# DIGITAL AUDIO: FROM LOSSLESS TO TRANSPARENT CODING

**Peter Noll** and **Tilman Liebchen**

Technische Universität Berlin
Einsteinufer 25, 10587 Berlin, Germany
Phone: +49 30 3142 3326; Fax: +49 30 3142 2514
Email: noll@ee.tu-berlin.de

**Abstract:** We have seen rapid progress in high-quality compression of wideband audio signals. Today's coding algorithms can achieve substantially more compression than was thought possible only a few years ago. In the case of audio coding with its bandwidth of 20 kHz and more, the concept of perceptual coding has paved the way for significant bit rate reductions.

However, multiple codings can reveal originally masked distortions. In addition, reproduction of critical music items shows that even the best systems can not be considered as truly transparent. Therefore lossless audio coding has become a topic of high interest both for professional and customer applications.

This paper will explain approaches to lossless and lossy compressions, in the latter case with emphasis on existing MPEG standards which have found a wide range of communications-based and storage-based applications. It will be shown that the very recent MPEG-2 *Advanced Audio Coding* (AAC) standard outperforms many other coding algorithms (including MPEG-1 coders). Finally, we will address the current MPEG-4 speech and audio coding standardization work which merges the whole range of audio from high fidelity audio coding and speech coding down to synthetic audio, synthetic speech and text-to-speech conversion.

## 1 Introduction

Wideband (high fidelity) audio representations including multichannel audio need bandwidths of at least 20 kHz. The conventional digital format of digital audio is PCM, with sampling rates of 32, 44.1, or 48 kHz and an amplitude resolution (PCM bits per sample) of 16 bit. Typical application areas for digital audio are in the fields of audio production, program distribution and exchange, digital sound broadcasting, digital storage, and various multimedia applications.

For archiving and processing of audio signals, highest quality formats with 96 kHz sampling and 24 to 30 bit amplitude resolution are under discussion. In some applications coding will have to be *lossless* - with compression factors around two to three as will be shown shortly. Upcoming Digital versatile Discs (DVD) with their capacity of 4,7 GB (single layer) or 8.5 GB (double layer) will be the appropriate storage devices for lossless-coded audio material. The capacity can be doubled if both sides of the discs are readable.

The *Compact Disc* (CD) is today's *de facto standard* of digital audio *representation*. On a CD with its 44.1 kHz sampling rate the resulting stereo net bit rate is 2 x 44.1 x 16 x 1000 ≡ 1.41 Mb/s (see Table 1). However, the CD needs a significant overhead resulting in a 49-bit representation of each 16-bit audio sample. Hence the total stereo bit rate is 1.41 x 49/16 = 4.32 Mb/s. Table 1 compares parameters of the Compact Disc and the *Digital Audio Tape* (DAT) with those of two more recent storage systems, Philips *Digital Compact Cassette (DCC)* and Sony's 64 mm optical or magneto-optical *MiniDisc* (MD), and with the parameters of the European Digital Audio Broadcast (DAB).

| Applications | Format | Sampling rate | Audio bit rate | Overhead bit rate | Total bit rate |
|---|---|---|---|---|---|
| **Compact Disc (CD)** | PCM | 44.1 kHz | 1.41 Mb/s | 2.91 Mb/s | 4.32 Mb/s |
| **Digital Tape (DAT)** | PCM | 44.1 kHz | 1.41 Mb/s | 1.67 Mb/s | 3.08 Mb/s |
| **Digital Cassette (DCC)** | MPEG-1 Layer I | 48 kHz | 384 kb/s | 384 kb/s | 768 kb/s |
| **MiniDisc (MD)** | ATRAC | 44.1 kHz | 292 kb/s | 718 kb/s | 1.01 Mb/s |
| **Digital Broadcast** | MPEG-1 Layer II, III | 48 kHz | 256 kb/s | 256 kb/s | 512 kb/s |

Table 1. Bit rates for various digital audio schemes (Stereophonic signals)

## 2 Lossless audio coding

Entropy coding methods as Lempel-Ziv, Huffman or arithmetic coding when applied directly to the audio signal are not very efficient because of the long-time correlations in the audio signal. Therefore, conventional *text* compression tools fail in the case of digital audio data. A preprocessing stage, which eliminates the statistical dependencies within the signal, leads to an almost uncorrelated source which is easier to code. Such decorrelation can either be achieved by linear prediction or by linear transforms [1]. The samples of the difference signal or the transform coefficients have to be quantized which causes inevitable errors in the output signal. Therefore, lossless coding has to be considered as a combination of conventional lossy coding and an additional transmission of the coding error. Fig. 1 shows the principle of a lossless coding scheme. The signal code $c$ is derived from an integer-valued input signal $x$, such as given by the 16-bit format of a CD, using a lossy compression algorithm. To achieve lossless compression, the difference $e$ between the input signal $x$ and the reconstructed signal $y$ is generated in the encoder by local decoding. Both the lossy compressed code $c$ and the error $e$, which serves as a correction signal in the decoding process, are transmitted digitally. In the decoder, the error signal $e$ is added to the decoded approximation, resulting in an output signal which is a perfectly reconstructed version of the input signal.

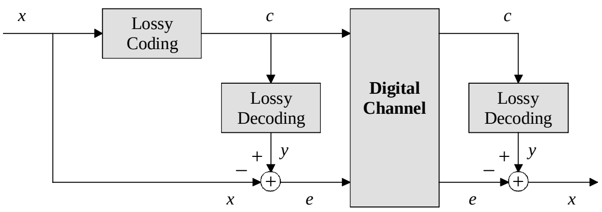


Fig. 1. Principle of lossless coding

It is obvious that the error signal depends on the lossy coding algorithm. If the compression ratio is high, the error signal $e = x - y$ will be large and correlated, and the coding problem passes from the original signal to the error signal. On the other hand, if the compression ratio is small, the error will be close to zero, but then the signal redundancy is not removed sufficiently. Thus, as a main goal in lossless coding, we have to find the best compromise between a high compression ratio in the lossy branch and an easily codeable signal in the correction branch.

As an example we consider a *lossless transform coding* system as it is shown in Fig. 2. From the quantized input signal $x(n) \in \mathbf{Z}$ a set of transform coefficients $t(k)$ is calculated using an arbitrary orthonormal transform $\mathbf{T}$ of block length $M$. The coefficients are scaled by $\alpha$ and quantized with an unitary quantization step size $\Delta = 1$, leading to an integer-valued transform coefficients $c(k)$ which are then entropy-coded and transmitted. Using a suitable transform, many of the coefficients are very small or even zero, and, moreover, they constitute an uncorrelated source. Hence, the integer-valued spectrum can be easily entropy-coded without taking into account joint probabilities.

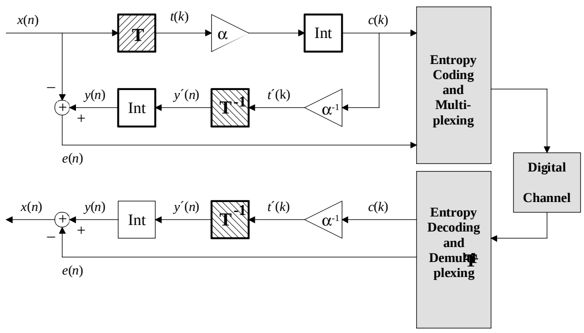


Fig. 2. Transform-based lossless coding

As a result of the quantization, which is equivalent to integer rounding, decoding of $c(k)$ does not lead to perfect reconstruction, although the integer spectrum itself is coded losslessly. Therefore, we have to check for possible errors by locally decoding $c(k)$ and by computing an error correction signal, if necessary.

After descaling with $\alpha^{-1}$ and applying the inverse transform $\mathbf{T}^{-1} = \mathbf{T}^{t}$, where $\mathbf{T}^{t}$ stands for the transposed matrix, we obtain a real-valued signal $y'(n)$ which, due to the quantization in the transform domain, is not the original $x(n)$. However, since we have assumed integer-valued input samples $x(n)$, there is no need to reconstruct the input signal exactly by a real-valued signal. If $|y'(n) - x(n)| < 0.5$, integer rounding of $y'(n)$ leads to a reconstructed integer signal $y(n)$ which is identical with $x(n)$. Otherwise, the input signal is not perfectly reconstructed, i.e. we have an error $e(n) \neq 0$. Of course, $e(n) = x(n) - y(n)$ is an integer signal, because $x(n)$ and $y(n)$ are integer values as well. The error signal $e(n)$ has to be transmitted in addition to the coefficients $c(k)$.

On the decoder side, $y(n)$ is calculated identically by descaling, inverse transformation and integer rounding. After adding $e(n)$, we have $y(n) + e(n) = y(n) + (x(n) - y(n)) = x(n)$, which is the desired original input signal.

## 3 Results of lossless coding

We have implemented a lossless transform coding system using an orthonormal DCT with variable block length [2]. Each integer-valued spectrum $c(k)$ is divided into groups of 32 adjacent coefficients. This partition proved to be most efficient. Since these groups have an almost Laplacian PDF, the codebook consists of several Rice codes [3]. Each group is coded using the code which leads to a minimum number of bits. A Rice code is in fact a Huffman code for a Laplacian PDF, which is determined by its standard deviation $\sigma$. Since Rice codes only exist for discrete values of $\sigma$, only the indices of the chosen codes have to be transmitted.

The inverse transform allows for the generation of the error signal, which is encoded using an arithmetic coder with a static model.

The input signal $x(n)$ is divided into blocks of $M$ samples. Each combination of $M$-, $M/2$- and $M/4$-point transforms is calculated, and for the corresponding block the most suitable combination is finally selected. The decoder applies the inverse transform after decoding of the coefficients. By adding the decoded error signal, the original signal is perfectly reconstructed. The computational complexity of the decoder is about half of that of the encoder, because only one transformation has to be performed.

In general, a higher block length leads to better coding results due to a better decorrelation of the source signal, except for signals with very fast varying statistics, pitched signals like speech or very transient signals like castanets. The adaptive block length mode finds the appropriate block length for each block.

Table 2 compares the results of our Lossless Transform Audio Compression (LTAC) algorithm with those obtained by the popular linear predictive coding algorithm "Shorten" [4], used with default parameters (polynomial prediction) combined with the -c 2 option for stereo files.

For most categories, the results of LTAC are significantly better than those of “Shorten” for the audio material of the SQAM disc [5]. The average value is 4.56 bit per sample (bps) instead of 4.98 bps. Other lossless audio coders based on linear prediction are able to achieve 4.83 bps [6] and 4.68 bps [7] for the same audio material.

| **Category** | **Shorten** | **LTAC** |
|---|---|---|
| | **(Public Domain)** | **TU Berlin** |
| Alignment signals | 6.29 | 6.06 |
| Artificial signals | 2.84 | 2.60 |
| Single instruments | 4.60 | 4.15 |
| Vocal | 5.35 | 4.83 |
| Speech | 5.45 | 5.36 |
| Solo instruments | 5.09 | 4.52 |
| Vocal & Orchestra | 6.73 | 6.14 |
| Orchestra | 5.33 | 5.07 |
| Pop Music | 6.37 | 6.03 |
| **Average** | **4.98** | **4.56** |

Table 2. Coding results (bits per sample) for the audio material of the SQAM disc [5]. The categories are based on the according SQAM sections. Maximum adaptive block length: $M = 4096$. LTAC version 1.61.

## 4 Lossy but subjectively transparent audio coding

### Speech and Audio Coding

First proposals to reduce wideband audio coding rates have followed those for speech coding [1]. Speech and audio coding are similar in that in both cases quality is based on the properties of human auditory perception. However, speech can be coded very efficiently because a speech production model is available, whereas nothing similar exists for audio signals (see Figure 3).

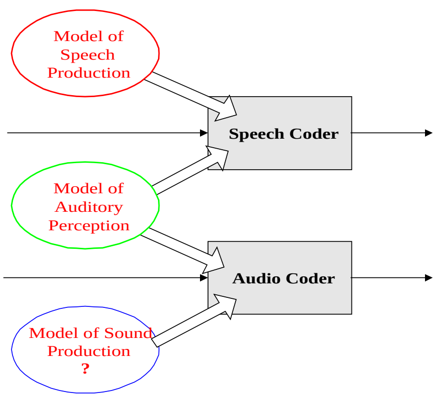


Fig. 3. Efficient speech and audio compression by employing models of perception and production

### Quality Measures

As a measure of quality, the most popular subjective assessment method is the mean opinion scoring where subjects classify the quality of coders on an N-point quality scale. The final result of such tests is an averaged judgement called the *mean opinion score (MOS)*. Two 5-point adjectival grading scales are in use, one for signal *quality*, and the other one for signal *impairment*, and an associated numbering [1]. The 5-point ITU-R impairment scale of Table 3 is extremely useful if coders with only small impairments have to be graded.

| Mean opinion score | Impairment scale |
|---|---|
| 5 | Imperceptible |
| 4 | Perceptible, but not annoying |
| 3 | Slightly annoying |
| 2 | Annoying |
| 1 | Very annoying |

Table 3. 5-point MOS impairment scale

### Auditory Perception

The inner ear performs short-term critical band analyses where frequency-to-place transformations occur along the basilar membrane. The power spectra are not represented on a linear frequency scale but on limited frequency bands called *critical bands*. The auditory system can roughly be described as a bandpass filterbank, consisting of strongly overlapping bandpass filters with bandwidths in the order of 100 Hz for signals below 500 Hz and up to 5000 Hz for signals at high frequencies. Twenty-five critical bands covering frequencies of up to 20 kHz have to be taken into account.

*Simultaneous masking* is a frequency domain phenomenon where a low-level signal (the maskee) can be made inaudible (masked) by a simultaneously occurring stronger signal (the masker), if masker and maskee are close enough to each other in frequency [8]. A *masking threshold* can be measured below which the low-level signal will not be audible. This masked signal can consist of low-level signal contributions, of quantization noise, aliasing distortion, or of transmission errors. The masking threshold varies with time. It depends on the sound pressure level (SPL), the frequency of the masker, and on the characteristics of masker and maskee. Take the example of the masking threshold for the SPL = 60 dB narrowband masker in Figure 4: around 1 kHz the five maskees (one of which is hidden behind the masker) will be masked as long as their individual sound pressure levels are below the masking threshold. The slope of the masking threshold is steeper towards lower frequencies, i.e. higher frequencies are easier masked. It should be noted that the distance between masker and masking threshold is smaller in noise-masking-tone experiments than in tone-masking-noise experiments, i.e., noise is a better masker than a tone. Without a masker, a signal is inaudible if its sound pressure level is below the *threshold in quiet* which depends on frequency and covers a dynamic range of more than 60 dB as shown in the lower curve of Figure 4. The distance between the level of the masker and the masking threshold is called *signal-to-mask ratio* (SMR). Within a critical band, coding noise will not be audible as long as its signal-to-noise ratio SNR(m), the signal-to-noise ratio resulting from an m-bit quantization, is higher than its SMR.

We have just described masking by only one masker. If the source signal consists of many simultaneous maskers, a *global masking threshold* can be computed that describes the overall threshold of just noticeable distortions as a function of frequency.

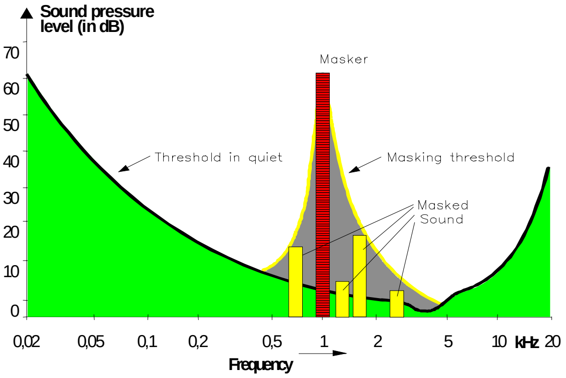


Fig. 4. Threshold in quiet and masking threshold
(Acoustical events in the gray areas will not be audible)

If the necessary bit rate for a complete masking of distortion is available, the coding scheme will be perceptually *transparent*, i.e. the decoded signal is then subjectively indistinguishable from the source signal or from another reference.

In practical designs, we cannot go to the limits of just noticeable distortion, since postprocessing of the audio signal (e.g., filtering in equalizers) by the end-user and multiple encoding/decoding processes in transmission links have to be considered. Moreover, our current knowledge about auditory masking is very limited. Generalizations of masking results, derived for simple and stationary maskers and for limited bandwidths, may be appropriate for most source signals, but may fail for others. Therefore, as an additional requirement, we need a sufficient safety margin in practical designs of such perception-based coders.

### Perceptual Coding

Digital coding at high bit rates is dominantly waveform-preserving, i.e., the amplitude-versus-time waveform of the decoded signal approximates that of the input signal. However, at lower bit rates, facts about the production and perception of speech and audio signals have to be included in coder design, and the error criterion has to be in favor of an output signal that is useful to the human receiver rather than favoring an output signal that follows and preserves the input waveform.

Basically, an efficient source coding algorithm will (i) remove redundant components of the source signal by exploiting correlations between its samples and (ii) remove components which are irrelevant to the ear. Irrelevancy manifests itself as unnecessary amplitude or frequency resolution; portions of the source signal which are masked need not to be transmitted.

The dependence of human auditory perception on frequency and the accompanying perceptual tolerance of errors can (and should) directly influence encoder designs; *noise-shaping techniques* can shift coding noise to frequency bands where that noise is not of perceptual importance. The noise shifting must be dynamically adapted to the actual short-term input spectrum in accordance with the signal-to-mask ratio and can be done in different ways. However, frequency weightings based on linear filtering, as typical in speech coding, cannot make full use of results from psychoacoustics. Therefore, in wideband audio coding, noise-shaping parameters are dynamically controlled in a more efficient way to exploit simultaneous masking and temporal masking. Figure 5 depicts the structure of a *perception-based coder* that exploits auditory masking. The encoding process is controlled by the signal-to-mask ratio (SMR) vs. frequency curve from which the necessary amplitude resolution (and hence the bit allocation and rate) in each critical band is derived. The SMR is the ratio of the sound pressures of signal and its masking threshold within a given frequency band. It is determined from a high resolution, say, a 1024-point FFT-based spectral analysis of the audio block to be coded. Principally, any coding scheme can be used that allows for a dynamic control by such perceptual information. Frequency domain coders are of particular interest since they offer a direct method for noise shaping.

## 5 ISO/MPEG Audio Coding

The MPEG-1 audio coding standard [9,10] has already become a universal standard in diverse fields, such as consumer electronics, professional audio processing, telecommunications, and broadcasting. It offers a subjective reproduction quality that is equivalent to compact disc (CD) quality (16 bit PCM) at stereo rates at and above 128 – 256 kb/s for many types of music.

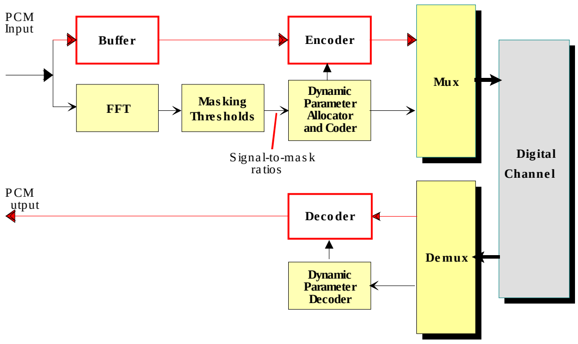


Fig. 5. Block diagram of perception-based coders

The structure of MPEG coders follows that of perception-based coders. In the first step the audio signal is converted into spectral components via an analysis filterbank; Layers I and II make use of a subband filterbank, Layer III employs a hybrid filterbank. Each spectral component is quantized and coded with the goal to keep the quantization noise below the masking threshold. The number of bits for each subband and a scalefactor are determined on a block-by-block basis. The number of quantizer bits is obtained from a dynamic bit allocation algorithm that is controlled by a *psychoacoustical model*. The subband codewords, the scalefactor, and the bit allocation information are multiplexed into one bitstream, together with a header and optional ancillary data. In the decoder the synthesis filterbank reconstructs a block of 32 audio output samples from the demultiplexed bitstream.

The Layer III hybrid filterbank approach has become quite popular, in particular in Internet applications (MP3). The structure of the switched hybrid filterbank is given in Figure 6. This filterbank achieves a higher frequency resolution closer to critical band partitions by subdividing the 32 subband signals further in frequency content by applying, to each of the subbands, a 6-point or 18-point modified DCT block transform, with 50% overlap; hence, the windows contain, resp., 12 or 36 subband samples.

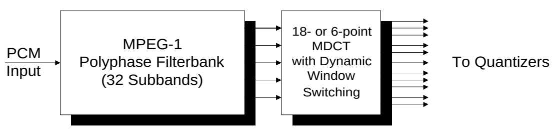


Fig. 6: Hybrid filterbank of MPEG-1 Layer III encoder

In addition it employs an analysis-by-synthesis approach, an advanced pre-echo control, and nonuniform quantization with entropy coding. A buffer technique, called *bit reservoir*, leads to further savings in bit rate.

## 6 MPEG Advanced Audio Coding

The MPEG-2 AAC standard employs high resolution filter banks, prediction techniques, and noiseless coding [11, 12]. It is based on recent evaluations and definitions of *tools (or modules)* each having been selected from a number of proposals. The self-contained tools include an optional preprocessing, a filterbank, a perceptual model, temporal noise shaping, intensity multichannel coding, time-domain prediction, M/S stereo coding, quantization, noiseless coding, and a bit stream multiplexer. The filterbank is a 1024-point modified discrete cosine transform (due to an 50% overlap, the transform is taken over 2048 windowed samples), the perceptual model is taken from MPEG-1.

The temporal noise shaping (TNS) tool plays an important role in improving the overall performance of the coder (see Figure 7). It performs a prediction of the *spectral* coefficients of each audio frame. Instead of the coefficients, the prediction residual is transmitted. TNS is very effective in case of transient audio signals since such transients (signal „attacks“) imply a high predictibility in the spectral domain. (Recall that „peaky“ spectra lead to a high predictibility in the time domain). Therefore the TNS tool controls the time dependence of the quantization noise. Time domain prediction is applied to subsequent subband samples in a given subband in order to further improve coding efficiency, in particular for stationary sounds (see Figure 7 again). Second-order backward-adaptive predictors are used for this purpose. Finally, for quantization an iterative method is employed so as to keep the quantization noise in all critical bands below the global masking threshold.

The MPEG-2 AAC standard offers high quality at lowest possible bit rates, it will therefore find many applications, both for consumer and professional use. The following figure shows MOS differences, with diffscore = 0 for the compact disc reference. For example, the AAC coder operating at 128 kb/s stereo rate is close to the MOS value of the reference (with a diffscore of around - 0.18). At that rate, the MPEG-1 Layer 3 coder (MP3 128) has a diffscore of almost - 1. Note also, that, at a rate of 96 kb/s, the AAC main coder performs better than the MPEG-1 Layer 2 coder at twice the rate (MP2 192).

## 7 MPEG-4 Audio Coding

Activities within MPEG-4 have aimed at proposals for a broad field of applications including multimedia. It is clear that communication services, interactive services and broadcast services will overlap in future applications. The new standard, which has become an international standard in early 1999, takes into account that *a growing part of information is read, seen and heard in interactive ways*. It supports new forms of communications, in particular for Internet and Multimedia applications and in Mobile Communications.

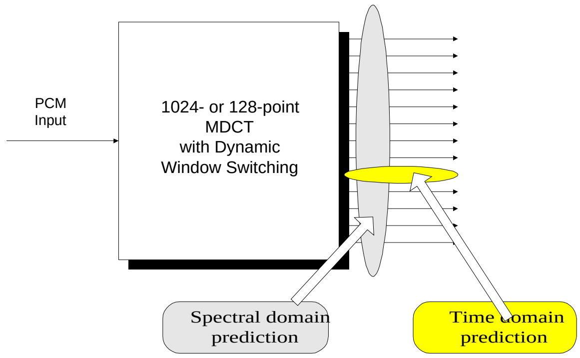


Fig. 7. Spectral and time domain prediction in MPEG-2 Advanced Audio Coding (AAC)

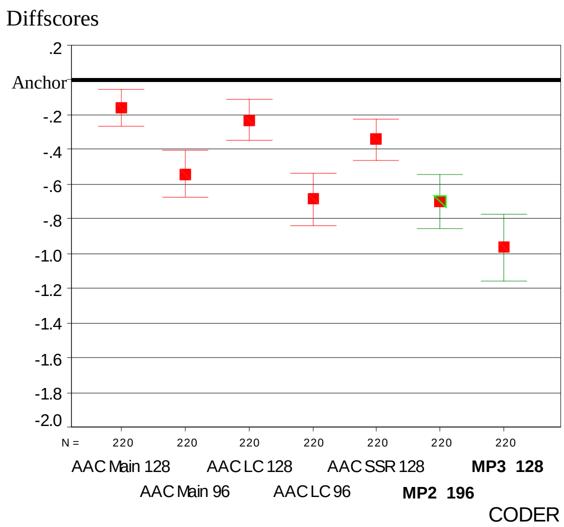


Fig.8: Subjective quality of AAC and MPEG-1 audio coders [13].

MPEG-1 and MPEG-2 have some main disadvantages: they offer only a very limited interaction and control over the presentation and configuration of the system. In addition, an integration of natural and synthetic content is difficult, and an access and transmission across heterogeneous networks is not well-supported. MPEG 4 is different: it represents an audio-visual scene as a composition of (potentially meaningful) objects and supports the evolving ways in which audiovisual material is produced, delivered, and consumed. For example, computer-generated content becomes part in the production of an audiovisual scene. In addition, interaction with objects with scene is possible. For example, it will be possible to associate a Web address to a person in a scene.

In the case of *audio*, MPEG-4 will merge the whole range of audio from high fidelity audio coding and speech coding down to synthetic speech and synthetic audio, supporting applications from high-fidelity audio systems down to mobile-access multimedia terminals. The following figures indicate the potential of MPEG-4: Figure 9 describes an audiovisual scene with a number of audio „objects“: the noise of an incoming train, an announcement, a conversation, and background music.

For example, the noise of the train can be described by an eight-channel representation. On the other hand, if the necessary bandwidth is not available, a one-channel representation - or no representation at all - could be used instead. Such a form of scalability will be very useful in future applications whenever audiovisual signals have to be transmitted to and via receivers of differing complexity and channels of differing capacity. In the case of the announcement, one-channel pseudo 3-D and echo effects could be added. The background music may have an AAC format, or it is of synthetic origin.

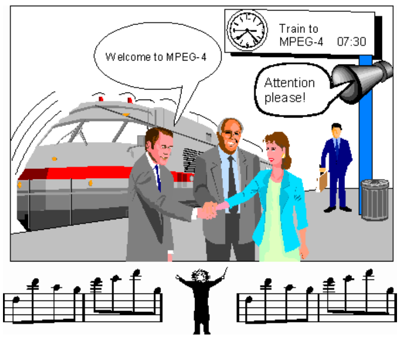


Fig. 9: Audiovisual scene [14]

In order to represent, integrate and exchange pieces of audio-visual information, MPEG-4 offers tools which can be combined to satisfy specific user requirements. A number of such configurations has been standardized. A syntactic description is used to convey to a decoder the choice of tools made by the encoder. This description can also be used to describe new algorithms and download their configuration to the decoding processor for execution. In the case of audio and speech the current toolset supports compression at monophonic bit rates ranging from 2 to 64 kb/s. Three *core coders* are used:

- a parametric coding scheme („vocoder“) for low bit rate speech coding (2 to 10 kbit/s)
- a CELP-based analysis-by-synthesis coding scheme for medium bit rates (4 to 16 kb/s)
- a transform-based coding scheme for higher bit rates (up to 64 kbit/s).

MPEG-4 not only offers simple means of manipulation of coded data such as time scale control, pitch change, but also a flexible access to coded data and subsets thereof, i.e. scalability of bit rate, bandwidth, complexity, and of error robustness. In addition, MPEG-4 supports not only natural audio coding at rates between 2 and 64 kb/s**,** but also text-to-speech conversion (TTS) and structured audio. Natural sounding TTS is obtained by combining conventional TTS synthesis with additional prosodic parameters. The standard offers also an interface between TTS and facial animation for synthetic face models to be driven from speech (“*Talking Heads*”).

Ultra-low bit rate coding of sound is achieved by coding and transmitting parameters of a sound model. MPEG-4 standardizes a sound language and related tools for structured coding of synthetic music and sound effects at rates of 0.01 to 10 kb/s. MPEG-4 does not standardize a particular set of synthesis methods, but a signal-processing language for describing synthesis methods. Any current or future sound-synthesis method may be described in the MPEG-4 structured audio format. The language is entirely normative and standardized, so that every piece of music will sound exactly the same on every compliant MPEG-4 decoder. The following Figure 10 indicates the range of bit rates offered by the new standard.

Transform-based audio coders show a very good performance at bit rates down to 16 kb/s, whereas speech coder perform clearly better at rates between 2.4 kb/s and 16 kb/s. Currently a number of speech coders is available with good performance in that range of bit rates. Both coder classes, however, do not offer solutions for audio coding at 4 - 16 kb/s.

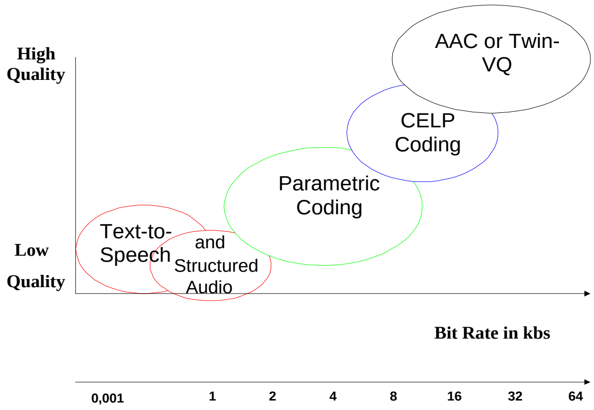


Fig. 10: Range of audio quality and bit rates in MPEG-4.

## 8 References

[1] N. S. Jayant and P. Noll, "Digital Coding of Waveforms: Principles and Applications to Speech and Video," Prentice Hall, 1984.

[2] Purat, M.; Liebchen, T.; Noll, P.: *Lossless Transform Coding of Audio Signals*. Presented at 102nd AES Convention, Munich, 1997.

[3] Bruekers, Oomen, v.d. Vleuten, v.d. Kerkhof: *Lossless Coding for DVD Audio*. Presented at 101st AES Convention, Los Angeles, 1996.

[4] Robinson, T.: *Shorten: Simple lossless and near-lossless waveform compression*. Technical report CUED/F-INFENG/TR.156, Cambridge University Engineering Department, December 1994.

[5] *Sound Quality Assessment Material - Recordings for subjective Tests*. Technical Centre of the European Broadcasting Union (EBU), 1988.

[6] Cellier, C.; Chênes, P.; Rossi, M.: *Lossless Audio Data Compression for Real Time Applications*. Presented at 95th AES Convention, New York, 1993.

[7] Craven, P.; Law, M.; Stuart, J.: *Lossless Compression Using IIR Prediction Filters*. Presented at 102nd AES Convention, Munich, 1997.

[8] E. Zwicker and R. Feldtkeller, Das Ohr als Nachrichtenempfänger. Stuttgart: S. Hirzel Verlag, 1967.

[9] P. Noll, "Digital Audio Coding for Visual Communications," Proc. of the IEEE, vol. 83, No. 6, June 1995.

[10] P. Noll, “MPEG Audio Coding Standards“, IEEE Signal Processing Magazine, Sept. 1997.

[11] ISO/IEC JTCl/SC29, "Information Technology - Generic Coding of Moving Pictures and Associated Audio Information - IS 13818 (Part 7, Audio)" , 1997.

[12] M. Bosi et al, “ISO/IEC MPEG-2 Advanced Audio Coding,“ J. Audio Eng. Soc., Vol 45, No. 10, S. 789 - 814, 1997.

[13] D. Meares, K. Watanabe, and E. Scheirer, “ Report on the MPEG-2 AAC Stereo Verification Tests”, MPEG Document 98/N2006 (Febr. 98).

[14] ISO/IEC/JTC1/SC29, MPEG Document 98/N2431 (Oct. 98).