

モバイルマルチメディア信号処理技術特集

モバイルマルチメディア信号処理技術概要

モバイルマルチメディアサービスの普及にはIMT・2000に見られる通信基盤の整備と、これに対応したマルチメディア信号処理技術の開発が不可欠である。

本稿では、信号処理技術の概要を時間軸上で述べるとともに関連するシステム技術を説明する。

なかの ひろたか えとう みのる
中野 博隆 栄藤 稔

1. まえがき

21世紀を迎えて、“次世代”移動通信網といってきたIMT・2000 (International Mobile Telecommunications・2000)のサービスも間近に迫ってきている。近年の通信路広帯域化により情報圧縮は必要ないとの意見はあるが、周波数帯域は有限であり、またWWWに見られる個人発信情報の指数的な増加を考えると今後のモバイルネットワークにおいては画像、音声、音響データを効率良く伝送するための情報圧縮技術は依然として中心的役割を果たすであろう。

本稿では、現在のドコモ基本サービスを担っている音声符号化技術およびIMT・2000へのサービス導入が考えられる画像・音声符号化技術を概説する。そして、これら符号化技術を統合し、網との接続を実現するシステム技術を述べる。

2. 画像符号化技術， 音響符号化技術， 音声符号化技術 の概観

信号をデジタル化して蓄積伝送しようとするとき、その波形をどのようにモデル化するのが最も重要である。

デジタル化された信号をそれぞれ統計的に独立の標本としてモデル化すると、パルス符号変調 (PCM: Pulse Code Modulation) 技術となる。前後する標本に相関を仮定すると、相関を利用した予測値との差分を符号化する差分PCM (DPCM: Differential PCM) 技術となる。1970年代から1980年代にかけての研究開発はPCM, DPCMの量子化と予測に関するものであった。画像、音響、音声の符号化技術はこの頃まで、入力波形をいかに忠実に符号化するかという波形符号化技術のもとで同じ舞台で語られてきた。

1960年代中頃に離散フーリエ変換 (DFT: Discrete Fourier Transform) の変形として離散コサイン変換 (DCT: Discrete Cosine Transform) が発明され、1980年代になり、符号化への応用が検討されたときから各符号化技術が分化を始める。利用できる要素技術が進歩し、それぞれ適した情報源モデルが異なり始めたからである。以下、画像、音響、音声の順に符号化の歴史、系譜を述べていきたいと思う。各論に入る前にマルチメディア符号化の特色を一つ述べる。それはメディア符号化が人間相手の符号化技術であるということである。画像・音響・音声符号化では、医用など一部の用途を除き波形は忠実に再現される必要はない。人間が許容できる範囲で歪みは許される。

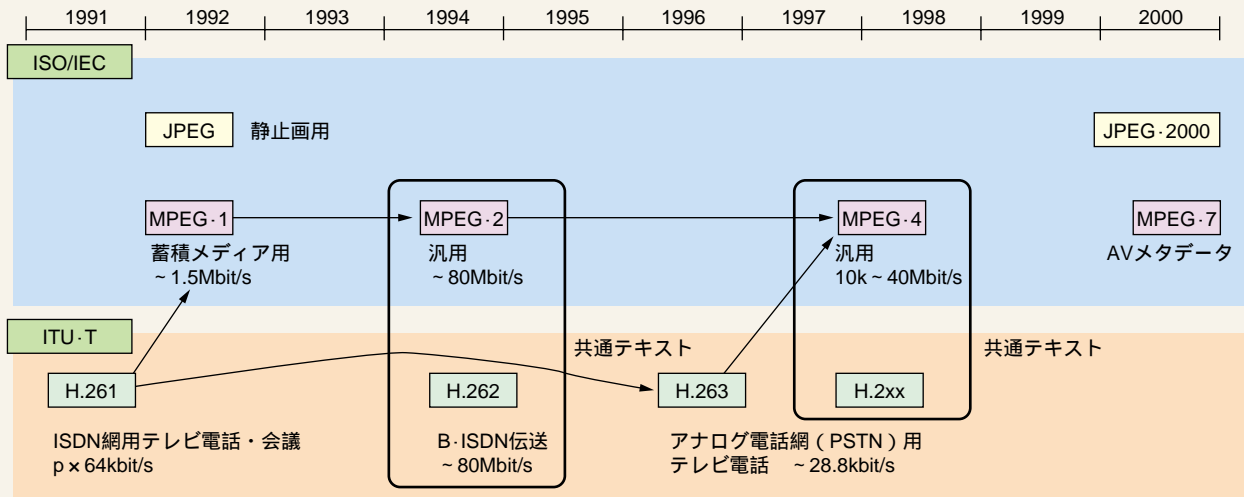
この人間が許容できる歪みとは何かをモデル化することが重要であり、この分野の符号化技術の真髄でもある。

2.1 画像符号化

近年の動画像符号化の原点は1990年に標準化が終了したN-ISDN (Narrowband Integrated Service Digital Network) テレビ電話標準H.261にある。この符号化方式では画面 (フレーム) 内の画素値 (ピクセルの値) の相関を利用するDCTと前後するフレーム間での相関を利用する動き補償の組み合わせから構成されている。現在も我々がTV会議で利用する符号化方式の多くはH.261である。その後の符号化方式はH.261の系譜に属している。これを図1に示す。圧縮率は動画で1/50程度、静止画で1/10程度である。

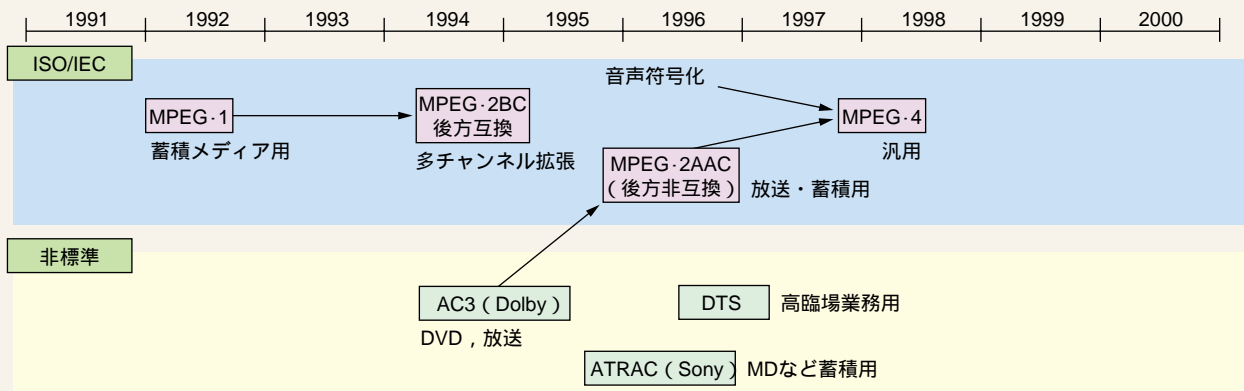
これはメディアが本来持っている冗長度に依存している。

H.261は実時間通信のために設計されたが、蓄積・放送の用途では、遅延時間が許される。この条件緩和を利用して動き補償が、前後するフレームを参照できる方式に拡張された。これがMPEG (Moving Picture Experts Group)・1である。MPEG・1はvideo CDなどに利用された。MPEG・1をベースにインタレース画像 (通常のテレビ放送は偶奇のフィールドで構成されている) を扱えるように拡張されたの



B-ISDN: Broadband Aspects of Integrated Services Digital Network
 IEC: International Electrotechnical Commission (国際電気標準会議)
 ISDN: Integrated Services Digital Network
 ISO: International Organization for Standardization (国際標準化機構)
 ITU-T: International Telecommunication Union·Telecommunication Standardization Sector (国際電気通信連合・電気通信標準化部門)
 JPEG: Joint Photographic Experts Group
 MPEG: Moving Picture Experts Group
 PSTN: Public Switch Telephone Network (公衆交換電話網)

図1 画像符号化技術



AAC: Advanced Audio Coding
 ATRAC: Adaptive Transform Acoustic Coding
 BC: Backward Compatible (後方互換性)
 DTS: Digital Theater System
 IEC: International Electrotechnical Commission (国際電気標準会議)
 ISO: International Organization for Standardization (国際標準化機構)
 MPEG: Moving Picture Experts Group

図2 音響符号化技術

がMPEG-2である。これにより、デジタルTV放送が可能となり、我々が眼にする多くの画像がMPEG-2で符号化されようとしている。2000年度ドコモサービス網で利用されようとしているMPEG-4は、これまでの符号化方式の集大成となっており、無線網で生じる耐誤り性能、インターネット

網の帯域変動に対するスケーラビリティ、コンピュータグラフィックスとの親和性などの拡張が施されている。H.261の系譜からは離れてJPEG (Joint Photographic Experts Group)*1は別の発展を遂げた。1992年に標準化されたJPEGはH.261と同じくDCTを用いた符号化方式であるが、まもなく標準化

を完了するJPEG2000にはサブバンド分割の変形であるウェーブレット変換が採用された。ここでは、符号化効率改善よりも解像度スケーラビリティなどの機能が重視されている。

MPEG-7は、圧縮技術ではなく

*1 JPEGもMPEGも本来は組織名である。転じて標準方式そのものを指す言葉となった。

XML (Extensible Markup Language) と呼ばれるメタ言語を用いてマルチメディアの中身、書誌事項を記述しようとするものである。

2.2 音響符号化

音響符号化は、後述の音声符号化と1980年代になり分化した。音響符号化と音声符号化も方式が大きく異なる。音声符号化が、発声による波形を符号化することに特化して進化したことに対して、音響符号化では、楽音など森羅万象の1次元連続波形を符号化することが必要であるためである。音響符号化では、前述の人間が許容できる歪みを最小可聴限とマスキングによりモデル化する。最小可聴限には低音、高音には聴覚感度が低いという耳の特性を利用する。マスキングは大きな音と周波数軸および時間軸で近傍音が聴き

取りにくい聴覚特性を利用する。これにより、圧縮率は1/5程度となる。

音響符号化の黎明期を作ったのがMPEG-1オーディオ^{*2}である。この次期は連続波形を滑らかに符号化するために、複数のフィルタを用いて入力波形を帯域の異なる信号に分割し、それぞれにビット配分を最適に行うという方式を用いていた。これは一般にサブバンド符号化と呼ばれている。ビデオ符号化がMPEG-2となり放送に適用されるに合わせて、MPEG-1オーディオも機能が拡張された。その結果、多チャンネル、多言語対応となった。このとき、MPEG-1オーディオとの互換性^{*3}を重視した方式がMPEG-2 BC (Backward Compatible) である。これと並行してDolby社が、サブバンド符号化によらない方式をAC3 (Audio Code Number3) として開発し

ていた。AC3はオーバーラップ窓を持つDCTであり、MPEG-2 BCよりも効率良かった。これは直感的にサブバンド符号化とDCT符号化の効率の違いとして説明できる。DCTは周波数分離特性、基底の直交性、波形再構成の完全性、以上の3点すべてを備えた変換であり、サブバンド分割はこの点で劣る。かねてより筆者はサブバンド分割を併用しないDCT符号化がオーディオ符号化に用いられると予測していたが、AC3の商用化がその端緒となった。MPEGではDolbyの提案を受けて、MPEG-2 AAC (Advanced Audio Coding) を完成させた^{*4}。MPEG-2 AACはサブバンド符号化ではなく、

*2 MPEGは蓄積・放送のためのマルチメディア情報の表現形式を標準化対象としているので、当然ながらオーディオ符号化も標準化方式として含まれる。

*3 Backward- Compatible 後方互換性と呼ばれる。

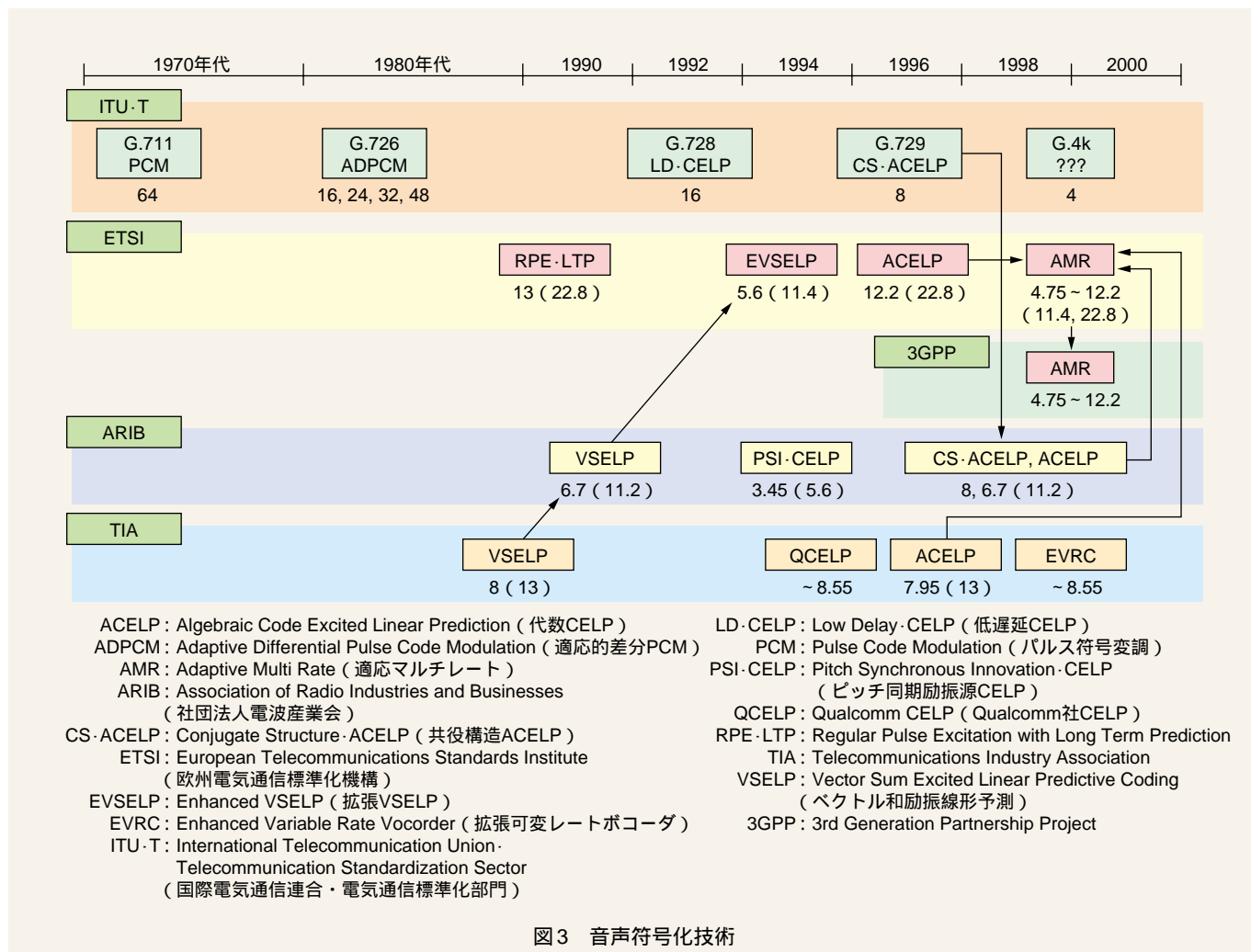


図3 音声符号化技術

DCT符号化とすることにより後方互換性を犠牲にしたが、Advanced Audio Codingの名が示すようにCD音質を達成するには最も効率の良い符号化方式として知られるようになった。

MPEG-4は、このMPEG-2AACを引継ぎ、さらに音声符号化方式を包含した汎用符号化方式として標準化されている。

2.3 音声符号化

通信基本サービスである電話のために特化された符号化技術が音声符号化である。このために音声符号化では、人の発声機構をモデル化し、声帯の振動、口蓋による変調が情報源モデルとして利用されている。図3に示すように、音声符号化では、ベクトル和励振線形予測（VSELP：Vector Sum Excited Linear Prediction）の登場以降、劇的に符号化効率が向上していることが分かる。それ以前のPCM、適応的

差分PCM（ADPCM：Adaptive Differential Pulse Modulation）符号化は発声機構のモデル化を行わない波形符号化技術であった。一方で1970年代から音源が簡単なパラメータでモデル化できる場合、偏自己相関（PARCOR：Partial Autocorrelation）や線スペクトル対（LSP：Line Spectrum Pair）により超低レート符号化が可能になることが知られていた。しかし背景音の混入、複数音源の存在など実用上では、モデル化が難しく、その品質に限界があった。VSELPをはじめとする符号励振線形予測（CELP：Code Excited Linear Prediction）符号化は以上の限界が、音源信号の極端なモデル化にあるという反省のうえに立って設計されている。音源信号の設計にモデル化できない雑音成分を含めることにより、波形符号化と上記音源モデル符号化の中庸を狙ったハイブリッド符号化である。CELP符号化のパリエーションの多くは、音源信号のモデル化の違いであり、パルス励振源のような音源を用

いることにより、少ない処理量で広い空間の探索を行う方法が、現在の音声符号化の主流となっている。IMT-2000では完成度の高いCELPである適応マルチレート（AMR：Adaptive Multi Rate）が用いられる。

現在の音声符号化の圧縮率は1/10程度であり、音響符号化よりも信号対雑音比（SNR：Signal to Noise Ratio）特性が悪い。

しかし、主観的音質は同等かそれ以上であり、その意味で効率が良いといえる。これは、発声機構のモデリングを利用していることによる。

3. システム技術

音声・音響・画像の符号化により基本ストリームと呼ぶデータ系列が生成されるが、それだけでは通信システムとしての機能は実現できない。基本ストリームのネットワーク適応（パケット化、通信路へのマップ、複数基本ストリームの多重化）と端末間シグナリ

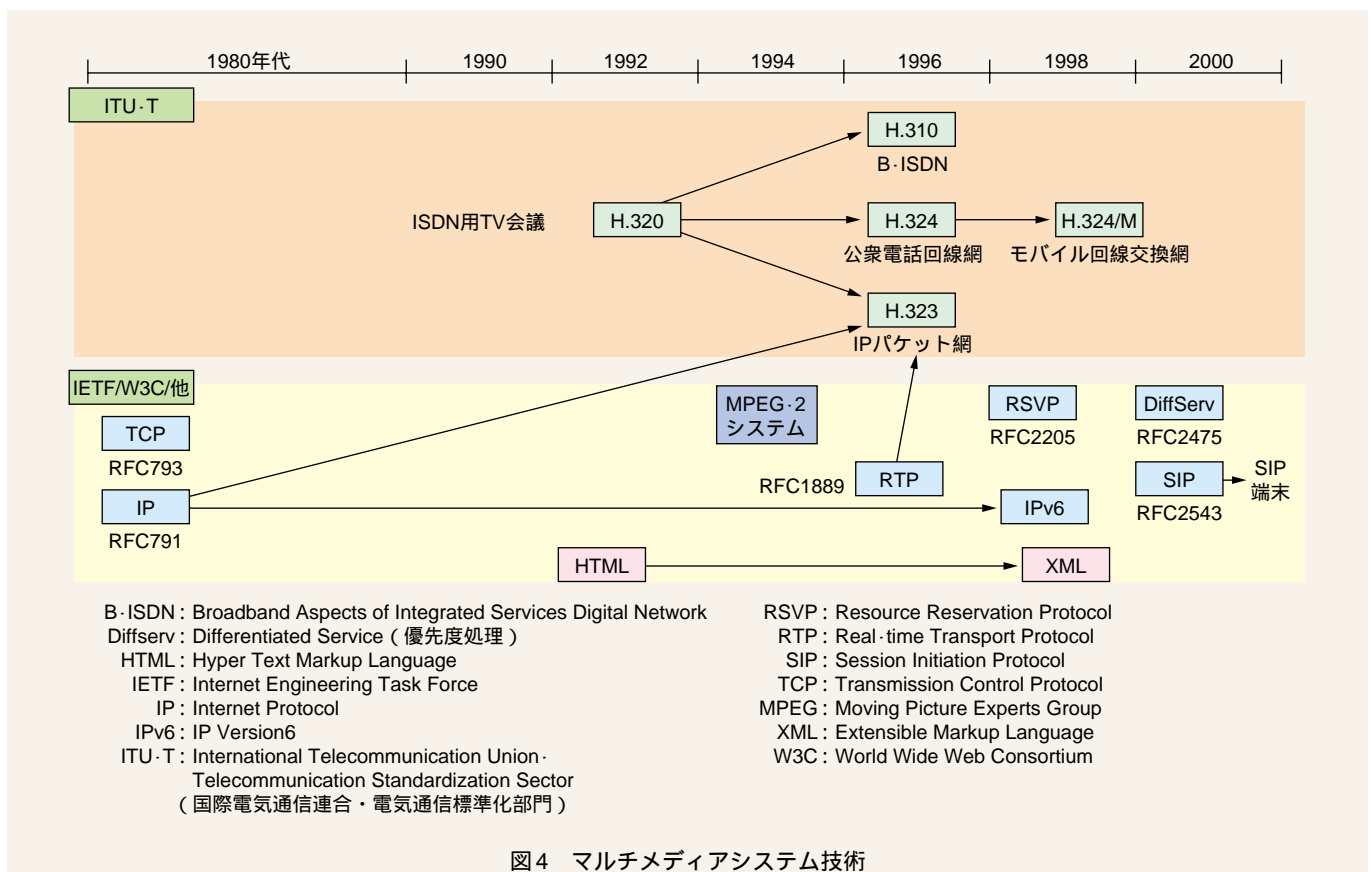


図4 マルチメディアシステム技術

ング（能力交換，同期）が必要となる．このような端末と配信に関する技術を標準化では“システム”と呼んでいる．このシステム標準化の経緯を図4に示す．

蓄積・放送分野ではMPEG-2システムが標準化されている．このシステム標準により，デジタルTV放送の複数番組の複数音響・画像データの多重化と同期が規定されている．一方で通信分野ではH.320が，国際電気通信連合・電気通信標準化部門（ITU-T：International Telecommunication Union・Telecommunication Standardization Sector）で1990年に制定された．これはN-ISDN用テレビ電話の勧告である．その後，B-ISDN（Broadband Aspects of Integrated Services Digital Network），公衆交換電話網（PSTN：Public Switched Telephone Network）およびIP（Internet Protocol）網に特化した端末・システムの検討が行われ，それぞれH.310，H.324，H.323として，1996年に制定されている．H.324は回線交換網の上に，効率良く音声・画像データを多重化し，伝送するシステムを定義している．これをベースに，移動通信の無線伝送に起因する伝送誤りへの耐性改善を目的として機能拡張を行ったものが3GPP（3rd Generation Partnership Project）規格3G-324Mである．IMT-2000でサービス予定されている映像端末・ビジュアルホンは，この3G-324Mに準拠した端末である．

IP上で通信端末のシステム規格を決めているのがH.323であり，インターネット電話の標準として知られている．

インターネットは，本来データ通信のために設計されたパケット交換網であり，データ通信プロトコルTCP/IPが1981年に制定されて以来，15年経て実時間メディア通信のためのプロトコルRTP（Real-time Transport Protocol）が制定された．これは近年のインターネットの普及によりマルチメディアデータをIP網で流通させる要求に沿った

ものである．インターネット電話を実現するうえで課題となるのが，遅延時間やパケットロスである．このために，IP上で帯域保証（RSVP：Resource Reservation Protocol），DiffServ（優先度処理）を行うとする動きがある．一方で，IETF（Internet Engineering Task Force）ではH.323と同等のシステムを実現するべく，HTTP（Hyper Text Transfer Protocol）上で新たな標準化，SIP（Session Initiation Protocol）が進められている．システム技術の進歩は著しく，IETF，W3C（World Wide Web Consortium）の動きは急であり，目が離せない．例えば，システムとメディアを繋ぐ技術としてXMLがある．XMLそのものは，マークアップ言語であり，マークアップ言語を定義するための言語にすぎないが，メディア記述（MPEG-7）やプロトコル記述を拡張性と互換性高く定義することができる．

4. 今後の展望

今後のマルチメディア符号化の課題を二つ挙げる．一つは，人が許容できるまたはできない符号化歪みとは何であるかをモデル化することである．計算コストへの懸念に臆することなく是非挑戦したいものである．もう一つは，分化した符号化アルゴリズムをどう最適に統合するかにある．音声・音響・画像符号化のアルゴリズムは各種存在し，それぞれに符号化速度，情報源に依存して得意・不得意分野があり，ユーザにとって使いやすいとはいえない．一方，システムに話を移すと，電話だけでなくすべてのメディア伝送をIP網上に構築する試みAll-IP化は，無線パケット網へもその適用が検討され始めようとしている．課題は，本来データ通信アーキテクチャであったIP網上でメディアを扱うことであり，ペイロードであるメディアとの親和性の良いインタフェースを議論する必要がある．

5. あとがき

本稿では，モバイルマルチメディア信号処理の概要を解説した．より深い解説が新世紀最初の本特集号の後続記事でなされるのでぜひ，一読いただければ幸いである．

用語一覧

AAC : Advanced Audio Coding	ISDN : Integrated Services Digital Network
AC3 : Audio Coder Number3	ISO : International Organization for Standardization (国際標準化機構)
ACELP : Algebraic Code Excited Linear Prediction (代数CELP)	ITU-T : International Telecommunication Union - Telecommunication Standardization Sector (国際電気通信連合・電気通信標準化部門)
ADPCM : Adaptive Differential Pulse Code Modulation (適応の差分PCM)	JPEG : Joint Photographic Experts Group
AMR : Adaptive Multi Rate (適応マルチレート)	LD-CELP : Low Delay-CELP (低遅延CELP)
ARIB : Association of Radio Industries and Businesses (社団法人電波産業会)	LSP : Line Spectrum Pair (線スペクトル対)
ATRAC : Adaptive Transform Acoustic Coding	MPEG : Moving Picture Experts Group
BC : Backward Compatible (後方互換性)	N-ISDN : Narrowband Integrated Service Digital Network
B-ISDN : Broadband Aspects of Integrated Services Digital Network	PARCOR : Partial Autocorrelation (偏自己相関)
CELP : Code Excited Linear Prediction (符号励振線形予測)	PCM : Pulse Code Modulation (パルス符号変調)
CS-ACELP : Conjugate Structure-ACELP (共役構造ACELP)	PSI-CELP : Pitch Synchronous Innovation-CELP (ピッチ同期雑音励振源CELP)
DCT : Discrete Cosine Transform (離散コサイン変換)	PSTN : Public Switched Telephone Network (公衆交換電話網)
DFT : Discrete Fourier Transform (離散フーリエ変換)	QCELP : Qualcomm CELP (Qualcomm社CELP)
Diffserv : Differentiated Service (優先度処理)	RPE-LTP : Regular Pulse Excitation with Long Term Prediction
DPCM : Differential PCM (差分PCM)	RSVP : Resource Reservation Protocol (帯域保証)
DTS : Digital Theater System	RTP : Real-time Transport Protocol
ETSI : European Telecommunications Standards Institute (欧州電気通信標準化機構)	SIP : Session Initiation Protocol
EVRC : Enhanced Variable Rate Vocorder (拡張可変ボコーダー)	SNR : Signal to Noise Ratio (信号対雑音比)
EVSELP : Enhanced VSELP (拡張VSELP)	TCP : Transmission Control Protocol
HTML : Hyper Text Markup Language	TIA : Telecommunications Industry Association
IEC : International Electrotechnical Commission (国際電気標準会議)	VSELP : Vector Sum Excited Linear Predictive Coding (ベクトル和励振線形予測)
IETF : Internet Engineering Task Force	W3C : World Wide Web Consortium
IMT-2000 : International Mobile Telecommunications - 2000 (次世代移動通信)	XML : Extensible Markup Language
IP : Internet Protocol	3GPP : 3rd Generation Partnership Project
IPv6 : IP Version6	