

氏名	Yu 俞	Rong 榮
授与学位	工学博士	
学位授与年月日	昭和61年3月25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) 情報工学専攻	
学位論文題目	不特定話者を対象とする日本語音声 中の母音認識に関する研究	
指導教官	東北大学教授 木村 正行	
論文審査委員	東北大学教授 木村 正行 東北大学教授 城戸 健一 東北大学教授 曾根 敏夫	

論文内容要旨

本研究は、不特定話者日本語音声中の母音認識を目指し、それを実現するための諸問題、特に個人差問題の解決を中心に行った。

第1章 序 論

本章では、まず本研究の背景と音声認識の現状を述べ、母音認識の音声認識における役割と重要性を論じ、本研究に用いる主な方法を簡略に述べた。最後に本論文の構成を説明した。

第2章 音声資料および分析方法

本章では、本研究において評価と実験のために用いた音声資料について説明した。また、音声資料の解析方法について述べた。

第3章 個人差問題の考察および解決方法についての検討

本章では、多数話者による音声資料を対象に、解析と認識実験の二面から個人差問題を検討した。主成分分析を用いて個人差問題について解析した結果、特徴ベクトル空間における母音の分布は個人差により非常に複雑な形を呈し、線形あるいは簡単な非線形的な変換では話者に関する正規化はほとんど不可能であることが分かった。パターンマッチングの方法を用いて、個人差の影響が含ま

れない完全クローズ認識実験と、個人差の影響が含まれる完全オープン認識実験を行った。その結果、どのデータセットでも、後者の認識率は前者より大幅に低下し、特に212単語音声については、認識率の低下は10%–28%までに達した。このことは個人差の影響の大きさを示す一方、従来の研究によく使われる、パターンマッチングによる認識方法の限界をも示している。

また、第3章では個人差問題を解決するための予備的研究の一つについて述べた。即ち、主成分分析を用いて話者の分類尺度を提案し、この分類尺度を用いて話者のクラスタリングを行い、マルチテンプレートを作成する方法を用いた母音認識について検討し良好な結果を得た。

第4章 母音の相対関係モデル

本章では、本研究の基礎となる母音の相対関係モデルについて述べた。従来の音声認識に使われる特徴量はほとんど絶対量であり、個人差の影響を受けやすい。本研究では、絶対量から相対量を求め、更に、ケプストラムベクトル空間における日本語5母音を相対関係（三項関係と二項関係）で結び付け、これらの相対関係の集合を相対関係モデルとして定義した。

TYPE 1相対関係はケプストラムベクトル空間における母音間の基本位置関係である。多数話者の音声資料についてTYPE 1相対関係を調べた結果、話者によってほぼ変わらない性質が明らかになった。

TYPE 2相対関係は低域の部分スペクトルに基づいて求められた相対関係である。4連鎖母音、文節などの音声資料について検討した結果、TYPE 2相対関係は強い調音結合を受けてもあまり影響されない性質を有し、しかも話者に関してもほぼ安定していることが分かった。

TYPE 3相対関係は主成分分析によって二次元まで圧縮された母音の特徴量から求められた。多数話者の音声資料について検討した結果、TYPE 3相対関係はおおまかでありながら、個人差と調音結合にあまり影響されない大局的な安定性がある。二次元の情報量のみにより、数多くの意味がある相対関係が求められることは、興味深いと思う。

部分帯域に基づく相対関係と全帯域に基づく相対関係の間に補完性があることは、これらの相対関係の分布から分かった。これは異なる相対関係の統合利用にとって有効な性質である。

パワースペクトルに関して、話者によらずほぼ安定した母音間の相対関係が求められる。これは母音についての相対関係の普遍性を示している。

第5章 最適候補系列を求めるアルゴリズム

本章では、母音の相対関係に基づいて、母音の候補 list から最適候補系列を求める三つのアルゴリズムを提案し、その説明と評価を中心に述べた。ここで提案する方法の立場は、母音を個々に評価し認識するより、母音系列全体に着目して評価する方が正確かつ信頼性の高い結果が得られることにある。本研究に用いるすべての音声資料に対して、all check 法、第一候補に基づく法、弛緩法の三つのアルゴリズムのそれぞれを用いて、初期候補 list から最適候補系列を求める実験を行った。その結果、三つのアルゴリズムとも正答率が高い最適候補系列が得られた。しかもこの正答率は初期候補 list の第一候補の正答率より大幅に上がっている。また、候補 list の構成に関する

検討から、listの長さは長ければ長いほど結果の信頼性が高い、長さは15（候補listに10個の母音が含まれる）以上が望ましい。また候補listに少なくとも4種類の母音が含まれることが必要であるが、各種類の母音の数のバランスが取れなくても結果にはあまり影響がない、などのことが明らかになった。

三つのアルゴリズムの性質を、結果の信頼性、母音系列の構成に関する適応性、初期候補listに対する依存度、候補の信頼性、計算量などの面から比較し、弛緩法に基づく方法がもっとも優れていることが示された。

第4章に述べた三つのTYPEの相対関係がそれぞれ単独に評価の基準として使われ、よい実験結果が得られた。また三つのTYPEの相対関係が同時に使われた時の実験結果は最もよい。これらの結果から、それぞれのTYPEの相対関係の有効性が検証された。特に部分スペクトルに基づくTYPE2相対関係、二次元特徴量に基づくTYPE3相対関係が単独で使われた時でもTYPE1の結果にあまり劣らない結果が得られるということは、興味深いことだと思う。

三つのTYPEの相対関係のそれぞれの特長をいかして統合する形で使われれば、よりよい結果が得られると思われる。

従来の方法と違って、外部標準パターンの情報にあまり依存しないで、主に認識対象それ自身の間の相対関係で候補の正当性が評価される。これによって入力データと外部標準パターンに現れる個人差をなくすあるいは縮小することができる。

第6章 相対関係に基づいた不特定話者日本語音声での母音認識法

本章では、相対関係モデルと弛緩法アルゴリズムに基づく不特定話者音声での母音認識法を提案した。まず入力データの一部分からなる母音系列に対して弛緩法アルゴリズムを施し、最適候補系列を求め、この正答率の高い系列を候補listの第一列に置き、一位候補に基づくアルゴリズムを用いてcheckを行い、正答率を更に高める。次に、求められた最適候補系列から、入力データに適応した新たな内部標準パターンを作成し、これを用いて母音認識を行う。それと同時に、入力データと内部標準パターン間に現れる個人性を追跡し、母音系列の更新および内部標準パターンの更新条件を獲得する。

従来の認識法と比べて、この母音認識法の特長は、ダイナミックに入力データの特徴を追跡し、必要に応じて、正答率が高い最適候補系列を求め、これに基づいて入力データに適応するような標準パターンが作られることである。このような過程は外部標準パターンの情報を使わないので、不特定話者母音認識法として、見込みのある方法である。

第7章 結 論

本章は結論であり、本研究の総括を行い、本論文のまとめを各章ごとに行った。

本研究の主な成果を下記にまとめて列記する。

1. 従来の音声認識に使われる特徴量がほとんど絶対量であったのに対し、本研究では絶対量から求められた相対的な特徴量を用いて、特徴量間における日本語5母音の関係を相対関係（三項関

係と二項関係)で表わす。母音間の基本関係を反映するTYPE 1, 調音結合の影響に強いTYPE 2, 大局的な安定性を持つTYPE 3の三つのTYPEの相対関係が定義され, それぞれの話者によってほぼ不変な性質と, 異なるTYPEの相対関係間に存在する補完性が示された。

2. 従来の認識手法はほとんど認識対象を孤立に扱うのに対して, 本研究では認識対象を系列ごとに評価する。その時外部標準パターンとは関係なく, 系列の中の母音間で互いに提供しあう相対関係そのものが評価される。本研究で提案されたアルゴリズムにより, 初期候補 list における真の候補の分布にほぼかわらず正答率の高い最適候補系列が求められる。特に弛緩法アルゴリズムは候補に対する評価を並列に行うイタレーションにより, 真の候補の信頼性(確率で表わす)を高めることができ, しかも計算の効率性, 信頼性, 適応性などの面でもよい性質を示した。

3. ダイナミックに入力データに適応して標準パターンを更新する認識方法を提案した。入力データの特性を追跡し, 随時に母音系列を更新し, 最適候補系列を求め, 内部標準パターンを更新する。この場合, 個人差が大きければ大きいほど, 母音系列と内部標準パターンが更新される可能性は大きくなる。また, 外部標準パターンの役割を, 三位までの候補を取り込むことだけにとどめることができる。これらは不特定話者母音認識にとっては望ましいことである。

以上に述べたことから, 本研究の最も重要な結論は, 本研究の結果に基づいた認識法は不特定話者音声中の母音認識法としてきわめて有効であるということである。この結論の基本的な根拠は, 認識の角度から見れば, 個人差というのは入力データと外部標準パターンの間にあらわれる差であり, そのため初期候補 list の第一歩の正答率が低くなる。しかし本研究の結果により, 初期候補 list の真の候補の分布にほぼ関係なく, 最終的に正答率が高い最適候補系列が求められるというところにある。

審査結果の要旨

日本語音声においては、母音の役割は特に大きい。母音が確実に認識されれば、子音の音響的環境が明らかとなり、また言語の文脈を認識に利用する大きな手掛りとなる。このような視点から母音認識の研究が行われてきたが、個人差や調音結合などの困難な問題があり、まだ十分な精度が得られていない。

著者は、同一話者が発声した日本語5母音の特徴空間における相対的な位置関係（相対関係）が、話者によらずほぼ不偏であることを見出し、その相対関係を評価基準とする不特定話者母音認識について研究した。本論文はその成果をまとめたもので7章よりなる。

第1章は序論である。第2章では、単音節、連鎖母音、文節及び単語中の母音など本研究で用いる9組（発声者計94人）のデータセットの構成と用途などについて述べている。

第3章では、主成分分析とパターンマッチングによる母音認識の実験から、第2章の各データセットにおける個人差について詳しく検討し、そのいずれにもかなりの個人差が見られることを確認している。

第4章では、不特定話者母音認識に母音間の相対関係を活用する立場から、異なる特徴空間に基づく3種の相対関係を定義し、各データセットを用いてそれらが個人差や調音結合によってどのように影響されるかを詳細に検討している。その結果、どのタイプの相対関係も個人差にはほぼ影響されないと見做し得るが、調音結合にはそれぞれ一部の相対関係がかなり影響されることを明らかにしている。これらは興味ある有用な知見である。

第5章では、第4章の相対関係を評価基準とする母音認識アルゴリズムを構成し、各データセットに対する認識実験によってその有用性を立証している。このアルゴリズムは、入力データとの単なるマッチングでは最大尤度にならない母音をも候補リスト中の他の母音との相対関係を用いてその信頼度を再評価することにより、認識精度を更に向上させることを始めて可能にしたもので本研究の重要な成果である。

第6章では、第5章の母音認識アルゴリズムを基にして、入力データの特性に常に適合するように動的に標準パターンを生成して個人差の影響を少なくする不特定話者母音認識法を提案し、それを各データセットに適用して高精度の認識結果を得ている。

第7章は結論である。

以上要するに本論文は、各母音を独立に扱う従来の認識手法に対し、話者に不偏と見做し得る母音間の相対関係を導入し、それらを評価基準とする不特定話者母音認識の手法を開発してその有用性を立証し、音声認識の分野に新たな知見を加えたもので、情報工学の発展に資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。