TR-I-0076

HMM－LR法を用いた文節認識における継続時間長制御パラメータ変換法の検討

Study of HMM Duration Parameter Adaptation for HMM－LR Speech Recognition System

平田　好充†　　川端　豪　　花沢　利行

Yoshimitsu Hirata, Takeshi Kawabata and Toshiyuki Hanazawa

1989. 3

概要

HMM-LR法による文節認識システムにおいて、音韻モデルの学習時と認識時の発話速度の違いに対応するために、音韻モデルの継続時間長パラメータを補正する方法を検討した。音韻モデルの学習は5240単語中の音韻を用いて行い、学習時と発話速度の異なる連続音声(文節)を認識の対象としてHMM-LR法による認識実験を行った。まず継続時間長パラメータを補正せずに実験を行い、認識誤りの傾向を調べた。次いで、①学習データ(単語)と評価データ(文節)の平均モーラ長の比から単純に線形補正した場合、②平均モーラ長の比から求めた折れ線関数で補正した場合、③音韻モデル毎の継続時間の伸縮率を用いた場合、について認識実験を行い各補正方法の比較･検討を行った。

なお本報告書は、豊橋技術科学大学情報工学課程4年次平田好充が平成元年1月9日～2月28日の期間にATR自動翻訳電話研究所において行った研究を、実務訓練報告書としてまとめたものである。

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

†豊橋技術科学大学

1 はじめに

単語発声データで学習されたHMM(Hidden Markov Model)を用いて、文節発声された音声をHMM－LR法を用いて認識する。この際、単語発声と文節発声では発話速度が異なるため、発声速度をどう補正するかが問題となる。そこで、発声速度の変化に伴う継続時間長パラメータの変換法についての検討を行った。ここでは、単語発声データから得られた継続時間長パラメータを補正せずに文節認識に適用した場合、発話の平均モーラ長の比から単純に線形補正した場合、平均モーラ長の比から求めた折れ線関数で補正した場合、および音韻モデル毎の継続時間の伸縮率を用いて補正した場合について認識実験を行った。その結果、1位の認識率はそれぞれ、85.3%、87.1%、86.4%、86.0%であった。検討の結果、/h/については継続時間長パラメータを[語頭]、[語中]で分けたほうが良いことが分かった。さらに、/kaigini/を含む文節における/g/の脱落については、継続時間長制御とは本質が異なる問題を含んでいるということも明かとなった。

2 HMM－LR法

高精度の連続音声認識を行うには言語情報の利用が不可欠である。プログラミング言語処理系の分野でよく知られているLR構文解析アルゴリズムを、自然言語のような曖昧な文法規則を取り扱えるように拡張した拡張LR構文解析アルゴリズムがTomita[1]により提案されている。LR構文解析アルゴリズムは、表駆動型の構文解析アルゴリズムであり、パーザの動作を記述してある動作表を参照しながら、入力に対するバックトラックなしに決定的に構文解析処理を行えるため、従来の構文解析アルゴリズムより高速な処理が可能である。本認識実験で用いた認識手法は、この拡張LR構文解析アルゴリズムを用いて次音韻を予測し、予測された音韻の存在確率をHMMにより照合し音声データを解析していくHMM－LR法[2]である。

3 HMMによる音声認識

3.1 HMM

HMM(Hidden Markov Model)は、出力シンボルによって一意に状態遷移先が決まらないという意味での非決定有限オートマトンとして定義される。そのため、出力シンボル系列が与えられても状態遷移系列は一意に定まらない。観測されるのは、シンボル系列だけであることからこの名がついている。音声認識に用いられるHMMは初期状態、最終状態が設定され、left-to-rightモデルと呼ばれる。図1にHMMの例を示す。

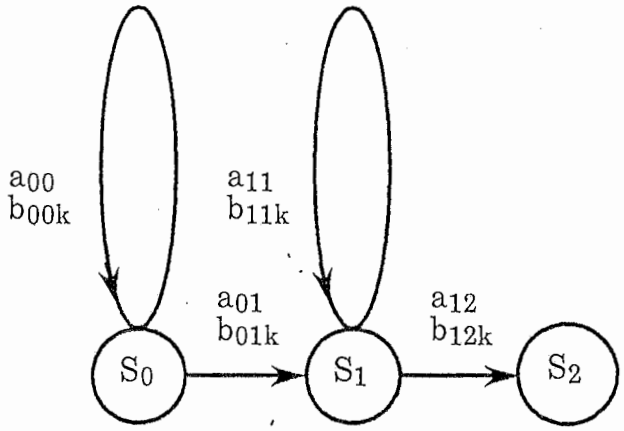


図1 HMMの例

図1に示すモデルのステート数は3であり、遷移を表すアークの数は4である。図中のa_{ij}は状態iから状態jへの遷移確率を表し、b_{ijk}は状態iから状態jへ遷移するとした場合の出力シンボルのラベルがkである確率(出力確率)を表している。音声認識において離散モデルの場合、ベクトル量子化のコードkに対応する。

HMMを用いて音声認識するには、あらかじめ必要なカテゴリー数だけHMMを用意しておき、未知の入力に対して最も大きな生起確率を発生するモデルのカテゴリーを認識結果とする。各カテゴリーのHMMは、Baum-Welch(Forward-Backward)アルゴリズム[3]を用いてそのカテゴリーを発生する生起確率が最大となるように学習しておく。

3.2 HMM音韻モデル

本認識実験では、大語彙の認識を容易にするために認識単位を音韻に設定している。音韻の種類を表1に、モデルの構造を図2に示す。なお、あるステートiにおける自己ループ上での出力確率b_{ii}と、そこからステートj(ただし、j=i+1)に向かうアーク上での出力確率b_{ij}を等しくおいている。

表1 音韻モデルの種類

摩擦音	/s/ /sh/ /h/ /z/
破擦音	/ch/ /ts/
破裂音	/p/ /t/ /k/ /b/ /d/ /g/
鼻音	/ng/ /m/ /n/ /N/
流音	/r/
半母音	/w/ /y/
母音	/a/ /i/ /u/ /e/ /o/
無声化母音	/i'/ /u'/
長母音, 二重母音	/aa/ /ii/ /uu/ /ee/ /oo/ /ei/ /ou/
拗音	/sy/ /hy/ /zy/ /cy/ /py/ /ky/ /by/ /gy/ /ngy/ /my/ /ny/ /ry/

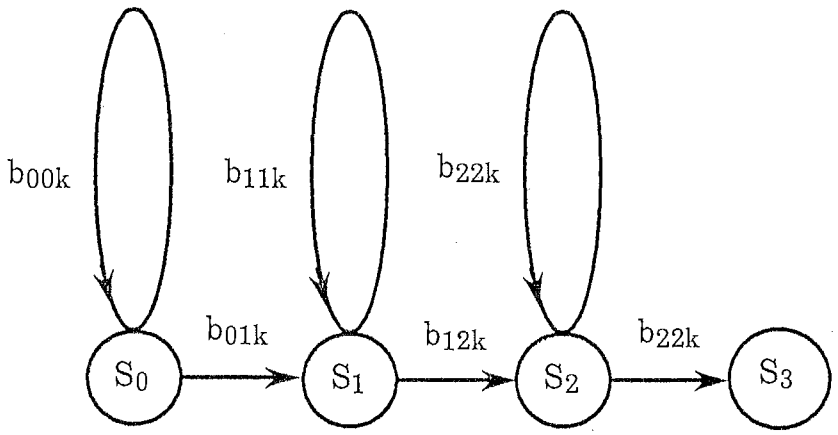
[語頭] [語中]でモデルを別に持つもの

/ch/ /ts/ /p/ /t/ /k/ /b/ /d/ /g/

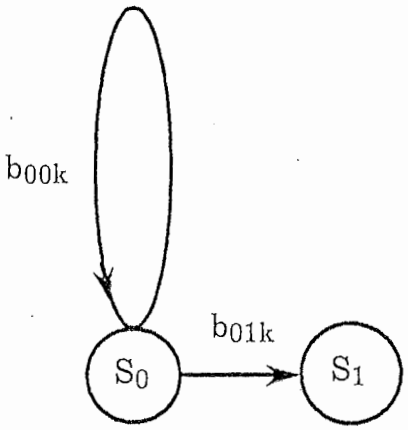
[語頭·語中][語尾]で継続時間長パラメータを別に持つもの

/N/ /a/ /i/ /u/ /e/ /o/ /aa/ /ii/

/uu/ /ee/ /oo/ /ei/ /ou/



(a)3ループモデル(音韻モデル)



(b)1ループモデル(無音、促音区間のモデル)

図2 音韻モデル

3.3 継続時間長制御

基本HMMでは、状態iから状態iにn単位時間とどまる確率d(n)は

$$d(n)=a_{ii}^{(n-1)}\times(1-a_{ii})$$

で表され、nの増加と共に指数関数的に減少する。しかし、音声の定常区間では対応する区間に長くとどまるのが理想的であり、基本HMMはこれをうまく表現できていない。そこで、ここではモデルの学習時に各音韻モデルに対応する学習データの音韻継続時間の最小、最大、平均を求めておく。さらに、Viterbiアルゴリズムによる最適状態遷移経路から、各状態にとどまった継続時間の平均と分散を求めておき、これらを継続時間長制御のパラメータとする。このパラメータを用いて、認識時にモデルの各状態毎の継続時間に応じたペナルティを与えることにより状態毎の継続時間の制御を行う。

4 文節認識実験

4.1 音声データ

1名の話者(NHKアナウンサー、男性)によって発声された日本語重要単語5240単語および音韻バランス216単語をサンプリング周波数12kHzでAD変換する。フレーム周期3ms毎に256点ハミング窓で切り出し、12次LPC分析を行い、①スペクトル(WLR)、②スペクトルの動的特徴(DCEP)、③パワーの各々の特徴量に関してVQコード列に変換する[4]。学習用データには、これらより切り出された音韻を用いた。

学習用データを発声した話者と同一の話者によって発声された25文章279文節を評価用データとした。発声内容は、国際会議の問い合わせを想定したものである。各文節は、1個の自立語の後に複数個の付属語が連鎖したものとなっている。評価用データは、学習用データと同様に分析されVQコード列に変換される。

4.2 文節文法の複雑性

実験に用いた文法の複雑性[5]を表2に示す。

表2 文節文法の複雑性

異なり語彙数	1035
タスクエントロピー	17.0
音韻パープレキシティ	5.9

4.3 実験

単語発声で得られた音韻毎のHMMモデルを使って文節発声された音声を認識する。単語発声と文節発声では発話速度が異なるため、この発声速度の変動を補正するように、継続時間長制御パラメータを変換する方法を検討した。

4.3.1 単語発声のパラメータをそのまま適用(無補正)した場合

単語データから得られた継続時間長制御パラメータを補正せずに文節認識に用いた。認識の結果、1位で認識されたものの文節認識率は85.3%であり、2位から5位までに認識されたものの文節認識率は表3に示すような結果となった。

表3 文節認識結果(無補正)

順位	正解数	累積正解数	認識率
1	238	238	85.3%
2	28	266	95.3%
3	8	274	98.2%
4	2	276	98.9%
5	0	276	98.9%

主な認識誤りのとして、次のようなものがあった。

- 語中の/g/の脱落

文節番号	正解文節	認識結果
No. 5	/kaigini/	/kai/
No. 30	/kaigi/	/kai/
No. 48	/kaigini/	/kaini/
No.101	/kaigino/	/kaino/
No.217	/kaigini/	/kaini/

/kaigi/を含む文節は14個現れる。

- 語尾の/made/において/d/が脱落するもの

文節番号	正解文節	認識結果
No. 21	/zimukyokumade/	/zimukyokumae/
No.164	/kitaoojiekimade/	/kitaoojiekimae/
No.177	/kaigijoumade/	/kaigijoumae/

語尾の/made/を含む文節は3個現れる。

- /i/と/o/の間に/y/が挿入するもの

文節番号	正解文節	認識結果
No. 42	/tetsuzukio/	/tetsuzukiyo/
No. 53	/youshio/	/youshiyo/
No. 80	/yokoushuudaio/	/yokoushuudaiyo/

No.104　/youshio/　/youshiyo/

No.131　/youshio/　/youshiyo/

No.171　/takushiio/　/takushiiyo/

- 語尾の/o/の脱落

文節番号　正解文節　認識結果

No. 66　/gohaQpyouo/　/gohaQpyou/

No. 70　/tourokuryouo/　/tourokuryou/

No. 97　/gozyuusyoo/　/gozyuusyo/

No.135　/kaizyouo/　/kaizyou/

- その他の置換、脱落、挿入誤り

置換 18個

脱落 3個

挿入 2個

4.3.2 線形補正

発話速度が変化した場合、発声速度の比を用いて線形に継続時間長パラメータを変換することが考えられる。学習用単語発声音声データにおける平均モーラ長は175ms、評価用に用いた文節における平均モーラ長は140msであるので、それらの比は0.797である。この線形変換を継続時間長パラメータに適用し、認識実験を行ったときの結果を表4に示す。

表4 文節認識結果(線形補正)

順位	正解数	累積正解数	認識率
1	243	243	87.1%
2	22	265	95.0%
3	6	271	97.1%
4	1	272	97.5%
5	2	274	98.2%

主な誤りとして、次のようなものがあった。

● 語中の/g/の脱落

文節番号	正解文節	認識結果
No. 30	/kaigi/	/kai/
No. 48	/kaigini/	/kaini/
No.101	/kaigino/	/kaino/

● /i/と/o/の間に/y/が挿入するもの

文節番号	正解文節	認識結果
No. 42	/tetsuzukio/	/tetsuzukiyo/
No. 53	/youshio/	/youshiyo/
No. 80	/yokoushuudaio/	/yokoushuudaiyo/
No.104	/youshio/	/youshiyo/
No.131	/youshio/	/youshiyo/
No.171	/takushiio/	/takushiiyo/

● 語尾の/o/の脱落

文節番号	正解文節	認識結果
No. 66	/gohaQpyouo/	/gohaQpyou/
No. 70	/tourokuryouo/	/tourokuryou/
No. 97	/gozyuusyoo/	/gozyuusyo/
No.135	/kaizyouo/	/kaizyou/

● 語頭の/h/と/sh/の置換

文節番号	正解文節	認識結果
No. 63	/hiyouwa/	/shiryouwa/
No. 73	/hiyouwa/	/shiryouda/

● その他の置換、脱落、挿入誤り

置換 14個

脱落 4個

挿入 2個

表4より1位の認識率は87.1%であり、継続時間長パラメータに補正を加えない場合よりも良くなっていることが分かる。補正を加えない場合に比べて、線形に

補正した方が認識結果が向上しているものの主なものには、/kaigi/を含む文節が3個、/made/を語尾とする文節が2個あり、次のように順位が向上している。

文節番号	正解文節	無補正での順位	線形補正での順位
No. 5	/kaigini/	4位	1位
No.101	/kaigino/	3位	2位
No.217	/kaigini/	3位	1位
No. 21	/zimukyokumade/	2位	1位
No.177	/kaigizyoumade/	2位	1位

これらの結果において線形補正の方が認識結果が向上しているのは、単語発声に比べ文節発声の方が/igi/の/i/(単語発声に比べて、文節発声の語中の/i/は平均76%短くなっている)の音や/made/の/e/(単語発声に比べて、文節発声の語尾の/e/は平均83%に伸縮している。)が短く発声されているため、無補正の場合に/g/や/d/が脱落したためと考えられる。逆に、無補正の方が認識結果が良いものの主なものを次に挙げる。

文節番号	正解文節	無補正での順位	線形補正での順位
No. 63	/hiyouwa/	1位	2位
No. 73	/hiyouwa/	2位	4位
No.233	/geNgo/	3位	5位
No.281	/agemasu/	1位	6位

/hiyouwa/で線形補正では/h/を/sh/に誤認識している。これは、語頭の/h/が本来あまり変化しないのに(語頭の/h/は平均0.93の縮み)、継続時間長パラメータを縮めたことによって、より長い音韻/sh/に認識されたと考えられる。また、/geNgo/は/geNkou/と認識され、/agemasu/は/agemasusou/と認識されている。これも、語尾の/o/、/u/の継続時間長パラメータを縮めたことによってより多くの音韻があると認識され(語尾の/o/は平均1.03倍の伸び、語尾の無声化/u/は平均0.98倍の縮み)、挿入誤りとなったと考えられる。

4.3.3 折れ線補正

平均モーラ長の比を用いて継続時間長パラメータを縮めた結果、4.2.2で示したように良くなったものもあれば、/h/、語尾の/o/、/u/等のように本来変化しない

ものまで縮めたため誤りが生じたものもあった。そこで、一般に元々短い音韻は発声速度が変化しても変化せず、長い音韻は発声速度が速くなると短くなる傾向があるので、ある長さ(ブレークポイント)より短い音韻の継続時間長パラメータは変化させず、それより長い音韻の継続時間長パラメータを縮めるという方法として、ここでは折れ線を継続時間長パラメータの補正関数とする。

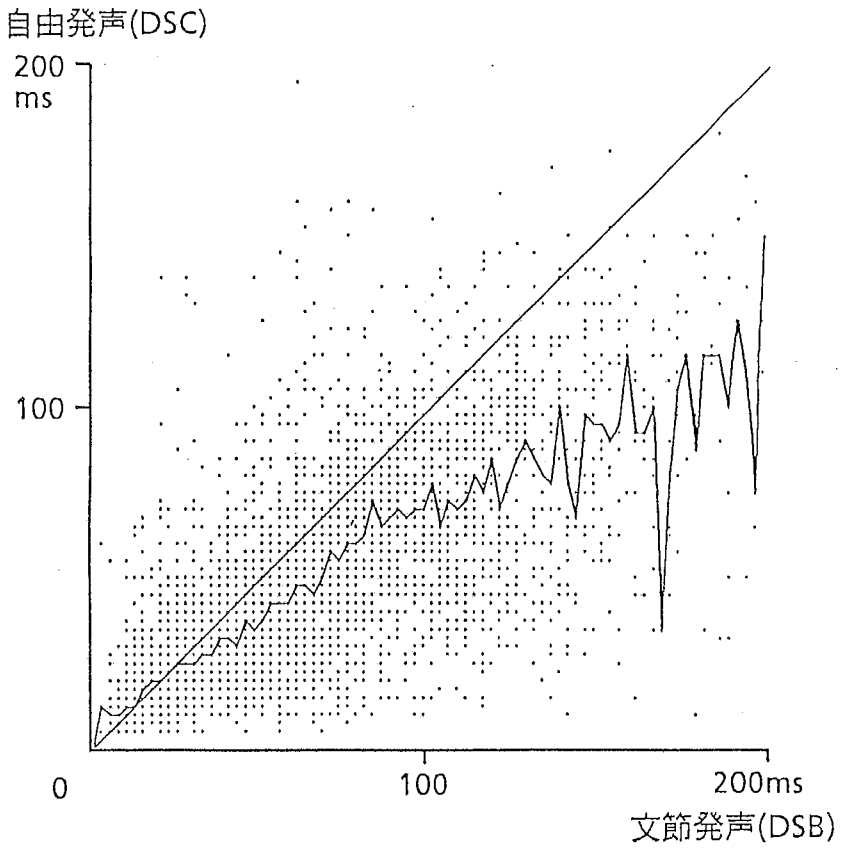


図3 イベント継続時間長の変化

図3は、「短い文節」と「自由発声」の二通りで発声された同じ文章について、対応するイベントの長さがどう変化するかをプロットしたものである。横軸の各点毎に求めたメジアンを折れ線で示す。実験において、学習用音声データである「5240単語」と評価用データである「短い音節」との間のこの様なイベント継続時間長の変化のメジアン関数を折れ線で近似することができれば良いのであるが、単語と文章では発声内容が異なるためイベント間の対応付けができない。そこで、発声内容が同じで発声速度が異なる3種類の音声データ間について継続時間長の変換関数である折れ線を求めておき、発声速度の比から学習用デー

タと評価用データ間の折れ線を推定することにする。3種類の音声データの種類と平均モーラ長を表5に示す。

表5 音声データの種類と平均モーラ長

記号	種類	平均モーラ長
DSA	長い文節	129.58ms
DSB	短い文節	140.04ms
DSC	自由発声	104.64ms

いま、メジアン関数をm(x)、求める折れ線関数をg(x)とする。ブレークポイントB以下の音韻は縮めず、それ以上の長さのものは変化させるので、

$$g(x)=x \qquad (x \leqq B) \tag{1}$$

$$g(x)=ax+b \qquad (x>B) \tag{2}$$

となる。x=Bでは(1)、(2)式の値は等しいとおくことによりb=(1−a)Bとなるから、(2)式は、(3)式のように書き換えられる。

$$g(x)=ax+(1-a)B \quad (x>B) \tag{3}$$

横軸xでのイベントの出現頻度をc(x)として、出現頻度の重み付きのm(x)とg(x)最小自乗誤差Eは、

$$E=\Sigma c(x)(g(x)-m(x))^2 \tag{4}$$

(4)式を最小にするaは、(3)式より

$$\frac{dE}{da}=\sum_{x>B}2c(x)\{ax+(1-a)B-m(x)\}(x-B)=0$$

よって

$$a=\frac{\Sigma c(x)\{(m(x)-B)(x-B)\}}{\Sigma c(x)(x-B)^2} \tag{5}$$

となり、aはBが定まれば一意に決まることが分かる。そこで、ブレークポイントBをある範囲で変化させ、(4)式を最小にするものを見いだせば良い。

図4に、図3に示したメジアン関数を頻度の重み付きで最小自乗近似した折れ線を示す。

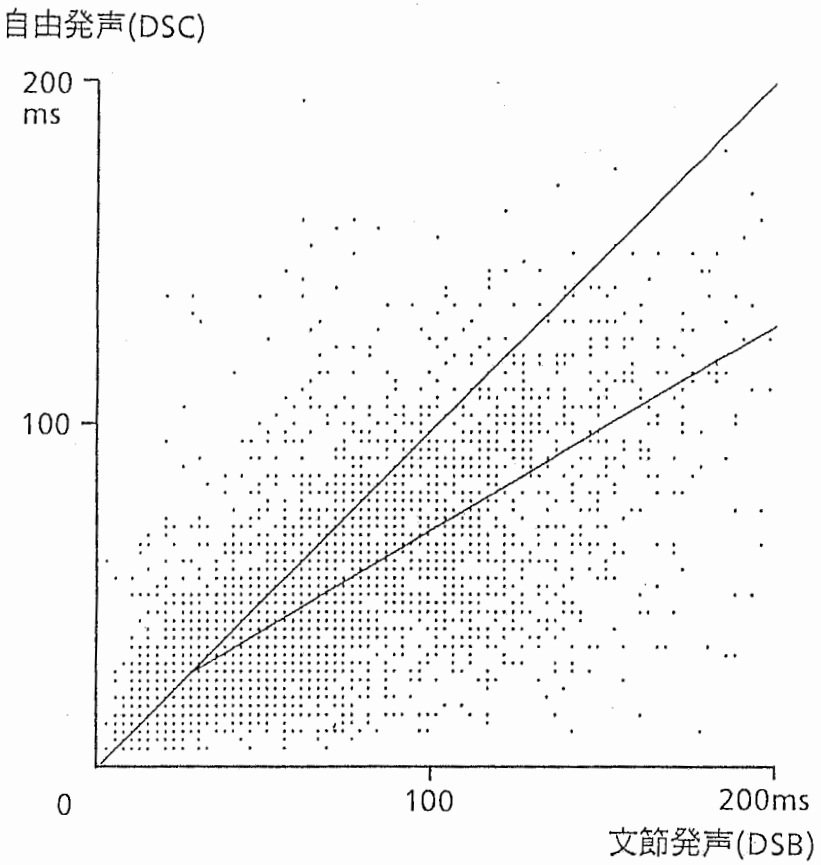


図4 イベント継続時間長の折れ線近似

DSB－DSC間、DSA－DSC間及びDSB－DSA間で、上に述べた折れ線を求めた結果の傾きaとブレークポイントBを、平均モーラ長の比をパラメータとして図5に示す。

学習データの「5240単語」と認識対象の「短い音節」(DSB)の平均モーラ長の比は0.797であるので、図5より必要とされる折れ線のブレークポイントBと傾きaの推定値は次のようになる。

ブレークポイント　30ms

傾き　0.665

この様にして推定された折れ線関数で、単語で求めた継続時間長パラメータを変換した。変換は、音韻モデルの平均継続時間長を折れ線で補正した後、補正前との比を各状態の継続時間に乗じることによって行った。

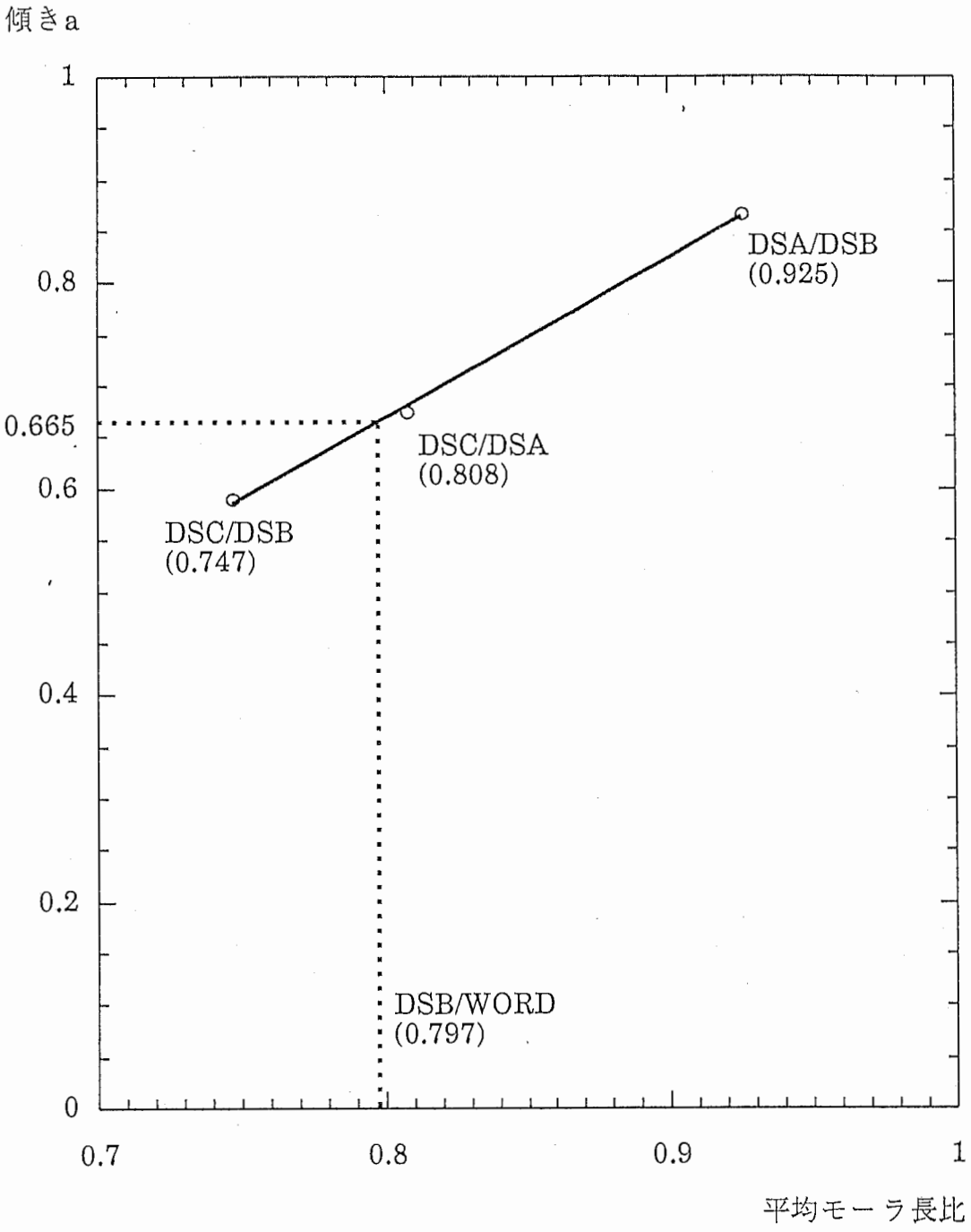


図5 (a) 平均モーラ長比と傾きaの関係

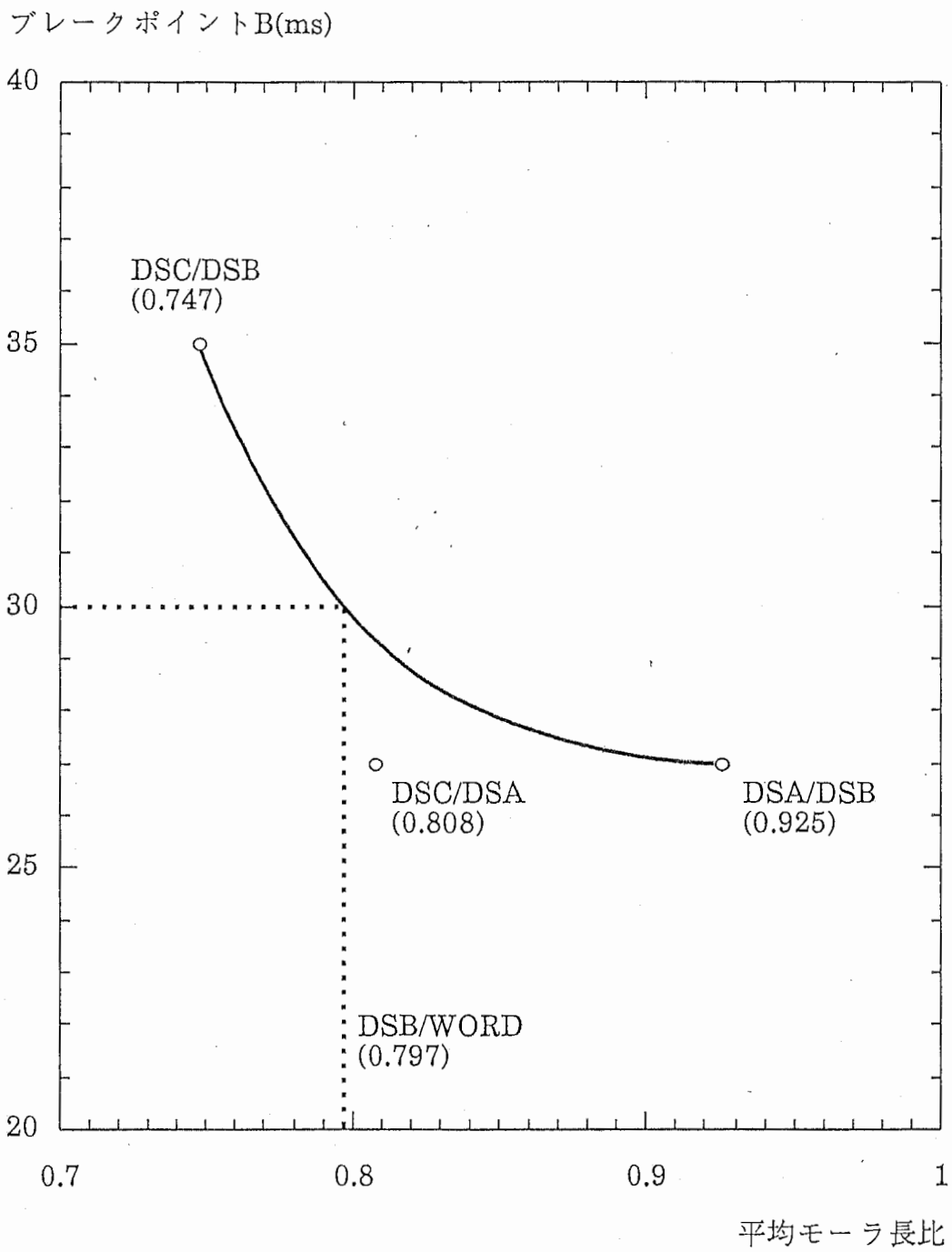


図5 (b) 平均モーラ長比とブレークポイントBの関係

認識結果を表5に示す。

表6 文節認識結果(折れ線補正)

順位	正解数	累積正解数	認識率
1	241	241	86.4%
2	24	265	95.0%
3	6	271	97.1%
4	2	273	97.9%
5	0	273	97.9%

主な認識誤りのとして、次のようなものがあった。

- 語中の/g/の脱落

文節番号	正解文節	認識結果
No. 48	/kaigini/	/kaini/
No.101	/kaigino/	/kaino/

- /i/と/o/の間の/y/の挿入

文節番号	正解文節	認識結果
No. 42	/tetsuzukio/	/tetsuzukiyo/
No. 53	/youshio/	/youshiyo/
No. 80	/yokoushuudaio/	/yokoushuudaiyo/
No.104	/youshio/	/youshiyo/
No.131	/youshio/	/youshiyo/
No.171	/takushiio/	/takushiiyo/

- 語尾の/o/の脱落

文節番号	正解文節	認識結果
No. 66	/gohaQpyouo/	/gohaQpyou/
No. 70	/tourokuryouo/	/tourokuryou/
No. 97	/gozyuusyoo/	/gozyuusyo/
No.135	/kaizyouo/	/kaizyou/

- /h/と/sh/の置換

文節番号	正解文節	認識結果
No. 63	/hiyouwa/	/shiryouwa/
No. 73	/hiyouwa/	/shiryouda/
No. 83	/hiyouwa/	/shiryouwa/

●その他の置換、脱落、挿入誤り

置換 15個

脱落 3個

挿入 3個

予想では、折れ線補正を行うことによって、短い音韻の場合は補正せず、長い音韻については補正を加えるため、無補正の場合と線形補正した場合の両者の長所がでると考えたが、認識結果を見ると1位の正解率は86.4%であり両者の中間の性能を示し、不本意な結果に終わった。誤りの中で/i/と/o/の間の/y/の挿入や語尾の/o/の脱落などは、継続時間長パラメータを無補正の場合及び線形補正の場合と同じ誤りであるが、無補正及び線形補正の両方よりも順位が低下したものとして、次のようなものがあった。

文節番号	正解文節	認識結果
No. 7	/tourokuo/	/tourokuryou/
No. 37	/saNkao/	/saNkou/
No. 82	/hiyouwa/	/shiryouwa/
No.233	/geNgo/	/geNkou/
No.263	/eigoni/	/zeNgoni/

折れ線近似を行うと、ある長さ(本認識実験の場合は76ms)以上の音韻については、最も伸縮率が大きくなる。したがって、これら語頭の/h/、語尾の母音など、本来あまり変化しないものを音韻の長さだけによって縮めた結果、直線補正よりも性能が低下したと考えらる。

4.3.4 音韻モデルごとに補正

4.3.3において、音韻モデルの長さのみに注目し、その継続時間長を折れ線関数で補正した。しかし、音韻によっては、/made/等の語尾の/e/は単語発声に比べて平均0.83倍に短くなっており、逆に/saNkao/等の語尾の/o/は平均1.03倍になって

おりほとんど変化していない。また、/hiyouwa/等の語頭の/h/は単語発声に比べてほとんど変化していない。このように、単純に継続時間長の短い音韻の継続時間長は変化させず、長い音韻の継続時間のみを縮めるという折れ線補正では無理があると考えられる。そこで、用意している音韻モデル毎に対応する音韻の学習用音声データと評価用音声データのそれぞれの継続時間の平均の比をその伸縮率として、継続時間長のパラメータの補正を行った。継続時間長パラメータの変換は、線形補正の場合と同様に、音韻の継続時間長と各状態ごとの継続時間長に伸縮率を乗じて行った。

認識結果を表7に示す。

表7 文節認識結果(音韻モデル毎に補正)

順位	正解数	累積正解数	認識率
1	240	240	86.0%
2	25	265	95.0%
3	5	270	96.8%
4	3	273	97.9%
5	0	273	97.9%

主な認識誤りのとして、次のようなものがあった。

- 語中の/g/の脱落

文節番号	正解文節	認識結果
No. 30	/kaigi/	/kai/
No. 48	/kaigini/	/kaini/
No.101	/kaigino/	/kaino/

- /i/と/o/の間の/y/の挿入

文節番号	正解文節	認識結果
No. 42	/tetsuzukio/	/tetsuzukiyo/
No. 53	/youshio/	/youshiyo/
No. 80	/yokoushuudaio/	/yokoushuudaiyo/
No.104	/youshio/	/youshiyo/

No.131　/youshio/　/youshiyo/

No.171　/takushiio/　/takushiiyo/

- 語尾の/o/の脱落

文節番号	正解文節	認識結果
No. 66	/gohaQpyouo/	/gohaQpyou/
No. 70	/tourokuryouo/	/tourokuryou/
No. 97	/gozyuusyoo/	/gozyuusyo/
No.135	/kaizyouo/	/kaizyou/

- /h/と/sh/の置換

文節番号	正解文節	認識結果
No. 63	/hiyouwa/	/shiryouwa/
No. 73	/hiyouwa/	/shiryouda/

- その他の置換、脱落、挿入誤り

置換 16個

脱落　4個

挿入　4個

結果をみると、全体としては音韻モデル毎に継続時間長を補正した場合と折れ線で補正した場合とでは、ほぼ同等の認識性能ということができる。音韻毎に継続時間を補正することによって、折れ線近似で認識した際に無補正及び線形補正結果より認識順位が低下していた5文節については全て順位の向上がみられた。逆に、折れ線補正より順位が低下したものもみられた。例えば、/kaigi/を含む音節2個(No.30、No.48)の順位が、折れ線補正したものよりそれぞれ1位づつ低下した。

5 考察

音韻毎に補正したことによって、No.7 /tourokuo/は2位から1位に、No.37 /saNkao/は3位から2位に、そしてNo.233 /geNgo/は6位以下から4位にそれぞれ折れ線補正のものに比べて順位が向上した。これらは、語尾の/o/(1.03倍)など本来縮まない音韻を縮めなかったことによって音韻毎の特徴が反映され、挿入誤りが

減少した例ということができる。音韻毎に補正したことによって、無補正では正解であった/hiyouwa/は/shiryouwa/に間違っている。これは、/h/の継続時間長パラメータが語頭と語中で分けていないことから、語頭、語中の全体での倍率(0.87倍)で補正したため、語頭の/h/(0.93倍)に合わなかったものと考えられる。そこで、/h/については[語頭]、[語中]で継続時間長パラメータを別に設けたほうが良いと考えられる。また、/agemasu/を/agemasusou/に間違っているが、これは語尾の/u/の無声化に起因するとも思われる。

次に、音韻毎の継続時間の伸縮率を使って音韻モデル毎に継続時間長パラメータを補正した場合でも、No.30 /kaigi/、No.48 /kaigini/の2文節は/g/の脱落によって、継続時間を折れ線で補正した場合よりも順位が低下した。これらの文節の/i/、/g/の部分が評価音声データにおける/i/、/g/の平均的な継続時間長よりも短く発声されているのであれば、伸縮率のより大きい折れ線補正のほうが、順位がよくなるということは考えられる。しかし、これら2文節について視察により付けられた音節境界(ラベル)から/igi/の部分についてそれぞれの音韻の長さを調べると表8に示すようになっており、平均的な継続時間長よりも短くはなっていないことが分かった。

表8 /igi/の部分の各音韻の継続時間長

文節番号	/i/	/g/	/i/
No.30	88ms (70)	55ms (50)	138ms (143)
No.48	73ms (70)	88ms (50)	48ms (70)

(( )内の値は評価用音声データにおける平均値)

継続時間長制御を行わずに正解データを与えてHMMで認識したところ、いずれの文節の場合も/g/のモデルにおいてはループを全く通過せず、先行あるいは後続する/i/が本来/g/であるべきところまで及んでいることが分かった。実際に両文節について/g/とラベル付けされた部分を[語中]の/g/のモデルと[語頭･語中]の/i/のモデルとで出力確率を計算したもを図6、図7に示す。

スペクトル(WLR)

b_{23}	8888888886666466666	8888888888888688888
b_{22}	8888888886666466666	8888888888888688888
b_{12}	6666666667777777777	8888888887777377777
b_{11}	6666666667777777777	8888888887777377777
b_{01}	7777777776666766666	7777777776666566666
b_{00}	7777777776666766666	7777777776666566666

スペクトルの変化(DCEP)

b_{23}	6666344666677777777	7777666777788888878
b_{22}	6666344666677777777	7777666777788888878
b_{12}	6666444666688888877	7777666666688888888
b_{11}	6666444666688888877	7777666666688888888
b_{01}	7777677888866666666	6666666444444444455
b_{00}	7777677888866666666	6666666444444444455

パワー(POW)

b_{23}	54220000000000000004	8888888888888888888
b_{22}	54220000000000000004	8888888888888888888
b_{12}	77666566511100133367	64443233200011000034
b_{11}	77666566511100133367	64443233200011000034
b_{01}	87666666665555555567	66666666665555556666
b_{00}	87666666665555555567	66666666665555556666

WLR+DCEP+POW

b_{23}	20000000000000000000	66665666666777677766
b_{22}	20000000000000000000	66665666666777677766
b_{12}	32220002000000001144	43221012100000000012
b_{11}	32220002000000001144	43221012100000000012
b_{01}	54444344422212222224	33332222000000000000
b_{00}	54444344422212222224	33332222000000000000

(a) /g/ (b) /i/

図6 文節No.30の/g/の部分を/g/と/i/のモデルで照合した場合の出力確率

スペクトル(WLR)

b_{23}	5855568888888888888866666666	7877776888888888888866667778
b_{22}	5855568888888888888866666666	7877776888888888888866667778
b_{12}	5655555666666666666655556666	8888888888888888888888887778
b_{11}	5655555666666666666655556666	8888888888888888888888887778
b_{01}	6766666777777777777776666666	7777778777777777777788887777
b_{00}	6766666777777777777776666666	7777778777777777777788887777

スペクトルの変化(DCEP)

b_{23}	7667766666666677777777667667	6777777777777788888888888888
b_{22}	7667766666666677777777667667	6777777777777788888888888888
b_{12}	7778877777777777777788668778	8888888888888888888888888888
b_{11}	7778877777777777777788668778	8888888888888888888888888888
b_{01}	6666666666666666666677777666	6666655555555555556666655665664
b_{00}	6666666666666666666677777666	6666655555555555556666655665664

パワー(POW)

b_{23}	66666666566666666666666666666	88888888888888888888888888888
b_{22}	66666666566666666666666666666	88888888888888888888888888888
b_{12}	88888777777788888888888888888	88877666666688888888888888888
b_{11}	88888777777788888888888888888	88877666666688888888888888888
b_{01}	88888888888888888888888888888	87777766667777777777777777777
b_{00}	88888888888888888888888888888	87777766667777777777777777777

WLR+DCEP+POW

b_{23}	24221023233344444444422223332	46566666666666666666666566567
b_{22}	24221023233344444444422223332	46566666666666666666666566567
b_{12}	45333224444444445444444334445	67666556566678888877766666666
b_{11}	45333224444444445444444334445	67666556566678888877766666666
b_{01}	45444444445555555555544444444	44422222222222222323233442332
b_{00}	45444444445555555555544444444	44422222222222222323233442332

(a) /g/ (b) /i/

図7 文節No.48の/g/の部分を/g/と/i/のモデルで照合した場合の出力確率

図中の数字は各アークでの出力確率であり数値の大きなものほど出力確率が大きいことを示し、横軸は時間(フレーム)を表している。この場合、No.30の/kaigini/の/g/の部分はパワーのスコアが低いために/g/よりも/i/で照合したほうが全体として出力確率が大きくなっている。No.48の場合も、パワーについてはNo.30ほどはっきりと差は現れていないが、全体としてはやはり/i/で照合したほうが出力確率は大きくなっていることが分かる。この他の/kaigi/を含む文節のうち、No.5、No.101、No.217についても/g/の部分について照合を行ったが、これらについても/i/のほうが/g/よりも出力確率が大きかった。したがって、出力確率の小さい/g/はできるだけ短く認識したほうが文節単位のスコアとしてはよくなるであろうし、/g/と認識するよりは/i/と認識したほうがスコアは高くなると考えられる。そこで、/g/の領域までを/i/となろうとするのを抑制する効果は/i/の継続時間長パラメータを最も縮める折れ線補正が最も大きいと考えられる。発話が速くなったことによる音韻の縮みということ以外にこれらの理由から、無補正の場合は/g/の脱落誤りが多く、直線補正、折れ線補正の順に/g/の脱落が減少したものと考えられる。このことが継続時間以外についての問題であるために、音韻毎に補正した結果、例え適切なパラメータに変換できたとした場合でも問題は解決しないことになる。/kaigi/を含む文節は14個あるが、/g/の脱落に関する誤りは/kaigi/、/kaigini/、/kaigino/等、/kaigi/に続く音韻数が1個以下の場

合であって、/kaigieno/、/kaigizyoue/、/kaigizyoumade/等のときはいずれの補正法でも誤っていなかった。これを解決するためには/kaigi/に続く付属語や/kaigi/自身が自立語の一部なのかどうか等の言語情報から何らかの対策を考える必要があるのかもしれない。

6 むすび

今回、単語発声された音声データから学習されたモデルを使って文節発声された音声を認識する際の継続時間長制御パラメータの変換法について検討を行い、発声速度の比から求まる折れ線関数での変換法を提案した。文節単位の認識実験からは、補正を行わなかった場合、発声速度の比で単純に線形補正を行った場合、折れ線で補正を行った場合、および音韻モデル毎の継続時間の伸縮率を用いて補正を行った場合で1位の認識率に大差は見られなかった。これは、今回の音声データでは伸縮率が0.8倍とあまり発声速度が速くなっていないためであると考えられ、発声速度が大きく変化した場合にはもう少し明らかな継続時間長制御の補正効果が期待できると思われる。

参考文献

(1) M.Tomita: "An Efficient Parsing for Natural Language - A Fast Algorithm for Practical Systems", Kluwer Academic Publishers (1986)

(2) 北, 川端, 斎藤: 「HMM音韻認識と予測LRパーザを用いた文節認識」, 信学技報SP88-88(昭和63年10月)

(3) S.E.Levinson, L.R.Rabiner, M.M.Sondhi: "An Introduction to the Application of the Theory of Probabilistic Functions of a Markov Process to Automatic Speech Recognition", AT&T Technical Journal Vol.62, No.4, (April 1983)

(4) 花沢, 川端, 鹿野: 「HMM音韻認識におけるセパレートベクトル量子化の検討」, 音講論集(昭和63年10月)

(5) 川端, 鹿野, 北: 「音韻パープレキシティの提案」, 音講論集(平成元年3月)