



HOKKAIDO UNIVERSITY

Title	連続発声母音の自動認識
Author(s)	板坂, 政昭; Itasaka, Masaaki; 宮腰, 政明 他
Citation	北海道大學工學部研究報告, 122, 83-90
Issue Date	1984-07-31
Doc URL	https://hdl.handle.net/2115/41879
Type	departmental bulletin paper
File Information	122_83-90.pdf



連続発声母音の自動認識

板坂政昭** 宮腰政明* 新保 勝*
(昭和59年3月31日受理)

Automatic Recognition of Continuously Spoken Vowels

Masaaki ITASAKA, Masaaki MIYAKOSHI and Masaru SHIMBO
(Received March 31, 1984)

Abstract

An automatic speech recognition system of continuously spoken vowels is proposed based on the mathematical phonetical theory. In order to identify vowel phonemes in speech sounds, parameters on a projective vowel plane are calculated from their local peak frequencies. Some experiments are carried out for consecutive vowels, consonant-vowel utterances, etc.

1. ま え が き

近年の産業用ロボットの目覚ましい普及などにみられる各分野での作業の機械化と自動化に伴い、人間-機械系の通信手段として音声を利用する音声自動認識システムへの要請は高まっている。

人間と機械との情報の交換に音声を用いることは、文字や図形など人間のもつ他の情報伝達の方法に比べて高速度に情報を送り出すことができ、手や足を使う作業と平行して同時に使用することが可能であるという長所がある。

図1は音声自動認識を方式別に示したものである。現在実用化されている方式は100~200程度の単語を認識する特定話者単語方式¹⁾並びに同じく特定話者による離散発声を対象とした単音節認識方式^{2),3)}などがあるが、これらはその発声方式や語いの制約の面から何れも用途が限られている。

音声認識の目ざすところは、不特定多数の話者により自然に発声された連続音声の認識システムの実用化である。この観点からすると、少数の要素の配列により全ての音声の表現が可能であ

実現容易 ← → 研究の方向

◎発声方法	離散発声 → 連続発声
◎話者条件	特定話者 → 不特定話者
◎識別単位	単語単位 → 単音節単位 → 音素単位

図1 音声自動認識の分類

*) 情報工学専攻 情報処理工学講座

***) 現在 松下電器産業株式会社

る音素（例えば東京方言では24個⁴⁾）を言語情報の最小単位とし、連続音声の中の個々の音素を正しく認識することが最も一般的な音声認識システムの実現につながると考えられる。

不特定話者が発声した連続音声の自動認識を困難にする主な原因の一つは、話者の個人差による音声波の音響的特徴の変動であり、他には連続発声による隣接音の影響（いわゆる調音結合）を受けることである。これらの解決すべき問題点が残されている為、上記システムの実現には至っていない。調音結合に対処する方法としては、佐藤・藤崎による母音のホルマント周波数のターゲットを推定する AbS モデルや白井^{5),6)}らや石崎⁷⁾による調音レベルでのターゲット推定モデル及び粕谷・横田・秋沢⁸⁾による動的ボタンを用いたターゲット推定モデルがあるが、何れも個人差や調音結合を完全に除去するには至っていないように思われる。

本論文では音素認識の立場から、不特定話者による連続音声入力データを分析し、数理音声学理論⁹⁾に基づき、話者の影響を受け難く母音の分類・識別に有効とされる射影的母音平面^{9),10)}上での連続発声母音の挙動を調べ、音素単位の母音音声認識方式について考察する。

2. 射影的母音平面と音素識別楕円

2.1 射影的母音平面

母音を分析・識別する上での特徴パラメータとしては、声帯の振動が発声時の口腔の伝達特性により強調されて得られる周波数成分であるホルマント周波数が有効であることが知られている。

ホルマントのうち低次のもの、特に第1、第2ホルマントを各々座標軸とした平面 (F_1 - F_2 平面) 上で母音（日本語5母音）のおおまかな分類が可能であり^{4),11)} この二つのパラメータを用いて識別する試みもある⁹⁾ が、この平面上では話者や環境の差による変動のため、5母音の領域にはかなりの重複がある。 F_1 - F_2 平面上での個人差による変動は5母音の相対関係を保持したまま平行移動する傾向がみられ、この平面に第3ホルマント (F_3) からなる座標軸を加えたホルマント空間 (F_1 - F_2 - F_3 空間) 内での分布では5母音の重複がほぼ除去されることが報告されている^{11),12)}

母音の3ホルマントを用いて母音の識別を行なう一つの手法として射影的母音平面^{9),10)}が知られている。これは3ホルマント周波数 ($F_1 < F_2 < F_3$) の対数差

$$a = \log (F_2 / F_1)$$

$$b = \log (F_3 / F_2)$$

を座標軸として構成されている。この3ホルマント周波数の対数差を取ることは、ウェーバー・フェヒナー則に關係して母音の分化に役立つとともに、話者不特定化にも効果があると考えられる。3ホルマント周波数の対数差を取る方式としては他に、パラメータの組み合わせと係数の取り方に修正を加えたPAS方式などがあるが、識別に関しては大きな差はみられないことが報告されている¹³⁾ この対数差による話者不特定化への効果は、後の藤崎他の研究¹²⁾により母音空間内での個人差による変動が、この空間を極座標表示したr方向に分布することからも確認される。射影的母音平面上での母音分布を図2に示す。これは男性話者4名、女性話者2名によるもの¹⁰⁾で、各母音を囲っている楕円は便宜的に描いたものである。

2.2 射影的母音平面上におけるホルマントの動き

日本語5母音間の調音結合による動きをみるために、次

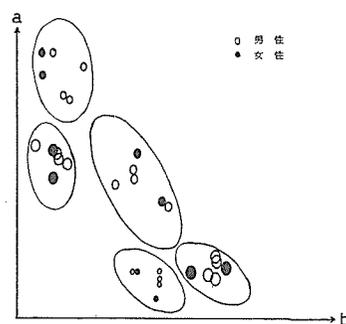


図2 射影的母音平面上の母音の分布

の四つの連続母音パターンを発声し、母音平面上での分布を調べた。

- (1) /aeioua/
- (2) /auoiea/
- (3) /aiueoa/
- (4) /aoeuia/

この四つのパターンにより全ての隣接音間の組み合わせが可能となる。⁴⁾ 図3は孤立発声時における特定男性話者の母音平面上の5母音の分布である。1点が1フレーム(29.3 ms, 1/2フレームのシフト)に対応し、約1秒間の発声による約70フレーム分がプロットしてある。図4は同じ

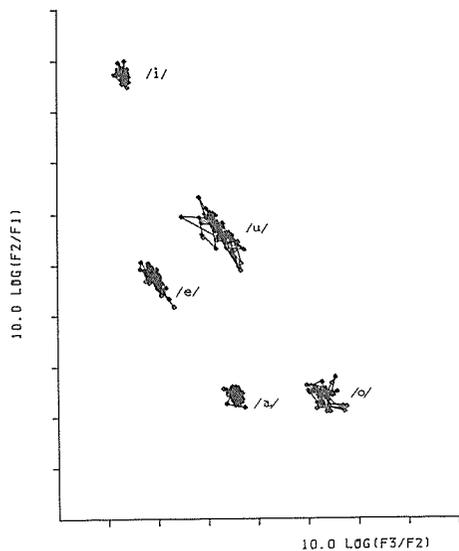


図3 孤立発声による母音分布

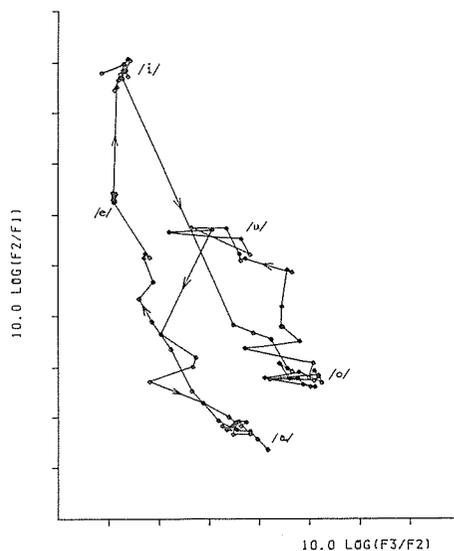


図4 /aeioua/

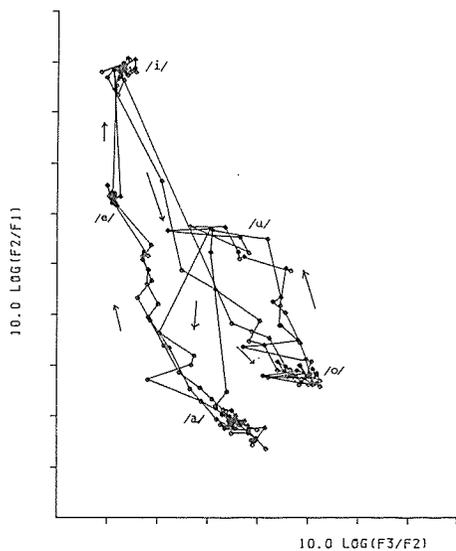


図5 /aeioua/+ /aeioua/

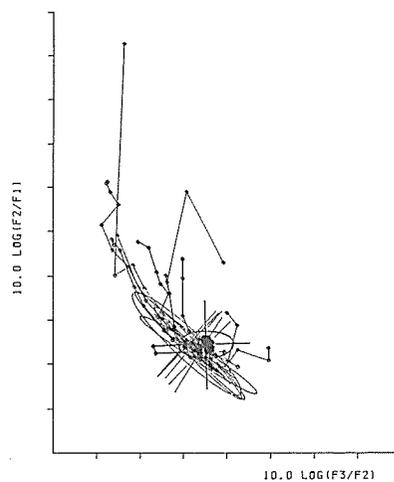


図6 /a/の調音結合による動き

話者の(1)の 패턴の発声で、/e/が後続母音/i/の影響を受けて途中から/i/に近い位置に分布しているものの、概して連続発声による大きな変動はみられない。図5は(1)の発声パターンを2回繰り返したものを同一平面上にプロットしたものであるが、同一話者に関しては同一の発声パターンに対する軌跡がほぼ一致することが確認される。さらに同じ話者の他のパターン(2)~(4)の発声結果から、/a/の始点と終点に若干の違いがあり、また後続母音により途中の経路に違いがみられるが、5母音の位置の変動はほとんどないことが確認される。

図6は話者Aの/a/に関する全ての連母音間の組み合わせによる動きと孤立発声による資料をプロットしたものである。図から/a/の近傍では隣接音が何であれ、ほぼ同じ方向に動くことが確かめられる。このことは他の母音についても予想される。

以上のデータをもとに、次節では各母音の領域を囲む音素識別楕円を構成する。

2.3 音素識別楕円

前節で調べた各母音の定常状態及び連続発声による動きを考慮して母音平面上に識別楕円を構成する。母音により、1~3名の発声による30~100フレーム分のデータを用い、統計的手法を利用して各母音の識別楕円の位置と大きさを決めている。

楕円を決定するにあたっては各母音の領域をできるだけ含み、しかも各楕円が交わらないことを目的としているが、母音によりバラツキ方に違いがあり、そのため楕円の大きさは母音ごとに条件が異なっている。

ここで識別の手段として楕円を導入することは、図6にみられる分布の様子などから、色におけるマッカダムの識別楕円¹⁵⁾にならったもので音の知覚現象における音素の二次形式表示の可能性を実験的に確かめることができる。

実際に構成した識別楕円とその作成に用いた音声データの分布を図7に示す。

3. 認識システムの概要

本研究で用いた認識システムの詳細については文献16)にゆずり、ここでは構成図だけを図8に示す。音声データはあらかじめテープデッキに録音したものをを用い、8.733kHzでサンプリング、12bitで量子化をしたのち、マイクロコンピュータシステム(Cromemco, M 68000)でソフトウェア的に処理している。装置の構造上、7.2秒間の音声データを一括して処理しており、

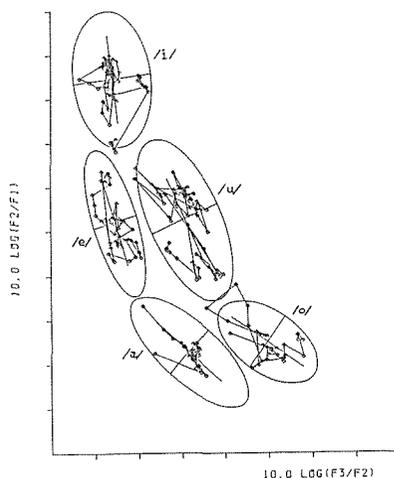


図7 識別楕円に用いた音声データ

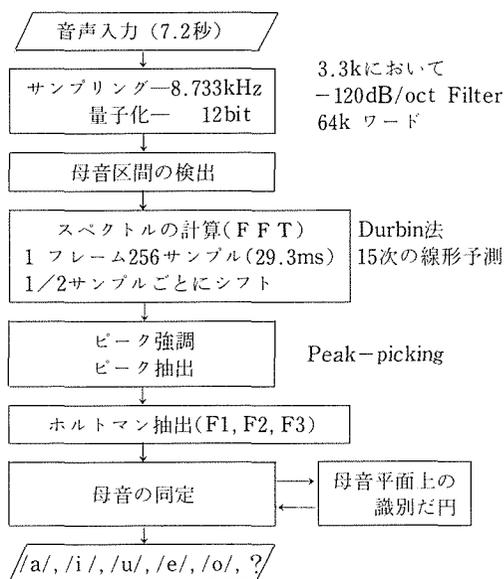


図8 システム構成

処理時間は約 30 分である。

4. 識 別 実 験

本研究で構成したシステムを用いて識別実験を行なった。実験に用いた音声資料は図 9 に示す通りである。

- 1) 連母音は成人男性話者 3 名が(1)~(4)のパターンを各 2 回発声 (/ a / × 48, 他は 24)
- 2) 子音母音型 (CV 型) 発声はカ行 (破裂音) サ行 (摩擦音) マ行 (通鼻音) 各々につき上記 4 パターンを成人男性者 1 名が各 2 回発声 (/ a / × 45, 他は 24)
- 3) 地点名は CV 型と同一の話者が各 5 回発声 (/ a / × 35, 他は 25)

なお地点名については 5 母音の個数ができるだけ平等になるように抽出している。

5. 実験結果と考察

- ◎ 連母音のパターン
 - (1) / a e i o u a /
 - (2) / a u o i e a /
 - (3) / a i u e o a /
 - (4) / a o e u i a /
- ◎ CV 型発声
 - (1) カ行 …………… 破裂音
 - カケキコクカ
 - カクコキケカ
 - カキクケコカ
 - カコケクキカ
 - (2) サ行 …………… 摩擦音
 - サセシソスサ
 - サスソシセサ
 - サシスセソサ
 - サソセスシサ
 - (3) マ行 …………… 通鼻音
 - マメミモムマ
 - マムモミメマ
 - マミムメモマ
 - マモムメミマ
- ◎ 地点名 (8 地点)
 - (1) 北大 [h o k u d a i]
 - (2) 愛媛 [e h i m e]
 - (3) 箱根 [h a k o n e]
 - (4) 岩手 [i w a t e]
 - (5) 松江 [m a t s u e]
 - (6) 小樽 [o t a r u]
 - (7) 静岡 [s h i z u o k a]
 - (8) 徳島 [t o k u s h i m a]

図 9 識別実験に用いた音声資料

5.1 実験結果 (表 1~表 3)

表は横軸が入力で縦軸が認識結果を表わす。なお、出力のうち、(?) は欠落誤りを意味する。また、誤り例の欄の下側の数字はフレーム数を示す。例えば、/(i)a/ が/(i)e a/ と出力された例では、/i/, /e/, /a/ のフレーム数が各々 5, 5, 11であることを示す。

表-1 連母音による識別実験(成人男性話者3名 表中下段はフレーム数を表わす)

OUT IN	a	i	u	e	o	(i)uo	(a)ei	(e)uo	uiu	iui	?
						(o)ui	(i)ea	(o)ue			
a	46						1 5-5-11				1
i		20				2 24-3-21 24-4-27	1 23-3-28			1 12-3-5	
u			22						2 4-6-6 3-4-5		
e		1		22				1 23-5-22			
o			2		18	2 20-6-18 16-8-11		2 29-5-0 29-4-4			

表-2 C V型の連続発声の認識結果(成人男性話者1名 表中Mはマ行, Sはサ行を表わす)

OUT IN	a	i	u	e	o	iu	ue	eue	o eo	ou	eo	?
							eu				oe	
a	45											
i		22				2 M19-1 M14-3						
u			23				1 M12-4					
e				20			2 M28-4 S17-4	1 S 4-5-7				1 s
o					18				1 S 7-3-5	1 S14-5	4 M	

表-3 地点名(8地点)の認識結果(成人男性話者1名)

OUT IN	a	i	u	e	o	ua	ae	ou	ei	oe	?
a	23					7	5				
i		24									1
u			23					1			1
e	1			22					1		1
o			2		19			2		1	1

(1) 連母音での認識誤りでは挿入誤りがほとんどを占め、しかもわたり音部が3フレーム以上続いた為に起きたと思われる例が16の誤り例のうち8例ある。

(2) CV型発声では破裂子音(カ行)は100%識別され、摩擦音(サ行)での誤り例が5例(表中S表示)、通鼻音(マ行)での誤り例(表中M表示)が8例ある。

(3) 地点名に関しては欠落誤りが連母音やCV型に比べて多く、また/u/, /o/に関係した挿入誤りが多くみられる。

5.2 考察

(1) わたり音部に現われる挿入誤りに関しては表や図10(/a i/→/a e i/と出力された例)にみられるように、わたり音部で誤認識された母音は前後の母音に比べてフレーム数が少ない(図10では/a/, /e/, /i/が各々23, 3, 28)ことから、書き換え規則⁹⁾などを適用することにより解決されると思われる。今回の実験ではデータが少なく、また連続発声による動きを調べることに主眼をおいた為、この種の規則は適用しなかったが、今後は検討する必要があると思われる。

(2) 図11の/hakone/の動きにみられるように母音から母音に直接移らないで大きく迂回して5回とも同じ軌跡を描くことは、迂回している部分に何らかの子音(ここでは/k/, /n/)の情報が含まれているのではないかと考えられる。この軌跡の特徴は、同時に実験を行なったもう一名の話者にもみられることが確かめられている。

このわたり音の方向から子音を推定する試みについては伊達¹⁴⁾による無声破裂子音に関する研究などがあるが、¹⁷⁾他の多数の話者についても同様の軌跡を描くことが確認できれば、さらに考察する必要があると思われる。

なお、本研究の一部は文部省科学研究費(課題番号57460207)の援助によって行なわれたことを付記する。

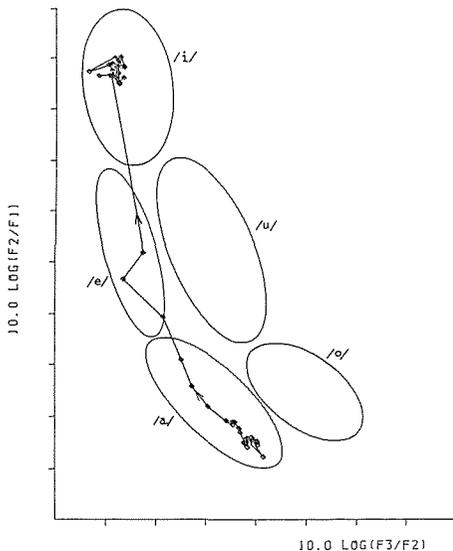


図10 /a i/が/a e i/と誤認識された例

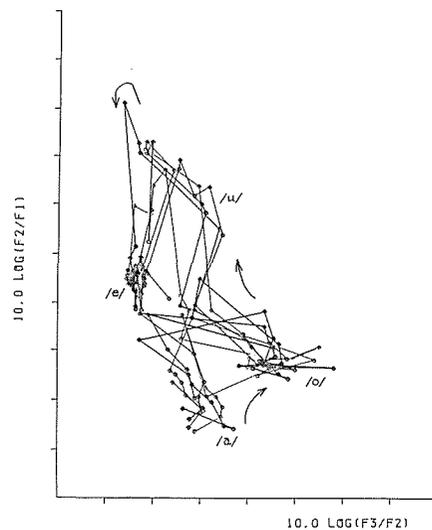


図11 /hakone/の5回の発声の母音平面上での動き

文 献

- 1) 齊藤収三：電子通信学会誌, 57-08 (1982), 514.
- 2) 吉田・迫江・千葉：音響学会研究発表会講演論文集, 3-2-6 (昭54-06), 607.
- 3) 似鳥寧信：北海道大学学位請求論文 (1983).
- 4) 新美康永：音声認識, 共立出版 (1979).
- 5) 白井・菅田：電子通信学会論文誌, Vol. 61-A, No. 5 (1978), 409.
- 6) 白井・小林・北村：音響学会音声研究会資料, S81-18 (1981).
- 7) 石崎 俊：電子通信学会論文誌, Vol. 61-A, No. 10 (1978), 956.
- 8) 柏谷・横田・秋沢：音響学会音声研究会資料, S81-17 (1981).
- 9) 近藤一夫：数理音声学序説, 東京大学出版会 (1964).
- 10) M. Iri: RAAG Memoirs, 3 (1962), 551.
- 11) 齊藤収三・中田和男：音声情報処理の基礎, オーム社 (1982).
- 12) 佐藤・藤崎：音響学会誌, 34 (1978), 177.
- 13) S. Itahashi and S. Yokoyama: パターン情報システム調査報告, PIPS-R, No. 8 (1974), 34.
- 14) T. Date: RAAG Memoirs, 4 (1968), 522.
- 15) D. L. MacAdam: Proc. of the I. R. E., 5 (1951), 468.
- 16) 板坂政昭：北海道大学大学院工学研究科情報工学専攻修士論文 (1984).
- 17) 新保・板坂・宮腰：北海道大学工学部研究報告, 119 (1984), 37.