

## 音声圧縮の原理と実際

BS デジタル放送では、ハイビジョン映像にふさわしい高音質のマルチチャンネルステレオ音声サービスが求められます。こうした音声サービスを実現する音声符号化方式としてわが国で採用されたのが、MPEG-2 Audio 規格の中の AAC (Advanced Audio Coding) 符号化方式です。MPEG-2 Audio には、AAC の他に BC (Backward Compatible) と呼ばれる方式もあり、日本の CS デジタル放送および欧州のデジタル放送で採用されています。

AAC は、BC に 2 年半ほど遅れて開発・標準化された方式で、BC に比べ約 2 倍圧縮効率が高いとされています。圧縮前の音声品質を 5 として 5 段階評価をしたときに、評価値の平均および最悪値が 4 以上になるために必要となる情報量は、表 1 に示すようになっています。BS デジタル放送では、これ以上の情報量を確保して、2 チャンネルステレオで 144kbps とすることが一般的になっています。

AAC では、音声信号のデータ量を効果的に削減するために、人間の聴覚特性を利用しています。まず、聴覚の特性について説明しましょう。

人は小さい音から大きい音まで、広い範囲の音を聞くことができるため、音声信号のダイナミックレンジは 90dB (デシベル) 以上 (最も小さい音のレベルと最も大きい音のレベルの比が 30000 以上) を再現する必要があります。現行の衛星放送の B モード音声では、48kHz でサンプリングした信号を±32768 のレベル (16 ビット、約 96dB) で量子化しています。音声信号のダイナミックレンジは、映像 (8~10 ビット) に比べてたいへん広いという特徴があります。

一方、駅のホームで会話をしている時のことを考えてください。電車が入ってくると、電車の音にかき消されて、会話が聞き取れなくなる現象が起こります。これがマスキング効果と呼ばれるものです。つまり、静かな環境で聴く小さな音と、大きな音が聞こえる環境での小さな音の聞こえ方が異なるため、これらを同等に扱うことは効率が悪いのです。

例えば、図 1 では、横軸に音の高さ (周波数)、縦軸にかき消す音に対するかき消される音の相対的な大きさが表されています。大きな音 (A) がある場合、音 (A) によるマスキングカーブ以下の音 (B) は、人の耳には聞き取れなくなります。したがって、符号化の際には、(A) の音のみを送れば、(B) の音を送らなくても、聞いたときに元の音との違いは区別できません。一方、カーブの外側にある (C) や (D) の音は聞き取ることができるので削除することはできません。

AAC ではこのような聴覚の特性を利用して、効果的な圧縮を行っているのです。

|             | AAC     | BC      |
|-------------|---------|---------|
| 5 チャンネルステレオ | 320kbps | 640kbps |
| 2 チャンネルステレオ | 128kbps | 256kbps |
| モノラル        | 64kbps  | 128kbps |

表 1 評価値の平均値が 4 以上になるために必要となる情報量

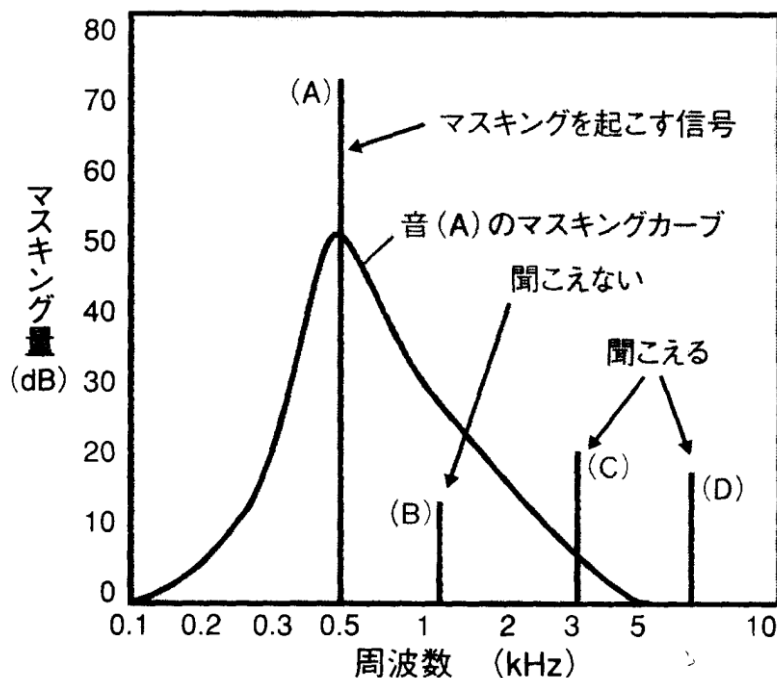


図1 マスキング効果

## 具体的な圧縮の手法

AAC では、人間の耳の聴覚特性であるマスキング効果を利用して、ビット割り当てを重み付けすることで、品質を高く保ちながらビットレートの削減を行います。このとき、音声信号を複数の周波数領域に分割して、各領域ごとに処理します。この処理方法によって、圧縮効果に差が生じます。

AAC では、一定の重なりを持ってある時間間隔ごとに音声入力信号を切り出し、DCT（離散コサイン変換）処理により周波数成分に分割します。これは、映像信号ではブロックごとに DCT 処理を行うことに相当します。DCT 法は、細かい周波数に分割して処理できること、その分割した帯域を一括りにする割合を人間の聴覚特性にあった形で実現可能であることから、AAC に採用されました。

AAC では、DCT されたデータをマスキング効果による聴覚心理モデルに基づいて重み付け量子化します。量子化に当たっては、小さい音の時には細かく、大きい音の時には多少粗く量子化するなど、音の大きさに応じて量子化の細かさを変化させることでデータの削減を図っています。その後、ノイズレス符号化を用いてさらにビットレートを削減します（映像の符号化で説明した可変長符号化に相当）。

DCT と量子化の組み合わせで情報圧縮を行っている点は、映像の圧縮と同じなのです。

## ステレオ符号化

2 チャンネルステレオ信号やマルチチャンネル信号を符号化する場合には、聴覚による音の到来方向の知覚の性質やチャンネル間の相関を利用したビット圧縮が可能です。

前者（音の到来方向の知覚）は、人が高い周波数成分を知覚する際、到着時間遅れより、

音の強さに影響を受けやすい性質を利用して、複数チャンネルの高い周波数の量子化係数を1つの信号にまとめて送ることにより、ビットレートを削減します。後者（チャンネル間の相関）は、左右チャンネルの音が類似して中央に定位する信号の場合は、左（L）チャンネルと右（R）チャンネルの信号を符号化する代わりに、和成分（L+R）と差成分（L-R）を符号化することでビット圧縮が可能になります。

AAC 方式は、低ビットレート音声から高品質音声まで対応できるよう、8kHz から 96kHz までのサンプリング周波数に対応しています。また、音声チャンネルの構成についても、デフォルトの音声モードであるモノ、ステレオ、5.1 チャンネルに限らず、チャンネル数、想定スピーカー配置などに柔軟に対応することができます。

また、AAC には三つの「プロファイル」が用意されています。これは想定されるアプリケーションをもとに機能を制約した、AAC のサブセットと考えてよいでしょう。

#### ① メインプロファイル

処理速度の速い CPU や大きなメモリ容量などが必要ですが、最高の音質が伝送できます。1 チャンネル当たり 64kbps 程度です。

#### ② LC (Low Complexity) プロファイル

メインプロファイルに比べてビットレートは多くなりますが、品質を同程度に保ったまま、プロセッサの演算量やメモリ容量などの回路規模を小さくすることができます。

#### ③ SSR (Scalable Sampling Rate) プロファイル

メイン、LC プロファイルのデコーダでは復号できません。入力段で四つの帯域に分割して処理を行います。デコーダ側で、ハードウェア規模に応じてデコードする帯域を選択できます。

BS デジタル放送では LC プロファイルが採用され、基本的な音声パラメータは表 7-4 に示すとおりになっています。

|           |                                                                    |
|-----------|--------------------------------------------------------------------|
| 音声符号化     | MPEG-2 ACC<br>LC (Low Complexity) プロファイル                           |
| サンプリング周波数 | 48kHz、44.1kHz、32kHz                                                |
| 量子化ビット数   | 16 ビット以上                                                           |
| 音声チャンネル数  | 1 (モノ)、2 (ステレオ)、4 (3/1)、5 (3/2)<br>5. 1 (3/2 + 低域強調用) およびこれらの組み合わせ |
| 参考ビットレート  | 144kbps / ステレオ                                                     |

表 2 BS デジタル放送の音声パラメータ

### AAC 以外の音声圧縮技術

音声圧縮技術には、AAC など、人間の聴覚特性を利用した非可逆圧縮と、完全に元の信号が復元できる可逆圧縮の 2 通りがありますが、後者はまだ圧縮率が低く、現在、研究が進められている分野です。以下に、現在実用化されている非可逆圧縮技術の主なものを示します。

|                                   |                                  |
|-----------------------------------|----------------------------------|
| MP3<br>(MPEG-1 オーディオレイヤー3)        | 各種携帯型プレイヤーが登場                    |
| MPEG-2 BC                         | 現在の CS デジタル放送に採用                 |
| MPEG-2 AAC                        | BS デジタル放送に採用                     |
| MPEG-4 AAC                        | MPEG-2AAC と互換性有り                 |
| ATRAC                             | MD に採用                           |
| ATRAC3                            | ATRAC の 2 倍の圧縮率を実現した改良版          |
| MS Audio<br>(Windows Media Audio) | ATRAC3 と同様、著作権処理も含めたシステムとの連動をめざす |
| Real Audio                        | 現在インターネットで最も普及している圧縮形式           |
| TwinVQ                            | 圧縮に時間がかかるが、圧縮率は高い                |