社団法人 情報処理学会 研究報告
IPSJ SIG Technical Reports

2006－SLP－64 (20)
2006／12／21

特定話者音節HMMの標準偏差補正と無音削除処理による認識率改善

西 宏之† 江藤 貫峰†

† 崇城大学情報学部
熊本市池田4丁目22の1
E-mail: †nishi@cis.sojo-u.ac.jp

あらまし 現在，音声認識においては前後の音素環境を考慮した不特定話者音素HMMを音響モデルとして用いる手法が主流であるが，これらのシステムでは，話者適応化のため長時間のトレーニング用発話が必要であり，音声認識に馴染みのないユーザへの障壁となっている．本報告では，特定話者音節HMMを用いて学習用発話の時間短縮と認識率の確保を試みた結果を述べる．日本語では110個程度の音節で，外来語を含むすべての単語を表現できるので，特定話者音節モデルを採用することで，学習用の発話を数分以内に完了できるという利点がある．その反面，単語発話や連続発話では，音節境界付近の音響パラメータが音節発話の場合とは大きく異なることから，単語数が多くなると認識率の劣化が避けられないという問題点がある．本報告では，音節境界付近の音響パラメータの変形を，音節HMMのパラメータの標準偏差を操作し，さらに無音区間を削除することで吸収し，認識率を改善する手法を提案する．はじめに，音節発話から得られた音節HMMをそのまま適用して認識率を確認し，次に標準偏差を種々の固定値とし，無音区間を削除して学習と認識を行った場合を対象に評価した．その結果，標準偏差を平均値の30％前後に設定すると，単語数20で，1位正解率88％，単語数50では1位正解率76％程度まで改善できることを示した．
キーワード 音声認識，音節HMM，特定話者，話者適応，トレーニング

Standard deviation control of acoustic parameters and pause deletion method for speaker dependent HMM

Hiroyuki NISHI† and Yasutaka ETOH†

† SOJO University Conputers and Infomations Science Department
4-22-1, Ikeda, Kumamoto city
E-mail: †nishi@cis.sojo-u.ac.jp

Abstract Speaker independent phoneme HMM that considers the phoneme environment is a main current in speech recognition systems. However, that system requires long time utterance for training. In this report, using the speaker dependent syllable HMM, the result of trying shortening the time of the utterance for training and improving the recognition accuracy is described. All of japanese words are organized by 110 syllables and utterance for training can be completed in several minutes. On the other hand, feature parameters in the syllable boundary of word utterance or continuous utterance are different from that of syllable utterence. Therefore, it is difficult to secure the hige recogniton accuracy by syrablle HMM. In order to solve the problem, the method that absorbe the transformation of HMM parameters by controlling the standard deviations of parameters is described. Referencing the original accuracy of raw syrablle HMM, the effect of the controlling the standard deviations of parameters is confirmed. The best score is obtained under the condition that the standard deviation is 30 % of the averave value.
Key words speech recognition, syllable HMM, speaker dependent, speaker adjustment, training

1. はじめに

現在，多くの音声認識システムでは，音響モデルとして前後の音素環境を考慮した不特定話者音素HMM(以下，トライフォンモデルと呼ぶ)が用いられている．トライフォンモデルは，HMMの出力確率，遷移確率ともに前後の音素環境を考慮して学習されるため，正解候補に対して高い尤度を出力し，認識率の確保が容易であるという利点がある．しかし，日本語で

用いられる音素の数は23個程度といわれている[1]が，前後の音素環境を考慮すると，モデルの数が膨大となる点が指摘されており，それに対処する手法として音節モデルが検討されている[2][3][4].

従来の研究では，フレームでなくセグメント単位入力の音節HMMに関するもの[2]，直前の音素でモデルを切り分けるもの[3]，モーラでなく厳密な意味の音節単位のモデルを用いるもの[4]などがある．従来の研究はいずれも，不特定話者モデルの構築を指向しており，モデルの混合および大量のデータを用いて学習することを前提としている．これらの方法では，一旦不特定話者モデルを構築した後，実際の利用者に対して，話者適応化のためトレーニング用発話が必要となる．市販の音声認識ソフトウエアでは数十分の話者適応用のトレーニング発話を必要とするものもあり，音声認識に馴染みのないユーザに対して大きな障壁となっている．

本報告では，特定話者音節HMMを用いて学習用発話の時間短縮と認識率の確保を試みた結果を述べる．日本語では110個程度の音節で外来語を含むすべての単語を表現できるので，特定話者音節モデルを採用することで，学習用の発話を数分以内に完了できるという利点がある．その反面，単語発話や連続発話では，音節境界付近の音響パラメータが音節発話の場合とは大きく異なることから，単語数の増加に伴い，認識率の劣化が避けられないという問題点がある．

これを改善するため，音節境界付近の音響パラメータの変形を，音節HMMパラメータの標準偏差を補正するとともに無音区間の削除を行うことで吸収し，認識率を改善する手法を提案する．

始めに音節発話から得られた音節HMMをそのまま適用した場合，認識率が極めて低くなることを確認する．次に標準偏差を一定値に設定することを想定し，どのような値に設定した場合に認識率が最もよくなるかについて実験的に検証した結果を述べる．最後に，単音節モデルでは促音や破裂音の直前に生じる無音区間をモデルに反映できないことから，学習用音声と評価用音声の両方から無音区間を削除することにより，認識率を向上できることを示す．

2. 音節モデルによる単語音声認識

2.1 トライフォンモデル

トライフォンモデルは，図1の上図に示すように，その音素の前に存在する音素と，後続する音素ごとに別個のHMMとしてモデル化する．[ ]は対象音素の前に存在する音素を，< >は後続する音素を表す．同じ音素でも，[ ]あるいは< >で示される音素が異なれば別の音素モデルとして表現する．

したがって，例えば「やま」という単語をトライフォンモデルの縦続接続で表現すると，図1の下図のようになる．この場合，音素モデルを子音，母音，撥音に分類して考えると次のような考え方で数えることができる．

(1) 母音　日本語の5母音に対し，前に来る音素は子音，母音，撥音，休止である．後続する音素もまた，子音，母音，撥音，休止である．したがって，音素環境を考慮した母音の数は24

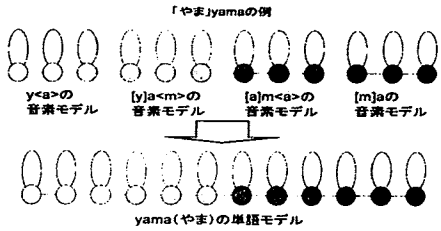

図1　トライフォンモデルの例

×5×24=2880となる

(2) 子音　子音の場合，前に来る音素は母音，撥音，休止であり，後続する音素は母音のみである．したがって，音素環境を考慮した子音の数は，7×17×5=595となる．

(2) 撥音　撥音の場合，前に来る音素は母音のみであり，後続する音素は母音，子音，休止である．したがって，音素環境を考慮した撥音の数は，5×1×23=115となる．

上記より，トライフォンモデルの数は3590個となり，これらをすべて含む学習データを収集するには長時間の発話を必要とすることが推測できる．

2.2 音節モデル

音節モデルは音節と同じ数，すなわち日本語の場合101個，外来語を含めても110個程度である．子音を3状態，母音を3状態として，その連鎖である音節モデルは6状態で表現されることから，図2に示すように，「やま」という単語は12状態から構成される．音節発話による学習は，少ない発話ですべての

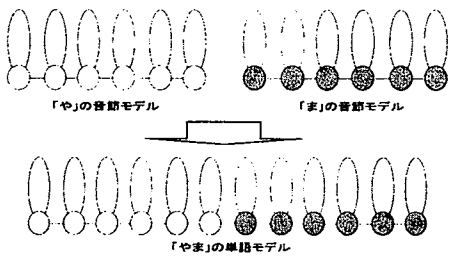

図2　音節モデルの例

特定話者データを収集できるという利点がある一方，そのままではいくつかの重大な問題点がある．次節では，トライフォンモデルと音節モデルの特徴を比較することで音節モデルの問題点を明確にする．

2.3 トライフォンモデルと音節モデルの特徴比較

表1にトライフォンモデルと音節モデルの特徴を比較した結果を示す．以下，表1において音節モデルが不利となる要因について詳細に説明する．

[出力確率の誤差]　音節モデルでは，音節発話された音声を学習データとして用いるので，前後の音節の音響的影響が含まれない．したがって，トレーニング発話の時間をできるだけ短くするため1回〜2回程度の発話音声を用いる場合，データ数が

表 1　トライフォンモデルと音節モデルの比較

比較項目	トライフォンモデル	音節モデル
必要なモデル数	3600 個程度 (×)	110 個程度 (○)
出力確率の誤差	小 (○)	大 (音節境界付近で) (×)
遷移確率の誤差	小 (○)	中 (△)
無音区間での誤差	小 (○)	大 (×)

少ないことから，特徴パラメータの分散値 (標準偏差) が極めて小さな値となるものと考えられる．一方，認識用単語発話音声では，音節境界付近で特徴パラメータは前後の音節の影響で変移し，結果的に，出力確率の大きな減少を招き，認識率を劣化させることが予想される．

[遷移確率の誤差] 学習時は音節発話，認識時は単語以上の単位での発話であることから，音節同士で比較した場合の継続時間が異なり，音節発話時の方が単語発話に比べて長くなることが予想される．この結果，両者で状態遷移確率が異なり，認識率を劣化させる可能性が考えられる．

[無音区間での誤差] ファイルとして与えられた学習用音節発話を切り出す際，データの先頭から探索して無音区間から音声区間への変化点を始端とし，データの最後尾から探索して無音区間から音声区間への変化点を終端点とする方法が一般的である．したがって，音節モデルの縦続接続により単語モデルを構築すると，破裂子音の直前に存在する無音区間の情報を音節モデルに組み込むことができないため，認識対象音声に含まれる，破裂子音前の無音区間時に尤度が下がり，認識率が劣化する．

2.4　音節モデルによる単語認識予備実験

本節では，音節モデルによる単語認識予備実験を行い，前節に述べた種々の誤差が認識率に及ぼす影響を定量的に評価する．実験条件を表 2 に示す．分析パラメータはパワーとケプストラムを用いた．音節モデルの接続で単語モデルを構築することを考慮し，音節境界付近の信頼性が低いことから，デルタケプストラムは使用しなかった．

表 2　予備実験条件

項目	実験条件
分析条件	8kHz サンプリング,16bit, ハミング窓 フレーム長：30ms，フレームシフト：6ms
区間検出しきい値	無音区間パワー値に 10dB を加えた値
分析パラメータ	パワー ケプストラム 10 次
認識対象単語セット	100 都市名
モデル用音声データ	電子協共通音声データベース 男性 1 名分の 101 音節データ
認識用単語音声データ	電子協共通音声データベース 男性 1 名分の 100 都市名 音声データ

特定話者を想定しているので，モデル用音節データの発話者と，評価用 100 都市単語音声データの発話者とは同一とした．

予備実験の結果を表 3 に示す．表 3 に示される認識予備実験

表 3　予備実験結果

	1 位正解率	3 位以内正解率	10 位以内正解率
結果	12 %	24 %	41 %

結果は極めて劣悪なものであり，実使用に耐えるものではない．この原因を表 1 および 2.3 の考察結果に立ち返り，実際のデータをもとに検証することで解決の方策を探ることとする．

2.5　単語認識予備実験結果の分析

2.3 では，音節モデルが不利となる要因として①出力確率の誤差，②遷移確率の誤差，③無音区間での誤差，について説明した．ここでは，これら 3 つの要因を定量的に考察する．

2.5.1　出力確率の誤差

2.3 で述べたように，音節モデル用発話では，特徴パラメータの分散値 (標準偏差) が極めて小さな値となるものと予想され，一方，認識用単語発話音声では，音節境界付近で特徴パラメータは前後の音節の影響で変移し，結果的に，出力確率の大きな減少を招くものと考えられるので，ここでは実際のデータでその予想を検証する．表 4 に 5 母音について，学習用音節発話と認識用単語発話を対象に，パワー値，ケプストラムの 1 次係数，2 次係数，3 次係数の平均値と標準偏差を示す．音節別，

表 4　5 母音の平均値と標準偏差

音節名	位置	項目	パワー (dB)	Cep 1 次	Cep 2 次	Cep 3 次
あ	単音節	平均値	90.9697	1.0543	0.1450	-0.3214
		標準偏差	1.6012	0.1728	0.0574	0.0334
	単語中 (あさひ)	平均値	77.6255	1.4335	0.2710	-0.0549
		標準偏差	8.0074	0.4452	0.2149	0.2400
い	単音節	平均値	80.1275	-0.3709	1.8337	0.8175
		標準偏差	3.7226	0.1547	0.1372	0.0388
	単語中 (せんだい)	平均値	63.9013	0.9522	0.7630	0.9041
		標準偏差	4.7079	0.1378	0.3025	0.3257
う	単音節	平均値	87.9507	0.7807	0.9290	0.2508
		標準偏差	0.5999	0.0535	0.0848	0.0590
	単語中 (はにゅう)	平均値	70.8695	1.0983	0.6319	0.5486
		標準偏差	5.3336	0.2823	0.2147	0.1917
え	単音節	平均値	84.3150	0.2623	0.5928	0.8190
		標準偏差	3.4698	0.0971	0.1129	0.0644
	単語中 (えにわ)	平均値	78.3085	0.2303	0.7918	0.9674
		標準偏差	5.2371	0.2191	0.2295	0.1491
お	単音節	平均値	85.5551	1.1966	1.4728	0.6036
		標準偏差	2.2498	0.1373	0.0324	0.0476
	単語中 (うおず)	平均値	81.4898	1.4823	1.2218	0.5154
		標準偏差	5.5480	0.3484	0.0833	0.2189

パラメータ毎に標準偏差の値を，単音節と単語中とで比較すると，大半は単語中の標準偏差の方が大きいことがわかる．中には数倍の差があるパラメータも見うけられ，単語発話ではパラメータの変動が大きく，誤認識の要因の一つとなっていることが類推される．

これらをさらに明確にするため，5 母音ごとに，単音節と単語中別に，パラメータごとに平均値と標準偏差の比を計算し，その比をパワーとケプストラム (10 次) の全データで平均した

表 5　全パラメータの平均値と標準偏差の比

母音名	分類	「平均値と標準偏差の比」の全平均
「あ」	単音節	0.1641
	単語中	2.9986
「い」	単音節	1.1892
	単語中	2.2653
「う」	単音節	0.1594
	単語中	0.9641
「え」	単音節	0.7361
	単語中	1.0477
「お」	単音節	0.1510
	単語中	1.8601

結果を表 5 に示す．表 5 より，全母音において，単音節より単語中の母音の標準偏差が大きいことがわかる．したがって，これらの状況から，表 3 の理由の一つとして，単音節発話で学習した音節モデルの縦続接続で単語 HMM を構築した場合に，実際の単語発話とのパラメータの誤差が大きく，出力確率が異常に小さな値となることが上げられる．

2.5.2　遷移確率の誤差

2.3 で述べたように，学習時は音節発話，認識時は単語発話以上の単位での発話であることから音節同士で比較した場合の継続時間が異なり，状態遷移確率の違いとなる．この差が尤度の低下をもたらすことが懸念される．そこで，実際のデータでその予想を検証する．5 母音ごとに，単音節と単語中別に，継続時間長を比較した結果を表 6 に示す．表 6 からは，「あ」，「い」，

表 6　継続時間長の比較

母音名	分類	継続時間長 (ms)
「あ」	単音節	138
	単語中	114
「い」	単音節	180
	単語中	114
「う」	単音節	108
	単語中	216
「え」	単音節	162
	単語中	132
「お」	単音節	150
	単語中	210

「え」では単音節の方が継続時間が長く，「う」，「お」では単語中の方が長い．いずれにしても，表 5 ほど単音節と単語中の差は明らかでない．したがって，継続時間長の誤差が表 3 の主要な原因であると断定することは難しい．

2.5.3　無音区間での誤差

2.3 に述べたように，音節モデルの縦続接続により単語モデルを構築すると，破裂子音の直前に存在する無音区間の情報を音節モデルに組み込むことができないため，認識対象音声に含まれる，破裂子音前の無音区間時に尤度が下がり，認識率が劣化すると考えられる．ここでは，この推測を提示するだけにとどめ，具体的な方策と定量的な検証は後に述べる．

2.6　認識率改善のための方策

上記分析結果に対応し，認識率を改善するための方策を考察する．

2.6.1　出力確率の誤差に対する方策

本質的な改善策は学習時と認識時の音響パラメータの特性を合わせることであるが，そのためには前後の音素環境を考慮して音節 HMM を構築する必要があり，2.1 に述べたようにトライフォンモデル同様，極めて多数の音節モデルを必要とする．そこで，本報告では，次のような考え方に基づく方策を検討する．

図 3 に示すように，「やま」という単語の単語モデルは音節「や」，「ま」の各々の音節モデルを縦続接続して得られる．した

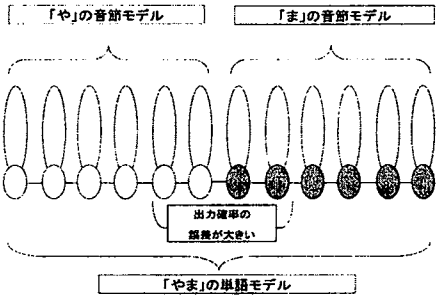

図 3　出力確率の誤差が大きくなる状態の例

がって音節「や」の最終の 1〜2 状態は次に続く音節「ま」の影響を受けて変形し，同時に音節「ま」の最初の 1〜2 状態は直前の音節「や」の影響を受けて変形する．そこで，本報告では，各音節の中心部分のパラメータは単音節，単語中とで大きな相違は生じないと考え，音節の境界付近の音響パラメータの変形を，音響パラメータの標準偏差を大きく設定することで吸収し，出力確率の大幅な劣化を避けることができるという仮説を立てる．これは表 5 に示した単音節と単語中との標準偏差の違いからも妥当な仮説であると考えられる．

後に述べる評価実験では，標準偏差の拡大の程度を変えて認識性能を評価し，最も認識性能が良くなる標準偏差を求めることとする．ただし，表 4 からわかるように音節やパラメータによって標準偏差の値には大きな差があり，標準偏差値を一律に一定値倍することはできない．そこで，本研究では，得られた標準偏差は用いずに，平均値に対して 1 より小さい一定の倍率を乗じて標準偏差を求めることとする．

なお，2.5.2 に述べたように遷移確率は，単音節と単語中とで大きな差がないことから本研究では特段の対処策は検討しない．

2.6.2　無音区間における誤差に対する方策

無音区間における誤差の解決のためには，以下の方法が考えられる．

[無音モデル挿入法]　破裂音など音節モデルの直前に無音が生じる子音が現れたら，その前に強制的に無音モデルを挿入する．

[無音区間削除法]　認識対象発話から無音区間を予め削除しておき，有音区間のみを対象に認識処理を行う．

[始終端検出法]　特段の処理を行わず，始端終端を検出して従来どおりの認識処理を行う．

上記のうち，無音モデル挿入法は，発話者による単語中の無音区間長のばらつきを吸収する無音区間モデルの設定が難しく，無音区間中の出力確率の設定にも課題が残る．無音区間削除法 [5] は一見無音部分の情報が欠落するため音声認識には不向きと思われるが，無音区間が音声認識処理に必要とされるのは，始終端の検出と促音の認識であり，それ以外の場面で無音区間が認識のための重要な要素となることは考えられない．むしろ無音区間が無いことで，無音区間における出力確率を設定する必要がなく，全体の処理量も削減されることから有効であると期待できる．以上の状況から，本研究では，始終端検出法をリファレンスとして，無音区間削除法の適用の効果を認識実験により検証する．

3. 単語音声認識実験

3.1 実験条件

前節に述べた方策の有効性を検証するための単語音声認識実験を行う．実験条件を表 7 に示す．上記実験条件のうち，「分析条件」～「認識対象単語セット」の条件は予備実験条件 (表 2) と同じである．予備実験では 1 名の男性話者で実施したが，単語音声認識実験は 5 名の話者を用いた．ただし，特定話者音声認識であるので，5 名の話者ごとに，個別に音節モデルの構築と特定話者音声認識実験を行い，個別の認識率の平均値により総合的な評価を行った．

表 7　単語音声認識実験条件

項目	実験条件
分析条件	8kHz サンプリング,16bit フレーム長：30ms フレームシフト：6ms 窓関数：ハミング窓
区間検出しきい値	無音区間パワー値に 10dB を加えた値
分析パラメータ	パワー ケプストラム 10 次
認識対象単語セット	100 都市名
モデル用音声データ	電子協共通音声データベース 男性 5 名分 (D01～D05) の 101 音節
認識用単語音声データ	電子協共通音声データベース 男性 5 名分 (D01～D05) の 100 都市名 全 100 都市名，50 都市名，20 都市名
認識性能評価方法	・1 位正解率 (認識率) ・3 位累積正解率 ・10 位累積正解率 上記 3 評価を 5 名の話者別 および 5 名分の平均で評価
認識性能改善方策 (1) [実験 1] ＜標準偏差の再設定＞	音響パラメータの 標準偏差を下記に設定 パラメータ平均値の 10 %,20 %,30 %,40 %,50 %
認識性能改善方策 (2) [実験 2] ＜無音区間の削除＞	上記方策 (1) の全条件下で 下記方法 A と方法 B を比較評価 方法 A：始終端を検出して単語認識 方法 B：全無音区間を削除して単語認識

参考のため，100 都市名すべてを認識対象単語セットとした場合の他，50 都市名，20 都市名を認識対象とした場合についても認識実験を行った．

認識性能評価方法としては，1 位候補が正解である 1 位正解率 (認識率)，3 位までの候補中に正解が含まれる 3 位累積正解率，10 位までの候補に正解が含まれる 10 位累積正解率を用いた．これらの評価方法を 5 名の話者別に行い，総合評価として 5 名分の平均でも評価した．

第一の認識性能改善方策 (実験 1) として，出力確率の補正のため，音響パラメータの標準偏差値を，音響パラメータ平均値の 10 %,20 %,30 %,40 %,50 %および 60 %の値に設定して認識実験を行った．設定の範囲は，表 5 より，大半の単語中の音節の標準偏差が，単音節中の標準偏差の 2 倍～10 倍の範囲の値を持つことから選定した．

第二の認識性能改善方策 (実験 2) は無音区間の削除を行った後に認識処理を行うことである．具体的には，第一の方策の全条件下で始終端を検出して単語認識を行う方法 A と，無音区間をすべて削除して単語認識を行う方法 B を実施し，その性能を比較評価する．

3.2 実験結果

3.2.1 実験 1～標準偏差の再設定

表 7 の認識性能改善方策 (1) として，音響パラメータの標準偏差値を変えて認識実験を行った結果を図 4～図 6 に示す．図 4 は 1 位候補の正解率，図 5 は 3 位までの累積正解率，図 6 は 10 位までの累積正解率である．5 名の話者 (D01～D05) 別の認識結果および全話者の平均値を示した．図 4 および図 5 より，標準偏差の値としてはパラメータの平均値の 30 %程度の時に最も認識性能がよいことが読み取れる．図 6 では 40 %が最も平均の性能がよいが，10 位以内の性能では実使用上の性能として不足があり，1 位および 3 位以内の結果を用いることが妥当である．また，話者 D04 が他の話者に比べて突出して認識性能がよいが，単音節発話の音響パラメータの特性と，単語中のそれとが偶然近い話者であったということであり，一般性を認めることはできない．評価結果としては平均値で議論することとする．実験 1 により，標準偏差の値としてはパラメータの平均値の 30 %程度の値に設定すればよいことがわかったが，性能的には平均値で 1 位正解率が 50 %に満たず，未対処の方法 (表 3) に比べると改善は著しいが，実使用上，充分な性能とはいいがたい．

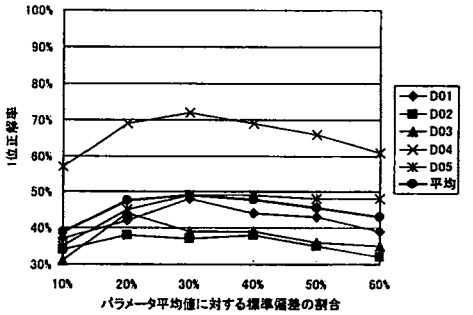

図 4　標準偏差を再設定した場合の認識結果 (1 位正解率)

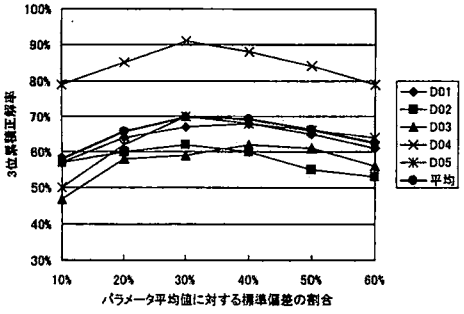

図 5　標準偏差を再設定した場合の認識結果 (3 位累積正解率)

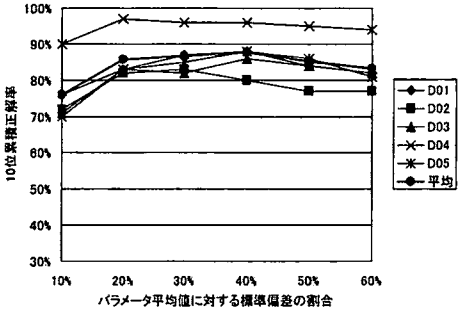

図 6　標準偏差を再設定した場合の認識結果 (10 位累積正解率)

3.2.2　実験 2〜無音区間の削除

表 7 の認識性能改善方策 (2) として，無音区間の削除を行った後に認識処理を行う方法の実験結果を図 7〜図 9 に示す．この実験では，前節の結果をふまえ，標準偏差は平均値の 30 %に固定して実施した．いずれも前節同様始端と終端を検出する方法を方法 A とし，無音区間をすべて削除した後に認識処理を行う方法 B とを比較して示した．図 7 は 100 都市名すべてを認識対象とした場合，図 8 は 50 単語に限定した場合，図 9 は 20 単語の場合の結果である．いずれも，方法 B は方法 A にくらべて，明らかに認識性能が優れていることが確認できる．

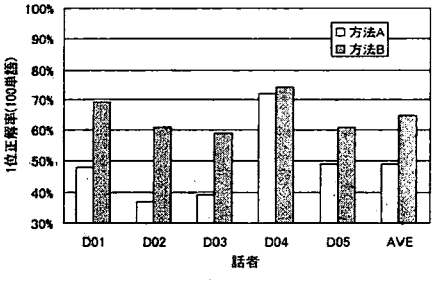

図 7　無音区間削除法 (方法 B) の認識性能 (100 単語)

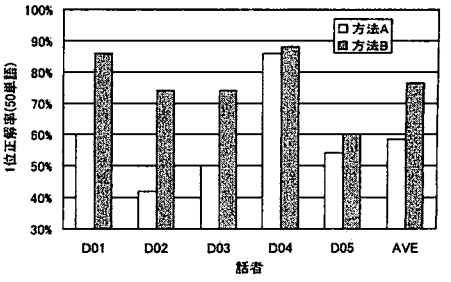

図 8　無音区間削除法 (方法 B) の認識性能 (50 単語)

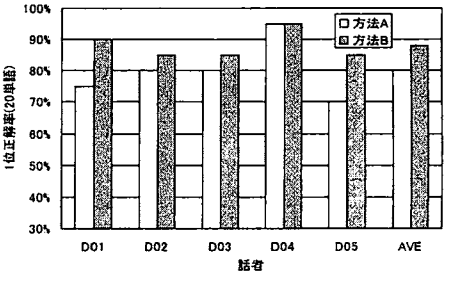

図 9　無音区間削除法 (方法 B) の認識性能 (20 単語)

4.　まとめと今後の課題

3.2 より，音節 HMM の標準偏差としては，平均値の 30 %程度とし，学習および認識処理に先立って，無音を削除することにより，特定話者音節 HMM を用いた単語音声認識率を大幅に改善できる見通しを得た．しかし，認識率が，100 単語で 70 %，50 単語で 80 %に満たないことから，実使用上の認識性能としては不満が残る．本報告では，音節境界付近の出力確率の低下を標準偏差を拡大することで対処したが，本質的な解決方法として，平均値自体の誤差を小さくする方法の検討が重要である．今後はそのような観点から音節 HMM を改良する検討を行う．

文　献

[1] 例えば，板橋秀一編著:"音声工学"，森北出版，pp.7，2005 年

[2] 中川聖一，花井建豪，山本一公，峯松信明:"HMM に基づく音声認識のための音節モデルと triphone モデルの比較"，信学論，D-II，Vol. J83-D-II，No.6 pp.1412-1421，2000 年 6 月

[3] 山本一公，池田太郎，松本弘:"コンパクトで高精度な音節モデルの検討"，音学講論，1-9-22，2002 年秋季研究発表会，pp.43，2002 年 9 月

[4] 緒方淳，有木康雄:"日本語話し言葉音声認識のための音節に基づく音響モデリング"，信学論，D-II，Vol. J86-D-II，No.11 pp.1523-1530，2003 年 11 月

[5] 西宏之，江藤貴峰:"音節 HMM 特定話者音声認識における音声区間検出方法の影響"，平成 18 年度電気関連学会九州支部連合大会，06-1P-03，pp.151，2006 年 9 月