

## 符号化ドメインにおける効率的 MP3-AAC 間変換方式の一検討

高木 幸一<sup>†</sup> 宮地 悟史<sup>†</sup> 滝嶋 康弘<sup>†</sup>

<sup>†</sup>株式会社 KDDI 研究所 〒356-8502 埼玉県ふじみ野市大原 2-1-15

E-mail: <sup>†</sup>ko-takagi@kddilabs.jp

あらまし 本稿では、MP3 ビットストリームから AAC ビットストリームへの効率的な変換方式について提案する。一般に、異コーデック間における変換を実現するためには“トランスコーディング”と呼ばれるデコードおよび再エンコードが必要となる。しかしながら同方式における再エンコードは音質の劣化を伴うばかりでなく、大きく時間を要すると言う問題がある。そこで、本稿では、MP3 のフレーム構造と量子化スケールサイズを AAC に継承する手法を提案する。同手法により、音質を劣化させることなく、AAC の符号化プロセスにおいて最も時間を要する繰り返しプロセスが簡略化でき、結果として高速化が実現可能となる。実験結果により、提案手法が高音質を保持したまま符号化ドメインにおける高速変換が可能であることを検証した。

キーワード MP3, AAC, 符号化ドメイン, トランスコーディング

### A Study on an Efficient MP3 - AAC Conversion on the Coded Domain

Koichi TAKAGI<sup>†</sup> Satoshi MIYAJI<sup>†</sup> and Yasuhiro TAKISHIMA<sup>†</sup>

<sup>†</sup>KDDI R&D Labs. Inc. 2-1-15 Ohara, Fujimino-shi, Saitama, 356-8502 Japan

E-mail: <sup>†</sup>ko-takagi@kddilabs.jp

**Abstract** In this paper, we propose an efficient conversion algorithm from an MP3 stream into AAC. Generally, this kind of conversion, “transcoding,” requires full-decoding and re-encoding. However, the re-encoding based transcoding process may cause quality degradation and take a longer time. This paper proposes a transcoding method where the AAC encoding process inherits the frame structure and the quantization scales from the MP3 bitstream. This enables a reduction of iterative procedure which requires the most of processing time of AAC encoding without incurring quality degradation. Experimental results show that the proposed method realizes high speed coded domain transcoding while maintaining a higher level of audio quality.

**Keyword** MP3, AAC, Coded domain, Transcoding

#### 1. はじめに

携帯電話などの小型のデバイスに対し各種圧縮オーディオデコーダが容易に実装できるようになってきている。そのため、デジタルオーディオ市場は急速な進化を遂げている。また市場の伸びと同時に、オーディオ圧縮方式も機能性・圧縮率の観点から進化したものがいくつか検討されつつある。

そのような中で、ISO で標準化された MP3[1,2]が有名である。MP3 で圧縮されたコンテンツが市場に多く存在すると同時に、MP3 デコーダが非常に多くのポータブル再生デバイスに実装されている。

一方で、MP3 とは互換性がないが、音質を保ったままより高い圧縮率を実現した符号化方式として AAC[3,4]が規定されている。同方式は、オンライン音楽配信や各種放送サービスなどで採用されている。また、CD 品質の音声を一チャンネルあたり約 48~64 kbps で実現でき、MP3 と比

較し、約 30%符号量を削減できると言われている。そのため、AAC に関連する製品やサービスが今後市場に多く出現することが予想される。AAC のデコーダのみを実装したデバイスでは MP3 で圧縮されたデータをそのまま再生することができない。そのため、すでに MP3 で圧縮されたファイルを AAC ファイルに変換したいと言う要望が生まれる。そこで、本稿では、MP3 を AAC に変換する手法について検討する。

MP3 を AAC に変換するための最も簡単な方法として、“非圧縮ドメインでのトランスコーディング”が挙げられる。これは、MP3 を完全にデコードし、その後で AAC に再エンコードすることを意味する。しかしながら、非圧縮ドメインにおける変換は、その処理に時間を要するだけでなく、品質の劣化を伴う可能性がある。同処理では、2種類の異なるエンコーダが1つのオーディオデータに適用されるため、符号化特性の違いによる異なる雑音が混入するためである。

ところで、ビデオについても同一の課題が存在する。実際に MPEG のビデオストリームを非圧縮ドメインでトランスコードした場合、映像品質が大きく劣化することが報告されている[6]。そこで、異なる MPEG ビデオ符号化方式間でトランスコードするために、劣化を最小限にとどめると同時に、それを高速に実現するための手法が数多く検討されている[7]。しかしながら、異なるオーディオ圧縮方式間のトランスコーディングについては報告例がない。

そこで本稿では、MP3 から AAC への変換をターゲットとし、同変換における音質の劣化を最小限にとどめつつ高速に変換するため、同変換を圧縮ドメインで可能とする手法を提案する。はじめに、MP3 と AAC の符号化パラメータとして利用されているサイド情報、スケールファクタ、MDCT 係数などについて、両符号化方式におけるそれらの違いおよび類似性について分析する。そして、その類似性を利用し、MP3 から AAC へ非圧縮ドメインを介し単純に再符号化する方法に対し高速かつ高品質に変換を実現するための方法について述べる。

以下、提案手法の概要および関連技術について 2 章に述べた後、3 章でそれを既存の手法と比較し評価する。そして、4 章で本稿をまとめる。

## 2. 提案手法

本節では、はじめにオーディオデータの変換に関する関連技術を概説し、その後、MP3 と AAC の概要について述べる。それから MP3 と AAC の量子化スケールサイズの類似性について述べ、最後にそれを利用し MP3 から AAC へ的高速変換を実現する提案手法について述べる。

### 2.1. 関連技術

これまで、デジタルビデオおよびオーディオデータの変換についてはさまざまな方法が提案されてきている。

まず、非圧縮ドメインにおける変換として最も多く検討がされてきているのが、サンプリング周波数変換技術である(例えば[5])。それらの多くは Shannon-染谷の標本化

定理をベースにしている。一方で、圧縮ドメインにおける変換について、MPEG-1 Audio Layer I/II のビットレート変換技術が議論されている[8]。ここでは、帯域制限による高周波成分の削減と再量子化により、音質を保ったまま高速にビットレートの削減を実現している。しかしながら、同方式は同一符号化方式内での変換技術であり、異オーディオ符号化方式間の直接変換技術は検討されていない。

### 2.2. MP3 と AAC

本節では MP3 と AAC の概要について述べる。

MP3 として知られている MPEG-1/2 LayerIII には多くの技術が導入されている。デコーダの構造を図 1 の左に示す。

MP3 では、Layer I/II 同様に原音を 32 個のサブバンド信号に分割する。その後、サブバンドごとにフレームを半分ずつオーバーラップさせながら、6 点 (short) または 18 点(long)の MDCT(modified DCT)を適用する。6 点の MDCT は特にプリエコーが起こるような時間的に急峻に変化する部分に対し適用される。そして、各サブバンド信号を臨界帯域の分布に近づけるために、スペクトル係数を臨界帯域ごとに分割しそれぞれにスケールファクタバンドを割り当てる。

一方、AAC は、MP3 を拡張し、さらに高品質な圧縮符号化を実現することを目的として検討された符号化方式である。そのため、AAC で利用されている多くの技術は MP3 でもすでに導入されている。AAC のエンコーダの構造を図 1 の右に示す。

AAC において MP3 と比較し音質が向上する理由の一つとして、MDCT フィルタバンクの導入が挙げられる。AAC ではフィルタバンクのサイズを 128 点と 1024 点の間で適応的に変動させることが可能である (256 点および 2048 点の窓を適用し、うち半分をオーバーラップさせる)。オーディオ信号では、高周波数解像度のフィルタバンクがしばしば必要とされるが、AAC では 1024 点のフィルタバンクを適用可能であるのに対し、

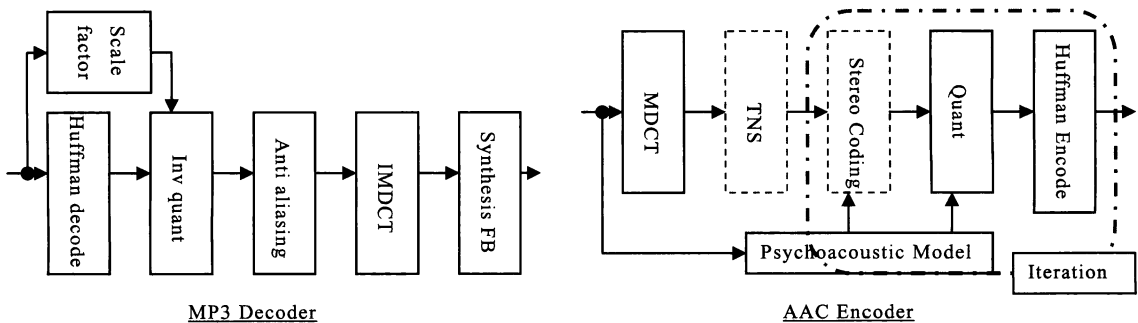


図 1. MP3 Decoder および AAC Encoder の構成

MP3は高くても576点までである。一方、過渡的な信号の場合は高い周波数解像度が必要とされないため、MP3,AACともに適応的に窓長をより短く設定することが可能となっている。この場合、周波数解像度の設定をMP3では192点で行うのに対し、AACでは128点で行うことが可能である。また、AACではMP3で採用されているQMFとMDCTのハイブリッド符号化が利用されないため、クロストークによるエイリアス削減を必要としない。

以上、MP3とAACについて述べたが、両者の違いの一つは周波数ドメインへの変換の方式である。さらに変換のための各フレームサイズも異なる。以上の理由により、圧縮ドメインにおいてMP3をAACに変換することは容易ではない。

一般に、MP3とAACの符号化では、同一の心理聴覚モデルを用いて符号ビットの割り当てが行われる。さらに、符号割り当ては一般的に反復演算を行うため、これに必要となる時間はエンコード中で大半を占める。したがって、この処理のための時間を削減することはオーディオ符号化全体の時間を大きく削減することにつながる。それゆえ、反復演算で決定されるAACの変数がMP3のパラメータから計算できることが望ましい。

### 2.3. 予備実験

前節で述べたとおり、MP3とAACのデコーダにおいてフィルタバンクの構造は大きく異なっている。したがって、同処理に関しては、パラメータを継承せずそのまま行うことが望ましい。一方で、両者のブロックダイアグラムには類似する点が多い。そこで、フィルタバンク適用直前に位置する再量子化プロセスに着目する。

まず、MP3の再量子化係数はスケールファクタバンドごとのスケールファクタ値を使って以下のように計算される。

$$xr[i] = is[i]^{\frac{4}{3}} \times 2^{\frac{1}{4}A} \quad (1)$$

$$A = \text{global\_gain}[gr] - 64 - 8\text{subblock\_gain}[w][gr] - 2(1 + \text{scalefac\_scale}[gr])(\text{scalefac}[cb][w][gr] + \text{preflag}[gr]\text{pretab}[cb])$$

ここで、 $xr[i]$ 、 $is[i]$ はそれぞれ量子化MDCT係数およびMDCT係数を意味し、 $i$ 、 $gr$ 、 $w$ 、 $cb$ はそれぞれMDCT係数、グラニューール、適用される窓、および符号帳のインデックスをそれぞれ示す。一方で、AACの再量子化係数もまたスケールファクタを利用し、以下の通り計算される。

$$xr[i] = is[i]^{4/3} \times 2^{\frac{1}{4}(\text{scale\_facbr}[g][sfb]-100)} \quad (2)$$

ここで、 $g$ 、 $sfb$ はそれぞれ窓関数適用グループおよ

びスケールファクタバンドのインデックスを意味する。式(1),(2)において2の指数部で示されるこれらの量子化スケール値の対数分布は図2のように表現される。MP3とAACの量子化スケールサイズに関して絶対的な値は異なるが、その変動に相関があることがグラフからわかる。

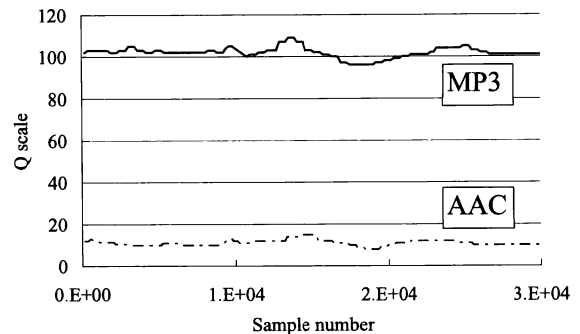


図2. 量子化スケールの変動

### 2.4. 提案手法

前節の予備実験の結果に基づき、本節ではMP3の窓構造および量子化スケール値をAACに継承する方法を利用したMP3→AAC変換手法について、その詳細を述べる。

1. 【初期化】 $N$ フレーム分のMP3データをバッファに格納する。MP3における変換の単位は1152sampleであり、AACは1024sampleである。これらの最小公倍数は9216sampleなので、 $N=8(=9216/1152)$ と設定する。これにより、MP3とAACのフレームサイズの違いの問題を吸収する。なお、 $N$ は8の倍数であればどのような値でもよい。
2. バッファ内のMP3データをデコードする。デコードの際に逆量子化スケール値を変数 $Q[i]$ ( $i$ はサンプル番号)に、また、各時刻におけるフレーム構造(longかshortか)を $\text{frame}[k]$ ( $k$ はMP3におけるフレーム番号)にそれぞれ保存する。
3. 2でデコードされたデータに対し、MDCTを適用する。当該時刻に対し抽出されたフレーム構造がshort blockであった場合、相当する時刻に対しshort windowを選択する。詳細を図3に示す。ここで、同図上部に書かれたMP3の各フレームにおいて、short blockとなるフレームをSと表記する。これを(同図下部に書かれた)AACのフレームに照らし合わせたとき、その大部分が含まれるフレームをshort windowが選択されるフレームとする(図でS'と表記)。

4. MP3から継承された量子化スケール値  $Q[i]$  を一次関数により変換し、それを  $Q'[i]$  とする。そして、この  $Q'[i]$  を AAC における量子化スケール値とし、これに基づき符号量を決定する。その後、心理聴覚モデルに従い、global\_gain 値の修正を行う。
5. ハフマン符号を適用し、AAC データを出力する。
6. 仮にバッファ内のデータがなければ次の  $N$  フレームの MP3 データをバッファに供給する。そして、2 から 6 の手続きが適用される。MP3 データがなくなり次第、変換プロセスは終了となる。

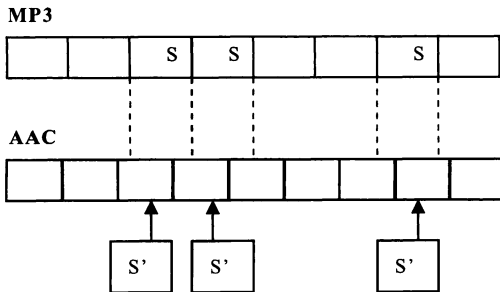


図3 MP3 と AAC の short block の対応

### 3. 性能評価実験

本節では、提案手法により MP3 から変換された AAC データ(a)の音質を、非圧縮ドメインにおけるトランスコーディング法により生成されたデータ(b)、および参考のために PCM から AAC 符号化されたデータ(c)と比較する。また、計算コストについてもあわせて評価を行う。図4は(a),(b),(c)の各データの生成手順を示している。

本実験は Intel Mobile Pentium 1.5GHz/512MB RAM, OS:WinXP PC 上で行った。300 秒のオーディオデータを2

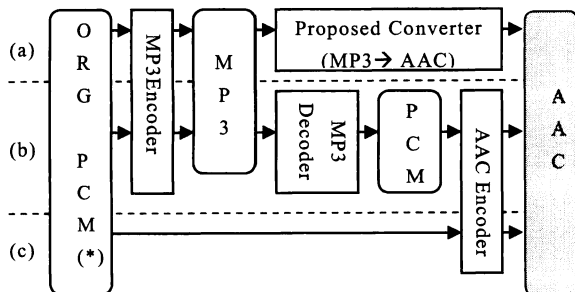


図4. 評価に利用したストリームの作成手順  
曲(ポップスとロック)準備しテストに用いた。MP3 のエンコード/デコードには LAME[10]を用い、AAC のエンコードは ISO のエンコーダを利用した。MP3 と AAC の

エンコードビットレートの組み合わせとして、それぞれ  
 <A> 128kbps, 96kbps  
 <B> 128kbps, 64kbps  
 <C> 96kbps, 64kbps  
 の3通りを用いた。これらはすべて 44.1kHz, モノラルで符号化したものである。

### 3.1. 音質評価

前節(a),(b),(c)各方法に対する音質を評価するために、10 人の評価者を用いて主観評価実験[12]を行った。はじめに、評価者にはオリジナルの PCM データを聴く(図4の(\*)印に相当)。その後、それぞれのデータの種別を被験者には知らせずに2つのデータ(オリジナル音/評価音)を聴き、オリジナル音と評価音の差を1から5の5段階で評価する。なお、評価値が高いほど評価音がオリジナル音に近いことを意味する。例えば、“4.00”は“差を知覚はできるが、気にはならない”を意味する。以上の方法により得られた結果を表1に示す。

表1. 主観評価結果

条件	(a)提案手法による変換	(b)再符号化による変換	(c)原音→AAC
<A>	4.10	3.75	4.25
<B>	3.80	3.55	4.00
<C>	3.75	3.25	

前述の通り、エンコードプロセスが2回適用されると、1回の場合と比較し音質は劣化する。したがって、この実験では、理論的には(c)の場合がベストケースとなる。その前提のもとで、提案手法(a)は(c)に非常に近い評価値が得られており、さらに、(b)と比較し高い値が得られていることがわかる。

なお、音質の違いを調査するために、テストセット<B>の場合の(a),(b),(c)の波形を図5に示す。同図において特に破線で囲んだ部分の(a)と(c)の違いは(b)と(c)の違いと比較し小さいことがわかる。実際に全曲に対する PSNR を測定すると(a)30dB, (b)21dB, (c)35dB となる。これらの値からも(a)の音質が(c)に近いことがわかる。

### 3.2. 変換速度評価

本節では(a)提案手法と(b)非圧縮ドメインでのトランスコード手法に対し、MP3 から AAC への変換速度に関する評価を行う。その結果を前述のテストセット<A>~<C>に適用した結果が表2である。同表より、提案手法(a)では従来手法(b)と比較し大幅に処理時間を削減できていることがわかる。

表 2. 変換速度

条件	(a)提案手法による変換	(b)再符号化による変換
<A>	42[sec]	74[sec]
<B>	40[sec]	72[sec]
<C>	40[sec]	70[sec]

#### 4. おわりに

本稿では、効率的なオーディオトランスコード手法に関する検討を行った。特に MP3 と AAC に着目し、それらのフレームサイズの違い、および量子化スケールの特徴の類似性に着目し、音質を劣化させることなく、両者を高速に変換する方法を提案した。また実験結果から、MP3 をデコードして AAC で再エンコードする方式と比較し、より高速に実現できることを示した。今後は MP3 から他のオーディオコーデックへの変換の検討を行う。

#### 文 献

- [1] ISO/IEC 11172-3:1993, "Coding of moving pictures and associated audio and digital storage media at up to about 1.5Mbit/s, Part3: Audio," 1993.
- [2] ISO/IEC 13818-3:1998, "Generic coding of moving pictures and associated audio, Part 3: Audio," 1998.
- [3] ISO/IEC 13818-7:2006, "Generic coding of moving pictures and associated audio, Part 7: Advanced Audio Coding," 2006.
- [4] ISO/IEC 14496-3:2005, "Information technology — Coding of audio-visual objects — Part 3: Audio," 2005.
- [5] S. Park, G. Hillman, R. Robles, "A novel structure for real-time digital sample-rate converters with finite precision error analysis," *Proc. IEEE ICASSP91*, Vol.5, pp. 3613 -3616, 1991.
- [6] A. T. Erdem, M. I. Sezan, "Multi-generation Characteristics of the MPEG Video Compression Standards," *Proc. IEEE ICIP*, Vol.II, pp.933-937, 1994.
- [7] J Zhang, A Perkis, N Georganas, "H. 264/AVC and Transcoding for Multimedia Adaptation," *Proc. the 6th COST*, 2004.
- [8] 中島康之, 米山暁夫, 菅野勝, 柳原広昌, "MPEG オーディオデータの符号化ビットレートスケラビリティ," *信学論 D-II*, Vol. 82, No.9, pp.1355-1364, 1999.
- [9] T. H. Tsai, C. Yen, "A high quality re-quantization / quantization method for MP3 and MPEG-4 AAC audio coding," *Proc. IEEE ISCAS*, Vol.3, pp. 851-854, 2002.
- [10] LAME Project, <http://lame.sourceforge.net/>
- [11] ISO/IEC 14496-5: 2001, "Information technology — Coding of audio-visual objects — Part 5: Reference software," 2001.
- [12] ITU-R BS.1116-1, "Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems," 1997.

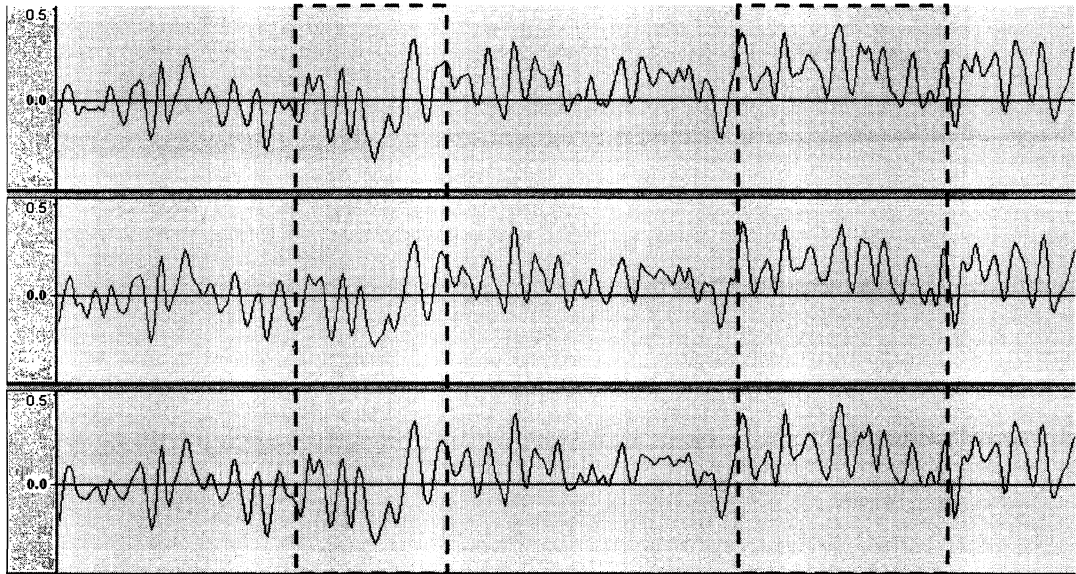


図 5 音声波形 (上から (a)提案手法による変換, (b)再符号化による変換, (c)原音→AAC)

- [4] ISO/IEC 14496-3:2005, "Information technology — Coding of audio-visual objects — Part 3: Audio," 2005.