様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

# 科学研究費助成事業　　研究成果報告書

平成　２７　年　　６　月　１２　日現在

機関番号：　１４７０１
研究種目：　基盤研究(B)
研究期間：　2012～2014
課題番号：　２４３０００７３

研究課題名（和文）聴覚の情報表現に基づく高度音声分析変換合成方式の研究

研究課題名（英文）Advanced speech analysis, modification and synthesis framework based on auditory information representations

研究代表者

河原　英紀（Kawahara, Hideki）

和歌山大学・システム工学部・教授

研究者番号：４０２９４３００

交付決定額（研究期間全体）：（直接経費）　　13,800,000円

研究成果の概要（和文）：音声には、文字で表されるテキスト情報に加え、話し手の個人性、感情、体調などの非言語情報や、声色やイントネーションを通じて意図などの表現を行うパラ言語情報が含まれている。さらに、悲鳴や叫び、声を用いた芸術などでは、従来の技術では分析不可能な物理属性に遭遇する。本課題では、これらの分析、変換、合成のための基盤となる音声信号処理の強力な方法が発明され実装・応用されるとともに、物理属性と心理的属性の関係を研究するユニークで強力な手段である音声モーフィングが、数理的に整理されて見通しが良く高い汎用性を有する方法として定式化されるなどの大きな成果が得られた。また、新たな枠組みの萌芽が得られている。

研究成果の概要（英文）：Speech consists of richer information than texts. They are non-linguistic information, such as individuality, emotional and physical states, and para-linguist information, which represents intention and/or expression using prosody. In cases of extreme voicing found in screaming, shout and artistic voice activities, conventional signal processing methods sometimes fail to handle unusual physical attributes found in such voices. New speech signal processing frameworks, which are capable of analyzing, modifying and synthesizing these speech sounds, were invented in this project. They are also implemented on tablets and used in various applications. The radically new formulation of voice morphing is a remarkable important achievement of this project, which provides unique and strong basis for investigating quantitative relations between physical attributes and psychological attributes in a mathematically sound manner. In addition, it emerged a prospective speech processing framework.

研究分野：　聴覚メディア処理

キーワード：　音声情報処理　音声分析　音声変換　標本化　テクスチャ　発声評価　音声訓練

１．研究開始当初の背景

処理の自由度の高い分析合成方式であるにも関わらず高い品質での変換と合成を可能にした STRAIGHT と、それに基づく音声モーフィング技術は、研究代表者らにより発明されたものであり、当時既に音声知覚や音声合成研究の分野における世界標準となっていた。さらに、2008 年に、STRAIGHT は数理的基盤が強化されて見通しが良くなるとともに効率化された TANDEM-STRAIGHT として、モーフィングは、2009 年に、外挿の場合にも破綻しない一般化された時変多属性モーフィングとして生まれ変わっている。

音声を文字以上のものとしている個人性、感情、年齢などを反映した非言語情報や、イントネーションにより意図を伝えるなどの機能を有するパラ言語情報、障害音声や悲鳴、叫び、強烈な歌唱表現による影響は、音声の物理属性を大きく変化させることがある。これらの研究に対して、TANDEM-STRAIGHT と音声モーフィングは、強力な基盤を提供する。しかし、特に人間に強い印象を与える音声の場合には、これらを含む既存の方法では、適切に分析することも処理することも困難な状況があった。また、音声モーフィングも、二つの音声資料の補間・補外に限られるなど、制約の多いものであった。

音声に含まれるテキスト情報の処理技術（音声認識、テキスト音声合成）が限定された範囲ではあるが実用レベルに近づきつつある状況を背景として、音声に含まれる非言語情報とパラ言語情報の処理技術が新たな挑戦を待つ課題として浮上していた。

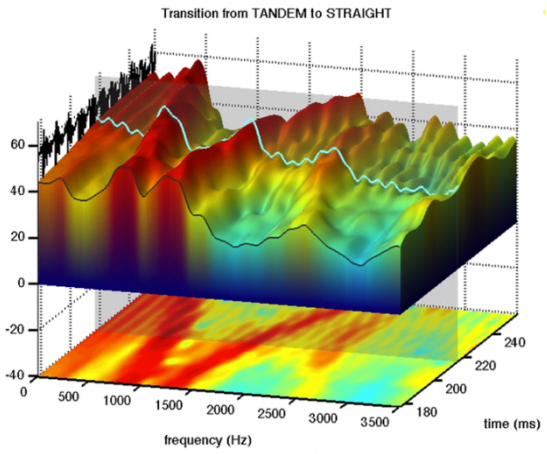


**図 1　TANDEM-STRAIGHT による表現**

２．研究の目的

この、音声に含まれる非言語情報とパラ言語情報に対する関心の高まりを背景として、それらを研究するための技術基盤の確立が本課題の提案時には、急務となっていた。本課題の目的は、世界最先端のレベルにある音声分析変換合成技術 STRAIGHT と音声モーフィングなどの周辺技術を、障害音声、歌唱音声、感情音声、悲鳴などの異常を知らせる音声など、当時の技術レベルでは十分な処理が困難な音声の処理にまで拡張し、音声を用いた人間と人間、人間と機械とのコミュニケーションに生ずる様々な困難を克服するための技術基盤を確立するとともに、それらの成果をツール群として適切な形で社会に還元することにある。

この目的を実現するために、具体的な目標として以下の 3 項目を設定した。(1)精密な音源情報の分析および合成技術の構築。(2)疎な標本からのスペクトル形状復元のための基底の構築法と頑健なパラメタ推定法。(3)PDA など、汎用の情報基盤上への支援環境のツール群の実装とそれらの提供。これらの達成目標の実現を軸として、実施にあたっては、早い段階からツールや基盤となる理論／技術の研究開発と、応用分野からのフィードバックのサイクルを回すように努めることとした。

３．研究の方法

これらの達成目標の実現に向けて、(1)基盤となる理論の構築とアルゴリズムの開発、(2)汎用情報基盤に適したアルゴリズムと実装技術の開発、(3)評価用データベースの構築および評価用応用アプリケーションの開発、(4)客観評価実験および主観評価実験の遂行、を担当するグループにより、相互の密接な連携の下に研究を進める。2008 年の発明により再構築された TANDEM-STRAIGHT とその周辺のツールを共通の研究基盤として用い、それらの構成要素を本プロジェクトでの研究成果に基づいて随時置き換えることで、効率の良いグループ間連携を可能とするとともに、研究の進展に応じて、逐次、成果を社会に還元するように努める。

(1) 基礎理論・アルゴリズムグループ
基本周波数推定法、スペクトルの疎な表現、非周期成分の表現と推定法に関する検討を進める。

(2) ツール・実装・アルゴリズムグループ
基礎理論・アルゴリズムグループのせいかを受けて、学術研究や教育・訓練などを支援するツールの開発、アプリケーション開発のためのライブラリ化、PDA および実時間アプリケーションの開発を進める。

(3) データベース・アプリケーション
基礎理論とアルゴリズムの開発に必要な音声資料を収録するとともに、応用の雛形となるアプリケーションを開発する。ただし、この項目は申請時からの減額により、最小限の規模に縮小する。

(4) 評価グループ
開発したアプリケーションなどの評価を、聴覚末梢系の最先端のモデルである動的圧縮型ガンマチャープフィルタバンクを用いた客観評価と、被験者を用いた主観評価により進める。この項目も、減額により最小限の規模に縮小する。

４．研究成果

当初の計画を大きく凌ぐ複数の予想外の成果が本研究により生み出された。これらは、当初計画の射程を拡大するものであり、新しい研究計画の提案につながっている。まず、それらを項目ごとに説明するとともに、最後に全体像について説明する。

(1) 時変多属性任意事例数モーフィング

TANDEM-STRAIGHT に基づく音声モーフィングは、音声の非言語情報（感情、個人性、表現、緊急性など）の研究ならびに応用のための強力な基盤である。目的とする非言語情報を体現している両極となる適切な音声資料を用意し、資料の個別の物理属性をモーフィングにより補間して作成した刺激を用いて主観評価実験を遂行するというパラダイムが、この技術により実行可能となった。このパラダイムを用いることにより、主観的で定性的な非言語情報と物理属性との関係を定量的に明らかにすることができる。しかし、これまでのモーフィングは、二つの事例間についてのみ定義される限定されたものであった。

このモーフィングが、研究協力者である Schweinberger 教授、研究員の Skuk 博士との協力を通じて、複数の事例間の一括処理を可能にした「時変多属性任意事例数モーフィング」という一般化された方法として、根本的に定式化しなおされたのである。この定式化は十分に抽象化されており、特定のアルゴリズム（今回の実装は TANDEM-STRAIGHT を用いている）以外の方法によるパラメタにも用いることができる応用可能性の高いものとなっている。さらに、この方法を使いやすいものとするように、支援用の GUI を有するツール群が開発された。

音声モーフィング技術の改良は当初計画で予定されていたものである。しかし、ここでの成果は、音声モーフィング技術を根本から書き換えるものであり、予想を超えた大きな発明である。しかも、高度に抽象化されているため、TANDEM-STRAIGHT に限定されることなく格段に広い範囲に応用することができるという、インパクトの大きな成果となっている。

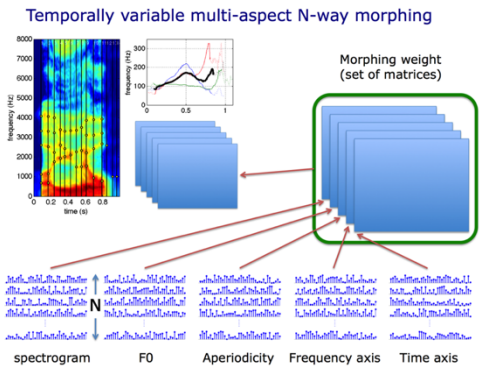


**図 2　拡張されたモーフィング**

(2) 時間分解能の高い音源情報の抽出法

目標とした、特異な音声の基本周波数を分析することのできる方法が発明された。強い感情を感じさせる発声や歌唱では、通常の音声の特徴である安定した周期性が大きく崩れている。具体的には、例えば複数の声帯振動がまとまりとなって、結果として基本周波数が高速に変動するという現象が存在する。本課題で発明された方法を用いることで、このような基本周波数の高速な変動や周期性の崩れを正確に追跡することができる。この方法は、簡単な波形処理に基づいており、高速で実時間処理にも適するという特徴を有する。TANDEM-STRAIGHT に限らず、応用範囲の広い、インパクトのある成果である。さらに、最終年度においては、統計モデルの導入により初期推定の頑健性を大きく改善し、さらに応用範囲を広いものとした。

実際、この方法の高速性を生かして、声から声道長を推定して対話的に即座にフィードバックするという応用プログラムが開発されている。

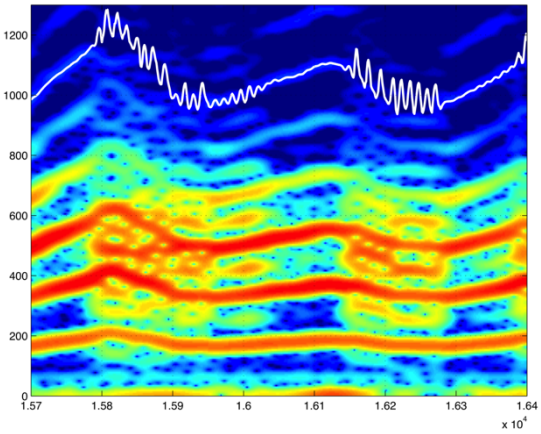


**図 3　基本周波数の高速変動（白線）**

(3) スペクトル情報の疎な表現

音声の発声機構の物理特性に基づいて、知覚的に重要なスペクトルピーク周辺の形状を回復する方法を導出した。この方法では、スペクトルの疎な表現として、LPC（線形予測子係数）、LSP（線スペクトル対）など数学的に等価なパラメタが求められ用いられる。この方法は、当初想定していた変換音声の品質改善だけではなく、声の明瞭度や好感度、声質の改善に応用できるとともに、音声モーフィングの自動化の基盤となるという、当初の計画では想定していなかったつながりが見えてきている。これは、本課題の後継プロジェクトにおいて検討すべき重要な可能性である。これは、当初予想を超える重要な成果と評価することができる。この可能性の一部を、後述する成果展開と社会還元につながる研究用ツールの一部として公開している。

(4) 音声パラメタの静的表現

この成果も、当初予想していなかったものであり、大きなインパクトを有する。パワースペクトル、瞬時周波数、群遅延は、信号の性質を表す基本的な物理量である。これら全て

に共通する新しい分析原理が見つかったのである。この原理は、最終年度での検討の過程で発見されたものであり、本課題の後継プロジェクトにおいて検討すべき重要な成果である。

周期的信号から求められるこれら上記のパラメタには、分析に用いる窓と分析対象である波形との相対的位置に依存する時間方向の変動と、周期性に依存する周波数方向の変動が含まれ、精密で安定な分析を阻害していた。STRAIGHT と TANDEM-STRAIGHT では、窓関数の巧妙な工夫と、最近の標本化理論の進歩を取り入れることにより、パワースペクトルのみについて、この問題を解決していた。この問題についての新しい解が、重要な音源情報である非周期成分の郡遅延に基づく表現の研究の過程で発見されたのである。

この原理を、以下の一文で表すことができる。「時間-周波数平面において、時間方向に基本周期の半分、周波数方向において基本周波数の半分の間隔を隔てて配置された４点におけるパラメタの値を、それぞれにおけるパワースペクトルの値を重みとした加重平均を求めることにより、周期的変動を実質的に取り除くことができる。」これが、新しい静的表現を求める原理である。

なお、この発見とは別に、パワースペクトルについて、Cheap Trick と名付けられた別の原理が発見され、オープンアクセス論文として公開された。これも実用的な効果が大きく、社会還元により大きなインパクトを与えることのできる成果である。

(5) タブレット環境への実装

申請時の計画から、採択時の減額により縮小していた本項目でも、大きな成果が得られた。STRAIGHT および TANDEM-STRAIGHT のライブラリ化が進められて、成果が公開された。ノート PC を含む計算機については、GPGPU の演算能力を生かすことで、実時間処理が可能な版が開発された。また、タブレットやスマートフォンなどのモバイル端末において使用できる版も作成され公開された。これは、当初計画を十分に満たす成果である。並行して、Cheap Trick に基づく WORLD という実装も公開された。このような実際に動作する実装を公開することは、有効な社会還元の方法の一つであり、大きなインパクトのある成果であるということができる。

(6) 応用の展開

本課題での基盤技術とアルゴリズムの開発と並行して、それらの応用が展開された。これらは計画で説明したサイクルの一環であるとともに、それ自身で有用なものであり、本課題の成果に挙げることができる。具体的には、様々な歌唱変換法、音声の明瞭度の改善法、実時間での声質変換法、喉頭摘出者の支援技術などが含まれる。

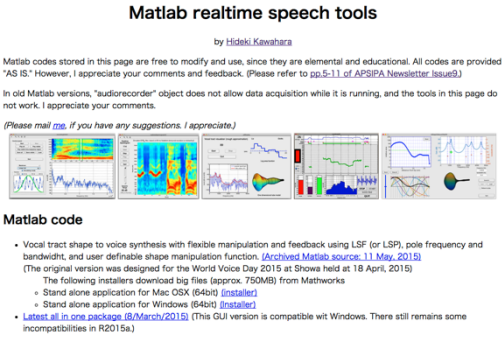


**図４　Matlab による実時間ツール群**

(7) ツール群の提供

TANDEM-STRAIGHT、拡張された音声モーフィングについて、利用を促進するための GUI に基づくツールを多数開発した。また、それらの効果を体験／理解してもらうためのデモ用のツールを併せて開発し、提供した。さらに、本課題の研究基盤として用いている科学技術計算環境である Matlab の入出力機能が実時間処理向けに大きく拡張されたことにより、当初の予定を超えて、多数のツール群を非常に効率良く生み出すことができた。これは、成果の社会還元に非常に効果的であり、外部状況の変化によるものではあるが、当初の予定を超えた大きなインパクトを有する成果となった。

以上を総合すると、本課題は、当初の目標を高い水準で満たすとともに、当初の計画では予定されていなかったインパクトの大きな成果を数多く生み出し、新たな可能性を拓くものであったということができる。

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　６件）


① Masanori Morise, Cheap Trick, a spectral envelope estimator for high-quality speech synthesis, Speech Communication, 査読有, Vol.67, 2015, pp.1-7.
DOI:10.1016/j.specom2014.09.003

② 河原英紀, 音声の実時間表示とモーフィングで探る声の多様性, 音声研究, 査読有, 18 巻, 3 号, 2014, pp.43-53
http://www.psj.gr.jp/jpn/publication/publication_vol18

③ Masanori Morise, Satoshi Tsuzuki, Hideki Banno, Seiji Ozawa, Muffled and Brisk Speech Evaluation with Criterion Based on Temporal Differentiation of Vocal Tract Area Function, IEICE TRANSACTIONS on Information and Systems, 査読有, Vol.E97-D, No.12, 2014, pp.3230-3233.
http://search.ieice.org/bin/summary.php?id=e97-d_12_3230

④ S.R. Schweinberger, H. Kawahara, A.

P. Simpson, V.G. Skuk, R. Zaeske, Speaker perception, WIRES Cognitive Science, 査読有, Vol.5, Issue-1, 2014, pp.15-25.
DOI: 10.1002/wcs.1261

⑤ Taiki Nishi, Ryuichi Nisimura, Toshio Irino, Hideki Kawahara, Controlling linguistic information and filtered sound identity for a new cross-synthesis vocoder, Acoustical Science and Technology, 査読有, Vol.34, No.4, 2013, pp.287-288.
DOI: 10.1250/ast.34.287

⑥ 中野 皓太, 森勢 将雅, 西浦 敬信, 山下 洋一, 基本周波数の転写に基づく実時間歌唱制御システムの実現を目的とした高品質ボコーダSTRAIGHTの高速化, 査読有, Vol.J95-A, No.7, 2012, pp.563-572.
http://search.ieice.org/bin/summary.php?id=j95-a_7_563&category=A&year=2012&lang=J&abst=

〔学会発表〕(計８２件)

① 鶴田 さくら,田中 宏,戸田 智基,ニュービッグ グラム,サクリアニ サクティ,中村 哲, 非可聴つぶやき強調音声の雑音環境下における明瞭性改善に関する検討, 日本音響学会 2015 年春季研究発表会, 東京（日本）, 2015 年 3 月 16-18 日.

② 小林 和弘,戸田 智基,ニュービッグ グラム,サクティ サクリアニ,中村 哲, 差分スペクトル補正に基づく歌声声質変換におけるパラメータ生成法に関する調査, 日本音響学会 2015 年春季研究発表会, 東京（日本）, 2015 年 3 月 16-18 日.

③ 牧野 奨平,坂野 秀樹,旭 健作, 声道断面積関数の変換による鼻声の声質改善手法に関する検討, 日本音響学会 2015 年春季研究発表会, 東京(日本), 2015 年 3 月 16-18 日.

④ 河原 英紀, 西村 竜一, 入野 俊夫, 高次対称性に基づく基本周波数推定法のモデル化と filled pause の分析への応用について, 電子情報通信学会音声研究会, 石垣市（沖縄）, 2015 年 3 月 3 日

⑤ Hideki Kawahara, Speech Analysis Modification and Synthesis tool STRAIGHT and extended voice morphing, Workshop for Auditory Research Software, ARO midwinter meeting, Baltimore(USA), 22 March, 2015.

⑥ Hideki Kawahara, Masanori Morise, Ken-Ichi Sakakibara, Tomoki Toda, Hideki Banno Ryuichi Nisimura, Toshio Irino, Excitation source design for high-quality speech manipulation systems based on a temporally static group delay representation of periodic signals, APSIPA ASC 2014, Siem Reap(Cambodia), 9-12 Dec. 2014.

⑦ Kou Tanaka, Tomoki Toda, Graham Neubig, Sakriani Sakti, Satoshi Nakamura, An Inter-Speaker Evaluation through Simulation of Electrolarynx Control Based on Statistical F0 Prediction, APSIPA ASC 2014, Siem Reap(Cambodia), 9-12 Dec. 2014.

⑧ Hideki Kawahara, STRAIGHT speech analysis, Tutorial of APSIPA ASC 2014, Siem Reap(Cambodia), 9-12 Dec. 2014.

⑨ 坂野 秀樹, 森勢 将雅, 河原 英紀, TANDEM-STRAIGHT の種々のデバイスへの実装と評価 ～スマートフォンから GPGPU まで～, 電子情報通信学会技術研究報告, vol. 114, no. 272, pp. 7-12, 白浜(和歌山), Oct. 23-24, 2014. (発表日 23 日)

⑩ Hideki Kawahara, Masanori Morise, Tomoki Toda, Hideki Banno Ryuichi Nisimura, Toshio Irino, Excitation source analysis for high-quality speech manipulation systems based on an interference-free representation of group delay with minimum phase response compensation, Interspeech 2014, Singapore(Singapore), 14-18, Sept. 2014.

⑪ 森勢 将雅,オープンソース音声合成システム WORLD の現状と課題, 情報処理学会音楽情報科学研究会（音声研究会連催), vol.2014-MUS-103, no. 68, pp. 1-6, 東京（日本), May 24-25, 2014. (発表日 25 日)

⑫ 河原 英紀, 溝渕 翔平, 森勢 将雅, 榊原 健一, 西村 竜一, 入野 俊夫, 非線形振動子による変調と近似時変フィルタに基づくグロウル系統の歌唱への実時間変換の定式化について, 情報処理学会音楽情報科学研究会, Vol. 2014-MUS-102, No.14, 東京（日本), Feb. 23-24, 2014. (発表日 23 日)

⑬ H. Kawahara, M. Morise, K. Sakakibara, Temporally fine F0 extractor applied for frequency modulation power spectral analysis of singing voices, MAVEBA 2013, Firenze(Italy), 16-18 Dec.2013, pp.125-128.

⑭ M. Sakaguchi, M. Kobayashi, R. Nisimura, T. Irino, H. Kawahara,

Spectrally estimated vocal tract lengths of singing voices and their contributing factors, MAVEBA 2013, Firenze(Italy), 16-18 Dec. pp.121-124, 2013.

⑮ Hideki Kawahara, Masanori Morise, Hideki Banno, and Verena G. Skuk, Temporally variable multi-aspect N-way morphing based on interference-free speech representations, APSIPA ASC 2013, Kaohsiung(Taiwan), OS.28-SLA.9, 31 Oct. 2013.

⑯ Yuri Nishigaki, Ken-Ichi Sakakibara, Masanori Morise, Ryuichi Nisimura, Toshio Irino and Hideki Kawahara, Controlling "shout" expression in a Japanese POP singing performance: analysis and suppression study, Interspeech2013, Lyon(France), pp.2905-2909, 2013. 28 Aug. 2013.

⑰ Hideki Kawahara, Masanori Morise, Tomoki Toda, Ryuichi Nisimura and Toshio Irino, Beyond bandlimited sampling of speech spectral envelope imposed by the harmonic structure of voiced sounds, Interspeech2013, Lyon(France), pp.34-38, 26 Aug. 2013.

⑱ Hideki Kawahara, Masanori Morise, Ken-Ichi, Sakakibara, Interference- free observation of temporal and spectral features in "shout" singing voices and their perceptual roles, SMAC-SMC 2013, Stockholm(Sweden), pp.256-263, 1 Aug. 2013.

⑲ Hideki Kawahara, Masanori Morise, Ryuichi Nisimura and Toshio Irino, Higher order waveform symmetry measure and its application to periodicity detectors for speech and singing with fine temporal resolution, ICASSP2013, Vancouver(Canada), 26-31 May, 2013, pp.6797-6801.

⑳ Tomoki Toda, Takashi Muramatsu, Hideki Banno, Implementation of Computationally Efficient Real-Time Voice Conversion, Interspeech2012, Portland(USA), 10 Sept. 2012.

〔図書〕(計　1件)

① Hideki Kawahara, (eds. Keikichi Hirose and Jianhua Tao), Speech Prosody in Speech Synthesis: Modeling and generation of prosody for high quality and flexible speech synthesis, Springer Berlin, 2015, 213

〔その他〕

ホームページ等

http://www.wakayama-u.ac.jp/~kawahara/STRAIGHTadv/index_j.html

http://www.wakayama-u.ac.jp/~kawahara/MatlabRealtimeSpeechTools/

http://ml.cs.yamanashi.ac.jp/world/

６．研究組織

(1)研究代表者

河原　英紀　　(KAWAHARA, Hideki)
和歌山大学・システム工学部・教授
研究者番号：　４０２９４３００

(2)研究分担者

入野　俊夫　　(IRINO, Toshio)
和歌山大学・システム工学部・教授
研究者番号：　２０３４６３３１

西村　竜一　　(NISIMURA, Ryuichi)
和歌山大学・システム工学部・助教
研究者番号：　００３７９６１１

戸田　智基　　(TODA, Tomoki)
奈良先端科学術大学院大学・
情報科学研究科・准教授
研究者番号：　９０４０３３２８

坂野　秀樹　　(BANNO, Hideki)
名城大学・理工学部・准教授
研究者番号：　２０３３５００３

榊原　健一　　(SAKAKIBARA, Ken-Ichi)
北海道医療大学・心理科学部・准教授
研究者番号：　８０３９６１６８

森勢　将雅　　(MORISE, Masanori)
山梨大学・大学院医学工学総合研究部・
助教
研究者番号：　６０５１００１３

(4)研究協力者

Stefan R. Schweinberger
Friedrich Schiller University of Jena
(独)・教授

Verena G. Skuk
Friedrich Schiller University of Jena
(独)・研究員