Multimedia Data and Its Encoding

M. Adnan Quaium

Assistant Professor
Department of Electrical and Electronic Engineering
Ahsanullah University of Science and Technology
Room – 4A07
Email – adnan.eee@aust.edu
URL- http://adnan.quaium.com/aust/cse4295
Tones and sounds are analog signals, the signals themselves are both time and value continuous and therefore cannot initially be displayed on a computer directly. This requires a conversion to time and discrete value, i.e., digital signals.

The conversion from an analog output signal into a digital, computer-displayable signal proceeds in three steps:
1. Sampling,
2. Quantization (rounding) and
3. Coding.
Analog-to-Digital Conversion

**Fig. 4.22** Analog-digital conversion.

\[ f(t) \quad \xrightarrow{\text{Sampling}} \quad f([t]) \quad \xrightarrow{\text{Analog-to-digital conversion}} \quad \left[ f([t]) \right] \]

- Continuous analog input signal
- Discontinuous analog signal
- Discontinuous discrete output signal

**Fig. 4.23** Step-by-step process of digitalization.
Audio compression procedures carry out the same processing steps as the human ear.

- The signal is first broken down with respect to frequency.
- Based on anatomical conditions, the human ear can only perceive acoustic signals with a sound pressure of between 0 dB and 120 dB and with a frequency range of 20 Hz to 20,000 Hz. This is the so-called auditory sensation area.
- Acoustic signals outside of the human auditory sensation area must not be encoded (irrelevance reduction).
Predictive Coding

- Knowledge about the previously sent or encoded signal is used to make a prediction about the signal to follow.
- The actual compression is made by only saving the difference between the signal and its prediction.
- This is smaller than the original signal and thus can be encoded more efficiently.

Spectral or Transform Coding

- Via the waveform of the signal a Fourier transformation is performed that transforms the signal in the frequency domain.
- Since the transformed representation of the signal changes more slowly, it is only necessary for a few samples to be transmitted.
- Transform encoders normally use a large number of sub-bands and view adjacent samples together in terms of frequency.
Sub-Band Coding

- The available audio spectrum is divided into individual frequency bands.
- In encoding it is used to an advantage that almost all of the frequency bands have much less information (or less important information) than the loudest band.
- Consequently, in compression the important bands are given more space than the unimportant tones, which in some cases may even be left out completely.
- The elaborate work of selection as to how many bits are assigned to which sub-band, is done by the encoder based on a so-called psycho-acoustic model.
- In addition to the actual audio data, the information must also be transmitted via the bit distributor. Sub-band coding is often understood as a special type of transform coding.
MPEG stands for Motion Picture Experts Group, a committee that was originally involved in the coding and compression of video and audio data.

The bit rate that should be transmitted with the signal in MPEG is taken as constant. This means that signals need not necessarily be displayed in the smallest possible space, but that the signal display can optimally utilize the given bandwidth.

The hardware implementation of a MPEG encoder, the configuration of the filter banks (including, e.g., the Fourier or cosine transformation, as well as information on the use of the psychoacoustic model*) is firmly defined in the standard and does not depend on the signal to be encoded. Nevertheless, this approach does not guarantee optimal efficiency and quality from the onset.

* The frequencies and sound signals that cannot be detected by the human ear are filtered out and not saved.
Fig. 4.36 Schematic flow of the MPEG audio encoding/decoding.
The MPEG-1 specification for audio encoding is divided into so-called layers. This enables the compression of a stereo signal in sampling rates of

<table>
<thead>
<tr>
<th>MPEG-1</th>
<th>Target bit rate</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Layer 1</td>
<td>192 kbps per channel</td>
<td>digital compact cassettes (DCC)</td>
</tr>
<tr>
<td>Layer 2</td>
<td>128 kbps per channel</td>
<td>digital radio (DAB), digital television (DVB) Audio to video CDs</td>
</tr>
<tr>
<td>Layer 3</td>
<td>variable bit rates 32-384 kbps</td>
<td>Internet, Audio-MP3</td>
</tr>
</tbody>
</table>

- Encoders for each MPEG layers are backwards compatible, i.e., layer 1 is the basis that all encoders and decoders (also designated as Codec) must provide.

- Decoders for layer 2 must automatically be able to implement layer 1 data, but not vice versa. The well-known MP3 files are encoded according to the procedure established in MPEG-1 layer 3.
• Via the filter bank – the input audio signal is divided into 32 equal-width, each is 750 Hz wide with a sampling rate of 48kHz.
• The filter bank takes an input sample (sample value) and breaks it down into its spectral components, which are each distributed over 32 sub-bands.
• An MPEG layer 1 data packet includes 384 samples, whereby 12 samples are grouped into each of the 32 sub-bands.
• Based on the psychoacoustic model implemented, the encoder allocates the number of necessary bits to each sample group.

• The filter bank and synthesis are lossy here, which is however not audible.
• Additionally, so-called aliasing effect occur. This means that significant overlapping takes place in adjacent frequency bands in the MPEG layer 1, because the frequency bands are not sharply defined.
Four different modes are defined in channel coding:

- **Single Channel coding** for the encoding of mono signals,
- **Dual Channel coding** for the encoding of two separate mono signals, e.g., bilingual audio,
- **Stereo Coding** for the encoding of a stereo signal, in which both stereo channels are however encoded separately, and
- **Joint Stereo Coding** for encoding a stereo signal, in which the data irrelevancies and redundancies between the two channels are used for compression.
MPEG-1 Layer 2

- An MPEG-1 layer 2 data packet contains 1,152 samples per channel and blocks of 3 each made up of 12 samples per subband (subframes).
- This results in an increased time resolution of the signal.
- Within a frame, the post-masking effect (temporal overlap) of human hearing can also be exploited.

- At the start of encoding there is an uncompressed audio signal sampled with a 48 kHz sampling rate and quantized with a 16 bit sampling depth.

- MPEG-1 layer 2 encodes the data in larger groups and limits bit allocation in medium and high sub-bands, since they are not as important for auditory perception.
- In this way, bit allocation data, scaling factors and quantized samples can be saved in a more compact form. This extends the available space for the important audio data.
MP3 used an additional *Modified Discrete Cosine Transformation (MDCT)* on the output of the original filter bank, thereby bringing about a drastic rise in resolution from 32 to a maximum of 576 sub-bands.
Among the improvements introduced in MP3 are:

- A compensation for the filter bank aliasing effect
- The Quantizer potentiates the input to 3/4 in order to distribute the signal-to-noise ratio more evenly over the value range of the quantization interval.
- Scaling factors are summarized in MP3 in bands. One band includes multiple MDCT coefficients and has a width similar to that of the human ear. In this way, quantization noise is covered and there are no more noise peak values.
- A bit reservoir is introduced, which can be used to increase the resolution. The encoder is only allowed to take bits from the reservoir that it had previously saved and stored there.
MPEG-1 Layer 3

Josie - direct from CD

After MP3 encode (160 kbps) and decode

Residual (after aligning 1148 sample delay)
In lossy MP3 compression, the choice of unsuitable basic parameters – such as the selection of a too low bit rate – result in disturbances and distortion in the form of artifacts.

These type of artifacts are distinguished from those that normally occur in the analog world for example in broadcasting.

A distinction is made between

- **distorted artifacts** (but no harmonic distortion),
- **noise artifacts** (with noise affecting only a certain frequency range) and
- **interference artifacts** (interference in relation to the audio signal is very significant as the characteristics of the interference signal can change every 24 milliseconds).
### MPEG-1 Data Compression Rate

<table>
<thead>
<tr>
<th>MPEG Layer</th>
<th>Compression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Layer 1</td>
<td>1:4</td>
</tr>
<tr>
<td>Layer 2</td>
<td>1:6, ..., 1:8</td>
</tr>
<tr>
<td>Layer 3</td>
<td>1:10, ..., 1:12</td>
</tr>
</tbody>
</table>
MPEG-2 is an expansion of the MPEG-1 standard.

It includes the following addition to MPEG-1:

- the additional sampling rates of 8 kHz, 11 kHz, 16 kHz, 22.5 kHz and 24 kHz.
- 3 extra audio channels that allow a 5 channel surround sound (left, middle, right and 2 spatial channels).
- support of a separate audio channel for low sound frequency effects (<100 Hz).
- support of extra audio information in different configurations (up to 8 multi-lingual channels). These can offer both multiple language transmission support for the hearing and visually impaired.
- use of variable bit rates for compression adjustment to the changing complexity of the audio information to be encoded.
To ensure compatibility, MPEG-2 data must be encoded in such a way that a conventional MPEG-1 decoder is capable of filtering out the left and right stereo channels from the five possible channels. The fixed data format in MPEG-1 is used for the compression of the left and right channel.