MPEG-4 ALS—International Standard for Lossless Audio Coding

Takehiro Moriya†, Noboru Harada, Yutaka Kamamoto, and Hiroshi Sekigawa

Abstract

This article explains the technologies and applications of lossless audio coding. NTT started research on lossless coding of audio signals and proposed the initiation of international standardization in 2002, aiming at high-quality services suitable for broadband networks. As a result of cooperative work with other organizations, the specifications of this technology were officially incorporated in the ISO/IEC MPEG standard published in March 2006 (ISO: International Standards Organization, IEC: International Electrotechnical Commission, MPEG: moving picture experts group).

1. Background

In parallel with the evolution of broadband networks and digital audio equipment, information rates for delivery and storage keep growing rapidly in response to demands for high-quality audio signals (high sampling rates, high word resolution, and multiple channels). NTT Communication Science Laboratories recognized the importance of lossless compression of audio signals and the standardization of its technology, considering interoperability, long-term maintenance, and clear status of intellectual property rights. The Laboratories took the initiative in promoting this technology and advancing it to a standard in MPEG (moving picture experts group), which is a working group within ISO/IEC (International Standards Organization, International Electrotechnical Commission).

NTT initiated discussion on the need for and requirements for such a standard and prepared the technical call for technologies. Through the normal standardization process, several improvements and integration efforts were applied to the initial reference model. The partners for this standardization included the Technical University of Berlin (Germany), RealNetworks Corp. (USA), and I2R (Singapore).

As a result of the final ballot in December 2005, the specifications of the lossless coding were officially established as “14496-4 3rd edition amendment 2 (ALS: audio lossless coding)” and published in March 2006 [1]-[5]. Note that, in parallel, MPEG-4 also established MPEG-4 SLS (scalable lossless coding), which provides the special functionality of lossy-to-lossless compression and bit-rate scalable compression. In this article, we focus on the more general compression tool: MPEG-4 ALS.

As shown in **Fig. 1**, MPEG audio standards have made significant contributions to communication systems including broadcasting, mobile services, and Internet services. Most of the audio coding standards, such as MP3 and AAC or the technology for the MiniDisc system[*1], are based on perceptual coding with a high compression ratio in exchange for minor waveform distortion at the decoder. These encoders carefully control the quantization distortion by utilizing the characteristics of the human ear. The waveform is different from the original, although it is perceptually very close to it.

Unlike perceptual coding, lossless coding ensures perfect reconstruction of the waveform without a sin-

† NTT Communication Science Laboratories
Atsugi-shi, 243-0198 Japan
E-mail: moriya@idea.brl.ntt.co.jp

*1 MiniDisc: System introduced by Sony Corporation as a digital replacement for conventional analog compact audio cassettes.

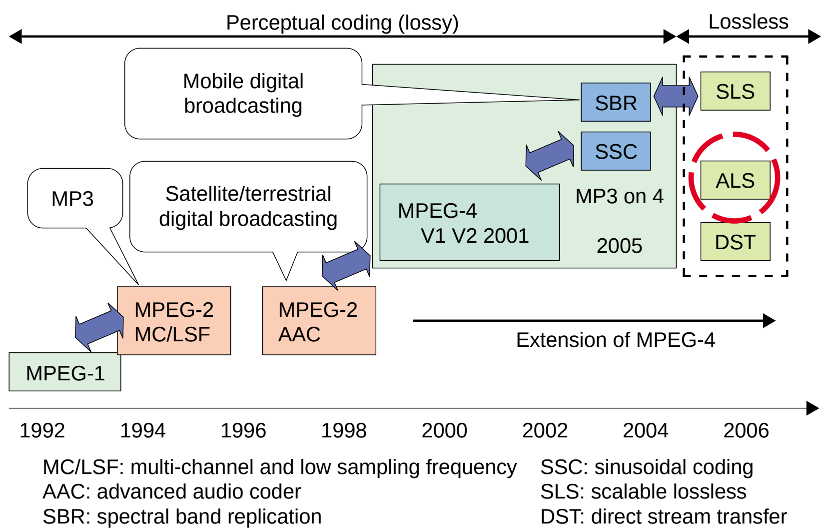


Fig. 1. History of MPEG audio.

Table 1. Comparison of perceptual coding and lossless coding.

	Perceptual coding	Lossless coding
Compression ratio	5–10%	15–70%
Difference from original	Perceptually no difference	Strictly no difference
Applications	Broadcasting, portable players	Storage, editing, other signals besides audio
Methods, standards	MPEG (MP3, AAC), MiniDisc	Some free software, SACD, DVD-audio

SACD: super audio compact disc

gle bit of difference. This property is very important for applications such as waveform editing and the archiving of high-quality audio signals. The cost of perfect reconstruction is a limited compression ratio and hence larger compressed file sizes, which range from 15 to 70% of the original file size depending on the statistical properties of the original waveform. These comparisons are summarized in **Table 1**.

2. Technical description

2.1 Fundamental principles

This technology is based on time-domain linear prediction. The fundamental processing of encoding and decoding is shown in **Fig. 2**. NTT Laboratories is one of the pioneer contributors to the development of linear prediction technology, which has been widely used as an essential tool in speech coding systems such as cellular phones and IP (Internet protocol) phones. Linear prediction analysis estimates prediction parameters that minimize the errors caused by prediction from a certain number of past samples. Integer values of the prediction error signal are transmitted to the decoder. The decoder can reconstruct the original waveform without a single bit error from the prediction parameters and the prediction error signal. The prediction error signal actually has small amplitude values and they can be compressed by means of an entropy coding method such as Rice code, as shown in **Table 2**. It is obvious that the smaller the amplitude, the shorter the code length. In parallel, prediction parameters are actually converted to PARCOR*2 coefficients and quantized. The coefficients are also compressed with a similar Rice code. PARCOR coefficients are convenient for quantization and stability checking. The prediction order can be adaptively set from 0 (no prediction) to 1023. Moreover, progressive order prediction is used at the initial samples of the starting frames of the random

*2 PARCOR (partial auto correlation): A set of predictive parameters invented by the Musashino Electrical Communication Laboratories of Nippon Telegraph and Telephone Public Corporation (now NTT) in 1972. This set has the property of stability and easy quantization, so it is widely used for speech coding and synthesis and in other signal processing areas.

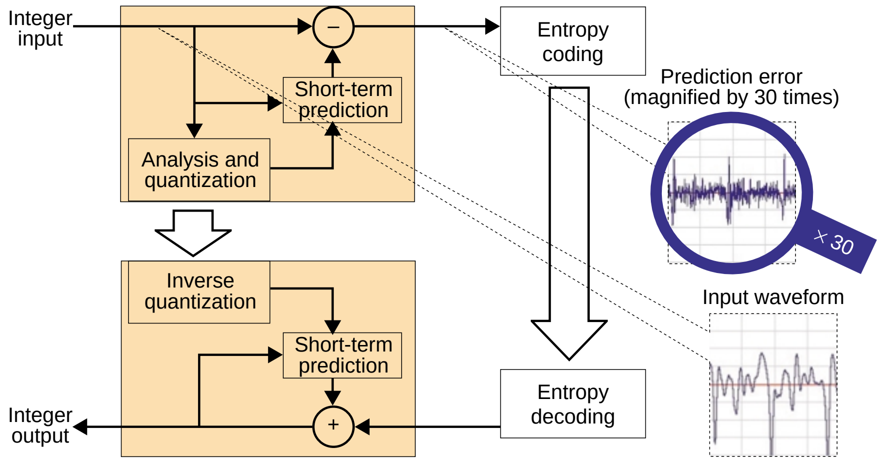


Fig. 2. Fundamental structure of encoder and decoder.

Table 2. Example of Rice code.

Amplitude	Binary code
4	111111110
3	1111110
2	11110
1	110
0	0
–1	10
–2	1110
–3	111110
–4	11111110

access points.

2.2 Long-term prediction and multi-channel prediction

Linear prediction utilizes the correlation between neighboring samples. Speech and audio signals sometimes have long-term correlation due to the pitch. In addition, there is inter-channel correlation between multiple channels. All these correlations can be utilized to enhance the prediction performance and hence reduce the bit rate. The extended prediction scheme is shown in **Fig. 3** [6], [7]. In the encoder, 5-tap (5 independent coefficients) long-term prediction and 3- or 6-tap multi-channel prediction are sequentially applied to the short-term prediction error signal to reduce the amplitude.

For long-term prediction, the best delay parameter from the previous sample and the associated weighting factor are determined. For multi-channel coding, channel-pair combinations are searched for to get the maximum inter-channel correlation. In addition, the relative delay parameter between the channel-pair and the associated weighting factors are determined. All these weighting factors are quantized and compressed by Rice code.

2.3 Floating-point signals

The extended scheme for floating-point signals, which makes use of the fundamental compression structure for the integer signal, is shown in **Fig. 4** [8]-[10]. The floating-point format is useful for professional mixing of music because there is no risk of overflow or underflow. However, this format cannot be compressed because its nominal value has no correlation between samples. In response, we invented a novel scheme that decomposes a floating-point signal sequence into an integer sequence and the remaining sequence. We also invented the ACF (approximate common factor) scheme, which has achieved a great improvement in the compression performance when the input floating-point value sequence is generated from an integer value sequence multiplied by a common number throughout the frame. The common number can be detected by means of rational approximations even though the input samples have errors due to truncation or other operations. The remaining sequence is further compressed by masked LZ[*3] compression, making full use of the properties inherited from the decomposition processes.

2.4 Other features

The ALS standard has some additional operation

*3 LZ (Lempel-Ziv): Universal lossless compression tool invented by Lempel and Ziv and used in ZIP. This tool adaptively updates the codebook depending on the input sequence and is useful for the compression of text and program source code.

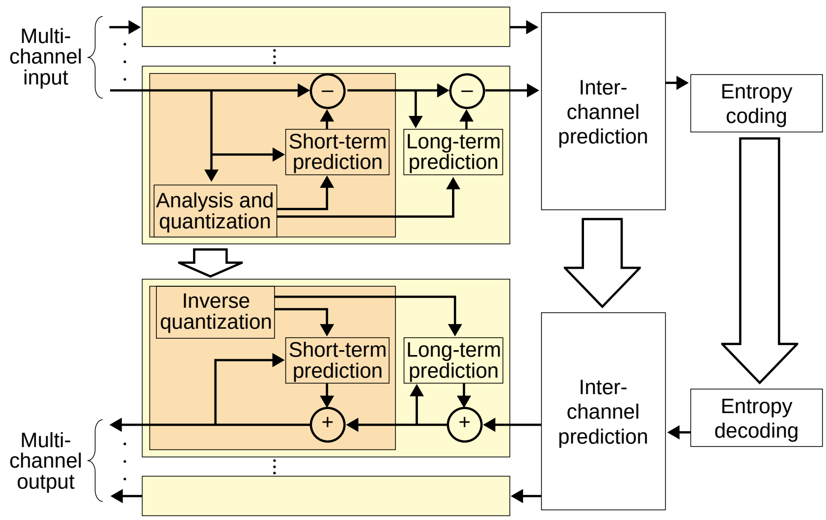


Fig. 3. Extensions for long-term and multi-channel prediction.

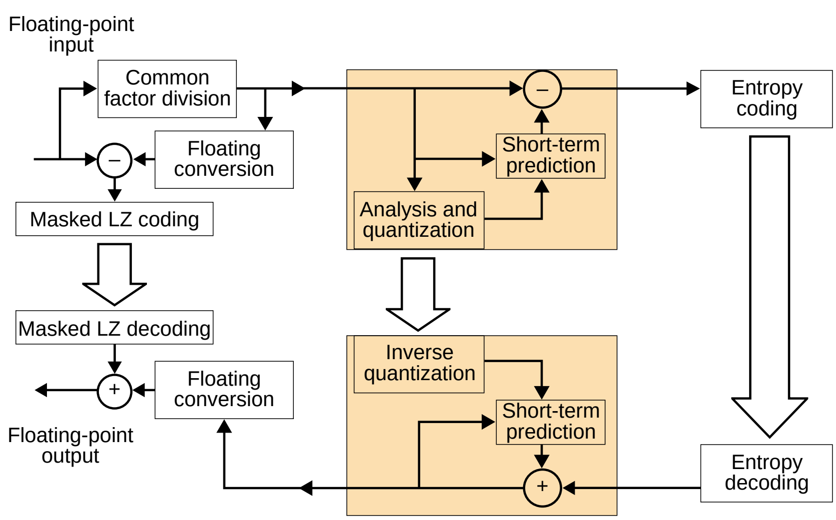


Fig. 4. Extension for floating-point input.

modes. One is a backward prediction mode, which attains efficient compression performance at the cost of longer processing times in the encoder and decoder. Another mode uses a hierarchical adaptive block switching mechanism for higher compression performance. Furthermore, there are several other control parameters at the encoder, which may be efficiently estimated in the future without losing conformity to the standard.

3. Practical evaluation

ALS is compared with other available lossless coding tools (free software and MPEG-4 SLS) in terms of compression ratio (the smaller the better) and decoding time (the smaller the better) in **Fig. 5**. In addition, the performance of NTT's proprietary optimized decoder is shown. We can see from this figure that the standard provides the state-of-the-art performance. Note that the compression performance of ALS outperforms that of general-purpose compression tools such as ZIP, so far as audio signals are concerned.

This standard accepts a wide range of waveform formats as input for compression, as shown in **Table 3**, and can be used for most of the audio-related applications. Encoding can be executable in real time while the music is being played back on a low-end

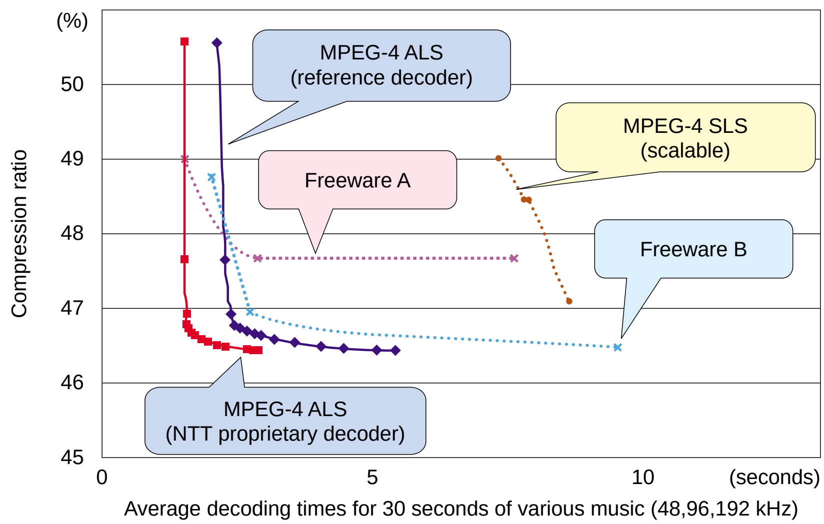


Fig. 5. Comparison of compression performances.

Table 3. Waveform formats supported by MPEG-4 ALS.

	MPEG-4 ALS	CD (reference)
Max. sampling frequency	192 kHz	44.1 kHz
Max. amplitude resolution	32 bits	16 bits
Max. number of channels	65,536	2
Formats	Integer and floating-point value	Integer

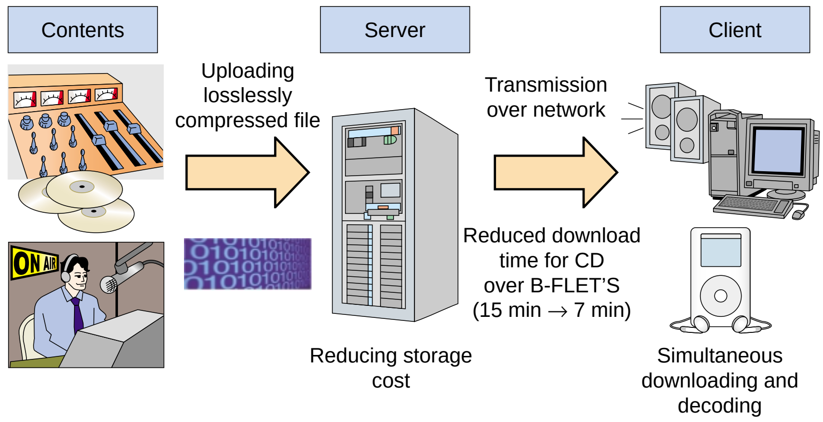


Fig. 6. Example of application scenario (music delivery).

personal computer, and decoding can be executed more than ten times faster than the playback time. In other words, the load on the central processing unit for decoding is negligible for most applications. For a typical server/client music delivery system (**Fig. 6**), the total download time can be significantly reduced because the download file size is significantly reduced, decoding time is less than the download time, and decoding can be executed in parallel with downloading. This holds true even if an optical fiber high-speed connection is used.

4. Future tasks

It is expected that this standard will be used for common tools for various applications and will continue to be maintained so that compressed files can be perfectly decoded even 100 years in the future. It is also expected that a consortium of essential patent holders will be organized for collecting and delivering the patent royalties.

NTT Communication Science Laboratories will continue to support the standardization of conformance and reference software and the enhancement of encoder performance. In parallel, NTT Communications will design and provide integrated delivery and archiving systems making use of practical software that complies with this standard. In addition, NTT Group companies will collaborate with partners or issue licenses to other organizations for various applications including professional audio editing tools, portable music players, and editing and archiving systems for medical and environmental data.

References

[1] "International Standardization of Lossless Coding Technology for Audio Signals," NTT Technical Review, Vol. 4, No. 4, p. 55, 2006.

[2] ISO/IEC 14496-3:2005/Amd.2:2006, Information technology—Coding of audio-visual objects—Part 3: Audio, Amendment 2: Audio Lossless Coding (ALS), new audio profiles and BSAC extensions, edition 2006-03-15.

[3] T. Moriya, D. Yang, and T. Liebchen, "Extended Linear Prediction Tools for Lossless Audio Coding," Proceedings of ICASSP 2004, pp. III-1008-1011, 2004.

[4] T. Liebchen, Y. Reznik, T. Moriya, and D. Yang, "MPEG-4 Audio Lossless Coding," Preprint paper #6047, 116th Audio Engineering Society Convention, Berlin, 2004.

[5] T. Liebchen, T. Moriya, N. Harada, Y. Kamamoto, and Y. Reznik, "The MPEG-4 Audio Lossless Coding (ALS) Standard Technology and Applications," Preprint paper #6589, 119th Audio Engineering Society Convention, New York, 2005.

[6] Y. Kamamoto, T. Moriya, T. Nishimoto, and S. Sagayama, "Lossless compression of multi-channel signals based on inter-channel correlation," IPSJ, Vol. 46, No. 5, pp. 1118 - 1128, 2005 (in Japanese).

[7] Y. Kamamoto, T. Moriya, N. Harada, T. Nishimoto, and S. Sagayama, "Intra- and Inter-Channel Long-Term Prediction in ISO/IEC MPEG-4 Audio Lossless Coding (ALS)," IEICE Trans. on Communications, Vol. J89-B, No. 2, pp. 214-222, 2006 (in Japanese).

[8] D. Yang and T. Moriya, "Lossless Compression for Audio Sources with IEEE Floating Point Format," Preprint 115th Audio Engineering Society Convention, #5987, 2003.

[9] N. Harada, T. Moriya, H. Sekigawa, and K. Shirayanagi, "Lossless Compression of IEEE Floating-point Audio using the Approximate-Common-Factor Coding," Preprint paper #6352, 118th Audio Engineering Society Convention, Barcelona, 2005.

[10] N. Harada, T. Moriya, H. Sekigawa, K. Shirayanagi, and Y. Kamamoto, "Lossless Compression of IEEE754 Floating-Point Signal in ISO/IEC MPEG-4 Audio Lossless Coding (ALS)," IEICE Trans. on Communications, Vol. J89-B, No. 2, pp. 204-213, 2006 (in Japanese).

Takehiro Moriya

Executive Manager, NTT R&D Fellow, Human and Information Science Laboratory, NTT Communication Science Laboratories.

He received the B.S., M.S., and Ph.D. degrees in applied mathematics and instrumentation physics from the University of Tokyo, Tokyo, in 1978, 1980, and 1989, respectively. After joining the Musashino Electrical Communication Laboratories of Nippon Telegraph and Telephone Public Corporation (now NTT) in 1980, he engaged in research on and the standardization of speech and audio coding. In 1989, he stayed at AT&T Bell Laboratories as a guest researcher. He is a member of the Acoustical Society of Japan (ASJ), the Information Processing Society of Japan (IPSJ), and the Institute of Electronics, Information and Communication Engineers (IEICE) of Japan and a fellow of IEEE.

Noboru Harada

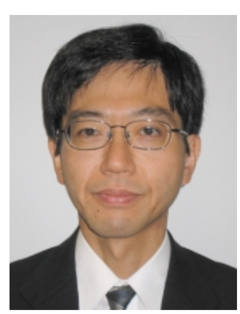
Research Scientist, Human and Information Science Laboratory, NTT Communication Science Laboratories.

He received the B.S. and M.S. degrees in computer science and systems engineering from Kyushu Institute of Technology, Fukuoka, in 1995 and 1997, respectively. He joined NTT Human Interface Laboratories, Tokyo, in 1997. His main research area has been lossless audio coding and high-efficiency coding of speech and audio. He is a member of ASJ, IEICE, and IEEE.

Yutaka Kamamoto

Researcher, Human and Information Science Laboratory, NTT Communication Science Laboratories.

He received the B.S. degree in applied physics from Keio University, Kanagawa, in 2003 and the M.S. degree in information physics and computing from the University of Tokyo, Tokyo, in 2005. Since joining NTT Communication Science Laboratories in 2005, he has been studying signal processing and information theory. He is a member of ASJ, IPSJ, the Society of Information Theory and its Applications, and IEICE.



Hiroshi Sekigawa

Senior Research Scientist, Human and Information Science Laboratory, NTT Communication Science Laboratories.

He received the B.S. and M.S. degrees in mathematics and the Ph.D degree in mathematical sciences from the University of Tokyo, Tokyo, in 1986, 1989, and 2004, respectively. He joined NTT in 1989. He was engaged in research on computer-aided design for LSIs from 1989 to 1994 and on symbolic and algebraic computation and symbolic-numeric computation from 1994 to 2000. In 2000, he was temporarily transferred to Business Communications Headquarters, NTT West, Osaka. Since his return to NTT Communication Science Laboratories in 2002, he has been engaged in research on symbolic-numeric computation and its application to science and engineering problems. He is a member of IPSJ, the Japan Society for Symbolic and Algebraic Computation, IEICE, and the Japan Society for Industrial and Applied Mathematics.