

文脈適応型音楽生成をともなうぬいぐるみインタラクション

米 澤 朋 子^{†*} ブライアン クラークソン^{††} 間 瀬 健 二^{†††}

本論文では、文脈に応じた音楽表現を用いた、新しいぬいぐるみインタラクションを提案し、その効果を評価するものである。様々なセンサを埋め込んだぬいぐるみとの接触インタラクションにより、あらかじめ与えたぬいぐるみの内部状態を変化させ、その状態に応じた音楽を生成するシステムを実装した。また、人間同士のコミュニケーションにおける、本システムの音楽表現による効果を評価した。

Tactile Sensor-doll Interaction with Context-aware Music Expressions

TOMOKO YONEZAWA,^{†*} BRIAN CLARKSON^{††} and KENJI MASE^{†††}

We present a sensor-doll capable of music expression as a sympathetic communication device. The doll has a computer and various sensors to recognize its own situation and the activities of the user. It also has the internal "mind" states to reflect different situated contexts. The user's multimodal interaction with the passive doll is translated into musical expressions that depend on the state of mind of the doll. Finally we evaluate the effect of music expression to the human communication using the tactile doll.

1. はじめに

人間とコンピュータのインタラクションにおいて、人間同士がしているコミュニケーションと同じ自然さを提供するようにするには、コンピュータをつねに決まった答えを返す箱ではなく、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とすることが考えられる。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多く察知し、認識することができれば、行われるインタラクションに豊かな要素を与えることができるようになると考えられる。最近提案された個人化ガイドシステムにおいても文脈適応を考慮したものがある。たとえば Abowd らによる Cyberguide システム¹⁾では教室内などのグループや個人による情報アクセスにおいて、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、Sumi 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、場所や時間あるいは行動パターンを状況として利用している。

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1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、その触覚をともなった擬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。同様に、大人にとっても擬人化によって身替わりや代理としての役割を持つ存在として影響を与える可能性がある。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。本文では、音楽コミュニケーションの実験のために試作したぬいぐるみシステム^{3),4)}について、システムの詳細と初期実験の結果を示し、評価観点を提案する。そして、ぬいぐるみを用いて音楽生成を行うことがコミュニケーションの表現として有効であることを検証する。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

[†] ATR 知能映像通信研究所
ATR Media Integration & Communications Research Laboratories

^{††} MIT メディア研究所 (米国)
MIT Media Laboratory

^{†††} ATR メディア情報科学研究所
ATR Media Information Science Laboratories

* 現在、日本電信電話 NTT サイバースペース研究所
Presently with NTT Cyber Space Laboratories

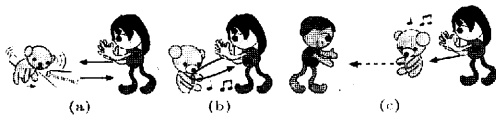


図1 擬人化ぬいぐるみによるコミュニケーション
Fig. 1 Communication with stuffed-toy.

センサなど様々なセンサと、それ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。このシステムが人間同士のコミュニケーションでどのような効果を発揮するかについて検証するため、音楽表現デバイスや表現音を変え比較実験を行った。

ぬいぐるみに類似するものとして近年ペットロボットなどが注目されているが、ロボットとのインタラクションにおいては、ロボットは自律的に活動して強い人格を表すためユーザとロボットの関係が対等になる。そのため、コミュニケーションの対象となり、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、身体的な活動性が低いぬいぐるみに注目した。図1(c)のような人間同士のコミュニケーションを支援する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音声出力や身体動作を用いないこととした。

本論文では、まず、2章においてこれまでに行われているいく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あげ、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章では、我々が開発したセンサを埋め込んだぬいぐるみシステムの内部処理、ハードウェア、そしてインタラクションにともなう音楽生成手法について具体的に記述する。評価実験の手法と結果について4章で紹介し、最後に5章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

ぬいぐるみや人形やロボット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人形型インタフェース研究が多方面からなされている。

KayらのVivarium Projectでは、Druin⁵⁾がコンピュータと人間のインタフェースにぬいぐるみとの接触を用いる試みを行った。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は、我々の研究と関連が深い。それらに対し我々は、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)」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。

また、おもちゃ市場ですでに取り扱われているセンサを導入した様々なぬいぐるみシステムが存在している。しかしそれらのシステムは、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すぎない。“ActiMates Barney”⁷⁾は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。単純な返答をするモードのみでなく、ビデオやコンピュータソフトとつなげて連動するぬいぐるみにもなる。ユーザスタディにおいて子供はある操作に一貫したリアクションを期待する傾向を見せつつも社会的状況によってはまったく異なるリアクションを期待していると報告している。

土井ら⁸⁾は音声を、また鈴木ら⁹⁾は非文節音を用い、ぬいぐるみとのインタラクションを提案している。前者は言葉によりエージェントとしての役割を果たし、後者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。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、なかでも音楽による表現を採用した。ユーザが、ぬいぐるみをアバタとして、またエージェントとして適応的に扱うことを目標とし、あらかじめコントロール可能な音楽マッピングを与え、状況に応じてそのマッピングを変更する。

非言語表現のうち音楽を用いる対話的なシステムとして、Camurriら¹⁰⁾やSuzukiら¹¹⁾が開発した自律的に動く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て提案をしている。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。我々はそ

れに対し、受動的なぬいぐるみを用いることで人間と人間の表現支援の役割に適応的に動作させることができると考え、コミュニケーションデバイスとしてセンサぬいぐるみを活用することを提案する。

3. システムデザイン

3.1 インタクションモデルと内部状態の設計

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間のインタクションを念頭に、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。ここで、ぬいぐるみとのインタクションに、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、それらを織り混ぜた豊かなインタクションを行うことができると思った。

はじめに、人間同士のコミュニケーションを参考として、入力イベントとして単純操作と状況変化の2種類を想定し、それに対し、直接的反応とモード反応の組合せを考える。すると次の4タイプに分類できる。

Type 1: 単純な入力に対して即座に反応をする

Type 2: 連続した入力群に対して即座に反応をする

Type 3: 単純な入力に対して反応のモードが変わる

Type 4: 連続した入力群に対して反応のモードが変わる

たとえば、Type 1は、子供を叩くと泣くなどの即座な反応を指す。Type 2は、ある種のジェスチャを見せることで乳児が笑うなどの反応があてはまる。Type 3は、思いやりのある言葉でその後の信頼関係が生まれるなど、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あげられる。Type 4は、ジェスチャやダンスなどにより、相手に意思を伝え「モード」が切り替わるといったことが考えられる。

上にあげた4種の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、A) コミュニケーションのモードになる内部状態とそれを変化させる機構、および、B) それぞれのモードにおいて入力を音楽出力に直接翻訳する機構、を設計した。まず、ぬいぐるみの内部状態をユーザとのインタクションの強さ・激しさ (Interaction Level = IL) の変化によって遷移させる。ぬいぐるみの気分などを表すため、乳児とのインタクションを想定し、5つの内部状態を準備した (表1参照)。

3.2 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、図2に示すような内部状態の有限オートマトンを準備した。

表1 インタクションレベル (IL) : 人形の内部状態
Table 1 Interaction Level (IL): Internal states of doll.

IL	内部状態	出力する音
L0	眠っているが周辺に興味がある状態	呼吸音を立てている
L1	ユーザとの出会いの段階	接触に反応して返事のような声を返す
L2	温かみのある、より親密なコミュニケーションの状態	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
L3	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、和音を持つ即興の曲をユーザと協調して演奏する
L4	コミュニケーション不能の暴走状態	リズムやメロディは存在するものの、混乱を表す音楽を出力

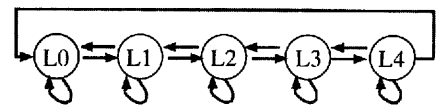


図2 インタクションレベルの遷移モデル
Fig. 2 Transition model of Interaction Level.

それぞれの内部状態において異なる種類の入力信号がイベントとして認識され、遷移のトリガとなる。そのため、各内部状態に応じたイベント認識機構が働いている。ぬいぐるみの各内部状態が、Level 0 (L0) から Level 4 (L4) として示されるようなインタクションの文脈を決定し、状態に応じた出力が得られる。

ぬいぐるみとユーザの間で想定されるいくつかのインタクションの中の、インタクションモデル (3.1節参照) の Type 3 と Type 4 にあたると考えられるものを、遷移信号のイベントとして用いる。Type 3 は出会いの段階でユーザに変化を気付きやすくする必要のある部分に主に採用した。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。たとえば、L1 のときにぬいぐるみが「抱き上げられる (Type 3)」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 L2 へ推移する。これは、L2 における親密なコミュニケーションに状況が変化したことをぬいぐるみが認識するための機構として用いられる。他にも、L2 の状態のときに、ぬいぐるみはつねに様々なタイプのリズミ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「リズミカルな入力 (Type 4)」があったと認識したとき、ぬいぐるみの内部状態は L3 へ推移する (表2参照)。

3.3 センサぬいぐるみ “Com-Music” の構成

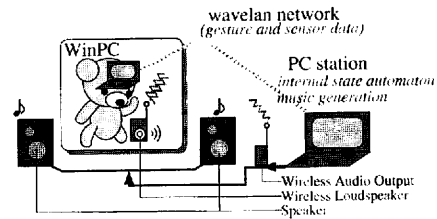
センサぬいぐるみの本プロトタイプにおいては、す

表 2 状態遷移信号とコミュニケーションのタイプ
Table 2 Transition signal and communication type.

IL	遷移信号	タイプ
L4	ぬいぐるみを抱いている状態から降りし、しばらく放っておく	Type 4
L0	ぬいぐるみのどこかに触れる	Type 3
L1	ぬいぐるみを抱き上げる	Type 3
L2	ぬいぐるみのどこかにリズムカルな入力を行う	Type 4
L3	同じジェスチャを繰り返したり、加速度が大きい運動などハードウェアに負荷のかかる無理な入力を行う	Type 4
L4	なだめる動作をする	Type 4
L3	しばらく音楽的な操作を行わない	Type 4
L2	抱き上げている状態からぬいぐるみを下ろす	Type 3
L1	ぬいぐるみに触れるのをやめ、放っておく	Type 4
L0		

すべてのセンサ入力の A/D 変換、前処理をぬいぐるみの PC で実行する。今回の実験システムでは、音楽生成や音響合成を行うぬいぐるみ外部の PC Station を用意し、センサのデータを無線 LAN によって送り出している。PC Station 内ではセンサのデータは音楽プログラミング言語 Max/MSP に合わせて MIDI 形式に変換し、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。PC Station の構成は無線 LAN の受信をする WinPC と内部状態遷移・状態ごとの音楽生成を行う Macintosh (G4, Max-MSP) を用いている。本構成は評価実験のためのラビッドプロトタイプに適している。外部の PC Station で行う計算処理を内蔵 PC で行うことで、ぬいぐるみ単体での実装は技術的に可能である[☆]。

センサぬいぐるみは、140 mm × 100 mm × 40 mm 程度の小さな PC (PentiumII 333 MHz, OS は Windows2000) と、無線 LAN カード、電力供給部分 (7.2 V, 2360 mAh の電池使用、ただし連続実験時には AC/DC 変換器を介し交流 100 V 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 (a) 参照)。センサは、ぬいぐるみの鼻の部分にカメラ (USB i/f)、耳にマイク (USB i/f) 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 (1 軸, ピエゾ素子型, 重力加速度) センサ, 手足それぞれに曲げセンサ (ピエゾ素子型), ぬいぐるみの内部と外部それぞれに温度セ



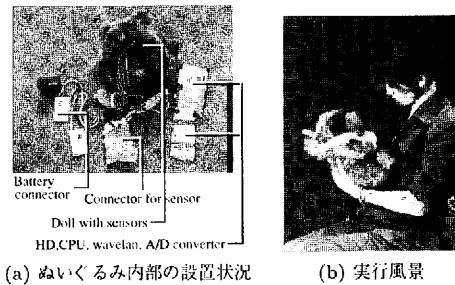
(a) システム構成



(b) センサの設置

図 3 システム構成

Fig. 3 System configuration.



(a) ぬいぐるみ内部の設置状況

(b) 実行風景

図 4 センサぬいぐるみ

Fig. 4 Sensor-equipped stuffed toy.

ンサ、鼻と臀部に接近センサ、頭部・腹部・背中・両手に圧力センサ (感圧抵抗型タッチセンサ) を設置した (図 3 (b) 参照)。これらのセンサは、単一のセンサ入力を用いたり、センサの複数の組み合わせで状況を判断するといったように、複合的に用いたりすることもある。たとえば、カメラの輝度値の差分がしきい値を超えると、目の前に動くものの存在を検知する (単一のセンサ値による判断) ほか、重力加速度センサの変化の差分と臀部の接近センサが一定のしきい値を超えることにより、抱き上げられていることを判断する (複合的な判断)。また他にも、鼻の接近センサがある程度のしきい値を超え、かつ他の部位にある圧力センサにかけられている力の総合がしきい値を超えている時、抱きしめるというしきい値を判断する。

図 5 に示すように、センサの値はまずパターン認識器に送られ、パターン認識器から得られたジェスチャ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ほか、センサからの値と同時に文脈適応型ジェスチャ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

☆ 音楽生成処理において市販のツールキット Direct Music を利用し、ほぼ同等の処理全体をぬいぐるみ内部の 1 台の PC で処理可能であることを確認している¹²⁾。

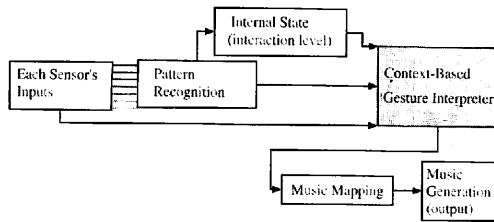


図5 センサぬいぐるみのデータ処理の流れ
Fig. 5 Diagram of data processing in sensor-doll.

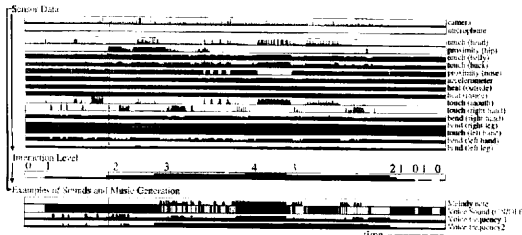


図6 センサ入力から IL、音楽生成までの流れの例
Fig. 6 Example of sensor inputs and corresponding output.

生成を行う。上記“Com-Music”システムの実行風景を図4(b)に示す。

3.4 文脈に応じた音楽表現

ここでは、唯一のぬいぐるみの出力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、たとえばメロディが同じであっても伴奏の和音が異なるとき違った響きを生む。我々はまず音楽制御要素：a)呼吸の音、b)声のような音、c)メロディの音、d)リズムの音、e)音全体のコントロール、を準備し、ぬいぐるみの気分としての内部状態と、それに適したセンサ情報の音楽マッピングを試みた(図6、表3参照)。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)。

ここから、あらかじめ想定したそれぞれの内部状態におけるIL(3.1節参照)に合った音楽要素を選択し、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L3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。たとえば、手を握り握手をするというユーザの入力が、L0においてはL1へ移行する遷移信号となり、一方、L1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、

表3 ILごとの音楽制御マッピング例
Table 3 Music mapping for each Interaction Level.

Level	入力	対応する音
L0	温度 カメラの輝度変化 or 音量変化 or 顔に接近	呼吸音の早さ 呼吸音の間隔
L1	手の曲げ L1の継続時間	声の音の中心周波数 声の反応遅れ減る
L2	曲げ入力の頻度 手の曲げ 両足の曲げ平均 抱きしめる 両手を握る	呼吸の早さ 声の音の中心周波数(音階上の音程) 呼吸音の音量 呼吸音のフィルタ .1 和音構成音数 .2
L3	リズムカルな入力 右手曲げ 左手曲げ 左手を握る 足の曲げ 頭を触る 腹か両足との接触 顔に接近	音楽のテンポ リズムタイプ (1-4) 全体の和音選択 (1-6) 声の音の中心周波数 メロディの音程 呼吸音の音量 全体の音量 全体の調 (1-12) 呼吸音のフィルタ メロディの音の長さ
L4	なだめる回数 頭に長く触れる	調の変更幅を減らす .3 和音構成音数

* 1: 反響フィルタの強さで楽音的な音へ連続的に変化させる。
* 2: 周波数帯域調整した呼吸音を複数重ねて和音を構成する。
* 3: なだめる回数により調を変更する random 幅を減らす。

L2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、L3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の音量に反映される(表3参照)。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合せにより、発生する音楽や音の響きが異なる。

3.5 リズムカルな入力の検出

ここで音楽演奏によるコミュニケーションへ遷移するというL2からL3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(規則をともなった)コミュニケーションへ変化することのきっかけとして、リズムカルな入力を用いることにした。またこのリズムカルな入力からテンポを検出し、L3ではそれに合ったリズムを反映することとした。

図7(c)にリズム検出手法を示す。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7(a))と、200ms前のその値との差分(図7(b))をとり、その差分があるしきい値を超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。このTap入力のインターバルを用いて、過去2,000msの検出窓内でインターバルを検出し、各インターバル値に対す

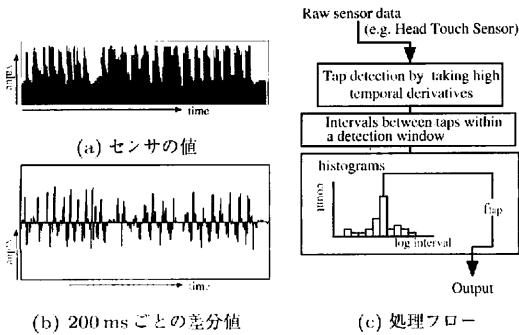


図7 リズム検出手法

Fig. 7 Rhythm extraction method.

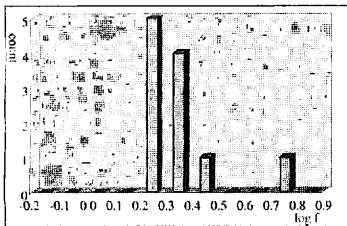


図8 入力間隔のヒストグラムの例

Fig. 8 An example of histogram of input interval.

るヒストグラムを逐次計算する。こうして、最も頻度が高いインターバル値を検出し、それをテンポと考える。なお、ヒストグラムの区間は \log 単位で軸をとり均等化する。

すなわちリズムカルな入力の間隔 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 [0.63 Hz] から 0.9 [7.9 Hz] まで 0.1 ごとに等分割している。

図7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8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、 $t_s = 10^{-0.25} \approx 0.56(\text{sec})$ ごとに入力が行われていると判断される。

3.6 予備実験

本システムの特徴は、1) コミュニケーションチャンネルの手段として音楽表現を用いた点、2) 音楽表現の入力手段にぬいぐるみを用いた点、3) 文脈に応じた表現を導入した点である。そこで、これらの効果を評価する必要がある。1) は音楽表現のコミュニケーションへの影響があり、2) の中には、音楽表現へのぬいぐるみという入力インタフェースを用いたことによる影響と、ぬいぐるみインタフェースに対し音楽表現という

フィードバック音を用いたことによる影響が混在している可能性が考えられる。

我々は予備実験として、このぬいぐるみシステムを公開展示し、システムに詳しくないユーザが本システム全般に対してどのような行動をとるか観察した。その結果、内部状態の変化に一喜一憂し幼児や乳児のように反応を見ながら親しみを持ってぬいぐるみに接する場面が多く見られ、上記3) について、内部状態が遷移することによって、ぬいぐるみシステムにコミュニケーションの相手(エージェント)としての面を持たせることにある程度成功したと考えられる。しかし、上記2) について、ぬいぐるみを音楽表現コントローラ(アバタ)として用いて、他者に表現して見せたり伝えたりするという人は、前者に比べ観察された例が少なかった。これは、公開展示の性質上、本システムを人と人とのコミュニケーションに用いる機会がなかったためだと考えられるが、そのような場面で人がどのように本システムを使い、あるいは受け止めるかを定量的・定性的に調べる必要がある。

そこで、上記予備実験では観察できなかった1) と2) の特徴について、次章で評価実験に取り組んだ。音楽表現が可能なぬいぐるみによってコミュニケーションの新しいチャンネルを提供できるかを調べるため、対話を導入した実験を試みた。

4. 評価実験

4.1 実験手法

本実験では、音楽を表現することができるぬいぐるみが、人と人とのコミュニケーション、特に対話に与える影響について明らかにするため、1対1の対話の場面を設定し、表現する音やぬいぐるみというデバイスの特性について調べた。

被験者を2つのグループに分け、ぬいぐるみや楽器を持つA群(Player)と何も持たないB群(Listener)とする。A群とB群から1人ずつ選んでペアを組み、いくつかの条件下で1分から5分間の会話をしてもらい、その対話の様子を記録する。会話の内容は、メロンについて、トマトについて、といった、単純な対象についてお互いの意見を述べ合うように指示した。

実験の目的は2つである。1つは、ぬいぐるみという接触インタフェースを用いることが会話に与える影響について明らかにすることであり、もう1つはその接触インタフェースが音楽出力をすることの会話に与える影響について明らかにすることである。

実験は大きく分けて2種類行った。ぬいぐるみや他のものを持つことによって発話の量・バランス・内容に

何らかの違いが出る可能性があると考えた。そのため、両実験とも、被験者 B (Listener とする) は何も持たない状態という条件を変えず、被験者 A (Player) は条件ごとにぬいぐるみや従来楽器 (ここではピアノ) により音を鳴らすことができる。そして Player のぬいぐるみや楽器への入力の様子を記録する。使用するぬいぐるみは Com-Music システムに準拠し、手足内部に曲げセンサ、胴体と頭部に圧力センサを入れる。しかし内部に PC を持たず、有線で外部の A/D 変換器につなぎ、Macintosh でセンサの入力の模様を記録する。同時に Player と Listener の発話、Player が鳴らす音をマイクで記録しておく。

実験 1 は、3 つのパートに分かれ、各パートで Player にそれぞれ、1) 音楽的な音が出るぬいぐるみ (Com-Music の L3 の音楽マッピングを省略したもの、これを条件 **doll** とする) を触りながら、2) 音が出るピアノ (条件 **piano**) を触りながら、3) 何も触らず (条件 **no_dev**)、に会話することを指示した。

実験 2 では、1) Player にセンサぬいぐるみを持ってもらい、ぬいぐるみが、メロディのような非連続の音階の音程を出す (**melody**)、2) 変化が連続的で声のような音を出す (条件 **voice**)、3) 何も音を出さない (条件 **no_sound**) ようにパートを分けて会話してもらい、上記同様に観察する。

被験者は 18~35 歳までの男女で、男性同士/女性同士を Player, Listener のペアにし、合計で 14 組 (28 人) の実験を行った。この組は互いに初対面か、会話をはじめの知人とした。被験者には、ぬいぐるみにセンサがあり、コンピュータを介して音が出せる仕組みを説明したあと、各条件ごとにぬいぐるみが出す音が異なることと、各音のマッピングとコントロール方法について各実験の前に示し、Player にはそれぞれのコントロールを試す時間を与えた。なお、条件 **piano**、条件 **no_sound** も同様に説明を行っているが条件 **no_device** では会話を行うことに関する指示のみ出した。特に、ピアノの条件では、個人の演奏スキルによる影響を小さくするため、「演奏してください」という指示ではなく、ただ触れながらもかまわないように、「自由に触れて音を出してください」という指示を出した。また、実験順序が結果に影響しないよう、話題とパートの順番をランダムにして実験を行った。実験の様子を図 9 に示す。

4.2 実施結果の抜粋

4.2.1 実験の概要

実験は 14 組中 13 組の被験者対において 5 分以上のやりとりが見られたため、300 秒で打ち切った。1 組



(a)

(b)

図 9 ぬいぐるみを用いた実験の様子
Fig. 9 Talking experiment with using doll.

のみ平均対話時間が 172.8 秒であった。実験の途中で話題に尽きしばらく Listener, Player とともに無言で話題を探る場面もいくつか見られ、話に困ると楽器やぬいぐるみを触る人もいた。楽器の演奏を披露する場面もあった (それに対し、ぬいぐるみの演奏を披露する場面は見られなかった)。Listener が Player にぬいぐるみによる音を出してほしいと頼む場面もあった。このように 5 分間の長さは話題の盛り上がりや沈滞などの多様な会話の状態を含む平均的データとして利用できるかと仮定する。その他、図 9(a) のようにぬいぐるみを相手に見せながら話す被験者や、図 9(b) のようにぬいぐるみを見ながら話す被験者がいた。しかしながら、文脈適応の音楽表現をともなった予備実験で見られたような、ぬいぐるみを相手として見るような、話しかけたり、反応を確かめたりするといった光景は見られなかった。

4.2.2 発声・操作時間データの分析

Listener の発話を L, Player の発話を P, 楽器やぬいぐるみの音 (Sound) を S で表記する。実験の条件 $A \in \{\text{doll, piano, no_device, melody, voice, no_sound}\}$ との組で、 L_A, P_A, S_A により 1 パート実験ごとの総発話時間を表す*。

分析 1: 実験 1 の条件 **doll** において、Player がぬいぐるみを使ってぬいぐるみに発話させた総発話時間 S_{doll} を Player と Listener の総発話時間の和 $P_{\text{doll}} + L_{\text{doll}}$ で割って正規化した値を求めたところ、被験者ペアごとの値は (平均, 分散, 最大, 最小) = (0.576, 0.053, 1.44, 0.25) となった。つまり、Player と Listener の発話の平均 $(P_{\text{doll}} + L_{\text{doll}})/2$ と同じ程度の、ぬいぐるみによる音の表現がなされているといえる。

分析 2: 次に、Player によるぬいぐるみやピアノの総発話時間 $S_{\text{条件}}$ に着目した。

(分析 2-1) 実験 1 の条件 **piano** でピアノが出し

* 注: 出力は音楽的表現や声のような音が含まれるが、用語を簡易にするため発話時間と呼ぶ。

た総発話時間 S_{piano} は平均 87 秒であったが、14 人中 5 人の Player については 10 秒未満で、それらのサンプルの平均は 3.76 秒、それ以外のサンプルの平均は 133.6 秒であった。このことから、会話の最中にピアノを弾かないグループがあったといえる。それに対し条件 doll でぬいぐるみの総発話時間 S_{doll} は平均 108 秒で、5~250 秒の間にほぼ均等に分散している。

(分析 2-2) 実験 2 で S_{melody} の 14 組の平均は 174.14 秒であるのに対し S_{voice} の平均は 124.75 秒で、T 検定により T 値 2.366 が求まる。これは危険率 $\alpha = 0.05$ の棄却限界値 2.160 を上回り、melody の音の方が voice の音より多く発音されているといえる。

分析 3: さらに、Player と Listener の総発話時間の比率に着目する。会話の主導権は、つねにどちらかが握るものではなく、同じ会話の中でも移り変わりがある。本実験は 5 分間の記録をしており、Listener と Player の会話量のバランスが十分に平均化されたと考える。条件 A における発話バランス $L_A/P_A = L/P_A$ を L/P 値とし、条件間の L/P 値の比の変化を調べ効果を示した。

(分析 3-1) 実験 1 では L/P_{piano} と L/P_{no_device} の比をプロットする (図 10) と、相関の強いまとまりのあるクラスタが見いだされる。そこでクラスタから大きく外れるサンプル (図 10 中、x で表す) を除いて傾向を求める。すると、外れ値を除き線形の関係が見いだせるのに対し、 L/P_{doll} と L/P_{no_sound} はデータに散らばりが見られる。ここから、Player が楽器を触りながら対話をして、何も触れなくても、話者の間にある発話量のバランスはごく一部の人を除き大きく変動しない。それに対し、ぬいぐるみを用いた会話を行うと、二者間の発話バランスに影響し、何も触れないときの比率が維持されない。

(分析 3-2) 実験 2 で、条件 melody と条件 no_sound の比は $y = x$ で近似でき、L/P 値はほぼ等しい (図 11 参照)。それに対し、条件 voice と条件 no_sound の比では傾きが変化し散らばりも増えている。これより、ぬいぐるみの出す音が melody のような音のときと何も音を出さないときでは、発話バランスの関係が変わらず、voice を出すときは発話バランスがずれて、ぬいぐるみを持たない Listener が話す割合が高くなる (または Player が黙る) といえる。

分析 4: 実験 1 の条件 no_device と実験 2 の条件 no_sound において L/P_{no_device} と L/P_{no_sound} の相関を調べたところ、相関係数は 0.04 でほとんど相関が見られない。よって、ぬいぐるみを持つことで Listener と Player の発話バランスは大きく変動している

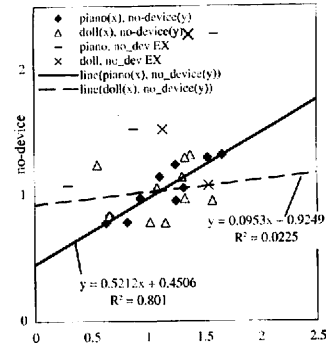


図 10 doll の L/P 値と piano の L/P 値の変化
Fig. 10 L/P value difference between “doll” and “piano”.

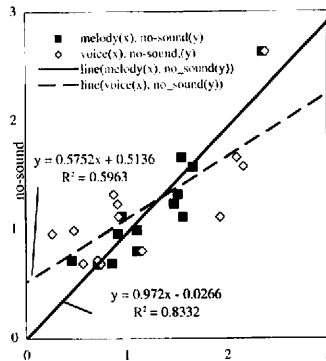


図 11 melody の L/P 値と voice の L/P 値の変化
Fig. 11 L/P value difference between “melody” and “voice”.

ことが分かる。

4.2.3 主観データの分析

実験終了後にアンケートを行った。それぞれの条件における表現や会話について、会話が楽しかったか・親しみを感じたか・うまく会話できたかなどの項目について S/D 法で 1~7 段階の主観評価とした。主な評価結果を表 4, 5 に示す。ここでは、質問項目に該当するほど点数が高い尺度になっている。

5. 考 察

5.1 ぬいぐるみメディアによる会話への影響

まず、分析 1 の結果より、Player は自らの発話と同等の時間長だけぬいぐるみとの接触によりぬいぐるみに発話 (音を出力) させており、会話の最中に音の出るぬいぐるみに触れることに強い抵抗感はないと考えられる。また分析 2-1 より、 S_{piano} で推測される、ピアノによりほとんど音を出さないグループ Φ と演奏をするグループ Θ の S の平均値をそれぞれ S_{doll} と比較すると、 $S_{piano\Phi} < S_{doll} < S_{piano\Theta}$ になる。これは、ピアノという従来型の楽器を使いながらの会話

表4 伝達達成感の主観評価の平均

Table 4 Average of communication achievement feeling.

実験1	Player			Listener		
	doll	pian	no_d	doll	pian	no_d
$P \rightarrow L$	4.57	4.71	4.57	4.93	5.21	4.85
(分散)	1.64	<u>2.84</u>	2.42	2.23	1.57	2.44
$L \rightarrow P$	5.21	4.36	5.07	4.64	4.28	5.28
(分散)	1.72	1.63	2.68	1.94	2.53	2.68

実験2	Player			Listener		
	melo	voic	no_s	melo	voic	no_s
$P \rightarrow L$	5.07	4.38	5.07	4.92	4.57	4.85
(分散)	2.53	1.76	2.07	1.61	1.80	1.98
$L \rightarrow P$	5.29	5.31	5.43	4.93	4.14	4.5
(分散)	1.45	0.73	1.96	1.15	2.13	2.73

pian: piano, melo: melody, voic: voice,

no_d: no_device, no_s: no_sound

表5 会話に関する主観評価の平均

Table 5 Average of subjective evaluation on conversations.

実験1	Player			Listener		
	doll	pian	no_d	doll	pian	no_d
親しみ(相手)	4.57	4.21	4.86	4.36	4.42	4.93
(分散)	1.18	0.95	1.36	2.71	1.19	1.61
会話しやすさ	3.85	<u>3.29</u>	4.64	3.93	<u>3.71</u>	4.5
(分散)	<u>3.82</u>	<u>3.14</u>	2.40	2.22	2.22	2.88
会話の楽しさ	4.64	4.42	5.21	5	4.28	4.85
(分散)	1.93	1.95	0.95	1.53	1.30	1.52

実験2	Player			Listener		
	melo	voic	no_s	melo	voic	no_s
親しみ(相手)	4.36	3.93	4.36	4.43	3.36	4.14
(分散)	0.40	0.23	0.55	0.88	1.48	0.75
会話しやすさ	4.5	4.14	4.79	4.28	3.36	4.07
(分散)	1.37	1.51	1.72	1.14	1.48	1.92
会話の楽しさ	4.71	4.85	4.69	4.92	4.07	4.57
(分散)	1.45	1.05	1.73	0.38	0.84	0.57

条件の表記は表4と同じ

には、量的に会話を補助するような演奏をすることが困難な人が少なくないことを示している。しかし、楽器に慣れていたり、楽器を弾くことにハードルを感じないで弾くことのできるグループは、ぬいぐるみによる発音よりも楽器の演奏が容易であることが推測できる。このことより、ぬいぐるみを用いることにより、楽器を弾きにくいユーザには、ぬいぐるみによって音楽を用いた表現のためのしきいが下げられると考える。

次に、表4および5に着目すると、pianoはListener, Playerともに会話のしやすさで最も評価の平均値(表5中下線, 3.29, 3.71)を下げ、値は分散した。また、「うまく話を伝達できたか」という項目のPlayerの評価においてもpianoの分散が大きい(表4中下線, 2.84)。これらはΦとΘの2つのグループが存在したことを裏付けると考える。また、会話のしや

すさの評価は全体的に大きく分散している。なかでもPlayerはdoll, pianoともに分散している(表5中下線, 3.82, 3.14)。ここで、Player側は、Listener側よりも、自らに音楽表現手段があることで影響を受け、その評価が分散したと考えられる。

最後に、分析4の結果と、分析3-1のL/P値より、ぬいぐるみを用いることにより、それぞれで異なる何らかの変化が見られる。よってぬいぐるみを持つことによって受ける、発話バランスへの何らかの影響(たとえば異なる習慣や背景を持つ人とその組合せ)があることが分かる。また、実験時にピアノの演奏を披露する場面が多く見られたが、その間発話はなく、会話する時間と演奏する時間が分かれていた。そのため、2者間の発話バランスに関して実質上はno_deviceと状況が変わらなかったことになり、分析3-1の L/P_{piano} の結果を説明できる。楽器によって会話をしながら演奏することは、その得意/不得意の差がつく。対照的に、ぬいぐるみによる演奏は容易で、音を聞きながらさらに話すということも可能であるため、発話バランスに多様な影響をもたらしたと考える。

5.2 ぬいぐるみの音楽表現における評価

分析3-2で、各条件どうしのL/P値の比が線形近似できたことより、出力される音の質や有無が発話バランスに一定の影響を及ぼすといえる。 L/P_{melody} と L/P_{no_sound} の比より、発話バランスは、melodyによる音を出力することによっては、ほとんど変化しないことが分かる。それに比べvoiceによる発音はL/P値比が高く(図11の点線の傾きが小さい)、melodyにより発話の邪魔を受けなかったPlayerがvoiceにより邪魔を受けたか、Listenerがvoiceにより発話の刺激を受けたかのどちらかが推測される。また、分析2-2より、声のような連続的な周波数変化をする音より、音楽的な非連続音程の方が発音に何らかの優位性があったと考えられる。さらに主観評価と見比べると、表4よりPlayerはvoiceでうまく話をできず、no_soundが最も話を理解できたと評価しているが、Listenerはmelodyのとき最も話を理解できたと感じ、自分も話を伝えられたと感じると同時に、表5より、やはりmelodyのとき、相手に対する親しみを感じたり、会話しやすく楽しいといった評価をしたり、また、表4, 5のvoiceの評価はPlayer, Listenerともに低く、音高の変動が言葉に似る連続的な音により発話の邪魔を受けたという上記の推測を説明することができる。

ここで、言葉ではない表現メディアとして、音楽を用いることにより、会話と並行した表現が行いやすくなったと考える。さらに、音楽的表現が及ぼす影響は、

表現する側ではなく表現を受容する側に顕著であることが分かる。

5.3 コミュニケーションチャンネルとしてのぬいぐるみ

ぬいぐるみを新しいコミュニケーションチャンネルとするためには、発話がうまく行われることと、発話により得る印象に付加を与えることが条件だと考える。I/P 値を導入した結果をまとめる。

評価実験において、音楽的な表現を用いることで個々に与える会話のしやすさ、相手への親しみといった印象は異なるが、その発話バランスには影響しない。しかしその伝達達成感では、Player は連続的な声のような音に大きく差をつけ、音がしないと同様の評価であったのと、さらに受け手の達成感は音がしないよりさらに評価が高かった。会話は1人で行うものではなく、Player のみではなく Listener に対しても発話を活性化させるか、現状を維持することができるため、音楽的表現が、会話に平行したコミュニケーションチャンネルとなりうることを示唆すると考える。

また、ぬいぐるみを持つことで様ではない影響を受けつつも、ぬいぐるみによる音楽表現を行うことは、従来楽器の初心者にあったハードルを軽減し、多くの音表現を行うことができるという結果が得られた。ぬいぐるみを持つことに慣れないことや抵抗を感じるといった悪影響と、ぬいぐるみを持ち表現することで楽器では得られない楽しみや相手への親しみという良い影響があると考え。このことより、音楽を用いた表現を行う際に、ぬいぐるみは楽器による表現を行えないユーザに有効であると考え。

5.4 文脈適応型状態遷移システムについて

3.6 節では、文脈適応性を含む評価項目を設定し、予備実験においてシステムの総合的な評価を行った。本システムをぬいぐるみ形態の文脈適応型システムとすることによって、全体としてユーザとの関係付けが可能であることは予備実験から示唆された。

文脈適応性には多数の要因が含まれているため、まず、ぬいぐるみの形態と音楽表現の評価項目について着目し、4 章で人と人との対話に状況を固定した評価実験を行い、Player と Listener 間のコミュニケーションへの影響について考察した。この実験では文脈によって内部状態が遷移しないようにした。前述したとおり、文脈適応のある予備実験ではぬいぐるみを擬人化してあやしたり相手をしたりしようと試みていた。それに対し、内部状態を固定した評価実験では、計測前の準備段階から計測中に至るまで、ぬいぐるみを操作したり、無意識的に触れたりする様子が多く観察さ

れた。状態遷移の有無の影響は大きいと推測される。

今後は、内部状態の遷移による文脈適応性の実現方法が、人間とぬいぐるみのインタラクションと人間同士のコミュニケーション支援において、どのように影響するのかを、要因別に調べる必要がある。それを視野に入れて、本節では、実験などを通じて得られた評価観点を整理し提案する。

各実験において利用者からは、状態の変化が分かりにくいとか、内部状態の遷移がスムーズでないなどを指摘する意見があった。これらを元にさらに深く考察すると、文脈適応の状態遷移システムの設計に関わる要因として、少なくとも、3-1) 内部状態遷移がもたらす音楽出力表現の変化の明瞭性、3-2) 内部状態遷移のトリガとなるインタラクションの種類、3-3) 内部状態の数、3-4) 内部状態の表現特徴（活発性以外）、といった評価観点が考えられる。今後これらの評価観点をふまえた実験を通して、自然なインタラクションが可能なシステム設計に反映することが可能となると思われる。

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“Com-Music”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ことについて考察した。最後に本システムの音楽表現によるインタラクションの実現を評価した。

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) 高度な状況認識・ジェスチャ認識、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。また、今回の評価実験において、ぬいぐるみによる音楽表現の効果を評価する手法として別のぬいぐるみを使用したため、Com-Music システムを直接使った実験ではない。そのため、Com-Music でも今回と同様の定量的評価を行うことが必要だと考えている。さらに、本システムの重要な要素である文脈適応型音

楽表現についての定量的評価は、今後の重要な課題である。

謝辞 本研究を進めるにあたって、元 ATR 知能映像通信研究所の中津良平社長、Seon-Woo Lee 氏、慶應義塾大学の安村通見教授はか多くの方々にご協力いただ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, *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, *Community Computing and Support Systems*, pp.138-155 (1998).
- 3) Yonezawa, T., Clarkson, B., Yasumura, M. and Mase, K.: Context-aware Sensor-Doll as a Music Expression Device, *Extended Abstracts of CHI2001*, pp.307-308 (2001).
- 4) 米澤朋子, Clarkson, B., 安村通見, 間瀬健二: め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*, pp.17-24 (2001).
- 5) Druin, A.: NOOBIE: The Animal Design Playstation, *SIGCHI Bulletin*, Vol.20, No.1, pp.45-53 (1987).
- 6) Johnson, M., Wilson, A., Kline, C., Blumberg, B. and Bobick, A.: Sympathetic Interfaces: Using Plush Toys to Direct Synthetic Characters, *Proc. CHI99*, pp.152-158 (1999).
- 7) Strommen, E.: When the Interface is a Talking Dinosaur: Learning Across Media with ActiMates Barney, *Proc. CHI 98*, pp.288-295 (1998).
- 8) 上井俊介, 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).
- 9) 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).
- 10) Camurri, A., Coletta, P., Ricchetti, M. and Volpe, G.: Synthesis of expressive movement, *Proc. International Computer Music Conference 2000 (ICMC 2000)*, pp.270-273 (2000).
- 11) Suzuki, K., Tabe, K. and Hashimoto, S.: A Mobile Robot Platform for Music and Dance Performance, *Proc. International Computer Music Conference 2000 (ICMC 2000)*, pp.539-

542 (2000).

- 12) 齊藤志志, 米澤朋子, 服部進実, 間瀬健二: めいぐるみ I/F による遠隔地間ノンバーバルコミュニケーションの実現, *情報処理学会研究報告ヒューマンインタフェース HI96-3*, pp.15-22 (2001).

(平成 13 年 5 月 28 日受付)

(平成 14 年 5 月 15 日採録)



米澤 朋子 (正会員)

1999 年慶應義塾大学環境情報学部卒業。2001 年同大学大学院政策・メディア研究科修士課程修了。1999 年 9 月～2001 年 3 月 ATR 知能映像通信研究所第二研究室に学外実習生として滞在。音楽インタフェースの研究を行う。同年日本電信電話(株)入社。以来、音声合成の研究開発に従事。日本 VR 学会、日本音響学会、日本音楽知覚認知学会各会員。



ブライアン・クラークソン

Brian Clarkson is a Ph.D. student at the MIT Media Laboratory working with Prof. Alex Pentland. His graduate research involves exploiting techniques from computational perception and machine learning to tackle problems in human behavior and sociology research. Brian has been involved in wearables research since 1998.



間瀬 健二 (正会員)

1979 年名古屋大学工学部電気工学科卒業。1981 年同大学大学院工学研究科修士(情報)課程修了。同年日本電信電話公社(現在 NTT)入社。以来、コンピュータグラフィックスおよび画像処理、そのヒューマンインタフェースへの応用の研究に従事。1988～1989 年米国 MIT メディア研究所客員研究員。1995～2001 年(株)ATR 知能映像通信研究所第二研究室室長。2001 年より ATR メディア情報科学研究所第一研究室室長。コミュニケーション支援のためのインタフェースエージェントの研究を推進している。人工知能学会 1999 年度論文賞、IEEE、ACM、日本 VR 学会、画像電子学会各会員、博士(工学)。