

■2群 (画像・音・言語) - 8編 (音声・オーディオ符号化)

3章 モバイル向け符号化方式

(執筆著者：小澤一範) [2009年12月 受領]

■概要■

既にサービスされていたアナログ方式と比べ同等以上の無線周波数利用効率を確保するために、1980年代後半から移動体通信のデジタル化(第二世代)を実現する検討が行われた。これには音声のデジタル圧縮は必須の重要技術であり、モバイル向けの音声符号化方式の標準化が必要とされた。要求条件としては、低いビットレートでも音質が良好なことだけでなく、モバイル環境特有の伝送路誤りや背景雑音への頑健性が求められた。各国の無線周波数や電波政策などの諸事情から変復調方式や速度が異なっていたため、欧州 GSM、北米 TIA、日本 ARIB において、独立に標準化が行われ、異なる符号化方式が選定された。

その後、1990年代後半からの第三世代移動体通信向けの標準化では、グローバルな標準を目指したものの、実際には W-CDMA 陣営と CDMA2000 陣営の2つに分かれることとなり、W-CDMA 陣営は 3GPP が中心となり音声符号化方式を標準化し、CDMA2000 陣営は 3GPP2 が中心となり方式の標準化が行われたため、両者のビットストリームレベルでの相互接続は困難となってしまった。なお、3GPP の標準化では、当初は 3GPP 独自方式で ITU-T の標準方式とは異なる方式であったが、一部に ITU-T と基本部分を共通化する動きもあり AMR-WB では 3GPP で選定された方式と基本部分が共通な方式が ITU-T で G.722.2 として採択された。

一方、オーディオ符号化はこれまで ISO/IEC JTC1/SC29/WG11 (通称 MPEG) で勢力的に標準化が行われてきたことでもあり、3GPP では W-CDMA 向けのオーディオ符号化については MPEG で標準化された方式を基本部分に採用する方針とし、これに誤り隠蔽機能などを追加した方式を 3GPP で採択した。

【本章の構成】

本章では、モバイル向け音声・オーディオ符号化方式の全体の俯瞰をし (3-1 節)、3GPP、3GPP2、ETSI/GSM、北米 TIA 及び日本 ARIB で標準化された音声符号化方式を述べ (3-2 節)、3GPP で標準化されたオーディオ符号化方式について述べる (3-3 節)。

■2群 - 8編 - 3章

3-1 符号化方式の一覧

(執筆者：小澤一範) [2009年12月 受領]

モバイル向けの音声符号化方式の一覧を表 3・1 に、オーディオ符号化方式の一覧を表 3・2 に、それぞれ示す。表には各方式について、ビットレート、標準化周波数、処理単位であるフレーム長、標準化機関、標準化の規格番号、用途を記載した。なお、用途で GSM、TDMA や PDC は第二世代の移動体通信を示している^{1,2)}。

表 3・1 モバイル向け音声符号化方式

方式	ビットレート (kbit/s)	標準化 周波数 (kHz)	フレーム長 (msec)	標準化 機関	規格 番号	用途
AMR-NB	4.75~12.2	8	20	3GPP	TS26.090	W-CDMA
AMR-WB	6.6~23.85	16	20	3GPP	TS26.190	W-CDMA
AMR-WB+	6~36 (モノ) 7~48 (ステレオ)	12.8~ 38.4	20	3GPP	TS26.290	W-CDMA
VSELP	7.95	8	20	TIA	IS-54	TDMA
VSELP	6.7	8	20	ARIB	STD-27	PDC
EVSELP	5.6	8	20	ETSI/ GSM	GSM06.20	GSM
RPE-LTP	13	8	20	ETSI/ GSM	GSM06.10	GSM
EFR	12.2	8	20	ETSI/ GSM	GSM06.60	GSM
EFR	7.4	8	20	TIA	IS-641	TDMA
EFR	6.7/8	8	20	ARIB	STD-27	PDC
PSI-CELP	3.45	8	40	ARIB	STD-27	PDC
QCELP	8, 4, 2, 1	8	20	TIA	IS-96	CDMA
QCELP	13.3, 6.2, 2.7, 1	8	20	TIA	IS-733	CDMA
EVRC	8.55, 4, 0.8	8	20	TIA	IS-127	CDMA
VMR-WB	13.3, 6.2, 2.7, 1.0	8, 16	20	3GPP2	C.S0052-A	CDMA
SMV	8.55, 4, 2, 0.8	8	20	3GPP2	C.S0030-0	CDMA

表 3・1 において、AMR-NB、AMR-WB、EFR は本編 1 章 1-3-9 で述べた ACELP 方式に基づいている。AMR-WB+は音声だけでなくオーディオ信号に対しても良好な性能を得るために音声信号は ACELP に基づき符号化され、オーディオ信号は MDCT などの要素技術を用いる TCX 符号化に基づき符号化される。

一方、VSELP、EVSELP、PSI-CELP、QCELP、EVRC、VMR-WB と SMV は CELP 方式に基づいている。なお、RPE-LTP はレギュラーパルス方式に基づくが、これはマルチパルス符号化においてパルス列を一定間隔ごとに配置するような拘束を設けることにより、パルス探索に必要な演算量を削減した方式である。

表 3-2 モバイル向けオーディオ符号化方式

方式	ビットレート (kbit/s)	標本化 周波数 (kHz)	フレーム長 (msec)	標準化 機関	規格 番号	用途
MPEG-4 AAC-LC	128 程度 (ステレオ) 最大:6*SF*CH	8~48	1024/SF	ISO/IEC 3GPP	ISO/IEC 14496-3, 3GPP TS26.401~ 411	W-CDMA
aacPlus	64 程度 (ステレオ)	8~48	2048/SF	ISO/IEC 3GPP	ISO/IEC 14496-3, 3GPP TS26.401~ 411	W-CDMA
Enhanced aacPlus	32 程度 (ステレオ)	8~48	2048/SF	ISO/IEC 3GPP	ISO/IEC 14496-3, 3GPP TS26.401~ 411	W-CDMA

表 3-2 のビットレート、フレーム長の欄で、“SF”の記載は標本化周波数（単位は kHz）を意味し、“CH”の記載はチャンネル数を意味する。オーディオ符号化方式は要素技術として、本編 1 章 1-2 で述べた MDCT や聴覚モデルを用いている。

■参考文献

- 1) 小澤一範, “デジタル移動通信のための高能率音声符号化技術,” トリケップス, 1992.
- 2) 守谷健弘, “音声音響符号化における標準,” 音響誌, vol.64, no.2, pp.114-118, 2008.

■2 群-8 編-3 章

3-2 音声符号化

(執筆者：小澤一範) [2009年12月 受領]

モバイル向けの音声符号化方式は、デジタル移動体通信に使用するために 3GPP, 3GPP2, ETSI/GSM, 北米 TIA 及び日本 ARIB において方式の標準化が行われ、選定試験が実施された。モバイル特有の環境条件や、携帯電話といった演算性能、メモリ量がそれほど高くない端末で音声コーデックを動作させるためのインプリ上での制約を考慮し、方式選定のクライテリアには、種々の C/I 条件であらかじめシミュレートしたビット誤りパターンを混入させた条件や背景雑音を重量させた条件などでの音質得点、ならびにインプリ容易性に関する演算量、データ ROM 容量及びデータ RAM 容量の各得点に対し、あらかじめ定めた重みづけを施して算出した合計得点が用いられ、合計得点が高い方式が選定された。

各標準化機関で選定された音声符号化方式のうち、本節では代表的な方式について特徴を述べる。

3-2-1 AMR-NB (Adaptive Multi-Rate - NarrowBand) (執筆者：伊藤博紀) [2009年12月 受領]

適応マルチレートと呼ばれ、ACELP 方式に基づく符号化方式である。1999 年に ETSI で実施した、伝送路誤りが高い環境での標準化コンテストにおいて、標準化選定された¹⁾。その後、3GPP で採用された。フレーム長 20 msec、先読み長 5 msec、4.75~12.2 kbit/s で 8 種類のビットレートをサポートし、フレーム単位でビットレートの切り替えが可能である。ビットレートを表 3・3 に示す。

ここで、12.2, 7.4, 6.7 kbit/s はそれぞれ、ETSI/GSM EFR, TIA EFR, PDC EFR と同一の仕様である。なお、低ビットレート時の音質を向上させるために、固定コードブックゲインの適応平滑化と適応パルス拡散処理を有する。また、VAD/DTX/CNG (音声検出/間欠伝送/疑似快適雑音生成) 機能もサポートする。デコード処理のブロック図を図 3・1 に示す。

表 3・3 AMR-NB のビットレート

ビットレート (kbit/s)
12.2, 10.2, 7.95, 7.4,
6.7, 5.9, 5.15, 4.75

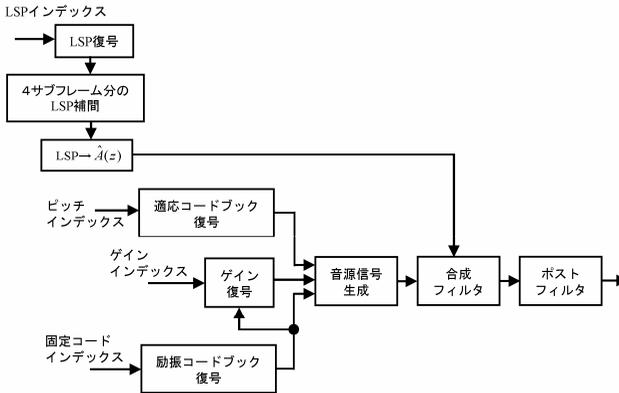


図 3・1 AMR-NB のデコード処理

3-2-2 AMR-WB (Adaptive Multi-Rate – WideBand)

(執筆者：伊藤博紀) [2009年12月 受領]

適応マルチレート広帯域と呼ばれ、ACELP方式に基づく7kHz帯域音声向けの符号化方式である。2001年に3GPPで標準化選定され³⁾、2002年にITU-TでG.722.2として採用された⁴⁾。フレーム長20msec、先読み長5msec、6.6～23.85kbit/sで9種類のビットレートをサポートし、フレーム単位でビットレートの切り替えが可能である。ビットレートを表3・4に示す。また、VAD/DTX/CNG(音声検出/間欠伝送/擬似快適雑音生成)機能もサポートする。

AMR方式で符号化するのは6.4kHz帯域の音声信号であり、6.4～7.0kHzの高域信号は、受信側にて白色雑音特性を持つ励振信号を、符号化時の線形予測フィルタでフィルタリングすることにより得ている。高域部の励振信号のゲインは、23.85kbit/sでは送信側で符号化時に算出した値を伝送するが、それ以外のビットレートでは、受信側で復号音声信号から推定した値を用いる。デコード処理のブロック図を図3・2に示す。

表 3・4 AMR-WBのビットレート

ビットレート (kbit/s)					
23.85,	23.05,	19.85,	18.25,	15.85,	
14.25,	12.65,	8.85,	6.60		

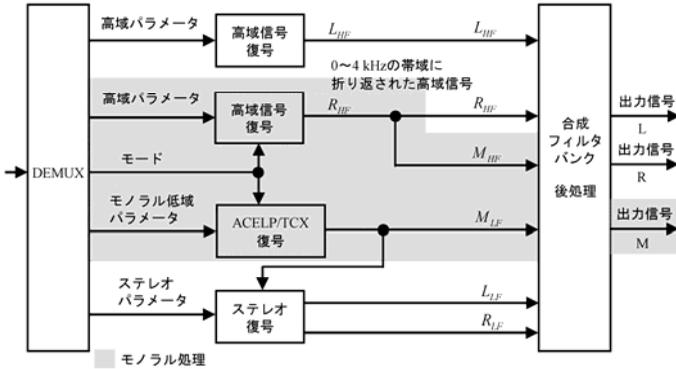


図 3-3 AMR-WB+のデコード処理

3-2-4 VSELP (Vector Sum Excited Linear Prediction)

(執筆者：小澤一範) [2009年12月 受領]

ベクトル和励振線形予測符号化と呼ばれ、I. Gerson らにより提案された⁶⁾。励振符号のビット数に相当する個数の基底ベクトルをもち、対応するビットの0, 1に応じて基底ベクトルに1または-1の重みを乗じた上で基底ベクトルを加算することにより、コードベクトルを生成する。この構造により、コードブック探索に必要な演算量と格納に必要なメモリ量の両者を大幅に削減している。さらに、伝送路誤りにより符号系列を誤って受信しても、基底ベクトルに乗ずる重みの極性が誤るだけなので、伝送路誤りによる劣化は少ない。また、音声データベースを用いて基底ベクトルを事前にトレーニングしておくことにより性能を改善している。受信側ではピッチの調波構造を利用して量子化雑音を重みづけするピッチプリフィルタを導入し、音質を改善している。1989年にTIAが、1990年に電波産業会(現ARIB)が、それぞれフルレートデジタル自動車電話/携帯電話向けの標準化コンテストを実施し、北米ではIS-54として13 kbp VSELP(音声7.95 kbit/s及び誤り訂正5.05 kbit/s)が、日本ではPDC向けに11.2 kbps VSELP(音声6.7 kbit/s及び誤り訂正4.5 kbit/s)が、それぞれ選定された。

日本で選定されたVSELPのデコード処理のブロック図を図3-4に示す。

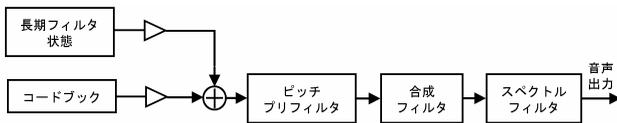


図 3-4 VSELPのデコード処理

3-2-5 EVSELP (Enhanced VSELP)

(執筆者：小澤一範) [2009年12月 受領]

ETSI/GSMの11.4 kbit/s ハーフレートデジタル自動車電話/携帯電話向けに1994年に選定された方式である⁷⁾。VSELPをベースに改良を施しビットレートを5.6 kbit/sに低減化した。主な改良点を次に示す。

VSELP では PARCOR 係数をスカラ量子化していたが、EVSELP では 3 分割スプリットベクトル量子化を導入しビットレート削減化した。送信側で聴感重みづけを行うための聴感重みづけフィルタを、VSELP での全極型フィルタから ITU-T G.728 LD-CELP で採用された極-零型フィルタに変更し、量子化雑音に対する重みづけ精度を上げて低ビットレートでの音質を改善した。

3-2-6 RPE-LTP (Regular Pulse Excited LPC-Long Term Prediction)

(執筆者：小澤一範) [2009 年 12 月 受領]

ETSI/GSM の 22.8 kbit/s フルレートディジタル自動車電話／携帯電話向けの標準方式であり、1987 年に選定された⁸⁾。元になった RPE はレギュラーパルス励振符号化と呼ばれ、マルチパルス励振符号化のパルス探索に必要な演算量を低減化するために、パルスの位置が一定間隔ごとに並ぶように拘束することにより、パルス位置の探索ならびに位置情報の伝送を不要にした方式である。ただし、そのためには短区間 (5 msec) ごとに先頭パルスの初期位相を求めて伝送する必要がある。

RPE-LTP は RPE の音質を改善するために、長期予測器を追加した方式である。20 msec フレームごとに音声線を線形予測分析し逆フィルタすることで得た線形予測残差信号に対し、5 msec ごとに開ループの 1 次ピッチ予測を行い、ピッチ予測残差信号に対しレギュラーパルスの初期位相を求める。レギュラーパルスの各振幅は適応 PCM により量子化され、PARCOR、ピッチ係数と遅延、初期位相、ブロック最大値、レギュラーパルスの各振幅が伝送される。音声には 13 kbit/s を、誤り訂正には 9.8 kbit/s を割り当てている。

3-2-7 EFR (Enhanced Full Rate)

(執筆者：小澤一範) [2009 年 12 月 受領]

フルレートディジタル自動車電話／携帯電話の音質を改善するために、欧州、北米、日本において、それぞれ EFR 音声符号化の選定が実施された。GSM/ETSI では 1995 年に ACELP をベースにした符号化が選定された⁹⁾。音声に 12.2 kbit/s を、誤り訂正に 10.6 kbit/s を割り当てている。EFR の音質は GSM フルレート及びハーフルレートよりも大幅に改善されており、誤りなし及び誤り率の少ない条件、背景雑音、タンデム接続条件のいずれも ITU-T G.726 ADPCM と同等以上の音質が得られる。

一方、TIA では TDMA デジタル移動通信システム (IS-136) 向けに EFR の標準化が実施され、1996 年に ACELP ベースの符号化が IS-641 として選択された¹⁰⁾。フレーム長はいずれも 20 msec であるが、同じ ACELP ベースでも GSM の EFR とはビットレート及び細部が異なっており、TIA では音声に 7.4 kbit/s が、誤り訂正に 5.6 kbit/s が割り当てられている点、適応コードブックと音源コードブックのゲインをまとめてベクトル量子化している点や、LSP のベクトル量子化法や音源コードブックの構成などが異なる。

日本では電波産業会で、1999 年及び 2000 年に PDC-EFR として 2 種の方式が認定された。どちらも ACELP に基づく。一方は、音声符号化に 8 kbit/s の ITU-T G.729 CS-ACELP を採用し、3.2 kbit/s の誤り訂正符号化などを新たに規定した。もう一方は、音声符号化に 6.7 kbit/s ACELP を新規に用いるが、誤り訂正は PDC フルレートと同一であり 4.5 kbit/s を割り当てた。どちらの EFR を用いるかは通信事業者により異なる。

なお、1999 年に 3GPP で第三世代ディジタル移動通信向けに選定された AMR-NB は 8 種

類のモードをもつが、12.2 kbit/s モードは ETSI/GSM の EFR と、7.4 kbit/s モードは TIA IS-641 EFR と、6.7 kbit/s モードは PDC-EFR と、それぞれ同一である。

3-2-8 PSI-CELP (Pitch Synchronous Innovation Code Excited Linear Prediction)

(執筆者：小澤一範) [2009年12月 受領]

ピッチ同期雑音励振源符号励振線形予測符号化と呼ばれる。4 kbit/s 以下の低ビットレートでの実用化をめざしフレーム長を 40 msec、サブフレーム長を 10 msec とし、ピッチ周期性のあるサブフレームで適応コードブックの遅延がサブフレーム長より短い場合は、雑音符号帳の波形を遅延の周期で繰り返しピッチ同期化（ピッチ周期化）することにより、ピッチ周期性を強調している。

一方、ピッチ周期性のないサブフレームでは適応符号帳のかわりに固定符号帳を選択し、ランダム成分を生成している。雑音符号帳は演算量及びメモリ量削減のために 2 段の符号帳の和で表現し、探索にはディレイドデジションの手法を導入し適応符号帳／固定符号帳から 2 個の候補を呼び選択し、それぞれの候補に対し雑音符号帳を探索することで、演算量を抑えながら性能を改善している¹¹⁾。デコード処理のブロック図を図 3・5 に示す。1993 年に電波産業会（現 ARIB）により実施された日本のデジタルハーフレート自動車／携帯電話向けの標準化コンテストにおいて、標準方式に選定された¹²⁾。誤り訂正を含めた合計のビットレートはフルレート方式の 1/2 の 5.6 kbit/s であるため、PSI-CELP では音声に 3.45 kbit/s を、誤り訂正に 2.15 kbit/s を割り当てている。

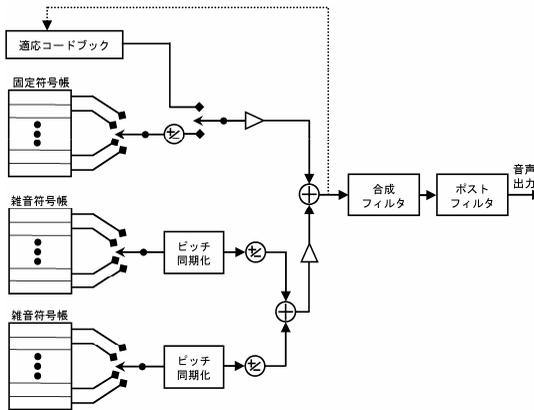


図 3・5 PSI-CELP のデコード処理

3-2-9 QCELP (Qualcomm CELP)

(執筆者：小澤一範) [2009年12月 受領]

8 kbit/s QCELP は、1993 年に TIA の CDMA デジタルセルラーシステム向け可変速度音声符号化標準 IS-96 として選定された¹³⁾。20 msec フレームでの入力音声の性質により、4 種のレート（8 kbit/s、4 kbit/s、2 kbit/s、1 kbit/s）から 1 つを選択する。レートの選択は現フレームの音声のエネルギーと背景雑音エネルギー推定値とを比較することにより行う。符号化

原理は CELP に基づいており、適応コードブック及び、乱数から構成される音源コードブックをもつ。可変速度のため、平均ビットレートは 4 kbit/s 以下となるが、音質は 8 kbit/s VSELP と同等と報告されている。

なお、さらに音質を改善するために、TIA では 1995 年に 13 kbit/s QCELP を IS-733 として選定した。20 msec フレームごとに 4 種のレート (13.3 kbit/s, 6.2 kbit/s, 2.7 kbit/s 及び 1 kbit/s) から 1 つを選択し、平均ビットレートを約 6.6 kbit/s に抑えている。

3-2-10 EVRC (Enhanced Variable Bit Rate Coder) (執筆者：小澤一範) [2009年12月 受領]

可変ビットレート符号化であり、フレーム当たりの伝送レートが入力信号の性質やシステムサイドからの要求により時間的に変化する。フレームあたりの伝送レートは、表 3・7 に示すように、レート 1 (171 ビット: 8.55 kbit/s に相当)、レート 1/2 (80 ビット: 4 kbit/s に相当)、レート 1/8 (16 ビット: 0.8 kbit/s に相当) の 3 種類あり、それぞれ、有声、無声、無音に対応して切り替わる。これにより固定ビットレート符号化に比べ平均的なビットレートを削減している。フレーム長は 20 msec, サブフレーム数は 3 で、ノイズサブレッサを内蔵している。各サブフレームのサンプル数は 53, 53, 54 でありレート 1 とレート 1/2 は RCELP (Relaxation CELP)¹⁴⁾ に基づき符号化する。RCELP はフレームごとに線形予測残差信号からピッチ遅延を求め、これをサブフレームごとに補間し補間後の遅延を用い時間軸を伸縮させた目標信号を計算し ACELP により音源符号帳を計算する。レート 1/8 では背景雑音に対しエネルギーの時間変化を符号化する。デコード処理のブロック図を図 3・6 に示す。

TIA は 1996 年に CDMA デジタルセルラー向けに EVRC を IS-127 として選定した¹⁵⁾。なお、3GPP2 は 2007 年に cdma2000 向けに EVRC-WB を選定した。7 kHz 帯域の音声を低域と高域の 2 つの帯域に分割し、最大 8.55 kbit/s で符号化している。

表 3・7 EVRC のビットレート

タイプ	ビットレート (kbit/s)
レート 1	8.55
レート 1/2	4
レート 1/8	0.8

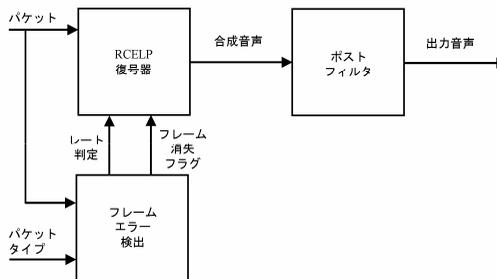


図 3・6 EVRC のデコード処理

3-2-11 VMR-WB (Variable-Rate Multimode WideBand) (執筆者：伊藤博紀) [2009年12月受領]

VMR-WB は、RCELP 及び ACELP 方式に基づく 7 kHz 及び 3.4 kHz 帯域音声向けの符号化方式である。2004 年に 3GPP2 で採用された¹⁶⁾。フレーム長は 20 msec、アルゴリズム遅延は 7 kHz 帯域時 33.75 msec、3.4 kHz 帯域時 35.0625 msec である。

入力信号を分析して、その特徴に応じたビットレート制御を行う符号化方式である。約 3.4 ~ 9.1 kbit/s の平均ビットレートにおいて、CDMA のシステムに応じた 5 種類の動作モードを持つ。AMR-WB の 12.65, 8.85, 6.6 kbit/s を含んでいるため、相互接続が可能である。音声の特徴分類やピッチ推定性能などの向上により、従来方式よりも高い符号化性能を実現している。

3-2-12 SMV (Selectable Mode Vocoder) (執筆者：伊藤博紀) [2009年12月受領]

EVRC の符号化性能を改善する方式として、TIA で標準化が開始され、2001 年に 3GPP2 で採用された¹⁷⁾。フレーム長 20 msec、先読み長 10 msec であり、8.55, 4, 2, 0.8 kbit/s の 4 種類のビットレートを持つ。8.55 kbit/s 及び 4 kbit/s は eX-CELP (eXtended CELP) 方式に基づいている。

入力音声はノイズ抑圧処理が施された後、有声音は高いビットレートで、無声音や背景雑音は低いビットレートで符号化される。有声音を符号化する 8.55, 4 kbit/s ではピッチゲインが大きい場合に、励振信号に多くのビットを割り当てるモードに切り替えることにより、符号化性能を向上させている。

■参考文献

- 1) 3GPP TS26.090, "AMR speech codec: Transcoding functions," 1999.
- 2) 3GPP TS26.190, "AMR wideband speech codec: Transcoding functions," 2001.
- 3) B. Besette, R. Salami, R. Lefebvre, M. Jelinek, J. Rotola-Pukkia, J. Vamio, H. Mikkola, and K. Jarvinen, "The adaptive multirate wideband speech codec (AMR-WB)," IEEE Trans. on speech and audio processing, vol.10, no.8, pp.620-636, 2002.
- 4) ITU-T Recommend. G.722.2, "Wideband coding of speech at around 16 kbit/s using adaptive multi-rate wideband (AMR-WB)," 2002.
- 5) 3GPP TS26.290, "Extended Adaptive Multi-Rate - Wideband (AMR-WB+) codec: Transcoding functions," 2004.
- 6) I.A. Gerson and M.A. Jasiuk, "Vector sun excited linear prediction (VSELPL)," IEEE Proc. ICASSP, pp.461-464, 1990.
- 7) I.A. Gerson and M.A. Jasiuk, "Techniques for improving the performance of CELP-type speech coders," IEEE JSAC, vol.10, no.5, pp.858-865, 1992.
- 8) J. Natvig, "Pan-European speech coding standard for digital mobile radio," Speech communication, pp.113-123, 1988.
- 9) K. Jarvinen, J. Vanio, P. Kapanen, T. Honkanen, and P. Haavisto, "GSM enhanced full rate speech codec," IEEE Proc. ICASSP, pp.771-774, 1997.
- 10) TIA/EIA/IS-641, "Interim standard, TDMA cellular/PCS radio interface-Enhanced full-rate speech codec," 1996.
- 11) 三樹聡, 守谷健弘, 間野一則, 大室仲, "ピッチ同期雑音励振源をもつ CELP 符号化 (PSI-CELP)," 信学論 A, vol.77-A, pp.314-324, 1994.
- 12) 大矢智之, 須田博人, 三木俊雄, "Pitch Synchronous Innovation CELP (PSI-CELP)-PDC ハーフレート音声 CODEC," 信学技報, RCS93-78, pp.63-70, 1993.
- 13) A. DeJaco, W. Gardner, P. Jacob, and C. Lee, "QCELP: The Northern American CDMA digital cellular variable

- rate speech coding standard,” Proc. IEEE Speech Coding Workshop, pp.5-6, 1993.
- 14) W.B. Kleijn, P. Kroon, and D. Nahumi, “The RCELP speech coding algorithm,” European Trans. on Telecommun., vol.5, no.5, pp.573-582, 1994.
 - 15) TIA/EIA/IS-127, “Enhanced variable rate codec, speech service option 3 for wideband spread spectrum digital systems,” 1997.
 - 16) 3GPP2 C.S0052-A, “Source-Controlled Variable-Rate Multimode Wideband Speech Codec (VMR-WB),” 2004.
 - 17) 3GPP2 C.S0030-0, “Selectable Mode Vocoder (SMV) Service Option for Wideband Spread Spectrum Communication Systems,” 2004.

■2 群 - 8 編 - 3 章

3-3 オーディオ符号化

(執筆者：野村俊之) [2009年12月 受領]

モバイル向けのオーディオ符号化方式は、主に ISO/IEC JTC1/SC29/WG11 (通称 MPEG) において汎用オーディオ符号化方式として標準化されているが、これらの方式をデジタル移動体通信で使用するために、3GPP などで標準化選定が行われた。モバイル特有の環境条件や携帯電話といった演算性能、メモリ量がそれほど高くない端末で動作させるために、低演算量モードでの動作などの各種制約がされている。また、無線伝送路での誤りを考慮し、誤り隠蔽機能などが追加されている。本節ではこれらのうち、代表的な方式について特徴を述べる。

3-3-1 MPEG-4 AAC-LC (Advanced Audio Coding-Low Complexity)

3GPP PSS (Packet-switched Streaming Service: パケット交換網向けストリーミングサービス) におけるオーディオ符号化方式である^{1,2)}。本編 2 章 2-5-4 「AAC」に記載の MPEG-4 AAC LC³⁾ あるいは LTP³⁾ と同一であり、サンプリング周波数は最大 48 kHz、チャンネル数はモノまたはステレオに制限されている。

3-3-2 aacPlus

低ビットレートオーディオ符号化方式である SBR (Spectral Band Replication) を AAC に組合せた符号化方式である^{1,2,4)}。SBR は本編 2 章 2-5-5 「HE-AAC」に記載の MPEG-4 HE-AAC 標準の基本方式であり、低周波数域から高周波数域を複製し、スペクトル形状を補正することにより、聴覚的な劣化を抑えたまま高周波数域に割り当てるビット数を削減している。aacPlus は Enhanced aacPlus のコア方式であり、モノラル信号、または、44 kbit/s 以上のステレオ信号の符号化に用いられる。

3-3-3 Enhanced aacPlus

3GPP PSS (Packet Switched Streaming) 規格で、低ビットレート化のために追加されたオーディオ符号化方式である^{1,2,4,5)}。基本方式は MPEG-4 HE-AAC にパラメトリックステレオ符号化方式を組合せた MPEG-4 HE-AACv2 方式である³⁾。

Enhanced aacPlus は、MPEG-4 HE-AACv2 方式に、携帯機器の実装を容易化するためのモノダウンミックス機能とサンプリング周波数変換機能、伝送路誤りへの性能を強化するための誤り隠蔽機能とが追加されている。

パラメトリックステレオ符号化方式では、左右チャンネル信号をモノラル信号にダウンミックスし、ステレオ情報はチャンネル間の信号比と信号相関などのパラメータ化することにより、符号化効率を改善する。なお、パラメトリックステレオ符号化方式は 44 kbit/s よりも低いビットレートで使用される。サンプリング周波数は最大 48 kHz、チャンネル数はモノまたはステレオであり、960 サンプルのフレームサイズは除外されている。Enhanced aacPlus のエンコーダ、デコーダのブロックならびに関連符号化方式との関係を図 3-7 に示す。

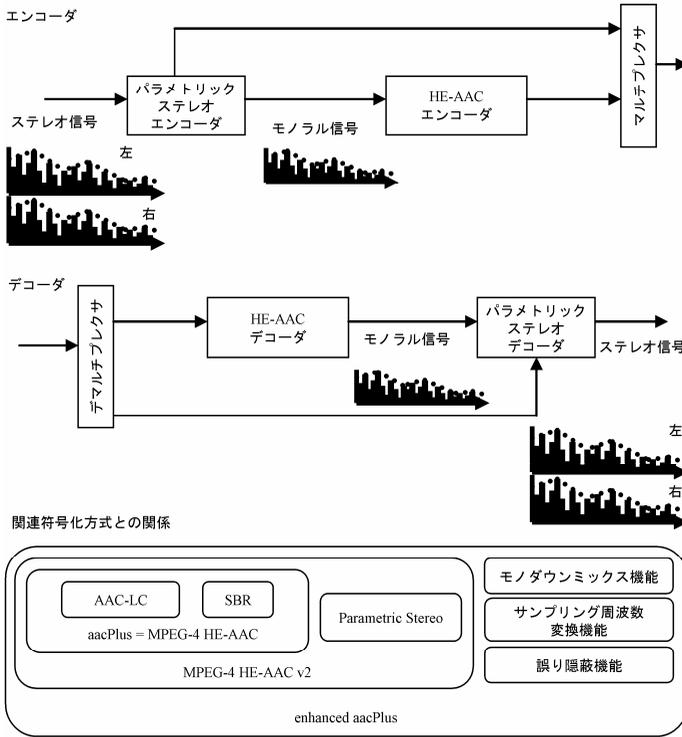


図 3・7 Enhanced aacPlus の動作ならびに関連符号化方式との関係

■参考文献

- 1) 3GPP TS26.401, "Enhanced aacPlus general audio codec: General description," 2004.
- 2) 3GPP TS26.403, "Enhanced aacPlus general audio codec: Advanced Audio Coding (AAC) part," 2004.
- 3) ISO/IEC 14496-3:2009, Information technology: Coding of Audio-Visual Objects -Part 3: Audio, 2009.
- 4) 3GPP TS26.404, "Enhanced aacPlus general audio codec: Spectral Band Replication (SBR) part," 2004.
- 5) 3GPP TS26.405, "Enhanced aacPlus general audio codec: Parametric stereo part," 2004.