

## 科学研究費助成事業 研究成果報告書

平成 26 年 6 月 16 日現在

機関番号：13501

研究種目：若手研究(B)

研究期間：2011～2013

課題番号：23700111

研究課題名(和文) 音声中の任意語彙検出の高度化と音声認識・音声ドキュメント内容検索への応用

研究課題名(英文) Improvement of Spoken Term Detection Technique and its Application to Speech Recognition and Spoken Document Retrieval

研究代表者

西崎 博光(NISHIZAKI, Hiromitsu)

山梨大学・医学工学総合研究部・助教

研究者番号：40362082

交付決定額(研究期間全体)：(直接経費) 3,200,000円、(間接経費) 960,000円

研究成果の概要(和文)：本研究の目的は、大規模音声データから任意の語彙が発話されている音声区間を特定する音声中の任意語彙検出(STD)技術の高度化を図ること、その技術を音声認識等に応用し精度を改善することにある。複数の音声認識結果、多数決信頼度、機械学習を利用するSTD手法を開発した。実験の結果、改良前のベースライン手法と比べてSTD精度が大幅に改善成功した。さらにSTDを音声認識に適用し、音声認識性能を改善することにも成功した。さらにSTDを実システム(電子ノート作成支援システム)に搭載しその有効性を評価し、企業等から高い評価を頂いた。

研究成果の概要(英文)：The goal of this study is refinement of the spoken term detection (STD) technique. An STD technology can detect speech intervals, where query terms are uttered, in lots of spoken documents. In addition, I also aim at adapting the STD method to other technologies such as speech recognition for improving their refinement. I developed the STD methods using multiple speech recognizers' outputs, confidence measures based on majority voting, and machine learning. In the experiment on STD, my techniques achieved improvement of the STD performance comparing to the baseline system. In addition, I adopted my STD technique to making recognition dictionary, which is necessary for speech recognition. In the result, my technique improves speech recognition performance. Furthermore, I implemented the STD technique to an electronic note-taking support system and evaluated its effectiveness of utilization of the STD technique. The system with STD is useful for retrieving words from multimedia data.

研究分野：総合領域

科研費の分科・細目：情報学・メディア情報学・データベース

キーワード：音声中の検索語検出 音声認識 音声ドキュメント検索 音声ドキュメント処理 機械学習 情報検索  
STD応用 言語モデル

## 1. 研究開始当初の背景

音声ドキュメント内容検索 (Spoken Term Detection, 以下 SDR と記す) は、古くは米国規格協会 (NIST) 主催の TREC 会議 (Text REtrieval Conference) のサブタスク (TREC-SDR) として行われてきた。SDR とは、ユーザの検索要求に関連する (1 ニュース記事や講義, 講演を単位とする) 音声ドキュメントを返すタスクである。これまでに、音声ドキュメントを音声認識することでテキスト化し、テキスト情報検索を行うことで実現する手法が多く提案されている。従って、音声認識誤りに如何に頑健な対策を施すかに焦点が当てられている。TREC-SDR では、比較的音声認識が簡易な英語や中国のニュース音声を対象にしている。ニュース音声に対する高い音声認識技術が確立されたため、TREC-SDR のタスクではテキスト検索と遜色ない結果を得るに至った。

近年では、同様に NIST が 2006 年に発話中の任意語彙検出 (Spoken Term Detection, 以下 STD と記す) タスクのコンペティション型ワークショップを開催している。STD は、ある音声 (ドキュメント) から、任意の語彙が発話されている区間を検出するタスクである。NIST の STD タスクは、TREC-SDR とは異なり、より音声認識が難しい電話会話等の話し言葉音声を対象にしている。

このように英語・中国語等では、音声データの検索技術開発のためのデータ整備が着々と進んでいるが、日本語のデータセットは 2006 年までは実現していなかった。そこで、情報処理学会音声言語情報処理研究会の「音声ドキュメント処理ワーキンググループ」(SDPWG) により、大規模音声データを対象とした音声ドキュメント内容検索テストコレクション、並びに音声中の任意語彙検出テストコレクションが開発された。これにより、日本語を対象とした SDR と STD 研究のための大規模なリソースが整いつつある。

そこで 2011 年に国立情報学研究所が主催する第 9 回 NTCIR ワークショップにおいて、正式に SDR と STD タスクが提案され、実施されることとなった。このことから、SDR・STD の研究が世界的に注目されていることが窺える。

現在、大規模な音声収録があらゆる場所で行われるようになった。例えば、裁判員制度の導入に伴う裁判記録、大学等での講義録音、議会の議事録、コールセンターの通話収集等が挙げられる。これらを瞬時に高精度で検索する技術の実用化はますます重要となって

くるのは明白である。特に企業は STD 技術に着目しており、特にコールセンターなどでオペレータの言動のチェックを守るために利用する動きもある。

## 2. 研究の目的

本研究の目的は大きく 2 つある。1 つ目は、大規模音声データから、任意の語彙が発話されている区間を特定する STD 技術の高度化を図ること、2 つ目はその技術を音声認識と SDR、その他に応用することで、これらの精度をさらに改善させることである。

### (1) STD 技術の高度化

科研費採択時、研究者が当時持っていた STD 技術では、SDPWG の STD テストコレクション (未知語タスク) を対象に、再現率約 65%、適合率約 60% の検索パフォーマンスを得ていた。しかし、誤検出 (本来は検出されてはいけないのに検出されてしまった誤り) が多く発生しているため、これを現行の半分に抑えること。すなわち、再現率を 70% 程度に保ったまま、適合率を 80~85% 程度まで高めることを目標とする。

### (2) STD の他技術への応用

STD 技術を音声認識の精度改善に応用する。これは、STD を用いることで、発話されている単語・フレーズを通常の音声認識手法よりも高い精度で検出することができることから、検出された語句を用いて言語モデルを話題適応化するというアイデアである。予備実験では、理想的に約 20% の音声認識性能の改善が見込めることが分かった。提案手法によって通常の音声認識性能よりも認識精度を改善できる技術を開発する。さらに、STD 技術を SDR の精度改善に利用することを試みる。

### (3) STD の有効性検証

最後に、申請当初は STD に関する基礎研究の開発のみを目標に据えていたが、STD の応用範囲は広いと考えて、この技術を組み込んだアプリケーションの開発も行い、STD 技術の有効性を示す。

## 3. 研究の方法

2 節(1)~(3) について、各方法を説明する。

### (1) STD 技術の高度化

本件については、主に以下の 3 つのことについて取り組んだ。

- ・複数の音声認識結果から得られる多数決信頼度を利用した STD の改善
- ・曖昧性尺度 (エントロピー) を用いた STD の改善
- ・機械学習を用いた STD の改善

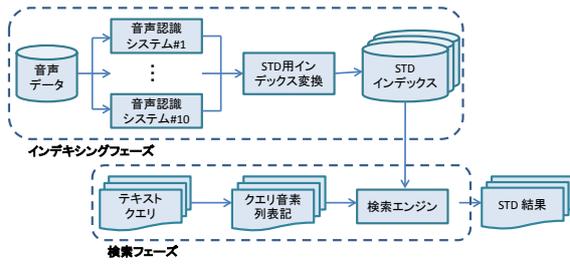


図1 DP法に基づくSTD処理の概要

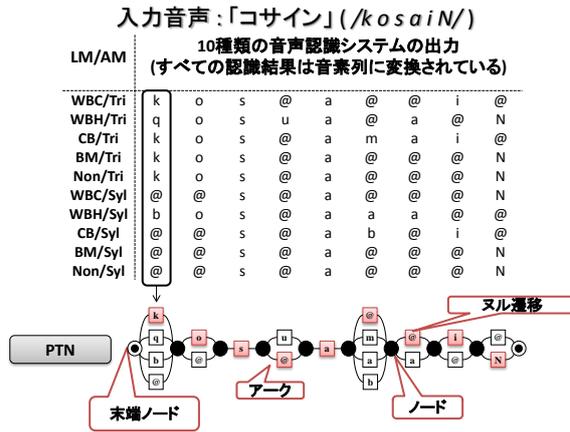


図2 PTN

以下、順に開発した手法を説明していく。

① 複数の音声認識結果から得られる多数決信頼度を利用したSTDの改善

科研費採択前は、図1, 図2に示すように、STD用のインデキシングフォーマットとして、複数の音声認識システムを利用した音素遷移ネットワーク (Phone Transition Network, PTN, 図2参照)を開発し、成果を挙げた。しかし、前述した通り誤検出が多く発生する。そのため、本研究では、これを抑制するための誤検出抑制パラメータを開発した。

図2に示すように、誤検出抑制パラメータは、複数の音声認識結果の多数決情報から求める。これは、同じ音素を一つの複数の音声認

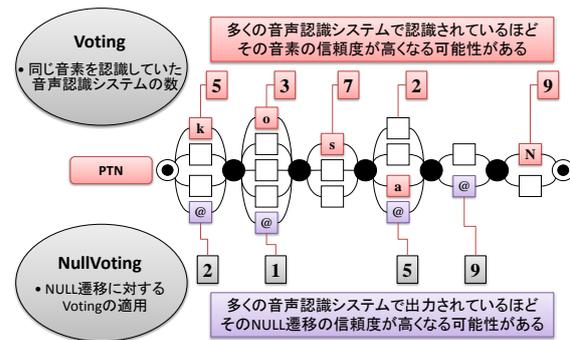


図3 誤検出抑制パラメータ

識システムが出力していれば、その信頼度は高く、また複数の認識システムから出力される音素の種類数が少ない方が信頼度は高いという考えに基づく。

このパラメータを検索エンジンに組み込んだ。検索エンジンでは、動的計画法 (DP) ベースの手法を採用しており、コスト計算にこの誤検出抑制パラメータを組み込む形で利用する。

② 曖昧性尺度(エントロピー)を用いたSTDの改善

本手法は、インデックス (PTN) のエントロピーと検索要求 (クエリ) のエントロピーを用いた曖昧性尺度に基づくパラメータをSTDエンジンに組み込むことで精度改善を図る方法である。

PTNのエントロピーは、誤検出抑制パラメータのうちの“Voting” (図1参照)を用いる。これは同じ音素を認識していた音声認識システムの数であり、これを用いて音素の事後確率を求める。これを用いてPTNのエントロピーを計算する。STDエンジンにより検索された検出区間のPTNエントロピーを求め、この値の大小により検出結果が適当であるかを検証する。

クエリのエントロピーを求めるためには、まず、音素単位に分割したコーパス (今回は、「日本語話し言葉コーパス」を利用) から、トライグラム近似を用いた統計的言語モデルを学習しておく。これを用いて、クエリの音素表記列からエントロピーを計算する。すなわち一種の“音素のつながりやすさ”を数値化する。この値をDPベースの検索エンジンに組み込む。

③ 機械学習を用いたSTDの改善

本手法では、上記2つのDPベースの手法とは異なり、単語の検出を条件付き確率場 (Conditional Random Fields, CRF) に基づいた統計的学習モデルによって行う方法である。検索の手法の概要を図4に示す。

この手法では、10種類の音声認識システムから得られる音素系列の書き起こしを利用して、CRFモデルを学習する。検索語は、一度音素列に変換され、さらにtriphoneに分解される。各triphoneに対応するCRFモデルを用いて、ある発話にそのtriphoneが含まれている確率を計算する。検索語を構成する全てのtriphoneの条件付き生起確率の総乗を、この検索語の検出スコアとする。素性には、unigram, bigram, trigramを用いる。

図5に示すようにbigramについては、current tokenを中心として前後1つのコン

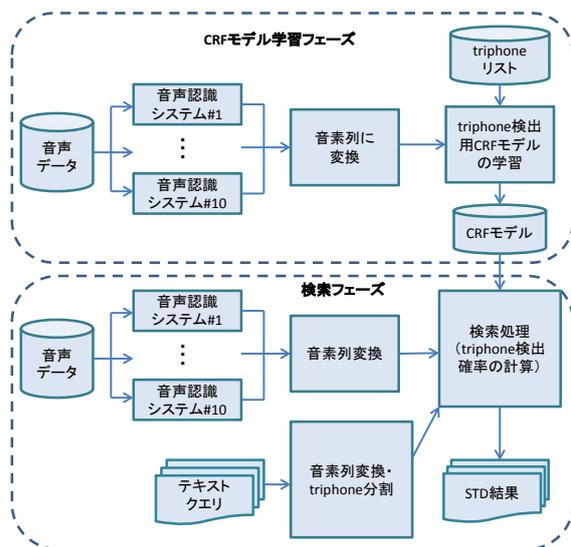


図4 CRFを用いたSTD処理の概要

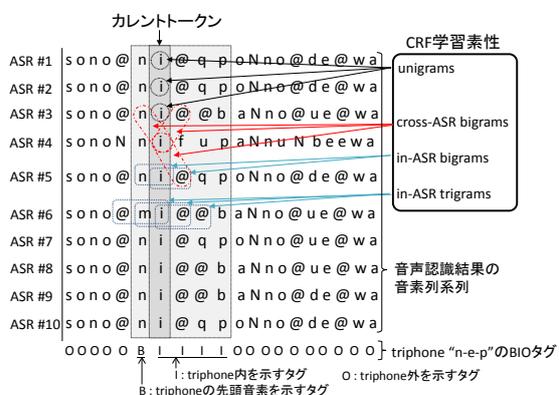


図5 CRFで用いる特徴量

テキストを考慮する。また、trigramも、2つ前までの履歴、前後1つずつ、後ろ2つまでのコンテキストを考慮している。さらに、認識システム間のbigramも素性に利用する。これによって、音素間の多様な音声認識誤りのパターンをCRFで学習させ、認識誤りを含んだ音素系列の書き起こしに対して頑強なtriphone検出を試みる。

## (2) STDの他技術への応用

### ① 音声認識への応用

STD技術を音声認識の精度改善に応用する。提案手法を図6に示す。提案手法では、音声認識で用いる言語モデルや認識辞書に登録する単語を取捨選択することで、モデルの話題を適応化し、その結果認識率を改善する。

まず、汎用的な言語モデルと認識辞書を用いて一度音声を認識する。汎用的な認識辞書に含まれている単語をクエリとして、STDを行う。その単語が検出されれば、再度辞書に

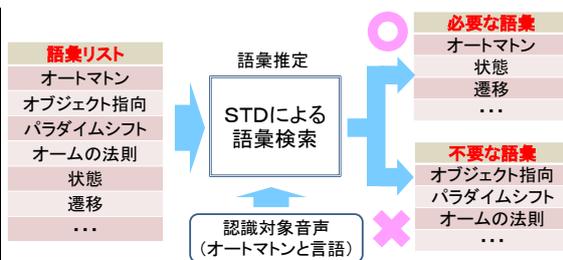


図6 STDを用いた話題適応化

登録、検出されなければ辞書に登録しない。STDによって単語を取捨選択するわけである。これを発話単位で行うことで、発話に適した認識辞書を構築することができる。

### ② STDを用いた音声ドキュメント検索

検索要求に含まれるキーワード(内容語)が検索対象の音声ドキュメント内で正しく認識されていなければ、検索ができない。そこで、音声認識誤りが含まれていてもキーワード検出がSTDで精度よくできれば、音声認識誤りに頑健なSDRが実現できるはずである。そこで、まず、検索要求に含まれるキーワードを取り出し、それをSTDで検出する。

### ③ STDの有効性検証

電子ノート作成支援システムを構築し、これにSTD技術を組み込むことで、STDの有効性を利用者の立場から検証する。

開発した電子ノート作成支援システムでは、講義音声の音声認識と録音、スクリーン(黒板)画像のキャプチャ、キャプチャ画像への手書き文字の加筆、ができる(図7参照)。ここで復習する際に、STDを用いた頭出しが行える機能を実装した。音声認識誤りがおきたとしても、指定したキーワードが話されている場所を的確に見つけることで、実際にノートの参照に有効かどうかを検証した。

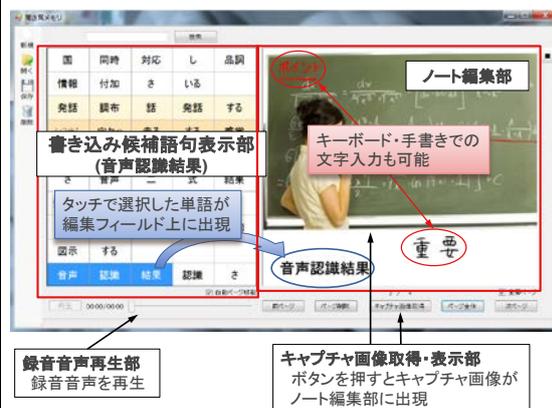


図7 電子ノート作成支援システム

#### 4. 研究成果

##### (1) STD 技術の高度化

① 複数の音声認識結果から得られる多数決信頼度を利用した STD の改善

評価に用いるデータは、NTCIR-9 のフォーマルランセットを用いる。評価指標は再現率・適合率 (Recall-Precision) カーブである。結果を図 8 に示す。

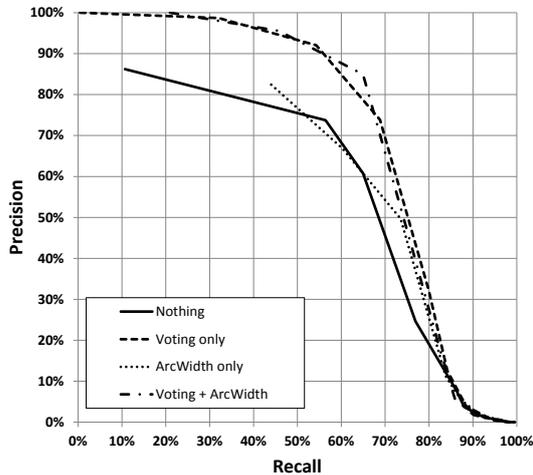


図 8 誤検出抑制パラメータの効果

図 8 で、グラフが右上に行くほど性能が良いことを示し、“Nothing”が誤検出抑制パラメータを利用しない、“Voting+ArcWidth”がパラメータ全てを利用する方法である。目標（再現率 70%、適合率 80%）には達しないまでも大幅な改善を得た。

② 曖昧性尺度 (エントロピー) を用いた STD の改善

まず、検出区間の PTN のエントロピーを用いた手法では、ほとんど改善が得られなかった (結果は割愛する)。これは誤検出抑制パラメータと特徴が似ているためであると考えられる。

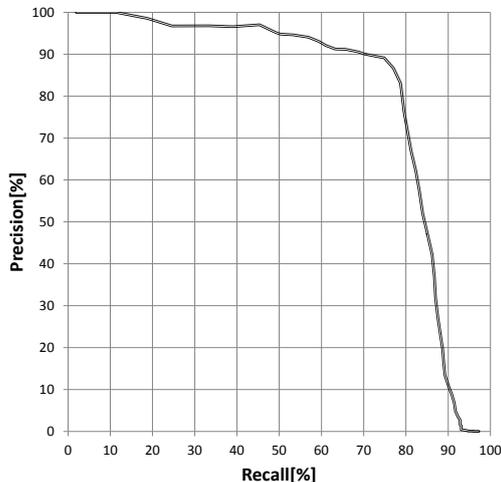


図 9 エントロピーの効果

つぎに、クエリのエントロピーを用いた場合の検索結果を図 9 に示す。図 8 に比べて大幅に検索結果が改善していることが分かる。本手法により検索性能の目標値を達成した。

③ 機械学習を用いた STD の改善

これは、「未知語テストセット」で評価を行った (①, ② と異なるデータ)。評価は F 値 (再現率と適合率の調和平均) を用いる。

実験の結果、CRF 単独の結果では、F 値が 53.8% であり、ベースライン (① の方法) の 73.1% を大きく下回る結果となった。しかしながら、提案手法とベースライン手法を組み合わせることで、F 値が 74.3% まで改善できたことから、提案手法の有効性を示すことができた。

(2) STD の他技術への応用

① 音声認識への応用

提案手法を大学講義音声で評価した。評価尺度は音声認識率である。何も行わないベースライン手法では音声認識率 52.3% だったが、提案手法では、音声認識率が 54.7% まで改善した。名詞のみに注目すると、ベースライン 42.5% が 48.7% と改善し、特に内容語の音声認識率改善に効果があることが分かった。すなわち、この方法は SDR や STD に有効であると言える。

② STD を用いた音声ドキュメント検索

実験の結果、提案手法の効果はほとんどなかった。これは、検索要求中のキーワードはほとんど音声認識されていたことによる。これは他の研究機関発表と同じ傾向である。しかしながら、未知語 (認識できない語句) を含んだ検索要求の場合には効果があると考えられるため、引き続き研究を行って有効性を調査する予定である。

(3) STD の有効性検証

電子ノート作成支援システムによる STD の有効を検証した。実験の結果、頭出しの効率化 (ノート参照にかかる時間の低下) や、メディア情報アクセスのしやすさなど、客観的・主観的に STD が有効であることが分かった。

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 13 件)

- ① 西崎博光, 太田晃平, 関口芳廣, 聞き耳メモリー: 音声認識結果を取捨選択して利用する電子メモ作成支援アプリケーション, 電子情報通信学会論文誌, Vol. J96-D, No.10, pp.2379-2389, 2013.10, 査読有 <http://ci.nii.ac.jp/naid/110009674111>
- ② Satoshi Natori, Yuto Furuya, Hirimitsu Nishizaki, Yoshihiro Sekiguchi, Spoken Term Detection Using Phoneme Transition Network from Multiple Speech Recognizers' Outputs, Journal of In-

formation Processing, Vol.21, No.2, pp.176-185, 2013.4, 査読有  
[https://www.jstage.jst.go.jp/article/ipsjji/p/21/2/21\\_176/\\_article](https://www.jstage.jst.go.jp/article/ipsjji/p/21/2/21_176/_article)

- ③ 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, 2013.2, 査読有  
<http://ci.nii.ac.jp/naid/110009537042>
- ④ Hiromitsu Nishizaki, Tomoyosi Akiba, Kiyooki Aikawa, Tatsuya Kawahara and Tomoko Matsui, Evaluation Framework Design of Spoken Term Detection Study at the NTCIR-9 IR for Spoken Documents Task, Journal of Natural Language Processing, Vol.19, No.4, 329-350, 2012.12, 査読有  
[https://www.jstage.jst.go.jp/article/jnlp/19/4/19\\_329/\\_pdf/](https://www.jstage.jst.go.jp/article/jnlp/19/4/19_329/_pdf/)
- ⑤ 西崎博光, 杉本樹世貴, 関口芳廣, 音声ドキュメント内容検索のための WEB を用いたドキュメント拡張, 情報処理学会論文誌, Vol.52, No.12, pp.3461-3470, 2011.12, 査読有  
<http://ci.nii.ac.jp/naid/110008719923>  
(8 件省略)

[学会発表] (計 17 件)

- ① 米倉千冬, 古屋裕斗, 澤田直輝, 名取賢, 西崎博光, 関口芳廣, 音声ドキュメントからの頻出発話語句の発見, 第 8 回音声ドキュメント処理ワークショップ講演論文集, pp.1-7, 2014 年 3 月 15 日発表, 愛知県・豊橋市民センター)
- ② 澤田直輝, 古屋裕斗, 名取賢, 西崎博光, 関口芳廣, STD システムへの音素間距離の導入方法の検討, 日本音響学会 2014 年春季研究発表会講演論文集, 3-Q5-11, pp. 213-216, 2014 年 3 月 12 日発表, 東京都・日本大学駿河台キャンパス
- ③ 古屋裕斗, 名取賢, 西崎博光, 関口芳廣, クエリのエン트로ピーを利用した STD 手法の検討, 日本音響学会 2014 年春季研究発表会講演論文集, 3-4-6, pp. 83-86, 2014 年 3 月 12 日発表, 東京都・日本大学駿河台キャンパス
- ④ 西崎博光, 古屋裕斗, 名取賢, 関口芳廣, 条件付き確率場を用いた音声中の検索語検出の検討, 日本音響学会 2013 年秋季研究発表会講演論文集, 2-P-26, pp. 217-220, 2013.9, 2013 年 9 月 26 日発表, 愛知県・豊橋技科大
- ⑤ 古屋裕斗, 西崎博光, 関口芳廣, 音声中の検索語検出のための音素遷移ネットワークのエン트로ピー分析, 第 7 回音声ドキュメント処理ワークショップ講演論文集, 2013 年 3 月 2 日発表, 愛知県・名古屋大

- ⑥ 古屋裕斗, 西崎博光, 関口芳廣, STD のための音素遷移ネットワークのエン트로ピーによる分析, 日本音響学会 2012 年秋季研究発表会講演論文集, 3-P-31, pp.203-204, 2012 年 9 月 21 日発表, 長野県・信州大
- ⑦ 古屋裕斗, 名取賢, 西崎博光, 関口芳廣, 音声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講演論文集, pp.1-8, 2012 年 3 月 3 日発表, 愛知県・豊橋技科大  
(10 件省略)

[産業財産権]

○出願状況 (計 1 件)

名称: 電子ノート作成支援装置および電子ノート作成支援装置用プログラム  
発明者: 西崎博光, 太田晃平, 関口芳廣  
権利者: 国立大学法人山梨大学  
種類: 特許  
番号: 特願 2012-236445  
出願年月日: 2012 年 10 月 26 日  
国内外の別: 国内

[その他]

ホームページ等

<http://www.alps-lab.org/>

6. 研究組織

(1) 研究代表者

西崎 博光 (NISHIAKI, Hiromitsu)

山梨大学・医学工学総合研究部・

助教

研究者番号: 40362082

(2) 研究分担者

なし

(3) 研究連携者

なし

(4) 研究協力者

関口 芳廣 (SEKIGUCHI, Yoshihiro)

山梨大学・医学工学総合研究部・

教授

研究者番号: 70020493