

画像視覚特性に基づく旋律自動生成の高度化に関する一検討

高山 伸也^{1,a)} 明堂 絵美¹ 酒澤 茂之¹ 愛澤 伯友²

概要：音楽付写真スライドショーが広く使用されているが、独立して作成された写真と音楽を同時に再生するだけでなく、視覚と聴覚の感覚が適合する提示ができれば、魅力的なサービスとなる。また、視覚に合わせて音楽が自動生成できれば、手軽に音楽付写真スライドショーを楽しめる。筆者らは、写真データを画像解析して視覚効果を自動付与し、付与された視覚効果に対応する旋律を自動生成して提示する写真再生システムを構築した。しかし、同システムでは、単音のみの旋律生成に止まっており、音楽表現が乏しいという課題があった。そこで本報告では、音楽のジャンルを予め指定し、多楽器で構成される伴奏と共に高度な旋律を生成することで、より豊かな音楽表現を可能とした写真再生システムを提案する。また、提案システムにおいて、主観評価試験を実施したので報告する。

1. はじめに

近年、デジタルカメラやカメラ付携帯端末の利用により、気軽にかつ簡単に写真を撮影することが可能となっている。また一方で、デジタルフォトフレームや写真共有型 SNS の普及により、個人が撮影した写真に、好みの音楽を付加してスライドショーやフォトムービーを作成し、友人同士で視聴したり、共有したりする機会も増加している [1]。

音楽付のスライドショーやフォトムービーを作成する際、個人が撮影した写真群に対して、既成の楽曲を単に貼り付ける方法が主流であるが、楽曲の一定の進行に合わせて、撮影時系列で整列された写真群が再生されるため、写真の遷移と楽曲の拍節が噛み合わない、写真の内容と楽曲の抑揚が一致しない等、違和感のあるスライドショーになることが少なくない。また、そのような違和感を解消しようとする、膨大な手作業での修正が必要なのが現状である。そこで、独立して形成された写真と音楽を同時に再生するだけでなく、視覚と聴覚の感覚が適合する提示が自動的にできれば、魅力的なサービスとなる。

音響データから、音量を示す音圧レベルや、音高を示す基本周波数等の音響特徴量を抽出し、楽曲に含まれる音楽的な表現を解析できれば、それらの表現に適合する視覚的な効果を自動的に付与することができる。しかし、ポピュラーミュージック等の既成の音楽コンテンツの多くは、多くの楽器を同時に演奏しているため、音楽的な表現が多く

含まれる主旋律の音高やクレッシェンド等で示される強弱成分のみ取り出すことが困難である。また、既成の音楽コンテンツを使用する場合には、著作権問題が生じ、ユーザが自由に利用しにくい問題がある。そこで、既成の音楽コンテンツを使用するのではなく、写真に適合した音楽を自動的に生成できれば有用であると考え、とりわけ、写真に付与するエフェクト（視覚効果）と音楽のメロディ（旋律）に着目し、視覚効果と旋律が適合するスライドショーの生成を目指す。筆者らは、写真データを画像解析して、顔画像の有無、数、占有率等を判定し、判定された顔画像の属性に応じて、視覚効果を自動付与し、付与された視覚効果に対応する旋律を自動生成して提示する写真再生システムを構築した [2]。しかしながら、同システムでは、単音のみの旋律生成に止まっており、音楽表現が乏しいという課題があった。

そこで本報告では、音楽のジャンルを予めボサノヴァに指定し、多楽器で構成される伴奏と共に高度な旋律を生成することで、より豊かな音楽表現を可能とした写真再生システムを提案する。また、同システムに対し、主観評価試験を実施したので、報告する。

2. 関連研究

これまでに、視覚と聴覚の感覚が適合する写真の提示を目的とした研究が報告されている [3][4][5][6]。文献 [3] では、楽曲データから抽出される基本周波数（ピッチ）や音圧等の音響特徴量を用いて楽曲全体の印象を推定し、再生する写真を自動的に選択する手法を提案している。しかしながら、ポピュラーミュージック等の既成の音楽コンテンツ

¹ 株式会社 KDDI 研究所
KDDI R&D Laboratories, Inc.

² デジタルハリウッド大学
The University of Digital Content

a) sh-takayama@kddilabs.jp

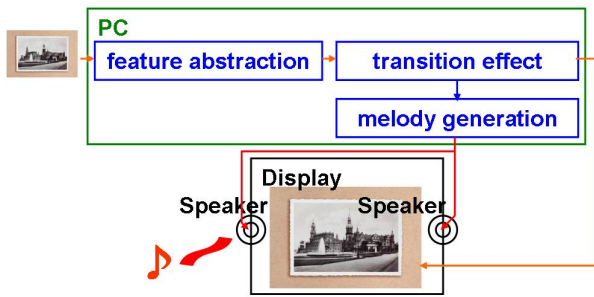


図1 本システムの構成

では、同一楽曲内から抽出される音響特徴量にかなり偏向があるため、抽出された各パラメータに応じて写真を自動選択すると、特定の写真のみが繰り返し再生されてしまう。文献[4]を用いれば、楽曲データから抽出されるビート(拍)情報を用いて楽曲における小節の先頭を自動判定し、再生する写真を遷移するタイミングを算出することが可能である。しかしながら、サビ等の旋律的に意味のある区間の始まりと自動判定された小節の先頭は必ずしも一致するとは限らないため、写真の遷移と楽曲の拍節が噛み合わない箇所が生じる場合が少なからずある。文献[5][6]では、楽曲データに付随する歌詞情報に基づいて、提示される写真が該当区間の歌詞の内容とできる限り近くなるように写真を配置し、スライドショーを自動的に生成するシステムを提案している。しかしながら、再生する写真の順序が強制的に変更されるため、写真のハイライトと楽曲の抑揚が一致しない等の課題がある。また、いずれの課題を解消するために、膨大な手作業での修正を必要なのが現状である。

そこで筆者らは、既成の音楽を用いるのではなく、写真に適合した音楽を自動的に生成することを目指し、表示される写真の視覚効果に対応する旋律を自動生成して提示する写真再生システムを構築した[2]。しかしながら、同システムでは、音楽表現が乏しいという課題があった。

そこで本報告では、ボサノヴァ風の伴奏と共に高度な旋律を生成することで、音楽表現の豊かな写真再生システムを提案する。また、主観評価試験を実施し、本システムの有効性について検証する。

3. 提案システムの構成

図1にシステムの構成を示す。まず、写真データを入力する。次に、入力された写真データを画像解析して、画像特徴量を抽出する。その後、算出された写真の輝度値に応じて和音列を決定し、該当区間に出力する伴奏を形成する。一方、入力された写真における顔画像の有無、顔画像の総数、顔画像の領域が写真全体に占める割合、各顔画像の傾き、各顔画像における視線の方向、各顔画像における笑顔度を同時に判定し、予め定義された条件と一致した場合は、その条件に対応する視覚効果を写真に自動的に付与する。

表1 画像特徴と視覚効果の対応表

条件式	視覚効果
$o < \theta_o, c \neq 0, \theta_{l_{min}} < l < \theta_{l_{max}}$	pinwheel
$q > 1, o < \theta_o, c = 0, l > \theta_{l_{max}}$	blast
$o > \theta_o, c = 0, \theta_{l_{min}} < l < \theta_{l_{max}}$	flashbulb* ¹
$o < \theta_o, c = 0, \vec{d} \in E, \theta_{l_{min}} < l < \theta_{l_{max}}$	turn
$q = 1, o < \theta_o, c = 0, \vec{d} \in U, \theta_{l_{min}} < l < \theta_{l_{max}}$	stretch up* ²
$q > 1, o < \theta_o, c = 0, \vec{d} \in U, \theta_{l_{min}} < l < \theta_{l_{max}}$	up* ³
$q = 1, o < \theta_o, c = 0, \vec{d} \in D, \theta_{l_{min}} < l < \theta_{l_{max}}$	stretch down* ⁴
$q > 1, o < \theta_o, c = 0, \vec{d} \in D, \theta_{l_{min}} < l < \theta_{l_{max}}$	down* ⁵
$q = 1, o < \theta_o, c = 0, l > \theta_{l_{max}}$	contract
$q = 1, o < \theta_o, c = 0, l < \theta_{l_{min}}$	relax* ⁶
$q > 1, o < \theta_o, c = 0, l < \theta_{l_{min}}$	fade out* ⁷

次に、付与された視覚効果に応じて、該当区間に出力する旋律を自動的に生成する。ここで、予め定められた時間が経過した場合は、入力データを次の写真に遷移する。最後に、効果が付与された写真と、生成された伴奏と旋律から成る音楽を、ディスプレイとスピーカから同時に提示する。なお、本システムでは、出力される音楽信号として、伴奏にはWAVを、旋律にはMIDIをそれぞれ使用する。

4. 画像特徴に基づく視覚効果自動付与

本提案手法では、上野らによって提案された手法[7]を用いて、入力された写真における顔画像の有無、属性を判定し、それらに対応する視覚効果を写真に付与する。ここでは、入力写真における顔画像の総数 q [個]、画像の領域が写真全体に占める割合(占有率) o [%]、各顔画像の傾き c [deg]、各顔画像における視線の方向 $\vec{d}$ 、各顔画像における笑顔度 l [%]を属性として用いる。例えば、入力された写真に顔画像が有り、その顔画像が、

$$o < \theta_o, c \neq 0, \theta_{l_{min}} < l < \theta_{l_{max}}$$

のように、傾いていると判定された場合は、[7]で適合度の高かった、 xy 平面で回転かつ拡大しながら写真が出現する「ピンウィール(pinwheel)」を視覚効果として付与する。ここで、 $\theta_o, \theta_{l_{min}}, \theta_{l_{max}}$ は各パラメータにおける閾値を示す。また、入力された写真に顔画像が有り、その顔画像が、

$$q > 1, o < \theta_o, c = 0, l > \theta_{l_{max}}$$

のように、複数で、且つ高い笑顔度だと判定された場合は、[7]で適合度の高かった、縦方向にストレッチして写真を表示する「爆破(blast)」を視覚効果として付与する。

*1 点滅しながら写真を表示する視覚効果

*2 x 軸方向にストレッチした後、 y 軸方向に緩やかに上昇しながら写真を表示する視覚効果

*3 y 軸方向に急上昇しながら写真を表示する視覚効果

*4 x 軸方向にストレッチした後、 y 軸方向に緩やかに下降しながら写真を表示する視覚効果

*5 y 軸方向に急下降しながら写真を表示する視覚効果

*6 拡大しながら写真を表示する視覚効果

*7 徐々に写真が消える視覚効果



図 2 視覚効果の付与例 (回転の場合)

表 2 画像の動きと音楽の変化との適合性

画像の動き	音楽の変化
上昇	上行旋律
下降	下行旋律
収縮	音量縮小
拡大	音量 拡大
爆発	破裂音
点滅	同音反復

また、入力された写真に顔画像が有り、その顔画像が、

$$q = 1, o < \theta_o, c = 0, l > \theta_{l_{max}}$$

のように、1 つで、且つ高い笑顔度だと判定された場合は、[7] で適合度の高かった、収縮しながら写真を表示する「収縮 (contract)」を視覚効果として付与する。このように、入力された写真における顔画像の属性が、表 1 で示す 11 条件の何れかであるかを判定し、右欄に示す視覚効果をそれぞれ付与する。なお、表 1 中の E は空間ベクトルにおいて垂直方向の成分が 0 の集合を、 U は空間ベクトルにおいて垂直方向の成分が正の集合を、 D は垂直方向の成分が負の集合をそれぞれ表している。ここで図 2 では、y 軸回転しながら写真を表示する「回転 (turn)」を視覚効果として付与する場合の例を示す。一方で、表 1 で示す 11 条件の何れにも該当しなかった場合は、11 種類の視覚効果の中からランダムに 1 つ選択し、視覚効果を付与する。

5. 視覚特性に基づく旋律自動生成

5.1 視覚効果と旋律との適合性

画像の移動方向と音楽の変化とが適合していると感じる感覚は、生理学上の根源的、本能的な問題ではなく、誕生以後の音楽経験や教育環境等によって大きく左右される [8][9]。一般的な日本の家庭環境、教育環境で育った人間は、画像や映像と音楽との感覚体験を通じて、画像におけるいくつかの動きと音楽の変化が適合した感覚であると捉えることを学習している [10][11]。特に、画像の動きと音楽の変化との適合性が顕著に見られるものを、表 2 に挙げる。これらの関連性は、海外を含む多くのアニメーション作品からの影響を強く受けている。

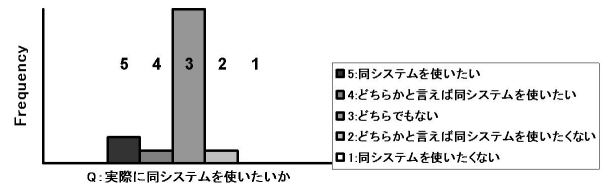


図 3 アンケート結果

表 3 アンケートにおける音楽経験別の回答傾向

図 3 中の回答番号	5	4	3	2	1
経験が多い人 (%)	25	12.5	62.5	0	0
経験が少ない人 (%)	0	0	87.5	12.5	0

表 4 本システムで使用した和音列

印象	和音列 (コード表記)
明るい	Dm9, G13, C9, C9
中立的	Dm9, Db9, CM9, CM9
暗い	Dm7(b5), Db9(b5), CM9, CM9

これらの知見をもとに、表 1 で示す視覚効果と単純な音型 (特定のルールに従った音の並び) との適合度を測定する予備実験を経て、各視覚効果と適合度の高い音型を生成するシステムを構築した [2]。

しかしながら、具体的な音楽スタイルを持たなかったために一般的な音楽性を持たず、図 3 で示す [2] で実施したアンケート結果が表すように、同システムを実際に使いたいと答えた被験者は少なく、表 3 が示すように、とりわけ音楽経験が比較的少ないに被験者に対して、満足のいく音楽を生成することができていないという結論を得た。そこで、今回は、音楽性を理解しやすいスタイルを持つ楽曲を生成することを目的とし、ボサノヴァを採用する。ボサノヴァとは、ブラジルの伝統的な大衆音楽で、伝統的なサンバを基礎として、新しい和音のパターンやジャズのコードから成る旋律を多く持つのが特徴である。

5.2 ボサノヴァ風伴奏の形成

本提案手法では、ボサノヴァの和音的特徴である和音の根音に対して、第 9 音 (根音より 9 度上の音)、第 11 音、第 13 音等の音を積み重ねた和音を使用し、伴奏における和音進行グループ (積重和音列) を作成する。表 4 で示すように、明るい印象を持つ積重和音列、中立的な印象を持つ積重和音列、暗い印象を持つ積重和音列のテーブルを予め作成しておき、入力された写真の輝度が高いと判定された場合は、明るい印象の積重和音列をベースとして伴奏を形成する。一方、入力された写真の輝度が低いと判定された場合は、暗い印象の積重和音列をベースとして伴奏を形成する。また、輝度が高いとも低いとも判定されなかった場合は、中立的な印象の積重和音列をベースとして伴奏を形成する。さらに、ギターを中心に、ベース、ピアノ、ドラムと、ボサノヴァを編成する楽器を伴奏に使用し、一般的に聞きやすい音楽表現を持った楽曲の生成を目指す。

表 5 視覚効果と音型の対応表

効果 ID	視覚効果	主となる音型
1	pinwheel	強調
2	blast	破裂
3	flashbulb	連続
4	turn	繰返
5	stretch up	上昇 (緩やか)
6	up	上昇 (急激)
7	stretch down	下降 (緩やか)
8	down	下降 (急激)
9	contract	高音
10	relax	低音
11	fade out	終了



図 4 順次進行のみで生成した旋律の例



図 5 提案手法に基づいて生成した旋律の例

5.3 ボサノヴァ風旋律の生成

本提案手法では、第 4 章で述べた方法によって付与された各視覚効果に応じて、表 5 に示すように、「上行」「同音反復」等の主となる音型を決定し、ボサノヴァ風の旋律を生成する。[2] では、図 4 の例で示すように、主となる音型に従って連続した音階の動き (順次進行) のみで旋律を生成した。しかし、実際の音楽では、順次進行と分散和音を混ぜて旋律を形成するのが通例であるため、本提案手法では、図 5 の例で示すように、順次進行と分散和音を混ぜ合わせると共に、主となる音型とは異なる別の要素を一部に加えて、より豊かな音楽表現を持った旋律の生成を目指す。また、ボサノヴァは拍節を移動させる「シンコペーション」が特徴的であるため、本提案手法では、図 5 のように、意図的にタイ (連続する 2 つの同じ高さの音を繋げること) を含ませることにより、シンコペーションを実現している。

6. 評価試験

本提案システムに対して、単一の視覚効果及び旋律に対して行う個別評価と、3 分程度のスライドショーコンテンツに対して行う総合評価の 2 種類の主観評価試験を実施した。被験者は 10 代後半から 20 代前半の学生で、女性 9 名、男性 28 名、合計 37 名から回答を得た。なお、一般的なボサノヴァ楽曲における特性を考慮し、本評価試験においては、1 枚の写真に割り当てられる時間を一律 6.4 秒に設定すると共に、伴奏にはギター、ベース、ピアノ、ドラムを、旋律にはピアノのみを使用した。また、本評価試験には、大型スクリーンと出力の大きいスピーカを使用し、被験者全員一斉に実施した。

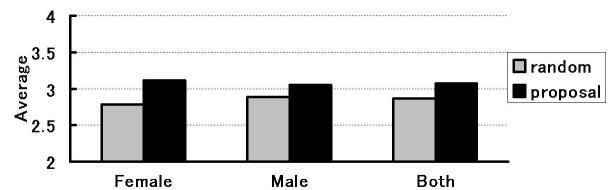


図 6 個別評価の平均値

表 6 各視覚効果における平均値

効果 ID	1	2	3	4	5
ランダム	2.46	2.72	2.68	2.98	3.43
提案手法	2.67	2.14	2.86	3.41	3.46
効果 ID	6	7	8	9	10
ランダム	2.75	3.40	2.72	2.97	2.69
提案手法	3.57	2.84	3.49	3.50	3.22

6.1 個別評価

個別評価試験では、提案手法に基づき付与された視覚効果とは異なるものをランダムに選択し、その視覚効果に応じて生成された旋律 (ランダム)、提案手法に基づき付与された視覚効果に応じて生成された旋律 (提案手法) の 2 種類を連続して不規則に配置し、スライドショーコンテンツを製作した。提示された視覚効果は音楽と適合していると感じますかという問いに対し、1:適合していない、2:どちらかと言えば適合していない、3:どちらとも言えない、4:どちらかと言えば適合している、5:適合しているの 5 段階で回答する試験を実施した。被験者は視覚効果及び旋律が付与された写真を 1 枚視聴する度に、評価値 1~5 を回答した。なお、本評価試験の総提示回数は 1442 回であった。

個別評価試験における評価値の平均を図 6 に示す。図 6 が表すように、ランダムに比べて提案手法の方が評価値が高いことが分かる。ランダム対提案手法で T 検定を実施したところ、女性で 1.9 %、男性で 4.4 % の有意水準で有意差が認められ、とりわけ女性の方が、無作為に選択された旋律に比べ、本提案手法に基づいて生成された旋律を好むことが示された。

各視覚効果における評価値の平均を表 6 に示す。なお、表 6 における視覚 ID は、表 5 におけるそれと対応している。表 6 が表すように、提案手法は、ターン (turn)、アップ (up)、ダウン (down) において、ランダムとの差が顕著に現れており、比較的動きの速い視覚効果に対して、有効な手法であると言える。また、[2] で実施した主観評価試験と比べると、とりわけ拡大 (contract) や収縮 (relax) において、ランダムとの差が開いている。これは、[2] では主となる音型のみから成る旋律を生成したため、例えば、収縮の場合を考えると、図 7 で示すような高音域の限定的な音しか持たない旋律であったものが、本提案手法では、主となる音型に加えてそれとは異なる別の要素も加えるにより、図 8 で示すような下降等、高音以外の音型も含む豊かな表現を持つ旋律が生成されたためだと考えられる。



図 7 [2] において生成した旋律の例 (収縮)



図 8 提案手法において生成した旋律の例 (収縮)

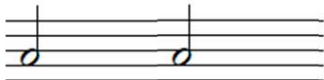


図 9 [2] において生成した旋律の例 (爆破)



図 10 提案手法において生成した旋律の例 (爆破)

一方、爆破 (blast) は、[2] で実施した主観評価試験と比べると、ランダムとの関係が逆転してしまった。予備実験の結果に基づいて、図 9 で示すような主に長音のみから成る旋律を生成したが、ボサノヴァでは一般的に長音を連続して使用しないことを考慮し、本提案手法では、長音の代わりに、破裂音として図 10 で示すような重音を取り入れた旋律を生成した。しかし、爆破の視覚効果である縦方向のストレッチは比較的速い動きではあるが、それでも 0.5 秒程度かかっており、約 0.2 秒の極めて短い時間で表現される破裂音とは時間的に適合しなかったため、結果的として提案手法における爆破の評価値が低下したと考えられる。

6.2 総合評価

総合評価試験では、提案手法のみを用いて 3 分程度のスライドショーコンテンツを製作した。提示された視覚効果は音楽と適合していると感じますかという問いに対し、1～5 の 5 段階で回答する MOS 評価を実施した。加えて、

- Q1: スライドショーの BGM として適切か
- Q2: 実際に本システムを使いたい
- Q3: 被験者の音楽経験

を問うアンケートを実施し、5 段階で回答を得た。なお、全ての回答は、コンテンツを最後まで視聴した後に行った。

総合評価試験における MOS 評価の結果を表 7 に示す。図 7 に示すように、MOS 値は 4 近くに達し、どちらかと言えば視覚効果と音楽とが適合していることが示されたものの、[2] で生成した順次進行から成る旋律のみの音楽に比べると、MOS 値が低下した。これは、表 5 に挙げた視覚効果と適合度の高い音型に加え、提案手法では、音楽の表現を豊かにするために該当音型とは異なる要素を加えたほか、5.2 節で述べたように、多楽器で構成される伴奏を同時に出力したことにより、視覚効果と適合度の高い、主となる音型を被験者が感じ難くなったためだと考えられる。

表 7 MOS 評価の結果

	女性	男性	総合
順次進行の旋律のみ	4.17	4.20	4.19
提案手法	3.56	3.93	3.84

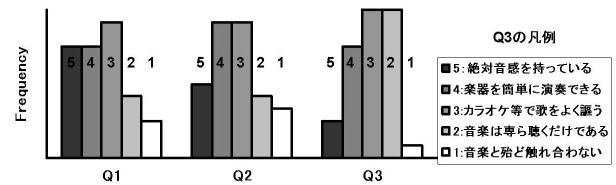


図 11 アンケートの結果

表 8 音楽経験が少ない被験者における Q2 の回答傾向

Q2 の回答	5	4	3	2	1
割合 (%)	8.3	37.5	25	16.7	12.5

また、総合評価試験におけるアンケートの結果を図 11 に示す。図 3 に示す順次進行から成る旋律のみの音楽を生成するシステムに対するアンケートに比べると、Q2 の問いに対して、5 または 4 と回答した被験者の割合が増加しており、豊かな表現を持つ高度な音楽を生成することで、実際のサービスに近づいたと言える。表 8 に示すように、とりわけ Q3 で 2 または 1 の何れかと回答した被験者の半数近くが、Q2 で 5 または 4 を選択しており、音楽経験が比較的少ないにユーザに対しても、満足のいく音楽を生成する手法に近づくことができたと考えられる。

しかし、一方で、Q2 の問いに対して、2 または 1 と回答した被験者の割合も増加している。これは、今回の生成する音楽のジャンルをボサノヴァに限定してしまったが故に生じた事象であることが想像され、今後は、ユーザが好む音楽のジャンルは人によって様々であることを考慮し、多くのジャンルの音楽を生成する仕組みが必要である。

7. まとめと今後の課題

本報告では、写真データを画像解析して視覚効果を自動付与し、付与された視覚効果に対応する旋律を自動生成して提示する写真再生システムにおいて、音楽のジャンルを予めボサノヴァに指定し、多楽器で構成される伴奏と共に高度な旋律を生成することで、より豊かな表現を持つ音楽の生成を実現した。同システムに対して主観評価試験を実施し、本提案手法に基づいて生成された旋律は、ランダムに選択されて生成された旋律に比べ、有意に優れることを確認した。また、本システムにおける視覚効果と音楽との適合度を評価した MOS 値は 3.84 であり、やや高い水準で両者が適合度されていることが示された。さらに、アンケートでは、実際に本システムを使いたい被験者の割合が増加し、豊かな表現を持つ高度な音楽を生成することで、実際

のサービスに近づけることができた。しかし、一方で、生成する音楽のジャンルをボサノヴァに限定してしまったが故に、生成された音楽を好まない被験者の割合も増加したため、ユーザが好む音楽のジャンルは人によって様々であることを考慮し、今後は、多くのジャンルの音楽を生成する仕組みを検討していく必要がある。

謝辞 本報告で実施した主観評価試験では、デジタルハリウッド大学に在籍する多くの学生にご協力を戴いた。この場を借りて同大学に深く深謝する。

参考文献

- [1] 総務省：平成 24 年版情報通信白書，ぎょうせい，pp.222-251(2012)。
- [2] 高山伸也，明堂絵美，酒澤茂之，愛澤伯友：写真スライドショー向け画像視覚特性に基づく旋律自動生成，映像情報メディア学会 2012 年年次大会，9(8)，pp.16-17(2012)。
- [3] Cai, R. Zhang, L. Jing, F. Lai, W. and Ma, W.-Y: Automated Music Video Generation Using Web Image Resource, Proceedings of IEEE International Conference on Acoustic, Speech and Signal Processing, pp.737-740 (2007)。
- [4] 宮島靖：Music Mosaic Generator:高精度時系列メタデータを利用した音楽リミックスシステム，第 15 回インタラクティブシステムとソフトウェアに関するワークショップ (WISS2007)，pp.13-18(2007)。
- [5] 舟澤慎太郎ら：歌詞情報を利用した Web 画像・楽曲連動スライドショー自動生成システム，情報処理学会研究報告，MUS，vol.84(13)，pp.1-6(2010)。
- [6] 青木圭子，石先広海，米山暁夫，小野智弘：楽曲スライドショーオンザフライシステムの構築，映像情報メディア学会 2011 年冬季大会，7(9)，pp.10(2012)。
- [7] 上野智史，明堂絵美，酒澤茂之，愛澤伯友：画像スライドショーにおける遷移効果決定のための方向印象推定の一検討，映像情報メディア学会 2011 年冬季大会，11(5)，pp.6(2011)。
- [8] 曽根敏夫：くらしと音，裳華房，pp.57(1991)。
- [9] 大橋力：音と文明，岩波書店，pp.167(2003)。
- [10] 岩宮眞一郎，大里俊晴，酒井博之，佐藤公信，清水富弘，城生佰太郎，田中宗隆，田中優子，彦坂裕，渡辺裕：音の百科事典，丸善，pp.70-650(2006)。
- [11] 岩宮眞一郎：音響サイエンスシリーズ 5 サイン音の科学，コロナ社，pp.138-163(2012)。