

一般社団法人電波産業会
デジタル放送システム開発部会
音声符号化方式作業班

地上デジタル放送方式高度化の音声符号化方式に関する中間報告（案）

提案方式（検討中）の概要

符号化音声入力フォーマット

項目	値
標本化周波数	48kHz
入力量子化ビット数	16bit 以上
対応する音声信号	オブジェクトベース チャンネルベース
最大入力音声チャンネル数	56ch
備考	48kHz 以外の標本化周波数は規定しない。

音声符号化方式

	MPEG-H 3D Audio	AC-4
準拠規格	ISO/IEC 23008-3	ETSI TS 103 190-2
プロファイル	Baseline	規定なし
ビットストリーム形式	MHAS 形式	raw_ac4_frame 形式

地上デジタルテレビジョン放送の高度化に関する音声符号化方式について、情報通信審議会放送システム委員会により示された要求条件をベースに提案方式の検討を行った。これまでの音声サービスに加え、多様なサービスを実現可能なオブジェクトベースを利用した音声サービスを想定し、提案方式と制約技術について検討状況を取りまとめた。

1 提案する技術方式

1.1 音声入力フォーマット

- (1) 音声信号の標本化周波数は、48kHz とする。
- (2) 入力量子化ビット数は 16 ビット以上とする。
- (3) 対応する音声信号はチャンネルベースと、オブジェクトベースとする。
- (4) 最大入力音声チャンネル数は、最大 56 チャンネルとする。

(理由)

- ・ 高度広帯域衛星デジタル放送のフォーマットを考慮した。
- ・ 現状の実運用動向を鑑み、標本化周波数は 48kHz のみとした。
- ・ 48kHz 以外の標本化周波数については、実運用動向から放送局設備へのインパクトが大きいこと、当面サービスが想定されていないことを鑑み、規定しないこととした。
- ・ 現状の実運用動向を鑑み、量子化ビット数は 16 bit 以上とした。
- ・ チャンネルベースに加え、多様な音声サービスを実現可能なオブジェクトベースを採用した。
- ・ 22.2 マルチチャンネル音響に対応し、かつ、オブジェクトベース音声を用いた差し替え音声によるサービスを考慮して、MPEG-H 3D Audio で規定されているレベル 4^{注1} で規定された最大入力音声チャンネル数^{注2}とした。

注 1 レベル：MPEG で規定されているデコーダの回路規模を決めるためのパラメータ。標本化周波数、最大入力信号数、最大同時デコード数、最大出力信号数が規定されている。

注 2 最大入力音声チャンネル数：これまでのチャンネルベースの放送サービスにおいて、二カ国語放送などの副音声は別のストリームで実施されてきたが、オブジェクトベースでは副音声の音声信号も一つのストリームとして伝送することを想定するために、最大入力音声チャンネル数は実際に再生される音声信号数よりも多く見積もられている。

1.2 音声符号化方式

(1) MPEG-H 3D Audio 規格 (ISO/IEC 23008-3)

項目	制約条件
プロファイル	Baseline プロファイル
ビットストリーム形式	MHAS 形式

(理由)

- ・ オブジェクトベースによる音声サービスが実現可能な符号化方式を採用した。
- ・ MPEG-2/4 AAC に比べて高効率な音声符号化方式であることを主観評価実験により確認した。
- ・ オブジェクトベースによる音声サービスを実施可能な最も回路規模が小さいプロファイルを選択した。

(2) AC-4 規格 (ETSI TS 103 190-2)

項目	制約条件
プロファイル	規定なし
ビットストリーム形式	raw_ac4_frame 形式

(理由)

- ・ オブジェクトベースによる音声サービスが実現可能な符号化方式を採用した。
- ・ MPEG-2/4 AAC に比べて高効率な音声符号化方式であることを主観評価実験により確認した。

2 要求条件と提案方式の整合性

地上デジタルテレビジョン放送の高度化における音声符号化方式に係る要求条件と提案する技術方式の整合性は以下の通りである。

(1) システム

要求条件	提案方式による適合性 (MPEG-H 3DA)	提案方式による適合性 (AC-4)
HDTV を超える高画質・高音質・高臨場感サービスを基本として、多様な画質のサービス等を提供できること。	音の位置を自由に座標でも配置できる高臨場感な3次元立体音声サービスに対応。オブジェクトベースに対応した多様なサービスが可能。	同左。
多様で柔軟な高機能サービスを提供できること。	オブジェクトベースに対応した多様なサービスが可能。	同左。
高齢者、障害者等様々な視聴者向けの放送サービスについても考慮すること。	オブジェクトベースに対応した多言語サービス、視覚及び視聴覚障害者用副音声サービス、高齢者用音声サービスを高効率に実現可能。	同左。
視聴者に違和感を与えない程度の映像・音声の遅延差であること。	システムレイヤ構成により実現可能。	同左。
高い実時間性を実現するため、できるだけ遅延時間を短くすること。	提案した符号化方式は短いブロック長を持つ。	同左。
放送の要件に応じて伝送パラメータの選択や組み合わせの指定を行うことができ、またそれに合わせて受信機制御ができる方式とすること。	音響メタデータにより受信機の制御が可能。	同左。
送出する映像、音声、デー	多くの音声モードやビット	同左。

タのフォーマットやビットレート、チャンネル数等を任意に選択、変更できること。	レートに対応。	
諸外国も容易に導入できるシステムとなるよう考慮すること。	提案方式が他国で導入済。 (提案した符号化音声入力フォーマットとは異なる仕様が主流)	同左。

(2) 放送品質

要求条件	提案方式による適合性 (MPEG-H 3DA)	提案方式による適合性 (AC-4)
放送サービスに応じて音声のフォーマットやビットレートを変更できること。	各種音声フォーマット・ビットレートに対応。	同左 ^注 。
UHDTV サービスに対応した、高音質・高臨場感な音声サービスに適した音質が望まれることを考慮し、できるだけ高い音質を保つこと。	最大22.2chの音声モードにより高臨場感音声サービスに対応。 主観評価実験により MPEG-2/4 AAC よりも高効率に符号化可能であることを確認。	

(3) 技術方式

要求条件	提案方式による適合性 (MPEG-H 3DA)	提案方式による適合性 (AC-4)
国際標準と整合した方式を用いること。	ISO/IEC 規格準拠。	ETSI 規格準拠。
高効率かつ高音質な符号化方式であること。	AAC よりも高効率に符号化可能。	同左。
多チャンネル音声放送をはじめとした、さまざまなサービス要件に柔軟に対応できる符号化方式である	最大 22.2ch に対応。 オブジェクトベースに対応した多様なサービスが可能。	最大 22.2ch に対応 ^注 。 オブジェクトベースに対応した多様なサービスが可能。

こと。		
現行設備や受信機への負担等を考慮して選定される種々の音声入力フォーマットに対応できること。	多くの入力フォーマットに対応。	同左。

注：参考3を参照

参考1 オブジェクトベースを用いた想定サービス

多様なサービスを実現するために、これまでのチャンネルベースに加えオブジェクトベースによる放送サービスを想定した。チャンネルベースとの違いとオブジェクトベースによる想定サービスについて補足する。

チャンネルベース音響

放送局の番組制作では、各音声素材（ナレーションのようなダイアログ、背景音、効果音等）を取音して最終的に一つ（または複数）のチャンネル構成（例えば、2ch ステレオや 5.1ch）にまとめて放送される。受信側ではチャンネル構成（2ch の場合は左チャンネルと右チャンネル）に対応するスピーカから再生することにより、制作意図のまま番組音声を楽しめる。このように制作時のチャンネル構成と受信機側の再生チャンネル構成が同一であることを前提に受信側のスピーカを直接ドライブする信号を制作/伝送する方式をチャンネルベース音響と呼ぶ。

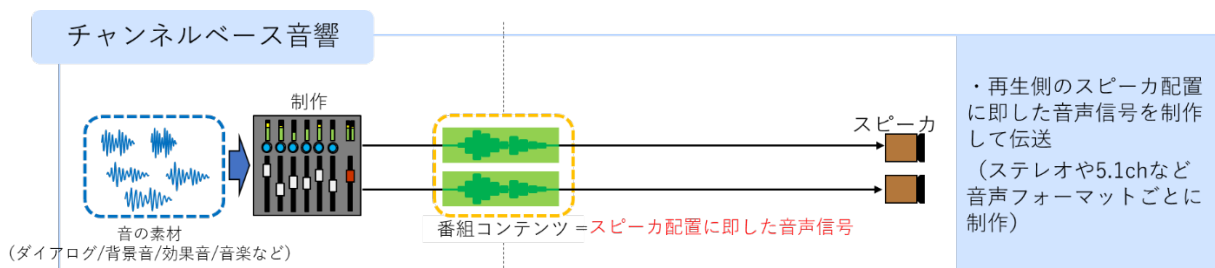


図 1 チャンネルベース音響

オブジェクトベース音響 (OBA)

オブジェクトベース音響ではナレーションや背景音、音楽や効果音を各々個別に位置やレベルなどを記述した音響メタデータとともに伝送し、受信側で音響メタデータを基に番組音声を再構成してスピーカをドライブする信号を出力する方式をオブジェクトベース音響と呼ぶ。

オブジェクトベース音響システムでは、ナレーションや背景音、音楽や効果音を個別に受信できるため、家庭のスピーカ配置へ最適化したり、ナレーションなどダイアログの音量を個別に調整することで聞き取りやすくしたり、多言語に対応することで日本語が非ネイティブの方に情報を伝えるなど、試聴環境や視聴者の好みに合わせた番組音声のサービスをきめ細かく提供できる。

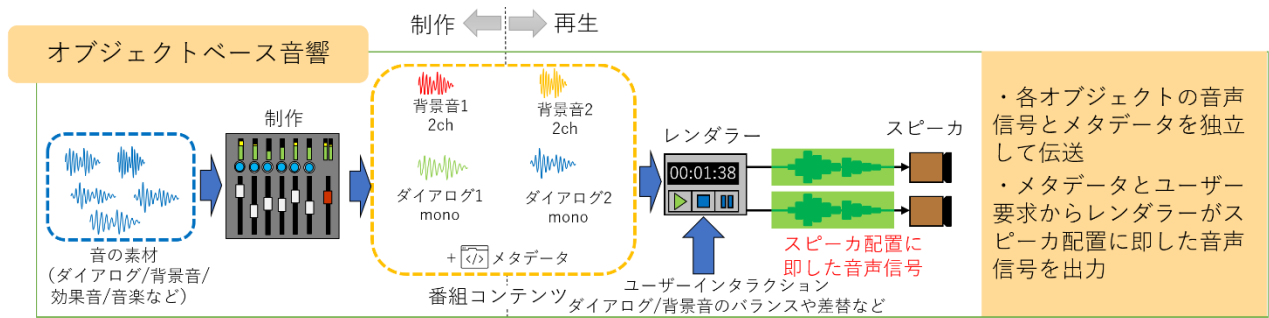


図 2 オブジェクトベース音響



図 3 オブジェクトベース音響で想定される音声サービス

参考2 最大入力音声チャンネル数の考え方について

提案技術では最大入力音声チャンネル数 56ch とした。これは、MPEG で回路規模を規定するためのパラメータであるレベル 4 に対応したものである。放送開始時期が明確になることで各放送事業者が必要とする最大入力音声チャンネル数が具体化するものと思われる。現時点では超高精細度テレビジョン放送(新 4K/8K)の音声符号化方式で採用されている、22.2 マルチチャンネルに対応し、かつ参考1のオブジェクトベースのサービスを想定した場合の最大入力音声チャンネル数について検討を行った。

表1にオブジェクトベースによるコンテンツ事例と入力チャンネル数の関係をまとめた。いくつかのコンテンツによっては、音声信号の合計数は56chを超える場合も想定されることが分かった。実運用を考慮していないことを鑑み、MPEGで規定されているレベル4で制約を与えることが妥当であるとした。

表1 オブジェクトベースのコンテンツ事例と入力チャンネル数

コンテンツ名	サービス事例	オブジェクト (Objects)	背景音 (MCA)	OBA	Total (Stereo×2+5.1×2)
国際放送21言語	国際放送21言語		42ch(2ch) : Stereo×42カ国語	42ch(2ch)	42ch
	OBAによる国際放送21言語	22ch(2ch) : 21カ国語 (mono)、緊急放送	2ch(2ch) : 背景音	24ch(4ch)	28ch(+Stereo x 2) 70ch(+Stereo x 42)
スポーツ番組 (解説)	異なる解説で聞きたい	6ch(2ch) : 解説 (普通、英語、裏トーク、専門家、音声解説)、緊急放送	24ch(24ch) : 背景音 (22.2)	30ch(26ch)	46ch
ドキュメンタリ (吹き替え)	出演者が日本語を話さないときに吹き替えで聞きたい、本人の声で聞きたい	7ch(2ch) : ダイアログ (日本語、英語、吹き替え、オリジナル、裏トーク、音声解説)、緊急放送	24ch(24ch) : 背景音 (22.2)	31ch(26ch)	47ch
音楽	歌手グループの誰か一人の歌声を聞きたい	11ch(3ch) : 歌声 (全員、歌手A～D : Stereo)、緊急放送	24ch(24ch) : 伴奏 (22.2)	35ch(27ch)	51ch
音楽	アレンジ違いの楽曲 (伴奏) を聞きたい。ナレーションはOFFにしたい。	15ch(14ch) : メインとなる楽器・演奏 (7.1.4)、ナレーション (日本語、英語)、緊急放送	36ch(12ch) : アレンジの異なる伴奏3種類 (7.1.4)	51ch(26ch)	67ch
スポーツ番組 (Home & Away)	スポーツ番組で応援しているチームよりの背景音や解説を聞きたい (+裏トーク+二カ国語)	7ch(2ch) : Home (日本語、英語、裏トーク、無)、Away (日本語、英語、裏トーク、無)、緊急放送	48ch(24ch) : Home (22.2背景)、Away (22.2背景)	55ch(26ch)	71ch

参考3 AC-4の要求条件の適合性に関する補足

欧米各国で採用されている AC-4 Level 3 で 22.2ch を符号化するためには、聴感的に重要度の低いチャンネルを適応的にまとめて、最大 17.1 個の座標付きモノラル信号として符号化する。復号側では、これを 22.2ch にレンダリングして出力を得る。ARIB での主観音質評価試験では良好な結果を得た。

更なる高音質化などのために、22.2ch をそのまま符号化する必要がある場合、同時最大デコードチャンネル数が AC-4 Level 3 では不足する。放送開始時期が明確になることで各放送事業者が必要とする最大入力音声チャンネル数や同時最大デコードチャンネル数などが具体化するものと思われ、これに適合する形で、現在の AC-4 規格上で"Reserved"となっている Level 4 として定義する必要がある。これは、符号化・復号アルゴリズムはそのままに、対応する最大チャンネル数の上限を明確化するものである。