

心理コミュニケーションゲームのための複合会話動作生成

中野 敦, 塩入 健太, 星野 准一
筑波大学大学院システム情報工学研究科

{nakano, shioiri}@graphic.esys.tsukuba.ac.jp, jhoshino@edu.esys.tsukuba.ac.jp

近年, CG キャラクタとの対話によってストーリーの進行するコンテンツが数多く見られるようになった. そのため CG キャラクタの表現力を向上することは, これらのコンテンツに没入するための重要な要素となっている. 従来の対話型コンテンツでは, CG キャラクタは台詞に対応付けられた振る舞いを行っており, 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た. そのため内面に存在するはずの心理状態を感じられない無機質な印象を利用者に与えてしまうという問題がある. そこで本研究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やしぐさを動的に複合することで心理状態を表現した振る舞いを生成する. この技術によって表現豊かな振る舞いを生成した結果を心理コミュニケーションゲームという対話型コンテンツ上で示す.

[キーワード] 心理状態, コミュニケーションゲーム, 会話動作

Generating Composite Motion in Conversation for Mental Communication Game

Atsushi Nakano, Kenta Shioiri, Junichi Hoshino

University of Tsukuba, Graduate School of Systems and Information Engineering

Mental communication with characters is important for many entertainment and edutainment applications. Human's mental statuses are complex, and mental behaviors typically consist of multiple components such as poses and many unconscious movements. In this paper, we propose the layered behavior synthesis technique for integrating mental behaviors and conversational gestures. We compose various mental behavior representations by considering the spatial consistency and co-occurrence probability. We also show an example of a game using the mental behavior synthesis technique.

[keyword] mental status, communication game, conversational motion

1 はじめに

近年, CG キャラクタとの対話によってストーリーの進行するコンテンツが数多く見られる. CG キャラクタの表現力を向上することは, これらのコンテンツに没入するための重要な要素である. 本稿では姿勢としぐさという2つの行動が持つ表現力に着目する. 姿勢は, その語が表すように体の構えであると同時に心の構えでもある. 例えば, 姿勢を正すという言葉は, 単に背筋を伸ばして正しい姿勢を保つという意味だけでなく, 心理状態をも整えることを意味している. しぐさは本人の内面から自然に生じる動作であり, 心理状態が作為無くその

まま相手に伝わるという特徴がある.

これまでにも CG キャラクタ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, Stone[1]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キャラクタのジェスチャを割り当てるシステム

として BEAT[2]がある。これらの手法では複数のモーショ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 Kopp[3]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、これは視聴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った問題がある。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。

そこで本研究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。

2章でジェスチャ、姿勢、しぐさの定義とそれらを組み合わせた複合動作を生成するためのモーションユニットの内容に触れる。3章では、複合動作を会話文や心理状態に合わせて生成するためのプランニング手法について述べ、4章で手法の有効性を評価するために用意したコミュニケーションゲームについて述べる。

2 動作の分類とモーションユニット

2.1 姿勢、しぐさ、ジェスチャによる動作の分類

“姿勢”や“しぐさ”は、社会心理学の身体動作分類において情動表出と呼ばれるものに分類され、情緒的な

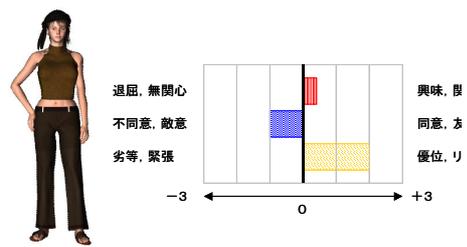


図1 姿勢の表す心理的な印象の定量化

状態や反応を示す表情や身振りのことを指す。発話の有無に関わらずその人の心理状態が無意識に現れるため、姿勢やしぐさの心理状態を動作生成に取り入れることは人間味のある心理状態表現を実現するために有効な手段であると考えられる。本稿では“姿勢”、“しぐさ”、“ジェスチャ”の3つに動作を分類し、発話に伴うジェスチャに、心理状態を表現する姿勢、しぐさを合成することで心理状態を表現した会話動作を生成する。

2.2 心理状態パラメータ

姿勢やしぐさ、ジェスチャの与える印象の定量化として心理状態パラメータを設定する。本稿では、心理状態パラメータとして“興味度”、“同意度”、“優位度”の3つのパラメータを設定する。これらのパラメータはPブル[4]によって姿勢から伝達される心理状態、姿勢に表れる心理状態として考えられている。しぐさも興味や緊張の度合いによって変化するため同様のパラメータで印象を表すこととした。

“興味度”は対象や会話内容に興味を持っているとき、関心がある状態では正の値をとり、退屈なとき、関心がない状態では負の値をとる。“同意度”は会話内容に同意しているとき、対象に友好的な心理状態のときに正の値をとる。反対に不同意なとき、敵対心を持っているとき負の値をとる。“優位度”は対象に対して優位であるとき、リラックスした心理状態であるときに正の値をとり、劣等を感じている状態、緊張している状態で負の値をとる。本稿では、3項目とも-3から+3の範囲で推移するものとした(図1)。

2.3 会話文と心理状態変化のための動作計画

“姿勢”、“しぐさ”、“ジェスチャを合成した振る舞いを本稿では複合動作と呼称する。この複合動作を会話文や心理状態の変化に合わせて自動生成するために動作計画を行う。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ションクリップに、心理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含め、モーションユニットとして定義す

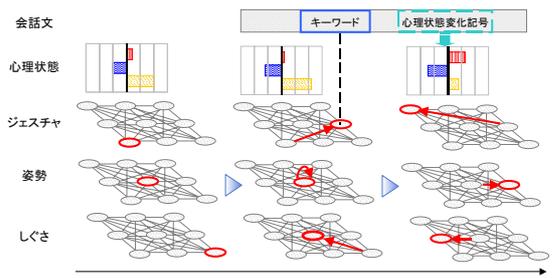


図2 複合動作のプランニング

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。各属性のネットワークの現状態であ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2）。モーションユニットネットワークの構築と動作計画の詳細については次章で述べる。

2.4 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心理状態パラメータ，ジェスチャに対応した語句，動作の継続と待機期間，拘束状態を定義する。

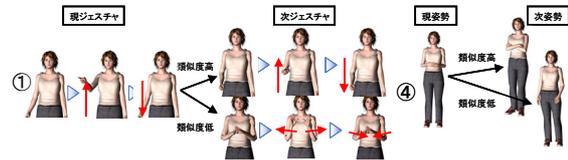
①[心理状態パラメータ]

各モーションユニットに心理状態パラメータを設定する。2.2節で述べたようにパラメータには対話者への興味を表す興味度，賛同の意を表す同意度，優越感を表す優位度を設定する。

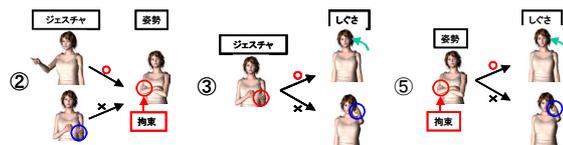
②[発話に伴うジェスチャ]

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[5]がある。また，表現したい空間や運動的なイメージを言語化し易いように，加工した心的表現に基づき，発話に伴うジェスチャは生成されるとするインタフェース仮説[6]がある。これらの仮説をもとに，会話文中の単語とジェスチャが関連付けられるものと

	ジェスチャ	姿勢	しぐさ
ジェスチャ	①使用する身体部位の類似度及びそれらの位置や軌道の類似度	②ジェスチャの主成分と姿勢の拘束の比較	③ジェスチャの主成分としぐさの主成分の比較
姿勢		④姿勢の類似度	⑤姿勢の拘束としぐさの主成分の比較



(a) 同モーションユニットネットワーク内の遷移



(b) 異なるモーションユニットネットワークの協調

図3 モーションユニットの静的な相関性

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。

③[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起した後にモーションユニットの状態が持続する時間幅とその後一定時間生起させずに待機する時間幅を定義し，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。

④[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の意味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て，モーションユニット間の協調性を考慮した動作計画を行う（図3(b)）。

3 階層化プランニング

キャラクタの発話や心理パラメータを入力としてジェスチャ，姿勢，しぐさの順にモーションユニットを選

択し、それらを合成することで動的に複合動作を生成する。本稿では、グラフィカルモデルを階層的に用いて動作計画モデルを構成する。グラフィカルモデルは変数間の依存関係を、変数をノード、関係をアークとするグラフによって記述する確率変動モデルである。はじめに動作のベースとなる姿勢ノードを選択する。次に選択された姿勢ノードとの依存関係をもとに発話に伴うジェスチャを選択する。最後に選択された姿勢ノード及びジェスチャとの依存関係をもとにしぐさを選択する。

3.1 モーションユニットネットワークの重み生成

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じる可能性を表す評価値を連続性値 ω^c (図3(a))、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 (図3(b))。これらは、システムの初期化の際に定めるため静的なパラメータとして用いる。また2.4節で述べ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。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 ϕ は次のような乗算の形式で定める。

$$\phi = \omega^c \omega^l \phi^{time} \phi^{state} \quad (1)$$

この ϕ の大きさに依存したモーションユニットの生起確率をソフトマックス法によって計算し、確率的に選択する。

3.2 複合動作のプランと実行

会話文や心理状態の変化によってリアルタイムで複合会話動作を生成するためにコマンドキューイング[7]と呼ばれる機能を実装した。これはキャラクターの未来の行動をスタックしておき、順次実行できるようにする機能である。会話文によっては一度の複合モーションユニットだけで表現できない場合があるため、このように複数のモーションユニットをスタックすることで会話に沿った一連の行動として表現する。またこの手法は既にスタックされたモーションユニット列の中に、新規のモ



図4 心理コミュニケーションゲーム

ーションユニットを挿入することで行動順序の変更が可能のため、本稿のような自律行動キャラクターの実現用途に適している。

4 心理コミュニケーションゲーム

本手法の有効性を確認するために図4に示すような評価用コンテンツを作成した。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。

実験のために、姿勢属性のモーションユニットを15種類、ジェスチャ属性のモーションユニットを7種類、しぐさ属性のモーションユニットを18種類用意した。姿勢の印象を表す心理状態パラメータは、前傾姿勢の場合は、後掲姿勢よりは、肯定的であると知覚されることや、腕組みや脚組みをしていない姿勢が肯定的な意味を持つとされる体の開放性についての文献[8][9][10]を基に定めた。しぐさは腕部、頭部、胴体に分類し、頭部や胴体は揺れの速度や振幅の大きさが心理状態パラメータの変化によって異なるデータを用意した。頭部の動きは興味や同意度が高くなるほど相手を注視し、反対に低くなるほど視線を逸らすものとした。胴体の揺れは優位度が高くなるほど大きくなり、反対に低くなるほど小さくなるものとした。CGキャラクターの持つ興味度、同意度は入力の無い場合は時間と共に下降し、優位度は時間

と共に上昇するものとした。

また台詞などによる心理状態パラメータの変化は心理状態パラメータ（興味度，同意度，優位度）の上昇，下降を示す△▽，▲▼，↑↓などの心理変化記号を会話文中に含めて，これらを解釈することで行った。例えば，「今日も疲れたなあ▽」，「やったあ！▲」，「ふうっ↑眠いなあ▽」というように使用する。記号は各パラメータを1上昇もしくは1減少することとした。また△2は△が2つ入力された場合に等しい，↓3は↓を3つ入力した場合に等しいというように心理変化記号の後に数字を付けた場合，倍数を表すものとした。

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）。交渉成功例の心理状態変化，失敗例の心理状態変化をそれぞれ図5に示している。この例ではプレイヤーは2度選択肢を選ぶことができる。成功例の場合，初めの選択肢では△3▲2，2つ目の選択肢では△3▲2という記号が入力されている。失敗例では，初めの選択肢で▽3▼3↑，2つめの選択肢では，▽2▼2↑が入力されている。

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キャラクター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，しぐさを変化させた例では，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。

5 まとめ

本稿では，発話に伴うジェスチャに加えて，心理状態を表現する姿勢やしぐさを複合することで心理状態を表現した振る舞いを生成する手法を提案した。また本手法を用いた心理コミュニケーションゲームにより有効性を確認した。

本稿では，ジェスチャ，姿勢，しぐさのモーションを合成する際に，腕部や脚部が他の身体部位に重なってしまうという問題が残されており，今後はそのようなモーション合成の問題についても検討していきたい。

また今後の研究の方向性として，自律行動するCGキャラクターだけでなく，オンラインゲームでプレイヤーが操作するCGキャラクターに適用することを考えている。

参考文献

- [1] Stone, M., DeCarlo, D., Oh, I., Rodriguez, C., Stere, A., Lees, A. and Bregler, C.: Speaking with hands: creating animated conversational characters from recordings of human performance, *ACM Transactions on Graphics* 23(3), pp. 506-513 (2004).
- [2] Cassell, J., Vilhjálmsón, H., Bickmore, T.: BEAT: the Behavior Expression Animation Toolkit, *Proceedings of SIGGRAPH '01*, pp. 477-486 (2001).
- [3] Kopp, S., Wachsmuth, I.: Synthesizing Multimodal Utterances for Conversational Agents, *The Journal Computer Animation and Virtual Worlds* 15(1): pp.39-52 (2004).
- [4] P. ブウル 著, 市河淳章 高橋超 編訳, 飯塚雄一 大坊郁夫 訳, “姿勢としぐさの心理学”, 北大路書房, 2001.
- [5] Butterworth, B. & U.Hadar: Gesture, speech, and computational stages: A reply to McNeill. *Psychological Review*, 96, pp.168-174 (1989).
- [6] Kita, S. & A.Ozyurek, Semantic coordination between speech and gesture cross-linguistically: Evidence for the interface representation of spatial thinking and speaking. *Journal of Memory and Language*, under review.
- [7] Norman, Donald, “The Design of Everyday Things”, Currency/Doubleday
- [8] Mehrabian, A., Inference of attitude from the posture, orientation and distance of a communicator, *Journal of Consulting and Clinical Psychology* 32, pp. 296-308, 1968
- [9] Mehrabian, A., Relationship of attitude to seated posture, orientation and distance, *Journal of Personality and Social Psychology* 10, pp. 359-372, 1968
- [10] Mehrabian, A., Friar, J.T., Encoding of attitude by a seated communicator via posture and position cues, *Journal of Consulting and Clinical Psychology* 33, pp. 330-336, 1969

