情報処理学会論文誌　Vol. 50　No. 2　514–523 (Feb. 2009)

認識候補から正解テキストへの翻訳に基づく講演音声ドキュメントのアドホック検索

秋葉友良[†1]　横田悠右[†1]

音声ドキュメントを対象としたアドホック検索は，大語彙連続音声認識を用いて対象音声ドキュメントをテキストへと変換すれば，既存のテキストを対象とする検索手法がそのまま適用可能である．その際に問題となるのは，音声認識誤りの影響による自動書き起しテキストの劣化の問題がある．この問題に対し，本研究では，音声認識による自動書き起しと人手書き起しの間の差異を「翻訳」によって補完する検索手法を提案する．提案手法の効果を調べるため，日本語話し言葉コーパス（CSJ）を検索対象としたテストコレクションで評価実験を行った．その結果，提案手法は，特に小さいサイズの文書を検索対象とするタスクにおいて良い性能を示すことが分かった．

An Ad-hoc Retrieval Method for Lecture Speech by Translating Recognition Candidates into Correct Transcriptions

Tomoyosi Akiba[†1] and Yusuke Yokota[†1]

This paper proposes an ad hoc retrieval method for spoken documents that uses a statistical translation technique. After transcribing the spoken documents by using a Large-Vocabulary Continuous Speech Recognition (LVCSR) decoder, a text-based ad hoc retrieval method can be directly applied to the transcribed documents. However, recognition errors will significantly degrade the retrieval performance. To address such a problem, the proposed method aims to fill the gap between the automatically transcribed text and the correctly transcribed text by using a statistical translation technique. To investigate the effectness of the proposed method, we conducted an ad hoc retrieval experiment targeting the Corpus of Japanese Spontaneous Speech.The experimental evaluation shows that the proposed method performs better than the baseline ad hoc retrieval method using only the transcribed text, especially for retrieval tasks with relatively small target documents.

1. はじめに

音声認識技術の高度化にともない，本来コミュニケーションの手段として用いられてきた話しことばによる音声を，知識や技術を伝達するメディアとして利用することが可能になってきた．我々が特別な道具を要することなく日常的に発する音声を，年々大容量化するストレージに蓄え，高速化するネットワークを介して容易にアクセス可能な「文書」として扱うことができれは，ことばを利用した人間の知的活動は飛躍的に拡大するであろう．このような「音声ドキュメント処理」にとって鍵となる研究課題は，検索，要約，コンテンツ生成，など多岐にわたるが，本稿では音声ドキュメントの検索に焦点を当てる．

音声ドキュメントの検索は，米国 NIST 主催の評価型ワークショップ TREC において，1997 年から 2000 年の間 Spoken Document Retrieval（SDR）Track において，大規模な評価実験が行われている[1)]．TREC SDR では，英語の放送ニュース音声を対象に，最終的には約 557 時間の音声ドキュメントを対象とした評価用テストコレクションが構築された．一方日本では，情報処理学会音声言語情報処理研究会の「音声ドキュメント処理ワーキンググループ」の活動として，「日本語話し言葉コーパス」[2)]（以下，CSJ と略す）を対象とした音声ドキュメント検索評価用テストコレクション（以下，CSJ テストコレクション）が構築されつつある[3),4)]．本研究では，CSJ テストコレクションを対象に評価実験を行う．

音声処理の分野において「検索」という日本語は，既知の語を見つけることを目的とした，いわゆる「キーワード検索」の意味で使われることが多い．また，TREC SDR 以降，音声ドキュメントを対象とした検索の研究は，「キーワード検索」を対象としているものがほとんどである[⋆1]．一方，情報検索や言語処理分野で主な研究対象であるのは，検索者の情報要求を表す表現（たとえば，キーワードリストや自然言語文）から，その情報要求を満たす内容を持つ文書を検索する「アドホック検索」である．前述の TREC SDR においても，初年度は「既知語の検索」（Known Item Search）をタスクとしたが，2 年目以降はアドホック検索をタスクとしてきた．

音声ドキュメントを対象としたアドホック検索は，大語彙連続音声認識を用いて対象音声ドキュメントをテキストへと変換すれば，既存のテキストを対象とする検索手法がそのまま

†1 豊橋技術科学大学
Toyohashi University of Technology

⋆1 特定のキーワードが含まれる文書を見つける検索タスクを扱った研究（たとえば，文献 5)）も行われているが，人手判定した正解文書を見つけるアドホック検索とはタスクが異なる点に注意されたい．

適用可能である．その際に問題となるのは，音声認識にともなう自動書き起しテキストの劣化の問題である．現在の音声認識の精度は，あらかじめ用意された原稿の読み上げ音声では 95%前後の単語認識率が得られているが，本稿で対象とする講演のような自発的発話では 50%〜70%程度の認識率であることが多く，まだ十分な認識精度は達成されていない．また，今後音声認識技術の発展により認識率が向上しても，100%の認識率を達成することは本質的に困難である．さらに，現在の典型的な大語彙連続音声認識は数万語の認識辞書で構成されており，辞書にない語を認識することはできないという認識語彙外語の問題もある．本研究では，音声ドキュメントのアドホック検索における，このような誤認識の影響への対処法に焦点を当てる．

音声ドキュメントを対象としたキーワード検索の従来研究では，認識誤りや認識語彙外語の問題に対処するため，単語よりも小さい単位である音素や音節などの subword を基本単位として認識を行い，それらの列とキーワードのマッチングによって検索を行うアプローチが提案されている[5)–7)]．しかし，正しく認識された場合は単語を基本単位とした方が一般に検索性能は高いことが知られており，そのため subword と単語を併用する手法も提案されている[8),9)]．

一方，本研究で対象とするアドホック検索は，検索クエリによって表される内容を持つ文書を見つけることが目的であり，必ずしも特定のキーワードを含む文書を見つけることが目的ではない．アドホック検索手法としてよく用いられるベクトル空間法は，文書と検索クエリをその内容を表現するベクトル空間にマッピングし，空間上の類似度を計算することで検索を行う．そのため，文書の索引付けには，その内容を構成する単位を索引とするのが有効であり，言語において意味を担う最小単位である単語で索引付けすることには意味があると考えられる．上述したキーワード検索における従来法の，単語より小さい単位（たとえば，subword）を併用して索引付けする手法は，単語部分で扱えなかった（誤認識した）部分を，単語とは別の原理（subword 単位の空間）で扱うことになり，単語で扱う部分と同様の精度で扱えるようになるわけではない．

本研究は，音声ドキュメントのアドホック検索を研究対象とし，誤認識を単語レベルで扱うことを可能にする新規の手法を提案する．提案法は，音声認識による自動書き起しと人手書き起しの間の差異を「翻訳」によって補完し，推定される正解単語による索引付けを行う．認識誤りを含む認識候補から直接単語を予測する点が従来法と異なる新規な点である．テキストを対象とした文書検索において，テキストに現れる表層的な単語に加えてその関連語で拡張したり（たとえば，疑似適合性フィードバック[10)]），単語の代わりに抽象化された概念で索引付けを行う手法（潜在的意味インデキシング[11)]）が提案されている．提案法は，認識結果（表層）に現れる単語の代わりに推定される別の単語（概念）で索引付けする点において，これらの手法に近いと考えられる．

一方，提案法を適用する際には，翻訳モデルを構築するために，音声認識結果と人手書き起しテキストの対からなる学習データが必要である．そのため，従来法とくらべて，学習データ構築にコストがかかるという問題がある．本稿では，必要な学習データの量についても議論する．

その他の関連研究を以下にまとめる．近年，情報検索の分野において，翻訳モデルを用いた検索モデルが提案され，単一言語文書検索（Mono-lingual IR）[12)]，言語横断文書検索（Cross-lingual IR）[13)]，単一言語質問応答（Mono-lingual QA）[14)]，言語横断質問応答（Cross-lingual QA）[15)] に適用されている．提案法は，翻訳モデルに基づく検索モデルを音声ドキュメント検索に適用した手法と位置付けることができる．Ringger と Allen[16)] は，人手書き起しテキストと自動書き起しのペアから認識誤りのチャネルモデルを学習し，音声認識の後処理で誤り修正に利用する手法を示している．本手法は，このチャネルモデルを音声認識のバックエンドとなるアプリケーション（アドホック検索）に直接利用した手法と位置付けることができる．

本稿の構成は，以下のとおりである．まず，2 章で，本稿の評価対象である CSJ テストコレクションの概要を述べる．3 章では，提案法である翻訳モデルを用いた音声ドキュメント検索手法について述べる．4 章では，CSJ テストコレクションを用いた提案法の評価実験について述べる．

2. 音声ドキュメント検索評価用テストコレクション

本章では，CSJ テストコレクションの概要を構成要素ごとに簡単に説明する．詳しくは，文献 3)，4) を参照されたい．

検索対象文書　「日本語話し言葉コーパス」（CSJ）のうち「学会講演」と「模擬講演」を検索対象としている．どちらも独話で自発発話である．両方の講演をあわせると 600 時間を超える．

検索クエリ　講演中の連続する 5 発話程度の可変長の区間を検索対象とする，39 問の検索クエリが設定されている．この検索クエリは複数人によって作成されており，検索対象（講義音声）を対象としたある程度多様な質問を含んでいると考えられる．

正解文書　上記 39 問の検索クエリについて，発話を単位とした可変長の連続する区間に対

して適合性判定が行われている．判定は，適合の程度により，適合（Relevant），部分適合（Partially Relevant），不適合（Irrelevant）のいずれかである．

自動書き起し　大語彙連続音声認識を利用して，検索対象の講演全体に対し，各上位 10 候補の自動書き起しが作成されている．言語モデル，音響モデルは，CSJ 検索対象文書集合の一部を除いた人手書き起しテキストおよび音声データで学習した closed なモデルを利用している．単語認識精度は平均で 78.6%である．

3. 翻訳モデルを用いた音声ドキュメント検索

提案手法は，音声認識による自動書き起しテキストに対し，人手書き起しされた場合に現れるであろう語を使って直接索引付けすることにより，自動書き起しと人手書き起しの差異を補完する．この索引付けには，自動書き起しテキストに現れる単語 e が，人手書き起しテキストにおいて単語 f として現れる確率 $t(f|e)$ を利用する．この確率を，統計的機械翻訳の用語にならって単語翻訳確率と呼ぶ．

3.1 単語翻訳確率の推定

単語翻訳確率 $t(f|e)$ の推定には，音声認識結果の自動書き起しテキストと，人手による書き起しテキストのペアによるパラレルテキストを用いる．

まず，パラレルテキストの両サイドを形態素解析し[⋆1]単語列を得る．次に，この単語列ペアに対し，編集距離を指標とする DP マッチングを適用する．その結果から，自動書き起しと人手書き起しが完全一致する単語どうしについてのみアライメントを抽出し，これを初期アライメントとする．残りの単語，すなわち自動書き起しと人手書き起しが一致しない単語間については，正しいアライメントを求めることは困難なので，次の 2 通りの方法で近似的かつ断片的なアライメントを求めた．

単純分配法　先の初期アライメントと交差・連結しないように，かつ自動書き起し側の各単語がちょうど 1 つの手動書き起し単語とアライメントされるように，可能なアライメントを与えた．その際，各アライメントの出現は 1 回と数えるのではなく，可能なアライメント候補に対して一様となるように断片的な回数で現れたとする．

たとえば，自動書き起しの列 $\cdots e_p e_{p+1} \cdots e_{p+l} e_{p+l+1} \cdots$ と，人手書き起しの列 $\cdots f_q f_{q+1} \cdots f_{q+m} f_{q+m+1} \cdots$ について，e_p と f_q，e_{p+l+1} と f_{q+m+1} に完全一致によるアライメントがある場合，単語アライメント (e, f) の出現回数 $tc(e, f)$ は以下のように与えられる．

$$
\begin{aligned}
&tc(e_p, f_q) = 1 \\
&tc(e_i, f_j) = \frac{1}{m} \quad (p < i \le p+l, q < j \le q+m) \\
&tc(e_{p+l+1}, f_{q+m+1}) = 1
\end{aligned}
$$

単純分配法では，本来のアライメントとしてありえない交差したアライメントも推定の際に考慮に入れてしまうという問題がある．そこで，交差しないアライメントだけから推定を行う次の方法を試みた．

非交差分配法　各単語アライメントどうしは交差しないという制約を設け，可能な非交差アライメントに対して等分配して，単語アライメントの出現回数をカウントする．たとえば，正解単語列「患者/側/も」と認識候補「感情/は/もう」の間の可能な非交差アライメントは，**図 1** のように 10 通り考えられるが，このうち（感情，患者）の単語アライメントは 6 通りのアライメントに現れるので，出現回数は 6/10 と与える．

これら単語アライメントをパラレルテキスト全体で収集し，最尤推定によりパラメータ推

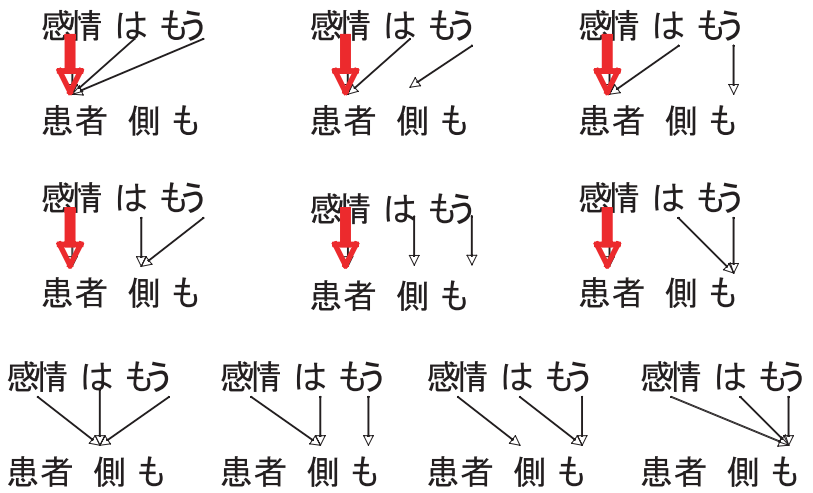

図 1　可能な非交差アライメント
Fig. 1　Possible no-crossing alignments.

⋆1 形態素解析により，活用語については基本型も同時に求め，アライメント作成には表層文字列を，単語翻訳モデルの構築には基本型を，それぞれ用いた．

定を行った．

3.2　単語翻訳確率を用いた音声ドキュメントの索引付け

ある音声ドキュメント D の自動書き起しに現れる単語集合を E_D とする．自動書き起しテキストからそのまま索引付けする場合，単語 $e \in E_D$ の D での単語頻度 $\mathrm{TF}_E(e, D)$ をもとにした統計情報を利用し，たとえば TF-IDF などの単語重みを利用して索引付けが行われる．一方，D を人手書き起しテキスト（単語集合 F_E）で索引付けする場合，同様に，単語 $f \in F_D$ の単語頻度 $\mathrm{TF}_F(f, D)$ をもとに索引付けが行われる．この $\mathrm{TF}_F(f, D)$ の期待値 $E(\mathrm{TF}_F(f, D))$ は，単語翻訳確率 $t(f|e)$ を用いて，以下のように求めることができる．

$$E(\mathrm{TF}_F(f, D)) = \sum_{e \in E_D} t(f|e)\mathrm{TF}_E(e, D) \tag{1}$$

さらに，翻訳モデルで扱えない（学習データに現れない）語を扱うためのスムージングとして，$\mathrm{TF}_E(e, D)$ との線形補間を行う．

$$\widetilde{\mathrm{TF}}_F(f, D) = \lambda E(\mathrm{TF}_F(f, D)) + (1 - \lambda)\mathrm{TF}_E(f, D) \tag{2}$$

この $\widetilde{\mathrm{TF}}_F(f, D)$ を用いて，音声ドキュメントの自動書き起しテキストの索引付けを行う．ただし，閾値 α を設けて，α 以下の $\widetilde{\mathrm{TF}}_F(f, D)$ となる単語では索引付けを行わない．

以上により正解テキストに含まれると期待される単語で索引付けされた音声ドキュメントに対し，既存のアドホック検索手法を適用し，音声ドキュメントの検索を行う．

4. 実　　験

4.1　検索タスク設定

2 章で述べたとおり，CSJ テストコレクションの正解は可変長の発話区間であり，固定の文書集合を対象とする既存のアドホック検索手法をそのまま適用できない．そこで，文献 4) にならって，あらかじめ検索対象の講演を重複のない固定長の発話区間に区切っておき，各区間を独立した文書と見なした文書検索タスクを設定した．

固定長区間としては，15 発話，30 発話，60 発話の 3 通りを試した．2,702 講演を区切った場合，疑似的な文書数はそれぞれ，60,202 文書（15 発話），30,762 文書（30 発話），16,060 文書（60 発話）となる．また，テストコレクションの正解発話が 1 発話でも含まれる区間を，本検索タスクにおける正解文書とした．ここで正解発話とは，「適合」と判定された発話（R 判定），「適合」あるいは「部分適合」と判定された発話（R+P 判定），のいずれかとする．

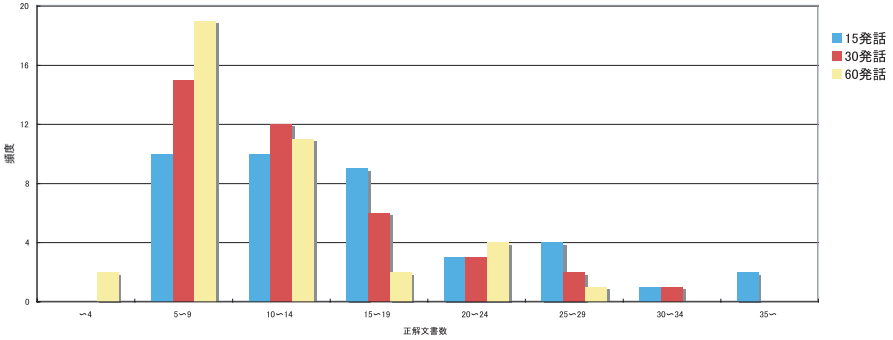

図 2　クエリあたり正解文書数の分布（R 判定）
Fig. 2　The distribution of the relevant documents.

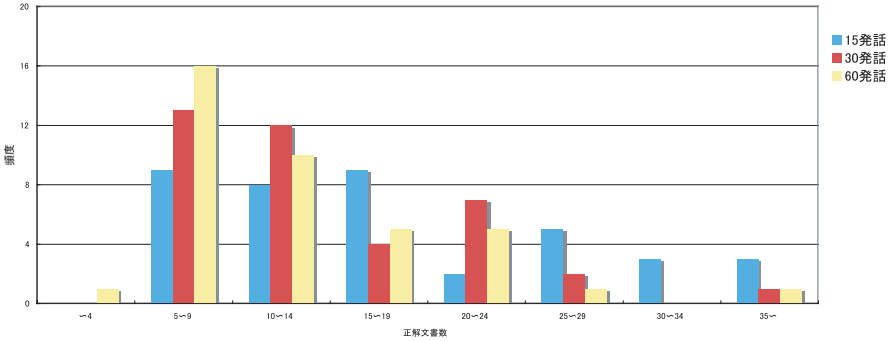

図 3　クエリあたり正解文書数の分布（R+P 判定）
Fig. 3　The distribution of the relevant documents.

CSJ テストコレクションの 39 検索クエリについて，この検索タスクにおける R 判定および R+P 判定正解文書数の分布を，それぞれ**図 2**，**図 3** に示す．

4.2　評 価 尺 度

検索性能の評価尺度には，0.0 から 1.0 まで 0.1 刻みの各再現率レベルにおける補間精度 IP を平均した 11 点平均精度 AP を，全検索クエリでマクロ平均をとった値を用いた[17]．

$$IP(x) = \max_{\{i|x \le R_i\}} P_i \tag{3}$$

$$AP = \frac{1}{11}\sum_{i=0}^{10} IP\left(\frac{i}{10}\right) \quad (4)$$

ここで，R_i および P_i は，それぞれ上位 i 番目までの検索結果に対する再現率と精度である．$IP(x)$ は，再現率 x に相当する精度を，再現率 x 以上における最大の精度で求めた，補間精度である．

この尺度は，様々な検索用途（種々の再現率レベル）における，検索文書数に対する適合文書の割合（検索精度）の平均値を示しており，値が大きいほど検索性能が優れていることを示している．たとえば，値 0.25 は，検索数 4 に対して 1 件の適合文書が平均的に得られることを表す．実際には，1 つの検索クエリに対して上位 1,000 件まで文書を検索して，11 点平均精度を近似的に求めた．

4.3　ベースライン

2 つのベースライン手法を設定し，提案手法との比較を行った．まず，自動書き起しテキストだけを用いた一般的な文書索引付け手法を実装した．また，自動書き起しテキストとして，音声認識結果の 1-best 候補だけを用いた場合と，10-best 候補までを連結して用いた場合を比較した．これらの手法を，以降の実験では「**認識結果のみ**」と記す．

次に，認識語彙外語や認識誤りへの簡易な対応手法として，音素列の類似した単語による文書拡張手法を実装した．まず，検索対象文書の人手書き起こしから，音声認識辞書に登録されていない語を特定した[⋆1]．そして，各認識語彙内語から編集距離を指標に最も類似している認識語彙外語の候補を求める．同じ距離の候補が複数ある場合は，それらをすべて選択する．そして，対象文書の認識結果に対して，対応する認識語彙外語も索引として利用する．この索引付け法を「**未知語拡張**」と記す[⋆2]．また同様に，認識語彙外語に限らずすべての単語について誤認識候補を求めて，文書拡張する手法も実装した．この手法を，「**全単語拡張**」と記す．**未知語拡張**は，誤認識の可能性のうち認識語彙外語だけを扱った手法であり，**全単語拡張**は，すべての誤認識候補を扱った手法である．

また参照手法として，人手書き起しテキストをそのまま索引付けに用いた場合とも比較した（「**人手書き起し**」と記す）．

そのほか，提案法を含むすべての手法に共通して，文献 4) と同様に，以下の設定で索引付けおよび検索を行った．索引付けの単位は形態素とし，活用語の場合は標準形に変換した．ストップワードの設定は行っていない．索引語重みには，文書長で正規化した TF-IDF 重み付け手法[18] を用いた．検索クエリに対しても同様に索引を抽出し，質問ベクトルと文書ベクトルの余弦で順序付けを行うベクトル空間モデルで検索を行った．実装には，文書検索エンジン GETA [19] を用いた．

4.4　提案法の実装

パラレルテキストには，2 章で述べた CSJ 音声ドキュメント検索テストコレクションの大語彙連続音声認識による自動書き起しテキストと，CSJ 書き起しを用いた．対応付けは，以下の手順によるテキストベースの DP マッチングで，CSJ で定義された「発話」ごとに対応付けを行った．

(1)　CSJ 書き起しおよび音声認識結果から，それぞれタグや形態素区切り情報を除去し，テキスト情報だけを抽出する．
ただし，書き起しの発話境界，および「認識結果」の文境界（音声認識の切出区切り）には，後の DP マッチングでそれらどうしが対応付けられることを期待して，共通の特殊な記号[⋆3]を挿入しておく．

(2)　両テキストをそれぞれ形態素解析し，形態素列を得る．

(3)　2 つの形態素列から，編集距離を指標に DP マッチングを行い，形態素単位の対応付けを得る．

(4)　形態素対応付けから，CSJ 書き起しの各「発話」に対応する認識結果形態素列を得る．

また，自動書き起しテキストとして認識候補の 1-best だけを使う場合と，10-best までの認識候補を用いて各発話に対し 10 対の対応付けされたパラレルテキストを使う場合，の 2 通りの方法を比較した．

翻訳モデルのパラメータ学習には，学習データが検索タスクに対してオープンとなるように，次のような交差検定の手法を用いた．まず，学習データである CSJ の 2,702 講演を，講演を単位としてランダムに同サイズの 10 ブロックに分割した．そして，ある 1 ブロック中の音声ドキュメントの索引付けには，残りの 9 ブロックのパラレルテキストから学習した単語翻訳モデルを用いる．この操作を，索引付け 1 ブロックと学習データ 9 ブロックの組合せを切り替えて，10 回繰り返した．

検索エンジンには GETA を用いた．GETA では索引付けの単語頻度は整数で与えなけれ

⋆1 ここでは，認識対象文書集合から認識語彙外語を漏れなく特定できたとする理想的な状況を仮定している点に注意されたい．
⋆2 この手法による未知語カバー率は 99.1%であった．
⋆3 実際には，「改行」を挿入した．直後の形態素解析において，「改行」は固有の一形態素として扱われる．

表 1　スムージング係数 λ による検索性能の比較
Table 1　The retrieval performances with respect to the smoothing coefficient λ.

学習データサイズ	31.4 万形態素		565 万形態素	
λ	1.0	0.5	1.0	0.5
15 発話	0.173	0.196	0.197	0.207
30 発話	0.215	0.244	0.237	0.249
60 発話	0.231	0.259	0.257	0.266

ばならないという制約があるため，実数の値をとりうる $\widetilde{\mathrm{TF}}_F(f, D)$ で索引付けする場合はすべての索引語頻度を定数倍（$1/\alpha$ 倍）し，小数点以下を切り捨てることで，閾値 α の制約を適用した．また，IDF は整数化した索引語頻度をもとに算出した．すなわち，$\widetilde{\mathrm{TF}}_F(f, D) \geq \alpha$ を満たす文書数を単語 f の文書頻度とした[⋆1]．

4.5　翻訳モデルの学習事例

パラレルテキストから学習された高頻度のアライメントの具体例を以下に示す（ただし，自動書き起し→人手書き起しの順）．

- 同音異義語
  感染→観戦，解放→開放，式→四季，創造→想像，下降→加工，ここ→個々，そこ→底
- 発音が類似
  研究→言及，要素→様相，構成→個性，計算→欠損，実験→事件，情報→譲歩，加工→確保，父→土

4.6　実 験 結 果

4.6.1　スムージングの効果

まず，式 (2) のスムージングの補完係数 λ を設定する効果について調べた．スムージングを行うことで，学習データに現れない語（すなわち翻訳モデルで本来の語を予測できない語）に対しても，認識結果の語をそのまま使って索引付けを行うことが可能となり，学習データに対するロバスト性が向上すると期待できる．**表 1** に，R 判定を正解として，$\lambda = 1.0$（スムージングなし）と $\lambda = 0.5$（翻訳モデルの予測と認識結果を同じ重みで考慮した場合）とした場合の検索性能の比較を示す．学習データサイズの違い（31.4 万形態素と 565 万形態素），および文書サイズの違い（15 発話，30 発話，60 発話）で比較を行った．いずれの場合もスムージングを用いた方が検索性能が向上した．また，学習データサイズが小さい方がスムージングの効果が大きい．以上の結果を考慮して，以降では $\lambda = 0.5$ に固定して実験を行った．

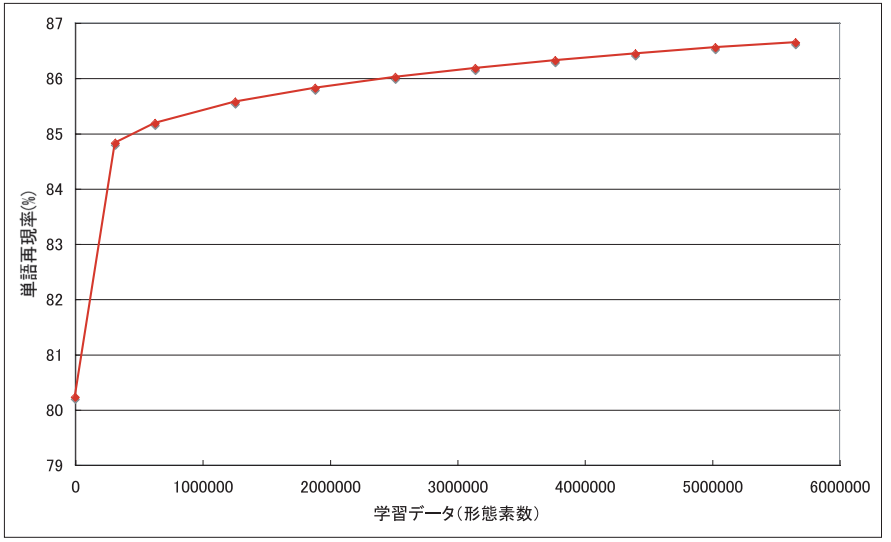

図 4　学習データサイズによる単語再現率の変化
Fig. 4　The effects of the quantity of training data.

4.6.2　学習データサイズの効果

翻訳モデルの学習に用いたパラレルテキストデータのサイズの効果を調べた．まず，翻訳モデルを用いて文書拡張することで，認識誤りを含む自動書き起しから本来の書き起しをどれだけ再現できるか（単語再現率）を調査した．単語再現率を以下のように定義する．

$$\text{単語再現率} = \frac{\text{正しく認識された，または，翻訳モデルで正解を予測できた単語数}}{\text{検索対象文書の人手書き起し単語数}}$$

翻訳モデルで予測する場合は，(単純分配法と同様に）正解単語アライメントにはさまれた誤認識単語のいずれかから，確率 0.01 以上で正解単語を予測できたときに再現できたとする[⋆2]．翻訳モデル学習には単純分配法を用いて，学習データサイズを 31.4 万形態素（0.5 ブロック）から 565 万形態素（9 ブロック）まで変化させた．結果を**図 4** に示す．翻訳モデ

⋆1 この方法では，IDF 値が実際より小さく見積もられてしまうが，今回は特に対策はしていない．

⋆2 削除誤りの場合は，再現不可能とした．

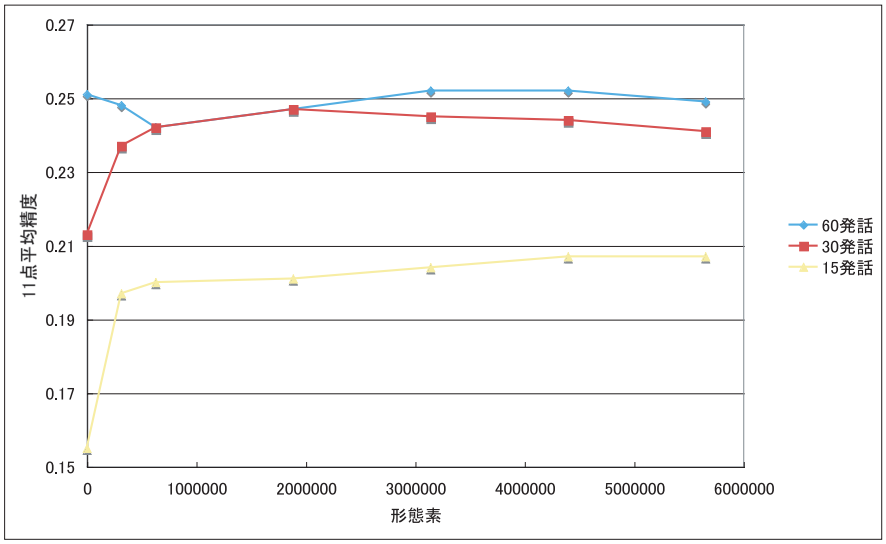

図 5　学習データサイズによる検索性能の変化
Fig. 5　The effects of the quantity of training data.

ルを用いることで，認識誤りの単語についても本来の単語が予測できていることが分かる．また，学習データサイズを増やすことで再現率も向上しており，検索性能の改善が期待できることが分かる．

次に，翻訳モデルの学習に用いたパラレルテキストデータのサイズによる検索性能の比較を行った．学習データサイズを変化させながら，**単純分配法**を用いて 1-best 候補から翻訳モデルの学習を行った．索引付けの閾値 α は 0.01 に設定した．正解は，R 判定の発話区間を用いた．結果を**図 5** に示す．データサイズが大きいほど性能は向上する傾向にあるが，およそ 63 万形態素を超えると，データサイズの増加に対してそれほど検索性能は向上しない．

4.6.3　閾値の設定と N-best 候補の利用

次に，学習データサイズ 31.4 万形態素とした場合の，同じ翻訳モデルについて，閾値 α の変化に対する検索性能を調べた．結果を**図 6** に示す．15 発話区間では閾値を小さくとり，なるべく多くの翻訳候補を用いた方が検索性能が高い．一方，30 発話区間，60 発話区間と検索対象の文書サイズを大きくするほど，より大きい閾値を設定して翻訳候補の絞り込みを行うことで検索性能が向上することが分かった．

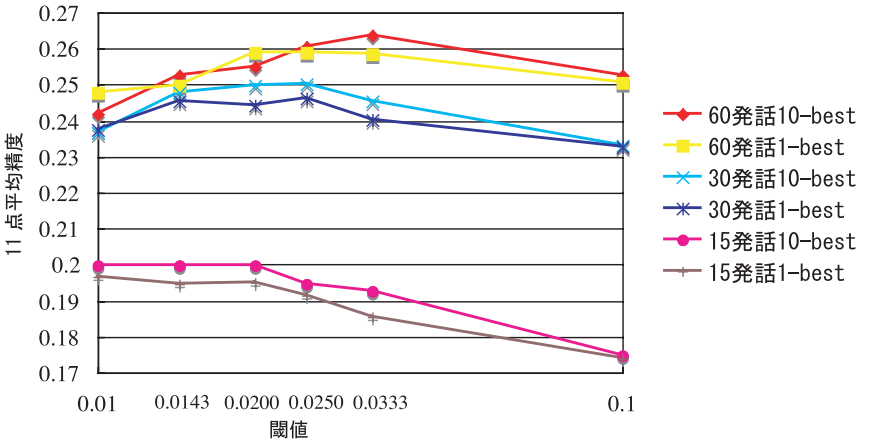

図 6　閾値の設定と認識 n-best の利用による検索性能の変化
Fig. 6　The effects of setting the threshold, with 1-best or 10-best candidates.

さらに，学習データとして，1-best 認識候補と正解書き起しのペアだけを用いる場合と，10-best 認識候補と正解書き起しの 10 種類のペアを用いる場合との比較を行った．図 6 に示すとおり，おおよそ 10-best 認識候補を使った場合の方が性能が高い．10-best 候補のパラレルテキストは，自動書き起しで候補を余計に生成するわずかなコストを除けば，1-best 候補とほぼ同じコストで入手可能なので，10-best 認識候補を学習データとして利用する方が優れているといえる．

4.6.4　従来手法との比較

最後に，従来手法と提案手法の各手法による検索性能の比較を行った．提案手法は，15 発話区間では閾値を 0.01 に，30 と 60 発話区間では閾値を 0.02 に設定した．また，学習データのサイズは 31.4 万形態素および 565 万形態素とし，認識候補のうち 1-best 候補と 10-best 候補を用いた場合を比べた．翻訳モデルの学習法として単純分配法と非交差分配法の 2 つの手法を比較した．正解には，R 判定および R+P 判定の発話区間を用いた場合の両方を調べた．結果を，それぞれ**表 2**，**表 3** に示す．

まず，ベースライン手法を比較すると，**認識結果のみ**に比べ**未知語拡張**と**全単語拡張**は，どの発話区間長でもそれぞれ 1，3 ポイント程度性能が低下していることが分かる．語の間の類似度だけを用いた文書拡張では，多くの不適切な語による索引付けがノイズとなり，性能が低下したと考えられる．また，ノイズの影響はより多くの拡張を行う**全単語拡張**の方が

表 2　R 判定を用いた検索性能の比較（11 点平均精度）
Table 2　11-point average precision for the compared methods using R degree of relevance.

索引付け	学習データサイズ	N-best 候補	15 発話	30 発話	60 発話
認識結果のみ		1-best	0.155	0.213	0.251
		10-best	0.177	0.225	0.256
未知語拡張		1-best	0.140	0.205	0.242
全単語拡張		1-best	0.124	0.185	0.223
翻訳モデル（単純分配法）	31.4 万	1-best	0.196	0.244	0.259
		10-best	0.200	0.250	0.255
	565 万	1-best	0.207	0.249	0.266
		10-best	0.206	0.252	0.258
翻訳モデル（非交差分配法）	31.4 万	1-best	0.195	0.244	0.257
		10-best	0.201	0.249	0.256
	565 万	1-best	0.207	0.250	0.266
		10-best	0.206	0.249	0.263
人手書き起し		—	0.180	0.249	0.305

表 3　R+P 判定を用いた検索性能の比較（11 点平均精度）
Table 3　11-point average precision for the compared methods using R+P degree of relevance.

索引付け	学習データサイズ	N-best 候補	15 発話	30 発話	60 発話
認識結果のみ		1-best	0.159	0.211	0.256
		10-best	0.179	0.227	0.261
未知語拡張		1-best	0.146	0.205	0.247
全単語拡張		1-best	0.126	0.184	0.228
翻訳モデル（単純分配法）	31.4 万	1-best	0.197	0.249	0.269
		10-best	0.202	0.255	0.266
	565 万	1-best	0.212	0.256	0.283
		10-best	0.210	0.259	0.272
翻訳モデル（非交差分配法）	31.4 万	1-best	0.197	0.249	0.267
		10-best	0.203	0.254	0.267
	565 万	1-best	0.212	0.258	0.282
		10-best	0.211	0.256	0.277
人手書き起し		—	0.181	0.249	0.305

大きい．

次に，ベースライン手法と提案手法の比較を行うと，31.4 万と 565 万形態素のどちらの学習データサイズにおいても，正解発話区間 15，30，60 発話のすべてのタスク設定において，提案手法はすべてのベースライン手法を上回る性能を示した．性能向上は，短い発話区間を正解とした場合に顕著である．特に，15 および 30 発話区間を正解とした場合，提案手法は人手書き起しテキストを用いた場合に匹敵する高い性能を示した．一方，60 発話区間を正解としたタスクにおいては，ベースライン手法との性能差は小さい．

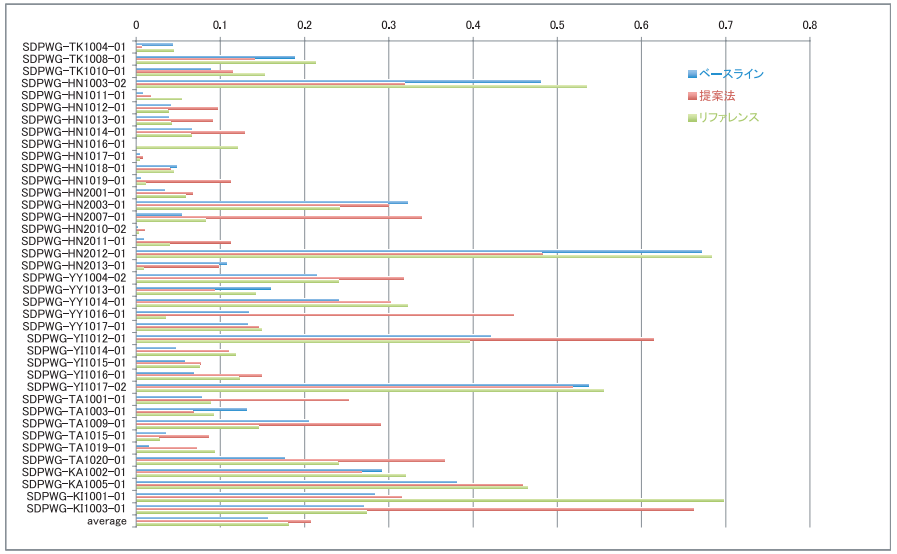

図 7　質問ごとの 11 点平均精度
Fig. 7　11-point average precision by query.

検索対象文書の語数が少ない場合，提案法は文書表現の拡張として働き，特に有効に機能したと考えられる．その効果は認識の複数候補を使う場合（**認識結果のみ**の 10-best）よりも大きい．一方，文書長が十分大きい場合は，その文書中の索引語だけで多様性が表現できるため，逆に誤った索引語を登録することによるノイズの影響が大きくなったと考えられる．

一方，単純分配法および非交差分配法の翻訳確率推定法の比較では，性能に大きな差は認められず，比較的単純な単純分配法でも十分な性能が得られることが分かった．

図 7 に，15 発話区間の R 判定正解を使った場合の，ベースライン（形態素のみ，1-best），提案手法（565 万形態素，1-best，単純分配法），リファレンス（人手書き起し）の各手法における，質問ごとの検索性能を示す．各質問についてほぼ同じ傾向の性能を示しているベースラインとリファレンスに対して，提案手法はそれを改善する場合と逆に性能が落ちる場合の差が顕著である．

これら 3 手法の性能（11 点平均精度の平均値）に有意な差があるかを調べるため，各手法

間で対標本の t 検定を行った．提案手法とリファレンスの間に有意な差は認められなかったものの，ベースラインと提案手法，ベースラインとリファレンスの間に有意差が認められた（p-value は，ベースライン–提案手法，ベースライン–リファレンス，提案手法–リファレンスの順に，0.00358，0.0310，0.123）．30 発話区間でも，同様に，ベースラインと提案手法，ベースラインとリファレンスの間に有意差が認められた（p-value は，それぞれ，0.0422，0.00275，0.487）．一方，60 発話区間では，ベースラインとリファレンスの間のみ有意であった（p-value は，それぞれ，0.225，0.00299，0.125）．以上より，提案手法は 15 および 30 発話区間の比較的短い発話区間を検索するタスクにおいて，リファレンスを使う場合と同様の検索性能が得られることが分かった．

5. ま と め

音声ドキュメントのアドホック検索手法として，音声認識による自動書き起しから人手書き起しへの翻訳モデルを利用した検索手法を提案し，CSJ テストコレクションで評価実験を行った．その結果，検索対象の文書サイズが小さい場合に有効に機能することが分かった．

本稿では，認識結果の候補として 10-best を用いたが，より大きな N-best や単語ラティスなどの多くの代替候補を用いた場合の効果や有効な手法についても今後検討したい．また，本稿では，TF-IDF ベースのアドホック検索の文脈において，翻訳モデルを期待 TF として再解釈して適用した．今後の課題として，翻訳モデルのような確率モデルをより直接的に扱える，言語モデリングに基づく検索手法に適用することが考えられる．

参 考 文 献

1) Garofolo, J.S., Voorhees, E.M., Stanford, V.M. and Jones, K.S.: TREC-6 1997 spoken document retrieval track overview and results, *Proc. 6th Text Retrieval Conference*, pp.83–91 (1997).
2) 前川喜久雄：『日本語話し言葉コーパス』の概要，日本語科学，Vol.15, pp.111–133 (2004).
3) 伊藤克亘，相川清明，秋葉友良，伊藤慶明，河原達也，南條浩輝，西崎博光，安田宜仁，山下洋一：音声ドキュメント検索評価のためのテストコレクションの試作，情報処理学会研究報告，SLP-064, pp.137–142 (2006).
4) 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，pp.73–80 (2007).
5) Moreau, N., Kim, H.-G. and Sikora, T.: Phonetic Confusion Based Document Expansion for Spoken Document Retrieval, *Proc. International Conference on Speech Communication and Technology* (*Eurospeech*), pp.542–545 (2004).
6) Saraclar, M. and Sproat, R.: Lattice-Based Search for Spoken Utterance Retrieval, *Proc. HLT-NAACL*, pp.129–136 (2004).
7) Iwata, K., Itoh, Y., Kojima, K., Ishigame, M., Tanaka, K. and Lee, S.W.: Open-Vocabulary Spoken Document Retrieval based on new subword models and subword phonetic similarity, *Proc. International Conference on Speech Communication and Technology* (*Eurospeech*), pp.325–328 (2006).
8) Yu, P. and Seide, F.: A Hybrid Word/Phoneme-based Approach for Improved Vocabulary-Independent Search in Spontaneous Speech, *Proc. International Conference on Speech Communication and Technology* (*Eurospeech*), pp.895–898 (2004).
9) Hori, T., Hetherington, I.L., Hazen, T.J. and Glass, J.R.: Open-vocabulary Spoken Utterance Retrieval using Confusion Networks, *Proc. International Conference on Acoustics Speech and Signal Processing*, Vol.IV, pp.73–76 (2007).
10) Attar, R. and Fraenkel, A.S.: Local Feedback in Full-Text Retrieval Systems, *Journal of the Association for Computing Machinery*, Vol.24, No.3, pp.397–417 (1977).
11) Deerwester, S.C., Dumais, S.T., Furnas, G., Landauer, T.K. and Harshman, R.A.: Indexing by latent semantic analysis, *Journal of the American Society for Information Science*, Vol.41, No.6, pp.391–407 (1990).
12) Berger, A. and Lafferty, J.: Information Retrieval as Statistical Translation, *Proc. 22nd Annual Conference on Research and Development in Information Retrieval* (*ACM SIGIR*), pp.222–229 (1999).
13) Xu, J. and Croft, W.B.: Evaluating a Probabilistic Model for Cross-lingual Information Retrieval, *Proc. 24th Annual Conference on Research and Development in Information Retrieval* (*ACM SIGIR*), pp.105–110 (2001).
14) Murdock, V. and Croft, W.B.: Simple Translation Models for Sentence Retrieval in Factoid Question Answering, *Proc. Workshop on Information Retrieval for Question Answering* (2004).
15) 清水　慧，秋葉友良，藤井　敦：統計翻訳に基づくパッセージ検索の言語横断質問応答への適用，言語処理学会第 13 回年次大会講演論文集，pp.1176–1179 (2007).
16) Ringger, E.K. and Allen, J.F.: Error Correction Via A Post-Processor For Continuous Speech Recognition, *Proc. International Conference on Acoustics Speech and Signal Processing*, pp.427–430 (1996).
17) 北　研二：情報検索アルゴリズム，共立出版 (2002).
18) Singhal, A., Buckley, C. and Mitra, M.: Pivoted document length normalization, *Proc. ACM SIGIR*, pp.21–29 (1996).
19) GETA: Generic Engine for Transposable Association. http://geta.ex.nii.ac.jp

(平成 20 年 5 月 25 日受付)
(平成 20 年 11 月 5 日採録)

秋葉　友良（正会員）

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）．ISCA，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．

横田　悠右

昭和 60 年生．平成 18 年詫間電波工業高等専門学校電子制御工学課卒業．平成 20 年豊橋技術科学大学情報工学課程卒業．同年豊橋技術科学大学大学院情報工学専攻入学，現在在学中．