

放送における音響信号処理

NHK放送技術研究所 西 隆司 渡辺 馨

衛星放送には放送衛星を用いたBS (Broadcasting Satellite) 放送と通信衛星を用いたCS (Communication Satellite) 放送があり、現行のBSテレビ放送では3チャンネルの標準テレビ放送と1チャンネルのMUSE (Multiple-Sub-Nyquist-Sampling Encoding) ハイビジョンテレビ放送が行われている。また、CS放送ではSkyPerfecTVとDirecTV Japanあわせて300チャンネル以上の標準テレビ放送が行われている。現行BSテレビ放送は映像信号をアナログで音響信号をデジタルで放送しており、CSテレビ放送は映像と音響信号を全デジタルで送る方式を採用している。またBS放送においても、2000年12月に映像と音響信号とも全デジタルで送るBSデジタル放送の開始が予定されている。このようにデジタル化が進む放送の中で、デジタル技術を駆使した音響信号の符号化技術、および音響再生技術の現状を中心に解説する。

本稿では、まず現行BS放送の音響符号化について述べ、次にBSデジタル放送の音響符号化について述べる。最後に高臨場感の音声放送を実現するための再生方式のうち、現行の3-1方式および研究レベルにあるマルチチャンネル音場再生について述べる。また、放送の分野では音響信号全般を指す言葉として「音声」を使用することが多く、以下音響信号の意味で音声を使用する。

現行BS放送における音声処理

現行のBS標準テレビ放送の音声は、4チャンネルの音声伝送が可能なAモードと2チャンネルの高品質音声の伝送が可能なBモードがあり、番組により切り換えて使用している。Bモードは国際的なスタジオ規格と同一のサンプリング周波数48kHz、量子化ビット数16ビットが用いられている。また、Aモードのサンプリング周波数は32kHzであり、14ビットで一様量子化した信号を準瞬時圧伸符号化により10ビットに圧縮して伝送して

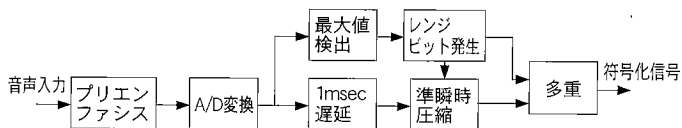


図-1 準瞬時圧伸符号化エンコーダ

いる。準瞬時圧伸符号化の構成を図-1に示す。

準瞬時圧伸符号化は、複数の音声サンプルを1ブロックとし、ブロックごとに音の大きさに応じて量子化の細かさを適応的に変化させる方式である。この符号化の小さい音と大きい音のデジタル符号の例を図-2に示す。小さい音の場合には、上位ビットは最上位ビット(MSB)と同じ値となるので、変化する下位ビットと、音の大きさ情報に相当するMSBと同じ値を持つ上位ビットの個数で表すことができる。一方、音が大きい場合は後述するマスキングを考慮してビット数を少なくすることができる。つまり、大きな音では下位ビットを切り捨てても、人はその劣化が分かりにくい。このように、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ことで、高音質を保ちながら伝送ビット数を減らすことができる。Aモードでは1ミリ秒32サンプルを1ブロックとしてブロックごとに音の大きさを表すレンジビットを計算し、レンジビットとレンジビットに従って10ビットに圧縮した音声データを伝送する。

MUSEハイビジョンテレビ放送の音声も、BS標準テレビ放送と同様に4チャンネル分の音声伝送が可能なAモードと2チャンネルの高品質音声のBモードがある。ハイビジョンMUSE方式に用いられる音声符号化は、差分PCM符号化と準瞬時圧伸符号化を組み合わせた準瞬時圧伸差分符号化である。準瞬時圧伸差分符号化の構成を図-3に示す。この方式は音声サンプルの差分値を計算し、この差分値を準瞬時圧縮する方式で、Aモードと

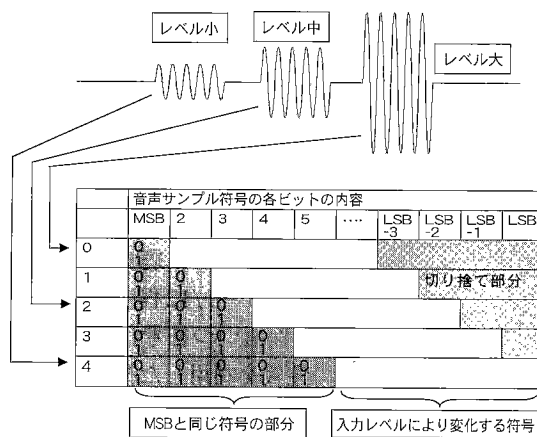


図-2 音のレベルとデジタル符号の例

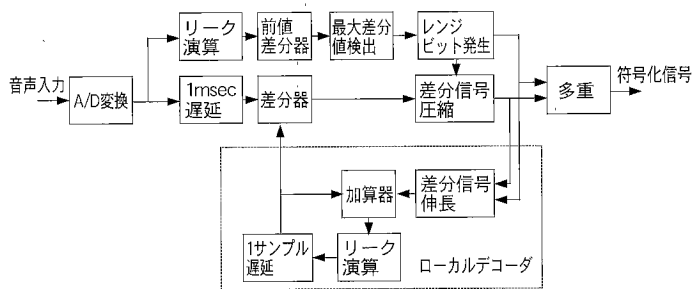


図-3 準瞬時圧伸差分符号化エンコーダ

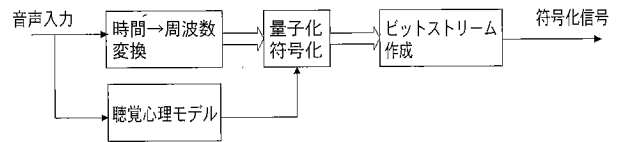


図-4 知覚音響符号化

		圧縮符号化	音声ビットレート (kbit/s)	サンプリング周波数 (kHz)	量子化ビット数 (bit)	最大チャンネル数 (ch)
BS	標準テレビ	Aモード 準瞬時圧伸	328/ch	32	14	4
		Bモード 一様量子化 (非圧縮)	768/ch	48	16	2
	ハイビジョン	Aモード 準瞬時圧伸差分	264/ch	32	15	4
		Bモード 準瞬時圧伸差分	536/ch	48	16	2
デジタル放送	BS	MPEG-2 AAC	72/ch (144/2ch)	32 44.1 48	16以上	5.1
	CS	MPEG-2 BC Layer II	128/ch	32 44.1 48	16	5.1

表-1 放送分野で用いられる音声圧縮符号化と符号化パラメータ

Bモードは各々8ビット、11ビットに差分データを圧縮して伝送している。これらの音質はBS標準テレビ放送の音質と同等であることが確認されている。

現行のBS、CSテレビ放送および次に述べるBSデジタル放送で用いられる音声圧縮符号化と符号化パラメータを表-1に示す。

🍎 音声の低ビットレート符号化処理 🍎

BSデジタル放送ではMPEG-2 AAC (Advanced Audio Coding) 符号化が用いられる。この符号化は、DCT (Discrete Cosine Transform) を用いた周波数変換と、聴覚のマスクングの性質を組み合わせた圧縮符号化である。周波数領域の処理を行うことで、聴覚の性質を効果的に利用することができ、1/6~1/12程度に大幅に圧縮しても、CD並みの音質を保ちつつ符号化することができる。このような聴覚の性質を利用した符号化は、一般的に知覚音響符号化といわれ、図-4の構成で表される。

ここでは、まず聴覚マスクングの性質について述べ、次にAAC符号化の基本的な処理を述べる。

<聴覚マスクングの性質>

マスクングは、たとえばプラットホームで普通に話をしていても、電車が入ってくると電車の大きな音にかき消されて、話が聴こえにくくなる現象である。図-5に示すように大きな音(A)がある場合、このマスクング特性のために、(B)の音が聴こえない。横軸は周波数、縦軸は大きな信号音(A)により聴こえなくなるマスクング量である。符号化の際には、(A)の音により(B)の音は聴こえないので、(A)の音のみを送れば(B)の音を送らなくても聞いた時には元の音と同じ音が聴こえる。一方、(A)の音があっても、(C)や(D)の音は聴こえるので、これらの音は送る必要がある。

音響信号自身によりマスクングされるレベルの例を図-6に示す。このレベル以下の音、たとえば網かけ部に示す量子化ノイズは、音響信号自身によりマスクされるので、聞いた時には元の音と同じ音が聴こえる。

<AAC符号化>

AAC符号化の構成を図-7に示す。AACはいろいろな技術要素からなるが、ここでは基本要素の聴覚マスクングとDCT変換符号化について述べる。

図-5に示すようにマスクング量は周波数に依存しているので周波数成分に分けてマスクングの性質を用いることで効率的な符号化が可能となる。周波数に分ける方法に、複数の帯域に分割する「サブバンド符号化」と、DCTを用いて直接周波数に変換する「DCT変換符号化」があるが、AAC符号化はDCT変換符号化を用いている。

AAC符号化は、まず音響信号をDCT変換により周波数スペクトルに分ける。このうち複数のDCTスペクトルを一括りにまとめる。一括りにするDCTスペクトルの間隔(スケールファクタバンドという)を聴覚の性質に合わせることで効率的に符号化することができる。一括りにしたスケールファクタバンド別に音の大きさを表すレベル成分を求めて、この情報を送る。また、マスクングレベルに基づいた割当ビット数に応じて個々のDCTスペクトルの量子化係数成分を計算する。この時、量子化が粗くなれば送るビット数が少なくなるが、図-6に示すひずみ成分(量子化ノイズ)が大きくなり、劣化が分かりやすくなる。一方、量子化が細くなれば、量子化ノイズが小さくなり劣化が分からなくなるが、送るビット数が大きくなる。この時、図-6のように量子化ノイズがマスクングレベル以下になるように最適なビット割当をすれば、音声信号自身により量子化ノイズがマスクされ、符号化による劣化を聴こえなくすることができる。多くの音響信号では、高い周波数領域のマスクングレベルが信号レベルよりも大きくなり、この帯域の信号

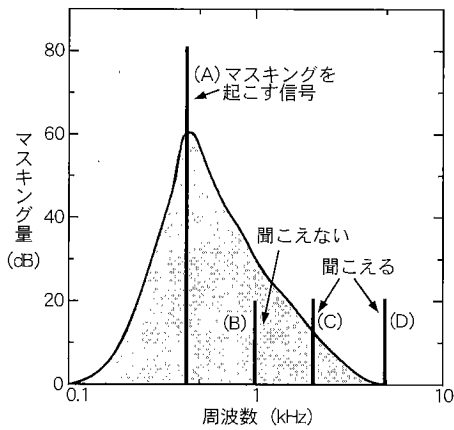


図-5 聴覚マスキングの例

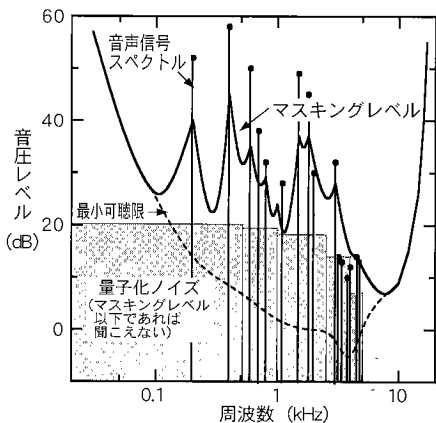


図-6 音響信号とマスキングレベル

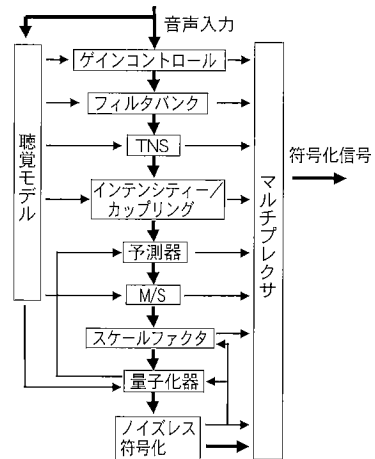


図-7 AACエンコーダの構成図

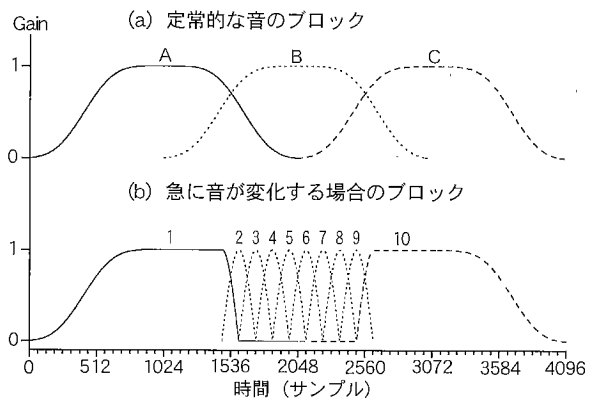


図-8 DCT変換符号化の時間サンプル構成

は送る必要がない。また、送る必要がある場合でも1～2ビット程度に圧縮が可能なが多い。

音響信号のDCT変換は、図-8 (a) に示すような隣接ブロックで50%オーバーラップさせてDCTする方法が用いられる。ブロックのオーバーラップを行わないとブロックのつながり部分に不連続が生まれノイズが起きるが、オーバーラップを行うことでスムーズに符号化することができる。また、TDAC (Time Domain Aliasing Cancelation) と呼ぶDCT変換手法を用いて、オーバーラップによる情報量の増加を防ぐことができる。しかし、サンプルの長いブロックでは、小さな音から急に大きな音になる場合に、小さい音の部分に、原音にはみられないプリエコーと呼ばれる音質劣化が起きる。このような場合は、図-8 (b) に示すように短いブロック長に切り替えを行うことで音質の劣化をなくすことができる。

AAC符号化では、圧縮量子化データをさらにハフマン符号化 (ノイズレス符号化) を用いて伝送ビット数を圧縮している。ハフマン符号は出現頻度の高いデータを短いビット列の符号語で表現し、頻度の低いデータを長いビット列の符号語で表して、量子化データに含まれる冗長性を削減することができる。

< BSデジタル放送とAAC符号化 >

AAC符号化は、所要ビットレートと回路規模および帯域幅に応じた復号の有無を基準に、3つのプロファイルがある。各プロファイルの特徴を示す。BSデジタル

放送ではこの中から音質とコストのバランスを勘案してLCプロファイルを用いることにしている。

(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あるが、最高の音質が伝送できるもので、メインプロファイルと呼ばれる。このプロファイルのみが予測器を持っており、1チャンネル当たり64kbps程度で高音質が実現できる。

(2) LCプロファイル

LC (Low Complexity) プロファイルはメインに比べて伝送ビットレートが多くなるが、プロセッサの演算量やメモリ容量などの回路規模が小さく音質とコストのバランスがよい。このため、さまざまな応用が可能で、BSデジタル放送で用いられる。LCプロファイルの符号化信号はメインプロファイルデコーダで復号できる。

(3) SSRプロファイル

SSR (Scalable Sampling Rate) プロファイルは、メインやLCプロファイルデコーダでは復号できないが、4つの音声の帯域幅に応じた復号ができる。サンプリング周波数が48kHzの時、6kHz、12kHz、18kHz、24kHzの4段階の帯域幅の音が復号できる。復号しない帯域のデータの逆DCTを省略することができ、小さい規模のデコーダも実現できる。

プロファイル以外にもMPEG-2 AAC符号化は各種の

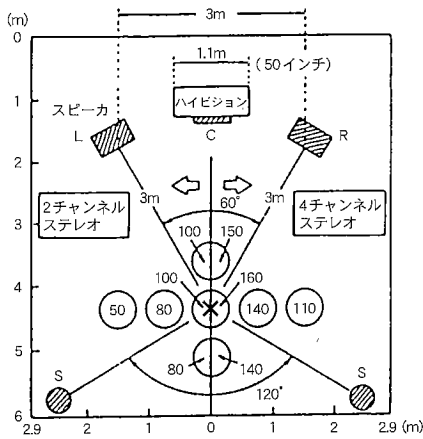


図-9 3-1方式のスピーカ配置および2チャンネルステレオと3-1方式との総合品質の比較

パラメータを持つが、BSデジタル放送では音声放送に最も適した音声パラメータが選定されている。サンプリング周波数は32kHz, 44.1kHz, 48kHz, 量子化ビット数は16ビット以上が用いられる。また音声チャンネル数は、モノ、2チャンネルステレオ音声から3-1方式などのマルチチャンネル音声が行える。また複数の符号化信号を多重でき、多言語放送や視覚障害者用放送も行える。

BSデジタル放送で用いられるLCプロファイルでは、1チャンネル当たり72kbps程度で符号化すれば、劣化が目立ちやすい音源においてもCD並みの放送品質が保てる。

高臨場感音場を再生するための音響処理技術

<3-1方式4チャンネルステレオ>

ハイビジョンの映像は、横長で精細度が高く、高い臨場感を得ることができるため、音響の面でもこの映像にふさわしい高い臨場感の音響再生方式が必要になる。従来の2チャンネルステレオに替わる、新しい音響再生方式の開発にあたって、次の要求条件を設定した。

- (1) 高品質な音声であること。
- (2) 2チャンネルステレオより高い臨場感が得られること。
- (3) 複数の人が同時に大きな画面で視聴する場合を想定し、広い聴取範囲で映像と音響の方向のずれが小さいこと、および、広い聴取範囲で高い臨場感が得られること。
- (4) 2チャンネルステレオに変換できること。

4チャンネル以下の再生方式を検討対象とし、種々のスピーカ配置、番組プログラムの条件下で聴感実験を行った。図-9にスピーカ配置を示すように、3-1方式は、聴取者の前方、左、中央、右の3チャンネル、後方の1チャンネルの音声信号を再生する方式である。なお後方の1チャンネルに関しては、この後方用の音声信号を同相、同レベルで2つのスピーカから再生することにより、再生音場の広がり感を増加させている。また中央にスピーカを置くことにより、視聴位置によらず正面方向の音響を常に安定に定位できることも3-1方式の効果と

して期待できる。2チャンネルステレオと3-1方式とを総合的な品質で比較した結果を図-9に示す。図中、○印は視聴位置、中の数字は総合品質の評価値を表し、2チャンネルステレオを中心軸上のベストの位置(×印の位置)で視聴した場合の基準評価値(100)に対する相対評価値を表している。左半分が2チャンネルステレオの評価値、右半分が4チャンネルステレオ(3-1方式)の評価値を表している。この結果から、4チャンネルステレオの総合的な品質は、中央位置および中央から1.5m離れた端でも2チャンネルステレオの中央位置のそれぞれ1.6倍、1.1倍であり、2チャンネルステレオ中央位置で視聴した場合より品質が良いことが分かる。また前後の視聴位置のずれが総合的な品質評価にほとんど影響していないことも分かる。したがって、3-1方式は、広い聴取範囲で、2チャンネルステレオより高い品質であることが結論された。

3-1方式4チャンネルステレオ(L, C, R, S)は2チャンネルステレオ(L', R')と

$$L' = L + 0.7C + 0.7S$$

$$R' = R + 0.7C + 0.7S$$

の変換式を使うことにより、変換できることが品質評価実験から確認されている。

現在、「NHK歌謡コンサート」、「高校野球」、「紅白歌合戦」等のハイビジョン番組で3-1方式4チャンネルステレオが放送されている。

映像と音声の総合的な品質の確保、電波の有効利用、番組制作や受信・再生機器に要するコスト等を総合的に判断した結果、ITU-R(国際電気通信連合無線通信部会)では、上述の3-1方式の他、後方の再生チャンネルを増やした3-2方式等のマルチチャンネル再生方式が勧告されている。

<マルチチャンネル音場再生手法>

ハイビジョン、3-1方式マルチチャンネルステレオ以後の映像、音声の両面で高い臨場感を持ったバーチャルな世界を家庭で再現することが、放送のニューメディア開発の大きな目標になっている。映像の奥行き感を再現する3次元立体映像の研究開発もこの流れに沿ったものと考えられる。

3-1方式のマルチチャンネルステレオよりも、さらに現実近く、臨場感のある音を視聴者に提供するにはどのように音を収録し、加工し、伝送し、再生すればよいかという視点からの研究を進めている。ここではその1つのアプローチとしてコンサートホール等の原音場を家庭のリビングルームで再現するための基本的な技術について述べる。

コンサートホールのステージ上で演奏された音は、図-10(a)に示すように、壁や天井で何度も反射を繰り返して、受聴者に届いている。図-10(b)に示すように、原音場の反射音の方向にスピーカを設置し、そこから反

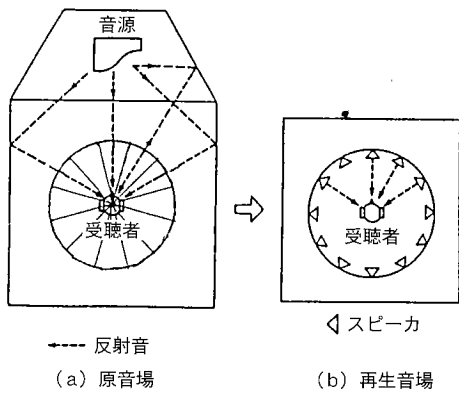
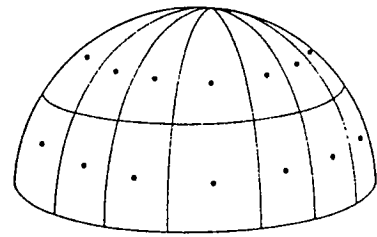


図-10 マルチチャンネル音場再生の概念図



・は分割の中心を表し、多チャンネル音場再生時にはこの位置にスピーカを配置する。

図-11 受聴点を中心とした音響空間の分割

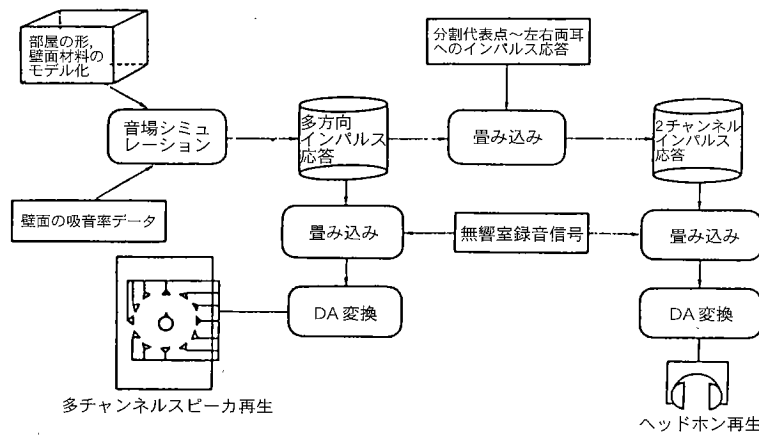


図-12 音場シミュレーションによって得られたインパルス応答からマルチチャンネルスピーカ再生までの流れ (2チャンネルヘッドホン再生の流れを含む)

射音をマルチチャンネルで再生すれば、コンサートホールの音場を、音環境の異なる家庭の部屋で再現することができると考えられる。原音場の反射音の振舞いは、音源から客席までの反射音の時系列特性（以後、インパルス応答と呼ぶ）から知ることができる。このインパルス応答によって、部屋という一種の伝送系の物理的な特徴を表すこともできる。たとえば、ホールの重要な物理パラメータである残響時間（響きの長さ）は、インパルス応答が減衰する速さに対応した物理量である。

マルチチャンネルの音響信号は、以下の3段階の手続きを経て得ることができる。

- (1) 原音場における受聴者の周りを図-11のように有有限個の空間に分割する。
- (2) それぞれの空間インパルス応答を求め、響きのない音源信号に畳み込む。
- (3) 再生音場で各方向に対応したスピーカから、上記(2)の音を再生する。

受聴者に入射する方向ごとのインパルス応答は、次の3つの方法のいずれかを使って求めることができる。

- (a) 音場で、鋭い指向性のマイクを使ってインパルス応答を測定する。
- (b) 計算機内に作った部屋のモデルをもとに、シミュレーションによってインパルス応答を予測する。
- (c) 1/10, 1/20といった部屋のスケールモデルを作り、

そのモデル内でインパルス応答を測定する。

上記 (b) の場合のシミュレーションによって得られたインパルス応答からマルチチャンネル再生の音が得られるまでの流れを図-12に表す。2チャンネルのインパルス応答は、受聴者を中心とした多方向のインパルス応答と、各方向から左右両耳までのインパルス応答を畳み込み合成することによって得ることができる。このマルチチャンネルから2チャンネルへの変換法も図-12に示している。

マルチチャンネル再生した音が、原音場の音をどの程度再現できているかを物理的、聴感的に評価することは重要であり、盛んに研究が進められている。残響時間等の音場のマクロな音響特性が原音場とマルチチャンネル再生音場とでよく一致することが種々の音場を用いた実験で確認されている。

マルチチャンネル再生の研究課題は、独立なチャンネル数、空間分割数、スピーカ配置、再生スピーカの音響特性等であり、それらを高臨場感再生の観点から最適化する研究が現在行われている。

参考文献

- 1) ISO/IEC JTC1/SC29/WG11 IS13818-7 MPEG-2 AAC Audio規格。
- 2) 黒住幸一: 放送におけるマルチチャンネルステレオ音声放送, 日本音響学会誌50 (11), pp.915-920 (1994)。

(平成11年8月5日受付)