MPEG Audio Compression Layer 3 (MP3)

Hai Tao
History

- MP3 originally developed by Fraunhofer IIS partly under EUREKA project EU147, Digital Audio Broadcasting (DAB). Then adopted as ISO-MPEG Audio Layer-3 (1991, There are Layer 1 and 2 also)
- For 44.1 KHz sampling, stereo 16 bit audio, the bit rate is 44100x2x16=1.41Mbit/s, Using MP3, good quality sound can be achieved even with >1:24 compression ration
- Layer 3 is the most advanced in all three methods

<table>
<thead>
<tr>
<th>Compression Ratio</th>
<th>By Layer</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:4</td>
<td>1:4</td>
<td>by Layer 1 (corresponds to 384 kbps for a stereo signal),</td>
</tr>
<tr>
<td>1:6...1:8</td>
<td>1:6...1:8</td>
<td>by Layer 2 (corresponds to 256..192 kbps for a stereo signal),</td>
</tr>
<tr>
<td>1:10...1:12</td>
<td>1:10...1:12</td>
<td>by Layer 3 (corresponds to 128..112 kbps for a stereo signal),</td>
</tr>
</tbody>
</table>
## Performance

### Typical performance for various sources

<table>
<thead>
<tr>
<th>sound quality</th>
<th>bandwidth</th>
<th>mode</th>
<th>bitrate</th>
<th>reduction ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>telephone sound</td>
<td>2.5 kHz</td>
<td>mono</td>
<td>8 kbps *</td>
<td>96:1</td>
</tr>
<tr>
<td>better than short wave</td>
<td>4.5 kHz</td>
<td>mono</td>
<td>16 kbps</td>
<td>48:1</td>
</tr>
<tr>
<td>better than AM radio</td>
<td>7.5 kHz</td>
<td>mono</td>
<td>32 kbps</td>
<td>24:1</td>
</tr>
<tr>
<td>similar to FM radio</td>
<td>11 kHz</td>
<td>stereo</td>
<td>56...64 kbps</td>
<td>26...24:1</td>
</tr>
<tr>
<td>near-CD</td>
<td>15 kHz</td>
<td>stereo</td>
<td>96 kbps</td>
<td>16:1</td>
</tr>
<tr>
<td>CD</td>
<td>&gt;15 kHz</td>
<td>stereo</td>
<td>112..128 kbps</td>
<td>14..12:1</td>
</tr>
</tbody>
</table>
Basic Diagram
Psycho-Acoustic Perceptual Model

- The basic perceptual model used in MP3 is that louder frequencies mask out adjacent quieter ones. People cannot hear a quiet sound at one frequency if there is a loud sound at another.
- This can be explained better by the following figures presented by Rapha Depke.
Audio perception

- Human perception of audio signal

![Diagram showing speech, music, and pain limits in the frequency-sound pressure plane.](image-url)
Perceptual model

- People can not hear a quiet sound at one frequency if there is a loud sound at another.
Perceptual model

- The louder the sound, the larger the mask area

- The bottom area is called absolute threshold of hearing (ATH), defining the weakest sound that can be heard
Filter Banks

- The audio signal passes through 32 filters with different frequency
Dominant band and the mask

- Dominant band is found and the corresponding mask is applied
Quantization of Audible Sound

- The components exceed the mask are quantized and encoded using the Huffman coding method
Joint Stereo

- Joint stereo coding takes advantage of the fact that both channels of a stereo channel pair contain similar information.
- These stereophonic irrelevancies and redundancies are exploited to reduce the total bitrate.
- Joint stereo is used in cases where only low bitrates are available but stereo signals are desired.
Encoder

- A typical solution has two nested iteration loops
  - Distortion/Noise control loop (outer loop)
  - Rate control loop (inner loop)
Rate control loop

For a given bit rate allocation, adjust the quantization steps to achieve the bit rate.

- This loop checks if the number of bits resulting from the coding operation exceeds the number of bits available to code a given block of data.
- If it is true, then the quantization step is increased to reduce the total bits. This can be achieved by adjusting the global gain.
Distortion control loop

This loop shape the quantization steps according to the perceptual mask threshold

- Start with a default factor 1.0 for every band
- If the quantization error in a band exceeds the mask threshold, the scale factor is adjusted to reduce this quantization error
- This will cause more bits and the rate control loop has to be invoked every time the scale factors are changed
- The distortion control is executed until the noise level is below the perceptual mask for every band
Decoder

- Decoder side relatively easier. The gain, scale factor, quantization steps recovered and used for reconstruct the filter bank responses
- Filter bank responses are combined to reconstruct the decoded audio signal