報 告

音と動画の意味的調和度の推定について

桑田 和也, 宝珍 輝尚, 野宮 浩揮

京都工芸繊維大学大学院工芸科学研究科

Study on the Estimation of the Semantic Harmony of Sound and Video Clips in the Audiovisual Material

Kazuya KUWATA, Teruhisa HOCHIN and Hiroki NOMIYA

Graduate School of Kyoto Institute of Technology, Goshokaido-cho, Matsugasaki, Sakyo-ku, Kyoto 606-8585, Japan

Abstract : It is experimentally clarified how the degree of the semantic harmony could be estimated in order to derive the degree of the general harmony. In the experiment, four patterns of audiovisual materials are used. These are created by combining the semantically matched and mismatched sound clips and the temporally matched and mismatched video clips. Subjects evaluate impression of the materials and semantic similarity of the sound and the video clips. This experiment clarifies that the degree of the semantic harmony could be estimated with certain accuracy by using the degree of the impressive similarity and that of the semantic one of the sound and the video clips.

Keywords : Audiovisual material, Harmony, Audio-visual interaction, Multiple linear regression analysis

1. は じ め に

近年, インターネットやパーソナルコンピュータの普及と進歩により, 画像・動画・音といったマルチメディアデータが身近なものとなってきている. そして, このようなマルチメディアデータを印象に基づいて相互に検索しようという試みも行われている [1]. ここでは, 例えば, 映像クリップを検索キーとしてその映像クリップに印象が合った音楽クリップの検索が可能である. しかし, 検索結果の音楽クリップは映像と無関係に再生されるのみで, それらの相互関係については考慮されていない.

ここで, 岩宮は, 映像と音・音楽を同時提示する場合において, 映像と音・音楽が持つ感性に影響を与える因子間の相互作用は, その処理レベルによって調和感から影響を受けることを示している [2]. 感性に基づくマルチメディアデータ検索システムでは, データベース中の映像や音・音楽の中からユーザの問合せに合致した度合い (合致度) の計算をシステム内で行う必要があり, 映像と音・音楽を同時提示する場合は, 両者の調和の度合いが印象にどの程度寄与するのかを明確にしなければならない.

また, メディア素材から受けるこのような調和感は, 意味的・印象的な側面での調和と時間的・構造的な側面での調和の2つのタイプの調和感に分類できることが指摘されている [2-5]. ここで, 意味的調和とは, 音・音楽と映像がそれぞれ持つ意味の類似性による調和である. 例えば, 鳩の鳴き声と鳩の映像は調和していることになる. 印象的調和とは, 音・音楽と映像がそれぞれ持つ印象の類似性による調和である. 例えば, 明るい音楽と明るい映像は調和していることになる. 時

間的調和とは, 音・音楽のテンポと映像の再生速度の類似性による調和である. 構造的調和とは, 音・音楽のアクセント構造と映像のアクセント構造の類似性による調和である.

著者らも, 一般的な映像作品と音・音楽を用いて, 音・音楽と映像との調和感がそれらを組み合わせた素材の印象に及ぼす影響, ならびに, 総合的な調和感における意味的・印象的調和感と時間的・構造的調和感の割合を明確化することを目的としてSD法による印象評価実験を行ってきた [11]. 印象評価を行う視聴覚素材として意味的・印象的調和感と時間的・構造的調和感がそれぞれ「合っている-合っていない」で分けられる4パターンの組み合わせ素材をいくつか作成し, その印象の評定をSD法により行った. この結果, 音・音楽と動画の組み合わせ素材の中で「自然な」「感動した」「美しい」という印象語の評点が高いものを得るには音・音楽と動画の総合的な調和感の評点が高くなければならないことや, 意味的・印象的調和感の評点と時間的・構造的調和感の評点による総合的な調和感の評点の推定が可能であることなどを確認してきた. ここで, 時間的・構造的調和度は, 音・音楽と動画のアクセントの同期の程度と再生速度から導出可能と考えられるが, 意味的・印象的調和度の導出については明確ではない.

そこで本論文では, 音・音楽と動画の意味的・印象的調和度の導出を目的として, まず, 音・音楽と動画の意味的な類似度について主観的な評価実験を行う. ここでは, 言語工学研究所のシソーラス [6] を参考にして意味の類似度の評価を行う. そして, ここで得られた音・音楽と動画が持つ意味の類似度と, 印象評価実験 [11] で得られた音・音楽の印象の評点と動画の印象の評点から, 音・音楽と動画を同時提示した際に受ける意味的・印象的調和感の評点を重回帰分析

を用いて推定する。

以降、まず、2.で先行実験[11]について概説する。次に3.で意味の類似度に関する研究について概説する。そして、4.で意味的類似度に関する実験を行う。5.で意味的・印象的調和度を推定し、6.で考察を行う。最後に7.でまとめる。

2. 先行実験[11]

著者らは、これまでに、以下を目的とした印象評価実験を行ってきた[11]。

- 音・音楽と映像との調和感がそれらを組み合わせた素材の印象に及ぼす影響の明確化
- 総合的な調和感における意味的・印象的調和感と時間的・構造的調和感の割合の明確化
- 音・音楽と動画のそれぞれの印象から、両者を同時提示した際の調和感の推定

2.1.では、この実験方法[11]について概説し、2.2.では、実験の結果について述べる。

2.1 実験方法

印象評定は以下の素材を以下の順序で行った。

- ① 音・音楽（約10秒）：20個
 - ② 動画（約10秒）：5個
 - ③ 音・音楽①と動画②を組み合わせた素材：20個
- 動画は、揺れているベル（電光看板）、踊っている男女（電光看板）、無数に湧き出る円、ジグザグに塗り重ねていくペンキ、端から崩れ落ちるタイルという動画である。

このような動画と組み合わせる音・音楽は、図1のように、音・音楽と動画が意味的・印象的に調和している／していないものと、音・音楽と動画が時間的・構造的に調和している／していないものを4パターンになるように組み合わせた。ここで、動画と時間的・構造的調和をしていない音・音楽素材は、動画と時間的・構造的調和をしている音・音楽素材の動画との同期とテンポをずらし、心理的に調和していないと著者の一人が感じるものを1種類作成する。また、動画と意味的・印象的に調和している・していない音・音楽素材は、素材集[12, 14, 15]の中から、著者の一人が意味的・印象

的に合っている・合っていないと思うものを選定する。例えば、「揺れているベル」の動画に対して、意味的・印象的に合っているものとして「鐘の鳴る音」を使用し、意味的・印象的に合っていないものとして「柱時計の振子の鳴る音」を使用する。1つの動画に対して、意味的・印象的に調和している音・音楽と意味的・印象的に調和していない音・音楽を使用し、時間的・構造的に調和しているものと調和していないものとを組み合わせ、意味的・印象的な調和と時間的・構造的な調和を同程度に考慮している。また、動画は素材集[13]の中から選出する。この素材集は、「光のイメージ」、「CGコンポジション」というカテゴリのものであり、カテゴリ内に多様な動画が収録されているが、素材集という性格上、強いメッセージ性を持たないものである。選定する音・音楽は、収録されている素材集[12, 14, 15]の中の「WORLD (short music)」や「自然」、「アニメ・ゲーム効果音」というカテゴリのものであり、演奏の中でも、組み合わせる動画と特に意味や印象が合っているものと合っていない部分を約10秒間抽出する。このような動画を12個、それぞれに対応する音・音楽を各4個、特定のカテゴリに偏らないように選定する。以上の素材から、(1)音・音楽と動画の調和感から組み合わせ素材の感動への影響が明確に現れるようにするため、「感動」の評定の絶対値が大きくない音・音楽、動画で、(2)図1の傾向が顕著に現れる組み合わせ素材のセットを選択したのが、本節の冒頭に記した5種の組み合わせ素材である[11]。

素材の提示順は素材の種類内でランダムであるが、音・音楽と組み合わせ素材については、1つの動画と組み合わせる4つのパターン間でそれぞれ比較しやすいように1セットにし、1セット内で順番に提示している。

上記の45個の素材に対してSD法で印象評定を行った。

①、②の単独提示する素材については、感性に基づくマルチメディアデータ検索システム[1]で用いられている16印象語対と「感動」に関する印象語対（表1参照）によって-2～2の5段階で評価している。③の組み合わせ素材については、前述の17個の印象語対に、「総合的な調和」に関する印象語対（完全にマッチしている－していない）、「意味的・印象的調和」に関する印象語対（意味的・印象的に調和している－していない）、ならびに、「時間的・構造的調和」に関する印象語対（時間的・構造的に調和している－していない）を追加した計20個の印象語対によって-2～2の5段階で評価している。

素材の評価は被験者にパソコン（17型液晶ディスプレイ）を用いて行ってもらった。素材はフリーのRealPlayer10.5で再生し、音・音楽はヘッドホンをつけて聴いてもらった。ヘッドホンはオーディオテクニカ製イヤフィットヘッドホンATH-EQ66を使用した。被験者は大学生の男子29名と女子5名の計34名であるが、男女間で印象語ごとに分散分析、及び、t検定を行った結果、平均値に有意な差がみられたので、以降は男子大学生29名の評点を使用する。

図1 視聴覚素材の調和の組み合わせ

2.2 実験結果

まず、2.1の印象評価実験で得られた17個の印象語対に対する評点の平均値を因子分析し、バリマックス回転を行って得られた因子負荷量を表2に示す。表2では、紙面の都合で、表1の印象語対の左側の印象語のみを示している。ここでは、固有値が1以上という基準で因子を得ている。それぞれの因子の因子負荷量の絶対値が大きい印象語をもとにして、5つの因子を「明快性」、「堅鋭性」、「自然性」、「力量性」、ならびに、「清涼性」と命名している [11]。

また、評価実験より、以下を明らかにしている [11]。

- ① 「自然な」、「感動した」、「美しい」といった印象語を含む「自然性因子」に含まれる印象語の評点が高い組み合わせ

表1 感性に基づくマルチメディアデータ検索システム [1] で使用されている16印象語対と「感動」に関する印象語対

NO.	印象語対
1	明るい — 暗い
2	うれしい — 悲しい
3	暖かい — 冷たい
4	緊張した — ゆったりとした
5	澄んだ — 濁った
6	大きい — 小さい
7	強い — 弱い
8	重い — 軽い
9	動的な — 静的な
10	美しい — 醜い
11	新鮮な — 古くさい
12	潤いのある — 濁いた
13	自然な — 不自然な
14	大胆な — 繊細な
15	単純な — 複雑な
16	鋭い — 鈍い
17	非常に感動した — 全く感動しなかった

表2 17個の印象語の評点に対する因子負荷量 (バリマックス回転後)

印象語	明快性	堅鋭性	自然性	力量性	清涼性
うれしい	0.944	0.125	0.256	-0.056	0.066
暖かい	0.900	-0.148	0.299	-0.011	-0.114
明るい	0.894	0.398	0.121	-0.103	0.076
潤い	0.826	0.079	0.274	0.061	0.208
鋭い	0.029	0.905	0.054	0.134	0.328
緊張した	-0.444	0.787	-0.197	0.325	-0.046
動的な	0.191	0.754	0.017	0.301	-0.501
新鮮な	0.195	0.730	-0.152	0.008	-0.097
重い	-0.603	-0.725	-0.008	0.185	-0.061
感動した	0.494	-0.108	0.853	0.043	-0.009
自然な	0.207	-0.088	0.809	-0.046	-0.095
美しい	0.539	-0.044	0.674	-0.079	0.409
大きい	0.017	-0.032	0.045	0.881	-0.010
強い	-0.138	0.282	-0.020	0.863	-0.128
大胆な	0.020	0.138	-0.300	0.532	-0.506
澄んだ	0.355	0.252	0.420	-0.059	0.752
単純な	0.014	-0.023	-0.105	-0.037	0.363
固有値	6.188	4.076	2.079	1.552	1.133
累積寄与率	36.402	60.379	72.607	81.735	88.399

わせ素材を得るには、音・音楽と動画の総合的な調和度の評点が高くなければならない。

- ② 音・音楽と動画の組み合わせ素材の総合的な調和度は、意味的・印象的調和度と時間的・構造的調和度によって推定が可能である。

従って、音・音楽と動画を同時提示する場合は、①より、両者の総合的な調和度を推定する必要があり、②より、そのためには意味的・印象的調和度と時間的・構造的調和度を推定する必要があることになる。

ここで、時間的・構造的調和度は音・音楽と動画のアクセントの同期の程度と再生速度から比較的容易に導出可能と考えられるが、意味的・印象的調和度の導出については明確ではない。

そこで、意味的・印象的調和度は音楽と映像がそれぞれ持つ印象や意味の類似性による調和である [2] ということに基づいて、本論文では、音・音楽と動画の意味的・印象的な類似度の導出について検討する。

3. 意味的類似度について

意味については、語を対象にした研究が主であるので、ここでは、語の意味的な類似度の研究について概説する。

3.1 語と語の意味的な関係

金田一春彦の日本語百科大事典 [7] によると、従来、語と語の意味関係は、ある特定の語と「意味」を完全に共有する語（同義語）、部分的に共有する語（類義語）、全く反対の意味を持つ語（反義語）などというようにとらえられてきた。また、語と語の意味関係は非常に複雑であり、図書館で本を探すときに、その本のタイトルや内容がどのジャンルに属しているかといったように、上位・下位の概念の情報処理が必要と考えられている。さらに、文中において、ある多義語がどの意味を指すのかという情報が必要であり、語と語の意味関係は単に個々の語だけでなく、語の体系あるいは構造の枠の中で考えなければならない。また、ある語が文中で用いられている時、その文中に用いられている他の語との関連で、特定の意味が出てくることがある。例えば、「君を駅まで送ろう」と「この本を彼に送ろう」という文章では、「送ろう」の意味が“送別”という意味と“配送”という意味で異なるといった具合である。

3.2 語と語の意味的類似度

3.2.1 シソーラスと語と語の意味的類似度

川島らは、シソーラスの木構造上の距離を利用した語と語の類似度の算出式を提案している [8]。一般的に、語Aと語Bの類似度は式 (1) で求められる。

$$\text{類似度} = \frac{2 \times A, B \text{の共通段数}}{A \text{の段数} + B \text{の段数}} \quad (1)$$

ここで、段数とはシソーラスの木構造の根を1段とし、それより意味クラスが1つ下位になるごとに1つずつ加算した

ものであり、共通段数は、語A、Bで共通する上位意味クラスの中で最も下の階層に位置するクラスの深さである。

ただし、2.で印象評価を行った音・音楽と動画の組み合わせに対し、式(1)に示す類似度計算法[8]を日本語語彙体系[9]や角川類語新辞典[10]を用いて試してみたところ、算出された意味的類似度は著者らの主観的な評価と大きく異なるものが数多く見られる結果となった。

3.2.2 語と語の関係の種類と意味的類似度

言語工学的研究所は、内部ノードにも語が存在するシソーラスの木構造上で、比較する語と語のカテゴリ間の枝の数によって類似度を算出する方法を提案している[6]。ここでは1組のノード間の距離(重み)を関係の種類ごとに表3のように定義している。

4. 意味的類似度評価実験

ここでは、音・音楽と映像が持つ意味の類似度を被験者が直接的に主観で評価し、その評点を用いて意味的調和度の推定を行う。

4.1 実験方法

2.1で評価を行った20個の組み合わせ素材のうち、「意味的・印象的に調和していて時間的・構造的に調和している組み合わせ素材」を5個と「意味的・印象的に調和してなくて時間的・構造的に調和している組み合わせ素材」を5個の計10個の組み合わせ素材について意味的類似度評価実験を行う。これらの素材を用いた理由は、時間的・構造的に調和していない組み合わせ素材は、それぞれ同じ動画に同じ音・音楽の再生速度を若干ずらしただけであるので、時間的・構造的に調和している組み合わせ素材と意味的類似度はほとんど変わらないと考えられるからである。

素材の提示順は、意味的・印象的に調和している／してい

ない素材間で同じ動画の組み合わせ素材が2個ずつあるのでそれぞれ1セットにし、比較しやすいようにセット内で順番にした上で、セットごとにランダムで提示する。

素材に対する評価は、3.2.2で述べた言語工学的研究所が提案している単語間の意味的類似度の算出方法[6]を参考に、音・音楽と動画の意味的類似度の算出に適応させるために表4のように解釈した5項目によってそれぞれ-2～2の5段階で評価する。ただし、狭義性・広義性については、組み合わせ素材における音・音楽と動画の狭義性・広義性が合っている程度を直接点数化するのが被験者にとって困難であったので、1つの組み合わせ素材における音・音楽と動画の抽象レベルをそれぞれ評価してもらい、式(2)で定義する音・音楽と動画の抽象レベルの差の大きさ(非類似度) D_{abs} として利用する。

$$D_{abs} = |W_s - W_m| \quad (2)$$

ここで、 W_s は音・音楽の抽象レベルの評点、 W_m は動画の抽象レベルの評点であり、 D_{abs} は0～4の5段階となる。

評価を行う環境は2.1と同様であり、被験者は男子大学生9名である。

4.2 実験結果

4.1の意味的類似度評価実験で得られた、各素材における意味的類似度に関する各項目の評点の平均値を図2に示す。ここで、同義性、反義性、関連性、共起性の評点は-2～2の範囲をとるが、抽象レベルの差の大きさの評点は0～4の範囲をとる。図より、全ての素材において、意味的に調和していない組み合わせ素材よりも意味的に調和している組み合わせ素材の同義性、反義性、関連性、共起性の評点は大きい。音・音楽と動画の抽象レベルの差の大きさについては不定である。

次に、同義性、反義性、関連性、共起性、抽象レベルの差の大きさの評点に対して因子分析を行った。バリマックス回

表3 言語工学的研究所が提案するノード間の距離の定義 [6]

関係	距離	内容
同義語	1	ここでは原則としてお互いに置き換えられるもの 例：犬→ドッグ
狭義語	2	ある語が示す意味の範囲に幅があるとき、狭く限定した方の意味の語 ここでは原則として属性の同じもの 例：犬→秋田犬(下位)
広義語	2	一つの言葉のもつ意味の範囲に幅があるとき、広い方の意味の語 ここでは原則として属性の同じもの 例：犬→哺乳類(上位)
反義語	2	意味が対立する語の関係 例：強いー弱い
関連語	4	ある程度の意味的な関連性を持つ語の関係 例：犬ーキツネ、自動車ータイヤ
共起語	2	係り受け関係を構成する語 主語ー述語の関係や修飾ー被修飾の関係等 例：本(を)ー読む、望遠鏡(で)ー見る

表4 音・音楽と動画の意味的類似度の評価項目

評価項目	解釈・評価基準
同義性	何が何をしているのか、動画の内容と音の内容が直接的に合っている程度 同義性がある2ー-2 同義性がない
狭義性・広義性	音・音楽や動画の内容の抽象レベルで、組み合わせた音・音楽と動画でそれぞれ評価する 狭義的(具体的)である2ー-2 広義的(抽象的)である
反義性	ある程度の関連性が認められる音・音楽と動画の意味の方向性が同じで程度で、関連性そのものが無ければ評価は0 意味的な方向が同じ2ー-2 意味的な方向が逆
関連性	音・音楽と動画に関連性があるかで、直接的に内容が合っているかは考慮しない 関連性がある2ー-2 関連性がない
共起性	動画の内容と音の内容の間で係り受け関係が成り立つか否か 係り受け関係がある2ー-2 係り受け関係がない

図2 音・音楽と動画の意味的類似度に関する項目の評点

((a) 揺れているベル (電光看板), (b) 踊っている男女 (電光看板), (c) 無数に湧き出る円, (d) ジグザグに塗り重ねていくペンキ, (e) 端から崩れ落ちるタイル)

転を行って得られた因子負荷量を表5に示す。ここでは、固有値が1以上という基準で因子を得ている。得られた1つの因子は同義性、反義性、関連性、共起性を説明変数とする因子であり、この因子を「意味的類似性」と呼ぶことにする。

5. 意味的・印象的な調和度の推定

2.1の実験で得られている音・音楽の印象の評点と動画の印象の評点、ならびに、4.1の実験から得られた音・音楽と映像が持つ意味的類似度の評点から、組み合わせ素材の意味的・印象的調和感の評点の推定を重回帰分析により試みる。

5.1 説明変数

まず、意味的・印象的調和度に含まれると考えられる印象的な類似度は、音・音楽と動画を同時提示する際に生じる視覚的な印象と聴覚的な印象の相互作用から感受されるものと想定し、意味的・印象的調和度を推定する重回帰式の説明変数には以下のような形の相互作用を含むものとする。

a. 音・音楽の印象の評点と動画の印象の評点の差の大きさ
音・音楽と動画の同じ印象語の評点の差が小さいほど調和感が高くなると考えられる。

b. 音・音楽の印象の評点と動画の印象の評点の積
音・音楽と動画の同じ印象語の評点と同じ方向 (符号) で絶対値が大きいくほど調和感が高くなると考えられる。

ここでは、多重共線性を回避するために、音・音楽と動画の同じ印象語の相互作用をそれぞれ表す17個の説明変数ではなく、2.2で述べた17個の印象語の評点を因子分析して得られた5個の因子の因子得点より、a.とb.を両方含んだ10個の説明変数を重回帰モデルに用いる。

また、意味的・印象的調和度に含まれると考えられる意味的な類似度についても、多重共線性を回避するために、意味的類似性因子 (表5) の因子得点と抽象レベルの差の大きさ D_{abs} を重回帰モデルの説明変数として用いる。

以上の説明変数をまとめて表6に示す。これらの説明変数を意味的調和度を推定する重回帰モデルに用い、変数選択により最終的に必要な説明変数だけを採用する。

5.2 加法回帰モデルによる推定

最終的な重回帰モデルは加法回帰モデルを目標とし、変数選択により説明変数が十分に少なくなれば、説明変数間の相

表5 意味的類似度の評価項目に対する因子負荷量 (抽象レベルの差の大きさを除く、バリマックス回転後)

項目名	意味的類似性因子
同義性	0.998
反義性	0.979
関連性	0.959
共起性	0.946
固有値	3.829
累積寄与率	95.726

相互作用を考慮した項の追加も試みる。加法回帰モデルはここでは式 (3) で表されるモデルである。

$$y = f_{F1-}(x_{F1-}) + \dots + f_{F5-}(x_{F5-}) + f_{F1*}(x_{F1*}) + \dots + f_{F5*}(x_{F5*}) + f_{SF}(x_{SF}) + f_{SDabs}(x_{SDabs}) + c \quad (3)$$

ここで、 y は被説明変数、 x_i は説明変数、 f_i は説明変数の変換を行う関数である。

次に、説明変数の変換を行う関数 f_i をそれぞれ決定するために、12個の説明変数からそれぞれ1つずつ意味的・印象的調和感の評点を目的変数にした単回帰分析を試み、自由度修正済み決定係数 R^2 の値が最も高い変換をその説明変数の変換関数 f_i とする。説明変数の変換のパターンは、説明変数に対する意味的・印象的調和感の評点の散布図を参考にした結果、3次以下の多項式、指数関数、対数関数、ロジスティック関数による変換のいずれかとする。表6の12個の説明変数を用いて、それぞれ単回帰分析での変数変換を試みた結果を表7に示す。表では、各説明変数で最大の自由度修正済み決定係数 R^2 を持つ変換の箇所をハッチングで示している。値が0のものは、目的変数との相関が低すぎて多項式の中の変数選択で必要とされる項が無かったものや、係数の初期値が適正に定まらなかった変換である。また、ここでの係数の p 値は必ずしも5%の水準を満たしていない。これは、加法回帰モデルにした際、係数の p 値は増減するためである。

式 (3) に自由度修正済み決定係数 R^2 の値が最も高い変換を行う関数 f_i をそれぞれあてはめ、係数の p 値 (0.05以下) とAIC (Akaike's Information Criterion) を基準にした変数増減法により必要な説明変数を選択した結果、ロジスティック変換を行った意味的類似性 x_{SF} のみが残った。しかし、ロジスティック変換を行った x_{SF} が意味的・印象的調和度の推定に有意であっても、ロジスティック関数内の係数で5%の水準を満たしていないものがあり、また、 x_{SF} 以外の変数を含めた多重ロジスティック関数によって回帰分析を行っても係数の有意水準が向上する傾向は見られなかった。そこで、表7より、ロジスティック関数による変換が最大の自由度修

正済み決定係数 R^2 の値を持つ説明変数は、多項式による変換では R^2 の値が若干低下する程度なので、式 (3) でロジスティック関数であった f_i を多項式に置き換えた式に変更し、再度、ここから必要な説明変数を変数増減法により選択した。この結果、式 (3) は式 (4) でまとめられ、意味的調和度 D_{SH} が推定される。

式 (4) の重回帰分析の自由度修正済み決定係数の値は0.817であり、分析の p 値は 2.10×10^{-7} である。

$$D_{SH} = c'_1 x_{SF}^3 + c'_2 x_{SF} + c'_3 \quad (4)$$

表7と R^2 の値が異なるのは、加法回帰モデルにする前の単回帰分析の時点では R^2 の値だけに着目し、説明変数の係数の p 値が0.05以上のものが残っていたためである。また、説明変数に関する値をまとめて表8に示す。分析全体の有意水準と各説明変量の有意水準はいずれも1%の水準で有意と認められる。多重共線性を表すトレランスの値は0.1を下回っているが、今回のような多項式に変換した場合の変数間ではある程度の従属性は当然であり、偏回帰係数の絶対値が極端に大きくないので大きな問題はないと考えられる。

また、式 (4) の意味的類似性 x_{SF} のみによる意味的・印象的調和度の推定と比較するために、表6で示した印象的な類似度に関する $x_{F1-} \sim x_{F5*}$ の10個の説明変数のみで意味的・印象的調和度を推定する。式 (3) から $f_{SF}(x_{SF})$ と $f_{SDabs}(x_{SDabs})$ を除いた式から、前述と同様の基準による変数増減法を用いて変数選択を行った結果は式 (5) でまとめられる。

$$D_{SH} = c''_1 x_{F1-}^3 + c''_2 x_{F1-} + c''_3 \quad (5)$$

式 (5) の重回帰分析の自由度修正済み決定係数 R^2 の値は0.471であり、分析の p 値は 1.73×10^{-3} である。ここでも表7と R^2 の値が異なるのは式 (4) と同様の理由である。また、説明変数に関する値をまとめて表9に示す。分析全体の有意水準と各説明変量の有意水準はいずれも5%の水準で有意と認められる。ここでも多重共線性を表すトレランスの値は0.1を下回っているが、式 (4) と同様の理由で大きな問題はな

表6 重回帰モデルに用いる説明変数

説明変数	内容
x_{F1-}	音の明快性－動画の明快性
x_{F2-}	音の堅鋭性－動画の堅鋭性
x_{F3-}	音の自然性－動画の自然性
x_{F4-}	音の力量性－動画の力量性
x_{F5-}	音の清涼性－動画の清涼性
x_{F1*}	音の明快性×動画の明快性
x_{F2*}	音の堅鋭性×動画の堅鋭性
x_{F3*}	音の自然性×動画の自然性
x_{F4*}	音の力量性×動画の力量性
x_{F5*}	音の清涼性×動画の清涼性
x_{SF}	意味的類似性
x_{SDabs}	音・音楽と動画の抽象レベルの差の大きさ

表7 単回帰分析において説明変数の変数変換を行った自由度修正済み決定係数 R^2

説明変数	多項式	指数関数	対数関数	ロジスティック関数
x_{F1-}	0.550	0.401	0.400	0.670
x_{F2-}	0.081	0.000	0.032	0.000
x_{F3-}	0.171	0.000	0.021	0.000
x_{F4-}	0.000	0.000	-0.047	0.000
x_{F5-}	0.000	0.000	0.000	0.000
x_{F1*}	0.187	0.195	0.180	0.272
x_{F2*}	0.118	0.145	0.137	0.160
x_{F3*}	0.063	0.000	0.000	0.000
x_{F4*}	0.000	0.000	0.000	0.000
x_{F5*}	0.000	0.016	0.000	0.014
x_{SF}	0.834	0.661	0.661	0.859
x_{SDabs}	0.130	0.000	0.000	0.000

いと考えられる。また、式 (5) において x_{F1-} のみが残っているのは、 x_{F1-} が明快性に関する因子であり、表2から分かるようにこの因子は印象を表す第一因子であり、印象を代表する因子であるとも考えられることから、この因子が選択されたものと考えられる。

5.3 相互作用を考慮した重回帰モデルによる推定

式 (4) を最終的な重回帰モデルとすると、表6で示した印象的な類似度に関する $x_{F1-} \sim x_{F5-}$ の10個の説明変数は意味的・印象的調和度の推定には必要がないことになる。そこで、単回帰分析における自由度修正済み決定係数 R^2 の値の高さや、5.2で式 (4) や式 (5) が導き出されたことから、意味的・印象的調和度の推定に必要な説明変数を x_{F1-} と x_{SF} に限定し、2つの説明変数の相互作用を考慮した重回帰モデルをいくつか試みた。その結果、式 (6) による推定モデルを提案する。

$$D_{SH} = c_1 x_{F1-}^3 + c_2 x_{F1-}^2 + c_3 x_{F1-} + c_4 x_{SF}^3 + c_5 x_{SF}^2 + c_6 x_{SF} + c_7 (x_{F1-} \cdot x_{SF})^3 + c_8 (x_{F1-} \cdot x_{SF})^2 + c_9 (x_{F1-} \cdot x_{SF}) + c_{10} \quad (6)$$

式 (6) の相互作用モデルから5.2と同様の基準による変数増減法を用いて変数選択を行うと、式 (6) は式 (7) でまとめられる。

$$D_{SH} = c_7 (x_{F1-} \cdot x_{SF})^3 + c_8 (x_{F1-} \cdot x_{SF})^2 + c_9 (x_{F1-} \cdot x_{SF}) + c_{10} \quad (7)$$

式 (7) の重回帰分析の自由度修正済み決定係数 R^2 の値は0.844、分析のp値は 2.82×10^{-7} である。説明変数に関する値をまとめて表10に示す。分析全体の有意水準と各説明変量の有意水準はいずれも5%の水準で有意と認められる。

6. 考 察

意味的な類似度に関する x_{SF} のみによる推定の式 (4)、印象的な類似度に関する x_{F1-} のみによる推定の式 (5)、ならびに、 x_{SF} と x_{F1-} の相互作用による推定の式 (7) から得られるそれぞれの意味的・印象的調和度の推定値と実測値をまとめて図3に示す。また、1つの動画に対する組み合わせ素材のセットごとに、式 (4)、式 (5)、ならびに、式 (7) の推定値と実測値との差を2乗し平均した結果を表11にまとめて示す。

意味的な類似度に関する x_{SF} のみによる重回帰分析では、自由度修正済み決定係数 R^2 の値が0.817と高いことや図3の傾向より、ある程度の精度の推定が期待できる。しかし、図3では、項目のラベルが9の素材から10の素材にかけて意味的・印象的調和度の実測値が大きく下がっているが、 x_{SF} のみによる推定ではどちらも同じ意味的・印象的類似度から同じ推定値を算出している。つまり、同じ動画と同じ音・音楽の組み合わせでも、時間的・構造的な調和をずらすことで意味的・印象的調和度も下がるような組み合わせ素材に対して、 x_{SF} のみによる重回帰分析では、時間的・構造的な調和をずらすことによる意味的・印象的調和度の低下に対応できない問題

表8 式 (4) による意味的調和度の重回帰分析結果

説明変数	偏回帰係数	p値	トレランス
x_{SF}^3	-0.915	8.44×10^{-04}	0.0817
x_{SF}	2.25	8.40×10^{-06}	0.0817
定数項	0.393	1.37×10^{-03}	NA

表9 式 (5) による意味的調和度の重回帰分析結果

説明変数	偏回帰係数	p値	トレランス
x_{F1-}^3	1.59	0.0359	0.0293
x_{F1-}^2	-3.56	0.0104	0.0293
定数項	1.93	6.72×10^{-4}	NA

表10 式 (7) による意味的・印象的調和度の重回帰分析結果

説明変数	偏回帰係数	p値	トレランス
$(x_{F1-} \cdot x_{SF})^3$	-0.999	1.53×10^{-3}	0.121
$(x_{F1-} \cdot x_{SF})^2$	-0.667	1.73×10^{-2}	0.377
$x_{F1-} \cdot x_{SF}$	1.98	1.31×10^{-6}	0.196
定数項	0.768	1.14×10^{-4}	NA

表11 意味的・印象的調和度の実測値と推定値との差の2乗の平均

	説明変数		
素材	x_{SF} のみ	x_{F1-} のみ	相互作用
揺れているベル (電光)	0.098	0.162	0.063
踊っている男女 (電光)	0.118	0.487	0.161
無数に湧き出る円	0.258	0.198	0.197
ジグザグに塗り重ねていくペンキ	0.141	0.973	0.057
端から崩れ落ちるタイル	0.193	0.515	0.170
合計	0.808	2.335	0.649

がある。また、同じ動画と同じ音・音楽の組み合わせで時間的・構造的な調和をずらすことで意味的・印象的調和度が下がるということは、「意味的・印象的に調和しているーしていない」という意味的・印象的調和感に関する印象語対と「時間的・構造的に調和しているーしていない」という時間的・構造的調和感に関する印象語対を用いた評価方法では、意味的・印象的調和感とは時間的・構造的な調和調和感に何らかの相互作用を受けていると考えられる。

印象的な類似度に関する x_{F1-} のみによる重回帰分析では、自由度修正済み決定係数 R^2 の値が0.471程度しかないことや図3の傾向、ならびに、表11で意味的・印象的調和感の実測値と推定値との差の大きさが他の重回帰モデルより大きいことから、あまり精度の良い推定が期待できない。また、表11から、「ジグザグに塗り重ねていくペンキ」の組み合わせ素材については、他の素材に比べて意味的・印象的調和感の実測値と推定値との差の2乗の値が非常に大きく、 x_{SF} のみによる重回帰分析の推定値と比較すると約7倍にもなる。これは、「ジグザグに塗り重ねていくペンキ」という内容の動画に、キュッキュッキュッという「マーカーで塗りつぶすよ

図3 意味的・印象的な調和度の推定結果の比較

素材番号：1～4－揺れているベル（電光看板），5～8－踊っている男女（電光看板），9～12－無数に湧き出る円，
13～16－ジグザグに塗り重ねていくペンキ，17～20－端から崩れ落ちるタイル

項目のラベルを4で割った余り：1－音・音楽と動画が意味的にも時間的にも合っている，2－音・音楽と動画が意味的にのみ合っている，3－音・音楽と動画が時間的にのみ合っている，0－音・音楽と動画が意味的にも時間的にも合っていない

うな音」や「救急車のサイレンの音」を組み合わせることにより生じる意味的調和感が，動画の印象と音・音楽の印象の類似性よりも，動画の内容と音・音楽の内容のシンボリックな意味の類似性による影響が強いためと考えられ，このような組み合わせ素材の推定には x_{FI-} のみによる重回帰分析は適さないと考えられる。

意味的な類似度に関する x_{SF} と印象的な類似度に関する x_{FI-} の相互作用による重回帰分析では，自由度修正済決定係数 R^2 の値が0.844と他の重回帰分析よりも高く，表11で実測値と推定値の誤差が著しく大きいものがないことから分かるように， x_{SF} のみによる推定と x_{FI-} のみによる推定の前述の問題点が互いに補われる傾向が見られる。また，他の重回帰分析より総合的に実測値と推定値との差が小さいことから，他の重回帰分析より精度の良い推定が期待できる。

7. ま と め

本論文では，印象評価実験[11]で得られた音・音楽の評点と動画の評点，ならびに，意味的類似度評価実験で得られた音・音楽と動画が持つ意味的類似度から，音・音楽と動画を同時提示した際に受ける意味的・印象的調和感の評点を推定する重回帰分析を試みた。この結果より以下の可能性を得た。

- 音・音楽と動画の意味的・印象的な調和度は，主に，音・音楽と動画の印象的な類似度よりも音・音楽と動画が持つシンボリックな意味の類似度に強く影響を受ける。
- 「意味的・印象的に調和している－していない」という意味的・印象的調和感に関する印象語対と「時間的・構造的に調和している－していない」という時間的・構造的調和感に関する印象語対を用いた評価方法では，意味的・印象的調和感は時間的調和感に何らかの相互作用を受けていると考えられる。
- 音・音楽と動画が持つ意味の類似度と，音・音楽と動画の印象（明快性）の類似度の相互作用による重回帰分析では，どのような組み合わせ素材でも一定の精度の推定が期待できそうである。

以上を検証するために，より多くの被験者やより多くのサンプルを用いた評価，ならびに，重回帰分析以外の手法による解析により，上記の可能性を検証することが今後の課題である。特に，動画サンプルに関しては，2.1で述べたように，カテゴリに偏りのないように選出した12個の動画から予備実験により5個を選定しているが，選定された5個に偏りがないとはいき切れず，今後，多くの偏りのないサンプルを用いた評価が必要である。また，これまでの考察は，男子大学生を対象とし，強い意味づけのない動画を対象として検討を

行ってきた。女性やより幅広い年齢層を対象とする検討、強い意味づけを持った動画を対象とする検討も今後の課題である。また、本論文では音・音楽と映像が持つ意味的・印象的類似度の定量化を主観的評価実験により試みた。ここで、個々のメディアデータが持つ意味を表現する汎用的な方法があれば、それを利用して2種のメディアデータの意味的・印象的類似度を求めることも可能ではないかと考えられる。このような個々のメディアデータが持つ意味の汎用的な表現方法に関する検討も今後の課題である。さらに、本実験では、音・音楽と動画の印象（明快性）の類似度と意味的な類似度の積の絶対値があまり大きくない条件で推定を行ったものであり、この絶対値が大きい条件での推定も今後の課題である。

謝 辞

本研究は、一部、文部科学省科学研究費補助金（課題番号：21500199）による。

参 考 文 献

- [1] Hochin, T., Tsuji, T.: Mutual Multimedia Access Using Kansei Factors, *Kansei Engineering International*, Vol. 2, No. 4, pp.9-18 (2001).
- [2] 岩宮眞一郎：音楽と映像のマルチモーダル・コミュニケーション、九州大学出版会（2000）。
- [3] 岩宮眞一郎：映像作品における視聴覚コミュニケーション、信学技報、HIP2002-42, pp.39-46（2002）。
- [4] Bolivar, V. J., Cohen, A. J. and Fentress, J. C.: Semantic and Formal Congruency in Music and Motion Pictures: Effects on the Interpretation of Visual Action, *Psychomusicology*, Vol. 13, pp.28-59 (1994).
- [5] Lipscomb, S. D.: The Perception of Audio - visual Composites: Accent Structure Alignment of Simple Stimuli, *Selected Reports in Ethnomusicology*, Vol. 12, pp.37-67 (2005).
- [6] 言語工学研究所：http://www.gengokk.co.jp/（2009.1）。
- [7] 金田一春彦、林大、柴田武：日本語百科大事典、大修館書店（1988）。
- [8] 川島貴広、石川勉：言葉の意味の類似性判別能力に関するシソーラスと概念ベースの性能比較、人工知能学会全国大

会（第18回）論文集，2D2-10，pp.152-154（2004）。

- [9] 池原悟，宮崎正弘，白井諭，横尾昭男，中岩浩巳，小倉健太郎，大山芳史，林良彦：日本語語彙体系，岩波書店（1997）。
- [10] 大野晋，浜西正人：角川類語新辞典，角川書店（1981）。
- [11] 桑田和也，宝珍輝尚：視聴覚素材における音と動画の調和について，日本感性工学会論文誌，Vol. 9, No. 1, pp.85-92（2009）。
- [12] ムーンポケット：音・辞典，データクラフト。
- [13] ムーンポケット：ビデオ素材辞典，データクラフト。
- [14] (株)ウィネット：Music Wizard, (株)ウィネット（1995）。
- [15] VOLT-AGE：バックの鬼 Digital Sound Clip Vol. 1 「響」，VOLT-AGE（1994）。

桑田 和也（非会員）

2007年京都工芸繊維大学工学部電子情報工学科卒業。2009年京都工芸繊維大学大学院工芸科学研究科情報工学専攻修了。同年、東京システムハウス株式会社入社。感性マルチメディアデータ検索、視聴覚素材の調和と

印象に関する研究に従事。

宝珍 輝尚（正会員）

1984年名古屋工業大学大学院修士課程修了。同年、日本電信電話公社入社。NTT電気通信研究所、福井大学、大阪府立大学を経て、2006年より京都工芸繊維大学大学院工芸科学研究科教授。博士（工学）。感性マルチメディアデータ検索、マルチメディアデータベースシステムの研究に従事。

日本感性工学会、電子情報通信学会、情報処理学会、日本データベース学会、IEEE、ACM、日本情報考古学会各会員。

野宮 浩揮（非会員）

2008年神戸大学大学院自然科学研究科博士課程修了。現在、京都工芸繊維大学工学部工学研究科助教。工学博士。この間、機械学習、画像認識、表情認識、マルチメディアデータベース等に関する研究に従事。