

表情認識における局所的な処理と全体的な処理

Partial and Holistic Processing in Facial Expression Recognition

赤 塚 瞳

Hitomi AKATSUKA

(日本女子大学大学院人間社会学研究科 心理学専攻博士課程前期1年)

要 約

顔認識の全体的処理モデルでは、顔は、構成要素が分離して認識されるのではなく、それらの特徴が統合され全体として認識されると考えられている。一方で、顔の構成要素に注意を向けるといった局所的な処理の存在も示唆される。本研究では、顔の構成要素がもたらす印象へのコントラスト感度と、表情の印象へのコントラスト感度を比較し、全体的処理モデルの妥当性を検討することを目的とした。この目的のために、各構成要素への感度を確率的に加重すると顔全体への感度と等しくなるのか、あるいは、表情全体への感度は各構成要素への感度の加重では説明できないのかを実験的に調べた。その結果、表情全体への感度は確率加重モデルにより予測できるか、あるいは個別の構成要素への感度と等しくなった。以上より、全体的処理モデルでは表情認識に関するコントラスト感度を十分に説明することができないことがわかった。

[Abstract]

One significant aspect of face processing is the holistic manner in which faces are processed. The human visual system is assumed to integrate facial features into a gestalt whole. In this study, we quantitatively examine the validity of the holistic processing model by utilizing a model based on the probability summation. In the experiments, we have examined whether the contrast sensitivity for detecting the facial expressions (happy or angry) from two facial parts is predicted by the probability summation of the contrast sensitivity for detecting each facial part. We found that the contrast sensitivity for detecting facial expressions can be predicted by the probability summation of the sensitivities for detecting each part, or it was lower than that from the prediction. These results indicate that the holistic processing model is not adequate to explain the contrast sensitivity for detecting the facial expressions.

1 序論

顔は、単なる身体の一部であるにも関わらず、たくさんの情報を含んでいる。見ただけでその人の性別やおおよその年齢、または感情を予測することができるのは、身体の中で顔だけである。顔は、構成要素（あるいは個々の特徴）が分離して認識されるのではなく、それらの特徴が統合され、全体として認識されると考えられている（Tanaka & Gordon, 2011）。こういった顔に関する処理を全体的処理（holistic processing）、顔に関する表現を全体的表現（holistic represen-

tation) と呼ぶ。つまり、顔の構成要素は同じであっても、構成要素の配置がずれるだけで、別の顔の「ひとまとまり」として認識される。顔認知は、目や口などそれぞれの部分の特徴でなく、その配置で見ることが最大の特徴である。全体的処理には1次処理と2次処理の2段階あり、顔そのものを検出することを1次処理、顔の配置を区別し、個々の顔として認識することを2次処理と示す (Maurer et al., 2002)。1次処理における顔を見る基準は構成要素の配置であり、構成要素が正しい配置にあるときに処理が進む。2次処理では、個々の顔における構成要素の配置の細かい違いをもとに、個人の顔が識別される。

顔が全体として表現されている一つ目の証拠として、倒立効果が挙げられる。つまり顔は、上下逆さまの場合に認識が非常に困難になることから、顔の構成要素の特徴は、独立的に知覚されるのではないと言える。サッチャー錯視で有名な倒立効果は、主に2次処理が阻害されることによって起きるとされている。成人の場合、倒立によってわかりにくくなる効果が強いのは2次処理による顔の区別であり (Freire et al., 2000)、全体処理と比べると部分処理では弱いとされている (Cabeza & Kato, 2000; Bartlett & Seacy, 1993)。サッチャー錯視でわからないのは顔が持つ印象（とくに不気味さ）であり、顔そのものは検出できる。このことから、サッチャー錯視による倒立効果は2次処理に関与していることが示される。

二つ目の証拠としては、顔の構成要素の記憶は、その要素のみを分離してテストした時よりも、顔全体の中に埋め込んで見せた時の方が、成績が向上するという点が挙げられる (Tanaka & Gordon, 2011)。この効果は、家などの顔ではない刺激や倒立顔では得られていないことから、顔が全体として表現されているという説と合致する。

一方で、私たちは顔やその構成要素の微妙な変化から感情を読み取ることができる。その際には、顔の全体的な表現に基づいて感情を読み取っているのでしょうか？それとも、顔の構成要素に注意を向けるといった、より局所的な処理を行っているのでしょうか？本研究では、顔の構成要素がもたらす感情的な印象と、顔全体がもたらす感情的な印象を比較することにより、この疑問について実験的に検討することを目的とする。その際には、確率加重 (probability summation) モデルを利用する。まず、顔の構成要素（例えば目と口）がどのような表情印象を与えるかを調べる。続いて、目と口を同時に提示したときにもたらされる表情印象について調べる。後者における印象形成において、前者における顔の構成要素がそれぞれ独立に寄与しているとすれば、確率加重で説明できると考えられる。この場合には、表情の認識は必ずしも全体的処理されているわけではないという結論になりうる。一方で、もし全体的処理が行われるとすれば、複数の構成要素を同時に提示した時に得られる表情印象は、構成要素が個別に（独立に）処理されると仮定した時、すなわち確率加重モデルに基づいて推定される表情印象よりも強くなることが予測される。本研究では、この点を検討することを主な目的とした。

2 実験

2-1 実験1

2-1-1 目的

表情を判断する際、顔の全体的表現に基づき全体的処理を行っているのか、あるいは、顔の構

成要素に注意を向けて局所的な処理を行っているのか、という疑問を検討するために、まず実験 1 で、顔の構成要素（目、口）がもたらす感情的な印象への感度を推定する。

2-1-2 方法

2-1-2-1 実験参加者

日本女子大学学生 20 名（20 ～ 22 歳）が参加した。すべての実験参加者が矯正視力を含む正常な視力を有していた。

2-1-2-2 装置

実験にはノート型コンピュータ（EPSON Endeavor NJ5950）を用いた。視覚刺激画像提示および反応の取得は、心理実験遂行用ソフトウェア SuperLab5.0（Cedrus 社）を用いた。

2-1-2-3 視覚刺激

表情データベース（ATRDB99, ATR-Promotion 社）より、ATR により実施された表情判断の得点が高いモデル（男女 1 名ずつ）を選んだ。各モデルから、それぞれ 2 種類（笑顔、怒り顔）の表情画像を選んだ。これらの画像からまず、顔の構成要素（目および口）を切り抜いて、500 × 500 ピクセルの白い背景に、構成要素を 1 枚の刺激画像につき 1 つずつ顔の位置関係を保った状態で貼り付けた画像を作成した。更に、Adobe Photoshop（Adobe 社）でコントラスト値を調整し、元の刺激画像を含めて 5 段階のコントラスト値で刺激画像を作成した（表 1）。画像のコントラストは、トーンカーブの出力値に基づいている。これはピクセルにおける輝度値の分布を変える操作であり、画像全体の平均的なコントラストを操作したことになる。以上を正立の刺激画像とし、同様に、これらを 180 度反転させたものを倒立の刺激画像とした。よって、男女モデル 2 種類 × 表情（笑顔あるいは怒り顔）2 種類 × 構成要素（目あるいは口）2 種類 × コントラスト値 5 段階 × 向き（正立あるいは倒立）2 種類の合計 80 枚の刺激画像が作成された。

表 1 表情画像のコントラスト値

トーンカーブの 出力値	コントラスト
230	0.1
180	0.3
140	0.45
100	0.61
0	1

2-1-2-4 手続き

実験では、実験参加者に対してこれらの刺激画像 80 枚をランダムな順番で、200 ミリ秒提示した。実験参加者は、強制 2 肢選択法によって、提示された刺激が笑顔だと判断したらキーボードの Y を押し、怒り顔だと判断したらキーボードの N を押した。正誤のフィードバックは行わなかった。この動作を、20 試行毎に休憩を挟みながら 80 試行続けた。実験は、十分な明るさのある室内で両眼視で行った。なお、実験前に練習試行を行った。練習試行は、表情データベース

(ATRDB99) から実験で使用したものとは異なる表情の画像を用いて作成した刺激画像を使用し、提示した。刺激画像の枚数を20枚とし、休憩は挟まなかった。

2-1-2-5 データ処理

得られる実験データは、各画像のタイプ（モデルのタイプ・笑顔／怒り顔の目あるいは口・正立／倒立）に応じて、コントラスト値に対する正答率となる。全実験参加者のデータをまとめ、以下のように処理する。まず、ある正答率（閾値）が得られるコントラスト値をもとめるために、得られた実験データにワイブル関数を非線形近似した。ワイブル関数は、以下の式で表される。

$$y = 1 - 0.5 \times e^{\left\{ -\left(\frac{x}{\alpha} \right)^{\beta} \right\}} \quad (\text{式1})$$

各変数は、

y: 表情の種類に関する正答率

x: 刺激画像のコントラスト

α : 閾値（正答率0.81になるコントラスト値）

β : ワイブル関数の傾き を表す。

非線形近似から得られた閾値 α （正答率0.81（81%）になるコントラスト値）に基づいて結果の考察を行った。

2-1-3 結果

実験参加者20名のそれぞれの正答数から、平均正答率を求め、ワイブル関数を用いて非線形近似した。非線形近似した結果は以下の通りである（図1）。

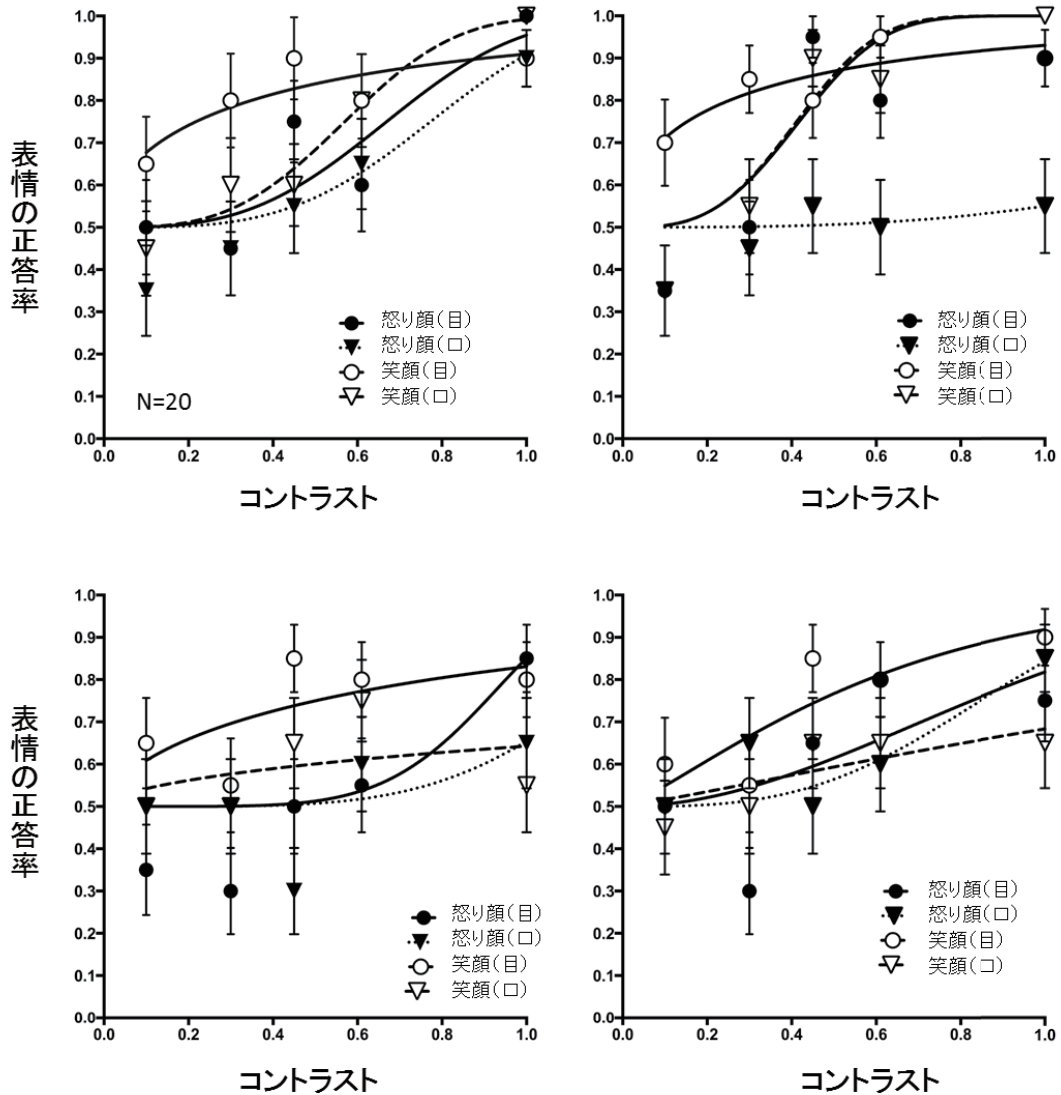


図1 実験1の結果。縦軸は構成要素別（目あるいは口，笑顔あるいは怒り顔）の正答率，横軸はコントラストを示す。上段は正立画像，下段は倒立画像の結果。曲線は近似したワイブル関数（式1），誤差棒は標準誤差を示す。

図1は，上段左が男性モデル正立画像の構成要素別（目／口・笑顔／怒り顔）の正答率，上段右が女性モデル正立画像の構成要素別の正答率，下段左が男性モデル倒立画像の構成要素別の正答率，下段右が女性モデル倒立画像の構成要素別の正答率を示したグラフである。正立画像・倒立画像ともに，非線形近似の曲線を見ると，どの構成要素でもコントラストが高くなるにつれて正答率は高くなっている。正立画像に関しては，男女ともに，笑顔（目）はコントラストが低くても高い正答率を示し，コントラストが高くなると正答率は高くなるものの，怒り顔（目）・笑顔（口）ほどの急激な上昇は見られない。コントラストが低いときは，笑顔（目）の方が笑顔（口）

よりも正答率が高かったが、途中で逆転し、コントラスト1.0では笑顔（口）の方が笑顔（目）よりも正答率が高くなっている。倒立画像では、男女ともに、正立画像に比べると全体的に低い正答率を示している。

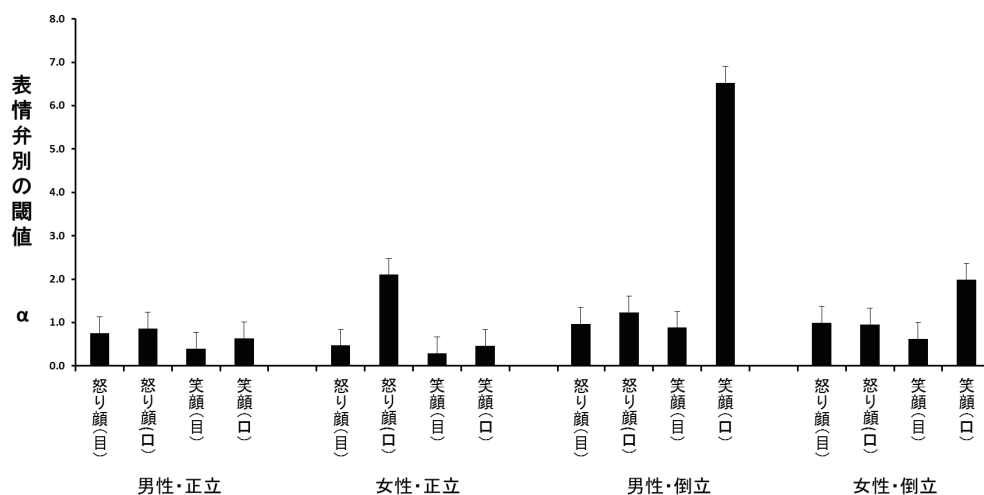


図2 各条件における表情弁別の閾値 α 。構成要素別の結果。誤差棒は標準誤差。

図2は、非線形近似により得られた α 値の結果である。 α 値は正答率が0.81（81%）になるコントラスト値である。よって、この値が小さい方が、より淡いコントラストでも表情がわかることを意味する。男性モデル倒立画像の笑顔（口）や女性モデル正立画像の怒り顔（口）の感度が低く、特に男性モデル倒立画像の笑顔（口）は際だって感度が悪い。逆に、男性モデル正立画像の笑顔（目）や女性モデル正立画像の怒り顔（目）・笑顔（目）は感度が高かった。

2-1-4 考察

ワイブル関数を用いて非線形近似したグラフを見ると、正立・倒立に関係なく、コントラストが高くなるにつれて正答率は高くなっており、表情判断は正確に行われている。正立画像においては、男女ともに笑顔（目）はコントラストが低くても正答率は高く、口よりも目がもたらす感情的な印象の方が強いように解釈できるが、男女ともにコントラスト1.0で最も高い正答率であるのは笑顔（口）である。このことから、感情的な印象を強くもたらすのは目であるとは言い切れないことがわかる。

2-2 実験2

2-2-1 目的

実験1で得られた結果から、確率加重モデルに従い、顔の構成要素が独立に印象をもたらすと

仮定した場合の予測を求める。そして、顔全体がもたらす感情的な印象が、確率加重モデルによる予測（すなわち全体的処理は行われていないとする予測）と合致するかどうかを検証する。また、倒立によってわかりにくくなる効果が特に強いのは2次処理による顔の区別で、全体処理と比べると部分処理では弱い。そのため倒立画像においては、個々の構成要素が独立に処理され、確率加重モデルと合致する結果が得られるのではないかと予測する。

2-2-2 方法

2-2-2-1 実験参加者

実験1と同じであった。

2-2-2-2 装置

実験1と同じであった。

2-2-2-3 視覚刺激

実験1と同じモデルの表情画像を用いた。これらの画像からまず、顔の構成要素（目および口）を切り抜いて、500 × 500 ピクセルの白い背景に、構成要素を1枚の刺激画像に顔の位置関係を保った状態で貼り付けた画像を作成した。このとき、構成要素の組み合わせは、「怒り顔の目 + 怒り顔の口」あるいは「笑顔の目 + 笑顔の口」という構成要素が矛盾しないものに加えて、構成要素が矛盾する「怒り顔の目 + 笑顔の口」「笑顔の目 + 怒り顔の口」という合計4種類とした。ただし、構成要素が矛盾する刺激画像の結果については、本稿では言及しない。実験1と同じように Adobe Photoshop でコントラスト値を調整し、元の刺激画像を含めて5段階のコントラスト値で刺激画像を作成した（表1）。以上を正立の刺激画像とし、同様に、これらを180度反転させたものを倒立の刺激画像とした。よって、男女モデル2種類 × 構成要素の組み合わせ4種類 × コントラスト値5段階 × 向き（正立あるいは倒立）2種類の合計80枚の刺激画像が作成された。

2-2-2-4 手続き

実験1と同様、実験参加者は強制2肢選択法により、提示された刺激が笑顔だと判断したらキーボードのYを押し、怒り顔だと判断したらキーボードのNを押した。正誤のフィードバックは行わなかった。

2-2-2-5 データ処理

構成要素の矛盾しない組み合わせの場合（「怒り顔目 + 怒り顔口」「笑顔目 + 笑顔口」）は実験1と同様に正答率を求めた。そして実験1のデータに確率加重モデルを当てはめ、実験2で利用した合成画像への正答率を予測した。続いて、実験2で得た実際のデータがこの予測と合致するかどうかを調べた。実験1（図1、図2）で求めた二組の α と β （それぞれ (α_1, β_1) (α_2, β_2) とする）から、目と口の画像を同時に提示した時に得られる正答率に関する予測を、以下の式により求めた。

$$y = 1 - 0.5 \times e^{-\left\{\left(\frac{x}{\alpha_1}\right)^{\beta_1} + \left(\frac{x}{\alpha_2}\right)^{\beta_2}\right\}} \quad (\text{式2})$$

各変数は、

y: 表情の種類に関する正答率 (予測)

x: 刺激画像のコントラスト

α_1, β_1 : 目の画像に対する閾値 (正答率 0.81 になるコントラスト値) と関数の傾き

α_2, β_2 : 口の画像に対する閾値 (正答率 0.81 になるコントラスト値) と関数の傾きを表す。

式2は、目に関する情報と口に関する情報が独立に処理されるという仮定の元、目と口を同時に提示した時の正答率がどのように変化するかを予測するモデルである。式2では確率加重 (probability summation) を計算しており (Wuerger, Hofbauer, & Meyer, 2003), 目と口が独立に処理されているとすれば (つまり全体的処理が行われていないとすれば), 実験2の結果はこのモデルに従うと考えられる。

2-2-3 結果

以下に示すグラフ (図3) は、非線形近似の結果 (図1, 図2) から「目+口」画像における正答率を確率加重 (式2) により予測し、さらに「目+口」画像における実験2のデータを加えたグラフである。

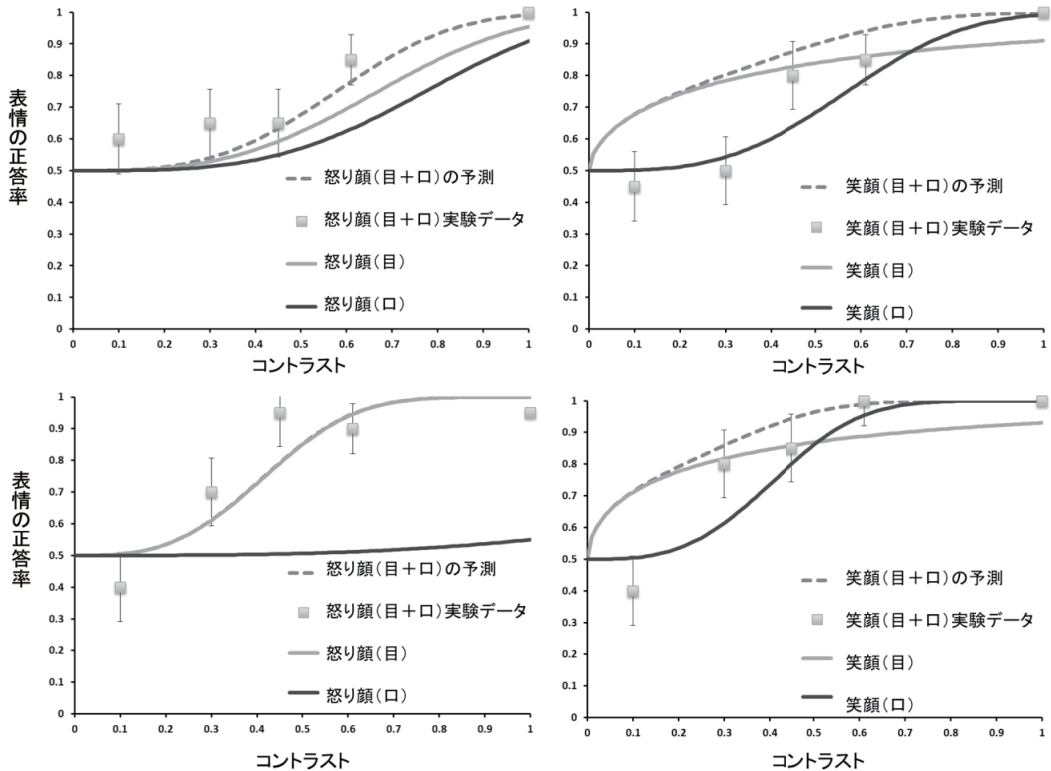


図3 正立画像における実験2の結果と実験1からの予測値。横軸はコントラスト、縦軸は正答率を示している。3種類の関数はそれぞれ、目の正答率（実験1）、口の正答率（実験1）、目と口を同時に提示した時の予測（式2の確率加重モデルより）を示している。実験データは、実験2（目と口を同時に提示）の結果。誤差棒は標準誤差。

図3は、上段左が男性モデル正立画像の怒り顔の実験データと予測の正答率、上段右が男性モデル正立画像の笑顔の実験データと予測の正答率、下段左が女性モデル正立画像の怒り顔の実験データと予測の正答率、下段右が女性モデル正立画像の笑顔の実験データと予測の正答率を示したグラフである。グラフは、それぞれの画像の「目」「口」「目+口」の実験データ、「目+口」の予測の正答率を示したグラフである。縦軸は表情の種類の正答率、横軸はコントラストである。点線は「目+口」の予測、灰色の曲線は「目」の実験データ、黒色の曲線は「口」の実験データ、四角い点は「目+口」の実験データを示す。

どの画像でも、コントラストが高くなるにつれ正答率は高くなっている。正立の怒り顔の場合、男女とも、コントラスト0.2あたりから「目+口」の正答率の予測が高くなる。「目+口」の実験データと予測値はほぼ同じであった。正立の笑顔の場合、怒り顔と比べるとコントラストが比較的低いときから「目+口」の正答率の予測が高くなっている。怒り顔と比べ低いコントラストで「目+口」の正答率の予測の値が極めて1に近くなっている。予測値は、「目+口」の実験データよりも、口だけの結果に近かった。

また図示はしていないが、倒立画像の場合は、怒り顔も笑顔も正立の場合よりも正答率の予測

の値は低く、実際のデータが確率加重の予測とほぼ合致していた。

2-2-4 考察

以上の結果をまとめる。まず一点目は、正立の怒り顔の場合である。正立の怒り顔の場合、「目+口」の実験データと予測値はほぼ同じであり、確率加重と合致していると考察できる。ただし、女性モデルでは、「目+口」の予測値と目の正答率がほぼ一致しており、コントラストが高くなっても口の正答率がほとんど高くない。よって、このモデルにおいては口の画像が表情判断をし難いものであり、口の正答率以外のデータがほぼ一致してしまうという結果が得られたのではないかと考察する。

二点目は、正立の笑顔の場合である。正立の笑顔の場合、予測値は、「目+口」の実験データよりも、口だけの結果に近かった。これは、表情の認識における二つの仮説（全体的処理仮説、確率加重仮説）とも合致しない結果である。この結果をもたらした一つの可能性は、たとえ複数の構成要素が提示された場合でも、実験参加者が目か口どちらかだけを見て判断したということである。図3を見ると、「目+口」の実験データの正答率は、コントラストが高くなるにつれて、口の正答率に近付いたことから、実験参加者は顔全体を見て判断したというよりも、笑顔の場合には口だけを見て、あるいは限りなく口に重点を置いて判断したことが考察できる。

三点目は倒立画像の場合である。「男性モデルの怒り顔」を除き、他のすべての倒立顔で同じ結果が得られていた。つまり倒立顔では、実験データと確率加重モデルからの予測に差はなかった。つまり、目と口が独立に処理されており、確率加重で説明できる。よって全体的処理は行われていないと結論づけられる。また、同じモデルの同じ表情で比べた場合、正答率の予測の値は正立の場合の方が高かったことから、倒立効果が出ていることが考察できる。

以上から、表情の種類が同じ構成要素を提示した場合、全体的処理が行われるという仮説は支持されなかったと言える。

3 総合考察

冒頭に述べた、表情を判断する際に、顔の全体的な表現に基づいているのであろうか、それとも、顔の構成要素に注意を向けるといった、より局所的な処理を行っているのであろうかといった疑問に対しては、以下のような結果が得られた。

まず、実験1より、顔の構成要素（目、口）がもたらす感情的な印象は、一般に口よりも目の方が強かった。加藤（1998）の顔の魅力度判断におけるパーツの魅力の影響の研究では、口は魅力度の高い顔の重要なパーツではなく、目はかなり重要なパーツであることが示されている。ただし、独立に提示したとき正立画像の笑顔の口というのは目よりも強い印象を与えていた。これは、口角の上がっているわかりやすい笑顔の口であり、怒り顔の口との区別がしやすかったことが関係すると考えられる。

実験2より、表情の種類が同じ構成要素を提示した場合、正立の怒り顔の場合、確率加重と合致していたことから、局所的処理が行われていた可能性が高いといえる。正立の笑顔においては、目と口を両方出すと表情の判断が難しくなった。目と口を同時に提示した実験データの正答率と

口の実験データの正答率が次第に重なったことから、笑顔の表情では、口の印象がとても強く、目から得られる情報がなくとも表情が笑顔だと判断することができるといえる。

今後の課題として、なぜ正立の笑顔の場合は本来正しい組み合わせの顔であるにも関わらず同時に提示すると表情がわかりにくくなったのか、別のモデルを使用して実験を行うことで確認していく必要がある。

また、本研究では、視覚刺激が妥当ではなかった可能性が考えられる。まず、顔の構成要素（目・口）の位置関係が離れすぎていた点である。また、輪郭を提示しなかったため、輪郭が表情認識の際に与えた影響も考慮する必要がある。加藤（1998）によれば、魅力度の高い顔の場合に、顔の輪郭は魅力度に影響を及ぼすことが指摘されている。構成要素以外の部分をぼかした画像を作成し視覚刺激を改善することで、輪郭が追加され、パーツの位置関係もオリジナルを保てるであろう。笑顔と怒り顔以外にも顔の表情の種類を増やし、恐れ、驚き、悲しみの表情などを検討することも視野に入れる。これらの結果から、感情印象生成における顔の実態について理解が進むことが期待される。

謝辞

本論文を作成するにあたり、指導教授の竹内龍人先生から、丁寧かつ熱心なご指導を賜りました。ここに感謝の意を表します。

[引用文献]

- Bartlett, J. C., & Seacy, J. (1993) Inversion and configuration of faces. *Cognitive Psychology*, 25, 281-316.
- Cabeza, R., & Kato, T. (2000) Features are also important: Contributions of featural and configural processing to face recognition. *Psychological Science*, 11, 429-433.
- Freire, A., Lee, K., Symons, L. A. (2000) The face-inversion effect as a deficit in the encoding of configural information: Direct evidence. *Perception*, 29, 159-170.
- 加藤隆・阿磨大介・森岡久美子・赤松茂(1998) 顔の魅力度判断におけるパーツの魅力の影響, 信学技報, HIP97- 53(1998-03)
- Maurer, D., Le Grand, R., & Mondloch, C. (2002) The many faces of configural processing. *Trends in Cognitive Sciences*, 6, 255-260.
- Tanaka, J. & Gordon, I. (2011) Features, configuration, and holistic face processing, In A. J. Calder, G. Rhodes, M. H. Johnson, J. V Haxby (Eds.), *The Oxford Handbook of Face Perception*, New York, NY: Oxford University Press.
- Wuerger, S., Hofbauer, M., & Meyer, G. F. (2003) The integration of auditory and visual motion signals at threshold. *Perception & Psychophysics*, 65, 1188-1196.