COMPARISON BETWEEN AAC AND HE-AAC MODELS FOR DESIGNING AN AUDIO CODEC

A.A. VALADI[1] **M. ALEXANDRU**[2]

Abstract: *The article discuses about audio signal encoder and decoders while analyzing the human auditory system along with its handicaps in receiving correctly the original transmitted waveforms; it also makes a brief introduction to the MPEG organization, its standards and the AAC, HE-AAC models taking in consideration its implementing possibilities on different platforms; in the final part of the paper a conclusive example is given on the importance of coding a signal with HE-AAC.*

Key words: *Codec, AAC, HE-AAC, Psychoacoustics.*

1. Introduction

Audio coding is used to compress digital signals so that it can reduce the amount of bits needed to transmit data in a digital format [2]. This is mainly used when the network bandwidth or the storage capacity of the devices is limited. The compression of data is based on encoding and decoding algorithms. The encoding part transforms the waveform into a coded representation, thereby compressing the audio signal. After that the obtained bit-stream needs to be restored eventually, this process representing the decoding part; in this way the signal transforms into an uncompressed data once again.

ISO-MPEG is a standard for high-quality, low bit rate video and audio coding. The standards audio part is composed by algorithms to reduce the bit-rate, while the quality of the audio signal remains fairly intact.

2. Psychoacoustics and Perceptual Coding

Psychoacoustics deals with the understanding of the perception of sound by the human auditory system. Algorithms are designed in a way to use this auditory handicap for its coding benefit; it takes advantage of psychoacoustic features to compress audio signals without degrading their quality, or doing it so in a way that a person won't hear noticeable differences.

The psychoacoustic model is based on many studies of human perception. The prior works have shown that the average human does not hear all frequencies the same. These occur due to the effect produced by different sounds in the environment and limitations of the human sensory system; thus unnecessary data may be cut out of an audio signal. A person can hear frequencies in the range from 20 Hz to 20,000 Hz, however this does not mean that all the frequencies are received in the same way. Usually, when a frequency is in

---

[1] Elettronika Research S.R.L., Braşov.
[2] Dept. of Electronics and Computers, *Transilvania* University of Braşov.

the lower or upper limit of the hearing range, it becomes more difficult to be received correctly by one.

One other observation forms the basis for modelling. Because humans hear lower frequencies, like those making up speech, with higher accuracy like high frequencies around 20 kHz, the ear has better capability in detecting differences in pitch at lower frequencies than at high ones (Figure 1). The frequency range from 20 Hz to 20,000 Hz can be broken up into critical bandwidths, which are non-uniform, non-linear, and dependent on the heard sound. Signals within one critical bandwidth are hard to separate for a human observer. A more uniform measure of frequency based on critical bandwidths is the Bark.

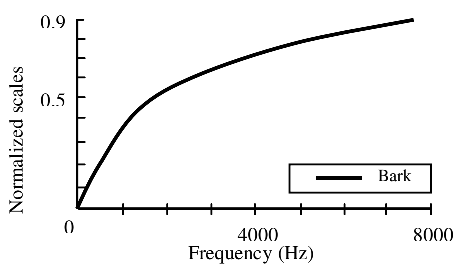

Fig. 1. *Matlab graphic illustration of auditory perception from 0 Hz to 8000 Hz*

Humans are not able to hear small differences in frequencies. It is very difficult to percept differences between signals of 1 kHz and 1,001 kHz. This process is even harder if the two sounds are received at the same time. Furthermore, the 1 kHz signal would also affect the human's ability to hear a signal that is in its immediate neighbourhood. This concept is known as masking. If the 1 kHz signal is stronger than the signal at its nearby frequencies, it will mask them, making them inaudible to the listener. For a masked signal to be heard, its power needs to be increased to a higher level than that of a threshold that is determined by the frequency of the masker tone and its strength.

If any frequency components around these maskers fall below the masking threshold, they can be discarded, thus giving birth to a wide range of algorithms for compressing an audio signal.

3. Encoding and Decoding

The encoding part is the one that transforms the compressed pulse code modulated (PCM) signal into a coded representation, thereby compressing the audio signal (Figure 2). The newly formed audio signal takes fewer bits to represent the original information because of the set of rules created by the encoder. This set of rules is a bit-stream that may be then transmitted.

The compressed audio signal eventually needs to be rebuilt to its original form (e.g. for playing back). In this process the coded audio signal is decoded so it will loose its compressed format and return to the original structure. The decoder receives the bit-stream and reconverts it into a PCM representation through audio synthesis, based on the set of rules held by the bit stream. The encoding/decoding process is presented in the figure below:

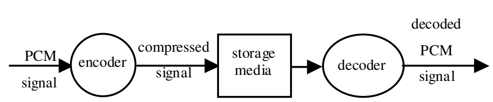

Fig. 2. *Encoding and decoding of a PCM signal*

The entire process which contains the compression and decompression steps may be ***lossless*** or ***lossy***. The lossless method consists of a bit exact restoring method of the original audio signal, thus the original and the decoded signal hold the same byte-sequence. Therefore, no audio information is lost during this process. On the other hand, the lossy method does not guarantee that the decoded signal is a replica of the original signal. In this way, the final form of the bit-stream may wary from the original bit map.

Lossy compression methods generally make use of characteristics and limitations of the human auditory system (e.g. masking), to encode the signal. In this way, the psychoacoustic models provide inaudible degradation of the quality of the original waveform. This degradation is a widely spread method of reducing audio data from the original signal, thereby reducing the amount of bits needed to represent the information.

4. MPEG-4 AAC and HE-AAC Standards

MPEG, mostly known by its nickname, stands for **M**oving **P**ictures **E**xperts **G**roup that has been set up in 1988 to develop generic standards for the coded representation of moving pictures, associated audio and their combination. It has developed many worldwide known standards but the most widely used audio compression formats are MPEG-1, 2, 3, 4.

The upper standards are the most spread for multiple reasons. It's utterly important to point out that they are defined as open standards; the specifications are available for anyone interested in implementing it. The standards [4], [5] consist of a collection of algorithms for coding video and audio. It also consists of many exemplary algorithms that an engineer could follow but it's all up to the person in need to decide on the best solution.

AAC (Advanced Audio Coding) is a lossy compression method which is included in the MPEG-2 and MPEG-4 specifications. With the AAC coding method it's possible to include 48 audio channels plus 16 low frequency effects.

High Efficiency AAC (HE-AAC or aacPlus) [3], is the next generation coding standard, and is a part of the MPEG-4 Part 3 specification. This coding method combines AAC and **S**pectral **B**and **R**eplication (harmonic redundancy in the frequency domain) and since its second version it includes the parametric stereo feature. The codec can be very usefully implemented for the compression of streaming data [1]. It can operate at very low bit-rates and is mainly used for internet radio streaming. Studies have shown that even a 48 kilobit-per-second stream is likely to have a higher quality than a 128 kilobit-per-second MP3.

5. Implementation and Platforms

An engineer has many possible methods in sight for implementing a platform on an electronic system. One first option is the running of software on a general-purpose processor. An interesting way to develop the software would have its basis on the IBM-PC architecture, since it's very common and widespread on the world market. When developing the software for the PC, there are some important choices to be made: for example an application that only uses the standard PC (8086) instruction set could be developed, or special-purpose multimedia instructions sets could be applied.

In most situations the use of personal PC, that are based on for example the Intel Pentium processors cannot be used, for its unitary cost is way to high for this kind of implementation. There are other options that can be considered depending on the production scale. For medium number of units forecasted to produce the best solution would be the implementation of a **D**igital **S**ignal **P**rocessor (DSP), since its unitary cost is lower then the cost of a general-purpose processor. However, when large amount of units are forecasted to go into production the best solution is a dedicated hardware application. The embedded systems are the functional integration of hardware and software for a specific application. In this case the hardware description, which is done with the help of **h**ardware **d**escription **l**anguage (HDL), may be accommodated on a **F**ield-**P**rogrammable **G**ate-**A**rray (FPGA), or detailed into a full-custom processor.

6. Application

One of the main advantages of the HE-AAC codec is the fact that it doesn't require large amount of bit-rate to ensure a high quality encoding of the audio information. This is particularly effective for the digital transmission of the encoded data from a studio to a transmitter (STL - **S**tudio to **T**ransmitter **L**ink). The relation between the necessary bandwidth, the bit-rate and the roll-off factor in the case of a PSK modulation is given by the $B = \dfrac{br\,(1+\alpha)}{\log_2 M}$ formula, where B is the bandwidth, br stands for bit-rate, α is the roll-off factor and M is the constellation number. Supposing the transmitter uses a QPSK modulator ($M = 4$), considering that the audio information is transmitted at a 64 kilobits-per-second rate (this will ensure a very good quality of the encoded data) and using a roll off factor of 0.1 (this is the usual value), the necessary bandwidth becomes 35.2 kHz. Thus, a very good quality of the audio stream can be ensured using a relatively narrow band for transmission and a relatively simple digital modulation method, which makes HE-AAC the ideal codec to be used for digital radio broadcasting.

DAB (**D**igital **A**udio **B**roadcasting) recommends this codec particularly for this kind of applications. DAB, also known as Eureka 147, is a digital radio technology for broadcasting radio stations, used in several countries, particularly in the UK and Europe. As of 2006, approximately 1,000 stations worldwide broadcast in the DAB format.

7. Conclusion

In this paper, the MPEG-4 AAC and HE-AAC standards were studied, resulting in a tutorial text covering basic aspects of the elements that have to be taken into consideration during a codec design and its implementation possibilities. Beside, some principles of Psychoacoustics were presented in the text, thereby enlightening the study of the general perceptual model.

The latter part of the work is dedicated on a comparison between the AAC and HE-AAC coding methods. Although it would seems that the HE-AAC holds a fundamental advantage over its predecessor, the fact is that the AAC is still widely implemented across many platforms because for the moment the HE-AAC is not an open source codec, thus making its implementation expensive for the time being.

References

1. Dalal, I.L.: *A Real-time AAC-type Audio Codec on the 16-bit dsPIC Architecture*. In: ICEEE 2007, Mexico City, Mexico, September 5-7, 2007. Available at: www.ee.cooper.edu/sprocom/files/ICEEE_07.pdf. Accesed: 20-03-2009.
2. Hanzo, L., Cherriman, P.J., Streit, J.: *Video Compression and Communications*. Second Edition, Wiley, 2007.
3. Wolters, M., Kjorling, K., Homm, D. Purnhagen, H.: *A Closer Look into MPEG-4 High Efficiency AAC*. Available at: www.telos-systems.com/techtalk/hosted/HE-AAC.pdf. Accesed: 05-04-2009.
4. ISO/IEC 13818-7:2006.
5. ISO/IEC 14496-5:2001/Amd 6:2005.