

# 音声符号化方式に関する 主観評価実験報告(中間報告)

2021年10月7日

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

# 音声符号化方式検討の目的

## 地上デジタル放送高度化に関わる音声符号化方式の比較検討

- 次世代地上デジタル放送の音声符号化方式として以前より研究開発・技術試験事務等で検討中、あるいは2020年3月の提案募集に対して応募があった以下の4方式に対して機能および性能の比較（オブジェクト符号化の比較を含む）を行う。

1. MPEG-4 AAC

2. MPEG-H 3D Audio

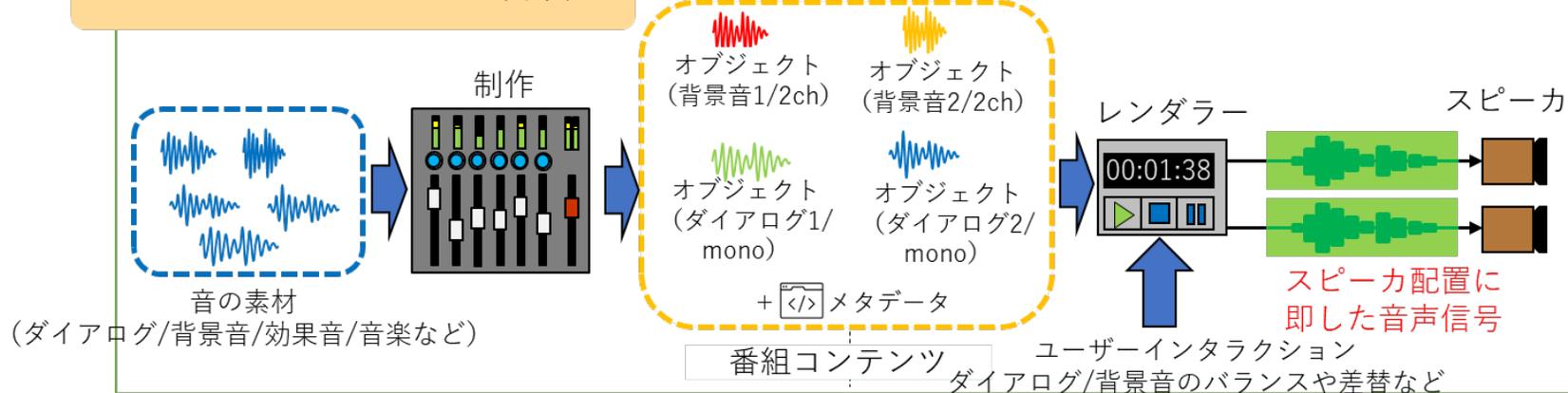
3. Enhanced AC-3

4. AC-4

# オブジェクトベース音響 (OBA) とサービスイメージ

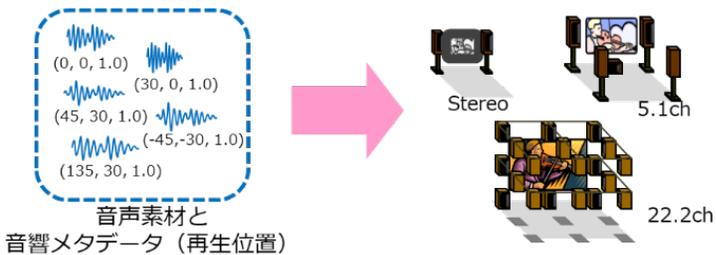
制作 ← 再生

## オブジェクトベース音響

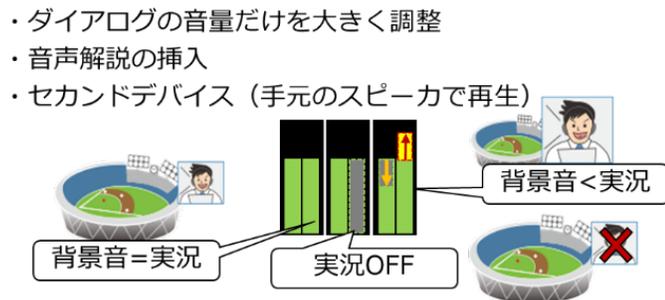


- ・各オブジェクトの音声信号とメタデータを独立して伝送
- ・メタデータとユーザー要求からレンダラーがスピーカ配置に即した音声信号を出力

## 再生環境 (スピーカ配置) への最適化



## アクセシビリティ改善



## 視聴者の好みに合わせたコンテンツ

- ・多言語、裏トークなどの多様な副音声 (ダイアログ差替え)
- ・ホーム&アウェイ

- ・視点切换
- ・アレンジ違いの音楽 (背景音差替え)
- ・ローカル局や課金サービスなどによる音声素材の追加



# 各符号化方式の検討の方向性

	AAC	MPEG-H 3DA	Enhanced AC-3	AC-4
OBA対応	×	○	○	○
同時再生数 (デフォルト)	MPEG-2 : 8 (ex.7.1ch) MPEG-4 : 24 (ex.22.2ch)	レベル3 : 16 (ex. 7.1.4+4obj)、 24 (オブジェクトのみ) レベル4 : 28 (ex. 22.2ch+4obj)	16 (ex. 7.1.4+4obj)	レベル3 : 18 (ex.7.1.4+6obj)
圧縮効率 * AAC以外の数値は参考情報 (勧告ITU-R BS.1548)	144kbps (ステレオ) / 320kbps (5.1ch) / 1.4Mbps (22.2ch) : MPEG-4	768kbps (22.2ch)	192kbps (ステレオ)	96kbps (ステレオ) / 192kbps (5.1ch) / 288kbps (7.1.4)
				
検討の方向性	<ul style="list-style-type: none"> <li>帯域の利用効率の観点から、参考情報とする。</li> </ul>	<ul style="list-style-type: none"> <li>22.2ch+objに対応しており、国内放送のすべての音声フォーマットに対応。</li> <li>帯域の利用効率が高い。</li> </ul>	<ul style="list-style-type: none"> <li>帯域の利用効率の観点から、参考情報とする。</li> </ul>	<ul style="list-style-type: none"> <li>帯域の利用効率が高い。</li> <li>放送要件として22.2chが採用され、22.2ch+obj対応が必要な場合、新規のレベル規定が必要。</li> </ul>

# 各符号化方式の主観評価実験

		AAC	MPEG-H 3DA	Enhanced AC-3	AC-4
主観評価実験	1. コア符号化の ビットレート	【実施】	【実施】	【実施】 ※ 22.2chを一度16chに変換しその後22.2chに戻す	【実施】 ※ 22.2ch+4obj対応のプロトタイプを使用
	2. レンダラーの方式比較	×	【今後実施】	【今後実施】	【今後実施】

# 主観評価実験 1 コア符号化のビットレート

- 目的

- 音声符号化方式4方式の符号化音の音質を主観評価実験により確認し、放送品質を満たすビットレートを導出

- ビットレート導出の考え方

- 視聴者がダイアログ制御などの機能を用いて音声オブジェクトの再生レベルを大きく変化させても放送品質を満たすように、**単一の背景音やモノオブジェクトだけで再生しても放送品質を担保するビットレート**を導出する。

OBAの所要ビットレート = (モノオブジェクトの所要ビットレート)  $\times i$  +  
( $N_1$ チャンネルの背景音の所要ビットレート)  $\times k_1$  +  
( $N_2$ チャンネルの背景音の所要ビットレート)  $\times k_2$  +  
... + (音響メタデータの所要ビットレート)

# 実験諸元

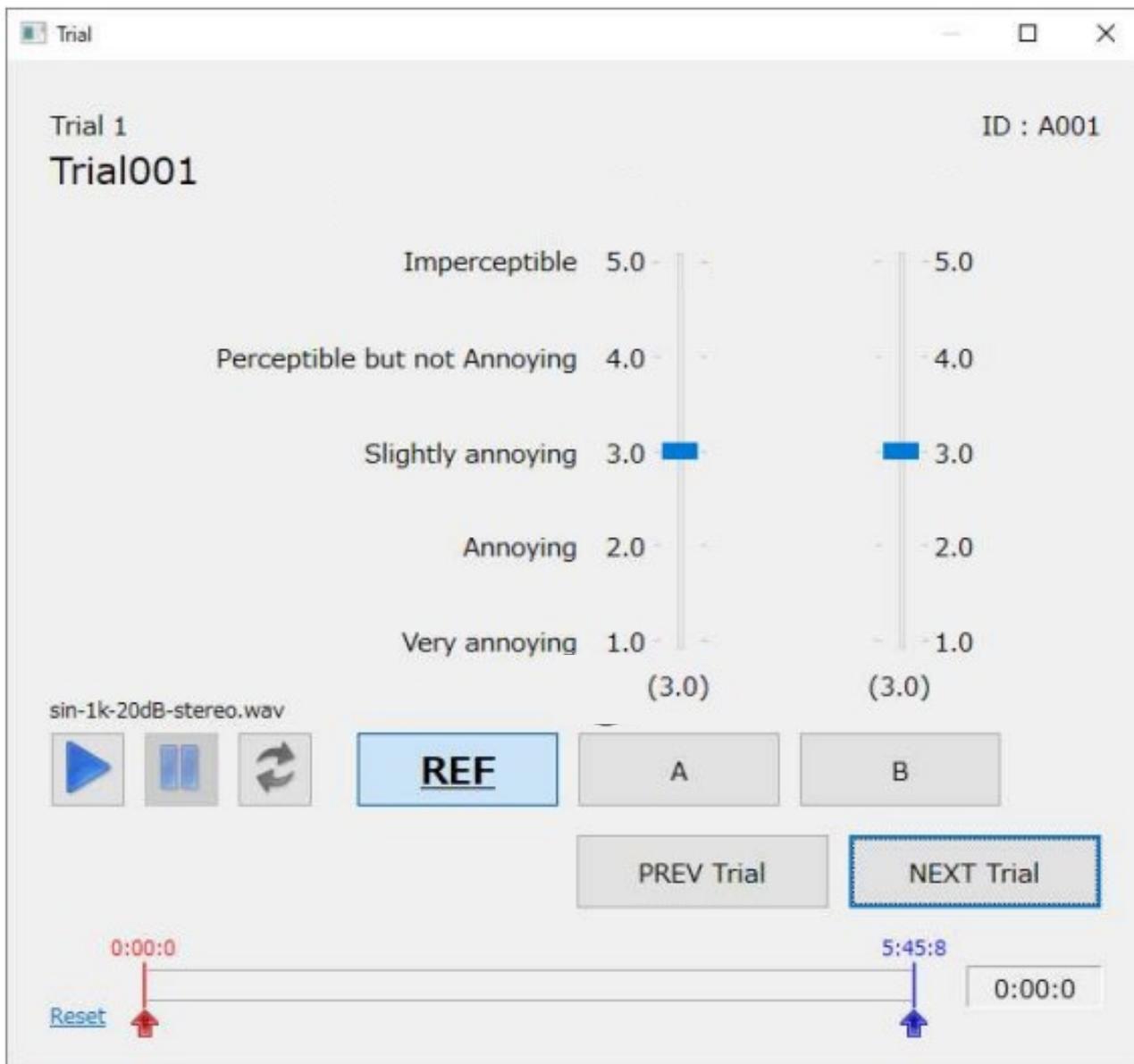
評価法	ITU-R BS.1116 隠れ基準付き三刺激二重盲検法
音声フォーマット	22.2マルチチャンネル音響 7.1.4 5.1chサラウンド ステレオ モノ
音源	各音声フォーマットごとに4音源を選出
符号化方式	4方式 (MPEG-4 AAC, MPEG-H 3DA, AC-4, Enhanced AC-3)
評定者数	音響専門家 34名 (ポストスクリーニングにより20名程度)
実験サイト	NHK技研、東京藝術大学 (ITU-R BS.1116準拠のリスニングルーム)
スケジュール	8月17日～9月28日 (現在分析中)

ポストスクリーニング：

評価結果から信頼度の低い評定者を除く処理

(例えば、原音に低い点をつける/でたらめに回答した分布と統計的な差がない等)

# ITU-R BS.1116 隠れ基準付き三刺激二重盲検法



- REF : 原音 (非圧縮)
- A/B : 一方に原音、もう一方に評価音 (圧縮音源)
- REFと比較しAとBの音源を評価

## 評価尺度

評点	REF(原音)との差が
5.0	差が分からない
4.0	差が分かるが、気にならない
3.0	差がわずかに気になる
2.0	差が気になる
1.0	差が非常に気になる

# 音源/ビットレート

音声フォーマット	使用音源
22.2マルチチャンネル音響	Applause (自然音)、Tenpura (自然音)、Etenraku (音楽)、MyKIngdom (ドラマ)
7.1.4	Ave Maria (音楽)、Choir at side (音楽)、Unfold (自然音)、Water burst (自然音)
5.1chサラウンド	Atami fireworks festival (自然音)、Applause (自然音)、Jazzquartet (音楽)、Glasses (自然音)
ステレオ	Glockenspiel (単楽器)、Feste Romane - La Befana (音楽)、German Narration - male (音声)、Harpsichord (単楽器)
モノ	Wind chime (単楽器)、Shamisen (単楽器)、Glockenspiel (単楽器)、English Narration - female (音声)

音声フォーマット	22.2マルチチャンネル音響			7.1.4			5.1chサラウンド			ステレオ			モノ		
	384	512	768	192	256	384	96	128	192	48	64	96	32	48	64
ビットレート [kbps]															
MPEG-4 AAC	X	X	0	X	X	X	X	X	0	X	X	0	X	X	0
MPEG-H 3DA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Enhanced AC-3	X	X	0	X	X	0	X	X	0	X	X	0	X	X	0
AC-4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

# 主観評価実験 2 レンダラー比較

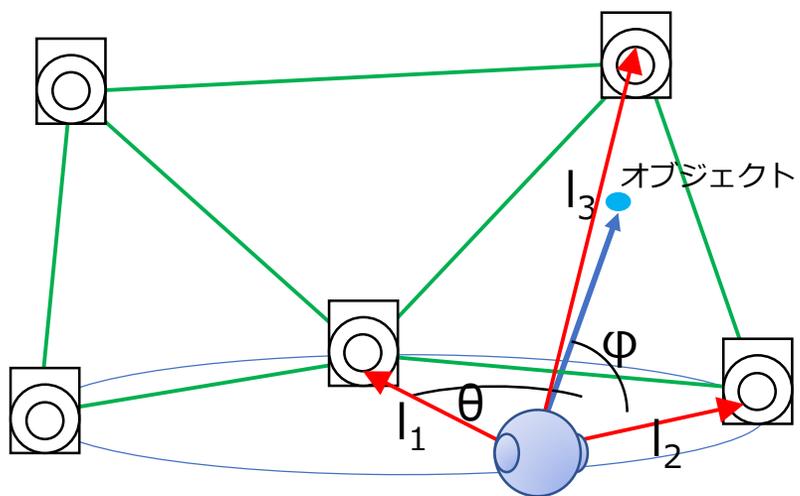
- 目的
  - レンダラーによる音質・音像定位精度・空間印象の違いを評価
- 検討課題
  - 座標系（極座標・直交座標）
  - パンニング則（VBAP／Triple-Balance）
    - 音の定位を変えるためのルール
  - チャンネル数変換
  - スピーカ位置補正

# レンダラー機能（音声信号のマッピング）

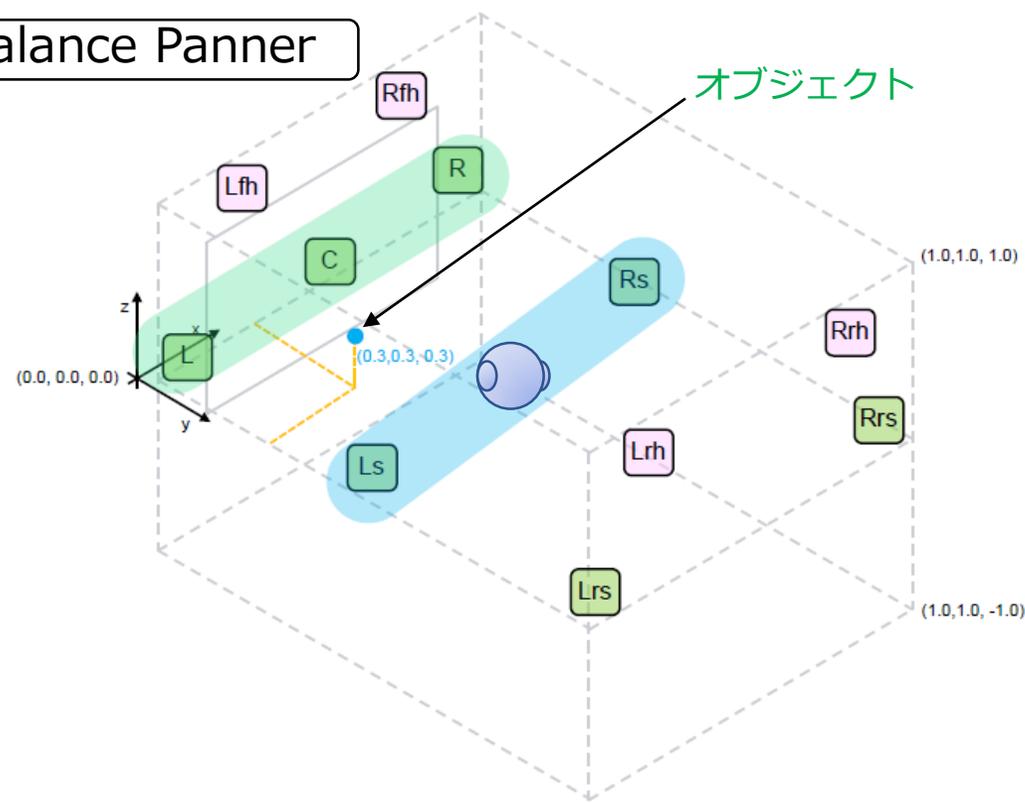
受信機側のスピーカ配置に即してダイアログなどのオブジェクトをマッピングする技術

- MPEG-H : **3D VBAP** (Vector Based. Amplitude Panning)
  - 極座標をベースとし、3スピーカによる領域に分割し各領域ごとに各スピーカーに分配するゲインをベクトル重み付けにより決定
- AC-4 : **Triple Balance Panner**
  - 直交座標をベースとし、オブジェクトの座標に対しx軸、y軸、z軸の各スピーカーに分配するゲインを決定

3D VBAP



Triple Balance Panner



# 検討スケジュール



## 主観評価実験スケジュールの詳細

- コア符号化のビットレート
  - 8～9月 実験
  - 10月 分析・まとめ
- レンダラー比較
  - 10月 計画立案
  - 11月 実験
  - 12月 分析・まとめ

## 最終報告構成案

### 背景と目的

- 音声符号化方式の標準化経緯
- 放送規格の採用状況
- 各国での採用状況

### 音響システムの概要

- チャンネルベース音響
- オブジェクトベース音響
- シーンベース音響

### 各音声符号化方式の概要

- 各音声符号化方式の技術比較
- 主観評価結果

# 參考資料

# 音声符号化方式検討 中間報告

2021.2.9

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

## 地上デジタル放送高度化に関わる音声符号化方式の比較検討

- 次世代地上デジタル放送の音声符号化方式として以前より研究開発・技術試験事務等で検討中、あるいは2020年3月の提案募集に対して応募があった以下の4方式に対して機能および性能の比較（オブジェクト符号化の比較を含む）を行う。

1. MPEG-4 AAC

2. MPEG-H 3D Audio

3. Enhanced AC-3

4. AC-4

## 新たな符号化方式が開発され圧縮率向上と高機能化を実現

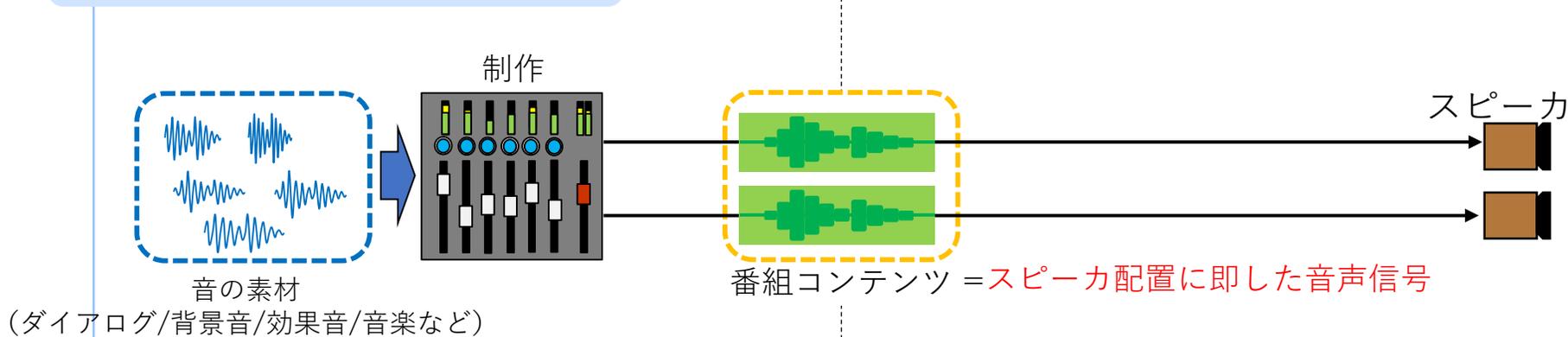
- 5.1chを超える **マルチチャンネル音声の高効率圧縮**
  - 例 22.2ch : 1.4Mbps (MPEG-4 AAC) ⇒ 768kbps (MPEG-H 3DA)
- **オブジェクトベース音響 (OBA)** を利用した音声サービス
  - ダイアログ\*1や背景音などの部品をそのまま圧縮・伝送し、視聴者の好みに応じて再生  
例：ダイアログ強調による高齢者の聞きやすいサービス
  - 受信側のスピーカ配置に即した音声信号に変換

\*1：対話を意味するが、  
テレビ/映画の台詞の意味で用いられる

	1990	2000	2010	2020	標準化	主な国内外の放送規格
MPEG-2 AAC		◆ 1997年規格化			MPEG	ARIB (地上デジタルテレビジョン放送/ BSデジタル放送ほか)
MPEG-4 AAC		◆ 2000年規格化			MPEG	ARIB (高度BSデジタル放送ほか)、DVB
MPEG-H 3DA			◆ 2015年規格化		MPEG	ATSC3.0、DVB
Enhanced AC-3		◆ 2005年規格化	◆ 2016年 オブジェクトベース対応		ETSI	ATSC、DVB
AC-4			◆ 2015年規格化		ETSI	ATSC3.0、DVB

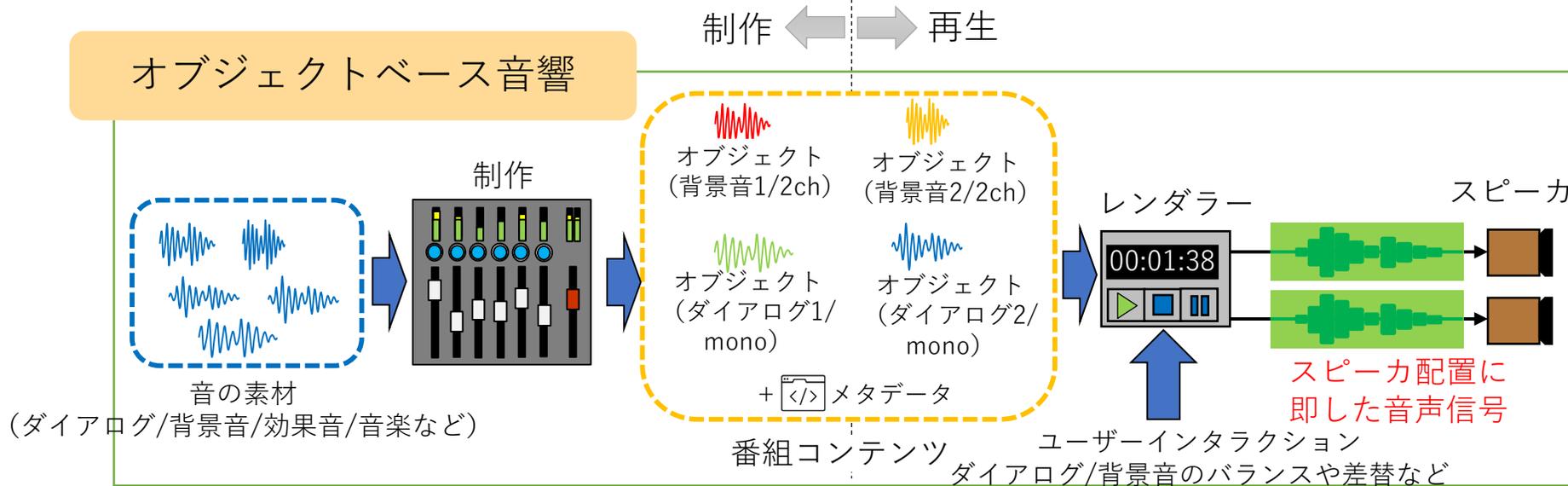
赤字はオブジェクトベース音響に  
対応した音声符号化方式

## チャンネルベース音響



- ・再生側のスピーカ配置に即した音声信号を制作して伝送  
(ステレオや5.1chなど音声フォーマットごとに制作)

## オブジェクトベース音響

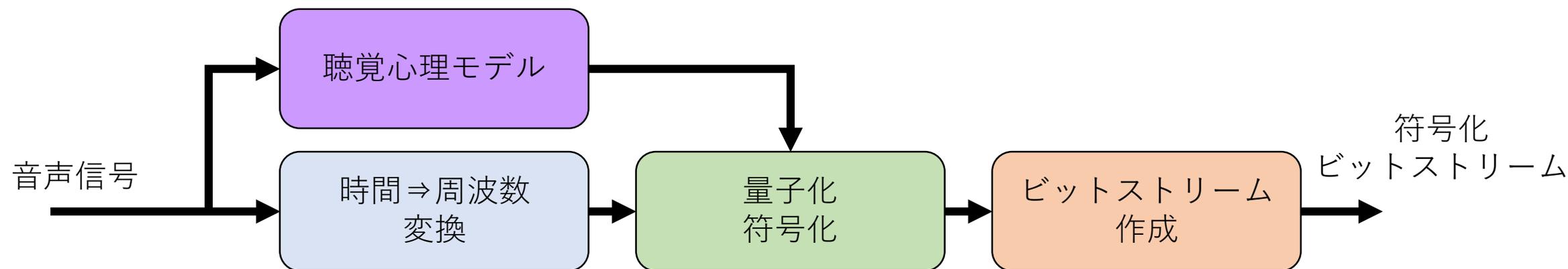


- ・各オブジェクトの音声信号とメタデータを独立して伝送
- ・メタデータとユーザー要求からレンダラがスピーカ配置に即した音声信号を出力

聴覚特性を利用した適応変換符号化方式（国内のデジタル放送用音声符号化方式として採用）

- 2チャンネル間の冗長性を削減したり、予測などの符号化ツール群を利用
- 標準でマルチチャンネル対応（MPEG-2：最大7.1ch、MPEG-4：最大22.2ch）
- 144kbps/ステレオ、1.4Mbps/22.2chで放送品質を実現（ARIBで評価実験を実施）
- **OBAは未対応**

聴覚特性を利用した音声符号化のブロック図



## MPEG-2 AACとMPEG-4 AACの違い

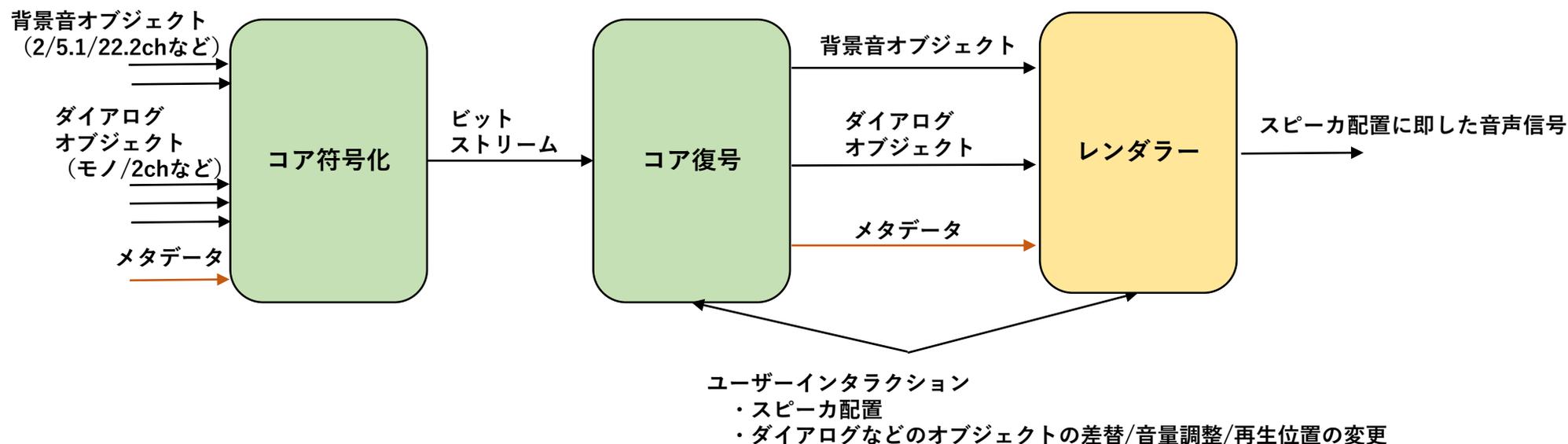
符号化ツールはほぼ同じものを使用

MPEG-4 AACは、PNS (Perceptual Noise Substitution) が追加

PNS:ノイズ成分が多い周波数領域の情報を間引く機能

## OBAに対応した音響音声符号化方式の構造

- コア符号化/復号とレンダラーから構成
  - ✓ コア符号化はダイアログや背景音など各オブジェクトの音声信号を独立に情報量圧縮
  - ✓ レンダラーはメタデータとユーザーインタラクションを基に、種々の機能を提供
    - 音声フォーマット変換、ダイアログの制御（差替や音量調整）など



OBAに対応音声符号化方式のブロック図

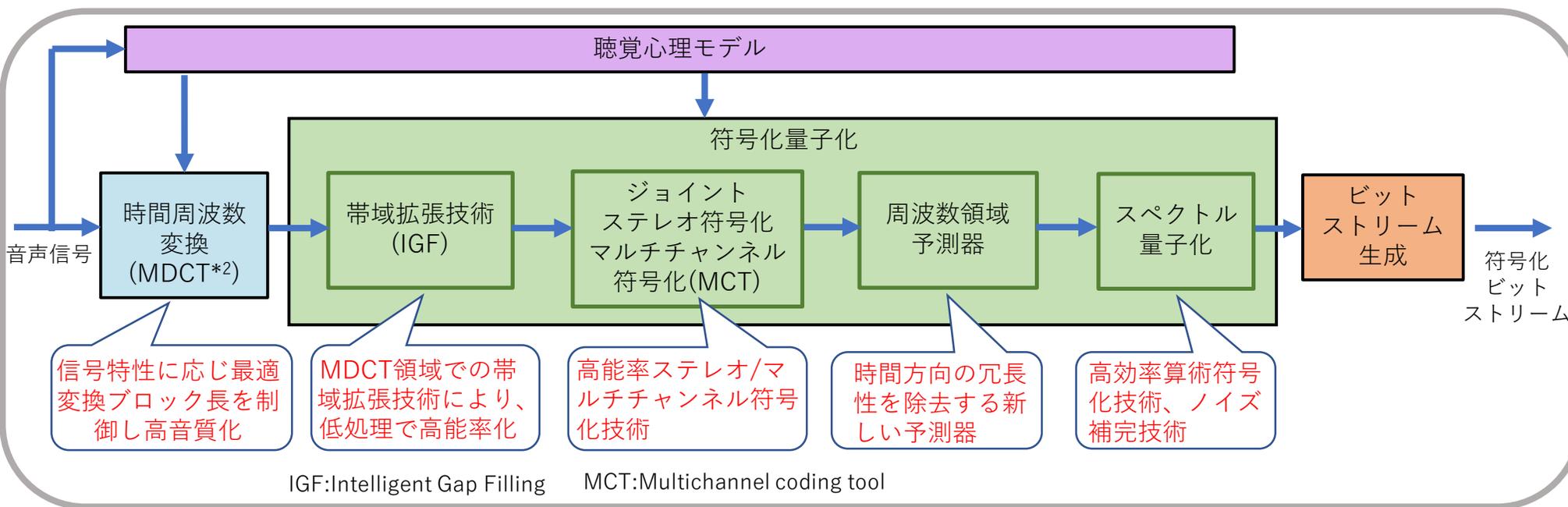
符号化方式毎にコア符号化とレンダラーの詳細技術が異なる

## 最新のMPEG 音声符号化方式 (2015年 MPEG標準化)

- AAC以降に標準化された方式の符号化ツールを積み上げ、高効率圧縮を実現  
768kbps/22.2chで放送品質を実現 (MPEG-4 AACの約50%) \*1
- プロファイルにより符号化ツールのサブセットを規定  
Baseline(BL)プロファイルまたはLow Complexity(LC)プロファイルが利用可能
- ATSC3.0/DVB放送規格に採用され、韓国ではLCプロファイルによる4K地上波放送を開始

\*1 : N19407, MPEG-H 3D Audio Baseline Profile Verification Test Report

## MPEG-H 3DAのコア符号化器のブロック図と主な高能率化



## プロファイルの関係



HOA : High-Order Ambisonics  
AR/VR用音声信号フォーマット

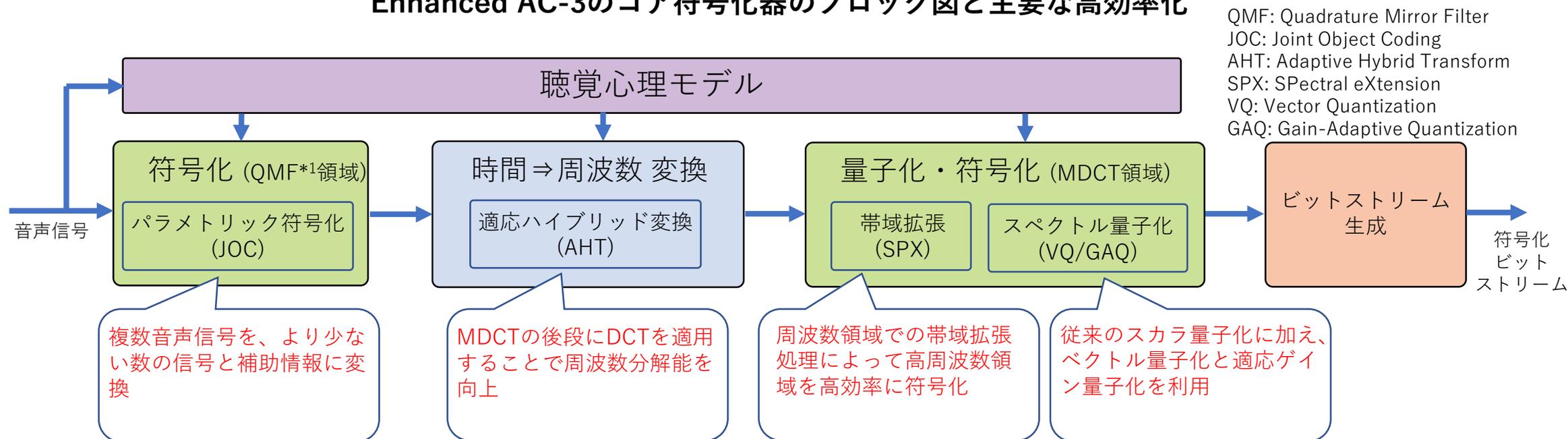
\*2 : 音声信号に対してブロックに分割してDCTを適用するため、ブロック歪を防ぐように修正した変換

現行の放送・インターネット配信サービスで普及し、OBAにも対応する音声符号化方式

- 最大7.1chに対応するチャンネルベース音声符号化方式として2005年にETSI標準化
  - 後にOBA（同時再生：最大16音声信号（LFE含む））に対応し、2016年にETSI標準化
- 欧州やブラジルなどで地上波放送に利用。国内のHybridcast規格にも採用

LFE : Low Frequency Effects

## Enhanced AC-3のコア符号化器のブロック図と主要な高効率化



\*1: 折り返し雑音が発生しないフィルタバンクの一種

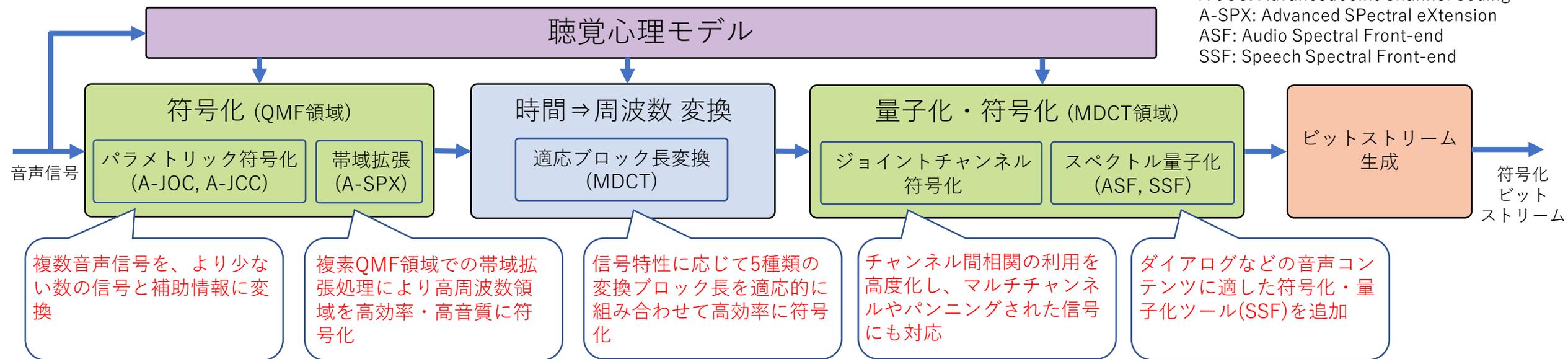
## OBAにも対応した多機能・高効率な最新の音声符号化方式（2015年 ETSI標準化）

- ATSC 3.0/DVBで採用され、北米・欧州・ブラジルで放送規格採用。米国では4K地上放送を開始。
- 192kbps/5.1ch、288kbps/7.1.4chで放送品質\*1
- 国内も含めTV受信機やスマートフォンなどに多数搭載済み

\*1 Dolby AC-4: Audio Delivery for Next-Generation Entertainment Services, [https://professional.dolby.com/siteassets/pdfs/dolby\\_-ac-4\\_audio\\_delivery\\_for--next-generation-entertainment-servicesvx.pdf](https://professional.dolby.com/siteassets/pdfs/dolby_-ac-4_audio_delivery_for--next-generation-entertainment-servicesvx.pdf)

### AC-4のコア符号化器のブロック図と主要な高効率化

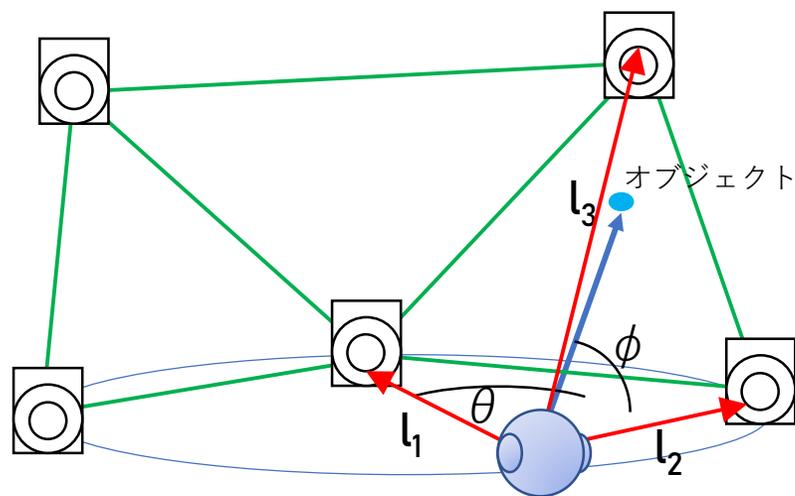
QMF: Quadrature Mirror Filter  
 A-JOC: Advanced Joint Object Coding  
 A-JCC: Advanced Joint Channel Coding  
 A-SPX: Advanced SPectral eXtension  
 ASF: Audio Spectral Front-end  
 SSF: Speech Spectral Front-end



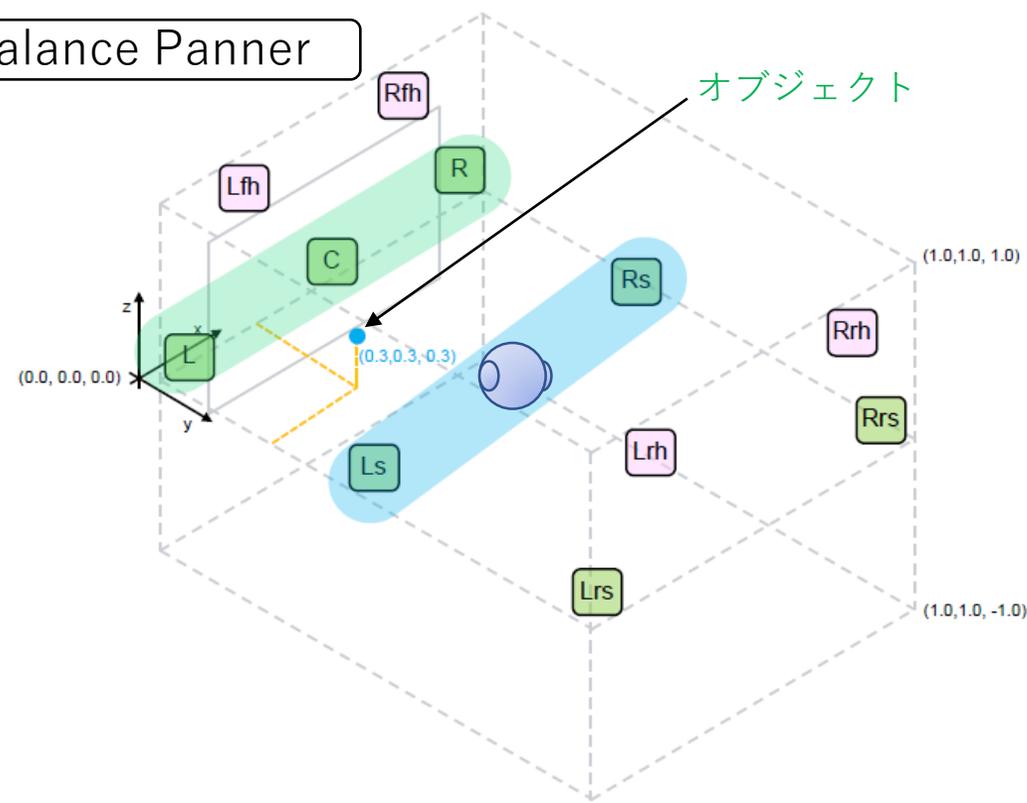
受信機側のスピーカ配置に即してダイアログなどのオブジェクトをマッピングする技術

- MPEG-H : **3D VBAP** (Vector Based. Amplitude Panning)
  - 極座標をベースとし、3スピーカによる領域に分割し各領域ごとに各スピーカーに分配するゲインをベクトル重み付けにより決定
- AC-4 : **Triple Balance Panner**
  - 直交座標をベースとし、オブジェクトの座標に対しx軸、y軸、z軸の各スピーカーに分配するゲインを決定

3D VBAP



Triple Balance Panner



# 符号化方式の比較表

項目	地上デジタルテレビジョン方式の高度化の要求条件	MPEG-2/4 AAC	MPEG-H 3D Audio	Enhanced AC-3	AC-4
放送規格		ARIB DVB	ATSC 3.0 DVB	ATSC DVB	ATSC 3.0 DVB
主な採用国 ※ 放送サービスを実施 無印は規格採用国		日本※ ブラジル※	ブラジル 韓国※	ブラジル※ 欧州※ インド※	ブラジル 北米（米国※） 欧州
同時再生可能な音声信号数 * 括弧内はサービス例	音声入力フォーマット及び符号化方式 音質	MPEG-2 : 8 (例 7.1ch) MPEG-4 : 24 (例 22.2ch)	レベル3 : 16 (例 7.1.4+4obj)、 24 (モノオブジェクトのみ) レベル4 : 28 (例 22.2ch+4obj)	16 (例 7.1.4+4obj)	レベル3 : 18 (例 7.1.4+6obj)
再生機能	高機能化/多様化 サービス	MPEG-4 :チャンネル差替/ 音量調整 (機能制限・制作制約あり)	オブジェクト差替/明瞭化/ 音量調整	オブジェクト差替/明瞭化/ 音量調整 (機能制限あり)	オブジェクト差替/明瞭化/ 音量調整
圧縮効率 * AAC以外の数値は参考情報 (勧告ITU-R BS.1548)	音声入力フォーマット及び符号化方式	144kbps (ステレオ) / 320kbps (5.1ch) / 1.4Mbps (22.2ch) : MPEG-4	768kbps (22.2ch)	192kbps (ステレオ)	96kbps (ステレオ) / 192kbps (5.1ch) / 288kbps (7.1.4)
IPストリーミング/ モバイル規格	高機能化/多様化	3GPP/DASH-IF/ Hybridcast	3GPP/DASH-IF/HbbTV	DASH-IF/HbbTV/ Hybridcast	DASH-IF/HbbTV
HDMI伝送規格	共通性/ インターフェース	IEC 61937-6 IEC 61937-11	IEC 61937-13	IEC 61937-3	IEC 61937-14 (IEC61937-9への変換も可能)
実用化状況		国内の衛星放送用TV、 地上放送用TV	韓国発売のTV、STB、 サウンドバー、AVR	TV、STB、スマートフォン、 サウンドバー、AVR	TV、STB、スマートフォン サウンドバー、AVR
ライセンス		RANDライセンス	RANDライセンス	RANDライセンス	RANDライセンス

## 課題

- 各符号化方式の品質評価が未実施
  - コア符号化
    - 各標準化団体で実施したデータは条件揃わず比較表の数値は参考情報
    - OBAのサービスを想定したコンテンツによる評価データがない
  - レンダラー
    - 再生品質に影響を与えるが、評価方法が定まっていない

## 今後の検討

- 主観評価実験（2021年12月に結果をまとめる）
  - 音源やビットレートなど条件を揃えてコア符号化の品質評価を実施
  - レンダラーの評価法を検討し、可能であれば品質評価を実施