

## ジェスチャー入力によるマルチモーダルインタラクティブシステムの試作

今野 潤<sup>†</sup>, 八木 正紀<sup>†</sup>, 山内 英明<sup>†</sup>, 安村 通晃<sup>†</sup>日立システムエンジニアリング<sup>†</sup>  
慶應義塾大学 環境情報学部<sup>†</sup>

次世代インターフェイスとして注目を集めているマルチモーダルインターフェイスでは、ジェスチャーや音声を入力手段として用いる。その研究の一環として、ユーザがコンピュータへの入力をジェスチャーによって行なう場合に、円滑なインタラクションをどのように実現すればよいかについて研究を行なった。

ジェスチャーを入力手段としたリアルタイムなインタラクティブシステムを試作し、被験者による評価実験を行なった。その結果、円滑なインタラクションを行なうためには、原則としてユーザのジェスチャー入力の様子を実画像で画面上に表示すること、ユーザへのフィードバックには複数のモダリティを併用すること、などが重要であることが分かった。

Prototyping of a Multimodal Interactive System  
through gesture-inputJun Konno<sup>†</sup>, Masanori Yagi<sup>†</sup>, Hideaki Yamauchi<sup>†</sup>, Michiaki Yasumura<sup>†</sup>Hitachi System Engineering<sup>†</sup>  
Keio University, Faculty of Environmental Information<sup>†</sup>

Multimodal interface, which attracts one's attention as a next generation interface, uses gesture-input and voice-input. We have designed and implemented a multimodal system which realizes the smooth interaction through gesture-input.

We have also made several applications and evaluated for them. We found that realistic image is important for the gesture-input and feedback to the user should use both sound and image.

# 1 はじめに

GUI (Graphical User Interface) に代わる次世代インターフェイスの候補として、マルチモーダルインターフェイスが注目を集めている。マルチモーダルインターフェイスは、音声やジェスチャーといった複数のモダリティ (伝達様式) によって情報の入出力を行なうインターフェイスである。コンピュータに疑似的な目や耳・口を持たせることで、人とコンピュータとのコミュニケーションの幅が広がり、かつコンピュータが積極的に人とのインタラクションを行なえるようになる。

このように、マルチモーダルインターフェイスでは様々なモダリティを用いて情報の伝達が行なえる。しかし、新たに導入したモダリティの活用の仕方が適切でない、かえってインタラクションの円滑な進行を妨げるおそれがある。そこで、ユーザがコンピュータへの入力にジェスチャーを用いる場合に、円滑なインタラクションをどのように実現すればよいかについて研究を行なった。

## 2 マルチモーダルインターフェイス

### 2.1 従来のインターフェイスとの比較

マルチモーダルインターフェイスは、従来のインターフェイスで用いられていたモダリティに加えて、音声やジェスチャーといったモダリティを活用する (図1)。この中で、音声やジェスチャーは、通常、人同士が意志の疎通を行なう際に用いるモダリティであり、対面コミュニケーションの成立に欠かせないものである。そのようなモダリティに対応することで、従来のインターフェイスと比較して、次のような利点がある。

- (1) 自然なインタラクションが可能である。
- (2) 新たに操作方法を学習したり、慣れたりする必要がない。
- (3) 操作対象に対する直接的な操作感が増す。
- (4) 身体障害者などの、利用可能なモダリティの限られる人に、代替手段を提供可能である。

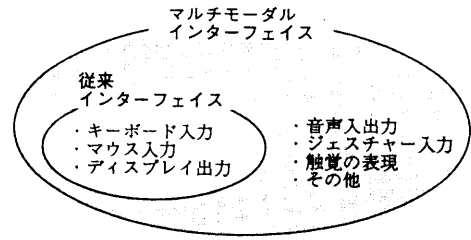


図1: インターフェイスとモダリティの関係

### 2.2 フィードバック

インターフェイスにおいて重要な要素の1つとして、フィードバックが挙げられる。

従来のインターフェイスでは、入力の自由度が狭く、入力装置によって伝達可能な情報の量が限られているため、入力によって生じるフィードバックも単純なものである (表1の a. と b.)。

しかし、マルチモーダルインターフェイスで扱う音声やジェスチャーといったモダリティでは、入力の自由度が大きいため、キーを押す行為やマウスを動かす行為に比べて入力される情報の種類が増えるとともに、情報の内容が複雑になる。そのため、どのような入力が行なわれたかをコンピュータで把握し、それをフィードバックによって適切にユーザに伝えることが難しくなる。

以下に、音声やジェスチャーが通常持っている情報の種類を示す。

- ・音声 … 話している内容、声の抑揚や音色
- ・ジェスチャー … 顔や手、身体の3次元の位置や形状、動き

表1: 各モダリティと対応するフィードバック

モダリティ	対応するフィードバックの例
a. キーボード	キーに対応した文字を表示
b. マウス	マウスカーソルを移動 クリックしたアイコンの選択
c. 音声	ユーザの話した言葉を表示
d. ジェスチャー	ユーザの入力の様子を表示

## 2.3 本研究で扱うモダリティ

本研究では、マルチモーダルインターフェイスで活用されるモダリティのうち、扱う対象をジェスチャーにしばって研究を行なった(図2)。今回試作したインタラクティブシステムでは、手や身体を使ったジェスチャーによって、ユーザがコンピュータとインタラクションを行なう。

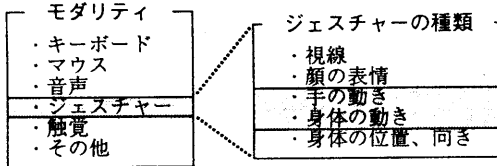


図2: 本研究で扱うジェスチャーの位置付け

## 2.4 ジェスチャーの入力方式

マルチモーダルインターフェイスにおけるジェスチャーの入力方式には、装着型と環境型がある。

装着型は、ユーザがデータグローブやデータスーツ、ヘッドマウントディスプレイ等を身に付ける方式であり、環境型は、ユーザの周囲の環境にビデオカメラなどの入力装置を設置する方式である。

本研究では、ユーザの動作が拘束されず、ユーザが自然な状態で入力できるという利点から、ジェスチャーの入力に環境型の方式を採用する。

## 3 ジェスチャーによるインタラクション

### 3.1 インタラクティブシステムの構成

ジェスチャーによるインタラクションを実現するためのインタラクティブシステムの構成を図3に示す。

#### (1) 入力部

環境型の方式として、ビデオカメラ(INDYに付属のIndyCam)によってジェスチャーを画像データとして入力する。画像データの

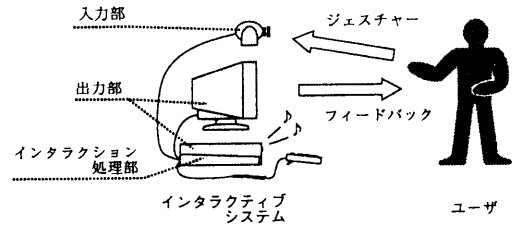


図3: インタラクティブシステムの構成

解像度は320×240ドットで、各ドットについて24bitの色情報(1677万色分)をもつ。

#### (2) インタラクション処理部

グラフィックスワークステーションINDYによって、ユーザからのジェスチャーを判定し、インタラクションの流れを制御する。

#### (3) 出力部

フィードバックには画像と音を利用するため、ディスプレイとINDY内蔵のスピーカーを用いる。

### 3.2 フィードバックがタスクの妨げになるインタラクション

ユーザがジェスチャーによる入力を行なっている場合には、そのジェスチャーの入力の様子を画面に表示することでフィードバックを行なうのが一般的である。しかし、ユーザが行なっているタスクの内容によっては、そのようなジェスチャーの入力過程のフィードバックがタスクの遂行の妨げになる場合があると考えられる。

フィードバックがタスクの妨げになる場合としては、以下のようなものが考えられる。

(1) 画面上に重要な情報が表示されていて、その画面にユーザのジェスチャー入力の過程などを重ね合わせて表示すると、情報取得の妨げとなる場合。

(2) タスクの内容が、観客に対してなにかを紹介するような種類のものであって、ユーザのジェスチャー入力の様子を表示すると、観

客の注意がそちらにそれてしまうおそれがある場合。

以上のことを考慮に入れ、本研究では、フィードバックがタスクの妨げになる場合とならない場合の2種類のインタラクティブシステムを試作する。

## 4 ジェスチャーの判定

本研究では、ユーザとコンピュータとが円滑にインタラクションを行なえることを重視するので、リアルタイム性を優先し、ジェスチャーの認識については正確さを追求しないことにする。そこで、リアルタイム性を実現するために、ユーザの動作を2次元のシルエットとして抽出し、そのシルエットからジェスチャーを判定した。

ユーザのシルエットの抽出（及び画像データの解像度の圧縮）は背景差分方式であり、内藤らの方式 [3] と基本的には同一である。ジェスチャーとしての判定は、抽出したシルエットの一部をインタラクションの種類に応じて局所的に参照することにより、実現した。

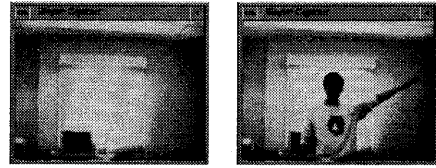
### 4.1 ユーザのシルエットの抽出

背景のみが映っている画像データと、同一背景に人が映った画像データとの差分をとる方法により、人と思える領域をシルエットとして抽出する。

ビデオカメラから入力される最初の画像データ（フレーム 0）には、人が映っていないものとする。ビデオカメラの視界に人が入ってくると、その時点で入力される画像データ（フレーム  $i$ ）において、人が映っている領域の色がフレーム 0 の色と異なる。そこで、フレーム 0 とフレーム  $i$  の対応する各ドットの色を比較することにより、画像データから人のシルエットを抽出することができる（図 4）。

### 4.2 画像データの解像度の圧縮

シルエットの抽出をする際に、画像データの解像度を  $320 \times 240$  ドットのままでフレーム 0 とフレーム  $i$  の比較を行なうと、ドット数が多いため計算に時間がかかる。そこで、計算量を減らした



(a) フレーム 0

(b) フレーム  $i$



(c) 2つのフレームの差分

図 4: シルエットの抽出

めに、画像データを  $10 \times 10$  ドットの領域ごとに 1 マスとして圧縮し、解像度を  $32 \times 24$  マスとした。

この結果、今回試作したインタラクティブシステムでは約 15 フレーム/秒の画像処理速度を実現している。

### 4.3 シルエットからのジェスチャー判定

本研究においては、インタラクションがユーザ - 仮想オブジェクト間で行なわれるものであることに着目し、正確さには欠けるが計算量が少なく高速なアルゴリズムによってジェスチャーを判定する。

インタラクションがユーザ - 仮想オブジェクト間で行なわれるものであるため、仮想オブジェクトの周辺にユーザのシルエットが存在する場合は、仮想オブジェクトがユーザの操作または存在により影響を受ける（図 5）。

## 5 システムの試作

### 5.1 フィードバックがタスクの妨げにならないインタラクティブシステム

図 6 に、試作したインタラクティブシステムの応用例を示す。ユーザのジェスチャー入力過程を、フィードバックとして画面に表示する時の表示パターンは、2種類用意した（詳細は 6.1 節で述べる）。

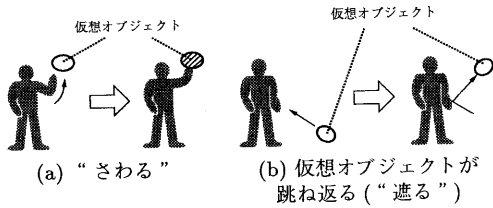


図 5: 仮想オブジェクトとユーザの関係

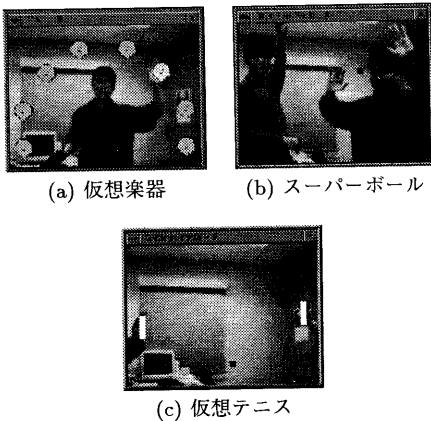


図 6: 試作したインタラクティブシステム (1)

(a) 仮想楽器

ドラムに似せた仮想オブジェクトをユーザが叩くことで、演奏を行なう。

最初に試作した時は、ドラムを叩いても音になるだけで、叩くことによる視覚的なフィードバックはなかった。

(b) スーパーボール

画面中を跳ね回っているボールを、ユーザが身体で遮ったり手で潰したりする。

初期には、ボールが画面の隅やユーザの身体に当たって跳ね返るのが見える以外に、ボールが跳ね返ることによる視覚的・聴覚的なフィードバックはなかった。

(c) 仮想テニス

画面の左右に表示されているラケットをユーザが手で上下に動かして、ボールを打ち合う対戦型のインタラクション。

仮想オブジェクトが返すフィードバックについては、ボールが跳ね回るという内容から、スーパーボールと基本的に同じである。

## 5.2 フィードバックがタスクの妨げになるインタラクティブシステム

3.2節で述べたタスクの性質をもとに、画面全体を使って絵を描くツールや、プレゼンテーション用のスライド切替えツールを応用例とするインタラクティブシステムを試作した(図7)。いずれも、ユーザのジェスチャー入力過程をフィードバックとして画面に表示する時の表示パターンを複数用意した(詳細は6.2節で述べる)。

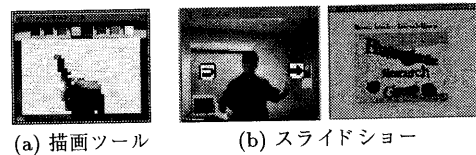


図 7: 試作したインタラクティブシステム (2)

(a) 描画ツール

指先を動かして、画面上に絵を描いていく。色の指定は、画面の上部のカラーパレットの選択により行なう。

色を指定した時や、絵を描き始めた時・描きやめた時に、視覚的・聴覚的なフィードバックにより状態の変化を知らせる。

(b) スライドショー

手で画面上のスライド操作ボタン(左側の window 内の仮想オブジェクト)を操作して、スライド(右側の window に表示されている絵)を切り替えていくプレゼンテーション用のインタラクティブシステム。

スライド操作ボタンを押した時に、視覚的・聴覚的なフィードバックが生じる。

## 6 試作したインタラクティブシステムの評価実験

### 6.1 フィードバックがタスクの妨げにならないインタラクティブシステム

数人の被験者に、5.1節で挙げたインタラクティブシステムを試用し、評価をしてもらった。

以下に各インタラクション固有の問題を挙げる。

#### (a) 仮想楽器

ドラムを叩いても、視覚的なフィードバックが生じないため、どのドラムを叩いたのかが分かりにくい。

このため、叩かれたドラムの色を変えて視覚的なフィードバックを追加することにより、解決した。

#### (b) スーパーボール

ボールが画面の隅やユーザの身体に当たって跳ね返るのは見えるが、跳ね返る際にボールに変化が見られず、音もしないため、ボールの存在感が感じられない。

このため、ボールが跳ね返る時には色を変化させるとともに、効果音を発生させることで視覚的・聴覚的フィードバックを追加して、解決した。

#### (c) 仮想テニス

ラケットの操作法が、一般の人が思い浮かべる操作法と食い違っていたため、操作に関するアドバイスなしでは被験者がラケットを自由に操作することができない。

これについては、操作法の変更により解決した。

ジェスチャーの入力過程をどのようにユーザにフィードバックするのが良いかという点については、シルエットのみ(図8(a))よりも実画像(ユーザの鏡像)が画面に表示されるパターン(図8(b))が良いという結果が得られた。

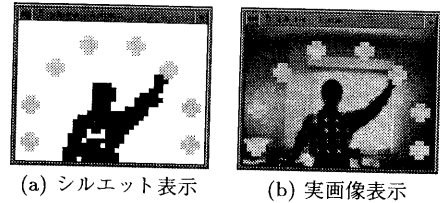


図 8: フィードバックの種類 (仮想楽器の例)

### 6.2 フィードバックがタスクの妨げになるインタラクティブシステム

12名の被験者に、5.2節で挙げたインタラクティブシステムを試用し、評価をしてもらった。

#### (a) 描画ツール

ジェスチャーの入力過程のフィードバックに関する表示パターンとして、図9の3種類を用意した。パターンAは手の指先の位置のみをカーソルで表示、パターンBは手の部分を薄暗いシルエット(画面の情報が透けて見える)で表示、パターンCは手の部分を実画像で表示している。

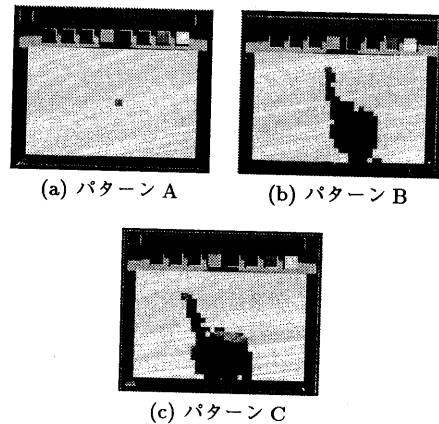
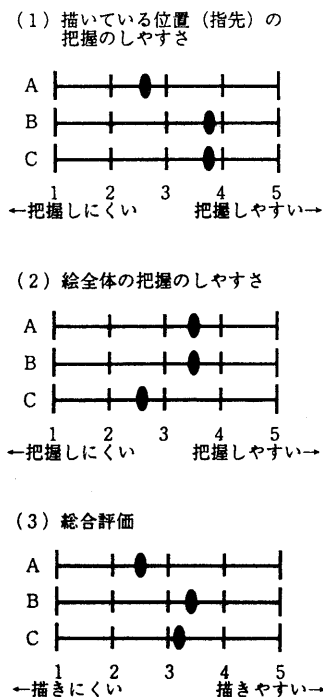


図 9: 描画ツールのフィードバックの種類

アンケートの5段階評価の集計結果を、表2に示す。表の値は、各項目についての被験者の評価の平均値である。表から、描いている位置を把握しやすいのはパターンAと

B、絵全体の様子を把握しやすいのは A と B であることが分かる。総合では、B が C に比べて若干絵を描きやすく、A は絵を描きにくいという結果が出ている。

表 2: 描画ツールに対する 5 段階評価



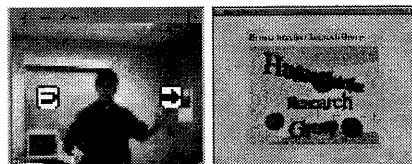
インタラクションに関しては次のような評価が得られた。

- 手や指の動きで、状態を変えられた方が良い。
- 片手でジェスチャー入力をするだけでなく、空いている手でキーボードやマウスを扱ったり、音声を併用できると良い。

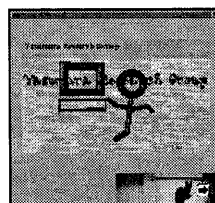
(b) スライドショー

ジェスチャーの入力過程のフィードバックに関する表示パターンとして、図 10 の 3 種類を用意した。パターン A はスライドを表示する window と操作ボタンのある window

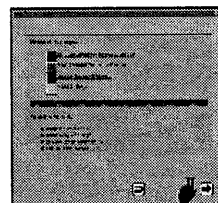
が別、パターン B, C はスライドと操作ボタンが同じ window である。ただし、パターン A, B ではユーザとその周辺の背景を全て実画像で表示、パターン C ではユーザの部分のみ実画像で表示している。



(a) パターン A



(b) パターン B



(c) パターン C

図 10: スライドショーのフィードバックの種類

アンケートの 5 段階評価の集計結果を、表 3 に示す。表の値は、各項目についての被験者の評価の平均値である。表から、スライドの切替え操作をしやすいのはパターン A と C であることが分かる。また、観客がスライドに集中しやすいのは、A であることが分かる。

さらに次のような意見も得られた。

- 観客の立場でいえば、スライドを操作しているところが見えていない方がスライドに集中しやすい。

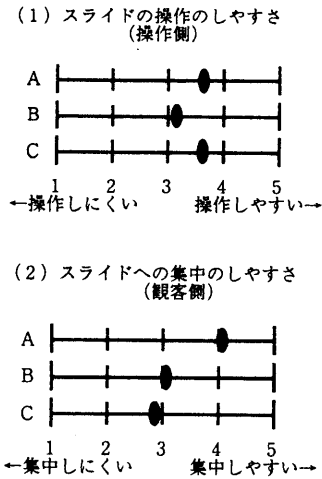
## 7 考察

評価実験の結果から、ユーザとコンピュータとで円滑なインタラクションを行なえるようにするためには、以下のことが重要であると思われる。

- (1) ユーザのジェスチャー入力の様子を表示する際には、実画像を用いる。

ユーザは、ユーザ自身の目に見えるもの

表 3: スライドショーに対する 5 段階評価



が自然であることを好むため、コンピュータの把握しているシルエットを表示するよりも、ビデオカメラで入力した実画像を表示する方が良い。

しかし、3.2節で述べたように、そのような表示がタスクの妨げとなる場合には、表示に関して特別な配慮が必要になる。

- (2) 複数のモダリティを併用して状態の変化をユーザに伝える。

仮想楽器やスーパーボールにおける結果が示すように、視覚もしくは聴覚単独のフィードバックだけでは、ユーザに伝わる仮想オブジェクトの存在感が不十分である。

- (3) ユーザが直観的に理解できる操作法にする。

仮想テニスでは、ラケットの操作法がユーザの予想するものと違っていた。そのため、ユーザが円滑なインタラクションを行なうことができなかった。

- (4) 入力のためのモダリティは複数用意する。

ユーザは、ジェスチャーがメインの入力手段の場合でも、補助的に他のモダリティを使用できることが望ましい。

## 8 まとめ

ユーザが入力するジェスチャーによって、リアルタイムでインタラクションを行なうことが可能なインタラクティブシステムを試作した。

さらに、試作したインタラクティブシステムの被験者による試用評価の結果、ジェスチャーによってユーザとコンピュータが円滑にインタラクションするために重要となる要因を、明らかにすることができた。

今後は、現在よりも複雑なジェスチャーの入力を行なえるようにするために、より高度なジェスチャーの認識方法を用いていくとともに、本研究がマルチモーダルインターフェイスの研究の一環であることから、ジェスチャーだけでなく他のモダリティも結合したインタラクションを考えていかなければならない。

## 9 謝辞

評価実験に快く参加して下さい、慶應義塾大学湘南藤沢キャンパスの学生の方々に感謝します。

## 参考文献

- [1] D.A. ノーマン、野島久雄訳：誰のためのデザイン？ 認知科学者のデザイン原論、新曜社 (1990).
- [2] M.W. クルーガー、下野隆生訳：人工現実 — インタラクティブ・メディアの展開一、トッパン (1991).
- [3] 内藤剛人、竹内彰一、所真理雄：視線を伴った表情合成システム、日本ソフトウェア科学会 WISS'93(1993).
- [4] 黒川隆夫：ノンバーバルインタフェース、オーム社 (1994).
- [5] 黒川隆夫：機械とのジェスチャー・コミュニケーションのデザイン、電子情報通信学会技術研究報告 [ヒューマン・コミュニケーション]、HC-93-82、pp.19-24(1994).