原著論文

視聴覚コンテンツの臨場感評価のための 視聴者の時系列感情分析および感情推定モデルの構築

木下 雄一郎, 伊藤 将亮, 小澤 賢司, 森勢 将雅

山梨大学

Time-series Analysis and Model Development of Audience Emotion for Evaluation of the Sense of Presence in Audio-visual Content

Yuichiro KINOSHITA, Masaaki ITO, Kenji OZAWA and Masanori MORISE

University of Yamanashi, 4-3-11 Takeda, Kofu, Yamanashi 400-8511, Japan

Abstract : The term *presence* is widely used to express the performances of audio-visual content and systems. Many advanced audio-visual systems for the perception of a higher degree of presence have been proposed. Meanwhile, methodologies for evaluating the sense of presence derived from these systems have also been discussed for the further evolution of content and systems. Although the emotion of audiences can be considered as one of the key factors in evaluating the sense of presence, the relationship between the time-series changes of the emotion and audio-visual content has not been examined so far. The authors firstly conduct subjective evaluation experiments to investigate this relationship. Based on the evaluation results and extracted audio-visual features, emotion estimation models are constructed to quantitatively estimate the time-series emotion of audiences, evoked from audio-visual content. The performance tests of the models demonstrated sufficient accuracy in estimating the emotion.

Keywords : *Sense of presence, Emotion, Audio-visual content, Emotion estimation model, Neural network*

1. は じ め に

昨今、映画やスポーツ番組などのコンテンツやAV (Audio-Visual) システムの性能を評価する上で「臨場感」という用語が重要視されている。また、近年のメディア技術の進歩に伴い、遠く離れている人間に臨場感を伝えることによる、相互理解や感動の共有に向けた期待が高まっている。これまでに、高い臨場感をもたらすコミュニケーションシステムに関する研究・開発が、映像や音響に関する分野を中心として行われてきた[1-4]。一方で、臨場感自体の評価に関する技術も、この分野の研究開発の進展に重要な役割を担っている。これまでに、質問紙を用いた臨場感評価手法に関する研究[5, 6]が行われている。また、評定尺度法や一対比較法を用い、臨場感について「臨場感があるか」という直接的な聞き方をした研究[7, 8]や、臨場感の多次元的印象構造を明らかにした研究[9, 10]もみられる。

また、筆者らのグループでは視聴覚コンテンツの臨場感に関わる映像および音の特徴量を検討した上で、コンテンツがもたらす臨場感を定量的に推定するモデルを構築してきた[11]。臨場感の辞書的意味は「あたかもその場に臨んでいるような感じ」であり[12]、この研究においても臨場感の定義として「その場にいる感じ」を用いている。一方、臨場感とコンテンツ視聴者の感情には一定の関わりがあることが、これまでの研究で明らかになっており[13]、筆者らは視聴者の感情についても焦点を当ててきた。この研究では、感情をRussellの円環モデル[14]によって定義しており、快・不快軸と覚醒

軸によって構成される2次元で表現している。しかし、これまでのモデル構築[11]や視聴者の感情に関する調査[13]はコンテンツ全体を対象にしたものであり、コンテンツ内における感情の時系列変化については扱われていない。

そこで、本研究では視聴覚コンテンツに対する視聴者の感情の時系列変化を感情評価実験により明らかにする。また、得られた時系列評価結果にもとづき、感情と臨場感との関係についても考察を行う。さらに、視聴覚コンテンツにおいて時刻の推移とともに変化する映像や音の特徴量から視聴者の感情を時系列として推定するモデルをニューラルネットワーク(NN)により構築する。本研究における感情および臨場感の定義は先行研究[11, 13]と同様のものとする。

本研究の貢献は、日常風景を記録した視聴覚コンテンツを用いて、視聴者の感情に関わる映像や音の特徴量、視聴者が感じる臨場感と視聴者の感情との関係を明らかにした点にある。また、本研究で構築するモデルは映像やVRなどのコンテンツ設計における指針として用いることができ、視聴者の感情を考慮した高い臨場感をもたらすコミュニケーションシステムの開発に寄与するものであると考える。

2. 感情評価実験

2.1 視聴覚コンテンツ

視聴覚コンテンツと視聴者の感情との関係を調査するために、被験者に視聴覚コンテンツを提示し、それに対する感情の程度を評価させる感情評価実験を行った。先行研究[7]では、映像・音像における動きの有無が臨場感の評価に影響する

(a) 7: passing roller coaster A (front)

(b) 13: fireworks (large quantity)

(c) 20: scene of playing Japanese archery

(d) 32: view of lakeside

図1 収集した視聴覚コンテンツの一例

一つの要因であることが示唆されている。そこで、本研究でも、映像・音像が動くか動かないかを基準にして、評価実験に用いる40種類の視聴覚コンテンツを選定した。なお、映像が動くか動かないかの判断は、観測者に対する対象の相対的な位置の移動の有無により行った。例えば、滝の風景は映像が動かないコンテンツとなる。コンテンツの映像はビデオカメラ (Sony, PMW-EX1R) による、Full HD画質 (1980×1080 / 60i) で録画し、音は原音場の情報を可能な限り再現するために、ダミーヘッド (高研, Samrai) を用いたバイノーラル録音を行った。記録方式は非圧縮 (PCM, 16 bit, 48 kHz) である。収集した視聴覚コンテンツの例を静止画像として図1に示す。

2.2 実験方法

実験では、選定した40種類のコンテンツを65インチディスプレイ (Sharp, LC-65RX1W) とヘッドホン (Sennheiser, HD650) により、被験者にランダムな順序でそれぞれ20～40秒間提示した。被験者はソファに座ってコンテンツを視聴した。これは、実際の家庭環境での視聴を想定したものである。ディスプレイと被験者の距離は、ITU-R Rec. Bt.710-4 [15] で勧告されている3H (2.4 m) とした。被験者は正常な視力

および聴力を有する20代前半の大学生の男女20名である。

評価には、Russellの円環モデル [14] にもとづき選定した2対の感情語で表現される感情「Activation—Deactivation」, 「Pleasant—Unpleasant」を用いた。本実験の被験者は全員が日本人であるが、それぞれの感情語に一对一で対応する日本語を選定するのは困難である。また、日本語1語を提示した場合、Russellのモデルにて定義されている本来の意味との間に齟齬が生じる可能性がある。そのため感情評価実験では、感情語はあえて英語のまま提示することとし、評価実験に先立って、それぞれの感情語について「活性、活動的」「非活性、非活動的」「快、感じの良さ、愉快」「不快、感じの悪さ、不愉快」のように複数の日本語訳を提示し、あらかじめ各感情語を概念として理解するよう教示した。

被験者は、提示されたコンテンツを視聴しながら、カテゴリ—連続判断法 [16] によって、時系列で変化する感情の程度を評価した。尺度は7段階のSD (Semantic Differential) 尺度を用い、その時点で感じている感情の程度を、キーボード上の1～7のキーを押しつづけることで回答した。ここでは、1が対の左側の感情に、7が右側の感情に対応している。被験者の回答は標準化周波数10 Hzで記録した。なお、刺激の提示開始後に臨場感を評価できるようになるまでは、いずれのキーも押さずに待機するように教示した。

2.3 実験結果

40種類のコンテンツに対する全被験者の平均評価値および各コンテンツを100 ms区間ごとに区切って算出した平均評価値の標準偏差を図2に示す。この結果から、「Activation—Deactivation」の感情と「Pleasant—Unpleasant」の感情はそれぞれ独立した評価が行われていることが確認できる。また、「2: passing train」や「33: ringing mobile phone」のように、時刻の推移にともなって評価値に変化が見られるコンテンツと、「9: passing vehicles in a tunnel」や「27: trees in the breeze」のようにコンテンツ全体を通じて評価値が定常的であるコンテンツの両方が確認された。

3. 視聴者の感情に関わる特徴量の検討

3.1 Activation—Deactivation

図1で例示した各コンテンツに対する「Activation—Deactivation」の全被験者の平均評価値を図3に青実線で示す。図3 (a) に示すコンテンツ「7: passing roller coaster A (front)」の開始21秒後付近でのローラーコースタ接近時や、図3 (c) に示すコンテンツ「20: scene of playing Japanese archery」の開始15秒後付近での弓道の矢の通過時など、映像や音像の動きが大きい場合にActivationの感情に顕著な増加が見られた。一方で、図3 (b) や3 (d) のコンテンツ「13: fireworks (large quantity)」 「32: view of lakeside」のように映像・音ともに定常的である場合は、評価値も定常的であった。そこで、本章では視聴者の感情評価に関わる特徴量の検討を行う。

視聴覚コンテンツの臨場感評価のための視聴者の時系列感情分析および感情推定モデルの構築

図2 平均感情評価値とコンテンツ全体における標準偏差

臨場感に関する先行研究 [11] では、臨場感に関連する映像特徴量として、移動物体が占める画素数の平均値、移動物体が占める画素数の変動量、映像を構成する色彩など 14 特徴量が、音特徴量として、音の大きさの感覚に対する尺度であるラウドネス [17]、音像の動きの有無、音の大きさの変動感、音の粗さなどの 6 特徴量が挙げられている。一方、感情に関する先行研究 [13] では、上述の特徴量から、感情への影響が小さい特徴量として、映像を構成する色彩のうち色相値に関する特徴量、輝度に関する特徴量、そして、移動物体が占める画素数の平均値を抽出している。ただし、これらの知見は、コンテンツ全体を対象とした特徴量にもとづくものである。ここでは、コンテンツ内における特徴量の時系列変化とそれに対する感情の変化を考慮することで、より詳細な検討を行う。

まず、各被験者の感情評価値の記録を確認したところ、頻繁に評価値が変化する場合でも 500 ms 程度は同一の評価値が継続することが確認された。そこでデータ量を削減するため、全てのコンテンツを 500 ms 区間ごとに区切った。そして、それぞれの区間について先行研究 [11, 13] で用いた

(a) 7: passing roller coaster A (front)

(b) 13: fireworks (large quantity)

(c) 20: scene of playing Japanese archery

(d) 32: view of lakeside

図3 感情の時系列評価結果の一例

各特徴量の平均値と 2 章の評価実験により得られた感情評価値の平均を算出し、これらの相関係数を確認した。その結果、図 4 (a) に示すようにラウドネスとの間に強い正の相関 ($r=0.74$) が確認された。このほか、音特徴量についてはラウドネス、音像の動きとの間にも相関 (それぞれ、 $r=0.63$, $r=0.42$) が見られた。以上より、音が大きくなる、音の大きさに素早い変動が起こる、音像が動く、といった場合に Activation の感情評価が増加していることが確認された。

図4 Deactivation—Activationの感情と各特微量との関係

図5 Unpleasant—Pleasantの感情と各特微量との関係

一方、映像特微量については、図4(b)に示すように平均彩度〔0, 255〕の範囲で表記)との間に弱い正の相関($r=0.40$)が確認されたほか、彩度の歪度との間にも負の相関($r=-0.46$)が確認された。しかし、移動物体が占める画素数の平均値とActivationの感情評価値との間には相関は確認されなかった($r=0.07$)。これはコンテンツ「27: trees in the breeze」や前述のコンテンツ32のように大局的には映像に動きは見られないが、定常的に木々や水面が揺れており、画素単位で見た場合に移動物体が占める割合が大きくなるコンテンツが存在していたことに起因すると考えられる。

以上のように、「Activation—Deactivation」の感情では、先行研究[13]で得られた知見と同様の結果を得た。

3.2 Pleasant—Unpleasant

「Pleasant—Unpleasant」の感情における全被験者の平均評価値を図3に緑実線で示す。図3(a)のコンテンツ7におけるローラーコースタの接近時など、ラウドネスが大きくなる場合にUnpleasantの感情の増加が見られた。「Activation—Deactivation」の場合と同様に、500 ms区間ごとに各特微量の平均値と平均感情評価値の関連を確認したところ、図5(a)に示すようにラウドネスとの間に弱い負の相関($r=-0.29$)が確認された。また、ラフネス、音像の動きとの間にも弱い負の相関(それぞれ、 $r=-0.29$, $r=-0.29$)が見られた。ただし、寄与率は高くなく、これらの特微量で「Pleasant—Unpleasant」の感情を説明できるとは言い切れない。一方、映像特微量については、図5(b)に示すように平均輝度〔0, 255〕の範囲で表記)に弱い負の相関($r=-0.36$)が確認さ

れたほか、輝度の歪度にも負の相関($r=-0.46$)が確認され、コンテンツの輝度が低くなる場合にPleasantの感情評価が増加する傾向があることが確認された。「Activation—Deactivation」の場合と同様、移動物体が占める画素数との間には相関は確認されなかった($r=-0.01$)。以上のように、「Pleasant—Unpleasant」の感情は「Activation—Deactivation」の場合と比較して、高い相関を示す映像や音の特微量は存在しない。このため、より複合的な要素によって決定される感情であると推察できる。

4. 視聴者の感情と臨場感との関係

4.1 Activation—Deactivation

次に、2章の評価実験により得られた感情評価結果と先行研究[18]における臨場感評価結果を比較する。なお、この先行研究では、評価実験に先立ち、臨場感の定義として「その場にいる感じ」が被験者に教示されている。図1に示した各コンテンツに対する臨場感評価結果を図3に灰色の実線で示す。臨場感評価結果は1～7の7段階のリッカート尺度によって評価されたものであり、「臨場感を感じた」という項目に対し1が「全くそう思わない」、7が「非常にそう思う」の評価に対応している。図3(a)の「7: passing roller coaster A (front)」, 図3(b)の「13: fireworks (large quantity)」, 図3(c)の「20: scene of playing Japanese archery」の開始15秒以降など、臨場感が高く評価されているコンテンツでは、Activationの感情も高く評価されている。一方、図3(d)の「32: view of lakeside」のように臨場感が低く評価されているコンテンツは、Deactivationの感情で評価されていることが読み取れる。さらに、コンテンツ7, 20において臨場感評価値が増加するタイミングで、Activationの感情も顕著に増加していることが確認できる。

本研究で使用した40種類の視聴覚コンテンツ全体について、感情評価と臨場感評価との関係を検証するため、全てのコンテンツを500 ms区間ごとに区切り、それぞれの区間について、平均感情評価値と先行研究[18]における臨場感評価値の平均を算出した。図6(a)は、全コンテンツの計2321区間について「Activation—Deactivation」の感情に対する評価値と臨場感評価値との関係をプロットしたものである。相関係数は $r=0.84$ であり「Activation—Deactivation」の評価値と臨場感評価値の間に強い正の相関を確認した。

4.2 Pleasant—Unpleasant

「Pleasant—Unpleasant」の感情では、図3(a)のコンテンツ7のように臨場感が高く評価されているコンテンツは、Unpleasantの感情で評価され、図3(d)のコンテンツ32のように臨場感が低く評価されているコンテンツは、Pleasantの感情で評価されていることが読み取れる。しかしながら、図3(b)のコンテンツ13も臨場感が高く評価されているコンテンツであるが、「Pleasant—Unpleasant」の評価値は中庸である。さらに、図3(c)のコンテンツ20において臨場感評

視聴覚コンテンツの臨場感評価のための視聴者の時系列感情分析および感情推定モデルの構築

図6 感情評価値と臨場感評価値との関係

価値が増加するタイミングで、Pleasantの感情も増加しているのに対して、コンテンツ7では、臨場感評価値が増加するタイミングで、Unpleasantの感情が増加していることが確認できる。このように、「Pleasant—Unpleasant」においては、「Activation—Deactivation」の場合と異なり、臨場感評価値との間に一定の関係を見出すことはできなかった。

図6(b)は、全てのコンテンツを500 ms区間ごとに区切り、これら計2309区間について「Pleasant—Unpleasant」の感情の平均評価値と平均臨場感評価値との関係をプロットしたものである。相関係数は $r = -0.35$ であり、「Pleasant—Unpleasant」の評価値と臨場感評価値の間に弱い負の相関が観測された。

5. 感情の時系列推定モデルの構築

5.1 構築方法

感情評価実験の結果をもとに、視聴覚コンテンツにおいて時刻の推移とともに変化する映像や音の特徴量から視聴者の感情を時系列として推定するモデルを構築する。筆者らのグループによる先行研究[11, 13]では、視聴覚コンテンツ全体の臨場感や感情を推定するモデルを構築するために、そのコンテンツ全体にわたって計測した映像および音に関する特徴量を算出し、これらと臨場感評価実験や感情評価実験の結果との関係をNNにより学習させた。本研究では、感情

の時系列評価値を推定可能とするよう、このモデルを拡張する。3章では、「Activation—Deactivation」の感情について、ラウドネスやラフネス、色彩の彩度など視聴者の感情に関わる特徴量が抽出された。その一方、「Pleasant—Unpleasant」の感情については、より複合的な要素によって決定される感情であることが推察された。そこで、特定の特徴量を選択的に使用するのではなく、あえて、先行研究[11]で使用された全ての特徴量を使用することとした。これらの特徴量を以下(V1)～(V14)、(A1)～(A6)に示す。

・映像に関する特徴量

- (V1) 色相値[0, 35]および[324, 359]の平均画素数
- (V2) 色相値[36, 107]の平均画素数
- (V3) 色相値[108, 179]の平均画素数
- (V4) 色相値[180, 251]の平均画素数
- (V5) 色相値[252, 323]の平均画素数
- (V6) 平均輝度
- (V7) 輝度標準偏差
- (V8) 輝度歪度
- (V9) 平均彩度
- (V10) 彩度標準偏差
- (V11) 彩度歪度
- (V12) 移動物体が占める画素数の平均値
- (V13) 移動物体が占める画素数の95パーセンタイル値と5パーセンタイル値との差
- (V14) 500 msごとの上述の差における標準偏差

・音に関する特徴量

- (A1) 平均ラウドネス推定値[sone]
- (A2) 平均シャープネス推定値[acum]
- (A3) 平均ラフネス[asper]
- (A4) 95%時間率騒音レベルL95に対する5%時間率騒音レベルL5の相対レベル[dB]
- (A5) 500 msごとの上述の相対レベルの標準偏差[dB]
- (A6) 音像の動き(動きの有無を1, 0で表現)

本研究では時系列評価値の推定を行うため、これらの各特徴量を、コンテンツ全体ではなく500 ms区間毎に計測し、それぞれの区間に対して算出する方法に変更した。ただし、特徴量(V13)、(V14)と(A4)、(A5)は直近の3 s区間について計測した。これは、「知覚的現在」が約3 sであるとされていることに由来するものである。また(V7)、(V10)については、500 ms間隔で計測した後、それらの標準偏差を直近の3 s区間について計算した。特徴量(A6)は、左右の音圧レベル差が連続する500 ms区間で2 dB以上異なる場合には「移動あり」と判断することとした。これは、ラウドネスの弁別限が約1 dBであることに由来する。音像の左右移動によりラウドネスが変化した場合、仮に、左側で1 dBの上昇があるとき、右側では1 dBの低下があると近似的に考えることができる。これは左右の音圧レベル差が2 dB変化したことに相当する。

これらの各特徴量にNNの入力層1ユニットを対応づけ、入力層20ユニット、中間層15ユニット、出力層1ユニット

のNNによって、「Activation—Deactivation」, 「Pleasant—Unpleasant」それぞれの感情について時系列推定モデルを構築した。NNの学習には前述の通り500 ms間隔で計測した特徴量を用い、それぞれの時点における全被験者の平均感情評価値を教師信号とした。各入出力値は[0, 1]の範囲に正規化して扱った。学習アルゴリズムには誤差逆伝播法を用い、平均二乗誤差が 1.0×10^{-3} 以下となるまで学習を行った。

5.2 構築結果

構築した両モデルそれぞれの有効性について検証を行った。モデルの汎化性能を検証する手法としてleave-one-out cross-validationを用いた。実験に用いた40コンテンツを、学習用39コンテンツと検証用の1コンテンツに分割し、学習用コンテンツを用いて構築したモデルで検証用サンプルを評価した。そして、40通りの分割の全てに対して汎化誤差を計算することでモデルの精度を検証した。

図7に40種類それぞれのコンテンツにおける、平均誤差および誤差の標準偏差を示す。モデルの出力値は[0, 1]の範囲であるが、ここでは7段階尺度における誤差に変換して示した。図より、「Activation—Deactivation」の感情については3コンテンツを除いて平均誤差が7段階評価値で1.0（尺度1段階分に相当）以下であることがわかる。また、半数以上の25コンテンツにおいて平均誤差が0.5を下回った。なお、全40コンテンツの誤差の平均は7段階尺度において0.46であった。「Pleasant—Unpleasant」の感情では、1コンテンツを除いて平均誤差が1.0以下であった。また、29コンテンツにおいて平均誤差が0.5を下回った。全40コンテンツの誤差の平均は0.42であった。一部の例外はあったものの、全体として、構築したモデルの感情推定性能は良好であると考えられる。

6. 考 察

6.1 Activation—Deactivation

図1に示した各コンテンツのうちActivationの感情に特徴的な増加が見られたコンテンツ「7: passing roller coaster A (front)」 「20: scene of playing Japanese archery」について、感情評価実験における平均評価値、感情推定モデルの出力値を3点移動平均した値、平均評価値との誤差それぞれを図8(a), 8(b)に青実線、赤実線、緑色の点線で示す。いずれも、感情評価実験における評価値に増加が見られたタイミングで、モデルによる推定値も増加しており、良好な推定が行われていることが読み取れる。

次に、モデル汎化性能の検証の結果、平均誤差が大きかったコンテンツ「24: view of a university building」 「27: trees in breeze」 「32: view of lakeside」のうち、コンテンツ27について、そのモデル出力値を図8(c)に示す。図より、感情評価実験で得られた評価値と比較して、モデルによる推定値がActivation側に寄っていることが読み取れる。これは、コンテンツ24, 32においても同様であった。これらのコンテンツはいずれも、映像・音像のどちらも動かないコンテンツであり、コンテンツ27, 32については、ほとんど無音である。また、コンテンツ24についても、時折、付近を通り過ぎる車の音や鳥のさえずりが聞こえる程度でありいずれも大きな音ではない。映像にも動きはなく、常に風景を捉えているものである。ただし、コンテンツ24, 32では水面が、コンテンツ24では木々がわずかではあるが定常的に揺れている。本モデルでは、移動物体の量を画素数で表現したため、大局的には映像に動きは見られない場合でも、木々や水面の揺れのように動きがある画素数が多い場合は、映像の多くの範囲が動いているものとして捉えられてしまう。そのため、モデルによる推定値がActivation側に寄ってしまったものと考えられる。

図7 構築した感情推定モデルの汎化性能検証結果

視聴覚コンテンツの臨場感評価のための視聴者の時系列感情分析および感情推定モデルの構築

本研究で選定した視聴覚コンテンツには、このように定常的な細かい揺れを含むコンテンツは多く含まれていなかったため、NNの学習が適切に行われなかったものと推察される。

6.2 Pleasant—Unpleasant

図1に示した各コンテンツのうち Pleasant の感情に特徴的な変化が見られたコンテンツ「7: passing roller coaster A (front)」 「20: scene of playing Japanese archery」について、そのモデル出力値を図9 (a), 9 (b) に示す。コンテンツ7では、感情評価実験における Pleasant の評価値が減少したタイミングでモデル推定値も減少していることが確認できる。また、コンテンツ20については、Pleasant の感情が増加したタイミングでモデルによる推定値も若干ではあるが増加していることが見て取れる。

次に、モデル汎化性能の検証の結果、平均誤差が大きかったコンテンツ「25: speakers reproducing music」について、そのモデル出力値を図9 (c) に示す。モデルによる推定値が感情評価実験の評価値と比較して、Unpleasant 側に寄っていることが確認できる。このコンテンツは、映像・音像のどちらも動かないコンテンツである。映像は常に正面に設置されたスピーカーを捉えており、コンテンツ全体を通して変動はない。一方、音は、スピーカーから流れている楽曲がそのまま収録されている。そのため、スピーカーから流れる楽曲の内容が本コンテンツから受ける感情に大きく寄与しているものと考えられる。本研究のモデル構築に使用した音特徴量 (A1) ~ (A6) は、楽曲聴取時など、複雑な要因によって形

成される感情の推定を想定したものではない。その結果、NNによる推定が困難であったものとする。

7. お わ り に

本研究では、まず、時系列で変化する視聴覚コンテンツと視聴者の感情や臨場感との関係を明らかにするために感情評価実験を実施した。実験の結果から、視聴者の感情に音のラウドネスや映像の彩度、輝度などの特徴量が関わっていることを明らかにした。また、臨場感と「Activation—Deactivation」の感情には、強い相関があることが明らかになった。次にニューラルネットワークを用い、映像、音の特徴量からコンテンツ視聴者の感情の時系列変化を推定するモデルを構築した。構築したモデルを検証し、十分な汎化性能を持つことを確認した。本研究で得られた知見や構築したモデルは、映像やゲーム、VRなどのコンテンツ設計における指針として用いることができる。また、視聴者の感情を考慮した高い臨場感をもたらすコミュニケーションシステムの研究・開発の進展に貢献するものであると考える。

謝 辞

本研究は、NICT 委託研究「革新的な三次元映像技術による超臨場感コミュニケーション技術の研究開発」の一環として行われた。コンテンツの映像・音特徴量の計測にご助力いただいた山梨大学工学部の塚原将太様に謝意を表す。

図8 モデルによるDeactivation—Activationの感情の時系列推定結果の一例

図9 モデルによるUnpleasant—Pleasantの感情の時系列推定結果の一例

参 考 文 献

- [1] Schroder, B. R. and Atal, B. S.: Computer simulation of sound transmission in rooms, Proc. IEEE, 7(3), pp.150-155, 1963.
- [2] Hamasaki, K., Nishiguchi, T., Okumura, R., Nakayama, Y. and Ando, A.: A 22.2 multichannel sound system for ultrahigh-definition TV, SMPTE Motion Imaging Journal, 117(3), pp.40-49, 2008.
- [3] Sugawara, M., Masaoka, K., Emoto, M., Matsuo, Y. and Nojiri, Y.: Research on human factors in ultrahigh-definition television to determine its specifications, SMPTE Motion Imaging Journal, 117(3), pp.23-29, 2008.
- [4] Okano, F., Arai, J., Hoshino, H. and Yuyama, I.: Real-time three-dimensional pickup and display system based on Integral Photography, Proc. SPIE, Conference on Novel Optical Systems Design and Optimization II, 3430, pp.70-79, 1998.
- [5] Witmer, B. G. and Singer, M. J.: Measuring presence in virtual environments: A presence questionnaire, Presence: Teleoperators and Virtual Environments, 7(3), pp.225-240, 1998.
- [6] Schubert, T., Friedmann, F. and Regenbrecht, H.: The experience of presence: Factor analytic insights, Presence: Teleoperators and Virtual Environments, 10(3), pp.266-281, 2001.
- [7] Ozawa, K., Chujo, Y., Suzuki, Y. and Sone, T.: Contents which yield high auditory-presence in sound reproduction, Kansei Engineering International, 3(4), pp.25-30, 2002.
- [8] 川除佳和, 海棠輝彦, 柴田啓司, 堀田祐弘: 符号化された映像と音響ビットレートが臨場感に与える影響の解析, 画像電子学会誌, 37(5), pp.671-677, 2008.
- [9] Ozawa, K., Chujo, Y., Suzuki, Y. and Sone, T.: Psychological factors involved in auditory presence, Acoustical Science and Technology, 24(1), pp.42-44, 2003.
- [10] 寺本渉, 吉田和博, 浅井暢子, 日高聡太, 行場次朗, 鈴木陽一: 臨場感の素朴な理解, 日本バーチャルリアリティ学会論文誌, 15(1), pp.7-16, 2010.
- [11] Kinoshita, Y., Fukue, K. and Ozawa, K.: Development of Kansei estimation models for the sense of presence in audio-visual content, Proc. 2011 IEEE International Conference on Systems, Man, and Cybernetics, pp.3280-3285, 2011.
- [12] 松村明編: 大辞林 第二版, 三省堂, 東京, 1995.
- [13] Nguyen, T. H. Y., 木下雄一朗, 小澤賢司, 郷健太郎, 森勢将雅: 視聴覚コンテンツの臨場感評価を目的とした視聴者の感情推定モデルの構築, 第15回日本感性工学会大会予稿集, C64, 2013.
- [14] Barrett, L. F. and Russell, J. A.: Independence and bipolarity in the structure of current affect, Journal of Personality and Social Psychology, 74(4), pp.967-984, 1998.
- [15] Rec. ITU-R BT.710-4: Subjective assessment methods for image quality in high-definition television, ITU, 1998.
- [16] Kuwano, S., Namba, S.: Continuous judgment of level-fluctuating sounds and the relationship between overall loudness and instantaneous loudness, Psychological Research, 47(1), pp.27-37, 1985.
- [17] Fastl, H. and Zwicker, E.: Psychoacoustics: Facts and models, Springer, New York, 2006.
- [18] Ozawa, K., Tsukahara, S., Kinoshita, Y. and Morise, M.: Development of an estimation model for instantaneous presence in audio-visual content, IEICE Transactions on Information and Systems, E99-D(1), pp.120-127, 2016.

木下 雄一朗 (正会員)

2007年 立命館大学大学院理工学研究科博士課程修了。同年, 山梨大学大学院医学工学総合研究部助教。2013年より同准教授。博士(工学)。主に, 感性情報処理, ヒューマンコンピュータインタラクションに関する研究に従事。日本感性工学会評議員, ヒューマンインタフェース学会, 情報処理学会, 日本知能情報ファジィ学会, ACM SIGCHI, IEEE 各会員。

伊藤 将亮 (非会員)

2015年 山梨大学工学部コンピュータ・メディア工学科卒業。現在, 山梨大学大学院医学工学総合教育部コンピュータ・メディア工学専攻所属。主に, 感性情報処理に関する研究に従事。日本音響学会会員。

小澤 賢司 (正会員)

1988年 東北大学大学院修士課程修了。同年 東北大学電気通信研究所助手。東北大学助教授, 山梨大学助教授を経て, 2007年 山梨大学教授。メディア感性工学, 音響信号処理などの研究に従事。日本音響学会理事。

森勢 将雅 (非会員)

2008年 和歌山大学大学院システム工学研究科博士後期課程修了。関西学院大学博士研究員, 立命館大学助教を経て, 2013年2月より山梨大学大学院総合研究部特任助教。博士(工学)。主に, 音声・聴覚情報処理に関する研究に従事。

日本音響学会, 電子情報通信学会, 情報処理学会, 日本騒音制御工学会各会員。