

Introducción al audio y la percepción humana

El oído humano tiene limitaciones tanto en el dominio del tiempo como en el de la frecuencia. En el dominio temporal, no puede detectar señales de muy corta duración (menores a 1 milisegundo), y además, las señales fuertes enmascaran otras más suaves que ocurren cerca en el tiempo. En el dominio frecuencial, ocurre algo similar: señales potentes en una determinada frecuencia pueden ocultar otras cercanas, más débiles. Esta propiedad se aprovecha en la compresión con pérdida basada en la percepción.

Aplicaciones del audio digital

El audio digital se utiliza en numerosos contextos: desde comunicaciones por voz (VoIP) hasta la transmisión de audio en CD, DVD, sistemas de radiodifusión digital como DAB y DVB, y plataformas de streaming por Internet. También está presente en los enlaces entre estudios de grabación y emisoras, en presentaciones cinematográficas y en tecnologías como MIDI, un estándar técnico para instrumentos musicales. MIDI no transmite sonido real, sino información estructurada como notas, duración y velocidad, que luego es interpretada por hardware o software para generar sonido.

Señales analógicas y su representación

Las señales de audio pueden analizarse tanto en forma de onda temporal como en forma de espectro de frecuencias. Las herramientas como Audacity permiten visualizar ambas representaciones, siendo útiles para entender cómo se procesa el sonido.

Modulación y digitalización del audio

Las señales analógicas de banda base pueden convertirse en digitales para su transmisión mediante distintas técnicas de modulación:

- Modulación analógica (AM, FM) para canales analógicos.
- Modulación digital (PSK, FSK, QAM) para canales analógicos con señales digitales.
- Modulación por pulsos (PCM, PAM) para digitalizar señales analógicas.
- Modulación por densidad de pulsos (PDM), como en los Super Audio CD.

La modulación por código de pulsos (PCM) es la base de la digitalización del audio. Existen variantes como DPCM, ADPCM, delta-modulación y delta-sigma modulación.

PCM: muestreo, cuantificación y codificación

El proceso de PCM incluye tres fases:

1. Muestreo: debe hacerse al menos al doble de la frecuencia máxima de la señal. La precisión depende del recorte de frecuencias y de la exactitud del reloj.
2. Cuantificación: puede ser uniforme (LPCM) o no uniforme (como las leyes A o μ).
3. Codificación: determina los bits por muestra.

Las bandas de frecuencia del audio suelen clasificarse en:

- Banda estrecha (300–3400 Hz, 8 kHz)
- Banda ancha (50–7000 Hz, 16 kHz)
- Banda súper ancha (50–14000 Hz, 32 kHz)
- Banda completa (20–20000 Hz, 48 kHz)

Codecs de audio

Un codec convierte y comprime señales de audio, ya sea para almacenamiento o transmisión. Algunos ejemplos incluyen:

- AMR-NB: con pérdida, diseñado para voz.
- Dolby Digital: con pérdida, usado en cine y televisión.
- Dolby TrueHD: sin pérdida, para entretenimiento doméstico.

Existen distintas clasificaciones de códecs según su fuente (voz, música, telefonía), su ancho de banda, su bitrate (habitualmente entre 4,8 y 16 kbps), si son con o sin pérdida, la calidad, la latencia, la licencia, el coste y su complejidad.

Compresión con pérdida y técnicas psicoacústicas

Los codecs con pérdida se basan en las propiedades del oído humano:

- Umbral de audición
- Enmascaramiento simultáneo (frecuencias fuertes tapan débiles)

Se aplican técnicas de:

- Dominio temporal: como codificación predictiva lineal (LPC), usada en voz.
- Dominio frecuencial: mediante transformadas como la MDCT, usada en códecs como MP3 o CELT.

Algunos códecs, como G.718, combinan ambos dominios.

Ratio de compresión

El ratio de compresión puede calcularse en función de la frecuencia de muestreo y los bits por muestra. Por ejemplo, la entrada de 64 kbps se puede comparar con cualquier bitrate de salida para obtener el ratio.

Estructura: por muestras vs por tramas

- Basado en muestras: cada muestra se codifica individualmente (e.g., PCM, ADPCM).
- Basado en tramas: se agrupan varias muestras para analizar correlaciones (e.g., G.723.1, G.729).

Latencia en códecs de audio

La latencia depende del tipo de aplicación:

- Baja latencia: necesaria en comunicaciones en tiempo real (VoIP, satélite).
- Alta latencia: aceptable en transmisión o almacenamiento (radio, vídeo).

Factores que influyen en la latencia:

- Procesamiento (hardware)
- Algoritmo: tamaño de trama, tiempo de anticipación (look-ahead)
- Cálculo total: el retardo del códec es el doble de la longitud de trama más el look-ahead.

Bitrate constante vs variable

- CBR (Constant Bitrate): más simple, pero menos eficiente.
- VBR (Variable Bitrate): ajusta el bitrate según la complejidad del audio, mejor calidad con menor tamaño, pero más complejo de codificar. Se usa tanto en compresión con pérdida como sin pérdida (ej. FLAC, Opus, AAC).

Canales de audio

Se pueden mezclar canales para optimizar el tamaño:

- Estéreo simple: dos canales independientes.
- Estéreo medio-lateral (Mid-side): divide en canal medio y canal lateral, ideal para señales similares en ambos canales.
- Estéreo por intensidad: codifica un solo canal y añade información direccional. Eficiente a bajo bitrate.

Ejemplo: MP3

- Versión: MPEG-1, MPEG-2, MPEG-2.5 Audio Layer III
- Solo define el decodificador, los codificadores varían (ej. LAME)
- Hasta 5.1 canales
- Soporta CBR y VBR
- Algoritmos: MDCT, subbandas
- Tuvo disputas de licencias y patentes

Otros ejemplos comunes:

- AAC: sucesor de MP3, parte de MPEG-2/4, con patentes
- Vorbis: libre, basado en MDCT, usado en Ogg/WebM
- Opus: libre, ideal para tiempo real, combina CELT y SILK

El caso del habla

El audio de voz tiene características distintas de la música:

- Voz sonora: con armónicos (distintos para hombres y mujeres)
- Voz no sonora: similar al ruido blanco

Objetivos de los códecs de voz:

- Mantener la inteligibilidad
- Permitir la identificación del hablante

Técnicas de compresión de voz

- Por forma de onda: baja complejidad, calidad media, 16–64 kbps (ej. PCM, ADPCM)
- Paramétrica (vocoders): alta compresión, menor calidad (ej. LPC: 1,2–4,8 kbps)
- Híbrida: combinación de ambas (ej. CELP, G.729, AMR, SILK)

Ejemplos destacados de códecs de voz

- SILK: de Skype, no libre
- iSAC: banda ancha, libre
- iLBC: banda estrecha, libre, usado en WebRTC
- AMR / AMR-NB / AMR-WB: adaptativos para redes móviles
- Speex: libre

G.711

Es el códec de referencia para voz:

- PCM de frecuencias vocales: $8 \text{ kHz} \times 8 \text{ bits} = 64 \text{ kbps}$
- Cuantificación no lineal (μ -law en Norteamérica y Japón, A-law en Europa)

Archivos y formatos de audio

Los archivos de audio pueden tener:

- Canales: mono, estéreo, multicanal
- Compresión:
 - Sin compresión: PCM puro
 - Sin pérdida: FLAC, Apple Lossless, WMA Lossless
 - Con pérdida: MP3, AAC, Vorbis

Contenedores comunes:

- WAV: basado en RIFF, contiene chunks con metadatos y puede incluir audio comprimido o sin comprimir
- Ogg, MP4: usados también para vídeo

Formatos físicos

- CD de audio:
 - Referencia: “Libro rojo”
 - PCM lineal, 44.1 kHz, 16 bits, estéreo
- Otros formatos:
 - SACD: usa modulación por densidad de pulsos
 - DVD-Audio, Blu-ray: usan codificación sin pérdida como Meridian Lossless Packing

Conectores de audio

- Solo audio:
 - Analógico: conectores de PC por colores
 - Digital: S/PDIF, admite PCM sin comprimir y sonido envolvente comprimido
- Audio y vídeo:
 - Analógico: SCART
 - Digital: HDMI