

SpokenQuery&Doc: 自由発話音声クエリからの情報アクセス

SpokenQuery&Doc: Information Access from Spontaneously Spoken Query

秋葉友良^{1*} 西崎博光²
南條浩輝³ Gareth Jones⁴

¹ 豊橋技術科学大学
¹ Toyohashi University of Technology
² 山梨大学
² Yamanashi University
³ 龍谷大学
³ Ryukoku University
⁴ Dublin City University

Abstract: This paper introduces the SpokenQuery&Doc task, which will be conducted in the next NTCIR evaluation. The SpokenQuery&Doc is a successor to the previous SpokenDoc and SpokenDoc-2 tasks evaluated at the past NTCIR workshops, which are also presented in this paper.

1 はじめに

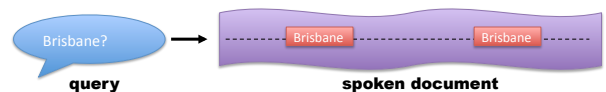
国立情報学研究所が主催する情報アクセス技術の評価型ワークショップ NTCIR(NII Testbeds and Communities for Information access Research) では、これまで様々な情報アクセスタスクの評価が行われてきたが、2011 年の NTCIR-9[7] より音声ドキュメント検索タスク SpokenDoc[1] が評価タスクとして採択され、最新の NTCIR-10[6] では 2 回目の評価タスク SpokenDoc-2[3] が実施された。

NTCIR-11 では、SpokenDoc の後継タスクとして SpokenQuery&Doc タスクを実施する。SpokenQuery&Doc では、これまでテキストで与えた検索クエリを自由発話音声で与えることに評価の焦点を当てる。一方、これまでのテキスト入力 STD および SCR タスクもサブタスクとして継続する。

2 NTCIR-9 SpokenDoc と NTCIR-10 SpokenDoc-2

多くの情報検索タスクが対象とするテキストと同様に、ラジオやテレビなどの放送や動画データなどに付随する音声にも豊富な言語情報が含まれている。動画配信サイト等を通して、音声を含むコンテンツは増加

• 音声検索語検出(Spoken Term Detection)



• 音声内容検索(Spoken Content Retrieval)



図 1: STD タスクと SCR タスク

の一途を辿りつつあるが、その言語情報へのアクセスはテキストのように容易ではない。音声データを、その文書としての側面に注目して「音声ドキュメント」と呼び、音声データを対象とした検索を「音声ドキュメント検索」[9] と呼ぶ。

第 9 回 NTCIR で実施した “IR for Spoken Documents (SpokenDoc)” タスク [1, 2] は、NTCIR 初の音声ドキュメント検索タスクであり、講演音声を対象とした 2 つの音声ドキュメント検索タスクが設定された (図 1)。

Spoken Term Detection (STD) 語をクエリとして与え、音声ドキュメント中からクエリが現れる位置を特定するタスク。計算効率 (索引に必要な空

*連絡先: 豊橋技術科学大学 情報・知能工学系
〒441-8580 愛知県豊橋市天伯町雲雀ヶ丘 1-1
E-mail: akiba@cs.tut.ac.jp

間コスト、検索時間コスト、など)と検索性能(精度と再現率)の2つの観点から評価を行った。

Spoken Document Retrieval (SDR) 質問文で表現した比較的長いクエリを与え、クエリと関連する音声セグメントを見つけるタスク。テキストを対象とした検索における内容検索に相当する。音声セグメントとして、講演全体(講演検索タスク)と、講演中の数秒程度の音声区間(パッセージ検索タスク)、の2種類の粒度を設定した。

SCRにおける講演検索タスクは、テキストを対象とした検索における文書検索に相当する。しかし、音声ドキュメント検索では講演のような大きな単位が検索されたとしても、検索結果を確認するためには音声の再生が必要となり、テキストのように全体をざっと一覧することができない。したがって、よりピンポイントに検索の適箇所(音声区間)を見つける技術が必要となる。これがパッセージ検索タスクを設定した理由である。

続く NTCIR-10 で実施した SpokenDoc-2 タスク [3, 8, 10] では、上の2つのタスクに加えてクエリが音声ドキュメント中に出現しないことを確認する iSTD タスクの新規設定や、より低い認識率の環境下での音声ドキュメント検索の評価を行った。

音声ドキュメント検索の性能は音声認識の精度に依存するため、タスクオーガナイズはタスク参加グループ共通で利用できる音声認識結果を用意した。これにより、参加グループの検索結果を共通の土台の上で比較することが可能になるとともに、音声認識システムを保有しない参加グループや、音声認識よりも検索手法に興味を持つ参加グループに対し、参加を容易にする環境を提供した。

3 NTCIR-11 SpokenQuery&Doc

SpokenDoc および SpokenDoc-2 では、テキストで与えたクエリから、音声データを対象とした情報検索タスクの評価を行った。NTCIR-11 SpokenQuery&Doc では、音声を扱う検索のもう一つの側面として、クエリを音声で与える検索にも焦点を当てる。

タスク設定の背景には、テキスト入力による情報検索の限界を音声入力によって克服したいという展望がある [4]。人はしばしば複雑な情報要求を抱くが、現状の情報検索システムでは入力インタフェース(テキスト入力)の制限によりユーザの情報要求を十分に汲み取ることができない。一方、音声入力には、ユーザへの負荷が少ない、入力速度が速い、思考を即時に表現できる、といった利点があり、ユーザが音声で思い付くまま情報要求を表現することで、これをそのままシス

今年の夏休みに始めて山に登山に行くことになったんですけども、あの登山は結構事故があつたや
はり夏になるとよくニュースで聞きますし、まあ
誰々が行方不明になったとか遭難したとかそ
ういう話があると思うんですけども、あの
山に登るときにはまあどういった心構えとい
うか、あの装備こういう装備が必要だとかこ
ういうものがあるといいよとかそういったあ
の山登りに関しての、山に登るときについ
ての心構えについて知りたいです。

図 3: 自由発話音声クエリの例

テムに入力する手段を提供する。一方、ユーザがその場で制限無く自由に発話した音声(自発音声)は、あらかじめ発話内容を調整してから発話した読み上げ音声に比べて、話し言葉の特徴(間投詞の頻出などの非流暢性、発話内容の非文法性、など)が多く現れるため、音声認識が難しい。図3に、予備実験 [11] で収集した自由発話音声クエリを示す。

NTCIR-11 SpokenQuery&Doc では、以下の3つのタスクを設定する。(図2)

SQ-SCR 音声クエリからの音声内容検索 (Spoken-query-driven Spoken Content Retrieval) タスク。SpokenQuery&Doc のメインタスクである。自由発話音声で表現した比較的長いクエリをシステムへの入力として、音声内容検索 (SCR) を行うタスク。このタスク実現のためのサブタスクとして、次の2つのタスクを実施する。

SQ-STD 音声クエリからの音声検索語検出 (Spoken-query-driven Spoken Term Detection) タスク。自由発話音声クエリ中に現れる語の音声区間を入力として、音声検索語検出 (SCR) を行うタスク。音声クエリから語の音声区間の切り出し結果は、タスクオーガナイズから提供する。

STD-SCR 音声検索語検出結果からの音声内容検索 (Spoken Content Retrieval using STD results) タスク。SQ-STD タスク参加者による検出結果を入力として、音声内容検索 (SCR) を行うタスク。

SQ-SCR および SQ-STD タスクでは、音声クエリに加えて、人手で書き起したテキストクエリも提供する。タスク参加者は、このテキストクエリを使って結果を提出してもよい。すなわち、SpokenDoc、SpokenDoc-2 と同様の STD タスクおよび SCR タスクも実施する。また、音声から音声を検索するタスクでは、必ずしも音声認識を用いたテキスト表現を介した検索アプロー

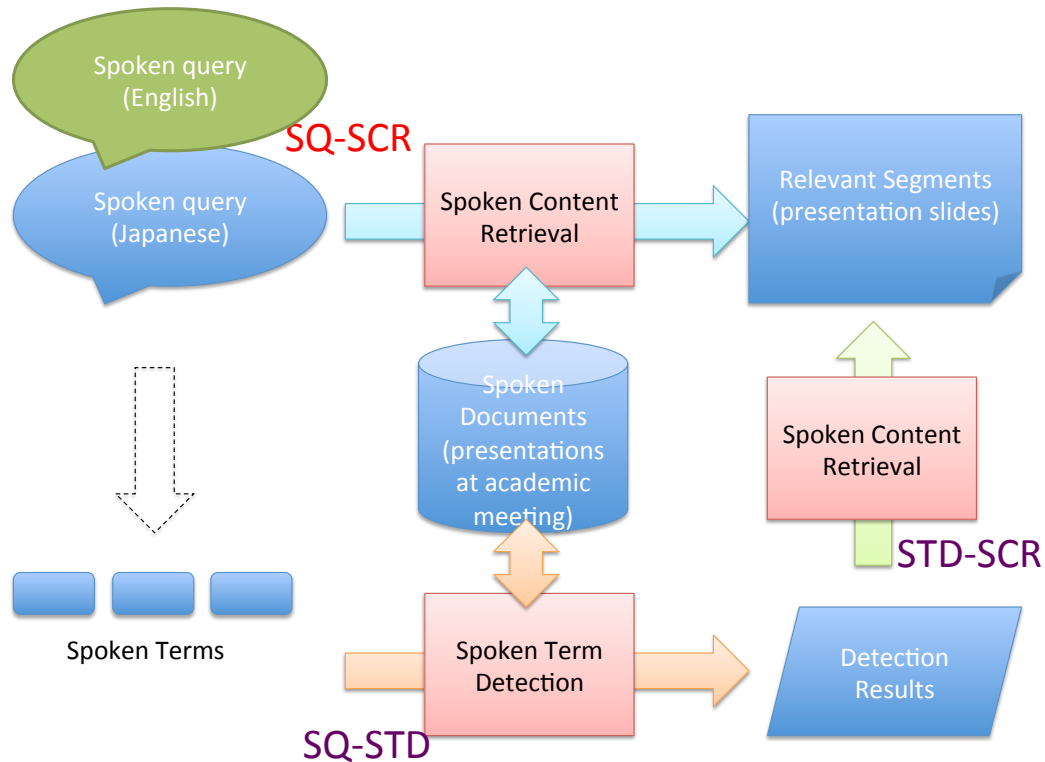


図 2: SpokenQuery&Doc で実施する 3 つのタスク

チを採用する必要はなく、音声を直接比較することも可能である [5]。タスクオーガナイザは、このようなアプローチによるタスク参加も期待している。

4 むすび

本稿では、過去 2 回の NTCIR で実施した SpokenDoc タスクと、次回 NTCIR-11 で実施する SpokenQuery&Doc タスクを紹介した。SpokenQuery&Doc のより詳しい情報は、以下の Web ページを参照されたい。

<http://www.nlp.cs.tut.ac.jp/ntcir11/>

参考文献

- [1] T. Akiba, et al. Overview of the IR for spoken documents task in NTCIR-9 workshop. In *Proceedings of The Ninth NTCIR Workshop Meeting*, pp. 223–235, 2011.
- [2] T. Akiba, et al. Designing an evaluation framework for spoken term detection and spoken document retrieval at the NTCIR-9 SpokenDoc task. In *Proceedings of International Conference on Language Resources and Evaluation*, 2012.
- [3] T. Akiba, et al. Overview of the NTCIR-10 SpokenDoc-2 task. In *Proceedings of The 10th NTCIR Conference*, pp. 573–587, 2013.
- [4] T. Akiba, A. Fujii, and K. Itou. Collecting spontaneously spoken queries for information retrieval. In *Proceedings of International Conference on Language Resources and Evaluation*, pp. 1439–1442, 2004.
- [5] T. K. Chia, K. C. Sim, H. Li, and H. T. Ng. A lattice-based approach to query-by-example spoken document retrieval. In *Proceedings of Annual International ACM SIGIR Conference on Research and development in information retrieval*, pp. 363–370, 2008.

- [6] H. Joho and T. Sakai. Overview of NTCIR-10. In *Proceedings of The 10th NTCIR Conference*, pp. 1–7, 2013.
- [7] T. Sakai and H. Joho. Overview of NTCIR-9. In *Proceedings of The Ninth NTCIR Workshop Meeting*, pp. 1–7, 2011.
- [8] 西崎, 秋葉, 相川, 伊藤, 河原, 胡, 中川, 南條, 山下. Ntcir-10 spokendoc-2 spoken term detection タスクの結果と知見. 日本音響学会秋季研究発表会講演論文集, pp. 107–110, 2013.
- [9] 秋葉. 音声ドキュメント検索: マルチメディアデータを対象とした音声言語情報検索. 情報の科学と技術, 63(1):21–27, 2013.
- [10] 秋葉, 西崎, 相川, 伊藤, 河原, 胡, 中川, 南條, 山下. Ntcir-10 spokendoc-2 spoken content retrieval タスクの結果と知見. 日本音響学会秋季研究発表会講演論文集, pp. 111–114, 2013.
- [11] 大島, 秋葉. 自発音声による情報要求の収集実験とその検索性能評価. 日本音響学会春季研究発表会講演論文集, pp. 233–234, 2013.