

音声入力型文書検索システムの開発とテストコレクションの構築

藤井 敦^{†,†††} 伊藤 克亘^{††,†††} 秋葉 友良^{††} 石川 徹也[†]

[†] 図書館情報大学

〒305-8550 つくば市春日 1-2

^{††} 産業技術総合研究所

〒305-8568 つくば市梅園 1-1-1 中央第2

^{†††} 科学技術振興事業団 CREST

E-mail: fujii@ulis.ac.jp

あらまし 本研究は、音声発話によって必要な情報を検索するためのシステムに関する研究開発を行った。従来の関連研究では、テキスト検索精度の向上が主要なテーマであり、音声認識精度の向上は対象とされてこなかった。本研究で提案するシステムの特長は、検索対象テキストを用いて音声認識用の言語モデルを作成し、利用する点にある。ユーザの発話は検索対象テキストに関連するものが中心なので、本システムは比較的高精度の音声認識を実現できる。さらに、システムを評価するためのテストコレクションを構築するために、既存のテキスト検索用コレクションを利用し、検索要求に関する音声発話データを収録した。当該テストコレクションを用いた実験によって本システムの有効性を確認することができた。

キーワード テキスト検索、音声認識、言語モデル、NTCIR コレクション、読み上げ音声

A Speech-Driven Text Retrieval System and its Evaluation Using a Test Collection

Atsushi Fujii^{†,†††}, Katunobu Itou^{††,†††}, Tomoyosi Akiba^{††}, Tetsuya Ishikawa[†]

[†] University of Library and Information Science

1-2 Kasuga, Tsukuba, 305-8550

^{††} National Institute of Advanced Industrial Science and Technology

1-1-1 Chuou Daini Umezono, Tsukuba, 305-8568, Japan

^{†††} CREST, Japan Science and Technology Corporation

E-mail: fujii@ulis.ac.jp

Abstract To facilitate retrieving information with spoken queries, we propose a speech-driven text retrieval system. In past research, no attempt has been made to improve speech recognition in the context of speech-driven retrieval. In our system, a language model used for speech recognition is produced based on a target text collection, so that user queries associated with the collection can be recognized with a high accuracy. We also produced a test collection to evaluate our system, for which we recorded dictated queries in an existing collection for text retrieval. We show the effectiveness of our system by way of experiments using this collection.

Keywords text retrieval, speech recognition, language models, the NTCIR collection, read speech

1 はじめに

近年の音声認識技術は、内容がある程度整理されている発話に対しては実用的な認識精度を達成できる。また、ハードウェア技術の発展にも支えられ、パソコン上で動作する商用/無償の音声認識ソフトウェアが存在する。そこで、既存のアプリケーションに音声認識を導入することは比較的容易になっており、その需要は今後ますます増加するだろう。

とりわけ、情報検索システムは歴史が長く主要な情報処理アプリケーションの一つであるため、音声認識を採り入れた研究も近年数多く行われている。これらは目的に応じて以下の2つに大別できる。

- 音声データの検索 [5, 12, 13, 14, 16, 17]
放送音声データなどを対象にした検索である。入力手段は問わないものの、テキスト(キーボード)入力を中心である。
- 音声による検索 [2, 3]
検索要求を音声入力によって行う。検索対象の形式は問わないものの、テキストが中心である。

すなわち、これらは検索対象と検索要求のどちらを音声データと捉えるかが異なる。さらに、両者を統合すれば、音声入力による音声データ検索を実現することも可能である。しかし、現在そのような研究事例はあまり存在しない。

音声データの検索は、TRECのSpoken Document Retrieval (SDR)トラック [4]で放送音声データを対象にしたテストコレクションが整備されていることを背景にして、盛んに研究が行われている。

他方において、音声による検索はカーナビゲーションシステムやコールセンターのようにキーボード入力を前提としない(バリアフリーな)アプリケーションを支える重要な基盤技術であるにも拘らず、音声データ検索に比べて研究事例は極端に少ない。

そこで、本研究は音声入力によるテキスト検索に焦点を当て、システムの研究開発を行った(2章)。さらに、テストコレクションを構築・整備し(3章)、本システムの評価に利用した(4章)。

2 本研究で提案する検索システム

2.1 背景と動機

音声による検索に関する従来のシステムでは、概して、音声認識とテキスト検索は完全に独立したモジュールとして存在し、単に入出力インタフェースで接続されているだけである。また、検索精度の向

上に焦点が当てられ、音声認識精度の向上は研究対象となっていないことが多い。

Barnettら [2]は、既存の音声認識システム(語彙サイズ20,000)をテキスト検索システムINQUERYの入力として利用して、音声による検索の評価実験を行った。具体的には、TRECの検索課題35件(101-135)に対する単一話者の読み上げ音声をテスト入力として利用し、TRECコレクションの検索実験を行った。Crestani [3]も上記35件の読み上げ検索課題を用いた実験を行い(通常のテキスト検索で用いられる)適合性フィードバックによって検索精度が向上することを示している。しかし、どちらの実験においても既存の音声認識システムを改良せずに利用しているため、単語誤り率は比較的高い(30%以上)。

それに対して、本研究では音声認識の精度向上にも焦点を当て、音声認識とテキスト検索の有機的な統合を目指してシステムの研究開発を行った。音声認識システムの多くは、内部データ(単語辞書など)を切替えることで、目的に応じた使い分けが可能である。そこで、検索対象テキストに基づいて内部データを作成し、利用することは自然な発想だろう。

統計的な音声認識システム [1]は主に音響モデルと言語モデルで構成され、両者は音声認識精度に強く影響する。音響モデルは音響的な特性に関するモデルであり、検索対象テキストとは独立な要素である。

言語モデルは音声認識結果(候補)の言語的妥当性を定量化するためのモデルである。しかし、あらゆる言語現象全てをモデル化することは不可能であるため、一般的には、与えられた学習用コーパスに出現する言語現象に特化したモデルを作成する。

本研究で提案するシステムにおいては、ユーザの発話は検索対象テキストに関連する内容である可能性が高い。そこで、検索対象テキストに基づいて言語モデルを作成すれば、音声認識の精度向上が期待できる。その結果、ユーザの発話が正しく認識されるので、テキスト入力に近い検索精度を実現することが可能になる。

音声認識の精度を高めることは、インタラクティブ検索を円滑に進めたり、発話通りの要求に基づいて検索が行われている安心感をユーザに与える上でも重要である。

2.2 システム構成

本研究で提案する検索システムの構成を図1に示す。本システムの特長は、検索テキストに基づいて音声認識精度を高めることで、音声認識とテキスト検

索の有機的な統合を実現する点にある。そこで、まずオフライン処理（破線矢印）によって、検索対象となるテキストコレクションから音声認識用の言語モデルを作成する。

オンライン処理では、ユーザが検索要求を発話すると、音響モデルと言語モデルを用いて音声認識が行われ、書き起こしが生成される。実際には、複数の書き起こし候補が生成され、尤度を最大化する候補が選択される。ここで、言語モデルはテキストコレクションに基づいて作成されているので、コレクション中のテキストに言語的に類似する書き起こしが優先的に選択される点に注意を要する。

次に、書き起こされた検索要求を用いてテキスト検索を実行し、検索結果をユーザに提示する。検索結果閲覧を支援するためには効果的な提示手法が必要である。しかし、出力インタフェースについては今回の研究では対象外とし、今後検討を行う。

なお、本システムは現在は日本語を対象に実装されているものの、原理的には対象言語を問わない。

以下、2.3、2.4 節で音声認識とテキスト検索についてそれぞれ説明する。

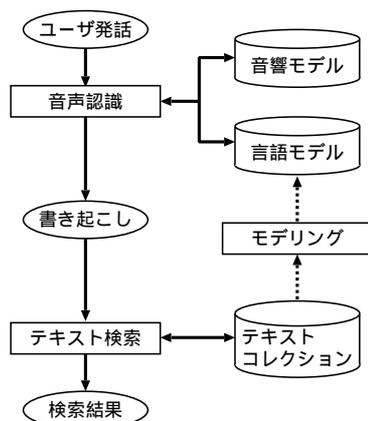


図 1: 音声入力型テキスト検索システムの構成

2.3 音声認識

音声認識には連続音声認識コンソーシアムの日本語ディクテーション基本ソフトウェアを用いた [6, 19]¹。本ソフトウェアは 2 万語規模の単語辞書を用いて、ほぼ実時間に近い動作で 90% の認識精度を実現できる。

音響モデルと認識エンジン（デコーダー）は、本ツールキット付属のものを変更せずに利用する。

他方において、統計的言語モデル（単語 N グラム）は検索対象のテキストコレクションに基づいて作成する。本ソフトウェアに付属されている関連ツール群

¹<http://www.lang.astem.or.jp/CSRC/>

や一般に利用可能な形態素解析システム「茶筌」[20] を併用することで、様々な対象に対して比較的容易に言語モデルを作成できる。すなわち、対象テキストから不要部分を削除するなどの前処理を行い「茶筌」を用いて形態素に分割し、読みを考慮した高頻度語制限モデルを作成する [18]。

2.4 テキスト検索

テキスト検索には確率的手法 [9] を用いた。本手法は、近年のいくつかの評価実験によって比較的高い検索精度を実現することが示されている。

検索要求が与えられると、索引語の頻度分布に基づいてコレクション中の各テキストに対する適合度を計算し、適合度が高いテキストから優先的に出力する。テキスト i の適合度は式 (1) によって計算する。

$$\sum_t \left(\frac{TF_{t,i}}{\frac{DL_i}{avglen} + TF_{t,i}} \cdot \log \frac{N}{DF_t} \right) \quad (1)$$

ここで、 t は検索要求（本システムでは、ユーザ発話の書き起こしに相当する）に含まれる索引語である。 $TF_{t,i}$ はテキスト i における索引語 t の出現頻度である。 DF_t は対象コレクションにおいて索引語 t を含むテキストの数であり、 N はコレクション中のテキスト総数である。 DL_i はテキスト i の文書長（バイト数）であり、 $avglen$ はコレクション中の全テキストに関する平均長である。

適合度を適切に計算するためには、オフラインでの索引語抽出（索引付け）が必要である。そこで「茶筌」を用いて単語分割、品詞付与を行う。さらに、品詞情報に基づいて内容語（主に名詞）を抽出し、単語単位で索引付けを行って転置ファイルを作成する。オンライン処理では、書き起こされた検索要求に対しても同様の処理で索引語を抽出し、検索に利用する。

3 テストコレクションの構築

3.1 背景と動機

研究開発した検索システムの性能を定量的に評価することは、問題点の分析や改善を行うために重要である。システムの評価には大きく分けて「実世界における試験運用」と「研究室における実験」がある。

前者は、入出力インタフェースのデザインや応答時間などの運用上の問題まで考慮して評価を行う必要がある。

それに対して、後者は比較的限定された設定のもとで行う評価である。すなわち、あらかじめ用意さ

れた検索要求に対してシステムが出力した検索結果を何らかの尺度によって評価する。そのためには、テストコレクションを構築し、評価用ベンチマークとして利用する手法が効果的である。

大規模なテストコレクションの作成には膨大なコストを要するものの、一旦作ってしまえばシステムの性能評価を繰り返すことが容易になる。そこで、被験者をその都度雇わなくても、様々な手法を比較評価しながらシステムを改善できる。

そこで、本研究は「研究室での実験」に焦点を当て、音声入力型テキスト検索システムのためのテストコレクション構築を目的とした。

Barnettら [2] は、TREC の検索課題 35 件を話者に読み上げてもらい、音声による検索要求データを作成した。そこで、当該データを TREC のテキストコレクションおよび適合性判定と併用することで、音声入力によるテキスト検索の精度を定量的に評価することが可能である。

また、一般公開されているテキスト検索コレクションを利用すれば、関連データを研究者間で共有することも容易になり、当該分野の発展に貢献できる。

ただし、研究室における実験の場合であっても、テスト入力として利用する検索要求は実世界の検索要求にできるだけ近づけることが好ましい。そのためには、ユーザ発話の特性について（少なくとも）次節に示す 5 つの観点から検討する必要がある。

3.2 ユーザ発話に関する特性

内容の詳細度 情報検索を広義に解釈すれば、ユーザが抱えている問題を解決できる情報を見つけ出すことである。すなわち、言語化されていない直観的な情報要求 (visceral need [15]) を検索質問 (query) に言語化してから検索を行い、検索したテキスト内容を理解する一連の処理である。

しかし、多くの研究では情報検索は狭義に解釈され、言語化された検索質問をシステムに対する直接の入力として扱う。例えば、キーワードは極端に具体化された検索質問の形態である。

しかし、音声入力型システムの場合は、キーボード入力では躊躇するような曖昧な要求でも、発話することは容易かもしれない。そこで、通常のテキスト入力検索に比べて、相対的に詳細度の低い検索要求まで考慮する必要がある。すなわち、音声認識の観点から見れば、短い単語から長い文まで一律に認識できる必要がある。

発話の長さ ユーザが一度に無理なく自然に発話できる長さ（単語数や時間）には限度があるため、極端に長い検索要求を利用することは現実的ではない。

これはテキスト入力型の検索システムについても当てはまる。しかし、テキスト入力の場合は、類似文書検索のように他の場所からダウンロード（コピー）した長いテキストも入力できるため、検索要求の長さに関する制約は比較的緩やかである。

内容（語彙）の多様性 適用範囲の広いシステムを目的とする場合、多様な検索要求を評価に用いることが好ましい。情報検索の研究分野では、経験的に数十以上の検索要求を利用することが推奨されている。音声認識の観点からは、検索要求に含まれる語彙の多様性に関する評価と捉えることができる。

発話スタイル 音声認識の評価では、読み上げ音声 (read speech)、自発音声 (spontaneous speech)、会話音声 (conversational speech) 等の発話スタイルを区別することが重要である。

検索システムの入力として利用する場合は、読み上げ/自発音声を区別する必要がある。Barnettら [2] が行ったように、既存の検索課題をそのまま発話すれば読み上げ音声である。他方において、話者が検索課題を理解し、課題達成に必要な検索要求を自分で考えて発話すれば、自発音声に近くなる。

話者の特性 対象を特定話者に限定するのか、不特定話者を想定するのかを考慮して、目的に応じた検索要求を作成する必要がある。また、性別や年齢などの多様性についても検討する必要がある。

以上 5 項目のうち、最初の 3 項目は音声認識とテキスト検索に共通の観点であり、残りの 2 項目は音声認識に固有の観点である点に注意が必要である。

3.3 対象にしたテキストコレクション

現在一般に入手可能な日本語テキスト検索用テストコレクションのうち、代表的なものを以下に挙げる。ただし、各コレクションの詳細は割愛する。

- NTCIR [7, 8]²
- IREX [11]³
- BMIR-J2 [10]⁴

IREX と BMIR-J2 が毎日新聞記事を検索対象テキストとしているのに対して、NTCIR は技術文書（論文抄録、科研費成果報告書概要）を対象にしている点

²<http://research.nii.ac.jp/ntcir/index-ja.html>

³<http://cs.nyu.edu/cs/projects/proteus/irex/>

⁴<http://www.ulis.ac.jp/~ishikawa/bmir-j2/>

が異なる．いずれのコレクションも本研究の目的に適しているものの，今回は NTCIR を対象にした．

広義には，NTCIR コレクションには多言語検索や自動要約用のコレクションも含まれる．しかし，今回は日本語検索コレクションのみを対象にした．そこで，以下では，NTCIR コレクションを日本語検索コレクションと同義で扱う．さらに，日本語検索コレクションには，NTCIR-1(予備版，公式版)，NTCIR-2(公式版) の 3 種類の異なるコレクションが存在する．

それぞれの検索課題数およびテキスト数を表 1 に示す．ここで，NTCIR-1 の予備・公式版に含まれるテキストは同一であり，NTCIR-2 公式版テキストは NTCIR-1 を完全に包含している．ただし，複数のコレクションに重複して含まれる検索課題はない．

すなわち，合計 132 件の検索課題を用いて音声発話による検索要求データを作成した．

表 1: NTCIR 日本語検索コレクションの構成

| コレクション名 | 検索課題数(課題番号) | テキスト数 |
|-------------|------------------|---------|
| NTCIR-1 予備版 | 30 (0001-0030) | 332,918 |
| NTCIR-1 公式版 | 53 (0031-0083) | 332,918 |
| NTCIR-2 公式版 | 49 (0101-0149) | 736,166 |

3.4 音声検索要求データの作成

図 2 に NTCIR コレクションに含まれる検索課題の例を示す．各検索課題は SGML 形式で記述された複数の項目で構成される．これらのうち，検索システムの評価に使用することが推奨されている項目は，主に<TITLE>，<DESCRIPTION>，<NARRATIVE>の 3 つである．これらは，いずれも検索要求を表現するものの，後半ほど詳細であり長い．

我々は 3.2 節の議論に基づいて，音声による検索要求を作成した．

まず，内容の詳細度と発話の長さの観点から，<DESCRIPTION>のみを利用した．すなわち，システムの入力としては，専門用語と一般語が混在する文単位の発話を想定する．

内容(語彙)の多様性については，NTCIR 検索課題の範囲に事実上制限される．しかし，情報処理，医学などの多分野に渡っている点に注意を要する．

発話スタイルに関しては，<DESCRIPTION>の読み上げを行った．今回のテストコレクション構築はパイロット的な側面もあるため，比較的音声認識を行いやすい対象を選択した．今後は対象の難易度を段階的に上げていき，最終的には検索課題に基づいた自発音声を収録することも検討している．

話者の特性については，男女 2 名ずつ合計 4 名(いずれも大学院生，年齢は 20 代)の話者に読み上げを依頼した．最終的には，合計 10 名を目標に収録を行う予定である．

収録環境は録音スタジオで，各発話は卓上型マイクを用いて DAT に収録した．132 件の<DESCRIPTION>は全て通して収録した(途中で一度休憩を入れた)．ただし，132 件をランダムな順序で記述した原稿を渡した．すなわち，疲れや慣れなどの身体的・心理的な影響によって，コレクションごとの発話状態がばらつかないようにした．読み間違えた場合には，正しく読めるまで同一検索課題を続けて発話してもらった．

なお，Barnett ら [2] が作成した TREC 検索課題の音声発話については，単一話者による読み上げであること以外の詳細(検索課題中のどの項目をどのような手順で収録したのかなど)は不明である．

4 実験

4.1 方法

本研究の検索システム(2 章)をテストコレクション(3 章)を用いて評価した．評価方法は，システムの入力として音声発話データを利用する点を除けば，NTCIR ワークショップにおける評価と同じである．

すなわち，各検索要求に適合するテキスト上位 1,000 件を出力し，再現率-適合率曲線と補間なし平均適合率に基づいて，以下に示す異なるシステム(手法)の比較評価を行った．

1. テキスト入力型のシステム

音声認識を 100%正しく行うシステム，すなわち理想的なシステムと見なすことができる．

2. 音声入力型システム

本研究で提案するシステムであり，NTCIR-2 公式版コレクションに基づいて作成された言語モデル(NTCIR モデル)を用いて音声認識を行う．NTCIR-2 公式版テキストは NTCIR-1 予備・公式版を包含しているので，検索対象によらずに全て当該モデルを使用した．

3. 音声入力型システム

本評価実験におけるベースラインとして利用する．日本語ディクテーション基本ソフトウェアに付属されている言語モデルを用いて音声認識を行う．当該モデルは，毎日新聞 75 か月分に基づいて作成された(新聞モデル)．

上記システムのうち，2 と 3 における違いは，言語モデルに利用するデータ(コーパス)だけに起因す

<TOPIC q=0101>
 <TITLE> B型肝炎</TITLE>
 <DESCRIPTION>遺伝子工学的手法によるB型肝炎ワクチンの開発について論じている文献</DESCRIPTION>
 <NARRATIVE>肝炎などのウイルス性疾患に対する安全かつ有効な予防法の確立は21世紀に向けての医療分野での重要な課題である。そのため、遺伝子工学的手法によるB型肝炎ワクチンの開発について論じていけば検索要求を満たす。開発されたB型肝炎ワクチンの物理化学的特性を論じているものやその免疫力増強に有効な免疫アジュバントについて論じているものも検索要求を満たす。しかし、遺伝子工学的手法に触れていない論文は不可。また、B型肝炎以外のワクチンも不可。</NARRATIVE>
 <CONCEPT>a. B型肝炎, b. 遺伝子工学的手法, c. ワクチン, 予防接種</CONCEPT>
 <FIELD>7. 医学・歯学</FIELD>
 </TOPIC>

図 2: NTCIR コレクション検索課題の例 (NTCIR-2 公式版, 課題番号 0101)

る。各言語モデルに使用した学習データの内訳を表 2 に示す。いずれの場合も同一のツール群を用いて、語彙サイズ 20,000 の単語トライグラムを作成した。

表 2: 言語モデル作成に使用したデータの比較

| | NTCIR-2 | 新聞記事 |
|-------|---------|------|
| 総形態素数 | 178M | 117M |
| 異なり語数 | 459K | 253K |

システム 2 と 3 では話者 4 名の発話を個別に使用したので、事実上、9 種類の異なる検索結果を比較評価した。

NTCIR コレクションの適合判定には、適合 (S, A), 部分適合 (B), 不適合 (C) のレベルがある。本実験では「適合」のみを正解と見なした。

4.2 結果と考察

まず、各システムの検索精度を概観するために、再現率-適合率 (recall-precision) 曲線をコレクションごとに図 3 ~ 5 に示す。グラフが右上に位置するほど検索精度が高い。

これらの図より、話者によらずに、本システムの検索精度は新聞モデルを用いたシステムの検索精度を大幅に改善していることが分かる。言い替えれば、本研究の趣旨すなわち検索対象テキストに基づく言語モデルの利用が妥当であることが分かった。

本システムの検索精度はテキスト入力型システムの精度には及ばないものの、NTCIR-1 予備版においては、かなり接近していることが分かる。

次に、検索精度と音声認識精度の関係を分析するために、補間なし平均適合率と音声認識誤り率を表 3 に示す。誤り率としては単語誤り率 (word error rate: WER) が一般的である。これは、式 (2) を用いて、正解発話に対する削除・挿入・置換の割合を単語単位で計算する尺度である。

$$\frac{\text{削除} + \text{挿入} + \text{置換}}{\text{全語数}} \quad (2)$$

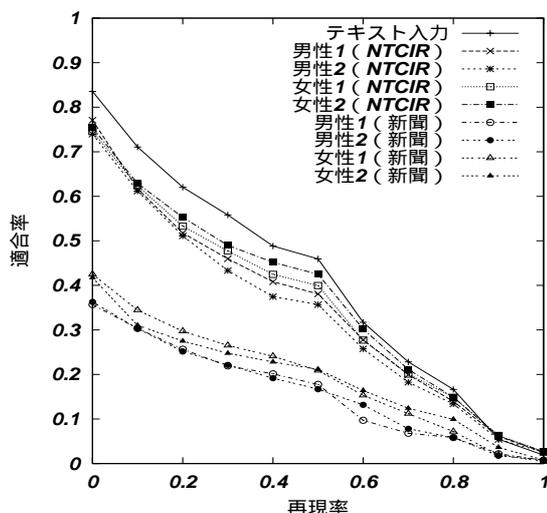


図 3: 再現率-適合率曲線 (NTCIR-1 予備版)

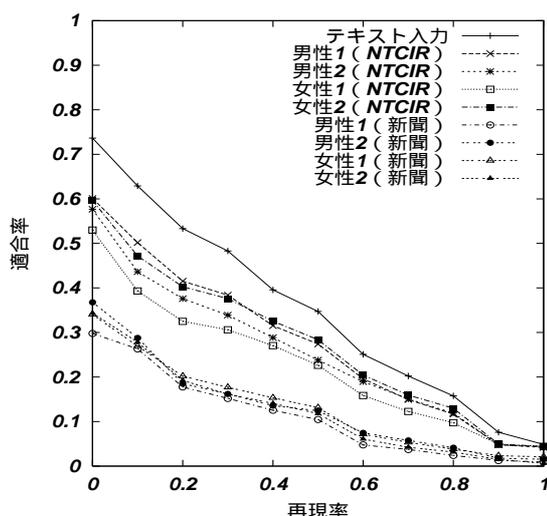


図 4: 再現率-適合率曲線 (NTCIR-1 公式版)

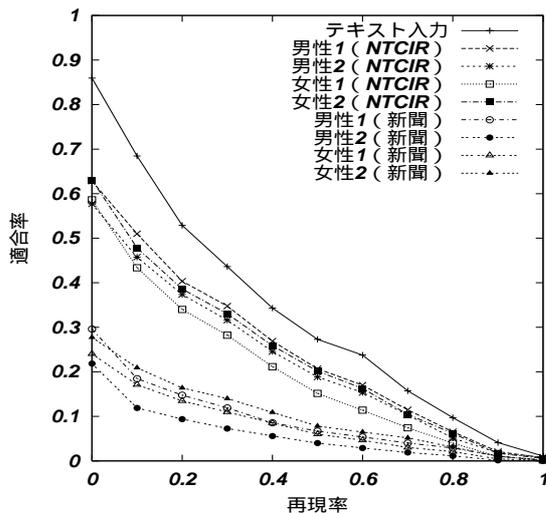


図 5: 再現率-適合率曲線 (NTCIR-2 公式版)

しかし、実際には音声認識結果中の単語全てが検索に利用される訳ではない。そこで、検索に利用された検索キーワードの単位でも誤り率 (keyword error rate: KER) を計算した。

概して、NTCIR モデルは新聞モデルの単語・キーワード誤り率を大幅に削減し、音声認識、検索精度 (平均適合率) の両面から効果的であることが分かる。この傾向は検索課題に専門用語を多く含む場合に、より顕著であった。

例えば、検索課題 0027 「シソーラスを用いたテキストの連想検索について」の認識結果と平均適合率を言語モデルごとと比較した結果を表 4 に示す。これは「シソーラス」や「連想検索」などの専門性の高い単語の認識誤りが平均適合率に強く影響する典型的な例である。

しかし、検索課題 0119 「日本人の生活価値観の変化」のような一般語だけで構成される発話は、どちらのモデルでも話者によらず正しく認識された。

他方において、NTCIR-1 予備版以外は、単語誤り率よりもキーワード誤り率の方が高いことも分かった。NTCIR モデルを用いると、検索対象テキスト (すなわち、技術文書) に含まれない表現に対する認識精度が低い。

例えば、NTCIR 検索課題には「～の文献が欲しい」という表現が頻出する。しかし「欲しい」という語が NTCIR モデルでは未知語となるため「～の文献が星」のように誤認識され、その結果「星」が名詞性のキーワードとして抽出されてしまった。

また、トライグラムモデルでは語の共起が考慮されるため、発話中の一部が誤認識されると周辺語も

連鎖的に誤認識される可能性がある。そこで、今後は検索対象テキストと検索要求との表現の不整合について対処する必要がある。

5 おわりに

音声認識とテキスト検索を統合し、音声発話によって高精度のテキスト検索を実現するシステムを提案した。本システムの特長は、検索対象テキストに基づいて音声認識用の言語モデルを作成することで音声認識精度を高める点にある。

さらに、システム評価用のベンチマークを構築するために、既存のテキスト検索用テストコレクションを用いて検索要求に対する音声発話データを作成した。

当該テストコレクションを用いた評価実験の結果、本システムの検索精度はテキスト入力型システムの精度に比べて若干劣るものの、既存の音声認識システムを拡張せずに利用した場合に比べてはるかに高いことが示された。今後は、音声発話データの規模拡張など行いながら、システムの性能をさらに向上させる予定である。

謝辞

NTCIR コレクションは国立情報学研究所の許諾を得て使用させて頂きました。関和広氏、伊堂史織氏 (図書館情報大学) には実験用データの整備を補助して頂きました。本研究の一部は日本学術振興会科学研究費補助金 (課題番号 12680406) の助成による。

参考文献

- [1] Lalit. R. Bahl, Frederick Jelinek, and Robert L. Mercer. A maximum likelihood approach to continuous speech recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 5, No. 2, pp. 179-190, 1983.
- [2] J. Barnett, S. Anderson, J. Broglio, M. Singh, R. Hudson, and S. W. Kuo. Experiments in spoken queries for document retrieval. In *Proceedings of Eurospeech97*, pp. 1323-1326, 1997.
- [3] Fabio Crestani. Word recognition errors and relevance feedback in spoken query processing. In *Proceedings of the Fourth International Conference on Flexible Query Answering Systems*, pp. 267-281, 2000.
- [4] John S. Garofolo, Ellen M. Voorhees, Vincent M. Stanford, and Karen Sparck Jones. TREC-6 1997 spoken document retrieval track overview and results. In *Proceedings of the 6th Text REtrieval Conference*, pp. 83-91, 1997.
- [5] S.E. Johnson, P. Jourlin, G.L. Moore, K. Spärck Jones, and P.C. Woodland. The Cambridge University spoken document retrieval system. In *Proceedings of ICASSP'99*, pp. 49-52, 1999.

表 3: 検索実験結果 (AP: 平均適合率, WER: 全単語誤り率, KER: キーワード誤り率)

| 入力方法 | NTCIR-1 予備版 | | | NTCIR-1 公式版 | | | NTCIR-2 公式版 | | |
|----------------|-------------|--------|--------|-------------|--------|--------|-------------|--------|--------|
| | AP | WER | KER | AP | WER | KER | AP | WER | KER |
| テキスト | 0.3877 | — | — | 0.3320 | — | — | 0.3118 | — | — |
| 男性 1 (NTCIR) | 0.3301 | 0.2123 | 0.2041 | 0.2609 | 0.1598 | 0.2120 | 0.2320 | 0.1605 | 0.2482 |
| 男性 2 (NTCIR) | 0.3145 | 0.2945 | 0.2789 | 0.2379 | 0.2228 | 0.2753 | 0.2119 | 0.2297 | 0.2956 |
| 女性 1 (NTCIR) | 0.3388 | 0.2055 | 0.2245 | 0.2116 | 0.1719 | 0.2690 | 0.1853 | 0.1841 | 0.2847 |
| 女性 2 (NTCIR) | 0.3507 | 0.1678 | 0.1429 | 0.2617 | 0.1380 | 0.2089 | 0.2213 | 0.1635 | 0.2555 |
| 男性 1 (新聞) | 0.1504 | 0.4658 | 0.5510 | 0.1030 | 0.3668 | 0.5759 | 0.0847 | 0.3918 | 0.5876 |
| 男性 2 (新聞) | 0.1536 | 0.6986 | 0.7278 | 0.1219 | 0.4529 | 0.6456 | 0.0512 | 0.5110 | 0.6606 |
| 女性 1 (新聞) | 0.1820 | 0.5514 | 0.6190 | 0.1213 | 0.3850 | 0.5854 | 0.0727 | 0.4021 | 0.5620 |
| 女性 2 (新聞) | 0.1803 | 0.4760 | 0.5374 | 0.1138 | 0.3341 | 0.5348 | 0.0941 | 0.3697 | 0.5803 |

表 4: 検索課題 0027 (シソーラスを用いたテキストの連想検索について) の音声認識結果と平均適合率

| 入力 | NTCIR モデル | 新聞モデル |
|------|-------------------------------------|-------------------------------------|
| 男性 1 | シソーラスを用いたテキストの連想検索について。 (0.6872) | 市長ら層を用いたテキストの連想を原作について。 (0.0109) |
| 男性 2 | シソーラスを用いたテキストの連鎖を検索について。 (0.2335) | 起訴をら三落ちたテキストの連鎖を検索について。 (0.0073) |
| 女性 1 | シソーラスを用いた。エキスの連想検索について。 (0.6644) | 小さな村落ちたが、って聞いても連想検索について。 (0.2491) |
| 女性 2 | シソーラスを用いた、テキストの連想検索について。 (0.6872) | 市長ら数を用いた。テキストの連投検索について。 (0.0119) |

- [6] T. Kawahara, A. Lee, T. Kobayashi, K. Takeda, N. Minematsu, S. Sagayama, K. Itou, A. Ito, M. Yamamoto, A. Yamada, T. Utsuro, and K. Shikano. Free software toolkit for Japanese large vocabulary continuous speech recognition. In *Proceedings of the 6th International Conference on Spoken Language Processing*, pp. 476–479, 2000.
- [7] National Center for Science Information Systems. *Proceedings of the 1st NTCIR Workshop on Research in Japanese Text Retrieval and Term Recognition*, 1999.
- [8] National Institute of Informatics. *Proceedings of the 2nd NTCIR Workshop Meeting on Evaluation of Chinese & Japanese Text Retrieval and Text Summarization*, 2001.
- [9] S. E. Robertson and S. Walker. Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval. In *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 232–241, 1994.
- [10] Tetsuya Sakai, Tsuyoshi Kitani, Yasushi Ogawa, Tetsuya Ishikawa, Haruo Kimoto, Ikuo Keshi, Jun Toyoura, Toshikazu Fukushima, Kunio Matsui, Yoshihiro Ueda, Takenobu Tokunaga, Hiroshi Tsuruoka, Hidekazu Nakawatase, Teru Agata, and Noriko Kando. BMIR-J2: A test collection for evaluation of japanese information retrieval systems. *ACM SIGIR FORUM*, Vol. 33, No. 1, pp. 13–17, 1999.
- [11] Satoshi Sekine and Hitoshi Isahara. IREX: IR and IE evaluation project in Japanese. In *Proceedings of the 2nd International Conference on Language Resources and Evaluation*, pp. 1475–1480, 2000.
- [12] Páraic Sheridan, Martin Wechsler, and Peter Schäubel. Cross-language speech retrieval: Establishing a baseline performance. In *Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 99–108, 1997.
- [13] Amit Singhal and Fernando Pereira. Document expansion for speech retrieval. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 34–41, 1999.
- [14] Savitha Srinivasan and Dragutin Petkovic. Phonetic confusion matrix based spoken document retrieval. In *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 81–87, 2000.
- [15] Roberto S. Taylor. The process of asking questions. *American Documentation*, Vol. 13, No. 4, pp. 391–396, 1962.
- [16] Martin Wechsler, Eugen Munteanu, and Peter Schäubel. New techniques for open-vocabulary spoken document retrieval. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 20–27, 1998.
- [17] Steve Whittaker, Julia Hirschberg, John Choi, Don Hindle, Fernando Pereira, and Amit Singhal. SCAN: Designing and evaluating user interfaces to support retrieval from speech archives. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 26–33, 1999.
- [18] 伊藤克巨, 山田篤, 天白成一, 山本俊一郎, 踊堂憲道, 宇津呂武仁, 山本幹雄, 鹿野清宏. 日本語ディクテーションのための言語資源・ツールの整備. 情報処理学会研究報告 99-SLP-26-5, 1999.
- [19] 鹿野清宏, 伊藤克巨, 河原達也, 武田一哉, 山本幹雄 (編). 音声認識システム. オーム社, 2001.
- [20] 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『茶釜』version 2.0 使用説明書. Technical Report NAIST-IS-TR99012, 奈良先端科学技術大学院大学, 1999.