Digital Audio Compression, Psychoacoustics and You

(but mostly Digital Audio Conversion and Psychoacoustics)

Martin McCrory

8 November 2005

What is Digital Audio Compression?

- reduction in size of an audio signal (over 1,400,000 bits of information per second in CD quality audio) to save digital space
- accomplished by eliminating certain pieces of information from the sound file that we can't hear or don't care about too much

How much space do we save?

- depends on the "bit rate" at which we choose to compress, and the encoder itself
- higher "bit rate" = more bits of information per unit time of data
  - example: 128kbps (de facto standard of audio compresion these days) = 128,000 bits of information per second, or about 1/11 the size of CD quality audio (sweet!)

How do we figure out what portions of the sound file to get rid of (or compress "aggressively")?

- eliminate sounds beyond a certain high and low bound (varies depending on desired file quality and content of file itself, often correlates to hearing range of humans)
- aggressively compress sounds that are masked
  - Simultaneous masking: when one sound, playing concurrently with another sound, interferes with the second sound's ability to be heard. often the masked sound will be compressed to a greater extent than the first, more important (and more prominent) sound)
    - example: hand clapping in a silent room: easy to hear. hand clapping on a busy street with lots of cars and people yelling and such: harder to hear.
  - Temporal masking: when a sound makes inaudible other sounds which are present immediately before or after the original sound.
    - complex harmonic tones tend to temporally mask tones centered around the first several harmonics
    - don't confuse with the ear's acoustic reflex (to protect from very loud sounds)!

What else can we do to reduce size without sacrificing too much of what we hear?

- Pulse-Code Modulation
  - This process functions somewhat like our sampling process in SPEAR
  - by "quantizing" the sound wave (like changing the "window size", sampling the

sound wave regularly at uniform intervals and creating a composite wave out of those "snapshots" of the sound wave), we discard portions of the sound wave that are less important to human hearing (but still leaving the general idea of the sound wave intact)

- quality of compressed sound depends greatly on the quantization size (goes along with "bit rate") and the type of sound being compressed
  - random sounds such as applause are difficult to compress because the quantization process will always remove "vital" portions of sound information
  - a more predictable sound sample (such as diatonic chords) can compress certain portions of the file aggressively and leave other portions of the file (the main harmonics of the chords) less compressed

Audio compression today!

- MP3 file format
  - Encoding audio in MP3 format results in a sound file that is compressed often to 1/12th the size of CD audio at a small, acceptable loss of quality
  - 128kbps compression is acceptable for most listeners, and 192kbps compression rivals CD quality for some listeners
- Ogg Vorbis file format
  - Much better than MP3 file format
  - Open source
  - listening tests have shows that higher compression Vorbis files correspond to lower compression MP3 files in terms of overall quality (128kbps Vorbis file compares to 192kbps mp3 file) WITHOUT sacrificing file size!
  - Don't use MP3.
  - Vorbis is better.

P.S. Ogg Vorbis