論文 / 著書情報
Article / Book Information

論題(和文)	話し言葉の音声認識と自動要約
Title	
著者(和文)	古井 貞熙
Author	SADAOKI FURUI
出典(和文)	第3回話し言葉の科学と工学ワークショップ講演予稿集, Vol. , No. , pp. 1-6
Journal/Book name	, Vol. , No. , pp. 1-6
発行日 / Issue date	2004, 2

話し言葉の音声認識と自動要約

古井　貞熙

東京工業大学大学院情報理工学研究科計算工学専攻

〒152-8552 東京都目黒区大岡山 2-12-1

E-mail: furui@cs.titech.ac.jp

あらまし　書き言葉を読み上げたような音声だけでなく、自然な話し言葉を音声認識し、発声者の意図を自動的に抽出するための、統計的枠組みの構築が求められている。そのため、話し言葉に対する音声認識精度を向上させるための、話し言葉コーパス（CSJ）を用いた音素モデルと言語モデルの構築、さらにそれらの話者および話題への適応化法に関する研究を行ってきた。音声を文字に変換することを目的としてきた従来の音声認識の枠組みを超えて、音声認識結果から発声者の意図を抽出する枠組みの研究を行った。話題を担う重要文や重要語を自動的に抽出して、発声者の意図を伝える要約文を自動的に生成する統計的方法を提案した。本方法を講演音声などに適用し、人が作成した要約文との比較により、その有効性を示した。話し言葉の音声変動をモデル化する方法として、ベイジアンネットの利用について検討し、雑音中での音声認識精度向上のため、これまでのケプストラムでなく、スペクトル領域での HMM を用いる方法について検討した。最後に、話し言葉コーパスを含む知識の体系化と利用などの、今後の音声認識研究の課題について述べる。

キーワード　話し言葉、コーパス、音声認識、音声要約、モデル適応化、ベイジアンネット、知識の体系化と利用

Spontaneous Speech Recognition and Summarization

Sadaoki FURUI

Tokyo Institute of Technology, Department of Computer Science

2-12-1 Ookayama, Meguro, Tokyo, 152-8552 Japan

E-mail: furui@cs.titech.ac.jp

Abstract　Although read speech and similar types of speech can be recognized with high accuracy, it is crucial to build a statistical framework for recognizing spontaneous speech and automatically extracting the intensions of the speakers for broadening the application of speech recognition.　From this point of view, we have been investigating methods for constructing acoustic and language models using the spontaneous speech corpus (CSJ) and adapting the models to each speaker.　We have also investigated the framework of extracting speakers' intensions using speech recognition results, instead of simply converting speech into text.　Important sentences and words are automatically extracted and concatenated to produce a summary which conveys intensions of the speaker based on a statistical method. The proposed method has been applied to various speech including lectures and evaluated in comparison with manually created summaries.　Effectiveness of the method has been confirmed.　For modeling acoustic variations of spontaneous speech, the Bayesian network method has been investigated.　In order to increase the robustness for noisy speech, a new method using HMMs in log-spectral domain instead of cepstral domain has also been investigated.　The paper finally presents future research topics of speech recognition research, such as systematization and application of knowledge resources including spontaneous speech corpora.

Keywords　Spontaneous speech, corpus, speech recognition, speech summarization, model adaptation, Bayesian net, systematization and application of knowledge resources

1. はじめに

書き言葉に近い音声の認識で比較的高い精度を得るのは難しくないが、話し言葉に関しては、まだ種々の実用に耐えるだけの十分な認識精度が得られていない。

コミュニケーションにおける音声の主要な役割は、自然に話しながら意図を伝達することである。自然な話し言葉には、「あのー」「えーと」などの不要語、言い直し、言いよどみ、言葉の省略、繰り返しなどが必ず含まれる。これまでの音声認識では、発声されたすべての言葉を正しく文字に変換することを目標にして来たが、話し言葉を対象とする場合は、不要な言葉や言い直しなどはむしろ無視し、発声者が伝えようとする内容を表す文を生成することが必要である。これまでの「正しい単語列」を最終目標とする理論ではなく、「正しい意味内容」を最終目標とする音声理解理論を構築することが求められている。

2. 話し言葉を音声認識するための基本技術の構築

1999 年から 5 年計画で開始された、科学技術振興調整費開放的融合研究推進制度による「話し言葉工学」プロジェクト[1]で作成された、学会講演音声のコーパス（CSJ: Corpus of Spoken Japanese）用いて、話し言葉を音声認識するための音素モデルおよび言語モデルの構築と、それを用いた評価実験を行ってきた。その結果、次のような成果が得られた[2]。

(1) 読み上げ音声から作成した音素モデルと、Web 上の講演録（講演音声を読みやすいように書き言葉に近い形で整形されているもの）から作成した言語モデルを用いた場合に 55%程度あった誤認識を、CSJ から作成した音素および言語モデルに置き換えることにより、30%程度にまで低下させることができた[3]。
(2) 発声速度、言い直し頻度、および未知語率が、話者による認識率の違いの主要な要因となっている[4]。すなわち、概して、早口の音声、言い直しの多い音声、および音声認識用辞書に登録されていない語彙を多く発声している音声の誤認識が大きい。
(3) 単語の連鎖確率に基づく言語モデルを用いているため、1%の言い直し頻度と未知語率の増加は、それぞれ、3.3%と 2.3%の単語正解精度の低下に拡大される[4]。
(4) フィラー（「えー」「あー」のような言葉）の回数も個人差が大きいが、フィラーを含む言語モデルを用いれば、フィラー自体の認識は特に難しくない。
(5) 話し言葉では、文の区切りが必ずしも明確でないため、書き言葉を読み上げた音声を認識する場合と異なり、文単位に区切らずに連続して音声認識（デコーディング）する方がよい[5]。認識結果から、文の区切りを推定することができる。
(6) 発声された音声の話題（分野）が学習に用いられたコーパスでカバーされていないときには、その分野の教科書などを用いて適応化するのが有効である[3]。

3. 音素モデルと言語モデルの話者適応化

音素モデルを話者に自動的に適応化することにより、音声認識性能が向上する。その際、音声認識結果には誤認識が含まれることがあるため、認識結果をそのまま用いて音素モデルを適応化すると、不適切なモデルが作られる可能性がある。そこで、音素を複数のクラスに分類し、各音素クラス内での適応化を、尤度最大化基準を満たす線形変換に拘束し（MLLR 適応）、それを一つの講演全体に対して繰り返すことにより、バッチ型教師なし適応を行う方法を提案した。これにより、誤認識があっても、認識性能を向上させることができることが確認された[3]。

言語モデルに関しても、話者や話している内容が変わるごとに、それに応じて適応化できることが望ましいが、そのために大量の（書き起こし）テキストを、個々に集めるのは現実的ではない。このため、スパースネスと音声認識誤りの問題に対処する目的から、単語を自動的にクラス化し、クラス言語モデルに基づいて、共通の基本的な言語モデルを、バッチ型教師なし適応する方法を提案した[6]。具体的には、一つの講演全体の認識仮説を用いて、クラス言語モデルを学習し、これを共通言語モデルと線形補間することによって、適応化言語モデルを作成して、講演音声を再認識する。これにより、音素モデルの適応化法と同様に、誤認識があっても、認識精度が向上することが確認された。さらに、音素モデルと言語モデルの適応化を組み合わせることにより、両者の効果が加算され、一層の性能向上が達成できることがわかった。

次に、言語モデルの適応化に関して、話題ごとの異なる言語モデルを作成して用いる方法を検討した[7]。2690 講演（約 750 万語、語彙数 30,678）の書き起こし全体を用いて、共通言語モデルを作成すると同時に、それらの講演を、単語の共起尺度に基づいてクラスタ化する。各クラスタごとに、単語言語モデルと、単語クラス言語モデルを作成する。音声認識対象の講演に対して、最小パープレキシティ原理に基づいて、一つあるいは複数のクラスタ依存言語モデルを選択し、共通言語モデルと線形補間する。補間係数は EM アルゴリズムによって自動的に決定する。こうして作成した言語モデルを用いて、講演音声を再認識する。認識実験の結果、全講演を 8 つにクラスタ化し、514 の単語クラスを用いて、補間係数を、認識対象の各文ごとではなく各講演内で共通に設定する場合に、最も高い適応効果が得られることが確認された。

4. 音声理解・要約の基本技術の構築

話し言葉音声を認識・理解するためには、話し言葉特有の性質に対処できる新しい方法を作り出すことが必要という視点から、すべての言葉を文字にしようとするこれまでのアプローチを超えた、新しいアプローチについて研究を進めた。すなわち音声理解の一つの形態として、音声の自動要約法について研究を進めた。発声者の意図が過不足なく表現された要約文が生成できれば、音声の意味表現ができた、すなわち音声が理解できたと考えられるからである。そこでまず、音声認識結果として得られる単語列から、話

題を担っている重要語（話題語）を自動的に抽出する研究を行った。重要語を選ぶ種々の尺度について検討した結果、情報検索の分野で用いられている tf-idf 尺度を変形した情報量尺度が最も優れていることがわかった[8]。次に、重要語を中心とする単語セットを、できるだけ文法的制約に従い、かつ意味が変わらないように最適に選び、それらを接続して要約文を作成する方法について研究を進めた。図 1 に、音声自動要約システムの構成を示す[9]。

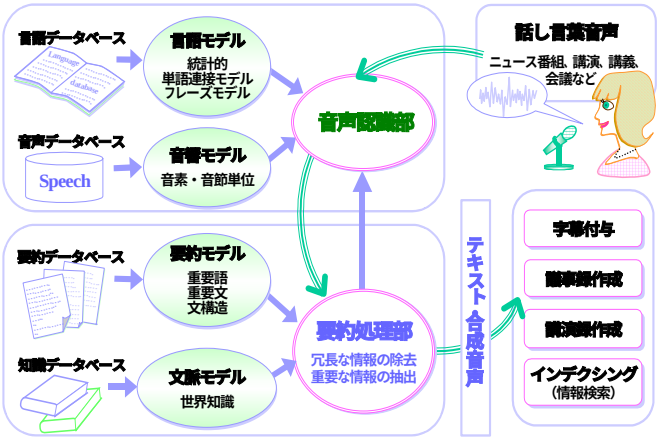

図 1　音声自動要約システムの構成

無数にある単語の組み合わせの中から、上記の目的を達成する単語セットを効率的に抽出するため、各単語の重要度スコア、要約文の統計的言語スコア、音声認識結果の信頼性、さらに単語間の係り受け関係をもとに、これらのスコアの総和が最大となる単語セットを、動的計画法を用いて自動的に選択する方法を提案した[10]。重要度スコアには、上記と同じ情報量尺度を用いている。統計的言語スコアには、音声よりも比較的簡潔な文体となっていると考えられる新聞の記事から計算した単語トライグラムを用いている。音声認識結果の信頼性は、認識結果として得られる単語グラフから計算される各単語の事後確率を用いている。単語の係り受け関係は、文脈自由文法に基づいて表現している。これら 4 種類のスコアは、それぞれ、話題を担う単語の重視、読みやすい文の生成、音声認識誤りが要約文に含まれることの防止、要約によって文の意味が変わってしまうことの防止に寄与する。

さらに、特に圧縮率が高い場合に効果的な方法として、上記の種々のスコアをもとに、多数の文の中からまず、意味的に重要で、音声認識結果としての信頼性の高い文を抽出し、抽出された各重要文に対して、上に示した方法でさらに圧縮（要約）する方法を提案した[11]。これらの方法に関して、放送ニュース音声や、学会講演の音声認識結果を用いて、文字数にして 10～70%に要約する条件で実験を行い、人が要約した結果に近い結果が得られることが確認された。

提案した要約法は、音声・言語コーパスさえあれば、言語に依存せずに適用することができるため、英語の放送ニュース音声にも適用して評価実験を行った。その結果、日本語に対する場合と同様に、英語の音声に関しても効果的であることが確認できた[12]。

上で述べた重要文抽出の方法は、講演全体の構成を考慮した方法でないため、この欠点を補うための新しい方法を検討した[13]。具体的には、特異値分解（Singular Value Decomposition; SVD）を用いる方法として、潜在的意味分析に基づく手法と、次元圧縮に基づく手法を検討した。検討に先立って、音声認識仮説から文の単位を自動的に切り出すため（文境界判定）、全てのポーズ区間の前後 2 単語ずつを含む単語列について、中心に文境界を仮定した場合としない場合の言語尤度の比を計算し、その値に閾値を設ける方法を適用した。各講演の音声認識仮説について、単語（名詞）を行に、文を列にもつ行列を、特異値分析することにより、単語と文の相互関連を捉え、潜在的な単語と文のクラスタリングができる。潜在的意味分析に基づく方法では、各特異値に対する最適な文を選択する。次元圧縮に基づく方法では、利用する特異値の数を減らすことによって次元圧縮され、特異値で重み付けられたベクトル空間に各文を射影し、各文にスコアを与えて重要文を抽出する。CSJ の 20 名の講演音声を用いて、50%に要約する実験を行った結果、次元圧縮に基づく方法が有効であることが、確認された。

今後、ニュースの自動字幕作成への適用の他、会議の議事録や講演録作成など、種々の応用分野への展開が期待されている。ニュース音声、講演、講義などを、音声認識技術を用いて自動的に要約してデータベース化することができれば、情報検索に限らず種々の用途に極めて有用であろう。

5. 音声から音声への要約技術

音声の要約結果をテキストで提示する場合は、音声認識誤りが避けられないため、読みづらくなるだけでなく、誤った情報をユーザに伝える可能性がある。また、韻律によって伝えられる話者の意図や感情情報を伝えることが難しい問題がある。元の音声をそのまま用いて提示すれば、それらの問題を回避することができる。

そこで、上で述べた音声要約手法を用いて、要約結果を音声で提示する「速聞きシステム」について研究を行った[14]。図 2 に、システムの構成を示す。要約音声は、テキストで出力された音声要約結果に対応する音声を、元の音声から切り出し接続して作成する。切り出した音声どうしをそのまま接続したのでは、音声の振幅や波形の差から、雑音が生ずる可能性があるので、音声端の振幅を徐々に減衰させ、前後の言語的関係に依存した長さを有するポーズ（無音区間）を挿入してから接続する。

要約音声を作成するのにふさわしい切り出し単位を調べるため、単語単位、文単位、文をさらにフィラーで区切っ

た単位、の３つの単位の選択により作成した要約音声について、「聞きやすさ」と「要約音声としてのふさわしさ」の 2 項目で、聴取実験による評価を行った。講演音声を用いた実験の結果、要約率 50%では、文単位選択による要約が最も有効であることが確認された。

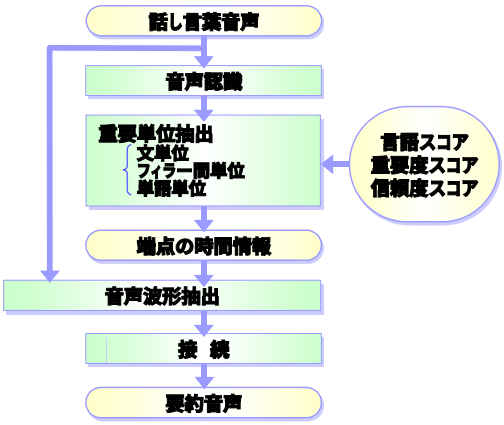

図 2　速聞きシステムの構成

6. ベイジアンネットによる音声のモデル化

2 章で述べたように、話し言葉音声の認識誤りを生ずる主たる要因の一つに、発声速度の変動がある。発声速度は、話者や文単位で変動するだけでなく、文の中、さらに単語の中でも音素ごとに変動する。しかも、認識に直接必要な音響特徴とは異なる時間尺度で変動する。音声認識に広く用いられている HMM では、状態を複数の要因に分解してモデル化できないなど、構造上の制限が存在する。このため、発声速度の時間的な変動に適応する方法として、発声速度変動を隠れ変数を用いてモデル化するベイジアンネットを用いた音響モデルについて研究を進めた[15]。動的ベイジアンネット（DBN; Dynamic Bayesian Network）は、HMM を含む様々な確率モデルを記述できる柔軟性があり、一連の学習・推論アルゴリズムが開発されている[16]。

提案したモデルにおける入力の 1 フレームに対応するベイジアンネットを図 3 に示す。ベイジアンネットでは、ノードは確率変数に対応し、ノード間の有向枝は、確率的な依存関係を示す。このモデルは、従来の HMM のパラメータ（音素と、音素 HMM 内の状態位置）に加えて、発声速度の状態を表す隠れ変数を持ち、その値に応じて音響特徴量の出力確率密度および遷移確率の制御が行われる。モデルの学習・評価（認識）時には、図 3 のネットワークは、与えられた音声の音響特徴量ベクトル列の長さに合わせて展開される。

提案手法では、音響モデルの学習および認識処理において、通常の音響特徴量に加え、発声速度を使用する。発声速度の計測法として、認識仮説を用いたモデルの強制アラインメントによる方法と、物理量である Enrate を用いる方法を検討した。実験の結果、CSJ および英語の会議音声の認識で、従来モデルよりも高い認識率が得られることが確認された。発声速度の計測法としては、英語音声の場合は、どちらの方法でも同様の効果が得られたが、日本語音声の場合は、Enrate よりも認識仮説を用いる方が効果的であった。

ベイジアンネットを用いた認識システムは、音声認識に特化した従来システムと比べると計算量が多い欠点があるが、様々な新しい確率モデルのプロトタイピングに適している。

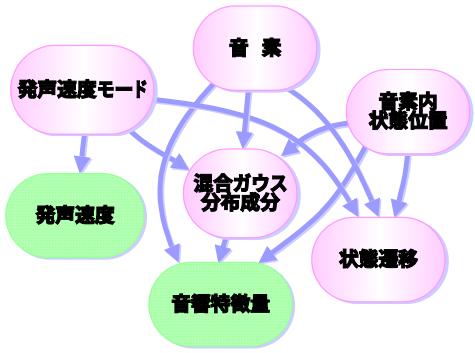

図 3　発声速度変動を隠れ変数を用いてモデル化するベイジアンネット

7. スペクトル領域の HMM による音声認識

現在のほとんどの音声認識システムで、MFCC（Mel frequency cepstral coeffcients）すなわちメル周波数に対する対するスペクトルを逆フーリエ変換したケプストラム、およびその動的特徴で表された空間上の HMM が用いられている。MFCC には種々の優れた性質があるため、長く用いられているが、欠点がないわけではない。このため、近年、対数スペクトル特徴量を直接的に用いる音声認識アルゴリズムが再検討されている。この方法には、人の聴覚器官で行われている分析に近い、ある周波数帯域に偏って音声に重畳している雑音に対して直接的に対処できる、などの特徴があるが、対数スペクトル特徴を直接用いて音声認識を行うと、特殊な雑音環境を除くと、一般的には MFCC を用いた場合よりも、低い認識性能しか得られない。

そこで、この認識性能の劣化の原因がどこにあるかを分析した結果、MFCC を用いる場合に効果的に行われている３つの正規化処理、すなわち、①各フレームの平均対数エネルギーの正規化（零次ケプストラムの除去）、②リフタリングによるスペクトル包絡の強調と平滑化、および③CMS（ケプストラム平均除去）によるスペクトル特性変動の正規化、を省くことが、対数スペクトル領域での性能低下をもたらしていることが明らかになった[17]。

このため、対数スペクトル領域での認識においても、これらの正規化処理を、一旦 MFCC 領域に変換して行うか、

対数スペクトル領域で等価的に行う方法を提案した。後者の場合の手順を、図 4 に示す。連続数字音声および CSJ を用いた評価実験の結果、いずれの方法の場合も、MFCC を用いる場合とほぼ同等の認識性能が得られることが確認された。さらに、スペクトル領域での処理の特徴を生かすため、スペクトル帯域ごとに、HMM から求まる対数尤度の値に重み付けを行う方法を導入した。具体的な重み付けとして、聴感的に重要と考えられるスペクトルピークに重みを付ける方法と、SNR に応じて重みを付ける方法について検討し、特に雑音下での音声認識において、認識性能の向上に有効であることを確認した。

ケプストラム係数には、各係数が直交化されているため、HMM に対角共分散行列を用いても、通常実用的に問題がないという大きな長所がある。スペクトルの隣接する帯域間には明らかに相関があるので、HMM に対角共分散行列を用いるのは理論的に無理があるが、上記の実験では、簡単のために、対角共分散行列を用いている。結果としてケプストラムと同等以上の結果が得られているので、非対角成分まで効果的に用いることができれば、さらに認識精度が向上する可能性がある。

図 4　対数スペクトル領域で正規化を行い、スペクトル特徴量を抽出する手順

8. これからの音声認識・理解

音声認識システムには、極めて多様な応用が期待できるが、任意の話題について、自然な話し言葉に対して、誰の声でも、どんな環境でも正しく音声認識できるためには、解決しなければならない課題が多く残っている。主たる課題は、話し言葉特有の言語的・音響的変動、周囲雑音や電話音声の歪み、音声の個人差などに対するロバストなモデルとアルゴリズムの構築である。その基礎となる話し言葉の大規模コーパスや、実環境での音声を大量に収録したデータベースの構築も極めて重要である。

これまでの音声認識では、主としてスペクトル包絡情報とパワー情報が用いられ、人が音声を認識するときに重要な役割を果たしている基本周波数などの韻律情報は、ほとんど用いられていない。限られたタスクの音声認識では、基本周波数の有効性が示されているが[18]、大語彙の話し言葉音声認識では、まだほとんど成功していない。CSJ のコア部分には、韻律情報が付与されているので、今後これらの解析や、音声認識、要約などへの利用技術の研究が進むと期待される。

テレビのニュース、学会や一般の講演、大学の講義、インタビューなど、毎日大量に生まれているこれらの音声ドキュメントを、録音しておいただけでは利用が難しい。このため、音声認識して文字化し、自動的に内容のインデックスをつけることが期待されている。今後の音声認識の研究は、単に音声を単語列に変換するだけでなく、発声者の意図の理解、話題語あるいはキーフレーズの抽出、内容の自動要約へと展開するであろう。音声の自動要約は、同じ言語の中での翻訳プロセスとしてモデル化することもできる。これらの技術の進展によって、誰が何を言ったかといった情報を含め、いろいろな人が持っている知識を即座に流通させ、活用することができるようになると期待される。さらに、それらの知識を、Web などの多様な知識と関連付け、体系化することができるかどうかが、21 世紀の重要な研究課題となると予想される。そのような体系化された知識が利用できるかどうかが、逆に、音声認識・理解技術の発展のかぎとなるであろう[19][20]。

コンピュータの小型化、高性能化、低価格化により、ユビキタス（遍在）計算とウェアラブル計算環境の時代が来るといわれている。このような環境下では、キーボード入力はなじまないので、音声認識がヒューマンインタフェースの基本的手段の一つとして広く使われるようになると予想される。その際、音声認識の多くは、個人が身につけ、個人に特化したマイクロホンおよびコンピュータで行なわれるようになり、そのシステム構成は、これまでとは大きく異なった形になると予想される。雑音などの環境情報は常時モニタでき、タスクに依存した知識は個人のコンピュータへ即座にダウンロードできるので、タスクへの適応性とロバスト性の向上が期待できる。ただし、それらの実現のためには、解決しなければならない課題も多く、今後の多角的な研究の展開が必要である[21]。

文　献

[1] 古井貞熙, 前川喜久雄, 井佐原均 “科学技術振興調整費開放的融合研究推進制度—大規模コーパスに基づく『話し言葉工学』の構築—“ 日本音響学会誌, vol.56, no.1, pp. 752-755, 2000

[2] 古井貞熙 ”話し言葉の音声認識・理解を目指して”電子情報通信学会技術報告, SP2000-47, 2000

[3] T. Shinozaki and S. Furui “Towards automatic transcription of spontaneous presentations” Proc. Eurospeech 2001, pp. 491-494, 2001

[4] T. Shinozaki and S. Furui “Analysis on individual differences in automatic transcription of spontaneous presentations” Proc. 2002 IEEE Int. Conf. Acoust., Speech, Signal Process., SP-P11.07, pp. I-729-732 , 2002

[5] T. Kawahara, H. Nanjo and S. Furui “Automatic transcription of spontaneous lecture speech” Proc. IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU2001), Madonna di Campiglio, Italy, 2001

[6] 横山忠介, 篠崎隆宏, 岩野公司, 古井貞熙 ”言語モデルのバッチ型教師なし適応化法” 電子情報通信学会技術報告, SP2002-151, 2002

[7] ルク ルシエ, エドワード ウィッテイカー, 古井貞熙 “話し言葉音声認識のための N グラム言語モデルの枠組みにおける種々の方法の検討” 電子情報通信学会技術報告, SP2003-141, 2003

[8] K. Ohtsuki, S. Furui, A. Iwasaki and N. Sakurai “Message-driven speech recognition and topic-word extraction” Proc. 1999 IEEE Int. Conf. Acoust., Speech, Signal Process., pp. 525-628, 1999

[9] 堀智織, 古井貞熙 “単語抽出による音声要約文生成法とその評価” 電子情報通信学会論文誌, J85-D-II, 2, pp. 200-209, 2000

[10] C. Hori and S. Furui “Advances in automatic speech summarization” Proc. Eurospeech 2001, pp. 1771-1774, 2001

[11] 菊池智紀, 古井貞熙, 堀智織 “重要文抽出と文圧縮による音声自動要約” 電子情報通信学会技術報告, SP2002-158, 2002

[12] C. Hori, S. Furui, R. Malkin, H. Yu and A. Waibel “Automatic speech summarization applied to English broadcast news speech” Proc. 2002 IEEE Int. Conf. Acoust., Speech, Signal Process., SP-L01.03, pp. I-9-12. 2002

[13] 広畑誠、新中庸介 “音声自動要約における SVD を用いた重要文抽出手法の検討” 日本音響学会秋季研究発表会, 2-6-17, 2003

[14] 新中庸介, 菊池智紀, 岩野公司, 古井貞熙 “音声自動要約を利用した講演速聞きシステムの検討” 情報処理学会研究報告, 2003-SLP-46(3), 2003

[15] 篠崎隆宏, 古井貞熙 “発話変動を考慮した隠れモード HMM による音声のモデル化 – 音声認識におけるベイジアンネットの応用 –“ 電子情報通信学会技術報告, SP2003-41, 2003

[16] J. Bilmes and G. Zweig “The Graphical Models Toolkit: An open source system for speech and time-series processing” Proc. 2002 IEEE Int. Conf. Acoust., Speech, Signal Process., ITT-P02.07, pp. IV-3916-3919 , 2002

[17] 西村義隆, 篠崎隆宏, 岩野公司, 古井貞熙 “周波数帯域ごとの重みつき尤度を用いた雑音に頑健な音声認識” 電子情報通信学会技術報告, SP2003-116, 2003

[18] 岩野公司, 関高浩, 古井貞熙 “雑音に頑健な音声認識のための韻律情報の利用” 情報処理学会研究報告, 2003-SLP-46(10), 2003

[19] 古井貞熙 “音声認識と知識の体系化” 日本音響学会誌, vol. 59, no. 10, pp. 589-590, 2003

[20] 古井貞熙 “21 世紀 COE プログラム「大規模知識資源の体系化と活用基盤構築」” 電子情報通信学会技術報告, SP2003-162, 2003

[21] 古井貞熙 “Ubiquitous／Wearable Computing 環境における音声認識の展望” 電子情報通信学会技術報告, SP99-114, 1999