移動通信の基盤技術 その1

1 高能率音声符号化の展望と課題

高能率音声符号化は，セル方式のデジタル移動通信システムを支える基盤技術である．本稿では，ここ数年で急速に進歩し実用段階に入った高能率音声符号化について，主な基本技術を説明するとともに，移動通信における実用例を紹介する．また，将来展望と解決すべき課題について述べる．

三木 俊雄（みき としお）

まえがき

高能率音声符号化は，セル方式のデジタル移動通信システムを支える基盤技術として，ここ数年で急速に進歩し実用段階に入った．移動通信分野では，符号化ビットレートの低減がシステムの大容量化と携帯電話機のバッテリ長寿命化に直接結びつくため，低いビットレートで高い音声品質を実現する高能率音声符号化の研究が極めて活発に進められている．

音声符号化の原点はPCM*であり，64kb/s程度のビットレートを必要とする．これに対し，デジタル移動通信の分野で主流であるCELP(Code-Excited Linear Prediction)系が数kb/sであることをみると隔世の感がある．この間，音声の冗長度を圧縮する多くの画期的な発明・発見がなされてきたことも事実であるが，残念なことに，他分野の技術者にとって音声符号化は難解でとっつきにくいとの印象を与えてしまったことも見逃せない．

本稿では，高能率音声符号化を少しでも多くの方々に理解いただくことを狙いとして，音声符号化研究の流れに触れつつ，主な基本技術を極力平易に説明したい．この中では，音声の冗長度圧縮のみならず，移動無線チャネルでは不可避の符号誤りに対する保護や，実現のための回路技術であるデジタル信号処理プロセッサ(DSP:Digital Signal Processor)についても述べることとしたい．

また，セル方式のデジタル移動通信システムで実際に使われているVSELP (Vector Sum Excited Linear Prediction) やPSI-CELP (Pitch Synchronous Innovation CELP)などの実用例を取りあげ，これらの基本技術がどのように活用されているかを示す．さらに，高能率音声符号化の研究開発の将来展望と解決すべき課題について触れる．

研究の流れ

電話用に研究開発されてきた主な音声符号化アルゴリズムを，ビットレートと標準化された年次に従って図1に示す．

固定電話網と移動通信網を適用領域とする2つの大きな流れがあることがわかる．

固定電話用の音声符号化アルゴリズムは，固定電話網のデジタル化に向けて，ITU-TS(旧CCITT*) における国際標準化を中心に進歩してきた．その特徴は，低遅延，高音質，データモデムに対する透明性などである．これらの要求条件を満たすものとして，64kb/sPCM[1]，32kb/sADPCM[2]が標準化された．

その後は，広帯域性を有する光ファイバの普及により低ビットレート化への意欲が薄れ，主たる適用領域がマルチメディア対応と移動通信への応用という方向に変化してきた．そこで，16kb/s LD-CELP方式*,[3]の標準化では低遅延，高音質のみを要求条件とし，さらに8kb/s 方

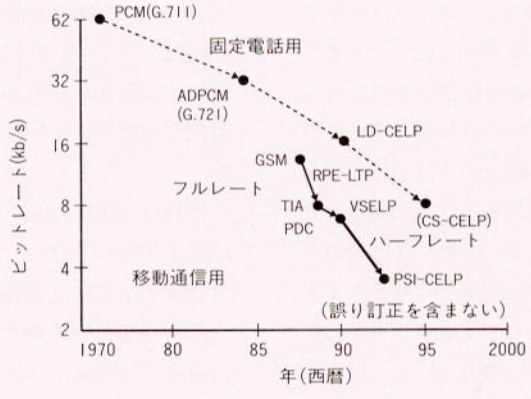

図1 音声符号化方式標準化の流れ

式の標準化では，遅延時間を10ms程度まで許すことになった．

これに対し，移動通信用の音声符号化アルゴリズムは，セル方式のデジタル自動車・携帯電話システムへの適用をねらいとして，日米欧の各地域で別々に標準化が行われてきた[4)~6)]．その特徴は，アナログ方式と同程度以上の音声品質，低ビットレート，誤り保護を含めた最適設計などである．また，エコーキャンセラの適用を前提に，ある程度の遅延時間を許容している．

低ビットレート化の勢いは急激で，ハーフレート化されたPDC*方式では，ついに3.45kb/sにまで至った．なお，標準化では符号誤りに対する耐性も評価されるため，全体の符号化ビットレートとしては2.15kb/sの誤り訂正符号化を含め，5.6 kb/s 方式として最適化されている．また，携帯電話機を実現するため，高速でかつ低消費電力のDSP が精力的に開発されている．

このように，現在では高能率音声符号化の主たる適用領域はセル方式の自動車・携帯電話であり，高音質でかつ符号誤りに耐え得る音声符号化と，これに適した誤り保護，その回路を実現するDSP が重要な基本技術となっている．

表 1　音声符号化の要素技術

技　術	説明・効果
ノイズキャンセラ	背景・周囲雑音の軽減
アダプティブコードブック	音声のピッチ（基本周期）成分の再現性向上
直交化処理	アダプティブコードブックと雑音コードブックを互いに直交化 独立探索により，コードブック探索の演算の演算量低減
多重ベクトル量子化	複数のベクトルの和で雑音コードブックを実現 符号誤りに強いCELP構造
コードブック予備選択	合成フィルタ前にコードブックを絞り込み コードブック探索・合成フィルタの演算量低減
聴覚重みフィルタ	聴覚の周波数特性に合った重みづけ 聴感雑音の低減
ポストフィルタ	主要周波数成分の強調 張りのある自然な音声再現
再マッピング（コード変換）	符号誤りの影響を軽減する符号化コードの変換

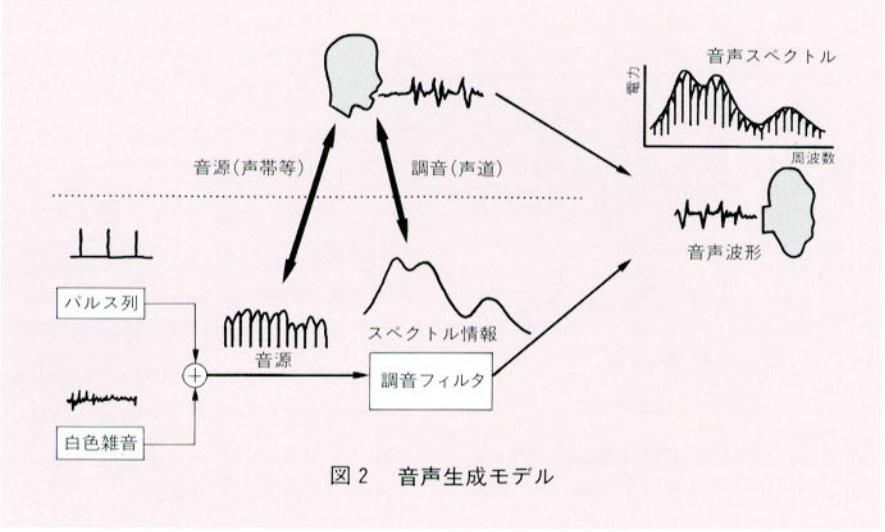

図 2　音声生成モデル

基本技術

本章では，高能率音声符号化の基本技術と，誤り保護，DSP について解説する．なお，基本技術を支える多くの要素技術については**表 1** にまとめて概説した．

■音声符号化の基本技術

(1)　音声生成モデル

音声の基本的な生成モデルを**図 2** に示す．音源には，声帯の振動のように規則的なパルス列で表せるものと，唇や舌の摩擦・破裂といった白色雑音で表されるものがある．調音フィルタは，声道（顎，舌，口腔など）の形を変化させることにより，細かな音色を調整する．音源・調音フィルタは時間とともに変化するが，20～30msの間はほぼ一定とみなして良い．高能率音声符号化では，音源と調音フィルタの特性を少ない情報量で効率的に表現することにより，原音声に近い復号音声が得られるように工夫している．音声生成モデルの詳細は文献 7)を参照されたい．

(2)　線形予測分析

通常，音声信号は**図 3** に示すように比較的ゆっくりと変化しているため，現在および過去の信号から将来の信号をかなり高い精度で予測することができる．予測のために現在および過去の信号に乗ずる係数を予測係数と呼び，予測係数を持つ回路を予測フィルタと呼ぶ．また，予測値と音声信号の差を予測残差と呼ぶ．多くの高能率音声符号化では，予測フィルタを上述の調音フィルタ，予測残差を音源とみなしている．

線形予測分析は，音声信号の相関特性などから予測係数を抽出し，調音フィルタを能率良く実現する方法である．実際には，予測係数をそのままの形で符号化して伝送せず，PARCOR* やLSP* といった数学的に等価で安定な別のパラメータに変換して伝送するのが得策である．予測係数が符号誤りに弱く，少しの誤差で音声品質が大きく劣化したり，場合によっては発振するためである．

(3)　マルチパルス駆動

図 4 に予測残差の一例を示す．大きなパルスと小さな雑音の和に見える．パルスの位置と大きさで予測残差，すなわち音源を表現する方法がマルチパルスである[8)]．符号化に与えられるビット数には限

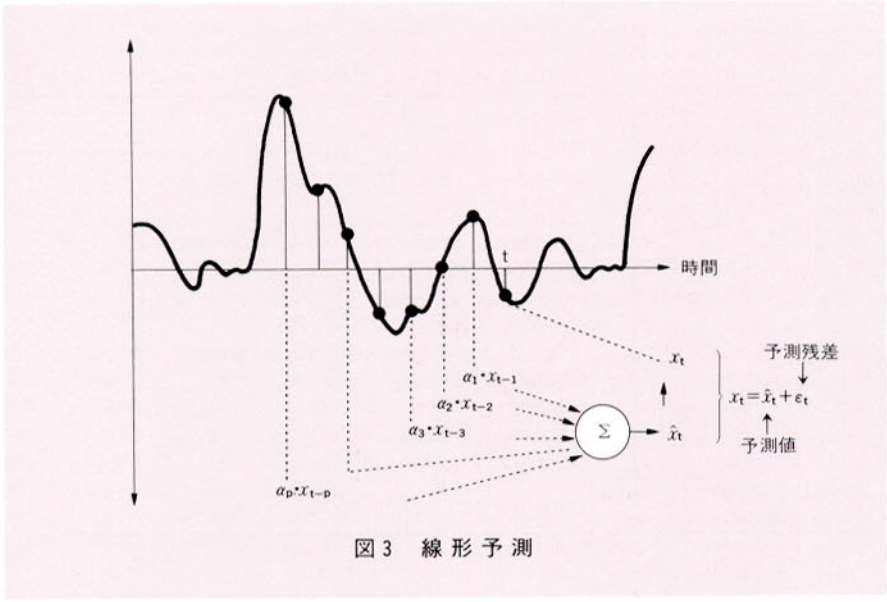

図3　線形予測

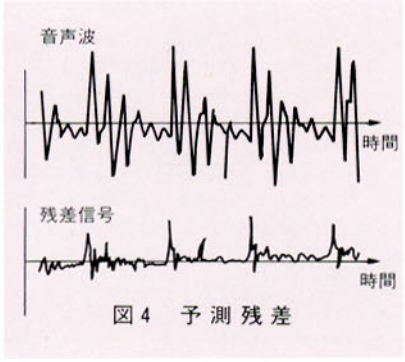

図4　予測残差

りがあるため，有限個の代表的なパルスのみを用いることが多い．また，個別パルスの代わりに，周期的に繰り返すレギュラーパルスを用いる方法もある．

高い音声品質が実現できることから一時期盛んに研究されたが，パルス探索の演算量が多いことや符号誤りに弱いなどの理由で，現在では次に述べるCELPに主役の座を譲っている．

(4)　雑音符号帳駆動

マルチパルスが音源をパルスの組合せで表現しているのに対し，CELPは符号帳と呼ばれる雑音系列の辞書の中から，音源として最もふさわしいものを見つけ，その番号を伝送するという方法である[9]．**図5**に符号器の構成を示す．

次節で説明する合成による分析法（A-b-S*法）を用いることから演算量が多く，CELP が提案された当初は実現不可能と考えられていた．しかし，様々な演算量削減方法が考案され，現在では高能率音声符号化の主流となっている．また，符号誤りが生じると全く異なる雑音系列を選択してしまうので，符号誤りに弱いとみられていたが，多重ベクトル量子化などを用いることにより，ADPCM よりむしろ符号誤りに強いこともわかってきた．

(5)　合成による分析法（A-b-S 法）

CELPやマルチパルスといった音源をモデル化する符号化方式の場合，最適なパルスや符号帳の番号を求める方法として次の２つが考えられる．

① 予測残差に最も近いパルスや符号帳番号を求める．

② 候補パルスや候補符号帳を予測フィルタで合成後，音声に最も近いものに対応するパルスや符号帳番号を求める．

①は演算量が少なく一見良い方法に思えるが，量子化雑音（入力音声と復号音声の誤差）が大きい．②は音声領域での誤差を最小にでき，予測フィルタの利得（図４の音声と予測残差の電力比）だけ，量子化雑音を小さくできるという長所がある．

この②の方法を合成による分析法（A-b-S 法）と呼び，高能率音声符号化を支える柱となっている．合成および入力音声との距離計算に要する膨大な演算量が欠点である．

■誤り保護の基本技術

(1)　ビット選別誤り訂正

図6は PSI-CELP に符号誤りを加えたときに生じる歪を示す．高いほど誤りの影響が大きいことを表している．この図から，符号誤りの影響（誤り感度）は各ビットにより大きく異なることがわかる．これらのビット全体を均一な誤り訂正符号で保護するのは，訂正効果と冗長度のバランスからみて賢明でない．

ビット選別誤り訂正（BS-FEC:Bit-Selective Forward-Error-Control）は，誤り感度に基づきビットのグループ化を行い，各グループの誤り感度に応じて強さ（冗長度）の異なる誤り訂正符号化を施すというものである[10]．概念的な構

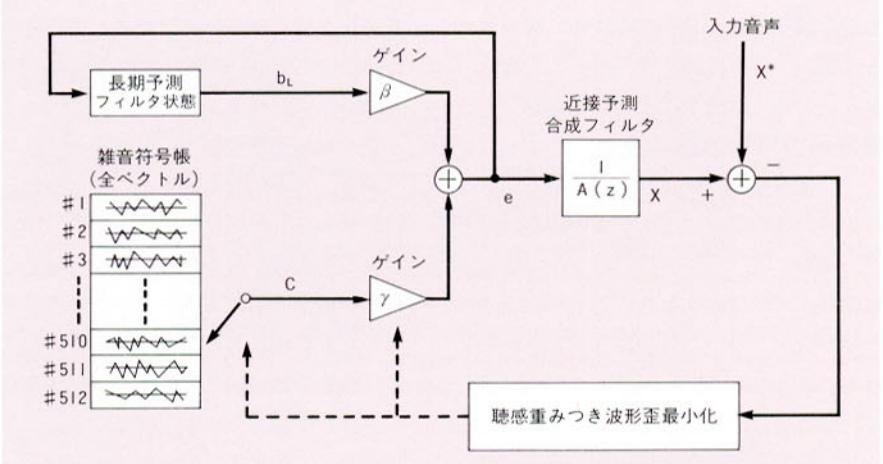

図5　CELP符号器

成図を図7に示す．ビット選別誤り訂正の採用により，必要な冗長度はおおむね20～30%程度削減できる．符号誤りがあると壊滅的な劣化を生じるような特に重要なビットには，さらにCRC*などの誤り検出符号を付加することもある．

ビット選別誤り訂正は，不均一誤り保護(UEP:Unequal Error Protection)と呼ばれることもある．

(2) 補　間

誤り訂正能力を超える符号誤りが生じた場合には，直前または直後の良好な復号音声や予測フィルタ係数を用いて，該当部分の音声を連続的につなぎ合わせることにより，復号音声の劣化を抑えることができる．これが補間である．

余り長い区間にわたって補間を続けると，弦を弾いたような不自然な音になるため，時間の経過とともに音量を抑え，最後にはスケルチとする工夫も取り入れられることが多い．図8に補間された復号音声の例を示す．

■デジタル信号処理プロセッサ

高能率音声符号化では，予測フィルタやベクトル量子化において，膨大な回数の積和演算（乗算と累加算）を行う．汎用プロセッサが苦手とするこの積和演算を，単一のマシンサイクルで実行するように設計されたのがデジタル信号処理プロセッサ(DSP)である．

図9にDSPの構成例を示す．データALU*では乗算器と加算器が結合され，各々1マシンサイクルで実行する．したがって，入力から出力まで2マシンサイクルかかる．これらを，パイプラインと呼ばれる流れ作業形式で動作させることができるため，実効的に1マシンサイクルで積和演算が実行される．また，乗算器へ2入力を供給するため，2系統のデータメモリとデータバスを備えている．また，データバスと命令バスが分離しており，1マシンサイクル内で命令の転送，データの転送，演算が同時にできるハーバードアーキテクチャとなっている．

半導体技術の進歩により現在では，音声符号化用で20MIPS*以上，画像符号化用で100MIPS以上という高速DSPが製造されており，PDCハーフレートもDSP1個で実現可能なレベルに到達している．

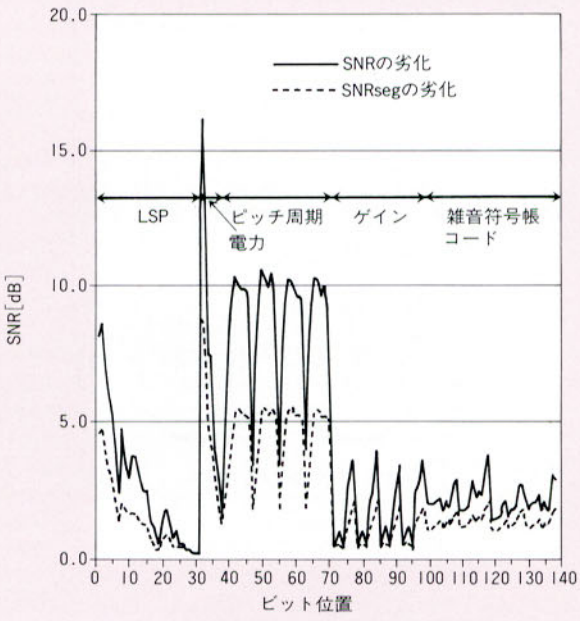

図6　符号誤り感度

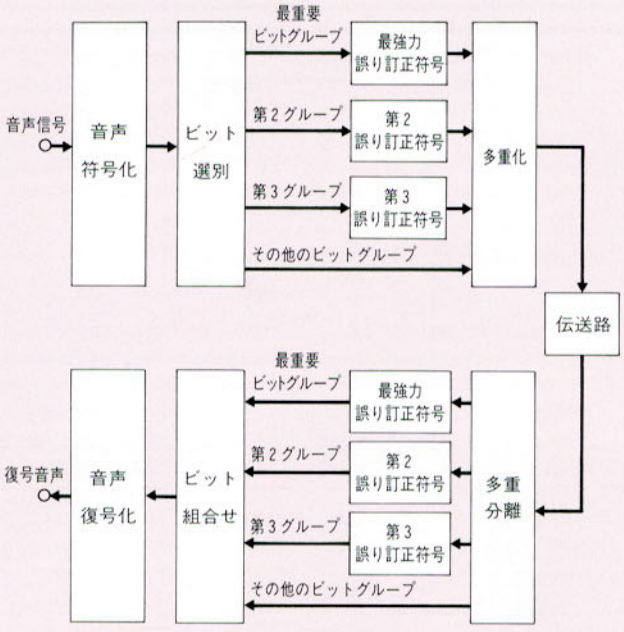

図7　BS-FECの原理

実用例

本章では，高能率音声符号化の実用例として，PDC*方式に採用されたVSELPとPSI-CELPを取りあげ，第3章の基本技術がどのように生かされているかを述べる．主要諸元を表2に示す．

■VSELP

VSELP[11]は，北米および日本のフルレート方式に採用されたもので，CELP系の

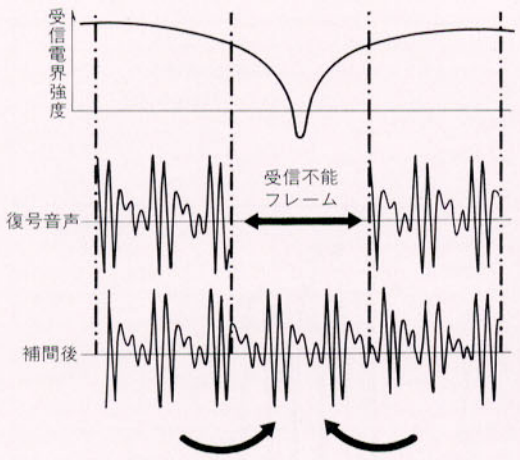

図 8　音声波形の補間

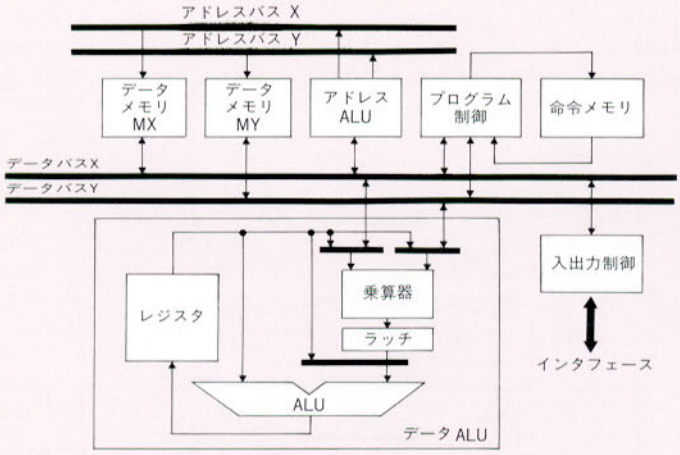

図 9　DSPの構成

表2　主要諸元

符号化方式		VSELP	PSI-CELP
ビットレート	音　声	6.7kb/s	3.45kb/s
	誤り訂正	4.5kb/s	2.15kb/s
処理量		7.8MOPS	18.7MOPS
フレーム長		20ms	40ms
遅延時間		48.1ms	85ms

符号化方式である．**図10**にVSELP 符号器の構成を示す．その大きな特徴は，VSELP の名前の由来にもなった雑音符号帳の構造にある．図5のCELPでは512本の候補が必要であったのに対し，図10のVSELP では9本しかない．これに±の符号を与えて和を求めることにより，512 通りの候補を作り出している．長所としては，①雑音符号帳の収納に必要なメモリ量が少ない，②符号誤りの影響が小さい，③合成フィルタの演算量が少ない，などがある．一方，量子化雑音が比較的大きいという短所もある．

また，表1の直交化処理を初めて導入した点も重要な特徴である．

■PSI-CELP

PSI-CELP[12]はPDC ハーフレートに採用されたCELP系の符号化方式であり，現時点で最も低いビットレートを達成している．**図11**にPSI-CELP符号器の構成を示す．その大きな特徴は，PSI-CELPの名前の由来にもなった雑音符号帳のピッチ同期化である．

図12にピッチ同期化の原理を示す．雑音符号帳を先頭から音声の基本周期であるピッチ周期分だけ取り出し，繰り返すような形に変形することにより，張りのある合成音声が得られる．ピッチ同期化は，量子化雑音を減らすこと，すなわち入力音声波形を忠実に再現することにはあまり貢献しないが，“らしさ”や“声の張り”が増加するので聞いた感じが随分良くなる．ピッチ同期化は，周期が短く，周期性がはっきりしている音声ほど効果的である．したがって，女性や子供のように高い声で特に著しい改善が見られる．

図13にVSELP とPSI-CELPの音声品質を示す．ビットレートが半減してもほぼ同じ品質が実現できている．また，図には現れないが，音の自然感でPSI-CELPが優れている．

将来展望

再び図1をご覧いただきたい．いずれの流れも，約5年ごとにビットレートが半減しているのがわかる．このペースが維持されれば，西暦2000年頃には移動通信網用に2kb/s を下回るクォータレートとも呼ぶべき低ビットレート音声符号化アルゴリズムが出現しても不思議はない．果たして本当だろうか．

ここでいくつかの問題点を整理してみたい．

まず第1に，骨格となる基本アルゴリズムについてである．今花盛りのCELP系アルゴリズムはどこまで低ビットレート化できるのか，CELPに代わり得る画期的新発明がここ数年内に出てくるのか，が焦点である．

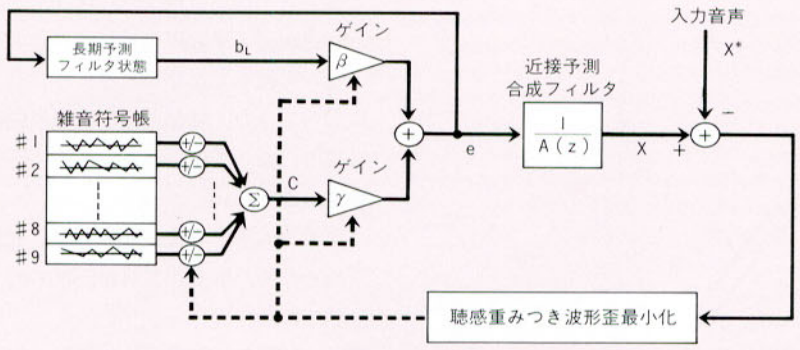

図10　VSELP符号器

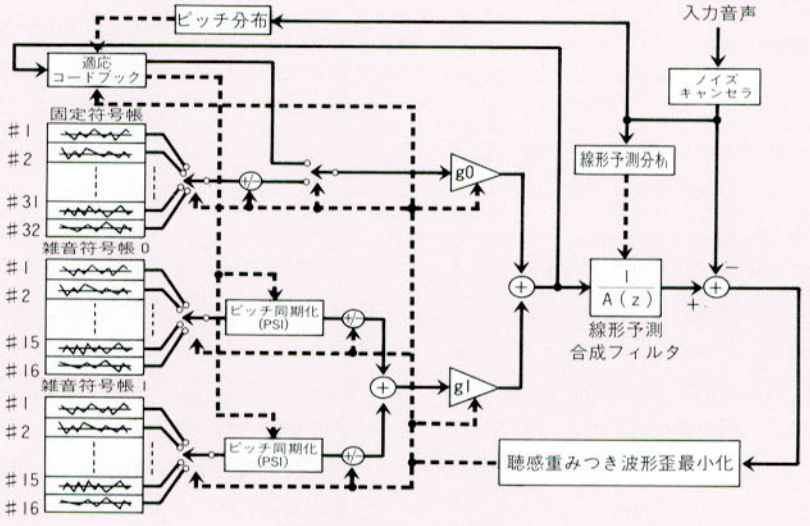

図11　PSI-CELP符号器

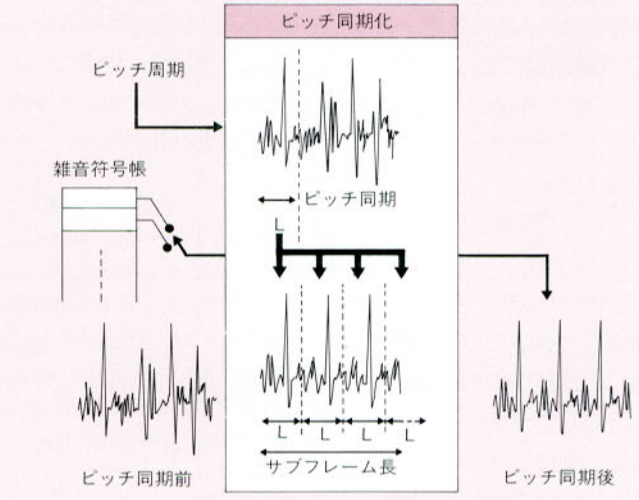

図12　ピッチ同期化の原理

第2に符号化遅延時間の問題である．これまで，ビットレートを半減させるごとに符号化遅延時間がほぼ倍増することを許容してきた．多くの音声サンプルを取り込み，じっくり分析した方が能率良く符号化できるからである．ところが，ハーフレートに至って，通話に支障がないとされる符号化遅延時間の最大値にかなり近づいてしまったため，これ以上大きくすることは許されない．

第3に処理量の問題である．符号化遅延時間同様，ビットレートを半減させるごとに処理量が数倍大きくなることを許容してきた．DSP の処理速度は半導体技術の発達とともに飛躍的に伸びてきたが，ここにきてやや頭打ちの状態にある．消費電力低減のためにも処理量を低く抑えたい．

第4に要求される音声品質である．携帯電話の普及とともにビジネスユースからパーソナルユースへと変化することが予想され，「無線だからこの程度で良い」とは言ってもらえず，「せめて固定電話並みの品質を」と要求されるようになるであろう．

第5に移動通信システムの容量の問題である．従来，大容量化が低ビットレート化を促進してきたが，変調方式や無線アクセス方式の進歩もあり，これ以上の低ビットレート化はそれほど必要ないとの雰囲気がある．

以上のことを総合すると，クォータレート化よりも，むしろ図1の2つの流れの最適統合的アプローチが当面の主流になると予想される．すなわち，

① 固定電話用音声符号化相当の高い音声品質と小さな遅延時間

② 移動通信用音声符号化相当の高い符号誤り耐性

を有し，

③ ビットレート4kb/s，処理量20 MOPS*，消費電力100mW 程度

の音声符号化アルゴリズムを，

④ CELP系の改良の範囲

で実現するというものである．

また，トラヒック密度や無線チャネル

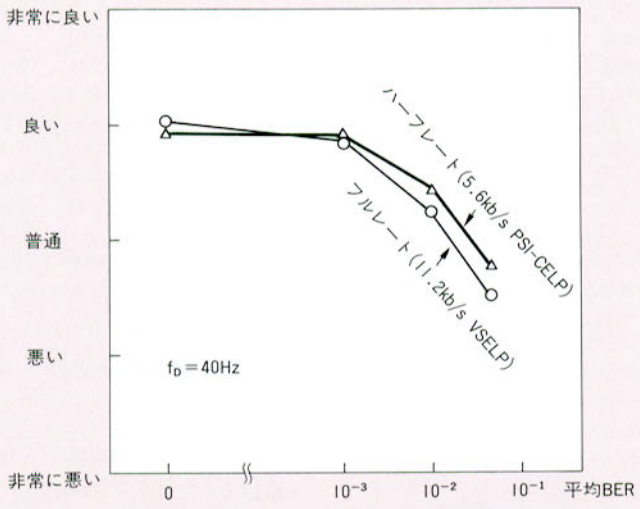

図13　フェージング下での符号化品質

状態，音声の有無や種別に応じて最適なビットレートを割り当てる可変レートアルゴリズムも盛んに研究されるだろう．これは，CDMA*/E-TDMA* など新しい無線アクセス方式やATM* にも親和性が良く，実効ビットレートを大幅に低減できることから無線周波数／ネットワーク資源の有効利用にもつながり，いずれ高能率音声符号化の主流となるものと予想される．

これまで述べてきたように，次世代の移動通信システムであるFPLMTS*の国際標準化の中で，高能率音声符号化がシステムの性能を左右する重要な基盤技術として議論の中心になることは間違いなく，今後も世界各国で活発に研究が進められると考えられる．

課　　題

最後に，高能率音声符号化全般が抱えている解決すべき課題について述べる．

■話者・言語依存性

ビットレートの低下に伴い，話者によって音声品質が極端に劣化する「人見知り」ともいうべき話者依存性現象が顕著となってきた．また，日本語，英語，独語といった言語によって音声品質がばらつく言語依存性も，国際標準化において大きな問題となっている．この原因は，

① 単一の音声生成モデルが必ずしも万人にあてはまらない
② コードブックサイズが有限であるため「既成服」では合わない場合がある
③ 言語を構成する母音・子音などの基本音韻が異なる

ことなどによる．この問題は音声符号化のみならず，音声認識など音声信号処理全体にとっても大きな問題であることから，より普遍的な音声生成モデルの登場が強く期待されている．

■音声品質評価方法

高能率音声符号化は，人間の聴覚特性を巧みに利用し，客観SNR*が低い復号音声で高い主観SNR を実現している．すなわち，「耳騙し」や「ものまね」に類する技術である．残念ながら「ものまね」の上手下手を量る絶対的な測定器や物差しはまだ開発されていない．したがって，復号音声の品質評価は最終的に主観評価（官能試験）によるしかなく，研究開発上大きな負担となるばかりか，評価結果の再現性に問題を残している．

最近では，客観評価用音声としての擬似音声や測定器としてのEPOQ*などが提案されているが，万能ではない．主観評価結果に普遍性を与えるための参照音声として長く使用されてきたMNR* も，その音色が高能率音声符号化音声と異なることから適切ではないと考えられるようになってきた．移動通信では不可避の符号誤りによって損なわれた音声の客観評価法は全くない．

高能率音声符号化の研究開発を促進するためにも，音声品質評価方法の確立が急務である．

■ノイズキャンセラ

入力音声に重畳されてくる背景／周囲雑音は符号化能率を劣化させるばかりか，符号化により「音化け」して違和感や不快感を与える．入力音声を歪ませることなく背景／周囲雑音のみを効率的に除去するノイズキャンセラの開発が必要である．

あ と が き

実用段階に入った高能率音声符号化技術について，基本技術や実用例を解説するとともに，将来展望と課題について述べた．本年３月にサービスを開始したデジタル移動通信システムをはじめ，高能率音声符号化技術は移動通信用を中心に急速な発展を遂げるものと期待される．PDC ハーフレートの標準化への貢献と同様，当社は本分野で世界トップレベルの研究開発を今後も強力に押し進めていく．

最後に，わかりやすさを優先した関係上，理論的な正確さに欠ける部分が多かったことを読者の皆様にお詫びする．少しでも多くの方に音声符号化への興味を持っていただければ幸いである．

文　　献

1) CCITT Recommendation G.712.
2) CCITT Recommendation G.721.
3) CCITT Recommendation G.731.
4) “デジタル方式自動車電話システム標準規格”，財団法人電波システム開発センタ，RCR STD-27B，(Dec.1992)
5) “Cellular System Dual-Mobile Station-Base Station Compatibility Standard”，IS-54，EIA/TIA Project No.2215 (1989)
6) “GSM Full Rate Speech Transcod-

ing", GSM 06.10 (1988)
7) 古井："ディジタル音声処理"，東海大学出版会
8) B.S.Atal and J.R.Remde: "A New Model of LPC Excitation for Producing Natural Sounding Speech at Low Bit Rates", Proc.ICASSP'82, S5.10 (1982)
9) M.S.Schroeder and B.S.Atal: "Code-Excited Linear Prediction (CELP):High-Quality at Very Low Bit Rates", Proc.ICASSP'85, pp.937-941 (1985)
10) H.Suda and T.Miki: "An Error Protected 16 kbit/s Voice Transmission for Land Mobile Radio Channel", IEEE J-SAC, vol.6, No.2, pp.346-352 (1988)
11) I.A.Gerson and M.A.Jasiuk: "Vector Sum Excited Linear Prediction(VSELP) Speech Coding at 8 kbps", Proc.ICASSP'90, pp.461-464 (1990)
12) T.Moriya, et.al.: "Pitch Synchronous Innovation CELP(PSI-CELP)", IEICE Trans. Fundamentals, vol.E76-A, No.7, pp.1177-1180, July (1993)

略語一覧

A-b-S	: Analysis by Synthesis
ALU	: Arithmetic Logic Unit
ATM	: Asynchronous Transfer Mode
CCITT	: Comite Consultatif International Telegraphique et Telephonique (仏語)
CDMA	: Code Division Multiple Access
CODEC	: Coder and Decoder
CRC	: Cyclic Redundant Code
EPOQ	: Equipment for Predicting indices of OPinions given on Quantizing distortion
E-TDMA	: Extended Time Division Multiple Access
FPLMTS	: Future Public Land Mobile Telecommunication Systems
LD-CELP	: Low Delay CELP
LSP	: Line SPectrum Pair
MIPS	: Mega Instruction Per Second
MNR	: Modulated Noise Reference
MOPS	: Mega Operation Per Second
MOS	: Mean Opinion Score
PARCOR	: PARtial autoCORrelation
PCM	: Pulse Code Modulation
PDC	: Personal Digital Cellular system
SNR	: Signal to Noise Ratio
SNRseg	: Segmental SNR

用語解説

LSP

線スペクトル対の英語名から作られた略語．予測フィルタを無損失で同相・逆相帰還させたときの共振周波数として得られ，予測係数と数学的に等価な係数である．係数間の大小関係が保証されているので，安定な予測フィルタが実現できる．予測係数の表現方法として最も能率の良いことで知られている．

PARCOR

偏自己相関係数の英語名から作られた略語．「線形予測分析」で述べた予測残差の自己相関係数として求められ，予測係数と数学的に等価な係数である．-1～+1の範囲にあることが保証されているので，安定な予測フィルタが実現できる．

PCM

本稿では，音声振幅を対数／指数的に瞬時圧伸し，線形量子化するものをいう．圧伸の方法として，日本・米国はμ-law，欧州はA-lawを用いている．いずれも人間の耳が小さな音の変化に敏感であることを利用して，小さな音は細かく，大きな音は大まかに表現している．音声サンプル当り5ビットくらいの情報圧縮効果がある．

PDC

日本におけるデジタル方式自動車電話システムとして，㈶電波システム開発センター(RCR)が標準化を進めている方式．平成3年に初版が策定され，平成5年にはハーフレートを含めたC改訂が予定されている．