

文脈に応じた音楽表現を伴うセンサぬいぐるみ

米澤 朋子^{†‡}, ブライアン クラークソン[§], 安村 通晃[‡], 間瀬 健二[†]

[†](株)ATR 知能映像通信研究所, [‡]慶応義塾大学大学院 政策・メディア研究科, [§]MIT Media Laboratory
{yone,mase}@mic.atr.co.jp, clarkson@media.mit.edu, yasumura@sfc.keio.ac.jp

1 研究の背景

我々はぬいぐるみとのインタラクションが文脈に応じた音楽を表現するシステムを新しいコミュニケーションツールとして提案する。ぬいぐるみにカメラやマイク、圧力センサなど様々なセンサを埋め込み、ぬいぐるみ内部に設置したPCによってユーザとぬいぐるみとの間に行われるインタラクションのレベルを文脈として判断する。本稿では、インタラクションの頻度・強度が音楽マッピングを変更し、状況や文脈によって異なる音楽を作る新しいタイプのぬいぐるみとのコミュニケーションについて説明する。



図 1: 擬人化ぬいぐるみコミュニケーション

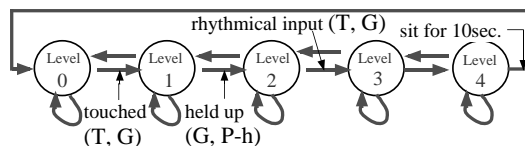
ぬいぐるみと人間のインタラクションは主体性を人間が司り、一方的に仮想的なコミュニケーションを展開しているという見方もできる。ここで我々は、ぬいぐるみとの、文脈を持ち、接触を中心としたマルチモーダルなインタラクションに着目した。そして、人間同士の非言語コミュニケーションを補う親しみやすいデバイスとして、ぬいぐるみを活用することを提案する。図 1(a) に示されるようにロボットは活動的であるために強い人格を持ち、ユーザとロボットのインタラクションにおける関係は対等に近くなる。そこで人間が主体性を持つ、図 1(c) の人間同士のコミュニケーションデバイスを実現するためのステップとして、図 1(b) に示すぬいぐるみと人間とのインタラクション部分を設計した。

Johnson ら [1] による “Swamped!” では、ぬいぐるみはユーザ入力をセンサから判断し、あらかじめ準備された視覚的な仮想世界で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る。“ActiMates Barney” [2] は子供の学習のパートナーとして製品化されている。ぬいぐるみが単純な返答をするモードのみでなく、テレビやコンピュータとつなげて連動することもできる。しかし、現段階では、他のぬいぐるみ製品と同様にあらかじめ言葉などの動作が録音されているため、反応の種類も乏しい。我々は、ぬいぐるみが、主体である人間自身の環境的要素になるよう考慮し、音や音楽をアクチュエータとすることで意味を限定しない独特の表現が導入できると考えた。そして、あらかじめ単純な反応の枠組みと

して音楽の要素を与え、ユーザがその反応の変化を楽しみながら、ぬいぐるみをコントローラのように操作したり、同時にぬいぐるみを擬人化してコミュニケーションを行うことを目指す。“Wearable Sensor System” [3] は、ユーザの身の回りのカメラやマイクにより HMM(隠れマルコフモデル)を用いて、庭や地下鉄、台所などユーザの状況を認識を行う。ここでのセンサによる状況分析の手法は、我々のシステムで多くのセンサにより複合的な状況認識を行う参考になると考えた。

2 システムデザイン

図 1(b) の人とぬいぐるみとの間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デザインした。まず、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した。これらのぬいぐるみの状況はユーザとのインタラクションの強さ (Interaction Level) に対応して遷移すると想定し、ぬいぐるみの内部状態を 5 つに分けた。Level0 は、眠っているが周辺に興味があり、呼吸音を立てる。Level1 はユーザとの出会いの段階で、接触に反応して返事のような声を返す。Level2 は温かみのある、より親密なコミュニケーションで、呼吸音が和音に変化し、返事のような声も音階に沿った音程になる。Level3 は音楽によるコミュニケーションの段階で、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。最後に Level4 はコミュニケーション不能の暴走状態で、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。



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

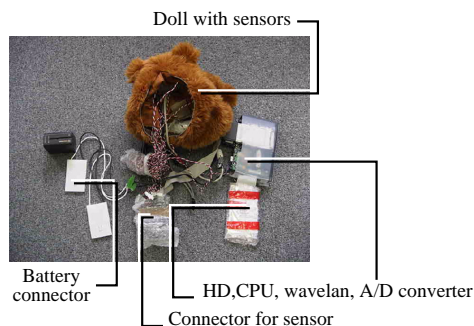


図 3: ぬいぐるみの内部

Sensor-Doll with Context-aware Music Expressions: 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

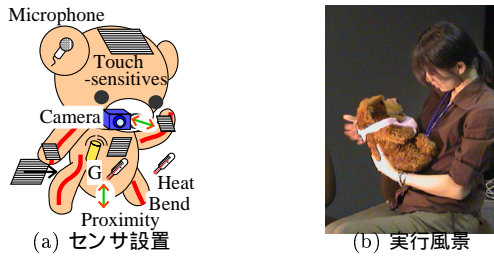


図 4: ぬいぐるみのセットアップと外観

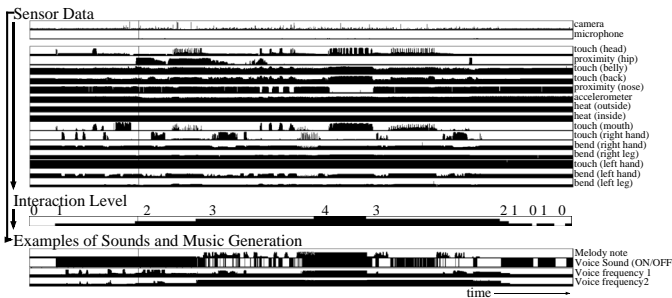


図 5: センサ入力から、音楽生成までの流れ

内部状態の有限オートマトンは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(図2)。ぬいぐるみの現在の内部状態が文脈を決定し、状態に合った出力を得る。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 が示す温かい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。また、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おり、何らかのリズムカルな入力があったと認識したするとぬいぐるみの内部状態は Level 3 へ推移する。そして Level 3 において音楽を表現メディアとするコミュニケーションを達成する。

ぬいぐるみは無線ネットワークカードを取り付けた小さな PC、電源、A/D 変換器、様々なセンサを内蔵している(図3)。ぬいぐるみの表面や内部にカメラ、マイク、加速度センサ、接近センサ、温度センサ(内外の温度差を図る)、複数の圧力センサと曲げセンサを埋め込んだ(図4(a))。センサの値と、パターン認識器を通じたジェスチャーやイベント、そしてそこで得られた現在の内部状態を、音や音楽を生成する別の PC Station へ無線 LAN によって送り出している。PC Station 内でこれらの値は MIDI に変換され、MAX/MSP を用い文脈適応型インタプリタを通して内部状態に応じた音楽要素マッピングにより音楽情報に変換される。これにより各入力に対応して音楽要素を操作することができる(図5参照)。

次に音楽や音の表現として、1) 全体の音量・音階・和音・テンポ、2) 呼吸の音の音量・間隔・フィルタの強さ、3) 声の音の音量・中心周波数・再生速度・反応の時間遅れ、4) メロディーの音量・音程・音の長さ、5) リズムの

音量・パターンという5種類の要素を準備した。これらの要素を用い、入力に対するそれぞれのぬいぐるみの内部状態に応じた音楽マッピングを試みた。周囲を包む環境としての表現として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音を送り、主体的な音を出力した。

3 考察

ぬいぐるみに対するユーザの行動を公開展示において観察したところ、反応を見ながら親しみを持って扱う場面や、逆に単純な音楽コントローラとして扱う場面が見られた。これは人間が主体性を持ち、ままごと遊びに見られるパートナーまたは分身のようにぬいぐるみを用いて、音楽を通じた表現活動を行っていたと考える。従来楽器と異なり、ぬいぐるみの内部状態によって同じ接触でも異なる出力を得る。ユーザの直接の表現ではなく、ぬいぐるみの内部状態による翻訳と、組み合わせによって違った響きを生む音楽独特の表現は、やんわりと相手にもものを伝える時などに有効である可能性もある。

また、カメラやマイクなどのセンサがあることを知りつつ、触覚インタラクションを試みる人が多く観察されたことから、ぬいぐるみというデバイスでは接触を含んだインタラクションが重要であると考えられる。

4 おわりに

我々は、“Com-music”により、ぬいぐるみをユーザの分身として、またパートナーとして用いた、接触を含むマルチモーダルインタラクションの実現を試み、状況に応じた音楽を生成するインタラクションを提案した。

今後は、ユーザ固有のジェスチャーや状況をシステムに学習させることを目指す。そのため、ぬいぐるみ内部の PC で、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、3) 音楽生成部分といったすべての処理を行うことを考慮し、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、内部状態の有限オートマトンを、ユーザの固有の変更可能な内部状態のオートマトンとし、音楽表現のマッピングをユーザが与えることができるシステムを目指す。それにより、ユーザ固有のより豊かな表現を持つ人間同士のコミュニケーションデバイスに拡張したい。

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] Johnson, M.P., Wilson, A., Kline, C., Blumberg, B. and Bobick, A., “Sympathetic Interfaces: Using Plush Toys to Direct Synthetic Characters”, Proceedings of CHI99, pp. 152-158, 1999.
- [2] Strommen, E., “When the Interface is a Talking Dinosaur: Learning Across Media with ActiMates Barney”, Proceedings of CHI98, pp. 288-295, 1998.
- [3] Clarkson, B., Mase, K. and Pentland, A., “Recognizing User Context via Wearable Sensors”, ISWC2000 Proceedings, pp. 565-569, 2000.