情報処理学会論文誌　Vol. 52　No. 12　3461–3470 (Dec. 2011)

音声ドキュメント内容検索のための WEB を用いたドキュメント拡張

西崎博光[†1]　杉本樹世貴[†2]　関口芳廣[†1]

音声ドキュメントの内容検索において，検索精度に影響する主な要因が音声認識誤りである．音声認識技術の改良により音声ドキュメントの認識を改善することができる．しかし，検索要求に未知語が含まれる場合は，その要求を満たす検索を行うことができない．そこで，本論文では，検索対象の音声ドキュメントの内容に関連する WEB ページを収集し，それを用いて検索対象のドキュメント拡張を行う方法を提案する．テストコレクションを用いた実験では，WEB ページによるドキュメント拡張は未知語の検索要求の場合に検索精度改善に効果があった．しかし，音声ドキュメントが持つ話題の多様性から，無関係なページも多く収集していることが確認できた．そこで，ドキュメントを内容ごとに分割し，分割されたセグメント単位で WEB ページを集めることで，検索対象のドキュメントの内容により即した WEB ページを収集する．これにより検索精度が改善でき，WEB ページを用いたドキュメント拡張の効果が見られた．

Document Expansion Using WEB for Spoken Document Retrieval

Hiromitsu Nishizaki,[†1] Kiyotaka Sugimoto[†2] and Yoshihiro Sekiguchi[†1]

In spoken document retrieval, the main factor affecting retrieval performance is speech recognition errors. Refining speech recognition technology can make improvement of speech recognition performance for spoken documents. However, if a query has out-of-vocabulary (OOV) words, we cannot get the spoken documents related to the query. This paper describes spoken document retrieval using document expansion based on WEB whose contents are similar to the spoken documents retrieved. The retrieval experiment showed that the document expansion worked well on OOV queries, but many irrelevant WEB pages were collected because of the variety of topics that spoken documents have. Therefore, each spoken document is automatically divided into some segments. And then, more similar WEB pages to the spoken document can be collected using the query derived from the segment. The improved document expansion achieved improvement of the spoken document retrieval performance.

1. はじめに

情報爆発時代を迎え，大量の情報から効率良く必要な情報を取り出す技術の確立が望まれている．大量の音声・映像ドキュメントから必要な情報を検索する技術もその 1 つである．

これらを検索するための検索要求は，テキスト入力で行われるのが一般的であったが，近年では音声入力による検索要求も行えるようになっており，スマートフォン等で利用できる．

これに先立ち，1990 年代から，アメリカ国立標準技術研究所（NIST）とアメリカ国防総省が主催する情報検索関連の評価型ワークショップ TREC（Text REtrieval Conference）が開催されている．1990 年代後半から 2000 年までの間，TREC の研究分野（トラック）の 1 つとして，音声ドキュメント検索トラック（Spoken Document Retrieval Track）が開催され，世界中で音声ドキュメント検索の研究が行われてきた．現在は，ビデオ検索の評価ワークショップ（TRECVID ⋆1）が開催されており，映像中に含まれる音声データを音声認識する等して，ビデオ検索精度の改善が図られている[1)]．

一方，日本では，2006 年に情報処理学会音声言語情報処理研究会のワークショップである「音声ドキュメント処理ワーキンググループ」が設立された．ワーキンググループは，『日本語話し言葉コーパス』（以下 CSJ と記す）を用いた音声ドキュメント検索用のクエリと，クエリに対する正解ドキュメントのセットを構築した[2)]．

この検索テストセットの構築により，日本でも大規模なデータを用いた情報検索技術開発の環境が整った．日本語の音声ドキュメント検索は，これまで，各研究機関が独自のデータベースを用いて研究を行ってきたが[3)]，各研究機関が共通して利用できる日本語の評価セットができたことから，音声ドキュメント検索研究のさらなる発展が期待できる．この評価セットは TREC SDR Track や TRECVID が採用しているニュース音声・映像とは異なり，

†1 山梨大学大学院医学工学総合研究部
Department of Research, Interdisciplinary Graduate School of Medicine and Engineering, Yamanashi University
†2 山梨大学大学院医学工学総合教育部
Department of Education, Interdisciplinary Graduate School of Medicine and Engineering, Yamanashi University

⋆1 http://www-nlpir.nist.gov/projects/trecvid/

話し言葉音声を対象としている．したがって，音声認識がより難しく，それに比例して検索の高精度化も難しい．

日本語に対しても将来的に，音声やビデオ等のマルチメディアデータが爆発的に増加していくことを考えると，これらの検索技術を開発することは喫緊の課題である．そこで本論文では，この検索テストセットを対象に，検索精度の高精度化を図ることを目的に，WEB ページを用いた新しい検索方法の提案を行う．

通常，音声ドキュメントを検索する場合，検索対象の音声データを音声認識することで音声波形を単語列等にシンボル化し，そこからドキュメントに対するインデックスを作成する．そのインデックスを手がかりにして検索を行う．この際，音声認識処理を用いるため，音声認識誤りや未知語問題（音声認識辞書に必要な単語が登録されていない）が，検索精度に大きく影響する．すなわち，検索のための入力クエリに含まれる単語が検索対象の音声ドキュメントの音声認識結果に出現していないと照合ができず，検索されるべきドキュメントが検索されないことになる．

これらの問題を改善するため，本研究ではインターネット上の WEB 情報を利用することで検索精度の改善を目指す．提案手法では，検索対象の音声ドキュメントのデータと内容が類似した WEB ページを音声ドキュメントのインデキシングに利用する．これによって，本来ならインデックスとして登録されてほしい単語が，音声認識誤りや未知語が原因で登録されない事態を回避することが狙いである．

簡単な検索実験の結果，1 音声ドキュメントごとに WEB ページの収集を行うドキュメント拡張により，未知語を含むクエリに対する検索精度が大きく改善された．しかし，クエリ全体での評価では，有意に改善されているとはいえない結果となった．

この理由として，音声ドキュメントが持つ内容の多様性により，真に必要とする WEB ページを収集できなかったことが考えられる．そこで，1 音声ドキュメント内に複数の話題が出現することに着目し，話題に沿ってドキュメント分割を行うことで話題ごとに WEB ページの収集を行う方法を考案した．これにより，より話題に適した WEB ページの収集を行うことができる．新しい検索実験の結果，音声ドキュメントを分割したセグメントごとに WEB ページの収集を行うことで，検索精度を改善することができた．

2. 関 連 研 究

音声ドキュメント検索問題を解決するために，90 年代後半からこれまで，数多くの検索手法が開発されてきた．

たとえば，Wechsler ら[4] は，音素認識器を使って音声を書き起こし，入力検索語の音素表記とのマッチングを行っている．Ng ら[5] は，音声の音素表記から作ったサブワードをインデキシングの単位として扱い，DP マッチングにより柔軟なマッチングを行っている．岩田ら[3] は，語彙に影響されない音声ドキュメント検索として，サブワード単位の認識結果を利用している．また，クエリとドキュメントにおいて，高速にサブワードどうしをマッチングさせる手法を提案している．

近年では，音声認識結果のラティス[6] やコンフュージョンネットワークを利用したり[7]，音声認識結果に信頼度を導入したりすることで[8]，検索性能改善を図っている研究例もある．

日本語でもテストコレクションが整備されたことから，この検索テストコレクションを対象にした研究も行われている．まず，文献 2) では，テストコレクションを利用した音声ドキュメント検索の基本的な検索精度が述べられている．また，Akiba ら[9] は，テキスト翻訳の技術を利用した語彙拡張を音声ドキュメント検索に応用し，テストコレクションでの有効性を示した．さらに，胡ら[7] は，音声認識結果のコンフュージョンネットワークを利用することで，検索精度が改善できることを示した．重安ら[10] は，インデックスに用いる単語形態として N 文字連鎖等の利用や，不要語の設定手法について提案している．

音声ドキュメント検索において，クエリ拡張を利用することで検索精度の改善を図っている研究例がある．たとえば Terao ら[11] は，検索に用いる音声クエリと関連性のある WEB の情報を利用することで，クエリ拡張を図っている．また Mamou ら[12] は，単語の発音に基づく拡張処理を提案している．

一方，検索対象の音声ドキュメントの誤認識や未知語問題を解決する方法として，音声ドキュメントを拡張することで検索精度の改善を図っている研究例もある．たとえば Singhal ら[13] の報告である．これはニュース音声を検索する際に，その音声と類似したデータをニュースコーパスから選択するという手法である．

以上をまとめると，音声ドキュメントのサブワード音声認識結果を用いる方法，ラティス等のより豊かな音声認識結果を用いる方法，音声ドキュメントや検索クエリを拡張する方法が提案され，有効性が示されている．本論文では，WEB の知識を用いて音声ドキュメントを拡張することで，検索性能の改善を図る．我々の提案手法は，既存の提案手法と簡単に組み合わせることでさらなる検索性能の改善を図ることができるが，組合せによる性能改善については別の機会で述べたい．

今回，我々がターゲットにしている検索対象のデータは，様々な話題を含んだ学会講演・模擬講演音声であるが，今後，様々な話題の音声（映像）ドキュメントが増加することを考

慮すれば，ドキュメント拡張の手法を導入する際に用いる類似コーパスとしては，WEB が適していると考えられる．

本論文の手法は，音声ドキュメント拡張の研究としては Singhal らの手法[13] と類似している．しかし，ニュース音声をターゲットとし，同じドメインの限られた量のニュースコーパスを用いてドキュメント拡張する文献 13) の手法よりも，膨大な WEB 上の情報を用いてドキュメント拡張する本論文の方が，適切なドキュメント拡張ができる可能性は高くなる．ただし，本論文の手法では，拡張したい音声ドキュメントの内容と異なる WEB ページが拡張に用いられる可能性も高くなるという難しさがある．

また，Terao らの研究[11] も WEB の情報をクエリ拡張に用いているが，ニュースの話題に限定しているため，本論文と比べると若干やさしい問題設定である．

このような問題をふまえたうえで，本論文では，WEB ページを用いたドキュメント拡張の手法を導入し，その有効性を確かめる．この手法では，いかに適した WEB ページを収集できるかが問題となる．これには，WEB ページを用いた言語モデル拡張の研究等で用いられる WEB 検索クエリの選定方法[14] が参考になるが，本研究では WEB 検索クエリの選定方法ではなく，音声ドキュメント分割を用いる方法を導入し，その有効性を示す．検索ユーザが欲する情報は，音声ドキュメントの一部分に含まれていることが多い．そこで，音声ドキュメントを分割し，その分割単位で WEB ページを集めることで，より適した WEB ページをドキュメント拡張に用いることができる．

WEB ページを利用したドキュメント拡張を行う利点として，3 つの点をあげることができる．1 点目は，音声認識誤りや未知語問題に対処できることである．外部の知識を用いて検索対象の情報を補完することで，クエリとドキュメントのミスマッチを抑えることができる．2 点目は，ユーザが欲する音声ドキュメント内に検索クエリの単語が含まれていなくても，WEB ページの利用によるドキュメント拡張によりその単語が補完される点である．最後の点は，WEB ページは日々増加しかつ更新もされているため，音声ドキュメントの内容と類似したページを見つけやすいという利点を持つ．特に検索対象の音声データベースに様々な話題の音声が含まれているときには拡張対象のデータとして WEB ページの利用は有効であると考えている．

3. 検索テストコレクションとその音声認識

3.1 テストコレクション

本研究では，情報処理学会音声言語情報処理研究会のワークショップである「音声ドキュメント処理ワーキンググループ」が CSJ[15] を対象に構築した，「音声ドキュメント検索テストコレクション」[16] を使用する．

「音声ドキュメント検索テストコレクション（以下，CSJ テストコレクションと記す）」とは，日本語の自然発話音声を対象とした音声ドキュメント検索を評価するためのテストセットである．CSJ テストコレクションには，CSJ に収録されている「学会講演」「模擬講演」の 2,702 講演（約 600 時間の音声データ）に対し，それらの講演，または講演の一部を検索するためのユーザの検索クエリと適合ドキュメントの組，および各講演の音声認識結果が含まれている．

ユーザの検索クエリは全 39 個であり，各検索クエリに対する正解が含まれている文章の場所（講演 ID とフレーズ ID）が人手によりタグ付けされている．タグには完全に正解と判断できる “適合” と正解とまでいえないまでも部分的に正解のヒントと判断できる “部分適合”，そしてまったく正解と関係ない “不適合” がある．音声認識結果には，第 1 候補（1-best）から第 10 候補（10-best）までの 10 個の候補文，およびそれらの音素情報が含まれている．また，未知語や音声認識誤りがない場合の検索性能を検証するため，各講演の人手による書き起こしデータも用意されている．

検索単位としては，講演を 1 ドキュメントとする単位と，ある一定の文章区間を 1 ドキュメントとする単位が考えられる[2]．本研究では前者を採用する．

3.2 テストコレクションの音声認識

本研究では，CSJ テストコレクションに含まれている音声認識データは使用せず独自に，音響モデルと言語モデルを用意し，音声認識を行った．その理由は以下の 2 点である．

1 点目として，付属の音声認識データは，80%以上の高い音声認識精度であり，未知語クエリの数も少ないが[2]，これは，2,702 講演を音声認識するために必要なモデルを完全にオープンな学習データから学習しておらず，音声認識を行う際に用いる音響モデル，言語モデルの学習データはクローズドであり，実用的ではないからである．

2 点目として，言語モデルの学習データを作成する際に用いる形態素解析システムを，インデキシングと検索システムに用いるものと同一にした方が実用的だからである．

これらをふまえて，本研究では，音声ドキュメントにオープンなデータも含まれるよう CSJ の学会講演集合のみから音響モデル，言語モデルを学習した．これにより，学習データは文献 2) で用いられている各モデルの学習データよりも少なくなっている．

音声認識に用いた音響モデルは，CSJ 学会講演から認識評価用セットを除いた学会講演 970 講演から学習したトライフォンである．使用する特徴量は，12 次元の MFCC，ΔMFCC，

表 1　音声ドキュメントの音声認識率とクエリに対する未知語率
Table 1　Speech recognition rates of spoken documents and out-of-vocabulary rate for query.

Corr.[%]	Acc.[%]	クエリに対する未知語率 [%]	未知語クエリ数
76.9	71.6	11.8	11

ΔΔMFCC と，Δ パワー，ΔΔ パワーの全 38 次元となっている．言語モデルの学習は，まず，音響モデルの学習データと同じ学会講演 970 講演の書き起こしを，形態素解析器 ChaSen 2.4.4（辞書には UniDic 1.3.9 を利用）で形態素解析を行う．それを用いて，語彙サイズ 17,000 の単語トライグラムを学習した．

CSJ の 2,702 講演に対する音声認識を行った結果を**表 1** に示す．表 1 において，“**クエリに対する未知語率**” とは，音声ドキュメントを検索するためのクエリに含まれる単語のうち，音声認識辞書に含まれていない単語の割合である．“**未知語クエリ数**” とは，全 39 個のクエリに対し，未知語を含んでいる検索クエリの個数である．

本論文で検索実験等に利用するデータは，表 1 のものである．

4. WEB によるドキュメント拡張

4.1 処理の概要

提案する WEB によるドキュメント拡張を用いた音声ドキュメント内容検索処理の概要を**図 1** に示す．

まず，検索対象の音声ドキュメントを大語彙連続音声認識システムを用いて音声認識する．このとき，音声認識システムの認識辞書に登録されていない単語が未知語となり，ユーザの検索クエリに未知語が含まれている場合，インデックスとの照合が不可能となる．

この認識結果から，話題の特定に不向きな単語を不要語[⋆1]として取り除き，インデックスを構築する．本論文では，これを “**認識インデックス**” と記す．

次に，WEB ページから作成するインデックスについて説明する．まず音声認識結果から，認識されたドキュメントと内容が類似した WEB ページを検索するためのクエリを作成する．ここで構築する WEB 検索用クエリの “質” によって，どれだけ内容が近い WEB ページを収集できるかが決定される．したがって，クエリの構築方法にも工夫が必要である．

作成したクエリを WEB 検索エンジンに入力し，WEB ページの収集を行う．このようにして集めた WEB ページから，認識インデックスを作成した場合と同様に不要語を取り除き，インデックスを構築する．本論文ではこれを “**WEB インデックス**” と記す．

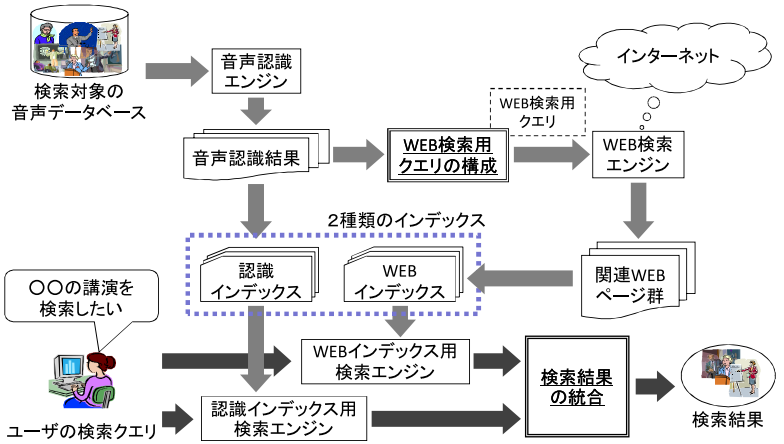

図 1　WEB によるドキュメント拡張を用いた音声ドキュメント内容検索処理の概要
Fig. 1　Spoken document retrieval framework using WEB document expantion.

このように，音声ドキュメントに対して 2 種類のインデックスが構築される．つまり音声認識結果から直接作成される “認識インデックス” と，関連する WEB ページから作成される “WEB インデックス” である．

検索処理では，それぞれのインデックスに対して照合処理を行うため，2 つの検索エンジンを用いる．ユーザの検索クエリを 2 つの検索エンジンに入力し，それぞれのエンジンから得られた検索スコアを統合することで，最終的な検索結果を得る．

4.2 WEB 検索用クエリの構成法

WEB 検索を行うためには，音声認識結果からキーワード抽出を行い，話題に合った適切な WEB 検索用クエリを構成しなければならない．しかし，音声認識結果が形態素に分割されていること，話題の特定に不向きな単語（機能語や一般語）が多く含まれていることから，話題に合った WEB 検索用クエリを構成することは容易ではない．

そこで，ある程度適切な WEB ページを収集できるように，WEB 検索用クエリの構成を工夫する．

WEB 検索用クエリを構成するために，音声ドキュメントの音声認識結果中に含まれる品詞を判定し，適宜単語の連結と削除を行うことで，実在する可能性が高い名詞 N-gram を

⋆1 本研究では，名詞と動詞，形容詞，形容動詞以外を不要語とした．

抽出する．この名詞 N-gram を WEB 検索用クエリとする．

具体的には，以下の手順で行う．

(1) 名詞 N-gram の作成：話題を特定する名詞が続く限りそれらを連結し，名詞 N-gram を作成する．N の大きさは任意である．このとき，Web 日本語 N-gram 第 1 版（“Google N-gram” と記す）[⋆1]を使用し，連結した名詞 N-gram が実在するかを確認する．Google N-gram は，一般に公開されている日本語の WEB ページで，Google がクロールしたものから抽出されている．抽出対象となった文数は約 200 億文で，出現頻度 20 回以上の 1〜7-gram が収録されている．すなわち，名詞 N-gram が Google N-gram に存在すれば，実在する文字列であるといえる．

連結された名詞 N-gram が Google N-gram に存在すれば，WEB 検索用クエリの候補とする．もし，存在しなければ，連結した名詞 N-gram の末尾の名詞単語を切り離し，再度 Google N-gram に存在するかどうかを確認する．この作業を，連結した名詞 N-gram が Google N-gram 中の N-gram と完全に一致するまで繰り返す．最後まで一致しなければ，この名詞 N-gram を構成する名詞は，WEB 検索用クエリ候補としない．これにより，音声ドキュメントの音声認識結果のテキストから，名詞 N-gram を抽出する．

具体的な処理の流れの例を**図 2** に示す．

(2) WEB 検索用クエリの構成：作成した名詞 N-gram のうち，式 (1) により求めた単語スコアの上位 5 種類の名詞 N-gram を採用する．$Score(w_i)$ が高くなる名詞 N-gram は，出現頻度が高く名詞 N-gram を構成する単語数が多いものである．これは，“同じドキュメント中で何度も繰り返し使用される単語は重要な単語である” という観点と，“名詞 N-gram が表す話題は，構成する単語が表す話題の積集合であり，構成単語数が多いほど話題を特定する能力が高い” という観点に基づいている．

$$Score(w_i) = \frac{1 + tf_i \times \log_{10}(1+N)}{\sum_{k=1}^{m} tf_k} \tag{1}$$

tf_i は N-gram w_i の出現頻度，N は N-gram w_i を構成している単語数，m はドキュメント内の N-gram の種類数である．

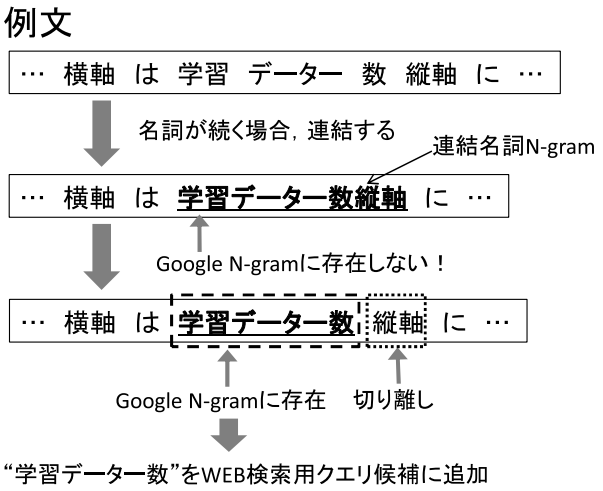

図 2　名詞 N-gram の作成例

Fig. 2　Example of making noun phrase from word sequence.

4.3　WEB ページの収集とインデックスの構築

4.2 節の手順で作成した WEB 検索用クエリを入力とし，“Yahoo! Web 検索 API [⋆2]” を使って音声ドキュメントに関連する WEB ページ群を収集する．

WEB ページの収集方法を**図 3** に示す．単語スコアの高い名詞 N-gram を優先して，名詞 N-gram の組合せを作成し，完全一致の AND 検索により WEB ページを収集する．収集する WEB ページ数は，1 音声ドキュメントに対して 50 ページとする．WEB 検索クエリによっては，目標とする件数の収集が困難な場合がある．その場合は，最も単語スコアの低い名詞 N-gram をクエリから取り除き，再度検索を行う．

集められた WEB ページからは不要語を除去しておく．1 音声ドキュメントから集められた WEB ページ群は，1 ドキュメントとして扱いインデックス化される．

4.4　ドキュメント検索エンジン

本研究では，検索システムとして汎用連想計算エンジン GETA [17] を利用する．GETA を利用することで，大規模なドキュメント–単語集合間の類似度を高速に計算することが可

⋆1 http://www.gsk.or.jp/catalog/GSK2007-C/catalog.html

⋆2 http://developer.yahoo.co.jp/webapi/search/

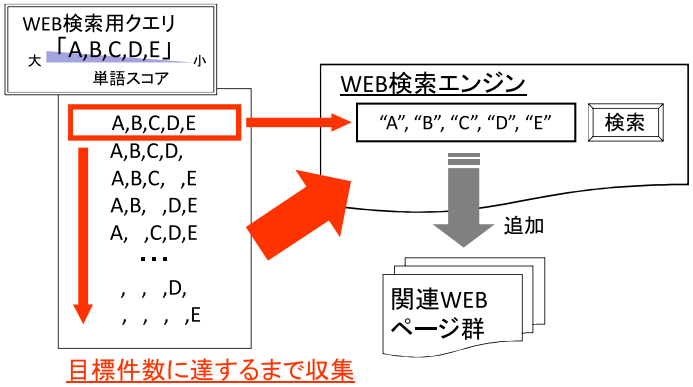

図 3　WEB ページの収集方法
Fig. 3　Collection method of WEB pages.

能である．

GETA には，類似度計算手法として TF・IDF や AND，SMART 法等の尺度があらかじめ用意されている．本研究では，それらの中から SMART 法[2),18)] を採用した．

検索結果は，類似度の高いドキュメントから順に出力される．

4.5　検索結果の統合

認識インデックスと WEB インデックス，それぞれのインデックスを用いたときの検索結果を統合し，最終的な結果を得る．

2 つのインデックスを用いて検索された音声ドキュメント d の最終的なスコア $sim(d)$ は，式 (2) に示すように，認識インデックスの検索結果の検索スコアと WEB インデックスの検索結果の検索スコアの線形補間により計算される．

$$sim(d) = (1-\alpha) \times sim(d|r) + \alpha \times sim(d|w) \quad (2)$$

ただし，$sim(d|r)$ は認識インデックス，$sim(d|w)$ は WEB インデックスを用いたときの音声ドキュメント d の検索スコアとなる．α は，各インデックスに対する重み係数である．実験ではこの α を 0.0 から 1.0 までの 0.1 刻みで変化させ，検索結果の統合を行う．

5.　音声ドキュメント検索実験

5.1　検索単位と評価尺度

検索の単位は講演単位とする．すなわち，1 ドキュメント 1 講演である．検索されたドキュメントの適合性の判定は “適合” と “部分適合” のドキュメントを正解とする．

各クエリに対する検索精度を計る尺度には，補間 11 点平均精度（Interpolated 11-points Average Precision）を用いる[19)]．

検索実験では，クエリごとに上位 1,000 件までの結果を出力し，クエリセットに対してクエリごとの補間 11 点平均精度で平均をとり評価する．

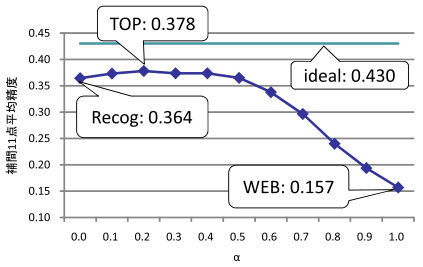

図 4　全 39 クエリ対する検索精度の変化
Fig. 4　11 point average values varying depend on α for all 39 queries.

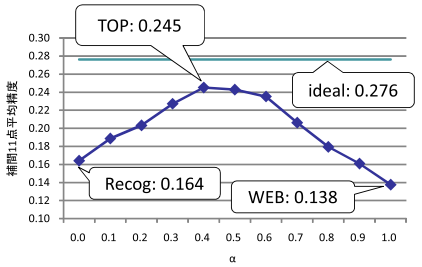

図 5　未知語クエリのみに対する検索精度の変化
Fig. 5　11 point average values varying depend on α for OOV queries.

5.2　ドキュメント拡張を用いた検索実験結果

認識インデックスと WEB インデックスの補間重み α に対する音声ドキュメントの検索精度を調べるための実験を行った．

実験結果を**図 4**，**図 5** に示す．図中の “Recog” と “WEB” は，それぞれ，認識インデックスのみを利用したとき，WEB インデックスのみを利用したときの検索精度である．“TOP” は重み α を変化させていったときの最大精度を示しており，“ideal” は，クエリごとに最も精度が高くなるように α を人手で決定した場合の検索精度で，最適な線形結合が実現された場合の理想値である．

図 4 の 39 クエリに対する検索精度の変化を見ると，α=0.2 の際に最大の精度 0.378 をとっている．また，WEB を用いないとき，すなわち $\alpha = 0$ のときの精度は 0.364 であった．このとき，39 のクエリ中 24 クエリで精度が改善されたが，有意差検定[⋆1]を行った結果，有意な改善は得られなかった．

一方で，図 4 がすべての検索クエリに対する結果だったのに対し，図 5 は音声認識辞書

⋆1 本論文では，結果の有意差を検証するための検定には，すべて符号検定を適用している．すなわち，39 クエリ（未知語を含んだクエリは 11）のうち，何クエリの検索精度が改善したのかという基準に基づいた検定である．

に対する未知語を含んでいるクエリに対する実験結果である．この結果を見ると，$\alpha = 0$ のとき 0.164，$\alpha = 0.4$ のとき 0.245 であり，11 クエリ中 9 クエリで検索精度が向上し，有意に改善されている（$p < 0.05$）．

したがって，4 章で述べた WEB を用いた手法は，未知語を含んだクエリに対しては有効であるといえる．すなわち，WEB ドキュメント拡張により未知語を補完できている．今回使用したテストコレクションの検索クエリには，全部で 12 種類の未知語と 1 種類の発声されていない単語が含まれている．本手法で集めた WEB ページの中には，これら 13 個の単語中 12 個の単語が含まれていた．このことから，未知語を含むクエリに対して検索性能が向上したと考えられる．

以上の結果より，音声ドキュメント検索において，WEB ページによるドキュメント拡張は，全 39 クエリでの平均精度の改善を有意に得られなかったものの，未知語を含む 11 クエリに対しては，8.1 ポイントの有意な改善が見られた．ただし，最適な重み係数をクエリごとに設定したときの理想値と比べてみると，まだ改善の余地が残っている．動的な重み係数決定手法を検討する必要がある．

全クエリの平均精度の改善が得られなかった理由としては，次の理由が考えられる．音声ドキュメントの中には複数の話題が含まれているものも多いため，適切な WEB 検索用クエリを設定できず，収集した WEB ページが音声ドキュメントの内容と異なっていることが考えられる．また，テストコレクションの検索クエリに対する正解情報の多くは，音声ドキュメントの一部分に含まれている．このことから，次章では，より適した WEB ページの収集を行うため，話題に基づいた音声ドキュメント分割手法を用いて WEB ページの収集を行い，それを利用したドキュメント拡張を行う方法を提案する．

6. ドキュメント分割を用いた SDR

6.1 音声ドキュメント分割方法

音声ドキュメントの分割には，Utiyama ら[20] が提案したテキストセグメンテーション手法を基にした分割手法を用いる．Utiyama らの手法は単語間に話題の分割境界を設定する手法であるが，これを拡張し文間に境界を設定するように工夫した．

連続する 3 発話を 1 文[⋆1]とし，文単位でセグメント分割を行う．文と文の間に接続ノードを置く．全接続ノード数はドキュメント内の文数+1 となる．接続ノード i と接続ノード j で囲まれた部分をセグメント S_{ij}（$i < j$）と呼びその話題コスト c_{ij} を式 (3) により算出する．話題コストは，2 つのノード間で囲まれたセグメント内に複数の話題が含まれていると大きくなる．この方法では動的計画法を用いて，話題コストの和が最小となるようにノードにセグメント分割境界を設定できる．なお，最初の接続ノードと最後の接続ノードで囲まれた範囲が 1 ドキュメントである．

$$c_{ij} = \sum_{l=1}^{length} \log \frac{length + k}{tf_l} + penalty \times \log W \tag{3}$$

ここで，$length$ はセグメント S_{ij} の総単語出現数，k はドキュメントの単語種類数，tf_l はセグメント S_{ij} に対する単語 w_l の出現数，$penalty$ は，セグメント分割時におけるペナルティ，W はドキュメントの総単語出現数である．また $penalty$ を設定することにより，尤度をコスト化し分割数を調節している．

例として，隣接する各文が S_{01} = “aaa”，S_{12} = “aaaaa”，S_{23} = “bbbb” であり，$penalty$ = 1.0 と設定した場合のセグメント分割を**図 6** に示す．この図において，a，b は単語を表す．このとき，ドキュメント全体から求める値 k と W は定数となり，それぞれ $k = 2$，$W = 12$ となる．

ドキュメント全体を 1 セグメントとする場合の話題コスト c_{03} を求める場合，単語 a の出現数は 8，単語 b の出現数は 4 となり，式 (3) を用いて計算すると c_{03} =5.1991 となる．すべての接続ノード間の話題コストを計算し，動的計画法を用いて最適なセグメント境界を探索すると，S_{02}，S_{23} の 2 つのセグメントに分割したときに，最も話題コストの和が小さくなる．

6.2 音声ドキュメント分割を考慮した WEB インデックス構築

音声ドキュメント分割を考慮した WEB インデックス構築の概要を**図 7** に示す．

まず，音声ドキュメント分割を用いて，音声認識結果を任意のセグメントに分割する．作成されたセグメントごとに，4.3 節で述べた手法を用いて WEB 検索用クエリの構成を行い，関連する WEB ページの収集を行う．

収集した WEB ページから，不要語を取り除き，セグメントごとにインデックスを構築する．すなわち 1 ドキュメントに対し，セグメント分割数分の検索対象が存在することになる．

6.3 音声ドキュメント分割を考慮した検索結果の統合

認識インデックスは，ドキュメント単位で構成されており，WEB インデックスは，セグ

⋆1 本論文ではこれを分割の最小単位とする．1 発話を最小の分割単位とすると，短い発話で 1 つのセグメントを構成してしまうため．

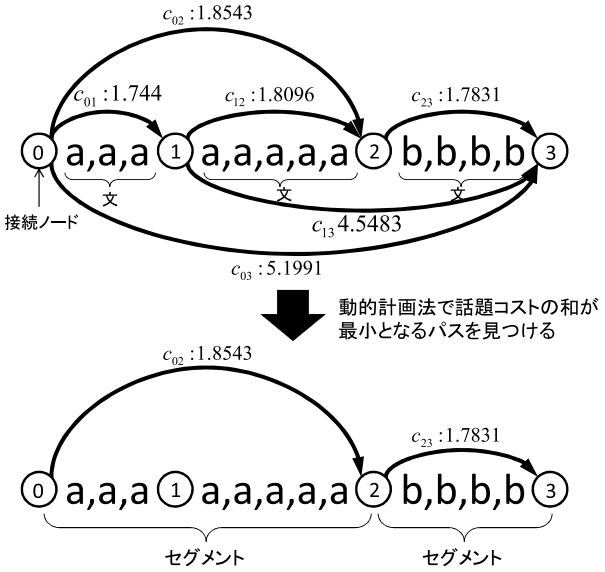

図 6　自動ドキュメント分割の例
Fig. 6　Example of automatic document segmentation.

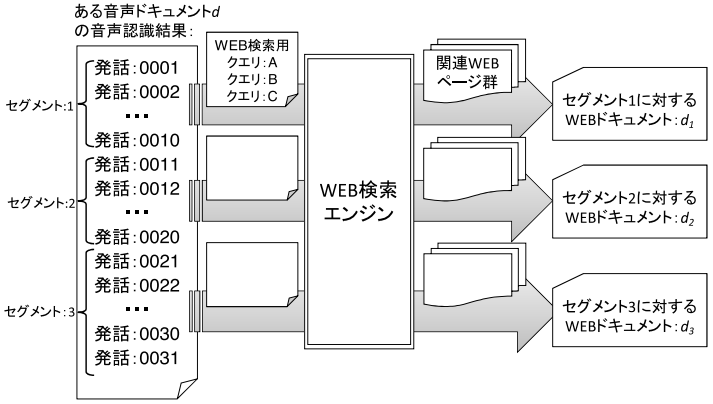

図 7　音声ドキュメント分割を考慮した WEB インデックス構築の概要
Fig. 7　Making WEB index using document segmentation.

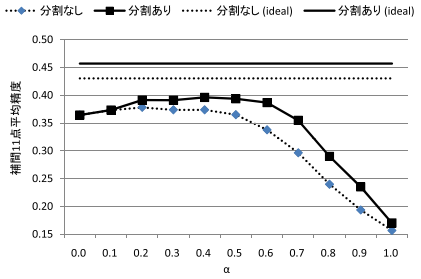

図 8　ドキュメント分割を用いた際の検索性能（全 39 クエリ）
Fig. 8　Retrieval performance with dynamic segmentation of spoken document for all 39 queries.

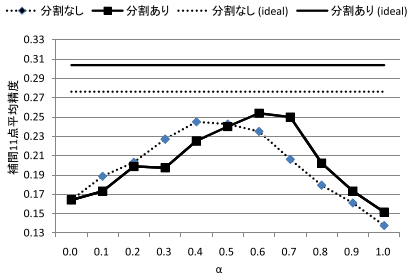

図 9　ドキュメント分割を用いた際の検索性能（未知語 11 クエリ）
Fig. 9　Retrieval performance with dynamic segmentation of spoken document for 11 queries including OOV words.

メント単位で構成されている．このことから，検索結果の統合を行う際に，両インデックス間で整合をとる必要がある．

そこで，式 (4) を用いて，認識インデックスと WEB インデックスにより得られた検索結果を統合し，最終的な結果を得る．すなわち，音声ドキュメント d の検索スコアは，認識インデックスの検索スコアと音声ドキュメント d に対して，最大となる WEB インデックスの検索スコアの線形補間により計算される．

$$sim(d) = (1-\alpha) \times sim(d|r) + \alpha \times \max_{s \in S} sim(d_s|w) \tag{4}$$

ここで，音声ドキュメント d を n 分割し，それぞれのセグメントごとに集めた WEB ページ群の集合を $S = \{d_1, d_2, \cdots, d_n\}$ とする．$sim(d|r)$ は認識インデックスでの音声ドキュメント d の検索スコア，$sim(d_s|w)$ は WEB インデックスでの音声ドキュメント d に対するセグメント s の検索スコアとなる．

6.4　分割を行った場合の検索実験結果

実験結果を**図 8**，**図 9** に示す．図 8 は，全 39 個の検索クエリ，図 9 は，未知語クエリに対するセグメント分割手法別の検索結果である．これらの図において，ドキュメント分割を行わなかった場合（これがこの実験でのベースラインとなる）と比較している．また，“ideal” は，検索クエリごとに最適な重み係数 α を手動で設定したときの結果である．

図 8 を見ると，すべての検索クエリに共通の重み係数を設定したとき，ドキュメント分割を行った場合で 0.401（$penalty = 0.5$，$\alpha = 0.4$）であり，ベースラインの 0.378（$n = 1$，

$\alpha = 0.2$）よりも検索精度は高くなったが，有意な改善ではなかった（39 クエリ中，19 クエリで精度改善）．しかし，ドキュメント拡張を行わないとき（$\alpha = 0$）の精度 0.364 と比べると，検索精度は有意（$p < 0.05$）に改善されている（39 クエリ中，26 クエリの精度改善）．

一方で，未知語が含まれているクエリのみに着目した場合においても（図 9），ドキュメント分割を行うことで，0.264（$penalty = 0.5$，$\alpha = 0.6$）とベースラインの 0.245（$n = 1$，$\alpha = 0.4$）より高い値（$p < 0.05$）となっている（11 クエリ中，9 クエリの精度改善）．

未知語のクエリのみに着目した場合，最大の検索精度が得られる α の値は，分割を行わない場合の $\alpha = 0.4$ と比べて分割を用いた場合は 0.6 と高くなっており，WEB インデックスがより重視されている結果となっている．したがって，ドキュメント分割を行った方が，より適した WEB ページが収集されていると考えることができる．また，未知語クエリについて，今回最適となった $\alpha = 0.6$ が適切か否かを判定するために，11 分割交差検定を行った[⋆1]．その結果，精度は 0.264 となり，図 9 と同じ結果になった．これは，どの試行でも最適な重みが $\alpha = 0.6$ となったためである．このことから，未知語クエリについては，$\alpha = 0.6$ を与えると，精度向上が期待できることが分かった．

また，“ideal” についてもセグメント分割を行うことで，検索精度が改善されており，ドキュメント分割を行った方が理想的な統合が行われた場合の改善率が高いことを示している．

以上のことから，音声ドキュメント分割を利用した WEB 収集の有効性が示された．ただし，“ideal” の精度と比べてみると，まだ改善の余地が残っている．

7. おわりに

本論文では，音声ドキュメントの内容検索において，WEB を用いた検索対象ドキュメントのドキュメント拡張手法について述べた．

音声認識結果を用いた音声ドキュメント検索の枠組みでは，未知語とそれによって引き起こされる音声認識誤りが検索精度に大きく影響する．そこで，未知語や音声認識誤りを補完するために，WEB ページを用いる方法を提案した．

初めは，検索対象のある 1 つの音声ドキュメントから 1 つの WEB 検索用クエリを作成し，WEB ページを収集する方法を試した．音声認識結果のみから作成したインデックスと，収集した WEB ページのみから作成したインデックスを用い，2 つのインデックスからの検索結果を線形補間することで，未知語クエリに対して検索精度が改善することを示した．しかし，1 つの音声ドキュメントから 1 つの WEB 検索クエリを作成し WEB ページを収集すると，音声ドキュメントに複数の話題が含まれているために，検索を助ける情報を含む WEB ページをうまく収集することができなかった．そのため，全クエリに対してはドキュメント拡張により精度が改善されたとはいえなかった．

そこで，音声ドキュメントに複数の話題が含まれていることを利用し，音声ドキュメントの自動分割に基づく WEB ページ収集を行った．検索実験の結果，WEB ページによるドキュメント拡張の効果をさらに高めることができた．

しかし，まだいくつかの課題が残る．実験では，検索クエリごとに最適な重みを設定することで，検索精度が大幅に向上させることが示された．今後は，クエリごとに最適な重みを決定する枠組みを検討する必要がある．

⋆1 11 個の未知語クエリ中，10 個のクエリで最適な重みを決定し，残り 1 個のクエリにその重みを適用して検索精度を計算し，これを 11 回繰り返す．

参考文献

1) Cheng, Y.J. and Chen, H.H.: Aligning Words from Speech Recognition and Shots for Video Information Retrieval, *Proc. TRECVID 2004* (2004).
2) Akiba, T., Aikawa, K., Itoh, Y., Kawahara, T., Nanjo, H., Nishizaki, H., Yasuda, N., Yamanashita, Y. and Itou, K.: Construction of a Test Collection for Spoken Document Retrieval from Lecture Audio Data, *IPSJ Journal*, Vol.50, No.2, pp.1234–1245 (2009).
3) 岩田耕平，伊藤慶明，小嶋和徳，石亀昌明，田中和世，李　時旭：語彙フリー音声文書検索手法における新しいサブワードモデルとサブワード音響距離の有効性の検証，情報処理学会論文誌，Vol.48, No.5, pp.1990–2000 (2007).
4) Wechsler, M., Munteaun, E. and Schauble, P.: New Techniques for Open-Vocabulary Spoken Document Retrieval, *Proc. ACM SIGIR'98*, pp.20–27 (1998).
5) Ng, K. and Zue, V.W.: Subword-based approaches for spoken document retrieval, *Speech Communication*, Vol.32, No.3, pp.157–186 (2000).
6) Cheng Pan, Y., Lin Chang, H., Chen, B. and Shan Lee, L.: Subword-based Position Specific Posterior Lattices (S-PSPL) for Indexing Speech Information, *Proc. INTERSPEECH 2007*, pp.318–321 (2007).
7) 胡　新輝，呉　友政，柏岡秀紀：Confusion Network を用いた音声ドキュメントの検索及び評価に関する研究，第 2 回音声ドキュメント処理ワークショップ講演論文集，豊橋技術科学大学メディア科学リサーチセンター，pp.85–90 (2008).
8) Kim, W. and Hansen, J.H.L.: Advances in SpeechFind: Transcript Reliability Estimation Employing Confidence Measure based on Discriminative Sub-word Model for SDR, *Proc. INTERSPEECH 2007*, pp.2409–2412 (2007).

9) Akiba, T. and Yokota, Y.: Spoken Document Retrieval by Translating Recognition Candidates into Correct Transcriptions, *Proc. INTERSPEECH 2008*, pp.2166–2169 (2008).
10) 重安幸治，南條浩輝，吉見毅彦：日本語講演音声ドキュメント検索における索引付けの検討，情報処理学会研究報告，2009-SLP-76, No.8, 情報処理学会 (2009).
11) Terao, M., Koshinaka, T., Ando, S., Isotani, R. and Okumura, A.: Open-Vocabulary Spoken-Document Retrieval Based on Query Expansion Using Related Web Documents, *Proc. INTERSPEECH 2008*, pp.2171–2174 (2008).
12) Mamou, J. and Ramabhadran, B.: Phonetic Query Expansion for Spoken Document Retrieval, *Proc. INTERSPEECH 2008*, pp.2106–2109 (2008).
13) Singhal, A. and Pereira, F.: Document Expansion for Speech Retrieval, *Proc. ACM SIGIR'99*, pp.34–41 (1999).
14) 増村　亮，伊藤　仁，伊藤彰則，牧野正三：WWW を利用した言語モデル適応のための検索クエリ構成の検討，情報処理学会研究報告，2010-SLP-76, No.10, 情報処理学会 (2010).
15) Maekawa, K.: Corpus of Spontaneous Japanese: Its Design and Evaluation, *Proc. ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition* (2003).
16) Akiba, T., Aikawa, K., Itoh, Y., Kawahara, T., Nanjyo, H., Nishizaki, H., Yasuda, N., Yamashita, Y. and Itoh, K.: Test Collections for Spoken Document Retrieval from Lecture Audio Data, *Proc. 6th edition of the Language Resources and Evaluation Conference* (*LREC*) (2008).
17) 高野明彦，西岡真吾，丹羽芳樹：連想に基づく情報アクセス技術：汎用連想計算エンジン GETA を用いて，情報の科学と技術，Vol.54, No.12, pp.634–639 (2004).
18) Singhal, A., Buckley, C. and Mitra, M.: Pivoted document length normalization, *Proc. ACM SIGIR'96*, pp.21–29 (1996).
19) 秋葉友良，相川清明，伊藤慶明，河原達也，南條浩輝，西崎博光，安田宜仁，山下洋一，伊藤克亘：音声ドキュメント検索テストコレクションの試作と基本検索性能評価，第 1 回音声ドキュメント処理ワークショップ講演論文集，豊橋技術科学大学メディア科学リサーチセンター，pp.73–80 (2007).
20) Utiyama, M. and Isahara, H.: A Statistical Model for Domain-Independent Text Segmentation, *Proc. 9th ECACL*, pp.491–498 (2001).

(平成 23 年 4 月 8 日受付)
(平成 23 年 9 月 12 日採録)

西崎　博光（正会員）

昭和 50 年生．平成 15 年豊橋技術科学大学大学院工学研究科電子・情報工学専攻修了．同年山梨大学大学院医学工学総合研究部助手．平成 19 年同大学助教．音声言語処理，音声インタフェースに関する研究に従事．博士（工学）．IEEE，日本音響学会，電子情報通信学会，日本教育工学会各会員．

杉本樹世貴

昭和 61 年生．平成 23 年山梨大学大学院医学工学総合教育部修士課程コンピュータ・メディア工学専攻修了．現在は，東芝ソリューション株式会社勤務．在学中は音声ドキュメント検索の研究に従事．平成 21 年 FIT ヤングリサーチャー賞受賞．

関口　芳廣（正会員）

昭和 23 年生．昭和 48 年山梨大学大学院工学研究科電子工学専攻修了．同年同大学工学部計算機科学科助手．現在，同大学大学院医学工学総合研究部教授．音声情報処理等の研究に従事．工学博士．電子情報通信学会，電気学会，日本音響学会各会員．