UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

PROGRAMA DE MAESTRÍA Y DOCTORADO EN INGENIERÍA
INGENIERÍA ELÉCTRICA - TELECOMUNICACIONES

COMPRESIÓN DE VOZ PARA LA TRANSMISIÓN DE COMUNICACIONES INALÁMBRICAS DE BANDA ANCHA

T E S I S
QUE PARA OPTAR POR EL GRADO DE:
MAESTRO EN INGENIERÍA

PRESENTA:
ING. JUAN CARLOS ZÁRRAGA ESTRADA

TUTOR PRINCIPAL
DR. VICTOR GARCIA GARDUÑO,
FACULTAD DE INGENIERÍA

MÉXICO, D. F. MAYO 2013

Jurado Asignado

Presidente:	Dr. Gomez Castellanos Javier
Secretario:	Dr. Psenicka Bohumil
Vocal:	Dr. García Garduño Víctor
1er. Suplente:	Dr. Matías Maruri José María
2do. Suplente:	Dr. Escalante Ramírez Boris

Lugar o lugares donde se realizó la tesis: Ciudad Universitaria, México, D.F.

TUTOR DE TESIS:

Dr. García Garduño Víctor

_____

Firma

Agradecimientos

A mi madre Juliana, quien me ha ayudado a encontrar la luz cuando todo es oscuridad.

A mi padre Juan Manuel, quien me ha aportado tantos consejos y valores durante mi andar en la vida.

A mis hermanos Guillermo y Sergio, quienes me han permitido aprender de ellos y me han alentado a lograr alcanzar mis metas.

A mis amigos, que nos apoyamos mutuamente en nuestra formación profesional.

A ti Gaby, que me has permitido caminar a tu lado durante este tiempo y por mostrarme que es mejor sonreírle a la vida. Y sobre todo gracias por ser mi inspiración.

A todos aquellos que participaron directa o indirectamente en la elaboración de esta tesis.

¡Gracias a ustedes!

Índice general

Resumen

A pesar de la aparición de sofisticados servicios multimedia, las comunicaciones de voz siguen siendo el medio predominante de las comunicaciones humanas. Con la introducción de servicios inalámbricos por internet a favorecido la transmisión de señales de datos y voz (VoIP- Voz sobre el protocolo IP).

El presente trabajo muestra diferentes técnicas de compresión de voz, siendo el principal propósito de los codificadores de voz el buscar mantener una mejora en la inteligibilidad y naturalidad de la voz, sin perder de vista la intensión de reducir la tasa de bits con el uso de algoritmos más robustos para la transmisión, almacenamiento o reproducción sobre redes inalámbricas.

En el primer capítulo se revisara las características generales de una señal de voz, así como la clasificación de los codificadores teniendo en cuenta medidas cualitativas y cuantitativas.

En el segundo capítulo se presentan diversas técnicas de codificación ideadas principalmente durante la década de 1980. Siendo las bases de la codificación que se utilizaran en las comunicaciones móviles para la cuarta generación.

En el tercer capítulo se muestran algunas técnicas de codificación cuyo auge comenzó con la telefonía celular. Recientemente con el uso de sistemas de comunicaciones inalámbricos de banda ancha ha permitido el empleo de aplicaciones innovadoras y servicios para sistemas móviles usados en la cuarta generación.

Si bien el ancho de banda que se utiliza se ha visto sobre explotado por la alta demanda es importante buscar a través de los códec híbridos un eficiente uso de éste.

Finalmente, en el cuarto capítulo se mostraran los requerimientos usados para realizar la comparativa entre distintos codificadores de voz, donde se adecuaron algunos archivos para llevar a cabo las evaluaciones de los diferentes codificadores a tasas de muestreo distintas y los resultados obtenidos fueron evaluados bajo criterios objetivos y subjetivos.

Contribución

Elaborar la programación de las evaluaciones objetivas y subjetivas en matlab, además de realizar la adecuación de algunos archivos de código previamente programados. Con el empleo de esta programación se realiza nuestro análisis comparativo entre las distintas técnicas de codificación, de una manera que nos facilita distinguir el comportamiento y percepción de la voz.

Objetivo

- Comprender y analizar distintas técnicas de compresión para la voz
- Evaluar a través de medidas objetivas y subjetivas la calidad de la voz reconstruida.

Capítulo 1

Antecedentes

La transmisión de señales aleatorias como la voz sobre los sistemas de radio comunicaciones es bastante compleja, pues existe una serie de inconvenientes, entre los que se encuentran la escasez de espectro disponible y los efectos nocivos de la propagación de la onda. El primer problema hace necesaria la eficiencia espectral, que obliga a establecer la comunicación con un flujo de datos de baja tasa, manteniendo un nivel de calidad comparable a la telefonía convencional (cuando no se presentan errores) y con mucha menor velocidad de transmisión para mantener un buen rendimiento, esto se logra con la codificación de la voz.

Siendo la UIT-T (Sector de Normalización de las Telecomunicaciones de la UIT - Unión Internacional de Telecomunicaciones) el órgano que estudia los aspectos técnicos, de explotación y tarifarios, y publica recomendaciones sobre los mismos, con miras a la normalización de las telecomunicaciones en el plano mundial.

En el primer capítulo revisaremos brevemente las características generales de una señal de voz, así como la clasificación de los codificadores. Se detallara el uso de la modulación por pulsos codificados para digitalizar una señal. Finalmente, se citaran algunos factores subjetivos y objetivos como medidas cualitativas y cuantitativas.

CAPÍTULO 1. ANTECEDENTES

1.1 Codificación de la voz

El objetivo de la codificación es la compresión de la señal, esto es, emplear el menor número de bits posibles para la representación digital de la señal de voz. La eficiente representación digital de la señal de voz hace posible conseguir la eficiencia del ancho de banda en la transmisión de la señal sobre una gran variedad de canales de comunicación.

Se entiende como codificación a la transformación de la señal analógica de voz en una señal digital para su transmisión, almacenamiento o reproducción, es decir, es el proceso mediante el cual se representa una muestra cuantificada, a través de una sucesión de números binarios. Existe un proceso inverso que se denomina decodificación, mediante el cual se reconstruyen la señal analógica, a partir de la señal numérica procedente.

1.2 Características de una señal de voz

La señal de voz ocupa un ancho de banda de 4 kHz, pero las componentes de alta frecuencia no son significativas para la comprensión de la voz por lo que en las redes telefónicas únicamente son indispensables la presencia de armónicos cuya frecuencia se halle entre 100 y 3400 Hz.

La energía de la voz está contenida en su mayor parte en las bajas frecuencias, lo que implica la calidad de las palabras. El pitch (frecuencia fundamental) varía de acuerdo a la persona. En voz masculina está ubicado en el rango de 50-250 Hz y para la voz femenina en el rango de 200-400 Hz.

Una señal de voz es considerada no-estacionaria. Sin embargo, si se toman pequeñas porciones de señales de voz cada 20 ms, la señal puede ser considerada estacionaria.

1.3 Clasificación de Códecs de voz

Los distintos métodos de codificación tratan de eliminar la redundancia de la señal y así poder reducir al mínimo el número de bits usados para codificar cada muestra. Los métodos de codificación de la voz pueden clasificarse en *codificación de forma de onda, vocoder y codificación híbrida.* Sus diferencias básicas explicitas comienzan con la calidad de la voz frente al rendimiento de tasa de bits en términos cualitativos.

En general, los códecs de forma de onda *(Waveform Coding)* son diseñados para codificar señales como tonos, conversación e incluso música. Estos pueden a su vez ser subdivididos en códecs de forma de onda en el dominio del tiempo y en el dominio de la frecuencia. Utilizan información redundante de las muestras de voz, de tal forma que no permiten una codificación eficiente, pero si una alta calidad de voz frente a la tasa de bits, en el orden de 32 kbps. Por tal motivo, no son útiles cuando se quiere codificar a bajas tasas de bits.

La señal codificada más representativa de forma de onda en el dominio del tiempo es la ley A (*A-law*), en el esquema de modulación por pulsos codificados (PCM), el cual ha sido estandarizado por la Unión Internacional de Telecomunicaciones (ITU) a 64 kbps, recomendación G711[13]. También es conocido el Diferencial Adaptativo PCM (ADPCM) a 32 kbps, estandarizado por la ITU en la recomendación G.726 [14].

Dentro de la codificación de forma de onda en el dominio de la frecuencia, puede subdividirse por sub-banda y por transformada. Para el caso de la sub-banda, la señal se somete a un corto tiempo de análisis espectral, donde se divide la señal en un número de componentes en frecuencias separadas y se codifican independientemente.

En el caso de la voz se emplean más bits para las frecuencias bajas con el fin de preservar el pitch y la información de los formantes. En el caso de la codificación de forma de onda en el dominio de la frecuencia por transformada consiste en una transformación por bloques, de forma que se realiza una transformación a un dominio diferente y se codifican los coeficientes de la transformación. Por ejemplo se utilizan las técnicas DCT (Discrete Cosine Transform) y DFT (Discrete Fourier Transform).

Tipo	Algoritmo de codificación
Codificadores de forma de onda	PCM (Pulse-Code Modulation) APCM (Adaptive PCM) DPCM (Differential PCM) ADPCM (Adaptive DPCM) DM (Delta Modulation) ADM (Adaptive DM) CVSD (Continuously Variable-Slope DM) APC (Adaptive Predictive Coding) SBC (Sub-band Coding) ATC (Adaptive Transform Coding)

Tabla 1.1: Clasificación de algunos Codificadores de forma de onda.

El principio de funcionamiento de los vocoders (*VOice CODERS*) es producir una señal de voz lo suficientemente entendible para transmitir un mensaje, independientemente de si la forma de onda se parece o no a la voz original.

El vocoder más utilizado es el de predicción lineal LPC (Linear Predictive Code), que se basa en obtener cada muestra a partir de una combinación lineal de las anteriores, con un filtro todo polos para modelar el tracto vocal. Los codificadores de fuente asumen un modelo para la señal de voz inteligible a muy baja tasa de bit, pero la calidad es baja lo que ocasiona que no suene natural.

Tipo	Algoritmo de codificación
Vocoders	Canal, Formante, Fase, Cepstral o Homomórfico LPC (Linear Predictive Coding) MELP (Mixed-Excitation Linear Prediction) STC (Sinusoidal Transform Coding) MBE (Multiband Excitation), MBE mejorada.

Tabla 1.2: Clasificación de algunos Vocoders.

La codificación híbrida es una mezcla de los dos tipos anteriores, mezclando la alta capacidad de compresión de los vocoders con la gran calidad de reproducción de los codificadores de forma de onda. Lo cual produce una señal de buena calidad con tasas de bit medias a bajas. Utilizan un modelo paramétrico de producción de la voz se trata de preservar las partes más importantes de la señal de entrada.

Presenta una degradación aceptable en presencia de ruido y errores de transmisión. La codificación se puede llevar a cabo en el dominio del tiempo como en el de la frecuencia.

Tipo	Algoritmo de codificación
Codificadores Híbridos	MPLP (Multipulse-Excited Linear Prediction) RPE (Regular Pulse-Excited linear prediction) RELP (Residual-Excited Linear Prediction) VSELP (Vector-Sum Excited Linear Prediction) CELP (Code-Excited Linear Prediction) ACELP (Algebraic CELP) CS-ACELP (Conjugated Structure ACELP)

Tabla 1.3: Clasificación de algunos Codificadores Híbridos.

1.4 PCM (Modulación por Pulsos Codificados - Pulse Code Modulation)

Muestreo

La modulación digital por pulsos codificados (*Pulse Code Modulation –PCM*) está basado en el teorema de Nyquist, donde se establece que una señal es muestreada uniformemente al menos al doble de la componente más alta de frecuencia. La componente en frecuencia de la señal de voz es de 4 kbps, por lo que se necesita muestrear la señal a 8000 muestras/s cada 1/8000 veces en un segundo (125μs), los valores muestreados son aún analógicos, por lo tanto, necesitan cuantizarse en un número fijo de niveles. Considerando 256 niveles de cuantización y cada muestra se puede representar por 8 bits. Así, un segundo de señal de voz puede representarse en 64kbits.

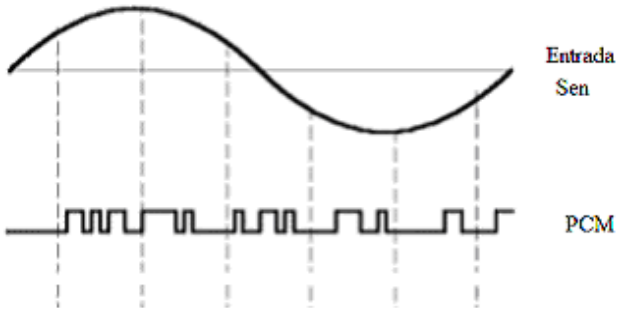


Figura 1.1: Modulación por Pulsos Codificados.

Cuantización

La cuantificación es el proceso que permite que una señal analógica sea representada digitalmente. En este proceso la amplitud de las muestras se cuantizan, dividiendo todo el rango en un conjunto finito de valores, y asignando el mismo valor de amplitud a todas las muestras que caen dentro de ese rango. Los cuantificadores uniformes, tienen los rangos de cuantización y los niveles distribuidos uniformemente. Son fáciles y menos costosos de implementar. Los cuantificadores logarítmicos, tienen los niveles de cuantización establecidos de forma logarítmica y se asigna una palabra código a cada región. Los cuantizadores logarítmicos más conocidos son la ley μ y ley A.

Cuantificación no uniforme - Ley μ

Esta característica de compresión está dada por:

$$y = C(x) = y_{max} \frac{\ln[1+\mu(|x|/x_{max})]}{\ln(1+\mu)} \operatorname{sgn}(x). \quad \text{.... (1.1)}$$

dada la función:

$$\log(1+z) \approx z \text{ si } z << 1,$$

en el caso de las señales pequeñas y grandes, respectivamente, tenemos la ecuación:

$$y = C(x) = \begin{cases} y_{max} \dfrac{\mu(|x|/x_{max})}{\ln\mu} & \text{si} \quad \mu\left(\dfrac{|x|}{x_{max}}\right) << 1 \\ y_{max} \dfrac{In[\mu(|x|/x_{max})]}{\ln\mu} & \text{si} \quad \mu\left(\dfrac{|x|}{x_{max}}\right) >> 1 \end{cases} \quad \text{.... (1.2)}$$

La función lineal de señales pequeñas $|x|/x_{max}$ y una función logarítmica de señales grande $\mu(|x|/x_{max})$ puede ser considerado el punto de ruptura entre ambas señales de operación con valor de 1.

El estándar de modulación de pulsos codificados (PCM) para un sistema de transmisión de voz recomienda μ=255 y R=8 [2].

La relación señal - ruido (SNR) de la ley μ puede estar representada por la siguiente ecuación:

$$\mathrm{SNR} = \frac{\int_{-x_{max}}^{x_{max}} x^2 p(x)dx}{(q^2/12)\int_{-x_{max}}^{x_{max}} (\ln \mu/y_{max})^2 x^2 p(x)\,dx} \qquad \text{.... (1.3)}$$

$$\mathrm{SNR} = \frac{1}{(q^2/12)(\ln \mu/y_{max})^2} = 3\frac{(2y_{max})^2}{q}.\left(\frac{1}{In\ \mu}\right)^2 = 3*2^{2R}.\left(\frac{1}{In\ \mu}\right)^2 \qquad \text{.... (1.4)}$$

Donde el factor $2y_{max}/q = 2^R$ representa el número de niveles de cuantificación y la ecuación anterior puede ser expresada en términos de dB obteniendo:

$$\mathrm{SNR}_{dB}^{\mu} = 6.02*R + 4.77 - 20\log_{10}(\ln(1+\mu)) \qquad \text{.... (1.5)}$$

dando un SNR de alrededor de 38dB considerando R=8 y μ = 255, del sistema estándar americano [2].

Cuantificación no uniforme - Ley A

La ley A fue normalizada por el CCITT, ahora conocido como ITU y que se utiliza en toda Europa:

$$y = C(x) = \begin{cases} y_{max}\dfrac{A(|x|/x_{max})}{1+\ln A}\,\mathrm{sgn}(x) & 0 < \dfrac{|x|}{x_{max}} < \dfrac{1}{A} \\ y_{max}\dfrac{1+\ln[A(|x|/x_{max})]}{1+\ln A}\,\mathrm{sgn}(x) & \dfrac{1}{A} < \dfrac{|x|}{x_{max}} < 1 \end{cases} \qquad \text{.... (1.6)}$$

Similar a las características de la ley μ, esta tiene una región lineal cerca del origen y una sección logarítmica que es el punto de ruptura $|x| = x_{max}/A$. Nótese, sin embargo, que en el caso de R=8 bits A<μ, de ahí la característica lineal-logarítmica de la ley A.

$2y_{max}/q = 2^R$ Representa el número de los niveles de cuantización,

$$\mathrm{SNR} = \frac{\int_{-x_{max}}^{x_{max}} x^2 p(x)dx}{(q^2/12)\int_{-x_{max}}^{x_{max}} ((1+\ln A)/y_{max})^2 x^2 p(x)\,dx} \qquad \text{.... (1.7)}$$

$$\mathrm{SNR} = \frac{1}{(q^2/12)((1+\ln A)/y_{max})^2} = 3\left(\frac{2y_{max}^{\ 2}}{q}\right)\left(\frac{1}{(1+\ln A)}\right)^2 = 3*2^{2R}.\left(\frac{1}{(1+In\ A)}\right)^2 \qquad \text{.... (1.8)}$$

Similar a la ley μ, el valor del SNR da alrededor de 38dB en el caso del estándar Europeo PCM para el sistema de transmisión de la voz con R=8 y A=87.56 [2].

$$\mathrm{SNR}_{dB}^{A} = 6.02*R + 4.77 - 20\log_{10}(1+\ln A) \qquad \text{.... (1.9)}$$

Simulación de los Cuantizadores

Para la cuantización uniforme los niveles de segmentación están distribuidos uniformemente, esto en ocasiones puede provocar que los errores de cuantización sean grandes y la señal representada digitalmente sea distinta a la original, mientras que para la cuantización logarítmica se pueden ajustar mejor la distribución de los niveles de cuantización, y por lo tanto, el error de cuantización será pequeño.

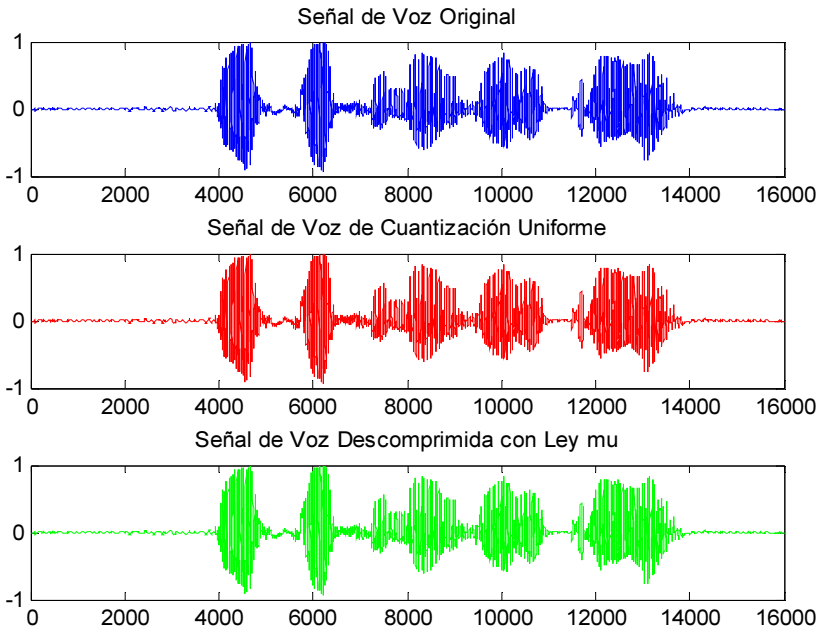


Figura 1.2: Una señal de voz cuantizada uniformemente y no uniformemente utilizando Matlab.

1.5 Evaluación de la Calidad

La calidad de la voz generada a partir de un codificador está en función de la tasa de bit, la complejidad, el retraso y el ancho de banda del mismo; factores que afectan la correcta percepción de la voz. Es importante darse cuenta de la fuerte interrelación que existe entre estos factores, siendo necesario, en muchas ocasiones, aceptar la degradación de uno o varios de ellos para conseguir la mejora de otro. Por ejemplo, un codificador con una tasa de bit baja suelen tener un mayor retraso que un codificador con una tasa de bit más alta. Además de ser de alta complejidad, suele tener peor calidad.

Tasa de bits

La tasa de bits, define el número de bits que se transmiten por unidad de tiempo a través de un canal, lo que implica que los codificadores de la voz al compartir el canal con otro tipo de información (datos), será necesario utilizar la menor tasa de bits posibles para no usar de forma excesiva el canal. Para aplicaciones que usan simultáneamente voz y datos se puede optar por usar un esquema de compresión de silencios como parte del estándar del código.

La compresión de silencios consiste en dos algoritmos: Por una lado un detector de actividad vocal (VAD; Voice Activity Detector), que determina si la señal de entrada es realmente voz o ruido de fondo. Si el detector determina que la señal es vocal, se codifica a una tasa de bits fija. Por el contrario, si determina que es ruido lo codifica a una tasa de bits baja. El segundo algoritmo es un generador de ruido (CNG; Comfort Noise Generation), que se usa en el receptor para reconstruir las principales características del ruido de fondo.

Latencia

El retraso es uno de los aspectos más importantes a la hora de implementar la voz, pues se busca minimizarlo. Dicho retraso es inherente a las redes de voz y es causado por un número de factores diferentes que intervienen en ellas. El retraso en un sistema de codificación de voz normalmente está formado por el retraso algorítmico, retraso por el procesamiento y el retraso por la comunicación.

Los codificadores de voz con una tasa de bits baja procesan las tramas una a una. Los parámetros de la señal son actualizados y transmitidos para cada trama. Además, para analizar la información correctamente, a veces es necesario analizarla más allá de los límites de la trama. Este proceso se califica como procesado hacia delante. Esto significa que antes de analizar la señal de voz, es necesario almacenar una serie de información. El retraso que se tiene como consecuencia de esto recibe el nombre de retraso algorítmico y este es inevitable para sistemas prácticos.

El procesamiento de la señal de voz genera un retraso debido a que el codificador requiere analizar la señal y el decodificador reconstruirla. Esto también depende de la velocidad del hardware.

La comunicación producida por la propia red está dada por la velocidad de transmisión de la misma, la congestión, y las demoras de los equipos de la red por donde se transmite la trama de información. Este retardo no afecta directamente la calidad de la voz, sino la calidad de la conversación. Hasta 100ms son generalmente tolerados, casi sin percepción de los interlocutores. Un efecto secundario, generado por las demoras elevadas, es el eco. El eco se debe a que parte de la energía de audio enviada es devuelta por el receptor.

La suma de estos tres retrasos se denomina retraso del sistema en un sentido. También se le conoce como latencia del códec. Valores máximos de hasta 400ms pueden ser admisibles sino hay ecos, aunque es preferible que este retraso esté por debajo de los 200ms. Si hay ecos, el máximo tolerable baja hasta los 25ms. De ahí el frecuente uso de canceladores de eco.

El jitter es una variante en las demoras, por lo que el receptor debe recibir los paquetes a intervalos constantes, para poder regenerar de forma adecuada la señal original. Estos receptores disponen de un buffer de entrada para recibir los paquetes a intervalos variables, y los entrega a intervalos constantes.

Ejemplo de algunos CÓDECS:

Algoritmo de muestreo/compresión	Demora típica
G.711 (64 kbps)	125 µs
G.728 (16 kbps)	2.5 ms
G.729 (8 kbps)	10 ms
G.723 (5.3 o 6.4 kbps)	30 ms

Tabla 1.4: El retardo de algunos algoritmos de compresión.

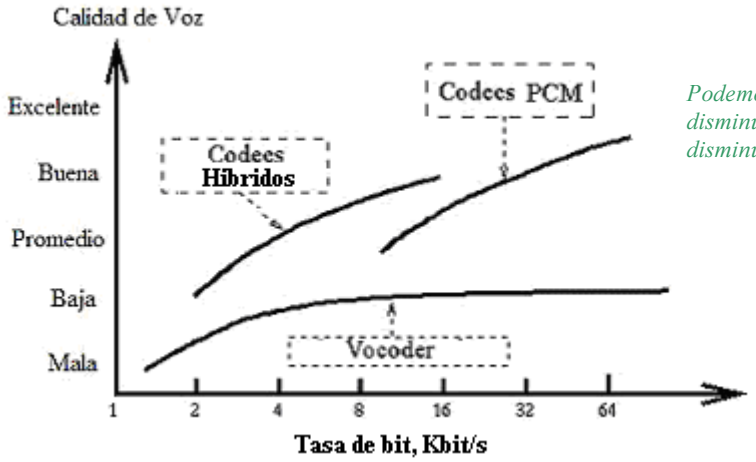
Ancho de Banda

El ancho de banda, es la cantidad de información que se transmite a través de una conexión en un tiempo determinado. El ancho de banda generalmente se determina en bits por segundo (bps). Para las redes inalámbricas, así como para los equipos y las tecnologías actuales el ancho de banda se ve limitado por el espectro de frecuencias, a pesar de que es pieza fundamental para el desempeño de la red.

1.6 Factores subjetivos y objetivos de la calidad

Los factores subjetivos de la calidad de voz, se basan en conocer directamente la opinión de los usuarios, pudiendo ser evaluada directamente (ACR–Absolute Category Rating) o en forma comparativa contra un audio de referencia (DCR–Degradation Category Rating). El MOS (Mean Opinion Store) es el promedio de los ACR medidos entre un gran número de usuarios. En las evaluaciones comparativas (DCR), el audio se califica también entre 1 y 5, siendo 5 cuando no hay diferencias apreciables entre el audio de referencia y el medido, y 1 cuando la degradación es muy molesta. El promedio de los valores DCR es conocido como DMOS (Degradation MOS).

La metodología de evaluación subjetiva más ampliamente usada es la del MOS, estandarizada en la recomendación ITU-T P.800[21]. Los factores objetivos, a su vez se subdividen en intrusivos (se inyecta una señal de voz conocida en el canal y se estudia su degradación a la salida) y no intrusivos (monitorean ciertos parámetros en un punto de la red y en base a estos se establece en tiempo real la calidad que percibiría un usuario).



Podemos apreciar cómo a medida que se disminuye la tasa de bit del codificador, disminuye la calidad de la señal de voz.

Figura 1.3: Relación entre la tasa de bit y la calidad de voz.

MOS Mean Opinion Store

El MOS consiste en una evaluación subjetiva de la calidad de la voz en un sistema [21]. Fue normalizado por la ITU y se le ha utilizado principalmente para medir la calidad en sistemas de comunicaciones celular. El proceso consiste en realizar una encuesta de opinión a un conjunto de usuarios de prueba, los cuales deben evaluar la calidad de la voz según la siguiente tabla:

MOS	Calidad	Clasificación de la voz
5	Excelente	Transparente
4	Buena	Red digital mejorada
3	Aceptable	Comunicaciones
2	Mediocre	Voz artificial
1	Mala	Muy Molesta

Tabla 1.5: MOS Mean Opinion Store.

SNR (Signal to Noise Ratio)

La relación señal-ruido (SNR) calcula la energía en la señal original con relación a la energía del ruido, donde el ruido es el error entre la señal original y la señal sintetizada. Las medidas de la calidad de la voz nos permiten cuantificar la distorsión de los sistemas de comunicaciones entre la entrada y salida en el tiempo o en el dominio de la frecuencia.

La relación SNR se puede definir como:

$$SNR = \frac{\sigma_{in}^2}{\sigma_e^2} = \frac{\sum_n s_{in}^2(n)}{\sum_n [s_{out}(n) - s_{in}(n)]^2}, \quad \text{.... (1.10)}$$

donde, Sin(n) y Sout(n) son la secuencia de entrada y salida de muestras de la voz, mientras que σ_{in}^2 y σ_e^2 son la varianza de la entrada de voz y la señal error de esta, respectivamente.

El valor de la relación señal a ruido está dominado por los segmentos de mayor energía de la voz. Por lo tanto, la fidelidad de la reconstrucción de la voz es una prioridad mayor que la de bajo consumo de energía de los sonidos sordos, cuando se calcula la media aritmética de la SNR, lo que puede ser expresada en dB como:

SNR^{dB} = 10 log10 SNR. .... *(1.11)*

La definición de los llamados segmentos SNR (SEGSNR):

$$SEG-SNR^{dB} = \frac{1}{M}\sum_{m=1}^{M} 10\log_{10}\frac{\sum_{n=1}^{N} s_{in}^2(n)}{\sum_{n=1}^{N}[s_{out}(n)-s_{in}(n)]^2}, \quad \text{.... (1.12)}$$

donde N es el número de muestras de la voz dentro de un segmento de 15 a 25 ms por lo general, es decir, 120 a 200 muestras a una velocidad de muestreo de 8 kHz, mientras que M es el número de segmentos entre 15-25 ms, sobre el cual se evalúa $SEGSNR^{dB}$.

BER (Bit Error Rate)

La Tasa de Error de Bit (BER) se define como la velocidad a la que se producen errores en un sistema de transmisión y es usualmente asociada con la calidad de señal. Es decir, este parámetro proporciona una indicación excelente del rendimiento de un enlace el cual está asociado con el número de bits o bloques incorrectamente recibidos, con respecto al total de bits o bloques enviados durante un intervalo específico de tiempo.

BER = número de bits erróneos / número total de bits enviados

Si el medio entre el transmisor y el receptor es buena, y la señal a ruido es alto, entonces la tasa de error será muy pequeña posiblemente insignificante y no teniendo ningún efecto notable sobre el sistema global. Sin embargo, si la señal a ruido es baja, entonces hay probabilidad de que la tasa de error de bits deba ser considerada.

Cabe señalar que cada tipo diferente de modulación tiene su propio valor de la función de error. Esto es porque cada tipo de modulación funciona de manera diferente en la presencia de ruido. En particular, los esquemas de modulación de orden superior (por ejemplo, 64 QAM, etc.) que son capaces de llevar a mayores velocidades de datos no son tan robustos en presencia de ruido. Formatos de modulación de orden inferior (por ejemplo, BPSK, QPSK, etc.) ofrecen velocidades más bajas, pero son más robustos.

FER (Frame Error Rate)

La Tasa de Error de la Trama (FER) es la probabilidad de que un bloque decodificado contenga errores, en cuyo caso es inutilizable para el decodificador fuente y por lo tanto se elimina.

Cada trama contiene bits CRC (Código de Redundancia Cíclica), que es un código de detección de errores usado para detectar cambios accidentales en los datos y la acción de corrección puede tomarse en contra de los datos presuntamente corrompidos.

Capítulo 2

Bases de la Codificación

Con la digitalización de la señal de voz, se comienza con el desarrollo de algoritmos de compresión más sofisticados para lograr reducir el número de bits necesarios para la transmisión. Esto da pie al surgimiento de distintas técnicas de digitalización de la voz. La codificación más sencilla es PCM que representa adecuadamente todo el rango de la señal de voz, pero que necesita un elevado número de bits.

Una posibilidad para reducir el número de bits necesarios en PCM es reducir el rango de la señal que se va a codificar, es cuando se hace uso de los predictores de codificación, es decir, no se codifica directamente la señal de voz, sino que se codifican diferencias entre dos puntos consecutivos de la señal. La señal de voz es una señal muy autocorrelacionada, de forma que al trabajar con las diferencias entre dos puntos consecutivos el rango de la señal se puede reducir.

Dentro del segundo capítulo se presentaran diversas técnicas de codificación ideadas principalmente durante la década de 1980. Siendo hoy en día las bases de la codificación para las comunicaciones móviles.

2.1 Predicción Lineal

La predicción lineal (LP) se aplica para obtener un modelo para la producción de señales de voz. El problema fundamental de la predicción lineal de la voz es la determinación del conjunto de coeficientes para cada trama que haga mínimo el error cuadrático medio de la predicción e(n). Una vez determinados los coeficientes para esa trama, si se aplica la excitación adecuada al sistema determinado por la función de transferencia se obtendrá a la salida una secuencia que reproducirá la trama en cuestión no exactamente en el dominio del tiempo, pero si en sus propiedades espectrales.

Con la codificación de la trama y el cálculo de los parámetros del modelo de la trama nos permitirán transmitir menos bits en relación a la trama original. El decodificador de la trama recibe estos parámetros e implementa un modelo de producción de voz para esa trama.

2.2 Diseño del predictor

Debido a la redundancia inherente en la voz, una muestra actual se puede predecir como una combinación lineal de *p* muestras anteriores de la voz.

$$\tilde{s}(n) = \sum_{k=1}^{P} a_k s(n-k), \qquad \text{.... (2.1)}$$

donde *p* es el orden de predicción, a_k representa los coeficientes del filtro de predicción lineal y $\tilde{s}(n)$ la predicción de las muestras de voz. La predicción del error *e(n)* es representada por:

$$e(n) = s(n) - \tilde{s}(n) = s(n) - \sum_{k=1}^{P} a_k s(n-k), \qquad \text{.... (2.2)}$$

$$e(n) = \sum_{k=0}^{P} a_k s(n-k),$$

donde $a_0 = 1$, al tomar la transformada z de la ecuación anterior, llegamos a:

$$E(z) = S(z) \cdot A(z), \qquad \text{.... (2.3)}$$

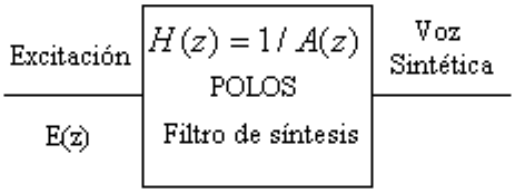


Figura 2.1: Modelo de la generación de voz.

observe que:

$$A(z) = 1 - \sum_{k=1}^{P} a_k z^{-k} = \sum_{k=0}^{P} a_k z^{-k}, a_0 = 1, \qquad \text{.... (2.4)}$$

puede ser expresado como:

$$A(z) = 1 - a_1 \cdot z^{-1} - a_2 \cdot z^{-2} - ... - a_P \cdot z^{-p} = (z - z_i)...(z - z_P), \qquad \text{.... (2.5)}$$

Lo que explícitamente se muestra es que este polinomio tiene sólo ceros, pero no polos.

Expresando la señal de voz $S(z)$ en términos de $E(z)$ y $A(z)$:

$$S(z) = \frac{E(z)}{A(z)} = E(z) \cdot H(z), \qquad \text{.... (2.6)}$$

La generación de la voz sintetizada utilizando la ecuación *(2.6),* también puede ser interpretada como la excitación del filtro de síntesis. Si el predictor elimina la redundancia de la señal de voz, reduciendo al mínimo la predicción residual, *e(n)* se convierte en impredecible, es decir, pseudo-aleatorio. Debido a la relación de $A(z) = H^{-1}(z)$, el filtro $_{A(z)}$ se refiere a menudo como el filtro inverso LPC.

El valor esperado (E) del error cuadrático medio de la predicción de la ecuación *(2.2)* se puede escribir como:

$$E\left[e^2(n)\right] = E\left[\left[s(n) - \sum_{k=1}^{P} a_k s(n-k)\right]^2\right] \qquad \text{.... (2.7)}$$

Con el fin de mejorar los coeficientes LPC se calcula la derivada parcial de la ecuación *(2.7)* con respecto a todos los coeficientes LPC, por lo tanto se establece que $\partial E / \partial a_i = 0$ para i = 1,. . . , *p*, se produce un conjunto de
p ecuaciones para la incógnita *p* para los coeficientes LPC a_i como:

$$\frac{\partial E\left[e^2(n)\right]}{\partial a_i} = -2 \cdot E\left\{\left[s(n) - \sum_{k=1}^{P} a_k s(n-k)\right] s(n-i)\right\} = 0, \quad \text{.... (2.8)}$$

dando,

$$E\left\{s(n)s(n-i)\right\} = E\left\{\sum_{k=1}^{P} a_k s(n-k)s(n-i)\right\} \qquad \text{.... (2.9)}$$

Al intercambiar el orden de la suma y el cálculo del valor esperado (E) en el lado derecho de la ecuación *(2.9)* llegamos a:

$$E\left\{s(n)s(n-i)\right\} = \sum_{k=1}^{P} a_k E\left\{s(n-k)s(n-i)\right\}, i = 1,..., p \qquad \text{.... (2.10)}$$

Observe de la ecuación anterior que

$$C(i,k) = E\left\{s(n-i)s(n-k)\right\}, \qquad \text{.... (2.11)}$$

La representación de los coeficientes de covarianza de la señal de entrada, nos permiten rescribir la ecuación *(2.10)* en una forma más concisa de la siguiente manera:

$$\sum_{k=1}^{P} a_k C(i,k) = C(i,0), i = 1,..., p. \qquad \text{.... (2.12)}$$

Calculo del coeficiente de la covarianza

En los códecs de baja complejidad o si la señal de entrada posee propiedades estadísticas estacionarias, es decir, que las estadísticas de la señal son invariantes en el tiempo, los coeficientes de covarianza se puede determinar mediante una secuencia de formación suficientemente larga. Entonces el conjunto de ecuaciones *p* de la ecuación *(2.12)* se puede resolver, por ejemplo, por eliminación Gauss-Jordan, o de manera más eficiente por el algoritmo iterativo de Levinson-Durbin. Esta técnica, que se refiere a menudo como predicción con adaptación hacia delante (forward-adaptive), implica que los coeficientes cuantificados en el codificador también deben ser transmitidos.

Otra alternativa es el uso de la predicción con adaptación hacia atrás (backward-adaptive), donde los coeficientes LPC no se transmiten al decodificador, sino que se recuperan de los segmentos anteriores de la señal decodificada. Una vez más, con el fin de garantizar el funcionamiento de los decodificadores, el codificador también utiliza los segmentos anteriores de la señal decodificada para determinar los coeficientes LPC.

Para tener una predicción eficiente el retraso asociado con la predicción con adaptación hacia atrás (backward-adaptive) debe ser lo más bajo posible, mientras que la calidad de la señal decodificada tiene que

ser lo más alta posible. Por lo tanto, esta técnica no se utiliza en aplicaciones de baja tasa de bits, donde la demora por lo general es alta y presenta alta distorsión de codificación reduciendo la eficacia del predictor.

Calculo del coeficiente del predictor

Una variedad de técnicas han sido propuestas para limitar el rango del cálculo de la covarianza, de los cuales los más utilizados son el método de autocorrelación y el método de covarianza.

Se destaca brevemente el método de autocorrelación, donde el término de error de predicción de la ecuación *(2.7)* es ahora limitado en el intervalo finito de $0 \le n \le L_a - 1$, en lugar de - ∞ <n <∞. Por lo tanto los coeficientes de la covarianza C (i, k) son calculados a partir del valor esperado:

$$C(i,k) = \sum_{n=0}^{L_a+p-1} s(n-i)s(n-k), i = 1,..., p, k = 0,..., p. \qquad \text{.... (2.13)}$$

Al establecer m = n - i, la ecuación *(2.13)* se puede expresar como:

$$C(i,k) = \sum_{m=0}^{L_a-1-(i-k)} s(m)s(m+i-k), \qquad \text{.... (2.14)}$$

Por lo tanto, $C(i,k)$ es la autocorrelación de corto plazo de la señal de entrada $s(m)$ evaluada en un desplazamiento de (i,k), esto es:

$$C(i,k) = R(i-k), \qquad \text{.... (2.15)}$$

$$R(j) = \sum_{n=0}^{L_a-1-j} s(n)s(n+j) = \sum_{n=j}^{L_a-1} s(n)s(n-j), \qquad \text{.... (2.16)}$$

Donde R(j) representa los coeficientes de autocorrelación de la voz. Por lo tanto el conjunto de ecuaciones *p* de la ecuación *(2.12)* puede ser reformulada como:

$$\sum_{k=1}^{P} a_k R(|i-k|) = R(i), i = 1,..., p. \qquad \text{.... (2.17)}$$

La ecuación *(2.17)* puede ser re-escrita de forma matricial como:

$$\begin{pmatrix} R(0) & R(1) & R(2) & \cdots & R(p-1) \\ R(1) & R(0) & R(1) & \cdots & R(p-2) \\ R(2) & R(1) & R(0) & \cdots & R(p-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R(p-1) & R(p-2) & R(p-3) & \cdots & R(0) \end{pmatrix} \cdot \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_p \end{pmatrix} = \begin{pmatrix} R(1) \\ R(2) \\ R(3) \\ \vdots \\ R(p) \end{pmatrix} \qquad \text{.... (2.18)}$$

La matriz de autocorrelación p×p es una matriz simétrica de Toeplitz, en donde los elementos a cada una de las diagonales son idénticos. La solución para resolver la ecuación *(2.18)* es un método recursivo conocido como algoritmo de Levinson-Durbin.

Algoritmo de Levinson-Durbin

El método de algoritmo de Levinson-Durbin, consiste en proceder recursivamente, empezando con un predictor de orden 1 y después incrementando el orden, usando las soluciones de orden menor para obtener la solución al siguiente orden superior, es decir, se comienza desde un orden menor hasta el orden deseado.

Esta es la matriz de autocorrelación de la señal de entrada no normalizada, usada por el algoritmo,

$$R = \begin{pmatrix} R(0) & R(1) & R(2) & \cdots & R(p-1) \\ R(1) & R(0) & R(1) & \cdots & R(p-2) \\ R(2) & R(1) & R(0) & \cdots & R(p-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R(p-1) & R(p-2) & R(p-3) & \cdots & R(0) \end{pmatrix} \qquad \text{.... (2.19)}$$

en donde,

$$R(0) = \sum_{i=0}^{n=1} x(i)x(i),$$

$$R(1) = \sum_{i=0}^{n=1} x(i)x(i+1), \qquad \text{.... (2.20)}$$

$$R(k) = \sum_{i=0}^{n=1} x(i)x(i+k), \qquad donde \qquad k = 0,1,...,n-1$$

Ya conociendo la matriz de autocorrelación, los coeficientes a_{nk} se calculan,

$$a_k(n) = \frac{R(n) + \sum_{i=0}^{n-1} R(n-k)a_{n-1}(k)}{R(0) + \prod_{i=1}^{n=1}(1 - a_i(k))^2} \qquad \text{.... (2.21)}$$

$$a_{n-1}(n-k) = a_{n-1}(n-k) + a_n(n)a_{n-1}(k) \qquad donde \qquad k = 1,2,...,n-1 \qquad \text{.... (2.22)}$$

Los coeficientes $a_n(k)$ describen el óptimo predictor lineal de orden k-ésimo.

2.3 Pitch

El pitch es una propiedad de la voz, que nos permite determinar la calidad percibida del sonido conocido como la "frecuencia fundamental". Nuestro sistema auditivo es capaz de percibir el "pitch" basado en la separación relativa de los armónicos. Un armónico es un múltiplo entero de la frecuencia fundamental cuando asumimos un escalado lineal en frecuencia.

Autocorrelación Estimación del Pitch

La función de autocorrelación se utiliza con frecuencia para la extracción del pitch. Una función de correlación es una medida del grado de similitud entre dos señales, es decir, mide lo bien qué la señal de entrada coincide con una versión desplazada en tiempo de sí misma. Los máximos de la función de autocorrelación se producen a intervalos del período del Pitch de la señal original.

Transformadas

La transformada Z en la codificación de la voz nos permite analizar y comprender la señal en el dominio de la frecuencia. La transformada Z en señales discretas tiene su equivalente en la transformada de Laplace para señales continuas y cada una de ellas mantiene su relación correspondiente con la transformada de Fourier. La transformada Z proporciona una representación útil para analizar las cualidades del espectro formado por los sistemas de un polo y / o ceros, y como la expresión más general de la transformada de Fourier.

La transformada de Fourier (FT) representa una señal en términos de exponenciales complejas. Como tal, la representación de la señal a través de la transformada de Fourier facilita algunos procesamientos.

La transformada Discreta de Fourier (DFT) es una transformación en frecuencia más utilizada en la forma de onda de la voz. La transformada DFT es una representación en Fourier de una secuencia de muestras que están igualmente espaciadas a lo largo del eje de la frecuencia de la FT, en lugar de ser una función continua.

Además del problema de no conocer la señal de voz en todos los tiempos, la voz es altamente no estacionaria, las estadísticas de la señal cambian con el tiempo. Aunque la señal de voz no es estacionaria, está es cuasi-estacionario en un pequeño segmento de voz (20 ms o menos) "casi" tiene las propiedades de una señal estacionaria. Por esta razón, la transformada de Fourier es utilizada para el procesamiento de las señas de la voz en pequeños segmentos. Un intervalo de tiempo específico (por ejemplo, 20 ms) se utiliza para el segmento, y el segmento es ponderado en consecuencia. Los parámetros de frecuencia resultantes se asocian con el segmento de tiempo de la voz que corresponde al centro del intervalo de análisis.

La transformada rápida de Fourier (FFT) es un conjunto de métodos que reorganiza los cálculos de la transformada DFT para reducirlos. El cálculo directo de la DFT requiere un número de multiplicaciones sobre el orden de N^2, mientras que la FFT reduce el número a la orden de N log N. El proceso utiliza la simetría y la periodicidad del factor exponencial para reducir los cálculos.

2.3.1 Simulación del Pitch en una señal de voz

Con una señal de voz de 2 [s] muestreada a 8 [kHz] en un intervalo de [1500,1739], el cual corresponde a una porción de sonido sonoro de la señal muestreada. Sobre una ventana de 20 milisegundos (240 muestras) en el dominio de la frecuencia, el tamaño de la ventana corresponde a 480 muestras por el teorema de Nyquist, con una resolución de 16.66 [Hz] por muestra.

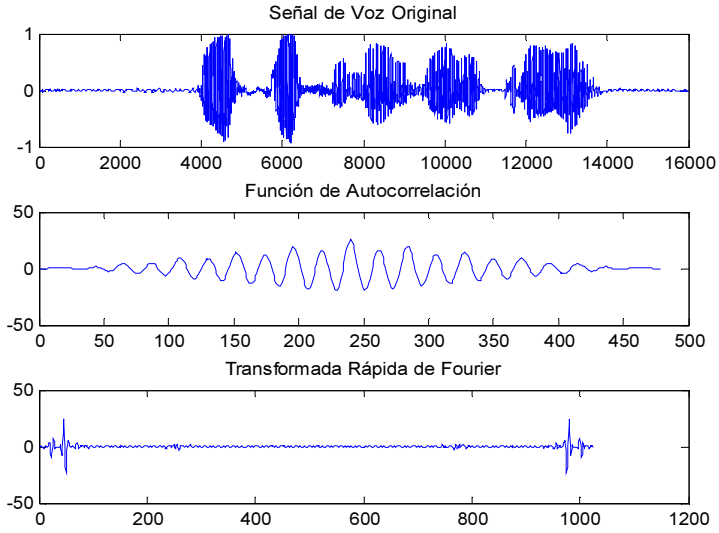


Figura 2.2: Análisis del Pitch de una señal de voz con Matlab.

>>Para el cálculo del pitch en el dominio del tiempo:

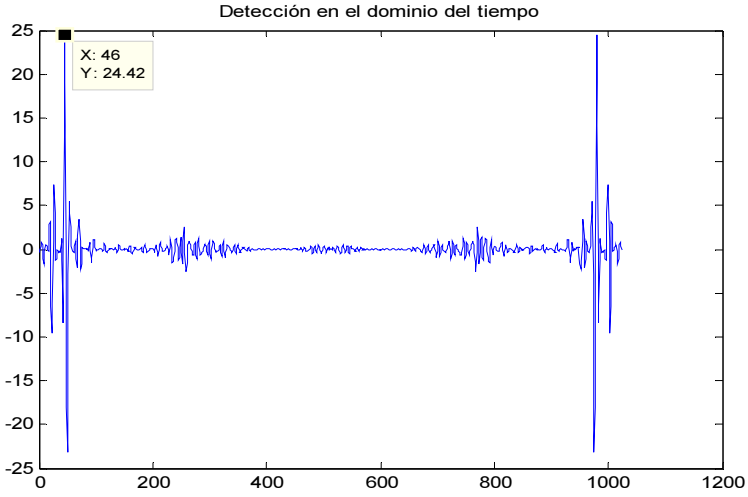


Figura 2.4: Detección en el dominio del tiempo.

Identificar el máximo local en el dominio de la frecuencia, se calcular el valor en Hz correspondiente a una muestra. En el dominio del tiempo, aplicando una FFT normal tendríamos 256 desplazamientos equivalentes a 8[kHz], por lo que para 1024 desplazamientos logramos una mayor resolución, es decir, ahora 1024 desplazamientos corresponden a 8 [kHz]. Entonces en una muestra tenemos:

$$Una\ muestra\ en\ Hz = \frac{(1[muestra]) * (8000[Hz])}{1024[muestras]} = 7.8125[Hz]$$

Finalmente el Pitch se obtiene calculando el total de Hz en la muestra 46, es decir en el máximo local:

$$Pitch = (46) * (7.8125[Hz]) \Rightarrow Pitch = 359.37[Hz]$$

>>Para el cálculo del Pitch en el dominio de la frecuencia:

Se identifica la distancia entre los máximos locales (los máximos locales se encuentran en la muestra 240 y la muestra 218).

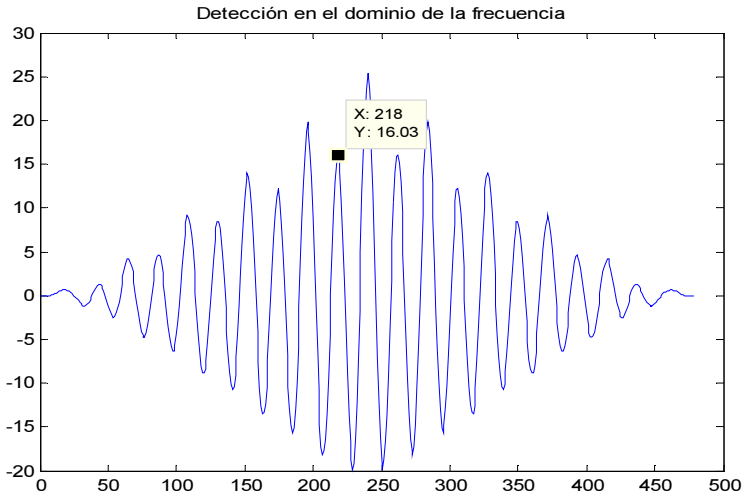


Figura 2.3: Detección en el domino de la frecuencia.

$$Num.muestras = 240 - 218 = 22[muestras]$$

Se obtiene una muestra en tiempo:

$$Una\ muestra\ en\ tiempo = \frac{1}{8000[Hz]} = 1.25x10^{-4}[s]$$

Calculamos el tiempo total de las muestras:

$$Muestra\ en\ tiempo = (Num.muestras) * (Una\ muestra\ en\ tiempo)$$
$$Muestra\ en\ tiempo = (22[muestras]) * (1.25x10^{-4}[s]) = 2.75x10^{-3}[s]$$

Sacando el inverso del tiempo total de muestras, conocemos el valor del Pitch:

$$Pitch = \frac{1}{Muestra\ \ en\ tiempo} = \frac{1}{2.75\ x10^{-3}[s]} \Rightarrow Pitch = 363.636\ [Hz]$$

Como se puede apreciar, ambos procesos usados en el dominio de la frecuencia o en el dominio del tiempo nos arrojan resultados muy similares para el cálculo del Pitch. Por lo tanto, ambos procesos permiten el cálculo del Pitch.

2.4 DM (Modulación Delta - Delta Modulation)

La Modulación Delta utiliza la codificación PCM, siendo la más sencilla modulación, consiste en codificar digitalmente la información mediante un sistema que codifica en forma diferencial cada muestra a 1 bit. En la Figura 2.5, se muestra un diagrama de bloques del modulador y demodulador delta. Donde podemos apreciar que el modulador delta se basa en la cuantización, es decir, en el cambio de la señal de muestra a otra en el valor absoluto de la señal en cada muestra. Para este modulador el integrador funciona como un predictor al intentar predecir la entrada x(t).

El término de error de predicción $x(t) - \bar{x}(t)$ en la predicción actual se cuantifica y se utiliza para hacer la predicción siguiente. Este error de predicción cuantificado (salida del modulador) es integrado en el receptor tal como es en el circuito de realimentación del mismo modulador. El receptor predice las señales de entrada y el filtro paso-bajas provee por sí mismo una medida aproximada de integración a su vez que suaviza y filtra la señal.

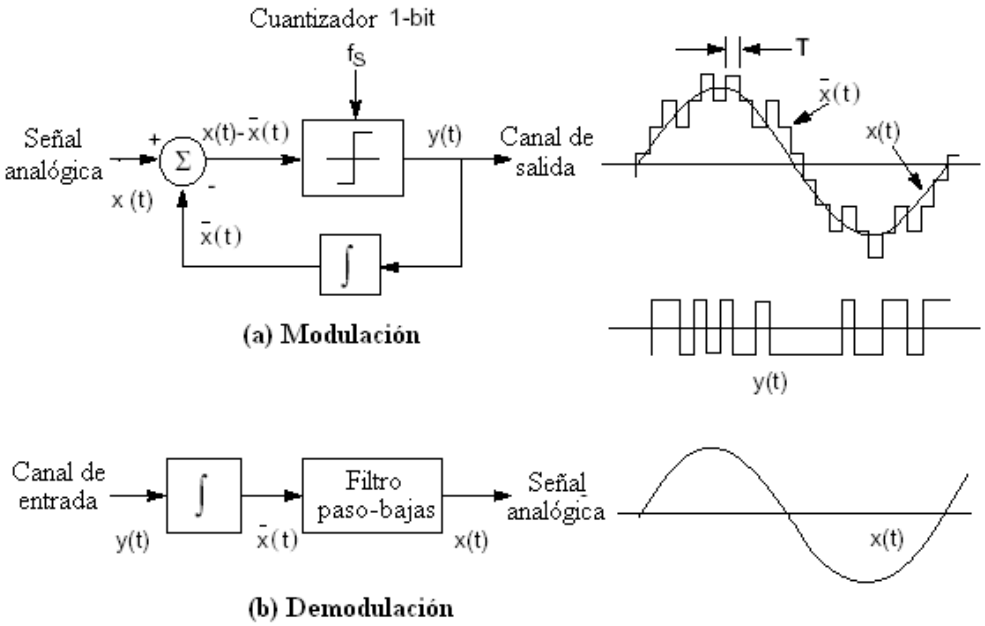


Figura 2.5: Modulación y Demodulación Delta.

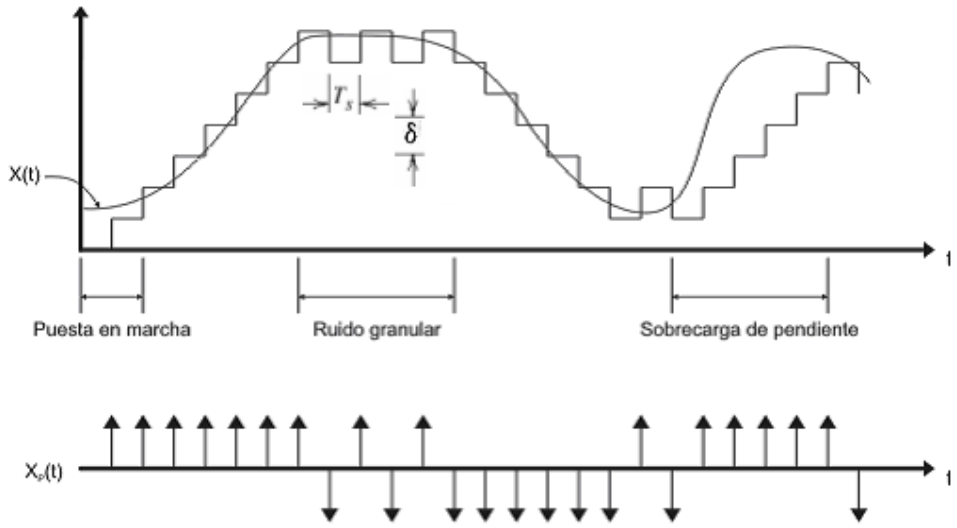


Figura 2.6: Puesta en marcha, ruido granular y sobrecarga de pendiente.

Sin embargo, esta modulación presenta algunos inconvenientes Figura 2.6 como es la puesta en marcha, el ruido granular, sobrecarga de pendiente y su rendimiento, es por lo tanto, dependiente de la frecuencia de la señal de entrada. En estos sistemas, los intervalos de ruido granular ocurren cuando la señal cambia muy lentamente, y la sobrecarga de pendiente, por el contrario, ocurre cuando la pendiente de la señal es muy elevada. Ambos problemas pueden ser contrarrestados ajustando el tamaño del escalón en forma adaptativa, en concordancia con la señal que ingresa al sistema. Además la tasa de muestreo es varias veces superior que la tasa de Nyquist para poder tener una predicción del valor anterior apropiada.

Simulación de una señal modulada en DM

Podemos apreciar que la aproximación de la señal de entrada (función Sen) es a través de una señal aproximada en escalera, donde solamente se utilizan dos niveles 1 o 0, donde en algunos casos la aproximación cae por debajo o por encima de la señal original, también se tiene que para esta señal las muestras no varían rápidamente de una a otra.

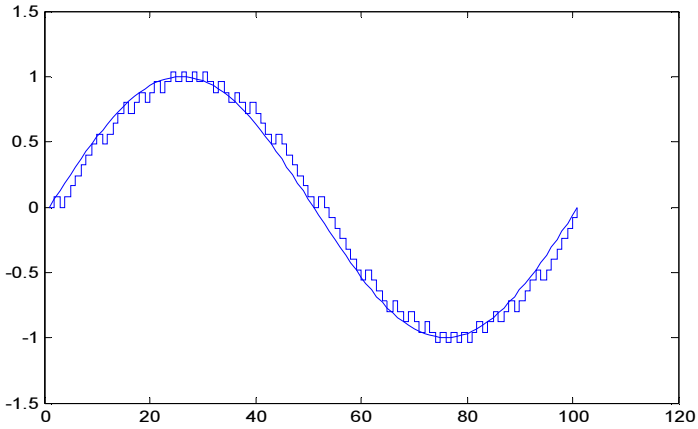


Figura 2.7: Se genera una señal que se modula en DM utilizando Matlab.

2.5 ADM(Modulación Delta Adaptativo - Adaptative Delta Modulation)

Los sistemas ADM se conocen también como "Sistemas de Modulación Delta de Variación Continua de Pendiente", CVSDM (Continuous Variable Slope Delta Modulation). La modulación ADM a diferencia de la modulación DM, esta modula el tamaño del escalón acorde con las características de la señal de entrada, tratando de evitar el ruido de sobrecarga de pendiente. En esta forma de modulación el escalón se ajusta en forma continua en vez de hacerlo en pasos discretos, utilizando los 4 últimos bits en la ganancia del integrador.

El tamaño del escalón es controlado por la ganancia de un circuito integrador y un dispositivo de ley cuadrática. Cuando la señal de entrada sea constante o varíe lentamente, el modulador estará evitando el "ruido granular" y la salida de pulsos de polaridad alternada. Debido al circuito de ley cuadrática, la ganancia del amplificador será incrementada sin importar la polaridad de los pulsos. El resultado neto es un incremento en el tamaño del escalón y una reducción en la sobrecarga de la pendiente. El demodulador de un sistema delta adaptativo, deberá tener un circuito de control adaptativo de ganancia similar al utilizado en el modulador.

La salida de un dispositivo de ley cuadrática con una sola frecuencia de entrada es cd y la segunda armónica. No se generan más armónicas que la segunda. Por consiguiente, se produce menos distorsión armónica.

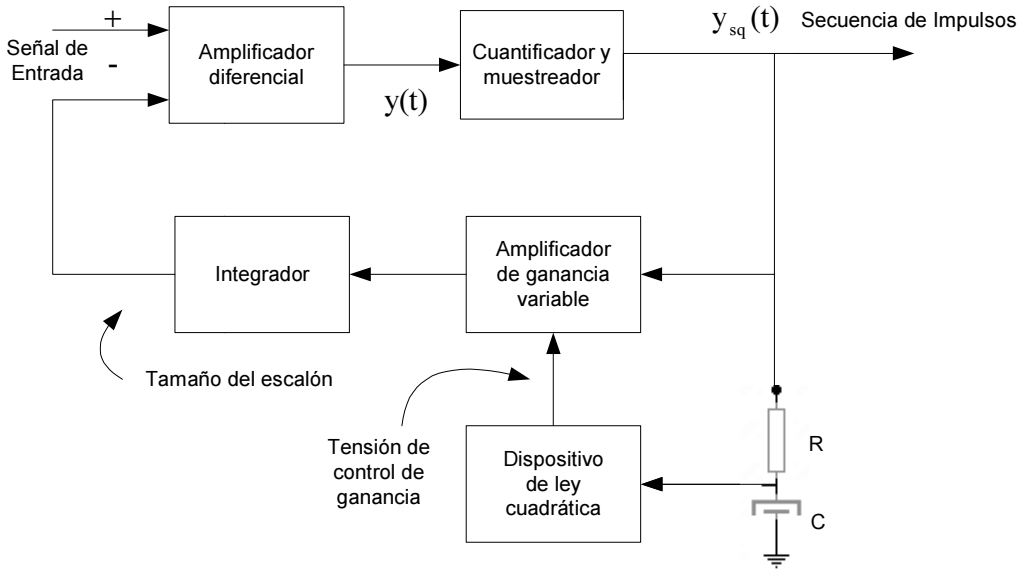


Figura 2.8: Modulador Delta Adaptativo.

2.6 DPCM (Modulación por Pulsos Codificados Diferenciales - Differential Pulse Code Modulation)

La modulación DPCM es un codificador de forma de onda que parte de la base del PCM, pero anexa funcionalidades basadas en la predicción de las muestras de la señal. Puesto que PCM no tiene en cuenta la forma de la onda de la señal a codificar, funciona muy bien con señales que no sean las de voz, cuando se tratan de señales de voz existe una gran correlación entre las muestras adyacentes.

La modulación de pulsos codificados diferenciales (DPCM) está diseñada para aprovechar la redundancia, de muestra a muestra, en las formas de onda de la voz. Con esta modulación la diferencia en amplitud de dos muestras sucesivas es la que se transmite en vez de la muestra original, de este modo se requieren menos bits para ser transmitidos.

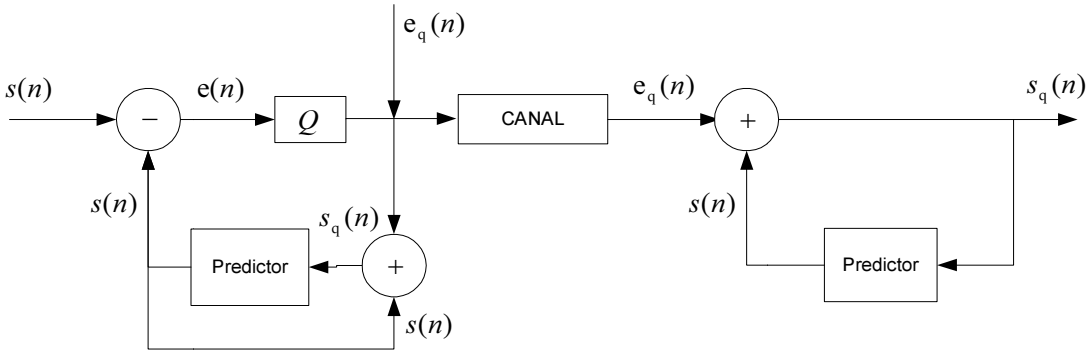


Figura 2.9: Modulación por Pulsos Codificados Diferenciales (DPCM).

En el diagrama anterior, se muestra que el decodificador cuenta con un bloque predictor, el cual ayuda a mitigar los efectos de errores de transmisión.

La varianza del error de predicción e(n) es normalmente inferior a la de la señal s(n), es decir, $\sigma e < \sigma s$, la tasa de bits necesaria para la cuantificación de e(n) puede ser reducida.

$$e(n) = s(n) + s(n).$$
$$e_q(n) = Q[e(n)]$$
$$s_q(n) = s(n) + e_q(n) \quad \text{.... (2.23)}$$

Si en un sistema DPCM aumentamos la frecuencia de muestreo, llega un momento en que dos muestras consecutivas tienen una amplitud tan próxima, que no se necesita más que un solo intervalo de cuantificación para cuantificar la diferencia.

Simulación de la modulación DPCM

El desempeño de este tipo de modulación DPCM es bastante bueno, puesto que transmite únicamente las diferencias entre las muestras. Esta señal en diferencia al tener un rango dinámico menor que la señal original, nos permite cuantificar a través de un número menor de niveles. En la figura siguiente podemos apreciar que la señal reconstruida por DPCM es casi idéntica a la señal original.

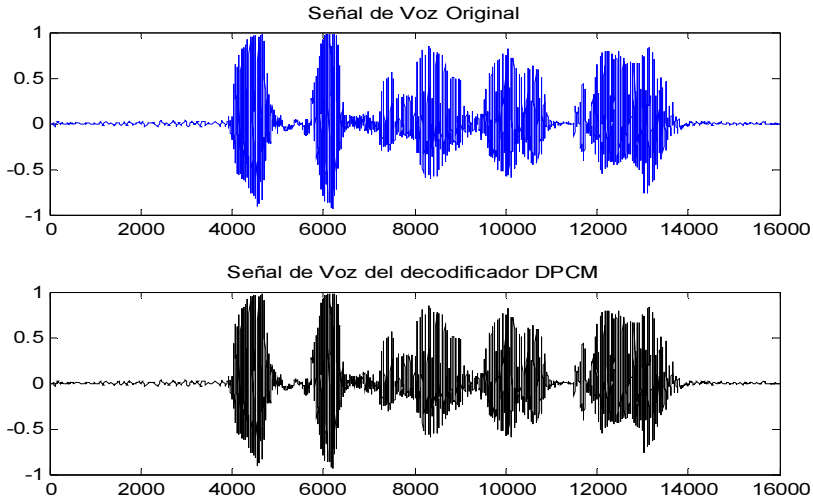


Figura 2.10: Una señal de voz se modulada en DPCM utilizando Matlab.

2.7 ADPCM (Modulación por Pulsos Codificados Diferenciales Adaptativos - Adaptative Differential Pulse Code Modulation)

Los ADPCM son codificadores de forma de onda que en vez de cuantificar la señal directamente como los codificadores PCM, cuantifican la diferencia entre la señal y una predicción hecha a partir de la señal, por lo que se trata de una codificación diferencial, porque se va incrementando o decrementando en función de la magnitud de las diferencias previamente codificadas. La DPCM usa escalones fijos, mientras que la ADPCM usa escalones variables para codificar la diferencia, es decir, se basa en ajustar la escala de cuantificación de forma dinámica para adaptarse mejor a las diferencias.

Se utilizan dos métodos para adaptar los cuantificadores y los predictores. El primero método es una adaptación hacia adelante (forward-adaptive) en donde los niveles de reconstrucción y los coeficientes de predicción se calculan en el emisor, usando un bloque de voz. Después son cuantificados y transmitidos al receptor como información. Tanto el emisor como el receptor usan estos valores cuantificados para hacer las predicciones y cuantificar el residuo. El segundo método es una adaptación hacia atrás (backward-adaptive) puede dar menores tasas de bits, pero son más sensibles a los errores de transmisión que la adaptación hacia adelante.

ADPCM es muy útil para codificar voz. La ITU propuso un estándar de codificación de voz telefónica a una velocidad de 32 kbps conocido como el estándar G726 [14].

Este estándar utiliza un esquema de adaptación hacia atrás (backward-adaptive) tanto para el cuantificador como para el predictor. El predictor tiene dos polos y seis ceros, por lo que produce una calidad de salida aceptable para señales que no son de voz. La ventaja del uso de codificadores ADPCM en aplicaciones es que nos permite reducir a 32kbps la tasa de transmisión, así como aliviar la congestión sin degradar la calidad de la señal.

2.7.1 Códec G726

El códec G.726 [14] (ADPCM) es una forma de modulación por pulsos codificados (PCM), con la finalidad de producir una señal digital con una tasa de bits inferior al PCM estándar.

Descripción funcional del códec G726

La tasa de transmisión del códec ADPCM es de 32 kbps, se ha especificado en la Recomendación de la ITU. La tasa de transmisión está determinada por una frecuencia de muestreo de 8 kHz, y cada muestra está representada por 4 bits. El codificador/decodificador se muestra en la Figura 2.11 y 2.12 respectivamente, y debido a que es esencialmente un códec de forma de onda, aparte de voz, también es capaz de transmitir señales de datos.

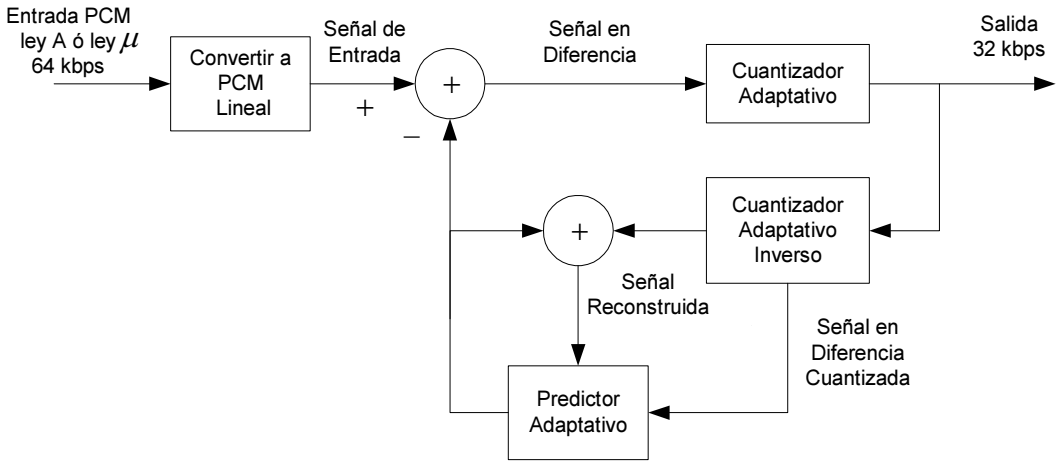


Figura 2.11: G.726 Codificador ADPCM.

En el diagrama de bloques Figura 2.11, se muestra como la señal PCM con compresión ley-A o ley-μ se convierte primero en el formato PCM lineal, la señal producida a la entrada del comparador se verá afectada por la resta de un señal producida por el predictor adaptativo con el fin de producir una señal en diferencia que tenga una menor varianza. Esta señal en diferencia pasará a través de un cuantizador de 4 bits de adaptación para luego ser transmitido. Por otro lado, la está señal retroalimentara el codificador para obtener una nueva señal producida por el predictor adaptativo.

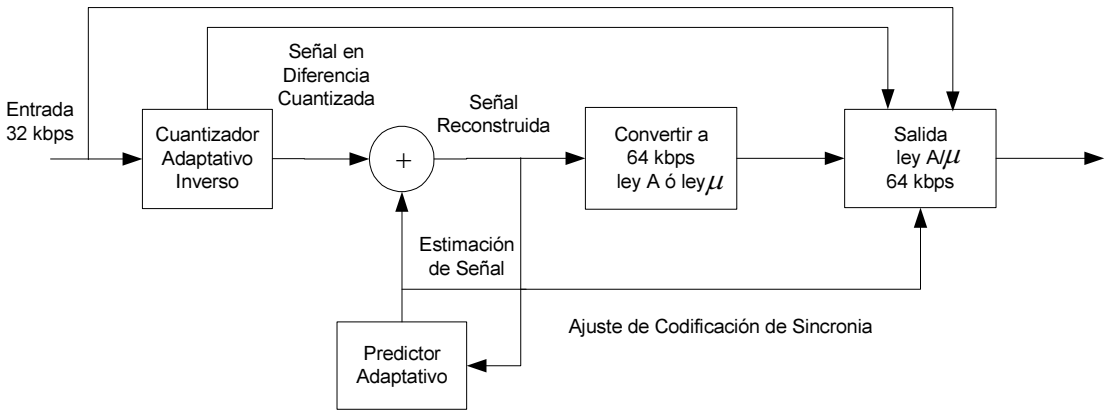


Figura 2.12: G.726 Decodificador ADPCM.

En el diagrama de bloques del decodificador, se puede apreciar que se usa un cuantizador de adaptación inversa, que entrega localmente la señal en diferencia cuantificada, que al añadirse con la estimación de la señal anterior producida por el predictor adaptativo producen una señal reconstruida. Sobre la base de la señal de diferencia cuantificada y la señal reconstruida, el predictor adaptativo estima la señal posterior. Esta señal reconstruida se cuantifica no uniformemente a través de la ley-A o ley-μ para obtener a la salida una señal de 64kbps.

Cuantizador de Adaptación

En el cuantizador de adaptación se utiliza un error de predicción o señal en diferencia d(k)=s1(k) – se(k), que está representado en logaritmo base 2 antes de la cuantificación y se resta por una señal y(k), generando un factor de escala de cuantización a la entrada/salida, ver siguiente tabla:

Rango de entrada del cuantizador $\log_2\|d(k)\| - y(k)$	$\|I(k)\|$	Rango de salida del cuantizador $\log_2\|d_q(k)\| - y(k)$
$3.16 - \infty$	7	3.34
2.78-3.16	6	2.95
2.42-2.78	5	2.59
2.04-2.42	4	2.23
1.58-2.04	3	1.81
0.96-1.58	2	1.29
-0.05-0.96	1	0.53
$-\infty$-0.05	0	-1.05

Tabla 2.1: Características de la escala del cuantizador adaptativo.

Factor de escala del cuantizador de adaptación

De este bloque deriva el factor de escala y(k) de la Tabla 2.1. Sus entradas son los valores de los 4 bits I(k) y el parámetro de control de la velocidad de adaptación a(k). La escala del cuantizador cambia rápidamente las características de las señales por las cuantiosas fluctuaciones.

El factor de escala rápida $y_u(k)$ es calculado recursivamente a partir del factor de escala logarítmica y(k) previamente introducido en logaritmo base 2.

$$y_u(k) = (1-2^{-5})y(k) + 2^{-5}W[I(k)],$$
$$y_u(k) \approx 0.97y(k) + 0.03 \cdot W[I(k)], \quad \text{.... (2.24)}$$

donde, $y_u(k)$ se limita al rango:

$$1.06 \le y_u(k) \le 10.00 \quad \text{.... (2.25)}$$

Es decir, $y_u(k)$ es una suma ponderada de y(k) e I(k), donde la parte dominante usualmente es y(k). El factor de pérdidas $(1-2^{-5}) \approx 0.971$ permite que el decodificador "olvide" el efecto de los errores de transmisión. El factor W(I) se especifica en la Recomendación G.726 como se ve en la siguiente Tabla:

$\|I\|$	7	6	5	4	3	2	1	0
$W(I)$	69.25	21.25	11.50	6.12	3.12	1.69	0.25	−0.75

Tabla 2.2: Definición del factor W(I). Copyright © CCITT G.726.

El valor del factor de escala lento del cuantizador $y_l(k)$ se deriva del factor de escala rápido $y_u(k)$ y del valor anterior del factor de escala lento $y_l(k-1)$, utilizado,

$$y_l(k) = (1-2^{-6})y_l(k-1) + 2^{-6}y_u(k),$$
$$y_l(k) \approx 0.984y_l(k-1) + 0.016y_u(k). \quad \text{.... (2.26)}$$

Entonces, de acuerdo con la Recomendación G.726, el factor de escala rápido y lento se combinan para formar el factor de escala y(k):

$$y(k) = a_l(k)y_u(k-1) + [1 - a_l(k)]y_l(k-1), \qquad \text{.... (2.27)}$$

donde el factor de control de la velocidad de adaptación está limitado por el rango de $0 \leq a_l \leq 1$, y tenemos $a_l \approx 1$ para señales de voz, mientras que $a_l \approx 0$ para señales de datos. Por lo tanto, para señales de voz el factor de escala rápida $y_u(k)$ domina, mientras que para las señales de datos el factor de escala lento $y_l(k)$ prevalece.

Control de velocidad de adaptación G.726

El control de velocidad de adaptación está basado en dos medidas del valor promedio de I(k). Es decir, d_{ms} describe un promedio a corto plazo de I (k), en tanto d_{ml}, constituye un promedio a largo plazo de este, lo que se define en el estándar G.726 como:

$$\begin{aligned} d_{ms}(k) &= (1 - 2^{-5})d_{ms}(k-1) + 2^{-5}F[I(k)] \\ d_{ml}(k) &= (1 - 2^{-7})d_{ml}(k-1) + 2^{-7}F[I(k)], \end{aligned} \qquad \text{.... (2.28)}$$

donde F[I(k)] está dado en la Recomendación G.726 como se especifica en la Tabla 2.3.

Como resultado del valor cero, la función con valores de ponderación F [I(k)], no toma en cuenta el valor de I(k).

$\lvert I(k)\rvert$	7	6	5	4	3	2	1	0
$F[I(k)]$	7	3	1	1	1	0	0	0

Tabla 2.3: Definición del factor F[I(k)]. Copyright © CCITT G.726.

De los promedios anteriores, la variable $a_p(k)$, que se utilizará en el factor de control de la velocidad de adaptación se define en la Recomendación G.726 como:

$$a_p(k) = \begin{cases} (1 - 2^{-4})a_p(k-1) + 2^{-3} & Si \quad |d_{ms}(k) - d_{ml}(k)| \geq 2^{-3}d_{ml}(k) \\ (1 - 2^{-4})a_p(k-1) + 2^{-3} & Si \quad y(k) < 3 \\ (1 - 2^{-4})a_p(k-1) & \text{De otra manera} \end{cases} \qquad \text{.... (2.29)}$$

Este factor $a_p(k)$, es incrementado y dentro de un largo plazo tiende hacia el valor de 2, si normalizamos la distancia promedio a corto y largo plazo: $[d_{ms}(k) - d_{ml}(k)]/d_{ml}(k) \geq 2^{-3}$, es decir, la magnitud de I(k) estará cambiando. Esto se debe a los dos factores de diferencia de la escala positiva y negativa de la ecuación *(2.29)*.

Por el contrario, el factor de control de la velocidad de adaptación a_p es disminuido y tiende a cero, si la diferencia de los promedios de error de predicción anterior de corto y largo plazo es relativamente pequeña, es decir, si I(k) es casi constante. Esto se debe a la acción continua del decremento del factor de 2^{-4} en la tercera línea de la ecuación *(2.2).* Por otra parte, si el factor de escala satisface y(k) <3, la cantidad $a_p(k)$ se incrementa y también tiende a 2.

Por último, el factor de control de la velocidad de adaptación a_1 utilizado en la ecuación *(2.27),* se obtiene mediante la limitación a_p acorde a la Recomendación G.726, como:

$$a_l(k)\begin{cases} 1 & Si \quad a_p(k-1) > 1 \\ a_p(k-1) & Si \quad a_p(k-1) \leq 1 \end{cases} \qquad \text{.... (2.30)}$$

Esto hace que la operación que limita el valor real de $a_p(k)$ no afecte, en tanto que es más grande que 1.

Predicción adaptativa y reconstrucción de la señal.

La estimación de la señal $s_e(k)$ y la señal de diferencia cuantificada $d_q(k)$, se muestran en la Figura 2.13.
Donde la función de transferencia del predictor se caracteriza por tener seis ceros y dos polos. Este modelo de predictor de polos y ceros hace que la envolvente espectral tenga una amplia variedad de señales de entrada de manera eficiente. Nótese, que el algoritmo de Levinson-Durbin es aplicable al problema de un modelo único para todos los polos. La señal reconstruida está dada por:

$$s_r(k-i) = s_e(k-i) + d_q(k-i), \quad \text{.... (2.31)}$$

donde la estimación de la señal $s_e(k-i)$ deriva de una combinación lineal de las muestras anteriores reconstruidas y las diferencias anteriores del cuantificador d_q, utilizando,

$$s_e(k) = \sum_{i=1}^{2} a_i(k-1)s_r(k-i) + \sum_{i=1}^{6} b_i(k-1)d_q(k-i), \quad \text{.... (2.32)}$$

donde los factores a_i, i = 1,..., 2, y b_i, i = 1,. . . , 6, representan los coeficientes de predicción. Ambos coeficientes de predicción $a_i(k)$ y $b_i(k)$ se calculan de forma recursiva mediante un algoritmo de gradiente.

Conforme a la Recomendación G.726 los coeficientes de predicción de segundo orden se especifican como:

$$a_1(k) = (1-2^{-8})a_1(k-1) + (3\cdot 2^{-8})\,\mathrm{sgn}[p(k)]\,\mathrm{sgn}[p(k-1)], \quad \text{.... (2.33)}$$

donde $a_1(k)$ es dependiente de $a_1(k-1)$, así como en la polaridad de las dos muestras consecutivas de la variable p(k), donde,

$$p(k) = d_q(k) + \sum_{i=1}^{6} b_i(k-1)d_q(k-i) \quad \text{.... (2.34)}$$

representa la suma de los actuales errores de predicción cuantificados $d_q(k)$ y la estimación de la señal a la salida del orden sexto del predictor, debido a los valores anteriores de $d_q(k)$, mientras usamos los coeficientes $b_i(k-1)$. Específicamente, cuando se actualiza, a_1 se incrementa en el segundo término de la ecuación *(2.33)*. Nótese, sin embargo, que esta adaptación es un proceso lento debido al factor de escala de $(3 - 1)\cdot 2^{-8}$.

El segundo coeficiente de predicción se especifica de la siguiente manera:

$$a_2(k) = (1-2^{-7})a_2(k-1) + 2^{-7}\{\mathrm{sgn}[p(k)]\mathrm{sgn}[p(k-1)] - f[a_1(k-1)]\mathrm{sgn}[p(k)]\mathrm{sgn}[p(k-1)]\}, \quad \text{.... (2.35)}$$

donde la función $f(a_1)$ está dada por:

$$f(a_1)\begin{cases} 4a_1 & Si \quad |a_1| \le 1/2 \\ 2\,\mathrm{sgn}(a_1) & Si \quad |a_1| > 1/2 \end{cases} \quad \text{.... (2.36)}$$

El predictor de sexto orden se actualiza como se muestra en la siguiente ecuación:

$$b_i(k) = (1-2^{-8})b_i(k-1) + 2^{-7}\,\mathrm{sgn}[d_q(k)]\mathrm{sgn}[d_q(k-i)] \quad para \quad i = 1,...,6, \quad \text{.... (2.37)}$$

donde los coeficientes del predictor están contenidos en el rango de $-2 \le b_i(k) \le 2$.

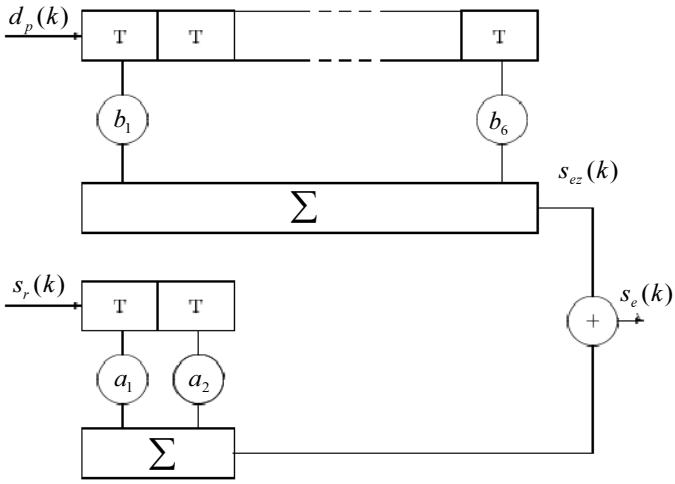


Figura 2.13: Adaptación seis-ceros, predictor dos polos del códec G.726.

Simulación de la modulación ADPCM

La codificación predictiva tiene el objetivo de aprovechar la redundancia que existe en la señal a codificar. Para la modulación ADPCM se usan escalones variables para codificar la diferencia, lo que nos permite comprimir las muestras, sin bajar la calidad de audio, siendo esta la mejora del DPCM.

En cuanto al desempeño se refiere vemos que la señal que paso a través del sistema ADPCM es idéntica a la señal original.

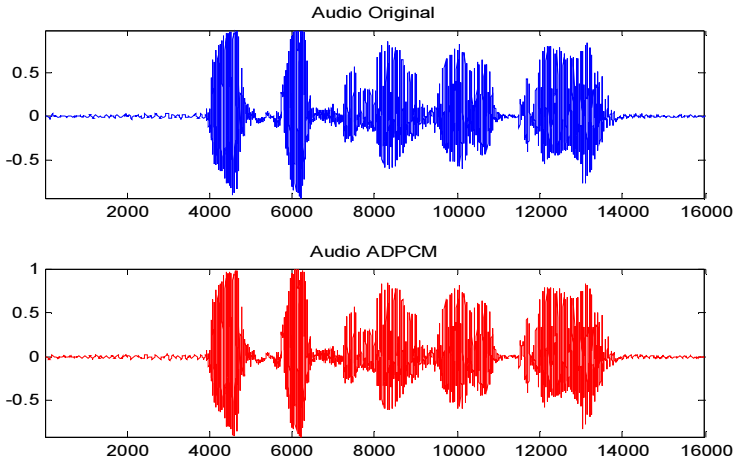


Figura 2.14: Una señal de voz modulada en ADPCM utilizando Matlab.

Capítulo 3

Codificación de la Voz en Banda Ancha

La banda ancha se ha convertido rápidamente en los últimos años en el principal vehículo de desarrollo para los diversos servicios, aplicaciones y contenidos. Sin embargo, resulta insuficiente la capacidad de ancho de banda en la actualidad, lo que ha propiciado la evolución de los mecanismos para la compresión de las señales de voz para la optimización de la información.

Las técnica de codificación de voz buscan mantener una mejora en la inteligibilidad y naturalidad de la voz, la intención es reducir la tasa de bits manteniendo la calidad y robustez del algoritmo frente a errores de transmisión.

Con el modelo CELP se incorporan codificadores como AMR (Adaptative Multi-Rate) para la optimización de audio y esto nos proporciona nuevas capacidades en los sistemas de comunicaciones futuros.

En el tercer capítulo veremos algunas técnicas de codificación cuyo auge comenzó con la telefonía celular y recientemente en los sistemas de comunicaciones inalámbricos de banda ancha ha permitido el empleo de aplicaciones innovadoras y servicios para sistemas móviles de la cuarta generación.

3.1 LPC (Codificación Predictiva Lineal-Linear Predictive Coding)

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

La principal limitación de los codificadores LPC es la consideración de que las señales de la voz son sonoras o no sonoras, de ahí que la fuente de excitación de la señal de voz para el filtro de síntesis de la predicción lineal sea un tren de pulsos (para señales sonoras) y ruido aleatorio (para señales no sonoras). Esta consideración es una simplificación demasiado grande para conseguir una buena calidad de la señal de voz.

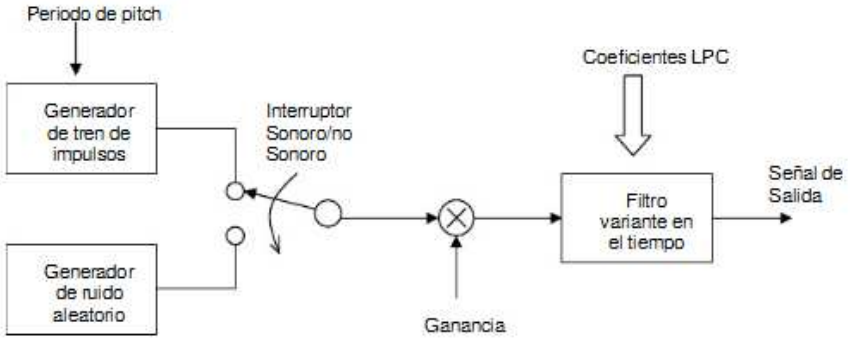


Figura 3.1: Diagrama del modelo simplificado de la producción de la voz.

3.2 Codificación predictiva mediante análisis por síntesis (AbS)

En 1982, Atal propuso un nuevo modelo de excitación conocido como excitación por multi-pulso (MP-LPC), siendo un modelo parecido al de los codificadores de forma de onda y no requiere ningún conocimiento a priori sobre si la señal de voz es sonora o no sonora. La excitación se modela por un número de pulsos cuyas amplitudes y posiciones son determinadas mediante un proceso en circuito cerrado, y no usando el error entre el residuo y su versión cuantizada, como lo hacen los codificadores que usan circuito abierto. Este modelo dio paso a una nueva generación de codificadores de voz mediante el análisis por síntesis capaces de producir alta calidad de señal de voz a tasas de bit sobre 10 kbps, llegando incluso a los 4.8 kbps. Esta nueva generación decodificadores usa el mismo filtro de síntesis todos polos.

Estos codificadores cuentan con una estructura básica en la cual la excitación es calculada minimizando el error porcentual ponderado entre la señal de voz original y la señal sintetizada. La diferencia reside en la forma de modelar la excitación. El modelo tiene una serie de parámetros que pueden variar los rangos de la señal de voz sintetizada. La complejidad de estos codificadores aumenta a medida que disminuye la tasa de bits. La estructura básica de un sistema de codificación LPC mediante análisis por síntesis se muestra en la Figura 3.2.

El modelo consta de los siguientes bloques:

El generador de excitación, el cual produce una secuencia de excitación que se ingresa al filtro de síntesis para producir la señal reconstruida en el receptor. Como puede apreciarse en el modelo existe un decodificador incluido dentro del codificador. Para optimizar la excitación, el método de análisis usa la diferencia entre la señal de voz original y la sintetizada como un criterio de error, y elige la secuencia de excitación que minimiza ese error ponderado. Los filtros de síntesis, son filtros lineales variantes en el tiempo, ya que sus coeficientes van cambiando en cada iteración del circuito. Podemos tener, un predictor de síntesis de corto plazo (STP, Short Term Predictor), que modela la envolvente espectral de la forma de onda de la señal de voz o un predictor de largo plazo (LTP, Long term Predictor) para modelar la estructura suave del espectro de la señal de voz. El minimizador de error, es el encargado de minimizar la diferencia entre la señal original y la señal sintetizada. El criterio más usado es el error cuadrático medio (MSE). Generalmente, se pasa el error por un filtro de ponderación, por lo que el ruido queda enmascarado por la señal de voz.

El procedimiento de decodificación se realiza pasando la señal de excitación decodificada a través los filtros de síntesis, proceso que da como resultado la señal de voz reconstruida. Cabe destacar que tanto en el codificador como en el decodificador, se genera una señal de voz sintetizada. Esto es necesario para actualizar los contenidos de memoria de los filtros de síntesis.

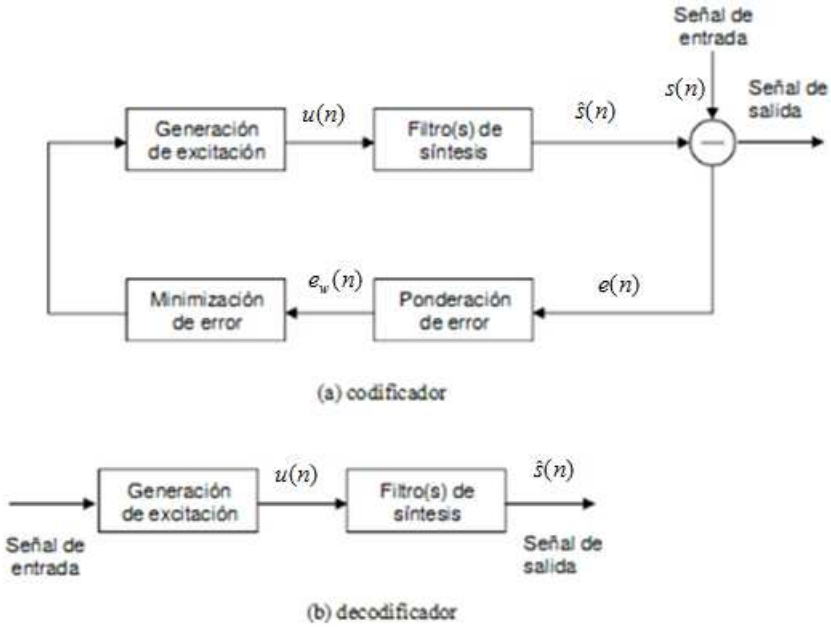


Figura 3.2: Modelo de Codificación LPC mediante análisis por síntesis.

Predictor de corto plazo

El predictor de corto plazo modela la envolvente espectral de un segmento de voz de longitud L de una muestra, está se puede aproximar mediante una función de transmisión de un filtro digital todo-polos de la siguiente forma:

$$H(z) = \frac{G}{A(z)} = \frac{G}{1 - P_\delta(z)} = \frac{G}{1 - \sum_{k=1}^{P} a_k z^{-k}} \quad \text{.... (3.1)}$$

donde

$$P_\delta(z) = \sum_{k=1}^{P} a_k z^{-k} \quad \text{.... (3.2)}$$

es el predictor de corto plazo, donde los coeficientes a_k son los coeficientes del predictor o los parámetros LPC. El número de coeficientes *p* es el orden del predictor.

Predictor de largo plazo

Mientras que el predictor de corto plazo modela la envolvente espectral del segmento de voz que está siendo analizado, el predictor de largo plazo o predictor de pitch, se usa para modelar la estructura suave de esa envolvente.

El filtrado inverso de la señal de voz de entrada elimina la envolvente del espectro de la señal, es decir, elimina algo de la redundancia de la voz tomando de la muestra de voz su valor predicho usando las *p* muestras anteriores. A esto se le denomina predicción de corto plazo. Sin embargo, el residuo de esa predicción todavía muestra considerables variaciones en su espectro, es decir, que todavía existen correlaciones de largo plazo entre muestras de la señal, especialmente en las regiones sonoras. Por tanto, aún existe alguna periodicidad (redundancia), relacionada con el periodo de pitch de la señal de voz original, que el análisis LP no puede eliminar. De ahí la necesidad de añadiendo un predictor de pitch al filtro inverso para elimina esa redundancia en el residuo de la señal y éste se convierte en ruido.

Se le llama predictor de pitch, ya que elimina la periodicidad de la señal o predictor de largo plazo, ya que su retraso está comprendido entre 20 y 160 muestras.

Este predictor de largo plazo es básico en los codificadores de voz con tasas de bits baja, como el CELP, donde la señal de excitación se modela con un proceso de Gaussiano y, por tanto, el predictor es necesario para asegurar que el residuo de la predicción sea lo más cercano posible a ruido aleatorio Gaussiano.

La forma general del filtro de correlación de largo plazo es:

$$\frac{1}{P(z)} = \frac{1}{1 - p_1(z)} = \frac{1}{1 - \sum_{k=m_1}^{m_2} G_k z^{-(\alpha+k)}} \quad \text{.... (3.3)}$$

La estabilidad del filtro de síntesis de pitch 1/P(z) no siempre está garantizada. La condición de estabilidad es $|G| \leq 1$. Por tanto, la estabilidad del filtro se puede conseguirse fácilmente fijando $|G| = 1$, cuando $|G| > 1$ la inestabilidad de este filtro no es tan perjudicial para la calidad de la señal reconstruida. El filtro inestable permanece durante unas tramas, pero al final, se encuentran periodos con el filtro estable, por lo que la salida no continua aumentando con el tiempo. Cuando se usa el predictor de largo plazo, el esquema general del codificador queda de la siguiente manera:

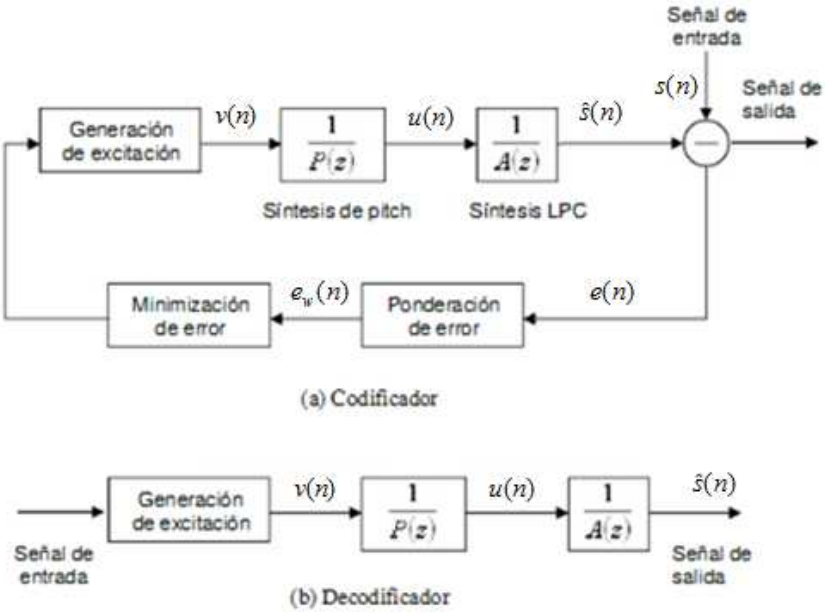


Figura 3.3: Diagrama de bloques de un codificador y decodificador de largo plazo.

Una mejora significativa se consigue cuando los parámetros LTP se optimizan dentro del circuito mediante el análisis por síntesis; es la aproximación por códigos adaptativos, el cálculo de los parámetros contribuye directamente al proceso de minimización del error ponderado.

Simulación del Códec LPC

La utilización de este tipo de codificación hace imposible reconocer, a partir de la voz sintetizada a la persona que origina la señal de voz. LPC basa su funcionamiento en dos tipos de sonidos con voz y sin voz, por lo que no puede representar los otros tipos de sonidos existentes, resultado de esto es que la voz sintetizada tenga una calidad muy inferior a la obtenida a través de las técnicas de PCM y ADPCM. Su principal ventaja del uso de LPC es su capacidad de producir voz inteligible a muy bajas velocidades.

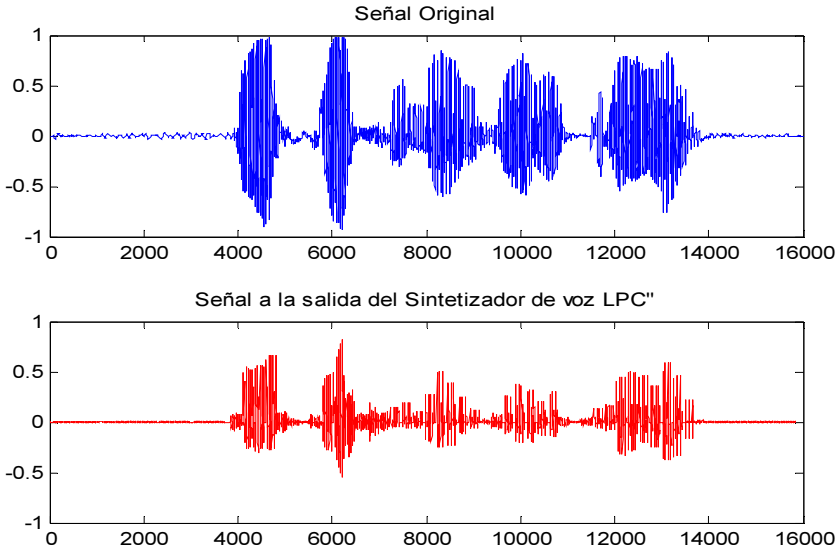


Figura 3.4: Una señal de Voz Sintetizada por un Predictor Lineal (LPC) utilizando Matlab.

3.3 Codificación en Sub-banda ADPCM

El funcionamiento de la codificación en sub-banda (SBC), se basa en que la señal de voz está inicialmente dividida en una serie de sub-bandas, que son codificadas por separado. La principal atracción de los códecs en sub-banda es que permiten una asignación de bit arbitraria que se aplicará a cada sub-banda de acuerdo a su percepción de importancia, por lo tanto, limita el ruido de cuantificación correspondientes a las sub-bandas en cuestión.

Entonces los bits de salida generados por los codificadores de sub-banda se multiplexan y se transmiten al receptor, donde después de demultiplexar y decodificar cada señal de sub-banda, la señal de banda completa se reconstruye mediante la combinación de las distintas sub-bandas. El éxito de esta técnica depende del diseño adecuado de la división de bandas en el análisis y la síntesis de filtros, que no se interfieran unos con otros en sus bandas de transición, es decir, evitar la introducción de la llamada distorsión por aliasing inducida por sub-bandas superpuestas debido a una frecuencia de muestreo lo suficientemente alta. Si, por otro lado, la frecuencia de muestreo es demasiado alta, el banco de filtros empleado genera un hueco espectral, una vez más, sufre la calidad de la voz. El organismo ITU ha estandarizado un códec de banda ancha conocido como G722 [15] modulación por impulsos codificados diferencial adaptativa por división de sub-bandas (SB-ADPCM) a una velocidad binaria de hasta 64 kbps.

El estándar del códec G722 abarca las siguientes especificaciones:

(1) La calidad de la voz es mejor que la empleada en PCM a 128 kbps que tenía como objetivo la calidad de la codificación de las señales de música.
(2) No hubo consideración para la transmisión de datos en banda vocal o en la señalización dentro de banda.
(3) El códec fue obligado a no tener una degradación significativa de calidad en un BER de 10^{-4} y tener un mejor rendimiento en 10^{-3}.
(4) El retardo total se ha especificado en menos de 4 ms.
(5) Era necesario para dar cabida a un canal de datos a costa de una reducción de la calidad de voz, por lo que se definieron los siguientes tres modos de funcionamiento: Modo 1 – voz sólo a 64 kbps; Modo 2 – 56 kbps de voz más 8 kbps de datos; y Modo 3 – 48 kbps de voz más 16 kbps de datos. Los dos últimos modos permiten un canal de datos mediante el uso de bits de la sub-banda inferior.

Modo	Velocidad de codificación de la voz	Velocidad del canal de datos
1	64 kbps	0
2	56 kbps	8 kbps
3	48 kbps	16 kbps

Tabla 3.1: Especificación del Códec G722.

Descripción funcional del codificador de Sub-banda ADPCM

En el diagrama de bloques que se muestra en la Figura 3.5, donde la banda completa de la señal de entrada $X(n)$ se divide en dos señales de sub-banda, es decir, la componente de banda superior $X_h(n)$ y la componente de banda inferior $X_L(n)$. La operación de división de banda es llevada a cabo por los filtros de espejo en cuadratura (FEC o en inglés *Quadrature Mirror Filter* (QMF)) sin aliasing.

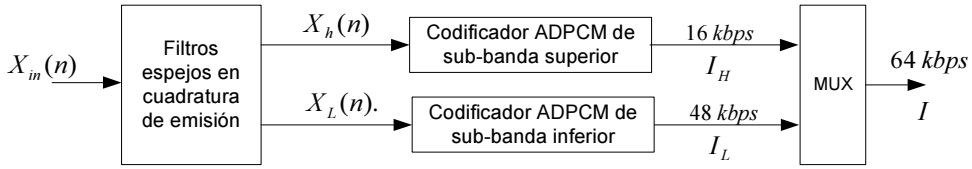


Figura 3.5: Diagrama de bloques del codificador G722 SB-ADPCM.

Filtros de espejo en cuadratura (FEC) de emisión

La etapa de FEC está constituida por dos filtros digitales no recursivos FIR de fase lineal, cuyo impulso de respuestas son simétricos. Estos filtros dividen la banda de frecuencia de 0-8000 Hz en dos sub-bandas: la sub-banda inferior a 0-4000 Hz y la sub-banda superior a 4000-8000 Hz.

La entrada del FEC de emisión, x_{in}, es la salida de la parte del audio de emisión, y se muestrea a 16 kHz. Las salidas, X_L y X_h, de las sub-bandas inferior y superior, respectivamente se muestrean a 8 kHz debido a la reducción a la mitad de su ancho de banda.

La banda inferior de 0-4000 Hz mantiene una proporción significativamente mayor de energía que la señal de banda superior. Por lo tanto, se codifica con 6 bits/muestra en la codificación ADPCM, es decir, se tiene 8 kHz * 6 bits/muestra = 48 kbps en el modo 1. La banda de menor importancia 4000-8000 Hz está codificada con 2 bits/muestra, es decir, a 16 kbps. Las señales resultantes se indican en la Figura 3.5 por la I_L y I_H, las cuales son multiplexadas para su transmisión por el canal digital

Codificador ADPCM de sub-banda inferior

En el diagrama de bloques del codificador ADPCM de sub-banda inferior Figura 3.6. La señal de entrada de la sub-banda inferior, X_L, tras la sustracción de una estimación de la señal de entrada, S_L, produce la señal en diferencia, e_L. Se utiliza un cuantificador adaptativo no lineal de 60 niveles para asignar 6 dígitos binarios al valor de la señal en diferencia y producir una señal a 48 kbps, I_L.

En el circuito de realimentación se suprimen los dos bits menos significativos de I_L para producir una señal de cuatro bits, I_{Lt}, que se utiliza para la adaptación del cuantificador y se aplica a un cuantificador adaptativo inverso de 15 niveles para producir una señal en diferencia cuantificada, d_{Lt}. La estimación de la señal, L, se suma a esta señal en diferencia cuantificada para producir una versión reconstruida de la señal de entrada de sub-banda inferior, r_{Lt}. Tanto la señal reconstruida como la señal en diferencia cuantificada se aplican a un predictor adaptativo que produce la estimación de la señal de entrada, S_L, completando así el circuito de realimentación.

El funcionamiento con 4 bits, en lugar de 6 bits, en el circuito de realimentación del codificador ADPCM de sub-banda inferior y del decodificador ADPCM, ofrece la posibilidad de insertar datos en los bits menos significativos, sin provocar un funcionamiento incorrecto del decodificador.

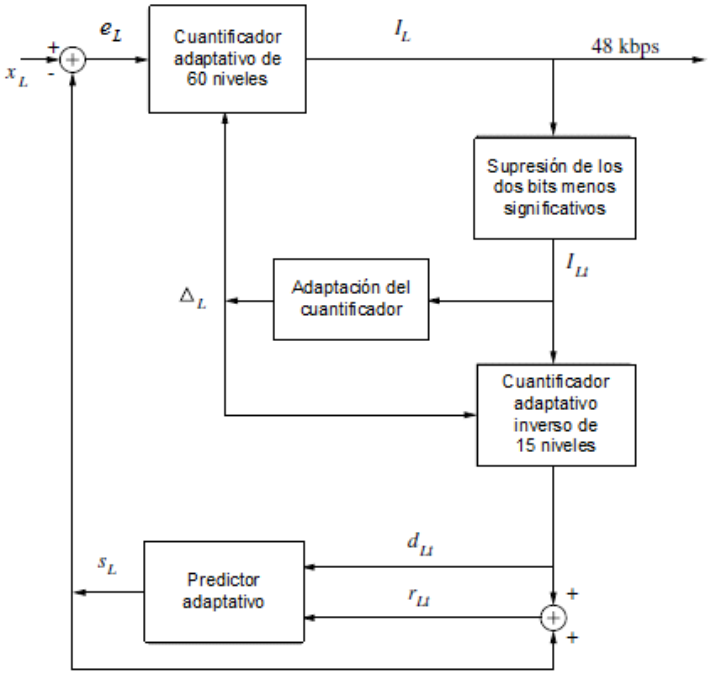


Figura 3.6: Diagrama de bloques del codificador ADPCM de la sub-banda inferior.

Codificador ADPCM de sub-banda superior

En el diagrama de bloques del codificador ADPCM de sub-banda superior Figura 3.7. La señal de entrada de la sub-banda superior, x_H, tras la sustracción de una estimación de la señal de entrada, S_H, produce la señal en diferencia, e_H. Se utiliza un cuantificador adaptativo no lineal de cuatro niveles para asignar dos dígitos binarios al valor de la señal en diferencia y producir una señal a 16 kbps, I_H.

Un cuantificador adaptativo inverso produce una señal en diferencia cuantificada, d_H, a partir de estos mismos dos dígitos binarios. La estimación de la señal, S_H, se suma a esta señal en diferencia cuantificada para producir una versión reconstruida de la señal de entrada de sub-banda superior, r_H. Tanto la señal reconstruida como la señal en diferencia cuantificada se aplican a un predictor adaptativo que produce la estimación de la señal de entrada, S_H, completando así el circuito de realimentación.

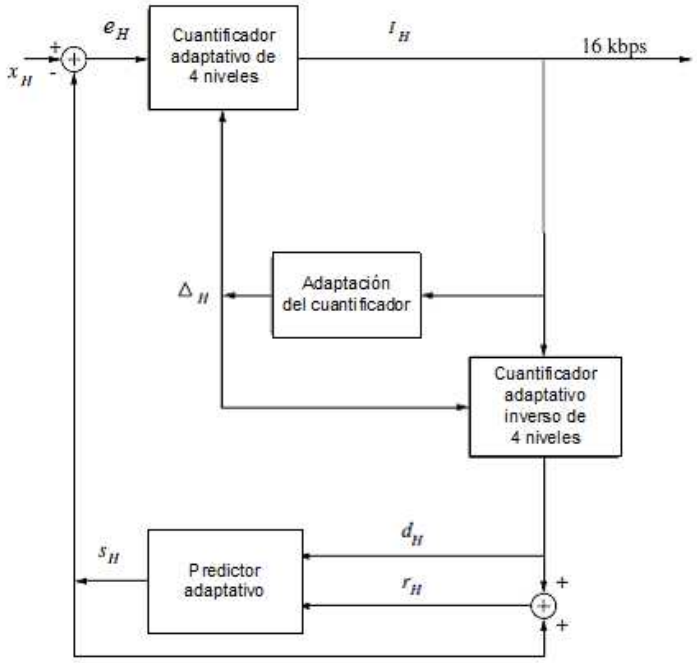


Figura 3.7: Diagrama de bloques del codificador ADPCM de sub-banda superior.

Multiplexor

El multiplexor (MUX) se utiliza para combinar las señales I_L e I_H procedentes de los codificadores ADPCM de las sub-bandas inferior y superior, respectivamente, para formar una señal compuesta de 64 kbps, *I*, con formato de octeto para su transmisión.

El formato de los octetos de salida, tras la multiplexación, es el siguiente:

$$I_{H1}\ I_{H2}\ I_{L1}\ I_{L2}\ I_{L3}\ I_{L4}\ I_{L5}\ I_{L6}$$

donde I_{H1} es el primer bit transmitido; I_{H1} e I_{L1} son los bits más significativos de I_H e I_L respectivamente.

Descripción funcional del decodificador SB-ADPCM

En diagrama de bloques del decodificador SB-ADPCM se muestra en la Figura 3.8.

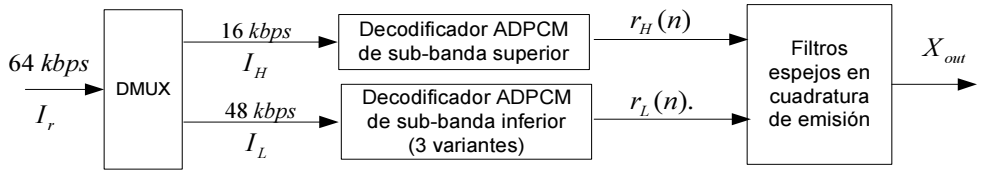


Figura 3.8: Diagrama de bloques del decodificador G722 SB-ADPCM.

Demultiplexor

El demultiplexor (DMUX) descompone la señal de 64 kbps recibida con formato octeto, I_r, en dos señales, I_{Lr} e I_H, que forman las entradas de palabra de código a los decodificadores ADPCM de las sub-bandas inferior y superior, respectivamente.

Decodificador ADPCM de sub-banda inferior

El trayecto que produce la estimación de la señal de entrada, S_L, incluida la adaptación del cuantificador, es idéntica a la parte de la realimentación del codificador ADPCM de sub-banda inferior. La señal reconstruida, r_L, se produce sumando a la estimación de la señal una de las tres posibles señales en diferencia cuantificadas, $d_{L,6}, d_{L,5}, d_{L,4}$ (d_{Li}), seleccionada según la indicación recibida del modo de funcionamiento.

Este decodificador puede funcionar en cualquiera de las tres siguientes variantes, según la indicación recibida del modo de funcionamiento.

Modo de funcionamiento	Señal diferencia cuantificada seleccionada	Cuantificador adaptativo inverso utilizado	Número de bits menos significativos I_{Lr}
Modo 1	$d_{L,6}$	60 niveles	0
Modo 2	$d_{L,5}$	30 niveles	1
Modo 3	$d_{L,4}$	15 niveles	2

Tabla 3.2: Variantes del decodificador ADPCM de sub-banda inferior.

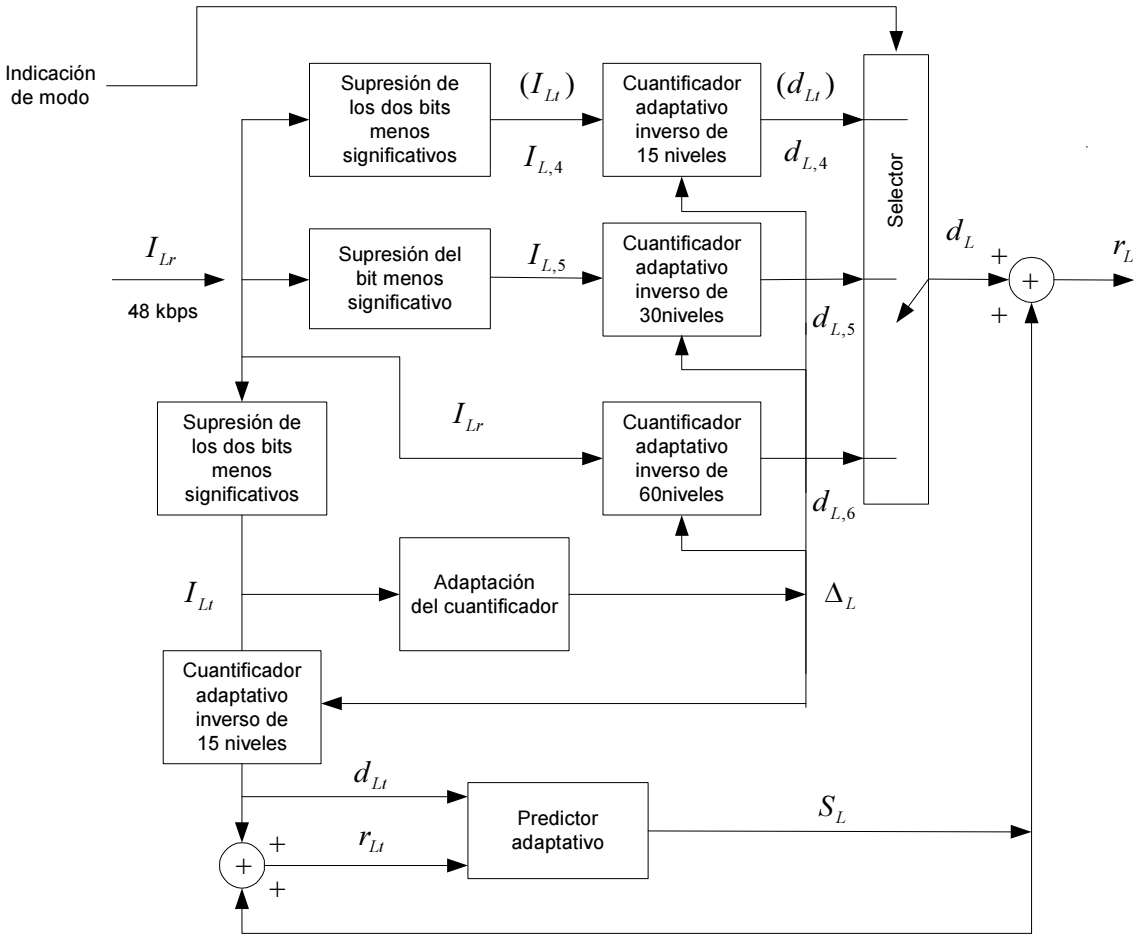


Figura 3.9: Diagrama de bloques del decodificador ADPCM de sub-banda inferior.

Decodificador ADPCM de sub-banda superior

Este decodificador es idéntico a la parte de realimentación del codificador ADPCM de sub-banda superior, el cual fue descrito anteriormente, siendo la señal reconstruida a la salida, r_H .

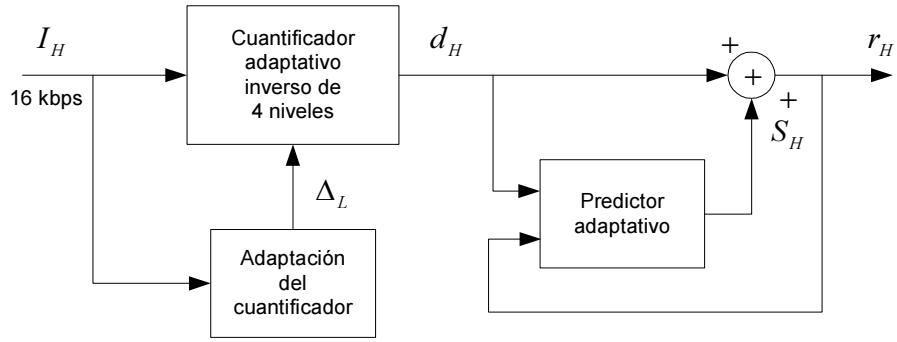


Figura 3.10: Diagrama de bloques del decodificador ADPCM de sub-banda superior.

Filtros de espejo en cuadratura (FEC) de recepción

Los FEC de recepción son dos filtros digitales no recursivos de fase lineal que interpolan las salidas, r_L y r_H, de los decodificadores ADPCM de las sub-bandas inferior y superior, de 8 kHz a 16 kHz, y que producen entonces una salida, X_{out}, muestreada a 16 kHz.

Simulación del Códec de Sub-Banda G722

En este sistema de codificación utilizamos la modulación por impulsos codificados diferencial adaptativa de sub-banda (ADPCM-SB) a una velocidad binaria de hasta 64 kbit/s. En esta técnica la banda de frecuencias se divide en dos sub-bandas (superior e inferior) Figura 3.11, y las señales de cada una se codifican utilizando ADPCM, y finalmente el multiplexor combina ambas señales obteniendo a la salida una sola señal (en el dominio de la frecuencia y en el dominio del tiempo) Figura 3.12.

Este tipo de señales se utiliza en diversas aplicaciones de voz de alta calidad.

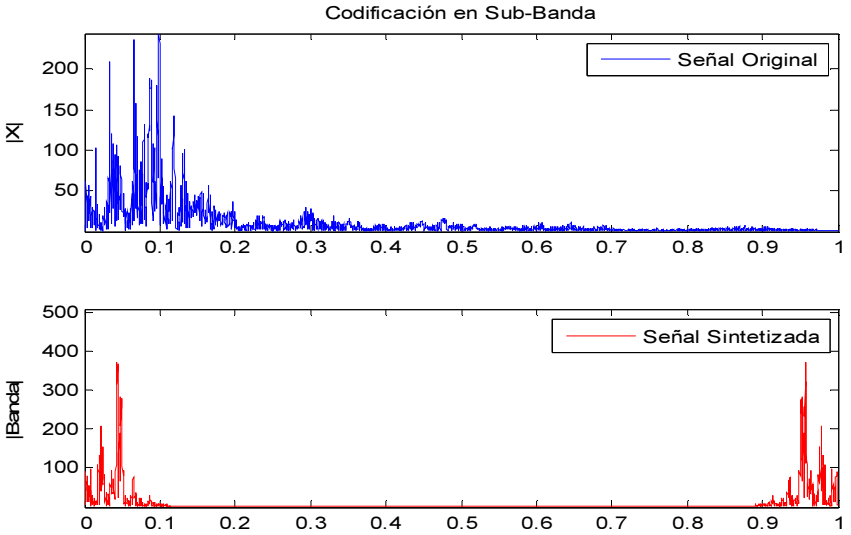
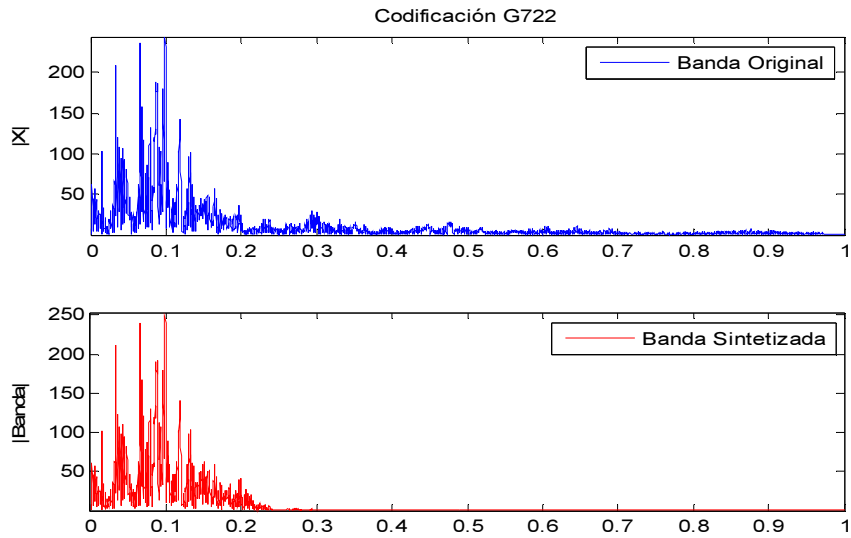


Figura 3.11: División de bandas superior e inferior utilizando Matlab.



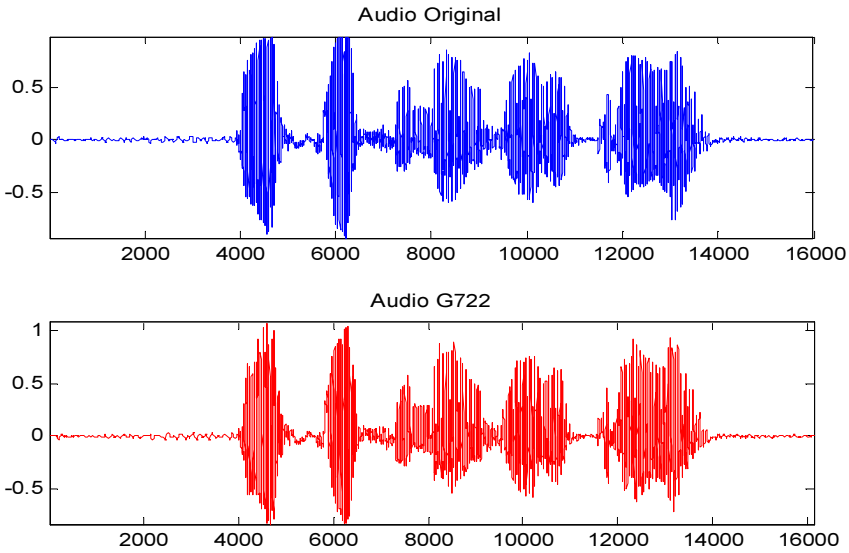


Figura 3.12: Codificación G722 utilizando Matlab.

3.4 Transformar la codificación de banda ancha a 32 kbps

El esquema de transformación por codificación para la voz en banda ancha propuesto por Quackenbush, en el cual se procesa 7 kHz de ancho de banda de la voz muestreada a 16 kHz. Este códec logra una compresión de 8:1[2], cuando se compara con la calidad de la señal de PCM. En su contribución, Quackenbush adopto el método de transformación por codificación propuesto por Johnston para señales de audio y con ella redujo la tasa de bits requerida.

Algoritmo de transformación por codificación

En el diagrama de bloques que se muestra en la Figura 3.13, se procesan 240 muestras por bloque, que corresponden a 15 ms a una tasa de muestreo de 16 kHz y se concatenan 16 muestras desde el bloque anterior, dando una longitud total por bloque de 256 muestras, las cuales son entonces transformadas a 128 coeficientes complejos utilizando la FFT, donde los coeficientes se cuantifican para la transmisión.

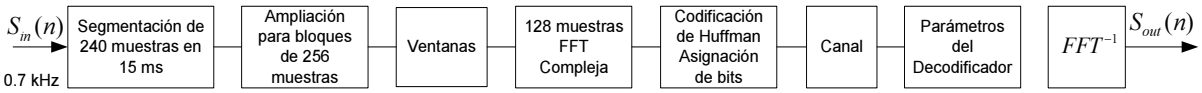


Figura 3.13: Diagrama de bloques de la transformación por codificación para la voz en banda ancha.

La codificación distribuye el ruido de cuantificación en el dominio espectral, tal que sus efectos perceptuales se reducen al mínimo mediante el ajuste de la relación señal a ruido a través de una cuantificación adecuada en la banda de frecuencia. Este proceso también podría llevarse a cabo convenientemente usando una etapa de FEC para dividirla banda de frecuencias dentro del ancho requerido por las sub-bandas, como hemos visto en el caso del códec G722.

Quackenbush siguió las sugerencias de Scharf con el fin de determinar el umbral tolerable al ruido T_i para la banda de frecuencia i, siendo i el llamado índice de banda crítica. Quackenbush optó por fija en vez de ajustar dinámicamente la evaluación de la energía de la banda crítica, determinando la energía C_i para la banda i desde el análisis a largo plazo de la voz.

Este modelo de enmascaramiento nos permite determinar la asignación de bits requerida en función de la frecuencia, lo que se lleva a cabo dinámicamente utilizando un procedimiento iterativo. En cada trama, 26 bits tienen un variante de tiempo asignada, la asignación de 16 bits para la más baja frecuencia de FFT, 4 bits indican el número de iteraciones durante el proceso de asignado de bits, lo que puede ser en consecuencia 16, 2 bits para la selección de uno de los cuatro codebooks de Huffman y 4 bits para la sincronización de la trama. Por lo tanto, en el supuesto de 2 bits/muestra de codificación, hay 480–76=454 bits para la codificación dinámica en el dominio espectral de los coeficientes de la FFT.

Parámetros	Nº de bits/15 ms
Frecuencia más baja de la FFT	16
Nº de iteraciones para asignación de bits	4
Selección de Codebook de Huffman	2
Tramas de sincronización	2
Coeficientes de la FFT	454
Total	480 bits/15ms = 32kbps

Tabla 3.3: Asignación de los bits para el códec de banda ancha a 32 kbps.

Después de un inicio tentativo de asignación de bits, la asignación de bits iterativo se activa. El espectro FFT se subdivide en 16 bandas de frecuencia y "sub-bandas" k = 1,...,11 son asignadas 6 líneas espectrales de la FFT, mientras que las "sub-bandas" k = 12, ...,16 son asignadas 12 líneas espectrales.

De lo anterior $11 \cdot 6 + 5 \cdot 12 = 126$ líneas espectrales son codificadas por la técnica iterativa, cabe destacar que la línea 0 tiene una asignación fija de 16 bits y la línea 127 no está codificada.

Entonces la máxima magnitud espectral de M_k de cada "sub-banda" k se encuentra y se cuantifica logarítmicamente, obteniéndose.

$$m_k = \{\log_2 M_k\}, \quad k = 1,...,16, \ldots. \quad (3.4)$$

Una vez que el máximo espectral de la "sub-banda" y las líneas espectrales están codificadas, la tasa de bits puede mejorar aún más mediante la codificación Huffman. Dos conjuntos de codebook se utilizan tanto para el máximo como para las líneas espectrales. Quackenbush[2] argumenta que esta técnica no mejoran dramáticamente la tasa de bits promedio global, pero reduce la tasa máxima.

La codificación de Huffman es una técnica sencilla, donde los mensajes se codifican en orden descendente y le asigna un código de longitud variable en base a su probabilidad de ocurrencia. Específicamente, los mensajes más frecuentes son codificados utilizando un bajo número de bits, mientras que los mensajes poco frecuentes pueden ser transmitidos utilizando códigos más largos.

Volviendo al proceso del máximo espectro de codificación de Huffman, las 16 "sub-bandas" de Quackenbush forman un codebook por separado. Para las líneas espectrales de un esquema más complejo se utilizan 1, 2 o 3 líneas espectrales complejas que se concatenan en una sola palabra antes de invocar a una determinada tabla decodificación de Huffman.

La elección de la tabla de decodificación de Huffman se rige por el valor de m_k y por lo tanto, parte de la información no tuvo que ser enviada al decodificador hasta que m_k fuera transmitida por todas las "sub-bandas". La elección de los codebooks de Huffman está resumida en la Tabla 3.4.

Condición	Nº de niveles de Cuantización	Codebook	Longitud del vector complejo
$15 < \lvert m_k \rvert$	1771	5	1(real)
$7 < \lvert m_k \rvert \le 15$	31	4	1
$3 < \lvert m_k \rvert \le 7$	15	3	1
$1 < \lvert m_k \rvert \le 3$	7	2	2
$0 < \lvert m_k \rvert \le 1$	3	1	3

Tabla 3.4: Esquema de codificación de Huffman sobre la máxima "sub-banda".

Específicamente, uno de los cinco codebooks de Huffman implicados en la Tabla 3.4 se invoca en cada una de las "sub-bandas". Todas las líneas espectrales pertenecientes a esta banda son codificadas por el mismo libro.

Si $\lvert m_k \rvert = 0$, no hay bits asignado a la banda.

Quackenbush utilizó el siguiente método de diseño iterativo de codebooks. Inicialmente un conjunto de cuantizadores lineales simples fueron usados con el fin de generar un histograma de las cantidades que son codificadas, como el máximo de la "sub-banda", la línea espectral, y la estimación de la tasa de bits generada. Entonces los codebooks de Huffman fueron generados sobre la base de estos histogramas tanto para el máximo de la "sub-banda" como para las líneas espectrales.

Estos codebooks se utilizaron luego dentro de una sesión posterior para estimar otra vez la tasa de bits, y por último una nueva serie de histogramas se genera de nuevo. Estas iteraciones pueden ser repetidas un número de veces con el fin de llegar a un conjunto casi óptimo de codebooks de Huffman.

Como se mencionó anteriormente, hay dos codebooks de Huffman, tanto para el máximo de la "sub-banda" como para las líneas espectrales. Primero, m_k es codificado tentativamente invocando ambos codebooks para cada máxima "sub-banda" y el resultante es una menor tasa de bits seleccionada. Esta parte de la información se indica también en la señal del decodificador. Entonces dependiendo de $|m_k|$, el conjunto correspondiente de codebook de la Tabla 3.4 se utiliza, verificando el número de bits generado por ambos codebooks. El número de bits generados también se almacena.

Quackenbush también sugirió un mecanismo de control de tasa de bits iterativo para mantener una tasa de 32 kbps o 480 bits por 15 ms. Si después de la primera codificación la tasa de bits no está entre 1.9 y 2 bits/muestra, correspondiente a 456. ..480 bits por trama, entonces una nueva codificación cíclica se produce. Un factor de escala entero de m se introdujo para controlar el número de niveles del cuantificador utilizado y en cada iteración las líneas espectrales cuantificadas se escalan por un factor de $2^{(1/m)}$, hasta que el número de bits generados esté entre 456 y 480. Explícitamente, si el número de bits generados es también bajo, el número de niveles de cuantificación se incrementa y viceversa.

El valor del factor de escala m es determinado de la siguiente forma: Si el número de bits producidos por la cuantificación de línea espectral inicial es inferior a 456, m se establece en 1, de lo contrario a -1. Esto podría implicar un cuantificador de línea espectral de escala hacia arriba o hacia abajo por un factor de 2 o 1/2, respectivamente. Durante las iteraciones posteriores, se observa si la dirección de la escala se mantiene y si el valor anterior de m se conserva. Sin embargo, si la dirección de la escala tiene que cambiar, debido aún paso de corrección previo ahora el número de bits generados se desvía del objetivo en la dirección opuesta, el valor de m es duplicado y el signo es activado, antes de que la siguiente iteración se lleve a cabo.

3.5 Codificación CELP (Predicción Lineal Extendida por el Código - Code Excited Linear Prediction) de banda ancha con división de Sub-banda

Los códecs CELP son exitosos en codificar señales de voz de banda estrecha, también han sido empleados en la codificación de banda ancha. Ordentlich y Shoham propusieron un códec de banda ancha a 32 kbps basado en CELP de bajo retraso, el cual alcanza una calidad similar de voz al códec G722 de 64 kbps en una complejidad más alta.

Codificación CELP de banda ancha basado en Sub-bandas

Uno de los problemas asociados con la codificación de voz de banda ancha es la incapacidad de los códecs para tratar de predecir las frecuencias altas, la baja energía de la banda de voz, la cual fue abordada por el códec G722 usando la codificación en banda dividida. Aunque esta banda es importante para mantener una mejorada inteligibilidad y naturalidad, solo contiene una pequeña fracción de energía de voz y por lo tanto, su contribución en tasa de bits tiene que ser limitada apropiadamente.

Esto es principalmente motivado por el hecho de que en un códec CELP de banda completa, la excitación es típicamente elegida sobre la base de proveer una buena regeneración de baja frecuencia, ya que la mayoría de la energía reside en esa banda.

De ahí que la región de alta frecuencia y baja energía podría no ser tratada adecuadamente en los códec CELP de banda completa, a menos que sean tomadas medidas apropiadas, tales como escoger vastos libros de códigos, que a su vez requieren de sofisticadas medidas para mitigar su complejidad.

Codificación de Banda inferior

Black argumentó que era necesario incorporar un LTP (Long Term Predictor) de adaptación hacia adelante para contrarrestar el efecto del error de retroalimentación que es potencialmente perjudicial para el análisis LPC de adaptación hacia atrás. Un filtro de ponderación perceptual convencional fue empleado y los retrasos no enteros del LTP fueron incorporados. Específicamente, una resolución de $1/3 \cdot 1/8$ kHz $\approx 41.67\ \mu$s fue usado entre retrasos de LTP de $19\frac{1}{3}$ y $84\frac{2}{3}$, mientras que en el rango 85… 143 ningún sobre-muestreo fue utilizado.

El retraso de LTP fue representado por 8 bits.

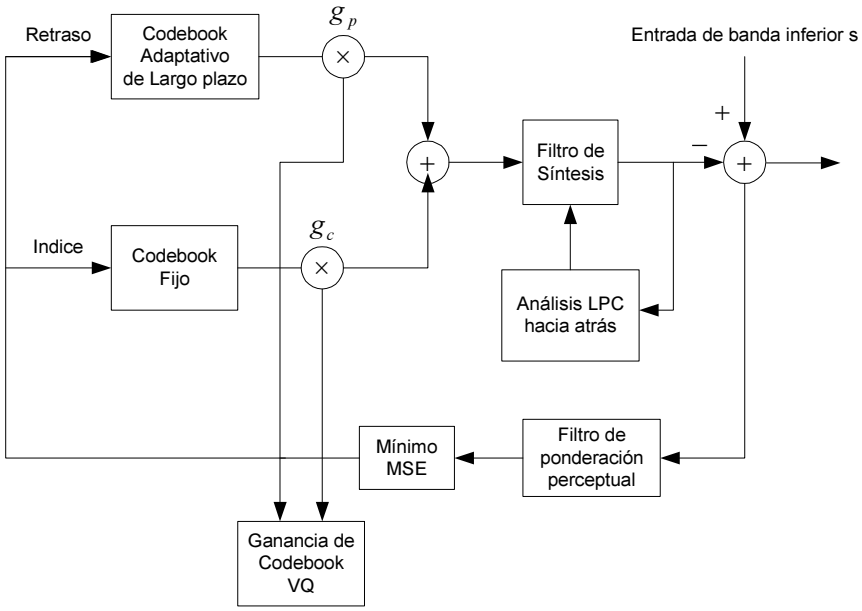


Figura 3.14: Diagrama de bloques del codificador de banda inferior de 16 kbps Códec CELP de sub-banda. Copyright © Blackl.

Codificación de Banda superior

La banda superior contiene típicamente menos estructura, como señal de ruido, la cual tiene un rango dinámico de variación lenta. Black propuso el uso de un predictor adaptativo hacia delante de sexto orden actualizado sobre un intervalo de 56 muestras, el cual está cuadruplicado en comparación a la banda inferior.

El análisis de LPC no intenta obtener una representación de forma de onda correspondiente a la señal de banda superior, meramente intenta modelar su envoltura espectral. Por tanto el decodificador regenera la señal de banda superior al excitar el filtro de síntesis LPC usando una escala aleatoria. La magnitud de este vector fue determinado por el codificador de filtro inverso de la señal de banda superior usando un filtro LPC de sexto orden y calculando la energía del residuo de más de 56 muestras.

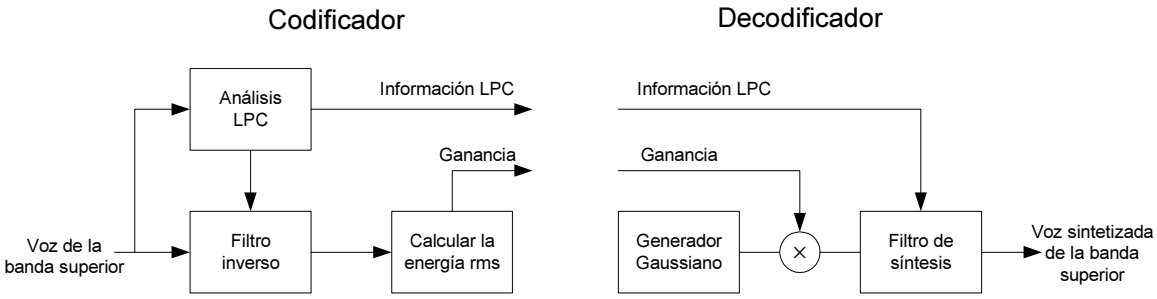


Figura 3.15: Diagrama de bloques del codificador/decodificador de banda superior de 16 kbps Códec CELP de sub-banda. Copyright © Black.

Los códecs CELP no suena lo suficientemente bueno[2], algunas de las mejoras para la calidad de la voz son:

- Los codebook estocásticos fueron instruidos para la voz real.
- Una medida de error más compleja se introdujo: Un filtro de ponderación de error, que se aplica a la señal de error antes de calcular el MSE: Así, el error de percepción se reduce al mínimo, según los resultados psicoacústicos.
- Otra idea es utilizar más de un codebook estocástico con el fin de minimizar aún más el error remanente.
- Las fracciones de los retrasos pueden mejorar la regeneración del tono. Un filtro de sobre-muestreo se utiliza para calcular las muestras intermedias de la señal de excitación pasado el filtro del tracto vocal, y el valor de retardo es por lo tanto, no más restringido a un múltiplo del periodo de muestreo.
- Mejores técnicas de cuantificación para los valores de los parámetros, utilizar la tasa de bits de manera más eficiente.

Esquema de distribución de bits

En la banda inferior predictiva hacia atrás ninguna información espectral de LPC es transmitida y de ahí que todos los bits son asignados para actualizar frecuentemente los codebook fijos y parámetros de los codebook adaptativos. La ganancia de los codebook fijos puede ser pronosticada por una técnica propuesta por Soheili, la cual pronostica la ganancia actual por el promedio de las tres ganancias cuantificadas precedentes.

Esta ganancia pronosticada Gp fue entonces usada para normalizar la ganancia del codebook fijo, por lo que se cuantifica la ganancia en vectores LTP en un proceso de circuito cerrado. Los seis coeficientes de LPC de banda superior fueron transformados a LSFs y un vector cuantificado con un total de 12 bits, mientras que el vector aleatorio de excitación fue cuantificado con 4 bits. El esquema de distribución de bits se muestra en la Tabla 3.5.

Parámetros	Bits	Actualizar (ms)	Tasa de bits(bps)
Banda inferior			
Retardo LTP	8	1.75	4571.4
Indice Codebook	8	1.75	4571.4
Ganancia VQ	8	1.75	4571.4
Banda superior			
LSFs	12	7	1714.4
Ganancia	4	7	571.4
Total			16000

Tabla 3.5: Asignación de bits de 16 kbps al Códec de banda ancha SB-CELP. Copyright © Black.

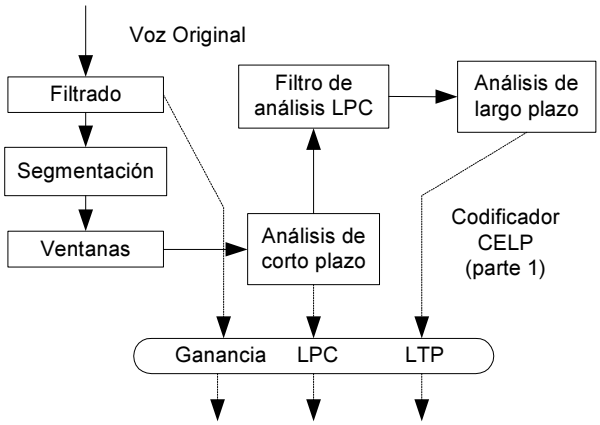


Figura 3.16: Diagrama de bloques del proceso de análisis de codificación CELP.

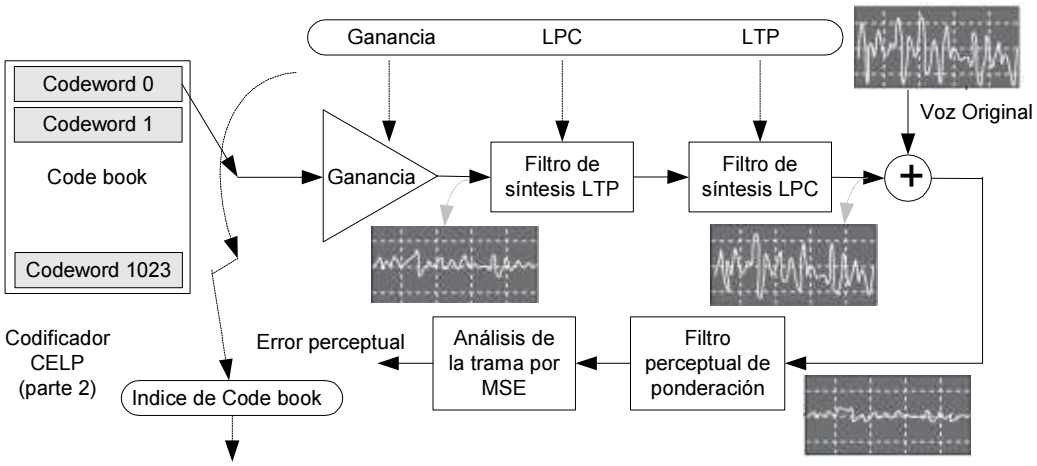


Figura 3.17: Diagrama de bloques del bucle de búsqueda de codebook CELP.

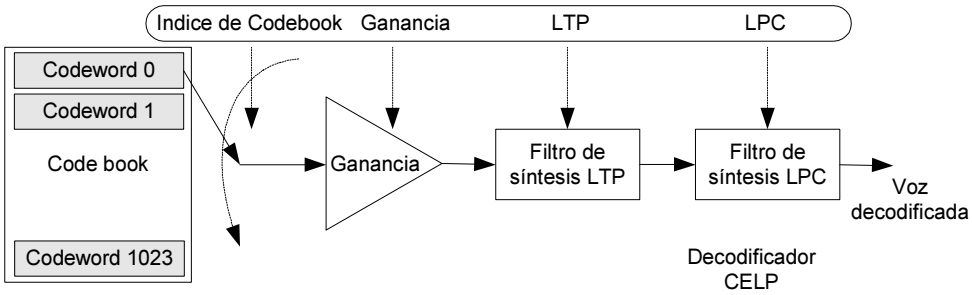


Figura 3.18: Diagrama de bloques del decodificador CELP.

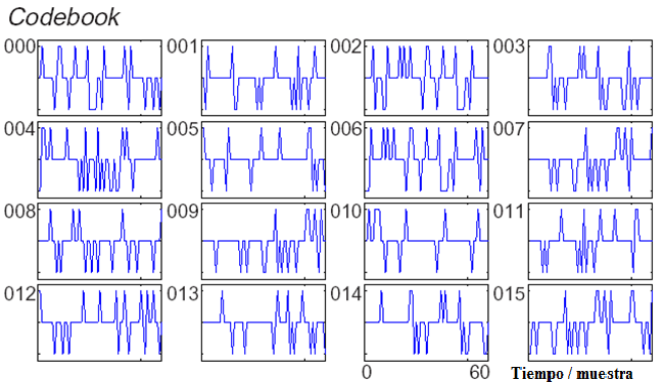


Figura 3.19: Ejemplo de codebook en CELP.

Simulación del Códec de CELP

CELP es una técnica hibrida de codificación, donde se combinan la codificación de forma de onda y la codificación por modelaje de la voz. En esta simulación podemos aprecia que la señal reconstruida a través del sistema CELP tiende a ser representado como ruido gausiano, a pesar de las diferencias de la señal original con respecto a la señal reconstruida esta última contiene la información de la señal original. Sin embargo, no es suficientemente buena la calidad de voz.

De esta manera CELP, además de enviar los parámetros que modelan el tracto vocal, la intensidad de la excitación, y la frecuencia de los formantes, también envía el código que permite obtener una aproximación al residuo de la señal de voz, que es parte importante para reconstruir con mayor fidelidad la señal.

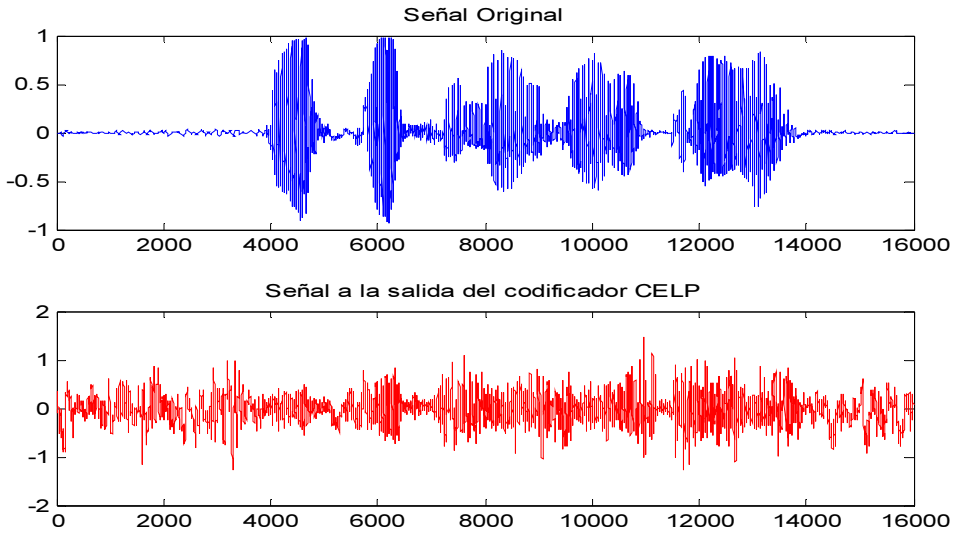


Figura 3.20: Codificación CELP utilizando Matlab.

3.6 Codificación ACELP (Predicción Lineal Extendida por el Código Algebraico - Algebraic Code Excited Linear Prediction) de banda ancha

El equipo de Sherbooke trabajo sobre un códec de predicción ACELP de banda ancha adaptativo hacia atrás a 32 kbps que utiliza el esquema convencional de CELP, excepto por el hecho de que se emplean dos generadores de excitación. Ambos generadores de excitación cuentan con un factor de ganancia separada. Por lo tanto, el vector de excitación conservaba una mayor flexibilidad en términos de las amplitudes de los pulsos y la búsqueda de los codebooks requiere de circuitos anidados para llegar a la excitación óptima.

El códec ACELP adaptativo hacia atrás utiliza filtros LPC de un orden de 32 y un predictor pitch de 3 pulsos, los cuales se actualizan cada 2 ms. La longitud de la trama de excitación es de16 muestras o de 1 ms, el alojamiento es de 4 pulsos por vector de excitación, cada uno codificado a 2 bits y las magnitudes de ± 1.

Por lo tanto, cada impulso requiere un total de 3 bits, y cada vector de alojamiento 4 pulsos, necesarios 12 bits de codificación. A las dos ganancias de los codebooks les fueron asignados a cada una 4 bits. Por lo tanto, los dos codebooks están permitiendo un total de 2· (12+4) = 32 bits, dando una tasa de bits de 32 kbps, mientras mantiene un retardo de 1 ms.

El esquema de asignación de los bits del códec se resume en la Tabla 3.6. El SEGSNR alcanzado por el códec a 32 kbps esta en el rango de 20–22 dB para señales de voz de banda ancha [2].

Parámetros	No. de bits/1 ms	Tasa de bits (kbps)
IndiceCodebook 1	4·(2+1)=12	12
IndiceCodebook 2	4·(2+1)=12	12
Ganancia Codebook 1	4	8
Ganancia Codebook 2	4	8
Total	32	32

Tabla 3.6: Códec ACELP de banda ancha adaptativo hacia atrás de 32 kbps.

Salami en una nueva contribución, utilizando una estructura ACELP de doble codebook, a través de una serie de técnicas innovadoras se propone a mitigar la creciente complejidad de los códecs debido a tasas de muestreo doble. Específicamente, el códec de banda ancha en tiempo real tiene la mitad del tiempo, es decir, 1/(16 kHz)= 62.5 μs para procesar las muestras.

En la tabla 3.7 se describe la propuesta del esquema de asignación de bits para una codificación ACELP adaptativa hacia adelante a 9.6 kbps. La longitud de la trama de actualización del LPC fue de 30 ms y un filtro de orden de 16 fue utilizado, cuantificando los LSFs con un total de 54 bits utilizando una capacidad de canal de 54/30 ms = 1.8 kbps. El retraso del pitch fue limita al rango de 40–295 y, en consecuencia 8 bits se utilizaron para su codificación. La ganancia LTP se cuantificó con 4 bits.

Parámetros	Actualización (ms)	No. de bits	Tasa de bits (kbps)
Filtro LPC	30	54	1.8
Retardo LTP	6	8	1.33
Ganancia LTP	6	4	0.67
IndiceCodebook 1	6	12	2
IndiceCodebook 2	6	13	2.17
Ganancia Codebook 1	6	6	1
Ganancia Codebook 2	6	3	0.5
Bits de relleno	30	4	0.13
Total	30	54 + (5·46) + 4 = 288	9.6

Tabla 3.7: Códec ACELP de banda ancha adaptativo hacia adelante.

Los dos codebooks en este esquema son diferentes. El primero contiene 4 pulsos convencionales entrelazados +1, -1,+1, -1 a 6 ms o 96 muestras del vector de excitación, donde el pulso m_i está definida como:

$$m_i^{(j)} = 3i + 12j, \quad i = 0,...,3, \; j = 0,...,7. \ldots. \quad (3.5)$$

Como puede deducirse de la ecuación anterior, hay ocho posibles posiciones para cada uno de los pulsos entrelazados y por lo tanto, un total de 12 bits por 6 ms = 2 kbps.

En la siguiente Tabla 3.8 se muestra un resumen de los códec antes mencionados.

Algoritmo de Codificación		Tasa de bits (kbps)	Retardo de codificación (ms)
G722	SB-ADPCM 0-4 kHz: 4-6 bit 4-8 kHz: 2bit	64	1.5
Quackenbush	Adaptativo 256-FFT	32	16
Laflamme	Banda Completa Adaptativo hacia adelante ACELP	16	15
Sanchez-Calle	Banda Completa Adaptativo hacia atrás ACELP	32	1
Salami	Banda Completa Adaptativo hacia adelante ACELP	9.6-14	30
Black	División de Bandas 0-4 kHz: 13.7 kbps Adaptativo hacia a atrás CELP 4-8 kHz: 2.3 kbps Vocoder	16	7

Tabla 3.8: Las funciones básicas de los códec de banda ancha.

3.7 Codificador por excitación de pulsos regulares y predicción de periodo largo (RPE-LTP Regular Pulse Excited - Long Term Predictor).

El codificador RPE-LTP (Regular Pulse Excitation - Long Term Prediction) es un codificador híbrido o paramétrico que utiliza un predictor lineal para eliminar la correlación entre tramas y que mejora la calidad de la señal con un predictor de retardo largo a través de una secuencia de excitación de patrones de pulsos regularmente espaciados. Este algoritmo de codificación se usa en las comunicaciones móviles GSM.

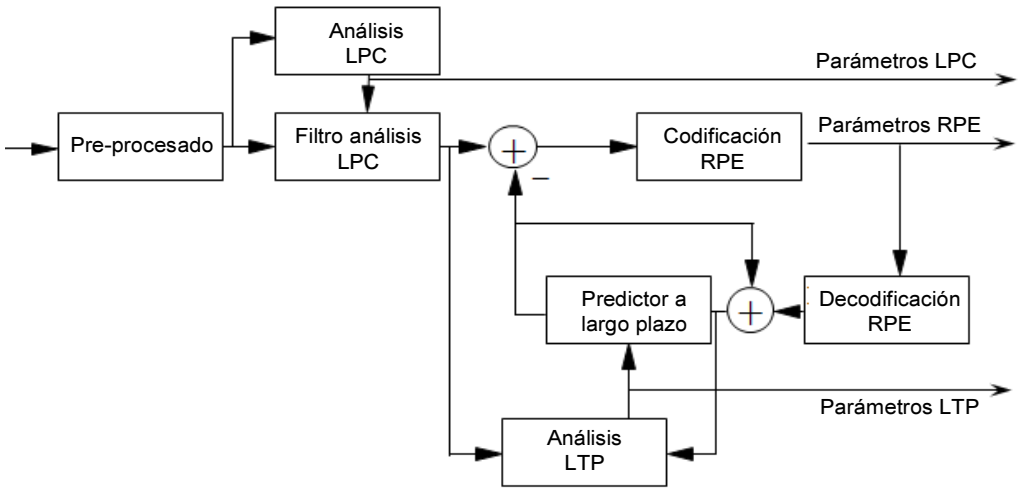


Figura 3.21: Diagrama de bloques del codificador RPE-LTP

En la Figura 3.21 se muestra el diagrama del codificador RPE-LTP. En éste se observan cuatro bloques funcionales. Estos son el pre-procesado, análisis LPC (Linear Prediction Coding), el filtro de predicción de retardo largo LTP (Long Term Predictor) y el cálculo RPE (Regular Pulse Exciting). A continuación se detalla el funcionamiento de estos bloques:

Pre-procesamiento; Cada 20 ms, se toman 160 valores de muestras del ADC (Analogic to Digital Conversion) y se almacenan en una memoria intermedia. El primer paso es el filtrado de la señal de entrada para eliminar el offset. A continuación se realiza otro filtrado que reduce el rango dinámico de la señal y eleva las zonas de los formantes a frecuencias altas. Este filtrado recibe el nombre de pre-énfasis de primer orden y se realiza porque el análisis LPC modela mal las amplitudes bajas de los formantes a altas frecuencias. La señal resultante se divide en bloques de 160 muestras que se almacenan en un buffer y se le aplica una ventana de Hamming para disminuir el efecto producido en el dominio de la frecuencia por la oscilación de Gibbs, causada por el truncamiento de la señal de voz fuera de la trama analizada.

Análisis LPC (Linear Prediction Coding); La señal de voz se divide en segmentos de 20 ms (160 muestras). A cada uno de estos segmentos se le aplica un análisis LPC de orden 8 (orden de predicción p=8). Se calculan nueve coeficientes de autocorrelación para obtener ocho coeficientes de reflexión a través del algoritmo de Durbin o utilizando la transformación inversa. Los coeficientes de reflexión se convierten en los LAR (Logarithmic Area Ratios), debido a sus propiedades de cuantización.

Filtro de predicción de retardo largo (LTP); Se evalúa cuatro veces por segmento, para cada 5 ms (40 muestras). Para cada subsegmento se calcula el factor de desplazamiento de relardo largo (pitch) y un factor de ganancia asociado. Como el parámetro de pitch puede tomar valores entre 40 y 120 se necesitan 7 bits para codificarlo. El factor de ganancia se codifica con 2 bits. El residuo del filtro LTP se calcula restando a la señal de residuo del filtro STP de una estimación de la misma que se calculó previamente a partir de la señal residuo STP reconstruida.

Codificación RPE; La señal residuo LTP se filtra con un filtro FIR, el propósito de este filtro perceptual es atenuar el espectro en frecuencia donde el error es perceptiblemente menos importante y amplificar aquellas zonas del espectro donde es perceptiblemente más importante. Con esto conseguimos una medida de error subjetiva significativa (propiedad de enmascaramiento del oído humano). La señal filtrada se submuestrea por un factor de 4, dando lugar a 4 secuencias entrelazadas de longitud 13. Se elige la secuencia de mayor energía como la representante de la excitación, secuencia RPE.

En el decodificador RPE-LTP se reciben los parámetros codificados y se procede a reconstruir la señal. En la Figura 3.22 se muestra el diagrama del decodificador RPE-LTP. En éste se tiene una estructura inversa a la del codificador. Los bloques funcionales son el decodificador RPE, el filtro de síntesis LTP, el filtro de síntesis STP y post-procesado.

La decodificación RPE; Consiste en decodificar y desnormalizar las muestras de la señal del residuo LTP. Posteriormente la frecuencia de muestreo se incrementa por un factor de 3 e insertando el resto de muestras como ceros.

En el filtro de síntesis LTP; Los parámetros del filtro LTP son cuantizados inversamente y se recupera un nuevo subsegmento del residuo STP.

En el filtro de síntesis STP; Los parámetros LAR (Logarithmic Area Ratios) se decodifican usando el cuantizador inverso, estos se interpolan en los bordes de la trama para prevenir cambios abruptos en las características de la envolvente de voz. Los parámetros interpolados se transforman en los coeficientes de reflexión.

El post-procesamiento; Se realiza usando el filtro inverso al filtro de pre-énfasis H(z), este reduce el rango dinámico de la señal y elevaba las zonas de los formantes a frecuencias altas. Con este filtro se obtiene la señal de voz decodificada.

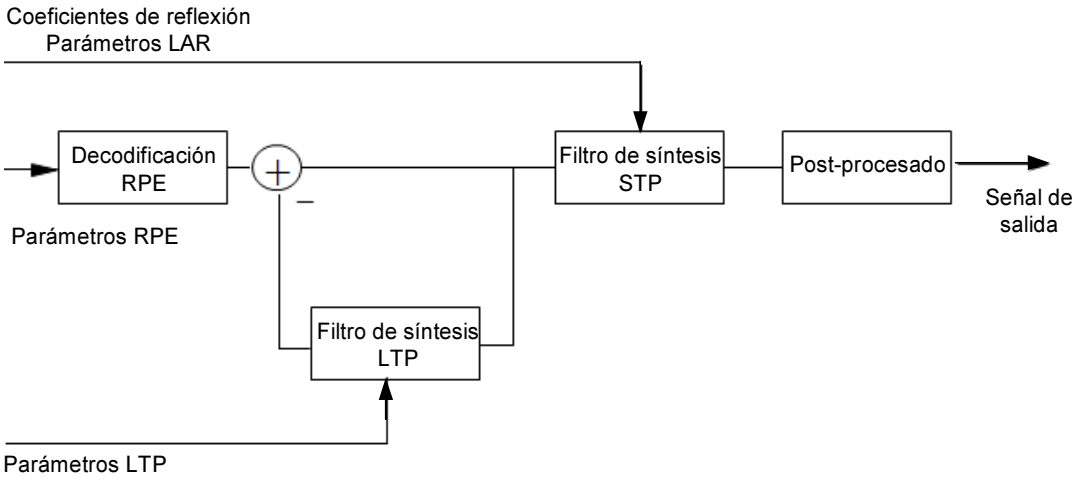


Figura 3.22: Diagrama de bloques del decodificador RPE-LTP

La codificación RPE (Regular Pulse Encoding) es un códec que transmite una serie de pulsos en vez de toda la señal, lo que reduce la complejidad y tiene mayor tasa de transmisión.

De cada subtrama de 40 muestras, se toman muestras alternadas y se construyen 3 tramas de 13 muestras.

Fuente	Descripción	Cada 5 ms (40 muestras)	Tasa de bits (kbps)
Filtro LPC	Parámetros LAR		36
Retardo LTP	Retardo (Delay)	7	28
	Ganancia	2	8
Señal de excitación	K máx	2	8
	V máx	6	24
Escala de Cuantización	13 muestras	3*13=39	156
Total			260 bits

Tabla 3.9: Bits transmitidos por el decodificador RELP-LTP

Simulación del Códec de RELP-LTP

El códec RELP-LPC nos proporciona un equilibrio entre el rendimiento y la complejidad.

En la Figura 3.23 se observa que la señal a la salida del codificador RELP-LTP tiene menor información, pero podemos apreciar que la forma es casi idéntica. La calidad de voz es buena con respecto a la señal origina, aún cuando la señal de salida del sistema RELP-LTP busca un rendimiento óptimo con respecto a la tasa de bits.

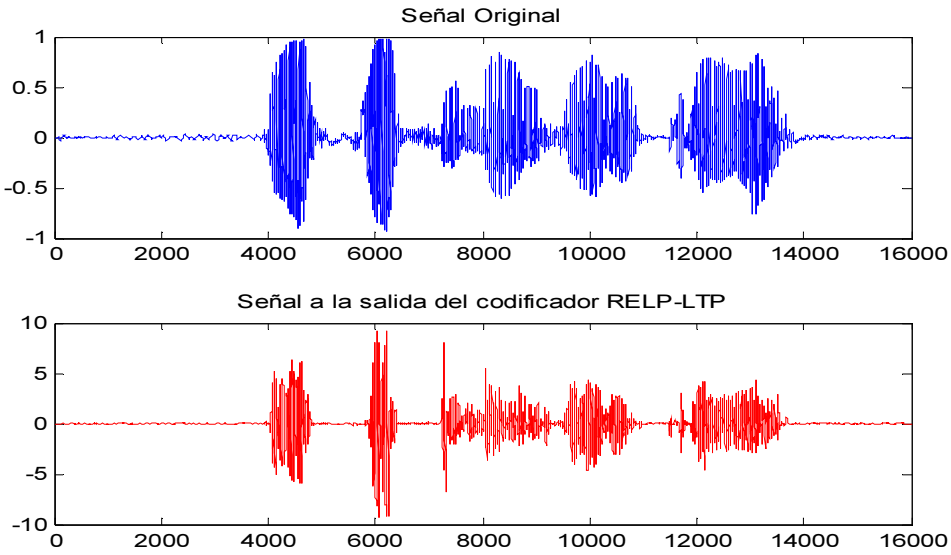


Figura 3.23: Codificación RELP-LTP utilizando Matlab

3.8 Códec G.722.1 de banda ancha

En años recientes la investigación de la codificación de voz se ha centrado en la banda ancha a 7 kHz. En este contexto se busca controlar la tasa de bits de forma adaptativa en un esfuerzo para encontrar la mejor relación en términos de la carga de las sub-portadoras, con la intención de incrementar la tasa de bits disponibles para mantener una alta tasa de codificación de voz y una alta calidad de la misma, mientras que también se mantiene una alta robustez frente a errores de transmisión.

El códec G.722 UIT a 64 kbps [15] es un estándar para la codificación de voz en banda ancha que se encuentra obsoleto, por lo que el códec de transformación Picture Tel (PTC), que se encuentra especificado en la recomendación UIT-TG.722.1[19] es el estándar actualmente para la codificación de audio en banda ancha. Este se basa en la llamada *Modulated Lapped Transform (MLT*), seguido por una etapa de cuantización psico-acústica y una codificación de Huffman para los coeficientes en el dominio de la frecuencia residual.

El G.722.1espera tramas de 320 muestras de audio PCM, obtenidas por muestreo de una señal de audio a una frecuencia de16 kHz con una resolución del cuantificador de 14, 15 o 16 bits. Además, las muestras de entrada se supone que contiene componentes de frecuencia de hasta 7 kHz. El estándar G.722.1 recomienda utilizar el códec a la salida sobre tasas de bits de 16, 24 o 32 kbps, generando longitudes de tramas de 320, 480 o 640 bits por 20 ms, respectivamente, para la cual el códec se ha optimizado.

El retardo total encontrado en una trama de audio, cuando pasa a través de este códec (que consta del codificador y descodificador) puede ser del orden de unos 60 ms, que es resultado de la técnica de superposición de la trama en el dominio de tiempo y el cálculo del retardo que es inherente al códec. Desde el empleo de la codificación de Huffman para codificar los coeficientes en el dominio de la frecuencia, la decodificación es muy sensible a los errores de bits. Por lo tanto, un solo bit de error puede representar que toda la trama de audio quede sin decodificar, por lo que la trama con el error simplemente se retransmite.

Descripción del Códec

Las etapas del procesamiento de señales incorporadas por la PTC, se muestran en el diagrama de bloques del codificador representado en la Figura 3.24. En la primera etapa del procesamiento, la señal de entrada PCM es mapeada desde el dominio del tiempo al dominio de la frecuencia, utilizando el MLT, un derivado de la DCT. Es bien sabido que la MLT puede emplearse en aplicaciones donde los efectos del bloque pueden causar graves distorsiones a la señal.

El MLT produce un bloque de 320 muestras en dominio de la frecuencia, con una resolución de 8000Hz/320= 25 Hz. Sólo los componentes de la señal con frecuencias de hasta 7 kHz están codificados, y corresponden a los coeficientes de la frecuencia con un índice inferior a 280, los demás coeficientes son descartados. Los coeficientes de MLT restantes están más agrupados dentro de las 14 regiones de igual anchura, cada una representa un rango de frecuencia de 500 Hz, y un alojamiento de 280/14=20 coeficientes. Para cada región de frecuencia, el RMS de la potencia da una estimación de la envolvente espectral.

Mediante el cálculo de una categorización inicial, se establece una determinada cuantificación y los parámetros de codificación son asignados a cada región. Como se representa en la Figura 3.24, se tiene un total de 16 categorizaciones tentativas y las asignaciones de bits son calculadas, de los cuales finalmente sólo el que hace uso de ellos se le asignan. Después de la asignación del mejor bit, los coeficientes del MLT son cuantificados y junto al código de Huffman los parámetros de las categorías son asociadas. Durante la última etapa del cálculo de los datos de salida, las señales descritas se multiplexan dentro de una trama de datos. La asignación de bits de una trama de datos típica a la salida del codificador PTC se ilustra en la Figura 3.25 para el caso de 320 bits por trama, es decir, 16 kbps.

Como se muestra en la Figura 3.24, el multiplexor (MUX) ordena los bits del código RMS, los bits de control de tasa y, finalmente los bits de código MLT en un flujo. La estructura de la trama se muestra en la Figura 3.25, junto con el número típico de bits necesarios para codificar la envolvente espectral y los coeficientes de la transformada. En cada trama, los primeros 5 bits son ocupados por el valor de índice (0) del RMS, seguido por los códigos de Huffman de los códigos diferenciales del RMS con índice 1, ..., 13 dentro del orden de la frecuencia espectral. Los siguientes 4 bits de cada trama son ocupados por los llamados bits de control de tasa. Entonces los índices del vector de código MLT son transmitidos, comenzando con la región de frecuencia 0. Inmediatamente después un código con índice de vector de longitud variable y los correspondientes bits del signo del coeficiente MLT son transmitidos, en orden de la frecuencia espectral.

Las etapas del procesamiento de las señales que constituyen el decodificador G.722.1 son esencialmente las operaciones inversas del codificador que se muestra en la Figura 3.24. La decodificación de una trama comienza con la reconstrucción de la envolvente espectral. A continuación, los 4 bits de control de tasa son decodificados, con el fin de determinar cuál de las 16 categorizaciones posibles han sido usadas para codificarlos coeficientes del MLT.

De la misma manera, como se generan las 16 categorizaciones en el codificador, estás son ahora también generada en el decodificador. Finalmente, la categorización utilizada en el codificador se emplea también en el decodificador. Después de desnormalizar por multiplicación todos los coeficientes de una región de frecuencias por sus valores en RMS, los coeficientes del MLT son reorganizados en bloques de 320 coeficientes, donde los primeros 40 coeficientes se establecen en cero, ya que pertenecen a las frecuencias por encima de los 7 kHz.

Entonces, el inverso del MLT (IMLT) se aplica a los coeficientes, generando 320 muestras en dominio de tiempo a la salida. Tanto el MLT y el IMLT se pueden descomponer a través de la transformada coseno discreta (DCT) y la inversa de DCT (IDCT), seguido por una ventana, la superposición y la operación añadir.

Debido a la codificación de Huffman la cual se aplica a los valores de la envolvente espectral, así como a los coeficientes del MLT, la información transportada por estas palabras de código (codewords) es extremadamente sensible a los errores de bits. Si el decodificador del canal no es capaz de corregir todos los errores de transmisión, el comportamiento de la recomendación del decodificador PTC es repetir los coeficientes del MLT de la trama anterior, en el caso de una trama única errónea, o para establecer los coeficientes del MLT a cero, lo que corresponde al silenciamiento de la señal de salida, siempre que la trama anterior también haya sido contaminada por los errores del canal.

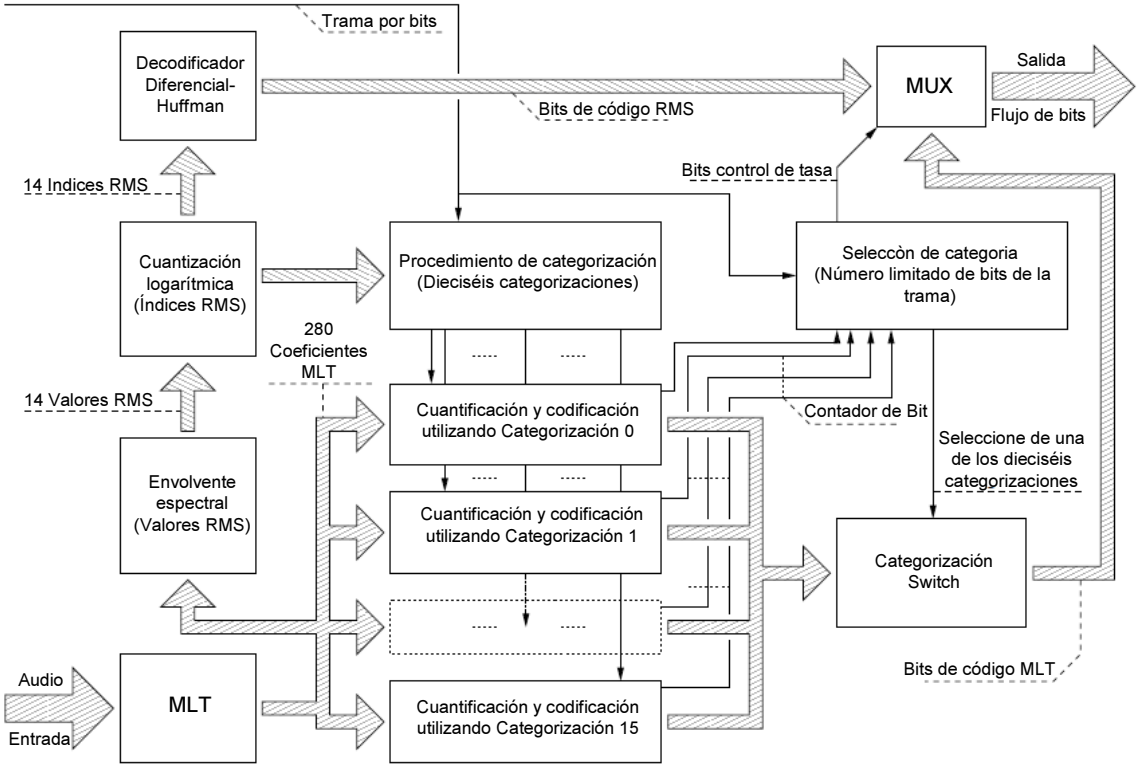


Figura 3.24: Diagrama de bloques del codificador G.722.1 Picture Tel.

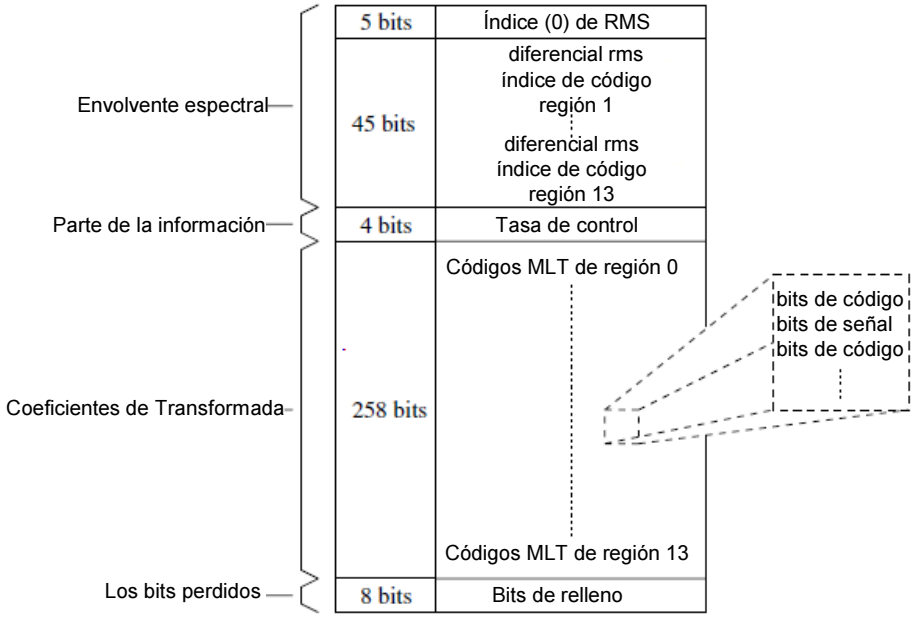


Figura 3.25: Estructura de una trama de código típica a una tasa de bits de 16 kbps (320 bits por trama de 20 ms).

3.9 Codificación AMR (Multi-Tasa Adaptativo - Adaptative Multi-Rate)

La codificación AMR es la compresión de datos de audio optimizados para la codificación de voz, adoptada en octubre de 1998 como el códec de voz estándar de 3GPP [6] (3d Generation Partnership Project) y ahora es ampliamente utilizado en GSM (Sistema Global para Comunicaciones Móviles). En su forma evolucionada, GSM es también conocida como UMTS (Universal Mobile Telecommunications System). El códec AMR (Adaptative Multi-Rate) es utilizado típicamente en redes celulares GSM. Hace uso de tecnologías DTX (Discontinous Transmition), VAD (Voice Activity Detection) para detección de actividad vocal y CNG (Confort Noise Generation). Provee una variedad de opciones en cuanto al ancho de banda que utiliza. Puede trabajar a velocidades de transmisión que varían entre 4.75 y 12.2 kbps.

Se basa en el modelo CELP, operando con ventanas de audio de 20 ms correspondientes a una cantidad de 160 muestras. Cada ventana es a su vez dividida en 4 sub-ventanas, de 5 ms (40 muestras) cada una. Para cada ventana se extraen los parámetros LP del modelo CELP (los coeficientes de los filtros LP), y por cada sub-ventana se obtienen los índices de los "codebooks" fijos y adaptivos, y las ganancias. Estos parámetros se cuantizan y se transmiten dentro de una trama con un formato pre-establecido en la recomendación. El códec AMR de banda ancha, desarrollado conjuntamente por Nokia y Voice Age, se ha normalizado para sistemas GSM y 3G de WCDMA en el año 2001[8].

3.10 Codificación AMR-WB (Adaptative Multi-RateWideband)

El códec de voz AMR-WB (Adaptive Multi-Rate Wideband) es capaz de soportar 9 posibles tasas de bits de codificación que varía desde 6.6 hasta 23.85 kbps y se ha convertido en un estándar de la 3GPP y la UIT-T en la recomendación G722.2 [17], la cual proporciona una calidad de voz superior en comparación con los códecs de voz de banda ancha convencionales de telefonía.

Cada trama de AMR-WB representa 20 ms de voz, produciendo 317 bits a una tasa de 15.85 Kbps, además de 23 bits de información de encabezado por trama. Los parámetros del códec en cada trama incluyen los llamados Inmitancia Espectral de Pares (ISP), el retardo del codebook de adaptación (retardo de pitch), el índice de excitación del codebook algebraico y el conjunto de vectores de cuantización de ganancia de pitch, así como las ganancias de los codebook algebraicos.

3.11 Codificación de Audio AMR-WB+

El códec AMR-WB+ es una técnica de codificación capaz de conseguir una calidad alta de audio a tasas excepcionalmente bajas. Este códec fue recientemente seleccionado por 3GPP y DVB para aplicaciones audiovisuales y de audio a bajas tasas de bits en las redes móviles. Los recientes avances, tanto en la compresión de fuente, así como en las redes inalámbricas y en las tecnologías de dispositivos móviles han permitido la introducción de servicios innovadores de multimedia enviados a dispositivos móviles.

La prestación de servicios multimedia móviles de alta calidad impone requisitos más exigentes, tanto en el diseño de un códec de fuente estereofónico, así como la red inalámbrica, cuando se pretende para una calidad de audio de percepción alta. El 3GPP ha definido una gama de servicios multimedia móviles de alta calidad, tanto para los sistemas de radio por paquetes genéricos (GPRS) como para las redes 3G, los cuales se benefician del aumento del rendimiento y reducen potencialmente las tasas de error en redes inalámbricas avanzadas.

Los estándares 3GPP correspondientes especifican los siguientes servicios avanzados:

- Servicio de mensajería multimedia (MMS) [3GPP TS 22.140]. Este servicio puede ser utilizado entre las terminales móviles, así como para la descarga de información de un servidor de contenidos a una terminal móvil y viceversa. A medida que las terminales multimedia son cada vez más sofisticadas impone retos de la interoperabilidad de las distintas redes.

- Servicio de streaming por paquetes conmutados (PSS) [3GPP TS 22.233]. Este servicio proporciona una trama punto a punto de los servicios de streaming multimedia con la ayuda del protocolo de transporte en tiempo real (RTP) para el transporte de audio interactivo en tiempo real y vídeo.

- Servicio de multimedia Broadcast / Multicast (MBMS) [3GPP TS 22.246]. Este servicio constituyen un servicio punto a multipuntos, donde normalmente el streaming se basa en el RTP y la descarga es empleada.

- Los subsistemas multimedia IP (IMS) de mensajería [3GPP TS 22.340].
- Servicio de presencia [3GPP TS 22.141].

El desafío que enfrenta el diseño del códec AMR-WB+, es que estos servicios incluyen la transmisión de música, voz y la mezcla de ambos, mientras que la tasas de bits disponibles en el contexto de las redes GPRS puede ser tan baja como 10 kbps. Por lo tanto, el organismo 3GPP ha estandarizado los códecs de audio para soportar la interoperabilidad de las redes GPRS y 3G, así como la de las terminales móviles. Sin embargo, cuando se libero el 3GPP versión 5 se había especificado como un códec sin audio, lo que era fácilmente capaz de satisfacer estos requisitos, porque los códecs de voz tienden a explotar las propiedades estadísticas de las señales de voz, mientras que las señales de audio por lo general exhiben diferentes propiedades estadísticas.

El códec AMR-WB visto anteriormente ofrece una calidad alta de 7 kHz de ancho de banda para señales de voz a una tasas de bits tan baja como 12.65 kbps, pero como era de esperar, no se desempeña bien cuando se codifica las señales de audio. Por lo tanto, se inició un proceso en el 3GPP para probar y seleccionar los códecs de audio con la versión 6, con el fin de emplearla en los servicios multimedia [6]. El códec AMR-WB para bajas tasas presenta un buen desempeño para audio y proporciona un mejor rendimiento, tanto para la voz como para contenidos mixtos. El códec AMR-WB+, posteriormente fue seleccionado para aplicaciones DVB como un códec opcional. Por tanto, el códec AMR-WB+ se incluye en la norma DVB para la entrega de señales de audio a través de paquetes RTP sobre redes IP.

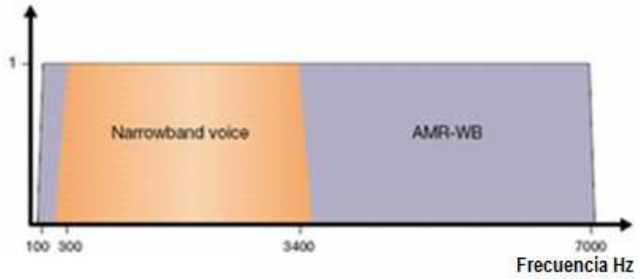


Figura 3.26: Ancho de banda para la voz en AMR-WB.

Códec AMR en UMTS

Originalmente desarrollado por la ETSI (Instituto Europeo de Normas de Telecomunicaciones) para su uso en las redes GSM, el códec de voz Adaptive Multi-Rate (AMR), fue aprobado para redes UMTS (Sistema de telefonía móvil celular) [norma IMT-2000 de la UIT]. Un códec de voz AMR adapta el nivel de protección del error sobre las condiciones de tráfico para que siempre se seleccione el canal óptimo y el modo del códec que ofrecer la mejor combinación de calidad de la voz y la capacidad del sistema. La estructura de la trama AMR, se divide en tres partes: La cabecera, información auxiliar y el núcleo de la trama.

La parte de la cabecera incluye el tipo de trama y los campos indicadores de calidad de trama. El tipo de trama puede indicar el uso de uno de los ocho modos de códec AMR para esa trama, una trama de ruido, o una trama vacía. El indicador de calidad de la trama indica si la trama es buena o mala.

La parte de la información auxiliar incluye el indicador del modo, las solicitudes del modo y los campos CRC del códec para los propósitos de detección de errores.

El núcleo de la trama se compone de los bits de los parámetros de la voz o, en caso de una trama de ruido de los parámetros del ruido de confort. Se utiliza para transportar los bits codificados y estos se dividen en clases. La división de clases es sólo informativa y proporciona información de apoyo para el mapeo de este formato genérico en formatos específicos.

La clase A contiene los bits más sensibles a errores y cualquier error en estos bits típicamente resulta en una trama de voz dañada, que no debe ser decodificado sin aplicar ocultamiento de error correspondiente. Esta clase está protegida por el códec CRC dentro de la información auxiliar del AMR. Las clases B y C

contienen bits donde el incremento de la tasas de error reducir gradualmente la calidad de la voz, pero es posible la decodificación de una trama de voz errónea.

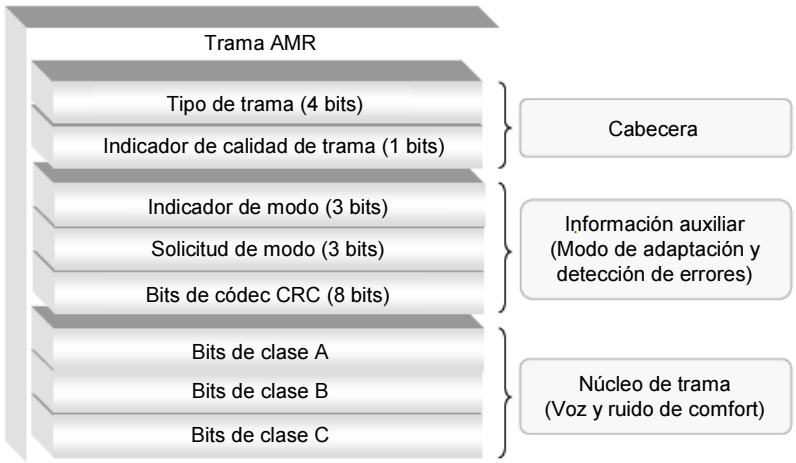


Figura 3.27: Estructura de la trama AMR.

Requisitos de audio en aplicaciones multimedia móviles

Veamos ahora un resumen de las especificaciones generales que deben cumplir el códec de audio de acuerdo con el organismo de estandarización 3GPP, cuando se comunica a través de redes GPRS o 3G. Una compensación atractiva tiene que encontrarse entre la tasa de bits requerida y la calidad del audio.

El decodificador tiene que ser capaz de recuperar la señal de audio en una calidad inobjetable, incluso en presencia de eventos de pérdida de paquetes en la transmisión por el canal inalámbrico.

En la Tabla 3.10 se resumen la tasas de bits que soportan los diversos portadores GPRS y 3G, donde en la tercera columna se indica la tasa de bits total del canal, mientras que la cuarta columna se representa la tasa de bits disponible para el códec de audio sin tener en cuenta la sobrecarga adicional impuesta por los protocolos de transmisión, tales como el protocolo de Internet.

		Contenido Audio		Contenido Audiovisual	
Servicio	Acceso Radio	BW del canal o tamaño del mensaje	Audio (tasa neta) o longitud del contenido	Ancho de banda del canal o tamaño del mensaje	Audio (tasa neta) o longitud del contenido
PSS	GPRS	36 kbps	24 kbps	36 kbps	~10 kbps
	UMTS	64 kbps	48 kbps	64 kbps	~14 kbps
				(128 kbps)	(~24 kbps)
MBMS Streaming	GPRS	36 kbps	<24 kbps	36 kbps	~10 kbps
	UMTS	64 kbps	<48 kbps	64 kbps	12 -16 kbps
				(128 kbps)	(~24 kbps)
MMS	GPRS o UMTS	100 KB (audio)	0.5 min a 24 kbps o 1 min a 14 kbps	75 KB (video) + 25 KB (audio)	20 s a 10 kbps
MBMS Descarga	GPRS o UMTS	300 KB (audio)	1.5 min a 24 kbps o 3 min a 14 kbps	225 KB (video) + 75 KB (audio)	60 s a 10 kbps

Tabla 3.10: Resumen de tasas de bits.

Descripción general del códec AMR-WB+

El códec de audio AMR-WB+ utiliza una técnica de codificación híbrida para ofrecer un alto nivel de calidad tanto para señales de voz como para las señales de música en tasas de bit que van desde 6-48 kbps. Está técnica de codificación híbrida incluye la técnica de codificación ACELP para la manipulación de las señales de voz, y está basada en la transformada en el dominio de la frecuencia para codificar eficientemente las señales de audio y de música.

El codificador AMR-WB+ decide primero la elección del modo más adecuada de codificación, es decir, entre CELP o codificación en transformada en base a cada trama. Esto permite que el códec proporcione una alta calidad de la señal reconstruida para una amplia gama de sonidos a una baja tasa de bits. Además, AMR-WB+ integra un modelo estéreo paramétrico con el fin de mejorar la percepción del usuario final de una reproducción de sonido de alta fidelidad a tasas de bits muy bajas.

El codificador AMR-WB+ es capaz de comprimir tanto las señales mono como estéreo. Mientras que el decodificador puede reproducir la señale original (mono o estéreo) del flujo de bits recibida, pero también es capaz de dar salida a una señal mono basado en el flujo de bits estéreo recibida. En el modo mono, las tasas de bits soportadas van desde 6-36 kbps, mientras que la tasa de bits en estéreo puede abarcar desde 8-48 kbps.

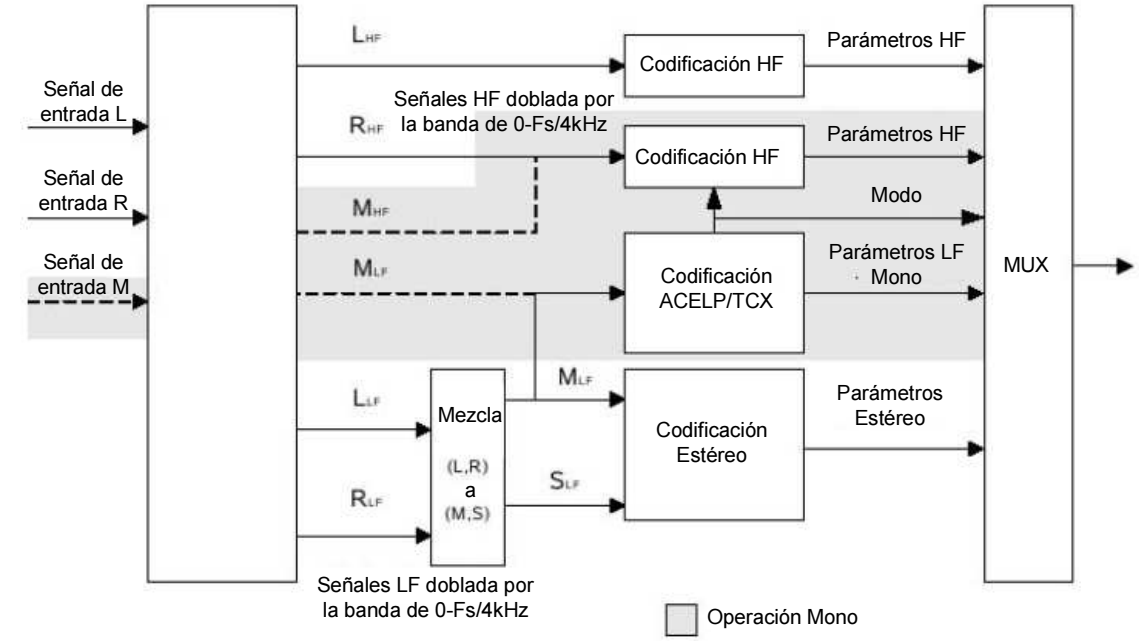
El codificador AMR-WB+ funciona a una frecuencia de muestreo de 25.6 kHz. La señal de entrada de audio es primero muestreada a una frecuencia de muestreo interna, y luego se dividen en dos bandas de igual ancho, que están críticamente submuestreadas a 12.8 kHz. Esto permite la integración eficiente del codificador de voz AMR-WB original, que opera a una frecuencia de muestreo de 12.8 kHz. El codificador procesa la señal de entrada en bloques de 2048 muestras, independientemente de la frecuencia de muestreo interna. Después de dividir la banda y del submuestreo crítico, la banda inferior y la banda superior son procesadas en bloques de 1024 muestras.

En el códec AMR-WB+ el bloque de 1024 muestras se le conoce como una super-trama. La super-trama en la banda inferior abarca desde 0 hasta 6.4 kHz, es codificada utilizando la codificación ACELP híbrida mencionada anteriormente y el modelo por transformada *TCX* (transform Coded Excitation). La super-trama en la banda superior barca desde 6.4 hasta 12.8 kHz se codifica utilizando 64 bits por 1024 muestras, empleando un método de Extensión de Ancho de Banda (BWE).

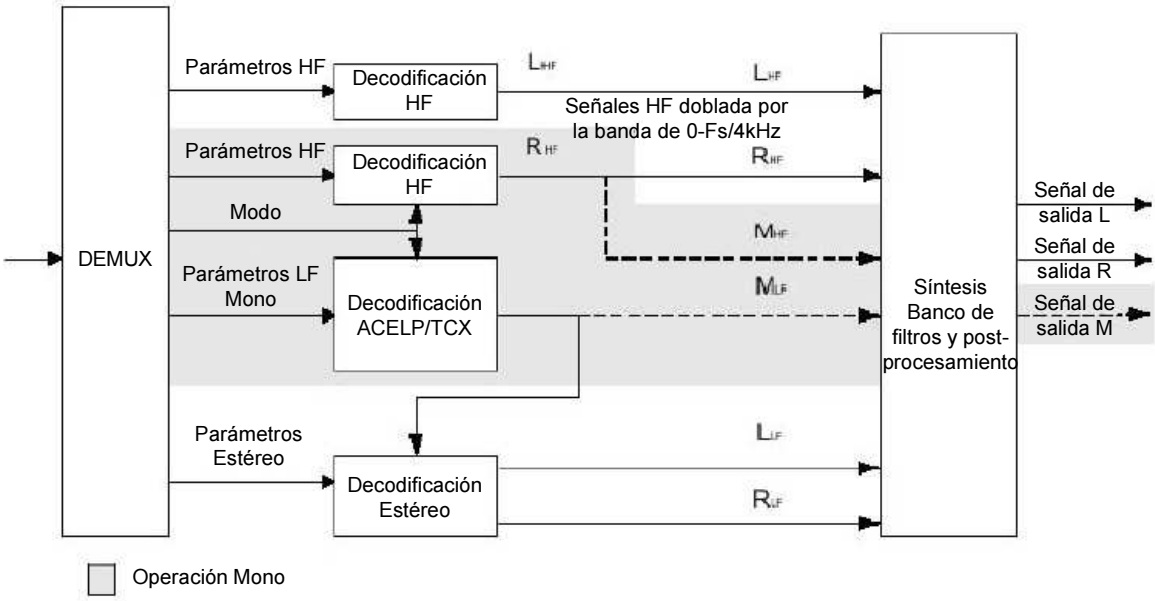
En la Figura 3.28 se muestra el diagramas del codificador y del decodificador AMR-WB+. Cuando se procesa una entrada mono, los esquemas de codificador y decodificador son limitados a la zona sombreada. Por el contrario, dentro del modo de operación estéreo el esquema completo se activa, es decir, las líneas punteadas se quitan. Las señales izquierda, derecha y centro están etiquetados como L, R y M, respectivamente, donde la señal del centro es creada a partir de los canales izquierdo y derecho de la señal estereofónica. Las señales de banda de baja frecuencia están denotadas como LF y señales de banda de alta frecuencia están representadas como HF. En la Figura 3.28 (a), el pre-procesamiento representa los filtros de acondicionamiento y la operación de muestreo, generando la señal requerida por el uso de la frecuencia de muestreo interna.

El análisis del banco de filtros divide la señal de entrada en señales de banda alta y baja. La operación de la mezcla descendente produce un canal lateral de los canales izquierdo y derecho de una señal de entrada estéreo. La banda baja de la señal de media (o mono) se codifica utilizando el modelo CELP/TCX, mientras que la banda alta de la señal de media (o mono) se codifica mediante la operación BWE.

En el modo de codificación por transformada, los coeficientes espectrales se cuantifican mediante una técnica conocida como vector de cuantización VQ. Nótese que en el modo por transformada (TCX) la trama se extiende con un segmento de "pre-análisis", que es necesaria para la superposición de las ventanas de transformada. El empleo de los diferentes modos de codificación y de las longitudes de trama ayuda al códec en el logro de una mejor codificación de los diferentes tipos de señales de entrada, tales como la voz y el audio.



(a)



(b)

Figura 3.28: (a) Diagrama de bloques del codificador AMR-WB+, (b) Diagrama de bloques del decodificador AMR-WB+.

Códecs de banda ancha

Códec	G.722	G.722.1	AMR-WB/G.722.2	G.711.1	G.729.1
Tasa de bits (kbps)	48, 56, 64	16, 24, 32	6.6, 8.85, 12.65, 14.25, 15.85, 18.25, 19.85, 23.05 y 23.85	64, 80, 96	8 a 32
Tipo	Sub-banda ADPCM	Codificación por Transformada	Predicción Lineal con Excitación por Código Algebraico (ACELP®)	Banda Ancha G.711	Banda Ancha G.729
Retardo (ms):					
Tamaño de Trama	0.125	20	20	11.875	<49
Cabecera	1.5	20	5		
Comentarios	Inicialmente diseñado para audio y videoconferencia, actualmente utilizado para la telefonía de calidad en VoIP.	Desempeño de la voz pobre en algunas condiciones de operación. Desempeño bueno para la música y videoconferencia	Estándar en común con 3GPP, gran inmunidad a los ruidos de fondo en ambientes adversos (por ejemplo celulares). Rendimiento de la voz a tasas de 12.65 kbps y superior.	Amplia el ancho de banda del codec G.711	Amplia el ancho de banda del codec G.729, y es compatible hacia atrás con este codec. Uso de audio de alta calidad.

Tabla 3.11: Códecs de Banda Ancha.

Capítulo 4

Evaluación

La compresión de la voz hoy en día se hace imprescindible con la incesante necesidad surgida a partir de una mayor demanda establecida por usuarios sobre las redes inalámbricas y redes celulares, no obstante la tecnología ha ido evolucionando rápidamente para ajustarse a tales necesidades lo que ha permitido el uso de banda ancha y quizás la pronta convergencia de ambas redes por medio de la tecnologías LTE.

El propósito de este cuarto capítulo es mostrar los requerimientos usados para realizar la comparativa entre distintos codificadores de voz, donde se adecuaron algunos archivos para llevar a cabo las evaluaciones de los diferentes codificadores a tasas de muestreo distintas y los resultados obtenidos fueron evaluados bajo criterios objetivos y subjetivos.

CAPÍTULO 4. EVALUACIÓN

4.1 Las pruebas

Las pruebas consisten en realizar un análisis de algunos codificadores de voz a través de evaluaciones objetivas y subjetivas, con el propósito de comparar la calidad de la voz reconstruida, donde el rendimiento de las redes inalámbricas depende de la alta eficiencia en la compresión de los códecs.

Codificadores que serán evaluados:

- ADPCM - Modulación por Pulsos Codificados Diferenciales Adaptativo
- G722
- LPC - Codificación Predictiva Lineal
- CELP - Predicción Lineal Extendida por el Código
- RELP- GSM - Codificador por excitación de pulsos regulares y predicción de periodo largo

Los criterios de evaluación los dividimos en dos partes; La primera parte consiste en un análisis objetivo donde se mide el comportamiento y funcionamiento del codificador/decodificador, y en una segunda parte se realizará un análisis subjetivo conforme a estadísticas de opinión.

Análisis objetivo:

- Relación señal-ruido (SNR)
- Tasas de bits para cada canal de voz
- Espectrogramas
- Ganancia

Análisis subjetivo:

- Estimador MOS (Mean Opinion Score)

4.2 Señal de entrada

A través del software de Matlab se adquirió la fuente de la señal de entrada de la voz en un formato .wav, grabado en PCM con una frecuencia de muestreo de 8 kHz a 8 bits. La señal de voz fue generada a través de un micrófono omnidireccional.

Las señales de entrada que se utilizaron para estas pruebas fueron dos; Una primera grabación con voz de hombre y una segunda con voz de mujer, ambas contenían el mensaje “Esta es una prueba de voz”.

```
% Se genera una señal de entrada con matlab
clear all
Fs=8000; % Frecuencia de muestreo
x=wavrecord(2*Fs,Fs,1);%,'int16'); % grabando voz y muestreando
wavwrite(x,Fs,'Prueba_Voz_H.wav'); %guardamos lo grabado como el archivo .wav
[s Fs]=wavread('Prueba_Voz_H.wav');
wavplay(s,Fs) %Reproducirlo
plot(s)
```

Características generales de la señal de entrada

Tiempo total es 2 s
Frecuencia de muestreo es 16000 Hz para cumplir con el criterio de Nyquist.
Tiempo es 62.5 μs por muestra.

Una trama se obtuvo con 100 muestras, por lo tanto tendremos 160 tramas.
2000 ms/20 ms = 100 muestras.

4.3 Especificaciones del hardware

Los dispositivos requeridos para llevar a cabo mis pruebas para la adquisición, reproducción y procesamiento de la voz son los siguientes.

La laptop:

Figura 4.1: Características físicas y lógicas de la laptop.

Micrófono:

- Omnidireccional (recepción de sonido en cualquier dirección)
- Respuesta en frecuencia: 50 Hz -16 kHz
- Sensibilidad: -60 dB + 3 dB
- Cable de 2,5 m con conector macho (plug) de 3,5 mm
- Modelo: MIC-505

Figura 4.2: Micrófono Modelo MIC-505

4.4 Especificaciones del software

El software utilizado para la programación de los distintos codificadores es Matlab® (Matrix Laboratory) un producto de la empresa The Mathworks Inc., empresa fundada en 1984.

Características:

MATLAB - The Language of Technical Computing
Versión 7.8.0.347 – Release: R2009a; version para PC: 32-bit (win32)
Compatible con Windows XP, Windows Vista y Windows 7

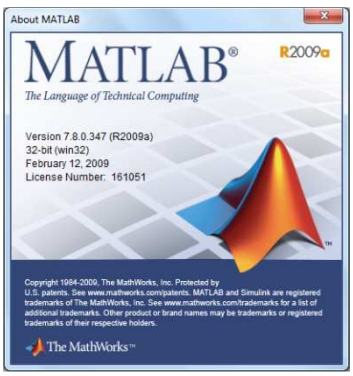


Figura 4.3: Software MATLAB

4.5 Simulación de Codificadores en Matlab

Las simulaciones de codificación se programaron sobre Matlab, habiendo utiliza algunos bloques pre-programados para dichas operaciones de codificación [23], los cuales nos permitieron disminuir los tiempos de implementación y, llevar a cabo las evaluaciones bajo los criterios objetivos y subjetivos.

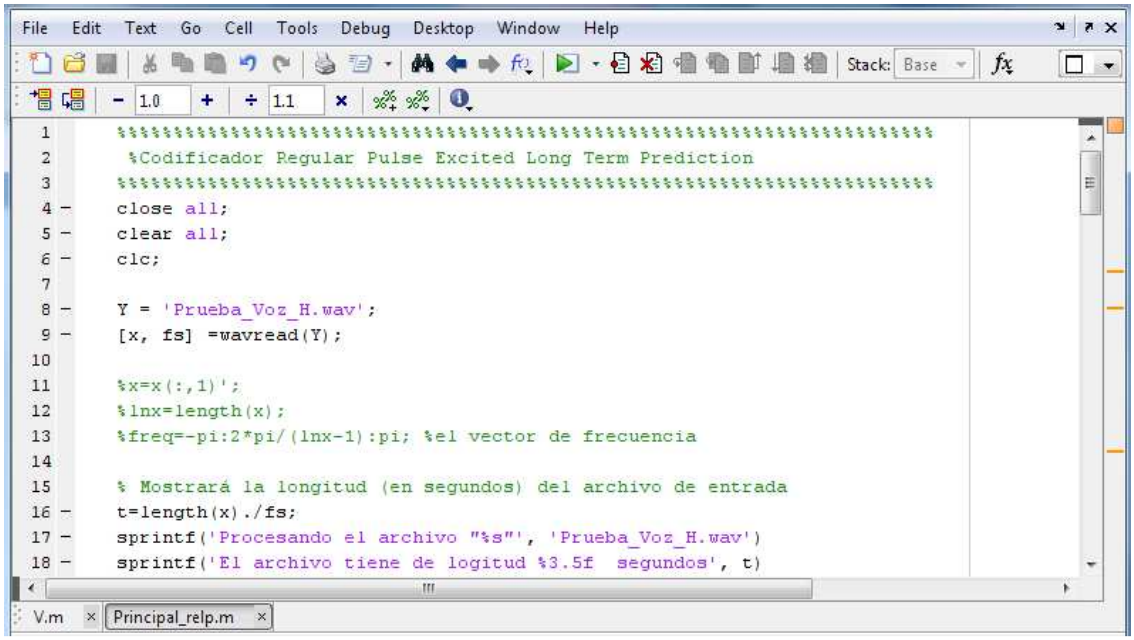


Figura 4.4: Codificadores de Voz programados en Matlab

4.6 Evaluación objetiva de la calidad de voz

Relación señal-ruido (SNR)

La relación señal-ruido (SNR) es el factor objetivo más comúnmente empleado para evaluar el desempeño de los códecs, debido a que nos permite calcular la energía en la señal original con relación a la energía del ruido, donde el ruido es el error entre la señal original y la señal reconstruida.

$$SNR = \frac{\sigma_{in}^2}{\sigma_e^2} = \frac{\sum_n s_{in}^2(n)}{\sum_n [s_{out}(n) - s_{in}(n)]^2}, \ldots (1.10)$$

donde, Sin(n) y Sout(n) son la secuencia de entrada y salida, respectivamente de las muestras de la voz.

En nuestro caso, se utilizó la ecuación (1.12), la relación señal-ruido por segmentos ($SEGSNR^{dB}$):

$$SEG - SNR^{dB} = \frac{1}{M}\sum_{m=1}^{M} 10\log_{10}\frac{\sum_{n=1}^{N} s_{in}^2(n)}{\sum_{n=1}^{N}[s_{out}(n) - s_{in}(n)]^2}, \ldots . (1.12)$$

donde, N es el número de muestras de la voz dentro de un segmento de 20 ms, es decir, 160 muestras a una velocidad de muestreo de 8 kHz, mientras que M es el número de segmentos.

De las pruebas hechas en Matlab se verificó la relación señal-ruido (SNR) por cada algoritmo de codificación. Primero se introdujo un archivo de voz (duración de 2 s) en formato .wav para ser codificado; Se obtuvo una señal codificada a la salida del programa, la cual conforme a la ecuación (1.12) se evaluó de manera objetiva (SEGSNR).

SEGSNR:

En la Figura 4.5, podemos apreciar que existe una tendencia muy similar para ambas gráficas con respecto a la relación señal-ruido por segmentos (SEGSNR).

De manera que para el primer códec LPC sus fluctuaciones se encuentran entre los 33 – 68 dB, mientras que para segundo códec RELP-GSM sus fluctuaciones están entre los 20 – 75 dB, está tendencia está asociada a la calidad de la voz y a la transmisión. Por lo tanto, si la tendencia es a la baja el error sería tan grande que resultaría imposible comprenden el mensaje y su transmisión seria a una corta distancia.

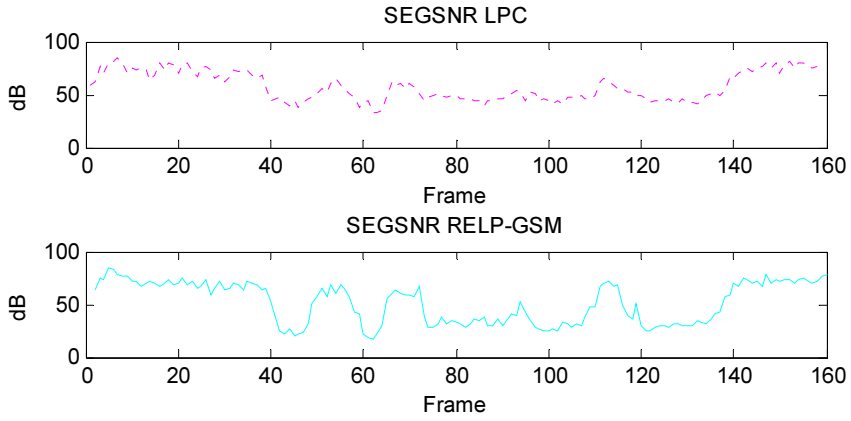


Figura 4.5: SEGSNR del Códec LPC y RELP-GSM utilizando Matlab
Voz de Hombre.

Desde el punto de vista de la calidad, el códec RELP-GSM cuenta con una mayor fidelidad con respecto al códec LPC.

Para poder tener una perspectiva de las relaciones señal-ruido por segmentos (SEGSNR) para el resto de los codificadores, se genero un gráfica Figura 4.6, con todos los resultados obtenidos, donde podemos apreciar que el códec ADPCM tiene un SNR alto en comparación con el resto de los códecs, pero su compresión es 1:2.

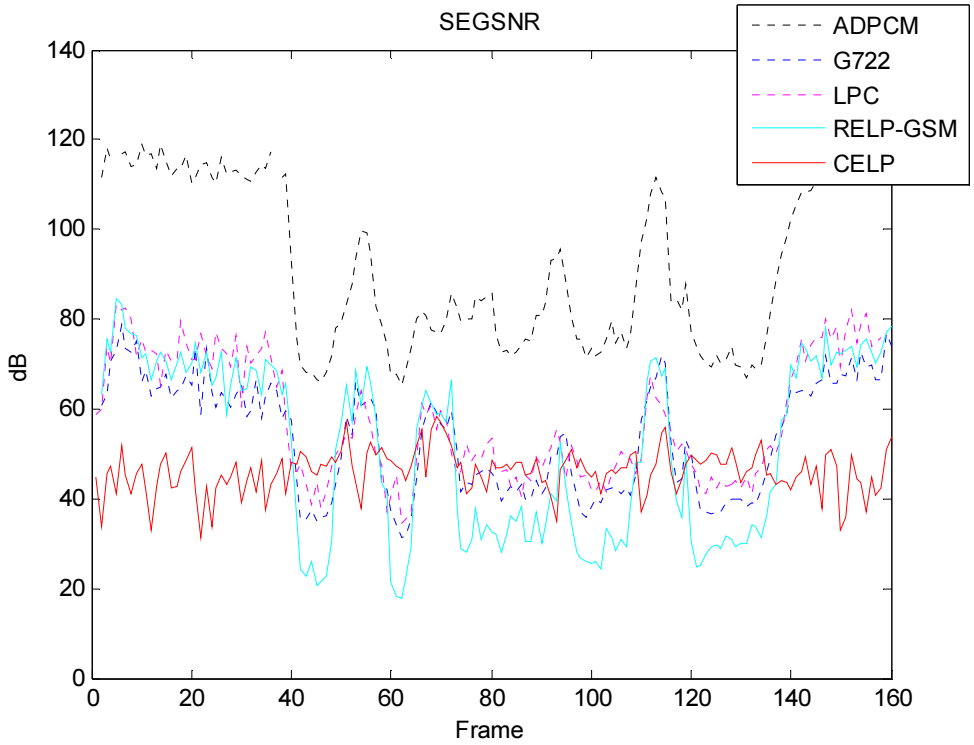


Figura 4.6: SEGSNR de diversos Códec utilizando Matlab
Voz de Hombre.

Por lo que respecta a la voz de mujer se puede apreciar un caso similar para los códecs, ver Figura 4.7, de igual forma el códec que tiene el mejor SNR es el códec ADPCM.

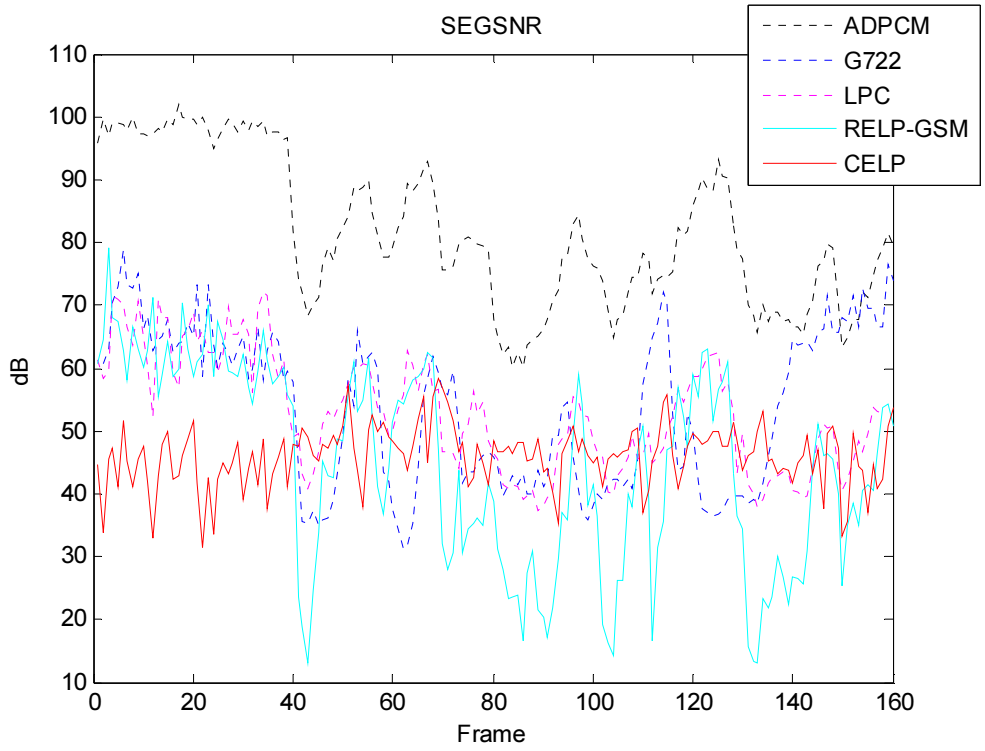


Figura 4.7: SEGSNR de diversos Códec utilizando Matlab Voz de Mujer.

A pesar de que el códec ADPCM tiene el mejor SNR este tiene una relación de compresión de 1:2. Está compresión es demasiado pequeña en relación a los demás códec, por lo que se considerara como punto de comparación solo para la calidad. Ahora bien, la relación de compresión del G722 es aproximadamente 1:4, La relación de compresión de LPC es aproximadamente 1:6.4 y para el códec RELP-GSM es aproximadamente 1:4.92, basados en estos criterios podemos decir, que el códec que proporciona una mejor SNR es el RELP-GSM.

Ancho de Banda

El *ancho de banda* de un canal es el rango de frecuencias que éste puede transmitir con razonable fidelidad. Por lo tanto, si aumentamos la velocidad de transmisión mediante la compresión en el tiempo de la señal en un factor de N, se transmitirá en 1/N del tiempo, y la velocidad de trasmisión será N. De esta forma, el índice de transmisión de un canal es directamente proporcional a su ancho de banda y este seria N.

Por ejemplo, para una señal se comprime en el tiempo en un factor de dos, se podrá transmitir en la mitad del tiempo, y la velocidad de transmisión se duplica. Sin embargo, la compresión por un factor de dos hace que la señal "oscile" dos veces más rápido, lo que implica que las frecuencias de sus componentes se dupliquen. Para transmitir sin distorsión esta señal comprimida, el ancho de banda del canal debe duplicarse.

Tasas de bits

Es importante obtener el cálculo de la tasa de transmisión porque de este depende el número de conversaciones que se mantendrán simultáneamente en un cierto ancho de banda, así como la calidad en que se transmitirán. Cuando se reduce la tasa de bits los requerimientos de ancho de banda también se reducen, lo que posibilita a la red poder manejar más conexiones simultáneas, pero esto incrementa la demora y la distorsión de la señales de voz.

Como se explicó en el primer capítulo PCM es el estándar G.711 sin compresión. Siendo la técnica base de todas las demás técnicas de codificación de la voz, pues generalmente se parte de una trama PCM para producir los demás estándares.

$$8000\left[\frac{muestras}{s}\right] * 8\left[\frac{bits}{muestras}\right] = 64 \text{ [kbps] para cada canal de voz}$$

Si bien en el segundo capítulo, se discutió la técnica ADPCM (Modulación por Pulsos Codificados Diferenciales Adaptativos) siendo el estándar de G.726. Para esta técnica solamente se requiere transmitir 4 bits de información en lugar de los 8.

$$8000\left[\frac{muestras}{s}\right] * 4\left[\frac{bits}{muestras}\right] = 32 \text{ [kbps] para cada canal de voz}$$

También se explicó el funcionamiento del códec G.722.1, donde se divide la banda de frecuencias de 0 a 8000 Hz en dos sub-bandas: la sub-banda inferior y la sub-banda superior, y se muestrea a 8 kHz.

$$4000\left[\frac{muestras}{s}\right] * 4\left[\frac{bits}{muestras}\right] = 16 \text{ [kbps] para cada canal de voz}$$

La codificación predictiva lineal LPC es un método de codificación de forma de onda visto previamente en el segundo capítulo. Las señales de voz al no variar significativamente nos permiten considerar como casi-estacionarias en periodos cortos de tiempo. Aprovechando esta predecibilidad para reducir la tasa de bits.

Se analizan 160 muestras en 20 ms.

$$8\left[\frac{bits}{muestras}\right] * 1/(0.8/1000)\left[\frac{muestras}{s}\right] = 10 \text{ [kbps] para cada canal de voz}$$

CELP es una técnica de codificación híbrida, que utiliza un modelo del tracto vocal muy similar al utilizado por LPC, hace uso de un libro de códigos que contiene una tabla con las señales de residuo típicas. Con esta codificación se logra obtener una calidad mucho mayor a la obtenida con LPC sin sacrificar mucho ancho de banda adicional.

Se analizaron cada 20 ms una muestra.

Banda inferior:

$$8\left[\frac{bits}{muestras}\right] * 1/(1.75/1000)\left[\frac{muestras}{s}\right] = 4.5714 \text{ [kbps]}$$

Retardo LTP: 4.5714 [kbps]
Índice Codebook: 4.5714 [kbps]
Ganancia VQ: 4.5714 [kbps]

Banda superior:

$$12\left[\frac{bits}{muestras}\right] * 1/(7/1000)\left[\frac{muestras}{s}\right] = 1.7144 \text{ [kbps]}$$

LSFs: 1.7144 [kbps]

$$4\left[\frac{bits}{muestras}\right] * 1/(7/1000)\left[\frac{muestras}{s}\right] = 0.5714 \text{ [kbps]}$$

Ganancia: 0.5714 [kbps]

Total: 16 [kbps] para cada canal de voz

Para la codificación RELP-GSM la señal de voz se pasa a través de un predictor lineal el cual elimina la correlación entre tramas. Si la predicción es bastante buena, la salida del predictor será aproximadamente ruido blanco. La idea de RELP es que una pequeña parte del residuo se transmite y a partir de él se reconstruye el residuo completo en el receptor.

Se analizaron cada 20 ms una muestra, con una excitación de 5 ms.

Predictor corto:
orden 8: 6, 6, 5, 5, 4, 4, 3, 3: 36 bits

Predictor largo:
ganancia (2 bits), posición (7 bits): 36 bits

Excitación regular: 13 pulsos, 4 posiciones.
posición (2 bits), ganancia bloque (6 bits): 32 bits
amplitudes 13 x 3 (39 bits): 156 bits

Total: 260 bits/20 ms:

13 [kbps] para cada canal de voz

Conforme a las tasas de bits obtenidas en cada una de las codificaciones se realiza una gráfica que muestra en bits promedios vs números de tramas, tanto para la voz de hombre como para la voz de mujer, ver Figura 4.8 y 4.9. Esta es una forma de estimar objetivamente la calidad del audio de los distintos codificadores, siendo la medida determinada por la tasa de muestreo y el número de bits por muestra.

Al analizar los resultados obtenidos, podemos determinar que la tasa de bits es un factor importante en cuanto a la compresión se refiere, pues es un parámetro útil por razones de almacenamiento, número de conexiones simultáneas, demora, distorsión y ancho de banda.

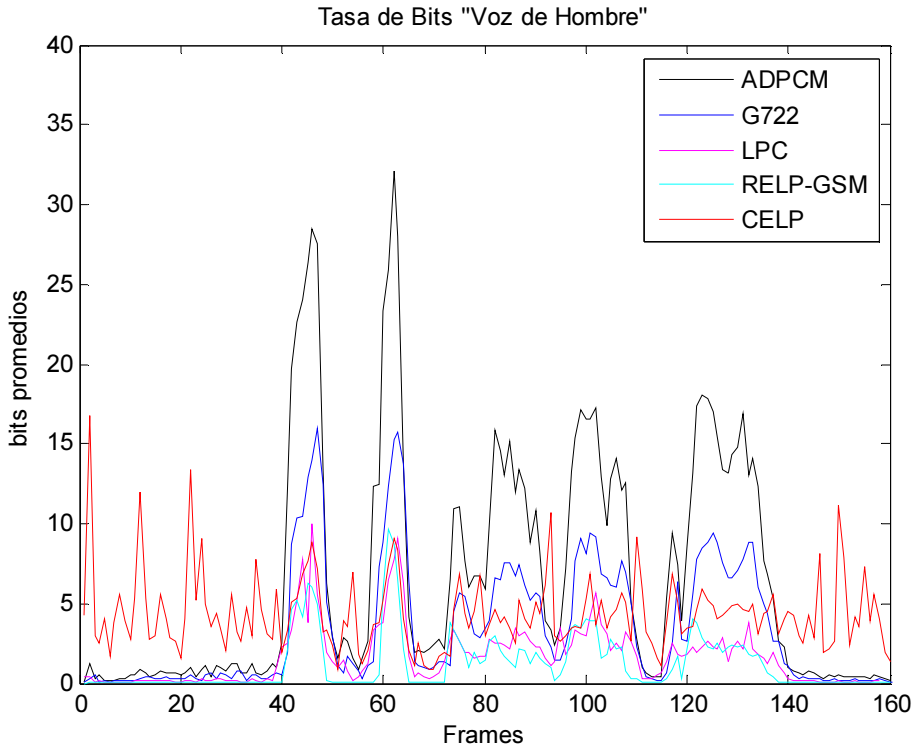


Figura 4.8: Tasa de Bits de diversos Códec utilizando Matlab Voz de Hombre.

Por lo que respecta a la voz de mujer se puede apreciar un caso similar para los códecs, ver Figura 4.9.

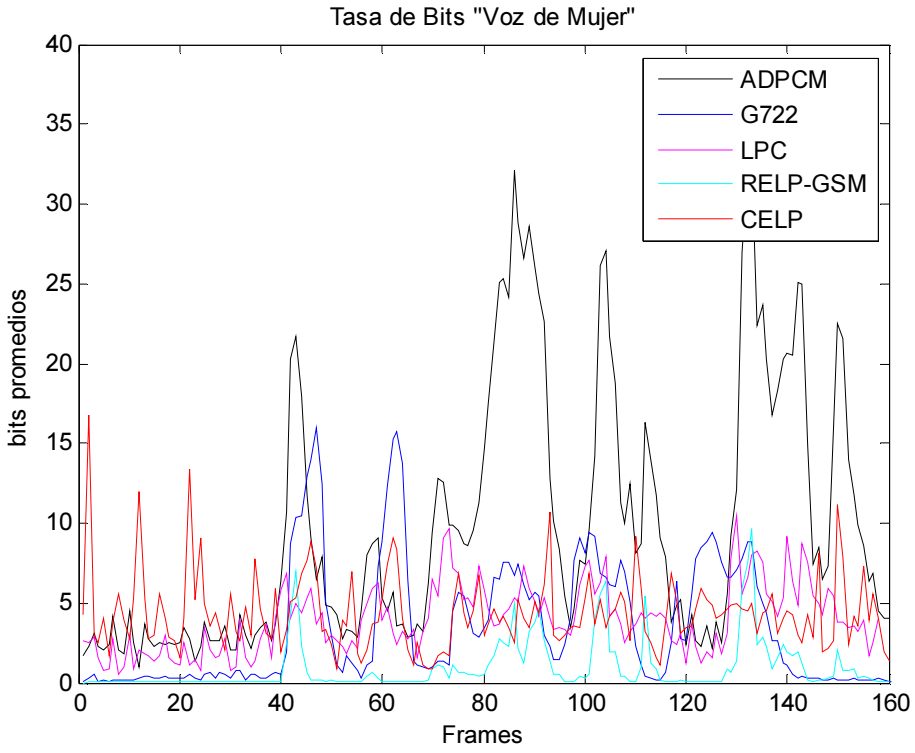


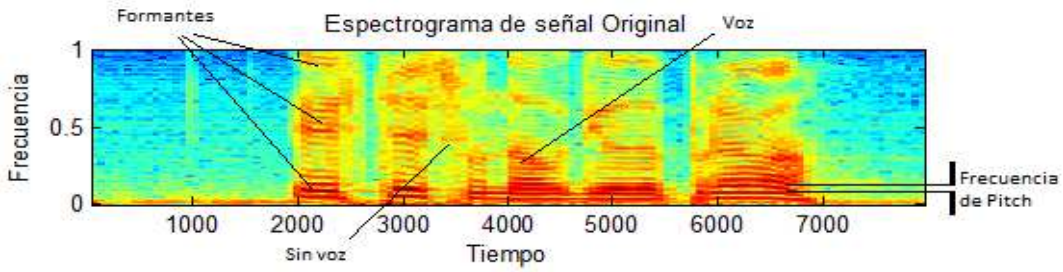
Figura 4.9: Tasa de Bits de diversos Códec utilizando Matlab Voz de Mujer.

Espectrogramas

Por otro lado, también realizamos la comparación desde el punto de vista espectral, debido a que las características de la señal de voz son variantes en el tiempo, la podemos representar mediante un espectrograma y darnos cuenta que los archivos son muy parecidos en vista de que la distribución de energía es similar para la señal original y la señal codificada.

En el espectrograma podemos ver que los formantes aparecen como franjas horizontales, mientras que los valores de amplitud en función de la frecuencia se representan en tonalidades de color naranja en sentido vertical. Siendo que un formante es el pico de intensidad en el espectro de un sonido; Se trata de la concentración de energía (amplitud de onda) que se da en una determinada frecuencia.

En la Figura 4.10, así como en la Figura 4.11 se representa el espectrograma de la señal original para la voz de hombre como para la voz de mujer, respectivamente.



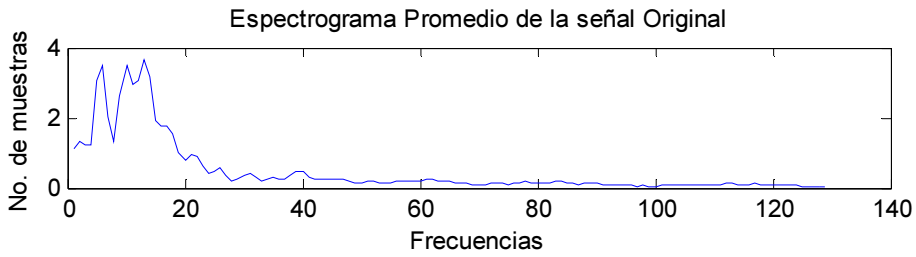


Figura 4.10: Espectrograma de la señal de voz de Hombre utilizando Matlab

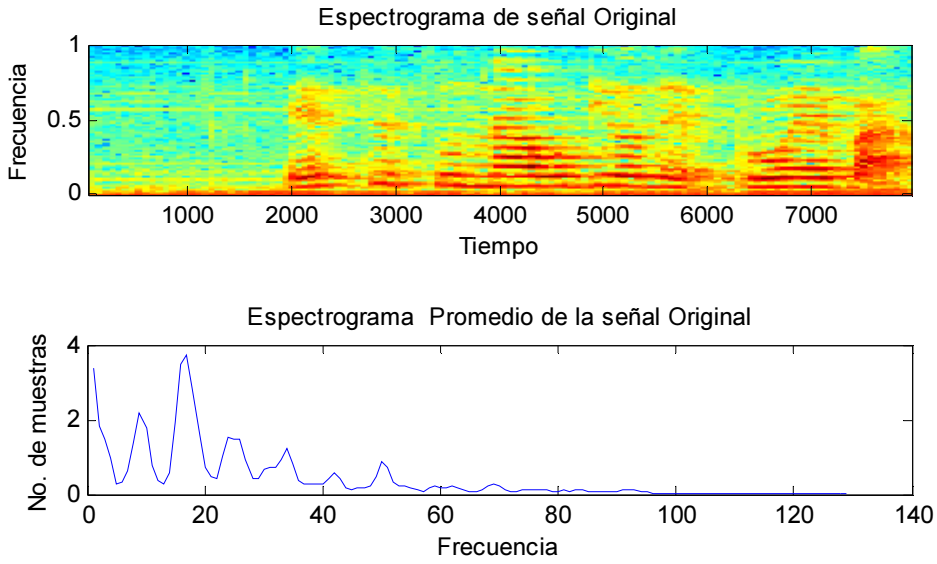


Figura 4.11: Espectrograma de la señal de voz de Mujer utilizando Matlab

Para cada espectrograma se puede visualizar el contenido en frecuencia de la señal en función del tiempo. En el eje vertical se encuentran las frecuencias discretas normalizadas entre 0 y 1, y en el eje horizontal equivale al tiempo (cada muestra es una ventana de análisis espectral).

Mientras que para el Espectrograma Promedio de la señal tenemos que en el eje horizontal corresponde al promedio de las frecuencias y en el eje vertical al número de muestras.

A continuación, se muestran los espectrogramas de la señal voz de hombre y de mujer al ser procesada por los distintos programas de codificación en Matlab. Para los espectrogramas de la voz de hombre se puede observarse que el audio presenta cierta distorsión en el dominio del tiempo, en tanto que en el dominio de la frecuencia se observa que se acentúan los componentes de baja frecuencia. En el caso de la voz de mujer se aprecia que existe una mayor distorsión para ambos dominios.

Sin embargo, la información principal de la voz se manifiesta cuando realizamos un análisis tiempo-frecuencia, a través de la función specgram de Matlab.

Estos son los espectrogramas de la señal de voz al ser procesada a través de los distintos códecs de voz:

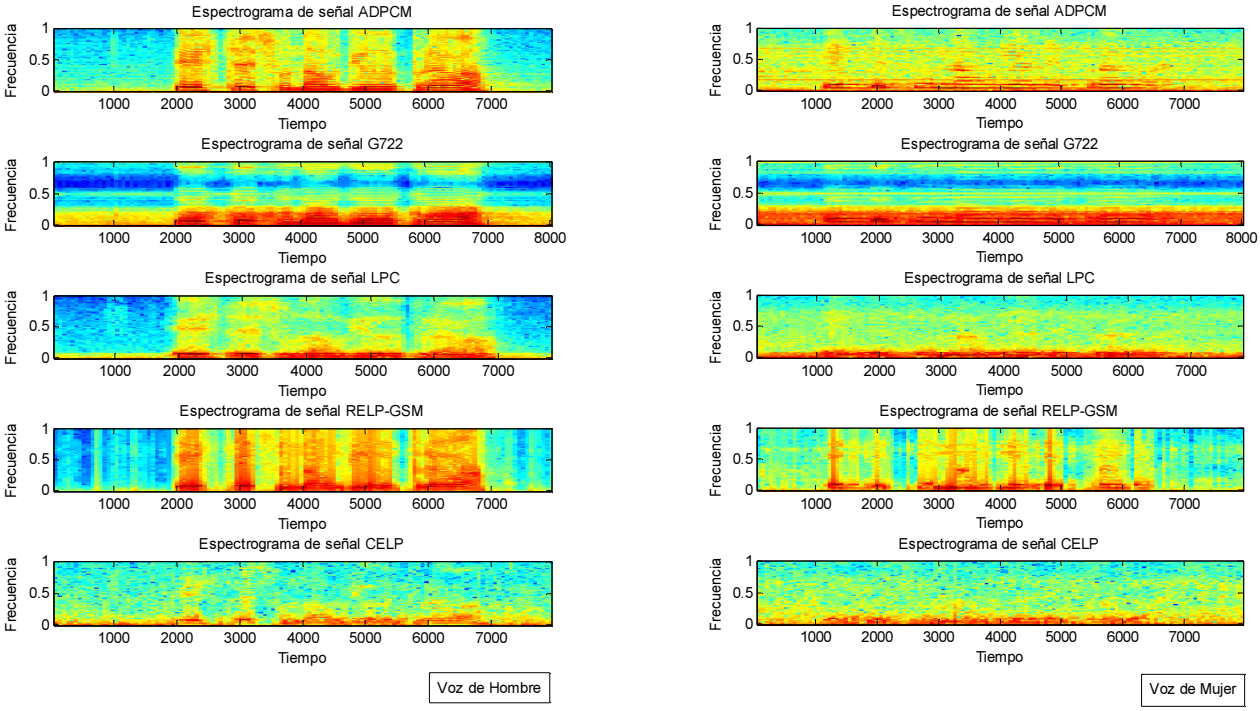


Figura 4.12: Espectrograma de la señal de voz de Hombre y de Mujer para diversos Códec utilizando Matlab

Se realiza una transformación en el dominio del tiempo al dominio de la frecuencia para ver las componentes de los datos de la señal de voz, donde utilizamos para este cambio la Transformada de Fourier Rápida (FFT). En la comparativa de los espectrogramas promedios de la señal podemos determinar que el códec RELP-GSM es quién presenta mayor número de muestras a una frecuencia menor de los 20 KHz.

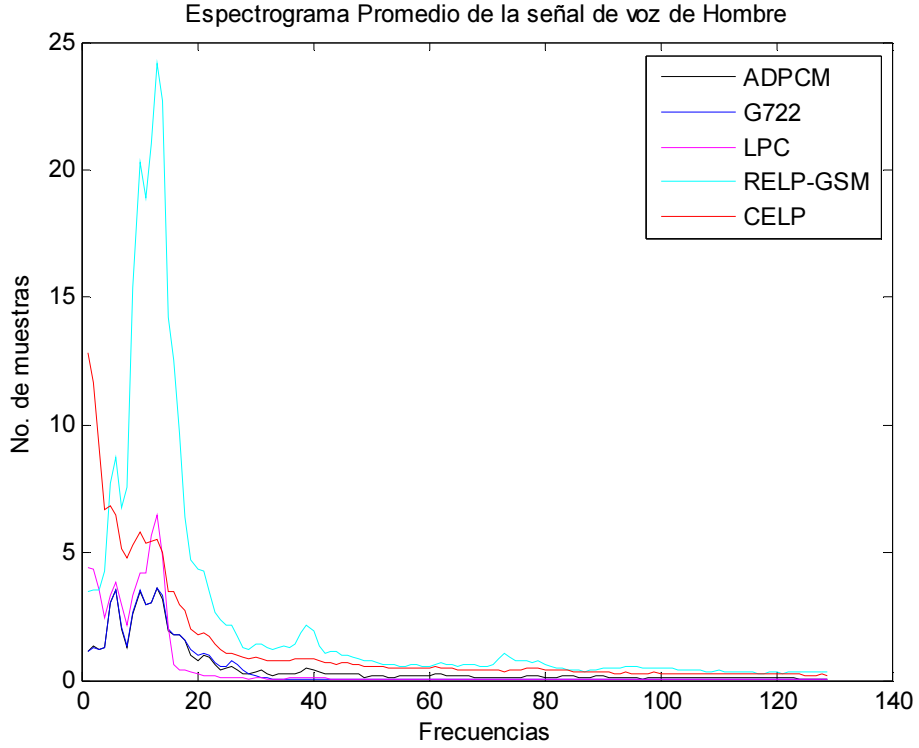


Figura 4.13: Espectrograma promedio de la señal de voz de Hombre de diversos Códec utilizando Matlab.

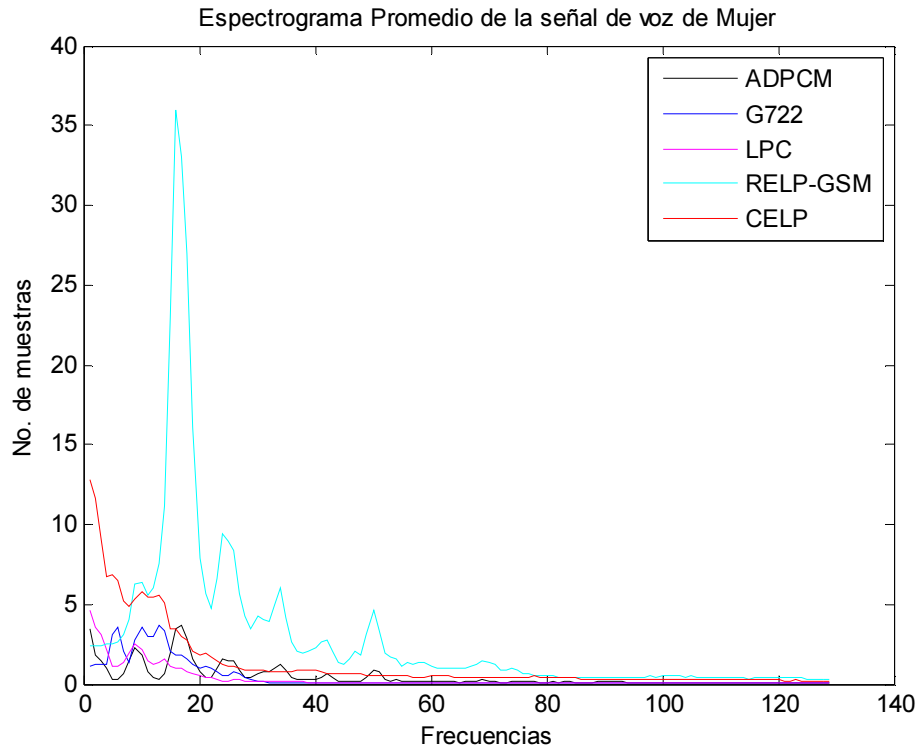


Figura 4.14: Espectrograma promedio de la señal de voz de Mujer de diversos Códec utilizando Matlab.

Ganancia

En la siguientes dos Figuras 4.15 y 4.16 podemos ver que la respuesta en frecuencia del sistema está presentada por varios picos, que corresponden a los formantes, los cuales son las frecuencias de resonancia del tracto vocal.

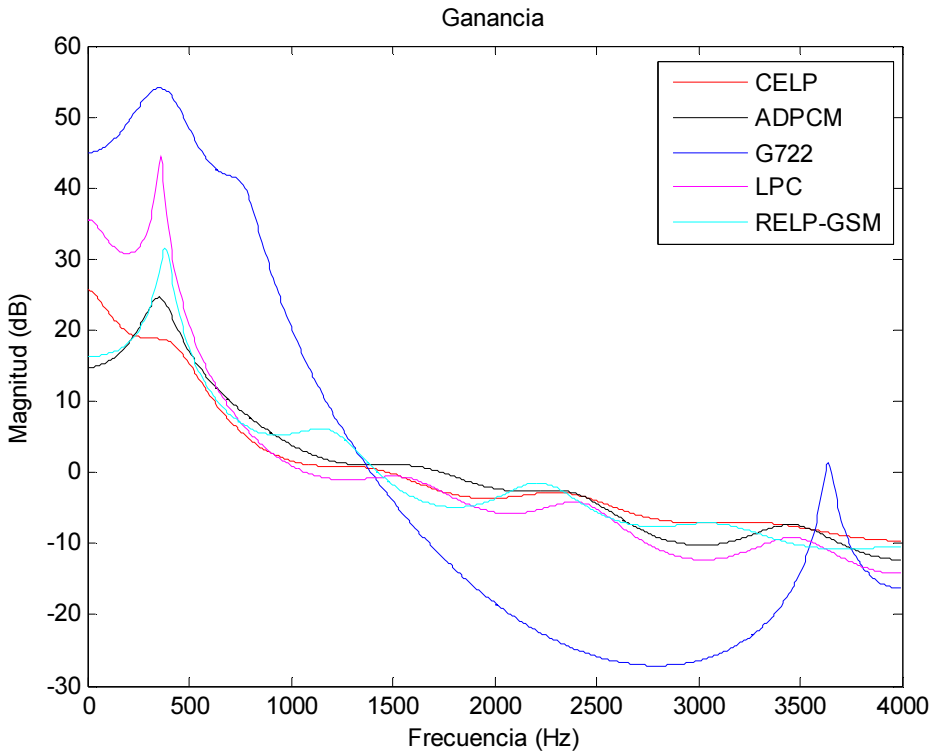


Figura 4.15: Magnitud (dB) vs Frecuencia (Hz) de la señal de voz de Hombre de diversos Códec utilizando Matlab

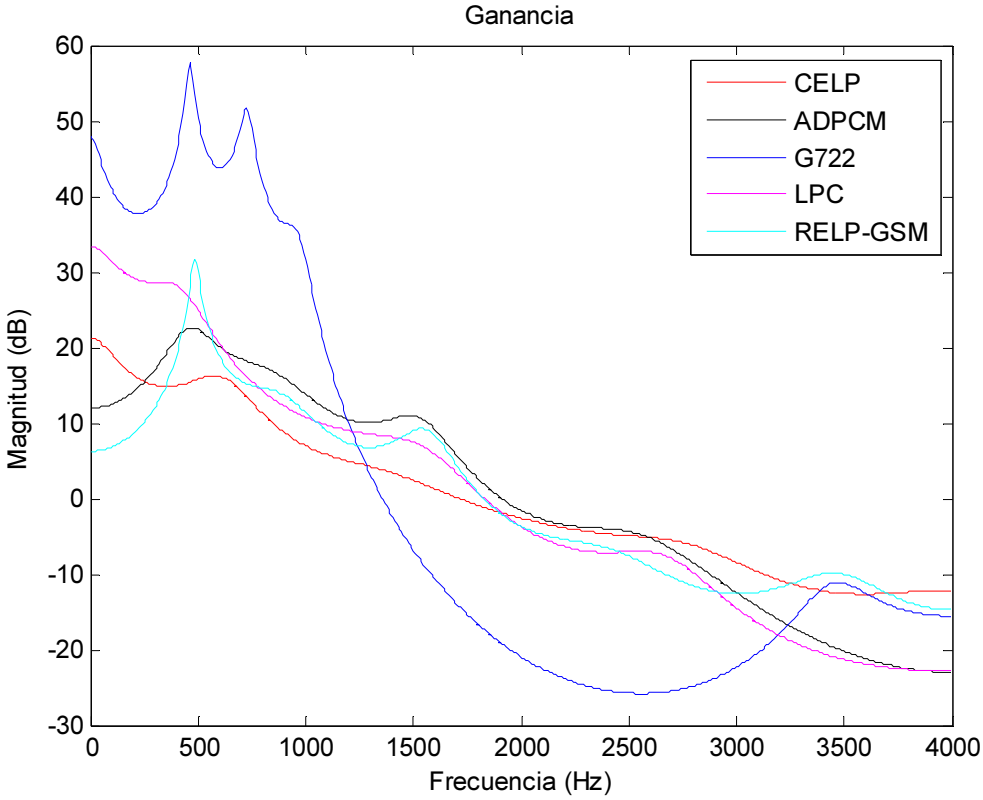


Figura 4.16: Magnitud (dB) vs Frecuencia (Hz) de la señal de voz de Mujer de diversos Códec utilizando Matlab.

4.7 Evaluación subjetiva de la calidad de voz

De las pruebas subjetivas que se realizaron se medirá la degradación sufrida por la señal al ser codificada. Conforme al uso del estimador MOS (Mean Opinion Score) tomaremos como referencia la señal original, y en base a ésta se asignara la calidad de la reconstruida.

En nuestra evaluación se recaudó la opinión de un conjunto de personas. La cual consistía en realizar una encuesta de opinión a cada usuario de prueba, con respecto a las señales reconstruidas en base al estimador MOS ver la siguiente Tabla 1.5:

MOS	Calidad	Clasificación de la voz
5	Excelente	Transparente
4	Buena	Red digital mejorada
3	Aceptable	Comunicaciones
2	Mediocre	Voz artificial
1	Mala	Muy Molesta

Tabla 1.5: MOS Mean Opinion Store.

En Figura 4.17 se puede ver una representación de la comparación de los distintos codificadores a través del uso del estimador MOS. La representación es un indicador obtenido a través de un sondeo realizado con 15 personas (usuarios de prueba) que nos compartieron su opinión en el desempeño de los distintos codificadores en cuanto a calidad de audio se refieren.

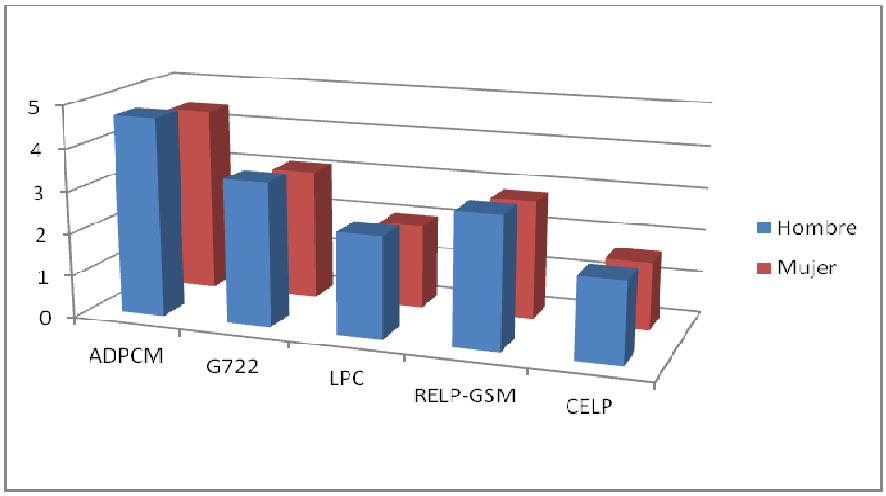


Figura 4.17: Medidas de calidad subjetivas, uso de MOS

En resumen, la calidad del audio procesado por los distintos codificadores de voz es sensible al tipo de voz, por lo que se puede apreciar una diferencia en la tonalidad de la voz de mujer con respecto a la voz de hombre.

Para la voz de hombre se aprecia una mejor calidad a través de cada uno de los códecs, siendo el algoritmo RELP-GSM el que cuenta con una mejor compresión, aunque las posibilidades de uso la limita aquellas en donde el usuario sólo requiera entender el mensaje, sacrificando un poco de calidad a cambio de un menor costo computacional, por ejemplo el retardo, el costo, la demora, etc. Por lo que respecta a la voz de mujer, la calidad es mejor cuando se utiliza el algoritmo RELP-GSM, tomando en cuenta la capacidad de compresión.

4.8 Aplicaciones

Actualmente, existen diversas conexiones inalámbricas, que se han establecido como estándares, pero pese a su innegable funcionalidad y éxito, dependen en gran medida de las tasas de compresión para mantener una calidad aceptable.

Existen diversos códecs de audio, los cuales cuentan con diferentes niveles de rendimiento, estos códecs están basados en alguna o algunas técnicas de compresión (ver Tabla 4.1).

Codec	Tasa de Bit (kbps)	Tecnología de Compresión
Tasa completa	13	RTE-LPC
EFR	12.2	ACELP
Tasa media	5.6	VSELP
AMR	12.2 - 4.75	ACELP
AMR-WB	23.85 - 6.60	ACELP

Tabla 4.1: Códec de audio.

Siendo que el principal objetivo de los códecs de audio es reducir la cantidad de datos digitales necesarios para reproducir la señal, esto resulta importante al momento de almacenar o transmitir la señal a través de un medio inalámbrico.

4.8.1 GSM Tasa Completa / Còdec RPE-LPC

El códec RPE-LPC o Excitación de Pulsos Regulares – Codificador Predictivo Lineal fue utilizado por primera vez con la tecnología GSM.

Este códec estas relacionado con dos códecs previos: RELP, Predicción Lineal de Excitación Residual y el MPE-LPC, LPC con Excitación de Multi-Pulsos. El primero tiene las ventajas de tener una complejidad relativamente baja como resultado del uso de la codificación de banda base, pero su rendimiento se ve limitada por el ruido tonal producido por el sistema. Mientras que para el segundo códec es más complejo, pero ofrece un mejor nivel de rendimiento. Por lo que el códec RPE-LPC mantiene un balance entre el rendimiento y la complejidad de la tecnología.

A pesar del trabajo realizado por el códec para proporcionar el rendimiento óptimo, las tecnologías sean desarrollado aun más y el códec de RPE-LPC sea visto con un ofrecimiento bajo de calidad para este tipo de tecnologías, y sea visto remplazado por el Códec AMR.

4.8.2 AMR-WR+

Con la aparición del códec AMR-WB+ se exhibe una mejor calidad de audio a tasas bajas, si se consideran distintos tipos de contenido. Como era de esperar, la calidad registrada por el contenido de la voz predominantemente es significativamente mejor que la registrada para los códecs de audio.

Características:

- Estándar G.722.2
- Tasa de Bit (kbit/s) – Soportar 9 posibles tasas de bits de codificación que varía desde 6.6 hasta 23.85kbps.
- Está basado en la técnica hibrida incluye la técnica de codificación ACELP (Algebraic Code Excited Linear Prediction) para la manipulación de las señales de voz, y en la transformada en el dominio de la frecuencia para la codificación eficiente de las señales de audio y de música.
- Tamaño de la trama 0.125 ms, 1.5 ms, 20 ms, 5 ms
- Calidad es comparada a 64 kbit/s
- Calidad alta de 7 kHz de BW a tasas de bits tan bajas como 12.65 kbps, no se desempaña bien con audio.
- Complejidad 10 MIPS < 15 MIPS
- Principales Aplicaciones ISDN, Video conferencias, Redes inalámbricas 3G.

El ancho de banda multi-tasa adaptable AMR-WB es el códec obligatorio en las redes GSM y WCDMA para servicios de conversación y multimedia cuando estos servicios evolucionan a voz de ancho de banda (~50 a 7 kHz).

4.8.3 Análisis de la Voz en redes

Con la migración de la voz a otras redes, dio como consecuencia los avances en las técnicas de procesamiento de señales y, posteriormente con la aparición de los procesadores digitales de señales o DSP, una vez que la voz se encontraba incorporar a las redes de paquetes, se da inicio el intento por reducir a través de algoritmos de compresión la cantidad de bits enviados.

Adicionalmente, como sabemos todos los medios de transmisión introducen ruido, que se traduce como pérdida de datos. Puesto que los mecanismos de retransmisión no son viables para la comunicación de voz, es necesario evaluar el codificador en condiciones de pérdida de información, lo que se conoce como la robustez del codificador.

Los componentes esenciales de la voz se muestran en la Figura 4.18, que hacer referencia a una típica conversación.

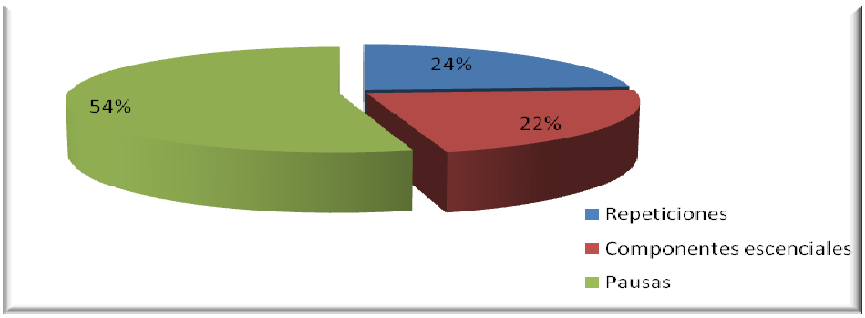


Figura 4.19: Gráfica de los componentes esenciales de la Voz.

La compresión de la voz busca satisfacer las siguientes características:

- Menor velocidad
- Mayor calidad (Medidas Subjetivas y Medidas Objetivas)
- Mayor eficiencia en el algoritmo (Complejidad-MIPS)
- Menor retardo en la compresión (<100ms)

Las redes inalámbricas de más éxito actualmente son las de telefonía celular. Sin embargo, con el aumento en la demanda y el costo la limita, es por esto que las redes de datos toman especial importancia y en poco tiempo las redes basadas en IP superaran su principal desventaja frente a estas tecnologías, la movilidad.

Por otro lado las redes inalámbricas no corresponden a ninguna de las generaciones celulares pues han sido desarrolladas con propósitos diferentes, pese a eso y debido a la integración de los dispositivos móviles y diversidad de servicios con creciente demanda, las redes celulares requieren cada vez más recursos, recursos que pueden ser provistos por redes basadas en IP.

La convergencia de las redes inalámbricas y las redes celulares pretende lograr la unión de los sistemas de comunicaciones que permitan una mayor expansión y compatibilidad entre las dos distintas tecnologías.

4.9 Redes LTE (Evolución a Largo Plazo - Long Term Evolution)

Con el crecimiento de las comunicaciones móviles, primero de la mano de GSM y últimamente con el despliegue definitivo de UMTS, surge una nueva generación de comunicaciones móviles, la cuarta generación o 4G, de la que el sistema LTE "Long Term Evolution", cuya primera especificación fue concluida por 3GPP y dando pie a la LTE-Advance [10].

Con el surgimiento de las redes LTE en un mercado tan competitivo y en desarrollo se busca llevar de forma rentable Internet móvil, lo que permitirá suministrar velocidad y eficiencia a diversas aplicaciones móviles.

En una comparativa con 3G, LTE nos proporciona:

- Hasta diez veces más velocidad.
- Una latencia de tres a cinco veces menor.
- Una eficiencia espectral de dos a tres veces mayor y que aumentará con mejoras de LTE.
- Pasar a todo a redes basadas en IP.
- Cambian la forma de cómo se diseñan, despliegan y gestionan las redes.
- Disminuye el coste total de propiedad de la red
- Reduce el consumo de energía y el volumen, suministrando una solución sostenible.
- Multiplica capacidad y flexibilidad para gestionar el crecimiento

A diferencia de 3G, LTE es todo IP. Esto es un aspectos relevantes de LTE porque todos los servicios, incluida la voz, se soportan sobre el protocolo IP (Internet Protocol), y que las velocidades de pico de la interfaz de radio se sitúan dentro del rango de 100 Mb/s y 1Gb/s. Con LTE se espera romper finalmente las barreras de la movilidad con una capacidad multimedia.

De hecho, se basa en IPv6 que soporta una cantidad enorme de direcciones IP adicionales y proporciona otras mejoras respecto a IPv4. Es una arquitectura más sencilla, escalable y económica. Y abre el acceso a nuevos segmentos de mercado como el uso de equipos inalámbricos móviles como laptop, ipod, ipad, etc.

Estos son algunos de los beneficios que nos proporciona LTE:

- Proporciona un ecosistema global con movilidad inherente.
- Ofrece fácil acceso y el uso de una mayor seguridad y privacidad.
- Mejora dramáticamente la velocidad y la latencia.
- Proporciona mejorado en tiempo real de video y multimedia.
- Crea una plataforma sobre la cual construir y desplegar productos y servicios.
- Reduce el coste por bit a través de una mejor eficiencia espectral.

Como se mencionó la compresión de la voz hoy en día se hace imprescindible con la incesante necesidad surgida a partir de una mayor demanda establecida por usuarios sobre las redes inalámbricas y redes celulares, no obstante la tecnología ha ido evolucionando rápidamente para ajustarse a tales necesidades lo que ha permitido el uso de banda ancha y quizás la pronta convergencia de ambas tecnologías con LTE.

Conclusiones

La intención de esta tesis es abordar algunas técnicas de compresión para la voz, presentando así las bases de la codificación de la voz y su acelerado impulso en las comunicaciones inalámbricas de banda ancha.

La codificación de la voz se puede resumir como el esfuerzo para reducir el ancho de banda de transmisión de la voz a través de una eficiente codificación que nos permita establecer las condiciones adecuadas o idóneas para que los paquetes sean transmitidos sobre una red inalámbrica, y poder representar de una forma mínima la señal de voz, mientras se mantiene un nivel aceptable de calidad de percepción de la voz decodificada, siendo este su principal objetivo.

Las razones del uso de las técnicas de compresión son diversas, pero debido al uso de las comunicaciones totalmente digitales ha permitido que el procesamiento de las señales de audio sea utilizado y la representación de la señal se vea manipulada antes de ser transmitida. Es importante utilizar sistemas que nos permitan reducir los requerimientos del espectro sin empobrecer la calidad de la transmisión, lo que nos permite que sea práctico y económico.

De acuerdo con lo expresado anteriormente, se puede decir que las dos razones fundamentales por lo cual las técnicas de compresión se usan son: Irrelevancia y redundancia. Una señal es irrelevante cuando su presencia no es perceptible o cuando no produce efecto alguno sobre el sistema, mientras que una señal es redundante cuando su presencia, aunque perceptible, no provee un aporte a la información ya conocida. De acuerdo con estas dos características, los distintos métodos de compresión se proponen eliminar del caudal de datos aquéllos que son irrelevantes y/o redundantes.

A lo largo de la historia el desarrollo de la tecnología de señales digitales, la fabricación de microprocesadores ha permitido la ejecución en tiempo real de algoritmos extremadamente complejos, garantizando así que la calidad del audio se vea apenas afectadas por la compresión. Estás técnicas de reducción de *redundancia* para aplicaciones sean vuelto necesarias para encontrar una manera adecuada de acomodar los cuantiosos flujos de datos a los medios de transmisión disponibles.

En este trabajo se llevaron a cabo simulaciones a través del software de Matlab de las cuales se evaluaron distintas técnicas de compresión, la evaluaron se realizó para medidas objetivas y subjetivas con el fin de describir el comportamiento de la voz, así como la percepción de la calidad de la misma. Los algoritmos que se utilizaron para realizar la comparación cuentan con parámetros que realmente influyen en la calidad de la conversación por ejemplo, el tamaño de los paquetes, el número de bits, etc.

En este contexto, se desarrolló un capítulo donde se analizaron y evaluaron las técnicas de compresión para la voz a través de bloques de programación (Matlab), concluyendo que la técnica RELP-GSM nos proporciona una calidad mucho mayor en comparación a las demás técnicas estudiadas, esto se logra debido a que esta codificación elimina la correlación entre tramas.

Como medidas objetivas se analizó la relación señal-ruido y se evaluó en cada uno de los distintos codificadores lo cual nos permitió calcular la energía de la señal con relación con la energía del ruido, donde pudimos apreciar que el códec ADPCM tiene un SNR alto. Sin embargo, su tasa de bits es mayor que el resto de los códec, por lo que utilizando una gráfica comparativa entre los codificadores y basados en la tasa de bits (un valor importante cuando se quiere transmitir), podemos decir, que el mejor codificador es RELP-

GSM. También se realizó la comparativa desde el punto de vista espectral y pudimos apreciar cómo se distribuye la energía en cada uno, y ver la intensidad del sonido (formantes).

Para las evaluaciones subjetiva se utilizó el estimador MOS (Mean Opinion Score), el cual toma como referencia la señal original, y en base a ésta se asignara la calidad de la distorsionada. De los resultados obtenidos podemos ver que los distintos codificadores si son un poco sensibles al tipo de voz, donde se aprecian diferencias en las tonalidades (voz de mujer y voz de hombre), siendo la voz de hombre de mejor calidad que la de la voz de mujer, además pudimos determinar que el algoritmo RELP-GSM cuenta con una calidad aceptable con respecto a su tasa de bits.

En los sistemas de comunicaciones inalámbricos es primordial el empleo eficiente del ancho de banda, de ahí la importancia del uso de las técnicas de compresión, que aunado a bajar las tasas de transmisión estas aumentan en complejidad y el retardo. En consecuencia surge el códec AMR-WB+ que es una de las principales técnicas para la transmisión y almacenamiento de contenido de audio a través de enlaces inalámbricos, que está basada en una técnica de codificación hibrida (codificación ACELP “Algebraic Code Excited Linear Prediction” y en la transformada en el dominio de la frecuencia).

Con la tecnología LTE se espera tener un aumento en la fluidez de la información, es importante mencionar que LTE no es una tecnología de estándar única, sino por el contrario se basa en estándares abiertos, lo que es igual a una diversidad de tecnologías y protocolos que permiten mayor compatibilidad con el resto de las tecnologías. Esta capacidad de adecuar viejas tecnologías con LTE garantiza costos reducidos de instalación y operación, ya que permite la adecuación de los equipos tecnológicos ya instalados y operativos, adaptarlos a la nueva tecnología.

Trabajos futuros

Llevar a cabo el análisis del desempeño de los códecs de voz sobre una red basada en tecnología Long Term Evolution (LTE). Buscando obtener algunos performance de esta tecnología, como son: el flujo de mensajes de datos, paquetes perdidos, mensajes de control de flujo, caídas de los enlaces, etc.

Con ésta evaluación obtendríamos estadísticas del desempeño de la tecnología sobre redes inalámbricas, así como las ventajas y desventajas de las mismas.

Conforme a la investigación, OPNET es un software que está diseñado para simular ambientes de prueba y, permite ver el comportamiento y rendimiento de una red LTE, pero la información que se inyecta al simulador es un modelo de tráfico (la representación de los códecs está basada en la tasa de bits), lo que implica que a la salida de la red obtendremos simplemente estadísticas de control de congestión de la red, lo que nos ayudara a determinar los efectos del uso de las redes LTE en un ambiente controlado.

Apéndice A

ITU-T P.862 (PESQ)

La recomendación ITU-T P.862 describe un método objetivo para predecir la calidad subjetiva de la voz. Siendo descrito el método de "evaluación de la calidad vocal por percepción" (PESQ, perceptual evaluation of speech quality), el cual es aplicable no sólo a los códecs vocales sino también a las mediciones de extremo a extremo. PESQ compara una señal inicial X(t) con una señal degradada Y(t) que se obtiene como resultado de la transmisión de X(t) a través de un sistema de comunicaciones (por ejemplo, una red IP). La salida de PESQ es una predicción de la calidad percibida por los sujetos en una prueba de escucha subjetiva que sería atribuida a Y(t).

En el primer paso PESQ consiste en una alineación temporal entre las señales iniciales X(t) y la degradada Y(t). Para cada intervalo de señal se calcula un punto de arranque y un punto de parada correspondientes. Una vez alineadas, PESQ compara la señal de entrada (inicial) con la salida degradada, utilizando un modelo por percepción, como el presentado en la Figura 6.1. Lo esencial en este proceso es la transformación de las dos señales, la inicial y la degradada, en una representación interna que es análoga a la representación psicoacústica de señales de audio en el sistema auditivo humano, teniendo en cuenta la frecuencia por percepción y la sonoridad.

El modelo PESQ termina brindando una distancia entre la señal inicial y la señal degradada. La que corresponde a su vez con una predicción de la MOS subjetiva. El resultado de PESQ es similar a la escala de MOS, un número único en una escala de –0.5 a 4.5, aunque en la mayoría de los casos la gama de las salidas estará entre 1.0 y 4.5, que es la gama normal de valores de MOS que suelen darse sobre la calidad de voz.

La descripción detallada del algoritmo es compleja, y puede verse en la Recomendación referenciada. El método PESQ es objetivo e intrusivo, ya que requiere del envío de una señal conocida de referencia para evaluar la calidad percibida de la voz, PESQ mide la calidad de un solo sentido.

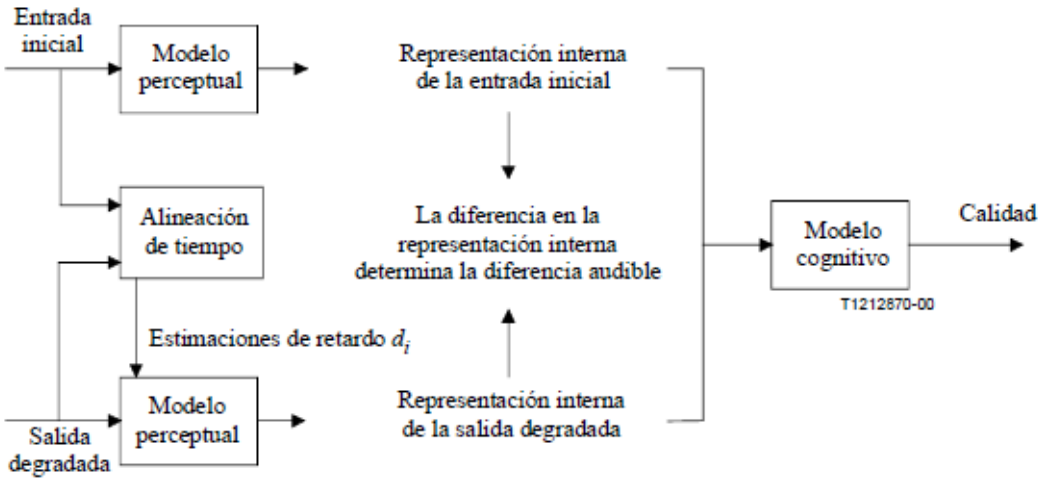


Figura 6.1: Visón general de los principios utilizados en PESQ.

El tratamiento llevado a cabo por PESQ se ilustra en la Figura 6.2. El modelo incluye las siguientes etapas:

- Nivel de alineación. Con el fin de comparar las señales, la señal de voz de referencia y la señal degradada, están alineados con el mismo nivel de potencia.

- Filtrado a la entrada. Los modelos PESQ y el compensador de filtrado se realizan en el auricular y en la red.

- Tiempo de alineación. El sistema puede incluir un retardo, que se puede cambiar varias veces durante una prueba por ejemplo voz sobre IP a menudo tiene retardo variable.

- Transformación Auditiva. La referencia y señales degradadas son pasadas a través de una transformada auditiva que imita las propiedades esenciales de la audición humana.

- Procesamiento de perturbación. Para los parámetros de perturbación se calculan utilizando promedios no lineales sobre áreas específicas de la superficie de error:

  - La alteración absoluta (simétrica): Una medida de error audible absoluto.
  - La alteración aditiva (asimétrica): Una medida de errores audibles, que son significativamente más alto que el de referencia.

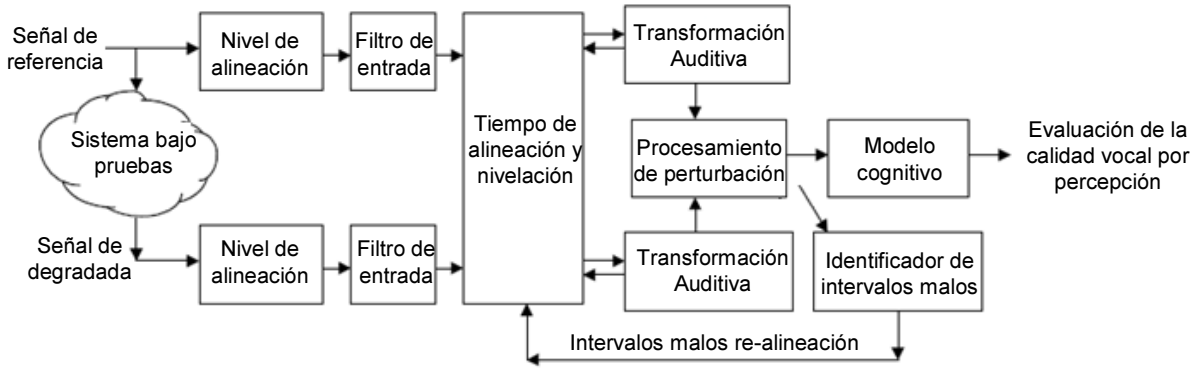


Figura 6.2: Estructura de PESQ.

PESQ se puede utilizar en una variedad de aplicaciones, incluyendo el aprovisionamiento, la puesta en marcha y para solucionar algún problema de redes, así como para realizar pruebas de equipos de telecomunicaciones. PESQ hace que sea posible obtener una opinión rápida de calidad y realizar pruebas exhaustivas de diferentes sistemas de comunicaciones, siendo utilizado con éxito para comparar tecnologías y escenarios de distorsión para las (redes móviles, VoIP y códecs de voz).

El impacto de los cambios a un algoritmo decodificación puede ser rápidamente investigados utilizando el modelo objetivo, incluso si su efecto es pequeño. El modelo también puede ser utilizado para explorar cómo la calidad varía con la velocidad de bits, el nivel de entrada o los errores del canal.

Apéndice B

PROGRAMACIÓN MATLAB

Estas son las pruebas programadas en Matlab para el codificador RELP-GSM, la programación es la misma para los demás codificadores.

```
%Reproducir las señales y almacenarlas

disp('Presione cualquier tecla para reproducir la voz original');
pause;
soundsc(x, fs);
pause(2);
disp('Presione cualquier tecla para reproducir la voz codificada');
pause;
soundsc(reconstruida, fs);

figure (1);
subplot(211),plot(x); title(['Señal Original']);
subplot(212), plot(reconstruida,'r'); title('Señal reconstruida');

wavwrite(reconstruida,fs,'Voz_RELP_GSM.wav');

Original= wavread('Voz_prueba10.wav');
Synth = wavread('Voz_RELP_GSM.wav');

%SNR en dB
figure(2);
e=x' - reconstruida
snr = 10*log10(sum(x.^2) ./ e.^2);
p2=promedio(snr);
plot(p2,'*:r');ylabel('dB');xlabel('Frame');
bits=snr;

%Espectroframa
figure(3)
specgram(reconstruida);ylabel('Frecuencia');xlabel('Tiempo');title('Espectrograma de señal
RELP-GSM');
save RELP_GSM2 synth_speech
save RELP_GSM1 p2

figure(4)
%Determinar los componentes de la frecuencia de los datos
xfft=abs(fft(reconstruida));
mag=20*log10(xfft);
mag5=mag(1:end/2);
plot(mag5,':c');ylabel('Magnitud (dB)');xlabel('Frecuencia (Hz)');
grid on
save RELP_GSM mag5

%Espectro promedio
figure(5)
P4 = mean(abs(specgram(reconstruida)'));
plot(P4);ylabel('No. de muestras');xlabel('Frecuencias');title('Espectrograma Promedio de
la señal Original');
save gsm3 P4
```

APÉNDICE B

```
%Bits promedios
figure(6);
a=abs(reconstruida);
b=(promedio(a));
c=b-min(b);
BIT=bits_codec/(max(c))
BIT5=(promedio(BIT*a));
plot(BIT5,':');ylabel('bits promedios');xlabel('Frames');
axis([0 160 0 40]); %% zoom in
save RELP_GSM4 BIT5

%Ganancia
figure(7)
ncoeff=2+fs/1000;
a=lpc(x,ncoeff);
b=lpc(reconstruida,ncoeff);
% Grafica de la respuesta a la frecuencia
[h,f]=freqz(1,a,512,fs);
subplot(2,1,1);
plot(f,20*log10(abs(h)+eps));
xlabel('Frecuencia (Hz)');
ylabel('Ganancia (dB)');

[h,f]=freqz(1,b,512,fs);
subplot(2,1,2);
plot(f,20*log10(abs(h)+eps));
xlabel('Frecuencia (Hz)');
ylabel('Ganancia (dB)');
```

Líneas de ejecución para graficar los resultados obtenidos.

```
load adpcm2 YY
load G7222 sub
load LPC2 outspeech2
load RELP_GSM2 synth_speech
load celp2 y

subplot(5,1,1)
specgram(YY);ylabel('Frecuencia');xlabel('Tiempo');title('Espectrograma de señal ADPCM');
subplot(5,1,2)
specgram(sub);ylabel('Frecuencia');xlabel('Tiempo');title('Espectrograma de señal G722');
subplot(5,1,3)
specgram(outspeech2);ylabel('Frecuencia');xlabel('Tiempo');title('Espectrograma de señal
LPC');
subplot(5,1,4)
specgram(synth_speech);ylabel('Frecuencia');xlabel('Tiempo');title('Espectrograma de señal
RELP-GSM');
subplot(5,1,5)
specgram(y);ylabel('Frecuencia');xlabel('Tiempo');title('Espectrograma de señal CELP');

load adpcm p4
load G722 p1
load LPC1 p5
load RELP_GSM1 p2
load celp1 p3

plot(p4,':k');ylabel('dB');xlabel('Frame');
hold on
plot(p1,':');ylabel('dB');xlabel('Frame');
hold on
plot(p5,':m');ylabel('dB');xlabel('Frame')
hold on
plot(p2,'c');ylabel('dB');xlabel('Frame');
hold on
plot(p3,'r')

load adpcm3 P1
load G7223 P2
load LPC3 P3
load gsm3 P4
load celp3 P5

plot(P1,'k');
hold on
plot(P2,'b');
hold on
plot(P3,'m');
```

```
hold on

plot(P4,'c');ylabel('No. de muestras');xlabel('Frecuencias');title('Espectrograma Promedio
de la señal Original');
hold on
plot(P5,'r');ylabel('No. de muestras');xlabel('Frecuencias');title('Espectrograma Promedio
de la señal Original');

load adpcm4
load G7224
load LPC4
load RELP_GSM4
load celp4

plot(BIT2,'k');
hold on
plot(BIT3,'b');
hold on
plot(BIT4,'m');
hold on
plot(BIT5,'c');
hold on
plot(BIT1,'r');

ylabel('bits promedios');xlabel('Frames');
axis([0 160 0 40]); %% zoom in
```

Es la función para segmentar una señal para su análisis

```
%Función para segmentar la señal
function [vector] =promedio(x)
n=length(x)+1
%n=14000+1;
w=100;
a=1;
y=0;
p=0;
while n>w
    b=x(y+1:w);
    v=mean(b);
    y=y;
    w=w+100;
    y=y+100;
 vector(a)=v;
 a=a+1;
end
%t=promedio(x)
%figure 2;
%plot(t,'.:');
```

Este es archivo de ejecución para el codificador ADPCM

```
%Comparar la señal de entrada y señal reconstruida del codificador ADPCM

close all;
clear all;
audio=wavread('Prueba_Voz_H.wav');
[Y,Fs,nbits] = wavread(' Prueba_Voz_H.wav');
Fs;
nbits;
tam=length(Y);
subplot(2,1,1)
plot(audio),axis('tight'),title('Voz Original');
y1 = adpcm_encoder(Y);
tam1=length(y1);
YY = adpcm_decoder(y1);
subplot(2,1,2)
plot(YY,'r'), axis('tight'),title('Voz reconstruida');
```

Referencias Bibliográficas

[1] Joint Coding Rate Control for Audio Streaming in short Range Wireless Networks.
Jelena Kovacevic, Dragan Samardzija, M. Temerinac,
IEEE transaction on Consumer Electronic, Vol. 55 No2, Mayo 2009. PP 486 – 491.

[2] Voice and Audio Compression for Wireless Communications.
Lajos Hanzo, F. Clare Somerville, Jason Woodard, John Wiley.
Segunda edición, Agosto 2007. PP 212.

[3] Sistemas de Comunicaciones Electrónicas.
Wayne-TomasiT.
Pearson Educación, 4ta Edición, 2003. PP 948.

[4] A Practical Handbook of Speech Coders.
Goldberg, R. G. "Frontmatter".
Ed. Randy Goldberg. Boca Raton: CRC Press LLC, 2000. PP 247.

[5] MATLAB Software for the code Excited Linear Prediction algorithm
The Federal Standard 1016.
Karthikeyan N. Ramamurthy, Andreas S. Spanias.
By Morgan & Calypool Publishers, 2da edición. Marzo 2010. PP 110.

[6] 3GPP TR 26.936 V10.0.0 (2011-03)
3rd Generation Partnership Project.
Technical Specification Group Services and System Aspects.
Performance characterization of 3GPP audio códecs. Marzo, 2011. PP 13.

[7] Signal Processing of Speech.
F.J. Owens,
Mac Millan New Electronics, Hong-Kong, 1993. PP 179.

[8] ARIB STD-T63-26.101 V3.3.0.
" Mandatory Speech Codec speech processing functions; AMR speech Codec Frame Structure
3GPP Organizational Partners (Release 1999)".
3GPP TS, Valbonne, France. Marzo2002. PP 20.

[9] Applications of digital signal processing to audio and acoustics.
M ark Kahrs, Karlheinz Brandenburg.
Kluwer Academic Publishers.
New York, Boston, Dordrecht, London, Moscow, 2002. PP 571.

[10] LTE: Nuevas Tendencias en Comunicaciones Móviles.
Ramón Agusti, Francisco Bernardo, Fernando Casadevall, Ramon Ferrús, Jordi Pérez- Romero, Oriol Sallen. © Copyright 2010. Fundación Vodafone España.

[11] B. Atal and S. Hanauer. Speech analysis and synthesis by linear prediction of the speech wave. J. Acoust. Soc. Am. 1971, PP. 637-655.

[12] B.S. Atal and J.R. Remde. A new model of lpc excitation for producing natural-sounding speech at low bit rates. IEEE Int. Conf. Acoust. Sp. Sig. Proc., 1982, PP. 614-617.

[13] Recomendación G.711.1 (09/12). Extensión incorporada de banda ancha para la modulación por impulsos codificados G.711. http://www.itu.int/rec/T-REC-G.711.1/es

[14] Recomendación G.726 (12/90). Modulación por impulsos codificados diferencial adaptativa (ADPCM) a 40, 32, 24, 16 kbit/s. https://www.itu.int/rec/T-REC-G.726-199012-I/es

[15] Recomendación G.722 (09/12). Codificación de audio de 7 kHz dentro de 64 kbit/**s.** https://www.itu.int/rec/T-REC-G.722/es

[16] Recomendación G.722 (07/03). Codificación en banda ancha de voz a unos 16 kbit/s utilizando banda ancha multivelocidad adaptativa. https://www.itu.int/rec/T-REC-G.722.2/es

[17] G.Imp722.2 (10/02). Implementors' Guide for G.722.2: (Wideband coding of speech at around 16 kbit/s using Adaptive Multi-rate Wideband, AMR-WB). https://www.itu.int/rec/T-REC-G.Imp722.2/es

[18] Recomendación G.729 (06/12). Codificación de la voz a 8 kbit/s mediante predicción lineal con excitación por código algebraico de estructura conjugada. https://www.itu.int/rec/T-REC-G.729/es

[19] Recomendación G.722.1 (05/05). Codificación de baja complejidad a 24 y 32 kbit/s para el funcionamiento manos libres en los sistemas con baja pérdida de tramas. http://www.itu.int/rec/T-REC-G.722.1/es

[20] http://www.numerix-dsp.com/appsnotes/APR8-sigma-delta.pdf MOTOROLA.

[21] Unión internacional de telecomunicaciones UIT-T P.862, P.800 y P.830. Métodos de evaluación objetiva y subjetiva de la calidad. http://www.itu.int/rec/T-REC-P.862-200102-I/en http://www.itu.int/rec/T-REC-P.800-199608-I/en http://www.itu.int/rec/T-REC-P.830/en

[22] Software OPNET. http://www.opnet.com/

[23] Matlab http://www.mathworks.com/