

孤立した顔パーツにおける情動認識

桑原 和也・野中 春花・山崎 美優

大正大学心理社会学部人間科学科

指導教員：井関 龍太

要旨：本研究は孤立した顔のパーツから情動を認識することが可能であるか明らかにすることを目的とした。実験参加者に孤立した目・鼻・口・耳と顔全体を提示し、人物がどのような情動を表出しているのか7つの情動（怒り・恐怖・幸せ・悲しみ・嫌悪・驚き・中立）から選択させた。実験の結果、口・顔全体の正答率がそれぞれ偶然の確率を上回った。このことから、口からは孤立した状態でも情動の認識が可能であることが明らかになった。

問 題

新型コロナウイルスの感染拡大を経て、近年は日頃からマスクを着用することが多くの人にとって習慣となっている。マスクを着用していると鼻や口は隠され、目元のみ露出している状態となる。このように顔全体が見えず1つのパーツのみが孤立して提示された状態で私たちはお互いの情動をうまく認識することができるのだろうか。

孤立した顔の一部分における情動認識に関する研究として Wegrzyn et al. (2017) がある。この研究は、顔の情動の認識が特定のパーツの特徴に依存するかどうかを明らかにすることを目的として行われた。刺激となったのは幸せ・悲しみ・怒り・恐怖・嫌悪・驚きと中立的な表情の6種類のうちいずれかを示す男性または女性の顔写真で、顔をタイル状に区切ってランダムに一部分を提示する実験と黄色や緑のマーカーで顔の一部分が覆われた状態で提示される実験の2種類があった。実験参加者の課題は、刺激として提示された顔写真が6種類の情動のうちいずれに当てはまるのかを推測することであった。結果からは、情動の表出を上手く認識するためには目と口が非常に重要であることが分かった。また、類似したテーマの研究として Blythe et al. (2023) が挙げられる。この研究は、孤立した身体の一部から情動を認識することができるか明らかにすることを目的として行われた。刺激となったのは顔が隠されて表情が見えない状態の10人の人物画像で、それぞれが怒り・恐怖・

幸福・悲しみ・嫌悪・驚きを示していた。各画像は手・腕・頭・胴体の4つの身体パーツに切り分けられた。実験参加者の課題は提示された身体パーツが6つの情動のうちいずれに当てはまるのかを推測することであった。結果からは、孤立した身体パーツには情動を認識するのに十分な情報が含まれていることが示唆された。また、身体全体が表す情動の方が身体パーツが表す情動よりも正確に認識することが可能であった。身体パーツごとの比較では手から最も正確に情動を認識しやすいことが分かった。

本研究は、孤立した顔のパーツから情動を認識できるのか明らかにすることを目的として行われた。怒り・恐怖・幸福・悲しみ・嫌悪・驚き・中立のいずれかを示した孤立した目・鼻・耳・口と顔全体を刺激とし、それぞれの正答率を比較する。Wegrzyn et al. (2017) と Blythe et al. (2023) によると、顔の一部分のみを切り取って提示されたときや身体をパーツごとに切り取って提示されたときでも、情動を認識することができる。そこで本研究では以下の3つの仮説を検討する。1点目は、顔全体が見えていなくても顔のパーツから情動を認識できるというものである。本研究は刺激となる目・鼻・耳・口を輪郭に沿って切り取った状態で提示するという点で、顔をタイル状に区分けした状態で提示した Wegrzyn et al. (2017) とは異なる。顔のパーツを輪郭に沿って切り取り、赤みやしわなど周りの皮膚から得られる情報を減らし、それぞれのパーツのみ提示した際にも同様に情動を認識できるのかを検討することが目的である。顔全体が見えていなくても顔のパーツから情動を認識できるという仮説が正しければ、孤立した顔のパーツのそれぞれが情動の認知に十分な情報を持ち、偶然の確率以上の正答率となることが予想される。2点目は、顔全体からの情動の認識の方が提示される面積が大きくそこに含まれる情報も多いため、顔のパーツからの情動の認識よりも正確であるというものである。本研究では顔のパーツを輪郭に沿って切り取った状態で、顔全体からの情動の認識と顔のパーツからの情動の認識を比較する。この仮説が正しければ、刺激として顔全体が提示された試行の方が孤立した顔のパーツが提示された試行よりも正答率が高いことが予想される。3点目は、目と口が表す情動の認識が鼻・耳からの情動の認識よりも正確であるというものである。Wegrzyn et al. (2017) の結果から目と口が情動の表出をうまく認識するために必要であることはわかっている。本研究では顔のパーツを輪郭に沿って切り取った状態でも同様の結果になるのか検討する。目と口が表す情動の認識が鼻・耳からの情動の認識よりも正確であるという仮説が正しければ、孤立した顔のパーツごとの結果を比較した際に目と口の正答率が高くなることが予想される。

方 法

実験参加者

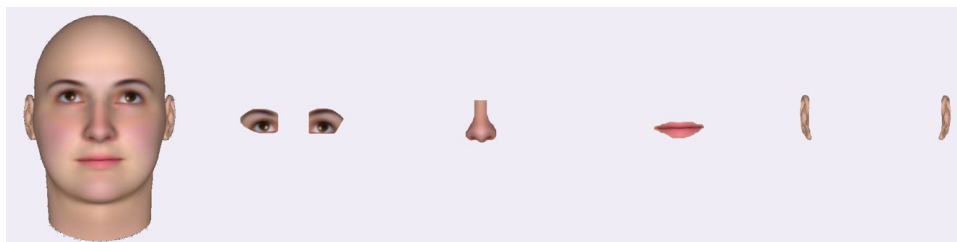
大学生の男性9名、女性13名、未回答1名の計23名が実験に参加した。参加者の平均年齢は21.05歳 ($SD = 1.58$) であった。

刺激と装置

刺激には Blythe et al. (2023) で用いられた画像を使用した。画像は BESST (Thoma et al., 2013) に収録されたものであった。画像は全部で 3,360 個あり、各画像には CG で合成された 80 名の男性または女性のうちいずれか 1 名の首から上が表示された。画像の人物 1 名につき、20 歳・40 歳・60 歳という 3 種類の年齢のバージョンがあり、それぞれが恐怖・嫌悪・幸せ・悲しみ・驚き・怒り・中立の 7 種類の情動を表した。また人物の表示角度は正面と右斜めの 2 種類があった。実験では 20 歳かつ正面の画像を用い、80 人からランダムに 10 人を選出して合計 70 枚の画像ファイルを抽出した。抽出した画像は、画像操作プログラムの GIMP を使用して目・鼻・口・耳を切り抜き、顔の一部分が表示されたパーツを作成した。孤立した目・鼻・口・耳と顔全体の計 5 種類の顔パーツを 10 人それぞれについて 7 種類の情動分用意した。その結果、刺激は全部で 350 個であった (図 1)。画像のサイズは 400 px×400 px であった。刺激の制御とデータの収集は lab.js (Henninger, Shevchenko, Mertens, Kieslich, & Hilbig, 2022) を用いて行われた。実験参加者は画面上のボタンをクリックして回答する必要があった。スマートフォンでは画面上に表示できる部分に限界があるため、参加者全員が各自のパソコンを使用して実験を行った。そのため提示された刺激の物理的な大きさは画面の大きさに応じて異なった。

図 1

顔パーツの例



注) 画像は例示のために新たに作成したもので実際に使用した刺激と異なる。

手続き

各試行は、画面中央に現れる注視点から始まり、その後刺激が提示された。注視点は 250 ms または 450 ms の間画面中央に提示された。刺激は白い背景に提示され、画面中央に顔画像、顔画像の上に「この画像は何の感情を表していると思いますか。」という文、顔画像の下に回答のための 6 つの情動の名称 (恐怖・怒り・嫌悪・幸せ・悲しみ・驚き) をそれぞれ枠で囲った長方形 (ボタン) を、それぞれ提示した (図 2)。

図 2

刺激提示画面の例



参加者は顔画像を見て、その画像が表していると思った情動のボタンをクリックするように教示された。刺激への反応には時間制限がなかった。出現する刺激の順序による回答への影響を統制するため、刺激の提示順序をすべてランダムにした。ボタンをクリックした後、再び 250 ms または 450 ms の間画面中央に注視点が出現した。以上の注視点から注視点までの行程を 1 試行とし、参加者は計 175 試行おこなった。

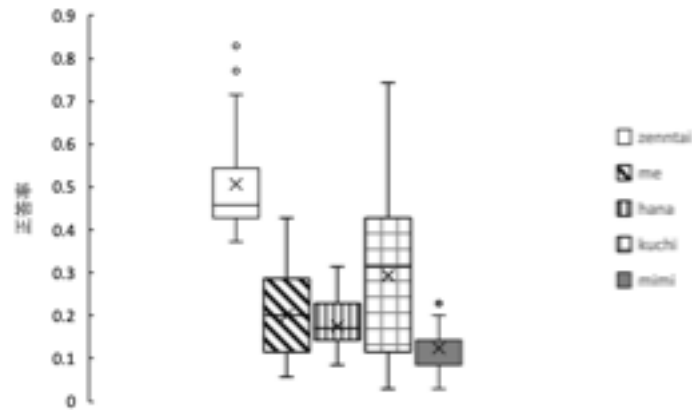
ボタンの並びによる回答への影響を統制するため、参加者を 2 つのグループに分けてカウンターバランスをとった。2 つのグループのボタンの並びは次の順序（左から右）であった。それぞれ「驚き・幸せ・悲しみ・恐怖・怒り・嫌悪」と「恐怖・怒り・嫌悪・驚き・幸せ・悲しみ」であった。一方のグループは 5 人の顔全体と他の 5 人の俳優の各パーツ（目・鼻・口・耳）が提示され、もう一方のグループは 5 人の各パーツ（目・鼻・口・耳）と他の 5 名の顔全体が提示された。そのため、参加者は各パーツの全体バージョンを見ることはなかった。刺激は 70 枚の顔全体、70 枚の目、70 枚の鼻、70 枚の口、70 枚の耳から構成された。これらの 70 枚の画像それぞれに、7 つの情動（中立・恐怖・怒り・嫌悪・幸せ・悲しみ・驚き）が描かれていた。実験は約 20 分で終了した。

結 果

図 3 に顔パーツごとの平均正答率を示した。本実験では、提示された刺激の情動と回答した情動が一致することを正答とし、顔パーツごとに正答をパーツごとの試行数である 35 で割った値を正答率とした。本実験では回答の選択肢が 6 つであったため、偶然正答する確率は $1/6$ (0.1666...) であった。以下の分析において、有意水準は 5% とした。これらのデータについて、対応のある t 検定を行った結果、顔全体の正答率は偶然を有意に上回った ($t(22) = 13.52, p < 0.001$)。目の正答率は偶然を有意に上回らなかった ($t(22) = 1.44, p = 0.16$)。鼻の正答率は偶然を有意に上回らなかった ($t(22) = 0.73, p = 0.48$)。口の正答率は偶然を有意に上回った ($t(22) = 3.13, p = 0.01$)。耳の正答率は偶然を有意に下回った ($t(22) = -3.85, p < 0.001$)。

図 3

顔パーツごとの正答率を示した箱ひげ図



パーツごとの正答率の比較のため、顔全体・目・鼻・口・耳を水準とした、一要因の参加者内分散分析を行った結果、パーツの主効果が見られた ($F(4, 88) = 46.59, p < 0.001, \eta_p^2 = 0.68$)。パーツの主効果について、Holm-Bonferroni 法を用いた多重比較を行った。その結果、顔全体の正答率は目の正答率を有意に上回った ($t(22) = 10.54$)。顔全体の正答率は鼻の正答率を有意に上回った ($t(22) = 13.03$)。顔全体の正答率は口の正答率を有意に上回った ($t(22) = 6.01$)。顔全体の正答率は耳の正答率を有意に上回った ($t(22) = 14.85$)。目と鼻の正答率の間に有意な差は見られなかった ($t(22) = 0.87$)。口の正答率は目の正答率を有意に上回った ($t(22) = 3.54$)。目の正答率は耳の正答率を有意に上回った ($t(22) = 2.90$)。口の正答率は鼻の正答率を有意に上回った ($t(22) = 2.61$)。鼻の正答率は耳の正答率を有意に上回った ($t(22) = 3.66$)。口の正答率は耳の正答率を有意に上回った ($t(22) = 3.95$)。

顔パーツごとに 7 つの情動に対しての正誤のパターンを調べるため、顔パーツごとの情動間の混同パターンを調べた。図 4~8 は顔パーツごとの情動間の混同行列、図 9 は混同行列 (図 4~8) の対角要素間の相関を示した。顔全体において、怒り、驚き、幸せ、嫌悪の正答率が高かった。恐怖は悲しみ、嫌悪を怒りとよく混同された。目において、怒りの正答率が高かった。恐怖は悲しみ、悲しみを驚き、幸せを悲しみ、嫌悪を怒りと混同された。中立は驚きとよく認識された。口において、幸せと嫌悪の正答率が高かった。恐怖は驚き、驚きを幸せとよく混同された。鼻と耳において、正答率が高い情動も混同された情動もなかったが、幸せと回答した割合が高かった。

図 4

顔全体の情動間の混同行列

顔全体						
怒り	0.34	0.02	0.02	0.07	0.52	0.02
恐怖	0.04	0.19	0.23	0.22	0.03	0.29
驚き	0.04	0.42	0.22	0.07	0.06	0.19
悲しみ	0.13	0.02	0.02	0.38	0.14	0.31
幸せ	0.58	0.03	0.03	0.01	0.32	0.02
嫌悪	0.17	0.02	0.01	0.36	0.26	0.18
中立	0.19	0.15	0.12	0.15	0.13	0.26
	怒り	恐怖	驚き	悲しみ	幸せ	嫌悪

図 5

目の情動間の混同行列

目						
怒り	0.04	0.09	0.15	0.19	0.52	0.01
恐怖	0.21	0.07	0.16	0.06	0.04	0.46
驚き	0.27	0.24	0.07	0.07	0.04	0.30
悲しみ	0.05	0.19	0.21	0.24	0.16	0.15
幸せ	0.09	0.27	0.05	0.26	0.17	0.16
嫌悪	0.04	0.10	0.02	0.39	0.39	0.06
中立	0.11	0.20	0.18	0.20	0.15	0.16
	怒り	恐怖	驚き	悲しみ	幸せ	嫌悪

図 6

鼻の情動間の混同行列

鼻						
怒り	0.27	0.16	0.07	0.12	0.31	0.07
恐怖	0.25	0.20	0.08	0.23	0.17	0.07
驚き	0.18	0.25	0.10	0.18	0.26	0.03
悲しみ	0.24	0.20	0.08	0.20	0.21	0.07
幸せ	0.20	0.21	0.03	0.22	0.28	0.06
嫌悪	0.22	0.15	0.09	0.19	0.25	0.10
中立	0.27	0.24	0.08	0.11	0.26	0.04
	怒り	恐怖	驚き	悲しみ	幸せ	嫌悪

図 7

口の情動間の混同行列

口						
怒り	0.37	0.13	0.05	0.25	0.12	0.07
恐怖	0.01	0.49	0.37	0.07	0.03	0.04
驚き	0.13	0.24	0.07	0.08	0.37	0.11
悲しみ	0.26	0.00	0.03	0.11	0.07	0.52
幸せ	0.64	0.03	0.00	0.20	0.07	0.06
嫌悪	0.14	0.04	0.04	0.46	0.29	0.03
中立	0.15	0.09	0.19	0.20	0.13	0.25
	怒り	恐怖	驚き	悲しみ	幸せ	嫌悪

図 8

耳の情動間の混同行列

耳						
怒り	0.27	0.28	0.10	0.12	0.15	0.06
恐怖	0.27	0.21	0.09	0.20	0.16	0.07
驚き	0.28	0.18	0.11	0.17	0.21	0.06
悲しみ	0.27	0.26	0.11	0.11	0.17	0.08
幸せ	0.24	0.20	0.12	0.13	0.20	0.10
嫌悪	0.29	0.20	0.09	0.16	0.17	0.09
中立	0.25	0.20	0.08	0.14	0.22	0.12
	怒り	恐怖	驚き	悲しみ	幸せ	嫌悪

注) 図 4~8 は、縦軸は提示された情動、横軸は参加者が回答した情動を示す。色は正答率の高低を表しており、色が濃いほど正答率が高い。

図 9

混同行列 (図 4~8) の類似性を示す表現類似度行列

目	鼻	口	耳	
0.40	0.35	0.54	-0.10	顔全体
	0.25	-0.01	-0.23	目
		0.21	0.62	鼻
			0.27	口

注) 色が濃いほど相関が正の方向である。

顔パーツごとの混同行列における類似度を定量化するために、表現類似度分析を行った。表現類似度分析とは、多変量解析によって求めた刺激に対する反応の類似性構造を抽出する方法である。その結果、顔全体と目の間に有意な相関が見られた ($t(40)=2.78, p=0.01$)。顔全体と鼻の間に有意な相関が見られた ($t(40)=2.32, p=0.03$)。顔全体と口の間に有意な

相関が見られた ($t(40) = 4.01, p < 0.001$)。顔全体と耳の間に有意な相関は見られなかった ($t(40) = -0.66, p = 0.51$)。目と鼻の間に有意な相関は見られなかった ($t(40) = 1.61, p = 0.11$)。目と口の間に有意な相関は見られなかった ($t(40) = -0.06, p = 0.95$)。目と耳の間に有意な相関は見られなかった ($t(40) = -1.51, p = 0.14$)。鼻と口の間に有意な相関は見られなかった ($t(40) = 1.41, p = 0.17$)。鼻と耳の間に有意な相関が見られた ($t(40) = 5.12, p < 0.001$)。口と耳の間に有意な相関は見られなかった ($t(40) = 1.78, p = 0.08$)。

考 察

本研究は、孤立した顔のパーツから情動を認識することが可能なのか明らかにすることを目的とした。本研究では3点の仮説を検証した。1点目は、顔全体が見えていなくても顔のパーツ特に目と口から偶然の確率以上に情動を認識できるという仮説であった。2点目は、顔全体からの情動の認識の方が顔のパーツからの情動の認識よりも正確であるという仮説であった。3点目は、目と口の情動の認識が鼻・耳よりも正確であるという仮説であった。実験の結果、顔全体と口からは偶然の確率以上に情動を認識することができた。目と鼻と耳からは偶然の確率以上に情動を認知することができなかった。顔全体の正答率は顔のパーツの正答率よりも有意に高かった。また、顔全体からは高確率で恐怖を悲しみ、嫌悪を怒りと混同された。目からは恐怖を悲しみ、悲しみを驚き、幸せを悲しみ、嫌悪を怒りと混同された。口からは恐怖を驚き、驚きを幸せと混同された。これらの結果から、顔全体からの情動の認識は他のパーツよりも正確であるが、口以外の顔のパーツから偶然以上に情動認識ができないことが示された。そのため、本研究の結果は、顔全体からの情動の認識の方が顔のパーツからの情動の認識よりも正確であるという仮説と一致した。しかし、目と鼻と耳からは偶然の確率以上に情動を認識できないという結果から、顔全体が見えていなくても、顔のパーツ、特に目と口から偶然の確率以上に情動を認識できる、目と口の情動の認識が鼻・耳よりも正確であるという2つの仮説とは一致しなかった。

このような結果となった背景には、提示された刺激がタイル状に区分けして切り取られた顔の一部分であったか、パーツの輪郭に沿って切り取られたものであったかという違いによるものだと考えられる。情動の認識において目と口が非常に重要であるという結果を示した研究では、顔をタイル状に区切った画像や、黄色や緑のマーカーで顔の一部分が覆われた状態の画像を刺激としていた (Wegrzyn et al., 2017)。それに対し本研究では、Blythe et al. (2023) の実験を基に、顔の目や鼻といったパーツをそれぞれ切り取ったものを刺激とした。そのため提示する面積とそこに含まれる情報が異なり、顔のパーツからは口だけが情動の認識を偶然の確率以上にできるという結果になったと考えられる。

引用文献

- Blythe, E., Garrido, L., & Longo, M. R. (2023). Emotion is perceived accurately from isolated body parts, especially hands. *Cognition*, 230, Article 105260.
- Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P. J., & Hilbig, B. E. (2022). lab.js: A free, open, online study builder. *Behavior Research Methods*, 54(2), 556–573.
- Thoma, P., Soria Bauser, D., & Suchan, B. (2013). BESST (Bochum emotional stimulus set): A pilot validation study of a stimulus set containing emotional bodies and faces from frontal and averted views. *Psychiatry Research*, 209(1), 98–109.
- Wegrzyn, M., Vogt, M., Kireclioglu, B., Schneider, J., & Kissler, J. (2017). Mapping the emotional face. How individual face parts contribute to successful emotion recognition. *PloS One*, 12(5), Article e0177239. (Blythe et al., 2023 の引用による)