連想モデルを用いる母音連鎖中の母音の認識

連想モデルを用いる母音連鎖中の母音の認識

高　良　富　夫*

Vowel Recognition Method for a Sequence of Vowels Using an Associative Model

Tomio TAKARA*

Abstract

In continuous speech, the feature of phoneme is ambiguous because of the coarticulation effect, which causes a difficulty of an automatic recognition of continuous speech. In the other hand, there is no coarticulation problem in human auditory.

In order to make the recognition easier, we propose an associative model of the human auditory process. It is assumed that the observed phoneme is associated with the acoustic characteristics of the three points (preceding, observed, and following points), and the association is linear.

The model was tested and it was found to be effective in a high noise condition. It was also found that the elements of the memory matrix were formed to clear the distinction between similar vowels, and that the spectrum reproduction error was sufficiently small.

Key Words : Recognition, Vowel, Continuous speech, Associative memory, Human auditory

1. まえがき

いくつかの音素を続けて発声した場合，発声器官の慣性などのため，音声の物理的な性質が平滑化され，隣接した音素は，互いに影響をおよぼしあって，その物理的特性が変化する（調音結合）。このため，連続音声中の音素のパターンの特徴は，孤立に発声された母音や音節中の音素に比較して不明瞭になっており，このことは，連続音声の自動認識を著しく困難なものにしている。

一方，人間の音声聴取過程において調音結合は，さほど問題にならない。この理由としては，文脈や文法上の前後関係の知識を利用して不明瞭になった音声を補償していることが挙げられるが，他方，文脈や文法とは独立に，前後の音韻情報で，この補償を行っていると考えられる現象がある。

例えば，前部と後部が同種の母音である母音の連鎖（対称形3連母音）VV_0Vにおいて，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，VV_0Vを聴取してV_0を同定すると，正聴率が95％程度になる[1]。この現象は，人間が連続音声中の母音を認識する場合，その直前・直後の音韻情報を利用していることを示している。

このような人間の音声聴取過程を模擬して調音結合の影響を除去し，連続音声認識を容易にしようとする試みとして，これまで，

(i)ホルマント周波数を特徴パラメータとする方法[1]，

(ii)調音パラメータを特徴パラメータとする方法[2]，

受付：1986年10月31日

*工学部電子・情報工学科

Dept. of Electronics & Information Engineering, Fac. of Eng.

(iii)スペクトルから特徴パラメータを抽出する方法[3]，などが提案されている。しかし，(i)，(ii)はいずれもパラメータの要素が独立に前後から影響を受けるとしたものであり，パラメータ要素の相互干渉は考慮していない。又，(iii)は，モデルの係数作成において，パラメータ要素相互の干渉を十分に考慮しているとはいえない。

そこで本論文では，人間の聴取機能と連想モデルの分散記憶との類似性に着目し，調音結合の影響を正規化する方法として，(iii)を一般化した線型連想モデルを用いる。ここでは，対称形3連母音VV$_0$Vの中央部の母音V$_0$を認識対象として，モデルを適用する。モデルの効果を検討するため，まず雑音下の音声の認識実験で，モデルを用いない方法との性能比較を行う。次に，モデルの係数（記憶行列）の要素の機能を定性的に検討し，最後に，モデルの補償特性をスペクトル上で定量的に評価する。

2. 原　　理

ここでは音声パターンの特徴ベクトルとして母音空間パラメータを使用し，これに連想モデルを適用するので，まず特徴ベクトルについて述べ，次に連想モデルについて述べる。

2.1 特徴ベクトル

図1に特徴ベクトル抽出のフローチャートを示す。

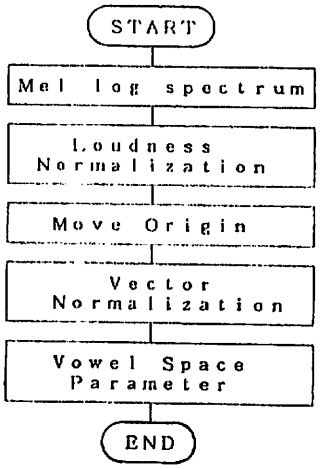


Fig. 1　Flow chart of extruction of a feature vector.

音声のスペクトル・パラメータとしては，比較的性能が良く，計算量の少ないメル対数スペクトル[4]を用いる。メル対数スペクトルは，パワースペクトルの周波数軸をメル尺度に変換し，振幅軸を対数尺度に変換したスペクトルである。後述の実験ではすべて，サンプリング周波数10kHz，量子化精度12ビットで音声波をA/D変換し，フレーム長25.6ms，フレーム周期20msでFFTを用いて20次のメル対数スペクトルを求めている。

第nフレームのメル対数スペクトルはベクトルを用いて

$$F_n=[F_n(1),\ F_n(2),\ \cdots,\ F_n(j),\ \cdots,F_n(J)]^T \quad (1)$$

と表記する。但し，Tは転置を表している。フレーム単位の音の大きさの正規化を次式で与える。

$$F'_n(j)=F_n(j)-\frac{1}{J}\sum_{j=1}^{J}F_n(j), \quad (2)$$
$$j=1\sim 20,$$
$$J=20.$$

式(2)は，音の大きさの正規化法の中ではメル対数スペクトルに特に有効である[4]

単母音（孤立発声の母音）のスペクトルの分布する空間（これを母音空間と呼ぶことにする）の重心に，式(2)のベクトルの原点を移動する。その方法は以下のとおりである。すなわち，母音クラス(i)に属する単母音のメル対数スペクトルを$V_k^{(i)}$と表すと，母音クラス(i)の重心$\overline{V}^{(i)}$は，

$$\overline{V}^{(i)}=\frac{1}{K}\sum_{k=1}^{K}V_k^{(i)}, \quad (3)$$
$$i=1\sim 5,$$

である。但し，Kは母音クラス(i)に属する単母音の数である。式(2)と同様に$\overline{V}^{(i)}$に音の大きさの正規化をほどこす。それを$\overline{V}'^{(i)}$と表せば，単母音全体の重心は，

$$\overline{\overline{V}}=\frac{1}{5}\sum_{i=1}^{5}\overline{V}'^{(i)} \quad (4)$$

である。$\overline{V}'^{(i)}$の原点を単母音全体の重心$\overline{\overline{V}}$に移し，ベクトルの大きさを1にする。すなわち，

$$\overline{V}''^{(i)}=\frac{\overline{V}'^{(i)}-\overline{\overline{V}}}{\|\overline{V}'^{(i)}-\overline{\overline{V}}\|}, \quad (5)$$
$$i=1\sim 5,$$

式(5)のベクトル群$\overline{V}''^{(i)}$，$i=1\sim5$（1,2,3,4,5はそれぞれ母音クラス /a/, /i/, /u/, /e/, /o/ に対応）は，母音空間を近似的に張っていると考えられるので，これらを母音空間の基本ベクトルと呼ぶことにする。ス

ベクトル F_n' の原点の移動およびベクトルの正規化は次式により行う。

$$F_n''=\frac{F_n'-\overline{\overline{V}}}{\|F_n'-\overline{\overline{V}}\|}. \tag{6}$$

F_n'' を母音空間に写像する。すなわち，

$$S_n(i)=\langle F_n'',\ \overline{V}^{(i)''}\rangle, \tag{7}$$

$$S_n=[S_n(1),S_n(2),\cdots,S_n(i),\cdots,S_n(5)]^T \tag{8}$$

但し，$\langle\cdot,\cdot\rangle$は内積を表している。

S_nは，スペクトル F_n'' を母音空間上でながめたものであるので，これを母音空間パラメータと呼ぶことにする。母音空間パラメータは，次元数が比較的少なく，かつ母音（不明瞭化した，対称形3連母音の中央部の母音を含めて）の特徴をよく表現していると考えられる。

2.2 連想モデルによる認識

人間が母音連鎖中の母音を認識する場合，その直前・直後の音韻情報をも利用している。このことを考慮して，ここでは，注目する時点の音韻の認識を，3時点（直前・注目点・直後）の音響特性からの連想としてとらえ，さらに，前後の音韻が注目点の音韻に線型に寄与すると仮定して，線型の連想モデルを構成する。

ここでは，母音連鎖の例として，対称形3連母音について検討する。対称形3連母音の3つの母音それぞれの母音中心（母音の特徴が最も明瞭な部分）の母音空間パラメータをそれぞれS_{n-1}，S_n，S_{n+1}とし，3連母音パターン全体の特徴 x を次式で表す。

$$x=\begin{bmatrix}S_{n-1}\\S_n\\S_{n+1}\end{bmatrix}. \tag{9}$$

5母音クラスに対応する5次元の単位ベクトル u_i，$i=1\sim5$をそれぞれ

$$\left.\begin{aligned}u_1&=[1,0,0,0,0]^T\\u_2&=[0,1,0,0,0]^T\\u_3&=[0,0,1,0,0]^T\\u_4&=[0,0,0,1,0]^T\\u_5&=[0,0,0,0,1]^T\end{aligned}\right\} \tag{10}$$

とする。但し，$i=1,2,3,4,5$はそれぞれ母音クラス /a/, /i/, /u/, /e/, /o/ に対応する。この単位ベクトルおよび対称形3連母音パターン x を用いると，連想モデルは次式で表される。

$$u=Mx. \tag{11}$$

但し，u は，xの中の S_nが属すべき母音クラスに対応する単位ベクトルである。Mは記憶行列と呼ばれ[5]，ここでは5行15列の行列である。

記憶行列 M が適当に与えられたものとすれば，対称形3連母音 x が入力されたとき，式(11)により，u が式(10)のいずれかのベクトルとして想起され，認識結果が得られる。

記憶行列 M は，次に示すように最小2乗解として得ることができる。まず，u および x の訓練用パターンの系列 ${}^ku$，$k=1\sim K$，および ${}^kx$，$k=1\sim K$ を行列形式で

$$U=[{}^1u\ {}^2u\cdots{}^ku\cdots{}^Ku], \tag{12}$$

$$X=[{}^1x\ {}^2x\cdots{}^kx\cdots{}^Kx] \tag{13}$$

と表す。但し，Uは5行 K列，Xは15行 K列の行列であり，${}^ku$ は，${}^kx$の中の S_nが属すべき母音に対応する単位ベクトルである。訓練用パターンのすべてについて式(11)が成立するとすれば，式(11)は行列の形式で

$$U=MX \tag{14}$$

とかける。式(14)のMの最小2乗近似解 $\hat{M}$は，

$$\hat{M}=UX^+ \tag{15}$$

で与えられる。[5] 但し，$(\cdot)^+$は一般逆行列を表す。

実際の認識では，xが入力されたとき，

$$\hat{u}=\hat{M}x \tag{16}$$

により $\hat{u}$ を求めるが，このとき，Mが近似解であることおよび x が線型独立でないことなどのため，$\hat{u}$ は必ずしも式(10)のいずれかに一致しない。しかし，$\hat{u}$ は u_i，$i=1\sim5$のいずれかの近似になっている[5]ので，$\hat{u}$ の要素を相互に比較して最大値の要素を求めることにより，$\hat{u}$ が u_i，$i=1\sim5$のいずれの近似であるかを決定することができる。

2.3 連想モデルの再成特性

前述の母音空間の基本ベクトルを直交化し，直交化された空間において連想モデルを適用し，それにより想起されたパラメータを音声スペクトルに変換する。得られた音声スペクトルを用いて，連想モデルによる音声スペクトルの再成誤差を評価する。

まず，母音空間の基本ベクトルを Gram-Schmidt の方法[5]により直交化する。式(3)の母音クラスの重心を表すベクトル群 $\overline{V}^{(i)}$，$i=1\sim5$から直交化された母音空間の基本ベクトル群 $\tilde{V}^{(i)}$，$i=1\sim5$を得る。方法は次の漸化式で与えられる。

(1) $i=1$のとき

$$\tilde{V}^{(1)}=\overline{V}^{(1)}. \tag{17}$$

(ii) $i=2\sim5$のとき

$$\tilde{V}^{(i)}=\tilde{V}^{(i)}-\sum_{j=1}^{i-1}\frac{\langle\tilde{V}^{(i)},\ \tilde{V}^{(j)}\rangle}{\|\tilde{V}^{(j)}\|^2}\tilde{V}^{(j)} \tag{18}$$

但し，$i=1$，2，3，4，5はそれぞれ/a/, /i/, /u/, /e/, /o/に対応する。

次に，次式により，この基本ベクトルの大きさを1に正規化する。

$$\tilde{V}^{(i)}=\frac{\tilde{V}^{(i)}}{\|\tilde{V}^{(i)}\|}. \tag{19}$$

以上により，$\tilde{V}^{(i)}$，$i=1\sim5$は，単母音の分布する空間の正規直交基底になる。

メル対数スペクトル$F_n^{\prime}$が入力されたとき，次式で$F_n^{\prime}$を母音空間に写像し，直交化母音空間パラメータS_nを得る。

$$S_n(i)=\langle F_n^{\prime},\ \tilde{V}^{(i)}\rangle, \tag{20}$$

$$S_n=[S_n(1),\ S_n(2),\ S_n(3),\ S_n(4),\ S_n(5)]^{\mathrm{T}} \tag{21}$$

このS_nに連想モデルを適用する。式(9)と同様に，対称形3連母音全体を

$$x=\begin{bmatrix}S_{n-1}\\ S_n\\ S_{n+1}\end{bmatrix} \tag{22}$$

と表すと，線型連想モデルは次式になる。

$$R=Mx. \tag{23}$$

但し，ここでRは，xの中のS_nが属すべき母音クラスの重心$\tilde{V}^{(i)}$の直交化母音空間パラメータである。換言すれば，Rは

$$R^{(i)}(j)=\langle\tilde{V}^{(i)},\ \tilde{V}^{(j)}\rangle, \tag{24}$$

$$R^{(i)}=[R^{(i)}(1),\ R^{(i)}(2),\ R^{(i)}(3),\ R^{(i)}(4),\ R^{(i)}(5)]^{\mathrm{T}} \tag{25}$$

で表される$R^{(i)}$，$i=1\sim5$のうちいずれかである。

式(23)の記憶行列Mは，式(12)～(15)と同様にして決定できる。すなわち，Rおよびxの訓練用パターンの系列をそれぞれ

$$R=[{}^1R\ \ {}^2R\cdots{}^kR\cdots{}^KR], \tag{26}$$

$$X=[{}^1x\ \ {}^2x\cdots{}^kx\ \cdots{}^Kx] \tag{27}$$

と表記すれば，全訓練用パターンに対する式(23)は，

$$R=MX \tag{28}$$

となる。式(28)の最小2乗近似解$\hat{M}$は

$$\hat{M}=RX^{+} \tag{29}$$

で与えられる。

xが入力されたとき，

$$\hat{R}=\hat{M}x \tag{30}$$

により$\hat{R}$が想起される。但し，

$$\hat{R}=[\hat{R}(1),\ \hat{R}(2),\ \hat{R}(3),\ \hat{R}(4),\ \hat{R}(5)]^{\mathrm{T}} \tag{31}$$

である。

$\hat{R}$を次式により変換し，母音スペクトル$\hat{F}$を再成する。

$$\hat{F}=\sum_{i=1}^{5}\hat{R}(i)\,\tilde{V}^{(i)}. \tag{32}$$

$\hat{R}$が正確に想起されれば，$\hat{F}$は$\tilde{V}^{(i)}$，$i=1\sim5$のいずれかに正確に一致する。

なお$\hat{F}$の再成誤差については第4章で検討する。

3. 認識実験

連想モデルの効果を調べるため，認識実験を行った。比較のため，まずモデルを用いない方法（スペクトル・マッチング法および$M^{(2)}$モデル（後述））により認識実験を行い，次に連想モデルを用いてClosed test およびOpen test を行う。以上は，雑音のない音声資料に対しての認識実験である。その後，雑音を付加した音声資料にして，スペクトル・マッチング法による実験および連想モデルを用いたOpen test を行い，両者を比較する。最後に，Closed test で作成された記憶行列を提示し，その行列要素の機能について定性的考察を行う。

認識実験で用いる音声資料は，日本人成人男性話者7名が普通の速さで各2回発声した対称形3連母音20種，すなわち／a i a／，／a u a／，／a e a／，／a o a／；／i a i／，／i u i／，／i e i／，／i o i／；／u a u／，／u i u／，／u e u／，／u o u／；／e a e／，／e i e／，／e u e／，／e o e／；／o a o／，／o i o／，／o u o／，／o e o／の計280個である。

式(9)におけるS_{n-1}，S_n，S_{n+1}としては，対称形3連母音の3つの母音中心1フレーム分を用いた。母音中心の位置は，母音空間パラメータの時系列パターンおよび波形を参照して視察により決定した。

式(3)における$V_k^{(i)}$は，日本人成人男性話者36名が発声した単母音を用いて，単母音の中央部4フレームのメル対数スペクトルを加算平均して求めた。

3.1 雑音のない場合

〔スペクトル・マッチング法〕

式(7)からわかるようにS_nの各成分$S_n(i)$，$i=1\sim5$は，スペクトル$F_n^{\prime}$の$\tilde{V}^{(i)\prime}$への類似度を表していると解釈することもできる。この場合，類似度の測度としては，S_nに対応するスペクトル$F_n^{\prime}$と平均母音

$\overline{V^{(i)}}$との間の方向余弦を用いていることになる。従って，次式

$$S_n(i)=\max_{1\le i\le 5} S_n(i) \qquad (33)$$

によって，iに対応する母音クラスがフレームnの認識結果であると決定することは，平均母音を標準パターンとする通常のスペクトル・マッチング法に他ならない。

ここでは，まず，対称形3連母音S_{n-1}，S_n，S_{n+1}のうち，S_nの1フレームだけに式(33)を適用する。これを方法M1とする。

次に，VV_0VのV_0内からS_nとS_nの直前・直後のフレーム計3フレームをとり，これを利用した実験も行った。方法M31では，この3フレームの母音空間パラメータを加算平均してできる1フレームの母音空間パラメータに式(33)を適用する。方法M32では，3フレーム個別に式(33)を適用し，その3つの結果の多数決をとる。

〔$M^{(2)}$model〕

前後の音韻(S_{n-1}，S_{n+1})からの効果を考察するためには，S_{n-1}，S_{n+1}を利用せずS_nだけを利用するようなモデルとも比較すべきである。

そこで

$$u=M^{(2)}S_n$$

型のモデルについても実験を行った。ここでは$M^{(2)}$は，上記連想モデルと同様にして作成する。

このモデルのClosed test 2で作成される行列$M^{(2)}$については3.3節でも検討する。

〔＋$V_0V_0V_0$Data〕

上記音声資料の中には，3連同一母音$V_0V_0V_0$が含まれていない。従って，S_{n-1}，S_{n+1}がS_nと異なっているという情報が連想モデルの認識率向上に寄与することが考えられる。そこで，これを確かめるため，音声資料に$V_0V_0V_0$型の資料を追加して実験を行った。但し，ここでは，これらの音声資料を準備することができなかったので，擬似的にこれを行った。すなわち，他のVV_0V型資料からVの中央部分を切り出して，これをV_nの部分におき，$V_0V_nV_0$型の資料を作成した。

〔Closed test1〕

同一音声資料セットを訓練用資料セットおよび認識用資料セットの両方に用いる。この場合，式(18)の$\hat{M}$が厳密解であれば，認識率は100%（完全な想起）になるはずである。しかし，前述のように$\hat{M}$は最小2乗近似解であるため，必ずしも完全には想起されない。

各話者の音声資料セット（40個）を用いて訓練，すなわち記憶行列$\hat{M}$の作成を行い，同じ音声資料を認識対象とする。これを各話者別々に行う。

〔Closed test 2〕

Closed testとしては他に，全資料を訓練用資料とし，同じ全資料を認識する実験も行った。ここで作成された記憶行列$\hat{M}$については，3.3節で詳細に検討する。

〔Open test〕

話者7名中3名の音声資料セットを訓練用資料セットとし，他の4名の音声資料セットを認識対象とする。これを，訓練用話者を入替えて3回行う。

以上の認識実験の認識率をまとめて表1に示す。スペクトル・マッチング法は，訓練を要しないので，Open testに含めた。

表1のスペクトル・マッチング法（Spectrum matching）では方法M1の方が方法M31，M32より良好である。すなわち，1フレームのみを用いた方が3フレーム用いるより良好である。この結果は，むしろ当然と考えられる。すなわち，VV_0VのV_0においては，その中央部のフレームS_nのパターンが最も明瞭であり，S_nから離れるに従って前後の母音の影響が強くなってパターンがボケる。従って，S_nの前後のフレームを利用することは，むしろ不利になる。実際，S_nの直前・直後でなく，S_nから2フレーム離れた部分を利用した場合は，認識率がさらに低下することを予備的実験で確認した。

Table1　Recognition score [%] of the test not using the noise.

	Spectrum matching					Associative
	M1	M31	M32	$M^{(2)}$ model	＋$V_0V_0V_0$ Data	model
Closed test 1	−	−	−	94.3	−	98.6
Closed test 2	−	−	−	94.3	96.1	96.4
Open test	93.6	92.1	92.1	92.9	94.1	94.8

$M^{(2)}$モデルでは，Open testにおいて，スペクトル・マッチング法（M1）より認識率が低い。従って，$M^{(2)}$モデルは連想モデルとしては不適当であると考えられる。

$V_0V_0V_0$型データを加えた場合（$+V_0V_0V_0$ Data）は，そうでない場合（Associative model）より認識率は低いが，スペクトル・マッチング法や$M^{(2)}$モデルよりは高い。

連想モデル（Associative model）のClosed test1の認識率がClosed test2の認識率より高いのは，Closed test1では，調音結合に含まれる個人性についても訓練されたためであると考えられる。

以上のように，モデルを用いる方法（Associative model,$+V_0V_0V_0$ Data）は，連想モデルを用いない方法（Spectrum matching，$M^{(2)}$ model）より認識率が高い。

このことから，ここで提案した方法は調音結合の影響を正規化するために効果があると考える。

3.2　雑音のある場合

前述の実験で用いた対称形3連母音の波形に雑音を付加し，これを用いて，前記と同様にOpen testを行った。但し，記憶行列は雑音を付加する前の音声資料を用いて作成した。

ここでは振幅が$-0.5 \leqq n(i) \leqq 0.5$である一様分布雑音を使用した。音声波形を$s(i)$，$i=1\sim I$とすれば，雑音を付加した音声波形$x(i)$，$i=1\sim I$は次式のようになる。

$$x(i)=s(i)+R\cdot n(i),\quad i=1\sim I. \tag{34}$$

$$R=\sqrt{12\times 10^{SNR/10}\times S^2} \tag{35}$$

但し，SNRはSN比であり，次式で与えられる。

$$SNR=10\log\frac{S^2}{N^2}, \tag{36}$$

$$S^2=\frac{1}{I}\sum_{i=1}^{I} s(i)^2, \tag{37}$$

$$N^2=\frac{1}{I}\sum_{i=1}^{I} n(i)^2. \tag{38}$$

ここでIは，音声継続時間長（サンプル）である。

認識実験結果の認識率を図2に示す。モデルを用いない場合（スペクトル・マッチング法）についても同様に実験を行い，併せて示した。ここでSN比∞dBというのは，前出の，雑音を付加する前の音声に対する実験結果である。

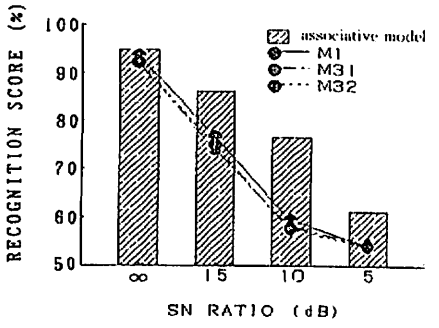


Fig. 2　Recognition score of the test using noise.

図2によれば，雑音の大きいとき（SN比15dB，10dBのとき）ほどモデルを用いる効果が大きいといえる。これは，連想モデルの，雑音に強いという特長を示すものであると考えられる。但し，雑音がかなり大きいとき（SN比5dBのとき）には，どの方法でも認識率はかなり低く，モデルの効果もさほど大きくないことがうかがわれる。

3.3　記憶行列

Closed test 2において作成された記憶行列を例にとって，記憶行列の要素の機能について考察する。表2に，記憶行列の例を示す。

便宜上，記憶行列Mを次式のように3つの行列に分ける。

$$M=[M^{(1)}M^{(2)}M^{(3)}]. \tag{39}$$

但し，$M^{(1)}$，$M^{(2)}$，$M^{(3)}$は5行5列の行列である。線型モデル式(10)は，式(39)を用いて次式のように表わされる。

$$u=Mx=M^{(1)}S_{n-1}+M^{(2)}S_n+M^{(3)}S_{n+1} \tag{40}$$

従って，$M^{(1)}$，$M^{(2)}$，$M^{(3)}$はそれぞれ，前部の音韻，中央部の音韻，後部の音韻の，uに対する寄与を表す係数であるといえる。

この寄与の大きさは，$M^{(j)}$の要素$m(j)$の2乗平均値$\bar{M}^{(j)}$で評価できる。但し，

$$\bar{M}^{(j)}=\sqrt{\frac{1}{25}\sum_{k=1}^{5}\sum_{i=1}^{5}(m(j))^2}, \tag{41}$$

$$M^{(j)}=(m(j)) \tag{42}$$

である。

表2の行列について，これを計算した結果は以下のとおりであった。

Table2　Memory matrix made in closed test2.

	$M^{(1)}$					$M^{(2)}$					$M^{(3)}$				
	/a/	/i/	/u/	/e/	/o/	/a/	/i/	/u/	/e/	/o/	/a/	/i/	/u/	/e/	/o/
/a/	−0.2460	0.3200	0.0524	−0.3652	0.0948	0.4548	−0.2042	0.0163	0.0317	−0.2951	−0.0348	−0.4430	0.2338	0.4832	0.1005
/i/	−0.3725	0.0149	0.2184	0.0542	0.2119	−0.1215	0.4560	−0.1438	−0.3450	−0.1194	−0.1341	−0.3332	0.2134	0.2599	−0.0637
/u/	0.1916	0.0029	−0.2039	−0.2869	0.1722	−0.0996	−0.3271	1.1700	0.0680	−0.4004	−0.1606	0.0319	−0.0126	0.2197	−0.0284
/e/	0.1488	0.1276	0.0901	−0.3586	−0.1397	−0.6877	−0.4721	0.2799	1.0970	0.3835	−0.0619	−0.0081	−0.1426	0.1191	0.1024
/o/	−0.1613	0.0990	0.1352	0.0285	−0.0843	−0.5166	−0.0161	−0.4887	−0.0633	1.1130	0.1243	−0.1573	0.1800	0.1209	−0.1683

$$\left.\begin{aligned}\overline{M}^{(1)}&=0.1990\\ \overline{M}^{(2)}&=0.4996\\ \overline{M}^{(3)}&=0.1982\end{aligned}\right\}\qquad (8)$$

この結果によれば，想起結果 $\hat{u}$ に対する寄与は，中央部の音韻からのものが最も大きく，前後からの寄与はその約半分であり，前・後はほぼ等しい。

このことから，想起パターンに対する効果は，S_n からのものが主であり，S_{n-1}，S_{n+1} からのものは，S_n からの想起をより明瞭にするため，補助的に働くと考えることができる。また「主」に対して「補助」効果が約半分であるということは，S_{n-1}，S_{n-1} の効果の和が「主」と同程度になるので，「補助」の効果が，わりと大きいともいえる。

次に，$M^{(2)}$ について定性的に検討する。$M^{(2)}$ の要素を相互に比較すると次のことがわかる。$M^{(2)}$ の，

(i) 対角要素の符号は正であり，その絶対値は各行で最大値である。

(ii) 各行で絶対値が2番目および3番目に大きい要素の符号は負であり，これらは各列において，その列に対応する母音クラスの隣接母音クラスに対応する要素である。但し，ここで隣接母音とは，図3において隣接する母音である。

(iii) 各行における残りの2要素の絶対値は比較的小さい。

S_n の各要素は対応する母音への類似度を表わしていることに注意すれば，以上のことから次のことがいえる。すなわち，S_n の要素の中で最大値である要素 $S_n(v)$ は，(i)により，式(8)の u の中の同じ要素を最大にするように働き，$S_n(v)$ は又，(ii)により，隣接母音に対応する要素に対しては，u において，これを小さくするように働く。従って，$M^{(2)}$ の要素は，隣接母音間の分離度をよくするように構成されているということがいえる。

比較のため，$u=M^{(2)}S_n$ 型モデルの $M^{(2)}$ を表3に示す。この場合は，上記のことが成立しているとは必ずしもいえない。

以上の効果が，認識実験における認識率にも現れていると考えられる。

$M^{(1)}$ および $M^{(3)}$ の各要素の機能については，今のところ，規則性が見い出せない。但し，$M^{(1)}$ および $M^{(3)}$ は相互に類似しておらず，むしろ相補的に機能しているように思われる。

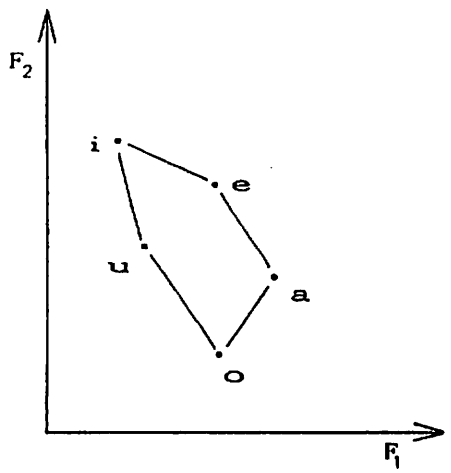


Fig. 3　Distribution of vowels in Formant frequency F1—F2 space.[6]

Table3 Memory matrix of the $M^{(2)}$ model.

	/a/	/i/	/u/	/e/	/o/
/a/	0.2541	−0.4242	0.2859	0.3231	−0.1885
/i/	−0.3301	0.2636	0.0846	−0.0751	0.0075
/u/	−0.1063	−0.2859	1.0855	0.0350	−0.3565
/e/	−0.7044	−0.4684	0.2195	1.1013	0.4443
/o/	−0.5093	−0.0745	−0.3543	0.0014	1.0267

4. 再成特性

想起された直交化母音空間パラメータから，2.3節に述べた方法により，音声スペクトル

$$\hat{F}_n=(\hat{F}_n(1),\ \hat{F}_n(2),\ \cdots,\ \hat{F}_n(j),\cdots,\hat{F}_n(20))^t \quad (44)$$

を再成し，その再成誤差を評価する。

前述のように，直交化母音空間パラメータが正確に想起されれば，$\hat{F}_n$は，平均母音$\bar{V}^{(i)}$のいずれかに正確に一致するはずであるから，再成誤差$\delta_n^{(i)}$を次式で定義する。

$$\delta_n^{(i)}=\sqrt{\frac{1}{20}\sum_{j=1}^{20}\{\hat{F}_n(j)-\bar{V}^{(i)}(j)\}^2} \quad (45)$$

母音クラス(i)に属すべき母音スペクトルの誤差$\delta_n^{(i)}$の平均$\bar{\delta}^{(i)}$は，

$$\bar{\delta}^{(i)}=\frac{1}{N}\sum_{n=1}^{N}\delta_n^{(i)} \quad (46)$$

である。但し，Nは母音クラス(i)に属すべき母音スペクトルの数である。$\bar{\delta}^{(i)}$の5母音にわたる平均値は，

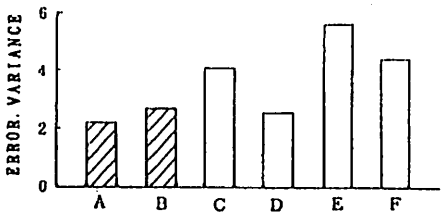
$$\bar{\delta}=\frac{1}{5}\sum_{i=1}^{5}\bar{\delta}^{(i)} \quad (47)$$

である。

式(47)の値を計算した結果を図4に示す。但し，ここでは，Open testにおける再成誤差およびClosed test 2における再成誤差を示した。又，比較のため，以下に述べるものも併せて示した。

〔単母音の分散〕

母音クラス(i)に含まれる単母音スペクトル$V_k^{(i)}$の，平均母音$\bar{V}^{(i)}$からの距離を，式(45)（但し，$\hat{F}_n(j)$を$V_k^{(i)}(j)$で置き換える）で評価し，その距離をその母音クラス内で平均したもの。この値は，単母音スペクトルの分布のバラツキ（分散）の大きさを表している。ここでは，前述の男性話者36名の単母音を使用して，これを計算した。



A: Closed test 2
B: Open test
} Reproduction Error of Associative Model

C: Original Parameter
D: Using Vowel Space Parameter
} Variance of Isolated Vowels

E: Original Parameter
F: Using Vowel Space Parameter
} Variance of Center Vowels

Fig. 4 Reproduction error of associative model, variance of isolated vowels, and variance of center vowels.

〔中央部母音の分散〕

同様に，対称形3連母音の中央部の母音のスペクトル分布の分散。これは，中央部母音の不明瞭さの平均値を表しているということができる。

これら単母音・中央部母音については，これをまず母音空間パラメータに変換し，次に，この母音空間パラメータからスペクトルを再成した場合についても，スペクトルの分散を計算した。

図4から以下のことがわかる。

(i) モデルを用いる方法のスペクトル再成誤差の平均値（A，B）は，多数話者の単母音スペクトルの分散（C，D）の同程度，又はそれ以下である。

それは，中央部母音スペクトルの分散（E，F）の約半分である。

(iii) 一度母音空間パラメータに変換して再成したスペクトルの分散（D，F）は，原スペクトルの分散（C，E）より小さい。

モデルを用いる方法では，連想モデルを適用する前に，母音空間パラメータに変換している。従って，それには(ii)の効果が含まれている。しかし，図4によれば，(ii)の効果を差し引いても山が成り立つことがわかる。

以上，線型の連想モデルを用いる認識法におけるスペクトルの再成誤差は十分小さく，それは，多数話者の単母音スペクトルの，平均母音スペクトルからの距離と同程度であるということがいえる。

5. むすび

母音連鎖中の母音の調音結合を正規化する一方法として，線型の連想モデルを用いる方法を提案し，その雑音下での認識特性，および記憶行列の要素の機能，スペクトルの再成特性を調べた。その結果，雑音レベルの高いときほど，このモデルを使用することの効果が大きいこと，および記憶行列のうち中央部母音の係数となる部分の要素は，隣接母音間の分離度を高めるように構成されていること，モデルのスペクトル再成誤差は十分小さく，多数話者の単母音の，平均母音からの距離と同程度であることが明らかとなり，この方法の有効性が示された。

しかしながら，雑音が無い場合において，モデルを用いる方法のOpen testの結果とスペクトル・マッチング法の差は必ずしも明瞭でない。この点に関しては他の論文[7]で検討されている。

今後の課題としては，特徴点の抽出を自動的に行い，このモデルを一般の音節連鎖へ適用することが挙げられる。

謝辞：本論文は，本学電子・情報工学科元教授故鉢嶺元助先生の連想記憶の研究と著者の音声認識の研究とを結合した新しいパターン認識方法に関する協同研究の成果の一部をとりまとめたものです。鉢嶺先生は，著者が本学に赴任して以来，計算機ハードウェアの教授方法，卒業研究の指導方法，およびパターン認識の研究，特に連想記憶に関して著者に御教示下さるとともに，同じ講座の教授として，個人的な問題に関しても多大なる御助言御指導を下さいました。先生は若くして教授になられたばかりであり，これからの研究成果・教育業績が大いに期待されていましたところ突然他界されたことは，同学を志すものとして，かえすがえすも残念でなりません。先生の御冥福をお祈り申し上げますとともに，これまでの御助言御指導に深く感謝し，本論文を鉢嶺先生に捧げます。

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) 高良，今井：“調音結合の線型モデルを用いる母音連鎖中の母音の認識”，信学論(A)，J65－A，4，pp.398－399（1982－04）．

(4) 高良，今井：“メル・ソーン・スペクトルを用いる母音識別”，信学論(A)，J65－A，8，pp.818－825（1982－08）．

(5) T.コホネン著，中谷和夫訳：“システム論的連想記憶”，サイエンス叢書（N－14），サイエンス社（1980－10）．

(6) 三浦種敏監修：“新版　聴覚と音声”，p.364，電子通信学会・コロナ社（1980－02）．

(7) 高良，福嶺，鉢嶺：“一般線型連想写像を用いる母音連鎖中の調音結合の正規化”，信学論(D)，J69－D，2，pp.261－263（1986－02）．