

再テスト法を用いた感情リアルタイム評定法の 信頼性の検討

櫻井優太^{*1}・清水 遵^{*2}

問題と目的

感情リアルタイム評定法の開発

感情や情動に関する主観的体験を測定する方法として様々な質問紙による自己評定尺度が開発されている。例えば、POMS (McNair, Lorr, & Droppleman, 1971; 横山・荒記・川上・竹下, 1990), 多面的感情状態尺度 (寺崎・岸本・古賀, 1992), 一般感情尺度 (小川・門地・菊谷・鈴木, 2000) などの多項目式の心理尺度や, Affect-grid (Russell, Weiss, & Mendelsohn, 1989) などの単一項目評定法が開発され, 種々の研究に用いられてきた。

これらの質問紙は, 感情や情動の生起と生理学的反応の関連性について検討する精神生理学的研究にも用いられている (レビューとして, Bradley, 2000; Cacioppo, Berntson, Larsen, Poehlmann, & Ito, 2000など)。しかし, 質問紙による自己評定法は感情喚起操作をおこなった後で実施される回顧的な方法であり, 感情喚起操作から評定まで一定の時間が経過することから, 喚起された感情を正確に反映していない可能性が考えられる。また, 質問紙による測定は時間的分解能が低く, 時間軸にそった変化を詳細にとらえることができない。そのため, 質問紙による評定値と生理指標や行動指標などの連続して測定される指標を対応させた分析をおこなうことは困難である。

そこで, 櫻井・清水は感情喚起操作を行いながら, 同時にその時の感情体験をリアルタイムかつ連続的に評定する方法を開発してきた (櫻井・清水, 2005; 櫻井・清水, 2007; 櫻井・清水, 2008)。櫻井・清水の方法では, 感情評定の入力装置としてジョイスティック装置を用い, スティックを左右に操作することで, その時点の感情状態を連続的に評定するというものであった。

櫻井らはこれまでに, この評定法の妥当性を検討してきた。櫻井・清水 (2008) では, 先行研究によって感情喚起刺激として妥当性が確認されていた「快感情喚起映像」と「不快感情喚起映像」を交互に呈示し, これらの映像によって喚起された感情がリアルタイム評定値に反映されるか検討した (実験1)。また, リアルタイム評定をおこなう条件と, それを行わない条件を比較し, リアルタイム評定をおこなうことが感情喚起に対してどのような影響を与えるのか検討した (実験2)。それらの結果, 本評定法は感情の変化を継続的に, 鋭敏に測定することができ, リアルタイム評定を課すことは喚起された感情に対して重大な影響は与えないことがわかった。

※1 コミュニケーション研究科博士後期課程 在籍

※2 コミュニケーション心理学科

感情リアルタイム評定法の信頼性

心理尺度の開発には、一般に信頼性の確認も必要となる。信頼性とは測定値の安定性や一貫性を示すもので、測定値に占める誤差の少なさを表す(村上, 2006)。心理尺度の開発において、その信頼性はクロンバックの α 係数を用いた内的一貫性の検討や、折半法による信頼性係数の検討、再テスト法による信頼性係数の検討から確認される。

リアルタイム評定法は一種の「単一項目評定尺度」であるため、クロンバックの α 係数や折半法による信頼性係数は算出できない。そこで、同一の参加者に、同一の感情喚起刺激を、同一の指示や実験条件で呈示することで、再テスト法による信頼性の検討を行う。このような条件では、参加者に複数回ほぼ同じ感情が喚起されることが予想されるため、本評定法が参加者(評定者)の感情状態を精度良く測定できるならば、複数回測定したデータは似通ったものになると予測できる。

方法

参加者

大学生10名が実験に参加した^{※3}。男性2名、女性8名であり、平均年齢19.8歳(SD=1.08)であった。

刺激

感情喚起刺激としてIAPSスライドセット(Lang, Bradley, & Cuthbert, 2005)を用いた。スライドはLangらの評定値(快適度)を用いて、IAPSのスライドセットから、それぞれのスライドの評定値が平均値+1SD以上のもの(快スライド)、平均値 \pm 1SD以内のもの(中性スライド)、平均値-1SD以下のもの(不快スライド)が抽出された。抽出には評定値の分散が小さいということも条件とした。「快」・「中性」・「不快」の3種のカテゴリーごとにそれぞれ20枚のスライドが抽出され、合計60枚のスライドが実験に用いられた。それぞれのスライドは1枚あたり5秒間呈示され、快・不快・中性のカテゴリーがそれぞれ5枚ごとに切り替わる合計5分間の一連の刺激系列が作成された。スライドのカテゴリーはランダムな順序に設定され、「中性・不快・快・中性・快・不快・不快・中性・快・中性・不快・快」であった。パワーポイントで制御された刺激スライドは、参加者の前方60cmの距離に設置された17型液晶モニタの画面全体に呈示された。

感情リアルタイム評定法^{※4}

感情評定の入力装置として、ゲーム用ジョイスティック装置(Microsoft社製サイドワインダープレジジョン2)を使用した。スティックの根元にプラスチック製の板を貼り付け、スティックが左右のみに動くように加工した。このスティックを右に傾けると「快」を、左に傾けると「不快」を表すものとし、スティックの左右にそれぞれ「快」・「不快」の凡例を表記した。傾ける角度を大きくするほど、当該の感情を強く感じていることを示すものとした。左右どちらにも傾けずにスティックをまっすぐ立てた状態は「中立」を示すものとした。

※3 本研究はリアルタイム評定法の妥当性に関する実験と平行しておこなわれており、実験には他の実験条件としてさらに10名が参加したが、それらのデータはここでは報告しない。

※4 実験ではリアルタイム評定法の他に覚醒度を評定する質問紙や生理指標が測定されたが、それらは本論文の主題とするものではないため、ここでは割愛する。

このジョイスティックをUSBでPCに接続し、プログラムによってジョイスティックとの通信をおこなった。このプログラムはサンプリング間隔0.1秒でスティックの角度を読みとり、ハードディスクに値を保存するように設定された。測定値はスティックを最も右側に倒した状態で65 535 (0xFFFF)、スティックを中立にした状態で32 767 (0x7FFF)、スティックを最も左側に倒した状態で0 (0x0000) の値をとるが、これらは「最も快」が100に、「中立」が0に、「最も不快」が-100になるように換算された。

手続き

各参加者は1週間の間隔をおいて2回実験に参加した。参加者は実験室の椅子に着席し、ジョイスティックの使用方法についての説明を受けた。続いて、ジョイスティック操作の練習用画面が呈示された。この画面はスティックの可動範囲を示す枠と、現在のスティックの角度を示す線分で構成されており、スティックを動かすとそれに対応して線分が左右に動くように設定された。参加者は現在のスティックの角度がどの程度「快」あるいは「不快」を示しているのかこの画面を見て確認しながら、スティック操作の練習をおこなった。十分に練習がなされた後に、スライドの呈示中は「連続して、リアルタイムに、現在の気持ちをスティックの角度で示し続けること」という教示を行った。

その後、参加者は5分間の安静状態におかれた。続いて一連のスライドが呈示され、参加者はジョイスティックを用いて自身の感情をリアルタイムに評定した。2回目の実験では、1回目と同一の刺激を同一の教示で呈示し、同様の測定を行った。

結果

測定に不備があった1名のデータを分析から除外し、9名のデータで分析を行った。

各参加者のリアルタイム評定のデータについて、実験回数とスライドカテゴリーごとに評定値の平均を算出した（図1）。このデータを用いて実験回数（1回目、2回目）×スライドカテゴリー（快、中性、不快）の2要因個人内計画の分散分析を行った。その結果、スライドカテゴリー要因の有意な主効果が認められ（ $F(2, 16) = 43.26, p < .001$ ）、引き続きTukeyのHSD法による多重比較をおこなったところ、全てのカテゴリーの組み合わせに有意な差が認められた（ $MSe =$

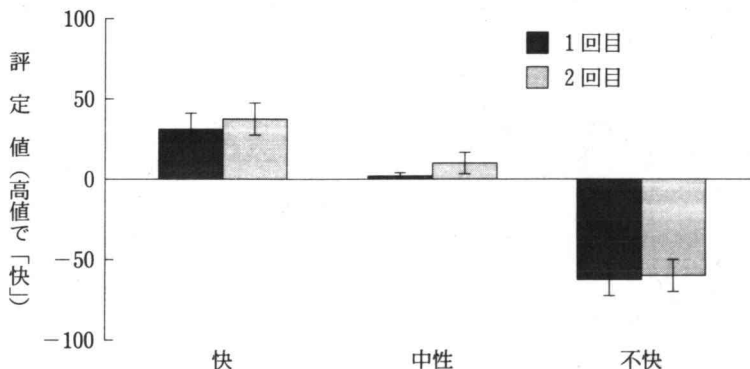


図1 スライド種別ごとのリアルタイム評定値の平均
（エラーバーは平均値の標準誤差）

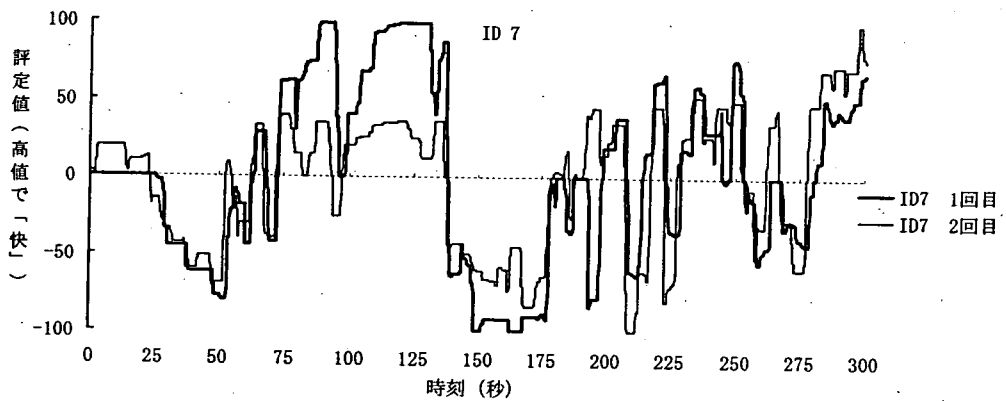
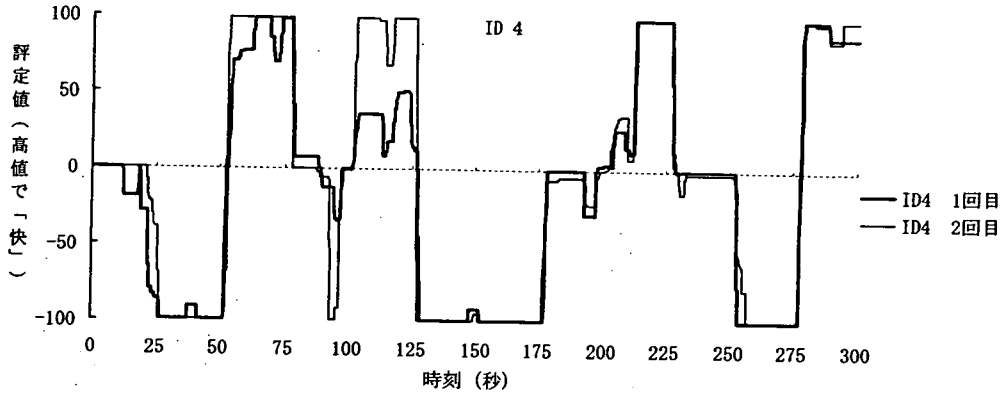


図2 参加者ごとのリアルタイム評価値の変動（代表例）

1041.16, $\alpha = .05$)。すなわち、「快」スライド、「中性」スライド、「不快スライド」の順で、「快適である」と評価されていた。実験回数の主効果 ($F(1, 8) = 2.44, ns$) および交互作用 ($F(2, 16) = 0.25, ns$) は有意ではなかった。

リアルタイム評価値のデータについて、参加者ごとに1回目に測定されたものと2回目のものを対にしてプロットした代表例を図2に示した。さらに、9名のデータを平均し同様にプロットした

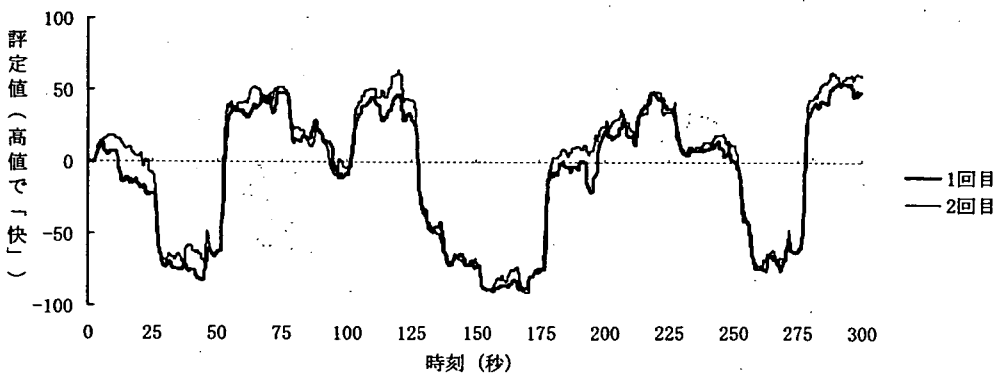


図3 リアルタイム評価値の平均

（図3）。各参加者の評定値の変動はスライドの切り替わりを反映したものになっており、また各回とも、それぞれの時刻で極めて類似した評定値が得られていることがわかった。

参加者ごとに1回目・2回目の間のリアルタイム評定値の相関係数を算出し、さらにその平均相関係数を算出した（表1）。リアルタイム評定のサンプル数（3001ポイント）を n として無相関の検定をおこなうと、全ての相関係数が0.1%水準で有意であった。平均相関係数は各相関係数をFisherの方法で z 変換し、その平均値を算出し、これを相関係数 r に変換する方法で計算した。この値も.914と非常に高い値を示し、各回の評定値が非常に類似した値を示していたことが確認された。

表1 参加者ごとの1回目の評定値と2回目の評定値の相関係数

参加者 ID	r
1	.714
2	.957
3	.929
4	.956
5	.869
6	.967
7	.731
8	.946
9	.906
平均相関係数	.914

考 察

スライドのカテゴリによるリアルタイム評定値の変動と複数回呈示による差異

リアルタイム評定の測定値がスライドの感情価を反映しているのかどうか、また複数回の呈示で評定値がどのように変動するかを検討するために、実験回数×スライドカテゴリの2要因個人内計画の分散分析を行った。その結果、スライドカテゴリの主効果のみが有意であり、リアルタイム評定の測定値はLang et al. (2005) の評定値と一致していた。すなわち、リアルタイム評定による測定値はスライドの感情価を反映したものになっており、リアルタイム評定法の妥当性を示唆する実験結果であった。

実験回数の主効果が有意ではなかったため、各参加者には各回でほぼ同等の感情喚起が行われていたと考えられる。従って、今回のデータを使用してリアルタイム評定法の信頼性について検討することができると思われる。

スライド呈示中のリアルタイム評定値の変動

参加者ごとに、各実験回数別にリアルタイム評定の測定データを時系列にプロットしたところ、1回目の測定値と2回目の測定値は極めて類似していることが分かった。参加者によっては個々のスライドを細かく評定する場合（代表例ID7）や、「快」や「不快」といったカテゴリごとに大まかに評価する場合（代表例ID4）も観察され、個人差が明確に現れていた。これらの個人差は、従来の質問紙のように時間的に細かい測定が行えない方法では得られない新しいデータである。

リアルタイム評定法の信頼性

2回測定されたリアルタイム評定のデータについて、参加者ごとに相関係数を算出し、2つの測定データの一致度を検討した。その結果、それぞれの参加者の相関係数は $r = .714 \sim .967$ の値をとり、各回のリアルタイム評定のデータは非常に一致する傾向が認められた。

本評定法のように感情の「状態」を測定することを目的とした尺度を複数回実施した場合、測定間の相関はむしろ中程度より低いことが望ましいとされている。その尺度が「特性」のように恒常的なものを反映しているならば、時間をおいて測定した場合にその時々状態の違いを反映しないため、測定間の相関が高くなる。従って、相関が高すぎる場合は「状態」の測定法としての妥当性

に疑問があると考えられている。

しかし、このような解釈は複数回の測定が実施される間に参加者の「状態」が十分に变化しており、逆に「特性」が変化していないという仮定に基づいている。今回の実験では各参加者に対して2回、ほぼ同じ感情を喚起する操作が行われた。そのため、各回の「状態」は極めて一致していたと考えられる。ここから、感情リアルタイム評定法の誤差が少なく信頼性が高いとすれば、各回の測定データは一致したものとなると予想される。前述のとおり、測定間の相関係数は非常に高い値となっており、本評定法の誤差の少なさ、すなわち信頼性の高さが示唆された。

本評定法は、時間の経過による感情の変化を詳細に測定できる方法である。さらに妥当性や信頼性について検討を進めるとともに、種々の感情研究に応用されることが望まれる。

謝 辞

本研究は平成20年度コミュニケーション心理学科清水ゼミ4年生、春日井美香さん、野原由美さん、引本佳那さん、山寺瑛美さんの協力で行われた。この場を借りて深く御礼申し上げます。

引用文献

- Bradley, M. M. (2000). Emotion and motivation. In Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (Eds.), *Handbook of psychophysiology*. Cambridge university press. Pp. 602-642.
- Cacioppo, J. T., Berntson, G. G., Larsen, J. T., Poehlmann, K. M., & Ito, T. A. (2000). The psychophysiology of emotion. In Lewis, M. & Haviland-Jones, J. M. (Eds.), *Handbook of emotions*. (2nd ed.) Guilford press. Pp. 173-191.
- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2005). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-6*. University of Florida, Gainesville, FL.
- McNair, D. M., Lorr, M., & Droppleman, L. F. (1971). *Profile of Mood States*. San Diego: Education and Industrial Testing Service.
- 村上宣寛 (2006). 心理尺度のつくり方 北大路書房
- 小川時洋・門地里絵・菊谷麻美・鈴木直人 (2000). 一般感情尺度の作成 心理学研究, 71, 241-246.
- Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect Grid: A Single-Item Scale of Pleasure and Arousal. *Journal of Personality and Social Psychology*, 57, 493-502.
- 櫻井優太・清水遵 (2005). ジョイスティックを用いた感情のリアルタイム評定法 生理心理学と精神生理学, 23, 142.
- 櫻井優太・清水遵 (2007). 感情のリアルタイム評定法の妥当性の検討 生理心理学と精神生理学, 25, 112.
- 櫻井優太・清水遵 (2008). ジョイスティックを用いた感情のリアルタイム評定法の作成と妥当性の検討 感情心理学研究, 16, 87-96.
- 寺崎正治・岸本陽一・古賀愛人 (1992). 多面的感情状態尺度の作成 心理学研究, 62, 350-356.
- 横山和仁・荒記俊一・川上憲人・竹下達也 (1990). POMS (感情プロフィール検査) 日本語版の作成と信頼性および妥当性の検討 日本公衆衛生雑誌, 37, 913-918.