

TR-IT-0094

自由発話音声からの HMnet の構造生成による  
音素認識の改善

山本 一公 小坂 哲夫

Kazumasa YAMAMOTO, Tetsuo KOSAKA

概要

HMnet の構造が発話形式に依存するかどうかは不明である。そこで、スポンテニアスデータの認識における、単語データから生成された HMnet とスポンテニアスデータから生成された HMnet との比較実験を行った。

スポンテニアスデータで HMnet を生成する場合、問題になると考えられるのがスポンテニアスデータの詳細なラベル付けについてと、スポンテニアスデータの音素バランスについてである。

今回は、構造はバランス単語データで作成し、パラメータだけをスポンテニアスデータで学習する方法、初期構造を単語データで作成しておき、途中からの状態分割をスポンテニアスデータを用いて行う方法、初期モデルでスポンテニアスデータに対するビターバイメントを取ることによって、それを詳細なラベルの代替として始めから状態分割を行う方法の 3 通りについてそれぞれ検討を行った。

©ATR. 音声翻訳通信研究所

©ATR. Interpreting Telecommunications Research Labs.

## 目次

1	はじめに	1
2	背景	1
2.1	スポンテニアス音声における問題点	1
2.2	使用するデータベースの問題点	2
2.3	実験条件の問題点	2
3	実験方法	3
3.1	予備実験	3
3.2	実験 1	3
3.3	実験 2	3
3.4	実験 3	5
4	結果	5
4.1	実験 1	5
4.2	実験 2	5
4.3	実験 3	7
4.4	考察	7
5	まとめ	9
6	今後の課題	9
7	プログラムについて	10
7.1	プログラムおよびディレクトリ説明	10
7.2	使用方法	11

## 表目次

1	スポンテニアスデータベースの種類	2
2	実験条件	3
3	学習用スポンテニアスデータ	4
4	認識用スポンテニアスデータ	4
5	予備実験の結果(初期モデル)	7
6	予備実験の結果(パラメータのみ学習)	7
7	実験 1 の結果	8
8	実験 2 の結果(1 回目の繰り返し)	8
9	実験 2 の結果(2 回目の繰り返し)	8
10	実験 2 の結果(3 回目の繰り返し)	8
11	実験 3 の結果	8
12	2620 単語で生成した HMnet の allophone	9
13	216 バランス単語・150 バランス文で生成した HMnet の allophone	9
14	実験 2 の 1 回目の繰り返しで生成された HMnet の allophone	9
15	実験 2 の 2 回目の繰り返しで生成された HMnet の allophone	9
16	実験 2 の 3 回目の繰り返しで生成された HMnet の allophone	9

## 図目次

1	隠れマルコフ網(HMnet)の例	1
2	実験方法 1	5
3	実験方法 2	6
4	実験方法 3	6

## 1 はじめに

近年、音声情報処理技術が急速に発展し、音声認識分野での研究の関心事は、大語彙、連続音声、不特定話者へと発展しつつある。さらに、認識タスクが高度化するにつれて、隠れマルコフモデル (Hidden Markov Model, HMM) に代表される確率的モデルが重要な手法となっている。本研究では、逐次状態分割法 (Successive State Splitting, SSS) により生成される、HMM の 1 つである隠れマルコフ網 (Hidden Markov Network, HMnet) のモデルを使用した (図 1)。

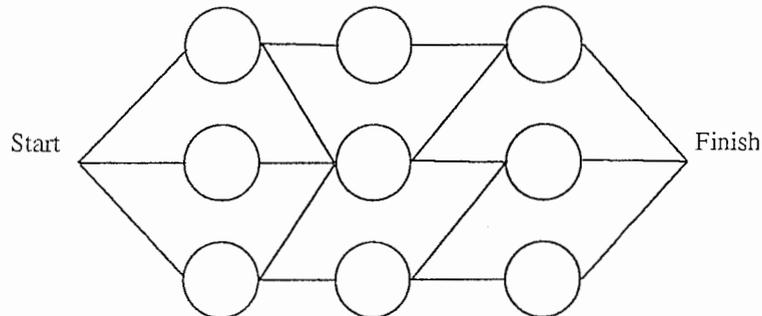


図 1: 隠れマルコフ網 (HMnet) の例

従来、HMnet の構造の生成は単語もしくは文節単位のデータで行われてきた。しかし、自由発話 (Spontaneous Speech) の認識を行う場合は、単語音声によって生成された HMnet で十分な認識率が得られるのか、それともスポンテニアス音声で HMnet を生成した方が認識率が向上するのか不明である。そこで、スポンテニアス音声を用いて HMnet を生成するいくつかの実験を行った。

スポンテニアス音声で HMnet を生成する場合に問題になると考えられるのが、スポンテニアス音声の詳細なラベル付けについてである。従来、HMnet の生成に用いられてきた単語音声は、詳細なラベル付けの行われたものである。しかし、大量のスポンテニアス音声に対して、同様に詳細なラベル付けを行うには膨大な時間を要するため、効率的ではない。もう一つの問題は、的確な状態分割が行われているかどうか、ということである。スポンテニアス音声の出現音素が偏っていれば的確な状態分割は行われず、結果出来上がる構造は偏ったものになってしまう十分な認識率は得られない。

以上のことから今回は、構造は音素バランス単語データで作成し、パラメータだけをスポンテニアスデータで学習する方法、初期構造を単語データで作成しておき、途中からの状態分割をスポンテニアスデータを用いて行う方法、初期モデルに対してスポンテニアスデータのビターピアライメントを取ることによって、それを詳細なラベルの代替として最初から状態分割を行う方法の 3 通りについてそれぞれ検討を行った。

## 2 背景

スポンテニアスデータを用いて HMnet を生成する場合、問題になるものとして次のようなものが考えられる。

### 2.1 スポンテニアス音声における問題点

単語データや読み上げデータに対して、スポンテニアスデータには、次のような問題点がある。

**言語的問題** スポンテニアスデータでは、文法的な崩れ、助詞落ち、倒置などといった言語的な変動が起りやすく、これに対処するためには、こういった変動を処理できる文法を作成しなければならない。また、「あの一」「えーと」などといった間投詞 (冗長語) や言い直し、言い誤りといった問題は音素変形の可能性を含んでいるため、音響モデルとも関係してくる。

**音響的問題** 音響的な問題の際たるものは、音素の脱落・変形などである。この中には、文尾の音素が不明瞭になる、母音が無音化してしまうことがある、音素カテゴリーの範疇外の音声が出現することがある、音素が脱落してしまう、調音結合の影響が大きく出る、といったようなことが含まれる。

また、言い淀み、強調、ポーズの長短などにより発話速度に変動が見られることもある。これは、Duration モデルに影響を与える。

他には、二人の話者の音声重なってしまう、音声区間の重畳という問題もあるが、現在のところデータベースでは扱っていない。

副次的に発生する問題 対話はデータ収録を中断することができないので、途中で失敗すると始めから対話をやり直す必要がある。そのため、多少の雑音(舌うち、紙やペンの音、息など)の混入が通常認められている。これらに対しては、非音声モデルを作成するなどして対処する必要がある。

今回の実験では、音響的問題の音素の脱落・変形の中で、HMnetのパラメータだけでは対処できないスペクトル変形を、構造を生成することによって対処することを考える。その他の問題については今回は考慮しない。

## 2.2 使用するデータベースの問題点

スポンテニアスデータのデータベースには表1に示されるように2種類のものがある。データベースの内容は、

- プロットにもとづく模擬対話。
- 発話の制約として、両話者の発話は重ならない。また発話は10秒以内。
- タスクは旅行に関する対話。

である。今回使用したのは音声言語データベースであり、通訳者を介して録音を行ったものであるため、スポンテニアス度は低い。

表 1: スポンテニアスデータベースの種類

	音声言語データベース	音声データベース
利用目的	音声翻訳(音声認識と翻訳)	音声認識
対話形態	日本語話者と英語話者の通訳者を介する対話	日本語話者同士の対話
収録環境	良好	やや雑音が多い
スポンテニアス度	低い	高い

## 2.3 実験条件の問題点

実験を行う際に比較しなければならない条件は、以下のようなものである。

話者	1話者 ⇔ 多数話者
ラベルデータ	詳細ラベル(ハンドラベル) ⇔ ビタービアライメント
構造生成用データ	単語音声 ⇔ 読み上げ音声 ⇔ 自由発話

従来HMnetは、話者は1話者、ラベルデータは詳細ラベルで生成されており、多数話者やビタービアライメントの影響が考察されていない。本来ならば、1話者、詳細ラベルで、構造生成用データの条件だけを変えて実験を行い、構造生成用のデータが構造にどのような影響を及ぼすかの検討を行うべきである。しかし今回の実験では、1話者でHMnetの構造を生成するだけの十分なデータを確保できなかったために多数話者を採用し、詳細ラベルも数量が不足しているためビタービアライメントをその代替として用いることとした。そのため、実験条件の話者・ラベル・構造生成用データの3つの問題がすべて混合されてしまい、実験結果に対しての考察が非常にやりにくくなってしまった。

今後、スポンテニアスにおける詳細ラベルが付加された時点で、問題を切り分けて実験する必要がある。

### 3 実験方法

実験条件を表 2 に、実験に使用したスポンテニアスデータの内わけを表 3,4 に示す。

表 2: 実験条件

分析条件	
Sampling-rate	12kHz (48kHz)
Window	Hamming window (20ms)
Frame period	5ms
Analysis	log power + 16-order LPC-Cep + $\Delta$ log power + 16-order $\Delta$ LPC-Cep
HMnet	
各状態の分布	混合ガウス分布 (混合数 : 5)、対角共分散行列
コンテキストの要因数	3 要因 (先行音素、当該音素、後続音素)
時間方向への状態分割制限	1 モデル当たりの状態数を最大 4 に制限

各実験の評価は次式で算出される認識率 (音素正解率) で行うものとする。

$$Acc = \frac{N_{phones} - N_{sub} - N_{ins} - N_{del}}{N_{phones}}$$

ここで、 $N_{phones}$  は正解音素系列中の総音素数、 $N_{ins}$  は音素挿入誤り数、 $N_{del}$  は音素脱落誤り数、 $N_{sub}$  は音素置換誤り数を表す。

実験は予備実験を含めて以下の 4 通りの方法で行う。

1. 予備実験として、初期モデルをそのまま使う場合と、初期モデルの構造を使ってパラメータだけをスポンテニアスデータで学習した場合の 2 通りの認識実験を行う。
2. 単語音声と音素バランス文で HMnet の構造を生成しておき、パラメータだけをスポンテニアスデータで学習する。
3. 初期モデルに対してスポンテニアスデータのピタービアライメントを取り、それを詳細ラベルの代わりとして、1 状態から状態分割を開始する。
4. 単語音声で HMnet の構造を 200 状態まで生成しておき、その後の状態分割はスポンテニアスデータで行う。

#### 3.1 予備実験

予備実験として、1 話者 (MHT) の 2620 単語 (音素数 15175、フレーム数 321005) で構造を作成したものに、2250 文 (話者 15 名で各話者 150 文) のデータ (音素数 159334、フレーム数 2834791) でパラメータを学習した HMnet を用いて、スポンテニアス音声の認識実験を行う。

また、その HMnet の構造を用いて 10 話者の 473 発声からなるスポンテニアスデータ (音素数 18779、フレーム数 265734) でパラメータだけを学習した HMnet を作成し、同様の認識実験を行う。

#### 3.2 実験 1

単語音声の HMnet は、1 話者 (M103) の 216 音素バランス単語と音素バランス文 150 文 (A,B,Cset 各 50 文) (単語と文を合わせて、音素数 10030、フレーム数 146367) で構造を作成したものである。状態数は 100,200,300,400,500,600 の 6 種類である。これのパラメータを 10 話者の 473 発声からなるスポンテニアスデータで学習する。

#### 3.3 実験 2

基本モデルとなる単語音声の HMnet は、1 話者 (MHT) の 2620 単語で構造を作成したものに、2250 文 (話者 15 名で各話者 150 文) のデータでパラメータを学習した HMnet である。この HMnet で、10 話者の 473 発声からなる

表 3: 学習用スポンテニアスデータ

話者	会話 ID	種別	発声数
FEMMA	TCC22074	B	9
	TAC22013	A	19
FHITA	TAS22001	A	12
	TAS22002	A	15
	TAS23002	B	24
FHITO	TAC22011	A	22
FMAKA	TCC23062	B	9
FMITE	TAC23014	B	33
	TAS32001	A	17
FYOAS	TAS32002	A	12
	TAS33002	B	24
FYOGO	TCC22031	A	9
MHIKO	TAC22014	A	15
	TAS12001	A	22
	TAS12002	A	19
	TBS12001	A	8
	TGS12001	A	14
	THS12001	A	14
	THS12002	A	14
MMASE	TRS12001	A	17
	TSS12001	A	18
	TSS12002	A	19
	TAS13002	B	27
	THS13001	B	39
	TRS13002	B	23
	MSHAK	TAS33014	B
	合計		473

表 4: 認識用スポンテニアスデータ

話者	会話 ID	種別	発声数
FYUYO	TAS12009	A	13
FYOMA	TAS12010	A	8
MMAKO	TAS32007	A	12
MNOSA	TAS32009	A	12
	合計		45

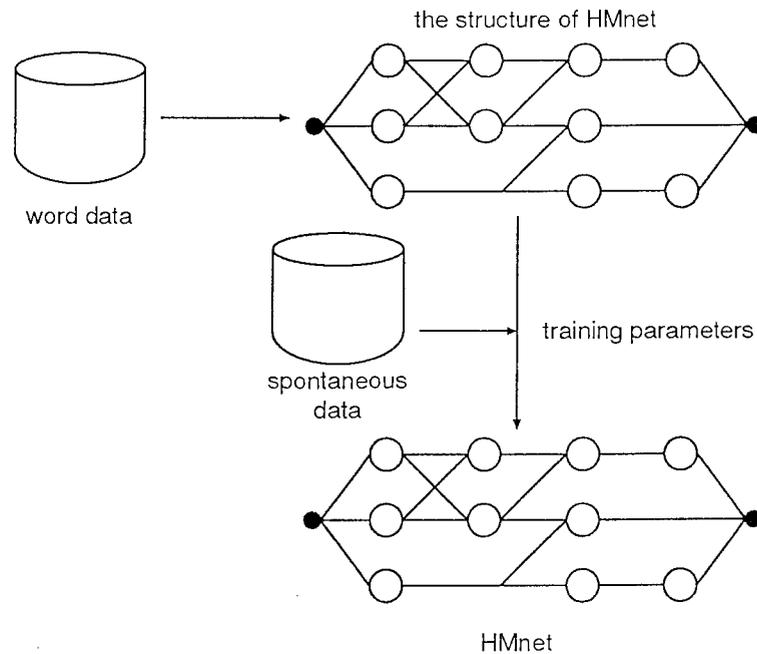


図 2: 実験方法 1

スポンテニアスデータのビターピアライメントを取り、そのデータを詳細ラベルの代わりに用いて SSS により HMnet を生成する。状態数は 100,200,300,400,500,600 の 6 種類である。HMnet のパラメータの学習も同じスポンテニアスデータで行う。

さらに、新しく生成された HMnet を基本モデルとすることで、繰り返し HMnet の生成を行う。

### 3.4 実験 3

単語音声の HMnet は、1 話者 (MHT) の 2620 単語で構造を作成したものであり、これのパラメータを 10 話者の 473 発声からなるスポンテニアスデータで学習する。そのパラメータを学習した HMnet で、同じスポンテニアスデータのビターピアライメントを取り、それを詳細ラベルの代わりとして 200 状態以後の状態分割を行う。よって状態数は 300,400,500,600 の 4 種類である。最終的なパラメータの学習も同じスポンテニアスデータで行う。

## 4 結果

表 5 と表 6 に予備実験、表 7 に実験 1、表 8～10 に実験 2、表 11 に実験 3 の音素認識実験の結果をそれぞれ示す。

### 4.1 実験 1

実験 1 は、予備実験と比較するとかなり認識率が悪くなっている。予備実験のパラメータをスポンテニアスデータで学習したものに比べても、かなり悪い。予備実験のモデルは音素バランス単語データで作成したものではなく、この実験のモデルは音素バランス単語と音素バランス文で作成したものである。パラメータは同じデータで学習されているので、認識率に影響を与えたのが構造であることは容易に考えられる。予備実験で使用したデータの話者とこの実験で使用した話者が異なるため、話者間の相違の問題もあるが、スポンテニアス音声では、日常会話中での出現頻度の高い音素がそのまま出てくるため、音素バランスは考慮しない方が好影響を与えている、ということも考えられる。

### 4.2 実験 2

実験 2 は、時間の関係上 3 回しか繰り返して実験を行うことが出来なかった。実験の結果、認識率は 1 回目で基本モデルより下がり、2 回目では 1 回目よりも上がり、3 回目ではまた 2 回目よりも下がった。この後、繰り返していくと認識率がどのように変動していくか、というのは今回の実験結果だけでは判断できない。

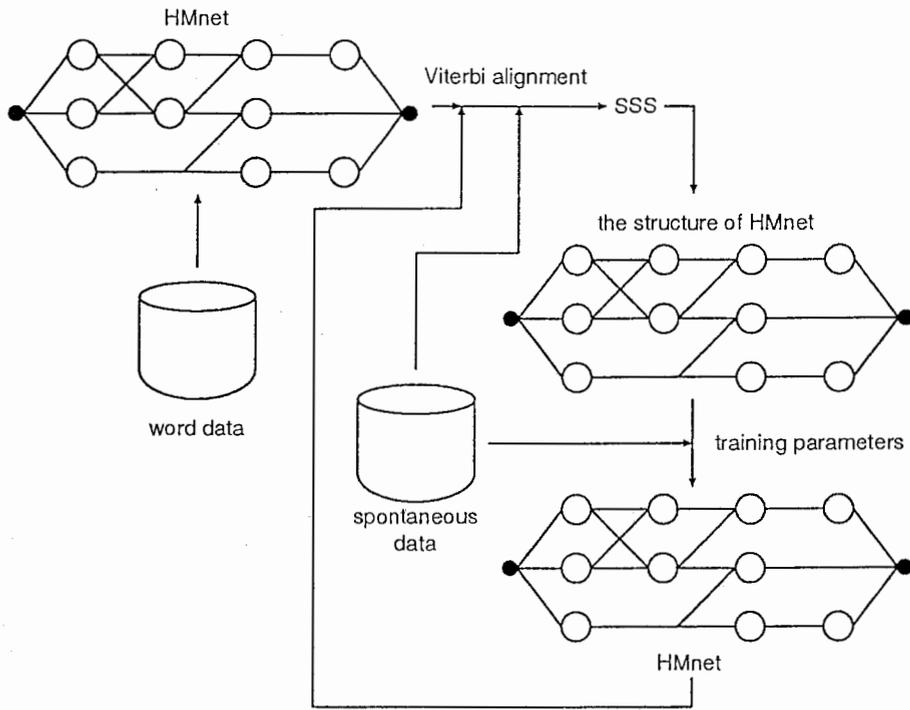


図 3: 実験方法 2

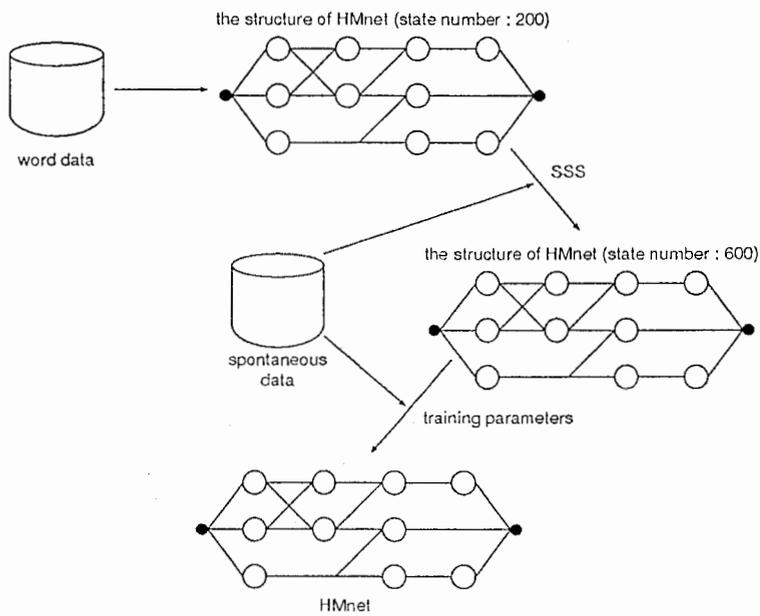


図 4: 実験方法 3

### 4.3 実験 3

実験 3 は、200 状態までの構造を生成した単語データに含まれていた音素環境がスポンテニアスデータに含まれていなかったために、状態分割が途中から出来なくなってしまった(正確には、状態分割中に推定されているガウス分布の平均や分散が 0 になってしまう現象が起こった)ため、状態数 300 のものしか生成されず、認識実験もそれだけで行った。状態分割が途中から出来なくなる問題を解決するには、状態分割プログラムそのものを変更する必要があるが、今回は時間の関係上それは出来なかった。

考察ができるだけの十分な実験結果を得ることが出来なかったため、この方法がどれほどの効果を持つかはわからなかった。

### 4.4 考察

考察として、今回実験で使用した HMnet の allophone を調べてみた。表 12～表 16 にそれを示す。HMnet はすべて状態数 200 のものについて調べた。

基本モデルとして使用した 2620 単語の HMnet と比較して今回生成した HMnet は、a,i,n,o が増加して、e,k,r が減少する傾向があるように見える。

認識実験の認識率が良くならない原因として考えられるのが、多数話者の影響である。1 話者では、状態分割はコンテキストの影響によって行われるが、多数話者ではコンテキストの影響以上に話者の影響が大きいとそれによって状態分割が行われてしまう。そうなるると的確な構造が生成されなくなり、そのことが認識率に影響を与えていると考えられる。

また、実験全体の結果を見ると、状態数が少ないほど認識率が良い傾向が見られる。本来ならば、状態数が多い詳細なモデルの方が認識率は上がるはずである。このことから、全体に学習データが不足していると考えられる。

表 5: 予備実験の結果(初期モデル)

タスク名	種別	話者	正解率	挿入率	脱落率	置換率
TAS12009	A	FYUYO	60.6	8.4	3.8	27.2
TAS12010	A	FYOMA	62.5	7.6	4.4	25.5
TAS32007	A	MMAKO	65.0	6.2	4.3	24.5
TAS32009	A	MNOSA	70.6	4.0	4.0	21.4
平均			64.9	6.5	4.1	24.6

表 6: 予備実験の結果(パラメータのみ学習)

タスク名	種別	話者	正解率	挿入率	脱落率	置換率
TAS12009	A	FYUYO	58.9	14.1	3.4	23.6
TAS12010	A	FYOMA	63.6	13.5	3.3	19.6
TAS32007	A	MMAKO	64.3	9.4	4.2	22.1
TAS32009	A	MNOSA	65.7	8.0	3.1	23.2
平均			63.1	11.0	3.5	22.4

表 7: 実験 1 の結果

タスク	種別	話者	状態数						平均
			100	200	300	400	500	600	
TAS12009	A	FYUYO	49.1	48.4	42.5	47.6	45.9	42.3	46.0
TAS12010	A	FYOMA	52.7	52.7	48.7	51.3	44.7	38.9	48.2
TAS32007	A	MMAKO	57.4	51.3	53.1	50.7	50.9	50.0	52.2
TAS32009	A	MNOSA	59.0	59.2	55.7	57.2	52.6	42.1	54.3
状態数平均			54.8	52.8	50.2	51.6	49.0	44.1	50.4

表 8: 実験 2 の結果 (1 回目の繰り返し)

タスク	種別	話者	状態数						平均
			100	200	300	400	500	600	
TAS12009	A	FYUYO	58.1	57.5	58.8	49.3	44.8	39.8	50.2
TAS12010	A	FYOMA	58.5	54.5	55.3	52.4	50.5	49.5	53.5
TAS32007	A	MMAKO	55.6	58.0	60.1	56.2	58.5	54.9	57.1
TAS32009	A	MNOSA	62.1	60.4	66.6	59.7	63.3	63.9	62.7
状態数平均			58.4	57.9	58.8	54.6	54.6	52.3	56.1

表 9: 実験 2 の結果 (2 回目の繰り返し)

タスク	種別	話者	状態数						平均
			100	200	300	400	500	600	
TAS12009	A	FYUYO	61.7	61.7	56.0	61.3	59.4	50.5	59.6
TAS12010	A	FYOMA	58.2	61.1	60.0	54.9	55.6	49.1	57.2
TAS32007	A	MMAKO	60.9	62.1	63.8	62.1	64.9	60.7	62.7
TAS32009	A	MNOSA	70.8	66.6	69.3	67.9	68.8	62.8	68.6
状態数平均			63.2	63.0	62.5	62.3	62.9	56.7	62.5

表 10: 実験 2 の結果 (3 回目の繰り返し)

タスク	種別	話者	状態数						平均
			100	200	300	400	500	600	
TAS12009	A	FYUYO	58.1	55.4	47.6	49.7	45.5	47.8	50.7
TAS12010	A	FYOMA	56.0	56.0	55.3	49.1	50.2	48.7	52.5
TAS32007	A	MMAKO	63.9	64.5	60.9	60.9	60.0	58.5	61.4
TAS32009	A	MNOSA	61.7	65.5	63.3	68.4	63.9	65.7	64.7
状態数平均			60.5	60.9	57.0	57.9	55.5	55.9	58.0

表 11: 実験 3 の結果

タスク名	種別	話者	正解率	挿入率	脱落率	置換率
TAS12009	A	FYUYO	54.9	14.9	2.1	28.0
TAS12010	A	FYOMA	61.8	12.4	2.5	23.3
TAS32007	A	MMAKO	62.9	11.2	2.9	23.0
TAS32009	A	MNOSA	67.5	7.8	3.1	21.6
平均			61.7	11.5	2.7	24.0

表 12: 2620 単語で生成した HMnet の allophone

-	a	b	ch	d	e	g	h	i	j	k	m	n	ng	o	p	q	r	s	sh	t	ts	u	w	z	zh
1	140	1	1	1	60	1	1	60	14	28	4	1	4	64	1	1	24	4	1	3	1	60	1	1	1

表 13: 216 バランス単語・150 バランス文で生成した HMnet の allophone

-	a	b	ch	d	e	g	h	i	j	k	m	n	ng	o	p	q	r	s	sh	t	ts	u	w	z	zh
1	153	1	1	1	37	2	1	109	12	19	3	12	6	159	1	1	7	2	1	6	1	116	1	1	1

表 14: 実験 2 の 1 回目の繰り返しで生成された HMnet の allophone

-	a	b	ch	d	e	g	h	i	j	k	m	n	ng	o	p	q	r	s	sh	t	ts	u	w	z	zh
1	174	1	1	4	64	1	1	76	12	9	12	15	1	290	1	1	11	5	1	5	1	62	1	1	1

表 15: 実験 2 の 2 回目の繰り返しで生成された HMnet の allophone

-	a	b	ch	d	e	g	h	i	j	k	m	n	ng	o	p	q	r	s	sh	t	ts	u	w	z	zh
1	148	1	1	2	46	1	1	145	6	27	4	12	1	216	1	1	14	4	1	4	1	37	1	1	1

表 16: 実験 2 の 3 回目の繰り返しで生成された HMnet の allophone

-	a	b	ch	d	e	g	h	i	j	k	m	n	ng	o	p	q	r	s	sh	t	ts	u	w	z	zh
1	250	1	1	2	24	1	1	96	10	9	12	7	1	135	1	1	10	2	1	6	1	30	1	1	1

## 5 まとめ

単語音声、読み上げ音声、スポンテニアス音声の各データでそれぞれ生成した HMnet の構造を用いて認識率の比較を行った。

認識結果としては、あまり良好な結果を得ることは出来なかった。

結果の良くない原因は、2.3節で述べたように、複数の要因が絡み合っているものと考えられ、また、実験結果から明確な傾向を見出すことができないので、現時点で結論を出すことは出来ない。しかし、HMnet の構造を多数話者で生成していることの影響は大きいと考えられる。

また、全体に状態数の少ない HMnet の方が認識率が高いことから、学習に用いているデータ量が十分でないと思われる。

## 6 今後の課題

全体にデータの量が少ない傾向が見られたので、データの量を増やして実験を行う必要がある。

また、1話者の詳細ラベルデータでの実験を行うことで、実験の条件比較を明確に行う必要がある。

### 謝辞

研究の機会を与えていただいた豊橋技術科学大学情報工学系計算機大講座中川聖一教授に感謝いたします。また、御指導いただいた ATR のみなさまに感謝いたします。

## 7 プログラムについて

この実験に使用したプログラムの説明と使用方法を示す。

## 7.1 プログラムおよびディレクトリ説明

```
~/bin/
  cxtlist.awk      学習データの CXT ファイルのリストを作る (実験 1)
  getcxtfile.sed  作成した CXT ファイルを格納するファイル名を生成する (実験 2)
  make_flist.csh  使用するスポンテニアスデータのファイルリストを生成する
  (make_flist.awk)                                     (実験 2)
  make_label.awk  Exe.viterbi_HMnet の出力を CXT ファイルに変換する (実験 2)
  make_result.csh 認識結果を状態数ごとに集計する
  (make_result.awk,make_table.awk)
  make_sample.csh スポンテニアス音声のパラメータ学習用データを作る

~/SSS/Work/
  train_spo.csh   繰り返し学習用シェル (実験 2)

~/SSS/Work/Recog/
  comp_exp.csh          初期モデルの構造でパラメータをスポンテニアス
  (Recog_comp.csh,mklist.csh) データで学習する予備実験の学習認識用シェル
  noadapt.csh          認識実験用シェル (実験 2)
  (Recog.csh,mklist.csh)
  noadapt_init.csh    初期モデルの認識実験 (予備実験) 用シェル
  (Recog.csh,mklist.csh)

/data/atrq06/xyamamot/
  Work/train.csh      HMnet の構造生成、学習用シェル (実験 1)
  Work/Recog/noadapt.csh 認識実験用シェル (実験 1)
  New/train.csh      HMnet の構造生成、学習用シェル (実験 3)
  New/Recog/recog.csh 認識実験用シェル (実験 3)

~/SSS/Work/          (実験 2)
  flist/             スポンテニアスデータのファイルリスト
  CXT/$n/$task      $n 回目の繰り返しで作成される CXT ファイル
  Spontaneous/$n/   $n 回目の繰り返しで作成される HMnet
  Model/Topology/mix_5/ 抽出した HMnet の構造
  Model/Spontaneous/mix_5/ パラメータ学習後の HMnet
  Recog/Score/$n/$state_num $n 回目の $state_num 状態のモデルの認識結果

/data/atrq06/xyamamot/
  Work/             学習用データなどの格納ディレクトリ
  HMnet/Model/     (実験 1)
  Original/        SSS により生成された HMnet
  Topology/mix_5   抽出した HMnetno 構造
  Spontaneous/mix_5 パラメータを学習した HMnet
  Recog/Score/$state_num $state_num 状態のモデルの認識結果

New/              (実験 3)
  CXT/             生成される CXT ファイル
```

```

Spontaneous/
  Topology/mix_5          構造を抽出した HMnet(混合数:5)
  Spontaneous/mix_5      パラメータを学習した HMnet
                          (ビターピアライメントを取るためのもの)
  Topology/mix_2          構造を抽出した HMnet(混合数:2)
  Spontaneous/mix_2      パラメータを学習した HMnet
                          (状態分割を再開するためのもの)

Model/
  Topology/mix_5          構造を抽出した HMnet(混合数:5)
  Spontaneous/mix_5      パラメータを学習した HMnet

Recog/
  Score/$state_num       $state_num 状態のモデルの認識結果

```

## 7.2 使用方法

- 予備実験(初期モデルをそのまま使用して、認識実験を行う)

```

cd ~/SSS/Work/Recog
noadapt_init.csh    % 結果は Score/init/200

```

- 予備実験(初期モデルの構造で、パラメータをスポンテナスデータで学習し、認識実験を行う)

```

cd ~/SSS/Work/Recog
comp_exp.csh        % 学習から認識まで一気にいきます。
                    % 結果は Score/init/comp

```

- 実験1(バランス単語とバランス文データで HMnet の構造を作成し、スポンテナスデータでパラメータを学習し、認識実験を行う)

```

cd /data/atrq06/xyamamot/Work
train.csh
cd Recog
noadapt.csh

```

- 実験2(基本モデルに対してスポンテナスデータのビターピアライメントを取り、それを詳細ラベルの代わりとして状態分割を行い、新たに生成される HMnet を基本モデルとして、繰り返し HMnet を生成する)

```

cd ~/SSS/Work/flist
~/bin/make_flist.csh % 学習データのファイルリストを作ります。
                    % データを差し替える場合は、train_tasklist を
                    % 書き換えて下さい。

cd ~/SSS/Work
make_sample.csh % パラメータ学習用データを作ります。
                % $DATAHOME/train_para.binary です。
train_spo.csh 3 % こうすれば最初から3回繰り返します。
                % 'train_spo.csh 2 4' とすれば、2回目からスターとして
                % 4回目まで繰り返します。

```

```

cd ~/SSS/Work/Recog
noadapt.csh 1 % 1回目の繰り返しの認識(100~600状態まで)
noadapt.csh 2 % 2回目の繰り返しの認識(100~600状態まで)
noadapt.csh 3 % 3回目の繰り返しの認識(100~600状態まで)

```

- 実験3(HMnetの構造を200状態までは単語データで生成し、それ以後はスポンテナスデータで状態分割を行う。パラメータもスポンテナスデータで学習し、認識実験を行う)

```

cd /data/atrq06/xyamamot/New
train.csh

```

```
cd Recog .  
recog.csh 300
```

この実験は、HMnet の状態分割が 360 状態付近からおかしくなるため、出力される HMnet のうち、400 状態以降のものは使用できない(すなわち、使用できるのは 300 状態だけである)。

また、状態分割中に出力される log ファイル (HMnet\_add.log) に手を加えないと (log ファイルの途中に挟まれる header を取り除く)、filling が出来ないので、シェルで自動的に実行できない (filling 自体ができないことがある)。