

顔表情の研究における表情判別システムの利用可能性^{※1}

松尾貴司・小川一美

A case study of using the facial expression recognition system

Takashi Matsuo and Kazumi Ogawa

要旨

本報告は、顔表情に関する研究の様々な場面で利用が期待される、画像処理による顔表情の判別システムの利用可能性を検証するための基礎的データを提供しようとするものである。Noldus社のFaceReaderという顔表情解析システムと、ATR顔表情画像データベースDB99に収録されている顔表情画像を用いて、FaceReaderの顔表情判別の精度について2つの検証をおこなった。1つは、ATRデータベースに付属する心理評定値との比較検討であった。もう1つは、FaceReaderが算出するデータとの対応をはかるために、新たにおこなった実験データとの比較検討であった。後者は、女子大学生20名に、Visual Analog Scaleを用いた評定法と強制選択法の2種類の方法により、顔表情を評定させるものであった。本報告に示されたデータでは、FaceReaderの顔表情判別の精度は欧米人の顔表情に関するデータに比べるとやや低いものであったが、今後の利用可能性が期待されるものであった。

キー・ワード：顔表情， 解説， 基本6表情， フェイス・リーダー（FaceReader）

はじめに

心理学は、人の顔表情について様々なアプローチをおこなってきた。一つは、人の顔表情についての系統発生的基盤，すなわち顔表情の起源（進化）に対する興味から生じるもので、このようなアプローチは、もともとエソロジストが示してきたものと言ってよいであろう。より心理学的なアプローチとしては、まず文化間比較の研究があげられる。これは、顔表情の普遍性に関する研究と言い換えてもよいであろう。これらの研究によって、顔表情の普遍性が認められた一方、文化によって異なる表出ルールが存在することも明らかになってきた。

これらは主として顔表情の表出に関するもので

あるが、一方、顔表情の解説，認知の正確性に関する研究も盛んに行われてきた。また、顔表情への発達のアプローチでは、顔表情の表出と解説の双方に関する研究がおこなわれてきたが、それらは情動の発達に関する研究の指標として顔表情を利用したものである。

顔表情そのものに対する興味からおこなわれる研究にせよ、指標として顔表情を利用する研究にせよ、顔表情を扱う研究では方法上の様々な問題に遭遇することになる。例えば、顔表情を独立変数とする研究では、刺激画像の妥当性を確保することは非常に重要であるが、容易なことではない。また、刺激の呈示方法や、刺激に対する反応をどのように測定するかも結果に影響をおよぼすことが考えられる。また、顔表情を従属変数とする研究の場合も、判別の客観性，正確性を確保することが大きな課題となる。

このような顔表情の研究における方法上の必要

※1 本研究は平成23年度・24年度愛知淑徳大学共同研究助成を受けておこなわれた。

性から、顔表情を客観的、量的に記述する方法が開発されてきた。中でも、Ekman & Friesen (1978)が開発したFACS (Facial Action Coding System)は、広く使われている。FACSは44の顔面筋動作をAU (Action Unit)として記述し、基本的な6表情を分類することが可能になるものである。

FACSのような記述システムが開発されたことによって、顔表情の客観的な記述が可能になったとはいえ、顔表情の判別には、かなりの時間と労力が必要であることには違いなかった。しかし、画像処理技術の進歩等によって、パーソナルコンピュータを用いて顔表情の判別をおこなう技術が開発されたことにより、状況は大きく変化してきた。もちろん、顔表情の研究方法上のすべての問題が解決されるわけではないが、顔表情の自動判別が研究にもたらす可能性は省力化以外にも期待される。

本報告では、Noldus社のFaceReaderという顔表情解析ソフトウェアを用いて、日本人の顔表情の自動判別について基礎的データを収集し、実用可能であるかを検証するとともに、顔表情の自動判別がどのような研究に利用可能かを考えてみたい。

検証1

標準的な顔表情として心理学などの実験でも用いられている、ATR顔表情画像データベースを用いて、FaceReaderによる表情の判別精度について検証した。

1. 方法

1.1. 顔表情画像

ATR顔表情画像データベースDB99に収録されている画像のうち、男性モデル6名、女性モデル4名の正面顔データを用いた。データベースには、真顔(本稿では、Neutral)、喜び(開口)(Happy (mouth open))、喜び(閉口)(Happy (mouth close))、悲しみ(Sad)、驚き(Surprised)、怒り(開口)(Angry (mouth open))、怒り(閉口)(Angry (mouth close))、嫌悪(Disgusted)、軽蔑、恐れ(Scared)の10種類の表情が含まれ

ているが、FaceReaderの分類項目にはない軽蔑を除く9種類の表情について分析をおこなった。

Neutralについては1枚のみが収録されており、それ以外については3枚以上の画像が含まれていたが、データベースに付属する心理評定実験で用いられた各3枚を分析対象とした。したがって、分析をおこなった全画像の枚数は、人物10×[Neutral×1+その他8表情×3]枚の計250枚であった。

1.2. FaceReaderのデータ

各画像(640×486ピクセルのビットマップ画像)をFaceReaderのImage Analysisにより解析した。使用したソフトウェアのバージョンは4.0.8、パーソナルコンピュータはDELL社のPrecision M6500であった。なお、解析時のFace ModelはEast Asianを選択した。

FaceReaderでは、7つの表情についてそれぞれ0～1の範囲でその強度が測定され、そのうち値が最も大きいものを画像の表情分類としている。

1.3. 比較対象データ

比較対象としてATRデータベースに付属の心理評定データを用いた(以下、ATR値)。ATR値は、各顔画像を見て、その顔が表していると思われる感情の強さを「幸福」「悲しみ」「驚き」「怒り」「嫌悪」「恐れ」「軽蔑」の7種類について、それぞれ1(全く表れてない)～7(強く表れている)の7段階で評定した27名の平均値である。この平均値に加えて、各画像について7種類のうち最も値が大きい感情をATR値の表情分類結果とした。また、各表情の強度については、1～7の値を0～1になるように変換して比較をおこなった。

2. 結果

2.1. FaceReaderによる解析の失敗

250枚すべての写真についてFaceReaderによる解析をおこなったところ、2名の男性モデル(M02およびM06)の画像で表情の分類に失敗する(Unknownとなる)ことが多数見られた。M02では25枚中14枚(56%)、M06では25枚中6枚(24%)が分類不能となったが、他のモデルの画像では3名で各1枚の分類不能があったのみで

あった(表1)。FaceReaderは、眼鏡をかけている場合は分析できないなど、いくつかの制約があることが分かっている。この2名の人物の場合、どのような特徴(あるいは、画像の特性)が原因であるかは明確でないが、今回は分析の対象から除外することにした。

また、M01とM02のモデルについては、真顔をNeutralと分類することができなかった。このことは、表情の判別に大きな影響を与える可能性があると考えられるので、M02、M06に加えてM01も分析対象から除外することにした。したがって、以下のデータは人物7×[Neutral×1+その他8表情×3]の計175枚の画像について分析をおこなった結果である。

2.2. 表情分類の精度

表2は、9種類の表情写真をFaceReaderがどのように分類したかを一覧にしたものである。一方、表3は、ATR値による表情分類の結果である。ATR値には、評定項目にNeutralがなく、今回使用しなかったContemptの項目が含まれているため、これらに分類された場合については、比率の計算から除外した。以下に、各顔表情の結果を記述する。

Happy 口の状態に関わらず、FaceReader、ATR値とも、すべてHappyと分類された。

Angry ATR値では、閉口の場合86%、開口の場合100%とかなり正確に分類されたが、

FaceReaderでは、閉口の場合は76%だったものの、開口では38%とかなり低い値となった。Scaredに分類されたもののがかなりあったが(33%)、Happyに分類されたものもあった(14%)。

Sad ATR値では、すべて正確に分類されたが、FaceReaderでは、24%しか正確に分類されなかった。Angryに分類されたものが多かったが(38%)、Neutralと分類されたものが19%あった。

Surprised FaceReaderでも(86%)、ATR値でも(95%)、かなり正確に分類されていた。

Scared ATR値では33%とかなり低い値であり、Surprisedに分類される割合が高かった(57%)。FaceReaderでは62%と、ATR値よりも高い値を示したが、同様にSurprisedに分類されることが多かった(29%)。

Disgusted ATR値(52%)、FaceReader(48%)とも半数程度が正確に分類された。ただし、ATR値ではSadと分類されたものが多かったが(43%)、FaceReaderではSadと分類されたものは1枚だけであった。

また、FaceReaderが各表情に分類したものが、当該の表情であった割合が、表2のPrecisionの値である。この値も、他の研究一例えば、Uyl & Kuilenburg (2005)では、0.80~0.97であったーと比べると必ずしも高いものとは言えないが、極端に低いとも言えないものであった。しかし、Neutralと分類されたものに、他の表情が多く含

表1 FaceReaderによる顔表情解析の失敗生起頻度

	Number of Photos	Classification of Neutral	Failure of Classification	Precision of Classification
F03	25	Match	0 (0%)	17 (68%)
F10	25	Match	1 (4%)	14 (56%)
F13	25	Match	0 (0%)	20 (80%)
F16	25	Match	0 (0%)	18 (72%)
M01	25	Unmatch	0 (0%)	12 (48%)
M02	25	Unmatch	14 (56%)	6 (24%)
M05	25	Match	0 (0%)	18 (72%)
M06	25	Match	6 (24%)	10 (40%)
M09	25	Match	1 (4%)	14 (56%)
M10	25	Match	1 (4%)	18 (72%)

まれていたことは、注意すべき点であるように思われる。

2.3. 表情の強度の評定

Neutralを除く8種類の表情について、人物7×3枚の計21枚の画像ごとの各表情の評定値をFaceReader, ATR値それぞれに示したものが図1である。ただし、Sadで1枚、Surprisedでは

2枚がFaceReaderで分析不能であったため、集計から除外された。

前述の分類の精度で、ATR値、FaceReaderとも高い値を示した、HappyとSurprisedは、強度のパターンも類似しており、当該の表情のみ値が大きくなり、他は非常に小さな値となっていた。

FaceReaderでは、特定の表情強度のみが強く

表2 FaceReaderによる顔表情の分類結果

	Happy	Angry	Sad	Surprised	Scared	Disgusted	Neutral	Unknown	Recall
Happy	42	0	0	0	0	0	0	0	1.00
<i>mouth close</i>	21	0	0	0	0	0	0	0	1.00
<i>mouth open</i>	21	0	0	0	0	0	0	0	1.00
Angry	3	24	1	3	7	2	2	0	0.57
<i>mouth close</i>	0	16	1	0	0	2	2	0	0.76
<i>mouth open</i>	3	8	0	3	7	0	0	0	0.38
Sad	0	8	5	0	0	3	4	1	0.24
Surprised	1	0	0	18	0	0	0	2	0.86
Scared	0	0	0	6	13	0	2	0	0.62
Disgusted	0	2	1	0	5	10	3	0	0.48
Neutral	0	0	0	0	0	0	7	0	1.00
Precision	0.91	0.71	0.71	0.67	0.52	0.67	0.39		

表3 ATR値に基づく顔表情の分類結果

	Happy	Angry	Sad	Surprised	Scared	Disgusted	Contempt	Unknown	Recall
Happy	42	0	0	0	0	0	0	0	1.00
<i>mouth close</i>	21	0	0	0	0	0	0	0	1.00
<i>mouth open</i>	21	0	0	0	0	0	0	0	1.00
Angry	0	39	0	0	0	2	1	0	0.93
<i>mouth close</i>	0	18	0	0	0	2	1	0	0.86
<i>mouth open</i>	0	21	0	0	0	0	0	0	1.00
Sad	0	0	21	0	0	0	0	0	1.00
Surprised	0	0	0	20	0	0	1	0	0.95
Scared	0	0	2	12	7	0	0	0	0.33
Disgusted	0	0	9	0	0	11	0	1	0.52
Neutral	4	0	2	0	0	0	1	0	
Precision *	1.00	1.00	0.66	0.63	1.00	0.85			

* Exclude the classification of Neutral stimulus

検出される傾向が見られるが、ATR値では、複数の表情の強度がある程度示されるパターンがあった。特に分類精度が低かったScaredやDisgustedで顕著に見られた。

検証 2

ATRデータベースに付属の心理評定データは、測定項目がFaceReaderと異なっていることもあり、比較するにはいくつかの問題があった。そこで、FaceReaderと同じような形でデータを収集し、比較検討することにした。本データの一部は、小川・松尾(2012)で報告されたもので、データ取得の手続きは同一である。

1. 方法

1.1. 顔表情画像

ATR顔表情画像データベースDB99に収録されている画像のうち、女性モデル2名(F03およびF16)の正面顔データを用いた。表情は、Neutral, Happy (mouth open), Sad, Angry (mouth close), Surprised, Scared, Disgustedの7種類について各1枚の画像を使用した。

1.2. 測定方法

画像呈示および評定は、Apple社のiPad上で自作アプリケーションを用いておこなった。評定方法は、強制選択法とVisual Analog Scale (VAS)による評定の2種類をおこなった。いずれの場合も、iPadを縦置きにした状態で、上半分に顔画像を呈示し、下半分にそれぞれの評定用ボタンを呈示した。

強制選択法では、「中立」「喜び」「悲しみ」「怒り」「驚き」「恐怖」「嫌悪」の7種類から、呈示された顔画像に最もよく当てはまると思われるものを1つ選択させた。

VASでは、顔画像に各表情がどれくらい表れていると思うかを、強制選択法と同じ7種類について、0 (Min)～100 (Max)の範囲でボタンをスライドさせる方法で評定させた。なお、ボタン位置の情報(数値)は、表示されていなかった。

1.3. 手続き

20名の女子大学生が、2つの評定方法を1日以上の間隔をあけてそれぞれおこなった。各評定方

法について、人物2×表情7の顔画像を各2回、ランダムに呈示し、合計28回の評定をおこなった。

2. 結果

2.1. 強制選択法

強制選択の結果は、1回ずつの評定結果を集計した。すなわち、モデル人物1名の1つの表情について、40回(評定者20名×2回)の評定をおこなったものとして集計した。

Happy, Angry, SurprisedおよびNeutralについては、2名のモデル人物とも90%～100%という高い割合で、正確に判別されていた。一方、ScaredとDisgustedは、比較的低い判別率となった。また、Sadについては、F16では100%であったのに対して、F03では68%と低くなった(表4, 表5)。

2.2. VAS

VASの値については、評定者1名の同一画像に対する2回の評定値の平均値を用いた。

20名の評定値の分布および、20名の評定の平均値とFaceReaderの評定値を、モデル人物ごとに示したものが図2および図3である。

強制選択法では正確に分類されていたHappy, Angry, Surprised, Neutralであったが、VASで見ると、いくつかの特徴が見られた。

Neutral FaceReaderでは、F03に対してAngryを検出していたが、本実験の参加者にはほとんど見られなかった。しかし、実験参加者の反応には、各個人で見るとNeutral以外の表情を検出している者がいた。

Happy どちらのモデルでも、実験参加者の反応にNeutralがかなり見られた。

Angry FaceReaderではAngryのみが強く検出されていたが、実験参加者の反応では、Disgustedもかなり大きな値になった。また、F16ではSadの値も大きくなっていった。

Surprised 実験参加者の一部で、F16のHappyの値が大きくなった。

強制選択法で判別率が低くなったScaredとDisgustedについては、VASの値の分布と強制選択の結果に対応が見られた。

Scared F03はFaceReaderでScaredがほとん

表4 F03モデルの顔表情の強制選択法による分類結果

F03	Happy	Angry	Sad	Surprised	Scared	Disgusted	Neutral	Recall	FaceReader's Choice
Happy	38	0	0	0	0	0	2	0.95	Happy
Angry	0	40	0	0	0	0	0	1.00	Angry
Sad	0	0	27	0	1	10	2	0.68	× <u>Angry</u>
Surprised	0	0	0	40	0	0	0	1.00	Surprised
Scared	0	0	3	8	22	7	0	0.55	× <u>Surprised</u>
Disgusted	0	23	0	0	0	17	0	0.43	Disgusted
Neutral	0	0	0	0	0	0	40	1.00	Neutral
Precision	1.00	0.63	0.90	0.83	0.96	0.50	0.91		

表5 F16モデルの顔表情の強制選択法による分類結果

F16	Happy	Angry	Sad	Surprised	Scared	Disgusted	Neutral	Recall	FaceReader's Choice
Happy	40	0	0	0	0	0	0	1.00	Happy
Angry	0	36	2	0	0	2	0	0.90	Angry
Sad	0	0	40	0	0	0	0	1.00	Sad
Surprised	3	0	0	37	0	0	0	0.93	Surprised
Scared	0	0	2	11	27	0	0	0.68	Scared
Disgusted	0	7	3	0	1	29	0	0.73	× <u>Scared</u>
Neutral	0	0	0	0	0	0	40	1.00	Neutral
Precision	0.93	0.84	0.85	0.77	0.96	0.94	1.00		

ど検出されなかったが、実験参加者では大きな値になったものもいた。また、VASの値が大きくなったDisgustedとSurprisedが強制選択でも多く選ばれていた。F16ではFaceReaderの値が比較的明瞭であったのに対して、参加者の反応は分散しており、強制選択でも多く選ばれたSurprisedの値が大きくなっていた。

Disgusted F03では、FaceReaderの値も実験参加者の反応もAngryの値がDisgustedと同様に大きくなった。強制選択においてもAngryが多く選ばれていることと対応していた。一方F16については、FaceReaderが特定の表情を強く検出していなかったのに対して、実験参加者はF03と同様にAngryとDisgustedに、さらにSadにも大きく反応していた。

Sad F03ではFaceReaderがSadをまったく検

出せずAngryの値が大きくなったのに対して、実験参加者ではSadの値とDisgustedが大きくなった。F16ではFaceReaderがSadの値と共にNeutralが大きくなっていたのに対して、実験参加者ではそのような反応は見られなかった。

おわりに

1. FaceReaderの顔表情分類の精度

検証1の結果を見る限り、FaceReaderの顔表情の分類精度は、十分に高いと言うことはできないものであった。これにはいくつかの原因が考えられるが、第一に、このシステムが西洋人の顔表情分類に最適化されて作られてきたことがあげられるであろう。もちろん、アジア人への対応も考えられてはいるが、十分なシステム構築が完了し

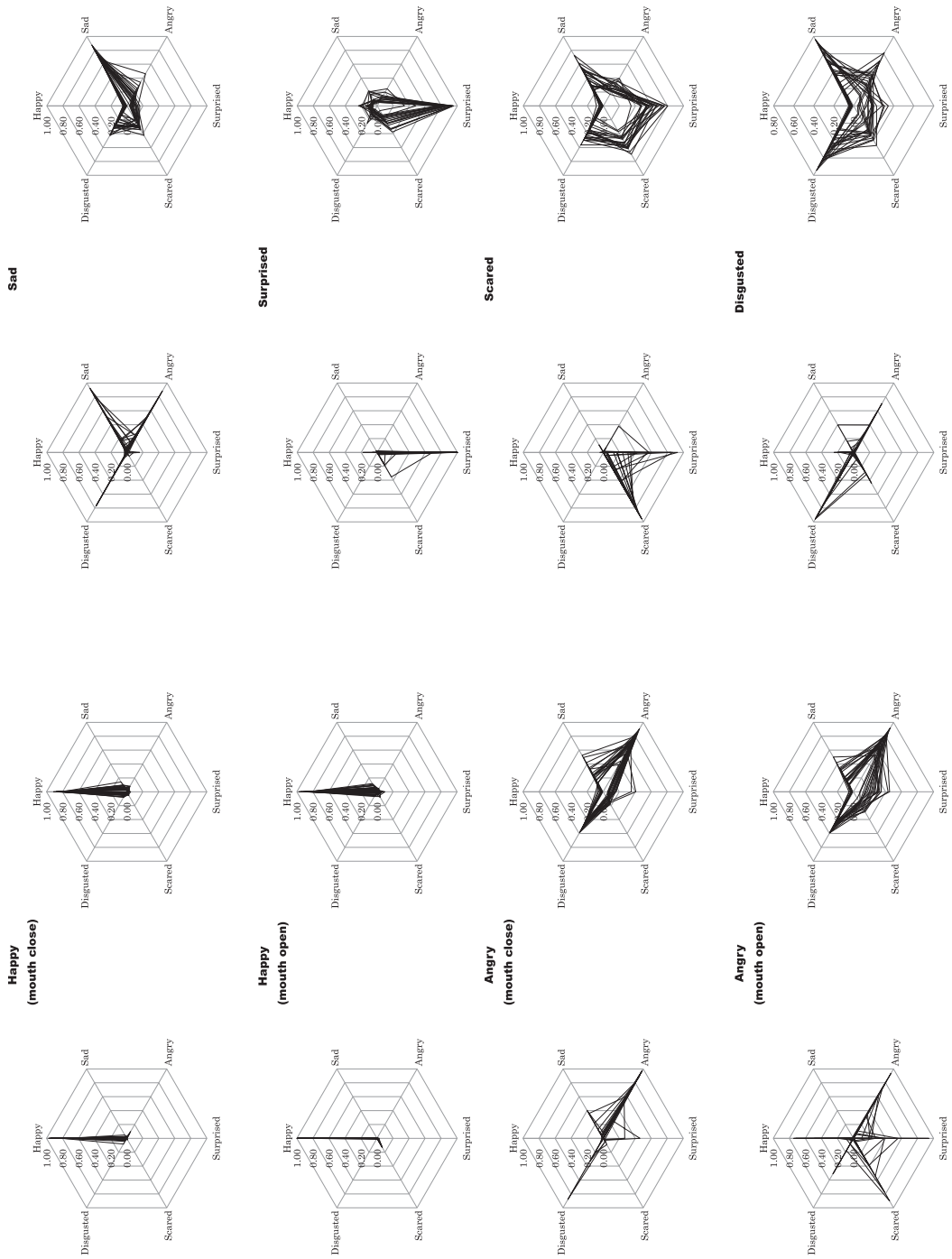


図 1 顔表情ごとの各顔画像 (21枚) に対する表情評定値の分布 (左: FaceReaderの値, 右: ATR値)

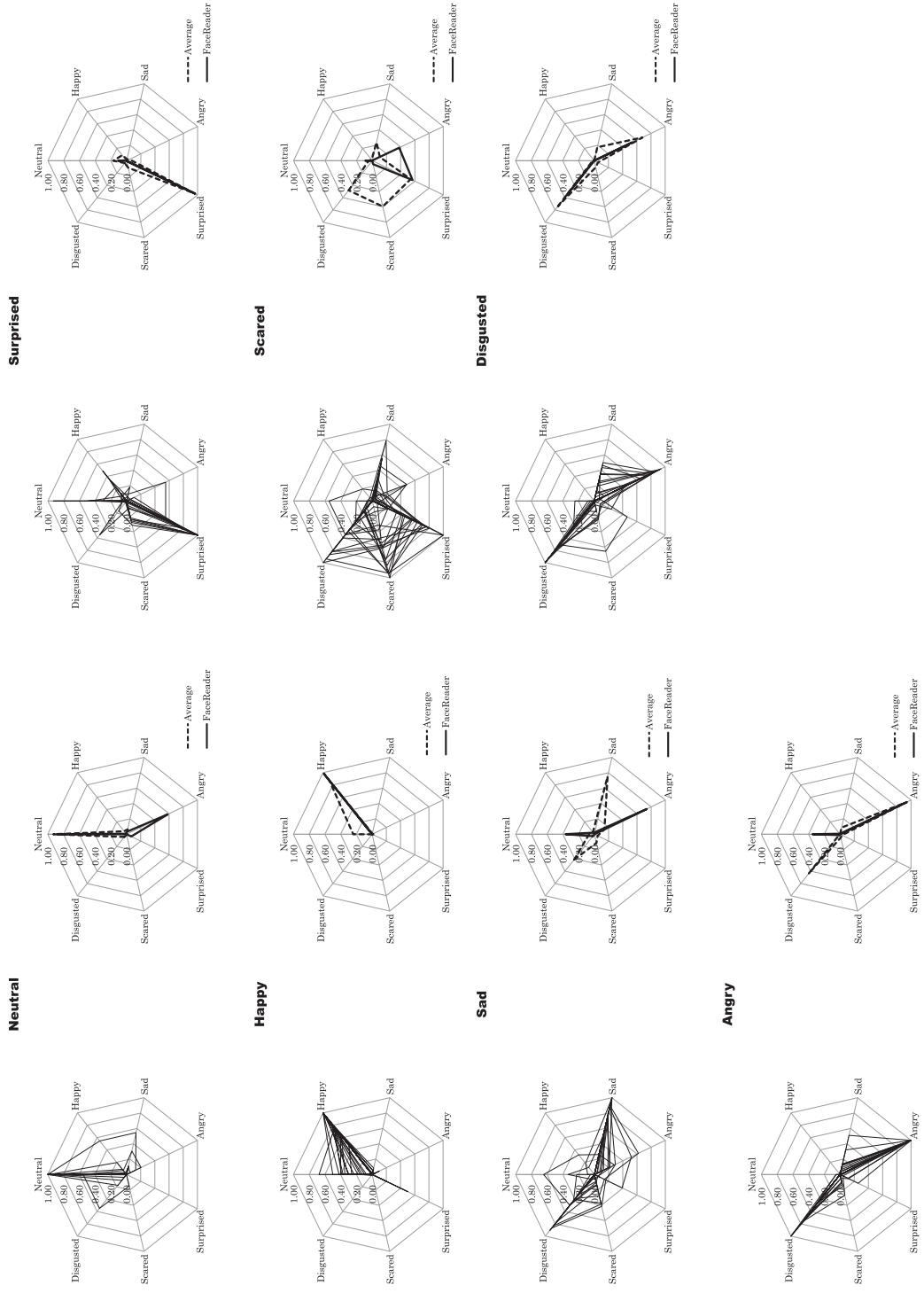


図2 モデルF03の各顔画像に対する表情評定値 (左: 20名の評定値分布, 右: 20名の評定平均値とFaceReaderの値)

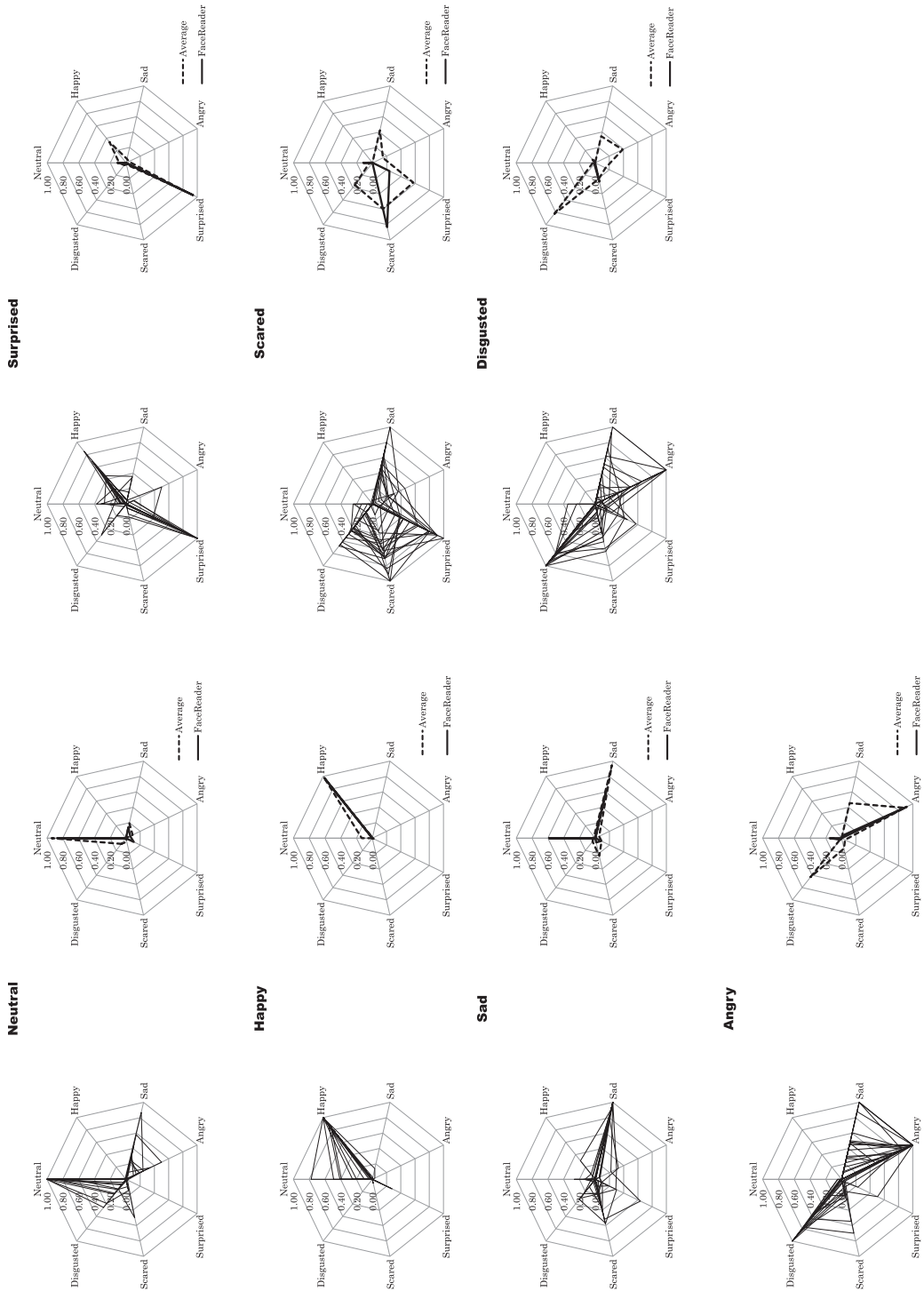


図3 モデルF16の各顔画像に対する表情評定値 (左: 20名の評定値分布, 右: 20名の評定平均値とFaceReaderの値)

ていない可能性もある。本研究ではバージョン4を使用して解析をおこなったが、旧バージョンによる解析では、更に精度が低くなる傾向が見られた。すでに次のバージョンが開発されているという情報もあり、さらなる精度の向上が期待される。

もう一つの問題は、解析対象の顔表情写真が適切であったかどうか、ということである。これは、表情研究では常につきまとう問題でもある。今回使用した顔画像に対する日本人による評定を見ても、ScaredやDisgustedに対する分類の正確性はそれほど高いものではなかった。もっとも、恐れや嫌悪が他の表情と判断されやすいというのは、他の研究でもしばしば見られるものではある（例えば、渋井・繁樹, 2005）。

また、各顔表情画像を正確に分類できた割合（Recall）に比べ、FaceReaderがある表情と分類した顔表情が実際にその表情であった割合（Precision）は、若干高くなっていた。しかし、Neutralではこの値が著しく低くなっていた（表2）。すなわち、Neutralでない顔表情がNeutralと分類されてしまったわけである。このことが表情分類にどのような影響を及ぼすかは明確ではないが、注意すべき問題であると考えられる。

2. 各表情の強度に基づく顔表情の判断

本報告で用いた、VASによって各表情の強度を評定するという方法は、実験参加者にとって、それほど困難な課題ではなかったようであり、比較的安定したデータが得られたように思う。

FaceReaderでは、各表情の強度が最も大きいものをその顔表情としているが、実際の表情解読は、もう少し複雑なプロセスが存在していると考えられる。本報告では、個別の判断プロセス、例えば、VASの値の分布と強制選択との関係などについては分析していないため、この点に言及することはできないが、今後の課題の一つである。

3. 表情研究における表情判別システムの利用可能性

表情研究において、FaceReaderのような表情判別システムを利用することの一つのメリットは、省力化である。これは主として、従属変数として顔表情の分類をおこなう研究において威力を発揮するであろう。このためには、さらにデータを集めて、分類精度が十分に得られることを確かめていくことが必要である。

また、独立変数として顔表情を利用する場合にも、選択した顔表情がどのような特徴をもっているかを客観的に示す方法として有益である。特に、複数の表情が混在した顔表情を刺激として用いようとする場合には、各表情の強度を量的に示すことができるのは有用であろう。

この他にも、FaceReaderでは、動画像やリアルタイムの画像から表情を分類することが可能であるため、表情のトレーニングや、より現実場面に近い表情変化に対する反応の研究などへの応用も期待できる。

文 献

- Den Uyl, M. J. & van Kuilenburg, H. (2005). The FaceReader: Online facial expression recognition. *Proceedings of Measuring Behavior* 2005, 589-590.
- Ekman, P. & Friesen, W. V. (1978). *Facial action coding system: A technique for the measurement of facial movement*. Palo Alto, CA: Consulting Psychologists Press.
- 小川一美・松尾貴司 (2012). 強制選択法と視覚的評価法による基本6表情の解読 電子情報通信学会技術研究報告, HCS2012-51, 19-23.
- 渋井進・繁樹算男 (2005). 表情の2次元空間配置モデルの検討 心理学研究, 76, 113-121.