

24
2e



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

**ESCUELA NACIONAL DE ESTUDIOS
PROFESIONALES**

"ARAGON"

**LA APLICACION DE
MICROPROCESADORES AL
PROCESAMIENTO DIGITAL
DE SEÑALES**

TESIS PROFESIONAL

Que para Obtener el título de
Ingeniero Mecánico Electricista

P R E S E N T A

JUVENTINO DELFINO ROSAS ESPEJEL

SAN JUAN DE ARAGON, EDO. DE MEXICO

1988

**TESIS CON
FALLA DE ORIGEN**



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 está 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.

INDICE

	Pág
AGRADECIMIENTOS.....	1
INTRODUCCION GENERAL.....	2
CAPITULO 1	
INTRODUCCION AL PROCESAMIENTO DIGITAL	
DE SEÑALES DE VOZ	
1.1 La señal de voz.....	7
1.2 Procesamiento digital de señales de voz.....	10
1.3 Estrategias de codificación.....	14
1.4 Sistemas de síntesis de voz.....	17
1.5 Sistemas de verificación e identificación de voz.....	18
1.6 Sistemas de reconocimiento de voz.....	18
1.7 Mejoramiento de la calidad de la señal.....	19
1.8 Ayuda para los incapacitados.....	19

CAPITULO 2

FUNDAMENTOS DEL PROCESAMIENTO DIGITAL

DE SEÑALES DE VOZ

2.1 Introducción.....	21
2.2 Descripción de los sistemas en el tiempo....	23
2.3 Transformadas de las señales.....	30
2.4 Muestreo.....	39
2.5 Proceso de producción de voz.....	45
2.6 Teoría acústica de la producción de voz.....	51
2.7 Modelos tubulares sin pérdidas.....	69
2.8 Modelos digitales para señales de voz.....	77
2.8.1 Modelo del tracto vocal.....	80
2.8.2 Modelo de Radiación.....	81
2.8.3 Modelo de Excitación.....	82

CAPITULO 3

ALGORITMOS PARA EL PROCESAMIENTO DIGITAL

DE SEÑALES DE VOZ

3.1 Modulación delta.....	89
3.2 Modulación delta adaptable.....	99
3.3 Predictores de orden alto en la modulación delta.....	107
3.4 PCM diferencial (DPCM).....	111
3.5 PCM diferencial adaptable (ADPCM).....	112
3.6 Codificación por sub-bandas.....	121

3.7 Vocoders	
3.7.1 Codificación predictiva lineal multi-pulso (MLPC).....	131
3.7.2 Codificación predictiva lineal exitada estocásticamente (SELPC).....	133
3.7.3 Vocoder verdadero.....	134
i) Vocoder de canal.....	137
ii) Vocoder cepstrum.....	140
iii) Vocoder con formantes.....	143
iv) Vocoder exitado sonoramente.....	144
3.8 Codificación por predicción lineal.....	147

CAPITULO 4

DISEÑO DE UN SISTEMA DE CODIFICACION ADPCM

4.1 Introducción.....	155
4.2 Estructura y descripción general.....	158
4.3 Diseño.....	163

CAPITULO 5

RESULTADOS, CONCLUSIONES Y RECOMENDACIONES SOBRE EL DESARROLLO DEL SISTEMA ADPCM

5.1 Introducción.....	202
5.2 Caracterización técnica y subjetiva.....	203
5.3 Estimado económico del proyecto.....	216

	Pág
5.4 Conclusiones.....	219
APENDICE A.....	222
APENDICE B.....	228
APENDICE C.....	230
APENDICE D.....	236
GLOSARIO.....	245
BIBLIOGRAFIA.....	246

INTRODUCCION

El procesamiento digital de señales es un área que en la actualidad tiene un gran crecimiento, en la última década se han desarrollado nuevos algoritmos ,nuevas tecnologías y se han encontrado nuevas aplicaciones de estas técnicas en muchos campos donde anteriormente no existían.

La tecnología utilizada en el procesamiento digital de señales es difícil de utilizar, porque implica tener que tratar con las matemáticas complejas, requiere hacer consideraciones críticas sobre velocidad, etc. Pero al mismo tiempo esta tecnología hace posible un amplio rango de aplicaciones, desde muñecas que hablan hasta radares, desde imágenes obtenidas por ultrasonido para aplicaciones médicas hasta sistemas de control adaptables para eliminar las patinadas de los automóviles, desde sintetizadores musicales hasta los sistemas de control industriales, etc.

La tecnología del procesamiento digital de señales ha avanzado desde los componentes discretos de semiconductor a la integración de gran escala (LSI) con densidades superiores a los 100,000 componentes por chip.

En esta década han aparecido dos nuevos productos integrados en un sólo chip, el procesador digital de señales (DSP), que es un microprocesador de propósito específico y los ASIC's que en el futuro se espera serán los integrados de más alto crecimiento para aplicaciones en el procesamiento digital de señales.

Para entender que son los ASIC's es necesario conocer antes la tecnología de construcción por bloques. Los bloques son integrados que ejecutan funciones primitivas específicas como son los multiplicadores de doble precisión, circuitos que ejecutan la transformada rápida de Fourier, etc. Cuando el diseñador utiliza la construcción por bloques, al hacer un diseño toma los bloques primitivos (integrados) y los conecta sobre una tarjeta, con esta técnica el diseñador se ahorra el tiempo de diseño de cada uno de los sistemas contenidos en los bloques primitivos (integrados) concretándose únicamente a diseñar la lógica para interconectar los bloques de manera que el sistema trabaje eficientemente.

Los ASIC's mantienen una gran afinidad con las técnicas de construcción por bloques, aunque en general los ASIC's tienen menor velocidad de trabajo tienen la ventaja de que el diseñador puede escoger los bloques necesarios para realizar su sistema dentro de un sólo chip, mejor que en varios chips como en el caso de las técnicas de construcción por bloques, facilitándose de esta manera el diseño de los sistemas para el

procesamiento digital de señales al no tenerse que realizar alambrado entre varios chips.

El desarrollo de una aplicación para el procesamiento digital de señales se comienza haciendo una lista de los requerimientos técnicos y de costos. En este proceso se deben de tener en cuenta el tiempo y el espacio, el uso de microprocesadores de punto flotante o de punto fijo (el TMS32010 es un microprocesador de punto fijo), la entrada/salida de las señales y la interface con otros sistemas.

La estimación del tiempo y del espacio se determina con una simulación completa del algoritmo en un lenguaje de alto nivel, en general los algoritmos para el procesamiento digital de señales gastan alrededor del 80% de los ciclos en ejecutar operaciones primitivas para el procesamiento de señales como son las correlaciones, convoluciones, transformadas rápidas de Fourier y filtrado.

En todas estas operaciones primitivas el producto de dos operandos y la acumulación del resultado era la peor consumidora de tiempo hasta que se desarrollaron los nuevos microprocesadores capaces de ejecutar la multiplicación y la acumulación del resultado en un sólo ciclo de reloj. El microprocesador TMS32010 que se utiliza para hacer el desarrollo del algoritmo ADPCM motivo de esta tesis tiene integrada esta última característica, aunque este

microprocesador tiene el defecto de tener instrucciones de control poco eficientes que consumen más ciclos de reloj que la multiplicación/suma.

En la actualidad están incurriendo en el campo del procesamiento digital de señales los nuevos microprocesadores de punto flotante, los cuales tienen algunos bits más de precisión, hacen las operaciones primitivas más rápido y utilizan menos espacio para el código con su aritmética de punto flotante al eliminar los escalamientos, el chequeo de límite de ciclos y el chequeo de ocurrencia de sobreflujos.

En general todas las señales que se deben procesar comienzan estando en el dominio del tiempo, por lo que primero deben ser digitalizadas y después introducidas a el sistema DSP para ser procesadas, en algunos casos la digitalización de las señales se realiza en un sistema de adquisición de datos separado del sistema del DSP. Y con frecuencia el sistema del DSP es parte de un sistema mucho más grande como un modem digital, algún sistema de reconocimiento de voz, etc.

Para el caso de esta tesis se utilizó el DSP TMS32010 por ser este el microprocesador que se tiene en el laboratorio del Departamento de Comunicaciones del IIE, para el diseño del ADPCM robusto de 32 Kb/s se utilizaron las técnicas de escalamiento, pero no hubo necesidad de checar la ocurrencia de sobreflujos. Los pormenores de este diseño se analizarán a través de los capítulos de esta tesis, destacándose en cada

caso los pros y los contras de cada una de las desiciones tomadas para la realización del ADPCM robusto de 32 Kb/s.

Es importante destacar que los proyectos de tesis como el presente, se realizan en el IIE, cofinanciados por el CONACYT con la finalidad de acrecentar la investigación científica del país, pero tratando de no desligar estas investigaciones de las aplicaciones tecnológicas. Los proyectos como el presente generan prototipos, que en su momento se transfieren al sector industrial para su producción, contribuyendo así al desarrollo de nuestro país en un sector muy importante como es el sector tecnológico.

CAPITULO I

INTRODUCCION

INTRODUCCION AL PROCESAMIENTO DIGITAL DE SEÑALES DE VOZ

1.1 Propósito de esta tesis

El propósito de esta tesis es mostrar como las técnicas de procesamiento digital de señales se aplican en problemas de la comunicación de voz. Este capítulo da una breve semblanza de preguntas como: cuál es la naturaleza de la señal de voz?, cuáles son algunas de las áreas de aplicación de la comunicación de voz en las cuales se han utilizado las técnicas de procesamiento digital de señales?, etc.

En el resto de la tesis se analizan los diferentes algoritmos que existen para lograr reducir el ancho de banda de la señal digitalizada que es transmitida y el desarrollo de un sistema codificador diferencial adaptable PCM (ADPCM) con sus resultados en las pruebas técnicas y subjetivas.

1.2 La señal de voz

El propósito de la voz es la comunicación. Hay varios modos de caracterizar el potencial de comunicación de la voz y una forma es en términos de la teoría de la información, la cual nos dice que la voz puede ser representada por su contenido de mensaje o información. Las ideas de la teoría de la información juegan un papel importante en los sistemas de

comunicación más sofisticados, sin embargo la comunicación de la voz basada en su forma de onda o un modelo paramétrico, es más usual para aplicaciones prácticas.

Para considerar el proceso de comunicación de voz primero es bueno pensar la representación del mensaje dentro del cerebro en una forma abstracta, luego a través del proceso de producción de voz la información del mensaje se convierte en una señal acústica. La información del mensaje se convierte primero en una serie de señales neuronales que controlan el mecanismo articulatorio (los movimientos de la lengua, labios, cuerdas vocales, etc.). Los articuladores se mueven en respuesta a las señales neuronales para ejecutar una secuencia de gestos, que al final resultan en una onda acústica que contiene la información del mensaje original.

La información que se comunica por la voz es intrínsecamente de naturaleza discreta, esto significa que puede ser representada por una concatenación de elementos de una colección de símbolos. Los símbolos que el sonido puede ser clasificado se llaman "fonemas", cada lengua tiene su propia colección de fonemas.

Uno de los principales intereses de la teoría de la información es la tasa a la cual la información es transmitida. En la voz, una estimación burda de la tasa de información se obtiene al notar las limitaciones físicas de la tasa de movimiento de los articuladores humanos, produciéndose

un promedio de fonemas por segundo. Si cada fonema se representa por un número binario, entonces con un código numérico de 6 bits se puede representar el total de los fonemas del lenguaje (el lenguaje tiene entre 30 y 50 fonemas). Además asumiendo una tasa promedio de 10 fonemas por segundo y despreciando la correlación de fonemas adyacentes obtenemos una estimación de 60 bits/seg para el promedio de la tasa de la información de la voz. Sin embargo el límite más bajo del contenido verdadero de información de la voz es considerablemente más alto a esta tasa. Esto es porque la estimación no toma en cuenta factores como la identidad y el estado emocional de la persona que habla, fuerza de la voz, etc.

En los sistemas de comunicación de voz, la señal de voz es transmitida, guardada y procesada de muchas maneras, esto conduce a una amplia variedad de representaciones de la señal de voz, pero en general hay dos intereses principales en cualquier sistema:

- 1.-Preservación del contenido del mensaje en la señal de voz.

- 2.-Representación de la señal de voz en forma conveniente para la transmisión o almacenado, en forma tal que se puedan hacer modificaciones a la señal de voz sin que se degrade en forma notable el mensaje del contenido.

La representación de la señal de voz deberá ser de forma tal que el contenido pueda ser extraído por el escucha de una manera fácil, o también de manera automática por una máquina diseñada con este fin.

1.3 Procesamiento digital de señales de voz

Un sistema de comunicación digital básico se puede representar como se ve en la figura siguiente:

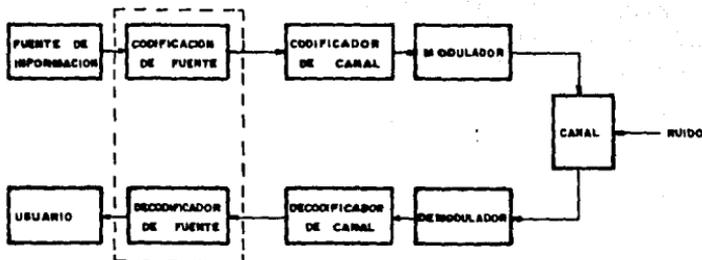


Fig 1.1 Sistema básico de comunicación digital

En la entrada del sistema tenemos una fuente de información que puede ser voz, datos, video; enseguida de esta parte viene la codificación de fuente en la que se tienen codificadores que reducen el ancho de banda de la señal de información, mezcladores (scrambler) para que no cualquier receptor pueda recuperar la señal sino solo al que se desea

que reciba la señal, etc. Después viene la sección de codificación de canal, que cumple la función de codificar la señal de manera que al final en el receptor se pueda recuperar la señal originalmente transmitida sin los errores que se producen por adición del ruido de canal, y además en algunos casos también se tiene que recuperar la señal de reloj en el receptor a partir de los datos transmitidos, para tener la sincronía adecuada en el receptor. Después de estas secciones viene el modulador el cual se encarga de acoplar la señal digital con las características y limitaciones del canal de transmisión esto se logra variando las características de la señal portadora de acuerdo con la señal moduladora (datos), en el canal se agrega ruido aditivo ya sea por factores externos o intrínsecos del canal, que degradan la señal.

En el receptor se encuentra el demodulador que se encarga de recuperar la señal digitalizada y codificada al separar la señal moduladora de la señal portadora, esta señal pasa al decodificador de canal que es el encargado de checar y corregir los errores generados en el canal de transmisión, es muy importante que este circuito este bien diseñado para la cantidad de errores que en promedio se generan en el canal de transmisión, ya que si es mayor la cantidad de errores generados que la soportada por el decodificador este no podrá recuperar la señal transmitida originalmente.

Después del decodificador de canal la señal pasa a él decodificador de fuente que recupera la señal comprimida por el codificador de fuente a una señal que se trata sea lo más fiel posible con la señal original de la fuente de información y de esta forma es presentada de manera comprensible para el usuario.

Del sistema de comunicaciones digital anterior, esta tesis se enfoca a el bloque correspondiente a la codificación de fuente, con el fin de comprimir el ancho de banda ocupado por la voz, para tratar de obtener un uso más óptimo del ancho de banda de el canal de transmisión.

Una vez comprimido el ancho de banda de la señal de voz se tiene espacio en el canal para transmitir más señales de voz, datos, etc. de forma multiplexada.

En el diseño de los codificadores digitales de señales de voz el esfuerzo se encamina a optimizar la acción mutua de cuatro parámetros: tasa de bits transmitida, calidad, complejidad y el tiempo de retraso.

Como la tasa de bits (ancho de banda) es reducida, la calidad de la naturalidad de la voz reconstruida cae, a menos que la complejidad del esquema de codificación sea incrementada. Pero la complejidad sube el costo y en muchos tipos de codificación esto incrementa la demora del procesamiento.

Para tasas altas como 64 Kb/s la calidad no es un problema. La modulación por pulsos codificados(PCM) producira alta calidad, esto quiere decir que poca gente puede distinguir al final de una línea telefónica que esta ha sido transmitida digitalmente. La onda de voz analógica es muestreada , cuantizada, codificada digitalmente y transmitida a 64 Kb/s. En el receptor, el proceso es invertido entonces por un decodificador.

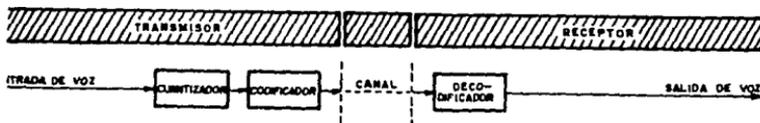


Fig 1.2 Modulación por pulsos codificados.

Se tiene un algoritmo de cuantización no uniforme para el PCM: usando un cuantizador alineal en el cual la fineza de la cuantización se incrementa como la magnitud de la voz de entrada decrece(ley- μ). Esta alinealidad favorece las bajas amplitudes, que predominan en la voz, cuantiza estas con error de cuantización imperceptible para efectos practicos. Este algoritmo explota una característica del oído humano: las grandes amplitudes enmascaran los errores de cuantización en algún grado.

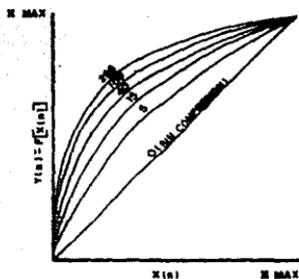


Fig 1.3 Relación entrada-salida para la ley- μ .

1.4 Estrategias de codificación

Para tasas bajas, sin embargo, son necesarias estrategias más elaboradas. La función general de estas estrategias es escudriñar la señal de voz más cuidadosamente, para eliminar las redundancias en la señal, y usar los bits disponibles para codificar las partes de la señal no redundantes de manera eficiente. Como la tasa de bits disponible es reducida a 32, 16, 8 y 4 Kb/s, las estrategias para remover las redundancias necesitan aún ser mas sofisticadas.

Por esto, considerando la aplicación de las técnicas del procesamiento digital de señales a los problemas de comunicación por voz, es útil enfocarse en tres tópicos

principales: la representación de señales de voz en forma digital, la implementación de las técnicas de procesamiento, y las clases de aplicación del procesamiento digital de señales.

En la representación digital de señales de voz se aplica el conocido teorema de muestreo que establece que las señales limitadas en banda pueden ser representadas por muestras periódicas tomadas a cuando menos dos veces la frecuencia más alta de la señal.

Hay muchas posibilidades para la representación discreta de señales de voz, pero estas representaciones pueden ser clasificadas en dos grandes grupos, llamadas representación de forma de onda y representación paramétrica. Ver figura (1.4).



Fig 1.4 Representaciones de la señal de voz.

La representación por forma de onda implica preservar la "forma de onda" de la voz analógica a través del muestreo y el proceso de cuantización. Por otro lado la representación paramétrica representa la señal de voz como la salida de un

modelo para la producción de voz. Los parámetros de ese modelo son convenientemente clasificados como parámetros de excitación y parámetros de respuesta del tracto vocal (ver fig. 1.4).

En la figura (1.5) se muestra una comparación de un número de diferentes representaciones de las señales de voz acorde a la tasa de datos requerido.

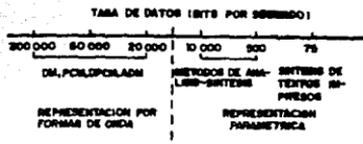


Fig 1.5 Rango de las tasas de bits para varios tipos de representaciones de voz.

Por supuesto la tasa de datos no es la única consideración para escoger la representación de la voz. Otras consideraciones son costo, flexibilidad de la representación, calidad de la voz, etc.

Por último la clase de aplicación es la mas importante consideración para escoger la representación de la señal y los métodos del procesamiento digital de señales aplicados subsecuentemente. La figura (1.6) muestra algunas de las muchas áreas de aplicación en las comunicaciones por voz.



Fig 1.6 Aplicaciones típicas de la comunicación por voz

Vale la pena una breve discusión de cada una de estas áreas.

1.5 Sistemas de síntesis de voz

Los sistemas de síntesis de voz son estimulados por la necesidad de un económico almacenaje digital de la señal de voz para los sistemas de respuesta computarizado. Un sistema de respuesta de voz computarizado es básicamente un servicio digital automático de información, el cual le puede preguntar

una persona por medio del teclado de una terminal, y el cual respondera la información deseada en forma hablada. Un ordinario teléfono "touch-tone" puede ser el teclado para un sistema semejante, las capacidades del servicio de información automático puede hacerse disponible sobre todos los teléfonos sin la necesidad de un equipo adicional especializado.

1.6 Sistemas de verificación e identificación de voz

Las técnicas de verificación e identificación de voz se usan para la autenticación e identificación de la persona que habla, de un grupo de posibles hablantes. El sistema de verificación debera decidir si la persona que habla es la persona que afirma ser. Este sistema es aplicable en situaciones requiriendo control de acceso a información o algunas áreas restringidas y varias clases de transacciones de crédito automáticas.

1.7 Sistemas de reconocimiento de voz

En general en estos sistemas hay una conversión de una onda acústica a un mensaje de información escrito equivalente. Como un ejemplo de estos sistemas podemos mencionar una impresora operada por voz y comunicaciones por medio de voz con las computadoras.

1.8 Mejoramiento de la calidad de la señal de voz

Hay muchas situaciones en que las señales de voz son degradadas de modo que limitan su efectividad para la comunicación. En estos casos las técnicas del procesamiento digital de señales pueden ser aplicadas para mejorar la calidad de la señal de voz. Como ejemplos se incluyen aplicaciones para remover la reverberación (o ecos) de la voz, o remover ruido de la señal de voz, o la restauración de la voz grabada en una mezcla de Helio-Oxígeno como la usada por los buzos.

1.9 Ayuda para los incapacitados

Una variedad de técnicas de procesamiento de señal han sido aplicadas para diseñar ayudas sensoriales y exhibidores visuales de información por voz como ayuda para enseñar a la persona sorda a hablar. En el Jet Propulsion Laboratory de California uno de los importantes centros espaciales en los Estados Unidos, se desarrolló una silla de ruedas que obedece ordenes orales. El parálítico puede ponerla en movimiento, detenerla, cambiar su velocidad y dirección diciendole lo que desea y controlandola con la voz, ver figura (1.7).

Por lo visto anteriormente es claro que existen una amplia variedad de tópicos, y cubrir cada uno de ellos con profundidad es extremadamente difícil, sin embargo, con la

semblanza anterior espero sirva para despertar la inquietud sobre el procesamiento digital de señales a la persona que lea esta tesis.



Fig 1.7 Silla de ruedas controlada por la voz

CAPITULO II

FUNDAMENTOS

FUNDAMENTOS DEL PROCESAMIENTO DIGITAL DE SEÑALES DE VOZ

2.1 Introducción

En la producción de la voz el sonido es generado en el sistema vocal, por la vibración de las cuerdas vocales, o también por la creación de una constricción en el flujo de aire. Los sonidos que son producidos están espectralmente formados por las características de transmisión del tracto vocal, que consiste de la laringe y la cavidad vocal, aunque en algunos casos el canal nasal también es implicado.

El proceso de la creación del sonido puede ser modelado como un sistema lineal variante en el tiempo, excitado por una fuente de señales cuasi-periódicas en los casos de voz sonora; o en el caso de voz sorda la excitación será una fuente de señales aleatorias con espectro plano en el dominio de la frecuencia.

El "análisis de voz" es el proceso de estimar los parámetros (variantes en el tiempo) del modelo de producción de voz, a partir de la señal de voz salida de este modelo.

La "síntesis de voz" es el proceso de producción de una señal acústica por controlar el modelo de producción de voz con el grupo de parámetros (variantes en el tiempo) apropiados. Si el modelo es lo suficientemente preciso y los parámetros son determinados con buena exactitud la salida resultante del modelo es para la mayoría de los casos indistinguible de la voz natural.

El análisis de voz es un ingrediente fundamental de casi todos los problemas técnicos de la comunicación por voz, cuando el sistema de análisis de voz es acoplado a su correspondiente sistema de síntesis de voz, forma un vocoder, que es un medio de transmisión y almacenaje eficiente de señales de voz.

Los sistemas de análisis de voz sirven como principio para los sistemas de reconocimiento automático de voz, para los sistemas de verificación automática de voz, etc. Así como hay muchas áreas de aplicación, hay muchas técnicas de análisis de voz, cada una basada en un modelo particular de producción de voz, aplicable a un grupo particular de problemas.

A pesar de las muchas investigaciones realizadas sobre los modelos y técnicas de análisis/síntesis de voz, los principios básicos no han cambiado de forma notable. Realmente lo que ha cambiado es la tecnología que puede ser aplicada a los problemas de comunicación por voz.

Hace veinticinco años el análisis/síntesis de voz fué hecho con hardware analógico, pero en los años 1960's los investigadores cambiaron del hardware analógico a las computadoras digitales para facilitar el diseño de sistemas más complejos através de la simulación en computadora. Al ir mejorando las computadoras y bajando su precio, los investigadores vieron que el procesamiento digital tenía muchas ventajas sobre las técnicas analógicas.

Hoy las técnicas usadas en tecnología y comunicación de voz son predominantemente digitales.

2.2 Descripción de los sistemas en el tiempo

En casi todas las situaciones donde hay procesamiento de información ó comunicación, es natural comenzar con una representación de la señal como un modelo de variación continua; la onda acústica en la voz humana es de esta naturaleza. En forma matemática representaremos este modelo como una función de la variable continua t , que representa el tiempo, y a la señal que varía su forma de onda de manera continua como $X_a(t)$, es también posible representar la señal $X_a(t)$ como una secuencia de números; en general designaremos a las secuencias como $x(n)$.

Si vemos en señales de voz, una secuencia puede ser pensada como una secuencia de muestras de la señal analógica tomadas periódicamente con un período de muestreo T , y es usual usar la notación $X_a(nT)$. La figura 2.1 muestra un ejemplo de una señal de voz representada como una señal analógica y como una secuencia de muestras a una tasa de muestreo de 8 Khz.

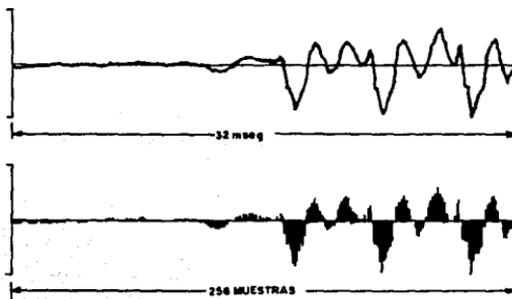


Fig 2.1 Representaciones de una señal de voz

Para el estudio de sistemas de procesamiento digital de señales de voz, encontraremos repetidamente algunas secuencias, por esto es importante conocerlas, algunas de ellas son mostradas en la figura 2.2

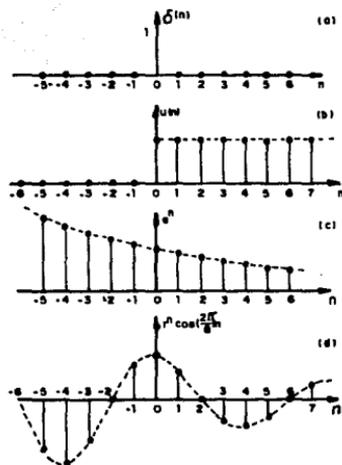


Fig 2.2 a)muestra unitaria b)escalón unitario
c)exponencial real d)coseno amortiguado

La secuencia impulso se define como

$$\begin{aligned} \delta(n) &= 1 & n &= 0 & (2.1) \\ &= 0 & & \text{en otros casos} \end{aligned}$$

La secuencia escalón unitario es

$$\begin{aligned} u(n) &= 1 & n &\geq 0 & (2.2) \\ &= 0 & n &< 0 \end{aligned}$$

Una secuencia exponencial es

$$X(n) = a^n \quad (2.3)$$

y si a es un número complejo entonces

$$a = r e^{j\omega_0}$$

y tenemos que

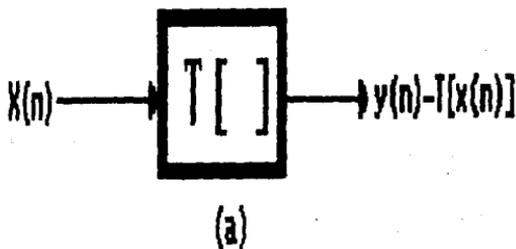
$$x(n) = r^n e^{j\omega_0 n} = r^n (\cos \omega_0 n + j \sin \omega_0 n) \quad (2.4)$$

Si $r=0$ y $\omega_0 \neq 0$, $X(n)$ es entonces una senoide compleja; si $\omega_0=0$, $X(n)$ es real; y si $r < 1$ y $\omega_0 \neq 0$, entonces $X(n)$ es una secuencia oscilatoria amortiguada exponencialmente. Las secuencias de este tipo se utilizan

especialmente en la representación de los sistemas lineales y en el modelado de la forma de onda de la señal de voz.

El procesamiento de señales implica la transformación de una señal a una forma que es más deseable, esto visto como un sistema discreto es la transformación de una secuencia de entrada en una secuencia de salida, esto se puede observar en la figura 2.3.a.

Muchos sistemas de análisis de voz son diseñados para estimar varios parámetros variantes en el tiempo a partir de la señal de voz, éstos sistemas tienen varias salidas, lo que quiere decir que a partir de una secuencia de entrada que representa la señal de voz, se obtiene un vector de secuencias como el que se muestra en la figura 2.3.b.



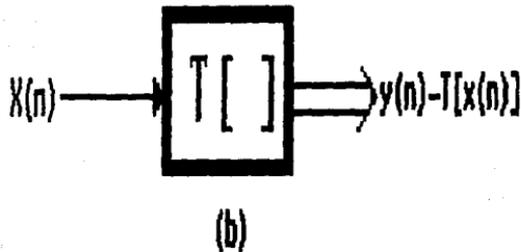


Fig 2.3 Diagrama a bloques de los sistemas de
a) una entrada/una salida b) una entrada/varias salidas

Los sistemas lineales invariantes en el tiempo son usados en el procesamiento de señales de voz, estos sistemas son completamente caracterizados por su respuesta impulso, la salida es calculada de la entrada $x(n)$ y la respuesta impulso $h(n)$.

$$y(n) = \sum_{k=-\infty}^{\infty} x(k)h(n-k)$$

$$y(n) = x(n) * h(n) \quad (2.5.a)$$

Una expresión equivalente es

$$y(n) = \sum_{k=-\infty}^{\infty} h(k) x(n-k)$$

$$y(n) = h(n) * x(n) \quad (2.5.b)$$

Los sistemas lineales invariantes en el tiempo son usuales en las operaciones de filtrado de señales de voz, pero un uso más importante es como modelos de producción de voz.

2.3 Transformadas de las señales

El análisis y diseño de los sistemas lineales se facilita al hacer la transformación de las señales y los sistemas al dominio de la frecuencia. Para comprender el porque se facilita el análisis y diseño de los sistemas lineales es necesario revisar la transformada z y la transformada de Fourier.

La transformada z de una señal $x(n)$ comparte muchos atributos con la transformada de Laplace de $x(t)$, el papel que juega la transformada z en los sistemas discretos en el tiempo es bastante similar al de la transformada de Laplace en los sistemas continuos en el tiempo.

Existen muchos teoremas y propiedades de la transformada z que son usuales en el estudio de sistemas discretos en el tiempo. Una lista de los teoremas más importantes es dado en el apéndice B de esta tesis.

La transformada z de una secuencia es definida por las ecuaciones

$$x(z) = \sum_{n=-\infty}^{\infty} x(n)z^{-n} \quad (2.6.a)$$

$$x(n) = \frac{1}{2\pi j} \int_C x(z) z^{n-1} dz \quad (2.6.b)$$

En estas ecuaciones la transformada z de $x(n)$ se define de acuerdo con la ecuación (2.6.a), se puede ver que en general $x(z)$ es una serie de potencias infinita de la variable z^{-1} , y las $x(n)$ representan los coeficientes en esta serie de potencias.

En general esta serie convergera a un valor finito solamente para ciertos valores de z .

Una condición suficiente para la existencia de la convergencia es

$$\sum_{n=-\infty}^{\infty} |x(n)| |z^{-n}| < \infty \quad (2.7)$$

El grupo de valores en los que la serie converge define una región en el plano z , llamada región de convergencia. Esta región de convergencia tiene la forma

$$R_1 < |z| < R_2 \quad (2.8)$$

La transformada inversa es dada por la ecuación 2.6.b, donde c es un contorno cerrado que contiene el origen del plano z y encierra la región de convergencia $x(z)$.

Aunque los teoremas de la transformada z parecen ser de forma similar a los teoremas de Laplace correspondientes usados en las funciones continuas en el tiempo, esta similitud no significa que la transformada z es en todo sentido una aproximación de la transformada de Laplace.

En realidad debemos entender que la transformada de Laplace es una representación exacta de las funciones continuas en el tiempo, y la transformada z es una representación exacta de una secuencia de números.

El método para relacionar las representaciones continuas y discretas de una señal es a través del teorema de muestreo.

Para usar los métodos del procesamiento digital de señales sobre una señal analógica, es necesario representar esta señal como una secuencia de números.

Esto se hace muestreando la señal analógica $x_a(t)$ periódicamente para producir la secuencia

$$x(n) = x_a(nT) \quad -\infty < n < \infty \quad (2.9)$$

donde n toma únicamente valores enteros. La figura 2.1 muestra la forma de onda de una señal de voz, y sus muestras correspondientes, con un período $T = 1/8000$ seg.

Veamos ahora la transformada de Fourier, su representación esta dada por las ecuaciones

$$x(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x(n) e^{-j\omega n} \quad (2.10.a)$$

$$x(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} x(e^{j\omega}) e^{j\omega n} d\omega \quad (2.10.b)$$

Estas ecuaciones pueden parecer un caso especial de las ecuaciones 2.6, pero de manera específica la transformada de Fourier se obtiene al restringir la transformada z al círculo unitario en el plano z, esto es, dejando que $z=e^{j\omega}$.

En la figura 2.4 la frecuencia digital variable ω tiene una interpretación de un ángulo en el plano z.

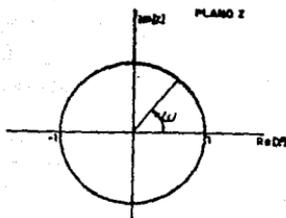


Fig 2.4 Circulo unitario en el plano z

La condición suficiente para la existencia de la transformada de Fourier se obtiene dejando que $|z|=1$ en la ecuación 2.7, obteniéndose como resultado

$$\sum_{n=-\infty}^{\infty} |x(n)| < \infty \quad (2.11)$$

La transformada de Fourier se obtiene al sustituir $z=e^{j\omega}$ en las ecuaciones de la transformada z . Algo que es importante resaltar de la transformada de Fourier de una secuencia es que $X(e^{j\omega})$ es una función periódica de ω , con un período de 2π . Esto se demuestra fácilmente sustituyendo ω por $\omega+2\pi$ en la ecuación 2.10.a.

En forma paralela, dado que $X(e^{j\omega})$ es la evaluación de $x(z)$ en el círculo unitario, vemos que $X(e^{j\omega})$ deberá volverse a repetir cada vez que se completa una vuelta alrededor del círculo unitario, esto es equivalente a que ω avance 2π radianes.

Si en la tabla del apéndice B de esta tesis sustituimos $z=e^{j\omega}$ en cada uno de los teoremas, obtenemos los teoremas correspondientes de la transformada de Fourier, aunque los resultados de realizar estas sustituciones son válidas únicamente si la transformada de Fourier realmente existe.

Con lo visto anteriormente ahora analicemos la transformada discreta de Fourier (DFT). Como en el caso de las señales analógicas, si tenemos una secuencia periódica con un período N , esto es

$$\bar{x}(n) = \bar{x}(n+N) \quad -\infty < n < \infty \quad (2.12)$$

entonces 2.12 se puede representar por una sumatoria de senoides mejor que por una integral como se hizo en la ecuación 2.10.b.

Una secuencia de longitud N puede representarse exactamente por una transformada discreta de Fourier de la forma

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi}{N}kn} \quad (2.13.a)$$

$$k = 0, 1, \dots, N-1$$

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{j\frac{2\pi}{N}kn} \quad (2.13.b)$$

$$n = 0, 1, \dots, N-1$$

Es importante tener en mente cuando se usa la transformada discreta de Fourier que todas las secuencias se comportan como si fueran secuencias periódicas cuando usamos

la transformada de Fourier.

La transformada discreta de Fourier es muy importante, por las siguientes razones

1.- La DFT, $X(k)$, se puede ver como una versión muestreada de la transformada z (o transformada de Fourier) de una secuencia de longitud finita.

2.- La DFT tiene propiedades similares (haciendo algunas modificaciones debido a su periodicidad inherente) a muchas de las propiedades de la transformada z y de la transformada de Fourier.

3.- Los N valores de $X(k)$, pueden ser calculados de manera eficiente (con un tiempo proporcional a $N \log(N)$) por los algoritmos conocidos como la transformada rápida de Fourier (FFT).

La transformada discreta de Fourier es ampliamente usada para calcular el espectro, las funciones de correlación y para implementar filtros digitales.

Con frecuencia la transformada discreta de Fourier también se usa en el procesamiento digital de señales de voz.

2.4 Muestreo

Para poder utilizar los métodos de procesamiento digital en señales como la voz, es necesario representar esta señal como una secuencia de números. Esto es hecho al muestrear la señal analógica $x_a(t)$ en forma periódica produciéndose la secuencia

$$x(n) = x_a(nT) \quad -\infty < n < \infty \quad (2.14)$$

con n tomando únicamente valores enteros.

Las preguntas que surgen de inmediato son : qué valor deberá tener la velocidad de muestreo? Hay alguna limitación en la velocidad a la que se puede muestrear? Podría tenerse la idea intuitiva de que el proceso de muestreo distorciona irremediablemente la señal original $x_a(t)$.

Sin embargo, de la figura 2.5 podemos extraer una conclusión muy importante en relación a la la frecuencia de muestreo y la máxima frecuencia de la señal muestreada.

Para poder extraer la señal original a partir de la señal muestreada, la primera banda lateral (o máxima frecuencia de f_m) no debe superponerse a la segunda banda lateral (o mínima frecuencia $f_s - f_m$), de lo contrario, el filtro pasabajas no

será capaz de eliminar por completo toda la segunda banda lateral.

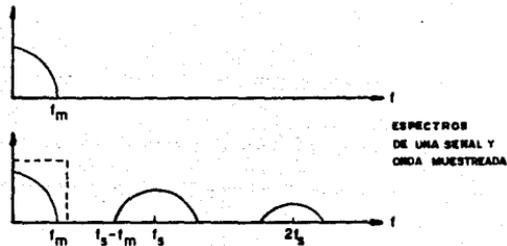


Fig 2.5 Espectros de una señal y
onda muestreada

Esto significa que la frecuencia mínima de la segunda banda lateral debe ser mayor que la máxima frecuencia de la primera banda lateral, es decir, debe de ser $f_m < f_s - f_m$, de donde :

$$f_s > 2f_m$$

Por lo tanto la frecuencia de muestreo de f_s debe ser por lo menos el doble de la máxima frecuencia de la señal para permitir la reconstrucción de la señal de filtrado.

Este principio es denominado Teorema de muestreo y es la base para la comunicación digital, este teorema conduce a la llamada velocidad de Nyquist.

Si se muestrea la señal de la figura 2.5 a una velocidad menor que la velocidad de Nyquist, se obtienen los espectros de la figura 2.6.

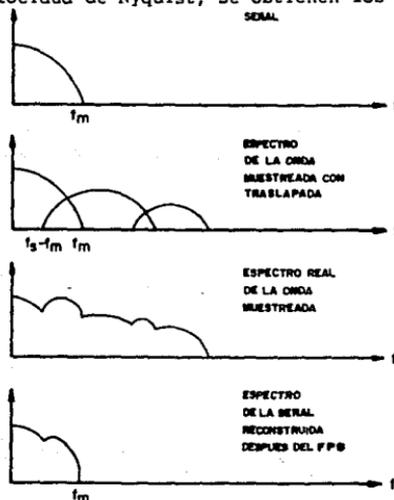


Fig 2.6 Espectros resultantes del muestreo a una frecuencia menor que la de Nyquist

- a) Señal b) Espectro de la onda con traslape
c) Espectro real de la onda muestreada

d) Espectro de la señal reconstruida despues del FPB

El resultado es que la frecuencias que normalmente estan fuera de la banda de frecuencias de la señal han sido desplazadas dentro de la banda de frecuencias y la señal será distorcionada (aliasing).

Dado que los filtros pasabajas (FPB) en la practica no son ideales y no tienen un punto de corte abrupto, la velocidad de muestreo necesaria en el sistema práctico es mayor que la velocidad de Nyquist. Por eso para muestrear una señal de voz de voz de frecuencia máxima de 3.4 Khz se emplea una velocidad de muestreo cercana a los 8 Khz.

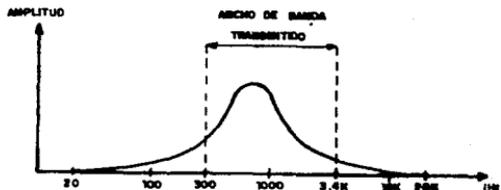


Fig 2.7 Espectro de una señal de voz promedio

El teorema de muestreo nos asegura que la representación de la secuencia muestreada sea única, esto es, que no exista otra secuencia que tenga la misma representación, esto quiere decir que no pasen a través de los mismos puntos de muestreo las dos secuencias.

El efecto de dos señales senoidales teniendo la misma representación se muestra en la figura 2.8. El término que describe esta confusión de identidad de las señales se llama "aliasing".

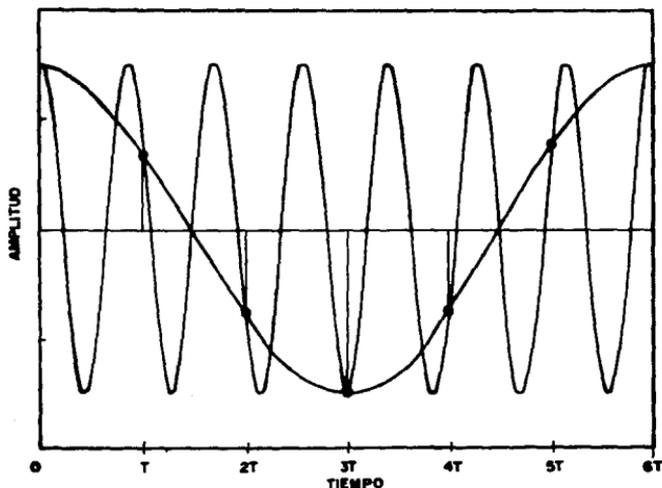


Fig 2.8 Dos señales senoidales muestreadas con período T teniendo la misma representación

La ambigüedad de esta situación se remueve al tener una frecuencia de muestreo de al menos dos veces la frecuencia de la señal.

2.5 Proceso de producción de voz

Para aplicar las técnicas del procesamiento digital de señales a los problemas de comunicación por voz, es necesario comprender los fundamentos de la producción de la señal de voz, en esta parte de la tesis obtendremos algunos modelos discretos (en el tiempo) para representar las señales de voz muestreadas.

Las técnicas básicas para el análisis y síntesis de la voz se pueden ver en términos de un modelo de forma de onda de la señal de voz, como una respuesta de un sistema que varía lentamente en el tiempo a una excitación periódica o una señal de ruido blanco Gaussiano.

Las señales de voz están compuestas de una secuencia de sonidos y la transición entre ellos sirve como una representación simbólica de la información. El arreglo de estos sonidos (símbolos) se gobierna por las reglas de cada uno de los lenguajes. Así, el estudio de estas reglas y sus implicaciones en la comunicación humana es por medio de la lingüística, y la clasificación y estudio de los sonidos de voz se hace por medio de la fonética.

En esta tesis no se hará una discusión detallada de la fonética y la lingüística, solo se analizará lo que es básicamente necesario para conocer la estructura de la señal de voz. Sin embargo, en el procesamiento digital de señales de voz con el fin de mejorar la calidad de la señal de voz o extraer la información de esta, es de gran ayuda conocer tanto como sea posible la estructura de la señal de voz, esto es, el modo en que la información es codificada en la señal de voz.

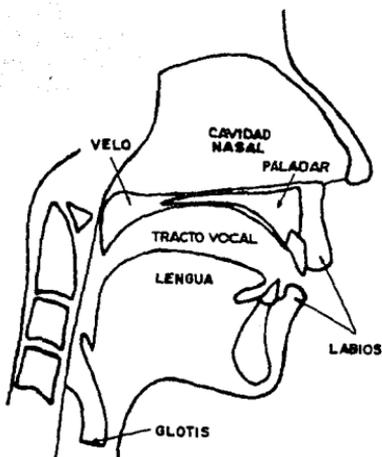


Fig 2.9 Vista en corte del mecanismo vocal humano

El mecanismo de producción de voz consiste en forma esencial de un tubo acústico, el tracto vocal, el cual es excitado por una fuente apropiada para generar el sonido deseado.

Este comienza entre las cuerdas vocales y finaliza en los labios, y consiste de la faringe, que es la conexión del esófago a la boca, y la cavidad oral. El tracto nasal comienza en el velo del paladar y finaliza en la nariz, cuando el velo es bajado, el tracto nasal es acoplado con el tracto vocal produciendo los sonidos nasales de la voz.

Estos sonidos de voz producidos pueden ser clasificados en dos diferentes clases dependiendo del modo de excitación. Los sonidos sonoros son creados al forzar el aire a través de la glotis con la tensión de las cuerdas vocales ajustada de manera que ellas puedan vibrar en una oscilación relajada, produciéndose pulsos de aire cuasi-periódicos que excitan el tracto vocal; como ejemplo tenemos las letras U,D,W,I,E.

Los sonidos fricativos o sordos son generados por formar una constricción en algún punto del tracto vocal (usualmente hacia el final de la boca), y forzando el aire a través de la constricción con bastante velocidad para producir turbulencia, creándose una fuente de ruido con espectro plano en el dominio de la frecuencia. La figura 2.10 es un ejemplo de una señal de voz conteniendo estos dos tipos de sonidos.

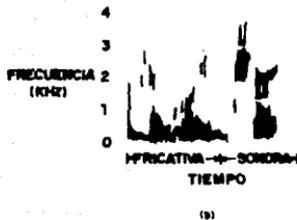
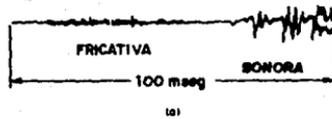


Fig 2.10 Ejemplo de una señal de voz mostrando diferentes clases de sonidos

Los tractos vocal y nasal en corte de sección de área no uniforme se muestran en la figura 2.11.

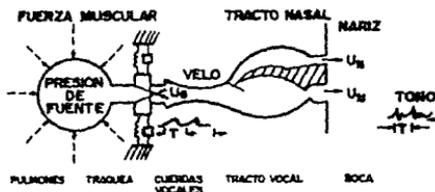


Fig 2.11 Diagrama esquematizado de un aparato vocal

El sonido generado, tal como se mencionó antes, se propaga a través de estos tubos, donde el espectro de frecuencia es formado por la selectividad de frecuencias del tubo.

Este efecto es muy similar a los efectos de resonancia observados en los órganos de tubos o en los instrumentos de viento. Para el contexto de la producción de voz, las frecuencias resonantes del tubo del tracto vocal son llamadas frecuencias de los formantes (formants) o simplemente formantes. Cada forma y dimensión del tracto vocal se caracteriza por su grupo de frecuencias formantes, de esta forma, las propiedades espectrales de la señal de voz varían en el tiempo así como la forma del tracto vocal varía.

Las características espectrales variantes en el tiempo de la señal de voz pueden ser mostradas usando un espectrógrafo de sonido. Este aparato produce un patrón en dos dimensiones llamado espectrograma, en el que el eje vertical corresponde a la frecuencia y el eje horizontal al tiempo.

En este patrón las zonas oscuras son proporcionales a la energía de la señal. Las frecuencias de resonancia del tracto vocal se muestran como bandas oscuras en el espectrograma.

En el espectrograma las regiones sonoras se caracterizan por tener una apariencia estriada, mientras que los intervalos de voz fricativa (sorda) son más llenos. El espectrograma de la expresión de la figura 2.10.a se muestra en la figura 2.10.b, comparando estas dos figuras se pueden correlacionar sus rasgos distintivos en el dominio de la frecuencia y del tiempo.

Los lenguajes también se pueden describir por medio de grupos de sonidos distintos llamados fonemas, estos incluyen vocales, diptongos, semivocales y consonantes. Cada uno de los fonemas puede ser clasificado como sonido continuo o discontinuo. Los sonidos continuos se producen con una configuración fija del tracto vocal exitada por una fuente apropiada.

Los sonidos continuos incluyen vocales y sonidos fricativos. En cambio los sonidos discontinuos son producidos por cambiar la configuración del tracto vocal, estos incluyen diptongos y semivocales.

2.6 Teoría acústica de la producción de voz

En la sección anterior se vió una descripción cualitativa de los sonidos de la voz y el modo en que estos se producen. Sin embargo para el análisis y síntesis de la voz es necesario tener una representación matemática de los procesos de análisis y síntesis de la voz.

Para realizar esta representación veamos lo que es la propagación del sonido.

El sonido es casi de hecho un sinónimo de la vibración. Las ondas de sonido se crean por la vibración y se propagan en el aire u otros medios por las vibraciones de las partículas

del medio. Las leyes de la Física sirven para describir la generación y propagación del sonido en el tracto vocal, entre estas leyes son importantes las leyes de la conservación de la energía, del momento y de la masa, así como las leyes de la termodinámica y la mecánica de fluidos. Todas estas leyes para nuestro caso las aplicaremos al fluido aire, que es el medio de la propagación del sonido de la voz.

Con estos principios físicos obtendremos un grupo de ecuaciones diferenciales para describir el movimiento del aire en el tracto vocal. La formulación y solución de estas ecuaciones es muy difícil, excepto para asepciones muy simples de la forma del tracto vocal y las pérdidas de energía en el sistema vocal.

Un sistema muy detallado de la teoría acústica debe de considerar los siguientes efectos:

- 1.- Variación en el tiempo de la forma del tracto vocal.
- 2.- Pérdidas debidas a la conducción de calor y la fricción viscosa en las paredes del tracto vocal.
- 3.- Blandura de las paredes del tracto vocal.
- 4.- La radiación del sonido en los labios.
- 5.- Acoplamiento nasal.
- 6.- La excitación del sonido en el tracto vocal.

Sin embargo una teoría que tome en cuenta todos estos efectos no existe aún , pero una discusión cualitativa dara una idea aproximada del comportamiento acústico del tracto vocal.

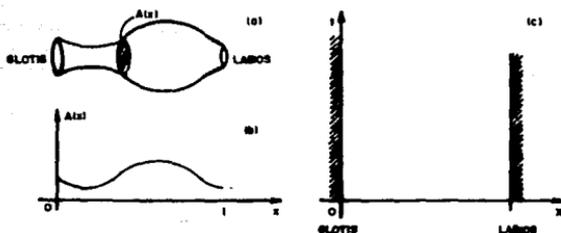


Fig 2.12 a) esquema del tracto vocal

b) función de área correspondiente

c) plano $x-t$ para la solución de la ecuación de onda

La configuración más simple para interpretar los términos de producción de voz se muestra en la figura 2.12.a. El tracto vocal es modelado como un tubo con área de corte no uniforme. Para las frecuencias de longitudes de onda comparables con las dimensiones del tracto vocal (menos de 4000 hz), se puede asumir que el plano de propagación de onda

está a lo largo del eje del tubo, otra asepción para simplificar el modelo es suponer que no existen pérdidas debidas a la viscosidad o conducción termal, en el volumen del fluido o en las paredes del tubo.

Apoyado en estas asepciones y las leyes de la conservación de la masa, momento y energía, M.R. Portnoff en su tesis de maestría " A Quasi-One-Dimensional Digital Simulation for the Time-Varying vocal tract" en el Instituto Tecnológico de Massachusetts, mostro que las ondas de sonido en el tubo satisfacen las ecuaciones:

$$-\frac{\delta p}{\delta x} = \rho \frac{\delta \left(\frac{u}{A}\right)}{\delta t} \quad (2.14.a)$$

$$-\frac{\delta u}{\delta x} = \frac{1}{\rho c^2} \frac{\delta(pA)}{\delta t} + \frac{\delta A}{\delta t} \quad (2.14.b)$$

donde:

$p = p(x,t)$ es la variación de la presión del sonido en la posición x y en tiempo t .

$u = u(x,t)$ es la variación en la velocidad del volumen del flujo en la posición x y en el tiempo t .

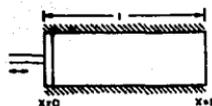
ρ es la densidad del aire en el tubo.

c es la velocidad del sonido

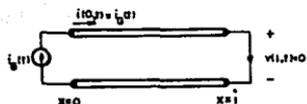
$A = A(x,t)$ es la función de área del tubo, esto es, el valor del área en el corte normal al eje tubo como una función de la distancia a lo largo del tubo y como función del tiempo.

Las soluciones de estas ecuaciones no son posibles de resolver en forma directa, excepto para la configuración más simple, sin embargo estas ecuaciones se pueden resolver por medio de los métodos numéricos.

Como un ejemplo de aplicación analizemos para el caso de un tubo uniforme sin pérdidas. Para este caso la función de área es constante para x y t (invariante en el tiempo con sección de corte transversal uniforme). La figura 2.13.a muestra un tubo con sección de corte transversal uniforme excitado por una fuente de velocidad de volumen de un flujo, se hace la asepción que en la salida del tubo no hay variaciones de la presión del aire.



(a)



(b)

Fig 2.13 a) tubo uniforme sin pérdidas con terminaciones ideales
b) línea de transmisión analógica correspondiente

Si $\Lambda(x,t)=\Lambda$ una constante, entonces las ecuaciones diferenciales parciales se reducen a:

$$-\frac{\delta p}{\delta x} = \frac{\rho}{\Lambda} \frac{\delta u}{\delta t} \quad (2.15.a)$$

$$-\frac{\delta u}{\delta x} = \frac{\Lambda}{\rho c^2} \frac{\delta p}{\delta t} \quad (2.15.b)$$

Resolviendo estas ecuaciones, y usando la analogía de la figura 2.13.b vemos que el tubo acústico uniforme se comporta como una línea de transmisión sin pérdidas terminada en un corto circuito en el final de la línea y excitado por una fuente de corriente, esto se muestra en la figura 2.13.b.

Si hacemos la representación de este sistema en el dominio de la frecuencia y resolvemos la función de transferencia de la relación de velocidades entre la entrada y la salida, veremos que esta tiene un número infinito de polos igualmente espaciados en el eje de frecuencias. Esto se muestra en la figura 2.14.a, y en forma matemática en la ecuación 2.16.

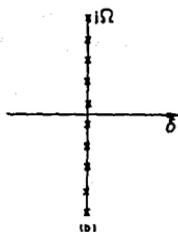
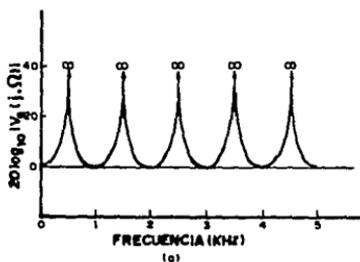


Fig 2.14 a) respuesta en frecuencia

b) localización de los polos para un tubo uniforme
sin pérdidas

Los polos en la figura 2.14.b corresponden a las frecuencias de resonancia de el sistema (ecuación 2.16), y por supuesto estas frecuencias de resonancia son llamadas frecuencias de los formantes cuando se considera el proceso de producción de voz. Como se vera más adelante los efectos de resonancias similares se observan al considerar la forma del tracto vocal.

La respuesta en frecuencia es una caracterización conveniente del modelo del sistema vocal. Con este modelo acústico muy simple de respuesta en frecuencia usado para la producción de voz, se pueden comenzar a considerar modelos más reales.

$$V_a(s) = \frac{2e^{-\frac{sL}{c}}}{1 + e^{-\frac{2sL}{c}}} \quad (2.16)$$

donde:

L = longitud del tubo

c = velocidad del sonido

La ecuación 2.16 muestra la solución de las ecuaciones 2.15 para el caso de un tubo uniforme sin pérdidas.

Ahora veamos como se comporta este modelo cuando consideramos que existen pérdidas en el tracto vocal. La ecuación 2.16 fue desarrollada considerando que no existían pérdidas en el tubo, pero en la realidad una parte de energía se pierde como resultado de la fricción viscosa entre el aire y las paredes del tubo. Si queremos incluir este efecto podríamos tomar las leyes básicas de la física y desarrollar otras ecuaciones nuevas, pero hacer esto es muy difícil por la dependencia de las pérdidas con la frecuencia.

Una forma de resolver este problema es modificar la representación del dominio de la frecuencia. En esta tesis no haremos el desarrollo de las ecuaciones por ser muy extenso, nos concretaremos a plantear y discutir cualitativamente lo que nos indican cada una de las ecuaciones y sus gráficas.

Primero consideraremos los efectos de la vibración de las paredes del tracto vocal. Las variaciones de la presión del aire dentro del tracto vocal causan que las paredes experimenten fuerzas variantes. Si consideramos que las paredes son elásticas, entonces el área del corte transversal del tubo cambiará dependiendo de la presión que exista en este.

Si asumimos que el área reacciona localmente, entonces el área $A(x,t)$ será función de $p(x,t)$. Por esto mismo las variaciones de presión serán pequeñas, y podemos tratarlas como una pequeña perturbación del área nominal, esto es:

$$A(x,t) = A_0(x,t) + \gamma A(x,t) \quad (2.17)$$

Donde $A_0(x,t)$ es el área nominal y $\gamma A(x,t)$ es una pequeña perturbación, esto se muestra en la figura 2.15.



Fig 2.15 Efectos de la vibración de la pared del tubo

Las ecuaciones que describen la propagación del sonido en una pared blanda, como el tracto vocal son:

$$\frac{\delta p}{\delta x} = \rho \frac{\delta \left(\frac{u}{\delta t} \right)}{\delta t} \quad (2.18.a)$$

$$\frac{\delta u}{\delta x} = \frac{1}{\rho c^2} \frac{\delta (p \Delta o)}{\delta t} + \frac{\delta \Delta o}{\delta t} + \frac{\delta (\gamma \Delta)}{\delta t} \quad (2.18.b)$$

Resolviendo la función de transferencia de la relación de velocidades entre la entrada y la salida ($V_a(s) = V_{entrada} / V_{salida}$) y considerando un tubo uniforme de 17.5 cm (como es el caso de la longitud típica del tracto vocal) obtenemos el diagrama de la figura 2.16.

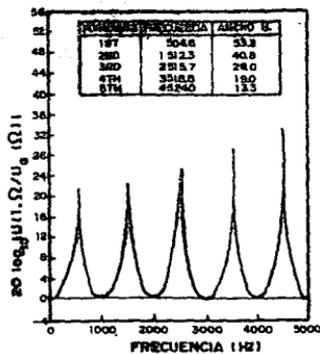


Fig 2.16 Respuesta en frecuencia de un tubo uniforme con paredes blandas y terminación en corto circuito (presión final =0)

Los resultados obtenidos en la figura 2.16 son similares a los obtenidos en la figura 2.14, pero las resonancias en el caso del tubo de paredes blandas no son infinitas como en el caso del tubo sin pérdidas.

Existen además otros efectos que se deben resaltar. En primer lugar vemos que las frecuencias centrales de resonancia son ligeramente más altas que para el caso del tubo sin pérdidas, y en segundo lugar el ancho de banda de las

resonancias no es cero como en el caso del tubo sin pérdidas, es por esto que el valor pico de la resonancia no es infinito. En el caso del tracto vocal estos efectos son debidos a las vibraciones de las paredes por las frecuencias bajas, ya que en las frecuencias altas existe muy poco movimiento de la masa de las paredes (las frecuencias centrales son ligeramente incrementadas y las resonancias de las frecuencias bajas abren su ancho de banda).

Los efectos debidos a la fricción viscosa y a la conducción térmica en las paredes son pequeños en comparación con los efectos de la vibración de las paredes. Los resultados de considerar estas pérdidas se muestran en la figura 2.17.

Las pérdidas viscosas y térmicas se incrementan con la frecuencia, teniendo su efecto más grande en las resonancias de las frecuencias altas, en cambio las pérdidas por vibración en las paredes son más pronunciadas a frecuencias bajas. En consecuencia las pérdidas por vibración de las paredes hacen subir las frecuencias de resonancia, mientras que las pérdidas térmicas y viscosas las hacen bajar.

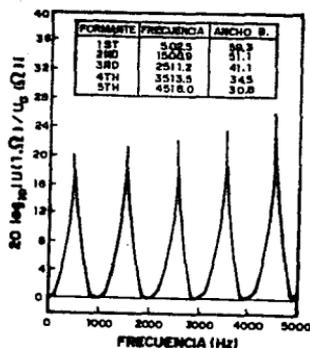


Fig 2.17 Respuesta en frecuencia de un tubo uniforme con paredes blandas, con pérdidas térmicas y por fricción, terminado en un corto circuito (presión final=0).

Por esto los efectos netos para las resonancias a frecuencias bajas es un pequeño corrimiento ascendente si se compara con los resultados del modelo de paredes rígidas y sin pérdidas. Además el efecto de las pérdidas viscosas y térmicas es pequeño comparado con las pérdidas por la vibración de las paredes (para frecuencias menores a 3-4 KHz).

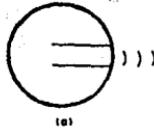
Entonces podemos desprestigiar estos efectos en la ecuación 2.18 y aún tener una buena representación de la transmisión de sonido en el tracto vocal.

En los ejemplos anteriores se asumió que la presión final en los labios era cero (en una línea analógica esto corresponde a un corto circuito).

En realidad el tubo del tracto vocal termina con una abertura entre los labios (o la nariz). Un modelo que considera este caso se muestra en la figura 2.18.a, que nos muestra los labios abiertos como un orificio en una esfera. A frecuencias bajas la abertura puede ser considerada como una superficie de radiación, con las ondas de sonido radiado difractadas por la esfera que representa la cabeza.

Los resultados de los efectos de difracción son complicados y muy difíciles de representar, sin embargo para representar las condiciones de frontera en los labios, lo único que se necesita es la relación entre la presión y la velocidad del volumen del aire en la superficie de radiación.

Si consideramos que la superficie de radiación (abertura de los labios) es pequeña comparada con el tamaño de la esfera (la cabeza), una muy razonable aproximación es asumir que la superficie de radiación es colocada en un plano de deflexión con extensión infinita, como se muestra en la figura 2.18.b.



**Fig 2.18 a)radiación desde un deflector esférico
b)radiación desde un deflector de plano infinito**

Aunque en esta tesis no se demuestra por falta de espacio. Las pérdidas por radiación en los labios son muy significativas en las frecuencias altas. La figura 2.19 muestra los resultados de respuesta en frecuencia de estas pérdidas.

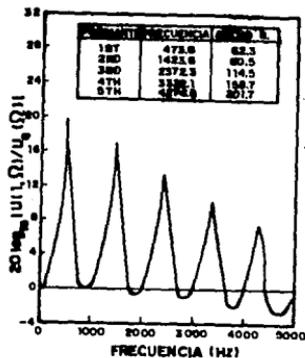


Fig 2.19 Respuesta en frecuencia de un tubo uniforme con paredes blandas, pérdidas térmicas y de fricción

Como se esperaba los mayores efectos ocurren en las frecuencias altas. El ancho de banda de la primera resonancia (formante) es determinado primordialmente por la pérdidas de las paredes, mientras que el ancho de banda del formante más alto es primordialmente determinado por las pérdidas de radiación. Los demás anchos de banda de los formantes son determinados por una combinación de estos dos mecanismos de pérdidas.

2.7 Modelos tubulares sin pérdidas

Lo visto en la sección anterior nos servirá para sacar un modelo de producción de voz basado en la asepción de que el tracto vocal puede ser representado como una concatenación de tubos acústicos sin pérdidas, esto se muestra en la figura 2.20.

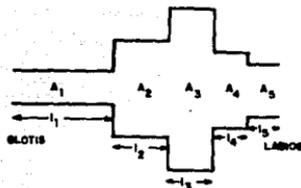


Fig 2.20 En cadenamiento de 5 tubos sin pérdidas

Las áreas de la sección de corte transversal $\{A_k\}$, de los tubos se escogen para aproximar lo más fielmente posible a la función de área $A(x)$ del tracto vocal.

Si un gran número de tubos de longitud corta se usan, entonces las frecuencias de resonancia de la concatenación de los tubos será aproximadamente la de un tubo con una función de área de variación continua.

Un hecho muy importante que se debe resaltar, es que el modelo del tubo sin pérdidas nos sirve para la transición entre los modelos continuos y los modelos discretos. Además el modelo del tubo sin pérdidas tiene muchas propiedades en común con los filtros digitales por lo que es conveniente empezar por ver un sistema compuesto de N tubos sin pérdidas para comprender esto. Cada uno de los N tubos tiene una longitud de $\Delta x = l/N$, donde l es la longitud total del tracto vocal. En este sistema las ondas de propagación se representan por retardos igual a $\tau = \Delta x/c$, que es el tiempo de propagación de la onda sobre la longitud de un tubo.

Comenzaremos por considerar la respuesta de este sistema a un impulso unitario entrada $\delta(t)$. El impulso se propagara a través de la serie de tubos, siendo en las juntas parcialmente reflejado y parcialmente transmitido. Lo más pronto que el impulso puede alcanzar la salida es $N\tau$ segundos.

Y la función de este sistema es:

$$\begin{aligned}
 V_a(s) &= \sum_{k=0}^{\infty} \alpha_k e^{-s(N+2k)\tau} \\
 &= e^{-sN\tau} \sum_{k=0}^{\infty} \alpha_k e^{-s2\tau k} \quad (2.19)
 \end{aligned}$$

Donde el factor $e^{-sN\tau}$ es el retardo de tiempo requerido para propagarse la onda a través de todas las N secciones, y la cantidad

$$V'_a(s) = \sum_{k=0}^{\infty} \alpha_k e^{-sk2\tau} \quad (2.20)$$

es la función lineal de un sistema cuya respuesta impulso es $V'_a(t) = V_a(t + N\tau)$, y representa las propiedades de resonancia del sistema.

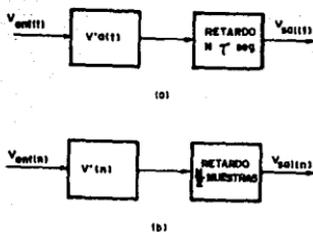


Fig 2.21 a) diagrama a bloques del modelo de un tubo acústico sin pérdidas. b) Sistema discreto equivalente

La figura 2.21.a es el diagrama a bloques del modelo del tubo sin pérdidas, mostrando la separación del sistema $V'a(t)$ de los retardos. Su respuesta en frecuencia $V'a(\Omega)$ es:

$$V'a(\Omega) = \sum_{k=0}^{\infty} \alpha_k e^{-j\Omega k 2\tau} \quad (2.21)$$

y se puede demostrar que

$$V'a\left(\Omega + \frac{2\pi}{2T}\right) = V'a(\Omega) \quad (2.22)$$

Las ecuaciones anteriores recuerdan la respuesta en frecuencia de un sistema discreto. Si la entrada al sistema es limitada en banda a frecuencias inferiores a $\pi/(2T)$, entonces podemos muestrear la entrada con un período de $T=2\tau$ y filtrar la señal de entrada con un filtro digital que tiene como respuesta impulso

$$\begin{aligned} V'(n) &= \alpha_n & n >= 0 \\ &= 0 & n < 0 \end{aligned} \quad (2.23)$$

Para un período de $T=2\tau$ el retardo $N\tau$ seg corresponde a un corrimiento de $N/2$ muestras. Así obtenemos el sistema discreto equivalente para entradas limitadas en banda que se muestra en la figura 2.21.b. Podemos notar que si N es un número par, $N/2$ es un entero y el retardo se implementa haciendo un corrimiento de la secuencia de salida del primer sistema (el caso de N impar generalmente no se

usa).

La transformada z de $V'(n)$ es

$$V'(z) = \sum_{k=0}^{\infty} \alpha_k z^{-k} \quad (2.24)$$

La gráfica de flujo de señales para el sistema discreto equivalente se obtiene de la gráfica de flujo del sistema analógico; cada variable en un nodo del sistema analógico se reemplaza por la secuencia de muestras correspondientes, y cada τ seg de retardo se reemplaza por $1/2$ de retardo de muestra ($\tau=T/2$). Esto se muestra en la figura 2.22.

Si observamos la figura 2.22.b notaremos que el retardo de la propagación se representa por una transmisión de $z^{1/2}$, y este retardo de $1/2$ de la muestra implica hacer una interpolación entre los valores de las muestras, lo cual es imposible de implementar de manera exacta. Por esto se hace la conversión de la configuración de la figura 2.22.b a la configuración de la figura 2.22.c.

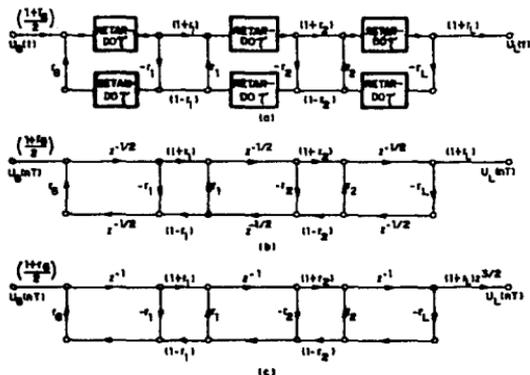


Fig 2.22 a)gráfica del flujo de señales de un modelo
 de tubo sin pérdidas para el tracto vocal
 b) sistema discreto equivalente
 c) sistema discreto optimizado

La configuración de la figura 2.22.c se deduce, al observar que el retardo alrededor de cualquier trayectoria cerrada en 2.22.b se preserva si los retardos de las ramas

inferiores son movidos arriba a sus ramas superiores correspondientes. Aunque el retardo total obtenido desde la entrada hasta la salida no es el correcto, en la práctica esto no tiene un gran efecto, y además teóricamente se puede compensar si se inserta un avance de $z^{N/2}$. En la figura 2.22.c se muestra esta compensación para la concatenación de tres tubos.

La ventaja de la forma de la figura 2.22.c es que se pueden escribir ecuaciones en diferencias para este sistema, y estas ecuaciones en diferencias se usan iterativamente para calcular las muestras de salida a partir de las muestras de entrada.

Aunque existen discusiones más profundas sobre las representaciones matemáticas del proceso de producción de la voz, básicamente se conservan los principios vistos hasta ahora.

2.8 Modelos digitales para señales para voz

Hasta ahora hemos visto como el sonido es generado en dos modos, y cada uno de estos modos da un tipo de salida, también vimos como el tracto vocal impone a la señal de excitación sus resonancias para poder producir los diferentes sonidos de voz.

Ahora es importante tener idea de la representación de las señales de voz en términos de un modelo terminal analógico como el mostrado en la figura 2.23.



Fig 2.23 Modelo de un sistema fuente para producción de voz

Este es un sistema lineal que tiene las propiedades de salida deseadas para producir voz cuando se controla por un grupo de parámetros relacionados con el proceso de producción de voz. Así este modelo es equivalente en su salida al modelo físico, pero con una estructura interna que no imita la física de la producción de voz.

Para producir señales de voz, el modo de excitación y las propiedades resonantes del sistema lineal deben de cambiar con el tiempo. En general las propiedades de la señal de voz cambian relativamente lentas, por esto para muchos sonidos de voz es razonable asumir que las propiedades de excitación y del tracto vocal permanecen fijas por períodos de 10 a 20 mseg. El modelo lineal analógico implica entonces a un sistema lineal que varía lentamente en el tiempo, excitado por una señal de excitación cuya naturaleza cambia desde pulsos cuasi-periódicos (voz sonora), hasta ruido aleatorio para voz fricativa (sorda). Los rasgos esenciales del modelo discreto del tubo sin pérdidas se muestra en la figura 2.24.

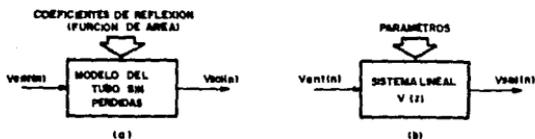


Fig 2.24 a)diagrama a bloques del modelo del tubo sin pérdidas b)modelo terminal analógico

En la figura 2.24.a el tracto vocal está caracterizado por un grupo de áreas, o en forma equivalente por los coeficientes de reflexión, como se muestra en la figura 2.22.c, donde podemos mostrar que la relación entre la entrada y la salida se puede representar por una función de transferencia:

$$V(z) = \frac{G}{1 - \sum_{k=1}^N \alpha_k z^{-k}} \quad (2.25)$$

donde G y $\{\alpha_k\}$ dependen de la función de área. En cuanto a la salida concierne cualquier sistema que tenga esta función de transferencia producirá la misma salida en respuesta a una entrada dada (esto no es estrictamente verdadero para sistemas variantes en el tiempo, pero las diferencias se pueden minimizar con una implementación cuidadosa), tomando el modelo la forma de la figura 2.24.b, conduciendonos a una implementación alternativa del filtro del tracto vocal.

La respuesta del modelo terminal analógico del tracto vocal debe de incluir una representación de los cambios de la función de excitación y los efectos de la radiación de sonido

en los labios. Por esto para desarrollar los modelos digitales para las señales de voz es bueno analizar cada una de las representaciones por separado y después combinarlas todas en un modelo completo.

2.8.1 El tracto vocal

Como ya se vió anteriormente las resonancias (formantes) de la voz corresponden a los polos de la función de transferencia $V(z)$. Pero existen dos modelos para representar el tracto vocal una con únicamente polos, el cual es una buena representación de los efectos del tracto vocal en la mayoría de los sonidos de voz, aunque la teoría acústica indica que los sonidos nasales y fricativos necesitan el otro modelo que tiene resonancias (polos) y antiresonancias (ceros).

Sin embargo B.S. Atal y S.L. Hanuer en su artículo "Speech Analysis and Syntesis by Linear Prediction of the Speech Wave", demostraron que el efecto de un cero en la función de transferencia puede sustituirse incluyendose más polos, por esto en la mayoría de los casos este planteamiento de agregar más polos a la función de transferencia es preferido para modelar el tracto vocal.

2.8.2 Radiación

Hasta este momento hemos considerado únicamente la función $V(z)$ que relaciona la velocidad del volumen del aire de la fuente, con la velocidad del volumen del aire en los labios. Pero nuestro oído realmente es sensitivo a las presiones y no tanto a las velocidades de los volúmenes, entonces necesitamos considerar un modelo que nos de la presión del aire en los labios, y en este modelo es necesario considerar los efectos de la radiación.

A frecuencias bajas la presión es aproximadamente la derivada de la velocidad del volumen del aire, así que para obtener una buena representación discreta de esta relación debemos de ser cuidadosos y usar técnicas digitales que nos eviten el traslape de colas (aliasing). La carga de la radiación se puede colocar en cascada con el modelo del tracto vocal, tal como se ve en la figura 2.25.



Fig 2.25 Modelo analógico terminal con adición de los efectos de la radiación en los labios

2.8.3 Excitación

Para completar el modelo analógico debemos generar una entrada apropiada al sistema de radiación del tracto vocal. Para esto es necesario recordar que los sonidos de voz se pueden clasificar en dos tipos, como sonoros o sordos (voiced/unvoiced), y para cada uno de estos sonidos las fuentes apropiadas son las que generan pulsos cuasiperiódicos en el caso de los sonidos sonoros, y ruido aleatorio en el caso de los sonidos sordos.

Para el caso de voz sonora, la forma de onda de la excitación deberá ser como la mostrada en la figura 2.26.

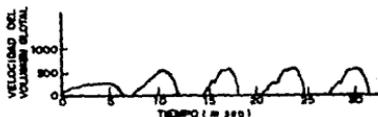


Fig 2.26 Velocidad del volumen glotal

Un modo que es conveniente para representar la generación de las formas de onda glotales se muestra en la figura 2.27.

En esta figura la generación del tren de impulsos produce una secuencia de impulsos unitarios espaciados por el tono fundamental que se desea. Después esta señal excita un sistema lineal con respuesta a impulso $g(n)$ que genera la forma de la señal glotal deseada, además se tiene un control de intensidad λ_v para esta excitación.

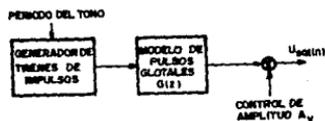


Fig 2.27 Generación de la señal de excitación para voz sonora

Para los sonidos fricativos (sordos) el modelo de excitación es más simple, y sólo se requiere de una fuente de ruido aleatorio y un parámetro de ganancia para controlar la intensidad de la excitación.

Se ha encontrado que la probabilidad de la distribución de las muestras de ruido no parecen ser críticas para las aplicaciones prácticas.

Integremos ahora cada uno de los modelos anteriores en un modelo completo como el que se muestra en la figura 2.28.

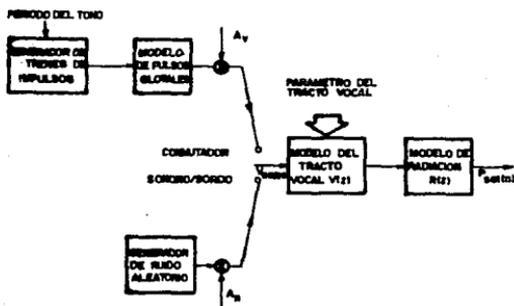


Fig 2.28 Modelo discreto completo para la producción de voz

Si dentro de este modelo conmutamos entre los generadores de excitación sonora/sorda podemos cambiar el modo de excitación de entrada al modelo. También podemos combinar los tres componentes de este modelo (pulsos glotales, radiación y

tracto vocal) y representarlos como una única función de transferencia con únicamente polos $H(z)=G(z)V(z)R(z)$, con esto se quiere remarcar que existe bastante libertad para hacer modificaciones.

Este modelo aunque es bueno, nos presenta algunas limitaciones, aunque ninguna de estas limita de manera severa su aplicabilidad práctica.

-La primera de estas limitaciones es la lenta variación de los parámetros. En los sonidos continuos como las vocales los parámetros cambian lentamente y el modelo trabaja muy bien, pero en los sonidos transitorios los parámetros no cambian tan rápido como debieran, haciendo que el modelo no tenga un comportamiento tan bueno, aunque todavía es adecuado.

En forma práctica se asume una representación de la señal de voz en segmentos de tiempo, haciendo la consideración de que estos segmentos son constantes por alrededor de 10-20 mseg (cuasiestacionarios).

-La segunda limitación es la falta de ceros en la función de transferencia del sistema, que hacen que el sistema no responda bien para los sonidos nasales.

-La tercera limitación es la dicotomía de las excitaciones sonora/sorda (voiced/unvoiced), por ser estas excitaciones inadecuadas para sonidos sonoro-fricativos.

Creo que es necesario enfatizar la necesidad de pensar estos modelos de dos maneras distintas, una como análisis de voz y otra como síntesis de voz.

En el análisis de voz el objetivo es encontrar técnicas para estimar los parámetros del modelo a partir de la señal de voz, estos parámetros serán la salida del modelo.

En la síntesis de voz, el modelo se usa para crear una señal de voz sintética por controlar el modelo con los parámetros apropiados.

Los dos puntos de vista anteriores en ocasiones pueden mezclarse, pero siempre es conveniente el pensarlos por separado.

CAPITULO III

ALGORITMOS

ALGORITMOS PARA EL PROCESAMIENTO DIGITAL DE SEÑALES DE VOZ

En la compresión del ancho de banda de la señal de voz existen dos estrategias a seguir tal como se menciona en el primer capítulo: la representación por forma de onda y la representación paramétrica.

Los algoritmos de compresión de voz basados en la representación por forma de onda tratan de eliminar las redundancias de la forma de la señal de voz, en cambio los algoritmos de compresión de voz basados en la representación paramétrica codifican únicamente los aspectos que perceptualmente son muy importantes en los procesos de la producción de la voz, y en el como escuchan las personas.

En la representación por forma de onda y en forma creciente de dificultad podemos mencionar las siguientes técnicas de procesamiento digital de señales de voz:

Modulación delta (MD)

Modulación delta adaptable (AMD)

Modulación Codificación de Pulsos Diferencial (DPCM)

Mod. Codificación Pulsos Adaptable Diferencial (ADPCM)

Codificación Adaptable por Subbandas (SBC)

3.1 Modulación Delta

Considerando el siguiente diagrama a bloques del sistema de modulación delta que se muestra en la figura (3.1).

Con la conversión Analógica/Digital (A/D), la señal analógica se muestrea en forma periódica. Los valores de las muestras, $x(n)$; $n = \dots, -2, -1, 0, 1, 2, \dots$, se comparan con un valor de muestra predicho, denominado $P_N(n)$, y la diferencia $E(n)$ se hace pasar por el cuantizador. Si $E(n)$ es pequeña durante la mayor parte del tiempo, de manera que la predicción es buena, se necesitan pocos bits para representar esta señal diferencia. El modulador delta utiliza un cuantizador de dos niveles, de manera que solo un bit se emplea para representar la señal. Estos dos niveles se llaman $\pm K$.

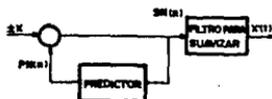
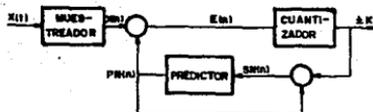


Fig 3.1

Modulación Delta: a) Transmisor b) Receptor

En el receptor, la señal diferencia cuantizada se agrega o se resta dependiendo del signo de K , a la salida del predictor para obtener la señal discreta $SN(n)$, de la señal muestreada $X(n)$. Estas estimaciones se pasan a continuación por un filtro pasa bajas con el objeto de suavizarlas y generar la estimación $X'(t)$ deseada.

Es evidente que se puede obtener un mejor desempeño usando más de dos niveles para el cuantizador. Este esquema más general se denomina modulación por pulsos diferencial (DPCM). La modulación delta es entonces un caso especial de DPCM, con solo dos niveles de cuantización.

El predictor empleado es generalmente una suma ponderada de cierto número de muestras pasadas estimadas. En éstos sistemas el predictor más simple es el que usa la estimación de la muestra previa sólo como una estimación de la muestra actual $PN(n)$. Para este caso se tiene:

$$PN(n) = h_1 SN(n-1)$$

El coeficiente h_1 es el factor de ponderación y es escogido para reducir en alguna medida el error de la estimación, para hacerlo tan pequeño como sea posible. El rendimiento del modulador delta depende del tipo de señal de entrada, la velocidad de muestreo, los niveles de cuantización, y la forma del predictor empleado.

Concentremos nuestra atención en algunas características muy generales de ellos. A causa de que la señal diferencial es cuantizada en dos niveles, aparece un ruido granular o de cuantización, similar al del PCM, que se presenta en la salida del receptor. Al contrario del PCM, donde el ruido se reduce usando más niveles de cuantización, ahora sólo puede reducirse muestreando con una mayor frecuencia. En los sistemas de

modulación delta se emplean velocidades de muestreo muy superiores a la velocidad de Nyquist que se utiliza en el PCM.

Además del ruido de cuantización, se presenta otro tipo de ruido, el cual se llama ruido de sobrecarga, y se presenta cuando los niveles de cuantización $\pm K$ son muy pequeños para seguir a una señal que varía rápidamente. Como ejemplo considerese una señal de entrada típica y sus muestras predichas de señal, como se muestra en la figura (3.7.b). Cuando se realizan las operaciones muestra por muestra, las estimaciones de las muestras $S_N(n)$ tienden a seguir la curva de $X(t)$, quedando unas veces abajo de ella, y otras, arriba como se ve en esta figura. El resultado es una señal que al recibirse sigue a la señal original, pero que introduce un ruido de cuantización.

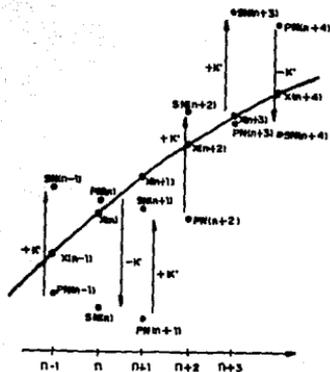


Fig 3.2 Señal típica y sus muestras predichas.

Utilizando en el receptor un circuito de retención para proporcionar algún filtrado, la señal resultante aparece como en la figura (3.3.a). Aquí se puede observar que el ruido de cuantización que se introduce es proporcional al tamaño del escalón K , por lo que al reducir K el ruido también se reduce. Sin embargo valores muy pequeños de K provocan el ruido de sobrecarga que ya se ha mencionado, lo cual se debe a la imposibilidad de que el modulador siga las grandes variaciones de la señal de entrada $x(t)$ en un intervalo pequeño de tiempo. Este fenómeno se puede observar en la figura (3.3.b).



Fig 3.3

Fuentes de ruido de la modulación delta:

a) Ruido de cuantización. b) Ruido de sobrecarga

Es obvio que existe un valor óptimo del tamaño del escalón K , este dependerá de las características de la señal de entrada $X(t)$, la velocidad de muestreo y el ruido total que pueda ser tolerado. La región de sobrecarga de la figura (3.3.b) se presenta debido a que el tamaño del escalón K fija un límite máximo a la pendiente de la señal de entrada que el modulador puede seguir.

Analizando más cuidadosamente, y teniendo una velocidad de muestreo de f_s muestras por segundo. El tiempo entre muestras es entonces $1/f_s$ segundos, como se ve en la figura (3.3.b). La máxima pendiente que se puede seguir corresponde a K unidades de amplitud en $1/f_s$ segundos, o sea, Kf_s . Esto es evidente de la región de sobrecarga de la figura (3.3.b). Pero si incrementamos K para reducir el ruido de sobrecarga, también se aumenta el ruido de cuantización. Esto da lugar para hacer una elección óptima de K .

En la figura (3.4) se muestra una figura típica del rendimiento de la modulación delta, la elección óptima del escalón K se muestra gráficamente. El rendimiento se mide en este caso como el recíproco del ruido (SNR), a medida que el ruido es reducido, el rendimiento mejorará.

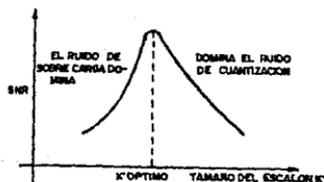


Fig 3.4 Curva de rendimiento de la modulación delta.

Un análisis cuantitativo nos hara ver mejor estos conceptos, para esto consideremos una señal senoidal de prueba:

$$X(t) = A \text{Sen} \omega_m t \quad (3.1)$$

Su máxima pendiente, en $t=0$ y en lugares espaciados por múltiplos de medio periodo ($1/2f_m$ segundos) a partir de este punto, es justamente $A\omega_m = 2\pi A f_m$, lo cual se muestra en la figura (3.5).



Fig 3.5 Señal senoidal de prueba.

Si ahora muestreamos muchas veces en un periodo:

$$f_s \gg f_m \quad (3.2)$$

Podrían usarse, diez o veinte muestras en un periodo que es una velocidad muy superior a la requerida según el teorema de

Nyquist.

Para esta señal senoidal la sobrecarga puede evitarse si Kfs es mayor o igual a la máxima pendiente:

$$Kfs > = Awm \quad (3.3)$$

Esto se muestra con frecuencia en forma normalizada al definir un tamaño relativo del escalón $K'=K/A$, es decir la razón entre el tamaño del escalón K y el máximo nivel de señal A . Para señales más reales (como la voz), con frecuencia representadas por las señales que varían en forma aleatoria, se selecciona en forma arbitraria como el máximo nivel que alcanza la señal un valor igual a cuatro veces la desviación estándar del promedio del nivel de la señal. Entonces el tamaño relativo del escalón puede definirse como $K'=4\alpha$ (α es la desviación estandar). Partiendo de esto, en términos del parámetro relativo del escalón, puede evitarse el ruido de sobrecarga para la señal senoidal al definir

$$K'fs > = Wm \quad (3.4)$$

Pero los grandes valores de K' (o de K) resultan en un ruido de cuantización notable. El ruido podría sin embargo, reducirse por medio de un muestreo a una mayor velocidad, o ya que es el ruido total el que interesa, el de cuantización podría reducirse a expensas de introducir el ruido de sobrecarga, reduciendo K' por debajo del valor crítico de

Wm/fs , válido para una onda senoidal. Este es el compromiso que se debe de adoptar (ver figura 3.4).

En la figura (3.4) podemos distinguir un rasgo muy importante, la agudeza del pico de la curva. Esto implica que el SNR es muy sensitivo al nivel de entrada, la calidad en el pico de la curva solamente puede ser mantenida sobre un estrecho rango del nivel de entrada, si el tamaño del escalón es fijo. Para alcanzar buena calidad, esto es, calidad comparable a un sistema PCM logaritmico de siete bits, para transmisión de voz se requiere una tasa de bits mucho más alta.

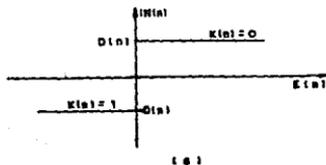
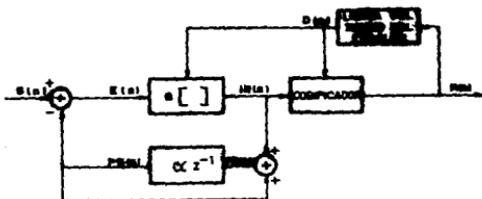
La principal ventaja de la modulación delta es su simplicidad. El sistema puede ser implementado con circuitos simples analógicos y digitales, porque solamente un bit es utilizado para codificar.

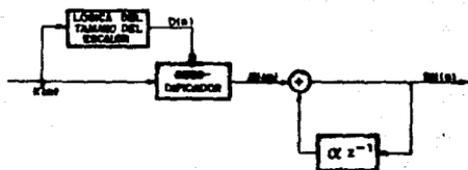
3.2 Modulación Delta Adaptable (ADM)

Del capítulo anterior, es natural suponer que los esquemas de cuantización adaptable podrían mejorar el desempeño de un modulador delta. Por esto es de gran interés este esquema que mejora la ejecución pero no incrementa de manera apreciable la complejidad del sistema modulador.

La forma general de estos sistemas se muestra en la figura (3.6).

Estos esquemas, en ausencia de errores, derivan la información del tamaño del escalón de la secuencia de la palabra código para el transmisor y el receptor.





(0)

Fig 3.6 Modulación delta con escalón adaptable.

a) codificador; b) decodificador

En este capítulo discutiremos el algoritmo de N. S. Jayant para el ADM. Como para el caso de un cuantizador multibit, el tamaño del escalón obedece la regla

$$D(n) = MD(n-1) \quad (3.5)$$

$$D_{\min} < D(n) < D_{\max} \quad (3.6)$$

En este caso, el multiplicador M es únicamente función de las palabras código presente y previa, $k(n)$ y $k(n-1)$. Puesto que $k(n)$ depende solamente del signo de $E(n)$, el cual está dado por

$$E(n) = S(n) - \alpha SN(n-1) \quad (3.7)$$

Además el signo de $E(n)$ puede ser determinado antes de la determinación del valor cuantizado actual de $IN(n)$ el cual debe esperar la determinación de $D(n)$ de la ecuación (3.5). El algoritmo para escoger el multiplicador del tamaño del escalón en la ecuación (3.5) es

$$M=P>1 \quad \text{si} \quad k(n)=k(n-1) \quad (3.8)$$

$$M=Q<1 \quad \text{si} \quad k(n) \neq k(n-1) \quad (3.9)$$

Estas ecuaciones nos indican que el tamaño del escalón se incrementará cuando las palabras código previa y presente sean iguales, y se decrementará cuando éstas palabras sean diferentes.

Esta estrategia de adaptación es motivada por los patrones de bits en la modulación delta. Ver figura (3.7). Observando la figura notaremos que los períodos de sobrecarga tienen corridas de 0's ó 1's. En los períodos de ruido granular se tienen secuencias alternantes de la forma ...0 1 0 1 0 1 ... en la figura (3.7.b) se muestra como la señal de la figura (3.7.a) podría ser cuantizada por un modulador delta

adaptable del tipo descrito. Por conveniencia los parámetros del sistema $P=2$, $Q=1/2$, $\alpha=1$ y el mínimo tamaño del escalón se muestran en la figura.

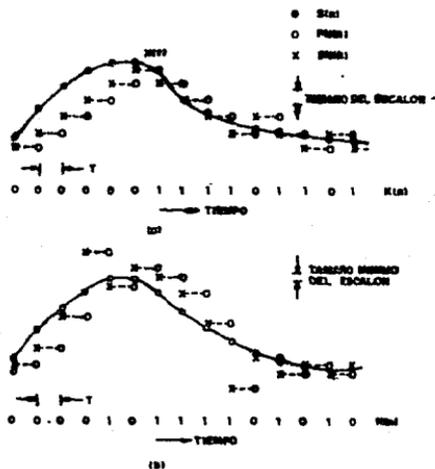


Fig 3.7 Modulación delta

a) escalón fijo b) escalón adaptable

Podría parecer que en la región de gran pendiente positiva todavía causa una corrida de 0's, pero en este caso el tamaño del escalón se incrementa exponencialmente siguiendo el incremento de la pendiente de la señal. En la región de granularidad a la derecha de la figura, esta otra vez señalado por una secuencia alternante de 0's y 1's, pero en este caso el tamaño del escalón cae rápidamente a el mínimo (D_{\min}) y permanece ahí tanto tiempo como la pendiente permanece pequeña.

Por esto el mínimo tamaño del escalón puede ser hecho mucho más pequeño que el requerido para el desempeño óptimo de un modulador delta, el ruido granular puede ser grandemente reducido. De la misma manera el máximo tamaño del escalón puede ser hecho mas grande que la máxima pendiente de la señal de entrada, reduciendo así el ruido de sobrecarga. Los parámetros de este sistema son P , Q , D_{\min} y D_{\max} . Los límites del tamaño del escalón deberan ser escogidos para dar el rango dinámico deseado. El radio D_{\max}/D_{\min} deberá ser lo suficientemente grande a fin de mantener un alto SNR sobre el rango deseado de los niveles de la señal de entrada. El tamaño mínimo del escalón deberá ser lo suficientemente pequeño para minimizar el ruido del canal desocupado. Jayant mostro que P y Q deben satisfacer la relación siguiente para tener estabilidad.

$$PQ < 1$$

$$(3.10)$$

Esto de estabilidad quiere decir mantener el tamaño del escalón a los valores apropiados para el nivel de señal de entrada (para reconstruir lo más fielmente posible la señal transmitida en el receptor). La figura (3.8) muestra los resultados de una simulación de señal de voz con $PQ=1$ para tres diferentes tasas de muestreo. De esta figura es evidente que el máximo SNR es obtenido para $P=1.5$; sin embargo, el pico de las tres curvas es muy extendido con el SNR comenzando a pocos decibeles a un máximo para

$$1.25 < P < 2$$

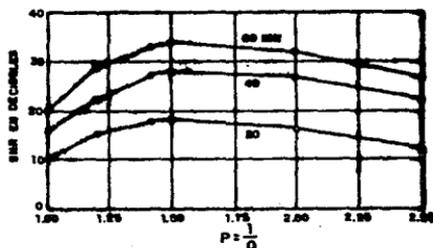


Fig 3.8 Razón señal a ruido de un modulador delta adaptable como función de p.

Los resultados de la figura (3.8) son dibujados en la figura (3.9) para comparar PCM, DM, ADM.

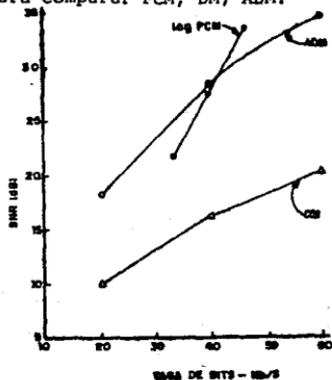


Fig 3.9 SNR vs tasa de bits para 3 esquemas de codificación usando una tasa de muestreo de 6.6 KHz.

La figura muestra que ADM es superior a DM por 8 dB a 20 kb/s y la ventaja del SNR se incrementa a 14 dB a 60 kb/s. Para el DM nosotros observamos alrededor de 6 dB de incremento cuando se dobla la tasa de muestreo, mientras que para el ADM el incremento correspondiente es de 10 dB. Si comparamos ADM con log PCM, notamos que para tasas abajo de los 40 kb/s ADM tiene mejor ejecución que que el log PCM. Para altas tasas de bits log PCM tiene un SNR más alto. Podemos ver que el ADM requiere 60 kb/s para alcanzar la misma calidad que del log PCM, teniendo este una tasa de bits de sólo 46 kb/s.

La mejorada calidad del ADM es alcanzada con poco incremento en complejidad. La adaptación del tamaño del escalón es hecha usando el flujo de bits transmitidos del sistema ADM, reteniendo la simplicidad básica de los sistemas de modulación delta. Para muchas aplicaciones ADM puede ser preferido a log PCM aún a expensas de una tasa de información ligeramente más alta.

3.3 Predictores de orden alto en la modulación delta

Por simplicidad los moduladores delta y moduladores delta adaptables usan un predictor fijo de primer orden de la forma

$$PN(n) = \alpha SN(n-1) \quad (3.11)$$

como se muestra en la figura (3.6). Para este caso la señal reconstruida satisface la ecuación

$$SN(n) = \alpha SN(n-1) + IN(n) \quad (3.12)$$

que es caracterizado por la función

$$H_1(z) = \frac{1}{1 - \alpha z^{-1}} \quad (3.13)$$

Que es el equivalente de un integrador (si $\alpha=1$).
Cuando $\alpha < 1$ se tiene un integrador con fuga.

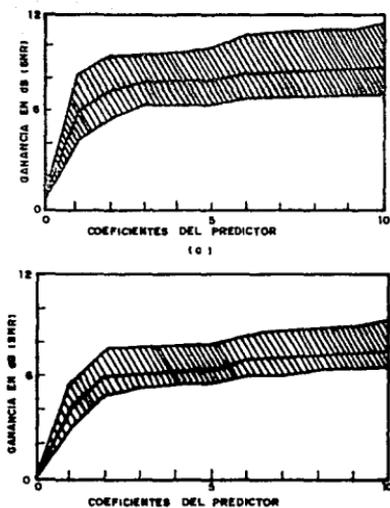


Fig 3.10 Ganancia óptima vs número de coeficientes del predictor.

a) voz filtrada en pasa-bajas; b) voz filtrada en pasa-banda

Los resultados mostrados en la figura (3.10) sugieren que para los sistemas de modulación delta, se puede obtener un gran SNR con un predictor de segundo orden, esto es con:

$$PN(n) = \alpha_1 SN(n-1) + \alpha_2 SN(n-2) \quad (3.14)$$

Y la señal reconstruida satisface la ecuación

$$SN(n) = \alpha_1 SN(n-1) + \alpha_2 SN(n-2) + IN(n) \quad (3.15)$$

que es caracterizada por función de transferencia

$$H_2(z) = \frac{1}{1 - \alpha_1 z^{-1} - \alpha_2 z^{-2}} \quad (3.16)$$

Empíricamente ha sido mostrado que un predictor de segundo orden da un desempeño mejorado cuando los polos de $H_2(z)$ son ambos reales, esto es

$$H_2(z) = \frac{1}{1 - az^{-1} - bz^{-1}} \quad 0 < a, b < 1 \quad (3.17)$$

El uso de un predictor de segundo orden con frecuencia se le llama "integrador doble". Su mejora sobre el integrador de primer orden puede ser hasta de 4 dB, dependiendo de la persona que hable.

Desafortunadamente, el uso de predicción de orden alto en ADM no es simplemente reemplazar el predictor de primer orden con un predictor de segundo orden, esto se complica porque el algoritmo de cuantización adaptable interactúa con el algoritmo de predicción. Por ejemplo para un predictor de segundo orden, el patrón de bits de para la condición de canal desocupado (sin transmitir) podría ser .. 0 1 0 1 0 1 ... ó ... 0 0 1 1 0 0 1 1 ... dependiendo de los valores de a_1 y a_2 y el pasado estado del sistema antes de que la entrada fuera cero. Esto sugiere un algoritmo de adaptación basado sobre más de dos bits consecutivos, si el tamaño del escalón cae a su valor mínimo para condiciones de canal desocupado.

El diseño de sistemas ADM con predictores de orden alto no ha sido extensamente estudiado. El uso de cuantizadores multi-bit simplifica el diseño a expensas de la necesidad de usar sincronización entre el transmisor y el receptor.

3.4 PCM diferencial (DPCM)

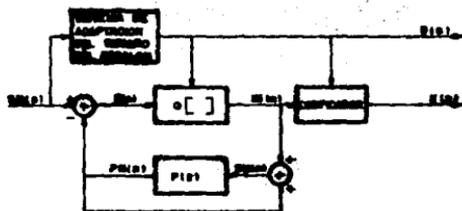
Cualquier sistema de la forma (3.6) puede ser llamado PCM diferencial. Los moduladores delta pueden ser llamados sistemas DPCM de un bit. Sin embargo, el término PCM diferencial es reservado para los sistemas con cuantización diferencial en los cuales el cuantizador tiene más de dos niveles.

De la figura (3.10), vemos que los sistemas DPCM con predictores fijos pueden mejorar el desempeño sobre la cuantización directa PCM sobre 4 a 11 dB. Una gran mejora ocurre yendo del sistema sin predicción a un sistema con predicción de primer orden, con pequeña ganancia adicional por incrementar el orden del predictor a 4 ó 5, después de lo cual la ganancia es despreciable.

Las variaciones de desempeño con el material de voz, junto con las variaciones en el nivel de la señal inherentes en el proceso de comunicación por medio de la voz, hacen que la cuantización adaptable sea necesaria para alcanzar el mejor desempeño sobre un amplio rango de situaciones con hablantes (personas con diferentes tonos de voz, energía, etc). Estos sistemas son llamados PCM diferencial adaptable (ADPCM).

3.5 PCM diferencial adaptable (ADPCM)

Lo visto anteriormente sobre cuantización adaptable es aplicable directamente a el caso del DPCM. La figura (3.11) muestra un cuantizador adaptable del tipo alimentación-directa (feed-forward) que es usado en un sistema ADPCM. Para esquemas de esta tipo, el tamaño del escalón cuantizado es proporcional a la varianza de la entrada a el cuantizador.



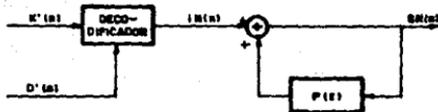
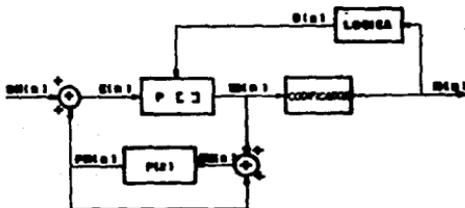


Fig 3.11 Sistema ADPCM con cuantización adaptable feed-forward
 a) codificador b) decodificador

Como la señal diferencia $E(n)$ es proporcional a la entrada, es razonable controlar el tamaño del escalón con $E(n)$ o, como en la figura (3.11) con la entrada $S(n)$. Como se vió anteriormente los procedimientos adaptables pueden dar una mejora de 5 dB en el SNR sobre un PCM ley- μ sin adaptación.

Esta mejora junto con los 6 dB que pueden ser obtenidos de la configuración diferencial con predictor fijo, significa que el ADPCM alimentación-directa (feed-forward) podría alcanzar un SNR de 10-11 dB más grande que el que puede ser obtenido con un cuantizador fijo del mismo número de niveles.

La figura (3.12) muestra como un cuantizador adaptable tipo "realimentado" (feedback) puede ser utilizado en un sistema ADPCM. Si por ejemplo la estrategia de adaptación de Jayant se usa, podemos esperar una mejora de 4-6 dB sobre el desempeño de un cuantizador PCM usando un cuantizador fijo que use la ley- μ y con el mismo número de bits.



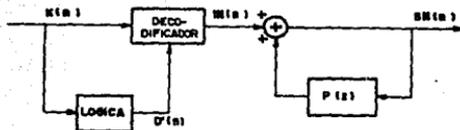


Fig 3.12 Sistema ADPCM con cuantización adaptable realimentada.

a) codificador b) decodificador

Además la adaptación del cuantizador da un rango dinámico mejorado y un SNR también mejorado. La principal ventaja del control realimentado (feedback) es que la información del tamaño del escalón es derivada de la secuencia de palabras de

código, así que no es necesario transmitir o guardar información adicional sobre el tamaño del escalón. Sin embargo, esto hace que la salida reconstruida sea más sensitiva a los errores en la transmisión. Con el control alimentación-directa (feed-forward), las palabras código y el tamaño del escalón sirven como representación de la señal original.

Existe además un sistema modulador ADPCM que es robusto, lo que significa que el sistema mantendrá su estabilidad a pesar de que se generen errores en el canal de transmisión.

Aunque esto incrementa la complejidad de la representación, hay posibilidad de transmitir el tamaño del escalón con protección contra errores, lo que significa una mejora de la calidad de la salida para transmisiones con una tasa de errores alta.

Analizemos como se realiza la protección contra errores producidos en el canal de transmisión, para esto definamos el factor de carga y como el cociente de la amplitud del escalón x y la desviación estandar α , que es el valor rms de la señal de entrada.

$$y = x/\alpha \quad (3.18)$$

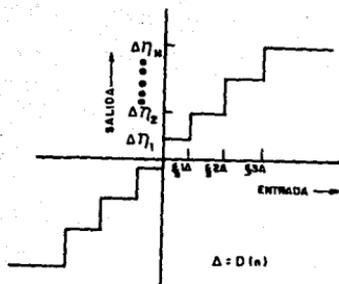


Fig 3.13 Cuantización adaptable.

Dada la forma del cuantizador y la información acerca del comportamiento de la entrada a un sistema ADPCM, el diseñador puede especificar un factor de carga, γ , que de una mezcla apropiada de ruido granular y ruido de sobrecarga. En telefonía las medidas rms de la señal de entrada fluctúan con el tiempo, la fluctuación depende sobre la persona que habla y su trayectoria de comunicación con el decodificador. Con un cuantizador fijo, x es constante y γ es inversamente proporcional a α que puede variar sobre los 40 dB. En un cuantizador adaptable, x varía con el tiempo de manera que reduce la dependencia de γ sobre α . En general para un

sistema ADPCM se cumple la siguiente relación

$$x(k+1) = M[IN(k)]x(k) \quad (3.19)$$

donde k implica el tiempo.

Aunque estos sistemas se consideran como sistemas de una palabra de memoria, y es importante apreciar que la memoria referida, es el número de registros requeridos para el funcionamiento del algoritmo. En el sentido de la teoría de la información, la memoria es infinita porque cada voltaje de sobrecarga, $x(k)$, depende sobre el pasado completo de la secuencia del código, dada por

$$x(k+1) = \prod_{m=0}^k M[IN(m)] x_0 \quad (3.20)$$

La ecuación (3.20) expone una desventaja del funcionamiento del algoritmo, su sensibilidad a los errores de la transmisión. Esto lo comprobamos si dejamos que x' e IN' sean las versiones decodificadas de x e IN , respectivamente y asumiendo que en el tiempo $\lambda < k$, $IN(\lambda) = i$ mientras que un error de transmisión causa que $IN'(\lambda) = j$. Si no existe más que este error, tenemos

$$x'(k+1) = \frac{M(j)}{M(i)} x(k+1) \quad (3.21)$$

cada error causa una compensación (offset) multiplicativa entre el transmisor y el receptor que en principio puede persistir indefinidamente, degradando la señal reconstruida.

Investiguemos ahora el procedimiento de adaptación para que el efecto de un error transmitido disminuya con el tiempo, para esto basémonos en la siguiente fórmula

$$x(k+1) = M[IN(k)] x^\beta(k) \quad (3.22)$$

Para disminuir los efectos de los errores de transmisión introducimos un nuevo factor exponencial β , y vemos como se comporta el codificador cuando existe un error generado en el canal de transmisión.

$$x(k+1) = \prod_{m=0}^k M[IN(m)] \beta^{(k-m)} x_0 \beta^{(k+1)} \quad (3.23)$$

Si en el tiempo λ , j es recibido en vez de i ; el sistema procederá de acuerdo con la siguiente fórmula para disminuir el offset generado por el error de transmisión:

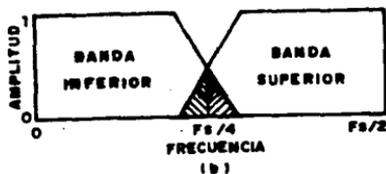
$$\frac{x'(k+1)}{x(k+1)} = \left[\frac{M(j)}{M(i)} \right] \beta^{(k-\lambda)} \quad (3.24)$$

Cuando $\beta < 1$, el offset debido a cada error decae exponencialmente con el tiempo, haciendo que el cuantizador sea más robusto en presencia de los errores de transmisión, que un cuantizador operando de acuerdo con (3.16), donde implícitamente $\beta = 1$.

3.6 Codificación por sub-bandas (SBC)

La codificación por sub-bandas representa el siguiente nivel en complejidad de los algoritmos de codificación de señales, este tiene su correspondiente ventaja en desempeño.

La figura (3.14) revisa la configuración básica para un diseño SBC de dos bandas. La señal de entrada $S(n)$ es asumida que esta en forma digital (PCM digital) y puede ser (opcionalmente) filtrada con un prefiltro pasa-banda, las razones serán discutidas más adelante. La señal de salida $x(n)$ se divide entonces en dos bandas igualmente espaciadas en frecuencia, por un filtro pasa-bajas $h_1(n)$ y un filtro pasa-altas $h_2(n)$. Cada sub-banda es entonces reducida en su tasa de muestreo por dos, esto es, si F_s es la tasa de muestreo de la señal de entrada, $F_s/2$ es la tasa de muestreo de cada una de las sub-bandas.



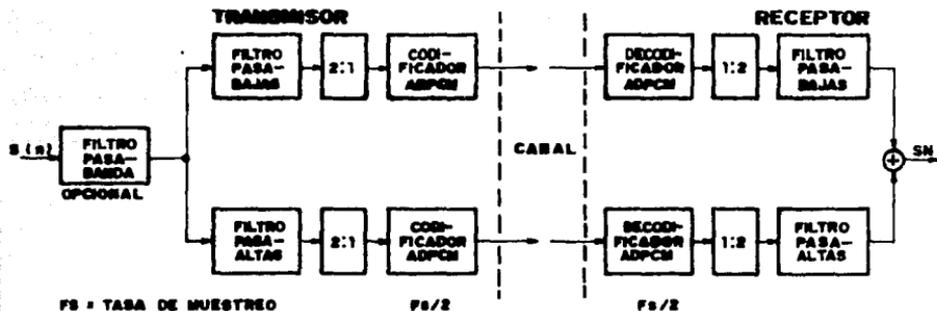


Fig 3.14 a) Diagrama a bloques de un codificador por sub-bandas

b) Descripción espectral de las sub-bandas

Las señales de cada una de las sub-bandas son después codificadas con un codificador tipo ADPCM y los bits de salida son multiplexados para almacenarse o transmitirse. En el receptor, las señales de las sub-bandas son decodificadas e

interpoladas a su tasa de muestreo original con la ayuda de filtros similares pasa-bajas y pasa-altas. La suma de las dos señales interpoladas de las sub-bandas, $SN(n)$, es la versión reconstruida de la señal de entrada $S(n)$ (ver figura 3.14).

Este proceso de dividir la señal en dos sub-bandas permite a cada banda a ser codificada con un diferente número de bits por muestra y con un tamaño del escalón adaptable independiente para cada sub-banda, con el fin de obtener una mejor calidad. Para una banda de voz telefónica (200 a 3200 Hz) muestreada a 8 Khz, la técnica de las dos bandas da alrededor de 3 ó 4 Kb/s de ventaja en la tasa de bits sobre el ADPCM en el rango de 24 KB/s.

La codificación en sub-bandas ofrece varias ventajas. El ruido de cuantización es constreñido en cada banda para prevenir enmascaramiento de un rango de frecuencia, por el ruido de cuantización en otro rango de frecuencia. Los cuantizadores adaptables son usados en cada banda, por esto las bandas con baja energía tendrán tamaños de escalones pequeños en sus cuantizadores y contribuirán con menos ruido de cuantización. Con una apropiada colocación de los bits en las diferentes bandas, la forma del ruido de cuantización puede ser controlado en frecuencia, lográndose una mejor calidad.

En las bandas de frecuencias bajas, donde la estructura del tono y los formantes debe ser conservada, se usa un gran número de bits por muestra, mientras que en las bandas superiores, donde ocurren sonidos fricativos en la voz, se necesitan pocos bits para cada muestra.

Una gran compresión en la tasa de bits (ó una alta calidad de ejecución para la misma tasa) sobre el esquema de dos bandas es posible con más de dos bandas. Diseños alternativos con espacios a la octava parte del original son posibles por sucesivamente subdividir las bandas internas en una "estructura de árbol". Estos diseños son una extensión lógica del diseño en dos bandas, y pueden tener una ventaja de ejecución de alrededor de 8 Kb/s sobre un ADPCM para voz (en el rango de 16 a 24 KB/s).

En el diseño de estos esquemas, un aspecto crítico e importante del SBC es el banco de filtros y su interacción con la reducción de la tasa de muestreo (decimación), y el subsecuente incremento de la tasa de muestreo (interpolación) de las señales en las sub-bandas, se debe de procurar que estos filtros sean bastante aproximados a los ideales para prevenir el fenómeno de traslape de colas de los filtros (aliasing), que reducen la calidad de la señal cuando se reconstruye.

La reducción de la tasa de muestreo es necesaria para mantener al mínimo la tasa total de bits de la señal codificada. Sin embargo esta reducción de la tasa de muestreo introduce términos del traslape de colas de los filtros, en cada una de las señales de las sub-bandas. Por ejemplo, en la banda baja la energía de la señal en el rango de frecuencia arriba de $F_s/4$ es doblada en el rango de 0 a $F_s/4$ y aparece como un traslape en esta señal, como se muestra en la región sombreada de la figura (3.14). Similarmente, para la banda superior cualquier energía de la señal en el rango abajo de $F_s/4$ es doblada arriba en la banda de $F_s/4$ a $F_s/2$. Este traslape mutuo de la energía de la señal entre las sub-bandas superior e inferior es llamada "fuga entre bandas" (leakage). La cantidad de fuga que ocurre entre sub-bandas es directamente dependiente del grado con que los filtros $h_l(n)$ y $h_u(n)$ se aproximan a los filtros ideales pasa-bajas y pasa-altas, respectivamente.

Para el proceso de reconstrucción, la tasa de muestreo de las sub-bandas es incrementada, por llenar con valores cero las muestras entre cada par de muestras de las sub-bandas. Esto introduce una repetición periódica del espectro de la señal en la sub-banda. Por ejemplo, en la banda inferior la energía de la señal de 0 a $F_s/4$ es doblada simétricamente alrededor de de la frecuencia $F_s/4$ en el rango de la banda superior.

Esta energía indeseada, referida como imagen es filtrada por el filtro pasa-bajas $h_1(n)$ en el receptor. La operación de filtrado interpola las muestras de valor cero que han sido insertadas entre las señales de la sub-banda, a valores que representan apropiadamente la forma de onda deseada. Similarmente, en la sub-banda superior una imagen es reflejada a la sub-banda inferior y filtrada por el filtro, $h_u(n)$.

El grado en que las imágenes son removidas por los filtros $h_1(n)$ y $h_u(n)$ es determinado por el grado en que éstos se aproximen a los filtros ideales. Este tipo de filtros se conocen como banco de filtros espejo en cuadratura (QMPB). De la relación de cuadratura de las señales de sub-banda en los QMPB, los componentes remanentes de las imágenes pueden exactamente ser cancelados por los términos de traslape introducidos. En la práctica esta cancelación se obtiene al nivel de ruido de cuantización de los codificadores.

Es deseable en la codificación SBC limitar en banda la señal de entrada antes de codificarla. Por ejemplo, en la voz hay una cantidad considerable de energía en el rango de frecuencia de 0 a 200 Hz. Esta energía contribuye a un incremento en el tamaño del escalón y por lo mismo a más ruido de cuantización en la sub-banda inferior.

Si la banda de voz telefónica (200 a 3200 Hz) es la de interés, entonces es deseable remover las frecuencias abajo de los 200 Hz y arriba de los 3200 Hz. Esto permite el uso de tamaños de escalón pequeños en la banda inferior, produciéndose menos ruido de cuantización.

3.7 VOCODERS

Los sistemas que usan las técnicas de la sección anterior se llaman codificadores de forma de onda, estos sistemas son distintos de los vocoders. Los codificadores de forma de onda usan algoritmos para producir a la salida del sistema una aproximación de la forma de onda de la señal de voz que entra al sistema.

En los vocoders se usa una descripción compacta de la entrada, digitalizando únicamente los parámetros de esta descripción compacta. La descripción se basa generalmente en una señal de excitación alimentando un filtro lineal, este es un modelo que intenta simular el proceso de excitación y modulación del mecanismo vocal humano. El resultado es una salida de sonido artificial, en la cual, las palabras son claramente comprensibles, pero la persona que habla no es siempre identificable.

La calidad del vocoder es tolerable en comunicaciones de seguridad militares que deben ser enviadas a muy bajas tasas de bits, 4 Kb/s o menos. Los algoritmos de codificación de los vocoders no son robustos, esto es, ellos pueden fallar cuando hay ruido de fondo o cuando varios usuarios usan el codificador simultáneamente.

Una aplicación prometedora del vocoder es en servicios de respuesta por voz, en la que el usuario codifica un mensaje para más tarde recuperarlo; recuperar el mensaje es meramente decodificar a una determinada calidad.

Algunos de los nuevos codificadores de onda a bajas tasas de bits, toman algunas de las tecnologías del vocoder. Estos codificadores híbridos o vocoders "blandos" son más robustos que los verdaderos vocoders; la meta a alcanzar es una calidad de comunicación que sea bastante buena para los servicios de rutina comerciales (transmisión y almacenado) a tasas de bits de 8 Kb/s y eventualmente 4 Kb/s. Claramente, reducir la tasa de bits por codificación requiere un gran escrutinio del proceso de la voz. Esto significa gran complejidad y gran retardo del procesamiento.

Veamos dos de los codificadores híbridos y el vocoder verdadero:

3.7.1 Codificación predictiva lineal multipulso (MLPC).

3.7.2 Codificación predictiva lineal exitada estocásticamente (SELPC).

3.7.3 Vocoder verdadero

Los codificadores híbridos pueden bajar más la tasa de datos, para esto alimentan al filtro lineal predictor (LPC) una señal de excitación optimizada, esta es una técnica que adopta mucha de la eficiencia del vocoder tradicional, y todavía es bastante flexible para seguir las sutiles propiedades de la forma de onda de la señal de voz. Esta técnica usa principios de codificación de forma de onda de alta calidad para optimizar la señal de excitación, en vez de usar la excitación rígida de dos estados del vocoder. Con las más avanzadas técnicas híbridas, solamente 1 a 0.5 bits por muestra son necesarios para codificar la señal de voz, resultando en tasas de bits de 8 a 4 Kb/s.

Optimizar la señal de excitación y los parámetros del filtro es un gran reto, ambos parámetros deben variar con el tiempo para alcanzar buena calidad de codificación y naturalidad, analicemos cada uno de ellos.

3.7.1 La codificación lineal multipulso (MLPC) es un buen candidato para la codificación a 8 Kb/s, en este código un número disponible de pulsos es alimentados como secuencia de excitación para un segmento de voz, podrían ser 10 pulsos para un segmento de 10 ms. La amplitud y localización de los pulsos está optimizado, pulso por pulso, en una malla de búsqueda cerrada. La tasa de bits reservada para la información de la excitación, es más de la mitad del total de la tasa bits de 8 Kb/s.

Esto no deja muchos bits para la información del filtro predictor lineal, pero con la sofisticación de la cuantización vectorial, la codificación de los parámetros de predicción pueden ser hechos bastante precisos.

Antes de continuar, veamos lo que es la cuantización vectorial, para entender mejor como funcionan estos esquemas.

En cuantización vectorial, el cuantizador ve en su tabla de memoria el dato que más cerradamente iguala una secuencia de muestras, por ejemplo, una secuencia de 40 muestras de largo. Este escoge la dirección del código guardado en memoria para el grupo de secuencias igualada. El cuantizador entonces transmite la dirección del código a el receptor, mejor que transmitir la secuencia de por ejemplo 40 muestras cuantizadas.

Si la memoria contiene 1024 secuencias para elegir de salida, solamente 10 bits son necesarios para transmitir la dirección del código para las 40 muestras ($2^{10} = 1024$), y la tasa de cuantización es por esto de solamente 0.25 bits por muestra, en la figura (3.15) se puede ver este sistema.

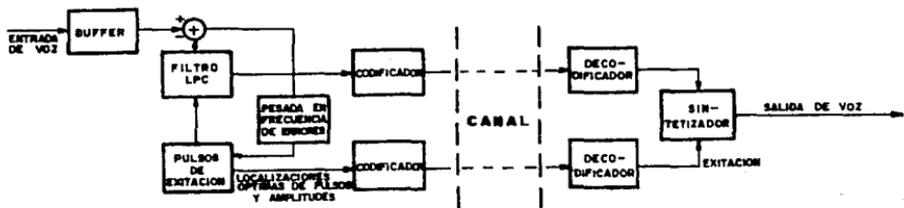


Fig 3.15 Codificación lineal predictiva multipulso.

3.7.2 La codificación predictiva lineal excitada estocásticamente (SELPC) es prometedora para la codificación a 4 Kb/s. En este caso el codificador guarda un repertorio de excitaciones para elegir, cada una es una aleatoria o estocástica secuencia de pulsos. La mejor secuencia es seleccionada por una malla de búsqueda cerrada. La cuantización vectorial en el filtro lineal predictor es casi una necesidad aquí para garantizar que bastantes bits quedan disponibles para los parámetros de excitación y predicción del

LPC. La cuantización vectorial asegura una buena calidad, por permitir bastantes opciones para la excitación y la tabla de códigos para el filtro.

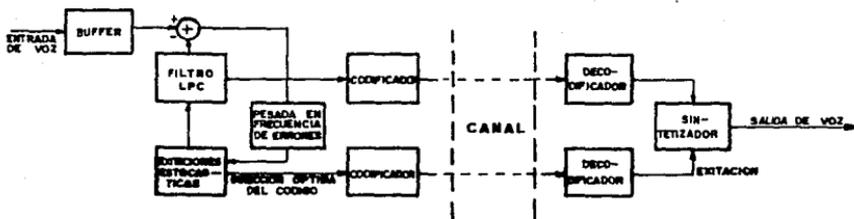


Fig 3.16 Codificación predictiva lineal
excitada estocásticamente.

3.7.3 Vocoder verdadero. Vocoder es un término derivado de las palabras VOICE (voz) y CODER (codificador). Básicamente los vocoders preservan el espectro de potencia del período mientras descartan la información de fase en la señal de voz. A diferencia de los moduladores delta, DPCM y ADPCM

estos no tratan de preservar la forma de la señal de voz. En vez de esto, la meta es preservar las propiedades significativas de la señal con la intención de sintetizar la señal en el receptor con un sonido muy parecido la señal original.

Analizando la señal de entrada, o un periodo corto del espectro los vocoders calculan los parámetros que describen un modelo simplificado del mecanismo de la producción de voz. Los modelos de los vocoders asumen que los sonidos de la voz caen en dos clases distintas: sonoras y sordas (voiced/unvoiced).

Los sonidos sonoros ocurren cuando las cuerdas vocales vibran y son caracterizados por el tono o tasa de vibración de las cuerdas vocales, así como por la estructura resonante del tracto vocal, formado por la garganta, la boca y las cavidades nasales.

En los sonidos de voz sorda, las cuerdas vocales no vibran. En vez de esto, resultan turbulencias de aire, al pasar el aire a través de una constricción estrecha formada por el órgano articulador, o también por el súbito escape de aire por los labios o lengua, creando un ruido acústico que excita el tracto vocal.

En la voz sonora, los articuladores crean condiciones de resonancia que concentran la energía en áreas particulares del espectro de potencia en frecuencia. Al contrario del caso sonoro, donde la energía ocurre en componentes discretas de frecuencia, durante la voz sorda la energía espectral es continua en frecuencia.

En general la sección de análisis del vocoder determina la estructura resonante del tracto vocal, estima el tono (pitch), y decide si el segmento de voz es sonoro o sordo. La sección sintetizadora usa estos rasgos de la voz para reconstruir de nuevo la forma de la señal, de manera que suene como la señal de entrada.

Los vocoders difieren en sus métodos para extraer los rasgos de la voz, así como también en sus métodos para reconstruir la voz usando estos rasgos. Aquí nos concretaremos a mencionar los tipos más importantes.

- i) Vocoder de canal
- ii) Vocoder cepstrum
- iii) Vocoder con formantes
- iv) Vocoder excitado sonoramente

i) Vocoder de canal. El analizador del vocoder de canal se muestra en la figura (3.17).

Una secuencia de filtros pasa banda es usado para dividir la señal de voz en canales de frecuencia. Los componentes de la señal en cada una de las frecuencias de los canales son primero rectificadas, usualmente por un circuito que use ley cuadrática. Después son integrados o filtrados con un filtro pasa-bajas, o ambos; esto es, para producir una estima continua de la amplitud del espectro de potencia de la voz en cada canal. Independientemente del análisis espectral, un extractor del tono determina el tono vocal y un detector de sonoridad determina si la señal de entrada es sonora o sorda.

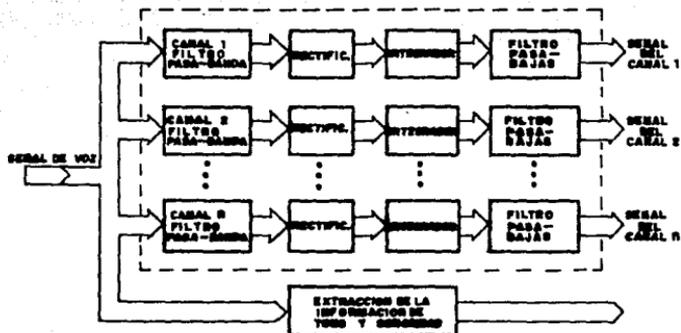


Fig 3.17 Analizador de un vocoder de canal.

Como se muestra en la figura (3.18), el sintetizador del vocoder de canal reconstruye la señal de voz usando una estima del espectro de potencia junto con el tono y la información de sonoridad.

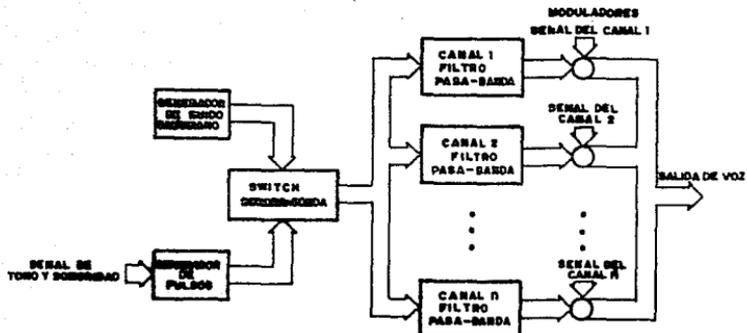


Fig 3.18 Receptor de un vocoder de canal.

Durante los segmentos sonoros, un generador de pulsos excita un banco de filtros similar al de la figura (3.18). Cada salida de los filtros es ajustada en un intento de hacer esta energía igual a la medida por el analizador del canal correspondiente. Para sonidos sordos, una fuente de ruido gaussiano excita el banco de filtros.

En los años 1950's el desarrollo de los vocoders fué enfocado principalmente a los aparatos de este tipo con 16 canales, se puso considerable énfasis para optimizar los filtros de los canales.

Sin embargo, los esfuerzos más recientes en el desarrollo de los vocoders están dirigidos a realizar una implementación completamente digital. Uno de los métodos es usar la transformada rápida de Fourier (FFT), y otro es con las técnicas usadas en los filtros digitales.

ii) Vocoder cepstrum. Una de las variaciones de la técnica con FFT es usar el concepto de filtrado homomórfico.

Para explicar como trabaja, consideremos un modelo que separa la información del tracto vocal de la excitación de tono en la voz sonora. Recordando que una señal de voz es el resultado de soplos de aire liberados periódicamente por las cuerdas vocales, que excitan la cavidad acústica formada por el tracto vocal. Por ésto, nuestro modelo de la señal de voz consiste de la convolución de tres componentes, que representan el tono, la forma del pulso glotal, y la respuesta impulso del tracto vocal. Este modelo puede ser representado por dibujar esquemáticamente el espectro de potencia de un segmento de voz sonoro, como se muestra en la figura (3.19).

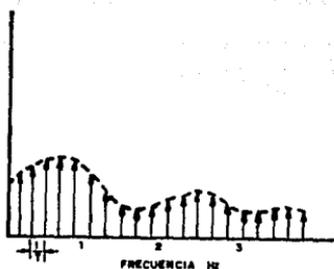


Fig 3.19 Espectro de potencia de la señal de voz.

Los impulsos representan el espectro de la excitación de voz espaciados a intervalos definidos por el período del tono.

En esta figura las flechas representan impulsos del espectro de potencia de la excitación. La energía es concentrada en la frecuencia de tono y sus armónicas. La envoltura, representada por la línea punteada, representa la función de magnitud que cambia relativamente lenta con la frecuencia.

El filtrado homomórfico implica tomar primero el logaritmo del espectro de potencia, para que la excitación y los efectos de la envoltura puedan ser más fácilmente separados. La transformada de Fourier del espectro de potencia logarítmico es calculada, produciendo como resultado

el cepstrum, como el que se muestra en la figura (3.20).

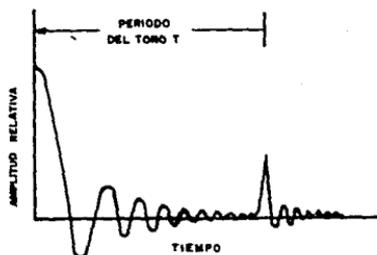


Fig 3.20 Cepstrum.

Esta señal se obtiene al tomar la transformada de Fourier del logaritmo del espectro de potencia de la señal de voz, se usa para separar la función de envoltura de la función de excitación.

El cepstrum se usa para separar la envoltura (configuración del tracto vocal) y las funciones de excitación (cuerdas vocales). La variación cercana al origen es debida casi completamente a los cambios lentos de la función de envoltura, mientras que la distancia entre el origen y la espiga predominante es el período de excitación.

Las muestras del cepstrum junto con la información del tono son transmitidas hacia el receptor. En el receptor las muestras son convertidas en una respuesta impulso, que es convolucionada con un tren de impulsos generado de la información del tono.

Los vocoders de canal y cepstrum operan entre 2400 y 4800 b/s, y ambos pueden producir voz de alta integridad, sin embargo, esta voz sincronizada se caracteriza por la degradación de la naturalidad y la reconocibilidad de la voz de la persona hablante.

iii) Vocoder con formantes. Este es similar al vocoder de canal. Teóricamente el vocoder con formantes puede operar a más baja tasa de bits que el vocoder de canal, porque éste usa pocas señales de control. En vez de enviar muestras de la envoltura del espectro de potencia, el vocoder con formantes transmite las posiciones de los picos de la envoltura del espectro, las posiciones de los picos son llamadas frecuencias de los formantes y corresponden a los polos en la respuesta de la cavidad vocal. Típicamente, un vocoder formante deberá identificar al menos tres formantes para la representación de los sonidos de voz, y deberá controlar las intensidades de los formantes. El control se implementa de varios modos dependiendo del método escogido para la reconstrucción del espectro de la voz.

Los vocoders con formantes pueden reproducir la voz a una tasa de datos de 1200 b/s, sin embargo éstos no fueron un éxito como los vocoders de canal, y permanecen más como una curiosidad de laboratorio que un aparato de comunicaciones práctico. Una de las principales razones de esta falla es que los vocoders con formantes tienen dificultad para calcular de forma precisa la localización de los formantes de la fuente voz humana. Sin embargo, los sintetizadores de los vocoders con formantes operados bajo el control de una computadora, han mostrado calidad vocal excelente, cuando una considerable cantidad de esfuerzo ha sido ejercida para obtener los trazos apropiados de los formantes. Por esto los sintetizadores de los vocoders con formantes se consideran para su uso en unidades de respuesta oral automáticas de los sistemas telefónicos.

iv) Vocoder exitado sonoramente. Se ha buscado evitar la de radación de la voz en los vocoders por eliminar la extracción del tono y las operaciones de detección de sonoridad. Uno de las más exitosas inovaciones es el vocoder exitado sonoramente (voiced-exited vocoders VEV), como el mostrado en la figura (3.21).

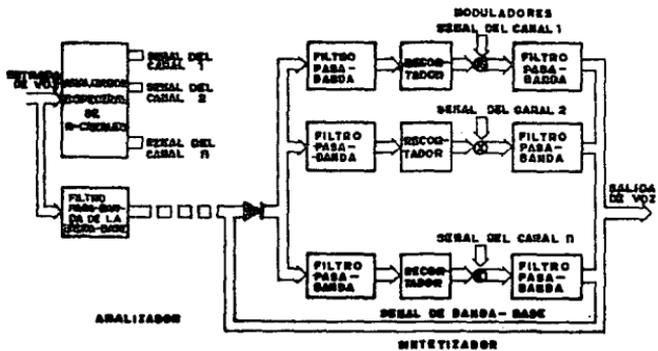


Fig 3.21 Vocoder exitado sonoramente.

El sistema no requiere separar la información de tono y no tiene que realizar decisiones sonora/sorda.

Este sistema usa una combinación híbrida de transmisión para el PCM en la banda de frecuencias bajas de la señal de voz, combinandolo con codificación de canal para la banda de altas frecuencias. El analizador no tiene que extraer el tono o hacer desiciones de sonoridad. Una señal de tono es generada en el sintetizador al rectificar, filtrar en banda baja, y recortar la señal de banda-base, creando una señal espectralmente plana con energía en los armónicos del tono. Estos vocoders se han diseñado para operar a 7200-9600 b/s, y su calidad es superior a la obtenida por el tradicional vocoder exitado por tono.

3.8 Codificación por predicción lineal (LPC)

Esta predicción intenta extraer los rasgos significativos de la voz desde la forma de onda en el tiempo, mejor que desde el espectro de frecuencia con que los vocoders de canal, con formantes, cepstrum, operan.

El sistema de predicción lineal usa una suma ponderada de n muestras pasadas para estimar la muestra presente. La ponderación es calculada para minimizar la energía promedio de la señal de error, que representa la diferencia entre las amplitudes de la señal de voz presente y la señal de voz predecida. A diferencia de los sistemas DPCM donde la ponderación es calculada una vez para toda la señal de voz, permaneciendo de ahí en adelante fija, las ponderaciones en el código de predicción son calculadas para segmentos cortos de señales de voz de 10 a 30 ms, cambiando como la estadística de la señal de voz varía.

Los sistemas de predicción lineal difieren también de los esquemas ADPCM. En los esquemas adaptables se transmite la señal de error, que representa la diferencia entre las muestras de señal presentes y predecidas. Los sistemas de predicción lineal transmiten únicamente características seleccionadas de la señal de error. Estos parámetros incluyen un factor de ganancia, información del tono, y la desición sobre información sonora-/sorda, que permite hacer una

aproximación de la señal de error correcta usada para evitar el filtro formado por los coeficientes del predictor y así generar una señal de voz sintética.

Las simulaciones en computadora de los codificadores de predicción lineal indican que estos aparatos pueden producir voz de mayor naturalidad que los vocoders tradicionales. Aunque el análisis es ejecutado en el dominio del tiempo, minimizando la señal de error, esto es equivalente a estimar el espectro de la señal de voz más exactamente alrededor de sus picos (polos) que de sus ceros.

Con esto se consigue eliminar una de las mayores deficiencias de los vocoders de canal y cepstrum. Sin embargo, la extracción del tono y una desición pobre de sonoridad, aún causan distorsiones perceptibles.

La importancia de este método esta en su habilidad para estimar en forma precisa los parámetros de la voz, y en su relativa velocidad de computación. Existen diferentes métodos para calcular los coeficientes del LPC:

- 1.- covarianza
- 2.- autocorrelación
- 3.- de rejilla
- 4.- filtro inverso
- 5.- estimación espectral
- 6.- máxima probabilidad
- 7.- producto interno

Pero todos se basan en un sistema de producción de voz, en la figura (3.22) vemos un sistema simplificado de producción de voz.

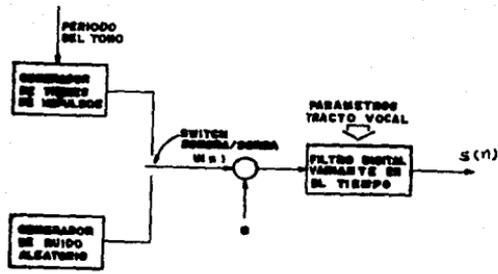


Fig 3.22 Diagrama a bloques de un modelo simplificado de producción de voz.

Este modelo es apropiado para el LPC. En este caso, los efectos del espectro de radiación compuesto, tracto vocal, y la excitación glotal son representados por un filtro digital variante en el tiempo cuya función de estado estable es de la forma

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{1 - \sum_{k=1}^p a_k z^{-k}} \quad (3.25)$$

El sistema es excitado por un tren de impulsos para voz sonora y por ruido aleatorio para voz sorda. Los parámetros de éste modelo son: clasificación sonora/sorda, período del tono para la voz sonora, parámetro de ganancia G , y los coeficientes $\{a_k\}$ del filtro digital, estos parámetros varían lentamente con el tiempo. Este modelo simplificado de unicamente polos es una representación natural de sonidos sonoros no-nasales, pero para sonidos nasales y sonidos fricativos la teoría acústica indica que una representación de la función de transferencia del tracto vocal con polos y ceros es lo más correcto. Sin embargo, como se demostrara, si el orden de p es bastante alto, el modelo con unicamente polos da

una buena representación para casi todos los sonidos de la voz. La mayor ventaja de este modelo es que el parámetro de ganancia G , y los filtros del coeficiente $\{a_k\}$ pueden ser determinados de una manera directa y computacionalmente eficiente por el método de análisis predictivo lineal.

Para el sistema de la figura (3.22), las muestras de voz $s(n)$ se relacionan con la excitación $u(n)$ por la ecuación en diferencias

$$s(n) = \sum_{k=1}^p a_k s(n-k) + Gu(n) \quad (3.26)$$

Un predictor lineal con coeficientes de predicción, α_k es definido como un sistema cuya salida es:

$$\bar{s}(n) = \sum_{k=1}^p \alpha_k s(n-k) \quad (3.27)$$

La función del sistema con predictor lineal de p^{th} orden es el polinomio

$$\varphi(z) = \sum_{k=1}^P a_k z^{-k} \quad (3.28)$$

El error de predicción, es definido como:

$$e(n) = s(n) - \bar{s}(n) = s(n) - \sum_{k=1}^P a_k s(n-k) \quad (3.29)$$

De la ecuación anterior se ve que la secuencia del error de predicción es la salida de un sistema cuya función de transferencia es:

$$\Lambda(z) = 1 - \sum_{k=1}^P a_k z^{-k} \quad (3.30)$$

Como se puede observar al comparar las ecuaciones (3.26) y (3.29) que si la señal de voz obedece el modelo de la ecuación (3.26) de manera exacta, y si $a_k = a_k$, entonces

$e(n)=Gu(n)$. Así, el filtro de predicción de error, $\Lambda(z)$, será el filtro inverso del sistema, $H(z)$ de la ecuación (3.25), esto es

$$H(z) = \frac{G}{\Lambda(z)} \quad (3.31)$$

Para el análisis de la predicción lineal, el problema básico es determinar los coeficientes del predictor $\{\alpha_k\}$ directamente de la señal de voz de manera que se obtenga una buena estima de las propiedades espectrales de la señal de voz usando la ecuación (3.25). Dado que por naturaleza la señal de voz es variante en el tiempo los coeficientes del predictor deberán ser estimados de segmentos cortos de la señal de voz. La idea básica es encontrar un juego de coeficientes del predictor que minimizen el error medio cuadrático de predicción sobre un segmento corto de señal de voz. Los parámetros resultantes son asumidos que son los parámetros del sistema $H(z)$, en el modelo de producción de voz.

Aunque existen muchos más algoritmos para el procesamiento digital de señales de voz, en general cada uno de ellos es una versión optimizada de los algoritmos que se han visto en este capítulo o también puede ser una combinación de dos o más algoritmos, con el fin de sumar sus ventajas para realizar la compresión del ancho de banda más grande posible y

también obtener una buena reconstrucción de la señal de voz en el receptor.

CAPITULO IV

DISEÑO

Diseño de un sistema de codificación

ADPCM

4.1 Introducción

El diseño de un sistema de codificación, comienza por establecer que aplicación tendrá el sistema, para ver cuales son los métodos de codificación que mejor trabajan para esta aplicación en particular. En nuestro caso la aplicación de la codificación será en las comunicaciones telefónicas.

Para este tipo de aplicación la configuración básica de cualquier codificación debe considerar las siguientes recomendaciones:

- 1.- De preferencia debe ser capaz de operar en tiempo real.
- 2.- Debe de tomar en cuenta la presencia de errores en el canal de transmisión, teniendo protección contra errores en el envío de información al receptor. Un error en el canal puede causar una compensación (offset) multiplicativa entre el nivel de la señal en el receptor y el nivel de la señal en el transmisor. Esta compensación puede persistir de manera indefinida si no se procura que el sistema tenga un mecanismo disipador de errores. Estos sistemas son llamados robustos

por su inmunidad a los errores en el canal de transmisión.

3.- Esencialmente debe ser de "calidad transparente", esto implica que la calidad de la señal reconstruida debe de ser indistinguible de la señal fuente sin codificar.

4.- Y para nuestro caso el sistema debe de implementarse en un microprocesador TMS32010.

Para tratar de cumplir con estas recomendaciones se eligió implementar la codificación en modulación por pulsos codificados diferencial adaptable (ADPCM), por ser una codificación que cumple con todas las recomendaciones vistas anteriormente. Existen otros métodos para codificar más complejos pero en general fallan en alguna(s) de las recomendaciones anteriores, por ejemplo los codificadores LPC aunque comprimen a una tasa muy baja (alrededor de 9.6 Kb/s), tienen el inconveniente que la señal de voz reconstruida, aunque es entendible no es una voz natural y en ocasiones no es posible reconocer la persona que habla.

Este problema lo presentan la mayoría de los algoritmos que se utilizan para comprimir la señal de voz a tasas muy bajas, y además también presentan el problema de que a menor tasa de bits el algoritmo es más complejo consumiendo relativamente mucho tiempo en realizar la reducción en la tasa de bits empleados.

Ahora existen muchos nuevos algoritmos que soslayan muchos de los inconvenientes de los antiguos algoritmos de codificación, pero su tecnología es muy avanzada y tendrían que realizarse estudios previos sobre el comportamiento de la señal de voz del idioma Español, para poder aplicarse en nuestro país.

Los artículos que se publican en las revistas especializadas sobre procesamiento digital de señales de voz solo dan la pauta a seguir sin profundizar demasiado en como se implementan éstos algoritmos en forma práctica. Por éstas razones se eligió el algoritmo de codificación ADPCM para comenzar el desarrollo del procesamiento digital de señales de voz en el Instituto de Investigaciones Eléctricas.

ADPCM es un método para codificar digitalmente con una tasa de bits en el rango de 24 a 48 Kb/s. En el rango de 24 Kb/s el ADPCM nos da una buena calidad de reproducción de la señal de voz, que es aceptable en aplicaciones como los sistemas de respuesta por voz.

Con una tasa de 32 Kb/s el sistema nos da una calidad telefónica con una reducción por un factor de dos en la tasa de bits sobre el PCM convencional con una tasa de 64Kb/s.

4.2 Estructura y descripción general

Para definir la estructura del cuantizador ADPCM fué necesario trabajar primero con cuantizadores de modulación delta y cuantizadores DPCM de 4 bits, esto se hizo con el fin de familiarizarse con la teoría y la implementación práctica de un sistema compresor de voz.

Los codificadores fueron simulados en lenguaje Basic sobre una computadora personal Televideo (compatible con las computadoras de la IBM). Como señal de entrada se usaron 256 muestras de un segmento de voz digitalizada real. En cada una de las simulaciones al final se graficaron las señales de entrada originales contra las señales reconstruidas. Algunas de las gráficas resultantes se muestran en en las figuras 4.1 a 4.3. En éstas figuras se puede observar como a mayor número de bits empleados, la señal reconstruida sigue más fielmente a la señal originalmente entrada.

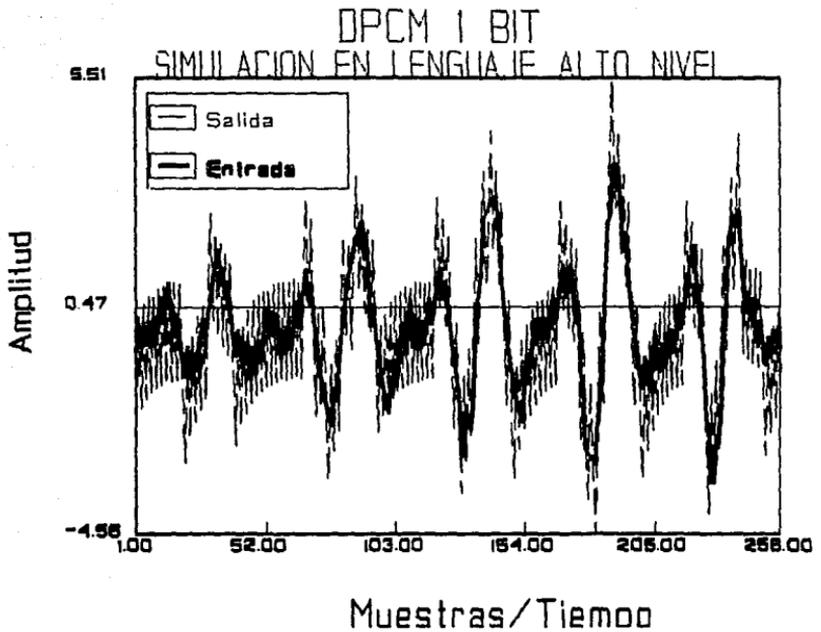


Fig 4.1 Simulación DPCM 1 bit

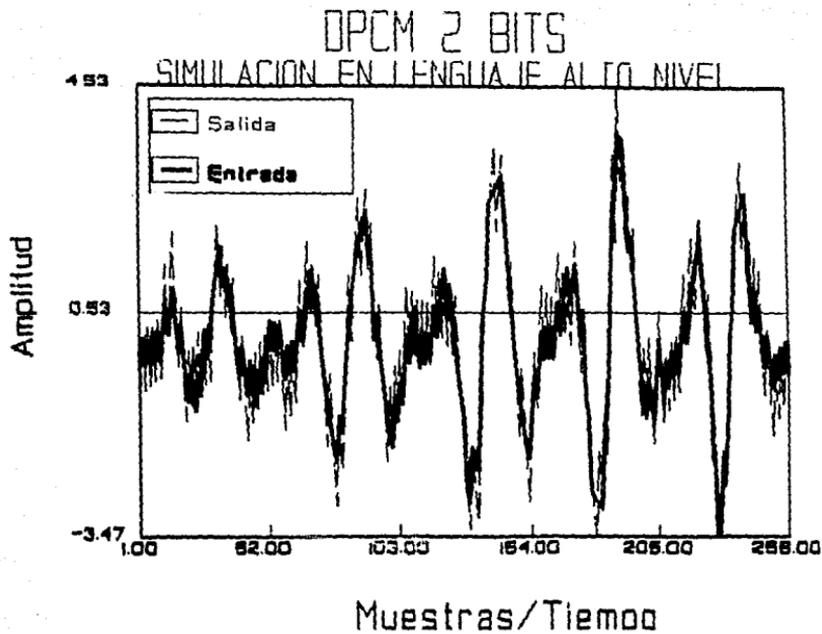


Fig 4.2 Simulación DPCM 2 bits

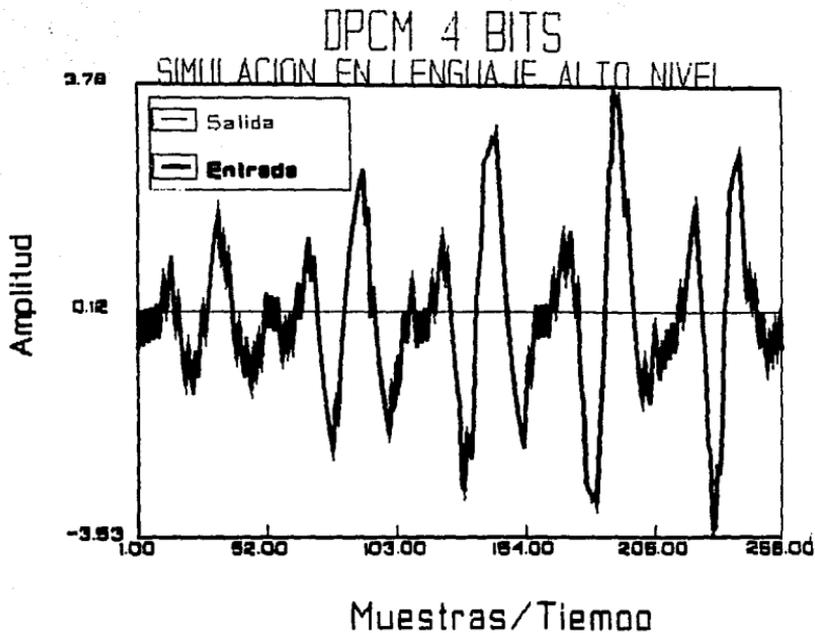


Fig 4.3 Simulación DPCM 4 bits

Una vez familiarizado con los cuantizadores delta y DPCM se comenzó a trabajar en el desarrollo del cuantizador ADPCM de la tesis. Es necesario recordar del capítulo 3 que existen dos tipos de cuantizadores ADPCM, uno de ellos es el cuantizador ADPCM alimentación-directa (feed-forward) y el otro es el ADPCM realimentado (feedback), éstos dos tipos de codificadores se muestran en las figuras 3.11 y 3.12 del capítulo 3.

Como ya se vió en el capítulo 3 en el cuantizador ADPCM alimentación-directa el tamaño del escalón es proporcional a la amplitud de la señal de entrada al cuantizador, y es necesario transmitir al decodificador la información sobre el control del tamaño del escalón, incrementandose la complejidad del algoritmo del cuantizador, sin embargo en este tipo de cuantizador ADPCM existe la posibilidad de transmitir la información sobre el tamaño del escalón con protección contra errores generados en el canal de transmisión, mejorándose con esto la calidad de la señal reconstruida cuando existe una tasa alta de errores en la transmisión.

El ADPCM realimentado tiene la ventaja de que la información sobre el tamaño del escalón se toma de la secuencia de palabras transmitidas, por lo que no es necesario transmitir información adicional sobre el tamaño del escalón hacia el decodificador. Sin embargo esta característica hace que la calidad de la señal reconstruida a la salida del

decodificador sea más sensitiva a los errores generados en el canal de transmisión.

4.3 Diseño

Para diseñar el cuantizador ADPCM fué necesario comprender primero como trabajan los sistemas de modulación delta con pendiente de variación continua (CVSD) fabricados en un integrado por la Motorola, el MC3417 y el MC3418. Este fué el primer paso para comprender como funciona un sistema de compresión de voz en forma práctica.

Aunque el diseño de este cuantizador es primordialmente analógico, y estos circuitos en aplicaciones reales tienden a presentar problemas de estabilidad y repetibilidad. En el laboratorio de comunicaciones del Instituto de Investigaciones Eléctricas se armó este sistema con su codificador y decodificador trabajando juntos. Se encontro que en general el funcionamiento de este sistema es bueno y no presento ningún problema de estabilidad.

En la figura 4.4 se muestra el diagrama básico de este sistema.

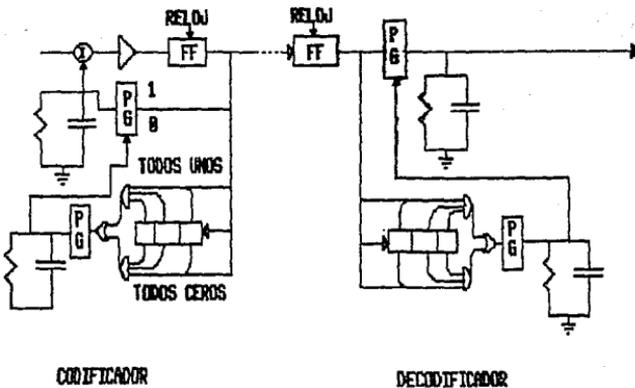


Fig 4.4 CVSD

El algoritmo de codificación básico CVSD fué descrito primero por Greefkes y Jager en 1968, después en los diseños subsiguientes se le agrego el control del tamaño del escalón. El CVSD toma la información sobre el tamaño del escalón de la secuencia de bits transmitidos. Como se puede observar en la figura 4.4 la lógica de adaptación monitorea los datos

transmitidos para detectar la ocurrencia de cuatro 1's sucesivos o cuatro 0's sucesivos.

Una secuencia de 1's indica que la señal de realimentación no esta incrementandose tan rápido como lo hace la señal de entrada, mientras que una secuencia de 0's indica que la señal de realimentación probablemente no esta decayendo tan rápido como la señal de entrada.

En todo caso, si hay secuencias de 1's ó 0's implicará que hay sobrecarga y el tamaño del escalón debe de ser incrementado, esto se hace capacitando el generador de pulsos₁. Generandose un pulso que incrementa el voltage del control del tamaño del escalón guardado en el capacitor C₁.

En el CVSD de la figura 4.4 se encuentra conectada en paralelo con el capacitor una resistencia, su función de esta es permitir que el voltage del control del tamaño del escalón decaiga con el tiempo. Esta combinación es conocida como "integración con fuga", implicando que el capacitor integra los pulsos de entrada, pero el resistor permite que el voltage de el tamaño del escalón acumulado se fugue (decaiga).

Una compensación (offset) a la salida de la terminal de recepción, causada por un error en el canal de transmisión, decaerá en el tiempo y será eliminada dentro de un período de pocas constantes de tiempo del segundo integrador.

La eliminación de esta compensación (offset) en la señal de salida no es tan importante para la señal de voz reconstruida que la persona oye a la salida del decodificador, pero si es necesaria para prevenir la saturación del decodificador porque distorcionará la forma de onda de la señal de voz reconstruida, estos integradores con fuga dan compensación a los pequeños desbalances en el tamaño del escalón que de otra manera conducirían a la saturación del decodificador. Si se utilizara una integración perfecta (sin fuga), un exeso de errores positivos o negativos podrían causar que la electrónica en el decodificador se sature y distorciona la forma de onda de la señal de voz.

Una vez comprendido el funcionamiento del CVSD, se eligió hacer un sistema ADPCM realimentado (feedback), porque aunque este sistema es más sensible a los errores generados en el canal de transmisión, se utilizaron las técnicas descritas en el capítulo 3 para hacerlo robusto (ecuación 3.22), esto significa que el sistema posee la capacidad de recobrase de los pasados errores generados en el canal de transmisión, durante los periodos libres de errores.

Se vió también la necesidad de optimizar el uso de las localidades de memoria. El sistema para control del tamaño del escalón del CVSD tiene que utilizar cuatro localidades de memoria para guardar los tres últimos bits de la secuencia transmitida y el bit presente, con el fin de hacer la

descisión de incrementar el tamaño del escalón.

En el sistema ADPCM sería necesario guardar las tres últimas palabras y la palabra presente. Por lo que para reducir las localidades de memoria utilizadas en la lógica del tamaño del escalón, se optó por utilizar la técnica creada por N.S. Jayant en su artículo "Adaptive Quantization With One-Word Memory", donde sólo se utiliza la palabra de memoria pasada para hacer la descisión de incrementar o decrementar el tamaño del escalón (esta técnica se describió en el capítulo 3).

Una vez definidos los parámetros básicos que contendrá el cuantizador, se utilizó la metodología para el diseño de aplicaciones propuesto en la Guía de Usuarios del TMS32010, el diagrama de flujo de esta metodología se muestra en la figura 4.5

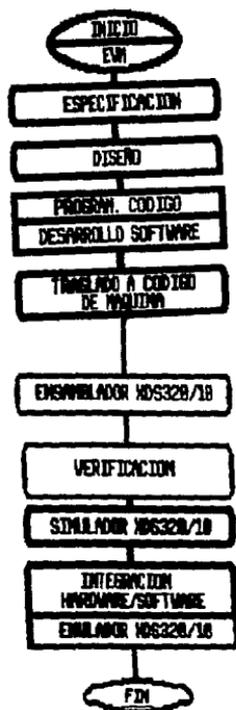


Fig 4.5 Diagrama de flujo para el diseño de una aplicación

En el diagrama se muestran las herramientas necesarias para un sistema basado en el microprocesador TMS32010. Sin embargo el Instituto de Investigaciones Eléctricas en un principio carecía de estas herramientas. Fué necesario conseguir los programas ensamblador XDS/320 y simulador XDS/320, éstos programas trabajan en las computadoras compatibles con las computadoras personales de la IBM. En el Instituto se cuenta con computadoras personales Televideo y Olivetti , las dos compatibles con las computadoras IBM.

El primer paso en el diseño del cuantizador fue crear un programa en lenguaje de alto nivel (Basic) que simulara el comportamiento del algoritmo ADPCM realimentado (feedback). Una vez creado ,se tuvo bastante cuidado en que los rediseños subsecuentes del cuantizador se acercaran a la implementación real del programa en el microprocesador. El programa completo se muestra en la figura 4.6.

```

10 REM *****
20 REM *      PROGRAMA PARA PROCESAR VOZ POR MEDIO DEL ALGORITMO ADPCM      *
30 REM *****
40 REM
50 REM INICIALIZACION DE LAS VARIABLES
60 DIM B(100)
70 GOSUB 1000
80 GOSUB 3000
90 MINP=8
100 MIN=8
110 DN1=0
120 DN1P=0
130 PN=0
140 PNP=0
150 AL=.85
160 GA=.98
170 SN=0
180 SP=0
185 WP=0
190 W=0
200 D=1
210 DV=1
215 DP=1
220 REM *****
230 REM *      SE INICIA LA CODIFICACION DE LA SENAL DE VOZ      *
240 REM *****
250 OPEN "DFCM.002" FOR INPUT AS #1
260 OPEN "ADPCM.003" FOR OUTPUT AS #2
270 FOR N=1 TO 256
280 INPUT #1,S
290 PN=SN*AL
300 E=S-PN
310 ES=E*DV
320 GOSUB 2000
330 ESN=IN+.5
340 EN=ESN*D
350 SN=EN+PN
360 GOSUB 2500
370 REM *****
380 REM *      SECCION DE DECODIFICACION DE LA SENAL DE VOZ      *
390 REM *****
400 GOSUB 3200
410 PRINT #2,SP
420 PRINT S,SP,N
430 NEXT N
440 REM *****
450 REM *      SECCION PARA GRAFICACION DE LA SENAL ORIGINAL DE VOZ      *
460 REM *      Y UN INSTANTE DESPUES LA SENAL SIMULADA CON ADPCM      *
470 REM *****
480 CLOSE #1
490 CLOSE #2
500 GOSUB 3500
510 END
1000 REM *****
1010 REM *      SUBROUTINA PARA GENERAR LOS ESPACIOS DE CUANTIZACION      *
1020 REM *****
1030 FOR I=1 TO 8
1040 B(I)=I
1050 NEXT I
1060 RETURN

```

```

2000 REM *****
2010 REM *           SUBROUTINA DE CUANTIZACION DE CUATRO BITS           *
2020 REM *****
2025 B(0)=0
2030 IF ES >=0 THEN GOTO 2150
2040 IF (ES<-8) THEN ES=-8
2050 H=ABS(ES)
2060 ES =H
2070 R=-1
2080 FOR I=1 TO 8
2090 IF ES <=B(I) THEN GOTO 2110
2100 GOTO 2140
2110 IN=B(I)*R
2120 ES=ES*R
2130 GOTO 2220
2140 NEXT I
2150 R=1
2155 IF (ES>8) THEN ES=8
2160 FOR I=1 TO 8
2170 IF ES <=B(I) THEN GOTO 2190
2180 GOTO 2210
2190 IN=B(I-1)
2200 GOTO 2220
2210 NEXT I
2220 RETURN
2500 REM *****
2502 REM *           SUBROUTINA DE CONTROL DEL TAMANO DEL ESCALON           *
2505 REM *****
2510 AB=ABS(IN+.5)
2520 IF (AB<4.5) THEN GOTO 2525
2525 MIN=-3
2530 GOTO 2560
2540 IF (AB>=4.5) THEN GOTO 2550
2550 MIN=8
2560 DN=DN1
2570 W=GA*DN
2580 DN1=MIN+W
2590 GOSUB 2800
2600 Q=FIX(DN)
2610 OFFSET=Q+32
2650 D=DELTA(OFFSET)
2660 DV=IDELTA(OFFSET)
2710 RETURN
2800 REM *****
2802 REM *           SUBROUTINA PARA DELTA MAXIMA Y DELTA MINIMA           *
2803 REM *****
2810 IF(DN1<=-31) THEN GOTO 2840
2820 IF(DN1>32) THEN GOTO 2860
2830 GOTO 2870
2840 DN1=-31
2850 GOTO 2870
2860 DN1=32
2870 RETURN
3000 REM *****
3010 REM * SUBROUTINA QUE GENERA LA TABLA DE ADECUACION DEL TAMANO DEL ESCALON *
3020 REM *****
3030 REM
3040 DIM DELTA(64)
3050 DIM IDELTA(64)
3060 N=10^.0375

```

```

3070 FOR I=1 TO 64
3080 X=I-1
3090 DELTA(I)=.1*N^X
3100 IDELTA(I)=1/DELTA(I)
3130 NEXT I
3140 RETURN
3200 REM *****
3210 REM *          SUBROUTINA DE DECODIFICACION DE LA SENAL          *
3220 REM *****
3230 REM
3240 REM          LLAMA LA SUBROUTINA DE DESICION DEL MULTIPLICADOR DEL
3250 REM          TAMANO DEL ESCALON
3260 ESNP=IN+.5
3280 ENP=ESNP*DP
3290 PNP=AL*SP
3300 SP=ENP+PNP
3310 GOSUB 4000
3320 RETURN
3500 REM *****
3510 REM *          SUBROUTINA PARA GRAFICAR LA SENAL ORIGINAL DE VOZ Y UN INSTANTE *
3520 REM *          DE TIEMPO DESPUES GRAFICAR LA SENAL SIMULADA CON ADPCM *
3530 REM *****
3540 INPUT "QUIERES QUE SE GRAFIQUE LA SENAL S/N";A$
3550 IF((A$="N")OR(A$="n")) THEN RETURN
3560 IF((A$="S")OR(A$="s")) THEN GOTO 3570
3570 CLS
3580 SCREEN 2
3590 LINE(20,0)-(532,160),,B
3600 FOR I=20 TO 160 STEP 25
3610 LINE(20,I)-(25,I)
3620 NEXT I
3630 PSET(20,80)
3640 OPEN "DPCM.002" FOR INPUT AS #1
3650 FOR I=1 TO 256
3660 INPUT #1,DATO
3670 DATOX=-DATO*20
3680 LINE-(2*I+20,DATOX+80)
3690 NEXT I
3700 N=0
3710 FOR I=1 TO 1000
3720 N=N+1
3730 NEXT I
3740 REM *****
3750 REM *          SE INICIA LA GRAFICACION DE LA SIMULACION CON ADPCM          *
3760 REM *****
3780 SCREEN 2
3790 LINE(20,0)-(532,160),,B
3800 FOR I=20 TO 160 STEP 25
3810 LINE (20,I)-(25,I)
3820 NEXT I
3830 PSET(20,80)
3840 OPEN "ADPCM.003" FOR INPUT AS #2
3850 FOR I=1 TO 256
3860 INPUT #2,DATO
3870 DATOX=-DATO*20
3880 LINE-(2*I+20,DATOX+80)
3890 NEXT I
3900 RETURN
4000 REM *****
4010 REM *          SUBROUTINA DE CONTROL DEL TAMANO DEL ESCALON PARA EL DECODIFICADOR *

```

```

4020 REM *****
4030 ABC=ABS(IN+.5)
4040 IF (ABC<4.5) THEN GOTO 4050
4050 MINP=-3
4060 GOTO 4090
4070 IF (ABC>=4.5) THEN GOTO 4080
4080 MINP=8
4090 DNP=DN1P
4100 WP=GA*DNP
4110 DN1P=MINP+WP
4120 GOSUB 4220
4130 QP=FIX(DNP)
4140 OFFSETP=QP+32
4180 DP=DELTA(OFFSETP)
4210 RETURN
4220 REM *****
4230 REM *      SUBROUTINA DEL DECODIFICADOR DE DELTA MAXIMA Y DELTA MINIMA      *
4240 REM *****
4250 IF (DN1P<=-31) THEN GOTO 4280
4260 IF (DN1P>32) THEN GOTO 4300
4270 GOTO 4310
4280 DN1P=-31
4290 GOTO 4310
4300 DN1P=32
4310 RETURN

```

Fig 4.6 Programa para simular un sistema
ADPCM realimentado

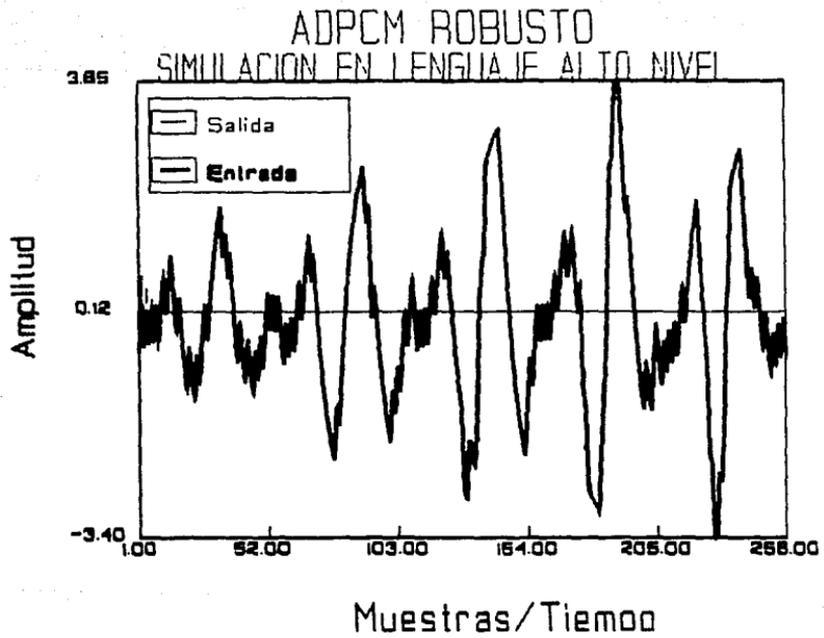
Los datos procesados por el programa fueron los mismos que se utilizaron en el DPCM de 4 bits. Los datos procesados se gráficaron en la figura 4.7.a, donde se compara la señal de entrada original contra la señal reconstruida.

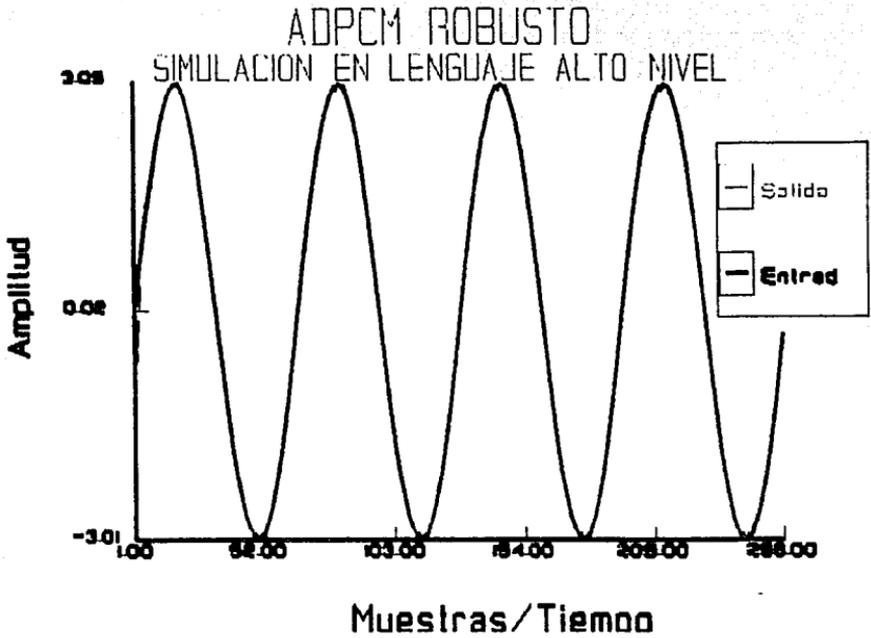
También se hicieron pruebas al comportamiento del programa simulador del ADPCM realimentado alimentándolo con una señal senoidal y una señal cuadrada.

Es interesante notar que si se observa la señal cuadrada reconstruida, se ve claramente el ruido de cuantización y el ruido de sobrecarga que se mencionó en el capítulo 3, figura 4.7.c.

En las gráficas se puede observar que el comportamiento del cuantizador ADPCM realimentado es bastante bueno. La señal reconstruida sigue fielmente la señal de entrada, con sólo algunas pequeñas diferencias debidas al ruido de cuantización y el ruido de sobrecarga.

1





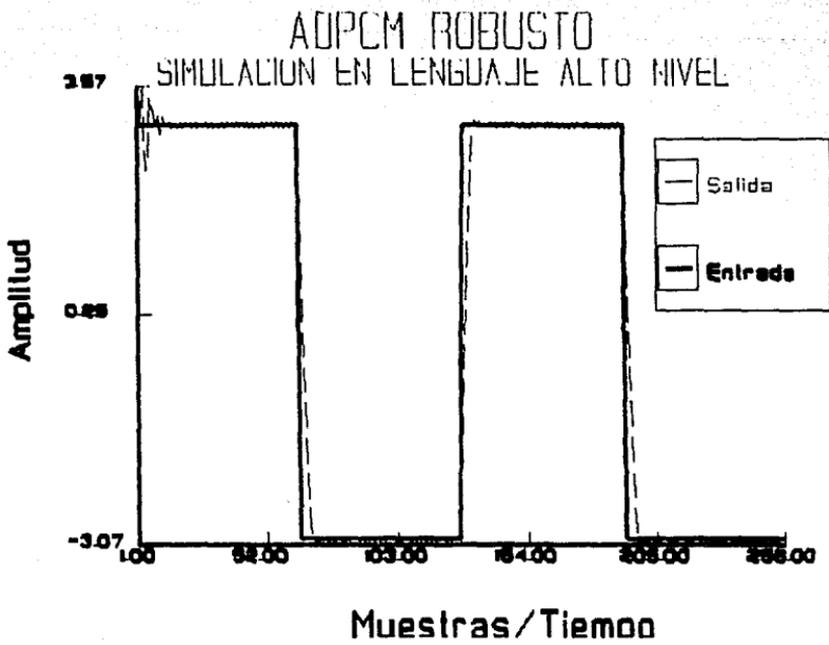


Fig 4.7 Señales de prueba al simulador del sistema

ADPCM realimentado

- a) señal de voz b) señal senoidal c) señal cuadrada

La figura 4.8 muestra el diagrama a bloques del último diseño del algoritmo ADPCM realimentado que se simuló en computadora. Este algoritmo se puede dividir en tres partes principales: el cuantizador adaptable ADPCM, el circuito predictor diferencial y el control de ganancia adaptable (tamaño del escalón) para el cuantizador. Analicemos el funcionamiento de cada una de estas partes:

4.3.1 Cuantizador adaptable

En esta sección del algoritmo, el valor predicho $PN(n)$ se sustrae de la señal de entrada $S(n)$

$$E(n) = S(n) - PN(n) \quad (4.1)$$

El valor $PN(n)$ se obtuvo previamente del predictor y es calculado a partir de la muestra anterior $n-1$. Después la señal diferencia $E(n)$ es escalada al rango de un cuantizador con un tamaño del escalón de 4 bits, quedando:

$$ESN(n) = DV(n) * E(n) \quad (4.2)$$

Esta señal diferencia escalada es cuantizada por el cuantizador adaptable para producir la palabra codificada $IN1(n)$, que es la palabra que se transmite al decodificador

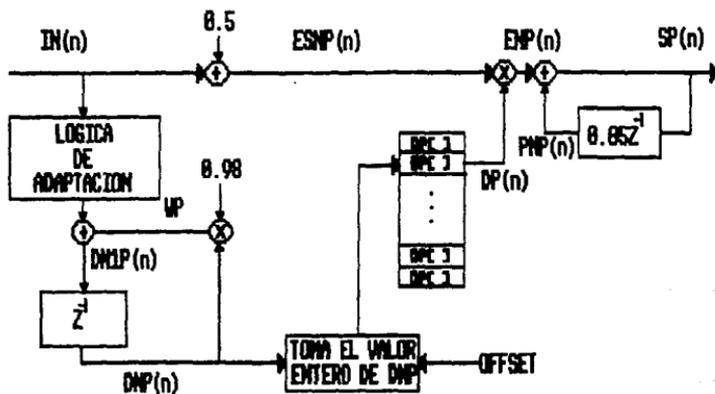


Fig 4.8.b Diagrama a bloques del algoritmo decodificador

ADPCM

En la ecuación 4.2 $DV(n)$ es un factor de escalamiento adaptable que es determinado de los datos de la muestra $n-1$. Esta combinación que se tiene de escalar por el valor de $DV(n)$

y luego realizar la cuantización con un tamaño de escalón fijo, es equivalente a tener un cuantizador con un tamaño de escalón adaptable (el cuál es inversamente proporcional a $DV(n)$).

La figura 4.9 muestra las características para un cuantizador con tamaño de escalón fijo de 4 bits (16 niveles). La señal de entrada $ES(n)$, que consiste de una parte entera y una parte fraccionaria, es convertida en uno de los 16 niveles de cuantización en rango de $-15/2$ a $+15/2$, con sus correspondientes palabras de código $IN1(n)$ en el rango de -8 a $+7$. La señal cuantizada $ESN(n)$ es relacionada con la palabra código $IN1(n)$ por la relación

$$ESN(n) = IN1(n) + 0.5 \quad (4.3)$$

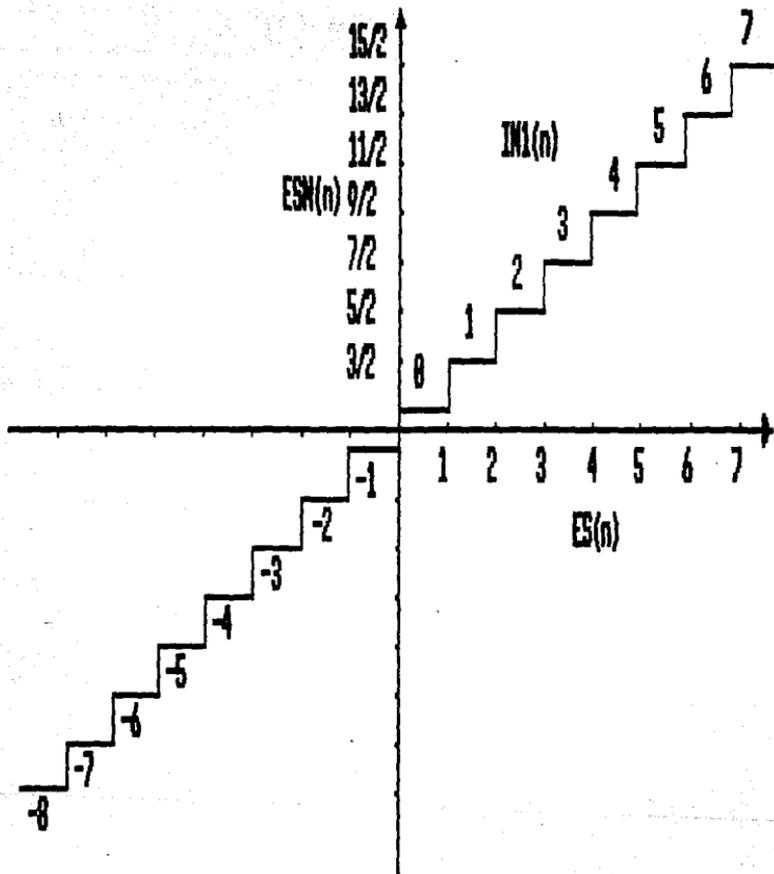


Fig 4.9 Cuantizador fijo PCM de 4 bits

A continuación la señal $ESN(n)$ es desescalada, por reescalar al multiplicar $ESN(n)$ por $D(n)$, donde $D(n)$ es inversamente relacionada con $DV(n)$ y directamente proporcional al tamaño del escalón del cuantizador adaptable. Así nos queda que

$$EN(n) = D(n) * ESN(n) \quad (4.4)$$

es la representación cuantizada de la señal diferencia $EN(n)$. Con el análisis de la ecuación 4.4 se termina el análisis de la sección del cuantizador adaptable ADPCM, ahora veamos cómo funciona la sección del predictor.

4.3.2 Circuito predictor diferencial de primer orden

Continuando con el análisis del codificador ADPCM, ahora veamos la sección predictora, para esto tomemos la suma de $PN(n)$ y $EN(n)$ para formar la representación cuantizada adaptable de la señal de entrada.

$$S(n) = PN(n) + EN(n) \quad (4.5)$$

Además el valor predécido $PN(n+1)$ y los factores para escalar el cuantizador $D(n+1)$ y $DV(n+1)$ se necesitan calcular para poder operar con la siguiente muestra $n+1$. Para calcular el valor nuevo del predictor, multiplicamos $SN(n)$ por GA veces y obtenemos $PN(n+1)$, en la codificación de las señales de voz a GA se le conoce como factor de fuga y usualmente se le asigna el valor de 0.85.

Después de realizados estos cálculos el valor de $PN(n+1)$ se guarda en una localidad de memoria para posteriormente ser utilizado en la muestra $n+1$. Esta sección se muestra en la parte inferior de la figura 4.8.a.

4.3.3 Control de ganancia adaptable del cuantizador

Como se puede observar en la figura 4.8.a para realizar el control del tamaño del escalón en la muestra $n+1$, se necesitan calcular los valores de $D(n+1)$ y $DV(n+1)$, pero hacerlo no es fácil.

Para poder entender como se realiza esto, es necesario recordar lo visto en el capítulo 3 (ecuación 3.22), referente a los sistemas ADPCM robustos (estos sistemas poseen la capacidad de recobrase de los errores pasados que se generaron en el canal de transmisión durante los periodos libres de errores), aquí nos concretaremos a utilizar esta técnica sin demostrarla.

El algoritmo de control del tamaño del escalón del cuantizador ADPCM robusto, motivo de esta tesis se basa en la ecuación

$$D(n+1) = [D(n)]^{\beta} * M[INl(n)] \quad (4.6)$$

donde $\beta = GA$, y el último tamaño del escalón calculado es elevado a la β potencia ($0 < \beta \leq 1$). Como se vió en el capítulo 3, β típicamente se le asigna un valor de 0.98 para los sistemas de codificación de voz.

$D(n)$ además de ser elevado a la β potencia, es escalado por el factor $M[.]$, el cual es función de la palabra código $INl(n)$. Si la palabra código $INl(n)$ ocupa los niveles superiores del cuantizador ($INl(n) = -8, -7, -6, -5, 4, 5, 6, 7$), entonces se le asigna a $M[.]$ un valor superior a la unidad, con el fin de incrementar el tamaño del escalón del cuantizador en la muestra siguiente. Si por el contrario la palabra código tiene un valor que ocupa los niveles inferiores del cuantizador ($INl(n) = -4, -3, -2, -1, 0, 1, 2, 3$), entonces se le asigna un valor menor que la unidad a $M[.]$, para reducir el tamaño del escalón del cuantizador en la muestra siguiente.

De este modo el algoritmo intenta continuamente adaptar el tamaño del escalón de manera que el rango dinámico del cuantizador adaptable sea el mismo que el rango dinámico de la señal $E(n)$.

Ahora, para poder implementar el algoritmo de la ecuación 4.6 en un sistema con microprocesador, obviamente necesitamos eliminar el exponente β . Para hacer esto, se recurre a los logaritmos, que nos indican que cuando se saca el logaritmo de un número con exponente, se convierte en un producto del número y el exponente.

En el caso de la ecuación 4.6 del logaritmo ADPCM robusto, hagamos que

$$d(n) = \log[D(n)] \quad (4.7)$$

$$m[IN1(n)] = \log [M(IN1(n))] \quad (4.8)$$

Por lo que la ecuación 4.6 nos queda

$$d(n+1) = \beta d(n) + m[IN1(n)] \quad (4.9)$$

Por otra parte, si hacemos $\beta d(n)=W$, la ecuación nos queda

$$d(n+1) = W + m[IN1(n)] \quad (4.10)$$

La implementación de la ecuación 4.10 se muestra en la parte derecha de la figura 4.8.a.

Donde $M[INl(n)]$ es función de $INl(n)$ y se determina por el algoritmo

$$m[INl(n)] = \begin{cases} 8 & \text{si } |INl(n) + 0.5| \geq 4.5 \\ -3 & \text{si } |INl(n) + 0.5| < 4.5 \end{cases} \quad (4.11)$$

Estos valores fueron calculados para que cada vez que el algoritmo determine que existe una sobrecarga incremental al doble el tamaño del escalón, y cada vez que la pendiente disminuya en determinada proporción decremente el tamaño del escalón en una tercera parte.

A diferencia de los otros algoritmos, este algoritmo aparte de realizar la función de incrementar y decrementar el tamaño del escalón, también posee la capacidad de hacer que el algoritmo sea robusto.

La salida del circuito de adaptación del tamaño del escalón es la señal $d(n)$, que la ecuación 4.7 muestra como el logaritmo del tamaño del escalón deseado.

Por esto, para obtener $D(n)$ y $DV(n)$, es necesario que se cumplan las relaciones

$$D(n) = \exp[d(n)] \quad (4.12.a)$$

$$DV(n) = \frac{1}{D(n)} = \exp[-d(n)] \quad (4.12.b)$$

para que seé eliminado el logaritmo a los parámetros $D(n)$ y $DV(n)$, al series aplicados una función exponencial.

En este logaritmo los exponenciales de las ecuaciones 4.12.a y 4.12.b son calculados usando tablas donde previamente se han colocado los valores apropiados, tal como se muestra en la figura 4.8.a.

En el sistema implementado con el microprocesador, los valores de las tablas son guardados en la memoria ROM. El valor entero de $d(n)$ es constreñido al rango de $-32 \leq [d(n)] < 32$ para la tabla con 64 valores y después de calculado $d(n)$ es necesario sumarle una compensación (offset) para direccionar correctamente la localización del valor exponencial en la memoria ROM.

El algoritmo usa 64 valores diferentes del tamaño del escalón $D(n)$ y $DV(n)$, estos valores se calcularon de tal manera que abarcaran el rango dinámico de la señal de entrada,

si ocurriera que la señal de entrada tuviera una amplitud más grande o más pequeña que la que puede abarcar el algoritmo, entonces, el tamaño del escalón se saturará respectivamente a los valores máximo ó mínimo de la tabla sin afectar de manera apreciable su funcionamiento.

Las direcciones calculadas de la tabla son guardadas para ser usadas en la muestra $n+1$, y así poder acceder los valores de $D(n+1)$ y $DV(n+1)$, cuando estos valores son requeridos por el algoritmo.

Veremos ahora como se escogieron los valores del tamaño del escalón guardados en la tablas de la memoria ROM, los valores de esta tablas sirven para calcular los valores de $D(n)$ y $DV(n)$ como lo indican las relaciones 4.12.a y 4.12.b.

Estos valores tienen una razón del máximo al mínimo tamaño del escalón de 256, $D[64]/D[1] = 256 = 2^8$ (8 bits) y $DV[1]/DV[64] = 256$. Con esto se obtiene un rango de adaptación de aproximadamente 48 dB (48 dB = $251.2 \rightarrow 256$), que es un valor apropiado para la codificación de las señales de voz.

Como las tablas tienen 64 valores del tamaño del escalón espaciados logarítmicamente, o sea que entre cada valor de la tabla existen $48/64=0.75$ dB de resolución, esto quiere decir entre los valores del tamaño del escalón contenidos en las tablas hay una razón de incremento de 1.09018, que

prácticamente la cerramos a un valor de 1.0902.

Con esta resolución que se cálculo, ahora podemos usar las relaciones 4.13.a y 4.13.b para calcular los valores contenidos en las tablas $D[i]$ y $DV[i]$ incrementando los valores de i , y resolviendo para cada uno de los 64 valores de cada una de las tablas.

$$D[i] = D[1] * (1.0902)^{i-1} \quad (4.13.a)$$

$$DV[i] = DV[1] * (1.0902)^{-i+1} \quad (4.13.b)$$

El máximo tamaño del escalón se calcula de manera que el máximo valor de $E(n)$ que se pueda presentar, escale a el máximo valor del cuantizador (4 bits= 2^4).

La figura 4.8.b muestra el diagrama a bloques para el decodificador robusto ADPCM, la entrada codificada $IN(n)$ es convertida en la señal diferencia $ENP(n)$ al sumarle 0.5 y escalarla por $DP(n)$. Después la señal de salida decodificada $SP(n)$ se obtiene como resultado de la acumulación de los valores de $ENP(n)$ en el circuito de predicción, en una manera semejante como se hace en el codificador del ADPCM para

calcular el nuevo tamaño del escalón.

En general el decodificador ADPCM duplica las características del diagrama a bloques del codificador ADPCM, por lo que no es necesaria una explicación más amplia sobre su funcionamiento.

Este algoritmo tal como se configuro es adecuado para implementarse directamente en el microprocesador TMS32010. El programa completo implementado en el microprocesador TMS32010 se muestra en la figura 4.10, este programa resulta de ensamblar el programa fuente originalmente editado, con el programa ensamblador XDS/320. La ventaja de esta presentación del programa es que nos muestra en sus columnas la información necesaria para comprender como esta estructurado el algoritmo.

En la primera columna se tiene de renglones de programa escritos, en la segunda se muestra el número de direcciones de memoria, en la siguiente columna tenemos el código de operación de cada una de las instrucciones guardadas en memoria, y en las columnas restantes tenemos el programa fuente como originalmente fué editado.

```

0001      >
*****  BLANK MISSING
*****  INVALID OPCODE
0002      *****
0003      *      PROGRAMA PARA LA CODIFICACION DE VOZ POR MEDIO DEL  *
0004      *      ALGORITMO "ADPCM"                                     *
0005      *****
0006      *
0007      *
0008      * SUBROUTINA PARA GENERAR LOS ESPACIOS DE CUANTIZACION
0009      *
0010 0000      AORG      0
0011 0000 F900      B      INICIO
          0001 009B
0012 0002 0270  RATE  DATA  >270
0013 0003 00FA  MODE  DATA  >FA
0014      *
0015      * TABLA DE LOS VALORES DE "DV"
0016      *
0017 0004 5000      DATA  >5000      * FORMATO Q12
0018 0005 4962      DATA  >4962
0019 0006 4350      DATA  >4350
0020 0007 3DBE      DATA  >3DBE
0021 0008 38A3      DATA  >38A3
0022 0009 33F3      DATA  >33F3
0023 000A 2FA7      DATA  >2FA7
0024 000B 2BB6      DATA  >2BB6
0025 000C 2818      DATA  >2818
0026 000D 24C7      DATA  >24C7
0027 000E 21BC      DATA  >21BC
0028 000F 1EF2      DATA  >1EF2
0029 0010 1C63      DATA  >1C63
0030 0011 1A09      DATA  >1A09
0031 0012 17E2      DATA  >17E2
0032 0013 15E8      DATA  >15E8
0033 0014 1418      DATA  >1418
0034 0015 126F      DATA  >126F
0035 0016 10E8      DATA  >10E8
0036 0017 0F82      DATA  >0F82
0037 0018 0E3A      DATA  >0E3A
0038 0019 0D0D      DATA  >0D0D
0039 001A 0BF8      DATA  >0BF8
0040 001B 0AFB      DATA  >0AFB
0041 001C 0A12      DATA  >0A12
0042 001D 093D      DATA  >093D
0043 001E 0879      DATA  >0879
0044 001F 07C6      DATA  >7C6
0045 0020 0721      DATA  >721
0046 0021 068A      DATA  >68A
0047 0022 0600      DATA  >600
0048 0023 0581  IOFSET DATA  >581
0049 0024 050C      DATA  >50C
0050 0025 04A1      DATA  >4A1
0051 0026 043F      DATA  >43F
0052 0027 03E5      DATA  >3E5
0053 0028 0393      DATA  >393
    
```

0054 0029 0347 DATA >347
NO\$IDT 32010 FAMILY MACRO ASSEMBLER

FC2.1-84.107 19:37:22 10-07-87
PAGE 0002

0055 002A 0302 DATA >302
0056 002B 02C2 DATA >2C2
0057 002C 0288 DATA >288
0058 002D 0252 DATA >252
0059 002E 0221 DATA >221
0060 002F 0200 DATA >200
0061 0030 0200 DATA >200
0062 0031 0200 DATA >200
0063 0032 0200 DATA >200
0064 0033 0200 DATA >200
0065 0034 0200 DATA >200
0066 0035 0200 DATA >200
0067 0036 0200 DATA >200
0068 0037 0200 DATA >200
0069 0038 0200 DATA >200
0070 0039 0200 DATA >200
0071 003A 0200 DATA >200
0072 003B 0200 DATA >200
0073 003C 0200 DATA >200
0074 003D 0200 DATA >200
0075 003E 0200 DATA >200
0076 003F 0200 DATA >200
0077 0040 0200 DATA >200
0078 0041 0200 DATA >200
0079 0042 0200 DATA >200
0080 0043 0200 DATA >200

0081
0082
0083

*
* TABLA DE LOS VALORES DE "D"
*

0084 0044 0333 DATA >333
0085 0045 06B0 DATA >6B0
0086 0046 03CE DATA >3CE
0087 0047 0425 DATA >425
0088 0048 0485 DATA >485
0089 0049 04EE DATA >4EE
0090 004A 055F DATA >55F
0091 004B 05DB DATA >5DB
0092 004C 0662 DATA >662
0093 004D 06F6 DATA >6F6
0094 004E 0797 DATA >797
0095 004F 0846 DATA >846
0096 0050 0905 DATA >905
0097 0051 09D5 DATA >9D5
0098 0052 0AB8 DATA >0AB8
0099 0053 0BAF DATA >0BAF
0100 0054 0CBD DATA >0CBD
0101 0055 0DE3 DATA >0DE3
0102 0056 0F24 DATA >0F24
0103 0057 1082 DATA >1082
0104 0058 11FF DATA >11FF
0105 0059 139E DATA >139E
0106 005A 1563 DATA >1563
0107 005B 1751 DATA >1751
0108 005C 196B DATA >196B
0109 005D 1BB6 DATA >1BB6
0110 005E 1E36 DATA >1E36

* FORMATO Q12

0112	0060	23E8		DATA	>23E8
0113	0061	2724		DATA	>2724
0114	0062	2D12		DATA	>2D12
0115	0063	2E85	OFSET	DATA	>2E85
0116	0064	32B7		DATA	>32B7
0117	0065	374A		DATA	>374A
0118	0066	3C47		DATA	>3C47
0119	0067	41B6		DATA	>41B6
0120	0068	47A4		DATA	>47A4
0121	0069	4E19		DATA	>4E19
0122	006A	5525		DATA	>5525
0123	006B	5CD2		DATA	>5CD2
0124	006C	6531		DATA	>6531
0125	006D	6E52		DATA	>6E52
0126	006E	7844		DATA	>7844
0127	006F	7FFF		DATA	>7FFF
0128	0070	7FFF		DATA	>7FFF
0129	0071	7FFF		DATA	>7FFF
0130	0072	7FFF		DATA	>7FFF
0131	0073	7FFF		DATA	>7FFF
0132	0074	7FFF		DATA	>7FFF
0133	0075	7FFF		DATA	>7FFF
0134	0076	7FFF		DATA	>7FFF
0135	0077	7FFF		DATA	>7FFF
0136	0078	7FFF		DATA	>7FFF
0137	0079	7FFF		DATA	>7FFF
0138	007A	7FFF		DATA	>7FFF
0139	007B	7FFF		DATA	>7FFF
0140	007C	7FFF		DATA	>7FFF
0141	007D	7FFF		DATA	>7FFF
0142	007E	7FFF		DATA	>7FFF
0143	007F	7FFF		DATA	>7FFF
0144	0080	7FFF		DATA	>7FFF
0145	0081	7FFF		DATA	>7FFF
0146	0082	7FFF		DATA	>7FFF
0147	0083	7FFF		DATA	>7FFF

0148 *
 0149 * INICIALIZACION DE LAS VARIABLES
 0150 *

0151	0084	0008	MINPA	DATA	>8
0152	0085	0008	MINA	DATA	>8
0153	0086	0000	DN1A	DATA	>0
0154	0087	0000	DN1PA	DATA	>0
0155	0088	0000	PNA	DATA	>0
0156	0089	0000	PNPA	DATA	>0
0157	008A	0000	SNA	DATA	>0
0158	008B	0000	SPA	DATA	>0
0159	008C	0000	WPA	DATA	>0
0160	008D	0000	WA	DATA	>0
0161	008E	1000	DA	DATA	>1000
0162	008F	1000	DVA	DATA	>1000
0163	0090	1000	DPA	DATA	>1000
0164	0091	0D9A	P85A	DATA	>D9A
0165	0092	0800	PT5A	DATA	>800
0166	0093	0008	MASK1A	DATA	>8
0167	0094	0008	MASK2A	DATA	>8

*
 * 1 EN Q12
 *

0169	0096	00FB	GAA	DATA	>FB
0170	0097	0009	NUEVO	DATA	>9
0171	0098	FFFD	MINU	DATA	-3
0172	0099	0020	P32A	DATA	>20
0173	009A	FFEL	MP31A	DATA	>FFEL
0174			*		
0175			* LOCALIDADES DE MEMORIA UTILIZADAS		
0176			*		
0177		0000	MINP	EQU	>0
0178		0001	MIN	EQU	>1
0179		0002	DN1	EQU	>2
0180		0003	DN1P	EQU	>3
0181		0004	PN	EQU	>4
0182		0005	PNP	EQU	>5
0183		0006	SN	EQU	>6
0184		0007	SP	EQU	>7
0185		0008	WP	EQU	>8
0186		0009	W	EQU	>9
0187		000A	D	EQU	>A
0188		000B	DV	EQU	>B
0189		000C	DP	EQU	>C
0190		000D	P85	EQU	>D
0191		000E	PT5	EQU	>E
0192		000F	MASK1	EQU	>F
0193		0010	MASK2	EQU	>10
0194		0011	MENOS	EQU	>11
0195		0012	GA	EQU	>12
0196		0013	NUEVE	EQU	>13
0197		0014	MINU3	EQU	>14
0198		0015	P32	EQU	>15
0199		0016	MP31	EQU	>16
0200		0017	E	EQU	>17
0201		0018	S	EQU	>18
0202		0019	ES	EQU	>19
0203		001A	ESN	EQU	>1A
0204		001B	IN1	EQU	>1B
0205		001C	EN	EQU	>1C
0206		001D	DN	EQU	>1D
0207		001E	AB	EQU	>1E
0208		001F	ONE	EQU	>1F
0209		0020	DN1B	EQU	>20
0210		0021	IN	EQU	>21
0211		0022	ESNP	EQU	>22
0212		0023	ENP	EQU	>23
0213		0024	ABC	EQU	>24
0214		0025	DNP	EQU	>25
0215		0026	DIR	EQU	>26
0216		0027	DIRD	EQU	>27
0217	009B	6800	INICIO	LDPK	0
0218	009C	7E02		LACK	RATE
0219	009D	6700		TBLR	0
0220	009E	4900		OUT	0,1
0221	009F	7E03		LACK	MODE
0222	00A0	6700		TBLR	0
0223	00A1	4800		OUT	0,0
0224			*		

0225
NO\$IDT

* SECCION PARA PASAR LOS DATOS DE LA MEMORIA DE PROGRAMA
32010 FAMILY MACRO ASSEMBLER PC2.1 84.107 19:37:22 10-07-87
PAGE 0005

0226 * A LA MEMORIA DE DATOS EN FORMA INDIRECTA
0227 *
0228 00A2 7E01 LACK 1
0229 00A3 501F SACL ONE * EL CONTENIDO DE ONE ES 1
0230 00A4 7016 LARK AR0,MP31
0231 00A5 7116 LARK AR1,22 * NUMERO DE DATOS - 1 = AR1
0232 00A6 7E9A LACK MP31A
0233 00A7 6880 LOAD LARP AR0
0234 00A8 6791 TBLR *- ,AR1
0235 00A9 101F SUB ONE
0236 00AA F400 BANZ LOAD * PRUEBA SI AR1 ES CERO
00AB 00A7

0237 *
0238 * SE INICIA LA CODIFICACION DE LA SENAL DE VOZ
0239 *

0240 00AC F600 ALOOP BIOZ ALOOP1
00AD 00B0
0241 00AE F900 B ALOOP
00AF 00AC
0242 00B0 4218 ALOOP1 IN S,2
0243 00B1 7F89 ZAC
0244 00B2 6A06 LT SN
0245 00B3 6D0D MPY P85
0246 00B4 7F8E PAC
0247 00B5 0E1F ADD ONE,14 * VALOR USADO PARA REDONDEO
0248 00B6 5C04 SACH PN,4 * PN=SN*AL
0249 00B7 2018 LAC S
0250 00B8 1004 SUB PN
0251 00B9 501F SACL E * E=S-PN
0252 00BA 6A17 LT E
0253 00BB 6D0B MPY DV
0254 00BC 7F8E PAC
0255 00BD 0E1F ADD ONE,14 * VALOR USADO PARA REDONDEO
0256 00BE 5C19 SACH ES,4 * ES=E*DV
0257 00BF F800 CALL QUANT
00C0 00D4

0258 * OUT IN1,2
0259 00C1 7F89 ZAC
0260 00C2 2C1B LAC IN1,12
0261 00C3 000E ADD PT5
0262 00C4 501A SACL ESN * ESN=IN+.5
0263 00C5 6A1A LT ESN
0264 00C6 6D0A MPY D
0265 00C7 7F8E PAC
0266 00C8 0E1F ADD ONE,14 * VALOR USADO PARA REDONDEO
0267 00C9 5C1C SACH EN,4 * EN=ESN*D
0268 00CA 201C LAC EN
0269 00CB 0004 ADD PN
0270 00CC 5006 SACL SN * SN=EN+PN
0271 00CD F800 CALL ESCAL * GOSUB 2500
00CE 008E

0272 *
0273 * SE INICIA LA DECODIFICACION DE VOZ
0274 *

0275 00CF F800 CALL DECOD
00D0 0123

0276
NO\$IDT

32010 FAMILY MACRO ASSEMBLER PC2.1 84.107

19:37:22 10-07-87
PAGE 0006

```
0277      * SALIDA DE LA VOZ PROCESADA *
0278      *****
0279 00D1 4A07      OUT      SP,2
0280      *****
0281 00D2
0282 00D2 F900      B          ALOOP      * REGRESA A TOMAR OTRA MUESTRA
0283 00D3 00AC
0284 00D4
0284 00D4 2419  QUANT    LAC      ES,4      * SUBROUTINA DE CUANTIZACION
0285 00D5 581B      SACH    IN1
0286 00D6 201B      LAC      IN1
0287 00D7 FA00      BLZ      NEG
0288 00D9 790F      AND      MASK1
0289 00DA FC00      BGZ      CCC
0290 00DB 00DE
0290 00DC F900      B          DDD
0291 00DD 00ED
0291 00DE 7E07  CCC      LACK    >7      * SI IN1>7 ENTONCES IN1=7
0292 00DF 501B      SACL    IN1
0293 00E0 F900      B          DDD
0294 00E1 00ED
0294 00E2 7F88  NEG      ABS
0295 00E3 1011      SUB      MENOS
0296 00E4 7910      AND      MASK2
0297 00E5 FC00      BGZ      EEE
0298 00E6 00E9
0298 00E7 F900      B          DDD
0299 00E8 00ED
0299 00E9 6A10  EEE      LT      MAS,2      * SI IN1<8 ENTONCES IN1=-8
0300 00EA 9FFF      MPYK    -1
0301 00EB 7F8E      PAC
0302 00EC 501B      SACL    IN1
0303 00ED 7F8D  DDD      RET
0304 00EE 2C1B  ESCAL    LAC      IN1,12      * SUB DEL TAMANO DEL ESCALON
0305 00EF 000E      ADD      PT5      * CONSIDERO FORMATO Q8 CON UN
0306 00F0 7F88      ABS      * CORRIMIENTO DE 4 BITS
0307 00F1 501E      SACL    AB
0308 00F2 6A1E      LT      AB
0309 00F3 8002      MPYK    >2
0310 00F4 7F8E      PAC
0311 00F5 5C1E      SACH    AB,4      * SALVA SOLO LA PARTE ENTERA
0312 00F6 201E      LAC      AB
0313 00F7 1013      SUB      NUEVE
0314 00F8 FD00      BGEZ    FFF
0315 00F9 00FE
0315 00FA 2014      LAC      MINU3
0316 00FB 5001      SACL    MIN
0317 00FC F900      B          GGG      * MIN=-3
0318 00FD 0100
0318 00FE 7E08  FFF      LACK    >8
0319 00FF 5001      SACL    MIN      * MIN=8
0320 0100 2802  GGG      LAC      DN1,8      * CAMBIO A FORMATO Q8
0321 0101 501D      SACL    DN
0322 0102 6A1D      LT      DN
0323 0103 6D12      MPY     CA
```

0324 0104 7F8E PAC * W=GA*DN
 NO\$IDT 32010 FAMILY MACRO ASSEMBLER PC2.1 84.107 19:37:22 10-07-87
 PAGE 0007

0325 0105 0E1F ADD ONE, 14 * VALOR USADO PARA REDONDEO
 0326 0106 6001 ADDH MIN
 0327 0107 5802 SACH DN1 * DN1=MIN+W
 0328 0108 F800 CALL CONST
 0109 0113

0329 *
 0330 * DIRECCIONAMIENTO DE LOS VALORES DE TABLAS DE "D" Y "DV"
 0331 *

0332 010A 281D LAC DN, 8
 0333 010B 5826 SACH DIR * AISLANDO LA PARTE ENTERA
 0334 010C 7E63 LACK OFSET
 0335 010D 0026 ADD DIR
 0336 010E 670A TBLR D * COLOCA EL VALOR EN D
 0337 010F 7E23 LACK IOFSET
 0338 0110 0026 ADD DIR
 0339 0111 670B TBLR DV * COLOCA EL VALOR EN DV
 0340 0112 7F8D RET

0341 *
 0342 * SUBROUTINA PARA CONSTREINIR EL VALOR DE DN1 A LOS RANGOS
 0343 * [32,-31]
 0344 *

0345 0113 2002 CONST LAC DN1
 0346 0114 FC00 BGZ HHH
 0115 011D
 0347 0116 0015 ADD P32
 0348 0117 FC00 BGZ SALID
 0118 0122
 0349 0119 2016 LAC MP31
 0350 011A 5002 SACL DN1
 0351 011B F900 B SALID
 011C 0122

0352 011D 1015 HHH SUB P32
 0353 011E FB00 BLEZ SALID
 011F 0122

0354 0120 2015 LAC P32
 0355 0121 5002 SACL DN1
 0356 0122 7F8D SALID RET

0357 *****
 0358 *SUBROUTINA DE DECODIFICACION*
 0359 *****
 0360 *

0361 0123 7F89 DECOD ZAC
 0362 0124 2C1B LAC IN1, 12
 0363 0125 000E ADD PT5
 0364 0126 5022 SACL ESNP * ESNP=IN1+.5
 0365 0127 6A22 LT ESNP
 0366 0128 6DOC MPY DP
 0367 0129 7F8E PAC

0368 012A 0E1F ADD ONE, 14 * VALOR USADO PARA REDONDEO
 0369 012B 5C23 SACH ENP, 4 * ENP=ESNP*DP
 0370 012C 6A07 LT SP
 0371 012D 6D0D MPY P85

0372 012E 7F8E PAC
 0373 012F 0E1F ADD ONE, 14 * VALOR USADO PARA REDONDEO
 0374 0130 5C05 SACH PNP, 4 * PNP=AL*SP
 0375 0131 2023 LAC ENP


```
0427 0161 670C      TBLR      DP
NO$IDT      32010 FAMILY MACRO ASSEMBLER  PC2.1 84.107  19:37:22 10-07-87
                                           PAGE 0009
```

```
0428 0162 7F8D      RET
0429          END
0002 ERRORS, 0000 WARNINGS, LAST ERROR AT 0001
NO$IDT      32010 FAMILY MACRO ASSEMBLER  PC2.1 84.107  19:37:37 10-07-87
                                           PAGE 0010
```

```
0001
***** <
***** BLANK MISSING
***** INVALID OPCODE
0002 0000
0003 0000
0004 0000
0005 0000
0006 0000
0007 0000
0008 0000
0009 0000
0010 0000
0011 0000
0012 0000
0002 ERRORS, 0000 WARNINGS, LAST ERROR AT 0001
ESTE PROGRAMA SE LLAMA EN LA VAX ADPCM.LST;2
```

Fig 4.10 Programa en lenguaje ensamblador TMS32010 del
algoritmo ADPCM robusto a 32 Kb/s

El programa y los valores de las tablas de $DV(n)$, $D(n)$ y $DP(n)$ ocupan 350 palabras de memoria ROM. El algoritmo sin los valores de las tablas ocupa 182 palabras de la memoria ROM, por esto el tiempo máximo que tarda el TMS32010 en ejecutar las instrucciones es de 36.4 μ seg.

El programa utiliza 28 localidades de memoria RAM y el microprocesador tiene integrados 288 bytes de memoria RAM, así que teóricamente el TMS32010 puede procesar 10 señales de entrada simultáneamente sin tenerle que colocar memoria RAM externa, la cual es manejada por el TMS32010 en forma más lenta.

El programa comienza con los valores de las tablas de D y DV (en el programa no se utiliza la nomenclatura de $x(n)$, sólo x para mayor facilidad de comprensión del programa). Las tablas fueron editadas en formato Q12, lo que quiere decir que se tiene 1 bit de signo, 3 bits para la parte entera y 12 bits para la parte fraccional, lo que nos indica que los valores de las tablas fueron escalados a valores menores a 8 (para comprender como funcionan los formatos Q referirse al apéndice C).

Después de los valores de las tablas se encuentra la inicialización de variables, éstos valores suponen que la señal de entrada se comportará en un principio de una manera determinada, sin embargo en general no todas las señales de entrada les funcionan estos valores de inicialización, y por

esto la señal que se predice no es fiel a la señal de entrada, haciendo que la señal reconstruida en el decodificador no se parezca a la señal original.

Este fenómeno se corrige por si sólo después de procesadas algunas muestras, el fenómeno se puede observar en el comienzo de las gráficas de las figuras 4.1-4.3.

Utilizando las pseudoinstrucciones EQU en el programa, se asignaron las localidades de la memoria RAM que se utilizaron para cada una de las variables. Una vez terminada la inicialización empieza realmente el funcionamiento del algoritmo para procesar las señales de voz, comprender esta parte del programa es directa, sólo es necesario poner atención en los formatos Q y los valores de redondeo utilizados.

Redondear los resultados en los cálculos de los microprocesadores de punto fijo como el TMS32010 es muy importante en aplicaciones de estos al procesamiento digital de señales, si los resultados no son redondeados causaran que el sistema sea inestable. Es por esto que a través del programa los resultados usualmente fueron redondeados.

CAPITULO V

RESULTADOS

RESULTADOS, CONCLUSIONES Y RECOMENDACIONES SOBRE EL DESARROLLO DEL SISTEMA ADPCM ROBUSTO

5.1 Introducción

En este capítulo presentaremos los resultados obtenidos utilizando el sistema de evaluación EVM y la tarjeta de interface analógica AIB, ambos de la compañía Texas Instruments. Estas dos tarjetas se encuentran en el Laboratorio del Departamento de Comunicaciones del IIE, se muestran en las figuras 5.1.a y 5.1.b. Todas las pruebas fueron realizadas con el equipo existente en este laboratorio.

A grandes rasgos las pruebas realizadas se presentan en dos bloques: pruebas técnicas y pruebas subjetivas. El primer bloque consiste de las pruebas de respuesta en frecuencia y de razón de señal a ruido de cuantización. El segundo bloque consiste de una apreciación subjetiva de la calidad de la señal de voz reproducida, después de ser procesada por el algoritmo ADPCM Robusto.

5.2 Caracterización técnica y subjetiva

Antes de ver como se hizo la realización de las pruebas técnicas y subjetivas se verá una descripción del equipo y del procedimiento empleado para la simulación en tiempo real del algoritmo ADPCM Robusto en el Laboratorio del Departamento de comunicaciones.

La tarjeta de la interface analógica AIB (Analog Interface Board), permite tomar muestras de la señal de entrada con una frecuencia programable entre 76.29 Hz y 34.48 KHz, convirtiéndolas en un número binario compatible con el formato Q₁₅, este número lo entrega al sistema de evaluación EVM para ser procesado por el microprocesador TMS32010. El único requisito que tiene que cumplir la señal de entrada es tener niveles máximos de voltaje de ± 10 Volts. La tarjeta de interface analógica se conecta al EVM (Evaluation Module) como lo muestra la figura 5.2.

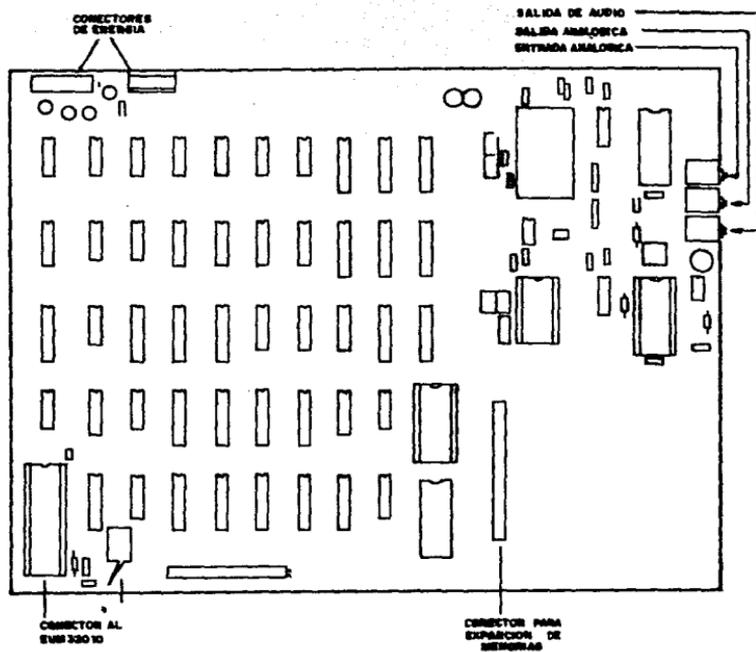


Fig 5.1.a Tarjeta de Interface Analógica (AIB)

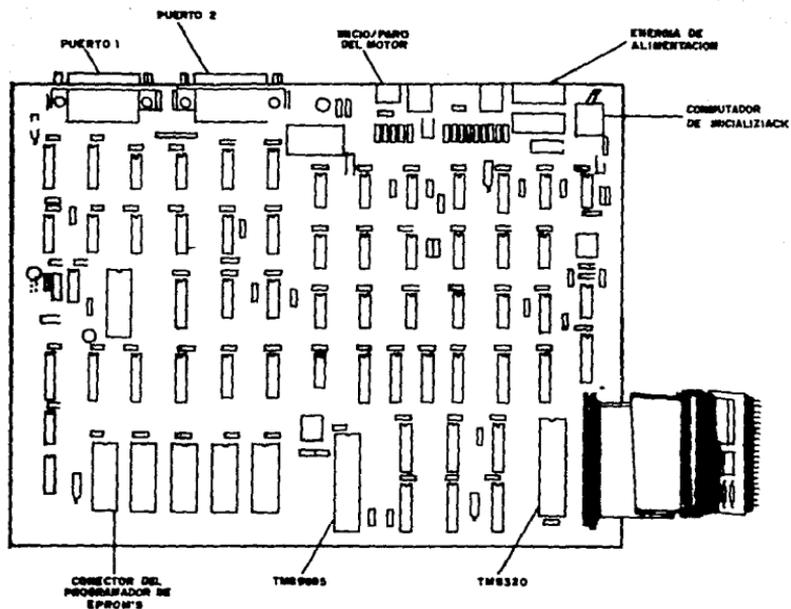


Fig 5.1.b Módulo de Evaluación (EVM)



Fig 5.2 Configuración del Sistema Emulador/AIB

El EVM es un sistema de evaluación de la Texas Instruments, con el cual es posible simular el comportamiento del microprocesador TMS32010 en condiciones de operación reales. El EVM sirve para desarrollar y remover fallas de los programas de los algoritmos, esta tarjeta tiene dos puertos con conectores RS-232 mediante los cuales se enlaza con una terminal de computadora (Puerto P1), y con el CPU de algún sistema de cómputo (Puerto P2).

Se tienen dos procedimientos para la simulación en tiempo real de los algoritmos en el Laboratorio del Departamento de Comunicaciones del IIE, en esta tesis se menciona uno de estos procedimientos:

Se conectan las tarjetas AIB y EVM a una computadora IBM PC/AT (Puerto P2) y a una terminal de computadora (Puerto P1) como se muestra en la figura 5.3.

FUENTE DE PODER

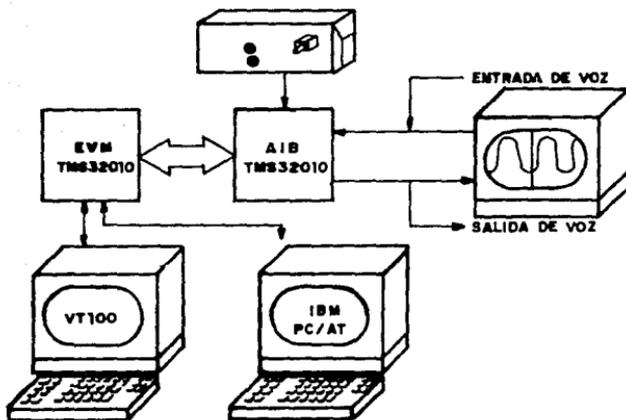


Fig 5.3 Sistema empleado para la simulación en tiempo real del Algoritmo ADPCM Robusto

El sistema permite por un lado la programación del microprocesador TMS32010 através de archivos editados en la computadora IBM PC/AT, y por otro lado permite el control del sistema de evaluación EVM através de la terminal de computadora VT100.

Para hacer las pruebas del programa ADPCM Robusto, se editó el programa del algoritmo en la computadora IBM PC/AT en lenguaje Turbo Pascal. Este programa debe de incluir como primer carácter el signo ">" para indicar al EVM que se inicia el programa, al final de las últimas instrucciones del programa se debe de escribir como último carácter del programa el signo "<" para indicar el fin del programa.

El programa antes de ser ejecutado en el EVM se simula por medio de los programas de soporte que proporciona la Texas Instruments para este fin. Primero se ensambla el programa por medio del programa XDS/320 Macro Assembler/Linker, en este paso se genera un programa que es apropiado para cargarlo en el programa XDS/320 Simulator, con el que se realiza la simulación paso a paso y en forma continua del algoritmo ADPCM Robusto, removiéndose las fallas existentes y depurándolo hasta que trabajó de manera correcta.

Es importante resaltar que cuando se presentaron fallas de ejecución en el programa del algoritmo ADPCM Robusto, estas fueron detectadas y corregidas por medio del programa simulador XDS/320 mejor que trabajando el programa directamente en la memoria del EVM, por ser este procedimiento más práctico de realizar.

Cuando el algoritmo trabajó de manera correcta en el programa simulador XDS/320, se procedió a realizar su ensamblado en la memoria del EVM para la realización de las pruebas del programa en tiempo real. Como primer paso se llama un programa contenido en al IBM PC/AT editado en BASIC llamado TTY, que sirve para hacer la transferencia asincrona del archivo contenido en la IBM PC/AT hacia la memoria del EVM.

Se corre el programa TTY, y este programa pregunta la velocidad de transmisión (SPEED ?), la velocidad se coloca a 9600 B/S que es la velocidad de recepción del puerto 2 del EVM, al mismo tiempo la terminal VT100 se coloca a una velocidad de transmisión de 9600 B/S, el EVM ajusta su velocidad de recepción automáticamente del puerto 1 a esta velocidad.

En la terminal VT100 también se teclea INIT, RETURN (CLOCK SOURCE INTERNAL ?), a esta pregunta se teclea E para indicar al EVM que el reloj será externo, enseguida se teclea ASM 2, y se deja esperando la terminal mientras se termina de dar instrucciones a la IBM PC/AT donde se presionan juntas las teclas CTRL y E (FILE ?), a esta pregunta se teclea el nombre del archivo que se transmitirá, enseguida se presiona RETURN, apareciendo en pantalla TRANSMIT OR RECEIVE ? , a esta pregunta se teclea T.

Ahora todo esta listo para el ensamblado del programa, se teclea RETURN primero en la terminal VT100 y después en la IBM PC/AT, y con esto se realiza el ensamblado del programa del algoritmo ADPCM Robusto en la memoria del EVM.

Una vez que el programa se tiene en la memoria del EVM se procede a ejecutarlo utilizando el comando RUN cuando se desea ejecutarlo sin puntos de interrupción, en este caso se puede detener la ejecución del programa accionando el conmutador de paro del EVM, o con el comando EX para ejecutar sólo una parte del programa.

Si es necesario en algún momento conocer el contenido de la memoria de datos o de la memoria de programa del microprocesador TMS32010 se usan los comandos DDM M N (Modify Program Memory) con los cuales mostrará el contenido de la localidad N de memoria solicitada y dejará el cursor frente a este por si se desea hacer algún cambio.

Para colocar los puntos de interrupción en el programa se emplea el comando SB N (Set Breackpoints) donde N es el número (entre 1 y 8) que corresponde al punto de interrupción en el EVM. Para limpiar los puntos de interrupción se utiliza el comando CB N (Clear Breackpoints).

Para realizar la prueba de respuesta en frecuencia se realizan los pasos descritos anteriormente para el ensamblado del algoritmo ADPCM y se ejecuta el programa con el comando

RUN. A este sistema (figura 5.3) se le inyecta una señal senoidal de ± 4 Volts y se varía la frecuencia de esta señal desde 200 Hz hasta 3.2 KHz, midiéndose la amplitud de la señal recuperada; la gráfica resultante se muestra en la figura 5.4.

Para evaluar el efecto de cuantizar una señal y el ruido generado por este efecto, se verá teóricamente como se genera este ruido en un sistema con un cuantizador uniforme. Representando las muestras cuantizadas $IN1(n)$ como:

$$IN1(n) = ES(n) + e(n) \quad (5.1)$$

donde $ES(n)$ es la muestra sin cuantizar y $e(n)$ es el error de cuantización (ver figura 5.5) Si definimos B como los bits contenidos en las palabras de código binario del cuantizador uniforme característico y Δ como el tamaño del escalón, entonces:

$$-\frac{\Delta}{2} \leq e(n) \leq \frac{\Delta}{2} \quad (5.2)$$

Ahora es posible relacionar la intensidad del ruido de cuantización a la intensidad de la señal de entrada, para esto se define el ratio de la señal a ruido de cuantización como:

$$\text{SNR} = \frac{\tau^2}{\tau_e^2} \quad (5.3)$$

Después de hacer algunos manejos matemáticos la ecuación 5.3 queda:

$$\text{SNR}(\text{dB}) = 6B - 7.2 \quad (5.4)$$

La ecuación 5.4 indica que por cada bit de la palabra código se incrementa en 6 dB la razón señal a ruido de cuantización; sin embargo lo visto anteriormente sobre el SNR sólo es válido para un sistema con un cuantizador uniforme, como es el caso de los moduladores delta y DPCM, no así de los moduladores ADPCM que trabajan con un cuantizador no-uniforme (adaptable).

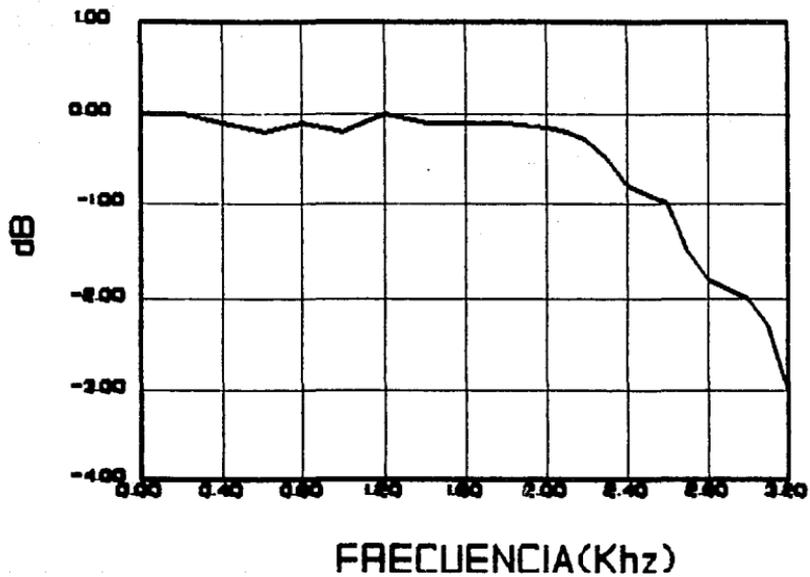


Fig 5.4 Respuesta en frecuencia del ADPCM Robusto

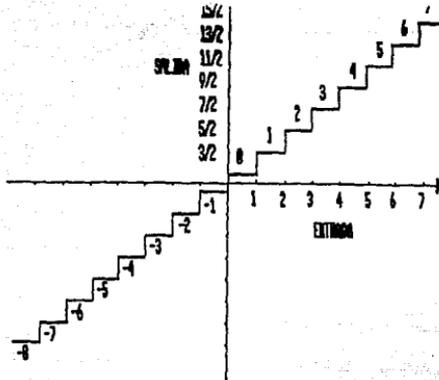


Fig 5.5 Cuantizador Uniforme Característico

Una buena medición de la calidad de la voz es esencial en los programas e investigaciones de los sistemas de procesamiento de voz. En la codificación de voz se necesitan mediciones de calidad para comprender las diferentes distorsiones presentes en la salida del decodificador. Sin embargo las evaluaciones de la calidad de la voz implica atributos psicológicos y subjetivos de la percepción humana.

La evaluación de la calidad de la voz nunca ha sido establecida matemáticamente de manera satisfactoria, la razón señal a ruido de cuantización (SNR), que es ampliamente usada para ambientes de transmisión/recepción de señales es una medida inefectiva de la calidad de la voz. Las mediciones Itakura-Saito usadas en la evaluación de la ejecución de los

codificadores de voz han mostrado ser más efectivos para la medición de la calidad de la voz, esto por tomar en cuenta algunos atributos subjetivos que el SNR no toma en cuenta, sin embargo en el Laboratorio del Departamento de Comunicaciones del IIE no se cuenta con equipo para realizar mediciones de este tipo.

La definición del SNR diverge de los rangos de calidad subjetiva debido a que las regiones con señales de gran amplitud dominan la influencia de las regiones de señales de pequeña amplitud en el cálculo del SNR.

Por lo visto anteriormente se puede concluir que para nuestro caso la evaluación subjetiva es más importante que la evaluación técnica y no tiene sentido realizar evaluaciones del SNR, por no ser estas mediciones lo suficientemente correlacionadas con la calidad de la voz percibida por el ser humano.

Las pruebas de la evaluación subjetiva se hizo basada en la apreciación humana de calidad de la voz a la salida del decodificador (tarjeta AIB), los resultados se muestran en la tabla de la figura 5.6.

	Ruido de Estática	Ruido de Bote	Integridad y Reconocibilidad
ADPCM Robusto	Mínimo	Nulo	Buena

Fig 5.6 Resultados de la evaluación subjetiva

5.3 Estimado económico del proyecto

Para la construcción física del ADPCM Robusto se sugiere la configuración mostrada en la figura 5.7.

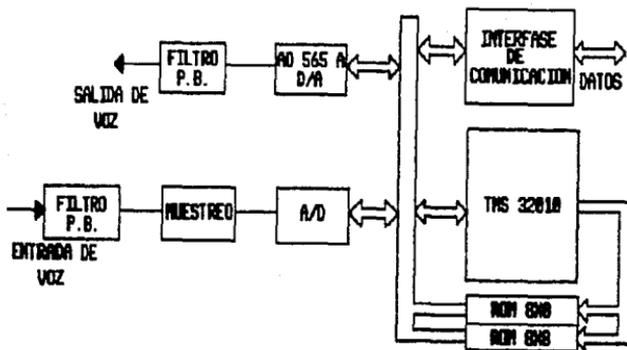


Fig 5.7 Configuración propuesta para la construcción física del ADPCM Robusto

Se utilizan dos circuitos integrados ROM, debido a que el bus de datos del sistema es de 16 bits y las memorias ROM sólo son de 8 bits.

La señal de entrada es digitalizada por el convertidor analógico-digital y mediante el bus de datos es llevado al TMS32010, el cuál lee el dato y ejecuta las instrucciones del algoritmo almacenado en la memoria ROM correspondientes al codificador ADPCM Robusto, el dato resultante es enviado mediante el bus hacia la interface de comunicación serie para su transmisión hacia el receptor (decodificador).

En el funcionamiento del sistema de la figura 5.7 como decodificador, el dato es leído a través de la interface de comunicación serie, llevado por el bus de datos hacia el TMS32010, el cuál lee el dato y ejecuta la parte del algoritmo del decodificador, el dato resultante de este proceso es enviado por medio del bus de datos hacia el convertidor digital-analógico que se encarga de la conversión de la señal numérica a una señal analógica (en nuestro caso voz).

El estimado económico del hardware se muestra en la siguiente tabla:

Componente	Precio (dol.)	Función
TMS32010	10.00	microproc.
ADC 80	13.45	convertidor AD
AD 565	12.40	convertidor DA
LF398	1.95	muestreador
2764(2)	12.00	EPROM
Interface IIE	50.00	comunicaciones
Circuiteria	10.00	lógica y comp.
TOTAL	109.80	

Como se puede observar en la tabla existe el dato de una interface que se esta desarrollando en el IIE, esta cumple una función analoga como el integrado 8251 que es una interface para comunicaciones programable, el costo dado es proximadamente el que tendrá cuando este terminada. No se utiliza la interface 8251 porque no puede transmitir palabras de 4 bits (sólo de 5 a 8 bists), que es lo que se requiere para el ADPCM Robusto.

El costo total del circuito es elevado (el integrado CVDS 3417, \$15.00, 32 Kbs/s), pero con la ventaja de poder utilizar el mismo esquema para algoritmos que reducen más la tasa de bits transmitidos a menos de 9.6 Kbs/s, que es el objetivo siguiente, y para el cuál no existen algoritmos integrados en el mercado como el 3417 que le puedan competir, esto se debe a lo complejo de los algoritmos para reducir la tasa de bits transmitidos, por lo que no se espera que en un futuro a mediano plazo se puedan diseñar estos integrados, siendo la

única alternativa el diseñarlos con microprocesadores como el TMS32010, en los cuales con solo cambiar el programa de la ROM se cambia el algoritmo utilizado, en el diseño de estos sistemas lo que implica dificultad es el diseño del software, el hardware no tiene mayores cambios entre algoritmos para procesar señales.

5.4 Conclusiones

El sistema ADPCM Robusto desarrollado en esta tesis, cumple con los objetivos que se trazaron al principio, calidad de la voz reconstruida transparente y ejecución del algoritmo en tiempo real. Este trabajo describe las diferentes etapas de diseño del ADPCM, desde su concepción inicial pasando por el desarrollo y depuración del sistema hasta sus pruebas en tiempo real.

En el futuro se espera diseñar un sistema para mezclar voz y datos en un mismo canal de transmisión, por esto, actualmente el departamento de Comunicaciones del IIE tiene en marcha un proyecto de un sistema de esta naturaleza a implementarse como accesorio de un multiplexor óptico diseñado en el mismo IIE.

Independientemente de los sistemas para mezclar voz y datos, como trabajo futuro de los codificadores de voz se prevee reducir el ancho de banda de la voz digitalizada a niveles de 9.6 Kb/s y menores. Para realizar esto se tendrá que trabajar mucho en el estudio del comportamiento de la voz en el habla Hispana, muy particularmente e independientemente de los trabajos realizados por los investigadores de habla Inglesa, con el fin de correlacionar los resultados con los obtenidos por ellos en su idioma.

De este modo se espera obtener niveles bajos de ancho de banda en los codificadores de voz, pero a diferencia de los sistemas actuales, con la característica adicional de que la voz a la salida del decodificador tenga "naturalidad" humana.

La aparición de los nuevos microprocesadores de punto flotante marcaran una nueva era en el procesamiento digital de señales. Los nuevos diseños de sistemas requeriran mucho menos tiempo y dolores de cabeza, los algoritmos se ejecutaran más rapido y ocuparán menos espacio de código, esto es debido a que en la programación de los nuevos microprocesadores de punto flotante a diferencia de los microprocesadores de punto fijo (como el TMS32010) no se tienen que realizar escalamientos, chequeo de ciclos y de sobreflujos.

En el caso de este trabajo, aunque sólo significa un pequeño paso en el mundo del procesamiento digital de señales, es de esperarse que en el futuro mexicano los sistemas basados en esta area se incrementen, al ser muy amplio el campo de aplicaciones prácticas.

APENDICE A
CONSIDERACIONES SOBRE EL CANAL
DE TRANSMISION

Es importante resaltar las limitaciones que presentan los canales de transmisión de las señales telefónicas, ya que las características de operación del canal influyen grandemente en la operación de los sistemas de compresión de ancho de banda de la voz. En particular veremos tres tipos de canales telefónicos que pueden ser representativos de los problemas que generan los canales de transmisión telefónicos en el desempeño de los sistemas de compresión de voz (en particular los VOCODERS).

1.- Canal telefónico convencional. Con frecuencia es referido con un ancho de banda de 4 Kc/s, sin embargo el canal proporciona mucho menos banda usable que los 4000 c/s. Las figuras (A.1) y (A.2) muestran la eficiencia de dos sistemas A.I.I.

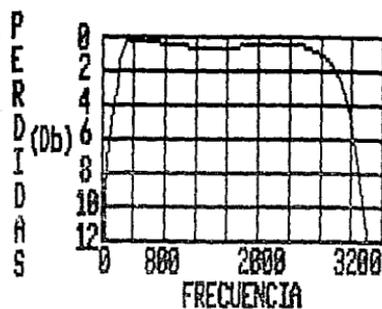


Fig A.1 Diagrama de un sistema típico pasa-banda.

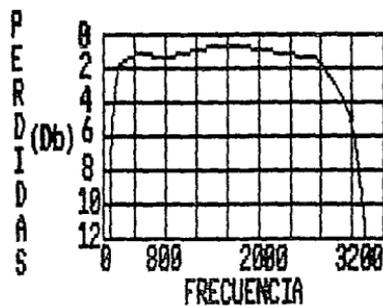


Fig A.2 Otro tipo sistema pasa-banda.

Observando notaremos que ninguno de los sistemas proporciona los 3 Kc/s entre los puntos con 3 dB de atenuación (ancho de banda), y solamente tenemos una respuesta plana entre los 400 c/s en el lado izquierdo y alrededor de 2500 c/s en el lado derecho. La figura (A.3) muestra una típica pérdida por inserción de un filtro del tipo empleado en los sistemas portadores.

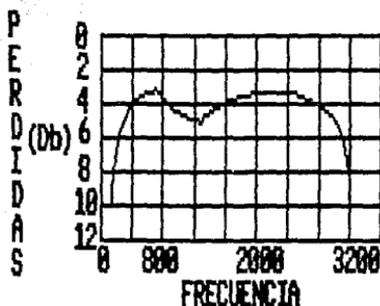


Fig A.3 Gráfica de pérdidas insertadas usando un sistema de filtros 561R de los laboratorios Bell.

Aquí el pasabanda es angosto y restrictivo, debe recordarse que los sistemas portadores en uso, fueron diseñados para acomodar señales analógicas donde las pequeñas variaciones de fase no dañan severamente la transmisión normal de voz. En las transmisiones de datos la distorsión por fase y retardo es indeseable, particularmente porque los modems de

datos emplean procesos que son sensibles a la fase como aspecto fundamental de su modo de operación.

En los canales telefónicos de larga distancia convencionales regularmente no se acomodan más de 2000 b/s, con tasas de error del orden de una parte en 10⁴, o mejores. Para las aplicaciones telefónicas, una condición de 0.1 a 0.2 por ciento de error puede ser considerado como el límite superior. Así que los vocoders de 2400 b/s, que usan este medio de transmisión, son considerados operando en el modo 'normal'.

2.- Canal telefónico degradado. En estas condiciones de canal, los circuitos de comunicación no soportarán la transmisión de un vocoder de 2400 b/s, pero pueden transportar una señal a 1200 b/s con una tasa de error menor al 5 por ciento. En este medio la banda de frecuencia es más estrecha, y tenemos una más grande distorsión por fase y por retardo, el ruido de impulso es alto, y los efectos de multi-trayectoria se presentan. Sobre estos circuitos, la calidad de y el reconocimiento de la señal codificada decrece, pero la inteligibilidad permanece al menos al 80 por ciento de lo mínimo aceptable. Los vocoders que se engranan a esta situación de transmisión se dice están en el modo de operación 'degradado'.

3.- Canales especiales de alta calidad. Esta categoría se refiere a los canales de ancho de banda extendidos, que soportan tasas de 4800, 7200, ó 9600 b/s. Estos canales tienen baja distorsión de fase, bajo nivel de ruido, buen acoplamiento por distorsión de retardo, y respuesta plana a través de la banda de paso. En algunas aplicaciones se utilizan canales gemelos, cada uno de los canales lleva una porción de la información. Los vocoders que operan sobre este tipo de canales se dice están en el modo de 'alta calidad'.

Canales de banda ancha. Los sistemas de voz digitalizada operando en canales de banda ancha sobre 30 a 50 Kh/s, se dice que están funcionando en modo de 'ultra-alta calidad'.

La presencia de ruido en la circuitería del canal de un vocoder causa una degradación severa de su calidad, particularmente en el rango de 70 a 300 Kc/s. Para el extractor del tono de voz, extraer el verdadero tono de la señal de voz solo es posible cuando la relación señal a ruido es muy buena.

Como se puede concluir, las características del canal influyen grandemente en el desempeño de los sistemas de compresión de voz. Por esto el sistema ADPCM robusto sólo puede funcionar correctamente en los canales de banda ancha.

APENDICE B
TABLA DE TRANSFORMADAS Z

TABLA DE TRANSFORMADAS Z

$X(s)$	$x(t)$ ó $x(k)$	$X(z)$
1	$\delta(t)$	1
e^{-kTs}	$\delta(t-kT)$	z^{-k}
$1/s$	$1(t)$	$z/(z-1)$
$1/s^2$	t	$Tz/(z-1)^2$
$1/(s+a)$	e^{-at}	$z/(z-e^{-aT})$
$a/s(s+a)$	$1-e^{-at}$	$(1-e^{-aT})z/(z-1)(z-e^{-aT})$
$1/(s+a)^2$	te^{-at}	$Tze^{-aT}/(z-e^{-aT})^2$
a^n	a^k	$z/(z-a)$
$a^n \cos kn$	$a^k \cos kn$	$z/(z+a)$

APENDICE C
ARITMETICA DE PUNTO FIJO
CON FORMATOS Q

Los cálculos del IMS32010 están basados en la representación de los números en complemento a dos con punto fijo. Cada número de 16 bits es evaluado con un bit de signo, i bits enteros y 15-i bits fraccionales. Entonces el número

00000010 10100000
 |-----punto decimal

tiene un valor de 2.625. Este número en particular se dice que está representado en formato D_8 (8 bits fraccionales). Su rango está entre -128 (1000000000000000) y 127.996 (0111111111111111). La precisión fraccional de un número en D_8 es de 0.004 (2^{-8}).

En algunas ocasiones es necesario manejar simultáneamente números con muy diversas magnitudes (algunos sólo enteros y otros únicamente fraccionales), entonces se debe de tener cuidado al sumarlos cuando no están representados en el mismo formato D .

Además, se debe de tomar en cuenta que el registro P del IMS32010 (donde queda el producto de las multiplicaciones) es de 32 bits y la memoria de datos es únicamente de 16 bits, así que, para guardar un producto en memoria se debe considerar el formato en que se desea guardar este para dar el

corrimiento de bits necesarios para llevar el producto a tal formato.

A continuación discutiremos algunas de las técnicas involucradas en el uso de este tipo de representación numérica.

a) Multiplicación.- Existe una gran variedad de situaciones que podemos encontrar cuando multiplicamos dos números.

primer caso.- fracción por fracción

Ejemplo 1: $0_{15} \times 0_{15}$

0 1000000000000000 = 0.5 en Q_{15}
A 0 1000000000000000 = 0.5 en Q_{15}
00 0100000000000000000000000000000000
|-----punto decimal

Para que el resultado se pueda almacenar en memoria en formato Q_{15} , es necesario recorrer un bit a la izquierda el producto obtenido y llevar a la memoria sólo los 16 bits más significativos (0 0100000000000000=0.25 en Q_{15}).

Para este caso el resultado deseado esta completamente en la mitad inferior del producto, entonces no es necesario ningún corrimiento al almacenar en memoria los 16 bits menos significativos del resultado. Debe notarse tambien de este ejemplo que toda la mitad superior del producto son bits redundantes del bit de signo.

Tercer caso... Enteros y fracciones

Ejemplo: $Q_{14} * Q_{14}$

01 10000000000000 = 1.5 en Q_{14}
 00 11000000000000 = 0.75 en Q_{14}
 0001 001000000000000000000000000000 = 1.125 en Q_{28}

↑----- punto decimal

Debe notarse que en este caso la magnitud máxima de un número en Q_{14} es muy próxima pero menor a 2, entonces la magnitud máxima del producto de dos números en Q_{14} es menor a 4. De lo anterior se puede obtener la siguiente regla.

El producto de un número con i bits enteros y f bits fraccionales, y un segundo número con j bits enteros y g bits fraccionales será un número con $(i+j)$ bits enteros y $(f+g)$

bits fraccionales. La mayor precisión posible para una representación de 16 bits de este número tendrá $(i+j)$ bits enteros y $(16-i-j)$ bits fraccionales.

b) Adición.- Como ya se vió, durante el proceso de multiplicación sólo es necesario ajustar el punto decimal después de la operación. Para la adición en cambio, se deben de considerar dos cosas, una es que ambos operandos tienen que estar representados en la misma notación Q , y la segunda consideración es que se debe manejar el formato Q conveniente para evitar sobreflujos en el resultado, o bien manipular en forma adecuada tales sobreflujos.

En la actualidad los nuevos microprocesadores de punto flotante ensayan el problema del sobreflujo, y no es necesario desperdiciar código para chequear si existe sobreflujo, haciendose más rápidos los programas de los algoritmos.

APENDICE D
PROCESADOR DIGITAL DE SEÑALES
TMS32010



El TMS32010 es el primer miembro de una nueva familia de procesadores de señales digitales (DSP) diseñados para las aplicaciones de alta velocidad. La familia TMS32010 contiene la primera microcomputadora MMS capaz de ejecutar 5 millones de instrucciones por segundo.

Para poder alcanzar esta velocidad el TMS32010 utiliza una modificación de arquitectura Harvard en la cual la memoria de datos y la memoria de programa se encuentran separadas, permitiendo así un traslape total de los ciclos de búsqueda y ejecución de cada instrucción. Además, en el TMS32010 se permite la transferencia de datos entre las memorias de programa y de datos.

El TMS32010 contiene los siguientes elementos.

Acumulador (ACC).- Acumulador de 32 bits.

Unidad aritmética lógica (ALU).- ALU con dos puertos de 32 bits.

Registros auxiliares (AR0 y AR1).- Dos registros auxiliares para direccionamiento indirecto de la memoria de datos y control del contador de ciclos. Los nueve bits menos significativos de cada registro son configurados como contadores bidireccionales.

Apuntador de registro auxiliar (ARP).- Registro de un sólo bit

que contiene la dirección del registro auxiliar en función.

Bus de datos (DBUS).- Bus de 16 bits conectado a la RAM de datos.

Apuntador de página de memoria de datos (DP).- Registro de un sólo bit que contiene la dirección de la página de RAM de datos (1 página=128 palabras).

RAM de datos.- 144 palabras de 16 bits de RAM para contener datos.

Registro de bandera de interrupción (INIF).- Registro de un sólo bit que indica cuando hay una solicitud de interrupción.

Multiplicador.- multiplicador hardware paralelo de 16x16 bits.

Registro de bandera de sobreflujo (OV).- Registro de un sólo bit que indica cuando ha ocurrido un sobreflujo en una operación aritmética.

Registro de modo sobreflujo (OVH).- Registro de un sólo bit que define el modo saturado ó no saturado en operaciones aritméticas.

Registro P (P).- Registro de 32 bits que contiene el producto de las multiplicaciones.

Bus de programa (PBUS).- Bus de 16 bits conectado a la memoria del programa.

Contador del programa (PC).- Registro de 12 bits que contiene la dirección siguiente de memoria de programa.

ROM del programa.- Sólo existe en la versión TMS320H10, 1536 palabras de 16 bits que contienen el código de programa.

Registros de corrimiento.- Uno es un registro de corrimiento a la izquierda variable de 0-15 bits al mover datos de memoria a la ALU. El otro registro de corrimiento actúa en el acumulador cuando se almacenan datos a la RAM de datos; este registro puede tener corrimientos a la izquierda de 0,1,4 bits.

Stack.- Registro de 4 por 12 bits para salvar el contenido del contador del programa en subrutinas ó llamadas de interrupción.

Registro T (T).- Registro de 16 bits que contiene el multiplicando durante las operaciones de multiplicación.

Para el caso del TMS32010, el cual no contiene memoria de programa interna, es posible direccionar hasta 4096 palabras de memoria de programa externa. Como el ciclo de máquina del TMS32010 es de 200 ns, la memoria de programa externa debe tener un tiempo de acceso máximo de 100 ns.

INSTRUCCIONES DEL ACUMULADOR

ABREVIADO	DESCRIPCION	No. CICLOS	No. PALABRAS	CODIGO DE OPERACION	
				15	0
ABS	VALOR ABSOLUTO AL ACUMULADOR	1	1	0 1 1 1 1 1 1 1 0 0 0 1 0 0 0	1
ADD	SUMA ACUMULADOR CON CORRIMIENTO	1	1	0 0 0 0 1 - S - 1 - - - -	D - - - -
ADDA	SUMA PARTE ALTA DEL ACUMULADOR	1	1	0 1 1 0 0 0 0 0 1 - - - -	D - - - -
ADDS	SUMA ACUMULADOR SIN EXTENSION DEL SIGNO	1	1	0 1 1 0 0 0 0 1 1 - - - -	D - - - -
AND	AND CON EL ACUMULADOR	1	1	0 1 1 1 1 0 0 1 1 - - - -	D - - - -
LAC	CARGA ACUMULADOR CON CORRIMIENTO	1	1	0 0 1 0 1 - S - 1 - - - -	D - - - -
LAD	CARGA ACUMULADOR CON DIRECCIONAMIENTO INMEDIATO	1	1	0 1 1 1 1 1 1 0 1 - - - -	K - - - -
OR	OR CON EL ACUMULADOR	1	1	0 1 1 1 1 0 1 0 1 - - - -	D - - - -
SACH	ALMACENA 16 BITS SUPERIORES DEL ACUMULADOR CON CORRIMIENTO	1	1	0 1 0 1 1 1 - X - 1 - - - -	D - - - -
SACL	ALMACENA 16 BITS INFERIORES DEL ACUMULADOR	1	1	0 1 0 1 0 0 0 0 1 - - - -	D - - - -
SUB	RESTA DEL ACUMULADOR CON CORRIMIENTO	1	1	0 0 0 1 1 - S - 1 - - - -	D - - - -
SUBC	RESTA CONDICIONALMENTE	1	1	0 1 1 0 0 1 0 0 1 - - - -	D - - - -
SUBH	RESTA DE LOS 16 BITS SUPERIORES DEL ACUMULADOR	1	1	0 1 1 0 0 0 1 0 1 - - - -	D - - - -
SUBS	RESTA DEL ACUMULADOR SIN EXTENSION DEL SIGNO	1	1	0 1 1 0 0 0 1 1 1 - - - -	D - - - -
XOR	OR EXCLUSIVA CON EL ACUMULADOR	1	1	0 1 1 1 1 0 0 0 1 - - - -	D - - - -
ZAC	CERO AL ACUMULADOR	1	1	0 1 1 1 1 1 1 1 1 0 0 0 1 0 0 1	1
ZALH	CERO AL ACUMULADOR Y CARGA LOS 16 BITS SUPERIORES	1	1	0 1 1 0 0 1 0 1 1 - - - -	D - - - -
ZALS	CERO AL ACUMULADOR Y CARGA LOS 16 BITS INFERIORES SIN EXTENSION DE SIGNO	1	1	0 1 1 0 0 1 1 0 1 1 - - - -	D - - - -

INSTRUCCIONES DE REGISTROS AUXILIARES Y APUNTADEOR DE PAGINA DE DATOS

MNEEMONICO	DESCRIPCION	No. CICLOS	No. PALABRAS	CODIGO DE OPERACION	
				15	0
LAR	CARGA REGISTRO AUXILIAR	1	1	0 0 1 1 1 0 0 R1	D
LARK	CARGA REGISTRO AUXILIAR CON DIRECCIONAMIENTO	1	1	0 1 1 1 0 0 0 R1	K
LAKP	CARGA APUNTADEOR DE RE- GISTRO AUXILIAR CON DI- RECCIONAMIENTO INMEDIATO	1	1	0 1 1 0 1 0 0 0 1 0 0 0 0 0 0 K	
LDP	CARGA APUNTADEOR DE PÁGI- NA DE MEMORIA DE DATOS	1	1	0 1 1 0 1 1 1 1 1 1	D
LDPK	CARGA APUNTADEOR DE PÁGI- NA DE MEMORIA DE DATOS CON DIRECCIONAMIENTO	1	1	0 1 1 0 1 1 1 0 0 0 0 0 0 0 0 K	
MAR	MODIFICA REGISTRO AUXI- LIAR	1	1	0 1 1 0 1 0 0 0 1	D
SAR	ALMACENA REGISTRO AUXI- LIAR	1	1	0 0 1 1 0 0 0 R1	D

OPERACIONES DE ENTRADA/SALIDA Y MEMORIA DE DATOS

MNEEMONICO	DESCRIPCION	No. CICLOS	No. PALABRAS	CODIGO DE OPERACION	
				15	0
DMOV	COPIA CONTENIDO DE UNA LOCALIDAD DE MEMORIA EN LA SIGUIENTE LOCALIDAD	1	1	0 1 1 0 1 0 0 1 1	D
IN	ENTRADA DE DATO DESDE UN PUNTO	2	1	0 1 0 0 0 1-PA	B
OUT	SALIDA DE DATO POR UN PUNTO	2	1	0 1 0 0 1 1-PA	B
TBLR	LEE DATO DE MEMORIA DE PROGRAMA A MEMORIA DE DATOS	3	1	0 1 1 0 0 1 1 1 1	D
TBLW	ESCRIBE DE LA MEMORIA DE DATOS A LA MEMORIA DE PROGRAMA	3	1	0 1 1 1 1 0 1 1 1	D

INSTRUCCIONES DE CONTROL

MNEMÓNICO	DESCRIPCIÓN	No. CICLOS	No. PALABRAS	CÓDIGO DE OPERACIÓN	
				15	0
DINT	DESABILITA INTERRUPTIO- NES	1	1	0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 1	
EINT	HABILITA INTERRUPTIONES	1	1	0 1 1 1 1 1 1 1 1 0 0 0 0 0 1 0	
LST	CARGA REGISTRO DE ESTADOS	1	1	0 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1	D
NOP	NO OPERACION	1	1	0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0	
POP	POP STACK AL ACUMULADOR	1	1	0 1 1 1 1 1 1 1 1 0 0 1 1 1 0 1	
PUSH	PUSH STACK DEL ACUMULA- DOR	1	1	0 1 1 1 1 1 1 1 1 0 0 1 1 1 0 0	
ROVM	QUIJA EL MODO SOBREFLUJO	1	1	0 1 1 1 1 1 1 1 1 0 0 0 1 0 1 0	
SOVM	PONE MODO SOBREFLUJO	1	1	0 1 1 1 1 1 1 1 1 0 0 0 1 0 1 1	
SST	ALMACENA REGISTRO DE ESTADOS	1	1	0 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1	D

INSTRUCCIONES DE REGISTROS P, T Y MULTIPLICACION

MNEMÓNICO	DESCRIPCIÓN	No. CICLOS	No. PALABRAS	CÓDIGO DE OPERACIÓN	
				15	0
APAC	SUMA P AL ACUMULADOR	1	1	0 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1	
LT	CARGA REGISTRO T	1	1	0 1 1 0 1 0 1 0 1 1 1 1 1 1 1 1	D
LTA	COMBINA LAS INSTRUCCIO- NES LT Y APAC	1	1	0 1 1 0 1 1 0 0 1 1 1 1 1 1 1 1	D
LTD	COMBINA LAS INSTRUCCIO- NES LT, APAC Y DMOV	1	1	0 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1	D
MPY	MULTIPLICA CON REGISTRO T Y PRODUCTO HACIA P	1	1	1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1	D
MPYK	MULTIPLICA REGISTRO T CON OPERANDO INMEDIATO	1	1	1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1	K
PAC	CARGA REGISTRO P EN EL ACUMULADOR	1	1	0 1 1 1 1 1 1 1 1 0 0 0 1 1 1 0	
SPAC	RESTA P DEL ACUMULADOR	1	1	0 1 1 1 1 1 1 1 1 0 0 1 0 0 0 0	

INSTRUCCIONES DE SALTO

MNEMOTICO	DESCRIPCION	No. CICLOS	No. PALABRAS	CODIGO DE OPERACION	
				15	0
B	SALTO INCONDICIONAL	2	2	1111100100000000	00001-DIRECCION DEL SALTO-1
BANZ	SALTA SI REGISTRO AUXILIAR NO ES CERO	2	2	1111010000000000	00001-DIRECCION DEL SALTO-1
BGFZ	SALTA SI ACUMULADOR ES MAYOR O IGUAL A CERO	2	2	1111100000000000	00001-DIRECCION DEL SALTO-1
BGZ	SALTA SI ACUMULADOR ES MAYOR QUE CERO	2	2	1111100000000000	00001-DIRECCION DEL SALTO-1
BIDZ	SALTA SI LA PUNTERA ES IGUAL A CERO	2	2	1111011000000000	00001-DIRECCION DEL SALTO-1
BLEZ	SALTA SI ACUMULADOR ES MENOR O IGUAL A CERO	2	2	1111101100000000	00001-DIRECCION DEL SALTO-1
BLZ	SALTA SI ACUMULADOR ES MENOR QUE CERO	2	2	1111101000000000	00001-DIRECCION DEL SALTO-1
BWZ	SALTA SI ACUMULADOR ES DIFERENTE DE CERO	2	2	1111110000000000	00001-DIRECCION DEL SALTO-1
BV	SALTA SI HAY SOBREFLUJO	2	2	1110101000000000	00001-DIRECCION DEL SALTO-1
BZ	SALTA SI EL ACUMULADOR ES IGUAL A CERO	2	2	1111111000000000	00001-DIRECCION DEL SALTO-1
CALA	LLAMA SUBROUTINA DESDE EL ACUMULADOR	2	1	0111111100011000	1111100000000000
CALL	LLAMA SUBROUTINA CON DIRECCIONAMIENTO INMEDIATO	2	2	1111100000000000	00001-DIRECCION DEL SALTO-1
RET	REGRESA DE SUBROUTINA O DE RUTINA DE INTERRUPCION	2	1	0111111100011010	

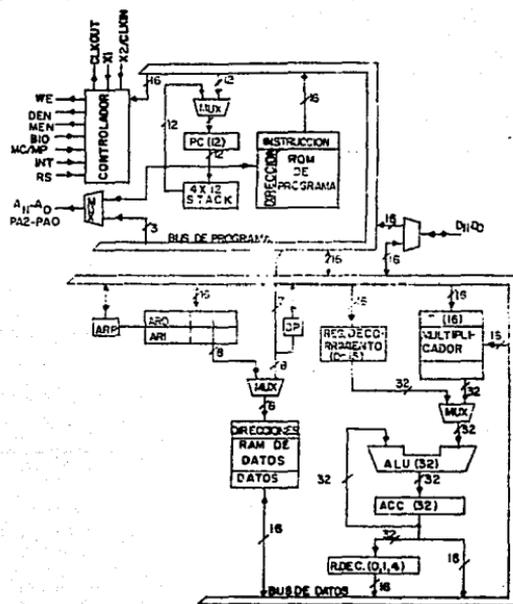


Fig D.1 Configuración interna del TMS32010

GLOSARIO

- Cuantización** Proceso de dividir el rango de valores de una señal en un número finito de subdivisiones
- Cepstrum** Señal obtenida al tomar la transformada de Fourier del logaritmo del espectro de potencia de la señal de voz
- Cuasiestacionaria (señal)** dice de la señal que varía lentamente con el tiempo
- Decimación** Técnica digital para reducir la tasa de muestreo
- DFT** Transformada discreta de Fourier
- FFT** Transformada rápida de Fourier
- DSP** Procesador digital de señales
- Espectrograma** Modelo cartesiano donde el eje de las abscisas corresponde al tiempo y el eje de las ordenadas a la frecuencia
- Fetch** Porción particular del ciclo de trabajo de un microprocesador en la que la localización de la siguiente instrucción es ejecutada
- Fonética** Rama de la Gramática que estudia los sonidos de voz
- Formantes** Frecuencias resonantes del tracto vocal
- Fonemas** Cada uno de los sonidos simples del lenguaje hablado
- Glotal** Referido a la Glotis
- Interpolación** Técnica digital para aumentar la tasa de muestreo
- Linguística** Ciencia general del lenguaje
- Muestreo** Proceso de obtener y grabar el valor de una señal en intervalos periódicos o aleatorios
- Neuronales** Referido a las neuronas del cerebro
- Página** Subdivisión de la sección de memoria
- Radiación** Emisión de ondas acústicas
- Rango dinámico** Cociente de la máxima a la mínima amplitud de una señal
- Ruido Granular** Ruido generado en un modulador diferencial por efecto de una señal con poca pendiente
- Ruido de sobrecarga** Ruido generado en un modulador diferencial por efecto de una señal de gran pendiente
- Sonido Fricativo (Sordo)** Sonidos cuya articulación resulta de cierta fricción o roce del aire en los órganos bucales
- Sonido Sonoro** Sonidos que durante su pronunciación van acompañados de una vibración de las cuerdas vocales
- Tono** Frecuencia fundamental del sonido
- Tracto vocal** Sistema bucal que consiste de la cavidad bucal y de la laringe

BIBLIOGRAFIA

- 1.- J.O. Smith y J.B. Allen
"Variable bandwidth Adaptive Delta Modulation"
The Bell System Technical Journal
Vol. 60, No. 5, Mayo-Junio de 1981
- 2.- C.M. Kortman
"Data Compression by Redundancy Reduction"
IEEE Spectrum, Marzo de 1967
- 3.- Jerry D. Gibson
"Adaptive Prediction for Speech Encoding"
IEEE ASSP Magazine, Julio de 1984
- 4.- R.E. Crochiere, R.V. Cox y J.D. Johnston
"Real Time Speech Coding"
IEEE Transactions on Communications
Vol. COM-30, No.4, Abril de 1982
- 5.- David Kahn
"Cryptology and the Origins of Spread Spectrum"
IEEE Spectrum, Septiembre de 1984
- 6.- Andres Buzo de la Peña y Federico Kuhlman
"Codificación de Señales de Voz"
Mexico 1983
- 7.- G.A. Frantz y R.H. Wiggins
"Design Case History: Speak & Spell Learns to Walk"
IEEE Spectrum, Febrero de 1982
- 8.- G.R. Doddington y T.B. Schalk
"Speech Recognition Turning Theory to Practice"
IEEE Spectrum
- 9.- H. Jeong y CH. Kwan Un
"A PCM/ADM y ADM/PCM Code Converter"
IEEE 1979 Acoustic, Speech & Signal Processing
- 10.- J.E. Abate
"Linear and Adaptive Delta Modulation"
Proceedings of the IEEE
Vol. 55, No. 3, Marzo de 1967

- 11.-V. Ramamorthy y N.S. Jayant
"Enhancement of ADPCM Speech by Adaptive Postfiltering"
AT&T Bell Laboratories Technical Journal
Vol. 63, No. 8, Octubre de 1984
- 12.-M.R. Buric, J. Kohut y J.P. Olive
"Speech Synthesis"
The Bell System Technical Journal
Vol. 60, No. 7, Septiembre de 1981
- 13.-Don H. Johnson
"Signal Processing Software Tools"
IEEE 1984 Acoustic, Speech & Signal Processing
- 14.- J.D. Gibson, S.K. Jones y J.L. Melsa
"Sequentially Adaptive Prediction and Coding of Speech Signals"
IEEE 1974 Acoustic, Speech & Signal Processing
- 15.-J.R. Boddie
"Digital Signal Processor"
The Bell System Technical Journal
Vol. 60, Septiembre de 1981
- 16.-B.H. Juang
"On Using the Itakura-Saito Measures for Speech
Coders Performance Evaluation"
AT&T Bell Laboratories Technical Journal
Vol. 63, No. 8, Octubre de 1984
- 17.-Ing. Sofía N. Galicia Haro
"Empleo de Canales de Comunicación en el Sistema
Eléctrico de la Compañía de Luz y Fuerza del Centro"
- 18.-Z.L. Budrikis y M. Hatamian
"Moment Calculations by Digital Filters"
AT&T Bell Laboratories Technical Journal
Vol. 63, No. 2, Febrero de 1984
- 19.- V. Ramamoorthy
"A Novel Speech Coder for Medium and High Bit Rate
Applications Using Modulo-PCM Principles"
IEEE Transactions on Acoustic, Speech and Signal Processing
Vol. ASSP-33, No. 2, Abril de 1985
- 20.-G. Mirchandani y R.C. Gauss
"Performance Characteristics of a Hardware Implementation
of the Cross-Talk Resistant Adaptive Noise Canceller"
IEEE 1986 Acoustic, Speech & Signal Processing

- 21.-De-Yuan Cheng y A. Gersho
 "A Fast Codebook Search Algorithm for NearestNeighbor
 Pattern Matching"
 IEEE 1986 Acoustic, Speech & Signal Processing
- 22.- J. Makhoul y M. Berouti
 "Adaptive Noise Spectral Shaping and Entropy Coding in
 Predictive Coding of Speech"
 IEEE 1979 Acoustic, Speech & Signal Processing
- 23.- Nuggehalli S. Jayant
 "Digital Coding of Speech Waveforms: PCM, DPCM, and
 DM Quantizers"
 Proceedings of the IEEE, Mayo de 1974
- 24.- A.W. Holk y W.W. Anderson
 "A Single-Processor LPC Vocoder"
 IEEE Acoustic, Speech & Signal Processing
- 25.-S. Cordray, D.L. Favin y D.P. Yorkgitis
 "Power Measurements"
 The Bell System Journal
 Vol. 60, No. 7, Septiembre de 1981
- 26.- Paul Wallich
 "Putting Speech Recognizers to Work"
 IEEE Spectrum, Abril de 1987
- 27.- D.J. Goodman y R.M. Wilkinson
 "A Robust Adaptive Quantizer"
 IEEE Transactions on Communications, Noviembre de 1975
- 28.- Gabriel Flores S.
 "Análisis de los Parametros de Sincronización de una
 Red Digital Telefónica"
 Mexico, 1983
- 29.- Alejandro Solorio y Enrique Viguri
 "Desarrollo de la Red Telefónica en México y su
 Futuro Inmediato"
 Mexico, 1983
- 30.- M. Dankberg, R. Iltis, D. Saxton y P. Wilson
 "Implementation of the Relp Vocoder Using the TMS32010"
 IEEE 1984 Acoustic, Speech & Signal Processing
- 31.-B.Bryden y H. Hassanein
 "Implementation of a Full Duplex 2.4 Kbps LPC Vocoder
 on a Single TMS-320 Microprocessor Chip"
 IEEE 1984 Acoustic, Speech & Signal Processing

- 32.- L.R. Morris
 "Structural Considerations for Large FFT Programs on the TI TMS32010 DSP Microchip"
 IEEE 1985 Acoustic, Speech & Signal Processing
- 33.- T.K. Miller y S.T. Alexander
 "An Implementation of the LMS Adaptive Filter Using an SIMD Multiprocessor Ring Architecture"
 IEEE 1985 Acoustic, Speech & Signal Processing
- 34.- C.R. Cole, A. Haoui y P.L. Winship
 "A High-Performance Digital Voice Echo Canceller on a Single TMS32020"
 IEEE 1986 Acoustic, Speech & Signal Processing
- 35.-D.P. Morgan y H.F. Silverman
 "An Investigation into the Efficiency of a Parallel Tms320 Architecture: DFT and Speech Filterbank Applications"
 IEEE 1986 Acoustic, Speech & Signal Processing
- 36.- W. Gass, R. Tarrand y G. Doddington
 "A Parallel Signal Processor System"
 IEEE 1986 Acoustic, Speech & Signal Processing
- 37.- C. Erskine
 "Architecture et Applications d'un Processeur de Traitement numérique du signal de la Seconde génération: le TMS32020"
 Traitement du Signal, Volume 2-No. 1, 1985
- 38.- Ch. Rogers, K-S. Min, S. Speier y J. Whitson
 "A Transportable TMS32010 Signal Processing System"
 IEEE 1985 Acoustic, Speech & Signal Processing
- 39.- Steve Sweitzer
 "A Low Cost FFT Chip Set"
 IEEE 1984 Acoustic, Speech & Signal Processing
- 40.- D.Y. Wong, D.A. Russo, C.D. Bergman y C.H. Lee
 "Signal Processing Software for a Voice Messaging System"
 IEEE 1985 Acoustic, Speech & Signal processing
- 41.- R.D. Hangartner y V.K. Jain
 "32 Kbs ADPCM/PCM Transcoder Using TI-320 DSP Microprocessor"
 IEEE 1985 Acoustic, Speech & Signal Processing
- 42.- Miguel Ríos
 "Integración de Señales de Voz y Datos en una Red Local Tipo Bus"
 Mexicon

- 43.-R.V. Cox
"A Comparison of Three Speech Coders to be Implemented on the Digital Signal Processor"
The Bell System Technical Journal
Vol. 60, No. 7, Septiembre de 1981
- 44.- Roger C. Wood
"On Optimun Quantization"
IEEE Transactions on Information Theory, Marzo de 1969
- 45.- D. Mitra y B. Gotz
"An Adaptive PCM System Designed for Noisy Channels and Digital Implementations"
The Bell System Technical Journal
Vol. 57, No. 7, Septiembre de 1978
- 46.- C.A. McGonegal, D.A. Berkley y N.S. Jayant
"Digital Signal Processor: Private Communications"
The Bell System Technical Journal
Vol. 60, No. 7, Septiembre de 1981
- 47.- D.J. Goodman y A. Gersho
"Theory of an Adaptive Quantizer"
IEEE Transactions on Communications, Agosto de 1974
- 48.- Zvi Orvach
"New Types of Adaptive Quantizers"
IEEE Transactions on Acoustic, Speech & Signal Processing
Vol. ASSP-32, No. 5, Octubre de 1984
- 49.- Debasis Mitra
"An Almost Linear Relation Between the Step Size Behavior and the Input Signal Intensity in Robust Adaptive Quantization"
IEEE 1979 Acoustic, Speech & Signal Processing
- 50.- S.R. Boullosa y S.P. Ruíz
"Espectro Promedio de Larga Duración del Idioma Español"
Mexico 1983
- 51.- R.E. Contolini, Ch.J. Mason y G.A. Sterling
"Multiple DSPs Provide Speed for Digital Servo Control"
Computer Design, Septiembre de 1987
- 52.- Special Report
"Squeeze Play: Carriers Quietly Tryings Compressed Voice, Users Could Soon Squeeze big Savings"
Data Communications, Julio de 1987
- 53.- D.Ll. Rice
"Friends, Humans, and Country Robots: Lend me Your Ears"
Telecommunications, Septiembre de 1986

- 54.- R.W. Shafer y J.D. Markel
 "Speech Analysis"
 IEEE PRESS, 1978 Editorial Board
- 55.- J.L. Flanagan, M. Schueder, N.S. Jayant y J.M. Tribolet
 "Speech Coding"
 IEEE Transactions on Communications
 Vol. COM-27, No. 4, Abril de 1979
- 56.- N. S. Jayant
 "Coding Speech at Low Bit Rates"
 IEEE Spectrum Agosto de 1986
- 57.- J.W. Bayless y S.J. Campanella
 "Voice Signals: bit-by-bit"
 IEEE Spectrum
- 58.- P.G. Edwards y John Clapper, Jr.
 "Better Vocoders are Coming"
 IEEE Spectrum, Septiembre de 1964
- 59.- C. Scagliola
 "Evaluation of Adaptive Speech Coders Under Noisy
 Channel Conditions"
 The Bell System Technical Journal
 Julio-Agosto de 1979
- 60.- P. Cumiskey, N.S. Jayant y J.L. Flanagan
 "Adaptive Quantization in Differential PCM
 Coding of the Speech"
 The Bell System Technical Journal
 Vol. 52, No. 7, Septiembre de 1973
- 61.- N.S. Jayant
 "Adaptive Quantization With a One-Word Memory"
 The Bell System Technical Journal
 Vol. 52, No. 7, Septiembre de 1973
- 62.- Arturo Veloz Guerrero
 "Un Sistema de Compresión de Voz Continua
 Implantado Sobre Microprocesador"
 Mexico 1986
- 63.- V. Marianov y R. Ramírez
 "Un Algoritmo Mejorado de Modulación Delta Para
 Digitalización de Voz a Tasas Bajas de Muestreo"
 Mexico 1986
- 64.- L.R. Rabiner y R.W. Schafer
 "Digital Processing of Speech Signals"
 Prentice-Hall Signal Processing Series, 1978

- 65.- Alan V. Oppenheim
"Applications of Digital Signal Processing"
Editorial Prentice-Hall, 1978
- 66.- Mischa Schwartz
"Transformación de Información Modulación y Ruido"
Editorial McGraw-Hill, 1983
- 67.- Fred J. Taylor
"Digital Filter Design Handbook"
Editorial Marcel Dekker Inc., 1983
- 68.- B.P. Lathi
"Introducción a la Teoría y Sistemas de Comunicación"
Editorial Limusa S.A., 1978