2008 年 6 月 16 日

音声符号化方式作業班

高度衛星デジタル放送　音声符号化方式
報告書案（技術的条件）

1. 暫定方式の概要

項目		内容
音声入力信号		
	最大入力音声チャンネル数	22.2 チャンネル（1 ストリームあたり）
	入力サンプリング周波数	48kHz
	量子化ビット数	16、20、24 ビット
	音声モード	モノラル、ステレオ、 マルチチャンネルステレオ （0-3/0/0-0.0 (3.0), 0-2/0/1-0.0 (3.0), 0-3/0/1-0.0 (4.0), 0-2/0/2-0.0 (4.0), 0-3/0/2-0.0 (5.0), 0-3/0/2-0.1 (5.1), 0-3/0/3-0.1 (6.1), 2/0/0-2/0/2-0.1 (6.1), 0-5/0/2-0.1 (7.1), 0-3/2/2-0.1 (7.1), 2/0/0-3/0/2-0.1 (7.1), 0/2/0-3/0/2-0.1 (7.1), 2/0/0-3/2/3-0.2 (10.2), 3/3/3-5/2/3-3/0/0.2 (22.2)）(注)、 2 音声（デュアルモノラル） ※下線を引いたモードが推奨音声モード （注）マルチチャンネルステレオの音声モード表記法として、チャンネル数を次のように表す。 上層—中層—下層.LFE 割り当てチャンネルがない層は0 とし、割り当てチャンネルがある層のチャンネル数を次のように表す。 フロント／サイド／バック なお、括弧内は従来の表記。
音声符号化方式（基本サービス用）		
	符号化方式	MPEG-2 AAC LC プロファイル ※AAC+SBR(Spectral Band Replication)の使用も可能
	符号化パラメータ	既存方式に順ずる ※ただし、音声入力信号の拡張に伴うパラメータの拡張を実施
音声符号化方式（非圧縮・ロスレス高音質サービス用）		
	符号化方式（非圧縮符号化）	リニア PCM 音声符号化
	符号化方式（ロスレス符号化）	MPEG-4 ALS（放送応用に適したサブセット化検討中）
	最大音声入力チャンネル数	5.1 チャンネル

2. 音声入力信号

(1) 最大入力音声チャンネル数

・ 最大入力音声チャンネル数は 22 チャンネル及び低域を強調する 2 チャンネル（以下、22.2 チャンネル）とする。

(2) 入力サンプリング周波数

・ 48kHz とする。

(3) 入力量子化ビット数

・ 16、20、24 ビットとする。

(4) 音声モード

・音声モードは、以下の通りとする。

＊<u>モノラル</u>
＊<u>ステレオ</u>
＊マルチチャンネルステレオ(注)

0-3/0/0-0.0 (3.0), 0-2/0/1-0.0 (3.0),
0-3/0/1-0.0 (4.0), 0-2/0/2-0.0 (4.0),
<u>0-3/0/2-0.0 (5.0)</u>, <u>0-3/0/2-0.1 (5.1)</u>,
<u>0-3/0/3-0.1 (6.1)</u>, 2/0/0-2/0/2-0.1 (6.1),
<u>0-5/0/2-0.1 (7.1)</u>, <u>0-3/2/2-0.1 (7.1)</u>,
<u>2/0/0-3/0/2-0.1 (7.1)</u>, 0/2/0-3/0/2-0.1 (7.1),
2/0/0-3/2/3-0.2 (10.2), <u>3/3/3-5/2/3-3/0/0.2 (22.2)</u>

＊2 音声（デュアルモノラル）

※下線は推奨音声モードを示す

（注）マルチチャンネルステレオの音声モード表記法として、チャンネル数を次のように表す。
上層—中層—下層.LFE
割り当てチャンネルがない層は0 とし、割り当てチャンネルがある層のチャンネル数を次のように表す。
フロント／サイド／バック
なお、括弧内は従来の表記。

（理由）

・ 最大入力音声チャンネル数 22.2 チャンネルは、HDTV を超える高精細度映像サービス（UHDTV : SMPTE 2036-1）に対応した音場空間として位置付けられており、高度衛星デジタル放送の音声入力フォーマットとしてふさわしいと考えられる。
・ 現行のデジタル放送でもサンプリング周波数 32kHz と 44.1kHz は使用されておらず、利用ケースを考慮して、サンプリング周波数は 48kHz のみとした。

（留意すべき事項）

・ 非圧縮・ロスレス高音質サービス用符号化方式のサンプリング周波数として、高音質化を考慮した 96kHz 採用に向けた検討が行われている。

3. 音声符号化方式

3.1. 音声符号化方式（基本サービス用）

(1) 準拠規格

- MPEG-2 AAC 方式（ISO/IEC 13818-7）LC プロファイルとする。
  - AAC＋SBR（Spectral Band Replication）方式（符号化サンプリング周波数は入力サンプリング周波数の半分）の使用も可能とする。

(2) 符号化における制約条件

- 最大音声符号化チャンネル数：22.2 チャンネルとする。

（理由）

- MPEG-2 AAC 方式は既存のデジタル放送の音声符号化方式として実績があり、最大 22.2 チャンネルの音声入力にも対応可能と考えられる。
- MPEG-2 AAC＋SBR 方式も、既存のデジタル放送での比較的低ビットレート応用における音声符号化方式として実績があり、マルチリンガル音声サービス等の用途において有用性があると考えられる。

3.2. 音声符号化方式（非圧縮・ロスレス高音質サービス用）

(1) 準拠規格

(ア) 非圧縮符号化方式

- リニア PCM 音声符号化方式とする。
  - リニア PCM 音声信号（AES3 データ）の伝送には、SMPTE302M に準拠した PES 伝送方式を用いる。（PES 伝送方式については参考 5.1 参照）
  - AES3 のユーザビット（192 ビット）を用いて、音声モード識別、ダウンミックス係数等の音声関連メタ情報を伝送する。

(イ) ロスレス符号化方式

- MPEG-4（ISO/IEC 14496-3）ALS 方式を候補とする。
  - ただし、MPEG-4 ALS 方式について、放送応用に適したサブセット化（符号化ツールおよびパラメータの制約）を行う。（詳細検討中）
  - パラメータ記述部（ALSSpecificConfig）の AUX 領域を用いて、音声モード識別、ダウンミックス係数等の音声関連メタ情報を伝送する。

(2) 符号化における制約条件

- 最大音声符号化チャンネル数：5.1 チャンネルとする。

・ サンプリング周波数： 96kHz の追加を検討する。（詳細検討中）

（理由）

・ 非圧縮音声符号化であるリニア PCM 音声符号化方式は、非圧縮・ロスレス高音質音声サービスの実現が可能であり、高度衛星デジタル放送の音声符号化方式の 1 つとしてふさわしいと考えられる。
・ ロスレス音声符号化である MPEG-4 ALS 方式は MPEG 国際標準であり、高能率な符号化を実現しつつ非圧縮・ロスレス高音質音声サービスの実現が可能である。そのため、リニア PCM 音声符号化方式と同様、高度衛星デジタル放送の音声符号化の 1 つとしてふさわしいと考えられる。ただし、MPEG-4 ALS の最終的な採用にあたっては、放送応用に適したサブセット化の検討が必要である。
・ リニア PCM 音声符号化方式については、リニア PCM 音声信号を AES-3 データとして格納し、SMPTE302M に準拠した PES 伝送を行う技術的実現性を机上で確認したほか、実証実験を通じてその有効性を確認した。
・ MPEG-4 ALS 方式については、技術的実現性を机上で確認した。また、ALS ストリームを含む MPEG-2 TS 生成実験により、TS 化の実現性を確認した。
・ 非圧縮・ロスレス高音質サービス用符号化方式においては、高音質化を考慮し、サンプリング周波数 96kHz の追加を検討する。ただし、96kHz の最終的な採用にあたっては、周波数有効利用の観点を考慮し、適用可能な運用の制約条件など、詳細の検討が必要である。

（留意すべき事項）

・ これまでデジタル放送の映像・音声の符号化に関しては、周波数有効利用の観点から、効率的な伝送技術の一つとして圧縮符号化方式の導入が図られてきた。周波数有効利用の観点は今回の衛星デジタル放送高度化においても同様であり、これを受けて「低ビットレートかつ高音質な符号化方式であること」が音声符号化方式の要求条件として求められている。
・ 「非圧縮・ロスレス高音質サービス」は原音を全く劣化なく放送するサービスを想定しているため、このサービスを実現する音声符号化方式として、いわゆる高能率圧縮符号化とは異なる概念の符号化方式、すなわち非圧縮音声符号化、ロスレス音声符号化を選定したところである。そのため、周波数有効利用に関する上記の要求条件との整合性において、これまでに無い大きな伝送容量を要し、効率的な伝送の点で課題があることを認識する必要がある。
・ 特に、非圧縮音声符号化であるリニア PCM においては、音声信号のベースバンドでのビットレート（例えば 5.1ch、48kHz、24bit の場合、約 6.9Mbps）を固定的に伝送路として割り当てる必要がある。

・ ロスレス（圧縮）音声符号化である MPEG-4 ALS は音源によりその程度は異なるものの、一般にリニア PCM に比べれば一定の周波数有効利用に資すると考えられる。ただし、そのビットレートは基本的に音源に依存することから、サービス運用上の柔軟性については、リニア PCM 同様の課題がある。また、トータルでの周波数有効利用に寄与するためには、音声を優先した映像・データとの統合的ビットレート制御技術が必要である。このような技術はまだ基本的なシミュレーション検討が始まった段階にあり、今後実用的に使用可能な機器開発など、さらなる総合的技術検討が必要である。

4. 解説

高度衛星デジタル放送高度化に伴う音声符号化方式の暫定方式案を策定するにあたり、その背景として、想定される利用ケース、検討の経緯（方式採否の理由等）を以下に示す。また、暫定方式を採用するにあたっての今後の課題も紹介する。

4.1 想定利用ケース

既存のデジタル放送で実現されている音声サービス（MPEG-2 AAC LC プロファイルを使用）に加えて、以下の利用ケースを想定し、暫定方式の策定を行った。

(1) UHDTV に対応した高臨場感音声サービス

- HDTV を超える高精細度映像サービス、いわゆる UHDTV に対応する音場空間として、最大 22.2 チャンネルの音声入力フォーマットを提案。
- 受信した 22.2 チャンネル音声を、家庭での様々な再生環境に応じて、柔軟な処理を行うことで 3 次元音再生が可能になる。
- 本サービスに伴う想定ビットレート： 1.60Mbps 程度（22.2ch の場合）
  - 72kbps x 22ch＋10kbps x 2ch（LFE）で算出

(2) 非圧縮・ロスレス高品質音声サービス

- 現行の BS デジタル放送及び広帯域 CS デジタル放送と同等又はそれ以上の音質が望まれることを考慮し、MPEG-2 AAC 圧縮符号化による音声品質を超える最高音質サービスを実現するため非圧縮およびロスレス音声符号化を提案。音声チャンネル数は最大 5.1 チャンネル。
- 非圧縮・ロスレス高音質音声サービスとして新たな音質表示の割り当てを提案。
- 本サービスに伴う想定ビットレート：
  ＜非圧縮音声符号化＞
  - 1.53Mbps（2ch、サンプリング周波数：48kHz、量子化ビット数：16bit の場合）
  - 最大 6.9Mbps（5.1ch、サンプリング周波数：48kHz、量子化ビット数：24bit の場合）
- ＜ロスレス音声符号化＞
  - 音源によりビットレートは変動するが、一般的に非圧縮音声符号化に比べて少ないビットレートでのサービスが可能。

(3) マルチリンガル音声サービス

- マルチリンガル音声サービスなど、1 サービスあたりの音声ストリーム数が増えるサービスを想定し、比較的低ビットレート（概ね 64kbps/ステレオ以下）にお

いて有効と考えられる AAC＋SBR 方式を提案。

4.2　検討の経緯

衛星デジタル放送の高度化に関する方式提案募集に対して、音声符号化関連では 2 社からの提案があった。提案内容は以下の通り大きく 3 つに分類される。

- 音声入力フォーマット（最大 22.2 チャンネルへの拡張）
- 圧縮符号化方式（既存方式に加えて AAC＋SBR 方式）
- 非圧縮符号化方式（リニア PCM 音声符号化・伝送方式）

また、原音を忠実に再現可能かつ高能率符号化が可能な「ロスレス符号化方式」について追加提案募集を行った結果、2 方式（MPEG-4 ALS 方式および AAL for Broadcasting 方式）の提案があった。

これらの提案について、以下の通り検討を行い、その採用の可否の判断を行った。

(1) 音声入力フォーマット

- 想定利用ケース、視聴環境について
  - 最大チャンネル数 22.2ch は、前述の通り、UHDTV に対応する音場空間として提案されている。
  - 視聴環境としては、必ずしも 22.2ch 分のスピーカ配置を前提とするものではなく、視聴者の環境に応じて柔軟なダウンミックス処理を行い、3 次元音再生を実現することも想定している。
  - 22.2 チャンネルの音場空間、およびダウンミックス処理による 5.1 チャンネル・2 チャンネルステレオでの音再生について、音声符号化方式作業班内でデモンストレーションを実施し、その効果を確認した。（参考 5.2 を参照）
- 技術的実現性について
  - 想定ビットレートは 22.2ch で 1.60Mbps 程度であり、1 トランスポンダ内でのビットレート割り当てや送出運用上の実現可能性は特に問題ないと考えられる。
- 96kHz サンプリング周波数について
  - 高度衛星デジタル放送の音声入力フォーマットとして、96kHz サンプリング周波数の採用可否について検討を行った。
    - 96KHz サンプリング周波数は、プロオーディオの分野では多くのデジタル機器で採用されており、家庭用次世代メディア（Blu-ray や HD-DVD など）でも採用されている。したがって、番組制作（ベースバンド）においては、パッケージメディアなどへの二次展開を考慮すると、96KHz サンプリング周波数は有効なフォーマットと考えられる。

✧ 一方、デジタル放送を対象とした場合、以下の考察となる。

- 圧縮符号化方式（MPEG-2 AAC 方式）は規格上 96kHz サンプリング周波数に対応しているが、現在の想定ビットレート（72kbps/チャンネル）では全帯域（～50kHz）を伝送することが難しく、圧縮効率上の課題がある。
- 非圧縮符号化（リニア PCM 音声）の場合、96kHz サンプリング周波数に対応させるためには、5.1ch サラウンドで約 13.8Mbps が必要となり、衛星デジタル放送の高度化で検討されている 1 中継器あたり HDTV4 番組（4TS）では、5.1ch サラウンドの伝送は事実上不可能と言わざるを得ない。
- 一方、ロスレス音声符号化を採用した場合、リニア PCM に比べて一般的に少ないビットレートでの伝送が可能であり、またリニア PCM でも例えば音声チャンネル数を 2ch とした場合、96kHz、24bit で 4.6Mbps となる。

✧ 以上により、今回の衛星デジタル放送高度化に限定した場合、基本サービス用としては 96kHz サンプリング周波数は不採用と判断した。一方、非圧縮・ロスレス高音質符号化用としては、高音質化を考慮し、96kHz サンプリング周波数の採用を継続検討することとする。ただし、周波数有効利用の観点から、その採用についてはメリットの明確化、適用可能な運用制約の詳細検討が必要である。

(2) 音声符号化方式（基本サービス用）

・ 想定利用ケースについて

➢ 提案方式として明示されてはないが、各提案の前提（基本サービスにおける音声符号化方式）として、既存のデジタル放送の音声符号化方式として使用されている MPEG-2 AAC 方式が想定されている。

➢ 一方、AAC+SBR 方式は比較的低ビットレート（概ね 64kbps/ステレオ以下）において有効な方式であり、マルチリンガル音声サービスなど、1 サービスあたりの音声ストリーム数が増えるサービスでの利用を想定している。例えば、メインチャンネルは MPEG2-AAC 方式のままとし、複数のサブチャンネルに AAC+SBR 方式を採用する利用ケースなどが想定される。

➢ なお、AAC＋SBR 方式は上述の通り利用ケースが限定されるため、サービス上、また受信端末上「オプション」の位置付けとなることが想定される。

・ 技術的実現性について

➢ MPEG-2 AAC 方式は既にデジタル放送に使用されており、基本的な問題はない。また AAC+SBR 方式も地上デジタルテレビジョン放送の携帯向けサービ

ス（ワンセグ）や地上・衛星デジタル音声放送等で既に使用されており、比較的低ビットレートの応用において実績がある。

- 最大入力音声チャンネル数 22.2ch に対応した AAC デコーダは、現行デジタル放送に比べて MIPS4 倍以上、メモリ 4 倍以上程度と想定され、2011 年時点でのハードウェア実現性はあると考えられる。同様に、AAC エンコーダについても実現性はあると考えられる。
- ただし、音声入力フォーマットの拡張（最大 22.2 チャンネル）に対応した MPEG-2 AAC 方式のチャンネル構成（Channel Configuration）が規格上規定されていないなど、今後 MPEG 規格改訂等の作業が必要となる。

(3) 音声符号化方式（非圧縮・ロスレス高音質サービス用）

・ 想定利用ケースについて

- 前述の通り、MPEG-2 AAC 圧縮符号化による音声品質を超える最高音質の音声サービスを想定。これは、とりわけ放送の音声品質向上に期待する視聴者や、オーディオ業界および団体等からの要望にこたえるため、家庭でもスタジオ品質の音声放送を楽しめるよう、最大 5.1 チャンネルの非圧縮・ロスレス高品質音声サービスを目指すものである。
- 既存のデジタル放送（MPEG-2 AAC 方式）においても、音質表示において「現行の衛星標準テレビ放送の B モードに相当する音声品質」の基準が示されているが、提案された非圧縮音声符号化方式（リニア PCM 方式）およびロスレス音声符号化方式（2 方式）は「非圧縮（あるいはロスレス）高音質」であることを特徴としている。そのため、本提案方式に対応した新たな音質表示モードの割り当てが併せて提案されている。

・ 技術的実現性について

- 1 トランスポンダ内でのビットレート割り当て、送出運用上の実現可能性
  リニア PCM 方式の場合、想定ビットレートは 2ch で 1.53Mbps、最大でも 5.1ch で 6.9Mbps であり、特に問題ないと考えられる。また、ロスレス音声符号化方式の場合、ビットレートは音源によって可変であるが、リニア PCM 方式の伝送方式である SMPTE 302M-2007 より低ビットレートであり、同様に問題ないと考えられる。なお、96kHz サンプリング周波数については、周波数有効利用の観点から、実現可能な範囲での運用制約を前提としてその採用に向けた検討を継続中である。
- ハードウェア実現性
  リニア PCM 方式については、実証実験を通じてその実現性を確認した（参考 5.3 参照）。
  ロスレス音声符号化方式については、ハードウェア実現性を机上で確認した

（参考 5.5 参照）。また、ロスレス音声符号化ストリームを含む MPEG-2 TS 生成実験により、TS 化の実現性を確認した（参考 5.7 参照）。

・ ロスレス音声符号化方式について

- 非圧縮・ロスレス高音質サービスが実現可能かつ高能率符号化が可能な方式として、ロスレス音声符号化方式の技術検討を行った。技術検討に際しては、追加提案募集を実施した結果、2 方式の提案があった。以下、方式選定の経緯、ロスレス音声符号化のメリット、そして同等のサービス要件を満たすリニア PCM 方式との併記に関する検討結果を記す。
- ロスレス音声符号化方式選定の経緯
  - 以下 2 方式について、方式選定を目的とした技術検討・その他検討を実施した。
    ①MPEG-4 ALS(Audio Lossless Coding)　（以下、ALS）
    ②AAL(ATRAC Advanced Lossless) for Broadcasting　（以下、AAL）
  - ALS 方式は MPEG 国際標準であり、圧縮率や MPEG-2 Systems との整合性に対するメリットが期待できる方式である。一方、AAL 方式は実装の容易性（ハードウェア演算量・規模が小さくできる）や IPR（ライセンス条件）面でのメリットを追求した方式である。この 2 方式について、方式選定を目的とした比較表を作成した（2 方式の技術概要および方式比較表については、参考 5.5 を参照）。
  - その結果、主に国際標準であることのメリットを理由として、ALS 方式をロスレス音声符号化方式の第 1 候補として選定した。（AAL 方式は第 2 候補）
  - ALS 方式を第 1 候補として選定したものの、ハードウェア演算量・規模の小ささも方式採用における重要項目の一つと考えられる。そのため、適正なハードウェア演算量･規模の実現を目的として ALS 方式の新プロファイル・レベルの検討を引き続き行うこととする。
  - なお、現在 ALS 方式のライセンス条件は確定していない（MPEG 国際標準のため RAND によるパテントプール設立準備中）。方式の実際の採用にあたっては、今後ライセンス条件の確認も重要である。
- ロスレス音声符号化採用のメリット
  - ロスレス音声符号化は、一般論として平均的に原データに対して数十%の圧縮率が得られ、帯域の有効活用が期待できる一方で、最悪値（例えば白色雑音を音源とした場合）はほぼ原データ同等の情報量となり、削減したビットレートを固定的に他のサービスに割り当てられない点がネックとして指摘されていた。逆に言えば、この点を克服できるシナリオの存在がロスレス音声符号化方式を採用する上で不可欠である。

- ✧ 検討の結果、ロスレス音声符号化の採用により帯域の有効活用が図れるシナリオとして、例えば以下のシナリオを確認した。①事前収録番組において、オンエア事前に番組内のビットレート最悪値が把握できれば、理論最悪値と番組最悪値との差分に応じた固定的なビットレート削減が可能
  ②将来的に音声を優先した映像・データとの統合的ビットレート制御可能な送出装置、およびそれを考慮した受信装置が開発されれば、瞬間的なビットレート変動にかかわらず、トータルでの帯域有効活用が可能
  ③将来的にダウンロード型（事前ファイル化）のサービスにおいては、平均圧縮率に相当する帯域有効活用が可能
- ✧ 一方、上記シナリオの実現可否の判断については、さらに詳細な検討が必要である。例えば①のシナリオは事前収録番組に限られ、さらに番組単位での帯域制御が可能な送出設備が前提となる。また②のように実現には時期的かつコスト面での課題を含むシナリオも含まれている。また、リニア PCM と比較して、ロスレス符号化を採用することによるデコーダのコストアップの可能性も指摘された。
- ✧ しかしながら、（メリットを十分に発揮できるかどうかはさておき）2011 年の段階でロスレス符号化自体の実現は技術的に問題ないこと、また、将来的にはロスレス音声符号化の採用による帯域の有効活用が期待できることから、結論として現時点でロスレス音声符号化方式を採用することとした。

- ➢ リニア PCM 方式との併記について
  - ✧ 上述の通り、ロスレス音声符号化のメリットを享受できるシナリオは当面限定的であり、事業者の運用形態によってはリニア PCM の方にメリットがあるケースも想定される（例えば生放送中心のサービス）。
  - ✧ 将来的にはロスレス音声符号化が主流となり得ると考えられるが、現時点ではリニア PCM の方がデジタル放送への適用において実証実験による実証済みの方式であり、またロスレス音声符号化に対して相対的処理量や実装コストが小さいため、方式併記によるデメリットも少ない。そのため、リニア PCM 方式とロスレス音声符号化方式を併記することとした。

4.3 今後の課題

(1) 音声入力フォーマット

- ・ 22.2ch 音場空間の定義

→　SMPTE へ UHDTV Audio（デジタルオーディオ特性ならびに 22.2 マルチチャンネル）に関する規格案を提出し、現在審議中。

→　IEC/TC100 において、22.2 マルチチャンネル音響方式を含む General Channel Assignment 規格を審議中。

(2) 音声符号化方式（基本サービス用）

・ 音声入力フォーマットの追加に伴う AAC 規格の拡張

- Channel Configuration
- その他

→　MPEG 規格の改訂提案を行い、現在審議中。

(3) 音声符号化方式（非圧縮・ロスレス高音質サービス用）

・ リニア PCM 方式：音声メタ情報の伝送

- ダウンミックス係数、音声モード識別等

→　AES-3 ユーザビット（192bit）の利用を検討。JEITA （インタフェース標準化グループ）とのリエゾンにより、IEC60958-3 への改訂提案を審議中。

※検討中の音声メタ情報案について、参考 5.4 を参照

・ MPEG-4 ALS 方式：放送応用に応じたサブセット化の検討、実証実験の実施

- 新しいプロファイル・レベルの検討（ツールやパラメータの制約）
- 伝送実験による実現性の検証（特に周波数有効利用シナリオの実現性）

→　上記検討・実施の後、MPEG 規格の改訂提案を行うことを検討。

・ 非圧縮・ロスレス音声符号化における 96kHz サンプリング周波数の検討

- 周波数有効利用の観点から、適用可能な運用制約の検討
- 96kHz サンプリング周波数追加に伴うメリットの検討

5. 参考

5.1 リニア PCM 音声の PES 伝送方式

リニア PCM 音声信号（AES3 データ）の伝送方式として、SMPTE 302M に準拠する PES 伝送方式を用いる。

5.1.1 準拠規格

SMPTE 302M-2007 “Mapping of AES3 Data into an MPEG-2 Transport Stream”

5.1.2 概要

- 複数の AES3 ストリーム（5.1ch の場合 3 つ）は 1 本の SMPTE AES3 ES に時分割多重される。
- SMPTE AES3 ES は MPEG-2 Systems（ISO/IEC 13818-1）準拠の PES パケットにマッピングされる。

5.1.3 PES パケット

(1) データ構造

・AES3 データを伝送する PES のデータ構造は以下の通りとする。

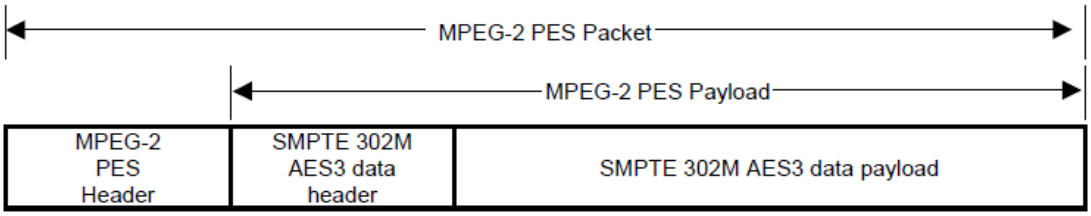


図 5.1-1 PES パケットのデータ構造

(2) PES ヘッダ

・MPEG-2 Systems（ISO/IEC 13818-1）に準拠する。

・ストリーム識別子：0xBD（プライベートストリーム 1）

・PTS_DTS_flags：01（PTS あり）

※PTS の値はビデオフレームに対応した PTS と同一とする。

(3) SMPTE 302M AES3 データヘッダ

・SMPTE 302M-2007 ”SMPTE AES3 data elementary stream header”の規定に準拠する。

・number_channels：最大 6 データチャンネル（AES3 データ 3 本）とする。

(4) SMPTE 302M AES3 データ本体

・SMPTE 302M-2007 ”SMPTE AES3 elementary streams”の規定に準拠する。

5.2　22.2 チャンネル音声デモンストレーション

5.2.1　22.2 マルチチャンネル音響方式の概要

22.2 マルチチャンネル音響は、超高精細かつ広視野大画面映像がもたらす臨場感や没入感に対応した音場再現を行う目的で開発された次世代のマルチチャンネル音響システムである。基本的な開発コンセプトは下記のとおり。

(1) 映像と音像の方向を一致させるために、広視野画面上に安定した音像定位を実現。
(2) 聴取者に対しあらゆる方向からの音の到来と音の上下感を実現。
(3) 臨場感や現実感の高い三次元音響空間再生
(4) 高品質な音場再生を受聴できるエリアが広い
(5) 既存のマルチチャンネル音響方式と互換性を有する

これらのコンセプトに基づき、下図のように上層に 9 チャンネル、中層に 10 チャンネル、下層に 3 チャンネルを配置し、LFE(低域効果チャンネル)を 2 チャンネルとした。

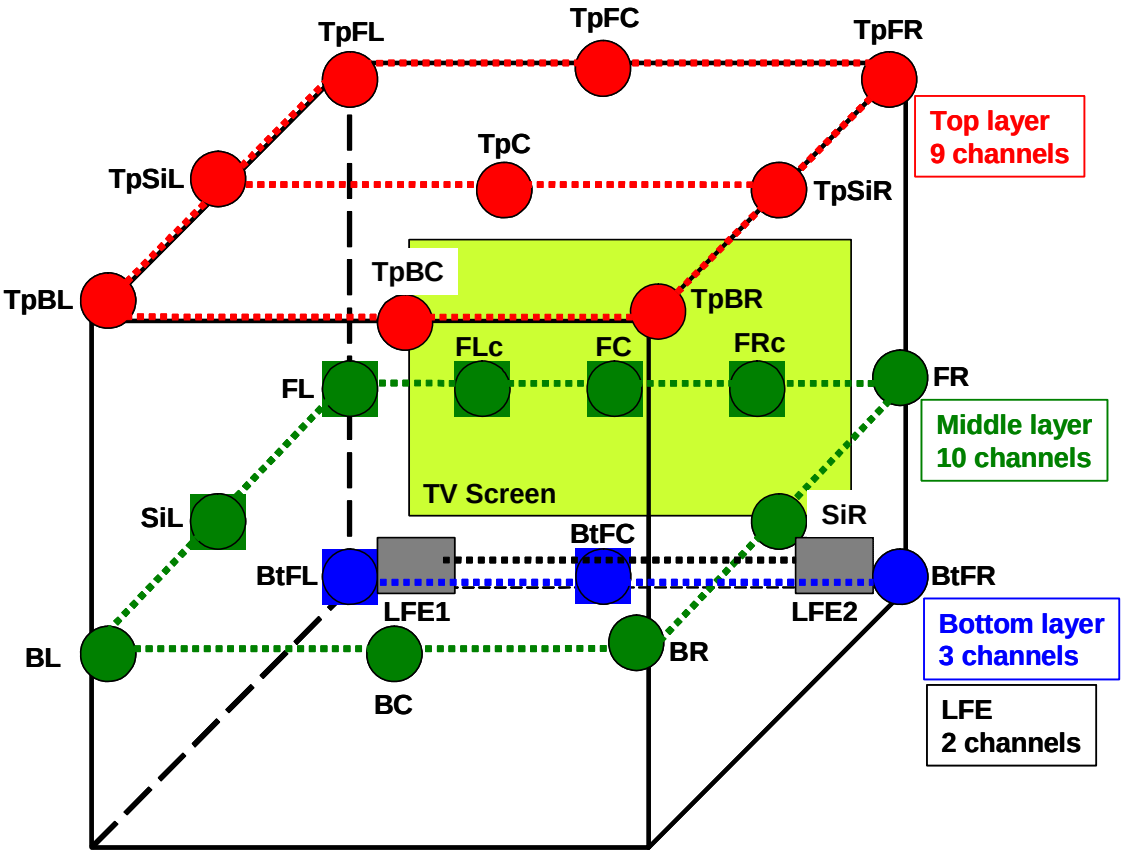


図 5.2-1　22.2 マルチチャンネル音響システム

5.2.2.　22.2 チャンネル音声の聴取確認

22.2 チャンネル音声の品質確認のため、NHK 技研において当作業班委員に対し 22.2 チャンネル音声のデモンストレーションを行った。また、22.2 チャンネル音声からダウンミックスした 5.1 チャンネル音声の品質確認も同時に行った。

・　実施日

平成 19 年 10 月 18 日

・実施場所

NHK 放送技術研究所

・デモンストレーション内容

1. シアター形式（画面サイズ 450 インチの SHV）での試聴（5.1 チャンネルダウンミックス音声含む）
2. 家庭などの小空間再生を想定した試聴（画面サイズ 50 インチの HDTV）（5.1 チャンネルダウンミックス音声含む）

・委員の講評

現行の 5.1 マルチチャンネル音声に比べ、シアター形式、小空間双方において 22.2 マルチチャンネル音声による空間再生品質が十分に高いことを確認した。

5.3　リニア PCM 音声符号化・伝送実証実験報告

高度 BS デジタル放送の音声符号化暫定方式の検討に関連して、ARIB 音声符号化方式作業班では SMPTE302M に基づくリニア PCM 音声信号の MPEG-2 システム伝送の特性確認を行った。

・　実施日

平成 19 年 12 月 13 日

・　参加者

浦野（音声符号化方式作業班主任、日本テレビ）、濱崎（NHK）、中山（NHK）

・　実験場所

NHK 技術研究所　CB209 実験室

・　実験項目

① エンコーダーデコーダの 5. 1 マルチチャンネル音声信号の各チャンネル間同期性能の確認

② エンコーダーデコーダの映像信号と音声信号の同期確認

③ 映像・音声素材による視聴確認

・　符号化器

➢　概要

✧AVC/H.264

✧SMPTE 302M（2ch/4ch/6ch/8ch）

✧エンコード＋デコードの遅延量　　約 800msec（メーカ公表値、標準モード）

✧デコーダの音声遅延用バッファ量　約 0.5sec

参考：MPEG2（video）、MPEG-1 Audio Layer II、MPEG-2 AAC の機能を持つ

➢　実証実験で使用した符号化パラメータ

✧TS RATE　　17Mbps

✧VIDEO RATE　　8.775Mbps（H.264　4:2:0 フォーマット）

✧AUDIO RATE　　6.912Mbps　（SMPTE302M 準拠　6ch モード）

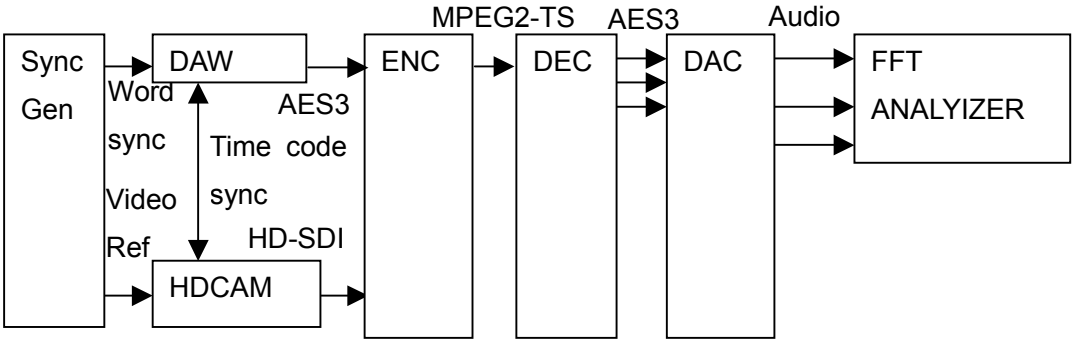
・　実験

① エンコーダーデコーダの 5.1 マルチチャンネル音声信号の各チャンネル間同期性能の確認

＜実験方法＞

音声フォーマット		24bit/48kHz　リニア PCM 信号（AES3 方式）
映像フォーマット		1920x1080（1080i）、16:9、（HD-SDI 方式）
音声テスト信号		広帯域ピンクノイズ（全チャンネルに同位相信号）
評価方法		チャンネル間の相互相関関数の測定
使用機器	エンコーダ/デコーダ	NTT エレクトロニクス　HVE9100/HVD9100
	デジタルオーディオワークステーション	フェアライト　Merlin
	ビデオ再生器	ソニー　HDW-250
	DA コンバータ	Emmlab　MarkIV　DAC
	シンクジェネレータ	ROSENDAHL Nanosyncs HD
	測定器	小野測器マルチパーパース FFT アナライザ CF-5220

＜実験システム＞



＜実験結果＞

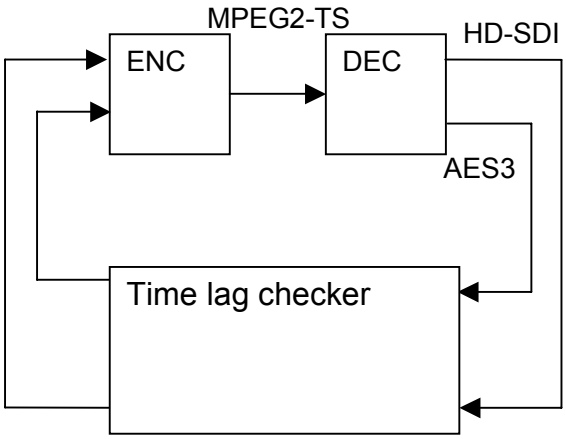
1ch に対して、2～6ch の時間差を 1ch アナログ信号とその他のアナログ信号の相互相関関数を測定しピーク位置を観測した。その結果、1ch と他のチャンネルの時間差はすべて 0μsec であった。

② エンコーダーデコーダの映像信号と音声信号の同期確認

＜実験方法＞

音声フォーマット		24bit/48kHz　リニア PCM 信号（AES3 方式）
映像フォーマット		1920x1080（1080i）、16:9、（HD-SDI 方式）
テスト信号		測定器の測定用信号を使用
評価方法		タイムコード（ＴＴＣ,ＬＴＣ）を利用した AV 時間差の測定
使用機器	エンコーダ/デコーダ	NTT エレクトロニクス　HVE9100/HVD9100
	測定器	YEM エレテックス　マルチフォーマットタイムラグチェッカー EDD 5130

＜実験システム＞



＜実験結果＞

電源の ON/OFF を数回繰り返し、AV 時間差を測定した。

１回目　0msec

２回目　0msec

３回目　0msec

４回目　0msec

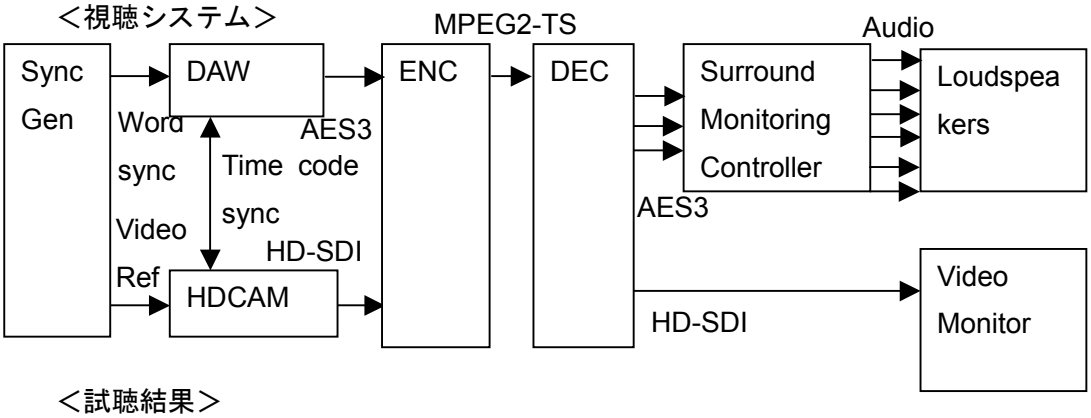
参考：測定開始時に 1msec の時間差が測定された。符号化器のリセットをしなかったためにエラーが生じたものと思われる。

③ 映像・音声素材による視聴確認

＜実験方法＞

音声フォーマット	24bit/48kHz　リニア PCM 信号（AES3 方式）	
映像フォーマット	1920x1080（1080i）、16:9、（HD-SDI 方式）	
テスト信号	NHK 制作の 5.1ch サラウンド番組集	
評価方法	専門家による視聴確認	
使用機器	エンコーダ/デコーダ	NTT エレクトロニクス　HVE9100/HVD9100
	デジタルオーディオワークステーション	フェアライト　Merlin
	ビデオ再生器	ソニー　HDW-250
	ビデオモニタ	ソニー　BVW-D32E1WU
	再生スピーカ	Musikelectronic Geithain RL901K
	サラウンドモニタコントローラ	GRACE design m906
	シンクジェネレータ	ROSENDAHL Nanosyncs HD

＜視聴システム＞



＜試聴結果＞

いくつかの番組を視聴したが問題なく再生できていることを確認した。

参考：AES3 のユーザビット（192 ビット）を用いた、音声モード識別、ダウンミックス係数等の音声関連メタ情報の伝送は、今回実施していない。

・　まとめ

高度 BS デジタル放送の音声符号化暫定方式の検討に関連して、SMPTE302M に基づくリニア PCM 音声信号の MPEG-2 システム伝送の確認を行った。実験結果より、現在、実現化されている符号化器で問題なく伝送できることを確認した。

高度 BS デジタル放送の暫定方式において、映像符号化方式として H.264 が提案されてている。本実験で使用した実機の場合、エンコード/デコードの処理時間として、約 800msec である。デコーダ側に 0.5 秒の音声遅延用バッファを利用することにより、映像と音声の同期ずれを補償している。

5.4　AES3 のユーザビットを用いたリニア PCM 音声関連メタ情報の伝送

リニア PCM 方式において AES3 のユーザビットを用いた音声関連メタ情報伝送の検討が行われている。以下、伝送することが望ましい音声メタ情報について検討状況を紹介する。

■リニア PCM 音声メタ情報例

データ構造	ビット数	ビット列表記
LinerPCM_Meta_descriptor(){		
component_type	8	uimsbf
matrix_mixdown_idx	3	uimsbf
future_reserve	13	uimsbf
}		

［コンセプト］

SMPTE302M では複数の AES３ストリームを一つの MPEG2-PES にした際のデコード情報を規定している。AES3 では 1AES ストリームをデコードするための情報をチャンネルステータスビットとして規定している。そこで、これらの情報を利用したうえで、放送に必要な音声メタ情報を AES3 のユーザビットに規定することとする。

［各規格のメタ情報の役割］

［識別子詳細］

component_type(音声モード識別); 8 ビット

　音声コンポーネント種別　ARIB STD-B10 第 2 部 表 6-43 参照＋提案分含む

matrix_mixdown_idx(ダウンミックス係数);3 ビット

　ダウンミックス係数　ARIB STD-B21 第 6 章 表 6-DM2 参照

future_reserve(将来利用のための予約);13 ビット

［コメント］

番組間音量差、チャンネル間音量差の改善を目的に、future_reserve の一部、もしくはすべてを reference_loudness_idx として ITU-R で現在審議されている参照ラウドネスを伝送し、受信環境のラウドネスコントロールを行うために使用できればと考える。

5.5　ロスレス音声符号化に関する提案方式（2 方式）の技術概要および方式比較

ロスレス音声符号化方式の技術方式提案募集に対して提案された 2 つの方式について、その技術概要および 2 方式の特長比較を以下にまとめる。

5.5.1　提案方式の技術概要

(1) MPEG-4 ALS(Audio Lossless Coding)

＜技術方式の概要＞

歪を伴わない圧縮符号化　ISO/IEC 14496-3 MPEG-4 ALS（Audio Lossless Coding）準拠.

- エンコーダ、デコーダで完全に整合がとれた確定的整数演算により、復号時に入力波形を完全再構成することを理論的に保証。
- 線形予測による隣接サンプル間の冗長性除去による情報圧縮。
- 予測誤差サンプルの振幅値をエントロピー符号化して振幅分布の偏りの冗長性除去による情報圧縮。
- 音声メタ情報はパラメータ記述部（ALSSpecificConfig）の Aux 部分を拡張して記述。

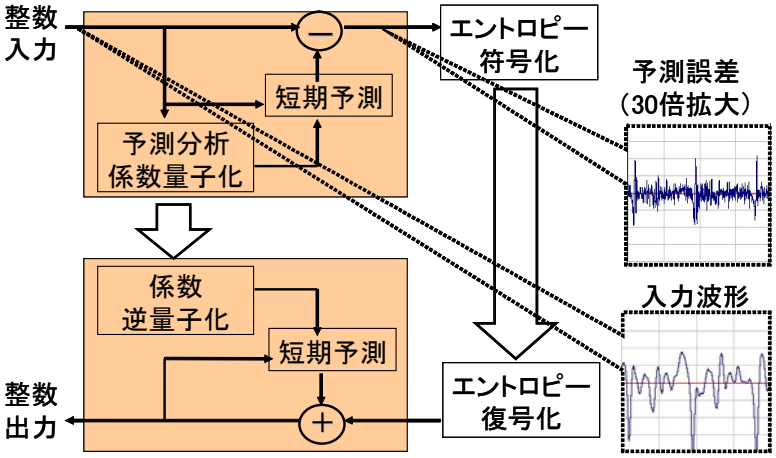


図 5.5-1　ALS のエンコーダ、デコーダの原理

＜放送システムへの適用＞

- MPEG-2 TS のオーディオパケットは PES パケットを含む。
- MPEG-2 システムの PMT パケットに MPEG-4_audio_extension_descriptor を含み、その中に AudioSpecificConfig と ALSSpecificConfig を含む。
- MPEG-2 システムの PES パケットにフレーム単位で ALS 準拠ペイロードを含む。

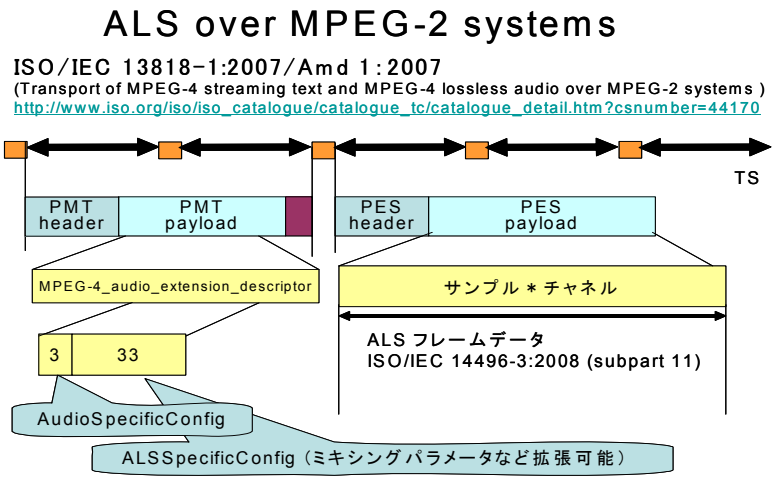


図 5.5-2　MPEG-2 Systems での伝送方法

(2) AAL(ATRAC Advanced Lossless) for Broadcasting

＜技術方式の概要＞

ロスレス符号化技術 ATRAC Advanced Lossless for Broadcasting (AAL for BC)

- 16bit 固定小数点演算を前提とした確定的整数演算により、プラットフォームを選ばずに入力波形を完全再構成することが可能
- 線形予測分析によるサンプル間相関除去により効率的な情報量圧縮が可能
- 必要かつ十分な線形予測次数による高速な符号化/復号化が可能
- 圧縮率の最悪値が確定しており、リニア PCM とほぼ同等である。具体的にはフレーム長 2048, 16bit, 2ch データの場合、リニア PCM の 100.07%

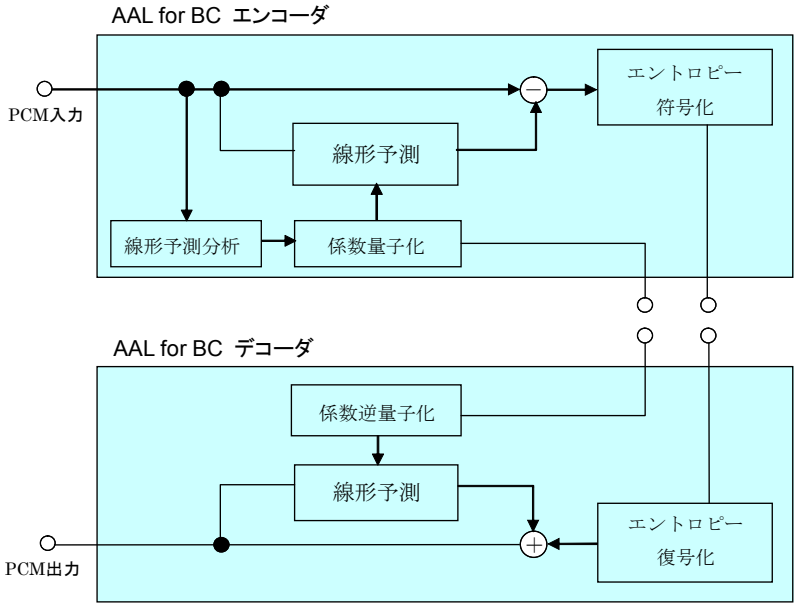


図 5.5-3　AAL for Broadcasting エンコーダおよびデコーダ ブロック図

＜放送システムへの適用＞

- MPEG-2 TS の PES パケットにて伝送する
- AAL for BC の Elementary Stream を以下のように MPEG2-TS の PES にマッピングした上で伝送を行う

  ーPES ヘッダは MPEG-2 Systems (ISO/IEC13818-1) に準拠

  ーストリーム ID はプライベート用のものを使用

  ービデオと同期した PTS を付加

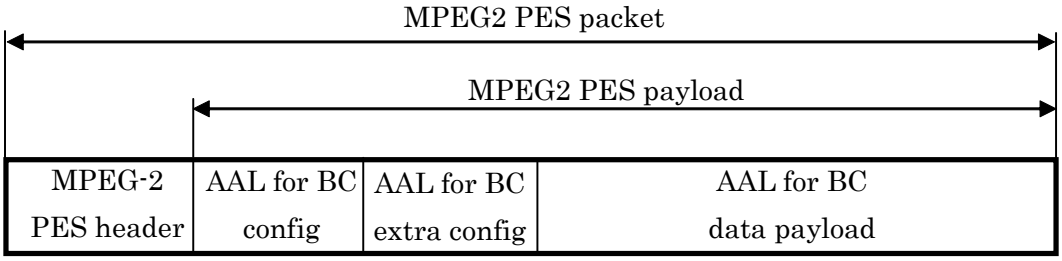


図 5.5-4　AAL for BC PES パケット構造

- AAL for BC 設定情報は AAL for BC 設定情報、AAL データペイロードを含む
- AAL for BC 設定情報は AAL for BC 基本設定情報、AAL for BC 拡張設定情報を含む
- AAL for BC 拡張設定情報はチャンネル配置情報を伝送可能
- AAL for BC 拡張設定情報はメタ情報領域を持ち、最大 2^{32} バイトまでのメタ情報を伝送可能

5.5.2　提案方式の方式比較

2 方式の方式比較表を表 5.5-1 に示す。なお、比較表には参考のためリニア PCM（SMPTE302M）についても記載している。

5.5.3　その他考慮すべき事項

2 方式の比較検討において、上記の比較表への記載事項のほか、考慮すべき事項として指摘された項目を以下に示す。

・ 国際標準化のリスク

  ➢ AAL は現時点で国際標準化されておらず、今後国際標準化機関に提案した場合、時期的に遅れる、あるいは国際標準化できないリスクがある。一方、ALS についても、新規規定のプロファイル・レベルは現時点で国際標準化されていないため、同様のリスクがある。

・ 帯域有効活用シナリオ

  ➢ ロスレス音声符号化方式である ALS、AAL のビットレートは音源によって変動

し、最悪の場合、圧縮効果が認められないケースが理論上あり得る。そのため、ロスレス音声符号化方式の採用にあたっては、音声ビットレートの変動においても総合的に帯域有効活用できるシナリオ、例えば映像やデータとの統合制御シナリオの存在が重要である。

表 5.5-1　提案方式（2 方式）の比較表

	ALS	AAL for Broadcasting	（参考） リニア PCM(SMPTE302M)
提案方式の特徴	国際標準準拠で、柔軟性、自由度圧縮性能に優れる。	実装が容易で無償ライセンス。最悪レートが PCM レートとほぼ同じ。	伝送蓄積などで使用実績が豊富
方式上の比較	線形予測最高 15 次、付加ツールで圧縮性能改善。多様な入力に対応。	線形予測 8 次、予測係数精度が 16 ビットで演算量低減。	
圧縮率　（注 1） （圧縮後のレート/PCM レート）	可変レート（入力信号依存） ファイル毎の平均値： 0.17%～106.32% フレーム毎の最悪値： 8.67%～106.32%	可変レート（入力信号依存）（注 2） ファイル毎の平均値： 26.65%～100.12% フレーム毎の最悪値： 33.84%～100.12%	固定レート（平均＝最悪） 16 bit のとき(20÷16＝)125%、24 bit のとき(28÷24≒)117%。
ハードウェアの実現性	低演算プロファイルで AAC-LC と同程度 25 MHz (48 kHz, 24 bit, 2 ch) @ARMulator ARM9E ROM 6 Kbyte, RAM 80 KByte @ARMulator ARM9E	容易 16.3 Mcyc/s (48 kHz, 16 bit, 2 ch) @NEC MP201 18.6 Mcyc/s (48 kHz, 24 bit, 2 ch) @NEC MP201 ROM 7 Kbyte, RAM 16.6 KByte @NEC MP201	非常に容易
国際標準との整合性	伝送は MPEG-2 準拠 符号化・復号化は MPEG-4 準拠 低演算プロファイル提案予定		SMPTE302M (2007)
IPR の扱い	RAND によるパテントプール設立準備中、コンテンツ課金なし。	対価請求はなし	無償
方式実現に向けた今後の課題	MPEG-4 Audio の改訂：新規の低演算量プロファイル、レベル定義	提案符号化方式の国際標準化 提案伝送方式の国際標準化	

（注 1）ロスレス音声符号化・圧縮性能実証実験（参考 5.6 章参照）での実測値

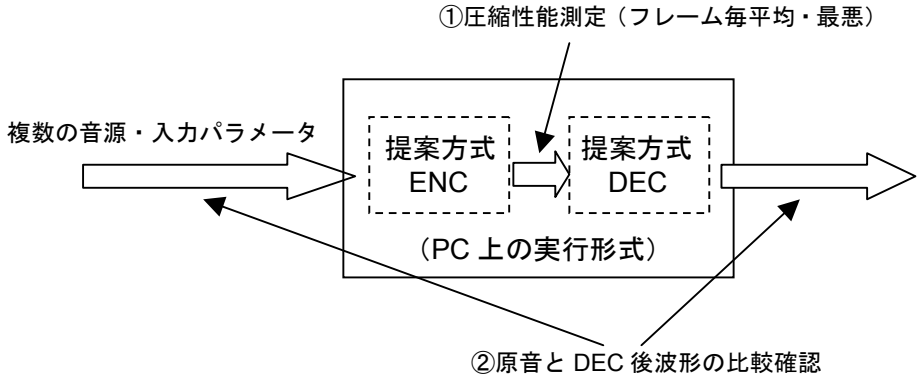
（注 2）今回測定に用いた AAL アルゴリズムは、20 ビット入力に対して、エンコード時に 4 ビット右シフトし、LSB 側 20 ビット詰めで 24 ビット処理を行い、デコード時には 4 ビット左シフトし、MSB 側 20 ビットを取り出し出力している

5.6　ロスレス音声符号化・圧縮性能実証実験報告

〇目的

高度衛星デジタル放送方式のロスレス音声符号化方式として提案された 2 方式（MPEG-4 ALS、AAL for Broadcasting）について、複数の音源・音声入力パラメータを用いて圧縮性能を測定し、その性能を実証する。

〇実験概要



・ ロスレス音声符号化提案 2 方式の CODEC（PC 上の実行形式ファイル）に対して複数の音源・入力パラメータを入力し、以下の測定を行う。
  - 原音データに対する、圧縮後の情報量の割合（値が小さいほうが効率がよい）のフレーム毎の平均値およびフレーム毎の最悪値を測定
    情報量の割合（%）＝（圧縮後の情報量）÷（PCM の情報量）×100
  - 原音とデコード後波形の一致の比較確認

・ 使用 CODEC の諸元
  - 2 方式ともにフレームサイズ＝2048 サンプル/フレーム
  - ALS サブセット：
    ✧48 kHz: 線形予測次数最大 15 次、MCC/JS, 3 段までのブロック分割、LTP
    ✧96 kHz: 線形予測次数最大 15 次、MCC/JS

〇評価音源

名称	入力パラメータ	提供者
MPEG 評価音源 x 15 種	48kHz 16bit 2ch 30sec	
MPEG 評価音源 x 15 種	48kHz 20bit 2ch 30sec	
MPEG 評価音源 x 15 種	48kHz 24bit 2ch 30sec	
MPEG 評価音源 x 15 種	96kHz 24bit 2ch 30sec	
シンセ	48kHz 24bit 2ch 184sec	SONY

室内楽	48kHz 16bit 2ch 44sec	SONY
ノイズ	48kHz 16bit 2ch 55sec	SONY
合成試験音源（一様分布白色雑音）	48kHz 16bit 1ch 10sec	NTT
合成試験音源（正弦波 1 kHz）	48kHz 16bit 1ch 10sec	NTT
合成試験音源（矩形波 1 kHz）	48kHz 16bit 1ch 10sec	NTT
合成試験音源（定数、無音）	48kHz 16bit 1ch 10sec	NTT
チェンバロ	96kHz 24bit 2ch 34sec	NHK
和太鼓	96kHz 24bit 2ch 43sec	NHK
ピアノ	96kHz 24bit 2ch 35sec	NHK
チェロ＋ピアノ演奏	96kHz 24bit 2ch 42sec	NHK
弦楽四重奏	96kHz 24bit 2ch 43sec	NHK
合奏曲 1	96kHz 24bit 5.1ch 30sec	NTT
合奏曲 2	96kHz 24bit 5.1ch 30sec	NTT

○担当およびスケジュール

項目	担当	スケジュール
実験計画の承認、音源の選定	音声符号化方式 WG	～6/2
実験準備（PC 実行形式 CODEC、音源、測定用スクリプト）	方式提案社	～6/6
測定実施	WG 主任	6/6、6/13（追加）

○実験結果

(1) 原音データとデコード後データの同一性の検証

今回実施した全てのデータにおいて、原音データとデコード後データが同一であることを PC 上でのファイル比較コマンドにより確認した。

(2) 各音源に対するフレーム平均値・フレーム最悪値

いずれも　単位：[%]

MPEG 音源 (48 kHz, 16 bit, stereo)	AAL 平均	AAL 最悪	ALS 平均	ALS 最悪
avemaria	40.39	55.46	38.54	52.97
blackandtan	56.60	74.63	55.11	73.82
broadway	50.50	76.27	48.57	73.68
cherokee	54.01	71.77	52.86	70.75
clarinet	48.47	62.40	46.96	61.19
cymbal	30.32	85.23	29.05	83.78
dcymbals	61.35	80.88	60.39	80.60
etude	43.78	57.85	41.63	56.03

flute	41.21	51.28	39.94	50.02
fouronsix	47.15	67.59	45.76	66.60
haffner	56.04	67.49	54.55	66.10
mfv	32.18	53.10	29.96	42.58
unfo	52.55	74.04	50.72	73.57
violin	49.56	63.72	47.68	61.74
waltz	54.01	76.23	52.55	75.52

MPEG 音源 (48 kHz, 20 bit, stereo)	AAL 平均	AAL 最悪	ALS 平均	ALS 最悪
avemaria	52.23	64.39	50.56	62.32
blackandtan	65.29	79.71	63.99	79.07
broadway	60.44	81.02	58.84	78.96
cherokee	63.16	77.42	62.13	76.62
clarinet	58.82	70.03	57.47	68.98
cymbal	39.98	88.18	38.88	87.07
dcymbals	69.09	84.71	68.33	84.51
etude	55.01	66.33	53.11	64.85
flute	52.98	61.14	51.85	60.07
fouronsix	57.59	74.12	56.26	73.30
haffner	64.93	74.01	63.63	72.90
mfv	45.60	62.64	43.66	54.00
unfo	61.98	79.23	60.30	78.88
violin	59.72	70.98	58.12	69.40
waltz	63.20	81.00	61.90	80.45

注）今回測定に用いた AAL アルゴリズムは、20 ビット入力に対して、エンコード時に 4 ビット右シフトし、LSB 側 20 ビット詰めで 24 ビット処理を行い、デコード時には 4 ビット左シフトし、MSB 側 20 ビットを取り出し出力している

MPEG 音源 (48 kHz, 24 bit, stereo)	AAL 平均	AAL 最悪	ALS 平均	ALS 最悪
avemaria	60.19	70.33	58.79	68.59
blackandtan	71.07	83.09	69.97	82.55
broadway	67.04	84.18	65.68	82.45
cherokee	69.30	81.18	68.42	80.50
clarinet	65.68	75.02	64.54	74.14
cymbal	49.93	90.15	48.99	89.19
dcymbals	74.24	87.26	73.59	87.08
etude	62.51	71.94	60.91	70.69

flute	60.82	67.63	59.86	66.70
fouronsix	64.66	78.43	63.53	77.73
haffner	70.78	78.34	69.68	77.40
mfv	54.67	68.87	53.03	61.65
unfo	68.31	82.69	66.90	82.38
violin	66.43	75.81	65.08	74.49
waltz	69.34	84.16	68.23	83.68

MPEG 音源 (96 kHz, 24 bit, stereo)	AAL 平均	AAL 最悪	ALS 平均	ALS 最悪
avemaria	52.48	61.30	50.90	58.69
blackandtan	52.68	87.87	45.45	60.87
broadway	58.63	76.77	57.44	73.47
cherokee	51.17	65.35	44.75	58.73
clarinet	49.76	59.20	42.75	50.22
cymbal	47.23	90.13	46.20	89.66
dcymbals	60.81	79.37	59.85	78.75
etude	54.10	64.05	52.40	61.38
flute	47.93	53.56	42.80	46.91
fouronsix	46.82	69.11	40.93	54.95
haffner	54.96	62.87	48.79	54.15
mfv	52.48	67.61	50.40	65.18
unfo	50.93	88.25	43.71	60.39
violin	51.07	57.57	45.89	51.33
waltz	51.60	88.98	44.37	61.04

独自音源（48 kHz）		AAL 平均	AAL 最悪	ALS 平均	ALS 最悪
Sony (2ch)	item1（シンセ）	73.58	86.98	72.00	86.08
	item2（室内楽）	38.22	44.08	37.01	42.50
	item3（ノイズ）	100.06	100.06	106.32	106.32
NTT (1ch)	Rand（白色雑音）	100.12	100.12	103.17	103.44
	Sine_1kHz（正弦波）	33.80	33.84	9.01	9.89
	Square_1kHz（矩形波）	88.40	88.43	36.07	42.92
	ZeroConst（定数）	26.65	63.77	0.17	8.67

独自音源（96 kHz）		AAL 平均	AAL 最悪	ALS 平均	ALS 最悪
NHK (2ch)	cembalo	73.48	82.49	72.03	81.68

	drum_japan	51.29	80.34	49.09	79.48
	Pf+cello	57.65	67.08	55.53	64.70
	piano	47.59	59.24	45.24	50.59
	quartet	50.94	60.15	48.22	54.54
NTT (5.1ch)	Ensemble1	41.51	46.60	39.68	43.03
	Ensemble2	39.35	42.70	37.38	40.83

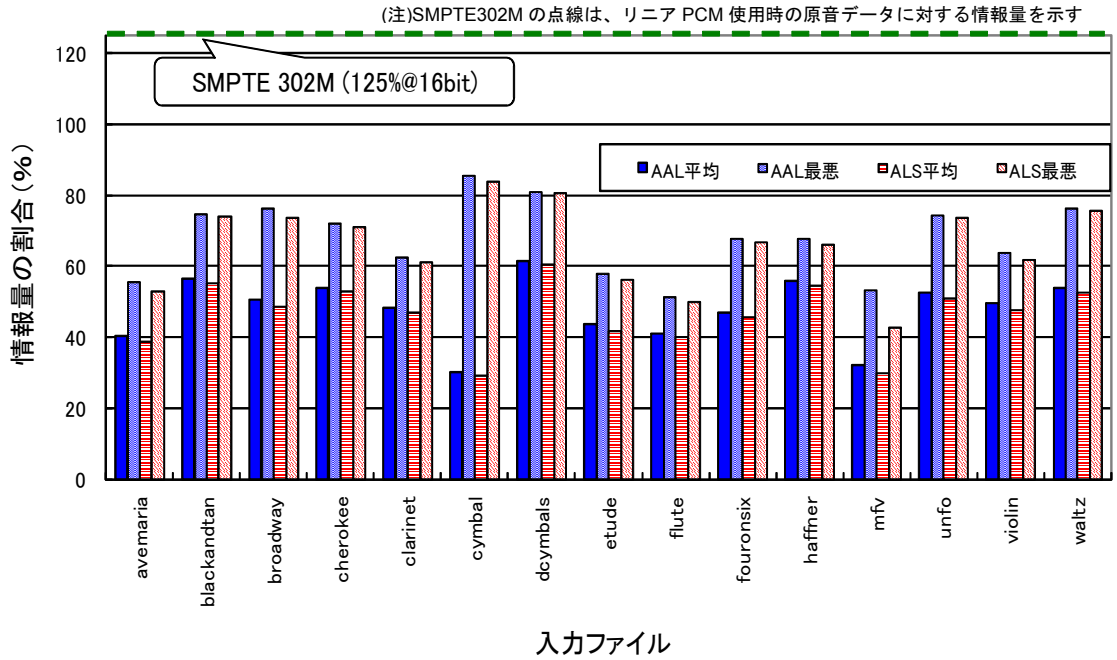


図 5.6-1(a)　MPEG 音源 15 種類（48kHz、16bit、stereo）

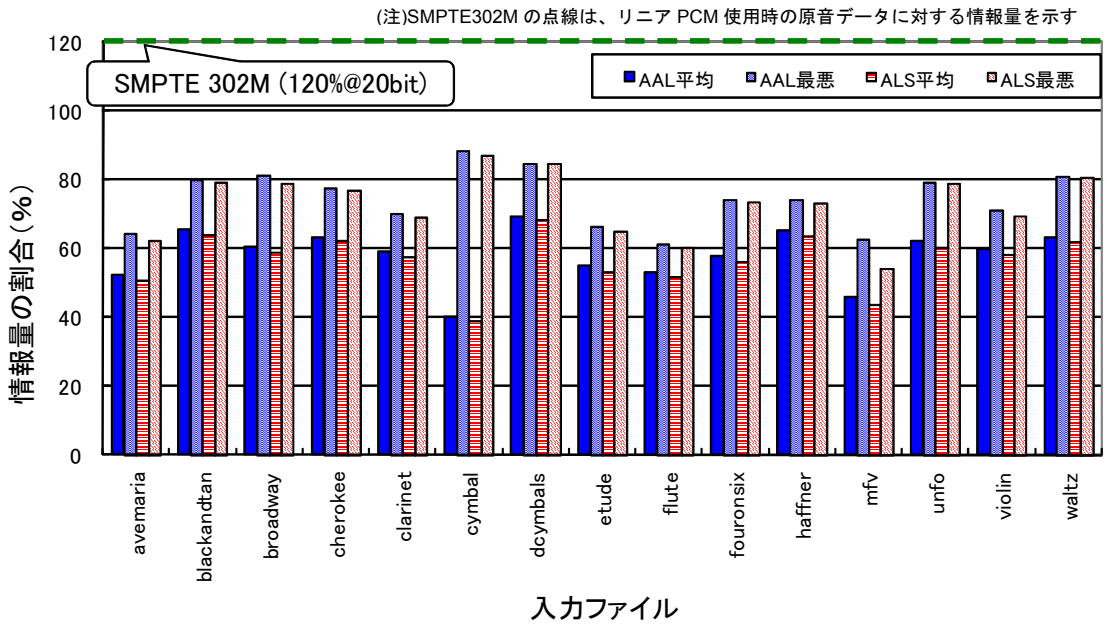


図 5.6-1(b)　MPEG 音源 15 種類（48kHz、20bit、stereo）

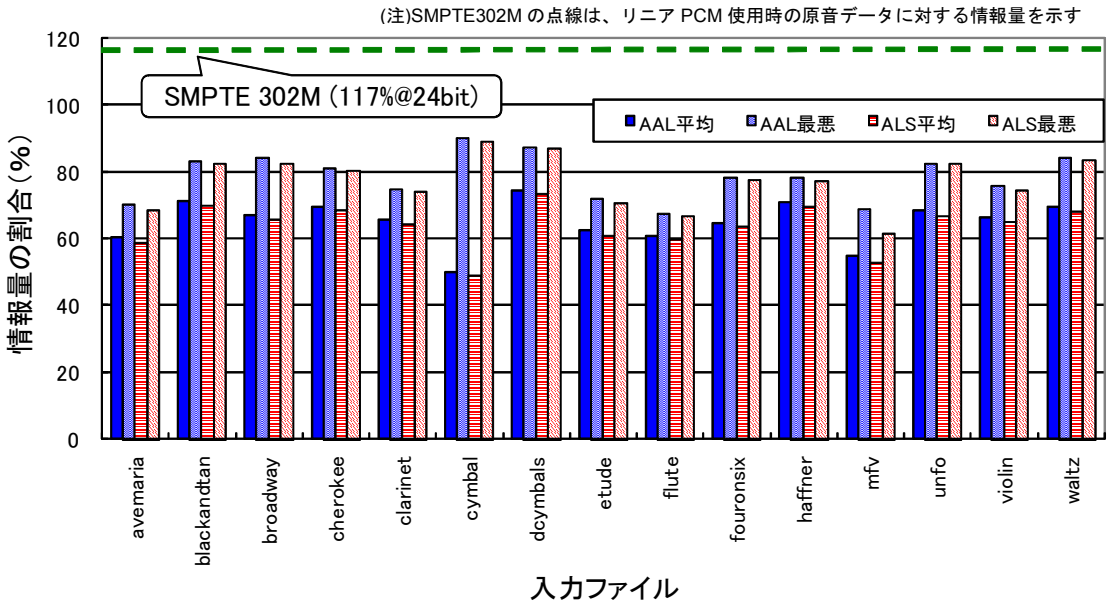


図 5.6-1(c)　MPEG 音源 15 種類（48kHz、24bit、stereo）

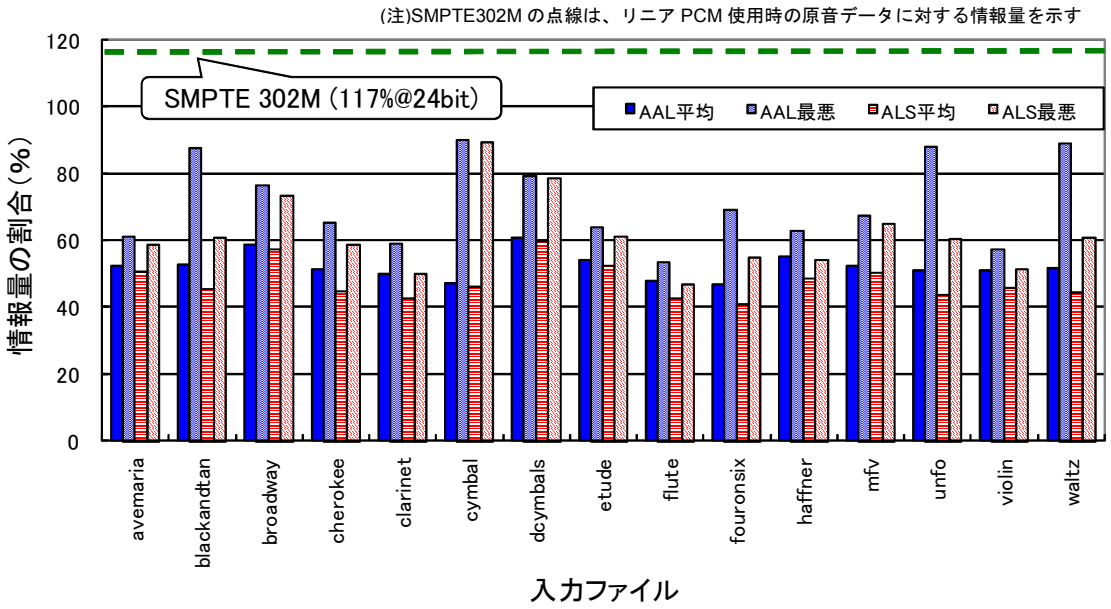


図 5.6-1(d)　MPEG 音源 15 種類（96kHz、24bit、stereo）

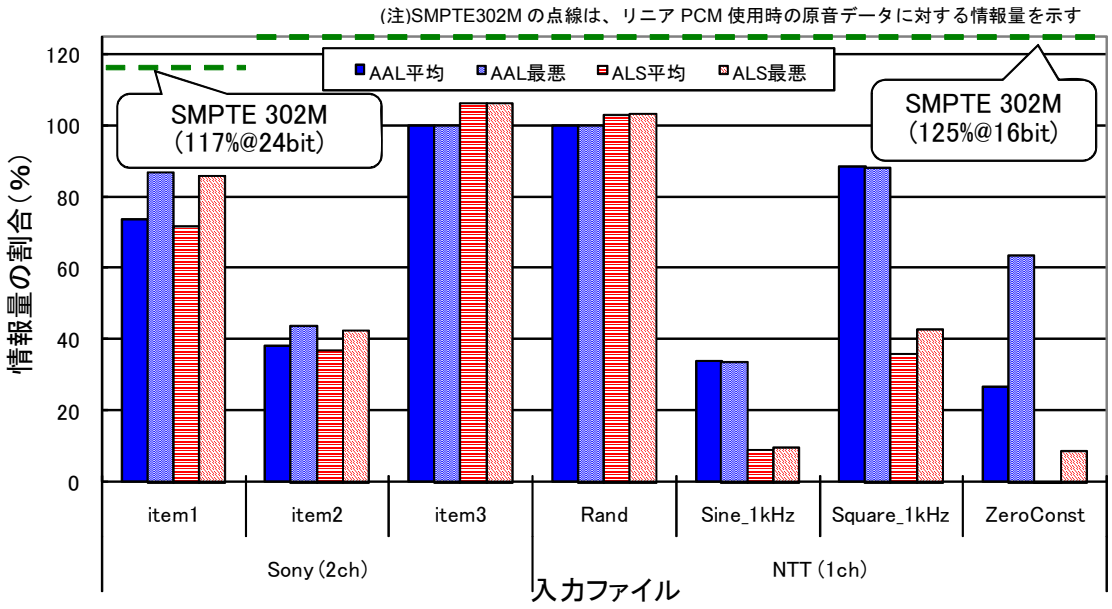


図 5.6-1(e)　独自音源（48kHz）

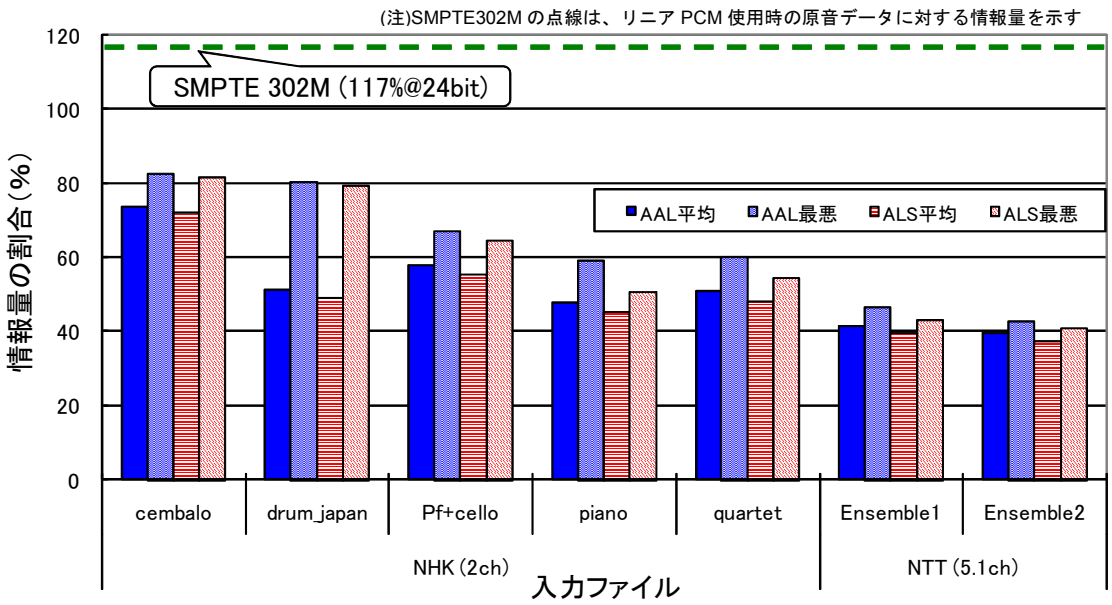


図 5.6-1(f)　独自音源（96kHz）

(3) 一音源内でのフレーム毎の圧縮性能比較

今回測定した音源のうち、典型例として 2 例（SONY 音源：室内楽、NHK 音源：和太鼓）について、音声フレーム毎の圧縮性能を以下にグラフで示す。

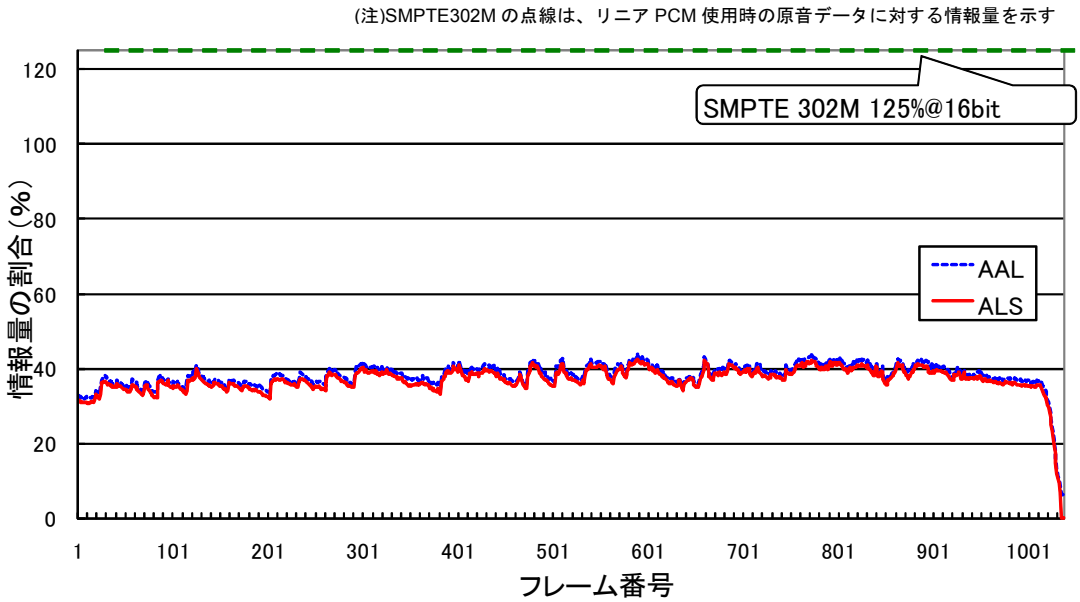


図 5.6-2(a)　フレーム毎の圧縮性能（室内楽：48kHz、16bit、2ch）

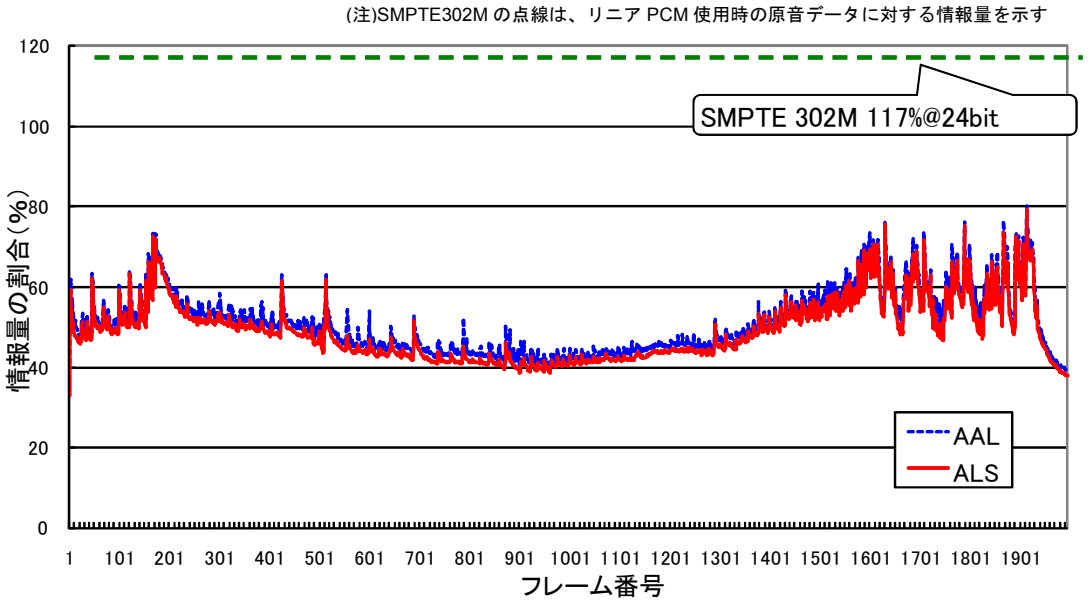


図 5.6-2(b)　フレーム毎の圧縮性能（和太鼓：96kHz、24bit、2ch）

5.7　ロスレス音声符号化・TS 生成実証実験報告

〇目的

ロスレス音声符号化方式である MPEG-4 ALS 符号化および AAL for broadcasting で音声信号をロスレス圧縮し､MPEG-2 システム準拠の TS を生成できることを確認する。 MPEG-4 ALS の詳細は付録 1、AAL for broadcasting の詳細は付録 2 を参照。

〇実施日

平成 20 年 6 月 13 日

〇参加者

浦野（音声符号化方式作業班主任、日本テレビ）、ソニー、ＮＴＴ、ＮＨＫ

〇実験場所

ARIB 会議室

〇実験項目

① MPEG-4 ALS で符号化されたビットストリームを含む TS の生成

MPEG-4 ALS で圧縮されたビットストリームによる MPEG-2 TS の生成が可能であることを下記により確認した。

- TS 用のエディタでストリーム内容確認
- 専用ソフトウェアでリアルタイム多重分離、復号、再生

② AAL for broadcasting で符号化されたビットストリームを含む TS の生成

AAL for broadcasting で圧縮されたビットストリームによる MPEG-2 TS の生成が可能であることを下記により確認した。

- TS 用のエディタでストリーム内容確認
- 専用ソフトウェアでリアルタイム多重分離、復号、再生

〇TS 作成付帯デモンストレーション　（MPEG-4 ALS のみ、付録 3 参照）

- ロスレス圧縮符号化音声と映像（H264/AVC）との TS レベルでの多重化と再生
- 音声を優先した映像との統合ビットレート制御による TS レベルでの多重化と再生

〇まとめ

ロスレス音声符号化方式による圧縮ビットストリームにより MPEG-2 システム準拠の TS の生成を確認した。また付帯デモンストレーションにより、音声のロスレス符号化の有効利用についての議論の具体化が喚起された。

付録１　MPEG-4 ALS で符号化されたビットストリームを含む TS の生成実験

〇目的

MPEG-4 ALS 符号化で圧縮し、圧縮ビットストリームにより MPEG-2 システム準拠の TS を生成できることを確認する。

〇符号化パラメータ

フレームサイズ：　8008 サンプル/フレーム

線形予測次数最大 15 次、ブロック分割最大 3 段、LTP、MCC/JS

〇実験結果

本実験では、まず 5 種類の音声信号（いずれも 48kHz,24bit,5.1ch）について ALS 圧縮性能を事前に測定した。その後、各音声信号における最悪ビットレートを固定的に割り当てて、TS を作成した。実験に用いた音声信号の ALS 圧縮性能を表に示す。

表　実験に用いた 5.1 マルチチャネル音声信号の MPEG-4　ALS 圧縮性能

音声入力信号 サラウンド収録 48kHz/24bit/5.1ch		PCM レート [Mbps]	再生時間 [sec]	平均 [Mbps]	最悪※ [Mbps]	標準 偏差
1	スポーツ（相撲）	6.912	326.5	3.939	4.614	0.373
2	オーディオドラマ（「浅間」）		354.8	3.815	5.120	0.489
3	音楽インドラマ（ジャズ）		271.1	4.304	5.159	0.579
4	音楽（パイプオルガン）		571.3	4.026	4.628	0.322
5	音楽インドラマ（ドラマ音楽）		258.3	3.688	5.101	0.927

※最悪レートは 8008 サンプル／フレーム×3 フレーム(約 500ms)を単位とした時の最悪値

作成した MPEG-2 TS の妥当性は、机上およびエンコーダ・多重化ツールの出力ログ、TS エディタによる内容の分析により確認した。さらに上記生成された TS を、PC を用いてソフトウェアによりリアルタイムで多重分離し、再生可能であることで TS の妥当性を確認した。

〇まとめ

MPEG-4ALS 符号化方式で音声信号をロスレス圧縮し、MPEG-2 システム準拠の TS を生成し、机上およびエンコーダ・MUX ツールの出力したログおよび TS エディタによる内容確認によって TS の妥当性の検証と、PC を用いた TS のリアルタイムに多重分離再生する実験により、MPEG-4　ALS を用いた TS の生成を確認した。

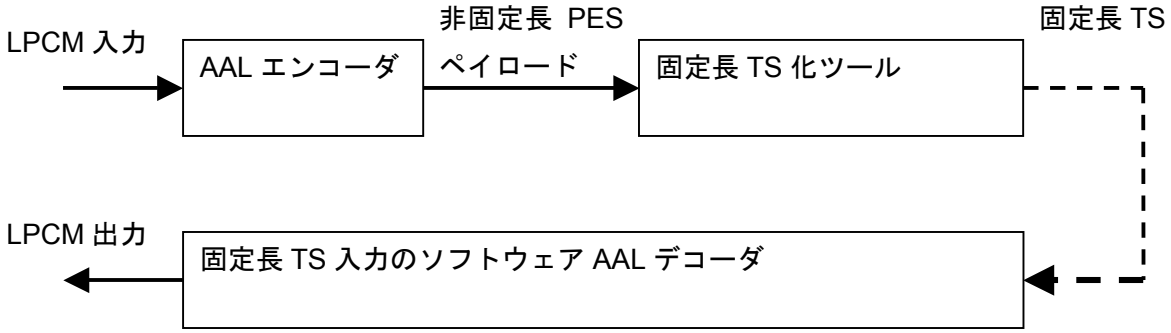
付録 2　ATRAC Advanced Lossless for Broadcasting の TS 作成実験

〇目的

高度衛星デジタル放送方式のロスレス音声符号化方式の検討に関連して、ATRAC Advanced Lossless for Broadcasting 符号化方式で音声信号をロスレス圧縮し、MPEG-2 システム準拠の TS を生成できることを確認する。

〇実験概要

ソフトウェア AAL エンコーダ、固定長 TS 化ツール、固定長 TS 入力のソフトウェア AAL デコーダを用いてＴＳの生成・復号・再生を模擬する。これにより、エンコーダ入力とデコーダ出力の一致検証を行い、ロスレスで復号できていることを確認する。このとき、エンコーダ入力をプリスキャンして最悪フレームサイズを算出しておき、この最悪フレームサイズに基づき固定長 TS 化に必要な NULL パケットを挿入して固定レート化を実現する。



〇符号化器

概要

－ AAL　(2ch)

－ デコーダの音声遅延用バッファ量　約 42.67msec に設定

〇実験に用いた音源

－ 48 kHz, 16 bit, 2ch の音楽信号

〇参考：実験に用いた 2 チャネル音声信号の AAL の圧縮性能

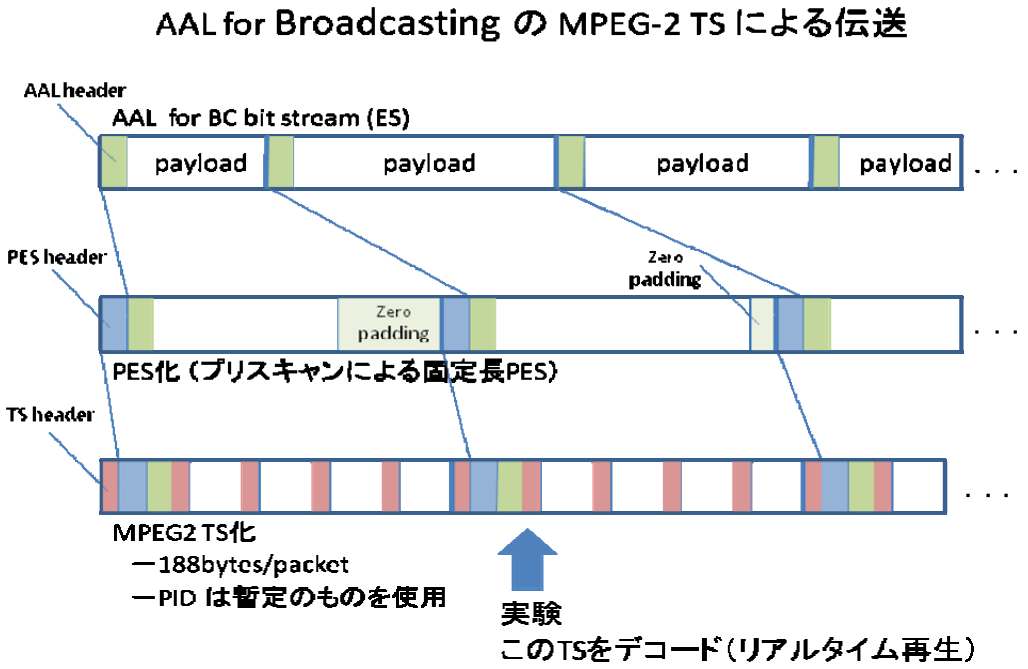
＜実験方法＞

音声入力信号	48kHz/16bit/2ch
音声符号化方式	ATRAC Advanced Lossless for Broadcasting 音声符号化
符号化パラメータ	フレームサイズ 2048 サンプル/フレーム 線形予測次数最大 8 次
使用機器	Windows PC

＜実験結果＞

音声入力信号 2 チャネルステレオ収録 48kHz/16bit/2ch		PCM レート [Mbps]	再生時間 [sec]	平均 [Mbps]	最悪※ [Mbps]
1	Sony item2	1.536	44.3	0.586	0.676
2	Sony item4		10.0	0.489	0.992

※最悪レートは 2048 サンプル／フレーム(約 42.66ms)を単位とした時の最悪値



〇まとめ

ATRAC Advanced Lossless for Broadcasting 符号化方式で音声信号をロスレス圧縮し、MPEG-2 システム準拠の TS を生成し、PC を用いた TS のリアルタイム再生する実験により、ATRAC Advanced Lossless for Broadcasting を用いて TS の生成が可能であることを確認した。

付録 3　MPEG-4 ALS と AVC/H.264 映像符号化との TS 作成、多重化、再生デモンストレーション

〇目的

TS 作成実験に付随して、映像との多重化 TS を作成し、実時間再生した。

〇実施項目

① ロスレス圧縮符号化音声と映像（H264/AVC）との TS レベルでの多重化と再生

② 音声を優先した映像との統合ビットレート制御による TS レベルでの多重化と再生

〇符号化・伝送パラメータ

- TS RATE　17Mbps
- AUDIO ES RATE　6.912Mbps－ALS 圧縮余剰[注 1)] (MPEG-4 ALS 準拠)
- VIDEO ES RATE [注 2)]　8.775Mbps＋ALS 圧縮余剰[注 1)] (H.264 4:2:0 フォーマット)
- 映像エンコード+デコードの遅延量 約 500msec に設定
- デコーダの音声遅延用バッファ量　約 500msec に設定

注 1）ALS 圧縮余剰＝入力 PCM レート（6.912Mbps）－ALS 圧縮後のレート

注 2）VIDEO ES RATE の数値はターゲットレート

〇デモに用いた音源・画像

- 音楽信号 5 種類（48 kHz, 24 bit, 5.1 ch）
- ハイビジョン・システム評価用標準動画像（1920x1080, 16:9）

〇デモ内容

① ロスレス圧縮符号化音声と映像（H264/AVC）との TS レベルでの多重化と再生

事前収録放送（完パケ）を想定して、当該プログラムにおける ALS 圧縮後の最悪レートに関する情報が事前に得られたとして、最悪レート相当を音声符号化側に割り当て、残りを映像符号化側に固定的に割り当てて多重化した TS を作成する。

例えば、音声信号「4. 音楽（パイプオルガン）」の場合では、ファイル全体の ALS 圧縮後の最悪レートが 4.628Mbps であったことから、最悪レートをもとに計算した ALS 圧縮余剰は 6.912-4.628=2.284Mbps となった。これにより映像の ES レートは 8.775Mbps＋2.284Mbps＝11.059 Mbps とした。

以上の手順により、ソフトウェアにより ALS ストリーム（音声）および H.264/AVC ストリーム（映像）をオフラインエンコードし MPEG-2 TS に多重化した。さらに上記生成された TS を、PC を用いてソフトウェアによりリアルタイムで多重分離し、再生可能であることを示した。

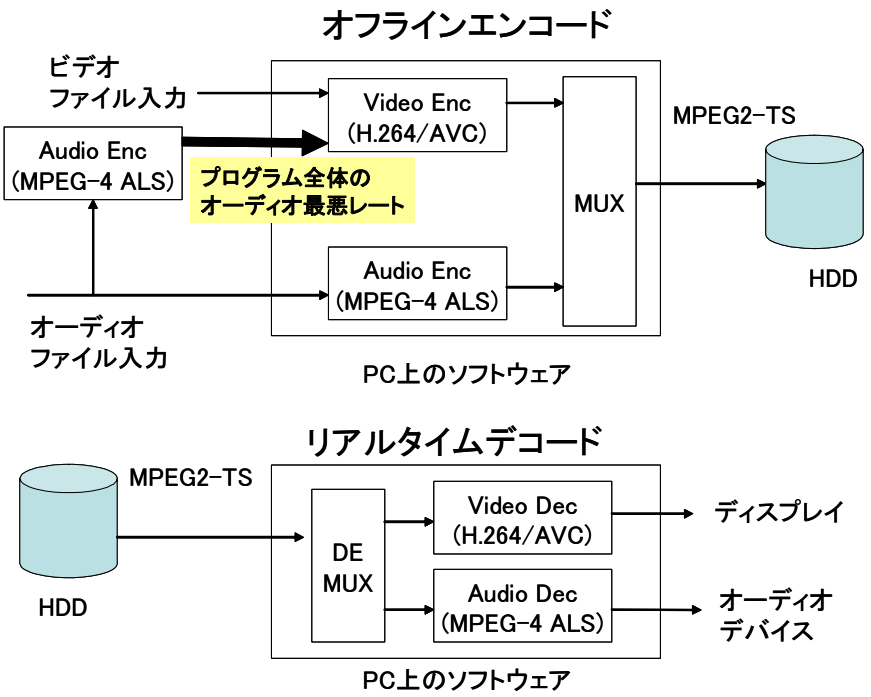


図 5.7- 1　デモシステム（余剰ビットレートをプログラム内固定割り当て）

②音声を優先した映像との統合ビットレート制御による TS レベルでの多重化と再生

事前に測定した入力音声信号に対する ALS 圧縮性能のデータ（フレーム毎の圧縮率）を用いて、500ms の GOP 単位ごとに、直前の GOP での ALS 圧縮後の余剰ビットレートを映像に追加で動的に割当てた(GOP 単位で映像のターゲットレートを制御)。これによって実効的には PCM レートから ALS 圧縮後の平均レートを差し引いた余剰ビットレート相当を映像に追加で割り当てることができる。例えば、「4. 音楽（パイプオルガン）」では、ファイル全体の ALS 圧縮後の平均レート（実測値）が 4.026Mbps であったことから、ALS 圧縮余剰は 6.912-4.026=2.886Mbps であった。余剰ビットレートを映像に動的に割り当てることで､映像 ES の平均レートはおよそ 8.775Mbps＋2.886Mbps＝11.661Mbps となった。

以上の手順により、ソフトウェアにより ALS ストリーム（音声）および H.264/AVC ストリーム（映像）をオフラインエンコードし MPEG-2 TS に多重化した。さらに上記生成された TS を、PC を用いてソフトウェアによりリアルタイムで多重分離し、再生可能であることを示した。

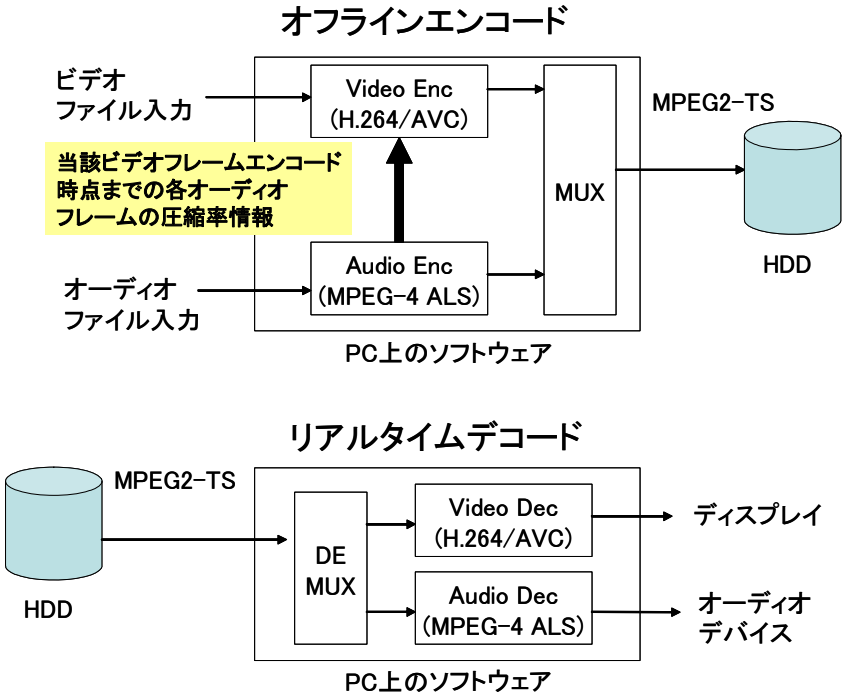


図 5.7-2　デモシステム（余剰ビットレートをフレーム単位で動的割り当て）

下記の３種の条件の TS を PS3 で多重分離し、映像部分だけを実時間再生した。

(1) リニア PCM音声符号化相当 (SMPTE302M, H.264/AVC 固定レート)
(2) ①の映像への固定割り当て (MPEG-4 ALS, H.264/AVC 固定レート)
(3) ②の映像への動的割り当て (MPEG-4 ALS, H.264/AVC 可変レート)

表 5.7-1　レート割り当ての実例（※音声入力ファイル「4.音楽（パイプオルガン）」の例）

音声 ALS のレートは実測値、映像のレートはターゲットレート

音声フォーマット (48kHz,24bit,5.1ch)	(1) リニア PCM 相当 (SMPTE302M)	8.064 Mbps (ES レート)
	(2) MPEG-4 ALS 固定レート	4.628 Mbps※ (プログラム毎の最悪レートに設定)
	(3) MPEG-4 ALS 可変レート	平均約 4.026Mbps※ (6.912－[ALS 圧縮余剰] Mbps)
映像フォーマット (1920x1080 (1080i)、 16:9)	(1) 固定レート (リニア PCM 相当に対応)	7.623 Mbps (ES レート)
	(2) 固定レート (ALS 固定レートに対応)	11.059 Mbps※ (音声圧縮後の余剰を割当て)
	(3)　可変レート (ALS 可変レートに対応)	平均約 11.661 Mbps※ (8.775＋[ALS 圧縮余剰] Mbps)
伝送フォーマット	MPEG-2　TS	