

オーディオメディアの 高品質圧縮符号化技術の動向

北脇信彦 筑波大学電子・情報工学会

はじめに

インターネットを介して高品質のステレオ音楽をパソコンで聞く。このようなサービスが現実のものとなっている(図-1)。音楽メディアは音声にくらべて信号帯域が広く情報量が大きいので、以前にはパソコン通信で音楽を送ることなどは考えられなかったが、最近の符号化技術のめざましい進展によって、インターネットのような低速ネットワークで扱えるほどの高圧縮符号化が可能となってきた。

また、オーディオメディアのマルチチャネルステレオ符号化技術が進展し、DVD (Digital Versatile Disc) などに適用されはじめている。これによって臨場感あるホームシアターが実現可能になった。

本稿では、このようなオーディオメディアの高品質圧縮符号化技術の動向について述べる。

品質を損なわずに圧縮符号化

オーディオメディアのデジタル表現

いろいろなコミュニケーションメディアにおける音声および音楽メディアの上限周波数を図-2に示す。最も基本的な電話音声では、経済性の面から3.4kHzを上限周波数としていた。通話の了解性を確保することと、ハンドセットによる通話ということを考慮したものである。ISDNで用いられる広帯域電話音声では、テレビ会議のようなスピーカによる出力ということを考慮して、より自然性のある7kHzを上限周波数としている。ラジオ放送では、AM (Amplitude Modulation) 放送が8kHz、FM (Frequency Modulation) 放送が15kHzを上限周波数としている。また、CD (Compact Disc) や DAT (Digital Audio Tape) などのパッケージメディアでは、22kHzを上限周波数としている。この

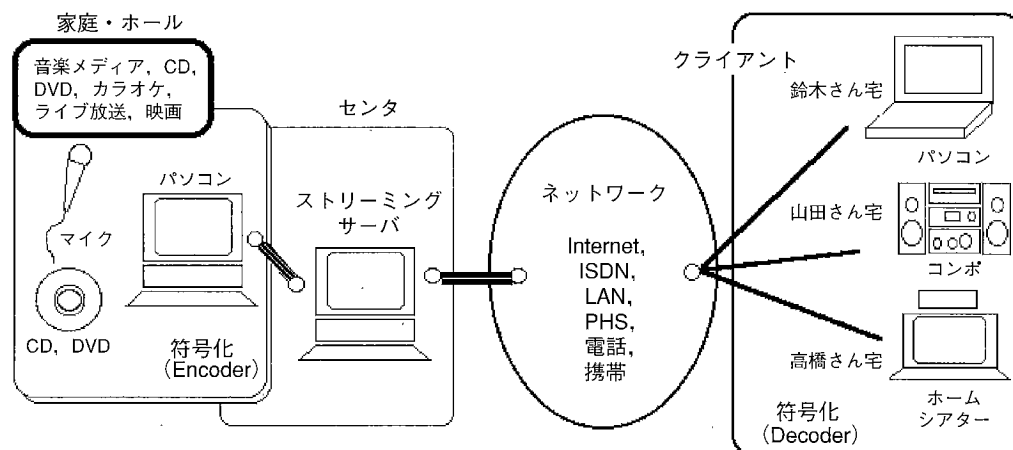


図-1 音楽メディア発信/配信システム

ことから、CDは電話に比べて約6.5倍の信号帯域を持っているといえる。

音声や音楽などのアナログメディアは標本化と量子化の操作によって、デジタル信号に変換される。すなわち、標本化によって時間的に飛び飛びの離散値に変換され、量子化によって振幅上で飛び飛びの離散値に変換される。標本化定理によれば、メディア信号が持つ上限周波数の2倍以上のサンプリング周波数で標本化すれば、完全に元の信号を復元できる。たとえば、電話音声の場合には、3.4 kHzを上限周波数とする音声を、8 kHzのサンプリング周波数で標本化し、各々の標本値を8bit/sampleで量子化している。したがってビットレートは、8 bit × 8,000 sample = 64 kbit/sとなる。この方式をPCM (Pulse Code Modulation) 方式という。

このようにしてデジタル化された音楽パッケージメディアのビットレートを表-1に示す。CDやDATのチャンネルあたりのビットレートは705kbit/sである。2チャンネルステレオでは1.41Mbit/sになる。電話の64kbit/sに比べて約22倍である。実際には、同期や誤り訂正などのためのオーバーヘッドがあるため、これのさらに2～3倍のビットレートを要する。なお、CDのデジタル表現がオーディオメディアに対するデジタル表現のデファクトスタンダードになっている。

音楽パッケージメディアとして用いられているMD (Mini Disc) やDCC (Digital Compact Cassette) では、圧縮符号化方式を用いてCDのビットレートの半分程度に情報量を圧縮している。MDで用いられている符号化方式はATRAC (Adaptive Transform Acoustic Coding) 方式といい、DCCで用いられている符号化方式はPASC (Precision Adaptive Sub-band Coding) 方式という。両方式とも可変長符号を用いているため、表-1におけるサンプル

あたりの量子化ビット数は平均で表されている。

オーディオメディアと圧縮符号化技術

オーディオメディアの符号化技術の構造を図-3に示す。ビットレートを低減する情報源圧縮符号化の技術、オーディオメディア信号をハイファイで立体再生するためのマルチチャンネルステレオ符号化技術、およびネットワークでの信頼性を高める通信路符号化技術からなっている。

2つのスピーカを用いた2チャンネルステレオ再生は、最も普及した空間音場の再現手段である。チャンネル数が多いほど、臨場感は向上する。大きな空間での音場再生は主に映画用音響として

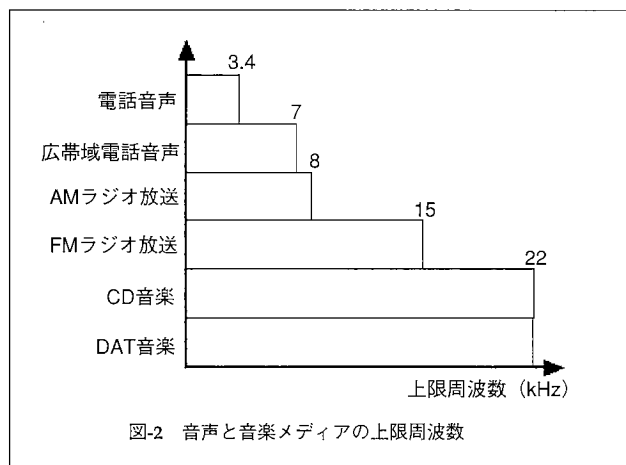
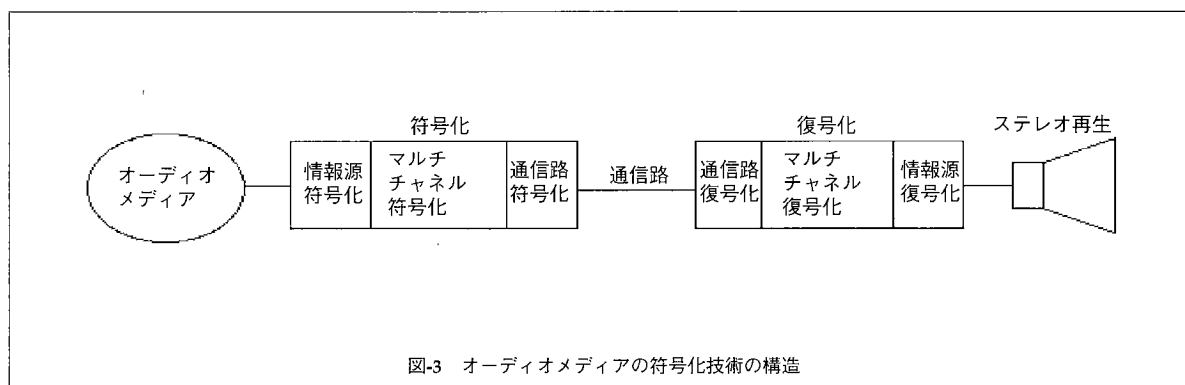
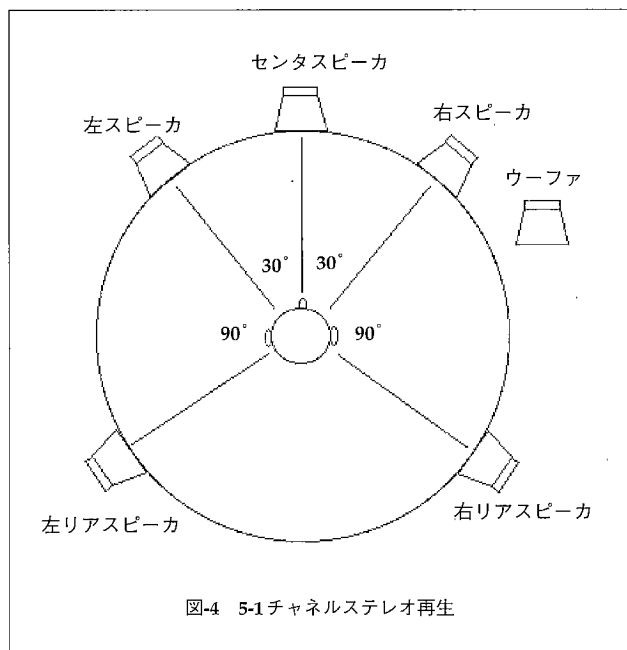


表-1 音楽パッケージメディアのビットレート

	標本化周波数 (kHz)	量子化ビット数 (bit/sample)	チャンネルあたりビットレート (kbit/s)	符号化方式
DAT	48	16	705 (総計2160*)	圧縮していない
CD	44.1	16	705 (総計1540*)	圧縮していない
MD	44.1	平均6.8	300	ATRAC
DCC	32 44.1 48	平均12 平均8.7 平均8	384	PASC

(*) 同期、誤り訂正などを含む





進展してきた。1952年に登場したシネラマでは、前方5チャンネル、側・後方2チャンネルの7チャンネル系、シネマスコープでは前方3チャンネル、側・後方1チャンネルの4チャンネル系が用いられている。また、1991年に標準化されたハイビジョンでは、前方3チャンネル、側・後方1チャンネルの3-1チャンネル系が用いられている。

典型的な例として前方3チャンネル、側・後方2チャンネルおよびウーファスピーカを持つ5-1チャンネルステレオ再生系を図-4に示す。センタスピーカは、セリフなど中央に位置する音を受け持つスピーカである。左右スピーカはメインとなるスピーカ、リアスピーカは効果音などの再生を受け持つサラウンドスピーカである。ウーファスピーカは重低音を受け持つスピーカで、低音は指向性がないのでどこに置いてもよい。

オーディオ符号化の標準化

標準方式

低ビットレートでのオーディオ符号化の研究は比較的新しい。64 kbit/sのビットレートで音楽信号を適応予測符号化する先駆的な装置（HiFi APC-AB：HiFi Adaptive Predictive

Coding with Adaptive Bit Allocation）が、1982年に開発された¹⁾。また、7 kHz帯域の広帯域音声を64 kbit/sで符号化する方式が、ITU-T^{*1}において勧告G.722（SB-ADPCM：Sub-band Adaptive Differential PCM）として1988年に標準化された²⁾。これは音声の主たるメディアとし、背景音楽も符号化できることをうたっている。

しかし本格的なオーディオ符号化の国際規格は、1988年に標準化作業を開始したISO/IEC^{*2}におけるMPEG^{*3}の活動を待たなければならなかった^{3), 4)}。表-2に標準オーディオ符号化方式の概要を示す。また、MPEG標準オーディオ符号化方式の構造を図-5に示す。なお、MPEG標準化におけるオーディオメディアの符号化は、MPEG Audioと呼ぶのが正式であるが、本稿では省略して単にMPEGと呼ぶ。

MPEG-1

1992年に標準化されたMPEG-1方式は、情報圧縮率と処理量が異なる3階層の符号化アルゴリズムを持つ構造になっている。レイヤIからレイヤIIIになるにしたがって符号化アルゴリズムは複雑になるが、それだけ品質も向上する。レイヤ間には互換性があり、どのレイヤで符号化されても、受信側で復号することができる。

MPEG-1方式は、帯域分割符号化および変換符号化技術をベースにしており、これ以降に標準化された方式の基礎になっている。サンプリング周波数は32, 44.1, 48kHz、ビットレートは128, 192, 256, 384kbit/s（ステレオ）である。2チャンネルステレオで、高サンプリング周波数のものはCDと同等の品質を有している。アプリケーションとしては、パッケージメディアを対象としている。

MPEG-2

MPEG-2の符号化アルゴリズムはMPEG-1と基本的には同じで、マルチチャンネルステレオ化を図ったものである。アプリケーションは放送、通信、映画を対象としている。

MPEG-2にはMPEG-1と互換性があるMPEG-2/BC（Backward Compatible）と、互換性を考慮していないMPEG-2/AAC（Advanced Audio Coding）とがある。MPEG-2/BCは図-4に示した5-1チャンネル系、MPEG-2/AACは7-1チャンネル系によりマルチチャンネルステレオ再生を行うことができる。MPEG-2/BCは、サンプリング周波数およびビットレートがMPEG-1とまったく同じMC（Multi Channel）/ML

^{*1} ITU-T：International Telecommunication Union-Telecommunication Sector,
国際電気通信連合-電気通信標準化部門

^{*2} ISO：International Standardization Organization,
国際標準化機構

IEC：International Electrotechnical Commission,
国際電気標準化会議

^{*3} MPEG：Moving Picture Experts Group,
動画専門家グループ

表-2 標準オーディオ符号化方式

年	標準方式	品質	特徴	主な用途
1988	ITU-T G.722	高品質電話	音声と背景音楽	ISDN
1992	MPEG-1	CDと同等の品質	2チャンネルステレオ	パッケージメディア
1994	MPEG-2/BC	CDと同等の品質	5-1チャンネルステレオ	放送、通信、映画
1997	MPEG-2/AAC		7-1チャンネルステレオ	
1999	MPEG-4	ラジオ・電話並みの品質	音声および音楽用汎用符号化方式	インターネット、携帯電話、テレビ電話

(Multi Lingual) アルゴリズムと、サンプリング周波数がMPEG-1の半分のLSF (Low Sampling Frequency) アルゴリズムとがある。LSFのビットレートは8～256kbit/sである。MPEG-2/BC-LSFは、MPEG-1と後方互換を保ちながら、最近のインターネット通信などの低速ネットワークサービスに応じようとして設けられたものである。MPEG-2/AACは、サンプリング周波数およびビットレートがMPEG-1と同じMC/MLアルゴリズムであるが、MPEG-1との互換性はない。互換性を考慮しないことによって高品質化を図り、DVDやホームシアターへの適用を狙っている。

MPEG-4

1999年初めの標準化を目指して検討が進められているのがMPEG-4である。MPEG-3は検討途上でMPEG-4に吸収されたため、欠番となっている。ビットレートによって分類したMPEG標準符号化方式の適用領域を図-6に示す。

MPEG-4は最近急速に需要が高まっている64kbit/s以下の低速ネットワークサービスに応じようとするものである。品質はCD品質よりもやや下のラジオ／電話品質を下回らないことを目指している。MPEG-1、MPEG-2との互換性は考慮されていない。アプリケーションも広範囲で、インターネットサービス、携帯電話、テレビ電話、パッケージメディア、狭帯域放送などが考えられている。

MPEG-4には3種類の符号化方式がある。①64kbit/s以下で帯域分割／変換をベースにした符号化 (Subband/transform-based coding), ②6-16kbit/sで合成による分析をベースにした符号化 (Analysis-by-synthesis coding), ③2-10kbit/sでパラメトリック符号化 (Parametric

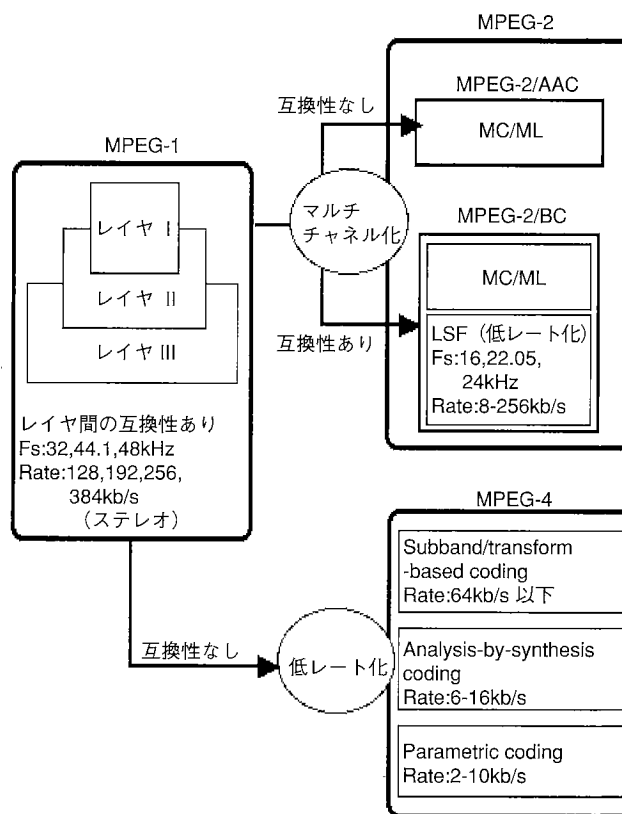


図-5 MPEG標準オーディオ符号化方式の構造

coding) をしたものである。①は狭帯域放送、②はインターネット音楽配信サービス、③は音声を中心としたサービスを主なアプリケーションの対象にしている。

②のA-b-Sをベースにした符号化方式のCommittee Draft (標準化勧告の原案) として、TwinVQ (Transform domain Weighted Interleave Vector Quantization: 変換領域重み付けインターリーブベクトル量子化) 方式⁵⁾ が1998年に選定された。

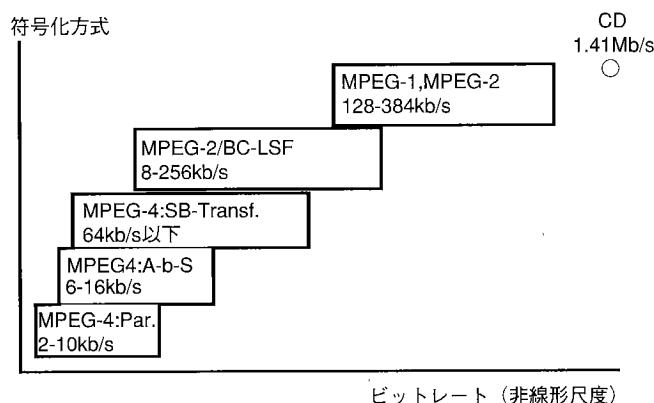


図-6 ビットレートによる符号化方式の適用領域

表-3 オーディオ符号化の要素技術

信号の冗長性	情報圧縮技術
オーディオメディアの統計的性質の利用	直交変換 帯域分割 フレーム内予測 (線形予測) フレーム間予測 ピッチ予測 (長期予測) 分析合成
聴覚特性の利用	帯域制限 非線形量子化 聴覚重み付け ポストフィルタ
統計的頻度分布の利用	可変長符号 可変長フレーム 無音区間圧縮 ベクトル量子化 (符号帳) インターリーブ

オーディオ符号化技術

基本技術

近年のオーディオ符号化技術は、符号化ひずみを評価関数として提案された種々の要素技術の組合せからなっている。オーディオメディアが持つ冗長性を、冗長性が発生する要因ごとに、メディアの統計的性質、聴覚特性、統計的発生頻度分布に分類し、それぞれの冗長性を抑圧する情報圧縮技術を表-3に示す。それぞれの情報圧縮技術の詳細な説明、および個々の符号化方式がどの要素技術を用いているのかは、紙数の関係でここでは述べないので、技術専門書を参照されたい。

Twin-VQ

MPEG-4で標準化作業が進められているTwinVQ方式の構成を図-7に示す。入力信号をMDCT (Modified Discrete Cosine Transformation) で周波数領域に直交変換する。これを線形予測技術によって求めたLSP (Line Spec-

trum Pair) 係数によって正規化する。スペクトル平坦化された信号を、次にパワー正規化し、インターリーブベクトル量子化する。すなわち、インターリーブによって係数の統計的分布に偏りを起こさせ、これをベクトル単位で量子化することによって、量子化の効率をあげる。TwinVQの音質はMPEG-2/BC-LSFの同じビットレートのものよりも優れている。

インターネット音楽発信/配信サービス

インターネットを用いた音楽発信/配信サービスでは、図-1に示したように、家庭、ホール、あるいはセンタで圧縮符号化した音楽ソースを、センタサーバのワークステーションに蓄積し、家庭や店舗のパソコンから、必要なときに希望の音楽を短時間で取り出して聴取するようになっている。

インターネット音楽サービスは、大きく分けて蓄積再生型と実時間再生型に分かれる。蓄積

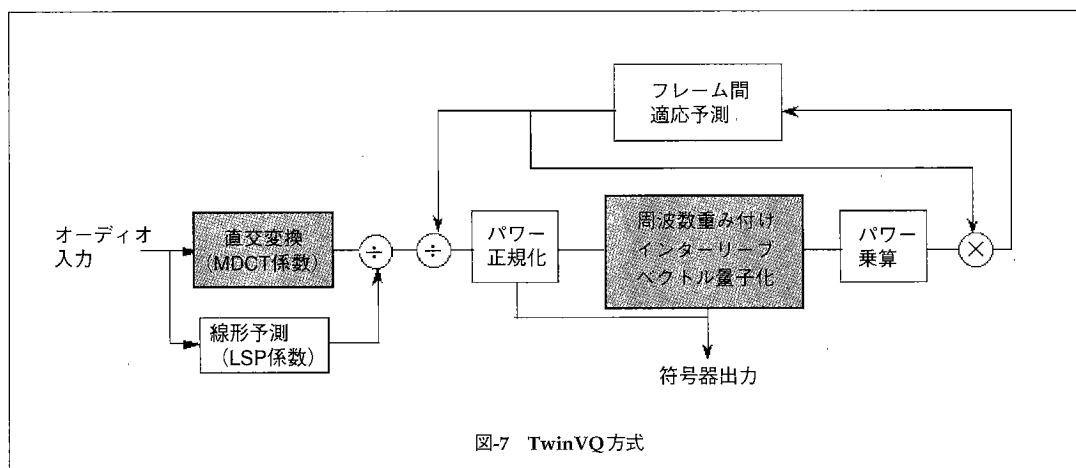


図-7 TwinVQ方式

表-4 オーディオ符号化方式とインターネット音楽サービス

符号化方式	インターネット音楽サービス
MPEG-2をベース	StreamWorks, ShockWave, NetShow
TwinVQをベース	VDOLive, NetShow, MobileMotion, music.co.jp, avex network, SoundVQ, PDB CD SAMPLER, MetaSound, WabTheater
AC-3 (Dolby社)をベース	RealAudio, LiquidAudio, NetShow

再生型はWAV形式などのファイルをサーバからダウンロードし、その後に再生を開始するものである。ファイルのダウンロードが終了するまでは音楽を聴取できないが、伝送路での誤りの訂正などができるので、実時間型に比べて高品質の音楽を提供できる。また、音声出力機能を有するパーソナルコンピュータであれば、安価に入手できるWWW (World Wide Web) ブラウザでサービスを受けられる。

実時間再生型は音楽データの伝送が始まると、すぐに音楽の再生が開始されるものである。したがって、待ち時間なく音楽を聴取することができる。しかし、圧縮符号化技術をもってしても、膨大な情報量を持つ音楽情報をインターネットを介して実時間で再生するためには、音楽の周波数帯域を狭めるなどして、多少品質を犠牲にしなければならないのが現状である。

インターネットの通信プロトコルは、通常TCP (Transport Control Protocol) が使われているが、インターネットでのさまざまなサービスを考慮してUDP (User Datagram Protocol), RTP (Real-time Transport Protocol), RTSP (Real-time Streaming Protocol) などが提案され、通信プロトコルに起因する品質劣化を最小にする技術も開発されている。

初期のインターネット音楽配信サービスではITU-T勧告G.722による符号化方式などが用いられていたが、品質は必ずしも満足できるものではなかった。その後、インターネットの普及とともに低ビットレートのオーディオ符号化技術も進展し、MPEG-1, MPEG-2/BC-LSF, TwinVQなどを用いて品質面での改善も急速になされた。いま話題のインターネット音楽サービスとそのシステムで用いられている符号化方式を表-4に示す。AC-3方式 (Dolby社) は標準方式ではないが、波形符号化、帯域分割符号化、変換符号化の技術を用いており、アルゴリズムの面からはMPEG-1, MPEG-2にきわめて近い。

また、PHSなどの無線通信で音楽サービスを

行うことを考慮して、符号誤り訂正などの通信路符号化を導入した方式も開発されている⁶⁾。

このようなミュージック・オン・デマンドシステムでは、音楽試聴レベルの高品質から低ビットの音楽蓄積まで、品質と圧縮率を用途に応じて選ぶことができる。また信号の周波数帯域は、電話帯域からCD帯域まで、用途に応じて選ぶことができる。

あとがき

本稿では、オーディオメディアの高品質圧縮符号化技術の動向について述べた。CD品質を維持したままでビットレートを低減する方向と、CD品質をやや下回るもののインターネットなどに適用できる極低ビットレート化の方向で、符号化方式が開発されてきたことを述べた。これらの方式はホームシアターなどの高品質臨場感音楽サービス、およびインターネット音楽発信/配信サービスなどへの適用が始まっている。

謝辞 研究を進めるにあたり筑波大学学内プロジェクトおよび電気通信普及財団から助成を受けたことを記して謝意を表する。

参考文献

- 1) 林, 誉田, 北脇: 適応予測を用いた高品質音響信号の高効率符号化, 電子情報通信学会論文誌, J67-A, No.4, pp.375-382 (1984)。
- 2) ITU, ITU-T Recommendation G.722-7 kHz Audio-Coding within 64 kbit/s.
- 3) Noll, P.: MPEG Digital Audio Coding, IEEE Signal Process. Mag., Vol.14, No.5, pp.59-81 (1997)。
- 4) 杉山, 岩垂, 野村: オーディオ符号化の国際標準, 日本音響学会誌, Vol.54, No.2, pp.145-153 (1998)。
- 5) Iwakami, N., Moriya, T. and Miki, S.: High Quality Audio Coding at Less than 64kb/s by Using Transform-domain Weighted Interleave Vector Quantization (TwinVQ), Proc. IEEE ICASSP95, pp.3095-3098 (1995)。
- 6) 森, 池田, 守谷, 若上, 金子: TwinVQによるPHS用ステレオ符号化, 電子情報通信学会情報・システムソサエティー大会講演論文集, pp.286 (1997)。

(平成10年8月20日受付)