

## 形態素情報と単語内位置情報を用いた 話し言葉音声認識のための音響モデル

五十川賢造<sup>†</sup> 篠田 浩一<sup>†</sup> 嵯峨山茂樹<sup>†</sup>

<sup>†</sup> 東京大学大学院情報理工学系研究科

E-mail: †{isogawa,k-shino,sagayama}@hil.t.u-tokyo.ac.jp

**あらまし** 話し言葉における発声変形をモデル化するため、文脈決定木を用いた状態クラスタリングにおいて、音素コンテキストの情報以外に品詞情報を利用する方法と音素の単語内位置情報を利用する方法とを提案する。品詞情報を利用する方法では、話し言葉の音声データから推定した単語継続時間長をもとに、品詞を2つのクラスタに分類する。各音素についてクラスタ別に初期モデルを構築し、音素コンテキストに関する質問に加えて所属クラスタに関する質問を用いた決定木による状態クラスタリングを行う。単語内位置情報を利用する方法では、各音素について語頭か語末かそれ以外かを考慮した初期モデルを構築し、音素コンテキストに関する質問に加えて語頭語末に関する質問を用いた決定木による状態クラスタリングを行う。CSJ(Corpus of Spontaneous Japanese) コーパスを用いた連続音声認識実験を行った結果、品詞情報を利用した方法では単語正解精度がテスト話者7名の平均で最大で2.4ポイント向上し、単語内位置情報を用いたモデルでは単語正解精度が最大で6.1ポイント向上した。

**キーワード** 話し言葉, 音響モデル, 品詞, 決定木

## Acoustic modeling using word contexts for spontaneous speech recognition

Kenzo ISOGAWA<sup>†</sup>, Koichi SHINODA<sup>†</sup>, and Shigeki SAGAYAMA<sup>†</sup>

<sup>†</sup> Graduate School of Information Science and Technology, The University of Tokyo

E-mail: †{isogawa,k-shino,sagayama}@hil.t.u-tokyo.ac.jp

**Abstract** In this paper, we study state clustering using word contexts for speech recognition. In spontaneous speech, poorly articulated words often cause recognition error. To improve the recognition performance, we add two questions used in the phonetical decision tree based state clustering. One is a question about parts of speech, and the other is a question about the position of phones within a word. To apply the question about parts of speech, we classify parts of speech into two classes based on the word's duration estimated by using the corpus of spontaneous speech. After making HMMs for each class, we carry out state clustering using a context decision tree with the questions about the classes. To apply questions about the position of phones within a word, we make HMMs for phones at the beginning of the word, those for phones at the ending of the word, and those for phones at the other positions, separately. Then we carry out state clustering using a context decision tree with questions about phone positions. We carried out speech recognition experiments using CSJ(Corpus of Spontaneous Japanese). In the best case, the word accuracy improved by 2.4 points with the use of the former method, and it improved by 6.1 points with the use of the latter method.

**Key words** spontaneous speech, acoustic model, part of speech, decision tree

### 1. はじめに

近年、HMMを用いた音声認識の研究が盛んであり、ていねいに発声された音声 (read speech) に対して90%を

越える単語正解精度が得られている。しかし講演音声、対話音声等の話し言葉 (spontaneous speech) に対する認識性能は未だ十分な域に達していない。例えば、講演音

声の単語認識精度は70%程度に留まっている。その原因として、話し言葉音声は朗読音声より発声変形が大きいこと、必ずしも文法的でない表現が用いられることなどが挙げられている[1]。

発声変形に対処するためには、変形要因に応じてHMMの状態もしくはモデルそのものを適切に切り替えるアプローチが有効である。Read speechの音声認識では、前後の音素コンテキストによる発声変形に対処するためにトライフォンHMMが用いられている。Spontaneous speechには音素コンテキストに依存する発声変形に加え、語尾の曖昧化など、前後の音素コンテキストのみで記述することが困難な発声変形が多い。発声変形そのものをとらえることは難しいため、他の研究では話速の変化[2][3][4]、尤度の変化[5]や単語境界[6]等を利用して発声変形が起きる音素を予測し、モデルのトポロジー[2][3][5]や状態クラスタリング法[4][6]の改良を行っている。

青野ら[7]は、会話音声に対して、朗読発話で学習した音響モデルが与える尤度と自然発話で学習した音響モデルが与える尤度とを比較した。その結果を品詞別に分類し、助詞や、文末表現に用いられる「～です」「～ます」等が所属する助動詞において、自然発話で学習した音響モデルが与える尤度と、朗読発話で学習した音響モデルが与える尤度との差が大きくなることを示した。この結果は、品詞情報が発音のなまけがおきるかどうかの指標に使える可能性を示したと考えることができる。

また、英語を対象とした研究では、曖昧化母音やアクセント記号などの情報の他に、単語境界も考慮した状態クラスタリング[6]や、音節内位置情報を利用した音響尤度の重みづけ[8]などの手法がある。認識の際に発声変形を考慮した音素体系を利用する、音素の位置情報を利用するといった方法は日本語に対しても有効であると考えられる。日本語を対象とした研究では、read speechに対してのラベリングに関するものではあるが、単語のモーラ数、モーラ位置別に母音モデルを構築し精度を向上させた研究[9]がある。この研究では特に4モーラ単語の語頭の“o”の精度が向上しており、語頭と他の部分との音声特徴量の間に異なりがある可能性を示している。また語尾については曖昧化しやすいことが分かっており、適切にモデル化できれば、精度の向上に繋がると考えられる。

本論文では決定木による状態クラスタリングにおいて、品詞情報と音素の単語内位置情報の2つを導入することを提案する。話速が大きい品詞で発声変形が生じやすいと考え、話し言葉のデータベースから推定した単語の継続時間長を用いて、全品詞を話速が大きい品詞とそうでない品詞に分け、その結果を状態クラスタリングに反映

させる。また、語頭語末の音素を別音素と見なして学習を行い、語頭語末に関する情報を反映させた状態クラスタリングを行う。2つの情報に関する質問は、音素コンテキストに関する質問の効果にも影響を与える可能性があるため、品詞、単語内位置それぞれにおいて、質問の適用順序を変えた2つの音響モデルを構築する。

CSJ(Corpus of Spontaneous Japanese)コーパスモニタ版[10]を用いた話し言葉音声の連続音声認識を行ったところ、品詞情報を用いた音響モデルでは単語認識精度がテスト話者7名の平均で最大で2.4ポイント向上した。また、単語内位置情報を用いたモデルでは単語正解精度が最大で6.1ポイント向上した。

本論文ではまず、次章で品詞情報の導入について述べ、第3章で単語内位置情報の導入について述べる。その後第4章で、提案手法の評価結果を述べる。

## 2. 品詞依存音響モデル

品詞情報の音響モデルへの導入は、

- (1) 品詞毎の継続時間長の推定
  - (2) 品詞を速く発声される“Fast”クラスとそうでない“Slow”クラスの2クラスに分類
  - (3) 品詞情報を用いた状態クラスタリング
- の順で行う。以下各ステップについて述べる。

まず、話し言葉のデータベースのテキストデータを形態素解析した単語列と、それに対応する音声データの間のアライメントをビタビアルゴリズムを用いて求める。この結果をもとに音素数で正規化を行った品詞毎の平均継続長を求める。コーパス中に出現する品詞 $K$ に属する単語 $w_K$ に対し、出現総数を $c_{w_K}$ 、音素数を $p_{w_K}$ 、 $i$ 番目のサンプルの継続長を $d_{w_K}^i$ とする。単語 $w_K$ の音素数で正規化された平均継続長 $\bar{d}_{w_K}$ は、

$$\bar{d}_{w_K} = \frac{1}{c_{w_K}} \sum_{i=1}^{c_{w_K}} \frac{d_{w_K}^i}{p_{w_K}} \quad (1)$$

となる。コーパス中の品詞 $K$ に所属する単語の総数を $C_K = \sum_{w_K} c_{w_K}$ とする。音素数で正規化された品詞 $K$ の平均継続長 $\bar{D}_K$ を、

$$\bar{D}_K = \frac{1}{C_K} \sum_{w_K} c_{w_K} \bar{d}_{w_K} \quad (2)$$

とする。次に、品詞を $D_K$ が短い順にソートし、上位 $n$ 個をFastクラス、それ以外をSlowクラスとする。 $n$ はデータベース全体の出現単語数 $C_{all}$ と、データベース内でFastクラスに属する単語の出現数 $C = \sum_{K \in \text{Fast}} C_K$ の比 $\gamma = \frac{C}{C_{all}}$ が適当な値になるように決定する。 $\gamma$ は実験から適切に決める。

Fastクラスの単語に属する音素とSlowクラスの単語

「の」 : n\* o\*  
 「私」 : w a t a s h i

図1 品詞情報を用いた辞書変更例: \*はFastクラスに対応する単語用の音素につけたラベル。1つの単語の中でFastクラス用のモデルとSlowクラス用のモデルが混在するエントリは作成されない。また、全ての単語は必ずSlowクラス、Fastクラスのどちらか片方にのみ所属する。

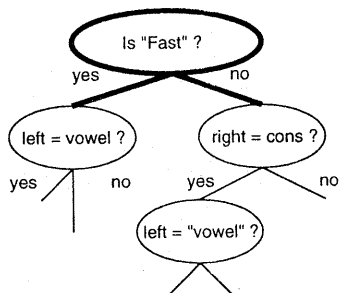


図2 品詞クラス別音響モデル用音韻決定木: 中心音素のクラスに関する質問が最初に適用される。

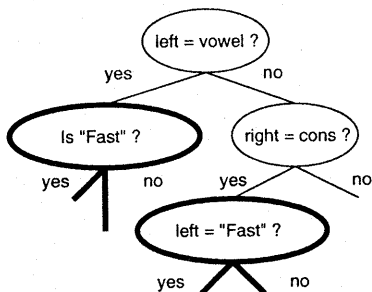


図3 品詞クラス依存音響モデル用音韻決定木: クラスに関する質問が適用されるノードは決定木クラスタリングの枠組みの中で決定される。

に属する音素は別の音素として初期モデルを構築する(図1)。初期モデルのトライフォンの学習を行った後、Slow、Fast両クラスでの、中心の環境非依存モノフォンの同じ状態を全てまとめ、決定木を用いた状態クラスタリングを行う。

話速によって隣接する音素コンテキストの影響が変わる可能性があるため、品詞クラスに関する質問が適用される前と後で、音素コンテキストに関する質問の効果が変化し得る可能性がある。そこで、決定木を用いた状態クラスタリングの際に、最初の質問を「中心エントリはFastクラスのものであるか」に固定し、以後は音素コンテキストに関する質問のみを用いる「品詞クラス別音響モデル」(図2)と、音素コンテキストに関する質問に加え「先行エントリはFastクラスのものであるか」「中心エントリはFastクラスのものであるか」「後続エントリは

淹れたて : < i r e t a t e >  
 入れた手 : < i r e t a > < t e >

図4 単語内位置情報を用いた音響モデルによる単語境界のモデル化例: < が語頭に対応する音素につくラベルであり、> が語末に対応する音素につくラベルである。語頭、語末を兼ねる音素は、語頭、語末、それ以外の音素いずれとも別の音素として初期モデルを構築する。

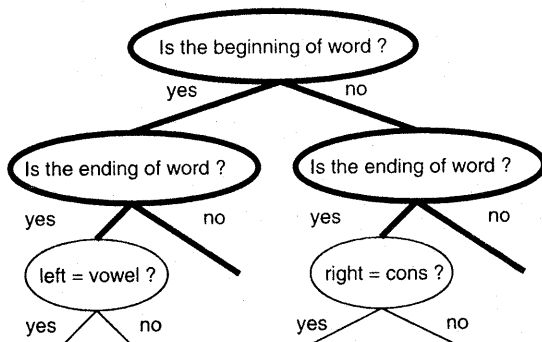


図5 単語内位置別音響モデル用決定木の例: 最初に中心音素の位置に関する質問が適用される。

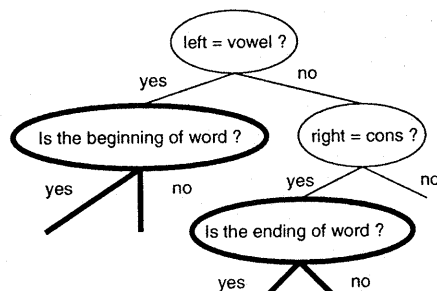


図6 単語内位置依存音響モデル用決定木の例: 中心音素の位置に関する質問が適用されるノードは、決定木クラスタリングの枠組みの中で決定される。

Fastクラスのものであるか」の3つの質問を決定木の任意のノードで用いる「品詞クラス依存音響モデル」(図3)の2つを構築する。

### 3. 単語内位置依存音響モデル

単語の先頭と終端の音素はそれ以外の音素とは別の音素として初期モデルを構築する。単語の境界は例えば図4のようにモデル化される。トライフォンを学習した後、中心の環境非依存モノフォンの同一である状態を全てまとめ、決定木を用いた状態クラスタリングを行う。この際、決定木の質問には、音素コンテキストに関する質問の他に「中心音素は語頭であるか」「先行音素は語頭であるか」「後続音素は語末であるか」「中心音素は語末であるか」の4つの質問を加える。

表1 実験条件

継続長推定用データ 分析条件	CSJ コーパスの全 391 講演 16kHz サンプリング, 30msec 窓幅, 10msec フレームシフト
特徴量	MFCC12 次元 + ΔMFCC12 次元 + Δ 対数パワー
学習データ	CSJ コーパスより 200 講演 (テスト話者を除く男性の講演)
HMM のパラメタ数	3000 状態, 16 混合, 対角共分散
テスト話者 (話者適応用データ)	CSJ コーパスより 7 講演分 a01m0007, a01m0035, a01m0074, a10m0117, a03m0100, a05m0031, a06m0134

表2 実験において、Fast クラスに選ばれた品詞

○ 動詞-接尾	○ 名詞-接尾-形容動詞語幹
○ 名詞-非自立-助動詞語幹	○ 記号-一般
○ 動詞-非自立	○ 名詞-非自立-形容動詞語幹
○ 名詞-非自立-一般	○ 助詞-副助詞
○ 名詞-非自立-副詞可能	○ 形容詞-非自立
○ 名詞-ナイ形容詞語幹	○ 助詞-副詞化
○ 名詞-接尾-サ変接続	○ 助動詞
○ 動詞-自立	

1型アクセントを考慮すると、語頭の子音と隣接する母音の間の音素環境は他の場合と大きく変わることが予想される。また、曖昧化が起こる子音についても、音素環境は他の場合と変わると考えられる。結果として品詞情報の場合と同じく、語頭語末に関する質問の前後で、音素環境に関する質問の効果が変わる可能性がある。そのため、最初に「中心音素は語頭であるか」「中心音素は語末であるか」の2質問を適用した後、音素コンテキストの質問に加え「先行音素は語頭であるか」「後続音素は語末であるか」の2質問を用いる「単語内位置別音響モデル」(図5)と、4つの質問を適用する場所に制限を設けない「単語内位置依存音響モデル」(図6)の2つを構築する。

#### 4. 音声認識実験

##### 4.1 実験条件

講演原稿の書き起こしを目的とした、対角共分散行列を用いた混合ガウス分布トライフォンHMMによる講演音声の認識実験を行った。モデルの学習・テストのためのデータにはCSJ(Corpus of Spontaneous Japanese)コーパスモニタ版(2001)[10]を用いた。認識エンジンにはJulius3.1[11]を用い、形態素解析にはChasen ver. 2.02[12]を用いた。言語モデルはCSJコーパスに付属している講演音声用言語モデル(2002-06; 京都大学)を用い、辞書についてもCSJコーパス付属の辞書を用いた。表1に実験条件を示す。CSJコーパスの音声データ全てを、書き起こしテキストをもとに500msec以上の無音区間を含まない20sec以下の区間に分割した。また、テキストデータを全て形態素解析し、未知語を含む区間は全て取り除いた。実験で用いる学習用ラベルは全て形態素解析後のテキストから作成した。

品詞のクラスタリングに用いるテキストと音声データとの間の対応付けには、音声認識システム[11]付属の性別非依存モノフォンモデル(混合数16)と、CSJコーパスに収録されている全講演のデータを用いた。形態素解

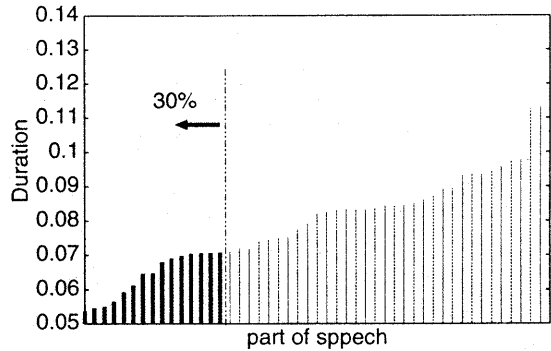


図7 形態素毎の平均継続時間長: 継続時間長が短い順にソートした。左か15本の太線が $\gamma = 0.3$ 、すなわちデータベース内ののべ出現単語総数の30%の単語がFastクラスとされる場合のFastクラスの形態素に相当する

析後のテキストデータと音声データ間のビタピアライメントを求め、その結果をもとに各単語の継続時間を推定した。2章で定義した閾値 $\gamma$ は0.3とした。Fastクラスに選ばれた品詞を表2に示す。また、品詞別の継続時間長の分布を図7に示す。

不特定話者モデルを初期モデルとして、7名のテスト話者それぞれについて、同一話者のデータ全てを使用した教師なし話者適応を行った。話者適応はMLLR(HTK, Ver. 3.0, HEAdapt[13])を用いた。

##### 4.2 実験結果

音素コンテキストに関する質問のみを用いて状態クラスタリングを行ったベースラインモデル、品詞クラスタ依存音響モデル、単語内位置依存音響モデルの3つを用いた実験結果を表3に示す。また、品詞クラスタ別音響モデルと、位置別音響モデルによる実験結果を表4に示す。また、表5にはモデル毎に、置換誤り(Sub.)、削除誤り(Del.)、挿入誤り(Ins.)のデータを示す。

品詞情報を用いたモデルでは、話者適応無しの品詞クラスタ別モデルで、認識精度(WA.)が45.1%から47.5%に向上した。それ以外のモデルは認識精度の変化がほとんど見られなかった。品詞クラスタ依存モデルより、品詞クラスタ別モデルの方が高い認識精度が得られた。クラスタリング方法や話者適応の有無にかかわらず、置換誤

表3 品詞クラスタ依存音響モデル、単語内位置依存音響モデルの認識実験結果 [%](PC.: Percent Correct, WA.: Word Accuracy)

モデル	ベースライン		品詞クラスタ依存		単語内位置依存	
-話者適応無し-						
話者	PC.	WA.	PC.	WA.	PC.	WA.
am035	55.4	42.5	53.5	41.5	57.2	47.4
am007	64.6	45.4	64.5	49.1	68.3	55.0
am074	67.2	56.9	66.0	53.2	70.0	60.9
am117	62.5	45.0	60.3	44.9	64.9	52.2
am100	59.7	48.0	57.6	48.9	61.3	52.9
am031	52.8	42.4	51.4	41.3	55.3	47.8
am134	50.0	42.9	50.5	42.0	52.7	47.5
全話者平均	58.9	45.1	57.6	45.0	61.5	51.4
-話者適応有り-						
話者	PC.	WA.	PC.	WA.	PC.	WA.
am035	59.6	46.4	60.0	48.7	63.0	54.1
am007	72.2	55.4	72.4	56.1	74.5	63.1
am074	72.7	63.5	70.6	62.0	74.9	68.2
am117	69.7	53.1	69.0	54.5	72.7	60.7
am100	65.8	55.3	63.4	54.7	66.8	59.8
am031	61.2	52.8	60.4	53.0	62.9	56.6
am134	57.3	51.1	55.1	48.4	59.9	54.3
全話者平均	65.9	53.2	65.0	53.6	68.3	59.3

表4 品詞クラスタ別音響モデル、単語内位置別音響モデルの認識実験結果 [%](PC.: Percent Correct, WA.: Word Accuracy)

モデル	ベースライン		品詞クラスタ別		単語内位置別	
-話者適応無し-						
話者	PC.	WA.	PC.	WA.	PC.	WA.
am035	55.4	42.5	56.1	45.0	56.4	46.2
am007	64.6	45.4	64.9	51.5	67.3	54.6
am074	67.2	56.9	65.7	56.9	65.3	57.3
am117	62.5	45.0	61.5	46.0	64.1	51.4
am100	59.7	48.0	59.2	51.0	62.7	56.4
am031	52.8	42.4	52.8	44.6	55.0	47.0
am134	50.0	42.9	50.8	45.3	53.7	48.3
全話者平均	59.0	45.1	58.7	47.5	60.8	50.8
-話者適応有り-						
話者	PC.	WA.	PC.	WA.	PC.	WA.
am035	59.6	46.4	60.4	49.5	61.9	53.4
am007	72.2	55.4	71.0	57.9	73.4	62.6
am074	72.7	63.5	72.8	65.4	72.3	65.5
am117	69.7	53.1	67.8	53.3	72.7	60.8
am100	65.8	55.3	62.2	53.3	67.9	62.2
am031	61.2	52.8	60.1	52.0	62.9	56.7
am134	57.3	51.1	54.2	49.4	59.7	55.0
全話者平均	65.9	53.2	64.4	53.7	67.9	59.2

りと挿入誤りは減少したが、削除誤りが増加した。

単語内位置情報を用いたモデルでは、全ての条件下でベースラインモデルを上回る認識性能を得た。単語認識精度は単語内位置別モデルに比べ、単語内位置依存モデルの方がわずかに高い。またクラスタリング方法や話者適応の有無にかかわらず、置換誤りと挿入誤りは減少

表5 モデル別認識結果 [%](PC.: Percent correct, WA.: Word Accuracy, Sub.: Substitution error, Del.: Deletion error, Ins.: Insertion error)

Model	PC.	WA.	Sub.	Del.	Ins.
-話者適応無し-					
ベースライン	58.9	45.1	33.5	7.5	13.9
品詞クラスタ別	58.7	47.5	32.0	9.3	11.2
品詞クラスタ依存	57.6	45.0	32.7	9.7	12.6
単語内位置別	60.8	50.8	30.9	8.4	10.0
単語内位置依存	61.5	51.4	30.0	8.6	10.1
-話者適応有り-					
ベースライン	65.9	53.2	28.0	6.1	12.6
品詞クラスタ別	64.4	53.7	27.9	7.7	10.7
品詞クラスタ依存	65.0	53.6	27.7	7.3	11.4
単語内位置別	67.9	59.2	25.1	7.0	8.7
単語内位置依存	68.3	59.3	25.1	6.6	9.1

表6 品詞クラスタ依存音響モデルによって誤りが改善した例

正解	利用する場面が多くなってきた
ベースライン	寄与する場面がをてました
品詞クラスタ別	寄与する場面が多くなってきた

し、削除誤りが増加した。

### 4.3 考察

#### 4.3.1 品詞情報を用いた音響モデル

品詞クラスタ依存音響モデルと品詞クラスタ別音響モデルは共に、単語正解率(PC.)が低下しているにもかかわらず、単語認識精度はほぼ同等以上となった。この結果は品詞情報が挿入誤りの削減に有効であることを示している。また、単語正解率が低下した話者の認識結果にも、表6のように文末表現の認識結果が改善した例も見られた。この結果は、品詞クラスタ依存音響モデルが文末表現の認識精度向上に効果がある可能性を示している。話者毎の結果を見ると認識精度が低下している話者が見られた。Fastクラスに分類すべき単語は話者に依存する可能性を示唆する。品詞クラスタ別モデルの認識精度が品詞クラスタ依存モデルの認識精度に比べ高かった理由は、話速により音素環境に関する質問の効果が変化するためであると考えられる。

#### 4.3.2 単語内位置情報を用いた音響モデル

状態クラスタリング方法と話者適応の有無にかかわらず、認識精度の向上が得られた。また挿入誤りの減少が見られるが、これは単語境界が適切にモデル化され、単語間への短い単語の挿入が起りにくくなったためと考えられる。認識結果を見ると音素数の少ない単語の誤りが改善されている例が目立った(表7)。語頭語末を考慮した状態クラスタリングにおいては、音素数が3以下の単語に対し、単語モデルを作成したことに類似した効果が得られるためであると考えられる。

表 7 単語内位置依存音響モデルによって誤りが改善した例

正解	機能 を 持っ て いる 可 能 性 が あ る と
ベースライン	機能 を 持っ て いる か の せ ん
単語内位置依存	機能 を 持っ て いる 可 能 性 が

## 5. おわりに

話し言葉音声における発声変形に対処するため、音響モデルの状態クラスタリングの際に、品詞の情報と音素の単語内位置の情報を用いる方法を提案した。品詞情報を用いる方法では単語正解精度はテスト話者7人の平均で最大で2.4ポイント向上し、単語内位置情報を用いる方法では単語正解精度は最大で6.1ポイント向上した。

今後の課題としては、品詞のクラスタリング方法の見直し、削除誤りの増加要因の解析、語頭・語末以外の単語内位置情報の検討、品詞を用いた状態クラスタリングと単語内位置情報を用いた状態クラスタリングの統合、本研究の手法の朗読音声に対する効果の検証等がある。

### 文 献

- [1] 河原 達也, “話し言葉音声認識の概観,” 第2回話し言葉の科学と工学ワークショップ研究発表資料集, Vol. 1, pp. 327-331, 2002.
- [2] 南條 浩輝, 河原 達也, “発話速度に依存した音素・音節モデル化の検討,” 話し言葉の科学と工学ワークショップ研究発表資料集, Vol. 1, pp. 340-347, 2001.
- [3] 吉田 明弘, 鈴木 基之, 牧野 正三, “自発音声の音素継続時間に基づいた音素認識の検討,” 日本音響学会 2002 年秋期研究発表会講演論文集, Vol. 1, pp. 91-92, 2002.
- [4] Christian Fügen, Ivica Rogina, “Integrating dynamic speech modalities into context decision trees,” Proc. ICASSP2000, Vol. 3, pp. 1277-1280, 2000.
- [5] 李 晃伸, 米良 祐一郎, 鹿野 清宏, “Julius によるマルチパス音韻モデルを用いた対話音声認識,” 第2回話し言葉の科学と工学ワークショップ講演予稿集, pp. 121-128, 2002.
- [6] Kai-Fu Lee, “Large-Vocabulary Speaker-Independent Continuous Speech Recognition: The SPHINX System,” CMU-CS-88-148, Carnegie Mellon University, Pittsburgh, PA, 1988.
- [7] 青野 邦生, 安田 圭志, 竹野 寿幸, 山本 誠一, 柳田 益造, “言語情報を考慮した発話スタイル依存音響モデル自動選択の予備検討,” 日本音響学会 2002 年秋期研究発表会講演論文集, Vol. 1, pp. 89-90, 2002.
- [8] Eric Fosler-Lussier, Steven Greenberg and Nelson

Morgan, “Incorporate contextual phonetics into automatic speech recognition,” Proc. ICPHS-99, pp. 611-614, 1999.

- [9] 前田 智広, 村上 仁一, 池原 悟, “モーラ情報を用いた音素ラベリング方式の検討,” 日本音響学会 2000 年秋期研究発表会講演論文集, Vol. 1, pp. 135-136, 2000.
- [10] 前川 喜久雄, 籠宮 隆之, 小磯 花絵, 小椋 秀樹, 菊地 英明, “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. 69-76, 2001.
- [11] 鹿野 清宏, 伊藤 克亘, 河原 達也, 武田 一哉, 山本 幹雄, “音声認識システム,” オーム社, 2001.
- [12] 松本 裕治, 北内 啓, 山下 達雄, 平野 善隆, 松田 寛, 浅原 正幸, “日本語形態素解析システム「茶筌」version 2.0 使用説明書 第二版,” Information Science Technology Report NAIST-IS-TR99012, 奈良先端科学技術大学院大学, 1999.
- [13] Steve Young, Dan Kershaw, Julian Odell, Dave Ollason and Valtcho Valtchev, Phil Woodland, “The HTK Book (for HTK Version 3.0),” Microsoft Corporation, 2000.