

コミュニケーションキャラクタの感情的振る舞い生成

服部裕介 中野敦 星野准一
筑波大学システム情報工学研究科

本稿では、対話型キャラクタのための姿勢、しぐさ、ジェスチャの複合による感情表現の生成システム(EMBOS:Emotional Behavior Organizing System)を提案する。本システムでは、感情を表現したモーションキャプチャデータに現れる振る舞いを統計的に分析し、姿勢やしぐさ、ジェスチャの表す妥当な感情の強度や頻度を決定する。これらの振る舞いの組み合わせをキャラクタに付与した感情パラメータの強弱の変化や会話内容に合わせてプランニングすることで人間味の感じられる振る舞いを生成する。また、提案システムの有効性を心理コミュニケーションゲームを名づけたキャラクタとの対話コンテンツ上で示す。

Emotional Behavior Synthesis for Communicative Characters

YUSUKE HATTORI ATSUSHI NAKANO JUNICHI HOSHINO
University of Tsukuba

Abstract-In this paper, a system to synthesize emotional expressions by integrating gestures (EMBOS: Emotional Behavior Organizing System) is proposed. In this system, behavior that shows up in motion capture data that express emotions are statistically analyzed to determine the appropriate intensity and frequency of occurrence of emotions that are expressed by pose, motion and gesture. By planning the combination of such behaviors to be consistent with changes in the intensity of emotional parameters that are attached to a character and the content of a dialogue, behaviors that are human-like are synthesized. In addition, the effectiveness of the proposed system is demonstrated in the dialogue content with a character that was named the mental communication game.

1 はじめに

近年、利用者と会話する機能を持つ対話型キャラクタは、ゲームのNPC(Non Player Character)、トークンヘッドと呼ばれる上半身キャラクタ、アプリケーションのガイド役など様々なエンタテインメントコンテンツやアプリケーションに用いられている。対話型キャラクタの振る舞いの表現力を向上することは、これらのコンテンツへの没入感を高めるために重要である。

しかし、これらのコンテンツにおける対話型キャラクタは、利用者からの問い掛けに対して、常に同じ姿勢のまま返答したり、何度も同じしぐさを繰り返したりするなど、内面の感情の変化を感じられず無機質な印象を与えているという問題がある。

そのため本稿では、キャラクタに付与した感情パラメータの変化や会話内容に合わせて姿勢やしぐさ、ジェスチャを適切にプランニングすることで心理的な振る舞

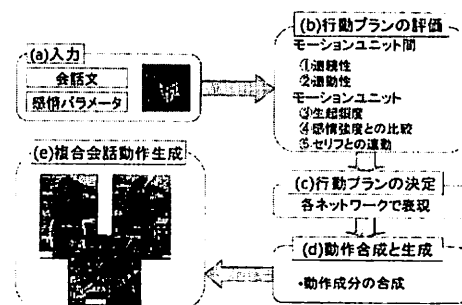


図1 EMBOS(Emotional Behavior Organizing System)の概要

いの表現力を向上する感情表現の生成システム(EMBOS: Emotional Behavior Organizing System)を提案する(図1)。

提案システムのポイントは2つある。まず、一つ目のポイントは、姿勢やしぐさ、ジェスチャの表す心理印象

を用いて会話動作を階層的にプランニングすることである。それらの振る舞いを複合して表現することで、より人間的で多様な振る舞いを生成することができる。本稿では会話内容を補足するジェスチャに加えて、心理状態を表現する姿勢、しぐさを動的に複合することで内面に存在する心理状態を表現した会話動作を生成する。まずモーションクリップにジェスチャ、姿勢、しぐさを合成するための情報を付加したモーションユニットをノードとしたネットワークを構築する。その後、心理状態や会話文に対してそれらの遷移系列をプランニングすることで心理状態を表現した複合的な振る舞いを生成する。

二つ目のポイントは、統計的な感情の定量化を行っていることである。感情は曖昧な概念であるため、その定量化は難しい問題である。従来研究でも主観的な定量化がなされてきた。しかし、我々は統計的に振る舞いの表す感情を定量化する。まずモーションアクターに同じシーンを様々な感情で演じて貰い、感情ステータスと対応付けのとれたモーションキャプチャデータのセットを用意する。そのモーションキャプチャデータのセットを分析し、振る舞いの含まれる回数や持続時間をもとに感情を表す度合いを定量化する。

キャラクターの感情変化に対応した振る舞いを生成するためには、それぞれの振る舞いの出現に適した感情パラメータを定める必要がある。しかし、感情は曖昧な概念であり、その定量化は難しい問題である。これまでも[1]や[16]によって感情的な振る舞いを生成する手法が提案されてきた。しかし、歩く、走る、物を取るといった周期的な動きや短期の動きの生成のみの適用であり、姿勢、しぐさ、振る舞いによる複合的な感情表現については検討が行われて来なかった。

そこで、我々は基本感情別に特定のシーンにおけるアクターの動きを計測し、そのデータを統計的に解釈することで感情値を定量化する。このシステムでは感情の変化に応じて適切な振る舞いを生成するために、6種類の基本感情と通常の状態の計7種類のモーションキャプチャデータを入力とする。これらのモーションキャプチャデータに表れる姿勢やしぐさ、ジェスチャの持続時間や出現頻度を統計的に分析する。これらの情報を元に振る舞いの表す感情の強度を決定する。

キャラクターの自律行動の計画は、[13]や[16]によってネットワーク構造を用いた手法が提案されている。しかし、これらは動きの滑らかさや脚の接地などといった拘束条件に重点が置かれており、会話動作や感情表現については触れられていない。我々は姿勢、しぐさ、ジェスチャを組み合わせて感情表現するために階層的なネットワークを構築する。一つの振る舞いを一つのノードと

したネットワークを各属性に構築する。それぞれのノードには、統計的に定量化された感情パラメータが付与される。このネットワークのノード間には、振る舞い同士の遷移し易さを表す連続性の評価値を与える。異なるネットワークのノード間には振る舞い同士の同時生起の可否を表した連動性の評価値を与える。次に、姿勢の遷移やしぐさの生起を自律的に促すために、振る舞いの生起頻度の評価値を設ける。そして、振る舞いとキャラクターの心理状態や会話内容との連動性を評価する。最後に、これらの評価を行いながら、現在の状態に最適な振る舞いを各属性のネットワークから選択する。それらの振る舞いを階層的に重ね合わせ最終的な振る舞いとして生成する。

2 関連研究

2.1 会話動作生成

CGキャラクターの表現力を向上するために、発話内容や音声に連動したジェスチャ生成や感情を表す顔表情生成など様々な研究が行われてきた。会話に伴う振る舞いについての従来研究では、[3]は、話が明確に区切られた語句から構成されることや、最も強調される共通のポイントでジェスチャとスピーチは同調するといった構造に着目して、適切で表現力のある会話動作を生成する技術を提案している。また入力されたテキストから、特定の単語や、単語間の関係を抽出してCGキャラクターのジェスチャを割り当てるシステムとして BEAT がある[4]。これらの手法では複数のモーションクリップの中から一つを順次選択して再生することで振る舞いが表現されているため常に同じ姿勢を取るなど、動きが単調に見えてしまうという問題がある。また[5]はジェスチャの時間的なタイミングにより身体の階層を考慮して複合的なタスクを生成する手法を提案している。しかし、これは視認している側に対して一度に一つの意味を示すことに重点が置かれており、発話に伴うジェスチャに姿勢やしぐさの表す印象を加えるといった複合的な振る舞いを生成することは検討されていない。

これらの研究では利用者へ明確に意図を伝えるという外面的な表現力の向上に重点が置かれている。その一方で内面的な表現力が軽視されてきた。例えば、従来の対話型コンテンツでは同じ台詞に対して同じ動きを繰り返すといった動きの単調さが見られる。そのため内面に存在するはずの心理状態を感じられず無機質な印象を受けるという問題がある。しかし、心理状態を表す姿勢やしぐさにより効果的に内面の心理状態に基づく行動を生成することができれば、このような問題の解決が見込める。

2.2 感情的な振る舞いの生成

感情的な振る舞いの生成はこれまでも試みられてきた。フーリエ原理を用いて感情的な振る舞いを生成する[1]や心理空間を情動方程式で表現し、ロボットに感情的な振る舞いをさせる[2]などがある。これらは歩行や走る動きといった周期的な動きには適用されているが、ジェスチャやしぐさなどの単発的な振る舞いには適用されていない。また、固視微動や前庭動眼反射といった目線などの顔表情に重点が置かれており、姿勢やしぐさのような振る舞いの表す感情には触れられていない。

2.3 振る舞いの分類と心理印象

そこで我々は振る舞いの表す心理印象を基に感情的な振る舞いを生成する。振る舞いは姿勢やしぐさ、発話に伴うジェスチャといった3つの要素に分けられる。

2.3.1 姿勢の表現する印象

“姿勢”とは身体の動きを示さない、立っている、座っている、腕を組んでいる、しゃがんでいる、ひざまずいているなど、身体の静止した状態を指しており、少なくとも1秒以上は持続する身体位置の変化とされている。姿勢の表現する印象についての従来研究として、前傾や後傾、横傾、両手腰当てや体の開放性といった要素が考察されている。

体の向きについては、地位の高い人に話しかけるとき、肩の向きは真っ直ぐであることが見出されている。両手腰当ての姿勢については、嫌いな人と話すとき、もしくは低い地位の人に話しかけるとき腰の両脇に手を当て、肘を突っ張る姿勢が有意に多く用いられることが見出されている。また体の開放性、これは腕組みや脚組み姿勢をしていない体が開放している状態を指すが、一般的に肯定的な意味を持つことが示唆されている。また、前傾や後傾については、好意のある人と会話するように求められたとき、後傾姿勢を有意に減少したことや、前傾姿勢が、後傾姿勢よりも、より肯定的だとして知覚されることが見出されている[6][7][8]。

また[10]は日本人被験者によって、前傾姿勢は、後傾姿勢よりも礼儀正しく柔軟であると判断されると述べている。このように前傾姿勢を取ることや、後傾姿勢の減少は、肯定的な態度を示すことが示唆されている。

2.3.2 しぐさの表現する印象

本稿における“しぐさ”は眉に手をやる、鼻に手をやる、口に手をやる、口を手で覆う、髪を掻き揚げるなどの自己接触行動や、あくびをする、腕時計を見る、のびをする、ひざをゆする、頭部の注視力や胴体の揺れ方といった無意識的な行動として定義する。これらのしぐさは、人を意識した対人向けのメッセージというよりも、本人の内面から自然に生じる動作だと考える。これらの

しぐさは次のような役割を果たしている。例えば、眉に手をやる、鼻に手をやる、口に手をやるなどのしぐさは、顔面の動きを相手の視線に触れないようにして、人に見られたくない表情を無意識のうちに人目に触れないようにガードしている場合が多い。このようにしぐさは心の状態を無意識に表現している。

また本稿では胴体や頭部の動きもしぐさとして捉える。胴体の動きには、リラックスした心理状態の際は緩やかになる、反対に緊張している時はぎこちなくなるといった変化が見られる。頭部の動きについては、[9]によって興味あると判断された刺激を長く見るといった精神状態との関連性が見出されている。

2.3.3 発話に伴うジェスチャ

発話の内容の補足や強調、意図的なメッセージ伝達といった意味のある動作についてはジェスチャとして扱う。発話に伴うジェスチャには、平手で示す、指差す、はやし立てる、なだめるなどの直示的ジェスチャや、形状や物体の軌道を表現する描写的ジェスチャ、上下あるいは左右に刻むように双方向的に小さく動き発話内容を強調する役割を果たすビートがあるとされている[11]。

3 EMBOS -感情的な振る舞いの生成システム

本システムの特徴は、姿勢、ジェスチャ、しぐさの組み合わせによって感情を表現することにある。キャラクタに付与した感情パラメータの変化に合わせて選択される姿勢やしぐさと、会話に伴うジェスチャを複合することで心理的な振る舞いの表現力を向上させる。これにより同じ姿勢で居続けたり、同じしぐさやジェスチャを繰り返したりといった振る舞いの単調さを改善できる。

本稿では、“姿勢”、“しぐさ”、“ジェスチャを合成した振る舞いを複合動作と呼ぶことにする。会話文や心理状態の変化に合わせて複合動作を自動生成するために、次のように動作計画を行う。まずキャラクタの位置や関節角といったモーションを構成する情報を含んだモー



図2 感情空間

ジョンクリップに、感情の状態や会話文に合わせて動作を複合するために必要な情報をペアにして、モーションユニットとして定義する。次に“姿勢”、“しぐさ”、“ジェスチャ”のそれぞれの属性で、モーションユニットをノードとしたネットワークを構成する。それらのネットワーク上の状態を現在のキャラクタの心理状態や会話文によって遷移させることで各属性が次に取るべき状態を決定する。姿勢、ジェスチャ、しぐさの各属性のネットワークで、現状態として選ばれているモーションユニットを合成することで、複合動作を生成する(図1)。

モーションを選択する際には、キャラクタの感情パラメータとモーションユニットの感情強度の類似度や会話との連動といった状態を評価する値、待機時間や持続時間によって得られる頻度の評価値、モーションユニット間の連続性と連動性の四つの評価値を用いて、これらの積を最終的な評価値とする。ネットワークの現在のノードからこの評価値をもとに姿勢、しぐさ、ジェスチャを選択し、選択された姿勢、しぐさ、ジェスチャを合成してキャラクタの複合された振る舞いとして出力する。振る舞いを生成する手順は以下の通りである。

[前準備]

1. モーションデータに種類(直立、指し示す)を注釈する
2. 各種類のモーションに感情強度を付加し、モーションユニットを生成する
3. モーションユニット間の連結関係をもとに階層化ネットワークを構築する

[実行時]

4. キャラクタの感情パラメータと会話を入力とし、適切な振る舞いをプランニングする。

3.1 感情空間

感情を表す軸やパラメータの設け方には諸説ある。本稿では[17]が顔表情の生成に用いた感情モデルをもとに「幸福」「恐怖」「悲しみ」「怒り」「嫌悪」「驚き」「通常」の7つのパラメータを感情の要素として用いる。顔表情の既存研究では「幸福」と「悲しみ」、「恐怖」と「怒り」は対の関係にあるとされているが、姿勢やしぐさなどの振る舞いは同じものが現出する可能性があるため、本稿では全て別の次元にあるものとして図2のような感情空間を用意した。この空間で、それぞれの感情は最小値0、最大値1の軸を持つ。この値の大小で振る舞いの印象を表現する。また、感情パラメータの値が低い場所は感情の現れないことを意味するため「通常」の状態としている。

キャラクタの感情の状態もこの空間上で表される。会話やインタラクションによって遷移し、振る舞いを選択する際にはCGキャラクタの感情パラメータと振る舞

いの感情パラメータを比較し、評価値の大きさを決定する。

3.2 モーションユニット

モーションクリップに含まれる時系列の位置や関節角の情報だけでは、言語や心理状態を踏まえた動作を計画するには不十分である。そのため心理状態を表す動作計画に必要な情報を含んだ単位をモーションユニットとして定義する。モーションユニットに含む項目は、動作の心理的な印象を表す感情強度、発話に伴うジェスチャ、動作の継続と待機期間、拘束状態を定義する。これらの項目を以下に詳しく説明する。

感情強度 各モーションユニットにその印象の強さを表した感情強度ベクトル $I=(i_1, \dots, i_n)$ 、 $i \in [0, 1]$ を設定する。この強度ベクトル I には、怒り、恐怖、喜び、悲しみ、嫌悪、驚き、通常の順で7つの値が設定される ($N=7$)。キャラクタも同様のベクトルを持ち、動作計画時にこの I の持つそれぞれの値と比較することで心理印象の近いモーションユニットであるかを評価する。

発話に伴うジェスチャ 会話文に合わせてジェスチャの配置をプランニングするためには、自然言語とジェスチャの対応付けが必要である。ジェスチャを分析した従来研究には、発話に伴うジェスチャは表現したい情報を言語化した後、その言語中に現れる単語の意味や表象に基づいて生成されるとする単語意味仮説[12]がある。このような仮説をもとに、会話文中の単語とジェスチャが関連付けられるものとして、発話に伴うジェスチャには対応した単語を設定する。

継続期間と待機期間 脚組みなどの振る舞いはある一定時間維持されることが多い。また髪を掻き揚げるといった動作は、一度行った直後に、再び続けて行うことは稀である。このような現象を成立させるために、キャラクタが過去に取った行動を評価しながら複合動作を生成する。具体的にはモーションユニットが生じた後にモーションユニットの状態が持続する時間幅 L_d とその後一定時間生起させずに待機する時間幅 L_w を定義し、持続時間の評価値を通常状態より高く、待機時間の評価値を通常状態より低く定めることで、持続的な振る舞いや非連続な振る舞いを生成する。

拘束状態 モーションクリップによっては身体部位が拘束されていて、合成すると動作の印象を変化させてしまう場合がある。例えば両手でジェスチャを行っている際に腕を組んだ姿勢を合成してしまい、腕を組んだ姿勢の表す印象が消失してしまう場合が挙げられる。そこで動作の表す主成分と拘束されている身体部位の集合を定義し、それらが重なることの無いようにモーションユニット間の連動性を考慮した動作計画を行う。

3.3 階層化プランニング

モーションユニット間の依存関係を次の4つのパラメータで表現する。初期化時に定まるパラメータとして同じネットワーク内でモーションユニットが連続して生起する可能性を表す評価値を連続性値 ω_c 。異なるネットワーク間でモーションユニットが同時生起する可能性を表す評価値を連動性値 ω_l と定義する。 ω_l は異なる階層のモーションユニット間で同じ身体部位を使用しているか拘束されている身体部位がある場合は0、そうでない場合は1の値を取る。これらはシステムの初期化の際に定められ、実行時に変化のない静的なパラメータとして用いる。また持続や待機時間といった時間経過により変動するモーションユニット自身の評価を ϕ_{time} 、動的に変動する心理状態や会話文などの入力される状態の評価を ϕ_{state} とする。 ϕ_{time} はモーション m の開始時刻を t_s とすると $t_s \leq t \leq t_s + L_d$ の区間ではモーションを持続させる必要があるため最大値、 $t_s + L_d < t \leq t_s + L_d + L_u$ の場合はモーションを生起させない区間であるため最小値、それ以外の場合はその中間の定常値を取る。 ϕ_{state} はキャラクタの感情パラメータとモーションユニットに設定されている感情強度の差を同じ振る舞いの属性の他のモーションユニットと比較した場合の順位を元に計算する。またジェスチャ属性の場合は設定された単語が会話文に現れた場合最大値、それ以外の場合は最小値を持つ評価値を乗算する。これらは、時間経過や入力により生じる状態変化を表すため、動的なパラメータとして用いる。

これらのパラメータはモーションユニットの生起評価という点で相互に依存しているため、モーションユニットの生起評価値は乗算の形式で定め、次式により評価値の最も大きいモーションユニット m を次のモーションユニットとして選択する。

$$\max_{m, m_d, M, m_u, eM_d} [\omega_c(m_d, m) \omega_l(m, m_u) \phi_{time}(m, t) \phi_{state}(m, t)] \quad (1)$$

ここで t は時刻、 m_d は現在実行しているモーション、 M は評価対象のネットワークのモーションユニットの集合、 M_d は M の上位階層のネットワークのモーションユニットの集合を表す。全階層について(1)式を評価し、各階層のモーションユニットの一つを選択する。次章はモーションキャプチャデータからこれらのパラメータを求める方法について述べる。

4 感情的な振る舞いの分析

本章では感情に基づくキャラクタの複合動作をプランニングするためのパラメータとして、モーションユニットの感情強度や持続時間、待機時間をモーションキャプチャデータから算出する方法について述べる。提案システムではモーションアクターに怒りや悲しみなどの感情ごとに演じて貰い、それらを計測したモーションキャプチャデータを元に感情を表現した振る舞いを生成する。まず、姿勢、しぐさ、ジェスチャのそれぞれに現れる振る舞いに名前を付けて注釈する。その注釈したデータからそれぞれの振る舞いについて感情強度や頻度を求める。

4.1 モーションの注釈

2.3 節の振る舞いの表す心理印象のゴウの内容に基づいてモーションデータに注釈を行う。感情ごとの振る舞いの特徴を分析するために同じシチュエーションを役者の方に異なる感情で演じ分けて貰い、そのモーションを計測する。シチュエーションは二人で「対話しているシーン」である。注釈は図3のようなモーションの注釈ツールを用いて行った。

4.1.1 振る舞いの分析

今回はフリーマーケットでの対話シーンという設定で行い、店員役のモーションキャプチャデータの注釈付

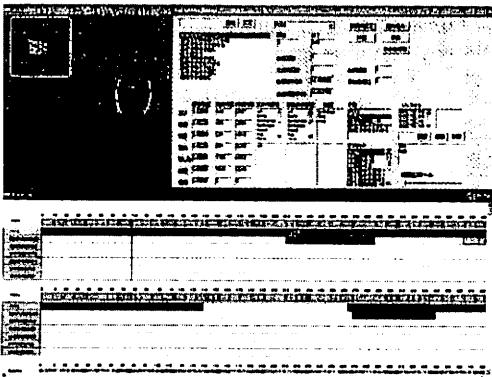


図3 モーションの注釈ツール

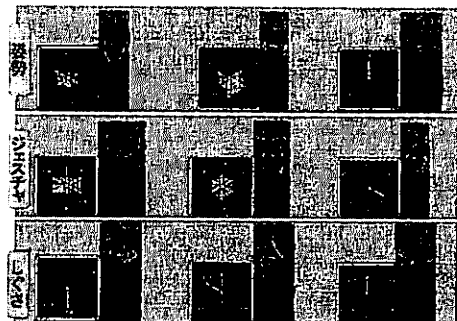


図4 感情強度の例

けを行った。注釈付けをすることで、各振る舞いに感情強度を設定する。注釈付けで得られた振る舞いの感情強度の例を図4に示す。

4.2 感情強度

振る舞いと感情を対応付ける統計情報から得られる要素として次の2つが挙げられる。

一つ目の要素は、様々な感情状態の間で比べたときの、振る舞いの特徴である。複数の感情状態において共通して現れる振る舞いは、特定の感情を印象付ける度合いは低いものと考えられる。反対に、特定の感情だけに出現する振る舞いは、その感情を表す特徴的な動きと考えられるため、感情を表す度合いは高いものと考えられる。本稿では、これを「振る舞いの普遍性」と呼ぶことにする。

二つ目の要素は、ある特定の感情状態の中で比較したときの、特徴的な振る舞いの重要性である。同じ感情状態でも、一般的には様々な振る舞いが行われる。しかし、同じ感情で繰り返し現れる振る舞いは、その感情を表す特徴的な振る舞いであり、感情を印象付ける度合いは高いものと考えられる。本稿では、これを「振る舞いの重要性」と呼ぶことにする。本システムではこの2つの仮定に基づいて統計情報から感情の強度を定める。

4.2.1 姿勢の感情強度

姿勢は一定時間持続する振る舞いであるため、持続時間をもとに感情強度を決定する。特定の感情 e において振る舞い m_d が出現した総フレーム数を F_e とすると、普遍性の強度 s_e は振る舞い m_d が全ての感情で出現する総フレーム数で割った値として、

$$s_e(m_d) = F_e(m_d) / \sum_{i=1}^N F_i(m_d) \text{ のように表す。感情 } e \text{ に出現}$$

する全ての振る舞い D の中での m_d の出現割合の順位を r とする。これは感情 e における重要度を表す。 s_e を値の大きさの順にソートした順位を n 、次のように姿勢の感情強度を定義する。

$$i_e(m_d, s_e(m_d)) = \alpha E(r(m_d)) + (1 - \alpha) E(n(s_e)) \quad (2)$$

ここで $E(x) = (D - x + 1) / D$ 、 $\alpha = [0, 1]$ とする。本稿では $\alpha = 0.3$ とした。第一項目は振る舞いの重要性を評価するための項である。第二項目は振る舞いの普遍性を評価するための項である。

4.2.2 しぐさ、ジェスチャの感情強度

しぐさやジェスチャは姿勢と異なり状態の持続は無いため姿勢の際には持続時間で評価していた部分を出現回数で評価する。感情 e における振る舞い m_d の出現回数を η_e とすると普遍性の度合いは全感情における振る舞い m_d の出現回数で割ったものとして

$$u_e(m_d) = \eta_e(m_d) / \sum_{i=1}^N \eta_i(m_d) \text{ のように表すことができる。 } n$$

を u の順位として姿勢の場合と同様に(2)式でしぐさとジェスチャの強度を定める。

4.3 持続時間と待機時間

感情 e に出現した振る舞い m_d の持続時間として、その平均値を用いる。待機時間 L_w は全体のモーションの時間長との比率関係から次のように求められる。

$$L_w = \frac{L_e - \sum L_d \bar{L}_d}{\sum L_d} \quad (3)$$

ここで L_e はその感情のモーションキャプチャデータの時間長から姿勢を注釈していない部分を除いた時間長とする。この持続時間と待機時間は感情別に計算する。

5 感情的な振る舞いの再構成

キャラクタの自律行動時には、喜び、怒りなどの特定の感情を示すだけでなく実際の人間と同じように中間的な感情の状態を設ける。この中間的な感情の状態に合うように持続時間や待機時間、姿勢の状態を混合した振る舞いを生成する。

キャラクタの感情パラメータを $I = (i_1 \dots i_N)$ とし、このパラメータに基づいた振る舞いを生成する。この中の通常状態だけは次のように定義する。

自律行動時のキャラクタの感情状態は基本的にモーションユニットに割り当てたものと同じであるが、通常状態は感情の表れない状態であるため、各感情の強度が低い場合に通常状態の強度が大きくなるように以下の

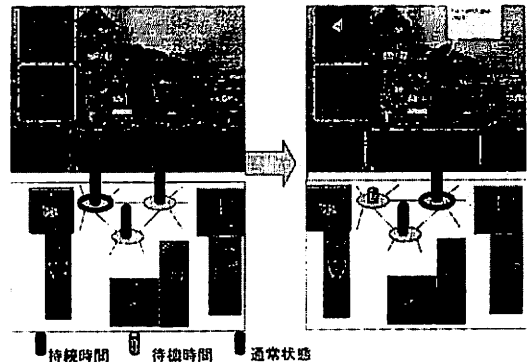


図5 コミュニケーションゲームへの適応例

ように定義する。

$$i_{neutral} = \begin{cases} 1 - \alpha & \alpha < 1 \\ 0 & otherwise \end{cases} \quad \alpha = \sum_{k=1}^{N-1} i_k \quad (4)$$

この通常状態の強度を含めた Γ に適した持続時間や待機時間、姿勢の補間を行い、その時点での中間的な感情を表現する。

5.1 感情に基づいた姿勢、顔度の補間

複数の感情に現れる姿勢を現在のキャラクタの感情に合った振る舞いにするために動作成分を補間する。モーションキャプチャデータを観察すると、同じ種類の姿勢でも感情が異なる場合には身体部位の位置や向きに変化がみられる。例えば同じ腕組みや両手に腰を当てる姿勢でも感情によって頭の傾きや背中丸み、身体の向きなどが異なる。そのため中間的な感情値に合った姿勢を生成する必要がある。RBF を用いた感情的な振る舞いの補間手法は[15]で提案されている。本稿では RBF を用いて以下のような式で姿勢を補間する。

$$m = \frac{\sum_{k=0}^{N-1} m_k \sigma(m_k) \Gamma(i_k, i_k)}{\sum_{k=0}^{N-1} \sigma(m_k) \Gamma(i_k, i_k)} \quad (5)$$

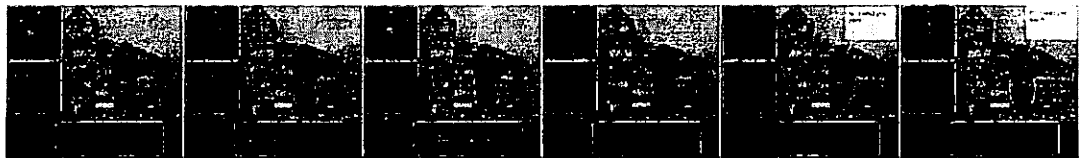
ここで σ は m_k が存在するとき 1、それ以外では 0 を表すものとする。 Γ は動径基底関数 $\Gamma(i_k, i_k) = \exp[-(i_k - i_k)^2 / r^2]$ を表す。

また、同じ種類の姿勢でも感情が異なると持続時間や待機時間は異なる。そのため(5)式の m を L という値に

6 実験

本手法の有効性を確認するために評価用コンテンツ(図5)を作成した。このコンテンツは心理状態を表現した複合動作をプレイヤーが見ながら交渉相手の心理状態を探り、適切な返答を選択して目的の品物を得るゲームである。本稿ではこのゲームを心理コミュニケーションゲーム (Mental Communication Game) と呼称する。従来のゲームでも心理状態を振る舞いで表現するゲームは存在したが、振る舞いによって意図を伝えることに重点が置かれていたため、現実の人と交渉しているような印象は受けなかった。このゲームは意図を伝えるジェスチャに加えて内面的な心理状態を表現した姿勢としぐさを合成する本手法を用いることで、初めて実現可能となったゲームである。

実験では、姿勢、しぐさは変化せずにジェスチャの変化のみで生成された行動例と、交渉に成功した場合の複合動作、失敗した場合の複合動作を比較した(図6)。これらの例を比較した結果、姿勢、しぐさの変化しない例では、CGキャラクタの考えていることが理解できず、機械的な印象を受けた。これに比べて姿勢、しぐさを变化させた例では、適切に心理表現をしているように感じられた。この結果、主観的な評価ではあるが姿勢の変化



(a) ジェスチャの変化のみ



(b) 交渉失敗



(c) 交渉成功

図6 : 実験結果の比較(左下のパラメータがモーションユニットネットワークのアクティブなノードのインデックスを示し、左上のパラメータは感情の強さを表します。)

置き換えて、同様に混合した持続時間や待機時間を用いる。 σ は $L_k > 0$ のとき 1、それ以外の場合は 0 を表すものとする。

させながら振舞うことは、心理状態を表現するために有効であることが確認できた。

7 まとめ

本研究では姿勢, しぐさ, ジェスチャの選択や生成に感情パラメータを用いることで感情を表現するシステムを提案した. 感情パラメータをもとに姿勢や動きの量を決めることで, 感情を表現できることを示した. 心理学の分野では姿勢は表情などに比べると, 感情を表す情報としての役割は低いとされているが, しぐさ, ジェスチャと組み合わせることで, 感情表現がより豊かになることを示した. 今後の課題としては, 最低何分のデータがあれば, 感情的なキャラクタを製作するために十分であるのかを検証していきたい. また, 姿勢, しぐさ, ジェスチャの自動注釈するために[14]らの手法を取り入れることも考えられる.

参考文献

- [1] UNUMA, M., ANJYO, K., TAKEUCHI, R. 1995. Fourier principles for emotion-based human figure animation. In *Proceedings of SIGGRAPH 1995*, 91-96.
- [2] MIWA, H., ITOH, K., MATSUMOTO, M., ZECCA, M., TAKANOBU, H., ROCCELLA, S., CARROZZA, C. M., DARIO, P., TAKANISHI, A. 2004. Effective Emotional Expressions with Emotion Expression Humanoid Robot WE-4RII. *Proceedings of the 2004 IEEE/RSJ Intl. Conference on Intelligent Robot and Systems*, 2203-2208.
- [3] STONE, M., DECARLO, D., OH, I., RODRIGUEZ, C., STERE, A., LEES, A. and BREGLER, C. 2004. Speaking with hands: creating animated conversational characters from recordings of human performance. In *Proceedings of SIGGRAPH 2004*, 506-513.
- [4] CASSELL, J., VILHJÁLMSSON, H., BICKMORE, T. 2001. BEAT: the Behavior Expression Animation Toolkit. In *Proceedings of SIGGRAPH 2001*, 477-486.
- [5] KOPP, S., WACHSMUCH, I. 2004. Synthesizing Multimodal Utterances for Conversational Agents. *The Journal Computer Animation and Virtual Worlds* 15(1), 39-52
- [6] MEHRABIAN, A. 1968. Inference of attitude from the posture, orientation and distance of a communicator. *Journal of Consulting and*

Clinical Psychology 32, 296-308

- [7] MEHRABIAN, A. 1968. Relationship of attitude to seated posture, orientation and distance. *Journal of Personality and Social Psychology* 10, 359-372.
- [8] MEHRABIAN, A., FRIAR, J.T. 1969. Encoding of attitude by a seated communicator via posture and position cues. *Journal of Consulting and Clinical Psychology* 33, 330-336.
- [9] BERLYNE, D. E. 1971. *Aesthetics and psychobiology*, Appleton-Century-Crofts.
- [10] BOND, M. H., SHIRAISHI, D. 1974. The effect of body lean and status of an interviewer on the non-verbal behavior of Japanese interviewees. *International Journal of Psychology* 9, 117-128
- [11] MCNEILL, D. 1992. *Hand and mind*. Chicago: The University of Chicago Press
- [12] BUTTERWORTH, B. & U.HARDAR. 1989. Gesture, speech, and computational stages: A reply to McNeill. *Psychological Review*, 96, 168-174
- [13] ARIKAN, O., FORSYTH, D. A. 2002. Interactive motion generation from examples. In *Proceedings of SIGGRAPH 2002*, 483-490
- [14] ARIKAN, O., FORSYTH, D. A., O'BRIEN, J. F. 2003. Motion Synthesis from Annotations. In *Proceedings of SIGGRAPH 2003*, 402-408.
- [15] ROSE, C., COHEN M. F., BODENHEIMER, B. 1998. Verbs and Adverbs: Multidimensional Motion Interpolation. *IEEE Computer Graphics and Applications*, Vol.18, No.5, 32-40.
- [16] Kovar, L., Gleicher, M., Pighin, F. 2002. Motion graphs. In *Proceedings of SIGGRAPH 2002*, 473-482.
- [17] EKMAN, P., FRIESEN, W. V. 1975. *Unmasking the Face*. Prentice-Hall