

**科学研究費助成事業 研究成果報告書**

平成 29 年 6 月 13 日現在

機関番号：13904

研究種目：基盤研究(C)（一般）

研究期間：2013～2016

課題番号：25330130

研究課題名（和文）自由発話音声による情報要求を活用したテキスト・音声・画像への情報アクセスの研究

研究課題名（英文）A research on information access targeting text, peech and images by making use of spontaneously spoken information need

研究代表者

秋葉 友良（Tomoyosi, Akiba）

豊橋技術科学大学・工学（系）研究科（研究院）・准教授

研究者番号：00356346

交付決定額（研究期間全体）：（直接経費） 3,800,000円

研究成果の概要（和文）：従来の音声入力情報検索は、テキスト入力の代替に過ぎず、音声固有の特長を活用していない。本研究では、音声の特徴である思考の即時表現可能性や省労力性・可搬性を活かした情報要求の表現方法を考察し、それを利用した自由発話音声を入力とする情報アクセス手法を開発することを目的とする。被験者実験により自発的に表現した音声クエリを収録し、再利用可能なテストコレクションを構築した。これを用いて、国内外から研究グループを集めて自由発話音声ドキュメント検索タスクを国際評価ワークショップで実施した。また、集めたデータを活用し自由発話音声の音響的特徴を活用した検索手法をはじめとする種々の新しい音声検索手法を開発した。

研究成果の概要（英文）：Since current speech driven information retrieval systems simply replace textual input to speech, it fail to exploit the advantage comes from nature of speech, which is most natural, easy, and efficient way for humans to express their own ideas. This research work aims to develop the information retrieval systems that make use of such advantages by taking spontaneously spoken queries as their input. We collected spontaneously spoken queries through our subjective experiment, which result in compiling a reusable test collection. By using the collection, we took place the shared task of spoken document retrieval from spontaneously spoken queries within an international evaluation workshop, to which many research groups attended from home and abroad. Furthermore, by using the collected data., we also developed several novel methods for speech-driven spoken document retrieval, including that exploits acoustic features extracted from spontaneous spoken queries.

研究分野：音声言語情報処理

キーワード：ボイスサーチ 自由発話音声 音声ドキュメント検索 情報検索 テストコレクション 音声中の検索  
後検出 音声内容検索

## 1. 研究開始当初の背景

音声・画像・ビデオの記録・編集機器の拡大、およびインターネットをはじめとする情報通信網の発展により、誰でも気軽にコンテンツを作成・公開することが可能となり、マルチメディアコンテンツの情報爆発が進行しつつあった。コンテンツの量が増えるにつれて、必要なコンテンツにたどり着くことは年々困難になり、効果的な情報アクセス技術の確立が至急の課題となっていた。

検索者が情報要求を表現する手段としては、テキストで入力するのが一般的であるが、少数のキーワードだけで情報要求を表現するには限界がある。一方、労力の少ない情報要求の表現手段として、音声で検索クエリを入力する方法が出現し始めてきた。しかし、Google ボイスサーチをはじめとする音声入力による検索方法は、キーボードによるテキスト入力の置き換えでしかなく、音声入力の長を十分に活用していない。これらはキーボード入力の代替として音声入力を使っているに過ぎないので、検索者は音声入力の前に十分に情報要求を吟味して文にまとめることを強いられることになる。

## 2. 研究の目的

表現手段としての音声の特長は、思考を即時に表現可能であること、思考の中断にならないこと、誰でも容易に労力なく短時間で、専用の入力機器を必要とせずどこでも利用できること、にあると考えられる。この長を活用し、検索者には情報要求を自由に発想することを許し、それを即時に音声発話して表現することを可能にするような、自由発話音声を入力とする情報アクセス手法を開発する。検索対象として、テキストだけでなく、音声データ、および画像データを対象とする。自由発話音声で表現した情報要求は、簡潔で十分に吟味された従来のキーワードリストなどとは異なり、冗長で漠然としたものである一方、大量で豊富な手がかりが得られることが期待できる。また、検索者が十分に言語化できていない漠然とした情報要求を持つ場合でも、まずは話し始める事で漸進的かつインタラクティブに情報にアクセスすることを支援する新たな利用方法が期待できる。

## 3. 研究の方法

音声の特長を活かした情報要求の表現方法を調査・分析した上で、新しい検索クエリ入力法を提案し、それを利用した情報アクセス手法を開発することを目的とする。この目的を実現するために、次の3つの課題を設定し研究を進めることで、提案法の有効性を明らかにする。

### (A) 自由発話音声による情報要求の収集実験と分析

既存の検索テストコレクションの検索課題に対応する自由発話音声を収集する。収集した自由発話検索クエリと、既存のテキストクエリの対について分析を行うことで、自由発話検索クエリを活用したテキスト検索手法を検討する。

### (B) 自由発話音声を情報要求に用いた音声ドキュメント検索システムの開発

自由発話音声を入力とするように、申請者が先行する科学研究課題で開発した音声ドキュメントの検索手法を拡張し、音声入力による音声データの検索を実現する。

### (C) 自由発話音声による画像データ表現の調査とそれを用いた画像検索システムの開発

画像を要求する音声言語表現を収集し、調査・分析を行う。分析結果を利用し、自由発話音声を情報要求とする画像検索システムを実装する。

## 4. 研究成果

本研究課題の目的である「自由発話音声による音声ドキュメント検索」を、本課題開始直後の2013年5月に国立情報学研究所の評価基盤構築プロジェクト NTCIR に提案し、2014年に実施する新規評価タスクとして採択されるに至った。また、2014年12月に継続評価タスクを引き続き提案し、2016年に実施する評価タスクとして再び採択された。そのため、NTCIRでの評価タスク実施の機会を活かした研究推進を目指すように研究計画を見直し、関連研究機関で共有する研究データを構築する課題(A)と、一研究機関として評価タスクに参加し知見を得る課題(B)に研究リソースを集中して研究課題を推進することにした。

### 4.1 自由発話音声による情報要求の収集実験と分析

自由発話音声クエリの収録方法について、既存のテキストクエリを元に擬似的に自由発話を発声する状況を作り出し収録する方法、及び検索課題を自由に発想させた上でそれを自由発話で表現した音声を収録する方法、を検討し、予備実験の結果、後者で収録を行うことにした。平成25年度には21人の被験者から42クエリ、平成27年度には22人の被験者から80クエリを収録した。後者80クエリに対しては、自由に発声したクエリを収録した後に、推敲して整った表現へと編集したクエリの収録も同時に行った。これにより、即時に表現したクエリと時間をかけて熟考したクエリの性質の差を調査することが可能になった。

自由発話音声クエリと編集後のテキストクエリについて、各クエリによる検索性能の違いを分析した。各クエリを構成要素(語)に分解して、それぞれの検索結果に対する貢献度を調べ、各クエリの有する利点と欠点を明らかにした。その結果、検索性能の評価指標である Mean Average Precision(MAP)で比較すると、音声クエリの 24.4 に対しテキストクエリが 26.0 と、平均的には後者の方が検索性能が高いが、個別に比較すると前者が後者を大きく改善している場合も多く見られ、音声クエリの特徴である冗長性や多様性が効果的に機能する場合があるという知見を得た。

収録した自由発話音声クエリを用いて、国立情報学研究所の評価基盤構築プロジェクト NTICR にて、音声ドキュメント検索の評価タスクを実施することになった。評価タスクは、平成 25 年 7 月～平成 26 年 12 月の SpokenQuery&Doc タスクと、平成 27 年 1 月～平成 28 年 6 月の SpokenQuery&Doc-2 タスクの 2 ラウンドにわたり実施した。第 1 ラウンドでは 11、第 2 ラウンドでは 6 の研究グループが、最終的に検索結果を提出した。参加者からの検索結果を利用して適合性判定を行うことで正解データを作成し、再利用可能なテストコレクションを構築した。これらの成果物は、NTICR プロジェクトを通して公開する予定である。

#### 4.2 音声ドキュメント検索システムの開発

構築した評価データを用いて、自由発話音声を利用した音声ドキュメント検索手法の開発を進めた。自由発話音声に含まれる言語情報に加えて、ピッチ、パワー、話速などの音響情報を抽出し、これらの検索処理に対する貢献度を分析した。

分析によって得た知見を利用し、音声クエリから得られる音響特徴を利用した検索手法を開発した。まず、音声クエリを音声認識し、各単語セグメントからピッチ、パワー、話速などの音響情報を抽出する。抽出した単語毎の音響特徴および言語特徴から、クエリの適合文書に対する検索貢献度を推定する機械学習モデルを構築する。そして検索処理時には、機械学習モデルで推定した検索貢献度を、重みとして検索モデルに組み込んで利用する。すなわち、音声クエリ中の検索に役に立つ部分を、音響的・言語的特徴から推定して、検索処理に利用する手法を開発した。

以下の結果は、検索モデルとしてクエリ尤度モデルを利用し、検索貢献度を Support Vector Regression により推定したものである。検索貢献度を利用しない baseline に比べ、検索性能が改善されているのがわかる。また、音響特徴、言語特徴単独でも貢献度推定の効果があり、両者を併用することで最も大きな改善が得られた。

|        | ベース<br>ライン | 音響特<br>徴 | 言語特<br>徴 | 音響・言<br>語特徴 |
|--------|------------|----------|----------|-------------|
| MAP(%) | 18.2       | 18.5     | 18.8     | 19.3        |

その他、音声ドキュメント検索手法の高度化について研究を進めた。音声中の検索後検出(Spoken Term Detection:STD)については、以前より研究を進めてきた距離順 STD 手法(Metric Subspace Indexing 法)について複数検索語の同時検出に拡張した。これにより、複数の検索語について、音声ドキュメントでの出現位置を同時に、かつ尤もらしい順に、検出することが可能になった。この結果を、以前から開発を進めてきた STD を前処理に用いる SCR 手法(STD-SCR 法)に利用して、検索結果の文書を漸進的に出力する新しい音声内容検索(SCR)システムを実装した。その他、STD については、軽い処理で検出した候補集合に対し、音響尤度・言語尤度を再計算することによって、効率的に検出精度を向上させる手法を開発した。

音声内容検索(Spoken Content Retrieval:SCR)については、現在の state-of-the-art である言語モデルに基づく検索モデルに対して、音声認識の湧き出し誤りに頑健な拡張手法を開発した。これを、前述の STD-SCR 手法に適用し、検索性能の改善を行った。また、音声ドキュメントの未知語や誤認識に対処する手法として、検索対象文書間の関係をグラフ構造で表現し文書拡張する手法を、音声ドキュメントに適用する手法を新たに開発した。

講義の音声ドキュメントに対して、併用するプレゼンテーションスライドに含まれるテキスト情報を利用して、検索性能を改善する手法を開発した。音声認識結果とテキストの組み合わせ方について、頻度レベル、確率レベル、検索スコアレベルで組み合わせる手法を検討し、確率レベルで組み合わせる手法が最も効果的であることがわかった。

検索結果の音声ドキュメントの提示方法として、音声に含まれるパラ言語情報・非言語情報を記号化する手法の検討も行った。言語以外の情報として笑声と関心に注目し、これらの音声巾での出現箇所を音響特徴から機械学習手法により推定し、その書き起こしの位置に当該記号を挿入する音声書き起こしシステムの開発を行った。

#### 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 1 件)

[1] 大野哲平, 金子泰輔, 秋葉友良, “部分距離空間上の索引を用いた DTW 距離順の音声巾中の検索語検出法”, 電子情報通信学会論文誌, 査読有, Vol .J97-D, No.3, pp.548-559, 2014.

[http://search.ieice.org/bin/pdf.php?lang=J&year=2014&fname=j97-d\\_3\\_548&abst=](http://search.ieice.org/bin/pdf.php?lang=J&year=2014&fname=j97-d_3_548&abst=)

〔学会発表〕(計 34 件)

[1] Tomoyosi Akiba, “Taking advantage of spontaneous speech for document retrieval: Lessons from the organization of evaluation tasks”, 5th Joint Meeting Acoustical Society of America and Acoustical Society of Japan, 2016 年 11 月, ホノルル(アメリカ)

[2] Tomoyosi Akiba, Hiromitsu Nishizaki, Hiroaki Nanjo, Gareth Jones, “Overview of the NTCIR-11 SpokenQuery&Doc-2 Task”, In Proceedings of the 12th NTCIR Conference, pp.167-179, 2016 年 6 月, 学術総合センター(東京都)

[3] Sho Kawasaki, Hiroshi Oshima, Tomoyosi Akiba, “Graph-based Document Expansion and Robust SCR Models for False Positives: Experiments at the NTCIR-12 SpokenQuery&Doc-2”, In Proceedings of the 12th NTCIR Conference, pp.196-199, 2016 年 6 月, 学術総合センター(東京都)

[4] 川崎祥, 秋葉友良, “認識語彙に依存しない音声内容検索におけるグラフに基づく文書拡張手法”, 日本音響学会春季研究発表会, pp.109-112, 2016 年 3 月, 桐蔭横浜大学(神奈川県)

[5] 佐藤聡介, 秋葉友良, “笑声・関心の自動検出に基づく音声書き起こしへの非言語・パラ言語ラベルの自動挿入”, 日本音響学会春季研究発表会, pp.137-140, 2016 年 3 月, 桐蔭横浜大学(神奈川県)

[6] 大島翔, 秋葉友良, “音響・言語特徴を用いた単語の重要度推定に基づく自然発話クエリからの音声内容検索”, 日本音響学会秋季研究発表会, 2015 年 9 月, 会津大学(福島県)

[7] 西崎博光, 秋葉友良, “音声ドキュメントに対する情報検索の今と共通評価基盤”, 電子情報通信学会音声研究会, 115(184), pp.21-26, 2015 年 8 月, 岩手県立大学(盛岡市)

[8] Yusuke Takada, Sho Kawasaki, Hiroshi Oshima, Hiroshi Kawatani and Tomoyoshi Akiba, “STD Score Combination with Acoustic Likelihood and Robust SCR Models for False Positives: Experiments at NTCIR-11 SpokenQuery&Doc”, In Proceedings of the 11th NTCIR Conference, pp.425-428, 2014 年 12 月, 学術総合センタ

ー(東京都)

[9] Tomoyosi Akiba, Hiromitsu Nishizaki, Hiroaki Nanjo, Gareth Jones, “Overview of the NTCIR-11 SpokenQuery&Doc Task”, In Proceedings of the 11th NTCIR Conference, pp.350-364, 2014 年 12 月, 学術総合センター(東京都)

[10] 大島翔, 秋葉友良, “自由発声した情報要求に含まれるキーワードの音響・言語的特長の調査”, 日本音響学会秋季研究発表会, pp.147-150, 2014 年 9 月, 北海学園大学(札幌市)

[11] 高田優介, 秋葉友良, “音響尤度と音節列間距離の組み合わせに基づく Spoken Term Detection の精度向上”, 日本音響学会秋季研究発表会, pp.129-130, 2014 年 9 月, 北海学園大学(札幌市)

[12] 佐藤聡介, 秋葉友良, “感情ラベル付き会議録自動作成のための笑声・関心の自動検出”, 日本音響学会秋季研究発表会, pp.131-132, 2014 年 9 月, 北海学園大学(札幌市)

[13] Sho Kawasaki, Tomoyosi Akiba, “Robust Retrieval Models for False Positive Errors in Spoken Documents”, In Proceedings of International Conference on Speech Communication and Technology, pp.1757-1761, 2014 年, シンガポール(シンガポール)

[14] 大野哲平, 秋葉友良, “部分距離空間索引に基づく距離順 STD の高速化と索引サイズの削減”, 日本音響学会春季研究発表会, 3-Q5-6, pp.197-200, 2014/3, 日本大学理工学部(東京都)

[15] 川崎祥, 秋葉友良, “単語共起を利用した偽陽性に頑健な音声ドキュメントの検索モデル”, 日本音響学会春季研究発表会, 3-Q5-5, pp.193-196, 2014/3, 日本大学理工学部(東京都)

[16] 加瀬健太, 秋葉友良, “適合性と文書構造を併用した音声ドキュメント検索における適合音声区間決定”, 第 8 回音声ドキュメント処理ワークショップ講演論文集, 2014/3, 豊橋市民センター(豊橋市)

[17] 大野哲平, 秋葉友良, “距離順 STD における索引の検索効率改善と複数検索語の同時検出”, 第 8 回音声ドキュメント処理ワークショップ講演論文集, 2014/3, 豊橋市民センター(豊橋市)

[18] 秋葉友良, “音声言語処理を利用した

情報検索評価タスク: NTCIR SpokenDoc から SpokenQuery&Doc へ”, 情報処理学会研究報告, 2014-SLP-100, 2014年2月1日, ホテルサンバレー富士見(静岡県)

[19] Tomoko Takigami, Tomoyosi Akiba, “Open Vocabulary Spoken Content Retrieval by Front-Ending with Spoken Term Detection”, In Proceedings of Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2013年10月, 高雄(台湾)

[20] 秋葉友良, 西崎博光, 南條浩輝, Gareth Jones, “SpokenQuery&Doc: 自由発話音声クエリからの情報アクセス”, 第5回インタラクティブ情報アクセスと可視化マイニング研究会, 2013年10月, 慶應大学日吉キャンパス(神奈川県)

[21] 加瀬健太, 秋葉友良, “音声ドキュメント検索における適合性と文書構造を併用した適合音声区間の自動決定”, 日本音響学会秋季研究発表会, 2-P-22, pp.203-206, 2013年9月, 豊橋技術科学大学(豊橋市)

[22] Teppei Ohno, Tomoyosi Akiba, “DTW-Distance-Ordered Spoken Term Detection”, In Proceedings of International Conference on Speech Communication and Technology, pp.3737-3741, 2013年8月, リヨン(フランス)

〔図書〕(計 1 件)

日本音響学会編, コロナ社, “音響キーワードブック”, 2016, pp.112-113, 秋葉友良, 「音声ドキュメント検索」

〔その他〕

NTCIR-11 Core Task: "Spoken Query and Spoken Document Retrieval (SpokenQuery&Doc)"

<http://www.cl.ics.tut.ac.jp/~sdpwg/index.php?ntcir11>

NTCIR-12 Core Task: "Spoken Query and Spoken Document Retrieval 2 (SpokenQuery&Doc-2)"

<http://www.cl.ics.tut.ac.jp/~sdpwg/index.php?ntcir12>

## 6. 研究組織

### (1) 研究代表者

秋葉 友良 (AKIBA TOMOYOSI)  
豊橋技術科学大学・工学研究科・准教授  
研究者番号: 00356346

### (2) 研究分担者

中川 聖一 (NAKAGAWA SEIICHI)  
豊橋技術科学大学・工学研究科・教授  
研究者番号: 20115893