Simulation of Digital Audio Compression

VLADISLAV SKORPIL, ABDULHAKIM ABUZAHO
Department of Telecommunications
Brno University of Technology
Purkynova 118, 612 00 BRNO, Czech Republic,
CZECH REPUBLIC

http://www.vutbr.cz

Abstract: This paper presents the digital data compression of multimedia transmission system. Digital audio compression is a scheme, which allows efficient storage and transmission of audio data. The main objective of this research is to show the program in MATLAB for Moving Pictures Expert Group (MPEG-1) audio encoder and decoder for layer-1, which will be analyzed in future on Digital Signal Processing (DSP) in real time. This work is constrained to the time-consuming sections of the MPEG-1 Layer-1 audio standard.

Compression is required for efficient transmission in order to send more data in the available bandwidth, or to send the same data in less bandwidth and more users can use it on the same bandwidth. It can also be used for storage purposes to compress more data and can compress for local storage, put details on cheaper media. It is also useful for progressive reconstruction, scalable delivery, browsing and as a front end to other signal processing.

Key-Words: - Compression, MATALB, MPEG-1, Digital audio, Simulation

1   Introduction

Sound, which is a pressure difference in air, when picked up by a microphone and fed through an amplifier converts to voltage level. The voltage is sampled by computer a number of times/second. For CD-audio quality, we need to sample 44100 times/second and each sample has a resolution of 16 bits. In stereo, this gives us 1.4Mbit/s and it can probably feel the need for compression.

To compress audio signal, Moving Picture Expert Group system (MPEG) tries to remove the irrelevant parts of the signal and the redundant parts of the signal. Parts of the sound, which are not audible can be discarded. For this purpose, MPEG audio uses psychoacoustic principles.

MPEG can compress to a bitstream of 32kbit/s to 448kbit/s (Layer-1). A raw PCM audio bitstream is about 705kbit/s, so this gives a maximum compression ratio of about 22. Normal compression ratio is more like 1:6 or 1:7. Unlike video, it is talking about no perceivable quality loss here. 96kbit/s is considered transparent for most practical purposes. This means that it will not notice any difference between the original and the compressed signal for rock'n roll or popular music. For more demanding stuff like piano concerts and such others, we will require to go up to 128kbit/s [1].

2   Signal compression

Signal compression is a process that decrease amount of data in order in to it's storage or transmission. The compression efficiency is given by so-called compression ratio P_k :

$$P_k = \frac{K_0}{K_k} \qquad (1)$$

where K_0 is the original data amount and K_k is the data amount after compression. "Data amount" term relates to " bit rate" term if compressed signal is about to be transmited. High compression ratio causes low bit rate needed for transmission of the same information in the same time-in other words, high P_k is associated with strong reduction of the bit (transfer) rate.

Compression techniques can be separated into two groups according to principle of their operation- lossless compression and lossy compression. Backup of the computer programs and data is the best-know utilization of the lossless compression algorithms. In this case every loss of information causes the crash. These algorithms remove certain redundancy algorithms from the compressed data and they generally have low compression ratio P_k [4]. These algorithms are used rarely for the audio compression [21].

So-called lossy algorithms have generally much higher ratio P_k but at the expense of loss of the information portion. Only these algorithms are efficient for the effective audio compression. Loss of data portion is not harmful because the audio signal has redundancy and irrelevancy whose omission dose not changes the whole subjective perception .

Great number of the compression techniques based on pieces of knowledge about perception has been designed which reach high P_k ratio with selectable subjective degradation of the signal quality. They are suitable for the digital television and sound broadcasting (e.g. all-European TV system DVB- digital Video Broadcasting), ISDN videoconferencing, Internet and data storage on media of DVD type (Digital Versatile Disc) or MD type (Minidisk)

3 Lossy compression

In 1948 Shannon introduced the source theory where accuracy of signal source representation is the main criteria [16]. Bit rate – information amount necessary for representation of the signal characteristics – depends on the distortion that it allow for their transmission, i.e. the accuracy that it want to use for the signal source specification. Well- known equation holds for information capacity of the ideal frequency limited channel [16]:

$$C = W \log_2 (1 + P/N) \quad [\text{bit/s}] \quad (2)$$

Where P is the maximum input signal power, W is the bandwidth, and N is the power of additive white Gaussian noise.

Information amount R that is generated by the signal source can be written using equation corresponding to the equation (2):

$$R = W \log_2 (S/N) \quad [\text{bit/s}]$$
only for the big S/N   (3)

It holds in case of the white Gaussian noise with the limited frequency band with width W and power S. In this case N means the estimation error-distortion. It represents a certain level of root means error between the signal {x(t)} and the signal estimation $\{\hat{x}(t)\}$ reconstructed in terms of the {x(t)} related data supplied by the bit rate R:

$$N = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} E\left[\hat{x}(t) - x(t)\right]^2 dt \quad (4)$$

D is used instead N in the modern information theory as symbol for the average distortion. Then the equation can be written

$$R(D) = W \log_2 (S/D) \quad [\text{bit/s}]$$
only for the big S/D   (5)

Problems with utilization of this equation appear in the cases of:

1)   Signal is not Guassian

2) Distortion dose not depend only on the difference $\{x(t)\} - \{\hat{x}(t)\}$ Distortion is specified by function different from the difference square function

3) Spectral density of the signal is not constant

Second and third case is interesting for us. They show that it can choose function different from the root mean error function for determination of signal distortion and it can reach lower bit rate that can be used for signal source specification. In the case of sound, pieces of knowledge of psychoacoustics can be used for error determination.

4 Psychoacoustics

Utilization of psychoacoustics knowledge has fundamental meaning for the lossy coding. Psychoacoustics is the scientific branch, which deals with research of human hearing. It studies principle of hearing, time frequency characteristics of ear, energetic viewpoints (threshold in quite and threshold of pain) and final interpretation depend on used measuring practising perception tests whose procedure and final interpretation depends on used measuring method. Measuring methods differ in psychoacoustics problem types[10].

Often used methods are:

•	Setting method – tested person sets e.g. level of pure tone as this tone is audible

•	Cueing method – tested person determines whether tone level increases or decreases

•	Yes – No method – tested person decides whether relevant signal is audible or not

5 Hearing Area

The hearing area is a plane in which audible sound can be displayed. The normal form of the hearing area is usually plotted in as frequency logarithmic scale on the X-axis and sound pressure level in dB linear scale on Y-axis.

The actual hearing area limits, that, hold for pure tones in steady state condition, lies between the thresholds in quit and the threshold of pain [10]. The components of music encompass a larger distribution in the hearing area are shown in figure by hatching. The high level border which known as the limit of damage risk that's very important in everyday life: reached at quite high sound pressure at very low frequencies is also indicated in the figure by the thin dotted line. area between threshold in quiet and threshold of pain. Also indicated are the areas encompassed by music and speech, and the limit of damage risk.

6 MATLAB Computing of the sound pressure level

The more powerful version was chosen for the determination of the sound pressure level in program in MATLAB. In relevant equation a maximum is found from all spectral lines in each subband and one pressure level gives by scalefactors. In fact, it means two loops, first is making a searching through spectral lines in each subband, second is searching just through subbands comparing to an appropriate pressure level and maximum from first loop.

In this section of program, only one loop is possible to see. It is just first step from both. second is comparing at the end of the psychoacoustic model, which is given together with calculating *SNR* (signal-to-noise ratio). It is more powerful as one loop is spared.

7   Calculation of the signal-to-mask ratio

According to the MPEG-1 standard, the signal-to-mask ratio for each sub band is given by

$$SMR_{sb}(n) = L_{sb}(n) - LT_{min}(n) \qquad [dB]$$
(6)

and the mean, which is computed for every sub band n.

8 SMR Calculation in MATLAB

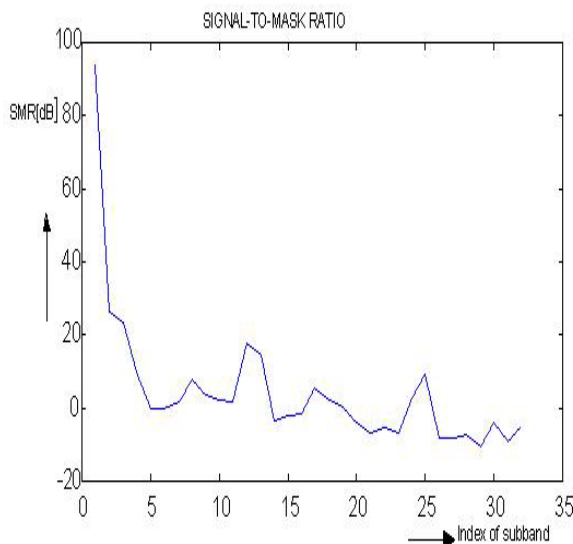This part of the program computes SMR in accordance with equation (6).

Fig. 1: SMR for tested signal

The curve on Fig. 1 shows the SMR in all sub bands. The SMR is used to determine the number of bits in each sub band. It is clear that in lower sub bands, more bits are transmitted than in higher ones. This depends on the input signal, which had six tones in the low-frequency spectrum

9 Bit Allocation

Under the MPEG-1 standard, each iterative procedure for calculating the bit-allocation vector per data frame requires several steps: first, calculating the minimal mask-to-noise ratio (MNR) for all sub bands, then selecting and updating the next quantization level for the sub band with the minimal MNR, then computing the new MNR of the specific sub band, and finally, updating the remaining bit count available for allocation.

Before adjusting to a fixed bit rate, two things must be determined: the **number of bits available for coding** and **the scale factors**. This number can be obtained by subtracting the following from the total number of bits *cb* available for one frame:

-    The number of bits needed for the header *bhdr* (32 bits),
-    The CRC check word *bcrc*, if it is used (16 bits),
-    The bit allocation *bbal,*
-    …And the number of bits required for ancillary data *banc*.

10   Conclusion

Taking the standard 12cm audio CD as an example, the digital output data rate from a 20kHz stereo CD is 1,411,200 bits/second. In terms of storage space, it would take 650MByte to store one hour of CD playtime and 1.5MHz of circuit bandwidth to carry that digital signal. Except within a studio environment the implementation of this and similar 16bit linear PCM format are deemed excessive and expensive. Digital audio data compression or bit rate reduction, as it is sometimes known, can reduce this data rate and hence the storage and transmission demands by a factor of between four and twelve.   This has a number   of   distinct   advantages.   For example, a combination of direct dial digital telephone lines (ISDN or SW56) and data compression offers broadcasters and other professional audio facilities alike an extremely economical and time saving means of networking high quality audio in real time. Additionally, it would be impossible without the application of some form of data compression to adapt certain

digital storage media (such as floppy disks) for use in digital audio cart machines.

References

[1] ISO/IEC 13818-7 Information Technology – Generic coding of moving pictures and associated audio information – part 7: Advanced Audio Coding (AAC), 1996, 131s.

[2] Bosi, M. and Brandenburg, K. ISO/IEC MPEG-2 Advanced Audio Coding. Journal of the Audio Engineering Society, Vol.45, 1997, No. 10, p. 789-814

[3] EN ISO/IEC 11172-5 Information technology – Coding of Moving Pictures and Associated Audio for Digital Storage Media up to about 1,5 Mbit/s – part 5: software simulation

[4] Serantes, C.and Pena, A. and Prelcic, N. A Fast NOISE-scaling Alogrithm for Uniform Quantization in Audio Coding schemes. IEEE Proceedings, 1997 , p 339-342

[5] Wei, X. and Shaw, M.: Optimum Allocation and Decompstion for High Quality Audio Coding. IEEE proceedings, 1997, p. 315-318

[6] Chan, W.Y. and Gersho, A. High Fidelity Audio Coding with Generalized Product Code VQ. Speech and Audio Coding For Wireless and Network Application, Kluwer Academic Publishers, 1998, p. 153-159

[7] Glchrist, N.: ATLANTIC audio: Preserving Technical Quality During Low bit rate Coding and Decoding. http://www.bbc.co.uk/atlantic/index.htm

[8] Ritscher, S. and Felderhoff, U. Cascading of Different Audio Codecs.100th AES Convention, Copenhagen, 1996,

[9] Kurth, F. An Audio Codec for Multiple Generation Compressions without Loss of Perceptual Quality. AES 17th conference, High Quality Audio Coding, Florence Italy, 1999

[10] zwicher, E.and Factl, H. psychoacoustics – Facts and Models. Berlin Heidelberg, springer-Verlag, 1990, 351 s.

[11] EN ISO/IEC 11172-3 information technology – coding of moving pictures and associated

[12] Pan, D. A Tutorial on MPEG/Audio compression. IEEE multimedia, summer 1995, p. 60-74

[13] Chan, D.Y.,Yang, J.F.and Fang, Ch.Ch.Fast Implementation of MPEG Audio Coder using Recursive Formula with Fast Discrete Cosine Transforms. IEEE Transactions on speech and Audio Processing, Vol. 4, No. 2, 1996, p. 144-148

[14] Brandenburg, K. and Stoll, G. ISO-MPEG-1 Audio: A generic standard for Coding of High-Quality Digital Audio. Journal of the Audio Engineering socity, Vol. 42, 1994, No 10,p. 780-792.

[15] Brandenburg, K. and Bosi, M. Overview of MPEG Audio: Current and Future Standards for Low-Bit-Rate Audio Coding. Journal of the Audio Engineering Society, Vol.45, 1997, No ½ ,p. 4-21

[16] Berger, T. and Gibson, D. Lossy source Coding. IEEE Transaction on Information Theory, Vol. 44, No. 6, 1998, p.2693-2723

[17] ITU-R BS.1116, Methods for the Subjective Assessment of small Impairments in Audio Systems Including Multichannel Sound Systems, Geneva Switzerland (1994)

[18] Boland, S.and Deriche, M. New Results in Low Bitrate Audio Coding Using a Combined Harmonic-Wavelet Representation. IEEE Proceedings, 1997 p. 351-354

[19] Sreenivas, T.V and Dietz, M. Vector Quantization of Scalefactors in Advanced Audio Coder (AAC). http://www.iis.fhg.de/amm/, 4 s.

[20] ISO/IEC 13818-3 Iformation technology – Generic Coding of Moving Pictures and Associated Audio Information – part 3: Audio, 1995, 101 s.

[21] Craven, P.and Gerzon, M. Lossless Coding for Audio Disc. Journal of the Audio Engineering Society, Vol. 44, 1996, no. 9, p. 706 –720

[22] Galko, M., Krbilová, I., Vestenický, P.: Blocking Probability Influence on the Single-Channel Service System Operation with Various Input Flows. Proceedings of TRANSCOM'97 Volume 2, University of Žilina, Žilina 1997, pages 37-40

[23] ABUZAHO, A. Digital Audio Compression. Doctoral Thesis, BUT Brno 2003

[24] Tomašov, P., Krbilová, I., Muzikářová, Ľ., Vestenický, P.: Operation Control of Railway Telecommunication Network. Proceedings of 10th International Conference TEMPT'97, Sofia 1997

[25] Vestenický, P., Krbilová, I.: Perspectives of Information Networks Development. Sborník přednášek celostátní konference s mezinárodní účastí TELEKOMUNIKACE '98. VUT, Brno 1998

[26] Krbilová, I., Vestenický, P.: Development Trends of Internet Services in Next Millennium. Zborník prednášok konferencie s medzinárodnou účasťou SLOVENSKO A INFORMAČNÁ SPOLOČNOSŤ. Slovenská elektrotechnická spoločnosť, Stupava 199

[27] Vestenický, P.: Solution of some technical problems in marker and marker locator development. Communications 6, 2004, No. 4, Pages 103 – 106. ISSN 1335 – 4205