

UNIVERSIDAD CARLOS III DE MADRID

RESUMEN DE PROYECTO FIN DE CARRERA

ESTUDIO DE TÉCNICAS DE OCULTACIÓN DE ERRORES
PARA SISTEMAS DE TRANSMISIÓN DE AUDIO EN TIEMPO
REAL

ARÁNZAZU ROBLES MOYA

JUNIO 2012

Contents

- 1 Introducción 3**
- 2 Métodos de Codificación de Audio 4**
 - 2.1 ADPCM 4
 - 2.1.1 Técnicas de Gestión de Errores 4
 - 2.2 Apt-x100 7
 - 2.2.1 Técnicas de Gestión de Errores 7
 - 2.3 AMR-WB+ 8
 - 2.3.1 Técnicas de Gestión de Errores 9
 - 2.4 AAC 10
 - 2.4.1 Técnicas de Gestión de Errores 10
 - 2.5 MPEG-1 Capa II y Capa III 13
 - 2.5.1 Técnicas de Gestión de Errores 13
 - 2.6 AC3 15
 - 2.6.1 Técnicas de Gestión de Errores 15
- 3 Simulación 17**

1 Introducción

En sistemas de transmisión de audio digital en tiempo real los efectos producidos por canales ruidosos deben ser evitados o al menos minimizados. Para poder mejorar la calidad de las comunicaciones de voz y los mecanismos de compresión, se pueden aplicar técnicas como interpolación, reconstrucción, repetición de forma de onda o silenciado. Además, teniendo en cuenta las características específicas de cada mecanismo se pueden definir nuevas técnicas que permitan una mejor gestión de los errores.

En este proyecto se presenta una visión general de las principales técnicas empleadas para la detección y ocultación de errores en varios sistemas de transmisión de audio en tiempo real. Los sistemas estudiados son:

- **Adaptive Differential Pulse-Code Modulation (ADPCM):** se utiliza por ejemplo, en “Digital Circuit Multiplication Equipment (DCME)” como técnica de compresión en enlaces de larga distancia, como comunicaciones por satélite o comunicaciones por cable submarinas.
- **Audio Processing Technology Apt-x100:** empleado en micrófonos digitales inalámbricos, Bluetooth, DBS-Radio, etc.
- **Extended Adaptive Multi-Rate Wideband (AMR-WB+):** forma parte de aplicaciones como por ejemplo, Servicio de Mensajería Multimedia (MMS), Subsistema Multimedia IP (IMS) y Packet-switched Streaming Service (PSS).
- **Advanced Audio Coding (AAC):** incluido por ejemplo en radio y televisión digital, DAB+ y DVB-S2, DVB-C2.
- **MPEG-1 Audio Layer II (MP2):** utilizado para audio broadcasting.
- **MPEG-1 Audio Layer III (MP3):** se emplea principalmente como estándar de facto como técnica de compresión para la transferencia y reproducción de música en reproductores de audio digital.
- **Adaptive Transform Coder 3 (AC3):** establecido en DVDs y discos Blu-ray Discs.

En el proyecto se incluye una simulación para el códec de audio AMR-WB+ que permite una evaluación de las diferentes sensibilidades frente a errores de cada bit dentro de la trama de audio codificada.

2 Métodos de Codificación de Audio

2.1 ADPCM

ADPCM proviene de las siglas en inglés Adaptive Differential Pulse-Code Modulation y está basado en la modulación por impulsos codificados (PCM). En ADPCM la señal diferencia calculada a partir de la señal de entrada y una estimación de dicha señal, es cuantizada y enviada al decodificador. Los elementos principales de este sistema son: el cuantizador adaptativo y el predictor adaptativo que utiliza un cierto número de muestras pasadas para realizar la estimación.

El proceso de codificación está representado en la Fig. 2.1

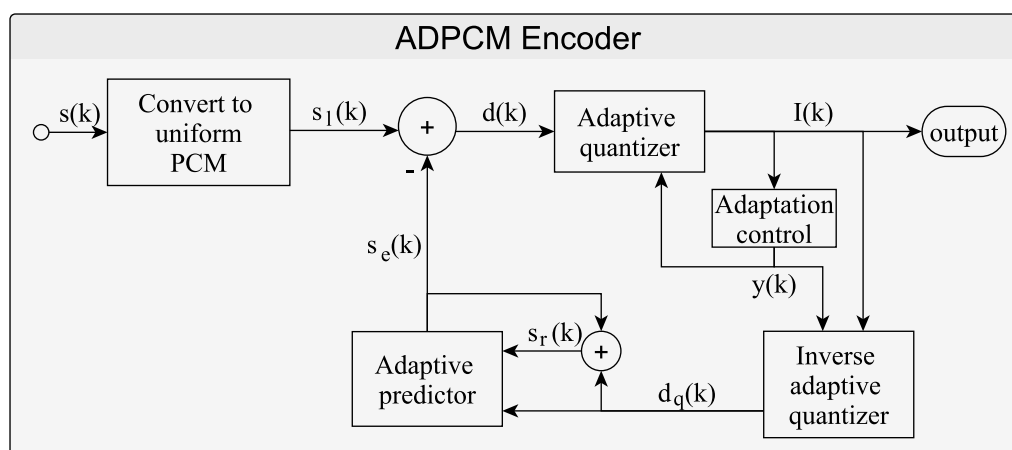


Fig. 2.1. Codificador ADPCM

2.1.1 Técnicas de Gestión de Errores

Las principales técnicas de gestión de errores para ADPCM son las siguientes:

Muting Convencional

El esquema de silenciado convencional de errores en ADPCM utiliza un código de comprobación por redundancia cíclica (CRC) para detectar errores en la secuencia recibida en el decodificador. Si algún error es detectado, se calcula la diferencia entre la actual señal recibida y la anterior. Cuando esa diferencia supera un cierto umbral (definido con anterioridad) la señal recibida se sustituye por la que produce un valor diferencia igual a cero.

“Super Muting” Esquema

Este mecanismo constituye una mejora del anterior esquema. Los valores diferencia que superan un cierto umbral son reemplazados por el valor con mayor probabilidad para el valor recibido. Los valores diferencia esperados $E(R_1R_2R_3R_4)$ se calculan según la ecuación (2.1).

$$E(R_1R_2R_3R_4) = \sum_{d_q(k)=-7}^7 d_q(k) \cdot P(d_q(k)) \cdot P_e(d_q(k)) / \sum_{d_q(k)=-7}^7 P_e(d_q(k)) \quad (2.1)$$

donde $d_q(k)$ corresponde al valor diferencia y $P(d_q(k))$ es la probabilidad de que ocurra el valor $d_q(k)$. La probabilidad del patrón de error actual se calcula empíricamente y está representada por $P_e(d_q(k))$.

Ambos mecanismos están representados en la Fig. 2.2.

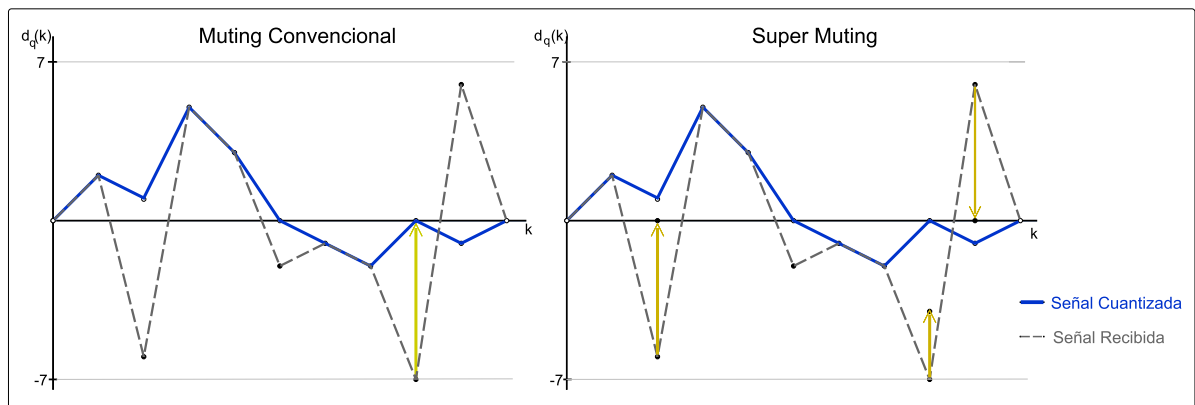


Fig. 2.2. Mecanismos de Muting para ADPCM

Esquema de Detección de Códigos para ADPCM

Se utiliza una estimación de la calidad del canal basada en el número de códigos recibidos con un valor diferencia absoluto mayor que un valor umbral definido de antemano. Si muchos códigos recibidos producen un valor diferencia que supera el umbral, implica que el canal está degradado. Si un valor diferencia mayor que el umbral es detectado y el canal es fiable, existe una probabilidad muy alta de que el valor recibido sea correcto. Por el contrario si el valor supera el umbral pero el canal está muy degradado, es muy probable que el valor sea erróneo.

Ocultación de Pérdida de Paquetes

En el siguiente método se propone la utilización de la repetición de la forma de onda como método de regeneración de la señal ante la pérdida de paquetes. De esta forma se consigue actualizar el cuantizador y el predictor en el decodificador y de este modo minimizar la degradación de la calidad de la señal de audio.

Decodificación basada en técnicas de Soft Decision

Esta aproximación consiste en la modificación del decodificador de modo que utilice softbits (información conjunta de bits recibidos y una estimación de su probabilidad instantánea) en lugar de hardbits (únicamente se utiliza la información recibida). Su aplicación para AD-PCM se traduce en la utilización del dominio lineal para representar la estimación de la señal diferencia en lugar de un dominio logarítmico. Esta representación lineal permite el uso de un estimador MS en lugar de las tablas de cuantización definidas para ADPCM, debido a que entre la calidad de la señal y el cuadrado de la señal residual existe una fuerte correlación. Se añade además otro nuevo estimador MS para añadir una mayor resiliencia frente a errores.

Detección de Click-Noise

Debido a que las señales de audio normales poseen pocos componentes en frecuencias altas y al hecho de que estos valores aumentan para señales de error, las frecuencias que corresponden a errores pueden ser detectadas y eliminadas mediante el uso de filtros paso alto.

Código de Corrección de Errores (ECC)

En este método se utiliza la fiabilidad de las palabras decodificadas para el control de la actualización del cuantizador inverso y para la estabilización del predictor en el receptor. El sistema está representado en la figura Fig. 2.3.

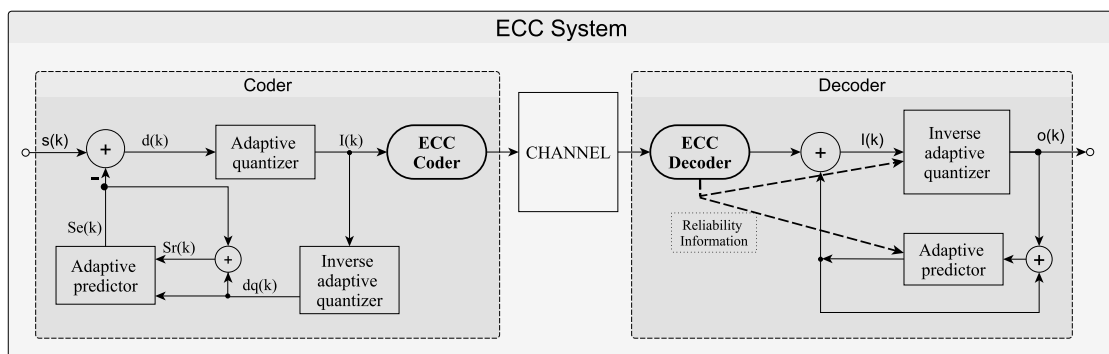


Fig. 2.3. Código de Corrección de Errores (ECC)

2.3 AMR-WB+

El codec de audio AMR-WB+ amplía las funcionalidades de AMR-WB y emplea una tecnología híbrida que incluye un modelo para codificación en el dominio del tiempo: ACELP (Adaptive Code Excited Linear Prediction) y otro modelo de codificación aplicable al dominio transformado: TCX (Transform Coded eXcitation). ACELP se emplea en la codificación de señales de voz mientras que TCX codifica sonidos enriquecidos como por ejemplo música.

La señal de entrada se procesa en bloques de 2048 muestras, se separa en dos bandas de frecuencia y se submuestra para realizar un muestreo crítico. Posteriormente, las bandas son segmentadas para formar bloques de 1024 muestras llamadas “superframes”. En la banda de altas frecuencias se codifica cada superframe con 64 bits, enviándose al decodificador únicamente la energía y la envolvente espectral. En la banda de bajas frecuencias se divide cada superframe en cuatro tramas de 256 muestras que se codifican posteriormente usando ACELP o TCX usando cuatro posibles modos:

- En una trama ACELP de 256 muestras
- En una trama TCX de 256 muestras (trama TCX corta)
- Como parte de una trama TCX de 512 muestras (trama TCX media)
- Como parte de una trama TCX de 1024 muestras (trama TCX larga)

El proceso de codificación puede observarse en la figura Fig. 2.5.

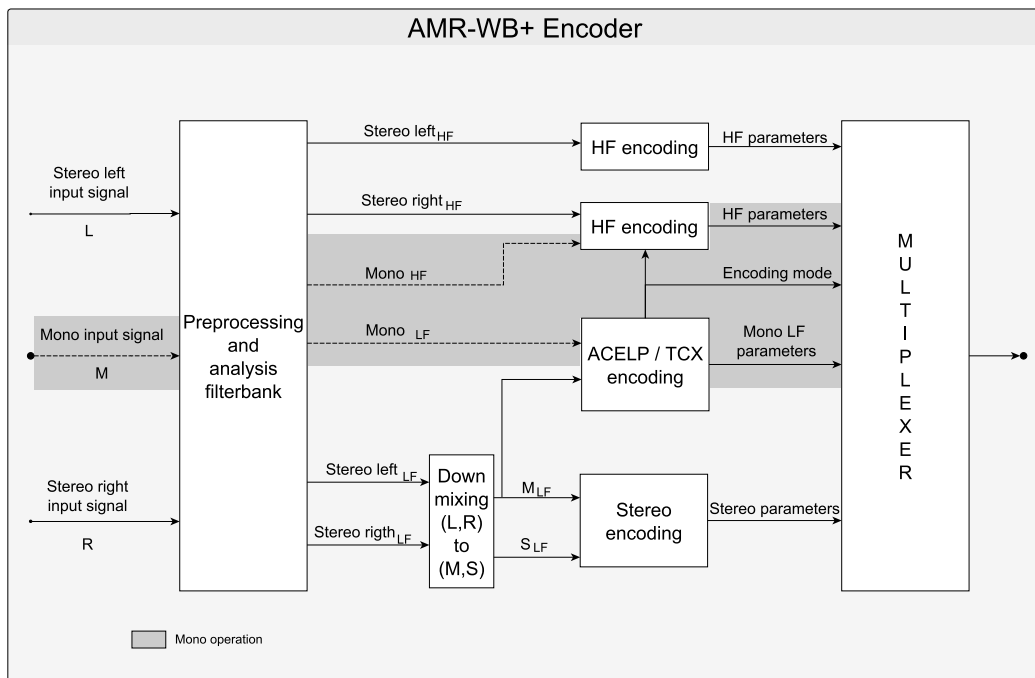


Fig. 2.5. Codificación en AMR-WB+

2.3.1 Técnicas de Gestión de Errores

El códec incluye en el decodificador un mecanismo de gestión frente a pérdida de paquetes en el que diferentes métodos son aplicados en función del tipo de codificación (mono o estéreo).

Decodificación de Señales Mono

Se aplican dos mecanismos:

- Extrapolación de modo.

Dicho mecanismo consiste en la obtención del modo de decodificación de las tramas perdidas a partir de los modos obtenidos con anterioridad. Los modos de decodificación son recibidos en el decodificador según el vector $\text{MODE} = (m_0, m_1, m_2, m_3)$ donde cada m_k representa el modo de cada trama de 256 muestras. Para ello el decodificador utiliza el vector BFI (Bad Frame Indicator) para marcar internamente las tramas perdidas. El vector tiene la forma $\text{BFI} = (\text{bfi}_0, \text{bfi}_1, \text{bfi}_2, \text{bfi}_3)$ donde $\text{bfi}_k = 0$ indica que la correspondiente trama dentro de la supertrama ha sido recibida correctamente, mientras que $\text{bfi}_k = 1$ indica que dicha trama no ha sido recibida. El mecanismo define varias reglas a seguir a la hora de recuperar los modos perdidos.

- Ocultación de tramas TCX defectuosas.

Para tramas TCX largas, se suponen señales casi-estacionarias y sus valores pueden ser interpolados a partir de los valores de las tramas precedentes, utilizando técnicas de extrapolación de fase y amplitud, y conformado del espectro.

Decodificación de señales Estéreo

Para la decodificación de señales estéreo se hace uso del indicador de trama defectuosa (bad frame indicator) junto con procedimientos específicos para las bandas media y baja.

2.4 AAC

El funcionamiento básico del códec AAC consiste en la eliminación de la redundancia de la señal de audio y en la eliminación de los componentes de la señal que son inaudibles para el oído humano. La representación espectral de la señal de entrada se obtiene en el bloque de análisis tiempo-frecuencia. Posteriormente se divide dicho espectro en factores de escala, que son cuantizados y codificados. Para obtener una estimación de los umbrales de enmascaramiento se aplica un análisis psicoacústico. Por último se multiplexan la información y los parámetros de codificación para formar el flujo de información codificada. El proceso de codificación está esquematizado en la figura Fig. 2.6.

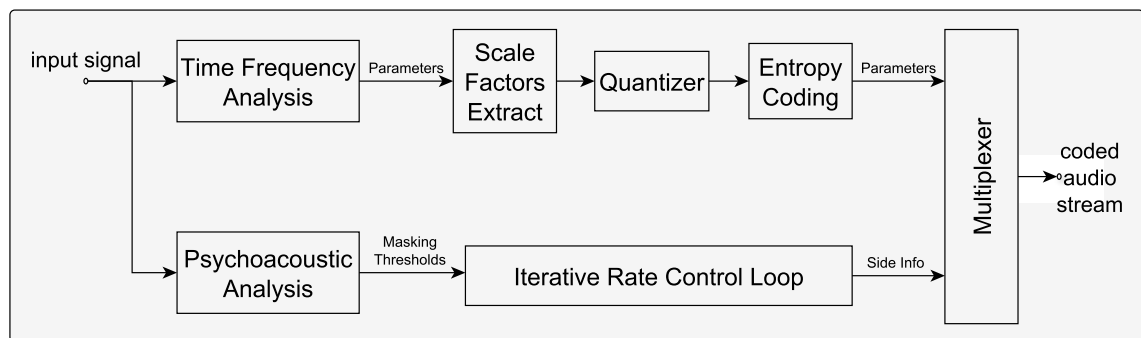


Fig. 2.6. Codificación para AAC

2.4.1 Técnicas de Gestión de Errores

Para el códec AAC existen varios mecanismos de gestión de errores, algunos de ellos están definidos en la documentación estándar de AAC. Los procedimientos descritos en este documento se basan en mejoras realizadas sobre la transformada de coseno discreta modificada (MDCT - Modified Discrete Cosine Transform) y en la gestión de los datos críticos.

Herramientas implementadas en AAC para proporcionar robustez frente a errores

- Reordenación de palabras código Huffman (HCR - Huffman Codeword Reordering). Las palabras código se ordenan según su prioridad de forma que las que tenga una prioridad más alta obtengan una mejor protección.
- Codificación reversible de longitud variable (RVLC - Reversible Variable Length Coding). Se utilizan códigos RVLC en lugar de utilizar códigos Huffman .
- Libro de códigos virtuales (VCB11 - Virtual Codebook 11). Los valores espectrales que superan el valor espectral máximo definido, se almacenan utilizando un libro de códigos específico. Dichos valores elevados representan normalmente ruido procedente de la transmisión y de esta manera se consigue limitar sus efectos en la codificación.

Método de reducción de “Click Noise”

Como primer paso de este procedimiento, el ruido a eliminar debe ser detectado. El ruido es una señal transitoria por lo tanto, el tipo de ventana utilizada en la codificación es ESS (Eight Short Sequence) donde la trama de 2048 muestras es dividida en ocho tramas cortas solapadas. El siguiente paso consiste en encontrar la trama dentro de las ocho posibles que contiene la distorsión debida a este tipo de ruido, para ello se hace uso de la entropía espectral. La trama errónea será la que contenga una entropía espectral mayor que un umbral definido a partir de un procedimiento de decisión bayesiana de minimización de error. Una vez detectado, el ruido debe ser eliminado. Esto se consigue estimando el ruido contenido en la trama corrupta utilizando la desviación media absoluta (MAD) para obtener umbrales blandos. Una vez se ha eliminado el ruido estimado se aplica un suavizado espectral para minimizar el efecto del ruido en las tramas adyacentes.

Recuperación robusta de datos críticos

AAC define un CRC que se aplica sobre los primeros 192 bits de cada trama. Los datos protegidos por el CRC se corresponden con cuatro tipos de datos diferentes:

- Campos constantes, K_n .
- Campos predecibles, E_n , que pueden ser calculados utilizando la información definida por la correlación sucesiva entre paquetes consecutivos R_n .
- Campos desconocidos, D_n
- Otros campos, O_n . Contienen datos que son menos críticos y cuyos valores pueden ser decodificados utilizando los métodos apropiados.

Si se produce una pérdida de paquetes se podrían recuperar los campos desconocidos aplicando la siguiente ecuación:

$$\hat{u}_{\text{MAP}} = \arg \max_{D \in \Omega_D} P(D|K, E, C, R_D, R_O, R_c) \quad (2.2)$$

donde C contiene la información sobre las tramas precedentes.

Procedimiento de resistencia frente a errores

En este último método se aplica intercalado de los datos dentro de las tramas, para conseguir una mejor protección de la información crítica sin la cual el proceso de decodificación no puede llevarse a cabo correctamente. Los mecanismos empleados para ello son: inserción de múltiples copias de la información crítica dentro de varios paquetes o formación de paquetes de datos independientes para cada tipo de información, de forma que estos puedan ser transportados a través de diferentes interfaces de red. Los datos críticos se enviarían utilizando el

protocolo TCP y el resto de la información se enviaría por medio de paquetes UDP.

Los esquemas propuestos para proporcionar mayor resistencia frente a errores son:

- Asignación de posiciones fijas dentro de cada trama para cada tipo de información.
- Codificación lineal de los factores de escala. Los valores de los factores de escala se aproximan utilizando un modelo lineal y un método de mínimos cuadrados.

Se emplean tres mecanismos de gestión de errores:

- silenciado de datos críticos, ya que si los paquetes se envían usando diferentes protocolos de red las pérdidas de datos críticos son muy improbables.
- aproximación lineal para recuperar los factores de escala.
- interpolación lineal de los coeficientes espectrales QMDCT a partir de muestras adyacentes.

2.5 MPEG-1 Capa II y Capa III

MPEG-1 es un mecanismo de compresión en la que la información irrelevante es eliminada utilizando técnicas perceptuales. Debido a las limitaciones en la percepción del oído humano, no se produce ninguna distorsión audible. MPEG-1 proporciona una arquitectura de tres capas con diferentes complejidades, retardo y calidad del audio en función de la aplicación. Cada capa implementa mejoras en la compresión, añadiendo complejidad al codificador y al decodificador.

- **Capa I** es la más simple. La señal de entrada es mapeada en 32 subbandas, los datos son formateados para incluirlos en bloques, se aplica asignación adaptativa de bits utilizando un modelo psicoacústico y se cuantizan los datos obtenidos.
- **Capa II** amplía las funcionalidades incluyendo factores de escala, un formato de trama diferente y utilizando un procedimiento de asignación de bits modificado.
- **Capa III** es la capa más compleja. Utiliza un banco de filtros con una resolución en frecuencia más precisa, un cuantizador no uniforme, segmentación adaptativa y codificación de entropía.

En la figura Fig. 2.7 se muestra el diagrama del proceso de codificación.

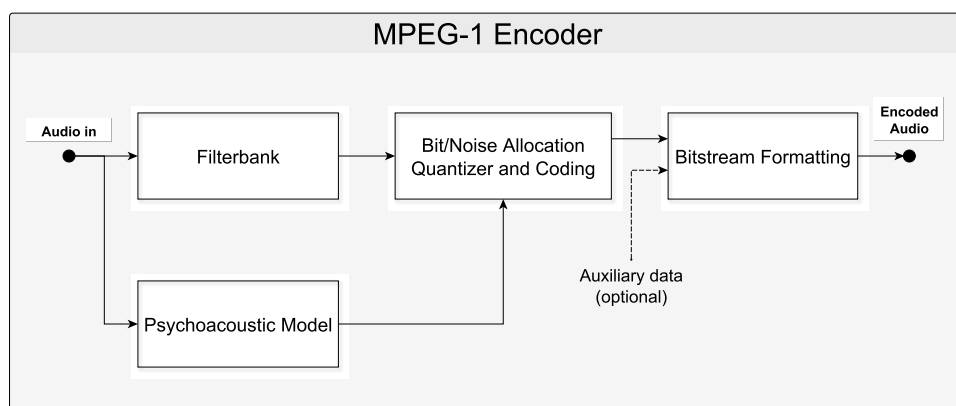


Fig. 2.7. Codificación MPEG-1

2.5.1 Técnicas de Gestión de Errores

Como método de detección de errores MPEG-1 proporciona un CRC opcional, controlado por la activación/desactivación de un bit (“bit de protección frente a errores”) situado al inicio de la trama. Si el valor de dicho bit es 0, se aplica detección de errores y la palabra de 16 bits que sigue a la cabecera es examinada. En las tres capas definidas para MPEG-1 se utiliza el mismo campo de CRC, sin embargo en cada una de ellas se protegen diferentes campos.

El algoritmo de detección empleado es CRC-16 con un polinomio generador de la forma:

$$x^{16} + x^{15} + x^2 + 1$$

Los bits incluidos en el cálculo del CRC para MPEG-1 Capa I se obtienen según:

$$4 * [\text{numero_canales} * \text{limite_intensidad_estereo} + (32 - \text{limite_intensidad_estereo})]$$

Para codificación mono, $\text{numero_canales} = 1$ and $\text{limite_intensidad_estereo} = 32$ por lo que el número total de bits es 128. Para **Capa II**, el cálculo incluye más variables y es por lo tanto más complejo. Para **Capa III**, los campos protegidos corresponden a la información adicional relativa a la codificación que siguen a la cabecera. El método empleado para el cálculo del CRC-16 se muestra en Fig. 2.8.

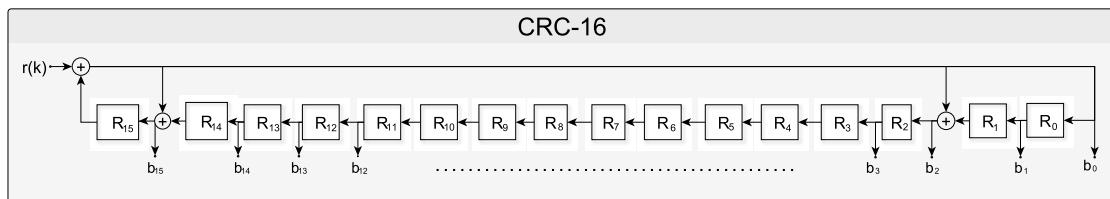


Fig. 2.8. Mecanismo de CRC para MPEG-1

En MPEG-1 el valor del campo CRC recibido se almacena en el decodificador. Posteriormente se calcula el valor del CRC correspondiente a la trama recibida, haciendo pasar sus bits a través del circuito anterior para obtener como resultado los valores b_0, b_1, \dots y b_{15} . Dichos valores son comparados con los valores almacenados en el decodificador, en el caso de que los valores sean diferentes es necesario aplicar técnicas de ocultación de errores.

2.6 AC3

La base del mecanismo de compresión del códec AC3, consiste en la generación de una versión digital de la señal de audio que utiliza un número mínimo de bits para su representación, mediante el empleo de técnicas de codificación perceptual. La descripción del proceso de codificación se muestra en la figura Fig. 2.9.

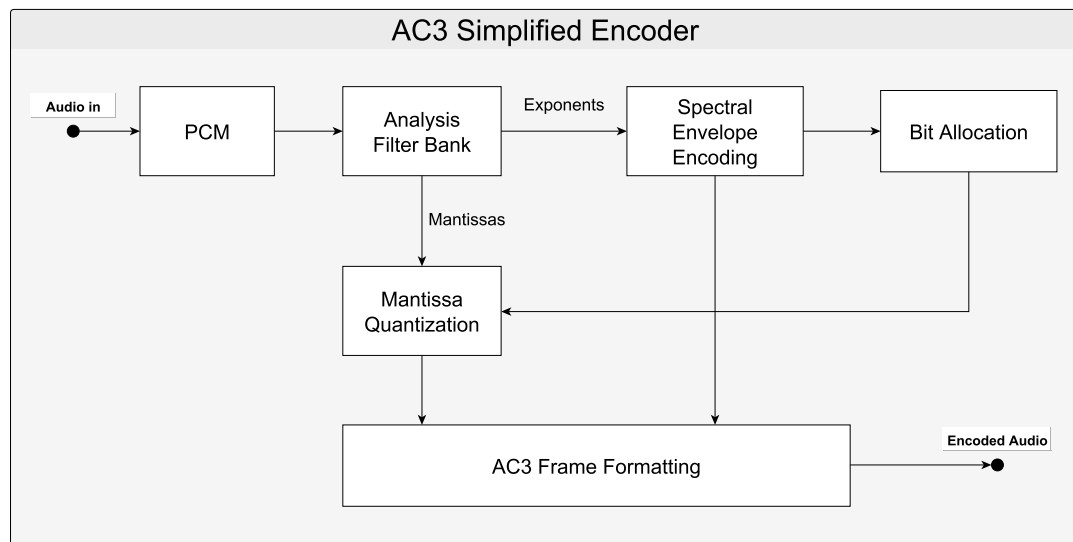


Fig. 2.9. Codificación en AC3

2.6.1 Técnicas de Gestión de Errores

Como técnica de corrección de errores, AC3 proporciona dos CRC que pueden ser utilizados para detectar la presencia de errores dentro de la trama. Sin embargo, el códec no implementa ningún mecanismo a aplicar cuando los errores son detectados.

Comprobación del CRC

Las dos palabras correspondientes al CRC incluidas en AC3 tienen una longitud de 16 bits. La primera de ellas (crc1) está localizada al inicio de la trama, después de la palabra de sincronización y la segunda (crc2) se encuentra en la parte final. Ambos CRC cubren diferentes segmentos de la trama: crc1 cubre los primeros 5/8 sin incluir la palabra de sincronización y crc2 cubre los 3/8 restantes. El diseño del códec asegura que al menos dos bloques de audio están contenidos en los primeros 5/8. Permitiendo que el primer CRC cubra una parte de la trama, se consigue que el proceso de decodificación pueda comenzar sin que sea necesario recibir la trama completa. La función polinómica utilizada en la generación de ambos CRC es la siguiente:

$$x^{16} + x^{15} + x^2 + 1 \quad (2.3)$$

La función que calcula la parte de la trama sobre la que se aplica el primer CRC es:

```
num_bits_crc1 = truncate( tamaño_trama/2) + truncate(tamaño_trama/8)
or
num_bits_crc1 = (int) (tamaño_trama >> 1) + (int) (tamaño_trama >> 3)
```

Para la implementación del cálculo del CRC existen varios métodos, uno de ellos consiste en la implementación hardware basada en Registros Lineales de Desplazamiento con Realimentación (LFSR). Los mecanismos LFSR incluyen varios registros de desplazamiento junto con un pequeño número de puertas XOR tal como se puede apreciar en Fig. 2.10. Para el cálculo del CRC en el decodificador, se hacen pasar todos los bits de la trama a través de los registros de desplazamiento, de forma que si la transmisión está libre de errores el valor de los registros una vez finalizado el proceso debe ser cero.

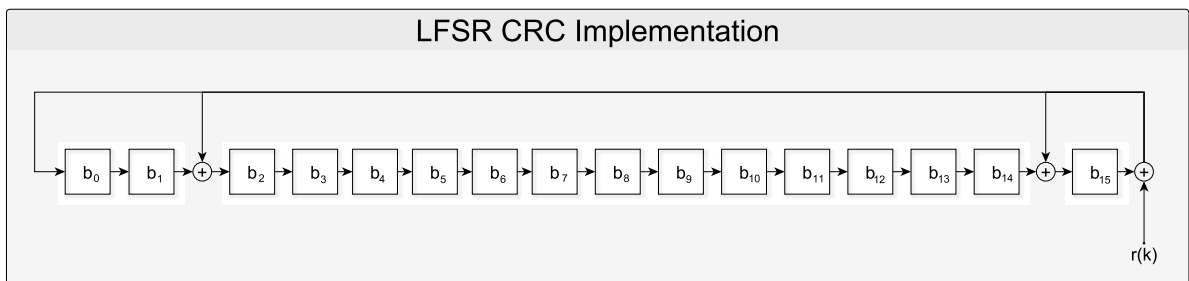


Fig. 2.10. Implementación de CRC basada en LFSR

3 Simulación

Con motivo de evaluar la respuesta de un códec de audio frente a errores de bit se realizó una simulación en Matlab. El códec elegido fue AMR-WB+ debido a su inherente robustez frente a errores y a sus buenas prestaciones para audio y voz. Con ayuda de la simulación se realizó una investigación de la sensibilidad frente a fallos de cada bit dentro de cada tipo de trama AMR-WB+. Como datos de entrada se utilizaron 13 ficheros de audio correspondientes a señales mono, codificadas usando codificación PCM de 16 Bits y con tasa de muestreo de 48kHz. El programa generado con Matlab simula la modificación de un bit dentro de cada tipo de trama en el fichero codificado, la posterior decodificación del fichero de audio modificado y la evaluación de la calidad perceptual de la decodificación, usando la herramienta “PEAQ Eval”. Como resultado de la simulación se obtiene el “valor medio del grado de diferencia objetivo” (ODG) para cada bit. Utilizando los valores ODG para cada uno de los 13 ficheros de entrada se obtienen cuatro gráficos que se corresponden con los cuatro tipos de tramas de AMR-WB+.

Los resultados se pueden agrupar en dos gráficos, el obtenido para tramas ACELP y el obtenido para tramas TCX cortas ya que los resultados referidos a tramas TCX son muy similares. Los gráficos obtenidos así como los grupos de bits con mayor sensibilidad frente a fallos se muestran en las figuras Fig. 3.1 y Fig. 3.2.

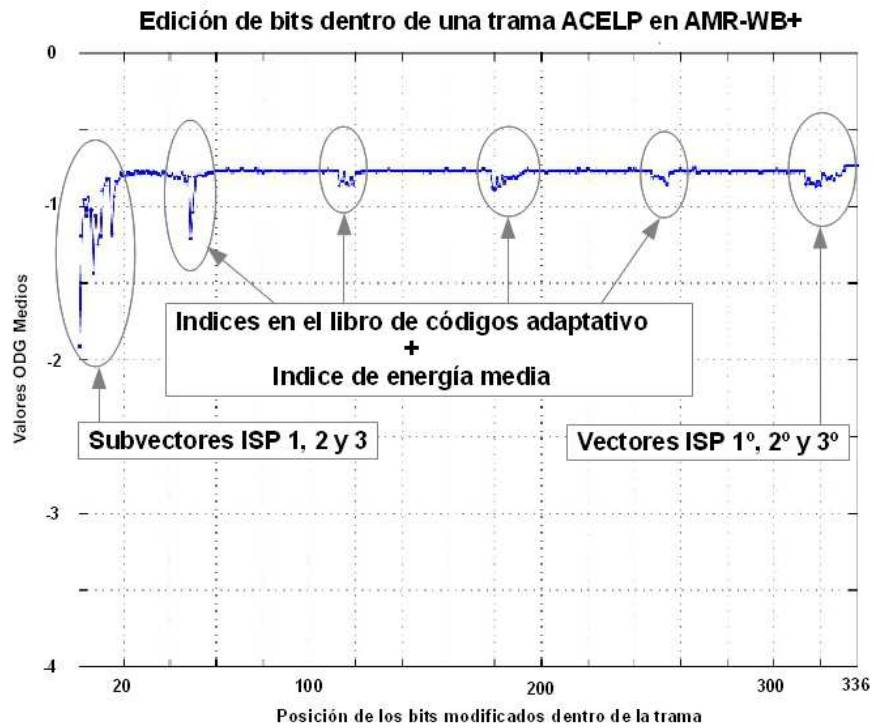


Fig. 3.1. Modificaciones en una trama ACELP

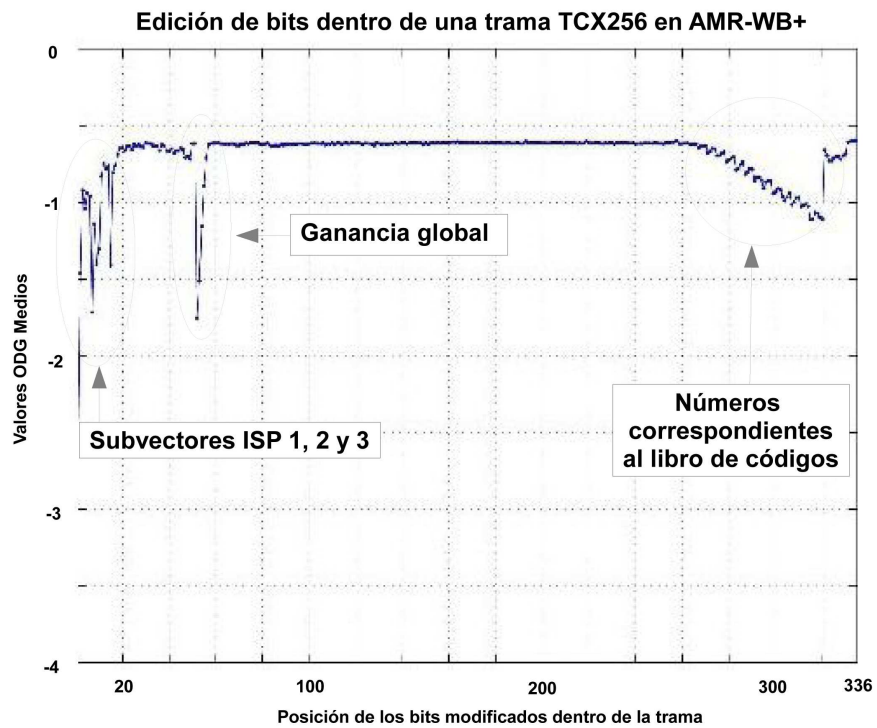


Fig. 3.2. Modificaciones en una trama TCX256