様式 C-19

科学研究費補助金研究成果報告書

平成２３年　３月３１日現在

機関番号：１４３０１
研究種目：特定領域研究
研究期間：　2006　～　2010
課題番号：１８０４９０４６
研究課題名（和文）
相互適応可能な実世界インタラクションのための計算モデル・システムの構築
研究課題名（英文）
Computational Models and Systems for Real World Human Machine Interaction
研究代表者
松山　隆司（MATSUYAMA TAKASHI）
京都大学・情報学研究科・教授
研究者番号：１０１０９０３５

研究成果の概要（和文）：２１世紀情報社会では、情報が量的に爆発するとともに質的な複雑さも増大し、人間と情報システムとの間に存在する様々なレベルのギャップが「デジタル・デバイドの普遍化」を招いている。本研究では、人間と情報システムとの自然なコミュニケーションによってこのギャップを解消することを目的とし、「息の合った」、「間合いの取れた」ヒューマン／マシン・インタラクションを実現するための計算モデルおよび、実世界インタラクションシステムの開発を行った。

研究成果の概要（英文）：Not only the quantitative explosion of information but also its qualitative increase in complexity widens gaps between information systems and human users in various levels. This trend eventually brings the universal spread of digital divides to every generation. The main goal of this research is to reduce those gaps by introducing natural communication between information systems and human users, and toward this goal, we have developed computational models for proper-timing coordination and interface systems for real-world human-machine interaction.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
２００６年度	13,300,000	0	13,300,000
２００７年度	14,300,000	0	14,300,000
２００８年度	18,200,000	0	18,200,000
２００９年度	18,200,000	0	18,200,000
２０１０年度	23,800,000	0	23,800,000
総　計	87,800,000	0	87,800,000

研究分野：　情報学
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：　実世界インタラクション、ハイブリッド・ダイナミカルシステム、マルチモーダル・インタラクション、プロアクティブ、対話・行動分析、視線・動作計測、相互適応

１．研究開始当初の背景

ソフトウェア・機器の多機能化、高機能化に伴い、情報システムの複雑性が爆発的に増大し、「デジタル・デバイドの普遍化」、すなわち、情報弱者だけでなく一般の人々にとっても情報システムの活用がむずかしくなるという現象が顕在化してきている。このことは、膨大かつ難解なマニュアルによって象徴されており、今後の情報社会の発展を妨げる１つの要因となる恐れがある。こうした問題を克服し、安心・快適な情報環境を構築するためには、人間と情報システムとの間に存在する様々なレベルのギャップを解消する必要があり、その解決手段として、人間との自然なコミュニケーションを通じて、適切な情報を適切なタイミングでユーザに提供でき

るような情報システムの実現が期待されている。

２．研究の目的

従来の多くのインタラクティブシステムでは、ユーザが命令・指示を与え、システムがそれに反応・応答するという「リアクティブなインタラクション」が用いられてきた。しかし、ユーザの興味、関心、意図といった、非明示的指示や無意識レベルでの心的状態の推定をリアクティブなシステムで実現するのは非常に困難である。したがって、システム側からユーザに対して能動的（プロアクティブ）な働きかけを行い、計測されたユーザの反応からその心的状態を推定するとともに、ユーザの漠然とした意図を顕在化させる必要がある。そしてそのためには、システム側からユーザに対する働きかけを適切に設計することで、システムとユーザの双方を同期・同調させるような、相互適応機能が重要となる。そこで本研究では、インタラクションのダイナミクスと、その相互適応的側面に注目し、「息の合った」、「間合いの取れた」プロアクティブ・インタラクションを実現するための計算モデルおよび、それに基づいた実世界インタラクションシステムの開発を目的とする。

３．研究の方法

研究目的で述べたような実世界インタラクションを実現するためには、人間同士のインタラクションにおいて、音声や視線、動作といったマルチモーダルなやり取りがいかに調整され、どのような機能を持つかを分析するとともに、分析から得られた知見を相互適応機能のモデル化やプロアクティブ・インタラクションの設計に活かすことが重要となる。そこで本研究では、以下で述べる４つの観点から研究を進めるとともに、それぞれの研究成果を有機的に連携させながら研究を展開した。

(1)マルチモーダル・インタラクションの高精度計測の実現：人間同士の自然なインタラクションを分析し、実世界インタラクションシステムを構築するためには、発話や動作などの各種信号を、人間を拘束することなく計測する必要がある。そこで本研究では、コンピュータビジョンや視聴覚情報統合の技術を発展させながら、人の立ち位置や顔位置、視線方向、話者の発話状態といった様々な情報を、非接触かつ高精度に推定する技術を開発する。

(2) マルチモーダル・インタラクションの分析：人間同士および人間と情報システムの間で行われる対話をはじめとしたインタラクションを、映像・音声・生体計測などの各種信号からアーカイブし、「発話タイミングや間の取り方」、「動的な提示情報に対する視線の反応特性と内的状態との関係性」など、インタラクションにおける人間のマルチモーダルな動的特性の分析を行う。

(3) プロアクティブ・インタラクションの設計：ユーザにおける心的／内的状態を、システム側からのプロアクティブな働きかけによって顕在化・推定するための、具体的な「プローブ」（状態を探るための手段）やインタラクションの枠組みを設計・考案するとともに、具体的な実世界インタラクションシステムとして、ユーザの心的状態推定に基づいて商品紹介などの情報提供を行う、大型ディスプレイ型の情報コンシェルジュシステムを開発する。

(4) インタラクションにおける相互適応機能のモデル化：ユーザとシステムとがインタラクションを行う際に、双方のダイナミクスが互いに同調されるような相互適応機能のモデル化を目指す。具体的な数理モデルとして、物理的現象記述に適した力学系モデル（連続的な計量空間における状態遷移を記述する微分方程式系）と、人間の心的・知的活動の記述に適した情報系モデル（離散的・順序的状態遷移を記述する記憶書き換え系）を統合した「ハイブリッド・ダイナミカルシステム」に基づくモデルを構築する。これにより、各種メディア信号間の複雑な時間的構造や、複数主体間が互いに相手に適応する際のタイミング調整の仕組みが表現可能になり、(1)の計測におけるマルチモーダル信号の統合手法や、(2)の分析の詳細な解釈、(3)の効果的なインタラクション設計につながると期待できる。

４．研究成果

(1)マルチモーダル・インタラクションの高精度計測の実現

①インタラクティブな情報提示システムのための非装着・非拘束な視線推定

大型ディスプレイを用いた情報提示システムにおいて、コンテンツをインタラクティブに制御するには、ユーザに計測機器を装着させることなく顔や視線の方向を推定し、ユーザの興味や反応を認識する必要がある。しかし、日常生活環境において、立ち位置や顔向きの変化に対してロバストな視線推定を精度良く行うには、カメラの高解像度化だけでは十分には対応できず、従来はユーザの立ち位置や顔向きを制限せざるを得なかった。これに対し本研究では、(i) カメラの首振り機構の導入による広範囲撮影、(ii) 視点の異なる複数台カメラの導入と推定結果の統合による多様な顔向きへの対応手法を提案した。

まず、情報提示システムに設置したカメラを、パン・チルト雲台によって制御すること

で自由な立ち位置での顔画像を取得し、虹彩検出を行うことでユーザの視線（ディスプレイ上の注目場所）をリアルタイムに推定する。この時、大型ディスプレイの周囲に設置したカメラでは顔向きの変化によってセルフオクルージョンが発生しやすいため、図１に示すように複数台のカメラを用い、各視点独立に処理した推定結果を統合することで、50 型ディスプレイから 1m 程度離れた立ち位置において、ディスプレイ上で平均 10cm 程度の誤差という推定精度を実現した。これは、カメラの高解像度化や詳細な眼球モデルの導入により、さらなる精度向上が可能である。

図 1　視線計測・情報提示システム

②視聴覚情報のタイミング構造に基づく話者検出および発話推定

複数人が発話する状況において話者検出や音声認識を行う技術は、マルチモーダル・インタラクションを実現する上で不可欠である。従来の話者検出手法としては、マイクロホンアレイを利用する方法があるが、音源定位の分解能以上に人物が近接している状況や、検出対象空間でのマイクロホン配置の制約によって、しばしば検出が困難となる。また、雑音環境下での発話に対して音声認識を行うには、音声認識の前段階で雑音抑制を行うことが重要となる。これらの状況・問題には、人の発話における口唇運動（口元の動き）と音声変化の間の共起性を利用する、視聴覚情報の統合手法が有効である。

従来の視聴覚情報の統合手法では、特徴抽出時のフレームを単位として、同一フレームや隣接フレームでの共起性や特徴量相関をモデル化する。ところが、/pa/などの破裂音の発声では口唇運動と音声の開始時刻がほぼ一致するのに対して、/a/などの母音では口唇運動の方が音声よりも先行するといったように、これらは必ずしも完全に同期するものではない。実際、人の知覚に許容される時間的なずれにも広がりがあることが知られている。そこで本研究ではこの点に着目し、「時区間ハイブリッド・ダイナミカルシステム」によって各メディア信号をモデル化することで、口唇運動と音声変化の間の系統的時間差を伴う時間的構造（タイミング構造）を、信号のダイナミクスが変化する分節点に基づき直接的に抽出し、分布として表現する手法を開発した。

図２に、この手法を映像中の話者検出に適用する際の流れを示す。実際の使用状況に合わせ、非発話者にも口の動きがある状況を設定して評価を行った結果、同一フレームもしくは隣接フレームの共起性に基づくモデルに比べ、提案手法ではより高精度に話者検出を行うことが可能であることが確認された。さらに、このモデルを非定常雑音環境下での音声推定に適用した結果、同一フレームの共起性に基づく手法に比べ、雑音抑制されたクリーン音声をより高精度に推定できることを確認した。本研究で提案したマルチメディア信号のタイミングモデルは、FIT ヤングリサーチャー賞を受賞するなど高い新規性を持つ。

本手法は、人物同士が非常に近接している場合や、カメラとマイクが各１台のみの状況でも適用可能であり、遠隔会議システムでの話者追跡による自動撮影だけでなく、アーカイブ化された映像コンテンツの分析などへの応用が期待できる。

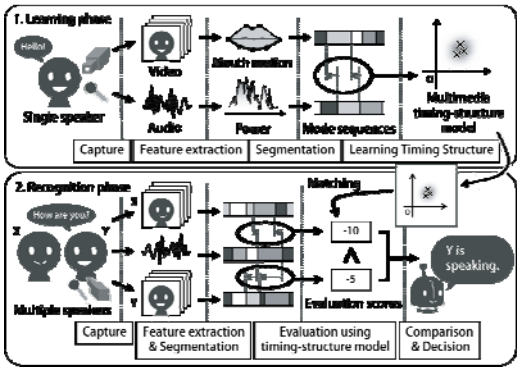

図 2　タイミング構造に基づく話者検出

(2)マルチモーダル・インタラクションの分析

①落語における視覚的「間合い」の解析

話者交替の間合いには、肯定や否定、同意や不同意など、発話の持つ意図によって異なる傾向が現れることが国内のいくつかの研究から明らかになっているが、さらに、聴覚だけでなく視覚モダリティも織り交ぜながら、自然な話者交替の間合いを表出している可能性がある。そこで、その具体例を探るために落語の映像解析を行った。実際の会話では、目的（議論、雑談など）、状況、話者の性格などのさまざまな要因が影響するため、会話分析によって特徴的な構造を見出すことがしばしば困難である。一方、落語においては (1)演者の動作やタイミングが観客を楽しませるために設計・最適化されている、(2) １人で複数の役柄を演じ分けながら会話情景を表現している、といったことから、視覚的な「間合い」に特徴的な構造が観察されることが期待できる。

特に落語では、頭部を左右にふりむける動作によって役柄交替が表現される。そこで、先行役柄の発話終了時刻からこの頭部動作開始までの間合いが、同じく寄席演芸である漫才の二者間会話の間合いと比べ、どのような類似性があるかを調べた。その結果、落語では漫才における後続話者の発話と同様に「負の間合い」、すなわち先行発話に対してオーバラップする頭部動作が積極的に用いられていることが確認された。また、現段階の分析データ数は十分ではないが、この間合いと発話内容（同意、非同意など）との関係を調べたところ、特に同意（肯定）応答の場合に、オーバラップの割合がより高まる傾向がみられた。この結果は、話者交替をはじめとする「間合い」の表出が視覚や聴覚などのモダリティによらないことを示唆しており、実世界インタラクションシステムを設計する際の指針となる。

② 対話相手の心的状態を顕在化する働きかけとその効果の分析

対話相手の意図や興味といった心的状態を探るために、人は相手に対して働きかけを行って反応を観察している。働きかけは音声による場合が多いが、対面対話においては視覚的手段も有効に利用されていると考えられる。そこで、働きかけの発話に頻繁に付随する相手への顔向け（視線の投げかけ）が、相手の心的状態を探る上でどのような役割を持つかについて分析した。このとき、対話相手の心的状態を綿密に理解する必要があり、働きかけと応答が高い頻度で出現するような、合意形成対話を分析対象とした。

図3は、働きかけ発話の終了から応答発話の開始までの時間間隔の頻度を示す。従来知見や(2)-①の落語の分析でみられた、不同意の応答タイミングは同意の応答タイミングに比べて遅くなるという傾向が、発話とともに顔向け動作を伴うような働きかけを行うことでさらに強まり、不同意の場合のタイミングが、同意の場合に比べ有意に遅くなることが明らかになった。

この知見は、ヒューマンインタフェース学会論文賞を受賞するなど学術的に高い評価を受けた。さらに、実際のインタフェースシステム設計という観点からも様々な適用方法がある。たとえば、システムが情報提示や質問などの働きかけをユーザに対して行う際に、顔向けのような「ユーザをみる行為」を共に用いることで、ユーザの反応をより強めることが期待できる。つまり、ユーザの興味や意図を、システムがより正確に推定する際のプローブとして利用することができると考えられる。

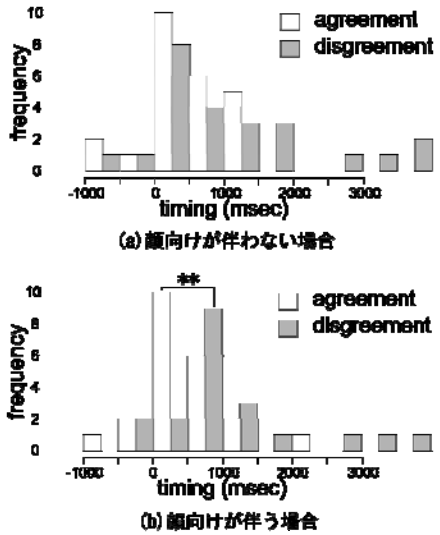

図 3 発話終了から応答発話開始まで時間間隔の頻度

(3)プロアクティブ・インタラクションの設計

システムが提示した複数の情報に対してユーザが好みのものを選択するシナリオにおいて、ユーザの心的／内的状態をシステムがプロアクティブに探るためのインタラクションを設計した。具体的には、(1)-①で述べた大型ディスプレイに選択肢（商品等）を提示し、これをユーザが閲覧する際に、システム側からの適切な働きかけを設計することにより、ユーザの注視状態や興味といった内的状態を推定する。以下では、このそれぞれの推定方法について述べる。

①Gaze Probing：イベント提示に基づく注視オブジェクト推定

情報提示システムにおいてディスプレイ上に提示されたオブジェクトのうち、ユーザが実際にどのオブジェクトを注視しているかという情報は、ユーザの興味や関心を知る上で重要な手掛かりである。しかしながら、その計測精度はユーザ側の自由度とのトレードオフであるため、計測された注視座標をそのまま用いるだけでは、ディスプレイ上に提示できるオブジェクト数を少なくしなければ、ユーザの注視オブジェクトを推定することが困難となる。そこで本研究では、提示するコンテンツに移動、停止といった特徴的な動き（イベント）を持たせ、この動きを「プローブ」として用いることで、注視対象と同調した眼球運動を引き出し、そのタイミングを分析することによって注視対象を推定する新たな手法「Gaze Probing」を考案した。

Gaze Probing では、計測誤差の影響を受けやすい眼球運動の変位パターンを直接用いるのではなく、眼球運動の反応時刻を抽出し、各コンテンツのイベント発生時刻との同期の程度を分析する。このように「ユーザがどの時刻のイベントに反応したか」という時間的な情報を用いることで、視線計測の精度を十分に望めない（ユーザの自由度の高い）状

況でも、従来手法に比べ、ユーザの注視オブジェクトを高精度に推定することが可能となる（図 4）。本提案手法の新規性は、パターン認識に関する国際会議の IBM Best Student Paper を受賞するとともに国内外の特許出願につながるなど高く評価された。

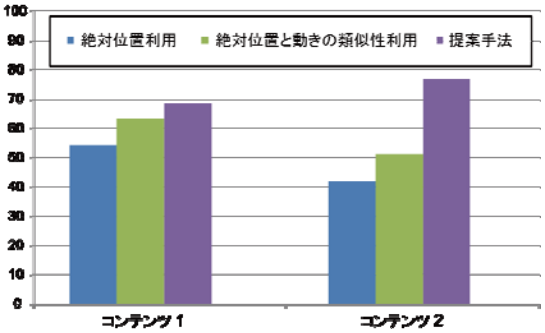

図 4　注視対象推定精度の比較（各々右端が提案手法）

②Mind Probing：能動的な働きかけと反応観察によるユーザの興味推定

ユーザの注視状態だけでなく、さらにその興味を推定する枠組みとして「Mind Probing」という手法を提案した。空間に並べられた様々な視覚情報からユーザが好みのものを選択する状況は、(i)全体をざっと見渡す状態と、(ii)いくつかの情報を吟味する状態とに分けられる。吟味状態では、興味を持った情報に対して能動的に注意を向けることが多く、この時に現れる内因性サッカード（急速な動きを持つ眼球運動）は、ユーザの興味を強く反映していると考えられる。しかし、ユーザの振る舞いを受動的に観察するだけでは、吟味状態であることの認識も、内因性サッカードと外因性サッカード（外部刺激に対する反応）の判別も困難である。そこで、これらを分離するためのプローブとして以下の情報提示方法を設計した。

複数のページから構成されるコンテンツを複数提示する状況を想定する。システムはまず、(i)コンテンツを 1 つ表示し、短い時間間隔でそのページを切り替え、続いて、ディスプレイの別の位置に他のコンテンツを同様に表示する。これを全てのコンテンツについて繰り返すことで、ユーザは一通りの情報を把握することができ、続くフェーズで吟味状態が生じやすくなることが期待できる。次に、(ii)これまでに提示したコンテンツを同じ場所に再び表示する。ただし、興味のあるコンテンツをユーザが十分に読み取ることができるように、ページ切り替えの時間間隔を長く設定する。これによって、サッカードは主に内因的に表出されることが期待できる。また、コンテンツ間のページ切り替えの順番は、ユーザに予想させないためにランダムとする。

このように提示されるコンテンツから被験者が 1 つを選択する実験を行ったところ、興味があるコンテンツを注視している場合に、他のコンテンツのページ切り替えへの反応が遅れる傾向があることを確かめた。この反応遅延（図 5）に基づいて、最も興味のあるコンテンツを推定したところ、従来研究で用いられている注視時間と注視頻度を用いる方法より高い精度が得られた（表 1）。

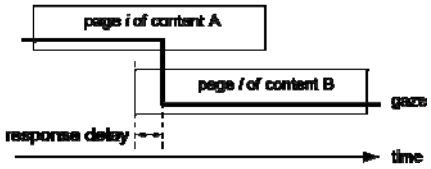

図 5　コンテンツ提示に対するサッカードの反応遅延

表 1　興味があるコンテンツの推定精度

Response delay	*Gaze duration*	*Gaze Freq.*
55.0%	35.0%	20.0%

(4)インタラクションにおける相互適応機能のモデル化

システムがユーザと「息の合った」会話を行うためには、ユーザの発話開始や終了タイミングを推定することによって自然な話者交替を実現する必要がある。しかし、実際には自然な話者交替をどのようにモデル化でき、「話がはずむ」、「息が合う」といった状態を、数理モデルを通じてどのように定量化できるかについて、国内外でも十分な知見がない。そこで、本研究では 2 者間対話に焦点を当て、対話の参加者それぞれを、発話状態と非発話状態とを切り替えることのできるハイブリッド・ダイナミカルシステム（HDS）として表現し、これらを相互結合することで、話者交替のモデルを構築することを試みた。

それぞれの HDS は発話／非発話状態を持ち、双方が自分の内部状態（興味）と相手の内部状態（推定分布）とのギャップに基づいて、発話／非発話状態の切り替えタイミングを調整するような相互適応機能を持たせた。その結果、相互結合された HDS は、自律的に話者交替を行うモデルとなっており、双方の同調度パラメータの設定によって発話頻度が変化するなど、人間同士の対話における話者交替や発話頻度と類似した特性をシミュレートできることが確認された。

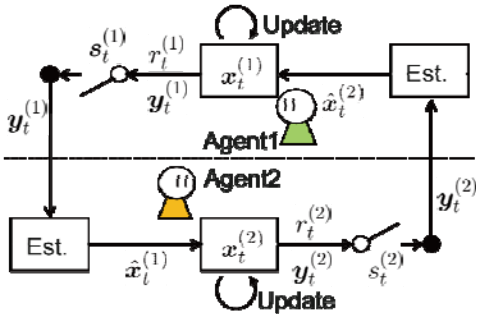

図 6　相互結合された主体間のインタラクション

(5) 以上の(1)から(4)の研究を通じて得られた成果は、商品紹介や観光案内、分析情報紹介などの様々な自律型情報提示システムの開発につながっており、本研究課題により、プロアクティブ・インタラクションに基づく実世界システムを開発する上での理論的・技術的基盤が構築された。

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１８件）
① T. Hirayama、J. B. Dodane、H. Kawashima、T. Matsuyama、Estimates of User Interest Using Timing Structures between Proactive Content-Display Updates and Eye Movements、IEICE Transactions on Information and Systems、査読有、Vol.E93-D、No.6、2010、pp.1470－1478
② 平山高嗣、大西哲朗、朴恵宣、松山隆司、対話における顔向けを伴う働きかけが同意・不同意応答のタイミングに及ぼす影響、ヒューマンインタフェース学会論文誌、査読有、Vol.10、No.4、2008、pp.385－394
③ 川嶋宏彰、松山隆司、時区間ハイブリッドダイナミカルシステムを用いたマルチメディア・タイミング構造のモデル化、情報処理学会論文誌、査読有、Vol.48、No.12、2007、pp.3680－3691
④ 平山高嗣、川嶋宏彰、西山正紘、松山隆司、表情譜:顔パーツ間のタイミング構造に基づく表情の記述、ヒューマンインタフェース学会論文誌、査読有、Vol.9、No.2、2007、pp.201－211
⑤ 川嶋宏彰、西川猛司、松山隆司、落語の役柄交替における視覚的「間合い」の解析、情報処理学会論文誌、査読有、Vol.48、No.12、2007、pp.3715－3728

〔学会発表〕（計４０件）
① H. Kawashima 、Speech Estimation in Non-Stationary Noise Environments Using Timing Structures Between Mouth Movements and Sound Signals 、 Interspeech2010 、 2010.9.27、Makuhari Japan
② R. Yonetani、Gaze Probing: Event-Based Estimation of Objects Being Focused On、The 20th International Conference on Pattern Recognition、2010.8.23、Istanbul Turkey
③ 平山高嗣、Gaze Mirroring: ユーザの興味を顕在化させるための注視模倣、電子情報通信学会技術報告（HCS)、2009.3.25、松江
④ H. Kawashima 、 Visual Filler: Facilitating Smooth Turn-Taking in Video Conferencing with Transmission Delay、ACM CHI 2008、2008.4.10、Florence Italy
⑤ 水口充、Mind Probing: システムの積極的な働きかけによる視線パタンからの興味推定、情報処理学会研究報告（HCI）、2007.9.28、大阪

〔産業財産権〕
○出願状況（計３件）
名称：注視対象判定装置及び注視対象判定方法
発明者：坂田幸太郎、前田茂則、米谷竜、平山高嗣、川嶋宏彰、松山隆司
権利者：パナソニック株式会社
種類：特許
番号：特願 2009-137647
出願年月日：2009.6.8
国内外の別：国内

〔その他〕
ホームページ等
http://vision.kuee.kyoto-u.ac.jp/

６．研究組織
(1)研究代表者
松山　隆司（MATSUYAMA TAKASHI）
京都大学・情報学研究科・教授
研究者番号：１０１０９０３５

(2)研究分担者
杉本　晃宏（SUGIMOTO AKIHIRO）
国立情報学研究所・知能システム研究系・教授
研究者番号：３０３１４２５６
（H18）
佐藤　洋一（SATO YOICHI）
東京大学・生産技術研究所・助教授
研究者番号：７０３０２６２７
（H18）
牧　淳人（MAKI ATSUTO）
京都大学・情報学研究科・助教授
研究者番号：６０３６２４１４
（H18→H19）
川嶋　宏彰（KAWASHIMA HIROAKI）
京都大学・情報学研究科・講師
研究者番号：４０３４６１０１
（H18→H21、H22：研究協力者）

(3)連携研究者
なし

(4)研究協力者
平山　高嗣（HIRAYAMA TAKATSUGU）
京都大学・情報学研究科・特任助教
研究者番号：１０４２３０２１
加藤　丈和（KATO TAKEKAZU）
京都大学・情報学研究科・特任研究員
研究者番号：３０３６２８５９