

# ぬいぐるみを用いた表現としての音楽生成の提案

米澤 朋子<sup>†‡</sup>, ブライアン クラークソン<sup>§</sup>, 安村 通晃<sup>‡</sup>, 間瀬 健二<sup>†</sup>

<sup>†</sup>ATR 知能映像通信研究所

<sup>‡</sup>慶応義塾大学大学院 政策・メディア研究科

<sup>§</sup>MIT Media Laboratory

{yone,mase}@mic.atr.co.jp, clarkson@media.mit.edu, yasumura@sfc.keio.ac.jp

## 概要

我々は、ぬいぐるみとユーザのインタラクションを文脈のある表現と見なし、様々なセンサーを埋め込んだぬいぐるみへの入力により音楽を生成するシステムを提案し、実装した。本稿では、インタラクションに予め5段階のレベルを与え、同時にぬいぐるみのジェスチャーを認識することにより、インタラクションの文脈に合わせた音楽マッピングの手法を紹介する。

## 1 はじめに

人間とコンピュータのインタラクションが自然なコミュニケーションであるためには、コンピュータは決まった答を返す箱ではなく、その時その時の文脈に応じた返答をする適応的なシステムであることが必要とされる。その文脈や状況と、それぞれの状況下でのユーザの行動とその意図をいかに多く把握しているかにより、行われるインタラクションが豊かな要素を与えることができるようになる。これまで、文脈適応型システムは特に個人化ガイドシステムの機構として議論されている。例えば Abowd による Cyberguide システム [1] は教室内などのグループでの利用により、個人情報と全体の情報をあわせ持ったモバイルコンピューティングの提案がされている。また、角らは美術館や見本市、学会における展示でのガイドエージェントシステム [2] を提案している。

それに対し我々は、人間同士の非言語コミュニケーションチャンネルの一つとして、文脈認識機構を持ったセンサぬいぐるみシステムを提案する。ぬいぐるみはその接触を通じて幼い子供が情操を養うのに擬人化して用いたり、その触覚を伴った疑似コミュニケーションでままごとを行ったりするのに用いられる他、大人にとっても同様に親しみやすい存在である。ぬいぐるみと人間のインタラクションの中には主体性を人間がつかさどり、一方的に仮想的なコミュニケーションを展開しているという見方もできる。ここで我々は、ぬ

いぐるみとの、文脈を持ちかつ接触を中心としたマルチモーダルなインタラクションに着目した。そして、人間同士のコミュニケーションを補助する親しみやすいデバイスとして、ぬいぐるみを活用することを提案する。文献 [3][4] で紹介したぬいぐるみとの音楽コミュニケーションシステムに加え、本文ではぬいぐるみシステムについての評価観点と手法について提案する。そして、ぬいぐるみを用いて音楽生成を行うことが表現として有効であるかについて、検証することができると考えている。

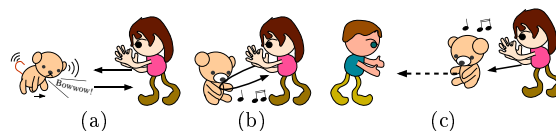


図 1: 擬人化ぬいぐるみコミュニケーション

ロボットとのインタラクションにおいては、図 1(a) に示されるようにロボットは活動的で強い人格を持つためユーザとロボットの関係が対等になる。これは人間同士のコミュニケーションを補助する用途ではむしろそれを妨げる可能性もある。そこで人間が主体性を持つことができる、図 1(c) のような人間同士のコミュニケーションデバイスを実現するためのステップとして、図 1(b) に示すぬいぐるみと人間とのインタラクションを実現するシステムを設計した。特に、ぬいぐるみがロボットやおしゃべり人形のような強すぎる主張をせず、主体となる人間自身の環境的要素になるよう考慮した。すなわち、ぬいぐるみのアクチュエータとして音や音楽のみで表現し、メッセージ性の強い表現や言葉を用いないこととした。

試作したぬいぐるみシステム “Com-music” は、カメラやマイク、圧力センサなど様々なセンサと、それ

Music Generation as Multi-modal Expressions using Sensor-Doll Tomoko Yonezawa, Brian Clarkson, Michiaki Yasumura and Kenji Mase

ATR MIC Research, Keio University, MIT Media Laboratory, 2-2 Hikari-dai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan E-mail: {yone, mase}@mic.atr.co.jp

らの情報を処理するためのPCをぬいぐるみの内部に持つ。ぬいぐるみはあらかじめ数種類の内部状態を与えられ、ユーザの入力行動のレベルを状況として認識し、内部状態を変化させる。異なる内部状態により音楽マッピングを変化させることで、同じ入力でも文脈により異なる反応をする。

本論では、まず、2節においてこれまでに行われている幾つかの文脈適応型インタラクション生成システムを例に挙げ、ぬいぐるみとのインタラクションを拡張させた本システム開発の背景について述べる。そして、3節では、我々が開発したセンサを埋め込まれたぬいぐるみシステムのコンセプトに沿った内部処理、ハードウェア、そしてインタラクションに伴う音楽生成手法について具体的に記述し、最後に4節においてぬいぐるみインタフェースと音楽表現によるインタラクションの表現方法についての考察をまとめる。

## 2 関連研究

前節に述べたように、人間とロボットとのインタラクションについての研究が行われる一方、ぬいぐるみや人形を親しみやすいコンピュータインタフェースとして用いるための研究も多く見られる。

Alan Kayら[5]のVivarium Projectでは、人工知能による電子生態系プロジェクトの一環として、コンピュータと人間のインタフェースにぬいぐるみを用いる試みを行っていた。人と同じくらいの大きさのぬいぐるみシステム“Noo-bie”は、表面の至る箇所に様々なセンサを埋め込まれ、ユーザの接触により腹部のディスプレイに様々な反応が映し出される。

SIGGRAPH98にて展示された、Johnsonら[6]による“Swamped!”システムは、センサを埋め込まれたぬいぐるみへのユーザ入力を判断し、あらかじめ構築されている視覚的な仮想世界の、場面場面に応じたキャラクターの行動をコントロールするものとして用いられている。ぬいぐるみを共感的インタフェースとして提案している点が、我々の研究に関連する点として注目できる。しかし「場面」という一種の文脈を導入しているにも関わらず、ここでは主に視覚的なコントロールに用いられており、インタフェースがぬいぐるみである必然性はない。

それに対し我々は、ままごと遊びにおけるぬいぐるみの存在が「パートナー(エージェント)」として、また「自分の代わり(アバター)」として見なされることに着目し、そういったぬいぐるみ独特の役割をインタフェースに導入することを提案する。

稲葉ら[7]、星野ら[8]による触覚センサスーツは、ヒューマノイドロボットの接触位置情報を検出する目的で開発された。同様に、ユーザのコミュニケーションとしての様々な接触を検知するために、ぬいぐるみの表面にセンサスーツを設置することも可能である。センサスーツは接触する位置情報を詳しく検出するためにはふさわしい手段である。しかし我々のシステムでは、例えば曲げセンサや接近センサなど、他の種類の接触にも対応したセンサの配置を考慮し、接触の検出が必要な位置を絞って圧力センサを配置した。

他にも、おもちゃ市場ですでに取り扱われている様々なぬいぐるみシステムが存在している。“My Real Baby”[9]は様々なセンサを埋め込まれた乳児の人形で、感情の種類をいくつか持つことによりその感情に応じた反応を返す。“Furby”[10]は様々なセンサを取りつけられ、ユーザの入力に対して言葉や簡単な動きで反応を返す、広く知られているぬいぐるみである。しかしこれらのシステムは、あらかじめ決められた言葉による反応など、設計された反応により予想されていたやりとりをする遊び相手としての存在に過ぎない。“ActiMates Barney”[11]は同様に子供の学習のための遊び相手として取り扱われているぬいぐるみシステムである。単純な返答をする“Stand-alone toy”モードのみでなく、テレビやコンピュータとつなげて連動するぬいぐるみにもなる。しかし現段階では、他のぬいぐるみ商品と同様に、あらかじめ言葉が録音されているため、反応の種類も乏しく、子供とぬいぐるみがコミュニケーションしているとは見ることは難しい。

土井ら[12]は音声入力型のぬいぐるみエージェントガイドシステムを提案している。これはぬいぐるみとのインタラクションに音声対話を導入している。親しみを持つことのできる相手と見るには、Tactileな入力に対するあらかじめ決められた言葉の応答よりも土井らのシステムは効果的であると考えられる。しかし実在するぬいぐるみをデバイスとして十分に活かしているとは言いがたい。鈴木ら[13]は、言葉ではなく非文節音を利用してぬいぐるみと対話することを提案している。これは言葉以外の新しいメディアを用いたインタラクションとして興味深い。しかしここで紹介されたシステムはユーザの感情を表現するツールやエージェントといった役割を持たない。

我々は言葉による表現はユーザ自身により行い、非言語の表現においてぬいぐるみの助けを借りる方法を検討し、音楽による表現を採用した。ここで我々は、あらかじめ準備する反応の枠組みとして音楽のパラメー

タを準備し、ユーザがその反応の変化を楽しみながら、コントローラのようにぬいぐるみを操作し、同時にぬいぐるみを生きている対象のように扱いコミュニケーションを行うことを目指す。

音楽を表現することを目的とした、対話的なロボットシステムとして、Camurriら [14] や Suzukiら [15] が開発したダンス共演ロボットがある。これらは、ユーザとの関係を音や音楽に変換しながら、コミュニケーションを図るといった提案をしている。しかしこれらはあくまでロボットと人間の対等な関係であり、そのインタラクションの主体はロボット側にある。他にも、自律的で単純な音楽表現をするシステムとして、Wassermanら [16] による“Roboser”がある。ロボットが感じた様々な状況を音楽に変換し、ロボットは周囲の状況（人間とのインタラクションもその一つとして含む）に対し音楽要素を返す。しかしここではユーザはロボットとの関係性などを反映したインタラクションを行うことはできない。

西本ら [17] 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、共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、人間が通常の生活で行き来するコミュニケーションに含まれる、思考活動のパターンでもある。子供のままごと遊びにおけるぬいぐるみの利用において「自分の相手」「自分の分身」として用いていることを、我々は文献 [17] の個人思考モード、意思疎通モードと共通していると見なし、ぬいぐるみをこれらの思考過程のサブプロセスであると捉え、各モードにおいて、ユーザの情操や思考活動の支援をする主体としてぬいぐるみを想定する。すなわち、個人思考モードにおいては、ユーザの相手となり、ぬいぐるみ-ユーザ間での相互コミュニケーションを行う。また、意思疎通モードにおいてはユーザの分身となり、他の人間とのコミュニケーションにおいてユーザの意思表示を行う、というものである。ここではセンサぬいぐるみを用いることで表現を豊かにすることを支援できると考える。

このように、我々は、一つはコントローラとしてのぬいぐるみ、もう一つは文脈に応じた異なる反応を持つ相手としてのぬいぐるみという、二つの役割に適応的に動作するコミュニケーションデバイスとしてのセンサぬいぐるみを提案する。動的なロボットとのインタラクションでは、ユーザはロボットと対等なやり取りを行おうとするため、コミュニケーションのツールとしてではなく強いパーソナリティを持つ対象として見

なされる。特に人間と人間のコミュニケーションツールとしてぬいぐるみを用いるためには、受動的なぬいぐるみの持つ役割の柔軟性を用いることが重要である。

多田ら [18] [19] による携帯・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では、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している。しかし本文のぬいぐるみは、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している。

Clarksonら [20] による“Wearable Sensor System”は、ユーザの洋服など身の回りにカメラやマイクを設置し、HMMを用いてユーザの状況認識を行うシステムで、庭や地下鉄、台所などの場所を特定することができる。HMMは先述の“Swamped!”システムでもジェスチャー認識に組み込まれており、多くのセンサによる複合的な状況認識で効果的である。文献 [21] のToy型インタフェースでは上記の技術をぬいぐるみに応用し、幼児期から生涯を共にする自動日記ウェアラブルシステムを紹介している。このToy型インタフェースの研究におけるセンサの種類・数や状況分析の手法などを基本に、我々のぬいぐるみシステムは構築されている。

Aibo[22] や R100[23] といった市販のペット・ロボット/パーソナル・ロボットは人の情操を伴うおもちゃとして人気がある。大野ら [24] は電子ペットを道具としてトレーニングに用い、ある意味では主体性を人間側に与えているとも考えられる。これらのシステムは自律的に動きロボット自身の表現を行うが、ぬいぐるみは表現支援の役割を果たすことができる。

### 3 システムデザイン

#### 3.1 ぬいぐるみとのインタラクションモデル

我々は、まず、図 1(b) に示されるような、人とぬいぐるみとのインタラクションを念頭に、センサを埋め込んだ文脈適応型ぬいぐるみをコミュニケーションに用いるためのデザインをした。

はじめに、人間同士のコミュニケーションを参考とするため、入力の種類・結果により以下の 4 タイプに分類した。

1. 単純な入力に対して即座に反応をする  
c1:  $Input_n \rightarrow Reaction_{n+\Delta}$
2. 連続した入力群に対して即座に反応をする  
c2:  $Input_{n \rightarrow n+i} \rightarrow Reaction_{n+\Delta}$
3. 単純な入力に対して反応のモードが変わる

c3:  $Input_n \rightarrow Mode_{n+\Delta}$

4. 連続した入力群に対して反応のモードが変わる

c4:  $Input_{n \rightarrow n+i} \rightarrow Mode_{n+\Delta}$

例えば, 1. は, 声をかけたら相手が振り向いたり, 子供を叩くと泣き出すなどの反応を指す. 2. は, ある種のジェスチャーを見せることで乳児が笑うなどの反応が当てはまる. 3. は, 思わずひどい言葉を吐くとその後の人間関係にひびが入るなど, コミュニケーションの「モード」が切り替わる場合が例として挙げられる. 4. は, お見合いの席で自分が発言した言葉群の総合が次回に相手が自分を見る見方となるような場合を例として挙げられるであろう.

よって, ぬいぐるみとのインタラクションに 1) コミュニケーションのモードとなるぬいぐるみの内部状態と, 2) それぞれの内部状態における即座な反応を準備することでより豊かなインタラクションを設計できると考えた.

### 3.2 ぬいぐるみの内部状態の設計

上に挙げたコミュニケーションのモデルに基づき, センサぬいぐるみに 1) コミュニケーションのモードになる内部モードとそれを変化させる機構, 2) それぞれのモードにおいて入力を音楽に直接翻訳する機構を設計した. まず, 仮に乳児とのインタラクションを想定し, ぬいぐるみの気分などを表す 5 つの内部状態を準備した.

Level 0: 眠っているが周辺に興味がある. 呼吸音を立てている.

Level 1: ユーザとの出会いの段階で, 接触に反応する. 返事のような声を返す.

Level 2: 温かみのある, より親密なコミュニケーション. 呼吸の音を和音に変化させ, 返事のような声も音階に沿った音程になる.

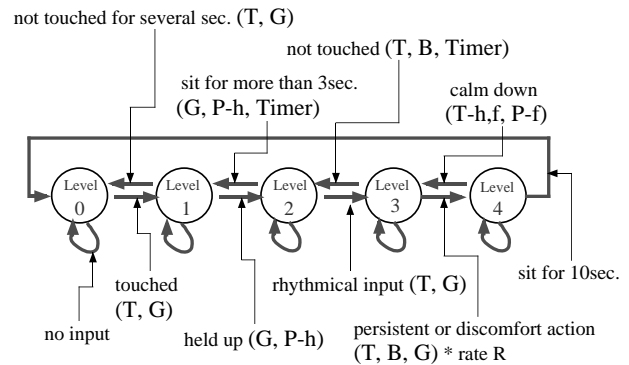
Level 3: 音楽によるコミュニケーションの段階. リズムやメロディー, 和音を持つ即興の曲をユーザと協調して演奏することができる.

Level 4: コミュニケーション不能の暴走状態. リズムやメロディーは存在するものの, 混乱を表す音楽を出力する.

ぬいぐるみのこれらの内部状態はユーザとのインタラクションの強さ・激しさ (Interaction Level) の変化によって遷移する.

### 3.3 内部状態の遷移モデル

上に述べた 5 つの内部状態を変化させるため, 内部状態の有限オートマトンを準備することにした. 図 2 に示すように, それぞれの内部状態においてそれぞれの異なる種類の入力信号がイベントとして認識され, 遷移のトリガーとなる. そのため, 現内部状態に応じたイベント認識機構が働いている. 現内部状態がぬいぐるみの Level 0 から 4 に示されるようなインタラクションの文脈を決定し, 状態に合った出力を得る.



G = G-force sensor, T = Touch sensor, P-h = Hip proximity sensor, P-f = Face proximity sensor, B = Bend sensor, T-h = Head touch sensor, T-f = Face touch sensor

図 2: Interaction Level の遷移

ここであらかじめ与えた遷移信号となるイベントは, ぬいぐるみとユーザの間で行われると想定されるいくつかのインタラクションの中から割り振った. この遷移モデルに基づき, あるイベントはある状態から他の状態への遷移信号となったり, 他のイベントは同じ内部状態を保つための信号となったりする. 例えば, Level 1 のときにぬいぐるみが「抱き上げられる」というイベントを認識したとき, ぬいぐるみの内部状態は Level 2 へ推移する. これにより, Level 2 における「温かみがあり親密である」コミュニケーションをぬいぐるみが認識し, 内部状態を変化させるという目標は達成されている. 他にも, Level 2 の状態の時に, ぬいぐるみは常に様々なタイプのリズムカルな入力を検出する認識器を動かしている. そして何らかのリズムカルな入力があったと認識したとき, ぬいぐるみの内部状態は Level 3 へ推移する. そして Level 3 においては「音楽を表現メディアとしたコミュニケーション」を達成する.

### 3.4 センサぬいぐるみの構成

センサぬいぐるみの本プロトタイプにおいて, すべてのセンサ入力をぬいぐるみの PC で内部処理できる. 今回の実験システムでは, 音楽生成や音響合成を行う

別の PC Station へ、センサのデータを無線 LAN によって送り出している。PC Station 内でセンサのデータは MIDI に変換され、そこで内部状態の遷移とそれぞれの内部状態に基づいた音楽生成操作を行う。それらのシステムを図 3 に示す。今後はすべての処理をぬいぐるみ内の PC で行う予定である。

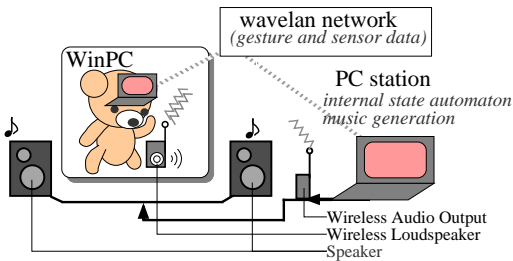


図 3: システム構成

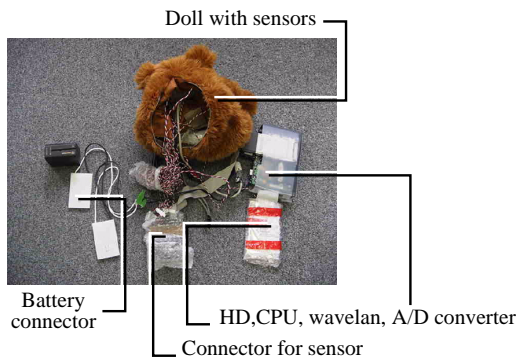


図 4: ぬいぐるみ内部の設置状況

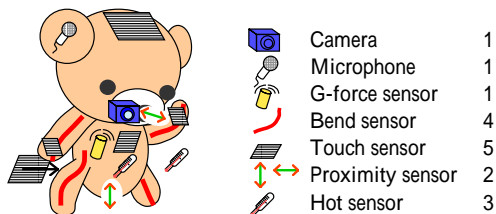


図 5: センサの設置

センサぬいぐるみは、140mm x 100mm x 40mm 程度の小さな PC (OS は Windows2000) と、無線 LAN カード (Wavelan)、電力供給部分 (7.2V, 2360mAh の電池使用、ただし実験時には AC/DC 変換器により交流 100V から直接電力供給)、A/D 変換器、そして様々なタイプのセンサを内側に含んでいる (図 4 参照)。センサは、ぬいぐるみの鼻の部分に USB カメラ、耳に USB マイクが埋め込まれている。他に、それぞれぬいぐるみ内部の、腹部に G-force (加速度) センサ、手足それぞれに曲げセンサ、ぬいぐるみの内部と外部それぞれに温度センサ、鼻と臀部に接近センサ、頭部・腹

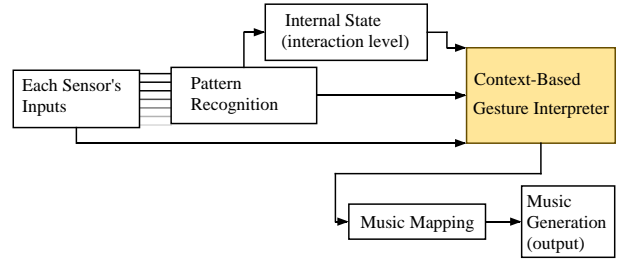


図 6: センサぬいぐるみのデータ処理の流れ

部・背中・両手に圧力センサ (タッチセンサ) を設置した (図 5 参照)。これらのセンサは、例えば抱きしめるというしぐさにおいて鼻の接近センサと他の圧力センサを組み合わせるといったように、複合的に用いられることもある。

上記のセンサの値は、まずパターン認識器に送られ、そこから得られたジェスチャーやイベントの信号は内部状態の遷移信号として用いられる他、センサからの値と同時に文脈適応型ジェスチャーインタプリタへ送られる。そこで内部状態に応じた音楽要素マッピングにより、各入力に対応した音楽生成を行う (図 6 参照)。

### 3.5 文脈に応じた音楽表現

ここで、唯一のアクチュエータとして用いられる音楽や音の表現のデザインについて述べる。音楽の要素は、例えばメロディーが同じであっても伴奏の和音が異なる時違った響きを生む。我々はまず下記の音楽制御要素を準備し、ぬいぐるみの気分としての内部状態とそれに応じた音楽マッピングを試みた (図 3.4 参照)。周囲を包む環境としての表現を実現するため、室内スピーカを用いると同時に、ぬいぐるみの内部に設置した小型スピーカへ無線で声のような音を送り、主体的な音を出力した (図 3 参照)。

1. グローバル: 全体の音量・和音の種類・キー・テンポをコントロール
2. 呼吸の音: 呼吸の間隔・音量・反響フィルタの強さ・音響フィルタにより作られた音の組み合わせの数
3. 声のような音 (フィルタをかけられ変化した声のサンプリング): 音量・フィルタにより作られる音の中心周波数・再生速度・反応速度 (delay)
4. メロディー: 音程・音の長さ・音量
5. リズム: 音量・リズムパターン

ここから、あらかじめ想定したそれぞれの Interaction Level (3.2 節参照) に合った要素を選択し、イベントやセンサからの直接の値に適宜マッピングした。マッピングのパターンは 30 種類ほどあるが、そのう

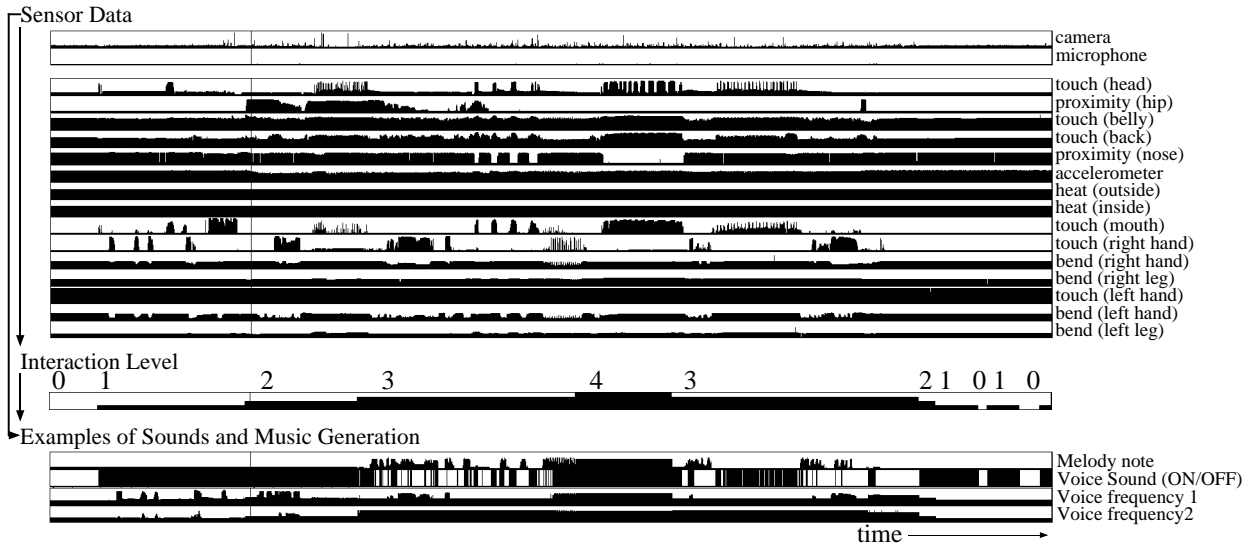


図 7: センサ入力から Interaction Level , 音楽生成までの流れ

表 1: 音楽制御マッピングの例

入力のタイプ	Interaction Level	音や音楽制御要素
両手を握る	2	和音の構成音数
	3	メロディーの音量
左手を曲げる	1	声の音の周波数
	2	音階に添った声の音程
	3	メロディーの音程

ち多くのマッピングは音楽によるコミュニケーションの段階として Interaction Level 3 で用いられている。

内部状態によってマッピングが変更するため、ぬいぐるみに対する同じ接触の種類でも異なる音楽や音を出力する機構になっている。例えば、手を握り握手をするというユーザの入力が、Level 0 においては Level 1 へ移行する遷移信号となり、一方、Level 1 においてはその触り方などが反映された返事のような声を出す。また、Level 2 においてはフィルタをかけられた呼吸の音がどのように和音を構成するかその構成音数に反映され、Level 3 においては音楽コミュニケーションの主体と考えているメロディーの音量に反映される (図 1 参照)。他にも様々な音楽や音の要素の制御をマッピング変更することで、1) 内部状態 と 2) その時その時の入力の組み合わせ により、発生する音楽や音の響きが異なる。

### 3.6 リズミカルな入力の検出

ここで音楽によるコミュニケーションへ遷移するという Level 2 から Level 3 への移行部分 (図 2 参照) に

ついて説明する。温かい親密なコミュニケーションから音楽的な (規則を伴った) コミュニケーションへ変化する際のきっかけとして、リズミカルな入力を用いることにした。またこのリズミカルな入力からテンポを検出し、Level 3 ではそれにあつたリズムを反映することとした。

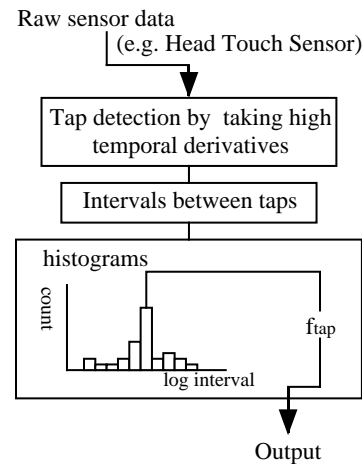


図 8: リズム検出手法

図 8 にリズム検出の手法を示す。頭部、背中といった接触センサや加速度センサなどからの直接入力値 (図 9(a)) と、200ms 前のその値との差分 (図 9(b)) をとり、その差分があるしきい値を越えて有意であると判断されるとき、叩いていると認識する。その後その叩いている入力の時間間隔を計算し、時間間隔ごとにヒストグラムを作成して、最も頻度の高い時間間隔を検出す



ることで、その時間間隔を検出されたテンポと考える。このヒストグラムは現在 2000ms の窓内で逐次計算している。

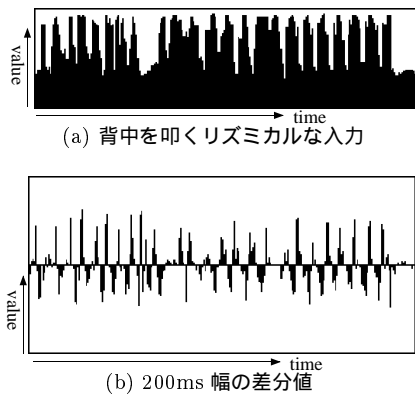


図 9: 背中の圧力センサからの値の例

すなわちリズムカルな入力の時間間隔  $t$  の逆関数を周波数  $f$  とし、対数周波数の値域を等分割してヒストグラムを作成した。

$$f = t^{-1} \quad (1)$$

$$\log_{10} f = \{-0.2, -0.1, 0.0, \dots, 0.9\} \quad (2)$$

上記のように  $\log_{10} f$  を  $-0.2$  [0.63Hz] から  $0.9$  [0.126Hz] まで  $0.1$  毎に等分割している。

図 9 に例示したデータから得た時間間隔のヒストグラムが図 10 である。 $\log_{10} f = 0.2 \sim 0.3$  の範囲で最も頻度の高い入力間隔があったと見なし、その平均  $0.25$  をとり、 $t_i = 10^{-0.25} \approx 0.5623(sec)$  毎に入力が行われていると判断される。

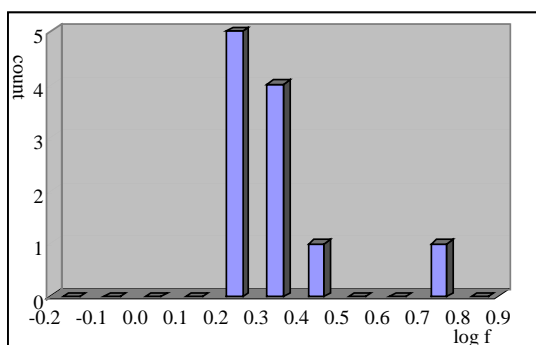


図 10: 入力間隔のヒストグラムの例

### 3.7 予備実験

我々はこのぬいぐるみシステムを公開展示し(図 11 参照)、システムに詳しくないユーザがぬいぐるみに対してどのような行動をとるか観察した。親しみを持って



図 11: 公開展示の様様

接する人の中には幼児や乳児のように反応を見ながらぬいぐるみを扱う人や、反対に単純な音楽コントローラとして扱う人がいたが、圧倒的に前者が多かった。中には、どこにカメラやマイクなどのセンサが埋め込まれているか等の情報を得ているにも関わらず、センサのないふさふさの部分で撫でて反応をみるなど、触覚インタラクションを試みる人も観察された。実行風景を図 12 に示す。



図 12: 実行風景

## 4 考察

### 4.1 コミュニケーションデバイスとしてのぬいぐるみ

本システムの目標は、図 1(b) に示したように、ぬいぐるみと人間の間で行われる非言語コミュニケーションにおいて人間が主体性を持ち表現活動を行うことを、音楽への翻訳を通じ実現することである。特にぬいぐるみによるまごとの遊びの中で観察される、自分の分身としての用途と、パートナーとしての用途が融合した使い方をされているかについて議論する。

即時に反応を返すコミュニケーションの場合、ユーザの表現を実現することが容易で、ぬいぐるみ自身の表現と見なされるのは時間遅れのある反応を得るコミュニケーションではないかと考えられる。ここで、Level 2, 3 では即時に音楽要素に反映するマッピング、Level 0, 1 では時間遅れのある反応を主にマッピングを与えることで、結果として Interaction Level に応じて徐々

にぬいぐるみから人へ主体性が移るデザインが実現すると考えている。実際、公開展示の観察を通して、ぬいぐるみの外形を利用することで、人が親しみをもちつつ主体的に表現をするインタラクションが生じていたことが観察された。

現システムにおける内部状態の遷移図は有限な形であるが、それをユーザ毎に変更可能な、ユーザ固有のぬいぐるみの内部状態として準備することで、ユーザが直接入力の変換フィルタを変更可能な人間同士のコミュニケーションデバイスとして拡張できると考える。

次に前節に述べた予備実験から、ぬいぐるみとの接触を含んだマルチモーダルインタラクションについて考察する。ぬいぐるみという受け身のデバイスに対して、様々な入力の可能性がある中で多くは感情を伴う様々な接触によるコミュニケーションの試みが観察された。これらの入力はだまかに 1) ぬいぐるみの注意を引く、2) 意図を持ってぬいぐるみに何らかのジェスチャーをさせるなどユーザ側からの入力を行う、3) 無意識にぬいぐるみに対する入力を行うという場合に分けられた。人間の無意識な身体動作 [25] は、例えば貧乏揺すりのようにその人の気分を無意識のうちに表わす。ぬいぐるみの背を叩くなどの無意識のリズミカルな入力が、ぬいぐるみの内部状態の遷移を引き起こすことにより、音や音楽のフィードバックが、ユーザ自身にその無意識の入力を気付かせる機能も有していると考えられるが、現時点では未調査である。

このように、ユーザの意図や意識に関わらず様々な入力をぬいぐるみの視点や状況から解釈し、フィードバックを返す本システムは、常にユーザの意図通りに表現活動を行う直接操作型のデバイスではない。しかし、ぬいぐるみの様子や反応を見ながらインタラクションする状態と、ユーザが主体性をもち表現する状態とが融合したことが、予備実験における、親しみをもちつつ主体的に表現をする様子より観察されたと考えられる。

#### 4.2 音楽コミュニケーションの可能性

本システムでは、既存の意味が固定的に与えられたメディアではなく、自由な表現と解釈の広い可能性を有する音楽をアクチュエータとして採用した。それにより、接触を中心としたマルチモーダルインタラクションを通じ音楽表現を行うことと、意図的もしくは無意識な非言語インタラクションが音楽やその組み合わせによって気付かされたり、喚起されることをねらっている。また、その音楽表現は従来楽器と異なり、ぬいぐるみの内部状態を通じた音楽表現となるため、ユー

ザによる直接的な表現ではない。これは、例えばやりわりと相手にもものを伝える時など、ぬいぐるみの内部状態による翻訳がユーザの表現の助けとなる場合に有効だと考えられる。

#### 4.3 ジェスチャーや文脈の認識機構

様々なセンサを取りつけられたぬいぐるみシステムを構築したが、現段階では単純なイベント認識器のみ使っている。今後はぬいぐるみ内部の PC を活用し、HMM によるジェスチャー認識、文脈認識をぬいぐるみの内部処理で行うことを検討している。HMM を用い新たなジェスチャーや状況のモデルをユーザが学習させることで、より複雑なジェスチャーや文脈・状況認識を行い、より正確な環境の情報やユーザの入力を得ることができる。その結果として、より多くの種類のぬいぐるみの内部状態を認識したり、より複雑な遷移図を準備することができ、インタラクションの種類を豊かにできると考えている。

例えば図 1(b) のモードでユーザがぬいぐるみに状況やジェスチャーを学習させたりそれを修正しておいて、それらを図 1(c) のモードで他の人間とのコミュニケーションデバイスとして用いるといったデザインが考えられる。

### 5 本システムの評価手法の提案

これまでの考察・議論を確かめるため、ここで本システムにおけるぬいぐるみインタラクションを評価する手法について考察し、我々がぬいぐるみを使用したことによって効果を得ると考えている点に焦点を当て、以下の評価実験を提案する。

#### 5.1 モードが変化することの効果

予備実験において、ユーザがぬいぐるみをコントローラとして扱う様子とコミュニケーションの相手のように扱う場面が観察された。それぞれの用途として用いられたり、それらを融合させて楽しむ様子と、Interaction Level が変化するシステムの構成の関連を確かめる。

ぬいぐるみが状況を判断し、それによって内部状態(モード)を変化させ、文脈適応型インタプリタによって表現形態を変える本システム構成が、ユーザインタラクションにおよぼす影響があるかどうか、またその中で音楽を用いた新しいインタラクションに慣れ親しむことができるかを調べる。

3.1 節で述べた入力の時間・結果により 4 タイプに分類したぬいぐるみとのインタラクションモデルを導入することで、そのインタラクションを豊かにすること



ができたかどうかについて評価する手法を下に述べる。

まずこれらのインタラクションモデルを大きく二つに分けると、即時の反応を得るモデルとインタラクションのモードを変化させるモデルに分類できる。このモードの変化がある場合と無い場合を比較することにより、上記の Interaction Level の遷移によるユーザの反応の変化を観察することができる。

そして同様に、入力の種類が即時であるか、それとも時間を伴った入力パターンを判断しているかという分類をした。即時の入力と応答というインタラクションモデルのみ準備した場合のユーザインタラクションを観察するのも評価方法として有効であると考える。

#### [ Level 変化と入力の多様性に関する比較実験 ]

モードの変化の存在する現在のシステムの効果を評価するため、モードが変化しないシステムを準備し、比較実験を行う。

- Interaction Level の遷移による効果の評価実験

5種類の Interaction Level を準備した現在のシステムと、音楽的な制御が中心となる Interaction Level 3 のみを持ち、その Level が変化せず、また音楽要素コントロールのマッピングも変化しないシステムを準備する。そしてそれぞれのシステム毎に被験者を準備し、Tactile なセンサや、音声、画像などのユーザインタラクションを記録する。

- 時間を伴った入力と即時の入力に対するそれぞれの反応を準備したことについての比較実験

様々な入力の種類を想定し準備したことによる効果を調べるため、上記の Interaction Level 3 のみの単純なシステムを用い、従来の Level 3 における音楽要素マッピングと、時間を伴った入力を受け付けず、単純な入力にのみ音楽の反応をマッピングするシステムを比較実験する。

#### 5.2 音楽コミュニケーションの評価

次に、既存の意味をもつ言葉ではなく、音楽をアクチュエータに用いたことによる効果について評価実験を行う。音楽を表現メディアに用いたことで、おしゃべり人形のようなコミュニケーション形態から変化したか、またその組み合わせが生む響きがインタラクションにおよぼす影響はあったかについて評価する。

おしゃべり人形は、大抵一つの入力に対し、時間的なまとまりを持った例えば言葉の列など一つの反応を返す。それに対し、音楽の自由な表現は一つ一つの要

素が一つ一つの入力によって制御され、言語表現における制御に当てはめると「あ」の声や「あ」から「い」への連続的な変化のパラメータなどを制御することになるため、直接比較しづらい。

そこで、現在マッピングされている音楽要素コントロールに対し、音楽以外の別の要素をマッピングすることでインタラクションに変化があらわれるかを観察する。ここでは、従来システムと比較するための二つのシステムを準備する。まず、Actimates Barney のようなおしゃべり人形を想定した言葉のマッピングを準備すると同時に、それらの中間のコントロールとして、あらかじめ時間を伴った音楽をマッピングする。

#### [ 言葉と音楽の反応による比較実験 ]

この実験では、モード変化のない比較的単純な Interaction Level 3 のみを用いる。具体的なコントロールの要素として、

1. 時間的なまとまりを持った言葉の列のマッピング
2. 時間的なまとまりを持った音楽 (楽曲の一部など) のマッピング
3. 本システムにおける音楽要素マッピング

の3種類を準備し、それぞれユーザインタラクションを観察する。

#### 5.3 ユーザインタラクションの分析

ここでは、センサからの入力によりユーザインタラクションを観察する手法を採る。図 3.4 に示すように、ぬいぐるみとの音楽コミュニケーションでは、1) 存在しているぬいぐるみ自身に対する入力 2) 音楽の反応に対する入力を分類することができる。そこから、音楽出力との照らしあわせを行うために、音楽出力との時間的な関係を計測する。また、比較実験においてユーザの入力がどの程度親しげであったかを、自由なインタラクションの全体の時間、抱き上げている時間、という主観的な状態に対する時間の計測のみではなく、Tactile センサの値の大きさ、入力時間の長さ、同時入力センサの数といった客観データにより判断することを現在考えている。

#### 6 おわりに

本論文では、ぬいぐるみとのインタラクションを導入した音楽コミュニケーションのためのシステム “Commusic” を紹介した。ぬいぐるみをユーザの分身として、あるいはパートナーとして、接触を含む様々なマルチモーダルインタラクションの実現を試み、それに

よりコミュニケーションのための音楽を生成する事について考察した。最後に本システムのモダリティを持つインタラクションの実現に対し新たな評価手法を提案した。

今後は、本システムを人間と人間のコミュニケーションデバイスとして用いるための拡張として、ユーザ固有のジェスチャーや状況を学習によってシステムに獲得させることを目指す。そのための課題として、ぬいぐるみ内部のPCによって1)HMMによる状況認識・ジェスチャー認識、2)内部状態の遷移と3)音楽生成部分といったすべての処理を行うことを考えており、現在は状況認識器に様々なジェスチャパターンを学習させる段階である。また、音楽表現のマッピングをユーザが与えることにより、より豊かな表現を持つユーザ固有のコミュニケーションデバイスが実現すると考えている。

## 謝辞

本研究を進めるにあたって、ATR 知能映像通信研究所の中津良平社長、Seon-Woo Lee 氏、他多くの方々にご協力頂いたことをここに感謝する。

## 参考文献

- [1] Abowd, G., Atkeson, C., Hong, J., Long, S., Kooper, R. and Pinkerton, M.: Cyberguide: A mobile context-aware tour guide, in *Wireless Networks*, Vol. 3, pp. 421-433 (1997).
- [2] Sumi, Y., Etani, T., Fels, S., Simonet, N., Kobayashi, K. and Mase, K.: "C-MAP: Building a context-aware mobile assistant for exhibition tours", in *Community Computing and Support Systems*, pp. 138-155 (1998).
- [3] 米澤朋子, Clarkson, B., 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 ヒューマンインタフェース HI92-3 (2001), to appear.
- [4] 米澤朋子, ブライアンクラークソン, 安村通晃, 間瀬健二: 文脈に応じた音楽表現を伴うセンサぬいぐるみ, 情報処理学会 Interaction2001 シンポジウム IPSJ Symposium Series (2001), to appear.
- [5] Yarger, L.: Vivarium History, <http://www.beanblossom.in.us/larry/vivHist.html>.
- [6] Johnson, M., Wilson, A., Kline, C., Blumberg, B. and Bobick, A.: Sympathetic Interfaces: Using Plush Toys to Direct Synthetic Characters, in *Proceedings of CHI99*, pp. 152-158 (1999).
- [7] 稲葉雅幸, 星野由紀子, 井上博允: 導電性ファブリックを用いた全身被覆型触覚センサースーツ, 日本ロボット学会誌, Vol. 16, pp. 80-86 (1998).
- [8] Hoshino, Y., Inaba, M. and Inoue, H.: Model and Processing of Whole-body Tactile Sensor Suit for Human-Robot Contact Interaction, in *Proceedings of 1998 IEEE International Conference on Robotics and Automation*, Vol. 3, pp. 2281-2286 (1998).
- [9] iRobot Inc.: My Real Baby, <http://www.irobot.com/mrb/index.htm>.
- [10] Hasbro Inc.: Furby, <http://www.furby.com/>.
- [11] Strommen, E.: When the Interface is a Talking Dinosaur: Learning Across Media with ActiMates Barney, in *Proceedings of CHI 98*, pp. 288-295 (1998).
- [12] 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, Vol. SIG-KBS-9904, pp. 55-60 (2000).
- [13] 鈴木紀子, 竹内勇剛, 石井和夫, 岡田美智男: 非文節音による反響的な模倣とその心理的影響, 情報処理学会論文誌, Vol. 41, No. 5, pp. 1328-1337 (2000).
- [14] Camurri, A., Coletta, P., Ricchetti, M. and Volpe, G.: Synthesis of expressive movement, in *Proceedings of International Computer Music Conference 2000 (ICMC 2000)*, pp. 270-273 (2000).
- [15] Suzuki, K., Tabe, K. and Hashimoto, S.: A Mobile Robot Platform for Music and Dance Performance, in *Proceedings of International Computer Music Conference 2000 (ICMC 2000)*, pp. 539-542 (2000).
- [16] Wassermann, K. C., Blanchard, M., Bernardet, U., Manzolli, J. and Verschure, P. F.: ROBOSER - AN AUTONOMOUS INTERACTIVE MUSICAL COMPOSITION SYSTEM, in *Proceedings of International Computer Music Conference 2000 (ICMC 2000)*, pp. 531-534 (2000).
- [17] 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).
- [18] 多田幸生, 西本一志, 前川督雄, 間瀬健二: 人や場を演出する音楽創奏システムの提案, 情報処理学会研究報告 音楽情報科学 2000-MUS-37, pp. 1-8 (2000).
- [19] 多田幸生, 西本一志, 前川督雄, 間瀬健二, 中津良平: CosTune: 状況に応じた事故演出を可能とする装着型楽器の実装, 情報処理学会 Interaction2001 シンポジウム IPSJ Symposium Series (2001), to appear.
- [20] Clarkson, B., Mase, K. and Pentland, A.: Recognizing User Context via Wearable Sensors, in *ISWC2000 Proceedings*, pp. 565-569 (2000).
- [21] 間瀬健二, Clarkson, B., 米澤朋子: 幼児期からのウェアラブル Toy 型インタフェース, 情報処理学会研究報告 ヒューマンインタフェース HI92-1 (2001), to appear.
- [22] Sony Corporation: Aibo, <http://www.aibo.com/>.
- [23] NEC : パーソナルロボット R100, <http://www.incx.nec.co.jp/robot/indexj.html>.
- [24] 大野彩子, 樋口文人, 安村通晃: 電子ペットを用いた対話型心拍トレーニング支援システム, in *WISS2000*, 第 24 巻, pp. 237-238, 日本ソフトウェア科学会 (2000).
- [25] Raffer-Engel, von W.: *Aspects of Nonverbal Communication*, Swets and Zeitlinger (1980), ノンバーバルコミュニケーション, 本名 伸幸, 井出 祥子, 谷林真里子, 大修館書店.