

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

伊藤 慶明^{1,a)} 西崎 博光² 中川 聖一³ 秋葉 友良³ 河原 達也⁴ 胡 新輝⁵
南條 浩輝⁶ 松井 知子⁷ 山下 洋一⁸ 相川 清明⁹

受付日 2012年5月31日, 採録日 2012年11月2日

概要: TRECにおいて Spoken Document Retrieval (SDR: 音声ドキュメント検索) の Track が 1996 年～2000 年に設定され, 2006 年には米国国立標準技術研究所 (NIST) を中心に音声中の検索語検出 (Spoken Term Detection: STD) タスクが設定され, 以降海外ではさかんに SDR, STD に関する研究が行われるようになった. 情報処理学会音声言語情報処理研究会 (SIG-SLP) で国内の音声ドキュメント処理研究の推進・活性化を目的として 2006 年に音声ドキュメント処理ワーキンググループを立ち上げ, これまでに SDR 評価用テストコレクションの構築を進めてきた. これに続き 2008 年から音声中の検索語検出の評価用テストコレクションの構築を開始し, 2009 年 10 月にテストコレクションを公開した. 以来, 様々な研究機関で利用されている. また 2011 年には国立情報学研究所が主催する NTCIR-9 での dry run に用いられ, formal run 用のテストコレクションへの展開も行った. 本論文ではこのテストコレクション構築にあたっての方針, テストコレクションの内容, ベースライン評価結果等の分析について説明するとともに, 構築したテストセットによって活性化した STD 研究の展開について述べる.

キーワード: テストコレクション, 音声中の検索語検出, 音声ドキュメント, 音声認識, 検索

Construction and Analysis of Test Collections for Spoken Term Detection

YOSHIAKI ITOH^{1,a)} HIROMITSU NISHIZAKI² SEIICHI NAKAGAWA³
TOMOYOSI AKIBA³ TATSUYA KAWAHARA⁴ XINHUI HU⁵ HIROAKI NANJO⁶
TOMOKO MATSUI⁷ YOICHI YAMASHITA⁸ KIYOAKI AIKAWA⁹

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

Abstract: Spoken Document Retrieval (SDR) was dealt with in one of tracks of TREC from 1996 to 2000. NIST supplied a task for Spoken Term Detection (STD) in 2006. Many researches have been conducted as for SDR and STD after these projects. A working group for spoken document processing of SIG-SLP (Spoken Language Processing) in IPSJ also aimed to activate the researches for spoken document processing, and developed a test collection for SDR so far. The working group started to develop a test collection for STD from 2008, and distributed the β version of the test collections in October 2009. The test collections was used for the test sets of a dry run in NTCIR-9 held by NII in 2011, and was extended to the test sets of a formal run. This paper reports the policy, the content and baseline evaluation results and their analysis of the collection in detail, and describes the research expansion by the test collections.

Keywords: test collections, spoken term detection, spoken documents, speech recognition, retrieval

1. はじめに

近年、パソコンのマルチメディア環境、高速なインターネット、大容量のHDDを搭載したビデオレコーダが普及し、撮り溜めたTV放送、長期間にわたり録画した家庭用ビデオ、動画サイトでのビデオコンテンツ、講義や教材用ビデオ等、音声・動画を含んだビデオコンテンツが増加・大容量化している。これにともないこれらの大量のデータから見たい・聞きたい部分を検索したいという機能が求められるようになった。音声を含むデータに対しては、音声認識技術を適用してデータを検索する方式が有望であり、音声ドキュメント検索 (Spoken Document Retrieval: SDR) としてすでに様々な研究が行われてきている。

SDRにおいては、ビデオや講義音声等の音声を含むデータを音声ドキュメントと呼び、複数あるいは大量の音声ドキュメントの中から、クエリに関連する音声ドキュメントを特定することをアドホック SDR あるいは単に SDR と呼ぶ。検索の基本的な枠組みでは、まず音声ドキュメントを単語ベースで音声認識しておき、その認識結果である単語列に対してテキスト検索の技術を用いて音声ドキュメントを特定する。性能を評価する際、音声認識では音声データの「質」(発話の丁寧さや、録音の精度等)に主に影響されるが、SDRでは音声データの「質」だけでなく「長さ」, 「正解箇所の数」にも影響される(たとえば、1時間の音声データから探す場合、10時間の音声データから探す場合、正解がまったく含まれていない場合、これらの検索性能の比較は困難)。このため SDR では共通の音声ドキュメント、クエリ、正解に基づいて評価が行われることが望ましい。

TREC (Text REtrieval Conference) においては、SDR の Track が 1996 年の TREC-6 から取り上げられ [1], [2], [3],

TREC-7~9 を経て 2000 年まで行われた [3]。これを機に海外では SDR に関しての研究が推進・活性化された。日本においても情報処理学会音声言語情報処理研究会 (SIG-SLP) において国内の音声ドキュメント処理研究の推進・活性化を目的として 2006 年に音声ドキュメント処理ワーキンググループ (SDPWG) を立ち上げ、これまでに SDR 評価用テストコレクションを構築しその公開を行った [4]。

アドホック SDR によりクエリと関連あるドキュメント群が特定できたとしても、その結果は一覧性・確実性に欠け、最上位のドキュメントでさえ、あるキーワードが含まれているかは実際に聞いてみないと確かめられない。検索語 (1 個以上の単語からなる言葉) が話されている箇所を音声の中から特定 (音声中の検索語検出: Spoken Term Detection: STD) したいというニーズは SDR において不可避である。

STD では音声ドキュメントを単語音声認識システムでいったん認識しておき、検索語がこの音声認識システムの辞書に登録されている単語 (既知語) ならば単語認識結果を検索すればよいが、辞書に登録されていない単語 (未知語) ならば、単語認識結果からは適切な検索は難しい。このため単語より小さな単位のサブワード (音素等) で音声ドキュメントを認識しておき、検索語が未知語である場合には検索語のサブワード系列と認識結果のサブワード系列どうしを照合する方式が一般的である。検索語が音声認識システムの未知語になる場合は多く [5], [6], 未知語の検索機能は不可欠である。米国国立標準技術研究所 (NIST) [1], [2] が 2006 年に STD を新たなテーマとして設定して以降、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り、2009 年以降の音声関連の国際会議 (INTERSPEECH, ICASSP, ASRU 等) で SDR, STD のセッションが組まれている。

このような状況をふまえ、日本語アドホック音声ドキュメント検调用テストコレクションに続き、SDPWG は日本語 STD 用テストコレクションの構築を 2008 年度から開始し、テストコレクションの β 版を策定し、2009 年 10 月の音声言語情報処理研究会で中間報告を行った [7]。研究会でのコメント・意見に基づき STD テストコレクションの改定を行い公開した [8]。より多くの研究者が利用できるよう、「SDR/STD 研究・開発を行う多様な利用者を想定し、複数かつ単純な検索語と正解セットの提供」を目指すもので、本テストコレクションは NIST のテストコレクションと比べ以下のような特徴を有する。

- NIST における音声ドキュメントが 1~3 時間と小規模であるのに対し、本テストコレクションでは近年の音声ドキュメント量の増大を前提に大規模化することに焦点を絞り、44~604 時間と大規模なものとした。
- NIST の検索語は、既知語・未知語と分けずコンテストの際に公開された。結果的に、STD システムのトータ

¹ 岩手県立大学
Iwate Prefectural University, Takizawa Iwate 020-0193, Japan

² 山梨大学
University of Yamanashi, Kofu, Yamanashi 400-0016, Japan

³ 独立行政法人情報通信研究機構
National Institute of Communication Technology (NICT) Koganei, Tokyo 184-0015, Japan

⁴ 龍谷大学
Ryukoku University, Otsu, Kyoto 520-2194, Japan

⁵ 豊橋技術科学大学
Toyohashi University of Technology, Toyohashi, Aichi 441-8122, Japan

⁶ 東京工科大学
Tokyo University of Technology, Hachioji, Tokyo 192-0982, Japan

⁷ 京都大学
Kyoto University, Kyoto 606-8501, Japan

⁸ 大学共同利用機関法人情報・システム研究機構統計数理研究所
The Institute of Statistical Mathematics, Tachikawa, Tokyo 190-0014, Japan

⁹ 立命館大学
Ritsumeikan University, Kusatsu, Shiga 525-8577, Japan

a) y-itoh@iwate-pu.ac.jp

るな評価を狙ったものであるが、本テストコレクションでは既知語と未知語の検索語セットを分け、既知語/未知語それぞれの検索語に対する検索方式/検索性能の比較が明確にできるようにした。また、公平な評価ができるよう、音声認識結果以外に、音響モデル、言語モデル、音声認識辞書の提供も行った。

本テストコレクションは、公開以来、様々な研究機関で利用されている。国立情報学研究所が主催する2011年のNTCIR-9では、SDPWGが提案したSDR/STDの評価タスクが受理され、SDPWGのメンバがこのタスクのオーガナイザとなり、本テストコレクションをdry run（試行用評価）に用いたほか、正式な評価を行うformal run用のテストコレクションへの展開も行った。

本論文では、まずNISTが2006年に設定したテストコレクションを概説し、日本語テストコレクション構築にあたっての方針、テストコレクションの内容、ベースライン評価結果とテストコレクションの分析結果について説明するとともに、構築したテストセットによって活性化したSTD研究の展開について述べる。

2. NISTにおけるSTD用テストコレクション [1], [2]

本章では2006年にNISTが設定したSTDタスクにおける、音声ドキュメント、検索語、正解、評価方式について簡単に説明する。

2.1 音声ドキュメント (Spoken documents)

音声ドキュメントは表1のとおりで、アラビア語（近代標準語とレバノン系）、中国語（標準語）、英語（米語）の3種の言語について、放送ニュース音声（Broadcast News）、電話での会話音声（Telephone Conversation）、会議音声（Roundtable Meetings）の3種の音声タイプについて、1～3時間のデータが用意された。

2.2 検索語 (Query terms)

1～4単語からなる検索語1,100個が用意された。3, 4-gramは100個、bigramは約400個、unigramは約500個

表1 NISTの評価用音声ドキュメント（言語/タイプ/時間）と検索語（数/出現回数）

Table 1 Spoken documents (language/speech type/length) and query terms (number/occurrence) for NIST evaluation.

		Arabic	Chinese	English
Broadcast News		1 hour	1 hour	3 hours
Telephone Conversation		1 hour	1 hour	3 hours
Roundtable Meetings		No	No	2 hours
Number of	Reference	1100	1120	1100
	Occurrences	5240	3684	14421

が選定された。3単語からなる検索語の場合、trigramを高頻度順に並べ、trigram中の各単語のunigram確率が高いものを除外し上位80個選定する。10個はコーパス外の最新のtrigramを選定し計90個とした（4-gramは10個）。bigramは3, 4-gramをbigramに展開したものの201個、trigramと同様な方法で189個、コーパス外のtrigramを展開した20個が選定された。

NISTのSTDタスクにおいてはコンテスト形式での評価が行われた（参加者はForbidden Data以外は学習用に利用可能）。このため検索語が既知語か未知語かは明確には定められていない。

2.3 正解と評価方式

ある検索語に対して、システムが出力した区間の中心時刻が、実際にその検索語が話されている位置との差が0.5秒以内であれば正解と見なされる。

評価は検出精度と速度については以下の項目が評価の対象となる。

- Indexing time
- Indexing memory consumption (high water mark)
- Index size (on disk)
- Search time for each term
- Searching memory consumption (high water mark)
- Computing platform benchmarks

STDにおいては性能が検索対象の音声ドキュメント長、正解の出現頻度、検索語の長さ等、様々な要素により影響される。今回、誤って検出する確率として性能を評価することで出現頻度等の問題を回避すべく、DET (Detection Error Tradeoff: ミスの確率とFA (False Alarm) の確率のカーブ) とATWV (Actual Term Weighted Value), MTWV (Maximum Term Weighted Value) 等の新たな指標が導入された。

3. 日本語版：音声中の検索語検出テストコレクション

本章では、SIG-SLPの音声ドキュメント処理ワーキンググループが構築した日本語版の音声中の検索語検出用テストコレクションについて説明する。

3.1 基本的な方針

すでにSDR, STD研究を行っている研究者だけでなく、新たに関連する研究を始める研究者も利用できるように、「SDR/STD研究・開発を行う多様な利用者を想定し、複数かつ単純な検索語と正解セットの提供」を目指すこととした。

研究向けに提供・公開したものは以下のとおりである。

- ① 検索語セット
- ② 正解データ

- ③ 音声認識システムで使用した音声認識辞書
- ④ 音声認識用音響モデルと言語モデル
- ⑤ 音声ドキュメントの認識結果

本テストコレクションにおける音声ドキュメントは日本語話し言葉コーパス (CSJ) [9] を検索対象ドキュメントとした。CSJ を所有している研究者・研究機関であれば利用可能である。音声認識環境が整っていない研究者等にも利用できるよう、音声認識結果以外に、音響モデル、言語モデル、音声認識辞書の提供も行った。

3.2 テストコレクション

3.2.1 音声ドキュメント

前節のとおり本テストコレクションにおける音声ドキュメントは CSJ とした。CSJ は実際の講演音声と模擬講演から構成されており、全部で 3,302 の音声データが収録されている。本テストコレクションでは、このうち学会講演 987 講演、模擬講演 1,715 講演の計 2,702 講演、604 時間の音声ドキュメントを検索対象データとし、これを「ALL」セットとする。ALL のうちの 177 講演 (学会講演 70, 模擬講演 107), 約 44 時間の音声ドキュメントを「CORE*¹」セットとする。CORE 中の講演には、豊富な研究用情報が付与されているため、これらの情報を利用した STD 研究も行うことが可能である。

NIST の音声ドキュメントが 1~3 時間と小規模であるのに対し、本テストコレクションでは近年の音声ドキュメント量の増大を前提に、44~604 時間と大規模なものにすることとした。転記テキストを有し大規模で誰もが利用可能な CSJ 以外の音声ドキュメントが少なかったため、また講演や講義音声のアーカイブ化が進む中で、これらのデータに対する実際の利用場面が想定されるため、本テストコレクションでは、講演音声の主とする CSJ を検索対象とした。NIST では放送ニュース音声、電話音声、会議音声の 3 種類を用いており、他の種類の音声については収集・書き起こしを含め今後の課題としたい。

3.2.2 検索語セット

検索語のセットとして以下の 4 つのセットを策定した。実際の検索語と出現頻度等は文献 [8] にリストアップされている。NIST の検索語は既知語・未知語と分けておらず、トータルな STD システムの評価を狙ったものであるが、本テストコレクションでは STD の方式の違いをより明確にするために、既知語の検索語セットと未知語の検索語セットとを分け、そのときの言語モデルまで提供することとした。これにより既知語/未知語それぞれに対する検索方式と検索性能がより明確に比較できるようになった。

(1) 既知語検索語セット

- a) ALL 講演用既知語セット：ALL 講演セットを対

象とした 100 検索語

- b) CORE 講演用既知語セット：CORE 講演セットを対象とした 50 検索語

(2) 未知語検索語セット：

- a) ALL 講演用未知語セット (未構築)
- b) CORE 講演用未知語セット：50 検索語

(3) 簡易評価のための CORE 講演用 50 検索語

以下、それぞれの検索語セットを説明する。

(1) 既知語検索語セット

a) ALL 講演用既知語セット：100 検索語

全音声ドキュメント 2,702 講演を対象として、音声認識システムの語彙に登録されている検索語を 100 個用意した。検索語は 1 語 (形態素) 以上からなる内容語とした。実際の検索場面を想定し、その検索語が検索において意味がある言葉であるよう TF・IDF 値が高い言葉を抽出し、最終的には重要と思われる検索語を人手を介して選定した (NIST では各単語は低頻度の出現率だが N-gram としては高頻度のものを機械的に選定)。検索語長に性能が左右されると想定し、4~11 モーラの検索語を各 12 個、12 モーラの検索語を 4 つ用意した。

b) CORE 講演用既知語セット：50 検索語

CSJ の CORE 講演を対象とし、比較的小規模な音声データを検索対象とした。検索語はすべて ALL 講演用既知語セットに含まれているものである。4~12 モーラの検索語をモーラ数ごとに 8 個程度用意した (9 モーラ以上は適切なものが少ないため、9 モーラは 6 個、10 モーラは 2 個、11, 12 モーラは 1 個とした)。

(2) 未知語検索語セット

検索語は 1 つ以上の形態素から構成されており、検索語を構成している形態素のうち、半数以上が未知語のものを未知語検索語として採用する。未知語は音声認識辞書の作り方によって異なる。辞書の作成方法については、次節で述べる。

a) ALL 講演用未知語セット

ALL 講演セットの中で、出現頻度が 2 回以下の検索語を選定したが、600 時間の音声から 1, 2 語を見つけるため非常に難しいタスクである。このセットについては、今後の課題としたい。

b) CORE 講演用未知語セット：50 検索語

未知語となる検索語は CORE 以外の CSJ 中で、出現頻度が少ないかまったく含まれないもので、CORE 講演に出現する名詞、複合名詞から選定した。

(3) 簡易評価のための CORE 講演用 50 検索語

CORE 講演中の 49 講演約 19 時間を音声ドキュメントとして、すでに文献 [6] で未知語として検索性能評価を行っている。この実験で用いた検索語、音声ドキュメントセットと性能を提供した。自分のシステムを簡易に評価できる

*¹ 文献 [9] の CSJ 「コア」には 201 講演含まれるが、本論文の「CORE」では対話・朗読の 24 講演を除いている。

表 2 奇数・偶数セットのデータ分布

Table 2 Analysis of training data for odd and even sets.

	講演数(学会, 模擬)	時間(学会, 模擬)	発話数	単語数
奇数セット	1,255 (458, 797)	281h (127, 154)	16.7 万文	343 万語
偶数セット	1,270 (459, 811)	287h (130, 157)	17.0 万文	351 万語
合計	2,525 (917, 1,608)	568h (257, 311)	33.7 万文	694 万語

表 3 CSJ 講演の音声認識率 (1-best) [%]

Table 3 Speech recognition rates for CSJ presentation speeches (1-best) [%].

	単語認識結果		音節認識結果	
	単語正解率	単語正解精度	音節正解率	音節正解精度
ALL 講演セット	74.1	69.2	80.5	73.3
CORE 講演セット	76.7	71.9	81.8	77.4

枠組みである。

3.2.3 音声認識

3.1 節で述べたように音声認識結果と認識の際のモデル等を提供する。本項では、音声ドキュメントを音声認識する際の認識方式、認識条件を説明する。音響・言語モデルの作り方や認識方式については同一にすることで公平な評価が行えると考えられる。本テストコレクションを用いて音声認識の性能や検索性能をベースライン性能と比較する場合には、以下の認識方式を参考にされたい。

CSJ 講演音声の音声認識をオープンな条件で行うため、次の方式を用いた。

- 各講演音声には固有の ID 番号が付与されている。この ID の下 1 桁の数字により 2 分割し、各々を偶数セット、奇数セットと呼ぶ。
- 偶数セット、奇数セット、それぞれから音響モデル、言語モデルを学習する。認識辞書は、奇数・偶数セット双方に共通で出現する語を登録する。ただし、CORE 講演は各モデルの学習や辞書作成に用いない。
- 偶数セットの音声認識は奇数セットから学習したモデルを、奇数セットの音声認識には偶数セットから学習したモデルを利用する。

奇数・偶数セットのデータを表 2 に示す。以下、詳細に説明する。

(1) 音声認識辞書

単語音声認識辞書に登録されている形態素数は約 27,000 語となった。形態素の定義は、Chasen with UniDic [10] (今回、Chasen 2.4.4, UniDic 1.3.9 を利用) に従った。CSJ の転記テキストをこの条件で形態素解析し、形態素に分割した。奇数セット、偶数セットに共通で出現し、かつ CORE を除く ALL 講演での出現頻度が 3 回以上のものを辞書に登録した。辞書サイズを大きくすると候補となる未知語自体が少なくなり未知語の検索語が設定しにくくなる。辞書サイズを小さくすると認識率が悪くなり評価が難しくなることから上記のような設定とした。音節認識の際の辞書は

すべての音節で、今回 265 個となった。

(2) 言語モデル

言語モデルは、単語・音節ともに、奇数・偶数セットごとに、(1) で形態素解析したテキストから Palmkit 1.0.32 を用いて単語 trigram モデルを学習した。

(3) 音響モデル

奇数・偶数セットごとに、triphone ベースの HMM を学習した。使用する音声特徴量は多くの文献で用いられている、MFCC(12 次) + ΔMFCC + ΔΔMFCC + ΔPower(1 次) + ΔΔPower の合計 38 次元とした。ガウス分布の混合数は 32、総分布数は 3,000 とした。

(4) 音声認識

音声認識エンジンには、Julius ver.4.1.3 を用いた。Julius のオプションにより、N-best 出力やコンフュージョンネットワークの出力を行うことができる。研究用に提供できる認識結果には、単語および音節認識結果の 10-best 出力とコンフュージョンネットワーク、ラティスの出力を含めた。前述した条件による ALL 講演セットと CORE 講演セットの音声認識率 (1-best) を表 3 に示す。

3.2.4 正解判定基準と評価指標

CSJ 中の転記テキスト上で検索語の文字列と一致する箇所を正解とした (表記の揺れはないものとし、同音異義語は正解としない)。NIST の正解判定基準では、正解音声区間の中心と検索語検出区間の中心との差が 0.5 秒以内とされているが、CSJ には単語境界情報が CORE にはあるが ALL にはないため、「検出した区間が正解フレーズ*2 (発話) に含まれていれば正解」とした。STD のアプリケーションを考えたとき、ユーザは検索語が検出された区間のみではなく、検索語を含むフレーズ全体を聞くことで内容確認を行うと想定され、今回の正解判定基準では、検索語が含まれているフレーズを特定できればよいとした。各検

*2 CSJ の転記テキストには、0.2 秒のポーズによって分割された単位で転記が行われており、この単位を Inter-Pausal Unit (IPU) と呼ぶ。本論文では特にフレーズと呼ぶことにする。

表 4 Yahoo キーワードランキングにおける固有名詞・外来語 (個: 30 個のうち)

Table 4 Proper nouns and loanwords in Yahoo Japan keyword ranking (the number among 30 keywords).

年	固有名詞	外来語	日本語(一般)
2012	22	11	4
2011	21	10	5
2010	24	12	5
2009	21	12	3

索語の正解フレーズリストを提供した。

STD においては性能が検索対象の音声ドキュメント長, 正解の出現頻度, 検索語の長さ等, 様々な要素により影響される. NIST では, 誤って検出する確率として性能を評価することで出現頻度等の問題を回避すべく, DET と ATWV, MTWV 等の新たな指標が導入された. 今回の検索指標としては ATWV 等についても検討したが, 挿入誤りの重みを決める β 等, 大規模な CSJ で適切な値を導くことが困難なことから, 再現率 (Recall), 精度 (Precision), F 値, および以下の式で表される MAP (Mean Average Precision) で評価することとした.

$$AP(q) = \frac{1}{N} \sum_{k=1}^{R_q} \delta_k \times precision(q, k) \quad (1)$$

$$MAP = \frac{1}{|Q|} \sum_{k=1}^{|Q|} AP(k) \quad (2)$$

検索語 q の平均適合率 $AP(q)$ を式 (1) で求める. q の正解数を N , 1 位の候補から出力し R_q 位ですべての正解が出力されたとし, k 位を出力した際, 正解ならば δ_k は 1, 不正解ならば 0 となる. AP は正解出力時の Precision の平均値を意味する. これを検索語ごとに求め, その平均を式 (2) より求め, これが MAP (Mean Average Precision) となる.

ATWV は音声ドキュメント長に依存しない指標として考案された (実際には前述のとおり FA の確率との重み β が不確定). MAP は同一の音声ドキュメント, 検索語に対しては性能を比較するうえで適切である一方, 音声ドキュメント長や検索語の出現頻度により MAP の値が変わるため, 様々な音声ドキュメント, 検索語に対してより適切な評価指標の開発は今後の課題としたい.

必要な資源については NIST 同様, 次の指標で評価できると考える.

- 索引に要する時間
- 索引付けの必要メモリと索引サイズ
- 検索語あたりの検索時間
- 検索に要するメモリ
- 用いた計算機の能力

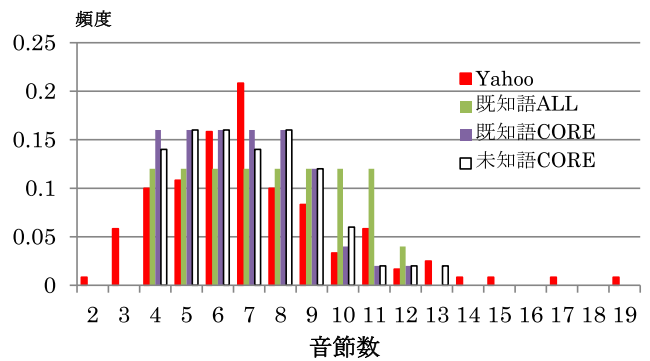


図 1 検索語の音節数と出現頻度

Fig. 1 Syllable numbers and occurrences among query term sets.

3.3 検索語の評価

本節では, 3.2.2 項で今回選定した検索語について検索に使われる言葉としての妥当性を分析・評価する.

表 4 に 2009~2012 年の 5 月 30 日近辺の Yahoo のキーワードランキング (30 位まで) における固有名詞と外来語の数を示す (外来語の固有名詞は両方にカウントした). これらの検索語の 7~8 割は固有名詞, 3 割強が外来語であった. そのほとんどが前述の音声認識システムの辞書では未知語となっていた. テストコレクションについては, 未知語の検索語は 9 割が固有名詞, 19 個 38%が外来語となっており, 検索語の実態と合っていると評価できる.

一方, 既知語の検索語では, 固有名詞は 24%, 外来語は 31%と実態より少ないが, 辞書に登録される高頻出単語を選定しているため, これらの言葉が選定されにくいのは当然ともいえる.

上記の 4 年分の検索語と, 既知語 ALL と CORE, 未知語 CORE それぞれのセットにおける検索語の音節数の分布を図 1 に示す. 既知語の 2 つのセットは音節数 4~8 で一定数とした. Yahoo の検索語では 3-11 音節に検索語が多く出現しており, テストコレクションの 3 つの検索語セットでも同様な部分を中心に選定されていることが分かる. 特に未知語の検索語の分布は, 実態との整合性が高い.

Yahoo の検索語では人名が多く, その音節数が 6, 7 になることからこの音節数の頻度が高くなっている. CSJ では人名は伏せられていることが多く人名を検索語とすることは難しい側面があった. この部分を除いては既知語, 未知語の検索語セットの音節数, 固有名詞・外来語の比率等の面からは, 結果的にはあるが今回の検索語セットは実際の検索語の選択状況に沿っていると考える.

4. ベースライン検索性能

本テストコレクションの単純な検索方式を用いてベースライン評価を行った. 以下, 既知語セットと未知語セットごとのベースライン評価結果について述べる.

表 5 既知語セットベースライン評価 [%]
Table 5 Baseline results for known query term sets [%].

	単語レベル完全一致			音素レベル完全一致 (参考)		
	再現率	精度	F 値	再現率	精度	F 値
ALL 講演セット	49.7	87.6	63.5	50.7	89.8	64.8
CORE 講演セット	54.1	93.6	68.6	55.3	96.9	70.4

4.1 既知語検索語セット

既知語の検索語セットの評価は、音声ドキュメントを前節で述べた条件下で音声認識を行い、検索語が正しく音声認識されているフレーズを検索結果として抽出した。ALL・CORE 講演セット双方とも同じ評価基準で行った。結果を表 5 に示す。参考までに、音素レベルでの完全一致 (検索語と音声認識結果を音素表記に変換し照合する) に基づく検索結果も表に併記した。表 5 の数値は、検索語ごとの平均値である。

ALL 講演セットよりも CORE 講演セットの方が音声認識率は高いため、検索性能としては CORE の方が良くなった。精度が高いのは、検索語に誤って誤認識される単語の事例が少なかったためである*3。単語レベルの完全一致よりも音素レベルでの完全一致の方が、同音異義語に誤認識された検索語をカバーできるため性能が良い。

4.2 未知語検索語セット

未知語の検索語の評価では、単語レベルの完全一致を行うことができない。そこで、各発話の音声認識結果を音素系列に変換し、検索語の音素系列との連続 DP によるスポッティングを行う。DP の際の局所距離は、Edit distance (編集距離) とし、置換、挿入、脱落誤りに対して、それぞれ距離を 1 と定義した。連続 DP の結果、平均距離が一定の閾値以下の場合に、その発話に検索語が含まれると判断する。

CORE 講演の音声認識を行う際、2 通りの認識方式で実施した。(1) は 3.2.3 項で説明した方式であり、(2) の言語モデルについては、学習テキストを音節列表記 (たとえば、“フーリエ”なら“fu u ri e”*)⁴に変換し、音節単位の trigram を学習して構築した。

- (1) 単語認識方式：認識単位を単語、言語モデルを単語 trigram とする。
- (2) 音節認識方式：認識単位を音節、言語モデルを音節 trigram とする。

上記の 2 つの認識方式を用いた際の CORE 講演の音声認識率と、ベースラインの検索性能を表 6、表 7 に示す。連続 DP で用いた閾値は、F 値が最適になる値を採用した。

表 6 のとおり、単語認識の方が音節認識よりも音節認識

表 6 CORE 講演の音節認識率 [%]

Table 6 Speech recognition rates for a CORE set in CSJ [%].

音声認識方式	正解率	正解精度
単語認識	86.5	83.0
音節認識	81.5	77.4

表 7 CORE 講演用未知語セットベースライン評価 [%]

Table 7 Baseline results for an unknown query term set for a CORE set [%].

音声認識方式	連続 DP		
	再現率	精度	F 値
単語認識	40.2	36.4	38.2
音節認識	56.4	51.2	53.7

率は約 5 ポイント高いが、未知語検索語が発話されている区間に対しては、音節認識方式の方が、検索語の発音により近い音素系列を出力することができるため、表 7 が示すように単語認識方式より高い未知語検出が可能だった。

4.3 テストコレクションの分析

本節では、テストコレクションのベースライン性能を示し、検索語の特性 (モーラ数と出現回数) と検索性能についての関連性を分析した。

4.3.1 ベースライン性能の Precision-Recall (精度-再現率) カーブ

本論文では、以下の 6 つの組合せについての Precision-Recall カーブを提供する。

- (1) CORE 講演に対して：未知語セットにおける単語認識と音節認識の検索結果
- (2) CORE 講演に対して：既知語セットにおける単語認識と音節認識の検索結果
- (3) ALL 講演に対して：既知語セットにおける単語認識と音節認識の検索結果

図 2 に上記の 6 つの条件における Precision-Recall カーブを示す。

図が示すように、検索語が既知語であると高い精度で認識されているため、単語認識結果を音素列に変換して用いた方が、以下の方式より検索性能が高くなる。

- 単語認識結果から直接検索語を検索する方式
 - 音節認識結果から音素レベルの連続 DP で検索する方式
- 検索語が既知語の場合、音声ドキュメントの大小 (ALL, CORE) に関係なく、同程度の Precision-Recall カーブと

*3 検索語が検出できなかった場合に精度を 0 としているため、表 4 の精度は悪いように見える点に注意。

*4 音節を構成する際には長母音はこの例のように母音を 2 回用いることで対応した。

表 8 単語・音節認識結果に対する DP による検索結果と検索語との相関

Table 8 Correlations between query term sets and F-measures obtained by using DP for word/syllable recognition results.

セット		認識単位	最大 F 値	MAP	F 値との相関		
					モーラ数 閾値固定	モーラ数 閾値最適	出現回数
CORE	既知語	単語	0.725	0.705	0.292	0.488	0.049
		音節	0.554	0.627	0.136	0.435	-0.051
	未知語	単語	0.382	0.532	-0.040	0.322	0.112
		音節	0.537	0.642	-0.030	0.237	0.096
ALL	既知語	単語	0.729	0.641	0.050	0.273	0.336
		音節	0.515	0.531	0.162	0.488	0.246

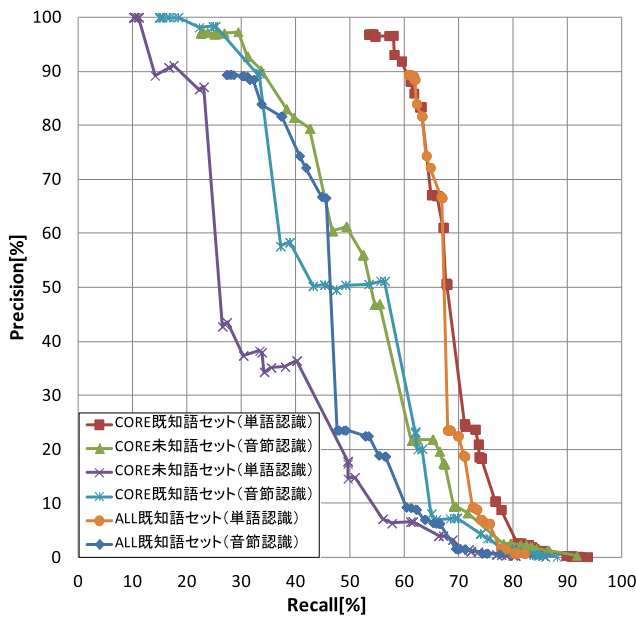


図 2 単語・音節認識結果に対する DP による検索結果
Fig. 2 STD performance using DP for word/syllable recognition results.

なったが、音声ドキュメントが大きい ALL の場合には CORE に比べ湧き出し誤りが増えるため高い Precision が得られず MAP が低下した。

検索語が未知語の場合、音節認識結果を音素列に変換して用いた方が、単語認識結果を使うよりも検索性能が高くなる。既知語と未知語と比べると処理方式が同じとなるため検索性能に大きな違いはない。

4.3.2 検索語と検索性能の分析

表 8 に示したように、検索語のモーラ数と F 値、検索語の出現回数と F 値との相関を各検索語セットについて求めた。この結果と相関を求めた際の散布図を分析し、以下の知見を得た。

(1) モーラ数と F 値について

モーラ数の多い検索語の方が検索性能は高いと一般に想定されるが、DP の出力をコントロールする閾値が一定の

場合には相関はなかった。一方、各検索語で閾値を最適にすると相関が得られた。詳細は以下のとおり。

- 単語認識から得られた音素列に対する DP において、CORE 既知語セットについては弱い相関があり (0.292)、ALL 既知語セットについては相関がなかった (0.050)。モーラ数の多い検索語は出現頻度が低く、単語認識に失敗していることが要因と推察する。
- 音節認識から得られた音素列に対する DP の結果では、モーラ数が多いほど性能が高くなると想定したが、全体の F 値を最大にする閾値に固定した場合 (表中：閾値固定)、相関は出なかった (0.136, -0.040, 0.162)。F 値最大の閾値だと厳しすぎて、9 つの検索語で正解がえられず F 値が 0 になっていた。そこで閾値を各正解で最適な F 値を示すよう設定すると (表中：閾値最適)、相関が見られるようになった (0.435, 0.237, 0.488)。

(2) 検索語の出現回数と F 値について

出現頻度と検索性能には相関がないと想定したが、出現回数が多い検索語は適切な言語モデルが構築され、正しい音声認識と、正しい検出が行われるため、出現回数が 50 回以上の検索語では F 値と弱い相関が認められた。詳細は以下のとおり。

- ALL 既知語セットでは、出現回数が多い (50 回以上) 単語は認識率が高くなる傾向にあり (言語モデルの出現確率が高いため)、出現回数と F 値で弱い相関 (相関値：0.336) が認められた。ただし、出現回数 50 回以下では相関値が -0.09 とほとんど相関はなかった。
- CORE 既知語セットでは、出現回数と F 値の相関は認められなかった。出現回数が多くない (すべての検索語が 50 回以下) ことが要因と考える (ALL 既知語セットの 50 回以下の検索語で相関がなかったことと同様)。

本テストコレクションのベースライン性能としては図 2 および表 8 を参照されたい。

5. STD テストコレクションの公開と研究利用状況

5.1 テストコレクションの公開と NTCIR

本テストコレクションは、2009年10月にβバージョンをアドホック検索のテストコレクションとともに「<http://www.cl.ics.tut.ac.jp/~sdpwg/>」で公開した。現時点で、本テストコレクションの利用申請は15件あり、公開して以来、様々な研究機関で利用され、新しい研究が展開されている。次節で、本テストコレクションの公開により波及した研究例・研究会等について概説するが、本章では新しい検索語セットの提供も行った NTCIR について述べる。

本テストコレクションの公開を契機として、SDPWG が働きかけて評価型の国際ワークショップ NTCIR-9 において IR for Spoken Documents Task が採択され、STD タスクと SDR タスクの評価型コンテストが 2011 年実施され、12 月にそのワークショップが開催された。NTCIR-9 では以下のテストコレクションの提供を行った [11]。以下、正式な評価を formal run と呼び、その実施に向けて formal run の数カ月前に、共通の形式での検索結果の提出とその予備評価が行われた。この予備評価を dry run と呼ぶ。

dry run 用検索語・正解セットと formal run 用検索語・正解セットは以下のとおりで、dry run 用の検索語・正解セットは本テストコレクションにより公開したものである。formal run 用のものは NTCIR-9 用に新たに構築したものであり、本テストコレクションとは異なり国立情報学研究所から評価用ツールとともに配布されている [21]。

- CORE 用既知検索語 50 語とその正解リスト
- CORE 用未知検索語 50 語とその正解リスト
- ALL 用既知検索語 100 語とその正解リスト

5.2 STD 研究の展開

本テストコレクションの公開により STD に関する研究が推進・活性化され、様々な研究発表が行われるようになった。本節では STD 研究の活性化を表すデータを示すとともに、本テストコレクションを用いた実際の研究事例と研究の展開を紹介する。

5.2.1 STD 関係の研究発表の増加

- 音声ドキュメント処理ワークショップ：
(<http://www.cl.ics.tut.ac.jp/~sdpwg/>)
2010 年 3 月 4 件
2011 年 3 月 6 件
2012 年 3 月 7 件
- 日本音響学会研究発表会
2010 年 春季 4 件 秋季 4 件
2011 年 春季 5 件 秋季 5 件
2012 年 春季 6 件

- NTCIR-9 (2011.12) 7 グループ

2010 年以前は国内では、5 つ程度の研究グループでしか STD の研究発表が行われていなかったが、現在では 10 以上 (山梨大学、立命館大学、豊橋科学技術大学 3、岩手県立大学、岐阜大学、龍谷大学、日立製作所、産業技術総合研究所、NICT) の研究グループが継続的に様々な研究を推進しており、毎年上記の研究発表会で研究発表がなされている。

5.2.2 テストコレクションの STD 研究利用状況

STD システムの基本的枠組みでは、前章の評価でも述べたように、音声ドキュメントを以下の 2 通りの方法であらかじめ認識しておく。

- ① 単語を認識単位とする大語彙連続音声認識
- ② 音節のようなサブワードを認識単位とするサブワード連続音声認識

検索語をテキストで与えられると、その検索語が①の大語彙連続音声認識の辞書中の既知語であれば、①の結果を利用することで高精度・高速に検索が可能である。検索語が辞書中に存在しない未知語であれば、①の結果を用いるよりも実際に発話された音素系列に近い②のサブワード連続音声認識結果を用いた方が高精度な検索が期待できる。しかし、一般にサブワード認識は誤認識を含んだサブワード列が出力されているため、検索性能は①の既知語検索に比べ低下する傾向がある。また、検索語が与えられてから、音声ドキュメント全体に対してフルテキストサーチや近似サーチを行うと検索時間を要することになる。

①の精度向上は音声認識技術そのものであり、高速化もすでにテキスト検索技術で確立されているため、STD の研究としては主に検索性能の向上と高速化が最近の主であり、実際に STD の性能向上を狙うもの [12], [13], [14] と STD の高速化を図るもの [14], [15], [16], [17], [18] が多く見受けられる。一方、STD を音声認識や SDR に活用する研究 [20] 等の新しい研究も見られるようになった。以下、それぞれについて簡単に述べる。

(1) STD の高精度化

STD の高精度化を狙う文献 [12], [13] の研究は、ともに音声ドキュメントをサブワードであらかじめ認識する際に、複数の音響モデル/複数の言語モデルを並行して使い、複数のサブワード認識結果を用意しておく。音声ドキュメントの各音声区間には複数のサブワード候補があるため、誤認識へ頑健性を高めるものである。

文献 [12] では、音響モデルは 2 種類、言語モデルは単語や音節等 5 種類を用意し、これら音響/言語モデルのすべての組合せ (2×5) の 10 システムで音声ドキュメントを認識しておく。10 本の音素同期型の Confusion network を構成することで候補の網羅性を高め、同一音素の候補数等を DP での局所距離とすることで誤検出を抑制し検索精度を向上させた。

表 9 NTCIR における CORE 講演未知語の STD 結果の評価 (文献 [11] の Table 4 の抜粋)

Table 9 An evaluation of STD performances for an unknown query term set and a CORE set in NTCIR (an excerpt of Table 4 in Ref. [11]).

	System	F 値 (max)	F 値 (spec.)	MAP	検索時間 [s]
	baseline	0.527	0.516	0.595	36.4
1	ALPS-1	0.725	0.708	0.837	13.5
2	IWAPU-1	0.644	0.628	0.772	3.5
3	AKBL-1	0.393	0.393	0.264	0.0017
4	NKGW-1	0.645	0.585	0.491	0.0016
5	NKI11-1	0.570	0.559	0.684	0.00094
6	RYSDT-3	0.521	0.334	0.469	2.9
7	YLAB-1	0.425	0.425	0.344	38.8

注) この結果は NTCIR の formal run set を用いて得られたものである。

文献 [13] では、サブワードとして音節, triphone, 半音素等, 時間長等が異なる複数のサブワードで音声ドキュメントを認識しており, 検索語に対し各認識結果を並列に検索しその結果を統合することで検索性能の向上を実現した. 2つのサブワードモデル間の音響距離を統計的に求め, 照合時の局所距離としている。

文献 [14] では, 音節認識する際に各フレームのサブワード HMM における各状態のスコアをすべて残しておき, 検索語が与えられたときに, 音素 trigram 等で粗く候補を絞った後, 音響スコアで再スコアリングすることで, 検索精度の向上と高速化を実現した。

(2) STD の高速化

高速化は一般にサブワード認識した結果を転置インデックスとして単語とその出現位置のペアを有しておくことで実現されるが, 単語の境界が不明なサブワードでは以下に述べるように様々な方法が提案されている。

文献 [15] では, Julius と SPOJUS の 2つの認識システムで音節認識を行い, 各々の N ベスト (5 ベスト程度) の結果と大語彙音声認識の結果の音節列に対し, n-gram をインデックスとして高速な検索を実現した. 置換誤りに対しては音響距離を割り当て, 脱落・挿入をインデックス構築の際に疑似的に作成することで検索精度を改善した。

文献 [16] では, 転置インデックスの代わりに Suffix array を使い, Suffix array では Suffix となるインデックスが長くなると検索時間を要するため, 6 程度の音素に Suffix の長さを固定することで高速なキーワード検索を実現した。

文献 [17] では, すべての音節 2-gram, 3-gram に対して事前に検索して候補を用意しておき, 検索語に含まれる音節 2-gram, 3-gram から候補区間を絞り込み, 連続 DP で詳細照合する候補を少数にすることで高速化を実現した。

文献 [18] では, 各音節について事前に類似位置を検出しておき, 検索語に含まれる音節列と音声ドキュメントの音節列の 2次元マトリックス上に, 検出しておいた音節を類

似度の高い順に配置し, その配置が一直線として検出できる区間を抽出する方式で, 従来にはない画像処理的な要素が加えられている。

(3) STD 研究における本テストコレクションの利用状況と STD 研究の展開

文献 [19] では, 音素や音節等のサブワードの代わりに, セグメントごとの音響特徴量を量子化した記号列で音声ドキュメントを表現しており, 新しい検索語検出のアプローチとして注目されている。

文献 [20] では, 音節認識結果に対し, 質問文を構成する単語を検索語とした STD を行い, 各語の出現情報を用いて各文書と質問文との関連度を計算し文書検索を行うもので, STD そのものの研究ではなく STD を応用した新しい研究の展開を示すものである。

これらの研究成果から得られた知見を簡単にまとめる. STD システムでは未知語の精度向上が重要課題であり, 未知語の精度向上のためには複数のサブワード認識の結果を利用し, 候補を様々保持しておくことで高い効果がある. 検索の高速化では, 転置インデックスや Suffix array のようなインデックスを効果的に用いることで検索性能の低下を最小限に抑えながら高速化が実現可能である. また, 1 段目に粗い検索で候補を絞り徐々に精度の高い検索を行う複数段階の検索方式も高速化に効果的である。

5.3 テストコレクションを用いた評価

表 9 にベースライン性能と NTCIR に参加したグループの検索性能を示す. F 値 (max) は F 値が最大値をとるところを事後的に決めた閾値, F 値 (spec.) は dry run から設定した閾値を示す. 前述のように 1, 2 は複数の認識ユニットから検索性能の向上を図るもので Exact matching に基づく Baseline よりも検索性能で向上している. 検索時間は 1 検索語あたりの平均検索時間を示している. 3~5 は高速化を図ったもので, Baseline よりも検索時間が 1,000 倍以上

向上しているのが分かる。5については局所距離に音素弁別特徴を取り入れているため性能面での向上も見られる。

上述のとおり、本テストコレクションのベースとした共通基盤上で性能を比較することで、各手法の利点、有効性を明確にすることができている。このように本テストコレクションはSTDの研究を適切な方向に、より効率的に推進する土台を与えるものと考えられる。

5.4 課題

(1) 評価方法の統一

今回、音声ドキュメント、検索語、正解箇所の統一したデータの提供を行い、共通の評価ができるようにしたが、NTCIR-9の結果を見ると必ずしも同一評価になっているわけではない。文献[12], [13], [14]では同じ音節を認識単位に使っているが、独自の音節リスト・認識システムを使っているためにその時点での検索性能が異なっており、一概に提出された検索性能だけで方法の優劣を比較できない側面がある。文献[13]では長母音があるだけで音節リストのエントリ数が文献[12]の倍近くなり、検索性能も異なると予想される。このほか、認識時のパラメータ等によっても検索性能は左右されるため、さらに厳密な制限での比較方式の検討が必要と考える。

(2) 検索時間の評価

検索時間が0.01秒のシステムと0.1秒のシステムでは10倍の検索時間の差であるが、ユーザには待ち時間の差は感じられないであろう。検索時間を1秒以内としたときの検索性能等で評価する方式も必要と考える。STDでは1段目は粗く検索し、2段目以降で精緻な検索を実施する方式が多い。STDの検索結果は一覧性がないため、得られた結果を1つずつ聞かないと確認できない。この確認時間中にも検索の精緻化を進めることが可能である。このような考え方から高適合候補だけ早く出力できることもSTDシステム実装上は有効であり、最上位の出力、あるいは10位以内の出力の適合率、検索時間での評価方式が必要であり、これらについては今後の課題としたい。

(3) 正解判定基準について

正解判定基準は現在、3.2.4項で述べたように「検出区間が正解フレーズに含まれる」としており正解位置にまで言及していない。「正解区間の中心と検出区間の中心との差が0.5秒以内」といったNISTと同様な厳密な基準における評価も求められる場合が想定され、厳密な正解区間の提供は今後の課題としたい。

6. おわりに

情報処理学会音声言語情報処理研究会における音声ドキュメント処理ワーキンググループがSDR評価用テストコレクションの構築に続きSTDの評価用テストコレクションの構築を行った。本論文ではこのSTDテストコレ

クションを概説し、本テストコレクションの特徴(TRECの1, 2時間の音声ドキュメントと比べ44~604時間と大規模とした点、既知語検索語・未知語検索語と分けた点等)を示した。またベースラインとなるSTDの性能を示すとともにその分析を行い、既知語/未知語の検索語の場合の適切な検索方式と検索語(モーラ長と出現頻度)と検索性能の相関を確認した。さらに本テストコレクションによって活性化されたSTD研究状況について述べた。

今後、本SDR/STDテストコレクションは幅広い専門分野の研究者によって利用され、国内外での本分野の研究・応用の活性化に寄与できるものと考えられる。また2012年度に国立情報学研究所主催のNTCIR-10において、継続してSDR/STDシステムの評価を計画しており、本テストコレクションをベースにより多くの研究者によるさらなる研究の進展を期待したい。

参考文献

- [1] 2006 Spoken Term Detection Evaluation, available from (<http://www.itl.nist.gov/iad/mig/tests/std/2006/index.html>).
- [2] NIST: The spoken term detection (STD) 2006 evaluation plan, (2006). available from (<http://www.itl.nist.gov/iad/mig/tests/std/2006/docs/std06-evalplan-v10.pdf>).
- [3] Garofolo, J.S., Auzanne, C.G.P., Voorhees, E.M.: The TREC spoken document retrieval track: A success story, *9th Text Retrieval Conference (TREC-9)*, NIST (2000).
- [4] 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, *IPSJ Journal*, Vol.50, No.2, pp.1234-1245 (2009).
- [5] Logan, B. and Van Thong, J.M.: Confusion-based query expansion for OOV words in spoken document retrieval, *Proc. ICSLP* (2002).
- [6] 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).
- [7] 伊藤慶明, 西崎博光, 胡新輝, 南條浩輝, 秋葉友良, 相川清明, 河原達也, 中川聖一, 松井知子, 山下洋一: 音声の中の検索語検出のためのテストコレクション構築—中間報告, 情報処理学会研究報告, Vol.2009-SLP-78, No.4, p.8 (2009).
- [8] 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, p.8 (2010).
- [9] 前川喜久雄: 『日本語話し言葉コーパス』の外観, Version 1.1, 入手先 (<http://www.kokken.go.jp/katsudo/seika/corpus/releaseinfo/040/overview.pdf>).
- [10] 形態素解析辞書 UniDic, 入手先 (<http://www.tokuteicorpus.jp/dist/>).
- [11] Akiba, T., Nishizaki, H., Aikawa, K., Kawahara, T. and Matsui, T.: Overview of the IR for Spoken Documents Task in NTCIR-9 Workshop, *NTCIR-9* (2011).
- [12] 古屋裕斗, 名取賢, 西崎博光, 関口芳廣: 音声の中の検索語検出における検出誤り抑制パラメータの検討, 第6回音声ドキュメント処理ワークショップ, SDPWS2012-11

- (2012).
- [13] 伊藤慶明, 岩田耕平, 小嶋和徳, 石亀昌明, 田中和世, 李時旭: 語彙制限のない音声文書検索における複数サブワードの統一検索語彙に依存した検索性能推定指標の導入, 情報処理学会論文誌, Vol.50, No.2, pp.524-533 (2009).
 - [14] 神田直之, 住吉貴志, 小窪浩明, 佐川浩彦, 大淵康成: 多段リスコアリングに基づく大規模音声中の任意検索語検出, 電子情報通信学会論文誌 D, Vol.J95-D, No.4, pp.969-981 (2012).
 - [15] 岩見圭祐, 山本一公, 中川聖一: 複数音声認識システムを併用した音節 n-gram 索引による検索性能の改善, 第 6 回音声ドキュメント処理ワークショップ, SDPWS2012-10 (2012).
 - [16] 勝浦広大, 桂田浩一, 入部百合絵, 森本容介, 辻 靖彦, 青木久美子, 新田恒雄: 放送大学の講義音声を対象とした高速キーワード検索の性能評価, 第 6 回音声ドキュメント処理ワークショップ, SDPWS2012-05 (2012).
 - [17] 齊藤裕之, 伊藤慶明, 小嶋和徳, 石亀昌明, 田中和世, 李時旭: 複数音節の事前検索結果を利用した音声中の検索語検出の高速化, 第 6 回音声ドキュメント処理ワークショップ, SDPWS2012-06 (2012).
 - [18] 金子泰輔, 秋葉友良: 部分距離空間上の索引を用いた STD における距離順計算の厳密化と非直線検出への拡張, 第 6 回音声ドキュメント処理ワークショップ, SDPWS2012-08 (2012).
 - [19] 松永 徹, 趙 國, 山下洋一: 音声ドキュメントの音響情報セグメント量子化を用いた音声検索語検出, 第 6 回音声ドキュメント処理ワークショップ, SDPWS2012-07 (2012).
 - [20] 瀧上智子, 秋葉友良: STD に基づく未知語に頑健な音声ドキュメント検索法, 第 6 回音声ドキュメント処理ワークショップ, SDPWS2012-12 (2012).
 - [21] NTCIR formal run, available from <http://research.nii.ac.jp/ntcir/permission/ntcir-9/perm-en-SPOKENDOC.html>.



伊藤 慶明 (正会員)

1989 年東京大学大学院工学系研究科航空工学専攻修士課程修了。同年川崎製鉄(株)に入社。1992 年より技術研究組合新情報処理開発機構に出向。2000 年より岩手県立大学准教授。博士(工学)。人工知能学会, 日本音響学会, 電子情報通信学会, IEEE 各会員。



西崎 博光 (正会員)

2003 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。2007 年同大学助教。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。



中川 聖一 (フェロー)

フェロー。

1976 年京都大学大学院博士課程修了, 同年京都大学工学部情報工学科助手, 1980 年豊橋技術科学大学情報工学系講師, 1990 年同大学情報工学系教授, 現在, 同大学大学院工学研究科情報・知能工学系教授, 電子情報通信学会



秋葉 友良 (正会員)

1995 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。2001 年独立行政法人産業技術総合研究所に組織移行。2004 年豊橋技術科学大学工学部助教。現在, 豊橋技術科学大学工学部准教授。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。



河原 達也 (正会員)

1987 年京都大学工学部情報工学科卒業。1989 年同大学大学院修士課程修了。現在, 同大学学術情報メディアセンター教授。音声言語処理, 特に音声認識および対話システムに関する研究に従事。IEEE, 日本音響学会, 電子情報通信学会, 人工知能学会, 言語処理学会各会員。



胡 新輝 (正会員)

1988 年ハルビン工業大学大学院通信システム学修士課程修了。1995 年東京大学より Ph.D. 取得。1995 年(株)富士ソフト, 2001 年(株)東芝研究開発センター。2003 年年 ATR 音声言語通信研究所, 2009 年より情報通信研究機構 NICT 音声コミュニケーション研究室で研究者として現在に至る。日本音響学会会員。



南條 浩輝 (正会員)

1999年京都大学工学部情報学科卒業。2001年同大学大学院情報学研究科修士課程修了。2004年同大学院情報学研究科博士後期課程修了。同年龍谷大学理工学部助手。2007年同助教。



松井 知子

1988年東京工業大学大学院修士課程修了。同年NTT(株)入社。1998年ATR音声翻訳通信研究所。2000年ATR音声言語通信研究所および音声言語コミュニケーション研究所に出向。2003年情報・システム研究機構統計数理研究所准教授。2008年同研究所教授。東京工業大学博士(工学)。IEEE senior member, 日本音響学会, 日本統計学会各会員。



山下 洋一

1982年大阪大学工学部電子工学科卒業。1984年同大学大学院修士課程修了。同年大阪大学産業科学研究所文部技官。1993年同助手。1994年同講師。1997年立命館大学理工学部助教。2001年同教授。2004年同大学大学院情報理工学研究科教授。博士(工学)。電子情報通信学会, 日本音響学会, 人工知能学会, ISCA, IEEE 各会員。



相川 清明 (正会員)

東京大学大学院工学研究科電子工学専門課程博士課程修了(工学博士)。NTT研究所, カーネギーメロン大学客員研究員, ATR主任研究員, NTT研究所音声認識グループ。マルチモーダル対話研究グループ。情報処理学会, 電子情報通信学会, 日本音響学会, IEEE, アメリカ音響学会各会員。