線型モデルによる母音連鎖中の母音の認識

線型モデルによる母音連鎖中の母音の認識

高良富夫* 今井　聖**

A Vowel Recognition Method for a Sequence of Vowels Based on a Linear Model.

Tomio TAKARA*, Satoshi IMAI**

Summary

Coarticuration of a sequence of vowels is normalized by a linear model of the coarticulation. An auditory response of V_0 in $V\ V_0\ V$ type connected vowels is modeled as a linear function of the acoustical features of the preceeding V, V_0, and the following V. The coefficients of the function are determined by the method of least squares using training data. The error rate of the recognition test using the model is lower by 7.5% than that of the test not using the model.

1　まえがき

我々が会話を行うとき，離散的な音韻の系列を発声しようと意図する。しかし，発声された音声は，発声器官の物理的制約のため，連続的なものとなる。この結果，連続音声中の音韻は，その前後の音韻の影響を強く受け，物理的特性（例えば周波数スペクトル）が，単独に発声された音韻（単音節など）の物理的特性から著しくかけ離れたものとなる。すなわち，同一音韻カテゴリの中の音声であるにもかかわらず，前後の音韻環境の違いにより異なる物理的特性を示したり，逆に，異なる音韻カテゴリに属すべき音声が，ほとんど相似の物理的特性を示したりする。このような現象は調音結合効果と呼ばれ，連続音声を機械で自動認識する場合，大きな障害となる。

一方，人間が連続音声を聴取する場合は，調音結合は，それほど問題にならない。この理由は，我々の会話においては，話題の前後関係や文法的な前後関係の知識によって，不明瞭になった音声の物理特性が補償されているためであると考えられる。しかし，さらに基本的な段階，すなわち聴覚機能の初期の段階においても，このような補償がなされていることが考えられる。

例えば，前と後に同一の母音があるような母音の連鎖 $V\ V_0\ V$ について，V_0 を切り出して聴取実験を行うと，正聴率は75%程度であるが，$V\ V_0\ V$ を聴取し V_0 を同定すると，正聴率が20%程度向上する。[1] このことは，人間が連続音声中の1つの母音を認識する場合，少なくともその母音の直前・直後からの，話題や文法上の情報ではなく，音韻情報を利用していることを意味している。

ここでは，人間におけるこのような音声認識の特性を模擬する機能的なモデルを提案し，音声自動認識に応用することを試みる。同様のモデルは，ホルマント

受付：1984年 4 月24日
*琉球大学工学部電子・情報工学科
**東京工業大学精密工学研究所

周波数を特徴パラメータとして桑原[1]が，又，調音パラメータについて石崎[2]がすでに提案している。しかし，いずれの場合も，パラメータ成分が独立に前後から影響を受けるとしたものであり，パラメータ成分の相互の干渉を考慮していない。又，使用したパラメータは，それを抽出するために複雑な処理を必要とし，良い精度で安定に抽出することが困難である。これに対し，ここで提案するモデルは，特徴パラメータ成分の相互の干渉をも考慮しているので，前後の音韻情報を利用するという意味で，より精度の高いモデルである。又ここで使用するパラメータは，音声スペクトルから簡単に得られ，かつ安定に抽出できる。

2　調音結合の線型モデル

2－1　モデル

一般に音節が連接して連続音声になるものとし，連続音声を

$$S^{(1)}S^{(2)}\cdots S^{(n-1)}S^{(n)}S^{(n+1)}\cdots S^{(N)} \tag{1}$$

と表記する。ここでNは音節数である。$S^{(n)}$は注目している時点nにおける音節の音響特性であり，ベクトルで表現できるものとする。ここでは連続音声の例として母音連鎖について検討する。

連続音声から$S^{(n)}$の部分だけを切り出して聴取すると，発声者の意図した母音に同定できない場合がある。しかしこの場合，前後の母音とともに$S^{(n)}$を聴取すると正しく同定できることが多い。このことから，時点nの母音の聴覚心理的特性Rは，その時点の母音の音響特性$S^{(n)}$と，その前後の母音の音響特性$S^{(n-1)}$および$S^{(n+1)}$の関数であると考えることができる。すなわち，

$$R = f(S^{(n-1)},\ S^{(n)},\ S^{(n+1)}) \tag{2}$$

である。ここでは，関数fとして，簡単のため，線型関数

$$R = MS + C \tag{3}$$

を仮定する。但し，

$$R = (R_1, R_2, \cdots, R_i, \cdots, R_I)^T, \tag{4}$$

$$S = \begin{bmatrix} S^{(n-1)} \\ S^{(n)} \\ S^{(n+1)} \end{bmatrix}, \tag{5}$$

$$M = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_i \\ \vdots \\ m_I \end{bmatrix}, \tag{6}$$

$$C = (C_1, C_2, \cdots, C_i, \cdots, C_I)^T, \tag{7}$$

$$S^{(l)} = (S_1^{(l)}, S_2^{(l)}, \cdots, S_i^{(l)}, \cdots, S_I^{(l)})^T,$$
$$l = n-1,\ n,\ n+1 \tag{8}$$

$$m_i = (m_{i,1}, m_{i,2}, \cdots, m_{i,3I}),$$
$$i = 1, 2, \cdots, I \tag{9}$$

である。但し，Tは転置を表し，Rと$S^{(l)}$とCは縦ベクトル，m_i，$i = 1 \sim I$は横ベクトルである。式(3)のベクトルの第i成分だけを書き出すと，

$$R_i = m_i S + C_i \tag{10}$$

である。

式(10)の係数m_iとC_iは，学習用パタンを用いて，モデルによる予測の2乗誤差が最小となるように決定する。これは次のように行う。まずm_iとC_iをまとめて，ひとつのベクトルで表記する。

$$b_i = (m_i, C_i)^T \tag{11}$$

第j番目の，学習用パタンと目標値をそれぞれ${}^jS$，${}^jR_i$，$j = 1 \sim J$とおき，全学習用パタンと目標値をそれぞれまとめて，

$$X = ({}^1S, {}^2S, \cdots, {}^jS, \cdots, {}^JS)^T \tag{12}$$

$$Y = ({}^1R_i, {}^2R_i, \cdots, {}^jR_i, \cdots, {}^JR_i)^T \tag{13}$$

とおくと，b_iは

$$\hat{b}_i = (X^T X)^{-1} X^T Y \tag{14}$$

で与えられる。これを

$$(\hat{m}_i, \hat{C}_i)^T = \hat{b}_i \tag{15}$$

とする。

3つの時点$n-1$，n，$n+1$の音響特性がそれぞれ$\tilde{S}^{(n-1)}$，$\tilde{S}^{(n)}$，$\tilde{S}^{(n+1)}$と観測されたとき，時点nの聴覚心理的特性の予測値$\tilde{R} = (\tilde{R}_1, \tilde{R}_2, \cdots, \tilde{R}_i, \cdots, \tilde{R}_I)$は，

$$\tilde{R}_i = \hat{m}_i \tilde{S} + \hat{C}_i \tag{16}$$

で与えられる。但し，

$$\tilde{S} = \begin{bmatrix} \tilde{S}^{(n-1)} \\ \tilde{S}^{(n)} \\ \tilde{S}^{(n+1)} \end{bmatrix} \tag{17}$$

である。$\tilde{R}$を用いて音韻を決定する。

2－2　母音空間パラメータ

音響特性 $S^{(t)}$ として，ここでは音声スペクトルから簡単に得られ，しかも音声スペクトルを良く表現し，次元数が少ないパラメータを次のように導入する。

スペクトルをベクトルで表現し，時点 t の音声スペクトルを $F^{(t)}$，母音クラス i の単母音の平均スペクトルを $\bar{V}_i$ とすると，そのパラメータの第 i 成分は

$$\left.\begin{aligned} S_i^{(t)} &= \frac{\bar{V}_i^{\top} F^{(t)}}{\|\bar{V}_i\| \cdot \|F^{(t)}\|} \quad , \\ i &= 1 \sim 5 \quad (/i/, /e/, /a/, /o/, /u/), \\ t &= n-1, \ n, \ n+1 \end{aligned}\right\} \quad (18)$$

で与えられる。但し，$\|\cdot\|$ はベクトルのノルムを表す。すなわち，このパラメータは，各平均母音スペクトル(ベクトル)方向への $F^{(t)}$ の方向余弦を成分としており，音声の特徴を母音空間上でながめたものといえる。このことから，これを母音空間パラメータと呼ぶことにする。一般に，連続音声は音節が連接したものであり，音節は，母音を中心にして構成されている。従って，母音空間パラメータは，連続音声の特徴をよく表現すると考えられる。しかもこのパラメータの次元数は5次元であり，比較的少ない。母音空間パラメータを用いた音声分析の例を図1に示す。この図は，音声資料 /aia/, /aea/, /aoa/, /aua/ を母音

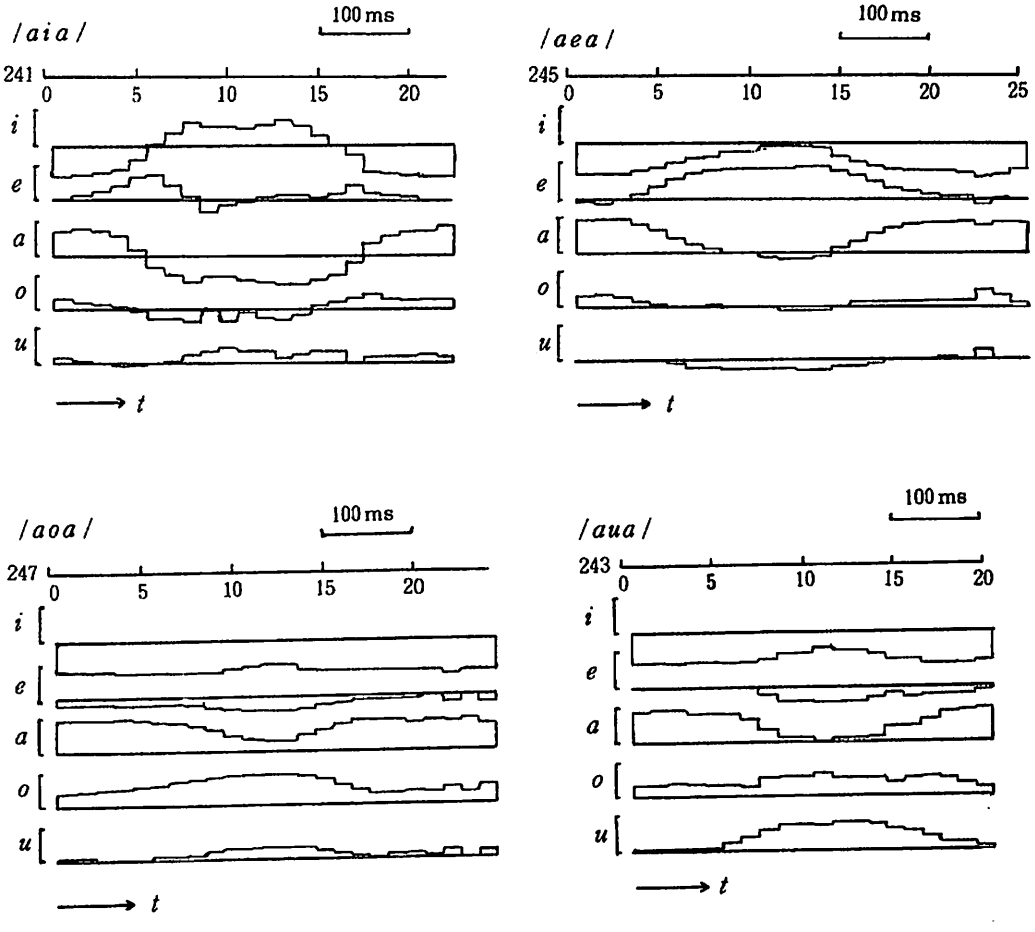

Fig 1. Examples of Vowel Space Parameter.

空間パラメータで分析し，パラメータの各成分を時系列パタンとしてプロットしたものである。

音声スペクトル $\boldsymbol{F}^{(i)}$ として，ここでは20次元のメル対数スペクトルを用いた。メル対数スペクトルとは，DFTパワースペクトル（10 kHzサンプリング，12ビット量子化，256点ブラックマン時間窓，FFTによる）をメル周波数尺度上で20等分割し，スペクトル値を各帯域内で平均し，これに帯域幅を乗じ，その値を対数化したものである。

単母音の平均スペクトルとして，ここでは，成人男性話者36名の発声した単母音のメル対数スペクトルの加算平均を用いた。

2－3　モデルの妥当性

連続音声の例として，対称形3連母音 $V\ V_0\ V$ をとりあげ，中央部の母音 V_0 を注目する時点の音韻として，上述のモデルを検討する。

図1に例示したような母音空間パラメータの時系列パタンと音声波形とを参照して，対称形3連母音の3つの母音中心を視察により検出する。前の V と，V_0，後の V の母音中心の母音空間パラメータ1フレーム分ずつをそれぞれ $\boldsymbol{S}^{(n-1)}$，$\boldsymbol{S}^{(n)}$，$\boldsymbol{S}^{(n+1)}$ とする。

成人男性話者1名が各2回発声した対称形3連母音40個を学習用資料として，式(12)の X を構成する。（$J=40$）目標値

$$ {}^{j}\boldsymbol{R} = ({}^{j}R_1, {}^{j}R_2, \cdots, {}^{j}R_i, \cdots, {}^{j}R_I) \qquad (19) $$

としては，V_0 の属すべき母音クラスの単母音スペクトルの母音空間パラメータを用いた。これらを用いて，式(14)と式(15)とにより $\hat{m}_i$ と $\hat{C}_i$ を求める。

式(16)の $\tilde{S}$ に学習用資料の値を代入した結果を表1に示す。

表の第1列は資料番号 j，第2列は目標値 ${}^{j}R_i$，第3列は式(18)の予測値 $\tilde{R}_i$ であり，第4列は予測誤差

$$ \varepsilon_j = \tilde{R}_i - {}^{j}R_i \qquad (20) $$

である。但し，このデータは母音空間パラメータの第3成分（$i=/a/$ に対応する成分）である。

予測誤差の2乗平均値 $\overline{\varepsilon_j^2}$ は0.0044であり，これは，目標値の2乗平均値 $\overline{{}^{j}R_i^2}=0.2678$ に比較して十分小さい。このことから，この例では，線型モデルでかなり良い予測値が得られているといえる。

学習用資料以外の資料に対するこのモデルの性能は，認識実験によって評価する。

Table 1　Example of the Predicted Value and the Residue Using the Linear Model. (Speaker YI, $i=/a/$)

$\overline{\varepsilon_j^2}=0.0044$　　$\overline{{}^{j}R_i^2}=0.2678$

資料番号 j	目標値 ${}^{j}R_i$	予測値 $\tilde{R}_i$	予測誤差 ε_j
1	−.7655	−.7416	−.0239
2	−.7655	−.7026	−.0629
3	−.0652	−.0862	.0210
4	−.0652	−.0578	−.0074
5	−.4455	−.3549	−.0907
6	−.4455	−.4097	−.0359
7	.1491	.0203	.1288
8	.1491	.0590	.0901
9	.7464	.7883	−.0419
10	.7464	.7763	−.0299
11	−.0652	−.0527	−.0125
12	−.0652	−.0410	−.0242
13	−.4455	−.5141	.0685
14	−.4455	−.6253	.1797
15	.1491	.2121	−.0631
16	.1491	.2220	−.0729
17	.7464	.8242	−.0778
18	.7464	.7151	.0313
19	−.7655	−.7495	−.0160
20	−.7655	−.7492	−.0163
21	−.4455	−.4525	.0070
22	−.4455	−.4915	.0460
23	.1491	.1500	−.0009
24	.1491	.1859	−.0368
25	.7464	.7536	−.0072
26	.7464	.6249	.1214
27	−.7655	−.6621	−.1034
28	−.7655	−.7299	−.0356
29	−.0652	−.1299	.0647
30	−.0652	−.1017	.0365
31	.1491	.1386	.0105
32	.1491	.1926	−.0435
33	.7464	.7769	−.0305
34	.7464	.6293	.1171
35	−.7655	−.7980	.0325
36	−.7655	−.8499	.0844
37	−.0652	−.0484	−.0168
38	−.0652	−.0728	.0076
39	−.4455	−.3498	−.0957
40	−.4455	−.3441	−.1014

3　認識実験

モデルの有効性を認識実験によって比較検討するため，まず，モデルを用いないで，スペクトル・マッチングにより認識実験する。次に，話者の単母音を目標値として，モデルを用いる認識実験を行い，モデルの有効性を示し，又，式(3)の行列Mの成分のうち有効なものを選択する。さらに，多数話者の平均母音を目標値として，簡単化したモデルおよび強制変数ありの変数増減法についての検討を行い，最後に，5名の話者の資料にモデルを適用し認識する。

3－1　スペクトル・マッチングによる認識

成人男性話者2名（話者YI，話者HA）が各2回発声した対称形3連母音の中央部の母音80個を認識対象とする。

中央部の母音中心を視察により検出し，その時点での母音空間パラメータを抽出する。母音空間パラメータの各成分$S_i^{(n)}$，$i=1\sim5$を比較し，

$$S_k^{(n)} = \max_{1\le i\le 5} S_i^{(n)} \qquad (21)$$

であるkに対応する母音クラスを認識結果とする。式(18)に示したように，$S_i^{(n)}$は，観測点のスペクトルと母音クラスiの平均スペクトルとの内積であるから，式(21)により，平均スペクトルを参照パタンとする認識を行うことができる。このとき，マッチングの測度は，式(18)ということになる。

この実験の結果，誤認識の数は，話者YIについては4，話者HAについては8の計12であった。ただしここでは，全入力数は80である。

3－2　単母音を目標値とする認識

話者2名のうち1名の音声を学習用資料とし，他方の話者の音声と学習用資料を認識する。学習用の話者を入替えて，これを2回行う。

(1)　基本重回帰分析

式(10)から式(15)までの手順は通常の重回帰分析と同じであるから，これを基本重回帰分析と呼ぶことにする。

対称形3連母音の3つの母音の母音中心を視察で検出し，その3点の母音空間パラメータを抽出し，それぞれを$S^{(n-1)}$，$S^{(n)}$，$S^{(n+1)}$とする。式(11)～式(15)により学習を行う。このとき，目標値$'_R$としては，学習用話者の単母音（単独発声の母音）の母音空間パラメータを用いる。式(19)で与えられる予測値$\tilde{R}$の成分を比較し，

$$\tilde{R}_k = \max_{1\le i\le 5} \tilde{R}_i \qquad (22)$$

であるkに対応する母音クラスを認識結果とする。

認識実験の結果を表2に示す。学習用資料を認識したときをclosed test，他方の話者の資料を認識したときをopen testと呼び，それぞれを集計して併せて示した。本実験の全入力数は，スペクトル・マッチングの実験の2倍である。スペクトル・マッチングの方法と比較すると，この方法では誤認識数が約半減している。

Table 2　Results of the Recognition Test Using the Isolated Vowels for the Targets.(a).

基本重回帰分析

誤認識数（カッコ内は全入力数）

学習用話者	入力用話者 YI	入力用話者 HA	
			closed test　6（80）
YI	2（40）	3（40）	open test　7（80）
HA	4（40）	4（40）	計　13（160）

(2)　変数増減法[3]

線型モデル

$$R_i = m_i S + C_i \qquad (23)$$

(10)

において，ベクトルSの各成分は，重回帰分析の用語では説明変数であり，R_iは目的変数である。説明変数の候補がたくさんあるとき，目的変数を最もよく説明する（又は予測する）変数の組合せを探す手法は，一般に「変数選択の問題」と呼ばれている。

Sの成分として，対称形3連母音の母音空間パラメータを使用すると，説明変数の数は15であるから，式

$$R = MS + C \qquad (24)$$

((3))

の行列Mの成分の数は75（＝15×5）である。ここで行列Mの成分の数を減少させることを考える。式(23)に変数増減法（詳細は文献(3)を参照）を適用し，説明変数を選択する。選択されなかった変数については，その変数の係数（ベクトルm_iの成分）を0とおく。

Table 3 Example of the Predicted Value and the Residue Using the Variable Increase－Decrease Method. (Speaker YI, $i=/a/$)

$\overline{\varepsilon_j^2}=0.0047$ $\overline{{}^jR_i^2}=0.2678$

資料番号 j	目標値 ${}^jR_i$	予測値 $\tilde{R}_i$	予測誤差 ε_j
1	−.7655	−.7512	−.0143
2	−.7655	−.7203	−.0452
3	−.0652	−.0790	.0138
4	−.0652	−.0819	.0167
5	−.4455	−.3429	−.1027
6	−.4455	−.4193	−.0263
7	.1491	.0214	.1277
8	.1491	.0510	.0980
9	.7464	.8221	−.0757
10	.7464	.7716	−.0252
11	−.0652	−.0515	−.0137
12	−.0652	−.0296	−.0356
13	−.4455	−.5189	.0733
14	−.4455	−.6217	.1762
15	.1491	.2228	−.0737
16	.1491	.2178	−.0687
17	.7464	.8393	−.0929
18	.7464	.7253	.0211
19	−.7655	−.7411	−.0244
20	−.7655	−.7384	−.0271
21	−.4455	−.4599	.0144
22	−.4455	−.4914	.0459
23	.1491	.1593	−.0103
24	.1491	.2023	−.0532
25	.7464	.7016	.0447
26	.7464	.6127	.1337
27	−.7655	−.6761	−.0894
28	−.7655	−.7470	−.0185
29	−.0652	−.1627	.0975
30	−.0652	−.0943	.0291
31	.1491	.1279	.0212
32	.1491	.1575	−.0084
33	.7464	.7753	−.0289
34	.7464	.6427	.1037
35	−.7655	−.7789	.0134
36	−.7655	−.8241	.0586
37	−.0652	−.0381	−.0271
38	−.0652	−.0534	−.0118
39	−.4455	−.3323	−.1133
40	−.4455	−.3430	−.1026

これを $i=1$ ～5について行う。変数選択の基準は，$Fin_1=Fout_1=0.2$，$Fin_2=Fout_2=2.0$ とする。

変数増減法により変数を選択し，選択された説明変数だけを用いて予測した例を表3に示す。この例は，表1と同じ話者YIの $i=3$ の成分（/a/ に対応）である。説明変数は15から4に減少した。誤差の2乗平均値は0.0047であり，表1の結果とほとんど変わらない。

認識実験の結果を表4に示す。行列Mの成分の数は，話者YIの音声で学習した場合，75から25へと減少し，HAで学習した場合は32へと減少した。誤認識の数は，基本重回帰分析の場合に比較して，closed testでは1だけ増加しているがopen testでは逆に1だけ減少しており，総数は同じである。

Table 4 Results of the Recognition Test Using the Isolated Vowels for the Targets.(b).

変数増減法

誤認識数（カッコ内は全入力数）

学習用話者	入力用話者 YI	入力用話者 HA	
			closed test 7 (80)
YI	2 (40)	3 (40)	open test 6 (80)
			計 13 (160)
HA	3 (40)	5 (40)	

行列Mの成分の数 YIで学習：25，HAで学習：32

以上の実験結果から，提案したモデルは，注目する時点の情報だけを用いるスペクトル・マッチングの方法より有効であるといえる。又，変数増減法により行列Mの成分数を半分以下にしても，モデルの性能はほとんど変わらないことがわかる。

次に平均母音を目標値とする認識実験を行い，認識率がさらに向上することを示し，又，行列Mの成分の意味を考察する。

3－3 平均母音を目標値とする認識

3－2節では，目標値 ${}^jR$ として，学習用話者の単母音の母音空間パラメータを用いたが，ここでは，多数話者の単母音平均スペクトルの母音空間パラメータを用いる。すなわち，母音クラス k の平均スペクトル

を $\bar{V}_k$ とすると，目標値の成分 R_i は，

$$R_i = \frac{\bar{V}_i^T \bar{V}_k}{\|\bar{V}_i\| \cdot \|\bar{V}_k\|} \qquad (25)$$

で与えられる。ここで $\bar{V}_k$ は，成人男性話者36名が各1回発声した単母音のメル対数スペクトルの加算平均である。

(1) 変数増減法

3－2(1)節で述べた方法と同様の方法で認識する。但し，ここでは平均母音を目標値とする。

認識実験の結果を表5に示す。これと表4を比較すると，誤認識数が著しく減少していることがわかる。

Table 5　Results of the Recognition Test Using the Averaged Vowels for the Targets.(a).

変数増減法

誤認識数（カッコ内は全入力数）

学習用話者	入力用話者 YI	入力用話者 HA		
			closed test	2 (80)
			open test	2 (80)
YI	0 (40)	2 (40)	計	4 (160)
HA	0 (40)	2 (40)		

行列Mの成分の数　YIで学習：30，HAで学習：39

この理由は次のように考えられる。すなわち，平均母音を目標値とする方法では，中央部母音の属すべき母音クラスに対応する ^jR の成分が最大値1になり，モデルの誤差に対する目標値の比が，単母音を目標値とする方法より大きくなるからである。^jR のその成分が1となること，及びそれが，取り得る値の最大値であることは，式(25)から理解できる。

(2) 簡単化したモデルによる方法

変数増減法では，統計的な基準値であるF基準値[3]により変数を選択したが，選択された変数の解釈は困難であった。ここでは，モデルを次のように簡単化し，行列Mの成分について考察する。

音響特性 $S^{(n-1)}$，$S^{(n)}$，$S^{(n+1)}$ のそれぞれの第 i 成分だけが心理的特性（目標値）R の成分 R_i に寄与するものとする。従ってモデルは，式(10)の代わりに，

$$R_i' = m_{i,i} S_i^{(n-1)} + m_{i,i+5} S_i^{(n)} + m_{i,i+10} S_i^{(n+1)} + C_i \qquad (26)$$

となる。このモデルの係数 $m_{i,j}$，$j = i, i+5, i+10$ は石崎が調音パラメータで導入した強調係数[2]と等価である。但し，石崎は係数の値を適当に与えたが，ここでは，$m_{i,j}$ と C_i を，式(11)～式(15)と同様にして，最小2乗法（基本重回帰分析）により決定する。

そして，

$$R_k' = \max_{1 \le i \le 5} R_i' \qquad (27)$$

である k に対応する母音クラスを認識結果とする。

認識実験の結果を表6に示す。この方法では，行列Mの成分の数は15であるが，誤認識の数は，変数増減法に比較して1増加しただけである。

Table 6　Results of the Recognition Test Using the Averaged Vowels for the Targets.(b).

簡単化したモデル

誤認識数（カッコ内は全入力数）

学習用話者	入力用話者 YI	入力用話者 HA		
			closed test	3 (80)
			open test	2 (80)
YI	0 (40)	2 (40)	計	5 (160)
HA	0 (40)	3 (40)		

行列Mの成分の数　15

(3) 強制変数のある変数増減法

変数増減法では，まず数組の変数（強制変数）を人為的に選択し，さらにその他の変数を統計的に（ここではF基準値で）選択することができる。

ここでは，上述の「簡単化したモデル」で使用した変数を強制変数として変数増減法を適用する。学習および認識の方法は前述の変数増減法と同様とする。

認識実験の結果を表7に示す。この結果を表5と比較すると，行列Mの成分の数は1～3だけ増加したが，誤認識数は1減少したことがわかる。

Table 7　Results of the Recognition Test Using the Averaged Vowels for the Targets.(c).

強制変数のある変数増減法
誤認識数（カッコ内は全入力数）

学習用話者	入力用話者 YI HA	
		closed test 1 (80)
YI	0 (40)　2 (40)	open test 2 (80)
		計 3 (160)
HA	0 (40)　1 (40)	

行列Mの成分の数　YIで学習：33，HAで学習：40

以上の実験結果から，行列Mの成分のうち「簡単化したモデル」の成分が大きな役割を演じ，他の成分を加えることによりモデルの性能をさらに高めることができるということがいえる。

3－4　話者5名の音声資料に対する認識実験

成人男性話者5名の音声資料について認識実験を行った。5名のうち2名はこれまでの実験でも使用した話者である。まず，2名の話者の音声で学習を行い，他の3名の音声と学習用資料とを認識する。次に，後者3名の音声で学習を行い，前者2名の音声と学習用資料とを認識する。音声資料は各話者につき40個であるから，全入力数は200個である。モデルの目標値としては平均母音を用いた。

認識実験の結果を表8に示す。スペクトル・マッチングによる結果も併せて示す。スペクトル・マッチングの方法に対して，簡単化したモデルでは，6.5%，変数増減法では7.5%認識率が向上した。

Table 8　Results of the Recognition Tests of Five Speakers.

〔誤認識数〕　　〔誤認識数（全入力数）〕

強制変数付変数増減法

学習用話者	入力用話者 YI	YA	MT	HA	KS	
						closed test 2 (200)
YI YA	0	0	0	2	0	open test 2 (200)
						計 4 (400)
MT HA KS	0	0	0	2	0	

行列Mの成分の数　YI，YAで学習：39，
MT，HA，KSで学習：34　　認識率 99.0%

簡単化したモデル

学習用話者	入力用話者 YI	YA	MT	HA	KS	
						closed test 5 (200)
YI YA	0	1	0	2	0	open test 3 (200)
						計 8 (400)
MT HA KS	0	1	1	3	0	

行列Mの成分の数　15　　認識率 98.0%

スペクトル・マッチング

YI	YA	MT	HA	KS	計　17(200)
4	2	2	8	1	

認識率　91.5%

4 検　討

対称形3連母音の中央部母音を認識する際，中央部母音のパラメータだけで認識する方法（スペクトル・マッチングによる方法）に比較して，前後の母音の情報をも取入れる方法（モデルを用いる方法）は有効であった。この理由は単に，認識に使用するパラメータの数，すなわち情報の量が，後者では前者より多いためであるとも考えられる。パラメータの数は，前者では，母音空間パラメータの成分の数5であり，後者では，例えば表7の場合，行列Mの成分の数33～40で，表6の場合は15である。

ここでは上述のことの真為を検討するため，中央部母音のパラメータだけを用いるが前述のスペクトル・マッチング法よりはパラメータの数を多少増して認識実験を行い，モデルを用いる方法と比較する。

2名の話者（YI，HA）の対称形3連母音を認識実験用資料とする。これまでの実験と同様にして中央部母音の母音中心の母音空間パラメータを抽出する。一方の話者の母音空間パラメータを母音クラスごとに加算平均し，これを参照パタンとする。他方の話者の母音空間パラメータを入力パタンとし，入力パタンと参照パタンとの間のユークリッド距離を計算し，入力パタンに最近隣の母音クラスを認識結果とする。参照パタン用の話者を入替えて，これを2回行う。この方法では，参照パタンのパラメータ成分の数は25（＝5母音×5母音空間パラメータ成分）である。

認識実験の結果を表9に示す。この結果を，表6および表7と比較すると，誤認識の数は，表6，表7より多いことがわかる。使用するパラメータの成分の数は，表6の簡単化したモデルでは，行列Mの成分の数が15，定数項が5で計20であった。今回，情報の量は多くなっているにもかかわらず，認識率は必ずしも高くなっていない。すなわち，前後からの情報という情報の「質」が重要であることが示されている。

これらのことから，連続音声の認識においては，前後の音韻の音韻情報をとり入れることは，認識率の向上に役立つ，ということが結論付けられる。

Table 9　Results of the Recognition Test Using the Reference Pattern of Vowel Space Parameter of the Center Vowels.

誤認識数（カッコ内は全入力数）

学習用話者	入力用話者 YI	入力用話者 HA	closed test　3（80）
YI	0(40)	4(40)	open test　6（80）
HA	2(40)	3(40)	計　9（160）

参照パタンの総次元数　25

5 むすび

調音結合の線型モデルを用いて母音連鎖中の母音を認識する一方法を提案した。まず，モデルとその計算法を説明した。次に，このモデルに使用する音響パラメータとして母音空間パラメータを提案した。母音連鎖の例として対称形3連母音をとりあげ，モデルの妥当性を検討した後，認識実験を行った。認識実験では，まず，スペクトル・マッチングの方法と比較し，モデルを用いる方法が有効であることを示し，次に，このモデルを簡単化し，モデルの係数成分について考察した。5名の男性話者の音声資料を用いた実験では，スペクトル・マッチングの方法で認識率91.5%であるのに対し，簡単化したモデルを用いる方法では98.0%，提案したモデルを用いる方法では99.0%となった。最後に，認識時に必要な情報の量について比較検討し，連続音声認識において前後の音韻情報を利用することが有効であることを示した。

今後の課題としては，母音中心を自動的に検出すること，非対称形3連母音や一般の音節連鎖へモデルを適用すること，音節中心だけでなく，より連続的に係数を与えること，非線型モデルへモデルを一般化することなどが挙げられる。

参考文献

1) 桑原・境：“連続音声中の母音連鎖における調音結合効果の正規化”，音響学会誌，29，2，pp.91－99（1973－02）．

2) 石崎　俊：“調音モデルを用いた調音結合の動的処理”，音響学会音声研資，S78－45（1978－11）．

3) 芳賀・橋本：“回帰分析と主成分分析”，日科技連出版社（1980－05）．