

集音マイクと咽喉マイクを併用した受動的・能動的
会話情報収集システムの開発

メタデータ	言語: ja 出版者: 静岡大学 公開日: 2020-04-13 キーワード (Ja): キーワード (En): 作成者: 西田, 昌史 メールアドレス: 所属:
URL	http://hdl.handle.net/10297/00027317

令和元年6月14日現在

機関番号：13801

研究種目：基盤研究(C)（一般）

研究期間：2016～2018

課題番号：16K01543

研究課題名（和文）集音マイクと咽喉マイクを併用した受動的・能動的会話情報収集システムの開発

研究課題名（英文）Development of Passive and Active Conversation Information Collection System Using Close-Talk Microphone and Throat Microphone

研究代表者

西田 昌史（Nishida, Masafumi）

静岡大学・情報学部・准教授

研究者番号：80361442

交付決定額（研究期間全体）：（直接経費） 2,700,000円

研究成果の概要（和文）：本研究では、高齢者を対象とし音声による雑談を交えて必要な情報を収集するシステムを開発した。本システムのユーザ入力として非言語音響情報だけを利用してシステムの応答を選択することで、人間に近い応答が得られることがわかった。また、ユーザの発話欲求度に応じて話題誘導を行う手法を提案し、今の話題から近い話題を遷移する対話戦略に比べて、対話満足度の低下を抑えつつ、必要な情報を収集できることが明らかになった。さらに、属格ペアを用いた話題誘導手法を提案し、従来の1単語のみの話題誘導を行う手法に比べて、より自然な話題の遷移を実現することができた。

研究成果の学術的意義や社会的意義

介護負担を軽減するための手段として高齢者の話し相手となる雑談対話システムが期待されているが、従来の多くのシステムはテキスト入力を前提としていた。それに対して本研究では、音声で雑談しながら体調などの必要な情報を収集できるシステムを開発した。本システムにより情報を収集するだけでなく、音声で会話することで認知症の予防にも効果が期待され、社会的意義は大きい。また、これまで対話のどのタイミングでどの話題に誘導すべきかの検討が不十分であったが、本研究で提案した発話欲求度と属格ペアによる手法により、自然でユーザの満足度を低下させず情報を収集できることは、学術的意義が大きい。

研究成果の概要（英文）：In this study, we developed a system that gathers necessary information through spoken chatting for elderly people. The response close to the human could be obtained by selecting the response of the system using only nonverbal acoustic information as user input of this system. We proposed a method to control a topic according to the user's utterance desire. The proposed method can suppress the lowering of the dialogue satisfaction and collect necessary information compared to the dialogue strategy which shifts the topic close to the present topic. We also proposed a method to control a topic using pair of genitive cases. The proposed method realized a more natural transition of topics compared to the conventional method of topic induction with only one word.

研究分野：音声情報処理

キーワード：高齢者 音声対話 雑談 情報収集 非言語音響情報 話題誘導 発話欲求度 属格ペア

1. 研究開始当初の背景

高齢社会の進展に伴い、介護者不足が深刻な問題となっている。介護負担を軽減するための一つの手段として高齢者の話し相手となる雑談対話システムが期待されている。しかし、高齢者を対象とした多くの雑談対話システムはテキスト入力を前提としている。音声対話システムとの対話内容から高齢者の体調などに関する情報を収集できれば、センサーだけでは到底実現できないような高度な見守りを実現できる可能性がある。システムが欲しい情報に関する質問ばかりしていたのではユーザ満足度が低下してしまうため、話題を適切にコントロールし、目的の情報を収集する必要がある。特に高齢者との自然な対話を実現するためには、言語情報だけでなく、音声に含まれる非言語音響情報の活用が重要であると考えられ、非言語音響情報を活用した聞き役雑談対話システムについて検討が必要である。一方、高齢者向け対話システムは話を聞くだけの役割にとどまらず、高齢者の持つ潜在的な情報を引き出すことによって、その情報を介護の現場に活用できる可能性を持っている。これまで雑談の中で問診に必要な質問が出やすいように、対話の誘導を行うことでより多くの情報を収集することが試みられた。また、対話誘導の中でも話題に着目し、概念辞書を利用した自然な話題誘導をすることで同様の効果が示された。しかし、話題誘導の方法については検討がされているが、対話のどのタイミングでどの話題に誘導すべきかの検討はまだ十分になされていない。また、雑談という広域なドメインにおいて、ユーザは必ずしもシステムが収集した情報についての話題を話してくれるとは限らない。それに対して会話の中で現在話されている話題から目的の話題まで移動させる話題誘導が検討されている。これらの多くは一つの話題は一つの単語から構成されており、対話では一単語のみで話題誘導を行うことは難しい。

2. 研究の目的

本研究では、聞き役雑談対話システムのための非言語音響特徴量の有効性、非言語音響情報を利用した話題誘導を行う情報収集対話システムの開発、属格ペアを用いた雑談対話の話題誘導といった大きく分けて3つのテーマに基づいて研究を実施した。以下、これらの3つのテーマごとに詳細について述べる。

(1) 聞き役雑談対話システムのための非言語音響特徴量の有効性

本研究では、ユーザ入力として簡易的な非言語音響情報だけを利用した統計的対話システムを提案し、その選択応答の自然さについて評価を行い、人間が選択した応答に近い評価が得られるかどうか、また対話システムにおける非言語音響情報の可能性を明らかにする。

(2) 非言語音響情報を利用した話題誘導を行う情報収集対話システムの開発

発話に含まれる非言語音響情報と、適切な話題誘導のタイミングには何らかの関係性があると考えた。本研究では、対話中の話題と非言語音響情報の関係を分析し、情報収集を目的とした対話システムへの活用可能性を検討する。また、発話欲求度を利用した話題選択アルゴリズムを実際の対話システムに組み込み、ユーザ満足度を低下させずに効率的に情報収集ができるかどうか検討を行う。

(3) 属格ペアを用いた雑談対話の話題誘導

本研究では、日本語の属格構造を用いた2単語による話題誘導を提案し、1単語のみで行われる話題誘導と比較し、その有効性を示す。

3. 研究の方法

(1) 聞き役雑談対話システムのための非言語音響特徴量の有効性

ユーザの発話音声データから発話区間中の平均発話音量、発話長、有音率の3種類の簡易的な音響特徴を抽出し、これらに対話システムの入力とする。有音率とは発話区間中の無音時間に対する有音時間の比率である。システムでは、これら3種類の非言語音響情報を測定し、学習データから推定した平均値と標準偏差に基づいて閾値を設定し、大、中、小の3段階に分類する。

ユーザ発話の入力が完了した後、システムは現状態から次状態へと状態遷移を行う。システムが遷移できる状態としては話題転換、質問、傾聴、自己開示、共感の5状態である。状態遷移が完了したとき、システムは遷移先の状態に関連した発話を生成する。

システムとの対話のターンごとにユーザが応答の自然性を評価し、その評価値が高い対話パターンがより多く出現するような対話コーパスを生成し、それに対話モデルの学習データとすることでユーザが自然と感じる応答選択規則を学習させる。学習の流れとしては、まずターンごとにシステム応答の自然性に対する5段階評価値を記録する対話実験を学習者ごとに行う(ステップ1)。その後、複数の学習者による評価値の総和が高い対話パターンがより高い確率となるような対話パターンの確率分布を推定する(ステップ2)。最後に、この確率分布に従って対話コーパスを生成し、単一の対話モデルを再学習する(ステップ3)。ステップ3が終了するとステップ1に戻り、学習済みの対話モデルを用いて再度対話実験を行い、学習用対話コーパスを更新する。

(2) 非言語音響情報を利用した話題誘導を行う情報収集対話システムの開発

人間同士が雑談を行うとき、相手が現在の話題についてより多くのことを話したのであればその話題をより深く掘り下げ、逆に話したくなさそうであれば大きく別の話題に話題転換することで相手の興味を探りながら雑談を行っている。ここではこのユーザの現在の話題への話したさを発話欲求度と呼ぶ。もし対話システムにも同様の対話戦略が利用できるのであれば、相手の発話欲求が高ければ概念距離上近い話題を選び、相手の発話欲求が低ければ概念距離上遠い話題に話題を転換することで、よりユーザ満足度を下げずに対話を続けることができると考えられる。なお、話題間の概念距離はその話題の単語間の概念距離を wordnet に基づいて算出する。加えて概念距離上近い話題か、遠い話題かどちらに遷移したとしてもどちらもシステムが誘導したい話題に概念距離上近づけることができれば、ユーザ満足度を下げない話題誘導が可能となると考えられる。

また、システムが遷移することができるドメインは被験者である学生が答えやすいように食事、生活習慣、予定、運動の4つとした。それぞれのドメインには関連した複数の話題が紐づいている。さらに各ドメインには収集したい情報を聞く質問が紐づいており、それらは適切な中間発話を挟むことによって違和感なく質問することができるようになっている。収集する情報は実家の場所、体調、日付、年齢に関連した情報とした。今回、次に遷移する話題が現在の話題から近いのか、遠いのかについては現在のドメイン内の発話をするか、ドメインを切り替えるかのどちらかとした。よってシステムが行うことができるアクションは現在のドメインにとどまるか、別のドメインに切り替えるか、情報収集の質問をするかの3種類とした。

(3) 属格ペアを用いた雑談対話の話題誘導

属格ペアとは「AのB」のように2単語のどちらかがもう一方を所有するような単語のペアを指し、日本語では多くの場合後ろの単語は前の単語に所属し、2単語間は「の」という格助詞で接続される。属格ペアを話題として用いると、「AのB」のように2単語間をつなげることができるだけでなく、「AのC」のように前の単語がさらに所有している別の単語も関係性のある話題として違和感なく発話することができる可能性がある。その性質を利用して「AのB」→「AのC」→「DのC」のように話題を連鎖させることができ、目的の話題まで移動することができる。図1に食事から生活までの話題誘導例を示す。ここでは属格の所有している側(前の単語)を親単語、所有されている側(後の単語)を子単語と呼ぶ。

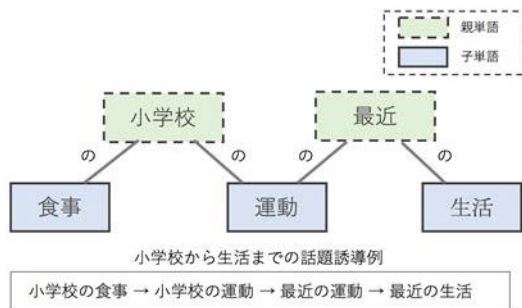


図1 食事から生活までの話題誘導例

4. 研究成果

(1) 聞き役雑談対話システムのための非言語音響特徴量の有効性

提案手法の有用性を示すために①その都度人間が判断した最適応答を返す Wizard 手法、②非言語音響情報から統計的対話制御を行った提案手法、③ランダムに応答を返すランダム手法、それぞれの手法に基づく3種類の対話システムを用意し、比較評価を行った。評価時には、客観的に対話の自然さを評価するために対話者と重複しない3名の評価者に対話の録音データを聞いてもらうことで評価を行った。

対話者6人の客観評価結果と6人の平均値を図2に示す。学習対話者及び学習外対話者それぞれ3名に対する平均評価値は多くの対話者において Wizard 手法、提案手法、ランダム手法の順に高い値となった。また、対話者6人の平均評価値を見ると、提案手法は人間が選択した Wizard 手法に対し1.1の低下にとどまった一方、ランダム手法に対し0.7の有意を示すことができた。よって、提案手法である簡易的な音響情報を用いた統計的対話システムは複数の学習者を用いた場合においても、ある程度自然と感じてもらえる対話制御が可能ということを示すことができた。



図2 対話者別の平均評価値

(2) 非言語音響情報を利用した話題誘導を行う情報収集対話システムの開発

発話欲求と話題の概念距離の関係を分析するために発話欲求に応じた話題転換を行う対話実験を Wizard of OZ 方式にて行った。対話中は、被験者は発話欲求度を 10 段階(0~100)でリアルタイムに入力する。Wizard はその発話欲求度に応じて 4 ターンごとに話題転換を行った。対話は 1 人の被験者に対し 10 分程度のものを次の 2 セッション行った。[セッション A] 話題転換時にユーザの発話欲求度が高い場合はより遠く的话题を選択し、発話欲求度が低い場合はより近く的话题を選択するケース、[セッション B] ユーザの発話欲求度が高い場合は近く的话题を選択し、発話欲求度が低い場合より遠く的话题を選択するケース。対話を行った後、被験者は対話満足度を 7 段階で評価した。なお、この実験を被験者 10 人に対して実施した。

話題が切り替わるタイミングにおける被験者の発話欲求度と、前後の話題間の概念距離との関係を図 3 に示す。セッション A では被験者の発話欲求度が高いほど Wizard は遠く的话题(概念距離がより 0 に近い)を選択するため、推定された回帰直線は右肩下がりとなった。一方、セッション B はその逆の戦略をとっているため右肩上がりとなっている。よってセッション A, B は話題選択の条件に則ったデータとなっていることが示された。しかし、回帰直線の傾きがあまり大きくないところから、wordnet の概念距離と人間の心理的尺度との間に開きがあることも確認した。

対話満足度の結果を図 4 に示す。セッション A より B のほうが対話満足度が 1.0 高くなっている。よって、ユーザの対話欲求が高い時に概念距離が近い話題を選択し、対話欲求が低い時に概念距離が遠い話題を選択することによってユーザ満足度が高くなることが分かった。

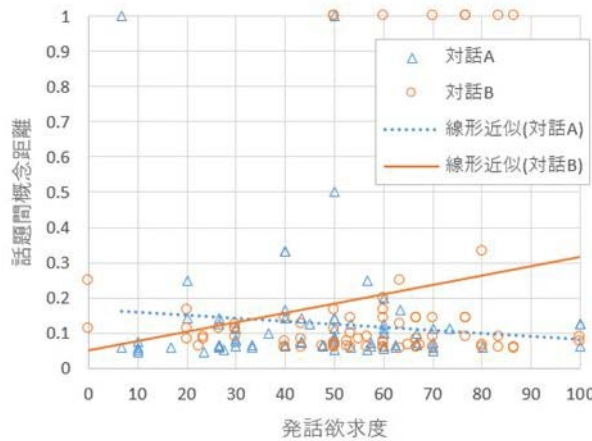


図3 発話欲求度と話題間概念距離の相関

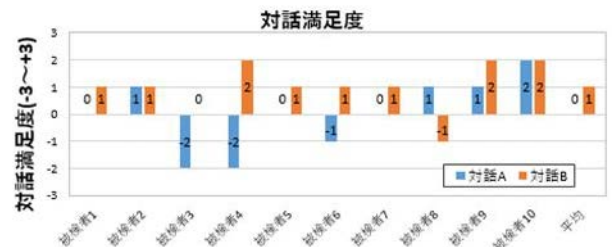


図4 対話満足度の結果

以上の結果から、現在の話題から遠い話題を選択してもユーザ満足度が低下しないのであれば、そのタイミングでシステムは取得したい情報の質問をすることができる。それ以外のタイミングでは現在の話題に近い話題を選択すれば、ユーザ満足度の低下を防ぎながら効率的な情報収集をすることができると思われる。

実験時のユーザの発話欲求度はキャリブレーションによって算出した平均発話長より高いか低いかで判定することとした。システムの有用性を示すために near, control, question の 3 種類のシステムと 12 ターン対話をしてもらった。near はドメイン内の話題を 1 つずつ発話していき、4 ターンに 1 度質問を行い、9 ターンに一度ドメインを切り替える。つまり、なるべく今の話題から近い話題に遷移することを目指して対話を行う。question システムはドメイン切り替えと質問を繰り返し、なるべく質問を多く行うような対話を行う。control システムは本研究の提案手法であり、推定された発話欲求度から話題誘導を行う。被験者にはランダムな順番でこれら 3 システムとの対話をしてもらい、1 システムとの対話が終わるたびに対話満足度と対話の自然さの評価値を 7 段階で記録してもらった。この比較実験を 20 代の男性被験者 6 名に対して行った。

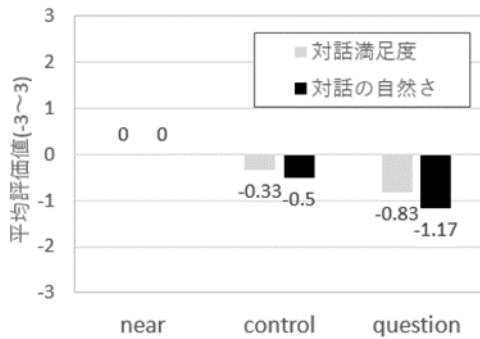


図5 実験システムごとのユーザの平均評価値

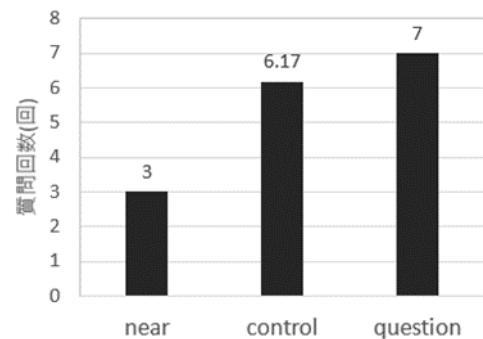


図6 実験システムごとの平均質問回数

名で実施した。

図5にユーザの評価結果を示す。対話満足度と対話の自然さのどちらの評価軸においても順序はnear > control > questionという結果となった。また、1セッション中にシステムが行うことができた質問の回数を図6に示す。質問の回数という評価軸においてはnear < control < questionという順序になった。本研究で目指すシステムはユーザ満足度をできるだけ下げずに情報収集をなるべく多く行うシステムである。提案手法であるcontrolシステムは、対話満足度をnearシステムから-0.33の低下にとどめつつ、questionシステムとほぼ同じ回数質問をすることができた。よって、提案手法は話題誘導を行いながら情報収集を行うシステムにおいて有用であることを示すことができた。

(3) 属格ペアを用いた雑談対話の話題誘導

属格ペアを用いた雑談対話システムの評価実験を行った。実験システムは抽出した親単語と子単語の類似度が高いパスを誘導パスとして利用した。ベースラインとして1単語のみの話題で話題誘導を行うシステムを用意した。ベースラインシステムはWord2Vecを用いて単語ベクトルが目的の話題との間にある単語を誘導パス上の話題として利用した。被験者5名に、話題誘導を行う2つのシステムとテキストチャットをしてもらい、話題の遷移の自然さを5段階で評価した。

実験結果を図7に示す。結果、提案手法は平均値でベースラインから0.8ほどの改善を得ることができた。一方、被験者1に関してはベースラインのほうが評価が高く、属格ペアを用いた話題誘導を自然とは思わない人もいたことがわかった。

以上の結果から、Wikipediaから自動抽出した属格ペアを用いた話題誘導によって、雑談対話のような広域ドメインにおいての自然な話題誘導を試みた結果、属格ペアによる話題誘導は1単語のみのシステムと比べ、より自然な話題の遷移を実現することができた。

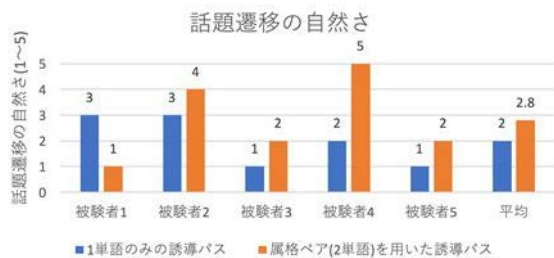


図7 話題誘導システムにおける比較実験結果

5. 主な発表論文等

[学会発表] (計24件)

- ① 鈴木 貴仁, 緒方 淳, 綱川 隆司, 西田 昌史, 西村 雅史, 咽喉マイクの大語彙音声認識のための知識蒸留, 情報処理学会第81回全国大会, 2019年
- ② T. Suzuki, J. Ogata, T. Tsunakawa, M. Nishida, M. Nishimura, Bottleneck Feature-mediated DNN-based Feature Mapping for Throat Microphone Speech Recognition, 10th Asia-Pacific Signal and Information Processing Association (APSIPA), 2018年
- ③ M. Abe, T. Tsunakawa, M. Nishida, M. Nishimura, Dialogue Breakdown Detection Based on Nonlinguistic Acoustic Information, IEEE 7th Global Conference on Consumer Electronics (GCCE), 2018年
- ④ 鈴木 貴仁, 緒方 淳, 綱川 隆司, 西田 昌史, 西村 雅史, 深層学習に基づく特徴量変換を用いた咽喉マイクの音声認識, 第123回音声言語情報処理研究会, 2018年

鈴木 貴仁, 緒方 淳, 綱川 隆司, 西田 昌史, 西村 雅史, 咽喉音による音声認識のための収録デバイスに関する検討, 日本音響学会 2018 年秋季研究発表会, 2018 年

R. Togai, M. Abe, T. Tsunakawa, M. Nishida, M. Nishimura, Analysis and Estimation of User's Motivation of Conversation by Using Acoustic Features, 15th International Workshop on Nonlinear Circuits, Communications, and Signal Processing (NCSP), 2018 年

S. Lin, T. Tsunakawa, M. Nishida, M. Nishimura, Conversational Speech Recognition Using Multiple Wearable Microphones, 15th International Workshop on Nonlinear Circuits, Communications, and Signal Processing (NCSP), 2018 年

阿部 元樹, 梅井 良太, 綱川 隆司, 西田 昌史, 西村 雅史, 言語情報と音響情報を併用した対話破綻分類器の構築, 情報処理学会第 80 回全国大会, 2018 年

大高 祥裕, 綱川 隆司, 西田 昌史, 西村 雅史, 咽喉マイクを用いた発話区間検出に基づく多人数会話音声認識, 情報処理学会第 80 回全国大会, 2018 年

鈴木 貴仁, 緒方 淳, 綱川 隆司, 西田 昌史, 西村 雅史, スペクトラム変換とボトルネック特徴量を用いた咽喉マイクの大語彙連続音声認識, 日本音響学会 2018 年春季研究発表会, 2018 年

阿部 元樹, 梅井 良太, 綱川 隆司, 西田 昌史, 西村 雅史, システム・ユーザ発話に着目した対話破綻検出, 第 120 回音声言語情報処理研究会, 2018 年

梅井 良太, 阿部 元樹, 綱川 隆司, 西田 昌史, 西村 雅史, 属格ペアを用いた雑談対話の話題誘導とその評価, 第 8 回対話システムシンポジウム, 2017 年

阿部 元樹, 梅井 良太, 狩野 芳伸, 綱川 隆司, 西田 昌史, 西村 雅史, 音響情報を利用した音声対話システムにおける破綻検出, 第 8 回対話システムシンポジウム, 2017 年

林 升柯, 綱川 隆司, 西田 昌史, 西村 雅史, 複数の装着型マイクを用いた多人数会話音声認識に関する検討, 日本音響学会 2017 年秋季研究発表会, 2017 年

梅井 良太, 阿部 元樹, 綱川 隆司, 西田 昌史, 西村 雅史, 非言語音響情報を利用した話題誘導を行う情報収集対話システム, 第 16 回情報科学技術フォーラム(FIT), 2017 年

阿部 元樹, 梅井 良太, 綱川 隆司, 西田 昌史, 西村 雅史, 個人差と対話行為を考慮した対話破綻検出に関する検討, 第 16 回情報科学技術フォーラム(FIT), 2017 年

S. Lin, T. Tsunakawa, M. Nishida, M. Nishimura, DNN-based Feature Transformation for Speech Recognition Using Throat Microphone, 9th Asia-Pacific Signal and Information Processing Association (APSIPA), 2017 年

Y. Otaka, T. Tsunakawa, M. Nishida, M. Nishimura, Voice Activity Detection Using Throat and Lavalier Microphones for Multi-Party Conversations, 14th International Workshop on Nonlinear Circuits, Communications, and Signal Processing (NCSP), 2017 年

阿部 元樹, 梅井 良太, 綱川 隆司, 西田 昌史, 西村 雅史, 非言語音響情報を利用した対話破綻検出に関する検討, 情報処理学会第 79 回全国大会, 2017 年

梅井 良太, 阿部 元樹, 綱川 隆司, 西田 昌史, 西村 雅史, 非言語音響情報を利用した話題誘導を行う対話システムの検討, 情報処理学会第 79 回全国大会, 2017 年

⑲ 阿部 元樹, 梅井 良太, 綱川 隆司, 西田 昌史, 西村 雅史, 対話破綻検出のための音響情報分析, 第 14 回情報学ワークショップ, 2016 年

⑳ 大高 祥裕, 綱川 隆司, 西田 昌史, 西村 雅史, 咽喉マイクとピンマイクの同時集音に基づく多人数会話における発話区間推定に関する研究, 電子情報通信学会音声研究会, 2016 年

㉑ 梅井 良太, 伊東 伸泰, 綱川 隆司, 西田 昌史, 西村 雅史, 聞き役雑談対話システムのための非言語音響特徴量に関する検討, 第 15 回情報科学技術フォーラム(FIT), 2016 年

㉒ 大高 祥裕, 西田 昌史, 綱川 隆司, 西村 雅史, 咽喉マイクとピンマイクの同時集音に基づく多人数会話における発話区間推定, 第 15 回情報科学技術フォーラム(FIT), 2016 年

6. 研究組織

(1) 研究分担者

研究分担者氏名: 西村 雅史

ローマ字氏名: Nishimura Masafumi

所属研究機関名: 静岡大学

部局名: 情報学部

職名: 教授

研究者番号(8桁): 60740363

(2) 研究協力者 なし

科研究による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。