ECE431: Mini-Lecture on MPEG Audio Compression

R. Mark Grover and Alp Kucukelbir

November 6, 2008

1 Some History of MPEG

MPEG stands for the *Moving Picture Expert Group*. The group was formed in 1988. They created standards, which are simply specifications without any architectural design.

MPEG-1: Video CD and MP3 Audio

MPEG-2: Digital TV

MPEG-3: HDTV (Not that popular)

MPEG-4: Multimedia for fixed and mobile web applications (e.g. DRM)

Each of the above standards has layers. For this lecture we primarily constrain ourselves to MPEG-1 Layer II Audio Encoding. For a complete history, please see [1].

2 Compression: Why is it possible?

A simplified view on this is to consider the various correlations that might exist in the original signal. Three major types of correlation may be exploited for compression purposes:

- SPATIAL CORRELATION: Spatially neighboring samples exhibit similar values. (Think about pixels in an high resolution image.)
- SPECTRAL CORRELATION: Multiple source signals exhibit spectral correlation. (Think about the left and right channels of a stereo audio signal.)
- TEMPORAL CORRELATION: Samples in adjoining *frames* exhibit similar values. (Think about a specific pixel in a video stream.)

2.1 Video Encoding Example

A Compact Disc (CD) has capacity 650MBytes. Compare the duration of a 30fps, 720×480 pixel video stream that can fit on this CD with and without compression. Assume that uncompressed NTSC video uses 2 bytes per pixel, and that compressed NTSC video results in a bit rate of 1.5Mbits/second.

Solution:
Without Compression:

Number of pixels per second $= 720 \times 480 \times 30 = 10368 \times 10^3$ pixels/second

$\Rightarrow$ We get a bit rate of $2 \times 10368 \times 10^3$ MBytes/second
RESULT = 31 seconds.

With Compression:
Given a bit rate of 1.5Mbits/second we can directly compute the capacity
RESULT = 58 minutes.

In practice, the bit rate varies from one implementation to the next, and hence the answer is closer to 74 minutes.

3 Lossless versus Lossy Compression

3.1 Lossless Compression

The main idea for our purposes is the following: Lossless compression allows for the *perfect* reconstruction of the encoded signal at the decoder.

Some techniques used are:

- Huffman encoding
- Airthmetic encoding

Common applications include ZIP archives and Tarballs.

3.2 Lossy Compression

Lossy compression is also an immense topic. In this tutorial we shall consider lossy compression as an encoding scheme that discards parts of the original signal. Hence the *perfect* reconstruction of the encoded signal is (in general) not possible. Although, we shall see that we can design encoding schemes to reproduce the original signal to a "perceivably correct" level.

In most cases, lossy compression schemes fall into two categories in which

1. Certain constraints on the input signal are assumed.
2. Information is assumed about usage of the decoded signal.

MPEG audio compression falls into this second category. The assumption is that a *human* will *listen* to the decoded audio signal, allowing the encoding scheme to exploit the many non-linearities of the *human ear*.[1] JPEG image compression behaves similarly where the non-linearities of the *human eye* are exploited.

4 MPEG-1 Layer II – The Encoder

Figure 1 presents a simplified outline of the MPEG-1 Layer II Audio encoder. Assume $x[n]$ is a stable signal, obtained by sampling $x_c(t)$ strictly above its Nyquist rate.

[1] More on this in a bit.

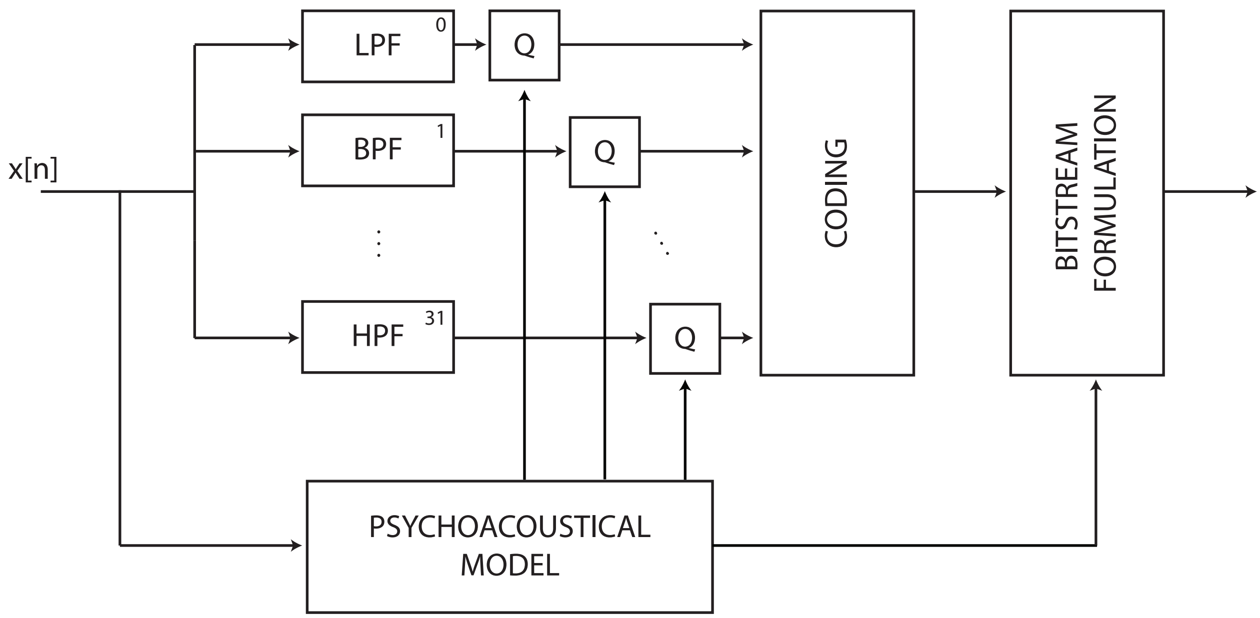

Figure 1: Simplified MPEG-1 Layer II Encoder

4.1 Filterbank

A couple of points regarding the filterbank:

- The filterbank divides the 0-20kHz audible spectrum into 32 equal width bins (with some overlap). This "resolution" of the audible spectrum is too coarse to perform any special operations.
- The filters are implemented as 512 tap Finite-Impulse Response (FIR) filters. This provides a good amount of control over the properties of the individual filters.
- The filter bank is critically sampled. For our purposes, it suffices to think of this as follows: for every 32 samples of $x[n]$, the filter bank generates 32 output samples, 1 for each filter block. This is done to minimize the number of bits to be encoded.

4.2 Psychoacoustic Model

The human ear is a nonlinear hearing device. In particular it exhibits, what we call, critical bands. The human ear tends to *blur* sounds within a critical band. These bands can be as thin as $\leq$ 100Hz in the lower frequencies, to as high as $\geq$ 4kHz in the higher frequencies.

So the main idea of the psychoacoustic model is to build a noise mask as a function of frequency for the audio signal to be encoded. Using such a mask, the psychoacoustic model can determine how many bits to use for quantization purposes, which brings us to the next stage.

For a much more thorough study of the psychoacoustic model (and MPEG audio coding in general) please refer to [2, 3, 4].

4.3 Quantization, Coding, and Bitstream Formulation

QUANTIZATION:
The quantization stage will vary with each cycle of the encoding process. Such quantizers are said to be *adaptive*. As the psychoacoustic model determines the optimal bit distribution, the quantization for each sub-band will be performed at a specific resolution.

CODING:
Here come the Error-Correcting Codes and general coding schemes. As a topic out of the range of this course we have left this part out. For those who are interested, consider taking ECE417 and ECE1501 (graduate).

BITSTREAM FORMULATION:
The most important thing to realize in this stage is that the quantization bit distribution information is also packaged along with the encoded audio samples. This is crucial, as it allows for the "blind decoding" of the encoded audio signal on the decoder end—i.e. the decoder has no knowledge of the psychoacoustic model and its inner operation scheme.

5 MPEG-1 Layer II – The Decoder

As mentioned above, no *a priori* information is required about the psychoacoustic model. It is completely up to the encoder to take care of the psychoacoustical signal processing. The decoder is thus able to decode any MPEG-1 Layer II audio signal regardless of the exact implementation of a psychoacoustic model on the encoder end.

Another crucial point is, due to the lower complexity, MPEG-1 Layer II decoders are cheap, which accounts for its wide usage in home video standards such as the VCD and DVD (without surround sound).

References

[1] (2008, November) MPEG - home page. [Online]. Available: http://www.chiariglione.org/mpeg/

[2] D. Pan, "A tutorial on MPEG/audio compression," *Multimedia, IEEE*, vol. 2, no. 2, pp. 60–74, Summer 1995.

[3] K. Pohlmann, *Principles of Digital Audio*. McGraw-Hill/TAB Electronics, 2005.

[4] V. Bhaskaran and K. Konstantinides, *Image and Video Compression Standards: Algorithms and Architectures*. Kluwer Academic Publishers Norwell, MA, USA, 1997.