Genesis of the MP3 Audio Coding Standard

Hans Georg Musmann

Abstract — *In 1988 the International Standardization Organization (ISO) established the Motion Picture Expert Group (MPEG) to develop a digital coding standard for video and audio signals in order to enable interactive video and audio signals on digital storage media. The MPEG Audio Group started with members from 14 research institutions in order to develop a digital audio coding standard guided by a chairman. As a result the MPEG-1, Layer I, Layer II and Layer III coding standards have been developed and proposed for coding of stereo audio signals at 2 x 192 kbit/s, 2 x 128 kbit/s and 2 x 64 kbit/s in 1992. Later, the abbreviation "mp3" or "MP3" was introduced in order to substitute the long name of the successful MPEG-1, Layer III coding standard. This paper describes the development of the MP3 coding standard and its essential components as contributed by the members of the MPEG Audio Group.*

Index Terms — Audio coding, MPEG standards, MP3

I. INTRODUCTION

By the initiative of Hiroshi Yasuda, NTT Japan, and Leonardo Chiariglione, CSELT Italy, the International Standardization Organization (ISO) established a Moving Picture Expert Group (MPEG) in 1988 to develop a digital coding standard for video and audio signals in order to provide interactive video and audio on a Compact Disk (CD).

At that time, the CD was the only low-price digital storage media with large storage capacity and short access time adequate for interactive communications. The problem was the bit rate of 1.5 Mbit/s which required the application of very efficient coding techniques that reduce the bit rate of a stereo audio signal down to 256 kbit/s preserving CD quality and leaving room for an encoded video signal with a bit rate of 1.2 Mbit/s.

Two MPEG Groups were set up for developing these coding techniques and two chairmen were appointed. D. Le Gall, Bell Communications Research became head of the Video Group and H. G. Musmann, University of Hannover, was head of the Audio Group. In addition a so-called Systems Group was established for considering the aspects of application systems. The chairman of this group was A. Simon.

In December 1988 an MPEG meeting was hosted by Thomson's research organization in Germany, Deutsche Thomson-Brandt in Hannover, and the first meeting of the MPEG Audio Group took place at the Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung of the University of Hannover.

A Call for Proposals was issued with the goal of establishing a unique audio coding standard for various applications like

- CD-ROM for audio and video
- DAT Recorder
- Digital Audio Broadcasting.

The audio coding algorithms considered for standardization, had to meet the system requirements of the ISO Proposal Package Description [1] as e.g.

- Input sampling rate: 32, 44.1, 48 kHz
- Input resolution: 16 bit uniform
- Monophonic bit rates: 32, 64, 96, 128, 192 kbit/s
- Stereo or bilingual bit rates: 128, 192, 256, 384 kbit/s
- Length of an addressable unit is less than 1/30 s
- Total encoding and decoding delay less than 80 ms

In response to the Call for Proposals, 14 audio coding algorithms were submitted in June 1989. Because of certain similarities between these coding proposals they were clustered into four development groups. The members of each group collaborated in developing one coding proposal. Table 1 shows the coding concepts and the members of the four groups. The four cluster proposals were submitted in October 1989.

TABLE I
PROPOSERS OF AUDIO CODING ALGORITHMS

Company	Country	Coding Concept
CCETT	F	Subband coding with more
IRT	D	than 8 subbands
Matsushita EIC	J	(MUSICAM)
Philips CE	NL	
AT & T	USA	Transform coding with
France Telecom	F	overlapping blocks
Fraunhofer Ges.	D	(ASPEC)
Deutsche Thomson-Brandt	D	
Fujitsu Limited	J	Transform coding with
JVC	J	non-overlapping blocks
NEC Corp.	J	(ATAC)
Sony Corp.	J	
BTRL	GB	Subband coding with less
NTT	J	than 8 subbands
		(SB / ADPCM)

In July 1990 implementations of the four proposals were tested extensively at the Swedish Broadcasting Corporation, Stockholm in cooperation with the MPEG Audio Group. Due to hardware failure of two coding systems only the MUSICAM and ASPEC coding system could be evaluated completely with respect to the defined subjective and objective tests.

Fig.1 shows a block diagram that is common to both, the ASPEC and MUSICAM coding algorithms. Left and right

H. G. Musmann is with the Department of Electrical Engineering and Computer Science, University of Hannover, Germany.

Contributed Paper
Manuscript received June 29, 2006 0098 3063/06/$20.00 © 2006 IEEE

Authorized licensed use limited to: IPFW. Downloaded on October 4, 2009 at 13:49 from IEEE Xplore. Restrictions apply.

channel of a stereo audio signal are encoded separately. The input signal of a single channel encoder is represented in Pulse Code Modulation (PCM) with 48 kHz sampling frequency and 16 bit per sample. In the first step, the signal is mapped from the time domain into a frequency domain representation either by a transform (ASPEC) or by a subband filterbank (MUSICAM).

The resulting coefficients are normalized in the second step with scalefactors, which have to be transmitted as side information. In the third step, the coefficients are quantized and entropy coded. In order to control the quantization, masking thresholds of the quantization error are calculated based on a psychoacoustic model. From the masking thresholds the bit allocation to the coefficients is derived. The bit allocation information selects one quantizer out of a set of possible quantizers. The bit allocation has to be transmitted as additional side information.

The subjective tests of the stereo sound quality at 2 x 128 kbit/s rate have shown that the sound quality of both coding techniques is very close to that of the PCM studio signal whereby that of ASPEC is slightly superior, especially for lower bit rates. On the other hand MUSICAM showed less decoding delay and less complexity and thus achieved a higher total scoring. Therefore it was agreed at the MPEG Meeting in Santa Clara, CA in September 1990 and at the subsequent ad-hoc meeting at AT&T in Murray Hill, NJ to combine both coding algorithms into one standard algorithm to improve the audio quality and to provide a layer-structure for the combined algorithm in order to support various applications. The chairman of the Audio Group described this situation and the planned procedure for developing the standard in a paper published in December 1990 [2]. Initially a four-layered structure had been planned, but one layer was dropped in December 1990 after an evaluation by Thomson had shown that the expected advantages of the third layer were only marginal, thus leaving only three layers.

An ad-hoc group for preparing the so-called "Committee Draft" of the standard was arranged with four members of the ASPEC group: H. Gerhäuser (FhG), N. Jayant (AT&T), J. Spille (DTB), J. Petit (CNET) and four members of the MUSICAM group: L. v. de Kerkhof (Philips), P. de Wit (Philips), G. Stoll (IRT) and Y. F. Dehery (CCETT). P. Noll (Technical University of Berlin) assisted this ad-hoc group.

In addition, an ad-hoc committee on software verification was formed with the goal to prepare a C-code of the coding algorithm as described in the draft. The chairman of this committee was D. Pan (DEC), the members were B. Aspromonte (Apple), D. Wright (Apple), J. Nelson (Brooktree), J. G. Fritsch (C-Cube) and H. Fuchs (University of Hannover).

The developed ISO/IEC MPEG-1 Audio Coding Standard [3] comprises components that need not necessarily be used, depending on the application and available bit rate. Three layers have been defined:

Layer I contains the basic mapping of the digital audio input into 32 subbands, fixed segmentation to format the data into blocks, a psychoacoustic model to determine the adaptive bit allocation and quantization using block companding and formatting. The recommended bit rate is 2 x 192 kbit/s.

Layer II provides additional coding of bit allocation, scalefactors and samples. Different framing is used. The recommended bit rate is 2 x 128 kbit/s.

Layer III introduces increased frequency resolution based on a hybrid filterbank. It adds a different (nonuniform) quantizer, adaptive segmentation and entropy coding of the quantized values. The recommended bit rate is 2 x 64 kbit/s. Joint stereo coding can be used as an additional feature to any of the layers.

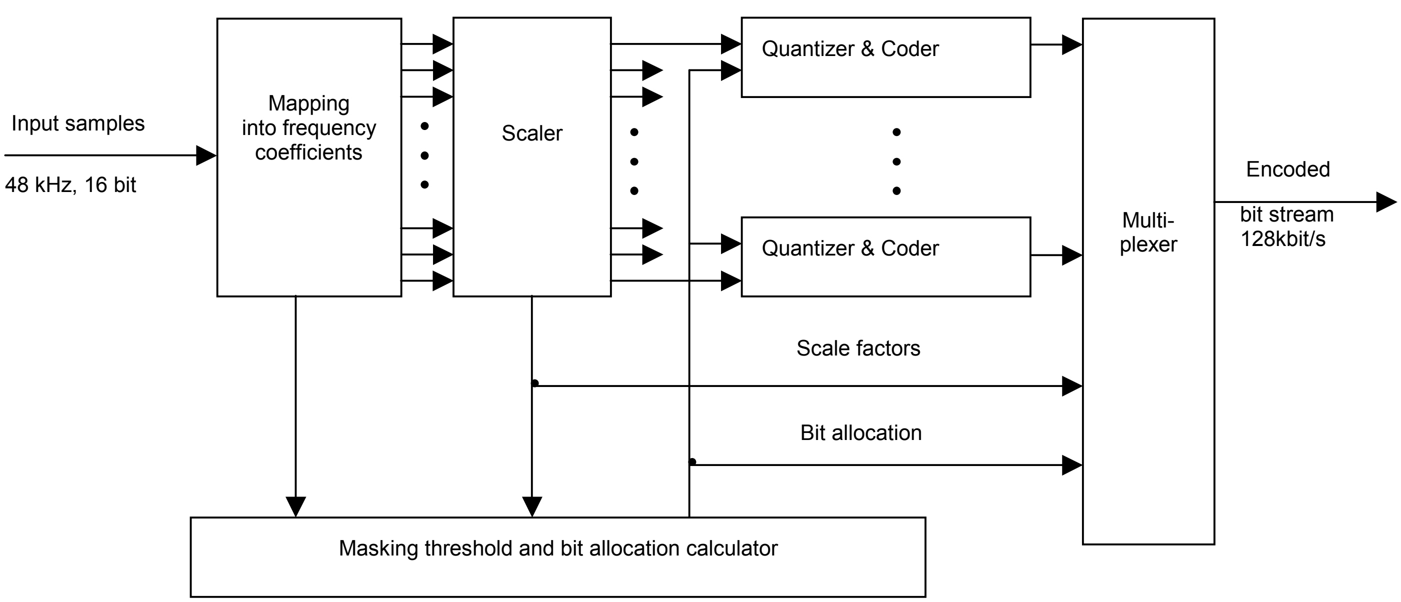


Fig. 1. Block diagram of a single channel encoder

Authorized licensed use limited to: IPFW. Downloaded on October 4, 2009 at 13:49 from IEEE Xplore. Restrictions apply.

The coding complexity increases from Layer I to Layer III, allowing a lower bit rate at the same audio quality. An MPEG-1 Audio Layer N decoder can be designed to decode bitstream data that has been encoded in Layer N and all layers below.

Chapter II describes the origins and developments of the basic coding components from the ASPEC and MUSICAM coding algorithm that merged into the standard algorithm. The basic coding components are the filterbank for subband decomposition and the psychoacoustic model for quantizer bit allocation. During the merging period, additional new ideas came up for improving the coding efficiency. These additional components and the frame structure of the encoded data are outlined. The final chapter III presents an overview of the temporal development of the standard.

II. BASIC CODING COMPONENTS

PSYCHOACOUSTIC MODEL FOR QUANTIZER BIT ALLOCATION

In order to avoid audible coding distortions the quantization error may not exceed the masking threshold of the ear, which depends on the actual characteristics of the masker, i.e. the audio signal itself. If the masking threshold is known, the quantizer step size can be adjusted such that the resulting bit rate is a minimum, but no quantization errors are perceptible.

The masking threshold can be estimated by use of a psychoacoustic model, which is composed of several components that contribute to the masking, see Fig. 2. Here, only the main components shall be explained.

The psychoacoustic model analyzes the audio signal and calculates the masking threshold M(n) as well as the sound pressure level S(n) for each subband n=1,2, ... of the encoder. The difference S(n) - M(n) gives the signal-to-mask ratio SMR(n) in dB which is used to control the quantization of the subband signal n.

In the first step, the frequency spectrum is calculated via a windowed Discrete Fourier Transform (DFT). The resolution of the spectrum required for calculating the masking is finer than that of the subband decomposition. For each DFT coefficient of a subband or group of subbands called scalefactor band, the masking threshold is computed. The minimum masking threshold M(n) within subband n results from an evaluation of three masking effects that have an influence on the global masking.

From E. Zwicker 1961 [4], D. Greenwood 1961 [5] and B. Scharf 1970 [6] it is known that auditory perception works with critical bands. These are overlapping bandpass filters with a bandwidth increasing at high frequencies. For computing the masking threshold the signal energy in a critical band has to be evaluated. The first papers which propose to exploit the masking in critical bands for encoding are those of M. R. Schroeder, B. S. Atal and J. L. Hall 1979 [7] and of M. A. Krasner 1980 [8] in application to speech signals.

J. P. Egan and H. W. Hake 1950 [9] as well as R. P. Hellmann 1981 [10] show that a tonal and a nontonal masker of the same level exhibit different masking. Therefore D. Krahé 1985 [11], [12] distinguishes between tonal and nontonal maskers for estimating the masking threshold in a coding system.

J. D. Johnston 1988 [13] suggests to use the "Spectral Flatness Measure" as defined in N. S. Jayant and P. Noll 1984 [14] in order to determine the tonal and nontonal components of a signal. B. Edler 1988 [15] proposes to use a prediction technique for the same purpose. A detailed description of such a predictive technique can be found in a patent by K. Brandenburg and J. D. Johnston 1989 [16].

After separating tonal and nontonal components, for each masker, a threshold function of simultaneous masking is computed. It describes the masking of neighboring masked frequencies caused by a masker.

Threshold functions of simultaneous masking have been known for many years, see J. P. Egan and H. W. Hake 1950 [9]. An early paper that makes use of simultaneous masking for speech coding is from J. Blauert and P. Tritthart 1975 [17]. The first papers that propose to exploit simultaneous masking for bit rate reduction in an audio coding system were published by D. Krahé 1985 [18], E. F. Schröder and W. Voessing 1986 [19], K. H. Brandenburg 1987 [20], J. D. Johnston 1988 [13], G. Theile 1984 [21], [22], G. Stoll 1988 [23] and R. N. J. Veldhuis, M. Breeuwer and R. G. van der Waal 1989 [24].

The global masking threshold of a frequency results from the sum of the masking thresholds originating from the tonal and nontonal maskers and the absolute threshold in quiet. A reference for the absolute threshold in quiet is E. Zwicker [25]. The frequency with minimum masking threshold within a subband determines the global masking threshold M(n) and the signal-to-mask ratio SMR(n) required for adjusting the quantizer step size.

In addition to the psychoacoustic model, presented in Fig. 2. the temporal masking is used to control the frequency resolution of the subband decomposition. From measurements of C. E. Robinson and I. Pollack, 1973 [26] as well as H. Fastl 1977 [27], it is known that an acoustic attack generates temporal masking before and after the attack. Again, D. Krahé 1986 [28] is one of the first authors who suggests considering temporal masking in an audio coding system.

The psychoacoustic model is not specified in the mandatory part of the standard although it has a strong impact on the sound quality of the encoded signal. Therefore, two psychoacoustic models are described in the informative part of the standard. Modifications of these models may also be used.

SUBBAND DECOMPOSITION

Layer I contains the basic mapping of the digital audio signal into 32 subband samples, segmentation of the sequence of subband samples into blocks and block companding by scalefactors. Layer II provides additional coding of bit.allocation, scalefactors and subband samples. The structure

Authorized licensed use limited to: IPFW. Downloaded on October 4, 2009 at 13:49 from IEEE Xplore. Restrictions apply.

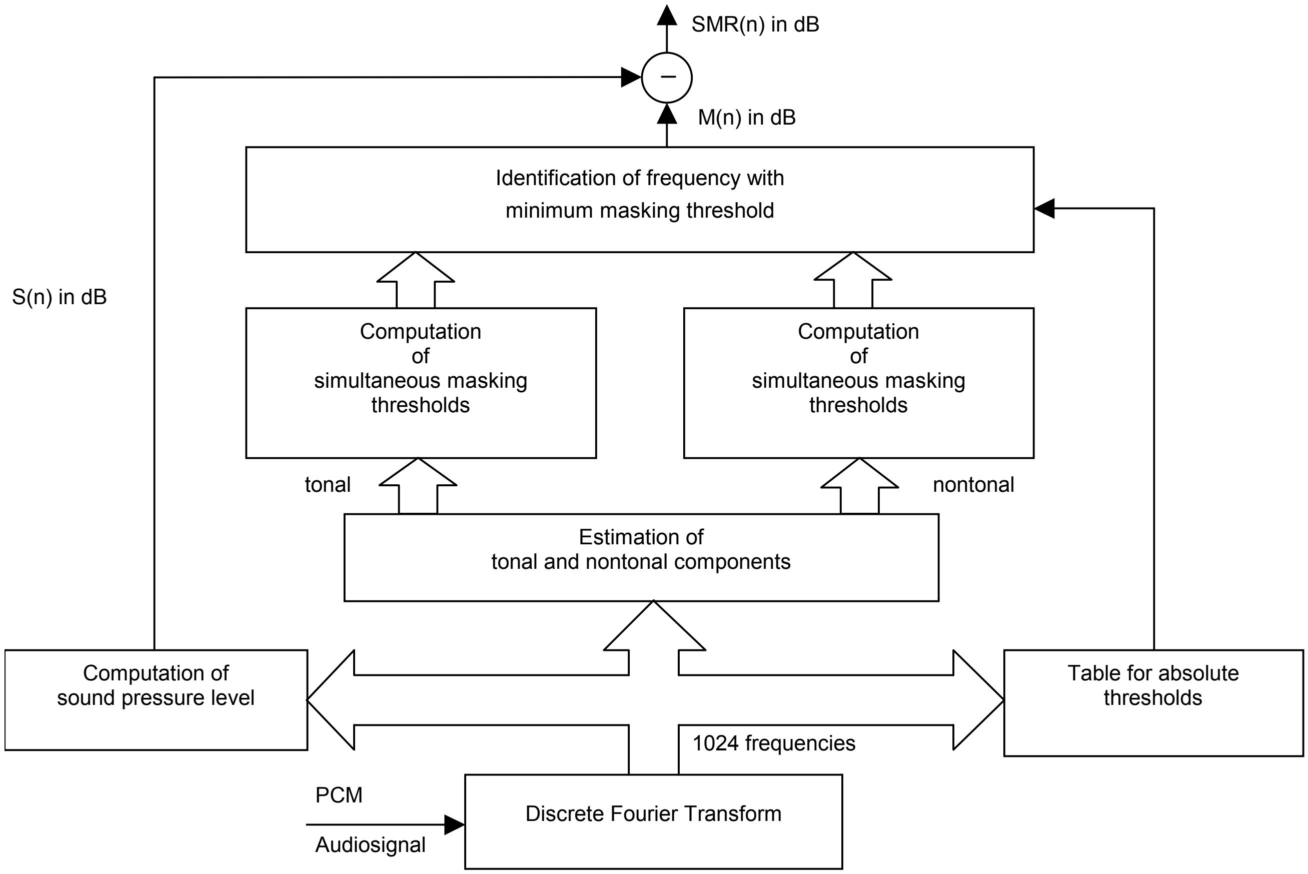


Fig. 2. Psychoacoustic model for calculating the masking threshold M(n) and the signal-to-mask ratio SMR(n) of subband n

and description of these two layers has mainly been contributed by the MUSICAM group [29], [30].

For subband decomposition, a polyphase filterbank with 32 bands of uniform bandwidth is applied. Since the frequency resolution of this decomposition does not correspond to that of critical bands, the psychoacoustic model uses a 1024 point Fast Fourier Transform (FFT) in order to provide a higher frequency resolution adequate for calculating the masking thresholds.

The structure of the polyphase filterbank was invented by J. H. Rothweiler in 1983 [31]. Special embodiments for such a polyphase filterbank are described by J. R. L. Masson and Z. Picel 1986 [32] as well as by J. B. Rault, Y. F. Dehery, J. Y. Roudaut, A.A.M.L. Bruekers and R. N. J. Veldhuis 1990 [33].

The use of 32 bands is a compromise between complexity and bit rate considered for Layer I and Layer II coding. This 32-band polyphase filterbank is extended by a Modified Discrete Cosine Transform (MDCT) to a so-called hybrid filterbank in Layer III coding in order to achieve a lower bit rate by higher frequency resolution.

Fig. 3 shows a block diagram of the ISO/MPEG Layer III encoder. Early papers on transform coding for speech and audio signals were published by R. Zelinski and P. Noll 1977 [34] and K. H. Brandenburg and H. Schramm 1983 [36]. Whereas in [34] a DCT with adaptive bit allocation is proposed, [36] describes an implementation of such a coder. A. Sugiyama, F. Hazu, M. Iwadare and T. Nishitani 1990 [35] investigated a transform with adaptive block size in order to cope with instationary signal statistics. The structure of the hybrid filterbank has mainly been developed by members of the ASPEC group [37], [39] with contributions by B. Edler, University of Hannover [38], [40].

The hybrid filterbank provides an increased frequency resolution by splitting each of the 32 bands into 18 frequency lines by use of a Modified Discrete Cosine Transform (MDCT) with overlapping blocks and a window length of 36. In the case of a sound attack the quantization can generate audible pre-echoes which are not masked by temporal masking of the ear. In order to avoid such audible pre-echoes adaptive windowing was proposed by B. Edler 1989 [38], which allows switching between an 18-point MDCT and a 6-point MDCT with a reduced window length of 12 preserving exact reconstruction. The window switching decision (WSD) is derived from the temporal masking of the pre-echoes. Several methods were proposed for estimating the temporal masking and calculating the WSD [16], [38], [39]. The 32 subband signals of the polyphase analysis filterbank contain aliasing components which are compensated in Layer I and Layer II coding by the polyphase synthesis filterbank in the decoder. In Layer III coding these aliasing components impair the coding efficiency of the subsequent MDCT. Therefore butterflies for aliasing reduction are introduced in the 32 subband signals according to a proposal of B. Edler 1992 [40].

Authorized licensed use limited to: IPFW. Downloaded on October 4, 2009 at 13:49 from IEEE Xplore. Restrictions apply.

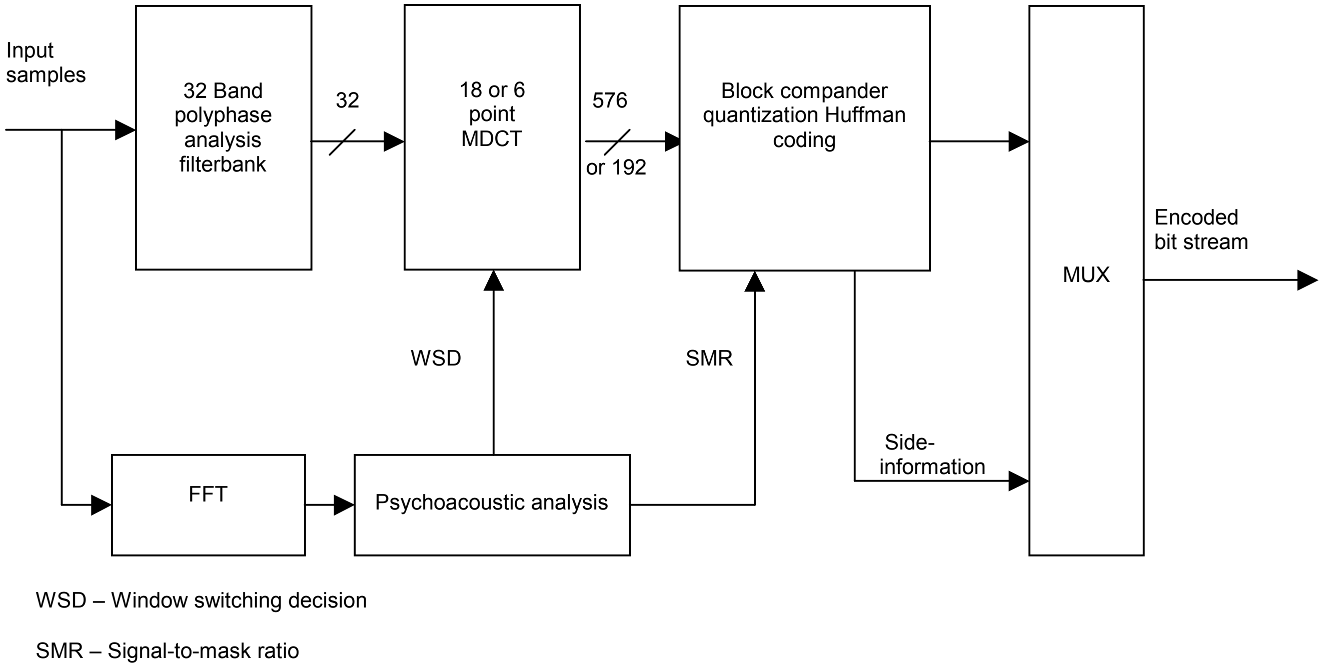


Fig. 3. Block diagram of the Layer III encoder

CODING OF AUDIO CHANNELS

The standard allows the coding of either single-channel (mono) or dual-channel (stereo) audio signals. Initially stereo signals were coded as two separate channels, both adaptively sharing the available data rate. In order to improve the coding efficiency two modes of joint stereo coding have been added.

Intensity stereo mode is limited to the upper frequency range above about 4 kHz and relies on the predominantly intensity-based directional hearing in that frequency range. Left and right channel signals are combined and reconstructed from the one composite signal and scalefactor information [41] in Layer I and II, while in Layer III directional information is explicitly encoded into the data stream.

The middle/side (MS) stereo mode as first described by Johnston in 1989 [42] is only available in Layer III. Here the two audio signals are coded and transmitted as the sum and difference signal of the left and right channel. Both MS stereo mode and intensity stereo mode can be used simultaneously in one frame separately for low and high frequency components.

FRAME STRUCTURE OF ENCODED DATA

Whereas in Layer I and Layer II, consecutive samples of a subband are grouped into blocks, in Layer III neighboring frequency lines of the MDCT are grouped into so-called scalefactor bands for scaling. Bit allocation to the quantizers is controlled by the signal-to-mask ratio (SMR) provided by the psychoacoustic model. Layer III adds Huffman coding and does the optimization in two iteration loops for each frame.

The data to be encoded and transmitted consists of the data for the frequency components and for side-information like scalefactors, sampling frequency or bit rate. The encoded data is formatted in frames. One frame contains the information for 384 PCM audio samples in Layer I or for 1152 PCM audio samples in Layer II or III. The frame structure and the semantics were developed and described for Layer I and Layer II coding by the MUSICAM group [41], [43] and were extended later in collaboration with the ASPEC group for Layer III coding.

Since Huffman coding results in a variable bit rate, which may exceed the available bit rate, a second iteration loop is required in order to reduce the bit rate in such a situation by increasing the quantizer step sizes. In the case that the available bit rate is not completely used for coding a frame, the remaining bits are saved in a bit reservoir for the next frame. The concept of this procedure goes back on K. H. Brandenburg 1987 [20].

III. TEMPORAL EVOLUTION OF THE COMMITTEE DRAFT AND STANDARD

The first important task of the MPEG Audio Group was the definition of the requirements arising from various prospective applications and the definition of the testing conditions for candidates of audio coding algorithms. Some of these aspects are described by G. Stoll and Y. F. Dehery 1990 [44].

After testing the four cluster proposals and evaluating the subjective and objective tests the ad-hoc group for drafting the standard was arranged in August 1990 at the Stockholm Meeting, see Fig.4. In December 1990 the structure of the so-called Committee Draft was available. A second subjective test to verify the newly defined Layer I, II and III took place at Swedish Broadcasting in Stockholm in May 1991. One month later at the Paris Meeting in June 1991, the description of

Authorized licensed use limited to: IPFW. Downloaded on October 4, 2009 at 13:49 from IEEE Xplore. Restrictions apply.

Layer I and Layer II was almost finalized. At that time the quality that was expected.

In order to prepare a revised version of the Committee Draft CD 11172, the following responsible editors were identified:

Layer I: L. van de Kerkhof
Layer II: G. Stoll
Layer III: K. H. Brandenburg with assistance of J. D. Johnston, G. Stoll and Y. F. Dehery
Psychoacoustic Model I : G. Stoll
Psychoacoustic Model II : J. D. Johnston

The next meeting was held in Santa Clara in August 1991. The main discussion concentrated on Layer III coding and the need for a hybrid filterbank for improving the sound quality.

In order to answer this question, a third subjective test was organized at Hannover in October 1991 by H. Fuchs [45] from the University of Hannover with support from the Swedish Broadcasting Corporation, from Deutsche Thomson-Brandt and Norddeutscher Rundfunk. As a result of this test it was agreed at the Kurihama Meeting in November 1991 to include the Layer III algorithm with the hybrid filterbank in the standard.

Hannover, December 1988
- *First meeting of MPEG Audio Group*
- *Call for proposals*

Rennes, May 1989
- *Definition of requirements and test conditions for the proposals*

Yokosuka, October 1989
- *Clustering of 14 proposals into 4 groups*

Stockholm, August 1990
- *First assessment of proposals by listening test and evaluation of objective criteria*
- *Concentration on MUSICAM and ASPEC algorithms*
- *Start of drafting the standard with a layer- concept*

Berlin, December 1990
- *Structure of the Committee Draft available*

Stockholm, May 1991
- *Second subjective test for verification*

Paris, June 1991
- *Description of Layer I and Layer II almost finalized*
- *Second subjective test reveals problems with Layer III*

Santa Clara, August 1991
- *Layer I and Layer II approved*

Hannover, October 1991
- *Third subjective test for verification of Layer III*

Kurihama, November 1991
- *Layer III approved*
- *Committee Draft available ISO / IEC 1992*
- *International Standard IS 11172-3 adopted*

Fig. 4. Milestones of the Committee Draft development

Thus the Committee Draft of the MPEG Audio Standard could be finalized by the end of 1991 as planned. At the same time Professor Musmann passed the chair of the MPEG Audio Layer III coding algorithm did not yet provide the sound Group to Professor Noll, who guided the final steps of the formal standardization by ISO and started the work on MPEG - 2 Advanced Audio Coding (AAC).

Tutorials on the described audio coding standard presenting more technical detail were published e.g. by K. H. Brandenburg and G. Stoll 1992 [46], D. Pan 1995 [47] and P. Noll 1995 [48]. The importance of this audio coding standard for digital broadcasting is discussed in a book, edited by R. De Gaudenzi and M. Luize, 1994 [49] with contributions by e. g. G. Stoll (IRT), Y. F. Dehery (CCETT), B. Grill (FhG) and L. Van de Kerkhof (Philips).

The Fraunhofer-Institute FhG-IIS in Erlangen, Germany immediately promoted the commercial use of the MPEG-1 contributions by Layer III Audio Coding Standard by developing software and hardware modules. This gave rise to the first applications within two years. Also, the Fraunhofer-Gesellschaft Erlangen introduced the abbreviation MP3 instead of MPEG-1 Layer III for marketing purposes. The great breakthrough came with its application to the Internet.

As it is typical for an ISO/IEC standard, the use of the standardized methods and technologies may not be free but subject to reasonable and nondiscriminatory licensing by the holders of respective intellectual property rights. Today two organizations, Thomson of France and Sisvel of Italy are handling the licensing of the ISO MPEG AUDIO CODING STANDARD.

ACKNOWLEDGMENT

The Audio Group needed three years from its first meeting to finalize the Committee Draft. In 1990 and 1991 the members of the Audio Group had to work very hard in order to meet the planned schedule. In addition, the MUSICAM and ASPEC group worked in strong but coordinated competition. In the end, we learned that due to the competition and merging of the results we had developed the first digital International Audio Coding Standard with remarkable data compression through cooperation.

It was a pleasure for the author to chair this Audio Group of the highest qualified audio coding experts in the world being competitors at the beginning and friends at the end.

Finally the author wants to express his thanks to Professor K. H. Brandenburg, Y. F. Dehery, Professor P. Noll, Dr. B. Edler, Dr. H. Fuchs, and Dr. E. F. Schröder for their careful review of this paper and valuable comments.

REFERENCES

[1] ISO / IEC ITC1 / SC2 / WG8 MPEG Document 89 / 128 “Proposal Package Description” [1] ISO / IEC ITC1 / SC2 / WG8 MPEG Document 89 / 128 “Proposal Package Description” 1989

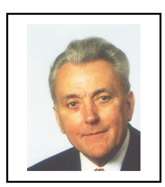
[2] H. G. Musmann, “The ISO audio coding standard”, Proc. IEEE Globecom, December 1990, vol. 1 pp. 405.1.1-405.1.

[3] ISO / IEC International Standard IS 11172-3 “Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1,5 Mbit/s - Part3: Audio”, 1992

[4] E. Zwicker, “Subdivision of the audible frequency range into critical bands”, J. Acoust. Soc. Am. 33, 1961, pp. 248

Authorized licensed use limited to: IPFW. Downloaded on October 4, 2009 at 13:49 from IEEE Xplore. Restrictions apply.

[5] D. D. Greenwood, “Critical bandwidth and frequency coordinates of the basilar membrane”, J. Acoust. Soc. Am. 33, 1961, pp. 1344-1356
[6] B. Scharf, “Critical bands in Foundations of Modern Auditory Theory”, ed. by J. V. Tobias, Acedemic Press, 1970, pp. 159-202
[7] M. R. Schroeder, B. S. Atal and J. L. Hall, “Optimizing digital speech coders by exploiting masking properties of the human ear”, J. Acoust. Soc. Am. 66, No. 6, 1979, pp. 1647-1652
[8] M. A. Krasner, “The critical band coder - Digital encoding of speech signals based on the perceptual requirements of the auditory system”, ICASSP 80, vol.5, April 1980, pp. 327-331
[9] J. P. Egan and H. W. Hake, "On the masking pattern of simple auditory stimulus", J. Acoust. Soc. Am. 22, No. 5, Sept. 1950, pp. 622-630
[10] R. P. Hellmann, “Asymmetry of masking between tone and noise”, Perception Psychophysics, vol. 11, No. 3, 1981, pp. 241-246
[11] D. Krahé "Ein Verfahren zur Datenreduktion bei digitalen Audiosignalen unter Ausnutzung psychoakustischer Phänomene", Rundfunktechnische Mitteilungen, Jahrg. 30, 1986, H. 3, pp.117-123
[12] D. Krahé "Verfahren zur Übertragung eines digitalen Audiosignals", German Patent No. DE 3506912, Feb. 1985
[13] J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria", IEEE Journal on Selected Areas in Communications, vol. 6, No. 2, February 1988, pp. 314-323
[14] N. S. Jayant and P. Noll, "Digital coding of waveforms", Prentice Hall, Englewood Cliffs, New Jersey, 1984
[15] B. Edler, "Prädiktive Teilbandcodierung mit Vektorquantisierung für Audiosignale hoher Tonqualität", Proc. ITG-Fachtagung "Hörrundfunk" 1988, pp. 223-228
[16] K. H. Brandenburg, J. D. Johnston, "Perceptual coding of audio signals" US Patent No. 5040217, filed on 18 Oct. 1989 C. J. Kaufman, Rocky Mountain Research Laboratories, Boulder, CO, personal communication, 1992.
[17] J. Blauert and P. Tritthart, "Ausnutzung von Verdeckungseffekten bei der Sprachcodierung", Proc. DAGA 1975, pp. 377-380
[18] D. Krahé "Neues Quellencodierungsverfahren für qualitativ hochwertige digitale Audiosignale", Proc. NTG-Fachtagung "Hörrundfunk" November 1985, pp. 371-381
[19] E. F. Schröder and W. Voessing, "High quality digital audio encoding with 3.0 bit / sample using adaptive transform coding", Proc. 80. AES-Convention 1986, Preprint 2321
[20] K. H. Brandenburg, "OCF-A new coding algorithm for high quality sound signals", Proc, ICASSP 1987, pp. 141-144
[21] G. Theile, "Verfahren zum digitalen Übertragen eines Rundfunkprogrammsignals", German Patent No. DE 3440613, November 1984
[22] G. Theile, M. Link and G. Stoll "Low bit rate coding of high quality audio signals", Proc. 82. AES-Convention, March 1987, Preprint 2432
[23] G. Stoll, M. Link and G. Theile "Masking-pattern adapted subband coding: Use of the dynamic bit-rate margin", Proc. 84. AES-Convention, March 1988, Preprint 2585
[24] R.N.J. Veldhuis, M. Breeuwer and R.G. van der Waal, "Subband coding of digital audio signals without loss of quality", Proc. ICASSP 1989, pp. 2009-2012
[25] E. Zwicker, "Psychoakustik", Springer Verlag Berlin, Heidelberg, New York, 1982
[26] C. E. Robinson and I. Pollack, "Interaction between forward and backward masking: A measure of the integrating period of the auditory system", J. Acoust. Soc. Am. 53, 1973, pp. 1313-1316
[27] H. Fastl, "Temporal masking effects: II. Critical band noise masker", Acoustica, vol. 36, 1977, pp. 317-330
[28] D. Krahé "Ein Verfahren zur Datenreduktion bei qualitativ hochwertigen digitalen Audiosignalen", Proc. 14. Tonmeistertagung 1986, pp. 434-445
[29] Y. F. Dehery, M. Lever and P. Urcun, "A MUSICAM source codec for digital audio broadcasting and storage", Proc. ICASSP Toronto-1991, pp. 3605-3608
[30] D. Wiese and G. Stoll, "Bitrate reduction of high quality audio signals by modelling the ear's masking thresholds", Proc. 89. AES-Convention, Sept. 1990, Reprint 2970
[31] J. H. Rothweiler, "System for digital multiband filtering", US Patent No. 4691 292, April 1983
[32] J.L.R. Masson and Z. Picel, "Signal analyzing and synthesizing filter bank system", US Patent No. 4799-179, Jan. 1986
[33] J.B. Rault, Y. F. Dehery, J. Y. Roudaut, A.A.M.L. Bruekers and R.N.J. Veldhuis, "Digital transmission system using subband coding of a digital signal", European Patent No. EP 0400 755 B1, May 1990
[34] R. Zelinski and P. Noll, "Adaptive transform coding of speech signals", IEEE Trans. Acoust. Speech and Signal Process., vol. ASSP-25, 1977, pp. 299-309
[35] A. Sugiyama, F. Hazu, M. Iwadare and T. Nishitani, "Adaptive transform coding with an adaptive blocksize (ATC-ABS)", Proc. ICASSP, May 1990, pp. 1093-1096
[36] K. H. Brandenburg and H. Schramm, " A 16 bit adaptive transform coder for real time processing of sound signals", Proc. EUSIPCO 1983, pp. 359-362
[37] K. H. Brandenburg J. Herre, J. D. Johnston, Y. Mahieux, E. F. Schröder, "ASPEC: Adaptive Spectral Perceptual Entropy Coding of high quality music signals", Proc. 90. AES-Convention, Febr. 1991, Preprint 3011
[38] B. Edler, "Codierung von Audiosignalen mit überlappender Transformation und adaptiven Fensterfunktionen", Frequenz, vol. 43, pp. 252-256, 1989
[39] K. Brandenburg and J. D. Johnston, "Second generation perceptual audio coding: The hybrid coder", Proc. 88. AES-Convention, March 1990, Preprint 2937
[40] B. Edler, "Aliasing reduction in subbands of cascaded filterbanks with decimation", Electronic Letters, vol. 28, No. 12, pp. 1104-1106, June 1992
[41] G. C. P. Lokhoff, Y. F. Dehery, G. Theile and G. Stoll, "Intensity stereo encoding and decoding in a transmission system", European Patent No. EP 0599 824 B1, May 1990
[42] J. D. Johnston, "Perceptual transform coding of wideband stereo signals", Proc. ICASSP 1989, pp. 1993-1996
[43] G. C. P. Lokhoff, "Decoder for decoding an encoded digital signal and a receiver comprising the decoder", European Patent No. EP 0660 540 B1, May 1990
[44] G. Stoll and Y. F. Dehery, "High quality audio bit-rate reduction system family for different applications", Proc. ICC, April 1990, pp.322.3.1 - 322.3.5
[45] H. Fuchs, "Report on the MPEG/Audio subjective listening tests in Hannover", ISO/IEC JTC1/SC29/WG11 doc. MPEG 91/331, November 1991
[46] K. Brandenburg and G. Stoll, "The ISO/MPEG Audio Codec: A generic standard for coding of high quality digital audio", Proc. 92. AES-Convention, March 1992, Preprint 3336
[47] D. Pan, "A tutorial on MPEG / audio compression, IEEE Multimedia, vol. 2, No. 2, 1995, pp. 60-74
[48] P. Noll, "Digital audio coding for visual communications", Proc. of the IEEE, vol. 83, No.6, June 1995, pp. 925-943
[49] R. De Gaudenzi and M. Luize, "Audio and video digital radio broadcasting systems and techniques", Elsevier, 1994M. Young, *The Technical Writer's Handbook.* Mill Valley, CA: University Science, 1989.



Hans Georg Musmann received the Dipl.-Ing degree in Electrical Engineering in 1962, and the Dr.-Ing. degree in 1966, both from the Technische Universität Braunschweig, Germany. From 1973 to 2003 he was Full Professor and Head of the Institut für Theoretische Nachrichtentechnik und Informationsverarbeitung at the Universität of Hannover, Germany. His main research field is video and audio coding for future communication systems. He was guest professor at the University of Hangzhou in China in 1981 and at the University of Pretoria in South Africa in 1984. H. G. Musmann is a member of the Akademie der Wissenschaften zu Göttingen, of the Academia Europaea Cambridge, of the Academia Scientiarum et Artium Europea, Wien and of the Braunschweigische Wissenschaftliche Gesellschaft. From 1988 to 1992 he chaired the ISO MPEG/Audio group which developed the ISO Audio Coding Standard MP3.

Authorized licensed use limited to: IPFW. Downloaded on October 4, 2009 at 13:49 from IEEE Xplore. Restrictions apply.