

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状態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ーであり、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的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、これは同調的表情的が全観察時間の 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、同調的表情的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的を再現するため、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的を再現する。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的決定

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 μ_i 、共分散行列を Σ_i 、その正規分布を $N_i(\mu_i, \Sigma_i)$ 、状態 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、2.5 節で用いた学習データの要素数 p に依存する。各状態において、

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

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、 $x_2 = a$ 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、

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

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

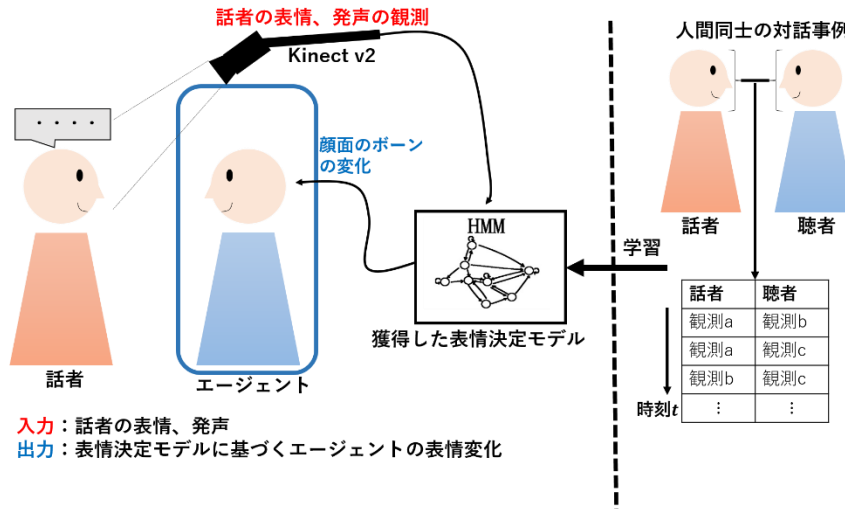


図 4 システムの全体像
Fig.4 Interaction System.

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

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

$P(\mathbf{a})$ 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.

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

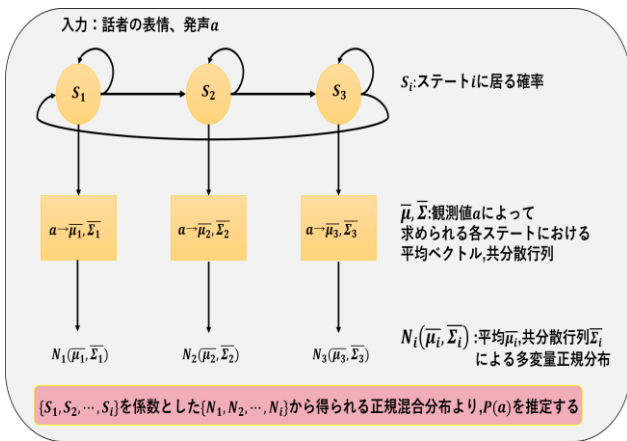


図 5 HMM による表情決定
Fig.5 HMM Based Facial expression Determination.

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的表情を再現できないことと, 実対話事例での同調的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