Vol.2013-SLP-99 No.3
2013/12/19

音声中の検索語検出における 事前検索・HMM 状態系列照合・リランキングの適用

髙橋仁基[†1]　紺野和磨[†1]　熊谷真純[†1]
李時旭[†2]　田中和世[†3]　小嶋和徳[†1]　石亀昌明[†1]　伊藤慶明[†1]

音声中の検索語検出(STD : Spoken Term Detection) において，隠れマルコフモデル(HMM : Hidden Markov Model) 状態系列間の照合方式を用いた STD における・高精度化方式を提案する．提案方式では，音声ドキュメントに対し，予め音節認識を行っておき，得られた認識結果に対してあらゆる 2 音節（音節バイグラム）での事前に検索を行っておく．クエリが与えられると，クエリの音節列を 1 音節ずつシフトさせながら 2 音節を作成し，2 音節事前検索の上位 K 件のみを照合対象データとして絞り込んだ上で，HMM 状態系列間の照合を行う．さらにリランキングを行うことによって，より高精度な検索の実現を図る．評価実験の結果，検索精度は従来方式と比べ，すべてのテストセットで約 10〜17 ポイントの精度向上が見られ，NTCIR-9 の Dry run を除くと，事前検索結果を導入してもほぼ精度低下なく 1.5 秒以下で検索可能であった．K = 1,000 とした場合，NTCIR-9 の Dry run を除き 3 つのテストセットにおいて，HMM 間照合をリランキングした場合と比べても，検索精度・検索時間で優位となった．以上より本提案方式の有効性を確認できた．

Improving Retrieval Accuracy by Applying Various Methods to Spoken Term Detection

JINKI TAKAHASHI[†1]　KAZUMA KONNO[†1]　MASUMI KUMAGAI[†1]
SHI-WOOK LEE[†2]　KAZUYO TANAKA[†3]
KAZUNORI KOJIMA[†1]　MASAAKI ISHIGAME[†1] and YOSHIAKI ITOH[†1]

We propose various methods for Spoken Term Detection (STD), which identifies the target utterances where query terms are spoken in spoken documents. In this paper, we apply three methods to STD to improve the retrieval accuracy. The experimental results demonstrated the methods worked well for open test collections.

1. はじめに

近年，携帯可能な大容量記憶デバイスが普及し，撮りためた大量の音声データの中から特定のキーワードを検索する機能が求められている．この機能の実現に対し，音声中の検索語検出 (STD : Spoken Term Detection) が有効であると考えられ，STD に関する研究が盛んに行われるようになった．米国の NIST により TREC において評価型ワークショップが行われた[1]．また，国立情報学研究所が主催する NTCIR Workshop 9 が 2011 年に[2]，NTCIR Workshop 10 が 2013 年に開催され[3]，STD 方式についての評価が行われた．STD とは，音声ドキュメント中で 1 単語以上の検索語（クエリ）が発話されている位置を特定するタスクであり，一般的には，次のようにクエリが辞書に登録されている既知語の場合と，未知語の場合とで別々に処理される．クエリが既知語ならば単語認識結果を用いて検索を行い，クエリが未知語である場合は単語認識では誤認識となり，正しく検索することが困難であるため，サブワード認識結果を用いてクエリのサブワード系列と照合する方式が用いられる．STD では辞書に登録されていない未知語の検索が重要であり，本提案方式も未知語の検索に焦点を絞り，サブワード認識に基づく STD システムをベースとしている．

サブワード認識に基づく STD システムでは，音声ドキュメント群を予めサブワードで音声認識し，サブワード系列の検索対象データとしておく．テキストでクエリが与えられると，クエリの音節（音素）系列をサブワード系列に変換し，サブワード系列の検索対象データと連続動的計画法 (連続 DP：Continuous Dynamic Programming) 等で照合を行う．連続 DP で照合を行う際，我々は検索精度を向上させるため，サブワード間の音響的な距離を導入している[4]．サブワードをベースとする未知語検索においては，一般的に以下の 2 つが問題となる．

① 既知語検索に比べ未知語検索の検索精度が劣る
② サブワード列照合を行うため，検索時間は音声ドキュメントの量に比例して多くなる

†1 岩手県立大学
Iwate Prefectural University
†2 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology
†3 筑波大学
University of Tsukuba

①の検索精度を向上させるために，我々は前述のようにサブワード間の音響距離を定義した他，リランキング方式を提案した[5]．このリランキング方式では高順位候補は正しく検出されており，高順位候補を含むドキュメントにはその検索語が複数含まれていると仮定し，高順位候補を含むドキュメント中の候補のスコアを優先するものである．さらなる検索精度の向上を目指し，本稿では[6]を参考としてサブワードを構成している状態系列での照合方式を導入する．状態系列での照合を行った場合，例えば triphone は一般に 3 状態で構成されているため，参照側・入力側とも 3 倍の系列同士の照合となり，triphone 間の照合に比べ 9 倍の計算時間を要することになる．上記問題②に述べたように連続 DP 等で全音声ドキュメントと照合する場合（以降，全照合，All CDP と示す），音声ドキュメントの長さに比例した検索時間を要することになる．例えば，日本語話し言葉コーパス (CSJ：Corpus of Spontaneous Japanese)[7] 中のコア 177 講演に対する 1 クエリあたりの検索時間は 0.38 秒，全 2702 講演では約 5.8 秒を要し，検索対象に比例した．状態系列間照合するとこの約 9 倍の検索時間が必要となる．音声ドキュメントの長さに依らずに高速にかつ検索精度の低下を抑えつつ検索を行うために，我々は既に音節バイグラム事前検索結果を利用したシステムの提案を行った[8]．これは事前にすべての音節バイグラムで音声ドキュメントを検索しておき，クエリが与えられるとそのクエリを音節バイグラムに分解し，各音節バイグラムについて上位 K 件の事前結果 (発話，IPU：Inter Portal Unit) を抽出し，その IPU にのみ連続 DP 照合を行い，高速な検索を実現した．この音節バイグラム事前検索方式を適用することで検索時間の増大を抑制する．

本稿では，まず 2 章で我々の STD システムについて概説し，次に上述の 3 つの提案方式を説明する．3 章で評価実験を通して本提案方式の有効性を示す．

2. 提案方式

2.1 システムの概要

本節では，我々が提案している主に未知語のクエリが与えられた場合のサブワードベースの STD システムについて概説する．図 1 に STD システムの概要を示す．

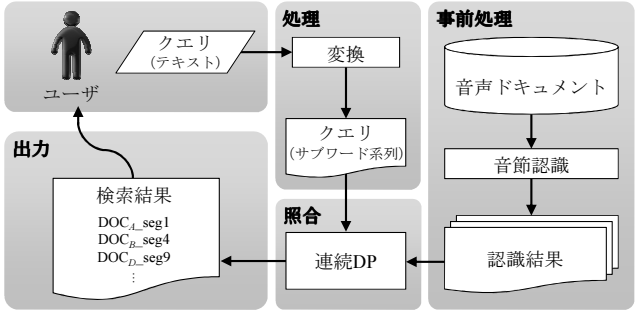

図 1　STD system 概要図

検索対象の音声ドキュメントを，ポーズにより発話毎にセグメンテーションし，発話毎にサブワード認識を行う．認識結果のサブワード系列をサブワード認識結果として予め保持しておく．本稿では，クエリはテキストで与えられるものとし，クエリは変換規則に則り自動でサブワード系列に変換できるものとする．このサブワード系列クエリと音声ドキュメントのサブワード系列群を連続 DP により照合を行う．照合時の局所距離に，サブワード間音響距離を用いることで，検索精度の向上を図っている．連続 DP の累積距離が小さい，即ち類似度が高い順にユーザへ候補発話 (IPU) を提示する．

2.2 提案方式の概要

図 2 に本稿で提案する 3 つの方式の概要を示す．図 1 の処理に加え，認識結果に対し，存在し得るすべての音節バイグラム(音節の 2 つ組) について，HMM (Hidden Markov Model) 間の音響距離を用いて連続 DP で照合を行う．ユーザからクエリが与えられると，クエリを音節バイグラムに変換し，事前に照合しておいた検索結果から上位 K 件の検索結果を照合対象 IPU 群として抽出する（① 音節バイグラムによる事前検索）．この IPU 群について従来の HMM 系列間ではなく，状態系列間で照合を行う（② HMM 状態系列間照合）．最後に出力された検索結果に対して，リランキングを行い，類似度が高い順にユーザへ候補区間 (IPU) を提示する（③ リランキング）．

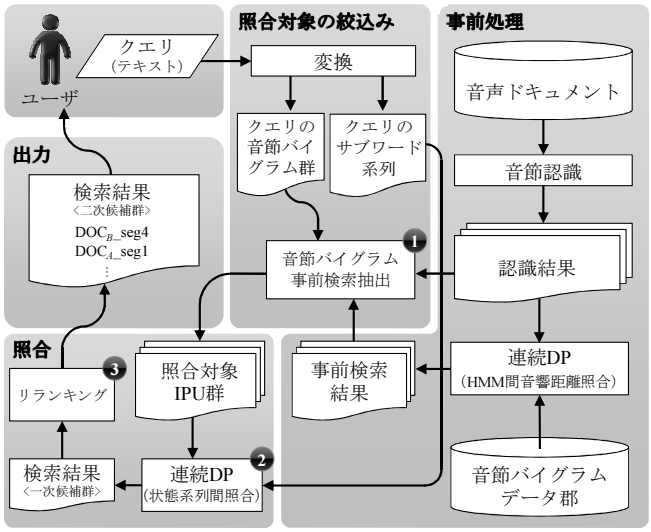

図 2　提案方式 概要図

2.3 音節バイグラム事前検索方式

2.3.1 音節バイグラムによる事前検索の作成

検索対象の音声ドキュメントに対し，事前に triphone 等のサブワードを用いて音節認識を行っておく．その認識結果に対して，存在し得る音節バイグラム (66,700 種) 全てを用いて連続 DP による照合を行っておく．本稿では照合

を行う際の局所距離には，triphone HMM 間の音響距離を用いる．各音節バイグラムについては，上位 K 件を保持させ，事前検索結果全体として，66,700 × K 個が保持される．

2.3.2 クエリの音節バイグラムの抽出

クエリが与えられると，クエリ中の音節列から音節バイグラムを，1 音節ずつシフトさせながら抽出し，クエリの音節バイグラム群を作成する．1 つのクエリに対して，（クエリの音節数 − 1）個の音節バイグラムを抽出する．例えばクエリが「カイセキ」であれば，「カイ」「イセ」「セキ」の 3 個の音節バイグラムを抽出する．

2.3.3 照合対象の IPU 群の抽出

2.3.2 で抽出したクエリの各音節バイグラムに対し，図 3 に示したように 2.3.1 で作成した事前検索結果の上位 K 件の候補 IPU 群を照合対象 IPU 群として抽出する．今回はクエリの各音節バイグラムの候補 IPU 群の論理和をとった．

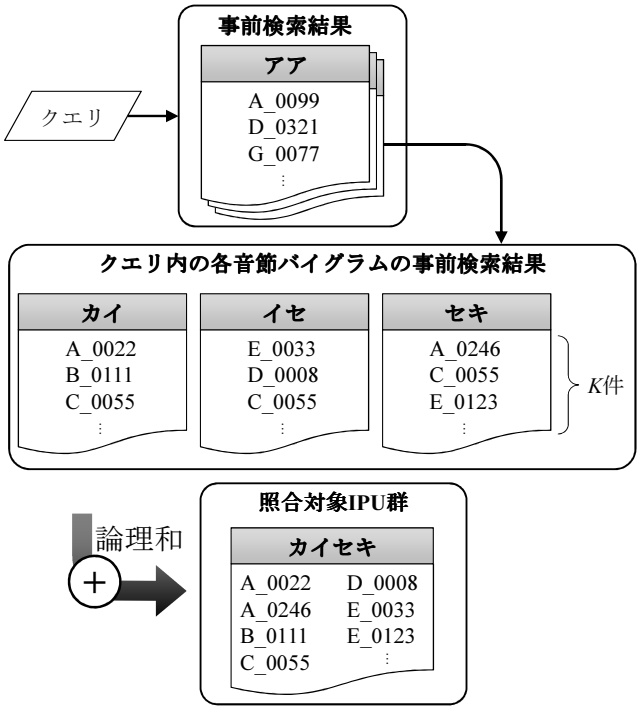

図 3　照合対象 IPU 群作成手順

2.4 HMM 状態系列間による照合方式

2.4.1 連続 DP による一次ランキング結果出力

2.2.3 で抽出した検索対象 IPU 群に対し，サブワード間ではなく HMM の状態系列間で連続 DP による照合を行う．従来，2 つのサブワード HMM 間の音響距離を用いて照合を行なっていた．提案方式では，HMM を構成している状態系列での照合を行うことで，詳細な照合を行い，検索精度の向上を図る．

我々のサブワード HMM は全て 3 状態で構成されており，状態共有により N 状態とした時，全ての状態間の距離行列 (N × N) を予め作成しておく．状態間距離は[9]に沿って以下の手順で求める．Dim 次元の特徴量として，μ_{smd}，σ^2_{smd} は状態 s，分布 m の d 番目の特徴量の平均，分散を表すものとする．

① 状態 s の M_1 個の分布中 m 混合目の分布と，状態 t の M_2 個の分布中 n 混合目の分布をそれぞれ取り出す

② ①の 2 つの分布間で式(1)により，バタチャリヤ距離を計算する

$$BD(s_m, t_n) = \frac{1}{4}\sum_{d=1}^{Dim}\left\{\frac{(\mu_{smd}-\mu_{tnd})^2}{\sigma^2_{smd}+\sigma^2_{tnd}} + \log\frac{(\sigma^2_{smd}+\sigma^2_{tnd})^2}{4\sigma^2_{smd}\sigma^2_{tnd}}\right\} \quad \cdots(1)$$

③ 全ての組み合わせ (M_1 × M_2) の中から距離が最小のものを，状態 s と状態 t の距離とする

$$LD(s,t) = \min\left(BD(s_1,t_1), BD(s_1,t_2), \cdots, BD(M_1,N_2)\right) \quad \cdots(2)$$

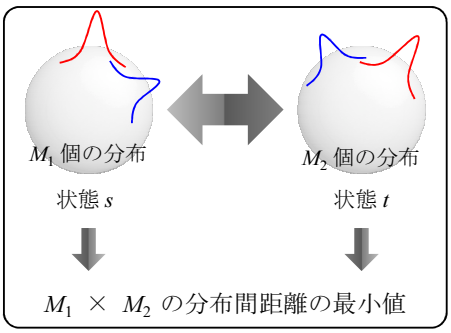

図 4　状態系列間距離の算出イメージ

状態間距離行列 (3,009 状態 × 3,009 状態) はメモリ上に保持し，連続 DP を行う際には局所距離としてこの行列を参照するのみである．サブワード系列との音声ドキュメントは予め状態番号系列に変換しておく．クエリが与えられるとそのサブワード系列を状態系列に展開して連続 DP による照合を行う．この連続 DP の距離に従って，第一次のランキングがなされる．

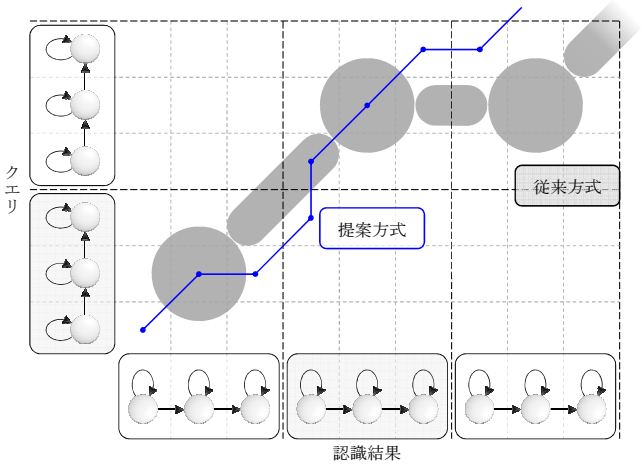

図 5　HMM 間照合と HMM 状態系列間照合

2.5 同一ドキュメント内の上位候補を利用したリランキング方式

2.5.1 リランキングによる二次ランキング結果出力

連続 DP を行った結果に対して，リランキングを行う．ユーザが入力する検索語は，特定のドキュメントに頻繁に出現する単語あると考えられるため，全体で上位となる候補を含むドキュメントの中では，その検索語が他にも話されていると仮定できる．そこと上位候補を含むドキュメントに対し,そのドキュメントに属する全ての候補の連続 DP の距離が有利になるように調整する[5].

音声ドキュメント d 内の候補発話に対してリスコアリングを行う．d 内の s 番目の発話 d_s（発話 ID）について d 内での順位が r 位であったとき，発話 d_s の距離を $D(d_s, r)$ と表す. リスコアリングは, リスコアリング後の距離を D' として以下の式により行う．

$$D'(d_s, r) = \begin{cases} D(d_s, r) & (r = 1) \\ \alpha \times D(d_s, r) + (1-\alpha) \times \dfrac{1}{r-1} \displaystyle\sum_{t=1}^{r-1} D'(d_s, t) & (r \neq 1) \end{cases} \quad \cdots (3)$$

上式では, 音声ドキュメント d 内で 1 位（最小）の場合，元々の距離をそのまま $D'(d_s, 1)$ に用い，2 位以下の場合は元々の距離 $D(d_s, r)$ と，1〜$(k - 1)$ 位までの候補について，リスコアリング後の距離の平均を線形結合することで求める．$\alpha\ (0 \leq \alpha \leq 1)$ は重み係数である．

これにより，高順位の候補を有しないドキュメント内では，候補の距離はあまり変動しないが，全体で高順位候補を有するドキュメント内では，高順位の小さい距離の影響を受け順位が上がることになる．このため，認識誤りで候補順位が下がってしまった候補発話の順位を上げることができ，検索精度の向上を図ることができる．

3. 評価実験

3.1 実験条件

音響モデルと言語モデルの学習には，評価に用いる CSJ 中のコア 177 講演を除いた学会講演と模擬講演の内，偶数講演データ（1255 講演，287 時間）を用いた．音響モデルは 3 状態の triphone HMM で構成し，HTK (Hidden Markov Model Toolkit)[10] を用いて学習を行った．モデル数は 7,946，状態共有を行って 3,009 状態とした．言語モデルは Palmkit[11]を用い，音節単位の前向き 2-gram と後ろ向き 3-gram を構築した．音声認識には Julius ver.4.1.5.1[12]を用いた．音響・言語モデルの構築条件を表 1 に示す．

表 1　音響・言語モデルの構築条件

標本化周波数	16 kHz
量子化	16 bit
音響特徴量	38 次元 (MFCC_E_D_A_Z_N) MFCC(12dim) + ΔMFCC(12dim) + ΔΔMFCC(12dim) + ΔPower + ΔΔPower
窓長	25 msec
フレームシフト	10 msec
分析窓	ハミング窓
音響モデル	Triphone Left-to-right HMM with 3 states (7,946 models / 3,009states)
言語モデル	Syllable bigram, trigram

3.2 評価用データ

評価には表 2 に示すように 2 つの評価セットを用いた．検索対象音声ドキュメントは, NTCIR-9 では CSJ のコア 177 講演（約 44 時間，53,892 発話），NTCIR-10 では音声ドキュメントワークショップでの講演音声(SDPWS：Corpus of Spoken Document Processing Workshop)（約 28.6 時間，40,746 発話）を用いた. クエリと正解情報は NTCIR で提供された，NTCIR-9 Spoken Doc Task における Formal run と Dry run 及び，NTCIR-10 Spoken Doc Task における Formal run と Dry run の 4 種類である（iSTD(inexistent Spoken Term Detection) タスク用のクエリを除く）.

表 2　評価用データセット

検索対象 音声データ	CSJ177 講演 (約 44 時間)	SDPWS (104 講演) (約 28.6 時間)
発話件数	53,892 発話	40,746 発話
クエリ	NTCIR-9 Spoken Doc Task Formal run 50 クエリ Dry run 50 クエリ	NTCIR-10 Spoken Doc Task Formal run 100 クエリ Dry run 32 クエリ

3.3 評価方法

NTCIR では IPU 単位（発話単位）で正解の判定が行われており，候補とした発話内にクエリが実際に含まれていれば正解となる．評価は主に検索精度と検索時間，必要メモリ量から行う．検索精度の評価には，式(4)(5) から算出される MAP (Mean Average Precision) を用いた．式(4)により，クエリごとに順位が上位の候補から出力した際，正解出現時の適合率を平均とすると，AP (Average Precision) が得られ，式(5)の通りこの AP を全てのクエリに対して平均すると MAP が得られる．

$$AP(q) = \frac{1}{c} \sum_{i=1}^{R} \delta_i \times precision(q, i) \quad \cdots (4)$$

$$MAP = \frac{1}{Q} \sum_{q=1}^{Q} AP(q) \quad \cdots (5)$$

処理時間の計測には，Intel 社の Xeon E3，メモリ 16GB の Linux マシンを使用し, C 言語関数の clock を用いて時間の計測を行った．

3.4 実験結果

3.4.1 HMM 状態系列間照合方式の評価

まず 2.4 で提案した HMM 状態間照合方式について評価を行う．全ての音声ドキュメントに対して，従来の HMM 系列間で照合した場合と，HMM 状態系列間で照合した場合の実験結果を表 3 に示す．

表 3　HMM 系列間照合と HMM 状態系列間照合

MAP (%)		HMM 間照合	HMM 状態系列間照合	差
NTCIR-9	Formal set	71.54	77.38	+ 5.84
	Dry set	63.76	68.16	+ 4.40
	Time(sec./query)	0.38	3.23	8.50 倍
NTCIR-10	Formal set	48.27	54.56	+ 6.29
	Dry set	54.78	64.20	+ 9.42
	Time(sec./query)	0.37	3.14	8.49 倍

表 3 が示すように, 4 種のテストセットに対して 4.4〜9.4 ポイントの精度向上が得られ，HMM 状態系列間での照合が有効であることが分かる．HMM 状態系列での照合は，HMM をクエリ側で 3 状態，音声ドキュメント側も 3 状態に展開するため, 原理的には 9 倍の計算時間が必要となる．実験の結果においても 8.5 倍の計算時間が必要であった．

3.4.2 事前検索方式の評価

前節の通り，状態間で照合すると計算時間を要するため，事前検索結果を利用して発話数を絞った上で照合を行うことにより，検索時間を削減する．クエリの各音節バイグラムに対し，事前検索の上位 K 件に絞った上で HMM 状態系列間の照合を行った結果を表 4 に示す．

表中，All CDP は前述の通り，音声ドキュメント全体に対し連続 DP 照合を行ったもので精度の上限を示している．

表 4　事前検索結果適応結果

Top K	NTCIR-9 MAP (%)			NTCIR-10 MAP (%)		
	Formal	Dry	Time (sec.)	Formal	Dry	Time (sec.)
All CDP 状態系列	77.38	68.16	3.22	54.56	64.20	3.14
6,000	77.28	67.04	1.24	54.52	64.17	1.45
5,000	77.28	66.71	1.09	54.53	64.14	1.29
4,000	76.60	66.09	0.92	54.53	64.13	1.11
3,000	76.50	64.35	0.74	54.52	64.12	0.89
2,000	75.64	62.51	0.53	54.53	64.00	0.64
1,000	74.85	58.08	0.28	53.77	63.71	0.35
All CDP HMM 系列	71.54	63.76	0.38	48.27	54.78	0.37

評価実験より，NTCIR-9 Formal set，Dry set，NTCIR-10 Formal set，Dry set 共に K = 6,000 であれば，All CDP と比べ，検索精度の低下がほぼ無く検索時間を削減できた．また NTCIR-9 の Dry run を除き，K = 1,000 において，HMM 系列照合と比較しても検索精度, 検索時間の両面で優った．

K = 3,000 のとき，NTCIR-9 の Dry run を除き性能低下は 1 ポイント以下に抑えながら計算時間を 1 秒以内で検索可能であることが分かる．この時の検索精度は HMM 系列間照合と比べ，NTCIR-9 Formal run, Dry run, NTCIR-10 Formal run, Dry run で，それぞれ 5.0, 0.6, 6.3, 9.3 ポイント良くなっており，状態系列照合と事前検索結果を適用することの有効性が確認できた．照合時に必要となる距離行列のメモリ容量については HMM 間距離行列が 241MB (7,946 × 7,946 × 4B), 状態間距離行列は 35MB (3,009 × 3,009 × 4B) と小さくなったが, 事前検索結果として K = 6,000 の時, 1.4GB, K = 2,000 の時，0.5GB を要することになる．

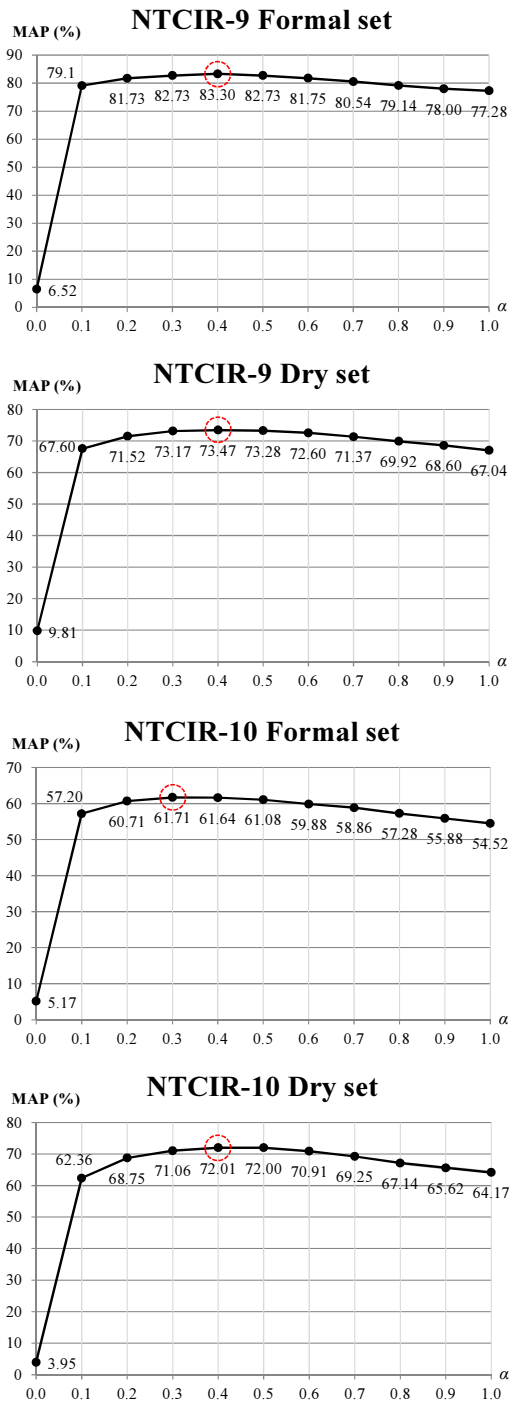

図 6　K = 6,000 のリランキングスコア

3.4.3 リランキング方式の評価

事前検索の K = 6,000 の時，状態系列照合を行った結果に対して，リランキングを行った式(3)の α を 0.0〜1.0 に変化させた時の結果を図 6 に示す．α = 1.0 のときは第 2 項の上位の候補の距離を全く使わず，α = 0.0 のときは第 1 項の当該候補の距離を全く使わないことを示している．

図 6 の 4 つのグラフのいずれにおいても，α = 0.3〜0.4 の時，最も良い精度が得られた．文献[5]においても α = 0.3 と α = 0.4 で最も良く，そのクロスバリデーションにより評価がなされていた．今回もクロスバリデーションによる評価を試みたが，結果としては α = 0.3〜0.4 で最良の精度が得られ，α が安定していることも確認できた．

事前検索の上位 K 件に対して HMM 状態系列の照合を行った一次ランキングに対して，α = 0.4 でリランキングを行った結果を NTCIR-9 については表 5 に，NTCIR-10 については表 6 に示す．

表 5　NTCIR-9 のリランキングスコア (α = 0.4)

Top K	Formal run MAP (%)			Dry run MAP (%)		
		リランキング	up (point)		リランキング	up (point)
All CDP 状態系列	77.38	83.56	+6.18	68.16	74.84	+6.68
6,000	77.28	83.30	+6.02	67.04	73.47	+6.43
5,000	77.28	83.38	+6.10	66.71	72.89	+6.18
4,000	76.60	82.36	+5.76	66.09	72.30	+6.21
3,000	76.50	82.19	+5.69	64.35	72.25	+7.90
2,000	75.64	82.30	+6.66	62.51	67.62	+5.11
1,000	74.85	80.17	+5.32	58.08	62.08	+4.00
All CDP HMM 系列	71.54	78.23	+6.69	63.76	71.83	+8.07

表 6　NTCIR-10 のリランキングスコア (α = 0.4)

Top K	Formal run MAP (%)			Dry run MAP (%)		
		リランキング	up (point)		リランキング	up (point)
All CDP 状態系列	54.56	61.71	+7.15	64.20	72.11	+7.91
6,000	54.52	61.64	+7.12	64.17	72.01	+7.84
5,000	54.53	61.65	+7.12	64.14	71.86	+7.72
4,000	54.53	61.59	+7.06	64.13	71.83	+7.70
3,000	54.52	61.60	+7.08	64.12	71.61	+7.49
2,000	54.53	61.48	+6.95	64.00	71.38	+7.38
1,000	53.77	60.00	+6.23	63.71	69.88	+6.17
All CDP HMM 系列	48.27	56.16	+7.89	54.78	62.39	+7.61

評価実験より，リランキングを行うことで全てのテストセット及び K の値において 4.0〜7.9 ポイント良くなっており，リランキング適用の有効性を確認できた．リランキングの計算時間は 0.01 秒未満，新たに必要となるメモリは 1MB にも満たないため，無視できるものとした．

3.4.4 考察

事前検索方式，HMM 状態系列間照合方式，リランキング方式を適用した場合の総合的な結果を表 7 に示す．従来方式の HMM 間照合と比べ，最も性能が良くなるのは状態間照合を行った上でリランキングを行ったものである．この場合には計算時間を要するため，事前検索結果を利用すると，K = 6,000 のとき，検索精度は従来方式と比べ，すべてのテストセットで約 10〜17 ポイントの精度向上が見られ，NTCIR-9 の Dry run を除くと，事前検索結果を導入してもほぼ精度低下なく 1.5 秒以下で検索可能，K = 3,000 のときは 1 秒以下で（精度低下は 1.3 ポイント程度まで）検索可能であった．K = 1,000 とした場合，NTCIR-9 の Dry run を除き 3 つのテストセットにおいて，HMM 間照合をリランキングした場合と比べても，検索精度・検索時間で優位となった．以上より本提案方式の有効性を確認できた．NTCIR-9 の Dry run については事前検索を導入した場合の精度低下が大きく，この理由と対策については今後の課題としたい．

4. おわりに

本稿では音声ドキュメントを予め音節認識した結果から事前検索を行い，クエリが与えられると事前検索結果の上位 K 件を照合対象として絞り込んで，HMM 状態系列での照合を行う．この結果をさらにリランキングを適用することで，速度を低下させることなく STD の高精度化を実現する方式を提案した．

状態系列での照合により計算時間が約 9 倍となるが，速度低下がない時，NTCIR-9 Dry run を除いて 8.6〜15.1 ポイントの精度向上を実現した．

今後は，NTCIR-9 Dry run の精度に対して，事前検索を適用した場合の精度低下の原因の追求と対策を検討していきたい．また，他のサブワードでの本提案方式の有効性を検証していきたい．

表 7　HMM 状態系列間照合，事前検索，リランキングの適用結果（計算時間は秒，他は MAP(%)）

MAP (%)		従来方式		提案方式				
		HMM 系列	計算時間	状態系列		事前検索	リランキング	計算時間
NTCIR-9	Formal run	71.54	0.38 (sec./query)	77.38	状態系列照合に対して		83.56	3.23
					K = 6,000	77.28	83.30	1.24
					K = 3,000	76.50	82.19	0.74
					K = 1,000	74.85	80.17	0.28
				HMM 系列照合に対して			78.23	0.38
NTCIR-9	Dry run	63.76	0.38 (sec./query)	68.16	状態系列照合に対して		74.84	3.23
					K = 6,000	67.04	73.47	1.24
					K = 3,000	64.35	72.25	0.74
					K = 1,000	58.08	62.08	0.28
				HMM 系列照合に対して			71.83	0.38
NTCIR-10	Formal run	48.27	0.37 (sec./query)	54.56	状態系列照合に対して		61.74	3.14
					K = 6,000	54.52	61.61	1.45
					K = 3,000	54.52	61.60	0.89
					K = 1,000	53.77	60.00	0.35
				HMM 系列照合に対して			56.16	0.37
NTCIR-10	Dry run	54.78	0.37 (sec./query)	64.20	状態系列照合に対して		72.11	3.14
					K = 6,000	64.17	72.01	1.45
					K = 3,000	64.12	71.61	0.89
					K = 1,000	63.71	69.88	0.35
				HMM 系列照合に対して			62.39	0.37

謝辞

本研究の一部は文部科学省学術研究助成基金助成金基盤研究(C)No.24500124 を受けて実施された.

参考文献

[1] Jonathan G. Fiscus, Jerome Ajot, John S. Garofolo, George Doddington, in Proceedings of SIGIR Workshop Searching Spontaneous Conversational Speech. Results of the 2006 spoken term detection evaluation (Sept. 2007), pp. 45–50.

[2] Tomoyosi Akiba, Hiromitsu Nishizaki, Kiyoaki Aikawa, Tatsuya Kawahara and Tomoko Matsui, "Overview of the IR for Spoken Document Task in NTCIR-9 Workshop", Proceedings of NTCIR-9 Workshop Meeting, pp.223-235 (2011.12.06-09)

[3] Tomoyosi Akiba, et al., "Overview of the NTCIR-10 Spoken Doc-2 Task", Proceedings of the NTCIR-10 Conference, 2013.

[4] 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”, 情報通信学会論文誌，Vol.48, No.5, pp.1990-2000, (2007).

[5] 紺野和磨，伊藤慶明，小嶋和徳，石亀昌明，田中和世，李時旭, “音声中の検索語検出における高順位ドキュメント優先方式の提案”, 日本音響学会春季研究発表会, 3-8-8, p.115-118, 4 pages (2013.09).

[6] Naoki Yamamoto, Atsuhiko Kai, "Using Acoustic Dissimilarity Measures Based on State-level Distance Vector Representation for Improved Spoken Term Detection", APSIPA 2013

[7] Corpus of Spontaneous Japanese http://www.ninjal.ac.jp/corpus_center/csj/

[8] 斉藤裕之，伊藤慶明，小嶋和徳，石亀昌明，田中和世，李時旭, “N-音節事前検索結果を用いた音声中の検索語検出における上位候補の高速検索”, 日本音響学会秋季研究発表会，3-1-2, 4 pages (2012.09).

[9] 谷藤史崇，伊藤慶明，小嶋和徳，石亀昌明，田中和世，李時旭, “適切なモデル間距離による音声中の検索語検出の精度向上”，日本音響学会春季研究発表会，2-P-58, 2 pages (2011.03).

[10] Hidden Markov Model Toolkit http://htk.eng.cam.ac.uk

[11] Palmkit http://palmkit.sourceforge.net

[12] 大語彙連続音声認識エンジン Julius http://julius.sourceforge.jp

[13] 伊藤慶明，西崎博光，中川聖一，秋葉友良，河原達也，胡新輝，南條浩輝，松井知子，山下洋一，相川清明，“音声中の検索語検出のためのテストコレクションの構築と分析”，情報処理学会論文誌，Vol.54, No.2, pp.471-483, (2013.02)

[14] 西崎博光，秋葉友良，相川清明，伊藤慶明，河原達也，胡新輝，中川聖一，南條浩輝，山下洋一，“NTCIR-10 SpokenDoc-2 Spoken Term Detection タスクの結果と知見”，日本音響学会秋季研究発表会，3-8-6, pp.107-110.