

対話事例に基づく機械学習を用いた同調的表情を提示する 対話エージェント

藪下剛史^{†1} 三武裕玄^{†1} 長谷川晶一^{†1}

概要: 対話エージェントが社会に溶け込み楽しい対話相手となるには、いかに人間のような振る舞いを自然に行えるかが重要となる。本研究ではエージェントの自然な振る舞いを決定する要素の一つである表情に注目し、人間と社会関係を形成する上で重要となる同調的表情を自然に行う手法を提案する。具体的には、実際の対話における話者と聴者の表情変化や音声を数値化したものを時系列データとして記録しHMMを学習する。次に得られた学習モデルに話者の表情と音声を入力して聴者の行動を推定し、対話エージェントの表情変化を生成する。

キーワード: 機械学習, 隠れマルコフモデル, 対話エージェント, 表情

Conversational Agent Learning facial contagion of actual conversation example

TSUYOSHI YABUSHITA^{†1} HIRONORI MITAKE^{†1}
SHOICHI HASEGAWA^{†1}

Abstract: Conversation agents are expected to be a life partners with social behaviors. It is important for Agents to give social and natural action like human. We propose a method to generate facial expression synchronized with the speaker's facial expression. The method includes to capture facial expression and voice volume, and learning HMM with the conversation data. Using the HMM as facial expression determination model, our conversation agent shows natural facial expression based on actual conversation example.

Keywords: Machine Learning, Hidden Markov Model, Conversational Agent, Facial expression

1. はじめに

近年、情報処理技術の発展により、エンターテインメントロボットをはじめとする人間とインタラクションを行うエージェントが社会に普及しつつある。例として客に対し接客・受付を行うロボットの「Pepper」[1]や、Web上でユーザーの要望や相談に応じてくれるAI対話型Webエージェント「Desse」[2]が挙げられる。このようなエージェントは、道案内や介護施設での対話相手など、人間社会に自然に溶け込み人間同様に社会的役割を果たすことが期待されている。エージェントが人間に好印象を与え、自然な対話が行えるかどうかは非常に重要であり、人間はエージェントとの対話のなかで違和感を覚えると相手に社会性や知性を感じられず、対話のモチベーションを削がれてしまう。そのため、人間に対し好印象を与え自然な対話を促す対話エージェントの研究が現在盛んにおこなわれている。

人間同士の対話は、話し言葉や文字といった言語インタラクションだけでなく視線やジェスチャーといった非言語インタラクションを交えて行われるので、対話エージェントとの自然な対話を実現するためには、言語インタラクションだけでなく非言語インタラクションを実現することが必

要である。その中でも、表情変化は非言語情報の伝達に重要であり、インタラクション中のエージェントが提示する表情には自然な対話を実現するための重要な要素が含まれると考えられる。

対話における人間の表情変化は様々である。そのなかに、同調的表情というものがある。例えば、話者が面白い話をしているときは、話者と聴者は笑顔になる割合が高くなり、悲しい出来事を伝えるときはお互い暗い顔になりやすいといったように、人は他者の表情に対してそれと同調的な表情を表出することがある[3]。この表情の同調的反応を、Hinsz&Tomhave(1991)[4]は顔面伝染(facial contagion)と呼んだ。

同調的表情は人間同士のインタラクションにおいてはコミュニケーションスキルの1つと考えられている。相手の笑顔を見続けていると自分も愉快的気分になり、同じように笑顔を作ってしまうといった同調的表情の表出は日常生活で頻繁に起こっている。このように相手に同調した表情を表出することによって、相手は自分に対して好感度や信頼感といった肯定的印象を抱くと推測される[5]。その結果信頼関係が生まれ社会的関係の形成、促進に繋がる。この観

^{†1} 東京工業大学
Tokyo Institute of Technology

点から、同調的表情を表出する対話エージェントは、表出しないものよりも人間からの信頼を獲得しやすいはずである。

また、同調的表情は話者の表情や語調に合わせて表出するものであるため、それらが計測できればどのような同調的表情を表出すべきか推測がしやすい。つまり、同調的表情は話者の表情や語調に合わせて表出するものなので、エージェントの適切な同調的表情の表出には対話相手の非言語情報（表情、韻律など）を計測しエージェントの表情を生成する必要がある。一方で、対話相手が笑顔を見せていようが怒り顔を見せていようが、笑顔で同調するほうがより好感を持たれるという研究結果[3]がある。これらより、エージェントが対話相手の話の内容を理解し、それに応じて複雑で様々な感情を表情に表さずとも、対話相手の表情変化に対し、適切に笑顔を返すだけでも好印象を与えられるエージェントが作成できると考えた。そこで本研究では、非言語情報を用いて人間との対話において同調的表情を表出する対話エージェントの作成を目指す。

2. 関連研究

2.1 非言語情報を用いる対話エージェント

非言語情報も用いてコミュニケーションを行う対話エージェントの研究は数多くなされている。Devaultらの研究[6]では話者の発話や視線、表情やタイミングを返すといったエージェントを作成した。この研究では、行動データや行動決定モデルを手で精巧に作成することでジェスチャーの選択やそれを話者に返すタイミングについては適切である一方、表情については微笑みといった感情を読み取れるほどのものは表出されず、表情変化に乏しいため無機質な印象を持つ可能性がある。

2.2 同調的表情

同調的表情に関する研究も数多くなされている。埴淵らの研究[5]では、対面コミュニケーションの中で、信頼関係などの社会的関係を形成し、促進し、維持させるためには他者の肯定的な印象を抱かせることが重要であり、そのためには意図的に他者の表情に同調して自分の表情を表出することが必要であることが指摘されている。また、佐藤らの研究[7]では、話者の表情を見ると常にそれに引きずられて自分の表情を表出とはならず、同調の割合は対話の全観察時間の20%以下であることも報告されている。これは、敢えて同調的表情の表出頻度を低くすることで、対面コミュニケーションが感情的に共鳴しすぎることを防ぐ効果がある。

3. 提案手法

2.2節で述べたように、同調的表情は話者の表情に対し必ずしも起こることではなく、その表出のさせ方も様々である。そのため、対話事例を注意深く観察したとしてもその

法則性をアルゴリズムとして書き下すことが容易であるとは限らず、また無意識的に行う事が多いので聞き取りや内省で法則性を見いだすことも難しい。

そこで本研究では、機械学習を用いることによって、人間同士の対話のなかで生じる同調的表情を再現することを目的とする。

まず対話エージェントの表情決定モデルを実対話事例から学習することで得る。なお、今回は時系列データを学習することからモデルとしてHMMを用いる。次にそのモデルに対して、表情認識ツールやマイクから取得した話者の表情変化と音声の値をモデルに入力し、それに応じた対話エージェントの表情を出力として得る。ここで、表情だけでなく発話も入力値として用いる理由は、人間は笑う際に表情変化だけでなく、声量にも変化がある可能性を考慮しているためである。そして、モデルから得られる出力値によって対話エージェントの顔面をインタラクティブに動かすことで同調的表情を再現する。システムの全体像を図1に示す。

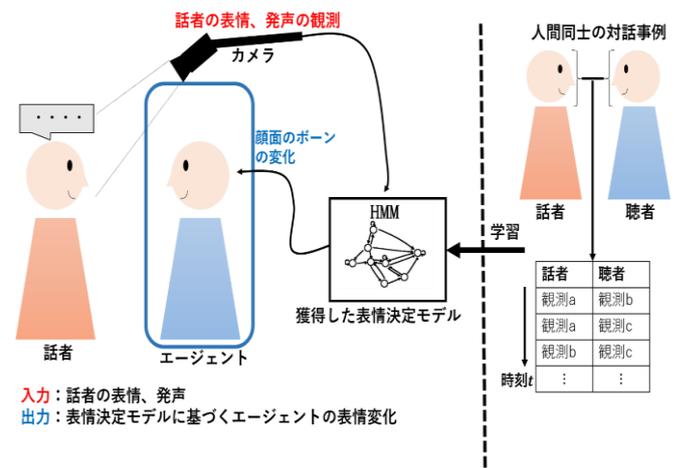


図1 システム全体像

Figure 1 The whole system.

3.1 実対話事例の収集

対話エージェントの表情決定の規範となる、人間同士の対話の様子を記録する実験を行った。1人の話者と、エージェント役として1人の聴者に対話を行ってもらい、その様子を記録した。今回の実験では、参加者の表情を記録するために表情解析ツールを用いた。また、発話を計測するためにマイクを使用した。

今回の実験では1対1の3分ほどの対話で、話者が聴者に対し一方的に話しかけ、聴者はそれに対し「うん」や「はい」といった最低限の応答を返すといった事例を対象とした。

3.2 実験環境

今回の実験では、表情解析ツールには「Kinect v2」[8]と「OKAOVision」[9]、マイクには「HYP-190H」[10]を用いた。

2人の参加者はマイクを装着し、話者側には Kinect v2, エージェント役側には OKAOVision をそれぞれの顔が十分計測できるように設置した。対話中の様子を図 2 に示す。

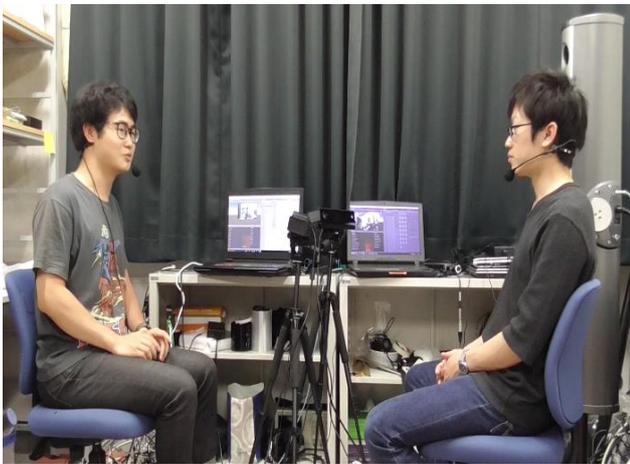


図 2 対話中の様子

Figure 2 State in conversation.

3.3 表情解析

マイクロソフト社が提供する Kinect v2 の機能の一つである High definition face tracking から、顔面における 17 個の特徴点の動き (Animation Units, 以下 AU) をリアルタイムで取得できる。このうち、作成する対話エージェントの表情を動かすために必要な 8 個の AU を学習に用いる。その AU の名称を表 1 に示す。これらの値と、マイクから得られるボリュームの値 [0,1] の 9 要素を時系列に並べたものを話者役の実験データ A とした。

また、オムロン株式会社が提供する OKAOVision の機能の一つである笑顔度推定から、カメラに映る人間の笑顔度を 0~100 までの数値として取得できる。この機能より得られる笑顔度を時系列に並べたものを実験データ B とした。そして、実験データ A と実験データ B の時系列をそろえたものを、機械学習に用いる学習データとした。

3.4 対話エージェントの表情

本研究で使用する対話エージェントのモデルは、キャラクター自体に親和的な印象を持ってもらえるように、微細な表情変化を可能とする「KizunaAI」[11]にした。

しかし、架空の 3D モデルであるエージェントの表情変化の仕方と、現実の人間の表情変化の仕方は必ずしも対応しているわけではない。例えば、エージェントの笑った表情を作る際、目が半開きの状態を維持させると違和感を覚える可能性が高い。これは、現実の人間の場合笑ったときに目が半開きであろうと違和感を覚えることは少ないが、現実の人間より造形が単純な 3D モデルのエージェントの多くは、目の半開き状態を維持することを想定して作られていないため、人間の表情変化では感じなかった違和感に気づくためである。そのため、現実の人間の表情の動きを記

録した AU 値をエージェントの顔面にそのまま適用し動かすと、エージェントの目が半開きの状態が続いた場合に違和感を覚える。その様子を図 3 に示す。

表 1 各 AU の名称とデータ範囲

Table 1 Name and data range of each AU.

Animation Units	名称	データ範囲
JawOpen	口を開ける	0,1
LeftcheekPuff	左頬を膨らませる	0,1
LefteyebrowLowerer	左眉を下げる	-1,1
LefteyeClosed	左目を閉じる	0,1
LipCornerDepressorLeft	唇の左端を下げる	0,1
LipCornerPullerLeft	唇の左端を引き上げる	0,1
LipPucker	唇を膨らませる	0,1
LipStretcherLeft	唇の左端を横に引く	0,1



図 3 人間と対話エージェントの表情の違い

Figure 3 Difference in facial expression between human and conversation agent.

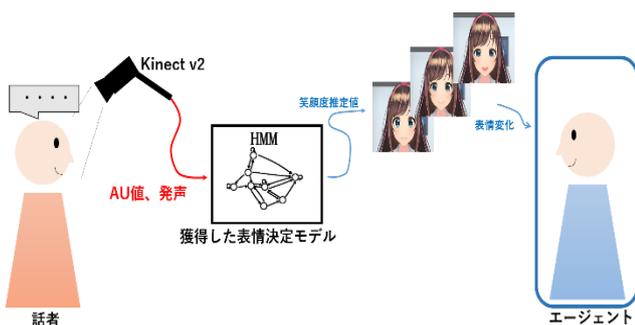
このような人間と 3D モデルの表情変化のずれを解消するために、対話エージェントの表情には、あらかじめ用意した 11 段階のキーフレーム間を補完しながら動かすキーフレームアニメーションを用いた。11 段階のキーフレームは、0 番目では無表情となっており、11 番目に向かって動かすにつれて笑顔になっていくものにした。その様子を図

4 に示す. なお, 同調的表情は笑顔に対して最も生じやすく, 怒りや悲しみといった表情に対しては同調しても好感度を得られにくいという検討[4]から, 今回は笑顔に特化した表情変化のみ行うことにした. そのため, 実験データ収集の際話者側には微細な表情変化をとらえるために Kinect の AU 値を用いているが, 聴者側の表情変化に基づいて動かすエージェントの表情変化は笑顔の度合いという一要素だけで良いので, OKAOVision の笑顔度を用いて一元的に制御することにした. この 11 段階の間を補充しながら動かすことで, 違和感のない自然な表情変化を再現する. この概要を図 5 に示す.



図 4 エージェントのキーフレーム

Figure 4 Keyframe of agent.



入力: 話者の表情、発声
出力: 表情決定モデルに基づくエージェントの表情変化

図 5 エージェントの表情変化

Figure 5 Facial expression change of agent.

3.5 HMM による機械学習

対話中における同調的表情は, 相手の表情変化を受けて表出されるものであることから, 実験データは時系列データとして扱う. また実験より, 同調的表情は対話中である程度の変化をもって度々起こることが見受けられた. そこで, 今回は時系列データを扱うことが出来, かつ特定の繰り返しパターンの発見に長けている HMM(Hidden Markov Model)を用いた. これを用いて 3.3 節で生成した学習データを状態数 50 として機械学習させたところ, 10 次元ベクトルの表情決定モデルが生成された. その内訳は, 1~8 次元目は表 1 で示した AU, 9 次元目はマイクのボリューム値, 10 次元目は笑顔度推定値である. 学習にかかった時間はおよそ 4~5 時間であった.

3.6 HMM による笑顔度推定値の計算手法

3.5 節で生成した表情決定モデルから出力値を得るための具体的な手法を示す. 獲得した表情決定モデルにおける状態 i の平均出力を μ_i , 共分散行列を Σ_i , その正規分布を $N_i(\mu_i, \Sigma_i)$, ステート i にいる確率を S_i とする. これらのベクトルまたは行列は, 用いた学習データの要素数 p に依存する. 各ステートにおいて,

$$X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \text{ with sizes } \begin{bmatrix} q \times 1 \\ (p-q) \times 1 \end{bmatrix} \quad (1)$$

が与えられたとする. これに従って

$$\mu_i = \begin{bmatrix} \mu_{i,1} \\ \mu_{i,2} \end{bmatrix} \text{ with sizes } \begin{bmatrix} q \times 1 \\ (p-q) \times 1 \end{bmatrix} \quad (2)$$

$$\Sigma_i = \begin{bmatrix} \Sigma_{i,11} & \Sigma_{i,12} \\ \Sigma_{i,21} & \Sigma_{i,22} \end{bmatrix} \text{ with sizes}$$

$$\begin{bmatrix} q \times q & q \times (p-q) \\ (p-q) \times q & (p-q) \times (p-q) \end{bmatrix} \quad (3)$$

とする. ここで x_2 は観測値であり, 話者の AU 値と音声である. $x_2 = a$ としたときの出力 x_1 の条件付き正規分布を求める. このとき得られる条件付き正規分布を $\bar{N}_i(\bar{\mu}_i, \bar{\Sigma}_i)$ とすると,

$$\bar{\mu}_i = \mu_{i,1} + \Sigma_{i,12} \Sigma_{i,22}^{-1} (a - \mu_{i,2}) \quad (4)$$

$$\bar{\Sigma}_i = \Sigma_{i,11} - \Sigma_{i,12} \Sigma_{i,22}^{-1} \Sigma_{i,21} \quad (5)$$

となる. 式(4), (5)より, $x_2 = a$ に対して出力 x_1 が選ばれる確率 $P(a)$ を得る.

$$P(a) = \sum_{k=1}^i S_k \bar{N}(a | \bar{\mu}_i, \bar{\Sigma}_i) \quad (6)$$

$P(a)$ によって得られた出力値を笑顔度推定値として, エージェントの顔面を動かす. 動作方法については後述する. そして, 観測値と出力値の組から, フォワードアルゴリズムを用いて表情決定モデルの各ステートの遷移確率の更新を行う. 本手法の概要を図 6 に示す.

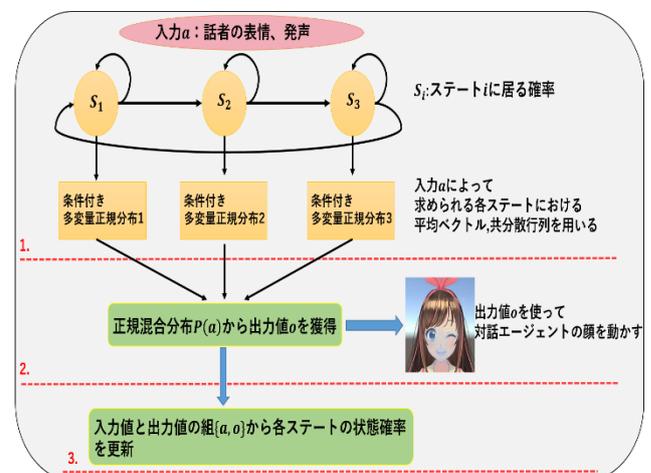


図 6 HMM による表情決定

Figure 6 HMM based Facial expression.

3.7 出力値によるインタラクティブ表情決定

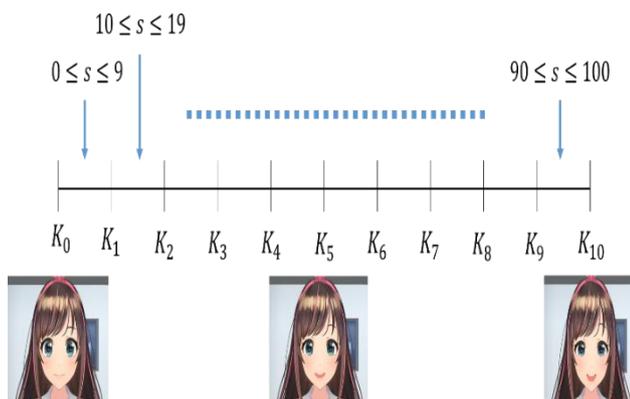
3.6節で得られた笑顔度推定値を用いて、エージェントの顔面の動作方法についての概要を図7に示す。得られた笑顔度推定値を s とおき、キーフレームアニメーションベクトルを $\mathbf{K}_n(n=0,1,\dots,10)$ とおく。 \mathbf{K}_n のそれぞれには、図7の表情を表出するエージェントの顔面ブレンドシェイプ値を保存してある。 s の値に応じて11段階のキーフレームアニメーション間を補完するために、 $[0,100]$ の範囲を11分割し \mathbf{K}_n を割り当てる。次に s の値と0~100の範囲の s の位置に応じて決める t を用いて

$$s' = \frac{s}{10} - t \quad (t = 0, 1, \dots, 9) \quad (7)$$

とし、最後に

$$\mathbf{S} = (1 - s')\mathbf{K}_t + s'\mathbf{K}_{t+1} \quad (8)$$

とすることで、 \mathbf{S} にはエージェントに表出させるべき顔面ブレンドシェイプ値が格納される。そして \mathbf{S} の値に応じて毎ステップエージェントの表情を動かすことにより、リアルタイムな表情変化を実現する。



例) $s = 15$ なら... K_1 と K_2 中間の表情が生成される

図7 表情の動作方法

Figure 7 How to operate facial expressions.

4. 実装

Unity[12]を使用し、提案手法による対話エージェントとインタラクションを行うデモを作成した。ディスプレイに対話エージェントの顔が十分映るよう配置し、ディスプレイの後ろにはデータ解析用にKinectとOKAOVisionをユーザの顔が十分映りこむように配置した。KinectによってユーザのAUを毎フレーム取得し、その値とマイクのボリューム値をHMMによって作成したモデルの入力値とした。また、このデモを体験する際は、評価用にOKAOVisionによる笑顔度も毎フレーム取得し、インタラクションの様子を動画で撮影しながら行うことにした。またユーザが対話エージェントの表情以外の部分について

違和感を覚える可能性を極力減らすために、対話エージェントには椅子に座っている姿勢を取らせ、常に体をゆっくりと揺らし呼吸しているように見せるアニメーションと、目のまばたき動作を実装した。このデモの様子を図8、図9に示す。

5. 評価

5.1 同調的表情の再現

作成したエージェントが実対話から学習できた同調的表情を再現するか確認するために、ユーザ1名がエージェントに話しかける実験を行った。同調的表情が再現できた様子を図10に示す。対話中は、エージェントが自発的に笑うことはあまりなく、ユーザの話が弾み笑顔になっていく変化を受けてからエージェントの笑顔が表出されているので、目的である同調的表情が再現されていることが分かった。また、エージェントは微細な笑いや大きな笑いなど笑い方にも多様性が見られ、ユーザの笑顔に対し笑顔を示さないことも度々見ることができた。



図8 デモの様子

Figure 8 Demo situation.



図9 デモの様子

Figure 9 Demo situation.



図 10 同調的表情的再現

Figure 10 Reproduction of facial contagion.

5.2 評価方法

提案手法の有用性を確かめるために、比較実験を行った。各実験条件とも、笑顔度からエージェントの表情を生成する部分は提案手法と同じとし、笑顔度を生成する手法として

- ・ランダム(RAND エージェント)
- ・ユーザの笑顔度に同期(IMI エージェント)
- ・ユーザの対話相手役の人間の笑顔度をリアルタイム計測して使用(WOZ エージェント)
- ・提案手法による笑顔度推定値(HMM エージェント)

を用意した。なお、IMI エージェントの笑顔度の同期は 400ms 遅らせて行う。これは、人が他者の表情変化を自身に対する応答と認識しやすい時間が 300ms から 1100ms であるためである[6]。

ユーザはこれら 4 種類の仕組みが異なるエージェントとそれぞれ対話を行い、「エージェントの中身が人間ではない」と気づくまでの時間を計測した。ユーザには、対話しているエージェントの中身が何なのかは一切説明せず、またユーザに提示するエージェントの種類の順番はランダムに入れ替えた。

次に、実験の効果を主観的な印象からも確かめるために、対話中の各エージェントに対しての印象を以下の 5 つの評価軸について 10 段階でアンケートに回答してもらった。

- Q1:話を聞いてくれているか
- Q2:自分の気持ちを汲み取ってくれているか
- Q3:親しみやすいか
- Q4:自分の話に興味を持ってくれているか
- Q5:機械的ではないか

5.3 結果

学内外から実験参加者を募り、6 名の被験者から回答を得られた。エージェント毎に、「エージェントの中身が人間ではない」と気づくまでの時間をグラフにしたものを図 11

に示す。図 11 より、どのユーザにおいても、IMI エージェントや RAND エージェントよりも HMM エージェントが比較的人間ではないと気づかれるまでの時間が長かった。一方、ほとんどのユーザにおいて、HMM エージェントは WOZ エージェントよりも早く人間ではないと気づかれてしまっていた。

次に、アンケートの各質問におけるエージェントに対しての点数の平均値を図 12 に示す。図 12 より、全体的に各エージェント間で大きく差がついた質問項目はなかった。Q1 や Q2 においては HMM エージェントの点数がわずかに高かったが、他の質問ではとりわけ良い点数を記録しなかった。

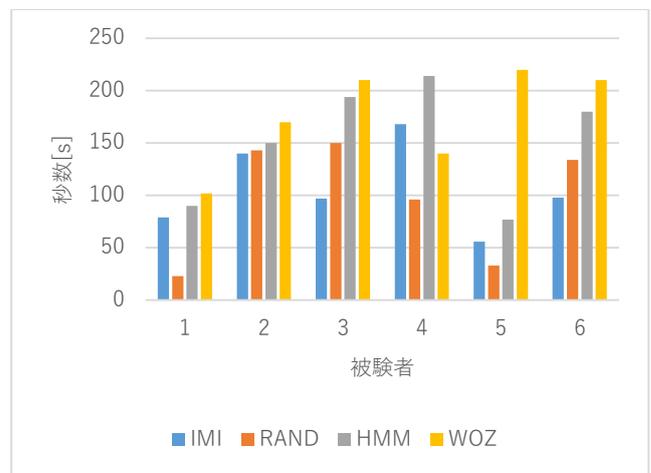


図 11 人間ではないと気づくまでの時間[s]

Fig.11 Time to notice that the opponent is not a human being.

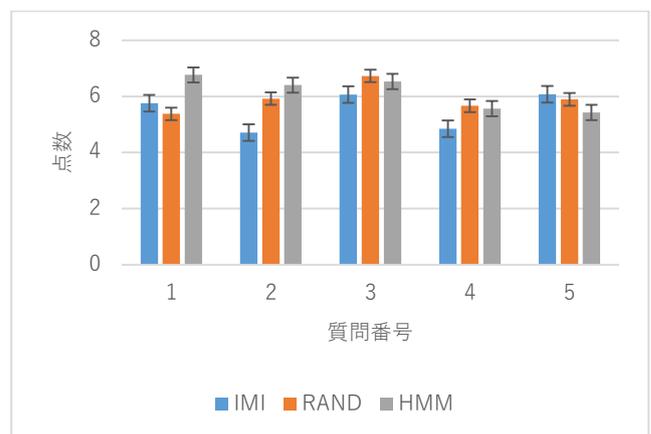


図 12 各質問におけるエージェントに対しての点数の平均値

Fig.12 The average value of points for agents in each question.

5.4 結果

5.1 節より、同調的表情的が機械学習によって再現できていることを確認した。

5.3 節の図 11 より、RAND エージェントが最も人間ではな

いと気づかれやすいことが分かった。これは、RAND エージェントはユーザの表情を認識していないので、対話中のユーザの表情変化に全く関係のない場面でも笑うことが多かったため、早い段階でユーザがエージェントに対し違和感を覚えたためだと考えられる。しかし図 12 のアンケート結果からは、Q1 においては RAND エージェントの点数がやや低い回答が多く見られたが、Q2 から Q4 の項目においては比較的高評価を得たため、一概に相手の表情変化を読み取らずに笑顔を作ることがユーザにとって悪印象に繋がると完全には言いきれない。笑顔のランダム生成が人間に与える印象について今後深く調査する必要がある。

また、相手の笑顔を模倣する IMI エージェントはユーザが笑顔でないときに笑顔になることは無いため、違和感を与えにくいと思われたが、人間ではないと気づかれた時間は HMM エージェントや WOZ エージェントより短い。アンケートの Q2 や Q4 においても、特に IMI エージェントの点数が低かったことから、ただ相手の笑顔をそっくり模倣するだけでは不自然な印象を与え、人間らしさを損なうことが分かった。

最後に、WOZ エージェントを除いて HMM エージェントが最も人間ではないと気づかれず時間が長く、そのまま対話が終了したケースが多かった。つまり、IMI エージェントと RAND エージェントに比べ、HMM エージェントはより人間らしいエージェントであることが明らかになった。一方、アンケートにおいて特に目立って良い評価を得られた質問はなく、HMM エージェントが人間に対し好印象を与えられるかどうかは今後更に改善、検証をしていく必要がある。

6. 今後の課題

確かに HMM エージェントは比較的人間らしい印象を与えやすいことが分かったが、図 11 よりほとんどの被験者にとって HMM エージェントは WOZ エージェントよりも人間ではないと気づかれやすいことが分かる。このことから、ユーザは HMM エージェントと実際の人間との挙動に明白な違いが感じられることが推測される。この差の原因はアンケート結果の評価軸からは見出すことが出来なかったが、これは対話中に HMM エージェントが稀に大げさに笑いすぎることや、同調的表情的の表出回数が少ない場面があったことが原因として考えられる。実験データを収集する段階でエージェント役となる聴者がいかに豊かにかつ自然な表情変化を行えるかによって、生成される表情決定モデルの精度が左右する可能性が考えられるため、今後実験の際にはエージェント役に適した人物を吟味する必要がある。

また、今回の実験は3分あたりの対話を対象としたが、より長い対話を実験データにすることで、更に細かい表情変化のパターンを学習が期待できる。また、エージェント役の聴者の表情変化の癖が表情決定モデルに現れるため、聴

者の演技によっては表情変化に個性を持たせたエージェントの作成が期待できる。

謝辞 本研究は科研費(17K17713)の助成を受けたものである。また本研究を進めるにあたり、日ごろから様々なアドバイスを頂いた長谷川晶一先生、三武裕玄先生に深謝する。また、快く被験者を引き受けて頂いた研究室の方々や、他研究室、学外からきて頂いた友人達にも、深く感謝の意を表す。

参考文献

- [1] “Pepper(一般販売モデル)”
<https://www.softbank.jp/robot/consumer/>
- [2] “AI 対話型 Web エージェント「Desse(デッセ)」”
<https://www.scsk.jp/product/common/desse>
- [3] 市川寛子. 二者対面コミュニケーションにおける同調的表情的表出. 筑波大学博士(行動科学)学位論文. 2008
- [4] Hinsz, V. B., & Tomhave, J. Smile and (half)the world smiles with you, frown and you frown alone. *Personality and Social Psychology Bulletin*, 17, 586-592, 1991
- [5] 埴淵俊平,伊藤京子,西田正吾. 同調的表情的表出を提示するインタフェースの提案-2 者間会話環境に向けて-. インタラクシオン 2010(ポスター発表・議論部門スタンダード), Vol., no.2010, pp.3, ポスター発表, 2010
- [6] David DeVault et al. SimSensei Kiosk: A Virtual Human Interviewer for Healthcare Decision Support, *Proceeding AAMAS '14 Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, Pages 1061-1068
- [7] Wataru Sato, Sakiko Yoshikawa: Spontaneous facial mimicry in response to dynamic facial expressions, *Cognition*, Vol.104, pp.1-18, 2007.
- [8] “Kinect Sensor for Xbox One”
<https://www.microsoft.com/en-us/store/d/Kinect-Sensor-for-Xbox-One/91HQ5578VKSC>
- [9] “オムロン OKAOVision”
<http://plus-sensing.omron.co.jp/technology/>
- [10] “ハンズフリーマイクロホン HYP-190H”
https://www.audio-technica.co.jp/mi/show_model.php?modelId=2531
- [11] “© Kizuna AI”
<http://kizunaai.com/>
- [12] “Unity”
<https://unity3d.com/jp>