

解説 音声処理技術とその応用

2. 音声符号化技術

Speech Coding Technology by Takehiro MORIYA and Kazunori MANO (NTT Human Interface Laboratories).

守谷 健弘¹ 間野 一 則¹

¹ NTT ヒューマンインタフェース研究所

1. はじめに

音声符号化技術とは、音声信号をデジタルの符号に変換し、伝送・記録する際に不必要な情報量を削減することあるいは効率的な表現へと変換することである。圧縮法には完全に元に戻る可逆な圧縮と、再生したときに歪みが残る非可逆な圧縮があるが、本稿では後者を中心に扱う。それは前者の圧縮率は 1/2 程度が限界であるのに対して、後者は聴覚特性も考慮すれば 1/10 以下の圧縮も可能であるために広く研究され、使われているためである。

近年の LSI 技術の急速な発展により、複雑な演算処理と多くのメモリを必要とするが高圧縮率の符号化手法が実現可能となってきた。そして用途に応じた標準規格が制定され、デジタル移動通信や多様な形態のマルチメディア通信に本格的に適用されつつある。

携帯電話、PHS、インターネット上の通信、テレビ会議、多重化装置、ボイスメール、音楽配信など、これらのサービスのすべてにおいて音声、音楽の符号化技術が普及のための大きな原動力の

1 つになっている。

2. 音声符号化の基本技術

2.1 音声符号化の分類

図-1 は代表的な音声符号化方式の情報量と開発あるいは標準化された年代の関係を示すものである。この図のように、音声符号化法は、大きく分けて次の 3 つに分類できる。

(1) 波形符号化

音声信号のデジタル符号化の原点である PCM (Pulse Code Modulation: パルス符号変調) や ADPCM (Adaptive Differential PCM: 適応差分 PCM) のように、原音と符号化音声とのサンプル間の歪みをできるだけ小さくするような符号化方式である¹⁾。ビットレートが高ければ品質はよいが、ビットレートが低くなると急速に品質の劣化が問題となる。圧縮効率は低いが、遅延やコストが少ないという利点があり、PCM は、汎用の有線商用電話の基幹伝送、ADPCM は最近 PHS (Personal Handy phone System) にも使用されている。

(2) ボコーダ(分析合成符号化)

音声をその生成モデルによって分析してパラメータを抽出し、それに基づいて音声信号を合成する符号化方式である。ボコーダでは、分析したパラメータを伝送するだけでよく、波形を忠実に再現する必要がないので、ビットレートを下げることが可能である。しかし逆に、符号化音声品質はモデルに依存し、ビットレートを高くしても品質向上が難しい。PARCOR (PARTIAL auto-CORrelation: 偏自己相関係数)、LSP (Line Spectrum Pair: 線スペクトル対)²⁾に代表される線形予測ボコーダ、MBE (Multi-Band Excitation: マルチバンド励振) 符号化³⁾、WI (Waveform Interpolation: 波形補間) 符号化⁴⁾、

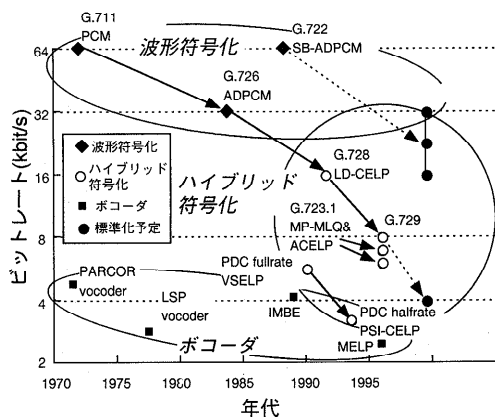


図-1 音声符号化の情報量と開発・標準化年代

MELP (Mixed Excitation Linear Prediction: 混合励振線形予測) 符号化⁵⁾ などがある。これらのボコーダは、玩具や音声応答装置、低ビットレート性の必要な船舶通信などに利用されている。

(3) ハイブリッド符号化

(1)の波形符号化と(2)のボコーダを組み合わせた方式で、信号のモデル化を行い、A-b-S (Analysis-by-Synthesis: 合成による分析) によって、原波形との歪を小さくするように、モデルのパラメータとパラメータで表現できない残りの情報の両方を伝送する方式である。低ビット化を狙った高能率波形符号化ともいえる。CELP (Code Excited Linear Prediction coding: 符号励振線形予測符号化)⁶⁾、マルチパルス符号化などがある。3章で述べる携帯電話のVSELP, PSI-CELP, GSM や、16kbit/s 以下のITU-Tの勧告の方式は、すべてこのハイブリッド符号化である。

2.2 音声の性質と特徴量の抽出

音声の高能率符号化技術は、どのような情報を音声から抽出し、どのようなモデルで音声を合成するかが課題となる。図-2に各種符号化のエンコーダブロック図を示す。図-2(a)は、波形符号化方式としてADPCM、図-2(b)は、ボコーダとして線形予測ボコーダ、図-2(c)は、ハイブリッド符号化としてCELPを示す。

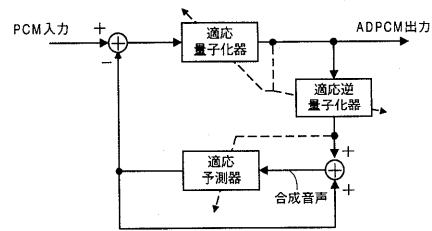
ADPCMでは、簡単な適応量子化器と適応予測器をもち、サンプルごとに更新される。それ以上の音声信号のモデル化はせずに、基本的に波形歪み最小化の構成となっている。

これに対して、線形予測ボコーダとCELPでは、5ms ~ 40ms程度のフレーム単位で音声の特徴を捉え、それを駆動音源と線形予測合成フィルタの形で表現する。合成フィルタの係数(10 ~ 14次)は、スペクトルパラメータとして抽出される。このパラメータは、音声の音色、すなわち、口や舌といった声道の形を表現している。入力となる駆動音源は、有声音の声帯振動や無声音の乱流雑音源、破裂音源に対応している*。線形予測ボコーダでは、有声区間では周期成分としてピッチ周期のパルス列、無声区間では白色雑音系列を用いる。CELPでは、周期成分として過去の駆動音源(適応符号帳)をピッチ周期で繰り返した系列

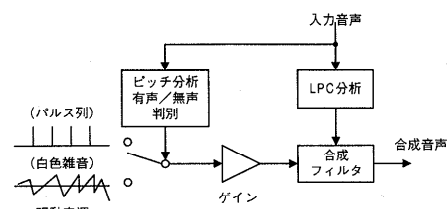
と雑音成分としてあらかじめ用意した雑音パターンを格納した雑音符号帳の中から選択された系列を加え合わせて用いる。

ボコーダは、スペクトル包絡、ピッチ、有声/無声判別、パワのパラメータで合成モデルを構成しており、情報圧縮効率が高く演算量も少ない。ただし、分析誤りによって合成音声に不自然な歪みが出たり、モデルで表現することのできない背景雑音によって劣化しやすいなどの品質上の問題がある。このため音声応答装置などには広く応用されているが、これまで双方向通信にはあまり利用されていなかった。

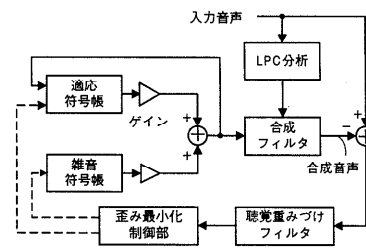
CELPの場合、適応符号帳の周期成分で表しきれないような部分や、あるいはピッチ抽出を誤った場合でも、その誤差をA-b-Sによって雑音成分で補うことになるので、ボコーダ方式より自然性の高い音声を得られる。ただし、伝送情報量を少なくしていくと雑音感が目立つようになる。ま



(a) 波形符号化 (ADPCM)



(b) ボコーダ (線形予測ボコーダ)



(c) ハイブリッド符号化 (CELP)

図-2 各種符号化エンコーダブロック図

* 各種パラメータおよび、音声の基本的な特徴抽出に関しての詳細は、たとえば、文献2)、16)、17)を参考のこと。

表-1 移動通信用音声符号化一覧

地 域	欧州 GSM			日本 ARIB		米国 TIA				
	標準名称	GSM フルレート	GSM エン ハンスト フルレート	GSM ハーフレート	PDC フルレート STC-27C	PDC ハーフレート STC-27C	TDMA		CDMA	
							フルレート IS-54	エン ハンスト フルレート IS-641	フルレート IS-96	エン ハンスト 可変レート IS-127
符号化方式名	RPE-LTP	EFR (ACELP ベース)	EVSELP	VSELP	PSI-CELP	VSELP	EFR (ACELP ベース)	QCELP	EVRC	
ビットレート 音声 [+誤り訂 正] (kbit/s)	13 + 9.8	12.2 + 10.6	5.6 + 5.8	6.7 + 4.5	3.45 + 2.15	7.95 + 5.05	7.4 + 5.6	8/4/2/0.8	8/4/0.8	
標準化年	1988	1996	1994	1990	1993	1989	1996	1994	1997	
品 質	有線商用電 話より悪い	有線商用電 話と同品質	有線商用電 話より悪い	有線商用電 話より悪い	PDCフルレ ートと同程度	GSMフルレ ートと同程度	ほぼ有線商 用電話と同 品質	IS-54より 悪い	IS-54より よい	
演算量(MOPS)	4.5MIPS	18.1WMOPS*	21.2WMOPS	7.8	18.7	8.4	13.54WMOPS	20		
フレーム長(ms)	20	20	20	20	40	20	20	20	20	
先読み(ms)	0	0	5	8	5	8	5	5		

* WMOPS: Weighted MOPS

た符号帳には最低でも 1024 個の候補ベクトルを蓄えておき、すべての候補を合成して入力との歪みを計算する必要がある。提案された 1985 年の段階では演算量的にもメモリ量的にも実用化にはほど遠い技術であった。その後、計算機ハードウェアの進歩と、CELP アルゴリズム自体の効率化、より簡単に演算量の少ない雑音パタンの構成法の研究が進展し、現在では、携帯電話やパソコン上でのソフトウェアでも動作可能な方式が登場してきている。

また、主観品質を向上させるため、聴覚重みづけフィルタによりマスキングの効果によって人間に聞こえにくいように量子化雑音の整形をしたり、ポストフィルタ(後処理フィルタ)によって、スペクトル包絡のピークやピッチハーモニックスの強調によって合成音声の主観品質を向上させている。

2.3 音声符号器の設計条件

音声符号器を設計する際に考慮すべき設計条件は、次のとおりである。

- 品質(音源の種類: 電話, 音楽, 信号帯域幅, 伝送誤りのある場合)
 - ビットレート(固定レート, 可変レート)
 - 遅延時間(双方向, 片方向, 実時間性)
 - 演算コスト(演算量, メモリ量, 消費電力)
- もちろん、品質がよく、ビットレート・遅延時間・演算コストが少ないのが理想であるが、符号化に要求される条件によって、それに最適な設計

を工夫しなければ効率が悪い。たとえば、自動車・携帯電話では、増大する需要に応じて電波のチャネルを確保するために、電話音声の品質が多少劣化していても、まず、ビットレートを低くする。さらに、伝送誤りには強く、双方向で通信可能な遅延、かつ、電池消費電力を少なくするために演算コストを切り詰めなくてはならない。TV 会議では、臨場感を得るためにもう少し帯域を広くとる必要があるし、また、インターネットでの音楽の試聴サービスでは、符号化はオフラインでよいので遅延は大きくてもかまわないが、品質を高く、復号器側での処理の小さいアルゴリズムが必要である。

現在の研究の最前線では、電話帯域(300Hz ~ 3.4kHz)で 4kbit/s、広帯域(100Hz ~ 7kHz)で 8kbit/s、楽音帯域(50Hz ~ 22kHz)で 24kbit/s 程度のビットレートでそれぞれ高品質アルゴリズムの検討が行われている。これは、標本化の 1 サンプルあたり 0.5bit 程度の情報量に相当している。

3. 音声符号化の標準化動向

しばらく前までは、ITU-T 勧告 G.711(PCM)以外の音声符号化は、企業内の専用通信や蓄積再生といった、どちらかという特定の範囲、特定の人の集まりに閉じた技術であった。しかし、携帯電話、インターネット上での通信においては、さまざまなシステムから不特定多数の人が利用す

表-2 ITU-T 勧告の音声符号化一覧

勧告番号	G.711	G.726	G.727	G.728	G.729	G.723.1	(G.4k)	G.722	(G.wb)
符号化方式	μ 則 PCM (パルス符号 変調)	ADPCM (適応差分 パルス符号 変調)	エンベデッ ド ADPCM	LD-CELP (低遅延符号 励振線形予 測)	CS-ACELP (共役構造一 代数的符号励 振線形予測)	MP-MLQ/ ACELP (マルチパル ス最ゆう量子 化/代数的符 号励振線形予 測)	(現在, 候補 募集中)	SB-ADPCM (帯域分割 ADPCM)	(現在, 候補 募集中)
符号器大分 類	波形符号化	波形符号化	波形符号化	ハイブリッ ド符号化	ハイブリッ ド符号化	ハイブリッ ド符号化	未定	波形符号化	未定
ビットレ ート (kbit/s)	64	32 (40, 24, 16)	40, 32, 24, 16	16 (12.8, 9.6)	8	6.3 5.3	4	64 (56, 48)	16 24, 32
帯域(kHz)	0.3-3.4	0.3-3.4	0.3-3.4	0.3-3.4	0.3-3.4	0.3-3.4	0.3-3.4	0.05-7	
勧告年	1972	1984, 1990	1990	1992	1996	1996	1999 予定	1992	1999 予定
品 質	商用電話 品質	商用電話 品質	商用電話 品質	商用電話 品質	商用電話 品質	ほぼ商用 電話品質	商用電話 品質	広帯域 AM 放送程度	16k: G.722(48k) 24k: G.722(56k) 32k: G.722(64k)
演算量 (MIPS)	1 以下	2 以下	2 以下	30	20 (10: Annex A)	20	商用 DSP 1 チップ相当	10	商用 DSP 1 チップ相当
フレーム長 (ms)	0.125	0.125	0.125	0.625	10	30	20 (10, 5...も可)	1.5	20 (10, 5...も可)
先読み(ms)	0	0	0	0	5	7.5	15	0	20
主用途	汎用音声通 信, 基幹伝送	汎用音声通 信, 基幹伝送 (PHS)	汎用音声通 信, 基幹伝送	企業内通信, TV 会議, 次 世代携帯電話	企業内通信, インターネット 電話, 次世 代携帯電話, DSVD モデム	企業内通信, インターネット TV 電話	企業内通信, 次世代携帯電 話, インター ネット電話	TV 会議, 教 育用高品質双 方向通信	TV 会議, 高品質 双方方向通信, オ ーディオ

るようになるので、標準化によって、広域で透過性の高い音声通信を実現する必要がある。

3.1 移動通信用の音声符号化

移動通信(自動車・携帯電話)においては、当初デジタル化を行うと、既存サービスであるアナログ周波数変調(FM)方式より電波利用効率が低くなってしまふという問題があった。利用効率を上げるために 8kbit/s 以下のビットレートで品質を確保し、かつ、伝送路符号誤り耐性、遅延条件、さらに携帯電話では消費電力の制限があり、これらの厳しい要求条件の中で符号器を設計しなくてはならなかった。

しかし、音声符号化技術とそれに合わせた効率的な誤り訂正符号化技術、さらに LSI 技術の急速な進展により、1987 年には日米欧でそれぞれ、移動通信のデジタル化の方針が打ち出された。表-1 に、各地域での移動通信用音声符号化一覧を示す。欧州では 1988 年、北米で 1989 年、日本では 1990 年に標準方式(フルレート)が選定された。欧州では RPE-LTP(Regular Pulse Excitation with Long-Term Prediction: 長期予測付等間隔パルス励振)符号化⁷⁾、北米、日本では VSELP(Vector Sum Excited Linear Predic-

tion: ベクトル加算励振線形予測)符号化⁸⁾が選定された。これらは、いずれもハイブリッド方式の符号化法である。

日本では半分の情報量のハーフレート符号化の標準化が進み、PSI-CELP(Pitch Synchronous Innovation CELP: ピッチ同期雑音励振源 CELP)符号化^{9), 10)}が選定された。この方式の音声符号化ビットレートは、3.45kbit/s であり、商用化された携帯電話用コーデックとしては世界で最も低いビットレートである。

欧州でもハーフレート化が進められたが、それと同時に、フルレートの品質を改善したエンハンスドフルレートの標準化が行われた。表以外にもさらに可変レート化の標準化も行われるようである。

北米では無線アクセス方式として TDMA(Time Division Multiple Access)と、CDMA(Code Division Multiple Access)の 2 方式があり、それぞれについて標準化が行われている。TDMA 方式では、固定ビットレートの音声符号化であるのに対し、CDMA 方式では、可変ビットレートの符号化を用いる。TDMA 方式については、当初ハーフレート化の検討が行われたが中

断し、TDMA、CDMA とともにフルレートの音質改善要請にしたがってエンハンスドフルレートの標準化が先に行われた。

3.2 ITU-T 国際標準

表-2 に、ITU-T 勧告の音声符号化一覧を示す。1970 年代の G.711 (64kbit/s PCM)、1980 年代の G.726 (32kbit/s ADPCM) に続いて、遅延時間の少ない汎用方式として G.728 (16kbit/s LD-CELP)¹¹⁾ の標準化が行われた。

ITU-T では、音声符号化への品質要求が高く、基本的にビットレートが下がっても品質は、商用電話品質として ADPCM 相当の品質を要求されている。G.728 に次いで、ADPCM と同等品質の 8kbit/s 音声符号化の標準化が行われ、フレーム長 10ms の G.729 (CS-ACELP)¹²⁾ となった。これは、基本品質とともに無線用途を考慮し、符号誤り耐性に強い設計となっており、伝送誤りのある条件においては、ADPCM と比較して、はるかに品質がよい。G.723.1 はテレビ電話などのマルチメディア通信用の符号化として選定されたものであり、6.3kbit/s、5.3kbit/s の 2 つのビットレートをもつ符号化方式である。ただし、フレーム長は 30ms であり遅延は大きい。

現在、さらにフレーム長を 20ms に抑え、品質は G.726 と同等という条件の 4kbit/s の符号化が課題となっており、候補募集中である。

3.3 そのほかの標準

情報圧縮効率を高くしようとすると、用途によって最適化基準が異なるために、個別の標準化の必要性が高くなる。1989 年には米国連邦政府の専用線標準方式として 4.8kbit/s の CELP が標準化された。その後、米国政府ではそれまでの 2.4kbit/s の線形予測ボコーダに代わる新たな方式の選定作業を行い、ボコーダ方式の一種である MELP を 2.4kbit/s の標準化方式に決定した。

これとは独立に 1989 年に INMARSAT による衛星通信用音声符号化標準として、IMBE (Improved Multi-Band Excitation) 方式¹³⁾ が標準化された。この方式は、CELP とはまったく異なり、スペクトル包絡を再現することを規範とするボコーダである。

4. 広帯域音声および楽音の情報圧縮

前章までは 0.3 ~ 3.4kHz 信号帯域の電話音声を対象にした情報圧縮符号化について紹介したが、7kHz 帯域 (AM 放送程度) や 15kHz 帯域 (FM 放送)、さらに 20kHz の HiFi 音楽の情報圧縮技術についても標準化や実用化が進んでいる。電話音声を狭帯域音声というのに対して、7kHz 帯域の音声は、広帯域音声と呼ばれ、より自然性の高い音声である。スピーカを用いた TV 会議や教育用の高品質の双方向通信や蓄積への適用が想定されている。一方、楽音の符号化は放送、CATV、CD-ROM などへの蓄積といった片方向のみのシステムへの適用が中心である。したがって楽音の符号化では、遅延時間の条件が緩いことや、復号器のみのコストが問題となるという特徴がある。

4.1 広帯域音声の符号化

表-2 に示したように、7kHz 帯域音声の符号化としては、64kbit/s の SB-ADPCM 方式が ITU-T 勧告 G.722 として標準化されている。この方式は遅延や処理量が少ないという利点があるが、音楽を入力したときの品質は必ずしも保証されていない。

現在、G.722 よりも低ビットの 16 ~ 32kbit/s での符号化法が ITU-T の課題として採り上げられており、2 ~ 3 年のうちに標準化される見込みである。この標準化において最も重要視されている点は、通信用途を考慮してフレーム長が 20ms 程度であること、音声と同時に楽音に対しても品質を確保していることである。

4.2 楽音の符号化

楽音の符号化では DAT (Digital Audio Tape: 片チャンネル 768kbit/s) の半分程度まで圧縮する手法が DCC (Digital Compact Cassette) や MD (Mini Disc) に採用され商品化されている。ISO では MPEG-1 (Moving Picture Expert Group) の標準として、64 ~ 192kbit/s の符号化を決定した¹⁴⁾。

また、この標準規格を 16kHz までの標準化周波数の入力や多チャンネルの入力に拡張する規格として MPEG-2 がある。その中で、1997 年にはマルチチャンネル (通常 5 チャンネルまで) の AAC (Advanced Audio Coder) と名づけた規格も標準化された。チャンネルあたり 64kbit/s で 48kHz で

サンプリングされた信号を符号化しても品質の劣化はほとんど検知できないことが確かめられている。

現在、MPEG-4として、8～48kHz帯域、2～64kbit/sの音声、楽音符号化方式(ツール群)も1998年の完成を目指して標準化作業中である。

5. マルチメディアへの応用と今後の課題

音声や楽音の符号化技術は携帯電話やインターネット電話を含む電話、音楽を鑑賞するなど、日常生活のさまざまな製品に欠くことのできない基本部品に应用されつつあり、これまでになく大きな影響を与えることは確実である。

インターネット電話、テレビ電話などパソコンを利用したネットワーク経由での音声通信では、従来は、独自の符号化方式であったが、相互接続性のニーズと標準化の進行により、標準化の符号化方式も利用可能となりつつある。たとえば、ITU-TのQoS非保証型LAN上でのビジュアル電話の勧告H.323では、音声コーデックとして、音声主体の場合には、遅延が少なく、高品質のG.729を用い、また、テレビ電話といったマルチメディア利用の場合には、G.723.1を用いることが推奨されている。今後これらの標準を搭載したパソコンが急速に広まるであろう。

ネットワーク上で音声通信を利用する場合、無音区間圧縮をすることにより、不必要なパケットを送信しなくてすむ。G.729とG.723.1に関しては、VAD(Voice Activity Detector)と呼ばれる音声区間検出器に関してすでにAnnex(付属資料)として勧告されている¹⁵⁾。

また、G.729のAnnex Aとして、DSVD(Digital Simultaneous Voice and Data: デジタル音声・データ同時通信)用に、10MIPS以下に演算量を削減したバージョンも勧告されている。

さらに、DSP用の固定小数点版のほかに浮動小数点版についても勧告が準備されている。これにより、パソコンを使った、ソフトウェアコーデックとして音声符号化を行うことができるので、ハードウェアを追加することなく、さまざまなシステムに適用可能になる。

今後、映像やほかのデータとの整合性をよくするために、圧縮以外のメリットをもつ音声符号化

の検討が必要である。たとえば、音声の編集加工性や、スケーラビリティ、ほかのメディア(テキスト、楽譜、静止画、動画など)との同時符号化、電子透かしなどのセキュリティ確保といったことが必要になるだろう。

一方、音声符号化自体の重要な課題として、背景雑音(車内、オフィス、話し声、BGM)下での品質向上やモバイル用途のための伝送誤り(ビット誤り、フレーム消失)耐性の向上技術の開発がある。広帯域の符号化に関しては、音声と音楽の両方に性能のよい手法の開発が必要である。

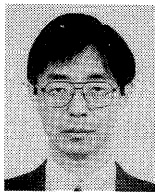
6. おわりに

音声の符号化技術の動向と最新の標準化活動を関連させて紹介した。通信システムの実用化には標準化が大きな役割を果たしている。また標準化方式は本来の用途以外の必ずしも標準化の必要のないシステムへも適用可能である。適用するにあたっては伝送情報量、品質、遅延、符号誤り耐性、ハードウェアコストなどを総合的に評価する必要がある。今後、音声符号化は、単体として使われるだけでなく、マルチメディアを駆使したコンピュータ通信の1つの機能としてほかの機能と組み合わせられて使われるようになると考えられる。

参 考 文 献

- 1) 安田浩編著：マルチメディア符号化の国際標準，276p.，丸善(1991)。
- 2) 斎藤収三，中田和男：音声情報処理の基礎，268p.，オーム社(1981)。
- 3) Griffin, D. R. and Lim, J. E.: Multiband Excitation Vocoder, IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol.36, No.8, pp.1223-1235 (1988)。
- 4) Kleijn, W. B.: Encoding Speech Using Prototype Waveforms, IEEE Trans. on Speech and Audio Processing, Vol.1, No.4, pp.386-399 (1993)。
- 5) McCree, A. V. and Barnwell III, T. P.: A Mixed Excitation LPC Vocoder Model for Low Bit Rate Speech Coding, IEEE Trans. on Speech and Audio Processing, Vol.3, No.4, pp.242-250 (1995)。
- 6) Shroeder, M. R. and Atal, B. S.: Code-Excited Linear Prediction(CELP): High Quality Speech at Very Low Bit Rates, Proc. ICASSP'85, pp.937-940 (1985)。
- 7) Natvig, J. E.: Evaluation of Six Medium Bitrate Coders for the Pan-European Digital Mobile System, IEEE JSAC, Vol.6, No.2, pp.324-331

- (1988).
- 8) Gerson, I. and Jusiuk, M.: Vector Sum Excited Linear Prediction(VSELP) Speech Coding at 8 kbit/s, Proc. ICASSP'90, 33.S9.3 (1990).
 - 9) 大矢, 須田, 三木: Pitch Synchronous Innovation CELP (PSI-CELP)-PDC ハーフレート音声 CODEC, 信学技報, RSC93-78 (Nov. 1993).
 - 10) 三樹, 守谷, 問野, 大室: 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, pp.314-324 (Mar. 1994).
 - 11) Chen, J.-H., Cox, R. V., Lin, Y.-C., Jayant, N. and Melchner, M. J.: A Low-delay CELP Coder for the CCITT 16 kb/s Speech Coding Standard, IEEE JSAC, Vol.10, No.5, pp.830-849 (1992).
 - 12) 特集: ITU 国際標準 8kbit/s 音声符号化方式, NTT R&D, Vol.45, No.4, pp.317-348 (1996).
 - 13) Hardwick, J. C. and Lim, J. S.: The Application of the IMBE Speech Coder to Mobile Communications, Proc. ICASSP'91, pp.249-252 (1991).
 - 14) 安田浩編著: MPEG /マルチメディア符号化の国際標準, 265p., 丸善 (1994).
 - 15) 問野: 低ビットレートの音声符号化, インターネットでも適用始まる, 日経コミュニケーション, No.230, pp.132-140 (Sep. 1996).
 - 16) 古井貞熙: デジタル音声処理, 238p., 東海大学出版会 (1985).
 - 17) Kleijn, W. B. and Paliwal, K. K.: Speech Coding and Synthesis, 755p., Elsevier (1995).
(平成9年9月9日受付)



守谷 健弘 (正会員)

1978年東京大学工学部計数工学科卒業。1980年同大学院修士課程修了。同年日本電信電話公社武蔵野通研入社。1989年東京大学より工学博士。1989年AT&Tベル研究所客員研究員。入社以来、音声、楽音の高効率符号化の研究に従事。現在、NTTヒューマンインタフェース研究所特別研究員。電子情報通信学会論文誌編集委員。IEEE SP ソサエティ技術委員会委員。情報処学会規格調査会オーディオ小委員会幹事。
e-mail:moriya@splab.hil.ntt.co.jp



問野 一則 (正会員)

1982年早稲田大学理工学部電気工学科卒業。1987年同大学院理工学研究科博士課程修了(工学博士)。同年日本電信電話(株)入社。以来、音声の高効率符号化の研究に従事。現在、NTTヒューマンインタフェース研究所音声情報研究部主幹研究員。電子情報通信学会、日本音響学会音声研究会幹事。e-mail:mano@splab.hil.ntt.co.jp

付録 2章の用語解説(欧文先頭, 50音順)

A-b-S :

音声の合成モデルを仮定し、再合成信号と入力信号の誤差を最小にするモデルのパラメータを探索によって求め、そのモデルの最適値とする分析法。

CELP :

線形予測合成フィルタを駆動する音源を雑音符号帳のベクトルパタンを用いてベクトル量子化する音声符号化法。

LSP :

線形予測係数と等価な音声のスペクトル概形を表現する周波数領域のパラメータ。すべての係数値が昇順であればその合成フィルタは安定であり、線形予測係数や PARCOR よりも量子化効率と補間効率がよい。

MBE :

正弦波重畳モデルによるボコーダの一種。音声の周波数スペクトルを複数の帯域に分け、有声帯域では、高調波成分を正弦波で表現し、無声帯域では、白色雑音で表現する。

MELP :

線形予測合成フィルタを駆動する音源を周波数帯域ごとに有声・無声判断し、ピッチ周期パルスと白色雑音の重みつき和によって表現するボコーダ。

PARCOR :

線形予測係数と等価な音声のスペクトル概形を表現するパラメータ。すべての係数の絶対値が1より小さければその合成フィルタは安定であり、線形予測係数よりも量子化効率がよい。

PSI-CELP :

CELP の有声音の符号化において、雑音符号帳ベクトルをピッチ周期で繰り返し、駆動音源とすることを特徴とする音声符号化法。

RPE-LTP :

線形予測合成フィルタを駆動する音源を等間隔位置のパルスとピッチ周期性を付与する長期予測器によって表現する音声符号化法。

VSELP :

CELP 符号化において、少数の基底ベクトルの加減算により雑音符号ベクトルを合成し、駆動音源とすることを特徴とする音声符号化法。

WI :

分析区間ごとに音声あるいは残差信号をピッチ単位で抽出し、高速変動成分と低速変動成分に分離して量子化し、前区間との補間によって音声を再合成するボコーダ。

スケラビリティ :

信号を階層的に符号化し、任意の条件(ビットレート、品質、帯域、演算量など)に応じて、符号系列の一部を用いて、復号を可能とする機能。スペクトルパラメータ :

信号の周波数特性を表現するパラメータ。

スペクトル包絡 :

信号の周波数振幅特性の概形。音声の場合、声道形状に対応して変化する。

多重化装置 :

音声を圧縮符号化して、1回線あたり複数の音声をまとめて伝送できるようにする装置。

適応予測器 :

入力信号または、伝送符号に応じて、予測誤差を小さくするように予測係数を制御する機能をもつ予測器。

適応量子化器 :

入力信号または、伝送符号に応じて、量子化歪みを小さくするように量子化のステップ幅、信号の正規化を制御する機能をもつ量子化器。

白色雑音 :

周波数振幅特性が一定な不規則信号。

パケット :

情報を伝送する時に、一定の大きさに分割もしくはまとめたもの。音声情報通信では、無音区間を利用して、伝送量を低減できる。

パワ :

信号の2乗平均。

ピッチ周期 :

人間の声の高さに対応する声帯の振動周期。基本周期とも呼ぶ。

ビットレート :

単位時間(1秒)あたりに伝送するビット数のこと。bit/sで表す。

ボコーダ :

音声を分析して、特徴量をパラメータ抽出し、それに基づいて合成する符号化法。波形そのものは再現されないが低ビット化が可能である。

マスキング :

音がほかの音によって聞こえにくくなること。符号化ではこの効果を利用して、量子化雑音を整形し、雑音感を低減することができる。

マルチパルス符号化 :

線形予測合成フィルタを駆動する音源を複数パルスの位置とその振幅によって表現する音声符号化法。