マルコフモデルを用いる破裂音認識

マルコフモデルを用いる破裂音認識

高良富夫* 喜友名 健** 鉢嶺元助*

Stop Consonants Recognition Using Markov Model

Tomio TAKARA, Tsuyoshi KYŪNA and Gensuke HACHIMINE

Abstract

This paper describes an automatic speech recognition system utilizing Markov's stochastic model. The recognition algorithm is obtained by generalizing the Mahalanobis-DP method to the first order of Markov model. A closed test evaluation experiment of the system showed a recognition score of 97.8 percent for stop consonants placed between vowels.

Key Words: Markov Model, Speech Recognition, Stochastic, Consonant, Dynamic Programming

1. まえがき

連続音声中の音声は，人間の発声・調音器官の慣性のため，著しく中性化（不明瞭化）したものとなっている。このため，連続音声中の音声を自動的に精度よく認識するシステムを構成することは，音声自動認識の研究において，現在でも困難な課題のひとつとなっている。

連続音声中の中性化音声を精度よく認識するためには，人間の聴覚機能がそうである[1]ように，音声波の注目する時点の情報だけでなく，その前後の情報をも利用することが効果的であると考えられる。実際，連想モデルを利用する方法[2]は，このような観点から，人間の聴覚機能を線型モデル化する方法であり，その有効性が示されている。

一方，確率・統計的には，時系列の前後関係を考慮することは，マルコフモデルにより定式化される。マハラノビス距離を用いるDPマッチング法（MDP法）[3]は，本来，不特定多数話者の音声に対処するために考案された単語音声認識法であるが，その数学的構造は，0次のマルコフモデルになっている。

そこで，本論文では，MDP法を一般化し，1次のマルコフモデルとして，これを連続音声認識に適用する。まず，この方法の定式化および，これを計算機上に構成するためのアルゴリズムを示し，次に，この方法を連続音声中の破裂音の認識に適用し，その有効性を示す。

2. 原理

2.1 0次のマルコフモデル

音節クラス s に属する音声データ $X^{(s)}$ を

受付：1986年5月9日

* 工学部電子・情報工学科

** 沖縄日本電気ソフトウェア（株）

$$X^{(s)}=x_1^{(s)}x_2^{(s)}\cdots x_j^{(s)}\cdots x_J^{(s)}, \tag{1}$$

$$x_j^{(s)}=(x_{1j}^{(s)},\ x_{2j}^{(s)},\ \cdots,\ x_{Nj}^{(s)})^T \tag{2}$$

と表す。但し，jはフレーム番号，Jは全体のフレーム数，Nは特徴ベクトルの次元数である。

$X^{(s)}$の生起確率は，式(1)の各特徴ベクトルが独立であるとすれば，

$$\begin{aligned}P^{(s)}(X^{(s)})&=P^{(s)}(x_1^{(s)}x_2^{(s)}\cdots x_j^{(s)}\cdots x_J^{(s)})\\&=P_1^{(s)}(x_1^{(s)})P_2^{(s)}(x_2^{(s)})\cdots P_j^{(s)}(x_j^{(s)})\\&\quad\cdots P_J^{(s)}(x_J^{(s)})\end{aligned} \tag{3}$$

となる。ここで，各特徴ベクトルは式(4)の分布をすると仮定する。

$$\begin{aligned}P_j^{(s)}(x_j^{(s)})&=\frac{1}{\sqrt{(2\pi)^N|V_j^{(s)}|}}\\&\cdot\exp\{-\alpha(x_j^{(s)}-\bar{x}_j^{(s)})^T(V_j^{(s)})^{-1}(x_j^{(s)}-\bar{x}_j^{(s)})\}\end{aligned} \tag{4}$$

但し，$\bar{x}_j^{(s)}$は$x_j^{(s)}$の平均ベクトル，$V_j^{(s)}$は分散共分散行列，αは定数であり，Tは転置を表す。

以上のような確率分布を与えるものを参照パタンとする。

次に，認識の過程について述べる。入力音声データAが，

$$A=a_1a_2\cdots a_i\cdots a_I, \tag{5}$$

$$a_i=(a_{1i},\ a_{2i},\ \cdots,\ a_{Ni})^T \tag{6}$$

であるとき，音声データAの音節クラスsに対する尤度$L(A,s)$は，式(7)となる。

$$L(A,s)=\max_F B(A,F;s), \tag{7}$$

$$B(A,F;s)=\prod_{k=1}^{K}P_{j(k)}^{(s)}(a_{i(k)}), \tag{8}$$

$$F=c(1)c(2)\cdots c(k)\cdots c(K), \tag{9}$$

$$c(k)=(i(k),\ j(k)) \tag{10}$$

従って，音声データAの属する音節クラスは，

$$L(A,s_M)=\max_s L(A,s) \tag{11}$$

であるs_Mに決定する。

式(7)の対数をとり，負符号を付けて，

$$D(A,s)=\min_F\{-\ln B(A,F;s)\} \tag{12}$$

$$\begin{aligned}&=\min_F\sum_k\{\alpha(a_{i(k)}-\bar{x}_{j(k)}^{(s)})^T(V_{j(k)}^{(s)})^{-1}\\&\cdot(a_{i(k)}-\bar{x}_{j(k)}^{(s)})+\frac{1}{2}\ln(2\pi)^N|V_{j(k)}^{(s)}|\}\end{aligned} \tag{13}$$

とする。

式(13)は，フレーム間距離$d(i,j)$として

$$d(i,j)=(a_i-\bar{x}_j^{(s)})^T(V_j^{(s)})^{-1}(a_i-\bar{x}_j^{(s)})+\frac{1}{2}\ln|V_j^{(s)}| \tag{14}$$

を用いて，通常のDPのアルゴリズムで効率よく解くことができる。この場合，音声データAの属する音節クラスは，

$$D(A,s_M)=\min_s D(A,s) \tag{15}$$

であるs_Mである。

以上は，式(14)の$(1/2)\ln|V_j^{(s)}|$だけを除けば，マハラノビス距離を用いるDPマッチング法[3]と同じである。

2.2　1次のマルコフモデル

0次のマルコフモデルでは，式(1)の各特徴ベクトルが独立であると仮定した。各特徴ベクトルが，1つ前の特徴ベクトルの状態に依存すると仮定すると，$X^{(s)}$の生起確率は，

$$\begin{aligned}P^{(s)}(X^{(s)})&=P^{(s)}(x_1^{(s)}x_2^{(s)}\cdots x_j^{(s)}\cdots x_J^{(s)})\\&=P_1^{(s)}(x_2^{(s)}|x_1^{(s)})P_2^{(s)}(x_3^{(s)}|x_2^{(s)})\\&\quad\cdots P_j^{(s)}(x_{j+1}^{(s)}|x_j^{(s)})\cdots P_{J-1}^{(s)}(x_J^{(s)}|x_{J-1}^{(s)})\end{aligned} \tag{16}$$

となる（1次のマルコフモデル）。ここで式(16)の各フレームの確率を，

$$\begin{aligned}P_j^{(s)}(x_{j+1}^{(s)}|x_j^{(s)})&=\frac{P_{yj}^{(s)}(x_{j+1}^{(s)},\ x_j^{(s)})}{P_{xj}^{(s)}(x_j^{(s)})}\\&=\frac{P_{yj}^{(s)}(y_j^{(s)})}{P_{xj}^{(s)}(x_j^{(s)})},\end{aligned} \tag{17}$$

$$\begin{aligned}y_j^{(s)}&=(x_j^{(s)T},\ x_{j+1}^{(s)T})^T\\&=(x_{1j}^{(s)},\ \cdots,\ x_{Nj}^{(s)},\ x_{1(j+1)}^{(s)},\ \cdots,\ x_{N(j+1)}^{(s)})^T\end{aligned} \tag{18}$$

と表し，$y_j^{(s)}$と$x_j^{(s)}$が式(19)，(20)のような分布をすると仮定する。

$$\begin{aligned}P_{yj}^{(s)}(y_j^{(s)})&=\frac{1}{\sqrt{(2\pi)^{2N}|V_{yj}^{(s)}|}}\\&\cdot\exp\{-\alpha(y_j^{(s)}-\bar{y}_j^{(s)})^T(V_{yj}^{(s)})^{-1}(y_j^{(s)}-\bar{y}_j^{(s)})\}\end{aligned} \tag{19}$$

$$\begin{aligned}P_{xj}^{(s)}(x_j^{(s)})&=\frac{1}{\sqrt{(2\pi)^N|V_{xj}^{(s)}|}}\\&\cdot\exp\{-\alpha(x_j^{(s)}-\bar{x}_j^{(s)})^T(V_{xj}^{(s)})^{-1}(x_j^{(s)}-\bar{x}_j^{(s)})\}\end{aligned} \tag{20}$$

但し，$\bar{x}_j^{(s)}$は$x_j^{(s)}$の平均ベクトル，$\bar{y}_j^{(s)}$は$y_j^{(s)}$の平均ベクトル，$V_{yj}^{(s)}$は$y_j^{(s)}$の分散共分散行列，$V_{xj}^{(s)}$は$V_{yj}^{(s)}$の$N+1$行および$N+1$列以降の成分を除いてできる$N\times N$行列である。式(19)，(20)のような確率分布を与えるものを参照パタンとする。

次に，認識の過程について述べる。0次のマルコフモデルと同様に，音声データAの音節クラスsに対する尤度$L_2(A, s)$を，

$$L_2(A, s)=\max_{F} B_2(A, F; s), \tag{21}$$

$$B_2(A, F; s)=\prod_{k=1}^{K} P^{(s)}_{j(k)}(a_{i+1(k)} \mid a_{i(k)}) \tag{22}$$

とし，式(21)の対数をとり，負符号を付けて

$$D_2(A, s)=\min_{F}\{-\ln B_2(A, F; s)\} \tag{23}$$

$$\begin{aligned}=\min_{F}\sum_{k}\{&\alpha(b_{i(k)}-\bar{y}^{(s)}_{j(k)})^{\mathrm{T}}(V^{(s)}_{yj})^{-1}(b_{i(k)}\\&-\bar{y}^{(s)}_{j(k)})+\frac{1}{2}\ln(2\pi)^{2N}\mid V^{(s)}_{yj}\mid\\&-\alpha(a_{i(k)}-\bar{x}^{(s)}_{j(k)})^{\mathrm{T}}(V^{(s)}_{xj})^{-1}(a_{i(k)}-\bar{x}^{(s)}_{j(k)})\\&-\frac{1}{2}\ln(2\pi)^{N}\mid V^{(s)}_{xj}\mid\},\end{aligned} \tag{24}$$

$$\begin{aligned}b_{i(k)}&=(a^{\mathrm{T}}_{i(k)}, a^{\mathrm{T}}_{i+1(k)})^{\mathrm{T}}\\&=(a_{1i(k)}, \cdots, a_{Ni(k)}, a_{1(i+1)(k)},\\&\quad \cdots, a_{N(i+1)(k)})^{\mathrm{T}}\end{aligned} \tag{25}$$

とする。

式(24)は，フレーム間距離$d(i, j)$を

$$\begin{aligned}d(i, j)=&(b_i-\bar{y}^{(s)}_j)^{\mathrm{T}}(V^{(s)}_{yj})^{-1}(b_i-\bar{y}^{(s)}_j)\\&+\frac{1}{2}\ln\mid V^{(s)}_{yj}\mid-(a_i-\bar{x}^{(s)}_j)^{\mathrm{T}}(V^{(s)}_{xj})^{-1}(a_i-\bar{x}^{(s)}_j)\\&-\frac{1}{2}\ln\mid V^{(s)}_{xj}\mid\end{aligned} \tag{26}$$

として，DPのアルゴリズムで効率よく解くことができる。この場合，音声データAの属する音節クラスは，

$$D_2(A, s_M)=\min_{s} D_2(A, s) \tag{27}$$

であるs_Mである。

2.3　DPマッチング[4]

ここでは，始点および終点が自由で傾斜制限のあるDPマッチングを用いる。

音声データXの第jフレームと音声データAの第iフレームとの間の距離を$d(i, j)$とし，(1, 1)点から(i, j)点までの累積距離を$g(i, j)$とすると，漸化式は式(28)～(31)で与えられる。

$$g(1, 1)=2d(1, 1) \tag{28}$$

$$g(i, 1)=g(i-1, 1)+d(i, 1),\quad 2\leq i\leq R_1 \tag{29}$$

$$g(1, j)=g(1, j-1)+d(1, j),\quad 2\leq j\leq R_1 \tag{30}$$

$$g(i, j)$$

$$=\min\begin{cases} g(i-1, j-2)+2d(i, j-1)+d(i, j)\\ g(i-1, j-1)+2d(i, j)\\ g(i-2, j-1)+2d(i-1, j)+d(i, j)\end{cases} \tag{31}$$

但し，R_1は始点の自由度である。

フレーム間距離$d(i, j)$は，参照パタンを作成するときには，絶対値距離

$$d(i, j)=\sum_{l}\mid a_{li}-x_{lj}\mid \tag{32}$$

を使用する。但し，x_{lj}およびa_{lj}はそれぞれ核パタン（後述）の第jフレームおよび参照パタン作成用音声データの第iフレームの，特徴ベクトルの第l成分である。ここで絶対値距離を使用したのは，その計算量が少ないためである。認識のときには，フレーム間距離として0次のマルコフモデルでは式(14)を，1次のマルコフモデルでは式(26)を使用する。

時間正規化距離は次式で与えられる。

$$D(A, X)=\min\begin{cases}\min\limits_{J-R_2\leq j\leq J}\dfrac{g(I, j)}{I+j}\\ \min\limits_{I-R_2\leq i\leq I}\dfrac{g(i, J)}{i+J}\end{cases} \tag{33}$$

ここでIおよびJはそれぞれ音声データAおよびXの全フレーム数であり，R_2は終点の自由度である。

以下に述べる実験では，$R_1=6$ (=30 ms)，R_2は参照パタン作成時には0，認識時には6とした。全ての音声データは同じフレーム数($I=J=21$)とした。

2.4　参照パタンの作成方法

参照パタンとして，平均ベクトルおよび分散共分散行列の系列を用いる。図1に，参照パタン作成のフローチャートを示す。これを各音節クラスにつき行う。音声は，すでに音響分析されて特徴ベクトルの系列に変換されているものとする。

まず，その音節クラスに属する音声データをひとつ入力する。これは差分ベクトルの系列に変換されて参照パタン作成が終了するまで記憶される。これを核パタンと呼ぶことにし，同一クラスに属する参照パタン作成用音声データの各フレームの割り付けのために使用する。ここで差分ベクトル系列とは，もとの特徴ベクトルの系列を

$$C=c_1\ c_2\cdots c_j\cdots c_{21} \tag{34}$$

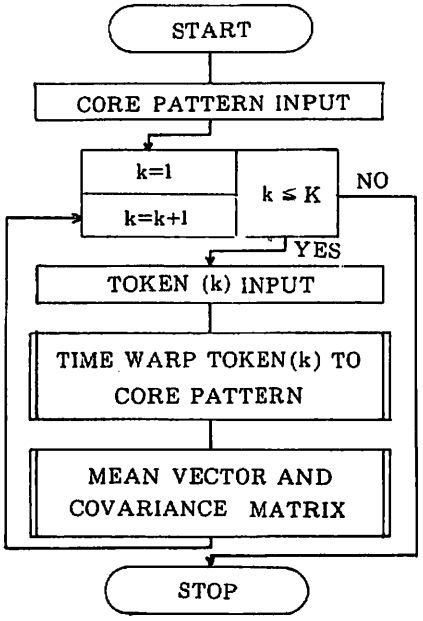

Fig. 1 Reference pattern generation.

としたとき，

$$C' = c'_1,\ c'_2\ \cdots c'_j\ \cdots c'_{21}, \tag{35}$$

$$c'_j = c_{j+1} - c_j \tag{36}$$

で与えられるベクトルの系列である。

次に，参照パタン作成用の音声データをひとつ入力する。これも同様に差分ベクトル系列に変換され，これと核パタンとのDPマッチングを行い，時間正規化距離が最小となるマッチング経路にそって，もとの特徴ベクトル系列を用いて，各フレームごとに平均ベクトルおよび分散共分散行列を作成する。

次の参照パタン作成用音声データを入力し，同様に，差分ベクトルを用いて核パタンとのDPマッチングを行い，以前の平均ベクトルおよび分散共分散行列と入力パタンとから新しい平均ベクトルおよび分散共分散行列を各フレームごとに作成する。

参照パタンのあるフレームにk回目に割り付けられた入力の特徴ベクトル，更新した平均ベクトルおよび分散共分散行列をそれぞれ，

$$x^{(k)} = (x_1^{(k)},\ x_2^{(k)},\ \cdots,\ x_n^{(k)},\ \cdots,\ x_N^{(k)})^T, \tag{37}$$

$$\bar{x}^{(k)} = (\bar{x}_1^{(k)},\ \bar{x}_2^{(k)},\ \cdots,\ \bar{x}_n^{(k)},\ \cdots,\ \bar{x}_N^{(k)})^T, \tag{38}$$

$$V^{(k)} = (V_{mn}^{(k)}) \tag{39}$$

とすると，更新は式(40)～(43)で行う。

$k=1$のとき

$$\bar{x}_n^{(1)} = x_n^{(1)}, \tag{40}$$

$$V_{mn}^{(1)} = 0 \tag{41}$$

$k \geqq 2$のとき

$$\bar{x}_n^{(k)} = \bar{x}_n^{(k-1)} + d_n^{(k)}/k, \tag{42}$$

$$V_{mn}^{(k)} = \frac{k-2}{k-1} V_{mn}^{(k-1)} + \frac{1}{k} d_m^{(k)} d_n^{(k)} \tag{43}$$

但し，

$$d_n^{(k)} = x_n^{(k)} - \bar{x}_n^{(k-1)} \tag{44}$$

以下同様に，参照パタン作成用音声データを次々と入力し，平均ベクトルおよび分散共分散行列を更新していく。

DPマッチングを行わない方法でも参照パタンの作成を行った。この方法では，入力の特徴ベクトルのフレーム番号と参照パタンのフレーム番号が同じになるようにフレームの割り付けを行った。これを，線型による割り付け方法と呼ぶことにする。

3. 認識実験

3.1 方 法

ここでは，破裂音の前後に母音を配して連続音声中の音韻を模擬した。また簡単のため，前後の母音は同一種とした。音声データは，破裂音音節30種の前後に同じ母音を配した30種×5母音=150個の連続音声を，成人男性話者3名が各1回発声したもの計450個である。破裂音音節30種を1セットとすると，15セット(=5母音環境×3名分)あることになる。この15セットに表1のようにセット番号を付ける。

これらの音声データは，5kHz低域通過フィルタを通した後，サンプリング周波数10kHz，精度12ビットでAD変換した。音声波にフレーム周期5ms，フレーム長25.6msでブラックマン窓をかけ，音響分析を行い，特徴ベクトルFMS（メル・ソーン・スペクトルの逆フーリエ変換）[3]の系列に変換する。実験では，視察で決めた子音から母音への変化点と，その前5フレーム，後15フレームの計21フレームを用いた。またFMSのうち第0成分を除いて，第1～3の成分を用いた。

表1　セット番号

セット番号	話　者	前後の母音
1	K.K.	/a/
2		/i/
3		/u/
4		/e/
5		/o/
6	T.K.	/a/
7		/i/
8		/u/
9		/e/
10		/o/
11	T.Y.	/a/
12		/i/
13		/u/
14		/e/
15		/o/

参照パタンの作成では，セット番号1のパタンを核パタンとした。

3.2　closed test

音声資料15セットで，各音節クラスの参照パタンを作成し，同じ15セットを入力パタンとして認識実験を行った。入力パタンと各音節クラスの参照パタンとのDPマッチング（フレーム間距離として，0次のマルコフモデルでは式(14)，1次のマルコフモデルでは式(26)を用いる）を行い，最小距離の音節クラスを認識結果とした。

比較のため，他に，フレーム間距離をユークリッド距離

$$d(i,\ j)=\sqrt{\sum_{l}(a_{li}-\bar{x}_{lj})^2} \qquad (45)$$

とする認識実験も行った。このとき参照パタンとしては，前述のDPによる作成法で作成した参照パタンのうち平均ベクトルの系列だけを用いた。

実験結果を表2に示す。

表2から次のことがわかる。フレーム間距離としてそれぞれ式(14)および式(26)を用いる0次のマルコフモデルおよび1次のマルコフモデルの方が，フレーム間距離をユークリッド距離とする場合よりも認識率は高い。参照パタンの作成は，線型（線型による割り付け方法）の方がDP（DPマッチングを用いる割り付け方法）よりも認識率は高い。また，1次のマルコフモデルの方が0次のマルコフモデルよりも認識率は高い。

3.3　資料に関する open test

音声資料15セットのうち14セットで各音節クラスの参照パタンを作成し，残り1セットを入力パタンとして認識実験を行った。入力パタンを入れ替えて，これを14回行ったので，全入力パタン数は420個である。

認識実験の結果を表3に示す。

平均認識率を比較すると，closed test では高い認識率が得られた1次のマルコフモデルが，資料に関する open test では0次のマルコフモデルよりも低い認識率になっている。ユークリッド距離をフレーム間距離とする認識方法は，closed test と同様に，0次のマルコフモデルおよび1次のマルコフモデルよりも低い認識率である。

この結果から，参照パタンとして平均ベクトルの系列だけを用いる方法より平均ベクトルおよび分散共分散行列の系列を用いる方法（0次，1次のマルコフモデル）が有効であることがわかる。

次に，参照パタンの作成を線型による割り付け方法で行った場合の結果を表4に示す。

表3と表4の結果を比較すると，closed test の結果と同じく，参照パタンの作成は線型で割り付けを行う方が高い認識率が得られた。

参照パタンの作成で線型による割り付け方法が高い認識率が得られたので，認識でも線型マッチング（入力の特徴ベクトルのフレーム番号と参照パタンのフレーム番号が同じになるようにマッチングを行う）を用いて実験を行ってみた。

表2　closed test の結果（認識率〔%〕）

参照パタンの作成方法	0次のマルコフモデル	1次のマルコフモデル	ユークリッド距離
DP	77.8	96.7	46.0
線型	85.3	97.8	—

表3 資料に関するopen testの結果(1)（認識率〔%〕）

入力パタンのセット番号	0次のマルコフモデル	1次のマルコフモデル	ユークリッド距離
2	60.0	33.3	36.7
3	53.3	60.0	30.0
4	53.3	53.3	46.7
5	50.0	43.3	30.0
6	60.0	36.7	46.7
7	50.0	43.3	50.0
8	70.0	60.0	66.7
9	53.3	43.3	36.7
10	80.0	56.7	53.3
11	56.7	56.7	30.0
12	53.3	60.0	30.0
13	56.7	66.7	40.0
14	60.0	73.3	43.3
15	53.3	50.0	56.7
平均	57.9	52.6	42.7

※参照パタンの作成は，DPマッチングを用いる割り付け方法で行った。

表4 資料に関するopen testの結果(2)（認識率〔%〕）

入力パタンのセット番号	0次のマルコフモデル	1次のマルコフモデル
2	56.7	43.3
3	66.7	56.7
4	56.7	53.3
5	56.7	50.0
6	66.7	46.7
7	53.3	50.0
8	66.7	50.0
9	60.0	60.0
10	70.0	56.7
11	63.3	53.3
12	53.3	56.7
13	63.3	63.3
14	70.0	73.3
15	46.7	53.3
平均	60.7	54.8

※参照パタンの作成は，線型による割り付け方法で行った。

表5 資料に関するopen testの結果(3)（認識率〔%〕）

入力パタンのセット番号	0次のマルコフモデル
2	56.7
3	53.3
4	53.3
5	50.0
6	63.3
7	46.7
8	63.3
9	50.0
10	70.0
11	66.7
12	53.3
13	66.7
14	63.3
15	53.3
平均	57.9

※参照パタンの作成は，線型による割り付け方法で行った。

※※認識で線型マッチングを使用した。

結果を表5に示す。このとき参照パタンは線型による割り付け方法で作成した。

表4と表5の結果を比べると，表4のDPマッチングを用いる認識の方が認識率が高い。

以上のことから，参照パタン作成時におけるフレームの割り付けは線型で行い，認識時における時間軸の正規化はDPで行う方が有効であるといえる。

参照パタンの作成においてDPを使うことが，本実験では，有効でなかった理由は次のように考えられる。すなわち，DPを使用すると，参照パタン作成用音声データのフレームはより類似の核パタンのフレームに割り付けられるので，作成される分散共分散行列の分散は一般に小さな値となる。従って，認識時における距離は，入力パタンが核パタンに似ていなければ，急激に大きな値となる。このとき核パタンが適当なものでなく，同一カテゴリの平均値（重心）からずれていると，同一カテゴリに属すべき入力パタンに対してもリジェクトする可能性が急激に増大する。その結果，誤認識が増大する。

DPを用いて参照パタンを作成する場合は，核パタンの選定をうまく行うことが必要であると考えられる。

4. むすび

マルコフモデルを用いる音声認識法を提案し，これを連続音声の破裂音の認識に適用した。認識実験の結果，closed testでは1次のマルコフモデルで97.8％の高い認識率が得られ，その有効性が示された。また，参照パタンの作成では，フレームの線型割り付けが有効であり，DPで割り付けを行う場合には，核パタンの選定が重要であることが示された。

しかし，open testにおいては，1次のマルコフモデルは，0次のマルコフモデルに比較して必ずしも良好な結果は得られなかった。これは，参照パタン作成のための音声データが少なかったことが一因であると考えられる。音声データが少ないときでも良好な参照パタンを作成するためには，得られた分散共分散行列の分散を単純に大きくすることや，前後のフレームを利用してスムージングすることなどが効果的と考えられるが，これは今後の課題とする。

また，参照パタン作成の際，ここではDP経路の“節”に条件付確率を割り付けたが，これはDP経路の“枝”に割り付けてもよいはずであり，この点についても今後検討する必要がある。

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”，日本音響学会誌，29，2，pp. 91−99（1973−02）.

(2) 高良，福嶋，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）.

(3) 高良，今井：“マハラノビス距離を用いるDPマッチングによる単語音声認識”，信学論（A），J66−A，1，pp. 64−70（1983−01）.

(4) Sakoe, H. and Chiba, S.: “Dynamic programing algorithm optimization for spoken word recognition”, IEEE Trans. Acoust., Speech & Signal Process., ASSP−26, 1, pp. 43−49 (1978−02).