Control Data Flow Graphs
- MP3 Decoder

Sue Tyerman

School of Computer Science
University of Adelaide

Introduction

• Motion Picture Expert Group
• Work on a number of standards
• ISO standard 11172-3 MPEG-1 Layer III
• Perceptual codec for compression of CD quality music
  ■ Uses knowledge of human auditory perception to assist in compression
• Common compression rate is from 1.4 Mbits/s down to 128 kbits/s
• Best used for bit rates of 64 kbits per channel
MP3 Decoder

Bitstream → Synchronisation and Error Checking → Huffman Code Bits, Huffman Information, Scalefactor Information → Huffman Decoding, Huffman Information Decoding, Scalefactor Decoding → Magnitude and sign → Requantisation, Reordering → Discrete Cosine Transformation → Joint Stereo Decoding → Alias Reduction, IMDCT → Inverse Modified Discrete Cosine Transformation → Frequency Inversion → Synthesis Polyphase Filterbank → left Pulse Code Modulated Signal, right PCM Signal

Rest of talk

- Some crucial elements behind MP3 coding/decoding
- The frame
Minimal audition threshold

- Hearing threshold of the ear is not linear
- Range between 20 Hz and 20 kHz, best between 2 kHz and 4 kHz, with normal speech between 500 Hz and 2 kHz
- Not necessary to code sounds situated outside this threshold

The masking effect

- MP3 encoder uses a psychoacoustic model modeling the behavior of the human ear
  - Acts like a bandpass filter with 26 strongly overlapped bands used to cover audible range (up to 24 kHz)
- Simultaneous masking: During strong sounds, you do not hear the weakest sounds. Strong/loud sounds can mask the quieter sounds.
- Temporal masking: Loud sounds can also mask softer sounds that occur immediately after the louder sound
- Don’t encode what you can’t hear
The bytes reservoir

- Frames can be different sizes (maximum of 2kBits)
- Some applications require a fixed bit rate while others do not
- Some passages cannot be compressed into one frame and maintain required quality
- Borrow from other frames that don’t use all their capacity
  - Repercussion of this is that a number (up to 9) frames may need to be decoded to decode a piece of music
  - Such frames occur in the bitstream earlier than the frame side information that references them.

The Joint Stereo coding

- Stereophonic signals have spatial information and can be compressed further using Intensity Stereo
  - Some frequencies are recorded as a monophonic signal followed by additional information in order to restore a minimum of spatialisation
- When left and the right channels are similar, middle (L+R) and side (L-R) channels are encoded instead of left and right
  - Allows reduction of the final file size by using less bits for the side channel
Huffman Coding

- At the end of the compression used to code information
- Creates variable length codes
- Higher probability symbols have shorter codes
- Have a unique prefix
  - Can therefore be decoded correctly in spite of their variable length
- Decoding step is very fast (via a lookup table)
- Save on average a bit less than 20% of space.

Structure of encoded data

- A musical piece is broken into a series of frames
- Each frame is independently analysed and codified with necessary decode information
- Each frame has a header with all necessary information to decode that frame
- Length of music coded into frames varies from 2 to 50 ms
- The data for each frame has 1152 samples broken down into two granules of 576 samples each
### MP3 Frame

<table>
<thead>
<tr>
<th>Header (4 Bytes)</th>
<th>Side Information (Single channel = 17 Bytes Dual channel = 32 Bytes)</th>
<th>Main Data</th>
<th>Ancillary Data</th>
</tr>
</thead>
</table>

#### Diagram:

- **Header Frame 1**
  - Sync Side info 1
  - Main 1
  - Main data 1

- **Header Frame 2**
  - Sync Side info 2
  - Main 2
  - Main data 2

- **Header Frame 3**
  - Sync Side info 3
  - Main 3
  - Main data 3

- **Header Frame 4**
  - Sync Side info 4
  - Main 4
  - Main data 4
Frame Structure - Roles

- Header
  - Information on layer, bitrate, sampling frequency and stereo mode
  - Sync word (12 bits long) to find the start of the frame in the bitstream
- Side Information
  - Information to decode the main data: Huffman table selection, scale factors, requantisation parameters, window selection

Frame Structure - Roles

- Main Data
  - Contains coded scale factor values and Huffman coded frequency lines
  - Length depends on window length, bitrate, scale factors and length of ancillary data
- Ancillary Information
  - User defined data, catered for but ignored in the standard
Interesting problem

- Modeling MP3 decoding has interesting problems
  - Bitrate, Sample rate, Framesize can vary
  - Main data can be shared between frames
  - Mix of multiple lookup tables and computation
  - Stereo vs mono signal
  - Window lengths vary (short or long), has an impact on storage, lookup and computation
    - Reordering of data to make use of lookup tables more efficient when short windows are used
  - Windows are overlapped with each other
  - Scaling factors at global, frame and subblock levels
- And all of these can vary frame to frame...

Initial CDFG Model
One interesting CDFG problem

- Atomicity
  - Realtime application
  - Outputs music as generated, according to calculated rates
  - Musical piece composed of multiple frames
  - One frame may use space in preceding frames
  - Don’t know the size of the main data till the side information of a frame is processed
  - Don’t know the size of the side information till the header of a frame is processed

What next?

- Fully come to grips with decoder
  - What, where and how is data used and transformed?
  - Lots of missing bits in initial model
- Model using CDFG
  - Hand
  - Design/CPN
  - What works?
  - What problems arise?
Some References

• Background:
  www.iis.fraunhofer.de/amm/techinf/layer3/

• The MPEG Home Page: www.chiariglione.org/mpeg/

• Master’s thesis and reference code:
  www.kmlager.com/thesis.php

• Background, useful links, reference code, etc:
  www.mp3-tech.org/