

# 歌唱音楽の心象風景映像への前景エージェントが 音楽映像視聴体験に与える影響

## Imaginary Scenery Video with Emotional Expression Agent Based on Music Lyrics

西野 歩真<sup>1\*</sup> 石田 真子<sup>1</sup> 米澤 朋子<sup>1</sup>  
Ayuma Nishino<sup>1</sup> Mako Ishida<sup>1</sup> Tomoko Yonezawa<sup>1</sup>

<sup>1</sup> 関西大学  
<sup>1</sup> Kansai University

**Abstract:** 本研究は、歌唱音楽における楽曲の音の流れや歌詞文脈がもたらす感覚の相互作用を最大限に発揮することのできる音楽映像体験の実現とその自動生成手法の提案を目指す。これまでに、楽曲の歌詞内容に連携した風景映像を生成するシステムを作成し、具象的な風景映像がよりユーザーの興味を惹きつけ、没入感や感情の揺動感もたらす可能性などを確認した。本稿では、心象風景映像を背景とし、その上に前景となる登場人物エージェントを配置し、楽曲の歌詞内容に連携した感情表出を行うシステムを提案する。まずエージェント映像の歌詞感情連携の効果を検証した結果、連携によりユーザーの楽しさ、没入感、感情移入感、物語性の体感をより高めることが示唆された。そして背景と前景のエージェントを組み合わせ、表現におけるそれぞれの歌詞内容連携の効果を検証した結果、エージェントの表現の歌詞内容連携と背景映像の歌詞内容連携のそれぞれが、楽しさ、没入感、感情移入感、物語性の体感を高めることが示唆された。

### 1 はじめに

歌唱音楽はリズム、和音、旋律などの音楽要素と感情や情景を伝える歌詞、および歌唱音声の表現で構成されており、聴取者によって様々な聴き方で楽しまれている。例えば、抽象的に音の流れを聴いて楽しんだり、具象的に歌詞から感じられる情景を思い浮かべて楽しんだりなど、聴き手の楽しみ方の抽象度次第で変わる。このように音と歌詞の要素によってもたらされる様々な感覚が相互的に作用することにより1つの音楽体験となる。

この体験を通して、聴取者の心の中には風景（背景）と登場人物（前景）で構成される心象風景が作り出されると考えられる。この心象風景が背景と前景で構成されるという想定は、J-POPにおいて歌唱音楽の歌詞の持つ物語性に由来して、想起される情景の中に主役となる登場人物の行為や感情が描かれることが多く見られることから考案した。

本研究では、この心象風景を音楽と映像の調和の観点から再現することを狙って映像生成することにより、音と歌詞の相互作用が最大限に発揮される音楽映像視聴体験の実現とその自動生成手法の提案を目指す。音

楽と映像の調和には、音楽と映像の印象やメッセージ性などの一致を意味する意味的調和と、拍節アクセントや変化タイミングなどの一致を意味する時間的調和の2つがあり、両調和の存在が音楽映像の主観評価を高めることが示されている [13, 8]。これら調和の在り方に着目して心象風景映像の生成に取り組むことにより、歌唱音楽の音と歌詞の要素の相互作用がもたらす効果を最適・最大化することができると考えた。

これまで心象風景の背景となる映像に関しては、音楽と映像のもたらす相互作用や、意味的調和・時間的調和が心象風景映像に与える影響について検討を行い、歌唱音楽に付加する映像の抽象度の違いが集中して聴取する側面を変化させる可能性 [17] や、付加する映像が具象的である場合、抽象的な場合よりも視聴者の興味を惹きつけたり、没入感や感情の揺動感を与える可能性などを確認した [16]。

本稿では、これまで行った背景に関する検討内容を踏まえて歌唱音楽の歌詞内容に基づく風景映像を作成した上で、新規に前景となる登場人物エージェントを実装し統合した心象風景再現映像生成システムを作成し、検証した。この前景エージェントは歌詞感情に基づいて決定される内部状態から、顔表情や身体的なふるまいを変化させることで感情表現を行う仕組みとなっている。このような音楽メディアを介したエージェン

\*連絡先：関西大学

〒569-1095 大阪府高槻市霊仙寺町 2-1-1  
E-mail:k173603@kansai-u.ac.jp

トを組み合わせることで、背景映像のみでは伝えられない細かな歌詞感情の表現を可能にするとともに、前景と背景の相互作用を生み出し、ユーザが体験から受ける感情的効果を増幅することを狙う。

## 2 関連研究

### 2.1 音楽と映像の相互作用と調和

音楽映像は、主体となる音楽とそれに伴う映像で構成されており、これらは互いの印象に影響を及ぼし合う。例えば、明るい印象の映像が音楽の印象を明るくする [12] ことや、寂しい印象の映像が楽しい印象の音楽と組み合わせられると映像が楽しい印象となる [3] こと、また、音楽が感情的写真によって喚起される感情体験を顕著に高める [1] ことなどが確認されている。このように、音楽と映像は相互的に作用し合っており、効果的に組み合わせられることで作品としての質を高める [19]。

また、音楽と映像の相互作用の発揮を検討する上で重要な要素となるのが音と映像の調和である。音と映像の調和を図ることは作品のもたらす感動を大きくする要因となり、実際に映像コンテンツの評価が音と映像の調和度と高い相関関係にあることも示されている [13]。音楽と映像の調和感には意味的調和と時間的調和の2つの側面があるとされており、意味的調和感は、音楽と映像のムードや印象が一致していることで得られ、時間的調和感は、音楽と映像のアクセントや変化のタイミングが一致していることで得られる [8, 2]。音楽映像において意味的調和感が強く見られるとユーザの興味維持感や没入感、感情揺動感を高める可能性 [16] や、端的な図形描写を有する単純な音楽映像刺激においては時間的調和感が強く見られると没入感、楽しさ、高揚感、リズムやビートの体感、歌詞文脈への集中感を高める可能性 [17] などが著者の過去の研究によって確認されている。また、意味的調和に関して、実際の映像作品を用いた調和感調査の研究事例によると、オリジナルの音楽と映像の組み合わせが最も調和感を感じさせる結果となったことが示されている [20]。このことから、音楽映像作品の製作において意図的に、または無意識的に、最適な意味的調和感を生むような音楽と映像の組み合わせを高く評価する人間の感覚があると言える。そして、時間的調和に関して、ディズニーのアニメーションで多用されている「ミッキーマウジング」という技法を例に挙げる。この技法は音と映像のアクセントを同期させるようにアニメーションを作るものであり、映像が音楽構造の理解を助ける役割を担う場合があるとされている [13]。これら2つの調和はどちらか一方の存在に依存するのではなく、共に適

切なバランスで存在して効果を発揮することで音楽映像体験の評価向上に繋がると考えられる。

本研究では、上記のような実際の映像作品からも確認される2つの調和の表現手法と適切なバランスを考慮しながら歌唱音楽の心象風景再現映像の生成に取り組み。その手法として、意味的調和感を生み出すために歌詞の意味に連携した内容の風景映像に、歌詞の感情を表情やふるまいによって表現する登場人物エージェントを組み合わせる映像を生成するシステムを作成した。また、時間的調和感を生み出すために歌詞の意味に対応する背景映像を歌詞のフレーズごとに切り替えて表示した。このシステムにより、従来の音楽映像視聴体験よりも効果的に音楽と映像の相互作用を引き出すことを狙う。

### 2.2 歌詞感情と物語性

歌唱音楽の要素の一つとしての歌詞は人の感情や情景、意味を伝えるなどの役割を持ち、その意味や文脈の理解を伴うことで聴取者の感情に様々な影響を与えることから、音楽体験において重要な要素として捉えられている [15]。このように歌詞が感情移入における重要な要素とされているのは、音の要素に合わせて伝えられるテーマやメッセージ性、物語性を有しているからであると考えられる。物語には、全体を通して起承転結などの脈絡が存在し、物語の主演となる登場人物と背景となる場面により構成される。このような物語が歌詞から伝わることにより聴取者はそのストーリーに没入し、喜びや悲しみなどの感情を引き起こしていると考えられる。実際に、キャラクターに対してユーザが共感する効果を狙い、広告などに物語が利用されるケースが存在する [11]。

本研究では、上述するような物語に聴取者が感動したり共感する効果を歌唱音楽の心象風景映像生成に応用し、歌詞の物語性をユーザに感じさせることで、より没入感や感情移入感を感じられる音楽映像視聴体験の実現を目指す。

### 2.3 感情表出を行うエージェント

メディア体験におけるエージェントの存在は、ユーザとのインタラクションを通して体験の印象や効果を変化させる。その変化はエージェントの目的や仕様によって様々であるが、感情表出を行うエージェントの存在は特にユーザの心に影響を与え、情動的な変化をもたらす。例えば、音楽聴取シーンにおいてその楽曲を共聴しているように感じさせるロボットエージェントにより、ユーザの楽曲に対する好感度を高める効果が確認されている [5]。さらに、楽曲のリズムを通して

ユーザと音楽コミュニケーションを伴う共聴体験を実現する仮想エージェントを実装し、そのエージェントの表情表出により体験の楽しさ、興奮、共感が高まることを確認した研究事例も存在する [18]. また、人間の心では複数の感情が同時多発的に発生するとされ [4, 6], これを表現するエージェントの感情表現手法に関する検討も進められており、6 感情を複合的に表現可能なコミュニケーションエージェントの感情モデルが提案されるなどしている [14]. このように、音楽メディアを介したエージェントが存在することによりもたらされる単純な効果だけでなく、エージェントの音楽に基づいて変化する感情表出がユーザの情動に与える影響について検討を行うことが、さらなる音楽体験の追究に繋がると考えられる.

これに対して本研究では、歌唱音楽の心象風景映像の前景となる登場人物エージェントに、歌詞に基づいて感情表出を行う機構を実装し、ユーザが歌詞内容や文脈から受ける情動的効果を増幅できるようなメディア体験の実現を狙う.

## 2.4 音楽映像の自動生成

今日、より評価の高い視聴体験をもたらすための音楽映像の自動生成手法の検討やその基礎的な研究が、意味的類似（音楽から感じられる感情や印象と画像・動画コンテンツの内容の類似）と時間的構造の観点から進められている. 意味的類似の観点では、機械学習により感情の関連性に着目した音楽映像を自動生成するシステム [10] や、歌詞情報から作成された音楽の属性情報を用いて画像との意味的類似性を推定するフレームワーク [9] が提案されるなど、機械学習的なアプローチが見受けられる. また、時間的構造の観点では、音楽と映像の時間構造と繰り返しパターンに着目し、任意の映像から音楽映像を生成するシステムが提案されている [7].

これに対して本研究では、歌唱音楽から得られる情報を基に、各楽曲から想起されるイメージにより適合した音楽映像を自動生成するアルゴリズムの実現を目指す.

# 3 心象風景映像生成システム

## 3.1 システム概要

本研究では歌詞の情景を再現する風景映像を背景とし、前景に歌詞感情を表現する登場人物エージェントを組み合わせた心象風景映像生成システムを提案する. システムは音源再生部と背景映像生成部、前景エージェント生成部で構成される. 音源再生部は Pure Data

Vanilla<sup>1</sup>を用いて作成し、音源の再生や、音源の時間経過に合わせた描写変化のタイミング統制を行う. また、背景映像生成部と前景エージェント生成部による描写は Processing<sup>2</sup>を用いて作成し、Open Sound Control 通信<sup>3</sup>によって音源再生部と連携させた. 入力する音源ファイルは、ビットレート 16 [bit], サンプリングレート 48000 [Hz], チャンネルはステレオの wav ファイルとした. 出力する映像は解像度 720 × 720 [pixels], フレームレート 60 [fps] とした.

## 3.2 背景映像生成部

背景映像生成部によって生成される背景映像は歌詞のフレーズごとの内容を再現した画像を繋ぎ合わせて作られる. 歌詞内容再現画像は、画像生成 AI の DALL-E3<sup>4</sup>を用いて生成する. ChatGPT<sup>5</sup>を介して、フレーズごとに区切った歌詞をそれぞれ入力し、その文章から人物描写のない風景のみの画像を作成するように指示する. そのようにして作成された画像を歌詞の内容と連携した画像として、全てのフレーズ分準備し、システムで再生できるように格納する. 各画像は作成に使用した歌詞フレーズごとに切り替わっていく. また、背景と前景を合わせた心象風景映像の立体感を演出するために、生成した背景画像は時間経過とともに左へ流れていくように移動する仕様となっている. 生成された背景映像の様子を図 1 に示す. このように、歌詞内容に連携した風景を表出することで、聴取者がその楽曲を聴いて想起するイメージを映像として再現することを狙う.



背景描写例1

背景描写例2

図 1: 背景となる風景映像

<sup>1</sup><https://puredatajapan.info>

<sup>2</sup><https://processing.org>

<sup>3</sup><https://opensoundcontrol.stanford.edu/>

<sup>4</sup><https://openai.com/dall-e-3>

<sup>5</sup><https://chat.openai.com/>

### 3.3 前景エージェント生成部

前景エージェント生成部では、楽曲の歌詞フレーズごとに分析された感情値に基づいて登場人物エージェントの内部に想定される感情を決定し、それに基づく3段階の表情変化や連続的な身体のふるまいによって感情表現を行うエージェントを生成する。エージェントの各感情表現の様子を図2に示す。エージェントは映像画面の左下に足を抱えて座った状態を横から見た構図で配置される。各歌詞フレーズごとの感情値はGoogleのNatural Language API<sup>6</sup>を用いて算出する。Natural Language APIでは意味の区切りごとに分割した歌詞入力に対し、歌詞のポジティブ度をスコア化して算出する。スコアは-1.0（ネガティブ）から1.0（ポジティブ）の間の数値（0.1刻み）で算出され、これを歌詞感情値として扱う。エージェントはこの歌詞感情値に基づいて表情やふるまいを変化させる。具体的には、エージェントの身体は首・頭部と体部に分かれており首・頭部の角度を変化させることで、視線を上下するような動作を行う。首・頭部の角度は歌詞感情値が大きくなるほど斜め上を見上げる状態に近づき、小さくなるほど俯く状態に近づく。また、顔表情は歌詞感情値が1) 正のとき：ポジティブ表情、2) 0のとき：ニュートラル表情、3) 負のとき：ネガティブ表情の3段階で変化する。この仕組みにより、背景のみでは伝え切ることの出来ない詳細な歌詞感情を表現し、生成される心象風景映像を聴取者の想起するイメージにより近づけることを狙う。また、本エージェントが単なる静止画ではなく、歌唱音楽の流れと同じ時間の中に存在することを表現するために、定期的に瞬きを行う仕組みを実装した。

最後に、上述の前景の登場人物エージェントを背景映像に重ね合わせて表示するように実装し、歌唱音楽の心象風景映像生成システムを作成した。（図3）



図2: 前景エージェントの感情表出



心象風景映像 描写例1

心象風景映像 描写例2

図3: システムによる心象風景映像

## 4 実験

### 実験概要

本実験は歌唱音楽の歌詞内容に基づいた風景映像（背景）と歌詞感情に基づいて感情表現を行う登場人物エージェント（前景）で構成される心象風景映像が音楽映像視聴体験に与える影響を明らかにすることを目的とし、1) まず前景となる登場人物エージェントの感情表現が音楽映像視聴体験に与える影響に関する実験1を実施し、2) 次に背景と前景を重ねて統合した心象風景映像が音楽映像視聴体験に与える影響に関する実験2を実施した。実験は、20歳-26歳（平均年齢22.5歳、標準偏差4.7）の日本語を母国語とする26名（男性13名、女性13名）を対象に実施した。

### 4.1 実験1: 登場人物エージェントの感情表現が音楽映像視聴体験に与える影響

#### 実験仮説:

- H1: 登場人物エージェントの感情表現が歌詞感情に連携している方が音楽映像視聴体験の楽しさがより高まる
- H2: 登場人物エージェントの感情表現が歌詞感情に連携している方が、より音楽映像視聴体験に集中してのめりこみややすくなる
- H3: 登場人物エージェントの感情表現が歌詞感情に連携している方が、より音楽映像視聴体験への感情移入しやすくなる
- H4: 登場人物エージェントの感情表現が歌詞感情に連携している方が音楽映像視聴体験を通して歌詞の物語性をより強く感じる

#### 実験条件:

実験は、登場人物の感情表現に関する要因A（A1: 歌詞感情に連携した感情表現, A2: 歌詞感情に逆に連携した感情表現, A3: 感情表現無し）による1要因3水

<sup>6</sup><https://cloud.google.com/natural-language?hl=ja>

準の被験者内実験計画とした。実験条件の提示順についてはカウンタバランスを考慮しラテン法格法に基づいて決定した。

A1では、登場人物エージェントが歌詞フレーズごとにその部分のポジティブ度の数値に連携した感情表現を行う（正に連携）。A2では、登場人物エージェントが歌詞フレーズごとにその部分のポジティブ度の数値に逆に連携した感情表現を行う（正とは反対の連携であり、例えば数値がポジティブな場合にネガティブな表現を行う）。A3では登場人物エージェントが歌詞フレーズごとのポジティブ度に関わらず感情表現を行わない。これら3つの条件を比較することで、感情表現の有無や、歌詞に連携した感情表現が体験へもたらす影響の違いを調べることができると考えた。

### 実験手順：

実験ではノートPC（Mac Book Air, 14inch）とヘッドフォン（音質と付け心地に問題の感じられないもの）を用意し、Google Formsを使用して音を伴う映像を視聴させた後、評価させる形式で実施した。被験者は全3種の映像に対してそれぞれ評価を行った。実験を開始する前に、題材となる歌唱音楽の歌詞を提示して読み込ませ、歌詞の感情について考えさせた。提示した歌詞は意味の区切りごとに分割しており、各パートごとの感情をポジティブ・ニュートラル・ネガティブのどの感情に当てはまるかを回答させる形で感情の理解を促した。今回使用した音源は「榎原敬之 - 遠く遠く<sup>7</sup>」の0:00-2:08の部分である。

### 評価方法：

仮説H1に対するQ1, 仮説H2に対するQ2, 仮説H3に対するQ3, 仮説H4に対するQ4, そして登場人物の作画や動作の妥当性に関するQ5, Q6の評価項目を設定し、Visual Analog scale（以下、VAS）法に基づき評価項目を0-99（0：全くそう感じない、99：とてもそう感じる）で回答させた。

- Q1. 音楽映像体験は楽しかった
- Q2. 音楽映像体験に集中してのめりこんだ
- Q3. 音楽映像に感情移入した
- Q4. 楽曲の歌詞にストーリー（物語性）があると感じた
- Q5. 登場人物の体や顔表情の動きは自然だった
- Q6. 登場人物の感情表現は歌詞の内容に相応しかった

### 結果：

実験により得られた各評価結果に対して繰り返しのある分散分析（ANOVA）を有意水準  $\alpha = 0.05$  で実施した。各評価項目について平均値及び標準偏差を図4に、分散分析の結果を表1に示す。

分析の結果、Q1-Q3について、条件間に有意差が確認され全て  $A1 > A2 > A3$  となった。つまり、音楽映像視聴体験の楽しさ、没入感、感情移入感において、登場人物エージェントの感情表現が歌詞感情に連携するのが最も評価が高く、感情表現が無い条件よりも感情表現が歌詞感情に逆連携している条件の方が評価が高いことが示された。また、Q4-Q6においても条件間に有意差が確認され、全て  $A1 > A3$ ,  $A2 > A3$  となった。つまり、音楽映像視聴体験の物語性の体感、登場人物の動作の人間らしさ、登場人物の感情表現の歌詞内容との連携性において、登場人物エージェントの感情表現が歌詞感情に連携している条件および逆連携している条件で、感情表現のない条件よりも評価が高まること示された。

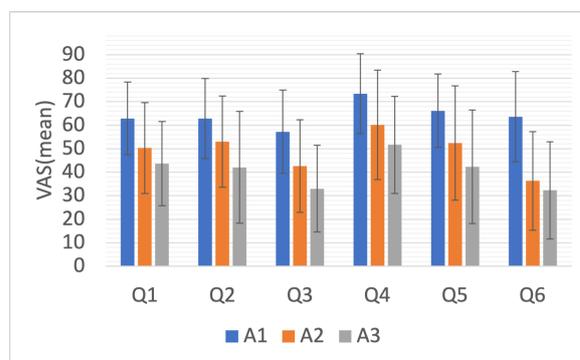


図 4: VAS における平均値と標準偏差 (実験 1)

表 1: VAS の分散分析結果 (実験 1)

	A		
	F	p	偏 $\eta^2$
Q1	29.441	<.001*	.541
Q2	19.482	<.001*	.438
Q3	25.388	<.001*	.504
Q4	16.443	<.001*	.397
Q5	12.335	<.001*	.330
Q6	22.144	<.001*	.470

\*:  $p < .05$

## 4.2 実験 2: 背景と前景を重ねて統合した心象風景映像が音楽映像視聴体験に与える影響

### 実験仮説：

H5：心象風景映像に重ね合わせた登場人物エージェントの感情表現が歌詞感情に連携していると、音楽映像体験の楽しさがより高まる

H6：心象風景映像に重ね合わせた登場人物エージェン

<sup>7</sup><https://youtu.be/KQNT6hbgf6g?si=IQP1hiCQyKkFifJs>

トの感情表現が歌詞感情に連携していると、より音楽映像体験に集中してのめりこみやすくなる

H7：心象風景映像に重ね合わせた登場人物エージェントの感情表現が歌詞感情に連携していると、より音楽映像体験に感情移入しやすくなる

H8：心象風景映像に重ね合わせた登場人物エージェントの感情表現が歌詞感情に連携していると、音楽映像体験を通して楽曲の物語性をより強く感じる

#### 実験条件：

実験は、前景となる登場人物エージェントの感情表現に関する要因 B（B1：歌詞感情に連携した感情表現、B2：歌詞感情に逆に連携した感情表現）と背景となる風景映像の内容に関する要因 C（C1：歌詞内容に連携した風景映像、C2：歌詞内容と無関係な風景映像）による 2 要因各 2 水準の被験者内実験計画とした。実験条件の提示順についてはカウンタバランスを考慮しラテン法格法に基づいて決定した。

B1, B2 は実験 1 における A1, A2 と同様の設定である。また、C1 では、歌詞の内容に基づいて生成された画像から作られた映像を視聴させ、C2 では、歌詞の内容とは無関係な画像を生成し、その画像から作られた映像を視聴させた。画像の変化タイミングは C1 と C2 で同じである。これら 2 要因 4 条件を比較し、歌詞内容に連携した風景映像と感情表現を行う登場人物エージェントが体験へもたらす相互作用の違いを調べることとした。

#### 実験手順：

被験者は実験 1 を終了した後、引き続き実験 2 に参加した。ノート PC とイヤホンを用意し、Google Forms を使用して音を伴う映像を視聴させた後、評価させた。被験者は全 4 種の映像に対してそれぞれ評価を行った。実験 2 で使用した音源は実験 1 と同様「榎原敬之 - 遠く遠く」の 0:00-2:08 の部分である。

#### 評価方法：

仮説 H5 に対する Q1, 仮説 H6 に対する Q2, 仮説 H7 に対する Q3, 仮説 H8 に対する Q4, そして背景の歌詞内容との連携性に関する Q5, 登場人物の感情表現の歌詞内容との連携性に関する Q6 の評価項目を設定し、VAS 法に基づき評価項目を 0-99 (0:全くそう感じない, 99:とてもそう感じる) で回答させた。

- Q1. 音楽映像体験は楽しかった
- Q2. 音楽映像体験に集中してのめりこんだ
- Q3. 音楽映像に感情移入した
- Q4. 楽曲の歌詞にストーリー (物語性) があると感じた
- Q5. 背景の映像は歌詞の内容に合っていた
- Q6. 登場人物の感情表現は歌詞の内容に相応しかった

#### 結果：

実験により得られた各評価結果に対して繰り返しのある分散分析 (ANOVA) を有意水準  $\alpha = 0.05$  で実施した。各評価項目について平均値及び標準偏差を図 5 に、分散分析の結果を表 2 に示す。

分析の結果、Q1-Q4, Q6 について、B1>B2 の有意差が確認された。つまり、音楽映像視聴体験の楽しさ、没入感、感情移入感、物語性の体感、登場人物の感情表現の歌詞内容との連携性の評価が、登場人物の感情表現が歌詞内容に逆連携するよりも正連携する方が高まることが示された。また、Q1-Q6 について、C1>C2 の有意差が確認された。つまり、音楽映像視聴体験の楽しさ、没入感、感情移入感、物語性の体感、背景映像の歌詞内容との連携性、登場人物の感情表現の歌詞内容との連携性の評価が、背景となる風景映像が歌詞内容と無関係であるより連携している方が高まること示された。そして、Q1 については、要因 B, C における条件間に有意な交互作用が確認され、交互作用グラフ (図 6) より、C1 水準において B1>B2, B1 水準において C1>C2, B2 水準において C1>C2 となった。C2 水準における B1 と B2 の間のみ有意差が認められなかった。つまり、音楽映像視聴体験の楽しさにおいて、風景映像が歌詞内容に連携している場合、登場人物の感情表現が歌詞感情に逆連携しているより、連携している方が評価が高まった。また、登場人物の感情表現が歌詞感情に連携している場合、および逆連携している場合においても、風景映像が歌詞内容に連携していないより、連携している方が評価が高まった。感情表現が逆連携の場合は、背景の連携の有無に差がなく、生じた違和感が背景の評価に影響したとも考えられる。

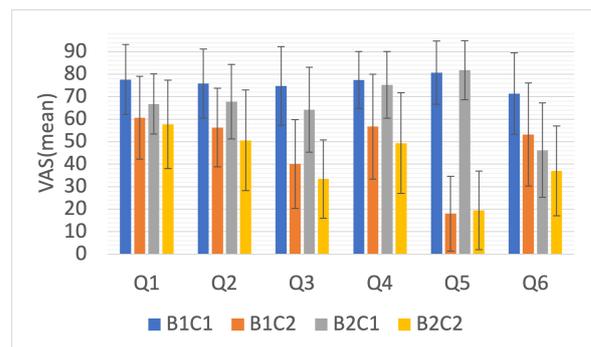


図 5: VAS における平均値と標準偏差 (実験 2)

表 2: VAS の分散分析結果 (実験 2)

	B			C			BC			交互作用の単純主効果
	F	p	偏 $\eta^2$	F	p	偏 $\eta^2$	F	p	偏 $\eta^2$	
Q1	9.581	.005*	.277	27.149	<.001*	.521	6.19	.020*	.198	B(c1),C(b1),C(b2)
Q2	9.775	.004*	.281	23.620	<.001*	.486	0.413	.526	.016	なし
Q3	17.750	<.001*	.415	84.206	<.001*	.771	1.118	.300	.043	なし
Q4	8.478	.007*	.253	54.312	<.001*	.685	1.419	.245	.054	なし
Q5	0.456	.506	.018	405.785	<.001*	.942	0.017	.898	.001	なし
Q6	29.33	<.001*	.54	16.143	<.001*	.392	4.136	.053	.142	なし

\*:  $p < .05$

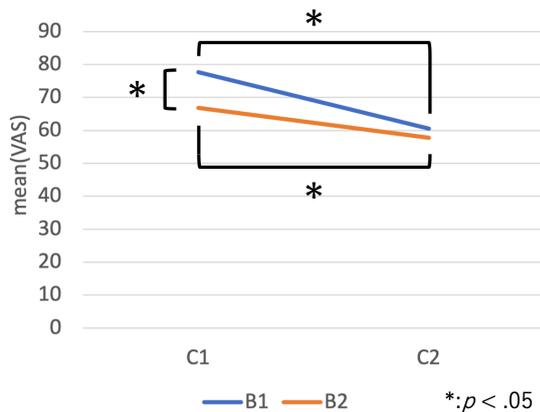


図 6: Q1 交互作用

## 5 考察

### 5.1 歌唱音楽に伴う登場人物エージェントによる影響 (実験 1 より)

実験 1 の評価項目 Q1-Q4 における結果より、仮説 1, 2, 3, 4 全てが支持された。

評価項目 Q1-Q3 において、登場人物エージェントの感情表現が歌詞感情に連携している条件、逆に連携している条件、そして感情表現のない条件の順に評価が高かったことから、歌唱音楽に伴う登場人物エージェントは歌詞感情に連携した感情表現を行うと、音楽映像視聴体験の楽しさ、没入感、感情移入感、物語性の体感をより高めることが示唆された。また、評価項目 Q4 においては、エージェントの感情表現が歌詞感情に連携している条件が、逆に連携している条件や感情表現のない条件よりも評価が高かったが、逆に連携している条件と感情表現のない条件の評価に有意な差は見られなかった。エージェントの感情表現の連携の正逆に関わらず感情表現がある方が、音楽映像視聴体験の楽しさ、没入感、感情移入感をより高めることが示唆された。

以上より、音楽映像視聴体験における楽しさ、没入感、感情移入感、物語性の体感が、歌唱音楽に伴う登場人物エージェントの歌詞感情に連携した感情表現によってより高まる可能性が示唆された。そして、評価項目 Q4 において、逆連携の条件と無表情の条件間に有意な差が見られなかったことから、物語性の体感は、エージェントが歌詞感情に連携した感情表現を行う場合に強く感じられる特有の感覚である可能性が示唆され、感情表現を行うエージェントによって変化する音楽映像視聴体験を検討する上で重要な要素であると考えられることから、歌詞感情の計算や文脈の分析手法ともにより深く検討していく必要があると考えられる。

また、評価項目 Q5, Q6 について、登場人物エージェントの感情表現が歌詞感情に逆に連携している条件や感情表現を行わない条件よりも、連携した条件の方が評価が高かったことから、本前景エージェントの歌詞感情に連携した感情表現がより人間らしく自然であり、歌詞感情との連携性が取れている可能性が示唆された。エージェントの作画やふるまいの妥当性については、個人の趣向などと併せてよりユーザの共通イメージに寄り添えるような描写手法を検討していく必要があると考えられる。

### 5.2 風景映像と登場人物エージェントによる心象風景映像 (実験 2 より)

実験 2 の評価項目 Q1-Q4 における結果より、仮説 5, 6, 7, 8 全てが支持された。

評価項目 Q1-Q4 について、登場人物エージェントの感情表現が歌詞感情に逆に連携している条件よりも、連携した条件の方が評価が高かったこと、そして、背景の風景映像が歌詞内容に無関係な条件よりも、歌詞内容に連携した内容である条件のほうが評価が高かったことから、歌詞内容に連携した背景の風景映像に、歌詞感情に連携した感情表現を行う登場人物エージェントを組み合わせると、音楽映像視聴体験の楽しさ、没入感、感情移入感、物語性の体感をより高めることが

示唆された。また、Q1において確認された交互作用より、水準 C2における要因 B の条件間以外の組み合わせでは、主効果で示された各条件の大小関係と同様の結果が得られた。このことから、歌詞感情に連携した感情表現を行う登場人物エージェントが高める音楽映像視聴体験の楽しさは、歌詞内容に連携した背景映像と組み合わせる場合において顕著に現れることが示唆され、これは前景と背景の両方が歌詞に連携していることによる相互作用が働くことに起因していると考えられる。

評価項目 Q5 について、要因 C においては背景の内容がランダムであるよりも歌詞内容に連携している方が評価を高める結果となったが、要因 B においては有意な差が確認されなかった。このことから、背景映像から感じられる歌唱音楽の情景としての印象は大きく、エージェントの感情表現の方法によって左右されにくいということが示唆された。本手法における背景映像が心象風景再現映像の印象を決定づける重要な要素であることを再確認し、今後も前景との関係性にも着目しながら深く検討すべきであると考えられる。

評価項目 Q6 について、歌詞感情に連携したエージェントの感情表現が歌詞内容に対してより相応しいと評価されたのは、背景が歌詞内容に連携している条件であったことから、やはり前景と背景がどちらも歌詞に連携していることで生じる相互作用が音楽映像視聴体験に良い影響をもたらしていることが考えられる。

一方で、今回の検証には複数の課題が残る。まず、登場人物エージェントの感情として、歌詞フレーズごとの感情値をそのまま反映したが、この数値は歌詞フレーズ前後の文脈を考慮しておらず、歌詞の物語性を十分に汲み取ることができていないと言える。これに関して、今後は歌詞全体の感情値や楽曲のテンポ、調性、および歌唱表現から読み取れる情報から文脈を分析する手法や、フレーズごとの意味や感情の流れを分析する手法について検討する必要がある。

また、今回のエージェントの表情はネガティブ表情、ニュートラル表情、ポジティブ表情の3段階で表現したが、今後、感情値に合わせた連続的な変化を行う仕組みを実装することで、より細かい感情描写を可能にすると共に、ユーザへの歌詞文脈の伝達にも活用することができると考えられる。

また、今回は単一的な心象風景映像生成について検討したが、歌唱音楽から想起される心象風景は各人によって多種多様であり、統一した共通イメージとして心象風景生成を行うのみでは音楽体験効果の最大化は難しいと考えられる。そのため、ユーザの年代や経験、趣向などの属性情報に基づいて心象風景生成の最適化を行う仕組みについて検討する必要がある。そして、本稿では歌唱音楽の歌詞情報を基に心象風景映像を生成する手法を提案したが、歌詞には言語としての意味伝

達の限界が存在し文化的側面による影響も少なくないと考えられる。そのため今後は、歌唱音楽のリズムや和音、旋律といった音楽の特徴、さらには歌唱音声の表現を心象風景映像に反映する手法について検討を行い、言語に囚われない音楽映像視聴体験の実現を目指す。

さらに、本手法は歌唱音楽から得られる事前情報を基に映像を生成し、出来上がったものをユーザに提供するという一方的なものであったが、今後は楽曲から生成される心象風景映像を視聴して移り変わるユーザの心境をリアルタイムで映像内容に反映し、心象風景映像やその視聴体験にさらなる変化をもたらすといった、ユーザとシステム間のインタラクション構造を組み込んだ新規手法の検討を行い、歌唱音楽の心象風景を通じた音楽映像体験の新たな価値の創出を目指す。

## 6 おわりに

本研究は歌唱音楽の心象風景の再現を狙った映像生成を行うことにより、楽曲の音の流れや歌詞文脈がもたらす感覚の相互作用を最大限に発揮することのできる音楽映像視聴体験の実現と、その自動生成手法の提案を行うことを目的とし、本稿では歌唱音楽の歌詞感情に基づいて感情表現を行う登場人物エージェント（前景）の検討、そして、その前景エージェントと背景となる歌唱音楽の歌詞内容に連携した風景映像の組み合わせによる心象風景映像生成システムの検討を行った。その結果、歌詞感情に連携した前景エージェントが音楽映像視聴体験の楽しさ、没入感、感情移入感、物語性の体感をより強める可能性が示唆された。また、歌詞内容に連携した背景となる風景映像と歌詞感情に連携した感情表現を行う前景の登場人物エージェントとの組み合わせが、音楽映像視聴体験の楽しさ、感情移入感、物語性の体感をより強める可能性が示唆された。今後は本検証における課題を解消すべく、歌詞の文脈を考慮した歌詞感情値の算出手法の検討や、言語や音楽ジャンルに囚われない心象風景映像の生成手法の検討を行い、さらなる音楽体験の価値を追究できるような心象風景自動生成システムの実現を目指す。

## 謝辞

本研究は、一部 JSPS 科研費 23K11278, 21K11968, 19K12090, 22K19792, 23K11202, および、2022 年度関西大学若手研究者育成経費の研究課題「エージェントを用いた共感的音楽体験共有の価値創造に関する研究」の助成を受け実施しその成果を公表するものである。

## 参考文献

- [1] Thomas Baumgartner, Michaela Esslen, and Lutz Jäncke. From emotion perception to emotion experience: Emotions evoked by pictures and classical music. *International journal of psychophysiology*, Vol. 60, No. 1, pp. 34–43, 2006.
- [2] Valerie J Bolivar, Annabel J Cohen, and John C Fentress. Semantic and formal congruency in music and motion pictures: Effects on the interpretation of visual action. *Psychomusicology: A Journal of Research in Music Cognition*, Vol. 13, No. 1-2, p. 28, 1994.
- [3] Annabel J Cohen. Associationism and musical soundtrack phenomena. *Contemporary music review*, Vol. 9, No. 1-2, pp. 163–178, 1993.
- [4] Mirjam Eladhari, Rik Nieuwdorp, and Mikael Fridenfalk. The soundtrack of your mind: mind music-adaptive audio for game characters. In *Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology*, pp. 54–es, 2006.
- [5] Guy Hoffman, Shira Bauman, and Keinan Vannun. Robotic experience companionship in music listening and video watching. *Personal and Ubiquitous Computing*, Vol. 20, No. 1, pp. 51–63, 2016.
- [6] Patrik N Juslin and John Sloboda. *Handbook of music and emotion: Theory, research, applications*. Oxford University Press, 2011.
- [7] Jen-Chun Lin, Wen-Li Wei, and Hsin-Min Wang. Automatic music video generation based on emotion-oriented pseudo song prediction and matching. In *Proceedings of the 24th ACM international conference on Multimedia*, pp. 372–376, 2016.
- [8] Scott D Lipscomb and Roger A Kendall. Perceptual judgement of the relationship between musical and visual components in film. *Psychomusicology: A Journal of Research in Music Cognition*, Vol. 13, No. 1-2, p. 60, 1994.
- [9] Xixuan Wu, Yu Qiao, Xiaogang Wang, and Xiaou Tang. Bridging music and image via cross-modal ranking analysis. *IEEE Transactions on Multimedia*, Vol. 18, No. 7, pp. 1305–1318, 2016.
- [10] Hong-Jiang ZHANG Xian-Sheng HUA, Lie LU. Automatic music video generation based on temporal pattern analysis. In *Proceedings of the 12th annual ACM international conference on Multimedia*, pp. 472–475, 2004.
- [11] 下村直樹. 広告における物語と感情: 自由回答による分析. 北海学園大学経営論集, 2011.
- [12] 岩宮眞一郎. 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, Vol. 48, No. 9, pp. 649–657, 1992.
- [13] 岩宮眞一郎. 音が映像作品の「でき」を決める. 日本音響学会誌, Vol. 64, No. 12, pp. 709–714, 2008.
- [14] 黒須亮成, 清水博貴, 橋本智己. コミュニケーションエージェントにおける感情モデルの提案. 知能と情報, Vol. 29, No. 1, pp. 501–506, 2017.
- [15] 森数馬, モリカズマ. 日常の音楽聴取における歌詞の役割についての研究. 対人社会心理学研究, Vol. 10, pp. 131–137, 2010.
- [16] 西野歩真, 石田真子, 米澤朋子. 歌唱音楽に付加する映像の表現様式と音楽との時間的調和の影響. Technical Report 14, 関西大学大学院総合情報学研究科, 関西大学大学院総合情報学研究科, 関西大学大学院総合情報学研究科/関西大学総合情報学部, jun 2023.
- [17] 西野歩真, 石田真子, 米澤朋子. 歌唱音楽の心象風景の視覚化によるユーザ音楽体験への影響. 研究報告コンピュータビジョンとイメージメディア (CVIM), Vol. 2023, No. 14, pp. 1–6, 2023.
- [18] 石田真子, 竹村響, 米澤朋子. リズム聴取で興奮・快感・内的時間を表すエージェントのもたらす共感性の検討. ヒューマンインタフェース学会論文誌, Vol. 25, No. 4, pp. 283–298, 2023.
- [19] 大串健吾, 桑野園子, 難波精一郎ほか. 音楽知覚認知ハンドブック-音楽の不思議の解明に挑む科学-. 北大路書房, 2020.
- [20] 劉沙紀, 矢萩徹, 大西英治, 岩宮眞一郎. 映像作品における音楽と映像の調和感と印象の連続測定. 音楽知覚認知研究, Vol. 21, No. 2, pp. 73–86, 2015.