Sound Formats *For Everyone*

David Thall
Insomniac Games

What is a sound file?

- A digital data format for the storage, interchange, and retrieval of sound data.
  - NOTE: This does not include playback...
    - Drivers and Applications are responsible for this!

So then what is sound data?

- A digital representation of a signal.
  - Most sound data is treated as a discrete function in 1 dimension (e.g., x(t)). Think Trigonometric Functions like cosine()/sine().
  - Can be represented in the time domain or frequency domain.
  - Usually stored as a succession of scaled and shifted ‘samples’ in the time domain (PCM).
  - Samples are quantized on the x-axis (time) and y-axis (value).

And how is that data organized?

- The sound data is organized alongside a header (unless the data is ‘raw’)
- Example Header Data might include:
  - Format: PCM, A-DPCM, MP3…
  - Data Type: Signed/Unsigned, Integer/Float
  - Bits / sample (depth), samples / frame (channels)
    - (OR) Bits / second (bit rate)
  - Size (excluded from MP3)
  - Frequency (included with PCM, A-DPCM)

PCM
Pulse Code Modulation

- Encodes a signal as a sequence of uniformly-spaced samples (fixed bits).
- Signal is ‘sampled’ on the x-axis
- Signal is ‘quantized’ on the y-axis
- Examples
  - Raw waveform data in AIFF, WAV, CDDA
  - Source waveform data in AC-3 encoding
  - Target waveform data in MP3 decoding

PCM Parameters

- Bits per sample (Bits)
  - For each bit, the number of possible quantization steps doubles. Thus, each significant bit doubles the dynamic range.
    - Huh? For each additional bit, you can manage twice the amount of input signal variation. This equates to 6 decibels of added dynamic range, independent of whether or not the quantization is linear or non-linear.
    - The area within the quantization step is noise.
    - bps = x bits/sample * y samples/second * z channels
    - kbps = bps / 1000

PCM Parameters

- Frequency (Sample Rate)
  - According to the Nyquist Theorum, a signal sampled at frequency ‘f’ can manage signals in the baseband [-f/2, +f/2], where f/2 is the Nyquist Frequency.
  - Frequencies within this range are indistinguishable from their aliases at multiples of the sampling frequency. Attempting to manage signals outside this range will result in audible aliasing and foldback.
- We can trade off bit depth for sample rate
  - Example: DSD (Delta Sigma Modulation, Direct Stream Digital) - SACD
  - Example: MLP (Packed PCM) – Dolby TrueHD (BD)
    - Only store differences on additional channels with less bits

DPCM / A-DPCM Adaptive-Differential Pulse Code Modulation

- Encodes a signal as a difference between each real sample and a prediction sample.
- Prediction sample is generated from the previous $x[n - c]$ decoded samples.
- Adapts the size of the quantization step according to current transient response (rate of change). Adds quantization noise!
- Example: VAG (4:1 compression ratio)

MP3
MPEG-1 Audio Layer 3

- Specification for decoding PCM audio from a perceptually-coded reduced data set.
  - Forward masking, JNDs, noise-shaping, joint-stereo
  - Encoding Practices: Sine+Noise, LPC-Resynthesis, Exciter <-> Resonator Feedback Delay Networks
  - Encoding generally involves frequency bin separation, analysis and mask-based filtering. Resulting bit rate is targeted by adaptive quantization. This does not affect frequency response!
  - e.g., LAME encoder/decoder
- Lossy (in CBR or VBR), just like DTS/AC-3

Why Compress?

- Size
  - As our games get larger, so do the sound asset requirements.
  - Memory is still limited.
- Channels (8++)
  - More and more people are purchasing multi-channel and surround decoding home theater systems. Simultaneously, game hardware continues to support these new formats. We will need to do the same!

Q/A

- How do we improve our compression codecs?
  - Larger compression ratios
  - Faster decoders
  - Less audible artifacts