

多人数・マルチモーダルインタラクション研究 のためのプラットフォーム構築

坊農真弓[†] 角康之^{††} 高梨克也^{†††}
岡田将吾^{††††} 菊地浩平[†] 東山英治^{†/*}

本研究発表では、多人数・マルチモーダル研究に対する我々のアプローチを紹介する。具体的には、日常的インタラクションデータの収録は「技術」「理論」「フィールド」の3つの側面を検討することが重要であることを指摘し、多人数・マルチモーダルインタラクション研究のためのプラットフォーム構築に向けた5～6時間程度の公開可能なサンプルデータ収録の詳細を報告する。

Establishing a Platform for Analysis of Multiparty and Multimodal Interaction

Mayumi Bono[†] Yasuyuki Sumi^{††} Katsuya Takanashi^{†††}
Shogo Okada^{††††} Kouhei Kikuchi[†] and Eiji Toyama^{†/*}

In this presentation, we introduce our approach toward multiparty and multimodal interaction studies. In particular, we point out the importance that we focus on three aspects, 'technology', 'theory' and 'real field', to collect naturally occurring interaction data, and report the details of collecting 5-6 hours sample data for establishing a platform for analysis of multiparty and multimodal interaction.

1. はじめに

我々の日常は、家庭内や友人間といったプライベートなインタラクションのみならず、老人医療施設や福祉施設といった医療福祉現場、カウンセリングや特別支援学級といった学校教育現場などの様々な機関が関わるインタラクションによって支えられている。少子高齢化に対する打開策や身体障害者の社会参画支援等は、多くの分野にとって緊急に対応すべき課題であり、実際にどのような解決すべき問題が存在しているのかは、日常的インタラクションデータの収録とその分析から詳らかにされていくべきである。つまり、情報機器に取り囲まれた我々の日常生活の現状やニーズに即したインタフェースデザインを議論するためには、実際のインタラクションデータの収録とその分析が必要不可欠であると考えられる。

本研究プロジェクト[a]の目標は、日常的インタラクションデータを収録する際の個人情報保護や将来的なデータ共有に向けた手続きといった諸問題を検討するための研究プラットフォームを構築することである。

近年、情報学を始めとした人と人とのコミュニケーションの理解に感心のある分野では、人の日常生活における様々なやり取りの情報を画像データや音声データといった形式で集積してアノテーションを付与し、その中から知識伝達の手法を探し出すといったプロジェクトが開始されつつある。しかしながら、そういったプロジェクトの多くは、人々の生き生きとしたやり取りを収録するために、常に個人情報保護やプライバシーの侵害の問題を回避するための工夫をデータ収録デザインに施す必要があり、そうした作業に多くの時間が割かれてきた。こういった背景は、データ収録を主とした様々な研究プロジェクトに対し、研究成果発表を遅らせ、またインタラクションデータの豊かさや多様性を削がせる結果を導く可能性がある。またこれまで日常的インタラクションデータの収録を中心課題としてこなかったメディア処理技術(例:データマイニング、ユビキタスコンピューティング)に関する研究者もここ10年、データ収録を重視するようになっており、個人情報保護やプライバシーの侵害の回避について議論が重ねられている。データ収録における個人情報保護とプライバシーの

[†] 国立情報学研究所
National Institute of Informatics

^{††} 公立はこだて未来大学
Future University Hakodate

^{†††} 科学技術振興機構さきがけ/京都大学
PRESTO, Japan Science and Technology Agency (JST)/Kyoto University

^{††††} 東京工業大学
Tokyo Institute of Technology

* 千葉大学
Chiba University

a) H21-H23年度、国立情報学研究所グラウンドチャレンジ、「情報環境を支える日常的インタラクションデータ収録のためのプラットフォーム構築」(研究代表者:坊農真弓)

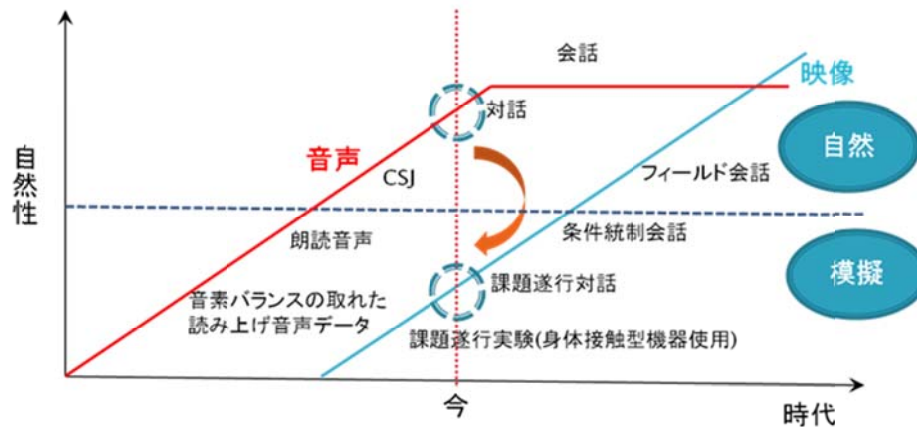


図 1 データのマルチモーダル化と自然性

侵害の問題と、日常的インタラクションデータの豊かさと多様性の維持は、今後統一的に検討されるべき課題である。

図 1 はデータのマルチモーダル化と自然性の関係を示したものである。従来、音声データは様々な研究目的で収録されてきたが、取扱いが容易になってきた映像データもインタラクション研究の素材として近年用いられるようになってきた。音声データはこれまで、より自然性の高い収録が目指されてきた。一方、映像データについても、音声データと比べると遅れを取っているが自然性の高い収録が目指されている。しかしながら、自然性の高いデータには、データの中に存在する個人情報の保護とプライバシーの侵害の問題が関わり、自然に起こった会話データよりも個人情報あまり含まれない模擬会話・実験会話の収録に踏みとどまるケースも多く見られる。本研究発表で提示する公開可能なサンプルデータ(多人数・マルチモーダルインタラクションコーパス)も同様の問題から実験会話である(3 節に詳述)。

2. 研究の位置づけ

2.1 アプローチ

上述した難問を解くため、本研究プロジェクトでは、これまで個々のプロジェクトが独自の手法で回避してきた個人情報に関する問題に対する統一的手法を確保した上で、日常的なインタラクションデータを収録・集積するためのモデルケースとなる手法を提示することを試みる。そして、それを広く公開することで日常的インタラクシ

ョンデータ収録・集積のためのガイドラインを整備する予定である(見込み、2012 年 3 月)。

本研究プロジェクトの戦略は、改めて大規模データを収録せず、これまでに国内外で収録されたデータの再検討を通し、そこに存在する問題を洗い出すことにある。近年、様々な研究プロジェクトがデータの収録を進めているが、分析にかかる時間を考えるとデータの量はこれ以上むやみやたらに増やすべきではなく、個人情報の保護等の問題が解決された質的に価値のあるデータが必要である。

具体的に本研究のアプローチは、(1)従来研究の調査、(2)従来研究に対する議論、(3)公開可能なサンプルデータの収録の 3 つからなる。本研究発表では主に(3)について手続きと現状を報告する。

2.2 研究の柱—「技術・理論・フィールド」

多人数・マルチモーダルインタラクション研究が主に扱っている日常的インタラクションデータは、図 2 に示したデータ収録に関わる 3 つの側面を検討することが重要である。

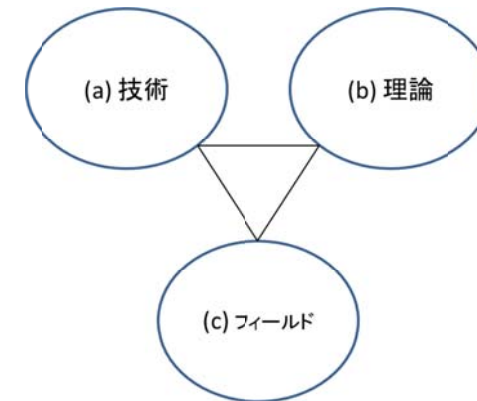


図 2 研究の柱

(a) 技術

データの収録ためには、現在利用可能な多種多様な機器の中から、多人数・マルチモーダルインタラクション研究に適した機器を選定する必要がある。当然、日常的インタラクションの流れを壊すことなくデータを収録することができるということは言うまでもないが、それに加え、インタラクション参加者によってインタラクシ

上で利用されている情報を収録可能な機器が有効であろう。またデータ収録後は、コンピュータ上での複数のデータを並列的に処理可能な環境の整備が必要になる[1].

(b) 理論

多人数・マルチモーダルインタラクション研究の理論的側面については、これまで言語学、社会学、人類学といった人文科学系分野で議論されてきた[2]. 多人数・マルチモーダルインタラクション研究は、最新のインタラクション収録・分析技術を利用して、どういった従来理論にアプローチするのかを明確にすることが重要である。

(c) フィールド

従来の言語・非言語コミュニケーション研究の多くは、統制が比較的取りやすい実験室環境でデータ収録をすることが多かった。それに対し、現在の多人数・マルチモーダルインタラクション研究は、ビデオカメラやマイクロフォンといった基本的なデータ収録機器の小型化・高性能化によって、日常的なインタラクション場面、いわゆる「フィールド」へ収録対象を移してきている。

図3は、医療福祉フィールドと学校教育フィールドを例にして、「技術」「理論」「フィールド」の関係を示したものである。技術と理論研究はすでに多くの蓄積があり、シーズが集まっている。一方、フィールドには情報技術やインタラクション理論を適用することで解決する可能性のある事象が存在し、ニーズが見え始めている。技術と理論を中心に据える研究ではシーズ出発型の研究、フィールドを中心に据える研究ではニーズ出発型の研究が進められている。

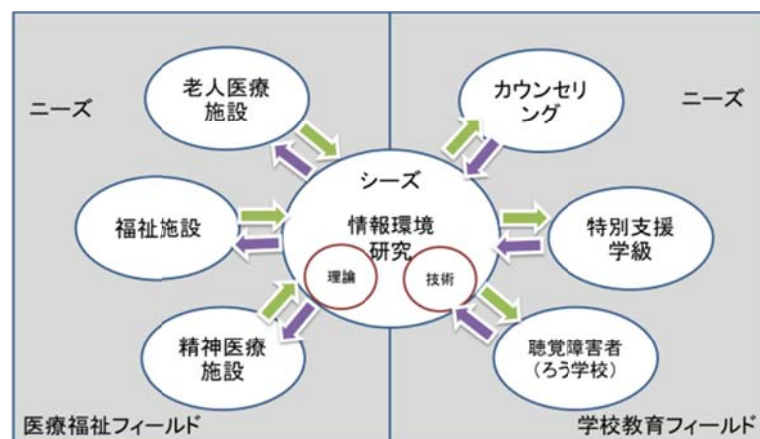


図3 フィールドの位置づけ

2.3 研究協力

2.1節で述べた(1)と(2)を進めるために、具体的に本研究プロジェクトでは、独自の視点でインタラクション研究を進めている研究者に声を掛け、多人数・マルチモーダルインタラクション研究の知見を集合させる場、いわゆる「プラットフォーム」を国立情報学研究所に構築している。

(1) 研究協力者(H23.4 時点) [b].

【NII】坊農真弓, 古山宣洋, 板橋秀一, 大須賀智子, 石本 祐一, 菊地浩平

【NII 客員教員】菊池英明(早稲田大学), 井上雅史(山形大学)

【その他】高梨克也(京都大学), 伝康晴(千葉大学), 榎本美香(東京工科大学), 細馬宏通(滋賀県立大学), 花田里欧子(京都教育大学), 市川薫(早稲田大学), 角康之(公立ほだて未来大学), 堀内靖雄(千葉大学), 小磯花絵(国立国語研究所), 前川喜久雄(国立国語研究所), 山川仁子(愛知淑徳大学), 長嶋祐二(工学院大学), 岡田将吾(東京工業大学)

(2) 協力研究プロジェクト(一部)(H23.4 時点)

研究協力者が個別に進めている研究プロジェクトの一部を本原稿末尾の付録に示した。それぞれの研究プロジェクトの研究代表者もしくは研究分担者が本研究プロジェクトに関わっている。各研究プロジェクト名の末尾に2.2節で挙げた技術・理論・フィールドのうちどれにより焦点を絞った研究を進めているかを参考までに記載した。

2.4 研究の進め方

2.1節で述べた(3)を進めるために、本研究プロジェクトでは5~6時間程度の公開可能なサンプルデータの収録し、それらに時間同期やアノテーションの処理を施し、コーパス作成を試みた。図4は研究の進め方とそこで議論となる項目の一部を示している。(a)から(c)は2.2節で挙げた技術・理論・フィールドとそれぞれ関連している。

3. 多人数・マルチモーダルインタラクションコーパスの作成

3.1 データ収録

コーパス作成のためのデータ収録はこれまで2度行った。一つは京都大学 IMADE ルーム[3]で、もう一つは NII スマートルームでそれぞれ実施した(以下、前者を「IMADE データ」、後者を「NII データ」と呼ぶ)。IMADE データは2010年7月14日に収録され、NII データは2011年7月28日に収録された。

3.2 収録協力者

IMADE データ、NII データともに、外部業者を通じて個人情報保護の説明をし、

b) 本原稿は3節で説明するサンプルデータ収録(多人数・マルチモーダルコーパス作成)に関わったメンバーで執筆した。

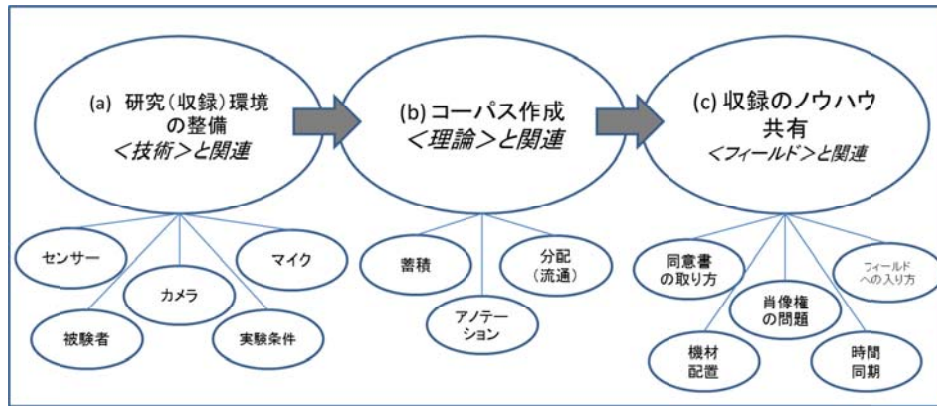


図 4 研究の進め方

条件に同意した収録協力者が参加した。IMADE データの本実験には初対面の 20 代女性を 7 名集め、課題によって 3 名と 4 名にグループ分けして収録に参加してもらった。また本実験の 1 週間前には実際の収録協力者の 2 倍にあたる 14 名に視線トラッキングオーディションに参加してもらい、視線計測装置(nac 社 EMR-9)による視線計測のしやすさを測定し、その中から 7 名の収録協力者を選定した。NII データの本実験には初対面の 20 代男性を 3 名集めて収録に参加してもらった。また IMADE データ同様、本実験の 1 週間前には実際の収録協力者 2 倍にあたる 6 名に視線トラッキングオーディションに参加してもらった。

3.3 収録課題

それぞれの収録には以下の 3 つの課題のうち、2 つずつの課題を利用した。

(1) アニメーション再生課題

この課題はこれまでジェスチャー研究[4][5][6]で頻繁に使われてきたものである。収録協力者 3 名のうち、1 名ないし 2 名は事前に 7 分程度のアニメーションを見ている。これは 1949 年にワーナーブラザーズ社によって制作されたショートアニメーションで、「Tweety and Sylvester」シリーズの「Canary row」という作品である。このアニメーションを見た収録協力者は、アニメーションの内容をなるべく詳しく他の二人に説明するように指示される。IMADE データと NII データ両方の収録で用いた。

(2) 砂漠課題

この課題は自分らが砂漠に不時着した飛行機の生存者と仮定し、提示された 12 のアイテムを収録協力者内で合意を取りながら、優先順位が高い順に並べ替えるものである。5 分程度で結果を出すように指示される。収録協力者は 3 名ないし 4 名にグル

ープ分けされた。IMADE データの収録で用いた。

(3) 自由対話課題(題目：学生時代の部活動の思い出)

この課題は収録協力者 3 名に自由に対話をさせるものである。自然に会話を開始させるために収録開始前にある一定の題目を提示するが、その題目以外の話に脱線しても構わないと指示される。NII データの収録で用いた。

図 5 は IMADE データ収録風景である。上述した(1)のアニメーション再生課題中に天井に設置したネットワークカメラで撮影したものである。この図の通り、収録参加者は等間隔に円陣形で肘掛のないパイプいすに着席し、課題を進めた。データ分析時には収録参加者を S1, S2, S3 と呼称して氏名等の個人情報を使用しない。図 5 のメンバーによる課題は、S2 と S3 がアニメーションを事前に見ており、S1 が見ていないという条件で進められた。S2 と S3 が頻繁に身振り手振りを加えながら発言し、S1 はうなずくなどして応答姿勢を取っているのが特徴的である[7]。

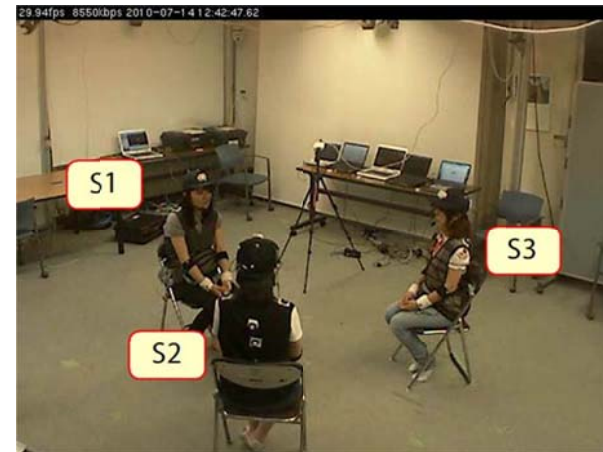


図 5 IMADE データ収録風景

3.4 収録機器

IMADE データ、NII データともに視線方向キャプチャには nac 社の EMR-9 を使用した。身体動作計測のためのモーションキャプチャには IMADE データでは nac 社の Mac 3D System を、NII データでは SPICE 社の OptiTrack をそれぞれ利用した。また、IMADE データの音声収録にはヘッドウォーンマイクロフォン(Shure 社 WH30XLR)と無線送受信システム(Shure 社 ULX)からなる最大 4 チャンネル同期収録可能な装置を利用し、NII データの音声収録には最大 4 チャンネル同期収録可能なミキサ(Avid

Technology 社 Pro Tools LE ミュージック・クリエーション/プロダクション・システム AVID 003 Factory)とアナログワイヤレスマイクフォン (Sony 社 UWP-V1) を利用した. IMADE データの映像収録のために 8 台のネットワークカメラ (AXIS Communications 社ネットワークカメラ AXIS210A) を使用した. 記録された映像は LAN ケーブルを通してサーバに蓄積し, avi ファイルに変換した上で保存している. また IMADE データと NII データともにハイビジョンビデオカメラ (Sony 社 HDR-XR550 等) と三脚を利用し, 各収録協力者の正面映像を収録している.

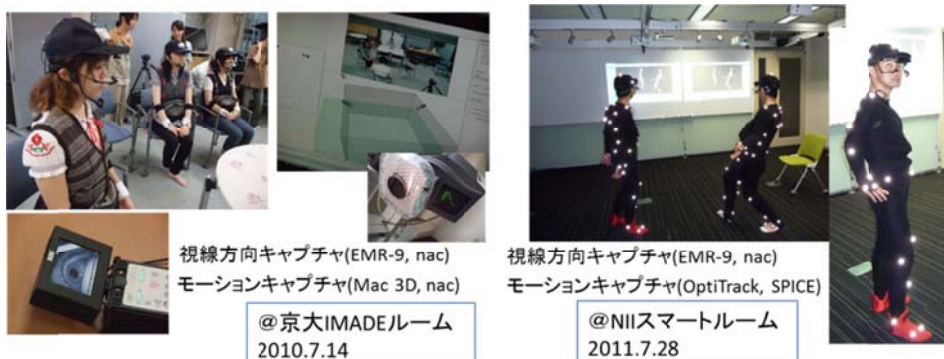


図 6 収録機器(左: IMADE データ, 右: NII データ)

3.5 ハンドアノテーションの付与

収録された IMADE データと NII データに対し, ルールに基づいて発話区間, ジェスチャー区間, 視線配布区間を手作業で区切り, それぞれに発話内容, ジェスチャー単位情報, 視線配布先情報のアノテーションを付与した. 作業にはマックスプランクインスティテュートが開発する ELAN[8]を用いた(図 7). また京都大学情報学研究科西田・角研究室では iCorpusStudio という多種多様な情報を一度に分析可能な環境の開発が行われている[9].

発話区間のセグメンテーションには間休止単位(Inter-Pausal Unit, IPU)を利用している. 今回は 200ms より長い無音によって区切られた一連の音声区間を発話と理解し, 発話区間を定めた[10].

ジェスチャー区間と視線区間のセグメンテーションにはジェスチャー研究で用いられるジェスチャー単位の考え方を利用している[11]. ジェスチャー単位とは, 手が膝や机の上(ホームポジションもしくはレストポジション)に置かれ, 全く動かない状態(ジェスチャーなし区間)から, 手がジェスチャーするための場所まで移動し(準備区

間: prep), ジェスチャーの核となる表現をし(ストローク: str), また膝や机の上に帰っていく(撤回区間: ret)といった一連の動作を記述する概念である. prep と str によって構成される動きの区間をジェスチャー句と呼び, ジェスチャー句は手がホームポジションに戻るまで新たに生成され連続される. 連続したジェスチャー句の集まりをジェスチャー単位と呼ぶ(図 8). 視線配布区間についても, ジェスチャー単位の考え方を援用し, アノテーションを付与している[12].

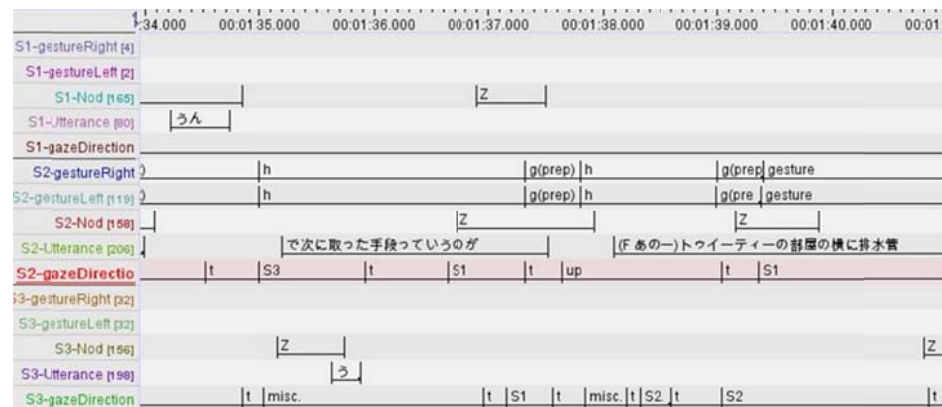


図 7 アノテーション例

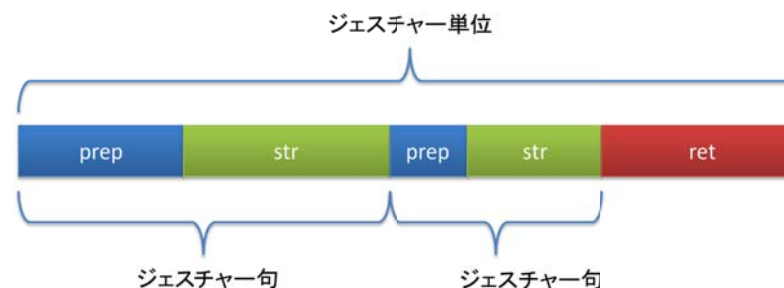


図 8 ジェスチャー単位の考え方

4. おわりに

本研究発表では、これまで個々のプロジェクトが独自の手法で回避してきた個人情報に関する問題に対する統一的手法を確保し、日常的なインタラクションデータを収録・集積するためのモデルケースとなる手法を提示するといった本研究プロジェクトのアプローチを述べた。また、5~6時間程度の公開可能なサンプルデータの収録および多人数・マルチモーダルコーパス作成の詳細を記述した。今後は、これらの情報を広く公開することで日常的インタラクションデータ収録・集積のためのガイドラインを整備していく予定である。これまでの活動及び今後の研究の進捗はホームページなどを通して随時公開していく予定である[13]。

謝辞 本研究プロジェクトを進めるためにご協力頂いた皆様に、謹んで感謝の意を表する。

参考文献

- 1) Yasuyuki Sumi, Masaharu Yano, and Toyoaki Nishida.: Analysis environment of conversational structure with nonverbal multimodal data, The Twelfth International Conference on Multimodal Interfaces and the Seventh Workshop on Machine Learning for Multimodal Interaction (ICMI-MLMI 2010), Beijing, China (2010).
- 2) 坊農真弓, 高梨克也共編/人工知能学会編: 知の科学 多人数インタラクションの分析手法, オーム社 (2009).
- 3) 中田 篤志, 角 康之, 西田 豊明: 非言語行動の出現パターンによる会話構造抽出, 電子情報通信学会論文誌, Vol.J94-D, No.1, pp.113-123 (2011).
- 4) McNeill, David.: *Hand and Mind*. University of Chicago Press, Chicago (1992).
- 5) Kendon, Adam.: *Gesture: Visible action as utterance*. Cambridge University Press, Cambridge, U.K. (2004).
- 6) McNeill, David.: *Gesture and Thought*. Chicago: The University of Chicago Press, Chicago (2005).
- 7) 井上卓, 角康之, 高梨克也: 状況説明会話における説明者間の発話とジェスチャーの引き取り, *インタラクション 2011* (2011).
- 8) ELAN <http://www.lat-mpi.eu/tools/elan/>
- 9) iCorpusStudio <http://www.ii.ist.i.kyoto-u.ac.jp/iCorpusStudio/index.html>
- 10) 榎本美香, 石崎雅人, 小磯花絵, 伝康晴, 水上悦雄, 矢野 博之.: 相互行為分析のための単位に関する検討, 人工知能学会研究会資料, SIG-SLUD-A402, pp.45-50. (2004).
- 11) 細馬宏通.: 非言語コミュニケーションのための分析単位 —ジェスチャー単位—, 人工知能学会誌, Vol.23, pp.390-396 (2008).
- 12) 坊農真弓.: 日本語会話における言語・非言語表現の動的構造に関する研究, ひつじ書房 (2008).
- 13) Bono lab. <http://research.nii.ac.jp/~bono/>

付録

- (1) H17-H22 年度, 情報爆発時代に向けた新しい IT 基盤技術の研究, 文部科学省科学研究費補助金特定領域研究, 【計画研究】 A03-00-02 「実世界インタラクションの分析・支援・コンテンツ化」(角, 研究分担者) <技術>
- (2) H22-H23 年度, 文部科学省科学研究費補助金基盤研究(B), 「参加者間のインタラクションに注目したミーティングの構造解析と支援」(角, 研究代表者) <技術>
- (3) H21-H23 年度, 文部科学省科学研究費補助金基盤研究(C), 「対話に潜むパターンを探り当てる手法の開発: 動的類似尺度を用いるアプローチ」(井上, 研究代表者) <技術>
- (4) H22-H23 年度, 文部科学省科学研究費補助金若手研究(B), 「インタラクションマイニングによる対話コンテキスト依存型ジェスチャー認識モデルの構築」(岡田, 研究代表者) <技術>
- (5) H20-H22 年度, 文部科学省科学研究費補助金基盤研究(B), 「対話における発話単位とその機能の認定に関する研究」(伝, 研究代表者) <理論>
- (6) H23-H25 年度, 文部科学省科学研究費補助金基盤研究(B), 「発話単位アノテーションに基づく対話の認知・伝達融合モデルの構築」(伝, 研究代表者) <理論>
- (7) H21-H23 年度, 文部科学省科学研究費補助金基盤研究(C), 「日本語と日本手話の「発話」に含まれる統合的關係と連鎖的關係のマルチモーダル分析」(高梨, 研究代表者) <理論>
- (8) H20-H22 年度, 文部科学省科学研究費補助金新学術領域研究(研究課題提案型), 「逸脱を吸収する社会実現に向けたコミュニケーションギャップ生成—解消機構の解明」(榎本, 研究代表者) <フィールド>
- (9) H23-H25 年度, 文部科学省科学研究費補助金基盤研究(B), 「精神医療現場における多相的コミュニケーションの共創支援~開かれた関係構築に向けて」(榎本, 研究代表者) <フィールド>
- (10) H21-H23 年度, 文部科学省科学研究費補助金基盤研究(C), 「介護施設において高齢者・介護職員間で交わされる身体動作を用いた空間表現の研究」(細馬, 研究代表者) <フィールド>
- (11) H21-25 年度, 科学技術振興機構戦略的創造研究推進事業(JST さきがけ), 「インタラクション理解に基づく調和的情報保障環境の構築」(坊農, 研究代表者) <フィールド>
- (12) H22-H23 年度, 文部科学省科学研究費補助金研究活動スタート支援, 「ろう者と聴者を含む多人数手話コミュニケーションでの参与枠組と言語・身体表現の関連」(菊地, 研究代表者) <フィールド>
- (13) H23-H25 年度, 学術研究助成基金助成金若手研究(B), 「学術領域における専門用語の手話表現とデータベース構築の研究」(菊地, 研究代表者) <フィールド>
- (14) H21-H25 年度, 科学技術振興機構戦略的創造研究推進事業(JST さきがけ), 「多人数インタラクション理解のための会話分析手法の開発」(高梨, 研究代表者) <フィールド/理論>
- (15) H23-H25 年度, 文部科学省科学研究費補助金基盤研究(B) 「ろう者・聴者の言語意識の改革を目指した「日本手話話し言葉コーパス」の構築」(坊農, 研究代表者) <フィールド/理論>