

機関番号：94301

研究種目：若手研究(B)

研究期間：2008～2010

課題番号：20700106

研究課題名(和文) 実体型擬人的媒体のマルチモーダル表現への個人化表情付与に関する研究
研究課題名(英文) Individualization of Multimodal Expression for Anthropomorphic and Embodied Media

研究代表者

米澤 朋子 (YONEZAWA TOMOKO)

株式会社国際電気通信基礎技術研究所・知能ロボティクス研究所・研究員

研究者番号：90395161

研究成果の概要(和文)：

本研究課題では、ロボットやエージェントなどの擬人的媒体において、個性を付与した様々な表現を実現することで、与えられた擬人的媒体のキャラクタに左右されない、本質的な「誰かからの」メッセージの「媒体」になりうるように変貌させることを目的とした。マルチモーダルフォントとして、日常挨拶表現を対象とし、様々な年齢層・性差を考慮した 30 名以上のデータを取得し、声の特徴だけでなく、動きの特徴、声色の表現幅といった個人表現を同時取得し分析できた。結果としてある程度の表現傾向を分類できることと、擬人的媒体などに載せる表現の単純化に前期分類を活かせることなど多くの知見を得た。

研究成果の概要(英文)：

Robots and virtual agents have each own (given) character for each product using their original multimodal data set of voices and motions. However, their given characters sometimes interrupt various expressions for communicative purposes. In this research theme, we have proposed a multimodal expression capturing method for individualization of the anthropomorphic embodied media.

Multimodal-FONT is the name of the database which this project has captured. As expressed in the name, each individual data of the multimodal expressions for daily greeting was captured with varied age groups and genders. Finally we could gather over 30 persons' multimodal expressions: voice, head and upper-body gestures, and directions of the gaze at the same time. The recorded data are analyzed using the prosody of the voice, motion characters, and gaze patterns, and variation widths of the whole expressions. As results, we could confirm that there are some kinds of expression patterns relevant or irrelevant to their gender groups or age groups.

交付決定額

(金額単位：円)

| | 直接経費 | 間接経費 | 合計 |
|--------|-----------|---------|-----------|
| 2008年度 | 1,500,000 | 450,000 | 1,950,000 |
| 2009年度 | 1,100,000 | 330,000 | 1,430,000 |
| 2010年度 | 700,000 | 210,000 | 910,000 |
| 年度 | | | |
| 年度 | | | |
| 総計 | 3,300,000 | 990,000 | 4,290,000 |

研究分野： 擬人化メディア, コミュニケーション科学, 音声

科研費の分科・細目： メディア情報学・データベース

キーワード： 擬人的媒体の表現, 日常挨拶マルチモーダル表現, 性差, 年齢差, 個人差, 個人内表現差, 個人間表現差, 表現のモーフィング

1. 研究開始当初の背景

ヒューマノイドロボットやペットロボットなど、ロボットの利用方法や波及効果の研究が行われる一方、各ロボットの持つキャラクターを個性化したままの使用方法が目立ち、動力を持つコミュニケーションメディアとしてのロボットに必要な表現プリミティブを十分に揃えないままの擬人的媒体が多い。仮想エージェントやロボットの研究や製品化が進む中、様々なキャラクター性をもった製品が世の中に広まりつつあり、それぞれ個性的なキャラクター(個人性)を持っており、を変更することは通常では必要ではない。

これに対し、単一のロボットに複数の個性表現を実現させることで、コミュニケーションの補助をしたりメッセージを伝達するなど、ロボット自身の個性ではない伝達内容自体に持たせるべき個性を付与することが必要だと考えた。例えば RobotPhone に見られるように、擬人的媒体をコミュニケーションメディアとして捉え利用しようとする研究もある。このような直接的な音声通信が伴うものでは、ロボットの動きが多少自然でなくても、音声の個性が「相手が誰であるか」を十分に認識させることが可能である。その一方で、せっかく動きを伝えてもそれは共通の記号としてのジェスチャに過ぎず、その人らしい動きやマルチモーダルな表現および、全体的なモダリティの相互作用(クロスモダリティ)に関しては反映されることはない。

音声合成データベースにおいて、表現の個人性を可能にするデータベースとして、Voicefont がある。音声のみの表現においても、生の音声データベースを用いる率が高いほうが表現の自然性は高い。また、合成音の調整手法には多大なデータをもとにした音素などのパラメータモデルが存在する。いずれにせよ、個人性(その人らしさ)を表現する上では、まずその人のデータを取得することが必須であるため、生のデータベースが存在するならば、それを極力有効活用するのが望ましい。

2. 研究の目的

本研究課題で提案したマルチモーダルフォントは、音声研究における voicefont の考え方を、複数モダリティに適用し、マルチモーダルデータベースの同時取得することにより、各モダリティの個人性のみならず、モダリティ間の相互作用(クロスモダリティ)が取得され表現に反映されることが期待できる。

(1) 異なる表現の連続性の実現

コーパススペースデータの再利用を検討する際、考慮すべき事項の中で多くを占めるのが連続性である。これは、二つのタイプの連続性の検討が考えられる。一つは、なるべく生のデータを用いるための接続部およびその近辺で考慮される連続性と、もう一つは、異なるタイプの表現を持つコーパス間での表現間変化における連続性である。

動作の連続性においては画像の連続性と同様に、ある同一パターンの動作であれば座標パラメータを連続的に変化させる手法がとれるが、音声の連続性には未だ問題点が多い。音声における連続性の検討についてを明らかにし、マルチモーダルデータの個人内変化を実現することを目的とした。

(2) 個性のあるマルチモーダルデータの取得

できるだけ生のマルチモーダルデータを同時に取得し、コーパスとすることを目的とした。これは音声コーパスでも言えることだが、できるだけ加工されていないデータを用いることで、自然性が高まり、表現性自体の議論を減らすことができる。

① 音声データの取得

② 動作データ(モーション)の取得

③ 視線データの取得

の3つを同時取得する枠組みを準備し、ロボットに反映できるデータセットとして活用することを目的とした。

視線表現は人間が意図的に行わない場合が多く、必ずしも個性が出るとは限らないが、その一方で、欧米人は日本人より相手の目をじっと見ながら話すといった文化の差が語られることは多い。このような視線表現における意識的・無意識的な視線表現も一つのマルチモーダルデータとして取り入れ、取得することとした。

(3) マルチモーダルデータの個性分析と利用

マルチモーダルデータはそのままの形式では当然活用には至らない。それ以前に、今後新規に個人性付与データを取得する際に、効率化を考慮しつつ、どの程度の生のデータを取得すべきかを検討するには、ある程度の「個性の傾向」を知る必要がある。

よってマルチモーダルデータのモダリティ間での関連性、性差や年齢差といった様々な個性要素、および個人としての表現要素の現れ方を分析しておくことが必要だと考えた。

3. 研究の方法

(1) 異なる表現の連続性に関する検討

連続性を検討するときに用いられるのは知覚試験、特に弁別実験が多く用いられる。

弁別実験では、どちらが表現 A に近いか表現 B に近いかが、の二者択一で被験者にこたえを求める方式である。弁別実験により明らかになるのは、人間の知覚においてどの程度の連続性が求められるか、また、弁別精度の曲線により、連続性が直線的か補正関数を必要とするものが確認されるものである。段階的印象評定だけでは得られない連続性知覚に関する結果が出る可能性を考慮して本手法を導入した。

(2) 個性付マルチモーダルデータの取得

個人性を表現するには、個人間で異なる表現を取得することが必要である。一方、個人間差の中に含まれる個人内表現幅として、個人内表現の多様化を測る際に、異なる種類の表現を取得し連続的な表現強度を調整可能にすることが重要である。

日常挨拶フレーズの複数表現を取得するため、会話相手として想定する対象を、

- ・ 同年代のごく親しい友人
- ・ 同年代の苦手な知人
- ・ 自分より年配の目上の人

の3種類として、それぞれの固定の人物をあらかじめ決定してもらった後、マルチモーダルデータの取得を毎回同様に行った。

以下の各モダリティのデータを同時取得するため、図1のようにモーションキャプチャカメラや視線用 USB カメラを設置した。各モダリティデータの取得に関しては下記のとおりである。

① 音声データの取得

個人性付きマルチモーダル表現合成に音声の特徴も用いることを考慮し、防音環境の部屋で実施した。音声取得用には、首後部で調整しと耳に付ける固定式ヘッドウォンマイク (AKG C 555 L) を参加者に装着し、ワイヤレスマイクセット (AKG WMS40PRO, 受信部: SR40, 送信部 PT40) により、動きながら発声することに苦痛がないよう、モーションキャプチャと同時に音声収録を行った。

② 動作データ (モーション) の取得

モーションキャプチャには、Natural point 社の Optitrack シリーズの SLIM:V100 カメラを 6 台構成し、ユーザの身体特徴ポイントを追跡することとした。なお、3次元キャプチャソフトウェアには、Natural point 社の TrackingTool(TM) を用いた。立位における胴体のブレを腰地点で固定する目的と、防音スペースの状況による制限により、背もたれのないパイプ椅子に座った状態での上半身の動きを記録した。

③ 視線データの取得

視線検出用 USB カメラに目が映りやすいよう、頭部のモーションキャプチャ用のマーカ付帽子のつばは、あらかじめ切り落としてある。画像処理による視線検出技術を用いて視線角度を検出し、エラー部分を手動修正することとした。

参加者の年齢層と性別ごとの人数を表1に示す。また、日常挨拶データの一部を表2に示す。

表1 世代差と性差を考慮したデータ取得

| 参加者 | 若年層 | 中間層 | 高齢層 |
|-----|---------|---------|---------|
| 年齢幅 | (19~27) | (40~50) | (65~74) |
| 男性 | 7名 | 5名 | 5名 |
| 女性 | 8名 | 6名 | 5名 |

表2 収録した日常表現フレーズの一部抜粋

| | 同年代向け用 | 目上向け用 |
|----|-----------------------------------------|-----------------------------------------------------|
| 挨拶 | こんにちは おはよう バイバイ | こんにちは おはようございます さようなら |
| 相槌 | うん. うん. そう. うそ!! 本当?? わかんない | はい. いいえ. そうですか. うそでしょう!! 本当ですか?? わかりません |
| 依頼 | お願い ごめんね | お願いします ごめんなさい |

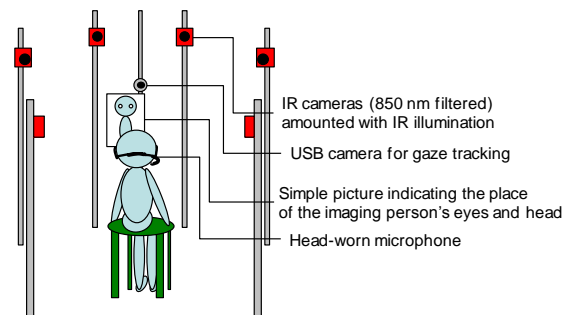


図1 データ取得用環境

(3) マルチモーダルデータからの個性分析

取得されたデータから、個人間表現差および個人内表現差を分析した。これは、傾向を把握するとともにマルチモーダルデータの単純利用のみならず、他者データ間での効率的利用を図るものである。

4. 研究成果

以下の課題項目別に結果をまとめる。

(1) 異なる表現の連続性に関する検討結果

まずは音声の知覚的検討の過去の結果からさらに連続性を自然にするための「表情付与」

に関する非線形性を検証した。その結果、音声表現の知覚においては線形性は弱く、表現の変化が表現Aと表現Bの midpoint に近づくにつれて表現の変化を感じやすいことがわかった。この結果は音声においてのみでなく、何らかの連続性を持つ表現の変化における表出手法に適用できるものであると考えられる。

(2) 個性付マルチモーダルデータの取得結果

次に、図2に示すように日常挨拶表現ごとのマルチモーダルデータ取得に成功した。また、想定する相手の違いによる異なる表現のもの、被験者によっては顕著であった。

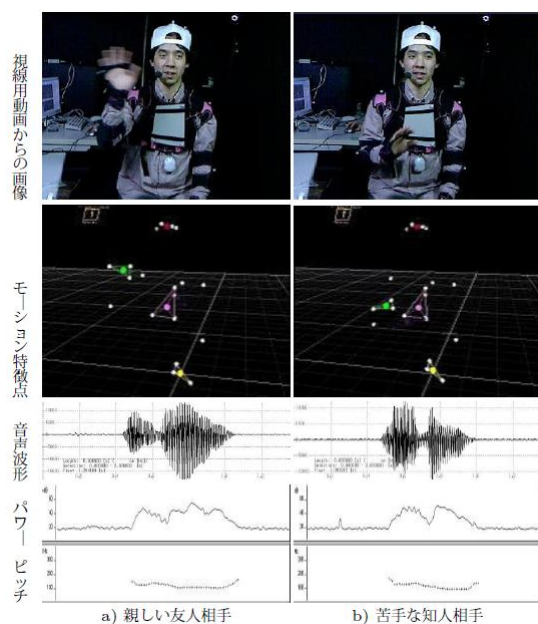


図2：マルチモーダルデータ「バイバイ」

(3) マルチモーダルデータからの個性分析

取得したデータに関する個性分析として、個人間表現差と個人内表現差（想定する相手に応じた表現差）を検証した。個人間差と個人内差が存在する中で、個人内差の度合いも個人間差に含まれる。

インタビューによる主観的・客観的個人性格特性等に応じて個人内差の幅が異なる傾向などが見られた。例えばグラウンド系運動部の現役監督や選手は声による表現が主で、ジェスチャは伴わないことが多く、想定する相手による表現の差は生じにくい。

また、年齢層と性差による傾向も見られた。例えば、女性の方が、親しい知人に対する表現が豊かで、苦手な相手に対する表現との差が態度に表れやすく、動きや音声だけでなく視線を合わせるか否かや顔表情にも違いが表れた参加者が多かったのに対し、中高年齢層男性では個人内表現差の少ない被験者が多かった。

様々な背景や性格によって個人ない表現

差は生じると考えられるが、個人間差は、その個性だけでなく、表3に表す表現幅タイプによっても表せると考えられる。

表3 個人内表現差によるタイプ分類

| | 個人内差 a 小 | 個人内差 a 大 |
|----------|----------|----------|
| 個人内差 b 小 | 平坦表現 | 率直表現 |
| 個人内差 b 大 | 丁寧表現 | 臨機応変 |

最後に各年度の成果についてそれぞれ報告する。

平成20年度の研究実績は、擬人的媒体に載せるモダリティのうちの音声表現に関する検討、および、予備実験における取得データの分析を行った。自然な表現を実体型擬人的媒体において実現するためには音声などの表現の連続性に関する知覚的検討が必要である。発表文献は、知覚的検討の過去の結果からさらに連続性を自然にするための「表情付与」に関する非線形性を検証したものである。この結果は音声においてのみ有効というよりは、何らかの時間的連続性を持つ表現における連続的な表現の表出手法に適用できるものであると考えられる。

また、マルチモーダルデータ取得の予備実験を一部行ったものの、繰り返し事項に述べたとおり、マルチモーダル表現をジェスチャとともに取得する環境（VICON ルーム）の使用において、所属組織内での調整がむずかしくなったため、別の取得方法として、センサによる体の動きや加速度を取得する手法を考案し予備検討を行った。この手法では、個人性を表す微妙な変化や差分を十分に表現しうるものではないことが考えられた。

（この検討に関しては平成21年度の検討に持ち越し、予算を繰り越した）

この問題点を解決する前に次年度の予定であった擬人的媒体の構成・構築や、擬人的媒体と人間のインタラクションにおける実際的なコミュニケーションのあり方について、個人化の前の土台として実験を行い、文脈適応を含めた基礎的なインタラクションの妥当性を検証した。これは、個人性表現を搭載した擬人的媒体における表現の検証だけでなく、中長期的利用における個人性評価に影響する重要な知見を得たと考えられる。

平成21年度の研究実績は、まず平成20年度に繰り越した個人化マルチモーダル表現のための、年代別・性別別の日常挨拶マルチモーダルデータの取得を行った。この際、平成20年度に問題となったマルチモーダルデータ取得環境を新たに構築し、新規追加データ等が取得しやすい環境を準備した。この実験では合計36名の参加者により、「シニア

層」「中間層」「若年層」の3種類の年代で、男女が各5名以上集まるようにしている。日常挨拶のマルチモーダルデータ同時取得実験とともに、別途音声のみの日常挨拶（平坦表現）を取得することで表現の幅を広げようと考えた。その結果、スポーツなどの趣味がある方々に共通する特徴として、声の抑揚のみにより表現を行う、ということが見られた一方、女性の半数程度はしぐさで声の表情不足をカバーするといった表現特徴がみられた。日常挨拶の想定相手を変化させ、相手に応じて抑揚が増減することもわかってきている。現在この分析は続行中であり、H22年度成果へつなげる予定である。

また平成21年度の純粋研究実績として、平成20年度のうちに行ってきた擬人的媒体と人間のインタラクションにおける実際的なコミュニケーションのあり方について、個人化の前の土台としての研究成果を発表した。これは、家族のような個性を表現するロボットと日常を共にすることを前提とし、そのロボットの表現やインタラクションがユーザにとって煩わしいものにならないよう、ユーザの音声や顔向き（視線）に応じて、ユーザの話しかけている相手を判定し、明示的にロボットに話しかけている時に要件などを伝え、それ以外の時に用件がある場合は、「話しかけ意図行動」を示し、ユーザに徐々に気付かせるという、気遣いを見せることが重要という結果をまとめたものである。この表現手法は、時に無遠慮であることが個人性表現に役立つ可能性が示唆されたという意味でも重要な知見である。

平成22年度の研究実績は、平成21年度に行った個人化マルチモーダル表現のための、年代別・性別別の日常挨拶マルチモーダルデータの取得実験の結果を分析し、研究会発表ではあるがマルチモーダルフォントと題して外部発表を行うとともに、ロボットへの適用に関する予備検討を行った。

特に、個人内・個人間変化を分析し、大局的に2種類の表現変化に分類を行うことを提言した。個人内表現差 a は同年代の相手で「親しい」・「苦手な」相手に対する態度の表現差を示し、個人内表現差 b は「同年代」と「目上の人」を相手とする時の態度の表現差を示す。この時、個人内表現差 ab の大小を分けるマトリクスから個性を導き出せるという、長期的な個性の示し方を発見した。

その一方で、単一モダリティやクロスモダリティ効果を分析したところ、個人の背景（例えば体育会系など）により、表現強度の高いモダリティが異なることがわかった。例えば屋外チームスポーツをやっている人は、日常挨拶にも大きな声を出し、身振りや手ぶりはあまり現れない。高齢層と若年層の女性

は身振りや手ぶりが出やすいが、そのストロークなどは年代や個性により異なる。

擬人的媒体における再現は選択部位的に予備検討にとどまったが、単純化の手法は体の部位の座標から計算される角度を適用するだけではなく、強調されるべき部位として肘以下の手の動きが反映されることが望ましい可能性が示唆されている。今後も部位反映に関する検討を継続する。

個人内表現差の増幅には中点を決定するのは取得データに基づく他ないが、個人間表現差の増幅のためには個人間変化量の中点がいずれにあるかを大量個性データにより検討するほうがより望ましいことも考えられる。しかしこれに対し、ある程度の人数による主観評価を求めることで、間接的に中点を決定できると結論付けた。

(4)今後の展望

昨今、マイクロソフトキネクトなど容易に簡略化された人間の特微量とモーションキャプチャが可能なシステムが研究終了タイミングで市場に出つつあり、新規データ取得手法として着目できる。

今後、新規個性を追加する枠組みを、大がかりなモーションキャプチャシステムに頼るだけではなく、これまでに取得されたデータの傾向による補間データと共に活用することで、簡易型キャプチャによる個人性データの取得方法を確立できると考えている。これにより、最低限のデータ取得とデータ補間方法を再度検討していく必要があると考えられる。データ取得の枠組みが変化する時にもフレキシブルに対応する補間手法の生成が、本研究課題の後に与えられた急務であると考えている。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計1件)

Tomoko Yonezawa, Hirotake Yamazoe, Akira Utsumi, and Shinji Abe, "Anthropomorphic awareness of partner robot to user's situation based on gaze and speech detection," IJAACS (International Journal of Autonomous and Adaptive Communications Systems), Inderscience, in press. (査読有)

[学会発表] (計5件)

米澤朋子, 鈴木紀子, 安部伸治, 間瀬健二, 小暮潔, "音声モーフィングによる歌声の声

色強度変化の知覚特性の分析,” 情報処理学会研究会, SIG HCI-MUS (2008-05), pp. 25-30, 2008. (査読無)

米澤朋子, 山添大丈, 内海章, 安部伸治, “ユーザの視線・発声に対するクロスモダリティ適応型のロボット行動設計,” HI 学会 SIGDE 研究会, no. 56, pp. 1-6, 2009. (査読無)

Tomoko Yonezawa, Hirotake Yamazoe, Akira Utsumi, and Shinji Abe, “Evaluating Crossmodal Awareness of Daily-partner Robot to User’s Behaviors with Gaze and Utterance Detection,” CASEMANS2009, pp. 1-8, 2009. [Best Paper Award] (査読有)

米澤朋子, 山添大丈, (1T-2 のみ神山祐一), 内海章, 安部伸治, “ユーザの視線・発声に対するクロスモーダルウェアネス～ユーザ状況把握を表すロボット,” HAI シンポジウム 2009, 口頭発表 2C-3 および体験デモ 1T-2, 2009, [Impressive Experience Award 受賞]. (査読有)

米澤朋子, “マルチモーダルフォント：日常表現における個人性データの取得に関する報告,” ヒューマンインタフェース学会第 71 回研究会, SIG-DE-05, pp. 53-56, 2011. (査読無)

6. 研究組織

(1) 研究代表者

米澤 朋子 (YONEZAWA TOMOKO)
株式会社国際電気通信基礎技術研究所・
知能ロボティクス研究所・研究員
研究者番号：90395161

(2) 研究分担者

なし

(3) 連携研究者

なし