論文 / 著書情報
Article / Book Information

論題(和文)	音声研究とディジタル信号処理
Title(English)	
著者(和文)	古井 貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	電子通信学会 ディジタル信号処理第二種研究会資料 DSP85-2,, , ,
Citation(English)	, , ,
発行日 / Pub. date	1986, 1
URL	http://search.ieice.org/
権利情報 / Copyright	本著作物の著作権は電子情報通信学会に帰属します。 Copyright (c) 1986 Institute of Electronics, Information and Communication Engineers.

ディジタル信号処理第二種研究会資料
資料番号 DSP85-2(1986-1)

音声研究とディジタル信号処理

古井　貞熙
（NTT基礎研究所情報通信基礎研究部）

1986年1月16日

社団法人 電子通信学会

電子通信学会ディジタル信号処理研究会資料
Transactions of the Committee on Digital Signal Processing, IECEJ
DSP85-2(January 16, 1986)

音声研究とディジタル信号処理
REVIEW AND PERSPECTIVE OF DIGITAL SIGNAL PROCESSING TECHNOLOGY RELATED TO SPEECH RESEARCH

古井 貞熙
Sadaoki FURUI

NTT 基礎研究所
NTT Research Laboratories

1．まえがき

音声研究とディジタル信号処理技術には、相互に密接な関連があり、音声分析、音声符号化、音声合成、音声認識等に代表される近年の音声情報処理の進展は、ディジタル信号処理技術の進歩に負う面が強いと同時に、ディジタル信号処理技術も、音声をその対象とすることによって進展して来たと言える[1]。FFT(Fast Fourier Transform)、線形予測分析、ケプストラム分析等はその代表例と言える。ディジタル信号処理技術には、極めて技巧を凝らした、アナログ処理では実現できないような信号処理技術が実現でき、信頼性が高いという特徴がある。この進歩には、電子計算機やLSIの発達も大きな力となっている。

ここでは音声情報処理において用いられている比較的新しいディジタル信号処理技術についてまとめて紹介してみたい。音声通信システムを構成する物理的な基本機能には、増幅・減衰・ろ波・等化・変復調・多重化・同期・交換などがあり、これらにも種々のディジタル信号処理技術が用いられているが、他の機会に紹介されると思われるので、ここでは省略した。またディジタル信号処理用LSIに関しても、他の機会にゆずることとした。音声波に重畳した雑音や反響除去の問題に関しても、重要な分野であるが頁数の都合で割愛させて頂いた。

2．音声生成のディジタルモデル

音声生成のプロセスは、音源の生成、声道の形による調音、口唇又は鼻孔からの放射の3作用に分離して考えることができる。さらに、電気・音響系の対応関係により、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を有するパルス又は三角波。
(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：6dB/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。

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

通常、図1に示すように、音源はパルス音源と雑音源で近似し、調音は全極型又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。従って$G(\omega)$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、スペクトルとして表現することが多い。声道による音声生成の物理的メカニズム、言語的制約、聴覚のメカニズム等により、音声信号には大きな冗長性がある。このため音声波に含まれる種々の特徴は、ディジタル信号処理の手法により、少ない情報量で表現することができる。

3．音声分析

3．1．スペクトル分析

音声生成のディジタルモデルによれば、音声の短時間スペクトルは、調音に対応し周波数とともにゆるやかに変化するスペクトル包絡

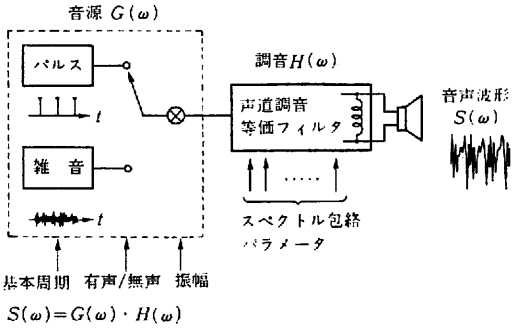


図1．音声生成機構の線形分離等価回路モデル

と、音源に対応し周波数とともに細かく周期的（有声音の場合）又は非周期的（無声音の場合）に変化する成分．即ちスペクトル微細構造の積（対数尺度では和）に分解して考えることができる。図2に、男性の／a／の場合の例を示す。

多くの統計的信号解析と同様に、音声のスペクトル包絡の抽出法は、ノンパラメトリック分析法（ＮＰＡ）とパラメトリック分析法（ＰＡ）に大別することができる。ＰＡは対象の信号に対してモデル化を行って、そのモデルを表現する特徴パラメータを抽出するものであり、ＮＰＡはこのような信号の性質に対応したモデル化を行わず、種々の信号に共通に適用できる分析法である。対象によくあったモデル化を行えば、ＰＡの方が能率的な分析や特徴表現を行うことができる。音声スペクトルの主な分析法と特徴を表１に示す[2]。

３．２．デイジタルサウンドスペクトログラム

図3に、男性が「朝鮮南部に」と発声したときの音声波形．短時間平均パワー、短時間スペクトル変化量[3]、基本周波数．及びサウンドスペクトログラムを示す。この図では、スペクトル変化量に基づいて自動的な音素セグメンテーションが行われている。サウンドスペ

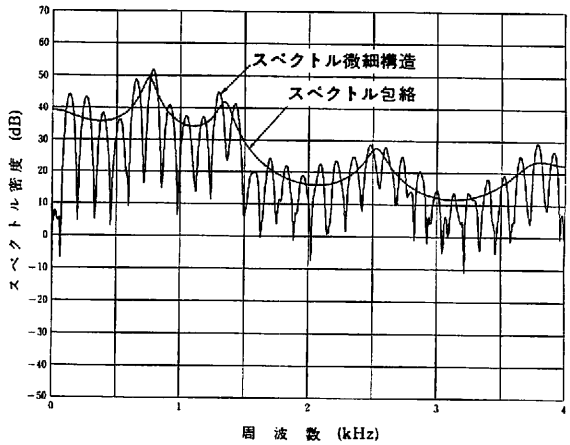


図２．男性の母音/a/のスペクトル包絡とスペクトル微細構造

表１．音声スペクトルの主な分析法とその特徴

分類	分析法	パラメータ	特徴
ノンパラメトリック分析 (NPA)	（ｉ）短時間自己相関分析	$\phi(m)$	スペクトル包絡と微細構造がたたみ込みの形で混在する．計算アルゴリズムが単純でハードウェア化が容易．
	（ii）短時間スペクトル分析	$S(\omega)$	スペクトル包絡と微細構造が積の形で混在．FFT によれば高速処理可能．
	（iii）ケプストラム分析	$c(\tau)$	スペクトル包絡と微細構造がケフレンシー領域において近似的に分離．２回の FFT と1回の対数変換必要．
	（iv）帯域フィルタバンク分析	フィルタ出力の RMS 値	スペクトル包絡の概形が求まる．アナログ実時間処理に適す．
	（v）零交叉数分析	零交叉数	(iv) との組合せによりホルマント周波数も近似的に求まる．ハードウェアが簡単．
パラメトリック分析 (PA)	（ｉ）合成による分析 (analysis-by-synthesis)	ホルマント，バンド幅など	モデルの精密化が可能．ホルマントの分析精度が高い．複雑な逐次近似が必要なため，処理時間大．
	（ii）線形予測分析 (LPC)		モデルは全極型スペクトルで単純．モデルの次数に等しい遅れの自己相関（共分散）から逐次近似によらず推定可能．
	（a）最尤スペクトル推定（自己相関法／逆フィルタ法）	α_i	合成フィルタの安定性の保証あり，時間窓が必要．計算量$\propto p^2$
	（b）共分散法	α_i k_i	合成フィルタの安定性の保証なし，安定化処理が必要．時間窓が不要なため，短時間分析に適す．計算量$\propto p^3$
	（c）PARCOR 分析	k_i	正規方程式の求解が格子形フィルタに埋めこまれる．積分特性の変更により，（a），（b）のいずれとも等価可能．計算量$\propto p^2$
	（d）LSP 分析	ω_i	量子化特性，補間特性ともに良好．ホルマントと類似の動きをする．計算量は PARCOR 分析よりやや多い．

ＦＦＴ：高速フーリエ変換
p：線形予測分析のモデルの次数

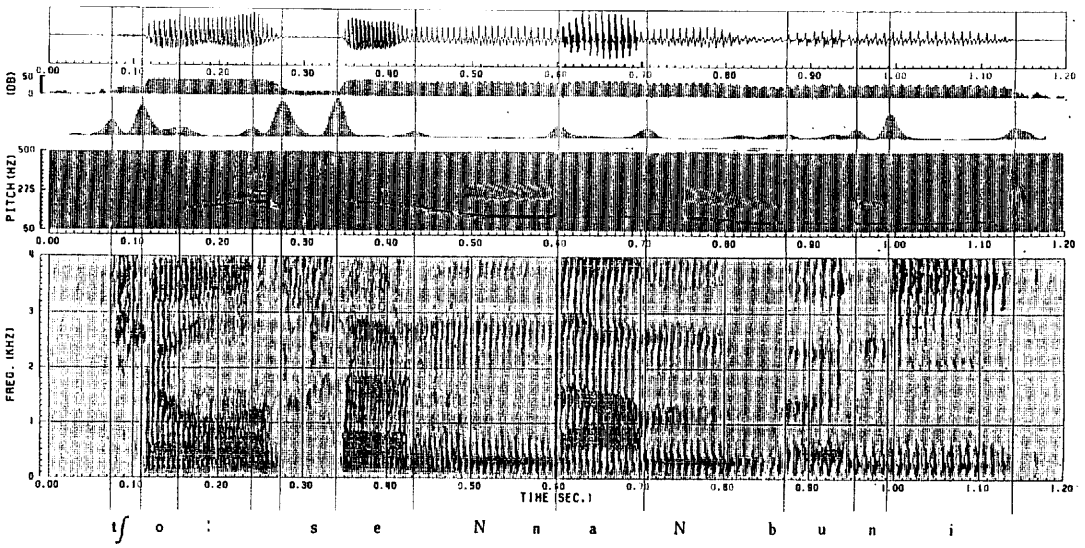


図3．音声波形．短時間平均パワー、短時間スペクトル変化量、基本周波数、及びサウンドスペクトログラム

クトログラムは、従来、帯域フィルタを用いたサウンドスペクトログラフによってアナログ的に分析表示されていたが、近年では、この図のようにＦＦＴを用いてデイジタル処理によって分析することが多い。

3．3．ケプストラム分析

ケプストラムは、短時間振幅スペクトルの対数の逆フーリエ変換として定義され、スペクトル包絡と微細構造を分離して抽出できる特徴がある。線形分離等価回路モデルを示す式(1)より、ケプストラム $c(\tau)$は、フーリエ変換を $\mathcal{F}$ の記号で表わすと、次のように計算される。

$$c(\tau)=\mathcal{F}^{-1}\log|S(\omega)|$$

$$=\mathcal{F}^{-1}\log|G(\omega)|+\mathcal{F}^{-1}\log|H(\omega)| \qquad (2)$$

τはケフレンシーと呼び、単位は時間である。

上式の右辺第１項はスペクトル上の微細構造であるので、高ケフレンシー部のピークとなり、第２項は周波数による変化のゆるやかなパターンであるので、低ケフレンシー部の特徴として現われる。このためケフレンシーに対するフィルター（リフター）により、両者を分離することができる。図4に、ケプストラム分析の手順を示す。ケプストラム分析は、一般に畳み込みの関係にあるものを和の形に変換し分離する処理である準同型分析（ホモモルフィック分析、準同型フィルタリング）の一種である。

ケプストラムの特殊なものとして、後述する線形予測分析によって推定された全極型音声生成システム関数

$$H(z)=1/\left(1+\sum_{i=1}^{P}\alpha_i\, z^{-i}\right) \qquad (3)$$

を音声信号のスペクトル密度とみなしたときのケプストラムを、ＬＰ

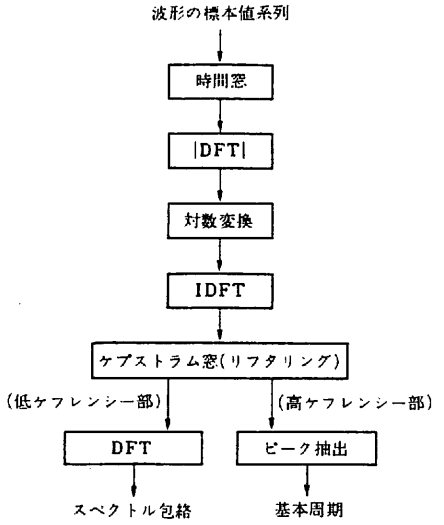


図4．ケプストラム分析の手順

Ｃケプストラムと呼ぶ。ＬＰＣケプストラムは、線形予測係数α_iに関する次のような再帰式によって得ることができる[4]。

$$\hat{c}_1=-\alpha_1$$

$$\hat{c}_n=-\alpha_n-\sum_{m=1}^{n-1}\left(1-\frac{m}{n}\right)\alpha_m\,\hat{c}_{n-m}\quad(1<n\leqq p)$$

$$\hat{c}_n=-\sum_{m=1}^{P}\left(1-\frac{m}{n}\right)\alpha_m\,\hat{c}_{n-m}\qquad(p<n) \qquad (4)$$

ＬＰＣケプストラムをフーリエ変換したときに得られるスペクトル包絡は、原型の（ＦＦＴ）ケプストラムによるものと類似しているが、

スペクトルのピークをより重視した形になっており、（ＦＦＴ）ケプストラムよりも少ない処理で求めることができる長所がある。

３．４．線形予測分析

ＰＡの代表例に線形予測分析があり、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[5][6]。

$$x_t + \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} = \varepsilon_t \qquad (5)$$

$\{\varepsilon_t\}$ 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、そのシステム関数が極のみを有することから全極モデルとも呼ばれる。

この方法は、音声波形のスペクトル密度 $f(\lambda)$ が全極型の有理スペクトル密度で表わされ、次の関数形を持つと仮定することと等しい[7]。

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{1}{\left| \prod_{i=1}^{p} \left(1 - \frac{z}{z_i}\right) \right|^2}$$

$$= \frac{\sigma^2}{2\pi} \frac{1}{\left| 1 + \sum_{i=1}^{p} \alpha_i z^{-i} \right|^2}$$

$$= \frac{\sigma^2}{2\pi} \frac{1}{A_0 + 2\sum_{i=1}^{p} A_i \cos i\lambda} \qquad (6)$$

ただし、z_i は $1 + \sum_{i=1}^{p} \alpha_i z^{-i} = 0$ の根で、

$$A_i = \sum_{j=0}^{p-|i|} \alpha_j \alpha_{j+|i|}, \quad \alpha_0 = 1, \quad (i = 0, \pm 1, \ldots, \pm p) \qquad (7)$$

である。このときパラメータ（$\sigma^2, \alpha_1, \alpha_2, \ldots, \alpha_p$）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。

線形予測係数（α_i）の解法としては、共分散法と自己相関法の２つの方法がある。最尤スペクトル推定法の解法は、自己相関法と等しく、この両者は、同一の線形システムを一方は時間領域で、他方は周波数領域で考察したものになっている。共分散法、自己相関法のいずれも線形連立一次方程式を解くものであり、共分散法の方程式はCholesky分解によって、相関法の方程式はDurbinの再帰的解法によって効率的に解くことができる。相関法の場合は解の安定性が保証されるが、共分散法の場合は保証されない。

Durbinの再帰的解法の過程でPARCOR係数（偏自己相関係数）が抽出される。PARCOR係数は線形予測係数よりも、解の安定性の保証が容易に得られ、より少ない情報量でスペクトル包絡が表現できる等の長所がある[8]。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よ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）が開発された[9]。線形予測分析による種々のスペクトルパラメータの相互関係を図５に示す[10]。

４．音声分析合成系

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）することを音声の分析合成と呼ぶ。

- 有声音／無声音の区別
- 基本周期
- 音源振幅

　・・・音源情報

- 線形フィルタ特性・・・・・・スペクトル包絡（調音）情報

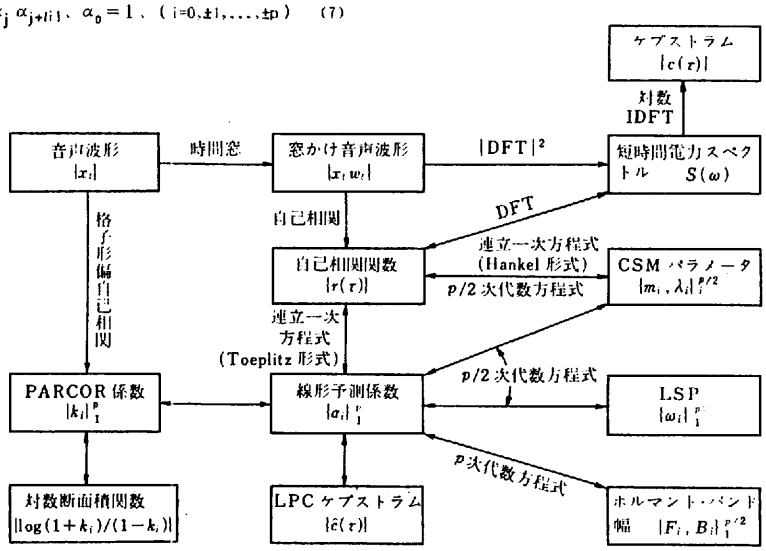


図５．線形予測分析による種々のスペクトルパラメータの相互関係

これまでの分析合成系の主な例を表2に示す[9]。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線形予測分析によるスペクトル包絡で正規化（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，即ち予測残差信号の相関関数（変形相関関数）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分析合成系の構成を図6に示す。

5．音声波形の符号化

5．1．波形符号化の原理

音声に含まれる音源と調音に関する情報を分離せず，波形自体を出来るだけ少ない情報量で忠実に表現しようとする方法を波形符号化と呼ぶ。波形符号化で用いられる情報圧縮技術とその背景としての音声信号と聴覚の性質を図7に示す[11]。分析合成では対象が単一話者の音声に限られたのに対し，波形符号化では任意の音を対象とすることができる特徴があるが，一般に分析合成よりも多くの情報量を必要とする。量子化ビット数と音声品質の関係を図8に示す[12]。ノイズシェイピングとは，図9に示すように，音声の周波数スペクトルよりもある程度低いしきい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，雑音を全体としてこのしきい値以下におさまるように調整することをいう[12]。4．8～9．6

表2．分析合成系の主な例

種類	スペクトル分析	特徴パラメータ	伝送容量
チャネル ボコーダ	帯域フィルタバンク分析	帯域フィルタの出力振幅	300Hz(アナログ伝送) 2400bps (ディジタル伝送)
ホルマント ボコーダ	帯域フィルタバンク分析，零交叉分析	帯域フィルタの出力振幅，零交叉数	300Hz 2400bps
パターンマッチングボコーダ	帯域フィルタバンク分析	音韻のスペクトルパターン	900bps
相関ボコーダ	短時間自己相関分析	自己相関関数 $\phi(m)$	400Hz
位相ボコーダ	帯域フィルタバンク分析	帯域フィルタの出力振幅，帯域信号の位相	1500Hz 7200～9600bps
最尤ボコーダ	最尤推定法	線形予測係数 α_i	5400bps
ホモモルフィックボコーダ	ケプストラム分析	ケプストラム $c(\tau)$	7800bps
PARCOR ボコーダ	偏自己相関分析	PARCOR 係数 k_i	2400～9600bps
線形予測 ボコーダ	共分散法	線形予測係数 α_i	3600bps
LSP ボコーダ	線スペクトル分析	LSP（線スペクトル対）ω_i	1600～4800bps

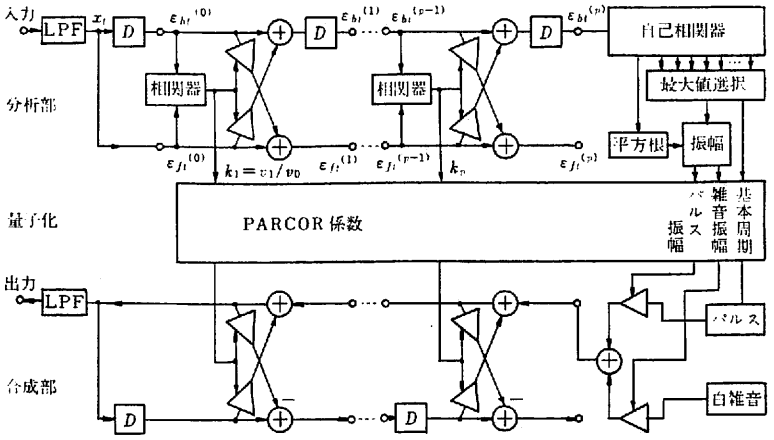


図6．PARCOR分析合成系の構成

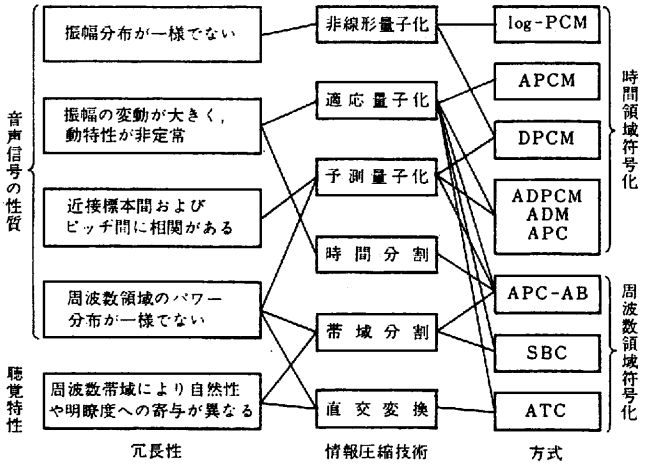


図7．波形符号化で用いられる情報圧縮技術とその背景としての音声信号と聴覚の性質

kbpsの範囲については、波形符号化と分析合成の長所を組み合わせた残差駆動線形予測ボコーダなどの方式が研究されている。

5．2．非線形量子化と適応量子化

音声振幅の統計的性質を用いて、振幅を非線形変換（通常対数変換）して圧縮し、その結果を線形に符号化する方法を非線形量子化と呼ぶ（log－PCM）。

さらに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のrms値に応じて時間的に変化させる方法を適応量子化と呼ぶ。（APCM） 実際にステップ幅を変化させる方法には、ブロック毎に行う前向き（フィードフォワード）適応方式と、直前の信号の符号化結果に応じて適応的に行う後向き（フィードバック）適応方式がある。

5．3．予測符号化

音声信号は隣接標本間のみならず、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（差分量子化（DPCM））、あるいはその相関を利用して直前の一定区間の時系列から予測した値と実際の標本値の差信号（予測残差）の符号化により情報圧縮することができ、これらの方法を予測符号化という。比較的近傍の標本値（8kHz標本化で4～20サンプル）を用いた予測をスペクトル包絡予測（短時間予測）、音声信号のピッチ間の予測をピッチ予測（長時間予測）と呼ぶ。

差分量子化の考えを極限にまで推し進め、1ビット量子化で近似できるまでに標本化周波数を高くした方式がデルタ変調（DM、ΔM）方式である。量子化のステップ幅を適応的に変化させる方式を適応デルタ変調（ADM）と呼ぶ。

適応予測を行うDPCMは、適応量子化を行うDPCM及びこの両者を行うDPCMと合わせて、ADPCMと呼ばれる。適応量子化と同様、適応予測にも前向き（フォワード型）と後向き（バックワード型）がある。前向き適応予測は音声信号に対してブロック毎に予測を行う方式で、前向き適応予測にピッチ予測を組み合わせた方式を適応予測符号化（APC）と呼ぶ。このブロック図を図10に示す[13]。

5．4．帯域分割符号化と適応変換符号化

音声を複数の周波数帯域に分割して符号化する方式を、帯域分割符号化（SBC）と呼ぶ。この方式は、人間の聴覚特性に対応してノ

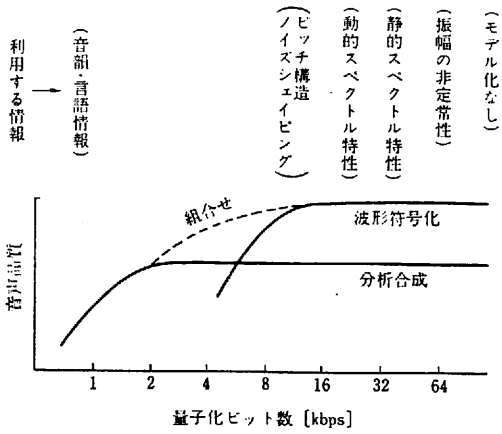


図8．波形符号化と分析合成系の量子化ビット数と音声品質の関係

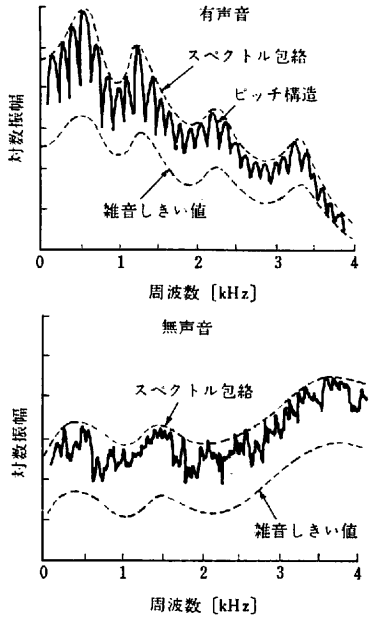
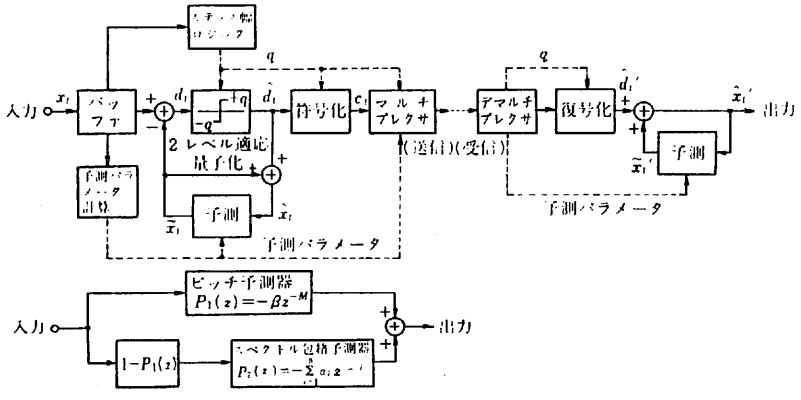


図9．スペクトル包絡と聴覚的マスキングによる雑音しきい値の関係

図10．適応予測符号化（APC）の原理



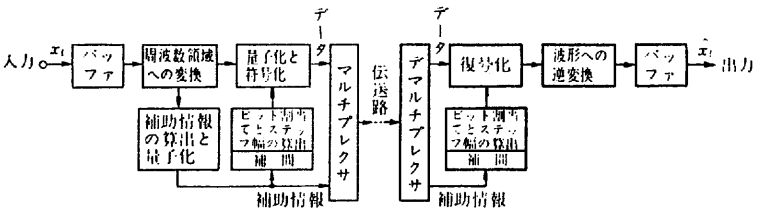


図11. 適応変換符号化（ＡＴＣ）の原理

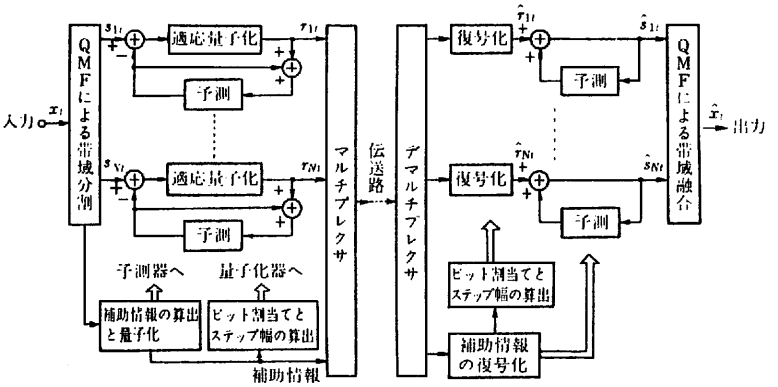


図12. 適応ビット割り当てＡＰＣ（ＡＰＣ－ＡＢ）の原理

イズシェイピングなどを考慮することが容易にでき、重要な帯域に多くの情報を割り当てることができる特徴がある。帯域通過フィルタ群は、帯域分割を２：１に限定すれば、ＱＭＦ（Quadrature Mirror Filter）[14]を用いることができ、処理が簡単でかつ折り返し歪が消える長所がある。

音声がほぼ定常と考えられる２０ms程度を１ブロックとして、ＤＣＴ等で周波数領域に直交変換した量を符号化する方式を適応変換符号化（ＡＴＣ）と呼ぶ。この方式のブロック図を図１１に示す[15]。

５．５．適応ビット割り当てＡＰＣ

適応ビット割り当てＡＰＣ（ＡＰＣ－ＡＢ）は、ＳＢＣとＡＰＣを組み合わせて情報圧縮をはかった方式で、基本的構成は図１２に示す通りである[16]。音声信号を帯域分割し、各帯域信号をベースバンド（低域）に変換したのち、線形予測分析を行って予測符号化（ＡＰＣ）する。残差波形に対して、残差パワーをもとに帯域ごとにビット割り当てを行うとともに、１ピッチ区間をさらに時間分割して、その区間内の残差パワーに応じて動的適応的にビット割り当てを行う。

５．６．マルチパルス駆動線形予測符号化

線形予測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、これによってＬＰＣ合成フィルタを駆動する方法をマルチパルス駆動線形予測符号化（ＭＰＣ）と呼ぶ[17]。複数のパルスの振幅と時間位置の最適な設定は、図１３に示すように合成による分析（Ａ－ｂ－Ｓ）の手法を用いた繰り返し演算によって行っている。

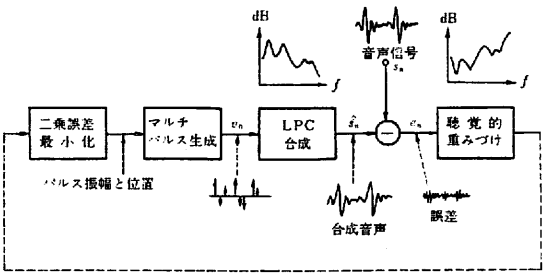


図13. A-b-S法によるマルチパルスの振幅と位置の決定アルゴリズム

５．７．位相等化処理

人間の聴覚は短時間位相に対して冗長性があるので、予測残差波形を位相等化（零位相化）することにより、波形の電力を時間的に集中させ、情報割り当てによって符号化効率を高めることができる。この位相等化処理を、マッチドフィルタの原理を用いて時間領域で行う方法が提案され、木符号（入力音声信号と符号化出力との誤差が最小となるように、最適な駆動音源信号系列を木構造を用いて探索し符号化する方式）と組み合わせる方式の有効性が確認されている[18]。

５．８．ベクトル量子化

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）と呼ぶ[19]。この原理を図１４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に関して、分布の偏りを考慮したクラスタリングの手法により、全体としての歪が最も小さくなるように、代表的なパタ

ーンを生成して蓄えておき、それぞれに符号を与えておく。この符号とパターン（コードベクトル、テンプレート）との対応を示す表を符号帳（codebook）と呼ぶ。符号化すべき入力の波形又はパラメータの組に関しては、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、その符号を伝送する。ベクトル量子化が高能率符号化になりうるのは、音声波形やパラメータの組に、連続性や偏りなど特殊性があって、それを利用するからである。

ベクトル量子化の一種であるが、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が、200bps程度の極低ビット音声符号化方式を目的として、検討されている[20]。

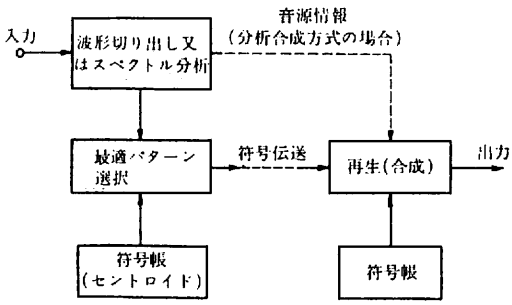


図14．ベクトル量子化の原理

6．音声合成

音声合成の方式は次の3種類に分類できる。

(1) 録音編集方式：人が発声した音声波を蓄積しておき、必要に応じてつなぎあわせて出力する。

(2) パラメータ編集方式：人が発声した音声波を分析してパラメータの系列で蓄積しておき、それをつなぎあわせて音声合成器を駆動し出力する。

(3) 規則合成方式：文字列あるいは音素記号列から、音声学的、言語学的規則に基づいて音声合成器を駆動し、音声を作り出す。

録音編集方式とパラメータ編集方式は、基本的に人が発声した音声のつなぎあわせであるので、出力できる語彙が蓄積した単位の組み合わせに限られるが、規則合成方式の場合は任意の言葉が出力できる。それだけ高度な制御規則が必要となるとも言える[21]。規則合成方式を用いて、日本語のテキストから音声を生成するシステムの構成を図15に示す[22]。

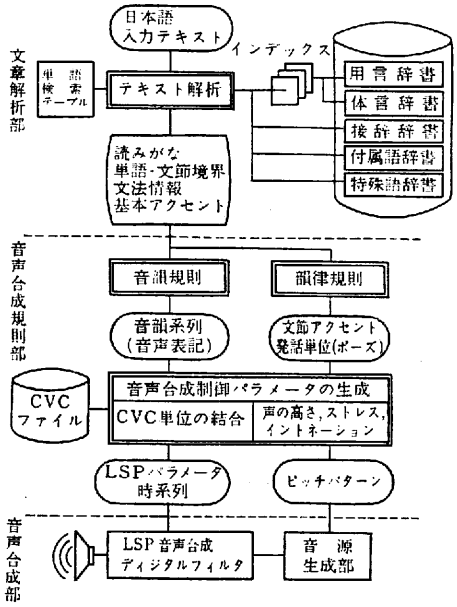


図15．CVC単位とLSP音声合成方式を用いた日本語テキスト音声合成システムの構成

現在規則合成方式又はパラメータ編集方式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もとづいており、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）回路の縦続又は並列接続で与えるホルマント合成器（ターミナルアナログ合成器）、又は線形予測分析に基づく分析合成系がよく用いられる。

ホルマント合成器における共振（極）回路は、次の形で与えられる。

$$H(z)=\sum_{k=1}^{2}\frac{1}{1-e^{T\theta_k}z^{-1}}\,\mathrm{Res}[H(\theta)]_{\theta=\theta_k}$$

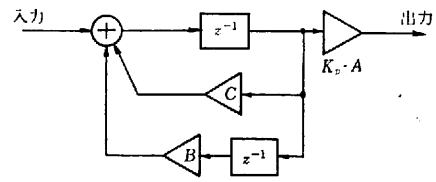
$$=K_p\left(\frac{Az^{-1}}{1+Cz^{-1}+Bz^{-2}}\right) \qquad (8)$$

ここで

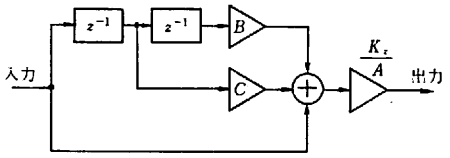
$$K_p=(\sigma_n^2+\omega_n^2)/\omega_n,\quad A=e^{-\sigma_n T}\sin\omega_n T$$

$$B=e^{-2\sigma_n T},\quad C=-2e^{-\sigma_n T}\cos\omega_n T$$

Tは信号の標本化周期、Res[・]は留数を示す。具体的な回路は図16に示すようになる。共振周波数$f_n=\omega_n/2\pi$[Hz]と帯域幅$b_n=\sigma_n/\pi$[Hz]を与えれば、回路定数を決めることができる。反共振（零）回路は、逆回路の関係を用いることによって、共振回路から容易に導かれる。図中で、$K_z=\omega_n/(\sigma_n^2+\omega_n^2)$である。



（a）共振（極）回路



（b）反共振（零）回路

図16．共振および反共振回路のデイジタルシミュレーション

7．音声認識

7．1．音声認識の原理

音声認識の形態は次のように分類できる。

- 孤立単語音声認識
- 連続音声認識
  - 連続単語音声認識
  - 音声理解（文音声認識、会話音声認識）

表3．線形予測分析に基づくスペクトル距離尺度

距離尺度	計算式
最尤スペクトル距離 (Itakura-Saito 距離)	$E=\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\log\frac{f(\lambda)}{g(\lambda)}+\frac{g(\lambda)}{f(\lambda)}-1\right)d\lambda$ $=\log\frac{u^{(f)}R^{(f)}}{u^{(g)}R^{(g)}}+\frac{u^{(g)}}{u^{(f)}R^{(f)}}\sum_{k=-p}^{p}A_k^{(f)}r_k^{(g)}-1$
対数尤度比距離	$H=\log\frac{\sum_{k=-p}^{p}A_k^{(f)}r_k^{(g)}}{R^{(g)}}$
予測残差	$N=\sum_{k=-p}^{p}A_k^{(f)}r_k^{(g)}$
cosh 尺度	$D=\frac{1}{2\pi}\int_{-\pi}^{\pi}2\{\cosh(\log f(\lambda)-\log g(\lambda))-1\}d\lambda$ $=\frac{u^{(g)}}{u^{(f)}R^{(f)}}\sum_{k=-p}^{p}A_k^{(f)}r_k^{(g)}+\frac{u^{(f)}}{u^{(g)}R^{(g)}}\sum_{k=-p}^{p}A_k^{(g)}r_k^{(f)}-2$
LPC ケプストラム距離	$L^2=\sum_{n=-\infty}^{\infty}(c_n^{(f)}-c_n^{(g)})^2$ $=\left(\log\frac{u^{(f)}R^{(f)}}{u^{(g)}R^{(g)}}\right)^2+2\sum_{n=1}^{\infty}(c_n^{(f)}-c_n^{(g)})^2$
WLR（重みつき尤度比） 尺度	$\mathrm{WLR}=\frac{1}{2\pi}\int_{-\pi}^{\pi}\left\{\left(\log\frac{f(\lambda)}{g(\lambda)}+\frac{g(\lambda)}{f(\lambda)}-1\right)\frac{f(\lambda)}{u^{(f)}}\right.$ $\left.+\left(\log\frac{g(\lambda)}{f(\lambda)}+\frac{f(\lambda)}{g(\lambda)}-1\right)\frac{g(\lambda)}{u^{(g)}}\right\}d\lambda$ $=2\sum_{n=1}^{\infty}(r_n^{(f)}-r_n^{(g)})(c_n^{(f)}-c_n^{(g)})$

表4．線形予測分析関連パラメータの記法

	標準パターン	入力音声
スペクトル包絡	$f(\lambda)$	$g(\lambda)$
パワー	$u^{(f)}$	$u^{(g)}$
自己相関係数	$r_j^{(f)}$	$r_j^{(g)}$
予測係数	$a_j^{(f)}$	$a_j^{(g)}$
最尤パラメータ	$A_j^{(f)}$	$A_j^{(g)}$
正規化残差	$R^{(f)}$	$R^{(g)}$
ケプストラム係数	$c_n^{(f)}$	$c_n^{(g)}$

$i=1,\cdots,p,\quad j=-p,\cdots,p$

p：線形予測モデルの次数

$n=-n_0,\ \cdots,\ n_0$

$A_j^{(f)}=\sum_{i=0}^{p-j}\ a_i^{(f)}\ a_{i+|j|}^{(f)},\quad a_0=1$

$$f(\lambda)=\frac{1}{2\pi}\cdot\frac{u^{(f)}R^{(f)}}{\sum_{k=-p}^{p}A_k^{(f)}e^{jk\lambda}}$$

$$g(\lambda)=\frac{1}{2\pi}\cdot\frac{u^{(g)}R^{(g)}}{\sum_{k=-p}^{p}A_k^{(g)}e^{jk\lambda}}$$

孤立単語音声認識は区切って発声された単語を認識するものであり、連続単語音声認識が連続した単語をすべて認識しようとするのに対し、音声理解は入力されたすべての言葉を認識することよりも、種々の言語的知識を用いて、その意味を理解することを目的とする。

7．2．スペクトル距離尺度

ほとんどの音声認識において、入力音声と標準パターンとの短時間スペクトルの距離又は類似度が認識の判定の基礎として用いられる。スペクトル分析には、3．1項で述べた種々の方法が用いられる。実際の距離尺度としては、NPAの場合はスペクトルの単純なユークリッド距離を用いることが多いが、聴覚の感度を考慮した重みづけを行ったり、判別分析、主成分分析等の統計的分析によって射影した、より低次元の空間での距離を用いる試みも行われている。

LPC分析に基づく距離尺度には、表3に示す種々のものが提案されている。ただしパラメータの定義は表4の通りである。WLR尺度は、ホルマントのようなスペクトルの山部に音声認識の重要な情報が存在しているため、スペクトルの山部における距離を強調するようにした距離尺度である[23]。ただし、$f(\lambda)$及び$g(\lambda)$はあらかじめスペクトルの全体的傾斜がなくなるように平坦化しておく。

7．3．DPマッチング

同じ人が同じ単語を発声しても、その長さはそのつど変わり、しかも非線形に伸縮する。このため、標準パターンと入力音声の全体としてのスペクトル距離を計算するためには、両者の同じ音素どうしが対応するように時間軸を非線形に伸縮する時間正規化（DTW）を行う必要がある。このためには動的計画法（DP）を用いることができ、このようにして標準パターンと入力音声を対応づけることを、DPマッチングという[24]。

対応すべき2つの時系列（スペクトルを示すベクトルの系列）を、次のように表わす。

$$\begin{cases} A=\mathbf{a}_1,\mathbf{a}_2,\ldots,\mathbf{a}_I \\ B=\mathbf{b}_1,\mathbf{b}_2,\ldots,\mathbf{b}_J \end{cases} \tag{9}$$

図17に示すようなA、Bからなる平面を考えると、時間伸縮関数、すなわちA、B両時系列の時間軸の対応づけは、この平面上の格子点$c=(i,j)$の系列

$$F=c_1,c_2,\ldots,c_k,\ldots,c_K \quad c_k=(i_k,j_k) \tag{10}$$

で表現することができる。2つのベクトル$\mathbf{a}_i$と$\mathbf{b}_j$との距離を$d(c)=d(i,j)$で表わすと、Fに沿った距離の和は、

$$D(F)=\sum_{k=1}^{K} d(c_k)w_k \Big/ \sum_{k=1}^{K} w_k \tag{11}$$

で表わすことができ、この値が小さいほどAとBの対応づけが良いことを示す。ここでw_kを市街化距離：$w_k=(i_k-i_{k-1})+(j_k-j_{k-1})$、$(i_0=j_0=0)$、とすると、

$$\sum_{k=1}^{K} w_k = I+J \tag{12}$$

となり、式(11)は

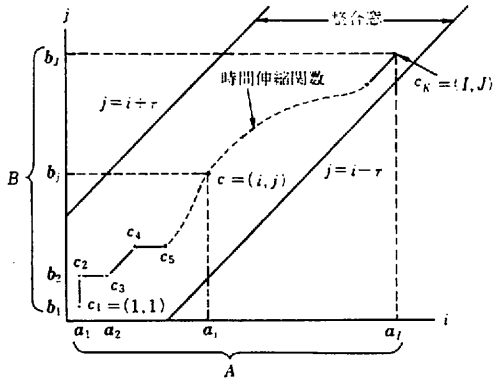


図17．A、B2つの時間軸の対応づけ

$$D(F)=\sum_{k=1}^{K} d(c_k)w_k / (I+J) \tag{13}$$

となって簡単化される。これをFに関する単調性、連続性、極端な伸縮を防ぐための整合窓の条件等のもとで最小化する。

部分点列$c_1,c_2,\ldots,c_k$　$(c_k=(i,j))$に対する部分和を考えると、

$$\begin{aligned} g(c_k)=g(i,j) &= \min_{c_1,\ldots,c_{k-1}} \Big[\sum_{\ell=1}^{k} d(c_\ell)w_\ell \Big] \\ &= \min_{c_1,\ldots,c_{k-1}} \Big[\sum_{\ell=1}^{k} d(c_\ell)w_\ell + d(c_k)w_k \Big] \\ &= \min_{c_{k-1}} \Big[\min_{c_1,\ldots,c_{k-2}} \Big(\sum_{\ell}^{k-1} d(c_\ell)w_\ell \Big) + d(c_k)w_k \Big] \\ &= \min_{c_{k-1}} \big[g(c_{k-1}) + d(c_k)w_k \big] \end{aligned} \tag{14}$$

これはDPの定式化になっている。上で述べた種々の条件とw_kの定式化を用いてこの式を書きかえると、

$$g(i,j)=\min \begin{bmatrix} g(i,j-1)+d(i,j) \\ g(i-1,j-1)+2d(i,j) \\ g(i-1,j)+d(i,j) \end{bmatrix} \tag{15}$$

となる。従って、$g(1,1)=2d(1,1)$、$j=1$　として、整合窓の範囲内でiを変えながら上式を計算し、次にjを増加させてj＝Jとなるまで同様の計算を繰り返せば、最後に$g(I,J)/(I+J)$として、A、B2つの時系列の時間正規化後の距離が求まる。

DPマッチングの方法としては、音素の始端・終端の検出位置の変動に対処するための端点フリーDP、計算量を削減するための圧縮DPやstaggered array DP、等の方法が提案されている。さらに、連続して発声された単語の認識のために、2段DPマッチング[25]、LB（level building）法、CW（clockwise）DP法、OS（one-stage）DP法、O(n)DP法、連続DP法等が提案されている。

7．4．SPLIT法

入力音声と標準パターンを5～20ms程度の細かい時間毎のスペクトルの系列で表わし距離を計算したのでは、語彙数が大きくなったときに膨大な計算量となってしまうため、5．8項で説明したベクトル量子化を用いて計算量を削減する方法が提案された。このときに用いられるコードベクトルを擬音韻標準パターン、この認識方法をSPLIT（Strings of Phoneme-Like Templates）法と呼ぶ[26]。SPLIT法による単語音声認識系のブロック図を図18に示す。各単語標準パターンは、擬音韻標準パターン（128～256種類）を示す符号の列で表現される。入力音声は短区間毎に全擬音韻標準パターンとの距離が計算され、これを距離行列の形で蓄える。次にこの距離行列を用いて、各単語辞書とのDPマッチングを行い、単語を決定する。

この方法をさらに進めて、単語標準パターンのみならず、入力音声の方もベクトル量子化する方法が提案されている[27]。この方法は、擬音韻のすべての対について距離をあらかじめ蓄えておく必要があるので、記憶容量が大きくなるが、入力音声のベクトル量子化を効率よく行うことによって、スペクトル距離の計算量を削減できる。

7．5．HMM法

各単語をその音素数程度すなわち5個程度の少ない状態の遷移図で表わし、各状態での擬音韻標準パターンの生起確率と状態間の遷移確率を与えて、入力音声をこのモデルに当てはめて認識する．hidden Markov model（HMM）に基づく方法が最近盛んに研究されている[28]。この方法は、確率モデルを与えるための学習処理が複雑になるが、認識時の処理量が少なくてすむ特徴がある。

図19は、不特定話者の単語音声を認識する場合の、従来の方法で各単語について複数のスペクトル時系列を標準パターン（multiple templates）として蓄える方法と、HMMによる方法を比較して示したものである。図20にHMMの例を示す。q_1が先頭の擬音韻、q_5が最後の擬音韻、$A=(a_{ij})$が遷移行列、$B=(b_{jk})$が生起確率行列を示す。

まず学習時には、各単語ごとに例えば100個（$k=1,\ldots,100$）の学習サンプル

$$O^{(k)}=(O_t^{(k)})_{t=1}^{T_k} \quad (T_k：フレーム数) \tag{16}$$

を用いて、

$$(A^*, B^*)=\underset{(A,B)}{\operatorname{argmax}} \prod_{k=1}^{100} \mathrm{Prob}(O^{(k)} \mid A, B) \tag{17}$$

となる(A^*, B^*)をBaum-Welchのアルゴリズムで決定する。

次に認識時には、例えば各10数字のモデルを(A_0, B_0)、(A_1, B_1)、・・・・、(A_9, B_9)、入力を$O=O_1O_2\cdots O_T$として、

$$(i^*, q^*)=\underset{(i,q)}{\operatorname{argmax}} \mathrm{Prob}(O, q \mid A_i, B_i) \tag{18}$$

となる(i^*, q^*)をViterbiのアルゴリズムで決定する。ここでqは状態系列である。

最近では入力音声をVQせずに、直接HMMに適用する方式（連続HMM法）も提案されている[29]。

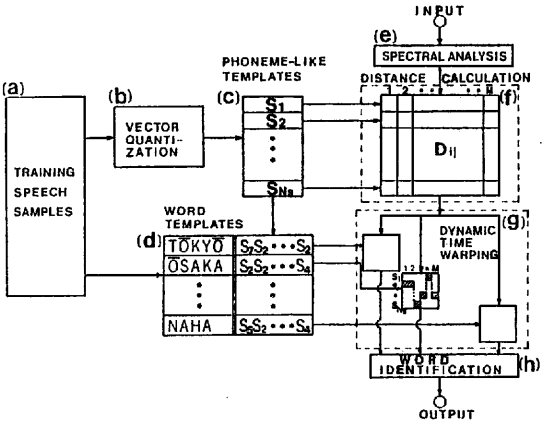


図18．SPLIT法による単語音声認識系のブロック図

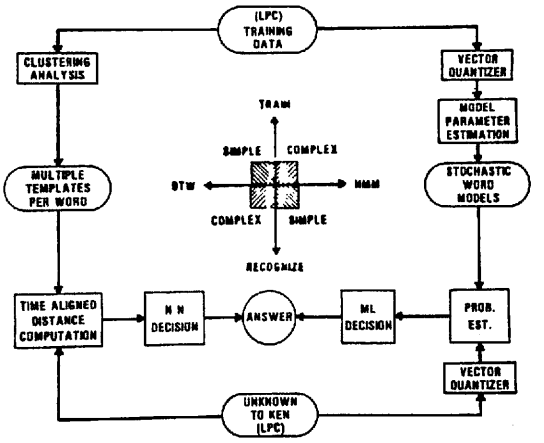


図19．複数テンプレートによる方法とHMMによる方法の比較

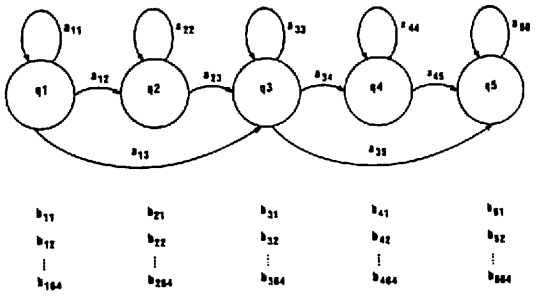


図20．HMMの例

8．話者認識

音声波に含まれる個人性情報を抽出して、誰の声かを判定することを話者認識という。話者認識と音声認識は本来表裏一体をなすべきものであるが、音韻性情報と個人性情報は複雑にからみあった形で音

声波（スペクトル）に含まれており、この両者を明確に分離して抽出する技術はまだ確立されていない。このため、音声認識では発声者を、話者認識では発声内容を固定することによって、高い認識率を得ているのが現状である。

話者認識に用いられる物理的特徴としては、スペクトル包絡、基本周波数、パワーなどの時間パターンやその統計的特徴がある。時間パターンを用いる方法は、音声認識で用いられる方法と類似しているが、安定した個人の特徴を強調するための、スペクトル等化などの処理が必要である。

スペクトルの動的特徴を用いた話者認識系のブロック図を図21に示す[30]。この方法では、10ms毎に抽出したLPCケプストラムの90msの区間での時間変化を直交多項式展開し、この展開係数の時系列で動特性を表現している。

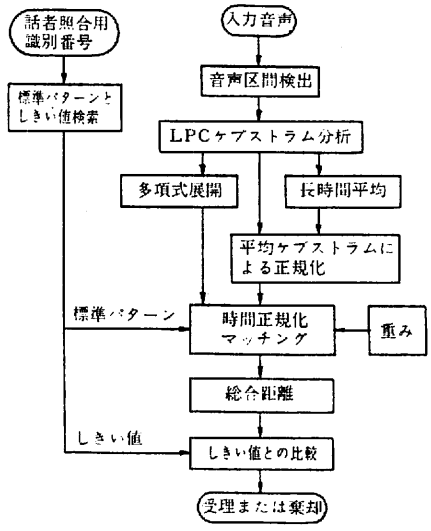


図21．スペクトルの動的特徴を用いた話者認識系の構成

9．今後の課題

9．1．過渡的信号の処理

音声波形およびそのスペクトル的特徴は常に時間的に変化しており、これまで定常と見なすことができるとして来た20ms程度以下の短区間でも、実際には変化している。しかも人間が音声を知覚する際には、この変化情報が重要な役割を果たしており、定常な部分よりもむしろ時間変化が極大となる時点（子音から母音への遷移部分）に最も重要な音韻性情報が存在していることが、実験的に確認されている[31]。しかし、このような過渡的な信号の性質を正しく安定に抽出する信号処理技術はまだ確立されてなく、今後の重要な課題であると言える。この技術が確立すれば、音声の符号化、合成、認識などに極めて大きな影響を与えると思われる。

これまでに、非定常波形の分析方法として試みられた方法の一つは、短区間分析法である[32]。これは、分析に用いる音声区間をできるだけ短くとることにより、変化に追随しようとするものであるが、分析区間を短くすると、結果が逆に不安定になる問題がある。その対策として、一般逆行列の手法を用いることにより安定性を増そうとする方法[33]、線形予測を階層的に適用する方法[34]などが提案されているが、いずれにしてもこれらの方法は短区間内での定常性を仮定しており、真の動的特性を分析しているとは言えない。これらの方法を実音声に適用する試みも行われているが、まだ有効な結果はもたらされていない。

分析区間内での過渡的特性を直接的に考慮する方法としては、すでに、予測係数が線形に変化することを許す非定常音声分析法[35]と、z平面上での極の位置が線形に変化することを許す極変動追随フィルタによる分析法[36]が提案されている。後者の方法は、1標本化時間幅の極の変動成分を、生成フィルタの逆フィルタのテイラー展開を利用することによって計算し、極位置を補正しようというものである。これらの方法が、信号の定常性を仮定したときの予測残差信号から非定常性の情報を抽出しようとしているのに対し、直接的にパラメータの非定常性を組入れた推定法が最近提案されている[37][38]。

一般的な過渡的信号の分析方法として最近注目されている方法に、Wigner分布を用いる方法がある[39]。これは次のように定義される。

相互Wigner分布：

$$W_{fg}(t,\omega)=\int_{-\infty}^{\infty} e^{-j\omega\tau} f\left(t+\frac{\tau}{2}\right) g^{*}\left(t-\frac{\tau}{2}\right) d\tau \qquad (19)$$

自己Wigner分布：

$$W_{f}(t,\omega)=\int_{-\infty}^{\infty} e^{-j\omega\tau} f\left(t+\frac{\tau}{2}\right) f^{*}\left(t-\frac{\tau}{2}\right) d\tau \qquad (20)$$

これは、FM音など単純な非定常信号に対してはその動特性をよく表現するが、音声のように複雑な高調波構造をもつ信号に対しては、高調波の相互作用で必ずしも良い結果が得られないようである。

音声波に含まれる重要な動的特徴は、上述のような短区間内の過渡特性だけではなく、50ms程度あるいはそれ以上の時間区間における変化も重要な情報をになっている。これらの特徴の表現には、20ms程度の区間内では定常性を仮定し、この区間を5～10ms程度ずつずらしたときの分析結果の時間変化を近似的に用いることができる。このような考え方にもとづいて検討されたのが、話者認識の項で述べた動的特徴による方法であり、この方法は単語音声認識系[40][41]においても有効性が確認されている。

9．2．知識処理との融合

ディジタル信号処理自体の問題ではないが、今後の音声研究の進展においては、ディジタル信号処理のレベルと知識処理とをいかにうまく融合させるかが重要な課題となると予想される。特に音声認識さらに音声理解においては、音声波形に対する種々の並列的なディジタル信号処理の結果と、音韻規則、辞書、構文、意味、一般的知識、文脈情報、等の大量な知識をいかにうまく表現し、いかにうまく統合して推論を進めるかが重要な鍵となろう。これは、あいまい性のある確率モデルと知識処理をいかに融合させるかという問題とも言える。このような観点から、エキスパートシステムとしての音声理解モデルがさかんに研究され始めているが[42][43][44]、まだこれまでの音声理解プロジェクトのレベルを超えるところまでは進んでいない。人工知能、音声処理、ディジタル信号処理の研究者の密接な協力関係が必要であり、これがそれぞれの分野での学問の発展に重要な刺激を与えると思われる。

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(昭60)

[2]板倉、東倉:"音声の特徴抽出と情報圧縮",情報処理、19、7、pp.644-656(昭53)

[3]嵯峨山、板倉:"音声の動的尺度に含まれる個人性情報"、音学春季講論、3-2-7(昭54)

[4]Atal,B.S.:"Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification",J.Acoust.Soc.Amer.,55.6,pp.1304-1312(1974)

[5]Markel,J.D. and Gray,Jr.,A.H.:"Linear prediction of speech", Springer-Verlag(1976)

[6]Atal,B.S. and Hanauer,S.L.:"Speech analysis and synthesis by linear prediction of the speech wave",J.Acoust.Soc.Amer.,50.2(pt.2),pp.637-655(1971)

[7]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、53-A、1、pp.35-42(昭45)

[8]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(昭48)

[9]板倉文忠:"線形予測係数の線スペクトル表現"、音響学会音声研資、S75-34(昭50)

[10]板倉文忠:"スペクトル符号化に基づく音声分析合成"、音響学会誌、37、5、pp.197-203(昭56)

[11]沖見、小塚:"音声符号化方式の開発動向"、施設、電電公社、34、12、pp.17-25(昭57)

[12]Crochiere,R.E. and Flanagan,J.L.:"Current perspectives in digital speech",IEEE Commun.Magazine,January,pp.32-40(Jan.1983)

[13]Atal,B.S. and Schroeder,M.R.:"Adaptive predictive coding of speech signals",Bell Syst.Tech.J.,49,8,pp.1973-1986(1970)

[14]Esteban,D. and Galand,C.:"Application of quadrature mirror filters to split band voice schemes",Proc.IEEE Int.Conf.Acoust.,Speech,Signal Processing,Hartford,CT,pp.191-195(1977)

[15]Zelinski,R. and Noll,P.:"Adaptive transform coding of speech signals",IEEE Trans.Acoust.,Speech,Signal Processing,ASSP-25,4,pp.299-309(1977)

[16]誉田、板倉:"音声の適応ビット割当て予測符号化"、音響学会音声研資、S80-2(昭55)

[17]Atal,B.S. and Remde,J.R.:"A new model of LPC excitation for producing natural-sounding speech at low bit rates",Proc.IEEE Int.Conf.Acoust.,Speech,Signal Processing,Paris,France,pp.614-617(1982)

[18]誉田、守谷:"位相等化処理を用いた音声符号化"、音響学会音声研資、S84-05(昭59)

[19]Gersho,A. and Cuperman,V.:"Vector quantization:A pattern-matching technique for speech coding",IEEE Commun.Magazine,December,pp.15-21(1983)

[20]白木、誉田:"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、S85-45(昭60)

[21]藤崎、広瀬:"規則による音声合成"、音響学会誌、37、5、pp.204-209(昭56)

[22]佐藤、匂坂、小暮、嵯峨山:"日本語テキストからの音声合成"、通研実報、32、11、pp.2243-2252(昭58)

[23]杉山、鹿野:"ピークに重みをおいたLPCスペクトルマッチング尺度"、信学論、J64-A、5、pp.409-416(昭56)

[24]迫江、千葉:"動的計画法を利用した音声の時間正規化に基づく連続単語認識"、音響学会誌、27、9、pp.483-500(昭46)

[25]Sakoe,H.:"Two-level DP-matching--A dynamic programming-based pattern matching algorithm for connected word recognition",IEEE Trans.Acoust.,Speech,Signal Processing,ASSP-27,6,pp.588-595(1979)

[26]管村、古井:"疑音韻標準パタンによる大語い単語音声認識"、信学論、J65-D、8、pp.1041-1048(昭57)

[27]鹿野清宏:"入力音声のベクトル量子化による単語音声認識"、音響学会音声研資、S82-60(昭57)

[28]Levinson,S.E. Rabiner,L.R. and Sondhi,M.M.:"Speaker independent isolated digit recognition using hidden Markov models",Proc.IEEE Int.Conf.Acoust.,Speech,Signal Processing,Boston,MA,pp.1049-1052(1983)

[29]Juang,B.H. Rabiner,L.R. Levinson,S.E. and Sondhi,M.M.:"Recent developments in the application of hidden Markov models to speaker-independent isolated word recognition",Proc.IFEE Int.Conf,Acoust.,Speech,Signal Processing,Tampa,FL,pp.9-12(1985)

[30]Furui,S.:"Cepstral analysis technique for automatic speaker verification",IEEE Trans.Acoust.,Speech,Signal Processing,ASSP-29,2,pp.254-272(1981)

[31]古井貞熙:"音声知覚におけるスペクトル変化情報の役割り"、音響学会聴覚研資、H85-6(昭60)

[32]河原、栃内、永田:"小区間の線形予測分析とその誤差評価"、音響学会誌、33、9、pp.470-479(昭52)

[33]柳田、木村、角所:"一般逆行列を用いた音声の線形予測分析"、信学論、J63-A、10、pp.665-671(昭55)

[34]白井、曹輪:"非定常過程に対するARモデルの適用に関する一考察"、音響学会音声研資、S81-75(昭57)

[35]中島、鈴木:"非定常態音声分析法"、音響学会音声研資、S80-42(昭55)

[36]赤木、飯島:"極変動追随フィルタの一構成法"、信学論、J67-A、2、pp.133-140(昭59)

[37]木竜、飯島:"局所非定常処理の検討について"、信学会パターン認識と学習研資、PRL84-32(昭59)

[38]今泉、樺沢:"パラメータの変動を仮定した時間依存Burg法"、信学会総合全大、1646(昭59)

[39]Claasen,T.A.C.M. and Mecklenbrauker,W.F.G.:"The Wigner distribution - A tool for time-frequency signal analysis -", Philips J.Research,35,6,pp.276-300(1980)

[40]古井貞熙:"音声スペクトルの動的特徴を用いた単語音声認識"、音響学会音声研資、S84-65(昭59)

[41]古井貞熙:"スペクトル変化強調による単語音声認識"、音響学会音声研資、S88-77(昭61)

[42]De Mori,R. Laface,P. and Mong,Y.:"Parallel algorithms for syllable recognition in continuous speech",IEEE Trans.Pattern Analysis,Machine Intelligence,PAMI-7,1,pp.56-69(1985)

[43]辻野、溝口、角所:"連続音声認識エキスパートシステムとその支援環境 - 音声認識と知識工学の融合 -"、情処学会知識工学と人工知能研資、41-3(昭60)

[44]片桐、古井:"スペクトルリーディングに基づく音韻特徴探索システム"、音響学会音声研資、S84-56(昭59)