

高 能 率 音 声 符 号 化 技 術

早稲田大学 山 崎 芳 男

1. まえがき

CDや衛星放送とデジタル化された音響信号が身近な存在となり、デジタルと意識されることすらなく広く使われるようになった。また電話の番号案内をはじめ玩具にまでデジタル化された音声サービスが使われたりと、12年前に本誌に高能率符号化の解説を書かせていただいた(「音響におけるデジタル技術入門」日本オーディオ協会誌 Vol. 17, No. 3)当時とは隔世の感がある。固体レコードや移動体を対象としたデジタル放送も現実のものとなりつつある今日、もう一度音響信号の符号化、とくに高能率符号化について考えてみることにしよう。

<節約したうえ質も良くする>

高能率符号化は文字通り何らかの手段により音響信号を出来るだけ能率良く伝送するための符号化技術である。一般に一旦デジタル化した信号をどこまで圧縮できるかといった観点から一段下の質を考えがちであるが、高能率符号化はそのような後ろ向きな?技術ではなく、むしろ伝送路は節約したうえ、質の向上を図ろうという前向きの符号化として捉えるべきものである。

<長島茂雄のあの独特の声>

長島茂雄があの独特の高い声で「いわゆる……」といった。と書けば彼、いや彼の声を知る多くのひと

はおおよその音を容易に思いうかべることができる。また信号を実際に送らなくとも山口百恵のCD35DH 51の何曲目と指定すれば、そのCDを再生することにより(なければ買ったうえで)、たった数文字分せいぜい100bitで膨大な信号を符号化、伝送したことになる。このように受信側に人間並、放送局のレコード室並の記憶・ライブラリーを用意すれば超高能率符号化が期待できる。今日の記憶媒体の驚異的高密度化と価格や消費電力の低減傾向をみると、案外近い将来、受け側の大容量記憶を前提とした伝送も実現するかも知れない。

<活字に相当する規格化音声>

合成音声の普及とともにこれを嫌う傾向も一部に根強いようだ。しかし文字において、洋の東西を問わず手書き文字と活字がうまく使い分けられているように、音響情報においても肉声と合成音声が共存する時期は遠くあるまい。もっとも日本でもワードプロセッサによる私信の是非の議論などあるようではあるが。

ともかく、すでに一部で実用化されているように速報性を要求されるニュース放送や各種の案内等にまず導入されることとなる。この場合、文意を解釈した自然な発声は不可欠な条件である。当然、活字の字体のようにいくつかの標準形が規格化されよう。

2. 音響信号の符号化

音響信号を情報理論的な見地で眺めてみよう。

<シャノンの美しい定理>

シャノンは信号の平均情報量 H と伝送路の伝送容量 C の関係が $H < C$ ならば無ひずみ伝送が可能であり、 $H > C$ ならばひずみを限りなく $C - H$ に近付けて伝送することが可能な符号化法が必ず存在するという美しい定理を導いている。

<無駄を省くか中身を削るか>

高能率符号化の導入に当たってはシャノンの定理に基づき信号の持つ冗長度を除去することにより、信号を劣化させることなく伝送路の節約を図る方法をまず検討すべきである。

信号の平均情報量と伝送容量が等しい状態、無ひずみ伝送の限界をレートひずみ限界(Rate Distortion Boundary)と呼ぶ。この限界をこえると大幅な伝送路の節約が可能である。たとえば20%のひずみを許容すると約50%の伝送路の節約が可能である。そこで人間の聴覚の性質を利用して、ある程度の信号劣化を許容して信号を大幅に圧縮する方法が考えられる。

<デジタル化された音響信号>

高品質音響信号の周波数帯域は15~20kHz、ダイナミックレンジは80~100dBにおよぶ。この信号を基本

の数bitに限定するといった符号誤りによる雑音が著しく不快な音にならないような配慮も必要となろう。また、冗長さの除去と同様に、高度な誤り制御は遅延が多くなるので、放送の誤り制御については、遅延の許容量も考慮する必要がある。

<盾と矛>

ところでデジタル信号処理のシステムパラメータの決定に当たっては、前述の信号と誤り制御の割り振りのように、しばしば立場の違いによる“矛盾”を感じることもある。

例えば高効率符号化を追求する立場にたてば、100kbit/sいやISDNのビットレート64kbit/sで十分高品質の音を送ることができると主張したい。一方、伝送路や記録媒体を使う立場にたてば、当然大きな入れものが欲しいと要求したくなる。少なくとも音響計測やソフトウェア制作には1チャンネル当たり1Mbit/sは欲しいといった具合に。

しかし、これはなにも矛盾した話ではないのである。なにが現れるか判らない最初の情報源符号化は十分大きなビットレートで行い、その後人間の聴覚機構に鑑み必要十分な信号に再符号化すればよいわけである。もちろんこれは同時に行ってもかまわない。

<角張ったデジタルだからこそ

自由度の大きな符号化を>

標準化周波数が帯域を決め、量子化特性がダイナミックレンジを決めるととらえられがちであるが、これは間違いであり、実は両者は密接に関係がある(本誌'84年4月号「大振幅ディザと高速標準化によるAD変換精度の向上」参照)。ディザの導入などの確な処理が施されていれば

量子化雑音は標準化周波数の1/2の帯域に一樣に分布する。最近のΣ-Δ型のように何らかの方法で量子化雑音のスペクトル分布を高域にかたよらせることもできる。したがって、たとえ1ビットでも標準化周波数を十分高くとれば100dB以上のダイナミックレンジを音響信号領域では確保することも可能である。

最近われわれの研究室では高域に集中したディザを導入したΣ-Δ型AD変換器を並列動作させることにより得られた高速(500k~2MHz)、1bitのデジタル信号をそのまま録音や測定に使っている。標準化周波数による高域の制約から開放されたアナログのセンスで扱うことができる自由度の大きい符号化方法である。DA変換も高速のまま行っているが、一般に使われている48kHz、16bit等の信号が必要な場合に初めてデジタルフィルタリングにより所望の信号に変換してやればよいわけである。

3. 各種の高効率符号化

表2に各種の高効率符号化方法を示す。前述のように、高効率符号化には信号のもつ冗長性を除去する信号劣化のない符号化方法と、聴覚特

性を利用してある程度の劣化を許容して信号を圧縮する方法、両者の組み合わせなどが考えられる。

また、直接時間領域の信号を符号化する方法と、一旦何らかの変換を行う変換符号化とがある。

<瞬時圧伸と準瞬時圧伸>

楽音や人の声などの音響信号の振幅は、時間とともに大幅に変化することはよく知られている。瞬時圧伸はこの性質を利用して変調時に振幅

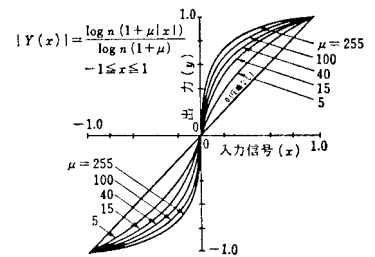


図3 μ法則の圧縮特性

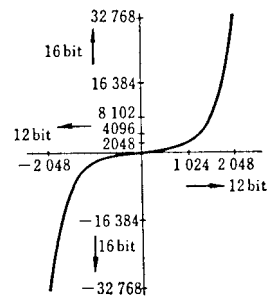
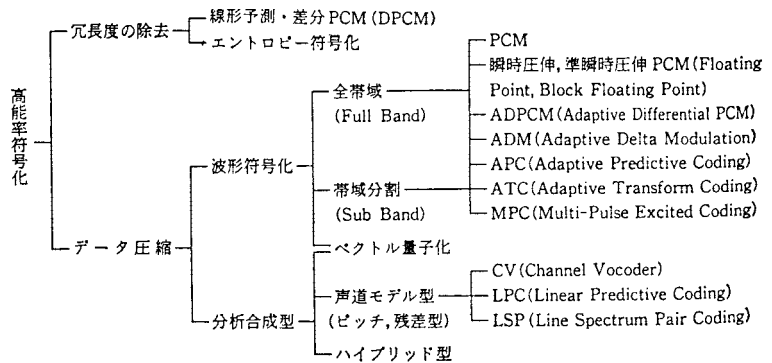


図4 DATの12bit 16bitの伸張特性

表2 高効率符号化の分類



を圧縮して、復調時に伸良する聴覚特性を利用した符号化である。アナログ伝送でも使用される方法で、早くから広く使われていた。

A/D・D/A変換段の圧伸は、量子化ステップを振幅に応じて変化させるので非一様量子化と呼ばれる。小振幅時には量子化ステップ幅を小さく、大振幅時には大きく設定する。これは 3.3×10^3 といった書き方をするのに相当するので、浮動小数点方式と呼ばれることもある。

代表的な瞬時圧伸則には対数特性を近似したA法則や μ 法則とよばれるものがある。図3に μ 圧縮則(伸張則は逆のカーブとなる)を示す。 μ 圧伸は電話の音声のPCMに広く使われており、またデジタル中継回線にはデジタル段で法則を近似した瞬時圧伸方式が導入されている。DATの長時間モードや8ミリビデオのPCM音声にもデジタル圧伸が使われている。DATでは16bitの信号を12bitにデジタル段で圧縮し、図4に示すように13折線の形を呈している。

<準瞬時圧伸>

ところで圧伸は瞬間瞬間の振幅に応じて圧縮伸長を行う瞬時圧伸方式と、ある範囲の信号を観察したうえ行う準瞬時圧伸方式とがある。

準瞬時圧伸は音響信号のレベルの変動速度が比較的ゆるやかな点に注目して、信号のある長さのブロックに分け、ブロック単位で観察して圧伸操作を行うものである。準瞬時圧伸されたデジタル信号は振幅の小さいブロックでは量子化ステップが小さく、大きなブロックでは量子化ステップが大きくなるのでブロック浮動小数点方式と呼ばれることもあ

り、能率の良い伝送方式である。あまりブロックを長くしすぎるとレベルの小さい部分でのSN比が問題となる。準瞬時圧伸は衛星放送のAモードやFM中継等に使われている。

<予測符号化>

予測符号化は図5に示すようにいくつかの標本値から現在の標本値を前向き、あるいは後向きに予測推定して、真の標本値と予測値との差(予測誤差という)を符号化して伝送するものである。

1標本前の値を予測値とする単純な差分PCMでも通常の楽音に対し1標本(16bit)当たり8~9bit、本格的な予測符号化では1標本当たり4~5bitで信号劣化のない伝送が可能である。ある程度ひずみを許容すれば、さらに大幅な伝送容量の

圧縮が可能である。

図6に差分PCMによる波形の変化と各ビットの使用率を示す。オーケストラのように比較的ゆっくり変化する信号に対しては、差分PCM信号にもう1度差分処理を施した2階差分により、さらに大幅なビット節減が可能である。逆にジャズアンサンブルのように動きの早い信号に対しては予測がむずかしく、あまり大幅な節減は期待できない。

<エントロピー符号化>

- 珍しい情報は長く、平凡な情報は短く -

図1に示したように音響信号は一般に大振幅に比べて小振幅が多く出現し、冗長性を取り除いた後ではさらにその傾向が強くなる。図7に6項の予測符号化を行った残差のレベ

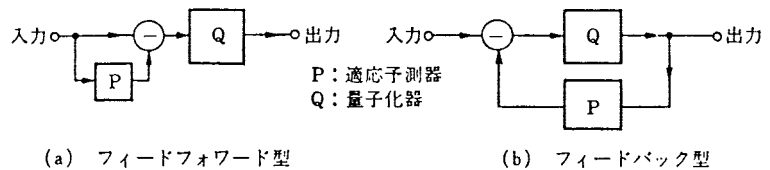


図5 予測符号化の構成

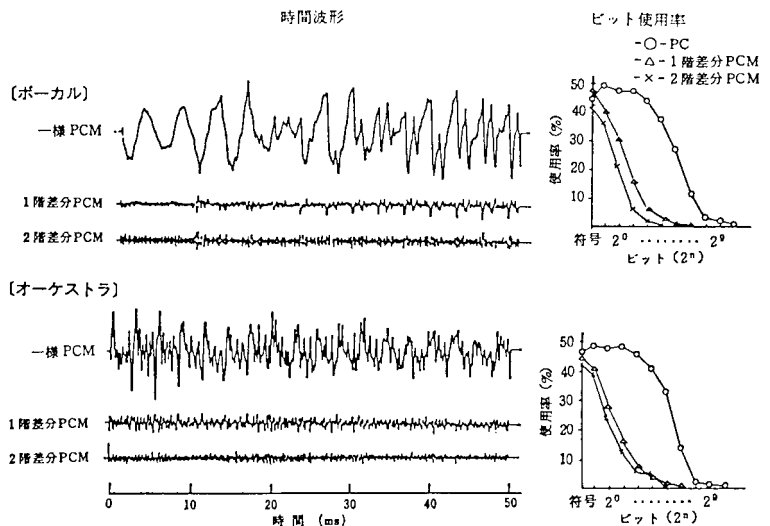


図6 差分PCMによる波形と各ビットの使用率

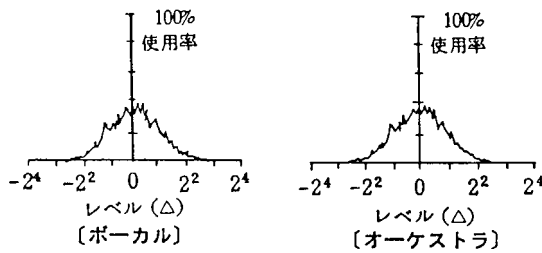


図7 予測符号化における残差成分のレベル分布

ル分布を示す。図1の原信号では8～10bitに広く分布していたが、図7では分布は低位の4～5bitに集中している。それともない低位ビットの使用率が大幅に増加している。そこでモールス符号のように出現頻度の高いレベルに短い符号、出現頻度の低いレベルに長い符号を与えることにより、伝送ビットレートをさらに節約することが可能である。このような符号化をエントロピー符号化と呼ぶ。前述のオーケストラの2階差分PCMのレベル分布に基づき割り当てた不等長符号（ハフマン符号）を適用すると、1標本当たりの平均情報量は2.979bitとなる。これはもともと16bitの信号が予測符号化により無ひずみで5bitに、さらにこの不等長符号によりひずみを与えることなく3bitで伝送可能なことを示している。

ただしエントロピー符号は不等長符号であるため、大容量の一時記憶素子と辞書（符号割当表）が必要である。伝送に使用する場合は大幅な

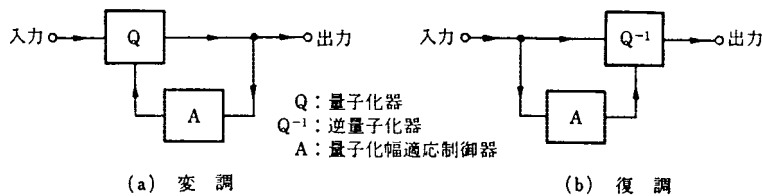


図8 APCMの構成

時間遅れが生ずる。

4. 適応型符号化

< 適応型 PCM

(APCM) >

前述のように音響信号の振幅や周波数分布は時間とともに比較的穏やかにでは

あるが大幅に変化する。この性質を利用して近傍の信号の性質に応じて量子化ステップ幅を変化させる符号化が適応型PCM (APCM)である。準瞬時圧伸やダイナミックエンファシス是一種のAPCMとみることができる。

図8に適応型PCM (APCM)の構成例を示す。APCMでは直前の符号により次の量子化ステップ幅を決定している。表3に3bit APCMの量子化ステップの係数の例 (Jayant らによる) を示す。この例では、直前の標本の正負を表す1bitを除いた他の2bitの量子化値の振幅が00と01の場合には量子化ステップ幅に0.9を乗ずることにより量子化ステップ幅を小さくし、10

表3 3bitのAPCMの量子化ステップの与え方の例

	符 号	係数P
正 の 値	0 1 1	1.75
	0 1 0	1.25
	0 0 1	0.9
	0 0 0	0.9
負 の 値	1 1 1	0.9
	1 1 0	0.9
	1 0 1	1.25
	1 0 0	1.75

の時には1.25, 11の時には1.75を乗ずることにより量子化ステップを大きくしている。このように適応型符号化では、たとえ量子化ビット数が少なくとも、ある程度時間をかければ様々な量子化ステップ幅が得られ、ダイナミックレンジはいくらでも拡大することが可能である。

< 適応型差分PCM (ADPCM) >

適応型APCMは差分PCM (DPCM)に適応形ステップ幅を導入したものである。すなわち、信号 $x(n)$ を直接量子化するのではなく、予測値 $\hat{x}(n)$ との差 $d(n)$ を適応量子化するもので、前述のAPCMよりも能率がよく中程度の質の高能率符号化

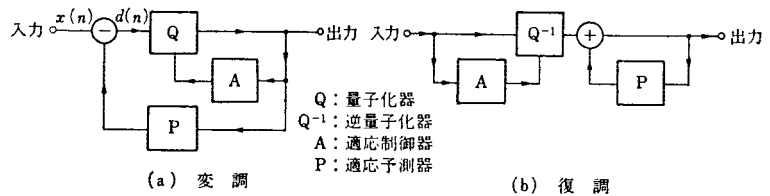


図9 ADPCMの構成

として大変有効な手法である。CD - や多機能電話機のメッセージ録音などの短時間の、テープ等を使わない固体録音機や案内放送、自動販売機等の各種の音声サービスに盛んに使われている。

図9にADPCMの構成、表4に

表4 適応量子化法におけるステップ幅係数 (Jayant による)

ビット数	APCM	ADPCM
2	0.6, 2.2	0.8, 1.6
3	0.85, 1, 1, 1.5	0.9, 0.9, 1.25, 1.75
4	0.8, 0.8, 0.8, 0.8, 1.2, 1.6, 2.0, 2.4	0.9, 0.9, 0.9, 0.9, 1.2, 1.6, 2.0, 2.4
5	0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 0.85, 1.2, 1.4, 1.6, 1.8, 2.0, 2.2, 2.4, 2.6	0.9, 0.9, 0.9, 0.9, 0.95, 0.95, 0.95, 0.95, 1.2, 1.5, 1.8, 2.1, 2.4, 2.7, 3.0, 3.3

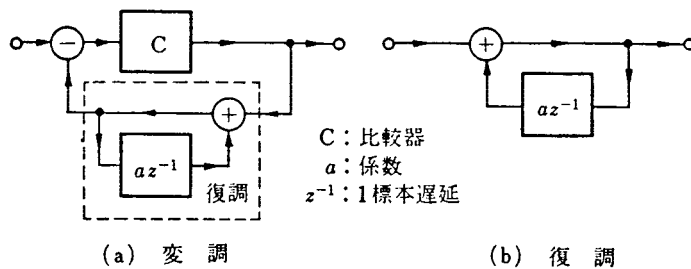


図10 ΔM の構成

2 ~ 5bitのAPCMとADPCMの係数の例を示す。APCMと同様に直前の符号の大小により量子化ステップ幅に乗ずる係数が異なり、次の量子化ステップ幅を決定する。

< ΔMと適応型 ΔM >

ΔM は信号を1bitで量子化する符号化方法である。図10に最も簡単なΔMの構成例を示す。このΔMは

DPCMの量子化ステップ幅の1bitにしたもので、標本値が1つ前の標本値より大きい場合は1, 小さい場合は0とする簡単な符号化方法である。

原理から明らかなように信号が急速に変化する場分には大きなひずみが生ずる。このひずみは標本化周波数を高く設定すれば減少するが、あ

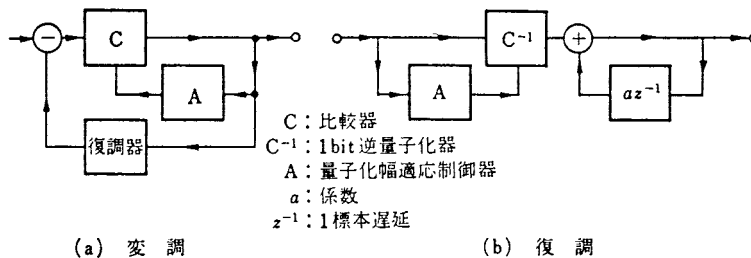


図11 適応型 ΔM の構成

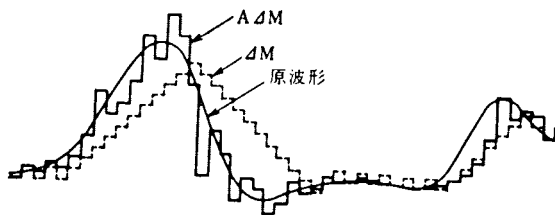


図12 ΔM と適応型 ΔM の波形

まり高く設定したのでは節約の意味がなくなってしまう。例えば標本化周波数を10倍に設定しても元の標本化間隔に10量子化

ステップ幅の変化にしか追従できない。PCMでは同じ10bit/標本で1,024量子化ステップ幅の変化にまで追従可能である。

適応型 ΔMは図11のように量子化ステップ幅を同じ符号が続く場合は1.5, 反転する場合は0.8というように適応させる1bitのADPCMである。図12はΔMと適応型 ΔMの信号への追従の様子を比較したものである。適応型 ΔMはハードウェアの構成が簡単な割にはよく原波形に追従しており、簡単なシステムには適した符号化方法である。

< 各種の符号化の比較 >

表5にオーケストラの演奏について行った各種の圧縮手法のSN比(原音との標本ごとの差のdB値の平均)と5段階の主観評価の結果を示す。表に示すように1標本当たり4 ~ 5bitすなわち, 129 ~ 200kbit/sで実用になる伝送が期待できる。

伝送の対象を音声に限定すれば, 前述の各種の高効率符号化や変換符号化, ベクトル量子化等の導入により200bit/s程度までの大幅な信号圧縮も可能である。

< 高効率符号化の規格 >

表6に規格化あるいは提案されている高・中品質音響信号を対象とした各種の符号化方法を示す。

なおCDの応用分野であるCD-に音楽の長時間化の手法として4 ~ 8bitのADPCMが採用されているが, これはむしろ294標本ごとに予測フィルターと量子化ステップ情報を伝送するブロック符号と呼んだ方が妥当だろう。

5. 音声の符号化

音声の分野では早くからディジタ

表5 各種圧伸手法（オーケストラ，標準化周波数32kHz）

手 法	4 bit/標本		5 bit/標本	
	SN比[dB]	主観評価	SN比[dB]	主観評価
μ圧伸	9.9	1	14.3	2
準瞬時圧伸	12.6	1	18.4	2
APCM	16.1	2	21.1	3
ADPCM	31.1	4	36.3	5

ただし，評価は5段階。1：許容できない～5：原音と遜色なし

表6 各種の符号化法

方 式	標準化周波数 [kHz]	量子化ビット数	信号帯域	伝送レート [kbit/s] 1CH当たり	特 徴
32k ADPCM (CCITT G721)	8	4	300Hz～3.4kHz	32	適応予測 2次IIR+6次FIR 適応量子化 DLQ(Dynamic Locking Quantizer)
64k Sub-Band ADPCM (CCITT G722)	16	4	50Hz～7kHz	64	Sub-Band符号化 低域(50Hz～4kHz) 6/5/4bit量子化 高域(4～7kHz) 2bit量子化
384k オーディオ符号化(CMTT)	32	11/10	50Hz～	384	14/11bit A則瞬時圧伸 14/10bit 準瞬時圧伸 (32サンプルごと)
衛星TV規格	32	14/10	50Hz～15kHz	512	14/10 準瞬時圧伸 14bit 直線
NHK衛星放送 Aモード	32	14/10	50Hz～15kHz	768	14/10 準瞬時圧伸
FM東京 FM多重放送	8	4	300Hz～3.4kHz	32	CCITT 32K ADPCM 準拠
CD-I	37.8	8	17kHz	309	ADPCM
	37.8	4		159	
	18.9	4	8.5kHz	80	
8mm VTR音声	31.5	8	50Hz～15kHz	354	アナログ対数圧伸(2:1) デジタル 10→8 bit 9折線圧伸
放送中継	32	11	50Hz～15kHz	768	13→11bit 7折線準瞬時圧伸

ル処理が導入されPCM - 24方式のような非一様量子化を取り入れた8kHz標準化，8bit量子化のPCM伝送が電話の局間中継に使われている。限られた伝送路を有効に利用すべく各種の高エネルギー符号化が使われている。また，伝送対象を単一話者の

音声に限定すること，さらに人間の発声機構に着目すること等により極く低ビットレートの伝送も試みられている。

< LPCボコーダー >

LPC (Linear Predictive Coding) 線形予測符号化は前述の予測

符号の一種であるが，ボコーダーは分析合成方式に属する音声符号化方式で，人間の音声の発生過程をシミュレーションした符号化方法である。

音声信号は有声音と無声音とに区別され，有声音の場合にはピッチ周期に対応したパルス列が，また無声音の場合には白色性雑音が声帯で生成される。この声帯で生成された信号で，声道をシミュレートする適応フィルターを駆動することにより合成音声を得るものである。したがって，有聲/無聲の区別，ピッチ，適応フィルターの係数のみを伝送すればよいので超低ビットレートで音声信号の伝送が可能である。

< MPC >

LPCボコーダーは，音声を有声音，無声音に区別するため，実際の音声で有声音と無声音が明確に区別できない部分では復号信号がいわゆる機械合成音的になる。MPC

(Multi Pulse excited Coding) は，LPCボコーダーを原形としながらも，声道をモデルとした適応フィルターに入力される音源をパルス列のみとし，パルス列のフィルター出力波形が，もとの音声信号になるべく近くなるよう，その位相，振幅をも制御するものであり，分析合成方式的な波形符号化を行っているものである。したがって，LPCに比べ，再生音声の忠実度が飛躍的に優れている。

< SB - ADPCM >

信号を帯域分割して符号化する方法である。それ自体でも高エネルギー音声符号化の一方式であるが，他の高エネルギー方式と組み合わせることでより有効な符号化となる。

SB - ADPCM (Sub - Band ADPCM)

M) サブバンド符号化は、例えば7 kHz 帯域の音声信号を高域、低域2つのサブバンドに分割し、それぞれADPCM方式で符号化する。この場合、音声エネルギーが多く存在する低域には情報量を多く割り当て精度の高い符号化を行い、音声エネルギーの少ない高域には情報量を少なく割り当てて符号化を行う。こうして7kHzの帯域全体として見たときの、高能率化と高品質化を実現しようとするものである。

他の音声符号化方式のような、電話用としてではなくテレビ会議用音声等へ応用が考えられている。本方式はCCITT標準方式として勧告されている。

< APC - AB >

APC - AB (Adaptive Predictive Coding with Adaptive Bit allocation)は音声信号を一定時間ごとのブロックに分け、ブロックごとの予測係数、ピッチ周期、ピッチ振幅の変化を求めて伝送を行うAPC符号化に、サブバンド符号化を組み合わせた方式である。音声信号は例えば3つの帯域に分けられ、それぞれAPC符号化が行われる。また、それぞれの帯域に存在するエネルギー分布に応じて適応的に情報量の割当を行い品質の向上を図っている。

6. む す び

以上高能率符号化について述べたが、音響信号のデジタル化に際し標準化周波数と量子化ビット数をプ

ログラムソースによらず一定の値に規格し、品質向上にはより高い標準化周波数とより多い量子化ビット数、すなわち伝送容量の拡大を求めざるを得ない従来の符号化を見直して、そこそこの伝送容量で人間の聴覚を十分カバーする自由度の大きい高能率符号化が検討されてもよい時期にきているのではないか。

伝送路の節約については多少夢物語めくが、冒頭に書いた“長島茂雄があの特徴の高い声で「いわゆる……」といった”。あるいは山口百恵のCD35DH51の3曲目と伝送するように実際に信号を送るのではなく受け側の大容量記憶を前提とした符号化方法の実現も遠くないのではないか。