新しい音声圧縮方式とその応用

及川 芳明[†], 正会員 渡辺 馨^{††}
 畠中 光行[†], 赤桐 健三[†]

MPEG-2 AAC (Advanced Audio Coding) は 1997 年 4 月に国際標準となった 1 チャンネル当たり 64 kbps で高音質に符号化を行うオーディオ符号化方式である。本稿では AAC の構成, 特徴的技術および音質評価テストとその結果について述べ, 最後に AAC の今後の展望について述べている。

1. ま え が き

1992 年に ISO/IEC JTC 1/SC 29/WG 11 MPEG (Moving Picture Experts Group) においてモノラルまたはステレオのオーディオ信号を高音質に符号化する方式として主に 128 kbps/ch で使用される MPEG-1¹⁾ が規格化された。その後マルチチャンネル信号を符号化できる方式や, 低いサンプリング周波数の信号を扱える方式の要請があり, MPEG-1 と後方互換を持つ MPEG-2 BC (Backward Compatibility)²⁾ が 1994 年に規格化された。しかし MPEG-2 BC は MPEG-1 との互換性を優先させたために, より低いビットレートで高音質を実現することができなかった。このような欠点を改善するため, MPEG-1 と互換性を持たないオーディオ符号化方式が規格化されることになった。

MPEG-2 AAC はこのような背景により規格化が行われた符号化方式で, 1997 年 4 月に ISO/IEC 13818-7³⁾ として国際標準となった 5 チャンネルの信号を 320 kbps (1 channel 当たり 64 kbps) で高音質に符号化するオーディオ符号化方式である。

本稿は AAC の構成, 特徴的技術および音質評価テストとその結果について述べ, 最後に AAC の今後の展望について述べる。

2. AAC の構成

AAC は 8~96 kHz までの 12 種類のサンプリング周波数に対応しており, チャンネル構成は標準でモノラルから 7 チャンネルまでに対応している。また, マルチチャンネルにおけるスピーカの位置と数を, 前方に何チャンネル, 横に何チャンネルおよび後方に何チャンネルと指定するこ

とも可能で, より柔軟なマルチチャンネルのチャンネル構成にも対応できるようになっている。

また, マルチチャンネルの信号を 2 チャンネルにダウンミックスするための係数を指定してデコーダ側に送ることができるようになっており, 2 チャンネルのみの再生装置での音質をコントロールすることが可能となっている。

さらに AAC は, デコーダの規模と音質のバランスをとるために以下の 3 つのプロファイルから構成される。

(1) メインプロファイル

ゲインコントロール(後述)以外のすべてのツールを使用し, 高音質を実現する。デコーダのプロセッサパワー, メモリ使用量が豊富なアプリケーションに主に使用される。メインプロファイルは LC プロファイルのビットストリームをデコードすることができる。

(2) LC (Low Complexity) プロファイル

メインプロファイルから予測器(後述)を除き, パラメータの制限を加えたもの。デコーダのプロセッサパワーおよびメモリの使用量がメインプロファイルに比較して小さいことが特徴である。

(3) SSR (Scaleable Sampling Rate) プロファイル

LC プロファイルにゲインコントロールを加えたもの。4 帯域の帯域分割フィルタを持つためデコードする周波数帯域を選択でき, デコーダのハードウェア規模を小さくすることができる。例えば, 最も低い周波数の帯域のみをデコードするデコーダは最も小さいデコーダ規模で実現できることになる。また 48 kHz のサンプリング周波数で 3 帯域のみをデコードするデコーダは, 18 kHz の再生帯域を確保しつつメモリー量を削減することが可能となる。

3. AAC の特徴的な技術

AAC は厳選された技術を実現したツール群⁴⁾を組合せることによって構成されている。図 1 にエンコーダの構成を示す。以下に各ツールの特徴的技術内容を示す。

3.1 ゲインコントロール

SSR プロファイルのみで使用されるツール。入力され

[†] ソニー株式会社 メディアプロセッシング研究所

^{††} NHK 放送技術研究所

“The Latest Standardized Audio Coding and Its Applications” by Yoshiaki Oikawa (Sony Corporation, Tokyo), Kaoru Watanabe (NHK Science and Technical Research Laboratories, Tokyo), Mitsuyuki Hatanaka and Kenzo Akagiri (Sony Corporation, Tokyo)

図1 エンコーダブロック図

た時間信号をPQF(Polyphase Quadrature Filter)帯域分割フィルタにより4分割し、最も低いバンド以外の信号をゲインコントロールする。ゲインコントロールは、例えば信号の振幅の小さい部分の振幅を増幅してエンコードし、デコードで増幅した部分の振幅を減衰させることによって、プリエコーと呼ばれる振幅の小さい部分の量子化雑音を抑えるものである。これによりカスケットなどの過渡的な信号の音質を向上させることができる。

また、デコードで4帯域のうちの何帯域をデコードするかを選択することにより、前述のスケラブルなデコードを実現することが可能となる。

3.2 フィルタバンク

入力時間信号はMDCT(Modified Discrete Cosine Transform)によりスペクトルデータ(MDCT係数)に変換される。変換は演算ブロック長を50%ずつオーバーラップして実行し、例えば2048サンプルを1024本のMDCT係数に変換する。

また、上述のプリエコーを抑えるためにMDCTの演算ブロック長を切り替えるブロックスイッチングという機構を用いている。フレームごとにロング/スタート/ショート/ストップのどれかのブロックとして切り替えることにより実現している。

メインおよびLCプロファイルでは定常的な信号の場合、MDCTの演算長を2048サンプルのロング、スタートまたはストップのブロックとし1024本のMDCT係数に変換する。一方、過渡的な信号の場合には256サンプルのショートのブロックとし128本のMDCT係数に変換する。ショートブロックでは8個連続で短い変換長を用いて

変換することにより、出力のMDCT係数の本数を1024本と、他のブロックと一致するようにしている。

SSRプロファイルの場合は、PQFで4分割されたものを各帯域ごとに変換するのでMDCT演算長は前述の値の1/4になる。しかし、4バンドの合計のMDCT係数は1024本で、他のプロファイルと同じになる。

また、MDCTの窓関数をフレームごとに、サイン関数による窓またはカイザーベッセル関数を元にした窓の2種類から選択して使用できるようになっている。

3.3 テンポラルノイズシェイピング(TNS)

通常、ノイズシェイピングは時間軸の信号をフィルタリング処理することにより周波数軸の雑音を制御するものであるが、TNSは周波数軸の信号であるMDCT係数を時間軸の信号とみなし、エンコードおよびデコードでLPCフィルタを通すことにより、時間軸上の振幅の大きいところに雑音を集中させ、男性の音声などの低いピッチ周波数を含む信号の音質を向上させる。フィルタリング処理する最高周波数とLPCフィルタの次数は、プロファイルにより異なる。

3.4 ステレオコーディング

2チャンネル以上のチャンネル構成の場合、チャンネル間の相関を利用して符号化効率を上げることが可能となる。AACでは以下の3つの方法を取り入れている。

(1) MS(Middle Side)ステレオ

左(L)チャンネルと右(R)チャンネルの信号を符号化するか、それぞれの和(L+R)と差(L-R)の信号を符号化するかをスケールファクタバンド(後述)ごとに選択して符号化する方法。左右のチャンネルの中央に定位する信号の符号化効率を上げることができる。

(2) インテンシティステレオ

高い周波数では左右の信号のパワーの差により音源の位置を感じるという特性を用い、左右の信号の和信号と左右のチャンネルのパワーの比を符号化する方法。

(3) カップリング

背景音を従来のマルチチャンネルの信号として符号化し、その背景音の中に定位させたい音をカップリングチャンネルとして符号化するボイスオーバーという方法を実現することができる。デコードで背景音に加え合わせるわけであるが、周波数軸で加え合わせるか、時間軸の信号で加え合わせるかで、2種類のカップリングチャンネルを持つことができる。ただし、SSRプロファイルはデコードの規模を小さくしたいためにカップリングチャンネルを持っていない。

3.5 予測器

メインプロファイルのショートブロック以外のフレームで使用される。48kHzサンプリングの場合、16kHzまでの各MDCT係数ごとに2次の格子型の予測器をそれぞれ持ち予測残差を符号化する。デコードでも同じ計算を行うことにより予測係数を送る必要がないようにしてある。

その代わりデコーダのプロセッサパワーおよびメモリー使用量を多く必要とする。この予測器により、ピッチパイプなどの定常的な信号の音質を向上させることができる。

3.6 聴覚モデル

MPEG-1 の Psychoacoustic モデル II の改良版が用いられている。どのブロックタイプにするかの決定結果と、各スケールファクタバンドごとの許容雑音レベルを出力する。

3.7 量子化器

人間の耳の特性であるクリティカルバンドをモデルに、1024 本の MDCT 係数を 49 個のスケールファクタバンドに分け、各スケールファクタバンドごとに正規化を行い、スケールファクタと正規化されたスペクトルデータを量子化する。ショートブロックの場合は 128 本の MDCT 係数を 14 個のスケールファクタバンドに分ける。

量子化雑音が聴覚モデルで求められた許容雑音レベル以内のレベルになるように、そのフレームで使用可能なビットを以下のような手続きで割り当てる。

(1) インナーイタレーションループ

量子化に必要なビット数が使用可能ビット数以内に収まるように量子化ステップ数を制御し、スペクトルデータの量子化を行うループ。スケールファクタおよびスペクトルデータの量子化には非線形量子化器が用いられる。

(2) アウターイタレーションループ

上記のインナーイタレーションループで求めた各スケールファクタバンドの量子化雑音が、聴覚モデルで求められた許容雑音レベルよりも大きくならないようにスケールファクタの値を制御するループ。

3.8 ノイズレスコーディング

以下の 4 種類の符号化方式が取り入れられている。

(1) パルスデータ

ショートブロック以外のフレームにおいて、4 本までの振幅の大きな MDCT 係数を小さい値に置き換え、デコーダへ位置情報と置き換えられた係数との差分情報を送る方法。これにより全体のダイナミックレンジを制限することができ、効率的にビット割り当てを行うことが可能となる。

(2) DPCM(Differential PCM)

スケールファクタの符号化に使用される。隣接するスケールファクタバンドのスケールファクタの差を非線形量子化する。先頭のスケールファクタはグローバルゲインとして DPCM のデータとは別に送られる。

(3) ハフマン符号化

DPCM して非線形量子化されたスケールファクタと正規化されて量子化されたスペクトルデータがハフマン符号化される。スペクトルデータは 2 本または 4 本のスペクトルデータごとにまとめられ、ハフマン符号化が行われる。コードブックは、スケールファクタ用に 1 種類、スペクトルデータ用に 11 種類を持つ。

(4) セクショニング

連続したスケールファクタバンドのスペクトルデータを同一のコードブックを使ってハフマン符号化できないかどうかを判定し、可能な場合どのスケールファクタバンドが同じコードブックを使用するかデータを符号化する。

3.9 マルチプレクサ

各ツールからの符号化データをまとめ 1 つのエンコードビットストリームを作成して出力する。ビットストリームの先頭にのみヘッダを持つ ADIF (Audio Data Interchange Format) と、各フレームにヘッダを持つ ADTS (Audio Data Transport Stream) の 2 種類のエンコードビットストリームを作成することができる。ADIF はコンピュータ上でのリアルタイムデコードなど、ビットストリームの編集を必要としないようなアプリケーションに使用される。

4. AACの主観評価とその結果

MPEG-2 AAC マルチチャンネル音声符号化の主観評価を、1996 年 9 月～10 月、NHK と BBC で行った。評価に使用したマルチチャンネル音声方式を図 2 に示す。この方式は、基本の左右 2 チャンネル L, R に前方の安定した音像定位を与えるセンタチャンネル C と、臨場感をさらに高めるための 2 つの後方チャンネル(左サラウンド, 右サラウンド)を付加した、5 チャンネル音声である。また、評価は映像なしで行った⁵⁾⁶⁾。

4.1 テスト符号化方式と音源

評価した符号化方式は次の 4 種類である。

MPEG-2 AAC メイン 320 kbps

MPEG-2 AAC メイン 256 kbps

MPEG-2 AAC LC 320 kbps

MPEG-2 BC Layer II 640 kbps

図 2 マルチチャンネル音声方式

表 1 テスト音源

番号	名前	内容
1	pitch pipe	ピッチパイプ
2	harpsichord	ハープシコード
3	triangle	トライアングル
4	cast pan 1	前方カステネットパン, 後方ノイズ
5	elliott 1	レストランの男女の会話, 音楽
6	mancini	オーケストラ・ストリング, シンバル, ドラム, ホルン
7	station master 1	蒸気機関車の男性の声
8	clarinet theatre	前方中央クラリネット, 劇場の歓声, 後方に雨音
9	thalheim 1	前方左ピアノ, 前方右サックス, 中央に女性の声
10	glock	グロッケンスピールとティンパニ

また、用いた音源は表 1 に示す 10 種類である。これらは選択パネルにより、劣化が目立ちやすいとして選定された音源である。

4.2 テスト方法

主観評価テストは、ITU-R 勧告 BS. 1116⁷⁾ に則って行った。評定者は、「基準試聴位置」で「基本音声品質」を評価した。

主観評価テストは、「隠れ基準付き三刺激二重盲検法」を用いた。この評価法を図 3 に示す。提示音は「基準音」、「評価音 A」、「評価音 B」である。評定者はこれらを自由に切り替えることができる。評価音 A と評価音 B の一方は符号化音、もう一方は隠れ基準で、ランダムに割り当てられている。音源は、評定者の希望により、何度でも繰り返し試聴することができる。評定者は各試行毎に、評価音 A と評価音 B の両方の基本音声品質を評価する。基準音との相違は、すべて劣化とみなす。各評定者は、図 4 に示す 5 段階劣化尺度で、小数点 1 桁まで評価を行った。また、評価音 A または評価音 B の 1 つは隠れ基準なので、少なくとも 1 つに評点 5.0 を与えるようにした。

4.3 テスト環境と評定者

NHK では、「放送技術研究所・試聴室 B 268」を用いて主観評価テストを行った。試聴室、スピーカの配置は図 5 の通りである。

NHK で 24 人、BBC で 32 人が評価テストに参加した。評定者は音響の専門家である。また、符号化の開発者は評定者から除外した。NHK の評価テストに参加した評定者の内訳は、5 名はミクサなどの番組制作者、残りの 19 名は音響技術者である。

4.4 評価結果

評価結果の分析は「diffgrade」を用いて行った。diffgrade は各試行毎に、符号化音の評点と隠れ基準音の評点の引き算から計算する。例えば、符号化音の評点 4.0 は -1.0 の diffgrade になる。

評定者の検定は 95% の片側 t 検定により行った。検定の結果、NHK の評定者 24 人中 16 人が、BBC の評定者

図 3 隠れ基準付き三刺激二重盲検法

図 4 ITU-R S 段階劣化 R 度

図 5 テストに用いた試聴室

32 人中の 23 人が信頼できると判断された。以降の分析は、信頼できると判断された評定者のみを使用した。また検定の結果、NHK と BBC の評価データは結合できないことがわかり、別々に分析した。

各符号化方式の評価結果を図 6～図 9 に示す。評価結果は、音源毎の平均評点と 95% 信頼区間を示している。この結果、すべての符号化は、ほぼ良好な音質であることがわかった。

また、評価結果を検討したところ、NHK と BBC の主観評価は同様の結果を示した。

320 kbps AAC メインプロファイルのマルチチャンネル符号化は、次の音質であることがわかった。

- ・640 kbps BC Layer II の品質よりも良い(危険率5%で統計的に有意な音源がある)。
- ・すべての音源に対して、ITU-R の放送品質(主観評価4.0 以上)を満足する。

4.5 SSR プロファイルの評価

上記の公式主観評価と別に、ソニーにより AAC SSR プロファイルの主観評価が行われた。この結果、SSR 320 kbps の音質は AAC LC 320 kbps とほとんど差は見られないことが報告されている⁸⁾。

5. AAC の展望

まえがきで述べたように MPEG-2 AAC は、MPEG-2 BC が達成できなかった ITU-R の放送品質(主観評価4.0 以上)を 5.1 マルチチャンネルにおいて 320 kbps から 384 kbps の低ビットレートで達成することを第1の目標として標準化が行われた。これが達成された結果、

- (1) マルチチャンネル
- (2) 放送音質確保
- (3) 一層の低ビットレート化

へと進むものと思われる今後のデジタル放送の標準化においては、MPEG-2 AAC 規格がまず第1の候補として検討されることになると期待される。現在最も MPEG-2 AAC のデジタル放送への適用について検討が進んでいるのは日本のデジタル放送規格策定であり、2000 年実用化計画が進められている BS デジタル放送に使用する方向で 1997 年 11 月末現在、規格化のまとめ作業が行われている。

BS デジタル放送ではビデオ、オーディオ、EPG (Electronic Programming Guide) 等に割り当てできるビット量の限界から、オーディオにはせいぜい 500 kbps 以下のビットレートしか期待できず、低ビットレートで高音質を達成できる MPEG-2 AAC の採用は、バランスのとれた放送方式を得るためには重要なキーとなると考えている。本誌が発刊されているであろう時点では、すでにオーディオを含む BS デジタル放送規格が正式決定しているものと思われる。日本のデジタル放送規格は、

- (1) 決定最終段階にある BS デジタル放送規格に加えて
- (2) 地上波デジタルテレビ放送規格
- (3) (2)と関連した伝送方式でのデジタル音声放送規格策定

が進行するスケジュールとなっている。これらの規格化においても MPEG-2 AAC が第一に検討されるものと思われる。特に現在電波利用の効率化が大変大きなテーマとなっていることから、MPEG-2 BC の半分のビットレートで同等以上の音質が得られる MPEG-2 AAC のメリットは大きい。例えば(3)としてあげたデジタル音声放送規格において、ビットレートが半分ですむという MPEG-2 AAC の存在は、伝送方式選定の過程にも影響を与える大

図6 AAC メイン 320 kbps 評価結果

図7 BC 640 kbps 評価結果

図8 AAC メイン 256 kbps 評価結果

図9 AAC LC 320 kbps 評価結果

きな意味を持つと考えられる。

日本のデジタル放送規格にはすでにCSおよびCATV規格が策定済みであり、これらにはMPEG-2 BC（実際の放送はマルチチャンネルではなくステレオで行われている）が採用されているが、これに加えMPEG-2 AACが加わることで、放送業者がそれぞれの特徴を生かし選択できるようになることも考えられる。

MPEG-2 AACは、現在規格策定中のインターネットをターゲットに含むMPEG-4においても64 kbps/chコーデックのためのツールとして規定されており、また40~16 kbps/chの領域でも候補として議論されているが、現在規格化の途上であり、もう少しの時間が必要である。すでにインターネットでは数多くのオーディオ符号化方式が実用化されており、今後もその趨勢は続くものと思われる。このような状況の中でMPEG-4は、最終的には1999年2月にInternational Standardとなる予定である。その時点でMPEG-4オーディオ方式がインターネットの世界でどのようなインパクトを与えるか、注目される場所である。

MPEG方式はMPEG-2 AACに限らず、主にすでにエンコードしたオーディオ情報をディストリビューットすることに主眼を置いているために、エンコードの複雑さはあまり気かけられていない。このためエンコードを含む機能を持たせなければいけない用途や、送られてきたオーディオ情報をディスクに記録するなどの用途のためには、エンコードの複雑さを考慮した方式との間での使い分けが行われることが考えられる。

以上のように、MPEG-2 AAC方式は日本において初めて取り入れられる最新の符号化方式であり、日本のデジタル放送が、技術的にも番組の内容の点からも、世界的に最高の放送規格となることを希望している。

〔文 献〕

- 1) ISO/IEC JTC1/SC29/WG11 IS 11172-3 MPEG-1 Audio 規格
- 2) ISO/IEC JTC1/SC29/WG11 IS 13818-3 MPEG-2 BC Audio 規格
- 3) ISO/IEC JTC1/SC29/WG11 IS 13818-7 MPEG-2 AAC Audio 規格
- 4) M. Bosi, et al: "ISO/IEC MPEG-2 Advanced Audio Coding", AES Conventions (1996)
- 5) Kirby, D.G. & Watanabe, K.: "Formal Subjective Testing of MPEG-2 NBC Multichannel Audio Coding Algorithm", AES 102nd Convention, preprint 4418 (Mar. 1997)
- 6) Kirby, D.G. & Watanabe, K.: "Report on the Formal Subjective Listening Tests of MPEG-2 NBC Multichannel Audio Coding", ISO/IEC JTC1/SC29/WG11/N1419 (Nov. 1996)
- 7) ITU-R Recommendation BS. 1116, "Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems", Geneva (1994)
- 8) Akagiri, K & Oikawa, Y.: "Listening Test Report of the MPEG-2 AAC SSR Profile Multichannel Audio Coding", ISO/IEC JTC1/SC29/WG11/N1704 (Apr. 1997)

おいかわ よしあき
及川 芳明 1986年、長岡技術科学大学大学院工学研究科修士課程電子機器工学専攻修了。同年、ソニー(株)に入社。現在、メディアプロセッシング研究所において、オーディオ符号化方式の研究開発に従事。

わたなべ かおる
渡辺 馨 1981年、東北大学大学院情報工学専攻修士課程修了。同年、NHKに入局。福岡放送局を経て、1984年より、放送技術研究所に勤務。現在、同所マルチメディアサービスに所属。主に、音響信号の高効率符号化、デジタル音声放送システムの研究に従事。正会員。

はたなか みつゆき
中 光行 1994年、日本大学大学院理工学研究科電気工学専攻修了。同年、ソニー(株)に入社。現在、メディアプロセッシング研究所においてオーディオ符号化の研究・開発に従事。

あかぎり けんぞう
赤桐 健三 1969年、国立明石高等工業専門学校電気工学科卒業。ベータ HiFi, 8mm ビデオのオーディオ規格、CD-ROM XA オーディオ規格、ミニディスクオーディオ規格の開発に従事。現在、オーディオおよびビデオ信号処理技術の開発に従事。ソニー(株)メディアプロセッシング研究所 A&V プロセッシングラボラトリー統括部長。