

審査の結果の要旨

氏名 ブライトフス ウェーナー

本論文は「Automatic Multimodal Behavior Generation for Embodied Virtual Characters based on Text (仮想キャラクタのためのテキストからの自動マルチモーダル行動生成)」と題し英文で記されており、6章から成る。

第1章「Introduction (序論)」では、顔と身体を有すマルチモーダルな仮想キャラクタエージェントは、コンピュータと人間間の自然で親しみのあるインタフェースの形態として重要になってきているが、発話テキストのデータに加えて身体動作の指示も与えなければならず大変であるという課題に対し、発話テキストから視線とジェスチャ動作を自動生成することにより、課題への対処を目指すという、本研究の目的を述べている。

第2章は「Related Research (関連研究)」であり、仮想エージェントのジェスチャの種別、エージェントの視線制御と、会話する仮想エージェントシステムのこれまでの研究例についてまとめている。ジェスチャの種別は、ビート動作(手をリズムカルに上下に振る)、適応動作(adapter: 会話者が顎やこめかみに触れるなど)、表象動作(emblem: 固有の意味を表現)、象徴動作(icon: 四角や丸などのように語彙の物理的な性質を表現)、隠喩動作(metaphor: 話者の考えを隠喩的に示す)、指差し動作の6種であるとし、本研究ではこれらの生成について扱うとしている。既存のシステムとの比較として、多くの既存システムは単体のエージェントの発話に伴う動作を扱っているが、本研究では複数エージェントでの会話状況を扱うとしている。また、発話テキストに現れる語彙に基づいてルールにより動作生成を行っているシステムも見られるが、本研究は会話における各エージェントの役割を考慮、複数ルールの相互作用を考慮してジェスチャ生成過程での適応的、動的な処理機能を導入している点が新規の特徴であるとしている。

第3章は「Automatic Behavior Generation System (自動行動生成システム)」であり、このシステムは、発話テキスト入力を処理する言語モジュール、視線とジェスチャ付与を行う行動生成モジュール、各種のアニメーションエンジンで表示し得る形式でエージェント行動情報を編成する出力モジュールとから成る。言語モジュールでは発話テキスト文の構文解析、品詞タグ付け、位置と方向に関係する用語の抽出、文の主題部(theme)と題述部(rheme)の同定等の処理を行う。

行動生成モジュールはエージェントの3種の役割、即ち、発話者、聞き手、その他の聴衆を考慮して適応的、動的にジェスチャを生成する。視線パターン生成に関しては、役割に基づき、24の決定ルールを備えている。ビートと指差し動作を別にして、ジェスチャは基本的に出現語彙に関連付けて生成されるが、82ジェスチャに対応する関連語彙は152の語彙集合にまとめられた計2012語から成る語彙辞書(部分を付録に記載している)を構成しており、これは存在するジェスチャ生成英語語彙辞書として最大規模のものである。ジェスチャの大きさ等は形容詞、副詞、否定形に対応して修飾される。ビート動作は初出の語彙に反応して生成され、これは聴衆の注意を喚起することになる。指差し動作に関しては、エージェントが存在する環境についての情報をシステムが持ち、テキスト中の物体名や指示代名詞と環境中の物体との対応をとり、指差し方向や物体との距離を考慮して指差し法(正確に指差すや曖昧な範囲を指し示すなど)を決定している。指差し動作生成に関しては、3次元空間で参照物体を定めて指差し

法を決めるアルゴリズムを備えた最初のエージェントシステムとなっている。

このモジュールでのジェスチャ生成が適応的、動的であるとは、ある語彙、ルールに基づいて生成候補となったジェスチャが、文に渡る相互の整合性と生成候補の優先度を考慮した一種のメタルールによる調整処理を経て決定されることを意味している。優先度に関しては、より個別的な意味を担うジェスチャが優先され、この観点から低い順からビート、適応動作、象徴動作 (icon)、隠喩動作 (metaphor)、表象動作 (emblem) となっている。このような動的処理により、従来より適切なジェスチャ生成が可能になるとしている。聞き手エージェントの動作生成に関しては、発話開始時に発話エージェントに視線を向け、題述部 (rheme) の主要フレーズに応じて頷きを行うなどである。82 種のジェスチャ生成を行うルールは総数で 200 以上となっており、既存システムを超える包括性を有している。

出力モジュールは視線やジェスチャ動作情報を付加した発話テキストデータを代表的仮想空間であるセカンドライフ上のエージェントを制御するエージェント行動記述言語といったような、特定の出力データ形式に編成する。この時、複数エージェントの空間的位置関係を考慮して視線や動作の具体的な方向が決められ、またエージェントのアニメーションエンジンで可能な所与のジェスチャへのマッピングも行われる。

第 4 章「Experiments and User Studies (実験とユーザスタディ)」では、作成したシステムについて 20 名ほどの被験者ユーザによる評価結果を示している。実験は、病院における医療専門家の 2 エージェントの会話場面と、セカンドライフ上の仮想家具ショップでの店員による顧客への商品説明場面を用いて行っている。前者では特に本システムによる視線の動きが効果的に生成されていること、後者では特に仮想 3 次元空間での指差し動作が効果的に生成されていることを示している。

第 5 章「System Structure (システム構成)」では、第 5 章に記しているモジュールから成るシステム構成を具体的に示し、別のコンピュータ上へシステム実装する方法を記している。

第 6 章「Conclusion (結論)」では、本論文の研究成果をまとめ、今後の可能性について言及している。

以上のように本論文は、仮想キャラクターエージェントの視線を含む動作をその発話テキストから自動的に生成する機構並びにシステムに関するものである。3 種の役割をもつ複数エージェントの会話を対象とし、視線の動きと 6 種別のジェスチャ動作生成を行うために、従来システムを超える規模のジェスチャ語彙辞書と動作生成ルールを有する包括性の高い機能を構成し、生成ルールによって得られる動作候補を文に渡る整合性と優先度によって適応的・動的に調整して動作を決定する機構や、3 次元仮想空間で参照物体を定め適切な指差し動作を生成する機構といったような新機能を開発し、それらの効果を被験者ユーザによる評価で確認している。これは仮想キャラクターエージェントによるマルチモーダル・メディア分野に貢献する研究成果と認められ、情報理工学における創造的実践の観点でも価値が認められる。

よって本論文は博士 (情報理工学) の学位請求論文として合格と認められる。