

遠隔コミュニケーションのための感性ロボットと人の表情対応付け

Mapping Human Face to Robot's Face for Tele-Communication

加納 政芳[†] 後藤 みの理[‡] 加藤 昇平[‡] 伊藤 英則[‡]
 Masayoshi Kanoh Minori Gotoh Shohei Kato Hidenori Itoh

1 はじめに

近年、携帯電話やインターネットの普及により、遠隔地間でコミュニケーションを行う機会が増えてきた。遠隔コミュニケーションにおける問題点として、非言語情報の不足があげられる。非言語情報とは、ジェスチャーや表情などの言語以外の情報である。人同士が対面して行うコミュニケーションでは、言語情報のみならず非言語情報のやりとりも重要な意味を持つ[1]。非言語情報が言語情報を補うことによって、互いの感情を正確に理解することができ、円滑なコミュニケーションを可能にする。特に顔における表情は、非言語情報を効果的に伝達することのできるメディアとして重要な役割を果たしている。例えば、我々は表情を通じて自分の心理状態を伝えることができ、その一方で、表情から相手の心理状態を読み取ることができる。

遠隔コミュニケーションにおける非言語情報を補うツールとして、テレビ電話があげられる。テレビ電話は、音声情報のみである従来の電話に比べて、顔の表情や視線、仕草などを相手に伝えることができる。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。テレビ電話のように、互いの映像を相互に伝送するシステムを、任意のネットワークコミュニケーションで採用すると、相手の心理状態の把握のみならず、相手の同定までできてしまう。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今日、不特定多数の相手に自分の顔をさらすことはセキュリティ上好ましくない。また、遠隔コミュニケーションの目的は、カウンセリングなどの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ため、今後、遠隔コミュニケーションの際に匿名性が必要となる状況も増えてくると考えられる。このため、相手に顔の造作情報を伝えることなく、表情のみを伝達できるようなシステムが必要である。本研究では、匿名性のある遠隔コミュニケーションの媒体として、感性会話ロボット ifbot [2, 3] を用いる。ロボットは実体を持っているため、立体感や実在感、空間共有感を出すことができ、それによって感性的因子をより良く表出できると考えられる[4]。また、ifbot は複雑な表情表出機構によって様々な表情を表出できるため、感性的因子の表出に向いているといえる。ここで、人と同様の顔構造を持ったロボットの表情と人の表情をマッピングするのであれば、抽出した人の表情パラメータをそのままロボットの表情制御パラメータとして用いることができる[4]。しかし、ifbot はエンタテイメント性を重視した漫画的な顔構造を持っており、人の顔構造とは異なるため、この手法を用いることができない。そこで本稿では、ニューラルネットワークを用いた ifbot と人の表情の対応付け手法を提案する。

2 ifbot

図1に ifbot の概観を示す。ifbot は身長 45cm、重さ 9.5kg であり、2つの腕を有し、車輪によって移動する。

図2に ifbot の表情表出機構の概要を示す。ifbot は表情表出のために、10のモータおよび101のLEDを持つ。モータは、首を2軸(同図 θ_{N1}, θ_{N2})、左右の目を2軸(同図 $\theta_{E1}^{(L)}, \theta_{E1}^{(R)}, \theta_{E2}^{(L)}, \theta_{E2}^{(R)}$)、左右のまぶたを2軸(同図 $\theta_{L1}^{(L)}, \theta_{L1}^{(R)}, \theta_{L2}^{(L)}, \theta_{L2}^{(R)}$)

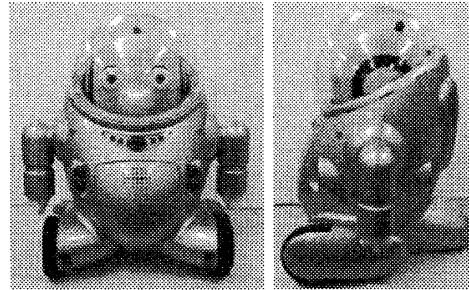


図1 ifbot の概観

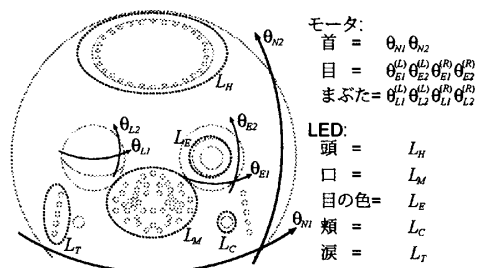


図2 ifbot の表情表出機構

表1 パラメータに使用する AU

AU 番号	顔部位の動き
1	眉の内側を上げる
2	眉の外側を上げる
4	眉を下げる
5	上まぶたを上げる
6	頬を持ち上げる
7	まぶたを緊張させる
12	唇両端を引き上げる
15	唇両端を引き下げる
23	唇を固く閉じる
27	口を大きく開く

$\theta_{L1}^{(R)}, \theta_{L2}^{(R)}$ の方向へ動作させる。LED は、頭部(同図 L_H)、口部(同図 L_M)、目部(同図 L_E)、頬部(同図 L_C)、涙部(同図 L_T)、耳部に配置され、頭部3色(橙、緑、赤)、口部1色(橙)、目部3色(緑、赤、青)、頬部1色(赤)、涙部1色(青)、耳部1色(橙)を発色する。これらの機構により、ifbot は表情豊かなコミュニケーションを行うことができる。

3 人表情の特徴量抽出

3.1 特徴点抽出

まず、人の顔画像から、顔表情を構成する各部位の特徴点を抽出する。特徴点の抽出には、「感性擬人化エージェントのための顔情報処理システム」の顔認識システム[5]を利用する。このシステムでは、眉、目、鼻、口に関して77個の特徴点を認識することができる。本稿では、これらの特徴点から、表1の10個のAUに関わる特徴点23個を使用する。対応する23個の特徴点を図3に示す。これらの点を制御することで、多様な表情を表現することができる。

つぎに、田中らの手法[6]を参考に、抽出した特徴点に

[†] 中京大学, Chukyo University

[‡] 名古屋工業大学, Nagoya Institute of Technology

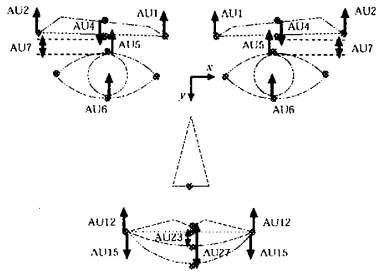


図3 AUと特徴点の対応

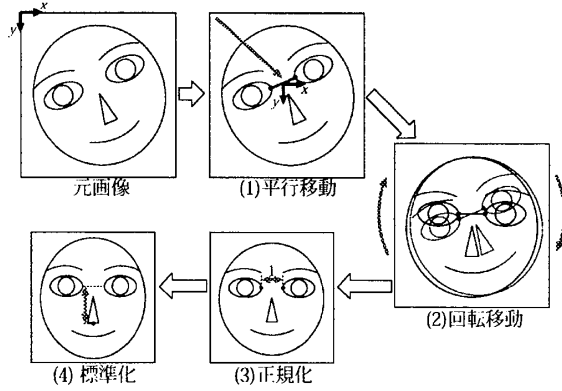


図4 特徴点抽出手順

以下の処理を施し、特徴点を補正する(図4)。

1. 平行移動: 両目頭の中点が原点となるように、平行移動する
2. 回転移動: 両目頭の2点を結ぶ直線の傾きが0になるように、原点を中心として回転移動を行う
3. 正規化: 各座標の値を目頭間の距離で除算する
4. 標準化: 原点から鼻先までの距離を利用し、縦幅を同一人物の平静表情のデータに統一する

こうして得られた座標点を特徴点として用いる。

3.1.1 人表情パラメータの決定

特徴点を用いて、表情の特徴量となる AU パラメータを決定する。各表情の位置情報において、それぞれの AU に関わる特徴点の平静表情からの変位に着目し、AU パラメータを以下のように定義する。

$$H_{AU_i} = \alpha_{AU_i} \frac{d_{AU_i}}{D_{AU_i}^{\max}} \quad (1)$$

$$(i = 1, 2, 4, 5, 6, 7, 12, 15, 23, 27).$$

ここで、 α_{AU_i} は $[0, 1]$ のパラメータ、 d_{AU_i} は AU_i の判定に使用される特徴点の平静表情からの変化量、 $D_{AU_i}^{\max}$ は変化量の最大値である。また、 $AU_1, AU_2, AU_4, AU_5, AU_6, AU_{12}, AU_{15}$ は特徴点の y 軸方向への変化量、 AU_7, AU_{23}, AU_{27} は特徴点間の距離の変化量を用いている。AU パラメータのうち、眉および目に関わる AU は左右独立に変化する。これを考慮し、人の表情特徴を表すパラメータ(人表情パラメータ)を以下のように定義する。

$$S_{\text{man}} = (H_{AU_1}^{(L)}, H_{AU_1}^{(R)}, H_{AU_2}^{(L)}, H_{AU_2}^{(R)}, H_{AU_4}^{(L)}, H_{AU_4}^{(R)}, H_{AU_5}^{(L)}, H_{AU_5}^{(R)}, H_{AU_6}^{(L)}, H_{AU_6}^{(R)}, H_{AU_7}^{(L)}, H_{AU_7}^{(R)}, H_{AU_{12}}, H_{AU_{15}}, H_{AU_{23}}, H_{AU_{27}})^T \quad (2)$$

4 ifbot の表情パラメータ

ifbot の表情パラメータは、10 のモータ出力値と各部位の LED 出力値から以下のように定義する。

$$S_{\text{ifbot}} = (\theta_{N1}, \theta_{N2}, \theta_{E1}^{(L)}, \theta_{E2}^{(L)}, \theta_{E1}^{(R)}, \theta_{E2}^{(R)}, \theta_{L1}^{(L)}, \theta_{L2}^{(L)}, \theta_{L1}^{(R)}, \theta_{L2}^{(R)}, L_H, P1_M, P2_M, L_E, L_C, L_T)^T \quad (3)$$

$$\theta_{L2}^{(L)}, \theta_{L1}^{(R)}, \theta_{L2}^{(R)}, L_H, P1_M, P2_M, L_E, \quad (3)$$

$$L_C, L_T)^T \quad (4)$$

ここで、 $\theta^{(i)}$ はモータの角度値であり、 L は LED の点灯をパターン分類した値であり、図2で示した部位に対応する。なお、口部の形状は、文献[7]の方法を用いて、自己組織化マップによってそれぞれの形状間の類似度を評価し、パターン($P1_M, P2_M$)で表現している。

5 表情マッピングシステムの構築

5.1 ニューラルネットワーク

人と同じ顔構造を持つロボットに人の表情まねをさせるならば、人の顔表情から抽出した AU パラメータをそのままロボットの表情表出機構に与えてやればよい[4]。しかし、ifbot 顔構造は、デフォルメされており、人の顔構造とは大きく異なる。そのため、人の AU パラメータをそのまま利用し ifbot の表情を制御することはできない。そこで、表情特徴の対応付けのためにニューラルネットワークを用いる。本稿では、3層ニューラルネットワークを利用する。入力データとして人の表情パラメータ s_{man} 、教師信号として ifbot の表情パラメータ s_{ifbot} を与えて学習させることによって、物理的に異なる表情パラメータ同士をマッピングする。また、ニューラルネットワークの汎化性能によって、学習データ以外の表情についてもマッピングできることが期待される。ネットワークのユニット数は、16, 48, 16とした。

5.2 パラメータの正規化

人の表情パラメータにおいて、眉および目を制御するパラメータは左右合わせて12個、口のパラメータは4個である。口を制御するパラメータの数に比べて、眉・目のパラメータ数が圧倒的に多いことが分かる。このパラメータ数の比率のままニューラルネットワークの学習を行うと、人表情の眉・目が表情特徴の多くの部分を決定することになってしまい、眉・目の微妙な変化が、ifbot の表情全体に大きく影響してしまう。そこで、パラメータ α_{AU_i} の値を以下のように定義することで、口の形状変化についても考慮できる学習を行う。

$$\alpha_{AU_i} = \begin{cases} 1.0 & (i = 12, 15, 23, 27) \\ 0.5 & (\text{otherwise}) \end{cases} \quad (5)$$

5.3 学習データ

ネットワークの学習は、感情をキーとして行う。感情的な表情は、表情の特徴を大きく変化させるので、学習データとして適していると考えられる。学習に用いた表情データを図5に示す。学習の入力データは、AからDの4人の人物が表出した平静、怒り、喜び、悲しみ、驚きの表情とし、教師データとして ifbot の平静、怒り、喜び、悲しみ、驚きの表情を与える。なお、ifbot の各表情と感情の対応付けは、文献[8]の結果を利用した。

6 システムの評価

6.1 評価方法

学習によって構築されたシステムをアンケートにより評価する。アンケートは、提示された、人の顔画像と ifbot の表情画像を見比べ、対応する組み合わせを回答する形式で行った。具体的には、アンケート回答者に、ある人物の顔画像5枚と、システムにそれらの顔画像を入力して得られた ifbot の表情画像5枚を、それぞれランダムな順序で並べて提示し、対応している表情組を5組作ることで行った。アンケートには、学習に使用した4人(A, B, C, D)の顔画像と、未使用の3人(E, F, G)の7つのサンプルが用意され、回答者はすべてのサンプルについて回答する。アンケートは、学生20人に対して実施した。サンプルに用いた未学習データの表情画像の対応関係を図6に示す。

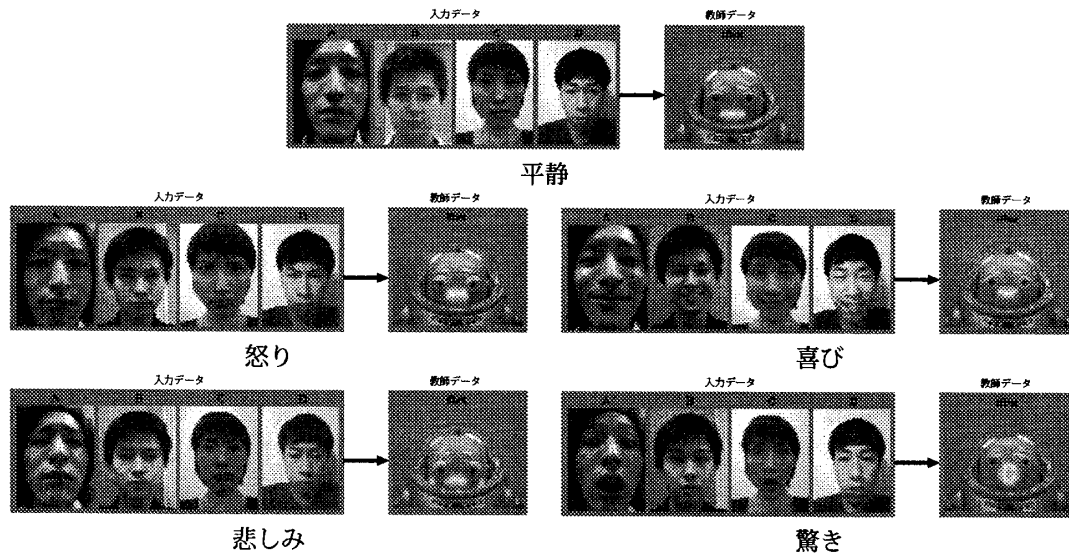


図5 学習データ

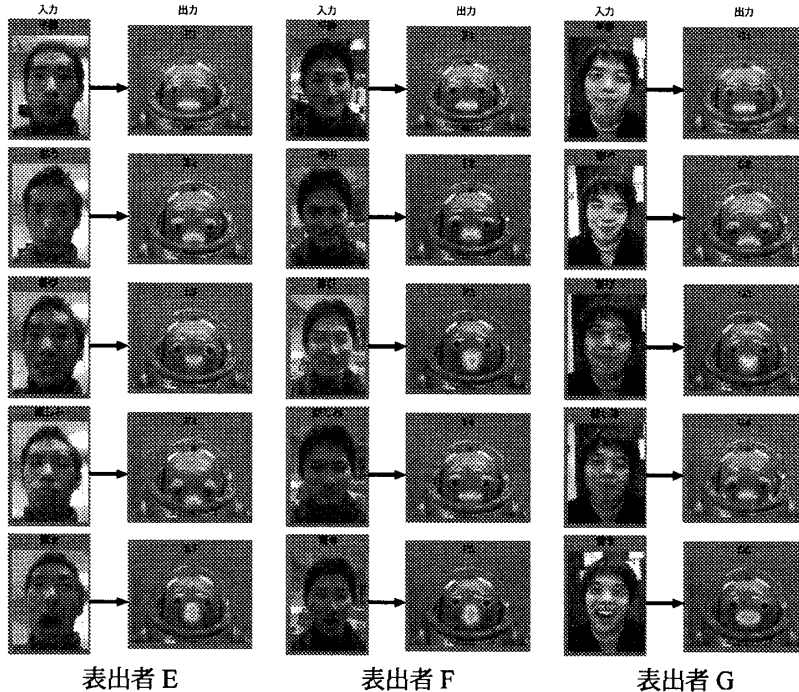


図6 未学習データの出力結果

6.2 結果と考察

表2から8にアンケート結果を示す。表中の数字は、アンケートにおいて各画像を組み合わせた回答者の人数である。太字で表記されている箇所は、システムの出力結果とアンケートによる評価結果が一致している箇所である。システムの入出力結果を正しく組み合わせることができた回答者の割合を正答率として右端の列に示す。同一人物の表情データごとの平均正答率を各表の右下に示す。

学習に使用した顔画像における平均正答率は、81%、使用していない顔画像における平均正答率は、69%であった。

まず、システムが出力した ifbot の表情について考察する。学習データ (表出者 A から D) については、ニューラルネットワークの学習により、正確に表情の対応付けが行える。未学習データ (図6参照) について見ると、平静、喜び、驚きの表情については、適切な感情表現ができていくことがわかる。悲しみの表情を見ると、表出者 F については、ifbot の表情は、涙を流してはいないが、眉の両端が下がっており悲しみの特徴を表現している。怒りの

表情については、3人とも感情を強く表出しているとは言いがたい。しかし、システムの出力した ifbot の表情を見ると、口角が下がるという怒りの口の特徴を表現しつつ、興奮状態を表す赤の LED が、頭部、目部で使用されており、怒りの印象を強めていることが見て取れる。これらのことから、システムの出力表情は、十分に感情を表現できていると考えられる。実際の利用時には、表情表出に加えて、言語情報も付加されるので、話し手の感情伝達の正確性は増すと考えられる。

つぎに、学習データのアンケート結果を考察する。

表出者 C (表4) については全員が正答であり、また、表出者 B (表3) および D (表5) についても、平静と悲しみについてやや誤答があるものの、高い正答率であった。表出者 A の結果 (表2) を見ると、怒りと悲しみが逆に対応付けられていることがわかる。実際に、図5で表出者 A の表情を確認すると、悲しみの表情では、眉の内側が引き上がり、口は震えているような印象を与えており、悲しみの表情の特徴を表現している。しかし、見方によっては、顔全体がやや上を向き、目が食い入るように凝

表2 アンケート結果(表出者 A)

		A1	A2	A3	A4	A5	正答率(%)
入力	平静	17	2	0	1	0	85
	怒り	3	2	0	15	0	10
	喜び	0	0	20	0	0	100
	悲しみ	0	16	0	4	0	20
	驚き	0	0	0	0	20	100
平均正答率							63%

表3 アンケート結果(表出者 B)

		B1	B2	B3	B4	B5	正答率(%)
入力	平静	13	1	0	6	0	65
	怒り	2	18	0	0	0	90
	喜び	0	0	20	0	0	100
	悲しみ	5	1	0	14	0	70
	驚き	0	0	0	0	20	100
平均正答率							85%

表4 アンケート結果(表出者 C)

		C1	C2	C3	C4	C5	正答率(%)
入力	平静	20	0	0	0	0	100
	怒り	0	20	0	0	0	100
	喜び	0	0	20	0	0	100
	悲しみ	0	0	0	20	0	100
	驚き	0	0	0	0	20	100
平均正答率							100%

表5 アンケート結果(表出者 D)

		D1	D2	D3	D4	D5	正答率(%)
入力	平静	12	0	3	5	0	60
	怒り	0	20	0	0	0	100
	喜び	0	0	16	0	4	80
	悲しみ	5	0	0	14	1	70
	驚き	2	0	1	2	15	75
平均正答率							77%

表6 アンケート結果(表出者 E)

		E1	E2	E3	E4	E5	正答率(%)
入力	平静	18	1	0	1	0	90
	怒り	2	11	0	7	0	55
	喜び	0	0	20	0	0	100
	悲しみ	0	8	0	12	0	60
	驚き	0	0	0	0	20	100
平均正答率							81%

表7 アンケート結果(表出者 F)

		F1	F2	F3	F4	F5	正答率(%)
入力	平静	15	0	1	3	1	75
	怒り	1	14	0	5	0	70
	喜び	1	1	17	0	1	85
	悲しみ	3	5	0	12	0	60
	驚き	0	0	2	0	18	90
平均正答率							76%

表8 アンケート結果(表出者 G)

		G1	G2	G3	G4	G5	正答率(%)
入力	平静	6	8	1	5	0	30
	怒り	4	4	12	0	0	20
	喜び	8	4	7	1	0	35
	悲しみ	2	4	0	14	0	70
	驚き	0	0	0	0	20	100
平均正答率							51%

7 おわりに

本稿では、人の顔画像から表情特徴を抽出し、ニューラルネットワークを用いてその表情を ifbot の表情で再現する表情対応付けを提案した。評価実験では、システムによる表情対応付けの妥当性が示された。また、本システムによって、人が他者の表情を読み取る能力以上の精度で、感情的な表情を表出できる効果も見られた。これは、別の言い方をすれば、感情表出を強める効果を持つということである。これに関しては、さらに実験データ数を増やし、検証していく必要がある。本システムの精度を向上させることによって、将来的には、匿名相談やカウンセリングといった場面への応用が期待できる。

謝辞

ifbot は株式会社ビジネスデザイン研究所の製品企画・総合プロデュースのもと、特に、形状デザイン、表情制御メカニズムはブラザー工業株式会社、および、表情制御ソフトはブラザー工業株式会社、ロボ株式会社、名古屋工業大学が共同開発した。関連各位に感謝する。

参考文献

- [1] 辻：感性の科学-感性情報処理へのアプローチ，サイエンス社(1997).
- [2] ビジネスデザイン研究所：http://www.ifbot.net.
- [3] 後藤 他：感性ロボットのための感情領域を用いた表情生成，人工知能学会論文誌，21(1)，55-62(2006).
- [4] 原，小林：顔という知能，共立出版(2004).
- [5] 感性擬人化エージェントのための顔情報処理システムの開発：http://www.hc.ic.i.u-tokyo.ac.jp/project/face/IPA.
- [6] 田中 他：ニューラルネットワークによる6基本表情認識の規則発見，PRMU，100(134)，71-78(2000).
- [7] 柴田 他：感情指定パラメータからの感性ロボットの表情生成，電学論C，125(12)，1852-1860(2005).
- [8] M. Kanoh et al.: Facial Expressions Using Emotional Space in Sensitivity Communication Robot "ifbot", IROS2004, 1586-1591(2004).

視しており、威嚇の表情とも読み取れる。これが、悲しみの表情が怒りの表情と混同された原因と考えられる。このように、主観評価では、表出者本人の感情が相手(アンケート回答者)にうまく伝達されていないことがわかる。他方、ニューラルネットワークを用いて表情マッピングする際には、本人の感情と対応した ifbot の表情を正規化したパラメータで対応付けて学習しているので、回答に窮するような表情表現についても適切に対応付けが行われていると考える。これを確認するために、未学習データのアンケート結果について見ていく。

表出者 E (表6), F (表7) については、怒りと悲しみの間でやや間違いが見受けられるが、おおむね高い識別率であった。表出者 G (表8) については、平静、怒り、喜びの正答率が非常に低かった。図6で表出者 G の表情を確認すると、驚きの表情を除いて、ほぼ同じような表情を表出している。このため、アンケート回答者の正答率が著しく低下したと考える。他方、ifbot と表出者の意図した感情との対応について見ると、非常によく対応づけられていることがわかる。このようにニューラルネットワークを用いることで、一般の人が他者の表情を読み取る能力以上の精度で、感情的な表情を表出できることがわかる。これは、遠隔診断において重要な要素であると考えられる。たとえば、専門家が、一般の人では読み取り間違えてしまうような表情を表出する患者を診察する際にも、本システムを用いることで適切な表情を表現しつつ、かつ、匿名性を保ちつつ診察できると予想される。