Ogg Vorbis

Masters's thesis in
Information- and Communication Technology

**Candidate:**
Hans Petter Sirevaag Selasky

**Title:**
Evaluation of perceptual sound compression with regard to perceived quality and compression methods

**Supervisor:**
Ola Torkild Aas, AUC

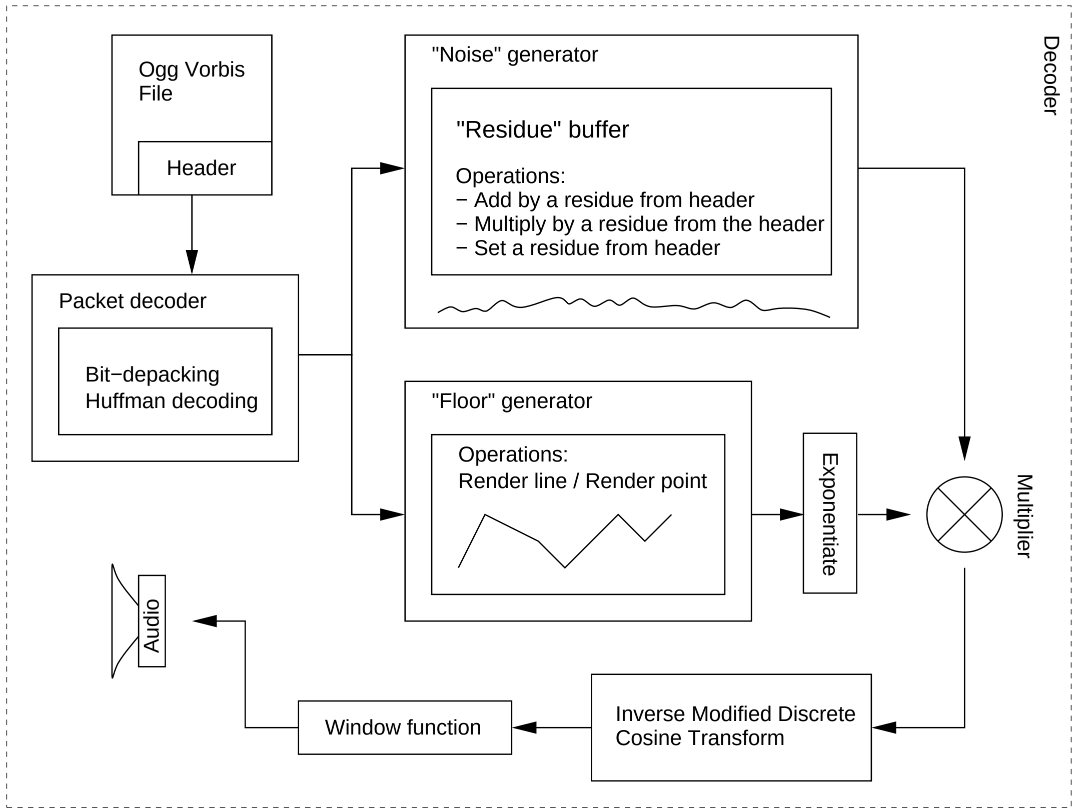
In the area of sound compression there exist several different codecs. In general there are two categories of sound codecs. There are so called "lossless sound codecs" that compress sound like data that must not be altered. Some examples are Flac, WavPack, Shorten, Monkey's Audio, Apple Lossless, Ogg Squish, Bonk, La, OptimFROG, LPAC, RKAU and many more. The typical compression ratio is 0.75 .. 0.50, relative to 1.0. The other general category of sound codecs is "lossy codecs", which will alter the sound. Some examples here are: Ogg Vorbis, mp3, mp4, WMA, QuickTime and AAC. The compression ratio of these sound codecs depends on the quality setting applied, but is typically 0.10 .. 0.04 relative to 1.0.

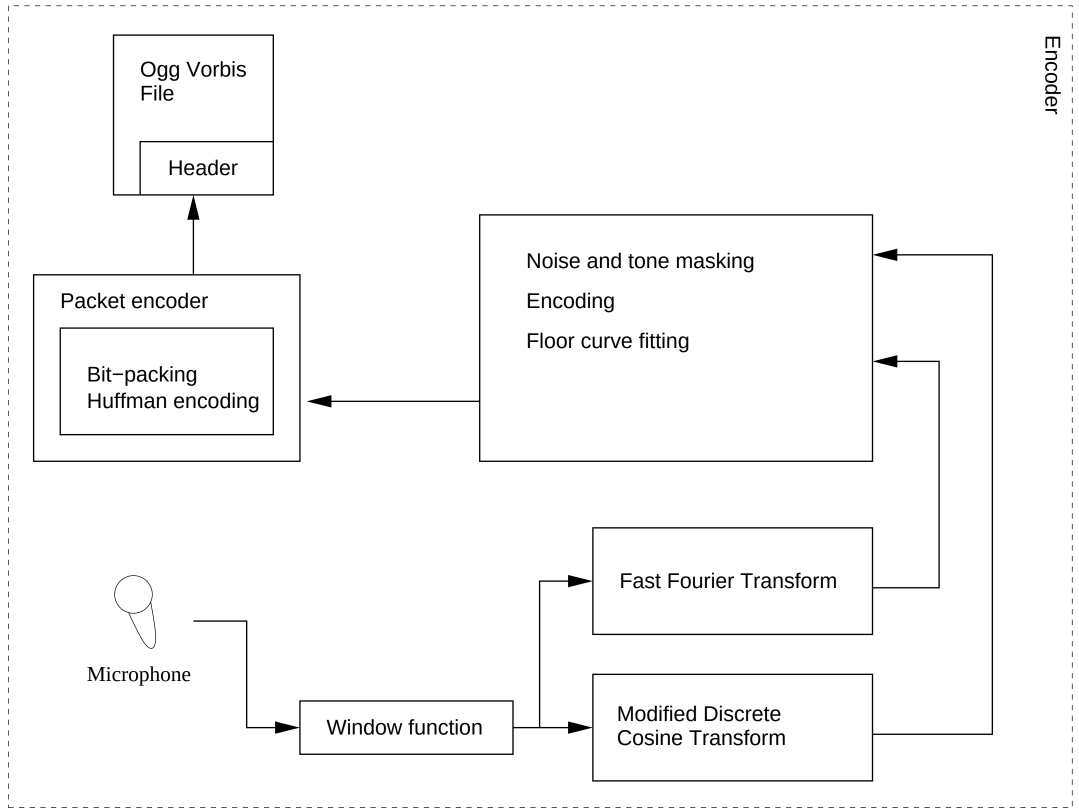
When Ogg Vorbis decodes sound, two parts are needed. These are called "residue" and "floor" curves. The "residue" is the fine detail of the audio cosine-spectrum that is left after that the "floor" has been divided out. The "floor" consists of a series of points. Between each point a line is drawn, using a function called "render_line()". The "floor" curve is stored in deciBel, dB, and is exponentiated before it is multiplied by the "residue". Then the result is passed on to an "Inverse Discrete Cosine Transform" and a "Window function", before it ends up in the speaker. The Ogg Vorbis synthesizer process is not very complicated. The "residue" buffer is composed of multiple "residues" that are transmitted in the Ogg Vorbis header. Composition means using addition and multiplication of the "residues" available. To decode a so called "audio frame", the decoder is passed a number that selects "residue" and a number that selects whether the "residue" should replace, add, or multiply the current "residue" buffer. Also a set of points are transferred to make up the "floor" curve. All values are transferred using Huffman coding.

When Ogg Vorbis encodes sound, the sound is first passed through a window function. Then the sound is analyzed in two ways. First using a Fast Fourier Transform. This transform reveals information about which sine-wave-frequencies that are strong, and is used to perform so called "tone masking". Second a modified discrete cosine transform is performed. This transform reveals information about the noise in the signal, and is used for so called "noise masking". The results from the transforms are combined to generate the "residue" and "floor" curves. Then all data is packed using bit-packing and Huffman encoding. In the end an Ogg Vorbis file is produced. This is the encoding process in a nutshell. There are a lot of more details and complicated formulas used in the encoding process, which I will not go into any details.

--HPS

http://www.vorbis.com




Grimstad, June 2006