtro paso altas y para hacer la compresión de amplitud requerida solamente se aumentan la potencia de la señal con menor amplitud, en especial las consonantes, al mismo nivel que las vocales sin introducir distorsión.

Esta técnica de realce de voz tiene un problema muy importante. Es muy difícil encontrar la frecuencia en la cual se debe realizar el filtrado paso altas. El valor de frecuencia en el que se encuentra el primer formante puede ser muy diferente en cada uno de los fonemas, de tal manera que fijar dicha frecuencia puede ir en detrimento de ciertos fonemas y en ventaja de otros. Experimentalmente Thomas⁴ reporta en un artículo que el mejor filtro paso altas tiene una frecuencia de corte de 110 Hz y una caída de 12 dB / octava y en otro artículo⁵ encontró que la frecuencia es 1500 Hz con 18 dB / octava.

II.4 REALCE DE VOZ DEGRADADA POR REFLEXIONES

Las reflexiones de una señal de acústica en un lugar cerrado producen eco (versiones retrasadas de la señal original). El retraso puede ir desde milisegundos hasta algunos segundos, dividiendo así al eco en cercano y lejano. El eco cercano es el producido por las reflexiones simples en los objetos, paredes y techo que se encuentran cerca del hablante y del micrófono, mientras que el eco lejano es

³ La distorsión se indica en terminos del número de dB en los cuales se ha reducido la amplitud máxima de la señal.

⁴ Thomas, I. B., y R. J. Niederjohn. "The Intelligibility of Filtered-Clipping speech in noise." *Journal of the audio Engineering Society* 18 (Enero 1970):299-330.

⁵ Thomas, I. B., y William J. Ohley. "Intelligibility Enhancement Through Spectral Weighting." *Proceedings of Conference on Speech Communications and Processing* (1972):360-363.

producido por múltiples reflexiones. El eco lejano se caracteriza por la pérdida rápida de potencia, aproximadamente 60 dB en un tercio de segundo. La forma como es percibido el eco en una conversación es muy diferente; el eco cercano afecta de manera muy importante a la respuesta en frecuencia de los fonemas, la voz se escucha hueca, en cambio, el eco lejano se escucha claramente como una versión retrasada de la misma señal de voz.

II.4.1 Reducción de Eco Lejano

La compresión de amplitud y muy especialmente el recorte central ha demostrado una gran eficiencia en la eliminación de ecos lejanos debido a que tienen baja amplitud. Cuando se utiliza el recorte central se eliminan las componentes de la señal que no alcanzan un valor determinado en amplitud. En algunos algoritmos la entrada se divide en varias bandas que pasan a través del recorte central por separado, de tal forma que la distorsión armónica que causa no afecte a toda la señal.

II.4.2 Reducción de Eco Cercano

El eco cercano presenta mayores complicaciones por sus características y por lo tanto la forma de abordar el problema ha sido totalmente diferente. Algunas formas de eliminar el eco cercano hacen uso de dos o mas micrófonos. Debido a que los micrófonos están en diferente posiciones el eco llega a cada micrófono en tiempos diferentes y por lo tanto la cancelación de frecuencias que provoca la mezcla de ambas señales son diferentes en cada entrada. La reconstrucción de la señal se lleva a cabo por la combinación de las bandas de frecuencia con mayor energía en cada micrófono. Otra manera popular utiliza la predicción lineal para reconstruir la señal, obteniendo los parámetros directamente de la señal con eco.

Considerando a la señal contaminada con eco como una combinación lineal de ambas es posible separarlas para eliminar el eco mediante la utilización del filtrado homomórfico.

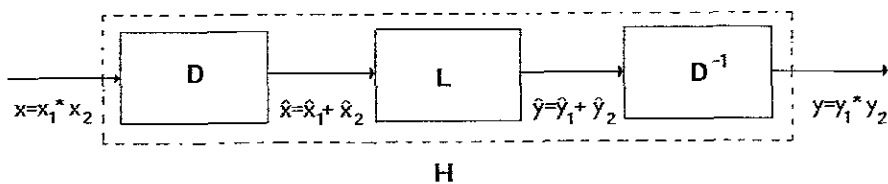


Fig. II.4.1 Esquema de la Deconvolución Homomórfica

De la figura podemos ver que el sistema H tiene la siguiente propiedad, si $x = x_1 * x_2$, entonces,

$$y = H(x_1 * x_2) = H(x_1) * H(x_2) = y_1 * y_2 \quad (II.2)$$

esta ecuación representa el principio de superposición en la convolución. El sistema D llamado sistema característico para la deconvolución homomórfica es no lineal y tiene la siguiente propiedad, si $x = x_1 * x_2$, entonces,

$$\hat{x} = D(x_1 * x_2) = D(x_1) + D(x_2) = \hat{x}_1 + \hat{x}_2 \quad (II.3)$$

de esta manera podemos ver que el sistema D convierte una convolución en una suma de las señales correspondientes. El sistema D^{-1} es el inverso de D y L es un sistema en el que si $\hat{x} = \hat{x}_1 + \hat{x}_2$,

$$\hat{y} = L(\hat{x}_1 * \hat{x}_2) = L(\hat{x}_1) + L(\hat{x}_2) = y_1 + y_2 \quad (II.4)$$

En general todos los sistemas homomorficos son iguales, sólo difieren en la parte lineal. Por lo tanto el diseño del sistema se reduce a encontrar el sistema lineal L adecuado.

Debido a que estamos hablando de sistemas digitales D representa a la transformada Z por cumplir con la característica ya mencionada. El segundo elemento de la transformación es un logaritmo complejo que produce una suma de logaritmos a partir del producto. Finalmente la salida $\hat{x}(n)$ se obtiene a partir de la transformada Z^{-1} que es la transformada inversa.

Si ahora aplicamos estas ecuaciones al caso específico de una señal con eco

$$x(n) = s(n) + \alpha s(n - n_0) \quad (II.5)$$

y la transformamos al dominio de Z ,

$$X(z) = S(z) (1 + \alpha z^{-n_0}) \quad (II.6)$$

ahora si aplicamos el operador lineal, es decir, el logaritmo obtenemos

$$\hat{X}(z) = \log[S(z)] + \log(1 + \alpha z^{-n_0}) \quad (II.7)$$

si $\alpha < 1$, ya que la amplitud de las reflexiones es siempre menor a la de la señal que llega directamente a los micrófonos desde la fuente, podemos escribir,

$$\hat{X}(z) = \log[S(z)] + \sum_{n=1}^{\infty} (-1)^{n+1} \frac{\alpha^n}{n} z^{-n_0 n} \quad (II.8)$$

Se podría demostrar que en la mayoría de los casos el primer termino no contribuye en gran medida al resultado ya que se hace cero rápidamente, mientras que el segundo termino genera una secuencia de pulsos a intervalos n_0 . Si representamos gráficamente (ver fig.II.4.2) la respuesta obtenida en el dominio de la frecuencia podremos notar que en una señal típica degradada por eco presenta un pico predominante debido a la secuencia de impulsos.



Fig. II.4.2 Salida típica del sistema D . Se puede notar el eco en la parte derecha de la imagen.

El sistema lineal que debe ser usado es invariante en la frecuencia,

$$\hat{y}(n) = g(n) \cdot \hat{x}(n) \quad (II.9)$$

Si se quiere eliminar el eco, debemos escoger a $g(n)$ para eliminar los picos que ocurren para $n > n_0$, o si lo que queremos es detectar el eco, debemos tomar $g(n)$ borrando los valores de $\hat{x}(n)$ para $n < n_0$. Es necesario hacer notar que a pesar de que no se necesita saber la magnitud del eco debido a su posición en el dominio de la frecuencia, si se requiere tener noción del tiempo de retraso para poder seleccionar a $g(n)$.

Hasta el momento se han visto algunas maneras para eliminar el eco, pero estos algoritmos presenta una solución que solamente funciona para el eco lejano o cercano de manera independiente. Es posible realizar una combinación de ambos métodos y reducir el eco en su totalidad o utilizar las características del eco para obtener un algoritmo que englobe ambos problemas.

El método propuesto por Allen [All77] utiliza la separación en bandas de las señales que se reciben en dos o más micrófonos, posteriormente con la ayuda de la función normalizada de crosocorrelación se filtra una de ellas en el dominio de la frecuencia y se regresa al del tiempo. El filtrado se lleva a cabo a partir de las propiedades espectrales de las señales. Si se colocan de manera adecuada los micrófonos es posible recibir la señal directa al mismo tiempo y el eco en tiempos diferentes en cada micrófono, de tal manera que si se comparan las señales captadas en pequeños lapsos, solamente tendrán parecido en la componente de la señal sin reflexiones.

ii.4.3 Reducción de Eco en Comunicaciones de Larga Distancia

El efecto del eco telefónico en los usuarios es muy complejo, puede depender de muchos factores, entre los que se encuentra los eléctricos, los acústicos y el humor de los usuarios. Dentro de esta área podemos hablar de cancelar o del supresor de eco, tareas que se llevan a cabo mediante la detección de la actividad de los usuarios en ambos extremos. El primer trabajo publicado al respecto pertenece a Sondhi [Soh67] de los Laboratorios Bell y el primer circuito integrado para esta aplicación apareció en el año 1969.

El hombre por naturaleza tiene una realimentación de lo que dice, debido a que uno escucha las cosas que dice. Esta característica ha sido parte natural de la telefonía, debido a que son inevitables las señales que se reflejan en conexiones que usan dos líneas en vez de cuatro. Tienen entre sus ventajas el ayudar al hombre a regular el volumen de la voz; pero si esta señal tiene un retraso considerable, este efecto no ayuda y se escucha un eco formado por nuestra propia voz que puede ser muy perjudicial para mantener una conversación fluida. Las comunicaciones de larga distancia y las que utilizan técnicas de codificación de canal son en general las que más sufren por este problema, ya que la gravedad del eco depende de su amplitud, frecuencia y retardo principalmente.

Las principales causas de producción de eco en telecomunicaciones son: el acoplamiento de impedancias en los convertidores de dos a cuatro hilos, el acoplamiento acústico en los teléfonos de 4 hilos en telefonía móvil y digital, la realimentación de la bocina al micrófono en los teléfonos.

La manera más usada para abordar el problema del eco ha sido la supresión. Un supresor de eco (ver fig. II.4.3.1) es la alternativa más simple para las comunicaciones "half duplex", bloqueando la señal en la dirección de la fuente inactiva. Debido a que había sido muy difícil detectar cuando se presenta voz o cuando no se presenta, el supresor de eco elimina algunas partes de la voz si la comunicación es simultánea.

En la actualidad el desarrollo en la tecnología de los DSP ha hecho más barata la cancelación de eco. Un filtro adaptable se puede usar para modelar la trayectoria

del eco y de esta manera sintetizar una copia del eco, la cual se resta de la señal transmitida para cancelar el eco. De esta manera, el eco es atenuado sin introducir gran distorsión y hace posible la comunicación simultánea.

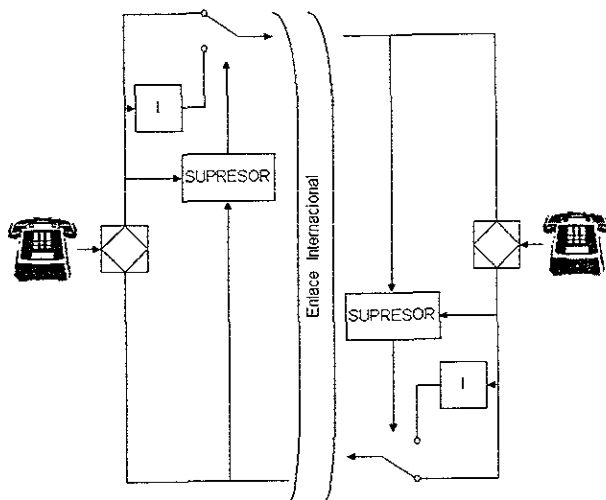


Fig. II.4.3.1 Supresor de Eco.

II.5 REALCE DE VOZ POR EL CAMBIO DE ESCALA

Como ya se ha dicho, este tipo de realce de voz, si bien, no tiene involucrado directamente ruido la voz procesada es preferida por una persona en circunstancia especiales. Una persona con algún tipo de deficiencia y las personas que están aprendiendo una nueva lengua son ejemplos claros de aplicaciones de este tipo de realce de voz.

II.5.1 Segmentación de la Forma de Onda

La forma mas fácil de realizar el proceso es aumentar o disminuir la velocidad de reproducción de la voz. En experimentos se ha encontrado que el máximo cambio no debe ser superior al 10%⁶, de no ser así se tendrá demasiada distorsión que al final hará a la voz ininteligible. El ser humano a pesar de tener limitaciones en la velocidad de la voz que puede producir, puede escuchar y entender voz

⁶ Lee, F. F. "Time Compression and Expansion of Speech by the Sampling Method " Journal of the Acoustical Engineering Society 20 (noviembre 1972): 738-742.

reproducida a velocidades mucho mayores que las que puede producir. El efecto que produce el cambio de velocidad en la reproducción se puede ver como una expansión en el espectro (ver fig. II.5.1) si se aumenta la velocidad o como una compresión si se reduce. De tal manera que se tiene un rango mas amplio en frecuencia dentro de la voz procesada.

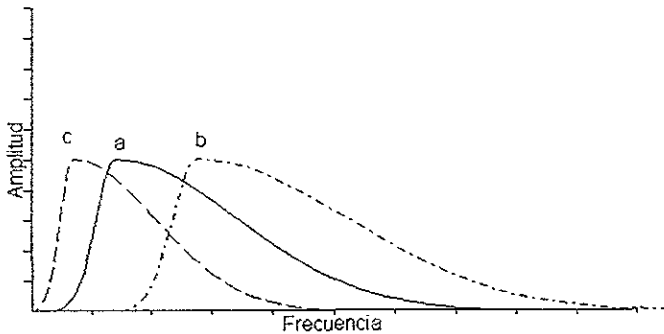


Fig. II.5.1 Espectro de una señal a diferentes velocidades de reproducción. a) señal sin modificación, b) señal con aumento en la velocidad y c) señal con disminución en la velocidad de reproducción.

Usualmente otra manera que se ha implementado para reducir la velocidad de reproducción es dividir la señal en pequeñas particiones descartando algunas de ellas periódicamente y reconstruyendo la señal con las partes que resten. Este proceso es muy sencillo y se puede utilizar para llevar a cabo compresión de voz. Es necesario tomar en cuenta que la duración de los segmentos eliminados no debe ser mayor a 100 ms ya que es esta la duración límite para mantener los fonemas (ver Apéndice A).

II.5.2 Segmentación Espectral

Otra manera que difiere de las ya mencionadas divide la señal en bandas de frecuencia por medio de un banco de filtros y posteriormente se aumenta o reduce la velocidad de reproducción de cada una de ellas, aumentando así la duración de algunos sonidos que son parte muy importante en la inteligibilidad de la voz, finalmente las diferentes bandas resultantes se combinan para obtener una señal realzada. Una de las ventajas de este método con respecto a los anteriores es la reducción en la distorsión total que sufre la señal, es posible modificar solamente ciertas bandas de la señal sin tener que hacerlo con toda ella.

Filtrado Adaptable

En este capítulo se dará una definición de filtro, se discutirán las diferentes clases de filtros que existen, pasando por los filtros digitales (IIR, FIR), el trabajo de Wiener se hará especial énfasis en las técnicas y procesos existentes en el filtrado adaptable como son el LMS y el RLS.

III.1 LOS FILTROS

El filtrado ha sido por años la parte más usada del procesamiento de señales. Los filtros se usan normalmente para eliminar o atenuar la parte del espectro que no nos interesa en una señal, de tal manera, que se le da realce a la parte de la señal seleccionada. Un filtro se puede ver como un sistema que puede discriminar lo que pasa a través de él. Por ejemplo un filtro de aire, tiene la capacidad de eliminar las impurezas en el aire y no dejarlas pasar, es decir, es un discriminador y selecciona solo la parte de interés. Los filtros tienen aplicación en diversos campos del conocimiento como la astronomía, la navegación, la sismología, la biomedicina y por supuesto en la ingeniería.

El filtrado se puede llevar a cabo utilizando circuitos analógicos o digitales. Los primeros manejan señales continuas en el tiempo y no requieren de los convertidores analógico/digital como los segundos, pero estos tienen todas las ventajas de los sistemas digitales.

III.1.1 Filtros Analógicos y Digitales

De la teoría básica de circuitos y de electricidad se sabe que los capacitores e inductores tienen una impedancia (Z) que depende de la frecuencia, por ejemplo para un capacitor,

$$Z_c = \frac{1}{j\omega C} \quad (III.1)$$

y para un inductor,

$$Z_L = j\omega L \quad (\text{III.2})$$

donde $\omega = 2\pi f$; si combinamos estos elementos con resistencias se puede construir un circuito cuya respuesta depende de la frecuencia (ver fig. II.1.1).

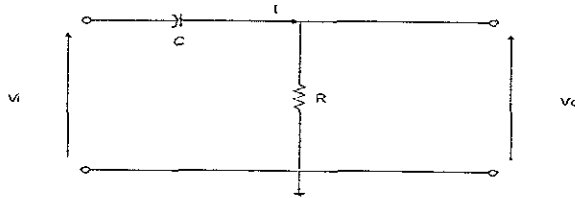


Fig. II.1.1 Filtro analógico paso altas.

Si aplicamos la ley de Ohm en el circuito de la figura II.1.1 podemos tener,

$$V_i = I(R + Z_C) \quad I = \frac{V_i}{R + \frac{1}{j\omega C}} \quad (\text{III.3})$$

despejando a la corriente y sustituyendo las impedancias si $V_o = IR$ obtenemos la función de transferencia,

$$H(j\omega) = \frac{V_o}{V_i} = \frac{R}{R + \frac{1}{j\omega C}} \quad (\text{III.4})$$

a partir de la ecuación anterior nos podemos dar cuenta que entre mayor sea la frecuencia, menor será la impedancia general del circuito. El filtrado analógico se realiza en señales continuas y nos da como resultado también señales continuas. Para su elaboración, además de los elementos que ya hemos mencionado, se utilizan amplificadores operacionales entre otros elementos. En general sufren de sensibilidad al ruido, no linealidades, un rango dinámico limitado, falta de estabilidad por cambio en las características de los componentes y no tienen la misma respuesta a entradas iguales. Por otro lado un filtro digital toma como entrada señales discretas en el tiempo y nos da también señales discretas.

Los filtros digitales que pueden ser implantados en un DSP, un microcontrolador o una computadora digital tienen dos ventajas muy importantes que los diferencian de los filtros analógicos. La primera es la reprogramabilidad, en el sistema digital es posible grabar otro algoritmo que tenga una respuesta en

frecuencia y en fase diferente sin tener que cambiar todo el hardware. La segunda, es posible actualizar los coeficientes de los filtros durante la operación del sistema, de esta manera es posible implementar los filtros adaptables

En general los filtros digitales pueden ser modelados por la siguiente ecuación y a partir de ella dividirlos en dos grandes grupos por la duración de su respuesta al impulso,

$$y(n) = \sum_{q=0}^Q b_q x(n-q) - \sum_{p=1}^P a_p y(n-p) \tag{III.5}$$

los límites en las sumatorias están restringidos a valores finitos para que el filtro sea realizable.

La primer clase de filtros digitales tiene a_p igual a cero para toda p y se llaman filtros FIR (**F**inite **I**mpulse **R**esponse) debido a que la respuesta al impulso decae hasta llegar a cero en con un número finito de muestras. A este tipo de filtros se les llama también no recursivos o MA (**M**oving **A**verage) ya que la salida es simplemente un promedio escalado de los valores de entrada.

Los filtros FIR se caracterizan por no tener realimentación, y su estructura para segundo orden se presenta en la figura II.1.1.

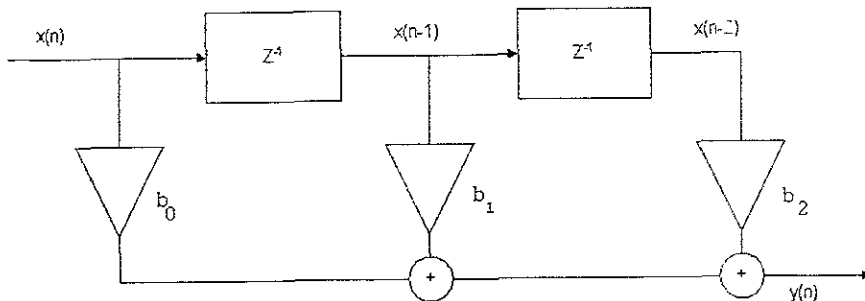


Fig. II.1.2 Estructura para un filtro FIR de segundo orden.

La ecuación general de los filtros FIR en el dominio de Z es,

$$H(z) = \frac{Y(z)}{X(z)} = b_0 + b_1 z^{-1} + b_2 z^{-2} + \dots + b_L z^{-L} \tag{III.6}$$

$$H(z) = \sum_{i=0}^k b_i z^{-i} \quad (III.7)$$

si utilizamos la transformada Z inversa podemos obtener la ecuación en diferencias,

$$y(n) = \sum_{i=0}^k b_i x(n-i) \quad (III.8)$$

La segunda clase de filtros, los filtros IIR (Infinite Impulse Response), completan la clasificación. Los filtros IIR incluyen a los filtros recursivos o AR (AutoRegresive) y a los ARMA (AutoRegresive Moving Average). Para los filtros AR $b_q = 0$ para $q = 1 \dots Q$. La estructura de un filtro IIR de segundo orden se presenta en la figura II.1.3.

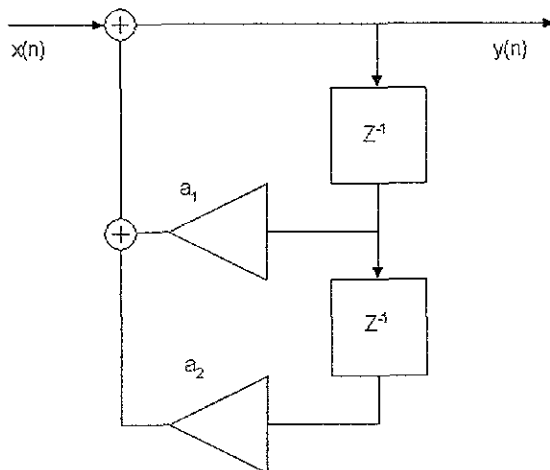


Fig II.1.3 Filtro IIR de segundo Orden.

Siguiendo la trayectoria de la señal de entrada podemos escribir el modelo matemático en diferencias para el filtro,

$$y(n) = x(n) + a_1 y(n-1) + a_2 y(n-2) \quad (III.9)$$

pasando al dominio de Z tenemos,

$$\begin{aligned} Y(z) &= X(z) + [a_1 z^{-1} + a_2 z^{-2}] Y(z) \\ &= X(z) \frac{1}{1 - a_1 z^{-1} - a_2 z^{-2}} \end{aligned} \tag{III.10}$$

y por lo tanto la función de transferencia es,

$$H(z) = \frac{Y(z)}{X(z)} = \frac{1}{1 - a_1 z^{-1} - a_2 z^{-2}} \tag{III.11}$$

Cuando la respuesta en fase del sistema es muy importante los filtros FIR son los mas indicados, es posible diseñarlos con fase lineal. Aunado a su fase lineal también los filtros FIR son estables y por lo tanto son los mas usados en sistemas que requieren un alto grado de control. En contraste, los filtros IIR tienen una respuesta en fase no lineal afectando a cada banda de frecuencias de manera diferente. En la mayoría de las aplicaciones que usan filtros IIR la prioridad es la búsqueda de gran calidad y una banda de transición muy pequeña con el menor número de etapas. Para un filtro IIR de un grado dado la eficiencia es del orden de 5 a 10 veces mayor que la de un filtro FIR.

Las razones antes mencionadas nos llevan fácilmente a pensar en los aspectos económicos de la puesta en marcha de un proyecto, y muy especialmente en el hardware y velocidad que se requieren para cumplir con ciertos parámetros que pueden cambiar la decisión entre usar un filtro FIR o uno IIR.

III.2 FILTRADO ADAPTABLE

Para el diseño de los filtros se pueden usar técnicas clásicas o modernas. La diferencia entre ellas radica en la utilización de propiedades estadísticas de las señales como un parámetro muy importante. En las técnicas modernas del filtrado se usan las características estadísticas tanto de las señales como del ruido y en la técnica clásica solo se hace uso de las características y contenido espectral del ruido y de las señales que nos interesan. Siendo las primeras técnicas las que nos dan la capacidad de satisfacer las características tan demandantes de los actuales sistemas de comunicaciones.

Ejemplos de las técnicas clásicas son el diseño de los filtros Butterworth y Chebyshev entre otros. Para que estos filtros funcionen bien es necesario que los

componentes espectrales de la señal y del ruido ocupen diferentes bandas en frecuencia. A principios de los años sesenta el diseño de filtros selectivos en la frecuencia fue hecho por medio de componentes analógicos como resistencias, inductores y capacitores; después, con el advenimiento de las computadoras digitales de propósito general creció el interés en la simulación de tales filtros, dando esto lugar a la evolución de los filtros digitales y su diseño en forma de programas de computadora y en circuitos.

En las técnicas modernas se ve tanto a la señal y al ruido como un proceso estocástico y se usan sus estadísticas para derivar una teoría del filtrado. El trabajo en esta área se inicia con Wiener en los Estados Unidos en 1942 y con Kolmogorov en la Unión Soviética en 1939. Wiener consideró el problema de la estimación de una señal continua descrita como un proceso estocástico el cual era degradado por ruido dando por resultado una ecuación integral conocida como Wiener-Hopf que es muy difícil de resolver mas que para algunos casos muy simples. Por otro lado Kolmogorov considero la optimización de un filtro lineal por medio de un proceso estocástico discreto. En general se llama filtrado de Wiener al filtrado derivado de las teorías de ambos científicos.

La teoría del filtrado de Kalman [Kal60], llamada así en honor al trabajo hecho por Kalman en 1960 soluciona un problema en la teoría de Wiener que consideraba que todos los procesos de interés eran estacionarios. La teoría del filtrado de Kalman es mas general y puede manejar situaciones en las que los problemas a resolver contienen señales no estacionarias intrínsecamente. La gran generalidad del filtrado de Kalman es obtenida mediante la adopción de una estructura variante en el tiempo para el filtrado óptimo de la cual se deriva el filtrado adaptable.

El primer trabajo en técnicas adaptables fue el hecho por Howells, Applebaum [How65] y colegas en la Compañía General Electric entre 1957 y 1960, años en los que diseñaron y construyeron un sistema para la cancelación de lóbulos secundarios en una antena; el sistema usaba una entrada de referencia proveniente de una antena auxiliar y un filtro adaptable de dos etapas. Posteriormente en la Universidad de Stanford en California durante el año de 1959 Widrow y Hoff desarrollaron el algoritmo del gradiente estocástico aproximado LMS (Least Mean Square) [Wid60] y el esquema de reconocimiento de patrones llamado Adaline (Adaptive linear threshold logic element) [Kof66]. Los trabajos continuaron en el Instituto de Aeronautica y Telemecanismos de Moscú, U.R.S.S., con Aizermann y colegas. Posteriormente en Gran Bretaña D. Gabor y colegas empezaron con el desarrollo de filtros adaptables [Gab60].

Para principios de los sesenta los trabajos y publicaciones en los sistemas adaptables se intensificaron de manera importante. En Francia en 1965, H. Mermoz publicó su disertación de doctorado titulada "Adaptive Filtering and Optimal Utilization of an Antenna", pero uno de los más importantes fue el de Lucky en los Laboratorios Bell con el desarrollo de un Módem de alta velocidad [Luc65] [Luc68] en el que las técnicas adaptables se usaron para la igualación del

canal, técnicas que mas tarde se usarían para la eliminación de ruido y cancelación de eco.

El primer sistema adaptable de cancelación de ruido propiamente fue desarrollado en 1965 en la Universidad de Stanford como parte de un trabajo final en el curso de Sistemas Adaptables. El proyecto consistía en la eliminación de la señal de 60 Hz que interfiere a la señal de un electrocardiograma. En esa misma década una serie de aplicaciones para los sistemas adaptables de cancelación de ruido fueron desarrollados, la eliminación de interferencias periódicas [Kau72] y la eliminación del eco en llamadas de larga distancia [Son67] por ejemplo.

Se puede decir que la aplicación de los sistemas adaptables es relativamente reciente, a pesar de que las bases formales existen desde principios de siglo. El auge en estas técnicas se ha dado en primer lugar por la necesidad que existe de su uso en los actuales sistemas de comunicaciones y por la posibilidad que dan los modernos DSP's y microcontroladores que hacen posible su implantación.

III.2.1 Características de un Sistema Adaptable

En general un sistema adaptable se parece a un ser vivo, reacciona de manera diferente dependiendo de las circunstancias en las que se encuentre, se adapta al medio. Los sistemas adaptables son no lineales, variantes en el tiempo y tienen las siguientes características:

- Se adaptan automáticamente para enfrentar los cambios del ambiente y los cambios requeridos por el sistema.
- Pueden ser entrenados para efectuar un filtrado específico y tomar decisiones en línea.
- Pueden extrapolar un modelo de comportamiento al tratar con nuevas situaciones después de haber sido entrenados sobre una cantidad finita de señales o un patrón frecuente de entrenamientos.
- Estos pueden corregirse asimismo. es decir, se pueden adaptar alrededor de cierta clase de defectos internos.
- Pueden ser descritos como un sistema no lineal con parámetros variables en el tiempo.
- Usualmente son más complejos que los sistemas no adaptables, pero ofrecen la posibilidad de un incremento en el desempeño del sistema, cuando las características de la señal de entrada son desconocidas o variables en el tiempo.

Una manera sencilla de clasificar a los sistemas adaptables como lo hizo Widrow [WID85] es como sistemas de lazo abierto o de lazo cerrado. Un sistema adaptable de lazo abierto toma en cuenta los valores de entrada como elementos que sirven para ejecutar un algoritmo que da como salida una serie de parámetros que sirven para ajustar al sistema, es decir, el criterio de adaptación se basa en

las características de la señal de entrada. Los sistemas de lazo cerrado, en cambio, utilizan la salida en un proceso de ajuste automático.

III.2.2 Sistemas de Lazo Cerrado

Debido a sus características los sistemas de lazo cerrado han sido los más utilizados y en la actualidad se utilizan en una gran variedad de aplicaciones.

Un sistema adaptable tiene en general la forma que se muestra en la figura II 2 2. La señal de entrada $x(n)$ se procesa para obtener una salida $y(n)$, la cual es restada de la señal deseada $d(n)$ para obtener un error $e(n)$ que se utiliza como entrada para el algoritmo adaptable.

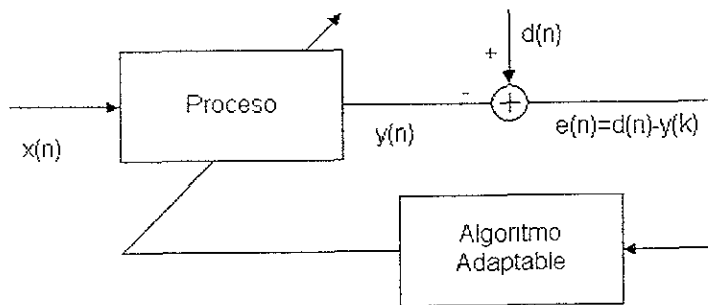


Fig. II.2.2 Sistema Adaptable

La estructura del sistema adaptable se ajusta mediante alguna medida de la señal de error de tal manera que se reduzca lo más posible. La predicción es una de las áreas más conocidas en las que se han utilizado los sistemas adaptables. El sistema tiene como señal de entrada a la señal deseada con un retraso de tal manera que la salida del algoritmo adaptable sea tal que pueda hacer tender a cero el error y de esta manera predecir la señal de entrada. La predicción es muy utilizada en sistemas de compresión y codificación de señales, así como en sistemas de reducción de ruido.

La identificación de canal ha sido otra área en la que los sistemas adaptables han dado buenos resultados. El sistema que se quiere identificar tiene como entrada la misma que el sistema adaptable y las salidas de ambos son las señales con las que se calcula el error. Una vez que el error es cero el sistema adaptable y el sistema a identificar tendrán la misma función de transferencia.

La aplicación que tiene mayor importancia para los fines de este trabajo es la cancelación de interferencias. En este sistema se dispone de dos entradas, una que se compone de la señal con ruido y otra que tiene ruido correlacionado de alguna manera con el ruido de la primera entrada. La finalidad del proceso adaptable es encontrar una señal lo mas parecida a la señal sin ruido.

III.2.3 Forma de los Filtros Adaptables

Dentro del filtrado adaptable existen tres estructuras de gran importancia y de gran uso. La estructura transversal que tiene la misma estructura que un filtro FIR. Otra estructura de mayor complejidad que es el filtro "lattice" o de cruz que toma su nombre de la semejanza que tiene con una rejilla y el filtro Lattice - Ladder. El filtro Lattice consiste de estructuras en cascada de dos entradas y dos salidas como se ve en la fig. II.2.3

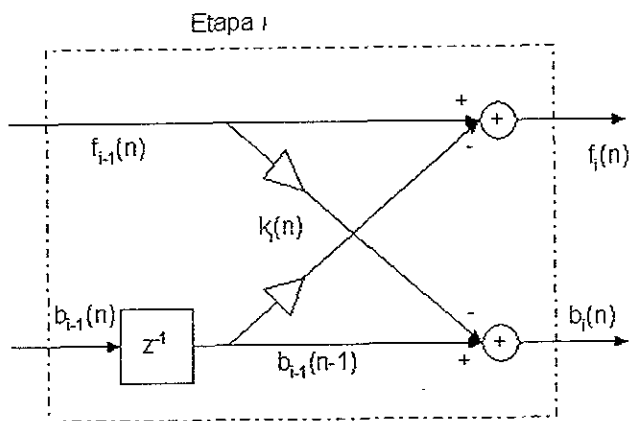


Fig II.2.3 Estructura Básica del Filtro Lattice o Filtro Cruz

Las ecuaciones que describen al filtro Lattice en realidad son muy sencillas.

$$f_i(n) = f_{i-1}(n) - k_i(n)b_{i-1}(n-1), \quad i = 1, 2, \dots, I-1 \quad (\text{III.12})$$

y (III.13)

$$b_i(n) = b_{i-1}(n-1) - k_i(n)f_{i-1}(n), \quad i = 1, 2, \dots, I-1$$

donde $f_i(n)$ y $b_i(n)$ son las señales de salida, $k_i(n)$ es el coeficiente de reflexión, i es el número de etapa y $I-1$ representa el número total de etapas. La señal de referencia $x(n)$ en la primera etapa se utiliza en ambas entradas con la que se

obtienen dos salidas dependiendo de la forma como cambian los coeficientes de reflexión.

III.2.4 Clasificación de los Algoritmos Adaptables

No existe una solución única para la resolución de los problemas del filtrado adaptable, en la actualidad existen muchos algoritmos. Cada algoritmo posee ciertas características que lo hacen muy bueno para una aplicación muy especializada. De la sección anterior sabemos que el objetivo del filtrado adaptable es la minimización del error para hacer una estimación óptima de los coeficientes. El criterio de minimización es un buen punto de partida para clasificar a los algoritmos adaptables.

III.2.4.1 Algoritmos Tipo Gradiente

Tienen como base principal las características estocásticas de las señales, principalmente la esperanza matemática del error cuadrático medio¹. La parte esencial del algoritmo es la de reducción del error mencionado de manera recursiva para ajustar los coeficientes del filtro.

Los algoritmos de pasos descendentes reducen la esperanza matemática del error cuadrático medio. El valor inicial es arbitrario, sin embargo, encuentra el mínimo en la superficie de error sin conocimiento previo y converge a la solución de Wiener como muchos otros algoritmos. Debido a que en la mayoría de los casos se desconocen los valores de entrada el algoritmo de pasos descendentes no puede ser utilizado. El algoritmo LMS resuelve mediante la utilización de gradientes instantáneos que no requieren la obtención de cuadrados, promedios o diferenciaciones el problema.

$$w(n-1) = w(n) + \mu x(n)e(n) \tag{III.14}$$

donde μ es el tamaño del paso, $w(0)$ es el vector de coeficientes y $e(n)$ es el error.

¹ El Error Cuadrático Medio definido por la siguiente ecuación, $\xi(n) = E[e^2(n)]$ representa una función cuadrática de los coeficientes que al graficarla se obtiene una superficie concava de forma hiperbólica. El error cuadrático medio también se puede representar en función de la matriz de autocorrelación R y de los coeficientes del filtro $w(n)$ y w^0

$$\xi(n) = \xi_{min} + [w(n) - w^0]^T R [w(n) - w^0]$$

III.2.4.2 Método de Mínimos Cuadrados

El algoritmo RLS (Recursive Least Squares) puede ser usado con un filtro transversal para dar mayor velocidad de convergencia y menor error que el LMS. Sin embargo, un inconveniente es el número de operaciones que se tienen que realizar, L^2 , donde L es el orden del filtro. El algoritmo FTF (Fast Transversal Filter), la versión mejorada del RLS reduce el número de operaciones a $7L$.

El método de mínimos cuadrados utiliza promedios en el tiempo n del error cuadrático o costo que consiste de una suma ponderada del error al cuadrado. En la derivación del algoritmo RLS se asume que el vector de coeficientes es constante durante un tiempo determinado.

$$\xi(n) = \sum_{i=1}^n \lambda^{n-i} e^2(i) \quad (\text{III.15})$$

Eliminación de Ruido

Dentro del presente capítulo hablaremos de las diferentes formas en las que se ha llevado a cabo la tarea de eliminar ruido, ya sea dentro de un ambiente de trabajo para evitar las molestias que este ocasiona, o el ruido que se desea eliminar para la posterior transmisión de una señal por un sistema de comunicaciones. Se hablará del Control Activo de Ruido (ANC) y la eliminación de ruido en el realce de voz.

IV.1 CONTROL ACTIVO DE RUIDO

La forma más fácil y tradicional de eliminar ruido es la utilización de barreras, como cuando se cierra una puerta para no dejar pasar el ruido exterior. Otra forma es la utilización de ductos y silenciadores para conducir el ruido y atenuarlo. Todas estas maneras de eliminar el ruido son técnicas pasivas que hacen uso de la pérdida de energía de una señal con el cambio de impedancia en un medio y la que sufre al propagarse por un medio especial. Si bien estas maneras de atenuar el ruido son muy eficientes tienen una limitante. La atenuación de los silenciadores pasivos se ve seriamente afectada cuando la longitud de onda del ruido es más grande que las dimensiones del silenciador.

Las técnicas del ANC han surgido como un buen complemento a las técnicas pasivas debido a su amplio campo de aplicación, se puede aplicar en problemas de acústicos en el aire o en agua, además de problemas de vibración. La forma en la que el ANC trabaja se basa en el principio de superposición y lo hace de la siguiente manera: un sensor primario recibe la señal a eliminar, la cual es procesada por medios electrónicos, acústicos o mecánicos produciéndose una señal de salida con igual amplitud y fase opuesta a la señal de entrada, de tal manera que al sumar tales señales se eliminan.

Al igual que las técnicas pasivas el Control Activo del Ruido también tiene limitantes. El ANC también está limitado en frecuencia, pero en este caso es en altas frecuencias; si hablamos de un proceso digital, el límite está en relación con la frecuencia máxima de muestreo de los convertidores analógico - digital y a la velocidad máxima del procesador, con lo que sí pensamos que las técnicas pasivas para poder atenuar señales de muy baja frecuencia requieren de elementos muy voluminosos y por ende costosos, y que el ANC es eficiente en bajas frecuencias podemos ver que son técnicas que se complementan.

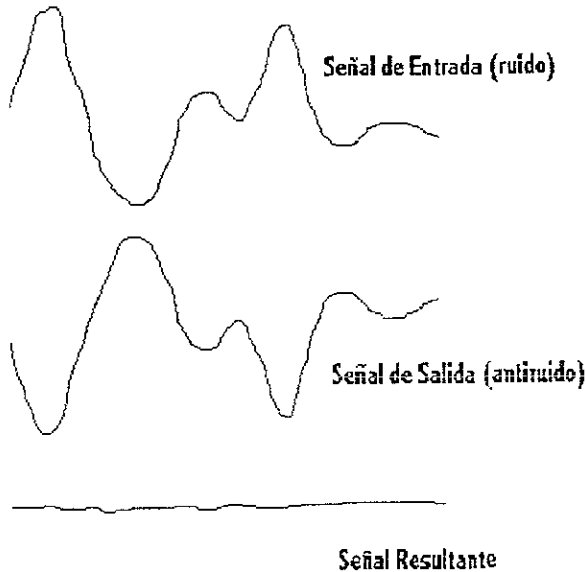


Fig. III.1.1 Principio de Operación del ANC

Dado que en la mayoría de los casos las características del ruido acústico son variantes frente al tiempo, la frecuencia, la fase, la amplitud y la velocidad del ruido son no estacionarias. Para poder trabajar con señales de estas características es necesario por lo tanto la utilización de técnicas adaptables en el ANC.

Aunado a la necesidad de las técnicas adaptables es necesario incluir al procesamiento digital. Para poder producir una señal (antiruido) que sea lo mas parecida a la señal original en su amplitud y que tenga la fase adecuada de tal manera que se puedan eliminar una a la otra se requiere de una precisión muy alta, la cual es más fácil de obtener digitalmente. Bien sabemos que el procesamiento digital entre sus ventajas tiene a la estabilidad y que se obtiene el mismo resultado a la misma entrada indefinidamente sin verse afectado por factores como la temperatura y el tiempo, siendo estos factores muy importantes en un sistema analógico, de ahí que sea deseable la utilización del procesamiento digital de señales para la obtención de un buen sistema de cancelación de ruido

La idea esencial del ANC no es nueva, fue descubierta en los años 30's, pero los medios para su desarrollo no se habían dado. La década de los años 80's y el desarrollo de los nuevos y poderosos DSP's hizo posible poner en marcha muchas ideas, no solo en la eliminación de ruido y el ANC, sino en el mundo del procesamiento de señales. Los nuevos DSP's fueron diseñados para realizar operaciones matemáticas a muy alta velocidad y de esta manera poder procesar señales digitales en tiempo real, y ya que tales dispositivos son programables fue posible la implantación de algoritmos adaptables a un bajo costo y darle actualidad

de nuevo a muchas técnicas que hasta el momento no se habían podido implementar.

IV.1.1 Aplicación del Control Activo de Ruido

Cuando usar una técnica del ANC y cuando no hacerlo depende únicamente de la eficiencia que se obtendría con este método en comparación con las técnicas pasivas. Utilizar el ANC en bajas frecuencias es muy adecuado, pero su utilización con señales de frecuencia alta es muy limitada debido a la necesidad de frecuencias de muestreo muy altas, velocidad de proceso y los modos que aparecen con señales de frecuencia muy alta, haciendo al ruido una señal muy complicada y difícil de procesar. Afortunadamente y como ya lo hemos mencionado las técnicas pasivas son muy eficientes en altas frecuencias y un muy buen complemento para el ANC.

Debido a las características y limitantes del ANC se ha utilizado en las siguientes áreas:

- *Automotriz*. En todo tipo de vehículos, de carga, militares y particulares. En sistemas de un solo canal y de dos canales con aplicaciones en sistemas silenciadores y en atenuación de ruido dentro de la cabina del operador entre otras.
- *Aparatos*. Sistemas de uno o dos canales con aplicación en sistemas de aire acondicionado, refrigeradores, lavadoras, limpiadores de vacío, etc.
- *Industria*. Ventiladores, ductos de aire, chimeneas, compresoras, bombas, en cabinas telefónicas, protectores para oídos, audífonos, etc.
- *Transporte*. Aviones, barcos, helicópteros, motocicletas, etc.

Dentro de todas estas áreas de aplicación nos interesa particularizar en las que tienen aplicación directa en la cancelación de ruido acústico en el aire, ya que son las que se relacionan directamente con el tema central de este trabajo. Para clasificar las aplicaciones del ANC dentro del ruido acústico en aire se puede hacer desde un punto de vista geométrico [War87]:

- *Ruido en Ductos*. Tuberías y Ductos usados en sistemas de calentamiento, ventilación, aire acondicionado, etc.
- *Ruido de interiores*. Ruido dentro de un espacio cerrado.
- *Ruido en espacios libres*. Ruido que se propaga en espacios abiertos.

- Protección personal para escuchar. Caso especial del ruido para interiores para dimensiones muy pequeñas

Cada uno de estos aspectos ha sido tratado especialmente por los investigadores, pero es el ruido en ductos el que más atención ha recibido debido a su importancia en la industria y potencial aplicación en edificios de oficinas, escuelas, salas de juntas y casas. La turbulencia que se genera por el rápido paso de los flujos de aire en curvaturas y esquinas dentro de una tubería o ducto genera ruido aleatorio de gran volumen donde se pueden aplicar técnicas de ANC con buenos resultados.

Para atenuar el ruido de interiores existen muchas maneras en las que es necesario saber si la fuente del ruido se encuentra dentro del área cerrada o fuera de esta, de manera que si esta dentro el sistema de cancelación de ruido se puede colocar cerca de la fuente teniendo especial cuidado en que no se tenga una realimentación que pueda afectar al desempeño del sistema en general. Cuando la fuente del ruido este fuera el problema se complica un poco ya que la distribución espacial del ruido puede tomar muchas formas, este es el caso del ruido dentro de un auto o dentro de la cabina de un helicóptero. Similarmente se puede tomar el caso del ruido en espacios abiertos, aquí tampoco sabemos el lugar exacto de donde proviene el ruido, de tal manera que es posible solamente la eliminación del ruido dentro de una zona delimitada. Es como si se formara una barrera en contra del ruido y así reducirlo para evitar la molestia a trabajadores en lugares especiales en una industria por ejemplo

Como un caso especial se puede tomar a la protección personal para escuchar debido a que el espacio en el que se desea eliminar el ruido es muy pequeño en comparación a la longitud de onda del rango de frecuencias de interés y a su directa aplicación a las comunicaciones. Pueden ser utilizadas en combinación con sistemas de comunicación para realzar la inteligibilidad en ambientes ruidosos con gran aplicación en ámbitos industriales, plantas manufactureras y en la milicia

IV.1.2 Sistemas de Control Activo de Ruido

Dependiendo del ancho de banda del ruido para el que están diseñados y de la forma en la que trabajan los sistemas, se pueden distinguir a los de control realimentado, control directo, de banda ancha y banda angosta. La turbulencia en un ducto que es totalmente aleatoria, por lo que su energía esta distribuida ampliamente a lo largo de la banda de frecuencias, es un ejemplo de ruido de banda ancha. Un compresor en un refrigerador, una bomba, una turbina y bombas de vacío que trabajan de forma rotatoria generan ruido que es casi periódico por lo que se concentra en una región muy estrecha de la banda de frecuencias, estos son ejemplos de ruidos de banda angosta. En este caso muy especial se han desarrollado técnicas muy específicas que están basadas en mediciones muy precisas mas que en valores estadísticos, en lugar de usar un sensor de referencia los ANC periódicos, como son llamados, utilizan un tacómetro para

medir la velocidad de rotación de la fuente de ruido y así generar la señal antiruido en la frecuencia básica de rotación y en las armónicas medidas

IV.1.2.1 Sistemas de Banda Ancha

Este tipo de sistemas se basan principalmente en la estructura que se muestra en la figura IV.1.2

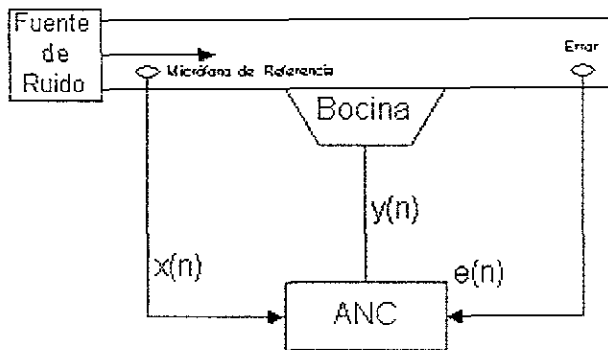


Fig. IV.1.2.1 Sistema ANC de banda Ancha de un solo Canal

El sistema trabaja de la siguiente manera, el primer micrófono (micrófono de referencia) mide el ruido en el ducto y envía la información al sistema ANC que genera un antiruido. El segundo micrófono capta la parte que no ha podido ser eliminada, el error, y envía la información como parte de la realimentación del filtro adaptable. Es importante mencionar que en este tipo de sistemas la distancia entre los elementos que la forman es muy importante.

El límite inferior de frecuencia que puede ser manejado por un sistema de este tipo depende de las fluctuaciones de presión provocadas por turbulencias: el límite superior, en cambio, está limitado por las dimensiones del ducto y por la velocidad de proceso del sistema.

IV.1.2.2 Sistemas de Banda Angosta

El ruido que se genera en la mayoría de la industria y que proviene de la maquinaria es ruido de banda angosta y se debe a la rotación de sus componentes. La manera de cancelar el ruido de banda angosta es muy diferente a la manera en como se hace para el ruido de banda ancha. Si se puede medir de alguna manera la frecuencia de rotación de la fuente del ruido se podrá generar

una señal a partir de esas mediciones Debido a la estreches del espectro del ruido y al clase de mediciones que se llevan a cabo la eliminación de ruido de banda angosta es mas eficiente acústicamente que la de ruido de banda ancha

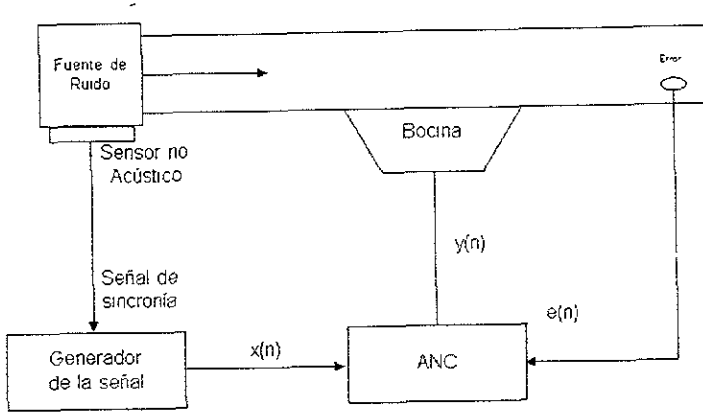


Fig. IV.1.2.2 Sistema de Control de ruido de Banda Angosta

IV.1.2.3 Sistemas con mas de un Canal

En lugares con una geometría muy complicada como por ejemplo una sala de juntas y un corredor dentro de un edificio, la aplicación de sistemas de un solo canal no tiene la suficiente eficiencia Debido a las múltiples trayectorias provocadas por la gran cantidad de fuentes y por la distribución de la sala es necesario utilizar mas de un canal para poder obtener buenos resultados. La posición de los micrófonos y de las bocinas es parte esencial para poder anular el complejo patrón de ruido que se genera.

IV.2 REALCE DE VOZ DEGRADADA POR RUIDO ADITIVO

Las maneras por las cuales se puede eliminar ruido y darle realce a una señal de voz son muchas, pero la mayoría de ellas utilizan alguna característica de la voz y del ruido para poder distinguirlos Dependiendo de la aplicación, el realce de voz puede incluir aspectos cualitativos como parte central (medición de la cantidad de ruido en la señal) o aspectos de percepción humana (como se escucha) ya que será a final de cuentas, en la mayoría de los casos, una persona quien decidirá cuando una señal tiene buena calidad.

Los primeros trabajos en esta área se remontan a los años 60. En ese año M.R. Schroeder [Sch60] fue el primero en utilizar el principio del Vocoder como elemento base para realzar la calidad de una señal de voz degradada por ruido aditivo. La idea principal fue la de generar una señal lo mas cercana a la señal original, pero con una envolvente que atenuara la señal entre los picos, es decir, la suavizara. Para 1964 Sievers y Sondhi [Sie64] retomaron la idea de Schroeder y con varias modificaciones hicieron una simulación. Aunque se observó que la idea era buena, la calidad de la voz que se obtuvo no lo fue

A través del tiempo muchas variantes de la idea básica de Schroeder han sido propuestas por varios investigadores. Weiss, Aschkenasy y Parsons [Wei74]; Boll [Bol79], McAulay y Malpass [Mca80], Ephraim y Malah [Eph84]; y Lim y Oppenheim [Lim79], pero todos tienen algo en común, la señal ruidosa se separa en varias regiones (canales) en la frecuencia mediante la utilización de un banco de filtros y posteriormente cada canal de la señal se atenúa por un factor que depende de la relación señal a ruido medida en dicha banda. La diferencia mas significativa entre todos los métodos es la manera de medir el nivel de ruido y de señal en cada canal.

Un método que es un poco diferente fue propuesto por Ephraim, Malah y Juang [Eph89] que aunque se puede clasificar también dentro de la misma categoría, tiene como característica particular el utilizar Modelos de Markov (HMM) para seleccionar la atenuación de las diferentes frecuencias en el ruido y en la señal de voz.

IV.2.1 Clasificación de los Métodos de Realce de Voz Degradada por Ruido Aditivo

De la misma manera que para el área de realce de voz existen múltiples maneras de clasificación, la parte que se encarga de la voz degradada por ruido también se puede dividir de muchas formas. Una manera de clasificación es una división basada en la forma de modelar la voz, podemos hablar de modelos estocásticos y de modelos de percepción humana. Si nos basamos en aspectos de percepción del hombre es posible dividir al realce de voz degradada con ruido en dos áreas básicas. La primer área toma como base la confusión que se presenta entre los sonidos fricativos y plosivos (ver Apéndice A) cuando se pierde la pausa existente entre los sonidos plosivos por el enmascaramiento que provoca el ruido. Cuando se presenta este tipo de confusiones la inteligibilidad de una palabra se reduce drásticamente. Por otra parte existen algoritmos en los que toda la atención se centra en las consonantes, ya que realzarlas puede dar mejor inteligibilidad a las palabras como ya hemos visto. Por ejemplo, si leemos la frase "Procesamiento de Señales" y se diera el caso en el que solo dispusiéramos de las vocales se leería " -o-e-a-ie-o -e -e-a-e-", pero si solamente contáramos con las consonantes leeríamos " Pr-c-s-m-nt- d- S-ñ-l-s". La última frase es mas comprensible y con la

ayuda de la capacidad del cerebro humano de completar espacios faltantes podríamos entender lo que se dice.

Otra manera muy importante de clasificar a los algoritmos de realce de voz es dependiendo del número de canales de entrada que se tenga. Se pueden tener sistemas que utilizan uno, dos o mas canales de entrada. Para sistemas que se basan en un solo canal se dispone de un solo micrófono, por lo que las características estadísticas del ruido se deben de medir durante los periodos de silencio de la conversación. En estos sistemas se asume al ruido como un proceso estacionario (ver Apéndice B). En los algoritmos basados en la utilización de dos canales, el sonido llega a cada entrada con una pequeña diferencia en tiempo (una de las entradas es normalmente la versión retrasada de la otra). Los métodos de dos o mas entradas se basan en dos hechos. El primer hecho es asumir que en el primer canal se tiene la señal de voz contaminada y en el segundo canal se tiene solamente ruido que esta correlacionado de alguna manera con el ruido en el primer canal. Normalmente se dispone de una barrera física entre los dos canales que evita que la señal de voz pueda ser recibida en ambos canales o se utilizan sensores muy directivos que forman haces evitando la mezcla de la señal y la voz.

Otra manera de clasificación no tan generalizada divide a los algoritmos en cuatro clases que difieren substancialmente en la forma de asumir el problema. La primera clase se basa en algoritmos que trabajan en el dominio de la Frecuencia (Análisis Espectral en Tiempo Corto). Esta técnica se basa en la supresión del ruido por medio de la resta del espectro del ruido estimado durante los periodos de silencio en sistemas de un solo canal, o bien el que se obtiene de un micrófono de referencia en un sistema de dos canales. La segunda clase de técnicas de realce se basan en el modelado de la voz y utilización de métodos iterativos. Estos sistemas se enfocan en la estimación de parámetros que puedan modelar a la voz, seguidamente de la síntesis de la señal ya libre de ruido mediante el filtrado de Wiener. En esta clase de métodos se requiere del conocimiento a priori de las características estadísticas tanto del ruido como de la voz y generalmente se obtienen esquemas iterativos. La tercera clase se basa en sistemas adaptables de cancelación de ruido. Tradicionalmente este tipo de sistemas se implementa usando un sistema de dos canales, ya sea en el dominio del tiempo o en el dominio de la frecuencia basado en algoritmos LMS (Least Mean Square). La última clase de realce se basa en modelos que consideran a la voz como periódica. Estos métodos emplean el seguimiento de la frecuencia fundamental en la voz mediante un sistema de un solo canal, o mediante filtros Comb adaptables a la magnitud del espectro armónico.

IV.2.2 Técnicas de Amplitud Espectral

Estas técnicas intentan realzar la voz degrada por ruido mediante la resta de una estimación del ruido. La manera en la que se lleve a cabo la resta nos lleva a varias maneras de formular el problema. Probablemente la más antigua y la mas sencilla debido a su sencillez en la implantación es la resta espectral que abarca

una serie de alternativas que difieren en los procesos anteriores y posteriores a la resta espectral. La diferencia puede ser la existencia o no, de coeficientes de peso pasados en la frecuencia o en la probabilidad de la voz y en la utilización de la rectificación o de un promedio al final del proceso

Todo el procesamiento en esta clase de métodos se lleva a cabo en el dominio de la frecuencia y trabaja directamente sobre la amplitud de la señal. Específicamente, el procesamiento trabaja con tramas de la señal original que son transformadas del dominio del tiempo al de la frecuencia, posteriormente se resta la magnitud estimada del ruido a la magnitud de la señal original y por último se regresa la señal procesada al dominio del tiempo.

IV.2.2.1 Resta Espectral

Dentro de este tipo de procesamiento se modela a la voz como un proceso aleatorio al cual se le ha agregado ruido aleatorio decorrelacionado con ella. Se supone al ruido como un proceso estacionario (ver Apéndice B) del cual se calculan sus características estadísticas de segundo orden durante los periodos de silencio en sistemas de un solo canal, o a partir de sensor de referencia en sistemas de más de un canal. La resta espectral es capaz de reducir los efectos del ruido mediante la resta de la media espectral estimada del ruido del espectro de la señal recibida.

Si x es la voz, y es la voz degradada y n es el ruido, la señal que se tiene antes del proceso se puede expresar como,

$$y(t) = x(t) + n(t), \tag{IV.1}$$

la resta espectral se puede definir de la siguiente manera,

$$|\hat{X}(f)|^b = |\hat{Y}(f)|^b - \alpha |\hat{N}(f)|^b \tag{IV.2}$$

donde $|\hat{X}(f)|^b$ es una estimación de la señal original $|X(f)|$, y $|\hat{N}(f)|^b$ es el espectro medio estimado del ruido. Para una resta de la magnitud $b=1$ y para resta de potencias $b=2$. El parámetro α controla la cantidad de ruido que se elimina de la señal. Como ya hemos mencionado, la media espectral del ruido se obtiene durante los periodos de silencio de la siguiente manera,

$$\overline{|N(f)|^b} = \frac{1}{k} \sum_{i=0}^{k-1} |N_i(f)|^b \quad (\text{IV.3})$$

donde $|N_i(f)|$ es el espectro de la i -ésima trama suponiendo de se tienen k tramas.

Para el realce de la señal en el dominio del tiempo la magnitud espectral $|\hat{X}(f)|$ es combinada con la fase de la señal original con ruido, $\theta_y(k)$, y ahora si, la señal se regresa la dominio del tiempo. Si bien, sabemos que el ruido no solamente afecta a la amplitud, sabemos también que afecta a la fase, pero evaluaciones en aspectos de percepción han demostrado que el oído humano es insensible al cambio de fase.

Debido a la variación del espectro del ruido, la resta espectral puede producir valores de potencia o magnitud espectral de signo negativo. Para evitar este problema algunos algoritmos utilizan una rectificación de media onda o de onda completa y otros procesan el resultado de la resta con una función de mapeo de la forma siguiente:

$$T\left[|\hat{X}(f)|\right] = \begin{cases} |\hat{X}(f)| & \text{si } |\hat{X}(f)| > \beta|Y(f)| \\ \beta|Y(f)| & \text{cualquier otro caso} \end{cases} \quad (\text{IV.4})$$

Valores típicos para β están desde 0.01 hasta 0.1, con lo que se obtiene una atenuación máxima de 10 a 20 dB en la potencia espectral.

IV.2.3 Filtrado de Wiener

EL filtro de Wiener es capaz de mejorar los resultados que se obtienen con la resta espectral al incorporar las características estadísticas no solamente del ruido, sino que utiliza también las de la señal original, es decir, de la señal sin ruido, por lo tanto podemos ver, que el filtro de Wiener no es un filtro adaptable a pesar de que la mayoría de ellos recaigan en su teoría. El filtrado de Wiener realiza dos suposiciones acerca de la señal y el ruido. Primero, supone que tanto la señal como el ruido son procesos estacionarios (ver Apéndice B) y segundo, que se conoce con anterioridad el espectro de ambas.

El error cuadrático medio en tiempo discreto para la reducción de ruido del filtro de Wiener se puede expresar de la siguiente manera,

$$w = (R_{xx} + R_{nn})^{-1} r_{xy} \quad (\text{IV.5})$$

donde R_{xx} y R_{nn} son las matrices de autocorrelación de la señal libre de ruido y del ruido respectivamente; y r_{xy} es el vector de croscorrelación de la señal con ruido y sin ruido. Transformando la ecuación anterior obtenemos el filtro de Wiener en el dominio de la frecuencia, que como se puede ver en el apéndice B la transformada de Fourier de la autocorrelación nos da como resultado la densidad espectral de potencia.

$$W(f) = \frac{P_{XY}(f)}{P_{XX}(f) + P_{NN}(f)} \quad (\text{IV.6})$$

dividiendo el numerador y el denominador por $P_{NN}(f)$, y restandole la relación señal a ruido $SNR(f) = \frac{P_{XX}(f)}{P_{NN}(f)}$ obtenemos la ganancia del filtro en la frecuencia en términos de SNR.

$$W(f) = \frac{SNR(f)}{SNR(f) + 1} \quad (\text{IV.8})$$

La respuesta en frecuencia del filtro de Wiener es un número real y positivo en el rango de $0 \leq W(f) \leq 1$. Cuando se tiene una relación señal a ruido muy alta $SNR(f) = \alpha$, o cuando es muy baja $SNR(f) = 0$, el filtro de Wiener tiene como resultados respectivamente $W(f) \approx 1$ y $W(f) \approx 0$, esto quiere decir que para ruido aditivo, el filtro de Wiener atenúa cada frecuencia proporcionalmente a la relación señal a ruido.

IV.2.4 Filtro Comb

El ruido puede tener muchas formas, y en ocasiones la voz puede ser degradada por procesos periódicos, por lo tanto el espectro del ruido también es periódico. Una forma de atacar este problema es utilizando filtros Comb adaptables.

En su forma mas simple un filtro Comb puede verse como un filtro que contiene uno o mas nulos en su respuesta espectral y que ocurren periódicamente a lo largo de toda la banda de frecuencias. Una aplicación clásica para esta clase de

filtros es la eliminación de la interferencia que ocasiona la línea de energía, el filtro Comb puede eliminar la componente principal de 60 Hz y sus armónicas.

Un filtro comb puede ser implantado a través de la estructura de un filtro FIR con la siguiente ecuación:

$$y(n) = \frac{1}{M+1} \sum_{k=0}^M x(n-k) \quad (IV.9)$$

y en el dominio de Z,

$$H(z) = \frac{1}{M+1} \frac{[1 - z^{-(M+1)}]}{(1 - z^{-1})} \quad (IV.10)$$

donde la respuesta en frecuencia esta dada por

$$H(w) = \frac{e^{-j\frac{wM}{2}} \operatorname{sen} w \left(\frac{M+1}{2} \right)}{M+1 \operatorname{sen} \left(\frac{w}{2} \right)} \quad (IV.11)$$

De la ecuación en el dominio de Z podemos ver que el filtro tiene los ceros en el círculo unitario en las posiciones

$$z = e^{j2\pi k(M+1)} \quad k = 1, 2, 3, \dots, M \quad (IV.12)$$

Es necesario notar que el polo en $z=1$ se anula con el cero que se encuentra en la misma posición, por lo tanto este filtro no tiene polos fuera de $z = 0$.

IV.2.5 Realce de voz Mediante la Función de Coherencia

IV.2.5.1 La Función de Coherencia

Como ya hemos visto, existen procesos de realce de voz que hacen su trabajo en el dominio del tiempo y otros que lo hacen en el dominio de la frecuencia, la función de coherencia es una medición que se hace en frecuencia. A la función de coherencia se le puede ver como un coeficiente de correlación en la frecuencia y se define como el valor normalizado de la densidad espectral de potencia entre dos o mas señales estacionarias en sentido amplio (ver Apéndice B),

$$\rho_{x_1 x_2}(f) = \frac{\gamma_{x_1 x_2}(f)}{\sqrt{\gamma_{x_1 x_1}(f) \gamma_{x_2 x_2}(f)}} \quad (IV.12)$$

donde $\gamma_{x_1 x_2}(f)$, es la densidad espectral de potencia cruzada y $\gamma_{x_1}(f)$ y $\gamma_{x_2}(f)$ son las densidad espectral de potencia de las señales. La densidad espectral de potencia de una señal es la transformada de Fourier de la función de correlación definida por,

$$r_{x_1 x_2}(k) = E[x_1(n)x_2(n-k)] \tag{IV.14}$$

Ya que trabajamos con una función normalizada, podemos observar que los valores que se obtienen al evaluar la magnitud de la función de coherencia al cuadrado (MSC) están dentro de los siguientes límites,

$$0 \leq |\rho_{x_1 x_2}(f)|^2 \leq 1 \tag{IV.15}$$

a lo largo de todas las frecuencias que componen a las señales procesadas.

De las ecuaciones anteriores nos podemos dar cuenta que la función de coherencia representa a la cantidad de energía que tienen en común las dos señales y que se relaciona a través de un filtro lineal.

IV.2.5.1 Reducción de Ruido

Primero, se considera que tenemos dos entradas; cada una compuesta de señal y de ruido, $x_i = s_i + n_i$, ($i = 1,2$) y se intenta obtener la señal \hat{s} , como se muestra en la figura IV.2.5.1.

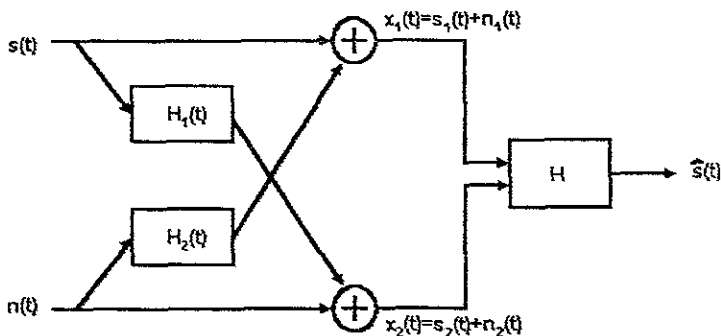


Fig. IV.2.5.2.1 Se suponen como entrada dos entradas $x_1(t)$ y $x_2(t)$ que se forman de señal y de ruido decorrelacionado temporalmente.

Dependiendo de la manera en la que se obtengan las señales es posible decir que las señales de voz están fuertemente correlacionadas en la banda de 0 a 4 kHz y,

dependiendo principalmente de la distancia entre los dos micrófonos y de la frecuencia, la correlación entre el ruido es muy pequeña. Para dos micrófonos separados 40 cm. uno del otro la coherencia es de 0.7 para frecuencias menores 700 Hz y decrece rápidamente para las frecuencias superiores¹

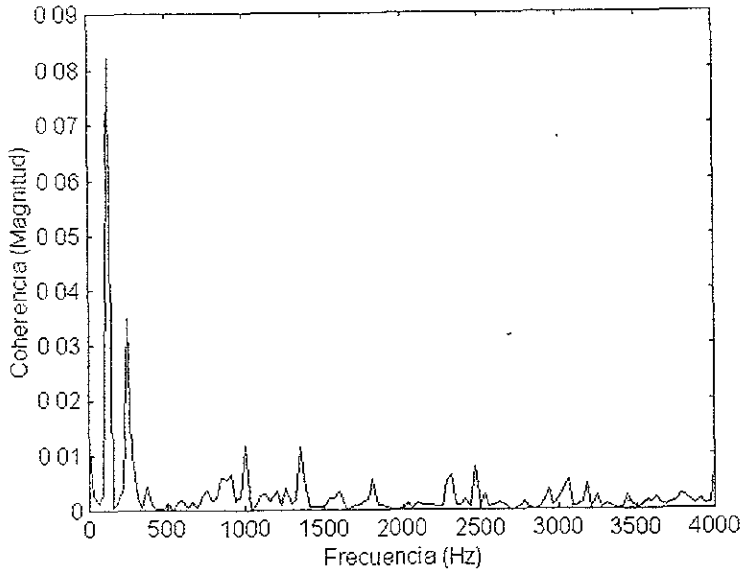


Fig. IV.2.5.2.2 Magnitud de la función de Coherencia del ruido. El ruido fue grabado en la calle a una distancia de aproximadamente 35 cm.

Suponiendo que la señal y el ruido son independientes, que el ruido esta decorrelacionado espacialmente y que el modulo de la función de coherencia de la señal es muy cercano a uno, podemos utilizar a la función de coherencia como un parámetro para discriminar la existencia de señal y de ruido en la señal. El proceso trabaja separando a las señales de entrada en tramas que son filtradas con el uso de una ventana de Hamming o Hanning por ejemplo. Posteriormente con la ayuda de la transformada de Fourier se obtienen el espectro y se calcula la función de coherencia que nos sirve como parámetro en el filtrado final de la señal.

¹ R. Le Bouquin y G. Faucon "Using de coherence function for noise reduction." IEE Proceedings-I. Vol. 139. No. 3. enero 1992.

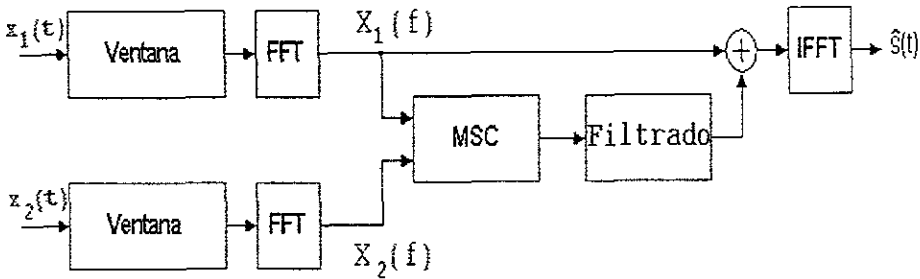


Fig. IV.2.5.2.3 Diagrama general del método de realce de voz utilizando la función de coherencia

La función de pesos se define de la siguiente manera:

$$\begin{aligned} \hat{S}_1(f) &= (MSC(f))^\delta X_1(f), & S_{min} < |\rho(f)| < S_{max} \\ \hat{S}_1(f) &= X_1(f), & |\rho(f)| \geq S_{max} \\ \hat{S}_1(f) &= (S_{min})^{2\delta} X_1(f), & |\rho(f)| \leq S_{min} \end{aligned} \quad (IV.16)$$

ya que suponemos que si el valor de la función de coherencia es muy cercano a uno lo que predomina es la señal y por lo tanto se deja pasar sin filtrado, y si el valor es muy cercano a cero se supone que lo que predomina es el ruido y por lo tanto se minimiza.

Al respecto del filtrado, considerando a $\gamma_{s_1}, \gamma_{s_2}, \gamma_{n_1}$ y γ_{n_2} las densidades espectrales de s_1, s_2, n_1 y n_2 respectivamente y, $\gamma_{s_1 s_2}$ y $\gamma_{n_1 n_2}$ las densidades espectrales cruzadas, podemos expresar la función de coherencia como,

$$\rho(f) = \frac{\gamma_{s_1 s_2}(f) + \gamma_{n_1 n_2}(f)}{\sqrt{[\gamma_{s_1}(f) + \gamma_{n_1}(f)]^2} \sqrt{[\gamma_{s_2}(f) + \gamma_{n_2}(f)]^2}} \quad (IV.17)$$

en el caso que nos interesa y con las suposiciones hechas la correlación entre el ruido es cero y la coherencia para la señal es muy cercana a uno, es decir que en el caso ideal las dos señales son la misma, podemos escribir por lo tanto,

$$\rho(f) \approx \frac{\gamma_{s_1}(f)}{\sqrt{[\gamma_{s_1}(f) + \gamma_{n_1}(f)]^2} \sqrt{[\gamma_{s_1}(f) + \gamma_{n_2}(f)]^2}} \quad (IV.18)$$

Ahora, si comparamos esta ecuación con el filtrado de Wiener si pensamos que queremos obtener s_1 a partir de x_1 , obtenemos la ecuación siguiente,

$$h(f) = \frac{\gamma_{s_1}(f)}{\gamma_{s_1}(f) + \gamma_{n_1}(f)} \quad (\text{IV.19})$$

si el ruido n_1 y n_2 tienen la misma densidad espectral $\rho(f) \approx h(f)$. Esto quiere decir que si tenemos una entrada x_1 , que pasa por un filtro con una función de transferencia $\rho(f)$, lo que obtenemos es un filtrado de Wiener.

Resultados y Conclusiones

A lo largo de este capítulo se presentaran los detalles de la implantación del algoritmo en MATLAB y en Ensamblador del TMS320C542, así como los resultados obtenidos. En general el proceso en la mayoría de los experimentos fue tomar las señales degradadas, procesarlas y observar los resultado tanto auditivamente como visualmente.

V.1 MATERIAL DE PRUEBA

Las señales de ruido que se utilizaron como material de prueba del algoritmo fueron grabadas por medio de una grabadora portátil en calles y avenidas comunes de la ciudad, en la mayoría de los casos en lugares con tráfico moderado. Además se obtuvieron archivos de Internet, en especial archivos de ruido en automóviles [Internet 1,2]. Las señales de voz en su mayoría se grabaron con la ayuda de una computadora personal de tal manera que se dispusiera de la señal sin ruido para poder llevar a cabo algunas mediciones de la eficiencia del algoritmo. Como complemento, se hicieron dos grabaciones en ambiente real con los micrófonos separados 40 cm. aproximadamente.

V.2 PRUEBAS CON EL PROGRAMA

Para realizar las pruebas en Matlab se programó el algoritmo y un ambiente de ventanas que nos permitiera interactuar. La interfaz con el usuario permite cambiar los parámetros y observar los resultados en el dominio del tiempo y en el dominio de la frecuencia con el espectrograma, además permite escuchar los resultados. La figura V.1 presenta una pantalla clásica del programa.

Los primeros experimentos se realizaron con la palabra "siete", se agrego ruido a la señal con la ayuda de la función "rand()" de Matlab y se obtuvieron los resultados que se muestran en la figura V.2 en el tiempo y V.3 en la frecuencia. Para el proceso se utilizó una ventana de Hanning de 256 puntos, $\max=0.81$, $\min=0.01$ y $\delta=2$.

Las figuras V.4 y V.5 presentan los resultados que se obtuvieron para la palabra "nueve" con los mismos parámetros y con ruido grabado en la calle. El espectro del ruido se presenta en la figura V.6. Es notoria la diferencia en los

resultados, se puede observar una gran mejoría, debido a que el espectro del ruido grabado en la calle esta distribuido de manera diferente.

En la figura V.7 se puede observar el efecto de alargamiento de los formantes que se presenta cuando se aumenta el tamaño de la ventana. En esta figura se presenta la palabrea "nueve" con los mismos parámetros y una ventana de 1024. Y en la figura V.8 como afecta el valor de δ en el filtrado, entre mayor sea el valor mas rígido es el filtrado. A pesar de que la forma de onda aún se parece a la original, el sonido se ha degradado tanto que es imposible reconocerla.

Finalmente, la figura V.9 muestra los resultados obtenidos con el ruido de un auto. Ya que en ambas entradas se tiene un ruido con una frecuencia en especial muy marcada (sonido ciclico del motor) el filtrado no la puede eliminar.

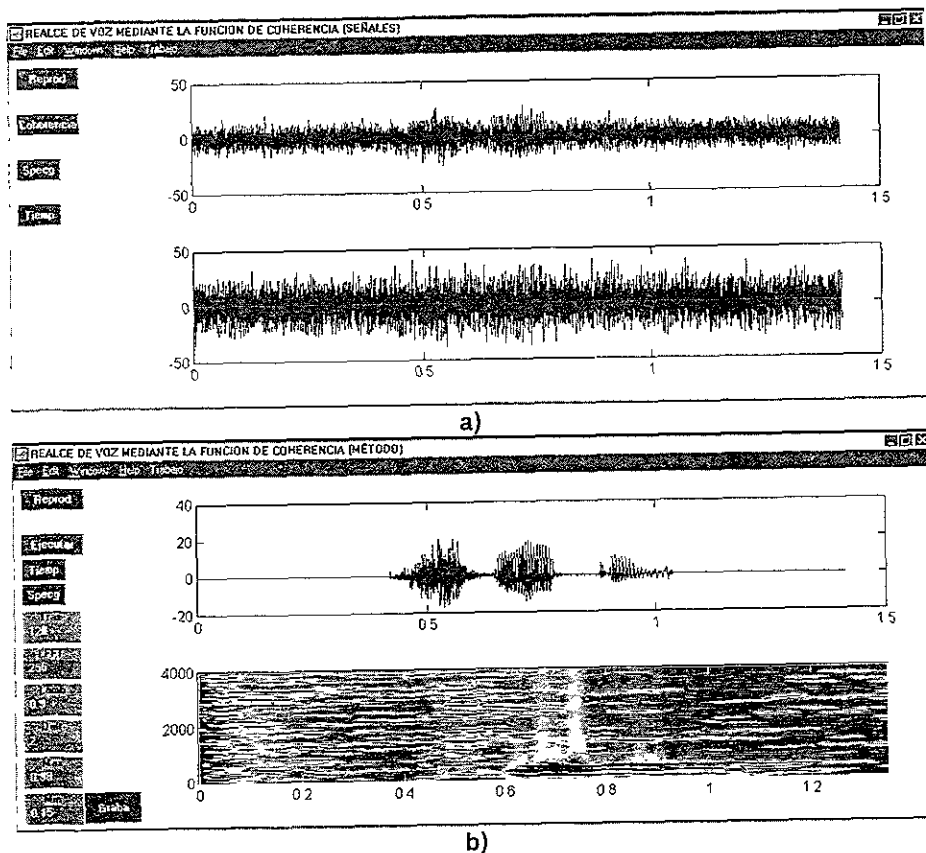


Fig. V.1 Pantallas del Programa en MATLAB. a) Dos señales de entrada. b) Señal original y Espectrograma de la señal procesada.

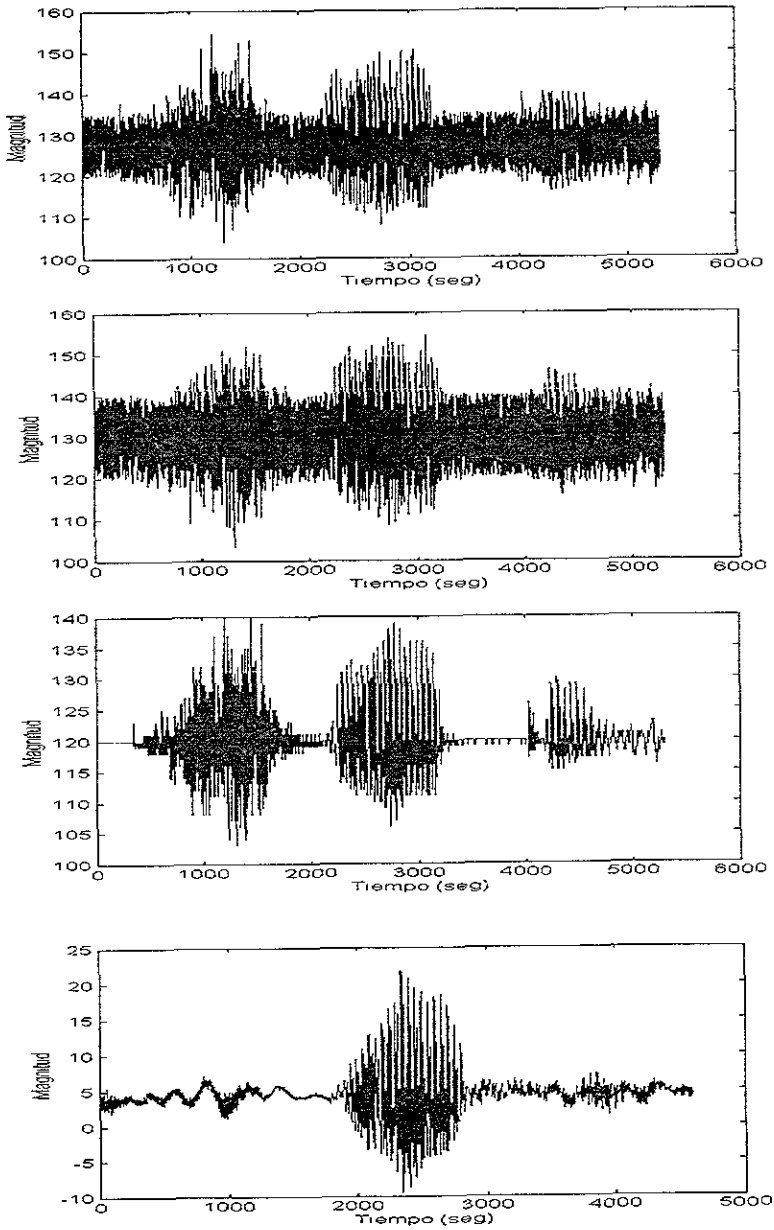


Fig.V.2 Experimento con la palabra "siete". a) y b) son las señales de entrada; c) la señal original y d) la señal procesada. . Los parámetros de entrada son: $\max=0.81$, $\min=0.01$ y $\delta=2.1$.

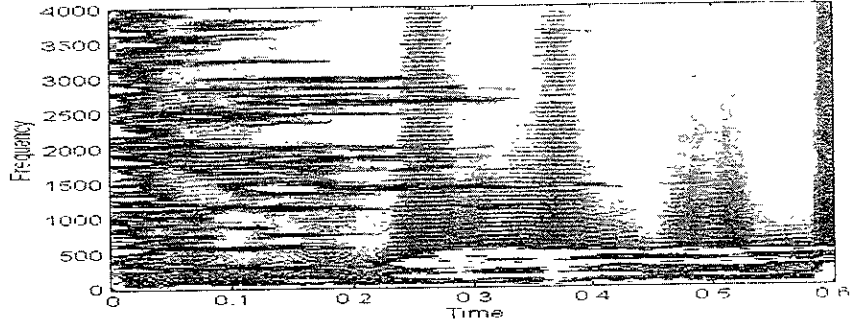
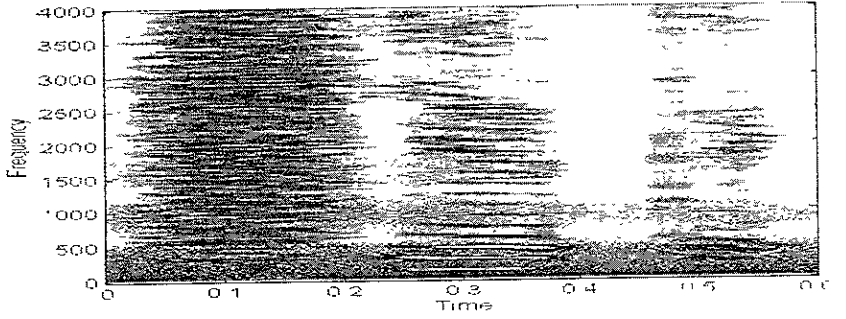
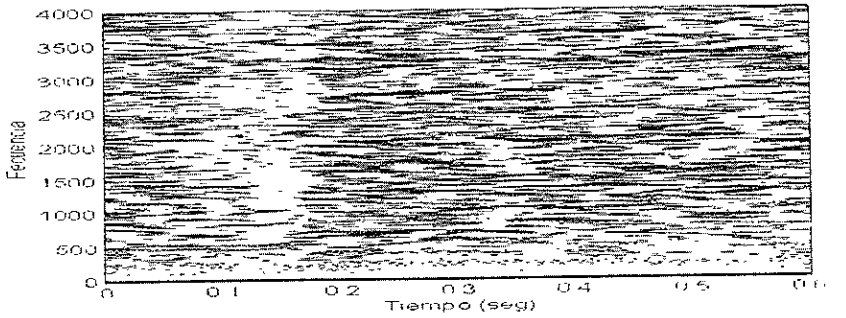


Fig.V.3 Espectrogramas de la palabra "siete". a) y b) son las señales de entrada; c) la señal original y d) la señal procesada. . Los parámetros de entrada son: $\max=0.81$, $\min=0.01$ y $\delta=2.1$.

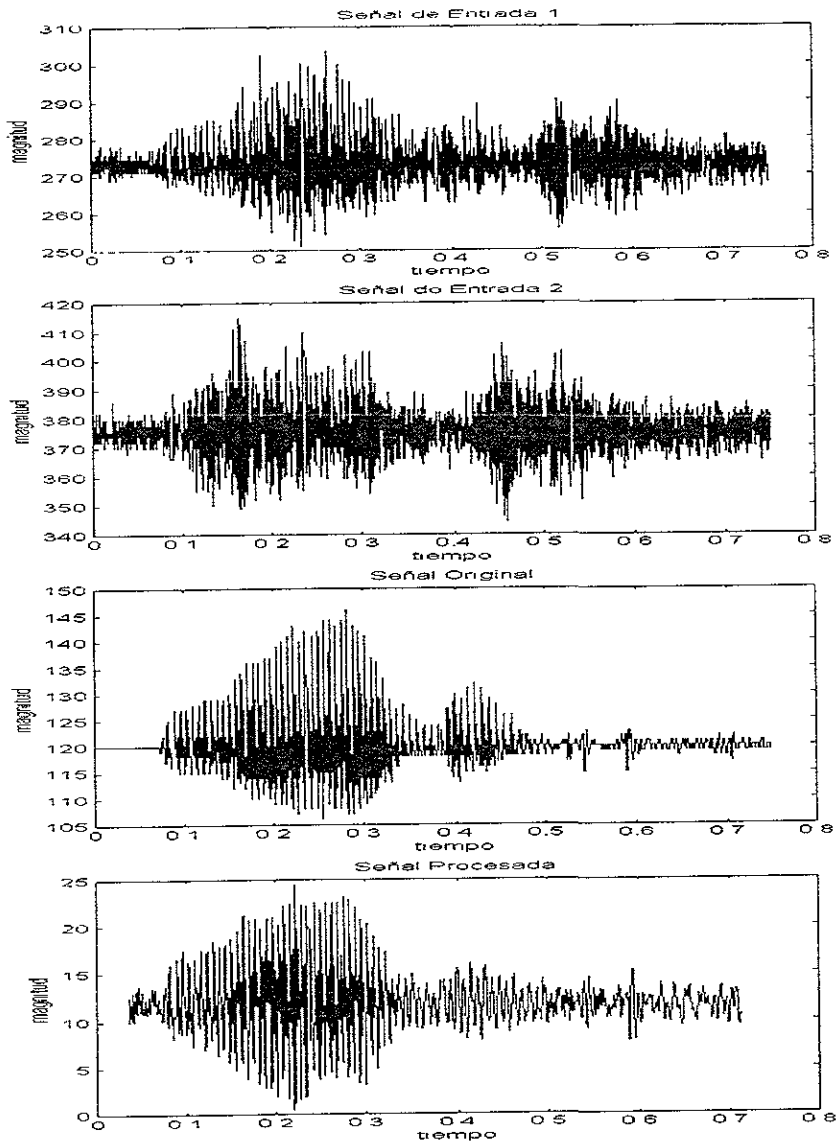


Fig.V.4 Experimento con la palabra "nueve" en el dominio del tiempo. a) y b) son las señales de entrada; c) la señal original y d) la señal procesada. Los parámetros de entrada son: $\max=0.81$, $\min=0.01$ y $\delta=2.1$.

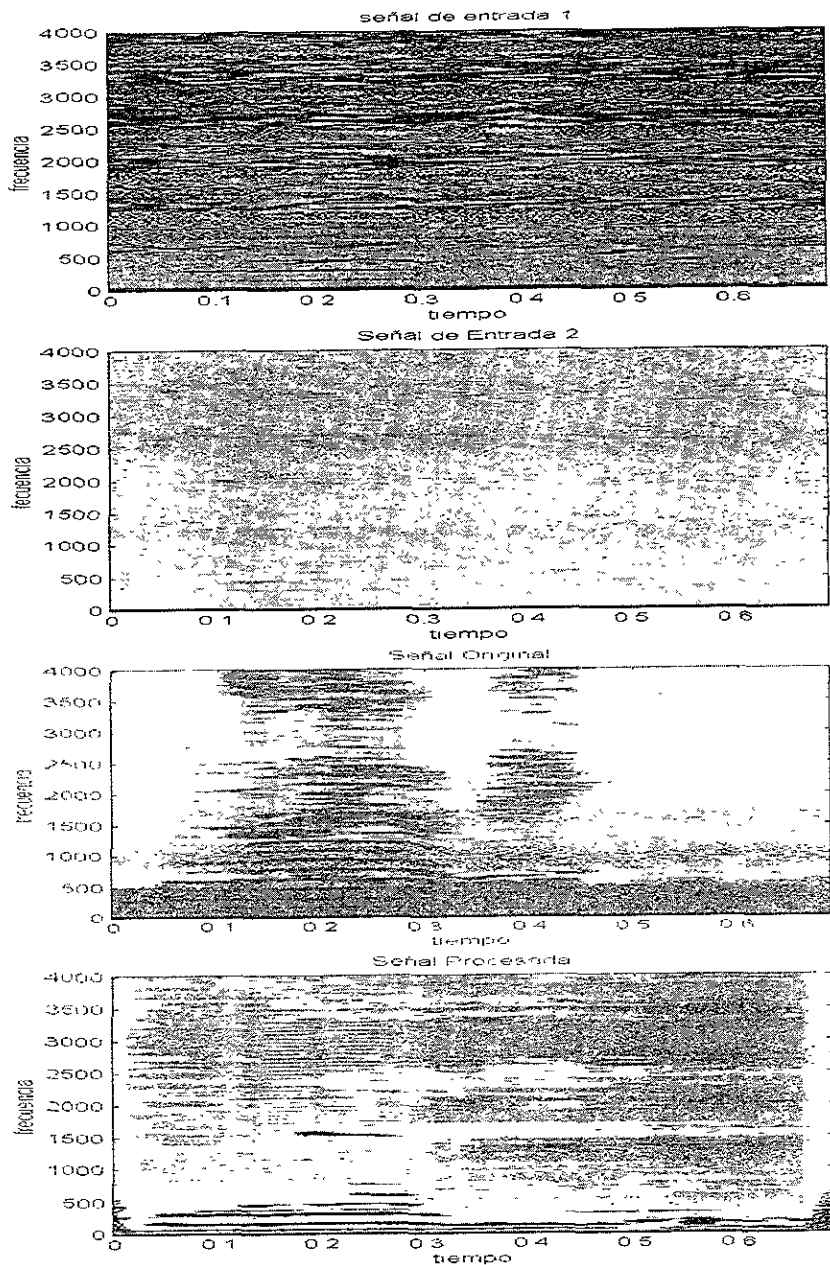


Fig.V.5 Experimento con la palabra “nueve” en el dominio de la frecuencia. a) y b) son las señales de entrada; c) la señal original y d) la señal procesada. Los parámetros de entrada son: $\max=0.81$, $\min=0.01$ y $\delta=2.1$.

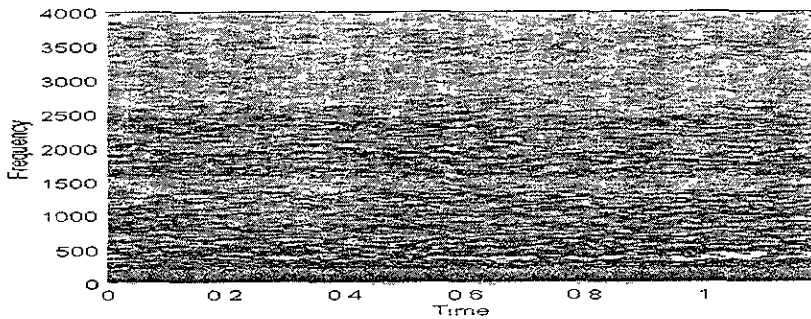


Fig.V.6 Espectrograma del ruido grabado en la calle.

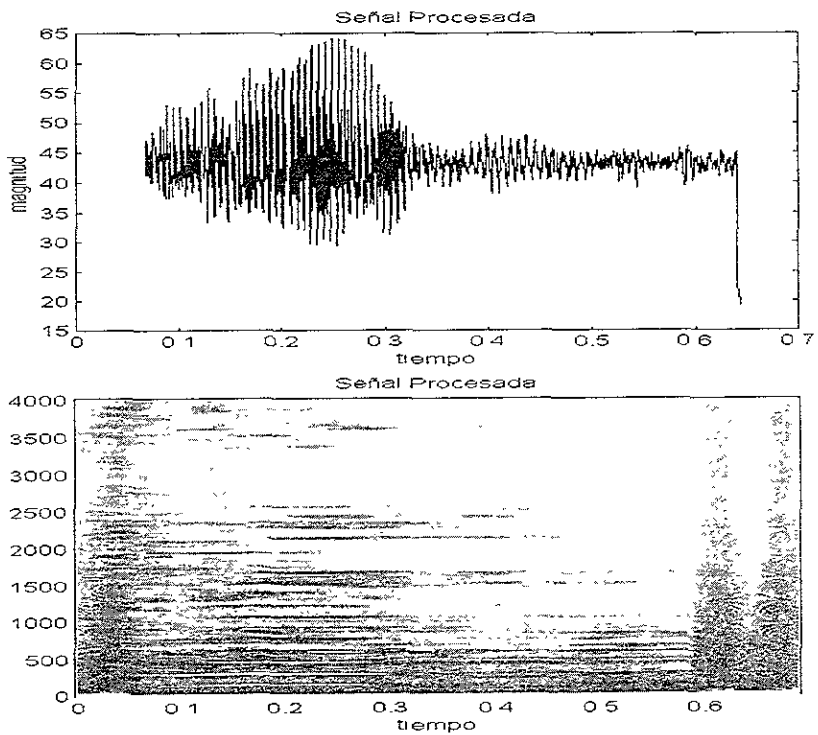


Fig.V.7 Resultados para la palabra "nueve" con una ventana de 1024 puntos.

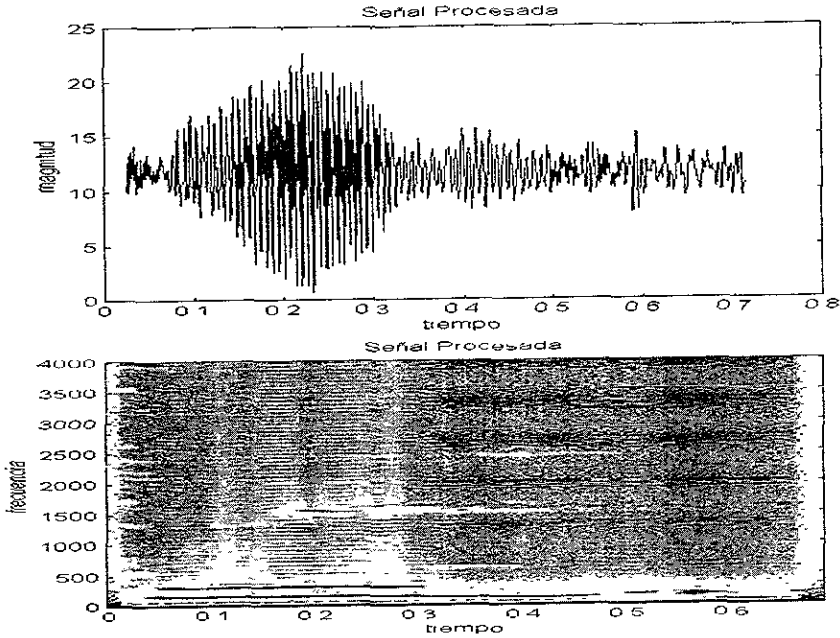


Fig.V.8 Resultados para la palabra "nueve" con una $\delta = 6$.

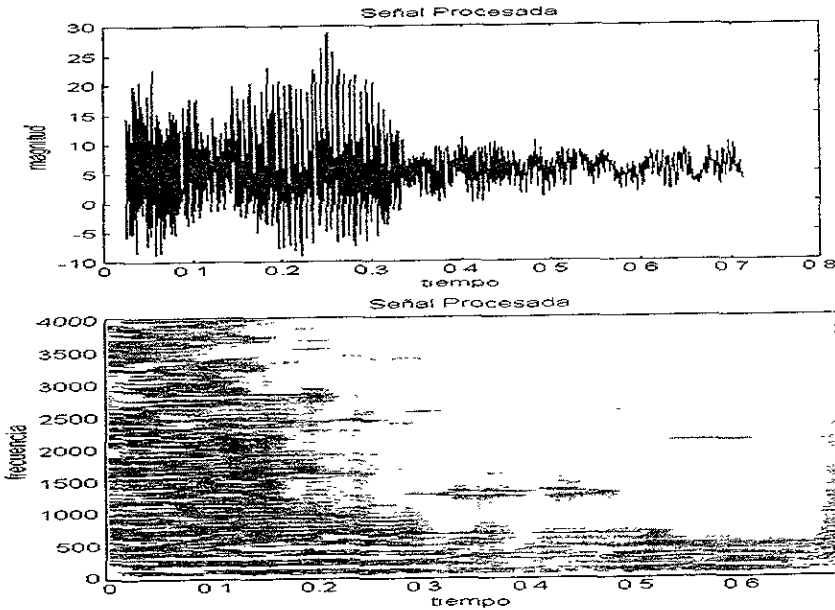


Fig.V.9 Resultados para la palabra "nueve" con una ventana con ruido de un auto Leopardo a 95 km/hr.

V.3 CONCLUSIONES

De las señales procesadas que se muestran se puede concluir que el algoritmo genera buenos resultados bajo circunstancias muy específicas y hasta cierto modo controladas. En un sistema de comunicaciones móviles real, lugar donde puede tener mayor aplicación el algoritmo, es muy difícil poder controlar todas las variables que influyen en el desempeño del sistema; por ejemplo, la posición que guardan los micrófonos con respecto a la persona que habla es muy variante al igual que las fuentes de ruido que pueden llegar a ser muy complejas debido a lo cual, los experimentos son solo una aproximación ya que se llevaron a cabo en un laboratorio.

Como ya se ha mencionado, la mayoría de los algoritmos de realce de voz son creados para casos muy específicos. La ventaja principal que presenta la función de coherencia como discriminante de ruido y voz es la de trabajar directamente con las muestras por una parte, y por otro lado no necesitar la detección de los periodos de silencio para obtener las características del ruido. Por lo tanto es posible pensar que tiene un ámbito mayor de aplicación ya que las suposiciones son muy pocas. Sin embargo, partir del hecho de que el ruido tiene una coherencia muy baja, casi cero, y que la misma medición en los micrófonos para el caso de la voz es casi uno, no siempre se cumple y restringe su aplicaciones. Para aproximarse a las suposiciones anteriores es necesario mantener una distancia considerable entre los micrófonos, alrededor de 40 cm, para que el ruido llegue en periodos de tiempo adecuados y se pueda tener en cada ventana solamente ruido decorrelacionado.

Al respecto de los parámetros del algoritmo, es posible decir que los valores tomados como máximo y mínimo para considerar a la señal como voz o ruido no son tan importantes como la longitud de la ventana y la rigidez con la que se lleve a cabo el filtrado. La longitud de la ventana esta directamente relacionado con la frecuencia de los formantes de la voz. 256 puntos es el valor que dio los mejores resultados; con una longitud menor se pierden algunos formantes y con valores superiores la señal procesada tiene mas posibilidad de contener ruido. De la misma manera, el valor de δ tiene una influencia muy marcada. Los mejores resultados se obtuvieron con valores entre 2 y 4. Por debajo de esos valores casi no existía filtrado y con valores superiores a pesar de que la forma de onda tenía una mejor apariencia la inteligibilidad era casi nula debido a la gran distorsión que sufre la señal.

Con el desarrollo actual de las telecomunicaciones y la gran demanda de servicios en los que las comunicaciones móviles han tomado un especial auge, la tendencia actual en el realce de voz es tomar en cuenta tanto al eco como al ruido en su conjunto como las fuentes principales de degradación. De tal manera, que la decorrelación espacial del ruidos, característica principal en la que se basa el algoritmo de la función de coherencia, la hacen una herramienta muy útil para resolver este problema y por lo tanto le dan actualidad.

CARACTERÍSTICAS FUNDAMENTALES DE LA VOZ

La comunicación oral se realiza a través del aire que sale de nuestra boca por medio de una onda longitudinal que cambia su amplitud de forma muy complicada mientras transcurre el tiempo y dentro de un rango muy amplio, desde niveles muy pequeños y poco notorios hasta niveles muy grandes que son muy fáciles de apreciar. Todo el proceso se lleva a cabo en los órganos productores del sonido que se pueden dividir en tres partes esenciales: el tracto pulmonar formado por los pulmones y la tráquea, la laringe que es el área que se encuentra por encima de la tráquea y por debajo de la faringe y por el tracto vocal, parte del sistema productor de voz que modula la señal, formado por la faringe y las cavidades bucal y nasal.

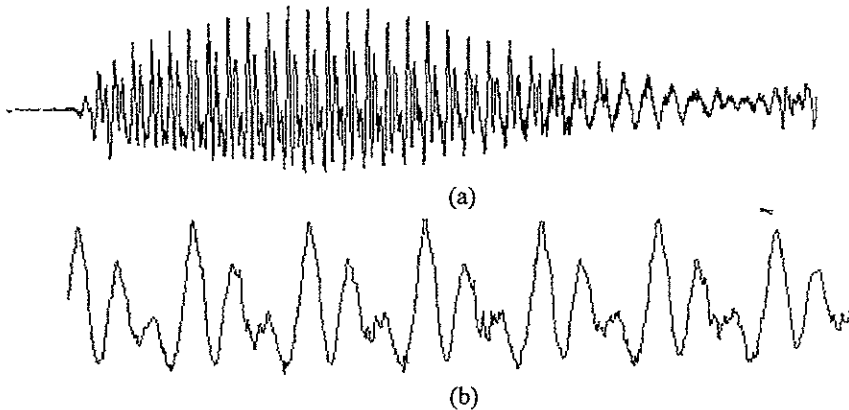


Fig. A.1 Ejemplos de formas de onda de voz. En (a) se puede ver una ventana de la forma de onda del fonema "e" y en (b) un acercamiento del mismo.

La amplitud de una señal promedio dentro de una conversación normal depende en gran parte de la lengua en la que se hable, pero generalmente podemos decir que tiene un promedio de 55 dB y que se distribuye a lo largo de 12 dB alrededor de la media.

A.1 Potencia de la Voz

Como ya hemos mencionado en la sección anterior el nivel en la forma de onda de la voz a lo largo de un tiempo determinado puede variar dependiendo de las circunstancias y cambios del hablante y del medio. En un lugar silencioso, el valor promedio es de 55 dB medidos con un micrófono a una distancia de un metro de los labios, aunque es necesario tomar en cuenta que la media en un hombre es superior en 4.5 dB a la de una mujer. Si tomamos en cuenta que podemos aproximar a una forma gaussiana el nivel de la voz a lo largo del tiempo con una desviación estándar de aproximadamente 3.8 dB.

En el caso de una conversación telefónica el valor medio en el nivel de la voz es de aproximadamente 63.1 dB. Si asumimos que la forma de onda se propaga de igual manera en todas las direcciones, la potencia radiada en una semiesfera con un radio de un metro será de 4π veces 63.1 dB que es $51.3 \mu W$. Ahora si consideramos que la potencia que se aprecia desde la parte trasera de la cabeza de la persona que habla es menor en 4 dB que enfrente de él, la potencia emitida es aproximadamente de 20 a $30 \mu W$.

A.2 Frecuencia de la Voz

De igual manera que la amplitud de la forma de onda, las componentes de frecuencia de la voz cambian continuamente a lo largo del tiempo afectada directamente por el tracto vocal, que como ya se ha dicho, funciona como un modulador. El comportamiento espectral del tracto vocal se modifica dependiendo de la forma que tome. Al mover la lengua se modifica la forma del tracto y junto con la posición de los labios y la mandíbula se produce un efecto de modulación. Si se midieran las componentes de frecuencia de la voz a lo largo del tiempo con la ayuda de un banco de filtro y se hiciera un promedio, podríamos obtener una gráfica como la que se muestra en la figura A.2. Sin embargo es necesario hacer notar que tales valores están muy relacionados con el idioma al que nos referimos y a si se han hecho mediciones en la voz de un hombre o de una mujer, pero en general la forma de la curva es muy similar.

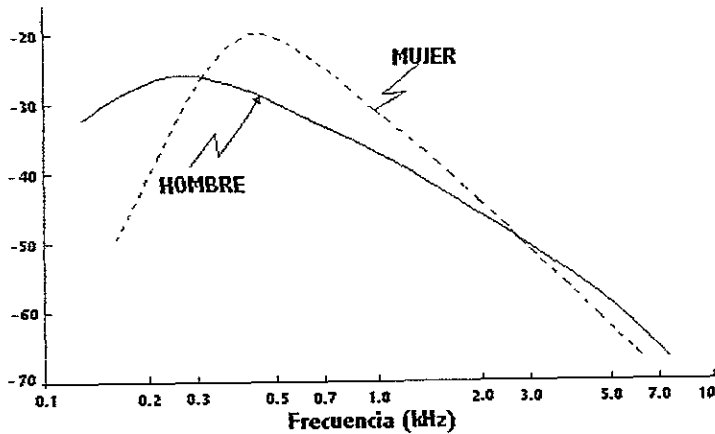


Fig A.2 Frecuencia Espectral de la voz.

La respuesta en frecuencia de un hombre es mayor en las bajas frecuencias, aproximadamente en 0.13 kHz, pero sin embargo en las mujeres la respuesta aumenta de forma muy rápida. De la figura se puede desprender que el 80 % de la energía de la señal esta contenida en frecuencias aproximadamente por debajo de 1 kHz.

A.3 Frecuencia Fundamental de la Voz

La forma de onda de la voz esta compuesta de dos partes; la primera que es una señal en la que la amplitud varia de manera aleatoria, y como segunda componente una parte periódica en la que se repite la misma forma de onda cíclicamente. Esta señal periódica esta formada de una frecuencia llamada fundamental, la menor, y de todas sus armónicas. La frecuencia fundamental corresponde a la frecuencia de vibración de las cuerdas vocales.

La frecuencia fundamental dentro de una conversación normal va cambiando continuamente y de forma paulatina en el tiempo. Como se muestra en la figura A.3 la desviación estándar de la frecuencia fundamental en la voz de una mujer es más del doble que para un hombre.

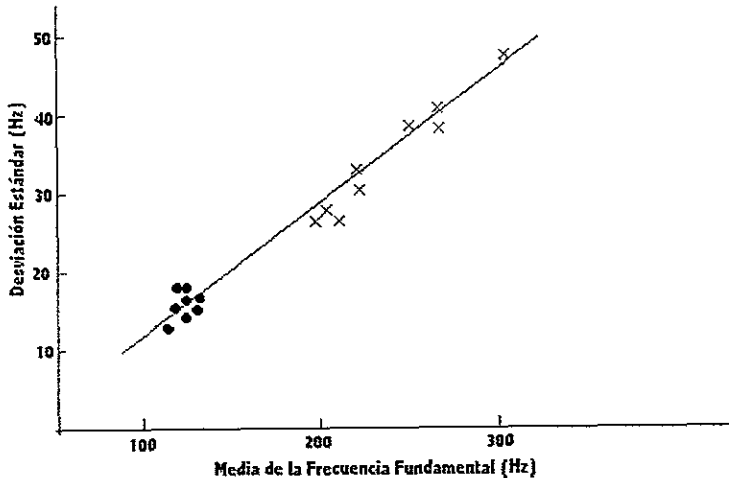


Fig. A.3 Media y Desviación Estándar de la Frecuencia Fundamental del hombre (●) y de la mujer (X).

A.4 Características fundamentales de los sonidos en la Voz.

La voz representa información lingüística que se describe por los fonemas que la forman. El número de los fonemas que forman una lengua depende de esta, pero en general van de 20 hasta 30 tipos diferentes que se clasifican en vocales y en consonantes.

A.4.1 Vocales.

Una vocal se genera con la resonancia (cada vocal se caracteriza por una resonancia en especial) del tracto vocal. Las cavidades oral y nasal se contraen en cierta forma junto con la excitación del flujo de aire y con la abertura y cierre de las cuerdas vocales cíclicamente generan el sonido. Algunos lenguajes llegan a tener hasta 10 vocales.

A las frecuencias predominantes que caracterizan a cada vocal se les llama formantes que se clasifican en primera, segunda, tercera y así continuamente a partir de la de mas baja frecuencia hasta la de mas alta frecuencia. La frecuencia de los formantes puede variar por factores tan diversos como es el sexo y muy especialmente por el fonema anterior y posterior a ellas, de tal manera que una vocal depende en gran medida por las primeras dos formantes.

A.4.2 Consonantes

Una consonante es formada por la turbulencia del flujo de aire que se genera muy cerca del punto más estrecho del tracto vocal debido a la articulación. Un flujo de aire en las cercanías de la cavidad oral genera silencios, las fricativas son generadas en el punto más estrecho y las africativas se generan con el flujo de aire entre los dos puntos anteriores. Una consonante puede ser sonora o sorda dependiendo de si las cuerdas vocales vibran o no.

En general las consonantes se clasifican por el lugar de articulación y se puede decir que cada fonema depende de la periodicidad de la forma de onda, de la frecuencia, de la duración y de la transición de estas.

A partir de un análisis de la duración de los fonemas muestra que la duración de las vocales casi no cambian y es de aproximadamente 70 msec y que las consonantes varían dependiendo de la clase entre 5 y 130 msec, de tal manera que una sílaba puede durar entre 75 y 200 msec.

SEÑALES ALEATORIAS Y DENSIDAD ESPECTRAL DE POTENCIA

Muchos de los fenómenos de la vida diaria se pueden representar mejor si se hace con modelos estadísticos debido a su naturaleza. Por ejemplo, la temperatura del aire y la voz humana que tienen una variación aleatoria en función del tiempo pueden ser modelados como señales de duración y energía finita.

B.1 Proceso Aleatorio

Por ejemplo supongamos que disponemos de generadores de ruido que tienen las mismas características y que generan una señal $x_i(t)$, de tal manera que es una señal aleatoria de potencia promedio finita. Al conjunto de todas las posibles formas de onda se les llama ensamble de la función del tiempo o proceso aleatorio. La forma de onda de cada generador en particular es diferente y se le llama realización del proceso aleatorio.

De manera similar, si pensamos en el ruido térmico en una resistencia, el voltaje que se mide en los extremos de la resistencia es una realización del proceso aleatorio consistente de todos los voltajes medidos en un conjunto de resistencias.

Al conjunto de todas las posibles mediciones se denota como $X(t, S)$, donde t representa al tiempo y S al conjunto de todas las posibles realizaciones. Cada realización en el proceso aleatorio se denota como $x(t, s)$. En general y por comodidad la variable que representa a las realizaciones no se escribe, de tal manera que un proceso aleatorio se denota como $X(t)$, y una de sus realizaciones como $x(t)$.

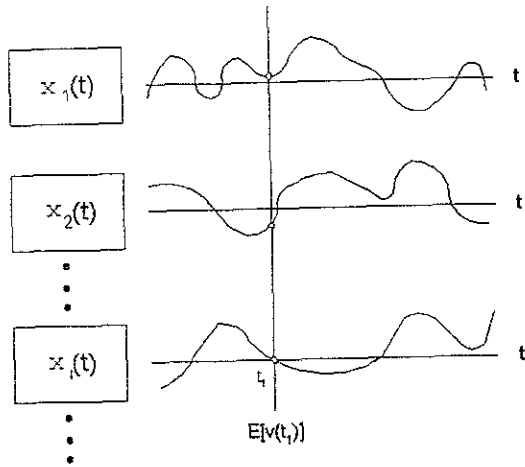


Fig. B.1 Representación de un proceso aleatorio con un conjunto de generadores de ruido aleatorio. Se observan las realizaciones sobre cada eje horizontal.

Considerando todo el conjunto de realizaciones, es decir, el proceso aleatorio en su conjunto durante cada instante de tiempo, por ejemplo t_i , podemos definir una variable aleatoria que se caracteriza estadísticamente con la función conjunta de probabilidad (PDF) que se denota como $p(x_{t_1}, x_{t_2}, \dots, x_{t_n})$, donde n es el número de realizaciones.

B.1 Proceso Aleatorio Estacionario

Supongamos que disponemos de un conjunto de muestras del proceso aleatorio $X(t)$, en $t = t_i$, $i = 1, 2, \dots, n$, y otro conjunto de muestras desplazadas por un tiempo τ , $X_{t_i+\tau} \equiv X(t_i + \tau)$, $i = 1, 2, \dots, n$. El segundo conjunto de variables tiene su propia función de densidad conjunta $p(x_{t_1+\tau}, x_{t_2+\tau}, \dots, x_{t_n+\tau})$. Si las dos funciones de densidad conjunta son iguales para todo τ , se dice que el proceso aleatorio es estacionario en sentido estricto. Es decir, las propiedades estadísticas de un proceso aleatorio estacionario no cambian a causa de un desplazamiento en el tiempo.

B.1 Densidad Espectral de Potencia

Un proceso aleatorio estacionario es un señal de energía infinita y por lo tanto la transformada de Fourier de Tal proceso no existe. La manera de caracterizar un proceso aleatorio en el dominio de la frecuencia es a través del teorema de Wiener-Khinchine,

calculando la transformada de Fourier de la función de autocorrelación del proceso. Al realizar esta operación estamos calculando la distribución de la potencia para cada frecuencia, al contrario de lo que obtendríamos con una transformada de Fourier simple en la que se obtiene la distribución del voltaje.

$$\Gamma_{xx}(F) = \int_{-\infty}^{\infty} \gamma_{xx}(\tau) e^{-j2\pi F\tau} d\tau \quad (\text{B.1})$$

y la transformada inversa de Fourier,

$$\gamma_{xx}(\tau) = \frac{1}{T} \int_{-\infty}^{\infty} \Gamma_{xx}(F) e^{j2\pi F\tau} dF \quad (\text{B.2})$$

De las ecuaciones anteriores podemos observar que:

$$\begin{aligned} \gamma_{xx}(0) &= \int_{-\infty}^{\infty} \Gamma_{xx}(F) dF \\ &= E(X_t^2) \geq 0 \end{aligned} \quad (\text{B.3})$$

Donde $\gamma_{xx}(0) = E(X_t^2)$ representa la potencia promedio del proceso aleatorio, que es el área bajo la curva de $\Gamma_{xx}(F)$, que como ya habíamos dicho, es la distribución de potencia a lo largo de la frecuencia, por esta razón se le llama Densidad Espectral de Potencia.

Si el proceso aleatorio es real, $\gamma_{xx}(\tau)$ es una función real y par y por lo tanto, $\Gamma_{xx}(F)$ también lo es. Si el proceso aleatorio es complejo, $\gamma_{xx}(\tau) = \gamma_{xx}^*(-\tau)$, entonces,

$$\begin{aligned} \Gamma_{xx}^*(F) &= \int_{-\infty}^{\infty} \gamma_{xx}^*(\tau) e^{j2\pi FF\tau} d\tau = \int_{-\infty}^{\infty} \gamma_{xx}^*(-\tau) e^{-j2\pi FF\tau} d\tau \\ &= \int_{-\infty}^{\infty} \gamma_{xx}(\tau) e^{-j2\pi FF\tau} d\tau = \Gamma_{xx}(F) \end{aligned} \quad (\text{B.4})$$

por lo tanto $\Gamma_{xx}(F)$ siempre es real.

Procesador Digital de Señales (DSP) TMS320C54x

El TMS320C54x es un DSP de punto fijo de la familia TMS320 de Texas Instruments. La arquitectura tipo Harvard modificada con la que cuenta le permite un gran ahorro en consumo de energía y un grado muy alto de paralelismo en sus instrucciones que junto con la gran versatilidad en los modos de direccionamiento y el conjunto de instrucciones le proporciona un gran desempeño en la resolución de diversos problemas en telecomunicaciones, computación, instrumentación e industria.

C.1 Descripción General

El TMS320C54x pertenece a la quinta generación de DSP de punto fijo de la familia TMS320 que hizo su aparición en el año de 1982 con el TMS320C10. El C54x tiene un alto grado de flexibilidad en operaciones y en velocidad. Combina una arquitectura Harvard modificada (un bus de memoria de programa, tres buses de memoria de datos y cuatro buses de direccionamiento), un CPU con aplicaciones específicas diseñadas en hardware, memoria y periféricos en el mismo chip, y un gran conjunto de instrucciones especializadas.

C.2 Características

CPU

- Arquitectura multibus con un bus de programa, tres de datos y cuatro de direccionamiento.
- Unidad Aritmética Lógica (ALU) y dos acumuladores de 40 bits.
- Multiplicador paralelo de 17 bits x 17 bits acoplado directamente a un sumador dedicado de 40 bits para la ejecución de la instrucción MAC en un solo ciclo.

- Unidad de selección y comparación (CSSU) para la operación suma/compra selecciona del código de Viterbi.
- Codificador de exponente para calcular el exponente del acumulador de 40 bits en un solo ciclo.
- Dos generadores de direcciones, incluyendo ocho registros auxiliares y la unidad Aritmética Lógica auxiliar.

MEMORIA

- Espacio de memoria direccionable de 192K palabras de 16 bits (64K de programa, 64K de datos y 64K de I/O).
- Configurable en el chip (K palabras).

Dispositivo	Memoria ROM	Memoria de Datos/Programa ROM	De acceso doble DARAM	De acceso sencillo SARAM
'C542	2	0	10	0

CONJUNTO DE INSTRUCCIONES

- De repetición, tanto de un bloque de instrucciones como de una sola instrucción.
- Para mover bloques de instrucciones para una mejor administración de la memoria.
- Con operadores de 32 bits.
- De lectura simultanea de dos o tres operadores.
- Aritméticas de escritura y lectura paralela.
- De escritura condicional.
- De regreso rápido de Interrupciones.

PERIFÉRICOS EN EL CHIP

- Generador de estados de espera programables vía software.
- Banco programable de conmutación.
- Reloj generado internamente por un PLL o de una fuente externa.
- Control de bus externo para deshabilitar el bus externo de datos, el bus de direccionamiento y las señales de control.
- Bus de datos con la característica de captura de bus.
- Temporizador programable.
- Puertos

Dispositivo	Puerto de Interfaz al Host	Puertos Seriales		
		Síncrono	Con memoria	TDM
'C542	1	0	1	1

POTENCIA

- Control en el Consumo mediante tres estados programables.
- Control para deshabilitar la señal CLKOUT.

VELOCIDAD

Dispositivo	Voltaje de Alimentación	Velocidad	Empaque
'C542	5 V	25 ns (40 MIPS)	144 pins TQFP
	3V / 3.3 V	25 ns / 20 ns (50 MIPS)	128 pins / 144 pins TQFP

C.3 Arquitectura

El 'C54x utiliza una arquitectura tipo Harvard modificada con ocho buses que maximiza la potencia de procesamiento. Teniendo por separado el espacio de memoria para datos y el espacio para programa permite el acceso a ambos al mismo tiempo lo cual provee un muy alto grado de paralelismo en las instrucciones. Por ejemplo, durante un solo ciclo de maquina se pueden realizar tres lecturas y una escritura en la memoria de datos.

C.3.1 Estructura de Buses

Como ya se ha mencionado, el TMS320C54x dispone de ocho buses de 16 bits. El bus de programa (PB) lleva el código de la instrucción y de los operadores desde la memoria de programa. Los tres buses de datos (CB, DB y EB) interconectan a varios elementos dentro de la arquitectura interna del DSP, entre los que se encuentran el CPU, el generador lógico de direcciones, el generador de direcciones de programa, los periféricos y la memoria de programa. Los buses CB y DB llevan los datos que fueron leídos, mientras que el bus EB lleva los datos que serán escritos en memoria. Los restantes cuatro buses, llamados de direccionamiento, (PAB,CAB,DAB y EAB) llevan las direcciones que se necesitan para ejecutar las instrucciones.

Aunado a los buses ya mencionados se tienen un bus bidireccional que accede a los periféricos y que se conecta con los buses DB y EB. El acceso a datos a través de este bus requiere de dos o más ciclos dependiendo del periférico al que se quiera llegar.

Con la ayuda de esta estructura de buses y de las dos unidades aritméticas lógicas (ARAU0 y ARAU1) que pueden generar hasta dos direcciones de memoria el TMS320C54x puede ejecutar instrucciones que contienen hasta tres operadores.

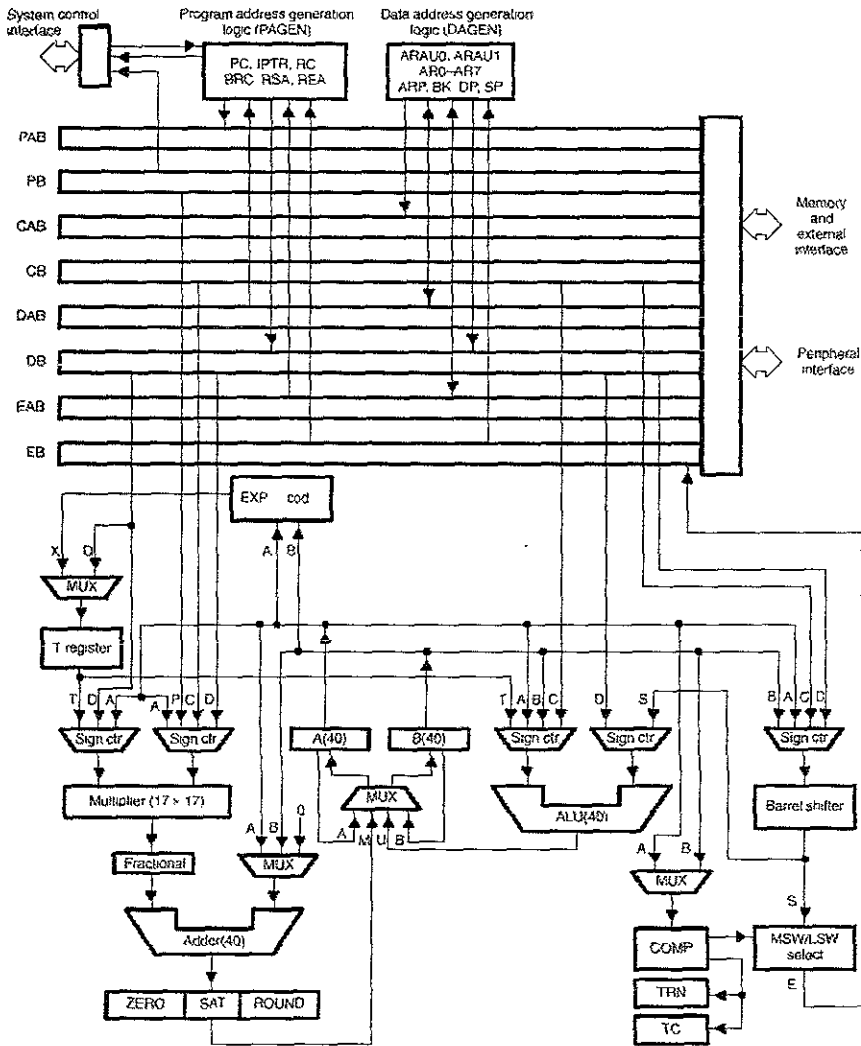


Fig. C.3.1 Diagrama de bloque de la arquitectura del TMS320C54x

C.3.2 Organización de la Memoria Interna

La memoria del 'C54x esta dividida en tres partes diferentes, la memoria de datos, la memoria de programa y la memoria de los dispositivos de entrada y salida (I/O). No importando la versión del dispositivo, todos disponen de memoria RAM y ROM y dependiendo de la versión se dispone de memoria RAM de acceso doble DARAM y/o de memoria de acceso sencillo SARAM. Como parte esencial en la estructura de la memoria se dispone de 26 registros internos del CPU y registros de periféricos mapeados en la memoria de datos sobre la pagina 0 para un fácil acceso como si fueran un dato mas.

Como parte importante en la seguridad de la información que reside en la memoria el 'C54x dispone de la opción de enmascaramiento de seguridad de tal manera que ninguna instrucción externa puede afectar al código en la memoria interna.

C.3.2.1 Memoria ROM

La memoria ROM es parte de la memoria de programa y/o de la memoria de datos dependiendo de la configuración. En los dispositivos que disponen de una memoria ROM muy pequeña solamente guardan ahí el programa de arranque que es muy útil para el arranque desde memoria RAM. Los dispositivos con mayor cantidad de memoria ROM por lo general son dispositivos especialmente cargados con código que es enviado por particulares para aplicaciones específicas.

C.3.2.1 Memoria DARAM y SARAM

La memoria DARAM se compone de varios bloque que pueden ser accedidos individualmente dos veces en cada ciclo por el CPU. La DARAM puede ser configurada como memoria de datos o de programa.

De igual manera la memoria SARAM está compuesta de varios bloques y por el contrario solamente se acceden una vez por cada ciclo ya sea para hace runa lectura o una escritura.

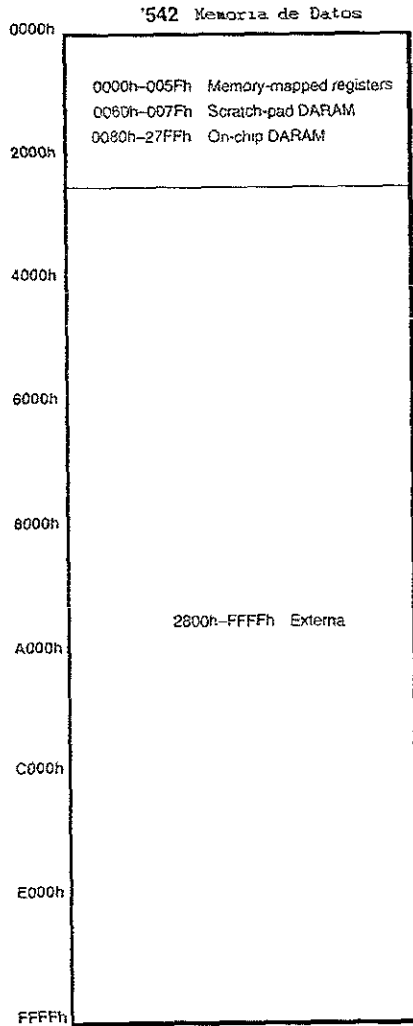
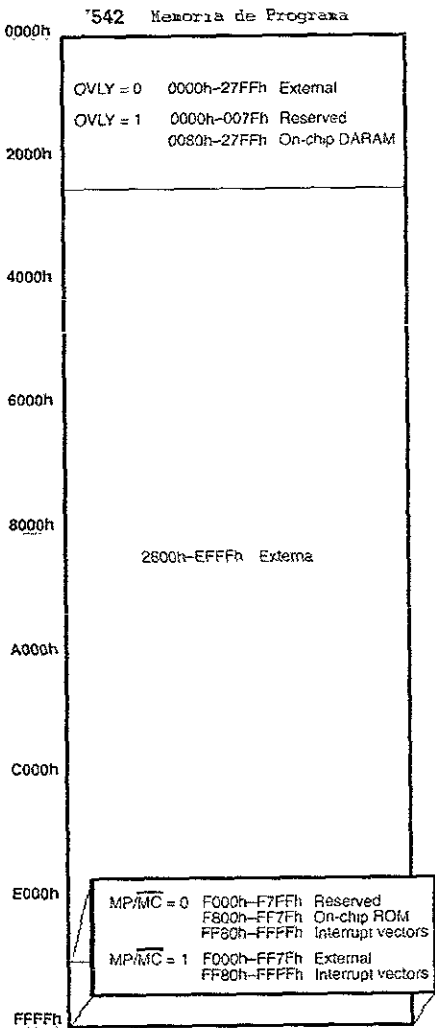


Fig. C.3.2.1 Mapa de Memoria del TMS320C54

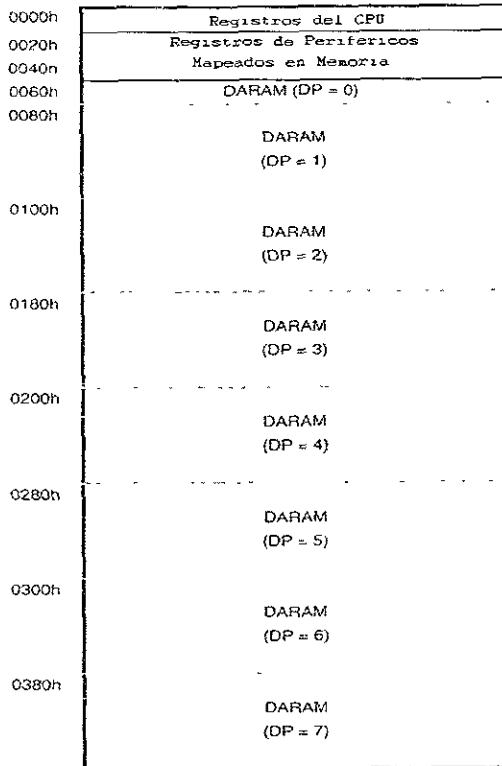


Fig. B.3.2.2 Registros Mapeados en Memoria

C.3.3 Unidad Central de Proceso (CPU)

El CPU está compuesto de la Unidad Aritmética Lógica (ALU), de dos Acumuladores, de un Operador de Desplazamientos, de un Multiplicador, de una Unidad de Comparación, Selección y Escritura y de dos generadores de direcciones, uno para la memoria de datos y otro para la memoria de programa.

C.3.3.1 Unidad Aritmética Lógica (ALU)

La ALU tiene 40 bits, trabaja con operaciones en complemento 2 y puede realizar operaciones Booleanas. Si así se requiriera, la ALU puede funcionar como dos ALU separadas de 16 bits cada una y funcionar simultáneamente.

C.3.3.3 Operador de Desplazamientos

Tiene una entrada de 40 bits conectada a los acumuladores o a la memoria de datos por medio de los buses CB y DB. Su salida de 40 bits se conecta directamente a la ALU o a la memoria de datos por medio del bus EB. Puede realizar desplazamientos de cero a 31 bits hacia la izquierda o de cero a 16 bits hacia la derecha dependiendo de la instrucción, de registro STI o del registro temporal T. Los bits menos significativos (LSB) se rellenan con ceros y los mas significativos (MSB) pueden rellenarse con ceros o con una extensión del bit de signo.

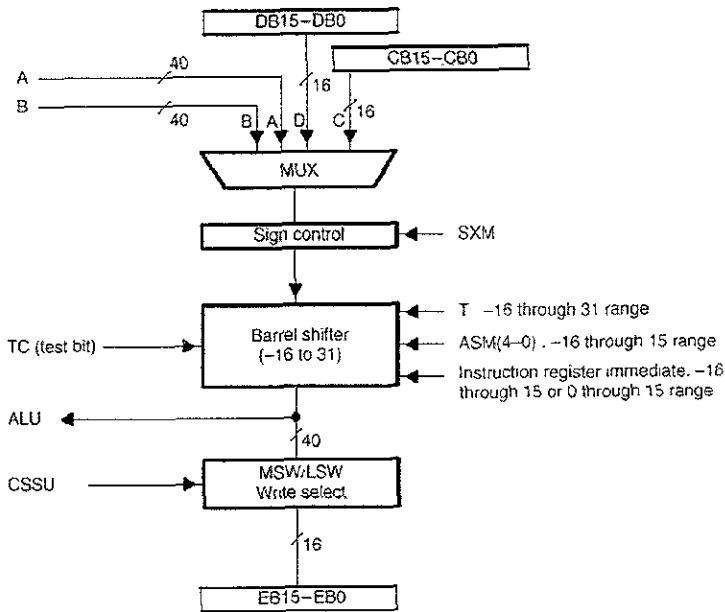


Fig. C.3.3.3 Operador de desplazamientos

C.3.3.4 Multiplicador /Sumador

Calcula una multiplicación en complemento 2 de dos números de 17 bits y una suma de una palabra de 40 bits en un solo ciclo. Esta constituido entre otros elementos por el multiplicador, el sumador y el registro temporal T que funciona como una de las entradas para el multiplicador junto con el acumulador A o un operando de la memoria de datos.

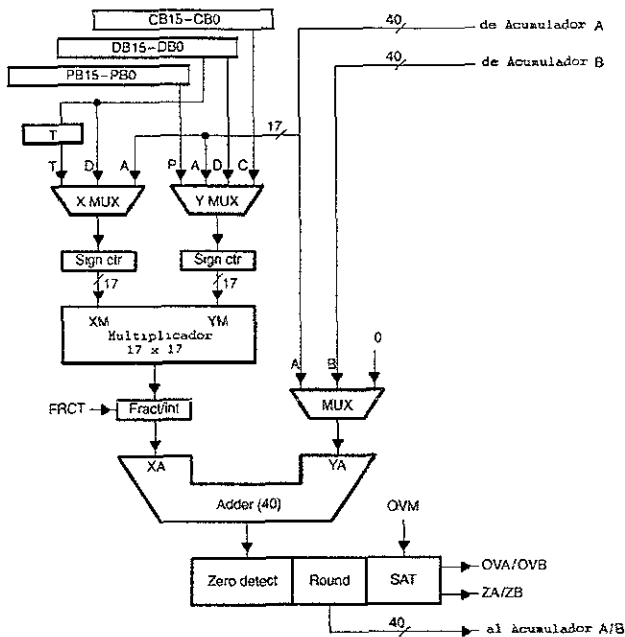


Fig. C.3.3.4 Diagrama Esquemático del Multiplicador / Sumador

C.3.3.5 Unidad de Comparación, Selección y Escritura (CSSU)

Realiza operaciones de comparación entre la parte alta y baja de los acumuladores (A y B) y seleccionar la parte mas grande para ser guardada en memoria. Permite a los bits (TC, bandera de control de estado del registro ST0) y al registro de transición TRN mantener una historia de los valores por los que han pasado. Con estas capacidades implementadas en hardware el CSSU permite realizar mucho mas rápido las mariposa en el algoritmo de Viterbi.

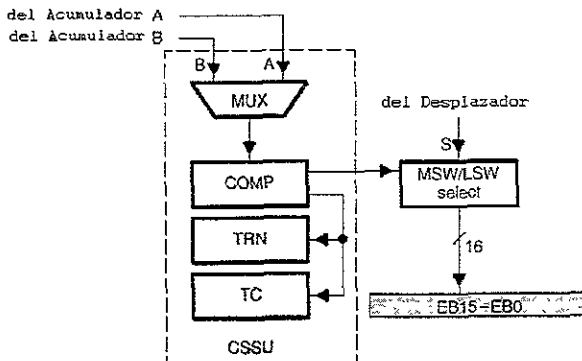


Fig. C.3.3.5 Unidades de Comparación, Selección y Escritura

C.4 Direccionamiento de Datos

El TMS320C54x tiene siete modos básicos de direccionamiento:

- Inmediato a través de la codificación de un valor en la instrucción.
- Absoluto a través de la codificación de la dirección en la instrucción.
- De acumulador, usando el valor guardado en el Acumulador para acceder a una parte de memoria de programa.
- Directa utilizando siete bits de la instrucción para codificar los siete bits menos significativos de la dirección. La dirección se completa con el apuntador de pagina (DP) o con el apuntador de pila (SP).
- Indirecta con el uso de los Registros Auxiliares.
- Con los Registros mapeados en memoria.
- De pila para administrar la entrada y salida de elementos de la pila.

Durante la ejecución de las instrucciones que utilizan direccionamiento directo, indirecto, con registros mapeados en memoria el generador lógico de direcciones de datos (DAGEN) calcula la dirección de los operadores en la memoria de datos.

C.5 Direccionamiento en la Memoria de Programa

El direccionamiento de la memoria de programa se hace a través del Contador de Programa (PC) que se obtiene del generador lógico de direcciones de programa (PAGEN). Generalmente el PAGEN incrementa el PC de manera secuencial, pero en el caso de instrucciones como interrupciones, llamado a subrutinas, saltos y reinicializaciones el valor del PC se guarda en la pila y se carga con la dirección que apunta a la primera instrucción de la subrutina, salto, etc. al termino de las operaciones de interrupción el PC se vuelve a carga con el valor guardado en la pila.

C.6 Operación del Pipeline

El pipeline de una instrucción se refiere a las operaciones que ocurren durante la ejecución de la operación. El 'C54x tiene un pipeline de seis niveles: prebúsqueda, búsqueda, decodificación, acceso, lectura y ejecución. En general cada vez que se tiene que ejecutar una operación durante un ciclo se operan seis, ya que mientras una instrucción es buscada otra ya esta siendo ejecutada. Cuando ocurre una discontinuidad en el PC, como se ha mencionado anteriormente, varios ciclos son perdidos para poder continuar con la secuencia lógica del pipeline.

C.7 Periféricos

Todos los Dsp's de la familia 'C54x tienen el mismo CPU, pero son diferentes en cuanto a los periféricos que tiene conectados. Las opciones de que se dispone son:

Pins de I/O de uso general (BIO y XF).

- Generador de estados de espera programable por software.
- Banco de conmutación lógico programable.
- Interfaz de host.
- Temporizador en hardware.
- Generador de reloj.
- Puerto serial.
 - Síncrono.
 - Amortiguado.
 - TDM

LISTADOS

D.1 MATLAB

```

function [xf, Cxy] = cohere1(s_o,x,y,f,fft,ventana,trasl,min,max,delt,lamb)
%
% COHERE1      Programa para el Realce de Voz.
%
% [xf, Cxy] = COHERE1(S_O,X,Y,FS,NFFT,VENTANA,TRASLAPE,MIN,MAX,DELTA,LAMB)
%
%           Utiliza la función de Coherencia como parámetro del filtrado. Primero se obtienen %los espectros de
las dos señales (x, y) con la transformada de Fourier, posteriormente se calculan %las Densidades Espectrales de Potencia
y a partir de ellas la Función de Coherencia.
%
%      s_o      señal original
%      x        primera señal degradada
%      y        segunda señal degradada
%      f        frecuencia de muestreo de las señales
%      nfft     tamaño de las ventanas
%      ventana  tipo de ventana
%      trasl    traslape utilizado
%      min      valor por debajo del cual se considera ruido solamente
%      max      valor por encima del cual se considera voz sin ruido
%      delt     parámetro del filtrado
%      lamb     valor utilizado para el calculos de las PDS (bebe ser muy cercano a 1)
%
%           Si no se especifican elementos de salida sedibuja la señal resultante.
%
%      Author: Eduardo Sanchez-Soto

error(nargchk(2,11,nargin));           %verifica que el número de parámetros sea el indicado

if (length(x)~=length(y)),
    error('X e Y deben ser de la misma longitud.')
end                                     % loa dos señales de entrada deben ser del mismo tamaño

sur(s_o,x)                             % se obtiene la SNR de las dos entradas
sur(s_o,y)

%x=x-mean(x);
%y=y-mean(y);
x = x(:);                               % Asegurarse que x sea un vector columna

```



```

y = y(:); % Asegurarse que y sea un vector columna
ventana = ventana(:);
%
n = length(x); % Numero de datos
nvent = length(ventana); % longitud de la ventana
k = fix((n-trasf)/(nvent-trasf)); % Numero de ventanas a procesar
% (k = fix(n/nwind) por trasf=0)
%
indice = 1:nvent; % Se calcula la primera ventana

xf=zeros(1,length(x));
gxx=ones(1,nfft)*0.4; gyy=gxx; gxy=gxx;
Cxy=gxx;
j=1;
Pxx = zeros(nfft,1); Xx2 = zeros(nfft,1);
Pyy = zeros(nfft,1); Yy2 = zeros(nfft,1);
Pxy = zeros(nfft,1); Xy2 = zeros(nfft,1); % se inicializan las variables

for i=1:k

    xw = ventana.*x(indice);
    yw = ventana.*y(indice); %Se pasan las señales por las ventanas

    Xx = fft(xw,nfft);
    Yy = fft(yw,nfft);

    if j==10000, % modificando este valor es posible evitar la aproximación en el
        Xx2 = (abs(Xx)).^2; % calculo de la PDS
        Yy2 = (abs(Yy)).^2;
        Xy2 = Yy.*conj(Xx);
        Pxx=Xx2;
        Pyy=Yy2;
        Pxy=Xy2;
        Cxy = ((abs(Pxy)/sqrt(Pxx.*Pyy)).^2);
        j=j+1
        gxx=Pxx; gyy=Pyy; gxy=Pxy;
    else
        Gxx = lamb.*gxx + ((1-lamb).*(Xx.*conj(Xx))); % calculo de la PDS
        Gyy = lamb.*gyy + ((1-lamb).*(Yy.*conj(Yy)));
        Gxy = lamb.*gxy + ((1-lamb).*(Xx.*conj(Yy)));

        gxx=Gxx; gyy=Gyy; gxy=Gxy;

        Cxy = (abs(Gxy ./ sqrt(Gyy.*Gxx)).^2);

        j=j+1;

        if j==100,
            j=1;
        end;
    end;

Xx=(Xx.*(min^(2*delt)*[Cxy <= min])) + ...
(Xx.*(Cxy.^delt .* [Cxy < max & Cxy > min])) + ...
(Xx.*[Cxy >= max]); % se hacen las comparaciones con los valores
% límite y se filtra la señal

xf(indice)=xf(indice)+fft(Xx,nfft); % se suma el resultado a la señal resultante
% trama por trama
indice=indice+(nvent-trasf); % se direcciona a la nueva ventana

end;

```

```

if (nargout >= 1), % la salida se pone a media cero cuando hay parámetros
    xf=detrend(real(xf),0), % de salida
    snr(s_o,xf+128) % se calcula de nuevo la SNR

elseif (nargout == 0), % do a plot
    newplot;
    plot(real(xf), grid
    xlabel('Tiempo'), ylabel('Señal Mejorada');
end; % si no hay parámetros de salida se grafica

```

```

function coh_fun(accion,var12)
%
% COH_FUN Función que genera un ambiente gráfico para la utilización de la función
% Cohere1 (Realce de Voz con la función de Coherencia)
%
% COH_FUN(ACCION,VAR12)
%
% Como primer paso, el programa solicita los dos archivos de entrada en formato WAV. A través de botones en
la ventana es posible observar el espectrograma de cada señal y regresar el despliegue al dominio del tiempo.
Presionando el botón "Cohere" se despliega otra ventana donde se pide el archivo de la señal sin ruido y se permite
ingresar los valores de los parámetros de la función COHERE1. Es posible grabar la señal resultante en un archivo
* MAT de MatLab.
%
% ACCION Indica la operación a realizar, por ejemplo. obtener el espectrograma,
% graficar la señal en el dominio del tiempo, o desplegar la ventana
% secundaria.
% VAR12 Mantiene el valor de la ventana activa.

```

```
% Autor: Eduardo Sánchez Soto
```

```

if nargin < 1
    accion='inicio';
end; % Verifica los parámetros de entrada

```

```

global Funcion_Coh, global Funcion_Coh1; global Funcion_Coh2; global Entradas,
global donde; global xf; global y; global x; global fx; global ty; global tx;
global fy; global fo; global control;

```

```
% se declaran variables globales
```

```

control=1;
ss_coh=1;ps_coh=1;FiguraSec_Coh1=1;so_coh=1;po_coh=1;

```

```

% _____
% Despliegue de la primera ventana _____
% _____

```

```
if strcmp(accion,'inicio'),
```

```
oldFigureNumber = watchon,
```

```

scrsz = get(0,'ScreenSize');
FiguraPrinc_Coh=figure('Position',[1 scrsz(3)/2 scrsz(3) scrsz(4)/2]);
set(gcf, ...
    'NumberTitle','off', ...

```

```

'Name','REALCE DE VOZ MEDIANTE LA FUNCION DE COHERENCIA (SEÑALES). ...
'backingstore','off',...
'Units','normalized');

m1 = uimenu(FiguraPrinc_Coh,'Label','Trabajo','BackgroundColor',[0 7 0.3 0 5]);
uimenu(m1,'Label','Par. Coh','Callback','figure');
uimenu(m1,'Label','Save','Callback','save');
uimenu(m1,'Label','Quit','Callback','exit',...
'Separator','on','Accelerator','Q');

coh_pushf = uicontrol('Style','Pushbutton',...
'Position',[9 250 50 25],...
'Callback','coh_fun("specg")',...
'String','Specg');

coh_pusht = uicontrol('Style','Pushbutton',
'Position',[9 200 50 25],...
'Callback','coh_fun("tiemp")',...
'String','Tiemp');

coh_pop = uicontrol('Style','Popup','String',...
'hsv\hot\cool\gray','Position',[9 100 50 25],...
'Callback','setmap');

coh_pop = uicontrol('Style','Pushbutton','String',...
'Coherencia','Position',[9 300 70 25],...
'Callback','coh_fun("cohe")');

play_pop = uicontrol('Style','Pushbutton','String',...
'Reprod','Position',[9 350 70 25],...
'Callback','coh_fun("play")');

%
% _____ Solicita los archivos de entrada y se hacen las gráficas _____
%

[nom_ar1,ruta1]=uigetfile('*.wav','Cargando Primer Archivo');

[x,fx]=wavread(nom_ar1);
x=x-mean(x);
x=x(:);
l=length(x);
tx=(1:l)/fx; % Se solicita el primer archivo

s1_coh=subplot('Position',[0.2 0.6 0.75 0.32]);
p1_coh=plot(tx,x);
ylabel('Magnitud 1');
xlabel('Tiempo (seg)'); % Se gráfica la primera señal

[nom_ar2,ruta2]=uigetfile('*.wav','Cargando Segundo Archivo');

[y,fy]=wavread(nom_ar2);
y=y-mean(y);
y=y(:);
l=length(y);
ty=(1:l)/fy; % Se solicita el segundo archivo

s2_coh=subplot('Position',[0.2 0.12 0.75 0.32]);
p2_coh=plot(ty,y);
ylabel('Magnitud 2');
xlabel('Tiempo (seg)'); % Se gráfica la segunda señal

```

ESTA TESIS NO DEBE
SALIR DE LA BIBLIOTECA

```
Funcion_Coh=[FiguraPrinc_Coh,s1_coh;p1_coh;s2_coh;p2_coh],  
Funcion_Coh2=[ss_coh,ps_coh],  
Funcion_Coh1=[FiguraSec_Coh1;so_coh;po_coh];
```

```
set(p1_coh,'ButtonDownFcn','coh_fun("actual",1));  
set(p2_coh,'ButtonDownFcn','coh_fun("actual",2));
```

```
%-----  
%-----Actualiza el valor de la ventana activa-----  
%-----
```

```
elseif strcmp(accion,'actual'),
```

```
    donde=var12;
```

```
%-----  
%----- Obtiene el Espectrograma -----  
%-----
```

```
elseif strcmp(accion,'specg'),
```

```
    FiguraPrinc_Coh=Funcion_Coh(1);  
    s1_coh=Funcion_Coh(2);  
    p1_coh=Funcion_Coh(3);  
    s2_coh=Funcion_Coh(4);  
    p2_coh=Funcion_Coh(5);
```

```
    FiguraSec_Coh1=Funcion_Coh1(1);  
    so_coh=Funcion_Coh1(2);  
    po_coh=Funcion_Coh1(3);  
    ss_coh=Funcion_Coh2(1);  
    ps_coh=Funcion_Coh2(2);
```

```
if donde==1
```

```
    figure(FiguraPrinc_Coh)  
    subplot(s1_coh);  
    plot(p1_coh);  
    [B,F,T]=specgram(x,512,fx,kaiser(500,5),475);  
    p1_coh=imagesc(T,F,20*log10(abs(B)));  
    axis('xy'), colormap(jet)  
    set(p1_coh,'ButtonDownFcn','coh_fun("actual",1)'),
```

```
elseif donde==2,
```

```
    figure(FiguraPrinc_Coh);  
    subplot(s2_coh);  
    plot(p2_coh);  
    [B,F,T]=specgram(y,512,fy,kaiser(500,5),475);  
    p2_coh=imagesc(T,F,20*log10(abs(B)));  
    axis('xy'), colormap(jet)  
    set(p2_coh,'ButtonDownFcn','coh_fun("actual",2)'),
```

```
elseif donde==3,
```

```
    figure(FiguraSec_Coh1)  
    subplot(so_coh);  
    plot(po_coh);  
    [B,F,T]=specgram(o,512,fx,kaiser(500,5),475);  
    po_coh=imagesc(T,F,20*log10(abs(B)));  
    axis('xy'), colormap(jet)  
    set(po_coh,'ButtonDownFcn','coh_fun("actual",3)'),
```

```
else
```

```
    figure(FiguraSec_Coh1)  
    subplot(ss_coh);  
    plot(ps_coh);
```

```

[B,F,T]=specgram(xf,512,fx,kaiser(500,5),475),
ps_coh=imagesc(T,F,20*log10(abs(B)));
axis('xy').colormap(jet)
set(ps_coh,'ButtonDownFcn','coh_fun("actual",4)');
end;

Funcion_Coh1=[FiguraSec_Coh1;so_coh;po_coh];
Funcion_Coh2=[ss_coh;ps_coh];
Funcion_Coh=[FiguraPrinc_Coh;s1_coh;p1_coh;s2_coh;p2_coh];

%-----
%-----Gráfica la señal en el tiempo-----
%-----

elseif strcmp(accion,'tiemp'),

FiguraPrinc_Coh=Funcion_Coh(1),
s1_coh=Funcion_Coh(2);
p1_coh=Funcion_Coh(3);
s2_coh=Funcion_Coh(4);
p2_coh=Funcion_Coh(5);

FiguraSec_Coh1=Funcion_Coh1(1);
so_coh=Funcion_Coh1(2);
po_coh=Funcion_Coh1(3);
ss_coh=Funcion_Coh2(1);
ps_coh=Funcion_Coh2(2);

if donde==1
figure(FiguraPrinc_Coh)
subplot(s1_coh);
plot(p1_coh);
p1_coh=plot(tx,x);
set(p1_coh,'ButtonDownFcn','coh_fun("actual",1)');
elseif donde ==2,
figure(FiguraPrinc_Coh);
subplot(s2_coh);
plot(p2_coh);
p2_coh=plot(ty,y);
set(p2_coh,'ButtonDownFcn','coh_fun("actual",2)');
elseif donde==3,
figure(FiguraSec_Coh1)
subplot(so_coh);
plot(po_coh);
po_coh=plot(tx,0);
set(po_coh,'ButtonDownFcn','coh_fun("actual",3)');
else
figure(FiguraSec_Coh1)
subplot(ss_coh);
plot(ps_coh);
ps_coh=plot(tx,xf);
set(ps_coh,'ButtonDownFcn','coh_fun("actual",4)');
end;

Funcion_Coh1=[FiguraSec_Coh1;so_coh;po_coh];
Funcion_Coh2=[ss_coh;ps_coh];
Funcion_Coh=[FiguraPrinc_Coh;s1_coh;p1_coh;s2_coh;p2_coh];

%-----
%-----Se despliega la segunda ventana, se solicita la señal original y se realiza el proceso de realce -
%-----

```

```

elseif strcmp(accion,'cohe'),

    oldfigNumber = watchon,

    scrsz = get(0,'ScreenSize')
    FiguraSec_Coh1=figure('Position',[1 scrsz(3)/6 scrsz(3) scrsz(4)/2]),
    set(gcf, ...
        'NumberTitle','off', ...
        'Name','REALCE DE VOZ MEDIANTE LA FUNCION DE COHERENCIA (MÉTOD0)', ...
        'backingstore','off',...
        'Units','normalized');

m2 = uimenu(FiguraSec_Coh1,'Label','Trabajo','BackgroundColor',[0.7 0.3 0.5]);
uimenu(m2,'Label','Par. Coh','Callback','figure'),
uimenu(m2,'Label','Save','Callback','save');
uimenu(m2,'Label','Quit','Callback','exit',...
    'Separator','on','Accelerator','Q');

%
uicontrol('Style','text',...
    'Position',[9 5 65 35],...
    'BackgroundColor',[0.7 0.3 0.5], ..
    'ForegroundColor','white',...
    'String','Min');
h_min=uicontrol('Style','edit',...
    'Position',[15 5 50 20],...
    'string','0.15');

uicontrol('Style','text',...
    'Position',[9 45 65 35],...
    'BackgroundColor',[0.7 0.3 0.5],...
    'ForegroundColor','white',...
    'String','Max');
h_max=uicontrol('Style','edit',...
    'Position',[15 45 50 20],...
    'string','0.98');

uicontrol('Style','text',...
    'Position',[9 85 65 35],...
    'BackgroundColor',[0.7 0.3 0.5],...
    'ForegroundColor','white',...
    'String','Delt');
h_delt=uicontrol('Style','edit',...
    'Position',[15 85 50 20],...
    'string','8');

uicontrol('Style','text',...
    'Position',[9 125 65 35],...
    'BackgroundColor',[0.7 0.3 0.5],...
    'ForegroundColor','white',...
    'String','Lamb');
h_lamb=uicontrol('Style','edit',...
    'Position',[15 125 50 20],...
    'string','0.95');

uicontrol('Style','text',...
    'Position',[9 165 65 35],...
    'BackgroundColor',[0.7 0.3 0.5],...
    'ForegroundColor','white',...
    'String','LFFT');
h_iffit=uicontrol('Style','edit',...

```

```

        'Position',[15 165 50 20],...
        'string','256');

uicontrol('Style','text',...
    'Position',[9 205 65 35], ...
    'BackgroundColor',[0.7 0.3 0.5],...
    'ForegroundColor','white',...
    'String','Tras');
h_tras=uicontrol('Style','edit',...
    'Position',[15 205 50 20],...
    'string','128');

h_pushf = uicontrol('Style','Pushbutton',
    'Position',[9 245 50 25],...
    'Callback','coh_fun("specg");',...
    'String','Specg');

h_pushnt = uicontrol('Style','Pushbutton',...
    'Position',[9 272 50 25],...
    'Callback','coh_fun("tiemp");',...
    'String','Tiemp');

h_graba = uicontrol('Style','Pushbutton',...
    'Position',[76 5 65 35],...
    'Callback','coh_fun("grab");',...
    'String','Graba');

Entradas=[h_min;h_max,h_delt;h_lamb;h_lfft;h_tras];

[nom_aro,rutao]=uigetfile(*.wav','Cargando Archivo Original');

[o,fo]=wavread(nom_aro);
o=o-mean(o);
o=o(:);
l=length(o);
to=(1:l)/fo;

so_coh=subplot('Position',[0.2 0.6 0.75 0.32]);
po_coh=plot(to,o);
ylabel('Magnitud');
xlabel('Tiempo (seg) Señal Original');

Funcion_Coh1=[FiguraSec_Coh1;so_coh;po_coh];

set(po_coh,'ButtonDownFcn','coh_fun("actual",3)');

%-----
ejecutar = uicontrol('Style','Pushbutton','String',...
    'Ejecutar','Position',[9 300 70 25],...
    'Callback','coh_fun("ejecuta");');

play_pop = uicontrol('Style','Pushbutton','String',...
    'Reprod','Position',[9 350 70 25],...
    'Callback','coh_fun("play");');

%-----
%--- Realiza la operación de realce-----
%-----

elseif strcmp(accion,'ejecuta'),

```

```

min=str2num(get(Entradas(1),'String'));
max=str2num(get(Entradas(2),'string'));
delt=str2num(get(Entradas(3),'string'));
lamb=str2num(get(Entradas(4),'string'));
lfft=str2num(get(Entradas(5),'string'));
tras=str2num(get(Entradas(6),'string'));

FiguraSec_Coh1=Funcion_Coh1(1);
so_coh=Funcion_Coh1(2);
po_coh=Funcion_Coh1(3);

xf = cohere1(o,x,y,fo,lfft,hanning(lfft),tras,min,max,delt,lamb),
xf=((xf-mean(xf));%/abs(max(xf))),
if control==1
figure(FiguraSec_Coh1);
ss_coh=subplot('Position',[0.2 0.12 0.75 0.32]);
ps_coh=plot(tx,xf);
ylabel('Magnitud');
xlabel('Tiempo (seg) Señal Procesada');
control=2;
else
ss_coh=Funcion_Coh2(1);
ps_coh=Funcion_Coh2(2);
figure(FiguraSec_Coh1);
subplot(ss_coh);
plot(ps_coh);
ps_coh=plot(tx,xf);
ylabel('Magnitud');
xlabel('Tiempo (seg) Señal Procesada');
end
Funcion_Coh2=[ss_coh;ps_coh];
Funcion_Coh1=[FiguraSec_Coh1;so_coh;po_coh];

set(ps_coh,'ButtonDownFcn','coh_fun("actual",4)'),

%-----
%---Reproduce las señales de trabajo-----
%-----

elseif strcmp(accion,'play')

if donde == 1
sound(x,fx);
elseif donde == 2
sound(y,fy);
elseif donde == 3
sound(o,fo);
else
sound(xf,fx);
end;

%-----
%---Graba en un archivo la señal resultante-----
%-----

elseif strcmp(accion,'grab'),
save 'salida.mat' xf fx;

end;

```


BIBLIOGRAFÍA

- [All77] Allen, J. B., D. A. Berkley, y J. Blauert. "Multimicrophone signal-Processing Technique to Remove Room Reverberations from Speech Signals." *Journal of the Acoustical Society of America* 62 (Octubre de 1977): 912-915.
- [Ata71] B. S. Atal y S. L. Hanauer. Speech analysis and synthesis by linear prediction of the speech wave. *J Acoustic Soc. Am.* 50(2):637-655,1971.
- [Bol79] S. F. Boll. Suppression of acoustic noise in speech using spectral subtraction. *IEEE ASSP*, 27(2):113-120, April 1979.
- [Bou92] R. Le Bouquin y G. Faucon. "Using de coherence function for noise reduction." *IEE Proceedings-I*, Vol. 139, No. 3, enero 1992.
- [Bur98] C. Sidney Burrus et all "Tratamiento de la Señal Usando Matlab" Prentice Hall, 1998.
- [Che98] Tsuhan Chen, "The Past, Present, and future of Speech Processiing",*IEEE Signal Processing Magazin* Mayo 1998.
- [Dud36] H. Dudley. *Synthesis Speech*. Bell Labs. Record, 15:98-102,1936.
- [Eph84] Y. Ephraim y D. Malah. Speech enhancement using a minimum mean square error short-time spectral amplitud estimator. *IEEE Trans ASSP*,32.1109-1121,1984.
- [Esc97] L. R. Escobar Salguero "Algoritmos de filtrado Adaptable: Implementación, Evaluación, Comparación y Aplicaciones en Telecomunicaciones" Tesis de Maestria DEPMI-UNAM
- [Eph89] Y. Ephraim, D. Malah y B. H. Juang. On the application of hidden Markov models for enhancing noisy speech. *IEEE Trans. ASSP*, 37(12):1846-1856, December 1989.
- [Gab60] D Gabor, W. P. L. Wilby, and R. Woodcock, "A universal nonlinear filter predictor and simulator which optimizes itself by a learning process," *Proc. Inst. Elec. Eng.*, vol. 108B, julio 1960.
- [Hay89] Simon Haykin "Modern Filters" MacMillan Publishing Company", USA 1989.
- [How65] P. Howells, "Intermediate frequency side-lobe canceller," U.S. Patent 3 202 990, Aug. 24 1965.
- [Ita68] F. Itakura y S. Saito. Analysis synthesis telephony based upon the maximum likelihood method. Editor Y. Kohasi, Editor, Reports of 6th Int. Cong. Acoust., pages C-5-5, C-17-20, Tokio, 1968.
- [Kal60] R.E. Kalma,"A New Approach to Linear Filtering and prediction Problems," *Trans. Of the ASME- Journal of Basic Engr.*, 35-45 marzo 1960.
- [Kau72] J. Kaunitz, "Adaptive filtering of broadband signals as applied to noise cancelling," Stanford Electronics Lab., Stanford Univ., Stanford Calif., Rep. SU-SEL-72-038, Agosto 1972 (Ph. D. dissertation).
- [Kof66] J. Koford and G. Groner, "The use of an adaptive threshold element

- to design a linear optimal pattern classifier," IEEE Trans Inform. Theory, vol IT-12, pp.42-50, Jan 1966.
- [Kry47] Kryter, K. D., J. C. R. Licklider, y S. S. Stevens. "Premodulation Clipping in AM Voice communication." Journal of Acoustical Society of America 19(January 1947):125-131.
- [Kem1791] W.V. Kempelen Le mecanisme de la parole, suivi de la description d'une machine parlante. J. V. Degen, Viena, 1791.
- [Kuo96] S M Kuo y D. R Morgan, "Active Noise Control Systems", Wiley Interscience 1996.
- [Lee72] Lee, F. F. "Time Compression and Expansion of Speech by the Sampling Method." Journal of the Acoustical Engineering Society 20 (noviembre 1972):738-742.
- [Lec89] Lecomte I. Lever M Boudy J Tassy A. "Car noise processing for speec input". ICASSP 89 1989 International Conference on Acoustic, Speech and Signal Processinf. Glasgow, UK pp 512-515, vol. 1 IEEE 23-26 mayo1989
- [Lic46] Licklider, J. C R. "Effects of Amplitud Distorsion upon the Intelligibility of Speech" Journal of Acoustical society of America 18 (Octubre 1946): 429-434.
- [Lim79] J. S. Lim y A V. Oppenheim Enhacements and barwidth compression of noisy speech Proc. IEEE, 67(12) 1586-1604, December 1979.
- [Luc65] R. Lucky, "Automatic equalization for digital communication," Bell Syst. Tech. J., vol. 44, pp. 547-588, Abril 1965.
- [Luc68] R. Lucky et al., "Principles of Data communication," New York, McGrawHill, 1968
- [Mar90] C R. Martins, et al "An Adaptive Noise Canceller And Its Iplementations on a DSP"IEEE
- [Mca80] R J McAulay y M L Malpass. Speech enhancement using a soft-decision noise suppression filter IEEE Trans ASSP, 28(2):137-145,April 1980
- [Roe87] Roe D. B., "Speech recognition with a noise-adapting codebook." Proceedings of ICASSP'87 Dallas, Texas. Pp 1139-1142. Abril 1987.
- [Sch65] M. R. Schroeder. US. Patent #3,180,936, filed 1960, issued April 1965.
- [Son67] Sondhi, M. M. "An Adaptive Echo Canceller." Bell system Technical Jorna 46 (marzo de 1967): 497-511.
- [Sor97] H. V. Sorensen y J. Chen, "A digital Processing Laboratory using the TMS320C30" Printice Hall 1997.
- [Sie64] Sievers y M. M. Sondhi. Sin publicación 1964.
- [Tho68] Thomas I. B. "The Influence of First and Second Formants on the Intelligibility of Clipped Speech." Journal of the Audio Engineering Society 16, (1968): 182.
- [War84] G E. Warnaka, L. A. Poole y J. Tichy, "Active Acoustic attenuators," Patente de E.U. 4,473,906, 25 septiembre 1984

- [Wei74] Weiss, Aschkenasy y Parson IEEE Symp. On Speech Recog., Pittsburgh, PA, 1974.
- [Wid60] B Widrow and M. Hoff, Jr., "Adaptive switching circuits," in IRE WESCON Conv. Rec., pt. 4, 96-104, 1960.
- [TI90] Digital Signal Processing Aplications with the TMS320 Family. 1990
- [TI97] TMS320C54x DSP CPU and Peripherals 1997
- [TI96] TMS320C54x DSKplus DSP Starter Kit 1996
- [TI97] TMS320C54x DSP Algebraic Instruction Set 1997

INTERNET

- 1 <http://spib.nce.edu/spib/>
- 2 <http://ftp.dcs.shef.ac.uk/share/spandh/ShATR>
- 3 http://amber.feld.cvut.cz/g_speech.htm
- 4 <http://www.speech.cs.cmu.edu/speech/>
- 5 <http://mambo.ucsc.edu/psl/speech.html/u>
- 6 <http://www.ircam.fr/equipement/analyse-synthese/oudot/>
- 7 <http://www.esca-speech.org>
- 8 <http://www.eurasip.org>
- 9 <http://www.ti.com/sc>