

プロアクティブヒューマンインターフェースの研究

- 第1報 人間型アクティブインターフェースの開発 -

Study on Proactive Human Interface - Development of Humanoid-type Active Interface -

○ 正倉爪亮(九大) 内田 誠一(九大) 谷口 倫一郎(九大) 長谷川 勉(九大)

Ryo Kurazume, Seiichi Uchida, Rinichiro Taniguchi, and Tsutomu Hasegawa
Kyushu University, 6-10-1, Hakozaki, Higashi-ku, Fukuoka

We are conducting research on “Embodied Proactive Human Interface”. The aim of this research is to develop a new human-friendly active interface based on a new driving mechanism named “Proactive Interface” and a physical device using robot technology. This paper introduces our motivation and the aim and the final target of this research. New humanoid-type proactive human interface and remote communication experiments are also presented.

Keywords: Proactive System, Active Human Interface, Humanoid, Intension

1 はじめに

我々は「プロアクティブヒューマンインターフェース」の研究を行っている。この研究の目的は、“人間からの詳細なデマンドに寄らない新たな駆動原理（プロアクティブ技術）”に基づき、従来の仮想メディアだけではなく“実体としてのロボット技術”の活用することで、人とコンピュータシステムに存在する垣根を低減化し、様々な人にとって自然で使いやすいヒューマンインタフェースの枠組みを提供しようというものである。

近年、コンピュータの低価格化と高速なインターネットの整備が進み、誰でもコンピュータとネットワークを手軽に使えるような状況になりつつあり、多くの個人的、社会的な活動がコンピュータのディスプレイ、キーボード、マウスを通して行えるようになってきている。しかし、現実にはデジタル・デバイドという言葉に象徴されるように、コンピュータシステムを使いこなせる人とそうでない人の間に格差が生じ始めている。このデジタル・デバイドの根底にある本質的な問題は、以下の2つに集約されると考えられる。

1. コンピュータシステムのインタフェースとして利用可能なメディアに制限があるため、コンピュータを通じた活動の多くは仮想化されており、実世界で行われる自然な活動との乖離が起こっている。現在のコンピュータシステムを用いた人間の活動は、対象となる問題をいかに仮想化できるかに依存しているが、人間の身体動作など適切な仮想化が難しいものも存在している。特に、物理的な実体を意識させるような（物理的な働きかけを可能にする）メディア技術が欠如しており、仮想化された世界での活動と実世界での活動の乖離を少なくするような技術が必要である。これを解決するためには、実世界の情報を様々なチャンネルで獲得する技術と、仮想化された情報をなるべく自然な形で人間に提示す

る技術の両者を確立し、仮想化された世界での活動を実世界での活動に近づけさせることが重要である。

2. コンピュータシステムを通して活動を行うためには、どのような作業や処理を行うかの詳細なデマンドを人間がシステムに陽に与えなければならぬ。しかし、前項で述べたように仮想化された作業と実世界の作業との乖離が大きいと、詳細なデマンドをシステムに与えることは容易ではなく、システムとのスムーズなインタラクションが困難になっている。人間の作業や処理の意図を理解することができれば、詳細なデマンドをコンピュータシステム側で事前にプランニングすることができ、人間の介入を削減することができるので、乖離の影響を少なくすることが可能である。さらに、システムが先回りして人間からのデマンドを予測し、システムと人間の間の情報交換の遅延を削減することで、スムーズなインタラクションが期待される。

以上の考察から、本研究では人間の活動の適切な仮想化を目指した情報チャネル増強のためのロボティクス技術と、人間の行動パターンや意図を理解して先回りし、人間からの詳細な直接的デマンドによらないヒューマンインタフェース技術（ここではプロアクティブ・ヒューマンインタフェースと呼ぶ）を組み合わせ、様々な人にとって自然で使いやすいヒューマンインタフェースの枠組みを提供しようというものである。

提案するプロアクティブインターフェースの実現例を図1に示す。この例では、遠隔地にいる2人の人間がコミュニケーションするとき、相手側のヒューマノイドロボットを自分の分身として使い、音声だけではなく、身振りや視線、表情なども伝送、再現することで、相手に自分の身体を意識させ、コミュニケーション能力を増強しようというものである。このとき、例えば会話や行動の履歴から操作者の意図を先読みし、予測し

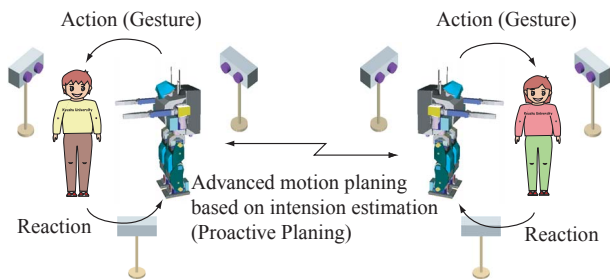


図 1: Concept of proactive human interface

た意図情報を実際の行動が生じる前に相手方に伝えて次時刻での再現に備えることができれば、実体を通したヒューマンインタフェースを、遅れやギャップの少ないより自然な動きに近づけることができると考えられる。また推定された意図に基づいて、相手方のロボットでその意図にあう行動を再現できれば（例えばその行動が実際には成されなかったとしても）自分の意図をより正確に伝えるための有効な情報チャンネルになる。また上述の例だけではなく、例えばコンピュータの入力装置に意図推定に基づく能動的な補助運動を利用することで、操作に不慣れな人や肢体に障害がある人にとって使いやすい入力装置が実現できる。技能伝達のための新たなメディアとしてのロボットの利用 [1], [2] や、操作者の意図推定を基にした動作補助 [3],[4] など、本研究には多くの先行研究があるが、特にコミュニケーション能力の補強を中心とし、先読みした意図による行動選択と実体を感じさせるヒューマンインタフェースの結びつけが本研究の特徴である。

2 人間型アクティブインターフェースの試作

2.1 人間型アクティブインターフェース

上述したプロアクティブインターフェース実現のための新たな情報提示ハードウェアの開発を目的として、遠隔コミュニケーションのための人間型ヒューマノイドロボットシステムを試作した。これは人間型ロボット (HOAP-2, 富士通オートメーション) をベースに、LCD やスピーカ、マイク、小型デジタルカメラを装備したものである (図 2)。このシステムは遠隔地の話者の会話や表情、動作を計測、伝送し、リアルタイムでヒューマノイドロボットにより歩行も含めた全身動作を再現する。ロボットが装着した液晶モニターやデジタルカメラは、遠隔地の話者の表情の計測と表示、または会話内容に関連したデータや会話そのものを表示するのに使用される。

2.2 歩行制御および全身動作の生成

インターフェースとしての実体に人間型ヒューマノイドロボットを用いた場合、2 足歩行動作を含む安定な全身運動の実現は最も基本的かつ重要な研究課題である。そこで我々がこれまでに 4 足歩行ロボットの安定歩行計画手法として開発した「3D 胴体揺動歩容方式」を、2 足歩行ロボットの制御問題に拡張し、安定

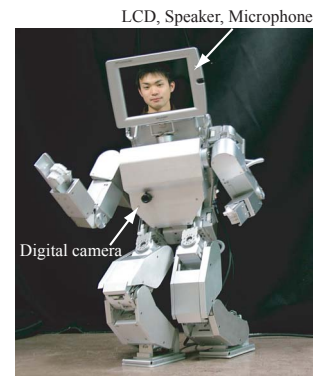


図 2: Embodied proactive human interface

重心軌道計画法 [5] や ZMP ヤコビを用いた多質点誤差補償法など安定歩行を実現するための運動制御手法を新たに開発した。さらにロボットに搭載した加速度センサ、角速度センサ、足裏力センサ等を拡張カルマンフィルタを用いて統合し、加速度の時間微分である躍度を利用した ZMP フィードバック系を構成することで、ある程度の凸凹面、傾斜面でも安定して歩行を継続できる「躍度フィードバック制御系」を構築した。さらにこれと梶田らによって提案された躍度予測制御系 [6],[7] と組み合わせ、実時間全方向安定歩行制御系を実現した。これら歩行制御法の詳細については別報で報告する。

開発した人間型アクティブインターフェースを用いた歩行実験の結果、75mm/秒の高速歩行や 30 度の傾斜面での安定直立動作 (図 3)、足裏力センサを用いた段差乗り越え (図 4)、その場旋回を含むオンライン全方向歩行動作 (図 5,6) などが実現できることを確認した。

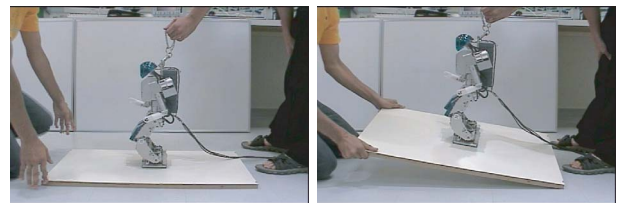


図 3: Stable standing on sloop

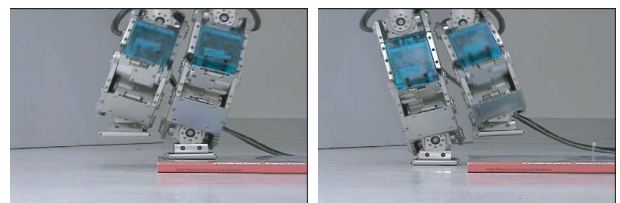


図 4: Step climbing

2.3 遠隔コミュニケーションのための予備実験

図 1 で示した遠隔コミュニケーションシステムを実現するための予備実験として、人間の上半身の動きをステレオカメラにより計測し、同様の動作を開発した人間型アクティブインターフェースによりリアルタイ



図 5: On-line walking pattern generation by joystick

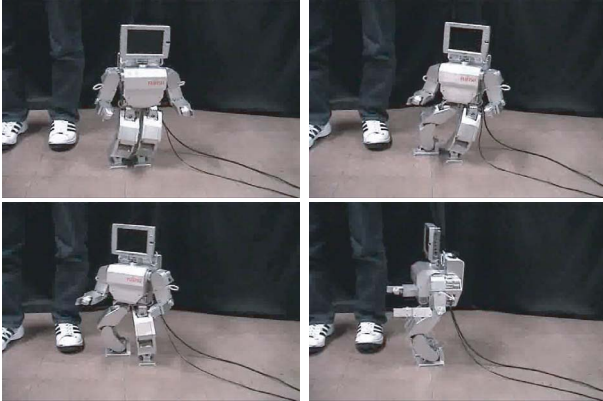


図 6: 90° turn motion

ムで行わせるシステムを構築し、動作再現実験を行った。本システムの処理の流れを以下に示す。

1. まず話者の前方に置かれたステレオカメラ (Point Gray 社製, BumbleBee) により距離画像を計測し、肌色抽出により両手と顔部分を切り出すことで、人間の手および顔の 3 次元位置を計測する (図 7)。
2. 人間の手と顔の相対位置情報を元に、ヒューマノイドロボットの身体モデルから手先指令位置を計算する。ただし人間とヒューマノイドロボットはサイズや自由度配置が異なることから、顔と手首の相対位置、および肘から手首への方向が一致するように変換した。
3. 手先の位置姿勢についてのインピーダンス制御系を構築し、刻々変化する目標手先位置に手先を自然に追従させる

図 8 に実験の様子を示す。操作者の動きを計測するシステムとロボットとは TCP/IP を用いて通信しており、インターネットを経由して遠隔地で動作の再現を行うことも可能である。実験の結果、ヒューマノイドは前項の躍度フィードバック制御系が有効に機能し、手先を自由に移動させてもバランスを崩すことなく、人間の全身動作を安定して再現できることがわかった。今後は歩行や踏み変えなどの下半身動作も加えた、人間の全身動作の再現実験を行う予定である。

3 行動意図の推定

3.1 行動意図推定のためのジェスチャのスポットティング認識

本節では、前述の人間型アクティブインターフェースでの利用を前提とした、人間の行動意図推定法につ

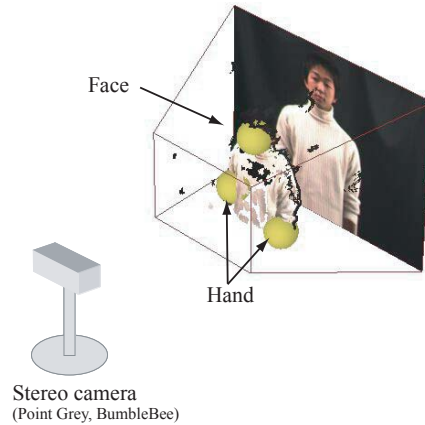


図 7: Measurement of human gesture for telecommunication system

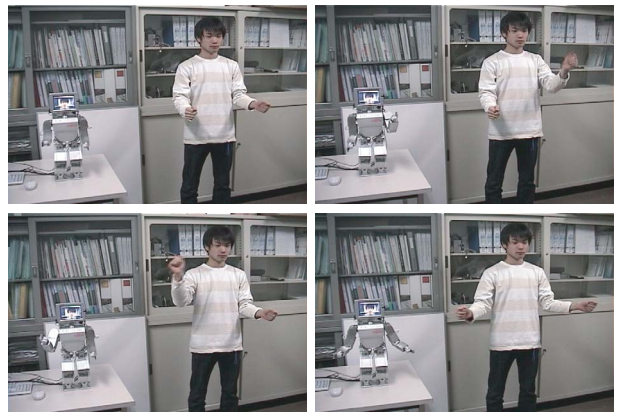


図 8: Imitation of human gesture

いて述べる。本手法は以下の 2 条件の満足を目指したジェスチャ認識法の一つである。まず第 1 は、ユーザの行動をなるべく初期の段階で認識可能という条件である。これは本手法がプロアクティブインターフェースにおける行動意図推定法として効果的に機能するための条件である。従来のジェスチャ認識法には一連の動作完了後に認識結果を出力するものもあるが、そうした手法では先読み処理は不可能である。これに対し、動作の初期の段階で認識結果が得られれば、その後もそのまま動作が継続すると考えることで、先読み処理が可能となる。

第 2 は、ジェスチャのスポットティング認識が可能という条件である。本研究で想定しているコミュニケーションタスク (図 1) では動作の開始/終了点が不明であるので、こうした条件が必要となる。

以上の 2 条件を満たす手法として、連続 DP に基づくジェスチャのスポットティング認識アルゴリズムを提案する。連続 DP [8] とは、DP マッチングを始端・終端自由型に拡張したもので、ジェスチャ認識にも利用されている [9]。なお、DP の確率的拡張である HMM もジェスチャ認識に利用されている (例えば [4])。しかし、HMM は基本的に始端 (もしくは終端) 固定であり、スポットティング認識に使うためには、事前のセグメンテーションやジェスチャ無し区間のモデル化などが必要となる点で不利である。

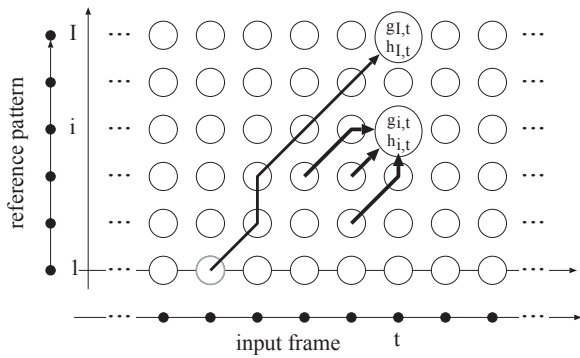


図 9: Gesture spotting based on continuous DP

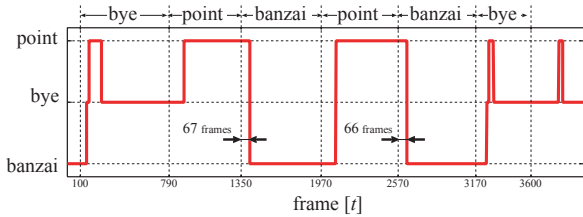


図 10: Result of gesture spotting

3.2 スポットイングアルゴリズム

本手法は，次式で計算される累積コスト $g_{i,t}$ ならびにサポート $h_{i,t}$ を用いてスポットイング認識を実行する（図 9）．

$$g_{i,t} = \min \begin{cases} g_{i-1,t-1} + 2d_{i,t} & \text{(a)} \\ g_{i-1,t-2} + 2d_{i,t-1} + d_{i,t} & \text{(b)} \\ g_{i-2,t-1} + 2d_{i-1,t} + d_{i,t} & \text{(c)} \end{cases}$$

$$h_{i,t} = \begin{cases} h_{i-1,t-1} \cap q(d_{i,t}) & \text{if (a)} \\ h_{i-1,t-2} \cap q(d_{i,t-1}) \cap q(d_{i,t}) & \text{if (b)} \\ h_{i-2,t-1} \cap q(d_{i-1,t}) \cap q(d_{i,t}) & \text{if (c)} \end{cases}$$

ここで， t は入力フレーム番号， i は標準パターンのフレーム番号である．また $d_{i,j}$ はフレーム t をフレーム i に対応させた場合の局所コストである．上式のように累積コスト $g_{i,t}$ は従来法 [8, 9] とほぼ同様に計算される．一方，サポート $h_{i,t}$ は前述の条件 1 のために新しく導入したものであり， $g_{i,t}$ 計算時の経路選択に応じて計算される．ここで $q(d_{i,t})$ は局所距離 $d_{i,j}$ がある閾値以下 (以上) の時 1(0) を返す関数である．従って (i, t) に至る最適経路上のすべての状態が標準パターンとよく類似していた場合に限り， $h_{i,t} = 1$ となることがわかる．

各時刻 t では，基本的にすべての標準ジェスチャパターンについて累積コスト $g_{i,t}$ を比較し，その最小値を与えたジェスチャを認識結果とする（経路長による正規化処理が入るが説明は省略する）．しかし，このままでは原理的にジェスチャ開始後 $I/2$ フレーム経過しないと認識結果が得られない．そこでそれ以前の過渡的なフレームにおいても，尤もな認識結果を得るためにサポートを用いる．具体的には，前述の累積コストの最小値がある値以上のフレーム，すなわち過渡的と判断されたフレーム t について，そのフレーム t のすべての標準パターンのサポート $h_{i,t}$ ($1 \leq i \leq I$) を

調査する．その結果，もし非零のサポートが唯一の標準ジェスチャパターンで存在した場合，現在の認識結果をそのジェスチャで置き換える．

図 10 は本手法により連続したジェスチャ動作を認識した結果である．想定したジェスチャは「さようなら」「万歳」「指差し」の 3 種である．いずれも $I = 300$ としたが，開始後 70 フレーム ($< I/2 = 150$) 程度で「万歳」のジェスチャを検出できていることがわかる．

4 おわりに

本報告では，現在我々が進めている「プロアクティブヒューマンインターフェース」の研究について紹介した．今後は現在開発している人間動作計測システム [10], [11] との統合や意図先読み手法の改良，および複数ヒューマノイドロボットを用いた多方向遠隔コミュニケーションシステムの開発を行う予定である．

参考文献

- [1] 横小路. テレオペレーションから機械メディアへの展開. システム/情報/制御, Vol. 44, No. 12, pp. 702–709, 2000.
- [2] 平松, 八木, 橋本, 谷内田. 対話者方向からのアビランズに基づくロボットアバターのためのジェスチャ生成. 日本ロボット学会誌, Vol. 21, No. 3, pp. 265–272, 2003.
- [3] 山田. 人間の行動意図を汲むロボットの技術. システム/情報/制御, Vol. 44, No. 12, pp. 696–701, 2000.
- [4] 山田, 山本, 酒井, 森園, 梅谷. メンテナブルな人間/ロボット共存システムによるヒューマン・エラー・リカバリー. 日本ロボット学会論文誌, Vol. 21, No. 4, pp. 86–92, 2003.
- [5] R. Kurazume, T. Hasegawa, and K. Yoneda. The sway compensation trajectory for a biped robot. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 925–931, 2003.
- [6] S. Kajita, O. Matsumoto, and M. Saigo. Real-time 3d walking pattern generation for a biped robot with telescopic legs. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 2209–2036, 2001.
- [7] S. Kajita, O. Matsumoto, and M. Saigo. A realtime pattern generator for biped walking. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 31–37, 2002.
- [8] 岡. 連続 DP を用いた連続単語認識. 日本音響学会音声研究会資料, S78-20, 1978.
- [9] 高橋, 関, 小島, 岡. ジェスチャー動画像のスポットイング認識. 電子情報通信学会論文誌, Vol. J77-DII, No. 8, pp. 1552–1561, 1994.
- [10] H. Yoshimoto, N. Date, D. Arita, and R. Taniguchi. Vision-based latency-free real-time human motion sensing. In *Proc. 10th Korea-Japan Joint Workshop on Frontiers of Computer Vision*, pp. 244–249, 2004.
- [11] 岩下, 倉爪, 辻, 原, 長谷川. Fast level set method を用いた複数移動物体の実時間追跡. 第 9 回ロボティクスシンポジウム講演会予稿集, p. 5B1, 2004.