音声言語情報処理 44－32
(2002. 12. 20)

講演音声認識のための言語モデルの教師なし適応

南條 浩輝† 河原 達也† 山田 篤††,††† 内元 清貴†††

† 京都大学 情報学研究科 知能情報学専攻
〒606-8501 京都市左京区吉田本町
†† (財) 京都高度技術研究所
〒600-8813 京都市下京区中堂寺南町 17
††† 独立行政法人通信総合研究所
〒619-0289 京都府相楽郡精華町光台 2-2-2
E-mail: †{nanjo,kawahara}@kuis.kyoto-u.ac.jp, ††yamada@astem.or.jp, †††uchimoto@crl.go.jp

あらまし 大語彙の話し言葉音声認識における言語モデルの教師なし話者適応について報告をおこなう．講演などの話し言葉においては，話題の他に文末表現などで発話の傾向やその発音が話者間で大きく異なるため，言語・発音モデルの話者性への適応が必要である．本稿では，教師なし言語モデル話者適応手法として (1) 認識結果を直接用いて適応する手法，及び (2) 発話文単位で類似テキストを選択しそれを用いて適応する手法，を提案する．その上で発音変動のモデル化についても検討し，話者適応の枠組みに統合することで，言語表現の傾向と発音変動の両方を同時にモデル化する．実際の講演の音声認識実験において提案手法それぞれの有効性を確認した．提案手法の統合の効果も確認し，単語誤り率を 4.4%改善できた．

キーワード 音声認識，話し言葉，講演，言語モデル，教師なし適応

Unsupervised Language Model Adaptation for Lecture Speech Recognition

Hiroaki NANJO†, Tatsuya KAWAHARA†, Atsushi YAMADA††,†††, and Kiyotaka UCHIMOTO†††

† Graduate School of Informatics, Kyoto University,
Yoshida-Hommachi, Sakyo-ku, Kyoto 606-8501 Japan
†† Advanced Software Technology & Mechatronics Research Institute of KYOTO (ASTEM)
17 Chudoji Minami-Machi, Shimogyo, Kyoto 600-8813, Japan
††† Comunications Research Laboratory
2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0289 Japan
E-mail: †{nanjo,kawahara}@kuis.kyoto-u.ac.jp, ††yamada@astem.or.jp, †††uchimoto@crl.go.jp

Abstract This paper addresses speaker adaptation of language model in large vocabulary spontaneous speech recognition. In spontaneous speech, the expression and pronunciation of words vary a lot depending on the speaker and topic. Therefore, we present unsupervised methods of language model adaptation to a specific speaker by (1) making direct use of the initial recognition result for generating an enhanced model, and (2) selecting similar texts for adaptation utterance by utterance. We also investigate the pronunciation variation modeling and its adaptation in the same framework. It is confirmed that all proposed adaptation methods and their combinations reduced the perplexity and word error rate in transcription of real lectures.

Key words automatic speech recognition, spontaneous speech, lecture speech, language model, unsupervised adaptation

1. はじめに

我々は、開放的融合研究「話し言葉工学」で構築されている大規模な日本語話し言葉コーパス (CSJ: Corpus of Spontaneous Japanese) [1] [2] を用いて講演音声の認識の研究をおこなっている。

音声認識においては一般に，話者依存モデルを作成・使用することで，話者独立モデルを用いる場合に比べて高い認識性能が得られる．話者依存モデルの作成は，モデル適応を用いるのが一般的である．音響モデルについては，CSJ の講演音声認識タスクにおいて話者適応による話者依存モデルの作成・評価がおこなわれており，その効果も確認されている [3] [4] [5].

一方，言語モデル適応に関する従来研究は主に，特定のタスクや話題への適応を目的としている [6] [7] [8]．講演のような話し言葉においては，文末表現やその発音は話者間で差が大きいため，言語モデルを話題だけでなく話者性に適応させることが重要である．

このような背景に基づいて，本稿では，言語モデルの教師なし話者適応について検討する．発音変動のモデル化についても検討をおこない，話者適応の枠組みに統合することで，各話者の言語表現の傾向と発音変動の両方を同時にモデル化する．

以下，2 章でテストセットとベースラインシステムについて述べる．3 章で認識結果を直接用いた言語モデル適応について，4 章でテキスト選択による言語モデルの適応について説明する．5 章では発音の変動をモデル化した上で話者適応の枠組みへの統合をおこない，それらの効果を調べる．

2. タスクとベースラインシステム

実験に使用した日本語話し言葉コーパス（CSJ）について簡単に説明し，次にテストセットと実験に使用した学習データについて述べる．

2.1 テストセット

CSJ は学会講演と特定の話題について独話した模擬講演から構成される．テストセットは表 1 に示す 10 名の話者の学会講演であり，文献 [4] と同一のものである．いずれも講演に熟練した男性話者による講演（学会講演）であり，原稿を用いずに話している．非公開の会合での講演も一部含まれており，それらは配布されているモニタ版には含まれていない（表 1 の A4, A5, B5）．テストセット音声はパワーと零交差数に基づいて自動的に分割し，これを入力発話として扱う．

2.2 言語モデル

言語モデルの学習には，2002 年の 2 月時点で利用できる講演（1099 講演）の書き起こしを用いた．形態素は，国立国語研究所で定義された短単位 [9] に基づいており，形態素解析システムは通信総合研究所で最大エントロピー法により CSJ を用いて統計的に学習されたものを用いる [10]．総形態素数は約 3.15M（ポーズモデル<sp>, <sil>を除いた場合は，約 2.82M）

表 1 テストセットとベースラインシステムによる結果
Table 1 Test-set lectures and baseline result

lecture ID	#morphemes (#pauses)	duration (min.)	WER (%)	perp-lexity	OOV (%)
A1	7355 (688)	28	38.5	72.40	1.77
A2	6109 (482)	27	31.3	83.72	2.05
A3	5269 (426)	23	39.2	58.90	2.39
A4	7747 (739)	42	33.4	67.78	2.49
A5	3561 (227)	15	29.7	68.94	1.83
B1	5413 (798)	30	22.5	55.33	1.20
B2	2843 (253)	12	24.4	61.21	2.18
B3	11781 (1334)	57	35.4	78.13	2.67
B4	3179 (350)	15	29.7	52.83	1.51
B5	3227 (238)	14	36.7	63.01	1.77
total	56484 (5535)	263	33.1	68.18	2.10

A1(A01M0035), A2(A05M0031), A3(A06M0134), A4(KK99DEC005), A5(YG99MAY005), B1(A01M0007), B2(A01M0074), B3(A02M0117), B4(A03M0100), B5(YG99JUN001)

表 2 音響・言語モデル学習データ
Table 2 Training data of acoustic and language model

	acoustic model	language model
#lectures	394	1099
data amount	60 hours	3.15M morphemes

音響モデルの学習データは学会講演のみ

である．ただし，<sil>は 1000msec 以上のポーズに，<sp>はそれ未満のポーズに割り当てている．これまで我々は形態素解析システムとして ChaSen を用いていきたが，今回の形態素体系の変更が音声認識に及ぼす影響はほとんどない [11].

語彙は学習コーパスに 4 回以上出現した形態素（16029 形態素）で構成し（カットオフ 3），これによるテストセットのカバレージは 97.9%であった．CMU-Cambridge SLM toolkit ver.2 を用いて逆向きの単語 3-gram モデルを作成し，ベースラインの言語モデルとした．back-off 平滑化には Witten-Bell 法を用いている．ベースラインモデルでは，表記が同一の形態素は同じエントリとして扱い，複数の読みがある場合は単語辞書に追加した．

2.3 音響モデル

音響モデルは混合連続分布 HMM（対角共分散）であり HTK で作成した．音声データ（16kHz, 16bit）をフレーム長 25msec のハミング窓，フレーム周期 10msec で音響分析をおこなった．各フレーム毎に MFCC（12 次元），ΔMFCC（12 次元），ΔPower（1 次元）を計算し，計 25 次元の特徴量ベクトルを求めた．

音素は 43 種類とし，各音素は 3 状態 left-to-right HMM（飛び越し遷移なし）でモデル化した．約 60 時間の男性話者の学会講演で学習をおこない，男性用の性別依存 PTM （Phonetic Tied-Mixture） triphone モデル [12]（129 コードブック x192 混合; 3000 状態）を作成した．

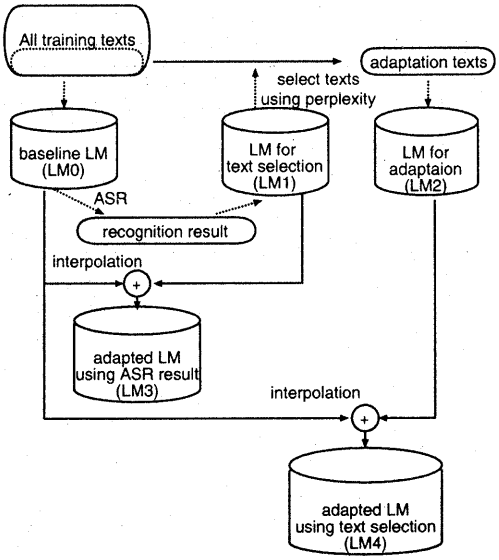

図 1　言語モデル適応の流れ
Fig. 1　Flowchart of language model adaptation

表 3　認識結果を用いた言語モデル適応の結果
Table 3　Result of adaptation using initial recognition result

	WER(%)	perplexity
baseline ($LM0$)	33.1	68.18
adapted using ASR result ($LM3$)	31.0	52.37

2.4　ベースラインシステムの性能

認識エンジンには Julius rev3.2 を用いた．各テスト講演に対する単語誤り率（WER）とテストセットパープレキシティを表 1 に示す．10 講演全体の単語誤り率は 33.1%，パープレキシティは 68.18 であった．パープレキシティの計算は，ポーズを含み未知語は除いておこなっているため，以前の我々の報告 [11] と直接比較はできない．単語誤り率の算出には，未知語は含むがポーズは含んでいない．

3.　認識結果を用いた言語モデル適応

まず，認識結果を用いた適応手法について述べる．各講演は比較的長く（テキストサイズが比較的大きく），各話者の話題や言い回しなどの発話の特徴を多く含んでいると考えられ，認識結果を用いることで話者性に適応できると考える．文献 [13] で，CSJ に対する言語モデル適応の検討がおこなわれており，認識結果を適応データとして用いることで改善が得られることが報告されている．ただし，適応パラメータは事後的に決定されている．

言語モデル適応の流れを図 1 に示す．まず，認識結果を用いてバックオフ単語 3-gram モデル（$LM1$）を作成する．その際，一度しか出現しなかった単語 2-gram 及び 3-gram エントリは除いた．次に，このモデル（$LM1$）とベースラインモデル（$LM0$）とを式 (1) に基づいて線形補間をおこない，適応モデルを得る．

original text
…で えー <sil> この 研究 は あのー <sp> 音声 認識
に おける <sp> 言語 モデル …
utterances
…で えー <sil> この 研究 は あのー <sp>
<sil> この 研究 は あのー <sp> 音声 認識 に おける <sp>
<sp> 音声 認識 に おける <sp> 言語 モデル …
<sp> 言語 モデル …

図 2　発話文単位の定義と分割
Fig. 2　Definition of utterance unit

$$P_{adapt}(w) = \lambda \cdot P_{rec}(w) + (1-\lambda) \cdot P_{base}(w) \qquad (1)$$

ここで，$P_{base}(w)$ はベースライン言語モデルによる単語列 w の確率，$P_{rec}(w)$ は認識結果から作成した言語モデルによる確率である．$P_{adapt}(w)$ は適応後のモデルによる確率であり，線形補間係数 λ は式 (2) の EM アルゴリズムを用いて推定する．

$$\hat{\lambda} = \sum_{i=1}^{N} \frac{\lambda \cdot P_{rec}(w_i)}{\lambda \cdot P_{rec}(w_i) + (1-\lambda) \cdot P_{base}(w_i)} \qquad (2)$$

ここで w_i は，適応ターゲット講演の正解単語列の i 番目の単語であるが，実際には正解単語列を用いることはできない．認識結果の単語列を類似テキストとして使用することを試みたが，補間係数 λ が大きく推定されてしまい，正解単語列に対するパープレキシティを下げることはできなかった．そこで，development-set を用いて λ の推定をおこなう．テストセット 10 講演を半分に分け，一方を development-set としてもう一方のテストセット講演（evaluation-set と表記する）のための線形補間係数 λ を推定する．本実験ではテストセットを表 1 に示すように A1 から A5 までと B1 から B5 までの二つに分けた．認識結果から学習した言語モデル（$LM1$）とベースライン言語モデル（$LM0$）の線形補間には，相補的バックオフアルゴリズム [14] を利用した．

この適応実験の結果を表 3 に示す．A1 から A5 を development-set，B1 から B5 を evaluation-set とした場合，線形補間係数 λ は 0.415，development-set と evaluation-set を入れ替えた場合も 0.415 であった．単純な教師なしの適応手法で 2.1%の単語誤り率の改善を得た．

4.　テキスト選択を用いた言語モデル適応

次に，テスト講演に類似したテキストデータに重みを付けて混合する適応手法について述べる．

4.1　テキスト選択手法

テスト講演に類似したテキストデータを選択するために，テスト講演の予稿や同じ学会でおこなわれた他の講演などの知識を用いることも考えられる．実際に我々も以前，予稿テキス

表 4　テキスト選択を用いた言語モデル適応の結果
Table 4　Result of adaptation using text selection

	WER(%)	perplexity
baseline ($LM0$)	33.1	68.18
adapted using text selection ($LM4$)	32.6	65.60

トを用いた言語モデル適応を試み，テストセット A1 に対して 0.5%，B1 に対して 3.0%の単語誤り率の改善を得た [15]．しかし，一般に予稿テキストをオンラインで入手するのは困難である場合が多い．

本稿では，事前知識を用いない類似テキストの選択手法を検討する．このような選択手法には，$tf \cdot idf$ などの話題依存の単語の頻度に基づく方法 [16] や，パープレキシティやカバレージに基づく方法がある．本研究では，類似尺度として単語 3-gram によるパープレキシティを採用する．これは，話者ごとの言い回しなどの発話の傾向を反映していると考えるためである．

4.2　発話文ごとのテキスト選択

テキスト選択は図 2 に示すように，ポーズで区切られた発話（発話文）ごとにおこなう．ポーズが 2 回出現した時点で区切り，発話文と定義する．次の発話文の開始位置は，一つ前のポーズに遡って設定する．これは，ポーズの前後の単語 3-gram がなくなってしまうのを防ぐためである．この処理の結果，1099 講演から 333087 発話文を得た．講演単位ごとのテキスト選択についても検討をおこなったが，パープレキシティの減少はみられなかった [17]．

4.3　実験と評価

この処理の流れも図 1 に示す．テスト講演の認識結果から言語モデル（$LM1$）を作成し(注1)，各学習データ（発話文テキスト）のパープレキシティを計算し距離尺度とする．その際，未知語も計算に含めている．パープレキシティがしきい値 th より低い発話文テキストを選択して適応データとし，言語モデル（$LM2$）を作成する．この言語モデルとベースラインの言語モデル（$LM0$）を線形補間することで各話者に適応した言語モデル（$LM4$）を作成する．適応パラメータ th 及び λは，development-set の各講演の最適なパラメータの平均値として推定した．

A1 から A5 を development-set，B1 から B5 を evaluation-set とした場合，パープレキシティしきい値 th 及び線形補間係数λはそれぞれ 110 と 0.472 であった．development-set と evaluation-set を入れ替えた場合は，それぞれ 92 と 0.479 であった．この，適応後のモデル（$LM4$）によるパープレキシティ及び単語誤り率を表 4 に示す．ベースラインに比べてパープレキシティを約 4%（68.18 → 65.60）減らすことができ，誤り率も 0.5%改善された．

適応用に選択された発話文テキストを詳細に分析したところ，『です ね』『で あのー』や『えーま』などのフィラーや文末表現を含むものが多くみつかった．それらは話者間での相違が顕著であった．これは，このような話者ごとの傾向に言語モデルが適応されたことを示している．

(注1)：ただし，ここでは cutoff はおこなっていないので前述の $LM1$ とは少し異なる．

表 5　発音変動を含めた言語モデル適応の結果
Table 5　Results of adaptation including pronunciation

	WER(%)	perplexity
baseline ($LM0$)	33.1	68.18
pronunciation ($LMP0$)	30.8	70.76
pronun + ASR result ($LMP3$)	28.8	53.69
pronun + text selection ($LMP4$)	30.4	67.77
adapted using all methods	28.7	53.20

5.　発音変動のモデル化

5.1　3-gram の枠組みでの発音変動のモデル化

次に，発音変動を適応の枠組みに含めることを検討する．話し言葉では，特につなぎ語や文末表現で発音が大きく変形し，問題となる．従来はこの問題に対して，単語辞書にエントリを追加することで対処をおこなっていた．CSJ では，表記形とその発音形が同時に表記されているため，ある形態素に関してどのような発音が実際におこなわれたかわかる．それぞれの形態素に関して，実際におこなわれた発音の全てを辞書に単純に登録した場合，短い形態素に様々な読みが付与されてしまい，認識時に悪影響を及ぼすことが指摘されている [18]．ベースラインシステムでは，単語辞書登録時に頻度の高くない読みや読み付与誤りの修正をおこなっていたが，このような処理をおこなわなかった場合は，単語誤り率は 39.7%（+6.6%）であった．

このような同じ単語の発音変動をより適切にモデル化するために，表記が同じ形態素でも発音が異なる場合（けれど〔ケレド〕と けれど〔ケード〕など）は言語モデル上で別のエントリとし，同じテキストデータから 3-gram モデルを作成した．なお文献 [19] では，CSJ タスクに対して汎用の形態素解析システム ChaSen を用いて，このような試みがなされているが，本研究では話し言葉である CSJ を用いて学習された形態素解析システムを利用している．

結果を表 5 の上段（$LMP0$）に示す．発音変動を 3-gram でモデル化することで，誤り率を 2.3%改善することができた．これは，発音変動がコンテキストに大きく依存していることを示している．その際，単語辞書上では読みに関する修正をおこなっておらず，頑健に動作していることがわかる．

5.2　発音変動を含めた言語モデル適応

発音変動を考慮した言語モデルに対して，本研究で提案した言語モデル適応手法を適用した．結果を表 5 の中段（$LMP3$, $LMP4$）に示す．発音変動を考慮しない場合と同程度の単語誤り率の改善が得られ，発音変動と話者依存性の両方を同時にモデル化できた．

5.3　提案手法の統合と評価

最後に各提案手法を統合し評価をおこなう．統合手法は図 1

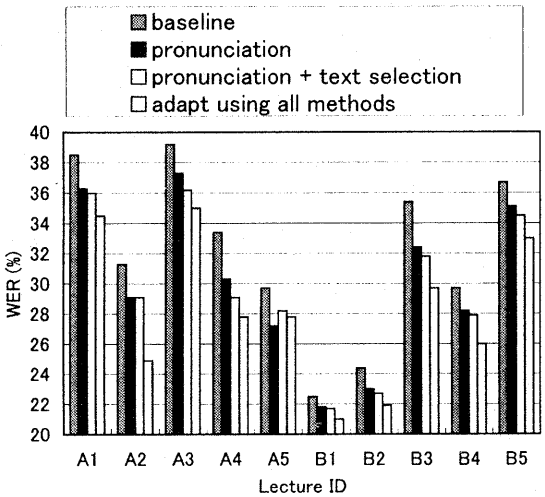

図 3　提案手法による各テストセット講演の単語誤り率（%）
Fig. 3　WER of each test-set lecture with proposed methods

に示す $LM4$ の導出とほぼ同じであるが，ベースラインのモデル（$LM0$）を，認識結果を用いて適応したモデル（$LM3$）で置き換えた上で，$LM4$ を得ている．その際，発音変動もモデル化に組み込んでいる．結果を表 5 の下段に示す．統合効果が確認され，単語誤り率を 33.1%から 28.7%にまで減少させることができた．図 3 に，各テストセット講演ごとの単語誤り率を示す．A5 以外の講演で，単語誤り率を着実に改善でき，最終的に 1%から 6%の単語誤り率の改善が得られた．

6. 言語モデル学習データの追加の効果

2002 年 10 月の時点で利用できる CSJ の書き起こしテキストが増加し，また形態素解析の学習アルゴリズムも改善したので，それらを用いて言語モデルの再構築をおこなった．表 6 に示すように，講演数で 2498 講演，形態素数で約 7.07M（ポーズを除いた場合は約 6.31M）と 2002 年 02 月の時点と比較するとデータ量が大幅に増加している．ただし，A タグ（数字やアルファベット）や？タグ（自信のない箇所）での表記形の選択が以前と異なる．また形態素解析の精度も向上しているため，形態素区切りは完全には一致していない．

言語モデルの作成の方法は，2002 年 2 月時点の言語モデル（ベースラインモデル）と同一である（2.2 章参照）．語彙は学習データに 4 回以上出現した形態素で構成し，サイズは 23192 となった．これによるテストセットのカバレージは 98.05%，これにより作成した言語モデルはパープレキシティは 62.64 であった．今回の音声認識の際には，発話の分割と認識を同時におこなう逐次デコーディング手法 [20] を用いた．

認識結果を表 7 に示す．形態素の単位が異なるため，単語誤り率での公正な比較はできない．そこで，文字誤り率（CER）でも比較をおこなった．新しい言語モデルによる単語誤り率は 29.3%であり，文字誤り率は 25.1%であった（表 7 の lexicon）．

表 6　言語モデル学習データ量の比較
Table 6　Comparison of training data amount of language model

	2002/02	2002/10
#lectures	1099	2498
#morphemes	3.15M	7.07M

表 7　言語モデルの性能の比較
Table 7　Comparison of language models

pronunciation	lexicon		3-gram	
LM	2002/02	2002/10	2002/02	2002/10
WER (%)	32.4	29.3	30.2	28.3
CER (%)	27.4	25.1	25.8	24.4

WER は単語の区切り等が異なるため，正確な比較ではない

2002 年 2 月時点の言語モデルでの文字誤り率は 27.4%であり，文字誤り率で 2.3%改善された．学習データ量の増加と同時に形態素解析の精度及び読み付与の精度も向上しているため，改善の要因は特定できないが，全体として精度のよいモデルが作成できた．

次に，発音変動のモデル化を単語 3-gram の枠組みに含める場合と含めない場合についての比較もおこなった．ここでも，発音変動を単語辞書に単純に追加した場合には，認識誤りが大きく増大した．そのため，ある単語（品詞まで考慮して）に対して，その読みが複数あった場合，発音の生起確率を求めた上でその値が 0.2 以下の発音は単語辞書のエントリから除いた．この結果が表 7 の lexicon の欄に示されている．5 章の結果と同様に，発音変動を考慮した場合（表 7 の 3-gram）の方が単語誤り率が低くなり（29.3% → 28.3%），発音変動のモデル化の効果を確認した．

7. ま と め

言語モデルを各話者の言語表現の傾向やその発音変動に適応させる手法について検討した．適応手法として，認識結果を直接用いる場合，及びテキストを選択しそれを用いる場合の実験を行った．各話者の言語表現の傾向に適応がおこなえ，単語誤り率の改善を得た．発音変動をモデル化し適応の枠組みに組み込むことで，言語表現の傾向と発音変動の両方に同時に適応がおこなえ，さらに単語誤り率の改善が得られた．10 講演平均で 4.4% （33.1% → 28.7%）の単語誤り率の改善を得た．

言語モデルの学習データを増加させて実験を行った場合も同様に，発音変動をモデル化することで認識率の向上が得られた．話者独立モデルで 28.3%の単語誤り率となり，発音変動のモデル化の効果を確認した．

今後は，このモデルを用いて話者の言語表現や発音変動だけでなく，話題への適応についても実験をおこなっていく予定である．

謝辞 本研究は，開放的融合研究『話し言葉工学』プロジェクトの一環としておこなわれた．アドバイスを頂きました東京工業大学の古井貞熙教授をはじめとして，ご協力を頂いた関係各位に感謝いたします．本研究を進めるにあたって貴重な御意見を頂きました，京都大学教授 奥乃博 先生に感謝いたします。

文　献

[1] 前川喜久雄. 言語研究における自発音声. 日本音響学会研究発表会講演論文集, 1-3-10, 春季 2001.

[2] 小磯花絵, 前川喜久雄. 『日本語話し言葉コーパス』の概要と書き起こし基準について. 情報処理学会研究報告, 2001-SLP-36-1, 2001.

[3] 篠崎隆宏, 古井貞熙. 日本語話し言葉コーパスを用いた講演音声認識. 情報処理学会論文誌, Vol. 43, No. 7, pp. 2098–2107, 2002.

[4] 南條浩輝, 河原達也. 発話速度に依存したデコーディングと音響モデルの適応. 電子情報通信学会技術研究報告, SP2001-103, NLC2001-68 (SLP-39-20), 2001.

[5] 緒方淳, 有木康雄. 音素事後確率に基づく信頼度を用いた音響モデルの教師なし適応化. 電子情報通信学会技術研究報告, SP2001-105, NLC2001-70 (SLP-39-22), 2001.

[6] 伊藤彰則, 好田正紀. N-gram 出現回数の混合によるタスク適応の性能解析. 電子情報通信学会論文誌, Vol. J83-DII, No. 11, pp. 2418–2427, 2000.

[7] 小林彰夫, 今井亨, 安藤彰男, 中林克己. ニュース音声認識のための時期依存言語モデル. 情報処理学会論文誌, Vol. 40, No. 4, pp. 1421–1429, 1999.

[8] M.Mahajan, D.Beeferman, and X.D.Huang. Improved Topic-Dependent Language Modeling using Information Retrieval Techniques. In *IEEE Int'l Conf. Acoust., Speech & Signal Process.*, Vol. 1, pp. 541–544, 1999.

[9] 小椋秀樹. 話し言葉コーパスの単位認定基準について. 話し言葉の科学と工学ワークショップ講演予稿集, pp. 21–28, Feb. 2001.

[10] 内元清貴, 井佐原均. 話し言葉コーパスの形態素解析. 話し言葉の科学と工学ワークショップ講演予稿集, pp. 33–38, Feb. 2002.

[11] 南條浩輝, 河原達也. 講演音声認識のための種々の形態素解析及び音響モデルの評価. 話し言葉の科学と工学ワークショップ講演予稿集, pp. 47–52, Feb. 2002.

[12] 李晃伸, 河原達也, 武田一哉, 鹿野清宏. Phonetic Tied-Mixture モデルを用いた大語彙連続音声認識. 電子情報通信学会論文誌, Vol. J83-DII, No. 12, pp. 2517–2525, 2000.

[13] 横山忠介, 篠崎隆宏, 古井貞熙. 講演音声を対象とした言語モデルの話者適応化. 日本音響学会研究発表会講演論文集, 3-9-6, 秋季 2002.

[14] 長友健太郎, 西村竜一, 小松久美子, 黒田由香, 李晃伸, 猿渡洋, 鹿野清宏:. 相補的バックオフを用いた言語モデル融合ツールの構築. 情報処理学会論文誌, Vol. 43, No. 9, pp. 2884–2893, 2002.

[15] 加藤一臣, 南條浩輝, 河原達也. 講演音声認識のための音響・言語モデルの検討. 電子情報通信学会技術研究報告, SP2000-97, NLC2000-49 (2000-SLP-34-23), 2000.

[16] T.Niesler and D.Willett. Unsupervised Language Model Adaptation for Lecture Speech Transcription. In *Proc. Int'l Conf. Spoken Language Processing (ICSLP)*, 2002.

[17] 南條浩輝, 河原達也. 講演音声認識における言語モデル適応の検討. 日本音響学会研究発表会講演論文集, 3-Q-16, 秋季 2002.

[18] M Ostendorf. Moving beyond the 'beads-on-a-string' model of speech. *In Proc. IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, 1999.

[19] 堤怜介, 加藤正治, 好田正紀. 講演音声認識のための音響・言語モデルの検討. 日本音響学会研究発表会講演論文集, 2-9-16, 秋季 2002.

[20] 河原達也, 加藤一臣, 南條浩輝, 李晃伸. 話し言葉音声認識のための言語モデルとデコーダの改善. 情報処理学会研究報告, 2001-SLP-36-3, 2001.