

解説

ロボカップ@ホームにおける日用品マニピュレーション

Everyday Object Manipulation in RoboCup@Home

杉浦孔明^{*1} 長井隆行^{*2} ^{*1}情報通信研究機構 ^{*2}電気通信大学

Komei Sugiura^{*1} and Takayuki Nagai^{*2} ^{*1}National Institute of Information and Communications Technology ^{*2}The University of Electro-Communications

1. はじめに

日本発のロボカップはロボットによるサッカー競技として有名である [1] が、ロボカップ@ホームリーグ（以下、ロボカップ@ホームと略記）についてはあまり知られていないかもしれない。ロボカップ@ホームは、サッカーやレスキューと並ぶロボカップのリーグのひとつであり、生活支援ロボットの競技である [2,3]。各チームのロボットは、日用品の探索、棚からユーザに言われたものを取ってくる、人を追従する等、日常生活に役立つ機能を制限時間内にどれだけ達成できるかを競う。2012 年のロボカップ@ホーム世界大会に出場したロボットを図 1 に示す。

ロボカップ@ホームでは、家庭・オフィス・スーパーマーケットなどにおけるロボットの応用を想定したタスクが設定されている。中心課題は、モバイルマニピュレーションとヒューマンロボットインタラクション (HRI) である。後述するように、未知環境における地図作成・移動、日用品の物体認識・把持、高騒音環境における音声認識などを含む。各タスクはベンチマークテストとして明文化されると同時に、複数の技術的課題を含んだストーリーになっており、観客を飽きさせないよう努力されている。

ロボカップ@ホームは 2006 年に始まり、我々は 2008 年より参加してきた。2009 年からは、世界大会のルール策定やジャパンオープン（日本大会）の運営にも参加している。2012 年メキシコで開催された世界大会には 9 カ国から 18 チームの参加があり、ジャパンオープンには 3 カ国から 10 チームの参加があった。各チームは 6~10 人程度で構成されていることが多い。

毎年行われる大会では、制限時間と相手チームの存在が、研究室でのデモとは違った緊張感をもたらしている。オブジェクトがレフェリーによりランダムに置かれる点や、未

知オブジェクトの把持が要求される点は、現状のマニピュレーション技術にとって簡単な問題ではない。音声対話の面からは、未知ユーザとのインタラクションは既知ユーザに比べてはるかに難しいので注意が必要である。

本稿では、マニピュレーションを中心としてロボカップ@ホームで対象とされる技術的課題と代表的なアプローチについて述べ、我々の取り組みを紹介する。

2. ロボカップ@ホームとは

2.1 競技の概要

ロボカップの他のリーグと同様に、例年、5 月にジャパンオープン、6 月に世界大会が開催されている。世界大会の日程は、2 日間のセットアップ期間と 4 日間のタスク期間からなる。セットアップ期間にフィールドやオブジェクトが発表されるので、効率的に環境地図構築、オブジェクト登録、練習を行うことが重要である。

2012 年のタスクの概要を表 1 に示す。タスクは、第 1 ステージ・第 2 ステージ・決勝にわかれている。第 1 ステージの総合得点により上位半数のチームが第 2 ステージに進み、第 2 ステージの上位 5 チームが決勝に進む。決勝では、各チームが自由デモを行う。デモに対する審査員の評価点と、それまでの得点を合計して最終順位が決定する。タスクは計 9 種類あり、各タスクには 5~10 分の制限時間と、1000~2000 点の配点が設定されている。



図 1 2012 年世界大会に参加したロボット

原稿受付

キーワード: RoboCup@Home, everyday manipulation, mobile manipulation, human-robot interaction

^{*1}〒 619-0289 京都府相楽郡精華町光台 3-5

^{*2}〒 182-8585 東京都調布市調布ヶ丘 1-5-1

^{*1}3-5 Hikaridai, Seika, Soraku, Kyoto

^{*2}1-5-1 Chofugaoka, Chofu-shi, Tokyo

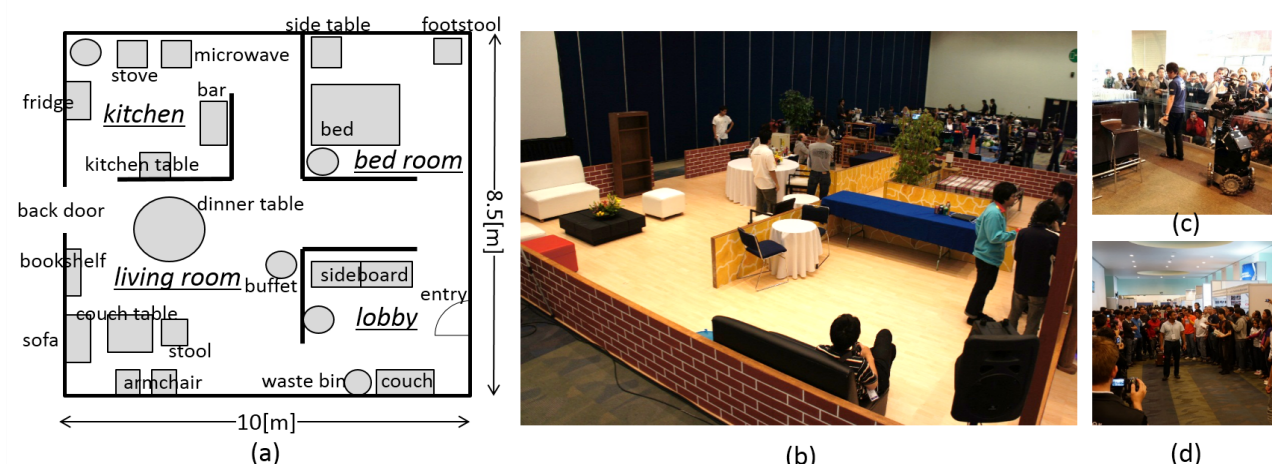


図2 2012年世界大会のタスク環境。(a)家具配置,(b)メイン環境(フィールド),(c) Restaurant タスクの環境,(d) Follow Me タスクの環境。

の命令を実行できた場合のみ満点を得られる。

- (1) 命令は3つの動作から構成される(例「キッチンテーブルまで行って、シーフードヌードルを取ってからフィールドから出て」)。オペレータは文生成器により生成された命令を読み上げるのみである。各動作を実行できれば200点が与えられる。
- (2) 命令には不十分な情報が含まれるため、確認発話を行なって曖昧性解消を行ったのち、動作を実行することが求められる。例えば、「飲み物をください」という命令の場合は、どこから取ってくるか、具体的な名称は何か、などの情報が欠けている。正しい確認発話を生成できれば300点、タスクを全て実行できれば600点(途中まで実行できた場合は300点)が与えられる。
- (3) テーブルがない状態で「コーラをテーブルにのせて」と命令されるなど、命令に実行可能でない部分が存在する。タスクを途中まで実行できた場合は300点、「コーラはありましたがテーブルはありません」のように問題を特定できれば200点が与えられる。

GPSR 文生成器は、動作・場所・オブジェクトをランダムに選択する。場所とオブジェクトはセットアップ期間に公表された名称の集合から選択される。動作は第1ステージのタスクで要求される動作のいずれかと定められている。動作を表す動詞(句)は類義語を含めて29種類あり、これらはルールブックの形態素解析結果から抽出されたものである。ただし、第1ステージのタスクが毎年更新されるため、動作候補数も変動する。単純に単語をランダムに組み合わせると非文(不自然な表現)が多く生成されるため、GPSR 文生成器ではテンプレートに基づく手法を採っている。テンプレート文のスロットに動作・場所・オブジェクトを組み合わせることで、数万種類の文を生成可能である。過去

の大会で用いられた GPSR 文生成器は公開されている[†]。

GPSR タスクは、ロボカップ@ホームタスクのなかでも多彩な機能を要求される。得点を取ることは非常に難しく、実際に2012年の世界大会では得点を取ることができたチームは3チームのみであった。ロボカップ@ホーム環境では信号対雑音比(SN比)が実質的に0dB程度であることを考えると、カテゴリ1の複文の音声認識は非常に困難である。

3. システム

3.1 DiGORO の概要

以下では、我々のロボットプラットフォーム DiGORO(図4(a))について概説する。なお、DiGOROの詳細については[7]を参照されたい。

● 自己位置推定および移動

レーザレンジファインダ(LRF)から得られる情報をもとに、ICP(Iterative Closest Point)[8]を用いて地図作成と自己位置推定を行う。経路計画にはRRT(Rapidly-exploring Random Tree)[9]を用い、全方位移動台車により移動する。

● 画像処理

CCDカメラ2台と赤外線TOF方式3次元距離測定カメラから得られた情報を統合して、RGB-Dデータを得ている[10]。

● マニピュレーション

[7]で提案した手法により物体認識を行い、認識された物体の重心座標を計算する。オブジェクトが遠い場合は台車で移動を行う。アームの経路計画では、13次元のコンフィギュレーション空間(右腕6自由度、左

[†]http://komeisugiura.jp/software/2011jp_GeneralPurposeTest.tgz

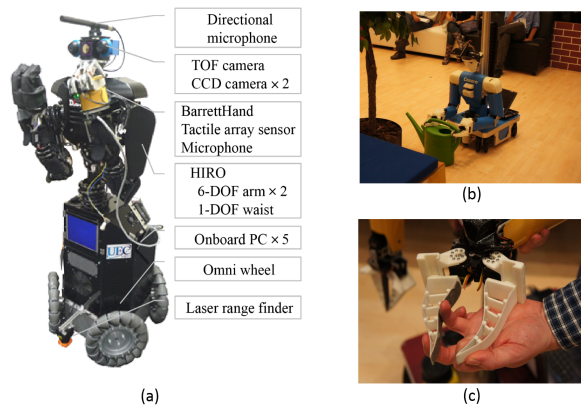


図4 (a) ロボット DiGORO, (b) 昇降機構を用いて床のオブジェクトを掴むロボット Cosero, (c) Cosero のハンド

腕 6 自由度, 腰 1 自由度) 上で経路を生成する [11].

● 音声対話

指向性マイクを用いることでユーザ方向以外の音声を除去する。さらに, Fujimoto らにより提案された手法 [12] を用いてノイズの逐次推定と抑圧を行う。音声認識および合成には, ATRASR [13] および XIMERA [14] を用いた。

3.2 他チームのシステム

図1に示すように, ロボカップ@ホームでは台車を移動手段として用いるロボットが多い。2011 年世界大会では 19 チーム中 18 チーム, 2012 年世界大会では 18 チームすべてが台車移動のロボットを用いていた。代表的な台車としては, Adept Technology(MobileRobots) 製 Pioneer シリーズ, Fraunhofer IAIS 製 Volksbot, Segway 製 RMP-100 などがある。RMP-100 は 2 輪のみで走行可能であるが, 緊急停止スイッチが押された場合に停止させるというルールがあるため, キャスターを取り付けるなど転倒予防策が必要である。

アームやグリップを持たないロボットの出場は可能であるが, スコア上有利ではない。上位 5 チームのうちアームやグリップを持たないロボットは, 2008 年 3 台, 2009 年 1 台であったが, 2010 年以降は 0 台である。既製アームとしては, Neuronic 製の katana 6M180 や Exact Dynamics 製 iARM などが用いられている。これまで Aldebaran Robotics 製の NAO を用いたチームもあったが, アームが届く範囲が極めて短いためオブジェクトの把持に苦労していた。

様々な高さに置かれたオブジェクトを把持できるように, アームの高さを上下させる機構を持つロボットが増えてきている (図 4(b))。オブジェクトは, 床, 棚, 机 (70cm 程度), 調理台 (85cm 程度) などに置かれるが, 単一のアームでは様々な高さに対応することが難しい。2010 年世界大会から, 昇降機構を用いて上半身を上下させるロボットが現れるようになった。ロボカップ@ホーム以外では, Willow

Garage の PR2 [15] やトヨタ自動車の HSR [16] など昇降機構を有している。

上位チームのロボットが把持可能なオブジェクトの重量は, 500g から 1000g 程度である。2011, 2012 年世界大会優勝のチーム Nimbro のロボット Cosero は, 小型軽量 (重量 32kg, フットプリント 59×44cm) ながら, 可搬重量は 1.5kg であった [5]。また, Cosero は Festo 社の FinRay をベースにした可変形状ハンド (図 4(c)) を用いてオブジェクトの把持を行なっている [17]。

ロボット開発用ミドルウェアとしては, ジャパンオープンでは RT ミドルウェア, 世界大会では ROS を採用するチームが多い。ROS はオープンソースのロボットライブラリであり, シミュレータ, 経路計画, 画像処理, 通信など, 豊富なライブラリに特徴がある。

4. コミュニティブマニピュレーション

1 節で述べたように, ロボカップ@ホームの中心課題はモバイルマニピュレーションと HRI である。これらの側面を持つタスクとして, ユーザがロボットに音声でマニピュレーションを指示するタスクが挙げられる。本稿では, このタスクをコミュニケーションと呼ぶ。以下では, コミュニティブマニピュレーションに関する我々の取り組みについて紹介する。

4.1 マニピュレーション命令の発話理解

ロボットが日常環境に浸透するに従い, コミュニケーション能力が課題となる。音声によるコミュニケーションはユーザにとって手軽であるというメリットがあるものの, 自然言語に含まれる曖昧性の処理は簡単ではない。前述したように GPSR タスクでは, 音声命令を解釈して動作する必要がある。指示文はランダムに生成されるため, 柔軟な行動生成能力が問われることになる。そこで, 我々は指示発話を解釈してロボットの行動に変換する枠組みを開発した [18]。

図5に処理の概略を示す。システムは, 音声命令の音声認識結果 (自然言語文) を入力として受け取り, 基本動作の組み合わせからなるプログラム (スクリプト) を出力する。まず, 形態素解析および構文解析により, 動詞と名詞の係り受け構造を求める。ただし, 動詞と基本動作列の対応, および基本動作群自体は開発者が与えるものとする。例えば「持ってくる」の場合, 移動, 探索, 把持などの動作列に対応する。動作を実行するには目標地点やオブジェクト情報などが必要である。これらは名詞と対応付けられた場所情報やセンサ情報から得られる。

ロボットの動作について知識を有していないユーザを対象とする場合, 動作と対応づけられていない動詞が使用される場合もあり得る。そのような場合, 日本語および英語の Wordnet [19] を同時に利用して, 既知の動詞の類義であるかを判定する。なお, 後述するように動作の学習も可能

である。

不足している情報がある場合は、すでに獲得している情報に依存した質問を生成する。例えば、「コップを持ってきて」という命令にはコップの場所の情報が不足しているため、「コップはどこにありますか?」という質問が生成される。実験の結果、質問を行うことで正しいスクリプトを生成できる割合の向上が達成できた。生成されたスクリプトを物理的に実行したときの動作成功率は 86%であった。

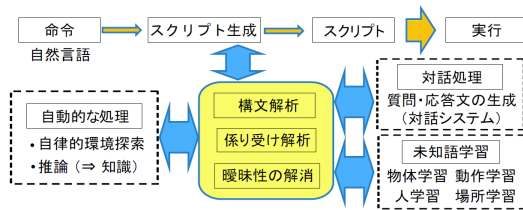


図 5 マニピュレーション命令の発話理解と動作生成

4.2 日用品マニピュレーションの模倣学習

前節では、基本動作をあらかじめ与える場合について述べた。ただし、さまざまな動詞と動作の対応を用意するコストは非常に大きい。そこで我々は、模倣学習の枠組みにより日用品を扱う動作を学習する手法の開発を行なっている。このような学習手法を構築することで、プログラミングスキルが必要とされないユーザフレンドリな動作教学方法を実現できる。ロボカップ@ホーム環境における家事動作の学習への適用例を図 6 に示す。

「X を Y にのせる」や「Z を回す」など参照点に依存した動作の模倣では、世界座標系での動作軌道の模倣に意味はなく、適切な座標系を推定し軌道を汎化しなければならない。このために、参照点に依存した隠れマルコフモデル (Reference-Point-Dependent HMM, 以下 RPD-HMM と略記) を開発した [20]。RPD-HMM は、物体位置の時系列を入力として、動作をモデル化するための最適な座標系を EM アルゴリズムにより推定し、軌道のモデルを学習している。

動作の生成時には、学習時とまったく同じように物体が配置されている訳ではないため、たとえ同じ「載せる」動作であっても、学習時の軌道をそのまま用いることは無意味である。学習した RPD-HMM は固有座標系において汎化されたものであるため、固有座標系 C から世界座標系 W へ変換する。RPD-HMM の位置・速度・加速度の平均ベクトルおよび共分散行列は、同時変換行列により C から W 上のモデルに変換される。HMM から滑らかな軌道を生成するために、音声合成の分野で用いられる HMM 軌道生成 [21] を用いる。実験の結果、7 回程度の教示で生成軌道の誤差が収束することが確認できた。

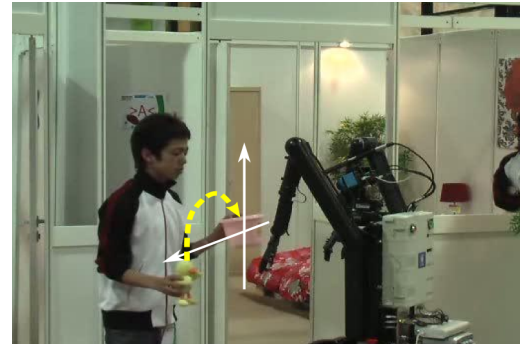


図 6 ロボカップ@ホーム環境における動作「捨てる」の学習

4.3 コミュニティブマニピュレーションにおける学習と統合

ここまでは、オブジェクト名の音素列やカテゴリなどの情報は与えられることを前提としてきた。また、一般的にロボットの対話処理機構では、実世界情報と音声認識は別々に処理されていることが多い [22]。一方、画像、動作、アフォーダンス、履歴などさまざまな実世界情報の学習と統合は知能ロボティクスの到達点の一つであろう。

このような背景から我々はコミュニケーション学習基盤 LCore [23] を開発している。LCore はマルチモーダル入力から学習されたモデルを用いてユーザの発話を理解する。図 7 に示す例において、全ての語彙は音素列も含め学習されたものである。LCore は、以下の発話理解スコア Ψ によりマルチモーダル発話理解を行う。

$$\Psi(s, a_k, O, \mathbf{q}^{(i)}) = \max_z (\gamma_1 B_S + \gamma_2 B_I + \gamma_3 B_M + \gamma_4 B_R + \gamma_5 B_H)$$

ここに、 z は各単語の概念構造への分割であり、 $(s, a_k, O, \mathbf{q}^{(i)})$ はそれぞれ、発話、行動、状況、トラジェクタに関するコンテキストを表す。また、 $\gamma = (\gamma_1, \dots, \gamma_5)$ は各項に対する重みであり、MCE 学習 [24] を用いて学習される。 B_S, B_I, B_M, B_R, B_H はそれぞれ、音声、視覚情報、予測軌道、動作-オブジェクト関係、行動コンテキストの尤度スコアを表す。モデルの詳細については [23] を参照されたい。我々は LCore を確認発話生成 [25] に拡張し、失敗動作を低減できるという知見を得ている。

5. おわりに

人口構造の変化や共働き世帯の増加とともに、日常環境で人間と共存するロボットへの期待が高まっている。例えば、掃除ロボットの国内市場規模は 2006 年に約 2 万台、2008 年に約 5 万台であったものが、2010 年には約 26 万台になっている。また、高齢化社会における QOL (Quality of Life) 向上が社会的に急務であることは論を待たない。日常環境への対応、ロボットによる自立支援の促進などは、ロボット開発において今後も重要な課題であろう。



図7 LCoreによるマルチモーダル発話理解

本稿では、日常環境における生活支援ロボットのベンチマークテストであるロボカップ@ホームと、それに関する我々の取り組みについて紹介した。生活支援ロボットの研究開発においては、独自の環境や評価尺度が用いられることが多く、一般的に手法同士の比較が難しい。一方、タスクを標準化することで比較評価のコストを低減すれば、コミュニティ全体の研究開発に貢献できるであろう。ロボカップ@ホームリーグとしては、タスクを標準化することによって、各ロボットを客観的に評価し競争的に開発を加速させることを目指していく。これまでの大会の動画はインターネット上に数多くアップロードされているので、「RoboCup@Home」などのキーワードで検索されたい。

参考文献

- [1] H. Kitano, M. Asada, Y. Kuniyoshi, I. Noda, and E. Osawa, "RoboCup: The Robot World Cup Initiative," Proceedings of the First International Conference on Autonomous Agents, pp.340–347ACM, 1997.
- [2] L. Iocchi, and T. van der Zant, "RoboCup@Home: Adaptive Benchmarking of Robot Bodies and Minds," Proceedings of the International Conference on Simulation, Modeling and Programming for Autonomous Robots, pp.171–182, 2010.
- [3] 杉浦孔明, "ロボカップ@ホームリーグ," 情報処理, vol.53, no.3, pp.250–261, 2012.
- [4] 岡田浩之, 大森隆司, "ロボカップ@ホーム: 人とロボットの共存を目指して," 人工知能学会誌, vol.25, no.2, pp.229–236, 2010.
- [5] J. Stückler, D. Holz, and S. Behnke, "RoboCup@Home: Demonstrating Everyday Manipulation Skills in RoboCup@Home," Robotics & Automation Magazine, IEEE, vol.19, no.2, pp.34–42, 2012.
- [6] <http://www.ai.rug.nl/robocupathome/>.
- [7] 中村友昭, アッタミムハンマド, 杉浦孔明, 長井隆行, 岩橋直人, 戸田智基, 岡田浩之, 大森隆司, "拡張モバイルマニピュレーションのための新規物体の学習," 日本ロボット学会誌, vol.30, no.2, pp.213–224, 2012.
- [8] F. Lu, and E. Milios, "Globally Consistent Range Scan Alignment for Environment Mapping," Autonomous robots, vol.4, no.4, pp.333–349, 1997.
- [9] S. LaValle, and J. Kuffner, "Rapidly-Exploring Random Trees: Progress and Prospects," Algorithmic and Computational Robotics: New Directions, pp.293–308, Wellesley, 2001.
- [10] M. Attamimi, A. Mizutani, T. Nakamura, T. Nagai, K. Funakoshi, and M. Nakano, "Real-Time 3D Visual Sensor for Robust Object Recognition," Proceedings of IROS, pp.4560–4565, 2010.
- [11] 伊東慶輔, 中村友昭, 長井隆行, "RRTを用いたコンフィギュレーション空間における双腕同時経路計画," 日本ロボット学会学術講演会, pp.1M2-2, 2010.
- [12] M. Fujimoto, and S. Nakamura, "Sequential Non-Stationary Noise Tracking Using Particle Filtering with Switching Dynamical System," Proceeding of 2006 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2006), pp.769–772, 2006.
- [13] S. Nakamura, K. Markov, H. Nakaiwa, G. Kikui, H. Kawai, T. Jitsuhiro, J. Zhang, H. Yamamoto, E. Sumita, and S. Yamamoto, "The ATR Multilingual Speech-to-Speech Translation System," IEEE Transactions on Audio, Speech, and Language Processing, vol.14, no.2, pp.365–376, 2006.
- [14] H. Kawai, T. Toda, J. Ni, M. Tsuzaki, and K. Tokuda, "XIMERA: A New TTS from ATR Based on Corpus-Based Technologies," Proceedings of Fifth ISCA Workshop on Speech Synthesis, pp.179–184, 2004.
- [15] J. Bohren, R. Rusu, E. Jones, E. Marder-Eppstein, C. Pantofaru, M. Wise, L. Mosenlechner, W. Meeussen, and S. Holzer, "Towards Autonomous Robotic Butlers: Lessons Learned with the PR2," Proceedings of ICRA, pp.5568–5575, 2011.
- [16] 寺田耕志, 橋本国松, 齋藤史倫, 池田幸一, 山本貴史, "生活支援ロボット HSR の開発と実証評価," 第13回システムインテグレーション部門講演会資料集, pp.2A1-2, 2012.
- [17] J. Stückler, R. Steffens, D. Holz, and S. Behnke, "Efficient 3D Object Perception and Grasp Planning for Mobile Manipulation in Domestic Environments," Robotics and Autonomous Systems, 2012.
- [18] 板谷純希, 中村友昭, 長井隆行, "ユーザとのインタラクションに基づく学習を利用したロボットのタスクプログラミング," 第25回人工知能学会全国大会資料, 2011.
- [19] G.A. Miller, et al., "WordNet: a Lexical Database for English," Communications of the ACM, vol.38, no.11, pp.39–41, 1995.
- [20] K. Sugiura, N. Iwahashi, H. Kashioka, and S. Nakamura, "Learning, Generation, and Recognition of Motions by Reference-Point-Dependent Probabilistic Models," Advanced Robotics, vol.25, no.6-7, pp.825–848, 2011.
- [21] K. Tokuda, T. Yoshimura, T. Masuko, T. Kobayashi, and T. Kitamura, "Speech Parameter Generation Algorithms for HMM-Based Speech Synthesis," Proceedings of ICASSP, pp.1315–1318, 2000.
- [22] R. Hartanto, A Hybrid Deliberative Layer for Robotic Agents, Springer, 2011.
- [23] N. Iwahashi, R. Taguchi, K. Sugiura, K. Funakoshi, and M. Nakano, "Robots that Learn to Converse: Developmental Approach to Situated Language Processing," Proceedings of International Symposium on Speech and Language Processing, pp.532–537, 2009.
- [24] S. Katagiri, B. Juang, and C. Lee, "Pattern Recognition Using a Family of Design Algorithms based upon the Generalized Probabilistic Descent Method," Proceedings of the IEEE, vol.86, no.11, pp.2345–2373, 1998.
- [25] 杉浦孔明, 岩橋直人, 柏岡秀紀, 中村哲, "言語獲得ロボットによる発話理解確率の推定に基づく物体操作対話," 日本ロボット学会誌, vol.28, no.8, pp.978–988, 2010.

杉浦 孔明 (Komei Sugiura)

2007年京都大学大学院情報学研究科博士後期課程修了。博士(情報学)。日本学術振興会特別研究員, ATR 音声言語コミュニケーション研究所研究員を経て, 2009年より情報通信研究機構研究員。ロボット対話, 機械学習, 推薦システムの研究に従事。計測自動制御学会, 人工知能学会, IEEEなどの会員。(日本ロボット学会正会員)

長井 隆行 (Takayuki Nagai)

1993 慶應義塾大学理工学部電気工学科卒。1995 同大学院理工学研究科修士課程修了。1997 同大学院博士課程修了。博士(工学)。慶應義塾大学理工学部訪問研究員を経て, 1998 電気通信大学電子工学科助手。2003 カリフォルニア大学サンディエゴ校客員研究員。2004 電気通信大学大学院電気通信学研究科助教授。現在同大学院情報理工学研究科准教授。2011より玉川大学脳科学研究所特別研究員を兼務。知能システム, 知能ロボティクスに関する研究に従事。日本ロボット学会, 情報処理学会, 電子情報通信学会, IEEE各会員。(日本ロボット学会正会員)