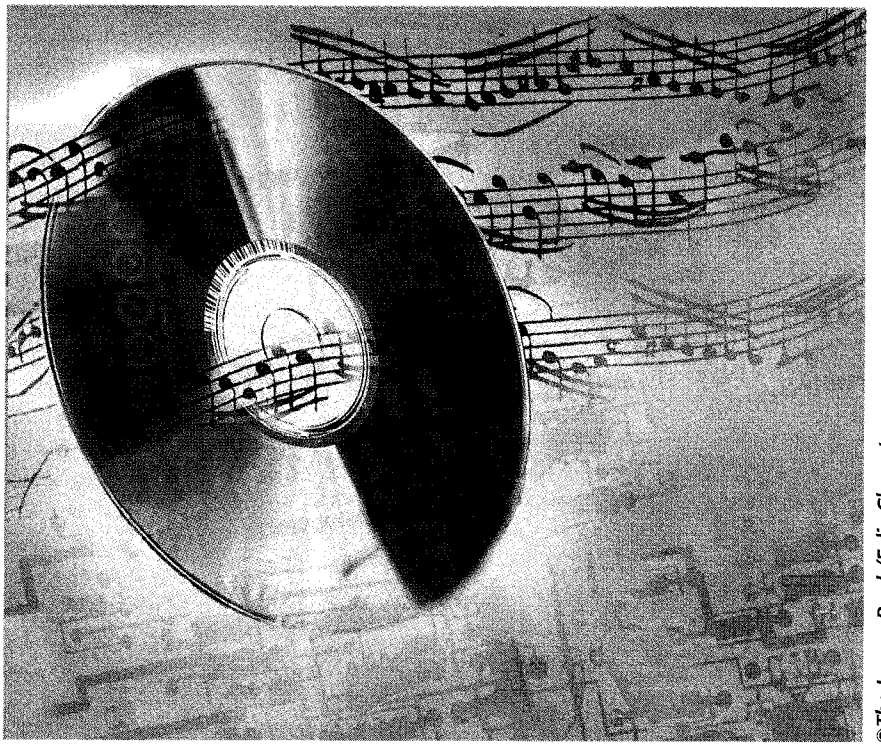
Peter Noll

MPEG Digital Audio Coding

Setting the Standard for High-Quality Audio Compression

The Moving Pictures Expert Group (MPEG) within the International Organization of Standardization (ISO) has developed a series of audio-visual standards known as MPEG-1 and MPEG-2. These audio-coding standards are the first international standards in the field of high-quality digital audio compression. MPEG-1 covers coding of stereophonic audio signals at high sampling rates aiming at *transparent* quality, whereas MPEG-2 also offers stereophonic audio coding at lower sampling rates. In addition, MPEG-2 introduces multichannel coding with and without backwards compatibility to MPEG-1 to provide an improved acoustical image for audio-only applications and for enhanced television and video-conferencing systems. MPEG-2 audio coding without backwards compatibility, called MPEG-2 Advanced Audio Coding (AAC), offers the highest compression rates.

Typical application areas for MPEG-based digital audio are in the fields of audio production, program distribution and exchange, digital sound broadcasting, digital storage, and various multimedia applications. In this article we describe in some detail the key technologies and main features of MPEG-1 and MPEG-2 audio coders. We also present a short section on the upcoming MPEG-4 standard, and we discuss some of the typical applications for MPEG audio compression.



©The Image Bank/Felix Clouzot

Dealing with Bit Rates

PCM Bit Rates

Typical audio signal classes are telephone speech, wideband speech, and wideband audio, all of which differ in bandwidth, dynamic range, and in listener expectation of offered quality. The quality of telephone-bandwidth speech is acceptable for telephony and for some videotelephony services. Higher bandwidths (7 kHz for wideband speech) may be necessary to improve the

1053-5888/97/$10.00©1997IEEE

Table 1. Basic parameters for PCM coding of speech and audio signals.				
	Frequency range in Hz	Sampling rate in kHz	PCM bits per sample	PCM bit rate in kb/s
Telephone speech	300 - 3,400[1]	8	8	64
Wideband speech	50 - 7,000	16	8	128
Mediumband audio	10 - 11,000	24	16	384
Wideband audio	10 - 22,000	48[2]	16	768

Table 2. CD and DAT bit rates (stereophonic signals, sampled at 44.1 kHz; DAT also supports sampling rates of 32 kHz and 48 kHz).			
Storage device	Audio rate	Overhead	Total bit rate
Compact Disc (CD)	1.41 Mb/s	2.91 Mb/s	4.32 Mb/s
Digital Audio Tape (DAT)	1.41 Mb/s	1.67 Mb/s	3.08 Mb/s

intelligibility and naturalness of speech. Wideband (high fidelity) audio representation including multichannel audio needs a bandwidth of at least 20 kHz. The conventional digital format for these signals is pulse code modulation (PCM), with typical sampling rates and amplitude resolutions (PCM bits per sample) as given in Table 1.

The compact disc (CD) is today's de facto standard of digital audio *representation*. With its 44.1 kHz sampling rate, the resulting stereo net bit rate on a CD is $2 \times 44.1 \times 16 \times 1000 \equiv 1.41$ Mb/s (see Table 2). However, the CD needs a significant overhead for a runlength-limited line code, which maps 8 information bits into 14 bits, for synchronization and for error correction, resulting in a 49-bit representation of each 16-bit audio sample. Hence, the total stereo bit rate is $1.41 \times 49/16 = 4.32$ Mb/s. Table 2 compares bit rates of the CD and the digital audio tape (DAT).

For archiving and processing of audio signals, sampling rates twice as large as those mentioned and amplitude resolutions of up to 24 b/sample are being discussed. Furthermore, lossless coding is an important topic in order not to compromise audio quality in any way [1]. The digital versatile (or video) disk (DVD) with its capacity of 4.7 GB (single layer) or 8.5 GB (double layer) is the appropriate storage medium for such applications.

Bit Rate Reduction

Although high bit rate channels and networks have become more easily accessible, low bit rate coding of audio signals has retained its importance. The main motivations for low bit-rate coding are the need to minimize transmission costs or to provide cost-efficient storage, the demand to transmit over channels of limited capacity such as mobile radio channels, and to support variable-rate coding in packet-oriented networks.

Basic requirements in the design of low bit-rate audio coders are firstly, to retain a high quality of the reconstructed signal with robustness to variations in spectra and levels. In the case of stereophonic and multichannel signals, spatial integrity is an additional dimension of quality. Secondly, robustness against random and bursty channel bit errors and packet losses is required. Thirdly, low complexity and power consumption of the codecs are of high relevance. For example, in broadcast and playback applications, the complexity and power consumption of audio decoders used must be low, whereas constraints on encoder complexity are more relaxed. Additional network-related requirements are low encoder/decoder delays, robustness against errors introduced by cascading codecs, and a graceful degradation of quality with increasing bit error rates in mobile radio and broadcast applications. Finally, in professional applications, the coded bit streams must allow editing, fading, mixing, and dynamic range compression.

We have seen rapid progress in bit-rate compression techniques for speech and audio signals [2-7]. Linear prediction, subband coding, transform coding, as well as various forms of vector quantization and entropy coding techniques have been used to design efficient coding algorithms that can achieve substantially more compression than was thought possible only a few years ago. Recent results in speech and audio coding indicate that an excellent coding quality can be obtained with bit rates of 0.5 to 1 b/sample for speech and wideband speech, and 1 to 2 b/sample for audio. In storage and packet-oriented transmission systems, additional savings are possible by employing variable-rate coding with its potential to offer a time-independent, constant-quality performance.

Compressed digital audio representations can be made less sensitive to channel impairments than analog ones if source and channel coding are implemented appropriately. Bandwidth expansion has often been mentioned as

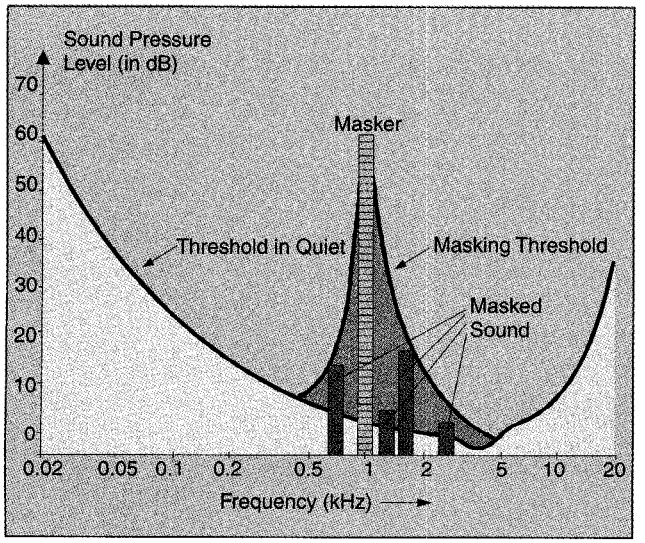
a disadvantage of digital coding and transmission, but with today's data compression and multilevel signaling techniques, channel bandwidths can be made smaller than those of analog systems. In broadcast systems, the reduced bandwidth requirements, together with the error robustness of the coding algorithms, will allow an efficient use of available radio and TV channels as well as "taboo" channels currently left vacant because of interference problems.

MPEG Standardization Activities

Of particular importance for digital audio is the standardization work within the ISO/IEC that is intended to provide international standards for a wide range of communications-based and storage-based applications. This group is called MPEG, an acronym for *Moving Pictures Experts Group*. MPEG's initial effort was the MPEG Phase 1 (MPEG-1) coding standard IS 11172, which supports bit rates of around 1.2 Mb/s for video (with video quality comparable to that of today's analog video cassette recorders) and 256 kb/s for two-channel audio (with audio quality comparable to that of today's CDs) [8].

The more recent MPEG-2 standard IS 13818 provides, in its video part, standards for high-quality video (including high definition TV (HDTV)) in bit-rate ranges from 3 to 15 Mb/s and above. In its audio part, multichannel audio coding with two to five full-bandwidth audio channels has been standardized. In addition, for stereophonic audio the range of sampling rates was extended to lower sampling frequencies for bit rates at, or below, 64 kb/s [9]. Part IS 13818-7 of that standard will offer a collection of very flexible tools for advanced audio coding (MPEG-2 AAC) for applications where compatibility with MPEG-1 is not relevant.

Finally, the current MPEG-4 work addresses standardization of audiovisual coding for applications ranging from mobile-access, low-complexity multimedia terminals to high-quality multichannel sound systems. The standard will allow for interactivity and universal accessibility, and it will provide a high degree of flexibility and extensibility [10].



▲ *1. Threshold in quiet and masking threshold (acoustical events in the gray areas will not be audible).*

In MPEG coding the encoder is not standardized, thus leaving room for improvements in the coding process.

Key Technologies in Audio Coding

Proposals to reduce wideband audio-coding rates have followed those for speech coding. Differences between audio and speech signals are manifold, however, audio coding implies higher sampling rates, better amplitude resolution, higher dynamic range, larger variations in power density spectra, stereophonic and multichannel audio signal representations, and, finally, higher quality expectations. Indeed, the high quality of the CD with its 16-b/sample PCM format has made digital audio popular.

Speech and audio coding are similar in that, in both cases, quality is based on the properties of human auditory perception. On the other hand, speech can be coded very efficiently because a *speech production model* is available, whereas nothing similar exists for audio signals.

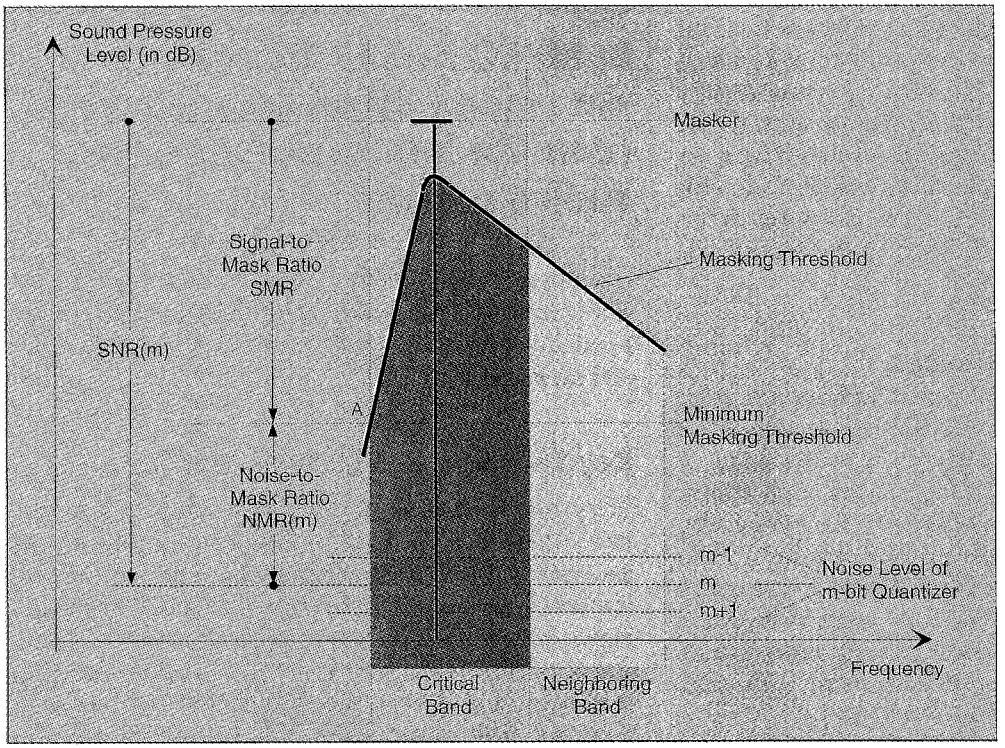
Modest reductions in audio bit rates have been obtained by instantaneous companding (e.g., a conversion of uniform 14-bit PCM into a 11-bit nonuniform PCM presentation), or by forward-adaptive PCM (block companding) as employed in various forms of near-instantaneously companded audio multiplex (NICAM) coding (ITU-R, Rec. 660). For example, the British Broadcasting Corporation (BBC) has used the NICAM 728 coding format for digital transmission of sound in several European broadcast television networks; it employs 32 kHz sampling with 14-bit initial quantization followed by a compression to a 10-bit format on the basis of 1 ms blocks resulting in a total stereo bit rate of 728 kb/s [11]. Such adaptive PCM schemes can solve the problem of providing a sufficient dynamic range for audio coding, but they are not efficient compression schemes since they do not exploit statistical dependencies between samples and do not sufficiently remove signal irrelevancies.

In recent audio-coding algorithms, four key technologies play an important role: perceptual coding, frequency-domain coding, window switching, and dynamic bit allocation. These technologies will be covered in the following sections.

Auditory Masking and Perceptual Coding

Auditory Masking

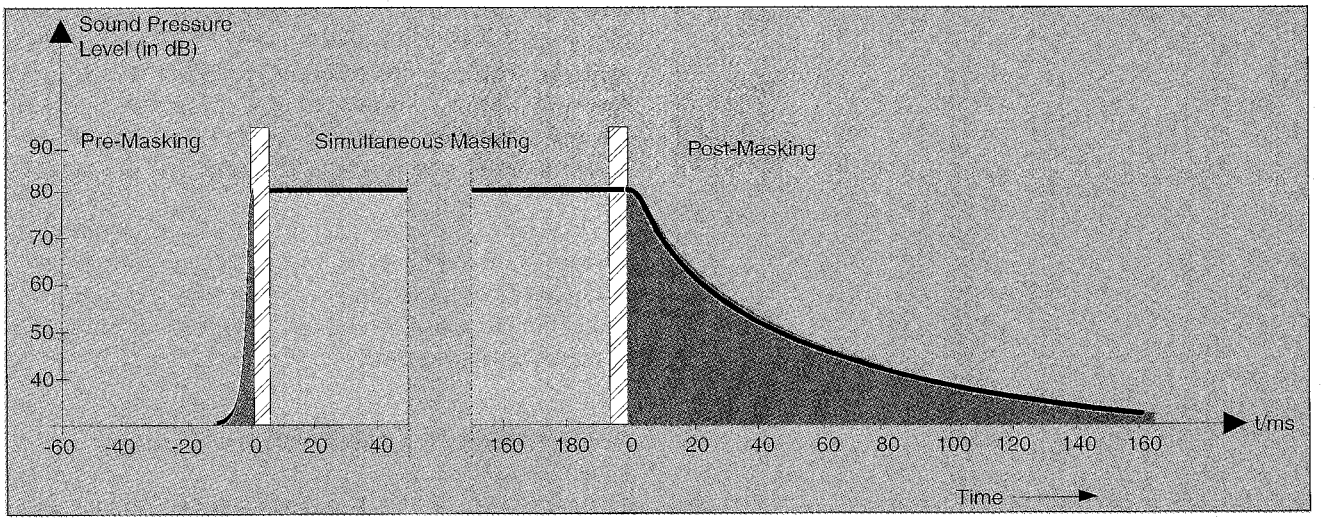
The inner ear performs short-term critical band analyses where frequency-to-place transformations occur along



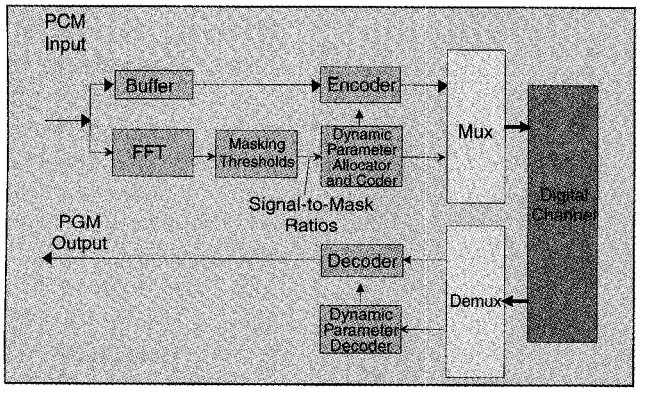
▲ *2. Masking threshold and signal-to-mask ratio (SMR) (acoustical events in the gray areas will not be audible).*

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 50 to 100 Hz for signals below 500 Hz and up to 5000 Hz for signals at high frequencies. Twenty-six critical bands covering frequencies of up to 24 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) as long as masker and maskee are close enough to each other in frequency [12]. Such masking is largest in the critical band in which the masker is located, and it is effective to a lesser degree in neighboring bands. A *masking threshold* can be measured and low-level signals below this threshold 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, in the context of source coding also known as threshold of just noticeable distortion (JND) [13], varies with time. It depends on the sound pressure level (SPL), the frequency of the masker, and on characteristics of masker and maskee. Take the example of the masking threshold for the SPL = 60 dB narrowband masker in Fig. 1: around 1 kHz the four maskees will be masked as long as their individual sound pressure levels are below the masking threshold. The slope of the masking threshold is steeper toward lower frequencies; i.e. higher frequencies are more easily 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. In MPEG coders both thresholds play a role in computing the masking threshold. Without a masker, a signal is inaudible if its sound pres-



▲ *3. Temporal masking (acoustic events in the gray areas will not be audible).*



▲ *4. Block diagram of perception-based coders (acoustical events in the gray areas will not be audible).*

sure 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 Fig. 1.

The qualitative sketch of Fig. 2 gives a few more details about the masking threshold: within a critical band, tones below this threshold (darker area) are masked. The distance between the level of the masker and the masking threshold is called the signal-to-mask ratio (SMR). Its maximum value is at the left border of the critical band (point A in Fig. 2), and its minimum value occurs in the frequency range of the masker and is around 6 dB in noise-masking-tone experiments. Assuming an m-bit quantization of an audio signal, within a critical band the quantization noise will not be audible as long as its signal-to-noise ratio (SNR) is higher than its SMR. Noise *and* signal contributions *outside* the particular critical band will also be masked, although to a lesser degree, if their SPL is below the masking threshold.

Defining SNR(m) as the SNR resulting from an m-bit quantization, the perceivable distortion in a given subband is measured by the *noise-to-mask ratio* (NMR):

$$\mathrm{NMR}(m) = \mathrm{SMR} - \mathrm{SNR}(m) \text{ (in dB)}.$$

NMR(m) describes the difference in dB between the SMR and the SNR ratio to be expected from an m-bit quantization. The NMR value is also the difference (in dB) between the level of quantization noise and the level where a distortion may just become audible in a given subband. Within a critical band, coding noise will not be audible as long as NMR(m) is negative.

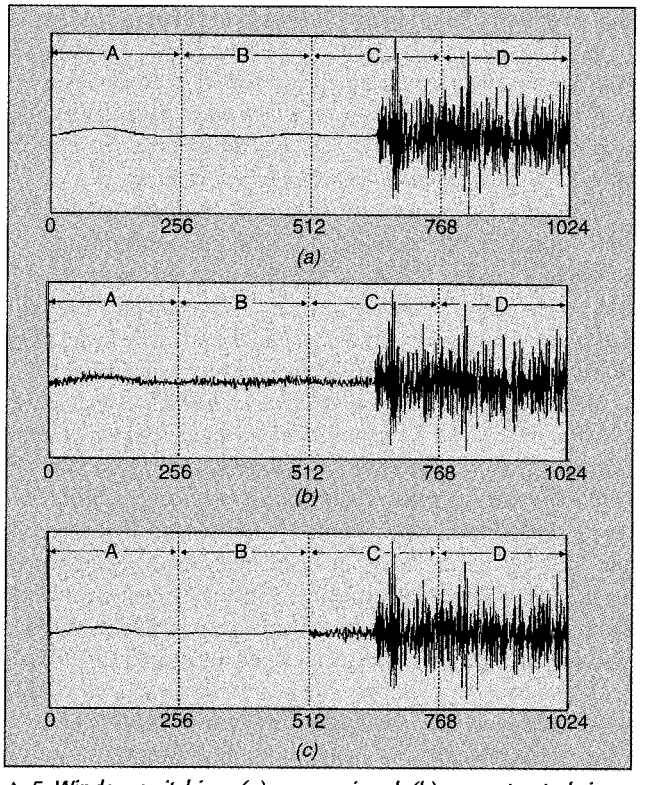
We have just described masking by only one masker. If the source signal consists of many simultaneous maskers, each has its own masking threshold, and a *global masking threshold* can be computed that describes the threshold of just-noticeable distortions as a function of frequency (see also the "ISO/MPEG-1 Audio Coding" section).

In addition to simultaneous masking, the time-domain phenomenon of *temporal masking* plays an important role in human auditory perception. It may occur when two sounds appear within a small interval of time. Depending on the individual SPLs, the stronger sound may mask the weaker one, even if the maskee precedes the masker (Fig. 3).

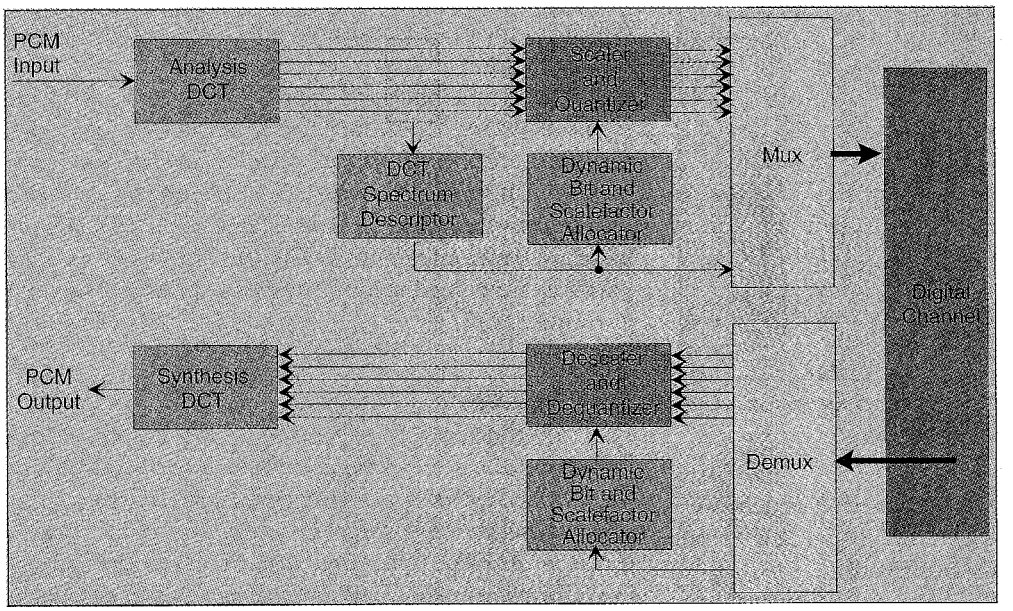
Temporal masking can help to mask pre-echoes caused by the spreading of a sudden large quantization error over the actual coding block (see the "Window Switching" section). The duration within which *premasking* applies is significantly less than one tenth of that of the *postmasking,* which is in the order of 50 to 200 ms. Both pre- and post-masking are being exploited in MPEG audio-coding algorithms.

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. The difference signal between input and output waveforms is then the basic error criterion of coder design. Waveform coding principles have been covered in detail in [2]. At lower bit rates, facts about the production and perception of 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 that are perceptually irrelevant to the ear. Irrelevancy manifests itself as unnecessary amplitude or frequency resolution; portions



▲ *5. Window switching: (a) source signal, (b) reconstructed signal with block size N = 1024, (c) reconstructed signal with block size N = 256. (Source: Iwadare et al. [25].)*



▲ 6. Conventional adaptive transform coding (ATC).

of the source signal that are masked do not need to be transmitted.

The dependence of human auditory perception on frequency and on the accompanying perceptual tolerance of errors can (and should) directly influence encoder designs; *noise-shaping techniques* can emphasize coding noise in frequency bands where that noise is not important for perception. To this end, the noise shifting must be dynamically adapted to the actual short-term input spectrum in accordance with the SMR, which can be done in different ways. However, frequency weightings based on linear filtering, as is 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.

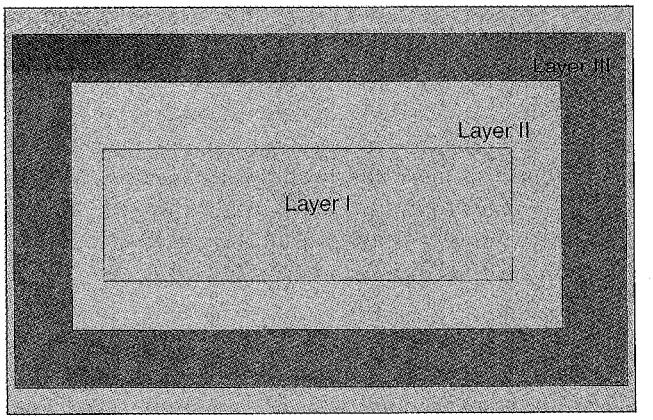
Figure 4 depicts the structure of a *perception-based coder* that exploits auditory masking. The encoding process is controlled by the SMR vs. frequency curve from which the needed amplitude resolution (and hence the bit allocation and rate) in each frequency band is derived. The SMR is typically determined from a high resolution, say, a 1024-point FFT-based spectral analysis of the audio block to be coded. In general, any coding scheme may be used that can be dynamically controlled by such perceptual information. Frequency domain coders (see next section) are of particular interest since they offer a direct method for noise shaping. If the frequency resolution of these coders is high enough, the SMR can be derived directly from the subband samples or transform coefficients without running a FFT-based spectral analysis in parallel [14, 15].

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. In practical designs, we cannot go to the limits of just-noticeable distortion, since postprocessing of the acoustic signal 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. It should be noted that the MPEG/Audio coding standard is open for better encoder-located psychoacoustic models since such models are not normative elements of the standard (see the "ISO/MPEG-1 Audio Coding" section).

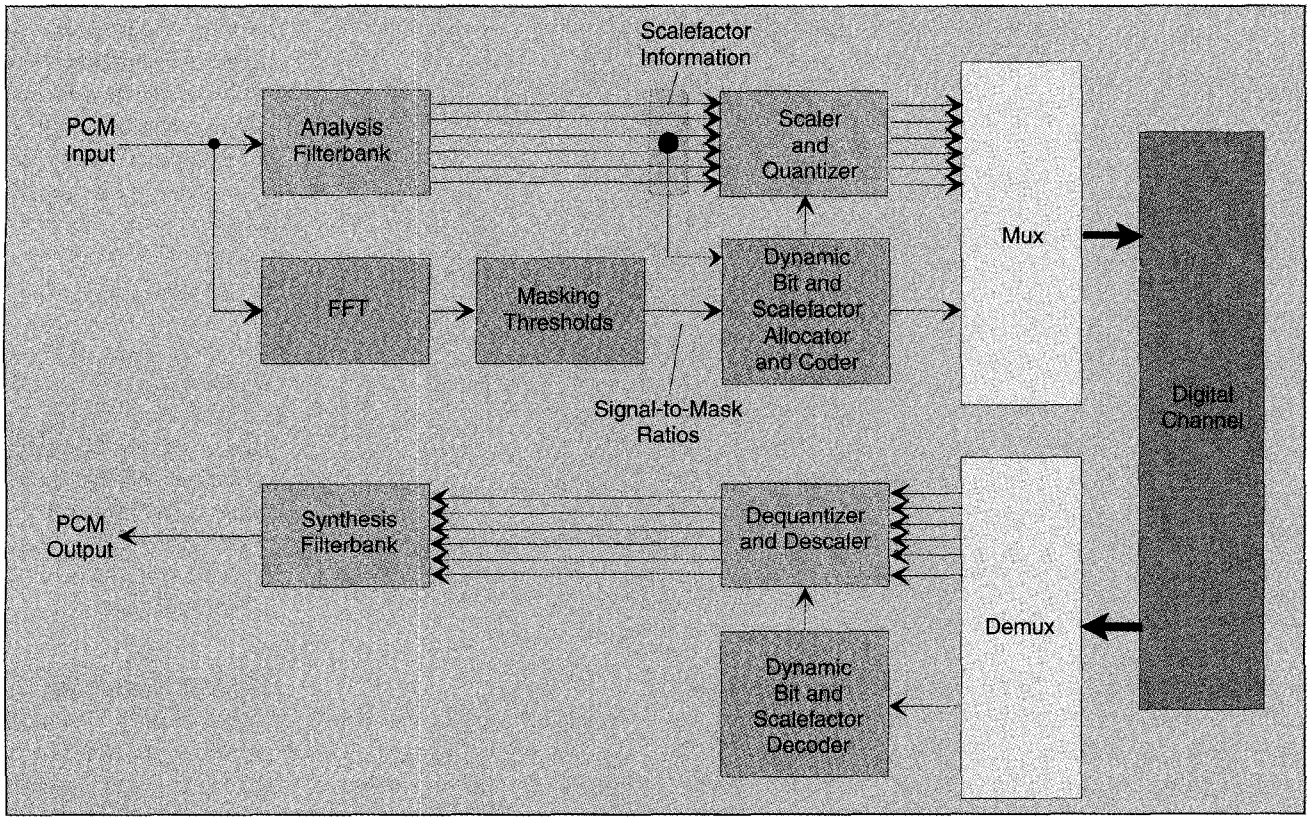
Frequency-Domain Coding

As one example of dynamic noise-shaping, quantization noise feedback can be used in predictive schemes [16, 17]. However, frequency-domain coders with dynamic allocations of bits (and hence of quantization noise contributions) to subbands or transform coefficients offer an easier and more accurate way to control the quantization noise [2, 14] (see also the "Dynamic Bit Allocation" section).

In all frequency-domain coders, redundancy (the nonflat short-term spectral characteristics of the source signal) and irrelevancy (signals below the psychoacoustical thresholds) are exploited to reduce the transmitted data rate with respect to PCM. This is achieved by splitting the source spectrum into frequency bands to generate nearly



▲ 7. Hierarchy of Layers I, II, and III of MPEG-1/Audio.



▲ *8. Structure of MPEG-1 audio encoder and decoder (Layers I and II).*

uncorrelated spectral components and by quantizing these components separately. Two coding categories exist, *transform coding* (TC) and *subband coding* (SBC). The differentiation between these two categories is mainly due to historical reasons. Both use an analysis filterbank in the encoder to decompose the input signal into subsampled spectral components. The spectral components are called subband samples if the filterbank has low frequency resolution, otherwise they are called spectral lines or transform coefficients. These spectral components are recombined in the decoder via synthesis filterbanks.

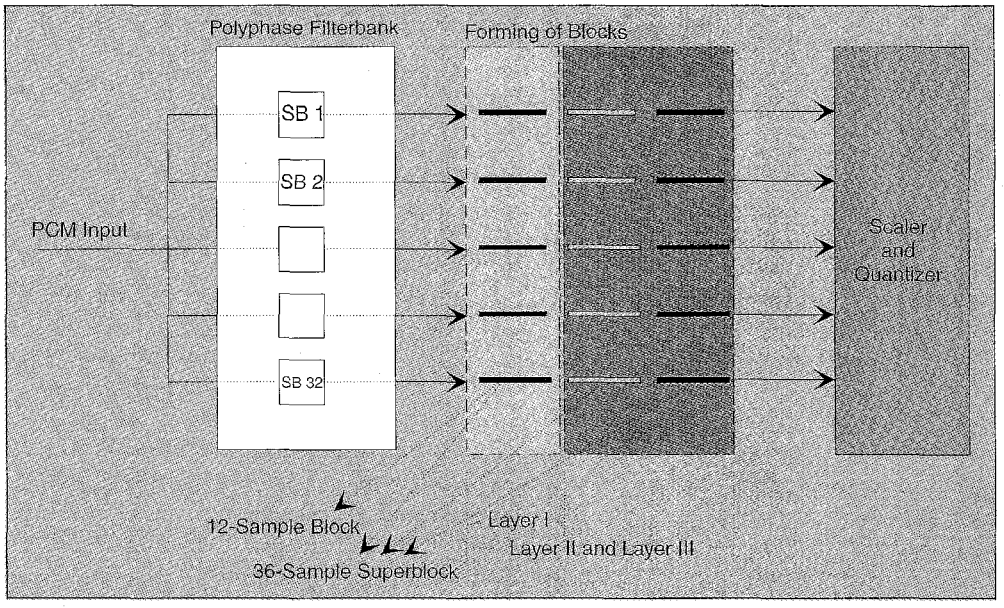
In SBC, the source signal is fed into an analysis filterbank consisting of M bandpass filters that are contiguous in frequency so that the set of subband signals can be recombined additively to produce the original signal or a close version thereof. Each filter output is critically decimated (i.e., sampled at twice the nominal bandwidth) by a factor equal to M, the number of bandpass filters. This decimation results in an aggregate number of subband samples that equals that in the source signal. In the receiver, the sampling rate of each subband is increased to that of the source signal by filling in the appropriate number of zero samples. Interpolated subband signals appear at the bandpass outputs of the synthesis filterbank. The sampling processes may introduce aliasing distortion due to the overlapping nature of the subbands. If perfect filters, such as two-band quadrature mirror filters or polyphase filters, are applied, aliasing terms will cancel and the sum of the bandpass outputs equals the source signal in the absence of quantization [18-21]. With quantization, aliasing components will not cancel ideally; nevertheless, the errors will be inaudible in MPEG/Audio coding if a sufficient number of bits is used. However, these errors may reduce the original dynamic range of 20 bits to around 18 bits [15].

In TC, a block of input samples is linearly transformed via a discrete transform into a set of near-uncorrelated

Table 3. Approximate MPEG-1 bit rates for transparent representations of audio signals and corresponding compression factors (compared to CD bit rate).

MPEG-1/Audio coding	Approximate stereo bit rates for transparent quality	Compression factor
Layer I	384 kb/s	4
Layer II	192 kb/s	8
Layer III	128 kb/s*	12

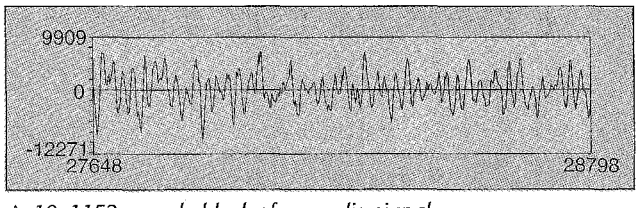
*Average bit rate; variable bit rate coding assumed



▲ *9. Block companding in MPEG-1 audio codecs.*

transform coefficients. These coefficients are then quantized and transmitted in digital form to the decoder. In the decoder an inverse transform maps the signal back into the time domain. In the absence of quantization errors the synthesis yields exact reconstruction. Typical transforms are the discrete Fourier transform or the discrete cosine transform (DCT), calculated via an FFT, and modified versions thereof. We have already mentioned that the decoder-based inverse transform can be viewed as the synthesis filterbank; the impulse responses of its bandpass filters equal the basis sequences of the transform. The impulse responses of the analysis filterbank are just the time-reversed versions thereof. The finite lenghts of these impulse responses may cause so-called block boundary effects. State-of-the-art transform coders employ a modified DCT (MDCT) filterbank as proposed by Princen and Bradley [20]. The MDCT is typically based on a 50% overlap between successive analysis blocks. Without quantization they are free from block boundary effects, have a higher transform coding gain than the DCT, and their basis sequences correspond to better bandpass responses. In the presence of quantization, block boundary effects are de-emphasized due to the doubling of the filter impulse responses resulting from the overlap.

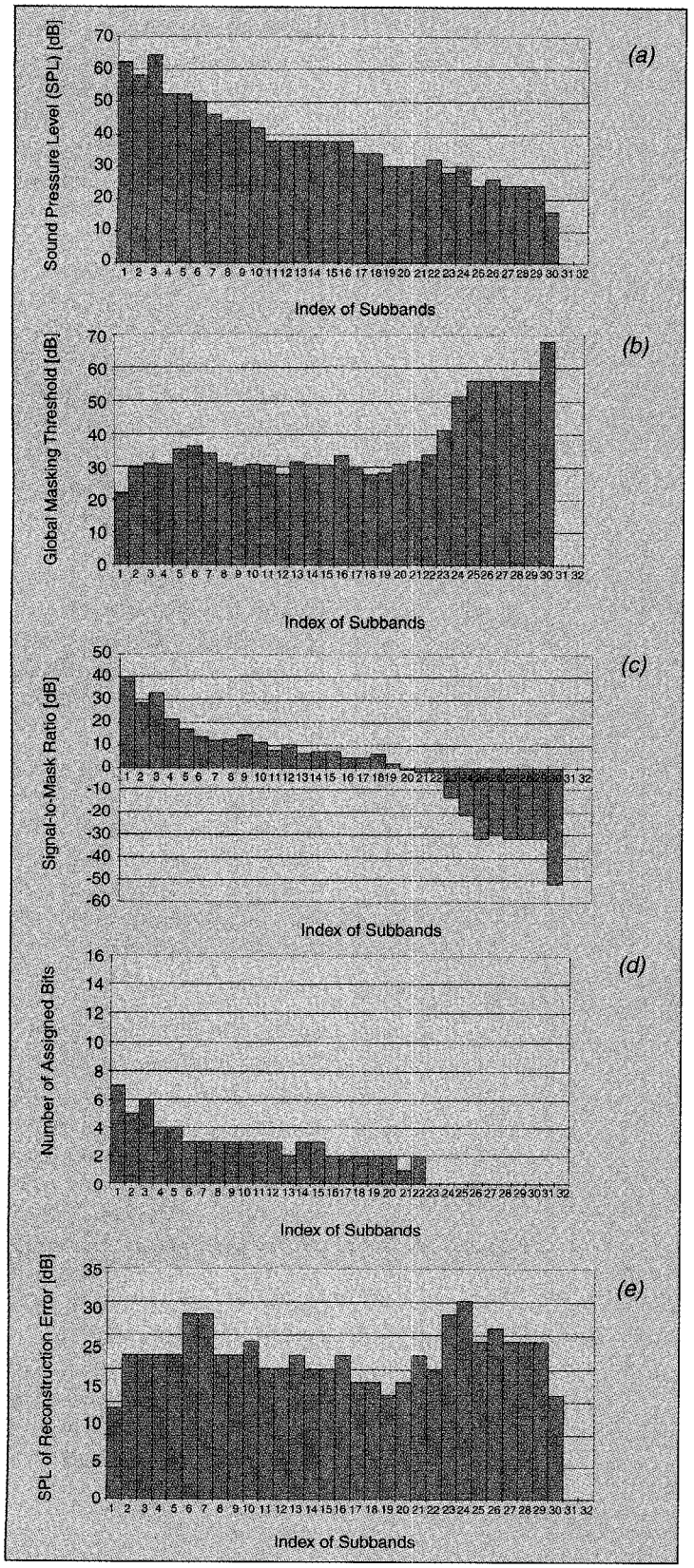
Hybrid filterbanks, i.e., combinations of discrete transform and filterbank implementations, have frequently been used in speech and audio coding [22, 23]. One of the advantages is that different frequency resolutions can be provided at different frequencies in a flexible way and with low complexity. For example, a high spectral resolution can be obtained in an efficient way by using a cascade of a filterbank (with its short delays) and a linear MDCT transform that splits each subband sequence further in frequency content to achieve a high-frequency resolution. MPEG audio coders use a subband approach in Layer I and Layer II, and a hybrid filterbank in Layer III.



▲ *10. 1152-sample block of an audio signal.*

Window Switching

A crucial part in frequency-domain coding of audio signals is the appearance of pre-echoes, which are similar to copying effects on analog tapes. Consider the case that a silent period is followed by a percussive sound, such as from castanets or triangles, within the same coding block. Such an onset ("attack") will cause comparably large instantaneous quantization errors. In TC, the inverse transform in the *decoding* process will distribute such errors over the block; similarily, in SBC, the decoder bandpass filters will spread such errors. In both mappings pre-echoes can become distinctively audible, especially at low bit rates with comparably high error contributions. Pre-echoes can be masked by the time-domain effect of premasking if the time spread is of short length (in the order of a few milliseconds). Therefore, they can be reduced or avoided by using blocks of short lengths. However, a larger percentage of the total bit rate is typically required for the transmission of side information if the blocks are shorter. A solution to this problem is to switch between block sizes of different lengths as proposed by Edler (*window switching*) [24]; typical block sizes are between $N = 64$ and $N = 1024$. The small blocks are only used to control pre-echo artifacts during nonstationary periods of the signal, otherwise the coder switches back to long blocks. It is clear that block size selection has to be based on an analysis of the characteristics of the actual audio-coding block. Fig. 5 demonstrates the effect in transform coding: if the block size is $N = 1024$ (Fig. 5b) pre-echoes are clearly (visible and) audible whereas a block size of 256 will reduce these effects because they are limited to the block where the signal attack and the corresponding quantization errors occur (Fig. 5c). In addition, pre-masking can become effective.



▲ *11. Frequency distributions of various important MPEG parameters taken from the audio block of Fig. 10. MPEG-1 Layer II coding with an overall bit rate of 128 kb/s. (a) Sound-pressure level (SPL) of input frame vs. index of subbands (each subband is 750 Hz wide); (b) Global masking threshold vs. frequency; (c): Signal-to-mask ratio vs. frequency; (d) Bit allocation vs. frequency; (e) SPL of reconstruction error vs. frequency.*

Dynamic Bit Allocation

Frequency-domain coding significantly gains in performance if the number of bits assigned to each of the quantizers of the transform coefficients is adapted to the short-term spectrum of the audio-coding block on a block-by-block basis. In the mid-1970s Zelinski and Noll introduced *dynamic bit allocation* and demonstrated significant SNR-based and subjective improvements with their adaptive transform coding (ATC) (see Fig. 6) [14, 26]. They proposed a DCT mapping and a dynamic bit-allocation algorithm that used the DCT transform coefficients to compute a DCT-based short-term spectral envelope. Parameters of this spectrum were coded and transmitted from which the short-term spectrum was estimated using linear interpolation in the log-domain. This estimate was then used to calculate the optimum number of bits for each transform coefficient, both in the encoder and decoder.

ATC had a number of shortcomings, such as block boundary effects, pre-echoes, marginal exploitation of masking, and low subjective quality at low bit rates. Despite these shortcomings we find many of the features of the conventional ATC in more recent frequency-domain coders. Examples of the very sophisticated bit-allocation strategies that MPEG audio-coding algorithms use will be described in detail in the "Layers 1 and 2" section.

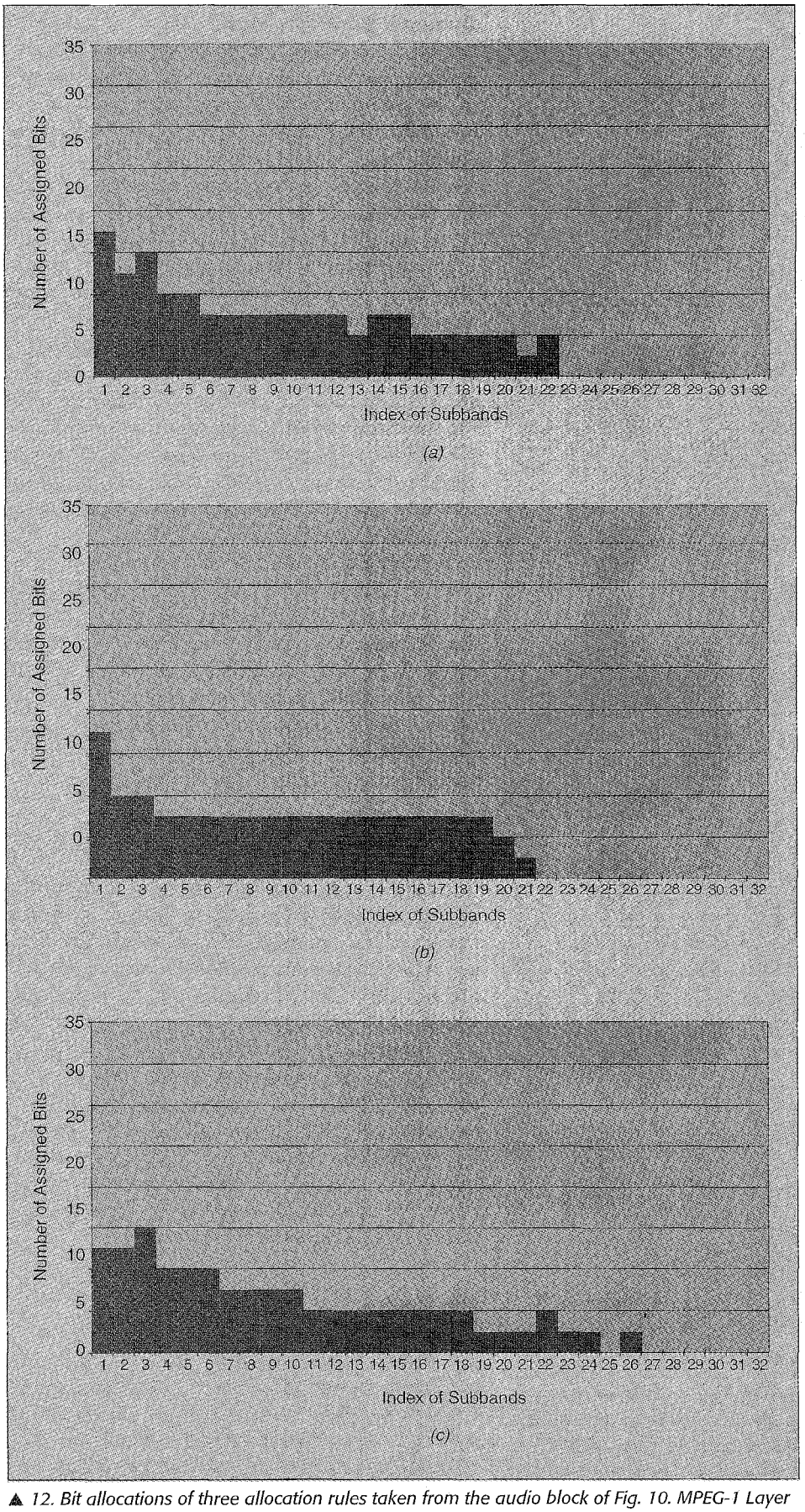
ISO/MPEG-1 Audio Coding

The MPEG audio-coding standard [8, 27-29] has already become a universal standard in diverse fields such as consumer electronics, professional audio processing, telecommunications, and broadcasting [30]. The standard combines features of MUSICAM and ASPEC coding algorithms [31, 32]. Main steps of development toward the MPEG-1 audio standard have been described in [29, 33]. MPEG-1 audio coding offers a subjective reproduction quality that is equivalent to CD quality (16-bit PCM) at stereo rates given in Table 3 for many types of music. Because of its high dynamic range, MPEG- aludio has potential to exceed the quality of a CD [30, 34].

The Basics

Structure

The basic structure of MPEG-1 audio coders follows that of perception-based coders (see Fig. 4). 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 and Layer III employs a hybrid filterbank. Each spectral component is quantized and coded with the goal of keeping the quantization noise below the masking



▲ *12. Bit allocations of three allocation rules taken from the audio block of Fig. 10. MPEG-1 Layer II coding with an overall bit rate of 128 kb/s: (a) Bit allocation (model 1); (b): Bit allocation (model 2); (c) Bit allocation for unweighted minimum mean-squared error.*

threshold. The number of bits for each subband and a scalefactor are determined on a block-by-block basis: each block has 12 (Layer I) or 36 (Layers II and III) subband samples (see the "Layers I and II" section). The number of quantizer bits is obtained from a dynamic bit-allocation algorithm (Layers I and II) that is controlled by a *psychoacoustic model* (see below). 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.

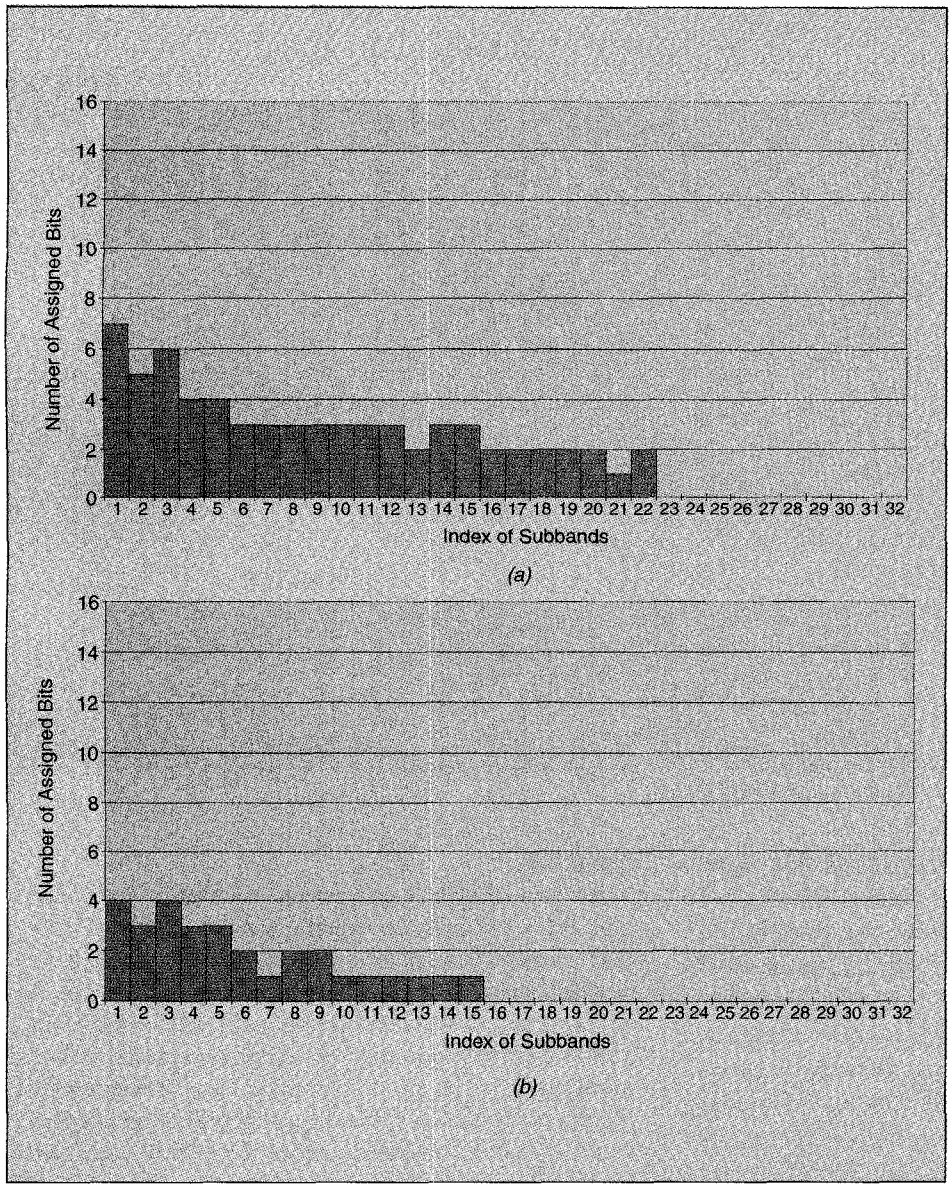
MPEG-1/Audio supports sampling rates of 32, 44.1, and 48 kHz and bit rates between 32 kb/s (mono) and 448 kb/s, 384 kb/s, and 320 kb/s (stereo and Layers I, II, and III, respectively).

Layers and Operating Modes

The standard consists of three layers (I, II, and III) of increasing complexity, delay, and subjective performance. From a hardware and software point of view, the higher layers incorporate the main building blocks of the lower layers (Fig. 7). A standard full MPEG-1 audio decoder is able to decode bit streams of all three layers. More typical are MPEG-1/Audio Layer X decoders (X = I, II, or III).

Stereo Redundancy Coding

MPEG-1/Audio supports four *modes*: mono, stereo, dual with two separate channels (useful for bilingual programs), and joint stereo. In the optional joint-stereo mode interchannel dependen-



▲ *13. Bit allocations for MPEG-1 Layer II coding with overall bit rates of 128kb/s and 64 kb/s.*

cies are exploited to reduce the overall bit rate by using an irrelevancy-reducing technique called *intensity stereo*. It is known that, above 2 kHz and within each critical band, the human auditory system bases its perception of stereo imaging more on the temporal envelope of the audio signal than on its temporal fine structure. Therefore, the MPEG audio-compression algorithm supports a stereo redundancy coding mode called *intensity stereo coding,* which reduces the total bit rate without violating the spatial integrity of the stereophonic signal.

In this mode the encoder codes some upper-frequency subband outputs with a single sum signal L + R (or some linear combination thereof) instead of sending independent left (L) and right (R) subband signals. The decoder reconstructs the left and right channels based only on the single L + R signal and on independent left- and right-channel scalefactors. Hence, the spectral shape of the left and right outputs is the same within each intensity-coded subband but the magnitudes are different [35]. The optional joint stereo mode will only be effective if the required bit rate exceeds the available bit rate, and it will only be applied to subbands corresponding to frequencies of around 2 kHz and above.

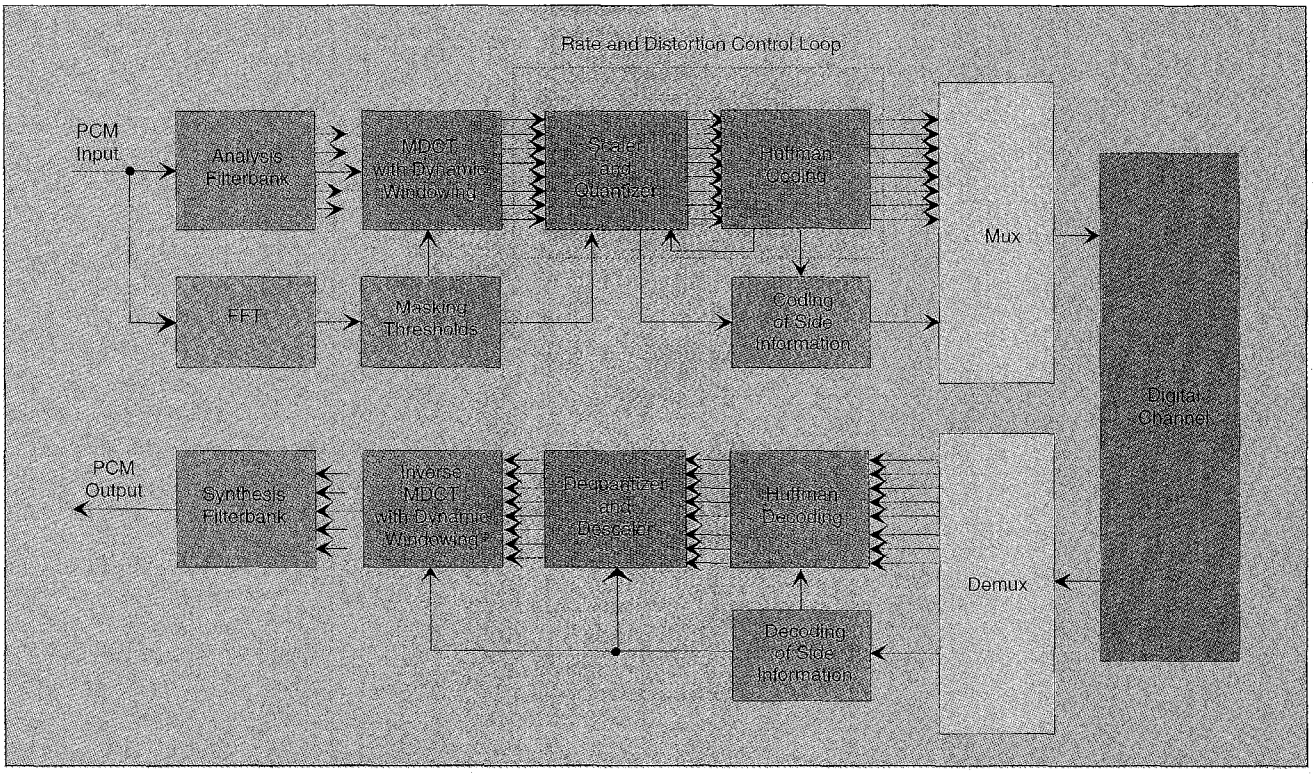
Layer III has an additional option: in the mono/stereo (M/S) mode the left- and right-channel signals are encoded as middle (L + R) and side (L − R) channels. This latter mode can be combined with the joint stereo mode.

Psychoacoustic Models

We have already mentioned that the adaptive bit-allocation algorithm is controlled by a psychoacoustic model. This model computes SMRs and takes into account the short-term spectrum of the audio block to be coded and knowledge about noise masking. The model is only needed in the encoder, which makes the decoder less complex; this asymmetry is a desirable feature for audio playback and audio broadcasting applications.

The normative part of the standard describes the decoder and the meaning of the encoded bitstream, but the encoder is not standardized, thus leaving room for an evolutionary improvement of the encoder. In particular, different psychoacoustic models can be used that range from very simple (or none at all) to very complex based on quality and implementability requirements. Information about the short-term spectrum can be derived in various ways; for example, as an accurate estimate from an FFT-based spectral analysis of the audio input samples, or, less accurate, directly from the spectral components as in the conventional ATC [14] (see also Fig. 6). Encoders can also be optimized for a certain application. All these encoders can be used with complete compatibility with all existing MPEG-1 audio decoders.

The informative part of the standard gives two examples of FFT-based models (see also [8, 29, 36]). Both models identify, in different ways, tonal and nontonal spectral components and use the corresponding results of tone-masks-noise and noise-masks-tone experiments in



▲ *14. Structure of MPEG-1 audio encoder and decoder (Layer III).*

the calculation of the global masking thresholds. Details are given in the standard experimental results for both psychoacoustic models that have been described in [36]. In the informative part of the standard a 512-point FFT is proposed for Layer I, and a 1024-point FFT for layers II and III. In both models the audio input samples are Hann-weighted. *Model 1*, which may be used for Layers I and II, computes for each masker its individual masking threshold, taking into account its frequency position, power, and tonality information. The global masking threshold is obtained as the sum of all individual masking thresholds and the absolute masking threshold. The SMR is then the ratio of the maximum signal level within a given subband and the minimum value of the global masking threshold in that given subband (see Fig. 2). *Model 2*, which may be used for all layers, is more complex: tonality is assumed when a simple prediction indicates a high prediction gain, the masking thresholds are calculated in the cochlea domain, i.e., properties of the inner ear are taken into account in more detail, and, finally, in case of potential pre-echoes the global masking threshold is adjusted appropriately.

Layers I and II

MPEG Layer I and II coders have very similar structures. The Layer II coder achieves a better performance, mainly because the overall scalefactor side information is reduced by exploiting redundancies between the scalefactors. Additionally, a slightly finer quantization is provided.

Filterbank

Layer I and Layer II coders map the digital audio input into 32 subbands via equally spaced bandpass filters (Figs. 8 and 9). A polyphase filter structure is used for the frequency mapping; its filters have 512 coefficients. Polyphase structures are computationally very efficient since a DCT can be used in the filtering process, and they are of moderate complexity and low delay. On the negative side, the filters are equally spaced, and therefore the frequency bands do not correspond well to the critical band partition (see the earlier section on auditory masking). At a 48 kHz sampling rate each band has a width of 24000/32 = 750 Hz; hence, at low frequencies, a single subband covers a number of adjacent critical bands. The subband signals are resampled (critically decimated) at a rate of 1500 Hz. The impulse response of subband k, $h_{sub(k)}(n)$, is obtained by multiplication of the impulse response of a single *prototype lowpass filter*, $h(n)$, by a modulating function that shifts the lowpass response to the appropriate subband frequency range:

$$h_{sub(k)}(n) = h(n)\cos\left[\frac{(2k-1)}{2M} + \varphi(k)\right];$$

$$M = 32; k = 1,2,\ldots 32; n = 1,2,\ldots,512$$

The prototype lowpass filter $h(n)$ has a 3 dB bandwidth of 750/2 = 375 Hz, and the center frequencies are at odd multiples thereof (all values at 48 kHz sampling rate). Therefore, the subsampled filter outputs exhibit a significant overlap. However, the design of the prototype filter and the inclusion of appropriate phase shifts in the

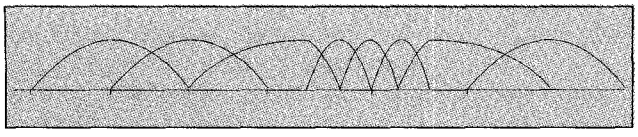
cosine terms result in an aliasing cancellation at the output of the decoder synthesis filterbank. Details about the coefficients of the prototype filter and the phase shifts $\varphi(k)$ are given in the ISO/MPEG standard. Details about an efficient implementation of the filterbank can be found in [15] and [36], and, again, in the standardization documents.

Quantization

The number of quantizer levels for each spectral component is obtained from a dynamic bit-allocation rule that is controlled by a psychoacoustic model. The bit-allocation algorithm selects one uniform midtread quantizer out of a set of available quantizers such that both the bit-rate requirement and the masking requirement are met. The iterative procedure minimizes the NMR in each subband. It starts with the number of bits for the samples and scalefactors set to zero. In each iteration step the quantizer SNR, SNR(m), is increased for the one m-bit subband quantizer producing the largest value of the NMR at the quantizer output. (The increase is obtained by allocating one more bit). For that purpose, the NMR, NMR(m) = SMR - SNR(m), is calculated as the difference (in dB) between the actual quantization noise level and the minimum global masking threshold. The standard provides tables with estimates for the quantizer SNR, SNR(m), for a given m.

Block companding is used in the quantization process, i.e., blocks of decimated samples are formed and divided by a *scalefactor* such that the sample of largest magnitude is unity. In Layer I, blocks of 12 decimated and scaled samples are formed in each subband (and for the left and right channel) and there is one bit allocation for each block. At a 48 kHz sampling rate, 12 subband samples correspond to 8 ms of audio. There are 32 blocks, each with 12 decimated samples, representing $32 \times 12 = 384$ audio samples.

In Layer II, in each subband a 36-sample *superblock* is formed of three consecutive blocks of 12 decimated samples corresponding to 24 ms of audio at 48 kHz sampling rate. There is one bit allocation for each 36-sample superblock. All 32 superblocks, each with 36 decimated samples, represent, altogether, $32 \times 36 = 1152$ audio samples. As in Layer I, a scalefactor is computed for each 12-sample block. A redundancy reduction technique is used for the transmission of the scalefactors: depending on the significance of the changes between the three consecutive scalefactors, one, two, or all three scalefactors are transmitted, together with a 2-bit *scalefactor select information*. Compared with Layer I, the bit rate for the scalefactors is reduced by around 50% [29]. Figure 9 indicates the block companding structure.



▲ *15. Typical sequence of windows in adaptive window switching.*

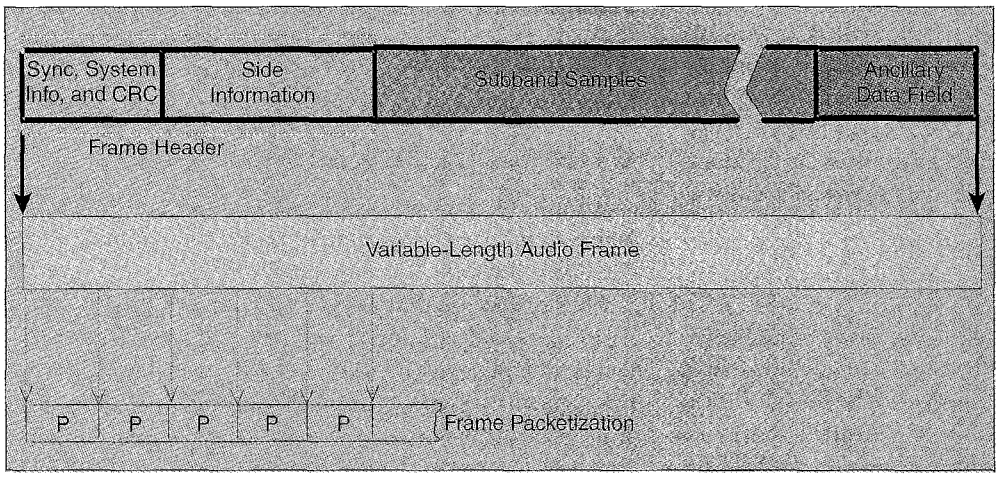
The coded bit streams must allow editing, fading, mixing, and dynamic range compression.

The scaled and quantized spectral subband components are transmitted to the receiver together with scalefactor, scalefactor select (Layer II), and bit-allocation information. Quantization with block companding provides a very large dynamic range of more than 120 dB. For example, in Layer II uniform midtread quantizers are available with 3, 5, 7, 9, 15, 31, ..., 65535 levels for subbands of low index (low frequencies). In the mid- and high-frequency region the quantizers have a reduced number of levels. For example, subbands of index 24 to 27 have only quantizers with 3, 5, and 65535 (!) levels. The 16-bit quantizers prevent overload effects. Subbands of index 28 to 32 are not transmitted at all. In order to reduce the bit rate, the codewords of three successive subband samples resulting from quantizing with 3-, 5-, and 9-step quantizers are assigned one common codeword. The savings in bit rate is about 40% [29].

Coding Examples

The following figures demonstrate the way MPEG-1 Layer II encodes audio signals. Figure 10 shows an individual 1152-sample audio block to be coded. Figure 11 shows the frequency dependencies of various important MPEG parameters; the frequency axes are divided in accordance with the subband separations. The sampling rate is 48 kHz, hence each subband index represents a subband of bandwidth 750 Hz. We have chosen an overall bit rate of 128 kb/s.

Figure 11(a) shows the frequency distribution of the sound-pressure level of the audio block. From this distribution and from the threshold in quiet a global masking threshold can be derived (Fig. 11(b)). For each subband, the SMR (in dB) is the difference between the level of the masker and the minimum value of the global masking threshold (Fig. 11(c)). Note that, for subbands of index 23 and higher, the signal power is significantly below that of the global masking threshold. Accordingly, the corresponding subband signals need not be transmitted. In the next step, the number of bits per subband quantizer is chosen such that its quantization noise is kept sufficiently below the global masking threshold (Fig. 11(d)). Therefore, the bit allocation for those subbands, which have to be transmitted, roughly follows the SMR. The spectrum of the reconstruction error is shown in Fig. 11(e) (Please take into account that the dB values of Fig. 11(e) cover only the range 0 to 35 dB). If we compare it with the spectrum of the global masking threshold, we note that the power of the reconstruction error is below the threshold,



▲ *16. MPEG-1 frame structure and packetization. Layer I: 384 subband samples; Layer II: 1152 subband samples. Packets P: 4-byte header; 184-byte payload field*

consequently, it is masked. Note also that the spectrum of the reconstruction error is identical to that of the input spectrum for subbands 23 and above because the corresponding subband signals are not transmitted.

In Fig. 12, three bit allocations for the same audio block are compared, employing (i) the psychoacoustic model 1 of the standard, (ii) the psychoacoustic model 2 of the standard, and (iii) the unweighted bit allocation. In this example, there are clear differences between the two models suggested in the MPEG standard. Note that the calculation of the unweighted bit allocation does not depend on masking thresholds. Nevertheless, the bit allocation resembles that of model 1, except that the unweighted bit allocation spends 3 bits for subbands of indices 23 and above where the signal power is well below that of the masking threshold.

Finally, in Fig. 13, we compare the model 1-based bit allocations for bit rates of 128 kb/s and 64 kb/s, again for the same audio block. Note that, at the lower bit rate, a lowpass version of the audio signal is reconstructed.

Decoding

The decoding is straightforward: the subband sequences are reconstructed on the basis of blocks of 12 subband samples taking into account the decoded scalefactor and bit-allocation information. If a subband has no bits allocated to it, the samples in that subband are set to zero. Each time the subband samples of all 32 subbands have been calculated, they are applied to the *synthesis filterbank*, and 32 consecutive 16-bit, PCM-format audio samples are calculated. If available, as in bidirectional communications or in recorder systems, the encoder (analysis) filterbank can be used in a reverse mode in the decoding process.

Layer III

Layer III of the MPEG-1/Audio coding standard introduces many new features (see Fig. 14), in particular a switched hybrid filterbank. 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. Layer III is the only layer that provides mandatory decoder support for variable bit-rate coding [37].

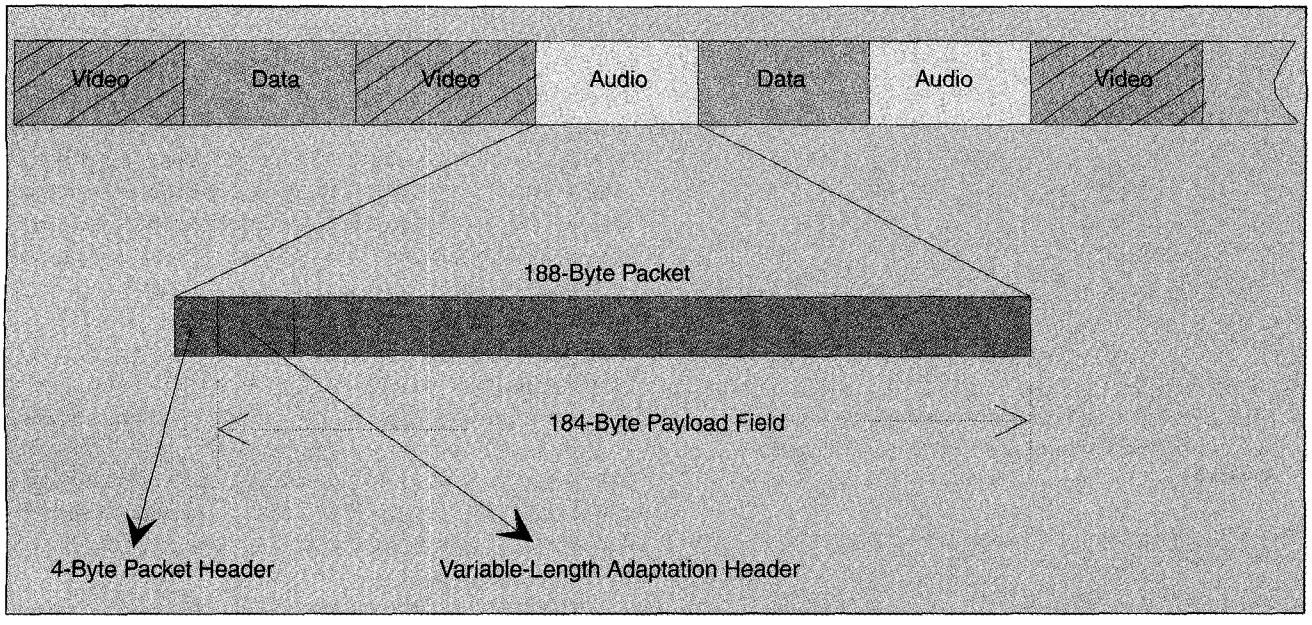
Switched Hybrid Filterbank

In order to achieve a higher frequency resolution closer to critical band partitions, the 32 subband signals are subdivided 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, respectively, 12 or 36 subband samples. The maximum number of frequency components is $32 \times 18 = 576$, each representing a bandwidth of only $24000/576 = 41.67$ Hz. The 18-point block transform is normally applied because it provides better frequency resolution, whereas the 6-point block transform provides better time resolution and is applied in case of expected pre-echoes (see the earlier section on window switching). In principle, a pre-echo is assumed when an instantaneous demand for a high number of bits occurs. Depending on the nature of potential pre-echoes, all or a smaller number of transforms are switched. Two special MDCT windows, a start window and a stop window, are needed in case of transitions between short and long blocks and vice versa to maintain the time-domain alias cancellation feature of the MDCT [21, 24, 36]. Figure 15 shows a typical sequence of windows.

Quantization and Coding

The MDCT output samples are nonuniformly quantized, thus providing both smaller mean-squared errors and masking because larger errors can be tolerated if the samples to be quantized are large. Huffman coding, based on 32 code tables and additional run-length coding, are applied to represent the quantizer indices in an efficient way. The encoder maps the variable wordlength codewords of the Huffman code tables into a constant bit rate by monitoring the state of a bit reservoir. The bit reservoir ensures that the decoder buffer neither underflows or overflows when the bitstream is presented to the decoder at a constant rate.

In order to keep the quantization noise in all critical bands below the global masking threshold (noise allocation) an *iterative analysis-by-synthesis method* is employed whereby the process of scaling, quantization, and coding



▲ *17. MPEG packet delivery.*

of spectral data is carried out within two nested iteration loops. The decoding follows that of the encoding process.

Frame and Multiplex Structure

Frame Structure

Figure 16 shows the frame structure of MPEG-1 audio-coded signals for both Layer I and Layer II. Each frame has a header; its first part contains 12 synchronisation bits, 20-bit system information, and an optional 16-bit cyclic redundancy check code. Its second part contains side information about the bit allocation and the scalefactors (and, in Layer II, scalefactor select information). As main information a frame carries a total of 32 × 12 subband samples (corresponding to 384 PCM audio input samples—equivalent to 8 ms at a sampling rate of 48 kHz) in Layer I, and a total of 32 × 36 subband samples in Layer II (corresponding to 1152 PCM audio input samples—equivalent to 24 ms at a sampling rate of 48 kHz). Please note that the frames are autonomous: each frame contains all information necessary for decoding. Therefore each frame can be decoded independently from previous frames—it defines an entry point for audio storage and audio editing applications. Please note also that the lengths of the frames are not fixed due to (i) the length of the main information field, which depends on bit-rate and sampling frequency; (ii) the side information field, which varies in Layer II; and (iii) the ancillary data field, the length of which is not specified.

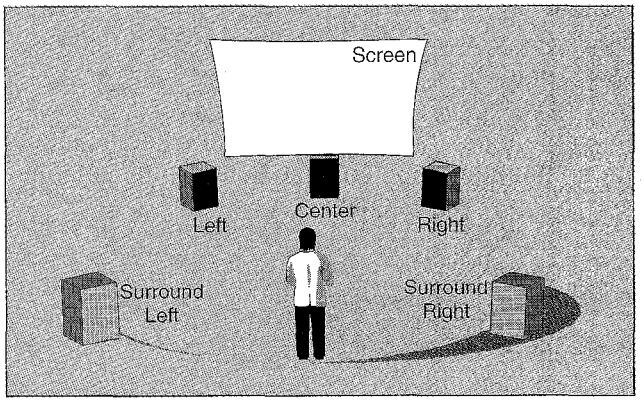
Multiplex Structure

The systems part of the MPEG-1 coding standard IS 11172 defines a packet structure for multiplexing audio, video, and ancillary data bitstreams in one stream. The variable-length MPEG frames are broken down into packets. The packet structure uses 188-byte packets consisting of a 4-byte header followed by 184 bytes of payload (see Fig. 17). The header includes a sync byte, a 13-bit field called the packet identifier to inform the decoder about the type of data, and additional information. For example, a 1-bit payload unit start indicator indicates if the payload starts with a frame header. No predetermined mix of audio, video, and ancillary data bitstreams is required; the mix may change dynamically; and services can be provided in a very flexible way. If additional header information is required such as for periodic synchronization of audio and video timing, a variable-length *adaptation header* can be used as part of the 184-byte payload field.

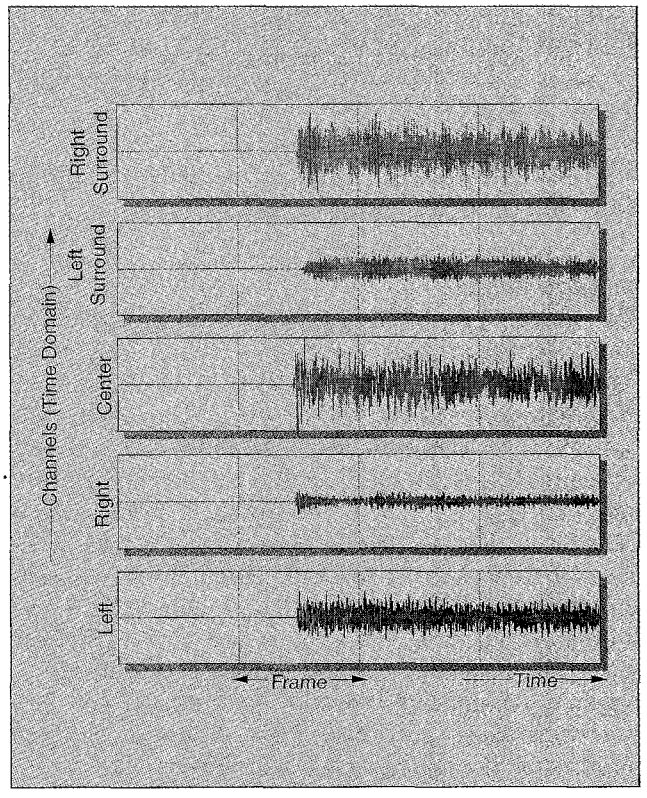
Although the lengths of the frames are not fixed, the interval between frame headers is constant (within a byte) throughout the use of padding bytes. The MPEG/Systems specification describes how MPEG-compressed

Table 4. Some multichannel loudspeaker configurations.

1 channel	1/0 configuration	center (monophonic)
2 channels	2/0 configuration	left, right (stereophonic)
3 channels	3/0 configuration	left, right, center
4 channels	3/1 configuration	left, right, center, mono surround
5 channels	3/2 configuration	left, right, center, surround left, surround right



▲ *18. 3/2 multichannel loudspeaker configuration (© Deutsche Telekom AG, Highlights aus der Forschung, 1996, with permission).*



▲ *19. Triangle sound representation in five channels (from top: RS, LS, C, R, and L) (©Deutsche Telekom AG, Highlights aus der Forschung, 1996, with permission).*

audio and video data streams are to be multiplexed together to form a single data stream. The terminology and the fundamental principles of the systems layer are described in [38].

Subjective Quality (MPEG-1; Stereophonic Audio Signals)

The standardization process included extensive subjective tests and objective evaluations of parameters such as complexity and overall delay. The MPEG (and equivalent ITU-R) listening tests were carried out under very similar and carefully defined conditions with around 60 experienced listeners, approximately 10 test sequences were used, and the sessions were performed in stereo with both loudspeakers and headphones. In order to detect even small impairments the 5-point ITU-R impairment scale was used in all experiments. Details are given in [39] and [40]. Critical test items were chosen in the tests to evaluate the coders by their *worst case* (not average) performance. The subjective evaluations, which had been based on triple stimulus/hidden reference/double blind tests, have shown very similar and stable evaluation results. In these tests the subject is offered three signals, A, B, and C (triple stimulus). A is always the unprocessed source signal (the reference). B and C, or C and B, are the reference and the system under test (hidden reference). The selection is neither known to the subjects nor to the conductor(s) of the test (double blind test). The subjects have to decide if B or C is the reference and have to grade the remaining one.

The MPEG-1 audio-coding standard has shown excellent performance for all layers at the rates given in Table 3. It should be mentioned again that the standard leaves room for encoder-based improvements by using better psychoacoustic models. Indeed, many improvements have been achieved since the first subjective tests had been carried out in 1991.

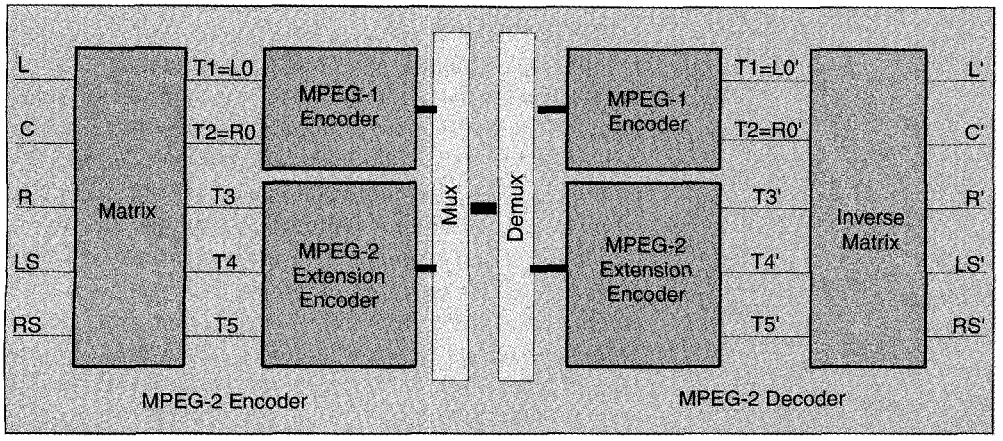
MPEG Audio Coding with Lower Sampling Frequencies

We have mentioned above that MPEG-1/Audio supports sampling frequencies of 32, 44.1, and 48 kHz. For applications with limited bandwidths (mediumband), lower sampling frequencies (16, 22.05, and 24 kHz) have been defined in MPEG-2 to bring bit rates down to 64 kb/s per channel and less [9]. The corresponding maximum audio bandwidths are 7.5, 10.3, and 11.25 kHz. The syntax, semantics, and coding techniques of MPEG-1 are maintained except for a small number of parameters (two tables in the decoder). Therefore, coding can be based again on Layers I, II, or III. The extension to lower sampling frequencies leads to higher frequency resolutions and hence to higher coding gains, partly because of better adaptations to the masking thresholds, and partly because side information becomes a smaller part of the overall bit rate. As in the case of coding wideband audio signals, the best audio quality is obtained with Layer III. Finally we note that some applications make use of sampling frequencies of 8, 11.025, and 12 kHz, which are outside the MPEG-2 standard.

MPEG Multichannel Audio Coding

Multichannel Audio Representations

A logical further step in digital audio is the definition of multichannel audio representation systems to create a realistic surround-sound field both for audio-only applications and for audiovisual systems, including video



▲ *20. Compatibility of MPEG-2 multichannel audio bit streams.*

conferencing, videophony, multimedia services, and electronic cinema. Multichannel systems can also provide multilingual channels or additional channels for visually impaired (a verbal description of the visual scene) and for hearing impaired (dialogue with enhanced intelligibility). ITU-R and other international groups have recommended a five-channel loudspeaker configuration, referred to as 3/2-stereo, with a left and a right channel (L and R), an additional center channel (C) and two side/rear surround channels (LS and RS) augmenting the L and R channels (see Fig. 18) (ITU-R Rec. 775). Such a configuration offers a surround-sound field with a stable frontal sound image and a large listening area. Figure 19 shows four blocks of a five-channel triangle audio signal (which is difficult to code).

Multichannel digital audio systems support p/q presentations with p front and q back channels, and also provide the possibilities of transmitting two independent stereophonic programs and/or a number of commentary or multilingual channels. Typical combinations of channels are given in Table 4.

ITU-R Recommendation 775 provides a set of downwards mixing equations if the number of loudspeakers is to be reduced (*downwards compatibility*). An additional low-frequency enhancement (LFE or subwoofer) channel, particularly useful for HDTV applications, can be optionally added to any of the configurations. The LFE channel extends the low-frequency content between 15 Hz and 120 Hz in terms of both frequency and level. One or more loudspeakers can be positioned freely in the listening room to reproduce this LFE signal. The film industry uses a similar system for their digital sound systems. A 3/2-configuration with five high-quality full-range channels plus a subwoofer channel is often called a 5.1 system.

In order to reduce the overall bit rate of multichannel audio-coding systems, redundancies and irrelevancy, such as interchannel dependencies and interchannel masking effects, respectively, may be exploited. In addition, components of the multichannel signal, which are irrelevant with respect to the spatial perception of the stereophonic presentation, i.e., those that do not contribute to the localization of sound sources, may be identified and reproduced in a monophonic format to further reduce bit rates. State-of-the-art multichannel coding algorithms make use of such effects. However, a careful design is needed, otherwise such joint coding may produce artifacts.

MPEG-2/Audio Multichannel Coding

The second phase of MPEG, labeled MPEG-2, includes in its audio part two multichannel audio-coding standards, one of which is forward- and backwards compatible with MPEG-1/Audio [8, 41-44]. *Forward compatibility* means that an MPEG-2 multichannel decoder is able to properly decode MPEG-1 mono- or stereophonic signals; *backwards compatibility* means that existing MPEG-1 stereo decoders, which only handle two-channel audio, are able to reproduce a meaningful basic 2/0 stereo signal from a MPEG-2 multichannel bit stream so as to serve the need of users with simple mono or stereo equipment. *Nonbackwards-compatible* multichannel coders will not be able to feed a meaningful bit stream into an MPEG-1 stereo decoder. On the other hand, nonbackwards-compatible codecs have more freedom in producing a high-quality reproduction of audio signals.

With backwards compatibility it is possible to introduce multichannel audio at any time in a smooth way without making existing two-channel stereo decoders obsolete. An important example is the European Digital Audio Broadcast system, which will require MPEG-1 stereo decoders in the first generation but may offer multichannel audio at a later point.

Backwards-Compatible MPEG-2 Audio Coding

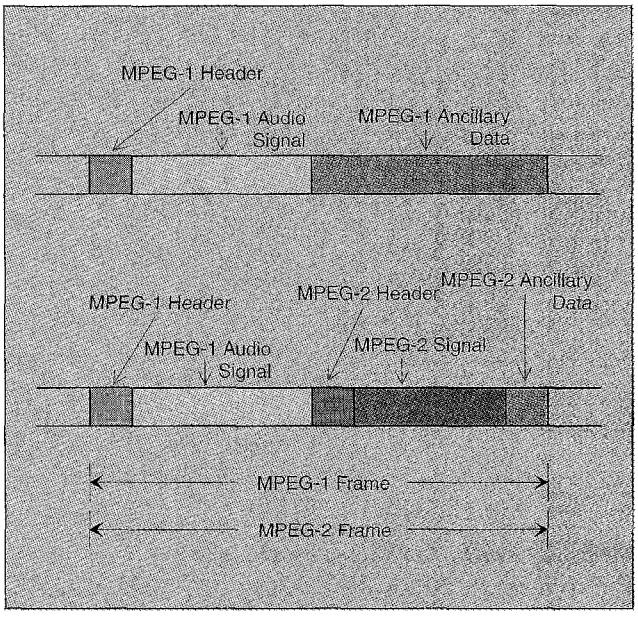
Backwards compatibility implies the use of compatibility matrices. A down-mix of the five channels ("matrixing") delivers a correct basic 2/0 stereo signal, consisting of a left and a right channel, L0 and R0, respectively. A typical set of equations is:

$$L0 = \alpha(L + \beta \cdot C + \delta \cdot LS)$$

$$\alpha = \frac{1}{1+\sqrt{2}};\beta = \delta = \sqrt{2}$$

$$R0 = \alpha(R + \beta \cdot C + \delta \cdot RS)$$

Other choices are possible, including L0 = L and R0 = R. The factors, α, β, and δ, attenuate the signals to avoid overload when calculating the compatible stereo signal



▲ *21. Data format of MPEG audio bit streams: (a) MPEG-1 audio frame; (b) MPEG-2 audio frame, compatible with MPEG-1 format.*

(L0,R0). L0 and R0 are transmitted in MPEG-1 format in transmission channels T1 and T2. Channels T3, T4, and T5 together form the *multichannel extension signal* (Fig. 20). They have to be chosen such that the decoder can recompute the complete 3/2-stereo multichannel signal. Interchannel redundancies and masking effects are taken into account to find the best choice. A simple example is T3 = C, T4 = LS, and T5 = RS. In MPEG-2 the matrixing can be done in a very flexible and even time-dependent way. Note , however, that the audio content of the extension signal is already delivered in the MPEG-1 audio stream (signals L0 and R0); this redundancy reduces the achievable compression rate.

Backwards compatibility is achieved by transmitting the channels L0 and R0 in the subband-sample section of the MPEG-1 audio frame and all multichannel extension signals (T3, T4, and T5) in the first part of the MPEG-1 audio frame reserved for ancillary data. This ancillary data field is ignored by MPEG-1 decoders (see Fig. 21). The length of the ancillary data field is not specified in the standard. If the decoder is of type MPEG-1, it uses the 2/0-format front left and right down-mix signals, L0' and R0', directly (see Fig. 22). If the decoder is of type MPEG-2, it recomputes the complete 3/2-stereo multichannel signal with its components L', R', C', LS', and RS' via "dematrixing" of L0', R0', T3', T4', and T5' (see Fig. 20).

Matrixing is obviously necessary to provide backwards compatibility; however, if used in connection with perceptual coding, "unmasking" of quantization noise may appear [45]. It may be caused in the dematrixing process when sum and difference signals are formed. In certain situations such a masking sum or difference signal component can disappear in a specific channel. Since this component was supposed to mask the quantization noise in that channel, this noise may become audible. Note that the masking signal will still be present in the multichannel representation but it will appear on a different loudspeaker. Measures against "unmasking" effects have been described in [46]. As an additional measure, MPEG-2's optional variable bit-rate mode can be evoked to encode difficult audio content at a momentarily higher bit rate.

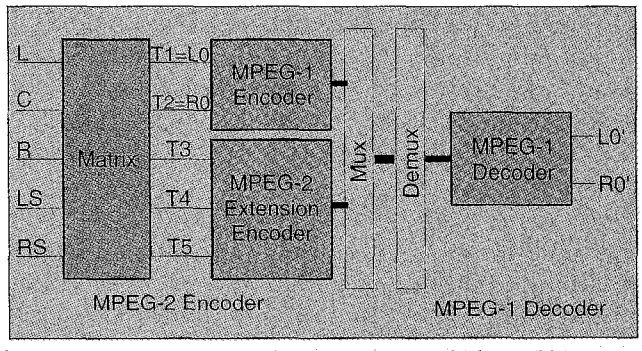
MPEG-1 decoders have a bit-rate limitation (384 kb/s in Layer II). In order to overcome this limitation, the MPEG-2 standard allows for a second bit stream, the extension part, to provide compatible multichannel audio at higher rates. Figure 23 shows the structure of the bit stream with extension.

MPEG-2 Advanced Audio Coding

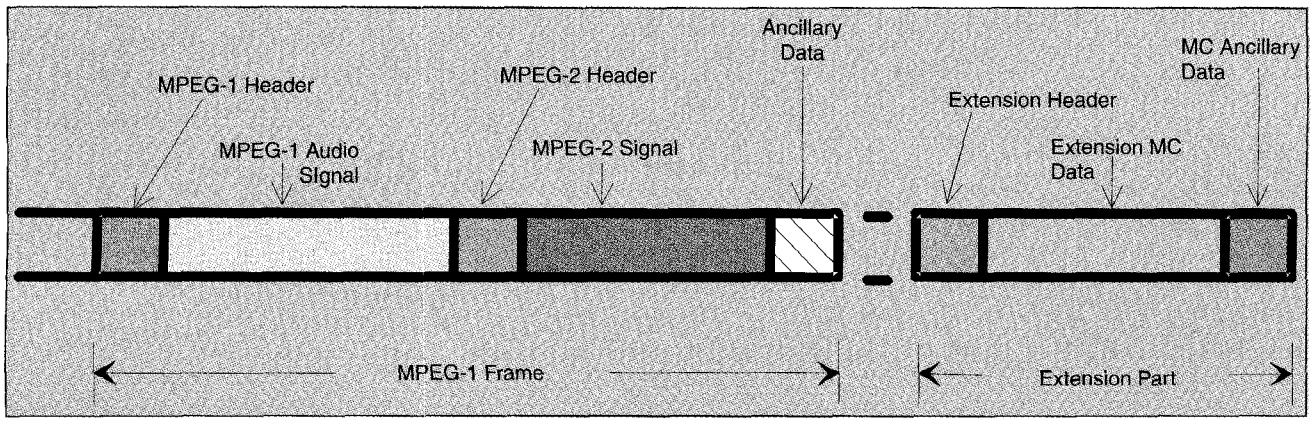
A second standard within MPEG-2 supports applications that do not request compatibility with the existing MPEG-1 stereo format. Therefore, matrixing and dematrixing are not necessary and the corresponding potential artifacts disappear (see Fig. 24).

The last two years have seen extensive activities in the optimization and standardization of a nonbackwards-compatible MPEG-2 multichannel audio coding algorithm. Many companies around the world contributed advanced audio-coding algorithms in an collaborative effort to come up with a flexible high-quality coding standard [43].

Tools. The MPEG-2 AAC standard employs high-resolution filter banks, prediction techniques, and noiseless coding. 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, prediction, M/S stereo coding, quantization, noiseless coding, and a bit-stream multiplexer (see Fig. 25). The filterbank is a 1024-line modified discrete cosine transform and the perceptual model is taken from MPEG-1 (model 2). The temporal noise shaping tool controls the time dependence of the quantization noise, intensity, and M/S coding and the second-order backward-adaptive predictor improves coding efficiency.



▲ *22. MPEG-1 stereo decoding of MPEG-2 multichannel bit stream.*



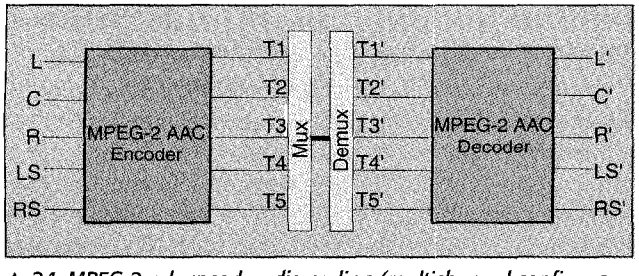
▲ *23. Data format of MPEG-2 audio bit stream with extension part for multichannel data.*

The predictor reduces the bit rate for coding subsequent subband samples in a given subband, and it bases its prediction on the quantized spectrum of the previous block, which is also available in the decoder (in the absence of channel errors). Finally, for quantization and noiseless coding, an iterative method is employed so as to keep the quantization noise in all critical bands below the global masking threshold.

Profiles. In order to serve different needs, the standard provides three profiles: (i) the main profile offers highest quality, (ii) the low-complexity profile works without prediction, and (iii) the sampling-rate-scaleable profile offers the lowest complexity. For example, in its *main profile* the filterbank is a 1024-line MDCT with 50% overlap (block length of 2048 samples). The filterbank is switchable to eight 128-line MDCTs (block lengths of 256 samples). Hence, it allows for a frequency resolution of 23.43 Hz and a time resolution of 2.6 ms (both at a sampling rate of 48 kHz). In the case of the long block length, the window shape can vary dynamically as a function of the signal.

The *low-complexity profile* does not employ temporal noise shaping and time-domain prediction (the prediction adds significantly to the complexing), whereas in the *sampling-rate-scaleable profile* a hybrid filterbank is used.

MPEG-2 AAC supports up to 46 channels for various multichannel loudspeaker configurations and other applications; the default loudspeaker configurations are the monophonic channel, the stereophonic channel, and the 5.1 system (five channels plus LFE channel).



▲ *24. MPEG-2 advanced audio coding (multichannel configuration).*

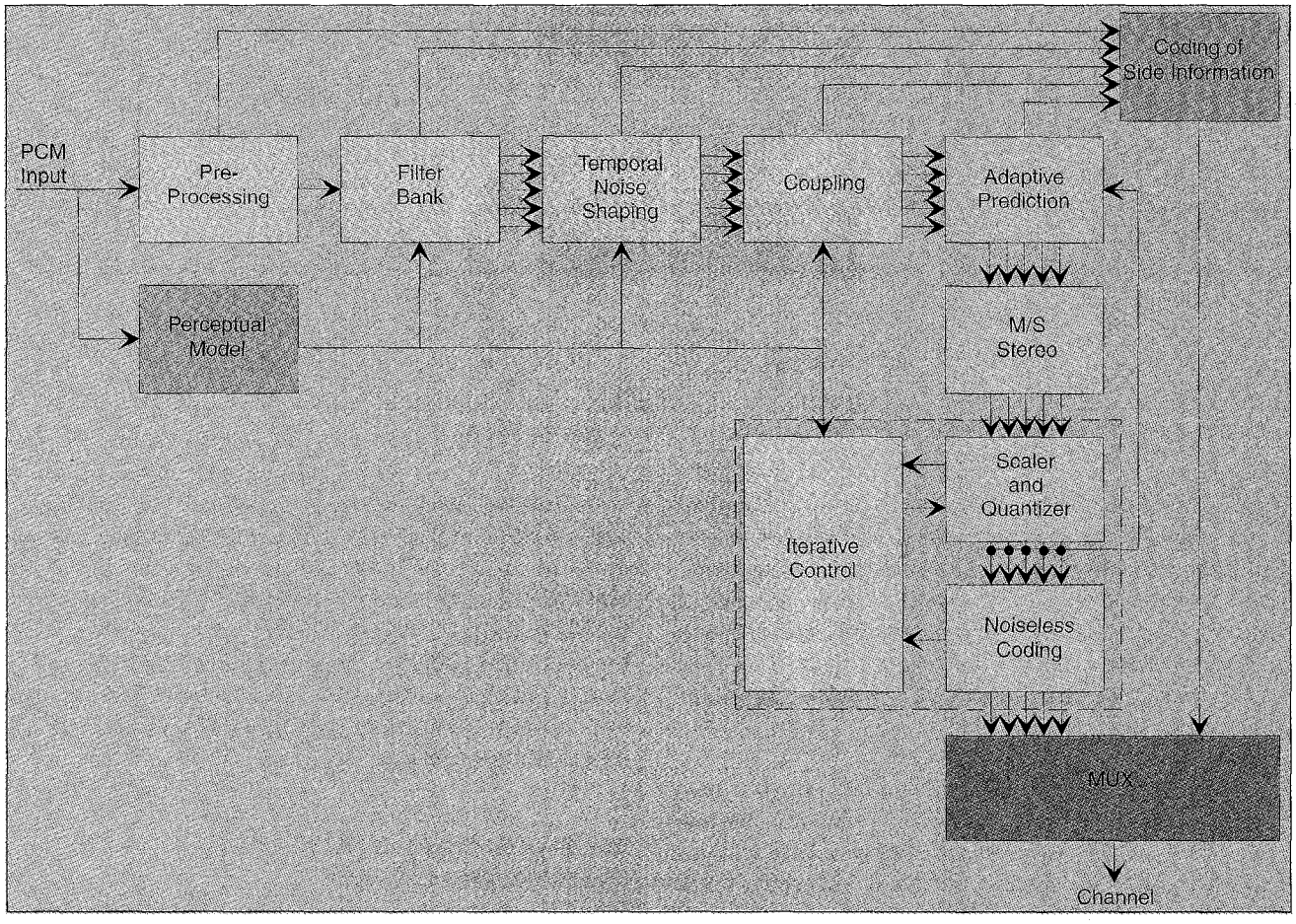
The above-listed selected modules define the MPEG-2 audio AAC standard that became an International Standard in April 1997 as an extension to MPEG-2 (ISO/MPEG 13818 - 7). A more detailed description of the MPEG-2 AAC multichannel standard can be found in the literature [43]. The standard offers high quality at the lowest possible bit rates between 320 and 384 kb/s for five channels; it will find many applications in both consumer and professional use.

Backwards Compatibility via Simulcast Transmission

If bit rates are not of high concern, a *simulcast transmission* may be employed where a full MPEG-1 bitstream is multiplexed with a full nonbackwards-compatible multichannel bit stream in order to support backwards compatibility without matrixing techniques (Fig. 26).

Subjective Tests (MPEG-2, Multichannel Audio Signals)

The first subjective tests, independently run at German Telekom and BBC (UK) under the umbrella of the MPEG-2 standardization process, had shown a satisfactory average performance of *nonbackwards-compatible* and of *backwards-compatible* coders. The tests had been carried out with experienced listeners and critical test items at low bit rates (320 and 384 kb/s). However, all codecs showed significant deviations from transparency for some of the test items [47, 48]. Recently, extensive formal subjective tests have been carried out to compare MPEG-2 AAC versions, operating, respectively, at 256 and 320 kb/s, and a backward-compatible MPEG-2 Layer II coder, operating at 640 kb/s [49] (a 1995 version of this latter coder was used, therefore its test results do not reflect any subsequent enhancements). All coders performed very well with a slight advantage to the nonbackwards-compatible 320 kb/s MPEG-2 AAC coder compared with the backwards-compatible 640 kb/s MPEG-2 Layer II coder. The performances of those coders are indistinguishable from the original in the sense of the EBU definition of *indistinguishable quality* [50].



▲ *25. Structure of MPEG-2 advanced audio coder (AAC).*

From these subjective tests, it has become clear that the concept of backwards compatibility implies the need for higher bit rates.

MPEG-4 Audio Coding

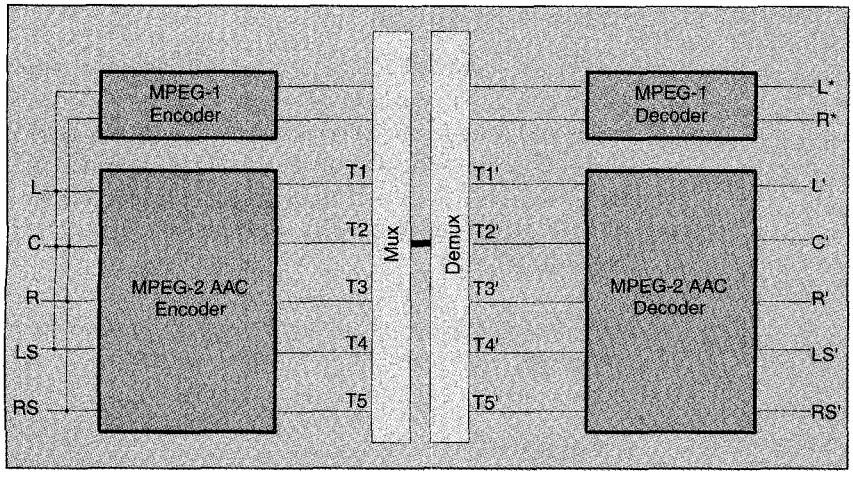
Activities within MPEG-4 aim at proposals for a broad field of applications including multimedia. (We note in passing that MPEG has started a new work item called "Multimedia content description interface" (in short "MPEG-7"). MPEG-7 does not cover coding, its goal is rather to specify a standardized description of various types of multimedia information. A typical application will be the search for video, graphics, or audio material in the sense of today's text-based search engines in the World Wide Web.) MPEG-4 will offer higher compression rates, and it 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. In order to represent, integrate, and exchange pieces of audio-visual information, MPEG-4 offers standard tools that can be combined to satisfy specific user requirements [51]. A number of such configurations may be standardized. A syntactic description will be 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.

The current toolset supports audio and speech compression at monophonic bit rates ranging from 2 to 64 kb/s. Three *core coders* are used:

- a parametric coding scheme for low bit rate speech coding (2 to 10 kb/s)
- an analysis-by-synthesis coding scheme for medium bit rates (6 to 16 kb/s)
- a subband/transform-based coding scheme for bit rates below 64 kb/s.

The three core coders have been integrated into a so-called verification model that describes the operations of encoders and decoders and that is used to carry out simulations and optimizations. In the end, the verification model will be the embodiment of the standard [51].

Let us also note that MPEG-4 will offer new functionalities such as time-scale changes, pitch control, editability, database access, and scalability, which allows one to extract from the transmitted bit stream a subset sufficient to generate audio signals with lower bandwidth and/or lower quality depending on channel capacity or decoder complexity. MPEG-4 will become an international standard in November 1998.



▲ *26. Backwards-compatible MPEG-2 multichannel audio coding (simulcast mode).*

Applications

MPEG/Audio compression technologies will play an important role in consumer electronics, professional audio, telecommunications, broadcasting, and multimedia. Here we describe a few typical application fields.

Main applications will be based on delivering digital audio signals over terrestrial and satellite-based digital *broadcast and transmission systems* such as subscriber lines, program exchange links, cellular mobile radio networks, cable-TV networks, and local area networks. [52]. For example, in narrowband integrated services digital networks (ISDN) customers have physical access to one or two 64-kb/s B channels and one 16-kb/s D channel (which supports signaling but can also carry user information). Other configurations are possible including $p \times 64$ kb/s (p = 1,2,3,...) services. ISDN rates offer useful channels for a practical distribution of stereophonic and multichannel audio signals. Because ISDN is a bidirectional service, it also provides upstream paths for future on-demand and interactive audiovisual *just-in-time* audio services. The backbone of digital telecommunication networks will be broadband (B-) ISDN with its cell-oriented structure. Cell delays and cell losses are sources of distortions to be taken into account in designs of digital audio systems [53]. A related application is Internet broadcasting, which will need significant compression rates as long as home computers are connected to the backbone networks via modems with typical bit rates between 14.4 kb/s and 33 kb/s.

In the field of digital storage on digital audio tape and (re-writeable) disks, a number of MPEG-based consumer products have recently reached the audio market. Of these products, Philips' Digital Compact Cassette (DCC) essentially makes use of Layer I of the MPEG-1 audio coder by employing its 384 kb/s stereo rate; its audio-coding algorithm is called PASC (precision audio subband coding) [15]. The upcoming DVD with its capacity of 4.7 GB relieves the pressure for extreme compression factors. It will open the possibilities of storing audio channels that have been coded in a lossless mode, and it will provide the necessary capacity for various forms of multichannel coding. The DVD will support stereophonic and (at least) 5.1-multichannel audio. In connection with video, the PAL version of the DVD (5625/50 TV system) will use MPEG audio coding with Dolby's AC-3 transform coding technique as an option [54-56], whereas the NTSC version (525/60 TV system) will be based on AC-3 with MPEG as an option. The overall audio bit rate is 384 kb/s.

A number of decisions concerning the introduction of digital audio broadcast (DAB) and digital video broadcast (DVB) services have been made recently. In Europe, a project group named Eureka 147 has worked out a DAB system that is able to cope with the problems of digital broadcasting [57]. ITU-R has recommended the MPEG-1 audio-coding standard after it had made extensive subjective tests. Layer II of this standard is used for program emission, and the Layer III version is recommended for commentary links at low rates. The sampling rate is 48 kHz in all cases, and the ancillary data field is used for program-associated data (PAD information) and other data. The DAB system has a significant bit-rate overhead for error correction based on punctured convolutional codes in order to support *source-adapted channel coding*, i.e., an unequal error protection that is in accordance with the sensitivity of individual bits or a group of bits to channel errors [58]. Additionally, error-concealment techniques are applied to provide a *graceful degradation* in case of severe errors. In the United States a standard has not yet been defined. Simulcasting analog and digital versions of the same audio program in the FM terrestrial band (88-108 MHz) is an important issue (whereas the European solution is based on new channels) [59].

The Hughes DirecTV satellite subscription television system and ADR (Astra Digital Radio) are examples of satellite-based digital broadcasting that make use of MPEG-1 Layer II. As a further example, the Eutelsat SaRa system will be based on Layer III coding.

Advanced digital TV systems provide HDTV delivery to the public by terrestrial broadcasting and a variety of alternate media and offer full-motion, high-resolution video and high-quality multichannel surround audio. The overall bit rate may be transmitted within the bandwidth of an analog UHF television channel. The United States Grand Alliance HDTV system and the European DVB system both make use of the MPEG-2 video-compression system and of the MPEG-2 system transport layer, which uses a flexible ATM-like packet protocol with headers/de-

scriptors for multiplexing audio and video bit streams in one stream with the necessary information to keep the streams synchronized when decoding (see Fig. 17). The systems differ in the way the audio signal is compressed: The Grand Alliance system will use Dolby's AC-3 algorithm, whereas the European system will use MPEG-2/Audio.

Conclusions

Low bit-rate digital audio is applied in many different fields, such as consumer electronics, professional audio processing, telecommunications, and broadcasting. Perceptual coding in the frequency domain has paved the way to high compression rates in audio coding. MPEG-1 audio coding with its three layers has been widely accepted as international standard. Software encoders, single DSP chip implementations, and computer extensions are available from a number of suppliers.

In the area of broadcasting and mobile radio systems, services are moving to portable and handheld devices, and new, third-generation mobile communication networks are evolving. Coders for these networks must not only operate at low bit rates but must be stable in burst-error and packet- (cell-) loss environments. Error-concealment techniques play a significant role. Due to the lack of available bandwidth, traditional channel coding techniques may not be able to sufficiently improve the reliability of the channel.

MPEG audio coders are controlled by psychoacoustic models that may be improved, thus leaving room for an evolutionary improvement of codecs. In the future, we will see new solutions for encoding. A better understanding of binaural perception and of stereo representation will lead to new proposals.

Digital multichannel audio improves stereophonic images and will be of importance both for audio-only and multimedia applications. MPEG-2/Audio offers both backwards-compatible and nonbackwards-compatible coding schemes to serve different needs. Ongoing research will result in enhanced multichannel representations by making better use of interchannel correlations and interchannel masking effects to bring the bit rates further down. We can also expect solutions for special presentations for people with impairments of hearing or vision which can make use of the multichannel configurations in various ways.

Current activities of the ISO/MPEG expert group aim at proposals for audio coding that will offer higher compression rates, and which will merge the whole range of audio, from high-fidelity audio coding and speech coding down to synthetic speech and synthetic audio (ISO/IEC MPEG-4). MPEG-4 will be the future multimedia standard. Because the basic audio quality will be more important than compatibility with existing standards, this activity has opened the door for completely new solutions.

Acknowledgments

The MPEG standards were created through the long-lasting efforts of a great many people coming from companies and research institutions from all over the world. Many of the contributors participated in the MPEG standards meetings. Their outstanding qualification, their enthusiasm, and countless collaborative efforts made these standards happen.

Peter Noll is a Professor of Telecommunications at the Technische Universität Berlin in Berlin, Germany.

References

1. A.A.M.L. Bruekers et al., "Lossless Coding for DVD Audio," 101st Audio Engineering Society Convention, Los Angeles, 1996, preprint 4358.
2. N. S. Jayant and P. Noll, "Digital Coding of Waveforms: Principles and Applications to Speech and Video," Prentice Hall, 1984.
3. A.S. Spanias, "Speech Coding: A Tutorial Review," *Proc. of the IEEE*, Vol. 82, No. 10, pp. 1541-1582, Oct. 94.
4. N.S. Jayant, J.D. Johnston, and Y. Shoham, "Coding of Wideband Speech," *Speech Communication* 11 (1992), pp. 127-138.
5. A. Gersho, "Advances in Speech and Audio Compression," *Proc. of the IEEE*, Vol. 82, No.6, pp. 900-918, 1994.
6. P. Noll, "Wideband Speech and Audio Coding," *IEEE Commun. Mag.*, Vol. 31, No. 11, pp. 34-44, 1993.
7. P. Noll, "Digital Audio Coding for Visual Communications," *Proc. of the IEEE*, Vol. 83, No. 6, June 1995.
8. ISO/IEC JTCl/SC29, "Information Technology - Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to About 1.5 Mbit/s - IS 11172 (Part 3, Audio)," 1992.
9. ISO/IEC JTCl/SC29, "Information Technology - Generic Coding of Moving Pictures and Associated Audio Information - IS 13818 (Part 3, Audio)," 1994.
10. ISO/MPEG, Doc. N0821, Proposal Package Description - Revision 1.0, Nov. 1994.
11. G.T. Hathaway, "A NICAM Digital Stereophonic Encoder," in "Audiovisual Telecommunications (Editor: N.D. Nigthingale), Chapman & Hall, 1992, pp. 71-84.
12. E. Zwicker and R. Feldtkeller, Das Ohr als Nachrichtenemfänger. Stuttgart: S. Hirzel Verlag, 1967.
13. N.S. Jayant, J.D. Johnston, and R. Safranek, "Signal compression based on models of human perception," *Proc. of the IEEE*, Vol. 81, No. 10, pp. 1385-1422, 1993.
14. R. Zelinski and P. Noll, "Adaptive Transform Coding of Speech Signals," *IEEE Trans. on Acoustics, Speech and Signal Proc.*, Vol. ASSP-25, pp. 299-309, August 1977.
15. A. Hoogendorn, "Digital compact cassette," *Proc. of the IEEE*, vol. 82, No. 10, pp. 1479-1489, Oct. 1994.
16. P. Noll, " On predictive quantizing schemes," *Bell System Technical Journal*, vol. 57, pp. 1499-1532, 1978.
17. J. Makhoul and M. Berouti, "Adaptive noise spectral shaping and entropy coding in predictive coding of speech." *IEEE Trans. on Acoustics, Speech, and Signal Processing* Vol. 27, No. 1, pp. 63-73, Feb. 1979
18. D. Esteban and C. Galand, "Application of Quadrature Mirror Filters to Split Band Voice Coding Schemes," *Proc. ICASSP*, pp. 191-195, 1987.
19. J.H. Rothweiler, "Polyphase Quadrature Filters, a New Subband Coding Technique," Proc. International Conference ICASSP'83, pp. 1280-1283, 1983.

20. J. Princen and A. Bradley, "Analysis/Synthesis Filterbank Design Based on Time Domain Aliasing Cancellation," *IEEE Trans. on Acoust. Speech, and Signal Process*. Vol. ASSP-34, pp.1153-1161, 1986.

21. H.S. Malvar, "Signal Processing with Lapped Transforms," Artech House Inc., 1992.

22. F.S. Yeoh, C.S. Xydeas, "Split-band coding of speech signals using a transform technique," *Proc. ICC*, 1984, Vol.3, pp. 1183-1187.

23 W. Granzow, P. Noll, C. Volmary, "Frequency-domain coding of speech signals," (in German), NTG-Fachbericht No. 94, VDE-Verlag, Berlin, 1986, pp. 150-155.

24. B. Edler, "Coding of Audio Signals with Overlapping Block Transform and Adaptive Window Functions," (in German), *Frequenz*, vol. 43, pp. 252-256, 1989.

25. M. Iwadare, A. Sugiyama, F. Hazu, A. Hirano, and T. Nishitani, "A 128 kb/s Hi-Fi Audio CODEC Based on Adaptive Transform Coding with Adaptive Block Size," *IEEE J. on Sel. Areas in Commun.*, Vol. 10, No. 1, pp. 138-144, January 1992.

26. R. Zelinski and P. Noll, "Adaptive Blockquantisierung von Sprachsignalen," Technical Report No. 181, Heinrich-Hertz-Institut für Nachrichtentechnik, Berlin, 1975.

27. R.G. van der Waal, K. Brandenburg, and G. Stoll, "Current and future standardization of high-quality digital audio coding in MPEG," *Proc. IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, N.Y., 1993.

28. P. Noll and D. Pan, "ISO/MPEG Audio Coding," *International Journal of High Speed Electronics and Systems*, Vol. 8, No. 1, pp. 69-118, 1997.

29. K. Brandenburg and G. Stoll, "The ISO/MPEG-Audio Codec: A Generic Standard for Coding of High Quality Digital Audio," *Journal of the Audio Engineering Society* (AES), Vol. 42, No. 10, pp. 780-792, Oct. 1994.

30. L.M. van de Kerkhof and A.G. Cugnini, "The ISO/MPEG audio coding standard," *Widescreen Review*, 1994.

31. Y.F. Dehery, G. Stoll, L. v.d. Kerkhof, "MUSICAM Source Coding for Digital Sound," 17th International Television Symposium, Montreux, Record pp. 612-617, June 1991.

32. K. Brandenburg, J. Herre, J. D. Johnston, Y. Mahieux, E.F. Schroeder: "ASPEC: Adaptive Spectral Perceptual Entropy Coding of High Quality Music Signals," *90th. Audio Engineering Societ Cyonvention*, Paris , preprint 3011, 1991.

33. H.G. Musmann, "The ISO Audio Coding Standard," Proc. IEEE Globecom, Dec. 1990.

34. R.G. van der Waal, A.W.J. Oomen, and F.A. Griffiths, "Performance comparison of CD, noise-shaped CD and DCC," *96th Audio Engineering Society Convention*, Amsterdam, preprint 3845, 1994.

35. J. Herre, K. Brandenburg, and D. Lederer, "Intensity stereo coding," *96th Audio Engineering Society Convention*, Amsterdam, preprint no. 3799, 1994.

36. D. Pan, "A Tutorial on MPEG/Audio Compression," *IEEE Trans. on Multimedia*, Vol. 2, No. 2, 1995, pp. 60-74.

37. K. Brandenburg et al., " Variable data-rate recording on a PC using MPEG-Audio Layer III," *95th Audio Engineering Society Convention*, New York, 1993.

38. P.A. Sarginson,"MPEG-2: Overview of the system layer," BBC Research and Development Report, BBC RD 1996/2, 1996.

39. T. Ryden, C. Grewin, S. Bergman, "The SR Report on the MPEG Audio subjective listening tests in Stockholm April/May 1991," ISO/IEC JTCl/SC29/WG ll: Doc.-No. MPEG 91/010, May1991.

40. H. Fuchs, "Report on the MPEG/Audio subjective listening tests in Hannover," ISO/IEC JTCl/SC29/WG 11: Doc.-No. MPEG 91/331, November 1991.

41. G. Stoll et al., "Extension of ISO/MPEG-Audio Layer II to multi-channel coding: The future standard for broadcasting, telecommunication, and multimedia application," *94th Audio Engineering Society Convention*, Berlin, Preprint no. 3550, 1993.

42. B. Grill et al., "Improved MPEG-2 audio multi-channel encoding," *96th Audio Engineering Society Convention*, preprint 3865, Amsterdam, 1994.

43. M. Bosi et al., "ISO/IEC MPEG-2 advanced audio coding," *101th Audio Engineering Society Convention*, preprint 4382, Los Angeles, 1996.

44. J.D. Johnston et al., "NBC-Audio-Stereo and multichannel coding methods," *101th Audio Engineering Society Convention*, Los Angeles, 1996, preprint 4383.

45. W.R.Th. Ten Kate, et al., "Matrixing of bit rate reduced audio signals," *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing* (ICASSP '92), Vol. 2, pp. II-205 - II-208, 1992.

46. W.R.Th. Ten Kate, "Compatibility matrixing of multi-channel bit-rate-reduced audio signals," *96th Audio Engineering Society Convention*, preprint 3792, Amsterdam, 1994.

47. F. Feige and D. Kirby, "Report on the MPEG/Audio Multichannel Formal Subjective Listening Tests, " ISO/IEC JTCl/SC29/WG 11: Doc. N 0685, March 1994.

48. D. Meares and D. Kirby, "Brief Subjective Listening Tests on MPEG-2 Backwards Compatible Multichannel Audio Codecs," ISO/IEC JTCl/SC29/WG 11: August 1994.

49. ISO/IEC/JTC1/SC29, "Report on the formal subjective listening tests of MPEG-2 NBC multichannel audio coding," Document N1371, Oct. 1996.

50. ITU-R Document TG 10-2/3, Oct. 1991.

51. IEC/JTC1/SC29, "Description of MPEG-4," Document N1410, Oct. 1996.

52. D.S. Burpee and P.W. Shumate, "Emerging residential broadband telecommunications," *Proc. IEEE*, Vol. 82, No.4, pp. 604-614, 1994.

53. N.S. Jayant, "High Quality Networking of Audio-visual Information," *IEEE Commun. Mag.*, pp. 84-95, 1993.

54. C. Todd et al., "AC-3: Flexible perceptual coding for audio transmission and storage," *96th Audio Engineering Society Convention*, Preprint 3796, Amsterdam, 1994.

55. R. Hopkins, "Choosing an American digital HDTV terrestrial broadcasting system," *Proc. of the IEEE*, Vol. 82, No. 4, pp. 554-563, 1994.

56. "The Grand Alliance," *IEEE Spectrum*, pp. 36-45, April 1995.

57. ETSI, European Telecommunication Standard, Draft prETS 300 401, Jan. 1994

58. Ch. Weck, "The error protection of DAB," Audio Engineering Society-Conference "DAB - The Future of Radio," London, May 1995.

59. R.D. Jurgen, "Broadcasting with Digital Audio," *IEEE Spectrum*, pp. 52-59, March 1996.