様式 C-19

科学研究費補助金研究成果報告書

平成 21 年 5 月 27 日現在

研究種目：若手研究(B)
研究期間：2006～2008
課題番号：18700177
研究課題名（和文）　ドメイン独立な話し言葉のモデル化に基づく音声認識の研究

研究課題名（英文）　Study on automatic speech recognition based on domain-independent modeling of spontaneous speech

研究代表者
秋田　祐哉（AKITA YUYA）
京都大学・学術情報メディアセンター・助教
研究者番号：90402742

研究成果の概要：会議などの話し言葉を対象とした音声認識では，認識用のモデルの構築に必要な話し言葉データの大規模収集が難しい.そこで,話し言葉の特徴を統計的にモデル化して，これに基づき言語モデルや発音モデルを話し言葉スタイルに変換する手法を検討した．本手法では話し言葉スタイルと正書体的なスタイルの表現・発音の比較を行い，確率的変換モデルを構築する．実際の音声における評価で，このモデルは言語予測性能や音声認識の精度を有意に改善することができた.

交付額

（金額単位：円）

	直接経費	間接経費	合　計
2006 年度	1,100,000	0	1,100,000
2007 年度	1,400,000	0	1,400,000
2008 年度	1,000,000	300,000	1,300,000
年度			
年度			
総　計	3,500,000	300,000	3,800,000

研究分野：音声情報処理
科研費の分科・細目：情報学　知覚情報処理・知能ロボティクス
キーワード：音声認識，話し言葉，言語モデル，発音モデル，スタイル変換

１．研究開始当初の背景

近年の音声認識の主要な研究対象として，講義・講演・会議・討論などの分野（ドメイン）における人間同士の音声コミュニケーション，いわゆる「話し言葉」音声がある．話し言葉音声認識は，書き起こしや字幕の自動作成にとどまらず，シーンの分割・分類や索引付け，要約作成といった高度な知的処理を実現するための基盤ともなるものである．これらの研究をいっそう活性化するためにも，さまざまなドメインの話し言葉音声認識の実現が期待されており，研究代表者も以前から会議や討論の音声認識に取り組んでいる．

話し言葉音声には，書き言葉と比べて多様な言語表現や非標準的な発音などの特有の特徴がみられる．したがって音声認識を実現するためには，ドメインに依存する話題などの特徴とともに，このような話し言葉の特徴もモデル化する必要がある．一般的な手法では，両特徴を区別せずに，認識対象とドメイ

ンが一致した話し言葉データを用いてモデル化を行う．しかし話し言葉データの収集は容易ではなく，話し言葉音声認識の研究は大規模なデータを整備できた特定のドメインに実質的に限定されてきた．音声認識の対象を広げるために，既存のドメインの異なる話し言葉データからでも話し言葉の特徴のモデル化が可能な，新たな方法論が求められている．

２．研究の目的

このような背景を踏まえ，本研究ではドメイン依存の特徴のモデル化と話し言葉の特徴のモデル化を分離し，ドメインに依存しない話し言葉のモデル化の可能性を追究する．具体的には以下の３つの課題について研究を実施する．

(1) 言語表現と発音を対象とした，標準的（正書体的）な場合と話し言葉の場合の特徴の比較・分析

(2) 音声認識における変換モデルの確立と，(1)に基づく実際のモデルの構築

(3) (2)のモデルの，さまざまなドメインにおける音声認識での有効性の検証

３．研究の方法

(1) ドメインごとの話し言葉特徴のモデル化
講演や討論などのいくつかのドメインにおいて，それぞれ話し言葉の言語表現および発音の統計的分析を行う．また，音声認識に組み込み可能な話し言葉変換モデルの枠組みについて，理論上および実装上の面から検討する．発音の変換については研究代表者のこれまでの研究で部分的に扱われているが，本研究ではより精緻なモデルの実現を目指す．

(2) 話し言葉モデルを用いた音声認識
(1)による話し言葉モデルについて，他のドメインへの適用に先だって，同一ドメインでの有効性を検証する．モデルの構築に用いたデータと同一のドメインの開発データを用意し，これに対する音声認識の性能評価と結果の分析を(1)にフィードバックし，モデルの枠組みの改良や洗練に役立てる．

(3) 異なるドメインへの話し言葉モデルの適用と分析
話し言葉モデルを，構築に用いたデータのものとは異なるドメインの音声認識に適用する．これにより，異なるドメインに適用する場合のモデルの有効性と問題点を明らかにする．

４．研究成果

(1) ドメインごとの話し言葉特徴のモデル化
国会会議や学会講演のドメインにおいて，話し言葉とその整形テキストを対応付けたデータベースをもとに統計的分析を行い，話し言葉に特徴的な表現とその発生条件（文脈）・頻度などを分析した．

表 1 は，総単語数 737K の国会データにおける，話し言葉に特徴的な表現とその頻度である．また，日本語話し言葉コーパス（CSJ）の学会講演（総単語数 496K）における同様の分析結果を表 2 に示す．多くのパターンは共通するものの，一方のみに特に多く見られるパターンや，頻度の傾向が異なるなどの違いも見られる．整形の基準が異なることもあるが，この結果は発話スタイルの違いを示唆するものともいえる．

このような統計的分析に基づいて音声認識用言語モデルを変換する話し言葉変換モデルについて，バックオフ的手法や補間手法，最大エントロピー法といった複数の統計的枠組みの検討を行った．国会音声における，言語モデルの予測能力の指標であるパープレキシティによる評価ではバックオフ的手法が最も高い性能を示し（図 1），これに基づいて本研究ではバックオフ的手法を枠組みとして採用した．

表 1:話し言葉に特徴的な表現と発生頻度
（国会会議における分析）

挿入		脱落		置換			
えー	5,623	を	586	てる	→	て いる	2,209
ですね	2,728	は	535	ていう	→	という	406
おー	2,434	が	233	やっぱり	→	やはり	246
あの	2,097	に	150	ん	→	の	240
あー	2,008	と	120	けども	→	けれども	235
まあ	1,953	いる	52	いろんな	→	いろいろ な	181
あのー	1,438	よう	40	けど も	→	けれども	163
その	1,171	です	34	けど	→	けれども	161
で	1,078	の	30	て	→	という	138
と	1,063	か	29	もん	→	もの	106

表 2:話し言葉に特徴的な表現と発生頻度
（学会講演における分析）

置換		削除		挿入	
ん→の	2,204	えー	5,601	い	606
って→と	1,673	で	1,409	を	475
けど→けれど	835	あの	1,327	が	199
てる→ている	726	ま	1,317	です	194
が→しかし	216	な ん	610	は	164
じゃ→では	155	こう	531	に	113
た ん→ました の	154	ですね	519	しかし	45

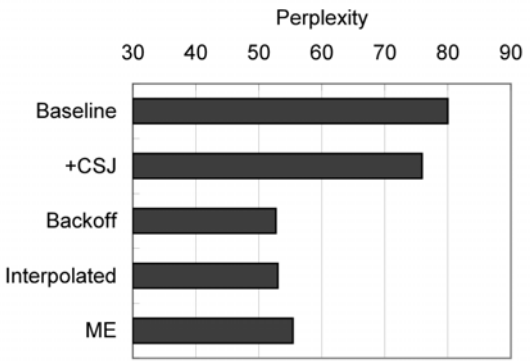

図 1:モデル化手法の比較

(2) 話し言葉変換モデルを用いた音声認識

(1)による話し言葉変換モデルについて，音声認識に対する有効性を検証した．さまざまな会議（すなわち話題）を含む国会音声の認識における，コーパス混合に基づく従来法（2 種類）と提案法による言語モデルのパープレキシティおよび単語誤り率（WER）を図 2 と図 3 にそれぞれ示す．

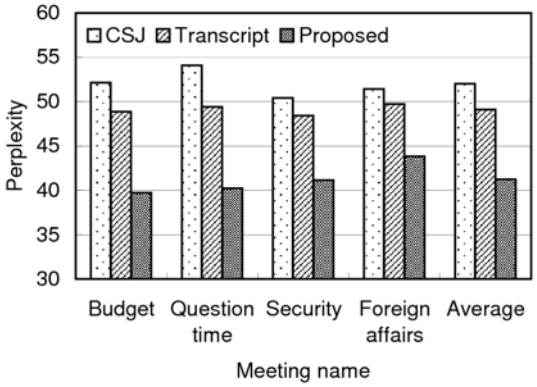

図 2:各言語モデルによる会議ごとのパープレキシティ

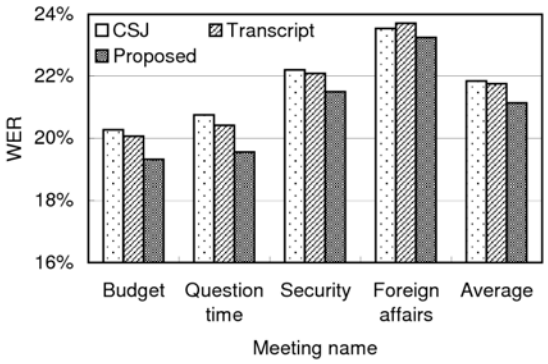

図 3: 各言語モデルによる会議ごとの単語誤り率

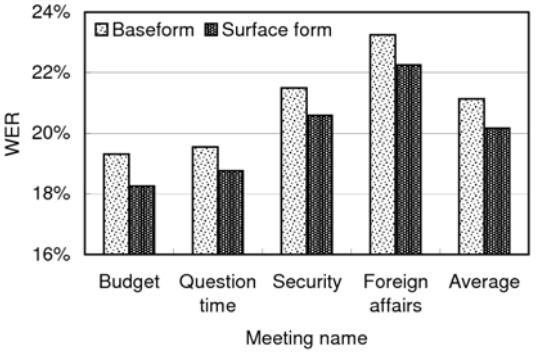

図 4:発音モデル変換の効果

これらの図より，提案法（Proposed）は従来法（CSJ/Transcript）に対して有意に性能を改善していることがわかる．また，従来の言語モデル構築法では少量データでは十分に対処することができない，話題の異なる（国会会議）音声においても言語的予測能力が大きく改善された．以上より提案法の有効性が示されたといえる．

(3) 発音の変換モデルを用いた音声認識

発音モデルの変換についても同様に評価を行った．発音モデルの変換モデルは CSJ の学会講演を用いて学習しており，国会音声とは異なるドメインである．変換前の標準的な発音（Baseform）と変換による話し言葉の発音（Surface form）のそれぞれによる単語誤り率を図 4 に示す．発音モデルの変換により性能が有意に改善され，異なるドメインであっても有効に機能することが実証された．

(4) 今後の展望

本研究の期間中における継続的な改良と，異なるデータによる評価を通じて，提案法が話し言葉音声認識において安定的に機能することを実証した．研究代表者らは，本研究の成果に基づき国会（衆議院）会議録の作成のための音声認識システムを構築するなど，本研究の実用化に取り組んでいるところである．

今後は，本研究の話し言葉モデリングと関連が深いと考えられる，話し言葉の自動整形などへの活用が期待される．

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計２件）

(1) Yuya Akita and Tatsuya Kawahara. Topic-independent Speaking-style Transformation of Language Model for Spontaneous Speech Recognition. Proc. IEEE Int'l Conf. Acoustics, Speech and Signal Processing (ICASSP), Vol.4, pp.33-36, 2007. （査読あり）

(2) Yuya Akita and Tatsuya Kawahara. Efficient Estimation of Language Model Statistics of Spontaneous Speech via Statistical Transformation Model. Proc. IEEE Int'l Conf. Acoustics, Speech and Signal Processing (ICASSP), Vol.1, pp.1049-1052, 2006. （査読あり）

〔学会発表〕（計６件）

(1) 秋田祐哉，三村正人，河原達也.会議録作成支援のための国会審議の音声認識システム.日本音響学会春季研究発表会，2009 年 3 月 19 日，東京都目黒区（東京工業大学).

(2) 秋田祐哉，三村正人，河原達也.会議録作成支援のための国会審議の音声認識システム.電子情報通信学会音声研究会，2008 年 12 月 10 日，東京都新宿区（早稲田大学).

(3) 秋田祐哉，河原達也.会議録作成のための話し言葉音声認識結果の自動整形.日本音響学会秋季研究発表会，2008 年 9 月 12 日，福岡県福岡市（九州大学).

(4) 秋田祐哉，河原達也.話し言葉スタイルへの統計的変換法の CSJ への適用.日本音響学会春季研究発表会,2008 年 3 月 19 日，千葉県習志野市（千葉工業大学).

(5) 秋田祐哉，河原達也.言語モデルと発音辞書の統計的話し言葉変換に基づく国会音声認識.電子情報通信学会音声研究会，2007 年 12 月 20 日,京都府精華町(NTT).

(6) 秋田祐哉，河原達也．言語モデルの話し言葉変換法の音声認識における評価．日本音響学会秋季研究発表会，2006 年 9 月 14 日，石川県金沢市（金沢大学).

６．研究組織
(1)研究代表者
秋田　祐哉（AKITA YUYA）
京都大学・学術情報メディアセンター・助教
研究者番号：90402742