

モバイルマルチメディア信号処理技術特集

音声符号化技術

デジタル移動通信基本サービスの根幹技術である、音声情報を高能率、高品質に圧縮伝送する方式について解説を行う。PSTN 通信で主に用いられている ITU-T 標準方式をはじめとし、移動通信用音声符号化方式である PDC や IMT-2000 の音声符号化について、その圧縮原理を概説するとともに、移動通信利用で最も問題となる伝送路符号誤りへの対策技術についても言及する。

なか 信彦 のぶひこ おあや ともゆき さいくさ まさひと はま とよかず
仲 彦 大矢 智之 三枝 正人 濱 豊和

1. まえがき

音声信号をデジタル信号で圧縮する音声符号化技術は、伝送効率の向上に寄与する。特に無線リソースが限られている移動通信で音声サービスを実現するため、ドコモもこれまで研究開発を進めてきた必須の技術である。現在では、移動通信で発展してきた高効率圧縮技術、低演算量化技術、誤り耐性技術は IP (Internet Protocol) 網をはじめ他の分野でも必要な技術となってきた。

本稿では、音声符号化技術として、移動通信用の音声符号化アルゴリズムの基本となっている、符号励振線形予測 (CELP: Code Excited Linear Prediction) について解説するとともに、ハイパートークとしてサービスが行われているデジタル自動車電話方式 (PDC: Personal Digital Cellular) 高音質フルレート音声符号化ならびに次世代移動通信 (IMT-2000: International Mobile Telecommunications-2000)、音声符号化適応マルチレート (AMR: Adaptive Multi Rate) を中心に代表的な音声符号化の概要を解説する。

2. 移動通信用音声符号化

2.1 CELP アルゴリズム

音声符号化には大きく分けて波形符号化、分析合成符号化、ハイブリッド符号化の3つがある。波形符号化はパルス符号変調 (PCM: Pulse Code Modulation) や適応的差分 PCM (ADPCM: Adaptive Differential Pulse Code Modulation) のように、信号の性質に依存せず、信号の波形をできるだけ忠実に符号化するものである。そのため、ビットレートが十分に高ければ高品質な符号化が可能であるが、ビットレートが低くなると急激に品質が劣化してしまう。一方で分析合成符号化は音声の生成モデルを仮定し、そのパラメータを分析・符号化する。そのため、ビットレートを低く抑えることができるが、音声品質は仮定する音声生成モデルに大きく依存し、ビットレートを上げて大きく品質を向上させることは難しい。ハイブリッド符号化は波形符号化と分析合成符号化を組み合わせた符号化方式であり、音声の生成モデルを仮定してそのパラメータを分析・符号化した後、パラメータでは表現できなかった残りの情報 (残差信号) に対して、さらに波形符号化する

方式である。ハイブリッド方式の代表的な方式の一つが CELP である。これは高能率かつ高品質な音声符号化を実現する一般的なアルゴリズムとして移動通信用音声符号化にも広く用いられている。

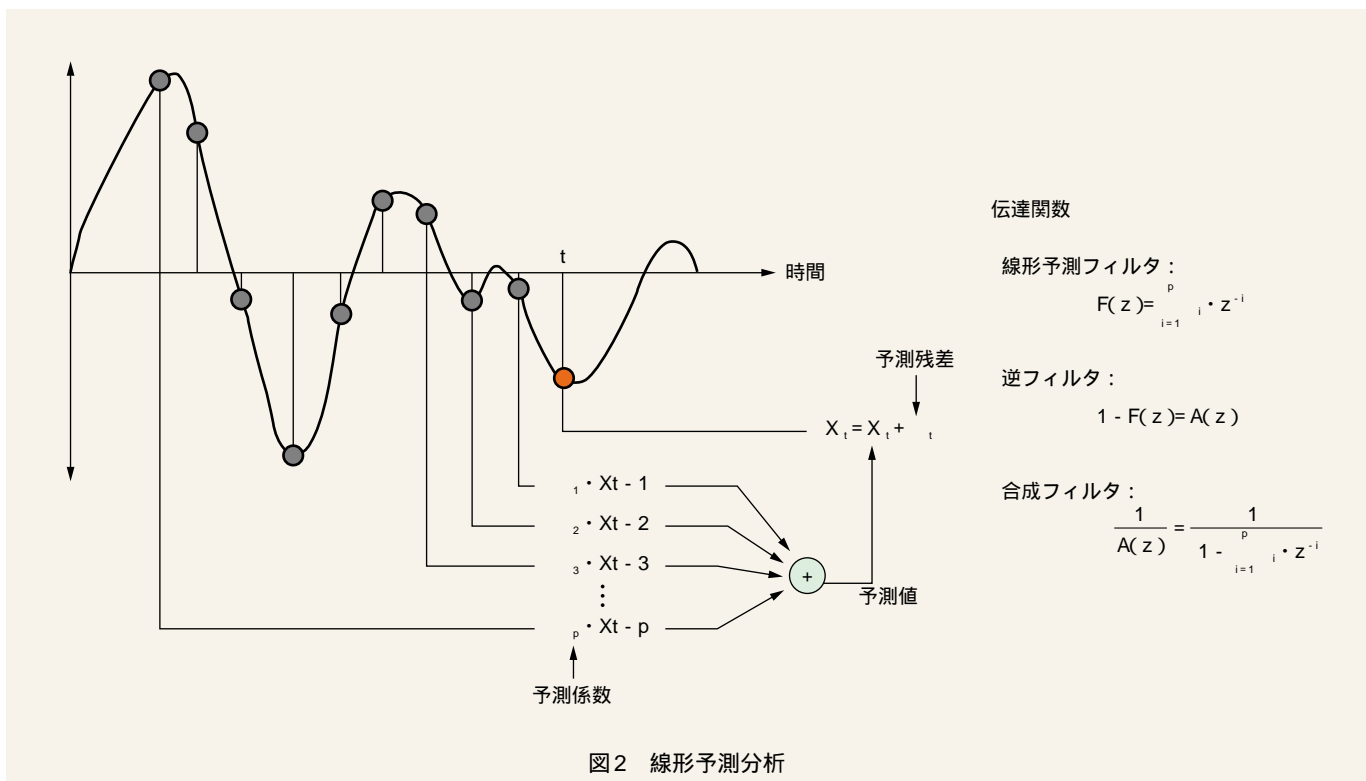
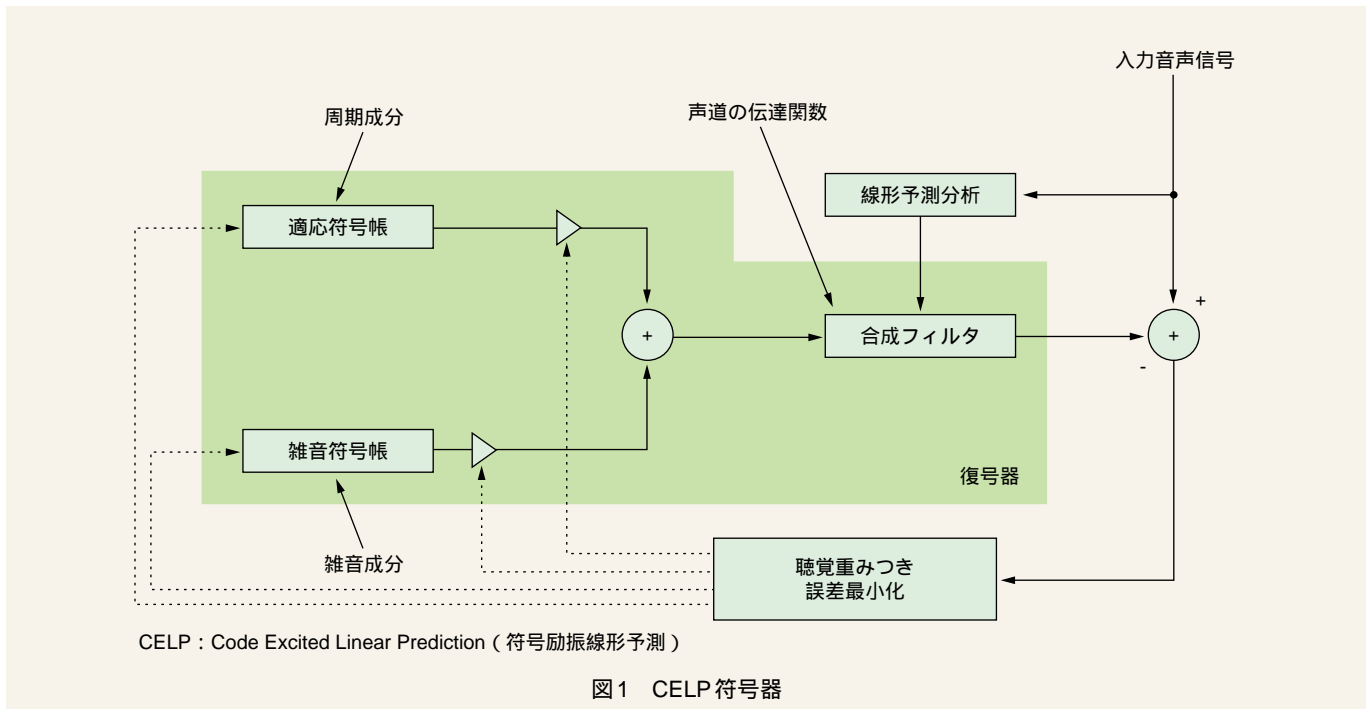
CELP 符号器のブロック図を図 1 に示す。CELP 符号器は内部に復号器と同じ構造を持つ。CELP 復号器は線形予測合成フィルタとそのフィルタを駆動するための励振信号を生成する2つの符号帳 (適応符号帳、雑音符号帳) で構成される。線形予測合成フィルタは音声信号のスペクトル包絡特性を表すため人間の声道に対応し、励振符号帳から生成される励振信号は声門を通過する肺からの呼気に対応する。すなわち、CELP は人間の発声メカニズムをモデル化したものといえる。

CELP 符号化で使用されている基本的技術について以下に説明する。

(1) 線形予測分析

線形予測分析は図 2 に示すように音声信号の時間的相関を利用し、過去の入力から現在の信号を予測するものである。予測信号と原信号との差分が予測残差である。

CELP 符号器では、音声の自己相関を計算し、Levinson Derwin Itakura 法などにより線形予測係数 a_k を求める。



電話帯域用符号化での線形予測係数の次数は通常10次である。線形予測係数はフィルタの安定度の判別が困難であるため、反射係数や線スペクトル対(LSP: Line Spectrum Pair)係数といった安定が保証された、等価な係数に変換して量子化され、伝送される。復号器では、伝送された a_i で合成フィル

タを構成し、予測残差で合成フィルタを駆動することにより復号音声を得る。合成フィルタの周波数特性は、音声のスペクトル包絡に対応する。

(2) 聴覚重み付けフィルタ

CEL P 符号器は内部に復号器と同じ構成を持ち、合成音声信号と入力音声信号の誤差が最小となるように各符号

帳のパターンとゲインを探索して符号化する。このような手法を合成による分析(A・b・S: Analysis・by・Synthesis)法と呼び、CEL P方式の特徴の一つとなっている。

A・b・Sでの誤差計算には、人間の聴覚特性を利用した重み付け誤差を用いる。聴覚重み付けフィルタは、線形

予測分析で求めた係数を用いた自己回帰移動平均 (ARMA : Auto Regressive Moving Average) 型のフィルタとして表現され、音声のスペクトル包絡を上下反転させた周波数特性を持つことにより、比較的知覚されやすいスペクトルの谷に相当する部分の量子化誤差を小さくする働きがある。

聴覚重み付けには非量子化線形予測係数を用いると特性が向上するが、演算量が大きくなる。そのため、従来は品質を多少犠牲にして量子化線形予測係数を用い、合成フィルタとの相殺により演算量を削減した例もあったが、現在は合成フィルタと聴覚重み付け合成フィルタ (これを聴覚重み付け合成フィルタと呼ぶ) のインパルスレスポンスを用いて計算するのが一般的である。

(3) 適応符号帳

適応符号帳は過去の励振信号をメモリに蓄積したものであり、動的に変化する。有声音のように励振信号が周期的な場合には、励振信号が声の高さに対応するピッチ周期によって繰り返されるため、適応符号帳により励振信号を効率的に表現することができる。ピッチ周期は、原音声と適応符号帳ベクトルを合成フィルタに通した出力との差が聴覚重み付領域で最小になるものが選択される。人間の音声のピッチ周期は通常 16 ~ 144 サンプルであるが、その範囲の中で探索される。ピッチ周期が短い場合には、オーバーサンプリングにより非整数周期まで量子化して、周波数分解能を向上させる。

誤差計算の演算量は非常に大きいため、通常はあらかじめ音声の自己相関を計算し、おおまかなピッチ周期を求めた上で、その周辺をオーバーサンプリングを含めて誤差計算を行うことで演算量を大幅に削減している。また、前回求めたピッチ周期の周辺のみを探索し、差分を量子化する手法も情報量と演算量の削減に有効な手法である。

(4) 雑音符号帳

雑音符号帳は適応符号帳により表現

できない残差信号を表現するため、非周期的なパターンを持つ。従来は、ガウス雑音や学習により求めた雑音信号を符号帳として持っていたが、現在は残差信号を数本のパルスで表現する代数字号帳が一般的に用いられる。これにより、雑音ベクトルの格納に必要であったメモリや適応符号帳との直交化演算および誤差計算の演算量が飛躍的に削減できることとなった。

(5) ポストフィルタ

ポストフィルタは復号音声を整形することにより主観品質を向上させる目的で、復号の最終過程で使用される。代表的なポストフィルタであるフォルマント強調フィルタはARMA型で聴覚重み付けフィルタの逆特性を持ち、スペクトルの谷を抑圧し、量子化誤差を知覚しづらくする効果がある。さらに、ポストフィルタには出力信号のスペクトルの傾きを補正するフィルタが加わる。

2.2 移動通信適用のための周辺技術

移動通信では無線伝送路の使用、屋外、移動中での使用といった特別な条件から、それらに適應するためのさまざまな周辺技術が用いられている。ここでは、それら周辺技術について概要を説明する。

(1) 誤り訂正技術

無線区間で生じる符号誤りを訂正するのが、誤り訂正符号である。音声符号化情報ビットの誤り感度 (そのビットが誤った場合の復号音声に与える歪みの大きさ) に応じて訂正能力の異なる訂正符号を使うビット選択前誤り訂正 (BS-FEC : Bit Selective-Forward Error Correction) もしくは不均一誤り保護 (UEP : Unequal Error Protection) を用いることにより、効率的に誤り訂正を行っている。

(2) 誤りコンシールメント技術

上記誤り訂正符号で誤りを訂正しきれない場合や情報が消失した場合などは、受信した情報を元に復号を行うと正しい復号ができない。このような場

合、過去の音声情報を利用したパラメータ補間により誤った区間の音声信号を生成し、聴感上の音声品質劣化を最小限に抑えるのが、誤りコンシールメント技術である。補間されるパラメータは線形予測係数、ピッチ周期、ゲインなどの時間的に相関が高いものが用いられる。

(3) 間欠送信

音声がない区間では、極端に少ない情報量か、もしくは送信を停止するのが間欠送信である。これは、移動機のバッテリー寿命の向上や、干渉の低減、統計多重効果などに有効である。VAD (Voice Activity Detector) は音声のパラメータを用いて有音無音を判定する。無音区間では、音声情報と比較して、非常に少ない情報量の背景雑音情報を元に背景雑音を生成し、間欠送信による違和感を軽減する。

(4) 雑音抑圧

前項で述べたように CELP アルゴリズムは人間の発声モデルを利用しているので、肉声以外の音、たとえば街頭の音などについては特性が劣化する。そのため、会話に必要な人間の声以外の雑音を抑圧することにより、会話音声の品質を向上することができる。

3. 音声符号化の歴史 (G.711からピッチ同期 雑音励振源 CELP まで)

表1は主な電話帯域音声符号化の規格化された年とビットレートである。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) においては、国際的な通話品質の確保という観点から汎用的な電話用音声符号化の規格化が行われてきた。一方、社団法人電波産業会* (ARIB : Association of Radio Industries and Businesses)、欧州電気通信標準化機構 (ETSI : European

*旧名称は財団法人電波システム開発センタ (RCR : Research and Development Center for Radio Systems)

表1 代表的な音声符号化

年代	ITU-T (固定網用符号化標準)	PDC用符号化標準(ARIB)	GSM(ETSI)	IMT-2000 (3GPP)	D-AMPS	CDMA-one/3GPP2	ISO/IEC
1970	1972 G.711(64kbit/s PCM)						
1980	1984 G.726 32kbit/s ADPCM		1988 RPE-LTP 22.8kbit/s		1988 VSELP 13kbit/s		
1990	1992 G.728 16kbit/s LD-CELP	1990 VSELP 11.2kbit/s					
	1996 G.729 8kbit/s CS-ACELP 1996 G.723.1 5.3kbit/s ACELP & 6.3kbit/s MPC-MLQ	1993 PSI-CELP 5.6kbit/s 1998 CS-ACELP, ACELP 11.2kbit/s	1994 EVSELP 11.4kbit/s 1996 EFR(ACELP)22.8kbit/s 1998 AMR(4.75, 5.15, 6.7, 7.4, 7.95, 10.2, 12.2kbit/s)		1996 EFR/ ACELP 13kbit/s	1994 QCELP (8k, 4k, 0.8kbit/s) 1997 EVRC (8k, 4k, 0.8k)	1999 MPEG-4 CELP
2000~	G.4k(未定)			1999 AMR			

- ACELP: Algebraic Code Excited Linear Prediction
- AMR: Adaptive Multi Rate (適応マルチレート)
- ARIB: Association of Radio Industries and Businesses (社団法人電波産業会)
- CELP: Code Excited Linear Prediction (符号励振線形予測)
- CS-ACELP: Conjugate Structure-ACELP (共役構造ACELP)
- D-AMPS: Digital Advanced Mobile Phone Services
- EFR: Enhanced Full Rate
- ETSI: European Telecommunications Standards Institute (欧州電気通信標準化機構)
- GSM: Global System for Mobile Communications
- IEC: International Electrotechnical Commission (国際電気標準会議)
- IMT-2000: International Mobile Telecommunications-2000 (次世代移動通信)
- 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)
- MLQ: Maximum Likelihood Quantization
- MPC: Multi Pulse Coding
- MPEG: Moving Picture Experts Group
- PDC: Personal Digital Cellular (デジタル自動車電話方式)
- PSI-CELP: Pitch Synchronous Innovation-Code Excited Linear Prediction (ピッチ同期雑音励振源CELP)
- QCELP: Qualcomm CELP (Qualcomm社CELP)
- VSELP: Vector Sum Excited Linear Predictive Coding (ベクトル和励振線形予測)
- 3GPP: 3rd Generation Partnership Project

Telecommunications Standards Institute), TIA (Telecommunications Industry Association) といった各地域標準化団体では、移動通信のデジタル化にあたり、無線回線容量確保の観点からより圧縮率の高い音声符号化方式をITU-T標準とは別にそれぞれの要求条件に合うよう策定を独自に行ってきた。表1に示すARIBの例では、1990年に早くも6.7kbit/s(音声符号化部のみ)のフルレート方式を選定し、3年後の1993年には3.45kbit/sのハーフレート方式を選定している。これらをはじめとする移動通信音声符号化には、単に音声を圧縮するだけでなく、誤り耐性、間欠送信技術といった移動通信特有の周辺技術についても同時に規格化された。このように音声の情報圧縮という観点では移動通信の分野で飛躍的に発展してきたが、近年ITU-Tも標準音声符号化のアプリケーションとして移動通信、マルチメディア通信を考慮した符号化を規格化するようになり、1996年に規格化されたのがPDC

高音質フルレートにも採用されているG.729である。

ここでは、代表的な電話用音声符号化方式について、その概要を以下に解説する。また、表2には本稿で取り上げた音声符号化の諸元を示す。

3.1 G.711 [1]

電話用途でもっとも基本となる音声符号化であり公衆交換電話網(PSTN: Public Switched Telephone Network)で採用されている。1音声サンプルを8ビットで非線形量子化し、64kbit/sで伝送する。量子化の方法としてA-lawと μ -lawの2種類が規定されており、日本や北米では μ -lawが採用されているので、 μ -law PCMと呼ばれることもある。

3.2 G.726 [2]

G.711に続きITU-Tで標準化されたのが現在のG.726(ADPCM)である。PHSに採用されている32kbit/sが一般的であるが、量子化ビット数に応じて

16, 24, 40kbit/sも規格化されている。波形符号化で1サンプルを4ビットで量子化するが、現音声信号を適応予測し、かつ量子化ステップも適応的に更新することにより効率の良い圧縮を実現している。また32kbit/sのADPCMは固定網並(toll quality)といわれる品質の基準としても用いられ、IMT-2000用音声符号化の要求品質でもある。

3.3 G.729 [3]

G.728に続きITU-Tで標準化された8kbit/sの音声符号化方式がG.729共役構造ACELP(CS-ACELP: Conjugate Structure-Algebraic Code Excited Linear Prediction)である。PSTNのみならず、移動網も当初から適用範囲として規格化された。LSP係数の符号化には2種類の予測を用い、その残差を他段VQで量子化する。雑音励振信号は4本のパルスの代数符号帳で表現される。代数符号帳のゲインは過去のフレームから予測し、その補正係数を量

表2 音声符号化の諸元

		ITU標準		PDC			IMT-2000
		G.711	G.726	フルレート	ハーフレート	フルレート拡張	AMR
方式		PCM	ADPCM	VSELP	PSI-CELP	CS-ACELP	MR-ACELP
ビットレート	合計	64kbit/s	32kbit/s	112.0kbit/s	5.6kbit/s	11.2kbit/s	12.2, 10.2, 7.95, 7.4, 6.7, 5.9, 5.15, 4.75kbit/s
	音声	64kbit/s	32kbit/s	6.7kbit/s	3.45kbit/s	8.0kbit/s	12.2, 10.2, 7.95, 7.4, 6.7, 5.9, 5.15, 4.75kbit/s
	誤り訂正	-	-	4.5kbit/s	2.15kbit/s	3.2kbit/s	-
処理量		1MIPS以下	2MIPS以下	7.8MOPS	18.7MOPS	約20MIPS	約15wMOPS
アルゴリズム遅延		0.125ms	0.125ms	28.125ms	45ms	25ms	25ms
フレーム長		-	-	20ms	40ms	20ms(音声は10ms)	20ms
その他		-	-	VAD/DTX機能	VAD/DTX機能 背景雑音抑圧機能	VAD/DTX機能 音声符号化はG.729	マルチレート 背景雑音生成 VAD/DTX機能

ADPCM: Adaptive Differential Pulse Code Modulation (適応的差分PCM)

AMR: Adaptive Multi Rate (適応マルチレート)

CS-ACELP: Conjugate Structure-ACELP (共役構造ACELP)

DTX: Discontinuous Transmission

IMT-2000: International Mobile Telecommunications-2000 (次世代移動通信)

ITU: International Telecommunication Union (国際電気通信連合)

MIPS: Million Instructions Per Second

MOPS: Million Operations Per Second

MR-ACELP: Multi Rate-ACELP

PCM: Pulse Code Modulation (パルス符号変調)

PDC: Personal Digital Cellular (デジタル自動車電話方式)

PSI-CELP: Pitch Synchronous Innovation-Code Excited Linear Prediction (ピッチ同期雑音励振源CELP)

VAD: Voice Activity Detector

VSELP: Vector Sum Excited Linear Predictive Coding (ベクトル和励振線形予測)

子化する。そのため、フレーム電力の伝送が不要である。また、ピッチ情報用の1ビットのパリティビットや共役構造のゲイン符号帳を用いることで誤り耐性を持たせている。固定網並の品質であることから、PDC高音質フルレート方式の音声符号化部の他、VoIP (Voice over IP) でも利用されている。また、低演算量化、VAD、高ビットレート版、低ビットレート版など、さまざまなAnnexによる拡張規格も存在する。

3.4 PDCフルレート方式[4]

欧米での移動通信のデジタル化に続き、日本のPDCフルレート方式に採用されたのが、6.7kbit/s (音声符号化のみ) のベクトル和励振線形予測 (VSELP: Vector Sum Excited Linear Predictive Coding) である。CELPをベースとし、励振ベクトルの直交化探索や聴覚重み付合成フィルタの簡略化などにより演算量を削減し、ベクトル和構造の雑音励振源によって誤り耐性を実現したため、米国の時分割多元接続方式 (TDMA: Time Division Multiple Access) フルレート方式やGSM (Global System for Mobile Communications) の

ハーフレート方式に同等のアルゴリズムが採用されたが、近年G.729などの同等のビットレートでより品質が高いものにとって変わられつつある。

3.5 PDCハーフレート方式[4]

移動通信の加入者の飛躍的な増大に対処するため、PDCハーフレート方式として採用されたのが、ピッチ同期雑音励振源CELP (PSI-CELP: Pitch Synchronous Innovation-Code Excited Linear Prediction) である。本アルゴリズムはドコモがARIBに提案し、コンテストの結果、採用されたものである。CELPを基本として、線形予測係数のマトリックス量子化、予備選択、分数周期のピッチ量子化、雑音符号帳のピッチ同期化、共役構造の雑音符号帳、ゲインの同時最適化、ディレイディシジョンといった技術によって、3.45kbit/sという高圧縮率を実時間で計算できるよう実現した。

4. PDC高音質フルレート方式^[4]

4.1 標準化

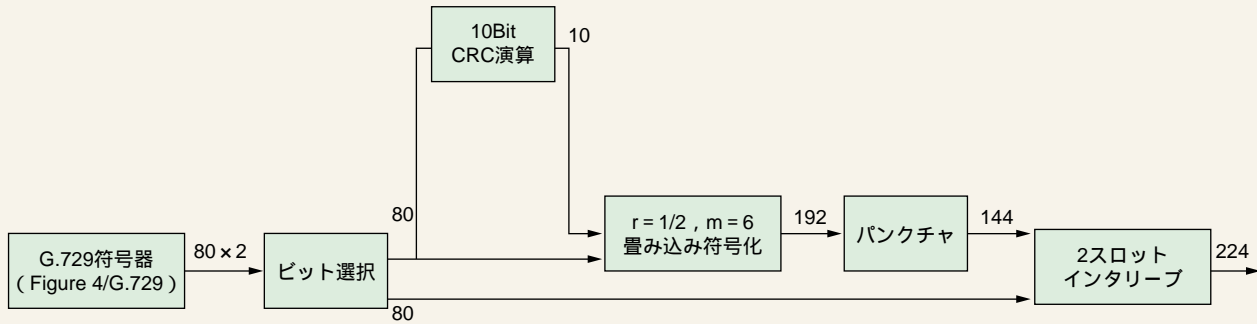
PDC高音質フルレートは、ARIBで

フルレート (VSELP) とハーフレート (PSI-CELP) の必須音声符号化方式に加えて、事業者ごとに選択可能なオプション音声符号化方式として1998年にRCR STD-27 H版改訂[4]にて標準化された。標準化における採用の可否は、事業者選択オプション音声符号化方式であるために、音声品質がVSELP、PSI-CELPより劣らないこととされた。

4.2 PDC高音質フルレート方式の特徴

PDCの通話音声品質向上を目的として開発したPDC高音質フルレート方式は、G.729を基本としてドコモ独自の誤り訂正技術および誤り補償技術を適用して、PDCシステムにおいて音声の送信に必要なビットレートを従来のフルレート (VSELP) と同一の11.2kbit/sとしつつ、より優れた音質を実現している。

G.729で生成される10msごとの音声符号化情報をPDCのエアインタフェースにあわせ、20msフレームで伝送路符号化が行われ、11.2kbit/sのフルレート信号として送信される。図3に、PDC高音質フルレート方式で用いる伝送路符号化の概略を示す。



CRC : Cyclic Redundancy Check (巡回冗長検査)

図3 PDC高音質フルレート方式の伝送路符号化

(1) 保護対象ビット選別

8kbit/s G.729符号化音声ビットストリームより、誤り感度に応じて保護対象と保護対象外ビットを選別する。選別された音声ビットストリーム2フレームを用い、20ms (160ビット)の伝送路符号化フレームを構成する。これを $r = 1/2$, $m = 6$ の畳み込み符号器で符号化する。さらに、畳み込み符号器の出力にパンクチャ、インタリーブを施し、224ビット/20msのフレームを作成する。

(2) 伝送路復号

伝送路復号では受信された系列を逆インタリーブした後、畳み込み符号の復号を行う。畳み込み符号の復号にはビタビ復号アルゴリズムを用いる。さらに、1ビットごとの信頼度情報と1シンボルごとのエンベロープ情報により距離メトリックの計算を行う軟判定復号を用いて特性の向上が可能である。誤り訂正復号後、復号系列から送信側と同様に巡回冗長検査 (CRC : Cyclic Redundancy Check) を計算する。CRCが一致しない場合はフレーム消失と判定され、誤り消失補償を適用する。

4.3 誤り消失補償技術

PDC高音質フルレートでは、G.729の誤り消失補償技術に加え、以下の技術を適用し、PDCへの適用性と通話品質の向上を実現している。

(1) 適応プリフィルタ制御

符号誤りが発生したフレームで異音が発生するのを抑制するために、バースト誤りの場合には、消失補償処理のピッチ強調処理を省略する。

(2) 適応符号帳の再構成

復号した現サブフレームのピッチ遅延、直前のサブフレーム復号時の適応および固定符号帳利得、直前のサブフレーム復号時に生成された置換用励振信号を用いて過去の励振信号を作成し、現サブフレームの適応符号帳ベクトルを再構成する。

(3) 励振信号の利得制御

誤りから復帰したフレームにおいて、適応および固定符号帳利得を調整し、誤りから復帰したフレームの復号音質を改善する。

これらの誤り補償技術により、PDC高音質フルレートはエラーフリーで、VSELP、PSI-CELPより有意に優れた音質、かつ誤り訂正に与えるビットが少ないにも関わらず、符号誤りがある場合にも、同等以上の音質を実現している。

5. IMT-2000 音声符号化 AMR^[5]

5.1 標準化

1997年にARIBにIMT-2000委員会が設立され、日本が世界に先駆けて第3世代の標準化を始めた。音声符号化

はIMT-2000委員会のCODEC作業班で選定作業が進められた。作業班では、委員各社が音声符号化方式を提案したため、評価要領を策定し、評価試験を実施した。評価試験の途中、1998年末ARIB、社団法人電信電話技術委員会 (TTC : The Telecommunications Technology Committee), TIA, ETSIにより第3世代移動通信の標準化プロジェクト3GPP (3rd Generation Partnership Project) が設立されたことにより、ARIBの評価結果を元に3GPP TSG-SA WG4 (CODEC)にて選定を行うことが合意された。ドコモも参加したこの評価試験により、他の候補音声符号化方式との優位性が認められ、AMRが3GPPの必須音声符号化方式として採用された。

5.2 アルゴリズム概要

AMRは、ACELPをベースとしたマルチレート音声符号化方式で、1998年にGSMの音声符号化方式として採用された。12.2kbit/sから4.75kbit/sまでの8つの符号化モードを有し、そのうち12.2, 7.4, 6.7kbit/sは、すでに地域標準として規格化されている音声符号化方式と共通アルゴリズムである。

アルゴリズムはG.729と同様であるが、マルチレート化のための工夫がされている。フレーム長は全モード20ms固定で、サブフレーム数、量子化ビット数によりマルチレートを実現

表3 AMRのビット配分

(TS 26.090 [5]より引用)

モード	パラメータ	第1サブフレーム	第2サブフレーム	第3サブフレーム	第4サブフレーム	フレーム合計
12.2kbit/s	LSP×2					38
	ピッチ遅延	9	6	9	6	30
	ピッチ利得	4	4	4	4	16
	代数符号	35	35	35	35	140
	符号帳利得	5	5	5	5	20
	合計					244
10.2kbit/s	LSP					26
	ピッチ遅延	8	5	8	5	26
	代数符号	31	31	31	31	124
	利得	7	7	7	7	28
	合計					204
7.95kbit/s	LSP					27
	ピッチ遅延	8	6	8	6	28
	ピッチ利得	4	4	4	4	16
	代数符号	17	17	17	17	68
	符号帳利得	5	5	5	5	20
	合計					159
7.40kbit/s	LSP					26
	ピッチ遅延	8	5	8	5	26
	代数符号	17	17	17	17	68
	利得	7	7	7	7	28
	合計					148
6.70kbit/s	LSP					26
	ピッチ遅延	8	4	8	4	24
	代数符号	14	14	14	14	56
	利得	7	7	7	7	28
	合計					134
5.90kbit/s	LSP					26
	ピッチ遅延	8	4	8	4	24
	代数符号	11	11	11	11	44
	利得	6	6	6	6	24
	合計					118
5.15kbit/s	LSP					23
	ピッチ遅延	8	4	4	4	20
	代数符号	9	9	9	9	36
	利得	6	6	6	6	24
	合計					103
4.75kbit/s	LSP					23
	ピッチ遅延	8	4	4	4	20
	代数符号	9	9	9	9	36
	利得		8		8	16
	合計					95

LSP : Line Spectrum Pair (線スペクトル対)

している(表3)。

線形予測係数は、12.2kbit/sでは1フレームにつき2回分析し、LSP係数の低次から順に2次ずつ分割して2×2の要素をLSP領域で予測した後、その残差をベクトル量子化する。それ以外のモードでは、1フレームにつき1回分析し、LSP領域で予測後低次から分割してベクトル量子化する。

長期予測タップは12.2kbit/sでは1/6、それ以外は1/3の非整数周期で探索し、フレーム内で差分子量子化される。

代数符号帳は、10本から2本の非零で大きさ1のパルスで構成される。また、PSI化と同等の効果が得られるピッチプリフィルタが符号帳ベクトル探索に適用される。符号帳利得は12.2kbit/sおよび7.95kbit/sでは適応符号帳、固定符号帳別個に量子化されるが、それ以外のモードではベクトル量子化される。復号器では合成音声にフォルマントポストフィルタおよび傾き補償フィルタを適用し、最終的な復号音声を得る。

AMRでは、移動通信に必要な周辺技術も同時に規定されている。間欠送信に必要なVADのアルゴリズムとして2つのオプションが用意されている。背景雑音情報(SID)は短期予測係数およびフレーム電力を35ビットで量子化し、一定間隔で伝送する。また、誤りが生じた場合のコンシールメントは要求条件が規定され、その処理例として、誤りの発生状況に応じた状態遷移による符号帳利得、短期予測係数など、符号化パラメータの補間処理

が規定されている。

また、IMT-2000のRAN(Radio Access Network)はツールボックスとして柔軟な設計ができるように規定されている。そのため、適応マルチレート(AMR: Adaptive Multi Rate)の符号化情報に対しRAN側で不均一誤り保護(UEP: Unequal Error Protection)を適用できるよう、符号化情報の重要度に応じたクラス分けが規定されている。なお、ISG(IMT-2000 Steering Group)ではこのクラス分けに応じた無線パラメータを規定している。

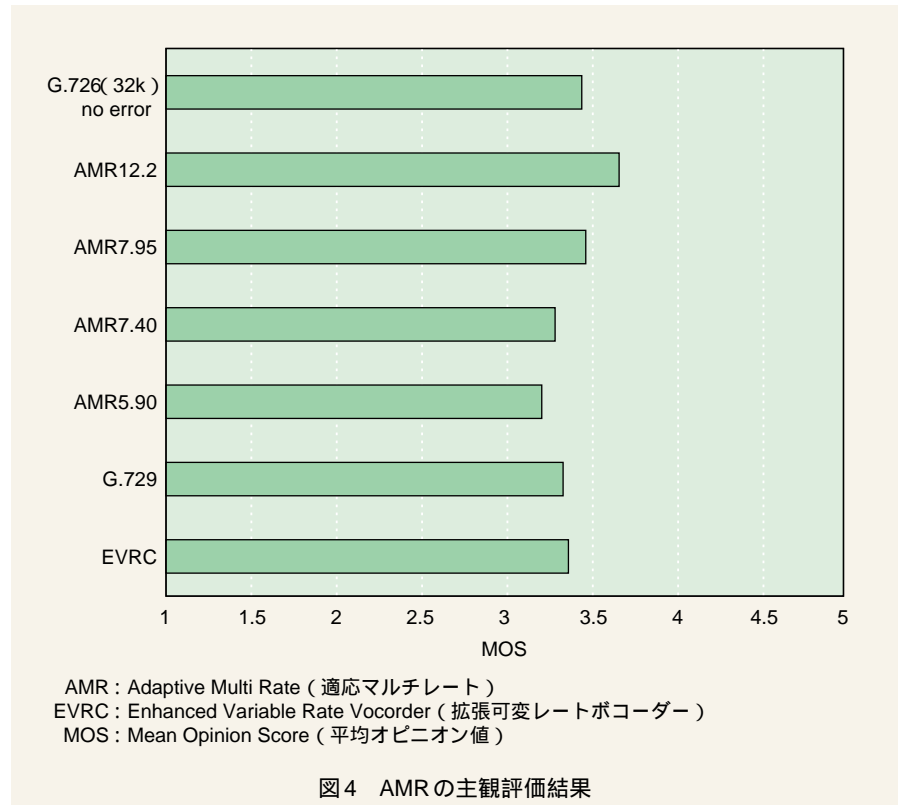
5.3 品質

ARIBの選定試験要領に沿ってドコモで実施し、3GPPに提出したAMRの主観評価結果の一部を図4に示す。これは広帯域符号分割多元接続方式(W-CDMA: Wideband Code Division Multiple Access)のBER(Bit Error Rate)が0.1%の条件で試験したものである(ただし、現在の無線の伝送方法とは若干異なる。)結果から、12.2kbit/sは他のどの符号化方式と比べても優れた品質であり、また他の符号化方式と同等のビットレートで比較しても同等以上であることが分かる。

さらに、AMRの品質は3GPP規格TR26.975 [6]でGSMでの品質が報告されており、現在W-CDMAにおける品質もドコモとNortel Networksが提供した誤りパターンを用いて3GPPで評価作業中である。また、ドコモで実施した単音明瞭度試験結果でも単音明瞭度80%を満足している。

5.4 電話以外の用途

AMRは、従来にない柔軟な構造とその優れた品質から、3GPPの回線交換用マルチメディア端末符号化規格3G-324M[7]の必須音声符号化としても採用されている。また、IETF(Internet Engineering Task Force)ではAMRをVoIPに適用するためのRTP(Realtime Transport Protocol)ペイロードフォーマットが規定されつつある



など、IMT-2000の音声サービスに限らず幅広い分野での適用が検討されている。

6. 今後の動向

現在、3GPPでは、AMRの広帯域版(~7kHz)であるAMR-WBが2000年10月に選定される予定であり、選定されたアルゴリズムはITU-Tの低レート広帯域音声符号化の候補となる予定である。また、ITU-Tでは4kbit/sのPSTN並品質音声符号化の標準化も進行中である。

一方で通信網のIP化の流れを受けて回線交換網と同等の電話サービスを提供するため、音声符号化のVoIPやストリーミングサービスへの適用も活発に議論されている。

VoIPの標準化はETSIのTIPHON(Telecommunication and Internet Protocol Harmonization Over Networks)プロジェクトやIETFのIPTEL(IP Telephony)、AVT(Audio Video Transport)が中心であるが、3GPPは移動網のIP化に向けてこれらの団体

と協調しつつ標準化を進めている。

7. あとがき

本稿では、移動通信用音声符号化技術を中心に音声符号化について解説した。ここで取り上げた音声符号化以外にも、マルチメディア用のITU-T標準G.723.1やcdmeOneのEVRC、国際標準化機構(ISO: International Organization for Standardization)/国際電気標準会議(IEC: International Electrotechnical Commission)のMPEG(Moving Picture Experts Group)-4などがある。これら音声符号化技術全般については電子通信情報学会の「音声符号化」[8]が詳しく、本稿でも記述の参考とした。

本稿で解説した音声符号化は、いまや移動通信にとどまらずインターネットを中心にさまざまな分野で応用されつつある。今後はより一般的な音声符号化について、高品質かつ柔軟性のある符号化方式の実現に向けて研究に取り組んでいきたい。

文 献

- [1] ITU-T Recommendation G.711, " Pulse code modulation (PCM) of voice frequency," November 1998.
- [2] ITU-T Recommendation G.726, " 40, 32, 24, 16 kbit/s adaptive differential pulse code modulation (ADPCM)," December 1990.
- [3] ITU-T Recommendation G.729, " Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction," March 1996.
- [4] ARIB STD-27H.
- [5] 3GPP 3G TS 26.090, " AMR Speech Codec; Transcoding functions," December 1999.
- [6] 3GPP 3G TR 26.975, " Performance characterization of the AMR speech codec," January 2000.
- [7] 3GPP 3G TS 26.111, " Codec for Circuit Switched Multimedia Telephony Services; Modifications to H.324," August 1999.
- [8] 守谷健弘, " 音声符号化," 電子通信情報学会, 1998.

用語一覧

A·b·S : Analysis·by·Synthesis (合成による分析)

ACELP : Algebraic Code Excited Linear Prediction

ADPCM : Adaptive Differential Pulse Code Modulation (適応的差分PCM)

AMR : Adaptive Multi Rate (適応マルチレート)

ARIB : Association of Radio Industries and Businesses (社団法人電波産業会)

ARMA : Auto Regressive Moving Average (自己回帰移動平均)

AVT : Audio Video Transport

BER : Bit Error Rate (ビット誤り率)

BS-FEC : Bit Selective-Forward Error Correction (ビット選択前誤り訂正)

CELP : Code Excited Linear Prediction (符号励振線形予測)

CODEC : Corder Decorder (符号化復号化)

CRC : Cyclic Redundancy Check (巡回冗長検査)

CS-ACELP : Conjugate Structure-ACELP (共役構造ACELP)

D-AMPS : Digital Advanced Mobile Phone Services

DTX : Discontinuous Transmission

EFR : Enhanced Full Rate

ETSI : European Telecommunications Standards Institute (欧州電気通信標準化機構)

EVRC : Enhanced Variable Rate Vocorder (拡張可変レートボコーダー)

GSM : Global System for Mobile Communications

IEC : International Electrotechnical Commission (国際電気標準会議)

IETF : Internet Engineering Task Force

IMT-2000 : International Mobile Telecommunications-2000 (次世代移動通信)

IP : Internet Protocol

IPTEL : IP Telephony

ISG : IMT-2000 Steering Group

ISO : International Organization for Standardization (国際標準化機構)

ITU : International Telecommunication Union (国際電気通信連合)

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)

LSP : Line Spectrum Pair (線スペクトル対)

MIPS : Million Instructions Per Second

MLQ : Maximum Likelihood Quantization

MOPS : Million Operations Per Second

MOS : Mean Opinion Score (平均オピニオン値)

MPC : Multi Pulse Coding

MPEG : Moving Picture Experts Group

MR-ACELP : Multi Rate-ACELP

PCM : Pulse Code Modulation (パルス符号変調)

PDC : Personal Digital Cellular (デジタル自動車電話方式)

PSI-CELP : Pitch Synchronous Inovation-Code Excited Linear Prediction (ピッチ同期雑音励振源CELP)

PSTN : Public Switched Telephone Network (公衆交換電話網)

RAN : Radio Access Network

RCR : 財団法人電波システム開発センタ

RTP : Real-time Transport Protocol

QCELP : Qualcomm CELP (Qualcomm 社 CELP)

SID : 背景雑音情報

TDMA : Time Division Multiple Access (時分割多元接続方式)

TIA : Telecommunications Industry Association

TIPHON : Telecommunication and Internet Protocol Harmonization Over Networks

TTC : The Telecommunications Technology Committee (社団法人電信電話技術委員会)

UEP : Unequal Error Protection (不均一誤り保護)

VAD : Voice Activity Detector

VoIP : Voice over IP

VSELP : Vector Sum Excited Linear Predictive Coding (ベクトル和励振線形予測)

W-CDMA : Wideband Code Division Multiple Access (広帯域符号分割多元接続方式)

3GPP : 3rd Generation Partnership Project