# DVB Audio

Leon van de Kerkhof
Philips Consumer Electronics
Email: Leon.vandeKerkhof@ehv.ce.philips.com

## Introduction

The introduction of the Compact Disc, already more than fifteen years ago, has brought high quality digital audio to the consumer. The audio coding used on the Compact Disc is 16-bit linear PCM, which means that for every audio sample 16 bits are used to represent the amplitude of the sample. This method of coding guarantees a very high audio quality, and the Compact Disc is regarded now as the "reference" digital audio system. However, for many transmission channels and storage media, the amount of digital data to be transmitted per second or stored is too high or not affordable. To make better use of the available transmission bandwidth or storage size, MPEG has developed and standardised a more efficient audio coding method.

DVB has adopted the international standards ISO/IEC 11172-3 and ISO/IEC 13818-3, better known as MPEG-1 Audio and MPEG-2 Audio respectively, for transmission of high quality audio accompanying MPEG-2 Video. This paper gives an overview of the MPEG Audio algorithms.

## The Standardisation Process

The International Standardisation Organisation (ISO) has recognised that it would be beneficial for consumers, hardware and software manufacturers and service providers to have a world-wide standard for audio-visual coding. They established, jointly with the International Electrotechnical Committee (IEC), a group of leading international experts in the field of audio and video coding, the so-called Moving Pictures Experts Group (MPEG).

The MPEG committee defines generic standards for coding of moving pictures and associated audio. The first set of international standards produced by this body, usually referred to as MPEG-1, aimed at digital storage media up to about 1.5 Mbit/s. Next to Video and Audio, there are three more parts: Systems, Conformance and Technical Report. The systems part defines a/o how multiple MPEG video and MPEG audio streams can be multiplexed into one stream, and how synchronous playback can be realised. The conformance part defines procedures to test validity of bitstreams and compliance of decoders to the standard. The technical report consists of a software decoder and an example software encoder, that helps interested parties to get quickly accustomed to MPEG Video and Audio. MPEG Audio is thus an integral part of a set of standards, which make it possible for every interested party to quickly design encoders and decoders, and to test conformance of their implementation to the standard, thereby guaranteeing interoperability of different implementations.

At the start of the development of the MPEG-2 standards, companies and institutes world-wide were invited to submit their audio and video coding algorithms to MPEG. For audio coding, in total 14 responses were received. Based on similarities between the algorithms, the

proposals were grouped into four clusters. In all four clusters, the proponents worked together to combine the strong parts of the proposals into one audio coding algorithm.
The four resulting algorithms were then repeatedly tested against each other. On the basis of criteria as sound quality at different bitrates, encoder and decoder complexity, coding delay and error robustness, the MUSICAM algorithm was selected out of four submitted algorithms by the MPEG committee to become the basis for the MPEG Audio standard. During the standardisation process, further improvements were incorporated in the algorithm. This finally lead to the MPEG-1 Audio standard that was accepted by the MPEG committee in 1992 and published in 1993. MPEG-1 Audio addresses high quality coding of monophonic and two-channel stereo signals.

In the mean time, the International Telecommunication Union started to develop a recommendation for the use of audio coding systems in broadcast applications. The process was comparable to the MPEG standardisation process, but now the requirements were solely directed towards use in broadcast environments. A number of different application areas were distinguished, each with their own set of requirements:

- Contribution/Distribution
- Emission
- Commentary links

Codecs were submitted by MPEG, Dolby, Aware, NHK and the Swiss PTT. After several rounds of subjective testing and evaluation of objective parameters as encoder/decoder complexity and delay, MPEG Audio Layer II was selected for the applications Contribution/Distribution and Emission. For the commentary links application MPEG Audio Layer III was selected.
This process finally lead to ITU-R Recommendation BS.1115. This recommendation directly applies to the DVB application.

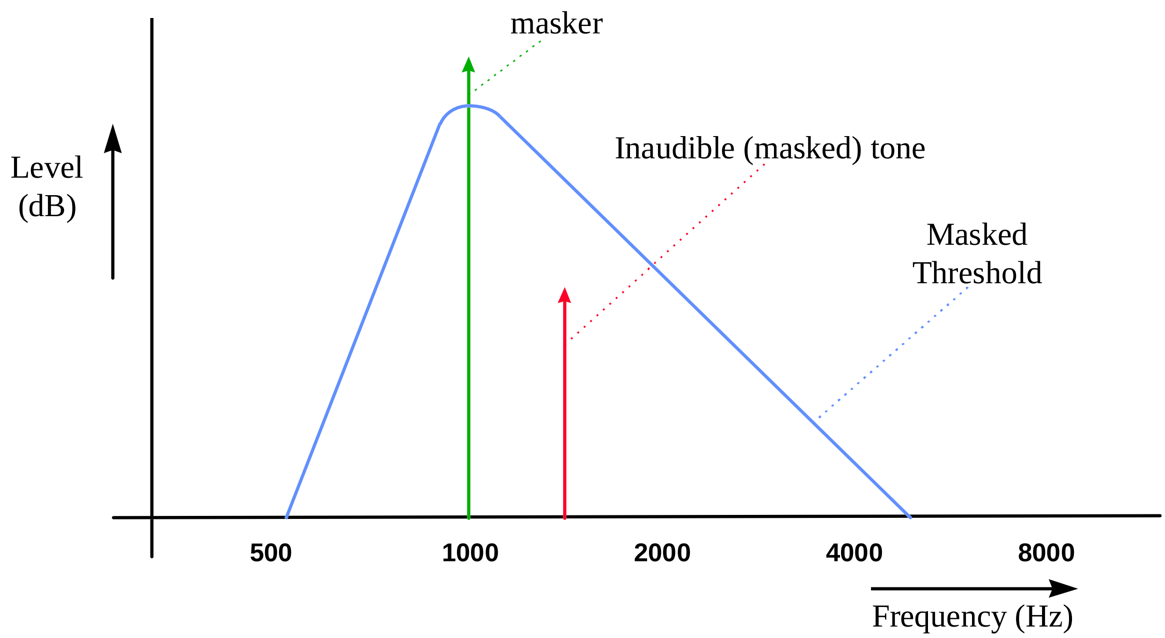
The second set of international standards produced by the MPEG group, known as MPEG-2, aims at broadcast quality video and multichannel audio. Again, MPEG-2 Video and MPEG-2 Audio are integral parts of a set of standards that also specify conformance testing and encoder and decoder software. The resulting standard was published in 1995.

With respect to audio coding, MPEG added 5.1 multichannel capability to the MPEG-1 standard, leading to the MPEG-2 Audio standard which was accepted by the MPEG committee in 1995. This was done in a scaleable way, compatible to MPEG-1 Audio. An improvement to the standard with respect to Pro Logic compatibility lead to a second edition of the MPEG-2 standard, that was accepted in 1997.

Because of its scaleability, the MPEG-2 Audio standard was also selected by the Digital Audio-Visual Council (DAVIC) for “Scaleable Audio”. This was published in release 1.3, 1997.

## Perceptual Coding

The MPEG Audio standard uses a technique called perceptual coding. A perceptual audio coder exploits a psychoacoustic effect known as masking. This is the effect that a tone that is close in frequency to another tone, but lower in level, cannot be heard; the lower level tone is masked by the higher level tone.

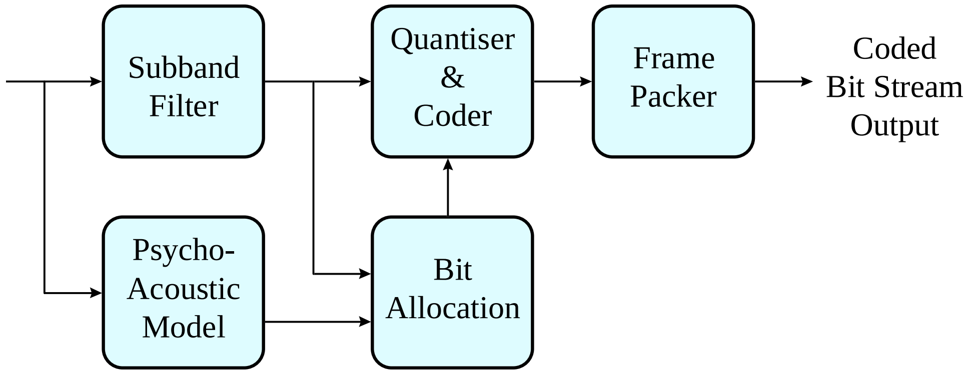



*Masking*

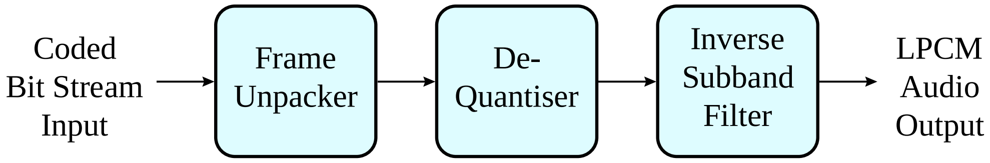
This effect also exists with more complex sounds, such as music and noise. Associated with every sound there exists a masked threshold curve, as a function of frequency, under which another sound cannot be detected by the human auditory system. This is a dynamic process. As the spectrum of the sound changes, this masking curve also changes. All digital audio systems are subject to noise produced in the quantisation process. Perceptual coders operate by shaping the spectrum of the quantisation noise as accurately as possible, so that it is kept under the masked threshold.

**Basic structure of the MPEG Audio subband coder**

A perceptual subband audio encoder constantly analyses the incoming audio signal. A psychoacoustic model dynamically determines the masked threshold, i.e., the threshold under which additional noise will not be audible by the human auditory system. The information from the psychoacoustic model is fed to a bit allocation module, that has the task to distribute the available bits on the channel in an optimal way over the frequency spectrum.
In parallel, the input signal is split into a number of frequency bands, called subbands. Each subband signal is scaled and then quantised in such a way that the quantisation noise introduced by the coding will not exceed the masked threshold for that subband. The quantisation noise spectrum is thus dynamically adapted to the signal spectrum. The "scale factors" and the information on the quantisers used in each subband are transmitted along with the coded subband samples.
The decoder can then decode the bit stream without knowledge of how the encoder determined this information. This keeps the decoder complexity low, and gives great flexibility in both the choice of encoders and the development of decoders. When new results in psychoacoustic research become available, more efficient and higher-performance encoders can be used with complete compatibility with all existing decoders. This flexibility has already paid off: current state-of-the-art encoders outperform the early encoders that were available during the standardisation process.




*MPEG Audio Encoder block diagram*




*MPEG Audio Decoder block diagram*

**Layers**

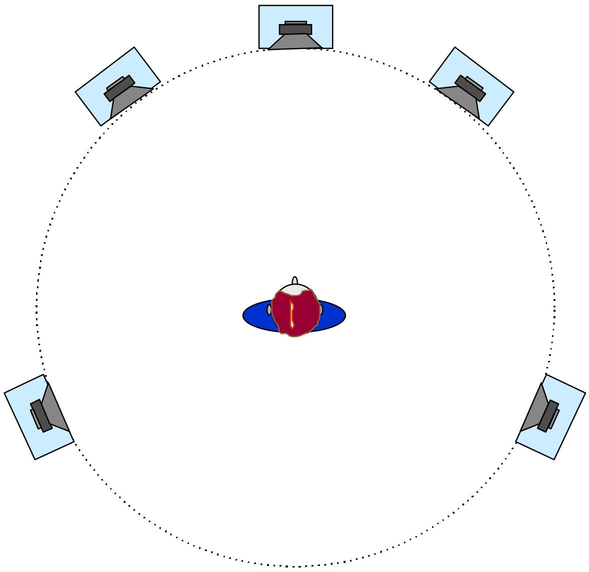
The MPEG Audio specification contains three different algorithms, called layers. The higher the layer, the higher the obtainable compression ratio, but also the higher the complexity, delay and sensitivity to transmission errors. Layer II is the layer that is optimised for broadcast applications. It features subband filtering with 32 equal-width subbands, adaptive bit allocation, and block companding. Bit rates range from 32-192 kb/s for mono, and from 64-384 kb/s for stereo. It has demonstrated excellent performance at 256 kb/s and, when using joint stereo, at 192 kb/s. At 128 kb/s (stereo) the performance is still acceptable for many applications.

Both MPEG Audio Layer I and Layer II are part of the DVB specification. The full range of bit rates for each layer is supported. All three sampling frequencies, 32, 44.1 and 48 kHz, are supported.

**MPEG-2 Extension to multichannel audio**

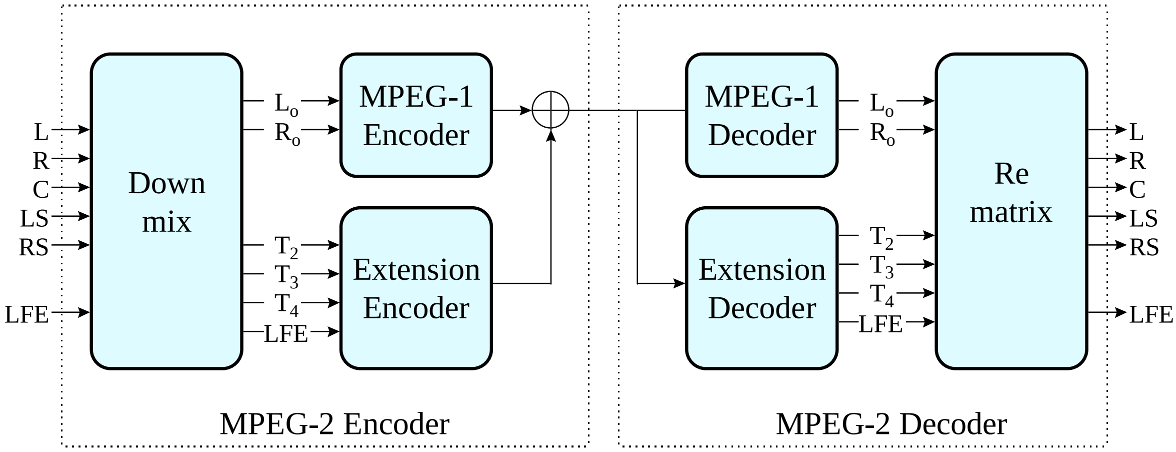
The ITU-R Task Group TG10-1 has worked on a recommendation for multichannel sound systems. The main outcome of this work is Recommendation BS.775, which says that a suitable multichannel sound configuration should contain five channels, where the five channels represent Left, Centre, Right, Left Surround, and Right Surround. If an optional low frequency enhancement (LFE) channel is used, the configuration is known as "5.1". The five-channel configuration is also referred to as `3/2', i.e., three front channels and two surround (rear) channels.

MPEG recognised the necessity to add multichannel capability according to ITU-R Recommendation 775 to the audio standard.



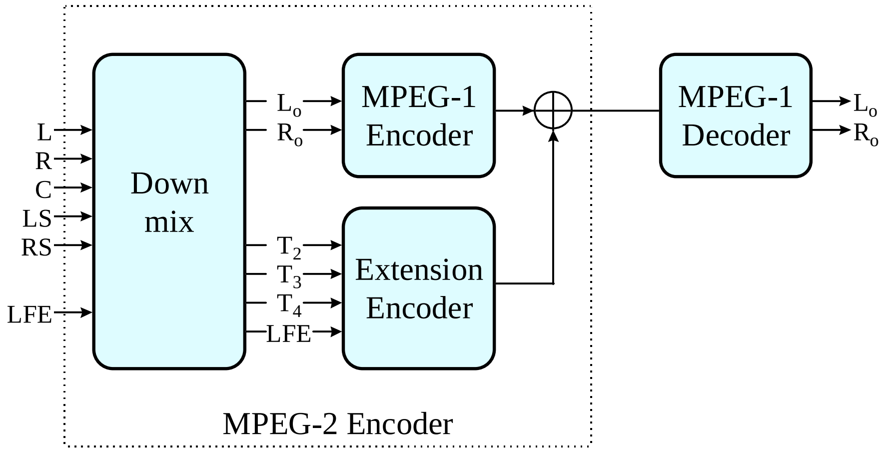
*Loudspeaker set-up according to ITU-R Rec. 775*

This was done in the second phase, leading to the MPEG-2 Audio standard. The extension to multichannel sound supports up to five input channels, the low frequency enhancement channel, and up to seven commentary channels in one bit stream. The extension is both forward and backward compatible with MPEG-1. Forward compatibility means that a multichannel decoder can properly decode a stereo bit stream. Backward compatibility means that a standard stereo decoder will produce a compatible stereo signal when decoding a multichannel bit stream.




*MPEG-2 Encoder/Decoder block diagram*

This is achieved with a truly *scaleable* approach, At the encoder the five input channels are downmixed to a compatible stereo signal. This compatible stereo signal is coded according to the MPEG-1 standard. All the information necessary to retrieve at the decoder end the original five channels is put in the MPEG-1 ancillary data field, which is ignored by MPEG-1 decoders. This additional information is conveyed in information channels $T_2$, $T_3$ and $T_4$, that typically contain the Centre, Left Surround and Right Surround channels, plus an LFE channel. An MPEG-2 multichannel decoder decodes the MPEG-1 part of the bitstream, plus the additional information channels $T_2$, $T_3$, $T_4$ and LFE. From this information, it can retrieve the original 5.1 channel sound.




*MPEG-2 Encoder - MPEG-1 Decoder compatibility*

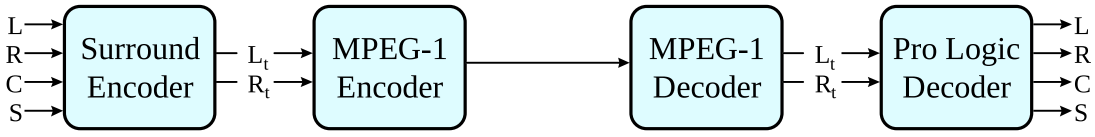
An MPEG-1 decoder, when fed the same bitstream, will decode the MPEG-1 part of the bit stream, and ignore all the additional multichannel information. It will thus reproduce the 2-channel downmix generated in the MPEG-2 encoder. In this way, compatibility with existing 2-channel decoders is obtained. Maybe even more important, this scaleable approach leads to low cost 2-channel decoders, even for multichannel services. With all other coding schemes, the 2-channel decoder for a multichannel service is basically a multichannel decoder that decodes all channels and generates a 2-channel downmix in the decoder.

The standard is very flexible in the sense that it incorporates a variety of techniques that the encoder can use to further improve audio quality.

For multichannel sound, only MPEG-2 Layer II is supported in the DVB specification.

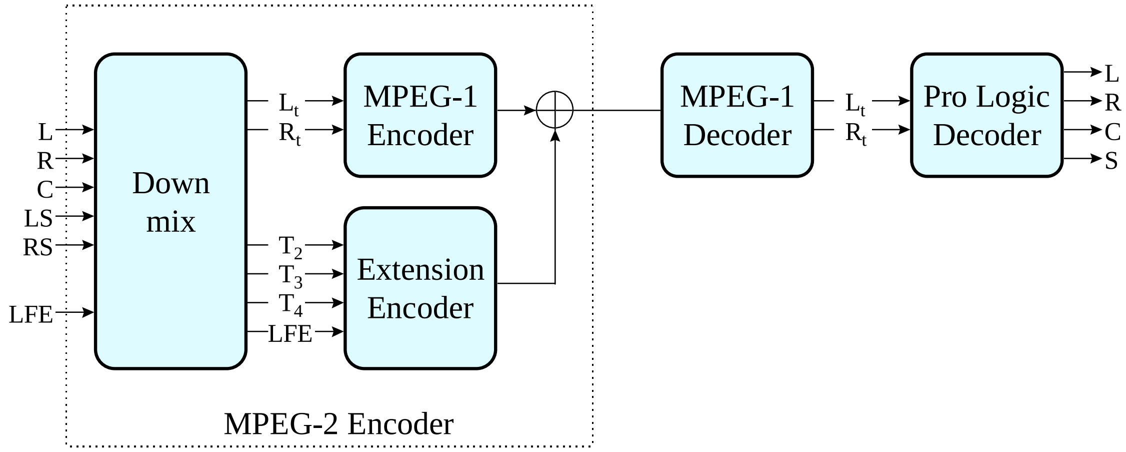
**Pro Logic compatibility**

When the source material is already surround-encoded (e.g., Dolby Surround), the broadcaster may choose to transmit this directly to the audience. One choice would be to transmit this material directly in the 2/0 (stereo only) mode. Surround-encoders operate basically by adding the centre signal in-phase to both the left and right signals, and by adding the surround signal out-of-phase to the left and right signals. In order to properly decode this information, the codec must preserve the amplitude and phase of the left and right signals with respect to each other. This is assured with MPEG coding by limiting any intensity-stereo coding to the frequencies above 8 kHz, since surround-encoding uses surround information only below 7 kHz.




*Transmission of surround material using MPEG-1 Audio*

When transmitting Multichannel source material, compatibility with existing (proprietary) surround decoders is assured by several means. The multichannel encoder can operate using a surround-compatible matrix. This will allow stereo decoders to receive the surround-encoded signal, with optional application to a surround decoder. A full multichannel decoder will rematrix all the signals to obtain the original multichannel presentation. This mode is supported in the MPEG-2 multichannel syntax, and consequently in the DVB specification.




*Matrixed surround compatibility using MPEG-2 Audio*

## IEC 61937

It is to be expected that, certainly in the beginning, many DVB receivers will not have a built-in digital multichannel decoder. Instead, players will have an output interface for the coded MPEG-2 Audio multichannel bitstream, that can be connected to an external decoder. Such an interface is currently being defined as IEC standard IEC 61937. It is based on the IEC 60958 interface. This interface is widely in use for transmission of digital audio in linear PCM format. By replacing the PCM audio samples by 16 bit data words and indicating that these data words are not valid PCM samples suitable for DA conversion, a data channel with a maximum capacity of 1536 kbit/s is obtained (at a sampling frequency of 48 kHz), which is more than sufficient for transmission of coded bit streams. Data is transmitted in bursts. The distance between the start of consequent bursts corresponds to the MPEG Audio frame length of 1152 PCM samples. The length of each burst corresponds to the bit rate, e.g. 24 x 384 bits when the bitrate equals 384 kbit/s at 48 kHz sampling frequency. Each burst is preceded by a preamble, that provides information on a/o the length of the burst.

## MPEG-2 Extension to Lower Sampling Frequencies

Next to an extension to multichannel, MPEG-2 Audio incorporates an extension of MPEG-1 Audio to lower sampling frequencies. This extension is aiming at obtaining an improved frequency resolution in a simple manner. By halving the sampling frequency, the frequency resolution improves by a factor of two, but the time resolution worsens by a factor of two. This results in a better quality for many stationary signals, and a lower quality for a few time-critical signals. The use of the halve sampling frequency is indicated in the bit stream by setting one bit in each frame header, the ID bit, to ‘0’. Further, the list of possible bit rates has been adapted to offer more choice in the lower bit rates, and the table of possible quantisers for each subband has been adapted to the higher frequency resolution.

The use of MPEG-2 Audio at lower sampling frequencies is eminently suited for secondary audio services in DVB.

**Dynamic Range Compression**

Most broadcast audio material is subject to dynamic range compression, either to make the material more suitable for listening conditions with a bad signal-to-noise ratio (e.g., car radio, background music), or to raise the average output level of a program in order to make it more pronounced with respect to competing programs. Because equipment for dynamic range compression is rather expensive, this is usually done on the transmitter side. Most of the digital compression techniques offer the possibility to transmit ancillary data along with the audio signal. This ancillary data could comprise gain factors for a Voltage Controlled Amplifier in the output circuit, which actually performs the dynamic range compression. The gain factors can be determined by the equipment at the broadcasting studio. The users can select, depending on their listening conditions, whether they want to listen to the uncompressed or the compressed signal. The level of compression can be chosen by the listener. MPEG coding syntax allows the implementation of such a system by providing an ancillary data field, within which the compression information may be transmitted.

Alternatively, the ISO/MPEG Layer I and II standard offers an even more attractive solution. By adjusting all scale factors with a single gain factor, the compression can be performed digitally in the decoder, with almost no computational overhead. The fact that the scale factors are valid for 8-ms blocks of data (at a sampling frequency of 48 kHz) is advantageous. The steps in gain that occur at the 8-ms block boundaries are effectively smoothened by the windowing action of the subband filter bank. The gain factors themselves can be derived in the decoder, as well, by using an energy estimate based on the scale factors. This has shown very good results.

**Compatibility with existing domestic hardware**

MPEG Audio is by far the most widespread audio coding algorithm in the world. By now, there are over 100 Million decoders in the market. The main applications are:

- Video CD players
- DVD Video players
- Windows PC’s
- DVB receivers
- DirecTV receivers
- Digital Audio Broadcast (DAB) receivers
- ISDN codecs
- Tapeless studio

Most MPEG Video decoder IC’s do have an on-chip MPEG Audio decoding functions as well. Many A/V receivers now support MPEG-2 multichannel (Sony, Marantz, Meridian, ...). By using a standard IEC 61937 interface for coded bit streams, multichannel decoders can be built in A/V receivers instead of settop boxes.

**Multilingual Support**

MPEG Audio supports two options for multilingual:

1. Separate audio streams for each language. Being an integral part of the MPEG set of standards, there is full support for this option in the MPEG system stream. This option is very flexible, because there are no restrictions on the different audio streams: there may e.g. exist a mix of 2-channel and multichannel streams in one MPEG System stream. The timestamp mechanism defined by MPEG assures accurate synchronisation between the video and any of the audio streams.
2. MPEG Audio also supports the option to embed up to 7 language channels in a multichannel audio stream. Such an audio stream may e.g. contain a multichannel programme at 384 kbit/s, plus a number of voice channels at 64 kbit/s. Each additional language takes 64 kbit/s instead of 384 kbit/s in case a complete multichannel audio stream were broadcast for each language. This is obviously very bit rate efficient.

**Cost of Implementation**

The MPEG Audio algorithm was developed with the broadcast application in mind. The most important characteristic of the broadcast application is that there are many decoders and only a few encoders needed. MPEG Audio has been developed from the basis as a non-symmetric system, in the sense that a high quality encoder is relatively complex, but the utmost attention has been paid to keep the decoder complexity low.

The MPEG-2 Audio algorithm is unique in its scaleable decoder concept. This concept ensures that it is possible to use a low-cost MPEG-1 Audio 2-channel decoder, even if the broadcast is in 5.1 channel sound. It also makes sure that a service can start with 2-channel transmission and 2-channel receivers, and that at any moment in time the transmission and/or the receiver can be upgraded to multichannel. All combinations of transmissions and receivers will remain working, of course only the combination of a multichannel transmission and a multichannel receiver will reproduce multichannel sound. The cost for this compatibility is a marginal increase in required bitrate for a multichannel transmission of 10% on average, about 20% worst case.
This scaleability is obtained by incorporating a 2-channel downmix in the coded bitstream. The advantage is that it is not necessary to either do simulcast transmission, which costs much more in bit rate, or to require all decoders to be basically multichannel, where the stereo decoder performs a downmix to 2-channel after decoding the full multichannel programme. In a market where the vast majority of the decoders will be 2-channel, the corresponding 2.5 times higher complexity - and thus cost - of the 2-channel decoder would be unexplainable to the consumer.

Another factor that makes MPEG Audio inherently cheaper than other audio coding systems is that most MPEG video decoder IC's have a built-in MPEG Audio decoder. This makes the cost of the audio decoder negligible.

**Production advantages**

Cascading (repeated encoding/decoding)
In a test by ITU-R, MPEG Audio Layer II showed the highest quality of all tested codecs in the case of cascading. In this test it was shown that even for the most critical material the MPEG Layer II codec was able to withstand up to 8 times repeated encoding/decoding without significant loss in quality. In fact, MPEG Audio Layer II was the only codec that

fulfilled the quality requirements. As a result, ITU-R Recommendation BS.1115 recommends the use of MPEG Layer II for the applications contribution and distribution.

For professional applications, a high bit rate up to more than 1 Mbit/s for one multichannel programme is supported. This bit rate allows repeated encoding/decoding stages with post-processing between the stages.

Editing
Near-seamless editing is possible in the coded domain with a resolution of 24 ms. Due to the characteristics of the subband filter as applied in MPEG Audio Layer II, a short cross-fade will automatically be generated at the editing points. This avoids clicks in the audio.

Meta Data
Meta data can be attached to the audio stream as 'ancillary data'. This data might comprise information for Dynamic Range Compression (DRC), Level control, etc.

## Conclusion

MPEG Audio Layer II has been designed with the broadcast application in mind.
Its main features are:

- High Audio quality
- Supports a wide range of channel configurations, of which the 2-channel stereo and the discrete 5.1 multichannel configuration are the most important.
- Scaleable decoder concept; low cost 2-channel decoder will reproduce 2-channel downmix from multichannel coded bitstream. This is highly bit rate *and* cost effective.
- Integral part of MPEG standards. Issues like A/V synchronisation, support for multiple audio streams etc. are inherently solved in the MPEG standards; many MPEG Video decoders contain an on-chip MPEG Audio decoder.
- Integral part of DVB specifications; all DVB settop boxes support MPEG Audio decoding.
- Open standard; both encoders and decoders widely available;
- Dynamic Range Control
- Dialogue normalisation
- Fully Pro Logic compatible
- Variable Bit Rate and Constant Bit Rate supported

These features make MPEG Audio the logical choice for DVB.