

報告番号	※甲	第	号
------	----	---	---

主 論 文 の 要 旨

論文題目 擬人的媒体の音声・身体動作によるマルチモーダル表現の表情強度に関する研究

氏 名 米澤 朋子

論 文 内 容 の 要 旨

人がコミュニケーションをする際には様々なモダリティを用いて表現している。日常のコミュニケーションでは、例えば、声の調子や、顔の表情の変化、意識的または無意識的な身体動作によって、細かい表現を行っている。表現を受け止める側は、そのような細かな表現から、無意識に意図や感情を統合して捉える。

現在、擬人的媒体の研究開発が進み、関連したシステムや商品も普及しつつある。これは、人が動物や無生物などの人ではないものをあたかも人のように錯覚したり扱ったりすることを利用したものである。このような研究やシステムの中には、仮想エージェントによる情報提示や、ロボットによるコミュニケーション補助・代替システムなど、擬人化による錯覚を利用した親和性の高いインタラクションを狙ったものがある。その際には、実際の間人や生物などと同様な形態や、内部機構、そして表現のそれぞれについて高い擬人性が望まれる。しかし現状では、それぞれの表現について部分的に議論されているに留まり、統合された擬人的表現に関する研究はほとんどない。ペットロボットやヒューマノイドロボット、アンドロイドといった、実体を伴う擬人的媒体において、ペットのよな親しみやすい形状やリアルな外観による擬人化だけではなく、統合的な観点からそれらのマルチモーダル表現を検討し適切に導入することが必要である。

本論文では、実体を伴う擬人的媒体における、音声と身体動作を通じたマルチモーダルな表現を検討し、知覚実験の分析により、表現における表情付けの強度（表情強度）に関する制御の有効性を議論する。擬人的媒体は、身体を持ち、その身体に基づき、人間と同様に音声、身体動作、顔の表情といった様々な人間の表現チャンネルを持つ。その中で、一般的な擬人的媒体の構造における制御を前提として、とくに音声と身体動作に焦点を当てる。各々の表現における自然性だけでは、必ずしもそれらを統合して構成されたマルチモーダル表現も自然になるとは限らない。

そこで、各々の表現手法を検討すると同時に、それらの組合せにおける効果（クロスモダリティ）についても検討する。この研究では、擬人的表現におけるタイミング、種類、強度という様々な表現の要素の中でも、とくに自然な表現への影響の大きい表情強度の制御に焦点を当てる。表情強度は、人間同士のコミュニケーション中では時々刻々と変化しながら伝えられる。このような人間の表現を模倣して擬人的表現を構成するためには、マルチモーダル表現および個々の構成要素を、望まれる強度で連続的に生成することが求められる。このため本論文では、まず擬人的媒体それ自身の検証を行い、次に擬人的表現における様々な表情強度を制御する手法を提案し、最後にその知覚的效果を検証する。そのため、以下の章を構成する。

第一章では、擬人的媒体におけるマルチモーダル表現に関する研究の背景と問題を提起し、本論文の目的と位置づけを明確化するとともに各章の位置づけを示す。

第二章では、擬人的媒体やマルチモーダル表現についての各々の関連研究を紹介しながら、本論文の位置づけと方向性を確認する。

第三章では、本論文で用いる擬人的媒体としてぬいぐるみに着目し、ぬいぐるみが介在するコミュニケーションにおける人の言動への影響を評価する。その親しまれやすい特性について実体性と外観という要素に着目し、人のコミュニケーション行動を観察することで、各要素がコミュニケーションに及ぼす影響を計測する。実験では、ぬいぐるみロボットによる動きと音声を通じた非対面コミュニケーションを設定し、i) 実存するぬいぐるみを用いた場合と二次元画面上に現れるぬいぐるみを用いた場合、ii) ぬいぐるみを用いた場合と内部機構がむき出しのロボットを用いた場合について、ユーザの発話および様々な行動をそれぞれ対比較する。その結果、ぬいぐるみの実体性と外観が、それぞれユーザの発話と動作に有効的に作用することが確認された。

第四章では、音声の表情強度の変化を滑らかに接続しながら表出する手法を検討しその有効性を評価する。まず、音声の中でもとくに声色における表情強度に着目し、表情のない平坦な歌声を始点とし表情付き歌声を終点とした音声モーフィングを適用して様々な強度の表情付き音声を自由に生成する手法を提案する。そして、この手法により作成された歌声音声に対する被験者実験の統計的分析により、i) 変化させるモーフィング率に応じて歌声の表情強度が補間されて知覚されたこと、およびii) 様々なモーフィング率においてそれぞれ差分の異なる刺激対を用いた弁別実験により、表情強度はカテゴリー知覚ではないことが確認されたことをそれぞれ示し、知覚的な連続性を実験的に示す。これに基づいて、iii) 知覚される表情強度を近似し、その逆関数を用いれば知覚される表情強度の変化が線形になるよう表情付けを調整できることを、実験結果により示唆する。最後に、自然性についての被験者実験を分析しiv) 表情強度の変化を補間することにより自然性の評価が高まることが示され、提案手法の有効性が確認された。

第五章では、自然に表情付けられた歌声表現を生成するための擬人化インターフェースとして、手人形を導入したシステム“HandySinger”について述べる。このシステムは、手人形インターフェースから得られるユーザの手のひらの動作に従い、連続的に歌声の表情強度を変化させる。手人形の身体動作としてユーザの手の形状を逐次認識するために、手人形インターフェースは7つのピエゾ式曲げセンサと2つの圧力センサを固定した手袋を内包する。提案システムの動作を確認するため、入力センサ値と表情強度出力を記録し、対応付けられた身体動作の表情付けの種類と強度に応じた歌声の制御が可能であることを確かめた。また、被験者実験を行い、センサ値の解像度や変化幅に個人差があったとしても、システム動作には十分な精度であることを確認した。上記のシステム評価によって、これらが機能して、擬人化された身体動作の入力により歌声の自然な表情付けを行うシステムが構築され、マルチモーダル表現の実験基盤が整った。

第六章では、コミュニケーションや情操的用途での実体を伴う擬人的媒体における、より豊かで自然な表現の可能性を追求するため、クロスモダリティ（異なるモダリティ間の関係性）を考慮した統合的な擬人的表現について検討する。歌声と身体動作のクロスモーダル表現について、知覚的側面から表情強度の影響や関係性を明らかにするため、異なるモダリティ間の1. 総合的な表情強度と2. 適合性について被験者実験を行った。その結果、統合的な表情強度の知覚は、身体動作による視覚的な表現の影響が圧倒的に強く、クロスモダリティの適合性の知覚は、モダリティ間で相当する組合せにおいてもっとも高いことが分かった。実験の分析より、i) 歌声の表情強度の変化幅を広げ、ii) 身体動作による表現との間の知覚的影響のバランスをとり、より自然な表現を実現させることを、擬人的媒体におけるクロスモーダル表現の設計指針として提案する。

第七章では本研究で得られた結果を総括する。擬人的媒体による自然なマルチモーダル表現を生成する手段の一つとして、被験者による知覚実験の統計的分析により導出した結論として、音声と身体動作とそれらの組み合わせの表情強度を制御する提案手法が有効であることを述べ、今後の課題と研究方向を展望する。

かつてより、擬人的媒体は情報提示などの目的志向で心理的負荷を減らすために用いられてきたが、現在では、実体性を伴うことを利用して、パートナーやペットなど生きている存在の代理の役割も務めるように発達しつつある。アンドロイドロボットや、動物型ロボットなど、目的に応じて求められる擬人性は異なるが、自然性が追求された高度な表現が必要とされていることには変わりはない。人に知覚される表情強度を考慮してマルチモーダル表現を構成することは、いずれの擬人的表現においても重要である。よって本論文で得られた知見は、擬人的媒体の表現を構成する上で非常に有益なものとして役立つことが期待される。