

氏名(本籍)	粕谷英樹(山形県)
学位の種類	工学博士
学位記番号	工博第218号
学位授与年月日	昭和45年3月25日
学位授与の要件	学位規則第5条第1項該当
研究科専門課程	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	連続音声の中の母音の音響的性質および認識に関する研究
	(主査)
論文審査委員	教授 城戸 健一 教授 大泉 充郎 教授 本多 波雄

論文内容要旨

第1章 緒 論

音声の自動認識・音声の合成の研究は、現在のところ音声情報処理工学の中心的な研究課題である。音声の自動認識の研究では、一般的な会話音声というよりはむしろ認識対象にある制限条件を付加して、たとえば数の限られた単語を識別するためのアルゴリズムを見出すことの方が、研究の方向性を明確にし得るし、そのような制限条件によって工学的な研究価値を失うものではない。識別対象をたとえば数字音声に限った場合、その対象のみに有効な識別アルゴリズムを求めることの方が実用化の観点からは有利であるかもしれないが、一般的な単語集合を識別対象とする場合は、音素・モーラ・音節のような言語学的単位での識別アルゴリズムを求めるのでなければ、識別システムの経済性・発展性の上で好ましいことではない。

本著では限られた数の単語音声の識別を目標として、言語音声の基本的単位である母音について、発声・聴き取りの両面からの接近によってその性質を明らかにし、その結果から帰納される単語中の母音を識別する2つの方法について論じる。

第2章 日本語単母音の音響的諸性質

母音の研究の上で最も基本的な日本語5母音を取扱う。7才の子供から25才の成人に到る118名の発声者による単独に発声された母音の低次の3つのホルマント周波数とピッチ周波数に着目し、母音の音響的特徴・個人性・弁別性について、調音レベルでの特性を考慮しながら詳細に検討した。本章で得られた知見は音声の自動認識の研究の基礎となるばかりでなく、音声の合成・伝送・知覚の研究にも貢献するものであるが、特に音声の自動認識の立場から重要な知見を述べる。発声者の年齢・性別にかかわらず、誤りなく単独発声の母音を識別するためには、低次の第1・第2ホルマント周波数だけでは不十分であって（特に成人男性の/e/と子供の/u/、成人男性の/a/と子供の/o/の間に境界領域の交叉が見られる）、更に第3ホルマント周波数あるいはピッチ周波数を考慮に入れなければならない（図1、図2）。ピッチ周波数あるいは第3ホルマント周波数は弁別の特徴として本質的に重要なパラメーターである。これらのことは線形判別関数による識別シミュレーションによって定量的に評価されている。

第3章 心理的聴空間における母音聴き取りのモデル

第2章で得られた音響分析的研究、特に母音弁別のためのホルマント周波数の検討結果をふまえて、母音の知覚過程を精神物理学の実験手段によって明らかにし、母音の弁別素性 (distinctive feature) およびその物理的関連量について考察し、ホルマント周波数の聴覚的意味づけを行なっている。

先ず物理音響刺激に対する心理的な聴覚パターンを表す心理的聴空間 (Psychological Auditory Space - PAS) および擬聴空間 (Pseudo - PAS) を導入して母音知覚のための静的モデルを設定した (図3)。これらの空間はユークリッド空間と仮定される。心理的聴空間は曖昧な母音も含んだすべての単母音の母音性 (vowel quality) と距離が対応する空間であり、擬聴空間は聴取者にとって標準的な5母音についてのみ母音性と距離が対応する空間である。すべての母音刺激 (S_i) は写像関数 (ϕ) によって pseudo - PAS に写像される (a'_i)。そこで定義されている標準母音 a_k^* ($k=i, e, a, o, u$) に対する効果は、その方向に長さ

$$|x_k| = d(a_k^*, a'_i) F[d(a_k^*, a'_i)]$$

だけシフトするベクトルで表され、最終的な刺激母音の音韻性を示す心理的聴空間座標 a_i は次式

によって与えられる。

$$a_i = a'_i + \sum_k \times_k$$

ここに $d(a_k^*, a'_i)$ は pseudo-PAS における a_k^* と a'_i のユークリッド距離であり、 $F[d(a_k^*, a'_i)]$ は距離の単調減少関数で、正規分布関数で近似される。このように母音知覚をモデル化し、PAS および pseudo PAS の構造（空間の次元数と各軸の言語学的意味づけ）多次元尺度化法によって自然発声の母音、合成母音について明らかにした。その結果心理的（擬）聴空間の次元数即ち母音の弁別素性の数は高々 3 であり、それは刺激音の種類に関係ないことが明らかとなった。又 2 つの軸は言語学的には拡散性 (diffuseness)、高音性 (acuteness) 素性に対応し、第 3 軸は聴取者によって異なる弁別素性である。写像関数 ϕ については、3 つのホルマント周波数の対数値の線形結合として表されることが分った。

第 4 章 連続音声の中の母音のホルマント周波数の性質

3 モーラ無意味単語 / hV₁C₁V₂C₂V₃/ の母音 V₂ について調音結合の観点から詳細に考察した。ここに V₁, V₃ は前母音の /i/ と後母音の /o/、C₁, C₂ は有声・無声の破裂音 /b, d(z), g, p, t(C), k/ であり、V₂ は日本語 5 母音 /i, e, a, o, u/ である。この章で明らかとなった主な知見を列挙すると次のようになる。

- (1) (調音結合の限界) 普通の発声速度のもとで母音の調音結合を考える場合、前後の子音およびもう一つ前後の母音との結合を考えるだけで充分である。
- (2) (調音結合によるホルマントシフト) 前後音素の F₂ locus あるいは target 周波数が高ければ母音 V₂ の第 2 ホルマント周波数は高く、低ければ低くシフトする。第 1 ホルマント周波数のシフト量は第 2 ホルマントに比べて小さい。
- (3) (前後音素との結合度の差) 前後の音素の種類に応じて結合の強さの違いは異なるが、平均的に見れば着目する母音の前の音素との結合の方が強い。
- (4) (発声のコマンド) 母音 V₂ の発声と同時に子音をはさんだ母音 V₃ の発声のコマンドが既に始まり、V₃ の発声に近づくにつれて次第に強くなる。又既に発声し終った V₁ の影響が V₂ の発声の終りまで現れる。
- (5) (第 1・第 2 ホルマント周波数のシフトの違い) 第 2 ホルマント周波数は前後の子音の調音の manner の違いよりは place の違いによるシフトの方が大きい。又前後の母音によるシフトは第 2 ホルマントでは大きく、第 1 ホルマントでは小さい。
- (6) (ターゲット周波数) 母音 V₂ のターゲット周波数は前後の母音・子音の種類によって異なる。即ちある母音を発声する場合にその母音に固有なターゲットがあって、前後の母音・子音はそのタ

ターゲットに向かって発声する場合の単なる摂動であるという解釈はなり立たない。

第5章 連続音声の中の母音の識別

第2章から第4章まで得られた知見をもとにして、単語音声の中の母音を識別する二つの方法について考察した。一つの方法は、母音区間の midpoint でのホルムント周波数のシフトを母音区間での始点・終点での値を知ることによって null context のそれに規格化し、規格化されたホルムント周波数にベーズ判定法を適用することによって母音を識別する方法である。この方法は規格化操作と識別の2段階よりなり、しかも計算あるいは処理が簡単であるが、識別率がホルムントの測定精度に強く依存する。又母音区間を正確に知る必要があり、特に母音連続あるいは拗音(半母音)と母音の結合の場合、境界の定義が問題となる。規格化の効果は前後の子音・母音の種類に応じて何種類かに分類し表にして持つことにより上げられるが、発声過程との対応を考えると不自然さは避けられない。

上に述べた方法を静的な(あるいは単母音参照的な)方法とすれば、もう一つの方法はむしろ動的な識別法である。つまり、実際我々が観測する現象としてのホルムント遷移の時間パターンを人間の発声過程の analogy から変形規則の形でとらえ、生成の形で表現された変形規則を逆にたどることによって識別がなされる。この方法は前に述べた方法に比べて処理が少々やっかいではあるが、前の方法で問題となった技術的・原理的困難さは少なくなる。単語においては、母音を識別することだけを考えたが、変形規則による識別法は子音あるいはもっと一般に音節・単語等の識別にも充分拡張でき、今後の phonology の研究と相俟って更に発展が期待されるものである。尚9名発声者による10都市名単語の識別シミュレーションを行ない、前者の方法では94.9%、後者では95.7% (多出力を正として)の識別正答率を得ている。

第6章 結 言

言語音声の基本的単位である母音の性質を、発声・聴き取りの両面から詳しく検討し、その基礎的な考察に基いて単語中の母音を識別するための識別アルゴリズムについて考察した。

本研究を行なうにあたり終始御指導戴いた東北大学城戸 健一教授、大泉 充郎教授、鈴木 久喜助教授、比企 静雄助教授に深く感謝致します。

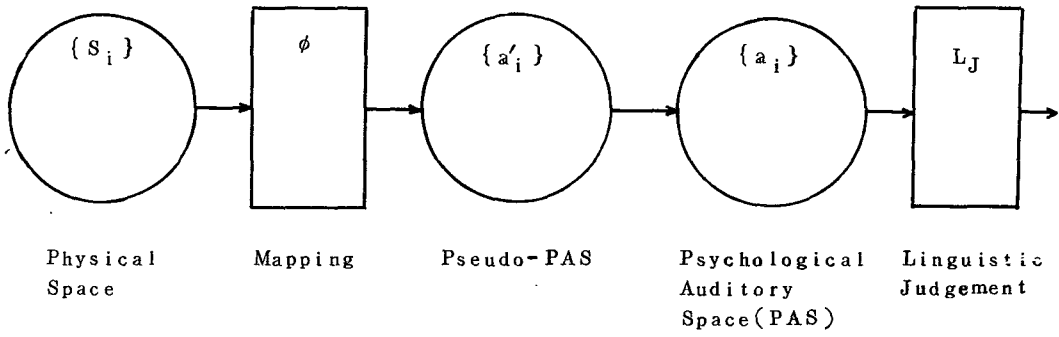


Fig.3 Static Model of Vowel Perception

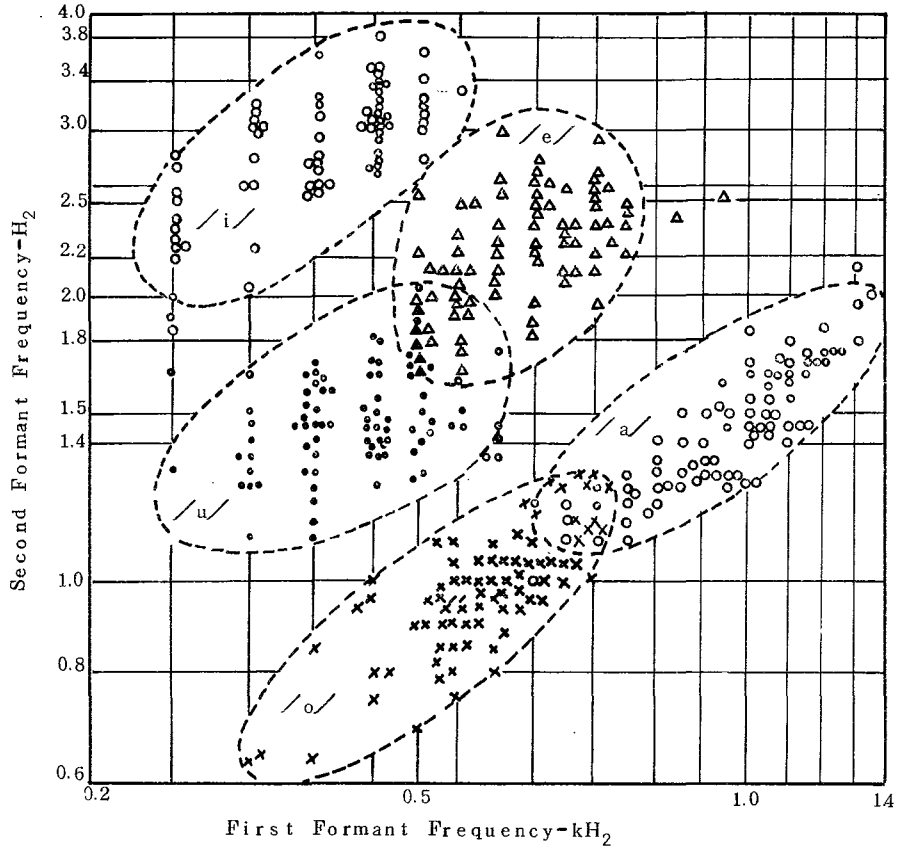
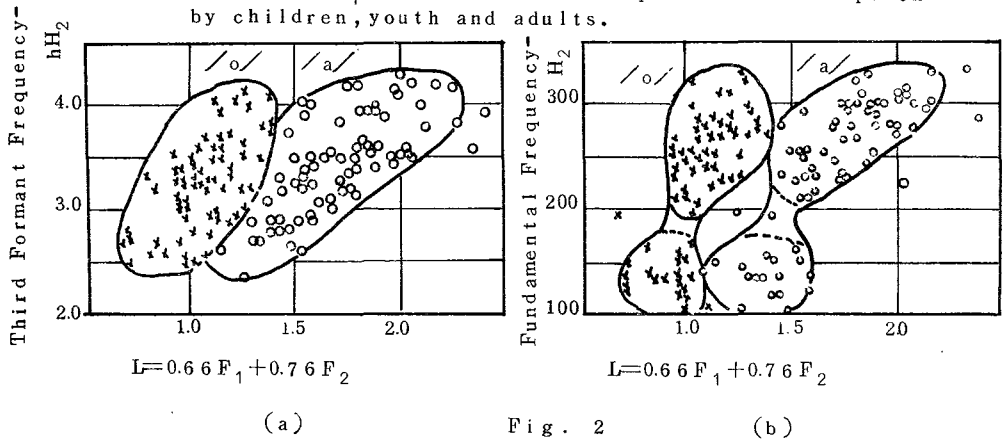


Fig.1 Second formant frequency (F_2) versus first formant frequency (F_1) diagram of five Japanese vowels spoken by children, youth and adults.



(a) Distributions of vowel measurements of /a/ and /o/ projected on the $L-F_3$ plane.

Fig. 2 (b) Distributions of vowel measurements of /a/ and /o/ projected on the $L-F_0$ plane.

審査結果の要旨

音声は、人間の行なう情報伝達の重要な媒体であって、それに関する研究は古くから行なわれて来たが、最近では、電子計算機に対する音声による情報の入出力という具体的な目標が現われ、研究の急速な進歩が要求されるようになった。しかしながら、音声の一般的な性質を論じ、連続音声の自動認識への道を開くには、未解決の問題が余りにも多すぎるのが現状である。このような状況にあることの一つの理由として、音声の音響的性質は、発声者により、前後の関係により、また、発声の仕方により大きく変化するため、多くの資料に基づく系統的な研究を必要とすることがあげられる。

本論文は、このような研究の段階で、著者が、日本語において重要な部分をなすと共に観測の比較的容易な母音部分を取り上げ、認識への応用を目的として、その音響的性質を明らかにするために行なった多数の資料に基づく体系的な研究の結果をまとめたもので、全6章からなる。

第1章は緒論であり、研究の背景と本論文の概要が述べられている。

第2章には、各年齢層の男女の発声した多数の単母音の分析結果と、それから結論されるホルマント周波数の正規化について述べてある。

第3章は、母音を人間が聴き取るときに、如何なる特徴量が如何なる形で用いられているかを推定するために行なった心理的聴空間に関する研究を述べたものである。本章は前章とともに、本研究においてホルマント周波数を主な特徴量として用いることの妥当なことを明らかにして、本研究の基礎を固めているのみならず、音声の聴取に関する新しい知見を与えるものとして注目されているものである。

第4章は、各種の調音結合によって母音のホルマントが如何に変化するかを明らかにした研究である。着目する母音が、その前後の子音を距てた外側の母音の影響を強く受けることを示し、それを体系的に整理して正規化の方法を明らかにしたのは、本研究が最初である。これによって、連続音声の中の母音は、そのホルマント周波数が得られれば高い率で認識されることが可能となったもので、この成果は高く評価されるべきである。

第5章は、本論文で得られた基礎的事実をもとにして、具体的な単語音声の中の母音の認識が可能となることを述べたもので、この研究の将来の発展の方向を示している。

第6章は結論である。

以上のように、本研究は、日本語音声の中の母音の音響的性質を体系的に明らかにすると共に、その自動認識に必要な基礎を与えたもので、音声情報処理に関する研究の進歩に寄与するところ少なくない。

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。