An introduction to MPEG Layer-3

K. Brandenburg and H. Popp
Fraunhofer Institut für Integrierte Schaltungen (IIS)

MPEG Layer-3, otherwise known as MP3, has generated a phenomenal interest among Internet users, or at least among those who want to download highly-compressed digital audio files at near-CD quality.

This article provides an introduction to the work of the MPEG group which was, and still is, responsible for bringing this open (i.e. non-proprietary) compression standard to the forefront of Internet audio downloads.

1. Introduction

The audio coding scheme *MPEG Layer-3* will soon celebrate its 10th birthday, having been standardized in 1991. In its first years, the scheme was mainly used within DSP-based codecs for studio applications, allowing professionals to use ISDN phone lines as cost-effective music links with high sound quality. In 1995, MPEG Layer-3 was selected as the audio format for the digital satellite broadcasting system developed by World-Space. This was its first step into the mass market. Its second step soon followed, due to the use of the Internet for the electronic distribution of music. Here, the proliferation of audio material – coded with MPEG Layer-3 (aka MP3) – has shown an exponential growth since 1995. By early 1999, ".mp3" had become the most popular search term on the Web (according to http://www.searchterms.com). In 1998, the "MPMAN" (by Saehan Information Systems, South Korea) was the first portable MP3 player, pioneering the road for numerous other manufacturers of consumer electronics.

"MP3" has been featured in many articles, mostly on the business pages of newspapers and periodicals, due to its enormous impact on the recording industry. This article explains the basic technology and some of the special features of MPEG Layer-3. It also sheds some light on the factors which determine the quality of the coded audio signal.

2. Why MPEG Layer-3?

MPEG Layer-3 emerged as the main tool for Internet audio delivery. Considering the reasons, the following factors were definitely helpful.

Open standard

MPEG is defined as an open standard. The specification is available (for a fee) to everybody. While there are a number of patents covering MPEG Audio encoding and decoding, all patent-holders have declared that they will license the patents on fair and reasonable terms to everybody. Public example source code is available to help implementers to understand the text of the specification. As the format is well defined, no problems with interoperability of equipment or software from different vendors have been reported – except from some rare incomplete implementations.

Availability of encoders and decoders

DSP-based hardware and software encoders and decoders have been available for a number of years – driven at first by the demand for professional use in broadcasting.

Supporting technologies

While audio compression is viewed as a main enabling technology, other evolving technologies contributed to the MP3 boom, such as:

- ⇨ the widespread use of computer sound cards;
- ⇨ computers becoming powerful enough to run software audio decoders and even encoders in real-time;
- ⇨ fast Internet access for universities and businesses;
- ⇨ the availability of CD-ROM and CD-Audio writers.

In short, MPEG Layer-3 had the luck to be the right technology available at the right time. In the meantime, research on perceptual audio coding progressed, and codecs with better compression efficiency became available. Of these, MPEG-2 Advanced Audio Coding (AAC) was developed as the successor of MPEG-1 Audio. Other – proprietary – audio coding schemes were also introduced, claiming a higher performance than MP3.

3. MPEG audio coding standards

MPEG, a working group formally named as ISO/IEC JTC1/SC29/ WG11, but mostly known by its nickname, Moving Pictures Experts Group, was set up by the ISO/IEC standardization body in 1988 to develop generic (i.e. useful for different applications) standards for the coded representation of moving pictures, associated audio and their combination. Since then, MPEG has undertaken the standardization of compression techniques for video and audio. Originally, its main goal was video coding together with audio coding for digital storage media. In the meantime, the MPEG audio coding standard found its way into many different applications, including

- ⇨ digital audio broadcasting (Eureka-147 DAB, WorldSpace, ARIB, DRM);
- ⇨ ISDN transmission for broadcast contribution and distribution purposes and for commentary links;
- ⇨ archival storage within broadcasting;
- ⇨ sound for digital television (DVB, Video CD, ARIB);
- ⇨ internet streaming (Microsoft Netshow, Apple Quicktime);
- ⇨ portable audio devices (mpman, mplayer3, Rio, Lyra, YEPP and many more);
- ⇨ storage and exchange of music files on computers.

The most widely-used audio compression formats are MPEG Audio Layer-2 and Layer-3 (see below for definitions) and Dolby AC-3. A large number of systems currently under development will use AAC.

MPEG-1

MPEG-1 is the name for the first phase of MPEG work, starting in 1988. This work was finalized with the adoption of the ISO/IEC standard IS 11172 in late 1992. The audio coding part of this standard (IS 11172-3) describes a generic coding system, designed to fit the demands of many applications. MPEG-1 Audio consists of three operating modes

Abbreviations

AAC	(MPEG-4) Advanced Audio Coding
ACTS	Advanced Communications Technologies and Services
ARIB	Association of Radio Industries and Businesses (Japan)
DAB	Digital Audio Broadcasting
DRM	Digital Radio Mondiale
DSP	Digital signal processor / processing
DVB	Digital Video Broadcasting
HDTV	High-definition television
IEC	International Electrotechnical Commission
ISDN	Integrated services digital network
ISO	International Organization for Standardization
JTC	Joint Technical Committee
MDCT	Modified discrete cosine transform
MPEG	Moving Picture Experts Group
PEAQ	Perceptual evaluation of audio quality
RACE	R&D in Advanced Communications technologies in Europe

called “Layers”, with increasing complexity and performance, named Layer-1, Layer-2 and Layer-3. Layer-3, with the highest complexity, was designed to provide the highest sound quality at low bit-rates (around 128 kbit/s for a typical stereo signal).

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 MPEG-2 Audio standard (IS 13818-3) was finalized in 1994 and consisted of two extensions to MPEG-1 Audio:

- ⇨ Multichannel audio coding, including the 5.1 channel configuration well known from cinema sound – this multichannel extension is done in a backward compatible way, allowing MPEG-1 stereo decoders to reproduce a mixture of all available channels.
- ⇨ Coding at lower sampling frequencies – this extension adds sampling frequencies of 16 kHz, 22.05 kHz and 24 kHz to the MPEG-1 sampling frequencies of 32 kHz, 44.1 kHz and 48 kHz, improving the coding efficiency at very low bit-rates.

MPEG-2 AAC

In early 1994, verification tests showed that new coding algorithms, without backward compatibility to MPEG-1, promised a significant improvement in coding efficiency. As a result, a new work item was defined that finally led to the definition of a new MPEG audio coding standard, MPEG-2 Advanced Audio Coding (AAC). The standard was finalized in 1997 (IS 13818-7). AAC is a second-generation audio coding scheme for generic coding of stereo and multichannel signals, supporting sampling frequencies from 8 kHz to 96 kHz and a number of audio channels ranging from 1 to 48.

MPEG-3

Originally, MPEG had a plan to define the video coding for HDTV applications in a further phase, to be called MPEG-3. However, later on it turned out that the tools developed for MPEG-2 video coding would also address the HDTV requirements, and MPEG gave up the plan to develop a special MPEG-3 standard. Sometimes MPEG Layer-3 (or “MP3”) is misnamed MPEG-3.

MPEG-4

MPEG-4 intends to become the next major standard in the world of multimedia. The first version was finished in late 1998 (IS 14496-3), and the second version at the end of

1999. Unlike MPEG-1 and MPEG-2, the emphasis in MPEG-4 is on new functionalities rather than better compression efficiency. Mobile as well as stationary user terminals, database access, communications and new types of interactive services will be major applications for MPEG-4. The new standard facilitates the growing interaction and overlap between the hitherto separate worlds of computing, electronic mass media (TV and Radio) and telecommunications. MPEG-4 audio consists of a family of audio coding algorithms – spanning the range from low bit-rate speech coding (down to 2 kbit/s) up to high-quality audio coding at 64 kbit/s per channel and above. Generic audio coding at medium to high bit-rates is done by AAC.

MPEG-7

Unlike MPEG-1, MPEG-2 and MPEG-4, MPEG-7 does not define compression algorithms. MPEG-7 is a content representation standard for multimedia information search, filtering, management and processing. MPEG-7 will be approved by July, 2001.

4. MPEG Layer-3 audio encoding

The following description of MPEG Layer-3 encoding focuses on the basic functions and a number of details necessary to understand the implications of encoding options on the sound quality. It is not meant to be a complete description of how to build an MPEG Layer-3 encoder.

Flexibility

In order to be applicable to a number of very different application scenarios, MPEG defined a data representation including a number of options.

Operating mode

MPEG-1 Audio works for both mono and stereo signals. A technique called joint stereo coding can be used to achieve a more efficient combined coding of the left and right channels of a stereophonic audio signal. Layer-3 allows both mid/side stereo coding and intensity stereo coding. The latter is especially helpful for lower bit-rates, but bears the risk of changing the sound image. The operating modes are:

- ⇨ single channel;
- ⇨ dual channel (two independent channels, for example containing different language versions of the audio);
- ⇨ stereo (no joint stereo coding);
- ⇨ joint stereo.

Sampling frequency

MPEG audio compression works on a number of different sampling frequencies. MPEG-1 defines audio compression at 32 kHz, 44.1 kHz and 48 kHz. MPEG-2 extends this to half the rates, i.e. 16 kHz, 22.05 and 24 kHz. “MPEG-2.5” is the name of a proprietary extension to Layer-3, developed by Fraunhofer IIS, which introduces the sampling frequencies 8 kHz, 11.05 kHz and 12 kHz.

Bit-rate

MPEG Audio does not just work at a fixed compression ratio. The selection of the bit-rate of the compressed audio is, within some limits, completely left to the implementer or operator of an MPEG audio coder. For Layer-3, the standard defines a range of bit-rates from 8 kbit/s up to 320 kbit/s. Furthermore, Layer-3 decoders must support the switching of bit-rates from audio frame to audio frame. Combined with the bit reservoir technology, this allows both variable bit-rate coding and constant bit-rate coding at any fixed value within the limits set by the standard.

Normative versus Informative

A very important property of the MPEG standards is the principle of minimizing the amount of normative elements in the standard. In the case of MPEG Audio, this led to the fact that only the data representation, i.e. the format of the compressed audio, and the decoder are normative.

Decoder considerations

Even the decoder is not specified in a bit-exact fashion. Instead, formulae are given for most parts of the algorithm, and compliance is defined by a maximum deviation of the decoded signal from a reference decoder, implementing the formulae with double-precision arithmetic accuracy. This allows us to build decoders running both on floating-point and fixed-point architectures. Depending on the skills of the implementers, fully-compliant high-accuracy Layer-3 decoders can be constructed with down to 20-bit arithmetic wordlength, without using double-precision calculations.

Encoder considerations

Encoding of MPEG Audio is completely left to the implementer of the standard. As a helpful guide-line, the ISO standards contain the description of example encoders. While these descriptions were derived from the original encoders used for verification tests, a lot of experience and knowledge is necessary to implement good-quality MPEG audio encoders. The amount of investment necessary to engineer a high-quality MPEG audio encoder has kept the number of independently-developed encoder implementations very low.

5. MPEG Layer-3 – the algorithm

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

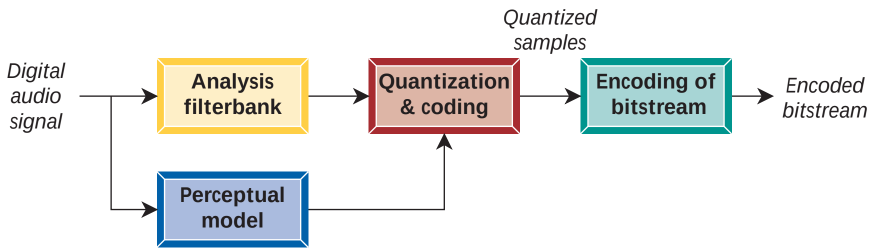

Figure 1
Typical MPEG Layer-3 encoder

Filterbank

The filterbank used in MPEG Layer-3 belongs to the class of hybrid filterbanks. It is built by cascading two different kinds of filterbanks, first a polyphase filterbank (as used in Layer-1 and Layer-2) and second, a Modified Discrete Cosine Transform (MDCT) filterbank. The polyphase filterbank fulfils the purpose of making Layer-3 more similar to Layer-1 and Layer-2. The subdivision of each polyphase frequency band into 18 finer sub-bands increases the potential for redundancy removal, leading to better coding efficiency for tonal signals. As a further positive result of the higher frequency resolution, the error signal can be better controlled, allowing a finer tracking of the masking threshold. The filterbank can be switched to a lower frequency resolution to avoid pre-echoes (see below).

Perceptual model

The perceptual model mainly determines the quality of a given encoder implementation. Since the original informative part in the standard was written, a lot of additional work has gone into this part of the encoder . 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 the 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.

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 and, in the case of very small numbers to be coded, in 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 because noise shaping has to be done to keep the quantization noise below the masking threshold, a global gain value (which determines the quantization step size) and scalefactors (which determine the 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).

While the inner iteration loop always converges (if necessary, by setting the quantization step size large enough to zero out all spectral values), this is not true for the combination of both iteration loops. If the perceptual model requires quantization step sizes so small that the rate loop always has to increase them to enable coding at the required bit-rate, both can go on forever. To avoid this situation, several conditions can be checked to stop the iterations more early. However, for fast encoding and good coding results, such a condition should be avoided. This is one reason why an MPEG Layer-3 encoder usually needs tuning of the parameter sets of the perceptual model for each bit-rate.

6. Quality Considerations

As explained above, the pure compliance of an encoder with an MPEG audio standard does not guarantee any quality of the compressed music. Audio quality differs between different items, depending on basic parameters including, of course, the bit-rate of the compressed audio and the sophistication of different encoders, even if they work with the same set of basic parameters. To gain more insight into the level of quality possible with MP3 and AAC, let us first have a look at typical artefacts associated with perceptual audio coders.

Common types of artefacts

Unlike analogue hi-fi equipment or FM broadcasting, perceptual encoders exhibit sound deficiencies when run at too low bit-rates or with the wrong parameters. These so-called “artefacts” are in most cases different from usual noise or distortion signals. Perceptual audio coding schemes such as MPEG Layer-3 introduce an error signal that can be described as a time-varying error at certain frequencies, which is not constrained to the harmonics of the music signal. The resulting music signal may sound:

- distorted, but not like harmonic distortions;
- noisy, but with the noise introduced only in a certain frequency range;
- rough, with the roughness often being very objectionable as the error may change its characteristics about every 24 ms.

Loss of bandwidth

If an encoder does not find a way to encode a block of music data with the required fidelity within the limits of the available bit-rate, it “runs out of bits”. This may lead to the deletion of some frequency lines, typically affecting the high-frequency content. Compared to a constant bandwidth reduction, such an effect becomes more objectionable if the effective bandwidth changes frame-by-frame (e.g. every 24 ms).

Pre-echoes

Pre-echoes are very common artefacts, in the case of perceptual audio coding schemes using high-frequency resolution. The name "pre-echo", although somewhat misleading, nicely describes the artefact, which is a noise signal occurring even before the music event that causes such noise.

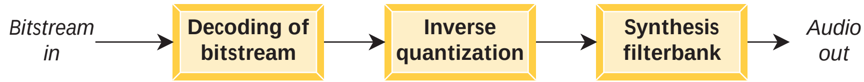

Figure 2
Decoder of a perceptual coding system

To understand the origin of pre-echoes, let us consider the decoder of a perceptual coding system (see *Fig. 2*). The reconstructed frequency lines are combined by the synthesis filterbank, consisting of a modulation matrix and a synthesis window. The quantization error introduced by the encoder can be seen as a signal added to the original frequency lines, with a length in time that is equal to the length of the synthesis window. Thus, reconstruction errors are spread over the full window length. If the music signal contains a sudden increase in signal energy (e.g. a castanet attack), the quantization error is increased as well. If such an attack occurs well within the analysis window, its error signal will be spread within the full synthesis window, preceding the actual cause for its existence in time. If such a pre-noise signal extends beyond the pre-masking period of the human ear, it becomes audible and is called "pre-echo".

There are a number of techniques to avoid audible pre-echoes, including variable bit-rate coding or a local increase in the bit-rate to reduce the amplitude of the pre-echo. In general, these artefacts belong to the "most difficult to avoid" category.

Roughness, "double-speak"

Especially at low bit-rates and low sampling frequencies, there is a mismatch between time resolution of the coder and the time structure of some signals. This effect is most noticeable on speech signals and when listening via headphones. As a single voice tends to sound like it has been recorded twice and then overlaid, this effect is sometimes called "double-speak".

7. Not all encoders are created equal

The MPEG standards do not prescribe the implementation of the audio encoder. In an extreme case, one could completely avoid implementing the perceptual model, decide not to use the scalefactors (and therefore the outer iteration loop), and do a very simple inner iteration loop. Such an encoder would be very fast (potentially much faster than any current encoder products), would be compliant with the standard, would even produce a nice audio quality for some signals, but would sound very bad with a large selection of music. While such a project would be easy to implement, it is much more difficult to build an encoder that offers very high audio quality across all types of music and even with the most exotic test items. In MPEG, testing had always aimed to verify sufficient encoder performance in worst-case scenarios. Nonetheless, the current MP3 encoders show remarkable differences in their ability to produce, in a consistent way, high sound quality at low bit-rates.

7.1. How to measure sound quality

Measuring the sound quality of perceptual audio codecs has developed into an art of its own, over the last ten years. Basically, there are three methods: Listening tests, simple objective measurement methods and perceptual measurement techniques.

Listening tests

To date, large-scale and well-controlled listening tests are still the only method available for comparing the performance of different coding algorithms and different encoders. With input from a number of broadcasters and the MPEG audio group, the ITU-R has developed a very elaborate set of rules for listening tests. These tests aim to stress the encoders under worst-case conditions, i.e. the testers try to find sound material which is most difficult to encode, and then evaluate the performance of the encoders under test for this material. This procedure is based on the observation that, in a lot of cases, coding artefacts become audible and even objectionable only after an extensive training period. Since the use of equipment based on audio compression technology (such as portable audio players) itself constitutes extensive training, everybody can become an expert listener over a period of time. Therefore, right from the beginning, encoders should be tuned better to satisfy the quality requirements of expert listeners.

Simple objective measurement techniques

Over and over again, people tried to get a measure of encoder quality by looking at parameters such as the signal-to-noise-ratio or bandwidth of the decoded signal. As the basic paradigm of perceptual audio coders relies on improving the subjective quality – by shaping the quantization noise over frequency (and time), leading to an SNR which is lower than is possible without noise shaping – these measurements defy the whole purpose of perceptual coding. As explained below, to rely on the bandwidth of the encoded signal does not show a very good understanding of the subject. Another approach is to look at the codec output for certain test signal inputs, such as transients or multi-tone signals. While the results of such a test may tell the expert a lot about the codec under test, it is very dangerous to rely solely on such results.

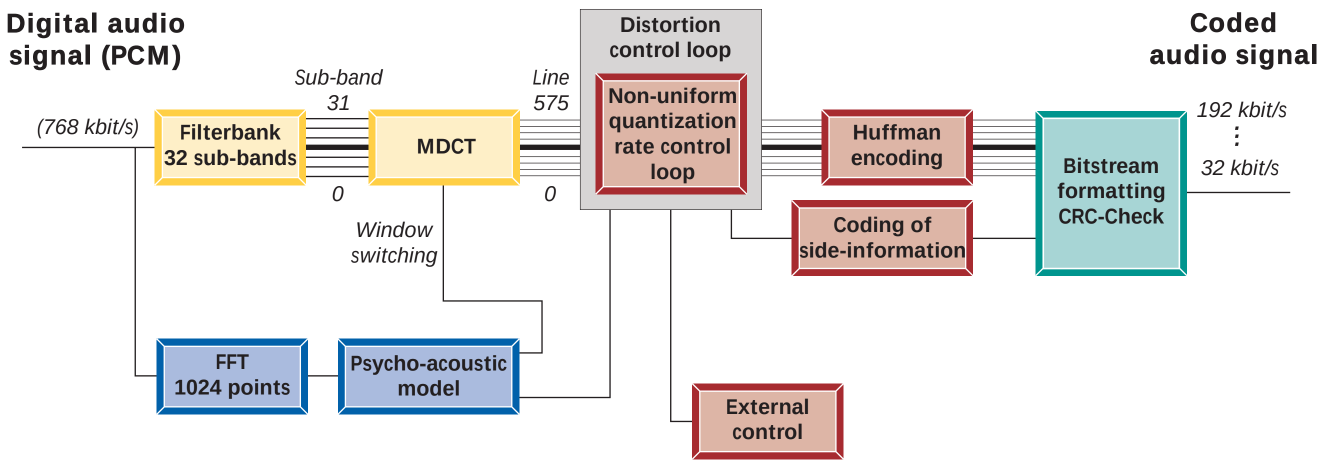

Figure 3
An MPEG Layer-3 encoder

Perceptual measurement techniques

For 15 years, there has been a lot of research into applying psycho-acoustic modelling to the prediction of sound quality and the audibility of certain artefacts. While the state of the art is not yet sufficient to make large-scale and well-prepared listening tests obsolete, perceptual measurement techniques have progressed to the point where they are a very useful supplement to listening tests and, in some cases, are already replacing them. The ITU-R Task Group 10/4 worked on the standardization of perceptual measurement techniques and finally produced a Recommendation for a system called PEAQ (Perceptual Evaluation of Audio Quality). This Recommendation defines a multi-mode system based on the collaborative efforts of all the leading laboratories working on perceptual measurement techniques.

7.2. The bandwidth myth

Reports about encoder testing often mention the bandwidth of the compressed audio signal. In a lot of cases this is due to misunderstandings about human hearing on the one hand and encoding strategies on the other hand.

Hearing at high frequencies

It is certainly true that a large number of (especially young) subjects are perfectly able to hear single sounds at frequencies up to and sometimes well above 20 kHz. However, contrary to popular belief, the authors are not aware of any scientific experiment showing beyond doubt that there is any listener (trained or not) who may detect the difference between a (complex) music signal with content up to 20 kHz, and the same signal with a bandlimit of around 16 kHz. There are some hints that a few listeners may have such capabilities, but the full scientific proof has not yet been given. Therefore, it is a reasonable strategy to limit the frequency response of an MP3 encoder to 16 kHz. Due to the brick-wall characteristic of the filters used in the codec, this can be done easily.

Please note, however, that this is not a general rule which can be applied to other types of audio equipment (in particular, analogue). Typical audio equipment has to support much higher frequencies in order to have the required perfectly flat frequency response up to 16 kHz: any deviation from the ideal straight line below the frequency cut-off point is very audible.

Encoding strategies

While a loss of audio HF response will produce a coding artefact, it does not necessarily mean that an encoder which produces a higher audio bandwidth will sound any better. There is, in fact, a basic trade-off depending on how the available bits are used. If they are used to improve the frequency response, they are no longer available to produce a clean sound at lower frequencies. Leaving this trade-off to the encoder often leads to a poor audio signal, with the high frequency cut-off point varying from frame to frame. According to the current state of the art, it is best to introduce a fixed bandwidth limitation.

Technically, MP3 can reproduce signal content up to the limit given by the actual sampling frequency. If there is a comparison, at the same bit-rate, between an encoder with a fixed limited frequency response, and another encoder with a much larger bandwidth, experience tells us that in most cases the encoder with the lower bandwidth produces better sounding compressed audio. However, there is a limit to this statement: at low bit-rates (64 kbit/s for stereo and lower), the question of the best trade-off in terms of bandwidth versus cleanness of sound is a hotly-contested question of taste. We have found that even trained listeners sometimes completely disagree about the bandwidth a given encoder should be run at.

7.3. Tuning for different bit-rates

As explained above, the double iteration loops do not converge if there is a mismatch between the coding requirements as given by the perceptual model and the bit-rate available to code a block of music. To avoid this situation, it is wise to set the parameters in the psycho-acoustic model in a way that the iteration loops will normally converge. This may require settings which lead to audible differences, but the final coding result is still better than the one from a perceptual model set to avoid any audible differences, combined with coding loops which do not converge in a sensible way. To achieve this balance between requirements from the perceptual model and the available bit-rate, the coding parameters have to be readjusted if the encoder is run at different bit-rates. Such tuning procedures are responsible for a large part of the effort being put towards developing an MP3 encoder.

Karlheinz Brandenburg received M.S. (Diplom) degrees in Electrical Engineering (1980) and in Mathematics (1982) from Erlangen University, Germany. In 1989, he gained a Ph.D. in Electrical Engineering, also from Erlangen University, for work on digital audiocoding and perceptual measurement techniques. From 1989 to 1990, he was with AT&T Bell Laboratories in Murray Hill, NJ, USA, where he worked on the ASPEC perceptual coding technique and on the definition of the ISO/IEC MPEG Layer-3 system. In 1990, he returned to Erlangen University to continue his research on audio coding and to teach a course on digital audio technology.

In 1993 Dr Brandenburg became head of the Audio/Multimedia department at the Fraunhofer Institute for Integrated Circuits (FhG-IIS). Since 2000, he has been teaching at Ilmenau Technical University and has become the director of the new Fraunhofer Group for Electronic Media Technologies in Ilmenau, Germany.

Karlheinz Brandenburg has received three awards from the AES for his work on perceptual audio coding and psycho-acoustics. He is a member of the technical committee on Audio and Electro-acoustics of the IEEE Signal Processing Society. He has worked within the MPEG-Audio committee since its beginnings in 1988. In recent years, he worked on MPEG-2 Advanced Audio Coding (standardized in 1997) and helped to organize the work for MPEG-4 Audio. He is a member of SDMI (the Secure Digital Music Initiative) and currently chairs the AES Standards Committee working group AESSC-06-04 on Internet Audio Delivery Systems. He has been granted 25 patents and has several more pending.

Harald Popp was born in Erlangen, Germany, in 1956. In 1981, he received an M.S. (Diplom) in Electrical Engineering from Erlangen University. From 1982 to 1984, he continued his work at Erlangen University and carried out a technology transfer project for advanced cable fault location. In 1984, he joined the *Fraunhofer Institut Integrierte Schaltungen (IIS)* at its inauguration and worked as a board-level hardware designer in various industrial projects. From 1987, he was responsible for the real-time audio coding systems of the IIS (LC-ATC, OCF, ASPEC, MPEG-Layer-3).

Today, Mr Popp is head of the Studio Department, focusing on effective DSP-based real-time implementations of audio and video coding schemes.

8. Conclusions

By using an encoder with good performance, both MPEG Layer-3 and MPEG-2 Advanced Audio Coding (AAC) can significantly compress music signals, while still maintaining CD or near-CD quality. Between the two systems, Layer-3 – with somewhat lower complexity – is the system of choice for current near-CD quality applications. AAC is its designated successor, providing near-CD quality at even larger compression rates, and enabling higher quality encoding and playback up to high definition audio at 96 kHz sampling rate.

Acknowledgements

The authors would like to thank all their colleagues at Fraunhofer IIS and MPEG Audio for all the wonderful collaborative work over the last 11 years, since the start of MPEG. Special thanks are due to Jürgen Herre, Martin Dietz, Oliver Kunz and Jürgen Koller for their helpful suggestions. It should be noted that part of the audio coding work at Fraunhofer IIS-A has been supported by the Bavarian Ministry for Economy, Transportation and Technology, and also the European Commission (within the RACE and ACTS programmes)

Bibliography

[1] K. Brandenburg, G. Stoll: **ISO-MPEG-1 Audio: A Generic Standard for Coding of High Quality Digital Audio**
"Collected Papers on Digital Audio Bit-Rate Reduction (Gilchrist and Grewin), AES, 1996.

[2] M. Dietz, H. Popp, K. Brandenburg and R. Friedrich: **Audio Compression for Network Transmission**
Journal of the AES, Vol. 44, No. 1-2, 1996.