An Overview of Ogg Vorbis

Presented by: Shi Yong

MUMT 611

Overview

- Ogg
  - A multimedia container format, contains audio and video data streams in a single file, similar to mov and avi.
  - Functions: framing, sync, error correction, positioning
  - Stream oriented, suitable for internet streaming.
  - Used with various Codecs:
    - Vorbis: Audio codec
    - Tremor: Fixed-point decoder
    - Theora: Video codec
    - FLAC: Free Lossless Audio Codec
    - Speex: Speech codec
    - OggWrit: Text phrase codec (subtitles)
    - Ogg Metadata: Arbitrary metadata format

Overview

- Vorbis
  - General purpose lossy audio compression algorithm/format, similar to MP3, WMA, etc..
  - Designed to be contained in a transport mechanism that provides framing, sync and error correction functions, such as Ogg (for file transport) or RTP (for webcast)
  - When used with Ogg, it is called Ogg Vorbis.
  - A great number of player now support Ogg Vorbis
  - Encode CD/DAT quality at below 48kbps
  - Wide range of sample rates: from 8kHz to 192kHz
  - Strong channel representations: monaural, polyphonic, stereo, quadraphonic, 5.1, up to 255 discrete channels

Overview

- Both Ogg and Vorbis are developed by the Xiph.Org Foundation, a non-profit corporation (http://www.xiph.org/)
- LGPL licensing:
  - open standard
  - open source
  - patent free
  - completely free for commercial or noncommercial use
- Software can be downloaded from http://www.vorbis.com/setup/

Encode

- Overview
  - Accepting input audio, dividing it into frames, compressing frames into raw packets
  - As a generic perceptual audio encoder, psycho acoustic model is used to remove redundant audio information
  - Function blocks is similar to most other lossy audio encoders, such as Time/Frequency Analysis, Psychoacoustic Analysis, Quantization and Encoding, Bit Allocation, Entropy Coding, etc.
- Codebook
  - Codebooks are defined and transmitted as a part of the audio stream
  - They must be sent before the audio packets, which typically cause a delay of 1s at 128kbps data rate

Encode

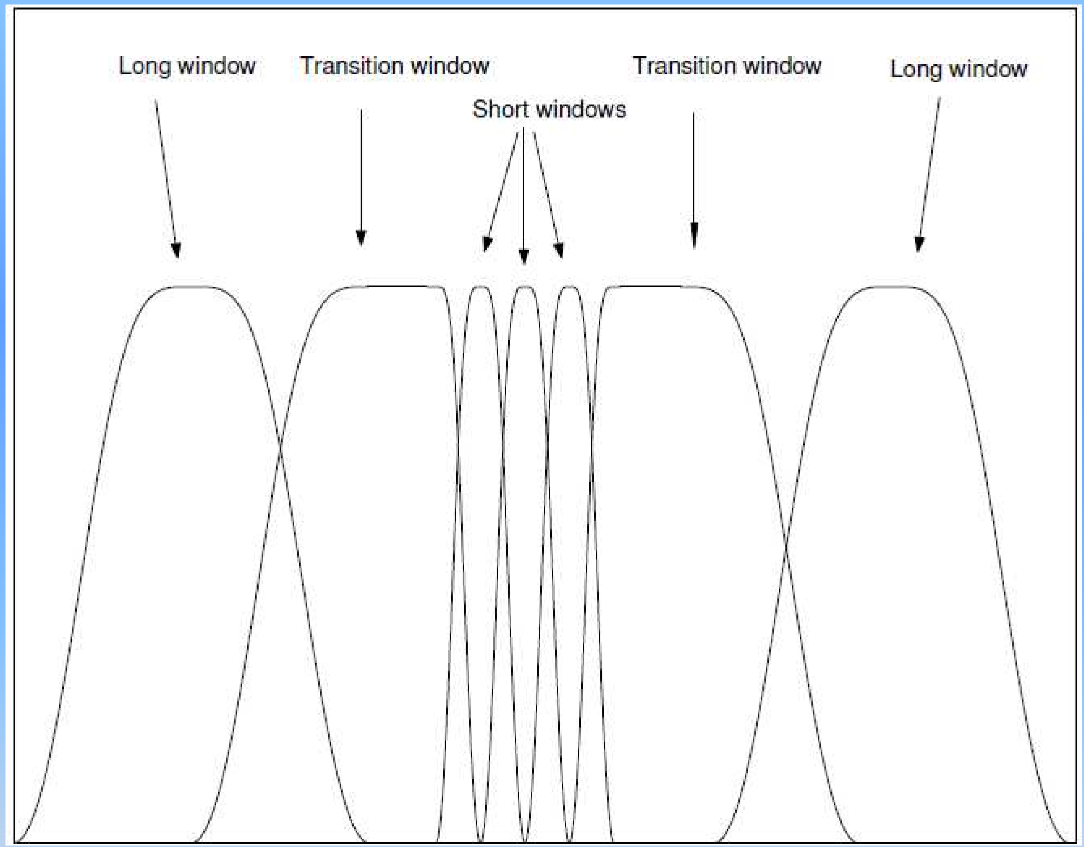
- Time-Frequency transform
  - MDCT (Modified Discrete Cosine Transform) combined with overlapping and windowing is used for this purpose.
  - Linear orthogonal lapped transform, based on the idea of time domain aliasing cancellation (TDAC)

$$X(m) = \sum_{k=0}^{n-1} f(k)x(k)\cos\left(\frac{\pi}{2n}\left(2k+1+\frac{n}{2}\right)(2m+1)\right), \text{ for } m = 0..\frac{n}{2}-1$$

$$y(p) = f(p)\frac{4}{n}\sum_{m=0}^{\frac{n}{2}-1} X(m)\cos\left(\frac{\pi}{2n}\left(2p+1+\frac{n}{2}\right)(2m+1)\right), \text{ for } p = 0..n-1$$

Encode

- Windows
  - Vorbis uses two windows: short and long.
  - The length of windows must be power of 2 (between 64 and 8192)
  - Short window can achieve better time resolution (fit for plosive sounds), while long window can achieve better spectral resolution
  - typical long, short and transition windows (cited from Erik's Master thesis)



Decode

- Vorbis format is defined by its decode specification
- Vorbis I specification can be found online: http://www.xiph.org/vorbis/doc/Vorbis_I_spec.pdf
- The design for embedded system can be "deviated" a little bit.
- A number of "component abstractions" perform specific functions in the decode pipeline
  - Blocksizes, modes, mappings, floors, codebooks, residues

Decode

- Decode
  - Accepting packets in sequence, decoding them, reconstructing spectrum data, synthesizing and reassembling audio frames
  - Flow chart:

    Audio packet
    $\Rightarrow$ Header decode
    $\Rightarrow$ Floor reconstruction
    $\Rightarrow$ Residue unpacking
    $\Rightarrow$ Channel coupling
    $\Rightarrow$ IMDCT
    $\Rightarrow$ Windowing
    PCM samples

Configuration

- Global configuration
- Mode
  - specify the encode method to a frame
- Mapping
  - Channel coupling description
  - A list of submaps
- Floor
  - Low resolution representation of the spectrum of a frame
  - Floor 0: packed LSP (Line Spectral Pair)
  - Floor 1: piecewise linear interpolated representation
  - Make use of entropy coding to save space
- Residue
  - Fine structure of the spectrum of a frame
  - Can be one of three packing/coding algorithms (number 0 to 2)
  - Also make use of entropy coding

Setup

- Decode Setup
  - Setup by using three header packets
    - Identification Header
    - Comment Header
      - Various field names
    - Setup Header
      - Configuration information
      - VQ and Huffman codebooks
  - Then followed by all audio packets (decoding and synthesis)

Decoding and Synthesis

- Decode procedure in detail (the same for all audio packets)
  1. Packet type flag
  2. Mode number
  3. Window shape
  4. Floor vectors
  5. Residue vectors
  6. Inverse channel coupling
  7. Generate floor curve
  8. Combine with residue (fine structure)
  9. Inverse transform of spectrum vector
  10. Overlap/add
  11. Store right hand-data of current frame for future lapping
  12. Get the result (the PCM audio data)

Floor Decode

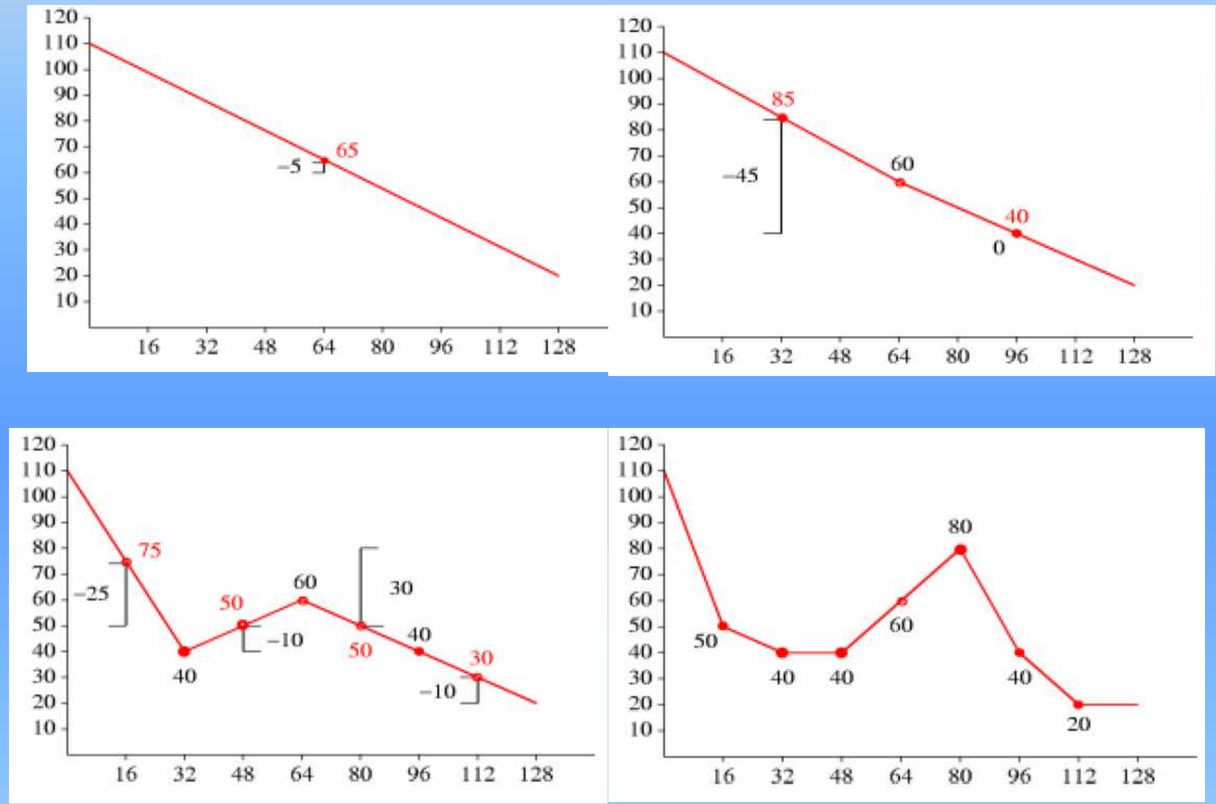
- Floor type 0:
  - Spectral envelope curve is encoded as the frequency response of the LSP (Line Spectral Pair) filter
  - Head decode
    - Configuration information
      - Six integer fields
  - Packet decode
    - Curve amplitude and LSP coefficients
    - Generate floor curve

Floor Decode

- Floor type 1
  - Spectral envelope curve is encoded as a piecewise straight-line.
  - Head decode
  - Packet decode
    - Amplitude value
    - Spectral curve synthesis

Curve Synthesis Example

- X = [0, 128, 64, 32, 96, 16, 48, 80, 112]
- Y = [110, 20, -5, -45, 0, -25, -10, 30 ,-10]



This example is cited from Vorbis I Specification by Xiph.org Foundation.