

## Tema 6: Codificación y compresión de audio.

- ◆ Introducción.
- ◆ Características del audio.
  - Digitalización.
  - Calidad de una señal de audio.
  - Parámetros específicos.
- ◆ Compresión de audio.
  - Calidad telefónica.
  - Calidad CD.

### 1. Introducción.

- Las secuencias de audio forman parte de las aplicaciones multimedia.
- El estudio de la codificación y compresión se puede enfocar en función de la aplicación:
  - ✓ Aplicaciones interactivas (audio-conferencia audio) → *codecs* simétricos.
  - ✓ Aplicaciones de difusión y reproducción de medios (TV digital, audio Hi-Fi, DVD, etc.) → *codecs* asimétricos
- Características de una señal de audio.
- Distintos tipos de calidad de audio.
- Técnicas de compresión de audio.

## 2. Características del audio

- ❑ Una señal de audio no es más que una onda acústica (variaciones de presión del aire).
- ❑ El rango de frecuencias audible está entre 20Hz y 20KHz.
- ❑ La relación de dos sonidos A y B se mide en decibelios:  $dB=20 \log_{10} (A/B)$ .
  - ✓ 0 dB: Menor sonido audible y 120dB: Umbral del dolor.
  - ✓ Rango dinámico:  $10^6$ .
- ❑ El oído es muy sensible a las variaciones de sonido de corta duración (ms) al contrario que el ojo humano.

3

### 2.1 Digitalización y cuantificación.

- ❑ La digitalización de las señales de audio se realizan mediante convertidores A/D.
  - ✓ Muestran la señal analógica de audio a una frecuencia determinada.
    - Según Nyquist:  $S_f \geq 2 B_w$
- ❑ Cuantificación: Las muestras obtenidas se codifican en un número finito de bits
  - ✓ Error de cuantificación (*quantification noise*).
  - ✓ Codificación lineal o logarítmica.
- ❑ PCM (*Pulse Code Modulation*).
  - ✓ Usado para la digitalización de señales de audio.
  - ✓ Parámetros:  $S_f$ , bits/muestra, niveles de cuantificación\*

4

## Digitalización: Interfaz MIDI

- MIDI (*Musical Instrument Digital Interface*).
  - ✓ Utilizado para codificar música (instrumentos).
  - ✓ Codifica los elementos básicos (notas, silencios, ritmos, etc.) en mensajes MIDI.
  - ✓ Cada instrumento tiene su propio código (hasta 127)
  - ✓ Un sintetizador interpreta los mensajes MIDI y produce la señal de audio correspondiente.
  - ✓ Ventaja:
    - Reduce mucho el ancho de banda necesario (factor de 1000 !!)
  - ✓ Inconvenientes:
    - Necesidad de un sintetizador en ambos extremos (calidad de sonido diferente).
    - Aplicable solo a música.

5

## 2.2 Calidad de una señal de audio.

- Voz (telefonía)
  - ✓ Se define para los servicios de telefonía digital.
  - ✓ Estándar G.711 (ITU): Codificación logarítmica.
    - Japón y USA: Transformación  $\mu$ -law.
    - Resto: Transformación A-law.
  - ✓ Parámetros:
    - Señal de audio de 3.5 KHz (BW).
    - $S_f = 8$  KHz
    - 8 bits/muestra.
    - Tasa de bits: 64Kbps (N-ISDN).
  - ✓ Otras técnicas de codificación y compresión:
    - DPCM y ADPCM,
    - G.72x,
    - GSM,
    - LPC y CELP,
    - etc.

6

## Calidad de una señal de audio.

- CD-Digital Audio.
  - ✓ Calidad de audio superior: Sonido Hi-Fi estereofónico.
  - ✓ Utiliza una codificación lineal. Las diferencias de amplitud deben ser respetadas por igual.
  - ✓ Parámetros:
    - Señal de audio de 20 KHz (BW).
    - $S_f = 41.1$  KHz
    - 16 bits/muestra.
    - Soporta estereofonía (dos canales)
    - Tasa de bits: 1.411 Mbps.
  - ✓ Otros estándares utilizan esta calidad de audio:
    - DAT (32.4 y 48 KHz),
    - MPEG (32, 44,1 y 48 KHz),
    - DVI,
    - etc.

7

## 2.3 Parámetros específicos.

- Tasa de bits (*throughput*):
  - ✓ Audio sin comprimir:
    - Calidad telefónica: 64Kbps.
    - Calidad CD: 1.411 Mbps.
  - ✓ Audio comprimido:
    - Calidad telefónica: 32, 16, 4 Kbps (ADPCM, CELP)
    - Calidad CD: 192 Kbps. (MPEG audio)
- Retardo de tránsito (aplicaciones interactivas)
  - ✓ Conversación:
    - Telefonía: < 25 ms (evitar echo).
    - 100 a 500 ms (sensación de tiempo real).

8

## Parámetros específicos.

- Varianza del retardo ( *jitter* ).
  - ✓ Es el parámetro más crítico para los  *streams*  de audio.
  - ✓ Solución:
    - Técnicas de ecualización del retardo.
      - Se suministra un tiempo adicional antes de comenzar la reproducción, almacenando los paquetes en un buffer de entrada.
    - Consecuencias:
      - Incrementamos el retardo total.
      - Necesitamos recursos de memoria para el buffer de ecualización.
  - ✓ Compromiso entre la capacidad de almacenamiento y el máximo  *jitter*  tolerable por la aplicación.
- Tasas de error:
  - ✓ Calidad telefónica:  $< 10^{-2}$ , Calidad CD:  $< 10^{-3}$

9

## Tema 6: Codificación y compresión de audio.

- ◆ Introducción.
- ◆ Características del audio.
  - Digitalización.
  - Calidad de una señal de audio.
  - Parámetros específicos.
- ◆ **Compresión de audio.**
  - **Calidad telefónica.**
  - **Calidad CD.**

### 3.1 Calidad telefónica

□ Codificación: ITU G.711

- ✓ Muestreo: 8KHz. Muestra: 16 bits. Codificación log.
- ✓  $\mu$ -law y A-law: Codifican una muestra PCM de 16 bits (14 bits) en 8 bits.
  - 8 segmentos: 0xFF, 0x1FF, 0x3FF, 0x7FF, 0xFFF, 0x1FFF, 0x3FFF, 0x7FFF

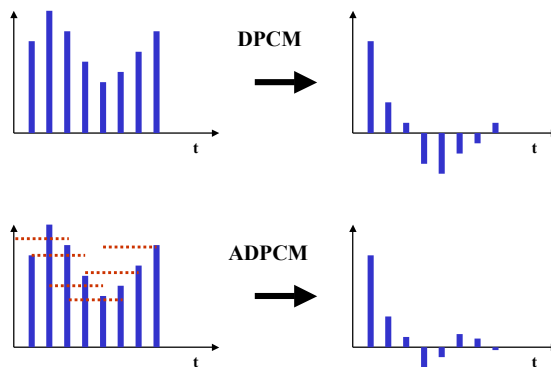
Cod. lineal	A-Law	Cod. Lineal	u-Law
000000wxyz	000wxyz	0000001wxyz	000wxyz
000001wxyz	001wxyz	000001wxyzab	001wxyz
00001wxyzab	010wxyz	00001wxyzabc	010wxyz
0001wxyzabc	011wxyz	0001wxyzabcd	011wxyz
001wxyzabcd	100wxyz	001wxyzabcde	100wxyz
01wxyzabcde	101wxyz	01wxyzabcdef	101wxyz
1wxyzabcdef	110wxyz	1wxyzabcdefg	110wxyz
	111wxyz		111wxyz

11

### Algoritmos de compresión (Voz)

□ Compresión G.721:

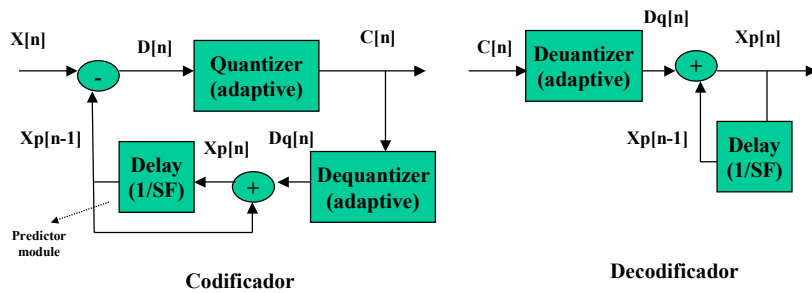
- ✓ ADPCM (*Adaptive Differential Pulse Code Modulation*).
  - Codifica la diferencia entre la muestra actual y una estimación basada en las últimas "n" muestras.
  - Muestreo: 8 KHz. Muestra: 16 bits. Diferencias: 4 bits.



12

## Ejemplo ADPCM

- Algoritmo ADPCM (IMA: *Interactive Multimedia Association*)
  - ✓ Algoritmo de dominio público. Calidad de audio e índice de compresión aceptables.
  - ✓ Sencillo y capaz de trabajar en tiempo real (*software*).
  - ✓ Índice de compresión: (PCMbites/4) a 1.



13

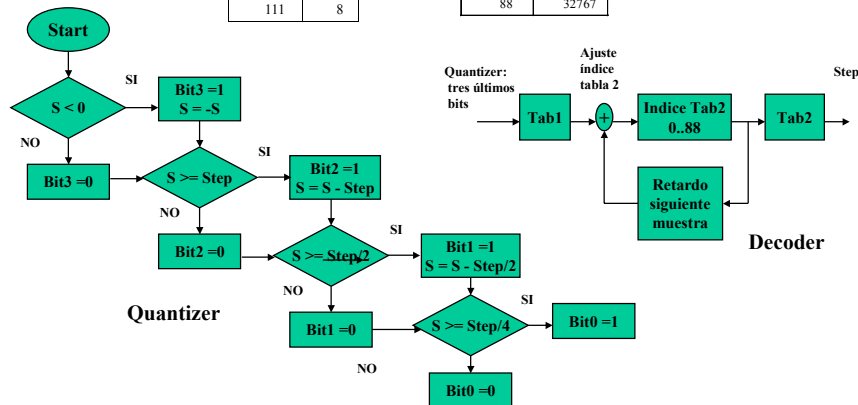
## IMA ADPCM

**Tabla 1**  
Ajuste sobre la tabla 2 para decidir las diferencias a aplicar.

B <sub>3</sub> B <sub>2</sub> B <sub>0</sub>	Ajuste
000	-1
001	-1
010	-1
011	-1
100	2
101	4
110	6
111	8

Índice	Step
0	7
1	8
2	9
...	...
50	876
51	963
...	...
88	32767

**Tabla 2**  
Distribuye el rango dinámico de la muestra PCM



14

## Calidad telefónica

### □ Otros:

- ✓ G.722: *Sub-Band ADPCM*.
  - Codifica señales de audio de hasta 7 KHz.
  - Descompone la señal en dos bandas de 4 KHz.
  - A cada banda le aplica ADPCM.
  - Muestreo: 16 KHz. Muestra: 14 bits
  - Tasas de bits: 48, 56 y 64 Kbps.
- ✓ G.723, G.726, G.727:
  - Variantes del G.721 (ADPCM).
- ✓ LPC (*Linear Predictive Coding*) US-FS-1015
  - Define un modelo analítico del aparato fonador.
  - Reduce cada segmento de audio a los parámetros del modelo que más se aproximan al original.
  - El decodificador recoge estos parámetros y sintetiza la voz correspondiente.
  - LPC-10E puede bajar hasta 2.4 Kbps.

15

## Calidad telefónica

- ✓ CELP (*Code Excited Linear Prediction*) US-FS-1016.
  - Es una versión mejorada del LPC.
  - Diferencia:
    - Utiliza un code-book con secuencias predefinidas para aplicarlas a cada frame de audio, eligiendo aquella que más se aproxima al original. Además, calcula los errores cometidos.
    - Se envían los parámetros y la versión comprimida de los errores.
  - Tasa de bits de hasta 4.8 Kbps.
- ✓ Variantes CELP:
  - GSM,
  - VSELP,
  - LD-CELP,
  - ITU G.729,
  - QCELP,
  - MELT, etc.

16

## Calidad telefónica

- Tabla resumen de algunos *codecs* de audio.

Año	Tasa de bits (Kbps)	Nombre	MOS
1972	64	PCM (PSTN)	4.4
1976	2.4	LPC-10	2.7
1984	32	G.721 ADPCM	4.1
1990	4.15	INMARSAT	3.2
1991	13	GSM	3.6
1991	4.8	CELP (US 1016)	3.2
1992	16	G.728 (LD-CELP)	4
1992	8	VSELP	3.5
1993	1-8	QCELP	3.4
1995	8	G.729	4.2
1995	6.3	G.723.1	3.98
1995	5-6	Half-Rate GSM	3.4
1996	2.4	New LPC	3.3

17

## 3.2 Calidad CD

- Estándares MPEG/audio (Estándar ISO)
  - ✓ MPEG (*Moving Pictures Expert Group*)
  - ✓ MPEG/audio ofrece altos índices de compresión, manteniendo la calidad del audio del stream original.
  - ✓ Son algoritmos de compresión con pérdidas\*.
- MPEG-1 /audio
  - ✓ Muestras: 32, 44.1 y 48 KHz.
  - ✓ Soportan uno o dos canales (diferentes modos de operación).
  - ✓ Tasas de bits: 32 a 256 Kbps/canal.
  - ✓ Índices de compresión: 2.7 a 24.
- MPEG-2 /audio
  - ✓ Compatibilidad hacia atrás con MPEG-1.
  - ✓ Diseñado para sistemas de sonido multicanal.

18

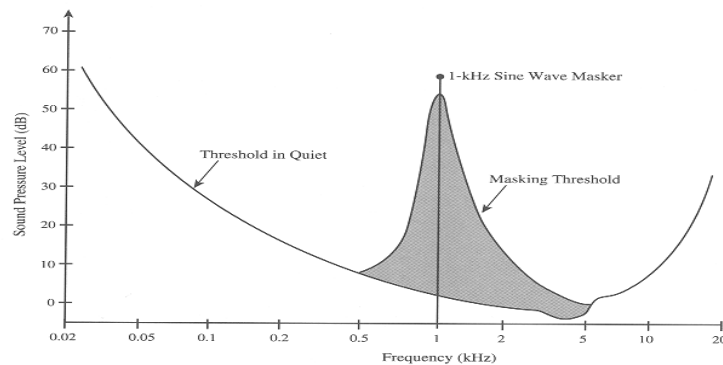
## MPEG-1 audio.

- ❑ El *stream* comprimido puede incluir información auxiliar (acceso aleatorio, avance y retroceso rápido, CRC, etc.)
- ❑ Arquitectura de tres niveles
  - ✓ MPEG-1 Nivel I:
    - El más sencillo. Tasa de bits 192 Kbps/canal. Aplicaciones: Philips DCC
  - ✓ MPEG-1 Nivel II:
    - Complejidad media. Tasa de bits 128 Kbps/canal. Aplicaciones: DAB, CD-I, Vídeo CD.
  - ✓ MPEG-1 Nivel III:
    - El más complejo. Ofrece la mejor calidad de audio con tasas de bits sobre 64 Kbps/canal. Está preparado para N-RDSI.
- ❑ Existen *codecs* hardware de los tres niveles para aplicaciones de tiempo real.

19

## MPEG audio: Fundamentos.

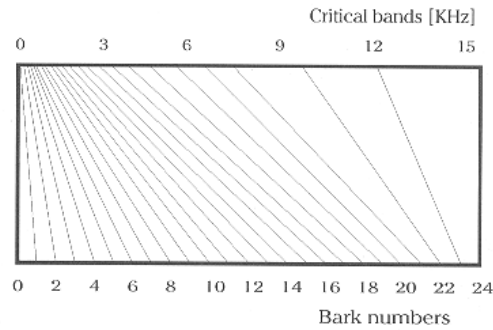
- ❑ Se basa en la capacidad de percepción que tiene el oído humano (modelos *psico-acústicos*)
  - ✓ Enmascaramiento de señales débiles (*noise masking*):



20

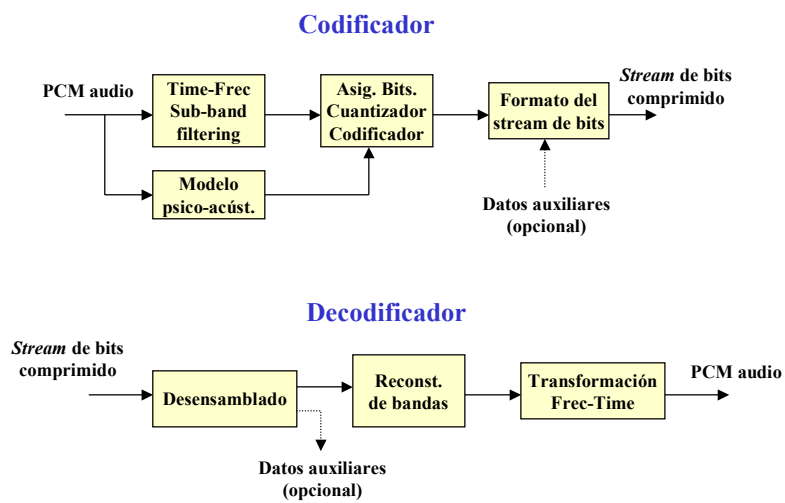
## MPEG audio: Fundamentos.

- Discriminación frecuencial limitada.
  - ✓ La agudeza (selectividad) del oído humano a baja frecuencia es muy superior que en altas frecuencias (*sub-band coding*)



21

## MPEG Audio: Diagrama de bloques



22

## MPEG-1 audio: Niveles.

- Nivel I:
  - ✓ Se divide la señal de audio en 32 bandas de 750 Hz.
  - ✓ Tasa de muestreo: 48 KHz. Tamaño de trama: 384 muestras
  - ✓ El umbral de enmascaramiento (SMR) se calcula con una FFT de 512 puntos (modelo psico-acústico).
  - ✓ Para cada sub-banda se escoge uno de los 15 cuantizadores definidos en función del SMR y la tasa de bits requerida.
- Nivel II:
  - ✓ Utiliza un tamaño de trama de 1152 muestras, una FFT de 1024 puntos (cálculo del SMR) y una cuantización más fina.
- Nivel III:
  - ✓ Incrementa la resolución en frecuencia de las 32 bandas (MDCT), utiliza un modelo psico-acústico más elaborado, y añade una etapa de compresión Huffman.

23

## MPEG Audio: Calidad de audio

- Parámetros de calidad objetivos:
  - ✓ MSE (*Mean Square Error*).
    - Calcula el error cuadrático medio entre la señal original y la reconstruida con el *codec*.

$$MSE = \frac{1}{N} \sum_{i=1}^N \left( s(i) - s'(i) \right)^2$$

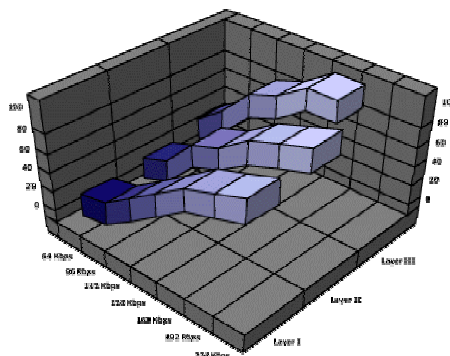
- ✓ SNR (*Signal-to-Noise Ratio*)
  - Relación logarítmica entre dos señales. Se utilizará para comparar la señal original con el error introducido por el *codec*.
  - Se expresa en decibelios (dB).

$$SNR(dB) = 10 \log_{10} \frac{\frac{1}{N} \sum_{i=1}^N s(i)^2}{MSE} \quad PSNR = 10 \log_{10} \frac{255^2}{MSE}$$

24

## MPEG Audio: Calidad de audio.

- Parámetros de calidad subjetivos:
  - ✓ MOS (*Mean Opinion Score*): MPEG define una serie de tests para determinar la calidad de audio generada por cada nivel.
- Resultados:
  - ✓ Fuente:
    - Estéreo, 16bits, 48KHz, 256 Kbps
  - ✓ Compresión 6:1
  - ✓ En condiciones de escucha óptimas, expertos en audición han sido incapaces de distinguir secuencias comprimidas de sus originales.



25