氏名	三輪譲二（みわじょうじ）
授与学位	工学博士
学位授与年月日	昭和54年3月27日
学位授与の根拠法規	学位規則第5条第1項
研究科，専攻の名称	東北大学大学院工学研究科 （博士課程）電気及通信工学専攻
学位論文題目	**単語音声の自動認識に関する研究** **ー不特定話者を目標としてー**
指導教官	東北大学教授 城戸 健一
論文審査委員	東北大学教授 城戸 健一　東北大学教授 二村 忠元 東北大学教授 木村 正行　東北大学教授 星子 幸男

論文内容要旨

第1章 序論

音声は，人間と人間との最も有効なコミュニケーションの手段であることから，音声を人間から機械への情報入力の手段として利用するために，音声認識の研究が行われている。しかし，現在までの研究は，話者の個人差の影響に対処することが困難であることから，話者を登録話者に限定したものが多く，一般的な話者を対象とした情報入力の手段として音声認識を利用することができない。このため，不特定話者を対象とした音声認識の研究は，現在重要な研究課題となっている。

また，対象話者の他に認識単位の選択も，一般的な音声認識を実現するために重要である。すなわち，認識対象単語の変更や追加を容易にし，さらに，連続音声認識へ拡張するためには，音素を認識単位とすることが必要である。

本論文では，一般的な音声認識を可能とするための研究として，不特定話者を対象として音素を認識単位とする単語音声自動認識システムを実際に構成し，認識実験により，単語音声認識の検討を行っている。

第2章 単語音声自動認識システムの構成

不特定話者を対象とし音素を認識単位とする単語音声自動認識システムは，従来の音声認識の研究成果をふまえて，音声分析，特徴抽出，音素認識，認識された音素系列の誤り訂正，及び単

語認識の5段階で構成するものとした。以下に，構成方針を示す。

音素認識のための特徴パラメータとして，音素認識のために有効性が確認されており，声道長の正規化に有効なホルマント周波数に対応する音声スペクトルのローカルピークを，主なパラメータとして用いている。また，個人差，調音結合，発声速度に影響されないで音素のセグメンテーションを行うために，音素の時間変化とよく対応する音声スペクトルの概略形を表わすパラメータを併用している。

音素認識では，認識誤りをできるだけ少なくするために，日本語の音形の性質を利用している。

音素認識の誤りは，言語情報としての音素結合規則や単語辞書を利用して訂正し，最終的な単語認識を行うものとしている。

音声資料として，音素の出現頻度などから，人口10万人以上の166都市名単語を用いている。

第3章　音声スペクトルの概略形と音形構造を利用した音素認識

音素認識では，前章のシステムの構成方針に従い，音声の言語としての性質を利用している。

音素認識に先立って，入力音声を，低尖鋭度（Q＝6）の帯域濾波器を対数間隔（1／6オクターブ間隔）で配列した29チャンネルの周波数分析装置によって分析し，10msごとに音声スペクトルを得る。

次に，音声スペクトルの最小二乗近似直線により，声帯音源に含まれる個人差の影響を除去した後，音声スペクトルのローカルピークと音声スペクトルの概略形を表わすパラメータを抽出する。スペクトルのローカルピークは，ホルマント周波数に対応しており，主に音素認識のために利用し，概略形を表わすパラメータは，音素のセグメンテーションに利用する。

音素のセグメンテーションでは，個人差，調音結合，発声速度によって，音素境界を誤らないようにするため，パラメータの静的な値によらず，パラメータの動的な性質である時間構造の極大点，極小点に着目して，セグメンテーションを行う方式としている。

音素認識では，音素が，調音様式，調音位置によってさまざまな音響的性質を示すため，全ての音素を1つのパラメータを用いてセグメンテーションを行うことが困難である。このため，日本語の音形構造を利用して，子音，半母音，母音，拗音の順に階層的に音素のセグメンテーションと認識を行う方式としている。

音素認識の方式の有効性を確認するために，音素認識のための標準パターンの作成に使用した成人男性5名の発声した20都市名単語中の音素の認識実験を行った。母音，半母音／w／，半母音／j／，子音の認識率は，それぞれ，89％，70％，85％，87％であり，音素認識に，音声スペクトルの概略形と音形構造を利用することにより信頼度の高い音素認識を行うことができることが確認された。

第4章　音素結合規則による音素系列の誤り訂正

自然言語の音形構造では，全ての音素が自由に結合できる訳ではなく，限られた音素同志の結合しかできない性質がある。このような性質は，音素認識の誤り訂正のために有効であるので，

音素の結合における言語上の規則である音素結合規則を用いて，認識された音素系列の誤り訂正を行う。

本研究では，音素結合規則として，一般の音素認識システムでも起こり易く，かつ，出現頻度の高い無声化と長母音のための規則を中心として構成している。また，音素結合規則では，音素系列中の音素の出現順序ばかりでなく，音素の出現位置と音素の持続時間の特徴を利用することにより，信頼度の高い誤り訂正ができるようにしている。

構成した音素結合規則の有効性を検討するために，規則を構成する際に使用した標準話者の成人男性5名及びその他の成人男性10名が発声した166都市名単語中の音素認識結果に対して，規則の適用実験を行った。この結果，無声化や長母音のための規則が正しく適用される割合は，80%以上であり，高い誤り訂正率が得られることが確かめられた。このように高い訂正率が得られた理由は，音素の位置別に，動的特徴によって安定に得られる音素の持続時間を利用して，無声化や長母音の判定を行ったためである。また，標準話者以外の話者が発声した音声資料に対しても，標準話者の音声資料と同程度の誤り訂正率が得られ，構成した規則が不特定話者に対しても有効であることが確認された。

音素結合規則を適用した後の段階では，標準話者5名とその他の話者10名の発声した166都市名単語中の音素認識率は，それぞれ，67.0%，63.5%であり，15名全体の音素認識率は，64.7%であった。さらに，15名が発声した166都市名単語中の17055音素に対する脱落と付加の割合は，それぞれ，4.3%，10.1%であった。

この段階で得られる認識音素系列が，脱落や付加もなく全ての音素が正しく認識された割合は，標準話者とその他の話者で，それぞれ，6.2%，4.2%にすぎず，この段階までの処理だけで，高い単語認識率を得ることは困難である。

第5章　単語辞書を利用した単語認識

前章で述べたように，音素結合規則を適用した後でも，認識音素系列中には，音素の脱落，付加及び認識誤りを含んでいる場合が多い。このため，単語辞書のもつ音素の配列の情報を利用することにより，最終的な単語認識を行う方法を用いている。

すなわち，認識音素系列と単語辞書項目との間の類似度を定義し，認識音素系列との類似度が最大となる単語辞書項目に対応する単語を認識結果とすることにより，認識誤りを含む音素系列から単語を認識するものとする。

単語辞書項目 $\mathbf{D}$（$D_1 D_2 \cdots D_I$）と認識音素系列 $\mathbf{W}$（$W_1 W_2 \cdots W_J$）との類似度は，$\mathbf{D}$ と $\mathbf{W}$ のマッチングをとったとき，音素間類似度 ℓ（D_i，W_j）の和の最大値として定義する。ここで，マッチングの際，次の制限を加えることにより，極端な対応づけを排除している。

(1) 単語境界同志は必ず対応する。（端点固定）

(2) 付加又は脱落は2連続以内である。

(3) 付加と脱落は連続して生起しない。

ここで，$\mathbf{D}$ と $\mathbf{W}$ との間の類似度の計算には，計算効率を高めるため動的計画法を用いている。

音素間類似度 $\ell(D_i, W_j)$ としては，音素の付加又は脱落の尤度を，音素の認識誤りと同一の尺度で表わすため，音素認識の結果から得られる音素の認識誤り確率の対数値の成分からなるConfusion Matrixを用いるものとした。

この音素間類似度を，コンテスト（前後の音素の種類や音素の位置）よって分けることにより，単語認識率にどのように影響するかを検討した。その結果，Matrixを，語頭，語中及び語尾に分け，／ki／，／kj／，／hi／，／iwa／の子音と半母音，及び撥音／N／を異音として扱い，さらに，付加の尤度は，前後の音素の種類に分けると，コンテキストを考慮しない場合より，記憶容量を多く必要とするが，単語認識率が改善されることが確認された。成人男性15名が発声した166都市名単語の認識率の場合は，5.6％改善され，85.7％と高い認識率が得られた。

第6章　単語音声自動認識システムの検討

前章までに述べた方法により構成した単語音声自動認識システムに対して，種々の検討を行った。

まず，音素間類似度を表わすConfusion Matrixを作成した標準話者15名と不特定話者10名の計25名の発声した単語音声の認識実験を行った結果，20都市名，51都市名，166都市名単語の認識率は，それぞれ，97.2％，95.0％，84.0％となり，音素を認識単位とし，不特定話者を対象とする現時点のシステムとしては，きわめて高い認識率が得られた。これより，前章までに述べた方法が，不特定話者に対して有効であることが確認された。

次に，システムの設計に用いた都市名単語以外の空港名と数字を認識対象単語に選び，認識実験を行った結果，都市名単語と同程度の単語認識率が得られ，音素記号で記述した単語辞書だけを変更することにより，認識対象単語を容易に変更できることが確認された。

第7章　結　　論

不特定話者を対象とし音素を認識単位とする単語音声の自動認識の研究の総括を行い，各章の結論をまとめた。

審 査 結 果 の 要 旨

電子計算機を利用した情報処理技術の一環として，人間と機械との音声による対話が望まれている。そのためには不特定話者の発声する音声の自動認識が必要であるが，同じ単語でも発声者ごとに音響的特徴が非常に異なるために，現在実用化されているシステムは全て，システムに登録された話者の発声する限られた数の単語を認識対象とするものである。著者は，不特定話者の発声する単語音声の自動認識を目標として研究し，音素単位の認識と単語辞書の利用により，話者を限定しない単語音声の自動認識が可能となることを実証した。本論文はこの研究をまとめたもので，全文7章よりなる。

第1章は序論である。

第2章では，認識システム構成の方針と，本研究で構成したシステムの概要，および認識実験に用いた音声資料について述べている。

第3章では，音声スペクトルの概略形の音形構造を利用する音素のセグメンテーションと認識について，実験結果とあわせて論述している。すなわち，音声スペクトルから個人差を除去するためにスペクトルの最小二乗近似直線を利用して得た特徴量を使い，音素のセグメンテーションと認識を，子音，半母音，母音，拗音の順に行う方法を提案し，それによる音素認識実験を行って，この方法が有用であることを示している。

第4章では，前章の方法によって得られた音素系列に含まれる誤りを音素結合規則によって訂正する方法を提案し，実験によりその有効性を明らかにしている。

第5章では，まず，音素認識結果から直接単語認識を行うことは困難であるが，言語情報を利用する手段として単語辞書を用いることにより，単語の認識が容易になることを論じている。次いで音素認識部で得られた音素系列と音素記号で書かれた単語辞書の各項目との類似度を計算し，最も類似度の高い辞書項目を認識出力とする本論文の方法を詳しく説明している。

第6章では，本研究によって構成された単語音声自動認識システムを，話者，認識対象単語，使用方法等の観点から詳細に検討し，本システムの方法が不特定話者を対象とすることに適しているばかりでなく，認識対象単語の変更を容易に行えるという特徴をも持っていることを確かめている。

第7章は結論である。

以上要するに，本論文は話者を限定しない単語音声の自動認識について研究し，他に例を見ない高い認識率を得ると共に，音声自動認識の研究に多くの知見を加えたものであり，情報工学並びに通信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。