

MPEG-2 AAC方式の運用に関する技術解説

AAC調整連絡会

1. まえがき

2004年1月頃より地上デジタルテレビジョン放送の5.1chサラウンド番組の視聴者から、音声再生不具合のクレームが届くようになった。放送事業者で調査した結果、送出エンコーダのAACパラメータのひとつであるTNS符号化ツールがONの際に、地上デジタル放送開始以前に設計されたAVアンプを地上デジタル放送用受信機に接続して5.1chサラウンド番組を視聴すると、音声の再生不具合が生ずる確率が高いことが判明した。

2004年3月、社団法人電子情報技術産業協会（以下、JEITAと記す）より放送事業者に対して、市場の混乱を回避する観点からTNS符号化ツールの運用制限の要請があり、放送事業者の自主的な対応として、2004年12月からはAACエンコーダの一部機能について「運用自粛」を行なったので、その後はAAC音声に関する特段の不具合は報告されていない。

2005年2月、JEITAと放送事業者は「視聴者保護」の観点から、「AAC調整連絡会」を設置し、AVアンプの現状調査や早期に包括的な「暫定対策」を行うための方策の検討、更に将来にわたり規格上明確化が必要と思われる事項の検討を開始した。

AAC調整連絡会では、これらの事項ならびにMPEG-2 AAC方式の実装に係わる留意点を検討し、結果をARIB STD-B32第2部およびARIB TR-B14第二編に反映するとともに、BPA¹、DRPほか関連の団体との情報共有を進めた。さらに符号化ツールおよびパラメータに関する機能確認用テストストリーム（以下、AACテストストリームと記す）を作成し、デジタル放送推進協会（Dpa）を通じて提供している。

JEITAでは、「運用自粛」解除によって音声再生に不具合の恐れがある2004年12月以前に製造されたAVアンプが、これまでにJEITA会員メーカー製品で78万台、会員外メーカー製品を合わせると約100万台が出荷されたと推定しており、有効な製品買い替え促進策が採られず自然減に任せた場合、これらAVアンプの残存台数が、JEITA会員メーカー機器について2012年で23万台、2016年で2.5万台あるものと予測した。

この予測をもとに、AAC調整連絡会では「運用自粛」の解除時期について検討し、各放送事業者は2016年12月31日を経過後、AACエンコーダの一部機能についての「運用自粛」を解除できるようにすることが適当であるとした。

本技術解説では、MPEG-2 AAC音声符号化方式の運用指針について、AACテストストリームをもとに検討した技術情報を提供する。また、今後のTNS運用を見据えたパラメータに関する検討についても述べている。

2. MPEG-2 AAC方式と符号化ツール

MPEG-2 AAC音声符号化方式のエンコード、ならびにデコード処理の概要を参考1に示す。MPEG-2 AAC方式には音質改善のためのオプションの符号化ツールとして、MSステレオ、イ

¹ 当時は、BSデジタル放送推進協会、Dpaに統合

ンテンシステレオ、TNS（TNS：Temporal Noise Shaping）等がある。エンコーダ側で、これらオプションの符号化ツールをONにした場合、デコーダではそれぞれに対応した演算処理が必要となる。

各オプションツールに対するデコーダの演算処理量は参考2に示すように、TNSをONとした際には、TNS OFF時と比べると著しく増加することが知られている。

すなわち、TNSは他のオプション符号化ツールに比べて多くのインストラクション数を要し、例えばM/Sステレオの約20倍のインストラクション数が必要であり、これはMPEG-2 AAC方式のデコード処理に必須なIMDCTの約4割に匹敵する処理量でもある。また実際のデコーダLSI実装形態によっては、さらに処理量が大きくなる可能性もある。

デコーダでの演算処理量が、その処理能力を超えると音声再生に不具合が発生するが、演算処理量はエンコーダに入力される音声信号の内容によって異なる。

そこで、AAC調整連絡会の技術検討WGを中心に、5種類のAACテストストリームについて、TNSの運用とビットレートの関係性を考察した。

3. AACテストストリームの目的

各テストストリームは、MPEG-2 AACデコーダ処理に関する以下の項目をチェックする目的で作成した。

① TNS運用による処理量の増加をチェックするもの（ストリームNo.1およびNo.4）

TNSは線形予測であり、予測成分を大きなフィルタ次数で表現しようとする、処理量がフィルタ次数に比例して増加することが知られている。フィルタ次数はLCプロファイルの場合最大12であり、5チャンネルの場合各チャンネルのフィルタ次数の合計（総次数）は最大60となる。

② 高ビットレート時におけるハフマンデコード処理量の増大をチェックするもの（ストリームNo.3およびNo.5）

ハフマンデコード処理は他の数値演算処理とは異なり、主として条件分岐処理を行う。AAC音声ビットストリームの各フレームのビット数をビットレートに換算した値（以下「瞬時ビットレート」という）と、ハフマンデコード処理量は基本的に比例する関係があるが、ESCコード等の特異値処理が存在することと、その組合せが多岐に渡るため、最大処理量の見積りが難しい。

③ TNS運用時のスケールレベル設定に起因する演算精度の不足をチェックするもの（ストリームNo.2）

デコーダ内のTNS逆フィルタのフィルタ利得によって、演算処理の小数点位置が変動するため、固定小数点型DSPでは変動量の最大値を予め規定する必要がある。なお、ストリームNo.2で計測した最大値は3ビットである。

4. TNS符号化ツールとテストストリームの瞬時ビットレート

MPEG-2 AACのデコード処理量に大きく影響するパラメータには、TNSの有無およびフィルタ次数、ならびに瞬時ビットレートがあり、これら2つのパラメータは、各々デコード処理量に対して比例関係にあることが知られている^(注1)。そこで、AACテストストリームのうち、TNS運用による処理量増加チェック用のテストストリームと、高ビットレート時におけるハ

フマンデコード処理量増加チェック用のテストストリーム（No.1、No.3、No.4）を用いて、TNSフィルタの総回数と瞬間ビットレートの2種類のパラメータを軸とした二次元分析を行った。図1は、これらのストリームの全てのAAC音声フレームでのTNSフィルタ総回数と瞬間ビットレートとの関係をプロットしたものである。

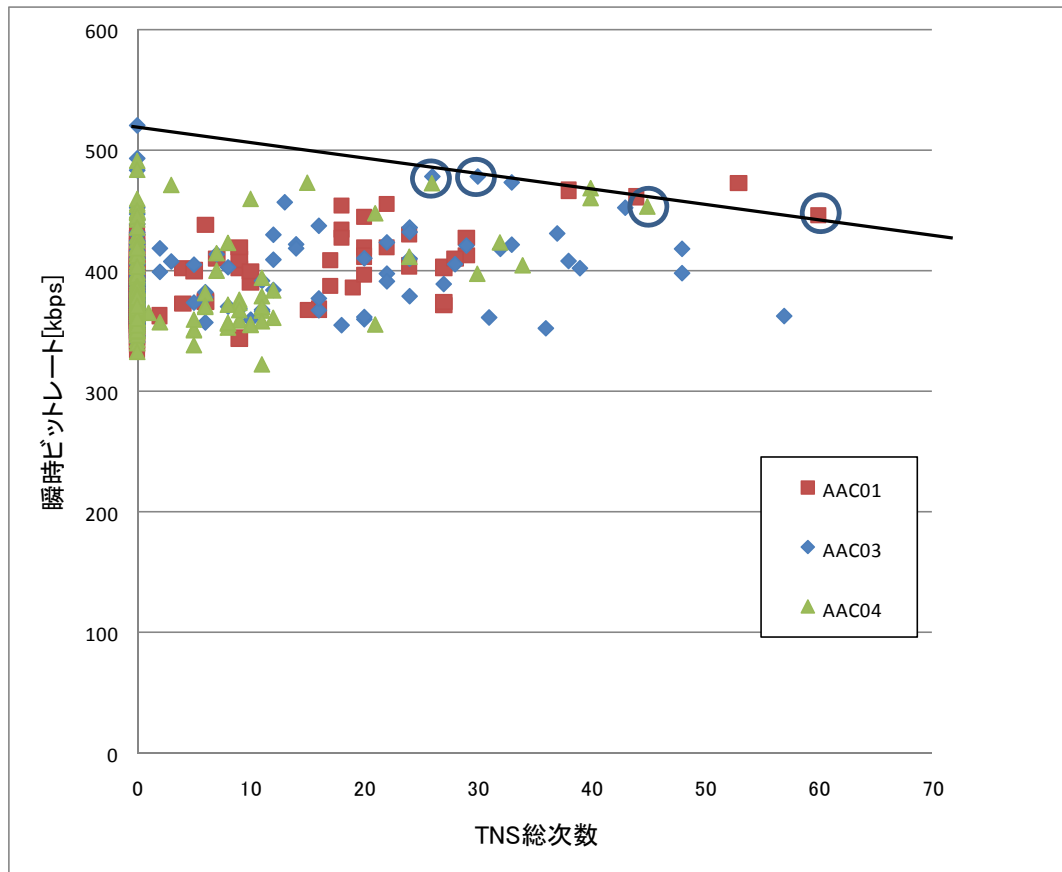


図1 AACテストストリーム（3種類）の全AAC音声フレームにおけるTNSフィルタ総回数と瞬間ビットレートの関係

図1の中に記した4か所の丸印は、TNS=ONの状態と過去に販売された一部の音声機器で音声再生の不具合が指摘されたAAC音声フレームに相当するもので、それぞれ以下の表に示すストリーム番号とフレーム番号のポイントである。

なお、図中に記載されていないが、これら3種類のAACテストストリームの平均音声ビットレートはいずれも384kbpsである。

ストリーム番号	フレーム番号	瞬間ビットレート	TNS総回数
1	20	445.50kbps	60
3	81	477.75kbps	26
3	152	477.75kbps	30
4	31	453.80kbps	45

(注1) 瞬間ビットレートとデコード処理量は統計的にはほぼ線形の関係にあるが、正確にはデコーダの実装方法に依存する部分は残る。特にフマンデコードの方法によっては瞬間ビットレートとデコード処理量の関係に特異点を持つ場合がある。

AACデコーダの処理量は、機器の実装に依存するものの、「TNSフィルタ総次数」「瞬間ビットレート」と、各々、独立に比例関係にある。従って、AACデコーダの処理量は両パラメータが共に高い場合（図1では右上方向）に大きくなる。図1の過去の不具合ポイントもTNSフィルタ総次数が高い場合は低い場合に比べて、より低い瞬間ビットレートで発生することが確認された。

瞬間ビットレートについては、現状運用されているAACエンコーダを調査した結果、①現状のAACエンコーダは必ずしも瞬間ビットレートの上限を制御していない、②現実論としては、瞬間ビットレートの上限は平均ビットレートの最大で2倍のビットレート（例えば平均ビットレートが384kbpsの場合、768kbps）を超える可能性はない、との情報が報告されている。

5. MPEG-2 AAC運用指針

ARIB STD-B32第2部およびARIB TR-B14第二編には、主としてMPEG-2 AAC方式の受信機やAACデコーダの実装に係わる留意点が記載されているが、送出側の運用規定については運用の目安として最大平均ビットレート（2chで256kbps、5.1chで384kbps）が記載されているだけで、運用指針が明示されていないため、送出側で安全な運用を行うための技術情報の提供が望まれてきた。

現在製造販売されている受信機やAVアンプは、前項に述べたAACテストストリームを処理しても不具合は生じないと考えられるが、過去に製造された受信機やAVアンプに不具合を起こさないよう留意が必要である。このためMPEG-2 AACの送出運用においては、基本的に図1に過去の不具合ポイントを直線で結び、この直線よりも低い領域でエンコーダの運用ができればよい。具体的には以下の運用指針に則って運用すればよいことになる。

(1)TNSの運用について

(運用指針例1)

TNSを運用する場合、図1において、丸印（過去の不具合ポイント）を結んだ直線よりも低い領域でのTNSフィルタ総次数および瞬間ビットレート、具体的には

$$\text{瞬間ビットレート} < 520 - 1.08N \quad (N : \text{TNSフィルタ総次数})$$

で運用する。

(運用指針例2)

TNSを運用する場合、最大瞬間ビットレート=445kbps以下で運用する。
(あるいはTNSを運用する場合、最大瞬間ビットレートを平均ビットレート（最大で384kbps）の1.2倍程度以下で運用する。)

上記の運用指針例は、個々のAACエンコーダの仕様に応じて、そのいずれかが選択されることを想定している。

(2)TNSを運用する際のスケールレベルについて

(運用指針例)

TNSを運用する場合、固定小数点演算デコーダの演算精度に配慮し、TNSフィルタの利得が大きくなり過ぎないように（変動量の最大値が3ビットを超えないよう）に運用する。

実際のエンコーダの運用にあたっては、当該AAC音声フレームのTNSをOFFにする、あるいはフィルタ係数を修正する等の方法が考えられる。ただし、個別のAACエンコーダで可能かどうかの検討は必要である。

以上に示した運用指針は現時点での「運用指針例」であり、将来の議論によっては、さらに修正の余地がある。なお、上記項目内の「受信機／AVアンプ」の定義は以下のとおりである。

1. 受信機の定義

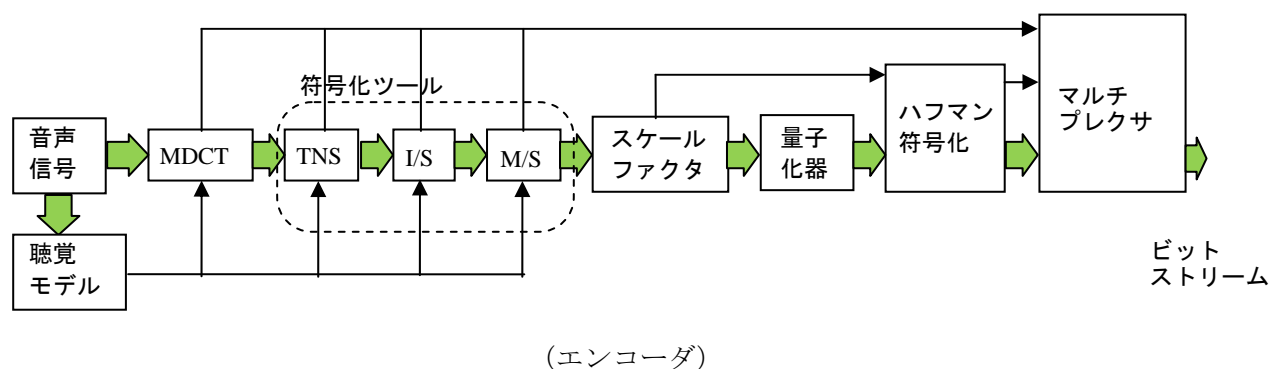
- ① ARIB STD-B1 に規定されている狭帯域CSデジタル放送用および高度狭帯域CSデジタル放送用受信装置
- ② ARIB STD-B21 に規定されているBSデジタル放送用および広帯域CSデジタル放送用および地上デジタル放送用受信装置（本規格に準拠する小型簡易受信装置および車載受信装置および携帯受信装置を含む）
- ③ JCTEA STD-007 に規定されているデジタル有線テレビジョン放送デジタルケーブルテレビジョン受信装置

2. AVアンプの定義

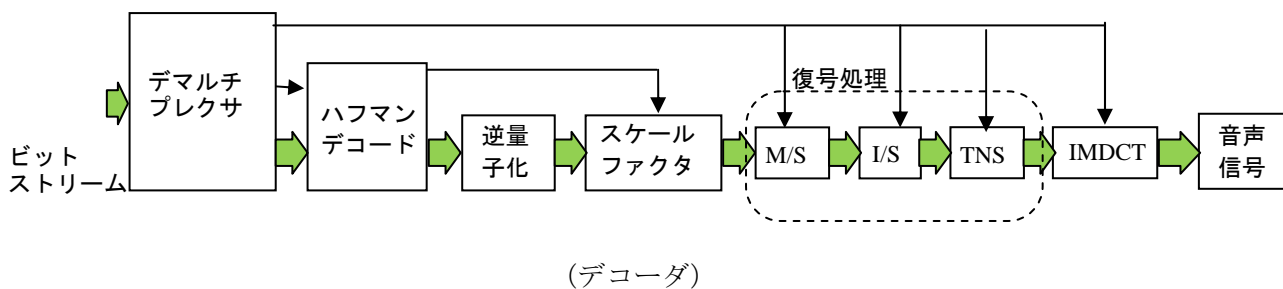
受信機から、デジタル入力によりAACをデコードするオーディオ機器

(参考 1) MPEG-2 AAC のアルゴリズムおよび符号化ツールの概要

AACのLCプロファイルでのアルゴリズムを図2に示す。時間域の音声信号は窓長2048もしくは256のMDCT（修正離散コサイン変換）を用いて、それぞれ1024点（long block）,128点（short block）の周波数域信号に変換される。この際、変換長は音声信号の聴覚特性に応じて切り替えられる。アタック音など急峻な変化のある信号の場合にはshort blockが用いられてプリエコーを抑制する。



(エンコーダ)



太い矢印はデータ、細い矢印は制御を表す

図2 MPEG-2 AAC LCプロファイルのアルゴリズム

符号化ツール群としてはTNS、I/S（インテンシティステレオ）およびM/SステレオがAACの圧縮符号化の性能を上げるために用意されている。これらは聴覚や信号の特殊な性質を利用してMDCT係数のビット数の削減を図ろうとするものである。

以下に3つの符号化ツール群、量子化、およびハフマン符号化について概要を述べる

① 符号化ツール群

(ア) TNS (Temporal Noise Shaping)

TNSは周波数軸のMDCT係数を時間軸の信号と見なし、エンコーダ、デコーダにおいてそれぞれMDCT係数に対して線形予測係数を用いたトランスバーサルフィルタリングおよび巡回型フィルタリング処理を行うことで、波形に含まれる量子化雑音を信号レベルの大きなところに集中させ、男性のスピーチなど低いピッチ周波数を含む信号の音質を向上させる。LCプロファイルの場合、フィルタの最大次数は12となる。なお、TNSは予想利得が閾値を超えたときだけ実行される。

(イ) インテンシティステレオ

高い周波数成分を聴くときに、左右の到来時間遅れよりも音の大きさの影響を受けやすい性質を利用して、複数のチャンネルの高い周波数の量子化係数をひとつの情報にまとめて送ることによりビットレートを削減する。

(ウ) M/Sステレオ

両チャンネルをそれぞれ符号化するか、和信号と差信号を代わりに符号化するかをスケールファクタバンド（近い周波数のMDCT係数を纏めたグループ）ごとに選択する方式。左右同位相、同振幅、例えば音楽のボーカルなどで符号化効率を高めることができる。

② 量子化

人間の聴覚特性である臨界帯域を模擬した49のスケールファクタバンドに1024のMDCT係数を分配する。スケールファクタバンドごとに正規化し、スケールファクタと正規化されたMDCT係数に対して量子化を行う。

③ ハフマン符号化

ハフマン符号化は可逆圧縮の代表的なアルゴリズムで、一定ビットごとに文字列を区切り、出現確率の高いパターンに対してより短い符号を与える。ハフマン符号の対応表はスペクトルデータ用に11種類用意されており、サブバンドごとに選択される。

参考文献

- (1) ISO/IEC 13818-7:2006 “Information technology – Generic coding of moving pictures and associated audio information – Part 7: Advanced Audio Coding (AAC)”
- (2) M. Bosi, et. al, “ISO/IEC MPEG-2 Advanced Audio Coding”, AES Convention (1996)
- (3) 及川他、「新しい音声圧縮方式とその応用」、映像情報メディア学会誌Vol.52、No.2（1998）

(参考 2) MPEG-2 AAC 方式のデコード処理量について

MPEG-2 AACツールの処理量に関する報告⁽¹⁾ から、デコード処理における、1フレーム(48kHz サンプルングの場合約23ミリ秒)あたりの演算回数 (インストラクション数) を抜粋して表1に示す。

TNS運用時のMPEG-2 AACデコード処理量は、TNS OFF時と比べると著しく増加することが知られている。TNSは他のオプションツールに比べて多くのインストラクション数を要し、M/Sステレオの約20倍 (=8130/854*2) のインストラクション数を要することがわかる。

さらに、実際のデコーダLSI実装形態によってはさらに処理量が大きくなる可能性があることに留意が必要である。

表1 各ツールにおけるAACデコーダの数値演算回数 (LCプロファイル)

ツール	必須/ オプション	インストラクション数 (注記なきものは1チャンネル当たり)
ハフマンデコード	必須	13657
逆量子化	必須	1708
M/Sステレオ	オプション	854 (1ステレオペア当たり)
インテンシティ ステレオ	オプション	0~854 (1ステレオペア当たり)
TNS ^(注)	オプション	8130 (フィルタ次数12)
IMDCT	必須	19968

(注) ARIB STD-B32 第2部 参考資料4「MPEG2 AAC方式の実装に係わる留意点」では、TNSの復号処理数に直接関連する事項として(1)フィルタ数、(2)フィルタ次数、(3)フィルタの長さ、(4)チャンネル数が記載されている。なお、TNSの復号処理数は、各チャンネル((4)チャンネル数)内に複数個((1)フィルタ数)存在するフィルタの処理数である((2)フィルタ次数)と((3)フィルタの長さ)の積の総和に依存する。

参考文献

- (1) S.R. Quackenbush, et. al. "Revised Report on Complexity of MPEG-2 AAC Tools", ISO/IEC JTC1/SC29/WG11 N2957 (Oct 1999)

※http://www.chiariglione.org/mpeg/working_documents.htm にて入手可能

(参考3) AAC テストストリームの目的および適用範囲 (リリースノートより抜粋)

(1) 目的

- ① AAC 音声のデコーダ実装に関して、ARIB STD-B32 第2部 (1.9 版) および TR-B14 第二編 (2.8 版) にて注意喚起されているツールおよびパラメータに関する機能確認用ストリームを提供することを目的とする。
- ② 具体的には、AAC 音声に関する以下の注意喚起事項に対応したストリームを提供する。
 1. TNS 使用時における復号側の処理量の増大
TNSは、その復号処理に大きな処理数が必要となる可能性のあるツールである。その原因となるTNS復号フィルタ処理数に直接関連する事項として、(1)フィルタ数、(2)フィルタの次数、(3)フィルタの長さ、(4)チャンネル数がある。
 2. TNS 使用時における復号側のスケールレベル設定に起因する演算精度の不足
例えば固定小数点演算を行うデコーダDSPの場合、フィルタによるオーバーフロー、アンダーフローを避けるためにscaling levelを大きく取ると、所要の演算精度を確保できない (TNSによる音質改善を効果的に働かせることができない) 場合がある。
 3. 高ビットレート時における復号側のハフマンデコード処理量の増大
ハフマンデコード処理量はビットレート (最大瞬時レート) が高くなるにつれて増大する傾向がある。
- ③ 本ストリームでは、ビットストリームが正確にデコードされる前提で、聴感上の異音が生じないことを確認する。
- ④ 本ストリームは、上記注意喚起事項に対する上限を規定するものではない。

(2) 適用範囲

- ① 本ストリームは、地上デジタルテレビジョン放送受信機の機能確認に適用する。
なお、地上デジタルテレビジョン放送を受信するケーブル STB、また地上デジタルテレビジョン放送受信機のデジタル音声出力を通じて AAC 音声を再生する録画機器および AV 機器も対象とする。
- ② 本ストリームの目的に鑑み、ARIB STD-B32 第2部および TR-B14 で規定される AAC 音声符号化を利用する地上デジタルテレビジョン放送以外のデジタル放送メディアにおいても、十分な考慮が望まれる。
- ③ AAC 音声の注意喚起に関する TR-B14 2.8 版改定 (2006 年 5 月) 後 18 ヶ月以降の新規受信機等に適用する。ただし、それ以前に発売の機器においても十分な考慮が望まれる。