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．

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 behaviours. 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.

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と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，自然な対話を実現するための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多くの割合を占めており，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な気分になり，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ので，それらが測定できればどのような同調

†1 東京工業大学
Tokyo Institute of Technology，Yokohama，Kanagawa 226-8503，Japan

的表情を表出すべきか推測がしやすい．つまり，話の内容といった言語情報ではなく，話者の表情や声の大きさといった非言語情報から自然な表情によるインタラクションを行える可能性がある．そこで本研究では，非言語情報を用いて人間同士の対話における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. 予備実験

2.1 話者の表情を真似る対話エージェント

同調的表情とは，つまるところ相手の表情の模倣である．そこで，ただ相手の表情を真似るだけでも好印象を抱ける自然なインタラクションが実現するか検証するために，一定のタイミングの遅延を伴って話者の表情を単に真似するだけの対話エージェントを作成した．エージェントの表情の遅延は，人間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 であるという検討[3]から，400ms とした．その様子を図 1 に示す．

この検証の結果，対話エージェントに話しかけていても，途中から明らかに「真似されているだけ」という感覚が強くなり，違和感を覚えてしまうことが分かった．このことから，積極的な対話を促し人間に好印象を与える対話エージェントを作成するには，同調する際に表情をただ模倣するのではなく，模倣のタイミング，度合いも再現する必要がある．しかし，正しい同調的表情を表出させるルールを人手で作成するためには，実対話を注意深く観察し，法則性を発見しアルゴリズム化する作業を伴うため，容易ではない．

そこで，本研究では実対話事例から HMM を用いた機械学習によって表情決定モデルを作成し，話者に対して正しい同調的表情を表出させることを目的とする．

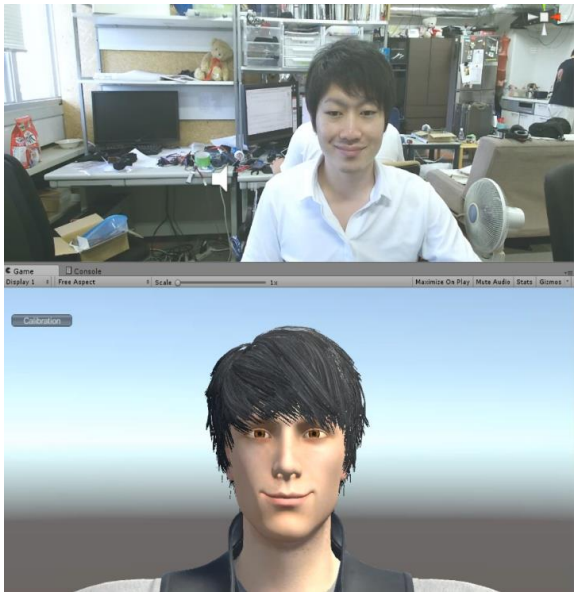

図 1　話者の表情を真似るエージェント

Fig.1　The Agent who imitates the expression of the speaker.

2.2 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．同調的な表情変化の要因には会話内容・相手への音声・表情変化などが考えられるが，本研究では会話内容に依らず相手の音声・表情変化のみによって引き起こされる表情変化を対象とする．このような表情変化の事例を効率良く収集するため，話者は聴者が理解できない言語を用いて話しかけるものとする．

2.3 実験環境

表情解析ツールには「Kinect v2」[6]，マイクには「HYP-190H」[7]を用いた．

2 人の参加者はマイクを装着し，2 人の間に Kinect を 2 台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．

2.4 表情解析

マイクロソフト社が提供された 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 11 個の AU を学習に用いる．その AU の名称を表 1 に示す．これらの値をマイクから得られるボリュームの値{0,1}と合わせて 12 要素とし，その 2 人分である 24 要素を HMM の学習データとした．

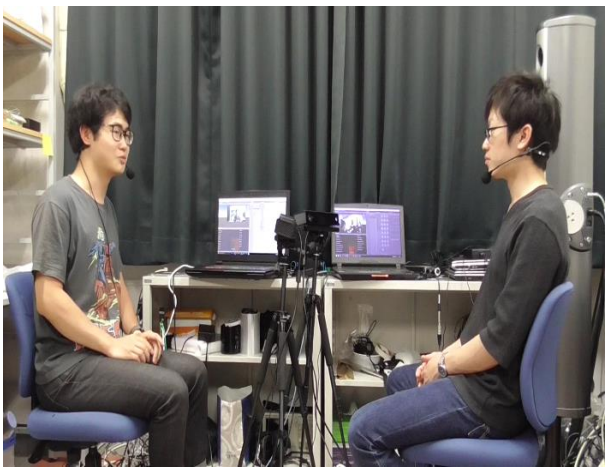

図 2　対話中の様子

Fig.2　State in conversation.

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
LipStretcherRight	唇の右端を横に引く	0,1
RighteyebrowLowerer	右眉を下げる	-1,1
RighteyeClosed	右目を閉じる	0,1

2.5 HMM の学習と結果

得られた学習データをもとに，HMM の学習を行った．学習には機械学習ライブラリ Accord．NET[7]を用いた．

実対話から学習で得られた HMM を用いて，対話エージェントの表情を生成した．作成した表情決定モデルにおける現在のステートでの確率密度関数から算出した出力した様子を図 3 に示す．図の左側が機械学習で得られた話者役のキャラクタであり,右側は聴者役のエージェントである．

生成されたインタラクションから，話者が微笑んでから伝染するようにエージェントが微笑むといった同調的表情の表出がしばしば見ることが出来た．また話者が微笑んだ際，エージェントは必ずしも同調を示すわけではなかったが，これは同調的表情が全観察時間の 20%以下しか表出されないという報告[5]から，実際の人間同士の対話に即していると考えられる．

この結果から，HMM による機械学習より獲得された表情決定モデルを用いることで，同調的表情の再現が可能であることが分かった．

図 3　表情決定モデルによる出力

Fig.3　Output by facial expression determination model

3. 提案手法

人間同士の対話のなかで生じる同調的表情を再現するため，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．次にそのモデルに対して，Kinect 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．そして，出力値によって対話エージェントの顔面の各ボーン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 4 に示す．

3.1 HMM によるインタラクティブ表情決定

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．2.5 節で獲得した表情決定モデルにおける状態iの平均出力を$\boldsymbol{\mu_i}$，共分散行列を$\boldsymbol{\Sigma_i}$，その正規分布を$\boldsymbol{N_i}(\boldsymbol{\mu_i}, \boldsymbol{\Sigma_i})$，ステート$i$にいる確率を$\boldsymbol{S_i}$とする．これらのベクトルまたは行列は，2.5 節で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，

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix} \text{ with sizes } \begin{bmatrix} q \times 1 \\ (p-q) \times 1 \end{bmatrix} \tag{1}$$

が与えられたとする．これに従って

$$\boldsymbol{\mu}_i = \begin{bmatrix} \boldsymbol{\mu}_{i,1} \\ \boldsymbol{\mu}_{i,2} \end{bmatrix} \text{ with sizes } \begin{bmatrix} q \times 1 \\ (p-q) \times 1 \end{bmatrix} \tag{2}$$

$$\boldsymbol{\Sigma}_i = \begin{bmatrix} \boldsymbol{\Sigma}_{i,11} & \boldsymbol{\Sigma}_{i,12} \\ \boldsymbol{\Sigma}_{i,21} & \boldsymbol{\Sigma}_{i,22} \end{bmatrix} \text{ with sizes}$$

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

とする．ここで$\boldsymbol{x}_2$は観測値であり，$\boldsymbol{x}_2 = \boldsymbol{a}$としたときの出力$\boldsymbol{x}_1$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$\overline{\boldsymbol{N_i}}(\overline{\boldsymbol{\mu_i}}, \overline{\boldsymbol{\Sigma_i}})$とすると，

$$\overline{\boldsymbol{\mu_i}} = \boldsymbol{\mu}_{i,1} + \boldsymbol{\Sigma}_{i,12}{\boldsymbol{\Sigma}_{i,22}}^{-1}(\boldsymbol{a} - \boldsymbol{\mu}_{i,2}) \tag{4}$$

$$\overline{\boldsymbol{\Sigma_i}} = \boldsymbol{\Sigma}_{i,11} - \boldsymbol{\Sigma}_{i,12}{\boldsymbol{\Sigma}_{i,22}}^{-1}\boldsymbol{\Sigma}_{i,21} \tag{5}$$

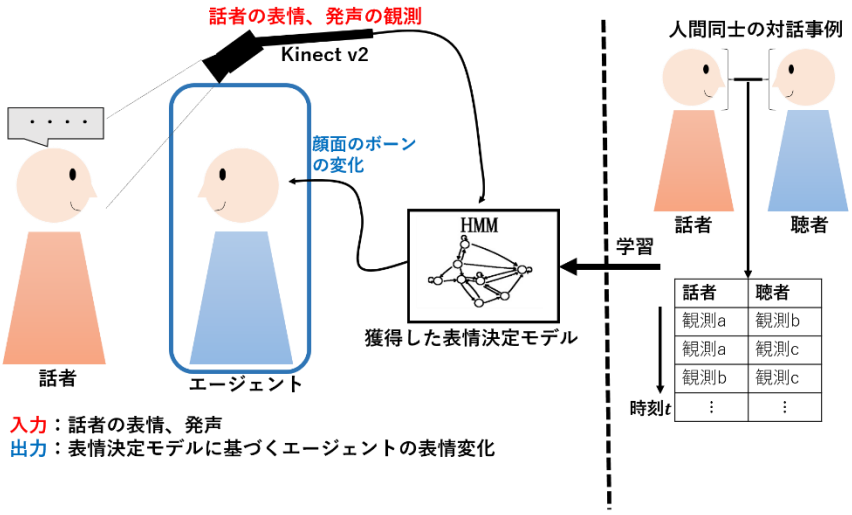

図 4　システムの全体像

Fig.4　Interaction System.

となる．式(4)，(5)より，$\boldsymbol{x}_2 = \boldsymbol{a}$に対して出力$\boldsymbol{x}_1$が選ばれる確率$P(a)$を得る．

$$P(a) = \sum_{k=1}^{i} S_k \overline{N}\left(a \middle| \overline{\boldsymbol{\mu}_i}, \overline{\boldsymbol{\Sigma}_i}\right) \qquad (6)$$

$P(a)$によって得られた出力値を，対話エージェントの顔面の各ボーンに適用することで，エージェントの表情を形成する．本手法の概要を図 5 に示す．

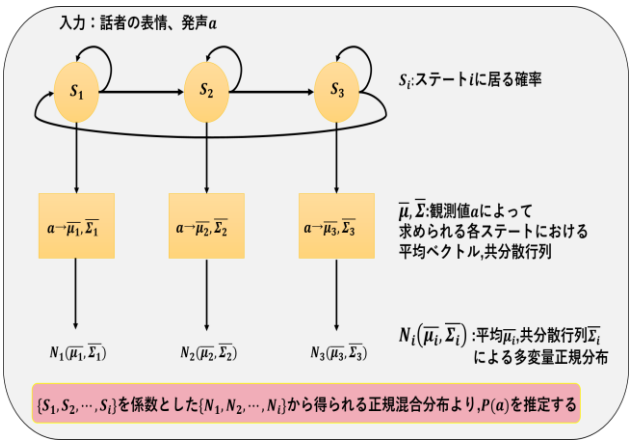

図 5　HMM による表情決定

Fig.5　HMM Based Facial expression Determination.

4.　今後の課題

予備実験では，相手の表情をただ真似るだけでは正しい同調的表情を再現できないことと，実対話事例での同調的表情を機械学習によって再現できることを確認した．

今後の課題としては，提案手法の実装，評価が挙げられる．また，同調的表情を学習した表情決定モデルを生成できたが，生成される笑顔にまだ若干の硬さが見られた．この原因としては，学習に用いた実対話データの不足，データ内の表情変化の乏しさや，出力値をキャラクタの顔面の各ボーンに適用する際のパラメータの最適化が行われていないなどが考えられる．よって，実験の試行回数の増加や，エージェントの作成の仕方も考慮していく必要がある．

謝辞

本研究は科研費（17K17713）の助成を受けたものである。

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- ,インタラクション 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)　Kinect Sensor for Xbox One
https://www.microsoft.com/en-us/store/d/Kinect-Sensor-for-Xbox-One/91HQ5578VKSC
7) 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