単語組み合わせ素性を用いたベクトル空間法による音声ドキュメント検索

瀧上 智子[1,a)]　秋葉 友良[1,b)]

概要：音声ドキュメント処理において，音声認識誤りや認識語彙外語 (未知語) への対応は，しばしば問題に上がる共通の課題である．音声ドキュメントの内容検索においても，検索対象の音声ドキュメントを音声認識する際に導入される認識誤りは検索性能を低下させる原因となる．しかし，従来法では誤りの無いテキストを対象とした検索手法をそのまま適用することが多かった．本研究では，ベクトル空間法を拡張して，誤りを含む文書を対象とした新しい検索モデルを提案する．提案法は，語の共起を利用することで確実な手掛かりを重要視することで，認識誤りが含まれる文書においても頑健に文書間類似度を求めることを可能にする．評価実験の結果，音声内容検索の典型的な手法，および筆者らが以前提案した STD を前処理に用いた音声内容検索法，どちらにおいても提案検索モデルの導入により検索精度が大きく改善することが分かった．

Spoken Content Retrieval by vector space model using term combination features

Tomoko Takigami[1,a)]　Tomoyosi Akiba[1,b)]

Abstract: How to deal with speech recognition errors and out-of-vocabulary (OOV) words is one of the challenging problems common in spoken document processing. Even for spoken content retrieval (SCR), one of the major causes for degrading the retrieval accuracy is the recognition error brought when transcribing the target spoken data into text. However, most of the previous works for SCR simply made use of the conventional retrieval methods, which were originally designed for error-free text data. In this work, we propose a novel retrieval model especially designed for text including errors, by extending conventional vector space model for text retrieval. The proposed method is unique in using the term co-occurrences in order to put emphasis on reliable clues in transcribed text for the similarity calculation, which results in working robust for documents including errors. Through our experimental evaluation, we found that the introduction of the proposed method into both the conventional SCR method and the STD-based SCR method, which previously had been proposed by us, did improve the retrieval accuracy.

1. はじめに

近年，マルチメディアコンテンツの情報爆発の進行に伴い，これらのコンテンツへアクセスするための手段が求められている．多くのマルチメディアコンテンツはメタデータが付与されておらず，従来のテキスト検索だけでは求める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して音声をテキストに変換した上でテキスト検索の適用が可能となる．このような音声に含まれる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」と呼ばれる．テキスト検索と異なり，音声ドキュメント検索では，音声認識で導入される認識誤り，特に音声認識の未知語への対処，が主な技術課題となる．

音声ドキュメント検索のうち，一つの検索語を入力として，その検索語が出現する音声ドキュメント中の位置を特定する問題は，音声中の検索語検出 (Spoken Term Detection,

[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

STD) と呼ばれる．STD は，テキスト検索における文字列照合に対応する問題である．一方，テキスト検索における内容検索に対応する問題は，音声内容検索 (Spoken Content Retrieval, SCR) と呼ばれる．SCR は，キーワードリストや自然言語文で表現した検索クエリを入力として，それが表す内容を含む文書を見つける問題である．

我々は，音声内容検索における未知語・認識誤りへの対策として，STD を前処理に用いた SCR 手法を提案している [1]．この手法は，検索クエリに含まれる語それぞれについて，検索対象文書に対して STD を実行しその出現位置を特定した後，語の出現頻度などの統計情報を用いて文書検査を行う手法であった．認識誤りや未知語やに頑健な STD 手法を前処理に用いることで，認識誤りや未知語に頑健な音声内容検索が実現できることが示されている．

一方，提案法では語の湧き出し誤りへの対応が求められる．大語彙連続音声認識で書き起こしたテキストを検索対象とする従来の典型的な音声内容検索手法では，本来音声ドキュメント中に出現する語を認識誤りのために見逃してしまう第１種の誤り (偽陰性, false negative) に対処する必要があるが，提案法ではさらに，本来出現していない語を STD で誤検出してしまう第２種の誤り (偽陽性, false positive) も考慮に入れる必要がある．また，従来のテキスト検索では検索対象文書にこれらの誤りが含まれていることは前提としていないため，音声内容検索にテキスト検索用の検索手法をそのまま用いることには限界があると考えられる．

本研究では，テキスト検索では生じないこれらの誤りを考慮に入れた，音声内容検索のための新しい検索モデルを提案する．提案手法は，テキストを対象とした検索モデルとして広く用いられているベクトル空間法を拡張し，単語の共起の情報も類似度計算の手がかりとして利用する．共起情報を用いることにより，文書の文脈から外れて出現する認識誤り語の文書間類似度計算への悪影響を排除するとともに，関連する語の共起情報をより積極的に類似度計算に利用することで，誤りを含む文書を対象とした文書検索の性能を改善することが期待できる．

2. STD を前処理に用いた音声ドキュメントの内容検索

2.1 音声内容検索の従来手法

一般的な音声ドキュメント検索システムでは，次のような処理を行っている．まず，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．そして，得られたテキストに形態素解析，ストップワード除去を適用して単語集合へと変換する．同時に高速化のために索引付けを行う．質問文が与えられると，これをドキュメントと同様に単語集合へと変換し，単語の文書中での出現頻度などの統計情報に基づいて，各文書と質問文との関連度を計算，関連度の高い文書から検索結果として出力する．

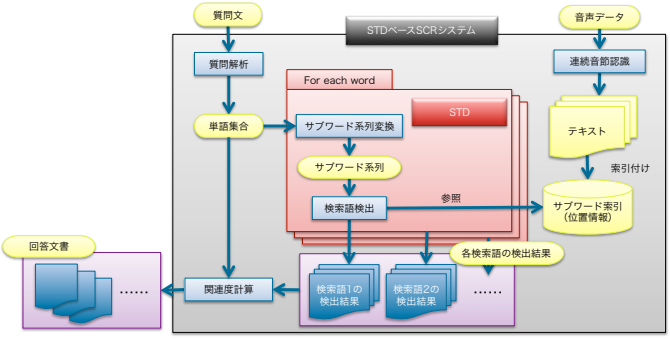

図 1 STD-SCR system.

従来の音声ドキュメント検索では，音声認識の結果にそのままテキストベースの検索を用いていたため，質問文に未知語が含まれていた場合には，その語を手がかりに検索を行うことはできない．これを解決するために，先行研究ではクエリまたはドキュメントの拡張手法 [2], [3], [4] などが提案されている．これらの手法は，正しく認識された語を手がかりとして関連語を拡張することで認識誤りに間接的に対処しており，実際にドキュメント中の未知語や認識誤りの音声区間は利用していない．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利用する手法が提案されている [5], [6], [7] ．一般的な文書検索が文書ベクトルの構成要素として単語を用いるのに対し，これらの研究ではサブワード n-gram を用いる．しかし，サブワード n-gram は言語知識を用いずに自動的に抽出した単位のため，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．

2.2 STD を前処理に用いた音声内容検索手法

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．そこで，検索質問文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，我々は，文書検索の前処理として STD を適用する手法を提案した [1]．提案システムを図 1 に示す．

基本的なシステムは次の通りである．まず，音声データに対し，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．与えられた質問文も単語集合に変換した後，各語をサブワード表現へと変換し，これらの語を検索語として，STD を行い，各語の出現頻度などの統計情報を得る．この統計情報に基づき，各文書と質問文との関連度を比較し，関連度の高い文書から検索結果として出力する．STD によって質問文中の各語の検出位置を求め，各検出位置を語の 1 回の出現と見なすことで文書ごとの各語の出現頻度を求めることができる．この統計量に基づき，各

文書と質問文との関連度の比較を行う．

提案法は音素 n-gram を手がかりとする先行研究 [5], [6], [7] とは異なることに注意されたい．提案法は，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っている．また，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 [8], [9]．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．

本研究ではサブワードとして音節を選択し，音声データに対し連続音節認識を行った．STD 手法には音節 bigram 索引と連続 DP マッチングとを併用する手法を用いた．

2.2.1 STD 手法

連続 DP マッチングは，STD において最も基本的な手法である．その距離行列の計算には以下の式を用いた．

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

ここで i は検索語の位置，j は検索対象の位置を表し，$d(a_i, b_j)$ 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．

連続 DP マッチングは高精度の検出が可能であるが，検索語が入力された後に検索対象の音声ドキュメント全体を走査する必要があるので実行時の計算コストが高い．SCR 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，全ての音声データに対して連続 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い，音声内容検索に用いるのは現実的ではない．また，検索語が増加するような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，見込みの無い発話については 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 $T = t_1 t_2 \ldots t_n$ に対し，音節 bi-gram $B_T = \{(t_i, t_{i+1}) \mid i = 1, 2, \ldots, n-1\}$ を作成する．索引はこの音節 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 $w = w_1 w_2 \ldots w_m$ に対して，同様にして検索語の音節 bi-gram B_w を作成する．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 はあらかじめ定めた閾値である．今回 θ は DP マッチングの閾値によらず一定であるとした．

$$
\frac{\sum_{b \in B_w} \delta(b \in B_u)}{|B_w|} \ge \theta \tag{2}
$$

$$
\delta(x \in X) = \begin{cases} 1 & (x \in X) \\ 0 & (\text{otherwise}) \end{cases} \tag{3}
$$

2.3 混合システム

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出来るが，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．まずオフライン処理として，従来法で用いる単語レベルの自動書き起こしテキストと，提案法で用いるサブワード系列の自動書き起こしテキストの二種類のテキストを得る．検索時には，クエリを単語集合に変換し，従来法，提案法それぞれの方法で二つの文書ベクトルを取得し，これを線形に結合して各文書の関連度を求める．

二つの文書ベクトルを基にした関連度の計算は以下の式によって行う．

$$
\begin{aligned}
sim(q,d) = {} & (1-\alpha)\{sim_{cSDR}(q,d)\} \\
& +\alpha\{(1-\beta)\, sim_{STD-SDR}(q_{IV}, d) \\
& +\beta sim_{STD-SDR}(q_{OOV}, d)\}
\end{aligned} \tag{4}
$$

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．

3. 音声内容検索のための検索モデル

文書検索においてクエリと各文書の類似度を求める手法としては，ベクトル空間モデルが広く用いられている．テキスト検索と異なり，音声ドキュメントの検索では，認識誤りのために各語の出現頻度は誤りを含んでいる可能性があるため，確実な手がかりとして利用することはできない．そこで，不確実な手がかりに頑健な，ベクトル空間法を拡張した検索モデルを提案する．

音声認識誤りによる文書ベクトルの誤りには 2 種類ある．一つは，本来は文書中に出現しているが，認識結果には出現しないという誤り，もう一つは，本来は文書中に出

現していないが，認識結果には出現しているという誤りである．前者は偽陰性 (false negative)，後者は偽陽性 (false positive) と呼ばれる．偽陰性は SCR の性能において再現率に影響し，偽陽性は精度に影響する．従来の SCR の研究では，主に偽陰性への対処について取り組まれていた．クエリや検索対象文書の拡張は，単語を増やすことで，偽陰性によって検出できなくなっている文書に対処している手法である．

一方，偽陽性についてはこれまであまり取り組まれていない．これは，クエリに関連しない文書に対して，クエリ中の複数の語が何度も誤って出現することは考えにくく，関連度の計算を行う段階で，淘汰されるためである．しかし，提案手法では STD を行っていることから，湧き出し誤りによって偽陽性が増加していると考えられる．そのため，偽陽性に対する積極的な対処が必要となる．また，短い文書に対する検索については，含まれる単語数が少ないことから，偽陽性の影響が大きく，この場合にも対処が必要であると考えられる．

3.1 単語組み合わせを素性に用いた検索モデル

クエリ中に含まれる重要語同士は文脈的に関連しており，一定の発話区間において，同時に出現しやすい語であると考えることができる．反対に，語が単独で出現している発話区間は，認識誤りである可能性が高くなる．そこで，単語同士が共起しているかどうかを類似度計算の素性に用いることを考える．

文書 d について，文書ベクトルは以下のように求められる．

$$v(d) = [tf(w_1, d), tf(w_2, d), \dots] \tag{5}$$

ここで $tf(w, d)$ は文書 d における単語 w の出現頻度を表す．このベクトルに次の共起情報ベクトルを連結する．

$$v_c(d) = [\delta(w_1, w_2, d), \delta(w_1, w_3, d), \dots, \delta(w_i, w_j, d), \dots] \tag{6}$$

$$\delta(w_i, w_j, d) = \begin{cases} 1 & (w_i \in d \text{ かつ } w_j \in d) \\ 0 & (\text{otherwise}) \end{cases} \tag{7}$$

$\delta(w_i, w_j, d)$ は単語 w_i と w_j が文書 d に同時に出現しているかどうかを表すデルタ関数である．共起情報ベクトル $v_c(d)$ のサイズは $|V|$ を語彙サイズとすると，$|V|^2 - |V|$ となる．クエリ $q = w_1, w_2, w_3$ に対する組み合わせ素性による文書ベクトルの例を図 2 に示す．

検索時には，システムはまず，文書 d に対してクエリ中に含まれる単語の出現頻度を取得し，$v(d)$ を得る．次にこの $v(d)$ に基づいて，$v_c(d)$ を求める．この二つのベクトルを連結してベクトル $[v(d), v_c(d)]$ を求め，このベクトル間でクエリと文書間の類似度を計算して[*1]，最終的な出力を得る．

	$v(d)$			$v_c(d)$		
	w_1	w_2	w_3	w_1w_2	w_1w_3	w_2w_3
d_1	3	0	1	0	1	0
d_2	1	2	2	1	1	1
d_3	0	0	1	0	0	0

図 2 拡張ベクトル空間モデルによる文書ベクトルの例

4. 実験

4.1 テストコレクション

NTCIR-9 SpokenDoc の SDR サブタスク [10] で用いられたテストコレクションに対して検索性能評価実験を行った．このテストコレクションは CSJ に収録されている音声のうち，2702 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．各クエリはパッセージと呼ばれる可変長区間の発話を正解としている．

このテストコレクションに対して次の 2 つのタスクで評価を行った．

講演検索タスク 検索の単位を講演とし，正解パッセージを含む講演を正解文書とする．この場合，検索対象文書数は 2,702 となる．

疑似パッセージ検索タスク [11] 各講演を予め先頭から 15 発話ずつの重なりの無い発話列に区切っておく．この 15 発話区間 (疑似パッセージと呼ぶ) を文書とみなして検索の単位とし，正解パッセージとオーバーラップする疑似パッセージを正解文書とする．この場合，検索対象文書数は 60,202 となる．

4.2 音声認識

検索対象である音声ドキュメントの音声認識結果には，以下の 2 種類の認識条件で得られた連続単語認識結果と連続音節認識結果を用いた．一つは NTCIR-9 SpokenDoc[10] 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．

もう一つの認識結果は言語モデルを新聞記事 75 ヶ月分で学習したもの [12] 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

[*1] 共起情報ベクトルについても IDF 重み付けを行った．

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．

4.3 比較手法

提案検索モデルの検索性能を調査するために，音声内容検索法として 2.1 節で述べた従来法 (以降，CSCR) と，2.2 節で述べた STD を前処理に用いる SCR 手法 (以降，STD-SCR)，2.3 節で述べた SCR と STD-SCR の混合システム (CSCR-STD-SCR) を実装し，それぞれについて検索モデルとして従来ベクトル空間モデルを用いた場合と提案検索モデルを用いた場合を比較した．

CSCR 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．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 の二種類を用いた．

STD-SCR の検索語検出には，2.2.1 節で述べた索引付けと連続 DP マッチングの併用を用いた．式 (1) の音節間距離 $d(a_i, b_j)$ には，文献 [13] で用いられた音響モデル間の Bhattacharyya 距離を用いた．沸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．

4.4 評価尺度

検索性能の評価尺度には MAP (Mean Average Precision) [14] を用いた．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．

$$MAP = \frac{1}{|Q|} \sum_{q \in Q}^{Q} AveP(q) \tag{8}$$

ここで，Q はクエリ集合を表す．$AveP(q)$ はクエリ q の平均精度を表し，以下の式で定義される．

$$AveP(q) = \frac{1}{|R_q|} \sum_{i=1}^{|O_q|} \left(\delta\left(O_{q_i}|R_q\right) \frac{\sum_{j=1}^{i} \delta\left(O_{q_j}|R_q\right)}{i} \right) \tag{9}$$

ここで，R_q はクエリ q の正解文書集合を表し，O_q はクエリ q に対するシステムの順序付き出力リストを表す．また，$\delta(o)$ は出力文書が正解かどうかを表すバイナリ値で，以下のように定義される．

$$\delta(o|R) = \begin{cases} 1 & (o \in R) \\ 0 & (o \notin R) \end{cases} \tag{10}$$

検索性能は全てのクエリ（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．matched の既知語数は約 27000 語，unmatched の 20000 語であり，それ以外の語が未知語となる．認識辞書が異なるため，matched と unmatched では IV（OOV）クエリの分類が異なり，OOV クエリの数は matched が 8 クエリ，unmatched が 41 クエリである．

4.5 実験結果

講演検索タスクに対する，従来法と単語組み合わせ素性を用いた提案法との検索性能の比較を表 1，表 2 に示す．全体的に従来法に比べ，提案法において性能の改善が見られ，単語の共起情報がより確実な手がかりとして利用できることが示された．特に，matched での改善が大きく，ALL クエリでは，CSCR(単語) で 11.4%，STD-SCR で 11.5%，CSCR-STD-SCR では 10.8% の改善であった．一方，unmatched では CSCR（単語）で 4.9%，STD-SCR で 0.4%，CSCR-STD-SCR で 7.7% と matched に比べ改善は小さい．しかしながら，CSCR-STD-SCR の OOV では 19.5% の改善が見られた．

次に疑似パッセージ検索タスクでの検索性能を表 3，表 4 に示す．疑似パッセージ検索タスクは講演検索タスクに比べて難しく，全ての手法で検索性能は低下している．しかし，講演啓作と同様に，提案手法による性能改善が見られた．matched の ALL クエリでは，CSCR(単語) で 24.3%，STD-SCR で 17.0%，CSCR-STD-SCR では 19.0% の改善，unmatched では，CSCR(単語) で 19.7%，STD-SCR で 13.6%，CSCR-STD-SCR では 20.6% の改善が見られた．

提案法は，講演検索タスクよりも疑似パッセージ検索タスクにおいて高い改善率が見られた．提案法は単語同士の共起の判定を文書の単位で行っているため，講演検索タスクでは講演全体での共起を，疑似パッセージ検索タスクでは 15 発話での共起を見ていることになる．そこで，講演検索タスクにおいて共起の判定をより小さい単位で行うことで性能の改善が見込めるのではないかと考え，単語の共起判定を行う範囲を変えたときの検索性能の変化を調べた．検索システムには STD-SCR を用い，STD の閾値を変化させて検索した結果を図 3 に示す．

STD の閾値が小さく，湧き出し誤りの少ない範囲では，100 発話で共起を見た場合が最も性能が良く，STD の閾値 0.2 の場合で 7.4% の改善が見られた．一方，STD の閾値を大きくし，湧き出し誤りの多い範囲では，小さい範囲で共起を判定した場合の性能が良く，STD の閾値 0.7，共起判定 1 発話で 20.8%，15 発話で 16.7% の改善が見られた．

表 1 講演検索タスク実験結果（matched）（共起判定単位:文書全体）

	従来法			提案法		
	ALL	IV	OOV	ALL	IV	OOV
CSCR（単語）	0.350	0.358	0.235	0.390	0.400	0.250
CSCR（音節 bi-gram）	0.214	0.213	0.226	0.230	0.230	0.227
STD-SCR	0.253	0.250	0.284	0.282	0.281	0.291
CSCR-STD-SCR	0.369	0.371	0.350	0.409	0.412	0.375

表 2 講演検索タスク実験結果（unmatched）（共起判定単位:文書全体）

	従来法			提案法		
	ALL	IV	OOV	ALL	IV	OOV
CSCR（単語）	0.307	0.373	0.149	0.322	0.393	0.150
CSCR（音節 bi-gram）	0.194	0.236	0.105	0.198	0.238	0.113
STD-SCR	0.273	0.291	0.237	0.274	0.290	0.241
CSCR-STD-SCR	0.349	0.384	0.277	0.376	0.399	0.331

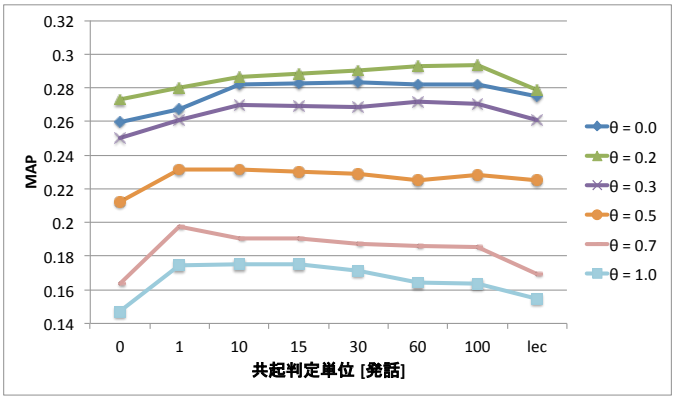

図 3 共起判定単位と検索性能（講演検索タスク）（unmatched）

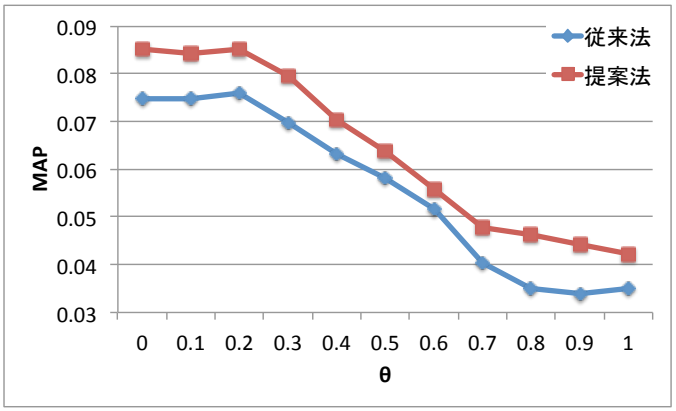

図 4 STD 閾値と検索性能（疑似パッセージ検索タスク）（unmatched）

次に，STD の閾値に対する従来法と提案法の性能を比較した．タスクは疑似パッセージ検索タスク，検索対象は unmatched，共起判定単位は文書 (15 発話) である．実験結果を図 4 に示す．閾値を下げて湧き出し誤りが増加する状況においても，提案法によって検索性能が改善されることが分かる（閾値 0.2：12.3%，閾値 0.8：32.4%）．

5. まとめ

本研究では，音声ドキュメントの内容検索において，検索対象の音声ドキュメントに対する音声認識誤りに対処するための，ベクトル空間法を拡張した新しい検索モデルを提案した．提案法は，語の共起を利用することで確実な手掛かりを重要視することで，認識誤りが含まれる文書においても頑健に文書間類似度を求めることを可能にする．評価実験の結果，音声内容検索の典型的な手法，および筆者らが以前提案した STD を前処理に用いた音声内容検索法，どちらにおいても提案検索モデルの導入により検索精度が大きく改善することが分かった．

参考文献

[1] 瀧上智子，秋葉友良：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，情報処理学会論文誌， Vol. 54, No. 2, pp. 1–12 (2013).

[2] Akiba, T. and Honda, K.: Effects of Query Expansion for Spoken Document Passage Retrieval, *Proceedings of International Conference on Speech Communication and Technology*, pp. 2137–2140 (2011).

[3] 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).

[4] 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).

[5] Chen, B., min Wang, H. and shan Lee, L.: Discriminating capabilities of syllable-based features and approaches of utilizing them for voice retrieval of speech information in Mandarin Chinese, *IEEE Transactions on Speeh and Audio Processing*, Vol. 10, pp. 303–314 (2002).

[6] Pan, Y.-c. and Lee, L.-s.: Performance Analysis for Lattice-Based Speech Indexing Approaches Using Words and Subword Units, *Trans. Audio, Speech and Lang. Proc.*, Vol. 18, No. 6, pp. 1562–1574 (2010).

[7] 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

表 3 疑似パッセージ検索実験結果（matched ）（共起判定単位:文書全体）

	従来法			提案法		
	ALL	IV	OOV	ALL	IV	OOV
CSCR（単語）	0.103	0.108	0.0387	0.128	0.134	0.0399
CSCR（音節 bi-gram）	0.0690	0.0696	0.0603	0.102	0.102	0.0103
STD-SCR	0.0707	0.0699	0.0828	0.0827	0.0820	0.0933
CSCR-STD-SCR	0.116	0.119	0.0665	0.138	0.140	0.108

表 4 疑似パッセージ検索実験結果（unmatched ）（共起判定単位:文書全体）

	従来法			提案法		
	ALL	IV	OOV	ALL	IV	OOV
CSCR（単語）	0.0927	0.112	0.0468	0.111	0.138	0.0452
CSCR（音節 bi-gram）	0.0659	0.0772	0.0421	0.0919	0.104	0.0659
STD-SCR	0.0744	0.0795	0.0639	0.0845	0.0885	0.0762
CSCR-STD-SCR	0.107	0.123	0.0743	0.129	0.148	0.0893

(2012).

[8] Jones, G. J. F., Foote, J. T., Jones, K. S. and Young, S. J.: Retrieving Spoken Documents by Combining Multiple Index Sources (1996).

[9] Wechsler, M., Munteanu, E. and Schäuble, P.: New techniques for open-vocabulary spoken document retrieval, *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '98, New York, NY, USA, ACM, pp. 20–27 (1998).

[10] 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).

[11] 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).

[12] 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).

[13] 山本一公，中川聖一：発話スタイルによる話速・音韻間距離・ゆう度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(11), pp. 2438–2447 (2000).

[14] 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, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)* (Chair), N. C. C., Choukri, K., Declerck, T., Doan, M. U., Maegaard, B., Mariani, J., Odijk, J. and Piperidis, S., eds.), Istanbul, Turkey, European Language Resources Association (ELRA) (2012).