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

2003－ＳＬＰ－49 (26)
2003／12／18

音節連鎖モデルによる大語彙連続音声認識

池田 太郎† 山本 一公† 松本 弘† 西谷 正信†† 宮澤 康永††

† 信州大学工学部 〒380-8553 長野市若里 4-17-1
†† セイコーエプソン株式会社 〒392-8502 諏訪市大和 3-3-5
E-mail: †{rps13,kyama,matsu}@sp.shinshu-u.ac.jp,
††{Nishitani.Masanobu,Yasunaga.Miyazawa}@exc.epson.co.jp

あらまし 本稿ではモーラモデルをベースとし，音節間の調音結合も表現する長いサブワード単位として音節連鎖モデルの検討を行なっている．全ての音節連鎖をモデル化するとモデル数が膨大となり推定精度の劣化を招くため，連続音節認識において誤った2音節連鎖から高頻度のものをモデルとして追加する．更に音節連鎖モデルの追加による学習データ不足に対応するためPTMと同様の分布共有を行なった．調音結合の影響をより強く受けると考えられる講演音声について提案法を検討した結果，誤り頻度100回以上の音節連鎖は，約300種で全誤りの約60%を占めることが分かった．この音節連鎖を音節モデルに追加することにより，単語正解精度は62.2%から64.4%に向上し，トライフォンモデルと比べ約1/3のパラメータ数で0.5%上回る認識性能が得られた．

キーワード 音響モデル，モーラモデル，音節連鎖モデル，PTM，講演音声

Large vocabulary continuous speech recognition by disyllable model

Taro IKEDA†, Kazumasa YAMAMOTO†, Hiroshi MATSUMOTO†, Masanobu NISHITANI††, and Yasunaga MIYAZAWA††

† Faculty of Engineering, Shinshu University 4-17-1 Wakasato, Nagano, 380-8553 Japan
†† SEIKO EPSON CORPORATION 3-3-5 Oowa, Suwa, 392-8502 Japan
E-mail: †{rps13,kyama,matsu}@sp.shinshu-u.ac.jp,
††{Nishitani.Masanobu,Yasunaga.Miyazawa}@exc.epson.co.jp

Abstract This paper proposes disyllable models to take into account of context dependency over longer phone sequence. Our sub-word models consist of both baseline monosyllable models and additional disyllable models. In our approach, we selected a subset of disyllables which causes a large part of recognition errors in continuous syllable recognition. Furthermore, to cope with limited database, we used phonetic tied-mixture (PTM) and modified minimum description length (MDL) criterion. The number of the disyllables with more than 100 errors was about 300, and they occupied about 60% of all errors. The proposed method outperformed a conventional triphone model in LVCSR on CSJ datebase.

Key words acoustic model, mora model, disyllable, PTM, CSJ

1. まえがき

近年の音声認識技術の発展により，静かな環境で丁寧に発声された読み上げ音声に対しては高い認識性能を得ることができるようになってきた．現在の音声認識システムでは，その音響モデルとして隠れマルコフモデル（HMM），音響モデルの単位として音素，特に前後の音素環境（コンテキスト）を考慮したトライフォンモデルを用いることが一般的となっている．トライフォンモデルではその膨大なモデル数に対し，モデルパラメータの推定精度確保と認識時の計算コストを削減するため，音素決定木に基づく状態共有が一般的に用いられ，高い認識性能を得ている．しかし，トライフォンモデルでは，状態共有により，平均的に調音結合への表現力は向上するが，音節などの長いサブワードモデルに比べ，精度の低い部分も生じる．また，認識時には単語境界処理のためにデコーダが複雑になるという問題点もある．

一方，日本語においては比較的音節数が少ないので，音節やモーラを音響モデルの単位として用いることも可能である．ト

ライフォンに比べモデル数が圧倒的に少ないので，各モデルに対する学習データが豊富にあり，各モデル毎に独立にガウス分布を持たせることが可能である．従って，音節間の調音結合は表現することができないが，音節内の調音結合の表現力はトライフォンモデルよりも優れていると考えられる．実際，モーラモデルとトライフォンモデルを比較した場合，ほぼ同等な認識精度が得られることが実験的に示されている [1]～[3]．基本的にコンテキスト独立モデルであるので，デコーダを単純化できるのも音節モデルの特長である．

最近の音声認識研究は，その研究の場を読み上げ音声から，人間同士のコミュニケーション等に用いられる話し言葉音声へと移しつつある．読み上げ音声で学習したモデルにより話し言葉音声を認識した場合，発話様式が読み上げ音声と異なるために，その変化に音響モデルが対応しきれず認識性能が著しく低下する．より自然な話し言葉音声で使用できるシステムを実用化するには，話し言葉特有の現象である発話速度の局所的変化や，言い直し，なまけ，非文法的な表現などを考慮しなければならない [4]．特に発話のなまけによって調音結合の影響がより長いコンテキストに及ぶことが予想され，その影響をモデル化できる長いサブワード単位の音響モデルを導入することで，認識率の改善を図ることができると期待される．話し言葉認識においては，トライフォンモデルよりも音節モデルの方が認識結果が優れているという報告もある [5]．しかし，より長いモデル単位として，音節連鎖や単語を取りあげると，モデル数が膨大となり，実際に学習することが不可能である．そこで，調音結合の影響を強く受ける部分に関してのみ，長いサブワードモデルを作成することを考える．

本稿では，長いサブワードモデルとしてモーラモデルを結合した音節連鎖モデルのうち，誤り頻度の高い音節連鎖のみをモデルセットに加える方法を検討する．ベースのモデルには調音結合の影響をより正確に表現できる事を考慮して，モーラモデルを用いる．誤り頻度の高い音節連鎖を選択するため，連続音節認識実験により誤って認識された音節を含む 2 音節連鎖の頻度を集計し，高頻度の連鎖をモデルとする．また，新たに音節連鎖モデルを追加することにより生じる学習データ不足の問題に対処するため，PTM [6] と同様の方法で分布を共有すると共に，MDL 基準による混合数の最適化を行なう．以上のモデルは話し言葉音声（講演音声）認識により評価を行なう．

2. 音節連鎖モデル

2.1 音節連鎖モデルの利点と問題点

音節連鎖モデルの構築はモーラモデルをベースに行なう．モーラモデル同士を結合させることでより長いコンテキストを表現するモデルを作成し，従来のモーラモデルのセットに追加する方法を用いる．

音節連鎖モデルには次のような利点・問題点がある．

- 調音結合：音節内だけではなく，音節間の調音結合も表現できる
- 長いサブワード単位：モデル単位が長く，音素モデル 2～4 つ分を一つのモデルとして表現するので，より長い調音結合の影響を考慮できる．また，挿入誤りの削減が見込める．
- コンテキスト独立：モーラモデルをベースとしているので，モデル自体はコンテキスト独立であり，認識器を単純化できる．
- 学習データ不足：新たにモデルを追加するため，モデルに対する学習データ量が相対的に減少する．

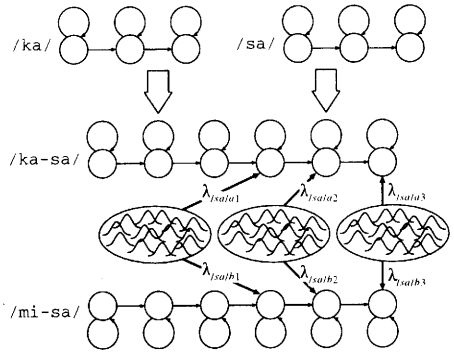


図 1 音節連鎖モデルでの分布共有

2.2 音節連鎖モデルの選択

音節連鎖モデルを追加する際，追加するモデル数や，追加による学習データ不足の問題を考慮しなければならない．追加するモデルが少ないと音節連鎖モデルの効果が発揮できず，逆に追加するモデルが多いと学習データ上の出現頻度の低いモデルについては十分に学習できない．さらに単音節モデルの学習データも減少するので，結果としてモデルパラメータの精度が低下すると考えられる．そこで，学習データに対して連続音節認識実験を行ない，誤った音節を含むコンテキストを音節連鎖として集計し，誤り頻度の高い音節連鎖を音節連鎖モデルとして選択する．

2.3 音節連鎖モデルの構造

音節連鎖の出現頻度は音節に比べ少なく，またモデルセットに新たに音節連鎖モデルを追加することによって学習データ不足の問題が生じる．そこで，モデルには PTM と同様の方法で，同じ音節の対応する状態間でガウス分布を共有し，混合重み λ は独立に学習する方法を用いる．一例を図 1 に示す．この図では，モーラモデル/ka/と/sa/から音節連鎖モデル/ka-sa/が構成される．同様に音節連鎖モデル/mi-sa/も構築されるが，このとき/ka-sa/の後ろ 3 状態と/mi-sa/の後ろ 3 状態は，元々同じモーラモデル/sa/であるので，対応する状態毎に分布を共有する (混合重み λ は独立)．

2.4 音節連鎖モデルの適用

学習用ラベルおよび認識用辞書に音節連鎖モデルを適用する際，複数の音節連鎖の候補が存在する場合には，誤りの多い音節連鎖を優先的に適用する．例えば，/to-o/と/o-i/が音節連鎖モデルとして追加され，且つ/to-o/の方が/o-i/よりも誤りが多い場合を考える．/to o i/というコンテキストに対しては，/to-o i/と/to o-i/の 2 通りの適用パターンが考えられるが，誤りが多いモデルを優先的に適用することから，/to-o i/を採用する．

表 1 分析条件

サンプリング周波数	16kHz
プリエンファシス	$1-0.97z^{-1}$
ハミング窓長	25ms
分析周期	10ms
特徴パラメータ	12 MFCC + 12 Δ MFCC + Δ pow

3. 音節連鎖誤り頻度の調査

モデル化する音節連鎖を選択するため，学習データに対し連続音節認識実験を行なう．その結果より，誤って認識された音節を含むコンテキストを誤り音節連鎖としてその頻度を集計する．

3.1 集計方法

誤り音節連鎖の集計方法を次に示す．

（1） 誤った音節とその先行音節：誤った音節と，先行音節との組合せを誤り音節連鎖とする方法（図 2(a)）．

（2） 誤った音節とその後続音節：誤った音節と，後続音節との組合せを誤り音節連鎖とする方法（図 2(b)）．

（3） 誤った音素により先行・後続音節を判断：この方法（図 2(c)）では誤った音節（CV）を音素 単位で考え，子音（C）が誤っていた場合は先行音節，母音（V）が誤っていた場合は後続音節と組み合わせて誤り音節連鎖とする．CV 両方が誤っている場合は，先行及び後続と組み合わせた両方の音節連鎖を誤りとする．単母音音節，促音 'q'，撥音 'N' が誤った場合は，音声学的特徴を考慮して，単母音音節と促音の場合先行音節，撥音の場合は後続音節と組み合わせて音節連鎖とする．この手法を用いることで，より音節間の調音結合による誤りを考慮できると考えられる．

なお，以上の集計で，サイレントとショートポーズとの連鎖は除外した．

3.2 実験条件

連続音節認識実験は，講演音声データベース "開放的融合研究「話し言葉工学」による「日本語話し言葉コーパス（モニター版 2002）」" [8] で行なう．音声の分析条件を表 1 に示す．学習データとして，学会講演における男性話者 115 名（約 17 時間）のデータを用いた．モーラ HMM は left-to-right 型で，HTK [9] を用いて学習した．モデルは母音，撥音，促音，無音は 3 状態，その他は 5 状態で，MDL 基準 [10] を用いて最大 64 混合まで各モデル各状態毎に混合数を最適化した 130 個のモデルを使用した．

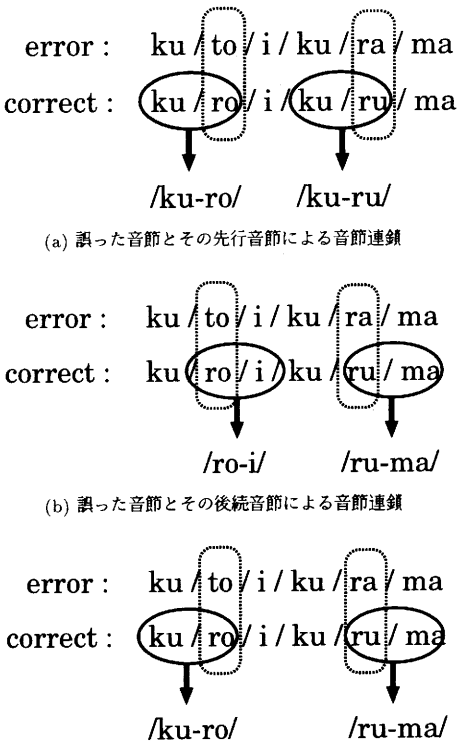
3.3 集計結果

連続音節認識の結果は，76.98%の音節正解率，67.38%の音節正解精度であった（表 5 の 'syl_MDL'）．この結果から音節の誤り傾向を集計した．

3.3.1 先行音節との連鎖

誤った音節とその先行音節による音節連鎖の誤り集計結果を表 2 に示す．

表 2(a) は音節連鎖の誤り回数毎に音節連鎖の種類数・誤り総数・全誤りに対する割合を示している．'音節連鎖数' は誤って認識された音節を含む音節連鎖の種類数，'誤り総数' は誤った音節連鎖の数を示し，'割合' は全誤りに対する割合を示す．これより，1 回以上の誤りを生じる音節連鎖は，論理的に可能な全音節連鎖数の 23%（$3863/128^2$）であり，誤り 100 回以上の音節連鎖の種類数はわずか 1.4%（$232/128^2$）で全誤りの約 50%を占めていることが分かる．



(a) 誤った音節とその先行音節による音節連鎖

(b) 誤った音節とその後続音節による音節連鎖

(c) 誤った音節を音素単位で考慮した音節連鎖

図 2 音節連鎖の構成方法

音声学的特徴別に誤りを幾つかのグループに分類した集計も行なった．誤り上位を表 2(b) に示す．'長母音' は長母音による母音の誤り，'N-xxx' は撥音 'N' が先行する場合の xxx の誤り，例えば xxx は摩擦音，無声破裂音などである．'q-無声破裂音' は先行音節が促音 'q' である場合の破裂音の誤りをそれぞれ表す．これより，文献 [5] と同様に，長母音は全エラーの約 14%を占め，極めて頻度の高いことが分かる．

3.3.2 後続音節との連鎖

誤った音節とその後続音節による音節連鎖の集計結果について，音節連鎖の誤り回数毎に集計した結果を表 3(a)，音声学的特徴別に幾つかのグループに分類したときの誤りに関して上位のものを表 3(b) にそれぞれ示す．

100 回以上誤っていた音節連鎖について注目し，誤った音節とその先行音節を集計した結果と，誤っていた音節とその後続音節を集計した結果を比較した．先行音節を考慮した集計結果には見られず，後続音節を考慮した結果に多く見られたのは，

表 2　先行音節を考慮した音節連鎖

(a) 音節連鎖のエラー頻度

誤り回数	音節連鎖数	誤り総数	割合
500 回以上	13	11380	9.8%
300 回以上	42	22035	19.0%
200 回以上	99	35850	30.9%
100 回以上	232	54342	46.9%
70 回以上	367	76680	66.3%
1 回以上	3863	115703	100%

(b) 音声学的特徴別のエラー頻度

	音節連鎖数	誤り総数	割合
長母音	103	15783	13.6%
N-摩擦音	20	856	0.7%
N-無声破裂音	22	889	0.7%
N-有声破裂音	16	1421	1.2%
N-鼻音	11	545	0.4%
q-無声破裂音	17	1391	1.2%

表 3　後続音節を考慮した音節連鎖

(a) 音節連鎖のエラー頻度

誤り回数	音節連鎖数	誤り総数	割合
500 回以上	13	10029	9.7%
300 回以上	41	20307	19.7%
200 回以上	90	32119	31.2%
100 回以上	251	54790	53.2%
70 回以上	366	64529	62.7%
1 回以上	3955	102953	100%

(b) 音声学的特徴別のエラー頻度

	音節連鎖数	誤り総数	割合
長母音	96	10917	10.6%
N-摩擦音	22	1183	1.1%
N-無声破裂音	24	391	0.3%
N-有声破裂音	17	678	0.6%
N-鼻音	12	769	0.7%
q-無声破裂音	17	998	0.9%

'母音-音節' という音節連鎖で，逆に後続音節を考慮した結果には見られず，先行音節を考慮した結果に多く見られたのは，長母音の誤りを含む '音節-母音' という形の音節連鎖であった．共通に見られる傾向としては，'su-ru' や 'de-su' など発話の語尾にあたる音節連鎖の誤りが多い．これは，講演音声の特徴である語尾のなまけによる誤りであると考えられる．

3.3.3　誤り音素を考慮した音節連鎖

誤った音節を音素単位とし，その先行，後続音節を考慮した集計を行なった．音節連鎖の誤り回数毎に集計した結果を表 4(a)，音声学的特徴別に幾つかのグループに分類したしたときの誤り頻度で上位のものを表 4(b) にそれぞれ示す．また、誤り頻度においての誤り総数と全誤りに対する割合をグラフで図 3 示す。誤り総数が先行／後続音節だけを考慮した場合に比べて増加しているのは，CV 両方が誤っている場合に先行／後続音節の両方を音節連鎖として考慮しているからである．

表 4　誤った音素位置を考慮した音節連鎖

(a) 音節連鎖のエラー頻度

誤り回数	音節連鎖数	誤り総数	割合
500 回以上	32	25479	19.0%
300 回以上	74	41515	31.0%
200 回以上	136	56584	42.3%
100 回以上	319	81695	61.1%
70 回以上	434	91318	68.3%
1 回以上	4152	133785	100%

(b) 音声学的特徴別のエラー頻度

	音節連鎖数	誤り総数	割合
長母音	103	23703	17.7%
N-摩擦音	24	2298	1.7%
N-無声破裂音	23	1048	0.8%
N-有声破裂音	18	1844	1.3%
N-鼻音	12	1975	1.5%
q-無声破裂音	23	2417	1.8%

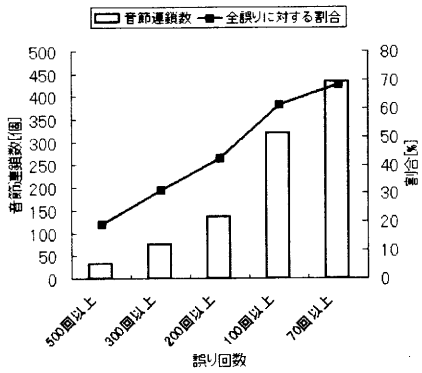


図 3　音節連鎖数と割合の関係

また，誤り総数で見ると，表 4(b) の結果は表 2(b) と表 3(b) をほぼ足し合わせた結果となっているのが分かる．この結果は誤りやすい音節の前後のコンテキストを考慮した結果であると考えられる．ここでもやはり長母音の誤りやなまけによる語尾の誤りが多く見られた．

4.　連続音声認識による評価

4.1　実験条件

音節連鎖モデルの効果を確かめるために，学習データに対して連続音節認識実験を行なった．実験条件については 3.2 節と同様で，今回作成した音節連鎖モデルは 6 状態〜10 状態となる．ベースとなるモーラモデルは MDL 基準を用いて最大 64 混合で混合数を最適化したモデル（誤りを集計するために用いたモデル）である．認識実験は，HTK のツールである HVite により行なう．

4.2　実験結果

音節連鎖モデルのベースとなるモーラモデルの認識性能を表 5 に示す．'syl' は従来のモーラモデル，'syl_MDL' は MDL 基

表 5　モーラモデルによる連続音節認識実験結果

type	mix	model	Gauss	Parm	Corr	Acc
syl	32	130	20160	1029k	76.21	66.78
syl	64	130	40320	2058k	79.61	71.17
syl_MDL	64	130	12714	650k	76.98	67.38

表 6　音節連鎖モデルによる連続音節認識実験結果

type	mix	model	Gauss	Parm	Corr	Acc
先行 2syl_100	64	362	12714	734k	82.98	68.93
後続 2syl_100	64	381	12714	740k	82.85	68.54
音素 2syl_200	64	301	12714	702k	82.26	68.59
音素 2syl_126	64	362	12714	734k	82.91	69.15
音素 2syl_100	64	491	12714	760k	83.21	69.53
音素 2syl_70	64	564	12714	792k	**83.54**	**69.82**

準を用いてガウス分布数をモデルの各状態毎に最大 64 混合まで最適化したモーラモデルをそれぞれ表す．また，'mix' はモデルの混合数（MDL を用いる場合は最大混合数），'model' はモデル数，'Gauss' は総ガウス分布数，'Parm' は遷移確率を含む総パラメータ数，'Corr' は音節正解率，'Acc' は音節正解精度を表す．結果より，32 混合から 64 混合へ混合数を増やすことで，Corr，Acc の両方で大きな改善が見られる．MDL 基準により混合数を最適化したモデルでは，'Gauss' と 'Parm' は約 3 分の 1 程度まで削減されるが，Corr，Acc は大きく低下する．

次に，音節連鎖モデルの認識性能を表 6 に示す．'先行 2syl_100' は，表 2(a) の結果より誤り回数 100 回以上の音節連鎖モデルをベースの 'syl_MDL' に追加したモデル，'後続 2syl_100' は，表 3(a) の結果より 100 回以上誤った音節連鎖モデルを追加したモデルを表す．'音素 2syl_200' は，表 4(a) の結果より 200 回以上誤った音節連鎖モデルを追加したモデルを表し，'音素 2syl_100' は同様に 100 回以上，'音素 2syl_70' は 70 回以上誤った音節連鎖をそれぞれ追加したモデルを表す．また，'音素 2syl_126' は，'先行 2syl_100' と比較するために追加モデル数を合わせた音節連鎖モデルを表す．結果より，ベースの 'syl_MDL' に比べ，音節連鎖を追加した全てのモデルで，Corr，Acc の両方で改善が見られる．モデル数がほぼ等しい '先行 2syl_100'，'後続 2syl_100'，'音素 2syl_126' で比較すると，'音素 2syl_126' が最も認識性能が良い．これは，音節連鎖選択の際，音素誤りを考慮した形で誤りを集計し選択した方法が，より正確に調音結合の影響を考慮することができることを示している．また，音素誤りを考慮した形で集計した音節連鎖モデルでは，モデル数増加に伴い Corr や Acc が改善されていることが分かる．これは音節連鎖モデルを追加することで音響モデルとしての性能が改善されることを示している．'音素 2syl_70' ではベースとなる 'syl_MDL' に比べ，Acc は 2.4%向上している．

表 7 に音節連鎖モデルを追加することによって改善された音節誤りの集計結果を示す．集計には '音素 2syl_100' の結果を用いた．音節連鎖モデルを追加することで，音節連鎖数・誤り総数ともに減少していることが分かる（図 4）．また，一回以上誤った音節連鎖の総数が 133785 個から 73536 個へと約半分に減少した．長母音の誤り総数は約 30%に削減された．

表 7　音節連鎖モデル追加後の誤り傾向

(a) 音節連鎖のエラー頻度

誤り回数	音節連鎖数	誤り総数	割合
500 回以上	3	2431	3.3%
300 回以上	13	6144	8.4%
200 回以上	35	11618	15.8%
100 回以上	115	22638	30.8%
70 回以上	232	32408	44.1%
1 回以上	4000	73536	100%

(b) 音声学的特徴別のエラー傾向

	音節連鎖数	誤り総数	割合
長母音	103	6782	9.2%
N-摩擦音	23	822	1.1%
N-無声破裂音	22	705	0.9%
N-有声破裂音	17	1055	1.4%
N-鼻音	12	633	0.8%
q-無声破裂音	19	656	0.8%

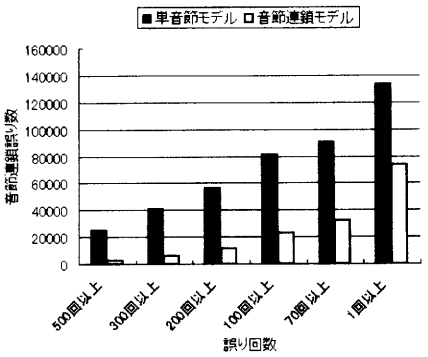


図 4　音節連鎖追加による誤り総数の減少

5. 大語彙連続音声認識による評価

前節では，モーラモデルと音節連鎖モデルについて音響モデルとしての性能を調べるために，学習データに対しクローズな実験を行なった．この節では各モデルについて大語彙連続音声認識実験によって評価を行ない，さらに現在主流となっているトライフォンモデルとの比較検討を行なう．

5.1 実験条件

テストには，学習に含まれない話者 7 名より，データベース附属の書き起こしの転記単位（200 ミリ秒以上の無音区間による区切り）に沿って切り出した発話から，各人毎に比較的長い 60 発話を選び，計 420 発話をテストセットとして用いた．デコーダには Julius rev.3.3p3 を用いた．

5.2 実験結果

ベースとなるモーラモデルの認識実験結果を表 8 に示す．学習データに対する連続音節認識実験結果で最も良い結果を得ていた 64 混合モデルでは，モデルのパラメータが多く推定精度が不十分なためか Acc は最も低く，一方 MDL 基準を導入した

表 8　モーラモデルによる連続単語認識実験結果

type	mix	model	Gauss	Parm	Corr	Acc
syl	32	130	20160	1029k	64.17	60.50
syl	64	130	40320	2058k	64.55	60.38
syl_MDL	64	130	12714	650k	66.36	62.24

表 9　音節連鎖モデルによる連続単語認識実験結果

type	mix	model	Gauss	Parm	Corr	Acc
先行 2syl_100	64	362	12714	734k	67.96	**64.41**
後続 2syl_100	64	381	12714	740k	67.72	63.87
音素 2syl_200	64	301	12714	702k	67.91	63.80
音素 2syl_126	64	362	12714	734k	67.94	64.23
音素 2syl_100	64	491	12714	760k	**68.31**	64.15
triphone	32	1577	42976	2192k	68.33	63.94

'syl_MDL' が最も高い認識精度を示した．MDL 基準により混合数を最適化することで約 2%Acc が向上し，標準的なトライフォンモデルを上回る認識性能が得られた．

音節連鎖モデルによる連続単語認識結果を表 9 に示す．表中の 'triphone' は標準的な状態共有トライフォンモデル（状態数 1343）を表す．結果より，連続単語認識実験でも全てのモデルで Corr と Acc は改善される．中でも先行音節を考慮した '先行 2syl_100' が最も良い認識性能を示した．ベースのモデルに比べ，Corr で 1.6%，Acc で 2.2%認識率が向上した．

連続音節認識実験の結果と連続単語認識実験の結果を比較すると，連続音節認識で良い認識性能を示したモデルが，必ずしも連続単語認識において良い性能を示すとは言えないことが分かる．これは，第一に連続音節認識実験は学習データに対するクローズな実験であることが原因と考えられる．また連続単語認識では言語モデルを用いるので，正解音節列の音響尤度が低く連続音節認識では不正解となる部分が言語尤度によって正解となる場合があり，その部分に関しては音節連鎖モデルを追加しても認識率には影響を及ぼさない．また，現在単語内でのみ音節連鎖モデルを適用しているため，追加した音節連鎖モデルが単語境界を跨いでいる場合には音節連鎖モデルの真価を十分に発揮できない．したがって，連続単語認識実験で高い認識率を得るためには，単語認識の結果から誤りを集計して単語認識に効果的な音節連鎖モデルを追加すると共に，単語境界を考慮した音節連鎖モデルの処理が必要であると考えられる．

6. ま と め

本稿では，誤り頻度の高い音節連鎖を選択し，ベースとなるモーラモデルを分布共有（PTM）して連結した音節連鎖モデルを提案した．ベースとなるモーラモデルに 300 程度の音節連鎖モデル追加することにより，従来のモーラモデルに比べ連続音節認識，連続単語認識の両方で認識精度の向上を得た．ベースとなるモーラモデルの混合数を MDL 基準で最適化した場合，標準的トライフォンの約 1/3 のパラメータ数のモデルセットでトライフォンを上回る認識精度を得ることができた．

今後は，単語境界を考慮した音節連鎖モデルの検討を行なう必要がある．

文　献

[1] 中川, 花井, 山本, 峯松, "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," 信学論誌, Vol.J83-D-II, No.6, pp. 1412-1421, 2000.

[2] 高橋, 中川, "コンテキスト依存音節 HMM の評価," 春季音響講論集, pp.97-98, 2001.

[3] 諸戸, 山本, 松本, "大語彙連続音声認識における音節モデルの改良," 春季音響講論集, pp.95-96, 2001.

[4] 河原達也, "話し言葉音声認識の概観," 信学技報, SP2000-95, 2000.

[5] 緒方, 有木, "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," 信学論誌, Vol.J86-D-II, No.11, pp.1523-1530, 2003.

[6] 李, 河原, 武田, 鹿野, "Phonetic tied-mixture モデルを用いた大語彙連続音声認識," 信学技報, SP99-100, 1999.

[7] 池田, 山本, 松本, 西谷, 宮澤, "音声認識における音節連鎖モデルの検討," 春季音響講論集, 1-4-3, 2003.

[8] 前川, "「日本語話し言葉コーパス」の構築," 話し言葉の科学と工学ワークショップ講演予稿集, pp.7-12, 2001.

[9] http://htk.eng.cam.ac.uk/

[10] K. Shinoda, D. Tran, K. Iso, "Efficient reduction of Gaussian components using MDL criterion for speech recgnition," Technical Report of IEICE, SP2001-83, 2001.