Vol.2013-MUS-98 No.10
Vol.2013-EC-27 No.10
2013/3/15

レコむし：画像と楽曲の印象の一致による楽曲推薦システム

佐々木 将人[†1,a)]　平井 辰典[†1]　大矢 隼士[†1]　森島 繁生[†1]

概要：入力した画像に対して感性的にマッチした楽曲を推薦するシステム，**レコむし** (**RECO**mmendation of **MUS**ic using an input **I**mage) を提案する．音楽を楽しむ上で，現在の情景は重要な要素の一つである．なぜなら，楽曲の印象がその情景と調和しているほど，楽曲を聴いたときの感動は増すためである．しかし，膨大な楽曲群の中から現在の情景に的確にマッチした楽曲を手動で探し出すことは容易ではない．そこで本研究では，AV(Arousal-Valence) 空間と呼ばれる心理空間に画像と楽曲を配置することで，情景に対する印象と楽曲に対する印象を対応付ける．**レコむし**はこの対応付けを用いることで，現在の情景に合った印象を与える楽曲をプレイリストとして推薦する．また，ランダム選曲による楽曲と比較することで，**レコむし**の評価を行った．

1. はじめに

1.1 情景に合った楽曲

音楽は，聴くという単純な動作だけで人に何らかの印象を与えることができる．さらに，その印象にはいくつもの種類があることが知られている [1]．また，聴覚に与えられる楽曲の情報と視覚に与えられる情景の情報が組み合わさることで，印象がさらに強まるという知見も示されている [2][3]．例えば，映画やドラマなどでは，BGM の挿入によってシーンへの印象が強調されている．しかし，無作為に音楽と情景を組み合わせればその印象が強調されるわけではなく，情景に調和した音楽を付加する必要がある [4]．

ここで，情景と音楽の調和について考える．情景と音楽の調和は，両者の雰囲気の一致による調和と，両者のシンボリックな意味の一致による調和の二種類に分けられる [5]．ここで，後者における情景や音楽のシンボリックな意味理解を行うためには，人手によるラベリングを行う必要がある．しかし，現在世の中の全ての楽曲に対して事前にラベリングを施すことは現実的ではない．よって，本研究では後者を対象外とし，前者の雰囲気の一致に基づく調和に注目する．本研究では，ユーザの聴きたい楽曲は現在の情景に合った楽曲であると仮定し，視覚と聴覚の調和による楽曲鑑賞を音楽の新たな楽しみ方として提案する．

[†1] 現在，早稲田大学 / JST CREST
Presently with Waseda Uniersity / JST CREST
[a)] joudanjanai-ss@akane.waseda.jp

1.2 本研究の必要性

近年の音楽コンテンツのデジタル化及び，メモリの大容量化に伴う携帯音楽プレイヤーの小型化により，個人の所有する楽曲数は増加し続けている．また，月額制の音楽聴き放題サービスなどもあり，インターネットを通じて多くの楽曲と触れ合うことができる．しかし，ユーザがそのような膨大な楽曲群の中から現在の情景に合った楽曲を探し出すことは困難である．なぜなら，現在の情景を具体的に表現することが困難なために検索ワード等を入力することが出来ないからである．その結果，ユーザは最近の試聴経験や音楽知識から情景に合った音楽を選曲するしかない．いくら携帯音楽プレイヤーが大容量化しても，普段聴く楽曲は頭に浮かぶ一部の楽曲からとなってしまう．これでは，メモリの大容量化を生かしきれていない上に，情景に合った音楽を自由に楽しむことは難しいといえる．

これは日本の音楽文化における，有名なアーティストの楽曲は出会いやすく評価されやすいという問題に深く関連している．例として 2012 年の年間オリコンチャートでは，1 位から 5 位までを同一のグループの楽曲が独占している．ユーザがすぐに思い浮かべることが出来る楽曲の大半が同一のグループのものだとすると，例えば物悲しい気分にあった楽曲を聴きたい場合であっても，実際に聴く楽曲の比重は同一のグループの楽曲の割合が高くなる．音楽プレイヤーに多くのジャンルの楽曲が入っていたとしても，ユーザの選択肢には偏りが生じてしまうのである．

この問題点に対し，我々はどのようなユーザでも現在の雰囲気に合う楽曲を自由に聴くことができるべきだと考える．そのために必要なことは，楽曲の推薦の基準が平等で

Vol.2013-MUS-98 No.10
Vol.2013-EC-27 No.10
2013/3/15

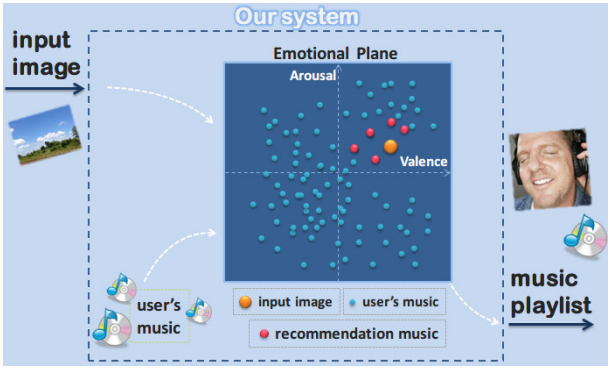

図 1 システムの概要

あるシステムの構築である．所持している情報量の差にとらわれない推薦を行うためには，言語情報を用いずに楽曲の特徴を表現する必要がある．それは，有名なアーティストの楽曲と無名なアーティストの楽曲との大きな差は，曲名やアーティスト名などの認知されている言語情報の差のみで，楽曲の内容や音響的な特徴に大きな差があるとは考えにくいからである．そこで，ユーザの欲求を言語情報以外の形で構築し，楽曲の音響特徴と比較する枠組みが可能であれば，どのようなアーティストの楽曲であろうと平等に評価される．このような新たな音楽との出会いを提供するシステムの構築が，本研究の最終的な目標である．

本研究では，AV 空間と呼ばれる心理空間に画像と楽曲を配置することで，情景に対する感じ方と楽曲に対する感じ方の対応を取る．そこに，聴きたい楽曲の印象をもつ情景を 1 枚の静止画として入力することで，現在の情景に合った楽曲推薦を行う．本稿では入力は静止画 1 枚であるが，パラメータ選択により動画像への拡張や，別メディアへの置換も，AV 空間を介することで実現可能となる．本研究のポイントは，楽曲を推薦する過程で，言語情報に落とし込む必要がないという点である．今回，新たな音楽推薦の形として画像情報を媒介とする音楽推薦を提案する．

2. 関連研究

2.1 従来の楽曲推薦手法

楽曲推薦には従来から多くの手法が提案されている．例えば，協調フィルタリングを用いた推薦手法では，他ユーザの楽曲評価を参考に楽曲の推薦を行う [6]．この手法は，有名なアーティストほど推薦されやすいので，社会の流行にのる上では大きな効果を持つ．しかし，推薦されるアーティストのバラエティが乏しいといった楽曲推薦上の問題がある．また，ユーザが好む楽曲と類似した音響特徴を持つ楽曲を推薦するコンテンツベース推薦手法もある [7]．しかし，精度が低いといった楽曲推薦上の問題がある．さらにいずれの手法でも，ユーザの置かれている状況に関わらず同じ基準で楽曲の推薦が行われてしまう．そのため，現在の状況に最適な選曲は困難である．

一方，ユーザの置かれた状況をアノテーション情報に落とし込み，歌詞や音楽要素と比較することによる楽曲推薦手法もある [8]．しかし，夕日や海岸といったアノテーション情報は検索クエリとして入力できるので，既存の楽曲検索システムを活用すれば推薦が可能といえる．さらに，歌詞のない楽曲への対応は困難であり，アノテーションにより記述できる情報も限られているため，現在の情景を正確には考慮できない．

2.2 AV 空間

本節では，情景に対する感じ方と楽曲に対する感じ方の対応を取るために，心理空間である AV 空間について記述する．Russell らは，人の感性を表現する空間として AV 空間を提案した [9]．AV 空間は，縦軸である Arousal 軸 (energeticcalm) と横軸である Valence 軸 (positive-negative) から成る二次平面である．Arousal は感情の興奮の度合いを表すために用いられる値である．また，Valence は感情のポジティブ-ネガティブの度合いを表す値である．この AV 空間上の座標である AV 値は人の感性や感情を表す．AV 空間の中心から遠いほど活発，穏やか，ポジティブ，ネガティブの度合いが強くなる．一方で，中心座標に近いほど，どの印象が特化しているとも言えないあいまいな印象となる．

また，Russell は AV 空間内と感情語の対応関係を示した．この AV 空間と，簡易的な感情語の分布を図 2 に示す．本研究では，AV 空間へ射影により画像，音楽の印象を定量的に解析する．これにより，情景に対する感じ方と楽曲に対する感じ方の対応を取り情景に合った楽曲の推薦を実現する．

3. 提案手法

本研究では，同一の心理空間に画像と楽曲を配置することで，両者の印象を対応させる手法を提案する．本章では，主に AV 空間への画像・楽曲の配置方法についての説明する．

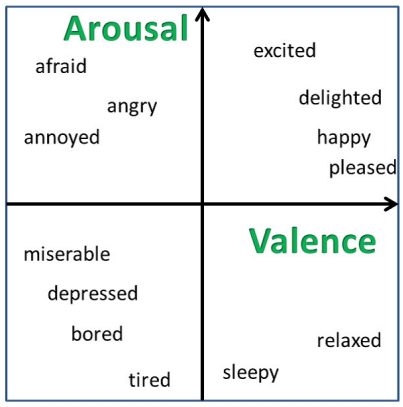

図 2 AV 空間

Vol.2013-MUS-98 No.10
Vol.2013-EC-27 No.10
2013/3/15

3.1 画像の配置

本節では，AV 空間への画像の配置について説明する．AV 空間に画像を配置するためには，画像を見たときに，人が感じる情報を体系化する必要がある．

3.1.1 色特徴量

Valdez らは，画像の特徴量である彩度と輝度が，Valence,Arousal との間に直接的な関係を持つことを実験により示した [10]．実験は，250 人の被験者にカラーパッチを見せ，その色に対してどのような感情を抱いたかを評価したものである．この実験で得られた輝度 Y，彩度 S と Valence V_1，Arousal A_1 との関係を式 (1)，(2) に示す．この関係を利用することで，情景の色味に対する印象を AV 空間へと反映できる．情景の色味を反映した AV 値の例を図 3 に示す．図 3 より，画像の色味により心理値に差が出ていることが確認できる．

$$V_1 = 0.69Y + 0.22S \quad (1)$$

$$A_1 = -0.31Y + 0.60S \quad (2)$$

3.1.2 形状特徴量

Jana らは画像のテクスチャ特徴量である Tamura 特徴量 [11] と Valence,Arousal との間に関連があることを示した [12]．Tamura 特徴量とは，テクスチャの方向性 (Direction)，粗さ (Coarseness) を表す画像特徴量である．この関係を用いるため，先行研究と同じ画像セットである IAPS Data set[13] を評価用のデータセットとして利用した．IAPS Data set は，心理学的知見に沿った膨大な評価実験により AV 空間へ配置された画像のセットである．正解 AV 値と Tamura 特徴量との正準相関分析により，式 (3)，(4) に示す Valence V_2，Arousal A_2 と Direction D，Coarseness C の間の関係式を得た．情景の形状による印象を反映した AV 値の例を図 4 に示す．図 4 より，画像の形状により心理値に差が出ていることが確認できる．

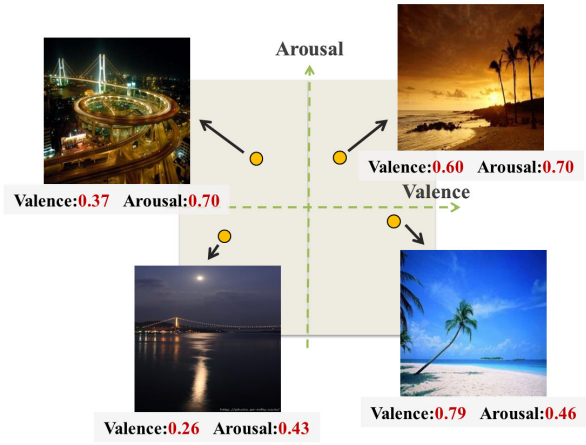

図 3 色特徴による心理値の例

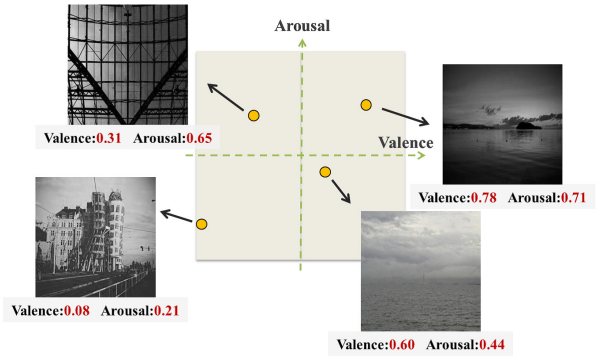

図 4 形状特徴による心理値の例

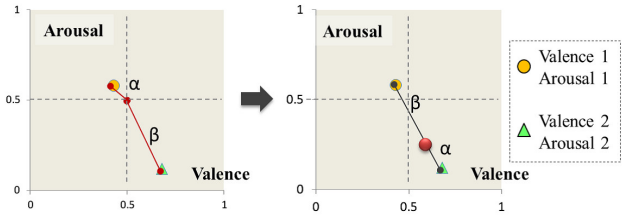

図 5 色と形状による AV 値の統合の例

$$V_2 = 4.57(D - 0.41) + 3.95 \quad (3)$$

$$A_2 = -0.29(C - 44.93) + 4.26 \quad (4)$$

3.1.3 色と形状による心理値の統合

式 (1)-(4) で求めた (V_1, A_1), (V_2, A_2) を，中心座標からの距離の比を用いて重み付けを行い統合する．統合の例を図 5 に示す．より強い印象を採用するため，中心座標から離れている AV 値に重みを乗せる．このようにして，入力画像を AV 空間へ配置することで，情景に対する印象を決定する．

3.2 楽曲の配置

本節では，AV 空間への楽曲の配置について説明する．画像の配置と同様に，楽曲を聴いたときに人が感じる情報についても体系化する必要がある．

3.2.1 主成分分析による特徴量の抽出

Eerola らは，29 次元の音響特徴量が楽曲のムードと関係することを示した．また，この特徴量を主成分分析することで得られる第一，第二主成分が，それぞれ Arousal，Valence と深い関連があることを実験により示した [14]．その実験とは，18-42 歳の 116 人の被験者に 110 楽曲を聴かせ，抱いた感情の評価を行ったものである．

本研究では，楽曲の AV 空間を構築するための主成分分析のための学習データとして，“RWC Music Database: Music Genre” [15] に収録されている 100 楽曲を用いた．RWC Music Database の楽曲はア・カペラの 1 楽曲を除くと，10 のジャンルと，33 のサブジャンルに分類され，ジャンルが均等に分かれている．音響特徴量を抽出する

Vol.2013-MUS-98 No.10
Vol.2013-EC-27 No.10
2013/3/15

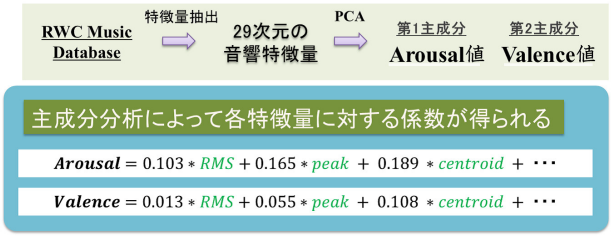

図 6 主成分分析による係数の抽出

際，我々は Eerola らによる先行研究と同様のツールである MIR toolbox Ver1.3[16] を用いた．得られた特徴量 f に対し，本研究では以下の式 (5) によって正規化を行う．主成分分析により算出した第二主成分までを採用することで各特徴量に対する係数を得た．主成分分析による係数の抽出の概要を図 6 に示す．これらの係数を各楽曲の 29 次元の特徴量に掛け合わせることにより，楽曲を AV 空間に配置することができる．このようにして，楽曲に対する印象を決定する．

$$f' = \frac{f - f_{ave}}{f_{std}} \tag{5}$$

ここで，f_{ave} は音響特徴量の平均値，
f_{std} は音響特徴量の標準偏差を表す．

3.2.2 印象に影響を与える音響特徴量の考察

3.2.1 で求めた Valence や Arousal に特に影響を与えている特徴量を，寄与率の大きい順に表 1 に示す．

表 1 Valence，Arousal との相関が高い特徴量

Valence	Arousal
chromagram centroid	rolloff
chromagram peak	spectrum flatness
zerocross	spectrum entropy
brightness	spectrum centroid
irregularity	flux

表 1 からわかる，Valence，Arousal に大きく寄与している特徴量を考察する． Valence に対しては，音階の成分の強さを表す chromagram の影響が大きいことが分かった．Chromagram の，重心を表す centroid と peak が強いことから，多くの音階が同時に鳴っている場合よりも，一つの音階が強く鳴っているほうが楽曲をポジティブに感じやすいといえる．また，1500Hz 以上の音域の割合を示す brightness が入っていることから，高音であるほうがポジティブに感じやすいことが確認できた．一方 Arousal に対しては，低音域の割合を表す roll-off が強いことから，ドラムなどの音が多く入っている楽曲を激しい楽曲と感じることがわかった．また，どの周波数を含むかを示す spectrum の平坦度，エントロピー，重心が強いことから，広い周波数の音が同じくらい大きく鳴っているほど，活発に感じることが確認できた．

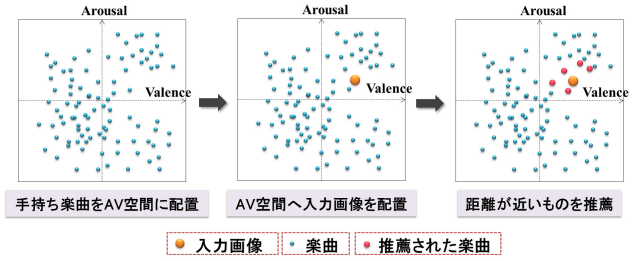

図 7 楽曲推薦の流れ

4. システムの実装

本章では，我々が実装した情景に合った楽曲を推薦するシステム，**レコむし** (**RECO**mmendation of **MUS**ic from **I**mage) の概要及びインタフェースについて説明を行う．

4.1 システムの概要

本研究で提案する楽曲推薦システム，**レコむし**による楽曲推薦の概要について説明する．

図 7 に，本システムの楽曲推薦の流れを示す．ユーザがウェブカメラを用いて撮影することにより画像が本システムに入力される．入力画像は 3.1 で説明した方法により AV 空間上に配置される．その後，同空間に事前に配置された楽曲群の各 AV 値に対して，入力画像の AV 値とのユークリッド距離を計算する．この距離が小さい楽曲 6 曲により，プレイリストが構成される．このプレイリストが本システムによる楽曲推薦の結果である．これによりユーザは本システムを通じて，情景に合った楽曲を楽しむことができる．

また，AV 空間内での入力画像の座標をドラッグすることにより，AV 値を自由に変更しながら楽曲をブラウジングできる機能を追加した．ドラッグの動きに合わせて楽曲の推薦が随時行われ，ドラッグ中はマウスから一番 AV 値の近い楽曲が流れる．これにより，従来の楽曲プレイリストと同様に，推薦結果に満足がいかない場合でも，楽曲を聴きながらユーザがインタラクティブに聴きたい曲を探すことができる．

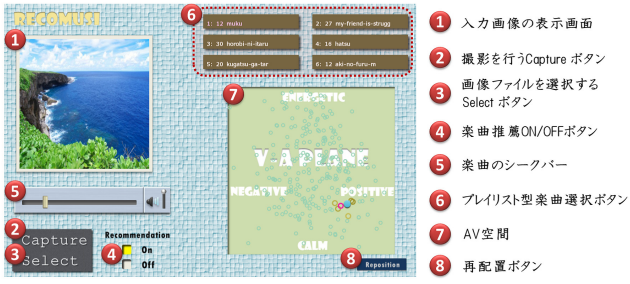

図 8 レコむしのインタフェース画面

Vol.2013-MUS-98 No.10
Vol.2013-EC-27 No.10
2013/3/15

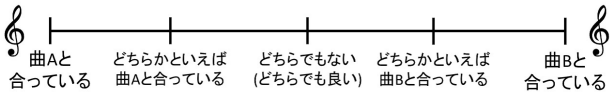

図 9 AB 法による 5 段階評価

4.2 インタフェース

楽曲推薦システムである**レコむし**のインタフェースについての説明を行う．

図 8 に本システムのインタフェース画面を示す．入力及び視聴のための基本機能として，入力画像の表示画面（図 8，①），ユーザが画像を撮影するための Capture ボタン（図 8，②），持っている画像ファイルを選択するための Select Pic ボタン（図 8，③），入力画像に対して楽曲推薦を開始するスタートボタン（図 8，④），推薦楽曲のシークバー（図 8，⑤），プレイリスト型楽曲選択ボタン（図 8，⑥）がある．入力された画像や，事前に登録した楽曲群は，インタフェース内の AV 空間（図 8，⑦）に配置される．この AV 空間は横軸が Valence，縦軸が Arousal に対応している．また，入力画像の位置は青い点で表され，推薦された楽曲は茶色の点で表示される．薄い緑色の点は，推薦された楽曲以外の楽曲を示す．

ドラッグによる移動で位置の変わった入力画像の点を，本来の AV 値に戻す際には，再配置ボタン（図 8，⑧）を用いる．

この移動機能により，ユーザは入力画像の AV 値を変化させることで，より好みの楽曲を楽しむことができるようになった．

5. 評価

本章では，提案手法の評価のために行った主観評価実験の方法，結果，考察について述べる．

5.1 主観評価実験

主観評価実験により提案手法の評価を行った．比較手法としてランダム選曲を用い，「提示された画像の印象に対してどちらの選曲が妥当であるか」を図 9 のようにして AB 法で 5 段階評価した．本実験は，20 代男性 15 名女性 2 名の計 17 名に対して 16 枚の画像を提示して評価してもらうことで行った．評価に用いた画像を図 10 に示す．

5.2 実験結果

各画像に対するスコアを図 11 に示す．スコアは 3 を中心として，提示された画像と本システムによる推薦楽曲が調和しているほどスコアを 5 に，ランダム選曲と調和しているほど 1 に近づく．実験の結果，16 枚の平均のスコアは 3.89 となり，本システムによる推薦楽曲の印象が，ランダムに選曲する場合に比べて入力した画像と感性的に対応が取れていることがわかった．

図 10 評価に用いた画像

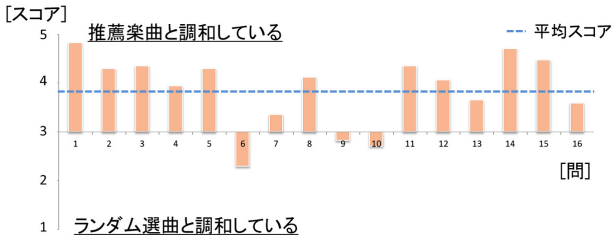

図 11 評価実験のスコア

5.3 考察

特に評価がランダム選曲と調和しているという結果が出た問 6, 9, 10 について考察を行う．図 12 に問 6, 9, 10 の表示画像と AV 値を示す．

問 6 では，画像が AV 空間の中心座標付近に位置している．中心座標付近では印象に特徴がないため，この部分は個人によって感じ方の変化が出やすいと考えることができる．一方，システムはどのような印象にも合わない楽曲を推薦してしまった．そこで，スキップボタンが押されることで，システムがユーザの好みを学習し，推薦結果を改善していく機能を追加することでこのような問題に対応できると考えている．

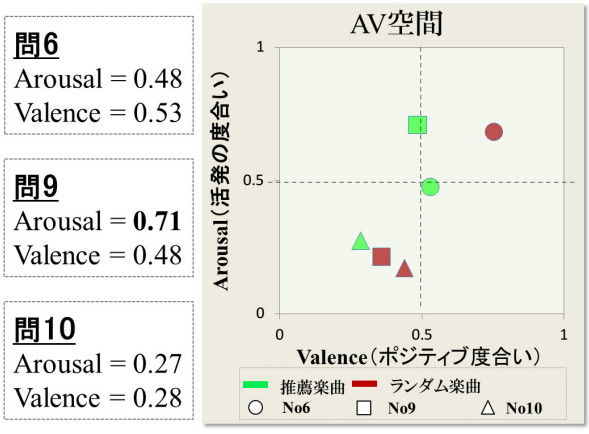

図 12 問 6, 9, 10 の入力画像と AV 値

問 9 では，橙色が高い Arousal 値として観測された．これは，3.1.1 で記述したカラーパッチによる実験では橙色は活発な印象と判断されたことによると考えられる．しかし，現実世界で目にする橙色は夕日や枯葉といった比較的穏やかな印象を与える景色に多いため，活発な印象を持つ推薦楽曲と合わずスコアが低くなったのではないかと考えられる．そこで，この問題を解決するために，新たな色尺度の導入を検討している．

問 10 では，推薦楽曲とランダム選曲による楽曲の心理値が近い値となった．推薦自体は成功したが，ランダム選曲による似たような印象の楽曲に劣ってしまったのではないかと考えられる．これは評価方法に関する問題点であり，今後，推薦楽曲と近い AV 値を持つ楽曲を比較対象としないことで改善可能である．

他の問では，提示された画像が鮮明に見える問のスコアが高くなっていることがわかる．これは，鮮明な画像ほどテクスチャ特徴量が精度良く求められるためだと考えられる．画像が暗く形状が判断しづらいものは，色のみの印象となるため，精度にバラつきが現れている．

6. 結論

本稿では，情景にマッチした楽曲の推薦を行うため，言語情報を用いずに楽曲を提示するシステム，**レコむし**の構築を行った．提案手法では AV 空間を用いることで，情景と楽曲の感じ方を架け橋に楽曲を推薦することを可能とした．これにより，視覚と聴覚の調和を楽しむことができる．さらに本研究により，感性を媒介情報として用いることで，その場の状況に相応しい音楽を見つけだす新しい音楽提示手法への道を切り開くことができた．本システムのインタフェースでは，楽曲を聴きながら心理空間上を移動するという今までにない音楽のブラウジング手法を提案した．現時点での応用法として，ドライブ中に見える景色に合う楽曲の推薦などに利用できると考えられる．また，提案手法を逆方向に用いることで，楽曲に感性的に合う画像の推薦も可能となる．応用例として，楽曲の入力による感性的な自動スライドショー生成や，自動 VJ 機能などが考えられる．

今後，新たな色特徴など考慮すべき画像特徴量の種類の増加やセグメンテーションなどの画像処理による注目物体の考慮による推薦精度の向上が考えられる．動画入力の対応による本システムの発展も検討している．また，スキップボタンを押すとシステムが学習を行い，推薦結果が個人向けにカスタマイズされていくインタラクティブな機能を加える事で，個人性も反映できるようなシステムの構築についても検討している．

本研究により，全ての楽曲との出会いを平等にすることで，人と音楽との新たな関係の実現に向けての大きな一歩を踏み出したい．

参考文献

[1] 大出 訓史, 今井 篤, 安藤 彰男, 谷口 高士, “音楽聴取における“感動”の評価要因——感動の種類と音楽の感情価の関係,” 情報処理学会論文誌 Vol.50(3), pp. 1111-1121, 2009.

[2] 岩宮眞一郎, “オーディオ・ビジュアル・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,” 音響学会誌 Vol.48(9),pp. 31-39, 1992.

[3] 古賀 広昭, 下塩 義文, 小山 善文, “画像にあった音楽の選定技術,” 映像情報メディア学会技術報告 Vol.23(59), pp. 25-32, 1991.

[4] 西山 正紘, 北原 鉄朗, 駒谷 和範, 尾形 哲也, 奥乃 博, “マルチメディアコンテンツにおける音楽と映像の調和に関する分析,” 情報処理学会研究報告 Vol.2007(15), pp.31-36, 2007.

[5] 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,” 音響学会誌 Vol.52, pp. 40-45, 1996.

[6] John S. Breese, David Heckerman, and Carl Kadie, “Empirical Analysis of Predictive Algorithms for Collaborative Filtering,” UAI’98 Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence, pp. 43-52, 1998.

[7] Keiichiro Hoashi, Kazunori Matsumoto, and Naomi Inoue, “Personalization of User Profile for Content-based Music Retrieval based on Relevance Feedback”, Proceeding MULTIMEDIA ’03 Proceedings of the eleventh ACM international conference on Multimedia, pp.110-119, 2003.

[8] 桐本 篤, 佐々木 史織, 清木 康, “風景画像とサンプル楽曲を用いた環境状況コンテクスト対応型音楽推薦システムの実現,” 情報処理学会 研究会報告, DBS-146, pp. 157-162, 2008.

[9] James Russell, “A circumplex model of affect,” Journal of Personality and Social Psychology Vol.39(6), pp. 1161-1178, 1980.

[10] Valdez Patricia, and Mehrabian Albert, “Effects of color on emotions,” Journal of Experimental Psychology: General Vol.123(4), pp.394-409, 1994.

[11] Hideyuki Tamura, Shunji Mori, and Takashi Yamawaki. “Textural features corresponding to visual perception,” IEEE Transactions on Systems, Man and Cybernetics Vol.8(6), pp.460-473, 1978.

[12] Jana Machajdik, and Allan Hanbury, “Affective Image Classification using Features Inspired by Psychology and Art Theory,” Proceeding MM ’10 Proceedings of the international conference on Multimedia, pp. 83-92, 2010.

[13] Peter J. Lang, Margaret M. Bradley, and B. N. Cuthbert, “International aective picture system (IAPS): Aective ratings of pictures and instruction manual.” Technical report A-6 University of Florida Gainesville, 2008.

[14] Tuomas Eerola,Olivier Lartillot, and Petri Toiviainen, “Prediction of multidimension alemotional ratings in music from audio using multivariate regression models,” In Proceedings of 10th International Conference on Music Information Retrieval, pp. 621-626, 2009.

[15] Masataka Goto, Hiroki Hashiguchi, Takuichi Nishimura, and Ryuichi Oka, “RWC music database: Music Genre Database and Musical Instrument Sound Database,” Proceedings of the 4th International Conference on Music Information Retrieval, pp.229-230, 2003.

[16] O. Lartillot, and P. Toiviainen, “MIR in matlab (II): A toolbox for musical feature extraction from audio,” Proceedings of the 5th International Conference on Music Information Retrieval, pp. 127-130, 2007.