



UNIÓN INTERNACIONAL DE TELECOMUNICACIONES

UIT-T

SECTOR DE NORMALIZACIÓN
DE LAS TELECOMUNICACIONES
DE LA UIT

G.722.2

(01/2002)

SERIE G: SISTEMAS Y MEDIOS DE TRANSMISIÓN,
SISTEMAS Y REDES DIGITALES

Equipos terminales digitales – Codificación de señales
analógicas mediante métodos diferentes de la MIC

**Codificación en banda ancha de voz a unos
16 kbit/s utilizando banda ancha multivelocidad
adaptativa**

Recomendación UIT-T G.722.2

RECOMENDACIONES UIT-T DE LA SERIE G
SISTEMAS Y MEDIOS DE TRANSMISIÓN, SISTEMAS Y REDES DIGITALES

CONEXIONES Y CIRCUITOS TELEFÓNICOS INTERNACIONALES	G.100–G.199
CARACTERÍSTICAS GENERALES COMUNES A TODOS LOS SISTEMAS ANALÓGICOS DE PORTADORAS	G.200–G.299
CARACTERÍSTICAS INDIVIDUALES DE LOS SISTEMAS TELEFÓNICOS INTERNACIONALES DE PORTADORAS EN LÍNEAS METÁLICAS	G.300–G.399
CARACTERÍSTICAS GENERALES DE LOS SISTEMAS TELEFÓNICOS INTERNACIONALES EN RADIOENLACES O POR SATÉLITE E INTERCONEXIÓN CON LOS SISTEMAS EN LÍNEAS METÁLICAS	G.400–G.449
COORDINACIÓN DE LA RADIOTELEFONÍA Y LA TELEFONÍA EN LÍNEA	G.450–G.499
EQUIPOS DE PRUEBAS	G.500–G.599
CARACTERÍSTICAS DE LOS MEDIOS DE TRANSMISIÓN	G.600–G.699
EQUIPOS TERMINALES DIGITALES	G.700–G.799
Generalidades	G.700–G.709
Codificación de señales analógicas mediante modulación por impulsos codificados (MIC)	G.710–G.719
Codificación de señales analógicas mediante métodos diferentes de la MIC	G.720–G.729
Características principales de los equipos múltiplex primarios	G.730–G.739
Características principales de los equipos múltiplex de segundo orden	G.740–G.749
Características principales de los equipos múltiplex de orden superior	G.750–G.759
Características principales de los transcodificadores y de los equipos de multiplicación de circuitos digitales	G.760–G.769
Características de operación, administración y mantenimiento de los equipos de transmisión	G.770–G.779
Características principales de los equipos múltiplex de la jerarquía digital síncrona	G.780–G.789
Otros equipos terminales	G.790–G.799
REDES DIGITALES	G.800–G.899
SECCIONES DIGITALES Y SISTEMAS DIGITALES DE LÍNEA	G.900–G.999

Para más información, véase la Lista de Recomendaciones del UIT-T.

Recomendación UIT-T G.722.2

Codificación en banda ancha de voz a unos 16 kbit/s utilizando banda ancha multivelocidad adaptativa

Resumen

Esta Recomendación describe un codificador y decodificador de banda ancha multivelocidad adaptativa (AMR-WB) de alta calidad destinado principalmente a las señales vocales de 7 kHz de anchura de banda. El AMR-WB funciona a numerosas velocidades binarias, que van de 6,6 kbit/s a 23,85 kbit/s. La velocidad binaria puede cambiarse en cualquier frontera de trama de 20 ms.

El anexo C a esta Recomendación incluye un lote integrado de soporte lógico del código fuente C que contiene la aplicación del codificador y decodificador de la presente Recomendación y de sus anexos A y B y apéndice I. En el anexo D se da un conjunto de vectores de prueba digitales destinado a los diseñadores. Dichos vectores constituyen la herramienta que proporciona una indicación del éxito en la implementación del códec.

El códec AMR-WB G.722.2 es el mismo que el AMR-WB 3GPP. Las especificaciones 3GPP correspondientes son la TS 26.190 para el códec de voz y la TS 26.194 para el detector de actividad vocal.

Orígenes

La Recomendación UIT-T G.722.2, preparada por la Comisión de Estudio 16 (2001-2004) del UIT-T, fue aprobada por el procedimiento de la Resolución 1 de la AMNT el 13 de enero de 2002.

PREFACIO

La UIT (Unión Internacional de Telecomunicaciones) es el organismo especializado de las Naciones Unidas en el campo de las telecomunicaciones. El UIT-T (Sector de Normalización de las Telecomunicaciones de la UIT) es un órgano permanente de la UIT. Este órgano 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.

La Asamblea Mundial de Normalización de las Telecomunicaciones (AMNT), que se celebra cada cuatro años, establece los temas que han de estudiar las Comisiones de Estudio del UIT-T, que a su vez producen Recomendaciones sobre dichos temas.

La aprobación de Recomendaciones por los Miembros del UIT-T es el objeto del procedimiento establecido en la Resolución 1 de la AMNT.

En ciertos sectores de la tecnología de la información que corresponden a la esfera de competencia del UIT-T, se preparan las normas necesarias en colaboración con la ISO y la CEI.

NOTA

En esta Recomendación, la expresión "Administración" se utiliza para designar, en forma abreviada, tanto una administración de telecomunicaciones como una empresa de explotación reconocida de telecomunicaciones.

PROPIEDAD INTELECTUAL

La UIT señala a la atención la posibilidad de que la utilización o aplicación de la presente Recomendación suponga el empleo de un derecho de propiedad intelectual reivindicado. La UIT no adopta ninguna posición en cuanto a la demostración, validez o aplicabilidad de los derechos de propiedad intelectual reivindicados, ya sea por los miembros de la UIT o por terceros ajenos al proceso de elaboración de Recomendaciones.

En la fecha de aprobación de la presente Recomendación, la UIT ha recibido notificación de propiedad intelectual, protegida por patente, que puede ser necesaria para aplicar esta Recomendación. Sin embargo, debe señalarse a los usuarios que puede que esta información no se encuentre totalmente actualizada al respecto, por lo que se les insta encarecidamente a consultar la base de datos sobre patentes de la TSB.

© UIT 2002

Reservados todos los derechos. Ninguna parte de esta publicación puede reproducirse por ningún procedimiento sin previa autorización escrita por parte de la UIT.

ÍNDICE

Página

1	Alcance	1
2	Referencias normativas.....	2
3	Definiciones, símbolos y abreviaturas.....	2
3.1	Definiciones.....	2
3.2	Símbolos	4
3.3	Abreviaturas	7
4	Descripción resumida	8
4.1	Descripción funcional de las partes audio	8
4.2	Preparación de muestras de voz	9
4.3	Principios del codificador de banda ancha multivelocidad adaptativa para voz	9
4.4	Principios del decodificador de multivelocidad adaptativa para voz.....	15
4.5	Secuencia e importancia subjetiva de los parámetros codificados.....	17
5	Descripción funcional del codificador.....	17
5.1	Preprocesamiento	17
5.2	Análisis y cuantificación de la predicción lineal.....	17
5.2.1	Presentación en ventana y cálculo de la autocorrelación	18
5.2.2	Algoritmo Levinson-Durbin.....	18
5.2.3	Conversión LP a ISP	19
5.2.4	Conversión ISP a LP	20
5.2.5	Cuantificación de los coeficientes ISP	21
5.2.6	Interpolación de los ISP.....	23
5.3	Ponderación perceptual	23
5.4	Análisis de la altura de tono en bucle abierto.....	23
5.4.1	Modo 6,60 kbit/s.....	24
5.4.2	Modos 8,85, 12,65, 14,25, 15,85, 18,25, 19,85, 23,05 y 23,85 kbit/s.....	25
5.5	Cálculo de la respuesta en tren de impulsos.....	25
5.6	Cálculo de la señal objetivo.....	26
5.7	Libro de códigos adaptativo	26
5.8	Libro de códigos algebraico	28
5.8.1	Estructura de libro de códigos	28
5.8.2	Indexación de impulsos	32
5.8.3	Búsqueda en libro de códigos.....	36
5.9	Cuantificación de las ganancias de libros de códigos adaptativo y fijo	40
5.10	Actualización de la memoria	41
5.11	Generación de ganancia de banda alta.....	41

	Página
6	Descripción funcional del decodificador 42
6.1	Decodificación y síntesis de la voz 42
6.2	Filtrado paso alto, escalaje en sentido ascendente e interpolación..... 45
6.3	Banda de altas frecuencias..... 45
6.3.1	Generación de excitación de banda alta 46
6.3.2	Filtro LP para la banda de altas frecuencias 46
6.3.3	Síntesis de banda alta..... 47
7	Atribución detallada de bits del códec de banda ancha multivelocidad adaptativa..... 47
8	Secuencias del retorno a la posición inicial 56
8.1	Descripción funcional..... 56
8.2	Definiciones..... 57
8.3	Retorno a la posición inicial del codificador 57
8.4	Retorno a la posición inicial del decodificador 58
9	Detector de actividad vocal (VAD) 58
9.1	Símbolos VAD 58
9.1.1	Variables VAD 58
9.1.2	Constantes VAD 59
9.1.3	Funciones..... 60
9.2	Descripción funcional..... 60
9.2.1	Banco de filtros y cálculo de niveles de subbanda 61
9.2.2	Detección de tono 63
9.2.3	Decisión VAD 63
10	Bibliografía 68

Recomendación UIT-T G.722.2

Codificación en banda ancha de voz a unos 16 kbit/s utilizando banda ancha multivelocidad adaptativa

1 Alcance

Esta Recomendación describe de forma detallada el establecimiento de la correspondencia entre bloques de entrada de 320 muestras de voz en formato MIC uniforme de 16 bits y bloques codificados de 132, 177, 253, 285, 317, 365, 397, 461 y 477 bits y entre bloques codificados de 132, 177, 253, 285, 317, 365, 397, 461 y 477 bits y bloques de salida de 320 muestras de voz reconstruidas. La velocidad de muestreo es de 16 000 muestras/s lo que da lugar a una velocidad binaria del tren binario codificado de 6,60, 8,85, 12,65, 14,25, 15,85, 18,25, 19,85, 23,05 o 23,85 kbit/s. El esquema de codificación de los modos de codificación multivelocidad es el llamado codificador de predicción lineal excitado por código algebraico, al que en adelante se hará referencia como ACELP. El codificador ACELP de banda ancha multivelocidad se denomina AMR-WB. El códec descrito en esta Recomendación utiliza además un detector de actividad vocal (VAD, *voice activity detector*) integrado.

Las aplicaciones previstas de la presente Recomendación son las siguientes: protocolo de transmisión de la voz sobre el protocolo Internet (VoIP, *voice over Internet protocol*) y aplicaciones Internet, comunicaciones móviles, aplicaciones RTPC, telefonía de banda ancha RDSI y videotelefonía y videoconferencia RDSI.

Además del algoritmo especificado en el cuerpo principal de la Recomendación G.722.2, los anexos A y B y el apéndice I proporcionan funcionalidades suplementarias que permiten la interoperabilidad con sistemas inalámbricos del sistema global para comunicaciones móviles (GSM, *global system for mobile communications*) y del proyecto de asociación de tercera generación (3GPP, *third generation partnership project*). Esas funcionalidades se concibieron en principio para estos sistemas, pero su utilización no está limitada a aplicaciones móviles. En otros dos anexos, el D y el E, se describen los vectores de prueba y la estructura de trama, respectivamente. Estos anexos se pueden implementar con independencia del cuerpo principal de la presente Recomendación, de acuerdo con los distintos requisitos de los sistemas que emplean el algoritmo AMR-WB:

- El anexo A describe aspectos del ruido de confort para la utilización del algoritmo AMR-WB en funcionamiento a velocidad controlada por la fuente. La implementación de este anexo es esencial para la interoperabilidad con sistemas inalámbricos GSM y 3GPP.
- El anexo B describe el funcionamiento a velocidad controlada por la fuente para el algoritmo AMR-WB. La implementación de este anexo es esencial para la interoperabilidad con sistemas inalámbricos GSM y 3GPP:
- El anexo D describe las secuencias de prueba digitales, que constituyen una herramienta de verificación indicadora del éxito en la implementación del códec AMR-WB.
- El anexo E describe la estructura de trama que se recomienda utilizar con los diferentes modos de funcionamiento del algoritmo AMR-WB.
- El apéndice I da un ejemplo de cómo ocultar los errores de tramas AMR-WB erróneas o perdidas.

Para facilitar su utilización, los códigos C ANSI con la descripción de bajo nivel de todas estas funcionalidades se han agrupado en un solo anexo, el anexo C. Si se produce alguna discrepancia entre cualquiera de las diferentes partes de la presente Recomendación y la implementación de esas descripciones del anexo C, prevalecerán las descripciones del anexo C.

En la cláusula 8 se describe un procedimiento de reinicio específico, llamado de retorno a la posición inicial. Se trata de una característica muy útil que sirve para llevar el códec a un estado inicial conocido, por ejemplo, a efectos de prueba). La cláusula 9 especifica el detector de actividad vocal (VAD) utilizado en este códec así como en el funcionamiento a velocidad controlada por la fuente (DTX) que se especifica en el anexo B.

2 Referencias normativas

Las siguientes Recomendaciones del UIT-T y otras referencias contienen disposiciones que, mediante su referencia en este texto, constituyen disposiciones de la presente Recomendación. Al efectuar esta publicación, estaban en vigor las ediciones indicadas. Todas las Recomendaciones y otras referencias son objeto de revisiones por lo que se preconiza que los usuarios de esta Recomendación investiguen la posibilidad de aplicar las ediciones más recientes de las Recomendaciones y otras referencias citadas a continuación. Se publica periódicamente una lista de las Recomendaciones UIT-T actualmente vigentes.

- [1] Recomendación UIT-T G.711 (1988), *Modulación por impulsos codificados (MIC) de frecuencias vocales*.

3 Definiciones, símbolos y abreviaturas

3.1 Definiciones

En esta Recomendación se definen los términos siguientes.

3.1.1 libro de códigos adaptativo: El libro de códigos adaptativo contiene los vectores de excitación adaptados a cada subtrama. El libro de códigos adaptativo se obtiene a partir del estado del filtro a largo plazo. El valor del retardo puede considerarse como un índice libro de códigos adaptativo.

3.1.2 libro de códigos algebraico: Libro de códigos fijo en el que se utiliza un código algebraico para poblar los vectores de excitación (vectores de renovación). La excitación contiene un pequeño número de impulsos distintos de cero con conjuntos entrelazados predefinidos de posiciones posibles. Las amplitudes y las posiciones de los impulsos del k-ésimo vector-código de excitación se pueden derivar del índice k aplicando una regla que no exige almacenamiento físico, ni siquiera mínimo, al contrario de lo que ocurre con los libros de códigos estocásticos en donde el trayecto desde el índice al vector-código implica la existencia de cuadros de consulta.

3.1.3 procedimiento antidispersión: Procedimiento de postprocesamiento adaptativo aplicado al vector de libro de códigos fijo para reducir los artefactos perceptuales de un vector de libro de códigos fijo disperso.

3.1.4 análisis de altura de tono en bucle cerrado: Se trata de la búsqueda en el libro de códigos adaptativo, es decir, un proceso de estimación del valor de la altura de tono (retardo) a partir de la voz de entrada ponderada y el estado del filtro a largo plazo. En la búsqueda en bucle cerrado, el retardo se indaga utilizando un bucle de minimización del error (análisis por síntesis). En el códec de banda ancha multivelocidad adaptativa, la búsqueda de la altura de tono en bucle cerrado se efectúa para cada subtrama.

3.1.5 coeficientes de forma directa: Es uno de los formatos de almacenamiento de parámetros de filtro a corto plazo. En el códec de banda ancha multivelocidad adaptativa, todos los filtros que se emplean para modificar muestras de voz utilizan coeficientes de forma directa.

3.1.6 libro de códigos fijo: El libro de códigos fijo contiene los vectores de excitación de los filtros de síntesis de la voz. El contenido del libro de códigos es no adaptativo (es decir, fijo). En el

códec de banda ancha multivelocidad adaptativa, el libro de códigos fijo se implementa utilizando un libro de códigos algebraico.

3.1.7 retardos fraccionarios: Conjunto de valores de retardo que tienen resolución de submuestra. En el códec de banda ancha multivelocidad adaptativa se utiliza una resolución de submuestra de 1/4 o 1/2 de muestra.

3.1.8 trama: Intervalo de tiempo igual a 2 ms (320 muestras a una velocidad de muestreo de 16 kHz).

3.1.9 frecuencias espectrales de inmitancia: (véase par espectral de inmitancia).

3.1.10 par espectral de inmitancia: Transformación de parámetros LPC. Los pares espectrales de inmitancia se obtienen descomponiendo la función de transferencia de filtro inverso $A(z)$ en un conjunto de dos funciones de transferencia, una de ellas con simetría par y la otra con simetría impar. Los pares espectrales de inmitancia (llamados también frecuencias espectrales de inmitancia) son las raíces de los polinomios del círculo unidad z .

3.1.11 retardos enteros: Conjunto de valores de retardo con resolución de muestra total.

3.1.12 filtro de interpolación: Filtro FIR utilizado para obtener una estimación de las muestras con resolución de submuestra, dada una entrada muestreada con resolución de muestra entera. En esta implementación, el filtro de interpolación tiene características de filtro paso bajo. El libro de códigos adaptativo consta por tanto de la excitación pasada interpolada con filtrado paso bajo.

3.1.13 filtro inverso: Filtro que elimina la correlación a corto plazo de la señal vocal. Este filtro modela una respuesta en frecuencia inversa del aparato vocal.

3.1.14 retardo: Demora del filtro a largo plazo. Este es normalmente el periodo de altura de tono verdadero, o su múltiplo o submúltiplo.

3.1.15 ventana de análisis de predicción lineal: Los coeficientes de filtro a corto plazo de cada trama se calculan utilizando las muestras de voz filtradas paso alto dentro de la ventana de análisis. En el códec de banda ancha multivelocidad adaptativa, la longitud de la ventana de análisis es siempre de 384 muestras. Para todos los modos se utiliza una sola ventana asimétrica con la que se genera un conjunto único de coeficientes LP. En el análisis se emplea indagación de 5 ms.

3.1.16 coeficientes de predicción lineal: Coeficientes de predicción lineal (LP), a los que también se denomina coeficientes predictivos lineales (LPC). Es un término descriptivo genérico que se refiere a los coeficientes de filtro a corto plazo.

3.1.17 modo: Cuando se utiliza solo se refiere al modo de códec fuente, es decir, a uno de los códecs fuente empleados en el códec AMR-WB.

3.1.18 búsqueda de altura de tono en bucle abierto: Es un proceso de estimación del valor de retardo cercano al óptimo directamente a partir de la entrada de voz ponderada. Se hace así para simplificar el análisis de la altura de tono y limitar la búsqueda de la altura de tono en bucle cerrado a un número pequeño de retardos en torno a los retardos estimados en bucle abierto. En el códec de banda ancha multivelocidad adaptativa se lleva a cabo una búsqueda de la altura de tono en bucle cerrado en subtramas alternas.

3.1.19 residuo: Señal de salida resultante de una operación de filtrado inverso.

3.1.20 filtro de síntesis a corto plazo: Este filtro introduce en la señal de excitación una correlación a corto plazo que modela la respuesta en tren de impulsos del aparato vocal.

3.1.21 filtro de ponderación perceptual: Este filtro se emplea en la búsqueda de análisis por síntesis de los libros de códigos. El filtro aprovecha las propiedades que tienen los conformadores

de enmascarar el ruido (resonancias del aparato vocal) ponderando el error, menos en regiones cercanas a las frecuencias del conformador y más en regiones alejadas de ellas.

3.1.22 subtrama: Intervalo de tiempo igual a 5 ms (80 muestras a una velocidad de muestreo de 16 kHz).

3.1.23 cuantificación de vector: Método de agrupación de varios parámetros y cuantificación simultánea de los mismos.

3.1.24 respuesta de entrada cero: Es la salida de un filtro debida a entradas anteriores, es decir, debida al estado presente del filtro, dado que se aplica una entrada de ceros.

3.1.25 respuesta de estado cero: Es la salida de un filtro debida a la entrada presente, dado que no han sido aplicadas entradas anteriores, es decir, porque la información de estado en el filtro es todo ceros.

3.2 Símbolos

Para los fines de esta Recomendación, se aplican los siguientes símbolos:

$A(z)$	Filtro inverso con coeficientes no cuantificados
$\hat{A}(z)$	Filtro inverso con coeficientes cuantificados
$H(z) = \frac{1}{\hat{A}(z)}$	Filtro de síntesis de la voz con coeficientes cuantificados
a_i	Parámetros de predicción lineal no cuantificados (coeficientes de forma directa)
\hat{a}_i	Parámetros de predicción lineal cuantificados
m	Orden del modelo LP
$W(z)$	Filtro de ponderación perceptual (coeficientes no cuantificados)
γ_1	Factor de ponderación perceptual
T	Retardo de altura de tono entero más cercano al retardo de altura de tono fraccionario en bucle cerrado
β	Coefficiente de prefiltro adaptativo (ganancia de altura de tono cuantificada)
$H_{h1}(z)$	Filtro paso alto de preprocesamiento
$w(n)$	Ventana de análisis LP
L_1	Longitud de la primera parte de la ventana de análisis LP $w(n)$
L_2	Longitud de la segunda parte de la ventana de análisis LP $w(n)$
$r(k)$	Autocorrelaciones de la palabra presentada en ventana $s'(n)$
$w_{lag}(i)$	Ventana de retardo para las autocorrelaciones (ampliación de la anchura de banda de 60 Hz)
f_0	Ampliación de la anchura de banda en Hz
f_s	Frecuencia de muestreo en Hz
$r'(k)$	Autocorrelaciones modificadas (anchura de banda ampliada)
$E(i)$	Error de predicción en la iteración i -ésima del algoritmo Levinson

k_i	Coefficiente de reflexión i -ésimo
$a_j^{(i)}$	Coefficiente de forma directa j -ésimo en la iteración i -ésima del algoritmo Levinson
$F_1'(z)$	Polinomio ISF simétrico
$F_2'(z)$	Polinomio ISF antisimétrico
$F_1(z)$	Polinomio $F_1'(z)$
$F_2(z)$	Polinomio $F_2'(z)$ con las raíces $z = 1$ y $z = -1$ eliminadas
q_i	Pares espectrales de inmitancia (ISP, <i>immittance spectral pairs</i>) en el dominio del coseno
q	Vector ISP en el dominio del coseno
$\hat{\mathbf{q}}_i^{(n)}$	Vector ISP cuantificado en la subtrama i -ésima de la trama n
ω_i	Frecuencias espectrales de inmitancia (ISF, <i>immittance spectral frequencies</i>)
$T_m(x)$	Polinomio de Chebyshev de orden m -ésimo
$f_1(i), f_2(i)$	Coefficientes de los polinomios $F_1(z)$ y $F_2(z)$
$f_1'(i), f_2'(i)$	Coefficientes de los polinomios $F_1'(z)$ y $F_2'(z)$
$f(i)$	Coefficientes de $F_1(z)$ o $F_2(z)$
$C(x)$	Suma polinómica de los polinomios de Chebyshev
x	Coseno de la frecuencia angular ω
λ_k	Coefficientes de recursión para la evaluación del polinomio de Chebyshev
f_i	Frecuencias espectrales de inmitancia (ISF) en Hz
$\mathbf{f}^t = [f_1 f_2 \dots f_{16}]$	Representación vectorial de las ISF en Hz
$\mathbf{z}(n)$	Vector ISF con el valor medio eliminado en la trama n
$\mathbf{r}(n)$	Vector ISF residual de predicción en la trama n
$\mathbf{p}(n)$	Vector ISF predicho en la trama n
$\hat{\mathbf{r}}(n-1)$	Vector residual cuantificado en la trama pasada
$\hat{\mathbf{r}}_i^k$	Subvector ISF cuantificado i correspondiente al índice de cuantificación k
d_i	Distancia entre las frecuencias espectrales de inmitancia f_{i+1} y f_{i-1}
$h(n)$	Respuesta en tren de impulsos del filtro de síntesis ponderada
$H(z)W(z)$	Filtro de síntesis ponderada
T_1	Entero más cercano al retardo de altura de tono fraccionario de la subtrama previa (1^a o 3^a)
$s'(n)$	Señal vocal presentada en ventana
$s_w(n)$	Señal vocal ponderada

$\hat{s}(n)$	Señal vocal reconstruida
$x(n)$	Señal objetivo de la búsqueda en libro de códigos adaptativo
$x_2(n), \mathbf{x}_2^t$	Señal objetivo de la búsqueda en libro de códigos algebraico
$res_{LP}(n)$	Señal residual LP
$c(n)$	Vector de libro de códigos fijo
$v(n)$	Vector de libro de códigos adaptativo
$y(n) = v(n)*h(n)$	Vector de libro de códigos adaptativo filtrado
$y_k(n)$	Excitación filtrada pasada
$u(n)$	Señal de excitación
$\hat{u}'(n)$	Señal de excitación enfatizada (resaltada) con ganancia escalada
T_{op}	Mejor retardo en bucle abierto
t_{min}	Valor de búsqueda de retardo mínimo
t_{max}	Valor de búsqueda de retardo máximo
$R(k)$	Término de correlación que se ha de maximizar en la búsqueda en libro de códigos adaptativo
$R(k)_t$	Valor interpolado de $R(k)$ para el periodo de demora entero k y la fracción t
A_k	Término de correlación que se ha de maximizar en la búsqueda en libro de códigos algebraico correspondiente al índice k
C_k	Correlación del numerador de A_k correspondiente al índice k
E_{Dk}	Energía del denominador de A_k correspondiente al índice k
$\mathbf{d} = \mathbf{H}^t \mathbf{x}_2$	Correlación entre la señal objetivo $x_2(n)$ y la respuesta en tren de impulsos $h(n)$, es decir, objetivo filtrado hacia atrás
\mathbf{H}	Matriz de convolución Toeplitz triangular inferior con diagonal $h(0)$ y diagonales inferiores $h(1), \dots, h(63)$
$\Phi = \mathbf{H}^t \mathbf{H}$	Matriz de correlaciones de $h(n)$
$d(n)$	Elementos del vector \mathbf{d}
$\phi(i, j)$	Elementos de la matriz simétrica Φ
c_k	Vector de renovación
C	Correlación del numerador de A_k
m_i	Posición del impulso i -ésimo
ϑ_i	Amplitud del impulso i -ésimo
N_p	Número de impulsos en la excitación de libro de códigos fijo
E_D	Energía del denominador de A_k
$res_{LTP}(n)$	Residuo de la predicción a largo plazo normalizado
$b(n)$	Señal utilizada para prefijar los signos en la búsqueda en libro de código algebraico

$s_b(n)$	Señal de signo para la búsqueda en libro de códigos algebraico
$d'(n)$	Objetivo filtrado hacia atrás con ampliación de signo
$\phi'(i, j)$	Elementos modificados de la matriz Φ , incluyendo información de signo
$z^l, z(n)$	Vector de libro de códigos fijo convolucionado con $h(n)$
$E(n)$	Energía de renovación con valor medio eliminado (en dB)
\bar{E}	Valor medio de la energía de renovación
$\tilde{E}(n)$	Energía predicha
$[b_1 b_2 b_3 b_4]$	Coefficientes de predicción MA
$\hat{R}(k)$	Error de predicción cuantificado correspondiente a la subtrama k
E_I	Energía de renovación media
$R(n)$	Error de predicción de la cuantificación de ganancia de libro de códigos fijo
E_Q	Error de cuantificación de la cuantificación de ganancia de libro de códigos fijo
$e(n)$	Estados del filtro de síntesis $1/\hat{A}(z)$
$e_w(n)$	Error ponderado perceptualmente de la búsqueda de análisis por síntesis
η	Factor de escalaje de ganancia para la excitación enfatizada
g_c	Ganancia de libro de códigos fijo
g'_c	Ganancia de libro de códigos fijo predicha
\hat{g}_c	Ganancia de libro de códigos fijo cuantificada
g_p	Ganancia de libro de códigos adaptativo
\hat{g}_p	Ganancia de libro de códigos adaptativo cuantificada
$\gamma_{gc} = g_c/g'_c$	Factor de corrección entre la ganancia g_c y la ganancia estimada g'_c
$\hat{\gamma}_{gc}$	Valor óptimo de γ_{gc}
γ_{sc}	Factor de escalaje de ganancia

3.3 Abreviaturas

En esta Recomendación se utilizan las siguientes siglas:

ACELP	Predicción lineal con excitación por código algebraico (<i>algebraic code excited linear prediction</i>)
AGC	Control de ganancia adaptativa (<i>adaptive gain control</i>)
AMR	Multivelocidad adaptativa (<i>adaptive multirate</i>)
AMR-WB	Banda ancha multivelocidad adaptativa (<i>adaptive multi-rate Wideband</i>)
CELP	Predicción lineal con excitación por código (<i>excited linear prediction</i>)
FIR	Respuesta en tren de impulsos finitos (<i>finite impulse response</i>)
ISF	Frecuencia espectral de inmitancia (<i>immittance spectral frequency</i>)

ISP	Par espectral de inmitancia (<i>immittance spectral pair</i>)
ISPP	Permutación de impulsos individuales entrelazados (<i>interleaved single-pulse permutation</i>)
LP	Predicción lineal (<i>linear prediction</i>)
LPC	Codificación predictiva lineal (<i>linear predictive coding</i>)
LTP	Predictor a largo plazo (o predicción a largo plazo) [<i>long-term predictor (or long term prediction)</i>]
MA	Media móvil (<i>moving average</i>)
S-MSVQ	Cuantificación de vector multietapa dividida (<i>split-multi-stage vector quantization</i>)
WB	Anchura de banda (<i>wideband</i>)

4 Descripción resumida

La presente Recomendación está estructurada como sigue:

- La cláusula 4.1 contiene una descripción funcional de las partes audio de las funciones de conversión analógico/digital (A/D) y digital/analógico (D/A).
- La cláusula 4.2 describe el formato de entrada del codificador AMR-WB y el formato de salida del decodificador AMR-WB.
- Las cláusulas 4.3 y 4.4 presentan una descripción simplificada de los principios del proceso de codificación y decodificación, respectivamente, del códec AMR-WB.
- En la cláusula 4.5 se da la secuencia y la importancia subjetiva de los parámetros codificados.
- La cláusula 5 presenta la descripción funcional de la codificación del códec AMR-WB, mientras que la cláusula 6 describe los procedimientos de decodificación.
- En la cláusula 7 se presenta de forma tabulada la atribución detallada de bits del códec AMR-WB.
- La cláusula 8 describe la operación de retorno a la posición inicial.

4.1 Descripción funcional de las partes audio

La conversión analógico a digital y digital a analógico comprenderá en principio los siguientes elementos:

- 1) MIC analógico a digital uniforme [1]
 - micrófono;
 - dispositivo de ajuste del nivel de entrada;
 - filtro de entrada para la eliminación de la dispersión por superposición de bandas laterales;
 - dispositivo de retención de muestras que muestrea a 16 kHz;
 - conversión analógico a digital uniforme en representación de 14 bits.

El formato uniforme se representará en complemento de 2.
- 2) MIC digital uniforme a analógico [1]
 - conversión de MIC uniforme de 14 bits/16 kHz a analógico;
 - dispositivo de retención;

- filtro de reconstrucción que incluye la corrección $x/\text{sen}(x)$;
- dispositivo de ajuste del nivel de salida;
- auricular o altavoz.

En el equipo terminal, la función A/D se puede efectuar:

- por conversión directa a formato MIC uniforme de 14 bits;

En la función D/A tiene lugar la operación inversa.

4.2 Preparación de muestras de voz

En el codificador se introducen datos que comprenden muestras con una resolución de 14 bits justificados a la izquierda en una palabra de 16 bits. El decodificador produce como salida datos con el mismo formato. Fuera del códec de voz se debe aplicar un procesamiento ulterior si los datos del tráfico aparecen con una representación diferente.

4.3 Principios del codificador de banda ancha multivelocidad adaptativa para voz

El códec AMR-WB consta de nueve códec fuente con velocidades binarias de 23,85, 23,05, 19,85, 18,25, 15,85, 14,25, 12,65, 8,85 y 6,60 kbit/s.

El códec se basa en el modelo de codificación con predicción lineal con excitación por código (CELP, *excited linear prediction*). La señal de entrada es preenfatzada utilizando el filtro $H_{\text{pre-emph}}(z) = 1 - \mu z^{-1}$. A continuación se aplica el modelo CELP a la señal preenfatzada. Se utiliza predicción lineal (LP) de 16 orden, o un filtro de síntesis a corto plazo, que viene dado por:

$$H(z) = \frac{1}{\hat{A}(z)} = \frac{1}{1 + \sum_{i=1}^m \hat{a}_i z^{-i}} \quad (1)$$

donde $\hat{a}_i, i = 1, \dots, m$ son los parámetros (cuantificados) de la predicción lineal (LP), y $m = 16$ es el orden del predictor. El filtro de síntesis a largo plazo o de altura de tono, viene dado normalmente por:

$$\frac{1}{B(z)} = \frac{1}{1 - g_p z^{-T}} \quad (2)$$

donde T es el retardo de la altura de tono y g_p es la ganancia de la altura de tono. El filtro de síntesis de altura de tono se implementa utilizando el procedimiento llamado de libro de códigos adaptativo.

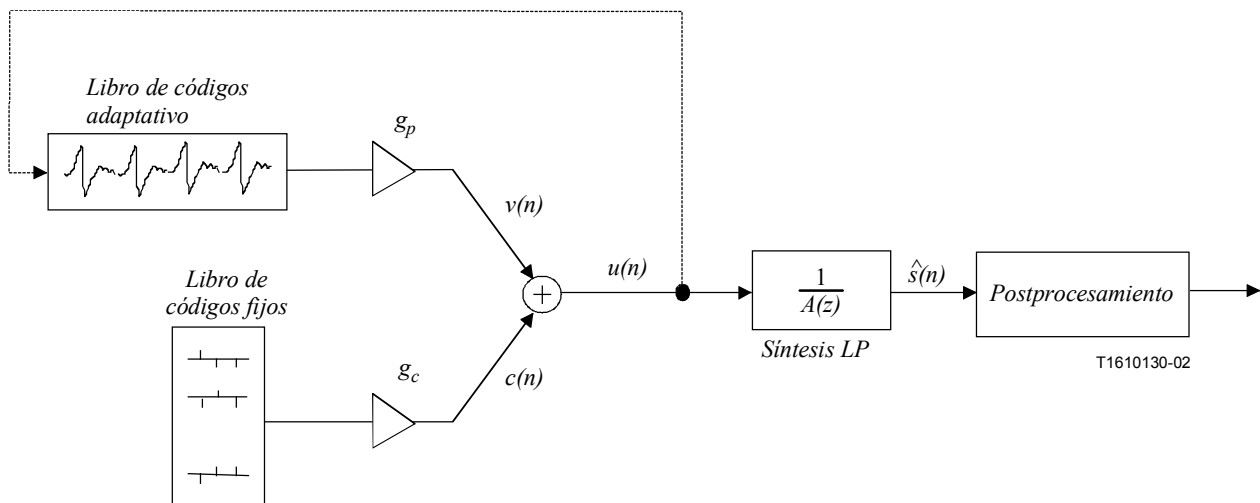


Figura 1/G.722.2 – Diagrama de bloques simplificado del modelo de síntesis CELP

En la figura 1 se muestra el modelo de síntesis de voz CELP. En este modelo, la señal de excitación a la entrada del filtro de síntesis LP a corto plazo se construye añadiendo dos vectores de excitación procedentes de los libros de códigos adaptativo y fijo (renovador). La voz se sintetiza introduciendo los dos vectores convenientemente elegidos de esos libros de códigos a través del filtro de síntesis a corto plazo. La secuencia de excitación óptima de un libro de códigos se elige aplicando un procedimiento de búsqueda de análisis por síntesis en el que el error entre la voz original y la sintetizada se minimiza de acuerdo con una medida de la distorsión ponderada perceptualmente.

El filtro de ponderación perceptual utilizado en la técnica de búsqueda de análisis por síntesis viene dado por:

$$W(z) = A(z/\gamma_1)H_{\text{de-emph}}(z) \quad (3)$$

donde $A(z)$ es el filtro LP no cuantificado, $H_{\text{de-emph}} = \frac{1}{1-0,68z^{-1}}$, y $\gamma_1=0,92$ es el factor de ponderación perceptual. El filtro de ponderación utiliza los parámetros LP no cuantificados.

El codificador efectúa el análisis de los parámetros LPC, LTP y del libro de códigos fijo a una velocidad de muestreo de 12,8 kHz. El codificador funciona en tramas de voz de 20 ms. En cada trama se analiza la señal vocal para extraer los parámetros del modelo CELP (coeficientes de filtro LP, índices de libros de códigos adaptativo y fijo y ganancias). En el modo 23,85 kbit/s se calculan, además de estos parámetros, los índices de ganancia de banda alta. Estos parámetros son codificados y transmitidos. En el decodificador, los parámetros se decodifican y la voz se sintetiza filtrando la señal de excitación reconstruida a través del filtro de síntesis LP.

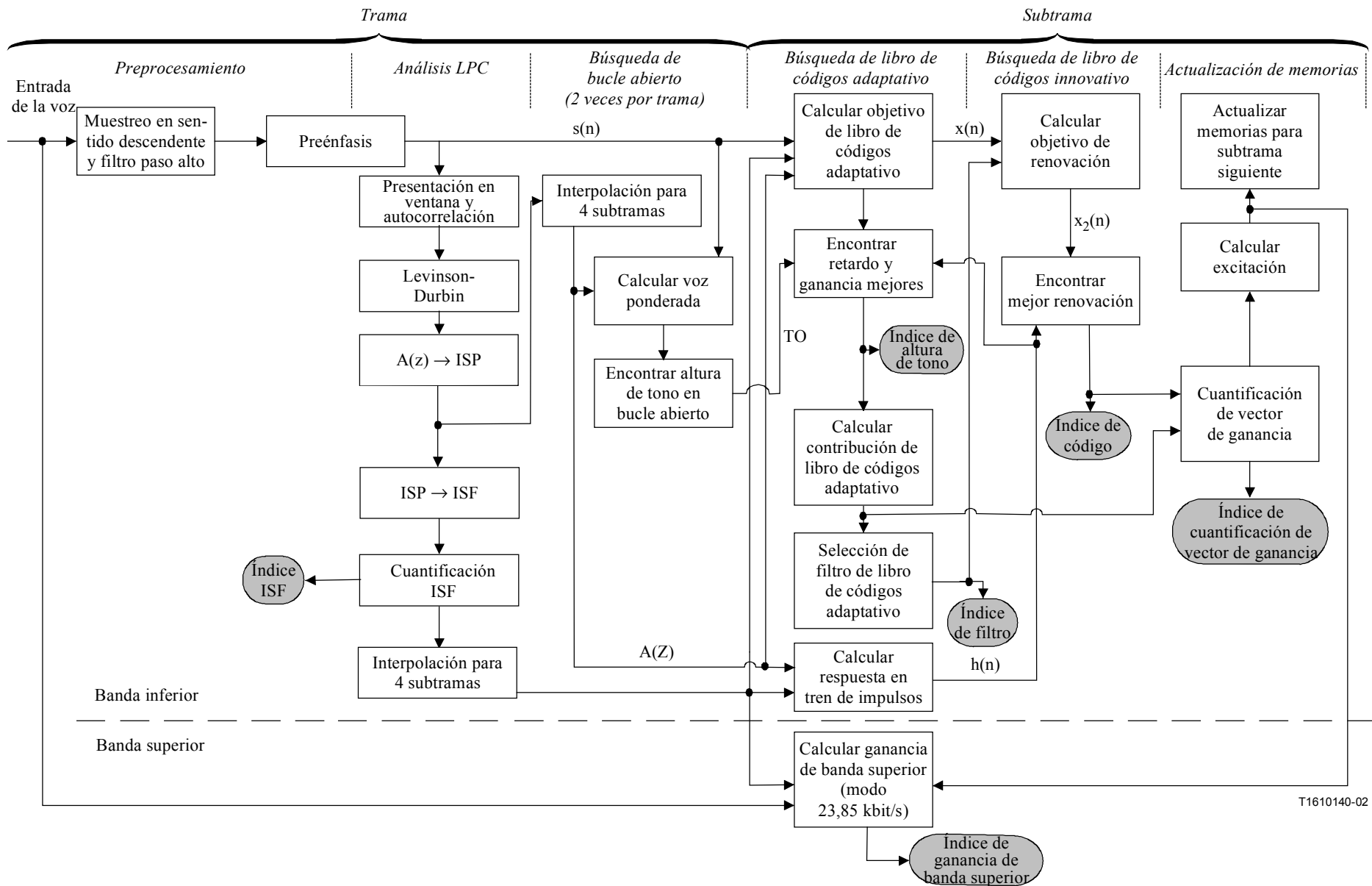


Figura 2/G.722.2–Diagrama de bloques detallado del codificador ACELP

En la figura 2 se muestra el flujo de la señal por el codificador. Tras efectuar una decimación, se lleva a cabo el filtrado paso alto con preénfasis. El análisis LP se hace una vez por trama. El conjunto de parámetros LP se convierte en pares espectrales de inmitancia (ISP) y el vector se cuantifica aplicando la cuantificación de vector multietapa dividida (S-MSVQ, *split-multi-stage vector quantization*) La trama de voz es dividida en 4 subtramas de 5 ms cada una (64 muestras a una velocidad de muestreo de 12,8 kHz). Los parámetros de libro de códigos adaptativo y fijo son transmitidos en todas las subtramas. La utilización de los parámetros LP cuantificados y no cuantificados o de sus versiones interpoladas depende de la subtrama. Se hace una estimación del retardo de la altura de tono en bucle abierto en subtramas alternas o una vez por trama de acuerdo con la señal ponderada perceptualmente.

A continuación se repiten las operaciones siguientes para cada subtrama:

- Se calcula la señal objetivo $x(n)$ filtrando el residuo LP a través del filtro de síntesis ponderada $W(z)H(z)$ con los estados iniciales de los filtros que han sido actualizados filtrando el error entre residuo LP y excitación (lo que equivale al procedimiento común de sustracción de la respuesta de entrada cero del filtro de síntesis ponderada de la señal vocal ponderada).
- Se calcula la respuesta en tren de impulsos, $h(n)$, del filtro de síntesis ponderada.
- Se lleva a cabo el análisis de la altura de tono en bucle cerrado (para encontrar el retardo y la ganancia de la altura de tono), con la señal objetivo $x(n)$ y la respuesta en tren de impulsos $h(n)$, buscando en torno al retardo de altura de tono en bucle abierto. Se utiliza una altura de tono fraccionaria con resolución de 1/4 o 1/2 de muestra (dependiendo del modo y del valor del retardo de la altura de tono). El filtro de interpolación de la búsqueda de altura de tono fraccionaria tiene una respuesta en frecuencia de paso bajo. Además, hay dos posibles características de paso bajo en el libro de códigos adaptativo, y esta información se codifica con 1 bit.
- La señal objetivo $x(n)$ se actualiza eliminando la contribución del libro de códigos adaptativo (vector-código adaptativo filtrado), y esta nueva señal objetivo, $x_2(n)$, se utiliza en la búsqueda en libro de códigos algebraico fijo (para encontrar la renovación óptima).
- Las ganancias de libros de códigos adaptativo y fijo se representan por un vector cuantificado de 6 o 7 bits [aplicando la predicción de media móvil (MA, *moving average*) a la ganancia de libro de códigos fijo].
- Por último, se actualizan las memorias de filtro (utilizando la señal de excitación determinada) para encontrar la señal objetivo en la subtrama siguiente.

En el cuadro 1 se muestra la atribución de bits de los modos del códec AMR-WB. En cada trama de voz de 20 ms se producen 132, 177, 253, 285, 317, 365, 397, 461 y 477 bits correspondientes a una velocidad binaria de 6,60, 8,85, 12,65, 14,25, 15,85, 18,25, 19,85, 23,05 o 23,85 kbit/s. En los cuadros 12a a 12i se da la atribución de bits entre los parámetros de códec de forma más detallada. Se señala que los bits más significativos (MSB, *most significant bits*) se envían siempre los primeros.

**Cuadro 1/G.722.2 – Atribución de bits del algoritmo de codificación AMR-WB
para trama de 20 ms**

Modo	Parámetro	1ª subtrama	2ª subtrama	3ª subtrama	4ª subtrama	Total por trama
23,85 kbit/s	Bandera VAD					1
	ISP					46
	Filtrado LTP	1	1	1	1	4
	Periodo de tono	9	6	9	6	30
	Código algebraico	88	88	88	88	352
	Ganancia de libro de códigos	7	7	7	7	28
	Energía banda alta	4	4	4	4	16
	Total					
23,05 kbit/s	Bandera VAD					1
	ISP					46
	Filtrado LTP	1	1	1	1	4
	Periodo de tono	9	6	9	6	30
	Código algebraico	88	88	88	88	352
	Ganancias	7	7	7	7	28
	Total					
19,85 kbit/s	Bandera VAD					1
	ISP					46
	Filtrado LTP	1	1	1	1	4
	Periodo de tono	9	6	9	6	30
	Código algebraico	72	72	72	72	288
	Ganancias de libro de código	7	7	7	7	28
	Total					
18,25 kbit/s	Bandera VAD					1
	ISP					46
	Filtrado LTP	1	1	1	1	4
	Periodo de tono	9	6	9	6	30
	Código algebraico	64	64	64	64	256
	Ganancias	7	7	7	7	28
	Total					

**Cuadro 1/G.722.2 – Atribución de bits del algoritmo de codificación AMR-WB
para trama de 20 ms**

Modo	Parámetro	1ª subtrama	2ª subtrama	3ª subtrama	4ª subtrama	Total por trama
15,85 kbit/s	Bandera VAD					1
	ISP					46
	Filtrado LTP	1	1	1	1	4
	Periodo de tono	9	6	9	6	30
	Código algebraico	52	52	52	52	208
	Ganancias	7	7	7	7	28
	Total					
14,25 kbit/s	Bandera VAD					1
	ISP					46
	Filtrado LTP	1	1	1	1	4
	Periodo de tono	9	6	9	6	30
	Código algebraico	44	44	44	44	176
	Ganancias	7	7	7	7	28
	Total					
12,65 kbit/s	Bandera VAD					1
	ISP					46
	Filtrado LTP	1	1	1	1	4
	Periodo de tono	9	6	9	6	30
	Código algebraico	36	36	36	36	144
	Ganancias	7	7	7	7	28
	Total					
8,85 kbit/s	Bandera VAD					1
	ISP					46
	Periodo de tono	8	5	8	5	26
	Código algebraico	20	20	20	20	80
	Ganancias	6	6	6	6	24
	Total					
6,60 kbit/s	Bandera VAD					1
	ISP					36
	Periodo de tono	8	5	5	5	23
	Código algebraico	12	12	12	12	48
	Ganancias	6	6	6	6	24
	Total					

4.4 Principios del decodificador de multivelocidad adaptativa para voz

En la figura 3 se muestra el flujo de la señal por el decodificador. En él se extraen del tren binario recibido los índices transmitidos. Los índices son decodificados para obtener los parámetros del codificador en cada trama de transmisión. Dichos parámetros son el vector ISP, los cuatro retardos de altura de tono fraccionarios, los cuatro parámetros de filtrado LTP, los cuatro vectores-código renovadores y los cuatro conjuntos de ganancias de altura de tono y renovadoras con vectores cuantificados. En el modo 23,85 kbits/s se decodifica también el índice de ganancia de banda alta. El vector ISP es convertido en coeficientes del filtro LP e interpolado para obtener filtros LP en cada subtrama. A continuación, en cada subtrama de 64 muestras:

- Se construye la excitación añadiendo los vectores-código adaptativos y renovadores escalados por sus ganancias respectivas.
- Se reconstruye la voz a 12,8 kHz filtrando la excitación a través del filtro de síntesis LP.
- La voz reconstruida es desenfatizada.

Por último, se aplica a la voz reconstruida un muestreo en sentido ascendente a 16 kHz y se añade la señal vocal de banda alta a la banda de frecuencias de 6 kHz a 7 kHz.

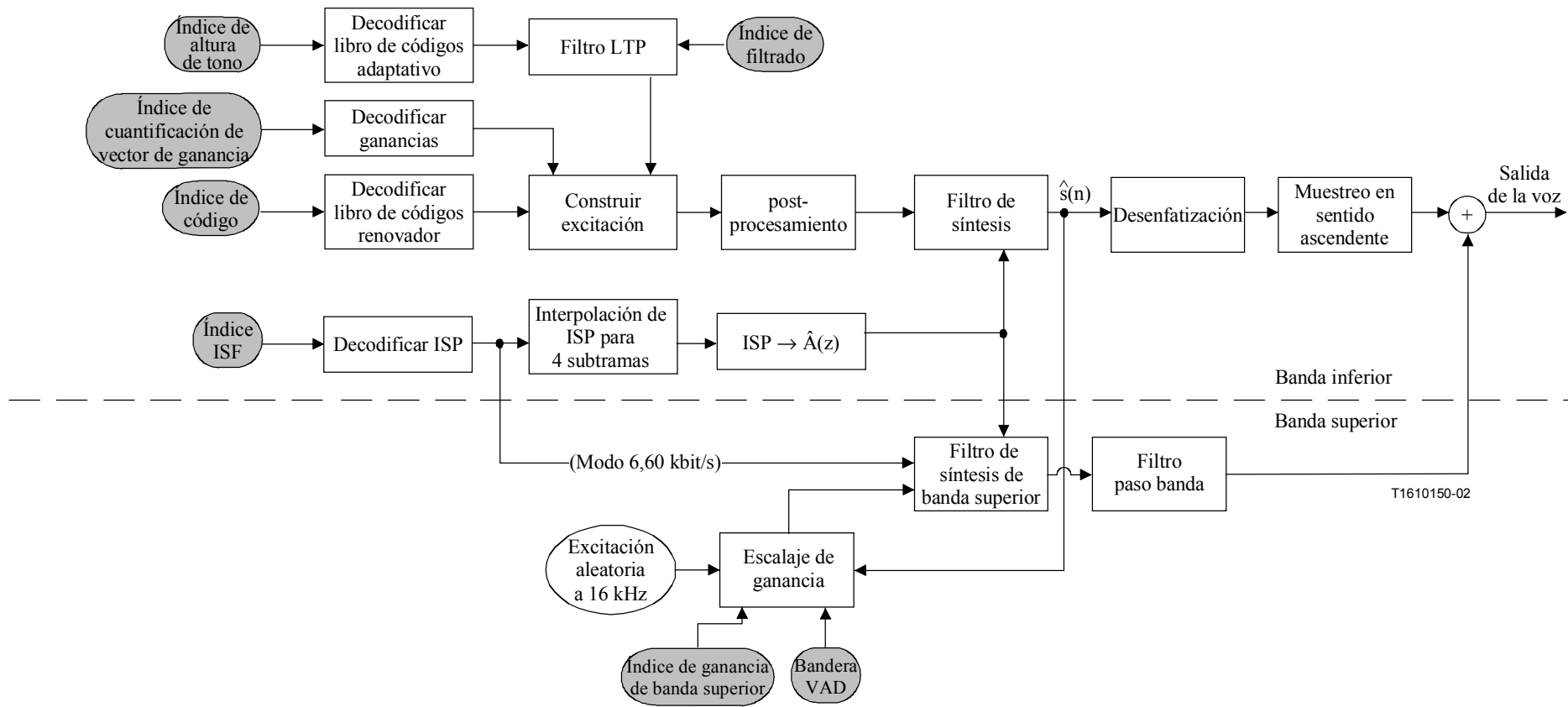


Figura 3 /G.722.2 – Diagrama de bloques detallado del decodificador ACELP

4.5 Secuencia e importancia subjetiva de los parámetros codificados

El codificador producirá la información de salida en una secuencia y con un formato único, y el decodificador debe recibir la misma información de la misma manera. En los cuadros 12a a 12i, se muestran la secuencia de los bits de salida y la atribución de bits para cada parámetro.

Los diferentes parámetros de la voz codificada y los bits de cada uno de ellos tienen una importancia desigual con respecto a la calidad subjetiva.

5 Descripción funcional del codificador

En esta cláusula se describen las diferentes funciones del codificador representado en la figura 2.

5.1 Preprocesamiento

El codificador efectúa el análisis de los parámetros LPC, LTP y del libro de códigos fijo a una velocidad de muestreo de 12,8 kHz. La señal de entrada, por tanto, ha de ser decimada de 16 kHz a 12,8 kHz. La decimación se lleva a cabo aplicando primero un muestreo en sentido ascendente por un factor de 4 y filtrando a continuación la salida a través del filtro FIR paso bajo $H_{\text{decim}}(z)$ cuya frecuencia de corte se halla en 6,4 kHz. Seguidamente, se aplica la señal un muestreo en sentido descendente por un factor de 5. La demora del filtrado se compensa añadiendo ceros al final del vector de entrada.

Después de la decimación, se aplican dos funciones de preprocesamiento a la señal antes del proceso de codificación: filtrado paso alto y preénfasis (y escalaje en sentido descendente de la señal).

(El escalaje en sentido descendente consiste en dividir la entrada por un factor de 2 para reducir la posibilidad de desbordamientos en la implementación de coma fija.)

El filtro paso alto sirve a modo de salvaguardia frente a componentes de baja frecuencia no deseados. Se utiliza un filtro con una frecuencia de corte de 50 Hz, y viene dado por:

$$H_{h1}(z) = \frac{0,989502 - 1,979004z^{-1} + 0,989502z^{-2}}{1 - 1,978882z^{-1} + 0,9799126z^{-2}} \quad (4)$$

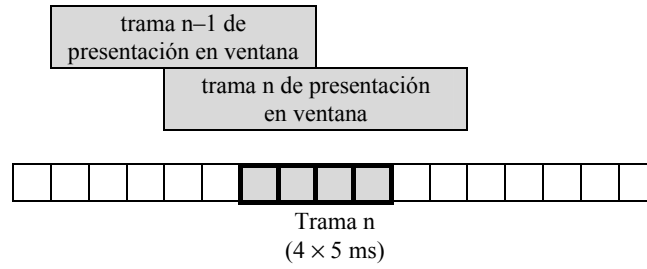
(El escalaje en sentido descendente y el filtrado paso alto se combinan dividiendo por 2 los coeficientes del numerador de $H_{h1}(z)$.)

En el preénfasis se utiliza un filtro paso alto de primer orden para enfatizar (resaltar) frecuencias superiores, y viene dado por:

$$H_{\text{pre-emph}}(z) = 1 - 0,68z^{-1} \quad (5)$$

5.2 Análisis y cuantificación de la predicción lineal

La predicción a corto plazo o análisis LP, se lleva a cabo una vez por trama de voz utilizando el procedimiento de la autocorrelación con ventanas asimétricas de 30 ms. En el cálculo de la autocorrelación se utiliza una tara de 5 ms. A continuación se muestra la estructura de la trama.



Las autocorrelaciones de las palabras presentadas en ventana se convierten en los coeficientes LP utilizando el algoritmo Levinson-Durbin. A continuación, los coeficientes LP son transformados al dominio ISP a efectos de cuantificación e interpolación. Los filtros interpolados cuantificados y no cuantificados se convierten de nuevo en los coeficientes de filtro LP (para construir los filtros de síntesis y ponderación en cada subtrama).

5.2.1 Presentación en ventana y cálculo de la autocorrelación

El análisis LP se lleva a cabo una vez por trama utilizando una ventana asimétrica. La ventana tiene su peso concentrado en la cuarta subtrama y consta de dos partes: la primera es la mitad de una ventana Hamming y la segunda es un cuarto de ciclo de la función de coseno Hamming. La ventana viene dada por:

$$\begin{aligned}
 w(n) &= 0,54 - 0,46 \cos\left(\frac{2\pi n}{2L_1 - 1}\right), & n = 0, \dots, L_1 - 1 \\
 &= \cos\left(\frac{2\pi(n - L_1)}{4L_2 - 1}\right), & n = L_1, \dots, L_1 + L_2 - 1
 \end{aligned} \tag{6}$$

donde se utilizan los valores $L_1=256$ y $L_2=128$.

Las autocorrelaciones de las palabras presentadas en ventana $s'(n), n = 0, \dots, 383$ se calculan aplicando la siguiente fórmula:

$$r(k) = \sum_{n=k}^{383} s'(n) s'(n-k), \quad k = 0, \dots, 16 \tag{7}$$

y se aplica una ampliación de anchura de banda de 60 Hz presentando en ventana retardada las autocorrelaciones que utilizan la ventana [3]:

$$w_{lag}(i) = \exp\left[-\frac{1}{2} \left(\frac{2\pi f_0 i}{f_s}\right)^2\right], \quad i = 1, \dots, 16 \tag{8}$$

donde $f_0 = 60$ Hz es la ampliación de la anchura de banda y $f_s = 12\,800$ Hz es la frecuencia de muestreo. Además, $r(0)$ se multiplica por el factor de corrección de ruido blanco 1,0001, lo que equivale añadir un suelo de ruido en -40 dB.

5.2.2 Algoritmo Levinson-Durbin

Las autocorrelaciones modificadas $r'(0) = 1.0001r(0)$ y $r'(k) = r(k)w_{lag}(k), k = 1, \dots, 16$, se utilizan para obtener los coeficientes de filtro LP $a_k, k = 1, \dots, 16$ resolviendo el sistema de ecuaciones:

$$\sum_{k=1}^{16} a_k r'(|i-k|) = -r'(i), \quad i = 1, \dots, 16 \tag{9}$$

El sistema de ecuaciones (9) se resuelve aplicando el algoritmo Levinson-Durbin [3]. Este algoritmo utiliza la siguiente recursión:

$$E(0) = r'(0)$$

Para $i = 1$ a 16 hágase

$$k_i = -\left[r'(i) + \sum_{j=1}^{i-1} a_j^{i-1} r'(i-j)\right] / E(i-1)$$

$$a_i^{(i)} = k_i$$

Para $j = 1$ a $i-1$ hágase

$$a_j^{(i)} = a_j^{(i-1)} + k_i a_{i-j}^{(i-1)}$$

$$E(i) = (1 - k_i^2) E(i-1)$$

La solución final viene dada por $a_j = a_j^{(16)}$, $j = 1, \dots, 16$.

Los coeficientes de filtro LP son convertidos a la representación ISP [5] a efectos de cuantificación e interpolación. Las conversiones al dominio ISP y el retorno al dominio de filtro LP se describen en las cláusulas 5.2.3 y 5.2.4.

5.2.3 Conversión LP a ISP

Los coeficientes de filtro LP a_k , $k=1, \dots, 16$ se convierten a la representación ISP a efectos de cuantificación e interpolación. Para un filtro LP de 16º orden, los ISP se definen como las raíces de los polinomios de suma y diferencia:

$$f_1'(z) = A(z) + z^{-16} A(z^{-1}) \quad (10)$$

y

$$f_2'(z) = A(z) - z^{-16} A(z^{-1}) \quad (11)$$

respectivamente. (Los polinomios $f_1'(z)$ y $f_2'(z)$ son simétrico y antisimétrico respectivamente). Puede demostrarse que todas las raíces de esos polinomios están en el círculo unidad y alternan entre sí [6]. $f_2'(z)$ tiene dos raíces en $z = 1$ ($\omega = 0$) y $z = -1$ ($\omega = \pi$). Para eliminar estas dos raíces se definen dos nuevos polinomios:

$$f_1(z) = f_1'(z) \quad (12)$$

y

$$f_2(z) = f_2'(z) / (1 - z^{-2}) \quad (13)$$

Los polinomios $f_1(z)$ y $f_2(z)$ tienen 8 y 7 raíces conjugadas en el círculo unidad ($e^{\pm j\omega_i}$), respectivamente. Por consiguiente, los polinomios pueden expresarse de la siguiente manera:

$$F_1(z) = (1 + a[16]) \prod_{i=0,2,\dots,14} (1 - 2q_i z^{-1} + z^{-2}) \quad (14)$$

y

$$F_2(z) = (1 + a[16]) \prod_{i=1,3,\dots,13} (1 - 2q_i z^{-1} + z^{-2}) \quad (15)$$

donde $q_i = \cos(\omega_i)$, siendo ω_i las frecuencias espectrales de inmitancia (ISF) y a [16] el último coeficiente del predictor. Las ISF satisfacen la propiedad de orden $0 < \omega_1 < \omega_2 < \dots < \omega_{16} < \pi$. En el dominio del coseno, q_i son los ISP.

Puesto que ambos polinomios $f_1(z)$ y $f_2(z)$ son simétricos, sólo es preciso calcular los primeros coeficientes 8 y 7 de cada polinomio, respectivamente, así como el último coeficiente del predictor.

Los coeficientes de estos polinomios se obtienen aplicando las siguientes relaciones recursivas:

para $i = 0$ a 7

$$\begin{aligned} f_1(i) &= a_i + a_{m-i} \\ f_2(i) &= a_i - a_{m-i} + f_2(i-2) \\ f_1(8) &= 2a_8 \end{aligned} \tag{16}$$

donde $m = 16$ es el orden del predictor, y $f_2(-2) = f_2(-1) = 0$.

Los ISP se obtienen evaluando los polinomios $F_1(z)$ y $F_2(z)$ en 100 puntos situados a intervalos iguales entre 0 y π y comprobando los cambios de signo. Cuando se produce un cambio de signo, ello significa la existencia de una raíz, y el intervalo del cambio de signo se divide entonces cuatro veces para acotar mejor esa raíz. Para evaluar $F_1(z)$ y $F_2(z)$ se utilizan los polinomios de Chebyshev [7]. En este método, las raíces se encuentran directamente en el dominio del coseno $\{q_i\}$. Los polinomios $F_1(z)$ y $F_2(z)$ evaluados en $z = e^{j\omega}$ se pueden expresar de la siguiente manera:

$$F_1(\omega) = 2e^{-j8\omega}C_1(x) \text{ y } F_2(\omega) = 2e^{-j7\omega}C_2(x) \tag{17}$$

con

$$C_1(x) = \sum_{i=0}^7 f_1(i)T_{8-i}(x) + f_1(8)/2, \text{ y } C_2(x) = \sum_{i=0}^6 f_2(i)T_{8-i}(x) + f_2(7)/2 \tag{18}$$

donde $T_m = \cos(m\omega)$ es el polinomio de Chebyshev de orden m -ésimo, y $f(i)$ son los coeficientes de $F_1(z)$ o $F_2(z)$, calculados utilizando las ecuaciones de (16). El polinomio $C(x)$ se evalúa en un determinado valor de $x = \cos(\omega)$ utilizando la relación recursiva:

para $k = n_f - 1$ a 1

$$b_k = 2xb_{k+1} - b_{k+2} + f(n_f - k)$$

final

$$C(x) = xb_1 - b_2 + f(n_f)/2$$

donde $n_f = 8$ en caso de $C_1(x)$ y $n_f = 7$ en caso de $C_2(x)$, con valores iniciales $b_{n_f} = f(0)$ y $b_{n_f+1} = 0$. Los detalles de la evaluación del polinomio de Chebyshev figuran en [7].

5.2.4 Conversión ISP a LP

Una vez cuantificados e interpolados los ISP, se convierten de nuevo al dominio de coeficientes LP $\{a_k\}$. La conversión al dominio LP se lleva a cabo como se indica a continuación. Los coeficientes de $F_1(z)$ y $F_2(z)$ se encuentran desarrollando las ecuaciones (14) y (15) con los ISP cuantificados e interpolados que ahora se conocen $q_i = , i = 0, \dots, m-1$, donde $m = 16$. Para calcular $f_1(z)$ se utiliza la siguiente relación recursiva:

para $i = 2$ a $m/2$

$$f_1(i) = -2q_{2i-2}f_1(i-1) + 2f_1(i-2)$$

para $j = i-1$ a 2

$$f_1(j) = f_1(j) - 2q_{2i-2}f_1(j-1) + f_1(j-2)$$

fin

$$f_1(1) = f_1(1) - 2q_{2i-2}$$

fin

con los valores iniciales $f_1(0) = 1$ y $f_1(1) = -2q_0$. Los coeficientes $f_2(i)$ se calculan de manera similar sustituyendo q_{2i-2} por q_{2i-1} y $m/2$ por $m/2-1$, y con las condiciones iniciales de $f_2(0) = 1$ y $f_2(1) = -2q_1$.

Una vez encontrados los coeficientes $f_1(z)$ y $f_2(z)$, se multiplica $F_2(z)$ por $1-z^{-2}$ para obtener $F'_2(z)$; esto es:

$$\begin{aligned} f'_2(i) &= f_2(i) - f_2(i-2), & i = 2, \dots, m/2-1 \\ f'_1(i) &= f_1(i) & i = 0, \dots, m/2 \end{aligned} \quad (19)$$

A continuación se multiplican $F'_1(z)$ y $F'_2(z)$ por $1+q_{m-1}$ y $1-q_{m-1}$, respectivamente; esto es:

$$\begin{aligned} f'_2(i) &= (1-q_{m-1})f'_2(i), & i = 0, \dots, m/2-1 \\ f'_1(i) &= (1+q_{m-1})f'_1(i) & i = 0, \dots, m/2 \end{aligned}$$

Por último, los coeficientes LP vienen dados por:

$$\begin{aligned} a_i &= 0.5f'_1(i) + 0.5f'_2(i), & i = 1, \dots, m/2-1 \\ &0.5f'_1(i) - 0.5f'_2(i), & i = m/2+1, \dots, m-1 \\ &0.5f'_1(m/2), & i = m/2 \\ &q_{m-1}, & i = m \end{aligned} \quad (20)$$

Lo anterior se deduce directamente de la relación $A(z) = (F'_1(z) + F'_2(z))/2$, teniendo en cuenta que $F'_1(z)$ y $F'_2(z)$ son polinomios simétrico y antisimétrico, respectivamente.

5.2.5 Cuantificación de los coeficientes ISP

Los coeficientes de filtro LP se cuantifican utilizando la representación ISP en el dominio de la frecuencia, esto es:

$$\begin{aligned} f_i &= \frac{f_s}{2\pi} \arccos(q_i), & i = 0, \dots, 14 \\ &= \frac{f_s}{4\pi} \arccos(q_i), & i = 15 \end{aligned} \quad (21)$$

siendo f_i las ISF en Hz [0,6400] y $f_s = 12\,800$ es la frecuencia de muestreo. El vector ISF viene dado por $\mathbf{f}^t = [f_0 f_1, \dots, f_{15}]$, donde t indica trasposición.

Se aplica una predicción MA de 1^{er} orden y se cuantifica el vector ISF residual utilizando una combinación de cuantificación de vector dividida (SVQ, *split sector quantization*) y cuantificación de vector multietapa (MSVQ, *multistage vector quantization*). La predicción y la cuantificación se llevan a cabo como sigue. Sea $\mathbf{z}(n)$ el vector ISF, al cual se le ha suprimido su valor medio, en la trama n . El vector residual de predicción $\mathbf{r}(n)$ viene dado por:

$$\mathbf{r}(n) = \mathbf{z}(n) - \mathbf{p}(n) \quad (22)$$

donde $\mathbf{p}(n)$ es el vector LSF predicho en la trama n . Se utiliza una predicción de media móvil (MA) de primer orden en la que:

$$\mathbf{p}(n) = \frac{1}{3} \hat{\mathbf{r}}(n-1) \quad (23)$$

donde $\hat{\mathbf{r}}(n-1)$ es el vector residual cuantificado en la trama pasada.

El vector residual \mathbf{r} de ISF se cuantifica utilizando cuantificación de vector multietapa dividida S-MSVQ. El vector se divide en dos subvectores $\mathbf{r}_1(n)$ y $\mathbf{r}_2(n)$ de dimensiones 9 y 7 respectivamente. Los dos subvectores se cuantifican en dos etapas. En la primera etapa se cuantifica $\mathbf{r}_1(n)$ con 8 bits y $\mathbf{r}_2(n)$ con 8 bits.

Para los modos 8,85, 12,65, 14,25, 15,85, 18,25, 19,85, 23,05 ó 23,85 kbit/s, los vectores de error de cuantificación $\mathbf{r}_i^{(2)} = \mathbf{r} - \hat{\mathbf{r}}_i, i=1,2$ se dividen en la etapa siguiente en tres y dos subvectores, respectivamente. Los subvectores se cuantifican utilizando las velocidades binarias descritas en el cuadro 2.

Cuadro 2/G.722.2 – Cuantificación de vector ISP para los modos 8,85, 12,65, 14,25, 15,85, 18,25, 19,85, 23,05 o 23,85 kbit/s

1. Vector ISP de 16 elementos de longitud no cuantificado				
2. Etapa 1 (\mathbf{r}_1) 8 bits			2. Etapa 1 (\mathbf{r}_2) 8 bits	
3. Etapa 2 ($\mathbf{r}^{(2)}_{1,0-2}$) 6 bits	3. Etapa 2 ($\mathbf{r}^{(2)}_{1,3-5}$) 7 bits	3. Etapa 2 ($\mathbf{r}^{(2)}_{1,6-8}$) 7 bits	3. Etapa 2 ($\mathbf{r}^{(2)}_{2,0-2}$) 5 bits	3. Etapa 2 ($\mathbf{r}^{(2)}_{2,3-6}$) 5 bits

Para el modo 6,60 kbit/s, los vectores de error de cuantificación $\mathbf{r}_i^{(2)} = \mathbf{r}_i - \hat{\mathbf{r}}_i, i=1,2$ se dividen en la etapa siguiente en 2 y 1 subvectores, respectivamente. Los subvectores se cuantifican utilizando las velocidades binarias descritas en el cuadro 3.

Cuadro 3/G.722.2 – Cuantificación de vector ISP para el modo 6,60 kbit/s

1. Vector ISP de 16 elementos de longitud no cuantificado		
2. Stage 1 (\mathbf{r}_1) 8 bits		2. Stage 1 (\mathbf{r}_2) 8 bits
3. Etapa 2 ($\mathbf{r}^{(2)}_{1,0-4}$) 7 bits	3. Etapa 2 ($\mathbf{r}^{(2)}_{1,5-8}$) 7 bits	3. Etapa 2 ($\mathbf{r}^{(2)}_{2,0-6}$) 6 bits

En el proceso de cuantificación se utiliza una medida de la distorsión, el ISP de errores cuadráticos. En general, para un subvector ISP de entrada o residual de error $\mathbf{r}_i, i=1,2$ y un vector cuantificado en el índice k , $\hat{\mathbf{r}}_i^k$, la cuantificación se realiza encontrando el índice k que hace mínimo

$$E = \sum_{i=m}^n \left[r_i - \hat{r}_i^k \right]^2 \quad (24)$$

donde m y n son el primer y el último elementos del subvector.

5.2.6 Interpolación de los ISP

El conjunto de los parámetros LP cuantificados (y no cuantificados) se utiliza para la cuarta subtrama mientras que la primera, la segunda y la tercera subtramas utilizan una interpolación lineal de los parámetros en las tramas adyacentes. La interpolación de los ISP se lleva a cabo en el dominio de \mathbf{q} . Sea $\hat{\mathbf{q}}_4^{(n)}$ el vector ISP en la 4ª subtrama de la trama, y $\hat{\mathbf{q}}_4^{(n-1)}$ el vector ISP en la 4ª subtrama de la trama pasada $n-1$. Los vectores ISP interpolados en las subtramas 1ª, 2ª y 3ª vienen dados por:

$$\hat{\mathbf{q}}_1^{(n)} = 0,55\hat{\mathbf{q}}_4^{(n-1)} + 0,45\hat{\mathbf{q}}_4^{(n)}$$

$$\hat{\mathbf{q}}_2^{(n)} = 0,2\hat{\mathbf{q}}_4^{(n-1)} + 0,8\hat{\mathbf{q}}_4^{(n)}$$

$$\hat{\mathbf{q}}_3^{(n)} = 0,04\hat{\mathbf{q}}_4^{(n-1)} + 0,96\hat{\mathbf{q}}_4^{(n)}$$

Se utiliza la misma fórmula para la interpolación de los ISP no cuantificados. Los vectores ISP interpolados se emplean para calcular un filtro LP diferente en cada subtrama (cuantificada y no cuantificada) aplicando el método de conversión ISP a LP descrito en 5.2.4.

5.3 Ponderación perceptual

El filtro de ponderación perceptual tradicional $W(z) = A(z/\gamma_1)/A(z/\gamma_2)$ tiene limitaciones específicas cuando se modelan la estructura conformadora y la inclinación espectral requerida al mismo tiempo. La inclinación espectral es más pronunciada en las señales de banda ancha debido a la amplia gama dinámica entre frecuencias bajas y altas. Una solución a este problema consiste en introducir el filtro de preénfasis a la entrada, calcular el filtro LP $A(z)$ en base a la señal vocal preenfatisada $s(n)$ y utilizar un filtro modificado $W(z)$ fijando su denominador. Esta estructura desvincula sustancialmente la ponderación conformadora de la inclinación.

Se utiliza un filtro de ponderación de la forma $W(z) = A(z/\gamma_1)H_{\text{de-emph}}(z)$, donde

$$H_{\text{de-emph}} = \frac{1}{1-\beta_1 z^{-1}} \text{ y } \beta_1 = 0,68.$$

Puesto que $A(z)$ se calcula en base a la señal vocal preenfatisada $s(n)$, la inclinación del filtro $1/A(z/\gamma_1)$ es menos pronunciada en comparación con el caso en que $A(z)$ se calcula en base a la señal vocal original. La desenfatisación se lleva a cabo en el extremo del decodificado, y por ello puede mostrarse cómo el espectro del error de cuantificación es conformado por un filtro que tiene una función de transferencia $W^{-1}(z)H_{\text{de-emph}}(z) = 1/A(z/\gamma_1)$. Así pues, el espectro del error de cuantificación lo conforma por un filtro cuya función de transferencia es $1/A(z/\gamma_1)$, calculándose $A(z)$ de acuerdo con la señal vocal preenfatisada.

5.4 Análisis de la altura de tono en bucle abierto

Dependiendo del modo, el análisis de la altura de tono en bucle abierto se efectúa una vez por trama (cada 10 ms) o dos veces por trama (cada 10 ms) para obtener dos estimaciones del retardo de la altura de tono en cada trama. Se hace así a fin de simplificar el análisis de la altura de tono y limitar la búsqueda de la altura de tono en bucle cerrado a un pequeño número de retardos en torno a los retardos estimados en bucle abierto.

La estimación de la altura de tono en bucle abierto se basa en la señal vocal ponderada $s_w(n)$ que se obtiene filtrando la señal vocal de entrada a través del filtro de ponderación

$$W(z) = A(z/\gamma_1)H_{\text{de-emph}}(z), \text{ donde } H_{\text{de-emph}} = \frac{1}{1-\beta_1 z^{-1}} \text{ y } \beta_1 = 0,68. \text{ Esto es, en una subtrama}$$

de tamaño L , la señal vocal ponderada viene dada por:

$$s_w(n) = s(n) + \sum_{i=1}^{16} a_i \gamma_1^i s(n-i) + \beta_1 s_w(n-1), \quad n = 0, \dots, L-1 \quad (25)$$

El análisis de la altura de tono en bucle abierto se efectúa en una señal decimada por dos. La señal decimada se obtiene filtrando $s_w(n)$ a través de un filtro FIR de cuarto orden $H_{\text{decim2}}(z)$ y aplicando a continuación a la salida un muestreo en sentido descendente por un factor de 2, para obtener la señal $s_{wd}(n)$.

5.4.1 Modo 6,60 kbit/s

El análisis de la altura de tono en bucle abierto se efectúa una vez por trama (cada 20 ms) para obtener una estimación del retardo de altura de tono en cada trama.

La realización del análisis de la altura de tono en bucle abierto es como sigue. Primero se determina la correlación de la voz ponderada decimada para cada valor de retardo de altura de tono d aplicando la siguiente fórmula:

$$C(d) = \sum_{n=0}^{128} S_{wd}(n) s_{wd}(n-d) w(d), \quad d = 17, \dots, 115 \quad (26)$$

donde $w(d)$ es una función de ponderación. El retardo de altura de tono estimado es el periodo de tiempo que maximiza la función de correlación ponderada $C(d)$. La ponderación enfatiza los valores del retardo de altura de tono más bajos reduciendo la probabilidad de seleccionarse un múltiplo del periodo de tiempo correcto. La función de ponderación consta de dos partes: una de énfasis del retardo de altura de tono bajo, $w_l(d)$ y otra de énfasis en torno al retardo de la trama previa, $w_n(d)$:

$$w(d) = w_l(d) w_n(d) \quad (27)$$

La función de énfasis del retardo de altura de tono bajo viene dada por:

$$w_l(d) = cw(d) \quad (28)$$

donde $cw(d)$ se define mediante un cuadro en la descripción computacional de coma fija. La función de énfasis en torno al retardo de la trama previa depende del retardo de altura de tono de las tramas de voz previas:

$$w_n(d) = \begin{cases} cw(|T_{old} - d| + 98), & v > 0,8 \\ 1,0, & \text{de otro modo} \end{cases} \quad (29)$$

donde T_{old} es el retardo de altura de tono filtrado mediano de 5 semitramas vocales previas y v es un parámetro adaptativo. Si la trama se clasifica como vocal por tener una ganancia en bucle abierto $g > 0,6$, se fija a continuación el valor de v a 1,0 para la trama siguiente. De otro modo, se actualiza el valor de v adoptando $v = 0,9v$. La ganancia en bucle abierto viene dada por:

$$g = \frac{\sum_{n=0}^{127} s_{wd}(n) s_{wd}(n - d_{\text{máx}})}{\sqrt{\sum_{n=0}^{127} s_{wd}^2(n) \sum_{n=0}^{127} s_{wd}^2(n - d_{\text{máx}})}} \quad (30)$$

donde $d_{\text{máx}}$ es el periodo de tiempo de altura de tono que maximiza $C(d)$. El filtro mediano se actualiza solamente durante las tramas vocales. La ponderación depende de la fiabilidad de los retardos de altura de tono antiguos. Si las tramas previas han contenido señales vocales sordas o silencio, la ponderación es atenuada mediante el parámetro v .

5.4.2 Modos 8,85, 12,65, 14,25, 15,85, 18,25, 19,85, 23,05 y 23,85 kbit/s

El análisis de altura de tono en bucle abierto se efectúa dos veces por trama (cada 10 ms) para obtener dos estimaciones del retardo de altura de tono en cada trama.

La realización del análisis de la altura de tono en bucle abierto es como sigue. Primero se determina la correlación de la voz ponderada decimada para cada valor de retardo de altura de tono d aplicando la siguiente fórmula:

$$C(d) = \sum_{n=0}^{63} s_{wd}(n)s_{wd}(n-d)w(d), \quad d = 17, \dots, 115 \quad (31)$$

donde $w(d)$ es una función de ponderación. El retardo de altura de tono estimado es el periodo de tiempo que maximiza la función de correlación ponderada $C(d)$. La ponderación enfatiza los valores del retardo de altura de tono más bajos reduciendo la probabilidad de seleccionar un múltiplo del periodo de tiempo correcto. La función de ponderación consta de dos partes: una de énfasis del retardo de altura de tono bajo, $w_l(d)$, y otra de énfasis en torno al retardo de la trama previa $w_n(d)$:

$$w(d) = w_l(d)w_n(d) \quad (32)$$

La función de énfasis del retardo de altura de tono bajo viene dada por:

$$w_l(d) = cw(d) \quad (33)$$

donde $cw(d)$ se define mediante un cuadro en la descripción computacional de coma fija. La función de énfasis en torno al retardo de la trama previa depende del retardo de altura de tono de las tramas de voz previas:

$$w_n(d) = \begin{cases} cw(|T_{old} - d| + 98), & v > 0,8, \\ 1,0, & \text{de otro modo} \end{cases} \quad (34)$$

donde T_{old} es el retardo de altura de tono filtrado mediano de 5 semitramas vocales previas y v es un parámetro adaptativo. Si la trama se clasifica como vocal por tener una ganancia en bucle abierto $g > 0,6$, el valor de v se fija a continuación en 1,0 para la trama siguiente. De otro modo, se actualiza el valor de v adoptando $v = 0,9v$. La ganancia en bucle abierto viene dada por:

$$g = \frac{\sum_{n=0}^{63} s_{wd}(n)s_{wd}(n-d_{m\acute{a}x})}{\sqrt{\sum_{n=0}^{63} s_{wd}^2(n) \sum_{n=0}^{63} s_{wd}^2(n-d_{m\acute{a}x})}} \quad (35)$$

donde $d_{m\acute{a}x}$ es el periodo de tiempo de altura de tono que maximiza $C(d)$. El filtro mediano se actualiza solamente durante las tramas vocales. La ponderación depende de la fiabilidad de los retardos de altura de tono antiguos. Si las tramas previas han contenido señales vocales sordas o silencio, la ponderación es atenuada mediante el parámetro v .

5.5 Cálculo de la respuesta en tren de impulsos

La respuesta en tren de impulsos, $h(n)$, del filtro de síntesis ponderada $H(z)W(z) = A(z/\gamma_1)H_{de-emph}(z)/\hat{A}(z)$ se calcula en cada subtrama. Esta respuesta en tren de impulsos se necesita para la búsqueda en los libros de códigos adaptativo y fijo. La respuesta en tren de impulsos $h(n)$ se calcula filtrando el vector de coeficientes del filtro $A(z/\gamma_1)$ ampliado con ceros a través de los dos filtros $1/\hat{A}(z)$ y $H_{de-emph}(z)$.

5.6 Cálculo de la señal objetivo

La señal objetivo de la búsqueda en el libro de códigos se calcula normalmente sustrayendo la respuesta de entrada cero del filtro de síntesis ponderada $H(z)W(z) = A(z/\gamma_1)H_{\text{de-emph}}(z)/\hat{A}(z)$ de la señal vocal ponderada $s_w(n)$. Esta operación se lleva a cabo en cada una de las subtramas.

Un procedimiento equivalente de cálculo de la señal objetivo, que se utiliza en este códec, consiste en filtrar la señal residual LP $r(n)$ a través de la combinación del filtro de síntesis $1/\hat{A}(z)$ y el filtro de ponderación $A(z/\gamma_1)H_{\text{de-emph}}(z)$. Tras determinar la excitación de la subtrama, se actualizan los estados iniciales de estos filtros filtrando la diferencia entre la señal residual LP y la señal de excitación. La actualización de la memoria de los filtros se explica en 5.10.

La señal residual $r(n)$ que se necesita para encontrar el vector objetivo se utiliza también en la búsqueda en libro de códigos adaptativo para ampliar la memoria intermedia de excitación pasada. Así se simplifica el procedimiento de búsqueda en libro de códigos adaptativo con demoras inferiores al tamaño de la subtrama de 64 como se explicará en cláusula 5.7. La señal residual LP viene dada por:

$$r(n) = s(n) - \sum_{i=1}^{16} \hat{a}_i s(n-i), \quad n = 0, \dots, 63 \quad (36)$$

5.7 Libro de códigos adaptativo

La búsqueda en libro de códigos adaptativo se lleva a cabo en cada una de las subtramas. Consiste en efectuar una búsqueda de altura de tono en bucle cerrado y calcular a continuación el vector-código adaptativo interpolando la excitación pasada en el retardo de altura de tono fraccionario seleccionado.

Los parámetros de libro de códigos adaptativo (o parámetros de altura de tono) son el retardo y la ganancia del filtro de altura de tono. En la etapa de búsqueda, la excitación es ampliada por la señal residual LP para simplificar la búsqueda en bucle cerrado.

En los modos 12,65, 14,25, 15,85, 18,25, 19,85, 23,05 o 23,85 kbit/s, en las subtramas primera y tercera, se utiliza un retardo de altura de tono fraccionario con resolución de 1/4 en la gama $[34, 127\frac{3}{4}]$, resolución de 1/2 en la gama $[128, 159\frac{1}{2}]$, y de enteros solamente en la gama $[160, 231]$. Para las subtramas segunda y cuarta se utiliza siempre una resolución de altura de tono de 1/4 en la gama $[T_1-8, T_1+7\frac{3}{4}]$, donde T_1 es el entero más cercano al retardo de altura de tono fraccionario de la subtrama previa (primera o tercera).

En el modo 8,85 kbit/s, en las subtramas primera y tercera, se utiliza un retardo de altura de tono fraccionario con resolución de 1/2 en la gama $[34, 91\frac{1}{2}]$, y de enteros solamente en la gama $[92, 231]$. Para las subtramas segunda y cuarta se utiliza siempre una resolución de altura de tono de 1/2 en la gama $[T_1-8, T_1+7\frac{1}{2}]$, donde T_1 es el entero más cercano al retado de altura de tono fraccionario de la subtrama previa (primera o tercera).

En el modo 6,60 kbit/s, en la primera subtrama, se utiliza un retardo de altura de tono fraccionario con resolución de 1/2 en la gama $[34, 91\frac{1}{2}]$, y de enteros solamente en la gama $[92, 231]$. Para las subtramas segunda y cuarta se utiliza siempre una resolución de alturas de tono de 1/2 en la gama $[T_1-8, T_1+7\frac{1}{2}]$, donde T_1 es el entero más cercano al retardo de altura de tono fraccionario de la primera subtrama.

El análisis de la altura de tono en bucle cerrado se lleva a cabo en torno a las estimaciones de la altura de tono en bucle abierto en cada una de las subtramas. En los modos 8,85, 12,65, 14,25, 15,85, 18,25, 19,85, 23,05 o 23,85 kbit/s, en la subtrama primera (y tercera) se busca la gama $T_{op} \pm 7$, limitada por 34...231. En el modo 6,60 kbit/s, en la subtrama primera se busca la gama $T_{op} \pm 7$, limitada por 34...231. Para todos los modos y todas las demás subtramas, el análisis de altura de tono en bucle cerrado se efectúa en torno a la altura de tono entera seleccionada en la subtrama previa, como se ha descrito anteriormente. En los modos 12,65, 14,25, 15,85, 18,25, 19,85, 23,05 ó 23,85 kbit/s, el periodo de tiempo de altura de tono se codifica con 9 bits en las subtramas primera y tercera y el periodo de tiempo relativo de las otras subtramas se codifica con 6 bits. En el modo 8,85 kbit/s, el periodo de tiempo de altura de tono se codifica con 8 bits en las subtramas primera y tercera y el periodo de tiempo relativo de las otras subtramas se codifica con 5 bits. En el modo 6,60 kbit/s, el periodo de tiempo de altura de tono se codifica con 8 bits en la subtrama primera y el periodo de tiempo relativo de las otras subtramas se codifica con 5 bits.

La búsqueda de la altura de tono en bucle cerrado se lleva a cabo minimizando el error ponderado cuadrático medio entre la voz original y la voz sintetizada. Para ello se maximiza el término:

$$T_k = \frac{\sum_{n=0}^{63} x(n)y_k(n)}{\sqrt{\sum_{n=0}^{63} y_k(n)y_k(n)}}, \quad (37)$$

donde $x(n)$ es la señal objetivo e $y_k(n)$ es la excitación filtrada pasada con una demora de k [excitación pasada convolucionada con $h(n)$]. Se señala que la gama de búsqueda está limitada en torno a la altura de tono en bucle abierto como se explicó anteriormente.

La convolución $y_k(n)$ se calcula para la primera demora en la gama buscada y, para las demás demoras de tiempo, se actualiza utilizando la relación recursiva:

$$y_k(n) = y_{k-1}(n-1) + u(-k)h(n) \quad (38)$$

donde $u(n)$, $n = -(231+17), \dots, 63$, es la memoria intermedia de excitación. Se señala que, en la etapa de búsqueda, las muestras $u(n)$, $n = 0, \dots, 63$, son desconocidas y se necesitan para retardos de altura de tono inferiores a 64. Para simplificar la búsqueda se copia la señal residual LP en $u(n)$ con objeto de hacer que la relación de la ecuación (38) sea válida cualquiera que sea el periodo de tiempo.

Una vez determinado el retardo de altura de tono entero óptimo, se prueban las fracciones de $-\frac{3}{4}$

a $\frac{3}{4}$ con un paso de $\frac{1}{4}$ en torno a ese valor entero. La búsqueda de altura de tono fraccionaria se

lleva a cabo interpolando la correlación normalizada en la ecuación (37) y buscando su máximo. Una vez determinado el retardo de altura de tono fraccionario, se calcula $v'(n)$ interpolando la señal de excitación pasada $u(n)$ en la fase (fracción) dada. [La interpolación se efectúa utilizando dos filtros FIR (funciones de sincronización presentadas en ventana Hamming); una para interpolar el término de la ecuación (34) con la sincronización truncada a ± 17 y el otro para interpolar la señal de excitación pasada con la sincronización truncada a ± 63]. Los filtros tienen su frecuencia de corte (-3 dB) a 6000 Hz en el dominio del sobremuestreo, lo que significa que los filtros de interpolación muestran respuesta en frecuencia paso bajo. Así pues, incluso cuando el retardo de altura de tono es un valor entero, la excitación de libro de códigos adaptativo consiste en una versión filtrada paso bajo de la excitación pasada durante el periodo de tiempo dado y no en una copia de la misma. Además, en el caso de retardos menores que el tamaño de la subtrama, la excitación de libro de códigos adaptativo se completa de acuerdo con la excitación pasada interpolada con filtro paso bajo y no repitiendo la excitación pasada.

Para mejorar la calidad de la predicción de la altura de tono en señales de banda ancha se utiliza un predictor de altura de tono dependiente de la frecuencia. Esto es importante en señales de banda ancha porque la periodicidad no necesariamente abarca la totalidad del espectro. En este algoritmo hay dos trayectos de señal asociados a los correspondientes conjuntos de parámetros de libro de códigos de altura de tono, y cada uno de esos trayectos consta de un dispositivo de cálculo del error de predicción de la altura de tono para calcular el error de predicción de un vector-código de altura de tono a partir de un dispositivo de búsqueda en libro de códigos de altura de tono. Uno de esos dos trayectos tiene un filtro paso bajo para filtrar el vector-código de altura de tono y el error de predicción de altura de tono se calcula para ambos trayectos de señal. El trayecto de señal que tiene el error de predicción de altura de tono calculado más bajo es el que se selecciona, junto con la ganancia de altura de tono asociada.

El filtro paso bajo utilizado en el segundo trayecto es de la forma $B_{LP}(z) = 0,18z + 0,64 + 0,8z^{-1}$. Se señala que se utiliza 1 bit para codificar el trayecto elegido.

Así pues, para los modos 8,85, 12,65, 14,25, 15,85, 18,25, 19,85, 23,05 ó 23,85 kbit/s, hay dos posibilidades de generar el libro de códigos adaptativo $v(n)$, $v(n) = v'(n)$ en el primer trayecto, o $v(n) = \sum_{i=1}^1 b_{LP}(i+1)v'(n+i)$ en el segundo trayecto, donde $\mathbf{b}_{LP} = [0,18, 0,64, 0,18]$. El trayecto que da

lugar a una energía mínima de la señal objetivo $x_2(n)$ definida en la ecuación (40) es seleccionado para el vector de libro de códigos adaptativo filtrado. Para el modo 6,60 kbit/s, $v(n)$ es siempre

$$v(n) = \sum_{i=1}^1 b_{LP}(i+1)v'(n+i)$$

La ganancia de libro de códigos adaptativo viene dada entonces por:

$$g_p = \frac{\sum_{n=0}^{63} x(n)y(n)}{\sqrt{\sum_{n=0}^{63} y(n)y(n)}}, \quad \text{limitado por } 0 \leq g_p \leq 1,2 \quad (39)$$

donde $y(n) = v(n)*h(n)$ es el vector de libro de códigos adaptativo filtrado [respuesta de estado cero de $H(z)W(z)$ a $v_i(n)$]. Para asegurar la estabilidad, se limita la ganancia de libro de códigos adaptativo g_p a 0,95, si las ganancias de libros de códigos adaptativos de las subtramas previas han sido pequeñas y los filtros LP de esas subtramas han estado cerca de la inestabilidad.

5.8 Libro de códigos algebraico

5.8.1 Estructura de libro de códigos

La estructura del libro de códigos se basa en el diseño con permutación de impulsos individuales entrelazados (ISPP, *interleaved single-pulse permutation*). Las 64 posiciones del vector-código se dividen en 4 pistas de posiciones entrelazadas, con 16 posiciones en cada pista. Los diferentes libros de códigos a diferentes velocidades se construyen situando un cierto número de impulsos con signo en las pistas (de 1 a 6 impulsos por pista). El índice del libro de códigos, o palabra de código, representa las posiciones y los signos de los impulsos de cada pista. Así pues, no se necesita almacenar el libro de códigos, ya que el vector de excitación del decodificador puede ser construido mediante la información contenida en el propio índice (sin cuadro de consulta).

Una característica importante del libro de códigos utilizado es que se trata de un libro de códigos dinámico formado por un libro de códigos algebraico al que sigue un prefiltro adaptativo $F(z)$ que refuerza los componentes espectrales especiales para mejorar la calidad de la voz sintetizada. Se utiliza un prefiltro propio de las señales de banda ancha en el que $F(z)$ consta de dos partes: una parte de mejora de la periodicidad $1/(1-0,85z^{-T})$ y una parte de inclinación $(1 - \beta_1 z^{-1})$, donde T es la parte entera del retardo de altura de tono y β_1 está relacionado con la circunstancia de que la

subtrama anterior sea una subtrama vocal y está limitado por $[0,0, 0,5]$. La búsqueda en el libro de códigos se efectúa en el dominio algebraico combinando el filtro $F(z)$ con el filtro de síntesis ponderada antes de la búsqueda en el libro de códigos. Así pues, la respuesta en tren de impulsos $h(n)$ ha de ser modificada para incluir el prefiltro $F(z)$. Esto es, $h(n) \leftarrow h(n) * f(n)$.

Las estructuras de los libros de códigos a diferentes velocidades binarias se dan más adelante.

5.8.1.1 Modo 23,85 y 23,05 kbit/s

En este libro de códigos, el vector de renovación contiene 24 impulsos distintos de cero. Todos los impulsos pueden tener las amplitudes $+1$ o -1 . Las 64 posiciones de una subtrama se dividen en 4 pistas, conteniendo cada una de ellas seis impulsos, como se muestra en el cuadro 4.

Cuadro 4/G.722.2 – Posiciones posibles de impulsos individuales en el libro de códigos algebraico, 23,85 y 23,05 kbit/s

Pista	Impulso	Posiciones
1	$i_0, i_4, i_8, i_{12}, i_{16}, i_{20}$	0, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60
2	$i_1, i_5, i_9, i_{13}, i_{17}, i_{21}$	1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, 49, 53, 57, 61
3	$i_2, i_6, i_{10}, i_{14}, i_{18}, i_{22}$	2, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62
4	$i_3, i_7, i_{11}, i_{15}, i_{19}, i_{23}$	3, 7, 11, 15, 19, 23, 27, 31, 35, 39, 43, 47, 51, 55, 59, 63

Los seis impulsos de una pista se codifican con 22 bits.

Esto da un total de 88 bits $(22+22+22+22)$ para el código algebraico.

5.8.1.2 Modo 19,85 kbit/s

En este libro de códigos, el vector de renovación contiene 18 impulsos distintos de cero. Todos los impulsos pueden tener las amplitudes $+1$ o -1 . Las 64 posiciones de una subtrama se dividen en 4 pistas, conteniendo cada una de las dos primeras cinco impulsos y cada una de las otras dos cuatro impulsos, como se muestra en el cuadro 5.

Cuadro 5/G.722.2 – Posiciones posibles de impulsos individuales en el libro de códigos algebraico, 19,85 kbit/s

Pista	Impulso	Posiciones
1	$i_0, i_4, i_8, i_{12}, i_{16}$	0, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60
2	$i_1, i_5, i_9, i_{13}, i_{17}$	1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, 49, 53, 57, 61
3	i_2, i_6, i_{10}, i_{14}	2, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62
4	i_3, i_7, i_{11}, i_{15}	3, 7, 11, 15, 19, 23, 27, 31, 35, 39, 43, 47, 51, 55, 59, 63

Los cinco impulsos de una pista se codifican con 20 bits. Los cuatro impulsos de una pista se codifican con 16 bits.

Esto da un total de 72 bits $(20+20+16+16)$ para el código algebraico.

5.8.1.3 Modo 18,25 kbit/s

En este libro de códigos, el vector de renovación contiene 16 impulsos distintos de cero. Todos los impulsos pueden tener las amplitudes $+1$ o -1 . Las 64 posiciones de una subtrama se dividen en 4 pistas conteniendo cada una de las pistas cuatro impulsos, como se muestra en el cuadro 6.

Cuadro 6/G.722.2 – Posiciones posibles de impulsos individuales en el libro de códigos algebraico, 18,25 kbit/s

Pista	Impulso	Posiciones
1	i_0, i_4, i_8, i_{12}	0, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60
2	i_1, i_5, i_9, i_{13}	1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, 49, 53, 57, 61
3	i_2, i_6, i_{10}, i_{14}	2, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62
4	i_3, i_7, i_{11}, i_{15}	3, 7, 11, 15, 19, 23, 27, 31, 35, 39, 43, 47, 51, 55, 59, 63

Los cuatro impulsos de una pista se codifican con 16 bits.

Esto da un total de 64 bits (16+16+16+16) para el código algebraico.

5.8.1.4 Modo 15,85 kbit/s

En este libro de códigos, el vector de renovación contiene 12 impulsos distintos de cero. Todos los impulsos pueden tener las amplitudes +1 o -1. Las 64 posiciones de una subtrama se dividen en 4 pistas conteniendo cada una de las pistas tres impulsos, como se muestra en el cuadro 7.

Cuadro 7/G.722.2 – Posiciones posibles de impulsos individuales en el libro de códigos algebraico, 15,85 kbit/s

Pista	Impulso	Posiciones
1	i_0, i_4, i_8	0, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60
2	i_1, i_5, i_9	1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, 49, 53, 57, 61
3	i_2, i_6, i_{10}	2, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62
4	i_3, i_7, i_{11}	3, 7, 11, 15, 19, 23, 27, 31, 35, 39, 43, 47, 51, 55, 59, 63

Los tres impulsos de una pista se codifican con 13 bits.

Esto da un total de 52 bits (13+13+13+13) para el código algebraico.

5.8.1.5 Modo 14,25 kbit/s

En este libro de códigos, el vector de renovación contiene 10 impulsos distintos de cero. Todos los impulsos pueden tener las amplitudes +1 o -1. Las 64 posiciones de una subtrama se dividen en 4 pistas conteniendo cada una de las dos primeras tres impulsos y cada una de las otras dos impulsos, como se muestra en el cuadro 8.

Cuadro 8/G.722.2 – Posiciones posibles de impulsos individuales en el libro de códigos algebraico, 14,25 kbit/s

Pista	Impulso	Posiciones
1	i_0, i_4, i_8	0, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60
2	i_1, i_5, i_9	1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, 49, 53, 57, 61
3	i_2, i_6	2, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62
4	i_3, i_7	3, 7, 11, 15, 19, 23, 27, 31, 35, 39, 43, 47, 51, 55, 59, 63

Cada dos posiciones de impulsos de una pista se codifican con 8 bits (4 bits para la posición de cada impulso) y el signo del primer impulso de la pista se codifica con 1 bit.

Los tres impulsos de una pista se codifican con 13 bits.

Esto da un total de 44 bits (9+9+13+13) para el código algebraico.

5.8.1.6 Modo 12,65 kbit/s

En este libro de códigos, el vector de renovación contiene 8 impulsos distintos de cero. Todos los impulsos pueden tener las amplitudes +1 o -1. Las 64 posiciones de una subtrama se dividen en 4 pistas conteniendo cada una de las pistas dos impulsos, como se muestra en el cuadro 9.

Cuadro 9/G.722.2 – Posiciones posibles de impulsos individuales en el libro de códigos algebraico, 12,65 kbit/s

Pista	Impulso	Posiciones
1	i_0, i_4	0, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60
2	i_1, i_5	1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, 49, 53, 57, 61
3	i_2, i_6	2, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62
4	i_3, i_7	3, 7, 11, 15, 19, 23, 27, 31, 35, 39, 43, 47, 51, 55, 59, 63

Cada dos posiciones de impulso de una pista se codifican con 8 bits (32 bits en total, 4 bits para la posición de cada impulso) y el signo del primer impulso de la pista se codifica con 1 bit (4 bits en total). Esto da un total de 36 bits para el código algebraico.

5.8.1.7 Modo 8,85 kbit/s

En este libro de códigos, el vector de renovación contiene 4 impulsos distintos de cero. Todos los impulsos pueden tener las amplitudes +1 o -1. Las 64 posiciones de una subtrama se dividen en 4 pistas conteniendo cada una de las pistas un impulso, como se muestra en el cuadro 10.

Cuadro 10/G.722.2 – Posiciones posibles de impulsos individuales en el libro de códigos algebraico, 8,85 kbit/s

Pista	Impulso	Posiciones
1	i_0	0, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60
2	i_1	1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, 49, 53, 57, 61
3	i_2	2, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62
4	i_3	3, 7, 11, 15, 19, 23, 27, 31, 35, 39, 43, 47, 51, 55, 59, 63

Cada posición de impulso de una pista se codifica con 4 bits y el signo del impulso de la pista se codifica con 1 bit. Esto da un total de 20 bits para el código algebraico.

5.8.1.8 Modo 6,60 kbit/s

En este libro de códigos, el vector de renovación contiene 2 impulsos distintos de cero. Todos los impulsos pueden tener las amplitudes +1 o -1. Las 64 posiciones de una subtrama se dividen en 2 pistas conteniendo cada una de las pistas un impulso, como se muestra en el cuadro 11.

Cuadro 11/G.722.2 – Posiciones posibles de impulsos individuales en el libro de códigos algebraico, 6,60 kbit/s

Pista	Impulso	Posiciones
1	i_0	0, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62
2	i_1	1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63

Cada posición de impulso de una pista se codifica con 5 bits y el signo del impulso de la pista se codifica con 1 bit. Esto da un total de 12 bits para el código algebraico

5.8.2 Indexación de impulsos

En 5.8.1 se dio el número de bits necesarios para codificar un cierto número de impulsos en una pista. En esta cláusula se describen los procedimientos utilizados para codificar de 1 a 6 impulsos por pista. La descripción se hará para el caso de 4 pistas por subtrama, con 16 posiciones por pista y una separación entre impulsos de 4 (lo cual es el caso para todos los modos excepto el modo 6,6 kbits/s).

Codificación de 1 impulso con signo por pista

El índice de la posición del impulso se codifica con 4 bits y el índice del signo con 1 bit. El índice de la posición viene dado por la posición del impulso en la subtrama dividida por la separación entre impulsos (parte entera de la división). El resto de la división da el índice de la pista. Por ejemplo, un impulso en la posición 31 tiene un índice de posición de $31/4 = 7$ y pertenece a la pista de índice 3 (4ª pista).

El índice de signo se fija aquí en 0 para signos positivos y en 1 para signos negativos.

El índice del impulso con signo viene dado por:

$$I_{1p} = p + s \times 2^M$$

Donde p es el índice de posición, s es el índice de signo y $M = 4$ es el número de bits por pista.

Codificación de 2 impulsos con signo por pista

En el caso de 2 impulsos por pista de $K = 2^M$ posiciones posibles (aquí es $M = 4$), cada impulso necesita 1 bit para el signo y M bits para la posición, lo que da un total de $2M + 2$ bits. Sin embargo, existe un cierto grado de redundancia debido a que el orden de los impulsos es irrelevante. Por ejemplo, situar el primer impulso en la posición p y el segundo impulso en la posición q es equivalente a situar el primer impulso en la posición q y el segundo impulso en la posición p . Se puede ahorrar 1 bit codificando solamente un signo y deduciendo el segundo signo del orden a partir de las posiciones en el índice. El índice viene dado aquí por:

$$I_{2p} = p_1 + p_0 \times 2^M + s \times 2^{2M}$$

donde s es el índice de signo del impulso en el índice de posición p_0 . Si los dos signos son iguales, la posición menor se fija en p_0 y la mayor en p_1 . Por el contrario, si los dos signos no son iguales, la posición mayor se fija en p_0 y la posición menor se fija en p_1 . En el decodificador se obtiene fácilmente el signo del impulso en la posición p_0 . El segundo signo se deduce a partir del orden de los impulsos. Si p_0 es mayor que p_1 , el signo del impulso en la posición p_1 es el opuesto al de la posición p_0 . Si no es este el caso, los dos signos se fijan iguales.

Codificación de 3 impulsos con signo por pista

En el caso de 3 impulsos por pista se puede utilizar una lógica similar a la del caso de 2 impulsos. Para una pista con 2^M posiciones se necesitan $3M + 1$ bits en vez de $3M + 3$ bits. Una manera sencilla de indexar los impulsos consiste en dividir las posiciones de pista en dos secciones (o mitades) e identificar una sección que contiene al menos dos impulsos. El número de posiciones de esa sección es $K/2 = 2^M/2 = 2^{M-1}$, que se puede representar con $M-1$ bits. Los dos impulsos de la sección que contiene al menos dos impulsos se codifican con el procedimiento de codificación de 2 impulsos con signo que requiere $2(M-1) + 1$ bits y el impulso restante, que puede estar en cualquier lugar de la pista (en cualquiera de las dos secciones), se codifica con los $M + 1$ bits. Por último, el índice de la sección que contiene los dos impulsos se codifica con 1 bit. Así pues, el número total de bits requeridos $2(M-1) + 1 + M + 1 + 1 = 3M + 1$.

Una manera sencilla de comprobar si dos impulsos están situados en la misma sección consiste en verificar si los bits más significativos (MSB) de sus índices de posición son iguales o no. Se señala que un MSB de 0 significa que la posición pertenece a la mitad inferior de la pista (0-7) y un MSB de 1 significa que pertenece a la mitad superior (8-15). Si los dos impulsos pertenecen a la mitad superior, han de ser desplazados a la gama (0-7) antes de codificados utilizando $2 \times 3 + 1$ bits. Esto puede hacerse enmascarando los $M-1$ bits menos significativos (LSB) con una máscara formada por $M-1$ unos (que corresponde al número 7 en este caso).

El índice de los 3 impulsos con signo viene dado por:

$$I_{3p} = I_{2p} + k \times 2^{2M-1} + I_{1p} \times 2^{2M}$$

donde I_{2p} es el índice de los dos impulsos situados en la misma sección, k es el índice de la sección (0 ó 1), e I_{1p} es el índice del tercer impulso de la pista.

Codificación de 4 impulsos con signo por pista

Los 4 impulsos con signo de una pista de longitud $K = 2^M$ se pueden codificar utilizando $4M$ bits. De manera similar al caso de 3 impulsos, las K posiciones de la pista se dividen en dos secciones (o mitades) conteniendo cada sección $K/2 = 8$ posiciones. En este caso, se denomina sección A a la sección que tiene las posiciones 0 a $K/2-1$ y sección B a la sección que tiene las posiciones $K/2$ a $K-1$. Cada sección puede contener de 0 a 4 impulsos. El cuadro que sigue muestra los 5 casos que representan el número de impulsos posibles en cada sección:

Caso	Impulsos en la en la sección A	Impulsos en la sección B	Bits necesarios
0	0	4	$4M-3$
1	1	3	$4M-2$
2	2	2	$4M-2$
3	3	1	$4M-2$
4	4	0	$4M-3$

En los casos 0 ó 4, los 4 impulsos de una sección de longitud $K/2 = 2^{M-1}$ se pueden codificar utilizando $4(M-1) + 1 = 4M-3$ bits (esto se explicará más adelante).

En los casos 1 ó 3, el impulso de una sección de longitud $K/2 = 2^{M-1}$ se puede codifica con $M-1+1 = M$ bits y los 3 impulsos de la otra sección se pueden codificar con $3(M-1) + 1 = 3M-2$ bits. Esto da un total de $M+3M-2 = 4M-2$ bits.

En el caso 2, los impulsos de una sección de longitud $K/2 = 2^{M-1}$ se pueden codificar con $2(M-1)+1 = 2M-1$ bits. Así pues, para ambas secciones se requieren $2(2M-1) = 4M-2$ bits.

El índice de caso se puede codificar ahora con 2 bits (4 casos posibles) suponiendo que los casos 0 y 4 están combinados. El número de bits necesarios a continuación para los casos 1, 2 ó 3 es $4M-2$. Esto da un total de $4M-2+2 = 4M$ bits. Para los casos 0 ó 4 se necesita 1 bit para identificar uno u otro caso, y se necesitan $4M-3$ bits para codificar los 4 impulsos de la sección. Añadiendo los 2 bits necesarios para el caso general, esto da un total de $1+4M-3+2 = 4M$ bits.

El índice de los 4 impulsos con signo viene dado por:

$$I_{4p} = I_{AB} + k \times 2^{4M-2}$$

donde k es el índice de caso (2 bits) e I_{AB} es el índice de los impulsos en ambas secciones para cada uno de los casos.

Para los casos 0 y 1, I_{AB} viene dado por:

$$I_{AB_{0,4}} = I_{4p_section} + j \times 2^{4M-3}$$

donde j es un índice de 1 bit que identifica la sección con 4 impulsos e $I_{4p_section}$ es el índice de los 4 impulsos de esa sección (que requiere $4M-3$ bits).

Para el caso 1, I_{AB} viene dado por:

$$I_{AB_1} = I_{3p_B} + I_{1p_A} \times 2^{3(M-1)+1}$$

donde I_{3p_B} es el índice de los 3 impulsos de la sección B ($3(M-1)+1$ bits) e I_{1p_A} es el índice del impulso de la sección A ($(M-1)+1$ bits).

Para el caso 2, I_{AB} viene dado por:

$$I_{AB_2} = I_{2p_B} + I_{2p_A} \times 2^{2(M-1)+1}$$

donde I_{2p_B} es el índice de los 2 impulsos de la sección B ($2(M-1)+1$ bits) e I_{2p_A} es el índice de los 2 impulsos de la sección A ($2(M-1)+1$ bits).

Por último, para el caso 3, I_{AB} viene dado por:

$$I_{AB_3} = I_{1p_B} + I_{3p_A} \times 2^M$$

donde I_{1p_B} es el índice del impulso de la sección B ($(M-1)+1$ bits) e I_{3p_A} es el índice de los 3 impulsos de la sección A ($3(M-1)+1$ bits).

Según se indicó, para los casos 0 y 4, los 4 impulsos de una sección se codifican utilizando $4(M-1)+1$ bits. Esto se hace dividiendo nuevamente la sección en 2 subsecciones de longitud $K/4 = 2^{M-2}$ (= 4 en este caso); identificando una subsección que contiene al menos 2 impulsos; codificando los dos impulsos de esa subsección con $2(M-2)+1 = 2M-3$ bits; codificando el índice de la subsección que contiene al menos 2 impulsos con 1 bit y codificando los 2 impulsos restantes, suponiendo que pueden estar en cualquier lugar de la sección, con $2(M-1)+1 = 2M-1$ bits. Esto da un total de $(2M-3)+(1)+(2M-1) = 4M-3$ bits.

Codificación de 5 impulsos con signo por pista

Los 5 impulsos con signo de una pista de longitud $K = 2^M$ se pueden codificar utilizando $5M$ bits. De manera similar al caso de 4 impulsos, las K posiciones de la pista se dividen en dos secciones A y B. Cada sección puede contener de 0 a 5 impulsos. Un procedimiento sencillo de codificar los 5 impulsos consiste en identificar una sección que contiene al menos 3 impulsos, y codificar los 3 impulsos de esa sección utilizando $3(M-1)+1 = 3M-2$ bits, y codificar los 2 impulsos restantes en toda la pista utilizando $2M+1$ bits. Esto da $5M-1$ bits. Hace falta un bit adicional para identificar la sección que contiene al menos 3 impulsos. Así pues, se necesita un total de $5M$ bits para codificar los 5 impulsos con signo.

El índice de los 5 impulsos con signo viene dado por:

$$I_{5p} = I_{2p} + I_{3p} \times 2^{2M} + k \times 2^{5M-1}$$

donde k es el índice de la sección que contiene al menos 3 impulsos, I_{3p} es el índice de los 3 impulsos de esa sección ($3(M-1)+1$ bits) e I_{2p} es el índice de los 2 impulsos restantes en la pista ($2M+1$ bits).

Codificación de 6 impulsos con signo por pista

Los 6 impulsos con signo de una pista de longitud $K = 2^M$ se codifican utilizando $6M-2$ bits. De manera similar al caso de 5 impulsos, las K posiciones de una pista se dividen en 2 secciones A y B. Cada sección puede contener de 0 a 6 impulsos. El cuadro que sigue muestra los 7 casos que representan el número de impulsos posibles en cada sección:

Caso	Impulsos en la sección A	Impulsos en la sección B	Bits necesarios
0	0	6	$6M-5$
1	1	5	$6M-5$
2	2	4	$6M-5$
3	3	3	$6M-4$
4	4	2	$6M-5$
5	5	1	$6M-5$
6	6	0	$6M-5$

Se señala que los casos 0 y 6 son similares con la salvedad de que los 6 impulsos están en secciones diferentes. De manera similar, los casos 1 y 5 así como los casos 2 y 4 difieren solamente en la sección que contiene más impulsos. Por consiguiente, estos casos se pueden emparejar y se puede añadir un bit adicional para identificar la sección que contiene más impulsos. Puesto que estos casos necesitan inicialmente $6M-5$ bits, los casos emparejados necesitan $6M-4$ bits teniendo en cuenta el bit de sección. Así pues, ahora hay 4 estados de casos emparejados, esto es (0,6), (1,5), (2,4) y (3), con 2 bits adicionales necesarios para el estado. Esto da un total de $6M-4+2 = 6M-2$ bits para los 6 impulsos con signo.

En los casos 0 y 6 se necesita 1 bit para identificar la sección que contiene 6 impulsos. De esa sección, 5 impulsos se codifican utilizando $5(M-1)$ bits (puesto que los impulsos están limitados a esa sección), y el impulso restante se codifica utilizando $(M-1)+1$ bits. Así pues, se necesita un total de $1+5(M-1)+M = 6M-4$ bits para este caso emparejado. Se necesitan 2 bits adicionales para codificar el estado del caso emparejado, lo que da un total de $6M-2$ bits. El índice de los 6 impulsos de este caso emparejado viene dado por:

$$I_{6p} = I_{1p} + I_{5p} \times 2^M + j \times 2^{6M-5} + k \times 2^{6M-4}$$

donde k es el índice del caso emparejado (2 bits), j es el índice de la sección que contiene 6 impulsos (1 bit), I_{5p} es el índice de los 5 impulsos de esa sección ($5(M-1)$ bits) e I_{1p} es el índice del impulso de la otra sección ($(M-1)+1$ bits).

En los casos 1 y 5 se necesita 1 bit para identificar la sección que contiene 5 impulsos. Los 5 impulsos de esa sección se codifican utilizando $5(M-1)$ bits y el impulso de la otra sección se codifica utilizando $(M-1)+1$ bits. El índice de los 6 impulsos de este caso emparejado viene dado por:

$$I_{6p} = I_{1p} + I_{5p} \times 2^M + j \times 2^{6M-5} + k \times 2^{6M-4}$$

donde k es el índice del caso emparejado (2 bits), j es el índice de la sección que contiene 5 impulsos (1 bit), I_{5p} es el índice de los 5 impulsos de esa sección ($5(M-1)$ bits) e I_{1p} es el índice del impulso de la otra sección ($(M-1)+1$ bits).

En los casos 2 ó 4 se necesita 1 bit para identificar la sección que contiene 4 impulsos. Los 4 impulsos de esa sección se codifican utilizando $4(M-1)$ bits y los 2 impulsos de la otra sección se codifican utilizando $2(M-1)+1$ bits. El índice de los 6 impulsos de este caso emparejado viene dado por:

$$I_{6p} = I_{2p} + I_{4p} \times 2^{2(M-1)+1} + j \times 2^{6M-5} + k \times 2^{6M-4}$$

donde k es el índice del caso emparejado (2 bits), j es el índice de la sección que contiene 4 impulsos (1 bit), I_{4p} es el índice de los 4 impulsos de esa sección ($4(M-1)$ bits) e I_{2p} es el índice de los 2 impulsos de la otra sección ($2(M-1)+1$ bits).

En el caso 3, los 3 impulsos de cada sección se codifican utilizando $(3(M-1)+1)$ bits en cada sección. El índice de los 6 impulsos de este caso viene dado por:

$$I_{6p} = I_{3pB} + I_{3pA} \times 2^{3(M-1)+1} + k \times 2^{6M-4}$$

donde k es el índice del caso emparejado (2 bits), I_{3pB} es el índice de los 3 impulsos de la sección B ($3(M-1)+1$ bits), e I_{3pA} es el índice de los 3 impulsos de la sección A ($3(M-1)+1$ bits).

5.8.3 Búsqueda en libro de códigos

La búsqueda en el libro de códigos algebraico se lleva a cabo minimizando el error cuadrático medio entre la voz de entrada ponderada y la voz de síntesis ponderada. La señal objetivo utilizada en la búsqueda de altura de tono en bucle cerrado se actualiza sustrayendo la contribución del libro de códigos adaptativo. Esto es:

$$x_2(n) = x(n) - g_p y(n), \quad n = 0, \dots, 63 \quad (40)$$

donde $y(n) = v(n) * h(n)$ es el vector de libro de códigos adaptativo filtrado y g_p es la ganancia de libro de códigos adaptativo no cuantificada.

La matriz \mathbf{H} se define como la matriz de convoluciones Toeplitz triangular inferior con diagonal $h(0)$ y diagonales inferiores $h(1), \dots, h(63)$ y $\mathbf{d} = \mathbf{H}^t \mathbf{x}_2$ es la correlación entre la señal objetivo $x_2(n)$ y la respuesta en tren de impulsos $h(n)$ (a lo que también se denomina vector objetivo filtrado hacia atrás), y $\Phi = \mathbf{H}^t \mathbf{H}$ es la matriz de correlaciones de $h(n)$.

Los elementos del vector \mathbf{d} se calculan mediante la expresión:

$$d(n) = \sum_{i=n}^{63} x_2(i) h(i-n), \quad n = 0, \dots, 63 \quad (41)$$

y los elementos de la matriz simétrica Φ se calculan mediante la expresión:

$$\phi(i, j) = \sum_{n=j}^{63} h(n-i) h(n-j), \quad i = 0, \dots, 63 \quad j = i, \dots, 63 \quad (42)$$

Si \mathbf{c}_k es el vector-código algebraico de índice k , el libro de códigos algebraico se investiga maximizando el criterio de búsqueda siguiente:

$$Q_k = \frac{(\mathbf{x}_2^t \mathbf{H} \mathbf{c}_k)^2}{\mathbf{c}_k^t \mathbf{H}^t \mathbf{H} \mathbf{c}_k} = \frac{(\mathbf{d}^t \mathbf{c}_k)^2}{\mathbf{c}_k^t \Phi \mathbf{c}_k} = \frac{(R_k)^2}{E_k} \quad (43)$$

El vector \mathbf{d} y la matriz Φ se calculan normalmente antes de la búsqueda en el libro de códigos.

La estructura algebraica del libro de códigos permite aplicar procedimientos de búsqueda muy rápidos, ya que el vector de renovación \mathbf{c}_k contiene tan solo unos pocos impulsos distintos de cero. La correlación del numerador de la ecuación (43) viene dada por:

$$C = \sum_{i=0}^{N_p-1} a_i d(m_i) \quad (44)$$

donde m_i es la posición del impulso i -ésimo, a_i es su amplitud y N_p es el número de impulsos. La energía del denominador de la ecuación (43) viene dada por:

$$E = \sum_{i=0}^{N_p-1} \phi(m_i, m_i) + 2 \sum_{i=0}^{N_p-2} \sum_{j=i+1}^{N_p-1} a_i a_j \phi(m_i, m_j) \quad (45)$$

Para simplificar el procedimiento de búsqueda, las amplitudes de los impulsos se predeterminan de acuerdo con una señal de referencia determinada $b(n)$. En este procedimiento llamado de amplitud de impulso seleccionado por la señal, el signo de un impulso en la posición i se fija igual al signo de la señal de referencia en esa posición. En el caso presente, la señal de referencia $b(n)$ viene dada por:

$$b(n) = \sqrt{\frac{E_d}{E_r}} r_{LTP}(n) + \alpha d(n) \quad (46)$$

donde $E_d = \mathbf{d}^t \mathbf{d}$ es la energía de la señal $d(n)$ y $E_r = \mathbf{r}_{LTP}^t \mathbf{r}_{LTP}$ es la energía de la señal $r_{LTP}(n)$ que es la señal residual tras la predicción a largo plazo. El factor de escalaje α controla el grado de dependencia de la señal de referencia con respecto a $d(n)$ y se reduce a medida que aumenta la velocidad binaria. En este caso, $\alpha = 2$ para los modos 6,6 y 8,85; $\alpha = 1$ para los modos 12,65, 14,25, y 15,85; $\alpha = 0,8$ para el modo 18,25; $\alpha = 0,75$ para el modo 19,85 y $\alpha = 0,5$ para los modos 23,05 y 23,85.

Para simplificar la búsqueda, se modifican la señal $d(n)$ y la matriz Φ de modo que incorporen los modos preseleccionados. Sea $s_b(n)$ el vector que contiene los signos de $b(n)$. La señal modificada $d'(n)$ viene dada por:

$$d'(n) = s_b(n) d(n) \quad n=0, \dots, N-1$$

y la matriz de autocorrelaciones modificada Φ' viene dada por:

$$\phi'(i, j) = s_b(i) s_b(j) \phi(i, j), \quad i = 0, \dots, N-1; \quad j = i, \dots, N-1$$

La correlación del numerador del criterio de búsqueda Q_k es ahora como sigue:

$$R = \sum_{i=0}^{N_p-1} d'(i)$$

y la energía del denominador del criterio de búsqueda Q_k es entonces:

$$E = \sum_{i=0}^{N_p-1} \phi'(m_i, m_i) + 2 \sum_{i=0}^{N_p-2} \sum_{j=i+1}^{N_p-1} \phi'(m_i, m_j)$$

El objetivo de la búsqueda es ahora determinar el vector-código que tiene el mejor conjunto de posiciones de impulso N_p suponiendo que las amplitudes de los impulsos han sido seleccionadas como se ha descrito anteriormente. El criterio de selección básico es la maximización de la relación Q_k mencionada más arriba.

Para reducir la complejidad de la búsqueda se utiliza un procedimiento de búsqueda conocido como procedimiento de búsqueda en árbol de primera profundidad, en el que se determinan a la vez las posiciones de N_m impulsos. De manera más precisa, los N_p impulsos disponibles se reparten entre M subconjuntos no vacíos de N_m impulsos respectivamente de tal modo que $N_1+N_2+\dots+N_m+\dots+N_M = N_p$. Una determinada elección de posiciones de los primeros $J = N_1+N_2+\dots+N_{M-1}$ impulsos considerados se denomina trayecto de nivel m o trayecto de longitud J . El criterio básico para un trayecto de J posiciones de impulso es la relación $Q_k(J)$ cuando sólo se consideran los J impulsos pertinentes.

La búsqueda empieza en el subconjunto 1 y avanza con los subconjuntos siguientes de acuerdo con una estructura en árbol en la que el subconjunto m se busca en el m -ésimo nivel del árbol. La búsqueda en el nivel 1 tiene por objeto considerar los N_1 impulsos del subconjunto 1 y sus posiciones válidas para determinar uno o varios trayectos candidatos de longitud N_1 que son los nodos del árbol en el nivel 1. El trayecto de cada nodo de terminación al nivel $m-1$ se amplía a la longitud $N_1+N_2+\dots+N_m$ al nivel m considerando N_m impulsos nuevos y sus posiciones válidas. Se determinan uno o varios trayectos candidatos ampliados de modo que constituyan nodos de nivel m . El mejor vector-código corresponde al trayecto de longitud N_p que maximiza el criterio $Q_k(N_p)$ con respecto a todos los nodos del nivel M .

Aquí se utiliza una forma especial del procedimiento de búsqueda en árbol de primera profundidad, en el que se buscan dos impulsos a la vez, es decir, $N_m = 2$, y esos dos impulsos pertenecen a dos pistas consecutivas. Además, en vez de suponer que la matriz Φ ha sido precalculada y almacenada, lo que requiere una memoria de $N \times N$ palabras ($64 \times 64 = 4k$ palabras), se aplica un procedimiento más eficaz desde el punto de vista de la memoria que reduce la necesidad de memoria. En ese procedimiento, la búsqueda se desarrolla de tal manera que sólo una parte de los elementos necesarios de la matriz de correlaciones se precalcula y almacena. Esa parte corresponde a las correlaciones de la respuesta en tren de impulsos correspondiente a las posiciones de impulso posibles en pistas consecutivas, así como a las correlaciones correspondientes a $\phi(j,j)$, $j = 0, \dots, N-1$ (esto es, los elementos de la diagonal principal de la matriz Φ).

A fin de reducir la complejidad del procedimiento se prueba un número limitado de posiciones posibles del primer impulso, al tiempo que se prueban combinaciones posibles de dos impulsos. Además, en caso de un número elevado de impulsos, se fijan algunos impulsos de los niveles superiores del árbol de búsqueda. Para estimar de manera precisa qué posiciones de impulso posibles se consideran para el primer impulso o bien para fijar algunas posiciones de impulso, se utiliza un "vector de estimación de la probabilidad de la posición del impulso" \mathbf{b} , que se basa en señales relacionadas con la voz. El p -ésimo componente $b(p)$ de este vector de estimación \mathbf{b} caracteriza la probabilidad de que un impulso ocupe la posición p ($p = 0, 1, \dots, N-1$) en el mejor vector-código que se está buscando. El vector de estimación \mathbf{b} es aquí el mismo vector que se utiliza para preseleccionar las amplitudes y que da la ecuación (46).

Los procedimientos de búsqueda son similares para todas las velocidades binarias. Se buscan dos impulsos al mismo tiempo, y esos dos impulsos corresponden siempre a pistas consecutivas. Esto es, los dos impulsos buscados están en las pistas T_0-T_1 , T_1-T_2 , T_2-T_3 , or T_3-T_0 .

Antes de buscar las posiciones, el signo de un impulso en una posición posible n se hace igual al signo de $b(n)$ en esa posición. A continuación se calcula la señal modificada $d'(n)$ incluyendo los signos predeterminados como se ha descrito anteriormente.

Para los 2 primeros impulsos (1^{er} nivel del árbol), la correlación del numerador del criterio de búsqueda viene dada por:

$$R = d'(m_0) + d'(m_1)$$

y la energía del denominador del criterio de búsqueda Q_k viene dada por:

$$E = \phi'(m_0, m_0) + \phi'(m_1, m_1) + 2\phi'(m_0, m_1)$$

donde las correlaciones $\phi'(m_i, m_j)$ han sido modificadas para incluir los signos preseleccionados en las posiciones m_i y m_j .

Para niveles subsiguientes se actualizan el numerador y el denominador añadiendo la contribución de dos impulsos nuevos. Suponiendo que se buscan dos impulsos nuevos en un nivel de árbol determinado con posiciones m_k y m_{k+1} de dos pistas consecutivas, el valor actualizado de R es como sigue:

$$R = R + d'(m_k) + d'(m_{k+1}) \quad (47)$$

y la energía actualizada es:

$$E = E + \phi'(m_k, m_k) + \phi'(m_{k+1}, m_{k+1}) + 2\phi'(m_k, m_{k+1}) + 2R_{hv}(m_k) + 2R_{hv}(m_{k+1}) \quad (48)$$

donde $R_{hv}(m)$ es la correlación entre la respuesta en tren de impulsos $h(n)$ y un vector $v_h(n)$ que contiene la adición de versiones demoradas de respuesta en tren de impulsos en las posiciones determinadas previamente. Esto es:

$$v_h(n) = \sum_{i=0}^{k-1} h(n - m_i)$$

y

$$R_{hv}(m) = \sum_{n=m}^{N-1} h(n)v_h(n-m)$$

En cada nivel del árbol se calculan de manera directa e inmediata los valores de $R_{hv}(m)$ para todas las posiciones posibles de cada una de las dos pistas que se prueban. De la ecuación (48) se deduce que sólo es necesario almacenar las correlaciones $\phi'(m_k, m_{k+1})$ correspondientes a posiciones de impulso de dos pistas consecutivas ($4 \times 16 \times 16$ palabras), junto con las correlaciones $\phi'(m_k, m_k)$ correspondientes a la matriz Φ (64 palabras). Así pues, la necesidad de memoria en la estructura algebraica presente es de 1088 palabras en vez de $64 \times 64 = 4096$ palabras.

Los procedimientos de búsqueda de los diversos modos de velocidades binarias son similares. La diferencia está en el número de impulsos y, en consecuencia, el número de niveles de la búsqueda en árbol. Para mantener una complejidad de búsqueda comparable a lo largo de los diferentes libros de códigos, se hace que el número de posiciones probadas sea similar.

A continuación se describe, a título de ejemplo, la búsqueda en el modo 12,65 kbit/s. En este modo se sitúan 2 impulsos en cada pista, lo que da un total de 8 impulsos por subtrama de longitud 64. Se buscan 2 impulsos a la vez, y esos 2 impulsos corresponden siempre a pistas consecutivas. Esto es, los 2 impulsos buscados están en las pistas T_0-T_1 , T_1-T_2 , T_2-T_3 , o T_3-T_0 . El árbol tiene 4 niveles en este caso. En el primer nivel se asigna el impulso P_0 a la pista T_0 y el impulso P_1 a la pista T_1 . En este nivel no se efectúa ninguna búsqueda y las dos posiciones de impulso se fijan en el máximo de $b(n)$ en cada pista. En el segundo nivel se asigna el impulso P_2 a la pista T_2 y el impulso P_3 a la pista T_3 . Se prueban 4 posiciones para el impulso P_2 frente a las 16 posiciones del impulso P_3 . Las 4 posiciones probadas de P_2 se determinan en base a los máximos de $b(n)$ en la pista. En el tercer nivel se asigna el impulso P_4 a la pista T_1 y el impulso P_5 a la pista T_2 . Se prueban 8 posiciones para el impulso P_4 frente a las 16 posiciones del impulso P_5 . De manera similar al nivel de búsqueda previo, las 8 posiciones probadas de P_4 se determinan en base a los máximos de $b(n)$ en la pista. En el cuarto nivel se asigna el impulso P_6 a la pista T_3 y el impulso T_7 a la pista T_0 . Se prueban 8 posiciones para el impulso P_6 frente a las 16 posiciones del impulso P_7 . Así pues, el número total de combinaciones probadas es $4 \times 16 + 8 \times 16 + 8 \times 16 = 320$. El proceso completo se repite 4 veces (4 iteraciones) asignando los impulsos a pistas diferentes. Por ejemplo, en la segunda

iteración, los impulsos P_0 a P_7 se asignan a las pistas $T_1, T_2, T_3, T_0, T_2, T_3, T_0$ y T_1 , respectivamente. De esta manera, el número total de combinaciones de posiciones probadas es $4 \times 320 = 1280$.

También a título de ejemplo de búsqueda, en el modo 15,85 kbit/s, se sitúan 3 impulsos en cada pista, lo que da un total de 12 impulsos. Hay 6 niveles de búsqueda en árbol y se buscan 2 impulsos en cada nivel. En los 2 primeros niveles se fijan 4 impulsos a los máximos de $b(n)$. En los 4 niveles siguientes, el número de combinaciones probadas es $4 \times 16, 6 \times 16, 8 \times 16$ y 8×16 , respectivamente. Se efectúan 4 iteraciones, lo que da un total de $4 \times 26 \times 16 = 1664$ combinaciones.

5.9 Cuantificación de las ganancias de libros de códigos adaptativo y fijo

Las ganancias de libros de códigos adaptativo (ganancia de altura de tono) y fijo (algebraico) se cuantifican utilizando un libro de códigos de 6 bits para los modos 8,85 y 6,60 kbit/s y utilizando un libro de códigos de 7 bits para todos los demás modos.

La cuantificación de la ganancia de libro de códigos fijo se efectúa utilizando la predicción MA con coeficientes fijos. La predicción MA de 4º orden se efectúa en la energía de renovación como sigue. Sea $E(n)$ la energía de renovación (en dB) con valor medio suprimido en la subtrama n , y dada por:

$$E(n) = 10 \log \left(\frac{1}{N} g_c^2 \sum_{i=0}^{N-1} c^2(i) \right) - \bar{E} \quad (49)$$

donde $N = 64$ es el tamaño de subtrama, $c(i)$ es la excitación de libro de códigos fijo y $\bar{E} = 30$ dB es el valor medio de la energía de renovación. La energía viene dada por:

$$\tilde{E}(n) = \sum_{i=1}^4 b_i \hat{R}(n-i) \quad (50)$$

donde $[b_1 \ b_2 \ b_3 \ b_4] = [0,5, 0,4, 0,3, 0,2]$ son los coeficientes de predicción MA y $\hat{R}(k)$ es el error de predicción de energía cuantificado en la subtrama k . La energía predicha se utiliza para calcular una ganancia de libro de códigos fijo predicha g'_c como en la ecuación (49) [sustituyendo $E(n)$ por $\tilde{E}(n)$ y g_c por g'_c], lo que se hace como se indica a continuación. Primero se obtiene la energía de renovación media aplicando la fórmula:

$$E_i = 10 \log \left(\frac{1}{N} \sum_{i=0}^{N-1} c^2(i) \right) \quad (51)$$

y a continuación se obtiene la ganancia g'_c

$$g'_c = 10^{0,05(\tilde{E}(n) + \bar{E} - E_i)} \quad (52)$$

El factor de corrección entre la ganancia g_c y la ganancia estimada g'_c viene dado por:

$$\gamma = g_c / g'_c \quad (53)$$

El error de predicción es:

$$R(n) = E(n) - \tilde{E}(n) = 20 \log(\gamma) \quad (54)$$

Los vectores de ganancia de altura de tono, g_p , y el factor de corrección γ se cuantifican conjuntamente utilizando un libro de códigos de 6 bits para los modos 8,85 y 6,60 kbit/s, y un libro de códigos de 7 bits para los demás modos. La búsqueda en el libro de códigos de ganancia se lleva a cabo minimizando la media cuadrática del error ponderado entre la voz original y la voz reconstruida, que viene dado por:

$$E = x^t x + g_p^2 y^t y + g_c^2 z^t z - 2g_p x^t y - 2g_c x^t z + 2g_p g_c y^t z \quad (55)$$

donde x es el vector objetivo, y es el vector de libro de códigos adaptativo filtrado y z es el vector de libro de códigos fijo filtrado. (Cada vector de ganancia de libro de códigos tiene además un elemento que representa el error de predicción de la energía cuantificado.) El error de predicción de la energía cuantificado asociado con las ganancias elegidas se utiliza para actualizar $\hat{R}(n)$. En la búsqueda sólo se tienen en cuenta los 64 vectores-código que están más próximos a la ganancia de altura de tono no cuantificada, g_p .

5.10 Actualización de la memoria

Se necesita una actualización de los estados de los filtros de síntesis y ponderación para calcular la señal objetivo de la subtrama siguiente.

Una vez cuantificadas las dos ganancias, la señal de excitación $u(n)$ de la subtrama presente viene dada por:

$$u(n) = \hat{g}_p v(n) + \hat{g}_c c(n), \quad n = 0, \dots, 63 \quad (56)$$

donde \hat{g}_p y \hat{g}_c son las ganancias cuantificadas de libros de códigos adaptativo y fijo, respectivamente, $v_i(n)$ es el vector de libro de códigos adaptativo (excitación pasada interpolada), y $c(n)$ es el vector de libro de códigos fijo (código algebraico que incluye la agudización de la altura del tono). Los estados de los filtros se pueden actualizar filtrando la señal $r(n) - u(n)$ (diferencia entre señal residual y señal de excitación) a través de los filtros $1/\hat{A}(z)$ y $A(z/\gamma_1)H_{\text{de-emph}}(z)$ de la subtrama de 64 muestras y guardando los estados de los filtros. Para esto harían falta 3 filtrados. Un procedimiento más sencillo, que sólo requiere un filtrado, es como sigue. La voz de síntesis local, $\hat{s}(n)$, se calcula filtrando la señal de excitación a través de $1/\hat{A}(z)$. La salida del filtro debida a la entrada $r(n) - u(n)$ es equivalente a $e(n) = s(n) - \hat{s}(n)$. Con esto, los estados del filtro de síntesis $1/\hat{A}(z)$ vienen dados por $e(n)$, $n = 48, \dots, 63$. La actualización de los estados del filtro $A(z/\gamma_1)H_{\text{de-emph}}(z)$ se puede hacer filtrando la señal de error $e(n)$ a través de este filtro para obtener el error ponderado perceptualmente $e_w(n)$. No obstante, la señal $e_w(n)$ se puede obtener de manera equivalente aplicando la fórmula:

$$e_w(n) = x(n) - \hat{g}_p y(n) = \hat{g}_c z(n) \quad (57)$$

Puesto que se dispone de las señales $x(n)$, $y(n)$, y $z(n)$, los estados del filtro de ponderación se actualizan calculando $e_w(n)$ mediante la ecuación (54) para $n = 48, \dots, 63$. De esta mane se ahorran dos filtrados.

5.11 Generación de ganancia de banda alta

Para calcular la ganancia de banda alta del modo 23,85 kbit/s, se filtra voz de entrada a 16 kHz a través de un filtro FIR pasobanda $H_{HB}(z)$ que tiene una banda de paso de 6,4 a 7 kHz. La ganancia de banda alta g_{HB} se obtiene aplicando la fórmula:

$$g_{HB} = \frac{\sum_{i=0}^{63} (s_{HB}(i))^2}{\sum_{i=0}^{63} (s_{HB2}(i))^2} \quad (58)$$

donde $s_{HB}(i)$ es la voz de entrada filtrada pasobanda y $s_{HB2}(i)$ es la síntesis de la voz de banda alta obtenida a partir de la excitación de banda alta $u_{HB2}(i)$ filtrada a través del filtro de síntesis de banda alta $A_{HB}(z)$ que se describe en 6.3.2.2.

6 Descripción funcional del decodificador

La función del decodificador consiste en decodificar los parámetros transmitidos (parámetros LP, vector de libro de códigos adaptativo, ganancia de libro de códigos adaptativo, vector de libro de códigos fijo y ganancia de banda alta) y efectuar la síntesis para obtener la voz reconstruida. A continuación se postprocesa la voz reconstruida y se le aplica un muestreo en sentido ascendente (y un escalaje en sentido ascendente). Por último, se genera la señal de banda alta en la banda de frecuencias de 6 a 7 kHz. En la figura 3 se muestra el flujo de la señal en el decodificador.

6.1 Decodificación y síntesis de la voz

El proceso de decodificación se efectúa en el siguiente orden:

Decodificación de parámetros de filtro LP: Los índices recibidos de la cuantificación ISP se utilizan para reconstruir el vector ISP cuantificado. Se lleva a cabo la interpolación descrita en 5.2.6 para obtener 4 vectores ISP interpolados (correspondientes a 4 subtramas). Para cada subtrama, el vector ISP interpolado se convierte al dominio de coeficientes de filtro LP a_k , que se utiliza para sintetizar la voz reconstruida en la subtrama.

Para cada subtrama se repiten los siguientes pasos:

- 1) **Decodificación del vector de libro de códigos adaptativo:** Se utiliza el índice de altura de tono recibido (índice de libro de códigos adaptativo) para hallar las partes entera y fraccionaria del retardo de altura de tono. El vector de libro de códigos adaptativo $v(n)$ se encuentra por interpolación de la señal de excitación pasada $u(n)$ (en el retardo de altura de tono) utilizando el filtro FIR descrito en 5.6. El índice de filtro adaptativo recibido se emplea para determinar si el libro de códigos adaptativo filtrado es $v_1(n) = v(n)$ o $v_2(n) = 0,18v(n) + 0,64v(n-1) + 0,18v(n-2)$.
- 2) **Decodificación del vector renovador:** Se utiliza el índice de libro de códigos algebraico recibido para extraer las posiciones y amplitudes (signos) de los impulsos de excitación y hallar el vector-código algebraico $c(n)$. Si la parte entera del retardo de altura de tono es inferior al tamaño de subtrama 64, se aplica el procedimiento de agudización de altura de tono equivalente a modificar $c(n)$ filtrándolo a través del prefiltro adaptativo $F(z)$, que consta de dos partes: una parte de mejora de la periodicidad $1/(1-0,85z^{-T})$ y una parte de inclinación $(1 - \beta_1 z^{-1})$, donde T es la parte entera del retardo de altura de tono y $\beta_1(n)$ tiene relación con la sonoridad de la subtrama previa y está limitado por $[0,0, 0,5]$.
- 3) **Decodificación de las ganancias de libros de códigos adaptativo y renovadores:** El índice recibido da el factor de corrección de ganancia de libro de códigos fijo $\hat{\gamma}$. La ganancia estimada de libro de códigos fijo g'_c , se obtiene como se describe en 5.8. Primero se obtiene la energía predicha de cada subtrama n aplicando la fórmula:

$$\tilde{E}(n) = \sum_{i=1}^4 b_i \hat{R}(n-i) \quad (59)$$

y a continuación se obtiene la energía de renovación media aplicando la fórmula:

$$E_i = 10 \log \left(\frac{1}{N} \sum_{i=0}^{N-1} c^2(i) \right) \quad (60)$$

La ganancia predicha g'_c viene dada por:

$$g'_c = 10^{0.05(\tilde{E}(n) + \bar{E} - E_i)} \quad (61)$$

La ganancia de libro de códigos fijo cuantificada viene dada por:

$$\hat{g}_c = \hat{\gamma} g'_c \quad (62)$$

- 4) **Cálculo de la voz reconstruida:** Los pasos que siguen son para $n = 0, \dots, 63$. La señal de excitación total se construye aplicando:

$$u(n) = \hat{g}_p v(n) + \hat{g}_c c(n) \quad (63)$$

Antes de la síntesis de la voz se lleva a cabo un postprocesamiento de elementos de excitación.

- 5) **Procesamiento antidispersión (modos 6,60 y 8,85 kbit/s):** Se aplica un procedimiento de postprocesamiento antidispersión adaptativo al vector de libro de códigos fijo $c(n)$ para reducir los artefactos perceptuales debidos a la dispersión de los vectores de libro de códigos fijo algebraico con tan sólo unas pocas muestras distintas de cero por subtrama. El procesamiento antidispersión consiste en una convolución circular del vector de libro de códigos fijo con una respuesta en tren de impulsos. Se utilizan tres respuestas en tren de impulsos prealmacenadas y se fija un número $impNr = 0,1,2$ para seleccionar una de ellas. Un valor de 2 corresponde a no modificación, un valor de 1 corresponde a modificación media y un valor de 0 corresponde a modificación fuerte. La selección de la respuesta en tren de impulsos se lleva a cabo de forma adaptativa a partir de las ganancias de libros de códigos adaptativo y fijo. Se emplea el siguiente procedimiento:

si $\hat{g}_p < 0,6$ entonces

$$impNr = 0;$$

en caso contrario y si $\hat{g}_p < 0,9$ entonces

$$impNr = 1;$$

en los demás casos

$$impNr = 2;$$

Hay que detectar el comienzo comparando la ganancia de libro de códigos fijo con la ganancia de libro de códigos fijo previa. Si el valor actual es más de tres veces el valor previo, se ha detectado un comienzo.

Si no hay comienzo e $impNr = 0$, se calcula el valor filtrado mediano de las ganancias de libros de códigos adaptativo actual y las cuatro previas. Si el valor es inferior a 0,6, $impNr = 0$.

Si no hay comienzo, el valor de $impNr$ se incrementa simplemente en un paso con respecto a la subtrama previa.

Si se declara un comienzo, el valor de $impNr$ se incrementa en una unidad cuando sea inferior a 2.

En el caso del modo 8,85/kbit/s, el valor de $impNr$ se incrementa en una unidad.

- 6) **Intensificador de ruido:** Se aplica una técnica de suavización de la ganancia no lineal a la ganancia de libro de códigos fijo \hat{g}_c para intensificar la excitación del ruido. En base a la estabilidad y la sonoridad del segmento vocal, se suaviza la ganancia de libro de códigos para reducir la fluctuación de la energía de excitación en caso de señales estacionarias. De esta manera se mejora la calidad de funcionamiento cuando el ruido de fondo sea estacionario.

El factor de sonoridad viene dado por $\lambda = 0,5(1-r_v)$ con $r_v = (E_v - E_c)/(E_v + E_c)$, donde E_v y E_c son las energías del vector-código de altura de tono escalado y el vector-código de renovación escalado, respectivamente. Se señala que, al estar el valor de r_v comprendido entre -1 y $+1$, el valor de λ está entre 0 y 1 . Se señala además que el factor λ está relacionado con el grado de ausencia de señales vocales con un valor de 0 para segmentos puramente vocales y un valor de 1 para segmentos puramente no vocales.

Se calcula un factor de estabilidad θ de acuerdo con una medida de la distancia entre los filtros LP adyacentes. El factor θ está relacionado aquí con la medida de la distancia ISP y está limitado por $0 \leq \theta \leq 1$, correspondiendo los valores de θ mayores a señales más estables.

Por último, el factor de suavización de la ganancia S_m viene dado por:

$$S_m = \lambda\theta \quad (64)$$

El valor de S_m se aproxima a 1 para señales no vocales y estables, que es el caso de las señales de ruido de fondo estacionario. Para señales puramente vocales o señales inestables, el valor de S_m se aproxima a 0 .

Se calcula una ganancia modificada inicial g_0 comparando la ganancia de libro de códigos fijo \hat{g}_c con un umbral dado por la ganancia modificada inicial de la subtrama previa g_{-1} . Si \hat{g}_c es mayor o igual que g_{-1} , se calcula g_0 decrementando \hat{g}_c en $1,5$ dB con el límite de $g_0 \geq g_{-1}$. Si \hat{g}_c es menor que g_{-1} , se calcula g_0 incrementando \hat{g}_c en $1,5$ dB con el límite de $g_0 \leq g_{-1}$.

Por último, se actualiza la ganancia con el valor de la ganancia suavizada como sigue:

$$\hat{g}_c = \theta g_0 + (1 + \theta)\hat{g}_c \quad (65)$$

- 7) **Intensificador de altura de tono:** Un procedimiento intensificador de altura de tono modifica la señal de excitación total $u(n)$ filtrando la señal de excitación de libro de códigos fijo a través de un filtro de renovación cuya respuesta en frecuencia enfatiza las frecuencias superiores más que las frecuencias inferiores y cuyos coeficientes están relacionados con la periodicidad de la señal. El filtro es de la forma:

$$F_{inno}(z) = -c_{pe}z + 1 - c_{pe}z^{-1} \quad (66)$$

donde $c_{pe} = 0,125(1 - r_v)$, con $r_v = (E_v - E_c)/(E_v + E_c)$ como se ha descrito anteriormente. El vector-código fijo filtrado viene dado por:

$$c'(n) = c(n) - c_{pe}(c(n+1) + c(n-1)) \quad (67)$$

y la excitación actualizada viene dada por:

$$u(n) = \hat{g}_p v(n) + \hat{g}_c c'(n) \quad (68)$$

El procedimiento anterior se puede desarrollar en un solo paso actualizando la excitación como sigue:

$$u(n) = u(n) - \hat{g}_c c_{pe}(c(n+1) + c(n-1)) \quad (69)$$

- 8) **Postprocesamiento de elementos de excitación (modo 6,60 kbit/s):** Se aplica un procedimiento de postprocesamiento de elementos de excitación a la excitación total $u(n)$ enfatizando (resaltando) la contribución del vector de libro de códigos adaptativo:

$$\hat{u}(n) = \begin{cases} u(n) + 0,25\beta\hat{g}_p v(n), & \hat{g}_p > 0.5 \\ u(n), & \hat{g}_p \leq 0.5 \end{cases} \quad (70)$$

Se utiliza el control de ganancia adaptativa (AGC, *adaptive gain control*) para compensar la diferencia de ganancia entre la excitación no enfatizada $u(n)$ y la excitación enfatizada $\hat{u}(n)$. El factor de escalaje de ganancia η de la excitación enfatizada se calcula como sigue:

$$\eta = \begin{cases} \sqrt{\frac{\sum_{n=0}^{63} u^2(n)}{\sum_{n=0}^{63} \hat{u}^2(n)}}, & \hat{g}_p > 0.5 \\ 1.0, & \hat{g}_p \leq 0.5 \end{cases} \quad (71)$$

La señal de excitación enfatizada con escalaje de ganancia $\hat{u}'(n)$ viene dada por:

$$\hat{u}'(n) = \hat{u}(n)\eta \quad (72)$$

La voz reconstruida de la subtrama de tamaño 64 viene dada por:

$$\hat{s}(n) = \hat{u}'(n) - \sum_{i=1}^{16} \hat{a}_i \hat{s}(n-i), \quad n = 0, \dots, 63 \quad (73)$$

donde \hat{a}_i son los coeficientes de filtro LP interpolados.

A continuación se pasa la voz de síntesis $\hat{s}(n)$ a través de un postprocesamiento adaptativo que se describe en la cláusula siguiente.

6.2 Filtrado paso alto, escalaje en sentido ascendente e interpolación

El filtro paso alto sirve a modo de salvaguardia contra componentes de baja frecuencia no deseados. La señal es filtrada a través del filtro paso alto $H_{h1}(z)$ y el filtro de desenfatización $H_{de_emph}(z)$.

Por último, se aplica a la señal un muestreo en sentido ascendente a 16 kHz para obtener la señal de síntesis de banda inferior $\hat{s}_{16k}(n)$. $\hat{s}_{16k}(n)$ se produce aplicando primero a la síntesis de banda inferior $\hat{s}_{12,8k}(n)$ un muestreo a 12,8 Hz e sentido ascendente por un factor de 5, filtrando a continuación la salida a través de $H_{decim}(z)$, y aplicando finalmente a la salida un muestreo en sentido descendente por un factor de 4.

(El escalaje en sentido ascendente consiste en multiplicar la salida del filtrado paso alto por un factor de 2 para compensar el escalaje en sentido descendente de la etapa de preprocesamiento.)

6.3 Banda de altas frecuencias

En la banda de frecuencias superiores (6,4-7,0 kHz), la excitación se genera para modelar las frecuencias más altas. El contenido de alta frecuencia se genera llenando la parte superior del espectro con un ruido blanco convenientemente escalado en el dominio de la excitación y convertido a continuación al dominio de la voz configurándolo con un filtro obtenido a partir del mismo filtro de síntesis LP utilizado para sintetizar la señal a la que se ha aplicado un muestreo en sentido descendente.

6.3.1 Generación de excitación de banda alta

La excitación de banda alta se obtiene generando primero ruido blanco $u_{HB1}(n)$. La potencia de la excitación de banda alta se fija iguala a la potencia de la excitación de banda inferior $u_2(n)$, lo que significa que:

$$u_{HB2}(n) = u_{HB1}(n) \sqrt{\frac{\sum_{k=0}^{63} u_2^2(k)}{\sum_{k=0}^{63} u_{HB1}^2(k)}} \quad (74)$$

Por último, la excitación de banda alta viene dada por:

$$u_{HB}(n) = \hat{g}_{HB} u_{HB2}(n) \quad (75)$$

donde \hat{g}_{HB} es un factor de ganancia.

En el modo 23,85 kbit/s, \hat{g}_{HB} se decodifica a partir del índice de ganancia recibido.

En los modos 6,60, 8,85, 12,65, 14,25, 15,85, 18,25, 19,85 y 23,05 kbit/s, g_{HB} se estima utilizando información de habla limitada por $[0,1, 1,0]$. Primero se obtiene la inclinación de síntesis e_{tilt}

$$e_{\text{tilt}} = \frac{\sum_{n=1}^{63} \hat{s}_{hp}(n) \hat{s}_{hp}(n-1)}{\sum_{n=0}^{63} \hat{s}_{hp}^2(n)} \quad (76)$$

donde $\hat{s}_{hp}(n)$ es la síntesis de la voz de banda baja filtrada paso alto $\hat{s}_{12,8k}(n)$ con frecuencia de corte de 400 Hz. La ganancia \hat{g}_{HB} viene dada entonces por:

$$g_{HB} = w_{SP} g_{SP} + (1 - w_{SP}) g_{BG} \quad (77)$$

donde $g_{SP} = 1 - e_{\text{tilt}}$ es la ganancia de la señal vocal, $g_{BG} = 1,25g_{SP}$ es la ganancia de la señal de ruido de fondo y w_{SP} es una función de ponderación fijada a 1, cuando el VAD está activado (ON) y 0 cuando el VAD está desactivado (OFF). Los límites de g_{HB} son $[0,1, 1,0]$. En el caso de segmentos de habla en los que está presente menos energía a altas frecuencias, e_{tilt} se aproxima a 1 con el resultado de una ganancia g_{HB} menor. Así se reduce la energía del ruido generado en el caso de segmentos de habla.

6.3.2 Filtro LP para la banda de altas frecuencias

6.3.2.1 Modo 6,60 kbit/s

El filtro de síntesis LP de banda alta $A_{HB}(z)$ se obtiene extrapolando el vector ISF cuantificado \mathbf{f} en el vector ISF de 20° orden \mathbf{f}_e . Primero se obtiene el valor máximo de la autocorrelación $C_{\max}(i)$ del vector diferencia de vectores ISF $f_{\Delta}(i) = f(i+1) - f(i), i = 1, \dots, 14$. A continuación se calcula un nuevo vector ISF a 16 kHz, $f'_e(i)$, como sigue:

$$f'_e(i) = \begin{cases} f(i-1), & i = 1, \dots, 15 \\ f'_e(i-1) + f'_e(i - C_{\max}(i) - 1) - f'_e(i - C_{\max}(i) - 2), & i = 16, \dots, 19 \end{cases} \quad (78)$$

Una aproximación del último elemento del nuevo vector ISF f_{e19} es actualizada en base a los coeficientes de frecuencia más baja. El nuevo vector diferencia de vectores ISF extrapolado, $f'_{e\Delta}(i)$ es:

$$f'_{e\Delta}(i) = c_{scale}(f'_e(i) - f'_e(i-1)), i = 16, \dots, 19, \quad (79)$$

donde c_{scale} aplica un escalaje a $f'_{e\Delta}(i)$ de manera que $f_e(19)$ sea igual a f_{e19} . Para asegurar la estabilidad, $f'_{e\Delta}(i)$ se limita como sigue:

$$f'_{e\Delta}(i) + f'_{e\Delta}(i-1) > 500, \quad i = 17, \dots, 19 \quad (80)$$

Por último, el vector ISF extrapolado f_e se obtiene de la siguiente manera:

$$f_e(i) = \begin{cases} f(i), & i = 1, \dots, 15 \\ f'_{e\Delta}(i) + f_e(i-1), & i = 16, \dots, 19 \\ f(16), & i = 20 \end{cases} \quad (81)$$

f_e se convierte al dominio del coseno para obtener q_e con una velocidad de muestreo de 16 000 Hz. El filtro de síntesis LP de banda alta $A_{HB}(z)$ se obtiene convirtiendo q_e en el filtro LP como se describe en 5.2.4 con $m = 20$.

6.3.2.2 Modos 8,85, 12,65, 14,25, 15,85, 18,25, 19,85, 23,05 o 23,85 kbit/s

El filtro de síntesis LP de banda alta $A_{HB}(z)$ es el filtro de síntesis LP de banda baja ponderado:

$$A_{HB}(z) = \hat{A}(z/0.8) \quad (82)$$

donde $\hat{A}(z)$ es el filtro de síntesis LP interpolado. $\hat{A}(z)$ se ha calculado analizando la señal con la velocidad de muestreo de 12,8 kHz, pero ahora se utiliza para una señal de 16 kHz. En la práctica, esto significa que la respuesta en frecuencia $FR_{16}(f)$ de $A_{HB}(z)$ se obtiene como sigue:

$$FR_{16}(f) = FR_{12,8}\left(\frac{12,8}{16}f\right) \quad (83)$$

donde $FR_{12,8}(f)$ es la respuesta en frecuencia de $A(z)$, lo que quiere decir que se establecerá la correspondencia entre la banda de 5,1 a 5,6 kHz del dominio de 12,8 kHz y la banda de 6,4 a 7,0 kHz del dominio de 16 kHz.

6.3.3 Síntesis de banda alta

$u_{HB}(n)$ se filtra a través de $A_{HB}(z)$. La salida de esta síntesis de banda alta $s_{HB}(n)$ se filtra a través de un filtro FIR pasobanda $H_{HB}(z)$ que tiene una banda de paso de 6 a 7 kHz. Por último, se añade s_{HB} a la voz sintetizada $\hat{s}_{16k}(n)$ para producir la señal vocal de salida sintetizada $\hat{s}_{output}(n)$.

7 Atribución detallada de bits del códec de banda ancha multivelocidad adaptativa

La atribución detallada de los bits en el codificador de banda ancha multivelocidad adaptativa para voz se muestra en los cuadros 12a a 12i. Esos cuadros muestran el orden en que son producidos los bits por el codificador de voz. Se señala que el bit más significativo (MSB) de cada parámetro del códec se envía siempre el primero.

Cuadro 12a/G.722.2 – Parámetros de salida de codificador fuente en orden de ocurrencia y atribución de bits en la trama vocal de 477 bits/20 ms, modo 23,85 kbit/s

Bits (MSB-LSB)	Descripción
s1	bandera VAD
s2 – s9	índice del 1er subvector ISP
s10 – s17	índice del 2º subvector ISP
s18 – s23	índice del 3er subvector ISP
s24 – s30	índice del 4º subvector ISP
s31 – s37	índice del 5º subvector ISP
s38 – s42	índice del 6º subvector ISP
s43 – s47	índice del 7º subvector ISP
subtrama 1	
s48 – s56	índice de libro de códigos adaptativo
s57	bandera de filtrado LTP
s58 – s68	índice 1 de libro de códigos para pista 1
s69 – s79	índice 1 de libro de códigos para pista 2
s80 – s90	índice 1 de libro de códigos para pista 3
s91 – s101	índice 1 de libro de códigos para pista 4
s102 – s112	índice 2 de libro de códigos para pista 1
s113 – s123	índice 2 de libro de códigos para pista 2
s124 – s134	índice 2 de libro de códigos para pista 3
s135 – s145	índice 2 de libro de códigos para pista 4
s146 – s152	ganancia de libro de códigos
s153 – s156	energía de banda alta
subtrama 2	
s157 – s162	índice de libro de códigos adaptativo (relativo)
s163 – s262	misma descripción que s57-s156
subtrama 3	
s263 – s371	misma descripción que s48-s156
subtrama 4	
s372 – s477	misma descripción que s157-s262

Cuadro 12b/G.722.2 – Parámetros de salida de codificador fuente en orden de ocurrencia y atribución de bits en la trama vocal de 461 bits/20 ms, modo 23,05 kbit/s

Bits (MSB-LSB)	Descripción
s1	bandera VAD
s2 – s9	índice del 1er subvector ISP
s10 – s17	índice del 2º subvector ISP
s18 – s23	índice del 3er subvector ISP
s24 – s30	índice del 4º subvector ISP
s31 – s37	índice del 5º subvector ISP
s38 – s42	índice del 6º subvector ISP
s43 – s47	índice del 7º subvector ISP
subtrama 1	
s48 – s56	índice de libro de códigos adaptativo
s57	bandera de filtrado LTP
s58 – s68	índice 1 de libro de códigos para pista 1
s69 – s79	índice 1 de libro de códigos para pista 2
s80 – s90	índice 1 de libro de códigos para pista 3
s91 – s101	índice 1 de libro de códigos para pista 4
s102 – s112	índice 2 de libro de códigos para pista 1
s113 – s123	índice 2 de libro de códigos para pista 2
s124 – s134	índice 2 de libro de códigos para pista 3
s135 – s145	índice 2 de libro de códigos para pista 4
s146 – s152	ganancia de libro de códigos
subtrama 2	
s153 – s158	índice de libro de códigos adaptativo (relativo)
s159 – s254	misma descripción que s57-s152
subtrama 3	
s255 – s359	misma descripción que s48-s152
subtrama 4	
s360 – s461	misma descripción que s153-s254

Cuadro 12c/G.722.2 – Parámetros de salida de codificador fuente en orden de ocurrencia y atribución de bits en la trama vocal de 397 bits/20 ms, modo 19,85 kbit/s

Bits (MSB-LSB)	Descripción
s1	bandera VAD
s2 – s9	índice del 1er subvector ISP
s10 – s17	índice del 2º subvector ISP
s18 – s23	índice del 3er subvector ISP
s24 – s30	índice del 4º subvector ISP
s31 – s37	índice del 5º subvector ISP
s38 – s42	índice del 6º subvector ISP
s43 – s47	índice del 7º subvector ISP
subtrama 1	
s48 – s56	índice de libro de códigos adaptativo
s57	bandera de filtrado LTP
s58 – s67	índice 1 de libro de códigos para pista 1
s68 – s77	índice 1 de libro de códigos para pista 2
s78 – s79	índice 1 de libro de códigos para pista 3
s80–s81	índice 1 de libro de códigos para pista 4
s82–s91	índice 2 de libro de códigos para pista 1
s92–s101	índice 2 de libro de códigos para pista 2
s102 – s115	índice 2 de libro de códigos para pista 3
s116 – s129	índice 2 de libro de códigos para pista 4
s130 – s136	ganancia de cuantificación de vector
subtrama 2	
s137 – s142	índice de libro de códigos adaptativo (relativo)
s143 – s222	misma descripción que s57-s136
subtrama 3	
s223 – s311	misma descripción que s48-s136
subtrama 4	
s312 – s397	misma descripción que s137-s222

Cuadro 12d/G.722.2 – Parámetros de salida de codificador fuente en orden de ocurrencia y atribución de bits en la trama vocal de 365 bits/20 ms, modo 18,25 kbit/s

Bits (MSB-LSB)	Descripción
s1	bandera VAD
s2 – s9	índice del 1er subvector ISP
s10 – s17	índice del 2º subvector ISP
s18 – s23	índice del 3er subvector ISP
s24 – s30	índice del 4º subvector ISP
s31 – s37	índice del 5º subvector ISP
s38 – s42	índice del 6º subvector ISP
s43 – s47	índice del 7º subvector ISP
subtrama 1	
s48 – s56	índice de libro de códigos adaptativo
s57	bandera de filtrado LTP
s58 – s59	índice de libro de códigos para pista 1
s60 – s61	índice de libro de códigos para pista 2
s62 – s63	índice de libro de códigos para pista 3
s64 – s65	índice de libro de códigos para pista 4
s66 – s79	índice de libro de códigos para pista 1
s80 – s93	índice de libro de códigos para pista 2
s94 – s107	índice de libro de códigos para pista 3
s108 – s121	índice de libro de códigos para pista 4
s122 – s128	ganancia de cuantificación de vector
subtrama 2	
s129 – s134	índice de libro de códigos adaptativo (relativo)
s135 – s206	misma descripción que s57-s128
subtrama 3	
s207 – s287	misma descripción que s48-s128
subtrama 4	
s288 – s365	misma descripción que s129-s206

Cuadro 12e/G.722.2 – Parámetros de salida de codificador fuente en orden de ocurrencia y atribución de bits en la trama vocal de 317 bits/20 ms, modo 15,85 kbit/s

Bits (MSB-LSB)	Descripción
s1	bandera VAD
s2 – s9	índice del 1 ^{er} subvector ISP
s10 – s17	índice del 2 ^o subvector ISP
s18 – s23	índice del 3 ^{er} subvector ISP
s24 – s30	índice del 4 ^o subvector ISP
s31 – s37	índice del 5 ^o subvector ISP
s38 – s42	índice del 6 ^o subvector ISP
s43 – s47	índice del 7 ^o subvector ISP
subtrama 1	
s48 – s56	índice de libro de códigos adaptativo
s57	bandera de filtrado LTP
s58 – s70	índice de libro de códigos para pista 1
s71 – s83	índice de libro de códigos para pista 2
s84 – s96	índice de libro de códigos para pista 3
s97 – s109	índice de libro de códigos para pista 4
s110 – s116	ganancia de cuantificación de vector
subtrama 2	
s117 – s122	índice de libro de códigos adaptativo (relativo)
s123 – s182	misma descripción que s57-s116
subtrama 3	
s183 – s251	misma descripción que s48-s116
subtrama 4	
s252 – s317	misma descripción que s117-s182

Cuadro 12f/G.722.2 – Parámetros de salida de codificador fuente en orden de ocurrencia y atribución de bits en la trama vocal de 285 bits/20 ms, modo 14,25 kbit/s

Bits (MSB-LSB)	Descripción
s1	bandera VAD
s2 – s9	índice del 1er subvector ISP
s10 – s17	índice del 2º subvector ISP
s18 – s23	índice del 3er subvector ISP
s24 – s30	índice del 4º subvector ISP
s31 – s37	índice del 5º subvector ISP
s38 – s42	índice del 6º subvector ISP
s43 – s47	índice del 7º subvector ISP
subtrama 1	
s48 – s56	índice de libro de códigos adaptativo
s57	bandera de filtrado LTP
s58 – s70	Índice de libro de códigos para pista 1
s71 – s83	Índice de libro de códigos para pista 2
s84 – s92	índice de libro de códigos para pista 3
s93 – s101	Índice de libro de códigos para pista 4
s102 – s108	ganancia de cuantificación de vector
subtrama 2	
s109 – s114	índice de libro de códigos adaptativo (relativo)
s115 – s166	misma descripción que s57-s108
subtrama 3	
s167 – s227	misma descripción que s48-s108
subtrama 4	
s228 – s285	misma descripción que s109-s166

Cuadro 12g/G.722.2 – Parámetros de salida de codificador fuente en orden de ocurrencia y atribución de bits en la trama vocal de 253 bits/20 ms, modo 12,65 kbit/s

Bits (MSB-LSB)	Descripción
s1	bandera VAD
s2 – s9	índice del 1er subvector ISP
s10 – s17	índice del 2º subvector ISP
s18 – s23	índice del 3er subvector ISP
s24 – s30	índice del 4º subvector ISP
s31 – s37	índice del 5º subvector ISP
s38 – s42	índice del 6º subvector ISP
s43 – s47	índice del 7º subvector ISP
subtrama 1	
s48 – s56	índice de libro de códigos adaptativo
s57	bandera de filtrado LTP
s58 – s66	índice de libro de códigos para pista 1
s67 – s75	índice de libro de códigos para pista 2
s76 – s84	índice de libro de códigos para pista 3
s85 – s93	Índice de libro de códigos para pista 4
s94 – s100	ganancia de cuantificación de vector
subtrama 2	
s101 – s106	índice de libro de códigos adaptativo (relativo)
s107 – s150	misma descripción que s57-s100
subtrama 3	
s151 – s203	misma descripción que s48-s100
subtrama 4	
s204 – s253	misma descripción que s101-s150

Cuadro 12h/G.722.2 – Parámetros de salida de codificador fuente en orden de ocurrencia y atribución de bits en la trama vocal de 177 bits/20 ms, modo 8,85 kbit/s

Bits (MSB-LSB)	Descripción
s1	bandera VAD
s2 – s9	índice del 1er subvector ISP
s10 – s17	índice del 2º subvector ISP
s18 – s23	índice del 3er subvector ISP
s24 – s30	índice del 4º subvector ISP
s31 – s37	índice del 5º subvector ISP
s38 – s42	índice del 6º subvector ISP
s43 – s47	índice del 7º subvector ISP
subtrama 1	
s48 – s55	índice de libro de códigos adaptativo
s56 – s60	índice de libro de códigos para pista 1
s61 – s65	índice de libro de códigos para pista 2
s66 – s70	índice de libro de códigos para pista 3
s71 – s75	índice de libro de códigos para pista 4
s76 – s81	ganancia de cuantificación de vector
subtrama 2	
s82 – s86	índice de libro de códigos adaptativo (relativo)
s87 – s112	misma descripción que s56-s81
subtrama 3	
s113 – s146	misma descripción que s48-s81
subtrama 4	
s147 – s177	misma descripción que s82-s112

Cuadro 12i/G.722.2 – Parámetros de salida de codificador fuente en orden de ocurrencia y atribución de bits en la trama vocal de 132 bits/20 ms, modo 6,60 kbit/s

Bits (MSB-LSB)	Descripción
s1	bandera VAD
s2 – s9	índice del 1er subvector ISP
s10 – s17	índice del 2º subvector ISP
s18 – s24	índice del 3er subvector ISP
s25 – s31	índice del 4º subvector ISP
s32 – s37	índice del 5º subvector ISP
subtrama 1	
s38 – s45	índice de libro de códigos adaptativo
s46 – s57	índice de libro de códigos
s58 – s63	ganancia de cuantificación de vector
subtrama 2	
s64 – s68	índice de libro de códigos adaptativo (relativo)
s69 – s86	misma descripción que s46-s63
subtrama 3	
s87 – s109	misma descripción que s64-s86
subtrama 4	
s110 – s132	misma descripción que s64-s86

8 Secuencias del retorno a la posición inicial

8.1 Descripción funcional

El códec de banda ancha multivelocidad adaptativa para voz se describe en una aritmética exacta a nivel de bit para facilitar la aprobación de tipos y también como prueba general del funcionamiento correcto del códec de banda ancha multivelocidad adaptativa para voz.

La respuesta del códec a una secuencia de entrada predefinida sólo puede preverse si las variables de estado internas del códec están en un estado predefinido al comienzo del experimento. Por consiguiente, el códec tiene que ponerse en un estado que se conoce como "estado inicial" antes de que se pueda efectuar la prueba exacta a nivel de bit. Esto se hace normalmente aplicando un reinicio (procedimiento en el que las variables de estado internas del códec se fijan a sus valores iniciales definidos). El modo de códec del codificador de voz y el decodificador de voz se fijará en el modo de códec probado por medios externos en el reinicio.

Para hacer posible un reinicio del códec en ubicaciones distantes se han definido tramas especiales de retorno a la posición inicial del codificador y el decodificador, lo que permite el retorno del códec a la posición inicial mediante señalización dentro de banda.

El procedimiento de retorno del códec a la posición inicial se define de tal manera que, cualquiera que sea el sentido (codificador o decodificador), las funciones de retorno a la posición inicial son invocadas tras el procesamiento de la trama de retorno a la posición inicial. La salida correspondiente a la primera trama de retorno a la posición inicial depende por tanto del modo de códec utilizado y del estado del códec cuando se recibe esa trama y, por ello, normalmente es desconocida. La respuesta del codificador a cualquier trama posterior de retorno a la posición inicial es, por definición, la trama correspondiente de retorno a la posición inicial del decodificador para el modo de códec utilizado. La respuesta del decodificador a cualquier trama posterior de retorno a la

posición inicial es, por definición, la trama de retorno a la posición inicial del decodificador. Este procedimiento permite el retorno a la posición inicial tanto del codificador como del decodificador desde cualquier lado, si se implementa una configuración en bucle, teniendo en cuenta la alineación de tramas adecuada.

8.2 Definiciones

Trama de retorno a la posición inicial del codificador: La trama de retorno a la posición inicial del codificador consta de 320 muestras idénticas, cada una de ellas con una longitud de 13 bits, con el bit menos significativo fijado en "uno" y todos los demás bits fijados en "cero". Cuando se escriben en palabras de 16 bits con justificación a la izquierda, las muestras tienen un valor de 0008 hexadecimal. El decodificador de voz tiene que producir esta trama como respuesta a la segunda y a cualquier trama posterior de retorno a la posición inicial del decodificador si se introducen en el decodificador al menos dos tramas de retorno a la posición inicial de manera consecutiva. La trama de retorno a la posición inicial del codificador es idéntica para todos los modos de códec.

Trama de retorno a la posición inicial del decodificador: Hay nueve tramas diferentes de retorno a la posición inicial del decodificado, que corresponden a los nueve modos del códec AMR-WB. Cuando se utiliza uno de estos modos de códec, la trama correspondiente de retorno a la posición inicial del decodificador es la respuesta natural del codificador de voz a la segunda y a cualquier trama posterior del de retorno a la posición inicial del codificador si se introducen en el codificador al menos dos tramas de retorno a la posición inicial de manera consecutiva. En el anexo C se dan los valores de los parámetros de cada trama de retorno a la posición inicial del codificador.

8.3 Retorno a la posición inicial del codificador

Cuando el codificador de banda ancha multivelocidad adaptativa para voz recibe a su entrada una trama de retorno a la posición inicial de codificador alineada exactamente con su segmentación de trama de voz interna, tienen lugar los siguientes eventos:

Paso 1

El codificador de voz realiza su operación normal incluyendo VAD y SCR y produce en su salida, de acuerdo con el modo de códec utilizado, una trama de parámetro de voz que por lo general es desconocida. Pero si el codificador de voz está en su estado inicial al comienzo de esa trama, la trama de parámetro de voz resultante es idéntica a la trama de retorno a la posición inicial del decodificador, que corresponde al modo de códec utilizado (esta es la manera según la cual se construyeron las tramas de retorno a la posición inicial del decodificador).

Paso 2

Tras la terminación satisfactoria de esa operación, el codificador de voz activa las funciones de retorno a la posición inicial para todos los submódulos que incluyen VAD y SCR y fija todas las variables de estado en su estado inicial. Cuando reciba la próxima trama de entrada, el codificador de voz empezará desde su estado inicial.

NOTA – La aplicación de una secuencia de N tramas de retorno a la posición inicial del codificador dará lugar al menos a N-1 tramas de retorno a la posición inicial del decodificador en la salida del codificador de voz.

8.4 Retorno a la posición inicial del decodificador

Cuando el decodificador de voz recibe a su entrada una trama de retorno a la posición inicial del decodificador, que corresponde al modo de códec utilizado, tienen lugar los siguientes eventos:

Paso 1

El decodificador de voz realiza su operación normal y produce en su salida una trama de voz, que por lo general es desconocida. Pero si el decodificador de voz está en su estado inicial al comienzo de esa trama, la trama de voz resultante es sustituida por la trama de retorno a la posición inicial del codificador. Este no sería el caso de forma natural, pero viene forzado por la presente definición.

Paso 2

Tras la terminación satisfactoria de esa operación, el decodificador de voz activa las funciones de retorno a la posición inicial para todos los submódulos que incluyen el generador de ruido de confort y fija todas las variables de estado en su estado inicial. Cuando reciba la próxima trama de entrada, el decodificador de voz empezará desde su estado inicial.

NOTA 1 – La aplicación de una secuencia de N tramas de retorno a la posición inicial del decodificador dará lugar al menos a N-1 tramas de retorno a la posición inicial del codificador en la salida del decodificador de voz.

NOTA 2 – Por definición, la primera trama de cada secuencia de prueba de decodificador debe diferir de la trama de retorno a la posición inicial del decodificador en al menos la posición de un bit dentro de los parámetros de LPC y la primera subtrama. Por consiguiente, si el decodificador está en su estado inicial, basta con comprobar tan sólo esos parámetros para detectar una trama de retorno a la posición inicial del decodificador subsiguiente. La definición se formula con miras al soporte de una implementación con respecto a la demora.

9 Detector de actividad vocal (VAD)

La función del algoritmo VAD es indicar si cada trama de 20 ms contiene señales que deberían ser transmitidas, por ejemplo, voz, música o tonos de información. La salida del algoritmo VAD es una bandera booleana (`VAD_flag`) que indica la presencia de esas señales. La bandera se utiliza en el codificador de voz AMR-WB y se refiere a ella el anexo B.

9.1 Símbolos VAD

Para los fines de esta cláusula, se aplican los siguientes símbolos.

9.1.1 Variables VAD

<code>bckr_est [n]</code>	estimación del ruido de fondo en la banda de frecuencias "n"
<code>burst_count</code>	longitud del cómputo de una ráfaga de conversación, utilizado por la adición del periodo de mantenimiento de VAD
<code>hang_count</code>	contador de periodos de mantenimiento, utilizado por la adición del periodo de mantenimiento de VAD
<code>level [n]</code>	nivel de la señal en la banda de frecuencias "n"
<code>new_speech</code>	puntero del codificador de voz, apunta a una memoria intermedia que contiene las últimas muestras recibidas de una trama de voz
<code>noise_level</code>	nivel de ruido estimado
<code>pow_sum</code>	potencia de entrada
<code>s(i)</code>	muestras de la trama de entrada
<code>snr_sum</code>	medida entre la trama de entrada y la estimación del ruido

speech_level	nivel de voz estimado
stat_count	contador estacionario
stat_rat	medida que indica estacionaridad de la trama de entrada
tone_flag	bandera que indica la presencia de un tono
vad_thr	umbral VAD
VAD_flag	bandera VAD booleana
vadreg	decisión VAD

9.1.2 Constantes VAD

ALPHA_UP1	constante para actualizar la estimación del ruido
ALPHA_DOWN1	constante para actualizar la estimación del ruido
ALPHA_UP2	constante para actualizar la estimación del ruido
ALPHA_DOWN2	constante para actualizar la estimación del ruido
ALPHA3	constante para actualizar la estimación del ruido
ALPHA4	constante para actualizar el nivel de señal medio
ALPHA5	constante para actualizar el nivel de señal medio
BURST_HIGH	constante para controlar la adición del periodo de mantenimiento de VAD
BURST_P1	constante para controlar la adición del periodo de mantenimiento de VAD
BURST_SLOPE	constante para controlar la adición del periodo de mantenimiento de VAD
COEFF3	coeficiente del banco de filtros
COEFF5_1	coeficiente del banco de filtros
COEFF5_2	coeficiente del banco de filtros
HANG_HIGH	constante para controlar la adición del periodo de mantenimiento de VAD
HANG_LOW	constante para controlar la adición del periodo de mantenimiento de VAD
HANG_P1	constante para controlar la adición del periodo de mantenimiento de VAD
HANG_SLOPE	constante para controlar la adición del periodo de mantenimiento de VAD
FRAME_LEN	tamaño de una trama de voz, 256 muestras (20 ms)
MIN_SPEECH_LEVEL1	constante para la estimación de la voz
MIN_SPEECH_LEVEL2	constante para la estimación de la voz
MIN_SPEECH_SNR	constante para la adaptación del umbral VAD
NO_P1	constante para la adaptación del umbral VAD
NO_SLOPE	constante para la adaptación del umbral VAD
NOISE_MAX	valor máximo para la estimación del ruido

NOISE_MIN	valor mínimo para la estimación del ruido
POW_TONE_THR	umbral de detección del tono
SP_ACTIVITY_COUNT	constante para la estimación de la voz
SP_ALPHA_DOWN	constante para la estimación de la voz
SP_ALPHA_UP	constante para la estimación de la voz
SP_CH_MAX	constante para la adaptación del umbral de VAD
SP_CH_MIN	constante para la adaptación del umbral de VAD
SP_EST_COUNT	constante para la estimación de la voz
SP_P1	constante para la adaptación del umbral de VAD
SP_SLOPE	constante para la adaptación del umbral de VAD
STAT_COUNT	umbral de detección estacionaria
STAT_THR	umbral de detección estacionaria
STAT_THR_LEVEL	umbral de detección estacionaria
THR_HIGH	constante para la adaptación del umbral de VAD
TONE_THR	umbral de detección de tono
VAD_POW_LOW	constante para controlar la adición del periodo de mantenimiento de VAD

9.1.3 Funciones

+	Adición
-	Sustracción
*	Multiplicación
/	División
x	valor absoluto de x
AND	AND booleano
OR	OR booleano

$$\sum_{n=a}^b x(n) = x(a) + x(a+1) + \dots + x(b-1) + x(b)$$

$$\text{MIN}(x, y) = \begin{cases} x, & x \leq y \\ y, & y < x \end{cases}$$

$$\text{MAX}(x, y) = \begin{cases} x, & x \geq y \\ y, & y > x \end{cases}$$

9.2 Descripción funcional

En la figura 4 se muestra el diagrama de bloques del algoritmo VAD. El algoritmo VAD utiliza parámetros del codificador de voz para calcular la bandera VAD (`VAD_flag`) booleana. Esta trama de entrada al VAD se muestrea a la frecuencia de 12,8 kHz y contiene por tanto 256 muestras. Las muestras de la trama de entrada (`s(i)`) se dividen en subbandas y el nivel de la señal (`level[n]`) se calcula en cada banda. Las ganancias de altura de tono en bucle abierto normalizadas son la entrada a la función de detección de tono, ganancias que se calculan mediante el análisis de la altura

de tono en bucle abierto del codificador de voz. La función de detección de tono calcula una bandera (*tone_flag*) que indica la presencia de un tono de señalización, señal vocal u otra señal fuertemente periódica. El nivel de ruido de fondo (*bckr_est[n]*) se estima en cada banda de acuerdo con la decisión VAD, la estacionaridad de la señal y la bandera de tono. La decisión VAD intermedia se calcula comparando la relación señal/ruido (SNR, *signal to noise ratio*) a la entrada. ($level[n] / bckr_est[n]$) con un umbral adaptativo. El umbral se adapta de acuerdo con las estimaciones de ruido y de voz a largo plazo. Por último, se calcula la bandera VAD añadiendo un periodo de mantenimiento a la decisión VAD intermedia.

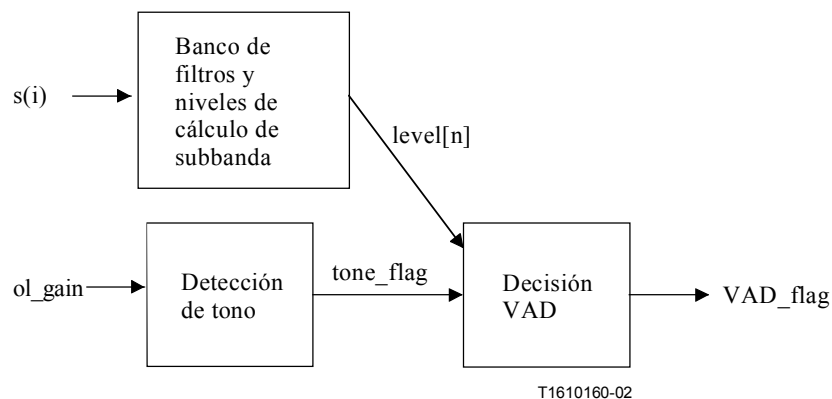


Figura 4/G.722.2 – Diagrama de bloques simplificado del algoritmo VAD

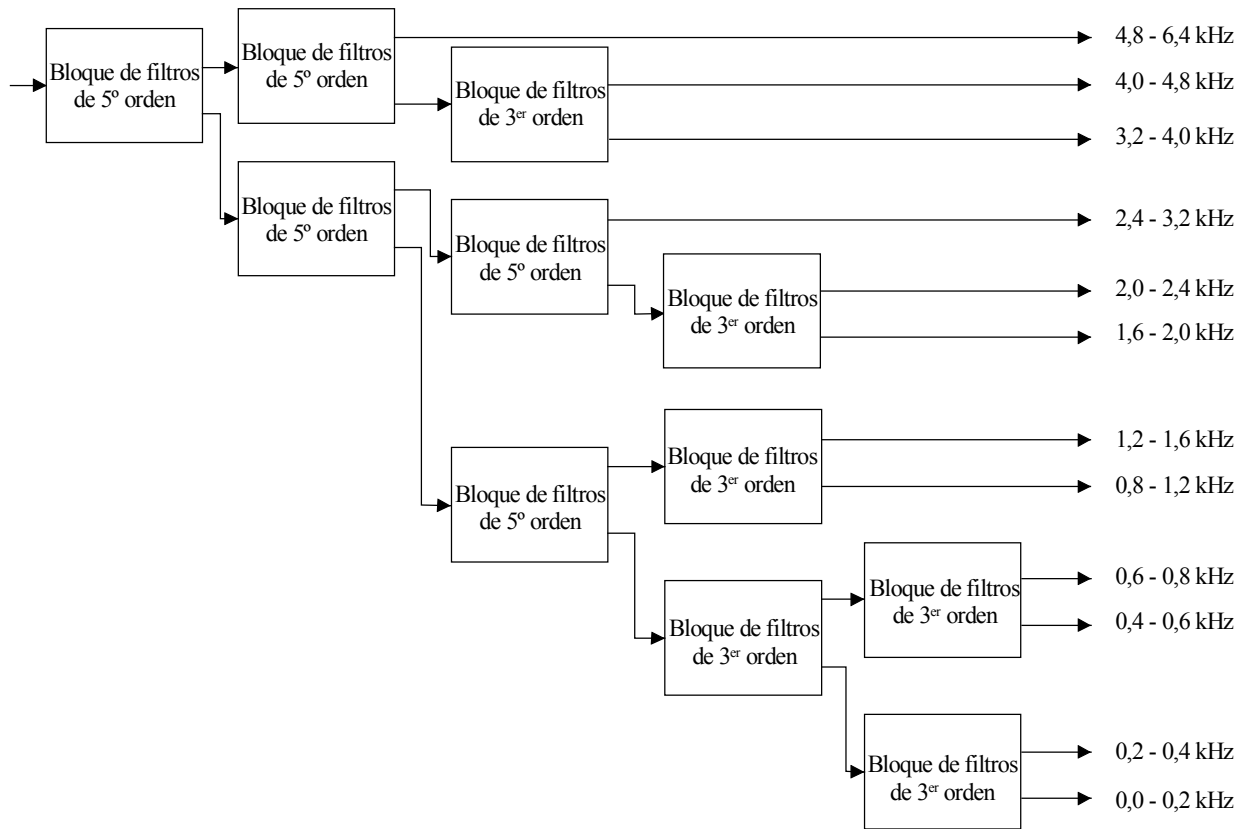
9.2.1 Banco de filtros y cálculo de niveles de subbanda

La señal de entrada se divide en bandas de frecuencias utilizando un banco de filtros de 12 bandas (figura 5). En el cuadro 13 se muestran las frecuencias de corte del banco de filtros.

Cuadro 13/G.722.2 – Frecuencias de corte del banco de filtros

Número de banda	Frecuencias
1	0-200 Hz
2	200-400 Hz
3	400-600 Hz
4	600-800 Hz
5	800-1200 Hz
6	1200-1600 Hz
7	1600-2000 Hz
8	2000-2400 Hz
9	2400-3200 Hz
10	3200-4000 Hz
11	4000-4800 Hz
12	4800-6400 Hz

La entrada al banco de filtros es una trama vocal a la que apunta el puntero *new_speech* del codificador de voz G.722.2. A los valores de entrada al banco de filtros se les aplica un escalaje en sentido descendente de un bit. De esta manera se garantiza un escalaje seguro, es decir, que no se puede producir saturación durante el cálculo del banco de filtros.



T1610170-02

Figura 5/G.722.2 – Banco de filtros

El banco de filtros consta de bloques de filtros de 5° y 3^{er} orden. Cada bloque de filtros divide la entrada en partes de paso alto y paso bajo y decima la frecuencia de muestreo por 2. El bloque de filtros de 5° orden se calcula como sigue:

$$x_{lp}(i) = 0.5 * (A_1(x(2*i)) + A_2(x(2*i+1))) \quad (84a)$$

$$x_{hp}(i) = 0.5 * (A_1(x(2*i)) - A_2(x(2*i+1))) \quad (84b)$$

donde:

$x(i)$ señal de entrada a un bloque de filtros

$x_{lp}(i)$ componente de paso bajo

$x_{hp}(i)$ componente de paso alto

El bloque de filtros de 3^{er} orden se calcula como sigue:

$$x_{lp}(i) = 0.5 * (x(2*i+1) + A_3(x(2*i))) \quad (85a)$$

$$x_{hp}(i) = 0.5 * (x(2*i+1) - A_3(x(2*i))) \quad (85b)$$

Los filtros $A_1()$, $A_2()$, y $A_3()$ son filtros todos los pasos de forma directa de primer orden, cuya función de transferencia viene dada por:

$$A(z) = \frac{C + z^{-1}}{1 + C * z^{-1}} \quad (86)$$

donde C es el coeficiente del filtro.

Los coeficientes de los filtros todos los pasos $A_1()$, $A_2()$, y $A_3()$ son $COEFF5_1$, $COEFF5_2$, y $COEFF5_3$, respectivamente.

El nivel de señal se calcula a la salida del banco de filtros en cada banda de frecuencias como sigue:

$$level(n) = \sum_{i=START_n}^{END_n} |x_n(i)| \quad (87)$$

donde:

n : índice de la banda de frecuencias

$x_n(i)$: muestra i a la salida del banco de filtros en la banda de frecuencias n

$$START_n = \begin{cases} -6, & 1 \leq n \leq 4 \\ -12, & 5 \leq n \leq 8 \\ -24, & 9 \leq n \leq 11 \\ -48, & n = 12 \end{cases}$$

$$END_n = \begin{cases} 7, & 1 \leq n \leq 4 \\ 15, & 5 \leq n \leq 8 \\ 31, & 9 \leq n \leq 11 \\ 63, & n = 12 \end{cases}$$

Los índices negativos de $x_n(i)$ hacen referencia a la trama previa.

9.2.2 Detección de tono

El objetivo de la función de detección de tono es detectar tonos de información, sonidos vocales y otras señales periódicas. La detección de tono utiliza ganancias de altura de tono en bucle abierto (ol_gain), normalizadas, que se reciben del codificador de voz. Si la ganancia de altura de tono es superior a la constante $TONE_THR$, el tono es detectado y se fija la bandera de tono como se indica a continuación:

```
si ( $ol\_gain > TONE\_THR$ )
     $tone\_flag = 1$ 
```

Se busca la altura de tono en bucle abierto y se calcula en consecuencia la bandera de tono dos veces en cada trama, salvo en el modo 6,60 kbit/s en el que se calcula una sola vez.

9.2.3 Decisión VAD

En la figura 6 se muestra el diagrama de bloques del algoritmo de decisión VAD.

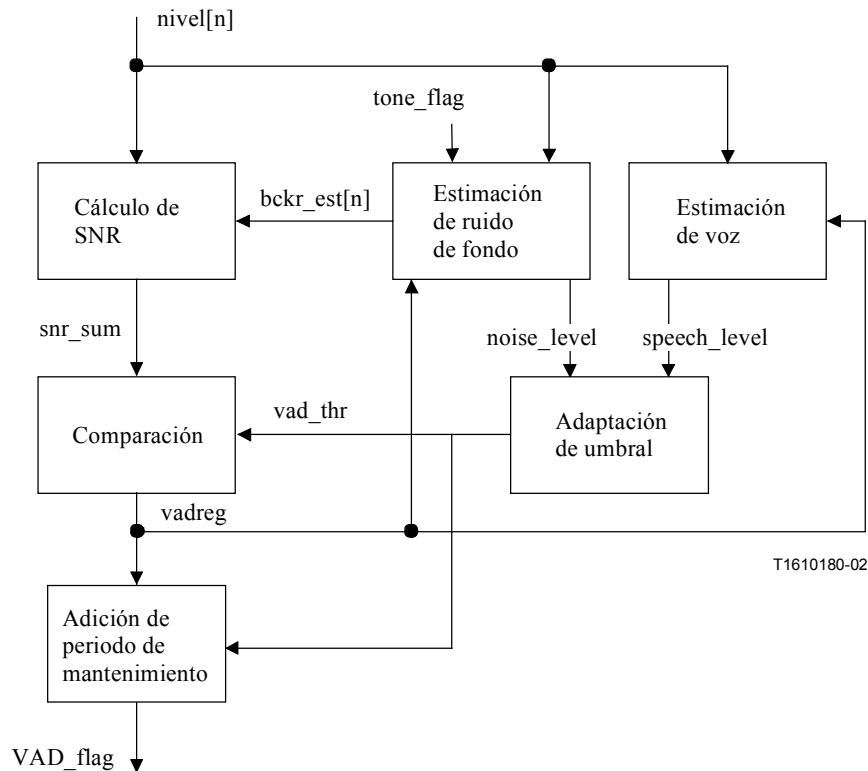


Figura 6/G.722.2 – Diagrama de bloques simplificado del algoritmo de decisión VAD

La potencia de la trama de entrada se calcula como sigue:

$$frame_pow = \sum_{i=0}^{FRAME_LEN} s(i) * s(i) \quad (88)$$

donde el puntero `new_speech` del codificador de voz apunta a las muestras $s(i)$ de la trama de entrada. La variable `pow_sum` es la suma de las potencias de las tramas en curso y previa. Si `pow_sum` es inferior a la constante `POW_TONE_THR`, la bandera de tono se fija en cero.

La diferencia entre los niveles de señal de la trama de entrada y la estimación de ruido de fondo se calcula como sigue:

$$snr_sum = \sum_{n=1}^{12} MAX(1.0, \frac{level[n]}{bckr_est[n]})^2 \quad (89)$$

donde:

- `level [n]` nivel de la señal en la banda n
- `bckr_est [n]` nivel de la estimación del ruido de fondo en la banda n

La decisión VAD se toma comparando la variable `snr_sum` con un umbral. El umbral (`vad_thr`) se adapta para obtener la sensibilidad deseada, dependiendo de los niveles de voz y ruido de fondo estimados.

El ruido de fondo medio se calcula añadiendo estimaciones del ruido en cada banda excepto en la más baja:

$$noise_level = \sum_{n=2}^{12} bckr_est[n] \quad (90)$$

Si SNR es inferior al umbral (MIN_SPEECH_SNR), el nivel de voz se incrementa como sigue:

```
Si (speech_level/noise_level < MIN_SPEECH_SNR)
    Speech_level = MIN_SPEECH_SNR * noise_level
```

El valor logarítmico de la estimación del ruido se calcula de la siguiente manera:

$$\text{ilog2_noise_level} = \log_2(\text{noise_level}) \quad (91)$$

Antes de calcular el valor logarítmico a partir de la estimación de la voz, se sustrae MIN_SPEECH_SNR*noise_level del nivel de voz para corregir su valor en situaciones de SNR baja.

$$\text{ilog2_speech_level} = \log_2(\text{speech_level} - \text{MIN_SPEECH_SNR} * \text{noise_level}) \quad (92)$$

El umbral de la decisión VAD se calcula como sigue:

$$\begin{aligned} \text{Vad_thr} = & \text{NO_SLOPE} * (\text{ilog2_noise_level} - \text{NO_P1}) + \text{THR_HIGH} + \\ & \text{MIN}(\text{SP_CH_MAX}, \text{MAX}(\text{SP_CH_MIN}, \text{SP_CH_MIN} \\ & + \text{SP_SLOPE} * (\text{ilog2_speech_level} - \text{SP_P1}))) \end{aligned} \quad (93)$$

donde NO_SLOPE, SP_SLOPE, NO_P1, SP_P1, THR_HIGH, SP_CH_MAX y SP_CH_MIN son constantes.

La variable vadreg indica decisión VAD intermedia y se calcula de la siguiente manera:

```
if (snr_sum > vad_thr)
    vadreg = 1
else
    vadreg = 0
```

9.2.3.1 Adición de periodo de mantenimiento

Antes de que se dé la bandera VAD final se añade un periodo de mantenimiento. La adición de ese periodo ayuda a detectar las terminaciones de baja potencia de ráfagas de conversación, que son subjetivamente importantes pero difíciles de detectar.

La bandera VAD se fija en "1" si han transcurrido menos de hang_len tramas con decisión "0" desde que se han detectado burst_len decisiones "1" consecutivas. Las variables hang_len y burst_len se calculan utilizando vad_thr como sigue:

$$\text{hang_len} = \text{MAX}(\text{HANG_LOW}, (\text{HANG_SLOPE} * (\text{vad_thr} - \text{HANG_P1}) + \text{HANG_HIGH})) \quad (94)$$

$$\text{burst_len} = \text{BURST_SLOPE} * (\text{vad_thr} - \text{BURST_P1}) + \text{BURST_HIGH} \quad (95)$$

La potencia de la trama de entrada se compara con un umbral (VAD_POW_LOW). Si la potencia es menor, la bandera VAD se fija en "0" y no se añade periodo de mantenimiento. La vad_flag se calcula como se indica a continuación:

```
Vad_flag = 0;
if (pow_sum < VAD_POW_LOW)
    burst_count = 0
    hang_count = 0
else
    if (vadreg = 1)
        burst_count = burst_count + 1
        if (burst_count >= burst_len)
            hang_count = hang_len
            VAD_flag = 1
```

```

else
    burst_count = 0
    if (hang_count > 0)
        hang_count = hang_count - 1
        VAD_flag = 1

```

9.2.3.2 Estimación de ruido de fondo

La estimación del ruido de fondo ($bckr_est[n]$) se actualiza utilizando niveles de amplitud de la trama previa. Así pues, la actualización se demora en una trama para evitar que el comienzo no detectado de ráfagas de conversación corrompa la estimación del ruido. La velocidad de actualización de la trama en curso se selecciona utilizando decisiones VAD intermedias ($vadreg$) y un contador de estacionaridad ($stat_count$) como sigue:

```

if (vadreg for the last 4 frames has been zero)
    alpha_up = ALPHA_UP1
    alpha_down = ALPHA_DOWN1
else if (stat_count = 0)
    alpha_up = ALPHA_UP2
    alpha_down = ALPHA_DOWN2
else
    alpha_up = 0
    alpha_down = ALPHA3

```

La variable $stat_count$ indica estacionaridad y su finalidad se explica más adelante en la presente cláusula. Las variables $alpha_up$ y $alpha_down$ definen la velocidad de actualización en sentido ascendente y en sentido descendente, respectivamente. La velocidad de actualización de cada banda "n" se selecciona como sigue:

```

if ( $bckr\_est_m[n] < level_{m-1}[n]$ )
    alpha[n] = alpha_up
else
    alpha[n] = alpha_down

```

Por último, la estimación del ruido se actualiza como se indica a continuación:

$$bckr_est_{m+1}[n] = (1.0 - alpha[n]) * bckr_est_m[n] + alpha[n] * level_{m-1}[n] , \quad (96)$$

donde:

- n: índice de la banda de frecuencias
- m: índice de la trama

El nivel de la estimación del ruido de fondo ($bckr_est[n]$) está limitado entre las constantes $NOISE_MIN$ y $NOISE_MAX$.

Si el nivel del ruido de fondo aumenta súbitamente, $vadreg$ se fija en "1" mientras que, normalmente, el ruido de fondo no se actualiza en sentido ascendente. Para recuperarse de esta situación se habilita la actualización de la estimación del ruido de fondo siempre que la decisión VAD intermedia ($vadreg$) sea "1" durante un periodo de tiempo suficiente y el espectro permanezca estacionario. La estacionaridad ($stat_rat$) se estima utilizando la siguiente ecuación:

$$stat_rat = \sum_{n=1}^{12} \frac{MAX(STAT_THR_LEVEL, MAX(ave_level_m[n], level_m[n]))}{MAX(STAT_THR_LEVEL, MIN(ave_level_m[n], level_m[n]))} , \quad (97)$$

donde:

STAT_THR_LEVEL: una constante

n: índice de la banda de frecuencias

m: índice de la trama

ave_level: nivel medio de la señal de entrada

Si la estimación estacionaria (*stat_rat*) es superior a un umbral, el contador estacionario (*stat_count*) se fija en el valor inicial definido por la constante *STAT_COUNT*. Si la señal no es estacionaria pero se ha detectado voz (la decisión VAD es "1"), se decrementa *stat_count* en una unidad en cada trama hasta que sea cero.

```
if (5 last tone flags have been one)
    stat_count = STAT_COUNT
else
    if (8 last internal VAD decisions have been zero) OR (stat_rat >
STAT_THR)
        stat_count = STAT_COUNT
    else
        if (vadreg) AND (stat_count ≠ 0)
            stat_count = stat_count - 1
```

Los niveles de señal medios (*ave_level* [*n*]) se calculan como sigue:

$$ave_level_{m+1}[n] = (1.0 - \alpha) * ave_level_m[n] + \alpha * level_m[n] \quad (98)$$

La velocidad de actualización (*alpha*) de la ecuación anterior se selecciona como se indica a continuación:

```
if (stat_count = STAT_COUNT)
    alpha = 1.0
else if (vadreg = 1)
    alpha=ALPHA5
else
    alpha = ALPHA4
```

9.2.3.3 Estimación del nivel de voz

Primero se calcula el nivel de entrada de la banda completa sumando los niveles de entrada en cada banda excepto en la banda más baja como sigue:

$$in_level = \sum_{n=2}^{12} level[n] \quad (99)$$

Se supone que una trama contiene voz si su nivel es suficientemente alto (*MIN_SPEECH_LEVEL1*) y la bandera VAD (*vadreg*) intermedia está fijada o el nivel de entrada es superior a la estimación del nivel de voz actual. Se busca el nivel máximo (*sp_max*) de las tramas *SP_EST_COUNT*. Si el número *SP_ACTIVITY_COUNT* de tramas de voz se encuentra dentro del número *SP_EST_COUNT* de tramas, la estimación del nivel de voz es actualizada por el nivel de señal máximo (*sp_max*). El pseudocódigo de estimación del nivel de voz es como sigue:

```
If (SP_ACTIVITY_COUNT > SP_EST_COUNT - sp_est_cnt + sp_max_cnt)
    sp_est_cnt = 0
    sp_max_cnt = 0
    sp_max = 0
sp_est_cnt = sp_est_cnt + 1
```

```

if (in_level > MIN_SPEECH_LEVEL1) AND ((vadreg = 1) OR (in_level >
speech_level))
    sp_max_cnt = sp_max_cnt + 1
    sp_max = MAX(sp_max, in_level)
    if (sp_max_cnt > SP_ACTIVITY_COUNT)
        if (sp_max > MIN_SPEECH_LEVEL2)
            if (sp_max > speech_level)
                speech_level = speech_level + SP_ALPHA_UP * (sp_max -
speech_level)
            else
                speech_level = speech_level + SP_ALPHA_DOWN * (sp_max -
speech_level)
        sp_max_cnt = 0
        sp_max = 0
        sp_est_cnt = 0

```

10 Bibliografía

- [2] SCHROEDER M.R. and ATAL B.S.: Code-Excited Linear Prediction (CELP): High quality speech at very low bit rates, in *Proc. ICASSP'85*, pp. 937-940, 1985.
- [3] RABINER L.R. and SCHAEFER R.W.: *Digital processing of speech signals*, Prentice-Hall Int., 1978.
- [4] ITAKURA F.: Line spectral representation of linear predictive coefficients of speech signals, *J. Acoust. Soc. Amer.*, Vol. 57, Supplement No. 1, S35, 1975.
- [5] BISTRITZ Y. and PELLERM S.: Immittance Spectral Pairs (ISP) for speech encoding, in *Proc. ICASSP'93*, pp. II-9 – II-12.
- [6] PALIWAL K.K and ATAL B.S.: Efficient vector quantization of LPC parameters at 24 bits/frame, *IEEE Trans. Speech and Audio Processing*, Vol. 1, no 1, pp. 3-14, 1993.
- [7] KABAL P. and RAMACHANDRAN R.P.: The computation of line spectral frequencies using Chebyshev polynomials, *IEEE Trans. on ASSP*, Vol. 34, no. 6, pp. 1419-1426, 1986.
- [8] JÄRVINEN K., VAINIO J., KAPANEN P., HONKANEN T., HAAVISTO P., SALAMI R., LAFLAMME C., and ADOUL J.-P.: GSM enhanced full rate speech codec, in *Proc. ICASSP'97*, pp. 771-774.
- [9] HONKANEN T., VAINIO J., JÄRVINEN K., HAAVISTO P., SALAMI R., LAFLAMME C., and ADOUL J.-P.: Enhanced full rate speech codec for IS-136 digital cellular system, in *Proc. ICASSP'97*, pp. 731-734.
- [10] HAGEN R., EKUDDEN E., JOHANSSON B., and KLEIJN W.B.: Removal of sparse-excitation artifacts in CELP, in *Proc. ICASSP'98*, pp. I-145-I-148.
- [11] 3GPP TS 26.171: AMR Wideband Speech Codec; General description.
- [12] 3GPP TS 26.190: AMR Wideband Speech Codec; Transcoding functions.
- [13] 3GPP TS 26.173: AMR Wideband Speech Codec; ANSI-C code.
- [14] 3GPP TS 26.174: AMR Wideband Speech Codec; Test sequences.
- [15] 3GPP TS 26.193: AMR Wideband Speech Codec; Source Controlled Rate operation.
- [16] 3GPP TS 26.194: AMR Wideband Speech Codec; Voice Activity Detector (VAD).

- [17] 3GPP TS 26.192: AMR Wideband Speech Codec; Comfort Noise Aspects.
- [18] 3GPP TS 26.191: AMR Wideband Speech Codec; Error Concealment of Lost Frames.
- [19] 3GPP TS 26.201: AMR Wideband Speech Codec; Frame Structure.
- [20] 3GPP TR 26.901: AMR Wideband Speech Codec; Feasibility study report.

SERIES DE RECOMENDACIONES DEL UIT-T

Serie A	Organización del trabajo del UIT-T
Serie B	Medios de expresión: definiciones, símbolos, clasificación
Serie C	Estadísticas generales de telecomunicaciones
Serie D	Principios generales de tarificación
Serie E	Explotación general de la red, servicio telefónico, explotación del servicio y factores humanos
Serie F	Servicios de telecomunicación no telefónicos
Serie G	Sistemas y medios de transmisión, sistemas y redes digitales
Serie H	Sistemas audiovisuales y multimedios
Serie I	Red digital de servicios integrados
Serie J	Redes de cable y transmisión de programas radiofónicos y televisivos, y de otras señales multimedios
Serie K	Protección contra las interferencias
Serie L	Construcción, instalación y protección de los cables y otros elementos de planta exterior
Serie M	RGT y mantenimiento de redes: sistemas de transmisión, circuitos telefónicos, telegrafía, facsímil y circuitos arrendados internacionales
Serie N	Mantenimiento: circuitos internacionales para transmisiones radiofónicas y de televisión
Serie O	Especificaciones de los aparatos de medida
Serie P	Calidad de transmisión telefónica, instalaciones telefónicas y redes locales
Serie Q	Conmutación y señalización
Serie R	Transmisión telegráfica
Serie S	Equipos terminales para servicios de telegrafía
Serie T	Terminales para servicios de telemática
Serie U	Conmutación telegráfica
Serie V	Comunicación de datos por la red telefónica
Serie X	Redes de datos y comunicación entre sistemas abiertos
Serie Y	Infraestructura mundial de la información y aspectos del protocolo Internet
Serie Z	Lenguajes y aspectos generales de soporte lógico para sistemas de telecomunicación