談話標識の抽出に基づいた講演音声の自動インデキシング

長谷川 将宏† 秋田 祐哉† 河原 達也†

講演音声において話題（セクション）の転換点で用いられる特徴的な単語（談話標識）を用いて，自動インデキシングを行う方法を提案する．本研究では，種々の講演の中でも流れが比較的明確で共通性のある学会講演を対象とする．学習データの講演の書き起こしからポーズ情報を用いてセクション境界候補を検出し，統計的言語モデルを用いて句点を挿入して，各セクションの先頭の 1 文を抽出する．その中に含まれる名詞から単語頻度と文頻度に基づいて談話標識を選定する．これらの過程は人手によるタグを必要としない教師なし学習により行われる．評価データの各文について談話標識の単語頻度と文頻度の統計量に基づく評価値を計算し，その合計が閾値以上であればインデックスを付与する．実際の講演音声の書き起こしと音声認識結果に対して評価を行った結果，再現率 85%程度（適合率は 20%程度）の精度でインデキシングできることを示す．

Automatic Indexing of Lecture Speech by Extracting Discourse Markers

Masahiro Hasegawa,† Yuya Akita† and Tatsuya Kawahara†

We address a method of automatic indexing for lecture speech using suggestive words that frequently appear in the initial sentences of sections, and we define such words as discourse markers. We deal with academic presentations because these presentations can be segmented into relatively distinct parts. At first, we segment transcriptions into sections with average duration of pauses in the lecture as a threshold. Next, each section is segmented into sentences by using a statistical language model. Then, discourse markers are selected from nouns based on term frequency and sentence frequency statistics. We evaluated these discourse markers with recall and precision rates on indexing task of lecture speech. Sentences are indexed if the sum of the term frequency and sentence frequency statistics of detected discourse markers exceeds a threshold. As a result, we achieved a recall rate of 85% with precision of 20%.

1. 緒 論

近年，計算機の性能が飛躍的に向上し，音声メディアをディジタルアーカイブとして保存できるようになった．しかし，音声アーカイブは一見して内容を把握することがテキストや映像以上に困難である．したがって，求める情報を効率良く検索するには，音声アーカイブにインデックスが付与されていることが必要であるが，インデックスを人手で付与することは手間と時間を要し，大量のデータに対して行うことは困難である．そこで，音声認識技術を利用することを考える．

本研究では講演音声を対象として自動インデキシングの検討を行う．講演には予稿があることが多いが，実際に予稿を読み上げる人はほとんどおらず，かなり自由な発話が行われ，話し言葉特有のくだけた言い回しや言い淀み，発話速度の変化や発声の怠けなどの種々の特徴が含まれる．そのため，現在の音声認識技術では講演のような話し言葉を高い精度で認識することができない．自動認識結果には誤りが多く含まれているため，そのまま講演録として使用できるレベルではなく，人手による修正や編集が必要である．したがって音声メディアの形で保存しておいたうえで，音声認識結果を利用してそのインデックスを自動的に付与することを考える．すなわち，本研究では，認識結果から要約を作成する[1),2)]のではなく，録音音声の参照を容易にするための自動インデキシングについて検討する．

認識率が十分に高くなくても話題同定や話題境界への分割ができる可能性は大きい．これまでに，放送ニュース[3)~6)]やボイスメール[7)]の話題分類などが研究されている．それらの大半が，特徴的なキーワードを抽出することによって話題の分類を行っており[8)]，ニュースやボイスメールのような多くの短い音声セグ

† 京都大学大学院情報学研究科知能情報学専攻
Graduate School of Informatics, Kyoto University

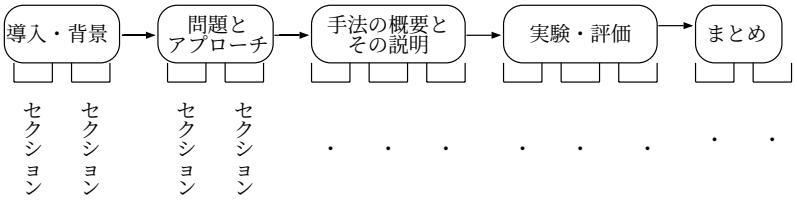

図 1 学会での研究発表（工学系）のモデル
Fig. 1 Model flow of academic presentation at technical conferences.

メント（長くて数分）からなる場合は効果的である．

しかしこの方法は，全体の話題は変わらずに細かい論点が次々と展開される講演や会議のような長い音声（数十分）に適用するのは困難である．すなわち，このような音声では話題に依存したキーワードの大半は全体で出現するが，全体の話題分類を行ってもあまり意味がない．一方，このような種類の長い音声ではインデックス機能の重要性が大きい[9),10)]．特に，セクションの境界にインデックスを付与できれば，スキップしながら聞きたい部分を探すことができるので，きわめて有用である．

本論文では，セクションの境界を検出することによって講演音声にインデックスを付与することを考える．従来の研究が話題依存なキーワードを利用していたのに対して，本研究では話題独立な談話標識に着目する．ここで，談話標識を講演や口頭発表の各セクションの冒頭に頻出する表現と定義し，人手によるタグ情報を前提としないで，セクションの境界を抽出する方法を提案する．

2. 講演音声の自動インデキシング

講演には，その分野や話題，長さ，スタイルなどによって様々な種類がある．使用する語彙や講演の流れはその種類によって異なるため，すべての種類の講演に対して一律にインデキシングを行うことは難しい．そこで本研究では，講演の流れが比較的明確で共通性がある学会での研究発表を対象とする．種々の学会講演が開放的融合研究プロジェクトで,「日本語話し言葉コーパス（CSJ）」[11)] として大規模に収集されており，これを利用する．

いくつかの学会での研究発表（主に工学系）の書き起こしを分析したところ，各講演の構成には一定のパターンが存在することが分かった．そのモデルを**図 1** に示す．多くの場合，導入・背景，問題とアプローチ，手法の概要とその説明，実験とその評価，まとめの 5 つの部分に大きく区分化でき，またこの順で述べられている．これらの各部分は，予稿において 1 つの（サブ）セクションから構成される場合もあれば複数の（サブ）セクションからなる場合もある．発表にスライドを使用する場合は複数のスライドがこのような（サブ）セクションを構成する．以下では，この比較的明確なまとまりの単位を単にセクションと記す．

本研究では講演音声をセクションごとに分割すること，すなわちセクション境界を検出することを目標とする．このインデックスはスキップや検索の際に有用である．本論文では行っていないが，スライドと連携すればマルチメディアアーカイブも実現できる[12)]．

このような境界を検出するために，ポーズやピッチなどの韻律情報も利用できる可能性がある[13)~15)]が，話の途中でも頻繁に長いポーズが挿入されたりするので，予備的な分析の結果，韻律情報のみでは十分な精度が得られないと判断した．

セクションの先頭の 1 文は，そのセクションで述べようとしている内容を短く端的に表している．たとえば,「本報告のアプローチについて説明します」,「次にその実験結果を示します」などといったものである．実際に，学会講演の書き起こしからセクションの先頭の 1 文を取り出してみたところ，この部分に頻出する特徴的な単語が存在することが分かった．たとえば「実験」や「説明」,「結果」,「背景」,「今回」,「最後」などである．本研究では，このようなセクションや話題の転換点を示すような単語を談話標識とよぶ．この談話標識を検出することで，セクションの先頭に対してインデキシングができると期待される．

本研究では，談話標識を講演の書き起こしテキストコーパスを利用して自動的に抽出する．膨大なコーパスに人手で談話標識のタグを付与することは大変な手間を要するので，そのようなタグを前提としない教師なし学習を実現する．また，学習テキストがセクションに分割されていることも仮定しない．抽出された談話標識は，講演音声の認識結果に対するインデックス付与に用いる．

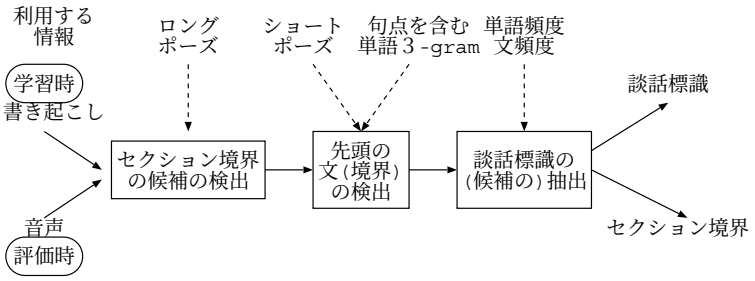

図 2 インデキシングの処理（学習時と評価時）の概要
Fig. 2 Overview of indexing processes (training & testing).

3. インデキシングのための談話標識の抽出

3.1 処理の概要

図 2 に全体の処理の流れと談話標識の抽出に使用する情報を示す．

まず，セクションの境界の候補を検出し，その最初の 1 文を抽出する．そして，その中から単語頻度と文頻度を計算し，それに基づいて談話標識を決定する．この過程では，ポーズ情報，N-gram 言語モデル，単語の出現頻度などの種々の情報源を利用する．以下に各処理について説明する．

3.2 ポーズ情報を用いたセクション境界候補の検出

話題の転換点であるセクション境界に挿入されるポーズ（ロングポーズ）は，セクション内の文の間に挿入されるポーズ（ショートポーズ）よりも，長いことが期待される．実際に文献 13) では，長いポーズがパラグラフ境界の主要な特徴として導出されている．そこで，ある閾値よりも長いポーズが挿入されている部分をセクションの境界の候補として抽出する．

後の処理で絞り込みを行うので，この段階ではできるだけ正しい境界がもれないように閾値を設定することが望ましい．すなわち，多少適合率が低くても再現率が十分に高いことが必要である．ただし，最適な値は話者によって異なる．発話速度が遅い人であれば挿入するポーズが全般に長く，速い人であれば短くなる．したがって，講演者の発話速度に応じた閾値を用いる．

ここでは，書き起こしテキストにあるポーズ情報をもとに，各講演ごとにポーズの平均の長さを求めて，これを閾値とした．ただし当該学習コーパス（日本語話し言葉コーパス）では，原則として 200 ms 以上の無音区間をポーズと判定している．

3.3 単語 3-gram モデルを用いた文境界の検出

インデキシングのために検出すべき部分は各セクションの先頭の 1 文とし，セクション境界候補を検出した後に，セクションの先頭の 1 文を取り出す．しかし，当該コーパスは句点などによって文単位に区切られておらず，また同コーパスによって構築された言語モデルを用いた音声認識結果にも句点は挿入されないので，これらを文単位に区切る必要がある．

本研究では，文境界の候補を検出するために，句点が含まれる Web 講演録から学習された単語 3-gram モデルを利用する．学習に用いた講演数は 81，テキストサイズは 1,692,802 語，異なり語数は 37,462 語である．これからカットオフを 1 にして単語 3-gram モデルを作成した．

文と文の間にはショートポーズが挿入されると仮定し，ポーズ部分での句点の有無による言語モデル尤度の差異に基づいて判定する[16)]．すなわち，ポーズの前の 2 単語 w_1，w_2 と，後の 2 単語 w_3，w_4 を取り出して，その 4 単語をそのまま並べた単語列 “w_1，w_2，w_3，w_4” の尤度 $P(w_1, w_2, w_3, w_4)$ と，句点を挿入した単語列 “w_1，w_2，句点，w_3，w_4” の尤度 $P(w_1, w_2$，句点，$w_3, w_4)$ を計算しそれらを比較する．後者の尤度の方が大きい場合は，その部分に句点を挿入し文境界と見なす．ただし，本実験では尤度のかわりに単語パープレキシティ $-\frac{1}{n}\log P(w_1 \cdots w_n)$ を用いた．

ここで，単純に尤度を比較するのではなく判定にマージンをとることとした．具体的には，句点を挿入しなかった場合のパープレキシティが句点を挿入した場合の 3 倍以内の値であれば，句点を挿入しないこととした．また，パープレキシティは未知語に対して大きい値を示すが，文末に現れるような単語が未知語であるとは考えにくいので，パープレキシティが 1000 以上の場合もその部分には句点を挿入しないこととした．本研究では，誤って短く区切られるよりは，ある程度以上長い方がインデックスの検出に有益であるのでこのようにした．

この手法を，AS00MAR011, AS00MAR015, AS00MAR020，AS00MAR026 の 4 つの講演を用いて評価した．この 4 講演の書き起こしから人手で文境界を判

表 1 単語 3-gram モデルを用いた句点挿入の結果
Table 1 Result of period insertion using word 3-gram language model.

講演	再現率	適合率
AS00MAR011	61/61(100%)	61/78(78.2%)
AS00MAR015	31/32(96.9%)	31/52(59.6%)
AS00MAR020	69/69(100%)	69/81(85.2%)
AS00MAR026	48/51(94.1%)	48/67(71.6%)
total	209/213(98.1%)	209/278(75.2%)

断し，正解を設定した．再現率と適合率の結果を**表 1** に示す．再現率はかなり高く，句点を挿入すべき部分には 100%に近い精度で正しく挿入できた．適合率は 60%程度から 85%程度とばらつきがあり平均すると 75%程度である．

3.4 単語頻度と文頻度を用いた談話標識の選択

学習セットから前 2 節で述べた手法により，各講演をセクション単位に区切り，セクションの先頭の 1 文を取り出した．次に，これらの文から談話標識となる単語集合を選択する．

談話標識として用いる単語は多くの講演に共通して出現し，話題独立であることが望ましい．本研究では，活用変化がない，話し言葉でも書き言葉とあまり変わらない，形態素解析結果が比較的信頼できるなどの理由から，名詞に着目した．ただし，ここでの名詞には「説明（する）」「実験（する）」などのサ変動詞の語幹も含まれている．また，名詞の中でも固有名詞や数詞は話題独立な単語でないと考えられるので，これらは除外した．

次に，単語頻度 tf と文頻度 sf を定義・計算する．名詞 w_i の単語頻度 tf_i は，セクションの先頭の 1 文として抽出された文の集合で名詞 w_i が出現する回数とする．文頻度 sf_i は学習セットの全講演のすべての文で名詞 w_i が出現する文の数とする．ある名詞 w_i について，tf_i の値が大きいとセクションの先頭の文によく出現していることを示し，sf_i の値が大きいと多くの文にまんべんなく出現していることを示す．したがって tf_i の値は大きく，sf_i の値が小さいものを談話標識として選択する．

情報検索で広く利用されている tf・idf 値[17] を参考にして，tf と sf を統合した評価尺度を式 (1) のように定義した．ここで N は全講演における文の総数である．

$$S_{w_i} = tf_i^a * \log\left(\frac{b * N}{sf_i}\right) \quad (1)$$

ベースラインの重みは a，b どちらも 1 とした．抽出された談話標識の例を**表 2** に示す．

表 2 抽出された談話標識の例
Table 2 Example of extracted discourse markers.

今日	研究	実験	目的	結果	我々
説明	評価	発表	今回	最後	まとめ

3.5 インデキシングの手法

新たな講演音声に対して，談話標識を利用して自動インデキシングを行う手法について述べる．インデキシングは音声認識した結果を用いて行う．

まず，講演音声のセクション境界の候補を検出する．音声認識に際しては，ポーズ長に基づいて音声を分割するので，この際に適切なポーズ長の閾値を用いればセクション境界の候補で区切ることができる．談話標識の学習時には書き起こしに記録されているポーズ長の平均値を閾値としたが，講演音声に挿入されているポーズ長の平均値を自動的に求めることは容易でないので，固定のポーズ長を用いることとした．後で絞り込みを行うことを考えると，この閾値をあまり厳しくする必要はないので，学習セットにおける平均のポーズ長で最低の値であった 500 ms を閾値とした．

次に文の終端の検出を行う．音声認識の言語モデルに句読点が含まれない場合は，3.3 節と同じ手法を適用して句点を自動挿入する．句読点がショートポーズに対応づけられて言語モデルに含まれている場合は，認識時に句点が挿入され，文の境界が検出される．

最後に 3.4 節で抽出された談話標識を用いてセクション境界の判定を行う．談話標識が音声認識結果に出現した場合に，1 文中に出現する各談話標識に対する単語頻度と文頻度に基づく評価値（式 (1)）の合計 $\sum_i S_{w_i}$ が一定の閾値（θ）を上回った場合にインデキシングを行う．

4. 評価実験

4.1 学習・評価データ

本研究では日本語話し言葉コーパス（CSJ）の一部を使用した．具体的には，**表 3** にある 65 件の口頭発表の忠実な書き起こしの ChaSen ver2.02 による形態素解析結果を談話標識の学習に使用した．このおよそ半分は日本音響学会の研究発表会から収集され，残りは他の学会から収集されたものである．ただし音声認識のための音響モデルと言語モデルの学習には学会以外の発表も含めてはるかに大量のデータを使用している[18]．

これとは別に**表 4** に示す 17 件の発表を評価セットとして用意した．これらは学習セットには含まれていない．講演の長さは 11 分から 15 分である．セクショ

表 3 学習データとして使用する講演の内訳
Table 3 Training set.

講演の種類	講演数
日本音響学会 春季&秋季研究発表会 (AS)	38
言語処理学会 年次大会 (NL)	4
国語学会 (JL)	5
音声学会 全国大会 (PS)	9
国立国語研究所内の種々の研究会 (KK)	6
融合研究の会合 (YG)	3
合計	65

表 4 評価データの一覧
Table 4 Test set.

	講演の長さ		正解
	時間	単語総数	インデックス数
AS99SEP008	12 分	1,943 語	12
AS99SEP009	11 分	1,680 語	11
AS99SEP011	13 分	2,541 語	8
AS99SEP014	11 分	2,067 語	9
AS99SEP015	12 分	1,871 語	7
AS99SEP018	13 分	1,628 語	8
AS99SEP019	14 分	1,926 語	12
AS99SEP037	12 分	2,158 語	14
AS99SEP039	13 分	2,138 語	11
AS99SEP027	12 分	1,460 語	11
AS99SEP097	12 分	2,508 語	16
PS99SEP025	27 分	5,372 語	18
AS00MAR011	12 分	1,903 語	13
AS00MAR026	14 分	2,487 語	10
NL00MAR007	15 分	2,644 語	8
NL00MAR033	13 分	2,974 語	10
NL00MAR081	15 分	3,059 語	12

ン境界の正解は人手により付与した．境界の数は 7 個から 18 個である．

評価値として，正しい境界の再現率（recall）と検出された境界の適合率（precision）の組合せである F-measure を用いる．F-measure は式 (2) で定義される．

$$F-measure(\alpha)=\frac{(1+\frac{1}{\alpha})*recall*precision}{\frac{1}{\alpha}*recall+precision} \quad (2)$$

正しい境界が検出されていないと検索ができないが，誤って検出された部分は検索時にスキップすればよいことから，再現率を重視する必要がある．ここでは，$\alpha = 10$ として再現率に 10 倍の重みをつけた場合を用いる．

4.2 談話標識の効果

学習セットの全講演から，パラグラフの先頭の 1 文として抽出された候補の形態素解析結果から，固有名詞，数詞を除いて名詞を抽出した結果，約 3,000 個の名詞が得られた．このように多数になった理由としては，セクション単位に区切る際に，適合率よりも再現率を重視し，誤った検出をかなり許したためである．これから式 (1) に基づいて，談話標識を 75 個抽出した．

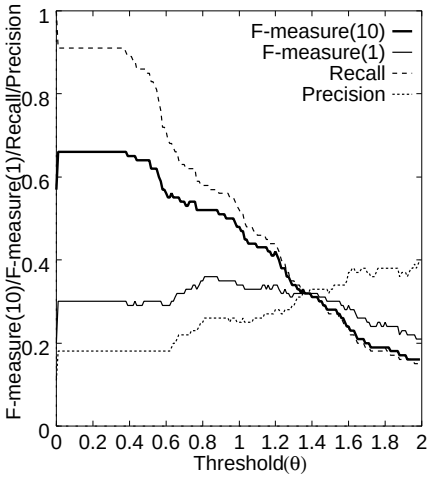

図 3 談話標識を用いたインデキシング結果
Fig. 3 Indexing performance using discourse markers.

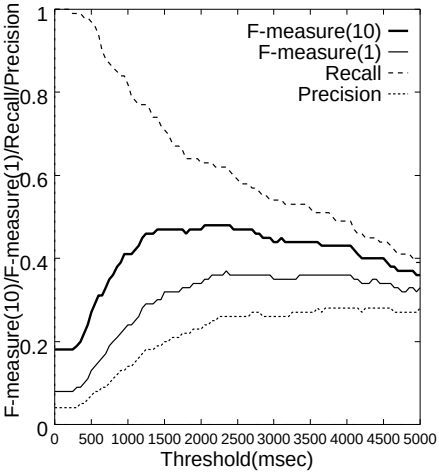

図 4 ポーズ長のみによるインデキシング結果
Fig. 4 Indexing performance using pause length only.

まず，談話標識を用いてインデキシングを行うことの基本的な有効性を確認した．75 個の談話標識を使用して，インデキシングの閾値 θ（3.5 節参照）を変化させながら再現率，適合率，F-measure(1)，F-measure(10) をプロットしたものを**図 3** に示す．比較として，単純にポーズ長のみを用い，ポーズ長が閾値より長い部分に対してインデキシングを行う手法も実行した．これは現在のテープレコーダでインデックスを付与する方法に相当する．ポーズ長の閾値を変化させた場合の結果を**図 4** に示す．

これらを比較することにより，本論文で提案した談話標識を用いる手法（図 3）の方が全般に高い性能を

得ていることが分かる．本インデキシングでは適合率よりも再現率が重要であるが，特に再現率（recall）の高い部分において談話標識を用いたインデキシングの方が適合率（precision）がかなり高くなっている．実際に再現率に重みをおいた F-measure(10) では両手法の差は顕著である．なおここでは，再現率の重み（α）を 10 とした F-measure(10) の場合を示しているが，$\alpha \geq 1$ においていずれの場合も提案手法の方が高い値を得ており，α を大きくするほどその差が大きくなる．

適合率は 20%程度と高くはないが，実際にこのようなインデキシングを人手（専門業者）で行うと 1 件の講演につき 5〜6 時間を要するとのこと（見積り）である．提案手法により，インデックスの候補を求めて，認識結果や音声区間と対応づけておけば，これらの候補から選択することによりインデキシング作業のコストが大幅に軽減できると期待される．

これより，統計的に抽出された談話標識を用いる提案手法の有効性が示された．

4.3 談話標識の評価尺度に関する検討

次に，式 (1) の重み a と b の値を変化させることを検討した．a の値として $\frac{1}{4}$，$\frac{1}{2}$，1，2，4 の 5 通り，b の値として $\frac{1}{6}$，$\frac{1}{5}$，$\frac{1}{4}$，$\frac{1}{3}$，$\frac{1}{2}$，1，2，3，4，5，6 の 11 通りを用いて比較した．その結果の一部を**図 5**，**図 6** に示す．ここでは F-measure(10) のみ示している．図 5 には a=1 として，b の値を変化させた場合を示しているが，性能に大きな変化は見られなかった．また図 6 では，b=1 として a の値を変化させているが，a の値を 1 より小さくしても大きくしても性能が低下している．これらの結果から，重みは a，b どちらも 1 とした．

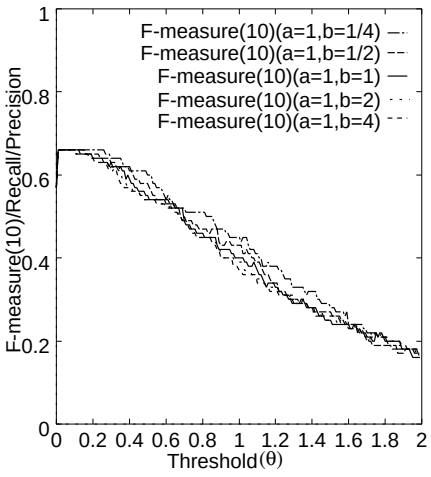

図 5 重み b の値のインデキシングへの影響
Fig. 5 Indexing performance by changing the weight b.

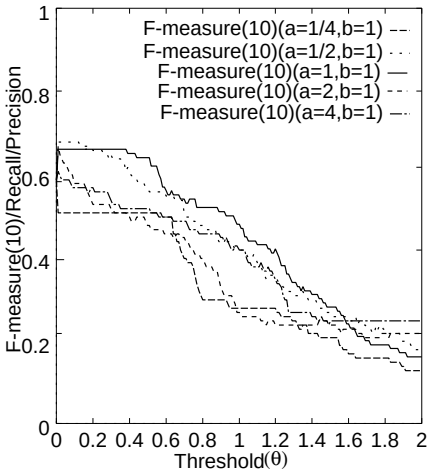

図 6 重み a の値のインデキシングへの影響
Fig. 6 Indexing performance by changing the weight a.

4.4 談話標識の数の影響

次に，談話標識の数による影響を調べた．用いる談話標識の数が 25 個，75 個，125 個の場合について，それぞれインデキシングを行った．これらは式 (1) の評価値 S_{w_i} の大きい順に抽出している．その結果を**図 7** に示す．

談話標識の数が 75 個の場合が最も F-measure(10) の値が高い．談話標識の数が少なすぎると，インデキシングすべき部分を抽出できないことが多くなり，また，多すぎると誤った抽出が多くなる．したがって，談話標識の数は 75 個とした．

4.5 音声認識結果に対するインデキシング

最後に，実際に講演音声を自動認識した結果にインデキシングし，手法の評価を行った．評価データとして表 4 の中から AS99SEP037，AS99SEP039，AS99SEP097，PS99SEP025，AS00MAR011，AS00MAR026，NL00MAR007，NL00MAR033，NL00MAR081 の 9 件の講演を用いた．

音声認識の際の音響モデルとして日本語話し言葉コーパスを用いて学習された 2000 状態 16 混合の triphone モデルを使用し，言語モデルとして同コーパスと Web 講演録を併用して学習した単語 3-gram モデルを使用した．認識エンジンは著者らの研究室で開発された Julius 3.1 である．単語認識精度は 60%から 70%程度である[18]．

図 8 に再現率，適合率，F-measure(10) を示す．比較として，書き起こしに対する結果も同時に示している．書き起こしに比べると，誤認識の影響を受けて再現率が低くなっているが，その低下の程度は認識誤り率に比べると相対的に小さい．また，図 4 のポーズ長

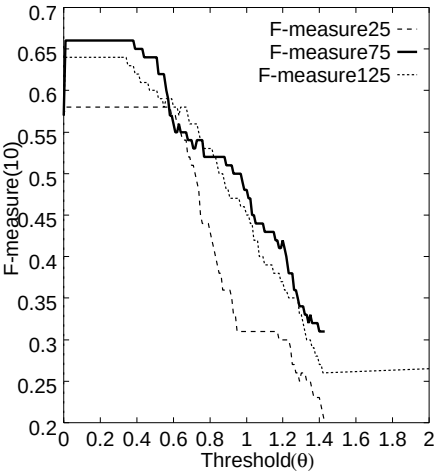

図 7 談話標識の数のインデキシングへの影響
Fig. 7 Indexing performance by changing the number of discourse markers.

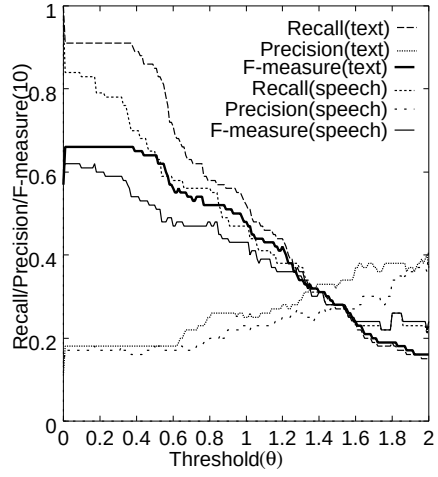

図 8 音声認識結果に対するインデキシング結果
Fig. 8 Indexing performance for speech recognition results.

のみを用いた方法よりも，依然高い性能を維持している．この結果は，談話標識を用いた統計的な評価値によるセクション境界の検出が頑健であることを示している．

5. 結　　論

本論文では，講演音声に対する自動インデキシングの方法を提案した．これは，談話標識として定義した，セクションの冒頭に頻出する特徴的な表現に注目している．談話標識は，人手によるタグをまったく必要としないで，完全に教師なしで統計的に学習する．本手法により再現率 85%，適合率 20%の性能を実現した．これは，長い音声に対する効率的な検索のためのインデックスの（半）自動的な付与として有望であると考えられる．本手法は 30%〜40%の音声認識誤りに対しても頑健であることが示された．

今後は，会議や討論といった他の種類の音声メディアへの適用を検討していく予定である．

謝辞 本研究は，開放的融合研究『話し言葉工学』プロジェクトの一環として行われた．東京工業大学の古井貞熙教授をはじめとして，ご協力をいただいた関係各位に感謝いたします．また，本研究に際してご指導，ご討論をいただいた，京都大学の奥乃博教授に深く感謝いたします．

参 考 文 献

1) 堀　智織，古井貞熙：話題語と言語モデルを用いた音声自動要約法の検討，情報処理学会研究報告，99-SLP-29-18 (1999).
2) 堀　智織，古井貞熙：かかり受け SCFG に基づく音声自動要約法の改善，情報処理学会研究報告，2000-SLP-34-42 (1999).
3) Imai, T., Schwartz, R., Kubala, F. and Nguyen, L.: Improved topic discrimination of broadcast news using a model of multiple simultaneous topics, *Proc. IEEE-ICASSP*, pp.727–730 (1997).
4) 横井謙太朗，河原達也，堂下修司：キーワードスポッティングに基づくニュース音声の話題同定，情報処理学会研究報告，95-SLP-6-3 (1995).
5) 鷹尾誠一，緒方　淳，有木康雄：ニュース音声に対するトピックセグメンテーションと分類，情報処理学会研究報告，98-SLP-24-8 (1998).
6) 大附克年，松岡達雄，松永昭一，古井貞熙：ニュース音声を対象とした大語彙連続音声認識と話題抽出，電子情報通信学会技術研究報告，SP97-27 (1997).
7) Jones, G.J.F., Foote, J.T., Jones, K.S. and Young, S.J.: Video mail retrieval: The effect of word spotting accurasy on precision, *Proc. IEEE-ICASSP*, pp.309–312 (1995).
8) McDonough, J., Ng, K., Jeanrenaud, P., Gish, H. and Rohlicek, J.R.: Approaches to topic identification on the SWITCHBOARD corpus, *Proc. IEEE-ICASSP*, Vol.1, pp.385–388 (1994).
9) Waibel, A., Bett, M., Metze, F., Ries, K., Schaaf, T., Schultz, T., Soltau, H., Yu, H. and Zechner, K.: Advances in automatic meeting record creation and access, *Proc. IEEE-ICASSP*, Vol.1, pp.597–600 (2001).
10) 秋田祐哉，河原達也：会議音声の自動アーカイブ化システム，情報処理学会研究報告，2000-SLP-34-11 (2000).
11) 小磯花絵，前川喜久雄：『日本語話し言葉コー

パス』の概要と書き起こし基準について，情報処理学会研究報告，2001-SLP-36-1 (2001).
12) 河原達也，石塚健太郎，堂下修司：発話検証に基づく音声操作プロジェクタとそれによる講演の自動ハイパーテキスト化，情報処理学会論文誌，Vol.40, No.4, pp.1491–1498 (1999).
13) Haase, M., Kriechbaum, W., Mohler, G. and Stenzel, G.: Deriving document structure from prosodic cues, *Proc. EUROSPEECH*, pp.2157–2160 (2001).
14) 野村和弘，河原達也，堂下修司：F0 パターンに基づく講義音声の文単位へのセグメンテーション，電子情報通信学会技術研究報告，SP99-13 (1999).
15) 笠原力弥，山下洋一：講演音声における重要文と韻律的特徴の関係，情報処理学会研究報告，2001-SLP-35-5 (2001).
16) 西村雅史，伊東伸泰，山崎一孝：単語を認識単位とした日本語の大語彙連続音声認識，情報処理学会論文誌，Vol.40, No.4, pp.1395–1403 (1999).
17) 長尾 真（編）：自然言語処理，岩波講座ソフトウェア科学 (1996).
18) 加藤一臣，南條浩輝，河原達也：講演音声認識のための音響・言語モデルの検討，情報処理学会研究報告，2000-SLP-34-23 (2000).

(平成 13 年 11 月 19 日受付)
(平成 14 年 4 月 16 日採録)

長谷川将宏

2001 年京都大学工学部情報学科卒業．現在，同大学院情報学研究科知能情報学専攻修士課程在籍．音声情報処理の研究に従事．

秋田 祐哉（学生会員）

2000 年京都大学工学部情報学科卒業．現在，同大学院情報学研究科修士課程在学中．音声認識・理解の研究に従事．

河原 達也（正会員）

1987 年京都大学工学部情報工学科卒業．1989 年同大学院修士課程修了．1990 年同博士後期課程退学．同年京都大学工学部助手．1995 年同助教授．1998 年同大学情報学研究科助教授．現在に至る．この間，1995 年から 96 年まで米国ベル研究所客員研究員．1998 年から ATR 客員研究員．1999 年から国立国語研究所非常勤研究員．2001 年から科学技術振興事業団さきがけ研究 21 研究者．音声認識・理解の研究に従事．京都大学博士（工学）．1997 年度日本音響学会粟屋賞受賞．2000 年度情報処理学会坂井記念特別賞受賞．情報処理学会連続音声認識コンソーシアム代表．電子情報通信学会，日本音響学会，人工知能学会，言語処理学会，IEEE 各会員．