

# プロアクティブヒューマンインターフェースの研究

## - 第3報 予測駆動型アクティブインターフェース実験 -

倉爪 亮 (九大) 内田 誠一 (九大) 谷口 倫一郎 (九大) 長谷川 勉 (九大)

### Study on Proactive Human Interface - Experiments of Prediction-based Active Interface -

\*Ryo Kurazume (Kyushu Univ.), Seiichi Uchida (Kyushu Univ.), Rinichiro Taniguchi (Kyushu Univ.), and Tsutomu Hasegawa (Kyushu Univ.)

**Abstract**— 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 communication experiment of the prediction-based active interface using the humanoid-type 2 legged robot named “PICO”.

**Key Words:** Proactive System, Prediction, 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 [1]。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。

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

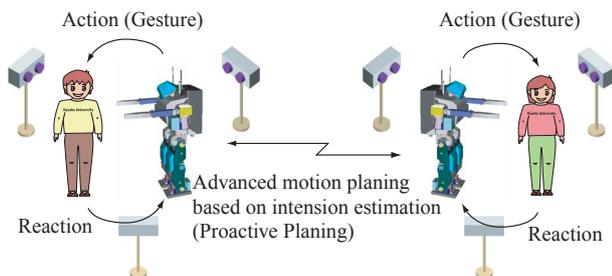


Fig.1 Concept of proactive human interface

がコミュニケーションを行うとき，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 [2], [3] や，操作者の意図推定を基にした動作補助 [4], [5] 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と行動意図の推定

##### 2.1 人間型アクティブインターフェース，“PICO”

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アクティブインターフェース “PICO” を試作した。このロボットは人間型ロボット (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，行動を計測，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行動を再現する。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 行動意図の早期認識

図1に示したように，本研究ではプロアクティブヒューマンインターフェースの一例として，“PICO” を介して遠隔地の人間が互いにコミュニケーションを行うタスクを想定している。その際，ネットワークの性能によっては，ユーザの行動情報の伝送に時間を要し，“PICO” による行動再現に遅れが発生する。また，口

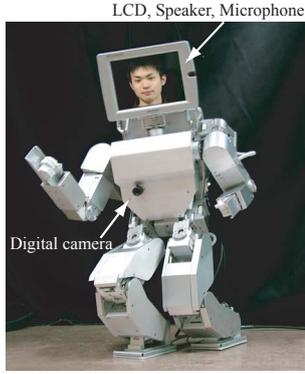


Fig.2 Embodied proactive human interface, "PICO"

ロボット自身の運動速度の限界によっても遅れが発生しうる。こうした遅れはさらに他方のユーザの行動の遅れを生むことになるので、円滑なコミュニケーションを著しく妨げる。

そこで本節では、この時間遅れを抑制する手法を提案する。この手法は、(i) 一方の人間の動作パターンを極力早期の段階で認識し、(ii) 認識結果を用いて今後の動作を予測し、(iii) その予測結果により他方の "PICO" を駆動する、という3段階からなる。例えば、まず人間が両手を挙げ始めた段階でそれを「万歳」と認識する。次に「万歳」の標準パターンを参照し、数秒後には両手が頭以上の高さまで上がるという予測を行う。最後にこの予測結果に基づいて "PICO" の動作を先回りして駆動する。この認識を利用した枠組みにより、単に現在の動作情報をそのまま使用した場合よりも時間遅れを低減できる。

本手法の説明に先立ち、その基本となる連続 DP [6, 7, 8] による動作 (ジェスチャ) 認識手法を説明する。以下では、システムにあらかじめ登録されてる標準ジェスチャパターンを特徴ベクトルの時系列  $R_{c,1}, \dots, R_{c,t}, \dots, R_{c,T_c}$  で表す。ここで  $c$  はジェスチャの種類を表す添字である。後述する実験では、各特徴ベクトル  $R_{c,t}$  は時刻 (フレーム)  $t$  での頭と両手の3次元位置からなる9次元ベクトルである。一方、認識対象とする連続的なジェスチャパターン (入力パターン) を、同様に特徴ベクトルの時系列  $I_1, \dots, I_\tau$  で表す。ここで  $\tau$  は現在の時刻を表す。このとき従来の連続 DP による認識では、次の漸化式を各入力フレーム  $\tau$  で計算する (図 3)。

$$g_{c,t}(\tau) = \min \begin{cases} g_{c,t-1}(\tau-1) + 3d_{c,t}(\tau) \\ g_{c,t-1}(\tau-2) + 2d_{c,t}(\tau-1) + d_{c,t}(\tau) \\ g_{c,t-2}(\tau-1) + 3d_{c,t-1}(\tau) + 3d_{c,t}(\tau) \end{cases} \quad (1)$$

ここで  $d_{c,t}(\tau)$  は入力パターンのフレーム  $I_\tau$  を標準パターンのフレーム  $R_{c,t}$  に対応させた場合の局所距離  $\|I_\tau - R_{c,t}\|$  である。漸化式 (1) に従って累積距離  $g_{c,t}(\tau)$  を各  $\tau$  のすべての  $c, t$  について計算することで、フレーム  $\tau$  での認識結果  $c^*$  が次のように得られる。

$$c^* = \underset{c}{\operatorname{argmin}} g_{c,T_c}(\tau) \quad (2)$$

この従来法を早期認識に用いるには問題がある。具体的には、認識規則 (2) を用いる限り、動作終了時点

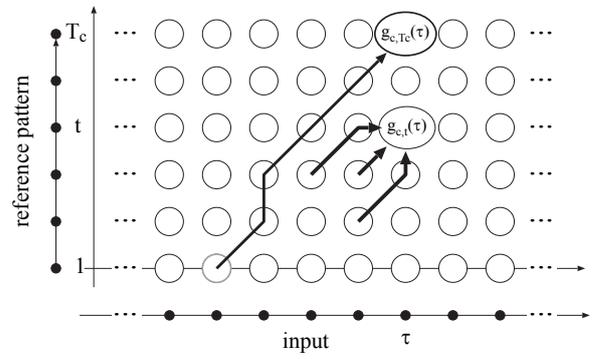


Fig.3 Gesture spotting based on continuous DP

付近になるまでその認識結果が得られない点が問題になる。これは標準パターン全体を入力パターン中にスポットティング認識していることに因る。

以上の従来法に対し本手法では、早期認識のために次の識別規則を用いる。

$$(c^*, t^*) = \underset{(c,t)}{\operatorname{argmin}} g_{c,t}(\tau)/t \quad (3)$$

識別規則 (2) との違いは、標準パターンの冒頭部の部分パターン  $R_{c,1}, \dots, R_{c,t}$  も認識の対象となっている点である。各部分パターンはその長さ  $t$  によって正規化された後、入力パターンと比較されることになる。この識別規則により、現フレーム  $\tau$  があるジェスチャ  $c$  の途中 (第  $t < T_c$  フレーム) だったとしても、その状況が認識結果  $(c^*, t^*)$  として正しく出力される。

今、現フレーム  $I_\tau$  が  $R_{c^*,t^*}$  に対応すると認識されたとすれば、 $\delta$  フレーム後には  $R_{c^*,t^*+\delta}$  に似た動作が現れると予測できる。すなわち、 $R_{c^*,t^*+\delta}$  が  $I_{\tau+\delta}$  の予測値となる。従って  $\delta$  をシステム全体の遅れ時間とすれば、 $I_\tau$  の代わりに  $R_{c^*,t^*+\delta}$  を他方の "PICO" に伝送することで、遅れの無い動作再現が可能となる。

識別規則 (3) を用いた際、2つのジェスチャ  $c$  と  $c'$  の部分パターンが類似していると、認識結果が  $c$  と  $c'$  の間で不安定になることがある。類似区間が比較的長い場合は、この不安定さはあまり問題にならない。これはそうした区間では  $R_{c,t+\delta}$  も  $R_{c',t+\delta}$  も類似しているの、どちらに認識されたとしても結局同じような予測値が得られるためである。むしろ非常に微小な区間 (数フレーム) だけが類似している場合の方が問題は顕著になる。例えば、ジェスチャ  $c$  を "vwxyz", ジェスチャ  $c'$  を "xyz" という、長さ3の時系列パターン (ここで  $v, w, \dots, z$  は特徴ベクトル) とする。この時、"vwxyzvwxyz" (ジェスチャ  $c$  の2回繰り返し) が入力されたとすると、第3フレームをジェスチャ  $c'$  の第1フレームと誤判定し、第4フレームの予測値として  $d$  を与えてしまう可能性がある。こうした微小部分パターンの曖昧性に因る不安定さを如何に解消するかについては今後の検討課題と考えている。

### 3. 遠隔コミュニケーションのための予備実験

前章で示した人間型アクティブインターフェース "PICO" と行動意図推定のためのジェスチャ認識技術を組み合わせ、意図推定に基づく行動予測駆動型アク

ティブインターフェースを試作し、遠隔コミュニケーションのための予備実験を行った。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複数カメラにより距離画像を計測し、肌色抽出により話者の手および顔の3次元位置を計測する(図4)。
2. 手と顔の相対位置の変化パターンを入力パターンとして、標準行動パターンと比較し、提案した早期認識の可能な連続 DP マッチング法により行動を分類する(図5)。
3. 分類された行動( $c^*$ )と、その行動の経過時間の予測値( $t^*$ )を“PICO”へ送信する。
4. “PICO”は、受信した行動と経過時間の予測値から、現時点及び次時刻での行動を予測し、再現する。
5. 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6に実験の様子を示す。操作者の動きを計測し動作を認識するシステムと“PICO”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を再現することも可能である。

次に計測、認識システムと“PICO”との間の通信に遅れがある場合、及び通信帯域が狭くデータが疎に送られてくる場合を想定し、(1) 認識システムを用いず、単純に計測された手先位置を送信し再現する手先追従フィードバック制御、(2) 現在の行動と経過時間を予測して送信し、それに基づいて行動を再現した場合の2種類について実験を行った。

まず、図7,8に遅れ時間の有無による動作再現性能の比較を示す。図7は遅れがない場合、図8は0.5秒だけ認識システムからロボットへの情報伝達に遅れがある場合であり、縦軸は右手の上下方向の位置、横軸は経過時間である。また図7,8の下図は、開始後8秒から25秒までを拡大したもので、万歳の動作(両手を挙げる)へ移る場面である。また両図とも、実際の人間の動作、単純な手先追従フィードバック制御、及び認識結果から予想された行動を再現した場合の3種類を比較している。ただし認識結果を用いる場合には、遅れ時間を補償するため、遅れ時間が既知であると仮定して、伝達された認識結果から0.5秒先の動作を再現している。

これより、単純なフィードバック制御では、情報伝達に遅れがある場合にはロボットの動作に定常的に遅れが生じるのに対し、認識結果から予想された行動を遅れを補償して再現することで、動作の遅れが動作開始後に徐々に解消できることがわかる(例えば図8の9秒目)。

次に図9にデータの転送周期を10分の1にした場合の動作再現性の比較を示す。図9下図は動作開始後15秒から30秒を拡大したものである。これより、転送周期が長い場合には、単純なフィードバックでは手先位置が急激に変化するが、認識結果を用いて動作を予測し再現することで、より滑らかな手先の動きが実現できることがわかる。



Fig.4 Measurement of human gesture for telecommunication system



Fig.5 Gesture recognition based on continuous DP matching

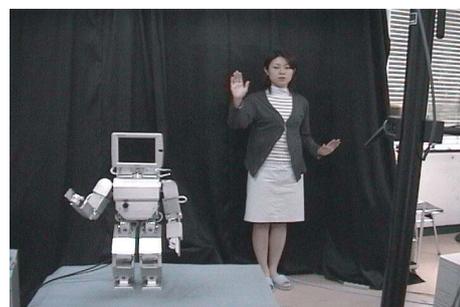


Fig.6 Imitation of human gesture

#### 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全身人間動作計測システム[9], [10]との統合やより長い周期での意図先読み手法の開発、フィードフォワード動作とセンサフィードバック制御の適切な切り替え、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] 倉爪, 内田, 長谷川, 谷口. 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 -. 日本機械学会ロボティクス・メカトロニクス講演会'04 講演論文集, pp. 1A1-H-76, 2004.
- [2] 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.
- [3] 平松, 八木, 橋本, 谷内田. 対話者方向からのアビアラン

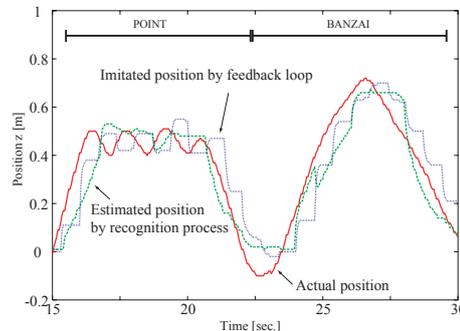
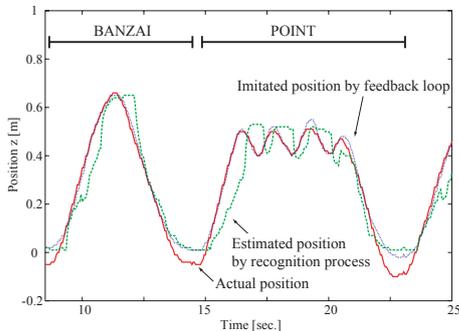
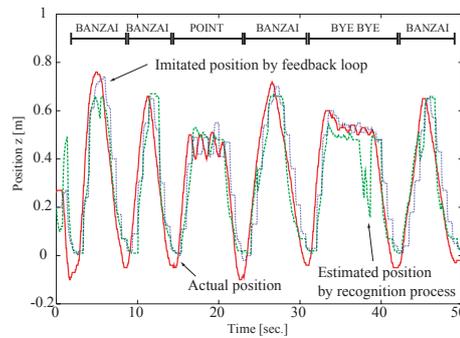
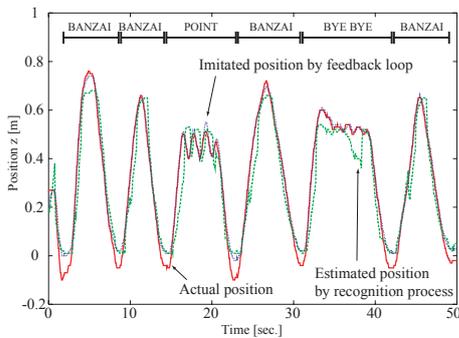


Fig.7 Imitation of human gesture in the case of no delay time

Fig.9 Imitation of human gesture with coarse data

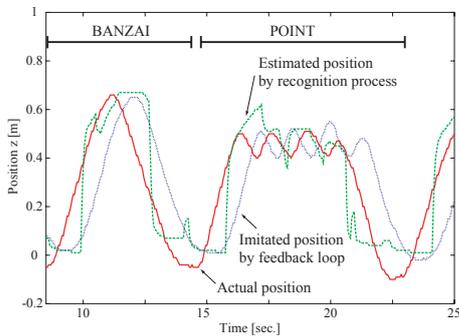
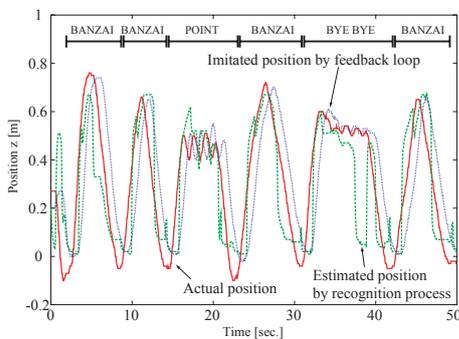


Fig.8 Imitation of human gesture in the case of 0.5 [s] delay time

- バリ。日本ロボット学会論文誌, Vol. 21, No. 4, pp. 86-92, 2003.
- [6] 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.
  - [7] 高橋, 関, 小島, 岡. ジェスチャー動画画像のスポットティング認識. 電子情報通信学会論文誌, Vol. J77-DII, No. 8, pp. 1552-1561, 1994.
  - [8] 呉, 木戸, 塩山. ジェスチャー認識のための連続 DP の改良. システム制御情報学会論文誌, Vol. 14, No. 6, pp. 283-290, 2001.
  - [9] 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.
  - [10] 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.

スに基づくロボットアバターのためのジェスチャー生成. 日本ロボット学会誌, Vol. 21, No. 3, pp. 265-272, 2003.

- [4] 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.
- [5] 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