

## WEB上の類似記事を利用した音声文書の認識性能の改善

伊藤 友裕<sup>†</sup> 西崎 博光<sup>‡</sup> 関口 芳廣<sup>‡</sup>

<sup>†</sup>山梨大学大学院医学工学総合教育部 〒400-8511 山梨県甲府市武田 4-3-11

<sup>‡</sup>山梨大学大学院医学工学総合研究部 〒400-8511 山梨県甲府市武田 4-3-11

E-mail: <sup>†</sup> yuu@alps.cs.yamanashi.ac.jp, <sup>‡</sup> {nisizaki,sekiguti}@ccn.yamanashi.ac.jp

あらまし 本稿では、音声文書をインデキシング目的で音声認識する際に、WEB上の類似ニュース記事を利用し、それを汎用的な言語モデルの話題適応化に用いることにより言語モデルや認識辞書を構築し、未知語に強固な音声認識を行う手法を提案する。また、タイプの異なる3種類の言語モデルを用い、その出力結果を多数決により組み合わせることで、最終的な音声認識性能の改善を図っている。音声認識システムでは、1) 新聞記事75ヶ月分より作成した語彙サイズ2万単語のモデル・辞書のみ、2) そのモデルと類似記事群から学習したモデルを融合させた言語モデル・辞書を用いた場合、3) 固有名詞をクラスとして学習した固有名詞クラス言語モデルを用いた場合、の3種類の言語モデルを使用する。それら3種類のそれぞれの認識結果から、多数決法により最終的な認識結果を導き出す。提案手法により、言語モデル・単語辞書に新聞から学習したモデルのみを用いた場合に比べ、単語正解率が47.0%から47.8%へ、単語正解精度は37.7%から39.5%へ、中でも特に名詞の正解率は44.7%から、46.4%へ、固有名詞の正解率は、43.9%から56.1%へと大幅な精度の向上がみられた。

キーワード 音声文書, 音声認識, WEB文書, 誤り訂正, 言語モデル適応

## Improving recognition performance of spoken documents using similar documents on the Internet

Yuusuke Itoh<sup>†</sup> Hiromitsu Nishizaki<sup>‡</sup> Yoshihiro Sekiguchi<sup>‡</sup>

<sup>†</sup> Graduate School of Medical and Engineering Science Department of Education, University of Yamanashi,  
4-3-11 Takeda, Kofu, Yamanashi 400-8511 JAPAN

<sup>‡</sup> Graduate School of Medicine and Engineering Science Department of Research, University of Yamanashi,  
4-3-11 Takeda, Kofu, Yamanashi 400-8511 JAPAN

E-mail: <sup>†</sup> yuu@alps.cs.yamanashi.ac.jp, <sup>‡</sup> {nisizaki,sekiguti}@ccn.yamanashi.ac.jp

**Abstract** This paper describes a technique of improving recognition performance of spoken documents by using a language model adaptation from similar documents on the Internet and combining various transcriptions from LVCSR systems. A language model and a dictionary made from the similar documents which may be relative to the spoken document give improvement of Out-of-vocabulary rate in the dictionary. We used three kinds of language models in a LVCSR system as follows: (1) a general 20K language model from newspaper articles (75 months), (2) a topic adapted language model using the similar WEB documents, (3) a class-based language model in which only proper nouns are classed. Three kinds of outputs from the LVCSR systems, where those three language models are used respectively, are combined by using a simple voting scheme. In an experimental result, the proposed method has improved the recognition performances comparing with the case of using the language model from the newspaper articles only. The word correct and accuracy rates were improved to 47.8% from 47.0%, to 39.5% from 37.7%, respectively. Especially, the correct rate in so far as proper nouns was dramatically gotten improved to 56.1% from 43.9%. These result showed that our technique was effective to transcribing the news documents automatically.

**Keyword** Spoken document, speech recognition, WEB document, error correction, language model adaptation

## 1. はじめに

近年、情報通信システムの発展に伴い、動画データの送受信が容易に行なわれるような環境が整いつつある。大量の電子化されたマルチメディアデータを誰もが容易に発信したり蓄えたりすることが可能になってきているが、あまりにも大量のデータであるため、それらを整理することが必要である。そのような背景下で、大量のマルチメディアデータから必要な情報を取り出すための検索技術の開発が盛んに行われている。一般的にマルチメディアデータの検索を行うためには、それらのデータに対して何らかのインデキシングを行わなければならない。情報検索等で利用するために MPEG7 等でマルチメディアデータへ付与するためのメタデータの規格化が検討されているが、大量のデータに対して人手でメタデータを付与することは困難である。そこでビデオ映像などを検索するためのメタデータを自動的に付与できれば、あまり人手などの労力が掛からずにインデックスを作成できる。

マルチメディアデータ、特に音声文書を検索するために必要な自動インデキシングを行う方法として、音声認識結果を利用する事が考えられる。しかし、自動でインデックスを付与するために音声認識技術を用いると、未知語（音声認識辞書に含まれない単語）や誤認識の問題に直面する。名詞、特に固有名詞など文書の特異性を表す重要単語は未知語になりやすく、その単語のみが認識誤りを起こすのではなく、その前後の単語にも影響を与えるため、未知語の含まれた音声文書は、音声認識での正確な書き起こしが難しいとされており、特にニュースなどの時事文書を扱う場合はなおさらである。

そこで、本稿ではニュース映像を対象とし、それらの自動インデキシング化のために、WEB上の類似ニュース記事を用いて、言語モデルや単語辞書を構築し、未知語に強固な音声認識を行う手法を提案する。音声認識の認識率向上のための研究はいくつか行なわれている。Ivanら[1]は、言語モデルを一般的な対話集から学習した言語モデルと、WEBからフィルター処理をして得た情報を適応化処理する事により、単語誤り率の減少を実現している。尾上ら[2]は、ニュース音声の認識誤りを訂正する際に、アナウンサーが読み上げる原稿には類似した記者原稿があることに着目し、DPマッチングを用いてアナウンサーが読んだニュースに内容が近い記者原稿を探し出し、1記事の認識結果全体を記者原稿に置き換えるという方法で訂正を行っている。また、松井ら[3]は、スポーツニュースを対象として、スポーツニュースに特化した言語モデル・音響モデルを用いた認識率の改善方法を提案している。Ringgerら[4]は、音声認識結果の間違ったデータを用い、そ

から認識エンジンの誤り傾向を導き出し、誤認識区間を訂正している。石川ら[5]の方法では、まず、音声認識結果の必要性を判断するために、認識結果を構文解析し、意味的距離を用いて構文木単位での訂正候補を決定し、テキストコーパスから音韻的な特性が類似している用例を検索し、置換を行う。置換処理の妥当性は意味的距離を用いて判断している。沖本ら[6]は、認識誤りは分かっていると仮定した上で、認識タスクと同じ内容の文章集合から用例を単語レベルによる DP マッチングにより検索し、文単位での置換処理を行っている。山口ら[7]は、誤り訂正個所の判別に N-gram と機械学習の手法を用い、その効果を示している。更に、伊藤ら[8]は、音声クエリー中の未知語を自動検出し、未知語検出を誤認識検出とみなせば、誤り訂正と同じということ、検索された文書中から、検出した未知語区間の音節列と類似している単語を探し出し置換するという方法を行なっている。これらの文献で提案されている認識率の改善手法で共通して言えることとしては、認識する音声と全く同じ、もしくは類似したタスクのコーパスを手動で用意し、利用していることである。こういったコーパスは言語モデルの学習に用いられる事が多いが、言語モデルの学習にはコーパスの学習量が多くなければニュース音声の様な発話には対応できない。

そこで本稿で提案する音声認識手法では、言語モデルの学習コーパスとして、WEB上に存在するニュース記事から自動抽出した類似記事群と毎日新聞の記事データ 75ヶ月分（1991年1月～1994年9月、1995年1月～1997年6月）より作成した語彙サイズ2万単語のモデルとを融合して用いることを検討している。音声データの類似記事群と75ヶ月分の新聞記事データとを融合し用いる利点としては、学習コーパスとして類似記事群を用いることにより、それらの類似記事にはその音声データ特有の名詞、特に固有名詞が含まれている可能性が高いということ、更に75ヶ月分の新聞記事データも利用する事により、汎用的な言葉にも対応できるため、音声データの未知語の軽減や単語のカバー率の向上といったことが挙げられる。

また、音声認識システムの言語モデルや単語辞書には、1) 新聞記事75ヶ月分より作成した語彙サイズ2万単語のモデル・辞書のみ、2) そのモデルと類似記事群から学習したモデルを適応化させた言語モデル・辞書を用いた場合、3) 固有名詞をクラスとして学習した固有名詞クラス言語モデルを用いた場合、の3種類の言語モデルを使用する。それら3種類のそれぞれの認識結果から、多数決法による最終的な認識結果を導き出す。

本稿で報告する実験では、特にニュース動画(音声)

に焦点を当てて実験を行ったが、WEB 文書より信頼できる限定されたタスクを自動抽出する事により、講演音声や国会答弁の音声<sup>1</sup>などへの技術の適用は可能であると考えており、このようなデータに対しても研究を進めている。今回はニュース映像に対する結果を報告する。

提案手法により、言語モデル・単語辞書に新聞から学習したモデルのみを用いた場合に比べ、単語正解率が 47.0%から 47.8%へ、単語正解精度は 37.7%から 39.5%へ、中でも名詞の正解率は 44.7%から、46.4%へ。固有名詞の正解率は、43.9%から 56.1%へと精度の向上がみられた。

## 2. システムの概要

### 2.1. 処理の概要

提案手法の処理の流れの概要図を図 1 に示す。

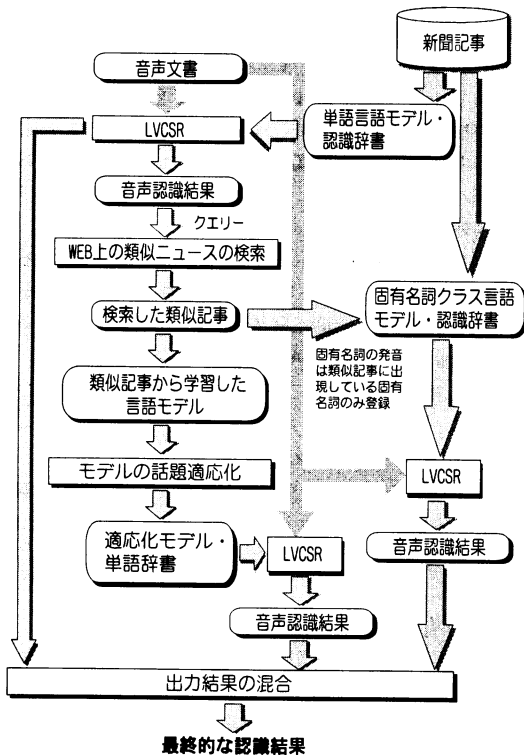


図 1 システムの処理の流れの概要

本提案手法では、まず、新聞記事から普通に学習した 2 万語彙の言語モデルと辞書(これを以下“新聞 LM”と記す)を用い、認識対象となる音声文書の音声認識

<sup>1</sup> これらには未知語になりやすい専門用語が多く含まれている。

結果を得る<sup>2</sup>。次に、その音声認識結果をクエリーとして用いることで、WEB 上のニュース記事集合から、認識対象の音声文書と類似している記事の検索・抽出を行う。得られた記事集合は、対象音声の内容に類似していると考えられるため、これから認識対象文書に特化した言語モデルを学習する(これを以下“類似 LM”と記す)。本研究では、この類似 LM をそのまま音声認識に用いるのではなく、新聞 LM と適応化したモデル(以下、これを“話題適応化 LM”と記す)を用いる<sup>3</sup>。線形補間を行う際の各モデルに対する重みは、回帰分析により学習データより最適値を求めた。

また、新聞 LM の学習の際に用いた学習コーパスに含まれる固有名詞を 9 種類に分類・ラベル付けし<sup>4</sup>、それから固有名詞のみをクラス化したクラス言語モデル(これを以下“固有名称クラス LM”と記す)を構築しておく。固有名詞クラス LM に対応する認識辞書を作成する際、固有名詞に対する発音表記に関して、WEB から得た類似記事集合に含まれている固有名詞のみの発音を辞書に登録することにしている<sup>5</sup>。

これら各言語モデルによって得られた 3 つの音声認識結果(音声認識結果 1、音声認識結果 2、音声認識結果 3)を DP マッチングにより単語ごとのアライメントをとり、多数決法により、最終的な認識結果を決定する。

以下提案手法の詳細について述べる。処理は、

- (1) WEB 上の類似記事の検出
- (2) 2 つの言語モデルの適応化処理
- (3) 固有名称クラス LM の構築
- (4) 各認識結果の多数決による統合の 4 つに大別される。

### 2.2. WEB 上の類似記事検出処理

WEB 上の類似記事を検索するためのクエリーには、対象音声文書の音声認識結果を用いる。この書き起こしに用いる音声認識システムとして、Julius[9]を用いる。

新聞 LM の学習には、毎日新聞の記事データ 75 ヶ月分(1991 年 1 月～1994 年 9 月、1995 年 1 月～1997 年 6 月)より作成の 2 万語彙の trigram を用いる。この言語モデルを用いた認識結果(音声認識結果 1)から、名詞のみを抽出し、それをクエリーとしインターネット上のニュースサイトから類似記事を検索する。なお、

<sup>2</sup> 以下、これを“音声認識結果 1”とする

<sup>3</sup> 話題適応化 LM を用いた認識結果を“音声認識結果 2”と表記する

<sup>4</sup> ChaSen の品詞体系による

<sup>5</sup> クラス言語モデルと認識辞書を用いて音声認識した結果を、“音声認識結果 3”とする

記事の究極の要約が見出しと考えられるため、音声認識結果 1 と WEB から検索したニュース記事の見出しの間で類似度の判断を行う。類似度の計算には、以下に示す式(1)を用いている。

$$sim(wl, wr) = \frac{\sum_t f(t_{wl})f(t_{wr})}{\sqrt{\sum_t f(t_{wl})^2} \sqrt{\sum_t f(t_{wr})^2}} \quad (1)$$

本研究では、 $wl$  は新聞 LM を用いた認識結果であり、 $wr$  はニュース記事の見出し、 $t$  は単語としている。つまり、 $f(t_{wl})$  は新聞 LM 中の単語  $t$  の出現数である。

$sim(wl, wr)$  が設定閾値以上の記事を類似記事としている。

また、得られた類似記事から学習した言語モデル(類似 LM)と新聞 LM を線形補間する際の最適な混合比の算出に用いるために、類似記事の記事数、語彙数および音声認識結果 1 の単語数を保存しておく。

## 2.3. 話題適応化の処理

### 2.3.1. 話題適応化 LM

話題適応化の処理は、さまざまな異なるトピックを持つ N-gram 言語モデルを任意に組み合わせることで言語モデルのタスク操作を実現する事である。

即ち、汎用性のある言語モデルとタスクの限定された言語モデルを組み合わせることによって、タスクが限定された話題の音声の認識率を向上させる。

最も単純なモデルの混合手法は、各々の学習元コーパスをつなぎ合わせて再学習する方法(コーパス結合)であるが、このコーパス結合は、巨大なコーパスを保持し、それをもとにモデルの再学習をする必要があるため、利便性に欠ける。よって、本研究ではより扱い易い方法としてよく用いられている手法として、各モデルにおける N-gram 確率を重み付きで結合する方法を採用した[10]。この方法は、特に少量のコーパスによる学習データの不足を補ったり、言語モデルのタスク適応などに用いられている方法である。

### 2.3.2. 話題適応化 LM を用いた音声認識

コサイン類似度により類似記事と判断されたニュース記事集合を用いて、類似 LM と類似記事単語辞書を作成する。また、類似記事全体の語彙の種類数、類似記事の記事数、また、音声認識結果 1 の総単語数といった情報を用いて、最適混合比の算出を行う。

混合比は、既知のデータを解析して実験的に求めら

れた以下の式を用いる。

$$\text{融合比(新聞 LM の割合)} = -(0.0016 * (\text{類似 LM の語彙数})) + (0.0015 * (\text{認識結果 1 の単語数})) + (0.0275 * (\text{類似 LM にしたニュース記事の記事数})) + 4.4335$$

この式から求めた最適混合比を用いて話題適応化 LM と辞書を作成する。

## 2.4. 固有名詞のクラスタリング処理

### 2.4.1. クラスタリング手法

発話文内に未知語が存在した場合、その単語のみが認識誤りを起こすだけではなく、その前後の単語も認識誤りを起こすことがある。

そのため、タスクに近いコーパスから学習した言語モデル等を利用することにより、未知語を出来るだけ減らす事を試みている。しかし、音声認識システムでは統計的言語モデルを用いている性質上、タスクが限定されている言語モデルの重みを大きくすることにより、隣接する単語の接続に柔軟性がなくなってしまうことが考えられる。新聞 LM のような汎用性のある言語モデルを用いた方が、固有名詞などの特異性を示す単語以外の認識率は高い場合も十分に想定できる。

そのため、WEB 上の類似記事ニュースを用いる利点である固有名詞のカバー率(限定されたタスクから取得)と、汎用的な新聞 LM を用いる場合の利点である一般的な単語接続や言語モデルの整合性の高さといった長所を組み合わせることで、さらに音声認識率の改善を図る。

### 2.4.2. 固有名詞クラス LM

新聞 LM の学習コーパスに含まれている固有名詞を 9 種類に分類・クラス化し、同じコーパス(新聞記事)からそれら 9 種類の固有名詞のクラス LM を作成する。

### 2.4.3. 固有名詞用辞書の作成

固有名詞クラス LM 用の辞書には、通常は新聞データに含まれる固有名詞すべての発音を辞書へ登録するが、今回の実験では、WEB から得た類似記事集合に含まれている固有名詞のみの発音を辞書に登録する。これは、認識対象の音声データに含まれる固有名詞のほぼすべてが類似記事に含まれているという前提の下、誤った音声認識結果が出現しないようにするための措置である。具体的な処理手順は、新聞 LM の辞書中の固有名詞の発音表記を全て削除し、その代わりに WEB 類似記事中に含まれている固有名詞の発音表記を追加する。これにより、誤った固有名詞に認識されることを防ぐ。

## 2.5. 多数決法による音声認識結果の決定

前述のように、3種類の言語モデルを用いた音声認識にはそれぞれ一長一短があるため、どのモデルの認識精度が格段に良くなるということはない。しかし、それぞれのモデルにおいて、単語の種類に対する得意、不得意が存在するため、これらを上手く組み合わせることで、さらなる認識率の改善を図る。

複数の音声認識システムを組み合わせる方法としては、ROVER法[11]や宇津呂らの機械学習を用いた手法[12]などがあるが、今回は単純なルールを適用した多数決法を用いることにする。

各言語モデルの出力結果に対して、単語単位でDPマッチングを行い、出力結果同士で単語のアライメントをとる。アライメントの基準には、固有名詞クラスLMを用いた認識器からの出力を用いる。これは、未知語が少なく、さらに汎用性のある新聞データから学習している言語モデルであるため無難な音声認識結果が期待できるためである。各セグメントに対して3つ(単語の種類としては1~3種類となる)の正解候補があり、この中で2票以上の得票が最も信頼できる認識単語であるとして選択する(規則①)。3つの認識結果がいずれも異なる単語に認識された場合は、各モデルが得意とする品詞の単語を正解とみなすことにする。

詳しく説明すると、固有名詞の認識は、話題適応化LMを用いた音声認識結果が優れているため、3つの認識結果がいずれも異なる単語に認識されている場合は、このモデルを用いた出力単語を認識結果(固有名詞)として正解とみなす(規則②)。また、新聞LMを用いた認識結果が、文法的(動詞・助詞・助動詞)に最も適切であると考えられるため、3つの認識結果がいずれも異なる場合、かつ話題適応化LMの出力結果が固有名詞意外である場合は、この言語モデルから得られた単語を正解とみなすことにする(規則③)。例を図2に示す。

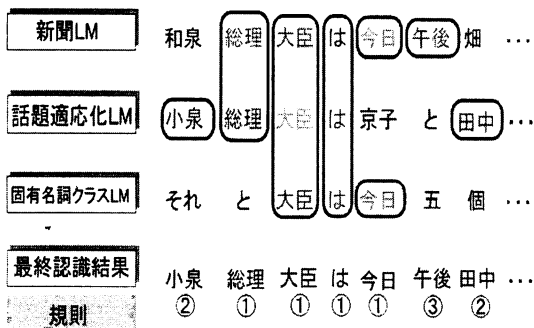


図2 多数決法を使った認識結果の決定例

## 3. 実験

### 3.1. 実験条件

音声認識システムとして、Julius ver.3.4.2を用いた。音響モデルは、5状態3ループ、性別依存(男性)、対角共分散、16kHzサンプリング、25msハミング窓、フレーム周期10ms、音素 triphone モデル(16混合、総状態数3000)、特徴ベクトル MFCC (12次元) + Δ MFCC + Δ POW (計25次元)である。

言語モデルとしては、新聞LMとして、毎日新聞の記事データ75ヶ月分(1991年1月~1994年9月、1995年1月~1997年6月)より作成の2万単語のtrigramを用いる。

今回の実験において、WEB上のニュース記事にはGOOのニュース記事速報を用いる<sup>6</sup>。

### 3.2. 類似記事検出実験

#### 3.2.1. ニュース記事数と実験データ

本実験で使用した音声データはNHK(日本放送協会)の「ニュース7」2004年、6月9日の「バス転落でけがの1人重体 福島県警」というトピックの収録音声である。発話数は7文、単語数は354単語、約2分の音声である。また、6月9日のニュース記事は660記事である。

#### 3.2.2. 実験結果

コサイン尺度による類似度計算の結果より求められた類似記事の抽出結果を表1に示す。( )内は、その記事中の正解の類似記事の数である<sup>7</sup>。

表1 類似記事検出実験結果

閾値	記事数(正解数)
無し(全記事)	660 (5) 個
0.1以上	26 (5) 個
0.2以上	5 (5) 個

コサイン尺度によって、認識結果と類似記事の見出しの類似度を算出して、類似記事を抽出した場合、閾値0.1の場合は、660記事を26記事に減らす事が出来て、閾値0.2の場合は、5記事まで類似記事を減らす事が可能であった。中でも、類似記事5記事全ての記事を含む事が出来、正解を保ちつつある程度記事数を減らす事が可能である。即ち、類似記事を精度良く抽出する事が出来る事が分かった。

類似記事が多すぎると処理の効率性が下がるが、言語モデルの融合のためには、ある程度多くの類似記事を含んでいる場合の方が適している。そこで類似記事抽出の際の閾値は0.1を用いるのが妥当であると考えられる。

<sup>6</sup> <http://news.goo.ne.jp>

<sup>7</sup> 正解類似記事は人間が妥当であると判断。

### 3.3. 音声認識精度比較実験

#### 3.3.1. 評価用音声データ

NHK ニュース「ニュース7」より2004年6月9日の10個の話題の収録音声データを評価に用いた。

なお、6月9日のニュース記事は660記事である。

#### 3.3.2. 実験結果

音声認識結果の単語正解率・正解精度を表2, 3に示す。

表2 全単語の音声認識率  
(Corr.:正解率, Acc.:正解精度)

	新聞 LM	話題適応化 LM	固有名詞クラス LM	多数決法
Corr.	47.0% (1462/3112)	41.0% (1274/3112)	45.9% (1429/3112)	47.8% (1488/3112)
Acc.	37.7% (1173/3112)	33.5% (1043/3112)	36.7% (1141/3112)	39.5% (1230/3112)

表3 名詞・固有名詞の音声認識率

	新聞 LM	話題適応化 LM	固有名詞クラス LM	多数決法
名詞正解率	44.7% (481/1077)	42.8% (461/1077)	42.9% (462/1077)	46.4% (500/1077)
固有名詞正解率	43.9% (50/114)	58.8% (67/114)	40.4% (46/114)	56.1% (64/114)

表2からWEB上の類似記事を利用して、言語モデル更新する事により、単語正解率は、新聞LMの47.0%(1462/3112)から、47.8%(1488/3112)へ、単語正解精度においては、新聞LMの37.7%(1173/3112)から39.5%(1230/3112)へ上がっている。

また表3から、重要な単語である名詞は、新聞LMの44.7%(481/1077)から46.4%(500/1077)へ、固有名詞は、43.9%(50/114)から56.1%(64/114)へと正解率が上がっている。

#### 4. おわりに

本稿では、WEB上のニュース記事を利用して、話題適応化LMを作り音声認識を行う方法、固有名詞をクラスタリングして音声認識を行う方法を提案している。

そして、新聞LMを含めた、これら三手法の結果を、多数決法を用いて新たに認識結果を出すという手法も提案した。

この手法によって、単語正解率が47.8%、単語正解精度39.5%が得られている。また、名詞の正解率は46.4%、固有名詞の正解率は、56.1%が得られた。

本稿では汎用性のある言語モデルとして毎日新聞より作成した語彙数2万単語の言語モデルを基準に使用しているが、話し言葉に対応した言語モデルを使用

する事によって[13]、更なる精度の向上が望めるのではないかと考える。

今後は、実際の検索システムで、この音声認識手法が使用できるかを評価する必要がある。

#### 謝辞

この研究は、平成16年度(財)人工知能研究振興財団研究助成金によって行われた。

#### 文献

- [1] Ivan Bulyko, Mari Ostendorf and Andreas Stolcke. Getting More Mileage from Web Text Sources for Conversational Speech Language Modeling using Class-Dependent Mixtures. HLT-NAACL 2003
- [2] 尾上 和穂, 今井 享, 安藤 彰男: “記者原稿を用いたニュース音声認識結果の修正法”, 日本音響学会平成10年度春期研究発表会, 1-6-6, pp.11-12, 1998.
- [3] 松井 淳, 世木 寛之, 小林 彰男 “スポーツニュースを対象とした音声認識”, 日本音響学会, 2001春期研究発表会, 2001.
- [4] Eric K. Ringger and James F. Allen. ROBUST ERROR CORRECTION OF CONTINUOUS SPEECH RECOGNITION. ESCA-NATO Robust Workshop'97, 1997
- [5] 石川 開, 隅田 英一郎: “テキストデータを使った音声認識誤りの訂正”, 自然言語処理, Vol.7, No.4, pp.205-228, 2000.
- [6] 沖本 純幸, 山本 博史, 隅田 英一郎, 菊井 玄一郎: “旅行会話基本表現コーパスを用いた認識誤り訂正の検討”, 情報処理学会研究報告, 2001-SLP-39-27, pp.159-164, 2001.
- [7] 山口 辰彦, 酒向 慎司, 山本 博史, 菊井 玄一郎: “信頼尺度に基づく音声認識検出および誤り訂正”, 秋季講演論文集, 3-Q-14, pp.153-154. 日本音響学会, 2003.
- [8] 伊藤 克亘, 藤井 敦, 石川 徹也: “未知語検出を用いた語彙統制のない音声検索”, 日本音響学会12年度春期研究発表会, 2002.
- [9] 河原達也, 李晃伸, 小林哲則, 武田一哉, 峯松信明, 嵯峨山茂樹, 伊藤克亘, 伊藤彰則, 山本幹雄, 山田篤, 宇津呂武仁, 鹿野清宏. “日本語ディクショ基本ソフトウェア(99年度版)の性能評価”. 情報処理学会研究報告, 2000-SLP-31-1, pp.9-16, 2000.
- [10] 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”. 情報処理学会研究報告, SLP-35-9, pp.49-54, 2001-02
- [11] J.G.Fiscus. “A post-processing system to yield reduced word error rates: Recognizer Output Voting Error Reduction (ROVER).” Proc. of the ASRU'97, pp.347-354, 1997.
- [12] 宇津呂武仁, 小玉康広, 渡邊友裕, 西崎博光, 中川聖一: “機械学習を用いた複数の大語彙連続音声認識モデルの出力の混合”, 電子情報通信学会論文誌, Vol.J87-D-II, No.7, pp.1428-1440, 2004.7
- [13] 河原達也, “『日本語話し言葉コーパス』を用いた音声認識の進展,” 第3回話し言葉の科学と工学ワークショップ講演予稿集, pp.61-65, 2004.