
CELP音声符号化

CELP Speech Coding

山根 淳*

Jun YAMANE

要 旨

携帯端末の普及やインターネット電話の広まりなどとともに、伝送帯域などの効率化を狙った低ビットレートの音声圧縮符号化技術に対する要求が高まっている。現在の主流はCELP(Code Excited Linear Prediction coding, コード励振線形予測符号化)方式である。このCELP方式の概要を説明するとともに、音声符号化方式の標準化動向およびCELP方式の処理量削減に着目したリコーの独自方式についても簡単に触れる。

ABSTRACT

As the Personal Digital Cellular systems, Internet phone systems, etc. are widely spread, the low-bitrate speech coding technology is strongly required. Many current speech coding methods are based on CELP(Code Excited Linear Predictive coding). The outline of CELP is explained with the current standardization of speech coding. Also the RICOH original speech coding method based on CELP focused on the reduction of the computational complexity is shortly presented.

1. 背景と目的

携帯電話、携帯端末の急速な普及、インターネットサービス利用者の増加とともに、限られた伝送帯域でより多くの音声通信を行うため、あるいは限られたメモリに

より多くの音声メッセージを蓄積するために、音声をデジタル化し高能率に圧縮する音声符号化方式が望まれている¹⁾。本稿においては、音声符号化方式の現状、代表的な音声符号化方式であるCELP(Code Excited Linear Predictive, 符号励振線形予測)方式²⁾、および我々の取り組みについて述べる。

* 研究開発本部 情報通信研究所
Information and Communication R&D Center,
Research and Development Group

2. 音声符号化方式とその現状

2-1 音声符号化方式の対象帯域

音声符号化を用いたアプリケーションは、その要求する音声品質に基づいた帯域の音声を用いる。帯域は大きく三つに分けることができる(Table 1 参照)例えば、CDやDVDといった高品質の音楽信号を対象とするアプリケーションは、20kHz以下の帯域を必要とする。帯域が広いほど高音質であり、基本的には帯域ごとに異なるアプローチで符号化方式の開発が行われている。本稿では、主に携帯電話などの音声通信や留守録などの蓄積系の音声アプリケーションに用いられている、電話帯域と呼ばれる3.4kHz以下の帯域における音声符号化を中心に述べる。

Table 1 Speech Bandwidth

	帯域	音質	用途
電話帯域音声	< 3.4 kHz	電話	通信, 携帯電話, 留守録
広帯域音声	< 7 kHz	AM ラジオ	広帯域電話, テレビ会議
楽音帯域音声	< 20 kHz	CD	DVD, デジタル放送

2-2 電話帯域音声符号化方式の種類

電話帯域の音声符号化方式は、大きく三種に大別することができる³⁾。(Table 2 参照。)波形符号化方式は、波形をできるだけ忠実に表現しようとする方式である。サンプルごとに量子化する方式や、いくつかのサンプルをまとめて量子化する方式がある。代表的な方式がサンプルごとに一様なステップで量子化するPCM(Pulse Code Modulation)方式である。音声をモデル化せずその信号対雑音比が最も小さくなるように符号化を行うため、十分なビットレートを与えれば高音質であるが、ビットレートの低下とともに急激に音質が劣化する。PHS(Personal Handy phone System)で用いられているADPCM(Adaptive Differential Pulse Code Modulation)は波形符号化方式に属する。

Table 2 Characteristic of Speech Coding

符号化の種類	特徴	方式名
波形符号化	波形を忠実に表現	PCM, ADPCM, DPCM
スペクトル符号化	スペクトルパラメータの符号化	LPC, PARCOR, LSP
ハイブリッド符号化	スペクトルパラメータ+スペクトル量子化	CELP, MELP

スペクトル符号化方式は、音声の周波数スペクトルを符号化し、音声合成モデルによって音声を合成する方式である。線形予測分析などを用いて音声のスペクトルパラメータを抽出し、分析後の残差信号からピッチ周波数などの駆動音源を抽出し、スペクトルパラメータと駆動音源を量子化することによって符号化を行うものである。スペクトル符号化方式は低ビットレートで明瞭な音質を得ることができるが、モデル化を行っているため音声の自然性や個性性は失われる。

ハイブリッド符号化方式は、上の波形符号化とスペクトル符号化を融合するような性格の符号化であり、波形符号化の自然性とスペクトル符号化の低ビットレート性を併せ持つ。CELP方式はこのハイブリッド符号化の代表的な方式であり、CELPの枠組みを用いたさまざまな方式(LD-CELP⁴⁾, VSELP⁵⁾, PSI-CELP⁶⁾, CS-CELP⁷⁾, ACELP⁸⁾など)が開発され、各所で標準採用されている。

2-3 標準化動向

音声符号化の標準化団体としては、ITU(International Telecommunication Union), ISO(International Standard Organization), および各デジタル携帯電話の標準化団体(日本ARIB, 欧州GSM, 北米TIAなど)がある。いずれも独自にニーズに合わせたビットレートなどの仕様による標準化を行っている。各団体と標準化方式との関係をFig.1に示す。現在は、(1)ITU-T低遅延4kbps符号化方式 (2)MPEG4(規則合成音声から高品質音声までカバーするツール群) といった標準化作業が行われている。

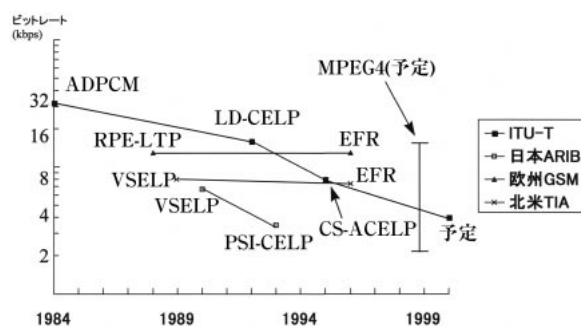


Fig.1 Standardization trend.

2-4 研究動向

今後の音声符号化の研究動向としては、以下のものが挙げられる。

(1)低ビットレート化

現在ITU-Tにおいて4kbpsの標準化が行われており、この標準を狙った方式が開発されているが、人工衛星を用いた携帯電話への応用などに向けた、2.4kbps⁹⁾あるいはそれ以下のビットレートの方式の開発¹¹⁾¹²⁾も行われている。

(2)広帯域符号化方式の低ビットレート化¹³⁾¹⁴⁾

これまでは電話帯域の音声を用いられていたテレビ会議システムやテレビ電話システムにおいて、より臨場感を高めるために広帯域音声を利用されるようになり、それと同時にこの帯域における低ビットレート化への要求も高まってきている。ITU-Tでは、16kbpsで、48kbps G.722(SB-ADPCM方式)と同等の品質を得ることのできる方式の標準化を行っている。

(3)音楽信号符号化方式の低ビットレート化

WWWにおける音楽サービスなど、ネットワークを

通じた音楽信号の伝送が行われるようになってきた。従来音楽信号の圧縮は楽音帯域の高品質の音声を対象に検討が進められていたが、最近では電話帯域においても音楽信号を良好に再生することのできる符号化方式の開発が行われている¹⁵⁾。

(4)演算量削減

PCによるソフトウェア処理や、安価なDSPでの実現を目指し、演算量削減の検討が行われている¹⁶⁾¹⁷⁾。

(5)ビット誤り耐性強化

通信路におけるビット誤りに対する耐性を強化するために、ビット誤りに強い方式の検討が行われている¹⁸⁾¹⁹⁾。

3. CELP 音声符号化方式

3-1 CELP 符号化方式の構成

電話帯域の低ビットレート音声符号化の主流はCELP方式である。CELPは人間の音声発生機構を音源成分(声帯)とスペクトル包絡成分(声道)とによってモデル化し、それぞれのパラメータを符号化する方式である。(Fig.2参照)スペクトル包絡成分は線形予測分析に基づく合成フィルタ、音源情報は予測残差という形で与えられ、予測残差は符号帳に蓄えた時系列符号ベクトルによって構成される。各符号ベクトルをフィルタに通して音声を合成し、入力波形に最も近いものを探索するという、いわゆるA-b-S (Analysis by Synthesis)法による閉ループ探索によって符号ベクトルの決定が行われる。

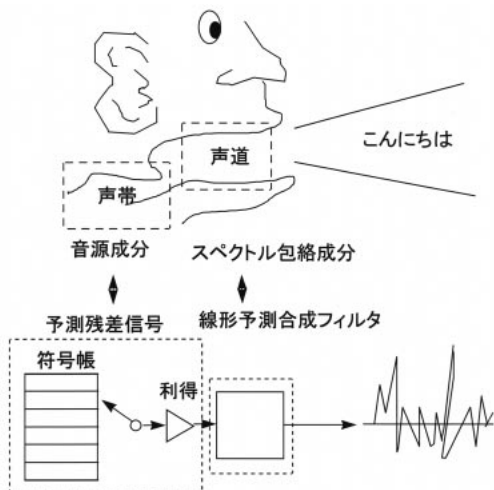


Fig.2 Speech synthesis model of a CELP coder.

3-2 CELP 符号化の動作

Fig.3にCELP 符号化システムの一般的な構成を示す。まず入力音声のある決まったサンプル数のフレームと呼ばれる単位に分割する。このフレーム長は、一般に音声定常であると考えられる20~30ms(サンプリング周波数8kHzで160~240サンプル)あたりの値が用いられることが多い。

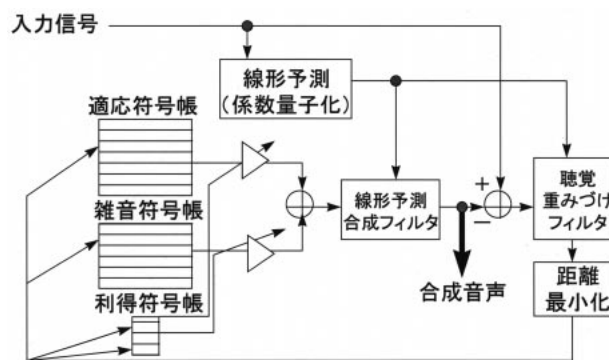


Fig.3 CELP speech coder.

分割されたフレームに対して、そのスペクトル特性に基づいた合成フィルタの係数を求めるために線形予測を行う。具体的には、フレームの自己相関関数を求め、Durbinの再帰解法などを用いて線形予測係数を求める。

線形予測係数は、量子化され、符号化されるが、この量子化は量子化効率の優れたLSP(Line Spectrum Pair, 線スペクトル対)パラメータやPARCOR(PARTIAL auto-CORrelation, 偏自己相関方式)パラメータ、反射係数などに変換して行われることが多い。

求めた線形予測係数に基づき、各フレームの予測残差成分(音源成分)の符号化のために符号帳探索を行う。一般に符号帳は、音声のピッチ成分を表現する適応符号帳と、その他の雑音的な成分を表現する雑音符号帳との二つの符号帳から構成される。また、符号帳探索は、フレームをさらに分割した単位(サブフレームと呼ばれる)に対して行われることが多い。

適応符号帳は、以前のフレームまでの残差成分を、あらかじめ設定された範囲の周期だけ繰り返して構成されたサブフレーム長のベクトル群によって構成される。(Fig.4参照)この時の繰り返し周期は、人間の音声のピッチ周波数(約50~400Hz)を考慮し、サンプリング周波数8kHzの場合で20~150サンプル(約2.5ms~18ms)付近の値を用いることが多い。このようにして構成された各ベクトルは、先に線形予測によって求められた合成フィルタに通され、入力音声と比較され、最も入力フレーム(サブフレーム)に近いものが選択される。一般に比較には、聴覚重みづけが用いられ、各適応符号ベクトルを任意にスカラ倍したとき入力音声に最も近くなるものを探索し、さらに利得成分を量子化するという、いわゆるGain-Shape型の量子化が行われる。

雑音符号帳は、固定的なサブフレーム長のベクトル群によって構成される。そのフィルタ出力を、入力信号から上で求めた適応符号ベクトルのフィルタ出力成分を除いたものと比較し、探索を行う。雑音符号帳についても、Gain-Shape型の量子化により、雑音符号ベクトルおよび利得成分の符号化が行われる。

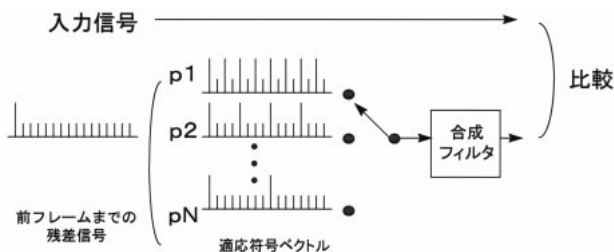


Fig.4 Construction of an adaptive codebook.

復号時は、各符号から、線形予測フィルタを構成し、雑音符号、適応符号、およびそれぞれの利得成分から予測残差信号を構成し、フィルタに通して再生音声を得る。

4. リコーの取り組み

音声符号化方式の主要な設計要素をFig.5に示す。ビットレートのほかに、処理単位長、処理量といった要素仕様がある。それぞれはFig.5に示すとおり、アプリケーションや実装対象のハードウェアに対する要求コストに影響を与える。また、それぞれが音質に影響を与え、基本的にはトレードオフの関係にある。音声符号化を決定・設計する場合は、構成しようとするアプリケーションや機器に対し、これらの仕様のどれを優先するかの見極めが重要になる。



Fig.5 Design factor of a speech coding.

リコーでは、CELPの持つ演算量の多さという欠点を克服する一方式として、雑音源の符号化を簡略化したCELP方式を開発した。本方式においては、入力音声を線形予測分析して合成フィルタを構成し、音声のピッチを表現する適応符号帳探索を行うまでは一般のCELP方式と同様である。ここで、一般的なCELPでは入力信号と適応符号ベクトルの合成フィルタ出力との差信号(以下参照信号と呼ぶ)に対して雑音符号帳探索を行うのに対し、本方式では参照信号を直接量子化することによって雑音源の符号化とする。Fig.6は一般的なCELPの動作であるが、リコー方式は、点線で囲まれた部分を簡略化していることになる。

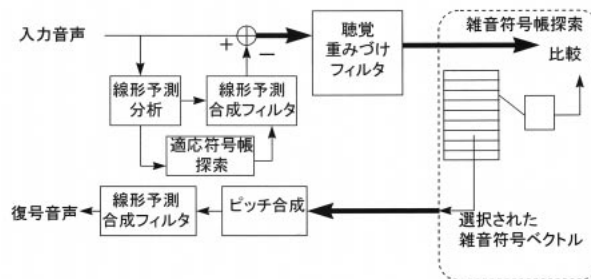


Fig.6 Block diagram of CELP coder/decoder.

5. おわりに・今後の展開

電話帯域低ビットレート音声符号化方式の現在の主流であるCELP方式の概略および標準化動向・リコーの取り組みについて述べた。今後は、独自方式をDSPに実装した場合の問題点を明らかにしていく予定である。

参考文献

- 1) 小野：音声符号化技術の最近の進展，日本音響学会誌，48，1，(1992)pp.52-59．
- 2) M. R. Schroeder, B. S. Atal：Proc. ICASSP 85，(1985)pp. 937-940．
- 3) 守谷：電話帯域音声符号化技術，日本音響学会誌，51，10，(1995)pp.778-783．
- 4) J-H. Chen：Proc. ICASSP 90(1990)pp. 453-456．
- 5) I. A. Gerson, M. A. Jasiuk：Proc. ICASSP-90，(1990)pp. 461-464．
- 6) 三樹他：ピッチ同期雑音励振源をもつ CELP 符号化 (PSI-CELP)，"信学論(A)，J77-A，3，(1994) pp. 314-324．
- 7) A. Kataoka et al.：Proc. ICASSP 94，II-97(1994)．
- 8) R. Salami et al.：Proc. ICASSP 94，II-94(1994)
- 9) 三樹，小口：日本音響学会平成5年度秋季研究発表会講演論文集，1 (1993)pp. 209-210．
- 10) 瀧澤，小澤：日本音響学会平成8年度春季研究発表会講演論文集，1 (1996)pp. 309-310．
- 11) 野村他：日本音響学会平成6年度秋季研究発表会講演論文集，1 (1994)pp. 261-262．
- 12) 佐藤他：1995年電子情報通信学会総合大会講演論文集，B(1995) pp. 510．
- 13) 佐々木，片岡：日本音響学会平成9年度春季研究発表会講演論文集，1 (1997)pp. 297．
- 14) 村島他：1997年電子情報通信学会総合大会講演論文集，情報・システム1 (1997)pp. 346-347．
- 15) 守谷他：1997年電子情報通信学会総合大会講演論文集，情報・システム1 (1997)pp. 354-355．
- 16) 大室，間野：1997年電子情報通信学会総合大会講演論文集，情報・システム1 (1997)pp. 350-351．
- 17) 小澤他：1997年電子情報通信学会総合大会講演論文集，情報・システム1 (1997)pp. 255．
- 18) 山浦他：日本音響学会平成9年度春季研究発表会講演論文集，1 (1997)pp. 283．
- 19) 江原他：1997年電子情報通信学会総合大会講演論文集，情報・システム1 (1997)pp. 256．