2004－ICS－137　(8)
2004／10／29

擬人化エージェントを介したコミュニケーションのための表情マッピングの獲得

角所　考[†a)]　　李　立群[††,†††]　　美濃　導彦[†]

Learning How to Mediate User's Facial Expressions by Lifelike Agent for Communication

Koh KAKUSHO[†a)], Liqun LI[††,†††], and Michihiko MINOH[†]

Abstract. 遠隔コミュニケーションにおいて，利用者の表情を擬人化エージェントによって伝達するエージェント媒介型表情コミュニケーションでは，利用者と擬人化エージェントという異なる顔による表情の間で，お互いに同一のもの同士を対応付ける表情マッピングが必要となる．この表情マッピングとして具体的にどのような対応付けが望ましいかは個人の主観に依存することから，本研究ではこれをエージェント媒介型表情コミュニケーションのための利用者とのインタラクションを利用して獲得することを目指す．このとき，獲得した表情マッピングが利用者の満足するものとなるには，表情マッピングの対応付けが，利用者の主観に一致したものとなっていることに加えて，それを実現する際の誤差が，利用者の許容範囲内に収まっている必要がある．本研究ではこれら２点を考慮し，表情マッピングに対して利用者の要求する対応付けと利用者の許容する誤差範囲を，利用者とのインタラクションによって得られる事例から推定すると共に，それらを実現する表情マッピングをRBF（radial basis functions）ネットワークを用いて学習する手法を提案する．

Keywords. 表情認識，顔画像処理，個人適応，事例ベース，コミュニケーション

1. はじめに

電子メールやチャット等，従来のネットワークコミュニケーションはテキストが中心であったが，相手の心理状態を理解するには，表情の伝達が必要となる．そこで，このような目的に対しては，お互いの映像を撮影して相互に伝送する遠隔会議システムが利用されている．しかし，我々の顔は，表情だけではなく，顔の造作も伝えるものであるため，上のような方法を任意のネットワークコミュニケーションで採用すると，表情と共に顔の造作も併せて伝達され，相手の表情の把握のみならず，相手の同定までができてしまうことになる．インターネットの普及により，コミュニケーションの相手が地球規模に広がりつつある一方，コミュニケーションの目的は，カウンセリングなど非常に個人的な事項に関わる領域にまで多様化されてきていることから，今後，ネットワークコミュニケーションの際に匿名性が必要となる状況が増えてくるものと考えられ，そのような状況において，表情と共に顔の造作情報が伝わることは望ましくない．

ネットワークコミュニケーションでの表情伝達における上のような問題への方策として，送受信者の顔を共に擬人化エージェント（あるいはアバタ）による合成顔に置き換えて表情を伝達する方法（以下 “エージェント媒介型表情コミュニケーション” と呼ぶ）が考えられる．実際，電子メールやチャットなどのテキストに基づくネットワークコミュニケーションでは，文字列で表情を表現するフェイスマークをタイプ入力したり，アイコン等で提示された表情の選択肢の中から適当なものを選択することによって表情を伝送する手法が既に利用されているが，送信者とは異なる顔で表情を伝える点でこれは広義のエージェント媒介型表情コミュニケーションであり，このようなコミュニケーションの形態が既に現実のものとなりつつあることを示している．

† 京都大学 学術情報メディアセンター，〒 606-8501 京都市左京区吉田本町
†† 京都大学大学院 情報学研究科 知能情報学専攻，同上
††† 現在，株式会社日立製作所
a) E-mail: kakusho@media.kyoto-u.ac.jp

エージェント媒介型表情コミュニケーションにおいて，上の例のように送信者自身がグラフィカルユーザインタフェース (GUI) 等を利用して送信すべき表情を直接入力する方法は，表情の入力が煩わしいという問題がある．また，聴者側の表情も伝えようとする場合，聴者がこのような方法で自分の表情を入力するということは，あまり現実的に想定できない．しかし，表情伝達を伴う人間同士のコミュニケーションにおいて，聴者の表情変化は話者にとって非常に重要であり，聴者の表情変化を見ながら発話内容を動的に変化させることは，我々が日常しばしば経験することである．したがって，表情の直接入力が不要で，かつ聴者表情の伝達も容易にするためには，送受信者の顔をカメラで撮影し，その表情を擬人化エージェントの表情によって再現する処理（以後 “表情マッピング” と呼ぶ）を導入することが必要となる．

表情マッピングでは，ある人物の実表情に対し，同一の表情をもつ別人（擬人化エージェント）の合成表情を生成することが必要となる．しかし，異なる顔によるどのような表情同士を同一表情とみなすかは個人の主観に依存する．従来手法 [1]- [6] では，これに関する人物の実表情とエージェントの合成表情との対応付けとして，なるべく一般的な観点から同一と思われるものをシステムの設計者が事前に定めた上で，これに基づく表情マッピングを実現しようとする，いわば設計者指向型のアプローチをとっており，実際に表情を表出している人物自身の観点においてその表情と同一の合成表情が生成される保証がない．

これに対し，筆者らは，エージェント媒介型表情コミュニケーションにおいて，表情表情を表出している送受信者が，自分の現在の表情と同じと感じられる表情を擬人化エージェントの合成表情として生成できる利用者指向型の表情マッピングを実現することを目指した研究を行っている．これまでのところ，エージェント媒介型表情コミュニケーションの過程で，擬人化エージェントの表情が自分の表情とは異なると利用者が感じたときに，随時エージェントの表情を GUI によって直接修正できる仕組みを導入し，これによるエージェントの表情の修正内容を事例として蓄積・再利用することにより，利用者の思い通りの合成表情が得られるような表情マッピングを獲得する手法を提案している [7]．しかし，この手法では，蓄積した事例に基づいて表情マッピングを行う際に，蓄積された事例の中から与えられた利用者の表情画像に類似したものを検索し，それらを補間することによって生成すべき合成表情を決定するため，事例の増加と共に類似事例検索のための時間が増加し，処理の高速化が難しい．また，事例の増加と共に利用者の思い通りの表情が生成される可能性は向上するものの，どの時点で利用者による直接修正が不要な完全自動化が実現されるのかがわからないため，表情マッピングの信頼性や安定性の点でも問題がある．

そこで本稿では，これらの問題を解決するためのアプローチを提案する．具体的には，利用者による擬人化エージェントの表情の直接修正結果に基づいて，表情マッピングを学習する際に，事例ベースではなく，RBF(Radial-Basis Function) ネットワークを利用することにより，追加学習が可能で，かつ学習結果の利用が高速な表情マッピングの学習を実現する．さらに，利用者による擬人化エージェントの表情の直接修正結果から，利用者の許容できる表情マッピングの誤差の大きさを推定し，これに基づいて RBF ネットワークを補正することによって，利用者の許容できる誤差範囲内で対応する合成表情を生成されることが保証された表情マッピングを実現する．

以下では，まず 2. において，本研究で実現を目指す表情マッピングの獲得処理の具体的な内容を定式化し，続く 3.，4. でこのための具体的な実現手法として，利用者の要求する対応付けを学習するための RBF ネットワークや，利用者の許容する誤差範囲の推定方法，RBF ネットワークの修正による許容誤差の達成方法について提案する．さらに 5. では実験結果を示し，6. で本研究の今後の課題について議論する．

2. 半自動化システムによる表情マッピングの獲得

2.1 表情マッピングの定式化

表情マッピングの入力となる利用者の実表情は，利用者の顔をカメラで観測することによって得られる．ここで，この顔画像を記述するパラメータ数を N とすると，利用者の任意の実表情は，これらのパラメータを成分とする N 次元ベクトル（これを “実表情ベクトル” と呼ぶ）$\boldsymbol{E} = (x_1, ...x_N)$ によって表現できる．このときのパラメータとしては，画像自体の画素値の並びや，主成分特徴，顔画像処理で抽出された造作特徴など，様々なものが考えられる [7]- [11]．

一方，擬人化エージェントの表情パラメータ数を M とすると，表情マッピングの出力となる擬人化エー

ジェントの合成表情も同様にこれらのパラメータを成分とする M 次元ベクトル（これを “合成表情ベクトル” と呼ぶ）$\boldsymbol{D} = (y_1, ...y_M)$ によって表現できる．このとき，表情マッピングは，実表情ベクトル空間中の各点を，合成表情ベクトル空間中の点に対応付ける写像 $\boldsymbol{D} = f(\boldsymbol{E})$ として表される．

このとき，実表情ベクトル空間中のどの点を，合成表情ベクトル空間中のどの点に対応付けるべきかは，人間の主観に依存する．例えば，図 1(a) のような実表情を，同図 (b) のどちらの合成表情に対応付けるべきかは一概には決まらず，個人によって判断が異なる．そこで，このような対応付けとして，利用者の要求するものを獲得することが本研究の目標となる．

(a) 実表情の例　(b)(a) に対応する合成表情の候補

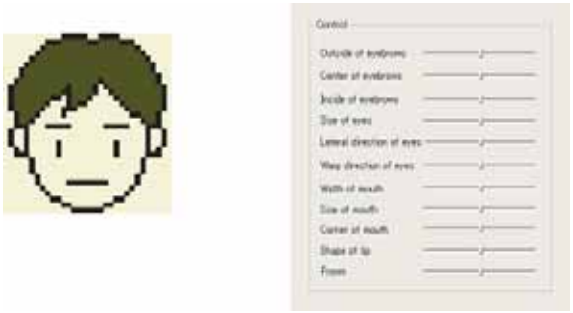
図 1　表情マッピングの主観依存性

2.2　利用者の直接介入による表情マッピングの修正

2.1 で述べた表情マッピングの写像として，利用者の要求するものを f，システムによって実現されるものを $\tilde{f}$ で表すことにすると，$f \neq \tilde{f}$ の場合には，利用者の表情 $\boldsymbol{E}$ に対して，$\tilde{f}$ の与える合成表情 $\tilde{\boldsymbol{D}} = \tilde{f}(\boldsymbol{E})$ と利用者の要求する表情 $\boldsymbol{D} = f(\boldsymbol{E})$ とが一致しない．一般に，どのような自動化システムにおいても，システムの設計ミスやプログラムのバグ，システムの仕様が利用者の要求と異なっている等の理由により，利用者の期待していないような挙動を示す可能性があることから，そのような状況が生じた場合には，利用者がシステムの動作に直接介入してその挙動を自分の望むように修正できる機能を備えていることがシステム設計上重要である．そこで，本研究で考えるエージェント媒介型表情コミュニケーションシステムにおいても，利用者の表情 $\boldsymbol{E}$ に対してシステムの生成する合成表情 $\tilde{\boldsymbol{D}}$ が $\boldsymbol{D} = f(\boldsymbol{E})$ と異なっており，利用者がこの違いを許容できない場合には，利用者が直接介入して $\tilde{\boldsymbol{D}}$ を $\boldsymbol{D}$ に修正できる仕組みを導入する．

このために，擬人化エージェントによるアニメーション生成等の際の表情の設定によく利用されている図 2 のような GUI の利用を考える．この GUI では，擬人化エージェントの合成表情を定める各パラメータ (この例では 11 個) に対応するスライダバーが用意されており，これを動かすことにより，利用者がエージェントの表情を直接修正できる．

以下では，エージェント媒介型表情コミュニケーションで行われる t 番目の表情マッピングにおける $\boldsymbol{E}, \boldsymbol{D}, \tilde{\boldsymbol{D}}$ をそれぞれ $\boldsymbol{E}_t, \boldsymbol{D}_t, \tilde{\boldsymbol{D}}_t = \tilde{f}(\boldsymbol{E})$ とする．ここで，上のような表情修正の仕組みを前提とすれば，$\tilde{\boldsymbol{D}}_t$ は $\boldsymbol{E}_t$ に対してシステムが生成する合成表情ベクトル，$\boldsymbol{D}_t$ は $\tilde{\boldsymbol{D}}_t$ を利用者が直接介入によって修正した結果に相当する．利用者が表情修正を行わなかった場合は，$\boldsymbol{D}_t = \tilde{\boldsymbol{D}}_t$ の状況に相当する．



(a) 擬人化エージェントの例　(b) 表情修正用スライダバー

図 2　表情修正のための GUI

2.3　表情修正を通じた表情マッピングの改善

上のような GUI を導入することによって，利用者が擬人化エージェントの合成表情を直接修正できるとしても，その修正内容が，それ以降のシステムの挙動に全く反映されなければ，利用者はエージェント媒介型表情コミュニケーションを行っている間，同じような表情修正を何度も実行する必要がある．そのようなエージェント媒介型表情コミュニケーションシステムは，利用者にとってシステムを利用するための労力が非常に大きいのに加え，合成表情を修正し続ける意欲も削がれてしまうことになる．

このようなことから，2.2 で述べたような利用者の直接介入による表情修正を導入する場合には，表情修正の結果がそれ以降の表情マッピングに反映され，表情修正の度に，表情マッピングの内容が利用者の要求に近づいていくような特性を実現する必要がある．このためには，表情修正を事例とした追加学習が必要となる．そこで，本研究では，2.1 で研究目標として述べた表情マッピングの獲得にあたって，このような利

用者の直接介入による表情修正結果を事例とした追加学習を，実現すべき課題の１つとして考える．

2.4 表情マッピングの許容誤差

前述のような追加学習を実現できれば，システムによる表情マッピングの写像 $\tilde{f}$ は，事例の増加と共に，利用者の要求する写像 f に近づいていくことになる．このとき，$\tilde{f}$ に対していつ利用者の直接介入による表情修正が完全に不要となるかは，$\tilde{f}$ の生成する合成表情 $\tilde{\boldsymbol{D}}_t$ と利用者の要求する合成表情 $\boldsymbol{D}_t$ の間に，利用者がどこまでの大きさの誤差を許容するかによって決まる．この誤差の大きさは，利用者が，どのような表情の相違を無視できる，あるいは無視できないと考えるかという利用者自身の主観に依存する．以上のことから，$\tilde{f}$ に利用者の直接介入が不要となるための条件は，次のように定めることができる．

擬人化エージェントの合成表情 $\tilde{\boldsymbol{D}}$，$\tilde{\boldsymbol{D}}'$ に対し，利用者がこれらの違いが無視できる程小さく，同じ表情とみなせると判断するとき，２つの表情は，利用者の観点において同じ表情カテゴリに属するといえる．このように，利用者が互いに同じであるとみなす合成表情のカテゴリを $S_k(k=1,...K)$ で表すことにする．ただし，K はカテゴリの数であり，K が大きい程，利用者は合成表情の違いをより細かく捉えていることになる．

ここで，エージェント媒介型表情コミュニケーションの任意の過程において，利用者が表出する可能性のある実表情の集合を $R=\boldsymbol{E}_t|t=0,1,2,...$ で表すと，$\boldsymbol{E}_t \in R$ に対する $\boldsymbol{D}_t$，$\tilde{\boldsymbol{D}}_t$ が常に同じ表情カテゴリ S_k に含まれるとき，$\boldsymbol{D}_t$，$\tilde{\boldsymbol{D}}_t$ の違いは利用者にとって無視できるほど小さいとみなせることになる．すなわち，システムによる表情マッピングの写像 $\tilde{f}$ が以下の条件を満足するとき，$\tilde{f}$ は，利用者の直接介入による表情修正の必要がないことが保証される．

$$\forall E_t \in R, \quad \exists k\, \boldsymbol{D}_t \in S_k,\, \tilde{\boldsymbol{D}}_t \in S_k \qquad (1)$$

この様子を，実表情ベクトル，合成表情ベクトルが共に１次元の場合を例として図 3 に示す．図で横軸が実表情のパラメータ，縦軸が合成表情のパラメータであり，これらの間の表情マッピングとして，利用者の要求する写像 f および，システムによる写像の例 $\tilde{f}_1$，$\tilde{f}_2$ を示してある．横軸上の領域 R_1，R_2，R_3 は R の部分集合であり，$R_1 \cup R_2 \cup R_3 = R$ である．この図において，$\tilde{f}_1, \tilde{f}_2$ と f とのユークリッド距離は全領域に渡って同じであるが，$\tilde{f}_1$ が式 (1) の条件を満足しているのに対し，$\tilde{f}_2$ は満足していない．この図からわかるように，写像 $\tilde{f}$ が式 (1) の条件を満足しているかどうかを判断するには，$\tilde{f}$ と f の間の誤差に加えて，$S_k(k=1,...K)$ を知る必要がある．そこで，本研究では，これを実現すべきもう１つの課題として考える．

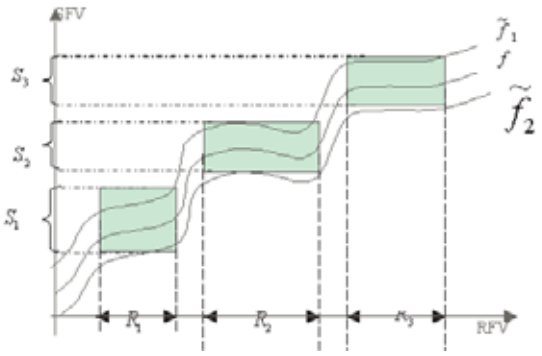


図 3　表情マッピングの許容誤差と表情カテゴリの関係

3. 表情マッピングの追加学習

3.1 ユーザによる表情修正を通じた正事例の獲得

2.3 で述べたように，本研究では，利用者の直接介入による表情修正結果を事例とした追加学習の実現を目標の１つとする．このための事例は，以下のようにして獲得することができる．

利用者の行うエージェント媒介型表情コミュニケーションにおいて，$\boldsymbol{D}_t = \tilde{\boldsymbol{D}}_t$，すなわちシステムの生成した合成表情に対して利用者が直接介入による表情修正をしなかった場合の $\boldsymbol{E}_t$ と $\tilde{\boldsymbol{D}}_t$，および $\boldsymbol{D}_t \neq \tilde{\boldsymbol{D}}_t$ のときの $\boldsymbol{E}_t$ と $\boldsymbol{D}_t$ は，それぞれ利用者が要求する表情マッピングの写像 f の正事例となることから，これらをエージェント媒介型表情コミュニケーションの過程を通じて獲得する．このとき，t 回の表情マッピングが行われた時点で，それと同数の t 個の正事例が獲得されることになる．以下では，このようにして得られた正事例の中で，τ 番目の正事例を構成する実表情ベクトルと合成表情ベクトルをそれぞれ $\mathcal{E}_\tau, \mathcal{D}_\tau$ で表すこととし（$1 \leqq \tau \leqq t$），それらの t 個の集合を，$P_t^E = \{\mathcal{E}_1, ..., \mathcal{E}_t\}, P_t^D = \{\mathcal{D}_1, ..., \mathcal{D}_t\}$ で表す．

3.2 正事例に基づく表情マッピングの学習

前節で得られた正事例から利用者の要求する表情マッピング f を学習する際に，なるべく少ない事例から f を正しく推定するには，f の性質を的確に表現したバイアスが必要である．表情マッピングにおい

て，実表情は顔面筋によって作り出されるものであるから，不連続な変化をすることは不可能であり，互いに似通った実表情は，互いに似通った合成写像に対応すべきと考えられることから，学習にあたっては，事例間を滑らかに補間するようなバイアスの利用が望ましいと考えられる．一方，2.3 で述べたような理由から学習を追加的に行うには，ある事例の学習結果が，その後の事例の追加の影響を受けない必要がある．

以上の要求を満足するような学習手法として，本研究では，RBF ネットワークを利用する．このネットワークは，図??に示すように，入力層，隠れ層，出力層の 3 層から成り，隣接層のユニット間はフィードフォーワード型の完全結合で結ばれている．入出力層のユニット数はそれぞれ実表情ベクトル，合成表情の次元数と同じ N，M であり，入力層の各ユニットの入力値および出力層の各ユニットの出力値はそれぞれ，実表情ベクトルおよび合成表情ベクトルの各成分の値に対応している．ここで隠れユニットの数を Q とすると，入力層に実表情ベクトル $\boldsymbol{E}$ が入力されたときの q 番目の隠れユニットの出力値 $h_q(\boldsymbol{E})(q = 1, \cdots, Q)$ は，そのユニット固有の定数ベクトル $\boldsymbol{C}_q$ に基づいて次式のように算出される．

$$h_q(\boldsymbol{E}) = \exp(-\frac{-\|\boldsymbol{E} - \boldsymbol{C}_q\|^2}{\sigma_q^2}) \qquad (2)$$

さらに，このときの m 番目の出力ユニットの出力値 $y_m(\boldsymbol{E})$ は，q 番目の隠れユニットから m 番目の出力ユニットへの結合荷重を w_{qm} として次式で定義される．

$$y_m(\boldsymbol{E}) = \sum_{q=1}^{Q} w_{qm} h_q(\boldsymbol{E}) \qquad (3)$$

結合荷重 w_{qm} の値は，入力ユニットに事例中の実表情ベクトル $\mathcal{E}_\tau (1 \leqq \tau \leqq t)$ が与えられたときに，出力ユニットの値が同じ事例中の対応する合成表情ベクトル $\mathcal{D}_\tau$ に一致するように，$Q = t$，$\boldsymbol{C}_q = \mathcal{E}_q \, (1 \leqq q \leqq Q = t)$ として，以下の値を最小とする値として算出する．

$$W = H^{-1} Y \qquad (4)$$

ここで W は qm 成分を w_{qm} とする $t \times M$ 行列，Y は τ 番目の行をベクトル $\mathcal{D}_\tau$ とする $P \times M$ 行列，H は τq 成分を $h_q(\mathcal{E}_\tau)$ とする $t \times t$ 行列で，H^{-1} は，H の逆行列である．

なお，式 (2) の定数 σ_q の値が小さすぎる場合には，事例間が滑らかに補間されない一方，大きすぎる場合は，事例の追加によって，他の事例の学習結果が影響を受ける．これらのことから，文献 [12] にしたがって，この値を $\mathcal{E}_q$ から 1,2 番目に近い事例までの距離の平均値によって定めた．

4. 表情マッピングの許容誤差の推定と実現

4.1 負事例の獲得

2.4 で述べたように，本研究のもう 1 つの目標は，表情マッピングに対する利用者の許容誤差を推定し，その範囲内の誤差の表情マッピングを実現することである．

2.4 の議論に基づけば，利用者が $\tilde{\boldsymbol{D}}_t$ を $\boldsymbol{D}_t$ を同じ表情であるとみなすのは，$\boldsymbol{D}_t \in S_k$ となる合成表情カテゴリ S_k に対して $\tilde{\boldsymbol{D}}_t \in S_k$ の場合であり，同じ表情ではないとみなすのは，$\tilde{\boldsymbol{D}}_t \in S_l \neq S_k$ の場合である．このうち後者の場合の $\boldsymbol{D}, \tilde{\boldsymbol{D}}_t$ は f の正事例，負事例に相当し，両者の間に S_k と S_l の境界が存在するはずである．

以上のことから，エージェント媒介型表情コミュニケーションにおいて，3.1 で述べたような方法によって，利用者の実表情 $\boldsymbol{E}_t$ に対して $\tilde{\boldsymbol{D}}_t \neq \boldsymbol{D}_t$ である場合の $\boldsymbol{D}_t$，$\tilde{\boldsymbol{D}}_t = \boldsymbol{D}_t$ である場合の $\tilde{\boldsymbol{D}}_t$ を f の正事例として獲得することに加え，さらに $\tilde{\boldsymbol{D}}_t \neq \boldsymbol{D}_t$ である場合の $\tilde{\boldsymbol{D}}_t$ を f の負事例として獲得し，$\boldsymbol{D}_t, \tilde{\boldsymbol{D}}_t$ の差に基づいて利用者の観点における合成表情カテゴリを求めることにより，利用者の許容する表情マッピングの誤差範囲を推定する．以下では，$\mathcal{E}_\tau$ と共に τ 番目の負事例を構成する合成表情ベクトルを $\tilde{\mathcal{D}}_\tau$ で表す．ただし，$\tilde{\boldsymbol{D}}_\tau = \boldsymbol{D}_\tau$ であるような τ に対しては負事例が獲得されないので，このような τ については $\tilde{\mathcal{D}}_\tau = \boldsymbol{0}$ と定める．

4.2 正事例と負事例に基づく許容誤差の推定

4.1 で述べた方法では，利用者の実表情 $\boldsymbol{E}_t$ に対する合成表情 $\tilde{\boldsymbol{D}}_t$ が利用者の要求する合成表情 D と異なる表情カテゴリに属するときに，利用者の直接介入による表情修正を通じて f の正事例 $\boldsymbol{D}_t$ と負事例 $\tilde{\boldsymbol{D}}_t$ を獲得することにより，互いに異なる表情カテゴリに属する合成表情ベクトルを得ることはできるが，$\tilde{\boldsymbol{D}}_t$ が $\boldsymbol{D}_t$ と同じ表情カテゴリに属する場合には $\boldsymbol{D}_t$ が得られないので，同じ表情カテゴリに属する合成表情ベクトルを獲得することはできない．

一方，3.2 で述べたように，利用者の実表情は連続的に変化し，かつ似通った実表情に対して対応付けられる合成表情はやはり似通っていると考えられるので，２つの合成表情が近い場合には，そうではない場合と比べて，両者が同じ表情カテゴリに属する可能性が高いと考えられる．また，利用者にとって同じ表情カテゴリに属する合成表情を，異なるカテゴリに属するものと考えても，許容誤差がより小さく見積もられることになるだけであり，利用者の要求する許容誤差の達成という点では問題がない．

以上から，3. の処理による表情マッピング f の学習がある程度進んだ時点で (この時点を $t = T$ とする)，それまでに獲得された正事例を構成する全ての合成表情 $\mathcal{P}_T^D$ を，それらの間の距離に基づいて，複数の表情カテゴリにクラスタリングし，さらに，このクラスタリングによって同じ表情カテゴリに分類された合成表情ベクトルの中で，同じ実表情に対する合成表情の正事例と負事例となるものが見つかった場合は，両者が別々の表情カテゴリになるようにカテゴリを再分割するという２段階の処理を実行する．

このうちの最初のクラスタリングには，最も単純な方法として，$\mathcal{D}_\tau \in P_T^D$ に対して次式で定義されるしきい値 θ を求め，２つの合成表情間の距離がこのしきい値より小さいものを同じ表情カテゴリに，そうでないものを異なる表情カテゴリに属するようにクラスタリングする単連結法を用いる．

$$\theta = \max_{\tau=1,\cdots,t} \delta_\tau \tag{5}$$

$$\delta_\tau = \min_{\substack{\tau'=1,\cdots,t \\ \tau' \neq \tau}} \{\|\mathcal{D}_\tau - D_{\tau'}\|\} \tag{6}$$

このようにして得られた合成表情ベクトルのクラスタ数を K' とする．このとき，k 番目の合成表情ベクトルのクラスタ $S_k(k = 1, \cdots, K')$ に対して，もし，$\mathcal{D}_\tau \in S_k$ かつ $\tilde{\mathcal{D}}_\tau \in S_k$ である場合には，S_k を，$\mathcal{D}_\tau$ と $\tilde{\mathcal{D}}_\tau$ の間で２つのクラスタに分割する．

以上の処理の結果得られたクラスタを利用者の主観による合成表情カテゴリであるとみなし，このときのクラスタ数を K とする．さらに，このときの合成表情クラスタ $S_k(k = 1, \cdots, K)$ に対応する実表情ベクトルのクラスタを $R_k(R_k = \{\mathcal{E}_\tau | D_\tau \in S_k\})$ で表す．

4.3 許容誤差の実現のための RBF ネットワークの修正

4.2 で得られた合成表情カテゴリ $S_k(k = 1, ...K)$ に対して式 (1) を満足するために，3. で得た RBF を次のように修正する．

まず，RBF ネットワークの隠れユニット数 Q をカテゴリ数と同じ $Q = K$ と定め，q 番目の隠れユニットに対する定数ベクトル $\boldsymbol{C}_q$ を，次式のように q 番目のカテゴリに属する正事例の重心にとる．

$$\boldsymbol{C}_q = \frac{1}{I_q} \sum_{\mathcal{E}_\tau \in R_q} \mathcal{E}_\tau \tag{7}$$

ただし，I_q は，$\mathcal{E}_\tau \in R_q$ を満たす実表情ベクトル $\mathcal{E}_\tau$ の個数である．

ここで，次式を満足する任意の実表情ベクトル $\boldsymbol{E}$ を，$\boldsymbol{E} \in R_q$ と分類する．

$$\|\boldsymbol{E} - \boldsymbol{C}_q\| \leqq d_q \tag{8}$$

$$d_q = \max_{\mathcal{E}_\tau \in R_q} \|\mathcal{E}_\tau - \boldsymbol{C}_q\| \tag{9}$$

ただし，$\|\mathcal{E}_{\tau'} - \boldsymbol{C}_q\| \leqq d_q$ かつ $\mathcal{E}_{\tau'} \notin R_q$ なる $\mathcal{E}_{\tau'} \in P_T^E$ が存在するとき，$\mathcal{E}_\tau \in R_q$ の中で，$\mathcal{E}_{\tau'}$ に最も近い 2 点 $\mathcal{E}_{\tau 1}, \mathcal{E}_{\tau 2}$ を求め，新たなカテゴリ $R_{q'}$ を設けて $\mathcal{E}_{\tau 1} \notin R_q, \mathcal{E}_{\tau 2} \notin R_q$ かつ $\mathcal{E}_{\tau 1} \in R_{q'}, \mathcal{E}_{\tau 2} \in R_{q'}$ とする．

このとき，カテゴリの増加に伴って RBF ネットワークにも q' 番目の新たな隠れユニットを設け，対応する定数ベクトル $\boldsymbol{C}_{q'}$ を次式で定める．

$$\boldsymbol{C}_{q'} = \frac{\mathcal{E}_{\tau 1} + \mathcal{E}_{\tau 2}}{2} \tag{10}$$

5. 実　　験

5.1 実表情ベクトルと合成表情ベクトル

上記で述べたような処理の有効性を評価するために，実験を行った．実表情ベクトルには利用者の顔画像の濃淡データをそのまま利用し，合成表情ベクトルには図 2 に示したような 11 個のパラメータから成る合成表情を利用した．表情マッピングの入力となる実表情ベクトルは，利用者の表情の変化のみを的確に捉えるものでなければならないが，上のような利用者の顔画像では，表情変化以外に，カメラに対する顔の位置や向きの変化によっても画像が変化する．そこで，この影響を取り除くために，利用者がタイプ入力した発言が相手側では合成音声で提示されるようなエージェント媒介型表情コミュニケーションを想定して，利用者がヘッドセットを装着しているものとし，このヘッド

セットに付けられた発光ダイオードのマーカを利用してカメラに対する利用者の顔の位置・向きに依存しない顔画像を獲得した．

この処理ではまず，発光ダイオードのマーカをヘッドセットに 3 点設置し，これらの位置を顔画像獲得用とは別にステレオカメラによって計測し，基準となる位置・向きに対する利用者の頭の移動量と回転角を得る．次に，その状態で得られた利用者の顔画像を，顔を平面と考えて，計測された移動量と回転角を相殺するように 3 次元空間中で並進・回転させてから画像平面上に再投影することによって，基準となる位置・向きにおける利用者の顔画像を近似的に復元する．さらにデータ量を削減するために，この顔画像を 64×64 画素に解像度を落として実表情ベクトルとして利用する．以上のような処理で得られた顔画像の例を図 4 に示す．

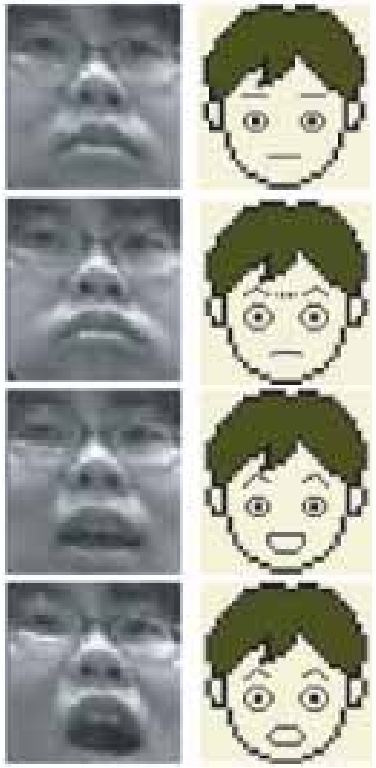
(a) 入力画像　　(b) 実表情ベクトル

図 4　実表情ベクトル

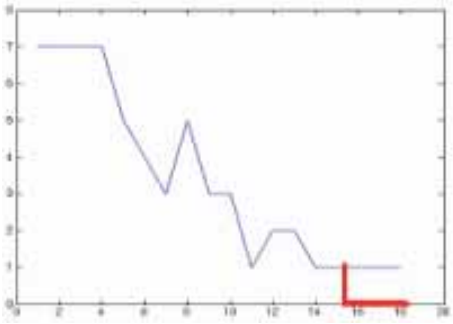
5.2 実験結果

前節のような実表情ベクトルと合成表情ベクトルに基づいて，被験者に図 5 のような 4 組の実表情と合成表情からなる表情マッピングをシステムに獲得させるようにインタラクションしてもらう実験を実施し，そのインタラクションの各段階で利用者がスライダバーを動かした大きさ，および利用者の不満足度を 5 段階（4：極めて不満足〜0：完全に満足）で答えてもらったときの値をグラフ化したものを同図 (b)，(c) に示す．

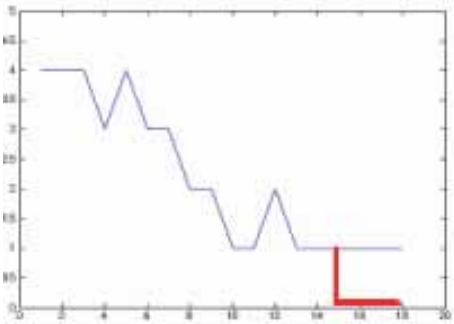
図 (a)，(b) 共に，小刻みな増減はあるものの，インタラクションを通じて下降する傾向を示している．また，この小刻みな増減についても，太線で示された箇所において，4. で述べた処理を行った以降は，共に値が零となっている．以上から，本稿で目標とした 2 つの課題である追加学習による表情マッピングの改善，ならびに，表情マッピングの許容誤差推定とその実現，の 2 つが概ね実現できていることが確認できた．



(a) 獲得すべき表情マッピング



(b) スライダバーの修正量



(c) 利用者の不満足度

図 5　実験結果

5.3 シミュレーションによる多様な状況の検証

前節では，4. で述べた処理によって利用者の許容誤差の推定と達成ができたが，実際のところこれがどの程度成功するかは，利用者の要求する表情マッピングのカテゴリの数や近接性によっても異なると考えられる．しかし，様々なカテゴリ数や近接度に相当する状況を全て被験者によって実験することは現実的ではないので，これについてはシミュレーションに基づく実験を行った．

上で述べたような実表情ベクトル，合成表情ベクトル空間中で事例点やその間の対応をランダムに生成した上で，そのような表情マッピングを要求するユーザとの計算機とのインタラクションを計算機でシミュレーションした．このときの 4. の処理の前後でのシステムによる表情マッピングの正解率の変化を表 1 に示す（上段：処理前，下段:処理後）．ただし，η は，隣接カテゴリ間の中心同士の距離に対する境界同士の距離の比を示し，η が小さいほど，隣接カテゴリ同士が互いに接近していることを表す．

表 1　表情カテゴリ数と近接性を変えた場合の結果

カテゴリ数	4	6	20
$\eta < 0.2$	0.7745	0.7376	0.5430
	0.8930	0.8301	0.7411
$0.2 < \eta < 0.5$	0.7939	0.7068	0.5844
	0.9096	0.8837	0.85
$0.5 < \eta < 0.8$	0.8283	0.7425	0.7248
	0.9544	0.9352	0.9124
$0.8 < \eta < 1$	0.9282	0.8735	0.9041
	0.9833	0.9754	0.9623

一般的に，カテゴリ数が増えるに従い，また，カテゴリ間の近接性が高まるに従い，それに基づく表情マッピングの許容誤差は小さくなり，これを実現することが難しくなる．表 1 においても，カテゴリ数の増加および η の減少と共に，正解率が低くなっており，このことを裏付けている．しかしながら，そのいずれの場合においても，4. で述べた処理によって，正解率が向上していることから，この処理の有効性が示されたといえる．

6. ま と め

本稿では，エージェント媒介型表情コミュニケーションのための表情マッピングにおいて，利用者の主観にあった表情の対応付けを獲得することを目指した．このような処理を実現する際の具体的な課題として，システムの利用過程における表情マッピングの改善，および利用者の許容する許容誤差の推定とその実現の 2 つを挙げ，半自動的なエージェント媒介型表情コミュニケーションの過程で得られる正事例，負事例を基に，これらを実現するための手法を提案した．

本稿で示した実験結果は，本稿で提案したアプローチの有効性をある程度示すものと考えるが，4. で示したクラスタリング手法や RBF ネットワークの修正方法にはまだ不完全な部分が残っており，表情のバリエーションを増やした場合など，多様な状況を考えた場合には，問題を生じる場合も存在する．今後は，これらの点を改善し，手法の完成度および一般性を高める必要があると考えている．

文　献

[1] H.Harashima,F.Kishino: Intelligent Image Coding and Communications with Realistic Sensations – Recent Trends –; *IEICE Trans. on Information Systems*, Vol.E-74, No.6 (1991).

[2] 海老原, 棚沢, 大谷, 中津, 岸野: 美術解剖学に基づいた仮想変身システムのための実時間人物表情再現; 信学論 (D-II),vol.J81-D-II, No.5, pp.841-849(1998).

[3] K.Ohba, G.Clary, T.Tsukada, T.Kotoku & K.Tanie: “Facial Expression Communication with FES,” *Proc. of Int. Conf. on Pattern Recognition* (1998).

[4] T.Otsuka, J.Ohya: “Converting Facial Expressions Using Recognition-Based Analysis of Image Sequences,” *Proc. Asian Conf. on Computer Vision* (1998).

[5] M.Cavazza, R.Earnshaw, N.Magnenat-Thalmann, D.Thalmann: “Motion Control of Virtual Humans,” *IEEE Computer Graphics and Applications*, Vol.18, No.5 (1998).

[6] L.Williams: “Performance-driven facial animation,” *Proc. ACM SIGGRAPH'90*, pp.235-242 (1990).

[7] 近藤 崇, 角所 考, 美濃導彦: “対話的に獲得される事例に基づく行為者指向の顔メディア変換,” システム制御情報学会論文誌, Vol.14, No.6, pp.308-315 (2001).

[8] K.Mase: “Recognition of Facial Expressions for Optical Flow,” *IEICE Trans. on Information Systems*, Vol.E-74, No.10 (1991).

[9] 松野, 李, 辻: “ポテンシャルネットと KL 展開を用いた顔表情の認識,” 信学論 (D-II),vol.J77-D-II, No.8, pp.1591-1600 (1994).

[10] 坂口竜己, 森島繁生: “画像の二次元 DCT を利用した実時間表情認識,” 電子情報通信学会論文誌 D-II, vol.J80-D-II,no.6,pp1547-1554 (1997).

[11] J.Haddadnia, K.Faez, M.Ahmadi: “N-Feature Neural Network Human Face Recognition,” *Proc. International Conference on Vision Interface*, pp.300-307 (2002).

[12] J.Moody, C.Darken: “Fast Learning in Networks of Locally-Tuned Processing Units,” *Neural Computation*, Vol.1, No.2, pp 281-294 (1989).