氏名	田中和世（たなかかずよ）
授与学位	工学博士
学位授与年月日	昭和59年7月11日
学位授与の根拠法規	学位規則第5条第2項
最終学歴	昭和45年5月 横浜国立大学工学部電気工学科卒業
学位論文題目	**音素的単位による音声の自動認識に関する基礎的研究**
論文審査委員	東北大学教授 城戸 健一　東北大学教授 木村 正行 東北大学教授 曽根 敏夫

論文内容要旨

最近の音声認識技術の進歩によって，限定された性能の範囲（話者登録型の限定語彙認識）においては，実用的装置が提供されるまでになった。しかし，制約の少ない高性能な音声認識システムの開発には，まだ多くの解明，解決されるべき問題が残されている。今日の技術水準を考察すると，音声認識技術の質的向上の主要な方向は，不特定話者の発声した音声の許容，連続発声した音声の許容，大容量語彙の許容ということになる。この3方向の技術向上にとって共通の基盤的要素となるのが，音声の音素レベルにおける高精度な識別技術である。本研究の主題はこの音素的レベルにおける自動認識に関するものであり，この問題を次のような基本要素に分割して論ずる。

(1) 音声信号波形から音韻特徴を表わす音響的パラメータを抽出すること，およびそれらのパラメータの評価。

(2) 音響的パラメータによる音韻の表現方法（表現モデルと領域の決定）。

(3) 未知入力音声から得られたパラメータ時系列の音韻系列へのカテゴリー化。

本論文では，上記の各段階で生ずる問題を考察，整理し，それらに対する具体的処理方法を提案する。手法の有用性は，実験によって検証される。議論は，取り分け，音声認識に固有な問題に重点が置かれる。これは，1960年代以前から続けられて来た音声認識に関する基礎的な研究と，最近の自動認識技術に関する研究との落差を埋めること，即ち，音声現象に固有な特性を自動認識システムに組み込むことに焦点を当てているためである。

本論文は6章よりなり，第1章は緒論である。緒論では，まず，本研究の背景，意義，目的について述べる。次に，音声認識システムの基本構成を考察し，その中において音素的単位の占める位置を明確にする。そして，高性能システムの実現にとって音素的単位の認識が最も有利であることを明らかにする。次に，音素レベルの識別システムが基本的に上記の3要素より成り立つことを述べ，その各部分においていかなる課題が存在するかを整理して具体的に示す。

第2章は，上記(1)のうち，特徴パラメータの抽出手法について述べる。本論文で用いる主要なパラメータは，ホルマントに相当するパワースペクトル包絡の局所ピーク（以後，擬似ホルマントと呼ぶ）である。ホルマント関連量は，従来，自動抽出が困難であったため，自動認識システムに使用されることは少なかった。しかし，その利点としては，話者や文脈による特性の変動を幾何学的枠組の中で捉えられ，直観性に優れること，また，耐雑音性に優れる点等がある。本章で提案する手法は，スペクトル包絡の時空間的な場のもつモーメントに基づいて，スペクトル場に掛けるガウス型窓を制御して擬似ホルマントを抽出するものである。この手法は，従来のホルマント抽出法のもつ問題点である解の収束性や逆追跡アルゴリズム等の困難を解消し，且つ，従来法では，抽出が不安定なため，その有用性がほとんど議論されて来なかった，ホルマントの強さと拡がりについても安定に抽出できることを示す。また，同様に安定な抽出が困難であったアンチホルマントについても，スペクトル包絡の局所谷として抽出できることを示す（これを以後，擬似アンチホルマントと呼ぶ）。図1は，本手法によって抽出されるパラメータの種類と処理過程の骨格を示したものである。

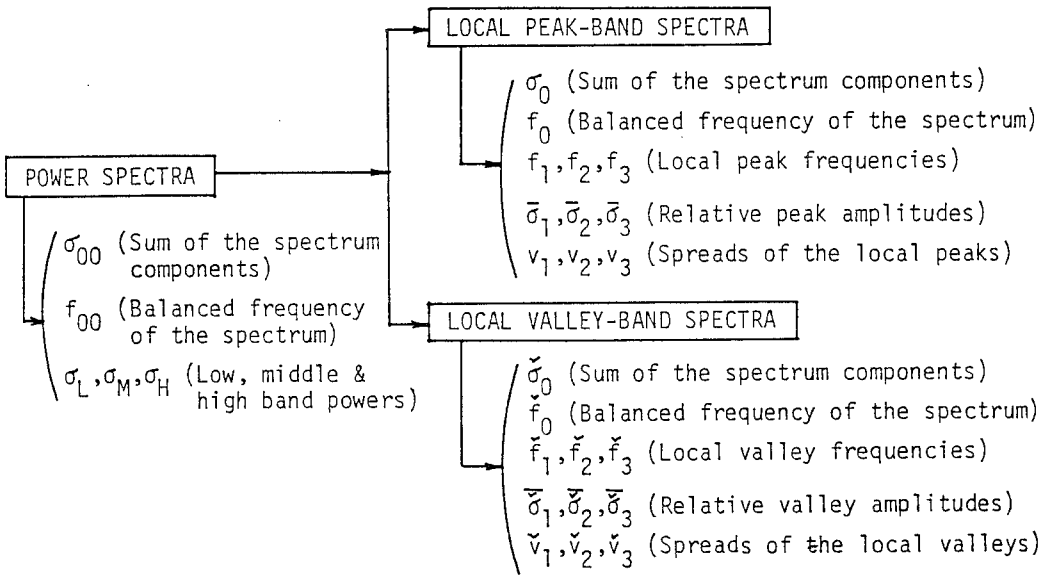

図1　パラメータの種類とその抽出過程の骨格

第3章は，上記(1)に関連したパラメータの分析・評価システムについて述べる。これは，同時に，音韻の音響的類別特性の調査研究にも相当する。本論文では，この問題を音韻特徴研究のための音声研究用データベース構築の一部としての観点から扱う。本章の前半では，本研究で作成した音声

データの収集・編集システムおよび分析・評価システムについて述べる。収集・編集システムは次の2つの部分よりなる。1つは，十分な標本化周波数と量子化精度によって音声波形を収集し，ファイル化する部分であり，もう1つは，この共通の仕様をもつファイルから各研究者が必要に応じてデータを編集・変換して使用者向ファイルを作成する部分である。分析・評価システムの概要のブロック図は，図2に示すようなものである。本章の後半では，このシステムによって破裂音，摩擦音の音響的特性を検討し，第2章で抽出した擬似ホルマントとアンチホルマント関連パラメータの有用性を示す。破裂音については，ホルマントの位置のみでなく，強さや拡がりも類別に有効なパラメータであること，また，摩擦音については，擬似アンチホルマント等を用いることによって，簡単なアルゴリズムで高い精度の類別が可能なことを示す。これらの結果は，従来の知識に比べて，直観性，数量化法および統計的側面のいづれをも満足させるものである。

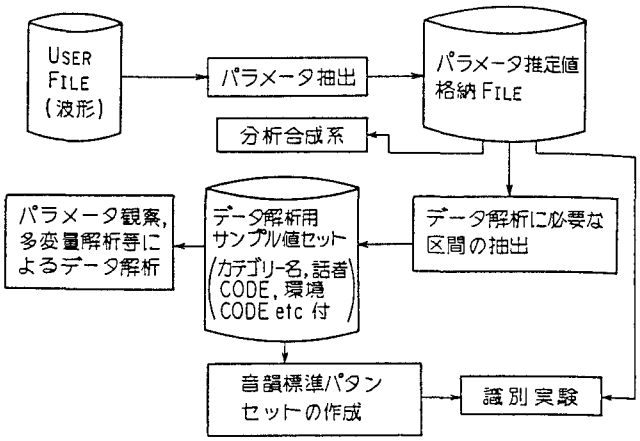

図2　パラメータの分析・評価システムのブロック図

第4章では，上記(2)の問題，即ち，音韻表現をいかなる空間においていかなる型によって表わし，また，その値（領域）をどのように決定するかという問題について論ずる。ここでは，音韻空間を音響的変量の時間的変動を含む行列パタンによって表わすことにする。これによって，全ての音韻を一様な型によって音韻空間内に定義でき，音韻表現量の解析的扱いが可能となる。各音韻が空間内で占める領域は，必ずしも単峰な凸領域を期待できない。従って，本方式では，領域はガウス型ポテンシャル関数の線形結合として定義される。学習サンプルセットとして与えられたサンプル分布から，このポテンシャル関数を決定する方法として，動力学的なクラスタリング手法を利用したアルゴリズムを提案する。この手法を破裂音の類別に適用して，その有用性を検証する。表1は，この手法によってクラスタリングされたサンプルセットの例である。このようにして得られた結果に基づいて，各音韻はそれぞれのサンプル分布に応じた適当な数のサブカテゴリーに分割されて代表される。従って，先験的に与えられたカテゴリーが無理なく bottom－up 的に表現される。この領域決定法は，普偏性のある手法であり，パラメータの次元数や与えられるカテゴリー数に依らない。本章の最後には，手法を逐次学習方式に拡張したアルゴリズムも示す。

第5章では，上記(3)の問題，即ち，未知入力音声より抽出されたパラメータ時系列を予め定められた音韻カテゴリーの系列へ変換する問題について論ずる。この変換は，基本的には，第4章で定められた音韻領域に基づいて時系列のカテゴリー化を行なうことであるが，このレベルの処理に固有の問題として調音結合がある。調音結合による音響的特性の処理は，次の2通りの場合に分けて

表1　無声破裂音サンプルセットのクラスタリングの結果。資料は男性8名（MA，MB，…，MH）が各2回発声した計240サンプル。

Input Categories		Clustering Result (Speakers)															
		MA1	MA2	MB1	MB2	MC1	MC2	MD1	MD2	ME1	ME2	MF1	MF2	MG1	MG2	MH1	MH2
/p/	pi	P2	P2	P2	P2	P2	P2	P2	P2	P2	P2	P2	P2	P2	P2	P2	P2
	pe	P3	P3	P3	P3	P3	P3	P3	P3	P3	P3	P3	P3	P3	P3	P3	P3
	pa	P1	P1	P4	P4	P4	P5	P4	P4	P4	P4	P4	P4	P1	P1	P4	P4
	po	P1	P1	P5	P4	P5	P5	P1	P1	P5	P5	P5	P5	P1	P1	P5	P4
	pu	P1	P1	P1	P1	P1	P1	P1	P1	P1	P1	P1	P1	P1	P1	P1	P1
/t/	ci	T1	T1	T1	T1	T1	T1	T1	T1	T1	T1	T1	T1	T1	T1	T1	T1
	te	T3	T3	T2	T3	T3	T3	T3	T3	T3	T3	T3	T3	T4	T4	T3	T3
	ta	T2	T2	T2	T2	T2	T2	T2	T2	T2	T2	T2	T2	T2	T2	T2	T2
	to	T5	T5	T2	T2	T2	T2	T2	T2	T2	T2	T2	T2	T4	T2	T2	T2
	cu	T1	T1	T1	T1	T1	T1	T1	T1	T1	T1	T1	T1	T1	T1	T1	T1
/k/	ki	K1	K1	K1	K1	K1	K1	K1	K1	K1	K1	K1	K1	K1	K1	K1	K1
	ke	K1	K1	K1	K1	K1	K1	K1	K1	K1	K1	K1	K1	K1	K4	K1	K1
	ka	K2	K2	K2	K2	K2	K2	K2	K2	K2	K2	K2	K2	K2	K2	K2	K2
	ko	K3	K3	K3	K3	K3	K3	K3	K3	K3	K3	K3	K3	K3	K3	K3	K3
	ku	K3	K3	K2	K2	K2	K2	K2	K2	K3	K2	K2	K2	K2	K2	K3	K3

考える必要がある。1つは，単音としての特性から不連続な変形を生ずる場合（例えば，母音の無声化，子音の口蓋化等）であり，これには，予め単独のカテゴリーを用意し，第4章で述べた手法によってその標準値を設定しておく必要がある。他の1つは，単音としての特性から連続的な変形を生ずる場合であり，変形は程度の差として現われる。この変形は，パラメータの動特性を利用して補正することが考えられる。本章では，主に後者の問題について考察し，観察結果とそれに基づいた補正手法を提案する。手法は，第4章の時空間音韻表現と連結しており，パラメータ時系列を時空間パタン（行列）の時系列と見なした処理を行なう。処理手法は，動力学的なものであり，時系列の行列による符号化，或いは量子化と捉えることもできる。行列による音韻空間表現により，従来試験的に行なわれてきた調音結合処理（主に母音を対象としている）を，子音等の非定常的音韻の調音結合の修復へと拡張することが可能となる。ここでは，語中の破裂音の識別に適用して，その有用性を検証する。

第6章は，本論文の結論である。結論では，各章の結果をまとめ，今後の研究の進むべき方向について述べる。

審 査 結 果 の 要 旨

近年実用化の緒についた音声自動認識は，話者や認識対象を大幅に限定しなければならないため，利用し得る範囲はまだ狭い。話者や発声内容に大きな制限を設けずに音声を自動認識するためには，認識単位を単語とする従来の方法では限界が低く，言語として見た音声の最小基本単位である音素を認識の単位として用いなければならないが，これは困難な問題である。著者は，音声自動認識の単位として音素もしくは音素に近いものを用いることを目的として研究を行い，新たな処理手法を提案した。本論文はその成果をまとめたもので，6章よりなる。

第1章は序論である。

第2章では，音声波形から音韻情報を担う音響的パラメータを抽出する手法を検討し，ホルマントとアンチホルマントの周波数のみならず，ホルマントの振幅と拡がりをも表現できるパラメータを推定する手法を提案している。

第3章では，音韻類別のための音響的特徴を代表するパラメータを知識源として蓄積し有効に利用することが必要であり，そのためには大量のデータ解析を要するという見地から本研究のために整備したデータ解析システムを説明し，破裂音と摩擦音の解析結果を示している。

第4章では，音韻表現にいかなる空間が適切であるかを論じ，時間要素を行方向にもつ行列によって表される空間を提案して，この空間に音韻の領域を設定する手法について述べている。これは，従来の主流であった時間要素を空間内の軌跡として捉える音韻表現とは異り，子音を含めた全ての音韻に適用し得るものである。本手法の効果は，破裂音への適用によって示されている。

第5章では，音響的パラメータの時系列として入力される音声を予め定められた音韻カテゴリーに変換する問題を扱い，音素的単位による音声自動認識の基盤を作っている。この変換においては，調音様式の変化に伴う単音としての特性からの不連続的変形や，調音器官の動きの遅れに伴う連続的な変形の処理が重要かつ困難な問題であるが，前者は第4章の手法によって作られたカテゴリーの追加により，後者は文脈をもつ動特性の利用により解決している。これによって普遍性のある定式化が行われたことは評価に値する。

第6章は結論である。

以上要するに本論文は，高水準の音声認識システムの基盤技術として，音素的単位による音声認識を周波数要素に時間要素を加えた空間を用いて行う手法を開発し，不特定話者の連続音声の認識に進歩を与えたもので，情報工学に寄与するところが少なくない。

よって，本論文は工学博士の学位論文として合格と認める。