Modeling of an MPEG Layer-3 Encoder and Decoder in Ptolemy

Literature Survey

Patrick Brown
EE382C – Embedded Software Systems
Spring, 2000

Abstract

MPEG Audio Layer-3, or “MP3,” has rapidly become the most popular format for the encoding and compression of digital audio and has found widespread application in both hardware and software. This literature survey provides a brief introduction to the MPEG (Moving Picture Experts Group) standard for audio encoding and explores the feasibility of formally modeling an encoder and a decoder. It also describes some of the available literature on the MPEG standard and associated research. The tremendous popularity of the standard has led to a large amount of literature, research, and source code related to the encoding and decoding of MPEG audio. The focus of the actual project will be a model of an MP3 encoder and decoder in Ptolemy [1]. This model will be used to generate C code that can be compared, in terms of speed and memory efficiency, to several of the widely used codecs currently available.

Introduction

MPEG Audio Layer-3, more commonly known as "MP3," is part of the set of standards created by the Moving Picture Experts Group (MPEG) [2]. Currently, there are three complete standards that have been created by MPEG and adopted by the International Organization for Standardization (ISO). They are

- MPEG-1, approved Nov. 1992
- MPEG-2, approved Nov. 1994
- MPEG-4, approved Oct. 1998 (version 1) and Dec. 1999 (version 2)

Each of these standards is made up of several parts. The three main parts to all three are systems, video, and audio. The "systems" section of the standard provides for the synchronization of the video and audio data. The overall focus of the MPEG standards is the compression of high-quality, synchronized audio and video to a data rate of approximately 1.5 Mbps [3]. Recently, however, there has been extremely rapid growth in audio-only applications of the standard. This is largely due to the high compression that is possible while maintaining very high sound quality. The following table [4] shows some of the most common compression ratios available using MP3 compression based on the relative quality of the resultant audio.

Sound Quality	Bandwidth	Mode	Bit-Rate	Compression Ratio
Telephone	2.5 kHz	mono	8 kbps	96:1
AM Radio	7.5 kHz	mono	32 kbps	24:1
FM Radio	11k Hz	stereo	56-64 kbps	26:1 – 24:1
Near CD	15 kHz	stereo	96 kbps	16:1
CD Quality	Over 15 kHz	stereo	112+ kbps	Up to 12:1

Within the MPEG Audio standard, there are three “layers.” Layer-1 represents the least complex encoding scheme and produces the lowest compression ratio, while Layer-3 is the most complex but produces the highest compression ratios. Layer-3 is used to produce bit-rates of 128 kbps and lower. Layer-1 and Layer-2 can produce compression ratios of up to 4:1 and 8:1, respectively [4].

Project Objectives

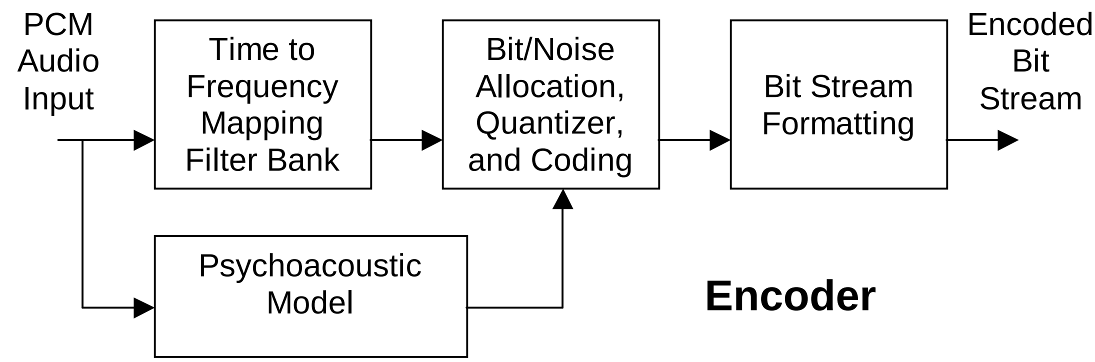
The final result of this project will be a model of an MPEG Layer-3 encoder and decoder created using Ptolemy. There are many source code samples available, including one example at the Fraunhofer Institute’s web site [4]. There are also many freeware, shareware, and commercial products, such as DLLs (dynamically linked libraries) and MP3 decoders/players. Although the encoding and compression scheme is an ISO standard, there seem to be a wide variety of implementations of the algorithm. Some of them are based directly on the Fraunhofer source example, rather than the ISO standard itself. As a result, there is also a wide range of performances among the different encoders. One of the commercial products claims to be as much as eight times faster than the others [7]. Once a model is produced in Ptolemy, the performance of various aspects of the encoding algorithm can be analyzed in terms of both speed and memory usage. Finally, the model will be used to generate C code which will be compiled and run on the x86 architecture and compared to the other products.

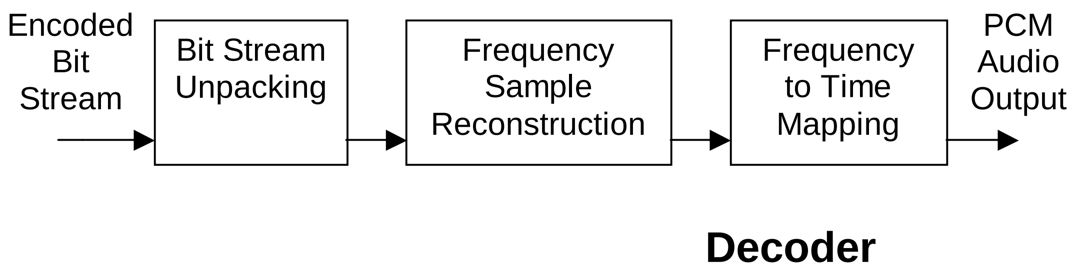
Project Context

MPEG Layer-3 is a relatively complex compression scheme. The IEEE Multimedia Journal article by Davis Pan [3] provides a detailed explanation of the process, but it does not present a formal model. By constructing a formal model in Ptolemy, it may be possible to improve upon data memory usage, code memory usage, or the overall speed of the encoding algorithms. In addition, MP3 audio-only hardware has begun to appear recently in the form of portable digital music players and car stereos. This increases the need for a formal model due to the much more strict memory and efficiency requirements of hardware implementations.

Implementation

The following diagram, from reference [3], shows the basic function of the encoder and decoder.





Modeling the encoder will be more difficult than the decoder. Normally, there is no need to encode digital audio in real-time, but decoding must be done in real-time because it is done during playback. Although Layer-3 is a complex encoding scheme, the abundance of available examples and source code should make the final project feasible.

A model of the “time to frequency mapping filter bank” should fall strictly within the SDF domain. This is the central component of the algorithm and involves using a poly-phase filter bank to divides the input signal into 32 equal width frequency sub-bands [3]. The “psychoacoustic model” is based on the way in which the human ear perceives sound. Like the main filter bank, it also uses a time to frequency mapping (Fourier transform) to divide the input signal into sub-bands, but it uses a finer frequency resolution. It then identifies which frequency components are the most important based on certain properties of the human auditory track. This block will most likely be modeled using SDF, also. The "bit/noise allocation, quantizer, and coding" block then uses the information generated by the psychoacoustic model, along with the total amount of energy present in each sub-band, to allocate bits to each of the 32 sub-bands [3]. In this way, the frequency content is compressed down to only those components that are actually necessary to reproduce the original sound. This must be modeled using DDF, because the amount of output produced by each sub-band will vary depending on the decisions made by the psychoacoustic model. Finally, the encoded audio is formatted to facilitate decoding. At this stage, ancillary data may also be added, such as copyright information and other text.

Decoding is much simpler, because it only involves adding the 32 sub-bands back together. It is important to note that this compression algorithm is not lossless [3]. However, it is designed to be "transparently lossy," meaning that the information that is discarded is (hopefully) imperceptible to humans. This claim is certainly subjective, and continues to be a source of debate among high-fidelity audio enthusiasts. A large number of tests have been performed in an attempt to decide this, and almost everyone agrees that there is no perceivable difference in sound quality between a 1.4 Mbps, uncompressed CD audio signal and a 192 kbps, MPEG Layer-3 compressed audio signal. Most of the disagreement is to whether or not the lower bit-rates (160 and 128 kbps) are noticeably different [4].

Additional Notes

The Layer-3 audio compression standard is largely unchanged between the different versions of the MPEG standards. Additions have been made, such as multi-channel support and lower compression rates (down to 8 kbps) [2]. The standard has been the subject of much controversy because it makes the illegal distribution of copyrighted material much easier. The recording industry is gradually accepting the standard, however, because it provides a feasible means of selling music over the Internet. Most people would agree that the benefits of the standard far outweigh the problems created by it.

References

1. "The Ptolemy Project," UC Berkely, Department of EECS, http://ptolemy.eecs.berkely.edu.
2. "The Moving Picture Experts Group Home Page," http://cselt.stet.it/mpeg.
3. Davis Pan, "A Tutorial on MPEG/Audio Compression," *IEEE Multimedia Journal*, vol. 2, no. 2, Summer 1995, pp. 60-74.
4. "MPEG Audio Layer-3," Fraunhofer-Gesellschaft Institute, http://www.iis.fhg.de/amm/techinf/layer3.
5. N. S. Jayant, E. Y. Chen, "Audio Compression: Technology and Applications," *AT&T Technical Journal*, vol. 74, no. 2, March-April 1995, pp. 23-34.
6. Bosi Brandenberg, "Overview of MPEG Audio: Current and Future Standards for Low Bit-Rate Audio Coding," *99th AES*, New York, October 1995, pp 4130.
7. "Xing Technology Corporation," http://www.xingtech.com/mp3.