

感性異種メディアデータ検索での問合せ結果提示における

動画の再生速度の影響について

桑田 和也 宝珍 輝尚

京都工芸繊維大学大学院工芸科学研究科

概要

本稿では、視覚素材と聴覚素材の調和が印象に及ぼす影響を明らかにすることを目的とし、一般的な視覚素材を用いてそれらの相互作用を実験的に明らかにする。実験には、一般的な映像作品と音・音楽を組み合わせて視聴覚素材を作成し、それを基準として、視聴覚素材の動画の再生速度を3倍、ならびに、1/2倍にした3パターンの視聴覚素材の評価を行う。この実験により、動画の再生速度を3倍にすることで、一般的に、素材の印象は「緊張した」「動的な」「鋭い」ものへ変化し、一方、動画の再生速度を1/2倍にすることで、素材の印象が「単純な」「静的な」「鈍い」ものへ変化することがわかった。また、素材ごとの平均値の比較によって「調和」に有意差が認められた素材については、再生速度によって「調和」が変動する方向へ「感動」の評価も変動していることが確認された。

On the Effect of Speed of Playing Video in the Query Result of Kansei

Cross-Media Retrieval

Kazuya KUWATA Teruhisa HOCHIN

Graduate School of Science and Technology, Kyoto Institute of Technology

Summary

For the purpose of clarifying the influence that the harmony of the video material and the sound material exerts on the impression, the influence is experimentally clarified by using a general audiovisual material. In the experiment, an audiovisual material is made by combining a general video one with a sound or music one. The video materials having three times and the half times of the playing speed as well as the original one are used as the video materials. This experiment clarifies that the impression of the material generally changes into "Strained", "Dynamic", and "Sharp" by tripling the playing speed of a video clip, on the other hand, the impression of the material changes into "Simple", "Static", and "Slow" by making the playing speed a half. In addition, comparing mean values of each material, if it is a material from which a significant difference was admitted in "Harmony", it is confirmed that the evaluation of "Impression" changes in the direction where "Harmony" changes according to the playing speed, too.

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、視覚と聴覚の間に生ずる相互作用の解明が強く望まれ、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[3]。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。

岩宮は、視聴覚素材における視覚と聴覚の相互作用を体系的にモデル化し、また、音と映像の「時間的調和」に及ぼす諸要因などについても研究を行っている[1]。しかし、岩宮は、ディスプレイ上を周期的に運動するボールといった単純な視覚素材を用いて視覚と聴覚との相互作用を分析する実験を行っているのみであり、より一般的な視聴覚素材についての影響は明らかではない。

そこで本研究では、一般的な映像作品と音・音楽を用いることで、視覚素材（通常、3倍、1/2倍の再生速度）と聴覚素材との「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を明らかにすることを目的とし実験を行った。花畑や竹林などの自然物、都会の夜景や電光看板などの人工物、宇宙の光や木星の公転などといった一般的な視覚素材と、それらと「調和」がとれていると判断される音や音楽を組み合わせる視聴覚素材を作成し、その印象の評定を行う。また、それらの動画の再生速度を基準とし、視聴覚素材の動画の再生速度を3倍、ならびに、1/2倍にした3つのパターンの視聴覚素材について印象評定を行うことで、動画の再生速度が視聴覚素材の印象や「調和感」に与える影響を考察する。

以降、2.では音と映像に関する研究について概説する。3.で官能検査実験について述べ、4.で実験結果を示し考察を行う。最後に5.でまとめる。

2. 感性に影響を与える諸要因

2.1. 視覚と聴覚の相互作用

岩宮は、視覚と聴覚の相互作用には、感性に影響を与える因子に処理レベルの差があり、図 2.1 に示すモデルによって表現している[1]。

このような因子の処理レベルは上位であるほど視覚と聴覚の相互作用は複雑になり、例えば下位の要因の素材に対する「明暗」では、単純に「明るい」印象同士の音と映像を組み合わせることで、お互いの「明るい」という印象を強め合う共鳴現象が見られるが、上位の要因の素材に対する「評価」では、「音と映像の調和」がとれていなければ、高い「評価」同士の音と映像を組み合わせても、お互い素材の「評価」が良くなるとは限らないのである。

2.2. 音と映像の時間的調和に及ぼす諸要因

2.1で述べた「調和」とは、音と映像のそれぞれが持つ印象もしくは意味の類似性による調和である「意味的調和」と、音楽のリズムと映像の動きの同期的な関係によって生じる調和である「時間的調和」の双方を統合したものである[1]。

このうち、音と映像の「時間的調和」に影響を及ぼす要因は次のようにまとめられる[1]。

(1) 調和

映像と音楽のリズムが同期していれば視聴覚素材の調和感をもたらす。さらに映像の動きの速さと音楽のテンポのバランスがとれていると調和感を増大させる。

(2) インパクト

調和感ほどではないが音と映像の時間的アクセントの同期が、視聴覚素材に「印象深さ」を与える。また映像速度の速い素材の方が「印象深い」と評価される傾向が認められる。

(3) 軽快さ

音楽のテンポが速いほど軽快さは増す。

(4) 複雑さ

遅い映像速度と速い音楽テンポ、速い映像速度と遅い音楽テンポといった速度感の異なる組み合わせであると「複雑な」印象となる。

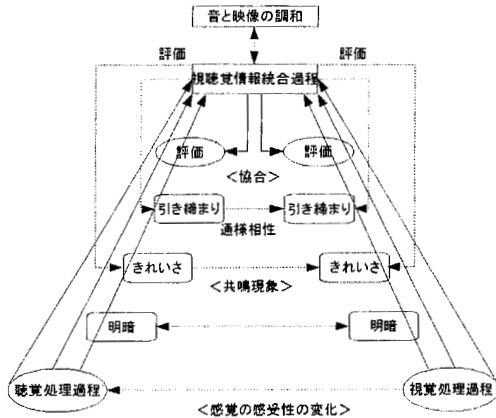


図 2.1 視覚と聴覚の相互作用のモデル化[1]

このような(1)~(4)をモデル化したもの[1]を図 2.2 に示す。

3. 実験方法

3.1. 印象評定を行う素材

本実験では、2.1 で述べた総合的な意味での「調和」がとれていると思われる視聴覚素材に対し、その動画部分の再生速度が視聴覚素材の印象にどのような影響があるのかを検討する。

印象評定を行う素材は以下の通りである。

- ① 音・音楽 (約 10 秒) : 8 個
- ② 動画 (約 10 秒) : 8 個
- ③ 音・音楽と動画を組み合わせさせた素材 : 8 個
- ④ ③の動画を 3 倍で再生させた素材 : 8 個
- ⑤ ③の動画を 1/2 倍で再生させた素材 : 8 個

以降、③、④、⑤の素材をそれぞれ通常再生素材、3 倍再生素材、1/2 倍再生素材と呼ぶ。また、音・音楽、動画の素材の詳細を表 3.1 に示す。なお、音・音楽と動画の素材の組み合わせには、同期化マルチメディア統合言語 : SMIL(Synchronized Multimedia Integration Language)により音・音楽と動画の再生を制御して同期させることで実現する。しかし、使用した再生プレイヤー(RealPlayer10.5)では SMIL2.0 の一部のモジュールが動作しなかったため、音・音楽の長さに合わせるための動画の繰り返し再生に

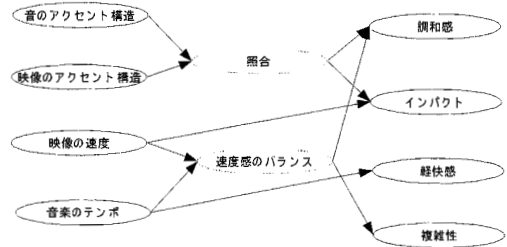


図 2.2 音楽と映像のアクセントの同期が映像作品の印象に及ぼす影響のモデル化[1]

表 3.1 音、音楽、動画の素材

素材名	素材概要	ファイル名	出典
音 1	宇宙船	05_39.wav	[4]
音 2	潮騒	01_36.wav	[4]
音 3	コマドリの鳴き声	01_21.wav	[4]
音 4	不気味で迫力のある音	09_16.wav	[4]
音楽 1	ジャズ	mm09.wav	[6]
音楽 2	お洒落な音楽	mm01.wav	[6]
音楽 3	明るいクラシック	s32k22s.wav	[5]
音楽 4	楽しげな音楽	s10k22s.wav	[5]
動画 1	宇宙の光	M01_007.MPG	[7]
動画 2	浜辺の鳥	M05_008.MPG	[7]
動画 3	竹林	M04_025.MPG	[7]
動画 4	公転する木星	M02_051.MPG	[7]
動画 5	上空からの夜景	M05_039.MPG	[7]
動画 6	バーの電光看板	M05_049.MPG	[7]
動画 7	一面の花畑	M04_007.MPG	[7]
動画 8	ジャグリング	video017.mpg	[3]

は手動で繰り返し回数を設定し、再生速度を制御する<speed>タグの代わりには、再生する動画ファイルを変換ソフト(SEffect 1.53)によって再生速度を変えたファイルを作成することで代替とした。

3.2. 実験手順

3.1 で述べた素材に対する印象評定は SD 法[3]

によって行う。①～②の単独表示する素材については感性の基本的な 16 印象語対、及び、感動に関する印象語対、実験時の被験者の気分、体調に関する印象語対の計 19 個の印象語対によって 5 段階で評価し、③～⑤の組み合わせ素材については、前述の 19 個の印象語対に「調和」に関する印象語対を 7 段階で評価したものを追加した計 20 個の印象語対によって評価する。

このような手順の実験を被験者にパソコン（17 型液晶ディスプレイ）で行ってもらい、メディア素材を RealPlayer10.5 で再生し、音と音楽はヘッドホンを付けて聴いてもらった。ヘッドホンは SONY 製オープンエアダイナミック型ステレオヘッドホン MDR-Q23SL、または、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 14 名と女子 2 名の計 16 名である。

3.3. 実験システム

印象評定システムは、印象評定の入力フォームから CGI を使って Java を実行させる Web ブラウザ上のシステムで行う。CGI(Common Gateway Interface)とは、Web サーバが Web ブラウザからの要求に応じてサーバ側でプログラムを起動させる仕組みである。Web サーバには AN HTTPD(ver.1.42.p)を使用している。

一つの素材の評価データは、素材名、19 個の印象語対（組み合わせ素材は「調和」に関する印象語対も追加）評点スコア、時間情報（送信時の年、月、日、時間、分）で構成されている。印象評定システムの画面の例を図 3.1 と図 3.2 に示す。

図 3.1 は素材選択画面であり、右フレームよりユーザ名、パスワード、気分、体調を入力し、ログインする。気分と体調については、ログイン後も実験途中で変更が可能である。ログイン後、左フレームより順次印象評定を行う素材を選択する。

素材を選択すると図 3.2 に示す素材の印象評定画面が現れる。この画面において、リンクをクリックして素材をプレイヤーで再生する。再生した素材に対し、印象語対に対して評価を行っていく。

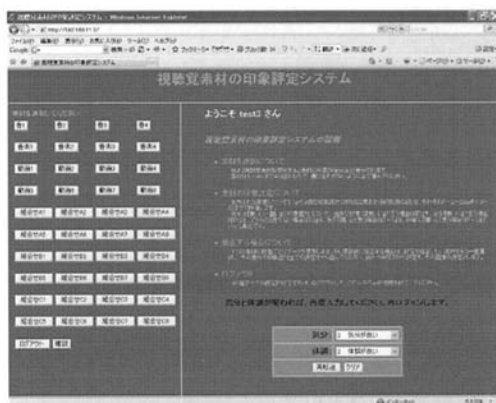


図 3.1 素材選択画面（ログイン後）



図 3.2 選択した素材の印象評定画面

4. 実験結果と考察

4.1. 評点スコアの解析

通常再生素材、3 倍再生素材、1/2 倍再生素材の組み合わせ素材について、印象語ごとの評点の平均値をレーダーチャートでプロットしたものを図 4.1 に示す。また、それらの値の間に有意差があるのかを有意水準 0.05 として t 検定を行った結果を表 4.1 に示す。

これらの結果より、一般的に、動画の再生速度を速くすることで視聴覚素材への印象は「緊張した」「鋭い」傾向が強くなり、動画の再生速度を遅くすることで「静的な」「単純な」「鈍い」傾向が

強くなっていることがわかる。

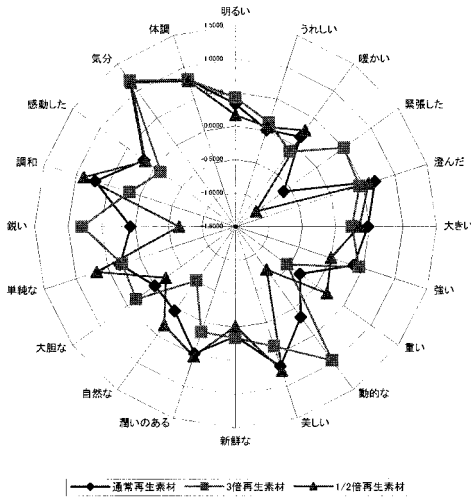


図 4.1 通常、3 倍、1/2 倍再生素材での平均評点

4.2. 因子分析

因子分析とは、多くの変数の値を少数の因子によって資料を説明する方法である。ここでは、全素材に対する因子を得るため「調和」に関する印象語対は因子分析に含めず、他の 19 印象語対の背後に潜む数個の共通因子を探り出して解析を行う。

対象とするデータは、音・音楽 8 個、動画 8 個、通常再生素材 8 個、3 倍再生素材 8 個、1/2 倍再生素材 8 個の計 40 行に、印象語対 19 列の 40×19 行列として被験者 16 名の平均値をスコアにしたものである。

このデータ行列の共通因子が持つ寄与量（固有値）と累積寄与率の第 7 因子までを表 4.2 に示す。この結果より、固有値が 1 以上であり、かつ、累積寄与率が 80% 以上となるまでの因子という両方の条件を満たす第 5 因子までを採用する。

共通因子数を 5 として因子分析を行い、バリマックス回転を行った後の因子負荷量を表 4.3 に示す。表 4.3 の因子の説明変数の箇所には色付けをしている。

表 4.1 p 値と検定結果

印象語対	通常-3 倍		通常-1/2 倍	
	p 値	有意差	p 値	有意差
明るい — 暗い	0.86	×	0.75	×
うれしい — 悲しい	0.77	×	0.91	×
暖かい — 冷たい	0.51	×	0.78	×
緊張した — ゆったりとした	0.036	○	0.35	×
澄んだ — 濁った	0.50	×	0.76	×
大きい — 小さい	0.41	×	0.65	×
強い — 弱い	0.79	×	0.28	×
重い — 軽い	0.55	×	0.23	×
動的な — 静的な	0.062	×	0.039	○
美しい — 醜い	0.33	×	0.81	×
新鮮な — 古くさい	0.93	×	0.58	×
潤いのある — 湯いた	0.32	×	0.90	×
自然な — 不自然な	0.14	×	0.47	×
大胆な — 繊細な	0.34	×	0.60	×
単純な — 複雑な	0.82	×	0.039	○
鋭い — 鈍い	0.0063	○	0.0050	○
完全にマッチしている — 完全にマッチしていない	0.23	×	0.72	×
非常に感動した — 全く感動しなかった	0.14	×	0.92	×
気分が良い — 気分が悪い	1.00	×	0.033	○
体調が良い — 体調が悪い	1.00	×	1.00	×

表 4.2 寄与量（固有値）と累積寄与率

因子	固有値	寄与率	累積寄与率
1	7.053	37.119	37.119
2	3.194	16.813	53.933
3	2.662	14.01	67.943
4	1.627	8.565	76.508
5	1.575	8.289	84.797
6	1.076	5.662	90.459
7	0.476	2.506	92.965

表 4.3 因子負荷量（バリマックス回転後）

印象語	評価性	明快性	力量性	堅鋭性	コンディション
明るい	0.160	0.973	-0.099	0.090	0.068
うれしい	-0.051	0.942	0.032	0.022	0.090
暖かい	0.317	0.800	-0.158	-0.332	0.005
緊張した	-0.585	-0.256	0.372	0.598	-0.024
澄んだ	0.895	0.240	-0.121	0.231	0.036
大きい	-0.009	-0.126	0.914	-0.059	0.051
強い	-0.313	-0.062	0.831	0.324	0.080
重い	-0.113	-0.676	0.565	-0.333	-0.169
動的な	-0.693	0.050	0.125	0.511	0.182
美しい	0.922	0.072	-0.110	-0.015	0.019
新鮮な	0.644	0.023	0.181	0.447	0.053
潤い	0.808	0.142	-0.243	0.044	0.107
自然な	0.813	0.052	0.051	-0.057	0.045
大胆な	-0.802	-0.076	0.490	0.082	-0.030
単純な	-0.139	-0.224	0.078	-0.136	-0.307
鋭い	0.098	0.071	-0.026	0.928	0.142
感動した	0.776	0.079	-0.021	-0.094	-0.040
気分	-0.108	0.037	0.121	0.072	0.981
体調	0.048	0.045	0.012	0.026	0.848

主因子の説明変数の決定方法を以下に示す。

- 各印象語対に対して、次の条件を共に満足する因子を選択する。
 - 因子負荷量の絶対値が最大である。
 - 因子負荷量の絶対値が 2 番目のものよりも 2 倍以上大きい。
- ある因子に対して、1.で印象語対が 1 個以上選択された場合、それらの印象語対がその因子を説明するものとする。
- 全ての因子に説明する印象語対が決定された場合、終了する。決定されていない因子がある場合、その因子に対して、因子負荷量が最大であるという条件のみで 1.から繰

り返す。

この方法で得られた説明変数と、それに基づいて因子の特徴をとらえた因子名を以下に示す。

- 評価性：[澄んだ—濁った]、[美しい—醜い]、[潤いのある—渴いた]、[自然な—不自然な]、[非常に感動した—全く感動しなかった]
- 明快性：[明るい—暗い]、[うれしい—悲しい]、[暖かい—冷たい]
- 力量性：[大きい—小さい]、[強い—弱い]、
- 堅鋭性：[鋭い—鈍い]
- コンディション：[気分が良い—気分が悪い]、[体調が良い—体調が悪い]

4.3. 組み合わせ素材の「調和」と他の印象語対の関係

通常再生素材、3 倍再生素材、1/2 倍再生素材の組み合わせ素材について、「調和」の評点ベクトルを説明変数とし、他のそれぞれの印象語対の評点ベクトルを目的変数として回帰分析を行うことでそれらの関係を調べた結果を表 4.4 に示す。また、5%の水準で相関が有意であると認められた印象語対の箇所には色付けを行っている。

表 4.4 より、「調和」が高くなるにつれて、「動的な—静的な」と「大胆な—繊細な」の印象語対が負の方向へ変化し、「美しい—醜い」、「自然な—不自然な」と「非常に感動した—全く感動しなかった」の印象語対が正の方向へ変化するので、視聴覚素材への印象が、「静的な」、「繊細な」、「美しい」、「自然な」、「非常に感動した」ものになることがわかる。これら 5 つの印象語対は、いずれも表 4.3 で示した「評価性」因子の説明変数、もしくは 5 つの因子の内でも「評価性」因子の因子負荷量の絶対値が最大であったものである。

また、「調和」と各因子得点との回帰分析の結果を表 4.5 に示す。ここでも、5%の水準で相関が有意であると認められた印象語対の箇所には色付けを行っており、「調和」と「評価性」因子に相関があることがわかる。

表 4.4 「調和」に対する印象語対の回帰分析

印象語	相関係数 R	有意 F	切片	回帰係数
明るい	0.2905	0.1684	0.4742	-0.2906
うれしい	0.3912	0.0587	0.2432	-0.2963
暖かい	0.0476	0.8253	0.1265	-0.0384
緊張した	0.3900	0.0596	-0.1344	-0.4705
澄んだ	0.3760	0.0702	0.4213	0.2599
大きい	0.0070	0.9743	0.3545	0.0039
強い	0.1999	0.3489	0.3365	-0.1216
重い	0.2162	0.3102	-0.3274	0.1819
動的な	0.5168	0.0097	0.4569	-0.5467
美しい	0.6562	0.0005	0.3722	0.4022
新鮮な	0.0017	0.9936	0.0917	-0.0009
潤い	0.3133	0.1361	0.2655	0.2101
自然な	0.5032	0.0122	-0.2809	0.3966
大胆な	0.5565	0.0047	0.2822	-0.4168
単純な	0.2771	0.1898	0.3649	0.1021
鋭い	0.2168	0.3088	0.1648	-0.1665
感動した	0.8672	0.0000	-0.1408	0.3752
気分	0.5116	0.0106	1.1846	-0.0129
体調	0.5116	0.0106	0.8096	-0.0129

表 4.5 「調和」に対する因子得点の回帰分析

主因子	相関係数	有意水準	切片	回帰係数
評価性	0.6168	0.0013	-0.4278	0.6019
明快性	0.3486	0.0950	0.2422	-0.3747
力量性	0.0428	0.8424	-0.0001	-0.0430
堅鋭性	0.2160	0.3108	0.2775	-0.2586
コンディション	0.3772	0.0692	-0.4248	-0.2593

「評価性」因子の中でも「感動」に関する印象語対が特に「調和」との相関が強いことは特筆すべき結果である。このことは、「評価性」因子

の説明変数の中でも、特に素材に対する評価的な因子を含んでいると推測される「感動」の印象語対が図 2.1 に示すようなモデルの最上位に位置する複雑な因子を含むものであり、同様に複雑な因子を持つと思われる「調和」と協動的な相互作用が生じていると考えられるのである。

4.4. 「感動」に関する協合現象

組み合わせ素材の「感動」の評点と「調和」の評点には特に強い相関が認められたという結果より、組み合わせ素材の「感動」の評点を、組み合わせ素材の「調和」の評点と音・音楽の「感動」の評点、ならびに、動画の「感動」の評点から推定する。重回帰分析により、全体としては、重相関係数 R は 0.972 で、有意水準 F は 0.0055 であった。回帰係数と p 値を表 4.5 に示す。分析全体の有意水準 F、各説明変量の p 値のいずれもが 5% の水準で有意であると認められる。

また、この重回帰分析の説明変量である組み合わせ素材の「調和」の評点を省き、組み合わせ素材の「感動」の評点を、音・音楽の「感動」の評点と動画の「感動」の評点だけで重回帰分析した結果を表 4.6 に示す。全体としては、重相関係数 R は 0.760 で、有意水準 F は 0.116 であった。分析全体の有意水準 F、各説明変量の p 値のいずれもが 5% の水準で有意であるとは認められない。

以上、得られた表 4.5 と表 4.6 の相違は、2.1 で述べたように、上位の処理レベルと考えられる「評価性」因子を含む「感動」の印象語は、組み合わせ素材の「調和」が高くなければ、「感動」の評点が共に高い音と映像を組み合わせても、「感動」の評点が高い組み合わせ素材が得られるとは限らない、ということを表していると考えられる。

表 4.5 組み合わせ素材の「感動」の評点を目的変量とした重回帰分析

説明変量	回帰係数	p 値
切片	-0.276	0.022
音:感動	0.470	0.013
動画:感動	0.314	0.022
調和	0.335	0.0066

表 4.6 組み合わせ素材の「感動」の評点を目的変量とした重回帰分析（「調和」抜き）

説明変量	回帰係数	p 値
切片	-0.100	0.577
音:感動	0.658	0.053
動画:感動	0.319	0.196

5. まとめ

本実験の結果より、まず、動画の再生速度を通常より速くすることで、一般的に素材への印象は「緊張した」「動的な」「鋭い」ものへ変化し、また、動画の再生速度を通常より遅くすることで、素材への印象が「単純な」「静的な」「鈍い」ものへ変化することがわかった。

次に、「調和」と「評価性」因子には正の相関があり、その「評価性」因子の中でも「感動」に関する印象語対が特に「調和」との相関が強いことがわかった。このことは、「評価性」因子の中でも素材に対する「評価」的な因子も含んでいると考えられる「感動」の印象語対は複雑な因子を含むものであり、同様に複雑な因子を持っていると思われる「調和」の印象語対と協動的な相互作用が生じていると推測される。

さらに、音と映像の組み合わせにおいては、「感動」の度合いは「調和」も大きく影響しているようである。

本研究の最終的な目的の一つは、『どのような組み合わせの素材が、動画の再生速度から「調和」に影響を受けるか』を検証することである。これを実現するには、今後、以下のような検討を行う必要があると考えられる。

- 本実験で得た「通常再生素材-3倍再生素材」間における調和の評点の差や、「通常再生素材-1/2再生素材」間における調和の評点の差に対し、それぞれ組み合わせに使った音と動画の評点との関係を調べる。
- 「調和」に関する印象語対を「意味的調和」と「時間的調和」に関する2つの印象語対に分けて評価を行う。
- 組み合わせ素材だけでなく、単独で表示した動画の通常、3倍、1/2倍の再生速度のものについても印象評定を行う。

また、「調和」の程度を様々に変えての、組み合わせ素材の「感動」における「調和」の関与に関係する研究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：18500163）による。

参考文献

- [1] 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、(2000.11)
- [2] 桶井 良幸、桶井 貞美：図解でわかる多変量解析、日本実業出版社、(2001.1)
- [3] 寶珍 輝尚、都司 達夫：印象に基づくマルチメディアデータの相互アクセス法、情報処理学会論文誌、Vol.43 NO.SIG2(TOD13)、pp.69-79 (2002.3)
- [4] VOLTAGE：バックの鬼 Digital Sound Clip Vol.1「響」、VOLTAGE、(1994)
- [5] (株)ウィネット：Music Wizard、(株)ウィネット、(1995)
- [6] (株)ウィネット：Music Wizard II、(株)ウィネット、(1996)
- [7] Moon pocket：ビデオ素材辞典、データクラフト、(2002)