

映像評定における時系列モデリング

野村 大 悠

1 はじめに

映像鑑賞から多種の感情が強く生起することは経験的に明らかである他に、Gross & Levenson (1995) や Sato, Noguchi & Yoshikawa (2007) の報告からも実証されている。更に、その感情の生起の要因は認知レベルに応じて大きく3つに分けられると考えられる。最も認知レベルが高い1つ目は、映像作品内の登場人物や物語要因に共感することで得られる感情 (Plantinga, 2013)、2つ目は状況に合わせたBGM効果などによる視覚刺激と聴覚刺激の相互作用による感情 (岩宮, 1992 野村・荒生・長谷川, 2019)、3つ目はショットと呼ばれる画面の切り替わりやカメラワーク、効果音演出などによる映像の構成要素から得られる感情 (Arijon, 1980, Christopher, 2011, 野村ら, 2019) である。特に、3つ目の映像の構成要素から得られる感情の計測には、刹那的に現れるカメラワークなどに対する反応を観測する必要がある。従来の心理学実験的測定法のうちオーソドックスな形態である質問紙法は、質問項目を設ければ設けるだけ、刺激の印象・特徴を多次元的にかつ包括的に測定することが可能である。一方で、刺激の呈示後の回顧的な方法である側面から、短時間に複数回現れる映像の構成要素単体の効果を測定するには不向きである。そこで、映像の鑑賞者の反応を時系列的に収集する方法として、櫻井・清水 (2008) は映像鑑賞中にジョイスティックを操作させる連続評定法を提案した。この方法は、刺激の鑑賞中にジョイスティックと呼ばれるヘリコプターの操縦桿のようなデバイスを、前後左右に傾けることで鑑賞者の意思表示をする方法であり、傾きの方向と角度によって鑑賞者の持つ意思の種類と程度を測定できるものである。野村

ら（2019）は、同様の評定法によって映像に対する興味度を時系列的に測定させた。時系列的に測定された興味度は、異なる映像刺激によって差があることを示した。

本研究の対象はこのような映像鑑賞中に得られる短時間的な印象の観測であり、それには時系列的な印象の測定が求められる。そして、測定されるデータが時系列であると、分析法にも同様に時系列的な観点が必要である。例えば、群間の差が時系列的にどのように推移していくかを求めたい場合を考えてみる。従来の一般的な分析方法である t 検定をはじめとした統計的仮説検定では、時系列データの一部の区間における条件間の差が有意かを判断できる。この時、設定する区間が広ければ、時系列データが均されてしまうため、先述の利他的な効果の有無が解釈しにくくなる。逆に、設定する区間を狭くすれば、実施される t 検定の回数が増えるために検定の多重性の問題が懸念される。

そこで、今回はベイズ推定による統計モデリングを用いる。研究者自身が独自に構築できる統計モデリングならば、前述の群間の差自体をパラメータとして設定することで分布の直接的な推定が可能である。この点において、区間の広狭による弊害を懸念する必要はなくなる。本稿の目的は、野村ら（2019）が統計的仮説検定で行った分析のうち2つをベイズ推定による統計モデリングを用いて再分析することである。即ち、群間の差の時系列分析と、映像の構成要素の発生と興味度推移の対応関係の2つの分析のモデル化である。心理学において広く用いられていた仮説検定とベイズ推定との有用性の議論は清水（2018）や大久保（2016）などをはじめ、多くの著作物で論じられている。特に、仮説検定法における帰無仮説の棄却による背理法のような考えが、検定結果の解釈を直観的でなくさせると批判されることが多い（松浦，2016）。本稿では再分析による従来との比較に重きを置いており、それらの手法そのものの優劣を考察するものではない。当該の手法によって時系列特性を踏まえた結果を得ることで、従来の報告に加え、映像鑑賞中に得られる短時間的な印象に対する知見を拡張することを狙いとしている。

2 既存の報告の到達点・課題

野村ら（2019）は、既存の映像作品から抽出した短い映像刺激を異なる条件刺激に編集し、条件間でジョイスティックによる時系列的興味度が異なるかを検証する2つの実験を報告している。

2.1 第一実験：実験参加者

実験参加者は大学生17名（男性3名、女性14名）であった。視力矯正をした者も居たが、視覚と聴覚に関して問題を報告した参加者は居なかった。

2.2 第一実験：手続き

第一実験では、Kingsman（2014）とTransporter2（2005）の冒頭シーンとアクションシーンを70秒間ずつの計4シーン（以下、KAS刺激、KOP刺激、TAS刺激、TOP刺激）を抽出し、それぞれ音声のみ呈示する条件（A条件）、映像のみ呈示する条件（V条件）、音声と映像両方を呈示する条件（AV条件）の3条件を用いて合計12個の映像刺激を同一参加者にランダムな順で呈示した。刺激の鑑賞中にジョイスティックを前後に倒すことで刺激全体を通した時系列的な興味度を測定した。A条件の興味度とV条件の興味度の和(AVSUM)とAV条件の興味度とを比較する分散分析とt検定を行った。また、映像刺激中に発生する映像の構成要素の発生頻度が、時系列を通してAV条件の興味度にどのように影響を与えるかを分析する重回帰分析を行った。

2.3 第一実験：結果と考察

AVSUMとAV条件の興味度を1秒ごとに比較する分散分析とt検定を行い、一部の区間では有意な差があった。有意な差が出る区間は映像内容の盛り上がりと対応していることから、視覚刺激と聴覚刺激単体の効果の和では説明しきれない視聴覚刺激の創発的特性を説いた。実験に用いた4つのシーンでは、攻撃による人物のリアクションや、効果音演出やカメラワークといった様々な視聴覚刺激が現れている。次に、これらの映像の構成要素の発生頻度のコーディングを行った。コーディングを行った構成要素は、映像内容

そのものに関する構成要素、聴覚に関する構成要素、視覚に関する要素の3つに分類した。映像内容そのものに関する構成要素として、主人公が相手に攻撃している場面が映されていること (Attack)、登場人物の表情が読み取れるほど顔面が映されていること (Face)、カメラの切り替えが行われていること (Shot)、攻撃の結果として人物や瓦礫が宙を舞っている様子が映されていること (Float) の4つを定義した。次に、聴覚に関する構成要素として、Shimamura (2013) の定義する物語世界の効果音が発生していること (SE_real)、非物語世界の効果音が発生していること (SE_fiction)、うめき声や叫び声を除いた登場人物のセリフが発生していること (Speech) の3つを定義した。最後に、視覚に関する構成要素として縦横方向のカメラの転回が行われていること (PanTilt)、カメラのズームインが行われていること (Zoomin)、カメラのズームアウトが行われていること (Zoomout)、カメラが対象物を追いかけるためにカメラごと追跡が発生していること (Track) の4つを定義した。11の構成要素と更に時間経過との計12の変数を説明変数としたステップワイズ法による重回帰分析を行い、興味度に影響を与える構成要素を割り出した (図1)。この時、分析はシーン毎ではなく4シーン全てを一括で対象にしたほか、分析区間を10秒に設定して行った。Attack、Shot、Zoominの3つが発生するほど興味度が増加し、Trackが発生するほど興味度が減少した。聴覚に関する3つの構成要素はいずれも興味度に影響を与える結果にならなかった。

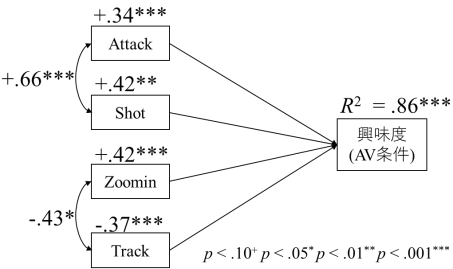


図1 野村ら (2019) のステップワイズ法による重回帰分析における映像の構成要素と興味度間の関係

2.4 第二実験：手続き

第二実験の実験参加者は大学生 20 名（男性 9 名、女性 11 名）であった。第一実験と同様に、視力矯正をした者も居たが、視覚と聴覚に関して問題を報告した参加者は居なかった。

2.5 第二実験：手続き

第二実験では、KAS 刺激と TOP 刺激を第一実験と同様に 70 秒間抽出し、それぞれ映像のみ呈示する条件（V 条件）、映像と BGM のみ呈示する条件（VM 条件）、映像と BGM と効果音を呈示する条件（VMS 条件）の 3 条件を用いて合計 6 個の映像刺激を同一参加者にランダムな順で呈示した。

2.6 第二実験：結果と考察

効果音が多量に発生している KAS 刺激では VMS 条件の興味度と VM 条件の興味度とに大きく差があったことに対し、効果音がほとんど発生していない TOP 刺激では VMS 条件の興味度と VM 条件の興味度とにほとんど差が発生していないというように、効果音の発生頻度の差異が条件間の差に同期していることを示した。

2.7 今回の統計モデリングによって得られる結果の差別点

野村ら（2019）の報告では、総じて、任意に指定した範囲の分析区間それぞれで条件間の差を明らかにしたこと、時系列データである興味度と非時系列データとの関係性を示したことが特徴であり、従来の仮説検定を活かした報告である。一方で、群間の興味度に有意差がある時点は刺激全体を通して少なく、有意差が発生していない区間についての考察が踏み込めないでいる。また、映像の構成要素の発生による興味度の対応関係についても、時系列的な観点による考察が必要だろう。これらの課題をベイズ統計モデリングで解決する。具体的には、群間の差の時系列的な推移を観測できること、時系列データに影響を及ぼす映像の構成要素の細かな影響を明らかにすることの 2 点を指針とした。

3 方法

本稿では、野村ら（2019）で報告された2つの分析をベイズ統計モデリングで再分析する。なお、下記のモデル式の中で頭文字が大文字になっている変数は野村ら（2019）で測定されたデータ、小文字になっている変数は推定するパラメータを指している。

3.1 モデルⅠ：視聴覚刺激が組み合わされることによる

創発得点を推定するモデル

1つ目の統計モデルの比較対象は、野村ら（2019）の70秒間の連続的な興味度を1秒間の区間毎にt検定を行うことで2条件の時系列データの差を比較した分析である。同様のアプローチを持つと思われるモデルを式（1）のように定義した。モデル式上の添え字*i*は参加者番号（1から17の整数）を指し、添え字*j*は時間（1から70の整数）を指している。

*i*番目の参加者の*j*秒目に測定されたAV条件の興味度 Av_{ij} とAVSUMの興味度 $Avsum_{ij}$ との間にある差を創発得点 add_j と定義するモデルである。創発得点 add_j のベイズ信頼区間に0が入っていなければ、 Av_{ij} と $Avsum_{ij}$ に十分な差があると判断できる。なお、 add_j と σ_j は無情報事前分布を用いて推定した。

$$Av_{ij} \sim \text{Normal}(Avsum_{ij} + add_j, \sigma_j) \quad \text{式 (1)}$$

3.2 モデルⅡ：映像の構成要素の興味度への影響力を推定するモデル

2つ目の統計モデルの比較対象は、野村ら（2019）の映像の構成要素の発生頻度が興味度に影響を与えているかを考慮する重回帰分析である。今回のベイズ統計モデリングでは式（2）のように、時間*j*ごとに発生する構成要素の発生秒数と時間経過要素*c*が興味度に影響を与えるかを推定するモデルを用いた。例えば、KAS刺激の8秒時点におけるFaceの発生秒数は0.73秒であるため、 $Face_8$ は0.73が代入される。従来の重回帰分析と同様のアプローチをしつつも、映像の構成要素が興味度に影響を与えるまでの時間的ラグを考慮した。なお、モデリングの際には発生秒数が0から1の実数に

収まるように標準化した。推定の結果、各映像構成要素の係数 k_i ($i = 1, 2, \dots, 11$) のベイズ信頼区間に 0 が含まれていなければ十分に影響を与える構成要素だと考えられる。なお、KAS 刺激における SE_fiction のように、発生頻度が 0 の構成要素や、最後の 1 秒間のみに発生する構成要素の場合は推定が出来ないため、そのような構成要素 i については式から除外した。また、 av_est_j ($j=1,2,3,4$)、 k_i 、 c 、 σ はいずれも無情報事前分布を用いて推定した。

$$av_est_j = k_1 \sum_{l=1}^4 Attack_{j-l} + k_2 \sum_{l=1}^4 Face_{j-l} + \dots + k_{11} \sum_{l=1}^4 Track_{j-l} + c * j$$

$$Av_j \sim Normal(av_est_j, \sigma) \quad \text{式 (2)}$$

4 結果と考察

4.1 モデル I：視聴覚刺激が組み合わされることによる創発得点を推定するモデル

式 (1) のモデルの推定にあたり、長さ 1500 のチェーンを 4 回発生させ、バーンイン期間を 150 とし、発生させた 5400 個の乱数を用いて事後分布に近似した。どのパラメータも \hat{R} が 1.05 を下回ったため、事後分布に収束したと判断した。

4 シーンの add_j の推移を図 2 に示した。 add_j の推移グラフのうち、黒の実線が中央値、濃い灰色の範囲が上下 50%、薄い灰色の範囲が上下 95% のベイズ信頼区間である。野村ら (2019) では同様の分析を、1 秒毎に独立した t 検定を実施しており、その t 検定で有意差が出た時点での t 値をピンク色の棒グラフで示した。全体的に、t 検定で有意差が発生した箇所と、推定値 add_j のベイズ信頼区間に 0 が入っていない箇所とは概ね一致した。

本モデルは、視聴覚刺激の組み合わせにおける創発得点を推定するモデルであり、野村ら (2019) による t 検定の結果とほぼ同様の結果となった。これはモデル I の発想は t 検定と同様に、条件間の差異そのものに注目する

アプローチであったからだと考えられる。推定値が0を含んでいない点は限られているが、その推移の様子が70秒間通して観察できるのは従来のt検定では到達できなかった点である。例えば、KOP刺激において45秒から60秒あたりでt検定による有意差と、モデリングによる推定値が正の実数に収束している点は同様である。しかし推定値は21秒付近から上昇して50秒あたりでピークを迎えている。21秒付近におけるKOP刺激では、それまでラジカセが再生されている様子が映されていたがカメラが引き、ヘリコプターが飛来して戦闘が始まるという、物語の始まりを想起させるシーンが描かれている。視聴覚刺激が組み合わさることによって生まれる創発得点 add_j は、A条件（あるいはV条件）とAV条件との間の映像内の情報量の差が開く時に高まると考えられる。再生されるラジカセという静的な印象から戦闘が始まる動的な印象への落差を与える映像内容によって創発得点が高まるとすると、この推移変動には映像内の描写のギャップが裏付けされている。同様の事例はTAS刺激における40秒付近とTOP刺激における30秒付近にも言える。

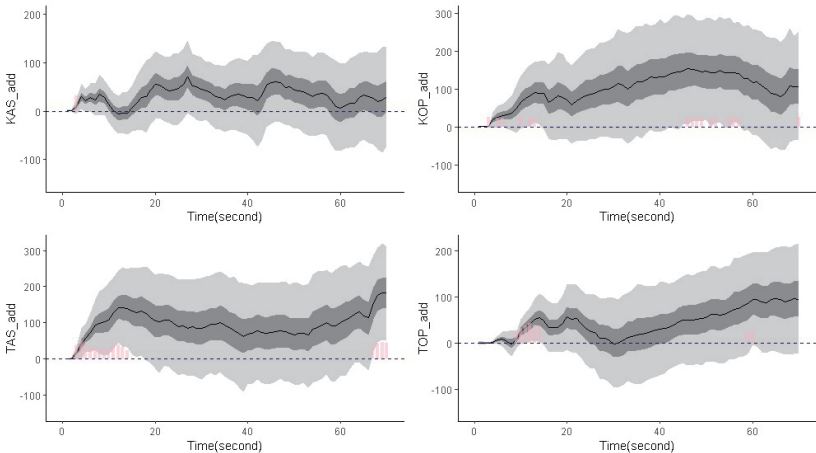


図2 モデル式(1)によるシーン別の add_j の推移

4.2 モデルⅡ：映像の構成要素の興味度への影響力を推定するモデル

式(2)のモデルの推定にあたり、長さ1500のチェーンを4回発生させ、バーンイン期間を150とし、発生させた5400個の乱数を用いて事後分布に近似した。どのパラメータも \hat{R} が1.05を下回ったため、事後分布に収束したと判断した。尚、最終的に考察対象にしたモデル式は4シーン毎に異なっている。それらの推定には同様の条件下で行われ、推定結果とモデル式は図3と図4に示した。それらに至る過程は下記の通りである。

式(2)によるモデルの推定の結果、係数 k_i のベイズ信頼区間が0を含まず正の実数に収束された構成要素は、KAS刺激においてはFace、SE_realの2つ、KOP刺激においてはSE_real、SE_fiction、PanTilt、Zoomin、Zoomoutの5つ、TAS刺激においてはShot、Float、SE_real、PanTilt、Trackの5つ、TOP刺激においてはFaceであった。一方で、係数 k_i のベイズ信頼区間が0を含まず負の実数に収束された構成要素はKOP刺激におけるSpeechとTOP刺激におけるAttackとTrackの計3つである。

続けて、各シーンの興味度の0を含まず正負の実数に収束した映像の構成要素と $c * j$ の和を計算する回帰式を定義し、各パラメータを推定した。推定にあたるチェーンの長さや事後分布の判断は式(2)と同様である。各シーンで推定された係数 k_i の推定結果を図3の通りであり、図中の線分は係数 k_i の95%のベイズ信頼区間を示したものである。新しく設定したモデル式と、推定された av_est_j の推移及び実際に測定されたAV条件の興味度の平均値を図4に示した。黒帯で描画されている線が av_est_j であり、緑の実線で描画されている線がAV条件の興味度の平均値の推移である。

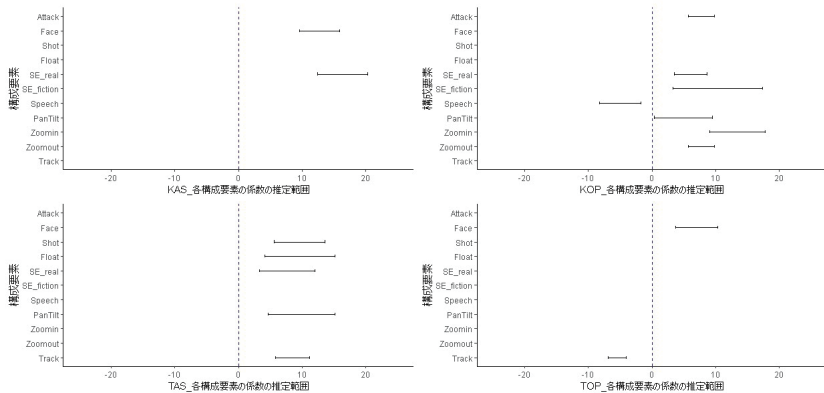


図3 モデル式(2)による係数 k_i の推定結果

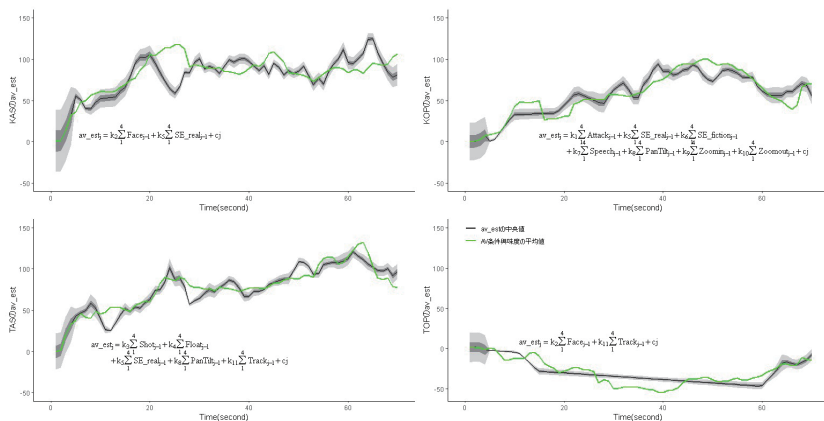


図4 av_est と実際の興味度の比較

野村ら(2019)では検出されなかった聴覚に関する構成要素が、KAS 刺激と KOP 刺激と TAS 刺激にて現れた。野村ら(2019)の第二実験では、KAS 刺激と TOP 刺激における効果音(SE_real と SE_fiction とともに含む)の効果を検証し、KAS 刺激では効果音の発生による興味度の上昇が見込まれたと報告された。今回の推定結果でも、KAS 刺激において SE_real の係数 k_5 が正の実数に収束したことは野村ら(2019)と同様である。また、野村ら

(2019) の第二実験では検討していない KOP 刺激と TAS 刺激についても、聴覚に関する構成要素の効果があることが示唆された。KOP 刺激でのモデルでは、SE_real と SE_fiction の係数 k_5 と k_6 がそれぞれ正の実数に収束した。累計発生秒数はそれぞれ SE_real が約 27.53 秒、SE_fiction が 3.95 秒である。なお、KOP 刺激の構成要素の発生秒数の平均値は 11.04 秒であり、最多の構成要素は PanTilt の約 40.86 秒である。TAS 刺激でのモデルでは、SE_real の係数 k_5 が正の実数に収束し、SE_real の累計発生秒数は約 18.33 秒であった。TAS 刺激の構成要素の発生秒数の平均値は 9.43 秒であり、最多の構成要素は Face の約 30.13 秒である。KOP 刺激も TAS 刺激も全体的に右肩上がりのグラフであるため、発生頻度の大小がそのまま係数の大きさに繋がるかということとそれに限らない。KOP 刺激における最多発生秒数を持つ PanTilt は係数が正の実数に収束しているが、TAS 刺激の Face の係数 k_2 は 0 を含んだ実数に収束している。この傾向は、各構成要素の発生頻度に対応して、興味度が増減したことへの証左である。推定された av_est_j と測定された興味度の平均値の推移も全体的に類似しており、モデルの妥当性を支持する結果となった。

5 まとめ

本稿では、野村ら (2019) の実験結果をベイズ推定による統計モデリングによって再分析を行い、その結果以下のような知見をもたらした。

モデル I において、群間の差の時系列的推移の可視化によって、グラフの傾きが負から正に反転する時点即ち差の生起時点を解釈出来るようになった。恐らくこの刺激や測定法に限らず、時系列データの群間の差の生起時点とは、従来の t 検定では有意差を検出しにくいと考えられる。差の発生時点とは映像内の変化が激しくなるタイミングだと考えられ、そこには物語的要素はもちろん、視聴覚的な変化も多分にあるだろう。このように、差の推移を観測出来ることは、野村ら (2019) の主張する視聴覚刺激の組み合わせによる創発性の説得力を増す支持材料となった。

一方で、モデルⅡでは、野村ら（2019）のステップワイズ法による重回帰分析の結果とは大きく異なった。これはモデルⅡでは全ての独立変数をモデルに組み込んでいるのに対し、最も適した独立変数の組み合わせを探索することを目的としたステップワイズ法とはアプローチが大きく異なっているためだと考えられる。加えて、参照データがシーン毎且つ1秒ごとといったように細断されていたこと、構成要素の影響に時間的ラグがあるとしたこと、といったように参照データの面でも回帰式の構成面でも違いがあった。結果として野村ら（2019）のステップワイズによる重回帰分析では検出できなかった聴覚に関する構成要素の影響力を示したこと、野村ら（2019）の第二実験では考慮出来なかった TAS 刺激においても聴覚に関する構成要素の影響力を示したことは今回のベイズ統計モデリングによって初めて明らかになった点である。しかし、野村ら（2019）と同様に、各刺激の映像の構成要素は複雑に共起していることから、一概に普遍的な因果関係があるとは言えないことは留意するべきである。

今回は、定義したモデルのサンプルデータへの適合度、引いてはモデルの妥当性には言及しなかった。本稿の目的が、従来の仮説検定とベイズ推定を用いた統計モデリング法との結果の比較である以上、モデルの妥当性を高めていくことは従来法との比較をするうえで解釈を妨げると考えたためである。また、モデルの妥当性を客観的に評価するために、AIC や WAIC、あるいは BIC や WBIC などの情報量規準が用いられる（松浦, 2016）。AIC と WAIC は、モデルの予測の良さを表し、BIC と WBIC は真のモデルへの近さを表している。今回のような階層モデル構造を定義する場合は、WAIC と WBIC を参照するのが良いとされる（渡辺, 2012）。今後、統計モデリングによる分析を主題とする際にはより良いモデル選択のためにこれらの情報量規準を参照することもあるだろう。しかし、WAIC や WBIC が本当に適用出来るかどうかという議論も活発であり、これらをモデル選択に用いる際には注意が必要である。

モデルⅠとモデルⅡは共に、野村ら（2019）とは同様のあるいは異なるアプローチに野村ら（2019）の課題を解決するための知見を拡張するものである。繰り返しになるが、ベイズ統計モデリングが従来の仮説検定よりも

優れた方法と断ずる訳ではない。時系列データは測定間隔や階層構造などの要因で複雑性が高く、分析する都合上繰り返しの検定の問題が生じる。そのため、モデル I のような時系列データの群間比較をするといった時系列データの特性が強く影響するような場合は、今回のようなベイズ統計モデリングが有力な選択肢の 1 つになるだろう。一方で、実験研究における仮説検定は論理的な説得力を強めるものであり（清水，2018）、印象評定の因子分析のような時系列データを伴わない場合は従来の検定法を採用して良いだろう。今後も時系列データを伴う実験研究において、今回導入したベイズ推定モデルを適宜取り入れることが望まれる。

引用文献

- Gross, J. J., & Levenson, R. W. (1995) . Emotion elicitation using films. *Cognition & emotion*, 9 (1) , 87-108.
- Sato, W., Noguchi, M., & Yoshikawa, S. (2007) . Emotion elicitation effect of films in a Japanese sample. *Social Behavior and Personality: an international journal*, 35 (7) , 863-874.
- Plantinga, C. (2013) . The affective power of movies. *Psychocinematics: Exploring cognition at the movies* (pp. 94-111) . Oxford University Press.
- 岩宮眞一郎 . (1992) . オーディオ・ヴィジュアル・メディアを通しての情報伝達における視覚と聴覚の相互作用に及ぼす音と映像の調和の影響 . 日本音響学会誌 , 48, 649-657.
- 野村大悠, 荒生弘史, & 長谷川智子 . (2019) . 映像に対する興味度の連続的評定法の提案と視聴覚相互作用の時系列特性の解析 . 日本感性工学会論文誌 , 18, 235-245.
- Arijon, D., 岩本憲児, & 出口丈人訳 . (1980) . 映画の文法—実作品にみる撮影と編集の技法, 紀伊國屋書店 .
- Christopher, K., 吉田俊太郎訳 . (2011) . マスターショット 100—低予算映画を大作に変える撮影術, フィルムアート社 .
- 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

- △評定法の作成と妥当性の検討 . 感情心理学研究 , 16, 87-96.
- 清水裕士 . (2018) . 心理学におけるベイズ統計モデリング . 心理学評論 , 61, 22-41.
- 大久保街亜 . (2016) . 帰無仮説検定と再現可能性 . 心理学評論 , 59, 57-67.
- 松浦健太郎 . (2017) . Stan と R でベイズ統計モデリング : Wonderful R 2 (Vol. 2) . 共立出版 .
- Shimamura, A. P. (2013) . Psychocinematics: Issues and directions. *Psychocinematics: Exploring cognition at the movies* (pp. 1-26) . Oxford University Press.
- Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957) . *The measurement of meaning*. University of Illinois press.
- 渡辺澄夫 . (2012) . ベイズ統計の理論と方法 . コロナ社 .