LOW BITRATE AUDIO CODING - STATE-OF-THE-ART, CHALLENGES AND FUTURE DIRECTIONS

Karlheinz Brandenburg
Ilmenau Technical University &
Fraunhofer IIS Arbeitsgruppe Elektronische Medientechnologie
Ilmenau, Germany

ABSTRACT

Perceptual encoding of high quality audio has found its way to many applications including digital radio, Electronic Music Distribution (EMD) systems and portable audio devices. An overview on the basics of high quality low bitrate audio coding will be followed by a look into currently widely used and newer, state-of-the-art coding systems like MP3 and MPEG-2 Advanced Audio Coding (AAC). The rapid deployment of older (1992) technologies (like MP3) followed by the news of new and improved algorithms (like AAC) raises the question about future improvements. The paper will analyse some candidates for such improvements and provide a view of some current research activities.

1. INTRODUCTION

High quality audio compression has found its way from research to widespread applications within a couple of years. Early research of 15 years ago was translated into standardization efforts of ISO/IEC and ITU-R 10 years ago. Since the finalization of MPEG-1 in 1992, many applications have beed devised. In the last couple of years, Internet audio delivery has emerged as a powerful cathegory of applications. These techniques made headline news in many parts of the world because of the potential to change the way of business for the music industry. Currently, among others the following applications employ low bit-rate audio coding techniques:

- Digital Audio Broadcasting (EUREKA DAB, WorldSpace, ARIB, DRM)
- ISDN transmission of high quality audio for broadcast contribution and distribution purposes
- Archival storage for broadcasting
- Accompanying audio for digital TV (DVB, ATSC, Video CD, ARIB)
- Internet streaming (RealAudio, Microsoft Netshow, Apple Quicktime and others)
- Portable audio (mpman, mplayer3, Rio, Lyra, YEPP and others)
- Storage and exchange of music files on computers

2. THE BASICS OF HIGH QUALITY AUDIO CODING

The basic task of a perceptual audio coding system is to compress the digital audio data in a way that

- the compression is as efficient as possible, i.e. the compressed file is as small as possible and
- the reconstructed (decoded) audio sounds exactly (or as close as possible) to the original audio before compression.

Other requirements for audio compression techniques include low complexity (to enable software decoders or inexpensive hardware decoders with low power consumption) and flexibility to cope with different application scenarios. The technique to do this is called perceptual encoding and uses knowledge from psychoacoustics to reach the target of efficient but inaudible compression. Perceptual encoding is a lossy compression technique, i.e. the decoded file is not a bit-exact replica of the original digital audio data.

Fig 1 shows the basic block diagram of a perceptual encoding system.

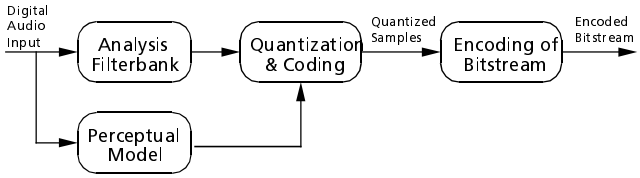

Figure 1: Block diagram of a perceptual encoding/decoding system.

It consists of the following building blocks:

- **Filter bank:** A filter bank is used to decompose the input signal into subsampled spectral components (time/frequency domain). Together with the corresponding filter bank in the decoder it forms an analysis/synthesis system.
- **Perceptual model:** Using either the time domain input signal and/or the output of the analysis filter bank, an estimate of the actual (time and frequency dependent) masking threshold is computed using rules known from psychoacoustics. This is called the perceptual model of the perceptual encoding system.
- **Quantization and coding:** The spectral components are quantized and coded with the aim of keeping the noise, which is introduced by quantizing, below the masking threshold. Depending on the algorithm, this step is done in very different ways, from simple block companding to analysis-by-synthesis systems using additional noiseless compression.
- **Encoding of bitstream:** A bitstream formatter is used to assemble the bitstream, which typically consists of the quantized and coded spectral coefficients and some side information, e.g. bit allocation information.

All current high quality low bit-rate audio coding systems follow the basic paradigm described above. They differ in the types of filterbanks used, in the quantization and coding techniques and in the use of additional features.

3. STANDARDIZED CODECS

MPEG (formally known as ISO/IEC JTC1/SC29/ WG11, mostly known by its nickname, Moving Pictures Experts Group) has been set up by the ISO/IEC standardization body in 1988 to develop generic (to be used for different applications) standards for the coded representation of moving pictures, associated audio, and their combination. Since 1988 ISO/MPEG has been undertaking the standardization of compression techniques for video and audio. The original main topic of MPEG was video coding together with audio for Digital Storage Media (DSM). From the beginning, audio-only applications have been part of the charter of the MPEG audio subgroup. Since the finalization of the first standard in 1992, MPEG Audio in its different flavours (mostly Layer-2, Layer-3 and Advanced Audio Coding) has delivered on the promise to establish universally applicable standards.

3.1 MPEG-1

MPEG-1 is the name for the first phase of MPEG work, started in 1988, and finalized with the adoption of ISO/IEC IS 11172 in late 1992. The audio coding part of MPEG-1 (ISO/IEC IS 11172-3, see [1] describes a generic coding system, designed to fit the demands of many applications. MPEG-1 audio consists of three operating modes called layers with increasing complexity and performance from Layer-1 to Layer-3. Layer-3 (in recent years nicknamed **MP3** because of the use of .mp3 as a file extension for music files in Layer-3 format) is the highest complexity mode, optimised to provide the highest quality at low bit-rates (around 128 kbit/s for a stereo signal).

The following paragraphs describe the Layer-3 encoding algorithm along the basic blocks of a perceptual encoder. More details about Layer-3 can be found in [1] and [2]. Fig 2 shows the block diagram of a typical MPEG-1/2 Layer-3 encoder.

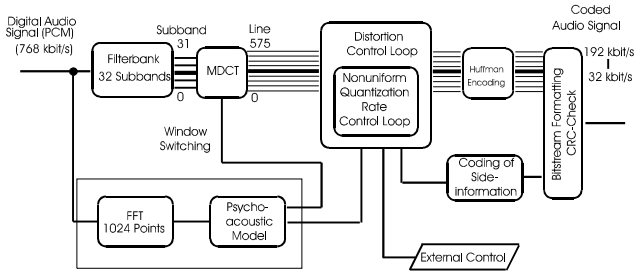

Fig. 2: Block diagram of MPEG Layer-3 (MP3) encoding

3.1.1 Filterbank

The filterbank used in MPEG-1 Layer-3 belongs to the class of hybrid filterbanks. It is built by cascading two different kinds of filterbank: First a polyphase filterbank (as used in Layer-1 and Layer2) and then an additional Modified Discrete Cosine Transform (MDCT). The polyphase filterbank has the purpose of making Layer-3 more similar to Layer-1 and Layer-2. The subdivision of each polyphase frequency band into 18 finer subbands increases the potential for redundancy removal, leading to better coding efficiency for tonal signals. Another positive result of better frequency resolution is the fact that the error signal can be controlled to allow a finer tracking of the masking threshold. The filter bank can be switched to less frequency resolution to avoid preechoes.

3.1.2 Perceptual Model

The perceptual model is mainly determining the quality of a given encoder implementation. A lot of additional work has gone into this part of an encoder since the original informative part in [1] has been written. The perceptual model either uses a separate filterbank as described in [1] or combines the calculation of energy values (for the masking calculations) and the main filterbank. The output of the perceptual model consists of values for the masking threshold or allowed noise for each coder partition. In Layer-3, these coder partitions are roughly equivalent to the critical bands of human hearing. If the quantization noise can be kept below the masking threshold for each coder partition, then the compression result should be indistinguishable from the original signal.

3.1.3 Quantization and Coding

A system of two nested iteration loops is the common solution for quantization and coding in a Layer-3 encoder. Quantization is done via a power-law quantizer. In this way, larger values are automatically coded with less accuracy and some noise shaping is already built into the quantization process. The quantized values are coded by Huffman coding. To adapt the coding process to different local statistics of the music signals the optimum Huffman table is selected from a number of choices. The Huffman coding works on pairs or quadruples. To get even better adaption to signal statistics, different Huffman code tables can be selected for different parts of the spectrum. Since Huffman coding is basically a variable code length method and noise shaping has to be done to keep the quantization noise below the masking threshold, a global gain value (determining the quantization step size) and scalefactors (determining noise shaping factors for each scalefactor band) are applied before actual quantization. The process to find the optimum gain and scalefactors for a given block, bit-rate and output from the perceptual model is usually done by two nested iteration loops in an analysis-by-synthesis way:

- Inner iteration loop (rate loop): The Huffman code tables assign shorter code words to (more frequent) smaller quantized values. If the number of bits resulting from the coding operation exceeds the number of bits available to code a given block of data, this can be corrected by adjusting the global gain to result in a larger quantization step size, leading to smaller quantized values. This operation is repeated with different quantization step sizes until the resulting bit demand for Huffman coding is small enough. The loop is called rate loop because it modifies the overall coder rate until it is small enough.

- Outer iteration loop (noise control loop): To shape the quantization noise according to the masking threshold, scalefactors are applied to each scalefactor band. The systems starts with a default factor of 1.0 for each band. If the quantization noise in a given band is found to exceed the masking threshold (allowed noise) as supplied by the perceptual model, the scalefactor for this band is adjusted to reduce the quantization noise. Since achieving a smaller quantization noise requires a larger number of quantization steps and thus a higher bit-rate, the rate adjustment loop has to be repeated every time new scalefactors are used. In other words, the rate loop is nested within the noise control loop. The outer (noise control) loop is executed until the actual noise (computed from the difference of the original spectral values minus the quantized spectral values) is below the masking threshold for every scalefactor band (i.e. critical band).

3.2 MPEG-2

MPEG-2 denotes the second phase of MPEG. It introduced a lot of new concepts into MPEG video coding including support for interlaced video signals. The main application area for MPEG-2 is digital television. The original (finalized in 1994) MPEG-2 Audio standard [3] just consists of two extensions to MPEG-1:

- Backwards compatible multichannel coding adds the option of forward and backwards compatible coding of multichannel signals including the 5.1 channel configuration known from cinema sound.
- Coding at lower sampling frequencies adds sampling frequencies of 16 kHz, 22.05 kHz and 24 kHz to the sampling frequencies supported by MPEG-1. This adds coding efficiency at very low bit-rates.

Both extensions do not introduce new coding algorithms over MPEG-1 Audio. The multichannel extension contains some new tools for joint coding techniques.

3.2.1 MPEG-2 Advanced Audio Coding

In verification tests in early 1994 it was shown that introducing new coding algorithms and giving up backwards compatibility to MPEG-1 promised a significant improvement in coding efficiency (for the five channel case). As a result, a new work item was defined and led to the definition of MPEG-2 Advanced Audio Coding (AAC) ([4], see the description in [5]). AAC is a second generation audio coding scheme for generic coding of stereo and multichannel signals.

Figure 3 shows a generic block diagram of a typical AAC encoder. Comparing this to Layer-3, the most visible difference is the addition of a number of new blocks. AAC follows the same basic paradigm as Layer-3. AAC encoders often use the same double iteration loop structure as described for Layer-3. The difference is in a number of details and in the addition of more flexibility and more coding tools.

3.2.2 Tools to enhance coding efficiency

The following changes compared to Layer-3 help to get the same quality at lower bit-rates:

- **Higher frequency resolution:** The number of frequency lines in AAC is up to 1024 compared to 576 for Layer-3
- **Improved joint stereo coding:** Compared to Layer-3, both the mid/side coding and the intensity coding are more flexible, allowing to apply them to reduce the bit-rate more frequently.
- **Improved Huffman coding:** In AAC, coding by quadruples of frequency lines is applied more often. In addition, the assignment of Huffman code tables to coder partitions allows for many more options.

3.2.3 Tools to enhance audio quality

There are other improvements in AAC which help to retain high quality for classes of very difficult signals.

- **Enhanced block switching:** Instead of the hybrid (cascaded) filterbank in Layer-3, AAC uses a standard switched MDCT (Modified Discrete Cosine Transform) filterbank with an impulse response (for short blocks) of 5.3 ms at 48 kHz sampling frequency. This compares favourably with Layer-3 at 18.6 ms and reduces the amount of pre-echo artifacts (see below for an explanation).
- **Temporal Noise Shaping, TNS:** This technique does noise shaping in time domain by doing an open loop prediction in the frequency domain. TNS is a new technique which proves to be especially successful for the improvement of speech quality at low bit-rates.

With the sum of many small improvements, AAC reaches on average the same quality as Layer-3 at about 70 % of the bit-rate.

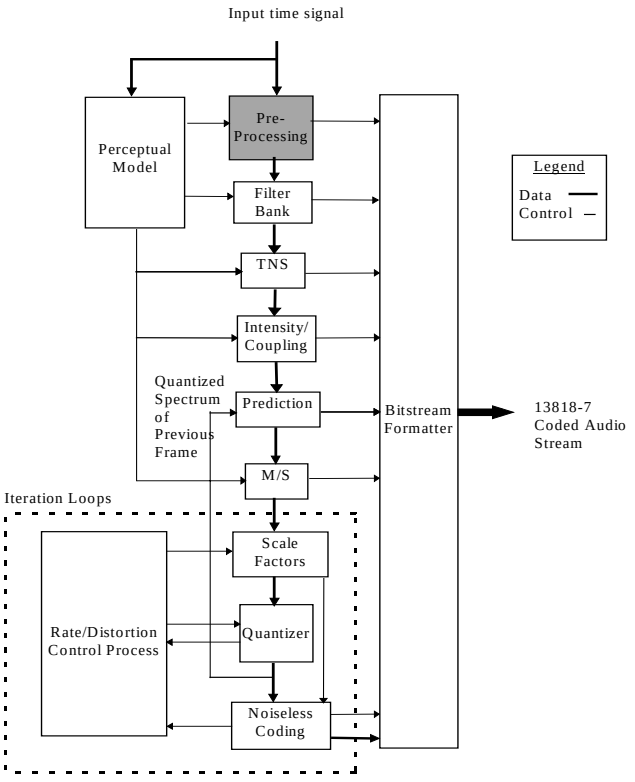

Fig 3: Block diagram of MPEG-2 Advanced Audio Coding

4. CANDIDATES FOR NEXT GENERATION CODECS

The basic paradigm of high frequency resolution audio codecs using variable length coding methods (e.g. entropy coding) has now been around for 14 years. The progress in enoding efficiency, best demonstrated by the advance in the state of the art from MPEG Layer-3 (MP3) to MPEG-2 Advanced Audio Coding (AAC), has recently somewhat slowed down. After three years, from all published tests results AAC is still the state-of-the-art encoding system if near transparent or transparent quality is desired.

Current work concentrates more on additonal flexibility (like the new MPEG-4 standard) or on lower bitrates. Among recent proposals, parametric coders (e.g. HILN, Harmonic, Individual Lines and Noise coding or the codec by QDesign) have shown the best promise to deliver nice sound quality at low bit-rates.

4.1. Other audio codecs proposed for Electronic Music Distribution:

The following audio codecs (mostly proprietary systems) have been named in the conjunction of future EMD systems. For most of them, no or nearly no independent data are available about the achievable audio quality. In independent tests done according to established test methods, none of them has shown an audio quality (at the same low bitrate) comparable to or surpassing the audio quality possible with the use of the MPEG-2 AAC format.

- Dolby AC-3 has been recommended by the ITU for the 5.1 multichannel sound of DTV systems.
- Lucent EPAC is an audio coding system similar to MPEG AAC. Wavelet based coding is used in addition to the MDCT filterbank.
- Sony ATRAC-3 is a recent system for EMD. The author is not aware of publications describing ATRAC-3 in detail.
- Microsoft WMA (Windows Media Audio) has been proposed for EMD, too.

4.2. Parametric coding:

One particularly interesting approach to high quality audio coding at very low bitrates is called "parametric coding". In parametric coding, instead of quantizing data directly representing the audio waveform (usually after transformation to a suitable target domain like the time-frequency representation given by an MDCT filterbank), data **describing** the signal are derived from the waveform. The decoding step consists of synthesizing a new waveform from these parameters. In MPEG-4, a system called HILN has found the way into the standard as a tool for scalable encoding at very low bitrates. HILN synthesizes audio from parameters on

- periodic components which are described by the way of pitch and harmonic content (H),
- Individual Lines (IL) describing additional frequency components and
- Noise (N) components which add up to describe the non-tonal parts of the signal.

A similar system taylored more to higher qualities (QDesign audio coding) is part of Quicktime Audio.

5. CONCLUSION AND FUTURE WORK

The current generation of high quality audio compression schemes delivers high quality audio from compressed signals at bit-rates of 128 kbit/s down to 64 kbit/s for a stereo signal. Currently, no techniques are known which could yield large improvements over these figures of merits. Current work on audio compression concentrates more on flexibility as needed for Internet multimedia or new multichannel applications than on improving on coding efficiency. The most interesting new work on audio compression is found in the area of music synthesis and hybrid natural coding / music synthesis. But even "traditional" audio compression has still many details to be solved and codec improvements, within the standards or as new systems, to be made. Improvements in the psychoacoustic model, in the encoding strategy or advances at points not considered today will certainly lead to better encoders in the future. There is hope that these improvements can be done within the constraints of todays coding standards, thus leading a compatible way into the future of sound reproduction.

References

[1] ISO/IEC IS 11172-3, Information technology -- Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s, Part 3: Audio, 1993

[2] K. Brandenburg, G. Stoll: ISO-MPEG-1 Audio: A Generic Standard for Coding of High Quality Digital Audio, in: Collected Papers on Digital Audio Bit-Rate Reduktion, N. Gilchrist and Chr. Grewin, ed., New York 1996, pp. 31 - 42

[3] ISO/IEC JTC1/SC29/WG11 MPEG, International Standard IS 13818-3 "Information Technology - Generic Coding of Moving Pictures and Associated Audio, Part 3: Audio".

[4] ISO/IEC IS 13818-7, Information technology -- Generic coding of moving pictures and associated audio information Part 7: Advanced Audio Coding (AAC), 1997

[5] K. Brandenburg and Marina Bosi, Overview of MPEG audio: Current and Future Standards for Low Bit-Rate Audio Coding, Journal of the Audio Engineering Society, Vol. 45, Jan/Feb 1997, pp. 4 - 21