論文 / 著書情報
Article / Book Information

論題(和文)	音声合成技術とその応用
Title(English)	
著者(和文)	古井 貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	テレビジョン学会誌, Vol. 41, No. 8, pp. 707-715
Citation(English)	, Vol. 41, No. 8, pp. 707-715
発行日 / Pub. date	1987,
権利情報 / Copyright	本著作物の著作権は映像情報メディア学会に帰属します。 Copyright (c) 1987 Institute of Image Information and Television Engineers.

2. 音声合成技術とその応用

古井貞熙†

1. まえがき

人間が音声を発声するメカニズムを電気回路などで実現し，音声を人工的に生成することを音声合成という．音声合成技術の背景には，その基本として音声符号化技術があり，広義には音声符号化を含めて音声合成という．文字などを用いる場合に比べて，音声合成による情報伝達には次のようなメリットがある[1]．

（1） 特別な注意や訓練なしで，誰でも容易に内容が理解できる．

（2） 移動したり作業したりしながら，また特に注意していなくても割り込んで情報を伝えることができる．

（3） 電話機が使えるので，特別な装置なしで遠隔地へ容易に伝えられる．

（4） 紙資源を消費しない．

逆に欠点としては，記録が残らない，拾い読みができない，大量かつ複雑な情報の伝達には適さない，などがある．近年，優れた音声合成方法が開発されたこと，LSIやコンピュータの発達によって比較的安価に装置が作れるようになったことなどにより，電話による種々の案内サービスなど広い範囲で音声合成が用いられるようになってきた．

以下では，音声合成と音声符号化の最近の技術動向について，特に基本技術に絞って解説してみたい．

2. 音声合成と音声符号化技術の概要

2.1 音声合成

音声合成の方法は，次の3種類に分類できる．

（a） 録音編集方式

人が発声した単語，文節，定型文章などをアナログまたはディジタル的に録音して，テープ，ディスク装置，LSIなどに蓄えておき，合成したい言葉に応じて，蓄えられた音声をつないで再生する．

（b） パラメータ編集方式

音声波形をそのまま蓄えておくのではなく，波形を分析して物理的パラメータの形で蓄えておき，これらのパラメータによって音声合成器を制御して音声を再生する．

（c） 規則合成方式

単語より小さい音素，音節などの基本単位を用いて，これらの結合法や韻律情報（アクセント，イントネーションなど）に関する規則により，音声合成器の制御パラメータを生成し，音声を出力する．

このうち，(a)と(b)では，出力する音声をあらかじめ人が発声しておく必要があるので，合成できる語彙が限られるが，(c)では任意の言葉を合成することができる．(a)と(b)の比較では，後者の方が音声を少ない情報量で蓄えることができるので，語彙数が大きい場合に適している．ただし(a)よりも(b)，さらに(c)と装置が複雑，高度化し，合成音の明瞭性，自然性などの品質が低下する．任意の文章を人間の声そっくりの自然な合成音で出力できるためには，まだ多くの研究が必要である．

2.2 音声符号化

音声符号化は，文字通り音声をできるだけ少ない情報量で表現するものであり，上述の録音編集方式とパラメータ編集方式における音声蓄積，規則合成方式における基本単位の蓄積の他に，音声のディジタル伝送や，いわゆるボイスメールサービスのための音声蓄積などの目的がある．音声が極めて少ない情報量で伝送できるようになれば，無線回線や国際間の長距離回線を経済的かつ有効に使えるほか，いわゆるISDN (integrated-services digital network)において，音声とデータを一括して扱えるようになり，アナログ伝送よりも容易に秘話性を高めることができるなど，多くのメリットが期待される．

音声符号化技術は，次の2種類に分類できる．

† NTT 電気通信研究所
2. "Trends of Speech Synthesis Technology and Its Application" by Sadaoki Furui(NTT Electrical Communications Laboratories. Tokyo)

（a） 波形符号化方式

音声の波形に含まれる聴覚的な冗長性を用いて，波形を能率的に表現する方法で，音声合成においては，録音編集方式による場合の情報圧縮に用いられる．

（b） 分析合成方式

音声の生成モデルに基づいて，音声をそのモデルパラメータに変換して表現する方法で，ボコーダ（vocoder）とも呼ばれる．音声合成では，パラメータ編集方式や規則合成方式の基本技術として用いられる．

音声の波形符号化のさきがけはPCM（Pulse Code Modulation）による方法で，64 kb/s程度の情報量を必要としたが，最近ではその数分の1の情報量で符号化する方法が研究されている．波形符号化よりも分析合成方式の方が少ない情報量で音声が表現できるが，音声の品質は低下する．波形符号化方式と分析合成方式に共通に適用できる優れた情報圧縮方法として，最近，ベクトル量子化法が盛んに研究されている．

3. 波形符号化の原理

波形符号化で用いられる基本的な情報圧縮技術は次の通りである．

（1） 非線形量子化

音声振幅の統計的性質を用いて，振幅を非線形（通常は対数変換）して圧縮し，そののち線形に量子化する．

（2） 適応量子化

音声振幅の動特性の非定常性に対処するため，量子化器のステップ幅を振幅に応じて時間的に変化させる．

（3） 予測符号化

音声信号は隣接標本間，さらに離れた点の間でも相関があるので，隣接標本間の差信号（差分）の符号化，あるいはその相関を利用して一定区間の時系列から予測した値と実際の標本値との差信号（予測残差）の符号化により情報圧縮する．

（4） 時間分割と帯域分割符号化

音声を複数の時間区間，あるいは周波数帯域に分割して，振幅の大きい区間や聴覚的に重要な帯域に多くの情報を割り当てる．

（5） 直交変換符号化

聴覚の冗長性にもとづき，音声がほぼ定常と考えられる20 ms程度を1ブロックとして，DCT（Discrete Cosine Transform）等で周波数領域に直交変換した量を符号化する．

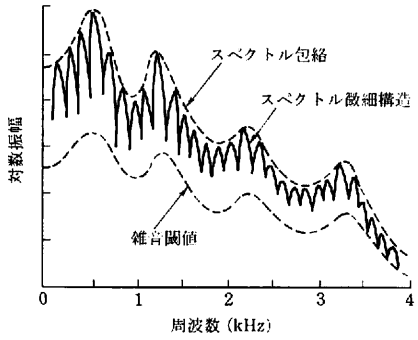


図1 聴覚的マスキングによる雑音閾値と，スペクトル包絡の関係（雑音を閾値以下に抑えれば知覚されない）

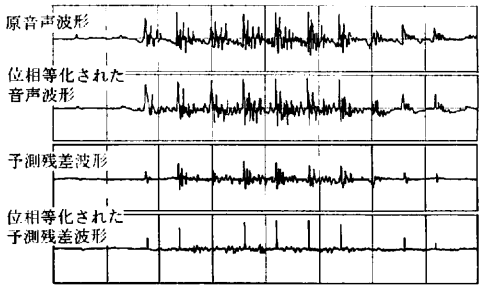


図2 男性音声の原音声と予測残差波形に対する位相等化処理結果の例（位相等化を施しても聴覚的な品質はほとんど変化しない）

さらに，聴覚的なひずみを増加させないで符号化効率を上げる方法として，次のような人間の聴覚特性を利用した処理が行われている．

（6） ノイズシェイピング

図1に示すように，音声の周波数スペクトルよりもある程度（15 dB程度）低い閾値以下の雑音は，聴覚的に知覚されないという現象（聴覚的マスキング現象）を利用して，雑音のスペクトルを整形・調整する処理である．これにより，同じSN比（信号対量子化雑音比）でも，聴覚的な雑音を低減することができる．

（7） 位相等化処理

人間の聴覚は位相に対して冗長性があるので，波形を位相等化（零位相化）することにより，図2に示すように波形の電力に時間的片寄り（集中化）を与え，電力の大きい部分に情報割当てを行うことによって符号化効率を高めることができる[2]．

4. 音声分析合成の原理

4.1 音声生成モデル

音声生成のプロセスは，次のように電気的な等価回路でモデル化することができる．

（1） 有声音源

声の大きさに対応するピーク値と，声の高さに対応する繰り返し周波数（基本周波数，逆数が基本周期）を有するパルスまたは三角波．

（2） 無声音源

声の大きさに対応する平均エネルギーを有する白色雑音．

（3） 調　音

単一共振・反共振回路の縦続/並列接続を表現する多段ディジタルフィルタ．

（4） 放　射

6 dB/oct の微分特性．

現在の音声情報処理ではこれをさらに簡単化して，音源 $G(\omega)$ と調音（共振・反共振特性）$H(\omega)$ の電気的等価回路が接続された形で，音声波形 $S(\omega)$ が生成されるとする，次のような線形分離等価回路モデルが主として用いられる．ω は角周波数である．

$$S(\omega)=G(\omega)H(\omega) \qquad (1)$$

通常，図3に示すように，音源はパルス音源と雑音源で近似し，調音は全極型または極零型のフィルタ特性で近似する．声帯波のマクロな周波数特性は，放射特性とともに声道フィルタ特性に一括して表現される．したがって $G(\omega)$ のマクロな周波数特性は平坦で，$H(\omega)$ は時変係数ディジタルフィルタで実現される．音声に含まれる重要な情報である音韻性情報は，主として $H(\omega)$ として表現される．音声波形に含まれる位相特性は人間の聴覚による音声の知覚には重要な役割を果たしていないので，実音声からの $H(\omega)$ の分析と再生においては，通常，振幅特性のみに着目し，

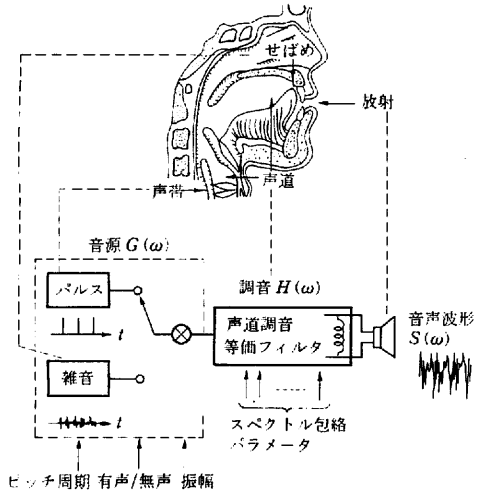


図 3　人間の発声器官と音声生成機構の線形分離等価回路モデル

スペクトルとして表現することが多い．

音声の分析合成では，通常音声波を線形分離等価回路モデルに基づいて分析して，音源と調音に関する次の特徴パラメータで表現し，これらの特徴パラメータを用いて音声を再生（合成）する．

有声音/無声音の区別
基本周期(パルス音源の繰り返し周期)　…音源情報
音源振幅(パルスおよび雑音の大きさ)

線形フィルタ特性…スペクトル包絡(調音)情報

スペクトル包絡の表現法によって，これまでに種々の分析合成系が提案されているが，近年では線形予測（LPC：Linear Predictive Coding）分析に基づくものが多い．

4.2 線形予測(LPC)分析

線形予測分析法には，音声波形やそのスペクトルの性質を極めて少数のパラメータで能率的かつ正確に表現でき，しかもそれらのパラメータが比較的簡単な計算で求まるという特徴がある．

この分析法では，現時点の標本値 x_t と隣接する過去の p 個の標本値との間に，次のような線形1次結合（線形予測モデル）が成り立つと仮定する[3)4)]．$\{\alpha_i\}$ を線形予測係数と呼ぶ．

$$x_t+\alpha_1 x_{t-1}+\cdots+\alpha_p x_{t-p}=\varepsilon_t \qquad (2)$$

$\{\varepsilon_t\}$ は，平均値 0，分散 σ^2 の互いに無相関の確率変数であり，線形予測残差と呼ばれる．上式が成り立つ時系列は，自己回帰（AR）過程とも呼ばれる．

この方法は，音声波形のスペクトル密度 $f(\omega)$ が全極型の有理スペクトル密度で表されると仮定することと等しい[5)]．このときパラメータ $(\sigma^2, \alpha_1, \alpha_2, \cdots, \alpha_p)$ は，その尤度が最大となるように推定することができるので，この方法は最尤スペクトル推定法とも呼ばれる．この方法は，スペクトルの山形部分を重視した推定法になっており，人間の聴覚特性と合致している．この方法は逆フィルタ法と呼ばれることもある．

4.3 PARCOR 分析合成系

線形予測分析法の一種として，より少ない情報量でスペクトル包絡が表現でき，音声合成回路の構成が比較的容易にできる PARCOR 方式[6)] が最近よく用いられている．PARCOR 係数（偏自己相関係数）を用いた分析合成系の構成を図4に示す．同図に示すように，一般に線形予測分析に基づく分析合成系では，入力信号の短時間スペクトルを分析フィルタで正規化（平坦化）することによって得られるスペクトル微細構造の相関関数，すなわち予測残差信号の相関関数によって，音源の周期性の有無と基本周期が抽出できる．PARCOR 分析合成系では，10～40 ms 毎の，10 個程

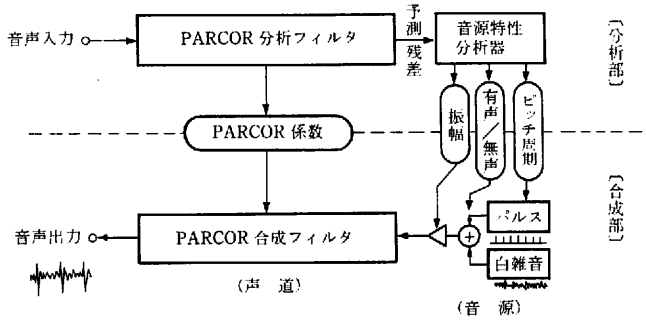


図 4　PARCOR 分析合成系の構成

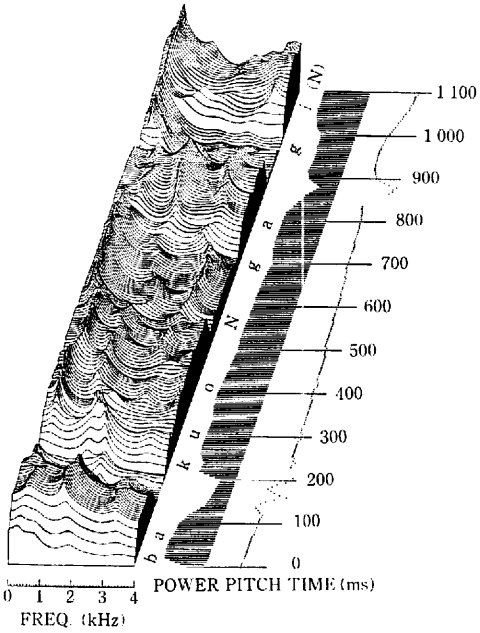


図 5　女性が/bakuoNga gi(N)…/と発声した音声を PARCOR 分析合成系により 2.4 kbps で符号化したときのスペクトル包絡，パワー，およびピッチ(基本周波数)の時間変化

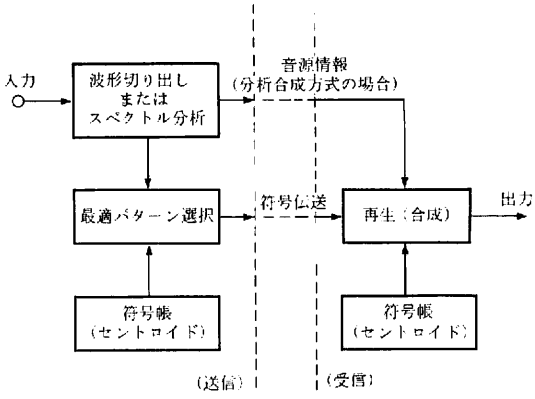


図 6　ベクトル量子化の原理

度の PARCOR 係数と音源に関するパラメータで音声が表現されるので，2.4 kb/s 程度の情報量（PCM の約 1/30）で音声を伝送・再生することができる．図 5 は，2.4 kb/s で符号化したときのスペクトル包絡，音声パワー，および基本周波数を，女性の「爆音が銀……」という音声の場合について示したものである．

PARCOR 係数は時間領域のパラメータであるが，さらに，より少ない情報量で能率良く音声の性質を表現するため，周波数領域のパラメータとして LSP（線スペクトル対）が開発されている[7]．LSP 方式を用いると，PARCOR 方式の約 60%の情報量で，同等の合成音を得ることができる[8]．

5.　ベクトル量子化の原理

波形符号化や分析合成系において，波形やスペクトル包絡パラメータを各サンプル値ごとに量子化せず，複数の値の列（ベクトル）をまとめて，1 つの符号で表現する方法をベクトル量子化（VQ：Vector Quantization）と呼ぶ[9]．この原理を図 6 に示す．あらかじめ一定時間の波形あるいはパラメータの組を多数用意して，これにクラスタリングの手法を適用し，代表的なパターン（セントロイド）の集合を生成する．代表パターンのそれぞれに符号を与え，この符号とパターン（コードベクトル，テンプレート）との対応を示す表，すなわち符号帳(codebook)を蓄える．音声の波形またはパラメータの組を，一定時間ごとに符号帳の各パターンと比較して，最も類似度の高いパターンを選び，その符号の列によって音声を符号化する．音声波形やパラメータの組には，連続性や分布の偏りなど特殊性があり，ベクトル量子化ではそれを利用するため，音声を少ない情報量で表現することができる．この方法は，情報速度・歪理論(Rate-distortion Theory）に基づく情報理論的な帯域圧縮の限界に漸近的に近づく方法として，近年脚光を浴びている．ベクト

表 1　波形符号化と分析合成方式の比較

項　目	波形符号化	分析合成
符号化情報	波形	スペクトル包絡情報 (短時間スペクトル) 音源情報 (ピッチ,振幅,有声/無声)
量子化ビット数 (kbps)	9.6～64 (中・広帯域)	2.4～4.8 (狭帯域)
対象とする音	任意音	単一話者の音声
問題点	量子化雑音の限界から符号化速度を下げるのが難しい.	周囲雑音による劣化 符号誤りによる劣化 音声品質の限界 処理が複雑
代表的方式	〔時間領域符号化〕 PCM, ADPCM, ⊿M, ADM, APC 〔周波数領域符号化〕 SBC, ATC 〔組合せ〕 APC-AB	チャンネルボコーダ ホルマントボコーダ 位相ボコーダ ケプストラムボコーダ LPC (最尤, PARCOR, LSP) ボコーダ

ル量子化を用いた音声分析合成系では，800 b/s 程度で了解性の高い音声が実現されている．

ベクトル量子化の一種であるが，それをさらに進めて，スペクトル包絡を表現する複数パラメータの時間パターンをセットとしてテンプレート（時空間パターン）化するマトリックス量子化，あるいはセグメント量子化が，200 b/s 程度の極低ビット音声符号化(分析合成系）を目的として，検討されている[10]．

6. 波形符号化と分析合成による音声符号化と音声合成の具体例

6.1 音声符号化の具体例

波形符号化と分析合成方式の比較を表 1 に示す[1]．分析合成方式では，原理的にその対象が単一話者の音声に限られるが，波形符号化では任意の音を対象とすることができる．ただし，一般に分析合成系よりも多くの情報量を必要とする．一般に音声符号化の種々の方式は，情報量(ビットレート)，符号化音声品質，装置の複雑さ，および符号化による遅延の 4 つの要因によって評価することができ，これらの要因の間にはトレードオフがある[11]．音声品質の中には，音声に雑音が重畳したり，伝送路に符号誤りが生じたときの品質の劣化の度合いなども含まれる．装置の複雑さは，それを実現するためのコストに密接な関係があり，一般に復号器よりも符号器の方がより複雑になる．音声を伝送せず単に蓄積する場合には，遅延を生じても問題はないが，伝送に用いる場合には，遅延があるとエコーを生じたり快適な通話を妨げる要因となる．

代表的な 6 種類の音声符号化方式の構成を図 7 に，これらの方式の4つの評価要因の概略を表2に示す[11]．MPC(Multipulse Linear Predictive Coding) は，LPC 分析合成系においてパルスと雑音による音源のモデル化を避け，有声音・無声音にかかわらず音源を複数のパルスによって表現して，これによって LPC 合成フィルタを駆動する方法である．CELP (Code-Excited Linear Predictive Coding)は，波形に含まれる音源ピッチの周期性に関する予測（長期予測）と，近接サンプル間の相関に関する予測（短期予測）を行って得られた予測残差波形を，ランダム雑音波形でベクトル量子化する方法である．MPC におけるマルチパルスを，ランダム雑音で置き換えたものと考えることもできる．

一般にビットレートが半分になると，同じ品質でも 1 桁複雑になる傾向がある．現在の LSI 技術をもってすれば，10 MIPS 程度の処理量であれば 1 チップ化することが可能であるが，それ以上になると複数のチップの組合せで実現することが必要である．ビットレートと音声品質の一般的関係を図 8 に示す[11]．8～16 kb/s の範囲においては，波形符号化と分析合成系の長所を組合せた方式が研究されており，雑音などによ

表 2　代表的な音声符号化方式の性能比較

符号化方式	用いられている符号化技術	情報量 (kb/s)	複雑さ (MIPS)	遅延 (ms)	音声品質
PCM (Pulse Code Modulation)	非線形量子化	64	0.01	0	高い
ADPCM (Adaptive Differential PCM)	適応量子化 予測量子化	32	0.1	0	高い
APC-BA (Adaptive Predictive Coding-Bit Allocation)	適応量子化 予測量子化 時間・帯域分割符号化	16	1	25	高い
MPC (Multipulse LPC)	分析合成における残差のマルチパルス化	8	10	35	中
CELP(Code-excited LPC)	分析合成における残差を雑音でベクトル量子化	4	100	35	中
LPC ボコーダ	分析合成	2	1	35	低い

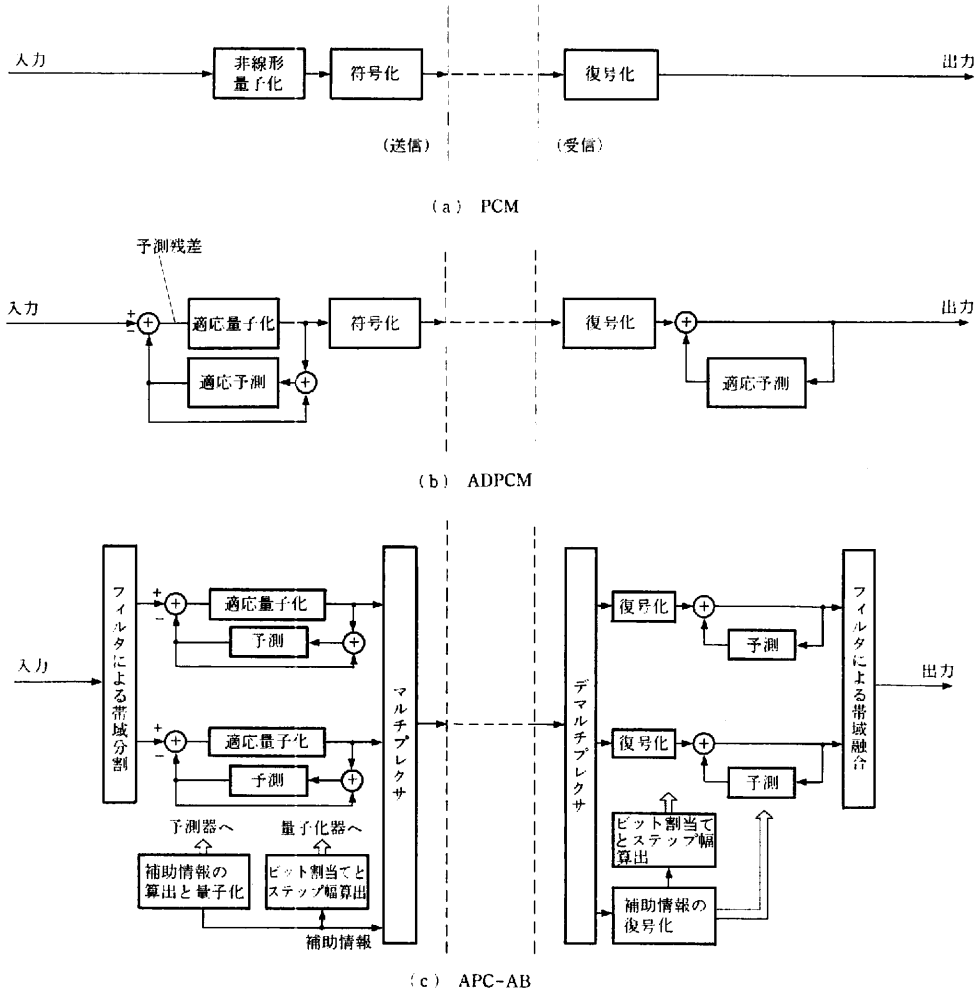


図 7 代表的な6種類の

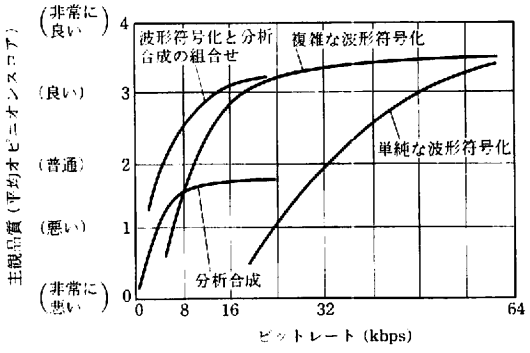


図 8 音声符号化におけるビットレートと符号化音声品質の関係

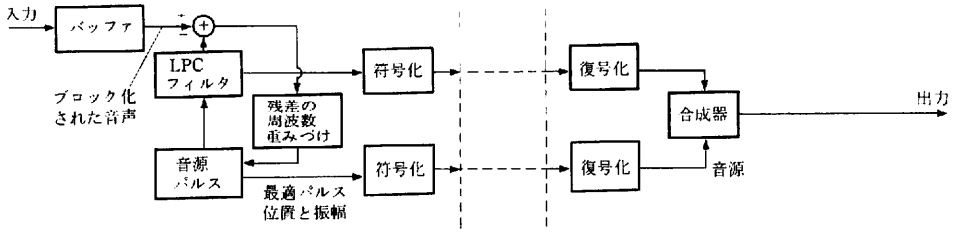
る品質劣化を避けることができる特徴がある．MPC や CELP はその例である．

なお従来の4 kHz電話帯域に代って，7 kHz帯域の信号を64 kb/sで高能率符号化し，自然性に優れた広帯域音声やAM放送規格の音楽信号を64 kb/sディジタル1回線で伝送するための研究も行われている．これが実現すれば，テレビ会議や拡声電話における臨場感のある高品質音声の伝送や，音楽などの放送に適用できると期待される[12]．

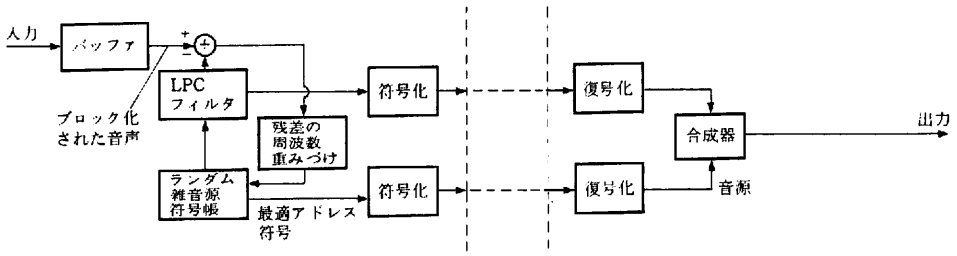
6.2 録音編集およびパラメータ編集方式による音声合成の具体例

符号化された波形あるいはパラメータを蓄える記憶装置の容量と，符号化・復号化部のハードウェア規模のトレードオフによって，種々の具体的方式が検討され，実用化されている．

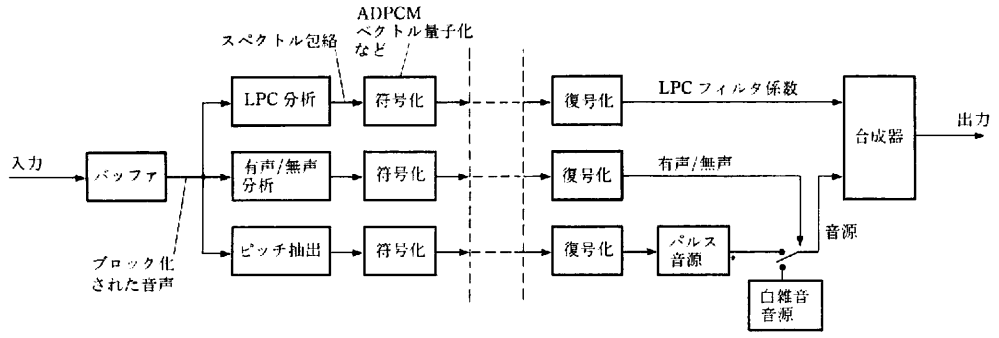
波形符号化（録音編集方式）と分析合成系（パラメータ編集方式）の技術的境界は連続的であるが，一般に後者の方がハードウェアが複雑になるかわりに記憶



(d) MPC



(c) CELP



(f) LPC ボコーダ

音声符号化方式の構成

容量が少なくてすむ．さらに分析合成系の方が，言葉と言葉の接続がなめらかになるようにパラメータを操作・調整したり，音声のピッチや発声速度を制御することも容易にでき，柔軟性のある音声合成が可能となる．語彙数が比較的少なく，定型の文章の組合せですむ場合には，波形符号化，あるいは単にアナログ録音による音声応答装置の方が経済的であるので，案内放送などにはこれが広く用いられている．

同じ言葉でも，文章中の位置によってイントネーションやアクセントが異なるので，波形符号化方式の場合は，1つの言葉に対して複数のタイプを蓄えておく必要がある．言葉と言葉の接続におけるポーズの入れ方も，自然性を出すには重要な要素のひとつである．

音声波形を蓄える記憶装置がランダムアクセスタイプのものであれば問題はないが，録音テープのようなものを用いると，安価ではあるが巻戻し，読出しに時間がかかるので用途が限られる．

7. テキスト合成の原理

規則合成方式を用いてテキストから音声を生成するシステムにおいて，合成すべき文字系列の他にアクセントやイントネーションに関する情報を外から与える場合と，システムが自動的にこれらの情報を生成する場合とがある．後者の場合を特にテキスト合成，あるいは文-音声変換と呼び，高度な言語解析機能が必要となるが，人間の文章朗読機能の工学的実現として近年多くの研究が行われている．

テキスト合成システムの構成例を図9に示す[13]．その具体的手順は，次の通りである．

(1) 書かれた文章の構文および意味解析（単語辞

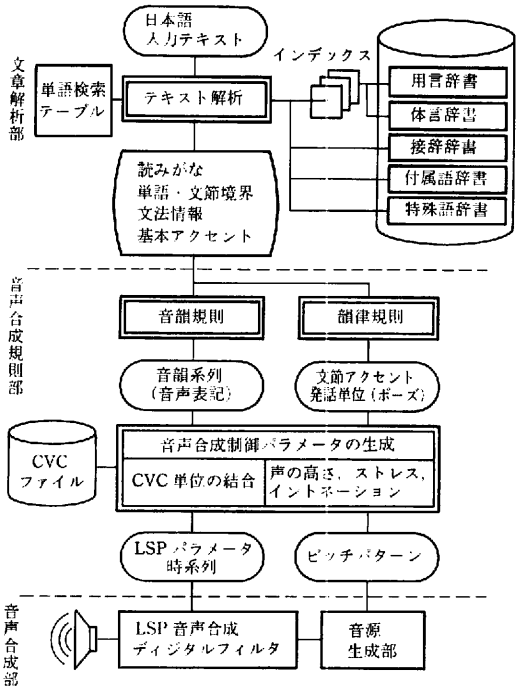


図 9　日本語テキストから音声を生成するテキスト合成系の構成

書の検索を含む).

(2)　各単語の発音記号，構文・意味情報および音韻規則による音韻記号への変換

(3)　各単語のアクセント位置，構文情報，韻律情報および韻律規則による継続時間長，アクセント，イントネーション，ポーズなどの決定

(4)　音声合成器の制御パラメータへの変換（合成基本単位の結合，ピッチパターンの生成など）

(5)　音声合成器による音声の生成

日本語は英語のような言語と違って，単語がつながって書かれるので，(1)では，単語に関する知識（辞書）に基づいて，単語境界を検出することが重要なステップとなる．この際，語の連続関係や意味を考慮しないと正解が得られないことも多い．(3)でも，語の係受け関係などを考えてアクセント的なまとまりを決め，文の構造上の切れ目の適当な位置にポーズを入れることが重要である．

(2)の音韻規則とは，「は」を/wa/と読む規則をはじめ，例えばasuka(飛鳥)の/u/のように，母音の/i/や/u/が無声子音に挟まると無声音になる「母音の無声化」や，単語が連なると読み方やアクセントが変化する現象（連濁など）の変形規則などであり，自然な合成音を生成するうえで重要な役割を果たす．

(4)においてピッチパターンを記述する方法としては，イントネーションに対応するフレーズ成分と，アクセントに対応するアクセント成分の和で表すモデルが広く用いられている[14]．

音声を生成する基本的単位としては，CV（子音Consonant-母音Vowel），VCV，CVCなどの音節が用いられ，これらの結合によって任意の音声を生成する．たとえば「規則」という単語をこれらの単位に分解すると，CV音節：ki+so+ku，VCV音節：ki+iso+oku，CVC音節：kis+sok+kuとなる．VCV音節を用いる場合は母音部で，CVC音節を用いる場合は子音部で基本単位が接続される．

音声の最も小さな基本単位は音素で，30～50種類であるが，この結合規則は極めて複雑になるので，通常これより大きい単位が用いられる．比較的大きい単位を用いる方が，基本単位の数は多くなるが，CまたはVに関して前後の音の影響（調音結合）を考慮した音節が与えられるので，比較的なめらかな品質の高い合成音を作ることができる．

日本語の中に出現するCV音節は約100種類，VCV音節は700～800種類，CVC音節は5000～6000種類である．CV音節とCVC音節を併用することにより，基本単位の数を1000種題程度に抑えて品質の高い合成音を生成する試みも行われている．英語では音節の種類が非常に多いので，音節を半分に分割した単位がよく用いられる．

テキスト合成で用いられている音声合成器のほとんどは，線形分離等価回路モデルに基づいており，スペクトル包絡（調音）情報をホルマント（共振）とアンチホルマント（反共振）回路の縦続または並列接続で与えるホルマント合成器，ケプストラム分析合成系や，線形予測分析に基づくPARCOR，LSPなどの分析合成系がよく用いられる．

8.　テキスト合成の具体例

テキスト合成を用いると，文字情報に基づいてどのような音声でも作り出すことができるので，電話回線を用いた銀行の振込通知サービスなどにおける振込人名義など，不特定な語の出力にすでに応用されている．また最近では，新聞や雑誌の記事の作成過程において，記事を合成音で自動的に読み上げることにより，校閲部門の読み合わせ作業の省力化と能率化をはかることが試みられている．

将来は，ニュースのテレホンサービスなどへの応用も考えられる．さらに，テキスト合成装置を文字読み取り装置と組合せれば，目が見えない人でも自由に本が読めるようになると考えられ，実用化が期待されて

いる．

9. むすび

最近の合成音の品質は，以前の機械音から徐々に自然性を増してきているが，文字から任意の音声が生成できるテキスト合成や，極めて少ない情報量による符号化音声の品質はまだ不充分である．文章の朗読は人間がなにげなくやっていることではあるが，実際には文章の意味や文脈を考慮した高度な機能であり，現在のテキスト合成における言語解析の能力は，人間の能力にはほど遠いものである．初めて見た言葉でも“常識”で読むことができる機能までを機械に与えるには，人工知能的研究の高度な進歩が必要である．

合成音の品質は，基本単位の品質とともに制御規則(制御情報と制御機構)によって大きく変わる．その声質は蓄えられている基本単位に完全に依存しており，1つのシステムで声質が変えられるものはない(男性，女性，子供のいずれかの声を選択できるものはある)．今後，用途や好みに応じて声質が自由に変えられる必要性が出てくると思われるが，このためには音声の個人性とは何かという基本的課題からのアプローチが必要である．

音声波形およびそのスペクトルは常に時間的に変化しており，その動特性が人間による音声知覚に重要な役割を果たしていることが確認されている．しかし，このような過渡的信号の性質を正しく安定に抽出し，再現できる音声符号化および音声合成技術はまだ確立されていない．人間が音声を発声するときの発声器官の観測が重要な手がかりを与えることが期待される．今後の重要な課題のひとつであるといえる．

音声の言語的性質を用いて，ディジタル音声分析のレベルと知識処理をいかにうまく融合させるかも，今後の重要な課題である．特に100～200 b/sあるいはそれ以下の極低ビットで音声を表現するためには，音韻的・言語的情報を用いることが必要であり，このような研究が不可欠であると思われる．

究極の音声伝送や音声蓄積の形態は，音声波に含まれる言語的情報などを記号化して伝送・蓄積する，認識伝送・蓄積になると考えられる．このためには音声符号化技術と音声認識技術との融合が必要である．

(昭和62年3月10日受付)

〔参考文献〕

1) 古井貞熙：“ディジタル音声処理”，東海大学出版会 (1985)
2) T. Moriya and M. Honda：“Speech Coder Using Phase Equalization and Vector Quantization”, Proc. Int. Conf. Acoust., Speech, Signal Processing, Tokyo, 33.6 (1986)
3) J. D. Markel and A. H. Gray, Jr.：“Linear Prediction of Speech”, Springer-Verlag, New York (1976)
4) B. S. Atal and S. L. Hanauer：“Speech Analysis and Synthesis by Linear Prediction of the Speech Wave”, J. Acoust. Soc. Amer., **50**, 2 (pt. 2), pp. 637-655 (1971)
5) 板倉，斎藤：“統計的手法による音声スペクトル密度とホルマント周波数の推定”，信学論誌，**53-A**, 1, pp. 35-42 (1970)
6) 板倉文忠：“新しい音声分析合成方式PARCOR”，日経エレクトロニクス，2.12, pp. 58-75 (1973)
7) 板倉，嵯峨山：“線スペクトル周波数をパラメータとした音声合成法とそのLSI化”，日経エレクトロニクス，2.2, pp. 128-159 (1981)
8) 広川，鈴木：“音声合成と楽音合成技術”，テレビ誌，**40**, 8, pp. 729-734 (1986)
9) A. Gersho and V. Cuperman：“Vector Quantization：A Pattern-matching Technique for Speech Coding”, IEEE Commun. Magazine, pp. 15-21 (Dec. 1983)
10) 日本，曽田：“音声の時空間パターン符号化の最適構成法”，音響学会音声研資，S 85-45 (1985)
11) N. Jayant：“Coding Speech at Low bit Rates”, IEEE Spectrum, pp. 58-63 (Aug. 1986)
12) 高正博：“通信における符号化”，テレビ誌，**40**, 8, pp. 709-711 (1986)
13) 佐藤大和：“音声合成”，数理科学，252(6月号)，pp. 42-48 (1984)
14) 藤崎博也：“音声の抑揚とそのメカニズム”，数理科学，252(6月号)，pp. 26-32 (1984)

古井 貞熙（ふるい さだおき）昭和45年，東京大学工学部計数工学科大学院修士課程修了．同年，NTT電気通信研究所入社．以後，同研究所において音声認識，話者認識，音声知覚などの基礎研究に従事．現在NTT基礎研究所情報通信基礎研究部第4研究室室長．工学博士．