

## SLP 音声ドキュメント処理ワーキンググループ活動報告

秋葉友良 (豊橋技術科学大学) 相川清明 (東京工科大学) 伊藤慶明 (岩手県立大学)  
原原達也 (京都大学) 南條浩輝 (龍谷大学)  
西崎博光 (山梨大学) 安田宜仁 (NTT) 山下洋一 (立命館大学)  
松井知子 (統計数理研究所) 胡新輝 (NICT/ATR)  
中川聖一 (豊橋技術科学大学) 伊藤克亘 (法政大学)

**あらまし** 「音声ドキュメント処理ワーキンググループ」は、情報処理学会音声言語情報処理研究会の下、音声ドキュメント処理研究の促進のため、平成18年4月から3年間の活動を行ってきた。本稿では、ワーキンググループでのこれまでの活動、今後の計画、および関連研究を報告する。

**キーワード** ワーキンググループ, 音声ドキュメント処理, データ収集, 情報検索, テストコレクション

### Progress Report of SLP Spoken Document Processing Working Group

Tomoyosi AKIBA (TOYOHASHI UNIV. OF TECHNOLOGY), Kiyooki AIKAWA (TOKYO UNIV. OF TECHNOLOGY), Yoshiaki ITOH (IWATE PREFECTURAL UNIV.), Tatsuya KAWAHARA (KYOTO UNIV.), Hiroaki NANJO (RYUKOKU UNIV.), Hiromitsu NISHIZAKI (UNIV. OF YAMANASHI), Norihito YASUDA (NTT), Yoichi YAMASHITA (RITSUMEIKAN UNIV.), Tomoko MATSUI (ISM), Xinhui HU (NICT/ATR), Seiichi NAKAGAWA (TOYOHASHI UNIV. OF TECHNOLOGY), and Katunobu ITOU (HOSEI UNIV.)

**Abstract** The Spoken Document Processing Working Group, which was organized in special interest group of spoken language processing, IPSJ (Information Processing Society of Japan), has been working for three years from April 2006, to promote the research activities about “spoken document processing.” This paper describes our activities so far and in future.

**Key words** Working Group, Spoken Document Processing, Data Collection, Information Retrieval, Test Collection

#### 1. はじめに

ことばには、二つの重要な機能がある [1]。一つは、相互作用の (interactional) な機能である。これは、人間がことばを用いて、他人に気持ちを伝えたり、社会的関係を構築・維持したりするための機能である。ことばには、もう一つ重要な機能がある。それは、業務的 (transactional) な機能である。これは、ことばを用いて、知識や技術・情報を伝達する機能である。

業務的な機能は、人類がことばを使いはじめた初期の段階では、口伝えという形で、話しことばが担っていただろう。しか

し、文字が発明されると業務的な機能は次第に書きことばが担うようになり、紙、印刷術といった発明によって、書きことばの業務的な機能は飛躍的にひろがってきた。

古代においては、業務的な機能を多く担っていたはずの話しことばは、近代以降においては、相互作用的な機能の面ばかりに使われているように思える。しかし、音声認識などの技術により洗練させることで、話しことばそれ自体を業務的な機能のメディアとして利用できるようになるのではないだろうか。そのような発明によって、書きことばより話しことばの方が表現しやすい内容の共有、再利用を促進し、人間の新たな知的活動

を刺激することになるのではないだろうか。

書きことばの世界で、特に情報管理の面を強調した形態を「文書」「ドキュメント」とよぶ。この名称にちなんで、我々は、平成 18 年 4 月に情報処理学会音声言語情報処理研究会の下に「音声ドキュメント処理ワーキンググループ」を発足させた。このワーキンググループでは、音声ドキュメント利用の第一歩である記号化(文字化)は音声認識技術で一定の成果をおさめているという前提のもとで、ドキュメントを処理するために不可欠となる技術開発を促進するため、研究のための基盤整備に取り組んできた。本稿では、本ワーキンググループのこれまでの活動を報告するとともに、今後の活動の方針について展望する。

## 2. 音声ドキュメント処理

書きことばを対象とした研究のうち、情報管理の側面に着目した処理を総称して「ドキュメント処理」と呼ぶことがある。ドキュメント処理の研究項目には、文書の交換、閲覧、加工などが含まれる。さらに、文書の構造化や文書の検索(タグを用いるものと全文検索の両方が含まれる)、質問応答、要約などの技術も含まれる。我々は、音声ドキュメント処理を音声メディアに対するこれらの処理を総称するものとして定義する。

ワーキンググループの活動は、音声ドキュメント処理を対象とした、大学・企業・研究機関での研究・開発を促進することを目的に、取りかかりとして次の 2 課題の達成を目標とした。**研究用音声ドキュメントの発掘** 音声ドキュメント処理研究のための基盤データとして、研究用途になるべく低コストかつ制約無く使用可能なデータが望まれる。それらのデータは、主要な研究課題への使用に耐えるか、音声ドキュメント処理の処理対象としての有効性、などの観点からも吟味される必要がある。そのような条件を満たす、既存データの発掘、新規データの構築、収集したデータの整備、などの作業を行う。**音声ドキュメント検索研究用テストコレクションの構築** 情報検索は、大量のドキュメントを処理するために、まず最初に考慮されるべき技術である。そのため、まずは音声ドキュメント検索研究のための基盤整備に取り組んだ。テキストを対象とした情報検索の研究に倣い、テストコレクションの構築を目指した。

各課題の活動状況について、それぞれ 3. 節、4. 節で述べる。

## 3. 研究用音声ドキュメントの発掘

音声ドキュメント処理研究のための基盤データとして利用可能なデータの調査を行った。データは、研究課題となる技術の実際の応用の場に合致するものであるとともに、なるべく低コストかつ制約無く研究利用可能であることが望まれる。我々はまず、コンテンツとして大学での講義および学会講演に着目し、各大学・研究機関が保有し研究利用可能な講義・講演音声のアンケート調査を行った。これらのデータは、学術的な内容検索、要約、e-learning 用途の視聴・ブラウジングシステムの研究開発に必要である。以下、現段階で利用可能、もしくは整備予定のデータについて報告する。

土屋らは、日本語講義音声コンテンツコーパス(CJLC)[2]を

公開している<sup>(注1)</sup>。このコーパスは、静岡大学、山梨大学、豊橋技術科学大学の 3 大学で実際に開講された講義(16 話者、22 科目、114 講義、3860 分)を収録・設計・編纂したものである。音声データ、書き起しテキストに加え、マルチメディアデータとしてスライドデータおよびビデオデータ、また、話者情報、収録環境情報、発話時間情報、重要発話、スライド切り替え時間情報などの付随データを含んでいる。

学会講演データとして、後述する(5.1 節)第 1 回、第 2 回音声ドキュメント処理ワークショップでの約 40 講演について、音声データ(ヘッドセットマイク、ピンマイク、ハンドマイク、など数種類)、スライドデータ(音声との同期情報含む)、ビデオデータ、を収録した。今後、利用方法を検討していく予定である。

## 4. 音声ドキュメント検索研究用テストコレクション

本ワーキンググループにて構築した音声ドキュメント検索用テストコレクションの概要について述べる。テストコレクションについてより詳しい情報は、文献[3]~[5]を参照されたい。

### 4.1 テストコレクションの概要

テキストを対象とした検索評価用テストコレクションは、一般に、検索対象文書集合、検索クエリ、正解文書集合から構成される。音声ドキュメントを対象とする場合は、さらに、人書き起しテキストおよび音声認識による標準的な書き起しも用意されていることが望ましい。

#### 4.1.1 検索対象

現時点でもっとも広く利用可能な日本語音声データとして、「日本語話し言葉コーパス」(CSJ)[6]を検索対象に選んだ。CSJは、数百時間以上の規模を持ち、書き起しも利用できる、唯一のコーパスである。テストコレクションでは、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 時間を越える。

#### 4.1.2 検索クエリ

上記検索対象の講演数は 2,702 である。講演を検索の単位(文書)とすると、既存の文書検索テストコレクションに比べて対象文書コレクションのサイズが極端に小さくなり、検索の問題設定が簡単になり過ぎる。そこで、講演中のおよそスライド 1 枚程度の説明に相当する連続する 1 分程度の区間を検索対象文書とする検索タスクを設定し、39 問の検索クエリが作成された。

#### 4.1.3 適合性判定と正解文書

対象文書集合から各検索クエリに適合する文書を判定し、正解文書を定める。ただし、上記した検索クエリの設計により、本検索タスクでの適合性判定は、発話を単位とした可変長の連続する区間に対して行なわれた。判定は、適合(Relevant)、部分適合(Partially Relevant)、不適合(Irrelevant)のいずれかである。検索クエリ 39 問に対する適合性判定の統計量を表 1 に

(注1) : <http://www.slp.ics.tut.ac.jp/CJLC/>

表 1 質問作成&適合性判定の概要

	質問あたり 判定区間数	質問あたり 異なり講演数	判定区間あたり 平均発話数
適合	11.18	7.90	10.39
適合& 部分適合	12.69	9.26	10.88

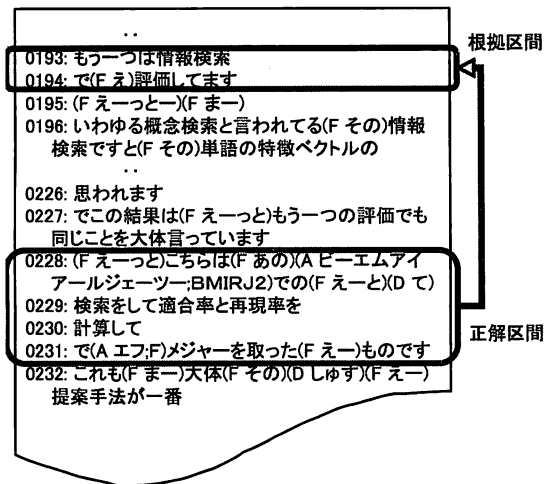


図 1 正解区間と根拠区間

示す。

適合性の判定に際し、根拠の存在は必須とされた。例えば、「情報検索性能を評価するにはどのような方法があるか知りたい」という質問に対し、「適合率」や「再現率」などが現れる発話は、それが情報検索の性能評価手法かどうか文書を見ただけでは判定できないので、適合ではない。適合性判定では根拠を表す発話区間も同時に特定されなければならない。ただし、根拠は、適合性判定対象の発話区間中に現れていなくても、同一講演の他の箇所、あるいは講演全体から読み取っても良いとされた。正解区間と根拠区間の例を図 1 に示す。また、構築した検索クエリと適合性判定の例を図に示す。

#### 4.1.4 音声認識による自動書き起し

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。

#### 4.1.5 TREC SDR テストコレクションとの比較

米国 NIST 主催の評価型ワークショップ TREC で構築された音声ドキュメント検索テストコレクション [7] との比較を表 2 に示す。言語 (英語-日本語) や対象文書 (ニュース放送-講演) の違いはあるものの、検索対象音声データのサイズはほぼ等しい。また、ニュース放送の 1 ストーリーが数分程度であるのに比べて講演は数十分と長く、したがって CSJ の総文書数は 8 分の 1 程度と少ない。しかし、CSJ SDR では講演中の 1 分程度の部分を探す検索タスクを設定してあるため、検索対象の粒度

表 2 TREC SDR と CSJ テストコレクションの比較

	TREC9 SDR	CSJ SDR
言語	英語	日本語
対象文書	ニュース音声	講演音声
時間	557 時間	623.6 時間
文書数	21,754	2,702 (30,762)
単語/文書	169	2,324.9 (204.2)
検索クエリ	50	39
書き起し	低品質 (WER 10.3%)	高品質
音声認識 WER	26.7%	21.4%

(CSJ の括弧内数字は、30 発話単位を文書とした場合。)

表 4 再定義した文書検索タスクの統計

疑似文書長	15	30	60	Lecture
疑似文書数	60,202	30,762	16,060	2,702
平均正解文書数	16.36	12.77	10.90	8.13

は同程度である。また、節で述べるように、講演中の 30 発話区間を文書とみなした検索タスクを設定すれば、文書数は約 3 万となり、TREC SDR と同程度の文書検索タスクが設定できる。

## 4.2 基本性能調査

構築したテストコレクションの難易度を調査することを目的に、ベースライン検索性能の評価を行った。ベースラインとなる手法として、音声書き起しテキストを索引付けに用いたテキストベースの文書検索手法を試した。また、人手書き起しテキストを索引付けに用いた場合と比較することで、対象が音声になることによる検索タスクの難易度の変化を調べた。

## 4.3 検索タスク設定

4.1.3 節で述べた通り、構築したテストコレクションの正解は可変長の発話区間である。現状では、可変長区間を対象とした検索タスクは難易度が高過ぎて、ベースライン評価としては適切ではないと考えた。そこで、あらかじめ検索対象の講演を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み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 発話および講演全体の 4 通りを試した。2702 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一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」と判定された発話とした<sup>(注2)</sup>。再定義したタスクの疑似文書数および検索クエリあたりの平均正解文書数を表 1 に示す。また、正解文書数の分布を図 2 に示す。

## 4.4 検索手法

書き起しテキストだけを用いた一般的な文書検索手法を実装した。自動または人手による書き起しテキストを形態素解析し、形態素の基本型によって索引づけした。索引語重みには、文書

(注2): 「適合」に加えて「部分適合」と判定された発話も正解とした場合も試したが、検索性能に大きな違いは見られなかった。

表3 クエリと適合性判定結果の例

発話区間	判定	根拠区間
<b>SDPWG-HN2010-02:</b> "煙草が体に及ぼす影響、有害性にはどのようなものがあるか?"		
などの炭水化物の取り過ぎによってビタミン (A ビーワン; B 1) 欠乏状態になる例が最近あるらしいので	適合	副腎皮質ホルモンの分泌を盛んにさせるストレスの増加や喫煙は/ビタミン (A シー; C) をより多く消費させるということなんです
たばこは/肺癌の七十二パーセント/喉頭癌の九十六パーセント/膀胱癌でさえ三十一パーセントの原因があると言われてます	適合	
ニコチンは/血管を収縮させ/血の巡りを悪くします	適合	煙には/ニコチン/さまざまな発癌物質/発癌促進物質/一酸化炭素/さまざまな線毛障害物質/その他/四千種以上の化学物質が含まれ/そのうち有害物質は/確認されただけでも/二百七十種あります
<b>SDPWG-HN2012-01:</b> "ワインの産地を知りたい。有名もしくは個人的に好まれている地方のワインについて、特に知りたい。"		
フランスのシャンパーニュ地方で造られた/発泡性のワインでことであります	適合	次は東の横綱フランスワイン
(F (? あ)) 更に (D ワ) ワイン以外で私大好きなのが (F あの) シャンパンなんですけれどもこれはフランスはシャンパーニュ地方の	適合	
(F あのー) 南フランスのバイヨンヌっていうところなんですけども/(F あの) 非常に田舎町です	部分適合	いうことですね (F あのー)(F ま) 結構ワイン/がおいしいと
<b>SDPWG-HN1014-01:</b> 講演音声の特徴について知りたい		
それから講演音声は脱み上げ音声のモデルよりも (F えーと) 対話音声のモデルに近い発話スタイルに/なっていると/いうことも (F まー) 言えると思います	適合	
(F えー) その一方で (F ま) 講演音声というものの特徴を考えていきますと (F えー)(F ま) 話し言葉の冗長的な表現というものを多く含みまして	適合	
講演である為に丁寧な/<FV> 口調で話されておりますので丁寧語が/各所入っております	部分適合	
<b>SDPWG-YY1016-01:</b> 尊敬されている人やものについて知りたい		
多分父親本当に尊敬してる人は父親だけだと思うんですけど	適合	
合掌造りの里とか曲がり屋とか	適合	.../(F あー) 私達が見て/とても尊敬 (D (? し))/に値すると思います
(F ま)(A ケー; K) 理事長	適合	(F あ) 尊敬する二人のトップを横軸に/話をしてみたい/と/思います/(F あ) まず
<b>SDPWG-YI1014-01:</b> DP マッチングを用いた研究を探したい		
でこれは二つのモジュールからなってますで第一段階で統合モジュール/これにより (A ディーピー; DP) マッチングを行ないまして各システムが出す単語列というものの対応を取ります	適合	
(F と) 本日の発表で/< 雑音 >/ (F えー)/主眼を置いているのはこの/(A ディーピー; DP) マッチングを (D おく) 連続 (A ディーピー; DP) を行なう際の距離尺度なんですここを/色々/< 雑音 >/変えてみようと考えています	適合	キーワードを/(F えー)(A ディーピー; DP) マッチング連続 (A ディーピー; DP) を行なった結果 (D き)/(F えー) 得られたパスというものはこのように (D も)
システムの方で音声区間抽出/(F え)(A ディーピー; DP) マッチングを行ない/整合経路の表示を行ないます/これが/その (A ディーピー; DP) マッチングをした時の結果の例/です	適合	
<b>SDPWG-TA1001-01:</b> "情報検索性能を評価するにはどのような方法があるか知りたい。"		
他方が (F あーの) 犠牲になるというような関係に基本的になりますでしたがって評価尺度の再現率と精度っていうのも普通は	適合	ですから (F (? ん))(F えーっと) いい検索システムというのは <FV> 両方の尺度ができるだけ高いと/いうことになります
通常の情報検索システムの/出力 (D2 と)/でよく使われる (F え) 平均精度/で (F えー) ランキングを評価する方法そして	適合	検索結果を評価する基準ですが/でこれに関しても二通り
でその日英検索の十一平均適合率を/取ると	不適合	(根拠がない)

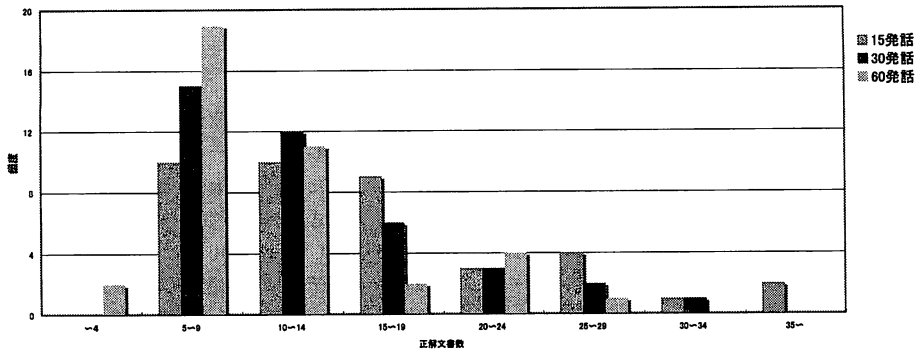


図2 検索クエリあたり正解文書数の分布

長で正規化した TF-IDF 重みづけ手法 [8] 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 [9] を用いた。検索対象文書のテキストには、音声認識結果の 1-best 候補だけを用いた場合、10-best 候補までを連結して用いた場合、人手書き起しテキストを用いた場合、の 3 種類について比較した。

#### 4.5 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  $AP$  について、全検索クエリで平均をとった 11 点平均精度を用いた。

$$IP(x) = \max_{\{i: x \leq R_i\}} P_i \quad AP = \frac{1}{11} \sum_{i=0}^{10} IP\left(\frac{i}{10}\right)$$

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。実際には、1 つの検索クエリに対して上位 1000 件まで文書を検索した。

#### 4.6 評価結果

図 3 に実験結果を示す。人手書き起しを用いる場合に比べて、音声認識による自動書き起し (1-best) を用いる場合、10 から 15%程度検索性能が低下することが分かる。これに対し、比較的短い文書を検索対象とする場合 (15 発話, 30 発話) は、認識結果の 10-best 候補を利用することで検索性能を改善することができている。一方、長い文書 (60 発話, 講演) を検索対象とする場合には、効果が見られない。

TREC SDR のストーリー検索精度は 0.45 前後と報告されている。ほぼ同条件の 30 発話検索タスクは、書き起しを使った場合でも 0.22 と半分程度の性能である。理由の一つとして、今回設定した検索タスクが人工的に生成した文書を用いているため、文書が内容的に一貫した単位となっておらず検索が困難になったことが挙げられる。その他、講演を対象としたことに起因する理由が考えられる。

第一に、学会講演は同じ研究分野の研究者を対象に行われるので、聴講者との知識の共有を前提に基本的な説明を省略する傾向がある。そのため、重要なキーワードが講演中に現れないことが多い。第二に、講演は、音声発話だけでなくスライド

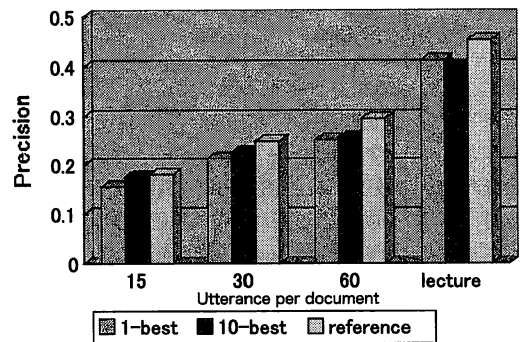


図3 自動書き起し (1-best, 10-best) および人手書き起しを用いた索引付けによる 11 点平均精度

併用して行われる。そのため、重要な用語を発声せずにスライドを指示するだけで進める場合も少なくない。例えば、研究の目的となるようなキーワードが表題の部分だけで述べられていることも多い。(しかし、CSJ の書き起しでは講演題目が削除されている。) また、「見れば分かる」実験結果の数字はスライドに掲載されるだけで、それに関するコメント (結果の差を述べるなど) だけが発話されることも多い。特に、全国大会など時間の短い (12 分程度の) 講演では、発表時間を補うためスライド資料を効率的に使うので、この傾向が大きい。

#### 4.7 テストコレクションを用いた研究事例

構築したテストコレクションの利用例として、ワーキンググループメンバーによる研究事例について述べる。

Akiba ら [10] は、音声ドキュメントを対象としたアドホック検索において、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に対し、統計的な翻訳モデルにより認識候補から正解テキストを予測して索引付けを行う検索手法を提案した。この研究では、テストコレクションの CSJ 書き起しテキストと認識結果をパラレルテキストとして利用し翻訳モデルの学習に利用している。また、上述した固定発話長の文書検索タスクを用いて手法の評価を行っている。

胡ら [11] は、音声ドキュメントの未知語検索のために、階層

的な言語モデルを使って単語と subword を結合した Confusion Network を作成し、これをベクトル空間モデルに適用したキーワード検索手法を提案している。この研究では、テストコレクションから検索キーワードを作成し、手法の評価を行っている。

#### 4.8 Term Detection テストコレクションの構築

前節までで述べたテストコレクションは、検索者の情報要求をキーワードリストや自然言語文で与え、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」タスクを対象とした。一方、与えられた用語 (term) が出現する音声データの位置を特定する「キーワード検索」タスクは、NIST Term Detection Project などをはじめ、近年でも活発に研究が行われている。そこで、CSJ を検索対象として、100 キーワード程度の Term Detection 用テストコレクションを構築することを計画している。

### 5. その他の活動

#### 5.1 音声ドキュメント処理ワークショップ

音声ドキュメント処理に関する研究発表の場として、豊橋技術科学大学メディア科学リサーチセンター<sup>(注3)</sup>主催、情報処理学会音声言語情報処理研究会 (SIG-SLP) 共催のもと、平成 19 年 2 月、および平成 20 年 2 月に、豊橋技術科学大学にて「音声ドキュメント処理ワークショップ」を開催した。

第 1 回ワークショップは、平成 19 年 2 月 26 日から 27 日の 2 日間にわたり開催した。一般発表セッションでは、講義データの収集・分析、e-learning、音声ドキュメント検索、言語モデル、マルチメディアコンテンツの生成、音声ドキュメントの認識・処理手法など、音声ドキュメント処理に関する研究テーマを網羅する多様なテーマで 22 件の発表が集まった。また、NHK 技研の八木伸行博士 [12]、神戸大学の有木康雄教授 [13] に、音声ドキュメントに対するメタデータの構築に関する特別講演をお願いした。

第 2 回ワークショップは、平成 20 年 2 月 29 日から 3 月 1 日の 2 日間にわたり開催した。音声ドキュメント処理の応用として期待される研究分野に関して、東京大学相澤清晴教授に「ライフログ」[14]、名古屋大学長尾確教授には「ディスカッションマイニング」[15]、の各テーマについて、特別講演をお願いした。一般発表セッションでは、音声ドキュメントの認識、質問応答・対話、検索、シーン検出・分割や、講義・講演処理、応用システムといった分野で 21 件の発表が行われた。

第 1 回第 2 回ともに、70 名近い参加者があり、会場は終始満席の状態で活発な議論が行なわれ、本分野への関心の強さが感じられた。引き続き平成 21 年 2 月には、第 3 回ワークショップを開催する予定である。

#### 5.2 「音声ドキュメント処理」特集号

音声ドキュメント処理に関する論文発表を促進するため、情報処理学会論文誌の特集号として「音声ドキュメント処理」特集号 (編集長: 中川聖一) を企画した。平成 20 年 5 月投稿締切で、29 件の投稿があった。編集作業を経て、平成 21 年 2 月号

に論文誌掲載の予定である。

### 6. まとめ

音声ドキュメント処理ワーキンググループのこれまでの活動について報告した。本ワーキンググループは、3 年間の活動で収集・構築したデータの有効利用を検討すべく、活動の継続を申請し、承認された。今後は、構築した検索テストコレクションの拡充およびそれを利用した研究の促進や、今後の研究動向に即した新規データの発掘・構築、および収集したデータの研究利用方法について議論していく予定である。

#### 文 献

- [1] G. Brown and G. Yule: "Discourse Analysis", Cambridge University Press (1983).
- [2] M. Tsuchiya, S. Kogure, H. Nishizaki, K. Ohta and S. Nakagawa: "Developing corpus of Japanese classroom lecture speech contents", Proceedings of International Conference on Language Resources and Evaluation (2008).
- [3] 伊藤克互, 相川清明, 秋葉友良, 伊藤慶明, 河原達也, 南條浩輝, 西崎博光, 安田直仁, 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, 安田直仁, 山下洋一, 伊藤克互: "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] T. Akiba, K. Aikawa, Yoshiaki Itoh, T. Kawahara, H. Nanjo, H. Nishizaki, N. Yasuda, Y. Yamashita and K. Itou: "Test collections for spoken document retrieval from lecture audio data", Proceedings of International Conference on Language Resources and Evaluation (2008).
- [6] K. Maekawa, H. Koiso, S. Furui and H. Isahara: "Spontaneous speech corpus of Japanese", Proceedings of LREC, pp. 947-952 (2000).
- [7] J. S. Garofolo, E. M. Voorhees, V. M. Stanford and K. S. Jones: "TREC-6 1997 spoken document retrieval track overview and results", Proceedings of the 6th Text Retrieval Conference, pp. 83-91 (1997).
- [8] A. Singhal, C. Buckley and M. Mitra: "Pivoted document length normalization", Proceedings of ACM SIGIR, pp. 21-29 (1996).
- [9] "汎用連想検索エンジン GETA", <http://geta.ex.nii.ac.jp>.
- [10] T. Akiba and Y. Yokota: "Spoken document retrieval by translating recognition candidates into correct transcriptions", Proceedings of International Conference on Speech Communication and Technology, pp. 2166-2169 (2008).
- [11] 胡新輝, 呉友政, 柏岡秀紀: "階層化言語モデルによる音声ドキュメントの検索", 情報処理学会研究報告, SLP-072, pp. 31-36 (2008).
- [12] 八木伸行: "映像・音声・言語解析を用いたテレビ番組のメタデータ制作", 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. 33-40 (2007).
- [13] 有木康雄: "音声を中心とするマルチメディアのメタデータ化", 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. 41-46 (2007).
- [14] 相澤清晴: "ライフログ: 体験の情報処理", 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, pp. 33-40 (2008).
- [15] 長尾確: "ディスカッションマイニング: 会議からの知識発見", 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, pp. 41-50 (2008).

(注3): <http://marc.tut.ac.jp>