

## フレーム間相関を用いた音韻HMM

高橋 敏 南 泰浩 松岡達雄 鹿野清宏

NTTヒューマンインタフェース研究所

現在のHMMの問題点の1つに、特徴ベクトル間の遷移に制約がない点が挙げられる。互いに出力確率が高い特徴ベクトル間の遷移は、学習データ中で観測されなかった遷移でも高い出力確率を与えられている。本稿では、2フレーム間の相関を用いて特徴ベクトルの遷移を制約し、不特定話者用HMMの特徴量分布を、入力話者に適した範囲に制約する新しいHMMを提案する。このモデルは、異なる音韻間の音韻特徴量分布の重なりを減少させ、認識率を向上させることができる。我々は既に、離散型不特定話者用HMMをもとに、VQコードのBigramを用いて遷移を制約する離散型Bigram 制約HMMを提案し、従来のHMMよりも性能がよいことを示した。本稿では、更に高い認識性能を得るために、この手法を半連続型HMM、連続型HMMに拡張した。これら3つのタイプのHMMを定式化し、連続音声中の音韻認識によって評価した。

## PHONEME HMMS USING FRAME CORRELATIONS

Satoshi Takahashi Yasuhiro Minami Tatsuo Matsuoka Kiyohiro Shikano

NTT Human Interface Laboratories

3-9-11 Midori-Cho, Musashino-Shi, Tokyo 180, Japan

One of the problems in Hidden Markov Modeling (HMM) is that there are no constraints on the feature vector transitions. All feature vector transitions which have high probabilities are allowed even if such transitions are not observed in the training data. This paper proposes new HMMs that use correlations between two frames to constrain the feature distributions of speaker-independent HMM to the region that is appropriate for an input speaker. This makes it possible to reduce the overlapping of feature distributions between different phonemes, and improve recognition performance. We previously proposed the bigram-constrained HMM based on the combination of the discrete HMM and the VQ-code bigram, and showed that it performed better than a conventional HMM. In this paper, the strategy is extended to semi-continuous HMMs and continuous HMMs to obtain better recognition performance. These three new types of HMMs are formulated and evaluated by phoneme recognition in continuous speech.

## 1. まえがき

不特定話者音声認識では、いかなる話者にも対応するために、様々な話者、様々な音韻環境のデータを使用し、モデルを作成する必要がある。隠れマルコフモデル (Hidden Markov Model: HMM) を用いた手法は、大量の音声データをもとに、統計的なモデル化を行なうことによって、話者や音韻環境による音韻特徴量のゆらぎを吸収し、従来の手法よりも高い認識性能を得ることができる。

不特定話者用HMMにおいて、各状態でモデルが表現する音韻特徴量の分布は、複数の話者、あるいは複数の話者クラスターの分布からなると考えられる。学習音声データのバリエーションを増加すると、未学習の特徴量空間は減少するが、一方で、各モデルが表現する音韻特徴量分布が次第に広がり、ある話者のある音韻の特徴量分布が、他の話者の異なる音韻の特徴量分布と重なることがしばしば起こる。この分布の重なりが、不特定話者音声認識の性能を劣化させている原因の1つであると考えられる。この問題は、現在のHMMが、状態ごとに音韻特徴量の静的な分布を表現しているにすぎず、常に学習に用いた全ての話者の音韻特徴量分布を用いて音韻を認識しているということに大きく起因する。1人の入力話者に対しては、その入力話者に適した一部の分布を考慮すれば十分であり、分布を制約することによって、音韻間の重なりを減少させることができる。

本報告では、不特定話者用HMMにおいて、広がって分布した多数話者の音韻特徴量分布を、2フレーム間の相関を利用して、入力話者に適した分布に制約するHMMを提案する。現在のHMMは、特徴量の遷移に関しては何ら制約がなく、互いに高い出力確率を持つ特徴量間の遷移は、学習データ中に観測されなかった遷移でも高い確率が与えられてしまう。我々は既に、離散型不特定話者用音韻HMMをもとに、フレーム間の相関情報としてVQコードのBigram情報を用いて、VQコードの遷移を制約するBigram制約HMM (Bigram-constrained HMM) を提案し、従来の音韻HMMよりも性能がよいことを示した<sup>[1][2]</sup>。本報告では更に、半連続型HMM、連続型HMMに拡張し、連続音声中の音韻認識によって、離散型Bigram制約HMMとの性能の比較を行なう。

## 2. フレーム間相関を利用したHMM

不特定話者用HMMの出力確率分布は、多数話者の分布からなると考えられるが、誤認識を減らすためには、入力話者に適した分布のみを用いて音韻を認識する必要がある。提案するモデルはフレーム間の相関を利用し、従来の不特定話者用HMMが表現する出力確率分布 (音韻特徴量分布) を入力話者に適した分布に制約するモデルである。静的な特徴量と同様に、動的な特徴量遷移にも含まれる音韻性情報や話者性情報を効果的に利用し、音韻間の分布の重なりを減少させ、認識率の向上を図る。

この考えに基づいて離散型、半連続型、連続型、それぞれのタイプのフレーム間相関を利用したHMMを定式化する。2フレームの音韻特徴量の相関情報として、離散型、半連続型ではVQコードのBigram、連続型では連続分布の条件付き確率を用いる。モデル作成に共通する大まかな処理の流れは以下の通りである。

- (1) 多数の学習用話者の音声データを用いて不特定話者用音韻HMMを作成する。
- (2) 学習用音声を用いてBigram (または条件付き確率分布) を計算する。
- (3) 上記2つのモデルの確率分布を合成し、前フレームの音韻特徴量を条件とする条件付き出力確率分布を得る。

### 2.1 離散型Bigram制約HMM

離散型Bigram制約HMMについては既に報告済みであるが<sup>[1][2]</sup>、半連続、連続型との比較のために、作成手順を以下に示し、ブロック図を図1に示す。

- (1) 多数の学習話者の音声データを用いて、ユニバーサルVQコードブックを作成し、これを用いて離散型不特定話者用音韻HMMを学習する。
- (2) 学習用音声を、ユニバーサルVQコードブックを用いてベクトル量子化し、Bigram確率を計算する。
- (3) 認識時に入力音声をベクトル量子化し、不特定話者用HMMの出力シンボル確率とBigram確率を

用いて、前フレームのVQコードで条件付けられた出力シンボル確率を求める。

$$b(c_j | c_i) = \frac{p(c_j | c_i) b(c_j)}{\sum_{m=1}^M p(c_m | c_i) b(c_m)} \quad (1)$$

ここで、ここで  $p(c_j | c_i)$  は、前時刻のフレームで  $i$  番目のVQコード  $c_i$  が出現した場合に、現時刻のフレームでVQコード  $c_j$  が出現するBigram確率である。 $b(c_j)$  は、従来の不特定話者用HMMにおいてVQコード  $c_j$  が出力する確率である。 $b(c_j | c_i)$  は、Bigram制約HMMにおいて、前時刻のフレームのVQコードが  $c_i$  であった場合に、VQコード  $c_j$  を出力する確率である。 $M$  はコードブックサイズである。この条件付き確率を用いて音韻を認識する。

## 2.2 半連続型Bigram制約HMM

半連続型HMM<sup>[3-5]</sup>は、混合連続分布型HMMの確率密度関数の平均値、分散を、すべての音韻モデル、すべての状態で"結び"としたモデルである。但し、各分布に対する重みは、モデルごと、状態ごとに最適に推定される。半連続型HMMにおいて、各分布の平均値を推定することは、ベクトル量子化(VQ)のコードブックの設計を行なっていることに等しく、モデルの学習中にVQとHMMパラメータの推定を同時に行なっている。従って、それらを独立に行なっている離散型HMMよりも、効率的にパラメータ表現されているといえる。

離散型Bigram制約HMMでは、VQコードのBigram情報を用いて出力シンボル確率を直接変更したが、半連続型Bigram制約HMMでは、各分布の平均値、分散を固定したまま、重み係数のみを変更する。以下に作成手順を示し、図2にブロック図を示す。

(1) 多数の学習話者の音声データを用いて、半連続型不特定話者用HMMを学習する。

(2) 半連続型HMMの各分布の平均値をセントロイドとみなし、ユニバーサルVQコードブックを作成する。

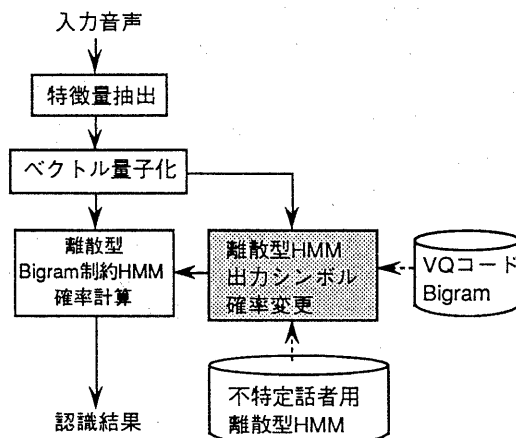


図1 離散型Bigram制約HMM

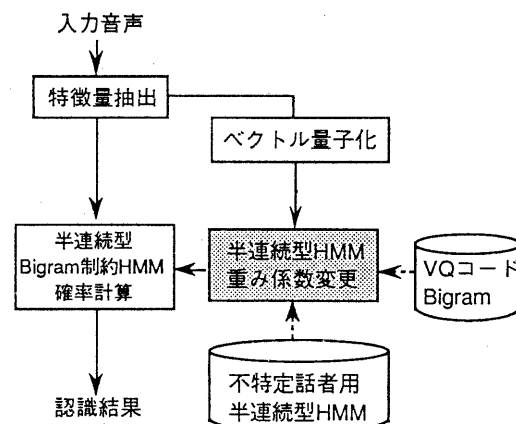


図2 半連続型Bigram制約HMM

(3) 学習音声をユニバーサルVQコードブックを用いてベクトル量子化し、Bigram確率を計算する。

(4) 認識時に入力音声をベクトル量子化し、Bigram確率を用いて、半連続型HMMの分布の重み係数を変更する。

$$\hat{\lambda}_j = \frac{p(c_j | c_i) \lambda_j}{\sum_{m=1}^M p(c_m | c_i) \lambda_m} \quad (2)$$

ここで、 $\lambda_j$  は変更前の、 $\hat{\lambda}_j$  は変更後の、 $j$  番目の分

布に対する重み係数である。 $c_i$  は  $i$  番目の分布の平均値をセントロイドにもつVQコードである。 $p(c_j | c_i)$  は、前時刻のフレームでVQコード  $c_i$  が出現し、現時刻のフレームでVQコード  $c_j$  が出現するBigram確率である。 $M$  は分布の総数 (=コードブックサイズ) である。

(5) 特徴ベクトル  $y$  に対する確率  $b(y)$  を、変更された重み係数を用いて、 $M$  個の連続分布の和によって求め、これを用いて音韻を認識する。

$$b(y) = \sum_{m=1}^M \hat{\lambda}_m b_m(y) \quad (3)$$

$b_m$  は  $m$  番目の確率密度分布を表す。

### 2.3 連続型Bigram制約HMM

連続型では、隣接する2フレームの特徴パラメータにおいて、前時刻のフレームの特徴パラメータを条件とする連続分布の条件付き確率分布と、不特定話者用HMMの確率分布の積によって得られる分布を用いて音韻を認識する。VQコードのBigramは使用しないが、ここでは離散型、半連続型にならない、仮に、連続型Bigram制約HMMと呼ぶことにする。モデル作成のブロック図を図3に示す。ここでは、不特定話者用HMMを混合ガウス分布で、条件付き確率を単一ガウス分布で近似する。

はじめに条件付き確率を計算する<sup>[6][7]</sup>。いま、ベクトル  $y$  が前フレーム  $y_1$ 、後フレーム  $y_2$  の2つの  $n$  次元の特徴ベクトルからなるとする。

$$y = (y_1, y_2) \quad (4)$$

$y$  は、平均値ベクトル  $\mu = (\mu_1, \mu_2)$  の多次元正規分布からのサンプルとする。共分散行列を

$$\Lambda = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix} \quad (5)$$

逆共分散行列を

$$V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} \quad (6)$$

とする。ただし、サブマトリクスは対角成分のみを考慮する。

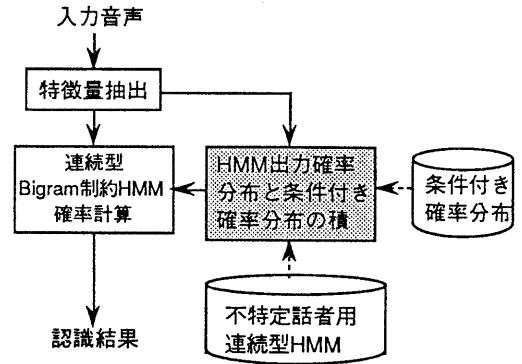


図3 連続型Bigram制約HMM

$$\Lambda_{11} = \begin{pmatrix} \varphi^{(1)}_1 & & 0 \\ & \ddots & \\ 0 & & \varphi^{(1)}_n \end{pmatrix}, \quad \Lambda_{22} = \begin{pmatrix} \varphi^{(2)}_1 & & 0 \\ & \ddots & \\ 0 & & \varphi^{(2)}_n \end{pmatrix},$$

$$\Lambda_{12} = \Lambda_{21} = \begin{pmatrix} \rho_1 & 0 \\ & \ddots \\ 0 & \rho_n \end{pmatrix} \quad (7)$$

$\varphi^{(1)}_i$  と  $\varphi^{(2)}_i$  は、それぞれ前後フレームにおける  $i$  次元の特徴パラメータの分散値である。また、 $\rho_i$  は  $\varphi^{(1)}_i$  と  $\varphi^{(2)}_i$  の相互相関係数である。

$z = y - \mu$ ,  $z_1 = y_1 - \mu_1$ ,  $z_2 = y_2 - \mu_2$  とし、 $p(y_1, y_2)$  が単一ガウス分布であると仮定すると

$$p(y_1, y_2) = \frac{|V|^{1/2}}{(2\pi)^{n/2}} e^{-\frac{1}{2} z^t V z} \quad (8)$$

ただし、 $z^t V z = z_1^t V_{11} z_1 + 2z_1^t V_{12} z_2 + z_2^t V_{22} z_2$  である。条件付き確率の定義から、

$$p(y_2 | y_1) = \frac{p(y_1, y_2)}{\int p(y_1, y_2) dy_2}$$

$$= \frac{e^{-\frac{1}{2} z^t V z}}{\int e^{-\frac{1}{2} z^t V z} dy_2}$$

$$= \frac{e^{-\frac{1}{2} (z_2^t V_{22} z_2 + 2z_1^t V_{12} z_2)}}{\int e^{-\frac{1}{2} (z_2^t V_{22} z_2 + 2z_1^t V_{12} z_2)} dy_2} \quad (9)$$

(9)式の分母の指数部を2次形式にした後、 $y_2$  で積分する。

$$\begin{aligned} & z_2^t V_{22} z_2 + 2z_1^t V_{12} z_2 \\ &= (z_2 + V_{22}^{-1} V_{21} z_1)^t V_{22} (z_2 + V_{22}^{-1} V_{21} z_1) - z_1^t V_{12} V_{22}^{-1} V_{21} z_1 \quad \dots (10) \\ & \int e^{-\frac{1}{2}(z_2^t V_{22} z_2 + 2z_1^t V_{12} z_2)} dy_2 \\ &= \frac{(2\pi)^{N/2}}{|V_{22}|^{1/2}} e^{\frac{1}{2} z_1^t V_{12} V_{22}^{-1} V_{21} z_1} \quad (11) \end{aligned}$$

従って、(9)式は以下ようになる。

$$\begin{aligned} p(y_2 | y_1) &= \\ &= \frac{|V_{22}|^{1/2}}{(2\pi)^{N/2}} e^{-\frac{1}{2} (z_2^t V_{22} z_2 + 2z_1^t V_{12} z_2 + z_1^t V_{12} V_{22}^{-1} V_{21} z_1)} \\ &= \frac{|V_{22}|^{1/2}}{(2\pi)^{N/2}} e^{-\frac{1}{2} (z_2 + V_{22}^{-1} V_{21} z_1)^t V_{22} (z_2 + V_{22}^{-1} V_{21} z_1)} \\ &= \frac{|V_{22}|^{1/2}}{(2\pi)^{N/2}} e^{-\frac{1}{2} (y_2 - (\mu_2 - V_{22}^{-1} V_{12} (y_1 - \mu_1)))^t V_{22} (y_2 - (\mu_2 - V_{22}^{-1} V_{12} (y_1 - \mu_1)))} \quad \dots (12) \end{aligned}$$

条件付き確率は、平均値  $\delta^{(p)} = \mu_2 - V_{22}^{-1} V_{12} (y_1 - \mu_1)$ 、共分散行列  $\Lambda^{(p)} = V_{22}^{-1}$  の単一ガウス分布になることがわかる。

$$\begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix} = I \quad (13)$$

より、

$$\begin{aligned} \delta^{(p)} &= \mu_2 - V_{22}^{-1} V_{12} (y_1 - \mu_1) \\ &= \mu_2 + \Lambda_{21} \Lambda_{11}^{-1} (y_1 - \mu_1) \quad (14) \end{aligned}$$

$$\Lambda^{(p)} = V_{22}^{-1} = \Lambda_{22} - \Lambda_{21} \Lambda_{11}^{-1} \Lambda_{12} \quad (15)$$

次に、この条件付き確率の単一ガウス分布と、音韻HMMの混合ガウス分布との積を計算し、新たなHMMの条件付き出力確率  $b(y_2 | y_1)$  を得る。

$$b(y_2 | y_1) = \sum_{m=1}^M \hat{\lambda}_m p(y_2 | y_1) b_m(y_2) \quad (16)$$

ここで、 $b_i(y_2)$  は現時刻のフレームの特徴パラメータ  $y_2$  の  $i$  番目の混合分布に対する出力確率である。 $\hat{\lambda}_i$  は、新たに計算される  $i$  番目の混合分布に対する重み係数である。 $M$  は混合分布数である。

(16)式において、条件付き確率  $p(y_2 | y_1)$  と音韻HMMの混合ガウス分布の  $i$  番目の分布  $b_i(y_2)$  との積は以下のように計算できる。

$$\begin{aligned} & p(y_2 | y_1) b_i(y_2) \\ &= \frac{|V^{(p)}|^{1/2}}{(2\pi)^{N/2}} e^{-\frac{1}{2} (y_2 - \delta^{(p)})^t V^{(p)} (y_2 - \delta^{(p)})} \\ & \quad \times \frac{|V^{(s)}|^{1/2}}{(2\pi)^{N/2}} e^{-\frac{1}{2} (y_2 - \delta^{(s)})^t V^{(s)} (y_2 - \delta^{(s)})} \quad (17) \end{aligned}$$

ここで、 $\delta^{(s)}$ 、 $V^{(s)}$  はそれぞれ、音韻HMMの  $i$  番目の混合分布の平均値ベクトルおよび逆共分散行列である。

先ほどと同様に指数部を2次形式にして展開すると、

$$= \frac{|V^{(p)} V^{(s)}|^{1/2}}{(2\pi)^N} e^{-\frac{1}{2} (y_2 - \hat{\delta})^t \hat{\Lambda}^{-1} (y_2 - \hat{\delta})} + \psi \quad (18)$$

ここで、

$$\psi = (\delta^{(p)} - \delta^{(s)})^t (V^{(p)} + V^{(s)})^{-1} (V^{(p)} V^{(s)}) (\delta^{(p)} - \delta^{(s)}) \quad (19)$$

平均値ベクトル

$$\begin{aligned} \hat{\delta} &= (V^{(p)} + V^{(s)})^{-1} (V^{(p)} \delta^{(p)} + V^{(s)} \delta^{(s)}) \\ &= (\Lambda^{(p)-1} + \Lambda^{(s)-1})^{-1} (\Lambda^{(p)-1} \delta^{(p)} + \Lambda^{(s)-1} \delta^{(s)}) \quad (20) \end{aligned}$$

共分散行列

$$\hat{\Lambda} = (V^{(p)} + V^{(s)})^{-1} = (\Lambda^{(p)-1} + \Lambda^{(s)-1})^{-1} \quad (21)$$

である。

また、(16)式における新しい重み係数  $\hat{\lambda}_i$  は、新たに生成される分布  $p(y_2 | y_1) b_i(y_2)$  の面積を考慮して決定する。

$$\hat{\lambda}_i = \frac{c_i}{\sum_{m=1}^M c_m} \quad (22)$$

$$c_i = \int p(y_2 | y_1) \lambda_i b_i(y_2) dy_2 \quad (23)$$

ここで、 $\lambda_i$  は変更前の重み係数である。

### 3. 評価実験

#### 3.1 連結学習と音韻ラベルなし評価法

これまでに述べた離散型、半連続型、連続型 Bigram制約HMMの性能を、連続音声中の音韻認識によって比較した。ベースとなる不特定話者用モデルは全て連結学習によって作成し、評価は音韻ラベルなし評価法<sup>[8]</sup>を用いた。よって、この実験では音

韻ラベルを一切使用していない。音韻ラベルなし評価法は、発声テキストをもとに音韻モデルを連結した後、評価する箇所の音韻モデルを全ての音韻モデルに置き換えながら文章全体の尤度を計算し、最大尤度を与える音韻を認識結果とする方法である。この評価法は、切り出された音韻の評価と違い、実際の連続音声認識に近い手法であり、また、タスクに独立であるという利点がある。

### 3.2 実験条件

使用したデータベースを表1にまとめる。連結学習に用いたデータは音響学会の連続音声データベースの音韻バランス文64人分である。各話者150文ずつ発声している（合計9600文）。評価は音響学会の連続音声データベースの対話読み上げ文13人分である。収録機関、発声内容が重ならないように823文を選択した。分析条件は、サンプリング周波数12kHz、ハミング窓長32ms、フレーム周期8msである。図4に音韻モデルの種類を示す。連結学習時には全ての発声テキストを覆うように43種類の音韻モデルを作成した。評価は、この中から長母音、二重母音、拗音等を除く23音韻に対して行なった。表2に、離散型におけるコードブックサイズ、半連続型におけるガウス分布の総分布数、連続型における各状態のガウス分布の混合数をパラメータごとに示す。連続型の出力確率は、共分散行列の対角成分のみ有する。HMMの構造は全て、4状態3ループである。

表1 実験に使用したデータベース

モデル	用途	使用データベース	文章数
不特定話者用HMM	学習	音響学会 連続音声データベース 音韻バランス文	9600文 (150文×64人)
	評価	音響学会 連続音声データベース 対話読み上げ文	823文 (計13人)
Bigram	入力話者用	対話読み上げ文 (HMMの評価用データとはテキスト独立)	各入力話者ごとに約50文 (3分間の音声)
	不特定話者用	HMMの学習用データ (音韻バランス文) より抜粋	2000文 (計40人)

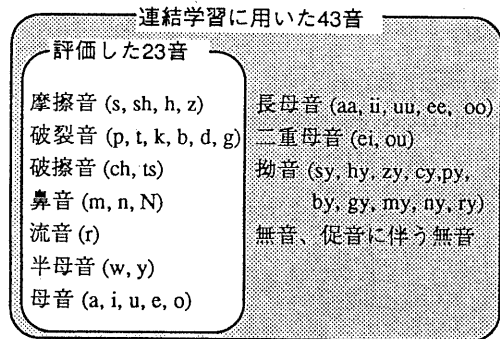


図4 音韻モデルの種類

表2 音韻特徴量

音韻特徴量	離散型 (コードブック サイズ)	半連続型 (総分布数)	連続型 (分布 /状態)
ケブストラム(16次)	256	256	4
Δ-ケブストラム(16次)	256	256	
Δ-パワー	64	64	

### 3.3 Bigramの作成

離散型、半連続型用のBigram、および連続型用の条件付き確率は、それぞれ、入力話者が発声した音声データのみから計算した入力話者用と、多数話者の音声データから計算した不特定話者用の2種類を用意した。入力話者用は、入力話者に限定したフレーム間相関情報を含んでおり、不特定話者用は、多数話者の一般的なフレーム間相関情報を含んでいる。図5に示すように、入力話者用Bigramからは入力話者用Bigram制約HMMを、不特定話者用Bigramからは不特定話者用Bigram制約HMMを作成することができる。表1に示すように、入力話者用は、音響学会の連続音声データベースの対話読み上げ文より、各入力話者が発声した平均50文（3分間の音声）を用いて作成した。この文章データは、HMMの評価用とはテキスト独立である。離散型、半連続型用のBigramを計算する場合、見かけ上のデータ量を増やすために Fuzzy VQ を使用した。不特定話者用は、音韻HMMの学習に用いた連続音声データ

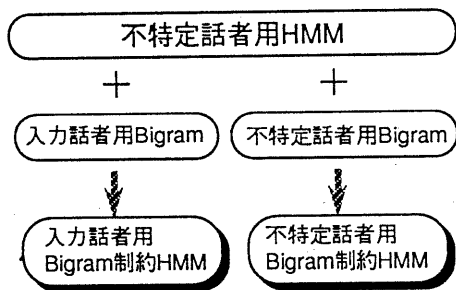


図5 Bigramの違いによるモデルの種類

ベースより、話者、あるいは発声テキストに片寄りがないように2000文を抜き出して作成した。

以前の検討結果から、Bigramを計算する際に、音韻ごとの出現頻度を正規化した方がよいことがわかっている<sup>1)</sup>。そこで、入力話者用、不特定話者用とも、発声テキストをもとに（教師あり学習）不特定話者用HMMを連結し、Viterbiセグメンテーションした後に音韻ごとにBigram（または条件付き確率）を計算した。

#### 4. 実験結果と考察

各モデルの評価用データに対する23音韻平均認識率を表3に示す。従来のHMMでは、半連続型、連続型の認識率がよく、離散型は若干悪い。Bigram制約HMMは、連続型の不特定話者用を除き、従来のHMMの性能を上回っている。また、入力話者用Bigram制約HMMは、不特定話者用Bigram制約HMMよりも性能がよい。更に、今回提案した半連続型

Bigram制約HMMは、以前提案した離散型Bigram制約HMMよりも性能がよくなっており、本モデルの有効性が示された。

しかし、連続型Bigram制約HMMでは、不特定話者用の条件付き確率を用いた場合に、もとのモデルの性能を下回っている。現在、条件付き確率は、単一ガウス分布で近似されており、相関行列も、対角成分（16個の相互相関係数）のみで表現されている。従って、相関情報が不十分であり、多数話者の分布を表現するまでに至っていないと考えられる。条件付き確率も混合分布で表現することを検討中である。

また、離散型、半連続型においても、不特定話者用Bigram制約HMMの認識率の改善が少ない。これは、認識対象が連続音声であることから、Bigramでは分布の制約が十分でないことが考えられる。よって、より制約を強めるために、Trigramを用いたモデルへの拡張が考えられる。また、入力話者用Bigram制約HMMでは、従来の話者適応技術と組み合わせ、適応化後にBigramを用いる方法が考えられる。

#### 5. むすび

2フレーム間の特徴パラメータの相関を利用し、不特定話者用HMMの分布を入力話者に適するように制約するBigram制約HMMを、離散型、半連続型、連続型について定式化し、評価実験を行なった。この手法によれば、不特定話者用HMMから、

表3 23音韻平均認識率 [%]

HMM Type		離散型	半連続型	連続型
従来の不特定話者用HMM		60.2	64.5	64.8
Bigram-constrained HMM	入力話者用	72.5	74.8	71.5
	不特定話者用	64.8	66.5	60.1

VQコードのBigram (離散型、半連続型の場合)、または前フレームの音韻特徴量の条件付き確率 (連続分布の場合) を用いることにより、不特定話者用Bigram制約HMM、あるいは入力話者に適応化した入力話者用Bigram制約HMMを得ることができる。

適応化用学習音声50文を用いた入力話者用モデルの実験の結果、半連続型Bigram制約HMMは、従来の半連続型不特定話者用HMMよりも性能がよく、更に、我々がこれまでに提案した離散型Bigram制約HMMよりも性能がよいことがわかった。一方、不特定話者モデルの実験では、半連続型Bigram制約HMMによって多少の認識率の向上が認められたが、今のところ、性能は不十分である。今後、関連情報をより詳細にすることにより、更に性能の向上が期待できる。

### 謝辞

日頃ご指導いただく古井特別研究室長、基礎研究所川端主任研究員、ご討論いただく皆様に感謝いたします。

実験では、日本音響学会連続音声認識データベースを使用した。

### [参考文献]

- [1] 高橋 敏, 松岡達雄, 鹿野清宏, "VQコードのBigramで制約したHMMによる音声認識", 信学技報 SP91-86 (1991-12)
- [2] S. Takahashi, T. Matsuoka, and K. Shikano, "Phonemic HMMs constrained by statistical VQ-code transition", Proc.ICASSP92, pp.553-556 (1992)
- [3] J. R. Bellegarda, D. Nahamoo, "Tied mixture continuous modeling for speech recognition", IEEE Trans. ASSP-38, 12, pp.2033-2045 (1990)
- [4] D. B. Paul, "The Lincoln tied-mixture HMM continuous speech recognizer", Proc.ICASSP91, pp.329-332 (1991)
- [5] X. D. Huang, "Phoneme classification using semicontinuous Hidden Markov Models", IEEE Trans. SP-40, 5, pp.1062-1067 (1992)
- [6] P. F. Brown, "The acoustic-modeling problem in

speech recognition", Ph. D dissertation, CMU (1987)

[7] C. J. Wellekens, "Explicit time correlation in Hidden Markov Models for speech recognition", Proc.ICASSP87, pp.384-386 (1987)

[8] 南 泰浩, 松岡達雄, 鹿野清浩, "不特定話者連続音声データベースによる連結学習HMMの評価", 信学技報 SP91-113 (1992-1)