

日本人の表情特性に応じた感情認知を行う キャラクターエージェント

大野つばさ^{†1} 小林夏美^{†1} 川合康央^{†1}

概要: 近年、コミュニケーションの相手が人間以外に広がりつつある。人間同士のコミュニケーションには言葉や身振り、手振り、そして表情などの非言語情報が大きく関わっている。しかし、人間と機械がコミュニケーションを取るのと、人間同士のコミュニケーションを取るのでは自然さが十分ではない。そこで本研究は、表情に応じて異なる反応を返すエージェントを作成し、人間の表情特性に応じて感情の認知を行い、感情を考慮した自然対話の実現を目指したキャラクターエージェントの開発を行った。

1. はじめに

近年、人工知能やロボットなどの分野で、人間との対話に関する開発・研究が進んでいる[1,2,3]。特に Apple 社の Siri や Google Assistant, Amazon Echo などのシステムは、スマートフォンやスマートスピーカーとして、私たちの日常生活に浸透している。この対話システムは大きく分けて 2 種類存在する。一つは、会社やホテルの受付などの「何かを達成するための会話」であり、タスク指向型対話システムと呼ばれる[4]。これらの会話システムは、ユーザ（人間）の身振り・手振りや表情を用いなくてもタスクの達成は可能である。一方で、タスク指向型対話システムとは逆に、会話を継続させることが目的となるものを、非タスク指向型対話システムと呼ぶ。非タスク指向型対話システムは、身振り・手振りや表情などの非言語情報も非常に重要なものとして扱われる。人間同士のコミュニケーションには、言語情報とともに、身振り・手振り等のジェスチャーや表情といった非言語情報を同時に用いて成り立っている[5]。人間は言語情報から相手の意図を把握し、相手の視線や表情、声のトーンから感情を予測・認知する。

現在、開発・研究が行われている音声対話システムでは、この非言語情報への対応が十分ではないため、人間同士の自然な対話の実現が課題となっている。人間と機械の会話では、一般的に機械の発話が機械音声となる。そのため感情によって変化しない機械の話し方は、人間の変化する感情に適合できないことが多い。例えば、人間が怒っている時に対応するのが、人間であれば怒りを落ち着かせるような有める話し方をするだろうが、機械は、設定された通りの話し方で会話するため、却って怒りを増やしてしまう可能性が高く、人間同士の自然な対話とは程遠いものとなる。相手の感情を認知できる人間だからこそ、相手の感情に合わせて声のトーンや話し方などを適応させることができるのである。

本研究では、非タスク指向型対話システムで非常に重要な要素である非言語情報の感情に着目し、ユーザの感情に応じて異なる表情・アニメーションや言葉を返すキャラクターエージェントの開発を行うこととした。本システムでは、ユーザの表情から感情を認知するという点に重点を置いている。ユーザの表情から感情を読み取るには、人間同士でも国、文化によって、認知する感情は差があるとされている[6]。本システムでは、Microsoft 社の FaceAPI[7]を用いて表情の分析を行い、レスポンスされた感情の値と日本人の表情を分析した値を用いて、日本人の表情特性に応じた感情認知を行い、人間同士の感情を用いた自然対話のような対話システムの実装を目的とする。

2. システムの構成

2.1 システムの概要

本システムは、まず始めにオープンソースの 3D モデリングソフト blender[8]を用いてキャラクターエージェントのモデリングを行った。次に、キャラクターエージェントには Miku Miku Dance[9]を用い、感情に応じて異なるモーションの作成と表情付けを行った。そのため、ユーザが喜んでると認知した場合には、キャラクターエージェントが飛び跳ねて両手を挙げ、表情も満面の笑顔となる。

モデリングを行ったキャラクターエージェントをゲームエンジンである Unity に取り込み、Miku Miku Dance で作成したキャラクターのモーションデータを、キャラクターエージェントに装着した Animation コントローラへと取り込んでアニメーションの設定を行った。また、同じく作成した表情モーションについても、各感情に関連付けを行っている(図 1)。

さらに、Unity 上で Microsoft 社の FaceAPI と IBM 社の Watson Conversation[10]の 2 つの API に接続して、処理を行うスクリプトコンポーネントを組み込んだ(図 2)。

^{†1} 文教大学



図 1 喜びのモーション

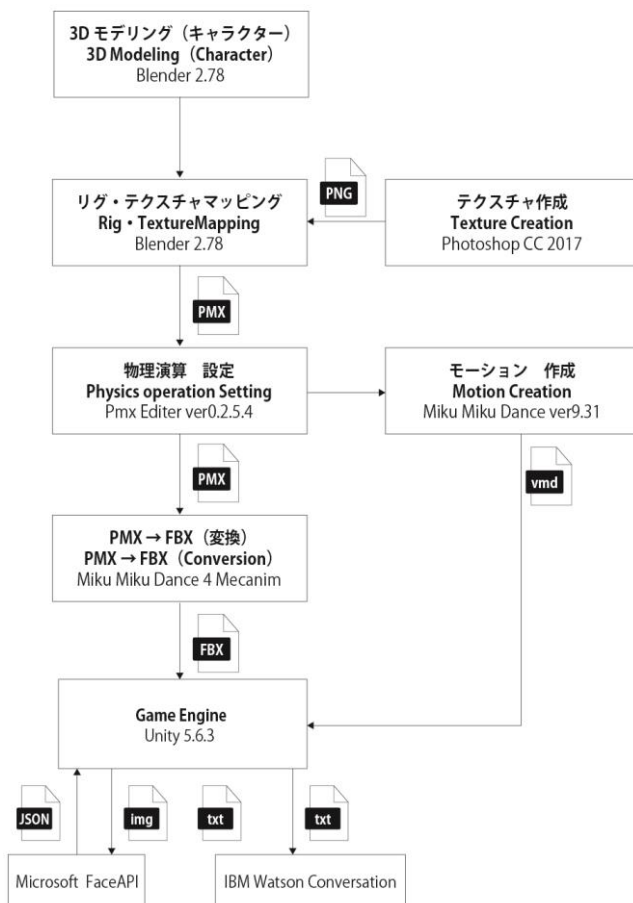


図 2 開発フローチャート

ユーザ表情の入力には、一般的な Web カメラを用い、取得された映像から FaceAPI の感情の値を求め、プロトタイプではその最も高い感情指数をユーザの感情とした。キャラクターエージェントはその認知した感情を元にアニメーションや Watson Conversation を用いた感情に応じた会話を返す構造となっている。感情の変化に応じて、キャラクターエージェントが話かけてくる言葉が異なるので、ユー

ザは自分の感情をキャラクターエージェントに理解されたという感覚を持って対話することができる。

2.2 キャラクターデザイン

本研究で用いたキャラクターエージェントは、既存の 3D モデルを用いず、一からモデリングを行っている。キャラクターの 3D 形状モデリングやリグ (骨)・テクスチャマッピング (肌・髪・服など) を、3DCG の統合環境である blender で作業した。テクスチャの作成には Adobe の Photoshop CC 2017 を用い、肌や髪、服のデザインを行った。また、髪の毛の動きなどの物理演算は、Pmx Editor を用いた。さらに、感情に応じたアニメーションのモーション・表情モーションの作成には Miku Miku Dance を用いた。

2.3 感情の測定方法

プロトタイプシステムでは、Microsoft の FaceAPI を用い、顔画像の表情から 8 つの感情を数値化したものを取得する。しかし、実際に値を取得してみると、Microsoft の FaceAPI にレスポンスされた値は、笑っているのに「中性」、怒っているのに「悲しみ」と誤認識される結果も少なくなかった。そこで、日本人の表情では感情判定の精度に課題があるのではないかと考えた。日本人の大学生 (20 代前半の男女、6 名) の顔を用いた数値測定を行い、日本人の表情特性に応じた感情認知を行うこととした。日本人大学生が怒り、喜び、悲しみ、驚きの 4 つの感情に応じた表情をした際の、FaceAPI の 8 つの感情の値を取得し、その際に得られた FaceAPI の 8 つの感情それぞれの値の平均値を係数として、4 つの感情 (怒り、喜び、悲しみ、驚き) に重み付けを行い感情の判定するシステムとした。

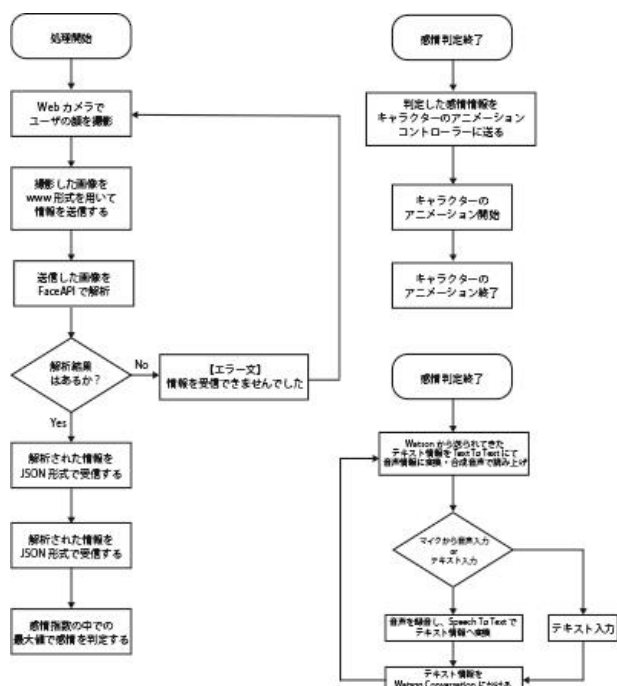
2.4 Watson Conversation を用いた理由

本研究には、IBM の Watson Conversation を用いて会話を行っている。類似する会話 API として、ドコモ [11] やリクルート [12] が開発した API が挙げられるが、感情情報をこれらの会話 API に送信しても、テキスト情報からの感情取得ができないため、感情に沿った返答を返すことは難しい。一方、Watson Conversation は会話パターンを学習させることが可能であるため、より人間同士の会話に近いものを開発できるのではないかと考え、本 API を採用した。

2.5 対話の手順

キャラクターエージェントとの対話には、Watson Conversation を使用した。まず、Watson Conversation には、事前に受信した感情情報に応じた返答について、事前に学習を行い、1 つの感情に 5 つの返答内容の中からランダムで応答するように準備を行った。手順としては、まず画像からユーザの感情の判定を行い、その判定された感情情報を Watson Conversation に送信する。ユーザの感情に応じた

Watson Conversation では、ユーザから入力されたテキスト情報を分析し、事前に学習させた熟語や返答パターンに照合してテキストに対する返答を行うようになっている。今回の開発では、Web カメラから表情を検出して、分析を行った後に、上記の方法で感情の判定を行う。認知した感情の情報を Watson Conversation に送り、1つの感情に対して5つ程度の返答内容の中からランダムで応答するようにした(図3)。



最初の感情判定から 10 秒が経過する毎に表情を用いた感情判定が行われ、その感情が前回の判定結果と異なった場合には、話の内容が変わる。これによって人間と対話しているような状態を再現した。感情判定の秒数間隔は容易に変更可能なため、ユーザの感覚に応じて応答間隔の変更が可能である。

キャラクターエージェントに会話をさせる方法については、Audio Source オブジェクトを利用して発音させ、口の動き（リップシンク）については、PMX ファイルから FBX ファイルへの変換に利用した MMD4 Mecanim の付属機能である MMD Speech Helper を使用し、発話内容に沿って口が動くように実装した(図 4).



システムの開発を行っていく工程で、システムをプレイ

したユーザから、感情が分かりにくいといったフィードバックがあったため、これを解決するため、プレイ画面のキャラクターエージェントの左側に各感情の割合が棒グラフ及び数値で表示され、各感情の値を視覚的に確認することが可能なものとした(図5)。またキャラクター上部には、感情に沿った顔文字を表示し、感情の可視化を行っている(図6)。このグラフ・数値の表示と感情の可視化を行う顔文字の配置により、ユーザは直感的に感情の判定結果を認識することが可能となった。

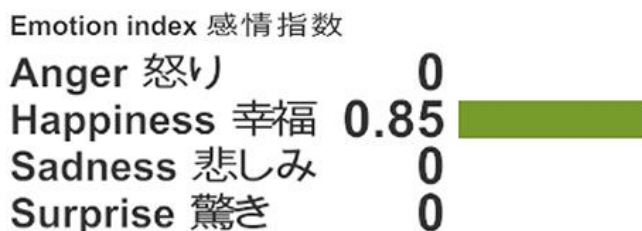


図5 感情の数値・グラフ表示



図6 ユーザの感情を表す顔文字

また、ユーザの顔の角度等で、感情の判定を失敗した場合には、赤の×マークをキャラクターエージェントの上部に表示し、10秒後に再度、表情の認識をするように設定を行った。

3. まとめ

今回は、人間と機械の対話を人間同士の非言語情報を用いた自然な会話に近づけることを目的として、非言語情報の中でも表情から求める感情に着目し、キャラクターエージェントの開発を行った。本キャラクターエージェントは、ユーザの表情から感情を認知し、それに応じて異なる反応を返すことが可能なものを開発することができた。

今回の日本人の表情特徴に応じた感情認知を行うエージェントの開発では、いくつかの課題が存在するものの、認知した感情に応じて話題が推移し、キャラクターの表情やアニメーションが変化するシステムの開発を行うことができた。声のトーンの課題や学習量の問題を解決することができると、人間同士の自然な対話に近づけることは容易だと考えている。また、今回は表情から感情を判定し、人間との会話に近づけることを目標として開発を行ってきたが、この会話エージェントに少しの手を加えるだけで、多種多様な使い方を考えることができる。例えば、楽天市場商品検索 API[13]などの販売商品をキャラクターエージェントが把握して、ユーザに商品を勧める。その勧めた商品を見たユーザの表情から感情を判定し、欲しいか欲しくないかを判定するなどの買い物お助けエージェントを製作することも容易である。つまり、機械がユーザの感情を把握することによって人間同士の雑談のような娯楽的要素だけでなく、人間の生活の困っている問題の解消にもつながる可能性を秘めているということが改めて実感できた。

また、日本人大学生が怒り、喜び、悲しみ、驚きの4つの感情に応じた表情をした際の、Microsoft社のFaceAPI[の8つの感情の値を取得し、その際に得られたFaceAPIの8つの感情それぞれの値の平均値を係数として、4つの感情(怒り、喜び、悲しみ、驚き)に重み付けを行い感情の判定するシステムとしたことによって、日本人の表情でもユーザの感情と数値の結果に差がなくなった。これによって、国や文化の違いによる表情特徴によって認知する感情の違いがあるということも確認することができた。

上記のことと同時に、今回の開発で以下の課題が見つかった。感情を考慮して対話の内容を変更することが可能であるが、合成音声のため、声の抑揚を変更することが不可能であった。そのため、キャラクターエージェントの返答内容と音声のトーンが一致せず、人間同士の自然な会話と比較して、やや違和感を持ったものとなった。人工音声に感情を取り入れられるかが課題となった。

また今回の感情判定方法にはAPIの数値を用いたが、日本人の各感情を表情にした写真や動画などを用いた画像情報に関する機械学習を行うことで、感情判定機の製作が可能のため、この方法での判定機での結果と本研究の数値を用いた判定機の結果を比較することで、新たな発見がある可能性が出てきた。

今後の研究では、上記の課題点の解決策を検討しつつより人間同士の自然な対話を行えるキャラクターエージェントの製作を進めていくこととする。

参考文献

- [1] 小野哲雄, 今井倫太, 石黒浩, 中津良平. 身体表現を用いた人とロボットの共創対話. 情報処理学会論文誌, 2001, vol.42, no.6, p.1348-1358.
- [2] 小野哲雄, 今井倫太, 江谷為之, 中津良平. ヒューマンロボットインタラクションにおける関係性の創出. 情報処理学会論文誌, 2000, vol.41, no.1, p.158-166.
- [3] 伊藤亮介, 駒谷和範, 河原達也, 奥乃博. ロボットとの音声対話におけるユーザの心的状態の分析. 情報処理学会研究報告音声言語情報処理, 2003, p.107-112.
- [4] 平沼英翔, 三武裕玄, 長谷川晶一. 実会話事例の機械学習により相手の話し方に合わせた音声応答を行う対話エージェント. エンタテインメントコンピューティングシンポジウム 2017 論文集, 2017, p.257-261.
- [5] 鈴井智史, 渡辺光司, 佐藤美恵, 春日正男, 羽鳥一夫. 顔器官における感情別特徴を利用した表情自動認識手法. 社団法人映像情報メディア学会
- [6] Rachael E. Jack, Oliver G. B. Garrod, Hui Yu, Roberto Caldara, Philippe G. Schyns. Facial expressions of emotion are not culturally universal. PNAS, 2012, vol.109, no.19, p.7241-7244.
- [7] “Microsoft Azure Face”, <https://azure.microsoft.com/ja-jp/services/cognitive-services/face/>, (参照 2018-7-30).
- [8] “Blender”, <https://www.blender.org/>,
- [9] “MMD (Miku Miku Dance)”, <https://www6.atwiki.jp/vpvpwiki/pages/187.html>
- [10] “Watson Assistant(旧 Conversation)”, <https://www.ibm.com/watson/jp-ja/developercloud/Conversation.html>,
- [11] “ドコモ自然対話 API”, https://dev.smt.docomo.ne.jp/?p=docs.api.page&api_name=natural_dialogue&p_name=api_usage_scenario, (参照 2018-12-20).
- [12] “Talk API”, <https://a3rt.recruit-tech.co.jp/product/talkAPI/>, (参照 2018-12-20).
- [13] “楽天市場系 API”, <https://webservice.rakuten.co.jp/document/#ichibaApi>,