Experimental evaluation of encoding parameters of MPEG-4 ALS for high-resolution audio

Shota Amada*, Yutaka Kamamoto†, Noboru Harada†, Ryosuke Sugiura†,
Takehiro Moriya†, Shoji Makino* and Takeshi Yamada*
* University of Tsukuba, Ibaraki, Japan
† NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation. Kanagawa, Japan
Email:amada@mmlab.cs.tsukuba.ac.jp
{kamamoto.yutaka, harada.noboru, sugiura.ryosuke, moriya.takehiro}@lab.ntt.co.jp
maki@tara.tsukuba.ac.jp takeshi@cs.tsukuba.ac.jp

***Abstract*—This paper describes experimental evaluations of encoding parameters that are appropriate for MPEG-4 Audio Lossless Coding (ALS) to compress high-resolution audio. MPEG-4 ALS Simple Profile defines the values of encoding parameters, such as the maximum sampling frequency and quantization bit depth, for making it easier to implement in the receiving applications. However, ALS Simple Profile does not define values for 96-kHz high-resolution audio. Therefore, we propose a range of values for 96-kHz high-resolution audio by extending the ALS Simple profile and evaluating its performance.**

I. Introduction

According to the Japan Electronics and Information Technology Industries Association (JEITA), high-resolution audio is defined as high quality digital audio of which sampling frequency or quantization bit depth exceeds the compact disc (CD) specification and both are equal to or higher than the CD specification [1]. The bit-rate or storage size of the high-resolution audio is several times that of the CD specification audio of the same recording time. General music CDs of 700-MB (700×1000×1000-byte) capacity can store about 66 minutes of CD specification audio, but same capacity can store only about 20 minutes of high-resolution audio with 96-kHz sampling rate and 24-bit quantization bit depth. For this reason, the needs for compression is increasing. However lossy compression such as MPEG-1/2 Audio Layer-3 (mp3) and MPEG-2/4 Advanced Audio Coding (AAC) makes digital distortion and decrease the sound quality. Lossy coding reduce high frequency components which humans hardly hear for reducing total code length. The high-frequency components are important for the people who select the high-resolution audio, it should not be omitted. If further processing is performed after lossy encoding, the distortion is emphasized and the loss is perceived in some case. Therefore lossy coding is inappropriate for compressing high-resolution audio and lossless compression is appropriate to maintain high sound quality of high-resolution audio.

Moving Picture Experts Group (MPEG), an international standardization organization dealing with moving images and audio data, has standardized MPEG-4 Audio Lossless Coding (ALS) as a lossless compression method of audio signals [2]–[4]. MPEG-4 ALS can handle many types of input signals including high-resolution audio such as one with sampling frequency up to about 4 GHz, quantization bit number up to 32 bit and up to 65536 channels. MPEG-4 ALS Simple Profile (Table I) defines the upper limit of parameters such as sampling frequency, quantization bit depth and number of channels for making it easier to implement in the receiving device compared to ALS without any profiles.

ALS Simple Profile has not defined these limits of parameters for high-resolution audio yet. Referring to the MPEG-4 ALS Simple Profile, for 48 kHz signals, maximum number of samples per frame is 4096 and maximum prediction order is 15. As shown in Fig. 1, when the sampling frequency is doubled to 96 kHz, doubling maximum number of samples per frame to 8192 and maximum prediction order to 31 makes it possible to process 96-kHz signals with the same time window as 48-kHz signals. Therefore, we verify the compression performance with these values of parameters and propose appropriate values of parameters for the 96-kHz high-resolution audio.

TABLE I
Level for the MPEG-4 ALS Simple Profile
(BS: Block Switching tool, MCC: Multi-Channel Coding tool)

Level	1	2	3	4
Max. number of channels	2	2	6	6
Max. sampling rate [kHz]	48	48	48	48
Max. word length [bit]	16	24	16	24
Max. number of samples per frame	4096	4096	4096	4096
Max. prediction order	15	15	15	15
Max. BS stages	3	3	3	3
Max. MCC stages	1	1	1	1

II. Overview of MPEG-4 ALS

In MPEG-4 ALS, lossless compression is performed by linear predictive analysis on input signals and expressing signal by prediction residual signals and partial auto-correlation

978-1-5090-4045-2/17/$31.00 ©2017 IEEE

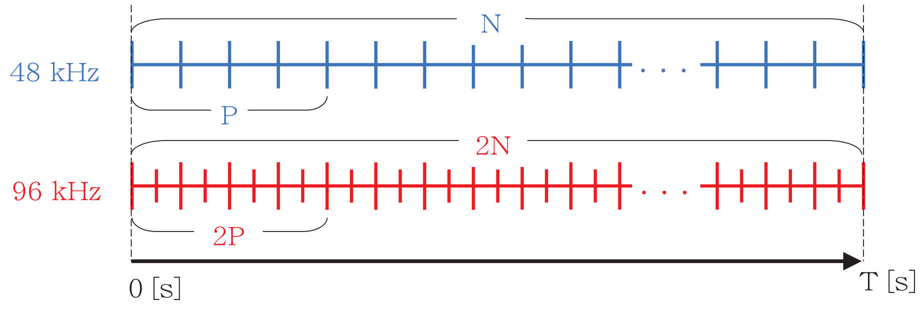

Fig. 1. Difference in number of samples at same time window.

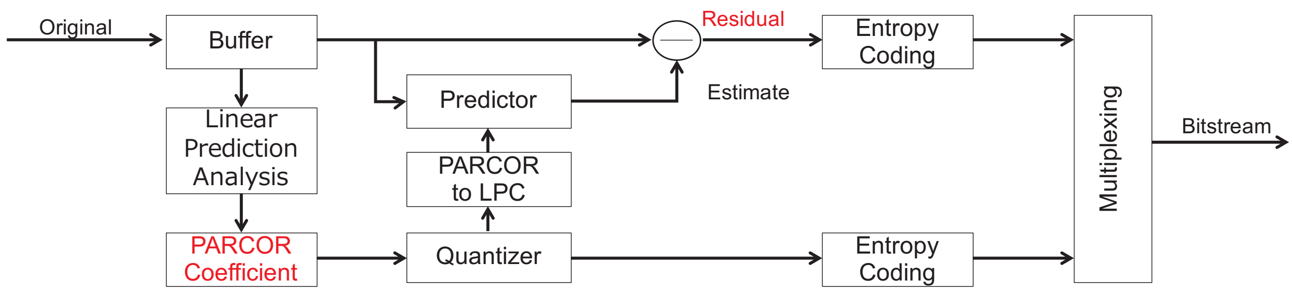

Fig. 2. Overview of MPEG-4 ALS encoder.

(PARCOR) coefficients. Fig. 2 shows overview of the encoder. There are trade-offs between the amount of information representing the prediction residual signal and PARCOR coefficients. Eq. (1) denotes estimated total code length per frames at the frame length is N and the prediction order is P, where $E(0) = \sum_{n=1}^{N}\{x(n)\}^2$, $x(n)$ $(n = 1, 2, \ldots, N)$ is input signal, k_i $(i = 1, 2, \ldots, P)$ are PARCOR coefficients and γ_i are the bits needed to represent the PARCOR coefficients [5].

$$C(P) = N\left\{\beta + \frac{1}{2}\log_2\left(\frac{E(0)}{N}\right) + \frac{1}{2}\sum_{i=1}^{P}\log_2(1-k_i^2)\right\} + \sum_{i=1}^{P}\gamma_i \quad (1)$$

Ignoring the first two constant terms of Eq. (1), then Eq. (2) is obtained. Estimate code length from Eq. (2), select the prediction order considered to be optimum at the limit below the maximum prediction order.

$$\tilde{C}(P) = \sum_{i=1}^{P}\left(\frac{N}{2}\log_2(1-k_i^2) + \gamma_i\right) \quad (2)$$

If the prediction order is increased, the amount of information for transmitting the prediction residual signals ($\frac{N}{2}\log_2(1-k_i^2)$) is reduced but the amount of information for transmitting the PARCOR coefficients (γ_i) is increased. The reverse is also true. Therefore, it is necessary to consider an appropriate prediction order that the total amount of information becomes smaller. The frame length also affects the compression performance.

Thus, it is necessary to assign appropriate prediction orders to each frame with appropriate frame length. Therefore, we investigated the appropriate prediction order and frame length when applying MPEG-4 ALS to high-resolution audio.

III. Experimental Evaluation of Appropriate Parameters

A. Experimental conditions

Table II shows the specifications of the audio that used in the experiment. We prepared 15 audio data including jazz, orchestra, opera and so on. 48-kHz audio data are generated by down-sampling 96-kHz audio data.

TABLE II
Specifications of input sound items.

Sampling rate [kHz]	48	96
Number of channels	2	2
Word length [bit]	24	24
Number of audio files	15	15
Recording time [s]	30	30
File size ($\times 10^6$)[byte]	8.64	17.28

B. Appropriate maximum prediction order and maximum frame length

In this experiment, optimal prediction order is defined as a prediction order by which the file size after encoding is minimized. Fig. 3 shows the process of investigating the encoded filesize with each frame length and prediction order. Perform this operation on 15 sound sources and find the optimum prediction orders at each frame length. The results are averages from the results of 15 sound sources.

Figs. 4 and 5 show the histograms of the optimal prediction orders that obtained by the exhaustive search within allowed

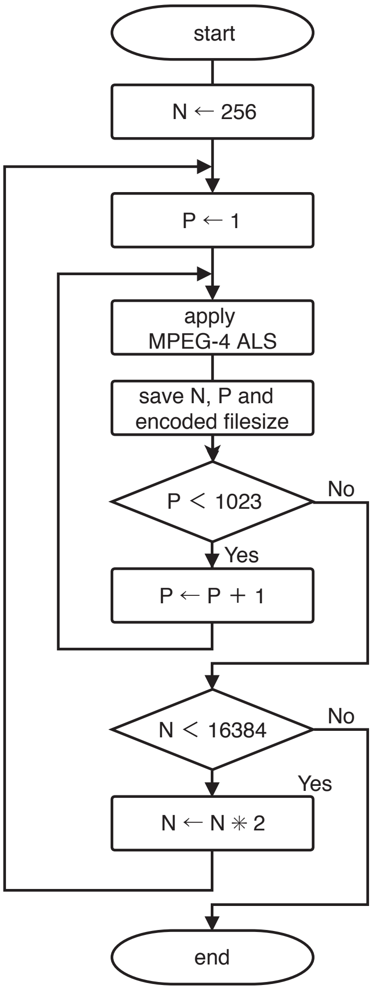

Fig. 3. Process of investigation the encoded filesize with each frame length and prediction order.

range of the prediction order for each frame length of 48-kHz and 96-kHz signals. Each figure plots the occurrence frequencies of optimal orders and peaks in each frame length. The peaks of the optimum prediction orders, for the 48-kHz signals, is located near 15 in each frame length. In the 96-kHz signals, peaks exceed 15 in every frame length and they are located less 31.

Figs. 6 and 7 is the graphs that superposed the curves of cumulative percentage about occurrence frequencies of optimal orders on Figs. 4 and 5, right vertical axis is cumulative percentage of the total number of occurrences. Tables. III and IV show the cumulative percentages at 15-th and 31-th order, 96-kHz and 48-kHz signal, each frame length. Increased cumulative percentages by increasing 15-th to 31-th order are greater at 96-kHz signals. Comparing the tables with the same time window, 15-th order with 48-kHz signals and 31-th order with the 96-kHz signals, each frame length are (1024, 2048, 4096) and (2048, 4096, 8192), these cumulative percentage are close. When the maximum prediction order is increased to 31 at the 96-kHz signals, optimal prediction orders could be handled at the same percentage as the maximum prediction order is 15 at the 48-kHz signals.

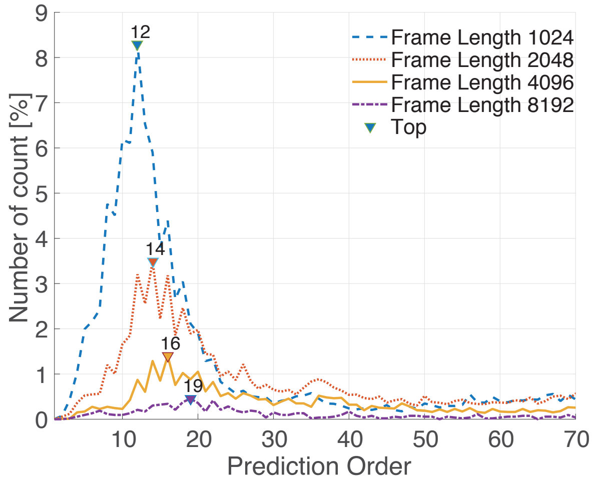

Fig. 4. Histograms of optimal prediction orders for each frame length of 48-kHz signals.

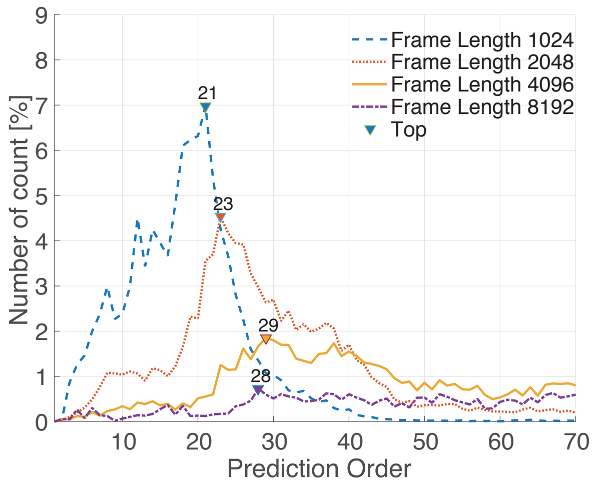

Fig. 5. Histograms of optimal prediction orders for each frame length of 96-kHz signals.

TABLE III
CUMULATIVE PERCENTAGES AT 48-KHZ SIGNALS.

Frame length	1024	2048	4096	8192
Cumulative percentages at 15-th order [%]	54	19	6	2
Cumulative percentages at 31-th order [%]	76	41	16	6
Increased percentages (15-th to 31-th) [%]	22	22	10	4

TABLE IV
CUMULATIVE PERCENTAGES AT 96-KHZ SIGNALS.

Frame length	1024	2048	4096	8192
Cumulative percentages at 15-th order [%]	35	10	3	2
Cumulative percentages at 31-th order [%]	93	57	20	7
Increased percentages (15-th to 31-th) [%]	58	47	17	5

Figs. 8 and 9 show the comparison of compressed file size when we applied MPEG-4 ALS to each sampling frequency signal while changing the frame length and maximum prediction order. The frame length with the minimum file size were respectively 2048 and 4096 samples for 48-kHz and 96-kHz signals. 8192-sample frame length is optimal

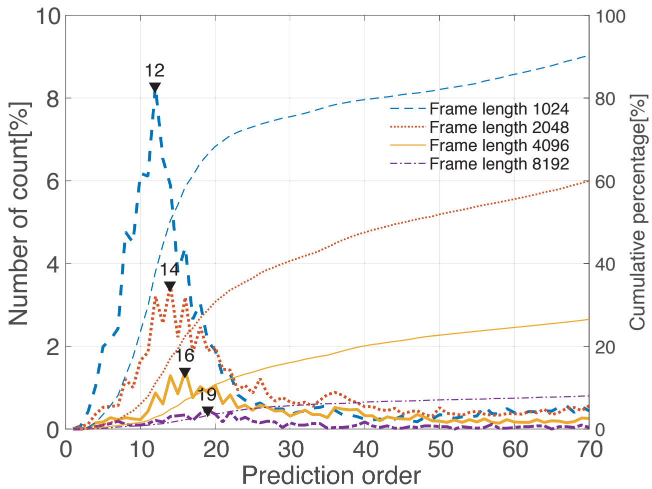

Fig. 6. Cumulative percentages and histograms about optimal prediction orders for each frame length of 48-kHz signals.

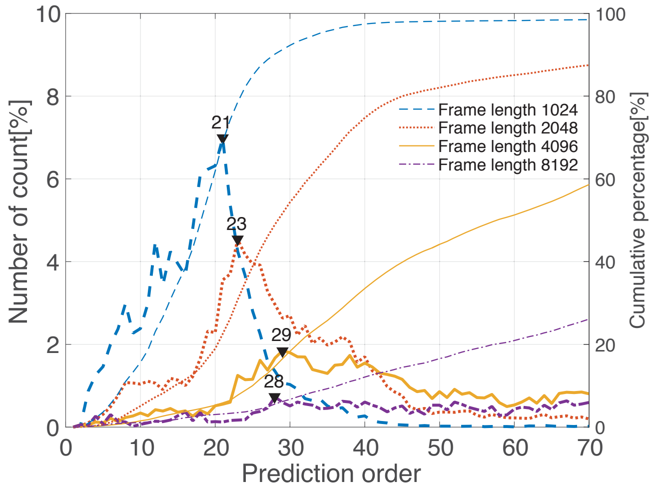

Fig. 7. Cumulative percentages and histograms about optimal prediction orders for each frame length of 96-kHz signals.

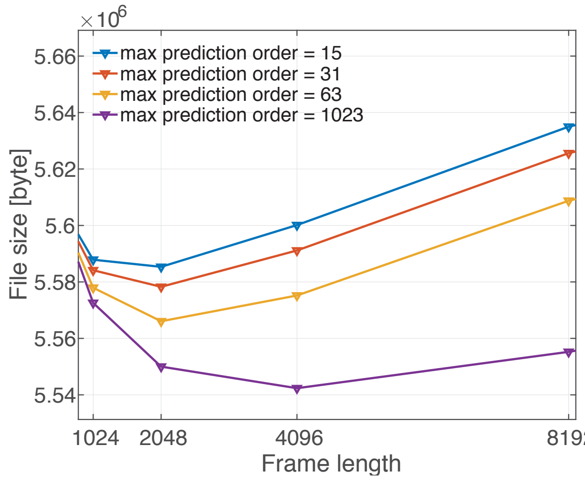

Fig. 8. The relationship between the encoded file size and frame length for 48-kHz signals.

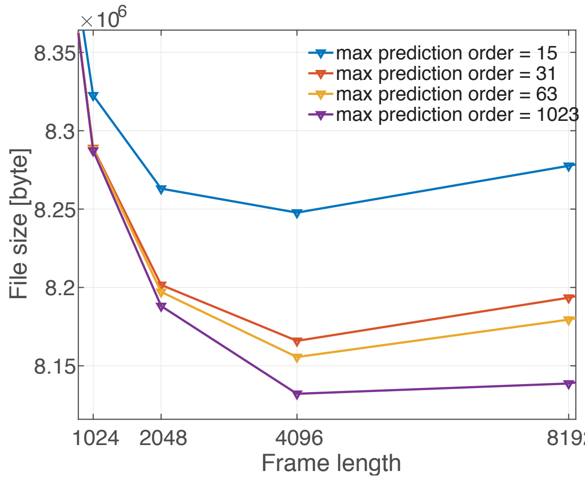

Fig. 9. The relationship between the encoded file size and frame length for 96-kHz signals.

in half of the audio files, so both 4096 and 8192-samples frame length are appropriate for 96-kHz signals. MPEG-4 ALS adopts block length switching, which divides a frame hierarchically into subblocks and can select and use optimal combination of block lengths. Thus if the maximum value of frame length is set to 8192, 4096-samples frame length is also able to use. The compressed file size is decreased when the maximum prediction order increased from 15 to 31. Ratio of this improvement in 96-kHz signals is greater than 48-kHz signal.

IV. Conclusion

For high-resolution audio contents, increasing the maximum prediction order and the maximum frame length can improve the compression ratio, but on the other hand, the amount of calculation and buffer size required for decoding and encoding increases. When MPEG-4 ALS was applied to high-resolution audio with 96-kHz sampling frequency, the appropriate frame length is 4096 or 8192. MPEG-4 ALS adopts block length switching, so it is reasonable to set the maximum value of frame length to 8192. As shown in the experiment result, it is expected that we can improve compression ratio more by extending the maximum prediction order to 31 for the 96-kHz signal.

In conclusion, when applying MPEG-4 ALS Simple Profile to 96-kHz high-resolution audio, the appropriate maximum frame length is 8192-samples and the maximum prediction order is 31.

References

[1] 25JEITA-CP42, 2014. http://home.jeita.or.jp/page_file/20140328095728_rhsiN0Pz8x.pdf

[2] ISO/IEC 14496-3:2009/Amd 3:2015, 2009 Information technology – Coding of audio-visual objects – Part 3: Audio

[3] T. Liebechen, T. Moriya, N. Harada, Y. Kamamoto, and Y. Reznik "The MPEG-4 Audio Lossless Coding (ALS) Standard - Technology and Applications," AES, 2005.

[4] T. Liebchen and Y. Reznik "MPEG-4 ALS : an emerging standard for lossless audio coding," IEEE, Data Compression Conference, 2004.

[5] Y. Kamamoto, *et al.* "Low-complexity PARCOR coefficient quantization and prediction order estimation designed for entropy coding of prediction residuals," 2013