

ランドマークで表現した顔表情を認識する際に意識的な観察行動の効果の検討

Investigating the Effect of Conscious Observational Behavior on Recognizing Facial Expressions Represented by Landmarks

徐 旣哲[†], 新川 広樹[†], 松香 敏彦[‡]

Kuangzhe Xu, Hiroki Shinkawa, Toshihiko Matsuka

[†] 弘前大学, [‡] 千葉大学

Hirosaki University, Chiba University

jokyotetsu@hirosaki-u.ac.jp

概要

本研究では、顔の機械学習で一般的に使用されるランドマークで表現された顔の動画刺激を使用し、人々がどの程度表情を認識できるか検討した。また、異なる表情を認識するために、どのような意識的な観察行動の違いがあるのか検討した。刺激として、あらかじめ6種類の表情（恐れ、怒り、嫌悪、悲しみ、喜び、驚き）の動画を68個の特徴点に変換したものを使用した。実験では全ての68の特徴点を表示するのではなく、意識的な観察行動を記録するため、マウスのカーソルの周辺の一部しか見えないように設定した。分析の結果、人々は68点で表現されている顔情報しか得られない状況下でも、ポジティブな表情に対しては機械学習と同様に高い精度で認識できることがわかった。一方、決して高いとは言えないものも、ネガティブな表情に対しては、機械学習よりも優れた認識精度を示した。また、観察者の性格特性と意識的な観察行動は、異なる表情の判断に有意に影響することも明らかにになった。

キーワード：表情認知, 意識的な観察行動, 顔表情, ベイズモデル

1. はじめに

表情認知に関する研究は数多く行われており、特に近年では観察行動との関係性についても多く検討されてきた。しかし、まだ解決していない疑問もいくつか残されている。まず、表情認知における正答率について、人間と機械学習では異なる結果が報告されている。機械学習では喜びや驚きなどのポジティブな表情に対しては人間と同様に高い正解率を示したが、悲しみや恐れなどのネガティブな表情では正確に認識できていない [1, 2]。その原因として、機械学習は人間の

ように画像や動画で目に見える情報をそのまま利用して分析するのではなく、ランドマークの抽出などを通じて顔の特徴を用いて分類していることが挙げられる [3]。しかし、このようなランドマークに基づく顔の特徴が人間でも正確に表情認知できるのか否かは、これまで検討されていない。

次に、表情認知と観察行動の関係性について、先行研究 [4] の結果により、視線が操作されても一部の性格特性と強く相関していることが示された。観察行動はおおよそ操作の指示に従ったが、実験参加者のほとんどが非常に短い時間においてその指示に従わなかったことが示された。つまり、眼球運動は意識的と無意識的の2つの種類が存在する可能性が示唆された。これまで多くの研究で使用してきたアイトラッカーで記録された視線データには、これら2つの種類の視線が混在している可能性があると考えられる。一方、実際に人間が表情認知を行う際の意識的な観察行動がどのように異なっているかについては、まだ検討されていない。

本研究では、前述の疑問点を検討するために、ランドマークで表現した顔の刺激動画を作成した。それらの刺激動画に対して、人間がどれだけ正確に認知できるかを行動実験によって検討した。さらに、人が異なる表情を認識する際の意識的な観察行動が表情判断に与える影響も検討した。

2. 実験

実験参加者：弘前大学大学生 41 名（男性 12 名、女性 29 名）が実験に参加した。

刺激画像：本実験では、AIST データベース [5] から 8 名の役者（男性 4 名、女性 4 名）による喜び、悲しみ、怒り、嫌悪、驚き、恐れ の 6 種類の動画表情を使用した。Dlib を用いて顔のランドマークのみを表示

する動画表情を作成した（ランドマーク数は先行研究を参考にして68個を採用した[3]）。

実験手続き：実験の各セッションの冒頭に、「6種類の表情の中から最も当てはまるものを選んでください」という指示を表示した。参加者は自身のタイミングでマウスをクリックし、0.5秒の注視記号「+」の後にランダムに刺激動画が呈示された。観察時間は固定せず、参加者が判断可能になるまで自由に観察することができた。マウスをクリックすると観察が終了し、評価ステージに移行した。評価ステージでは、呈示された動画に最も当てはまる項目の一つを選択するように求められ、その後、7段階での確信度の評価を求めた。評価が終了すると、次のセッションに移行した。

課題開始直後、カーソルの周辺部分以外はマスクで隠されていた。参加者は観察したい部位にマウスのカーソルを合わせることによって確認することができた。この仕様により、観察したい部位に移動されたカーソルの動きが「意識的な観察行動」を反映していると考えられる。

全てが終了後、参加者には日本語版の性格特性チェックリストへ回答を求め、各自の性格特性情報を測定した。

データ処理：意識的な観察行動の情報が反映されたカーソルの動きデータを整理した。具体的には、psychopyで記録されたカーソルの座標データをPCのスクリーン解像度に変換し、画面上の刺激の位置と重ね合わせた。その後、先行研究[6]を参考にして、標準偏差が10ピクセルのガウスフィルターを適用した。得られたデータにはエリアマスクを適用し、特定のエリアでの重みデータを算出した。エリアマスクは先行研究[7]を参考にし、顔の上部、中部、下部の三つのエリア（上：中：下=1:0.8:0.6）に分類した。

3. データ解析

3.1 分析 A. 表情判断と観察行動

表情判断の選択肢は全部6個があるため、その特徴から階層的バイズ多項ロジスティックモデルを構築した。

$$\prod_{1 \leq k \leq 6} \text{Bernoulli}(y_k | \text{logit}^{-1}(\mu_{ij})) \quad (1)$$

$$= \prod_{1 \leq k \leq 6} \begin{cases} \text{logit}^{-1}(\mu_{ij}) & \text{if } y_k = 1, \text{ and} \\ 1 - \text{logit}^{-1}(\mu_{ij}) & \text{if } y_k = 0. \end{cases} \quad (2)$$

$$\mu_{ij} = \alpha_{kg} + \sum_{1 \leq g \leq 3} G_{ijg} \cdot \beta_{kg} + r_{ig}^{subj} + r_{jg}^{stim} \quad (3)$$

上記式の k は選択された表情（喜び、悲しみ、怒り、嫌悪、驚きと恐れ）、 μ_{ij} は多項ロジスティックを推定するためのパラメータである。 i, j はそれぞれ参加者と刺激動画のインデックスを示している。 G_{ijg} は参加者 i の刺激 j への観察行動 g （上部、中部、または下部）を表している。 r_{ig}^{subj} と r_{jg}^{stim} はそれぞれ参加者と刺激動画のランダムエフェクトを表している。

モデル推定は Rstan で行った。固定効果とランダム効果を推定するための弱事前情報はそれぞれ Normal(0,10) と Gamma(10,10) に設定した。有意差判定は HDI を用いた。

3.2 分析 B. 表情判断と性格特性

目的変数は表情選択のみであるため、分析 A と同様に多項ロジスティックモデルを用いた。ただし、説明変数に観察行動を用いるのではなく、それらを性格特性に置き換えた。その式は以下になる。

$$\mu_{ij} = \alpha_{kp} + \sum_{1 \leq p \leq 5} P_{ip} \cdot \beta_{kp} + r_{ip}^{subj} + r_{jp}^{stim} \quad (4)$$

P_{ip} は参加者 i の性格特性 p (bigfive) を表している。モデル推定も Rstan で行い、推定するための弱事前情報と各パラメータは分析 A と同様に設定した。

4. 結果

図1はモデルによる各刺激動画の表情に対する選択傾向の推定結果を示している（破線はチャンスレベルを示している）。ポジティブな表情（喜びと驚き）に対しては正しく認識する傾向が見られたが、ネガティブな表情を正確に認識する傾向は高いとは言えない。特に恐れ表情に対して、喜びや驚きと誤って認識する確率はチャンスレベルを上回っていた。

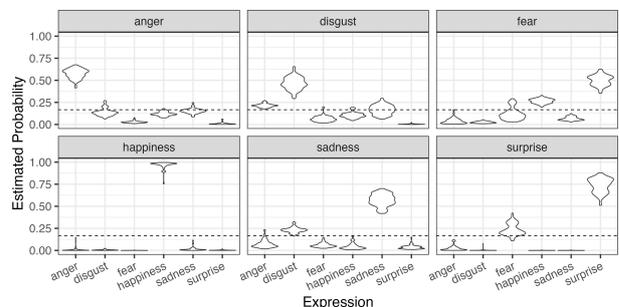


図1 表情に対する推定選択傾向の事後分布

表情判断と観察行動：観察行動と表情判断の有意な関係性を表1にまとめた。結果から、顔の中央部を

長く見ると恐れを正確に認識する傾向が高いことが示された。一方、顔の下部を長く見ると喜びを悲しみや嫌悪に誤認する傾向があり、顔の上部を長く見ると恐れを驚きと誤認する傾向が示された。また、顔の下部を長く見ると喜びを正しく認識しない傾向や、顔の上部を長く見ると喜びや怒りを正確に認識しない傾向も確認された。

表1 観察行動と表情判断の有意相関表

Emotion	Judgment	Area	Mean	95%HDI
Happiness	Happiness	Upper	-767.144	-1238.152 ~ -289.434
		Low	-1016.051	-1650.876 ~ -481.775
	Disgust	Low	856.070	61.954 ~ 1810.830
		Sadness	Low	1098.640
Fear	Fear	Middle	161.977	12.126 ~ 320.643
	Surprise	Upper	164.602	12.879 ~ 315.883
	Happiness	Middle	-258.489	-427.295 ~ -104.806
Anger	Anger	Upper	-230.716	-391.424 ~ -76.173
	Happiness	Middle	-280.260	-508.078 ~ -10.772

表情判断と性格特性：表2では、観察者の性格特性と表情認識の間に示された結果を表情ごとにまとめた。結果から、協調性が高い人ほど喜びと驚きを正しく認識する傾向が示された。また、外向性が高い人ほど驚きを正しく認識する傾向が見られた。一方、勤勉性と外向性が高い人は悲しみを喜びや怒りと誤認する傾向があり、神経質性が高い人は恐れを驚きと誤認する傾向が示された。さらに、外向性が高い人は悲しみを悲しみと認識しない傾向があり、神経質性が高い人は恐れを恐れと認識しない傾向も示された。

表2 性格特性と表情判断の有意相関表

Emotion	Judgment	Predictor	Mean	95%HDI
Happiness	Happiness	Agree.	0.812	0.395 ~ 1.321
	Anger	Agree.	-1.472	-2.375 ~ -0.634
	Sadness	Agree.	-0.736	-1.376 ~ -0.197
Sadness	Happiness	Consc.	0.385	0.004 ~ 0.771
		Neuro.	-0.353	-0.693 ~ -0.025
	Anger	Extra.	0.346	0.044 ~ 0.670
	Sadness	Extra.	-0.184	-0.346 ~ -0.004
	Surprise	Neuro.	-0.410	-0.774 ~ -0.094
Surprise	Surprise	Agree.	0.218	0.001 ~ 0.443
		Extra.	0.266	0.059 ~ 0.465
	Anger	Extra.	-0.881	-1.547 ~ -0.271
Anger	Fear	Agree.	-0.255	-0.484 ~ -0.031
Disgust	Surprise	Neuro.	-0.701	-1.320 ~ -0.084
Fear	Disgust	Consc.	-0.304	-0.552 ~ -0.045
		Neuro.	0.171	0.013 ~ 0.344
	Anger	Extra.	-0.620	-1.089 ~ -0.179
		Fear	Neuro.	-0.339

5. 考察

本研究では、64点の顔のランドマークで表現された表情の認識について検討した。その結果、ポジティブな表情の認識精度は高く、ネガティブな表情の認識精度は低いことが確認された。この結果は、深層学習を用いた表情認知の結果と一致している。

また、意識的な観察行動からは、人々が異なる表情に対し特定の特徴部位に着目することが分かった。具体的には、ポジティブな表情では下部、ネガティブな表情では上部が重要な役割を果たしていることが示された。つまり、口と目は異なる表情を表現するのに関与していると考えられる。

しかし、本研究では刺激動画の観察時間に制限を設けておらず、各表情における特定の部位への優先度を明確にすることができなかった。今後は呈示時間を制限し、有限な時間内で各表情をどのように観察するかを再度確認することが重要である。このような調査を行うことで、表情認知における部位の優先度や重要性についてさらなる理解が深まるだろう。

文献

- [1] Minaee S, Minaei M, Abdolrashidi A. Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network. *Sensors* (Basel). 2021 Apr 27;21(9):3046. doi: 10.3390/s21093046.
- [2] Tzirakis, P., Trigeorgis, G., Nicolaou, M.A., Schuller, B., Zafeiriou, S. End-to-End Multimodal Emotion Recognition Using Deep Neural Networks. *IEEE Journal of Selected Topics in Signal Processing*. 2017, 11, 1301-1309. DOI: 10.1109/JSTSP.2017.2764438.
- [3] Sagonas, C., Tzimiropoulos, G., Zafeiriou, S., Pantic, M. 300 Faces in-the-Wild Challenge: The First Facial Landmark Localization Challenge. 2013 IEEE International Conference on Computer Vision Workshops, Sydney, NSW, Australia, 2013, pp. 397-403, doi: 10.1109/ICCVW.2013.59.
- [4] Xu, K., Matsuka, T. Effects of observers' characteristics on impression formation on face. *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. 2018, 1293-1298. DOI: 10.23919/APSIPA.2018.8659469.
- [5] Fujimura, T., Umemura, H. Development and validation of a facial expression database based on the dimensional and categorical model of emotions. *Cognition and Emotion*. 2018, 32:8, 1663-1670, DOI: 10.1080/02699931.2017.1419936
- [6] Caldara, R., Miellet, S. iMap: a novel method for statistical fixation mapping of eye movement data. *Behavior Research Methods*. 2011; 43, 864-878. DOI: <https://doi.org/10.3758/s13428-011-0092-x>
- [7] Calvo, M.G., Fernández-Martín, A., Gutiérrez-García, A. et al. Selective eye fixations on diagnostic face regions of dynamic emotional expressions: KDEF-dyn database. *Sci Rep* 8, 17039 (2018). <https://doi.org/10.1038/s41598-018-35259-w>