

平成21年 5月25日現在

研究種目：若手研究（B）
 研究期間：2006～2008
 課題番号：18700173
 研究課題名（和文） 大学の講義・講演音声のコンテンツ化とその内容評価に関する研究
 研究課題名（英文） Study on Construction of Lecture Speech at University and Its Evaluation
 研究代表者
 西崎 博光（HIROMITSU NISHIZAKI）
 山梨大学・大学院医学工学総合研究部・助教
 研究者番号：40362082

研究成果の概要：

本研究では、大きく2つの技術を開発した。一つは収録した大学の講義や学会などの講演音声処理し、それらの高精度に検索できる技術の開発、二つ目は、収録した講義・講演音声から、その話し方の評価を（半）自動で行う技術の開発である。講義・講演音声の検索は Web を利用する全く新しい方法を開発した。話し方の評価は、音響的な要因と言語的な要因双方から分析を行い、話し方のよし悪しを評価するための枠組みを構築した。

交付額

(金額単位：円)

	直接経費	間接経費	合計
2006年度	1,500,000	0	1,500,000
2007年度	1,000,000	0	1,000,000
2008年度	1,000,000	300,000	1,300,000
年度			
年度			
総計	3,500,000	300,000	3,800,000

研究分野：音声言語情報処理

科研費の分科・細目：情報学・知覚情報処理，知能ロボティクス

キーワード：音声ドキュメント検索，講義音声，音声データベース，話し方評価

1. 研究開始当初の背景

国立大学の法人化に伴い、各大学が社会の教育要求水準を満たしているのか等、大学自体の質の向上が要求されるようになってきた。今後は、学生にとって分かりやすく面白い授業を行うことは教員にとって必須課題であると言えよう。

一方で、情報通信技術（ICT）の発展により、テキスト情報のみならず、様々な音声や映像をデジタルデータとして蓄積・配信することが可能になっている。最近では、各大学で e-Learning の一貫として授業の講義ビデオや講演ビデオの収録が盛んに行われて

いる。しかし、大量に講義等を録画してもそれらを有効に利用するには人手による多大な労力が必要であり、せっかく収録したビデオがあまり有効利用されていないのが現状である。

各大学間における単位互換、生涯教育、インターネットなどの遠隔講義に見られるように、教育環境は年々変化しており、将来的にはこういった e-Learning による学習の比重は上がってくると考えられ、収録した講義や講演音声のコンテンツ化が課題として挙がってくるのは容易に想像できることである。

2. 研究の目的

本研究では 1. で述べた研究背景を鑑み、教員の講義ビデオなどから音声データを抽出し、それらに基づいた授業評価が(半)自動で行えないかを検討する。

これまでの事前調査で、学生に対して授業に対するアンケートを採って見たところ、声の大きさや抑揚などが授業の分かりやすさに影響することが分かっている。また、教員の熱意が声に現れることも調査結果として判明した。

そこで、本研究課題では大きく3つの研究目的を設定した。

(1) 教員等の発話者の声から考えられるあらゆる特徴を抽出し、その特徴と講義・講演評価がどのように関係しているかを調査する。この結果を元に、講義・講演における話し方の(半)自動評価に利用する。この評価結果を元に、教員や講演者へのフィードバックを行い、講義・講演の質を向上させるために役立つシステムを開発することが目的の一つである。

(2) 学生が講義終了後に復習したり、また講演などを再度聴講したりする場面が十分に想定できるが、こういった要望に応えるためにも講義や講演を自動でインデキシング化しておき、ユーザの所望する内容を表す情報を音声言語情報処理技術を用いて手軽に精度よく検索(情報検索技術)することも、研究目的の一つに据える。

(3) 講義・講演音声の話し方の自動評価、インデキシング化を行うにあたり共通の技術として重要である音声認識の精度改善手法に関しても、重点的に研究を行うことも研究の目的である。

3. 研究の方法

(1) 講義・講演音声の話し方自動評価

本項目では、主にデータ収集と音声特徴量の調査・分析を行う。

大学教員に協力を仰ぎ、講義音声データを

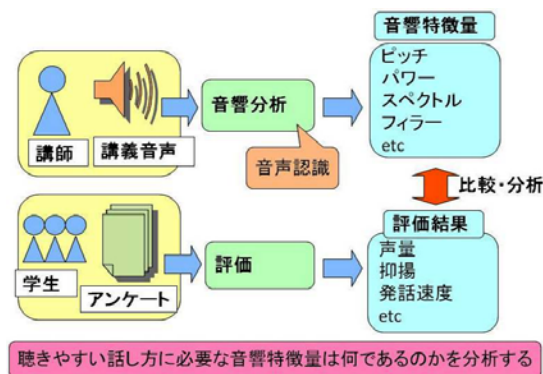


図 1. 話し方の分析

収集する。さらに、既存の学会等の講演音声データを被験者に聴取してもらい、その話し方についてのアンケートを取ることで、“音声”と“聴き手の印象評価”のデータセットを大量に用意する。

用意したデータを用いて、様々な話し方の特徴量と印象評価を関連付け、聴き手の印象に与える影響が大きい特徴を見出していく。図 1 に研究手法の概要を示す。

現在、様々な音声特徴量が、音声認識や感情分析、韻律分析などの研究分野で提案されている。

そこで、以下の事柄について研究を行う。

①音響的特徴

音声の音響的な特徴には、声の高さ・大きさ・話速・明瞭度などがある。音声から抽出した物理的特徴との関連性を明らかにする。また、音声認識の出力結果と講義評価との関連性も調査する。

②言語的特徴

言語的特徴としては、フィラー(間投詞)・言い直し・ポーズの入れ方などが考えられる。特にフィラーの使い方の有効性を検討する。

また、フィラーの使い方を自動的に評価するシステムの開発も行う。

③講義音声のデータベース化

また、収集した大学講義音声データベース化し、研究目的での公開を行うことで、講義ドキュメント処理の研究促進に役立てる。

(2) 講義・講演音声の検索技術の開発

通常、音声データを検索するようにするために、音声認識技術に頼ることが多い。音声を認識して文字列に変換し、それをテキスト検索することによって、大量の講義や講演音声を検索することができる。それゆえ、音声認識精度が検索精度に大きく影響する。

本研究では、音声認識技術を利用することを前提に、下記の項目について研究を行う。

①音声認識率別の検索精度調査

音声認識精度が検索精度にどれくらい影響するのかを、認識率別に調査する。

②インデキシング単位の調査

各音声ドキュメントに対して、インデキシングを作成する際、単語単位、文字単位、サブワード等様々な検索語単位が考えられる。有効なインデキシングの単位を調査する。

③Web によるドキュメント拡張の効果

Web によるドキュメント拡張を行うことで音声ドキュメントを高精度に検索する手法を開発する。

検索したい講演音声集合の各講演と内容が類似している Web データを利用してインデキシングを行うことで、未知語や認識誤りに頑健な検索方法を提案する。概要図を図 2 に示す。

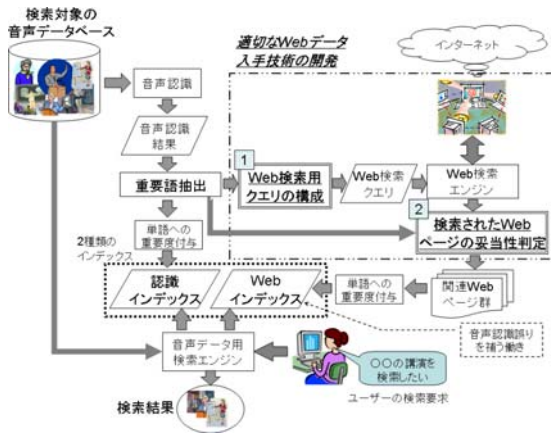


図 2. 研究内容概略図

(3) 音声認識精度の改善

講義評価や音声の検索には、音声認識技術が欠かせない。したがって、この精度を改善することが課題となる。本研究では、大きく2つの項目について研究を行う。提案手法の概要を図3に示す。

①Webを用いた言語モデルの話題適応化

講義音声は、話題が限定されていることがほとんどである。音声認識精度を改善する方法として、言語モデルの話題適応化が良く利用されている。

適応化処理は、あらかじめ音声認識対象の音声内容が良く分かっており、かつ適応化に利用できるテキストが存在する場合には非常に有効に働く。しかし、講義・講演の話題はある程度（例えば講義名等で）の話題限定ができて、適応化に利用できるデータがない。

本研究では、インターネット上に存在する大量のテキストデータに着目した。講義名などを利用して、その講義と関連するテキストデータを大量に収集し、それを用いて言語モデルを適応化する。この話題適応化言語モデルを利用することで、音声認識率の改善を目指す。

②語彙推定処理の開発

通常、大語彙連続音声認識を行う場合は、大語彙を音声認識辞書に登録するのが一般的である。これは、できるだけ多くの語彙をカバーすることで、未知語による誤認識等の致命的な問題を回避するためである。

しかし、音声認識の精度は、語彙のカバー率が同じであるならば、辞書に登録されている語彙数が小さい方が高くなる。講義・講演などの発話内容が狭い話題に限定されている場合、多くても数千種類の語彙しか使われていない。

そこで、本研究では、音声認識辞書に登録する語彙を推定することで、講義・講演音声の認識で用いる辞書のサイズをコンパクトにする。これにより、音声認識精度の改善が

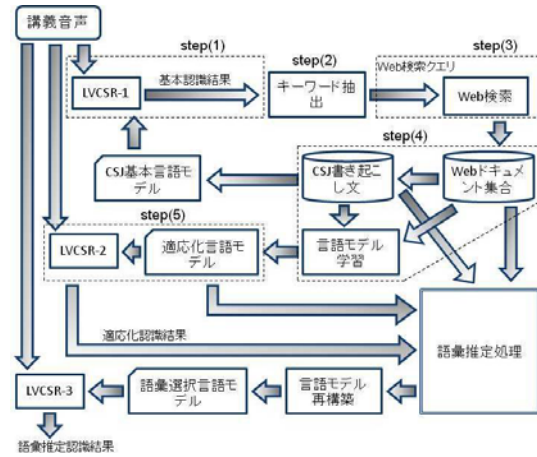


図 3. Web を用いた音声認識処理

期待できる。

具体的には、潜在的意味解析 (LSA) と単語の発音情報に基づく手法を開発する。

4. 研究成果

(1) 講義・講演音声の話し方自動評価

①音響的特徴の分析

講義音声に関する次の要因の含まれる、ピッチ、パワー、発話速度と講義音声の印象評価との関連を調査した (学会発表文献⑧⑫⑬⑯⑰)。

分析の結果、パワーやピッチなどが音声の印象評価に大きく影響していることが分かった。この結果を用いて講師音声の特徴を視覚化したものが図4である。縦軸がテンポであり、上にいくほど適切な速さで発話している。横軸は、声の質で明瞭性等に対応する。

また、音声認識率との関連性も調査し、認識率と声の評価に高い関連性があることも分かった。

講義評価アンケート単体の分析や、一般的に「良い話し方はこうである」という書籍はあるが、本研究で行っているような音声に含まれる物理的特徴量と評価の関係を明らかにした研究は他にない。

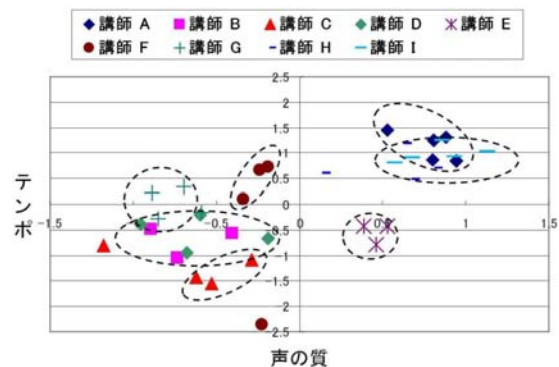


図 4. 講師音声の特徴

今後は、さらに詳細な分析を行い、大学等の講義評価支援に役立てていきたい。

②言語的特徴の分析の評価システム

話し言葉の言語的特徴の一つであるフィラー（間投詞）の研究を行った。この研究では、以下の研究成果を挙げた。すなわち：

1. フィラーが講演・講義の聞きやすさや理解しやすさに影響していることを明らかにした（学会発表文献⑩），
2. フィラーの使い方を自動で評価するシステムを開発（学会発表文献⑤），
3. 話し言葉中のフィラーの検出精度を改善する方法を開発（学会発表文献①⑦），

の3項目である。

1. においては、まず、アンケート調査により、聞きやすさや理解しやすさを基準に、発話中に含まれる最適なフィラーの数、長さ等の条件を求めた。その条件を利用し、原音声に対して人工的にフィラーの数や長さなどを調整した加工音声を使って、聞き比べ実験を行った。実験結果は、加工音声の方がよりよい印象であった。

2. においては、1. の実験でフィラーが講義や講演の聞きやすさに大きく影響していることから、フィラーの使い方を評価できるシステムができれば、講演や講義が改善できるのではないかと考えた。重回帰分析を利用したシステムでは、人間の評価と6割以上一致する評価システムを構築することができた。

3. では、話し言葉中のフィラーを自動で検出する方法を開発した。SVMによる機械学習法を導入することで、大幅に検出精度を改善することができた。

③講義音声のデータベース化

収録した講義音声を、他大学収録分も含めてデータベース化し（Corpus of Japanese Lecture Contents）、研究利用を目的に公開した¹。このデータベースは、今後の音声ドキュメント処理の研究発展に大いに役立つと思われる。（雑誌論文②③⑤，学会発表文献⑩⑪⑬）

（2）講義・講演音声の検索技術の開発

情報処理学会音声言語情報処理のワーキンググループが制定した「日本語話し言葉コーパスを対象とした検索テストコレクション」（雑誌論文①⑥，学会発表文献⑭⑰）を用いて、打規模な講演音声データベースからの検索精度を調査した。

①音声認識率別の検索精度調査

インデキシングを行う際に用いた音声認識結果において、その認識率別に検索精度の変化を調査した。結果を図5に示す。この図は、5k, 7k, 10k, 12k, 17k サイズの辞書を使って音声認識を行い、その出力結果を用いて

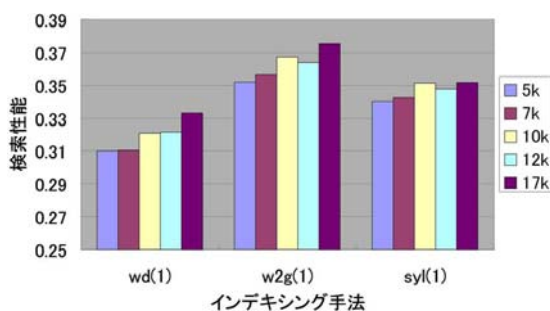


図5. 認識率別検索精度

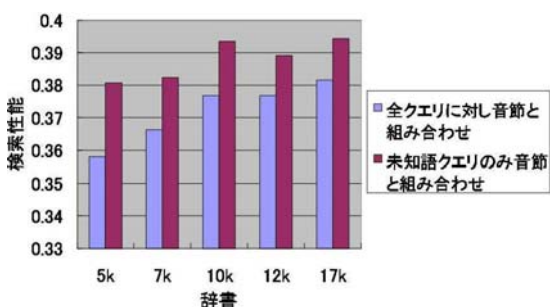


図6. 音節インデックスの組合せ効果

検索実験を行ったときの結果である。語彙サイズが小さくなるほど認識率が低下する。この結果より、音声認識精度の低下が検索精度に悪影響を与えていることが分かる。

②インデキシング単位の調査

インデキシングの単位を変えて実験を行った。単位として、「形態素」、「文字の2つ組」、「音節」を調査した。

図6に実験結果を示す。結果より、「形態素」(wd(1))と「文字の2つ組」(w2g(1))の組み合わせが有効であることがわかった。また、音節インデックス(syl(1))も、音声認識誤りや未知語に効果があり、前者の組み合わせにさらに音節インデックスを組み合わせることで、検索精度が大きく改善した(図6)。

③Webによるドキュメント拡張の効果

提案手法では、音声認識結果から直接作成した認識インデックスと類似Webテキストから作成したWebインデックス、それぞれのインデックスから検索された結果を、重み付で混合することで検索精度を改善する。

図7にWebによるドキュメント拡張を利用したときの検索精度を示す。

横軸は、混合時の重みである。

図7に示すように、Webを利用することで、検索精度が改善しており、ドキュメント拡張の効果が大きいことが分かる。

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

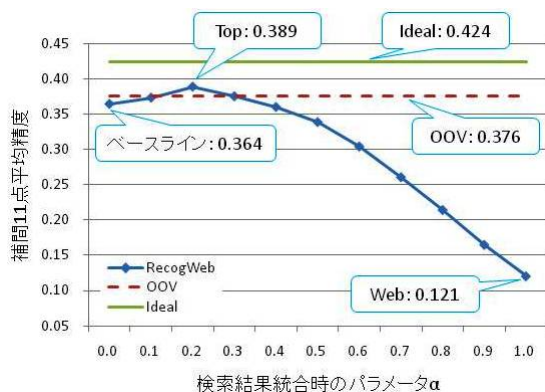


図 7. 音節インデックスの組合せ効果

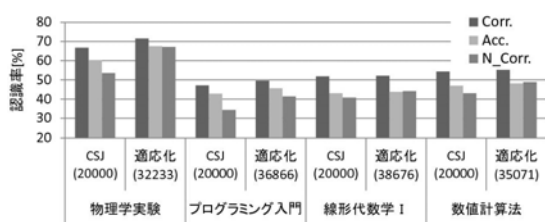


図 8. 音節インデックスの組合せ効果

(3) 音声認識精度の改善 (学会発表文献⑨)

① Web を用いた言語モデルの話題適応化
4 つの講義を対象に、Web を用いた言語モデルの適応化実験を行った。

図 8 に実験結果を示す。Web を用いた適応化は、音声認識率の改善に大きな効果があることが示された。

② 語彙推定処理の開発 (学会発表文献③)
次に、語彙推定処理を行い、音声認識辞書に登録する語彙を削減する効果を確認する。

実験の結果、若干ではあるが音声認識率が改善した。音声認識率の改善はわずかではあるが、語彙を削減したことにより、認識処理に必要なメモリの削減 (削減率約 50%)、音声認識処理の高速化 (20%以上) に効果が見られた。

語彙推定処理は、従来の音声認識のアイデア (すなわち辞書のサイズを大きくする) に逆行する斬新な考え方である。認識率の改善率はわずかであったが、処理速度の高速化など、様々な利点が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 8 件)

- ① T. Akiba, K. Aikawa, Y. Itoh, T. Kawahara, H. Nanjo, H. Nishizaki, N. Yasuda, Y. Yamashita, K. Itoh, "Construction of a Test Collection for Spoken Document

Retrieval from Lecture Audio Data", 情報処理学会論文誌, Vol.50, No.2, pp.501-513, 2009, 査読有

- ② 土屋雅稔, 小暮悟, 西崎博光, 太田健吾, 山本一公, 中川聖一, "日本語講義音声コンテンツコーパスの作成と分析", 情報処理学会論文誌, Vol.50, No.2, pp.448-459, 2009, 査読有

- ③ S. Kogure, H. Nishizaki, M. Tsuchiya, K. Yamamoto, S. Togashi, S. Nakagawa, "Speech Recognition Performance of CJLC: Corpus of Japanese Lecture Contents", the Proceedings of the 9th annual conference of the International Speech Communication Association, pp. 1554-1557, 2008, 査読有

- ④ K. Kobayashi, M. Somiya, H. Nishizaki, Y. Sekiguchi, "Is a Speech Recognizer Useful for Characteristic Analysis of Classroom Lecture Speech?", Proceedings of the 9th annual conference of the International Speech Communication Association, pp. 1341-1344, 2008, 査読有

- ⑤ M. Tsuchiya, S. Kogure, H. Nishizaki, K. Ohta, S. Nakagawa, "Developing Corpus of Japanese Classroom Lecture Speech Contents", Proceeding of the 6th edition of the Language Resources and Evaluation Conference, 2008, 査読有

- ⑥ T. Akiba, K. Aikawa, Y. Itoh, T. Kawahara, H. Nanjo, H. Nishizaki, N. Yasuda, Y. Yamashita, K. Itoh, "Test Collections for Spoken Document Retrieval from Lecture Audio Data", Proceeding of the 6th edition of the Language Resources and Evaluation Conference, 2008, 査読有

- ⑦ H. Nishizaki, M. Sohmiya, K. Kobayashi, Y. Sekiguchi, "The Effect of Filled Pauses in a Lecture Speech on Impressive Evaluation of Listeners", Proceedings of the 8th Annual Conference of International Speech Communication Association, pp.2673-2676, 2007, 査読有

- ⑧ H. Nishizaki, Y. Sekiguchi, "Word Error Correction of Continuous Speech Recognition Using WEB Documents for Spoken Document Indexing", Lecture Notes in Artificial Intelligence, No.4285, Springer, pp. 213-221, 2006, 査読有

[学会発表] (計 17 件)

- ① 宗宮充宏 (西崎博光), "音声認識結果に基づく SVM を用いた講義・講演音声中のフレーズ検出の検討", 日本音響学会 2009 年

- 春季研究発表会, 3-Q-19, pp.247-248, 2009年3月19日, 東京工業大学
- ②前沢慎吾 (西崎博光), “音声ドキュメント検索用テストコレクションにおける音節インデックスの効果”, 情報処理学会第71回全国大会, 4Q-7, pp. 127-128, 2009年3月11日, 立命館大学
- ③藤原裕幸 (西崎博光), “講義音声認識のためのLSAを利用した語彙推定手法の検討”, 第3回音声ドキュメント処理ワークショップ, pp.83-90, 2009年2月28日, 豊橋技術科学大学
- ④杉本樹世貴 (西崎博光), “検索対象と類似性の高いWebページを利用した音声ドキュメント検索の検討”, 第3回音声ドキュメント処理ワークショップ, pp. 33-38, 2009年2月27日, 豊橋技術科学大学
- ⑤宗宮充宏 (西崎博光), “講義・講演音声を対象としたフィラーの使い方自動評価システム”, 日本音響学会2008年秋季研究発表会, 2-P-19, pp.399-400, 2008年9月11日, 九州大学
- ⑥小林健司 (西崎博光), “講義音声を対象とした音声評価と認識率の関係”, 日本音響学会2008年秋季研究発表会, 2-P-18, pp.397-298, 2008年9月11日, 九州大学
- ⑦宗宮充宏 (西崎博光), “話し言葉音声中のフィラー検出精度と音声認識率の関連性”, 第7回情報科学技術フォーラム, E-024, pp.191-192, 2008年9月3日, 慶應義塾大学
- ⑧小林健司 (西崎博光), “講義音声の自動評価のための各種特徴量の調査”, 第2回音声ドキュメント処理ワークショップ, pp.143-148, 2008年3月1日, 豊橋技術科学大学
- ⑨徳田翔 (西崎博光), “講義音声認識のためのWEB文書を用いた言語モデルの適応化と語彙選択”, 第2回音声ドキュメント処理ワークショップ, pp.97-104, 2008年3月1日, 豊橋技術科学大学
- ⑩小暮悟 (西崎博光), “日本語講義音声コンテンツコーパスの構築と講義音声認識手法の検討”, 第2回音声ドキュメント処理ワークショップ, pp. 7-14, 2008年2月29日, 豊橋科学技術大学
- ⑪小暮悟 (西崎博光), “日本語講義コンテンツコーパスの構築と分析”, 日本音響学会2007年秋期講演発表会, 1-3-4, pp.13-16, 2007年9月19日, 山梨大学
- ⑫小林健司 (西崎博光), “講義音声の特徴分析”, 日本音響学会2007年秋期講演発表会, 1-P-13, pp.429-432, 2007年9月19日, 山梨大学
- ⑬小林健司 (西崎博光), “講義音声の分析と評価”, 日本音響学会春期講演発表会, 1-Q-14, pp.291-292, 2007年3月13日,

- 芝浦工業大学
- ⑭秋葉友良 (西崎博光), “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, pp.73-80, 2007年2月28日, 豊橋技術科学大学
- ⑮小暮悟 (西崎博光), “講義コンテンツの収集・分析および講義音声の認識手法に関する検討”, 第1回音声ドキュメント処理ワークショップ, pp.1-8, 2007年2月27日, 豊橋技術科学大学
- ⑯宗宮充宏 (西崎博光), “講義・講演音声自動評価のための音響・言語的分析”, 第1回音声ドキュメント処理ワークショップ, pp.9-16, 2007年2月27日, 豊橋技術科学大学
- ⑰伊藤克亘 (西崎博光), “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”, 情報処理学会音声言語処理研究会, 2006-SLP-64, pp.137-142, 2006年12月21日, 名古屋大学

6. 研究組織

(1) 研究代表者

西崎 博光 (HIROMITSU NISHIZAKI)
山梨大学・大学院医学工学総合研究部・助教
研究者番号: 40362082

(2) 研究分担者

該当なし

(3) 連携研究者

該当なし