

# 幼児期からのウェアラブルと Toy 型インタフェース

間瀬 健二<sup>†</sup>

ブライアン・クラークソン<sup>†‡</sup>

米澤 朋子<sup>†§</sup>

<sup>†</sup>ATR 知能映像通信研究所

〒 619-0288 京都府相楽郡精華町光台 2-2-2

{mase,clarkson,yone}@mic.atr.co.jp

<sup>‡</sup>MIT メディア研究所

<sup>§</sup>慶応義塾大学大学院 政策・メディア研究科

**概要:** ウェアラブルコンピュータがもつ、常時 on, ハンズフリーという特徴は、日常生活のなかで起こるイベントや経験を自動記録したり、ジェスチャなどのノンバーバルな手段による自己表現の支援に適している。一方、これらの特徴が生かせるタスクに幼児期の監視, エピソード記録, 自己表現補助, 情操教育などがある。そこで幼児期から生涯を共にするウェアラブルシステムを指向して、幼児むけに、装着するのではなく、つかず離れずに存在できる、センサーを多数組み込んだセンサー人形を試作した。これをもとに intimacy-oriented なインタフェースを議論する。

**キーワード:** ウェアラブル・コンピュータ, toy インタフェース, センサー人形, 生涯システム, 自動日記

## Lifetime Wearable and Toy Interface

Kenji Mase<sup>†</sup>

Brian Clarkson<sup>†‡</sup>

Tomoko Yonezawa<sup>†§</sup>

<sup>†</sup>ATR Media Integration & Communications Research Laboratories

<sup>‡</sup>Media Laboratory, MIT

<sup>§</sup>Graduate School of Media and Governance, Keio University

**Abstract:** The well-known characteristics of wearable computers are “always on” and “hands free,” which are suitable as an automatic recording system of events and experiences and a self expression supporting system through non-verbal languages. Those systems are plausible for the use of infants and children at keeping a watch on, recording episodes, educating sentiment, and providing self expressive tools. We have prototyped a sensor-equipped doll as a semi-wearable device for children, which is oriented toward a part of the life-time wearable system. This paper also discusses the intimacy-oriented interface.

**Keywords:** wearable computing, toy-interface, sensor doll, lifetime system, automatic diary

### 1 はじめに

ウェアラブルコンピュータがもつ、常時 on, ハンズフリーという特徴は、どこでもすぐに情報検索ができたり、実世界との重畳による電子マニュアルの閲覧など Augmented Reality や Mixed Reality の

枠組みでアプリケーションがしばしば説明される。これらは大抵、いつでも、どこでも、どのような体勢でも、情報をブラウズできるというウェアラブルの利点に重点をおいた、情報を受容するシステムである。一方、情報を記録したり、発信するアプリケーションは、少しの事例が紹介されているのみで

ある。これは、情報受容のための入力デバイスの単純さに対して、情報入力・発信のための入力には適当なデバイスが未発達であることが大きな理由と考えられる。

情報発信のための入力デバイスの研究には、片手保持キーボードの開発、常時装着型キーボード [1] の研究、あるいは音声認識の利用 (例えば [2]) などがある。PDA (携帯端末) 上にパーソナルエージェントを搭載する [3] ことで、PDA を知的な入力デバイスとして利用することは、手軽さを提供するための、間接的なアプローチと捉えることができよう。一方、記録する情報を文字に限らなければ、ビデオカメラの小型化により、自分が見ている映像を発信することはたやすい。ウェアラブルの初期からの研究者である Steve Mann [4] は、自分が見ている映像を広く公衆に発信することで、自己の安全が向上できるとしている。Thad Starner は、装着したカメラ映像で手話認識 [5] を行ない、それをテキストに変換する翻訳アプリケーションを提案している。これは情報発信における、自己表現の支援の要素が強い。

本文では、ウェアラブルなシステムの特徴を生かして、日常生活のなかで起こるイベントや経験を自動記録や蓄積したり、ジェスチャなどのノンバーバルな手段による自己表現の支援に利用することを考察する。これらの記録や表現の支援は、大人にとっては、記憶想起支援や対話支援などが考えられるが、幼児期においてもこれらの特徴が生かせるタスクであることを示す。すなわち、監視、エピソード記録 [6]、自己表現補助、情操教育などである。これらのタスクは、幼児期に限らず老年期まで生涯を通じて関心のあるものばかりである。具体的には、幼児期での利用を考慮して、人形型のインタフェースを有する半ウェアラブルなシステムとして、センサーを多数組み込んだセンサー人形を試作したので、その構成を紹介する。このセンサー人形は、現在、生涯日記の自動記録のための状況学習認識機構と、人形とのままごと遊びモデルを用いて、接触表現から音楽表現に翻訳する自己表現支援機構 [7] の 2 つの機能を実装している。

以下、2 節では、人形型ウェアラブルの提案とセンサー人形の構成を詳しく紹介し、3 節で生涯日記の自動記録のための状況学習認識機構を説明する。4 節で、自己表現支援機構の機能・動作について簡単に説明し、5 節で人形型ウェアラブルと生涯ウェアラブルについて考察する。なお、4 節の音楽コミュニケーションのための自己表現支援の詳細について

は文献 [7] にゆずり、本文では、システム全体との関係など簡単に紹介するにとどめる。

## 2 幼児期からのウェアラブル

### 2.1 人形型エージェント

ウェアラブルコンピュータは、コンピュータを着用することで携帯型の計算・記憶能力の拡張が可能になるだけでなく、ウェアラブルシステムにエージェント機能を組み込むことで次のような利点をもたらすと考えられる。すなわち、(1) ユーザ状況を随時認識し、(2) ユーザの必要を自動的に判断して的確なサービスをし、さらに (3) ユーザの使い勝手に合わせた適応的インタフェースを提供することが可能になる。

エージェントが特徴として有する知的・自律的な動作は、そのユーザの興味や意図を正しく認識することによって、効果をもたらす。不十分な状況の理解で興味や意図を推定することは、人間にとっても困難な課題である。コンピュータエージェントが状況をよりただしく理解するために、常に行動を共にするウェアラブルなシステム上で動作することは、最も効果的な手段である。とくに、常に監視されているユービキタスなシステムや生体にチップを埋め込む方法よりも、心理的、肉体的に非侵襲であるという利点がある。

このようなインタフェースと状況認識型のサービスを提供するエージェントは、ウェアラブルでなく携帯型の PDA に実装することも考えられる [3]。しかし、そのような小型の携帯 PDA は、ポケットや鞆の中、手のひら、机の上などを行き来するのみで、ユーザとの接触の機会は少なくその強度もあまり強くない。ユーザの状況を知るときに、体温や接触強度、ユーザの視点や環境などの重要な手がかりを得るには従来型の PDA では限界がある。とくに幼児期にはそのようなものは相手にされないこともある。合理的な状況よりもむしろ感性的な状況がしばしば重要な意味をもつユーザやそのような環境におかれたユーザにとって、いかにもコントローラという外観は不利である。たとえ、PDA 型がビジネスシーンで適当といっても、テーマパークのような場所では、そうとは限らないのである。

そこで、エージェントを搭載したウェアラブルなシステムとして、携帯 PDA との中間ともいえる、人形型を提案する。人形型は PDA に比べユーザの

接触機会が多く、よりユーザの感性的な状況認識がたやすくなることが期待される。また、エージェントを搭載したシステムは、機能的に自律性や個性などの擬人性を有することになるが、外見的にも擬人性を有することで、それらの機能へのアクセスが容易になることが想定される。たとえば、エージェントに周囲の状況を見せるときに、顔をそちらに向けてやるなどの動作が自然に生ずるであろう。



(a)



(b)

図 1: センサー人形 ( (a) 初期システム, 腹にカメラとマイクが見える, (b) 第二世代, 鼻の穴にカメラレンズが見える )

## 2.2 生涯ウェアラブル指向のセンサー人形

人形型のシステムには、“ActiMates Barney[8]”や、Furby などの商品化されたものや、“Swamped!” で使われたセンサー人形 [9] などがある。Furby などは、機能も限られているため単体で機能するが、Barney や Swamped の人形では、センサーデータを無線でホストコンピュータに送り、そこで学習教材や物語の仮想体験などのアプリケーションを動か



図 2: 人形を分解した様子

している<sup>1</sup>。

我々がめざしている、イベントや経験を自動記録や蓄積したり、ジェスチャなどのノンバーバルな手段による自己表現の支援に利用することを考えると、人形はワイヤレスで完全に自立して動作する必要がある。そのため、我々は人形本体に、状況認識などの高度なパターン認識機能、データ蓄積機能、変換表現生成機能などが実現できるように、一定性能をもった PC を搭載することにした。また、現在は研究試作段階であり、今後の拡張性も考えて、汎用 OS で動作する環境を整える。

## 2.3 センサー人形の内部

実際に制作したセンサー人形は、図 1 のような市販のぬいぐるみの中身をくりぬいて、代わりに PC とセンサー類をつめている。図 2 がその内部構造である。

初期試作機では、USB ビデオカメラ (160 x 120 pixels)、USB マイクロフォン、標準シリアル接続の 3 軸角加速度センサおよびサブノート PC を搭載し、スタンドアロン型のシステムを構築し、自動日記の実験を行なった。第 2 世代の試作機 (図 2) では、人形とユーザの接触から状況を獲得するために、PCMCIA カード型の 16ch の Analog/Digital 変換器を導入し、センサーとして piezo 圧電素子を利用した圧力センサと曲げセンサ、赤外近接センサ、温度センサなどを人形の各部に装着し、内蔵するコンピュータをカードタイプ (140mm x 100mm x 40mm) の PC (Win2K) として小型化をはかった。

<sup>1</sup> Barney は単体でも動作するが、その場合、Furby と同じく単純な反応になる。

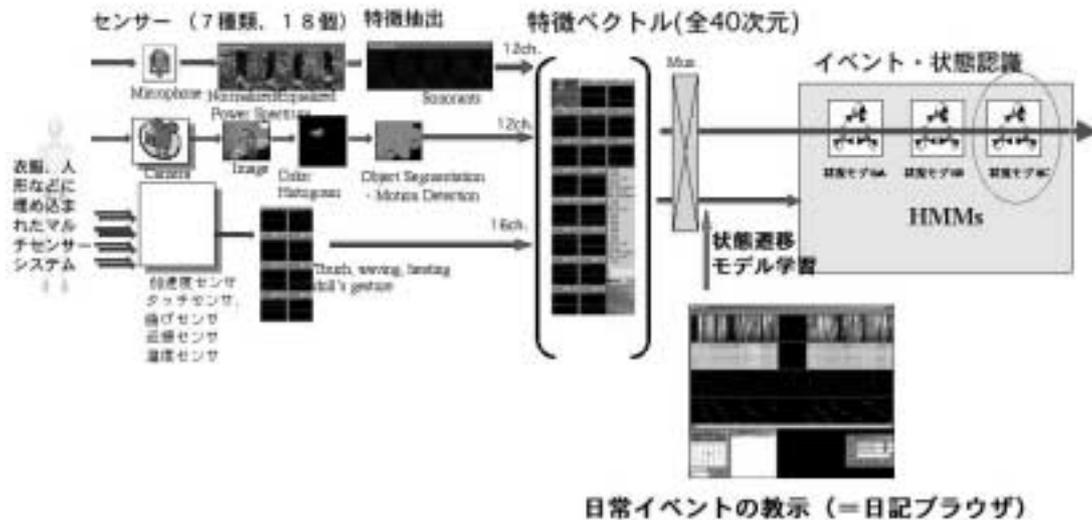


図 3: 認識システムダイアグラム

接続ポートの都合で、角加速度センサーの代りに1軸の並進加速度センサーを用いた。さらに磁気ディスク(9GB)を搭載して、日記の原データとなるイベントの記憶も可能とした。センサーは合計7種類18個である。無線LANにより外部コンピュータ・インターネットとの接続も可能である。

このセンサー人形システムは、以下に述べる生涯日記の利用においては、持ち歩いてイベントを記録した後、ステーションコンピュータから無線LAN経由でリモートログインして、新規イベントがあればそれをマニュアルで教示して、認識システムである隠れマルコフモデル(HMM)を学習させて次の利用に備える。また、音楽コミュニケーションのための自己表現用途においては、現在のところ、室内での利用を前提とし、センサーデータを前処理したものを無線LANでステーションコンピュータに送りそこで、センサーからの状況認識と状況に応じたセンサーから音への変換処理を行ない、室内スピーカーと人形につけた無線スピーカーをならすようにしている。なお、それぞれのアプリケーションにおいてはセンサーデータの利用の仕方が多少異なる。

以下、それぞれのシステムについて詳しく説明する。

### 3 自動イベント記録による生涯日記

我々が記憶と呼ぶときに、予定や計画を覚える展望的記憶と過去を振り返る回想的記憶(あるいは反省的記憶)があるという[6]。エージェントの役割も、予定や計画を通知したり推奨する機能と、過去の事象を想起するのを支援する機能がある。日記は回想的な記憶であるが、その事象を記録するのは1日の終わりや一連のできごとが一段落した後ということが多く、事実があいまいになったり感想や感動も薄れてしまうという欠点がある。また、記録をする場合に客観視するため、その場を体験できないこともある。子供の運動会でカメラやテレビカメラをもった親が映像を撮るのに夢中で、その場では勝敗がわからなかったということがある。主体視しつつ、想起したいような日記的事象を自動的に記録するシステムが必要である。

エピソード記憶とその想起支援の研究にはLammingらのForget-me-not[10]がある。彼らはユービキタスコンピューティング環境の中で、携帯端末を使って個人の活動を記録するシステムを作り、その記録が想起の手がかりになることをしめた。川嶋ら[6]は画像処理により、特定の人物や動作を検出エピソードとなる場面を記憶する方法を提案している。Clarksonら[11]は、画像、音声、加速度計のデータからHMMを使い状況を監視し、教えた状況を検出する手法を提案している。

我々は、前述のセンサー人形の入力を全部利用して一旦は内部磁気ディスクに記録し、あとでユーザがラベルをつけ HMM を学習させて特定の類似パターンを抽出してエピソードとして残すシステムをまず構築した。そのために、全イベントデータをブラウジングしてラベル付を補助するツールを用意した。教示が進みモデルが豊富になれば、このラベル付の作業は次第に不要になると考えられる。こうして日常的なイベントは自動的にラベル付されるシステムができる。

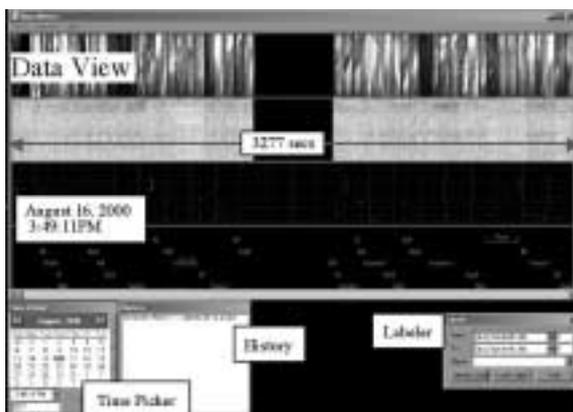


図 4: 日記データのブラウジング

### 3.1 HMM 状況認識システム

人形内の認識システムの処理ダイアグラムを図 3 に示す。画像信号と音声信号は、初期処理をしてそれぞれ 12 次元のデータに圧縮し、他のセンサー (16ch) と合わせて、40 次元の特徴ベクトルを構成する。

ここで用いる画像特徴は、輝度 (Y) とコンポーネント輝度 (R,G) の画像モーメント (0 次, 1 次, 2 次) を用いた。輝度を例にすると、平均輝度値、輝度重心座標 (x, y)、輝度分散の 4 次元データを計算する。R 信号、G 信号も同様にして、12 次元のデータを得る。また、音声特徴は、音量、継続係数、および 0~4KHz 帯を 400Hz ごとに区切ったスペクトルパワーとして、12 次元のデータを得る。これらは、対象に比較的依存しない特徴を得ることができるように選んだ。これらの特徴量の詳細と、音声および画像のみを用いた状況認識の初期実験については文献 [11] を参照のこと。

図 4 は日記データのブラウジングをしている画

面のスナップショットである。メインビューは上段からビデオ画像データ (全サンプルをずらして重畳表示している)、音声スペクトル、その他センサー (16ch) の入力、ラベル領域である。そのほか、カレンダーやラベル付けのための窓がある。図 5 は、時間軸を広げて画像と音声データのモニター部を拡大強調して表示している。人間の音声に反応したスペクトルが現れており、これが音声発声区間抽出の特徴として用いることができる。

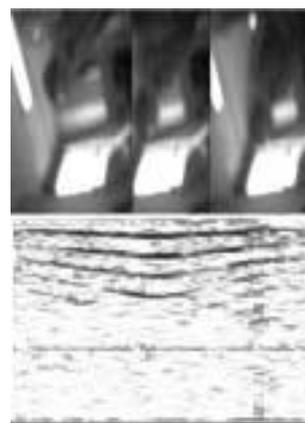


図 5: 画像・音声データの例 (上段: 部分画像, 下段: 音声スペクトル)

パターン学習はブラウザ (図 4) を使ってイベント全体をチェックして、ある種のイベントがあった区間を選択してラベル付けをする。同じモデルで学習させたいイベントには同じラベルを付ける。そうしておいて、ラベルごとのデータ列を使って HMM のモデル化計算を行なう。ここで、状況を認識する HMM のモデルを学習させる際に、全てのセンサーデータを評価対象とすると、予期しないセンサーのパターンを強く学習して、総合的に認識率の低下を招く可能性がある。そこで、本システムでは、HMM を学習させる際にチャンネルを選択できるようにしてある。

### 3.2 初期実験

図 6 は、あるイベント (label: motion) を選択して、学習を行なったときの、全区間に対する “motion” モデルの尤度値を計算した例である。印で現した教示区間に対して、その区間の尤度値が高くなっているのがわかる。また、教示しなかった区間についても 2ヶ所ほど、値が高くなっているのがわかる。こ

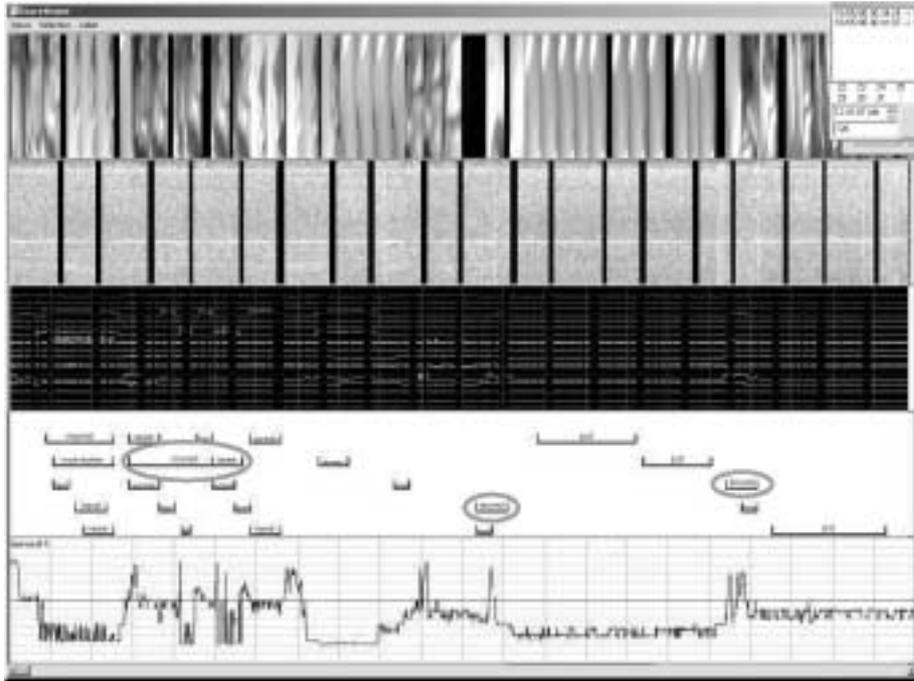


図 6: HMM 教示の例：最下段が、複数の教示データ（印）によってできたモデルが出力する尤度値

れは、ユーザの教示漏れか教示の誤りにあたる。

通常、パターン認識の実験において、このようなケースは、認識器による過剰認識か認識漏れの認識誤りにあたるが、教示は全区間について絶対的な真値 (ground truth) を提供しているわけではないので、ユーザの教示を真値に近づけるための指針になる。例えば、ラベルがついていないのに尤度値が高い区間は、気がつかないでいたがそのラベルのイベントが起きていたことを確認できるかもしれない。逆に、その区間を調べても尤度の高い理由が見つからないときは、教示に使ったパターン区間を再検討する必要がある。

#### 4 音楽による自己表現支援

機能的には携帯型のコンピュータ端末でありソフトウェアエージェントとしてのプログラムの動作主体にすぎないが、ぬいぐるみの形態がもたらす属性は、ユーザが親密な関係を構築するのに効果的であることが期待できる。日常的な活動に同伴することで、人形がコミュニケーションを支援するツールとして展開できるようになる。

西本ら [12] は、人間のグループ思考では、思考主体の遷移に基づいて、個人思考モード、意思疎通

モード、協同思考モードを行き来するグループ思考モデルを提案した。これは、コミュニケーションを生業とする人間が通常の生活で行き来する思考活動のパターンでもある。我々は、人形エージェントとのコミュニケーションが、この思考過程のサブプロセスであると考え、各モードにおいて、パートナーであるユーザの思考活動を支援する主体となることを想定する。

すなわち、個人思考モードにおいては、ユーザと同格の作業者として、ユーザと人形間で意思疎通のための相互のコミュニケーションを行ない、意思疎通モードにおいては、ユーザと人形が協同して（1個のグループ行動として）、人間の協同作業予定者に意志を表示する、というものである。ここで、人間ユーザにとっての人形との協同作業の目的は、自己を的確に表現することにある。エージェントであるセンサー人形をうまく使って、表現を豊かにすることを支援できると考える。

我々は、ことばによる表現は自分自身で行なうことにし、非言語の表現において人形の助けを借りる方法を検討し、文献 [7] で提案している。これは言い換えると、知的（理性的）な主張はことばで表現しつつ、感性的な主張は、人形に対するジェスチャ、姿勢、接触などのさまざまなノンバーバル言語により表現し、それを人形が認識理解して、音楽というメ

ディアで翻訳して補強して表現し、自己のノンバーバルな表現と組み合わせて相手に伝えるという、新しいコミュニケーションの枠組みの提案である。

音楽はその外的状況と相補的に協調して、その場面を作り出す効果をもっているため、このようなコミュニケーションを補強する手段として適していると考えられる。携帯・ウェアラブル型による、場を演出する音楽演奏支援装置の研究には、文献 [13] のような、個人のインタフェースはコントローラに徹し、セッションの創出に工夫をする研究があるが、本文の人形は、そのコントローラ部を、ユーザに適応的で親密なインタフェースにすることを主眼とする。

#### 4.1 実験システム構成

日記システムの実験では、学習パターンの教示を除くランタイム処理においては、システムはスタンドアロン型の構成とした。それに対し、本節の音楽による自己表現支援の実験システムでは、音楽表現部の実装は無線ネットワークを介して別のステーション PC 上で行ない、音響信号はその PC で生成している。ラピッドプロトタイピングのために、音楽生成のために必要な部品（ソフト、ハード）をぬいぐるみとは別にすることにしたためである。

また、前述のように状況に合わせて対話モードを切り替えることを実現するために、内部的には決定的有限状態遷移のモデルを導入し、人形との一人遊びや、人形と協調して表現をするモードの切り替えを外部とのインタラクション状況で制御するようにした。異なるモードでは、同じ入力に対する反応が異なるように設計している。なお、入力データのうち、画像と音声については、輝度の 0 次元モーメントと、音声のボリュームと、1KHz 帯のスペクトル強度の 3 次元データのみを使っている。

### 5 考察

生涯日記の応用においては、センサー人形は、記録の自動スイッチと記録センサーを兼ね備えている。しかしセンサー人形が、ユーザに近すぎて、鑑賞に適した映像にならない場合もありうる。その場合、外部カメラなどのリモートコントローラとしてのぬいぐるみの応用という構成も考えられる。ユービキタスなセンサ環境においては、ぬいぐるみと一緒に写ることが、さらにぬいぐるみとの親密性を増すことになろう。あるいは、よそに置かれているぬいぐる

みが、状況を認識して、記録映像を撮影することも可能である。このように、装着しないカメラによる記録のメリットは大きいと思われる。

ところで、日記として最も価値のある事象は、非日常的なものである。非日常的、突発的な事象を前もって教示することは不可能に近いが、次の可能性がある。(i) 日常的イベントの補集合をとる、(ii) 類似の非日常的イベントを教示データとして用いる、(iii) 特別なスイッチを用意する、などである。

現在は、生涯日記システムと自己表現支援システムを、同じシステムプラットフォームの上に別のアプリケーションシステムとして発展させている。日記システムの HMM による状況認識機構は、自己表現支援の応用において人形の状況モード制御に利用できるが、自己表現支援における音楽表現が、生涯日記システムの動作に、効果的なフィードバックを返すかどうかまだ不明である。人形との親密な関係づくりが、日記記録システムのインタフェースの適応的個人化にポジティブなフィードバックを返すようにするための機構について、さらに検討が必要である。



図 7: センサー人形と遊ぶ様子

### 6 おわりに

ぬいぐるみ型の人形に、コンピュータと多種多様なセンサーを詰め込みイベントを認識記録したり、イベントや状況に合わせて音楽によってコミュニケーションを支援することを目指したシステムを試作し、その構成や、基本的考え方などについて述べた。現在は、別アプリケーションシステムとして動作するが、全体を統合した生涯パートナーとしてのシステムをめざしたいと考えている。

人形型のインタフェースはウェアラブルとは呼ばないかもしれない。しかし、将来のウェアラブルの形態は、幼児期から形を変えつつ生涯を共にするようなシステムを指向することになると考えられる。その際、幼児期には、装着するのではなく、本文で提案したような、つかず離れずに存在しかつ身体性を有している人形型が適していると考える。その利点の検証については今後の課題である。

- [12] 西本一志, 角康之, 門林理恵子, 間瀬健二 中津良平: “マルチエージェントによるグループ思考支援”, 電子情報通信学会論文誌, J81-D-I, 5, pp. 478-487(1998). 論文.
- [13] 多田幸生, 西本一志, 前川督雄, 間瀬健二 中津良平: “人や場を演出する音楽創奏システムの提案”, 情処研報 音楽情報科学 2000-MUS-37, pp. 1-8(2000).

## 謝辞

日頃ご指導いただき、ATR 知能映像通信研究所 酒井保良会長ならびに中津良平社長に感謝します。また、角康之、西本一志、Seon-woo Lee の各氏をはじめ、日頃熱心に討論していただく研究員の皆様に感謝します。

## 参考文献

- [1] 福本雅朗, 外村佳伸: “Wireless fingering: 人体を信号経路に用いた常用装着型キーボード”, 情処論文誌, 39, 5, pp. 1423-1430(1998).
- [2] 土井俊介, 角康之, 間瀬健二, 中村哲 鹿野清宏: “音声対話型パーソナルガイドエージェントシステム”, 人工知能学会第 47 回知識ベースシステム研究会, volume SIG-KBS-9904, pp. 55-60(2000).
- [3] 角康之, 間瀬健二: “実世界コンテキストに埋め込まれたコミュニティウェア”, 情報処理学会論文誌, 41, 10, pp. 2679-2688(2000).
- [4] S. Mann: “Smart clothing: The wearable computer and wearcam”(1997).
- [5] T. Starner and A. Pentland: “A wearable-computer based american sign language recognizer”, First International Symposium on Wearable Computing(1997).
- [6] 川嶋稔夫: “記憶想起支援とウェアラブルコンピュータ”, 第 5 回知能情報メディアシンポジウム論文集, pp. 107-108(1999).
- [7] 米澤朋子, BrianClarkson, 安村通晃 間瀬健二: “ぬいぐるみインタフェースによる音楽コミュニケーション”, 情報処理学会研究報告 HI92-3(2001).
- [8] E. Strommen: “When the interface is a talking dinosaur: Learning across media with actimates barney”, Proceedings of CHI 98, pp. 288-295(1998).
- [9] M. Johnson, A. Wilson, C. Kline, B. Blumberg and A. Bobick: “Sympathetic interfaces: Using plush toys to direct sythetic characters”, Proceedings of CHI99, pp. 152-158(1999).
- [10] M. Lamming and M. Flynn: ““Forget-me-not” Intimate computing in support of human memory”, Proceedings of International Symposium on Next Generation Human Interface '94FRIEND21, pp. 150-158(1994).
- [11] B. P. Clarkson, K. Mase and A. Pentland: “Recognizing user context via wearable sensors”, ISWC2000(2000).