TR-I-0018

Hidden Markov Model を用いた日本語有声破裂音の識別

Recognition of Japanese Voiced Stops Using Hidden Markov Models

花沢 利行　川端 豪　鹿野 清宏

Toshiyuki Hanazawa, Takeshi Kawabata and Kiyohiro Shikano

1987.12

概要

音韻を認識単位とするHMM(Hidden Markov Model)による音声認識手法の確立を目指し、日本語有声破裂音/b/ /d/ /g/ の識別実験によってHMMの学習回数や状態数、状態間をつなぐ弧の構成法等を検討した。各破裂音は成人男性の発声した5240単語中から切り出した。HMMの学習は各破裂音に対し約200サンプルを使い、10回程度の繰り返し学習でほぼ収束した。有声破裂音に対するHMMは4状態、3ループ以上が必要なことがわかった。状態間をつなぐ弧としてヌルアーク、タイドアーク等は識別率において大差なかった。また、語頭･語中別や後続母音別にHMMを作成して識別実験を行った。語頭･語中別のHMMは/g/のように語頭･語中で発声が異なるものに対しては有効であった。後続母音別のHMMでは学習データ数の不足が問題となることがわかった。後続母音別のHMMと後続母音で分けていないHMMとのcomposite modelによる識別実験も行ったが後続母音で分けていないものと同程度の識別率しか得られなかった。男性3名に対する語頭･語中でHMMを分けた識別実験では3名の平均識別率94.4%を得た。

ATR Interpreting Telephony Research Laboratories
ATR 自動翻訳電話研究所

- 目次 -

1. はじめに

発声状況等の違いによる音声の変動を統計的に表現できるHHM(Hidden Markov Model)を音韻認識に応用することを検討した。今までにHHMを用いた音声認識手法として、モデルの単位を単語とするもの[1][2]、単音節とするもの[3]、音韻とするもの[4][5]等が試みられている。自動翻訳電話のように大語彙を音声認識対象とする場合、音韻を認識単位としたHMMを用いる方式がその組合せとして単語や文節を扱う事ができ、最も柔軟性に富んでいる。そこで我々は音韻を単位としたHMMによる音声認識手法の確立を目指し、今回は有声破裂音/b//d/ /g/の識別実験によって、HMMの状態数や、各状態間を結ぶ弧の構成法および、語頭･語中別や後続母音別にHMMを作成したときの効果等の検討を行ったので報告する。

2. 音声データ

成人男性3名(MAU, MHT, MNM)の発声した重要語5240単語と、音韻バランス216単語を使用した。これらのデータには訓練されたラベラーにより音韻ラベルが付けられている。[6]

2.1 分析　上記の音声データを12KHzでサンプリングし窓長21.3msec、周期3msecのハミング窓で切り出し、高域強調をかけた後、12次のLPC分析を行った。

2.2 コードブックの作成　音韻バランス216単語を用い、/b/ /d/ /g/ を含めた全音声区間に対しPWLRを距離尺度としたベクトル量子化(VQ)を行い256個のベクトルからなるコードブックを作成した。

2.3 HMMの学習用データ　5240単語中の偶数番目の単語から音韻ラベルを参照して/b/ /d/ /g/ の区間を切り出した(拗音、促音を除く)。ここで言う/b/ /d/ /g/ の区間とは閉鎖(closure)＋破裂(burst)＋気音(aspiration)の区間である。切り出したデータは前述したVQコードブックを用いてコード番号列に変換し、学習用データとする。

2.4 HMMの評価用データ　5240単語中の奇数番目の単語から/b/ /d/ /g/ の区間を切り出し、前述のVQコードブックを用いてコード番号列に変換し、評価用データとする。

3. Hidden Markov Model

図1.1に示すようにHMMは数個の状態(s_0,s_1,s_2)と各状態間を結ぶ弧によって構成される。各弧には、その弧を通って状態間を遷移する遷移確率aと各コード番号

を出力する出力確率bがパラメータとして与えられている。HMMは状態間を遷移することによって遷移確率aと出力確率bによって定められた確率で様々なコード番号列を出力する。

3.1 HMMの弧の構成　図1.2に示すようにHMMの状態間を結ぶ弧の代表的なものとして、次の3種がある。簡単に説明する。

①通常の弧　その弧を通ったときに出力確率bに従ってコード番号を出力する。図1.2に示すようにここでは、この弧を実線で表すことにする。
②ヌルアーク(null arc)　その弧を通ったときに出力をしない弧である。したがって、ヌルアークのパラメータは遷移確率aだけである。図1.2に示すようにここでは、この弧を点線で表すことにする。
③タイドアーク(tied arc)　複数の弧の組で出力確率bを共通とするものをタイドアークと呼び、図1.2に示すようにタイドアークとする弧を直線で結んで表すことにする。

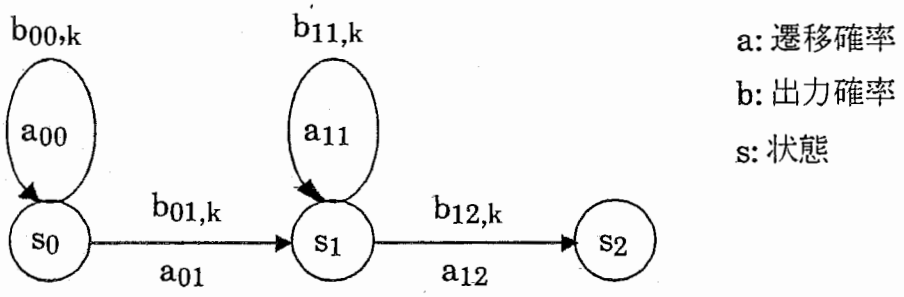


図1.1 Hidden Markov Model

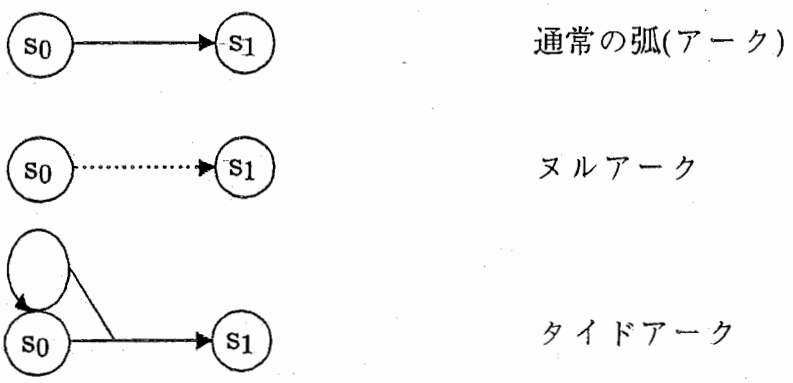


図1.2 HMMの状態間を結ぶ種々の弧

3.2 HMMの学習　/b/ /d/ /g/ 各々のHMMは、前述した学習データを使い、forward-backwardアルゴリム[7]により遷移確率a_{ij}(i,jは状態の番号)と出力確率

$b_{ij,k}$(kはコード番号)の学習を次の様に行う。学習後の遷移確率と出力確率をa_{ij}, $b_{ij,k}$と書くと、

$$a_{ij} = \frac{\sum_t \alpha(i,t-1)\cdot a_{ij}\cdot b_{ij,k}\cdot \beta(j,t)}{\sum_j \sum_t [\alpha(i,t-1)\cdot a_{ij}\cdot b_{ij,k}\cdot \beta(j,t)]}$$

$$b_{ij,k} = \frac{\sum_{t:o_t=k} \alpha(i,t-1)\cdot a_{ij}\cdot b_{ij,k}\cdot \beta(j,t)}{\sum_t \alpha(i,t-1)\cdot a_{ij}\cdot b_{ij,k}\cdot \beta(j,t)}$$

ここでα, βはそれぞれ前向き確率、後ろ向き確率と呼ばれ、次のように定義されている。
学習データのコード番号列を$o_1 o_2,\cdots,o_t,\cdots,o_T$としたとき、

$$\alpha(i,t) = \mathrm{Prob}(o_1 o_2,\cdots,o_t,\ \mathrm{destination} = s_i)$$

$$\beta(j,t) = \mathrm{Prob}(o_{t+1} o_{t+2},\cdots,o_T,\ \mathrm{origin} = s_j)$$

遷移確率a_{ij}と出力確率$b_{ij,k}$の学習の初期値は以下のように設定した。

○遷移確率a_{ij}の初期値　状態iから出る弧の数をnとしたとき$\sum_{j=1}^{n} a_{ij} = 1$, $a_{i1} = a_{i2} = \cdots = a_{in} = 1/n$と設定した。すなわち同じ状態iから出る弧の遷移確率には一様な値を与えた。
○出力確率$b_{ij,k}$の初期値　出力確率は学習データ中の各コード番号の出現頻度によって設定した。すなわち、$b_{ij,k}$ = (学習データ中に含まれるコード番号kの数)/(学習データ中に含まれるデータの総数)とし、各々の弧に等しい初期値をあたえた。

図2に/b/ /d/ /g/ 各々のHMMの学習過程での学習データの平均対数生起確率、すなわち各々の学習データがHMMから出力される確率の対数値の和を全学習データ長の和で正規化した値の変化の様子を示す。この図において縦軸が平均対数生起確率、横軸が学習の繰り返し回数である。この図から確率の収束に要する学習回数は音韻やデータ数によって異なるが10回程度の学習でほぼ収束していることがわかる。

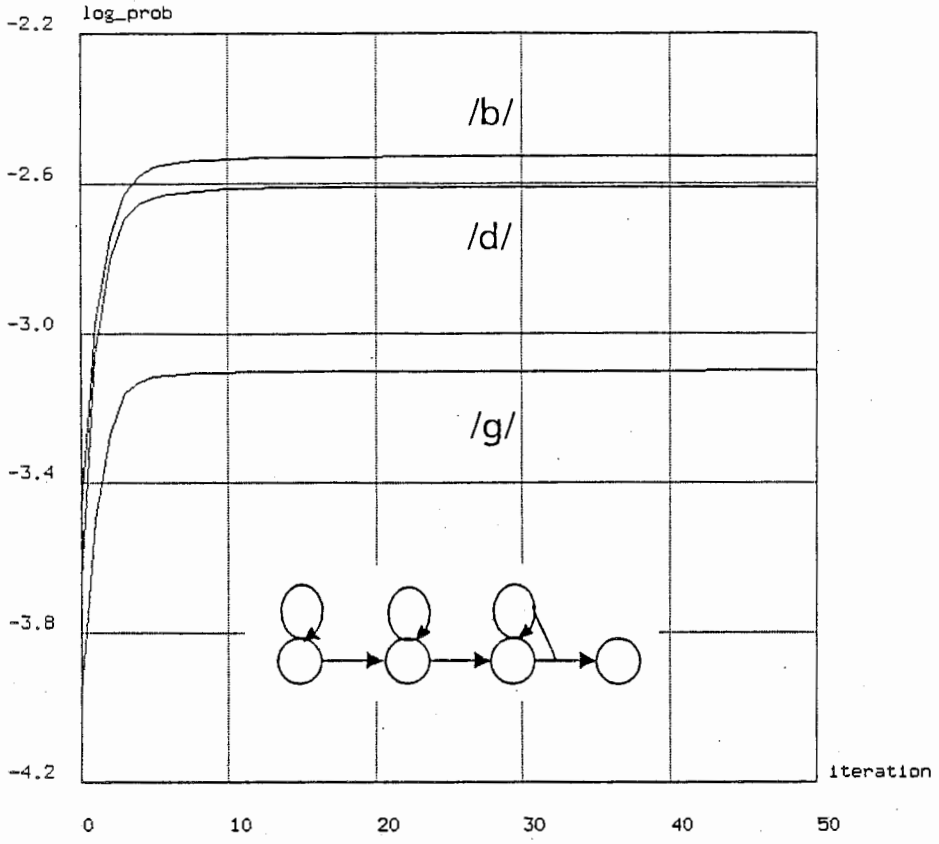


図2. HMMの学習過程における平均対数生起確率の変化 (話者:MAU)

3.3 HMMによる/b/ /d/ /g/の識別　評価データのコード番号列が/b/, /d/, /g/ 各々のHMMから出力される確率をtrellisアルゴリズムによって求める。そして各々のHMMから出力される確率を比較し、最大の確率を示したHMMが表現する音韻を認識結果とする。

但し、学習データに含まれないコード番号列が評価データに現れると、そのコード番号の出力確率が0になり誤認識を起こすので、今回の実験では出力確率bij,kの最低値を10^{-6}と設定して識別実験を行った。

4.　/b/ /d/ /g/ を表現するHMMの状態数の検討(実験1)

4.1 実験方法　HMMの状態数や構造を決める際に、HMMの各状態と音声事象との対応付けをしたいと考えている。有声破裂音の場合、①先行する音韻からの渡り②クロージャー③破裂及び母音への渡りという音声事象が考えられるのでHMMの状態数も3状態前後が適当であると思われる。そこで図3に示すように状態数の異なるHMMを用いて/b/ /d/ /g/ の識別実験を行い、状態数の違いが識別率に及ぼす影響を調べた。この実験は次の2通りのデータについて行った。

①学習･評価データとして /b/ /d/ /g/ の区間のみを用いた実験。

②学習･評価データとして/b//d//g/の区間に後続母音の区間を15msec(5フレーム)付加したものを用いた実験。但し、HMMは後続母音別に分けることはせず各破裂音に対して1つのHMMのみとする。

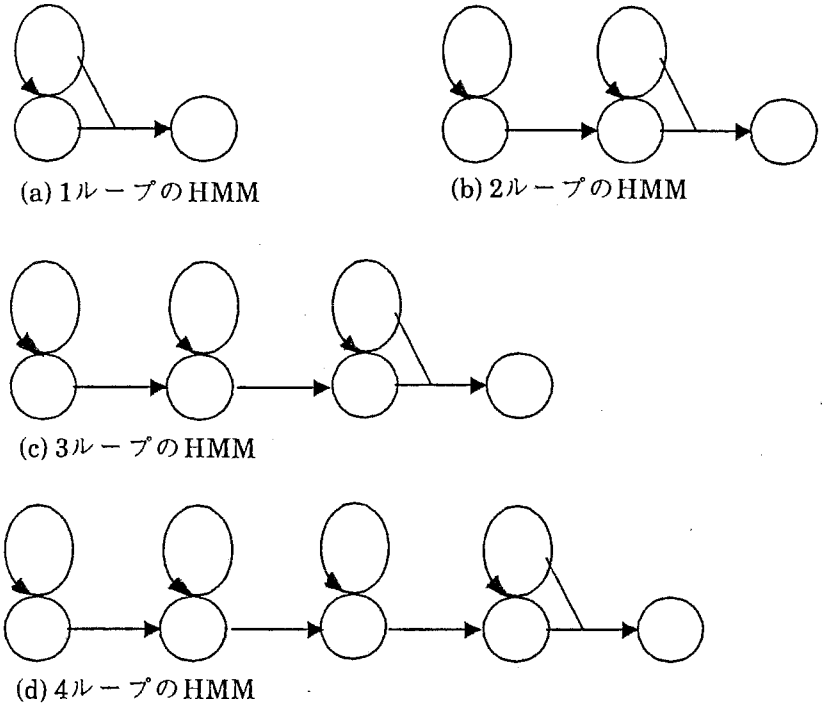


図3. 状態数の異なる種々のHMM

4.2 実験結果　表1.1に話者MAUに対する実験①、すなわち学習･評価データともに破裂音の区間のみを用いたものの識別結果を示す。この表から4状態で3ループを持ったHMMが一番高い識別率を示し、状態数が1つ多いHMMもほぼ同等の識別率を得ている事がわかる。一方、1ループしか持たないHMMでは他の3つのHMMに比べてかなり低い識別率であった。2ループを持つHMMでは1ループのHMMからはかなりの改善が見られるが3ループ以上を持つHMMと比べると識別率はやや低い。これは有声破裂音を表現するためにはループを持つ状態が2個以下では不十分であるためと考えられる。

表1.2に話者MAUに対する実験②、すなわち学習･評価データともに破裂音の区間に後続母音の区間を15msec付加した区間を用いたものの識別結果を示す。破裂音の区間のみで学習･識別を行った実験①と比べて全てのHMMにおいて識別率が向上しており、後続母音の特徴も/b//d//g/の識別に対して有効な情報であることがわかる。後続母音別にHMMを分けない場合でも識別率が向上するのはHMM内部で後続母音に対するマルチテンプレートを構成するためと考えられる。実験②においても3ループを持ったHMMにおいて最も高い識別率が得られ後続母音15msecを含んだモデルにおいても同程度の状態数でよいことがわかる。

表2.1に破裂音の区間のみを用いたもの、表2.2に後続母音の区間を15msec付加したものの識別実験の結果の混同表を示す。HMMは4状態3ループのものである。誤り分析のために語頭と語中に分けて表示してある。参考のためcloseデータによる実験、すなわちHMMの学習データを識別にも用いた実験の結果も示す(表2.1.c, 表2.2.c)。これらの表から語頭の/g/で識別率が低いことがわかるが、これは/g/の発声が語中では/ŋ/となるため1つのモデルで表現するには発声のバリエージョンが大きくなりすぎたためと考えられる。

図4.1-4.3は、横軸に正答カテゴリのモデルからの生起確率、縦軸に他カテゴリのモデルで生起確率が一番高いものをとりプロットしたものである。HMMは4状態3ループのものを用いている。これらの図を見ると後続母音の区間を付加することによって生起確率の分布が全体的に下方にシフトしている、すなわち他カテゴリのモデルからの生起確率が下がり識別率の向上につながったことがわかる。

以上、実験①②の結果から有声破裂音をHMMで表現するためにはループを持つ状態が3個以上必要なことがわかった。また有声破裂音の識別には後続母音のデータが有効であり後続母音別にHMMを分けない場合でも識別率が向上することが確認できた。

表1.1　有声破裂音のためのHMMの状態数の検討結果
(破裂音の区間のみ、話者:MAU)

HMMの状態数	識別率(%)			
	/b/	/d/	/g/	平均
2状態1ループ	83.3	87.2	79.4	82.8
3状態2ループ	86.3	89.9	83.3	86.2
4状態3ループ	88.1	91.1	85.7	88.0
5状態4ループ	86.8	89.9	86.9	87.7

表1.2　有声破裂音のためのHMMの状態数の検討結果
(後続母音のデータを15msec付加、話者:MAU)

HMMの状態数	識別率(%)			
	/b/	/d/	/g/	平均
2状態1ループ	89.9	95.5	83.3	88.9
3状態2ループ	91.2	96.6	89.3	91.9
4状態3ループ	92.1	96.6	90.9	92.9
5状態4ループ	91.2	97.2	90.1	92.4

表2.1　/b/ /d/ /g/の識別実験における混同表
(4状態3ループのHMM、破裂音の区間のみ、話者:MAU)

HMM 入力	/b/	/d/	/g/	データ数	エラー数	識別率(%)
/b/	**200**	24	2	227	27(1)	88.1
(語頭)	**45**	13	0	59	14(1)	76.3
(語中)	**155**	11	2	168	13	92.3
/d/	14	**163**	2	179	16	91.1
(語頭)	0	**65**	2	67	2	97.0
(語中)	14	**98**	0	112	14	87.5
/g/	19	17	**216**	252	36	85.7
(語頭)	6	16	**45**	67	22	67.2
(語中)	13	1	**171**	185	14	92.4
平均	-	-	-	658	79(1)	88.0

注:( )内の数はデータ長が短いためにリジェクトされたデータ数を示す。

表2.1.c　/b/ /d/ /g/の識別実験における混同表(closeデータ)
(4状態3ループのHMM、破裂音の区間のみ、話者:MAU)

HMM 入力	/b/	/d/	/g/	データ数	エラー数	識別率(%)
/b/	**201**	13	5	219	18	91.8
(語頭)	**53**	5	1	59	6	89.8
(語中)	**148**	8	4	160	12	92.5
/d/	3	**199**	1	203	4	98.0
(語頭)	0	**67**	1	68	1	98.5
(語中)	3	**132**	0	135	3	97.8
/g/	9	15	**236**	260	24	90.8
(語頭)	5	14	**49**	68	19	72.1
(語中)	4	1	**187**	192	5	97.4
平均	-	-	-	682	46	93.3

表2.2　/b/ /d/ /g/の識別実験における混同表
(4状態3ループのHMM、後続母音のデータを15msec付加、話者:MAU)

HMM 入力	/b/	/d/	/g/	データ数	エラー数	識別率(%)
/b/	**209**	16	2	227	18	92.1
(語頭)	**52**	7	0	59	7	88.1
(語中)	**157**	9	2	168	11	93.5
/d/	4	**173**	2	179	6	96.6
(語頭)	0	**66**	1	67	1	98.5
(語中)	4	**107**	1	112	5	95.5
/g/	14	9	**229**	252	23	90.9
(語頭)	9	6	**52**	67	15	77.6
(語中)	5	3	**177**	185	8	95.7
平均	-	-	-	658	47	92.9

表2.2.c　/b/ /d/ /g/の識別実験における混同表(closeデータ)
(4状態3ループのHMM、後続母音のデータを15msec付加、話者:MAU)

HMM 入力	/b/	/d/	/g/	データ数	エラー数	識別率(%)
/b/	**211**	5	3	219	8	96.3
(語頭)	**59**	0	0	59	0	100
(語中)	**152**	5	3	160	8	95.0
/d/	0	**203**	0	203	0	100
(語頭)	0	**68**	0	68	0	100
(語中)	0	**135**	0	135	0	100
/g/	9	4	**247**	260	13	95.0
(語頭)	6	3	**59**	68	9	86.8
(語中)	3	1	**188**	192	4	97.9
平均	-	-	-	682	21	96.9

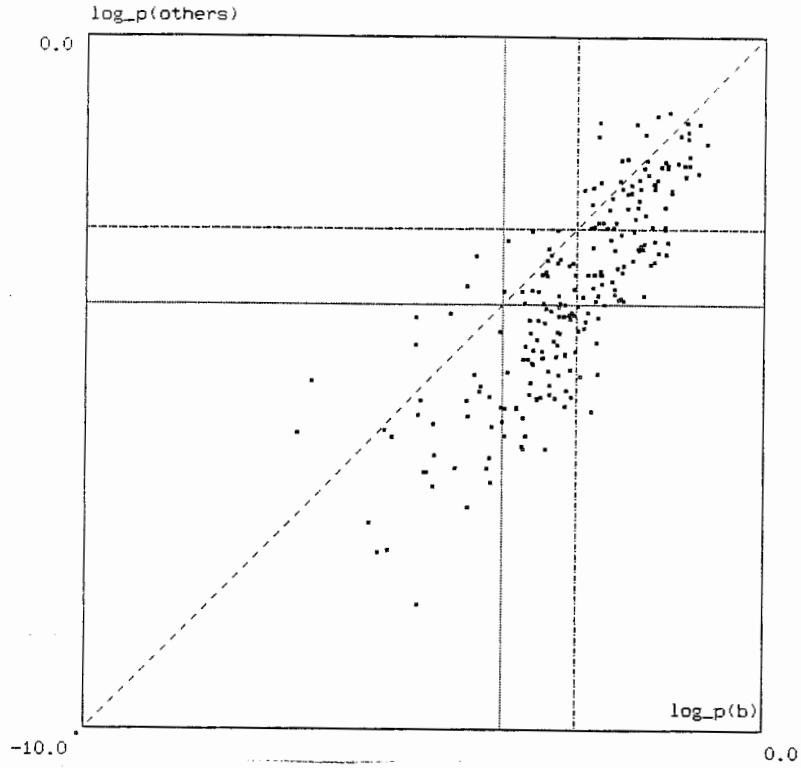


図4.1.1　正答カテゴリのモデルからの確率対他カテゴリのモデルからの最大確率の二次元プロット(入力… /b/)
(破裂音の区間のみ、話者:MAU)

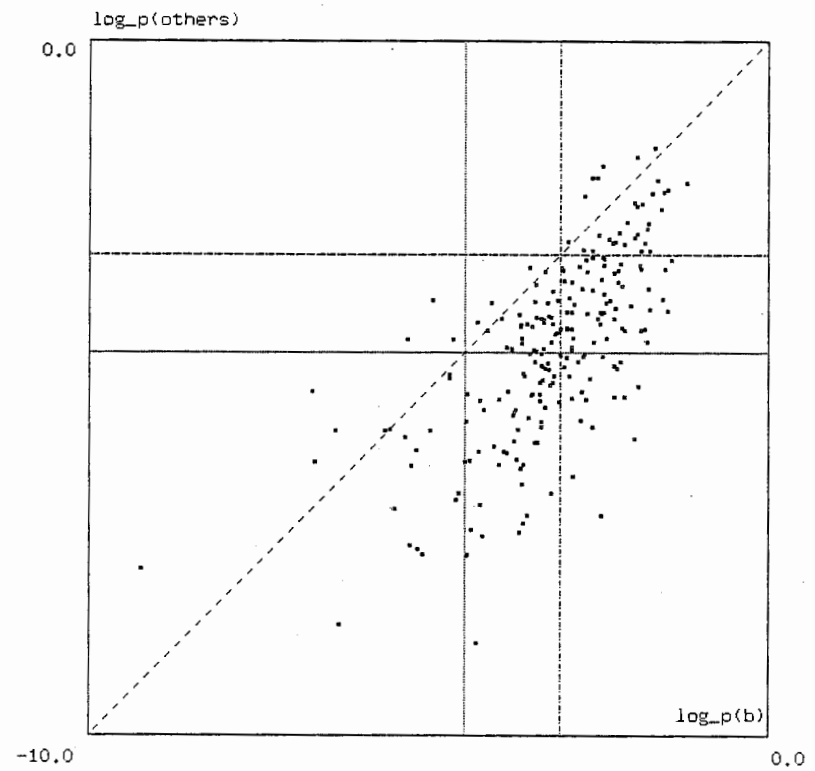


図4.1.2　正答カテゴリのモデルからの確率対他カテゴリのモデルからの最大確率の二次元プロット(入力… /b/)
(後続母音のデータを15msec付加、話者:MAU)

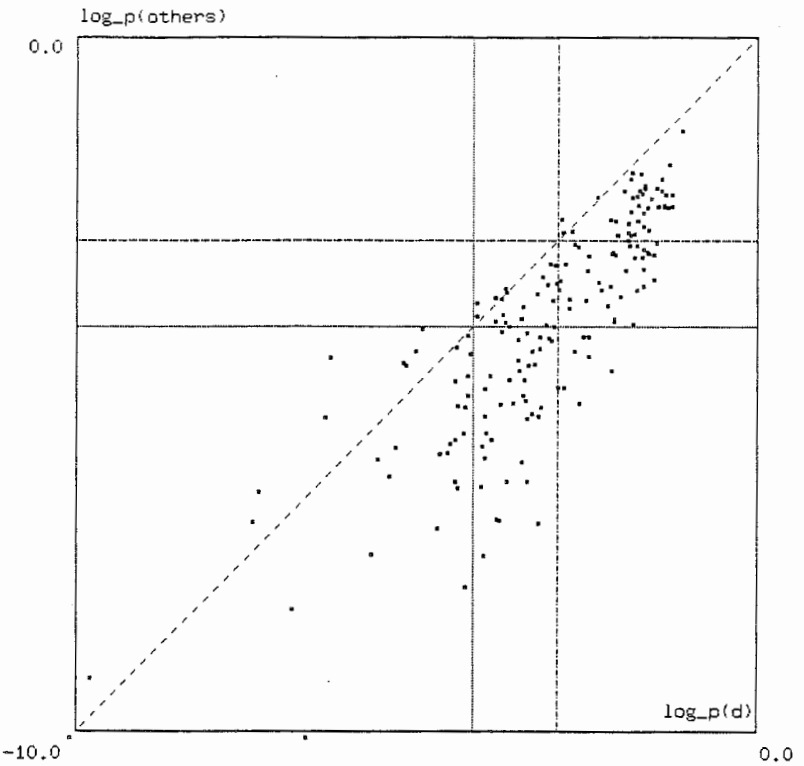


図4.2.1. 正答カテゴリのモデルからの確率対他カテゴリのモデルからの最大確率の二次元プロット(入力…/d/)
(破裂音の区間のみ、話者:MAU)

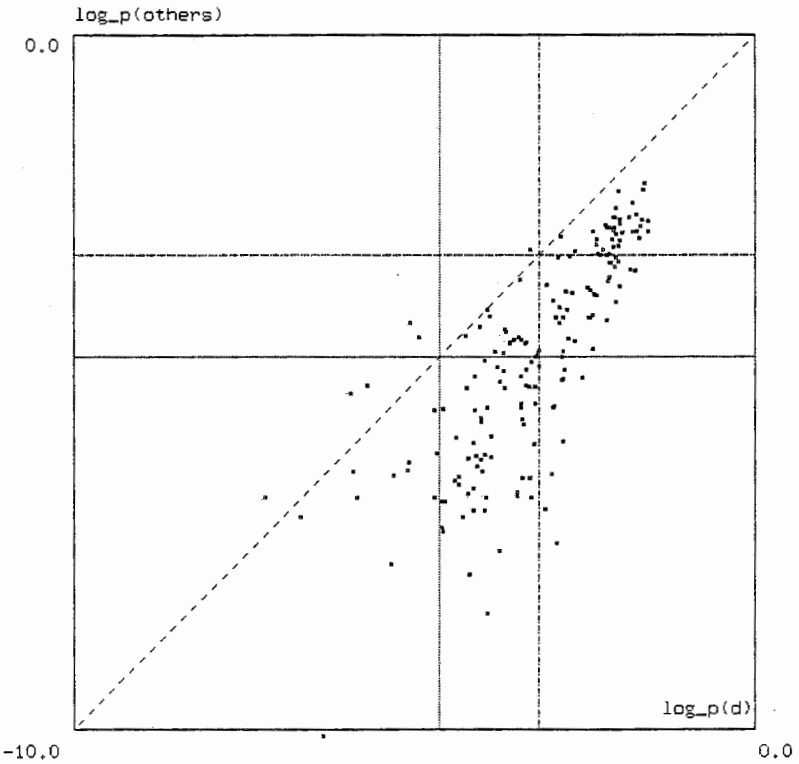


図4.2.2. 正答カテゴリのモデルからの確率対他カテゴリのモデルからの最大確率の二次元プロット(入力…/d/)
(後続母音のデータを15msec付加、話者:MAU)

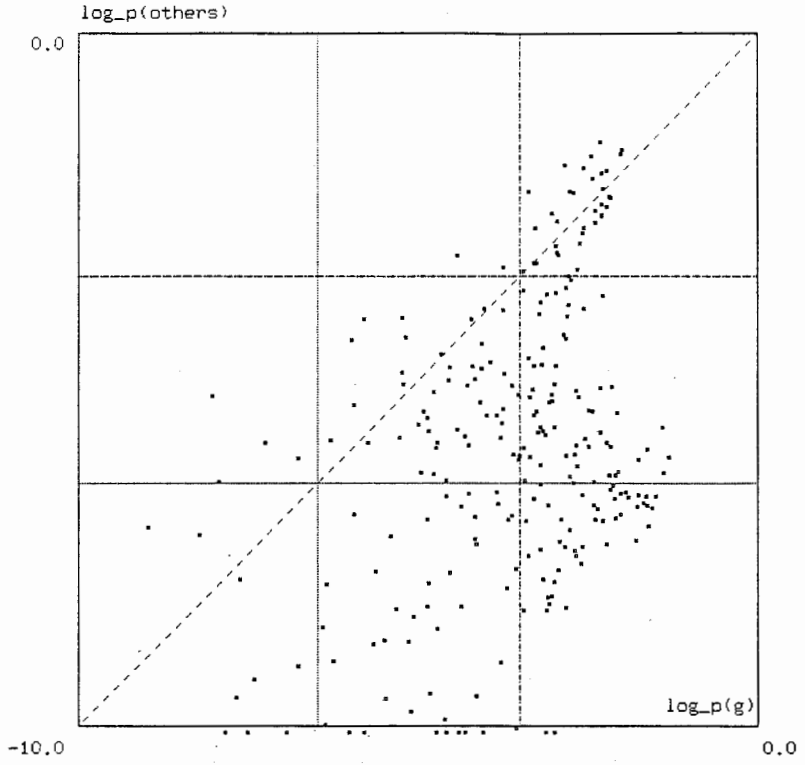


図4.3.1　正答カテゴリのモデルからの確率対他カテゴリのモデルからの最大確率の二次元プロット(入力…/g/)
(破裂音の区間のみ、話者:MAU)

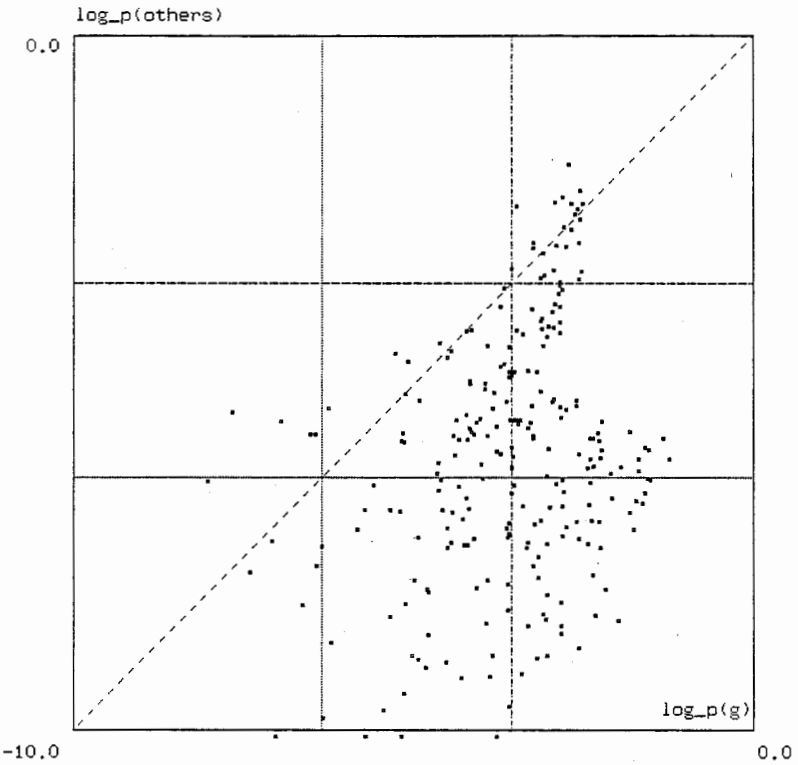


図4.3.2　正答カテゴリのモデルからの確率対他カテゴリのモデルからの最大確率の二次元プロット(入力…/g/)
(後続母音のデータを15msec付加、話者:MAU)

5. HMMの各状態間結ぶ弧の構成に関する検討(実験2)

5.1 実験方法 実験1の結果から有声破裂音をHMMで表現するためにはループを持つ状態が3個以上必要なことがわかった。そこで本章では3ループを持つHMMに対し、図5に示すように状態間を結ぶ弧の構成として3.1節で述べたような出力確率を共通とするタイドアーク、出力を出さないヌルアーク等の比較や、ループを持つ状態の最後にループを持たない状態を設けた時と、設けない時の違い等を検討する。有声破裂音の音声事象とHMMの状態との対応についても検討する。この実験で用いたデータは学習･評価データ共に /b/ /d/ /g/ の区間に後続母音を15mse付加したデータである。

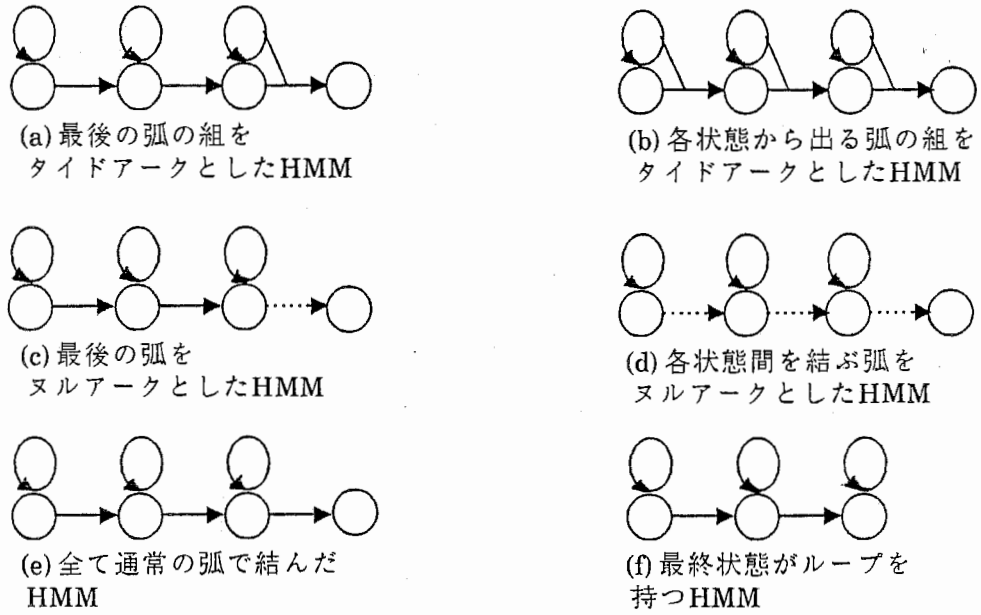


図5. 弧の構成が異なる種々のHMM

5.2 実験結果 表3に話者MAUに対する識別実験の結果を示す。この表から各HMMにおいて/b/ /d/ /g/ の平均識別率は大差ないことがわかる。但し最終状態がループを持つHMMは他のHMMと誤り傾向が異なっていることがわかる。すなわち最終状態がループを持たないHMM(a)-(e)では語頭の/b/の識別率が88%前後、語頭の/d/が98%前後であるのに対し、最終状態がループを持つHMM(f)では語頭の/b/の識別率が81.4%、語頭の/d/が91.0%となっている。逆に語頭の/g/の識別率は最終状態がループを持たないHMM(a)-(e)では77%前後、最終状態がループを持つHMM(f)では91.0%であった。このように誤り傾向が異なるのは以下に述べるように最終状態がループを持たないHMM(a)-(e)と最終状態がループを持つHMM(f)では、有声破裂音の音声事象とHMMの状態との対応関係が異なり、特に語頭のデータではこの傾向が著しいためと考える。

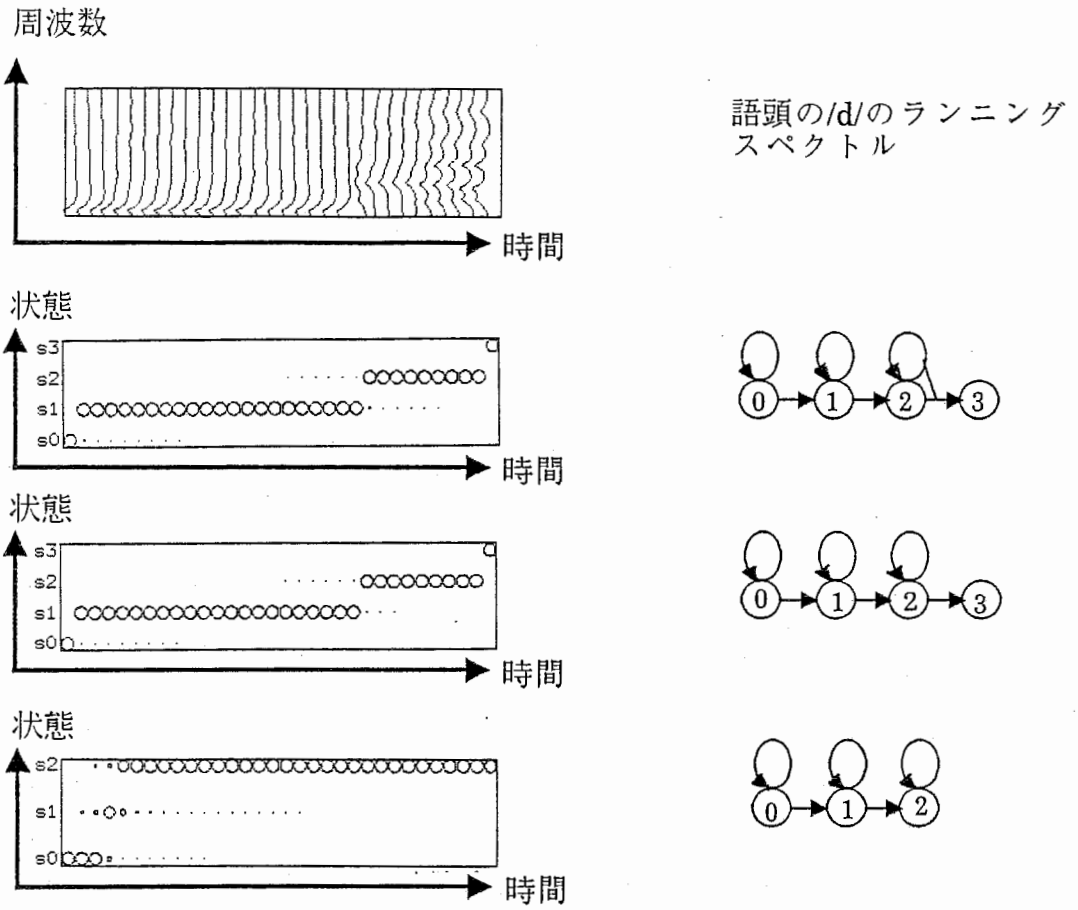
表3　HMMの弧の構成に関する検討結果
(後続母音のデータを15msec付加、話者:MAU)

HMMの弧の構成法	識別率(%)			
	/b/ (語頭/語中)	/d/ (語頭/語中)	/g/ (語頭/語中)	平均
タイドアーク(a)	92.1 (88.1/93.5)	96.6 (98.5/95.5)	90.9 (77.6/95.7)	92.9
タイドアーク(b)	91.2 (88.1/92.3)	96.6 (100/94.6)	89.3 (76.1/94.1)	91.9
ヌルアーク(c)	90.7 (88.1/91.7)	96.6 (98.5/95.5)	90.1 (76.1/95.1)	92.1
ヌルアーク(d)	90.3 (88.1/91.1)	96.6 (100/94.6)	89.3 (79.1/93.0)	91.6
通常の弧(e)	90.7 (89.8/91.1)	96.1 (98.5/94.6)	89.7 (74.6/95.1)	91.8
最終状態ループあり(f)	90.3 (81.4/93.5)	92.7 (91.0/93.8)	92.9 (91.0/93.5)	91.9

注: 表中の(a)~(f)は図5中の(a)~(f)に対応している

図6に有声破裂音の音声事象とHMMの状態との対応関係を示す。図中で一番上が、語頭から切り出した/d/のランニングスペクトルであり、下の3つの図は縦軸がHMMの状態、横軸が時間を示し、各HMMが/d/のスペクトルに対応するコード番号を出力する際に確率的な分布を持って状態を遷移する様子を円の面積で表したものである。この図から、最終状態にループを持たない上の2つのHMMでは状態s_1のループがクロージャー、s_2のループが破裂から母音にかけての区間への一応の対応がみられる。これに対し最終状態にループを持つHMMでは最終状態のループで大部分のデータを出力しており、音声事象とHMMの状態との対応が不適切である。これは最終状態のループでは遷移確率が恒常的に1であるため遷移確率が1より小さい他のループに対し遷移分布が偏ってしまうためと考えら

れる。



各時間における状態分布を円の面積比で表している

図6. 有声破裂音の音声事象と各HMMの状態との対応関係

以上まとめると、状態間を結ぶ弧として通常の弧、タイドアーク、ヌルアークを用いた場合どれも大差ないことがわかった。但し、最終状態にループを持つHMMでは最終状態のループに遷移分布が偏る傾向があり、音声事象と状態との対応付けに悪影響を及ぼすことがわかった。

6. 語頭·語中別に作成したHMMによる/b//d//g/の識別実験(実験3)

6.1 実験方法 実験1の結果から語頭の/g/は識別率が低いことがわかった。この問題を解決するため/b//d//g/の各破裂音に対し語頭と語中のデータで別々にHMMを作成し、識別実験を行った。HMMは最後の一組の弧だけをタイドアークにした構造(図5における(a))を使い、実験1と同様に次の2通りのデータについて行った。

①学習·評価データとして/b//d//g/の区間のみを用いた実験。

②学習·評価データとして/b//d//g/の区間に後続母音の区間を15msec(5フレーム)付加したものを用いた実験。

6.2 実験結果 表4.1に破裂音の区間のみを用いた実験の結果、表4.2に後続母音の区間を15msec付加したデータを用いた実験の結果の混同表を示す。この表において、語頭の/b/が語中の/b/と認識されても正解としている。語頭·語中を分けていない実験の結果である表2.1, 表2.2と比較してみると、いずれの場合も語頭の/g/の識別率が大幅に向上していることがわかる。後続母音の区間を付加したデータを用いた実験では95%の平均識別率を得た。

表4.1　語頭･語中別に作成したHMMによる/b/ /d/ /g/ の識別実験の混同表
(破裂音の区間のみ、話者:MAU)

HMM / 入力	/b/ (語頭/語中)	/d/ (語頭/語中)	/g/ (語頭/語中)	データ数	エラー数	識別率(%)
/b/	**203**	17	6	227	24(1)	89.4
(語頭)	**49/2**	5/0	2/0	59	8(1)	86.4
(語中)	**6/146**	0/12	2/2	168	16	90.5
/d/	10	**166**	3	179	13	92.7
(語頭)	0/0	**63/2**	2/0	67	2	97.0
(語中)	0/10	**5/96**	0/1	112	11	90.2
/g/	16	4	**232**	252	20	92.1
(語頭)	5/0	1/1	**60/0**	67	7	89.6
(語中)	0/11	0/2	**1/171**	185	13	93.0
平均	-	-	-	658	57(1)	91.3

注:( )内の数はデータ長が短いためにリジェクトされたデータ数を示す。

表4.1.c　語頭･語中別に作成したHMMによる/b/ /d/ /g/ の識別実験の混同表
(closeデータ、破裂音の区間のみ、話者:MAU)

HMM / 入力	/b/ (語頭/語中)	/d/ (語頭/語中)	/g/ (語頭/語中)	データ数	エラー数	識別率(%)
/b/	**208**	7	4	219	11	95.0
(語頭)	**56/1**	2/0	0/0	59	2	96.6
(語中)	**1/150**	1/4	1/3	160	9	94.4
/d/	2	**200**	1	203	3	98.5
(語頭)	0/0	**66/1**	1/0	68	1	98.5
(語中)	1/1	**3/130**	0/0	135	2	98.5
/g/	7	0	**253**	260	7	97.3
(語頭)	4/0	0/0	**64/0**	68	4	94.1
(語中)	0/3	0/0	**0/189**	192	3	98.4
平均	-	-	-	682	21	96.9

表4.2　語頭･語中別に作成したHMMによる/b/ /d/ /g/ の識別実験の混同表
(後続の母音の区間を15msec付加、話者:MAU)

HMM / 入力	/b/ (語頭/語中)	/d/ (語頭/語中)	/g/ (語頭/語中)	データ数	エラー数	識別率(%)
/b/	**210**	14	3	227	17	92.5
(語頭)	**53/1**	2/1	2/0	59	5	91.5
(語中)	**4/152**	0/11	0/1	168	12	92.9
/d/	2	**177**	0	179	2	98.9
(語頭)	0/0	**64/3**	0/0	67	0	100
(語中)	0/2	**3/107**	0/0	112	2	98.2
/g/	10	4	**238**	252	14	94.4
(語頭)	4/0	1/0	**62/0**	67	5	92.5
(語中)	0/6	0/3	**0/176**	185	9	95.1
平均	-	-	-	658	33	95.0

表4.2.c　語頭･語中別に作成したHMMによる/b/ /d/ /g/ の識別実験の混同表
(closeデータ、後続の母音の区間を15msec付加、話者:MAU)

HMM / 入力	/b/ (語頭/語中)	/d/ (語頭/語中)	/g/ (語頭/語中)	データ数	エラー数	識別率(%)
/b/	**211**	2	6	219	8	96.3
(語頭)	**57/1**	0/0	1/0	59	1	98.3
(語中)	**1/152**	0/2	2/3	160	7	96.5
/d/	0	**203**	0	203	0	100
(語頭)	0/0	**68/0**	0/0	68	0	100
(語中)	0/0	**3/132**	0/0	135	0	100
/g/	3	3	**254**	260	6	97.7
(語頭)	**2/0**	**1/0**	**65/0**	68	3	95.6
(語中)	0/1	0/2	**0/189**	192	3	98.4
平均	-	-	-	682	14	97.9

7. 後続母音別に作成したHMMによる/b/ /d/ /g/の識別実験 (実験4)

7.1 実験方法 より高精度に音韻を表現するために各音韻に対しコンテキスト別にHMMを作成する試みがなされている。[4][5]

ここでは、各破裂音に対して後続母音別にHMMを作成し、識別実験を行った。実験に使ったHMMの構造は図7に示す様に最後の一組の弧をタイドアークとしたHMMである。実験は次の3通りを行った。

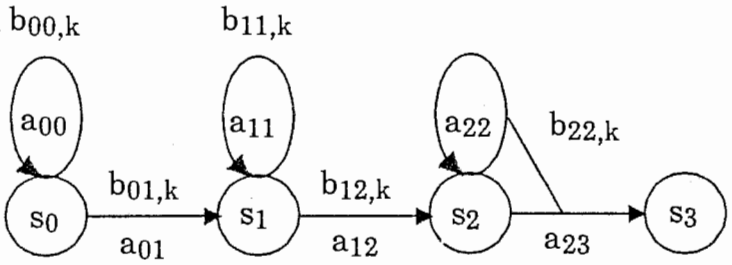


図7 最後の一組の弧をタイドアークとしたHMM

①学習･評価データ共に破裂音の区間のみを用いて後続母音別に作成したHMMによる識別実験。

②学習･評価データ共に破裂音の区間に後続母音の区間を15msec付加したデータを用いて後続母音別に作成したHMMによる識別実験。

③学習･評価データ共に破裂音の区間のみを用いて後続母音別に作成したHMMの遷移確率a_{ij}、出力確率$b_{ij,k}$と、同様のデータで後続母音を分けずに作成したHMMの遷移確率a_{ij}、出力確率$b_{ij,k}$との荷重平均をとって作成したHMM(composite model)による識別実験。これは学習データ数が少ない時のHMMの出力確率と遷移確率の平滑化の方法として文献[4][5]に上げられている方法を参考にしたものである。

遷移確率a_{ij}と出力確率$b_{ij,k}$の荷重平均は以下の様にとった。

composite modelの遷移確率と出力確率をa'_{ij}, $b'_{ij,k}$, 後続母音別に作成したHMMの遷移確率と出力確率をa_{ij}, $b_{ij,k}$, 後続母音を分けずに作成したHMMの遷移確率と出力確率を$\overline{a}_{ij}$, $\overline{b}_{ij,k}$, と書くと、

○遷移確率

$a'_{00} = \overline{a}_{00}$, $a'_{01} = \overline{a}_{01}$, $a'_{11} = \overline{a}_{11}$, $a'_{12} = \overline{a}_{12}$,

$a'_{22} = \omega * a_{22} + (1 - \omega) * \overline{a}_{22}$, $a'_{23} = \omega * a_{23} + (1 - \omega) * \overline{a}_{23}$,

○出力確率

$b'_{00,k} = \overline{b}_{00,k}$, $b'_{01,k} = \overline{b}_{01,k}$, $b'_{11,k} = \overline{b}_{11,k}$, $b'_{12,k} = \overline{b}_{12,k}$,
$b'_{22,k} = \omega * b_{22,k} + (1 - \omega) * \overline{b}_{22,k}$

ここでωは混合比を決める定数であり、$\omega = 0.25, 0.5, 0.75$ の3通りについて実験を行った。

7.2 実験結果　表5.1 に破裂音の区間のみを用いた実験①の結果を示す。後続母音で分けていない実験の結果である表2.1 と比較して、closeデータでは識別率が向上しており、後続母音別にHMMを作成することによってより高精度に各破裂音が表現されていることがわかる。一方、openデータでは逆に識別率が落ちている。これは後続母音別にHMMを分けたことによる学習データ数の不足が原因と考えられる。

表5.2 に後続母音の区間を15msec付加したデータを用いた実験②の結果を示す。後続母音で分けていない実験の結果である表2.2 と比較して、この実験においてもcloseデータでは識別率が向上し、openデータでは落ちる傾向が見られた。

表6 に実験③すなわち、composite modelにおける混合比ωを変化させたときの識別実験の結果を示す。また、表7に$\omega = 0.25$ としたときの識別結果の混同表を示す。単に後続母音別に作成したHMMを用いた実験①と比較して識別率の向上が見られるが、後続母音で分けていないものとは同程度の識別率しか得られていない。また、各混合比において識別率はほとんど変わらないことがわかる。

以上の様な検討結果から、後続母音別にHMMを作成することによってcloseデータに対してはより高精度に各破裂音が表現されるが、openデータに対しては学習データ数の不足が問題となることがわかった。また、出力確率と遷移確率の平滑化の方法としてcomposite modelによる識別実験を行ったが後続母音で分けていないものと同程度の識別率しか得られなかった。

表5.1　後続母音別に作成したHMMによる/b/ /d/ /g/の識別実験の混同表
(破裂音の区間のみ、話者:MAU)

HMM / 入力	bi(39)	be(21)	ba(53)	bo(43)	bu(63)	de(29)	da(90)	do(84)	gi(31)	ge(57)	ga(96)	go(37)	gu(39)	データ数	エラー数	識別率
bi	23	2	0	1	2	0	0	1	0	5	0	0	0	34	6	82.4
be	2	11	2	1	2	3	0	0	0	0	0	0	0	21	3	85.7
ba	2	0	47	2	10	2	3	0	0	0	0	0	0	66	5	92.4
bo	0	2	3	25	9	0	1	2	0	0	0	1	0	43	4	90.7
bu	0	3	10	17	16	1	8	7	0	0	0	0	0	63	17(1)	73.0
b														227	35(1)	84.6
de	0	0	0	0	1	15	8	3	0	0	0	0	0	27	1	96.3
da	1	0	1	0	2	8	60	6	0	0	3	0	0	81	7	91.4
do	0	0	0	3	5	0	3	60	0	0	0	0	0	71	8	88.7
d														179	16	91.1
gi	2	1	0	0	1	1	0	0	19	7	0	0	0	31	5	83.9
ge	4	0	0	0	1	2	0	0	7	33	1	0	5	53	7	86.8
ga	0	1	0	0	0	7	1	0	0	2	74	4	7	96	9	90.6
go	0	0	0	2	1	0	0	2	1	3	8	17	7	41	5	87.8
gu	2	0	4	0	0	1	3	0	1	4	7	3	6	31	10	67.7
g														252	36	85.7
bdg														658	87(1)	86.8

表5.1.c　後続母音別に作成したHMMによる/b/ /d/ /g/の識別実験の混同表
(closeデータ、破裂音の区間のみ、話者:MAU)

HMM / 入力	bi(39)	be(21)	ba(53)	bo(43)	bu(63)	de(29)	da(90)	do(84)	gi(31)	ge(57)	ga(96)	go(37)	gu(39)	データ数	エラー数	識別率
bi	39	0	0	0	0	0	0	0	0	0	0	0	0	39	0	100
be	1	19	0	0	1	0	0	0	0	0	0	0	0	21	0	100
ba	0	0	51	2	0	0	0	0	0	0	0	0	0	53	0	100
bo	0	0	2	39	2	0	0	0	0	0	0	0	0	43	0	100
bu	0	2	6	9	38	0	3	4	0	0	0	0	1	63	8	87.3
b														219	8	96.3
de	0	1	0	0	0	28	0	0	0	0	0	0	0	29	1	96.6
da	0	0	0	0	0	3	85	2	0	0	0	0	0	90	0	100
do	0	0	0	0	1	1	3	79	0	0	0	0	0	84	1	98.8
d														203	2	99.0
gi	0	0	0	0	0	0	0	1	27	3	0	0	0	31	1	96.8
ge	1	0	0	1	0	0	0	0	2	52	0	0	1	57	2	96.5
ga	0	0	0	0	0	6	1	0	0	1	84	0	4	96	7	92.7
go	0	0	0	1	0	0	0	0	0	0	3	29	4	37	1	97.3
gu	1	0	3	0	0	0	0	1	1	1	3	1	28	39	5	87.2
g														260	16	93.8
bdg														682	26	96.2

表5.2　後続母音別に作成したHMMによる/b/ /d/ /g/の識別実験の混同表
(後続の母音の区間を15msec付加、話者:MAU)

HMM / 入力	bi(39)	be(21)	ba(53)	bo(43)	bu(63)	de(29)	da(90)	do(84)	gi(31)	ge(57)	ga(96)	go(37)	gu(39)	データ数	エラー数	識別率
bi	30	0	0	0	0	0	0	1	0	3	0	0	0	34	4	88.2
be	4	14	0	0	2	1	0	0	0	0	0	0	0	21	1	95.2
ba	0	0	58	0	3	0	4	1	0	0	0	0	0	66	5	92.4
bo	0	0	1	35	3	0	1	2	0	0	0	1	0	43	4	90.7
bu	0	0	5	6	38	0	1	13	0	0	0	0	0	63	14	77.8
b														227	28	87.7
de	0	0	0	0	0	19	7	1	0	0	0	0	0	27	0	100
da	0	0	1	0	0	3	72	2	0	0	3	0	0	81	4	95.1
do	0	0	0	3	3	0	1	64	0	0	0	0	0	71	6	91.5
d														179	10	94.4
gi	1	0	0	0	0	0	0	0	27	3	0	0	0	31	1	96.8
ge	2	0	0	0	0	0	0	0	3	43	1	0	4	53	2	96.2
ga	1	0	1	0	1	0	3	0	0	1	87	1	1	96	6	93.8
go	0	0	0	1	0	0	0	1	0	2	6	22	9	41	2	95.1
gu	1	0	1	2	4	0	1	0	0	4	5	2	11	31	9	71.0
g														252	20	92.1
bdg														658	58	91.2

表5.2.c　後続母音別に作成したHMMによる/b/ /d/ /g/の識別実験の混同表
(closeデータ、後続の母音の区間を15msec付加、話者:MAU)

HMM / 入力	bi(39)	be(21)	ba(53)	bo(43)	bu(63)	de(29)	da(90)	do(84)	gi(31)	ge(57)	ga(96)	go(37)	gu(39)	データ数	エラー数	識別率
bi	38	0	0	0	0	0	0	0	0	1	0	0	0	39	1	97.4
be	0	21	0	0	0	0	0	0	0	0	0	0	0	21	0	100
ba	0	0	53	0	0	0	0	0	0	0	0	0	0	53	0	100
bo	0	0	0	43	0	0	0	0	0	0	0	0	0	43	0	100
bu	0	0	3	3	55	0	0	1	0	0	1	0	0	63	2	96.8
b														219	3	98.6
de	0	0	0	0	0	29	0	0	0	0	0	0	0	29	0	100
da	0	0	0	0	0	3	86	1	0	0	0	0	0	90	0	100
do	0	0	0	0	0	0	0	84	0	0	0	0	0	84	0	100
d														203	0	100
gi	0	0	0	0	0	0	0	0	30	1	0	0	0	31	0	100
ge	2	0	0	0	0	0	0	0	1	52	1	0	1	57	2	96.5
ga	0	0	0	0	0	0	2	0	0	0	92	0	2	96	2	97.9
go	0	0	0	0	0	0	0	0	0	0	1	35	1	37	0	100
gu	0	0	1	0	1	0	0	0	0	2	2	2	31	39	2	94.9
g														260	6	97.7
bdg														682	9	98.7

表6　composite model における混合比の検討結果
(破裂音の区間のみ、話者:MAU)

混合比ω	識別率			
	b	d	g	平均
通常の後続母音別(実験①)	84.6	91.1	85.7	86.8
0.75	87.2	91.6	86.1	88.0
0.5	87.2	91.6	86.1	88.0
0.25	88.1	90.5	86.6	88.1
0(HMMを分けていない)	88.1	91.1	85.7	88.0

表7　composite modelによる/b/ /d/ /g/の識別実験の混同表(ω = 0.25)
(破裂音の区間のみ、話者:MAU)

HMM / 入力	bi(39)	be(21)	ba(53)	bo(43)	bu(63)	de(29)	da(90)	do(84)	gi(31)	ge(57)	ga(96)	go(37)	gu(39)	データ数	エラー数	識別率
bi	29	2	0	0	1	0	0	1	0	1	0	0	0	34	2	94.1
be	2	16	0	0	1	1	0	0	0	0	0	0	1	21	2	90.5
ba	2	2	47	4	5	0	5	1	0	0	0	0	0	66	6	90.9
bo	0	0	4	25	13	0	0	1	0	0	0	0	0	43	1	97.7
bu	1	3	11	13	19	2	6	7	0	0	0	0	0	63	16(1)	74.6
b														227	27(1)	88.1
de	0	0	0	0	1	21	2	3	0	0	0	0	0	27	1	96.3
da	2	0	1	0	2	18	51	6	0	0	1	0	0	81	6	92.6
do	0	0	1	2	7	1	0	60	0	0	0	0	0	71	10	85.9
d														179	17	90.5
gi	2	0	0	0	0	0	0	2	22	3	0	1	1	31	4	87.1
ge	3	1	0	0	1	0	0	0	7	38	2	0	1	53	5	90.6
ga	0	1	1	0	1	6	1	0	1	12	64	4	5	96	10	89.6
go	0	0	0	3	0	0	0	2	0	4	8	21	3	41	5	87.8
gu	3	0	2	1	0	1	2	1	0	7	7	3	4	31	10	67.7
g														252	34	86.5
bdg														658	78(1)	88.1

8. 話者3名に対する/b/ /d/ /g/ の識別実験 (実験5)

今までの実験1, 2, 3, 4は男性1名(MAU)に対して行ってきたが、他の2名の男性話者(MHT, MNM)についても識別実験を少し行ったので報告する。

7.1. 実験方法　学習·評価データ共に破裂音の区間に後続母音の区間を15msec付加したデータを用い、語頭·語中でHMMを分けないものと、分けたものの2通りの実験を行った。ここで用いたHMMの構造は図5における(a)の型であり、4状態、3ループを持つものである。

7.2. 実験結果　表8に話者3名分(MAU, MHT, MNM)の実験結果を示す。識別率は、話者や破裂音の違いによってバラツキがあるが、語頭·語中でHMMを分けた実験において話者3名の平均で94.4%の識別率を得た。

表9.1,　表 9.2 に話者MHTに対する識別実験の混同表を示す。
表10.1, 表10.2 に話者MNMに対する識別実験の混同表を示す。

表8.　話者3名に対する/b/ /d/ /g/ の識別実験の結果
(後続母音の区間を 15msec付加)

話者	識別率(%) (HMMは語頭·語中で共通)				識別率(%) (HMMを語頭·語中で分けた)			
	/b/	/d/	/g/	平均	/b/	/d/	/g/	平均
MAU	92.1	96.6	90.9	92.9	92.5	98.9	94.4	95.0
MHT	96.2	98.2	97.2	97.2	94.2	98.8	98.4	97.2
MNM	87.5	92.7	92.6	90.9	87.0	90.4	94.9	91.1
平均	91.9	95.8	93.6	93.6	91.2	96.0	95.9	94.4

表9.1　/b/ /d/ /g/の識別実験における混同表
(4状態3ループのHMM、後続母音の区間を15msec付加、話者:MHT)

HMM / 入力	/b/	/d/	/g/	データ数	エラー数	識別率(%)
/b/	**200**	7	1	208	8	96.2
(語頭)	**57**	2	0	59	2	96.6
(語中)	**143**	5	1	149	6	96.0
/d/	2	**167**	1	170	3	98.2
(語頭)	0	**66**	1	67	1	98.5
(語中)	2	**101**	0	103	2	98.1
/g/	5	2	**247**	254	7	97.2
(語頭)	4	1	**62**	67	5	92.5
(語中)	1	1	**185**	187	2	98.9
平均	-	-	-	632	18	97.2

表9.2　語頭·語中別に作成したHMMによる/b/ /d/ /g/の識別実験の混同表
(4状態3ループのHMM、後続の母音の区間を15msec付加、話者:MHT)

HMM / 入力	/b/ (語頭/語中)	/d/ (語頭/語中)	/g/ (語頭/語中)	データ数	エラー数	識別率(%)
/b/	**196**	9	3	208	12	94.2
(語頭)	**53/3**	2/0	1/0	59	3	94.9
(語中)	**2/138**	0/7	2/0	149	9	94.0
/d/	1	**168**	1	170	2	98.8
(語頭)	0/0	**64/2**	1/0	67	1	98.5
(語中)	1/0	**0/102**	0/0	103	1	99.0
/g/	3	1	**250**	254	4	98.4
(語頭)	1/0	0/0	**66/0**	67	1	98.5
(語中)	0/2	1/0	**2/182**	187	3	98.4
平均	-	-	-	632	18	97.2

表10.1　/b/ /d/ /g/の識別実験における混同表
(4状態3ループのHMM、後続母音の区間を15msec付加、話者:MNM)

HMM / 入力	/b/	/d/	/g/	データ数	エラー数	識別率(%)
/b/	**189**	15	12	216	27	87.5
(語頭)	**46**	6	7	59	13	78.0
(語中)	**143**	9	5	157	14	91.1
/d/	12	**165**	1	178	13	92.7
(語頭)	2	**64**	1	67	3	95.5
(語中)	10	**101**	0	111	10	91.0
/g/	14	5	**237**	256	19	92.6
(語頭)	6	4	**57**	67	10	85.1
(語中)	8	1	**180**	189	9	95.2
平均	-	-	-	650	59	90.9

表11.2　語頭･語中別に作成したHMMによる/b/ /d/ /g/の識別実験の混同表
(4状態3ループのHMM、後続母音の区間を15msec付加、話者:MNM)

HMM / 入力	/b/ (語頭/語中)	/d/ (語頭/語中)	/g/ (語頭/語中)	データ数	エラー数	識別率(%)
/b/	**188**	16	12	216	28	87.0
(語頭)	**48/4**	2/0	5/0	59	7	88.1
(語中)	**2/134**	1/13	2/5	157	21	86.6
/d/	12	**162**	5	178	17	90.4
(語頭)	2/0	**59/2**	4/0	67	6	91.0
(語中)	1/9	**2/98**	0/1	111	11	90.1
/g/	11	2	**243**	256	13	94.9
(語頭)	2/3	2/0	**60/0**	67	7	89.6
(語中)	1/5	0/0	**0/183**	189	6	96.8
平均	-	-	-	650	58	91.1

9. むすび

音韻を認識単位としたHMMの検討の一段階として成人男性1名の発声した5240単語中から切り出した有声破裂音/b//d//g/の識別実験を行いHMMの状態数、各状態間を結ぶ弧の構成法等について検討した。有声破裂音を表現するHMMの状態数としてはループを持った状態が3個以上必要なことがわかった。また後続母音の区間を15msec程度付加したデータを用いて学習・認識した方が破裂音の区間のみを用いるよりも高い識別率が得られることがわかった。

各状態間を結ぶ弧の構成法として、通常の弧、タイドアーク、ヌルアークは識別率において大差がなかった。但し、最後の状態にループを持つHMMでは、初期値として遷移確率と出力確率を各弧に対して一様な値を与えて学習を行った場合、状態の遷移が最終状態のループに偏る傾向があることがわかった。

次に後続母音別にHMMを作成して識別実験を行った。closeデータに対してはより高精度に各破裂音が表現されるが、openデータに対しては学習データ数の不足が問題となることがわかった。また、出力確率と遷移確率の平滑化の方法としてcomposite modelによる識別実験を行ったが後続母音で分けていないものと同程度の識別率しか得られなかった。

最後に、他の2名の男性話者を加えた3名の話者に対して、語頭と語中のデータで別々にHMMを作成し/b//d//g/の識別実験を行った。この結果、3名の話者で平均94.4%の識別率を得た。

参考文献

[1]Rabiner,L.R., Levinson,S.E. : "ASpeaker-Independent, Sytax-Directed, Connected Word Recognition System Based on Hidden Markov Models and Lebel Building", IEEE Trans. Acoust., Speech, Signal Processing vol. ASSP-33, No.3. (June 1985.)

[2]Sugawara,K., Nishimura,M., Toshioka,K., Okochi,M., and Kaneko,T. : "Isolated word Recognition Using Hidden Markov Models", ICASSP85 (March 1985)

[3]菅原,大河内,年岡,金子:『遷移を制限したマルコフモデルを用いた音声認識』,音講論集(昭和59年10月)

[4]Schwartz,R., Chow, Roucos,S., Kransner,M., Makhoul,J. : "Improved Hidden Markov Modeling of Phonemes for Continuous Speech Recoginition", ICASSP84 (1984)

[5]Schwartz,R., Chow,Y., Kimball,O., Roucos,S., Kransner,M., Makhoul,J. : "Context-Dependent Modeling for Acoustic-Phonetic Recognition of Continuous Speech", ICASSP85 (March 1985)

[6]武田,匂坂,片桐:『音声データベース構築のための音韻ラベリング』,音講論集(昭和62年3月)

[7]Levinson,S.E, Rabiner,L.R., and Sondhi,M.M. : "An Introduction to the Application of the Theory of Probabilistic Functions of a Markov Process to Automatic Speech Recognition", The Bell System Technical Journal, vol.62, No.4. (April 1983.)