

ヒューマンインタフェースとしてのジェスチャー認識

澤田 秀之^{1*}、橋本 周司¹、松島俊明²

¹ 早稲田大学 理工学部 応用物理学科

〒169 東京都新宿区大久保3-4-1

TEL:(03)5286-3233 FAX:(03)3202-7523

e-mail: sawa@shalab.phys.waseda.ac.jp, shuji@shalab.phys.waseda.ac.jp

² 東邦大学 理学部 情報科学科

〒274 千葉県船橋市三山2-2-1

e-mail: matusima@is.sci.toho-u.ac.jp

* 日本学術振興会特別研究員

あらまし：意志や感情の伝達においては、言語以上に非言語的手段によるところが大きいといわれる。身振り手振りなどのジェスチャーはその代表的なものである。データグローブ、ポジションセンサおよび小型3次元加速度センサを用いて、5指の屈曲、空間的な手の配置、腕のダイナミカルな動きを検出することにより、ジェスチャー認識をおこなった。現在はジェスチャーによる柔軟なヒューマンインタフェース実現に向け、手話の認識およびデータベースの構築、音楽制御への応用を試みている。

A Study of Gesture Recognition as Human Interface

Hideyuki Sawada^{1*}, Shuji Hashimoto¹ and Toshiaki Matsushima²

¹ Department of Applied Physics, School of Science and Engineering, WASEDA University
3-4-1, Okubo, Shinjuku-ku, Tokyo, 169, Japan

² Department of Information Sciences, TOHO University

* Research Fellow of the Japan Society for the Promotion of Science

Abstract: Gesture plays an important role in our daily life as nonverbal media for emotional human communication. This paper proposes a gesture recognition algorithm using datagloves, position sensors and acceleration sensors to understand hand postures and human dynamic movements. Two experimental systems are also described for the realization of new human-machine interface through gestures.

1. はじめに

我々の日常生活における情報伝達においては、身体メディアによる非言語的手段が音声による言語的手段に劣らぬ機能を持っているといわれる[1][2]。身振り手振りなどのジェスチャーはその代表的なものであり、特に意志や感情の伝達においては大きな役割を果たしている。ジェスチャーは日常の対話において自然に現れるばかりでなく、言語障害者あるいは言葉の通じない外国人の間の会話では主要なコミュニケーション手段となっている。また音楽では、指揮者は指揮法に基づいた身振りや指揮棒の動きによって楽音表現を演奏者に伝え、演奏者は楽器を通してこの表現を音響に変換する。さらに、演奏者の楽器は動作から音響への感性的な変換器であるといえる。通常、人間に対して用いられるこれらのジェスチャーが機械によって理解できれば、伝達者の感情や情緒などの微妙な表現にも対応できる自然な入力手段となり得るであろう。

筆者らはこれまでに画像処理やデータグローブ、加速度センサを用いたジェスチャー認識手法を検討してきた[3]-[7]。前者の2つは、空間的な手の配置あるいは5指の屈曲などを基にしてジェスチャーを認識するものである。一方後者では、人間の意志や感情が、手などの位置よりもむしろ身体に加えられる力に顕著に現れると考え、運動加速度の直接検出によりジェスチャー認識も試みた。

本論文では、これら3次元加速度センサとデータグローブによる手法を統合したジェスチャー認識について述べる。さらに実時間手話認識への拡張と手話単語データベースの構築、およびジェスチャーによる音楽制御への応用について報告する。ジェスチャーの形状ばかりでなく動作のダイナミクスを考慮することにより、従来手法に比べてより多くのジェスチャー認識が可能となったばかりでなく、従来より問題となっていたジェスチャー

の始まりと終わりの検出もロボストに行うことが可能となった。

2. ジェスチャー計測と特徴パラメータ

2-1. 3次元加速度センサ

加速度の計測には、ジャイロをはじめとして様々な手法が用いられているが、ジェスチャー認識に用いるには、手軽に扱えて、かつ手先などに装着した場合に自由な運動が行える必要がある。また、腕の運動加速度は極めて大きく通常の運動でも最大値は20Gを越えるため、測定レンジも大きくなければならない。ここではおもりの変位を測定する簡単な加速度センサを用いている。

図1にその内部構造を示す。大きさは縦20mm、幅15mm、高さ12mmで、重量は12.5gと小型軽量なため、容易に体の一部に取り付けることができる。センサ内部にはx, y, z軸方向の加速度を独立に検出するために、各々の軸方向で金属おもりが圧電素子に接続してある。測定範囲は、-25~+25Gであり、加速度にともなう圧電素子の電圧が出力される。この出力は増幅されA/D変換ボードを通じて12ビットのデジタル値として計算機に取り込まれる。感度は1Gあたり約5mVである。なおドリフト除去のために、遮断周波数3Hzの低域除去フィルタをアンプ入力前に通している。

加速度センサからは、x, y, z方向の加速度値が $a_x(t)$, $a_y(t)$, $a_z(t)$ として得られる。なお本研究でセンサを手の甲に固定して使用する場合には、指先の向きがセンサx軸の負の向きと

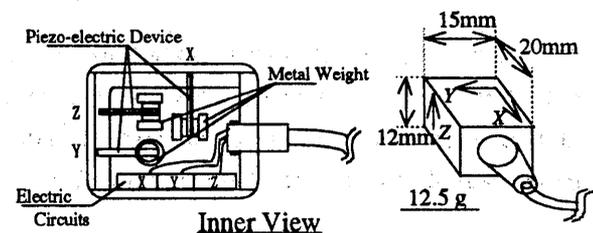
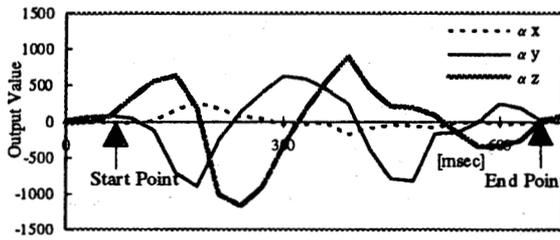
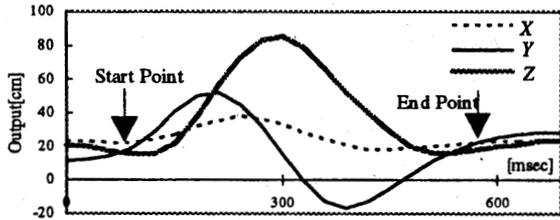


図1 3次元加速度センサの構造



(a) 加速度センサの出力



(b) ポジションセンサの出力

図2 「円を描く」動作による各センサの出力例

なるよう統一している。図2(a)に、y-z面で円を描く動作をおこなったときのセンサ出力値の例を、後述のポジションセンサからの出力値との比較で示す。各軸方向に取り付けられたおもりの変位を測定するという本センサの構造上、3軸の加速度値の分離が不十分で干渉がみられる。この3次元加速度ベクトルの時系列パターンから、ジェスチャーの認識を行うのであるが、ダイナミカルな動作から得られる加速度データの場合、単なるベクトル系列のパターンマッチングでは高い認識率が得られない。ジェスチャー運動の大域的な特徴を抽出する必要がある。ここでは運動の直感的な把握を容易にするために、次のような3つの2次元ベクトルに基づいて運動の特徴を抽出している。

$$\begin{aligned} \mathbf{A1}(t) &= (a_y(t), a_z(t)), \quad \mathbf{A2}(t) = (a_z(t), a_x(t)) \\ \mathbf{A3}(t) &= (a_x(t), a_y(t)) \end{aligned} \quad (1)$$

ここで $\mathbf{A1}(t)$ 、 $\mathbf{A2}(t)$ 、 $\mathbf{A3}(t)$ は、それぞれy-z平面、z-x平面、x-y平面へ射影した加速度ベクトルである。データ取り込み周波数 f_{in} は30Hzとしている。これは急激に方向や速度を変えるような運動の計測においては不十分な

サンプリング周波数ではあるが、ここでは最小限の加速度データを用いて実時間でジェスチャー認識を行うことに重点をおいた。図3に、y-z面の加速度軌跡の例を示す。

各射影加速度ベクトルの時系列から、表1に示す11個の運動特徴パラメータの抽出を行う。ここでは、実時間ジェスチャー認識を可能とするために最低限必要と思われる運動の特性として、運動によって生じる加速度変化の大きさ、回転運動の方向、動作の向きを用いている。ここで得られる特徴は、主として動きの動的特性となる。

表1 加速度センサの特徴パラメータ

Pd	動きの激しさの特性 (加速度ベクトルの変化)
Pg	回転方向特性 (ベクトル間の外積)
Pr	主動作の方向特性 (ベクトル外接四角形の縦横比)
Pa0 ~ Pa7	方向の分布特性 (ベクトルの方向分布)

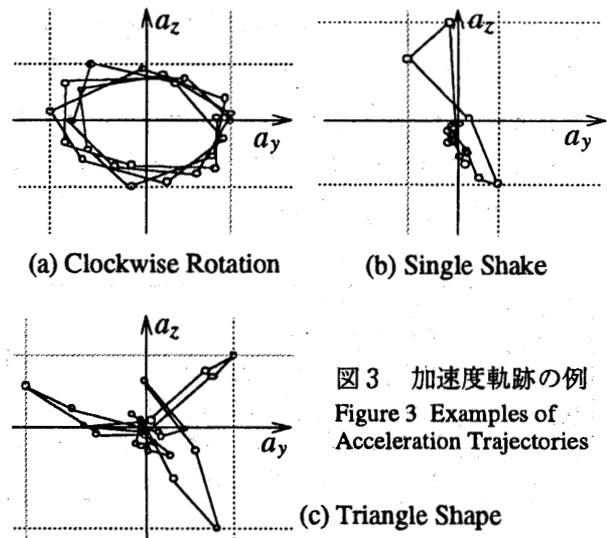


図3 加速度軌跡の例
Figure 3 Examples of Acceleration Trajectories

2-2. データグローブ

前述の加速度センサは身体のダイナミカルな動きを計測するには適しているが、ジェスチャー、特に手話においては手指の静的な形

状も大きな意味を持っている。そこで本システムでは、データグローブとポジションセンサを用いて手の形状および3次元位置の実時間計測をおこなっている。本データグローブは、曲げセンサを指裏側部分に配置することによって指の曲げ角に応じた出力が得られる。更に手の甲部分に固定されたポジションセンサによって、手の3次元位置 x, y, z と、姿勢としてピッチ、ロール、ヨー角も得ることができる。ポジションセンサによる3次元位置座標は、地上を基準とする人間の座標系と一致し、鉛直上方をz軸正の向き、手前側をx軸正の向き、左をy軸正の向きと設定した。

データグローブ及びポジションセンサからの出力をもとに、表2に示す18個の特徴パラメータが抽出される。これらのパラメータは、あらかじめ使用者の体型に合わせてジェスチャー空間を設定することによって、規格化されたものである。使用者はまず、前後、左右、上下の位置限界値と最大運動速度の入力を行った上で、手の初期位置の設定を行う。ポジションセンサから得られる位置パラメータR5-R7は、この初期位置を基準として算出される。さらに手形状のパラメータR0-R4は、あらかじめ指の曲げ伸ばし動作を行うことによって規格化が行われる。

表2 データグローブの特徴パラメータ

R0-R4	五指の曲げ角
R5-R7	手の3次元位置
R8-R10	過去10入力からの平均速さ
R11-R13	過去10入力からの速度絶対値の平均
R14-R16	手の姿勢(ピッチ、ロール、ヨー角)
R17	直線軌跡からのズレ

3. ジェスチャー認識

ジェスチャー認識には、加速度ベクトル及びデータグローブから得られる特徴パラメータを利用し、認識アルゴリズムは個人対応型としている。システムを使用する前にあらかじめ学習モードで個人の運動特徴の標準パ

ターンを作成してから認識モードに入る。これにより特定動作者の認識が実現できるばかりでなく、認識させたいジェスチャーを自由に選ぶことができる。

まず、学習モードで認識させたい動作のサンプルを M 回づつ入力し、前記の各運動特徴パラメータについての平均値 E_{α}^g および標準偏差 μ_{α}^g を次式で求め、各ジェスチャー g に対する標準パターンデータを作成する。ここでは試行回数 M は5とした。

$$E_{\alpha}^g = \frac{1}{M} \sum_{i=1}^M V_{\alpha}^{g,i} \quad (2)$$

$$\mu_{\alpha}^g = \sqrt{\frac{1}{M} \sum_{i=1}^M (V_{\alpha}^{g,i} - E_{\alpha}^g)^2} \quad (3)$$

V : Parameter Values
 α : P's and R's

認識モードでは未知動作を行うたびに各パラメータについて標準パターンの平均値 E_{α}^g との二乗誤差を計算し、それぞれ標準偏差 μ_{α}^g の二乗つまり分散値で割って規格化したものを重み付き誤差とする。これを各標準パターンごとに特徴パラメータ α について加え合わせたものをパターンの相違度 e_g とする。

$$e_g = \sum_{\alpha} \epsilon_{\alpha}^g = \sum_{\alpha} \frac{(V_{\alpha}' - E_{\alpha}^g)^2}{(\mu_{\alpha}^g)^2} \quad (4)$$

V' : Parameter Values

相違度が最小で、かつ下式(5)に従って決めるしきい値 T_h よりも小さいものを標準パターンから選び認識結果とする。 T_h は、対象とするジェスチャーの標準パターン間の最小距離に相当する。

$$T_h = \min_{i \neq j} \left\{ \sum_{\alpha} \frac{(V_{\alpha}^{g_i} - V_{\alpha}^{g_j})^2}{(\mu_{\alpha}^{g_j})^2} \right\} \quad (5)$$

for all i, j
 g_i, g_j : Gesture i, j

縦振り、横振り、斜め振り、時計回り、反時計回り、強いひと振り、星を描く、三角形を描く、ハートを描く、方向指示、胸の位置

に静止、といったジェスチャーについて認識実験をおこなった結果、100%の認識が可能であった。日常生活でよく用いられる手振りは、これらのジェスチャーのうち、縦振り、横振り、斜め振り、回転、強いひと振り、方向指示といった単体の動作、或いはそれら幾つかの組み合わせが多いと考えられる。一方で、画像処理などを用いた場合に実時間認識が困難と思われる、星形、三角形、ハート型といったジェスチャーが3次元加速度パターン及びポジションデータから認識可能であることがわかった。

4. ジェスチャー認識の応用

4-1. 手話データベースの構築と

手話単語認識

現在日本語手話には、約3000単語あるといわれている。しかし多くの手話は、いくつかの単純なジェスチャーの組み合わせで表すことができる[8]。図4に手話で用いられる代表的な動作方向を示した。今回はこの11の動作方向を基本動作として取り上げ、これに手形状および動作位置を加えることにより手話単語の認識をおこなっている。

更にわれわれはこの性質に着目し、基本動作の組み合わせをもとにした手話データベース

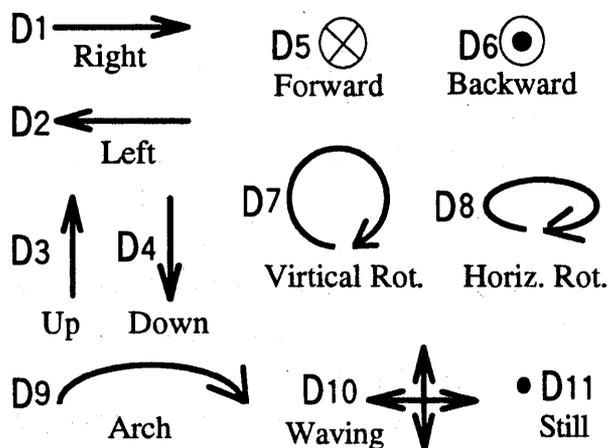


図4 代表的な手話の基本動作方向

あなた You			: 手話意味
S	bsbbb	uuuuu	: 右手/左手形状
		s=Straight, b=Bend, u=無視	
	D5	D11	: 右手/左手動作
E	bsbbb	uuuuu	: 右手/左手形状
楽しい Happy			: 手話意味
S	sssss	sssss	: 右手/左手形状
	D3	D4	: 右手/左手動作
	sssss	sssss	: 右手/左手形状
	D4	D3	: 右手/左手動作
E	sssss	sssss	: 右手/左手形状
北 North			: 手話意味
S	bssbb	bssbb	: 右手/左手形状
	D4	D4	: 右手/左手動作
	bssbb	bssbb	: 右手/左手形状
	D2	D1	: 右手/左手動作
E	bssbb	bssbb	: 右手/左手形状

図5 ジェスチャーデータベース記述例

スの構築を試みている。データベースには手話単語が、前述の基本動作とそれに付随する手形状データの列によって記述される。図5に本データベースにおける手話単語記述例を示す。本システムでは、両手にそれぞれデータグローブ、手の甲部分に固定したポジションセンサ及び加速度センサを装着することによって基本動作の認識をおこない、手話のデータベース登録、手話単語の認識、データベースからの手話単語の意味検索をおこなう。

図6に手話認識システムの流れを示す。まず時系列データから特定のジェスチャーを認識するためには、その始まりを検出しなければならない。ここではジェスチャーの開始は、ポジションセンサからの位置情報と加速度の絶対値の大きさを判断する(図2(a),(b)参照)。ポジションセンサでは、初期位置からの移動が検出されたときに手話動作の始まりとみなされる。加速度からの判定では、使用者の非動作時の手ぶれの大きさをあらかじめ測定しておき、これよりも大きな加速度入力

があった場合に、ジェスチャー開始の候補として選ばれる。このどちらかの信号を検出した時点でジェスチャー認識アルゴリズムが実行される。このアルゴリズムで対象とするジェスチャーは、図4で示した11個の基本動作の列および手形状である。この認識結果はジェスチャーテーブルにそれぞれジェスチャー基本動作パターン列としてシーケンシャルに書き込まれて行き、手話データベースを参照することによって手話単語が認識される。現在までに、日本語50音の指文字と、自己紹介や日常生活でよく使われる30程度の手話単語の登録をおこない、その実時間認識を実現している。11の基本動作単体についてはほぼ100%の認識が得られているが、例えば図5に示した「北」のようにD4→D2と一つの手話単語内で動作の方向が変わる場合、その変化点における誤認識パターンの排除が現在の課題となっている。

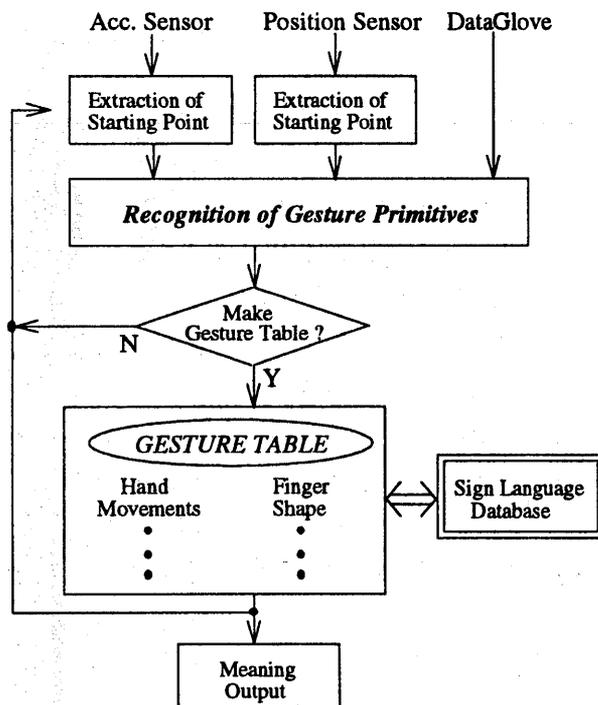


図6 手話単語認識の流れ

4-2. ジェスチャーによる音楽制御

これまでも人間の動作により音楽演奏を制御するシステムの試みは多い[9]-[12]。しかしながら、これらの多くは専門家向けに特化されたシステムであり、システムの大きさ、使用環境の制限、使用方法の複雑さ、などの問題がある。ここでは提案したジェスチャー認識を応用して、操作が簡単で誰にでも使用できる音楽演奏制御システムの製作を試みた。

大きなジェスチャーからは大きな音が作られる、あるいは指で指し示した方向に音源が生成される、音楽に合わせて手拍子をするといったように、ジェスチャーとそこから生成される音響の間には、ある程度納得のいく関係が存在する。その一方で、音色、ピッチ、音長、残響、エフェクタといった音生成パラメータを操作することによって音楽を生成していく過程では、使用者が自分の好みのジェスチャーを用いることができれば、楽器固有の操作法にとらわれることなく音楽を創作していくことが可能となるであろう。本システムでは、前者を演奏制御ルールとして持ち、また後者については対応関係を学習させることによって、演奏者が実時間で音楽、音響を生成し、演奏に介入することが可能となる。

試作したシステムの流れを図7に示す。左手にデータグローブを、ここでは加速度センサを2個用いて、一つを右手に他の一つをジェスチャーを認識しやすい足や頭などの身体部分に取り付ける。

まず、演奏開始、終了、コマンド選択といった音楽制御システムの実行に必要な、システムコマンドジェスチャーを登録する。このコマンドジェスチャーと演奏制御ルールジェスチャーにより、使用者はまず音色を選択し、空間に配置することによって音響の生成を行う。演奏開始コマンドにより楽音が生成されてゆくが、ここでは演奏者の指揮動作から抽出されるテンポ、音量情報に従って演

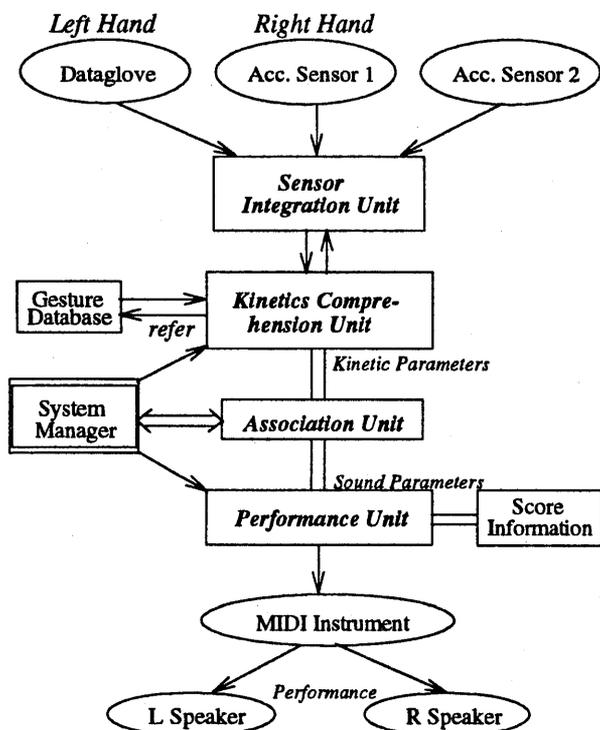


図7 音楽制御システム

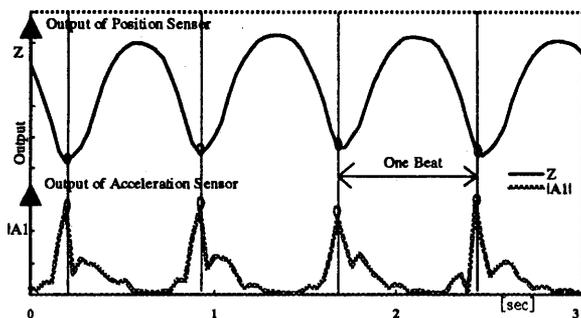


図8 演奏テンポの検出

奏が進行していく。演奏進行中には、コマンド選択のジェスチャーによって、テンポが保持されたまま、ジェスチャーと音響パラメータの関係の学習、音響パラメータの操作などが可能である。これによって自由なジェスチャーを用いて3次元空間の音響、楽音を生成していくことができる。パラメータ変更後には、演奏者が指揮動作によって再度演奏の進行に介入することが可能である。介入の際

には、演奏者は計算機の演奏に合わせて指揮動作をおこなって行くが、両者のテンポの同期が起こった時点から徐々に演奏の主導権が人間の演奏者に移って行く。これによりスムーズな演奏を可能にしている。図8にテンポ検出の例を示す。

5. まとめ

加速度センサ、データグローブおよびポジションセンサを用いたジェスチャー認識とその応用について報告した。ここでは、シンプルなハードウェア構成と単純な認識アルゴリズムによる実時間ジェスチャー認識を提案した。これを手話認識および音楽の制御に用いることにより、ジェスチャーを用いたマン・マシンインタフェースの入力部としての手がかりが得られたと考えている。

ジェスチャーは人間の微妙なニュアンスを表すのに適したメディアである。その一方でジェスチャーが果たす役割の大きい微妙な感情表現などのパターン化は困難である。ここで述べたジェスチャー認識手法では容易に取得データの時間分解能を上げることが可能であり、身体に及ぼされる細かな力のパターンと感情表現との関係を詳細に調べて行くことが今後の課題であると考えている。また、本手話認識システムを用いて手話動作を解析、分類することによって、力パターンと手形状データをもとにした手話辞典の構築を目指している。

このようなジェスチャー認識の適用により、音楽、手話ばかりでなく多くの分野に利用できる感性豊かなマン・マシンインタフェースが実現できると思われる。例えば、ジェスチャー認識を用いてジェスチャーとコマンドを対応付け、マウスでポインティングしながら片手のジェスチャーで操作するCADシステムなどが考えられる。

本研究の一部は、文部省科学研究費（重点
(2)「感性情報」課題番号06212225及び「人工
現実感」課題番号07244221）によっている。

文 献

- [1] 黒川隆夫「ノンバーバルインタフェース」、
電子情報通信学会編、オーム社、1994
- [2] 金山宣夫「世界20ヶ国ノンバーバル辞
典」、研究社出版、1983
- [3] Morita,H., Ohteru,S. and Hashimoto,S.,
"Computer Music System that Follows A Human
Conductor", IEEE Computer, Vol.24, No.7, pp.45-
53, 1991
- [4] Sato.A., Harada,T., Hashimoto,S. and Ohteru,S.,
"Singing and Playing in Musical Virtual Space",
Proc. of International Computer Music
Conference(ICMC), pp.289-292, 1991
- [5] Sawada,H. and Hashimoto,S., "Musical
Performance System Using 3D Acceleration
Sensor", Multimedia Modeling Towards
Information Superhighway, World Scientific, 1995,
pp.293-306
- [6] 澤田秀之、橋本周司「加速度センサを用い
たジェスチャー認識と音楽制御への応用」、
電子情報通信学会論文誌 VOL.J79-A No.2,
pp452-459, 1996
- [7] Sawada,H., Onoe,N. and Hashimoto,S.,
"Acceleration Sensor as an Input Device for
Musical Environment", Proc. ICMC, pp 421-424,
1996
- [8] Kim, J., Jang, W., and Bien Z., "A Dynamic
Gesture Recognition System for the Korean Sign
Language (KSL)", IEEE Trans. on Systems, Man
and Cybernetics -Part B:Cybernetics, Vol.26, No.2,
April, 1996
- [9] Keane,D. and Gross,P., "The Midi Baton", Proc.
of ICMC, pp.151-154, 1989
- [10] Katayose,H., Kanamori,T., Kamei,K.,
Nagashima,Y., Sato,K., Inokuchi,S. and Simura,S.,
"Virtual Performer", Proc. of ICMC, pp.138-145,
1993
- [11] Rubine,D. and McAvinney,P., "The
Videoharp", Proc. of ICMC, pp.49-55, 1988
- [12] Rokeby,"Body Language", ACM Siggraph Art
Show, Atlanta(1988)