



UNIVERSIDAD TECNOLÓGICA DE LA MIXTECA

Codificador de Voz a 16 Kbps
Utilizando la Técnica
RPE-LTP

Tesis Profesional
Que para Obtener el Título de
Ingeniero en Electrónica
presenta :

Alberto Antonio Vázquez Cruz

Trabajo dirigido por :
M.C. José Antonio Moreno Espinosa

Acatlima, Huajuapán de León, Oaxaca, Marzo del '97.

Hay un hecho que domina toda la psicología del lenguaje :

El pensamiento excede a la expresión. La palabra, con ser de todos los medios el más afortunado, no constituye sino un esfuerzo, nunca satisfecho por hacer del sonido el traductor fidelísimo de nuestra vida mental. Un esfuerzo , hemos dicho, y nada más. Pero gracias a él, tiene el lenguaje la movilidad y la inquietud de las cosas que viven.

Aníbal Ponce.

Tesis presentada el 20 de Marzo de 1997
ante los siguientes sinodales :

Ing. Hugo Suárez Onofre

M. C. Hiram Ochoa Arriaga

M. C. Enrique Guzmán Ramírez

Asesor

M.C. José Antonio Moreno Espinosa

DEDICATORIAS

Tesis dedicada especialmente a :

Guadalupe quien me ha enseñado a valorar que un triunfo sin esfuerzo, no es un triunfo.

Mis padres Teresa y Juan por su comprensión y ayuda durante todos estos años, y por haberme dado lo más valioso de ellos, la vida.

Mis hermanos

Silvia

Alejandro

Carlos

Salvador

Javier

Arturo

Quienes siempre me han brindado su apoyo y sugerencias para llevar a cabo de la manera más conveniente mis inquietudes.

Y los pequeños más traviosos e inquietos, Miguel Angel y Fernando.

Merced Cruz Cortés (q.e.p.d.) por su cariño y dedicación.

U. T. M. 6962

Alberto

AGRADECIMIENTOS

Deseo dar mi agradecimiento a las siguientes personas :

M. C. José Antonio Moreno Espinosa por su tiempo, interés y conocimiento que aportó como asesor de la tesis.

Ing. Hugo Suárez Onofre quien leyó el borrador preliminar de la tesis y proporcionó numerosas sugerencias y puntos de vista, por lo que todos sus comentarios fueron de mucha utilidad.

Miguel Angel Ramírez Jiménez compañero y amigo de siempre, quien participo en la implementación de tareas y en el desarrollo de la tesis.

SINODALES

Ing. Hugo Suárez Onofre

M.C. Hiram Ochoa Arriaga

M. C. Enrique Guzmán Ramírez

A quienes deseo agradecer su participación como miembros del jurado y sus acertadas observaciones.

RESUMEN

Se presenta un codificador con predicción de término largo excitado por pulsos regulares (RPE-LTP) a 16 Kbps, el cual procesa segmentos de voz de 20 ms. que extrae de un archivo previamente grabado con extensión [wav] (similar a la voz de la línea telefónica comercial) y su formato es PCM con 16 bits por muestra, frecuencia de muestreo de 8 KHz, y velocidad de transmisión de 64 Kbps.

Antes de procesar cada una de las tramas de 20 ms. se extraen los bytes de la cabecera del archivo [wav], y luego se lee el área de datos con segmentos de 160 muestras.

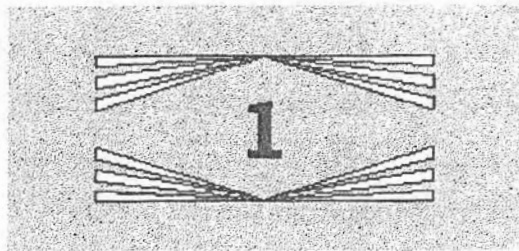
Después, por medio de los diferentes procesos de análisis en el codec RPE-LTP (proceso de autocorrelación, algoritmo de Schur, y filtro de análisis, entre otros) se obtiene a la salida las tramas codificadas, por lo que se transmiten al decoder a una velocidad de 16 Kbps.

U. T. M 6902

CONTENIDO

1.0 INTRODUCCION	1
2.0 ANTECEDENTES	
2.1 Introducción	7
2.2 Descripción del aparato fonador	10
2.3 Clasificación de los sonidos	12
2.4 Modelos numéricos de la voz	16
3.0 CODIFICADORES DE FORMA DE ONDA	
3.1 Introducción	27
3.2 Modulación por Pulsos Codificados (PCM)	28
3.3 Codificación Diferencial	33
3.3.1 Codificador Diferencial PCM (DPCM)	35
3.3.2 Modulación Delta (MD)	39
3.4 Codificación Adaptiva	44
3.4.1 Modulación Delta Adaptivo (ADM)	47
3.4.2 PCM Diferencial Adaptivo (ADPCM)	50
4.0 TECNICAS BASADAS EN LA PREDICCIÓN LINEAL	
4.1 Introducción	55
4.2 Codificador de Predicción Lineal (LPC)	56
4.3 Excitación Multi-pulso(MPE) y MPE-LTP	63
4.4 Excitación por Pulsos Regulares (RPE)	70
4.5 Excitación Codificada (CELP)	79

5.0	IMPLEMENTACION DEL ANALIZADOR RPE-LTP	
5.1	Esquema general del codificador	85
5.2	Funcionamiento general del codificador	88
5.2.1	Escalamiento	88
5.2.2	Autocorrelación	90
5.2.3	Algoritmo de Schur	91
5.2.4	Conversión Parcor-LAR	91
5.2.5	Cuantizador/Codificador	92
5.2.6	Filtro de análisis de término corto	93
5.2.7	Decodificación / interpolación de parámetros LAR	93
5.2.8	Conversión LAR-Parcor	93
5.2.9	Filtro de predicción de término largo	94
5.2.10	Sub-segmentación y selección de rejilla	96
5.2.11	Cuantización de la secuencia RPE	99
5.2.12	Cuantización del pulso máximo RPE	99
5.2.13	Formato de la trama codificada	99
5.2.14	Protección contra errores	109
5.3	Pseudocódigo de los principales módulos	110
6.0	CONCLUSIONES Y PERSPECTIVAS	125
	APENDICE	137
	GLOSARIO TECNICO	143
	REFERENCIAS	147



INTRODUCCION

En el medio ambiente de las comunicaciones digitales actuales, se ha dado una nueva revolución en la comunicación del mensaje hablado, en la que *la calidad y la fiabilidad* de los nuevos servicios son radicalmente mejores. El punto decisivo está en que las señales digitalizadas pueden ser procesadas de tal forma que se aumenta el abanico de posibilidades de comunicación y almacenamiento de información y prestación de servicios completamente novedosos. Dentro de ellas destaca el tratamiento digital de la señal de voz, como un área en la que se centran importantes grupos de estudio y en el que se investigan y desarrollan herramientas cada vez más precisas, enfocadas a reducir la velocidad de transmisión a tasas menores de 16 Kbps.

Justificación y Motivación

La mayor parte de las teorías matemáticas que soportan los modelos del habla, en los cuales se basan las nuevas técnicas, fueron desarrollados a lo largo de las últimas décadas. Si bien, es ahora cuando cobran importancia, ya que sus exigencias de cálculo habían demorado su aplicación real hasta la aparición de los potentes microprocesadores de procesamiento digital de señales que se ofrecen en el mercado.

Los métodos de codificación de voz se han estandarizado a 64 y 32 Kbps, y los codificadores en estas tasas de transmisión, son los que se usan en las redes telefónicas, y todavía no hay estándares para transmitir velocidades más bajas. Pero es bien conocido que la codificación PCM a 64 Kbps de la señal de voz no elimina mucha redundancia de la que existe en esta señal. Debido a ello han surgido una gran variedad de técnicas que permiten codificar la señal de voz a tasas menores. Estas técnicas se pueden dividir en dos tipos : *codificación de forma de onda y codificación de fuente.*

Una técnica que actualmente ha logrado buena aceptación es la ADPCM a 32 Kbps e incluso el CCITT (Comité Consultivo Internacional de Telefonía y Telegrafía) ha emitido^{<1>} algunas recomendaciones al respecto. La tendencia es reducir aún más la tasa de transmisión para lo cual se necesita el estudio de técnicas más eficientes cuya tarea principal sea preservar la calidad de voz.

Las redundancias introducidas en la señal de voz durante el proceso de producción de la voz, hace posible codificarla en bajos porcentajes de bits. Sin embargo, nuestro sistema auditivo no es

igualmente sensible a distorsiones en diferentes frecuencias, lo que a su vez origina que tenga un rango limitado.

Las técnicas de codificación de voz toman ventajas de éstas propiedades para reducir la tasa de bits para la transmisión de voz.

Como una ventaja importante, el porcentaje de bits de transmisión a 16 Kbps es idóneo para una gran variedad de aplicaciones.

Dentro de los campos de aplicación más importantes destacan los esfuerzos dedicados a síntesis (conversión de texto a voz), reconocimiento (de locutores y fonemas) y codificación. Otras posibilidades de aplicación están en torno a la radiocomunicación móvil, transmisión de voz en redes de conmutación de paquetes o encriptado de voz, entre otras. Para todo ello se hace necesario reducir el ancho de banda convencional en el envío de voz, memoria, canales de transmisión, etc.

Se ha observado en base a lo anterior, que la codificación de voz a bajo régimen, con calidad de voz próxima a la telefónica digital comercial, es una nueva necesidad para los servicios telemáticos, por lo que los bajos porcentajes de bits en la transmisión de voz juegan un papel muy importante, ya que proporcionan nuevas *capacidades* en sistemas de comunicación futuras.

Objetivo

Es por eso que el objetivo de éste trabajo es implementar un sistema de codificación de voz para reducir la velocidad de transmisión a 16 Kbps, con características que le permitan ser compatible con los

sistemas de transmisión de telefonía digital comercial, almacenar mensajes y no degradar la calidad subjetiva de la señal de voz.

De las aplicaciones que tiene la codificación de voz y que se mencionaron en párrafos anteriores puede observarse un gran enfoque a telefonía digital, de aquí la necesidad de que el sistema propuesto posea alto grado de compatibilidad, lo cual se refiere a que éste sistema pueda conectarse a diferentes equipos de comunicación como en el caso de :

- ♦ *Aparatos de usuario.*
- ♦ *Sistemas de transmisión.*
- ♦ *Sistemas de conmutación.*
- ♦ *Planeación y operación de redes bajo diferentes protocolos de señalización.*

En éste ambiente no serán raros los casos de transcodificación o de cambio en la velocidad de transmisión, ni de errores debido a la falta de sincronización.

Contenido de la tesis

Para explicar el funcionamiento y realización de éste sistema, la tesis se ha dividido como sigue :

En el capítulo dos se da una visión general de la producción de voz en el tracto vocal y la clasificación de sonidos que en base a el se generan, los cuales se utilizan para excitar el filtro de síntesis. También en este apartado se establecen diferentes modelos numéricos a partir de las cavidades fonadoras del tracto vocal.

En el capítulo tres se revisan brevemente algunas técnicas utilizadas en la codificación de voz, desde hace algunos años para resolver el problema de transmisión digital de la voz con el mínimo de información, pero cuidando la calidad de la señal en lo que se refiere a la inteligibilidad e identificación del interlocutor entre otras cosas.

De la investigación y desarrollo de tales técnicas, surgen teorías interesadas no en la codificación de la forma de onda de la señal de voz, sino en la codificación de la fuente que la genera. Este nuevo concepto, revoluciona la eficiencia en la codificación de la voz. Originando diversos métodos de acuerdo a las necesidades de transmisión.

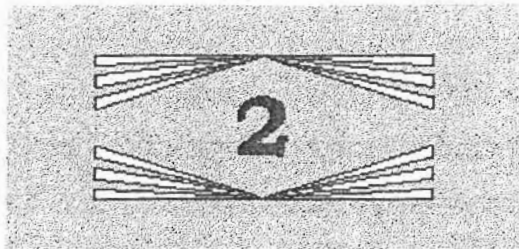
En términos generales, si se desea transmitir a muy bajas velocidades, debe utilizarse un codificador de fuente y la calidad que se espera es tal que no se puede reconocer la identidad del interlocutor y en ocasiones es difícil entender el mensaje.

Por otra parte, si lo que se desea es una buena calidad, puede emplearse desde codificadores de forma de onda, con altas tasas de transmisión y baja complejidad, hasta codificadores híbridos con velocidades medias y relativamente alta complejidad.

Algunos de estos métodos se describen en el capítulo cuatro, así como los fundamentos de la teoría de predicción lineal.

Como resultado de una investigación de las técnicas de codificación de voz a 16 Kbps, en el capítulo cinco se resume la arquitectura del codificador seleccionado.

Finalmente en el capítulo seis se exponen las perspectivas y conclusiones del proyecto.



ANTECEDENTES

2.1 Introducción

La facultad que tenemos los seres humanos para emitir sonidos^{<2>} que conforman la voz (series de sonidos articulados), supone un encadenamiento de actos de diversas índoles :

- ♦ **Psicológicos.** La implicación sensorial que tiene la creación de una imagen acústica-motriz.
- ♦ **Fisiológicos.** Los movimientos orgánicos necesarios para producirlos.
- ♦ **Físicos.** Las vibraciones atmosféricas causadas por dichos sonidos.

El proceso del habla distingue tres niveles :

- ♦ **Nivel lingüístico** : Comprende el uso de la palabra como la manifestación de las ideas y exteriorizaciones de los sentidos, estableciendo un método de clasificación y manejo de los sonidos.
- ♦ **Nivel fisiológico** : Abarca la generación de voz, analizando las funciones orgánicas que dan origen al sonido articulado.
- ♦ **Nivel físico** : Es el fenómeno en sí, considerando las características acústicas del sonido en general.

Entre las cualidades físicas del sonido se encuentran :

a) **Tono** : Que es la altura musical del sonido y depende directamente de la frecuencia de las vibraciones, o sea, el número de oscilaciones completas por unidad de tiempo.

b) **Intensidad** : Grado de fuerza con la que se produce el sonido y se manifiesta como la amplitud de la vibración que físicamente guarda relación con la energía de impulso que produjo la oscilación y decrece proporcionalmente con el cuadrado de la distancia.

c) **Timbre** : Forma física de la oscilación debido a la propiedad de superposición de ondas, que origina la formación de ondas complejas (no senoidales puras) cuya característica es la intervención simultánea de varias ondas secundarias (sobretonos o hipertonos que acompañan al tono fundamental y se diferencian de éste por su frecuencia (cabe aclarar que si los sobretonos son múltiplos del tono fundamental, reciben el nombre de armónicos), intensidad relativa y fase. La cantidad y existencia de sobretonos

depende de la naturaleza y forma del cuerpo sonoro, así como las resonancias u obstáculos que encuentre la onda a su paso.

d) **Cantidad** : Tiempo de vida del sonido antes de extinguirse y se relaciona con el número y duración de los impulsos de excitación, así mismo, el efecto conjunto de todos los anteriores, otorga una característica de sonido intrínseca a cada objeto y medio de propagación según su naturaleza, provocando distintas impresiones sensoriales al emitirse en igualdad de circunstancias.

La voz humana es el resultado de un complejo proceso orgánico que se inicia en la corteza cerebral y por el cual el hombre es capaz de emitir sonidos por la acción conjunta de varios órganos cuyos mecanismos permiten adoptar todas las cualidades físicas de los sonidos que articulados forman la voz.

La generación de voz en el ser humano se realiza en tres etapas :

- ♦ **Respiración** : Durante la respiración, el aire aspirado y contenido en los pulmones, sale de ellos por los bronquios y de la tráquea a la laringe, obligado por la presión del diafragma y la reducción de la cavidad torácica, formando una columna de aire.
- ♦ **Fonación** : Cuando la columna de aire pasa de la traquea a la laringe, donde se encuentran las cuerdas vocales (músculos gemelos y elásticos a modo de pliegues), que al tenderse o aflojarse, forman una abertura triangular (glotis) que se abre o cierra y la presión de aire obliga a vibrar, poniendo de esta forma a vibrar consigo la columna de aire.
- ♦ **Articulación** : La columna de aire en vibración sale de la laringe por la faringe a la boca, donde el movimiento de labios, mandíbula y

lengua modifican en conjunto la forma y espacio de las cavidades faríngea, nasal y bucal, haciendo resonadores, dificultando o bien desviando la salida del aire. Finalmente se produce el sonido articulado con la salida del aire al exterior.

2.2 Descripción del aparato fonador

Un esquema general del aparato fonador humano^[3] se muestra en la figura 2.1, en el se aprecia que dependiendo de la tensión muscular de los tendones de las cuerdas vocales, pueden dejar pasar el aire, como en el caso de la respiración, estar ligeramente más cerrados, como en los sonidos fricativos, o estar vibrando, como en los sonidos sonoros o vocálicos.

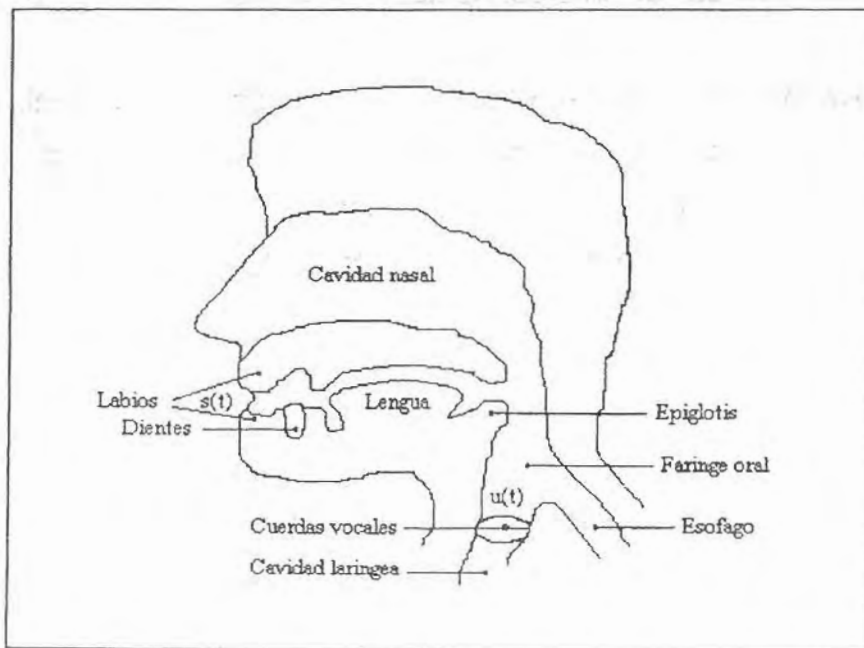


Figura 2.1 Esquema de los órganos fonadores en sonidos sonoros.

En el último caso la tensión esta graduada de tal forma que el flujo proveniente de los pulmones los force a abrirse, lo que a su vez hace disminuir la presión de la cavidad laríngea y por consiguiente las cuerdas vocales vuelven a cerrarse.

Con las cuerdas vocales vibrando la presión acústica crece hasta un determinado valor para a continuación volver a la presión atmosférica media.

La señal $u(t)$ de la figura 2.2 describe las variaciones de presión, en un intervalo sonoro la señal está formada por secuencia de pulsos sensiblemente triangulares. El periodo de la señal $u(t)$ se le denomina *periodo fundamental*, aunque no es muy adecuado, ya que cambia en forma gradual en las zonas sonoras. El filtrado de ésta señal por las cavidades fonadoras produce la señal de voz $s(t)$.

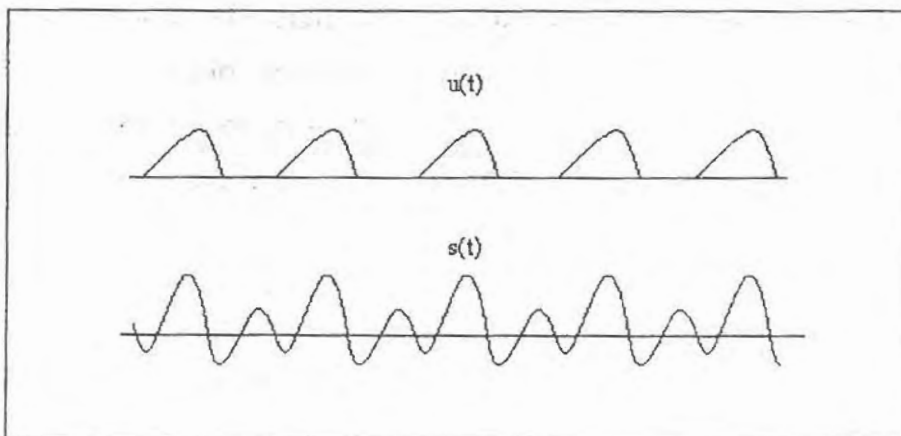


Figura 2.2 Variaciones de presión del aparato fonador con un sonido sonoro.

Después del paso de aire a través de la apertura de las cuerdas vocales el flujo del aire puede ir directamente a la boca, a la

cavidad nasal o a ambas. Las distintas alternativas dependen de la posición de la úvula o velo paladar.

En un intervalo corto de aproximadamente 20 ms. el aparato fonador puede considerarse estático. En ésta situación la señal de voz puede modelarse como la salida de un filtro acústico aplicando a su entrada la señal apropiada. Este filtro es aproximadamente lineal invariante durante ese intervalo. Las propiedades del filtrado pueden cambiarse variando la posición de la mandíbula, de la lengua y de los labios. Diferentes funciones de transferencia de éste filtro corresponden a diferentes timbres de voz.

Cada sonido se produce con una determinada configuración del aparato fonador. Así, la cavidad oronasal tomará una forma determinada que depende de la posición de la lengua, la mandíbula, los labios, la posición de la úvula y las condiciones de las cuerdas vocales, en algunos casos se encontrará vibrando con un determinado periodo o habrá una excitación fricativa. Diferentes sonidos se producirán mediante configuraciones distintas del aparato fonador. La secuencia de sonidos utilizada para comunicarnos depende de la lengua hablada, en nuestro caso el Español.

2.3 Clasificación de los sonidos

Los sonidos elementales del lenguaje articulado están constituidos por *fonemas* (unidades diferenciadoras indivisibles y abstractas), que son modelos o tipos ideales de sonidos que no pueden descomponerse en unidades menores y poseen cualidades que los distinguen de los demás, siendo portadores de una intención significativa diferente.

Las múltiples variantes de sonidos (alófonos) que provocan los fonemas vecinos al momento de articularse son :

- ♦ Sonorización o ensordecimiento. Por la tensión muscular ejercida.
- ♦ Nasalidad. Por la salida simultánea del aire por la boca y la nariz.
- ♦ Transiciones entre vocales y consonantes, individualmente o entre si. Por la abertura o modificación articulatorias.

Las variantes alteran en conjunto la pronunciación pero no su significado.

El material fonético del Español se compone de veintinueve sonidos básicos (cinco vocales, dos semivocales, dos semiconsonantes y veinte consonantes) de las cuales solo veintidós son fonemas (cinco vocales y diecisiete consonantes) y los siete restantes son alófonos siendo su representación aproximada el alfabeto ortográfico de treinta grafías. Ocho de ellos no se utilizan (en función de su significado) porque representan :

- ♦ Sonidos mudos {h}.
- ♦ Sonidos compuestos {x}.
- ♦ Sonidos Repetidos {c,q}.
- ♦ Alófonos {l,v,z}.

o fueron introducidos únicamente con propósitos de simulación con otras lenguas {w} y sólo forman parte del estilo en la escritura como representación de una lengua, siendo su valor únicamente estético y no funcional.

La siguiente clasificación de los sonidos se hace en base a la configuración del aparato fonador y está dada por :

1. **La acción de las cuerdas vocales :** En las cuales los sonidos pueden ser sonoros o fricativos. Los sonoros son los que se producen con las cuerdas vibrando, la señal de excitación es localmente periódica y el paso del aire puede ser a través de la cavidad oral, de la nasal o de ambas. Los fricativos son los que se producen con las cuerdas vocales abiertas.

2. **La posición del velo del paladar :**

a) **Los sonidos orales o bucales.** Son producidos cuando el velo está cerrando la entrada o la cavidad nasal. El aire sale únicamente a través de la cavidad oral. Ejemplo de estos sonidos son {a}, {p}, {b}.

b) **Los sonidos nasales.** Se producen cuando el velo cierra la entrada a la cavidad oral. El aire solo sale a través de la cavidad nasal. Como ejemplo de estos sonidos tenemos la {m}. se observa que la {m} y la {b} se diferencian por la posición del velo del paladar, la configuración de los demás órganos fonadores es prácticamente idéntica.

c) **Sonidos oronasales.** Tanto el conducto nasal como el vocal están abiertos. Un ejemplo de éste sonido lo encontramos en la vocal {e} cuando se pronuncia {humanam/e/nte}.

3. **El lugar de articulación :** El lugar de articulación es aquel en donde la lengua se aproxima más al paladar. De ésta forma se tienen sonidos labiales, labiodentales, dentales, interdientales, alveolares, paladares y velares. En la figura 2.3 se pueden apreciar las diferentes zonas bucales.

4. **El Modo de articulación :** Indica el grado de resistencia que imponen las cavidades oronasales al paso del aire.

Algunos sonidos son :

- ♦ Abiertos. Como las vocales.
- ♦ Medio cerrados. Como {s}, {f}, {x}.
- ♦ Cerrados o interruptos. Como {p}, {t}, {k}, {b}.

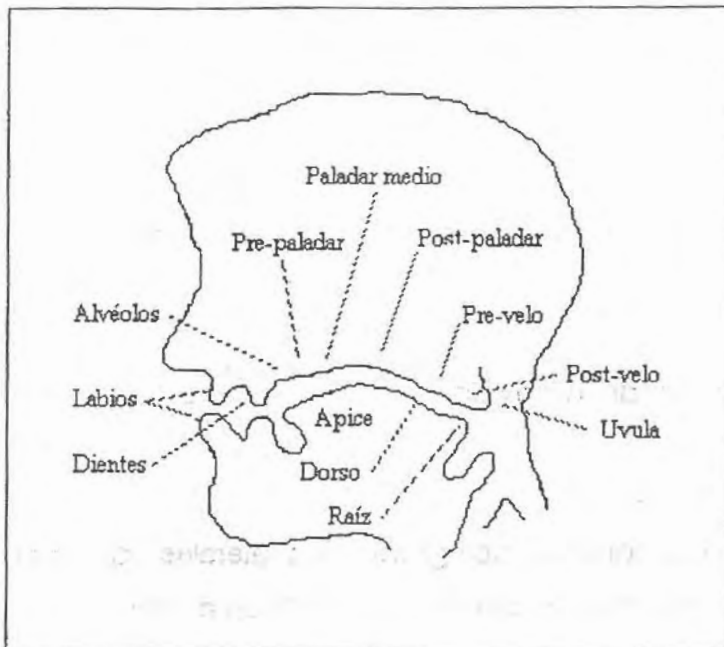


Figura 2.3 Zonas bucales.

5. Las vocales y consonantes :

a) Las vocales (a, e, i, o, u) son sonidos que se pronuncian con configuraciones más abiertas de la cavidad oral. Si la lengua se aproxima mucho al paladar, se tienen las vocales de pequeña abertura o cerradas tales como {i}, {u}. Si la lengua se separa más se tiene la vocal abierta {a}. Estos sonidos se producen con las cuerdas vocales vibrando.

Por otra parte, las consonantes se clasifican en :

- ♦ **Oclusivas.** Cuando se da un cierre completo de los órganos articulatorios {p},{t},{k}.
- ♦ **Fricativas.** Cuando se da un estrechamiento en los órganos articulatorios produciéndose turbulencia , {f},{s},{x}.
- ♦ **Africadas o semioclusivas.** Cuando al cierre completo de dos órganos articulatorios sucede una pequeña abertura. Una consonante africada consta de dos fases, en la primera hay una oclusión y en la segunda una fricación. Ejemplo la {y} de {con/y/uge}.
- ♦ **Nasales.** Cuando la cavidad bucal está cerrada y la nasal abierta. {m},{n}.
- ♦ **Líquidas.** Comprende dos grupos, las **laterales**, tal como la {}, en cuya pronunciación el aire sale por un lado o ambos de la cavidad oral, y las **vibrantes** cuya característica es una o más vibraciones del ápice de la lengua, como un ejemplo de éste tipo se tiene a {r}.

En la producción de un sonido consonántico sonoro, las cuerdas vocales vibran a una frecuencia inferior que en el caso vocálico.

2.4 Modelos numéricos de la señal de voz

Esencialmente cualquier aparato o máquina puede producir voz sintética o artificial, si es capaz de simular o duplicar efectivamente las propiedades acústicas del sistema fonatorio humano y cuenta con alguna forma de controlarlas (figura 2.4).

Por motivos prácticos se requieren modelos fáciles de producción de voz para su implementación en una computadora. Estos modelos numéricos tienen, además, una relación muy interesante y a la vez importante con los modelos de tubos.

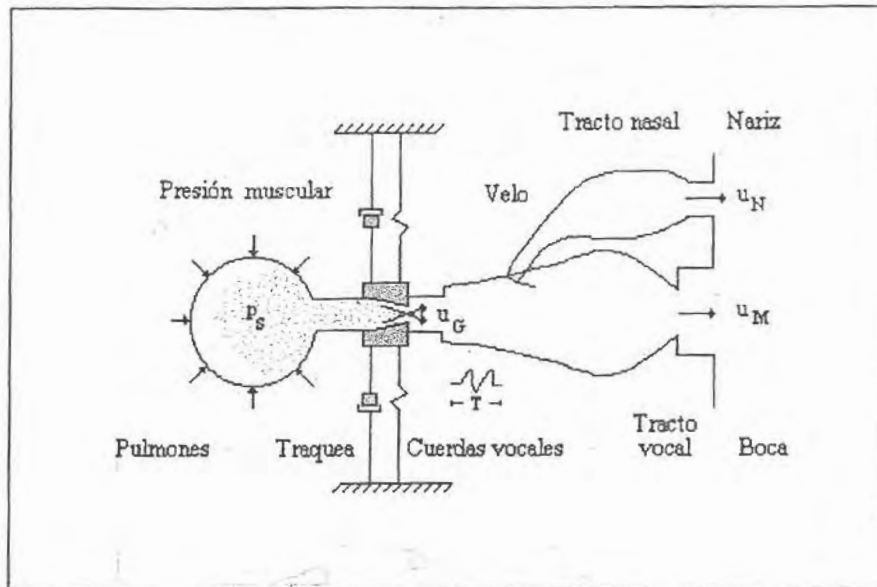


Figura 2.4 Equivalente mecánico del sistema fonatorio.

En general la relación que hay entre el espectro de un segmento de voz de unos 20 ms. y la forma del aparato fonador, que se supone invariante durante este intervalo, es altamente no lineal de ahí que :

Dos configuraciones similares pueden dar lugar a espectros muy diferentes y también dos configuraciones muy semejantes pueden producir espectros muy parecidos.

Una teoría afirma que los mecanismos del habla humana han evolucionado de tal manera que el sistema de comunicación sea robusto. Así los sonidos serían producidos por configuraciones del aparato fonador tales que pequeños cambios accidentales causarían poco efecto acústico.

También esta teoría abona el hecho que los cambios en el modo de la articulación tienen mayor importancia acústica que los cambios en el lugar, esto sugiere que se puede controlar con mayor precisión el primer factor que el segundo.

i) Modelo de Tres tubos

Las cavidades fonadoras se pueden aproximar por medio de tubos de sección uniforme conectados en serie. Todos los tubos son de longitud L igual, y la sección del tubo i -ésimo es A_i . En la figura 2.5 se muestran las aproximaciones para la cavidad oral :

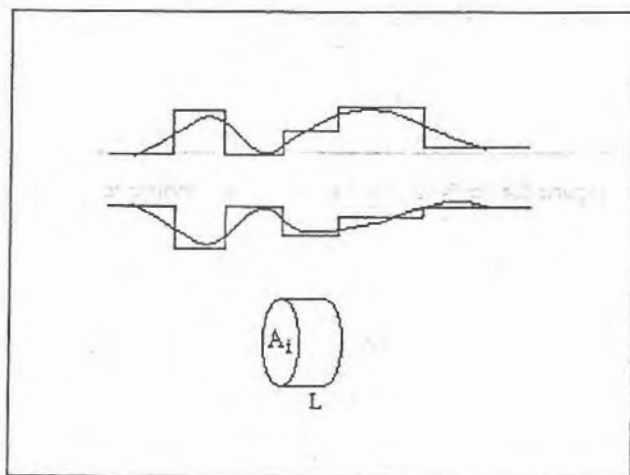


Figura 2.5 Aproximación de una cavidad fonadora mediante tubos.

El modelo de propagación es de ondas planas. Para la sección i con área A_i y longitud L , y la velocidad volumétrica en un punto distanciado x de su extremo izquierdo en el instante t se puede expresar como :

$$u_i(x, t) = u_i^+ \left(t - \frac{x}{c} \right) - u_i^- \left(t + \frac{x}{c} \right) \quad 2.1$$

donde :

$u_i(x, t)$: Velocidad volumétrica determinada por dos ondas.

u_i^+ : Onda propagándose a la derecha.

u_i^- : Onda Propagándose a la izquierda.

t : Tiempo.

x : Distancia.

La velocidad volumétrica resultante es la diferencia entre u_i^+ y u_i^- , considerando a :

$$u_i(t) = A_i v(t)$$

con

A_i : Area de la sección i .

$v(t)$: Velocidad de una partícula de aire.

La presión en cualquier punto de la sección i se expresa como :

$$p_i(x, t) = \frac{\rho c}{A_i} \left[u_i^+ \left(t - \frac{x}{c} \right) + u_i^- \left(t + \frac{x}{c} \right) \right] \quad 2.2$$

para

$$\rho = 1.2 \frac{\text{mg}}{\text{cm}^3} \quad (\text{Densidad del aire})$$

$$c = 340 \frac{\text{m}}{\text{s}} \quad (\text{velocidad del sonido})$$

Las condiciones del contorno entre la sección i y la $i+1$ son :

$$u_i(L, t) = u_{i+1}(0, t)$$

$$p_i(L, t) = p_{i+1}(0, t)$$

que junto con 2.1 y 2.2 resultan :

$$u_i^+(t - \tau) - u_i^-(t + \tau) = u_{i+1}^+(t) - u_{i+1}^-(t) \quad 2.3$$

$$\frac{\rho c}{A_i} [u_i^+(t - \tau) + u_i^-(t + \tau)] = \frac{\rho c}{A_{i+1}} [u_{i+1}^+(t - \tau) + u_{i+1}^-(t + \tau)] \quad 2.4$$

en la cual

$$\tau = L / c$$

es el tiempo de propagación de uno a otro extremo de la sección.

De las ecuaciones (2.3) y (2.4) se puede escribir la ecuación de las ondas salientes de la unión en función de las incidentes :

$$u_{i+1}^+(t) = (1 + r_i)u_i^+(t - \tau) + r_i u_{i+1}^-(t)$$

$$u_i^-(t + \tau) = -r_i u_i^+(t - \tau) + (1 - r_i) u_{i+1}^-(t)$$

Por lo que el *coeficiente de reflexión* entre las secciones i y la sección $i+1$ se define como :

$$r_i = \frac{A_{i+1} - A_i}{A_{i+1} + A_i} \quad \text{para } [r_i] \ll 1$$

El coeficiente r_i alcanza el valor absoluto inferior a uno, cuando una de las secciones tiene un área muy grande comparada con la otra.

En la figura 2.6 se muestra un modelo de tres tubos. En éste modelo se toma en cuenta que después de la unión labial la onda no se refleja, y que también antes de la unión glotal la onda va solo hacia la derecha.

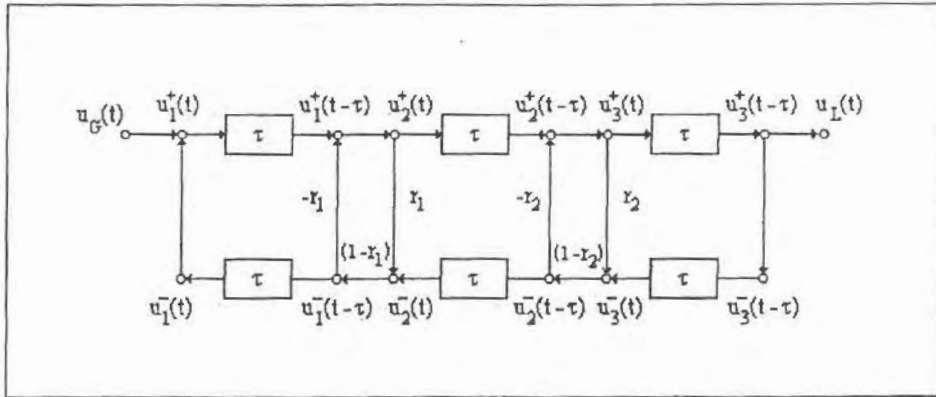


Figura 2.6 Modelo de tres tubos de una cavidad acústica.

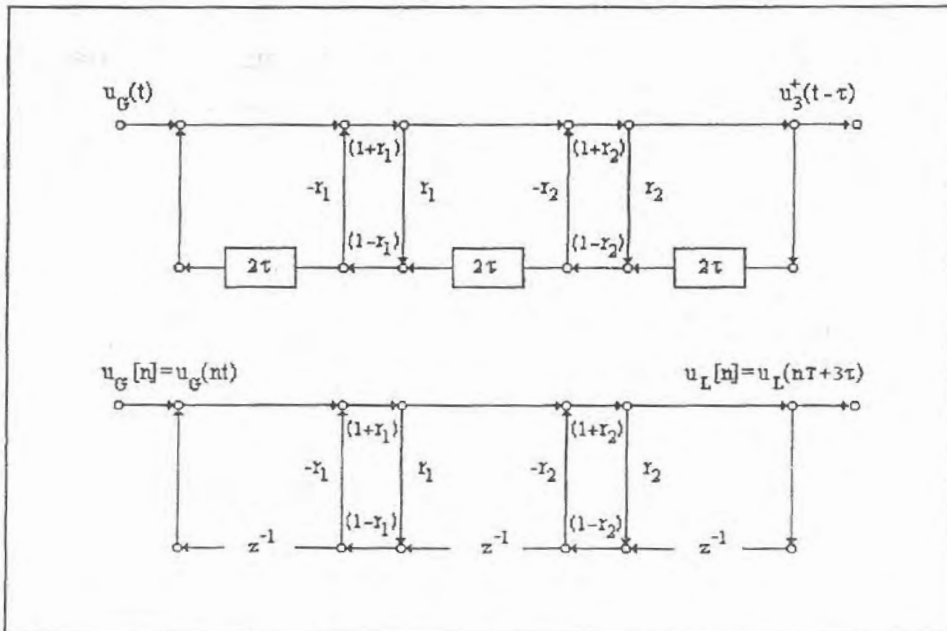


Figura 2.7 Transformación del modelo de tres tubos y obtención del modelo discreto.

En la parte superior de la figura 2.7 se tiene al modelo analógico de tres tubos y en la parte inferior el modelo discreto. En el modelo discreto se tiene que si se muestrea la señal de entrada con un periodo 2τ entonces el filtro discreto se obtiene reemplazando los retardos analógicos por los discretos. La secuencia $u_G(nT)$ permitirá recuperar la señal $u_G(t)$ si la frecuencia de muestreo es superior al doble de su ancho de banda.

Características del modelo de tres tubos

- ♦ La función de transferencia es de solo-polos. Diferentes sonidos pueden producirse cambiando la posición de los formantes, es decir, de los polos.
- ♦ La relación entre la posición de los polos y los coeficientes del filtro son altamente no lineales.
- ♦ Si a partir de la señal de voz se calculan los coeficientes del filtro, entonces a partir de éstos se pueden calcular los coeficientes de reflexión.
- ♦ De los últimos se pueden calcular las áreas de las secciones del modelo, salvo un factor de escala ya que los coeficientes de reflexión dependen de cocientes entre áreas.

El modelo discreto incluye por tanto ceros, además de los polos.

Si solo nos interesa la entrada y salida del modelo discreto se puede implementar mediante diferentes estructuras todas ellas equivalentes, admitiendo que la aritmética que se utilice es de precisión finita.

ii) Síntesis de formantes

La técnica consiste en procesar fonéticamente un texto que se desea generar; haciendo una codificación del mismo, particionando en fonemas. Esta técnica emplea un equivalente electrónico de la fisiología, modelando cada uno de los formantes constitutivos de los fonemas en la voz (figura 2.8).

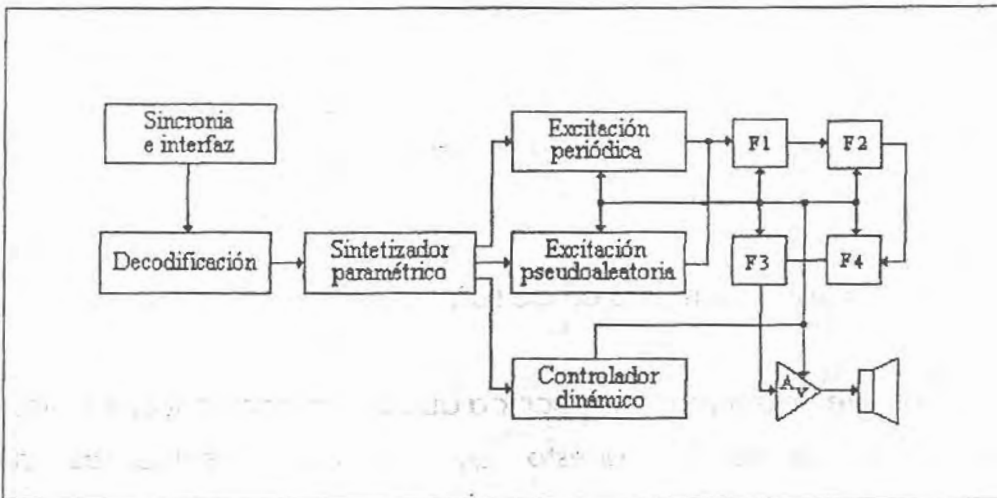


Figura 2.8 Sintetizador de formantes.

iii) Codificación de Predicción Lineal

En esta técnica se requiere un análisis previo de la señal que se desea reconstruir, codificándola directamente de la señal original por un complejo algoritmo matemático, en parámetros específicos que son almacenados digitalmente empleados en la simulación. La codificación de parámetros como coeficientes variables pueden deducirse de las estructuras de predicción lineal y que son recuperadas simultáneamente con la información de voz para

reconstruir la señal original a través de un conversor Digital-Analógico (figura 2.9).

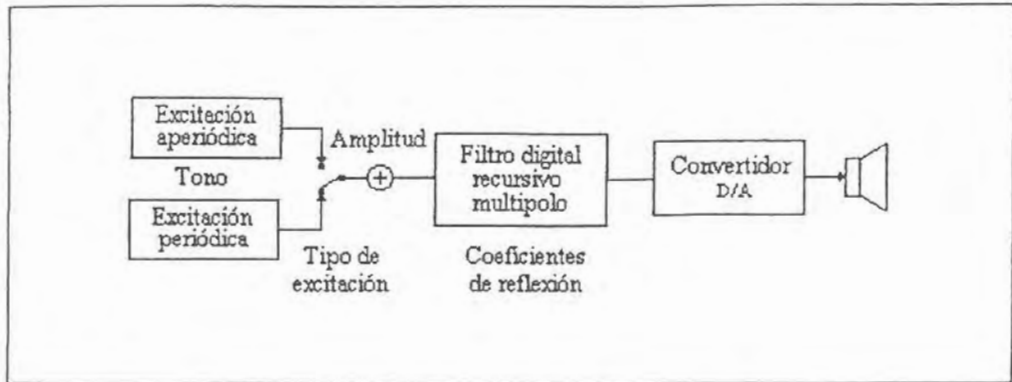


Figura 2.9 Sintetizador de Predicción Lineal.

iv) Digitalización y compresión de señales

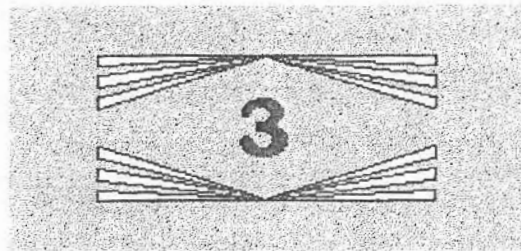
Estrictamente hablando, ésta técnica queda clasificada fuera de los grupos de simulación, puesto que la voz obtenida es la reproducción analógica de una señal grabada, comprimida y almacenada digitalmente. El único paso consiste en la aplicación de un algoritmo de compresión desarrollado por Forest Mozer en la Universidad de California empleando y combinando varias técnicas de compresión.

- ♦ Primero se interpola la digitalización hasta que todos los periodos de tono tengan un número fijo de ellos (quitando o agregando puntos si es necesario).
- ♦ Posteriormente se remueven los periodos de tonos redundantes sustituyendo N periodos similares por periodos únicos repetidos N veces.

- ♦ Después por Modulación Delta Adaptiva (ADM), se utiliza la diferencia de amplitud entre dos digitalizaciones sucesivas añadiendo un valor delta (el valor previo), en lugar de los valores reales en dichos puntos conservando solo la información distinta y útil.

- ♦ A continuación se elimina la componente de posición de la señal por medio de un ajuste de ángulo de fase, tomando cada segmento de tono de la señal y encontrando su serie de Fourier correspondiente. En ella el ángulo de fase no determina ni afecta la inteligibilidad de la voz, generándose una señal con simetría de espejo y poca amplitud para, al menos la mitad del periodo.

- ♦ Por último se hace un ajuste de cero en medio periodo que consiste en eliminar la porción restante de la señal con poca amplitud y reducirla a silencios, sin efecto notable puesto que la mayor parte de la energía se encuentra en la mitad central del periodo de tono. La reproducción de la voz se logra recuperando separadamente la información almacenada de control y datos a través de un convertidor digital/analógico.



CODIFICADORES DE FORMAS DE ONDA

3.1 Introducción

Las representaciones digitales de la señal de voz juegan un papel importante en las telecomunicaciones, así también, como en el almacenamiento de la información. Debido a esto, se han desarrollado diferentes técnicas de codificación digital evaluadas con respecto a la calidad subjetiva, tasa de transmisión y complejidad del algoritmo de codificación. La figura 3.1^{<4>} ilustra la relación entre la calidad de voz y el porcentaje de bits para algunos codificadores. En el diagrama, la calidad de voz se expresa como pobre, regular, buena y excelente.

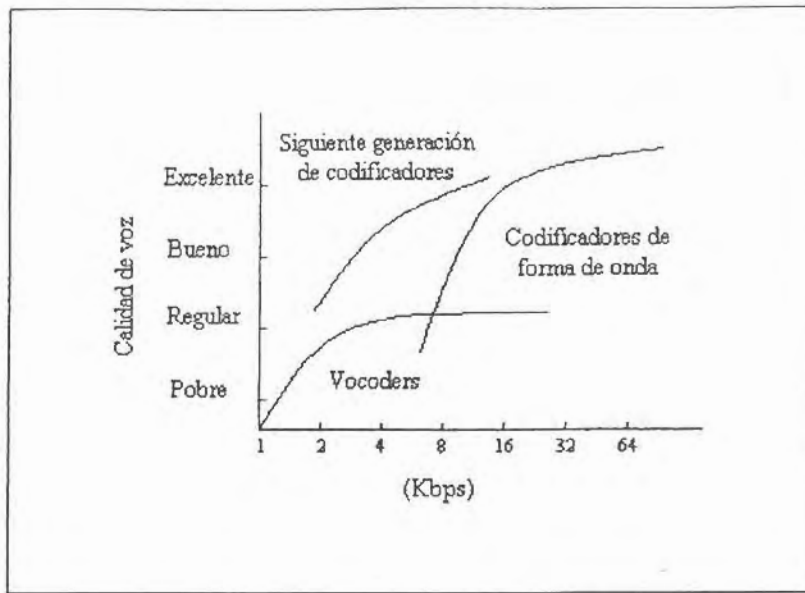


Figura 3.1 Calidad de voz para diferentes tipos de codificadores.

A continuación se revisarán algunos algoritmos de codificación de voz para reducir la velocidad de transmisión. Tal revisión no será exhaustiva pero las técnicas observadas representan los primeros intentos por disminuir la velocidad sin degradar significativamente la calidad de la señal.

3.2 Modulación por Pulsos Codificados (PCM)

La más antigua familia de codificadores se basa en la utilización de un cuantificador y la asignación de un código a cada nivel cuántico.

La digitalización de la señal de voz en PCM se muestra a un porcentaje ligeramente mayor que la frecuencia de Nyquist (normalmente 8 KHz). La señal muestreada resultante tiene una alta correlación de muestras adyacentes. Lo que significa que la señal

no cambia rápidamente de una muestra a la siguiente, con la ventaja de que la diferencia entre tales muestras poseen una variación que es aún más pequeña que la variación de la señal misma.

Si a cada muestra se asocia un nivel discreto y representativo del intervalo al que pertenece la magnitud de tal muestra, se necesitan 2^N niveles de cuantización para proporcionar una relación señal a ruido mayor o igual a 30 dB.

En un cuantificador lineal se requieren $N=13$ bits, es decir, $L=2^{13}$ niveles, para que el ruido de cuantificación no sea perceptible. Por lo que la velocidad de transmisión obtenida es de $8 \times 13 = 104$ Kbps.

A estos codificadores se les conoce como PCM Lineal.

Se observa que la velocidad de transmisión depende del número de niveles del cuantificador y del ancho de banda de la señal.

En la figura 3.2 puede apreciarse el diagrama a bloques del cuantizador PCM Lineal, en el se observan los diferentes elementos que lo constituyen.

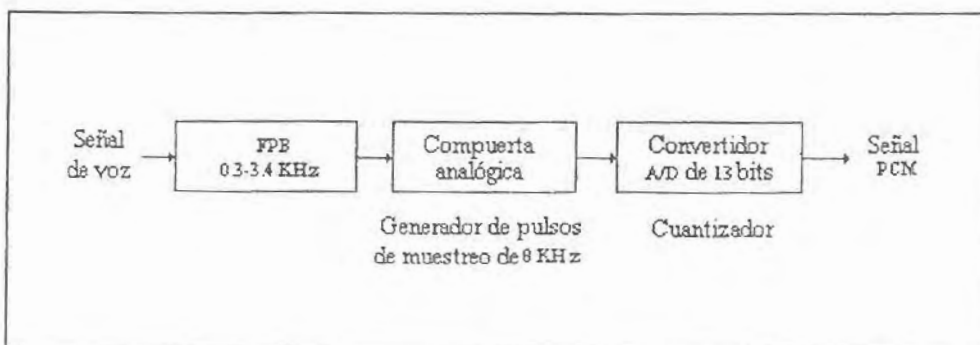


Figura 3.2 Cuantizador PCM lineal.

Las fuentes de error en un sistema PCM son :

- ♦ Error de cuantificación o de redondeo.
- ♦ Error de detección.

Este último es causado por errores de detección de la señal binaria en el receptor.

Para el procesamiento de voz es importante tener la señal cuantificada uniformemente. Esto es necesario porque todas las teorías se aplican a sistemas lineales y las no linealidades introducen complejidades que no son amenas para el análisis. La cuantificación uniforme a pesar de ser una operación no lineal, tiende a serlo cuando se incrementa el número de niveles hasta un determinado límite.

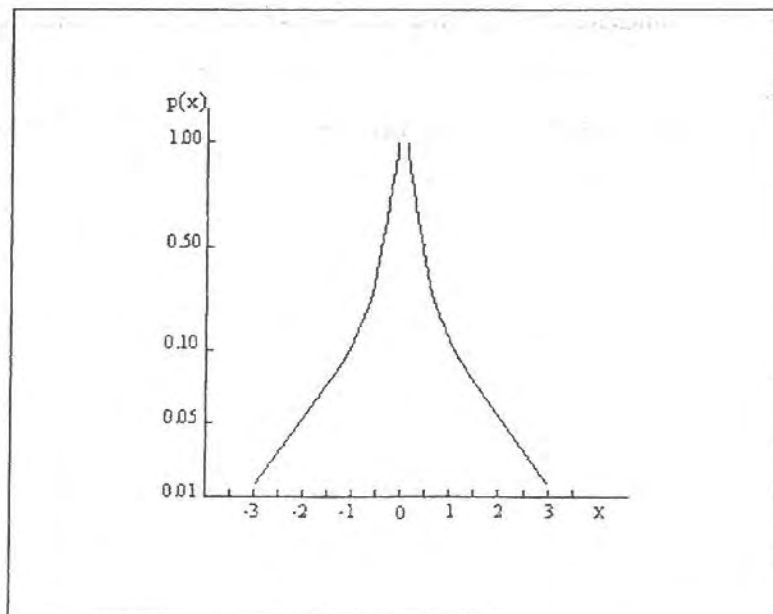


Figura 3.3 Función de densidad de la señal de voz.

El efecto de cuantización se modela tomando en consideración que la señal cuantizada es el producto de la suma de la señal analógica más una pequeña cantidad de ruido de cuantización uniformemente distribuido.

La función de densidad de la señal de voz (figura 3.3) hace suponer que una representación no lineal, por ejemplo logarítmica, proporciona una mejor relación señal a ruido sobre un amplio intervalo de amplitudes de entrada (figura 3.4) y de ahí que sea más usual que la cuantización lineal.

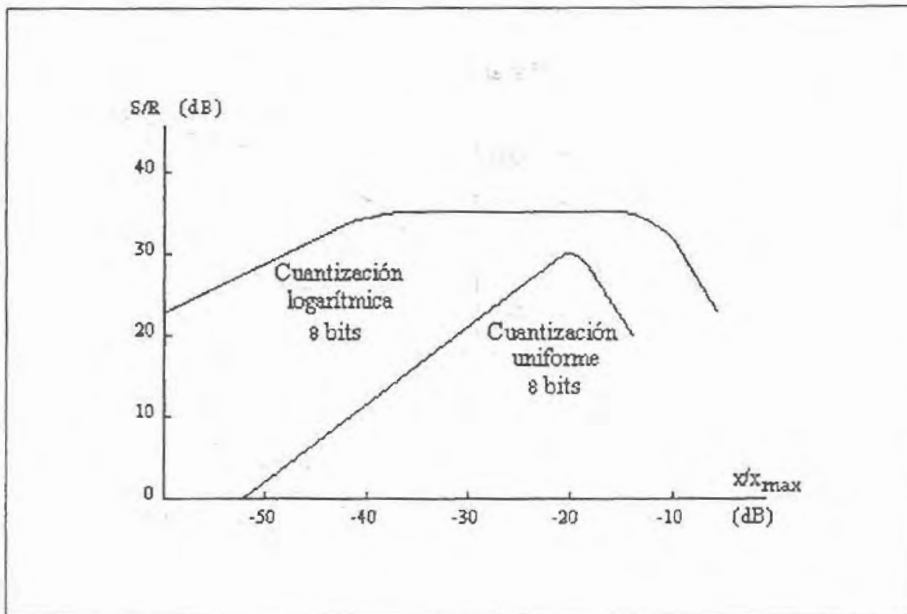


Figura 3.4 Dependencia de la relación S/R respecto a x/x_{max} .

En los sistemas telefónicos es común la codificación PCM logarítmica para transmitir señales con un ancho de banda cercano a 4 KHz y la frecuencia de muestreo es de 8 KHz.

El número de bits de estos cuantificadores es de siete u ocho, por lo que la velocidad de transmisión es de 56 ó 64 Kbps respectivamente.

La figura 3.5 caracteriza la relación logarítmica por medio de la aproximación de la ley A, con la cual se logran 8 bits de código equivalentes a 13 bits con cuantización lineal.

Al utilizar el PCM logarítmico la tasa de transmisión disminuye a :

$$8 \text{ bits/muestra} / 125 \mu\text{s/muestra} = 64 \text{ Kbps}$$

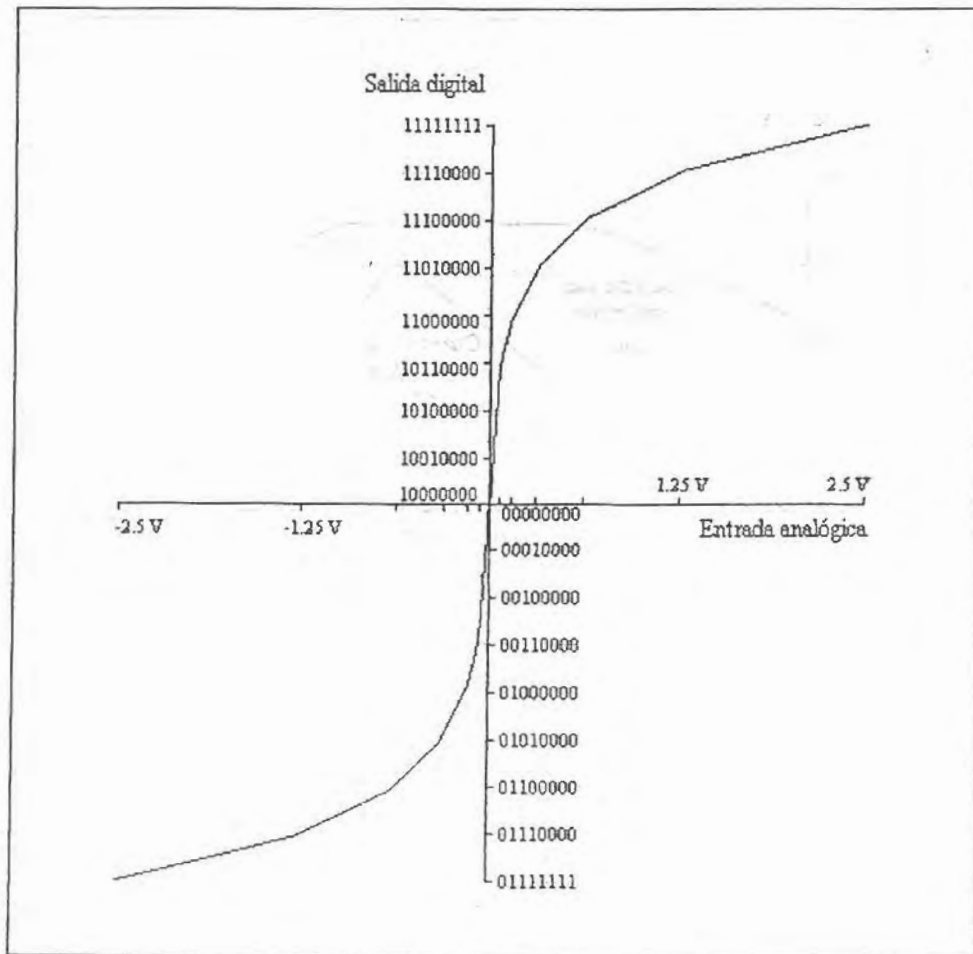


Figura 3.5 Ley A.

3.3 Codificación Diferencial

La estructura de un codificador diferencial se representa en la figura 3.6. En la parte superior se tiene la etapa de análisis y en la parte inferior la de síntesis.

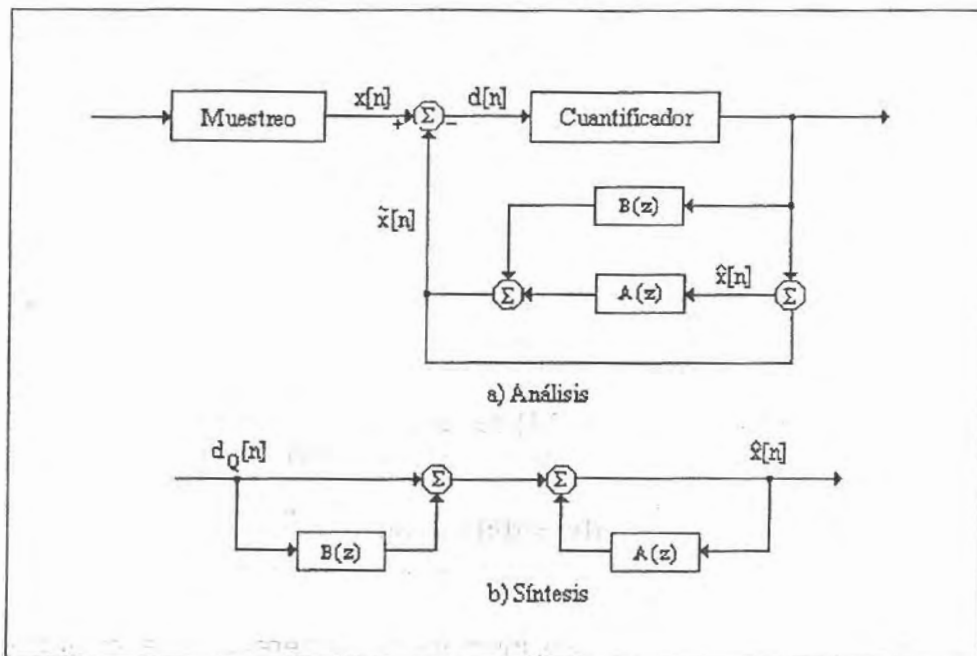


Figura 3.6 Estructura de un codificador diferencial.

Las funciones de transferencia de los filtros son de tipo FIR :

$$A(z) = \sum_{i=1}^p a_i z^{-i}$$

$$B(z) = \sum_{i=1}^q b_i z^{-i}$$

Los filtros y el cuantificador en el codificador diferencial pueden ser tanto adaptivos como fijos. Cuando el filtro y el cuantificador son adaptivos, la adaptación puede o no transmitir información lateral.

Analizando la figura 3.6 se obtiene :

$$\hat{x}[n] = \tilde{x}[n] + d_Q[n] \quad 3.1$$

Considerando que la secuencia en la salida del cuantificador es la entrada más el error de cuantización.

$$d_Q[n] = d[n] + e_Q[n] \quad 3.2$$

$$d[n] = x[n] - \tilde{x}[n] \quad 3.3$$

y sustituyendo (3.2) y (3.3) en (3.1) se tiene

$$\hat{x}[n] = x[n] + d_Q[n]$$

La señal reconstruida $\hat{x}[n]$ en el instante n se diferencia de la original $x[n]$ en el error de cuantificación en el mismo instante $e_Q[n]$.

La relación potencia de señal a potencia de ruido de cuantificación es :

$$SNR = \frac{s^2[n]}{e^2[n]}$$

Si a la SNR, se multiplica y divide por $d^2[n]$ que es la potencia de la señal diferencia resulta :

$$SNR = \frac{s^2[n] d^2[n]}{d^2[n] e_Q^2[n]}$$

Por consiguiente la SNR de un codificador diferencial se expresa como el producto de *ganancia de predicción*, denotado por G y por la relación señal a ruido del cuantificador

$$SNR = G(SNR)_{PCM}$$

en donde

$$G = \frac{s^2[n]}{d^2[n]}$$

Para reducir el número de niveles del cuantificador se necesita aumentar la ganancia de predicción. Esto se logra haciendo que la señal $d[n]$ de diferencia tenga una potencia pequeña. La señal $d[n]$ es la diferencia entre la señal original y la predicción que se hace de ella a partir de la información de muestras anteriores.

3.3.1 Codificación DPCM

Al codificar la salida de un cuantizador, como en la figura 3.7a, se obtiene una importante variación del PCM, el cual se conoce como PCM diferencial (DPCM).

La modulación diferencial de pulsos codificados (DPCM), en una forma simple utiliza la muestra de voz actual para predecir la próxima y almacena o transmite el error de predicción, es decir, la diferencia entre la muestra real y la muestra estimada. En base a esto se reduce la tasa de transmisión, ya que solo se transmite la

diferencia entre muestras y no su magnitud, lo cual requiere una menor cantidad de bits para mantener la exactitud de cuantización.

- La señal diferencia es análoga al error de predicción. En particular en las zonas sonoras muestra una estructura de picos, es decir, la señal tiene amplitudes importantes cuando hay excitación glotal, y mucho menores cuando este no es el caso. Por este motivo los cuantificadores suelen ser no adaptivos.

La diferencia $d[n]$ se somete a una cuantificación que generalmente es no uniforme y se codifica a $N \geq 2$ bits.

La entrada al cuantizador es una señal :

$$e(nT_s) = x(nT_s) - \hat{x}(nT_s) \quad 3.4$$

La cual es la diferencia entre la muestra de entrada $x(nT_s)$ no cuantizada y una predicción de esta, denotada por $\hat{x}(nT_s)$. El valor predicho se produce al usar el predictor cuya entrada como se verá, consta de una versión no cuantizada de la señal de entrada $x(nT_s)$. La señal de diferencia $e(nT_s)$ se le conoce como *error de predicción*, dado que es la cantidad por la cual el predictor *falla* en predecir en forma exacta la entrada.

La función no lineal $Q(\bullet)$ define la entrada y salida característica del cuantizador, la salida del cuantizador se representa como :

$$\begin{aligned} V(nT_s) &= Q[e(nT_s)] \\ &= e(nT_s) + q(nT_s) \end{aligned} \quad 3.5$$

Donde $q(nT_s)$ es el error de cuantización de acuerdo a la figura 3.7a, la salida del cuantizador $V(nT_s)$ se suma al valor $\hat{x}(nT_s)$ predicho para producir la entrada del transmisor.

$$U(nT_s) = x(nT_s) + V(nT_s) \tag{3.6}$$

Sustituyendo la ecuación (3.5) en la (3.6)

$$U(nT_s) = x(nT_s) + V(nT_s) \tag{3.7}$$

Sin embargo de la ecuación 3.4 se observa que $x(nT_s) + e(nT_s)$ es igual a la señal de entrada $x(nT_s)$, por lo que, si se vuelve a escribir la ecuación 3.7 queda como :

$$U(nT_s) = x(nT_s) + e(nT_s) + q(nT_s)$$

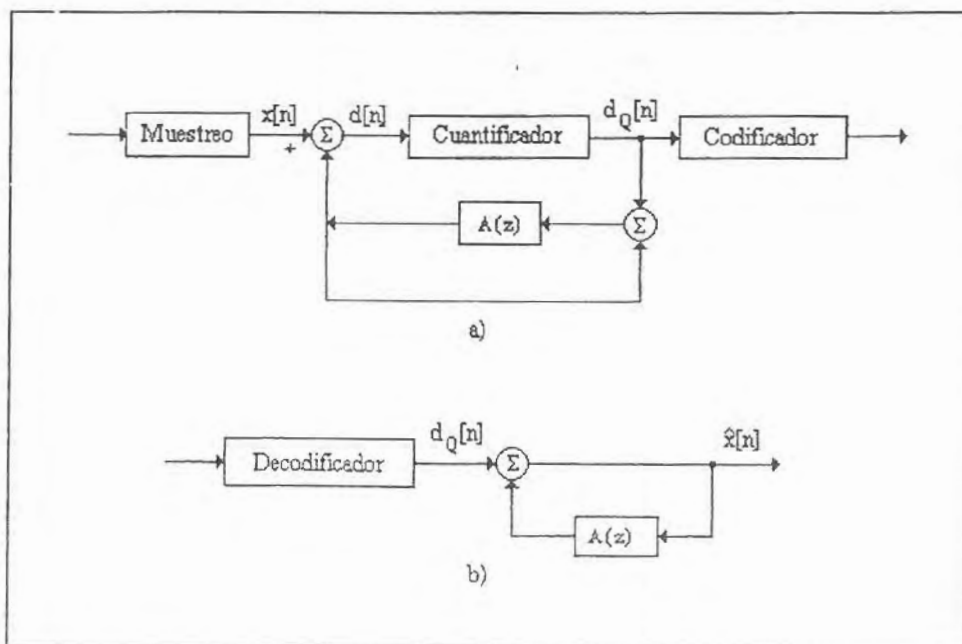


Figura 3.7 Esquema DPCM. a) Transmisor. b) Receptor.

Lo cual representa una versión cuantizada de la señal de entrada $x(nT_s)$. Es decir, no corresponde a las propiedades del predictor, la señal $U(nT_s)$ cuantizada en la entrada del predictor difiere de la señal $x(nT_s)$ original de entrada con el error de cuantización.

De acuerdo a lo anterior, si la predicción es buena, la variación del error $e(nT_s)$ de predicción será más pequeño que la variación de $x(nT_s)$, así que un cuantizador con un número dado de niveles de representación se puede ajustar para producir un error de cuantización con una variación más pequeña que podría ser posible si la señal de entrada $x(nT_s)$ se cuantizará directamente como un sistema PCM estándar.

El receptor utilizado para reconstruir la versión de la señal de entrada cuantizada se muestra en la figura 3.7b, la señal de error $d_q[n]$ se recupera y se introduce al mismo filtro pero con la función de transferencia inversa :

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

y se obtiene la señal de voz estimada $\hat{x}[n]$.

El esquema DPCM también incluye un predictor fijo. Los coeficientes se estiman utilizando técnicas de predicción lineal para un conjunto amplio de señales de voz. Por lo que se considera como un predictor medio.

La naturaleza del predictor es la de un filtro con respuesta impulsiva finita (FIR) cuya función de transferencia es :

$$p(n) = \sum_{k=1}^p a_k z^{-k}$$

Donde a_k ($1 \leq k \leq p$) son los coeficientes de predicción.

P	a1	a2	a3
1	0.85		
2	1.13	-0.38	
3	1.10	-0.28	-0.80

Tabla 1 Coeficientes para un predictor fijo.

Para un predictor fijo de orden $p=1, 2$, ó 3 los coeficientes son los de la tabla 1^{<5>}.

La ganancia de predicción aumenta con el orden del predictor. Experimentalmente se ha determinado que llega a unos 8 dB con predictores de orden 6. Aumentando el orden no se consiguen ganancias mayores, esto es, G tiende a saturarse conforme se aumenta el orden del predictor.

3.3.2 Modulación Delta

Un caso particular del sistema DPCM se da cuando el número de bits asignado por muestra de la señal de error es uno, a este nuevo esquema se le conoce como **Modulación Delta (MD)**.

En su forma básica, el MD proporciona una aproximación en forma de escalera para la versión re-muestreada de una señal de entrada,

como se ilustra en la figura 3.8. La diferencia entre la entrada y la aproximación se cuantiza en dos niveles ($\pm\delta$), correspondientes a las diferencias positivas y negativas.

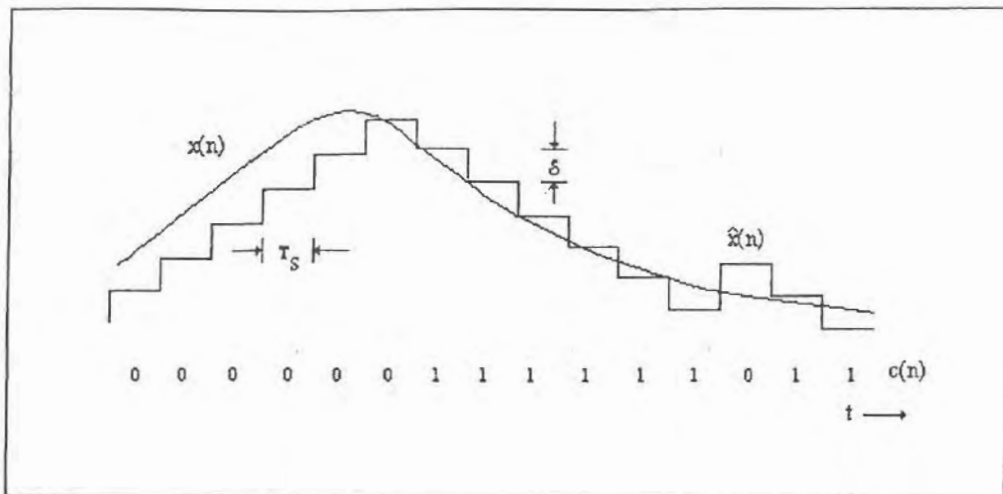


Figura 3.8 Modulación Delta.

Así, si la aproximación está por abajo de la señal se incrementa por δ , por el contrario si está por encima se disminuye. Lo anterior da como resultado que la señal no cambie demasiado rápido de muestra en muestra, δ denota el valor absoluto de la representación de dos niveles del cuantizador de un bit. Los dos niveles se indican en la figura 3.9.

El paso de cuantización Δ se relaciona con δ por $\Delta=2\delta$.

Cuando el orden del predictor se reduce a $p=1$, entonces en el decodificador se tiene :

$$\hat{x}(n) = a_1 \hat{x}(n-1) + \hat{d}(n)$$

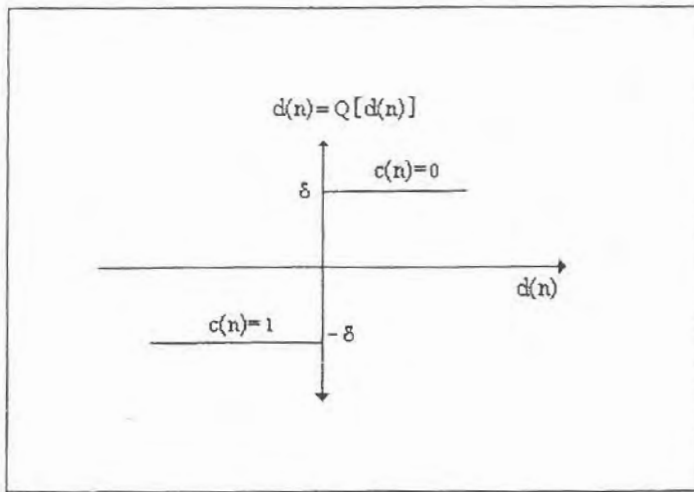


Figura 3.9 Cuantizador Delta.

Pero si $a_1 \cong 1$ entonces ésta ecuación es el equivalente digital de la integración, es decir, representa la acumulación de incrementos positivos o negativos de la magnitud Δ . a_i tiene un valor típico de 0.98.

Denotando la señal de entrada como $x(t)$ y su aproximación como $u(t)$, entonces, el principio básico de la modulación delta se puede formalizar con las siguientes relaciones :

$$\begin{aligned}
 e(nT_s) &= x(nT_s) - \hat{x}(nT_s) \\
 &= x(nT_s) - u(nT_s - T_s)
 \end{aligned}
 \tag{3.8}$$

$$b(nT_s) = \delta \operatorname{sgn}[e(nT_s)]
 \tag{3.9}$$

$$u(nT_s) = u(nT_s - T_s) + b(nT_s)
 \tag{3.10}$$

$e(nT_s)$ es el error de predicción entre el valor de la muestra presente $x(nT_s)$ de la señal de entrada y la última aproximación de ella, llamada :

$$\hat{x}(nT_s) = u(nT_s - T_s)$$

La cantidad binaria es el signo del error $e(nT_s)$ excepto para el factor δ .

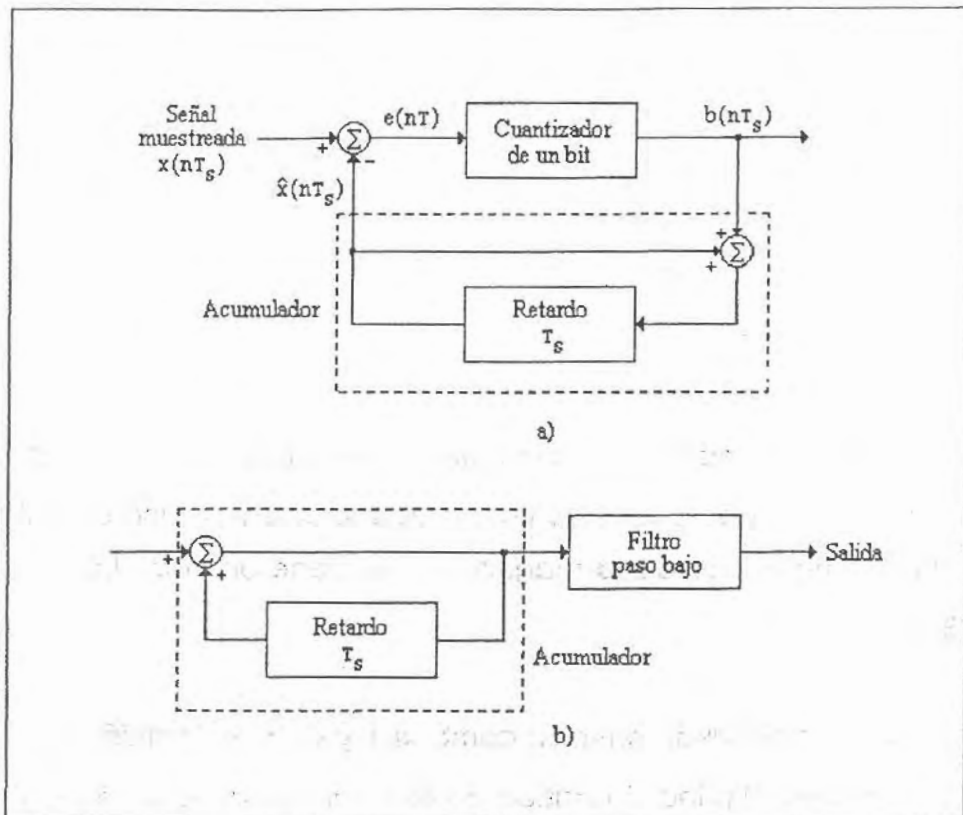


Figura 3.10 Sistema MD. a) Transmisor. b) Receptor.

Un modulador delta consta de un cuantizador de dos niveles y un acumulador, interconectados como se observa en la figura 3.10a.

Suponiendo que el acumulador se encuentra inicialmente en cero, y al resolver las ecuaciones 3.8-3.10 resulta :

$$u(nT_s) = \delta \sum_{i=1}^n \text{sgn}[e(iT_s)]$$

$$= \sum_{i=1}^n b(nT_s)$$

En cada instante de muestreo el acumulador incrementa la aproximación a la señal de entrada por $\pm\delta$, dependiendo de la salida del modulador.

En el receptor (figura 3.10b) la aproximación $u(t)$ se reconstruye al pasar la secuencia de pulsos positivos y negativos a través de un acumulador de forma similar a la del transmisor.

Al reconstruir la señal se observa que una variación rápida de la amplitud de la señal de entrada no es alcanzada por el modulador, a tal efecto se le conoce como *distorsión de sobrecarga*.

Por otro lado, cuando la señal de entrada permanece en un valor constante, la salida del cuantizador será una secuencia alternada de unos y ceros, originando una variación de la señal reconstruida alrededor de un valor constante. Este tipo de error de cuantización constante se le llama *ruido granular*.

Para caracterizar correctamente la señal de voz mediante este método, habrá que compensar la poca información transmitida (1 bit) aumentando la velocidad, por esta razón el bloque de conversión analógico-digital debe re-muestrear la señal de entrada a frecuencias del orden de 40 KHz.

3.4 Codificación Adaptiva

La reducción en el número de bits por muestra de 8 a 4 (como el utilizado en el PCM estándar) involucra el uso combinado de cuantización y predicción adaptiva. El término adaptivo significa tener respuesta a los cambios de nivel y de espectro de la señal de voz. Un esquema de codificación que utiliza la cuantización y predicción adaptiva se le denomina modulación por código de pulsos diferencial adaptiva (ADPCM).

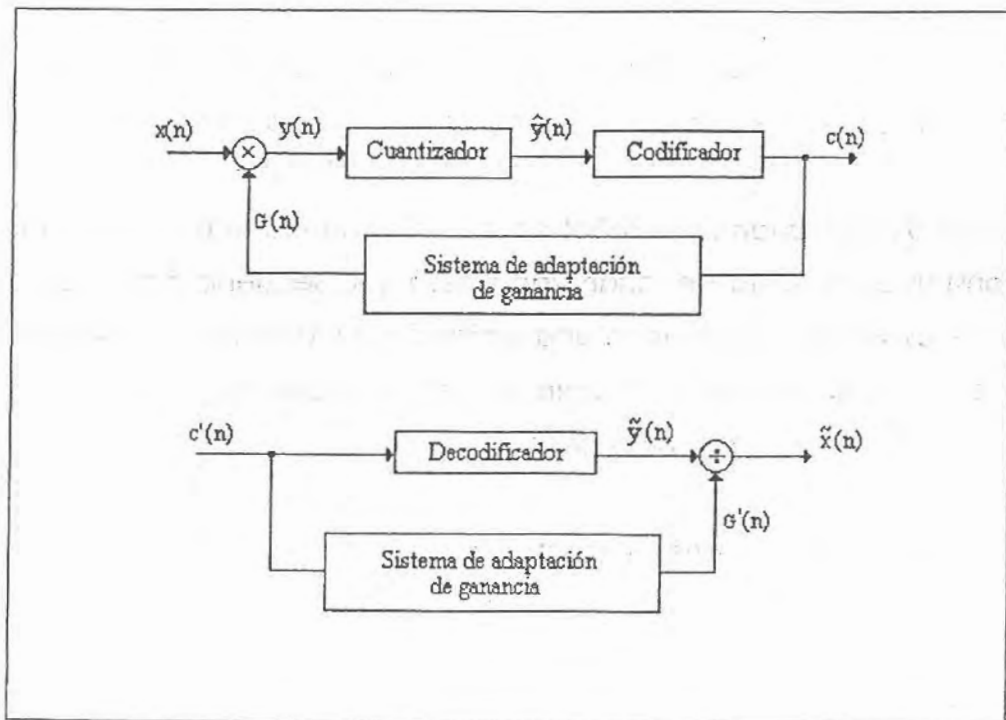


Figura 3.11 Adaptación de Ganancia.

La cuantización adaptiva se refiere a un cuantizador que opera con un tamaño de paso variante en el tiempo $\Delta(nT_s)$. En cualquier tiempo indefinido el cuantizador adaptivo asume una transferencia característica uniforme.

Una técnica para reducir los errores de cuantización es adaptar las variaciones de la señal de entrada con el intervalo de operación del cuantizador, de dos formas :

- Una de ellas es ajustar la magnitud de la señal de entrada mediante un sistema de adaptación de ganancia en un lazo de retroalimentación (figura 3.11), lo que permite al cuantizador permanecer fijo. La ganancia es inversamente proporcional a la magnitud codificada de la muestra anterior.

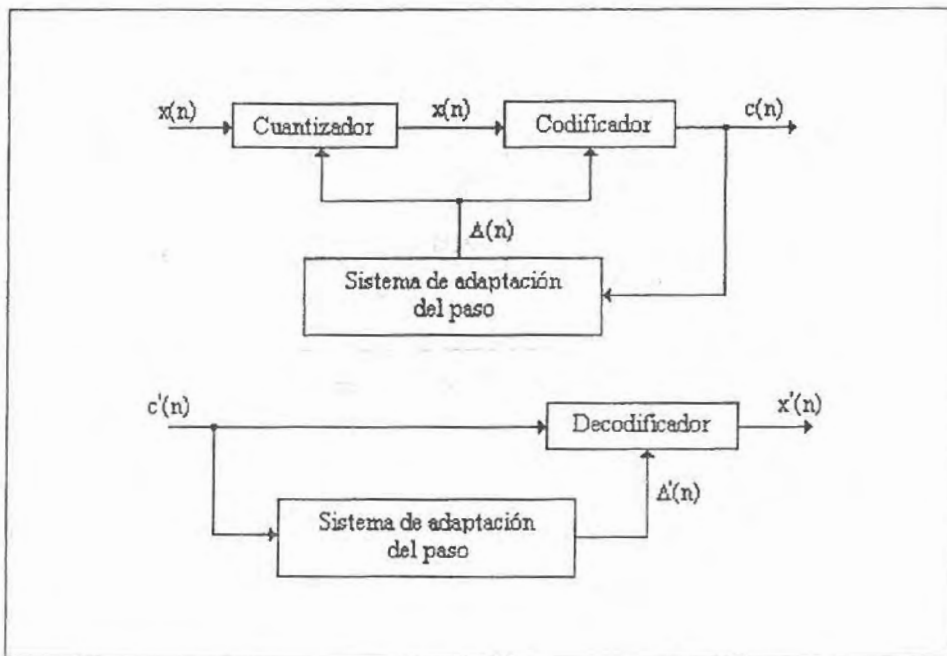


Figura 3.12 Adaptación del paso.

- Otro sistema mantiene la señal de entrada con su propia magnitud, pero en cambio, ajusta el intervalo de funcionamiento del cuantizador en un lazo de retroalimentación (figura 3.12) en donde el paso actual $\Delta(n)$ es el producto del paso anterior $\Delta(n-1)$ por un multiplicador de ajuste que a su vez, es función de la

magnitud del código anterior $|c(n-1)|$.

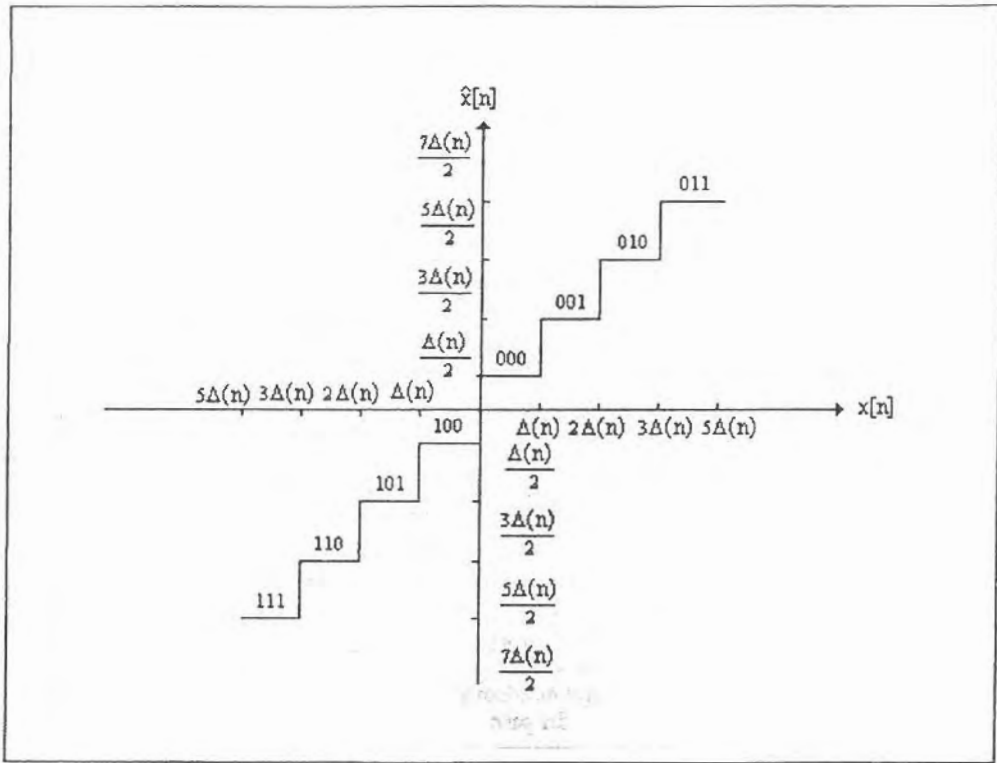


Figura 3.13 Características de E/S de un cuantizador adaptivo.

El cuantizador uniforme adaptivo utilizado (de tres bits por ejemplo), se observa en la figura 3.13.

Si el más significativo de los bits de la palabra código es el signo y el resto la magnitud, entonces la salida del cuantizador será :

$$\hat{x}(n) = \frac{1}{2} \Delta(n) \operatorname{sgn}\{c(n)\} + \Delta(n) c(n)$$

Como una consideración práctica es necesario imponer los límites $\Delta_{min} \leq \Delta(n) \leq \Delta_{max}$ para que un procesador de aritmética finita pueda calcularlo sin problemas.

La relación A_{min} / A_{max} controla el intervalo dinámico del cuantizador y los multiplicadores óptimos⁶⁶ para diversos cuantificadores se observan en la tabla 2.

BITS	TIPO DE CODIFICADOR	
	PCM	DPCM
2	0.60, 2.20	0.80, 1.50
3	0.85, 1.00, 1.00, 1.50	0.90, 0.90, 1.25, 1.75
4	0.80, 0.80, 0.80, 0.80, 1.20, 1.60, 2.00, 2.40	0.90, 0.90, 0.90, 0.90, 1.20, 1.60, 2.00, 2.40
5	0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 1.20, 1.40, 1.60, 1.80, 2.00, 2.20, 2.40, 2.60	0.90, 0.90, 0.90, 0.90, 0.95, 0.95, 0.95, 0.95, 1.20, 1.50, 1.80, 2.10, 2.40, 2.70, 3.00, 3.30

Tabla 2 Multiplicadores óptimos para diferentes cuantificadores.

3.4.1 Modulación Delta Adaptiva

Para un cuantizador diferencial de un bit (Modulación Delta) adaptable, el multiplicador óptimo será :

$$M = P > 1 ; \text{ si } c(n) = c(n - 1)$$

$$M = P < 1 ; \text{ si } c(n) \neq c(n - 1)$$

Para éste caso $c(n) = 0, 1$.

A este nuevo esquema se le conoce como *Modulación Delta Adaptiva (ADM)*.

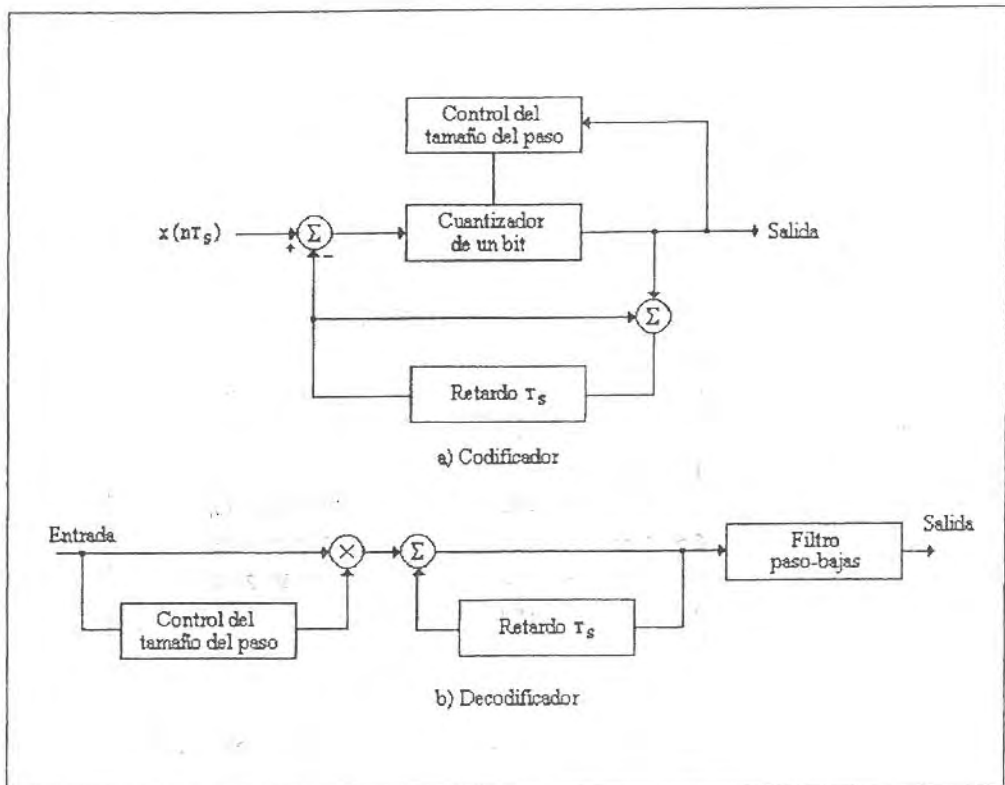


Figura 3.14 Modulador Delta Adaptivo.

Tal estrategia de cuantización fue motivada por la observación de los patrones de código de la Modulación Delta, al hacer que el tamaño del paso tome una forma variante en el tiempo. Es decir, durante una rápida variación del segmento de la señal de entrada el tamaño del paso se incrementa y cuando es más lenta disminuye, de esta forma el tamaño del paso se adapta al nivel de la señal de entrada. Por ejemplo, puede observarse que en periodos de distorsión de sobrecarga, el código resultante es una secuencia de

ceros o de unos. Durante periodos de granularidad las secuencias son alternaciones de 010101...

Hay algunos tipos de ADM, dependiendo del tipo de esquema utilizado para ajustar el tamaño del paso. En un tipo, se considera un conjunto discreto de valores para el tamaño del paso. En el otro tipo, se toma en cuenta un rango continuo para la variación del tamaño del paso. En la figura 3.14 se dibuja un diagrama del transmisor y receptor de un sistema ADM.

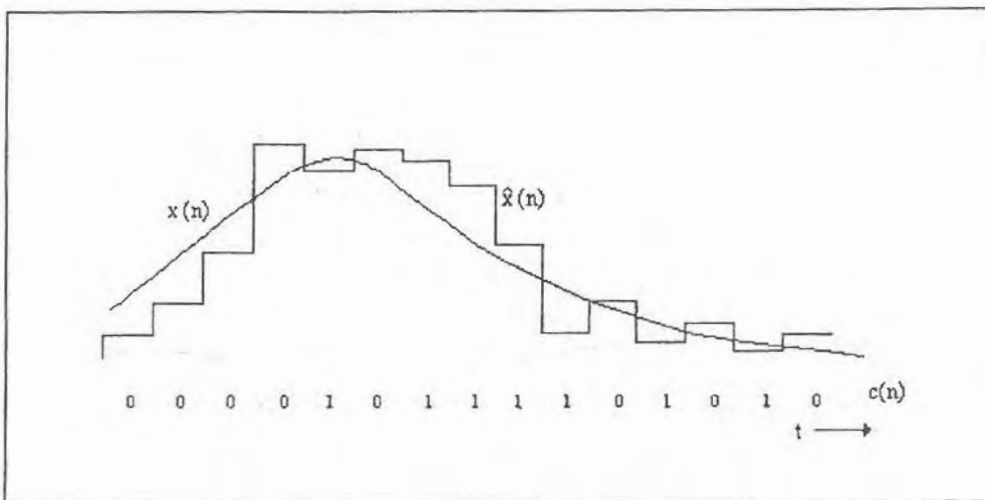


Figura 3.15 Modulación Delta Adaptiva.

Por otra parte la figura 3.15 muestra como puede cuantizarse una forma de onda mediante la modulación delta adaptiva. Los parámetros P y Q deben satisfacer la relación $PQ \leq 1$ para garantizar la estabilidad del sistema; de esta forma $P=2$ y $Q=1/2$ son valores simples con los que puede evitarse el uso de multiplicadores (sustituidos por registros de corrimiento). Sin embargo, los valores que proporcionan mayor relación señal a ruido son $P=1.5$ y $Q=1/P^{5/2}$.

Comparado con el PCM, el sistema ADM transmitiendo a una velocidad de 40 Kbps desarrolla la misma calidad.

3.4.2 PCM Diferencial Adaptivo

Otro ejemplo de cuantización adaptiva es el PCM diferencial adaptivo (ADPCM) con el que se logran tasas de transmisión de 32 Kbps, a partir del código generado por un codificador PCM. Tal sistema está recomendado por el CCITT⁽¹⁾ para su utilización en sistemas de telefonía.

Es claro que los sistemas DPCM con predictores fijos proporcionan un mejoramiento de 4 a 11 dB utilizando una cuantización directa (DPCM), pero la más grande mejora ocurre yendo de la no predicción a la predicción de orden 1, 4 ó 5.

El uso de la predicción adaptiva en ADPCM se justifica porque las señales de voz son inherentemente no estacionarias, fenómeno que se manifiesta por sí mismo en el hecho de que la función de autocorrelación y la densidad de potencia espectral de la señal de voz son funciones variantes en el tiempo de sus respectivas variables. Esto implica que el diseño de predictores para tales entradas debería también ser variante en tiempo, es decir, adaptivo.

Hay dos esquemas para realizar la predicción adaptiva :

- ♦ Predicción adaptiva con estimación hacia adelante (APF). Las muestras cuantizadas de la señal de entrada se utilizan para derivar las estimaciones de los coeficientes del predictor.

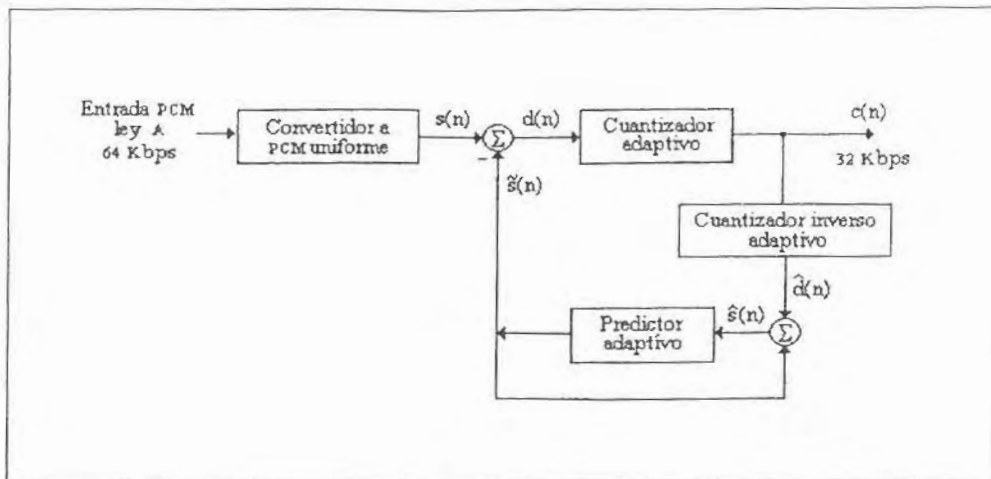


Figura 3.16 Codificador ADPCM.

- ♦ Predicción adaptiva con estimación hacia atrás (APB). Las muestras de la salida del cuantizador y el error de predicción se utilizan para derivar las estimaciones de los coeficientes del predictor.

En el codificador (figura 3.16), después de la conversión de la señal de entrada, PCM ley A de 64 Kbps, a PCM de cuantización uniforme, se obtiene una señal de diferencia $d(n)$ por la sustracción de la señal de entrada $s(n)$ menos una señal estimada $\tilde{s}(n)$. El cuantizador adaptivo asigna los cuatro bits de código ADPCM a cada muestra para ser transmitidos al decodificador.

El cuantizador inverso genera a partir de los cuatro bits, una señal de diferencia cuantizada $\hat{d}(n)$, a la cual se le suma la señal estimada $\tilde{s}(n)$, para producir la versión reconstruida de la señal de entrada $\hat{s}(n)$.

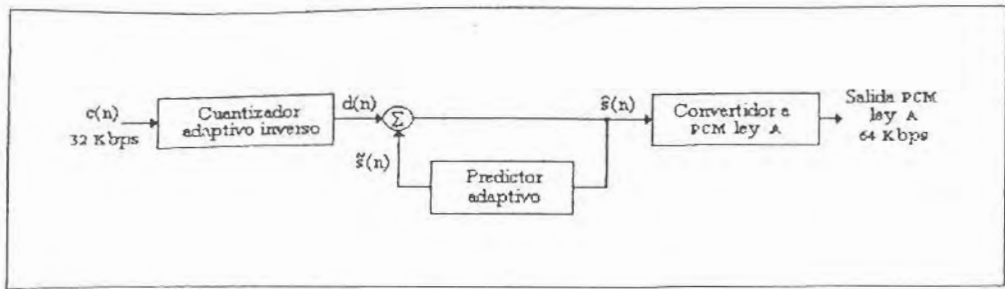


Figura 3.17 Decodificador ADPCM.

Tanto la señal reconstruida $\hat{s}(n)$ como la señal de diferencia cuantizada $\hat{d}(n)$ se procesan por el predictor adaptivo.

El decodificador (figura 3.17) incluye una estructura idéntica a la porción de alimentación del codificador y un cuantizador PCM uniforme -PCM Ley A-. El cuantizador adaptivo y el predictor adaptivo son los elementos principales del codec.

El cuantizador adaptivo del CCITT es un cuantizador hacia atrás, en donde la operación adaptiva se efectúa mediante una red de retroalimentación que opera como control automático de ganancia.

En el cuantizador adaptivo (figura 3.11), el ajuste se logra multiplicando la entrada $x(n)$ por un coeficiente adaptivo $G(n)$, llamado factor de escala, y se actualiza de dos formas :

- ♦ Rápido, para señales que producen grandes fluctuaciones de la señal de diferencia $d(n)$ (ejemplo : voz).
- ♦ Lento, para señales que producen pequeñas fluctuaciones de la señal de diferencia $d(n)$ (por ejemplo datos en la banda de voz o tonos).

El factor rápido de escala $G_r(n)$ es modificado por :

$$G_r(n+1) = m\{c(n)\}G_r^\beta(n)$$

donde $c(n)$ es el código generado por el cuantizador, $m\{\}$ es un multiplicador función de $c(n)$ y $\beta = 31/32$ una constante.

El factor lento de escala $G_l(n)$ se genera a la salida de un filtro paso bajas cuya entrada es $G_r(n)$. Finalmente el factor de escala $G(n)$ se tiene como una combinación lineal de los factores rápido y lento de escala.

El codec propuesto por el CCITT utiliza un predictor adaptivo con una configuración polos-ceros (ARMA). Se trata de combinar la sección solo polos (AR) de segundo orden y la sección solo ceros (MA) de sexto orden, en ambas los polos y ceros son adaptivos.

La sección solo-ceros es un filtro de respuesta impulsiva finita (FIR) de sexto orden, estable aún en presencia de errores de transmisión. La sección solo polos se ha diseñado con ciertas restricciones en los valores de los coeficientes del filtro para limitar la sensibilidad a los errores de transmisión.

La función de transferencia $H(z)$ del codificador es :

$$H(z) = \frac{\hat{s}(z)}{\hat{d}(z)} = \frac{1 + \sum_{i=1}^{\delta} b_i z^{-i}}{1 - \sum_{i=1}^2 a_i z^{-i}}$$

donde a_i ($i=1, 2$) y b_i ($i=1, 2, \dots, \delta$) son dos conjuntos de coeficientes adaptivos, para polos y ceros respectivamente.

El algoritmo de adaptación de coeficientes propuesto por el CCITT, trata de minimizar secuencialmente la potencia de la señal de diferencia $d(n)$.

Para el predictor de segundo orden

$$a_1(n) = (1 - 2^{-8}) a_1(n-1) + 3 * 2^{-8} \operatorname{sgn}(p(n)) \operatorname{sgn}(p(n-1))$$

$$a_2(n) = (1 - 2^{-7}) a_2(n-1) + 2^{-7} [\operatorname{sgn}(p(n)) \operatorname{sgn}(p(n-2)) + f(a_1(n-1)) \operatorname{sgn}(p(n)) \operatorname{sgn}(p(n-1))]$$

donde

$$p(n) = \hat{d}(n) + \sum_{i=1}^{\delta} b_i(n-1) \hat{d}(n-i)$$

$$f(a_i) = \begin{cases} 4a_i & |a_i| \leq 0.5 \\ 2 \operatorname{sgn}(a_i) & |a_i| > 0.5 \end{cases}$$

con las siguientes restricciones para garantizar estabilidad

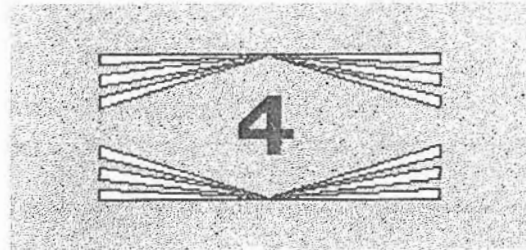
$$|a_2(n)| \leq 0.5$$

$$|a_1(n)| \leq 1 - 2^{-4} - a_2(n)$$

y para el predictor de sexto orden

$$b_i(n) = (1 - 2^{-8}) b_i(n-1) + 2^{-7} \operatorname{sgn}(\hat{d}(n)) \operatorname{sgn}(\hat{d}(n-1)) \quad \forall i=1, 2, \dots, \delta$$

Donde $b_i(n)$ está implícitamente limitado a $|b_i(n)| < 2$.



TECNICAS BASADAS EN LA PREDICCIÓN LINEAL

4.1 Introducción

Una segunda clase de codificadores de voz utilizan un conocimiento *a priori* acerca de como se generó la señal en la fuente. La idea es que ciertas parámetros físicos de la señal generada se pueden procesar y después reproducir, esto implica que la señal se debe modelar y parametrizar adecuadamente. A Las técnicas que explotan las características de la señal se les conoce como *codificadores de fuente*. Los codificadores de fuente para voz se les conoce como *vocoders* (contracción de las palabras voice coders).

4.2 Codificación por Predicción Lineal (LPC)

El filtro de predicción lineal fue desarrollado por Wiener en 1949. Sus primeras aplicaciones al estudio de la señal de voz fueron obra⁴⁹ de Saito e Itakura en 1966, y de Atal y Schroeder en 1967 y 1968. Markel y Gray lo consagraron definitivamente a los sistemas de codificación del habla con la realización del vocoder LPC en 1976.

En la figura 4.1 se muestra un modelo clásico de vocoder para la producción de voz, el modelo supone que el mecanismo generador de sonidos (fuente de excitación) es lineal, y la forma de excitación precisa dependerá si son o no sonidos vocalizados.

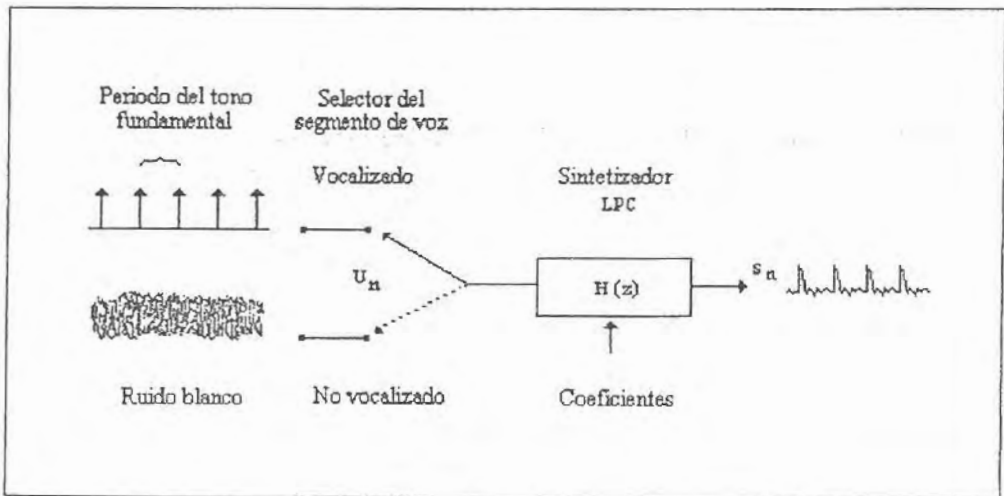


Figura 4.1 Modelo de producción de voz.

La generación de un sonido vocalizado se modela como respuesta del filtro del tracto vocal excitado con una secuencia de impulsos (pulsos muy cortos) espaciados por un periodo fundamental (*pitch*). Por otra parte, un sonido no vocalizado se modela como la respuesta de un filtro vocal excitado por una secuencia de ruido

blanco. El filtro del tracto vocal es variante en el tiempo, pero en un intervalo de tiempo corto, de una veintena de ms. se puede considerar como estático. Sus coeficientes pueden proporcionar una representación adecuada para el segmento de entrada vocalizado o no-vocalizado.

Un vocoder de predicción lineal consta de un transmisor y un receptor. El transmisor realiza la parte de análisis de la señal de entrada, bloque por bloque. Normalmente cada bloque es de 10 a 30 ms. de longitud, en el cual el proceso de producción de voz puede considerarse como estacionario. Los parámetros resultantes del análisis llamados coeficientes del filtro de error de predicción, parámetros vocalizado/no-vocalizado, y periodo del pitch, proporcionan una descripción completa para un segmento en particular de una señal de voz. Una representación digital de los parámetros constituyen la señal transmitida.

Por su parte, el receptor primero realiza la codificación, seguido por la síntesis de la señal de voz. El resultado del análisis y síntesis es una reproducción del sonido artificial de la señal de voz original. Un vocoder con calidad de reproducción pobre se tolera en el caso de seguridad en comunicación militar, donde se requiere una tasa de transmisión baja (4 Kbps o menos).

Análisis de predicción lineal

Se ha visto que la potente capacidad de representación del análisis de predicción lineal^[10], unido a su facilidad de cálculo y a su interpretación intuitiva, la convierten en una herramienta fundamental en el procesado de la señal de voz. La interpretación temporal del predictor es sencilla :

Calcula un conjunto de coeficientes, que predicen el valor de una muestra como una combinación lineal de las anteriores.

Sea $s'(n)$ la señal predicha, $s(n)$ la señal original, y $\{a_k\}$ los coeficientes de predicción lineal :

$$s'(n) = \sum_{k=1}^M a_k s(n-k)$$

$$e(n) = s(n) - s'(n)$$

La señal $e(n)$ es el error residual o error de predicción, y es la diferencia entre la señal original y la predicha.

Sustituyendo y tomando la transformada z de la expresión :

$$S(z) = E(z) + S(z) \sum_{k=1}^M a_k z^{-k}$$

$$H_{LPC}(z) = \frac{S(z)}{E(z)} = \frac{1}{1 - \sum_{k=1}^M a_k z^{-k}}$$

Llamando $A(z)$ al filtro inverso, es decir :

$$A(z) = 1 - \sum_{k=1}^M a_k z^{-k}$$

A este conjunto de coeficientes, se le puede dar otra interpretación, en el dominio de la frecuencia, más intuitiva en su aplicación al análisis de la señal de voz. La representación del módulo de la transformada de Fourier del filtro, construido con los coeficientes de

predicción lineal, es la envolvente espectral del fragmento de voz, a partir del cual se han calculado.

La señal de error residual es el resultado de eliminar de $s(n)$ toda la información espectral capturada por el análisis de predicción lineal, y en el caso ideal de su espectro sería plano. Sin embargo, el filtro LPC es tan solo una aproximación espectral y $e(n)$ conserva además, toda la información referente al carácter sonoro/sordo, a la forma de onda temporal y la no estacionariedad de $s(n)$.

La señal de error residual refleja con mayor nitidez la periodicidad de la señal primitiva. Como se ha eliminado de ella mucha información, que estaba contenida en $s(n)$, su varianza es menor, y su codificación puede realizarse más eficientemente.

Existen varias técnicas para el cálculo de los coeficientes LPC, a partir de la minimización del error cuadrático. Las más utilizadas son el método de la correlación y de la covarianza.

El filtro $A(z)$, es de fase mínima con lo cual su inverso es siempre estable.

Orden del filtro

La elección del orden del filtro LPC es un compromiso entre exactitud en la reproducción del espectro, los recursos de computación y el ancho de banda de transmisión.

Generalmente debe haber un número suficiente de polos para representar adecuadamente todos los formantes del ancho de

banda de la señal, y, además de dos a cuatro polos adicionales para reflejar la presencia de algún cero fortuito introducido por la glotis o el tracto vocal y la forma general del espectro. Normalmente se asume que es suficiente con 3 ó 4 polos para reproducir adecuadamente los ceros y otros efectos de la glotis. Se suelen utilizar, para una frecuencia de muestreo de 8 KHz, entre 10 y 16 polos.

En cualquier caso, al tratarse de señales enventanadas (segmentos de voz multiplicadas por la ventana de Hamming) existe una limitación adicional en la exactitud con la que puede representarse el espectro del habla, ya que llega un punto, en el que el predictor comienza a emular la estructura lobular, consecuencia del enventanado de las señales de voz.

Método de autocorrelación

Para la obtención de los coeficientes del filtro de predicción lineal, volvemos a su interpretación temporal, según la cual, predecimos una muestra como combinación lineal de las M muestras anteriores, siendo M el orden del filtro. Los coeficientes del filtro se obtienen, minimizando la señal diferencia, $e(n)$, entre la original y la predicha, para todo el segmento sometido al análisis.

$$e(n) = s(n) - \sum_{k=1}^M a_k s(n-k)$$

y el error cuadrático, siendo $a_0 = -1$:

$$E = \sum_{n=0}^{L-1} e^2(n) = \sum_{n=0}^{L-1} \left[\sum_{i=0}^M a_i s(n-i) \right]^2 = \sum_{n=0}^{L-1} \sum_{i=0}^M \sum_{j=0}^M (a_i s(n-i) s(n-j) a_j)$$

llamando R_{ij} a :

$$R_{ij} = \sum_{n=0}^{L-1} s(n-i)s(n-j)$$

lo podemos expresar como :

$$E = \sum_{i=0}^M \sum_{j=0}^M a_i a_j R_{ij}$$

minimizando con respecto a los coeficientes de predicción,

$$\frac{\partial E}{\partial a_i} = 0 \quad \Rightarrow \quad \sum_{j=0}^M a_j R_{ij} = 0, \quad a_0 = -1$$

da lugar a un sistema de ecuaciones :

$$a_0 = -1 \quad \Rightarrow \quad \sum_{j=1}^M a_j R_{ik} = R_{0k}, \quad k = 0, \dots, M$$

La señal de entrada se supone enventanada con una ventana de longitud L .

$$s(n) = 0, \quad \begin{cases} n < 0 \\ n > L-1 \end{cases}$$

y se puede escribir :

$$R_{ij} = \sum_{n=0}^{L-1} s(n-i)s(n-j) = \sum_{n=0}^{L-1-|i-j|} s(n)s(n+|i-j|) = \sum_{n=0}^{L-1-|i-j|} s(n)s(n-|i-j|) = r(|i-j|)$$

y el sistema queda :

$$\sum_{i=1}^M a_i r(|i-j|) = -r(j), \quad j = 1, \dots, M$$

siendo

$$r(l) = \sum_{n=0}^{L-1-l} s(n)s(n+l), \quad l \geq 0$$

o expresado en notación matricial

$$\begin{bmatrix} R_0 & R_1 & R_2 & \cdot & \cdot & \cdot & R_{p-1} \\ R_1 & R_0 & R_1 & \cdot & \cdot & \cdot & R_{p-2} \\ R_2 & R_1 & R_0 & \cdot & \cdot & \cdot & R_{p-3} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ R_{p-1} & R_{p-2} & R_{p-3} & \cdot & \cdot & \cdot & R_0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \cdot \\ \cdot \\ \cdot \\ a_p \end{bmatrix} = \begin{bmatrix} R_1 \\ R_2 \\ R_3 \\ \cdot \\ \cdot \\ \cdot \\ R_p \end{bmatrix}$$

$$\|R\| \cdot \|a\| = \|R_1\|$$

$$\|a\| = \|R\|^{-1} \|R_1\|$$

La matriz $\|R\|$ es Toeplitz, y para la solución puede emplearse la recursión de Levinson-Durbin.

Limitaciones

En los esquemas de codificación en los que aparece un filtro de predicción lineal, este modela el efecto de coloreado espectral del tracto vocal y de la glotis, sobre la voz.

En la naturaleza las transiciones espectrales son siempre continuas y suaves, y es necesario garantizar una evolución similar en la voz sintética, para no introducir distorsiones auditivas. Pero, para la

obtención de los coeficientes del filtro, con un grado de complejidad aceptable, es necesario imponer restricciones al modelo ideal de producción del habla :

- ♦ En primer lugar, la evolución de los parámetros no es continua, sino discreta, de tal forma que es preciso trabajar con tramas de la señal de voz en las que se asume un comportamiento estacionario (en cuanto a características espectrales se refiere) para toda la trama, y poder trabajar así con coeficientes constantes en lugar de funciones temporales.
- ♦ El modelado espectral de la fuente generadora de voz, puede ser simulado por un filtro de todos polos, esta simplificación tiene unas repercusiones notorias en la complejidad del procedimiento del cálculo y del régimen de transmisión, y gracias a que el oído humano es más sensible a la fiel reproducción de los picos espectrales que a los valles, resulta válida.

Los polos de un filtro son muy sensibles a la cuantificación de sus coeficientes por esto será necesario realizar transformaciones que tengan un comportamiento más robusto, y que garanticen la estabilidad del filtro, en cualquier situación.

4.3 Excitación Multi Pulso (MPE) y MPE-LTP

El concepto de codificación por predicción lineal excitado por multi-pulsos^[11] fue propuesto por B. S. Atal en 1982. En particular, este tipo de técnica ha demostrado que tiene buena calidad de voz en transmisión baja y media de bits.

En 1984, Atal propuso un mejoramiento de su técnica, el cual se basó en la combinación con técnicas de predicción de término largo (LTP).

Algoritmo básico LPC-MPE

El modelo LPC con excitación MPE es una nueva forma de mejorar la calidad de voz sintética proporcionada por el sintetizador de voz LPC.

La figura 4.2 muestra el diagrama a bloques del método de análisis por síntesis para la determinación de las amplitudes y posiciones de los pulsos de la señal de excitación.

En este diagrama las muestras de voz original $s(n)$ se comparan con las muestras correspondientes de la señal sintética $s'(n)$, la diferencia entre estas señales se pondera para tomar en cuenta las propiedades del oído humano. Entonces, la señal de excitación se determina al minimizar la energía E' del error ponderado $e'(n)$ en intervalos cortos de tiempo (5 ó 10 ms.).

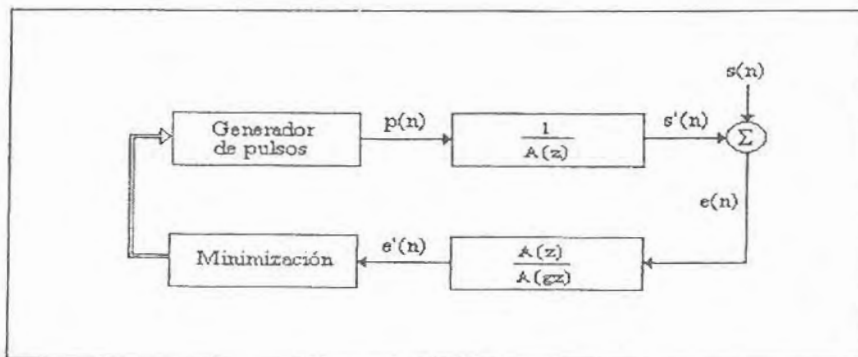


Figura 4.2 Método original MPE.

Se tiene que $S(z)$, $S'(z)$, $E(z)$, y $E'(z)$, denotan la transformada z de las señales $s(n)$, $s'(n)$, $e(n)$, $e'(n)$, respectivamente. Y considerando que :

$$E(z) = S(z) - S'(z) \quad 4.1$$

$$E'(z) = E(z) \frac{A(z)}{A(gz)} \quad \text{con} \quad 0 < g < 1 \quad 4.2$$

Si $R(z)$ y $P(z)$ denotan las transformadas z del error residual de la señal $r(n)$ y la señal de excitación multi-pulso $p(n)$, respectivamente.

$$R(z) = S(z)A(z)$$

$$P(z) = S'(z)A(z)$$

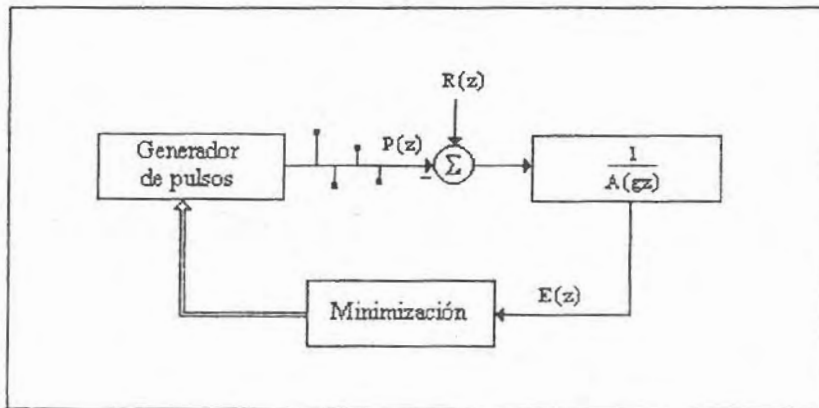


Figura 4.3 Método simplificado MPE.

Al combinar las relaciones (4.1) y (4.2) :

$$E'(z) = [R(z) - P(z)] \frac{1}{A(gz)} \quad 4.3$$

La relación 4.3 sugiere un método alternativo para la determinación de la excitación multi-pulso (figura 4.3).

Con lo referente a los algoritmos, se han propuesto algunos de ellos para determinar las posiciones y amplitudes óptimas de cada uno de los pulsos, y generalmente se basan en la minimización del error E' con respecto a las posiciones del pulso y a partir de esto se obtiene un conjunto de ecuaciones no lineales que no se resuelven directamente.

Las soluciones sub-óptimas de este problema se mejoran mediante procedimientos iterativos, donde cada iteración da resultado a una nueva posición de un pulso. Para cada pulso nuevo, la amplitud del pulso previo se optimiza junto con la amplitud del pulso actual, y el error también se minimiza.

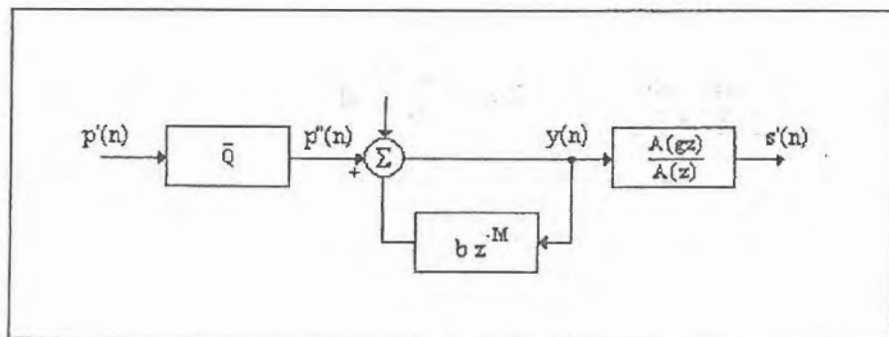


Figura 4.4 Modelo de síntesis con predictor de pitch.

La señal de excitación multi-pulso obtenida por este tipo de procedimiento permite mejorar significativamente la calidad perceptual de la señal de voz en la síntesis del codificador LPC, produciendo en forma natural los sonidos de voz en tasas de transmisión medias.

Codificador MPE-LTP

La segunda propuesta de Atal^[12] consiste en la introducción de un predictor de pitch en el modelo de excitación multipulso, para reducir la correlación observada en el cambio del periodo de un pitch a otro de la señal de excitación multi-pulso. El modelo de síntesis correspondiente se observa en la figura 4.4, a este modelo se le puede asociar dos variantes de modelos de análisis :

- ♦ Un modelo en el que los predictores de término corto y largo se colocan frente al codificador MPE, para que opere en una segunda señal residual $e(n)$ (figura 4.5), y

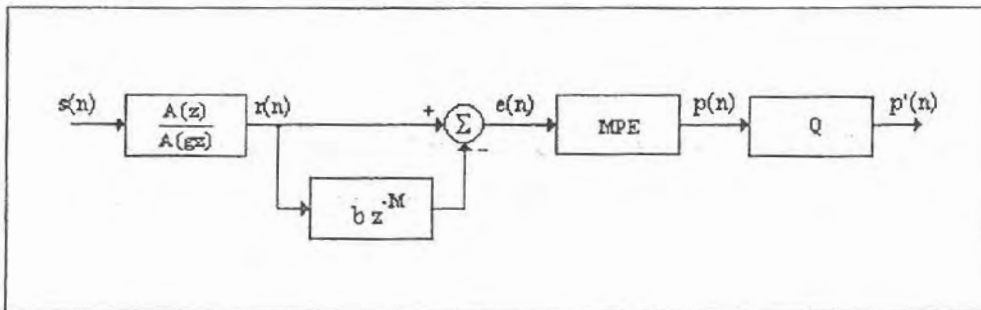


Figura 4.5 Modelo de análisis con alimentación hacia adelante.

- ♦ Un modelo basado en un codificador transversal predictivo, en donde el bloque MPE reemplaza al clásico cuantizador, y se incluye en el lazo de predicción del pitch (figura 4.6). El modelo tiene un mejor desempeño que el anterior, y por consiguiente el enfoque será sobre este.

En el analizador (figura 4.6) la señal de entrada se procesa por bloques de 160 muestras (20 ms.). Para cada bloque se determinan

los coeficientes de predicción de término corto, utilizando el método de autocorrelación y generando un filtro de orden 16.

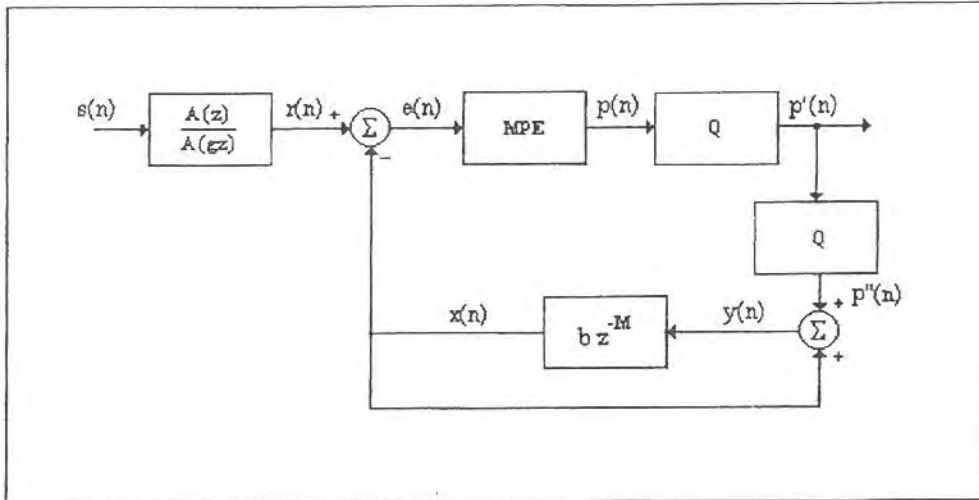


Figura 4.6 Modelo de análisis con lazo de retroalimentación.

El procedimiento de cálculo también utiliza coeficientes de ponderación para así introducir un efecto de enmascaramiento de ruido.

El enmascaramiento de ruido se refiere a que el oído no percibe de igual forma el ruido dentro de todo el espectro, por lo que la degradación subjetiva de la calidad obtenida al superponer ruido a las frecuencias de los formantes, es mucho menor que cuando este ruido se superpone a los valles del espectro. El efecto subjetivo de enmascarar el ruido en frecuencias de mayor energía de la señal de voz se reflejará en los codificadores para optimizar los resultados.

Después la señal de voz se filtra inversamente con un predictor óptimo, produciendo una señal de error residual $r(n)$.

Cada bloque de 20 ms. de la señal de excitación multi-pulso $p(n)$, se divide en 4 sub-bloques de 5 ms. cada uno. Para cada sub-bloque de muestras $r(n)$ se calcula la función de autocorrelación :

$$R(k) = \sum_{n=1}^{40} r(n)r(n-k), \quad k = 41, \dots, 120$$

La búsqueda para la localización del valor pico M lo realiza :

$$R(M) = \underset{k=41}{\overset{120}{\text{Max}}} R(k)$$

Se observa que el valor de M se busca fuera del bloque actual y no necesariamente corresponde al periodo del pitch de la señal de voz, pero si posiblemente a un múltiplo de ese periodo.

La ganancia del predictor de término largo para el sub-bloque actual se calcula por :

$$b = \frac{\sum_{n=1}^{40} r(n)r(n-M)}{\sum_{n=1}^{40} r^2(n-M)}$$

Para cada muestra del sub-bloque actual se escribe :

$$x(n) = b y(n-M)$$

$$e(n) = r(n) - x(n)$$

Donde $y(n)$ representa la señal residual reconstruida.

La señal $e(n)$ representa el error entre la señal residual y el valor predicho, siendo esta última la analizada por el codificador MPE

generando el tren de pulsos $p(n)$, el cual se codifica para originar la secuencia $p'(n)$.

Por ultimo, la señal $p'(n)$ se envía al receptor (figura 4.4) y localmente se utiliza para reconstruir la versión estimada o señal sintetizada $y(n)$ a partir de la señal $r(n)$.

4.4 Excitación de Pulsos Regulares (RPE)

Conceptualmente, el algoritmo de codificación RPE^[13] es un procedimiento de aproximación de la señal residual. La estructura básica del codificador se observa en la figura 4.7.

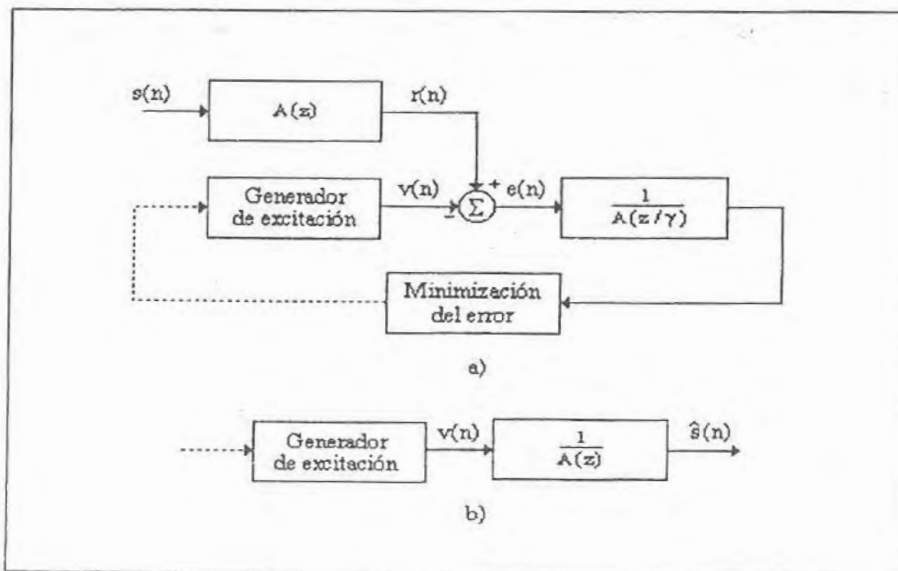


Figura 4.7 Diagrama a bloques del codificador con excitación de pulsos regulares.

La señal residual $r(n)$, se obtiene filtrando la señal de voz $s(n)$ a través del filtro variante en el tiempo $A(z)$ de orden p . Este filtro se

caracteriza mediante las técnicas de predicción lineal descritas anteriormente.

La diferencia entre la señal residual $r(n)$ y un cierto modelo de señal residual $v(n)$ alimenta al filtro de ponderación $1/A(z/\gamma)$.

Este filtro, cuya función es pesar el error, juega el mismo papel que el filtro de ponderación de error de los codificadores con excitación multi-pulso.

La diferencia pesada resultante $e(n)$, se eleva al cuadrado y se acumula para utilizarla como medida para determinar la efectividad del modelo propuesto $v(n)$ de la señal residual $r(n)$.

La secuencia de excitación $v(n)$, se determina para segmentos con L muestras de acuerdo al siguiente procedimiento.

Para cada segmento se requiere una versión *re-muestreada* de un determinado vector óptimo $b = (b(1), \dots, b(Q))$, con longitud $Q < L$. Cada segmento de la señal de excitación contiene Q muestras equidistantes de amplitud diferente a cero, mientras que el resto de las muestras son iguales a cero. La separación entre muestras diferentes de cero es $N = L/Q$.

La duración de un segmento de tamaño L es de 5 ms. típicamente. Cada segmento de excitación consta de N conjuntos de Q muestras equidistantes con magnitud distinta a cero, originando N secuencias de excitación candidatas.

La figura 4.8 muestra las posibles rejillas de excitación para un segmento con 40 muestras y separación $N=3$. En esta figura, las posiciones de los pulsos se marcan con una línea vertical y las muestras cero por puntos.

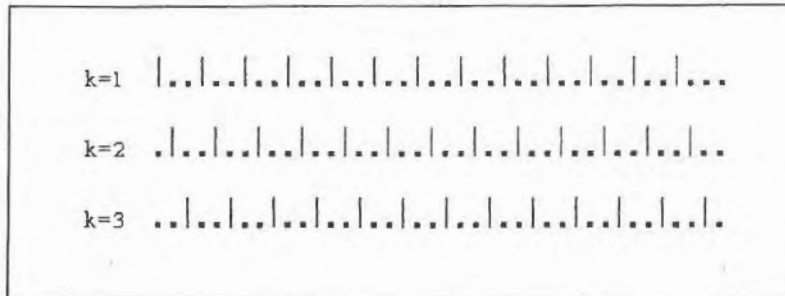


Figura 4.8 Posibles rejillas de excitación con $L=40$ y $N=3$.

Si k ($k = 1, \dots, N$) denota la fase de la versión re-muestreada del vector $b^{(k)}$, por ejemplo cuando la posición de la primera muestra tiene amplitud diferente de cero, es necesario calcular para cada valor de k , las amplitudes $b^{(k)}()$ que minimizan el error cuadrado acumulado.

El vector que logre el mínimo error se selecciona y transmite. El procedimiento de decodificación se observa también en la parte inferior de la figura 4.7.

Algoritmo de Codificación

Sea M_k la matriz Q por L con los elementos :

$$\begin{aligned} m_{ij} &= 1 && \text{si } j = i * N + k - 1 \\ m_{ij} &= 0 && \text{otro} \end{aligned}$$

para

$$\begin{aligned} 0 \leq i \leq Q-1 \\ 0 \leq j \leq L-1 \end{aligned}$$

El segmento de excitación o vector fila $v^{(k)}$ correspondiente a la k -ésima rejilla de excitación está dada por :

$$v^{(k)} = b^{(k)} M_k \quad 4.4$$

Sea H una matriz triangular superior de L por L cuya j -ésima fila ($j=0, \dots, L-1$) contiene la respuesta truncada $h(n)$, del filtro de ponderación de error $1/A(z/\gamma)$ causada por un impulso unitario $\delta(n-j)$. Esto es :

$$\begin{bmatrix} h(0) & h(1) & \dots & \dots & h(L-1) \\ 0 & h(0) & \dots & \dots & h(L-2) \\ 0 & 0 & \dots & \dots & h(L-3) \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & \dots & h(0) \end{bmatrix}$$

Si e_0 denota la salida del filtro de ponderación debido al contenido en la memoria (por ejemplo la salida como resultado del estado inicial del filtro) del intervalo anterior, entonces la señal $e(n)$ producida por el vector de entrada $b^{(k)}$ se describe como :

$$e^{(k)} = e^{(0)} - b^{(k)} H_k \quad k = 1, \dots, N \quad 4.5$$

donde

$$e^{(0)} = e_0 + rH \quad 4.6$$

$$H_k = M_k H \quad 4.7$$

y el vector r representa el residuo $r(n)$ del segmento actual.

Como el objetivo es minimizar el error cuadrado :

$$E^{(k)} = e^{(k)} e^{(k) t} \quad 4.8$$

Donde t denota la transpuesta.

Para una determinada fase (k), las amplitudes óptimas $b^{(k)}()$ se pueden calcular a partir de 4.5 y 4.8 considerando que el producto $e^{(k)} H_k^t$, que de cierta forma mide el error, tiende a cero, de aquí :

$$e^{(k)} H_k^t = e^{(0)} H_k^t - b^{(k)} H_k H_k^t \quad \text{con } e^{(k)} = 0$$

$$b^{(k)} = e^{(0)} H_k^t [H_k H_k^t]^{-1} \quad 4.9$$

Sustituyendo 4.9 en 4.5 y el resultado en 4.8, la expresión para el error E resulta :

$$\begin{aligned} e^{(k)} &= e^{(0)} - e^{(0)} H_k^t [H_k H_k^t]^{-1} H_k \\ e^{(k)} &= e^{(0)} \left[I - H_k^t [H_k H_k^t]^{-1} H_k \right] \\ E^{(k)} &= e^{(0)} \left[I - H_k^t [H_k H_k^t]^{-1} H_k \right] e^{(0) t} \end{aligned} \quad 4.10$$

Se selecciona el vector $b^{(k)}$ que origina el mínimo valor de $E^{(k)}$ para toda k . El vector resultante de óptima excitación $v^{(k)}$ se caracteriza completamente por su fase k y el correspondiente vector de

amplitudes $b^{(k)}$. El procedimiento completo comprende la solución de N conjuntos de ecuaciones lineales dadas en 4.9.

Aunque el efecto de ponderación de ruido se puede escuchar, el mecanismo real que hay detrás de este efecto no es muy claro. El parámetro γ del filtro de ponderación, determina la cantidad de potencia de ruido en las regiones de los formantes del espectro de la voz.

La ponderación de ruido reduce la relación señal a ruido pero mejora la calidad perceptual de voz. Para una frecuencia de muestreo de 8 KHz, un valor típico para γ se encuentra entre 0.8 y 0.9.

Además del valor de γ , el orden del filtro de ponderación de ruido tiene cierta importancia.

Como primera alternativa, los coeficientes a_k y el orden p de $1/A(z/\gamma)$ son iguales a los del predictor $A(z)$, pero puede calcularse un predictor de orden $q < p$ y usar los q coeficientes para definir el filtro de ponderación.

El hecho de que el filtro de ponderación sea variante en el tiempo, proporciona una contribución significativa a la complejidad del procedimiento de análisis, ya que el sistema de ecuaciones lineales a resolver se construye completamente a partir de la respuesta impulsiva de este filtro.

Es evidente que la complejidad de cálculo sería considerablemente menor, si el filtro de ponderación se eligiese de tal forma que la matriz a invertir, no dependiera de los datos de corto tiempo. Para

lograrlo, se escoge un filtro de ponderación $1/C(z/\gamma)$.

$$\frac{1}{C(z/\gamma)} = \frac{1}{1 + \sum_{k=1}^q c_k \gamma^k z^{-k}}$$

Donde $\{c_k\}$ son los coeficientes del predictor fijo de bajo orden, utilizado en los sistemas DPCM y se basan en las características espectrales promedio de la señal de voz.

El valor seleccionado para γ fué 0.8 y al utilizar los coeficientes de predicción $\{c_k\}$ para filtros fijos con diferentes órdenes ($q=1$ a 3), se observa que los efectos de los filtros $1/A(z/\gamma)$ y $1/C(z/\gamma)$ son casi equivalentes.

Este resultado se utiliza para reducir significativamente la complejidad del codificador.

El procedimiento de análisis para el codificador RPE necesita la solución de N conjuntos de ecuaciones lineales, donde N representa el espacio entre pulsos sucesivos en un segmento del modelo de excitación. Sin embargo, el problema de invertir las matrices $H_k H_k^t$ puede solucionarse de una manera más eficiente.

El algoritmo puede reconfigurarse para forzar al producto $H_k H_k^t$ en 4.9 a una simple matriz de Toeplitz la cual es independiente de la fase k .

Sea

$$h(n) = \gamma^n g(n) \quad n = 0, 1, 2, \dots$$

La respuesta impulsiva del filtro $1/A(z/\gamma)$, donde $g(n)$ es la respuesta impulsiva del filtro todo polos $1/A(z)$. Para valores de $|\gamma|$ menores a 1, $h(n)$ converge más rápidamente a cero que $g(n)$ y como resultado la matriz L por $2L$ construida con $h(n)$ se puede aproximar por una matriz de Toeplitz triangular superior H .

$$H = \begin{bmatrix} h(0) & h(1) & \dots & h(L-1) & 0 & \dots & \dots & 0 \\ 0 & h(0) & \dots & h(L-2) & h(L-1) & \dots & \dots & 0 \\ 0 & 0 & \dots & h(L-3) & h(L-2) & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0 \\ 0 & 0 & \dots & h(0) & \dots & \dots & h(L-1) & 0 \end{bmatrix} \quad 4.11$$

Al sustituir 4.11 en 4.6 y 4.7 los vectores e_0 , $e^{(0)}$ y $e^{(k)}$ en 4.6 y 4.5 serán ahora de longitud $2L$, mientras que los vectores $v^{(k)}$ y r en 4.4 y 4.6 mantienen su longitud L .

Después de sustituir 4.11 en 4.6 y 4.7 y descartando la aproximación de orden cero e_0 en 4.6, entonces 4.5 se convierte en :

$$e^{(k)} = e_0 + rH - b^{(k)}H_k \quad e_0 = 0 \quad 4.12$$

$$e^{(k)} = rH - b^{(k)}H_k$$

y la ecuación 4.9 en :

$$b^{(k)}[H_k H_k^t] = [e_0 + rH]H_k^t \quad e_0 = 0$$

$$b^{(k)}[H_k H_k^t] = rHH^t M_k^t \quad 4.13$$

si $S = HH^t$

y enfatizando que

$$H_k H_k^t \cong r_0 I$$

$$\text{donde } r_0 = \sum_{i=0}^{L-1} h^2(i)$$

$$\text{entonces } b^{(k)} = \frac{1}{r_0} r S M_k^t \quad 4.14$$

Es simplemente una versión a escala de la señal residual r ya que r_0 , S , y M_k^t son constantes, lo cual simplifica dramáticamente el cálculo de las rejillas de excitación.

De la misma forma, la selección de la rejilla óptima, basada en la minimización del error de aproximación 4.10, se simplifica al sustituir 4.12 en 4.8.

$$E^{(k)} = [rH - b^{(k)} H_k] [H^t r^t - H_k^t b^{(k)t}]$$

$$E^{(k)} = rHH^t r^t - rHH_k^t b^{(k)t} - b^{(k)} H_k H^t r^t + b^{(k)} H_k H_k^t b^{(k)t}$$

$$\text{pero } H_k^t = H^t M_k^t \quad H_k = M_k H$$

$$E^{(k)} = rHH^t r^t - rHH^t M_k^t b^{(k)t} - b^{(k)} M_k HH^t r^t + b^{(k)} H_k H_k^t b^{(k)t}$$

$$\text{de 4.13 } rHH^t M_k^t = b^{(k)} H_k H_k^t$$

$$E^{(k)} = rHH^t r^t - b^{(k)} H_k H_k^t b^{(k)t} - b^{(k)} M_k HH^t r^t + b^{(k)} H_k H_k^t b^{(k)t}$$

$$E^{(k)} = rHH^t r^t - b^{(k)} M_k HH^t r^t$$

De la transpuesta de 4.13 $M_k^t HH^t r^t = H_k H_k^t b^{(k)t}$

$$E^{(k)} = rHH^t r^t - r_0 b^{(k)} b^{(k)t}$$

De aquí :

$$\min\{E^{(k)}\} = \max\{b^{(k)} b^{(k)t}\}$$

De esta forma el procedimiento anterior es ahora extremadamente simple.

4.5 Excitación Codificada (CELP)

El codificador estocástico excitado por código (CELP)⁴⁹, como el resto de los codificadores paramétricos, modela la señal de voz dividiendo ésta en segmentos, que se asumen estacionarios. De cada segmento se extrae un conjunto de parámetros, a partir de los cuales se sintetiza en el receptor una señal de voz, que conserva la naturalidad y la caracterización del locutor, además de ser perfectamente inteligible.

El sintetizador⁴⁴ estocástico de voz es idéntico al utilizado en los codificadores por predicción adaptiva (APC). Consiste en dos filtros lineales recursivos y variantes en el tiempo, cada uno con un predictor en su lazo de retroalimentación (figura 4.9).

El modelado de la voz se realiza mediante los dos filtros de predicción lineal y una señal de excitación, formada con vectores de ruido blanco seleccionados de una librería o recetario.



BIBLIOTEC

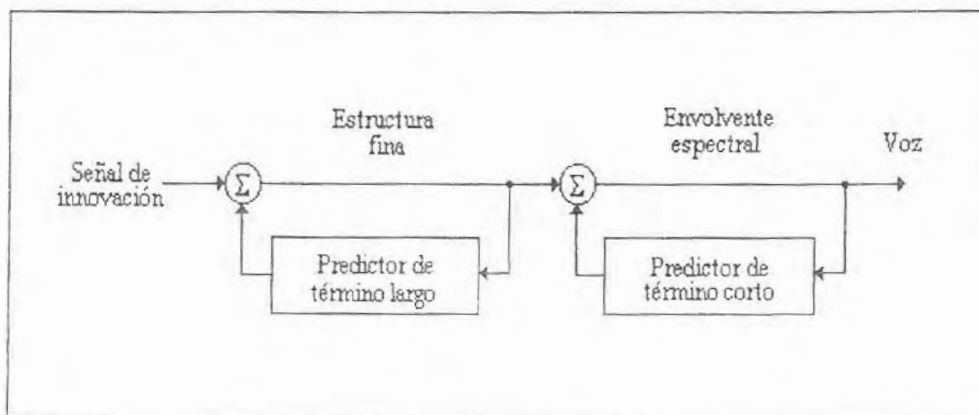


Figura 4.9 Modelo de síntesis de voz con dos tipos de predictor.

El primer lazo de retroalimentación incluye un predictor de término largo que genera la periodicidad del tono vocal de la señal de voz, o en su caso, reproduce las redundancias más marcadas que aparecen en el tramo de voz sometido a proceso, respecto a tramos anteriores. En el modelo de producción de voz, simula el efecto de vibración de cuerdas vocales. La introducción de este filtro evita la distorsión sonoro-sordo que suele ser fuente continua de conflictos.

El segundo, es un predictor de término corto para restaurar la envolvente espectral, es decir, modela el coloreado espectral introducido por el tracto vocal, y se caracteriza por 16 coeficientes de predicción lineal determinados cada 10 ms.

El segmento de la señal de voz a codificar tiene 5 ms. de duración. Cada segmento consta de 40 muestras de voz si la frecuencia de muestreo es de 8 KHz. Una tasa de transmisión de $\frac{1}{4}$ bit/muestra corresponde a 1024 secuencias posibles (10 bits) con una longitud de 40 muestras para cada segmento.

En un codificador estocástico la excitación se realiza a través de vectores de cuantización.

La señal de error obtenida de filtrar la señal de voz por los filtros inversos del filtro de predicción lineal y del predictor largo presenta aspecto de ruido, pero conserva aún información de la señal original :

Toda la información referente a la forma de onda temporal está relegada en $e(n)$, y si bien no es indispensable para la inteligibilidad del mensaje, es fundamental para mantener la naturalidad y las peculiaridades asociadas

al locutor en el habla reproducida. Además, $e(n)$ conserva aún parte del módulo del espectro de la señal.

Para extraer la información que aún queda en ella, se recurre a la cuantificación vectorial, la cual se considera como un sistema utilizado para el mapeo de una secuencia de vectores continuos o discretos almacenados en un *recetario*, del cual se selecciona la secuencia o vector óptimo.

Un detalle básico en el diseño de un cuantificador vectorial es la medida de distancia utilizada. En el caso de codificación de voz, la medida apropiada sería una medida auditiva, buscar el vector de excitación de la librería que permita sintetizar una señal de voz que suene lo más parecida posible a la original, para un receptor humano.

Los modelos auditivos de actual disposición son aún precarios, y la única forma de garantizar que ambas señales suenen igual, es reproduciendo la forma de onda de la señal de voz primitiva. Para lo cual, se recurre a un sistema de análisis mediante síntesis, en el que se compara el fragmento de voz original con las posibles voces sintetizadas por todo el conjunto de vectores de cuantificación, filtrándolos a través del filtro de síntesis, y seleccionando aquel que minimiza la energía del error, atendiendo a criterios perceptuales de enmascaramiento auditivo sencillos.

Para ello el emisor y el receptor disponen de una misma biblioteca de vectores sonoros y fricativos almacenada, y el emisor envía al receptor el índice y ganancia del vector óptimo. El sistema de análisis mediante síntesis, permite por otra parte la compensación de los fallos y limitaciones en el análisis de término largo y del predictor corto.

- El procedimiento para seleccionar el vector o la secuencia óptima se muestra en la figura 4.10.

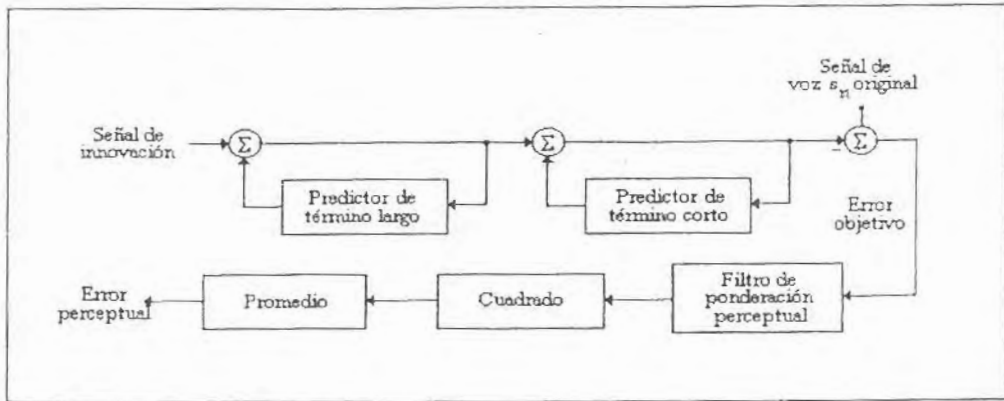


Figura 4.10 Procedimiento de selección de la secuencia de innovación óptima.

Cada miembro del *recetario* de segmentos proporciona 40 muestras de la señal de *innovación*. A cada muestra de la señal innovación se le escala por un factor de amplitud que es constante los 5 ms. del segmento y cambia a un nuevo valor cada 5 ms. Las muestras escaladas se filtran a través de dos filtros recursivos, uno para introducir la periodicidad de la voz y el otro la envolvente espectral.

Las muestras de voz regeneradas en la salida del segundo filtro, se comparan con las correspondientes muestras de la señal de voz original en una señal de diferencia. La señal de diferencia representa el error objetivo que, posteriormente, se procesa a través de un filtro lineal para atenuar todas aquellas frecuencias donde el error es perceptualmente menos importante y amplificar las frecuencias donde es más importante.

La función de transferencia de tal filtro es :

$$W(z) = \frac{1 - \sum_{k=1}^p a_k z^{-k}}{1 - \sum_{k=1}^p \alpha_k \alpha_k z^{-k}}$$

Donde a_k son los coeficientes de predicción lineal, $p=16$ y α es un parámetro para controlar el peso del error como función de la frecuencia.

Un valor apropiado para α es :

$$\alpha = e^{-2\pi 100/f_s}$$

f_s es la frecuencia de muestreo. El error cuadrático medio ponderado se determina al elevar al cuadrado y promediar las muestras de la señal de error en la salida del filtro que pondera al error, para cada segmento de 5 ms. La secuencia de innovación óptima para cada segmento, se selecciona por una búsqueda exhaustiva con el objeto de minimizar el error ponderado.

Construcción de recetarios óptimos

Un recetario, con las limitaciones de tamaño, debe proporcionar un muestreo tan denso como sea posible del espectro de secuencias de innovación. En principio, las palabras código podrían ser códigos de bloque que están óptimamente localizados en una hiperesfera de un espacio de 40 dimensiones (representando 40 muestras en cada segmento de 5 ms).

Un recetario determinístico, si se construye apropiadamente, debería actuar al menos igual, si no mejor, que un recetario

aleatorio y su naturaleza determinística facilitaría la búsqueda de la secuencia de innovación óptima para cada segmento de voz. Sin embargo, por lo regular es muy difícil diseñar un recetario determinístico óptimo. Como una aproximación inicial se escoge un recetario aleatorio en el cual cada palabra código se construye con números aleatorios de distribución Gaussiana con varianza unitaria.

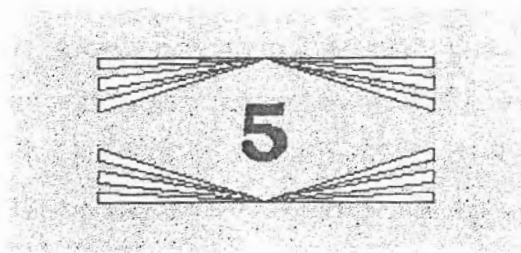
Una observación más profunda del error de predicción, muestra que la suposición de la distribución Gaussiana es válida casi en cualquier situación, excepto para ráfagas de consonantes no vocalizadas y para algunos periodos del tono fundamental durante la transición de sonidos no vocalizados a sonidos vocalizados, o en la transición de regiones en silencio a regiones vocalizadas.

Cada muestra v_n de la secuencia de innovación en un recetario Gaussiano, se puede expresar como una serie de Fourier de N funciones coseno ($N = 20$).

$$v_n = \sum_{k=0}^{N-1} c_k \cos(\pi kn / N + \phi_k) \quad n = 0, 1, \dots, 2N - 1$$

Donde c_k y ϕ_k son variables aleatorias independientes, ϕ_k con distribución uniforme entre 0 y 2π , y c_k con distribución de Rayleigh y función de densidad :

$$p(c_k) = c_k e^{-c_k^2/2} \quad c_k > 0$$



IMPLEMENTACION DEL ANALIZADOR RPE-LTP

5.1 Esquema general del codificador

Recientemente se han propuesto varias técnicas nuevas para codificar la voz con alta calidad a 16 Kbps^{<15>}, orientadas a los sistemas telefónicos digitales comerciales o para los sistemas de radio móvil celular digital entre otros.

En principio la calidad de cualquiera de estas técnicas no deben degradar mucho la calidad subjetiva de voz, es decir, la calidad no debe ser menor de la que ofrecen los sistemas móviles analógicos a 900 MHz. Además se requiere que el algoritmo de codificación sea robusto a las variaciones del nivel o del espectro de la señal de voz.

Se puede esperar una amplia variedad de espectros debido a la variación de locutores, micrófonos y efectos de transmisión de la red telefónica, pero el codificador debe ser robusto al ruido ambiente y a las señales de voces múltiples.

En el caso de una comunicación en tandem (comunicación paralela), durante una conversación entre dos unidades móviles, puede tolerarse, cierta degradación de la señal, sin embargo debe ser una comunicación inteligible en todos los aspectos.

La frecuencia de muestreo para el sistema será de 8 KHz para facilitar la incorporación funcional de la red pública de telefonía conmutada.

La norma básica del codificador de voz será probablemente un transcodificador entre el sistema PCM de cuantización uniforme a 13 bits y una sucesión de bits con una tasa de transmisión a 16 Kbps para radiarse en el sistema móvil.

Por el lado de la red, el codificador proporcionará la transcodificación hacia y desde un sistema PCM con cuantización basada en la ley μ o en ley A.

Los criterios, análisis y pruebas de voz para determinar la mejor técnica de codificación de voz a 16 Kbps bajo estas y otras restricciones^{<15>} se describen en el apéndice.

La figura 5.1 muestra el diagrama a bloques del esquema general del codificador RPE-LTP^{<12>}, ^{<16>} y ^{<24>}. En el se distinguen las diferentes etapas que lo conforman, desde la segmentación hasta la codificación y multiplexación de los parámetros característicos de la voz.

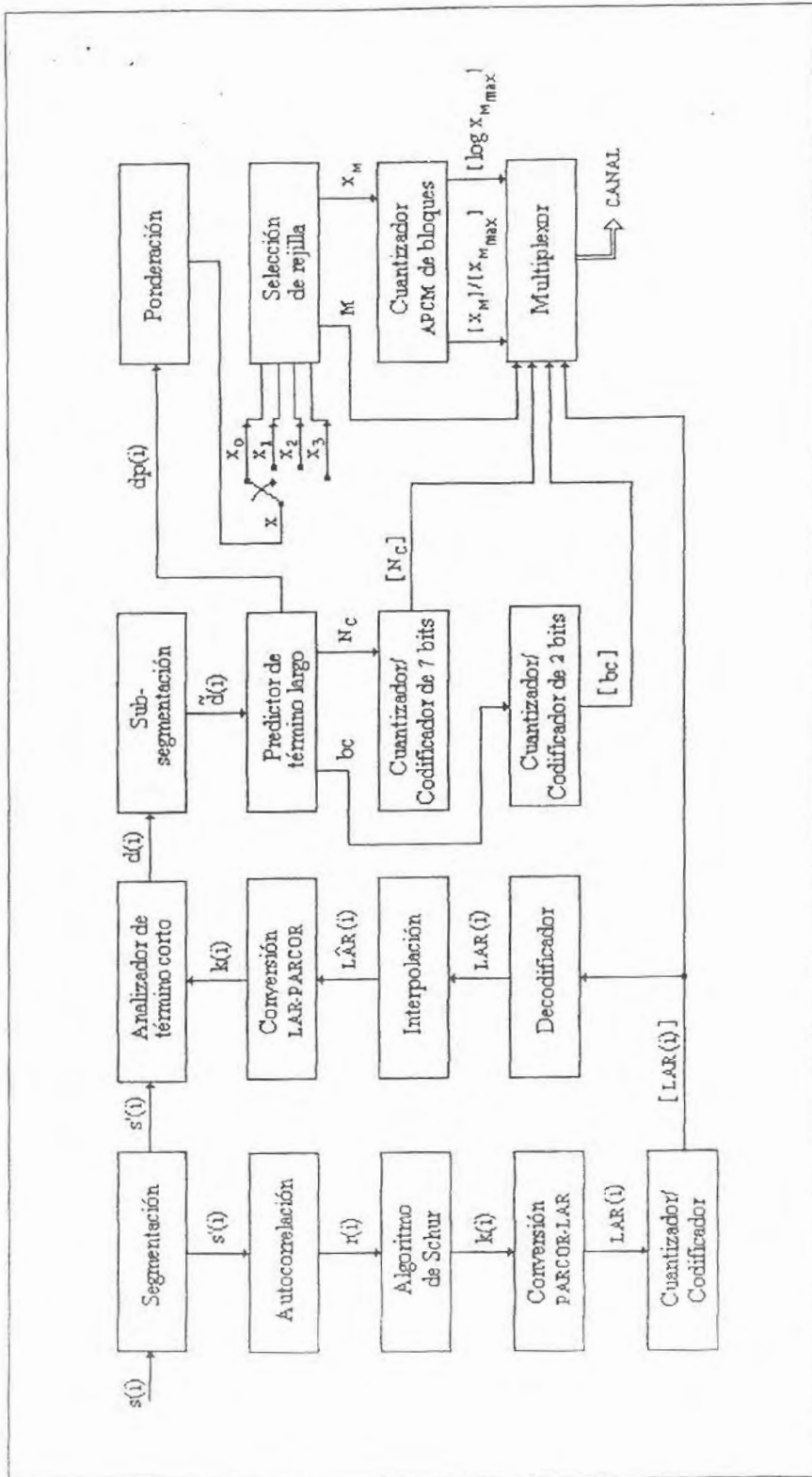


Figura 5.1 Diagrama a bloques del codificador RPE-LTP.

Se observa que para multiplicar dos cantidades bajo la aritmética de punto fijo, se debe considerar en primer lugar la precisión disponible en el procesador ya que por lo regular las palabras tienen una longitud de 16 bits. Por lo tanto, la exactitud de las operaciones será entonces de 15 bits más un bit de signo. El bit más significativo indica el signo.

En una multiplicación los factores pueden escalarse a un intervalo de posibles valores. Si la magnitud absoluta de tal intervalo es menor a uno, entonces la magnitud absoluta del producto será también menor a uno. De esta forma al multiplicar dos cantidades de 15 bits más signo, el producto será de 30 bits más signo con lo cual puede truncarse hasta la precisión de los factores.

Ejemplo :

Si se tienen dos factores con el formato anterior, al multiplicarlos el producto contiene dos bits con el signo y 30 bits para la magnitud.

```
Factor 1  sxxxxxxxxxxxxxx
Factor 2  syyyyyyyyyyyyyyy
producto  ss////////////////////
```

Tal cantidad puede expresarse con la exactitud de los productos considerando un bit de signo y las 15 cifras binarias más significativas de la siguiente forma :

```
Truncamiento  s////////////////
```

Bajo esta técnica cualquier cantidad x deberá estar en el intervalo:

$$-2^{15} \leq x \leq 2^{15} - 1.$$

De esta forma se efectúan las multiplicaciones en la mayoría de los procesos.

No se utiliza el traslapamiento entre segmentos de voz porque se encontró que estas operaciones no mejoran la calidad perceptual de voz producida por el codec. Esto se puede explicar al considerar que la longitud del bloque (20 ms.) hace bastante pequeñas las inexactitudes de la estimación espectral, es por eso que se optó por tomar una ventana rectangular sin traslape.

5.2.2 Autocorrelación

A continuación se realiza el análisis espectral usando el método de autocorrelación para así determinar los primeros $p+1=9$ valores a partir de las muestras de voz del bloque actual $s(n)$ ($n=1,\dots,160$).

$$r(i) = \sum_{j=0}^{L-1} s(j)s(j-i) \quad i = 0, 1, \dots, 9.$$

Puesto que el primer valor de la función de autocorrelación $r(0)$ es el de mayor valor, se utiliza para normalizar a todos los otros, de tal forma que el valor de $r(0)$ sea lo más cercano a 2^{15} , entonces se calcula un factor de normalización u por el cual se dividirán todos los valores $r(i)$.

$$u = r(i) / 2^{15}$$

Esta normalización permite que los valores de autocorrelación mantengan la precisión del microprocesador (1+15) bits.

Para lograr que el cociente de la división de dos cantidades menores a 2^{15} tengan la misma precisión, es necesario multiplicar al dividendo por 2^{15} antes de la división.

5.2.3 Algoritmo de Schur

Por otra parte, mediante el algoritmo de Schur^{<17>} se calculan los coeficientes de reflexión (o coeficientes de correlación parcial PARCOR). Para cada bloque de 20 ms. el algoritmo proporciona un conjunto de 8 coeficientes $k(i)$ ($i = 0, 1, \dots, 7$), los cuales se codifican y multiplexan dentro de la trama transmitida. Alternativamente puede utilizarse el algoritmo de Le Roux Gueguen^{<18>}. Este último resulta apropiado para el cálculo con precisión finita.

Teóricamente, el dominio para cada coeficiente de Parcor es:

$$-1 \leq k(i) \leq 1,$$

pero en la práctica se puede limitar, lo cual incrementa la eficiencia en la codificación.

5.2.4 Conversión Parcor-LAR

Los coeficientes PARCOR se convierten en razones de área logarítmica (LAR) con el propósito de reducir la sensibilidad a los efectos de cuantización^{<19>}.

$$LAR(i) = \log_{10} \left[\frac{1 + k(i)}{1 - k(i)} \right]$$

Tal transformación se aproxima por cinco segmentos lineales :

$$LAR(i) = \begin{cases} r(i) & ; |r(i)| < 0.675 \\ \text{sgn}[r(i)] [2|r(i)| - 0.675] & ; 0.675 \leq |r(i)| < 0.95 \\ \text{sgn}[r(i)] [8|r(i)| - 0.6375] & ; 0.95 \leq 1 \end{cases}$$

LAR No	CUANTIZACION DE PARAMETROS LAR			CODIFICACION BITS/LAR
	MIN	MAX	PASO	
1-2	-1.60	1.55	0.05	6
3	-1.00	0.55	0.05	5
4	-0.55	1.00	0.05	5
5-6	-0.70	0.80	0.10	4
7-8	-0.30	0.40	0.10	3

Tabla 1 Cuantización de parámetros LAR.

5.2.5 Cuantizador/Codificador

Finalmente para cuantizar las razones de área logarítmica, es necesario observar las distribuciones de amplitud para los ocho coeficientes de reflexión y asignar mayor cantidad de bits a los de menor orden.

Por esta razón los coeficientes LAR se cuantizan individualmente con diferentes cuantizadores de paso uniforme como puede apreciarse en la tabla 1.

5.2.6 Filtro de análisis de término corto

El segmento de voz actual de 20 ms. equivalente a $T_0 = 160$ muestras se mantiene en memoria hasta calcular los parámetros LAR, es entonces cuando se alimenta al filtro de término corto de estructura celosía y respuesta impulsiva finita (FIR) de orden $p=8$ cuyos coeficientes se han procesado de acuerdo a :

- ♦ Decodificación e interpolación de los parámetros LAR.
- ♦ Conversión de los parámetros LAR a Parcor, y
- ♦ Conversión de parámetros Parcor a coeficientes de reflexión (k_i).

5.2.7 Decodificación e interpolación de parámetros LAR

Las razones de área logarítmica (parámetros LAR) se decodifican, debido a que pueden ocurrir transitorios no deseados; si los coeficientes del filtro se cambian rápidamente, se aplica una interpolación lineal en la transición entre dos conjuntos sucesivos de parámetros LAR.

5.2.8 Conversión LAR-Parcor

Una vez que los coeficientes LAR son interpolados, se reconvierten a coeficientes de reflexión para derivar los coeficientes a_i ($i = 0, \dots, 8$) que caracterizarán al filtro de análisis de término corto.

El ciclo de cálculo descrito anteriormente (5.2.1 - 5.2.6) se repite cada 20 ms. y generan una secuencia de 160 muestras de la señal de error de predicción $d(k)$.

5.2.9 Filtro de predicción de término largo (LTP)

La correlación de término largo de la señal residual se evalúa de la siguiente forma :

Como cada bloque contiene 160 muestras correspondientes a 20 ms., la correlación de término largo se procesa 4 veces (segmentos de 5 ms. de la señal), por conveniencia se denota el subíndice como $k = 0, \dots, 3$ como el número de cada bloque, así las muestras pertenecientes al n -ésimo bloque de la señal residual se describe por $r(40k + n)$ con $n = 0, \dots, 39$.

Para cada uno de los cuatro bloques se determina el retardo de correlación de término largo $M(k)$ ($k = 0, \dots, 3$) y también se le asocia un factor de ganancia $b(k)$ ($k = 0, \dots, 3$). La estimación de tales parámetros se realiza en tres pasos :

(1) El primero de ellos consiste en la evaluación de la correlación cruzada del segmento actual de la señal residual $r(40k + n)$ con las muestras del segmento previo.

$$RX(l) = \sum_{n=0}^{39} r(40k + n) (40k + n - l) \quad l=40, \dots, 120$$

La correlación cruzada se evalúa para retardos de :

$$40 \leq l \leq 120$$

(2) El paso dos, es la búsqueda de la posición L del valor pico de la función de correlación cruzada dentro del intervalo :

$$RX(L) = \underset{40 \leq l \leq 120}{\text{Max}} RX(l)$$

El retardo de correlación $M(k)$ del bloque actual se le asigna el valor de L .

$$M(k) = L$$

(3) El último paso es evaluar el factor de ganancia $b(k)$ de acuerdo a:

$$b(k) = \frac{\sum_{n=0}^{39} r(i)r(i-L)}{\sum_{n=0}^{39} r^2(i-L)}, \quad i = 40k + n$$

i	Nivel de decisión	Nivel de cuantización
	DLB(i)	QLB(i)
0	0.2	0.10
1	0.5	0.35
2	0.8	0.65
3	1.0	1.00
Para precisión finita :		
0	6553	3278
1	16384	11468
2	26212	21299
3	32767	32767

Tabla 2 Ganancia y cuantización de $b(k)$.

Los tres pasos anteriores se repiten para cada segmento de 5 ms.

El retardo de correlación de término largo $M(k)$, solamente puede tomar valores dentro del rango de 40 a 120, codificando cada valor

con 7 bits, los cuales suman 28 bits por trama transmitida al decodificador.

La ganancia del predictor de término largo $b(k)$ se codifica cada una con 2 bits, utilizando para esto los valores de ganancia y cuantización que se muestran en la tabla 2.

5.2.10 Sub-segmentación y selección de rejilla

Para el proceso de sub-segmentación, ponderación y selección de rejilla, hay dos formas para calcular los vectores de excitación candidatos :

COEFICIENTES DEL FILTRO DE PONDERACION	
k	w(K)
6	1.000000
5,7	0.700790
4,8	0.250793
3,9	0.000000
2,10	-0.045649
1,11	-0.163560
Para precisión finita :	
k	w(K)
6	32767
5,7	22963
4,8	8217
3,9	0
2,10	-1495
1,11	-5359

Tabla 3 Respuesta impulsiva del filtro para los vectores de excitación.

- Una de ellas es por medio de la versión especial de la técnica RPF2⁽¹³⁾, la cual se logra dividiendo cada secuencia de la señal residual $d(i)$, con longitud K , en cuatro sub-segmentos con $K/4 = 40$ muestras cada uno. Estos sub-segmentos se aplican a un filtro de ponderación lineal e invariante en el tiempo con una respuesta impulsiva $w(k)$ de orden 11 (tabla 3), con el propósito de calcular los vectores de excitación candidatos (ecuación 4.14).

De la convolución del sub-segmento ($K/4$ muestras) con el filtro de orden n , resulta un segmento de longitud $K/4+n-1$ muestras.

Debido a que la operación de optimización RPE se aplica individualmente a cada sub-segmento, la operación de filtrado debe ser tal que no se incremente el número de muestras. Por ello en la convolución solo se calculan las $K/4$ muestras centrales.

- Otra forma de calcular el vector de excitación candidato es a través de la aplicación de un diezmado por un factor de tres con tasa de muestreo adaptiva como a continuación se detalla.

Cada sub-segmento filtrado de longitud $K/4$ se descompone en tres secuencias candidatas de pulsos regulares de longitud $K/12$ (figura 5.2) de acuerdo con :

$$x_m(l) = x(k_0 + m + 3l)$$

Donde

$$l = 0, 1, \dots, k/12 - 1 \text{ y } m = 0, 1, 2$$

k_0 define el inicio de la secuencia actual y m la fase de la rejilla diezmada.

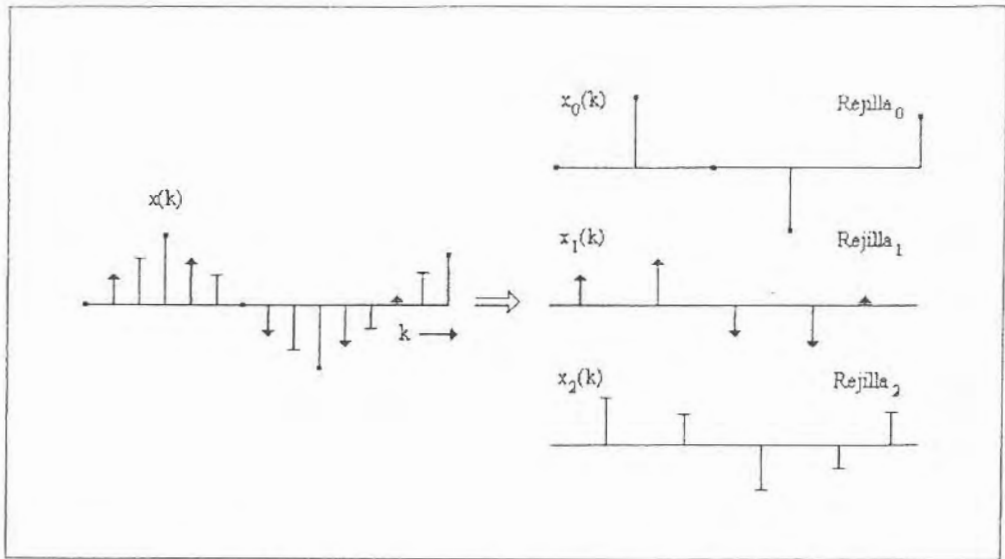


Figura 5.2 Determinación de las secuencias de excitación.

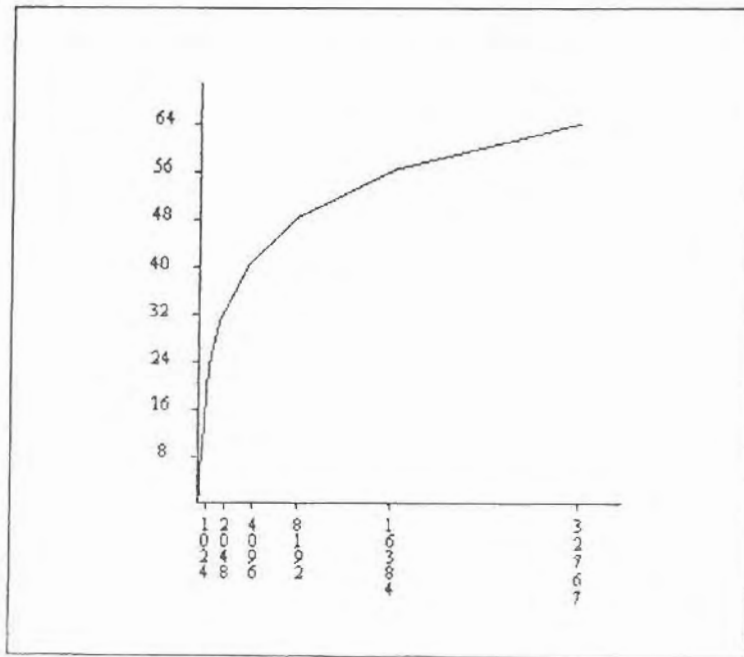


Figura 5.3 Ley A modificada para pulsos máximos.

El candidato óptimo $x_m(l)$ se selecciona, de acuerdo con el criterio de error cuadrático medio, tomando la secuencia de mayor energía.

$$E_M = \max_m \sum_{l=1}^{K/12} x_m(l) \quad m = 0, 1, 2$$

5.2.11 Cuantificación de la secuencia RPE

Finalmente la secuencia RPE seleccionada se cuantiza por la técnica de modulación por impulsos codificados (PCM) de bloque adaptivo.

Cada bloque de $K/12 = 13$ muestras, se normaliza de acuerdo al valor absoluto de la muestra máxima del bloque.

Las muestras escaladas se cuantizan por un cuantizador no uniforme de tres bits del tipo Max^{<20>}.

5.2.12 Cuantización del pulso máximo RPE

La muestra máxima del bloque se cuantiza por un cuantizador logarítmico de seis bits (ley A modificada), figura 5.3. El parámetro de la rejilla se codifica con dos bits.

5.2.13 Formato de la trama codificada

La distribución de los bits para cada trama a transmitir se resume en la tabla 4. La trama de la tabla 5 contiene la distribución de bits a transmitir la cual recibe el codificador para procesarla y al final obtener la señal sintetizada.

Es importante señalar que el formato de distribución de bits propuesto, aún no contiene los bits de protección contra errores. Sin éstos últimos, la tasa de transmisión neta es de 13.0 Kbps, y al agregarlos la tasa de transmisión total será de 16 Kbps, la cual es finalmente la tasa de transmisión requerida para el sistema.

Los detalles correspondientes a la decodificación de cada conjunto de parámetros de las tramas transmitidas se encuentra en la tesis complemento de esta, denominada :

Decodificador de voz a 16 Kbps utilizando la técnica RPE-LTP^[21].

Así mismo, ahí también se encontraran los resultados de las pruebas subjetivas obtenidos y los formatos utilizados para realizar dichas pruebas.

DISTRIBUCIÓN DE BITS POR TRAMA			
8	Parámetros LAR	36	bits
4	Retardo de correlación	28	bits
4	Ganancia del predictor	8	bits
4	Posiciones de rejilla	8	bits
4	Muestras máximas	24	bits
52	Muestras RPE	156	bits
	Protección contra errores	60	bits
Número de bits por trama		320	bits
Tasa neta		13.0	Kbps
Tasa total		16.0	Kbps

Tabla 4 Distribución de los bits en cada trama transmitida.

BIT	SIGNIFICADO
0	Parámetro LAR 1 bit 0
1	Parámetro LAR 1 bit 1
2	Parámetro LAR 1 bit 2
3	Parámetro LAR 1 bit 3
4	Parámetro LAR 1 bit 4
5	Parámetro LAR 1 bit 5
6	Parámetro LAR 2 bit 0
7	Parámetro LAR 2 bit 1
8	Parámetro LAR 2 bit 2
9	Parámetro LAR 2 bit 3
10	Parámetro LAR 2 bit 4
11	Parámetro LAR 2 bit 5
12	Parámetro LAR 3 bit 0
13	Parámetro LAR 3 bit 1
14	Parámetro LAR 3 bit 2
15	Parámetro LAR 3 bit 3
16	Parámetro LAR 3 bit 4
17	Parámetro LAR 4 bit 0
18	Parámetro LAR 4 bit 1
19	Parámetro LAR 4 bit 2
20	Parámetro LAR 4 bit 3
21	Parámetro LAR 4 bit 4
22	Parámetro LAR 5 bit 0
23	Parámetro LAR 5 bit 1
24	Parámetro LAR 5 bit 2
25	Parámetro LAR 5 bit 3
26	Parámetro LAR 6 bit 0
27	Parámetro LAR 6 bit 1
28	Parámetro LAR 6 bit 2
29	Parámetro LAR 6 bit 3
30	Parámetro LAR 7 bit 0

Tabla 5 Formato de la trama propuesta.

31	Parámetro LAR 7 bit 1
32	Parámetro LAR 7 bit 2
33	Parámetro LAR 8 bit 0
34	Parámetro LAR 8 bit 1
35	Parámetro LAR 8 bit 2
36	Retardo de correlación 0 bit 0
37	Retardo de correlación 0 bit 1
38	Retardo de correlación 0 bit 2
39	Retardo de correlación 0 bit 3
40	Retardo de correlación 0 bit 4
41	Retardo de correlación 0 bit 5
42	Retardo de correlación 0 bit 6
43	Factor de ganancia 0 bit 0
44	Factor de ganancia 0 bit 1
45	Fase de rejilla 0 bit 0
46	Fase de rejilla 0 bit 1
47	Pulso máximo de rejilla 0 bit 0
48	Pulso máximo de rejilla 0 bit 1
49	Pulso máximo de rejilla 0 bit 2
50	Pulso máximo de rejilla 0 bit 3
51	Pulso máximo de rejilla 0 bit 4
52	Pulso máximo de rejilla 0 bit 5
53	Pulso 0 rejilla 0 bit 0
54	Pulso 0 rejilla 0 bit 1
55	Pulso 0 rejilla 0 bit 2
56	Pulso 1 rejilla 0 bit 0
57	Pulso 1 rejilla 0 bit 1
58	Pulso 1 rejilla 0 bit 2
59	Pulso 2 rejilla 0 bit 0
60	Pulso 2 rejilla 0 bit 1
61	Pulso 2 rejilla 0 bit 2
62	Pulso 3 rejilla 0 bit 0
63	Pulso 3 rejilla 0 bit 1

Tabla 5 Continuación.

64	Pulso 3 rejilla 0 bit 2
65	Pulso 4 rejilla 0 bit 0
66	Pulso 4 rejilla 0 bit 1
67	Pulso 4 rejilla 0 bit 2
68	Pulso 5 rejilla 0 bit 0
69	Pulso 5 rejilla 0 bit 1
70	Pulso 5 rejilla 0 bit 2
71	Pulso 6 rejilla 0 bit 0
72	Pulso 6 rejilla 0 bit
73	Pulso 6 rejilla 0 bit 2
74	Pulso 7 rejilla 0 bit 0
75	Pulso 7 rejilla 0 bit 1
76	Pulso 7 rejilla 0 bit 2
77	Pulso 8 rejilla 0 bit 0
78	Pulso 8 rejilla 0 bit 1
79	Pulso 8 rejilla 0 bit 2
80	Pulso 9 rejilla 0 bit 0
81	Pulso 9 rejilla 0 bit 1
82	Pulso 9 rejilla 0 bit 2
83	Pulso 10 rejilla 0 bit 0
84	Pulso 10 rejilla 0 bit 1
85	Pulso 10 rejilla 0 bit 2
86	Pulso 11 rejilla 0 bit 0
87	Pulso 11 rejilla 0 bit 1
88	Pulso 11 rejilla 0 bit 2
89	Pulso 12 rejilla 0 bit 0
90	Pulso 12 rejilla 0 bit 1
91	Pulso 12 rejilla 0 bit 2
92	Retardo de correlación 1 bit 0
93	Retardo de correlación 1 bit 1
94	Retardo de correlación 1 bit 2
95	Retardo de correlación 1 bit 3
96	Retardo de correlación 1 bit 4

Tabla 5 Continuación.

97	Retardo de correlación 1 bit 5
98	Retardo de correlación 1 bit 6
99	Factor de ganancia 1 bit 0
100	Factor de ganancia 1 bit 1
101	Fase de rejilla 1 bit 0
102	Fase de rejilla 1 bit 1
103	Pulso máximo de rejilla 1 bit 0
104	Pulso máximo de rejilla 1 bit 1
105	Pulso máximo de rejilla 1 bit 2
106	Pulso máximo de rejilla 1 bit 3
107	Pulso máximo de rejilla 1 bit 4
108	Pulso máximo de rejilla 1 bit 5
109	Pulso 0 rejilla 1 bit 0
110	Pulso 0 rejilla 1 bit 1
111	Pulso 0 rejilla 1 bit 2
112	Pulso 1 rejilla 1 bit 0
113	Pulso 1 rejilla 1 bit 1
114	Pulso 1 rejilla 1 bit 2
115	Pulso 2 rejilla 1 bit 0
116	Pulso 2 rejilla 1 bit 1
117	Pulso 2 rejilla 1 bit 2
118	Pulso 3 rejilla 1 bit 0
119	Pulso 3 rejilla 1 bit 1
120	Pulso 3 rejilla 1 bit 2
121	Pulso 4 rejilla 1 bit 0
122	Pulso 4 rejilla 1 bit 1
123	Pulso 4 rejilla 1 bit 2
124	Pulso 5 rejilla 1 bit 0
125	Pulso 5 rejilla 1 bit 1
126	Pulso 5 rejilla 1 bit 2
127	Pulso 6 rejilla 1 bit 0
128	Pulso 6 rejilla 1 bit 1
129	Pulso 6 rejilla 1 bit 2

Tabla 5 Continuación.

130	Pulso 7 rejilla 1 bit 0
131	Pulso 7 rejilla 1 bit 1
132	Pulso 7 rejilla 1 bit 2
133	Pulso 8 rejilla 1 bit 0
134	Pulso 8 rejilla 1 bit 1
135	Pulso 8 rejilla 1 bit 2
136	Pulso 9 rejilla 1 bit 0
137	Pulso 9 rejilla 1 bit 1
138	Pulso 9 rejilla 1 bit 2
139	Pulso 10 rejilla 1 bit 0
140	Pulso 10 rejilla 1 bit 1
141	Pulso 10 rejilla 1 bit 2
142	Pulso 11 rejilla 1 bit 0
143	Pulso 11 rejilla 1 bit 1
144	Pulso 11 rejilla 1 bit 2
145	Pulso 12 rejilla 1 bit 0
146	Pulso 12 rejilla 1 bit 1
147	Pulso 12 rejilla 1 bit 2
148	Retardo de correlación 2 bit 0
149	Retardo de correlación 2 bit 1
150	Retardo de correlación 2 bit 2
151	Retardo de correlación 2 bit 3
152	Retardo de correlación 2 bit 4
153	Retardo de correlación 2 bit 5
154	Retardo de correlación 2 bit 6
155	Factor de ganancia 2 bit 0
156	Factor de ganancia 2 bit 1
157	Fase de rejilla 2 bit 0
158	Fase de rejilla 2 bit 1
159	Pulso máximo de rejilla 2 bit 0
160	Pulso máximo de rejilla 2 bit 1
161	Pulso máximo de rejilla 2 bit 2
162	Pulso máximo de rejilla 2 bit 3

Tabla 5 Continuación.

163	Pulso máximo de rejilla 2 bit 4
164	Pulso máximo de rejilla 2 bit 5
165	Pulso 0 rejilla 2 bit 0
166	Pulso 0 rejilla 2 bit 1
167	Pulso 0 rejilla 2 bit 2
168	Pulso 1 rejilla 2 bit 0
169	Pulso 1 rejilla 2 bit 1
170	Pulso 1 rejilla 2 bit 2
171	Pulso 2 rejilla 2 bit 0
172	Pulso 2 rejilla 2 bit 1
173	Pulso 2 rejilla 2 bit 2
174	Pulso 3 rejilla 2 bit 0
175	Pulso 3 rejilla 2 bit 1
176	Pulso 3 rejilla 2 bit 2
177	Pulso 4 rejilla 2 bit 0
178	Pulso 4 rejilla 2 bit 1
179	Pulso 4 rejilla 2 bit 2
180	Pulso 5 rejilla 2 bit 0
181	Pulso 5 rejilla 2 bit 1
182	Pulso 5 rejilla 2 bit 2
183	Pulso 6 rejilla 2 bit 0
184	Pulso 6 rejilla 2 bit 1
185	Pulso 6 rejilla 2 bit 2
186	Pulso 7 rejilla 2 bit 0
187	Pulso 7 rejilla 2 bit 1
188	Pulso 7 rejilla 2 bit 2
189	Pulso 8 rejilla 2 bit 0
190	Pulso 8 rejilla 2 bit 1
191	Pulso 8 rejilla 2 bit 2
192	Pulso 9 rejilla 2 bit 0
193	Pulso 9 rejilla 2 bit 1
194	Pulso 9 rejilla 2 bit 2
195	Pulso 10 rejilla 2 bit 0

Tabla 5 Continuación.

196	Pulso 10 rejilla 2 bit 1
197	Pulso 10 rejilla 2 bit 2
198	Pulso 11 rejilla 2 bit 0
199	Pulso 11 rejilla 2 bit 1
200	Pulso 11 rejilla 2 bit 2
201	Pulso 1 2 rejilla 2 bit 0
202	Pulso 12 rejilla 2 bit 1
203	Pulso 12 rejilla 2 bit 2
204	Retardo de correlación 3 bit 0
205	Retardo de correlación 3 bit 1
206	Retardo de correlación 3 bit 2
207	Retardo de correlación 3 bit 3
208	Retardo de correlación 3 bit 4
209	Retardo de correlación 3 bit 5
210	Retardo de correlación 3 bit 6
211	Factor de ganancia 3 bit 0
212	Factor de ganancia 3 bit 1
213	Fase de rejilla 3 bit 0
214	Fase de rejilla 3 bit 1
215	Pulso máximo de rejilla 3 bit 0
216	Pulso máximo de rejilla 3 bit 1
217	Pulso máximo de rejilla 3 bit 2
218	Pulso máximo de rejilla 3 bit 3
219	Pulso máximo de rejilla 3 bit 4
220	Pulso máximo de rejilla 3 bit 5
221	Pulso 0 rejilla 3 bit 0
222	Pulso 0 rejilla 3 bit 1
223	Pulso 0 rejilla 3 bit 2
224	Pulso 1 rejilla 3 bit 0
225	Pulso 1 rejilla 3 bit 1
226	Pulso 1 rejilla 3 bit 2
227	Pulso 2 rejilla 3 bit 0
228	Pulso 2 rejilla 3 bit 1

Tabla 5 Continuación.

229	Pulso 2 rejilla 3 bit 2
230	Pulso 3 rejilla 3 bit 0
231	Pulso 3 rejilla 3 bit 1
232	Pulso 3 rejilla 3 bit 2
233	Pulso 4 rejilla 3 bit 0
234	Pulso 4 rejilla 3 bit 1
235	Pulso 4 rejilla 3 bit 2
236	Pulso 5 rejilla 3 bit 0
237	Pulso 5 rejilla 3 bit 1
238	Pulso 5 rejilla 3 bit 2
239	Pulso 6 rejilla 3 bit 0
240	Pulso 6 rejilla 3 bit 1
241	Pulso 6 rejilla 3 bit 2
242	Pulso 7 rejilla 3 bit 0
243	Pulso 7 rejilla 3 bit 1
244	Pulso 7 rejilla 3 bit 2
245	Pulso 8 rejilla 3 bit 0
246	Pulso 8 rejilla 3 bit 1
247	Pulso 8 rejilla 3 bit 2
248	Pulso 9 rejilla 3 bit 0
249	Pulso 9 rejilla 3 bit 1
250	Pulso 9 rejilla 3 bit 2
251	Pulso 10 rejilla 3 bit 0
252	Pulso 10 rejilla 3 bit 1
253	Pulso 10 rejilla 3 bit 2
254	Pulso 11 rejilla 3 bit 0
255	Pulso 11 rejilla 3 bit 1
256	Pulso 11 rejilla 3 bit 2
257	Pulso 12 rejilla 3 bit 0
258	Pulso 12 rejilla 3 bit 1
259	Pulso 12 rejilla 3 bit 2

Tabla 5 Continuación.

5.2.14 Protección contra errores

Para proteger los bits más significativos de cada trama contra los errores de transmisión, se pueden utilizar varias técnicas como los códigos suaves RS y la de cuadratura (QAM) <22> y <23>, entre otras, tomando la que más se adapte a la optimización de la transmisión a 16 Kbps.

El objetivo de optimizar el esquema de codificación, es lograr una realización promedio satisfactoria con una cantidad mínima de bits agregados a los 13 Kbps. De esta forma, la prevención de la protección contra errores consta de dos elementos :

- ♦ Corrección selectiva de errores hacia adelante.

Se sabe que varios parámetros de los codificadores de voz con tasas medias y bajas de bits son sensibles a errores de bit. En el codec RPE-LTP, los parámetros que más degradan la calidad de la señal de voz cuando ocurren errores son las razones de área logarítmica, debido a que su valor caracteriza la función de transferencia del conducto vocal por espacio de hasta 20 ms.

Como consecuencia, para un manejo efectivo de una cantidad limitada de bits de redundancia por trama, la salida del codec se divide en clases de acuerdo a su importancia y a diferentes grados de protección.

Por lo que el problema a resolver consiste en determinar :

- i) La cantidad total de bits de redundancia requeridos.

- ii) La cantidad necesaria de clases.
- iii) El número de bits y el grado de protección en cada clase.
- iv) El algoritmo de protección contra errores.

En lo anterior se observa que es un trabajo complejo, así que inicialmente se hace una selección preliminar para reducir el número de variables, al considerar :

- a) Tres clases.
- b) Una protección de tres niveles usando códigos de convolución.

♦ Detección de errores

Las estadísticas de error residual esperadas en la interfaz del codec, aún después de la intercalación será de un error natural, con una fuerte correlación entre las clases de bits.

En particular, cuando hay un error no corregible en las clases más sensibles, es muy probable que la trama completa del codec se borre o no se recupere. Esta situación se puede detectar al incluir unos cuantos bits de paridad en la trama del codificador y el efecto subjetivo de la trama borrada o perdida disminuye al extrapolar los parámetros de tramas de voz anteriores.

En base a esto, el procedimiento anterior solo se aplica en periodos de tiempo limitado.

5.3 Pseudocódigo de los principales módulos

Los algoritmos de las diferentes etapas del codificador se implantaron utilizando lenguaje C. En la tabla 6 se observan las

principales variables utilizadas en la implementación de los algoritmos, y a continuación se encuentra el pseudocódigo de la mayoría de las funciones que constituyen el codificador.

VARIABLE	SIGNIFICADO
s	Arreglo de 160 muestras extraídas del archivo de entrada
FAC	Coefficientes de autocorrelación
r	Coefficientes de reflexión
LAR	Parámetros LAR
LARc	Parámetros LAR codificados
LARpp_j	Parámetros LAR actualmente interpolados y decodificados
LARpp_j_1	Parámetros LAR anteriormente interpolados y decodificados
LARp	Coefficientes de reflexión estimados
d	Señal residual
bc	Factor de ganancia codificada
Nc	Retardo de correlación cruzada codificado
dpp	Señal estimada por el filtro de análisis de término largo
e	Señal residual del filtro de análisis de término largo
x	Secuencia de muestras ponderadas
XM	Secuencia óptima de pulsos regulares seleccionada
Mc	Fase de rejilla diezmada y codificada
XMc	Secuencia óptima RPE codificada
Xmaxc	Muestra máxima codificada de la secuencia RPE
XMp	Secuencia RPE decodificada
ep	Señal residual de término largo reconstruida

Tabla 6 Variables principales del programa de codificación.

```

Procedimiento INICOD
Comienza
  Lee( s ) // Lee tramas de 160 muestras de voz. //
  Llama Autocorrelación( s, FAC )
  Llama Coeficientes_de_reflexión( FAC, r )
  Llama Transformación_r_a_LAR( r )
  Llama Cuantización_codificación( LAR )
  Desde i ← 0 hasta 3 hacer
    Comienza
      Llama Interpolación_de_LAR ( LARpp_j_1, LARpp_j,
LARp )
      Llama Transformación_LARp_a_rp( LARp )
      Llama Filtro_de_análisis( s, LARp, d )
    Termina // Desde //
  Llama Cálculo_de_Parametros_LTP( d, bc, Nc )
  Llama Análisis_de_term_largo( bc, Nc, d, dpp, e )
  Llama Filtro_de_ponderación( e, x )
  Llama Selección_de_rejilla_RPE( x, XM, Mc )
  Llama Cuantización_APCM( XM, XMC, Xmaxc )
  Llama Selección_secuencia_RPE( Mc, Xmp, ep )
  Escribe( LARc, Nc, bc, Mc, xmaxc, Xmc )
Termina // INICOD //

Procedimiento Autocorrelación
Comienza
  // Búsqueda del valor máximo de la secuencia de entrada. //
  smax ← 0

```

Figura 5.4 Pseudocódigo de los principales módulos del codificador.


```

Desde k ← 0 hasta 159 hacer
Comienza
    temp ← abs( s[ k ] )
    Si ( temp > smax ) entonces
        smax ← temp
Termina      // Desde //

// Cálculo de los 8 coeficientes por el método de autocorrelación. //
p ← 8
Desde k ← 0 hasta p hacer
Comienza
    sum ← 0
    Desde i ← 0 hasta 159 - k hacer
        sum + ← s[ i ] * s[ i + k ]
    FAC[ k ] ← sum
Termina      // Desde //
Termina      // Autocorrelación //

Procedimiento Coeficientes_de_reflexión
Comienza
    // Recursión de Schur con aritmética de 16 bits. //
    Si ( FAC[ 0 ] = 0 ) entonces
        Comienza
            Desde i ← 0 hasta 8 hacer
                FAC[ i ] ← 0

```

Figura 5.4 Continuación.

```

    salir
Termina // Condición //
// Inicialización del arreglo P y K para utilizarlos en la recursión. //
Desde i ← 1 hasta 7 hacer
    K[ i ] ← FAC[ i ]
Desde i ← 0 hasta 8 hacer
Comienza
    P[ i ] ← FAC[ i ]
Desde n ← 1 hasta 8 hacer
Comienza
    temp ← P[ 1 ]
    temp ← abs( temp )
    r ← Divide( temp, P[ 0 ] )
    Si ( P[ 1 ] > 0 ) entonces
        r ← (-r)
    Si ( n = 8 ) entonces
        salir
    temp ← Multiplica( P[ 1 ], r )
    P[ 0 ] ← Suma( P[ 0 ], temp )
Desde m ← 1 hasta 8 - n hacer
Comienza
    temp ← Multiplica( K[ m ], r )
    P[ m ] ← Suma( P[ m+1 ], temp )
    K[ m ] ← Suma( K[ m ], temp )
Termina // Desde //
Termina // Desde //
Termina // Coeficientes_de_reflexión //

```

Figura 5.4 Continuación.

Procedimiento Transformación_r_a_LAR

ComienzaDesde $i \leftarrow 1$ hasta 8 hacerComienzatemp \leftarrow rtemp \leftarrow abs(temp)Si (temp < 22118) entoncestemp \leftarrow tempOtro si (temp < 31130) entoncestemp \leftarrow sgn[temp * (2 * abs(temp) - 22118)]Otrotemp \leftarrow sgn[temp * (8 * abs(temp) - 208890)]Termina // Desde //Termina // Transformación_r_a_LAR //

Procedimiento Cuantización_codificación

Comienza

LAR[0] = Paso(20480, 0, 31, -32)

LAR[1] = Paso(20480, 0, 31, -32)

LAR[2] = Paso(20480, 2048, 15, -16)

LAR[3] = Paso(20480, -2560, 15, -16)

LAR[4] = Paso(20480, 94, 7, -8)

LAR[5] = Paso(20480, -1792, 7, -8)

LAR[6] = Paso(20480, -341, 3, -4)

LAR[7] = Paso(20480, -1144, 3, -4)

Termina // Cuantización_codificación //

Figura 5.4 Continuación.

Procedimiento Decodificación_de_LAR

Comienza

LARp[0] = Pasod(0, -32, 13107)

LARp[1] = Pasod(0, -32, 13107)

LARp[2] = Pasod(2048, -16, 13107)

LARp[3] = Pasod(2560, -16, 13107)

LARp[4] = Pasod(94, -8, 19223)

LARp[5] = Pasod(-1792, -8, 17476)

LARp[6] = Pasod(-341, -4, 31454)

LARp[7] = Pasod(-1144, -4, 29708)

Termina // Decodificación_de_LAR //

Procedimiento Transformación_LARp_a_rp

ComienzaDesde i ← 1 hasta 8 hacerSi (LARp[i] < 0) entoncesComienzaSi (LARp[i] = -32768) entonces

temp ← 32767

Otro

temp ← - (LARp[i])

Si (temp < 11059) entonces temp ← temp * 2Otro si (temp < 20070) entonces

temp ← temp + 11059

Otro

temp ← Suma(temp / 4, 26112)

Figura 5.4 Continuación.

```

        LARp[ i ] ← temp
    Termina          // Condición //
    Otro
    Comienza
        temp ← LARp[ i ]
        Si ( LARp[ i ] = - 32768 ) entonces
            temp ← 32767
        Otro
            temp ← - ( LARp[ i ] )
        Si ( temp < 11059 ) entonces
            temp ← temp * 2
        Otro si ( temp < 20070 ) entonces
            temp ← temp + 11059
        Otro temp ← Suma( temp / 4, 26112 )
        LARp[ i ] ← temp
    Termina          // Condición //
    Termina          // Desde //
Termina          // Transformación_LARp_a_rp //

```

Procedimiento Filtro_de_análisis

```

Comienza
    Desde k ← 0 hasta 39 hacer
    Comienza
        Desde i ← 0 hasta 8 hacer
        Comienza
            ui ← u[ i ]

```

Figura 5.4 Continuación.

```

        rpi ← rp[ i ]
        z ← Multiplica( rpi, di )
        z ← Multiplica( rpi, ui )
        di ← Suma( di, z )
    Termina // Desde //
    s[ i ] ← di
Termina // Desde //
Termina // Filtro_de_análisis //

Procedimiento Cálculo_de_parámetros_LTP
Comienza
    lmax ← 0
    Nc ← 40
    Desde i ← 40 hasta 120 hacer
    Comienza
        sum ← 0
        Desde n ← 0 hasta 39 hacer
        Comienza
            sum ← sum + r[ 40 * k + n ] * r[ 40 * k + n - 1 ]
            Si ( sum > lmax ) entonces
                Comienza
                    Nc ← l
                    lmax ← sum
                Termina // Condición //
        Termina // Desde //
    Termina // Desde //

```

Figura 5.4 Continuación.

```
lpot ← 0
Desde k ← 0 hasta 39 hacer
  Comienza
    ltemp ← dp[ k - Nc ] / 8
    lpot ← lpot + ltemp ^ 2
  Termina // Desde //
lpot ← lpot / 2
Si ( lmax <= 0 ) entonces
  Comienza
    bc ← 0
    salir
  Termina // Condición //
Si ( lmax => lpot ) entonces
  Comienza
    bc ← 3
    salir
  Termina // Condición //
// Cuantización y codificación de la ganancia bc. //
Desde bc ← 0 hasta 2 hacer
  Comienza
    Si ( R <= Multiplica( s, DBL[ bc ] ) entonces
      salir
    bcs ← bc
  Termina // Desde //
Termina // Cálculo_de_parámetros_LTP //
```

Figura 5.4 Continuación.

```

Procedimiento Análisis_de_término_largo
Comienza
  Si ( bc = 0) entonces
    Bp ← 3277
  Otro si (bc=1) entonces
    Bp ← 11469
  Otro si (bc=2) entonces
    Bp ← 21299
  Otro
    Bp ← 32767
  Desde k ← 0 hasta 39 hacer
  Comienza
    dpp[ k ] ← Multiplica( Bp, dp[ k - Nc )
    e[ k ] ← Resta( d[ k ], dpp[ k ] )
  Termina // Desde //
Termina // Análisis_de_término_largo //

```

```

Procedimiento Filtro_de_ponderación
Comienza
  Desde k ← 0 hasta 39 hacer
  Comienza
    Lres ← 0
    i ← 0
    H ← -134
    Lres ← Lres + Paso( i, H )
    i ← 1

```

Figura 5.4 Continuación.


```
H ← -374
Lres ← Lres + Paso( i, H )
i ← 2
H ← 0
Lres ← Lres + Paso( i, H )
i ← 3
H ← 2054
Lres ← Lres + Paso( i, H )
i ← 4
H ← 5741
Lres ← Lres + Paso( i, H )
i ← 5
H ← 8192
Lres ← Lres + Paso( i, H )
i ← 6
H ← 5741
Lres ← Lres + Paso( i, H )
i ← 7
H ← 2054
Lres ← Lres + Paso( i, H )
i ← 8
H ← 0
Lres ← Lres + Paso( i, H )
i ← 9
H ← -374
Lres ← Lres + Paso( i, H )
i ← 10
```

Figura 5.4 Continuación.

```

        H ← -374
        Lres ← Lres + Paso( i, H )
        Termina // Desde //
Termina // Filtro_de_ponderación //

Procedimiento Selección_de_rejilla_RPE
Comienza
    Em ← 0
    Mc ← 0
    j ← 1
    Desde m ← 0 hasta 3 hacer
    Comienza
        Si ( m = 0 ) or ( m = 3 ) entonces
            Comienza
                j ← i + 1
                temp ← x[ m + 3 * j ] / 2
                sum ← sum + temp ^ 2
            Termina // Condición //
        Otro
            Comienza
                temp ← x[ m + 3 * j ] / 2
                sum ← sum + temp ^ 2
            Termina // Condición //
    Termina // Desde //
    Si ( sum > m ) entonces
        Comienza

```

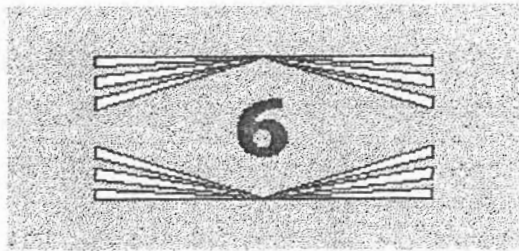
Figura 5.4 Continuación.

```

    Mc ← m
    Em ← sum
    Termina      // Condición //
Termina          // Selección_de_rejilla_RPE //

Procedimiento Cuantización_APCM
Comienza
    Desde i ← 0 hasta 12 hacer
    Comienza
        temp ← xM[ i ] / 2 ^ temp
        temp ← Multiplica( temp, temp1 )
        temp ← temp / 4096
        XMc ← temp
    Termina      // Desde //
temp ← 0
xmax ← 0
Desde i ← 0 hasta 12 hacer
Comienza
    temp ← XM[ i ]
    temp ← abs( temp )
    Si ( temp > xmax ) entonces xmax ← temp
Termina      // Desde //
temp ← xmax / 512
Desde i ← 0 hasta 5 hacer
    temp ← temp / 2
Termina      // Cuantización_APCM //
```

Figura 5.4 Continuación.



CONCLUSIONES Y PERSPECTIVAS

En este trabajo se implementó un codificador híbrido con predictor de término largo excitado con pulsos regulares (RPE-LTP)^{<24>}, derivado de la combinación de los codificadores de predicción de término largo con excitación multi-pulso (MPE-LTP), y del codificador de predicción lineal excitado con pulsos regulares (RPE-LPC).

El RPE-LTP intenta preservar la calidad de voz al reducir la tasa de transmisión de 64 a 16 Kbps, además modela el conducto vocal a través de sus parámetros utilizados para generar la señal sintetizada.

Cabe mencionar que la Conferencia Europea de Administraciones Postales y de Telecomunicaciones seleccionó al codificador RPE-LTP para el servicio de telefonía móvil celular digital y sus características

están reportadas en la recomendación GSM 06.10 y CEPT/TM3 1/03/11.

Las técnicas propuestas como alternativas de codificación orientadas a los sistemas de radio móvil celular digital, mejoran la calidad de voz ofrecida por el sistema FM con compresión analógica a 900 MHz. Debido a que cada uno de los codificadores candidatos tenía que cumplir con ciertos requerimientos.

En el apéndice se dan las principales características de los codificadores candidatos que participaron en la selección para el servicio de telefonía móvil celular digital.

Las características de los tres codificadores se observan en la tabla 1.

En los codificadores no hay mayores diferencias con lo que respecta a las siguientes funciones básicas :

- ♦ Longitud de las tramas
- ♦ Forma de análisis
- ♦ Filtro inverso

La principal diferencia está en el modelo de pulsos y en los métodos de cuantización que utilizan cada uno de ellos.

El algoritmo MPE representa las muestras de la señal residual por un pequeño número de pulsos. Las posiciones y amplitudes se optimizan para minimizar el criterio de error de ponderación. Este algoritmo también incluye un predictor de término largo para incrementar la eficiencia de codificación. Debido a lo anterior el MPE-LTP se considera como un algoritmo de codificación de complejidad alta.

Por otra parte los algoritmos RPE representan la señal residual como un tren de pulsos, seleccionado con un número menor de posiciones, pero con más pulsos que el esquema MPE, aun así, se le considera como un algoritmo de complejidad media.

CODIFICADOR	RPE-LPC	MPE-LTP	RPE-LTP
	Esquema básico	Predictor	Nuevo esquema
Tasa neta	14.8 Kbps	13.2 Kbps	13.0 Kbps
Tiempo Codificado por trama	19.5 ms.	20 ms.	20 ms.
Ventana	24.375 ms. Hamming traslapada	20 ms. Rectangular sin traslape	20 ms. Rectangular sin traslape
Orden del filtro de análisis	12	8	8
Algoritmo	Schur	Le Roux- Guegen	Schur
Codificación de coeficientes	52 bits	28 bits	36 bits
Tipo de filtro	Celosía	Forma directa	Celosía
análizador/sintetizador			
Modelo de excitación	Pulsos regulares (Método simplificado)	Multi-pulso con predictor de término largo	Pulsos regulares (Método simplificado)
Número de pulsos	52 / trama 4 bits / pulso	24/ trama 3 bits / pulso	52 / trama 3 bits / pulso

Tabla 1 Características de los codificadores RPE-LPC, MPE-LTP y del nuevo esquema RPE-LTP.

Propiedades del codec RPE-LTP

En particular, el RPE-LTP presenta las siguientes propiedades :

- ♦ Con una tasa de 13.0 Kbps, el codificador logra una calidad de voz mejor que la producida por el RPE-LPC a 14.77 Kbps, y que el MPE-LTP a 13.2 Kbps en condiciones libres de error.
- ♦ La inclusión del predictor LTP asegura un menor aumento en la sensibilidad de errores. Sin embargo, si la reducción en la tasa de transmisión se utiliza como protección contra errores, se asegura una mayor robustez.
- ♦ El retardo potencial de este nuevo esquema es de 5 ms. menos que el original (RPE-LPC), dado que no se necesita el traslapamiento entre tramas.
- ♦ La complejidad del nuevo esquema es similar a la del esquema original.

Por estas razones el CEPT decidió adoptar al nuevo esquema RPE-LTP como un estándar para los codificadores con una tasa de transmisión a 16 Kbps que operen para los sistemas móviles de telefonía celular digital.

En el diseño de un codificador se debe de considerar el ambiente hostil en el que operan los sistemas de radio móvil, esto es porque las señales pueden sufrir distorsiones o desvanecimientos en la transmisión, y para contrarrestar tales efectos se considera un canal con tasa de errores de bit (BER) de hasta 10^{-2} y continuos errores de ráfaga. En lo que respecta al comportamiento del codificador RPE-LTP operando bajo las condiciones anteriores será de una nota de

opinión promedio superior a 4.0 o incluso hasta 4.3^{<25>} al operar en buenas condiciones.

Calidad subjetiva y desviación estándar

Una forma de medir la calidad subjetiva de la señal de voz en cualquier codificador es por medio de los valores de la *nota de opinión promedio* (MOS) que son muy difíciles de duplicar al repetir un experimento (en un ambiente hostil o en buenas condiciones), y varios factores contribuyen a esta variabilidad.

Otra cantidad utilizada para asegurar el valor de una medición de la opinión es la *desviación estándar* de las opiniones proporcionadas por una población de participantes. Por ejemplo, en la codificación PCM de 64 Kbps la nota de opinión promedio es de 4.56 con una desviación estándar de 0.57^{<26>}.

Es importante aclarar que al realizar las evaluaciones de los codificadores, no se tomen en cuenta la nota de opinión promedio de las voces infantiles porque de los resultados obtenidos se observa una gran deficiencia en la calidad del codificador al utilizar dichas voces.

En sistemas de comunicación digitales, la alta calidad de voz siempre puede definirse en términos de la nota de opinión media, cuando se encuentra dentro del intervalo de 4.0 y 4.5.

Comparación de codificadores a 16 Kbps

La comparación del codec RPE-LTP únicamente se hizo con aquellos que tienen un desempeño aceptable (en cuanto a calidad de voz

se refiere) y que están diseñados para transmitir a una velocidad de 16 Kbps.

Es importante señalar que en nuestro país no se le ha dada mucha importancia al desarrollo, implementación y mejoras de las técnicas de codificación por tal motivo los trabajos realizados son realmente escasos o no existen.

Debido a lo anterior, el codificador se comparó con otros desarrollados en Europa o Estados Unidos. Esta comparación básicamente se hace en base a la nota de opinión promedio, complejidad del algoritmo y aspectos de implementación.

Los codificadores con los que se comparó el codec RPE-LTP son :

- ♦ Predictor de término largo con excitación multi-pulso (MPE-LTP).
- ♦ Codificador de predicción lineal excitado con pulsos regulares (RPE-LPC).
- ♦ Codificador excitado por código o codificador estocástico (CELP).

El codificador MPE-LTP utiliza más ciclos para procesar sus instrucciones, lo cual le lleva un poco más de tiempo para realizar la síntesis de voz . Aunque tiene la ventaja de utilizar menos pulsos que cualquiera de los codificadores RPE para excitar al filtro de síntesis, su MOS es de 3.27.

Por otra parte, el codificador RPE-LPC combina la simplicidad de un codificador RELP (Predictor Lineal Excitado con la señal Residual) en banda base con la alta calidad de voz de otros esquemas de codificación, de esta forma es como obtiene una mejora en su

calidad de voz sintetizada y una MOS de 3.54 , un poco más superior al anterior.

En el codificador estocástico (CELP), se analiza la señal de excitación para modelar la diferencia p rceptual ponderada entre la seal de voz y su predicci n. Los procedimientos eficientes incluyen algoritmos r pido desarrollados para encontrar la excitaci n  ptima desde una librer a que contiene los vectores de excitaci n.

En t rminos de MOS muestra una baja puntuaci n debido particularmente al hecho de que su degradaci n inherente se toma como ruido de codificaci n por los escuchas, es por eso que su MOS es de 3.8 operando en buenas condiciones.

A manera de comentario, este codificador es uno de los m s prometedores para transmitir a velocidades m s bajas, manteniendo una calidad comparable con la lograda por sistemas que transmiten a mayor velocidad. Solo que hay una peque a desventaja : la implementaci n de su algoritmo en tiempo real.

Con el RPE-LTP, la mejora de la calidad de la seal s ntetica producida se logra al modificar el esquema b sico RPE-LPC, incluyendo un predictor de t rmino largo para tener mayor robustez contra los errores de transmisi n o recepci n, y su complejidad es similar a la del RPE-LPC. La diferencia b sica entre los dos codificadores esta en que el RPE-LPC solo utiliza tres secuencias candidatas para la excitaci n del filtro de s ntesis y el RPE-LTP utiliza 4 secuencias lo que a su vez da una mejor calidad de voz sintetizada sin aumentar la complejidad en el calculo de las secuencias. El MOS logrado con codificaci n RPE-LTP es de 4.3 operando en buenas condiciones de transmisi n.

En base al MOS de cada uno de los codificadores anteriores, se observa en el codificador RPE-LTP la nota de opinión más alta, la cual se encuentra dentro del rango de un codificador de alta calidad y por ello su uso es totalmente recomendable como un sistema de comunicación comercial para sistemas de radio móvil o de telefonía, por lo que puede decirse que se logró el objetivo de implementar un codificador con buena calidad de voz sintetizada y aunque sea un buen esquema realizado no se descartan las mejoras que se le pudieran hacer para funcionar más eficientemente, esperando que estas mejoras ayuden a contribuir a lograr nuevos esquemas futuros con tasas de transmisión más bajas para eficientar el ancho de banda en las transmisiones sin tener mucha degradación en la calidad de la señal reproducida. En la referencia²¹ se encuentran las pruebas subjetivas realizadas, el proceso para llevarlos a cabo y los resultados obtenidos.

Análisis de la señal de voz con el software MATLAB

A través del desarrollo del codec se hicieron varias pruebas para analizar el comportamiento de la señal de voz en cada uno de los bloques, pero solo se explicará un bloque, porque en todos los demás se lleva un procedimiento similar para analizar las gráficas de la señal y los errores que en ella se pudieran encontrar.

El software utilizado fue el MATLAB y a continuación se explica el bloque de autocorrelación.

- ♦ Primeramente, se leyó el área de datos del archivo [wav] colocándose en vectores de 1024 muestras para su análisis. Se tomo el tamaño de 1024 muestras porque es la máxima cantidad que aceptan los buffers del MATLAB y que pueden leer o escribir.

- ♦ De cada vector de 1024 muestras, se tomaron 20 ms. de señal de voz y se autocorrelaron con la ayuda de una función predefinida (no todas las funciones del codec se encuentran en las librerías del MATLAB, la mayoría de ellas hay que implementarlas).

- ♦ De la autocorrelación anterior se obtienen 9 coeficientes de reflexión, ya que se utiliza un filtro de orden 8. En esta señal se observaron dos detalles importantes dependiendo del segmento de señal procesada :

a) Si el segmento de 20 ms. correspondía a una señal periódica entonces la señal presentaba un periodo fundamental y múltiplos del mismo bien definidos.

b) En cambio, si el segmento correspondía a un sonido fricativo, el periodo ni los múltiplos del mismo tenían una buena definición, es decir, no son tan claros como en el tramo de señal sonora.

Detalles como estos son los que se observan al procesar la trama a través de los diferentes bloques del codec. También por medio de las gráficas y los resultados se observa el comportamiento de la señal a la salida de cada proceso para saber si se está realizando lo adecuado o no y corregir los posibles errores.

Mejoras en el codificador RPE-LTP

El codificador presenta algunos factores que se pueden mejorar, como es el caso del volumen de la señal sintetizada, ya que casi disminuye entre un 30 y 40 por ciento con respecto a la señal original. En forma de sugerencia, para compensar esta pérdida se puede agregar un amplificador al final del filtro inverso de tal forma que aumente la ganancia de la señal sintetizada.

Otra mejora es la selección adecuada del algoritmo de protección contra errores para la implementación en tiempo real, lo cual asegura una mayor robustez en la transmisión-recepción de tramas de voz.

También puede disminuirse de 3 a 2 la posición de la rejilla en el vector de excitación y observar si hay mejoras o no en la señal producida.

Implementación en tiempo real

El algoritmo RPE-LTP implementado tiene una ventaja importante con lo que respecta en su aplicación, porque puede implementarse directamente en cualquier tarjeta de procesamiento digital de señales como es el caso de las tarjetas de la familia TMS320CXX, para propósitos de transmisión en tiempo real sin ningún problema ya que solo se necesitaría configurar la tarjeta y no el algoritmo.

Problemas encontrados en la implementación del codec

Hubo varios problemas en la implementación del codec, como por ejemplo, la selección adecuada del algoritmo RPE, ya que el algoritmo original presenta variantes que afectan de manera directa la relación señal a ruido de la señal, lo cual influye en la degradación de la propia señal. Aunado a lo anterior cada variante del RPE utiliza una cantidad considerable de operaciones para resolver las matrices cuyo producto final es el vector de excitación óptimo que contiene la máxima energía para excitar el filtro de síntesis.

Por lo anterior se necesitó de otro algoritmo RPE para evitar el aumento de la complejidad en el codec, aplicándose el RPE versión

simplificada. En este algoritmo se aplica un diezmado (desplazamiento) por un factor de tres a cada sub-segmento seleccionado y directamente se extrae la secuencia de máxima energía sin utilizar matrices.

Otro problema surgió cuando se implementaron los diferentes cuantizadores de parámetros, como resultado de un diseño no adecuado. El problema consistió también en que el proceso ocupaba una cantidad de tiempo considerable, por lo que se rediseñaron todos los cuantizadores para optimizar el tiempo de ejecución.

En la lectura del archivo [wav] se debe de comprobar de alguna manera la cantidad de bits que ocupa su cabecera, porque su longitud de bytes difiere con las cantidades encontrados en algunas referencias y esto puede ser problema para que el área de datos no se lea ni se grabe correctamente y aunque el archivo este bien procesado por el codec no lo podrá leer la grabadora de sonidos y no se puede saber si la señal reproducida esta libre de errores hasta que se hagan las correcciones.

La lectura de las muestras puede causar problemas en el proceso de cada trama si no se hace considerando la cantidad de bits asignados a cada una de ellas, es decir, si cada muestra se grabo con 8 bits, la muestra se podrá leer como byte, pero si se grabo con 16 bits para tener una mejor resolución de audio entonces se debe de manejar como un entero.

Aplicación del codec en medios telefónicos comerciales

Finalmente, una aplicación real e inmediata que puede tener el codec implementado sería su implantación en cualquiera de las centrales telefónicas distribuidas en nuestro país, con lo cual se estaría eficientando el ancho de banda de los canales de recepción y transmisión de dichos sistemas, sin mayor degradación en la señal de voz que es el objetivo fundamental que se persigue en los codificadores enfocados a telefonía comercial. Además, la implantación no resultaría costosa porque solo se deben adquirir dos tarjetas de procesamiento digital de señales para utilizarlas en la codificación o decodificación de las tramas transmitidas o recibidas, y configurarlas, lo cual no resulta muy difícil.

APENDICE

En este apartado se describen los criterios utilizados para seleccionar el codificador de voz de los sistemas de radio móvil celular digital, y una comparación entre las alternativas propuestas.

Objetivos del sistema

- ♦ Desde el punto de vista del usuario, el requerimiento básico en cuanto a la calidad promedio, debe ser al menos tan buena como la que ofrece el sistema analógico a 900 MHz.
- ♦ El sistema no debe alterar, de manera significativa, a la red telefónica o a la infraestructura existente para los servicios móviles celulares, además debe ser compatible con la red Digital de Servicios Integrados.

- ♦ La frecuencia de muestreo tendrá que ser de 8 KHz, lo cual permitirá convivir fácilmente con la Red Pública de Telefonía Conmutada.

- ♦ Debido a las conversiones de 4 a 2 hilos, ocurren reflexiones en los extremos de una conexión telefónica. Normalmente, estas reflexiones se atenúan lo suficiente en la red y no causan problemas. Sin embargo, si el retardo de la conexión en un sentido excede los 50 ms., no será suficiente la atenuación de la red para evitar ecos que ocasionen problemas a los usuarios. El retardo global en un sistema de radio móvil, tiene dos contribuciones principales:

- i) Para lograr una calidad de voz aceptable a tasas de transmisión por debajo de 16 Kbps, generalmente se utilizan esquemas de codificación de bloque. Por lo tanto, se espera que el retardo de transmisión sea considerable, dependiendo del tamaño del bloque, la complejidad del algoritmo y la velocidad del procesador.

- ii) Es necesario utilizar las técnicas de intercalado de bits para combatir el efecto de los errores de ráfaga. Esta es una técnica que requiere que los bits a transmitir, se almacenen y transmitan en una nueva secuencia para dispersar los errores. El retardo que se introduce de esta forma puede ser de varias tramas del codificador de voz.

Con el fin de evitar el retardo excesivo se fijó un límite no superior de 65 ms. para cada uno de estos componentes.

Los seis codificadores candidatos que se presentaron fueron evaluados por El Grupo Europeo de Expertos en codificación de voz, y se dividen en dos clases : Codificadores de Sub-bandas y Codificadores Excitados por Pulsos.

PARAMETRO	SBC-APCM<27>	SBC-APCM<28>	SBC-APCM<29>	SBC-ADPCM<30>
Señales de las Sub-Bandas	14 Kbps	14 Kbps	10 Kbps	15 Kbps
Codificación	Max	Max	Max	ADPCM
Asignación de bits	Adaptiva 1 - 4	Adaptiva 0, 2 - 5	Adaptiva 0 - 5	Banda 1-6 : 4, 3, 2, 2, 2, 2, bits
Información lateral	1 Kbps	1 Kbps	3 Kbps	0 Kbps
Representación espectral	Cuantización vectorial	Cuantización vectorial	PCM logarítmico	
Tamaño del "recetario"	128 x 6 bits	128 + 64 x 8 + 4 + 4 bits	14 x 3 bits	
Tasa neta	15 Kbps	15 Kbps	13 Kbps	15 Kbps
Sincronía				1 Kbps
Protección contra errores	1 Kbps	1 Kbps	3 Kbps	
Longitud de la Trama	15 ms	20 ms	16 ms	1 ms
No. de bandas Total/en uso	8 / 6	16 / 14	16 / 14	8 / 6
Tipo de filtro	QMF	Paralelo	Paralelo	QMF
Etapas	32, 24, 16	64	80	32, 24, 16
Ganancia del bloque	PCM Ley A	Log PCM	Log PCM	

Tabla 1 Parámetros de los algoritmos SBC.

Codificadores de sub-bandas

- ♦ Tres variantes del codificador de Sub-Bandas con codificación PCM adaptivo de bloque para la señal de las sub-bandas (SBC-APCM), y
- ♦ Un codificador de Sub-Bandas que utiliza codificación adaptiva hacia atrás ADPCM para la señal de las sub-bandas (SBC-ADPCM).

Los parámetros principales de estos codificadores se exhiben en la tabla 1.

PARAMETRO	RPE-LPC <31>	MPE-LTP <32>
Tasa neta	14.77 Kbps	13.2 Kbps
Protección contra errores	1.23 Kbps	2.8 Kbps
Trama de voz	19.5 ms	20 ms
Ventana	24.375 ms	20 ms
Orden del filtro de análisis	12	8
Algoritmo	Schur	Le Roux Gueguen
Codificación de coeficientes	52 bits	28 bits
Estructura del filtro	Celosía	FIR directa
Modelo de excitación	Pulsos regulares (Método Simplificado)	Multi-pulso con predicción de término largo
No. de pulsos/trama	52	24

Tabla 2 Parámetros de los codificadores con excitación de pulsos.

Codificadores excitados por pulsos

Participaron dos codificadores basados en la técnica de predicción lineal :

- ♦ Un codificador por predicción lineal con excitación simplificada de pulsos regulares (RPE-LPC), y
- ♦ Un codificador con excitación multi-pulso con predicción de término largo (MPE-LTP).

En la tabla 2 se tienen las principales características de estos dos codificadores.

Comparación entre sistemas

En las siguientes tablas se observan las características de calidad, retardo y complejidad de los algoritmos anteriores. Es conveniente recordar que las condiciones de evaluación están sujetas al funcionamiento en sistemas de radio móvil.

CODIFICADOR	NOTA DE OPINION PROMEDIO
SBC-APCM <27>	2.98
SBC-APCM <28>	2.46
SBC-APCM <29>	3.14
SBC-ADPCM <30>	2.92
RPE-LPC <31>	3.54
MPE-LTP <32>	3.27
FM Compresión analógica	1.95

Tabla 3 Calidad de voz promedio.

CODIFICADOR	LONGITUD DE LA TRAMA	RETARDO SUPONIENDO TRANSMISION SERIE
SBC-APCM <27>	15 ms	< 40 ms
SBC-APCM <28>	20 ms	< 45 ms
SBC-APCM <29>	16 ms	35 ms
SBC-APCM <30>	1 ms	7 ms
RPE-LPC <31>	19.5 ms	< 40 ms
MPE-LTP <32>	20 ms	36 ms

Tabla 4 Propiedades de retardo.

CODIFICADOR	RAM	ROM	SUMAS	MULTIPLIC.	TOTAL
SBC-APCM <27>	0.820 K	3.40 K	0.6 M	0.6 M	1.2 M
SBC-APCM <28>	1.660 K	5.50 K	0.9 M	0.5 M	1.4 M
SBC-APCM <29>	0.350 K	3.50 K	0.8 M	0.7 M	1.5 M
SBC-APCM <30>	0.450 K	2.20 K	1.0 M	0.9 M	1.9 M
RPE-LPC <31>	0.550 K	4.50 K	0.8 M	0.7 M	1.5 M
MPE-LTP <32>	0.340 K	12.6 K	3.5 M	1.4 M	4.9 M

Tabla 5 Complejidad y requerimientos de memoria.

GLOSARIO TECNICO

En este glosario se encuentran algunos de los términos más frecuentemente utilizados en el medio de codificación de voz.

Banda base : La banda de frecuencia que una señal ocupa cuando se genera inicialmente.

BER (Bit Error Rate) : Fracción de una secuencia de bits del mensaje que contienen error.

Bits de redundancia : Bits extra insertados en una cadena de datos digitales, para habilitar la detección o corrección de errores.

Codec : Abreviación de codificador-decodificador.

Código convolucional : código en el cual no hay correspondencia de bit a bit entre bits de datos y bits codificados.

Codificadores de forma de onda : Son aquellos codificadores que tratan de reproducir lo más fiel posible la forma de onda de la señal. Ejemplos de estos codificadores son PCM, DM, DPCM, ADM, y ADPCM, entre otros.

Codificadores de fuente : Codificadores que extraen las características o parámetros más importantes de la señal. Entre ellos encontramos a los siguientes codificadores . LPC, MPE, MPE-LTP, CELP, y RELP.

Compansión : Proceso por el cual el rango dinámico de voz se comprime antes de transmitirla y se expande después de su detección. El nombre es una contracción de compresión y expansión. Las leyes de compansión estandarizadas por el CCITT son la ley μ utilizada en Estados unidos y Japón, y la ley A utilizada en Europa y el resto del mundo. Estas leyes permiten comprimir o expandir los 13 bits de cuantización lineal a 8 bits de cuantización logarítmica obteniendo la misma calidad de la señal y utilizando menos bits para cuantizar cada una de las muestras.

Corrección de error : La corrección de bits en una trama de datos digitales que se ha dañado durante la transmisión.

Cuantización : Proceso para transformar muestras analógicas de una señal en un número finito de posibles valores.

Detección de error : Detección de bits en una trama de datos digitales que se ha dañado durante la transmisión.

Enmascaramiento de ruido : Cuando un sonido evita la percepción de otro se dice que lo enmascara. El enmascaramiento depende del nivel relativo de ambos sonidos y de sus componentes frecuenciales. El enmascaramiento de un tono por otro es más efectivo cuando sus frecuencias son próximas.

Errores de ráfaga : Errores que ocurren en bits adyacentes.

Filtro de análisis : Filtro de predicción cuya entrada es la secuencia de muestras de voz original y su salida es el error de predicción.

Filtro de síntesis : Es el filtro de análisis en forma inversa.

GSM (Grupo Especial de Móviles) : Grupo establecido en 1982 por la CEPT, Conferencia Europea de Administraciones Postales y de Telecomunicaciones y tiene como finalidad definir un único sistema de comunicaciones móviles para Europa.

Ley μ : En la ley de compansión μ , la característica del compresor $c(x)$, es continua, aproximando una dependencia lineal en x para bajos niveles de entrada y una logarítmica para altos niveles. Específicamente se describe por :

$$\frac{c(x)}{x_{max}} = \frac{\ln(1 + \mu|x|/x_{max})}{\ln(1 + \mu)} \quad 0 \leq \frac{|x|}{x_{max}} \leq 1$$

Un valor práctico para μ es 255.

Ley A : En la ley de compansión A, la característica del compresor $c(x)$ es discreta, formada de un segmento lineal para bajos niveles de entrada y un segmento logarítmico para altos niveles. Esta ley se describe como :

$$\frac{c(x)}{x_{max}} = \begin{cases} \frac{A(x)/x_{max}}{1 + \ln A} & 0 \leq \frac{|x|}{x_{max}} \leq \frac{1}{A} \\ \frac{1 + \ln(A|x|/x_{max})}{1 + \ln A} & \frac{1}{A} \leq \frac{|x|}{x_{max}} \leq 1 \end{cases}$$

Un valor práctico para A es 87.56.

Pitch : Es el periodo fundamental característico de cada persona y su unidad de medida son los Hz.

Recursión de Levinson-Durbin : Este algoritmo permite resolver un sistema de ecuaciones de una manera eficiente. La idea central se basa en calcular el predictor de primer orden, $p=1$, y a partir de él los de ordenes superiores, hasta el orden p .

S/N : Relación señal-ruido, y es la proporción de ruido que hay en la señal.

Señal fricativa : Intervalos de señal con apariencia ruidosa. En estos intervalos las cuerdas vocales no vibran.

Señal residual ($e(n)$) : Señal generada por medio de un filtro de análisis, y es la diferencia entre la señal original y la predicha.

Señal sonora : Intervalos de señal localmente periódicos que se producen cuando las cuerdas vocales están vibrando.

Trama : Segmento con una cierta cantidad de muestras de la señal a ser analizada o sintetizada y que normalmente se encuentra en el rango de 15 a 50 ms.

Voz sintética : Señal de voz producida a la salida de un filtro de síntesis, caracterizado con parámetros extraídos de la señal de voz original.

<8> G. Cannalire, F. Rosa

A 64/32 Kbps single chip converter using encoding technique
in accordance with CCITT G.721 recommendation
Signal Processing III : Theories and Applications,
EURASIP 1986, pp. 415-418.

<9> Moreno Pérez Sofía

Introducción al codificador estocástico excitado por código
en el dominio temporal.
Dpto. de Señales, Sistemas y Radiocomunicaciones ETSI.

<10> J. Makhoul

Lineal Prediction : A tutorial Review
Proc. IEEE, Vol. 63, No 4, Apr. 1975, pp. 561-580.

<11> B. S Atal, J. R. Remde

A new model of LPC excitation for producing
natural-sounding speech at low bit rates
Proc. IEEE, Int. Conf. Acoust. Speech Signal Processing,
Apr. 1982 pp. 614-617.

<12> C. Galand, E. Lancon, M. Rosso, J. Mendez

A new architecture of multi-pulse excited
linear predictive coder
Signal Processing III : Theories and applications,
EURASIP, 1986, pp. 435-438.

<13> P. Kroon, E. F. Deprettere, R. J. Sluyter

Regular-Pulse Excitation - A novel Approach to Effective and
Efficient Multipulse Coding of Speech
IEEE ASSP-34, No. 5, Oct. 1986, pp. 1054-1063.

- <14> M. R. Schroeder, B. S. Atal
Code-Excited Linear Prediction (CELP)
High-Quality speech at very Low Bit Rates
Proc. Int. Conf. on Acoustics, Speech and Signal
Proc., Vol. 3, Mar. 1985, pp. 937-940.
- <15> J. E. Natving
Pan-European speech coding standard
for digital mobile radio
Speech Communication, Jul. 1988, pp. 113-123.
- <16> P. Vary, R. Hofmann, K. Hellwing, R. J. Sluyter
A Regular-Pulse Excited Linear Predictive Codec
Speech Communication, Jul. 1988, pp. 209-215.
- <17> J. Schur
Über Potenzreihen, die im Innern des
Einheitskreises beschränkt sind
J. Reine angewandte Mathematik,
Band 147, 1917, pp. 205-232.
- <18> J. L. Roux, C. Gueguen
A Fixed Point Computation of Partial Correlation Coefficients
IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-25
No. 4, Jun. 1977, pp. 257-259.
- <19> R. Viswanathan, J. Makhoul
Quantization properties of transmission parameters
in linear predictive systems
IEEE Trans. Acoust., Speech, Signal Processing,
Vol. ASSP-23, Jun. 1975, pp. 309-321.

<20> J. Max

Quantization for minimum distortion

IEEE Transactions on Information Theory, Mar. 1960, pp. 7-12.

<21> Ramírez Jiménez Miguel A.

Decodificador de voz a 16 Kbps utilizando la técnica RPE-LTP

Tesis profesional, UTM, Huajuapán de León, Oaxaca, 1997.

<22> Atungiri S. A., Sweeney P., Saheili R., Kondoz A. M.,

Evans B. G.

Soft RS for half rate GSM channel

Journal Electronics Letters, Vol. 27, No 2, pp. 176-178, 1991.

<23> Weeb W., Hanzo L., Salami R. A., Steele R.

Does 16-QAM Provide Alternative to a Half-Rate GSM
Speech Codec ?

41 st. IEEE vehicular technology conference, Gateway to the
future technology in motion (cat, No. 91CH2944-7)
pp. 511-516, 1991.

<24> P. Vary, R. J. Sluyter, C. Galand and M. Rosso

RPE-LTP Codec - The Candidate for the GSM

Radio Communication system

Int. Conf. On Digital Land Mobile Radio Communications,
30 June - 3 July, Venice, pp. 507-516, 1987.

<25> G. S. de Brito

Low Byte Rate Speech for the Pan European Mobile
Communication System

IEEE Int. Conf. on Digital Land Mobile Radio Comm.
1988, pp. 147-152.

-
- <26> N. S. Jayant, P. Noll
Digital Coding of Waveforms
Principles and Applications to Speech and Video
Prentice Hall.
- <27> Centro Studi e Laboratori Telecomunicazioni (CSELT).
- <28> Electronics Laboratory (ELAB).
Technical University of Trondheim, Norway.
- <29> Ellementel Utveekling AB, Sweeden.
- <30> Britsh Telcom, U. K.
- <31> Philips Kcommunications Industrie AG, Nürnberg, F. R. G.
- <32> IBM Laboratory, La Graude, France.