UNIVERSIDAD AUTONOMA DE NUEVO LEON

FACULTAD DE INGENIERIA MECANICA
Y ELECTRICA

DIVISION DE ESTUDIOS DE POST-GRADO

COMPRESION DE VOZ PARA SU TRANSMISION EN REDES DE DATOS

POR

LAURA ESPINOSA CAMACHO

TESIS

EN OPCION AL GRADO DE MAESTRO EN
CIENCIAS DE LA INGENIERIA
CON ESPECIALIDAD EN TELECOMUNICACIONES

SAN NICOLAS DE LOS GARZA, N. L. SEPTIEMBRE DE 2003

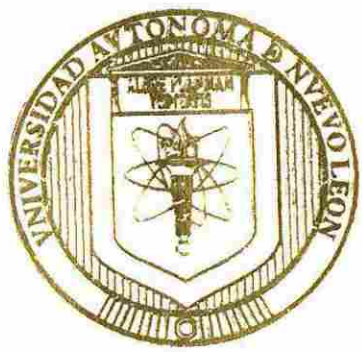
COMPRESION DE VOZ PARA SU TRANSMISION EN REDES DE DATOS

L.E.C.

UNIVERSIDAD AUTONOMA DE NUEVO LEON

FACULTAD DE INGENIERIA MECANICA
Y ELECTRICA

DIVISION DE ESTUDIOS DE POST-GRADO



COMPRESION DE VOZ PARA SU TRANSMISION EN REDES DE DATOS

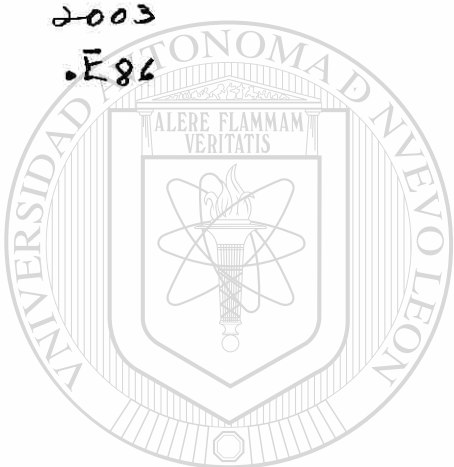
POR

LAURA ESPINOSA CAMACHO

TESIS

EN OPCION AL GRADO DE MAESTRO EN
CIENCIAS DE LA INGENIERIA
CON ESPECIALIDAD EN TELECOMUNICACIONES

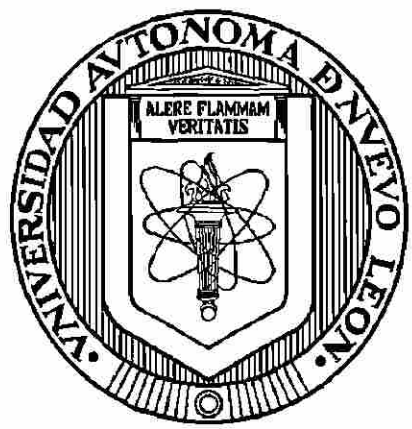
SAN NICOLAS DE LOS GARZA, N. L. SEPTIEMBRE DE 2003



UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE INGENIERÍA MECÁNICA Y ELÉCTRICA

DIVISIÓN DE ESTUDIOS DE POST-GRADO



COMPRESIÓN DE VOZ PARA SU TRANSMISIÓN EN REDES DE DATOS

POR

LAURA ESPINOSA CAMACHO

TESIS

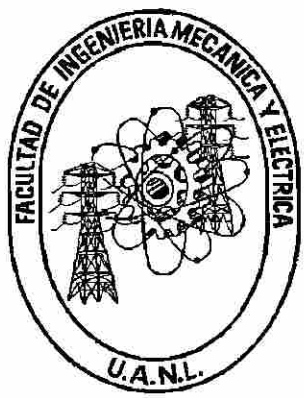
EN OPCIÓN AL GRADO DE MAESTRO EN CIENCIAS DE LA INGENIERÍA CON ESPECIALIDAD EN TELECOMUNICACIONES

SAN NICOLÁS DE LOS GARZA, N.L. SEPTIEMBRE DEL 2003

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE INGENIERÍA MECÁNICA Y ELÉCTRICA

DIVISIÓN DE ESTUDIOS DE POST-GRADO



COMPRESIÓN DE VOZ PARA SU TRANSMISIÓN EN REDES DE DATOS

POR

LAURA ESPINOSA CAMACHO

TESIS

EN OPCIÓN AL GRADO DE MAESTRO EN CIENCIAS DE LA INGENIERÍA CON ESPECIALIDAD EN TELECOMUNICACIONES

SAN NICOLÁS DE LOS GARZA, N.L. SEPTIEMBRE DEL 2003

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN
FACULTAD DE INGENIERÍA MECÁNICA Y ELÉCTRICA
DIVISIÓN DE ESTUDIOS DE POST-GRADO

Los miembros del comité de tesis recomendamos que la tesis **"Compresión de voz para su transmisión en redes de datos"**, realizada por la alumna Ing. Laura Espinosa Camacho, matrícula 775431 sea aceptada para su defensa como opción al grado de Maestro en Ciencias de la Ingeniería con especialidad en Telecomunicaciones.

El Comité de Tesis

Dr. José Antonio de la O Serna
Asesor

Dr. Marco Tulio Mata Jiménez
Coasesor

M.C. Raúl Alvarado Escamilla
Coasesor

Vo. Bo.
Dr. Guadalupe Alan Castillo Rodríguez
División de Estudios de Post-grado

Septiembre del 2003

DEDICATORIA

A Dios por permitir vivir este momento.

A mis Padres y mis hermanos por todo el apoyo brindado.
Mis sobrinos: David, Francisco Javier y próximo por nacer.

AGRADECIMIENTOS

Agradezco a la Universidad Autónoma de Nuevo León por el apoyo brindado en mis estudios de postgrado, así mismo agradezco a la Facultad de Ingeniería Mecánica y Eléctrica.

Mi gratitud al Dr. José Antonio de la O Serna por la asesoría que me brindo en este trabajo, la confianza y motivación que me dio para el desarrollo del mismo.

A los profesores sinodales: Dr. Marco Tulio Mata Jiménez y MC. Raúl Alvarado Escamilla.

A los profesores del Plan Doctoral, sin dejar de mencionar el apoyo que me brindaron el Dr. Ernesto Vázquez y Dr. Efraín Alcorta.

A todos los compañeros del Programa Doctoral por los grandes momentos que pasamos juntos.

Mis amigas: Ángeles Carrera, Irma Rosario Valadez, Mónica Saenz, Patricia Catalina Muñoz Báez y Verónica Agustin.

Resumen

Compresión de voz para su transmisión en redes de datos

Publicación No. _____

Laura Espinosa Camacho

Universidad Autónoma de Nuevo León

Facultad de Ingeniería Mecánica y Eléctrica

Profesor Asesor: Dr. José Antonio de la O Serna

Septiembre, 2003

El presente trabajo muestra diferentes algoritmos de compresión de voz siendo la codificación o compresión de voz un campo que interesa en obtener una representación digital compacta de señales, es decir, reducir la máximo la cantidad de información de transmisión.

Varios de los algoritmos que se mencionarán durante esta tesis han sido adoptadas en la industria de las telecomunicaciones, como por ejemplo: telefonía celular.

Así mismo los métodos de codificación que se discutirán están enfocadas a las comunicaciones digitales de la voz en telefonía.

Se describirá y se analizará un algoritmo de codificación de voz, LPC siendo este último una comparación con el modelo de producción de voz, donde la voz es modelada por pulsos cuasi periódicos o bien ruido aleatorio.

Este algoritmo nos proporciona robustez y precisión en cuanto al cálculo de los parámetros que caracterizan al sistema. Incluiremos la obtención del espectro de una señal de voz, el cálculo de error de predicción sobre un segmento de una señal además de mostrar los métodos para encontrar la solución a un sistema de ecuaciones lineales y finalmente la reconstrucción de una señal de voz. Podemos decir que el predictor lineal permite la representación de 100-200 muestras de una señal para obtener un promedio de 10-15 coeficientes. Estos coeficientes pueden ser usados para calcular la respuesta de tracto bucal.

Contenido

Lista de figuras

Lista de tablas

Capítulo 1

Introducción

El campo de la digitalización de la voz ha sido y continúa siendo un área de interés. En esta investigación se presentan varios algoritmos de compresión de voz que van a depender de su calidad de funcionamiento.

Desde los ochenta se han adoptado algoritmos de codificación de voz a tasa bajas de bits aplicados a las redes de telefonía celular. Así mismo se establecieron estándares de telefonía celular digital. En Norte América (TIA IS-54) se utilizan códigos híbridos a una tasa de 8Kbps. En Europa (GMS) se utiliza 13Kbits/seg, el algoritmo de excitación de pulso regular y el JDC en Japón. Los estándares empleados en comunicaciones militares han sido utilizados a una tasa de bits de 4.8Kbis/s, a 6.4Kbits/s a si mismo el código de excitación multibanda ha sido empleado en el Sistema Satelital Marítimo Internacional (Sistema INMARSAT-M) y el sistema Satelital Australiano (Sistema AUSSAT). Finalmente hay proyectos para incrementar la capacidad de la red celular introduciendo algoritmos a una tasa media de bits en el GSM en Europa, el JDC en Japón y el TIA IS-54, estándar utilizado en Norte América.

Los algoritmos de compresión de voz se dividen en dos categorías: aquellos que codifican la forma de onda de voz lo más exacto posible (codificadores de onda) y aquellos que procesan la forma de onda para codificar los aspectos perceptuales significativos del proceso de habla y escucha (Codificadores Paramétricos), generalmente se llegan a transmitir algunos parámetros como por ejemplo: coeficientes de filtro, valores de

ganancia, índices de tablas, los cuales se calculan en el codificador y se transmiten al decodificador. El receptor recibe estos parámetros para reconstruir la voz.

El algoritmo de codificación de voz se evalúa basándose en los siguientes parámetros: una tasa de bits, la calidad de voz reconstruida (codificación), la complejidad del algoritmo, el retraso introducido y la robustez de los algoritmos para errores de canal e interferencia acústica. Por ejemplo, la implementación en tiempo real de algoritmos a una tasa baja de bits se implementará en un procesador digital de señales, capaz de ejecutar 12 o más millones de instrucciones por segundo (MIPS). El retraso (la codificación más la decodificación) que se introduce en algoritmos es generalmente de 50 a 60 ms. En los sistemas de codificación de voz robustos se incorporan algoritmos de corrección de error para proteger la información perceptual en canales con ruido.

1.1 Definición del problema

El objetivo de este proyecto es desarrollar algunos codificadores de voz que sean capaces de generar la voz de alta calidad a tasas bajas de bits. Muchos de esos codificadores incorporan mecanismos para representar las propiedades espectrales de la voz. Varios de esos algoritmos han sido adoptados en telefonía celular.

La codificación de voz o compresión de voz es un campo que se interesa en obtener una representación digital compacta de señales de voz, es decir, reducir al máximo la cantidad de información transmitida. Al mismo tiempo se desea conservar la calidad de la voz reproducida y la complejidad del codificador-decodificador. La codificación de voz involucra el muestreo en el tiempo y la cuantificación en amplitud como procesos previos a la codificación.

Los métodos de codificación que se discuten en esta tesis están enfocados a comunicaciones digitales de la voz en telefonía. Un canal telefónico tiene 4Khz y se emplea un ancho de banda de voz de 3.4Khz, la cual se muestrea a 8Khz. Además en esta tesis nos referimos a una tasa binaria media (8-16Kbits/s) baja (8-2.4Kbits/s) y muy baja (<2.4Kbits/s).

La codificación de voz a tasas medias de bits se lleva a cabo usando un proceso de análisis-síntesis. En la etapa de análisis, la voz se representa en forma compacta de parámetros, la cual se codifica correctamente, en la etapa de síntesis esos parámetros son decodificados y usados en conjunto con otros mecanismos para formar la voz. El análisis puede ser de lazo abierto o de lazo cerrado. En el análisis de lazo cerrado, los parámetros se extraen y se codifican minimizando una medida (generalmente es el error cuadrático medio) de diferencia entre la señal original y la señal reconstruida. Además en el análisis de lazo cerrado se incorpora una síntesis, de aquí que este proceso se le conoce como análisis-síntesis.

1.2 Organización de la tesis

En la primera sección damos una breve descripción de las propiedades de la voz y continuamos con una revisión de medidas de funcionamiento. En el capítulo 2 discutimos los métodos de codificación de onda y se empieza con una descripción general de los métodos de cuantificación vectorial y escalar y después se discuten los códigos de onda. En el capítulo 3 presentamos los algoritmos para compresión de voz y los métodos de análisis-síntesis. En el capítulo 4, se especifica los tipos de redes de alta velocidad, como por ejemplo: Frame Relay, VoIP, etc cuya aplicación es muy adecuada en la actualidad para el manejo de la voz y datos, también incluiremos la comparación con plataformas de transmisión, además de determinar algunas metodología de transporte y estrategias de manejo de tráfico en función de retardos, ancho de banda, procesamiento/rendimiento de dispositivos de la red la cual es óptima para el transporte de la voz, además presentamos una descripción de las redes voz y datos implementadas. En el capítulo 5 se describe y se analiza un algoritmo de codificación de voz.

1.3 Objetivos de la tesis

El objetivo general de esta tesis es mostrar el algoritmo de Codificación Lineal Predictiva (LPC). Los objetivos específicos son:

1. Mostrar el tráfico de voz a través de una infraestructura de red.
2. Mostrar las exigencias de un tráfico integrando voz y datos.
3. Comparación de la calidad y tasa de bits de los algoritmos de codificación de voz.
4. Mostrar la razón de bits del codificador LPC utilizando propiedades del idioma español.

Capítulo 2

Técnicas de codificación

2.1 Antecedentes Teóricos

La comunicación de voz actualmente es uno de los servicios más dominantes en las redes de telecomunicaciones. En un principio se pretendía que las redes procesaran únicamente aplicaciones de datos, sin embargo en la actualidad existen interfaces para aplicaciones de voz, las cuales bajo ciertas circunstancias operan aceptablemente. Se pretende que la voz empaquetada no use el canal de comunicación a toda su capacidad. Algunos productos del mercado emplean técnicas de compresión ADPCM, CELP, ACELP y tecnología de detección de actividad de voz para utilizar más eficientemente el canal de transmisión. Por los motivos anteriores, en este capítulo se introducen los conceptos necesarios para entender el proceso del habla y los métodos para representar digitalmente las características espectrales en tiempo real para la forma de onda de la voz, los cuales son generalmente conocidos como técnicas de codificación.

2.2 Propiedades Básicas de la voz

El proceso de la voz humana se puede descomponer en dos partes.

- Generación del sonido, formado por el encadenamiento de fonemas (vocales y

consonantes).

- Propagación por el espacio.

La voz humana se produce por medio del aparato fonatorio [1]. Este está formado por los pulmones como fuente de energía en la forma de un flujo de aire, la laringe que contiene las cuerdas vocales, la faringe, las cavidades oral (o bucal) y nasal y una serie de elementos articulatorios (los labios, los dientes, el paladar y la lengua). Ver Figura 2.1

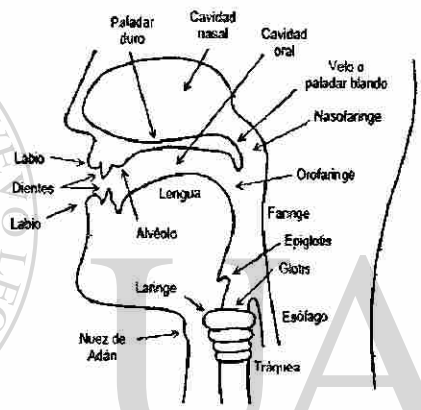


Figura 2.1: Corte esquemático del aparato fonatorio humano

Las cuerdas vocales son, en realidad, dos membranas dentro de la laringe orientadas de adelante hacia atrás (Ver Figura 2.2).

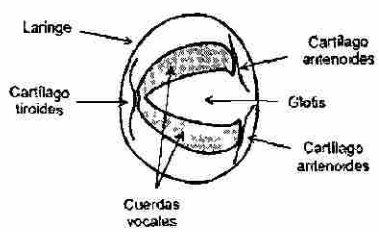


Figura 2.2: Corte esquemático de la laringe según un plano horizontal

La apertura entre ambas cuerdas se denomina glotis. Cuando las cuerdas vocales se encuentran separadas, la glotis adopta una forma triangular. El aire pasa libre-

mente y prácticamente no se produce el sonido. Es el caso de la respiración. Cuando la glotis comienza a cerrarse. El aire que la atraviesa proveniente de los pulmones experimenta una turbulencia, emitiéndose un ruido de origen aerodinámico, conocido como aspiración. Al cerrarse más las cuerdas vocales comienza a vibrar a modo de lengüetas, produciendo un sonido sonoro, es decir, periódico. La frecuencia de este sonido depende de varios factores como por ejemplo: el tamaño y la masa de las cuerdas vocales, de la tensión que se aplique y de la velocidad del flujo de aire proveniente de los pulmones. A mayor tamaño, menor frecuencia de vibración, lo cuál explica el porqué en los varones, cuya glotis es, en promedio, mayor que en las mujeres, la voz es en general más grave. A mayor tensión la frecuencia aumenta, siendo los sonidos más agudos. Así, para lograr emitir sonidos en el registro extremo de la voz es necesario un mayor esfuerzo vocal.

2.3 Clasificación de los sonidos de la voz

Los sonidos emitidos por el aparato fonatorio puede clasificarse de acuerdo con diversos criterios que tienen en cuenta los diferentes aspectos del fenómeno de emisión. Estos criterios son:

1. Según su carácter vocálico o consonántico.
2. Según su oralidad o nasalidad.
3. Según su carácter tonal (sonoro) o no tonal (no sonoro)
4. Según el lugar de articulación e) Modo de articulación f) Posición de órganos articulatorios.
5. Duración.

En este trabajo vamos a enfocarnos exclusivamente en la clasificación de la voz según su carácter sonoro o no sonoro.

Un sonido sonoro[1] es una señal cuasi periódica, llamándose pitch al periodo fundamental de esta señal. En una persona normal el rango del pitch puede estar entre 50 - 400Hz. Las mujeres y los niños tienden a producir una frecuencia promedio de pitch entre 120-500Hz, mayor que la de los hombres ya que las cuerdas vocales son más pequeñas. Los sonidos sonoros tiene una energía mucho mayor que la de los sonidos no-sonoros[2], existiendo entre ellos un gran margen dinámico de 30dB. De hecho desde el punto de vista de energía, los sonidos no-sonoros son del mismo orden del ruido. La estructura física de las cuerdas vocales, genera unas frecuencias de resonancia que quedan de manifiesto en las vocales. La señal de voz se considera como una serie de pulsos o de ruido aleatorio, dependiendo si el sonido es sonoro o no sonoro.

2.4 Producción del habla

El oído humano distingue los diferentes sonidos en base a los espectros en fracciones de tiempo. Debido a las limitaciones del organismo la producción del habla y del sistema auditivo, el sistema de comunicaciones humano tiene un ancho de banda limitado aproximado de 7-8Khz.

Para los sonidos sonoros, como las vocales, el tracto vocal[3] actúa como una cavidad resonante. Para las mayorías de las personas las frecuencias de resonancia se centra alrededor de 500 Hz. Esta resonancia produce picos grandes en el espectro de la voz, las cuales se conocen como formantes.

Los formantes contienen casi toda la información de la señal, variando de acuerdo

[1] Un sonido sonoro es emitido por una sola vibración de las cuerdas vocales sin ningún obstáculo entre la laringe y la apertura nasal, ejemplo las vocales, pero existen consonantes que también lo son: como ejemplo: "b", "d", "m"

[2] Aquellos sonidos producidos sin vibraciones glotales, se denomina no sonoros. Varios de ellos son el resultado de la turbulencia causada por el aire pasando a gran velocidad por un espacio reducido, como las consonantes: "s", "z", "j", "f"

[3] El término de tracto vocal es comúnmente utilizado por los ingenieros. Muchas veces se utiliza para referirse a la combinación de las tres cavidades (cavidad faringeal, oral o bucal y cavidad nasal) y otras, para referirse a todo el sistema de producción del habla

a la vocal pronunciada o lo que es lo mismo, en función de la forma de la cavidad bucal (según la posición de la lengua, dientes, labios). Este hecho significa que el tracto vocal puede modelarse por un sistema lineal de solo-polos. Los sonidos sonoros también exhiben los efectos de vibración de las cuerdas vocales. El efecto de esta vibración es introducir una cuasi-periodicidad de la voz.

Tanto el pitch como los formantes son parámetros característicos del estudio en frecuencia. El pitch nos dice si el locutor es masculino o femenino y la posición de los formantes identifica las vocales.

Desde el punto de vista de análisis temporal, la voz no es un proceso estacionario, ya que sus parámetros característicos (media, varianza, correlación) varían con el tiempo. Sin embargo, si escogemos trozos cortos de voz (20-40ms), se puede considerar un proceso estacionario, ya que los parámetros estadísticos se mantienen relativamente constantes durante esos periodos.

1. Un sonido sonoro son emitido por una sola vibración de las cuerdas vocales sin ningún obstáculo entre la laringe y la apertura nasal, ejemplo las vocales, pero existen consonantes que también lo son: como por ejemplo: "b","d", "m", etc

2. Aquellos sonidos producidos sin vibraciones glotales, se denomina no sonoros. Varios de ellos son el resultado de la turbulencia causada por el aire pasando a gran velocidad por un espacio reducido, como las consonantes: "s", "z", "j", "f".

3. El término de tracto vocal es comúnmente utilizada por los ingenieros. Muchas veces se utiliza para referirse a la combinación de las tres cavidades (cavidad faringeal,oral o bucal y cavidad nasal) y otras, para referirse a todo el sistema de producción del habla.

2.5 Análisis Temporal de la voz

La caracterización de la voz y el establecimiento de sus parámetros temporales son importantes, ya que se aplican directamente a la forma de onda de la señal, a diferencia

de los métodos de análisis en el dominio de la frecuencia, que se basan en el estudio de la representación espectral.

Algunas técnicas útiles y además sencillas para este análisis son la tasa de cruces por cero, la energía y la función de autocorrelación.

La señal de voz presenta grandes variaciones de amplitud (sonidos sonoros/sonidos no-sonoro). La energía en periodos cortos de tiempo refleja bien estas variaciones, permitiendo distinguir entre sonidos sonoros y no-sonoros.

Es evidente, que el factor clave del análisis temporal es la ventana hamming: su forma y su duración. Si la ventana es muy larga, entonces la energía no cambiará mucho de un segmento al siguiente, con lo que no reflejará correctamente los cambios de energía de la señal. Por otro parte, si es muy corta, la energía variará demasiado, con lo que tampoco nos servirá. Lo que buscamos es un filtrado paso bajo, de tal forma que en la energía quede una señal relativamente suave, sin variaciones bruscas, pero que represente adecuadamente la variación de la energía de la voz.

2.5.1 Tasa de cruces por cero

Se entiende cruces por cero cuando dos muestras sucesivas tienen signo distinto. Para señales de banda estrecha, la tasa de cruces por cero nos da una idea de su contenido en frecuencia. Sin embargo, las señales de voz no son de banda estrecha. El contenido espectral de los sonidos sonoros se concentra por debajo de los 3Khz, mientras que los sonidos no-sonoros tienen un contenido espectral en frecuencias altas considerables. Por tanto, parece obvio, que una señal de frecuencia baja (sonidos sonoros) tendrá menor tasa de cruces por cero que otra de frecuencia más alta (sonidos no-sonoro).

Otro problema que se puede ocultar en la tasa de cruces por cero es la existencia de un nivel de frecuencias muy bajas debido a errores que surge al muestrear una señal de voz.

2.5.2 Discriminación voz-ruido

Uno de los principales problemas en el análisis de la voz es poder distinguir cuando existe una señal de voz y cuándo no, como mencionamos anteriormente los sonidos no-sonoros son parecidos al ruido.

El problema de localización del principio y final de una palabra es difícil, sobre todo en sistemas de reconocimiento de voz. La técnica que se utiliza es la energía local y tasa de cruce por cero. Cuando el principio o final corresponda a un sonido sonoro, basta con calcular su energía local, que será muy superior a la del ruido (silencio). Si la palabra empieza o termina por sonidos no-sonoro, entonces la energía es muy parecida a la del ruido. Sin embargo, los sonidos no-sonoros tienen un contenido en frecuencia alta muy superior a la del ruido. Por tanto, utilizaremos como criterio para distinguirlo del ruido la tasa de cruces por cero.

2.6 Análisis espectral de la voz

Muchas veces resulta más interesante el estudio en frecuencia de la voz que en tiempo, por ejemplo al detectar la frecuencia fundamental o pitch, que como sabemos, caracteriza a la persona o que intenta reconocer las vocales que pronunció a partir del estudio de sus formantes. Primeramente, introduciremos unos conceptos básicos de análisis espectral de señales, sin profundizar en detalles teóricos.

2.7 Muestreo

En muchas de las aplicaciones es necesario que la voz esté en un formato digital, para que pueda ser procesada, acumulada o bien transmitirse. Sin embargo la voz procesada digitalmente es más flexible y muestra más ventajas para poder encriptarla. El interés de la codificación de voz o compresión de voz consiste en obtener una representación digital compacta de la señal de voz para llevar a cabo una transmisión eficiente.

La codificación de voz involucra el proceso de muestreo. Para poder procesar digitalmente una señal análoga es necesario convertirla en una señal discreta. El proceso de muestreo es el proceso que convierte una señal continua en el tiempo en una señal discreta, además el rango infinito de amplitudes se reduce a un conjunto finito de posibilidades. La forma de onda de una señal muestreada puede ser representada de la siguiente forma:

$$s(n) = s_a(\hat{n}T) \tag{2.1}$$

Donde s_a es una señal análoga, $\hat{n}$ es un entero y T es el periodo de muestreo o la diferencia de tiempo de dos muestras adyacentes, la cual es determinada por el ancho de banda o la frecuencia más alta en una señal de entrada. El teorema de muestreo establece que si una señal $s_a(t)$ tiene una transformada de Fourier limitada en banda, dada por:

$$S_a(j\omega) = \int_{-\infty}^{\infty} s_a(t)e^{-j\omega t}dt \tag{2.2}$$

Tal que $S_a(j\omega) = 0$ para $\omega \geq 2\pi W$, entonces la señal análoga puede ser reconstruida a partir de sus muestras si se cumple la siguiente condición $T \leq 1/2W$ donde W es la frecuencia de Nyquist.

El efecto de muestreo puede verse en la Figura 2.3. Como podemos ver en la Figura 2.3(b), 2.3(c) la transformada de Fourier que resulta al muestrear la señal análoga limitada en banda dada en la Figura 2.3(a), se obtiene una replica de la señal análoga en cada múltiplo de la frecuencia de muestreo.

Por lo tanto, antes de muestrear una señal análoga debemos tener cierta información general sobre el contenido frecuencial de la señal. En redes de telecomunicaciones, la señal de voz se limita en un rango de 300-3400Hz y se muestrea a 8000Hz.

Otro concepto básico relacionado con la codificación de la voz es la cuantización.

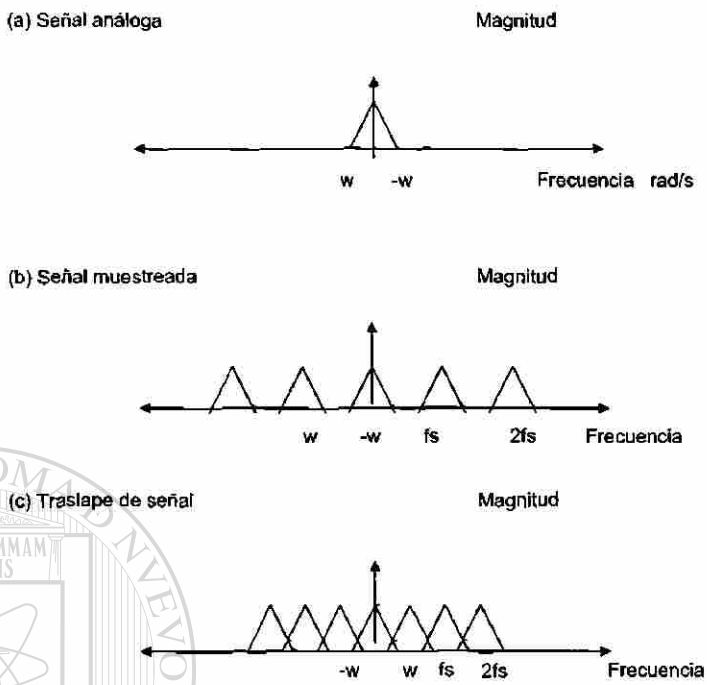


Figura 2.3: Efecto de muestreo (a) Señal original, (b) Señal muestreada T<1/2W, (c) Traslape de señal T>1/2W

2.8 Cuantización

El objetivo de la codificación de voz es representar la voz con un número mínimo de bits, pero sin deteriorar su calidad.

La cuantización o representación binaria convierte una señal de amplitud continua en una señal de amplitud discreta, con valores discretos divididos en regiones con anchos uniformes. A la diferencia entre la entrada no-cuantificada y la salida cuantificada se le conoce como ruido de cuantización. Para minimizar el error, existen diferentes métodos de cuantización que mencionaremos más adelante.

La distancia entre los niveles de amplitud es conocida como tamaño de escalón y generalmente se representa por Δ. Cada nivel de amplitud discreto n se representan por una palabra código $c(n)$, para su transmisión.

Si consideramos que todos los valores de amplitud discreta de una señal se representa

por un mismo número de B bits y la frecuencia de muestreo es f_s, tenemos que la tasa de bits que se transmiten en un canal de comunicación está dada por:

$$T_c = Bf_s \text{ bit/seg} \tag{2.3}$$

La única manera de reducir el flujo de bits en un canal es reduciendo el tamaño del código B. Sin embargo, una longitud reducida significa tener menos niveles en el conjunto de valores discretos de amplitudes, separados por intervalos más grandes.

En la Figura 2.4 se ilustra el concepto de cuantización de 3 bits en dos maneras distintas. En la Figura 2.4(a) se representa el rango de valores divididos en 8 regiones (2^3), así se utilizan la combinación binaria de los 3 bits. En la Figura 2.4(b) se ilustra la relación de entrada y salida, mientas que la entrada es continua, la salida toma únicamente valores discretos. El ancho de cada escalón es constante por lo tanto, la cuantización es uniforme.

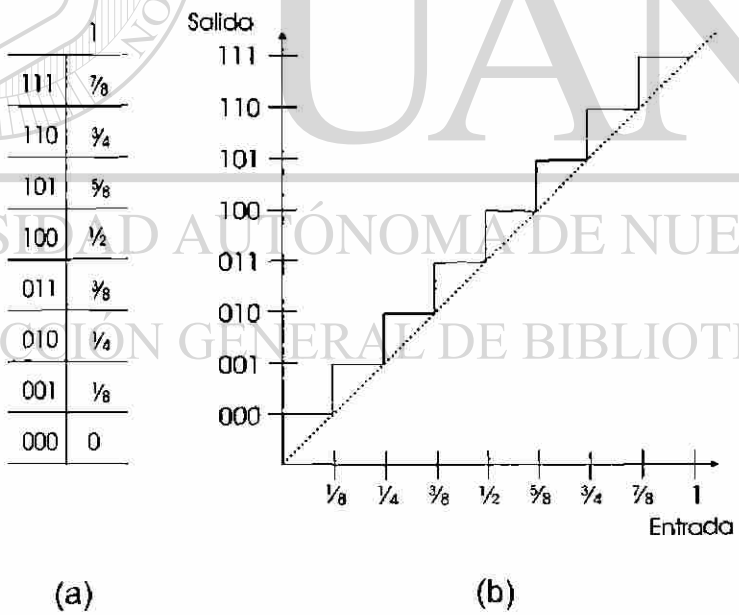


Figura 2.4: (a) y (b): Concepto de cuantización

Como mencionamos anteriormente, el proceso de muestreo convierte una señal continua en una secuencia de muestras discretas cuantizadas en amplitud y codificadas

en una secuencia binaria. Consideremos distintas técnicas de cuantización en el dominio del tiempo como por ejemplo: Cuantización uniforme, Cuantización logarítmica, Cuantización no-uniforme y Cuantización vectorial.

2.8.1 Cuantización uniforme

Los cuantizadores uniformes son aquellos en los cuales la distancia entre todos los niveles de reconstrucción es igual, además no se hace ninguna consideración acerca de la señal que se va a cuantizar. Es por esta razón que no tiene un buen rendimiento perceptual.

La representación de entrada y salida de un cuantizador uniforme se muestra en la Figura 2.5. Como se muestra en la figura, todos los intervalos cuantizados (escalones) son del mismo tamaño. En la cuantización uniforme hay únicamente dos parámetros a considerar: El número de niveles cuantizados y el tamaño del escalón cuantizado , por lo general el número de niveles se elige para ser de la forma $(2^B * \Delta)$ y B se elige para cubrir el rango de muestras de entradas.

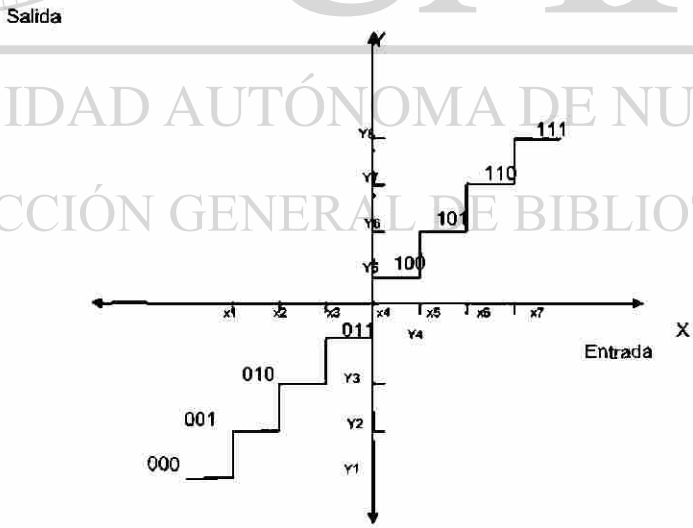


Figura 2.5: Características de transferencia de un cuantizador uniforme

Actualmente, en las redes de comunicaciones se utiliza un cuantizador uniforme de

13 bits para dar una calidad de voz aceptable semejante a la de transmisión telefónica de voz [3].

2.8.2 Cuantización logarítmica

Para proveer el amplio rango dinámico de una señal de voz se sugieren dos métodos de leyes dominantes conocidos como ley μ y ley A. Las señales de voz pueden tener un rango dinámico alrededor de 60dB, de tal modo que son necesarios un gran número de niveles de reconstrucción para que el cuantizador uniforme tenga buena calidad de voz. Debido a la distribución probabilística de la amplitud requiere que la resolución del cuantizador sea más fina en las partes de baja amplitud de la señal. Es obvio, que un cuantizador uniforme no usa uniformemente los niveles de reconstrucción. Esta situación se puede mejorar si la distancia entre los niveles de reconstrucción se aumenta en la medida que la amplitud de la señal aumenta.

Un método simple de lograr un mejor uso de los niveles de reconstrucción es pasar la señal a través de compresor con características logarítmicas antes de cuantizarla. El compresor es un dispositivo que acepta un rango dinámico grande y reduce este rango de acuerdo a una ley de compresión. A la salida del sistema la señal pasa a través de un expansor, que hace la función inversa del compresor, tomando la señal comprimida para restablecer su forma original. El proceso completo se le conoce en inglés como *Companding*, y es usado frecuentemente en la transmisión de voz. El compresor logarítmico hace que la relación señal-cuantización sea proporcional a la amplitud de la señal. La ley de compresión usada en una red telefónica en Norteamérica es la ley μ, y se expresa en la siguiente ecuación:

$$|\nu| = \frac{V \ln\left(1 + \frac{\mu|X|}{V}\right)}{\ln(1+\mu)}, \qquad \mu > 0 \tag{2.4}$$

El valor de $\mu = 225$ ha sido seleccionado para cuantizar la voz en 8 bits.

La ley A de compresión es utilizada en redes telefónicas en Europa, se define de la

siguiente forma:

$$\begin{aligned} |\nu| &= V\left[\frac{1+\ln A|X|}{1+\ln A}\right] \qquad \frac{1}{A} \leq X \leq 1 \\ |\nu| &= V\left[\frac{A|X|}{1+\ln A}\right] \qquad 0 \leq X \leq \frac{1}{A} \end{aligned} \tag{2.5}$$

donde el valor de A es 87.6.

2.8.3 Cuantización no uniforme

Cuando las muestras de entrada no están distribuidas uniformemente, es posible mejorar la relación señal a ruido, usando cuantización no uniforme. La cuantización no uniforme presenta varias ventajas en la codificación de voz por dos razones. Primeramente, un cuantizador no uniforme tiene una mejor función de distribución de probabilidad [3], por lo tanto produce una más relación señal a ruido que un cuantizador uniforme. Segundo, la disminución de las amplitudes de la señal, contribuyen a una mejor comprensibilidad de la voz y son cuantizadas con más presión en un cuantizador uniforme.

2.8.4 Cuantización vectorial

Un conjunto de datos se codifica en un bloque o en forma de vector. A este proceso se le conoce como cuantización vectorial también conocida como cuantización en bloque. En el vector de cuantización vamos a considerar un vector aleatorio de dimensión N.

$$X = \begin{bmatrix} \underline{x}(1) \\ \underline{x}(2) \\ \vdots \\ \underline{x}(N) \end{bmatrix} \tag{2.6}$$

Donde $\underline{x}(i)$ son variables aleatorias reales.

El cuantizador mapea el vector aleatorio x a otro vector aleatorio $\underline{Y}$ de dimensión N.

$$\underline{Y} = \begin{bmatrix} \underline{y}(1) \\ \underline{y}(2) \\ \vdots \\ \underline{y}(N) \end{bmatrix} \tag{2.7}$$

Este mapeo también lo podemos expresar de la siguiente forma:

$$\underline{Y} = Q(x) \tag{2.8}$$

El conjunto $\underline{Y}$ se llama tabla de código o plantilla de referencia donde L es el tamaño de la tabla de código. Para diseñar una tabla de código en un espacio de N dimensiones se particiona en L regiones o celdas C_i, $1 \leq i \leq L$ y un vector Y_i se asocia con cada celda C_i. El cuantizador asigna entonces el vector Y_i de la tabla de código si x está en C_i.

$$q(X) = Y_i \qquad \text{si } X \in C_i \tag{2.9}$$

Estos vectores tiene una distribución espacial y toman únicamente uno de L valores en R^N. La función de distribución de probabilidad consiste de L impulsos de un hiperplano de dimensión N.

En la Figura 2.6 se representa un esquema de vector de cuantización, el cual consiste de un cuantizador de dimensión N. La entrada de los vectores se forma de muestras consecutivas. El cuantizador mapea la entrada de un vector $N \times 1$: $s_i = [s_i(0)\, s_2(1) \ldots s_i(N-1)]^T$ hacia el canal $\{U_n, n = 1, 2, \ldots, L\}$. Por razones prácticas consideremos que en el canal no existe ruido, es decir, $(U_n = \hat{U}_n)$. El conjunto de vectores código diccionario consiste de L vectores, $\{\hat{s}_n = [\hat{s}_1(0)\hat{s}_2(1) \ldots \hat{s}_N(N-1)]^T, n = 1, 2, \ldots, L\}$, el cual permanece en la memoria del transmisor y receptor.

2.9 Codificación de onda

El convertir una señal análoga en una señal digital es generalmente conocido como codificación. En este caso, nuestro estudio está enfocado al análisis de los codificadores

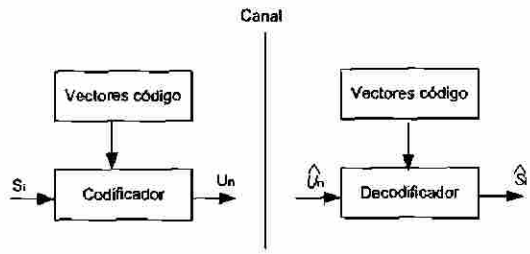


Figura 2.6: Esquema de un vector de cuantización en forma de bloque

de voz.

En Codificación de onda intentamos codificar una señal de voz aprovechando las características espectrales y temporales de esta señal. Por otra parte, la Codificación de voz involucra representar la señal de voz por un conjunto de parámetros. De allí que el cálculo de los parámetros de una trama de voz, y su codificación eficiente en forma digital para su posible transmisión o almacenamiento en cualquier medio de transmisión, sea el problema perceptual de la codificación de la voz. En las técnicas de codificación, una señal de voz en su forma análoga se filtra, seleccionando un ancho de banda de 3-4Khz, y se muestrea a 8000 muestras por segundo para evitar el traslape.

El objetivo principal de un codificador de voz es comprimir la señal, es decir, emplear pocos bits en la representación de voz en forma digital. En las pasadas cuatro o cinco décadas se han propuesto una variedad de técnicas de codificación de voz [5]. Vamos a describir los más importantes de esos métodos, la cual puede ser subdivididos en categorías generales como: Codificador de Onda, Codificador de Voz y Codificadores Híbridos.

Los métodos para representar digitalmente las características temporales y espectrales de las señales de voz se conocen como métodos de codificación de onda. Estos codificadores de onda a comparación de los vocoders (del inglés Voice CODEC, Codificadores de voz) resultan ser más robustos, en el sentido de que estos trabajan mejor con una variedad de señales y operan con una mayor tasa de bits a comparación de los vocoders. En esta sección vamos a describir varios tipos de codificación de onda en el

dominio del tiempo y en el dominio de la frecuencia. Estas técnicas han sido utilizadas ampliamente desde los años 50's en el área de telecomunicaciones.

2.10 Codificadores de onda en el dominio del tiempo

2.10.1 PCM (Modulación de pulsos codificados)

Esta técnica de codificación fue la primera en desarrollarse y también en representar en forma digital una señal de voz.

El proceso PCM es una técnica de codificación que utiliza técnicas de muestreo y cuantización, además de involucrar el proceso de sincronización. La sincronización se refiere a una medida de tiempo, en la cual la información es transmitida.

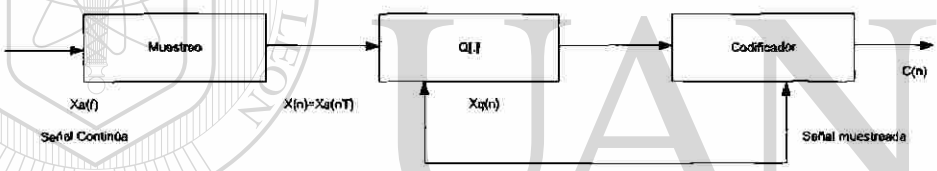


Figura 2.7: Sistema PCM

El PCM es la forma más simple de codificación de onda, en la voz de banda angosta se muestrea a 8000 veces por segundo. Entonces cada muestra de la voz es cuantizada. Si el cuantizador es lineal se usa entonces 12 bits por muestra, dando un flujo binario de 96 Kbits/seg. Para codificadores de voz se encontró que con 8 bits por muestra es suficiente para obtener una buena calidad, dando una relación de bits de 64 Kbits/seg. Existen dos codificadores PCM no lineal que fueron estandarizado en los 60's: La Ley μ en América y la Ley A en Europa. Estos codificadores (también llamados codec) muestran excelente calidad y bajo retraso, ambos son ampliamente usados. Esta técnica la podemos resumir en tres pasos de A/D: Muestreo, Cuantificación y Codificación.

Hay que aclarar que el PCM es un método de codificación y no de modulación con señales de radiofrecuencia. El proceso PCM simplemente cuantifica la amplitud de

cada señal a uno de 2^B niveles, generando B bits/muestra, con un flujo de información total de $2WB$ bits/seg, por $2W$ muestras/seg.

El estándar PCM es utilizado hoy en día en una red pública telefónica, la cual opera con una tasa de muestreo de 8000 muestras/seg. Esta tasa es adecuada para transmisión de voz en una red telefónica, de 8 bits a 64000 bits/seg [5].

2.10.2 Modulación por Pulsos Codificados Diferenciales (DPCM)

La diferencia entre PCM y DPCM es que este último puede ser utilizado a una tasa binaria más baja que la del PCM con la misma calidad de reproducción. Cuando se codifica la voz, existe una alta correlación entre las muestras adyacentes. Esta correlación puede ser utilizada para reducir la tasa de bits resultantes. Un método simple de hacerlo es transmitiendo solamente las diferencias entre cada muestra. Esta señal de diferencia tendrá un rango dinámico más bajo que la señal original de voz, por lo tanto puede ser cuantizada utilizando un número menor de niveles de cuantización.

Vamos a considerar una representación práctica del sistema DPCM, como se muestra en la Figura 2.8. En esta configuración el predictor se implementa con un lazo de retroalimentación al cuantizador, la salida del bloque del predictor es $\bar{s}(n)$, la cual representa la señal muestreada $s(n)$ modificada por un proceso de cuantización, a la salida del predictor tenemos que:

$$\bar{s}(n) = \sum_{i=1}^{M} \hat{a}\tilde{s}(n-i) \tag{2.10}$$

Donde $\hat{a}$ son los coeficientes de Predicción lineal y se eligen para minimizar el error cuadrático medio.

En la entrada al cuantizador tenemos la siguiente diferencia:

$$e(n) = s(n) - \bar{s}(n) \tag{2.11}$$

y $\tilde{e}(n)$ es la salida. Cada valor del error de predicción de cuantización $\tilde{e}(n)$ se codifica

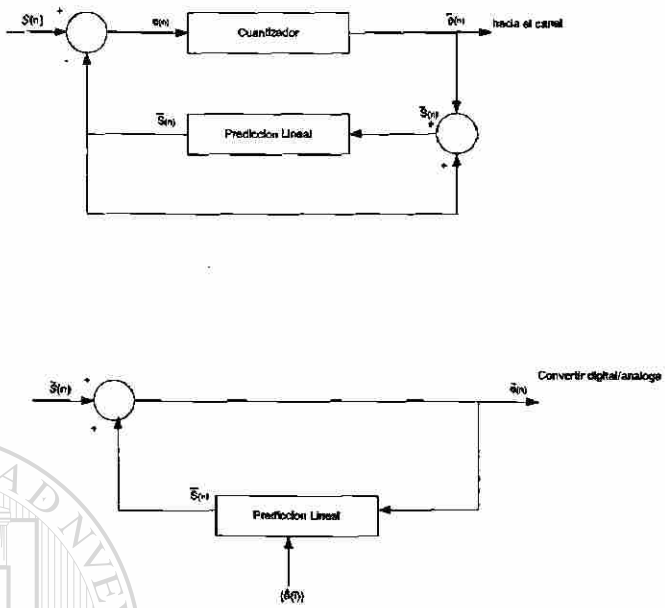


Figura 2.8: Un sistema DPCM (a) Codificador (b) Decodificador

en una secuencia de dígitos binarios y se transmite por el canal hacia el receptor. El error cuantizado $\tilde{e}(n)$ se agrega al error predicho $\bar{s}(n)$ para obtener $\tilde{s}(n)$.

2.10.3 Modulación de pulsos codificador diferencial y adaptable (ADPCM)

Con DPCM el predictor y el cuantizador permanecen fijos todo el tiempo. Se puede alcanzar mayor eficiencia si el cuantizador se adapta a las estadísticas cambiantes del residuo de predicción. También, se puede lograr mayores ganancias si el mismo predictor se adapta a la señal de voz. Esto aseguraría la minimización continua del error cuadrático medio de predicción independientemente del hablante y la señal de voz. Existen dos métodos para adaptar los cuantizadores y los predictores: adaptación hacia delante y adaptación hacia atrás. Con el primer método de adaptación los niveles de reconstrucción y los coeficientes de predicción se calculan en el transmisor, usando un bloque de voz. Después, se cuantizan y transmiten al receptor como información

lateral. El transmisor y el receptor usan estos valores cuantizados para realizar las predicciones y cuantizar el residuo. Para el método de adaptación hacia atrás, los niveles de reconstrucción y los coeficientes de predicción se calculan usando la señal codificada, ya que esta señal es conocida en el transmisor y al receptor no se necesita transmitir ninguna información lateral, de tal forma que el predictor y el cuantizador pueden actualizarse en cada muestra. Este tipo de técnica pueden producir tasas binarias de bits más bajas pero son más sensible a los errores de transmisión que las técnicas de adaptación hacia delante [23].

La modulación diferencial adaptable por codificación de pulsos (ADPCM) es muy útil para codificar la voz a tasas medias de bits. La CCITT ha formalizado un estándar para codificar la voz por teléfono en 32 Kb/s[10] utilizando una adaptación hacia atrás en el cuantizador y predictor.

A continuación se describirá los codificadores en el dominio de la frecuencia.

2.11 Codificación en el dominio de la frecuencia

En los codificadores en el dominio de la frecuencia, la señal de voz es dividida en componentes de frecuencia y codificada en cada banda de frecuencia utilizando diferentes número de bits, por ejemplo, a las señales de baja frecuencia se les asignan más bits que a las de alta frecuencia, con el fin de conservar el formato de información y el pitch de la señal.

2.11.1 Codificación por sub-bandas

Este método de análisis utiliza un banco de filtros, las cuales divide la señal de voz en N sub-bandas de frecuencia usando un banco de filtros paso-banda, la salida de cada filtro es muestreada y codificada. En el receptor las señales son demultiplexadas, decodificadas, para luego ser multiplexadas en el lado transmisor y luego se suman las sub-bandas para producir la señal de salida. El proceso de codificación introduce ruido de cuantización mientras que el proceso de muestreo y decodificación introduce

distorsión por traslape debido al solapamiento entre las bandas. Además las bandas de frecuencia baja contienen más energía espectral en señales voceadas por lo que se utilizan más bits en este tipo de señales que los utilizados en señales de frecuencias altas. En la Figura 2.9 se muestra un codificador por sub-bandas.

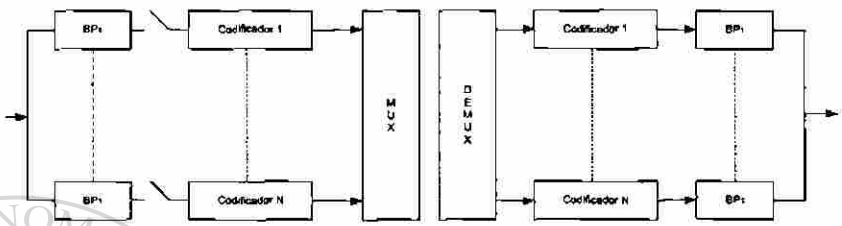


Figura 2.9: Codificador por sub-bandas

La codificación de sub-bandas ha encontrado uso en canales con ancho de banda grandes y de alta calidad utilizados en teleconferencias. Estos sistemas usan los codificadores descritos por los siguientes estándares: Código de sub-banda de AT&T y el estándar G.722 de la CCITT [3].

2.12 Conclusiones

Durante este trabajo se mencionó las técnicas de codificación, como por ejemplo: PCM. Este tipo de codificadores proporcionan una alta calidad de voz a una tasa de bits del orden de 32Kbits/seg pero no son útiles cuando se requiere codificar a baja tasas de bits.

Este tipo de codificación no sigue la filosofía del vocoder en general, sino que simplemente muestrea la voz. A partir de PCM se desarrolla DPCM y el ADPCM que fueron propuestos como estándares por la CCITT (Internacional Consultive Comité for Telephone and Telegraph). En el grupo de vocoder se encuentran los codificadores que sí tienen la naturaleza de la señal a codificar, en este caso la voz, y aprovecha las características de la misma para ganar en eficiencia. Además de permitir trabajar con baja tasa de bits. Gracias a la flexibilidad de los sistemas digitales, se pudo experi-

mentar con formas más sofisticadas de representación de voz. Las investigación ha sido encaminada a conseguir codificadores que utilicen anchos de banda cada vez menor mientras que la calidad de la voz sea cada vez mejor. Con esto se permite utilizar con más eficiencia y eficacia los canales de transmisión y se facilita la encriptación.

Capítulo 3

Codificadores de voz

3.1 Introducción

El objetivo principal de un codificador de voz es comprimir la señal, es decir, emplear pocos bits en la representación de voz en forma digital. En las pasadas cuatro o cinco décadas una variedad de técnicas de codificación de voz se han propuesto, analizado y desarrollado. En los vocoders, la descripción paramétrica [6] del sistema bucal humano puede tomar una variedad de formas, ya sea en el dominio del tiempo o en el dominio de la frecuencia. Ejemplos de codificadores que operan en el dominio de la frecuencia son: el codificador de canal y el codificador de formantes. Esos codificadores envían coeficientes espectrales, describiendo las frecuencias y amplitudes de formante. Por otra parte, los codificadores en el dominio del tiempo incluyen la autocorrelación, la función ortogonal y el cepstrum. Además en nuestro análisis veríamos los codificadores híbridos. En estos últimos se combinan las técnicas de los codificadores de onda con los vocoders, con el propósito de obtener una alta calidad de la voz a bajo bit-rates (inferior a 8Kbits/seg.). En estos codificadores las muestras de la señal de entrada se dividen en bloques de muestras (vectores) que son procesados como si fueran uno solo. Llevan a cabo una representación paramétrica de la señal de voz para tratar que la señal sintética se parezca lo más posible a la original. Algunos de los codificadores

que veremos en este capítulo son: Predicción lineal por excitación de residuo (RELP). Prediccion lineal por excitación Multi-pulso (MPLP). Predicción lineal por excitación de códigos (CELP). Excitación Multibanda.

Entre las aplicaciones principales de los codificadores de voz se encuentran registrar mensajes, encriptar la voz para su transmisión en bandas de radio HF, radio celular digital, o circuitos telefónicos análogos [22]. Las técnicas de codificación básicas son: a) Codificación de canal b) Codificación de formantes c) Codificación homomórfica d) Codificación por Predicción lineal.

3.2 Codificación de canal

Este codificador de canal es uno de los más antiguos de los codificadores y fue estudiado e implementado por Dudley en 1939, cuando desarrolló una técnica de procesamiento digital de señal con calidad de comunicación de voz a un rango de 2400bps [6]. En el codificador de canal se emplea un banco de filtros paso-banda, cada uno tiene un ancho de banda de 100Hz a 300Hz. Además se utilizan 16-20 filtros de respuesta impulsional finita que cubren una banda de audio de 0-4 Khz. Los filtros de ancho de banda angosta son empleados para las bandas de frecuencia bajas y los filtros de ancho de banda amplios son usados para bandas de frecuencias altas. La señal de salida se rectifica y se pasa a través de un filtro paso bajo para encontrar la envolvente de la señal, después se muestrea y se transmite al receptor. El receptor hace lo opuesto al transmisor. Los anchos de banda de los filtros aumentan con la frecuencia debido a que el oído humano responde linealmente a la escala logarítmica de las frecuencias. En la Figura 3.1 se muestra un diagrama de un codificador de canal.

3.3 Vocoder homomórfico

Este tipo de vocoder está basado en un esquema de análisis-síntesis para obtener la secuencia de excitación. Al usar el análisis Cepstral (Aplicación de una transformada

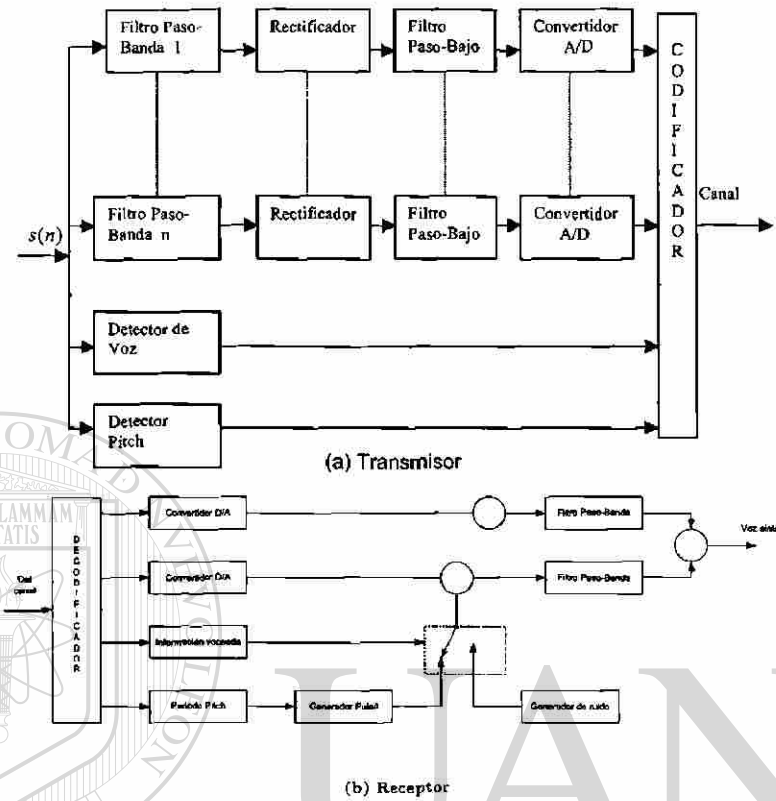


Figura 3.1: Diagrama de bloque de un codificador de canal: Transmisor y receptor

inversa de Fourier), la voz sintetizada se genera en el codificador excitando el sistema bucal. La diferencia entre la voz sintetizada y la original constituye una señal de error, la cual es espectralmente ancha haciendo resaltar las bajas frecuencias y minimizando la señal de excitación. Esta secuencia de excitación se calcula sobre 4 ó 5 bloques dentro de una duración de la trama de voz, lo que significa que la excitación se actualiza más frecuentemente que en el sistema bucal. En la Figura 3.2, se muestra un diagrama de análisis-síntesis de un sistema homomórfico.

En la síntesis, la IFFT nos da la respuesta impulsional del sistema bucal, la cual se convoluciona con la excitación para producir la voz sintética.

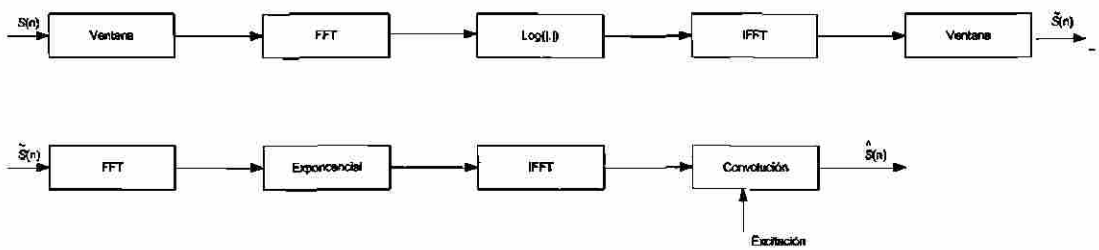


Figura 3.2: Sistema de análisis-síntesis de un sistema Homomórfico

3.4 El vocoder por formantes

Lo podemos ver como un canal de codificación que calcula los primeros tres o cuatro picos llamados formantes en el espectro de un segmento de voz y su ancho de banda. Esto porque en la voz, la información no se distribuye a través de todo el rango del espectro sonoro (20Hz a 20KHz), sino que más bien tiende a concentrarse en esos tres o cuatro picos (200 a 3400Hz.) que por lo general son suficientes para cubrir el rango espectral de la voz, otorgando una buena claridad en la voz. La información de esos picos más el periodo pitch se codifica y se transmite al receptor.

Uno de los objetivos principales de este codificador es determinar la ubicación y amplitud de los picos espectrales y transmitir esta información en vez de transmitir todo el envolvente espectral. Para una trama de voz, cada formante se caracteriza por un filtro digital de dos polos, la cual podemos expresar de la siguiente forma:

$$\theta_k(z) = \frac{\theta_k}{(1 - \rho_k e^{j\omega_k} z^{-1})(1 - \rho_k e^{-j\omega_k} z^{-1})} \tag{3.1}$$

Donde θ_k es un factor de ganancia, ρ_k es la distancia del polo evaluado en el plano complejo, y $\omega_k = 2\pi F_k T$, es la frecuencia del k-ésimo formante en Hz. y T es el periodo de muestreo en segundos. El ancho de banda se determina por la proximidad del polo hacia el círculo unitario. Los formantes pueden ser calculados por predicción lineal o bien por análisis cepstral.

3.5 Codificación de predicción lineal

Una de las técnicas más conocidas es la de predicción lineal (LPC) ya que es uno de los métodos más autoritarios para analizar la voz y además, una de las técnicas más ampliamente investigadas en los últimos 20 años [6].

El método de predicción lineal ha sido usado en numerosas problemas relacionados al procesamiento de señales. En procesamiento digital de voz, se usa el método LP para la transmisión, reconocimiento, y codificación de la voz, entre otras aplicaciones.

El objetivo del análisis de LP es calcular los parámetros de la señal como el tipo de excitación para calcular el periodo pitch y los parámetros de ganancia. Este codificador considera que el tracto bucal puede ser descrito por un filtro con solamente polos de respuesta impulsional infinita, H(z).

$$H(z) = \frac{G}{1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_p z^{-p}} \tag{3.2}$$

G: Ganancia.

Los bloques de voz de aproximadamente 20ms se guardan y se analizan en el vocoder para determinar los coeficientes de predicción a_i [21]. Luego este vector de coeficientes se cuantiza y se transmite al receptor.

Se puede comenzar describiendo el modelo del sistema y continuar con una descripción de algoritmo de predicción lineal.

Bajo el esquema LP, el sistema de producción de la voz se representa en la Figura 3.3. En este sistema se tienen dos estados de excitación (tren de impulsos y un ruido aleatorio).

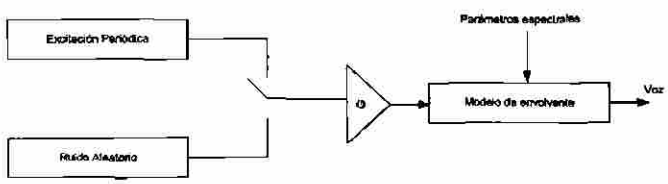


Figura 3.3: Modelo de reproducción de voz en un vocoder

Como se muestra en la Figura 3.3, los vocoders intentan reproducir una señal que se escuche como la voz original. En la parte de transmisión se analiza la voz y se extraen los resultados del modelo de excitación, esta información se envía al receptor donde se sintetiza la voz. El resultado es que se produce voz inteligible a muy baja tasa de bits.

Este es el tipo de vocoder más utilizado porque emplea un modelo de reproducción de los otros vocoders, lo que difiere es en la determinación del modelo de tracto vocal, el cuál puede ser descrito por un filtro todos polos como se representa en la Figura 3.4.

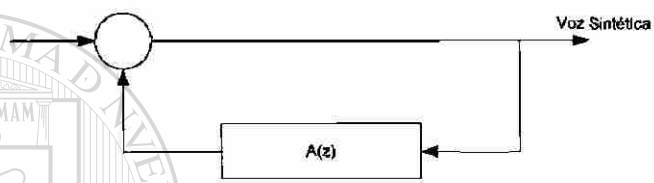


Figura 3.4: El tracto vocal puede ser descrito como un filtro con puros polos

El modelo de sistema de síntesis puede ser representado usando una función en el dominio z:

$$S(z) = \frac{g}{1 - A(z)} X(z) \tag{3.3}$$

Donde g representa la ganancia y $A(z)$ esta dado por $A(z) = \sum_{i=1}^{p} a_i z^{-i}$. Los parámetros del sistema de excitación de la (3.3) son desconocidos y deberán ser determinados por un conjunto finito de muestras de voz. Los coeficientes de $A(z)$ son obtenidos usando predicción lineal (LP). En codificación de predicción lineal, la ventana de análisis es típicamente de 20-30 ms, donde sus parámetros se actualizan cada 10-30ms. Una trama de voz se divide en subtramas cada 5ms, donde los parámetros de subtramas se obtienen por interpolación lineal de parámetros de tramas adyacentes. El codificador por predicción lineal es uno de los más populares de los vocoders debido a que el modelo del tracto bucal con únicamente polos funciona muy bien. Puede ser utilizado para obtener claridad en la voz a un flujo binario de 2.4Kb/s.

3.6 Codificación híbrida

Como mencionamos anteriormente, existen dos tipos de codificadores de voz: codificador en forma de onda y el vocoder. Los codificadores de forma de onda tratan de mantener la forma de onda de la señal a codificar. Son capaces de producir voz de alta calidad a flujos binarios medios, del orden de los 32Kb/s. Pero no pueden ser utilizados para codificar la voz a tasas binarias bajas. Por otra parte los vocoders tratan de producir una señal que suene como la original, reduciendo así el flujo binario.

3.6.1 Predicción lineal por excitación de residuo (RELP)

Este vocoder fue propuesto en la mitad de los años sesenta. Opera en un rango de 6 y 9.6 Kbits/s. Además comprime el ancho de banda a 800Hz, su objetivo es complementar la información que no haya sido capturada por el análisis de LP; como por ejemplo, la fase, la información pitch y los ceros debido a sonidos nasales[17]. Sin embargo el concepto de codificación de predicción residual también ha sido utilizado en ADPCM y en Codificación de Predicción Adaptable, como se vio en el capítulo anterior. Además el vocoder RELP confía en el hecho de que las componentes de bajas frecuencias de voz no son percibidas. En el receptor la señal residual banda base se procesa para que tenga un espectro lineal plano. El diagrama de bloque de un vocoder RELP codifica el residual en el dominio de la frecuencia usando FFT, como se muestra en la Figura 3.5(a). En este sistema, el FFT calcula las magnitudes y fases de las componentes de frecuencia del residuo dentro de la banda base. Estos se codifican y se transmiten hacia el receptor. La calidad de voz del vocoder RELP se limita por la pérdida de información al filtrar la señal residual en banda base. El código de predicción lineal presente evita este problema usando un modelo de excitación (Q) eficiente que puede ser óptimo para el acoplamiento de la señal y la percepción.

Los codificadores RELP se utilizan comúnmente para dar una buena calidad de voz a una razón de 9.6Kb/s.

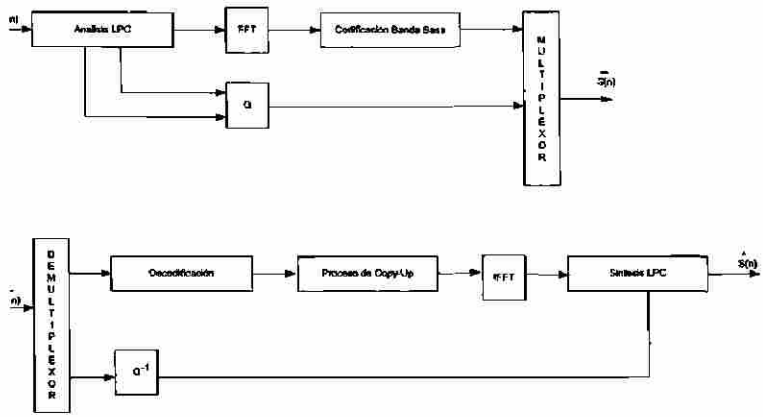


Figura 3.5: Diagrama del vododer RELP

3.6.2 Predicción lineal por excitación multi-pulso (MPLP)

Este algoritmo establece una secuencia de pulsos múltiples espaciados irregularmente, como se muestra en la fFigura 3.6. Este algoritmo utiliza 4-6 pulsos cada 5ms. Además es más claro que el vocoder clásico de predicción lineal porque el MPLP codifica tanto la amplitud como la ubicación de los pulsos. Este algoritmo produce buena calidad de voz en razones bajas de bits como de 10Kbits/seg. Una de sus aplicaciones es en la telefonía celular.

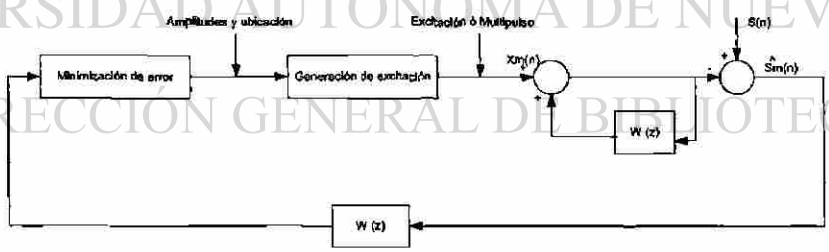


Figura 3.6: Análisis de MPLP

3.6.3 Predicción Lineal por excitación de códigos (CELP)

Como se dijo anteriormente el problema principal de los vocoders LP es el modelo simple que se utiliza en la excitación. Los métodos para resolver este problema son el

codificador de pulsos múltiples y la predicción lineal por excitación de códigos.

En el codificador CELP, la voz se pasa a través del conjunto en cascada formado por el predictor del tracto bucal y el predictor del "pitch". La salida de este predictor es una buena aproximación al ruido gaussiano. Esta secuencia de ruido es cuantizada y transmitida al receptor. Los codificadores de pulso múltiples la cuantizan utilizando una serie de pulsos ponderados. El codificador CELP usa cuantización vectorial [14] por lo que transmite únicamente el índice de la palabra código que produce la voz de mejor calidad. La búsqueda de la palabra código se lleva a cabo utilizando una técnica de análisis-síntesis, como se muestra en la Figura 3.7. La voz es sintetizada para cada palabra código, luego la palabra código que produce el menor error es seleccionada como excitación. La medida de error utilizada es ponderada de tal forma que la palabra código seleccionada, produzca la mejor calidad de la voz.

La búsqueda de la palabra código es computacionalmente intensiva, pero se han desarrollado algoritmos rápidos de tal forma que el codificador CELP pueda ser implementado en tiempo real, utilizando procesadores de señales digitales (DSPs). Actualmente, esta técnica es uno de los métodos más efectivos de obtener voz de alta calidad a razones bajas de bits. Por ejemplo el estándar Federal de U.S. FS-1016 describe el codificador celp que opera a 4.8Kbits/seg para ancho de banda angosta o voz vía telefónica[24].

3.6.4 Excitación multibanda (MBE)

El objetivo de este vocoder de excitación multibanda (MBE) es obtener una buena calidad de voz[17]. El codificador de excitación multibanda utiliza el mismo modelo de producción de voz de dos etapas de los vocoders tradicionales (es decir, la voz se produce por la excitación del modelo del tracto bucal). El modelo del tracto bucal es el mismo que se utiliza en los vocoders, pero el método de excitación es radicalmente distinto. Con los vocoders tradicionales la excitación se establece por una secuencia de sonidos ya sean sonoros o no sonoros. La excitación es una secuencia de ruido para

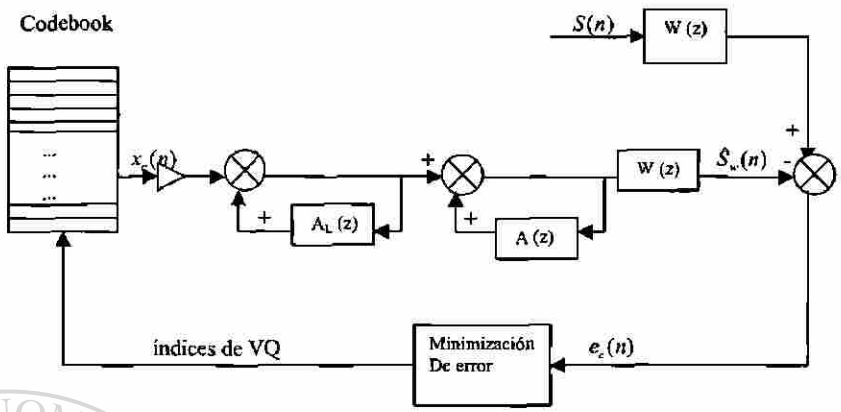


Figura 3.7: Análisis-Síntesis de un codificador Celp

los sonidos no sonoros y un tren de impulsos para los sonidos sonoros. En el vocoder MBE, por lo contrario, la excitación se divide en varias sub-bandas. En cada una de ellas el espectro de la voz se analiza para determinar si es o no sonoro. Al igual que los vocoders, los segmentos sonoros se codifican utilizando un tren de impulsos y los no-sonoros utilizan ruido. Esto permite que la voz codificada aparezca con características simultáneas de sonoridad/no-sonoridad como la voz real.

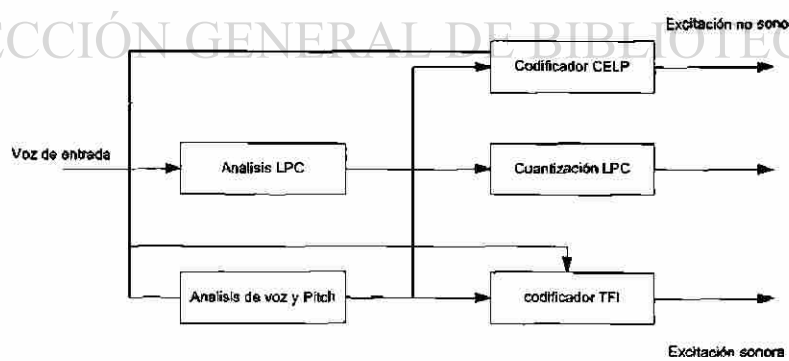


Figura 3.8: Codificador TFI

3.7 Conclusiones

En este capitulo se visualizó la tercera parte de los codificadores, codificadores híbridos. Estos combinan elementos de codificación de forma de onda y los vocoders para brindar una calidad de voz aceptable en el rango medio y bajo de velocidades de salida (3-16Kbits/seg), como se mencionó anteriormente, mantienen el modelo de tracto vocal de los vocoders, en particular la representación LPC. La diferencia básica entre distintos codificadores de este grupo, está la forma de onda de representación de la señal de excitación. Por ejemplo en el codificador RELP, la excitación viene dada por una serie de impulsos regulares, el codificador CELP utiliza cuantización vectorial para una señal de excitación, el codificador VSELP emplea dos cuantizadores VQ sumados, el codificador MBE emplea la codificación multi-banda del espectro de excitación.

De los tres tipos de codificadores, los codificadores híbridos son los que han encontrado más aplicaciones en los sistemas de comunicaciones actuales. Su importancia refleja en los más de diez estándares distintos que se han establecido en los últimos años a nivel mundial, y se espera que esta tendencia continúe en los próximos años.

Capítulo 4

Redes

4.1 Introducción

Hoy en día las redes de paquetes de la nueva generación pueden transportar voz, datos y vídeo sobre una infraestructura común.

Cuando se trata de integrar de voz y datos sobre una misma red, hay que tener en cuenta las diferentes características del tráfico de voz y datos; por una parte, la voz necesita de un retardo constante en la red, mientras que los datos pueden fluir a distintos ritmos, encargándose el receptor de reordenarlos; por otra lado la voz admite cierta distorsión de la señal ya que el ser humano es capaz de entender un mensaje aunque presente algunas alteraciones; mientras que la transmisión de datos requiere una alta calidad porque pueden producirse errores que pueden ser fatales para su compresión. Una de las técnicas básicas para la integración de voz sobre una red de datos es digitalizado mediante algoritmos como PCM (64Kbits/seg), ADPCM (32Kbits/seg) u otro que consigue una transmisión aceptable. Así, el tráfico de voz (procedente de un teléfono o un PBX) y datos se puede mezclar sobre una línea de transmisión punto a punto, consiguiéndose ahorros muy importantes. El ancho de banda se puede asignar dinámicamente, en función de la actividad o inactividad de los canales para la optimización del enlace; también, se puede reservar uno determinado para garantizar la

transmisión de voz.

En este capítulo se introducen distintas tecnologías de transmisión de voz actualmente estandarizadas, iniciamos con especificar redes que se clasifican en redes de circuitos y redes de paquetes[20].

El proceso de conmutación que se utiliza en los dos tipos de redes se describe a continuación.

4.1.1 Redes de conmutación de circuitos

El mejor ejemplo que podemos poner para entender este tipo de red es precisamente una red telefónica. El hecho de contar con un circuito para su uso exclusivo permite a las redes de conmutación de circuitos ser muy adecuadas para el transporte de tráfico sensitivo al retardo y que por lo tanto requiera su transmisión casi en tiempo real. Este es el caso de servicios como la voz, que al ser digitalizados con la técnica PCM produce un flujo constante de datos (64Kbps en el caso de la voz) que deben llegar sin problemas y en el extremo distante reconstruimos de nueva cuenta la señal. Dentro de las redes de conmutación de circuitos aún se puede definir dos tipos de circuitos. Por un lado están los circuitos conmutados y por otro los circuitos dedicados o privados. El primer tipo de circuitos corresponde a los ya mencionados con una llamada telefónica. Otro ejemplo de redes de conmutación de circuitos es la Red Digital de Servicios Integrados (ISDN), en la que además de centrales telefónicas y medios de transmisión digitales tenemos que los equipos instalados en los usuarios son también digitales. Los circuitos privados, también conocidos como líneas privadas, son conexiones que se establecen de manera permanente entre dos usuarios de la red. Este es el caso de las redes que operan bajo la técnica de multiplexaje por división de tiempo (TDM, Time División Mutiplexing). En México, la primera red de este tipo fue la Red Digital Integrada de Telmex.

4.2 Redes de conmutación de paquetes

En redes de conmutación de paquetes los usuarios, al igual que las redes telefónicas cuentan con un acceso al elemento de conmutación de paquetes que en este caso se denomina simplemente conmutador de datos. La información a enviar es dividida en unidades de información denominadas paquetes, tramas o en el caso de tecnologías modernas celdas. Cada paquete está formado básicamente de dos partes. Primeramente se tiene un poco de información adicional denominada encabezado (overhead). El encabezado contiene la información necesaria para el enrutamiento y la verificación de errores. La segunda parte de cada paquete es precisamente el campo de información en el que se transporta la información útil. La longitud de cada paquete dependerá del protocolo que en específico se maneje en una red particular, puede variar entre 53 bytes y 1500 bytes[1].

En la tabla 4.1 presentaremos un resumen de las características que diferencian a las técnicas de conmutación de circuitos y conmutación de paquetes.

4.3 Estandarización

En el mundo de las redes de telecomunicaciones existen diversas entidades de estandarización, algunas con funciones de carácter internacional, otras regionales o bien nacionales. Sin embargo y de forma general, todas persiguen lo mismo. Coordinan las acciones y esfuerzos de todos los participantes del mercado con el fin de salvaguardar las inversiones mediante la emisión de estándares que especifican por escrito los detalles que los interesados han acordado.

Conmutación de circuitos	Conmutación de Paquetes
• Adecuado para la comunicación de voz • Se requiere compatibilidad punto a punto entre terminales. • Sujeto a Bloqueo. • Retardo grande en el establecimiento de llamada. • Prácticamente sin retardo de transporte. • Moderadamente preciso. • Uso ineficientemente de los recursos de la red.	• Solamente para datos. • Conversión de velocidad, código y protocolos hechos por la red. • Prácticamente sin bloqueo. • Tiempo pequeño de establecimiento de llamada. • Retardo de milisegundos de transporte impuestos por la red. • Altamente preciso • Uso eficiente de los recursos de la red.

Tabla 4.1: Comparación de técnicas de enrutamiento

Las responsabilidades de las organizaciones son las siguientes:

- Desarrollo de estándares para equipo y sistemas.
- Coordinación de la información necesaria para la planeación y operación de los servicios de telecomunicaciones.

Ejemplo de Organizaciones:

ANSI:

- T1.606, Marco de operación
- T1.606 Descripción del servicio.
- T1.606 Administración de la congestión.

- T1.618 Nivel de enlace de datos.
- T1.617 Señalización

UIT-T (Antes CCITT)

- I.122 Marco de operación.
- I.233 Descripción del servicio.
- I.370 Administración de la congestión
- Q.922 Nivel de enlace de datos.
- Q.933 Señalización

Frame Relay Forum

Se constituye a Principios de 1991, hoy en día cuenta con cerca de 100 miembros con presencia mundial. Fue formado para estimular y soportar el desarrollo e implementación de los productos, servicios y estándares de Frame Relay.

Documentos del foro

FRF.1.1	Contrato de Implementación de usuario a red, Enero 19,1996.
FRF.2.1	Implementación de Frame Relay Red a Red julio10, 1995.
FRF.3.1	Implementación de Encapsulación Multiprotocolo, junio 22, 1995
FRF.4	Implementación de un circuito virtual switcheado, Diciembre 1994
FRF.5	Implementaciones de red Frame Relay/ATM, Diciembre 20 1994.
FRF.6	Servicio Frame Relay a Cliente, Marzo 1994.
FRF.9	Compresión de datos sobre Frame Relay , Enero 22,1996.
FRF.10	Implementación de voz sobre Frame Relay, Mayo 5,1997.
FRF.1.2	Fragmentación de Frame Relay, Diciembre 15,1997.

4.4 Estándares de codificación de voz

Para llevar a cabo una interrelación entre las diferentes redes de telecomunicaciones y estándares de algoritmos de codificación de voz. Existe una organización como la

ITU (Unión Internacional de Telecomunicaciones), el instituto estándar de telecomunicaciones Europa (ETSI) y la asociación industrial de telecomunicaciones (TIA), cada una de ellas definen y publican los estándares de telecomunicaciones. En general, los artículos disponibles en los estándares incluyen la documentación que describe al algoritmo.

4.5 Modelo de referencia de interconexión de sistemas abiertos, modelo OSI

En el modelo OSI se considera principalmente los protocolos, el cual establecen una conexión lógica antes de transferir información. Este modelo consta de 7 capas. Las capas del 1 al 3 comprenden funciones para el transporte de información de un sitio a otro. Estas funciones sirven para construir una red de comunicación. No entraremos a detalle para el resto de las capas porque no utilizan protocolos para las capas 4 a 6. A continuación se describe brevemente las capas del sistema OSI (Open System Interconnection).

4.5.1 El nivel físico

El nivel Físico maneja la transmisión de bits a lo largo de la conexión física, especificando los aspectos eléctricos, mecánicos y funcionales de la comunicación.

Un protocolo de nivel uno generalmente define:

- El tipo de conector.
- Los niveles de voltaje de las señales de los pines.
- El nombre de los pines y sus funciones.
- Los procedimientos mediante las cuales se utilizan a las señales para manejar el transporte de los bits a través del medio.

4.5.2 El Nivel de Enlace de Datos

La tarea del protocolo del nivel de enlace es manejar la transferencia de las unidades de datos (tramas) de un sistema a otro. En este protocolo se definen las siguientes características:

- Administración de enlace
- Control de errores
- Control de flujo
- Recuperación en caso de falla.

4.5.3 El nivel de red

El protocolo de nivel de red maneja la transferencia de unidades de datos conocidas como paquetes. Los protocolos de nivel físico y de enlaces funcionan en una base de conexión a conexión y de enlace a enlace respectivamente, mientras que el protocolo de nivel de red funciona en una base de punto a punto a través de la red pública de datos. Este tipo de nivel juega un papel más de red y con las siguientes funciones:

- Establecer conexiones punto a punto.
- Direcciona y enruta punto a punto.
- Realiza control de flujo punto a punto.
- Libera las conexiones de la red.
- Se recupera de fallas que se presentan en el nivel de enlace.
- Proporciona servicios opcionales de la red.
- Proporciona funciones de diagnóstico de la red.

4.6 Voz sobre redes de paquetes

A diferencia de las redes telefónicas, donde para cada conversación se establece un vínculo "estable" y "seguro", las redes de datos admiten pérdidas de paquetes. En encapsulamiento de voz, datos y video es encapsulado en paquetes y enviado, sin confirmación de recepción de cada paquete. Puede que haya un porcentaje de paquetes que no llegan al destino, en este caso puede escucharse como interrupciones en la voz.

La demora de estos paquetes puede deberse a los siguientes factores:

1. Algoritmos de compresión
   - G.711 (64Kbits/seg) que presenta un retraso de 125μ
   - G.728 (16 Kbits/seg) que presenta un retraso de 2.5ms
   - G.729 (8 Kbits/seg) que presenta un retraso de 10 ms
   - G.723 (5.3 o 6.4 Kbits/seg) que presenta un retraso de 30ms
2. Procesamiento
   - Implementación de protocolos.
3. Red
   - Velocidad de transmisión.
   - Congestión
   - Demora de los equipos de red (routers, gateway, etc).

Para poder transmitir las muestras codificadas de voz sobre redes de datos, es necesario armar "paquetes". Cada paquete tiene una información mínima de información (bytes) de control: Cabecera de paquete, origen, destino, etc. (Más adelante se describirá la estructura de paquetes o tramas que se utilizan en una red por Frame Relay).

4.7 Funcionalidad del X.25

En 1976, cuando nace X.25, se empleaban exclusivamente circuitos telefónicos analógicos, los cuales presentaban bastante ruido blanco y ruido impulsivo. Esta pobreza de calidad ocasionaba una probabilidad alta de errores en la transmisión. Por otro lado, los equipos clientes de la red X.25 no eran precisamente muy inteligentes, pues su capacidad de procesamiento era baja en comparación con los equipos modernos de cómputo [1]. Las razones anteriores exigían un protocolo de comunicaciones dentro de la red que ofreciera la robustez y por ende la confiabilidad ante tales circunstancias. De esta forma, el nivel 2 de X.25 certifica que cada trama sea recibida adecuadamente en cada enlace y en caso contrario se solicita la retransmisión, a nivel de paquete, es decir a nivel 3 se verifica una vez más la integridad de cada paquete. Evidentemente todas estas funciones de control de errores y de supervisión de la comunidad se traducen en un incremento de la información de encabezado reduciendo la eficiencia de la comunicación. Actualmente podemos confiar en medios de transmisión como por ejemplo las fibras ópticas cuya tasa de errores (BER, Bit Error Rate) es tan baja como 1×10^{-14}. Los elementos que ahora se comunican en las redes de datos cuentan con una alta inteligencia que les permite tomar control sobre algunas de las funciones que antes eran responsabilidad exclusiva de la red, ahora las computadoras son en realidad parte de la red.

4.8 Funcionamiento de Frame Relay

Aquí aparece la segunda tecnología, Frame Relay, que toma ventaja de la creciente inteligencia que ofrecen los elementos que se intentan conectar en los extremos de la red. Asimismo, es requisito para emplear Frame Relay el contar con enlaces digitales de alta confiabilidad y calidad. El protocolo Frame Relay no incluye funciones de recuperación de tramas por errores ni tampoco lleva control de flujo de información. Si se presentan problemas durante la transmisión, será responsabilidad de los dispositivos

de los extremos, y de los protocolos de capas superiores en los cuales reside la recuperación de las tramas perdidas. Frame Relay es un protocolo basado en tramas y orientadas a conexión. Esto significa que a través de la red se establece un circuito virtual permanente (CVP) para llevar la información de un usuario a otro. El circuito virtual lo que define es la trayectoria a seguir de un usuario al otro. Sin embargo, un mismo enlace físico, comparte entre sí varios enlaces lógicos. Esto permite enrutar las tramas a lo largo de la red sin necesidad de contar con muchos enlaces físicos. Las tramas se reciben en cada nodo y de acuerdo a tablas de enrutamiento, se envían al siguiente nodo de la manera inmediata. El procesamiento sobre cada trama es mínimo, lo que aumenta es el desempeño de la red. Las funciones que realizan los nodos de la red Frame Relay incluyen: Proveer el acceso de la red, envío en orden de las tramas, enrutamiento y multiplexaje.

Para los extremos de la red, quedan las funciones de manejo de errores, acuse de recibo y solicitud de transmisión. La verificación de errores que se efectúa en Frame Relay es un cálculo de CRC (Compromiso, Recuperación, Concurrencia) en cada trama, este cálculo se encarga de coordinar las interacciones de múltiples partes, sin embargo al detectar una trama con CRC errada esta se descarta.

Además de los errores, una trama puede ser descartada por motivo de congestión en la red. De hecho se tiene dos bits, el FECN (Forward Explicit Congestion Notification) y el BECN (Backward Explicit Congestion Notification) para dar a conocer al otro extremo el estado de congestión. El que recibe esta notificación puede optar por dejar de enviar tramas o por seguir enviando. Sin embargo, si se continúa generando información, se corre el riesgo de que esas tramas sean descartadas. Además se tiene un bit (Discard Elegibility) con el que se etiqueta a una trama como descartable, es decir, ante situaciones de congestión se descartaran primero las tramas con este bit en "1".

4.9 Estructura de trama de Frame Relay

Dependiendo del campo de información, la longitud de las tramas puede variar, sin embargo la longitud nunca deberá exceder los 8189 Bytes.

Elemento	Formato
Bandera	1 Byte con la forma: 01111110
Encabezado	Puede ser de 2,3 o 4 Bytes
Información	Hasta de 4096Bytes
Chequeo de errores (FCS)	2 Bytes
Bandera	1 Byte con la forma: 01111110

Tabla 4.2: Estructura de las tramas de Frame Relay Elemento Formato

- Bandera ("Flag"): Todas las tramas inician y terminan con una bandera, la cual tiene la siguiente estructura: 01111110. Esta secuencia le permite al receptor identificar el inicio de una trama y poder sincronizarse al flujo de tramas.
- Encabezado ("Header"): Este campo generalmente es de 2 bytes, sin embargo puede tener también una longitud de 3 ó 4 bytes. Aquí se concentra la información referente a la dirección o identificador, congestión y elegibilidad para descartar.
- Campo de Información("Information field"): Este campo contiene los datos de usuario y consiste en un número entero de octetos. El tamaño máximo por omisión es de 262 Bytes y el mínimo tamaño es de 1 byte. Sin embargo el valor de 1600 bytes es fuertemente recomendado para aplicaciones como interconexión de LAN's,
- Secuencia de verificación de trama("Frame Check Sequence"): Consiste en 2 Bytes que se utilizan para verificar que la trama se ha recibido sin error, utiliza un chequeo cíclico redundante o CRC con el polinomio X16+X12+X5+1 definido por el UIT-T.

Bandera	Cabecera	Campo de información	FCS	Bandera

Tabla 4.3: Trama de Frame Relay

4.10 Comparación de Frame Relay y X.25

El Frame Relay es simplemente una técnica de conmutación de paquetes diseñada para operar en redes digitales de alta calidad y confiabilidad. Se le considera una evolución de X.25 ya que Frame Relay resulta ser más eficiente que el X.25. Al querer transmitir un paquete de datos del origen al destino con Frame Relay tenemos que las tramas son de mayor tamaño. Al no verificar las transmisiones entre los nodos, se reduce dramáticamente la cantidad de overhead (cabecera de la trama) y se aumenta considerablemente la carga computacional hecha en una unidad de tiempo.

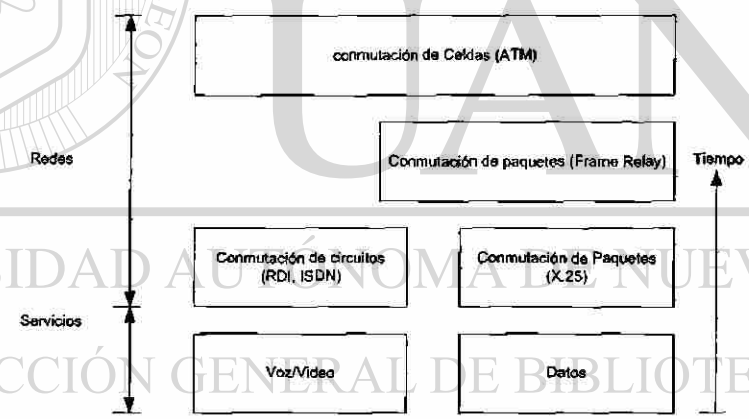


Figura 4.1: Panorama general de la tecnología

El Frame Relay tiende el puente hacia el futuro en el que se hacen converger todos los servicios en una sola red digital.

Entre sus aplicaciones actuales están las redes privadas de voz y datos. Además provee acceso a Internet. Aprovecha las ventajas de los medios digitales y maneja tráfico de alta velocidad por ráfagas, involucra un procedimiento mínimo en los nodos de la red, reduce el número de puertos y de líneas de acceso, y por último ha tenido muy

buena aceptación en la industria y su estandarización ha alcanzado cierta madurez.

Aunque por Frame Relay típicamente se transmiten protocolos de datos, se han implementado en los multicanalizadores de Frame Relay algoritmos que asignan prioridades a los paquetes que reciben. Al digitalizarse la voz, esta es tratada como datos y es enviada empaquetada al igual que los datos. La primera prioridad es asignada a la señal de voz y la segunda prioridad es asignada a las señales de datos.

A continuación mostramos las ventajas y desventajas de la transmisión de voz sobre Frame Relay.

Ventajas	Desventajas
• Uso integrado de la red de datos con voz.	• La calidad de voz varía de un momento a otro.
• Posibilidad de transmisión de faxes.	• Se produce eco y retardo.
• Diferentes velocidades de compresión 32,16,8,4.8 y 2.4K.	• Tamaños de tramas entre 20 y 83 bytes producirá pequeños silencios, si se pierde una trama.
• Se efectúa un ahorro del 30 al 50%, dependiendo del número de llamadas que se efectúen entre sitios.	• VoFR funciona cuando el carrier ofrece priorización de tráfico, QoS (Quality of Service).

Algunos productos del mercado utilizan técnicas de compresión ADPCM, CELP, ACELP y tecnologías de detección de actividad de voz, para utilizar mas eficientemente el canal de transmisión.

Standard	Año	Tipo de Codificación	Relación de Bits(Kbits/s)	Retraso de Algoritmo(ms)
ITU-G.711	1972	PCM	64	0.125
ITU-G.721	1984	ADPCM	32	0.125
ITU-G.726	1991	VBR-ADPCM	16,24,32 y 40	0.125
ITU-G.728	1992	LD-CELP	16	0.625
JDC Japannese Full-rate		VSELP	6.7	20
GSM Half-rate	1994	VSELP	5.6	24.375
ITU-G.723	1995	MLQCELP	5.27/6.3	37.5
American DOD FS1016	1990	CELP	4.8	45
American DOD FS1015	1984	LPC-10	2.4	22.5(mínimo)

Tabla 4.4: comparación de estándares utilizados en codificación de voz.

Capítulo 5

Codificación de Predicción Lineal

5.1 Introducción

En los últimos años se ha hecho un esfuerzo en desarrollar métodos de codificación de voz para distintas aplicaciones, como por ejemplo, comunicaciones móviles, la transmisión de voz sobre Internet involucrando la eliminación de redundantes información en la señal de voz, llevándonos a manejar una buena calidad en las señales de voz en una relación de bits tan bajos como 8Kbits/seg. Los algoritmos de voz son cada vez más sofisticados y más demandantes en términos del número de cálculos requeridos por segundo. Además, existe una implementación eficiente de los algoritmos de codificación de voz en procesadores de señal digital (DSPs).

5.2 Descripción general del codificador

En este capítulo se describe el diseño, implementación y desempeño de una propuesta de codificador de voz, modelo LPC. Este codificador propuesto funciona en el rango de 5.3 a 12.4 Kbits/seg para manejar varias calidades subjetivas de voz además de mostrar una variedad de parámetros de la voz, usando el método de Predicción Lineal, con ejemplos típicos de señales de voz.

Una de las técnicas más útil para el análisis de la voz es el método de predicción lin-

eal (LPC, Linear Prediction Coding). Este método ha sido una de las más dominantes para el cálculo de los parámetros básicos de la voz, como por ejemplo, detectar la frecuencia fundamental o pitch, los formantes, el contenido espectral de los sonidos todos estos cálculos nos encamina a utilizar el LPC para representar la voz a una transmisión de bits muy bajo.

La predicción lineal se relaciona con el modelo de producción de la voz donde la voz puede ser modelada por pulsos cuasi periódicos para la voz voceada o bien ruido aleatorio para la voz no-voceada. El método de predicción lineal nos proporciona robustez, confianza y precisión en cuanto al cálculo de los parámetros que caracterizan el sistema.

En los años 70's la técnica de predicción lineal fue aplicada para las señales de voz y al final resultó ser una técnica útil en el procesamiento de voz, después sus areas de aplicación fueron en reconocimiento y codificación de voz. El codificador de predicción lineal se refiere a una serie de cálculos para el modelo de la voz.

5.3 Principios Básicos del análisis de Predicción Lineal

El modelo apropiado para la discusión del análisis de predicción lineal se muestra en la Figura 5.1

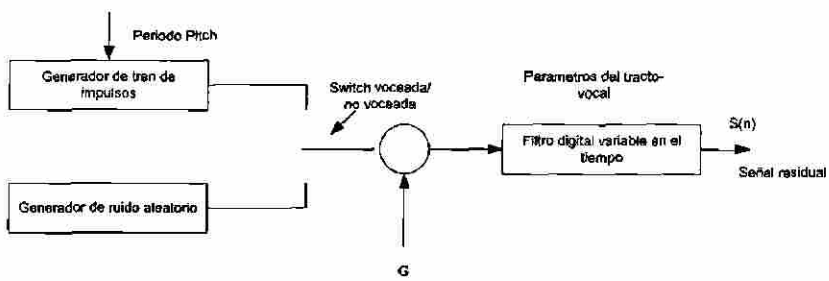


Figura 5.1: Diagrama de bloque simplificado para la producción de la voz

En este caso el efecto de radiación, el tracto vocal y la excitación glotal está rep-

resentado por un filtro digital variante en el tiempo, cuya función estado-estable se representa de la siguiente forma:

$$H(Z) = \frac{A(z)}{B(z)} = \frac{G}{1 - \sum_{k=1}^{p} a_k z^{-k}} \tag{5.1}$$

Como mencionamos anteriormente, el sistema es excitado por un tren de impulsos para señales de voz voceadas o un ruido aleatorio para señales no voceadas. Los parámetros de este modelo son:

1. Clasificación de señales de voz (voceada/no voceada)
2. Periodo Pitch para señales voceadas.
3. Parámetro de ganancia
4. Los coeficientes del filtro digital.

La señal de salida que se representa en la Figura 5.1 se muestra a continuación por una ecuación a diferencias simple:

$$S(n) = \sum_{k=1}^{p} a_k s(n-k) + Gu(n) \tag{5.2}$$

El problema básico del análisis de predicción lineal es determinar un conjunto de coeficientes de predictor, estos coeficientes deberán ser evaluados sobre segmentos cortos de una señal de voz. La metodología básica a seguir, es encontrar un conjunto de coeficientes de predicción que minimice el error cuadrático medio ver (Apéndice C) sobre un segmento corto de una señal de voz.

Durante el análisis de la predicción lineal un valor de p=12 es generalmente suficiente para señales de voz voceadas y no voceadas. Matemáticamente, el predictor lineal se describe por medio de la siguiente ecuación:

$$\tilde{x}(n) = a_1 x[n-1] + a_2 x[n-2] + \cdots + a_p x[n-p]$$

$$= \sum_{k=1}^{p} a_k x[n-k]$$

Donde $\tilde{x}[n]$ es la señal predicha en los instantes n y $a_1, a_2, \ldots, a_p$ son los coeficientes de predicción. En la Figura 5.2 se muestra una interpretación gráfica de Predicción Lineal.

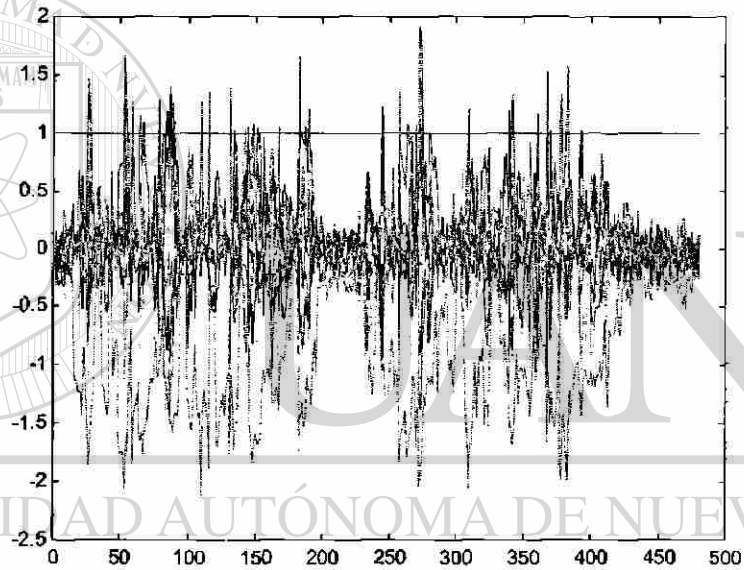


Figura 5.2: Interpretación Gráfica de Predicción Lineal

Para el cálculo de los valores de autocorrelación (ver Apéndice C) de una señal de voz, resulta imposible e impráctico calcular una sumatoria infinita para la obtención de los valores de autocorrelación, es por eso que es necesario el cálculo de los coeficientes, cada 10-30ms. Esos valores se calculan primero multiplicando la señal de voz por una función de ventana (por ejemplo, una ventana Hamming), de duración N muestras, como se muestra en la Figura 5.3. En el caso de señales voceadas, la función de autocorrelación exhibe picos correspondientes a los periodos pitch.

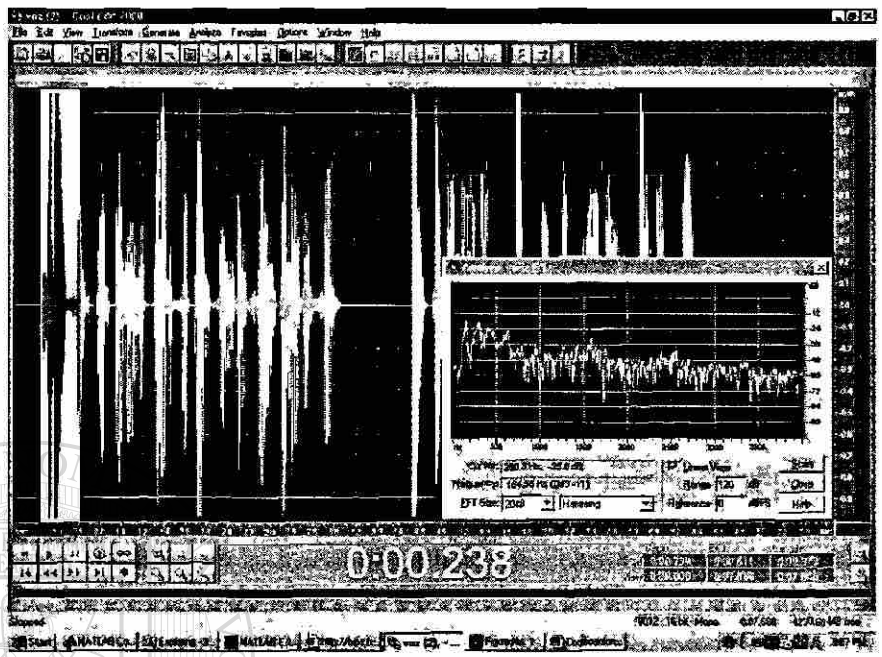


Figura 5.3: Señal Original y el cálculo de la ventana hamming para el segmento de la palabra: "Gracias".

En la Figura 5.4, se muestra la reproducción de la frase "Gracias por llamar a Fundación Televisa",el cual fue basado en el algoritmo LPC, teniendo 66,500 muestras, conviene mencionar que la grabación se realizó por un solo locutor: hombre y en condiciones de ausencia de ruido de fondo [Apéndice B].

Algo muy importante en el desarrollo de cualquier sistema de compresión es la evaluación de su desempeño; dicha evaluación se efectúa bajo dos perspectivas distintas:

1. Evaluaciones Cualitativas (o subjetivas)

2. Evaluaciones Cuantitativas (medidas numéricas de error).

Las cualitativas se centran en evaluar mediante test y cuestionarios, el desempeño de un sistema compresor auxiliándose de un auditorio de personas previamente seleccionado y entrenado. Dichas evaluaciones toman como referencia a un sistema o compresor ya en uso para asignar una calificación a un nuevo codificador o compresor

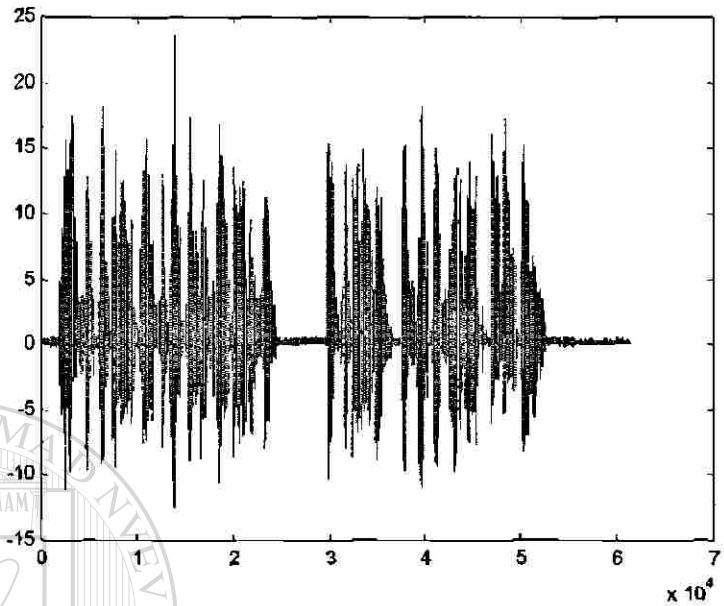


Figura 5.4: Ejemplo de una voz masculina en español

con respecto a este. Las métricas más usadas son: MOS (Mean Opinion Score), DAM (Diagnostic Acceptability Measure) y MRT (Modified Rhyme Test).

Por otro lado las evaluaciones cuantitativas se refieren al uso de expresiones matemáticas que determinan la calidad de la señal de salida (con respecto a la señal original sin comprimir). Las más usadas son: Signal to Noise Ratio (SNR) SNR Segmental Euclidean Distance.

Conclusiones : Los picos de la señal ocurren en los puntos correspondientes al cierre glotal, cuando la amplitud de la señal alcanza un máximo. En esos puntos el predictor encuentra más dificultad para modelar la señal de voz. Para un predictor ideal, la señal de error deberá consistir de un tren de impulsos a frecuencias pitch, la cual tiene un espectro plano o blanco. En el caso de señales no voceadas, la minimización del error cuadrático medio resulta una señal de error cercano al ruido blanco, la cual tiene un espectro plano.

La señal de error es fácilmente calculada (ver Apéndice C). Así la señal de error

puede ser obtenida cuando la señal de original es procesada a través de un filtro digital todos-ceros la cual es el inverso del filtro de todos polos (El filtro todos polos tiene

$$B(z) = 1H(z) = \frac{1}{A(z)}$$

únicamente potencias en z en el denominador de una función de transferencia y los ceros son las raíces del polinomio del numerador).

La principal atracción de un análisis de predicción lineal es que nos ofrece una gran exactitud y velocidad de cálculo. En conclusión, la teoría de entendimiento del método ha sido sujeta a una intensa investigación en años recientes y, como resultado es altamente avanzada y bien explicado. Basándonos en esta teoría, se han involucrado grandes aplicaciones del análisis de predicción lineal para el procesamiento de la voz. Numerosos esquemas han sido diseñados para el cálculo de todos los parámetros básicos de la voz basándonos en el análisis de predicción lineal, como por ejemplo, espectro, estimación de formantes, detección pitch, y el cálculo del pulso glotal.

La principal desventaja del análisis de predicción lineal es que en un modelo de todos polos es usado como una de la función de transferencia el tracto vocal, como quizá es esperado, este tipo de análisis es capaz de describir razonablemente bien la función de transferencia durante sonidos no nasales y sonidos como vocales. Sin embargo una función de transferencia general de tracto vocal real tiene ambos polos y ceros en la función de transferencia y además un exacto modelo análisis o síntesis de producción de voz deberá ser de tipo polos-ceros.

Los resultados obtenidos durante esta tesis no han sido todo satisfactorios, teniendo en cuenta las limitantes que se presentaron durante el desarrollo de esta tesis. Es por eso que este trabajo fue basado en el análisis LPC ya que permite representar la señal de voz y las características espectrales de forma precisa y eficiente, mediante parámetros obtenidos con cálculos sencillos, siendo la herramienta de trabajo el Matlab como lenguaje de programación, y ser adecuada para este tipo de tratamiento de señales y por su amplio uso en la industria.

Los resultados de este trabajo entrega un primer paso hacia los algoritmos de

codificación de voz. Así, las futuras investigaciones deberán orientarse a este tipo de sistemas, los cuales deberán aplicar los métodos antes mencionados. Uno de los objetivos en este campo es la codificación a velocidades de transmisión muy baja (menos de 1200 bits-por segundo). Se han desarrollado algoritmos usando representaciones espectrales con alta correlación. Este estudio fue mostrar una técnica residual de codificación (LPC) para bandas estrechas.

5.4 Investigaciones futuras

Actualmente hay aplicaciones basadas en ISDN (Integrated Service Digital Network) que también han creado un interés creciente por la codificación de Banda Ancha. Se han desarrollado técnicas de bajo retardo en los dominios temporal y frecuencia en Banda Ancha. Otro campo de investigación es el de codificación canal-fuente y su posterior aplicación a la codificación de la señal de voz. Entre las conclusiones más importantes que se obtuvieron son:

- El sistema LPC posee las características que la hacen adecuada para enfrentar el problema del envío de voz con calidad telefónica a través de una red conmutada de paquetes.
- Se necesita seguir investigando y simulando para ir a los límites de las capacidades del LPC y con ello diseñar nuevos algoritmos de compresión de más baja velocidad de bits.
- LPC puede ser la base para otros sistemas de compresión de voz.
- Este sistema se puede usar como base para implementar servicios especiales de mensajes de voz o simplemente adicionarse en aplicaciones como por ejemplo, correo de voz.

Anexo A

Acrónimos y Abreviaturas

ADC	Anolog to Digital Converter
ADPCM	Adaptive Differential Pulse Code Modulation
APC	Adaptive Predictive Coding
AR	Autoregresive (model) Bps Bips per second
CCITT	International Consultative Committee for Telegraphy and Telephony
CELP	Code-Excited Linear Prediction
CO	Central Office
CODEC	Coder/Decoder
CVSD	Continuos Variable Slope Delta Modulation
DB	Decibel
DMR	Digital Mobile Radio
DPCM	Differential Pulse Code Modulation
DSP	Digital Signal Processor
ECC	Error Control Coding
FEC	Forward Error Correction
FS	Federal Estándar
GSM	Global System for Mobile Communications
HZ	Hertz
IIRF	Infinite Impulse-Response Filter

INMARSAT	International Maritime Satellite
IPS	Instructions per Second
JDC	Japanese Digital Cellular
Kb/s	Kilobits per second
LD	Low delay
LD-CELP	Low Delay Code-Excited Linear Prediction
LP	Linear Prediction
LPC	Linear Predictive Coding
LSP	Line Spectrum Pair
Ms	millisecond
MBE	Multiband Excitation
MFLPS	Million Floting Point Operations per Second
MIPS	Million Instructions per Second
PBX	Private Branch Exchange
PCM	Pulse Code Modulation
RELP	Residual- Excited Linear Prediction
SNR	Signal to Noise Ratio
SNRseg	Segmental Signal to Noise Ratio
TFI	Time-Frequency Interpolation
VQ	Vector Quantization
VRS	Voice Response Systems
VSELP	Vector Sum-Excited Linear Prediction
X.25	A CCITT Standard

Anexo B

Espectros de señal analizados

El número de aplicaciones de la teoría de la señal es muy variado, y en algunas ocasiones se encuentran casos prácticos en los que se puede aplicar algún concepto estudiado y muchas veces para analizar un experimento resulta excesivamente complejo y solo se puede encontrar en un laboratorio muy especializado.

En la realización de esta tesis utilizamos un instrumento básico necesario para la realización de prácticas sencillas.

El material necesario para trabajar con señales de audio es una PC, una tarjeta de sonido y el software necesario para manipular este tipo de señales, en este caso utilizamos el Matlab y el cooledit (ver figura B.1)

El proceso que seguimos para trabajar con señales de audio será el siguiente:

1. Capturar una señal de audio.
2. Procesar una señal de audio utilizando el algoritmo de codificación de voz (LPC).
3. Escuchar la señal codificada.

B.1 Capturar una señal de audio.

El primer paso a realizar, fue contar con una tarjeta de sonido, micrófono y la instalación correcta de software. Después se procedió a grabar las señales de audio con

duración de 7 segundo, la frase que se grabo para procesar la señal fue: "Gracias por llamar a Fundación Televisa".

B.2 Pruebas de voz.

Seleccionamos una señal de voz, empleando los siguientes parámetros:

Con el comando Wavread reproducimos en Matlab algunos archivos de audio con el formato (.wav) y analizamos las diversas características técnicas de los archivos. En particular para el archivo voz.wav con un tamaño de 64000 bytes obtuvimos lo siguiente:

[y,Fs,Format]=Wavread

Este comando carga el archivo de audio y nos muestra la siguiente información técnica:

Y: Este vector almacena las muestras del archivo teniendo una longitud de 64000 Bytes.

Fs: Nos da la razón de muestreo, en este caso es de 8,000Hz.

Format: Es un vector de elementos con los siguientes datos.

F(1) nos indica que los datos tiene formato PCM.

F(2) nos muestra el tipo de canales, en este caso es mono.

F(3) nos da la razón de muestreo, 8000Hz.

Graficando el vector Y, obtenemos:

B.3 Pruebas de verificación

Las pruebas de verificación que se realizaron con las señales de voz se llevaran a cabo de la siguiente manera:

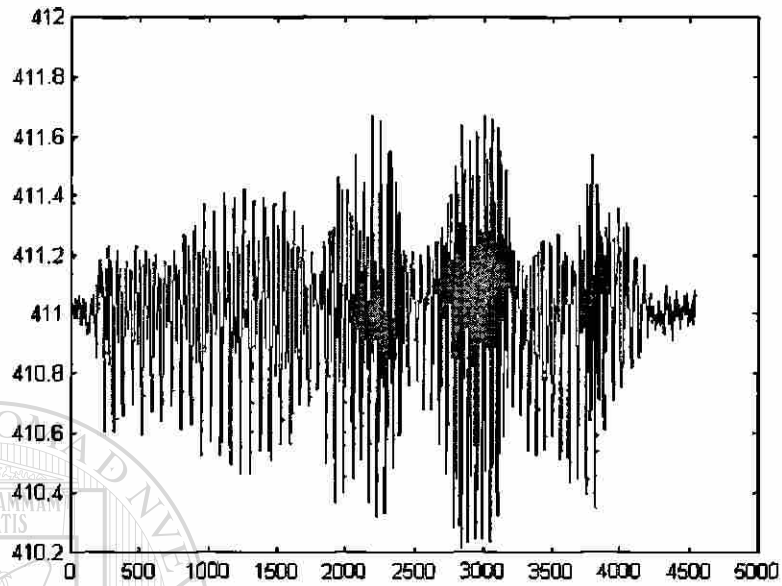


Figura B.1: Muestras

- Verificación de señales de voz.
- Verificación de la función de la ventana Hamming "
- Verificación de la función FFT

Primero se seleccionó una trama de la señal, después se procedió a trabajar con la base de datos de voz, para verificar el comportamiento de las funciones principales de procesamiento, al pasar por la función de una ventana hamming, se pudo verificar que la función estaba trabajando de una forma correcta, como se puede apreciar en la figura B.2

B.4 Verificación de la función FFT

Una vez que la ventana hamming, se aplicó correctamente a la señal de voz, se paso a través de la función FFT, la cual se puede apreciar en la figura B.3, obtuvimos el espectro de la señal.

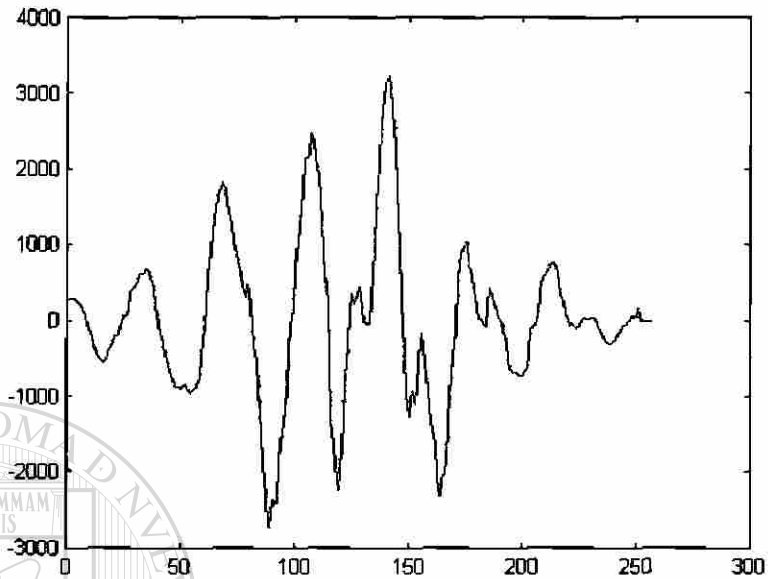


Figura B.2: Señal de voz

La función que se muestra a continuación se proceso una señal de voz, aplicando el algoritmo de codificación de voz.

% generar el siguiente fragmento de voz. x=wavread

% Se generó una matriz con vectores de 128x480 s=reshape(x,128,480);

% determinación de los coeficientes de Predicción Lineal. [A E]=lpc (s,12);

%respuesta de una señal estimada sintesis=zeros (128,480); for j=1:480; input= zeros (128,1);

%parametrización de la energía de la señal. input(1)=sqrt (E(j)); síntesis (:,j)=filter (1,A(j,:), input); end; sal=10*reshape(real(sintesis),61440,1);

%reproducción de la señal codificada. sound(sal,3200);

Anexo C

El error de predicción sobre un segmento de una señal de voz.

$$e(n) = x(n) - \tilde{x}(n) \tag{C.1}$$

Para minimizar el error cuadrático medio entre una señal de voz actual y los coeficientes de predicción puede ser determinado resolviendo un conjunto de ecuaciones lineales. Los coeficientes de predicción se calculan cada 10-30ms. El problema de predicción Lineal es determinar los coeficientes para minimizar el error cuadrático medio, sobre un número específico de muestras.

$$E = \sum_n e^2[n] = \sum_n [x[n] - \tilde{x}[n]]^2 = \sum \left[x[n] - \sum_{k=1}^{p} a_k x[n-k] \right]^2 \tag{C.2}$$

Si es minimizado para una elección apropiada de los coeficientes, entonces la derivada parcial de con respecto a cada uno de los coeficientes deberá ser cero, es decir,

$$\frac{\partial E}{\partial a_j} = -2 \sum_n x[n-j] \cdot \left[x[n] - \sum_{k=1}^{p} a_k x[n-k] \right] = 0 \tag{C.3}$$

entonces

$$\sum_{k=1}^{p} a_k \sum_n x[n-j] \cdot x[n-k] = \sum x[n] \cdot x[n-j] \qquad j = 1, 2, \ldots, p \tag{C.4}$$

La expresión de arriba nos representa un conjunto de p ecuaciones lineales para las incógnitas. Además es posible encontrar una solución a un sistema de ecuaciones si estas ecuaciones son lineales. Afortunadamente existen dos métodos para encontrar la solución a estos sistemas de ecuaciones, conocidos como: método de autocorrelación y autocovarianza.

Los límites de la sumatoria de las expresiones $\sum x[n-j]\cdot x[n-k]$ y $\sum x[n]\cdot x[n-j]$ en la ecuación 4 no los hemos especificado. Supongamos que la señal es estacionaria con energía finita, por supuesto no es para una señal de voz, el rango de la sumatoria es de $-\alpha$ hasta $+\alpha$, $x_n]$ empieza a ser definido como cero para $n<0$, (causal) entonces:

$$\begin{aligned}\sum_{n=-\alpha}^{\alpha} x[n-j]\cdot x[n-k] &= \sum_{n=-\alpha}^{\alpha} x[n-j+1]\cdot x[n-k+1] \\ &= \sum_{n=-\alpha}^{\alpha} x[n]\cdot x[n+j-k]\end{aligned} \tag{C.5}$$

Además el sistema de ecuaciones puede ser escrito como:

$$\sum_{k=1}^{p} a_k \sum_{n=-\alpha}^{\alpha} x[n]\cdot x[n+j-k] = \sum_{n=-\alpha}^{\alpha} x[n]\cdot x[n-j], \qquad j=1,2,\ldots,p \tag{C.6}$$

Los coeficientes del sistema de ecuaciones muestran los valores de autocorrelación de una señal de voz para un tiempo específico. Si $R(k)$ es una matriz cuyos valores de autocorrelación para k muestras, es:

$$R(k) = \sum_{n=-\alpha}^{\alpha} x[n]\cdot x[n+k] \tag{C.7}$$

El sistema de ecuaciones lo podemos representar como:

$$\begin{bmatrix} R(0) & R(1) & R(2) & \cdots & R(p-1) \\ R(1) & R(0) & R(1) & \cdots & R(p-2) \\ R(2) & R(1) & R(0) & \cdots & R(p-1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R(p-1) & R(p-2) & R(p-3) & \cdots & R(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \\ R(3) \\ \vdots \\ R(p) \end{bmatrix} \tag{C.8}$$

Esta es una matriz simétrica donde todos los elementos de la diagonal son iguales. Esto es conocido como matriz Toplitz y es considerado como un método muy eficiente debido a Durbin y Levinson [16] existente para la solución de este tipo de ecuaciones ya que no requiere mucho esfuerzo computacional y es generalmente necesario para la solución de un sistema de ecuaciones lineales. La señal de voz no es conocida todo el tiempo es imposible e impráctico calcular una sumatoria infinita requerida para obtener los valores de autocorrelación, es por eso que es necesario el cálculo de los coeficientes a_k , cada 10-30ms. Esos valores se calculan primero multiplicando la señal de voz $x[n]$ por una función de ventana (por ejemplo, una ventana Hanning), N muestras de duración. Los valores de autocorrelación se calculan de la siguiente forma:

$$R(k) = \sum_{n=0}^{N-1} \{W[n] \cdot x[n]\} \cdot \{W[n+k] \cdot x[n+k]\}, \qquad k = 0, 1, 2, \ldots, p \qquad \text{(C.9)}$$

En el análisis de una señal de voz, el método de cálculo de los coeficientes de predicción se le conoce como método de autocorrelación. Típicamente una ventana de duración 20-30ms (200 muestras en una relación 10Khz) es usada en un rango de una trama de 10-20mseg. La función de la ventana consiste en reducir el error de predicción en el comienzo y fin de un segmento. Los errores de predición por lo general surgen en el comienzo de un intervalo de $(0 \leq n \leq p-1)$ y en el fin de intervalo $N \leq n \leq N+p-1$.

Función de Autocorrelación

Para el calculo de valores de una señal estacionaria, cuyo tiempo es desfasado en muestras, se define de la siguiente manera:

$$R(k) = \sum_{n=-\alpha}^{\alpha} x[n] \cdot x[n+k] \qquad \text{(C.10)}$$

Si consideraremos intervalos finitos de una señal. La función de autocorrelación deberá ser apropiado para calcular segmentos (tramas) sucesivas y multiplicar esta señal obtenida con una ventana, de ancho, la cual tiene un intervalo de $(0, N-1)$ y cero fuera de ese intervalo. La función de autocorrelación aplicado a un segmento se define

de la siguiente manera:

$$R_m(k) = \sum_{n=0}^{N-1} \{x_n] \cdot W[n]\} \cdot \{x[n+k] \cdot W[n+k]\} \tag{C.11}$$

Anexo D

Reconstrucción de una señal de voz

De acuerdo a lo anterior podemos decir que el predictor lineal representado es esencialmente una técnica de codificación de onda en el dominio del tiempo que quizá permite 100-200 muestras de una señal de voz para representar 10-15 coeficientes. Estos coeficientes pueden ser usados para calcular la respuesta de tracto vocal. La señal de error $e[n]$ se calcula fácilmente usando los coeficientes de predictor :

$$\begin{aligned} e[n] &= x[n] - \sum_{k=1}^{p} a_k x[n-k] \\ &= x[n] - a_1 x[n-1] - a_2 x[n-2] - \cdots - a_p x[n-p] \end{aligned} \tag{D.1}$$

Si la señal de error se conoce y los coeficientes de predicción lineal $\{a_i, i = 1, 2, \ldots, p\}$, es posible reconstruir la señal original exactamente de la señal predicha $\tilde{x}[n]$, es decir,

$$x[n] = e[n] + \tilde{x}[n] = e[n] + \sum_{k=1}^{p} a_k x[n-k] \tag{D.2}$$

Tomando la transformada Z , obtenemos:

$$\begin{aligned} X(z) &= E(z) + \left[\sum_{k=1}^{p} a_k z^{-k}\right] X(z) \\ X(z) &= E(z) / \left(1 - \sum_{k=1}^{p} a_k z^{-k}\right) = H(z) \cdot E(z) \end{aligned} \tag{D.3}$$

Donde $X(z)$ y $E(z)$ son las transformadas z de $x(n)$ y $e(n)$ respectivamente y

$$H(z) = 1/\left(1 - \sum_{k=1}^{p} a_k z^{-k}\right) \tag{D.4}$$

Es la función de transferencia de un sistema digital o filtro la cual contiene únicamente potencias de z en el denominador y por esa razón que uno se refiere uno a sistema de todos-polos. (Los polos son las raíces del polinomio del denominador en z). La ecuación (D.3) nos muestra que una señal de voz $x[n]$ puede ser vista como la salida de un filtro todos-polos cuando la entrada es una señal de error $e[n]$ además de que el filtro $H(z)$ nos muestra la respuesta del tracto vocal, $e[n]$ nos indica la excitación del tracto vocal. Un cálculo de la envolvente del tracto vocal, $H(z)$ puede ser obtenida colocando $z = e^{j\omega t}$ en la función de transferencia de $H(z)$ del predictor lineal de todos polos, es decir,

$$|H(\omega)| = \left|1/\left(1 - \sum_{k=1}^{p} a_k e^{-j\omega kT}\right)\right| \tag{D.5}$$

El espectro se obtiene evaluando $|H(\omega)|$,considerando varios valores de ω, como se muestra en la ecuación (D.5).

La señal de error $e[n]$ se calcula fácilmente usando los coeficientes de predicción lineal, de la ecuación (D.4), tenemos que:

$$E(z) = \frac{1}{H(z)} \cdot X(z) = A(z) \cdot X(z) \tag{D.6}$$

Donde

$$A(z) = \frac{1}{H(z)} = \left\{1 - \sum_{k=1}^{p} a_k z^{-k}\right\} \tag{D.7}$$

Bibliografía

[1] Asertek. Redes de conmutación de paquetes: X.25 y frame relay. Texto de apoyo.

[2] Bishnu S. Atal, Vladimir Cuperman, and Allen Gersho, editors. *Speech and Audio Coding for Wireless and Network Applications* Kluwer Academic Publishers, 1993.

[3] John C. Bellamy. *Digital Telephony.* Wiley-Interscience, 1990.

[4] Andes Buzo, Augustine H. Gray, RObert M. Gray, and John D. Markel. Speech coding based upon vector quantization. *IEEE Transactions on Acoustic, Speech and Signal Processing,* Octubre 1980.

[5] David H. Crawford and Emmanuel Roy. *Techniques for Real-Time DSP Implementation of Speech Coding Algorithms* Addison-Wesley, 1986.

[6] John R. Deller, John G. Proakis, and John H. L. Hansen. *Discrete-time processing of speech signals.* McMillan Publishing Company, 1993.

[7] Bob Edgar. *PC Telephony.* Flatiron Publishing, 1995.

[8] Harvey Fletcher. *Speech and Hearing in Communication* Acoustical Society of America Publications, 1995.

[9] Juan L. Fuentes. *Gramática Moderna de la lengua española* LIMUSA, 1999.

[10] Sadaoki Furui. *Digital Speech Processing, Synthesis and Recognition* Marcel Dekker, 1989.

[11] Armando González González. C & d united / global one.

[12] A. Nejat Ince. *Speech coding, synthesis of speech signals* Kluwer Academic Publishers, 1992.

[13] Bernhard E. Keiser and Eugene Strange. *Digital Telephony and Network Ingration* Addison-Wesley, 1984.

[14] W. Bastiaan Klejin, Daniel J. Krasinski, and Richard H. Ketchum. Fast methods for the celp speech coding algorithm. *IEEE Transactions on Acoustic, Speech and Signal Processing*, 38, Agosto.

[15] Motorola. Voice technologies for IP and frame relay networks.

[16] F. J. Owens. *Signal Processing of Speech* McGraw Hill Text, 1993.

[17] K. K. Paliwal and T. Svendsen. A study of three coders (sub-band, relp, and mpe) for speech with additive white noise. In *IEEE International Conference on Acoustics, Speech, and Signal Processing* 1992.

[18] Gordon E. Pelton. *Voice Processing.* McGraw Hill Text, 1993.

[19] Lawrence R. Rabiner and Ronald W. Schafer. *Digital processing of speech signals* Prentice Hall, 1978.

[20] John D. Spragins, Krzysztof Pawlikowski, and Joseph Hammond. *Telecommunications: Protocols and Design* Pearson Education POD, 1991.

[21] Charles W. Therrien. *Discrete Random Signals and Statistical Signal Processing* Prentice Hall, 1992.

[22] F. A. Westall, R. D. Johnston, A. V. Lewis, and Denis Johnston, editors. *Speech Technology for Telecommunications* Chapman and Hall, 1993.

[23] Robert G. Winch. *Telecommunication Transmission System* McGraw Hill, 1990.

[24] Richard L. Zinser and Steven R. Koch. Celp coding at 4.0 kb/sec and below: Improvement to FS-1016. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1992.