# 音響ロスレス符号化MPEG-4 ALSにおけるハイレゾ音源向け線形予測次数最適化に関する検討と考察*

☆天田将太（筑波大）, 鎌本優, 原田登, 杉浦亮介, 守谷健弘（NTT）, 山田武志, 牧野昭二（筑波大)

## 1 はじめに

近年、ハイレゾ音源 [1] と呼ばれる高精細デジタルオーディオが普及しつつある。ハイレゾ音源は高品質である一方、その情報量は同録音時間の CD 音源と比べて数倍となる。一般的な音楽用 CD 容量である 700 MB= $700 \times 1024 \times 1024 \times 8$ bit 分の記録をすると CD スペック (44.1 kHz, 16 bit) では約 70 分間を記録できる。しかしこれを 96 kHz, 24 bit ハイレゾで記録すると約 21 分間しか記録できない。そのため圧縮の重要性がより高まるが、音質を高く保つためには可逆圧縮が必要である。ハイレゾ音源を効率よく伝送するために可逆圧縮方式の圧縮率を改善していくことが求められる。

動画や音声データを扱う国際標準化団体である MPEG (Moving Picture Experts Group) では音響信号の可逆符号化方式として MPEG-4 Audio Lossless Coding (ALS) [2–4] を規格化している。

MPEG-4 ALS は、サンプリング周波数最大約 4 GHz、量子化ビット数 32 bit に対応するなど CD スペックの音源のみならず、ハイレゾを含む幅広い入力信号を扱える。ただし受信端末への実装を考慮し、MPEG-4 ALS Simple Profile が定義されている (Table 1)。Simple Profile は最大チャネル数やサンプリング周波数、量子化ビット深度等を限定することで演算量やメモリ量を制限している。

本稿では ALS Simple Profile をサンプリング周波数 96 kHz のハイレゾ音源に対応させるために性能を調査し、ハイレゾ音源の圧縮に適した符号化パラメータを検討した結果について報告する。

## 2 MPEG-4 ALS のハイレゾ音源適応

### 2.1 MPEG-4 ALS

MPEG-4 ALS では入力信号について線形予測分析を行い、予測残差信号と、PARCOR 係数 (partial auto-correlation) の 2 つを用いて表現することで可逆圧縮を行っている。ここで予測次数を大きくすれば予測残差信号を表現するために必要な情報量は小さくなるが、PARCOR 係数を表現するために必要な情報量は大きくなってしまう。つまり予測残差信号のための情報量と PARCOR 係数のための情報量にはトレードオフの関係がある。よって全体の情報量がより小さくなる適切な予測次数を検討する必要がある。MPEG-4 ALS ではフレームごとに予測次数を適応的に変化させることができるため、適切な予測次数を求めることにより圧縮性能を改善することができる。また入力信号はフレームに分割されて処理されるため、フレーム長は圧縮性能に影響を与える。

Table 1 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 |

### 2.2 適切な最大フレーム長と最大予測次数

我々は MPEG-4 ALS にハイレゾ音源を適用する際に適切な最大予測次数とフレーム長の調査を行ってきた [5]。最大予測次数と最大フレーム長は用途に応じた選択をすることが重要であり、圧縮率の向上を優先するならば最大予測次数 31 次、最大フレーム長 8192 が適切である。また、演算量やバッファサイズの増加を抑え、短時間計算および省消費電力を優先するならば最大予測次数 15 次、最大フレーム長 4096 が適切である。

### 2.3 ブロック分割機能と長期予測機能

MPEG-4 ALS ではブロック分割機能 [7] が採用されており、フレームを階層的にサブブロックに分割し、最適なブロック長の組み合わせを選択し利用できる。定常な信号には長いブロック長を割り当て、非定常な信号に短いブロック長を割り当てることで圧縮性能を改善することができる。

また長期予測機能 [7] も採用されている。音声信号や音楽信号には長期の相関を持つ場合があり、このとき予測残差も相関を持つことになる。長期予測機能

*Experimental evaluation of the linear prediction order of MPEG-4 ALS for High-resolution audio. by Shota Amada (University of Tsukuba), Yutaka Kamamoto, Noboru Harada, Ryosuke Sugiura, Moriya Takehiro (NTT), Takeshi Yamada and Shoji Makino (University of Tsukuba).

ではこの相関を利用してさらに予測残差信号の振幅値を縮小できる。本稿における実験では、ブロック分割機能と長期予測機能両方を用いている。

### 2.4 最適予測次数選択

MPEG-4 ALS ではフレームごとに予測次数を適応的に変化させることができ、予測次数は圧縮後の符号量の推定式をもとに選択される [6]。フレーム長 $N$, 予測次数 $P$ のとき、信号フレーム $x(n)$ の予測残差信号の分散は次式のように表せる。

$$\sigma_P^2 = \frac{1}{N}\left(\sum_{n=1}^{N}\{x(n)\}^2\right)\prod_{i=1}^{P}(1-k_i^2) \quad (1)$$

ここで $k_i \quad (i = 1, 2, \cdots, P)$ は PARCOR 係数である。信号がガウス分布に従う場合のエントロピーを $H_G(\sigma_P^2)$、ラプラス分布に従う場合のエントロピーを $H_L(\sigma_P^2)$ とすると、それぞれ

$$H_G(\sigma_P^2) = \log_2\left(\sqrt{2\pi e\sigma_P^2}\right) \quad (2)$$

$$H_L(\sigma_P^2) = \log_2\left(\sqrt{2\ e^2\sigma_P^2}\right) \quad (3)$$

となる。いずれの場合もエントロピーは分散 $\sigma^2$ に依存するため、

$$H(\sigma_P^2) = \beta + \frac{1}{2}\log_2(\sigma_P^2) \quad (4)$$

と表せる。$E(0) = \sum_{n=1}^{N}\{x(n)\}^2$ とおくと、1 フレームあたりの予測残差信号の符号量 $C_e(P)$ は次式のように表せる。

$$\begin{aligned} C_e(P) &= H(\sigma_P^2)N \\ &= N\left\{\beta + \frac{1}{2}\log_2\left(\frac{E(0)}{N}\right) + \frac{1}{2}\sum_{i=1}^{P}\log_2(1-k_i^2)\right\} \end{aligned} \quad (5)$$

式 (5) に量子化された PARCOR 係数を表すために必要な符号量 $\sum_{i=1}^{P}\gamma_i$ を加えたものが 1 フレームあたりの符号量 $C(P)$ となる。

$$C(P) = N\left\{\beta + \frac{1}{2}\log_2\left(\frac{E(0)}{N}\right) + \frac{1}{2}\sum_{i=1}^{P}\log_2(1-k_i^2)\right\} + \sum_{i=1}^{P}\gamma_i \quad (6)$$

さらに式 (6) から定数項を除くと次式のようになる。

$$\tilde{C}(P) = \sum_{i=1}^{P}\left(\frac{N}{2}\log_2(1-k_i^2) + \gamma_i\right) \quad (7)$$

この符号量の推定式を用いて各フレーム毎に用いる予測次数を選択している。

## 3 適切な重み係数の調査

### 3.1 調査内容

式 (7) に示した推定符号量は、予測残差信号の分布がガウス分布やラプラス分布に従うと仮定している。この分布の仮定と実際の信号の分布に差がある場合を考慮し、予測残差の符号量に重み係数 $\omega$ をかけることで調整を行なっている。

$$\tilde{C}(P) = \sum_{i=1}^{P}\left(\omega N\log_2(1-k_i^2) + \gamma_i\right) \quad (8)$$

このとき $\omega = 0.50$ がガウス分布、ラプラス分布における理論値である。96 kHz ハイレゾ音源を適用する場合における重み係数 $\omega$ の実験的な最適値について調査を行った。

### 3.2 実験条件

実験に用いた音源の情報を Table 2 に示す。ジャズやオーケストラ、オペラを含む 15 音源を用いており、実験の結果はこれら 15 音源それぞれの結果を平均した値を示している。実験時に利用した各パラメータの組み合わせを Table 3 に示す。ALS Simple Profile で許可されているように、ブロック分割機能によりフレームを 3 回まで階層的に分割したときのブロック長の組み合わせを利用できる。また、長期予測機能も利用している。

Table 2 Specifications of input sound items.

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

Table 3 Combination of each coding parameter.

| | |
|---|---|
| Max. Frame length | 4096, 8192 |
| Max. Prediction order | 15, 31 |
| Max. Block switching stages | 3 |
| Bias parameter $\omega$ | 0.01, 0.02, ..., 2.00 |

### 3.3 圧縮後ファイルサイズと重み係数

96 kHz 音源に MPEG-4 ALS を適用する際に重み係数を変化させた場合の圧縮後ファイルサイズを比較する。最大フレーム長 4096、最大予測次数 15 次および最大フレーム長 8192、最大予測次数 31 次についてプロットした結果を Fig. 1 に示す。最大予測次数が 15 次の場合は理論値である $\omega = 0.50$ 付近でファイルサイズの減少は落ち着いている。対して最大

予測次数が 31 次の場合は $\omega = 0.50$〜$0.9$ の区間でもファイルサイズはゆるやかに減少し、$\omega = 0.90$〜$1.5$ の区間でさらに減少の度合いは大きくなっている。圧縮後ファイルサイズが最小となるのは、最大フレーム長 8192、最大予測次数 31 次において $\omega = 1.55$ の場合である。

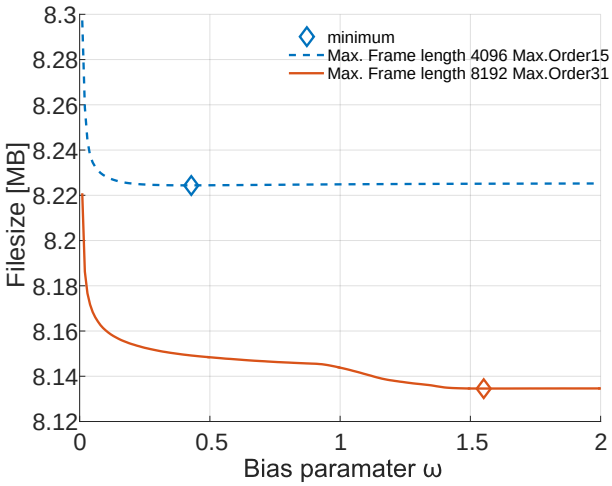


Fig. 1　The relationship between the encoded file size and bias parameter for 96-kHz signals.

### 3.4　選択予測次数の割合と重み係数

96 kHz 音源に MPEG-4 ALS を適用する際に重み係数を変化させた場合の選択予測次数の割合を比較する。最大フレーム長 8192、最大予測次数 31 次の $\omega = 0.50, 0.90, 1.55$ のそれぞれについて選択された予測次数の割合を表すヒストグラムを Fig. 2-4 に示す。$\omega = 0.50$ と $\omega = 0.90$ では割合に大きな変化はないが、$\omega = 1.55$ では最大予測次数である 31 次の割合が急増していることが確認できる。選択予測次数が大きくなると演算量が増加するため、$\omega = 1.55$ では演算量が急増していると考えられる。

### 3.5　演算量と重み係数

96 kHz 音源に MPEG-4 ALS を適用する際に重み係数を変化させた場合の演算量の概算を比較する。$\omega = 0.50$ の時を 100% として割合でプロットした結果を Fig. 5 に示す。前述した選択予測次数の割合からの予想通り、$\omega = 0.90$〜$1.55$ の区間で演算量の増加の割合は大きくなっていることが確認できる。

### 3.6　圧縮効率と重み係数

96 kHz 音源に MPEG-4 ALS を適用する際に重み係数を変化させた場合の圧縮効率を比較する。ここでいう圧縮効率とは、演算量あたりのファイルサイズの削減率を表す。$\omega = 0.50$ の時を 100% としてプロットした結果を Fig. 6 に示す。$\omega = 0.50$〜$0.90$ ではゆるやかに減少していき、$\omega = 0.9$〜$1.55$ の区間では大きく減少していることが確認できる。

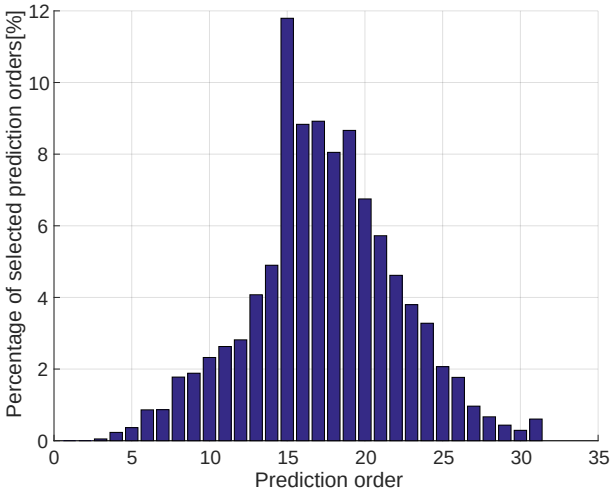


Fig. 2　Percentage of selected prediction order for 96-kHz signals. Bias parameter $\omega = 0.50$

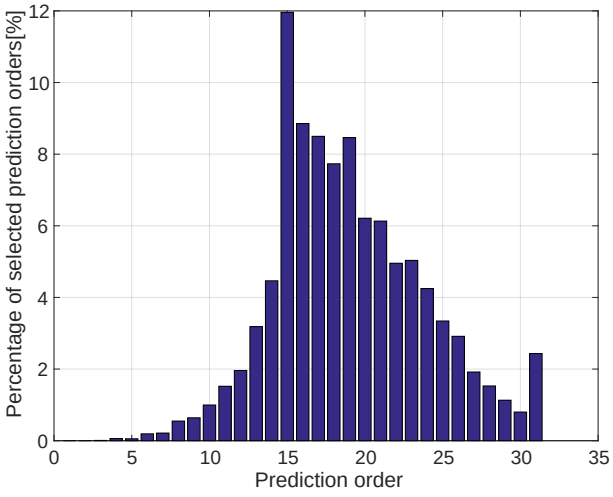


Fig. 3　Percentage of selected prediction order for 96-kHz signals. Bias parameter $\omega = 0.90$

### 3.7　適切な重み係数

最大予測次数が 31 次のとき圧縮率が最高となる重み係数は $\omega = 1.55$ である。しかし演算量も考慮して圧縮効率を考えると $\omega = 1.55$ ではファイルサイズの削減率に対して演算量の増加率が大きいため、適切ではない。最大予測次数が 15 次の場合ファイルサイズ最小となる重み係数は $\omega = 0.50$ 付近であること、最大予測次数が 31 次の場合では $\omega = 0.50$〜$0.90$ であれば圧縮効率は良いことを考慮すると、MPEG-4 ALS に 96 kHz ハイレゾ音源を適用する際に適切な重み係数は理論値 $\omega = 0.50$ だと考えられる。

## 4　おわりに

MPEG-4 ALS では、フレーム長 $N$、予測次数 $P$ のときの圧縮後の符号量を推定することで予測次数を

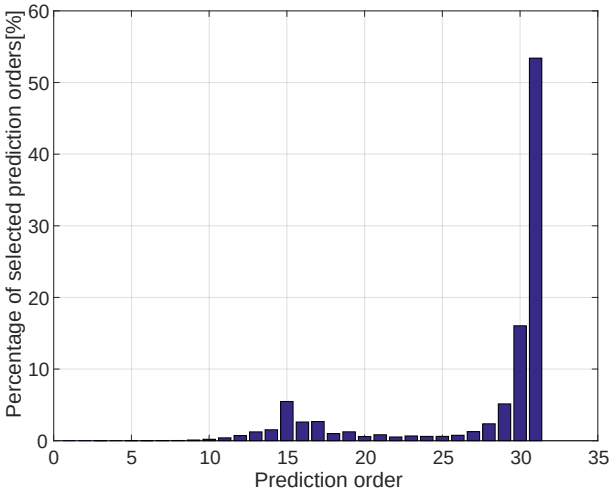


Fig. 4　Percentage of selected prediction order for 96-kHz signals. Bias parameter $\omega = 0.155$

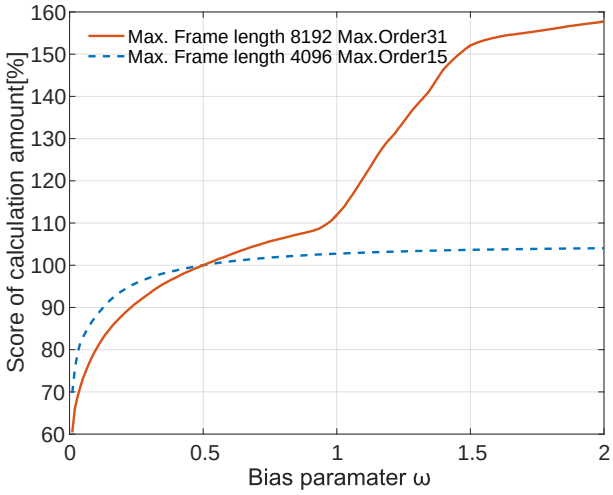


Fig. 5　The relationship between the score of calculation amount and bias parameter for 96-kHz signals.

選択している。このとき予測残差の符号量にかかる重み係数の値の適切値が 96 kHz 音源において変化するかを調査した。圧縮率の視点で見た重み係数の最適値は最大予測次数が 15 次の場合では $\omega = 0.50$ 付近であり理論値に近いが、最大予測次数が 31 次の場合では $\omega = 1.55$ と理論値から大きく離れた値となった。$\omega = 1.55$ を用いた場合の選択予測次数は最大予測次数付近に集中しており、$\omega = 0.50$〜$0.90$ の区間と $\omega = 0.90$〜$1.55$ の区間では演算量の増加の度合いが異なることを確認した。$\omega = 0.90$〜$1.55$ の区間では圧縮率の改善量に対する演算量の増加量が大きく、この区間の重み係数は適切ではない。$\omega = 0.50$〜$0.90$ の区間では圧縮率の改善量に対する演算量の増加度は良いためこの区間の重み係数を用いるべきであり、最大予測次数が 15 次の場合では $\omega = 0.50$ 付近で圧縮率最良であることを考慮すると、MPEG-4 ALS に

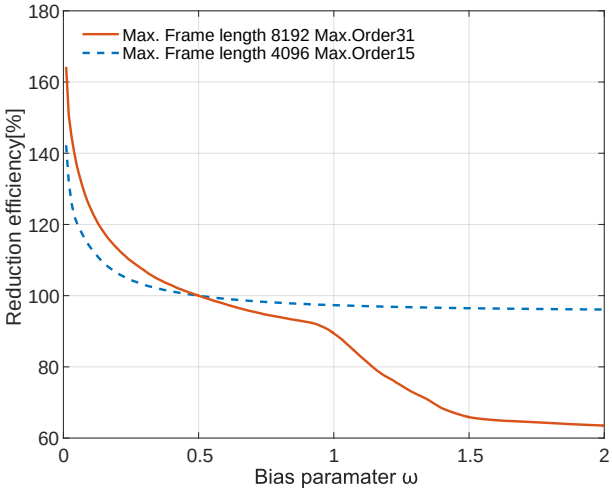


Fig. 6　The relationship between the reduction efficiency and bias parameter for 96-kHz signals.

96 kHz ハイレゾ音源を適用する場合に適切な重み係数は $\omega = 0.50$ 付近と考えられる。

圧縮率の改善量のみを考えた場合の最適重み係数は $\omega = 1.55$ であったが、$\omega = 0.50$ 付近における圧縮率と比較して劇的な改善ではなく、演算量の増加度を考慮すると適切な重み係数は $\omega = 0.50$ 付近であると思われる。

## 参考文献


[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,*et al.* "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] 天田将太, 他" 音響ロスレス符号化 MPEG-4 ALS のハイレゾ音源適応の検討と考察," ASJ, 2017.

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

[7] 鎌本優, 他 " ロスレス・オーディオ符号化 MPEG-4 ALS の高性能化," NTT 技術ジャーナル, x2008.