

音声検索語検出を前処理に用いた 未知語や認識誤りに頑健な音声ドキュメント検索

瀧上 智子^{1,a)} 秋葉 友良^{1,b)}

受付日 2012年5月31日, 採録日 2012年11月2日

概要: 音声ドキュメント処理において, 未知語や音声認識誤りへの対処は不可欠である. 本研究では, 音声ドキュメント検索に対し, 未知語・認識誤りへの対策が比較的検討されている音声検索語検出の手法を前処理として取り入れる手法を提案する. 提案手法はまず, 連続音節認識結果に対し, クエリを構成する単語を検索語とした音声検索語検出を行い, 各語の出現情報を取得し, 次にこの出現情報を用いて各文書とクエリとの関連度を計算し, 文書検索を行う. 提案手法の有効性を調べるために, 講演音声を対象とした音声ドキュメント検索テストコレクションを用いて, 大語彙連続音声認識結果に対してそのままテキストベースの検索手法を適用する従来法との比較実験を行った. 提案法は, クエリに未知語が含まれる場合でも, 検索性能を落とすことなく頑健に検索できることが分かった. 特に, 音声ドキュメントの単語正解率が60%程度と低い場合では, 未知語を含むクエリの性能を有意に改善した. さらに, 従来法と提案法を相補的に統合することにより, 検索性能がより改善されることが分かった.

キーワード: 音声ドキュメント処理, 音声ドキュメント検索, 音声検索語検出

Front-ending Spoken Document Retrieval with Spoken Term Detection Robust for OOV and Missrecognized Words

TOMOKO TAKIGAMI^{1,a)} TOMOYOSI AKIBA^{1,b)}

Received: May 31, 2012, Accepted: November 2, 2012

Abstract: How to deal with speech recognition errors and out-of-vocabulary (OOV) words is one of the challenging problems in spoken document processing. To deal with the problem in spoken document retrieval (SDR), we propose the method that incorporates spoken term detection (STD) into the spoken document retrieval process. The proposed method firstly performs STDs for each term appeared in a given query topic, and then the detection results are used to calculate the relevancy of the retrieved document against the topic. To investigate the effectiveness of the proposed method, we conducted the experimental evaluation using the SDR test collection targeting Japanese lecture speech by comparing with the conventional SDR method using LVCSR and text-based document retrieval. The result showed that the proposed method did not degrade the retrieval performance even if the query included OOV terms. It also showed that it improved the retrieval performance of OOV queries significantly when word correct rate of the spoken document was relatively low (around 60%). Furthermore, the integration of the proposed and the conventional methods further improved the retrieval performance.

Keywords: spoken document processing, spoken document retrieval, spoken term detection

1. はじめに

近年, 音声や動画の記録・編集機器の拡大や, 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, マルチメディアコンテンツの情報爆発が進行しつつある. それ

¹ 豊橋技術科学大学
Toyohashi University of Technology
^{a)} takigami@nlp.cs.tut.ac.jp
^{b)} akiba@nlp.cs.tut.ac.jp

にともない、これらのマルチメディアコンテンツに対する検索技術は必要不可欠となっている。しかしこれらのコンテンツには、ファイル名やタイトル以外にはメタデータが付与されていないことが多く、従来のテキストベースの検索技術だけでは、目的のコンテンツにたどり着くことは困難である。一方、音声を含むコンテンツの場合には、大語彙連続音声認識技術を利用することで、言語情報を利用した検索が可能となっている。この音声言語情報を対象とした検索技術は「音声ドキュメント検索 (Spoken Document Retrieval; SDR)」と呼ばれ、マルチメディアコンテンツの情報爆発時代に必要不可欠な技術である [1]。テキスト検索と異なり、音声ドキュメント検索では音声認識での認識誤りや、大語彙連続音声認識の認識語彙外語 (未知語) への対処が不可欠であり、これらの問題が音声ドキュメント検索における最も大きな課題である。

従来の音声ドキュメント検索では、音声認識結果の自動書き起こしテキストに対し、そのままテキストベースの文書検索技術を用いる手法が一般的である。しかしこの手法は音声認識の性能に大きく影響を受ける。認識を誤った語や未知語は、検索の手がかりとして利用できないため検索性能が低下するのである。特に、固有名詞は検索において重要な手がかりとなる語であるが、未知語となりやすい語であり、正しく認識されないことが多い。また、認識誤りを改善するためには、音声認識の適応を行う必要があるが、音響モデルや言語モデルの学習のために同一ドメインのリソースを入手する必要がある、高価である。特に、日々増え続ける大量のコンテンツを扱うためには、検索に備えた音声ドキュメントの事前処理での効率化も求められることから、個々のドキュメントにあわせた音声認識の適応は難しい。

本研究では、この問題に対処するために、未知語・認識誤りに対する様々な対策が検討されている音声検索語検出 (Spoken Term Detection; STD) の手法を利用した SDR 手法を提案する。STD は誤りを含む音声ドキュメント中から、任意の検索語を検出するタスクである。認識誤りや未知語があっても、目的の単語を検出するための様々な手法が研究されている。提案法は、この STD を行うことにより未知語や認識を誤った語を文書検索の手がかりとして利用する。音声ドキュメント検索のための処理は、検索に備えた検索対象ドキュメントの事前処理 (オフライン処理) と、クエリが与えられた後の文書検索処理 (オンライン処理) に二分される。オフライン処理には、音声認識や索引付け処理などがあり、オンライン処理には、質問解析や、文書検索、関連度計算などの処理がある。提案法は、オフライン処理として連続音節認識結果、または大語彙連続音声認識結果を音節に変換することなどから音声ドキュメントに対する音節列を得る。続いて、オンライン処理における文書検索の前処理として、クエリを構成する各単語につ

いて音声ドキュメント中に現れる位置を STD の手法を用いて特定し、その結果である単語出現情報を利用して文書検索を行う。STD を用いることにより、未知語や認識誤りに頑健な検索が可能となる。

本研究では SDR システム内で用いる STD 手法として、一般的な手法である連続 DP マッチングと、音節 bi-gram 索引を用いた 2 種類の高速化手法、の合計 3 種類の手法を採用した。連続 DP マッチングは STD 手法として最も一般的であるが、検出に時間がかかるという問題がある。SDR では 1 クエリあたりに複数の検索語が含まれており、それぞれの語で連続 DP マッチングを行うと、検索時間の増加が問題となる。それに加えて、クエリ拡張などの SDR の性能改善のための手法を行うことが難しいという問題も生じてしまう。これに対処するために、音節 bi-gram 索引を用いた STD の高速化を行うことで、検索性能を低下させることなく、検索時間を改善することを目指した。

提案法は未知語を利用して検索性能を向上させることができるが、もともと既知語で正しく認識されていた語に関しては湧き出し誤りによって検索に悪影響を及ぼす場合がある。一方で従来法は未知語を利用することはできないが、既知語に関しては有効に活用でき、提案法と相補的な関係にある。そこで、従来法と提案法を組み合わせることで、さらなる検索性能の向上を目指した。音声認識には大語彙連続音声認識と連続音節認識の 2 種類を用い、2 種類の索引を事前に作成する。与えられたクエリに対し、従来法、提案法のそれぞれの方法で各文書の文書ベクトルを作成し、関連度のスコアを線形に結合して、最終的なスコアによって文書のランク付けを行う。

提案法の性能を確かめるため、CSJ を対象として、検索対象のドキュメントに対して音声認識の適応を行った場合と、適応を行わなかった場合との 2 つの条件の下で検索実験を行った。音声認識の適応を行った場合には、提案法は全体での検索性能は従来法に及ばなかったものの、未知語を含むクエリにおいては、提案法は従来法を上回る検索性能を示した。一方、適応を行わなかった場合には、従来法は著しく性能が低下したのに対し、提案法は適応を行った場合と同等の性能を示した。また、従来法と提案法を混合したシステムは、最も高い検索性能を示した。

本論文の構成は以下のとおりである。2 章で関連研究について説明する。3 章では音声ドキュメント検索の一般的な手法について述べ、4 章で提案法について、5 章で従来法と提案法を混合する手法について説明する。6 章で実験の条件と結果を示し、最後に 7 章でまとめを述べる。

2. 関連研究

音声ドキュメントを対象とした検索のうち、入力した検索語 (クエリ、パターンなどと呼ぶ) が音声ドキュメント中で現れる位置を特定する問題は、音声中の検索語検出

(Spoken Term Detection; STD) と呼ばれ、音声情報処理の分野では活発に研究が行われている問題である。1997年には、米国 NIST 主催の評価型ワークショップ TREC の音声ドキュメント検索トラック (SDR Track) [2] において、参加機関の音声ドキュメント検索手法の評価が行われた。TREC SDR Track で最初に設定されたタスクは、クエリに含まれる単語の出現を見つけ出すタスクで、現在の STD に近いタスクである。また、2006年に NIST は再び Spoken Term Detection を研究課題として設定し [3]、それ以降未知語の検出を重視した STD の研究がさかんに行われるようになった。日本においても、情報処理学会音声言語処理研究会のワーキンググループにおいて、STD の評価用テストコレクションが整備され [4]、2012年には評価型ワークショップ NTCIR の SpokenDoc タスクのサブタスクとして参加グループの手法の評価が行われた [1]。

STD は “Known Item Retrieval” と呼ばれ、検索者が検索の対象 (用語) をすでに知っている状況 (ナビゲーション的な質問) を想定したタスクである。しかし人が検索を行う実際の場面では、知りたい事項に対して漠然としたイメージしか持っていない場合は多く、人の曖昧な情報要求 (インフォーマショナルな質問) から関連情報を見つける技術が必要とされる。このような状況における検索タスクは「内容検索」と呼ばれ、知りたい内容を表現した文やキーワードリストなどのクエリから、その内容を含む未知の文書を見つけることを目的とする。STD に対して、音声ドキュメントに対する内容検索は SDR (Spoken Document Retrieval) と呼ばれる。前述の TREC の音声ドキュメント検索トラック (SDR Track) においても、2年目以降は内容検索を対象に参加グループの SDR 手法の評価が行われた。日本においても、STD と同様に情報処理学会音声言語処理研究会のワーキンググループにおいて、SDR 評価用テストコレクションが公開され [5]、NTCIR の SpokenDoc タスクのサブタスクとして参加グループの各 SDR 手法の評価が行われている [1]。

STD の研究では、認識結果・検索単位へのサブワードの利用や、認識結果の複数候補の利用、誤りを許した一致判定法 (Soft Match) など、未知語や認識誤りへの対応手法が数多く提案されている [6]。SDR の研究では、単語単位の音声認識結果に対し、そのままテキストベースの文書検索手法を適用する手法が一般的で、STD に比べて未知語・認識誤りに直接対策を行った研究は少ない。SDR では文書中の全単語が手がかりとなり、少量単語に認識誤りがあっても比較的頑健な動作が期待されるためである。実際、高認識率下では正解書き起こしに対する検索と遜色ない性能を示し、SDR は「解決した」問題とする結論が出されている [2]。しかし、実際の利用場面では高認識率を期待できないデータを扱う必要も想定される。また、多様なデータのそれぞれについて十分な適応処理を行うことは

コストの面から困難である。さらに、文書検索の大きな手がかりとなる固有名詞は未知語となりやすいことから、特定のクエリでは検索性能が極端に低下してしまうことになる。このようなことから、SDR においても、未知語・認識誤りへの直接的な対策手法の適用が重要になると考えられる。認識誤りや未知語の問題を解決する方法の1つは、文書拡張あるいはクエリ拡張を行うことである。音声ドキュメント検索に対して、Web ページを利用してドキュメント拡張を行う手法 [7], [8] や、疑似適合性フィードバックを利用する手法 [9] が提案されている。これらの手法は、正しく認識された語を手がかりに関連語を拡張することで認識誤りを間接的に補完しようとする対処法であり、実際にはドキュメント中の未知語や認識誤りの生じた区間は利用していない。また、音声認識誤りに対して頑健な検索を行うため、複数認識候補を用いて生成した単語ラティスを対象に検索する方法が提案されている [10]。しかし、この手法では複数認識候補に出現しない語は利用することができないため、未知語には対処できないという問題がある。一方、未知語や認識誤りの音声区間も利用するために、サブワードの認識結果を検索の手がかりとして利用する手法が提案されている [11], [12], [13], [14]。単語を手がかりとして用いる一般的な文書検索に対し、文献 [11], [13], [14] では固定長でオーバーラップありの音素 n-gram を、文献 [14] では教師なし学習で獲得した可変長でオーバーラップなしの音素 multigram を検索時の手がかりとして利用する*1。音素 n-gram や音素 multigram は、サブワード音素認識結果から言語知識を用いずに自動的に抽出した単位であるため、単語に比べると検索時の手がかりとしては有効性が限られると考えられる。実際、これらの手法では大語彙連続音声認識で得た単語と併用することで検索性能の向上を試みている。一方、本研究で提案する手法は、同様にサブワードの認識結果を用いているものの、まず STD の適用によって単語の出現を検出した後に、検出した単語を手がかりとして文書検索を行う点が異なる。

本研究の提案手法と類似した手法として、音声ドキュメントに対してワードスポッティングを行った結果を利用する SDR 初期の研究があげられる [15], [16]。これらは、高速性が要求される検索処理に対して、ワードスポッティングの処理速度に問題点があるため、比較的小規模なデータにしか適用されておらず、その後現在まで適用例は報告されていない。本研究では、STD 研究の進展により様々な高速・高精度な STD 手法が開発されていることを利用して、大規模なデータを対象とした音声ドキュメント検索への適用を検討した。

その他、未知語の問題に直接対策を行った手法として、統計的機械翻訳の技術を用いて、認識誤りを含む文を誤り

*1 文献 [14] では、音素 n-gram の方が音素 multigram よりも効果的であったと報告している。

のない文へと翻訳する手法が提案されている [17]. しかしこの手法には, 翻訳モデルの学習のために認識誤りを含む文と誤りのない文との対応付けを行ったパラレルコーパスが大量に必要となる.

3. 一般的な音声ドキュメントの内容検索手法

音声ドキュメントの検索の一般的なシステムは, 音声認識結果の自動書き起こしテキストに対して, そのままテキストベースの検索手法を適用するものである. システム図を図 1 に示す. 一般的な音声ドキュメント検索システムでは以下のような処理を行っている.

(1) オフライン処理

音声ドキュメントに対し大語彙連続音声認識を適用し, 自動書き起こしテキストを得る. 得られたテキストに対し形態素解析およびストップワード除去を適用して, 単語集合へと変換する. 同時に処理の高速化のため, 索引付けを行う.

(2) 質問解析

クエリに形態素解析, およびストップワード除去を適用して単語集合に変換する.

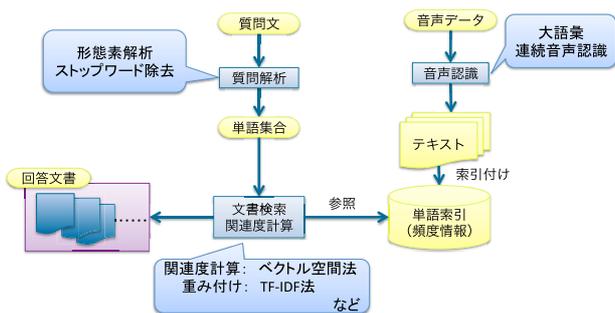


図 1 一般的な SDR システム

Fig. 1 Conventional SDR system.

(3) 文書検索

検索語の文書中での出現頻度などの統計情報に基づいて, 各文書とクエリとの関連度を計算し, 関連度の高い文書から検索結果として出力する. 関連度の計算としては, TF-IDF 重み付けを用いたベクトル空間法が広く用いられている.

ベクトル空間法は, クエリと文書を同じ索引語の重みベクトルで表現することで, 文書がどれだけクエリに適しているかを, ベクトル間の類似度の問題として扱うモデルである. ベクトル間の類似度の計算には様々な尺度があるが, 本研究では内積を用い, 以下の式で求める.

$$sim_{cSDR}(q, d) = \frac{1}{len(d)} \sum_i^{|q|} w_{q_i} w_{d_i} \quad (1)$$

ここで, w_{q_i} , w_{d_i} は i 番目の索引語の重みを表し, $len(d)$ は文書 d の長さ (単語数) である.

4. STD に基づく音声ドキュメントの検索手法

従来の音声ドキュメント検索の手法は, 認識結果をそのまま利用するため, 未知語や認識誤りの語を利用できないという問題がある. 検索クエリに未知語や認識誤りの語が含まれている場合でも, それらを文書検索の手がかりとして利用するため, オンラインの文書検索の前処理として STD を適用する手法を検討した. 提案システムを図 2 に示す. 基本的なシステムは次のとおりである.

(1) オフライン処理

音声ドキュメントに対し大語彙連続音声認識, あるいはサブワードの連続音声認識を適用し, サブワード系列からなる自動書き起こしテキストを得る. また, ステップ (3) で高速化する場合に索引付けを行う.

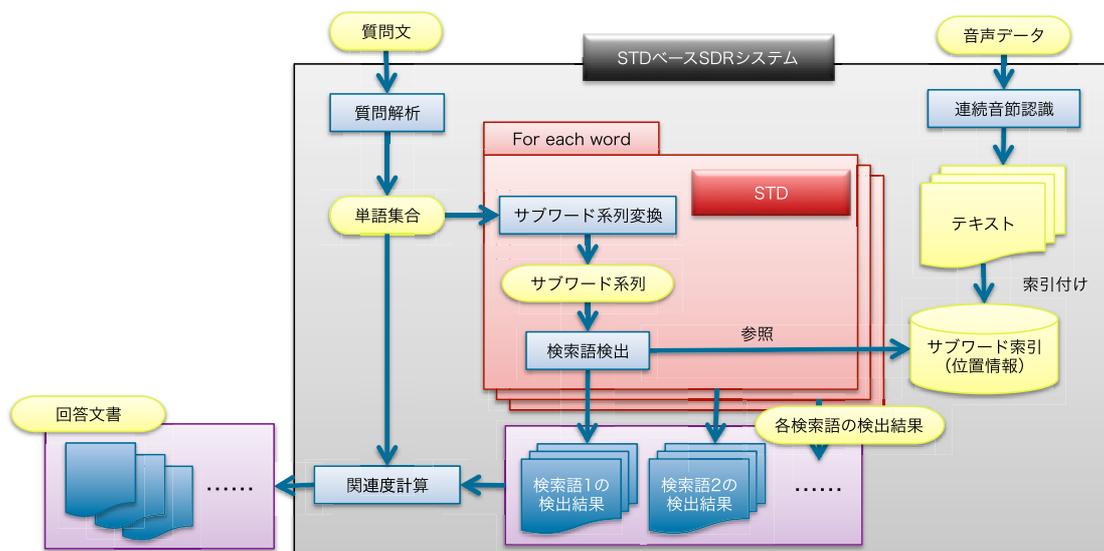


図 2 STD ベース SDR システム

Fig. 2 STD based SDR system.

(2) 質問解析

クエリに形態素解析, およびストップワード除去を適用して単語集合に変換する. そして, 各単語のサブワード列表現を得る.

(3) 検索語検出

クエリ中の各語について, 自動書き起こしテキストを対象に STD を行い, 各語の検出結果を得る.

(4) 文書検索

各語の検出結果を語の出現とし, 文書中での出現頻度などの統計情報を得る. これに基づいて各文書とクエリとの関連度を計算し, 関連度の高い文書から検索結果として出力する.

本研究ではサブワードとして音節を選択し, 音声ドキュメントに対し連続音節認識を行った. STD 手法には最も基本的な手法である連続 DP マッチングと, 4.1.2, 4.1.3 項で述べる索引付けを用いた手法とを実装した. STD によって検出された語の出現をもとに各文書の文書ベクトルを作成し, 従来法と同様にベクトル空間法を用いてクエリとの関連度を計算する.

$$sim_{STD-SDR}(q, d) = \frac{1}{len(d)} \sum_i^{|q|} w_{q_i} w_{d_i} \quad (2)$$

STD では “すべての単語数” を求めることが困難なため, $len(d)$ には各文書の総音節数を用いる.

4.1 STD 手法

4.1.1 連続 DP マッチング

STD の最も基本的な手法である連続 DP マッチングでは, 距離行列の計算に次の式を用いた.

$$\begin{aligned} D_{0,j} &= 0 \quad (0 \leq j \leq J) \\ D_{i,0} &= \infty \quad (1 \leq i \leq I) \\ D_{i,j} &= \min\{D_{i,j-1}, D_{i-1,j-1}, D_{i-1,j}\} + d(a_i, b_j) \quad (3) \end{aligned}$$

ここで i は検索語の位置, j は検索対象の位置を表し, I は検索語長, J は検索ドキュメント長を表す. $d(a_i, b_j)$ は音節 a_i, b_j 間の距離, $D_{i,j}$ は i, j における累積距離である. そして累積距離 $D_{I,J}$ を検索語長 I で正規化し, その値が閾値以下である場合を語の検出とした. 本論文では, 音節間距離 $d(a_i, b_j)$ として文献 [18] で用いられた音響モデル間の Bhattacharyya 距離を用いた.

4.1.2 索引付け + 連続 DP マッチング

連続 DP マッチングは検索に時間がかかるという問題がある. SDR では 1 クエリあたりに複数の検索語が含まれているため, すべての音声ドキュメントに対して連続 DP マッチングを行うと, 検索語数に比例して検索時間が増大してしまう. また, さらに検索語が増加するクエリ拡張などを行うことが難しいという問題もある.

そこで, 見込みのない発話については DP マッチングを

行わず, スキップすることで高速化を図る. まず, オフライン処理において音声ドキュメントから音節 bi-gram 索引を作成する. 音節 t で構成されるあるテキスト $T = t_1 t_2 \dots t_n$ に対し, 音節 bi-gram は次のように作成される.

$$B_T = \{(t_i, t_{i+1}) \mid i = 1, 2, \dots, n-1\} \quad (4)$$

索引は音節 bi-gram と, ドキュメント ID・発話 ID のペアとを保持している.

検索時には音節列で構成された検索語 $w = w_1 w_2 \dots w_m$ に対して, 検索語の音節 bi-gram B_w を作成する.

$$B_w = \{(w_i, w_{i+1}) \mid i = 1, 2, \dots, m-1\} \quad (5)$$

そして検索語の音節 bi-gram 集合 B_w と, ある発話 u の音節 bi-gram 集合 B_u について以下の式が成立する発話に対してのみ連続 DP マッチングを行った. θ_1 はあらかじめ定めた閾値である. 今回 θ_1 は DP マッチングの閾値によらず一定であるとした.

$$\frac{\sum_{b \in B_w} \delta(b \in B_u)}{|B_w|} \geq \theta_1 \quad (6)$$

$$\delta(x \in X) = \begin{cases} 1 & (x \in X) \\ 0 & (\text{otherwise}) \end{cases} \quad (7)$$

4.1.3 索引のみを用いた STD

4.1.2 項では索引を用いることにより, DP マッチングの実行回数を減らすことで検索速度の向上を図った. しかしながら, 依然 DP マッチングを行っているため, 従来法に比べると検索に時間がかかってしまう. そこで, さらに高速な検索を行うため, DP マッチングを行わない索引付けのみを用いた STD システムを構築した. しかし, 音節 bi-gram 索引のみでは, 検索時の手がかりが少ないため, 索引には音節 bi-gram と, 間の 1 音節をスキップしたスキップ bi-gram を用いる. スキップ bi-gram はテキスト $T = t_1 t_2 \dots t_n$ から以下のように作成される.

$$SB_T = \{(t_i, t_{i+2}) \mid i = 1, 2, \dots, n-2\} \quad (8)$$

2 種類の索引はマージして, 1 つの索引 ($I_T = B_T \cup SB_T$) として扱う. 検索語からも音節 bi-gram, スキップ bi-gram を作成し, 検索 bi-gram 集合 I_w を作成する. そして, ある発話 u の bi-gram 集合 I_u に対し, 以下の式が成立するとき, その発話において語が出現した (Detect) とする. ここで θ_2 はあらかじめ定めた閾値である.

$$\frac{\sum_{b \in I_w} \delta(b \in I_u)}{|I_w|} \geq \theta_2 \quad (9)$$

$$\delta(x \in X) = \begin{cases} 1 & (x \in X) \\ 0 & (\text{otherwise}) \end{cases} \quad (10)$$

表 1 音節 bi-gram とスキップ bi-gram を併用した索引の例
Table 1 Example of indices using syllable bi-grams and skip bi-grams.

	音節 bi-gram	スキップ bi-gram	音節 bi-gram & スキップ bi-gram
検索語 o N se N	o-N, N-se, se-N	o-se, N-N	o-N, N-se, se-N o-se, N-N
置換誤り o N sa N	o-N, <u>N-sa</u> , <u>sa-N</u>	<u>o-sa</u> , N-N	o-N, <u>N-sa</u> , <u>sa-N</u> <u>o-sa</u> , N-N
挿入誤り o N se a N	o-N, N-se, <u>se-a</u> , <u>a-N</u>	o-se, <u>N-a</u> , <u>se-N</u>	o-N, N-se, <u>se-a</u> , <u>a-N</u> o-se, <u>N-a</u> , <u>se-N</u>
脱落誤り o _ se N	<u>o-se</u> , se-N	<u>o-N</u>	o-se, se-N o-N

音節 bi-gram とスキップ bi-gram とを併用することによって、挿入誤り、脱落誤りに対してより頑健な検索が可能となる。表 1 に索引の例を示す。ここで、下線部は認識誤りを含む索引を表している。認識誤りによって、生成される bi-gram が変化するため、音節 bi-gram 単体の使用 (表 1, 2 列目) では、手がかりとなる bi-gram 数が減少してしまう。たとえば、正解の索引 3 つに対し、挿入誤りで手がかりとなるのは 2 索引である。特に検索語の音節数が小さい場合には、検索性能に大きく影響を与えてしまう。そこで、手がかりを増やすためにスキップ bi-gram を用いる。スキップ bi-gram を用いることで、手がかりとなる bi-gram 数を増やすことができる。たとえば、正解の 3 索引 (音節 bi-gram) と 2 索引 (スキップ bi-gram) の合計 5 索引に対し、挿入誤りではそれぞれ 2 索引と 1 索引で合計 3 索引が手がかりとして利用可能である。また、音節 bi-gram とスキップ bi-gram という 2 つの索引を区別なく扱うことで、さらに多くの bi-gram を検索時に利用することができるようになり、検索の頑健性が向上すると考えられる。たとえば、正解の合計 5 索引に対し、挿入誤りでは合計 4 索引が手がかりとなる。

5. 従来法と提案法との混合

提案法は未知語や認識誤りの語については、効果的に検索を行うことができるが、正しく認識された語に対しては湧き出し誤りによる悪影響が生じるという問題がある。一方で、従来法は正しく認識された語は有効に利用することができ、提案法とは相補的な関係にある。そこで、従来法と提案法の混合システムを構築し検索性能の改善を目指す。混合システムのシステム図を図 3 に示す。システムの処理は以下のとおりである。

(1) オフライン処理

(a) 単語索引の作成

音声ドキュメントに対し大語彙連続音声認識を適用し、自動書き起こしテキストを得る。得られたテキストに対し形態素解析およびストップワード除去を適用して、単語集合へと変換する。同時に

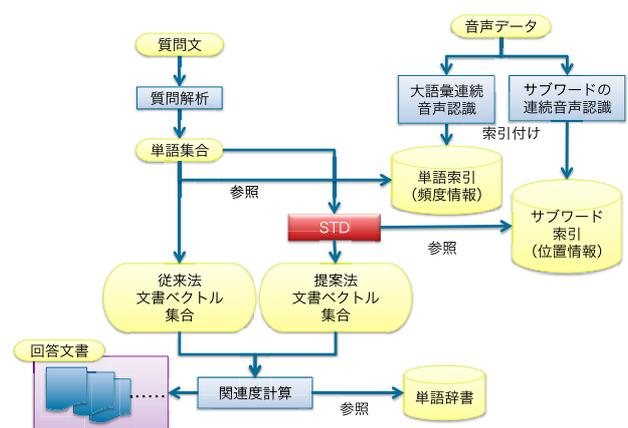


図 3 混合システム

Fig. 3 Hybrid SDR system.

処理の高速化のため、索引付けを行う。

(b) サブワード索引の作成

同様に、音声ドキュメントに対してサブワードの連続音声認識を適用し、サブワード系列の自動書き起こしテキストを得る。ステップ (3b) の STD システムに応じて、索引付けを行う

(2) 質問解析

クエリに形態素解析、およびストップワード除去を適用して単語集合に変換する。

(3) 文書ベクトルの作成

(a) 従来法

単語索引を参照して各文書の文書ベクトルを得る。

(b) 提案法

質問中の各語について、サブワード系列の自動書き起こしテキストを対象に STD を行い、各語の検出結果を得る。検出結果を語の出現とし、各文書の文書ベクトルを得る。

(4) 文書検索

文書ベクトルに基づいて、各文書とクエリとの関連度を計算し、関連度の高い文書から検索結果として出力する。

2 つの文書ベクトルを基にした関連度の計算は以下の式

表 2 認識結果のデータ

Table 2 Information about the recognition results.

	matched		unmatched	
	単語認識	音節認識	単語認識	音節認識
音響モデルコーパス	講演音声 (CSJ)			
言語モデルコーパス	講演音声 (CSJ)		新聞記事	
認識辞書語彙数	約 27,000 語	263 音節	20,000 語	263 音節
未知語率 (全講演)	1.09%	—	9.56%	—
単語正解率	74.1%	—	59.5%	—
音節正解率	83.0%	80.5%	80.6%	75.5%
OOV クエリ数	8 クエリ		41 クエリ	

によって行う.

$$\begin{aligned}
 sim(q, d) = & (1 - \alpha) \{sim_{cSDR}(q, d)\} \\
 & + \alpha \{(1 - \beta) sim_{STD-SDR}(q_{IV}, d) \\
 & + \beta sim_{STD-SDR}(q_{OOV}, d)\} \quad (11)
 \end{aligned}$$

ここで, α, β は線形結合のパラメータで, α は提案法と従来法のどちらを重視するかを表し, β は提案法について, 未知語と既知語のどちらを重視するかを表す. sim_{cSDR} , $sim_{STD-SDR}$ はそれぞれ式 (1), 式 (2) で示した従来法と提案法で用いる関連度の計算式に基づいたスコアであり, q_{IV} と q_{OOV} はクエリ中の既知語のみの集合, 未知語のみの集合を表している.

6. 実験

6.1 実験条件

NTCIR-9 SpokenDoc の SDR サブタスクで用いられたテストコレクションを用いて評価実験を行った. このテストコレクションは CSJ に収録されている 2,702 講演を対象としており, 本研究では dry-run で用いられた 39 クエリと, formal-run で用いられた 86 クエリ, 計 125 クエリを用いた. また, 検索は講演単位で行い, 正解発話区間を含む講演を検索できれば正解とした. 各クエリの平均正解文書数 (講演数) は 9.44 文書で, システムは 1 クエリあたり最大で 1,000 文書の出力を行う.

検索対象である音声ドキュメントの認識結果には, 2 種類の認識セットを用いた. 1 つは NTCIR-9 SpokenDoc [1] で提供されているものである (以降 matched と呼ぶ). matched は, 発話スタイルがマッチした講演の音声とその書き起こしを利用して音声認識の音響モデル, 言語モデルを学習しており, ドキュメントに対する適応処理を行ったという条件となっている. 音響モデルは音素 triphone, 言語モデルには大語彙連続音声認識は単語 tri-gram, 連続音節認識は音節 tri-gram が用いられている. 大語彙連続音声認識の単語正解率は 74.1%, 音節正解率は大語彙連続音声認識で 83.0%, 連続音節認識では 80.5% である. 適応処理を行っているため, 高い音声認識率が得られている.

もう 1 つの認識結果は言語モデルを新聞記事 75 カ月分で学習したもの [19] である (以降 unmatched と呼ぶ). これは, ドキュメントに対する適応処理を行っていない場合に相当する. 大語彙連続音声認識は単語 tri-gram, 連続音節認識は音節 tri-gram が用いられている. 音響モデルには上述のモデルと同じものを用いた. 新聞モデルでの大語彙連続音声認識の単語正解率は 59.5%, 音節正解率は大語彙連続音声認識で 80.6%, 連続音節認識では 75.5% である. 適応処理を行った matched に比べ, 音声認識の性能は低下している. 表 2 に認識モデルのデータを示す.

ベースライン手法には, 検索対象音声ドキュメントを大語彙連続音声認識して得られた自動書き起こしテキストに対して, テキスト文書検索を適用する従来手法を実装した (従来 SDR). 認識結果は 1-best の候補のみを使用し, TFIDF 重み付けによるベクトル空間法で検索を行った. 索引付けには単語と音節 bi-gram の 2 種類を用いた. また, 提案法との比較のため, 音節認識結果から作成した音節 bi-gram に対する検索も行った.

提案法 (STD-SDR) の検索語検出には連続 DP マッチング (DP), 索引付けと連続 DP マッチングの併用 (index + DP), 索引のみを使用するもの (indexSTD) という 3 種類のシステムを実装した. 検索対象には連続音節認識結果を用い, DP マッチングの音節間距離には音響モデル間の Bhattacharyya 距離 [18] を用いた. 湧き出し誤りの影響を抑えるために検索語長を 2 音節以上に制限することとし, また, 検索語長が 2 音節のときには完全一致のみを検出対象とした. STD の閾値は 2 分割のクロスバリデーションによって決定し, この閾値で検出した語の出現頻度を用いて, ベースライン手法と同様に TF-IDF 重み付けによるベクトル空間法で検索を行った.

さらに, 5 章で述べた混合システムを実装した. 混合システムの従来法には単語索引を用い, 提案法の STD には連続 DP マッチングを用いた. 線形結合パラメータは STD の閾値の決定と同様に, 2 分割のクロスバリデーションによって行った. STD の条件は上述のとおりである.

評価システムの条件を表 3 に示す.

表 3 評価システム条件
Table 3 System settings.

	従来 SDR			STD-SDR			混合システム
	単語索引	bi-gram (LVCSR)	bi-gram (syllable)	DP	index + DP	indexSTD	
認識結果	単語認識		音節認識			単語認識 音節認識	
索引付け	単語	音節 bi-gram		—	音節 bi-gram	音節 bi-gram skip bigram	単語 —
関連度計算	ベクトル空間法						
重み付け	TF-IDF 重み付け						

表 4 実験結果 (MAP) (matched)
Table 4 Experimental results (MAP) (matched).

	人手書き起こし			音声認識結果			検索時間 [sec/query]
	ALL	IV	OOV	ALL	IV	OOV	
従来 SDR (単語索引)	0.478	0.476	0.513	0.357	0.365	0.233	0.0259
従来 SDR (音節 bi-gram 索引 (単語認識))	0.249	0.244	0.318	0.215	0.214	0.231	0.462
従来 SDR (音節 bi-gram 索引 (音節認識))				0.177	0.174	0.215	0.462
STD-SDR (DP)	0.385 [‡]	0.381 [‡]	0.442 [‡]	0.264 ^{‡#}	0.263 ^{‡#}	0.270	727.324
STD-SDR (index + DP)	0.385 [‡]	0.382 [‡]	0.431	0.253 ^{‡#}	0.250 ^{‡#}	0.284	6.769
STD-SDR (indexSTD)	0.381 [‡]	0.380 [‡]	0.400	0.258 ^{‡#}	0.261 ^{‡#}	0.218	0.8015
混合システム	—	—	—	0.360 ^{‡#}	0.368 ^{‡#}	0.245	—

表 5 実験結果 (MAP) (unmatched)
Table 5 Experimental results (MAP) (unmatched).

	人手書き起こし			音声認識結果		
	ALL	IV	OOV	ALL	IV	OOV
従来 SDR (単語索引)	0.478	0.486	0.461	0.295	0.377	0.126
従来 SDR (音節 bi-gram 索引 (単語認識))	0.249	0.253	0.242	0.196	0.231	0.123
従来 SDR (音節 bi-gram 索引 (音節認識))				0.183	0.190	0.171
STD-SDR (DP)	0.385 [‡]	0.380 [‡]	0.395 [‡]	0.276 ^{‡#}	0.298 ^{‡#}	0.231 ^{‡#}
STD-SDR (index + DP)	0.385 [‡]	0.389 [‡]	0.378 [‡]	0.277 ^{‡#}	0.294 ^{‡#}	0.241 ^{‡#}
STD-SDR (indexSTD)	0.381 [‡]	0.383 [‡]	0.378 [‡]	0.262 ^{‡#}	0.286 ^{‡#}	0.213 [‡]
混合システム	—	—	—	0.336 ^{‡#}	0.381 ^{‡#}	0.246 ^{‡#}

6.2 実験結果

正解書き起こし, 音声認識結果に対して, 従来法と提案法を用いて検索を行った結果を表 4, 表 5 に示す. 検索性能の評価尺度には MAP (Mean Average Precision) [20] を用いた. 検索時間は 1 クエリあたりの平均検索時間である. MAP はクエリセット中の各クエリの平均精度を平均した値で, 以下の式で求められる.

$$MAP = \frac{1}{Q} \sum_{i=1}^Q AveP(i) \tag{12}$$

ここで, Q はクエリ数を表し, $AveP(i)$ は i 番目のクエリの平均精度を表し, 以下の式で定義される.

$$AveP(i) = \frac{1}{Rel_i} \sum_{r=1}^{N_i} (\delta_r \cdot Precision_i(r)) \tag{13}$$

ここで, r は i 番目のクエリの出力における検索ランクを, N_i は i 番目のクエリで検索された文書数. Rel_i は i 番目のクエリの正解文書数を表している. また, δ_r は r 番目の

出力文書が正解かどうかを表すバイナリ値, $Precision_i(r)$ は i 番目のクエリにおいて, r 番目までの出力に対する精度を表す.

検索性能はすべてのクエリ (ALL) と, 既知語のみで構成されるクエリ (IV), 未知語を含むクエリ (OOV) に分けて評価し, 未知語に対するシステムの頑健性を調査した. ここで既知語とは大語彙連続音声認識システムの認識辞書に登録されている語であり, 未知語は認識辞書に登録されていない語である. したがって, matched において既知語数は約 27,000 語, unmatched においては 20,000 語であり (表 2 認識辞書語彙数), それ以外の語はすべて未知語となる. 認識辞書が異なるため, matched と unmatched では IV (OOV) クエリの分類は異なっている (表 2 OOV クエリ数). また, 各手法の MAP 値について Paired t-test の両側検定を用いて統計的検定を行った.

6.2.1 matched 認識結果に対する実験結果

matched に対する実験結果を表 4 に示す. ここで †, ‡, #

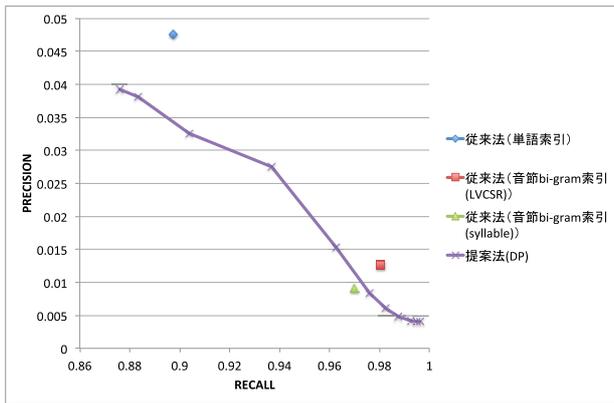


図 4 関連度計算前での検索性能 (matched)

Fig. 4 Search performance at the stage before ranking (matched).

はそれぞれ従来 SDR (単語索引), 従来 SDR (音節 bi-gram 索引 (単語認識)), 従来 SDR (音節 bi-gram 索引 (音節認識)) に対して有意水準 5% の有意差がある性能改善を示している. 表 4 より, すべてのクエリの平均 (ALL) では提案法の検索性能はいずれも従来法 (単語索引) の性能には及ばないことが分かる. これは, 検索時に漢字情報などが欠落してしまうことなどが原因であると考えられる. しかし, 音節 bi-gram を用いた従来 SDR と提案法とを比較すると, 同じ音節認識結果を用いたものだけでなく (0.177 と 0.264), 音節正解率の高い単語認識結果を用いた場合 (0.215 と 0.264) でも提案法の性能は有意に向上している. これは, 単なる固定長の n-gram 索引を用いている従来法に比べ, 提案法では STD を行うことにより単語の音節列という情報を効果的に利用できているためと考えられる. また, 単語索引を用いた従来法では未知語を利用できないことから, OOV クエリでの性能が低下するのに対し, 提案法では性能低下は見られず, 未知語の影響を受けない提案法の頑健性が確認できた (従来法 ALL: 0.357 → OOV: 0.233 に対し, 提案法 (DP) ALL: 0.264 → OOV: 0.270).

各手法で利用する検索の手がかり表現自体の性能を比較するために, 従来法の単語索引, 音節 n-gram 索引, および提案法の STD 結果が単純に出現する文書検索候補集合 (すなわち, ベクトル空間法で関連度計算を行う前の段階での文書候補集合) の, 正解文書に対する精度と再現率を調査した. 図 4 に結果を示す. 従来法の単語索引は未知語や認識誤りの影響で他の手法に比べ再現率がやや低いが, 精度は最も良い. 従来法の音節 bi-gram 索引はいずれも, サブワードを用いることで, 未知語や認識誤りに対処できているため, 再現率は高くなっている. しかし, 湧き出し誤りも多くなり精度が低下している. 後段の関連度計算では多くの文書から正解を絞り込む必要があるため, 最終的な性能が低下してしまったと考えられる. 一方, 提案法は STD の閾値を変化させることにより, 再現率の低下を抑え

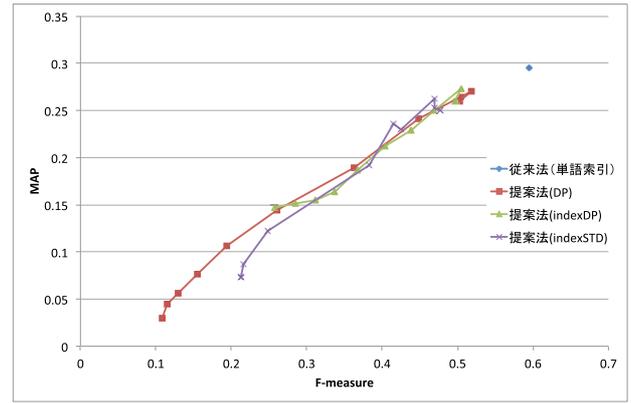


図 5 STD と SDR の性能関係 (matched)

Fig. 5 Relationship between STD and SDR performance (matched).

つつ, 従来法 (音節 bi-gram 索引) よりも高い精度を達成できている.

提案法間の性能を比較すると, 索引付けと DP を併用する提案法 STD-SDR (index + DP) は約 100 倍検索速度が向上し, 索引のみを用いる提案法 STD-SDR (indexSTD) ではさらに約 8 倍検索速度が向上し索引付けを用いた近似的な STD にもかかわらず, SDR の検索性能を低下させることなく検索性能の向上を達成することができた.

図 5 はクエリ中の検索語の検出性能 (F 値) と SDR 性能との関係である. 提案法での STD 閾値を変化させ, STD の性能 (F 値) を変化させると, それにともない SDR の性能 (MAP) も変化し, STD 性能と SDR 性能との間に強い正の相関があることが分かる. また, STD 性能をさらに向上させた先に従来法 (単語索引) の値 (F 値: 0.68, MAP: 0.357) があることから, F 値が 0.7 程度の STD 手法が実現できれば, 従来法に匹敵する性能を示すことが見込まれる.

6.2.2 unmatched 認識結果に対する実験結果

提案法の利点は未知語や認識誤りの多い文書に対する頑健性である. しかし, matched ではドキュメントに対し音声認識モデルが適応しており, 未知語率が低く, ある程度の認識率が得られている. したがって, matched に対して提案法を用いるメリットは少ない. そこで, 検索対象のドキュメントに適応していない言語モデルを用いて音声認識器を作成した. この認識器を用いて認識した音声ドキュメントに対して, 各システムの性能評価を行った (表 5). 表 4 と同様に右肩の †, ‡, § は従来法の各手法との間に有意水準 5% の有意差のある性能改善を示している.

人手書き起こしに対しては, 検索対象が matched と同一であるため, ALL の性能は一致している. ただし, IV (OOV) クエリの検索性能は異なっている. これは, 6.1 節で述べたように両モデルで用いている認識辞書が異なっているため, 各クエリの IV/OOV の分類が異なるからである. matched での OOV クエリ数は 8 クエリであり,

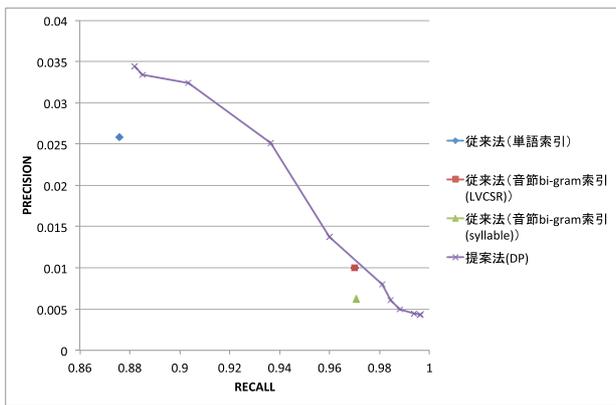


図 6 関連度計算前での検索性能 (unmatched)

Fig. 6 Search performance at the stage before ranking (unmatched).

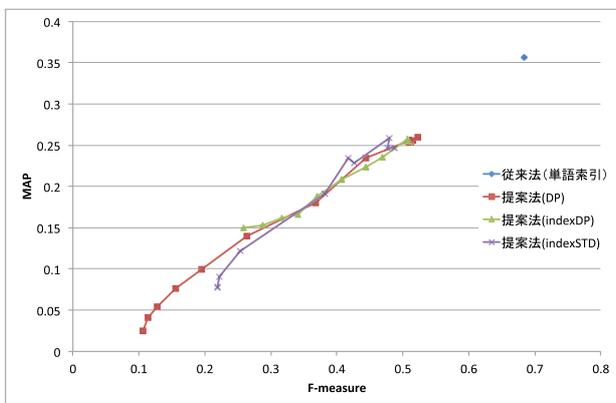


図 7 STD と SDR の性能関係 (unmatched)

Fig. 7 Relationship between STD and SDR performance (unmatched).

unmatchedでのOOVクエリ数は41クエリである(表2)。

unmatchedに対する検索では、提案法は全体性能では従来SDR(単語索引)に及ばないものの、従来法SDR(単語索引)と提案法の3つのシステムとの間に有意な差は見られなかった。また、単純なn-gram索引を用いた従来法(従来SDR(音節bi-gram索引(単語認識/音節認識)))を有意に上回る性能を示し、STDを用いることの利点を示した。従来法がOOVクエリで性能低下するのに対し、提案法は性能を維持しており、また従来法に対し有意な性能向上が見られたことから、matchedと同様に提案法の未知語に対する頑健さが確認された。

matchedと同様に関連度計算を行う前の段階での文書候補集合について、正解文書に対する精度と再現率を調査した結果を図6に示す。関連度計算前の性能について、matchedと同様の傾向を示しており、特にunmatchedにおいては同精度における再現率について提案法は従来法の単語索引を上回っている。したがって、より良い関連度計算の手法を利用することができれば、従来法の性能を上回ることも可能であると考えられる。

また、図7より、unmatchedにおいてもSTDとSDR

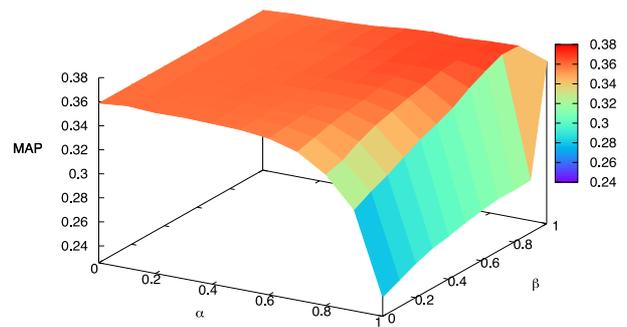


図 8 結合パラメータとシステム性能 (matched)

Fig. 8 SDR performance by varying parameter α and β (matched).

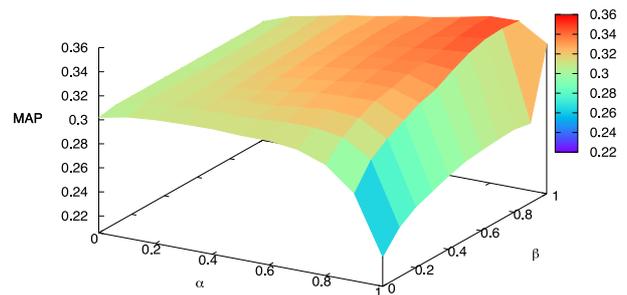


図 9 結合パラメータとシステム性能 (unmatched)

Fig. 9 SDR performance by varying parameter α and β (unmatched).

の性能間には強い正の相関が確認された。

matchedとunmatchedでの性能を比較すると、従来法ではmatchedに比べてunmatchedでの性能が著しく低い。従来法は音声認識結果の単語をそのまま索引として利用するため、単語正解率が低いunmatchedでは正しく索引付けされる語が少ないことが原因であると考えられる。したがって、従来法で検索性能を向上させるためには、検索対象に対して言語モデルを適応する必要がある。一方、提案法では両モデルの検索性能はほぼ同等である。提案法で用いている音節認識の言語モデルには音節trigramを用いている。音節のtrigramの確率は単語trigramに比べ、異なるコーパスにおける差異は少ないと考えられる。実際、両認識結果間の音節正解率にはほとんど差は見られない。したがって、提案法ではモデル間の性能差が見られなかったと考えられる。つまり、提案法は言語モデルを検索対象に適応することなく、様々な検索対象に対して安定した検索性能を示すことができると考えられる。

6.2.3 混合システムの評価

両認識結果に対して、相補的な関係である従来法と、提案法を線形結合した混合システムについて、線形結合パラメータ α 、 β を変化させながら性能を調査した。実験結果を図8、図9に示す。ここで、パラメータ α を増加させることは、従来法よりも提案法を重視することを示し、 β を増加させることは、提案法において、既知語よりも未知語

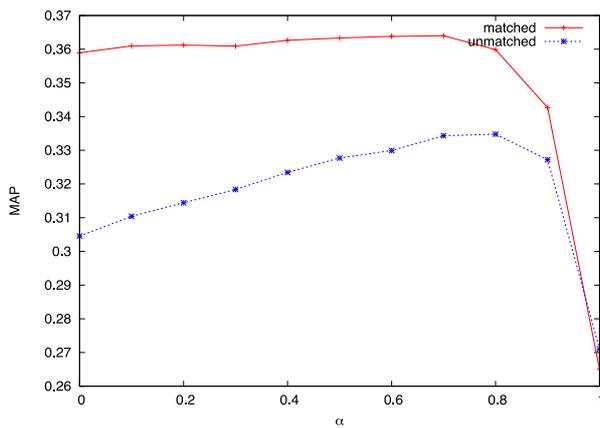


図 10 1 パラメータ性能 ($\beta = 0.5$)

Fig. 10 Relationship between parameter α and SDR performance.

を重視することを示している。特に、 $\alpha = 0$ は従来法を、 $\alpha = 1$, $\beta = 0.5$ は提案法を示す。

β を 0.5 に固定し、 α のみに着目した性能を図 10 に示す。matched の場合には、クエリの未知語率が低い (8/125 クエリ) 効果は薄い、unmatched に関しては α を増大させるにつれ性能は向上する。しかし、極端に α を大きくすると、従来法の情報をもっとく使わないこととなり ($\alpha = 1.0$)、性能は低下する。

次に β に着目すると (図 8, 図 9), α が小さい間は影響はないが、 α が大きくなると、提案法の影響が強くなり、 β を増大させるほど性能が向上する。また、matched と unmatched どちらの場合でも $\alpha = 0.5 \sim 0.9$, $\beta = 0.6 \sim 1.0$ 付近の広い領域で安定して高い性能が得られていることから、既知語についても提案法の STD を併用することで認識誤りに対処できることが確認された。ただし、matched については、単語認識率が高く、従来法の単語検索精度が良いため、STD による補完の効果は小さい。また、matched, unmatched それぞれに対して、2 分割のクロスバリデーションを行ってパラメータ α , β を決定し、従来法との性能の比較を行った (表 4, 表 5)。matched では性能に有意な改善は見られなかったが、unmatched では混合システムの性能 (0.336) は、従来 SDR (単語索引) の性能 (0.295) を有意に上回る性能を示した。

7. まとめ

本研究では未知語や認識誤りに頑健な音声ドキュメントの内容検索を行うために、音声検索語検出の結果を取り入れる SDR 手法を提案した。提案手法は連続音節認識の結果に対し、SDR の前処理としてクエリ中の各単語について STD を行いその出現位置を特定し、この単語の出現情報を利用して文書検索を行うものである。STD を行うことで未知語や認識を誤った語を文書検索の手がかりとして利用することができる。本研究では、STD 手法として、一般的な

連続 DP マッチングと、索引を用いた高速化手法とを実装した。索引を用いることで、連続 DP マッチングの欠点である検索速度を改善し、かつ検索性能の向上も達成した。

音声認識器を検索対象に適応させたモデルと、適応させていないモデルそれぞれについて講演音声を対象とした検索実験を行った結果、提案法は未知語に対して頑健であり、また、検索対象にあわせた音声認識の適応処理を必要としないことが確認された。実際のシステムの利用場面では、ドキュメントに対し、十分な適応処理を行えない場合も十分考えられ、このような状況において提案手法は有用であると考えられる。

STD を用いた提案法は、検索に時間がかかることが問題であると考えていたが、本研究では索引付けを用いることで検索の高速化が可能であることを確認できた。今後はクエリ拡張による検索性能の改善にも取り組んでゆく予定である。

参考文献

- [1] 秋葉友良, 西崎博光, 相川清明, 河原達也, 松井知子, 伊藤慶明, 胡 新輝, 中川聖一, 南條浩輝, 山下洋一: NTCIR-9 SpokenDoc: 音声検索語検出と音声ドキュメント検索の評価枠組みの設計, 情報処理学会研究報告, Vol.2010-SLP-84, No.18 (2010).
- [2] Garofolo, J.S., Auzanne, C.G.P. and Voorhees, E.M.: The TREC Spoken Document Retrieval Track: A Success Story, *Proc. TREC-9*, pp.107-129 (1999).
- [3] National Institute of Standards and Technology: Spoken Term Detection Evaluation Portal, available from <http://www.nist.gov/speech/tests/std/>.
- [4] 西崎博光, 胡 新輝, 南條浩輝, 伊藤慶明, 秋葉友良, 河原達也, 中川聖一, 松井知子, 山下洋一, 相川清明: Spoken Term Detection のためのテストコレクション構築とベースライン評価, 情報処理学会研究報告, Vol.2010-SLP-81, No.13 (2010).
- [5] Akiba, T., Aikawa, K., Itoh, Y., Kawahara, T., Nanjo, H., Nishizaki, H., Yasuda, N., Yamashita, Y. and Itou, K.: Construction of a Test Collection for Spoken Document Retrieval from Lecture Audio Data, *Journal of Information Society of Japan*, Vol.50, No.2, pp.501-513 (2009).
- [6] 秋葉友良: 音声ドキュメント検索の現状と課題, 情報処理学会研究報告, Vol.2010-SLP-82, No.10 (2010).
- [7] 杉本樹世貴, 西崎博光, 関口芳廣: 音声ドキュメント検索における Web ページを用いたドキュメント拡張の効果, 情報処理学会研究報告, Vol.2009-SLP-76, No.11 (2009).
- [8] 宇野 有, 伊藤彰則, 伊藤 仁, 牧野正三: 音声ドキュメント検索のための WWW を用いたインデクス改善, 第 4 回音声ドキュメント処理ワークショップ講演論文集, No.9 (2010).
- [9] Akiba, T. and Honda, K.: Effects of Query Expansion for Spoken Document Passage Retrieval, *Proc. International Conference on Speech Communication and Technology*, pp.2137-2140 (2011).
- [10] Chia, T.K., Sim, K.C., Li, H. and Ng, H.T.: A lattice-based approach to query-by-example spoken document retrieval, *Proc. 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.363-370 (2008).

- [11] Chen, B., Wang, H.M. and Lee, L.-S.: Discriminating capabilities of syllable-based features and approaches of utilizing them for voice retrieval of speech information in Mandarin Chinese, *IEEE Trans. Speech and Audio Processing*, Vol.10, pp.303-314 (2002).
- [12] 大橋宏正, 柘植 覚, 北岡教英, 武田一哉, 北 研二: クエリ拡張と音節認識の統合による音声ドキュメント検索, 日本音響学会春季研究発表会研究論文集, pp.259-262 (2012).
- [13] Pan, Y.-C. and Lee, L.-S.: Performance Analysis for Lattice-Based Speech Indexing Approaches Using Words and Subword Units, *IEEE Trans. Audio, Speech and Lang. Process.*, Vol.18, No.6, pp.1562-1574 (2010).
- [14] Ng, K. and Zue, V.W.: Subword-based approaches for spoken document retrieval, *Speech Communication*, Vol.32, No.3, pp.157-186 (2000).
- [15] Jones, G.J.F., Foote, J.T., Jones, K.S. and Young, S.J.: Retrieving Spoken Documents by Combining Multiple Index Sources (1996).
- [16] Wechsler, M., Munteanu, E. and Schäuble, P.: New techniques for open-vocabulary spoken document retrieval, *Proc. 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.20-27 (1998).
- [17] Akiba, T. and Yokota, Y.: Spoken Document Retrieval by Translating Recognition Candidates into Correct Transcriptions, *Proc. International Conference on Speech Communication and Technology*, pp.2166-2169 (2008).
- [18] 山本一公, 中川聖一: 発話スタイルによる話速・音韻間距離・ゆがみの違いと音声認識性能の関係, 電子情報通信学会論文誌, Vol.J83-D-II, No.11, pp.2438-2447 (2000).
- [19] 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: 日本語ディクテーション基本ソフトウェア (99年度版) の性能評価, 情報処理学会研究報告, Vol.SLP-31-2, No.137-7 (2000).
- [20] Akiba, T., Nishizaki, H., Aikawa, K., Kawahara, T. and Matsui, T.: Designing an Evaluation Framework for Spoken Term Detection and Spoken Document Retrieval at the NTCIR-9 SpokenDoc Task, *Proc. 8th International Conference on Language Resources and Evaluation* (2012).



秋葉 友良

昭和 40 年生。平成 7 年東京工業大学大学院システム科学専攻博士課程修了。同年通産省電子技術総合研究所入所。平成 13 年独立行政法人産業技術総合研究所に組織移行。平成 16 年より豊橋技術科学大学工学部助教授。現在、豊橋技術科学大学工学部准教授。自然言語処理、音声言語処理の研究に従事。博士 (工学)。電子情報通信学会、人工知能学会、日本音響学会、言語処理学会各会員。



瀧上 智子

昭和 62 年生。平成 23 年豊橋技術科学大学工学部情報工学課程卒業。同年同大学大学院工学研究科情報・知能工学専攻入学、現在、在学中。