2D-5

入力画像に感性的に一致した楽曲を推薦するシステム

佐々木　将人　　平井　辰典　　大矢　隼士　　森島　繁生
早稲田大学 / JST CREST

1 はじめに

音楽のデジタル化に伴い個人が持ち歩ける楽曲の数は増加している．一方，所有する大量の楽曲の中から最適な１曲を探し出すことは困難である．具体的に聴きたい楽曲はないが，音楽を楽しみたい場合，視覚と聴覚の調和を感じられるような楽曲を聴くと心地良い気分になることは容易に想像できる．しかし現状ではシャッフル機能や手作業で楽曲を逐一検索しなければならない．そこで本研究では，その場の情景に感性的に一致した楽曲及びプレイリストを自動で推薦するシステムを提案する．これにより，視覚と聴覚が調和した楽曲を聴くことができ，音楽の楽しみが広がる．

2 関連研究

2.1 従来の楽曲推薦

従来の楽曲推薦手法の手法を紹介する．1 つは，協調フィルタリングを用いた推薦手法で，他ユーザの楽曲評価を参考に楽曲の推薦を行う．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．しかし，いずれの手法もユーザの視覚情報が未考慮のため，状況に適した最適な選曲が困難である．

一方，ユーザの置かれた環境をアノテーション情報に落とし，歌詞やメロディと比較することによる楽曲推薦も行われているが，歌詞のない楽曲には完全に対応はできない上に，アノテーションにより記述できる情報は限られているため，正確に現在の情景を考慮できない．

2.2 感性の考慮手法

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[1]．AV 空間は，Arousal 軸(energetic-calm)と Valence 軸(positive-negative)の二軸から成る二次平面である．この AV 空間上の座標値である AV 値は人の感性や感情を表す．また，AV 値間の距離が近いほど，類似した印象を受ける．

Affective Music Recommendation System based on Input Image
Shoto SASAKI Tatsunori HIRAI Hayato OHYA and Shigeo MORISHIMA
Waseda University / JST CREST

そこで，本研究では AV 空間を用いて，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ことで情景に合った楽曲の推薦を実現する．

3 提案手法

3.1 AV 空間における画像の配置

AV 空間に画像を配置するためには，画像から感じられる情報を体系化する必要がある．

Valdez らは，画像の特徴量である彩度と輝度が，Arousal，Valence と対応関係を持つことを提唱した[2]．この輝度 Y，彩度 S と Arousal A，Valence V との関係を式(1)，(2)に示す．

$$A = -0.31Y + 0.60S \quad (1)$$

$$V = 0.69Y + 0.22S \quad (2)$$

この相関関係を利用することで，画像の彩度，輝度を AV 空間へ反映させる．

また，本研究では評価用のデータセットとして， AV 空間へ配置済みの画像のセットである IAPS データセット[3] を利用した．IAPS を正解データとして，画像のテクスチャ特徴量である Tamura 特徴量[4]と Arousal，Valence との間の関係を正準相関分析で求めた．得られた Tamura 特徴量の direction D 及び coarseness C と Arousal，Valence との関係を式(3)，式(4)に示す．この相関関係により，画像のテクスチャの direction 及び coarseness を AV 空間へ反映させる．

$$A = -0.29(C - 44.93) + 4.26 \quad (3)$$

$$V = 4.57(D - 0.41) + 3.95 \quad (4)$$

輝度，彩度を元に算出した AV 値とテクスチャ特徴量を元に算出した AV 値を，AV 空間の中心からの距離の比を用いて統合する．このようにして，情景を AV 空間へプロットすることで，情景に対する印象を決定する．

3.2 AV 空間における楽曲の配置

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．

Eerola らは，音響特徴量を主成分分析することにより得られる第一，第二主成分が，Arousal，Valence と高い相関があることを示した[5]．

本研究では，音楽的ジャンルが均等に分散している RWC Music Database[6]を，楽曲の AV 空間を構築するための主成分分析用の学習データ

として用いた．音響特徴量の抽出は，Eerola らによる先行研究と同様のツールである MIR toolbox Ver1.3[7]を用いた．得られた特徴量 f に対し，式 (5) によって正規化を行う．ここで，f_{ave} は音響特徴量の平均値を，f_{std} は音響特徴量の標準偏差を表す．

$$f' = \frac{f - f_{ave}}{f_{std}} \qquad (5)$$

主成分分析により算出された第二主成分までを採用することで得た係数を，各楽曲の音響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．

3.3 システムの概要

図 1 に本システムの GUI 図を示す．まず，ユーザがカメラを用いて静止画を撮影すると，3.1 節で記述した手法で画像が AV 空間上に配置される．その後，ユーザが所有する楽曲を基に，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算出する．この距離が小さい楽曲 6 曲により，楽曲推薦プレイリストが構成される．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 空間の入力画像の点をドラッグすることにより，入力画像の AV 値を自由に変更することが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，推薦された楽曲が気に入らなかった場合は，楽曲を聴きながらインタラクティブにユーザの聴きたい曲を探すことができる．

4 評価実験

提案手法を評価するため，主観評価実験を行った．比較手法としてランダム選曲を用い，

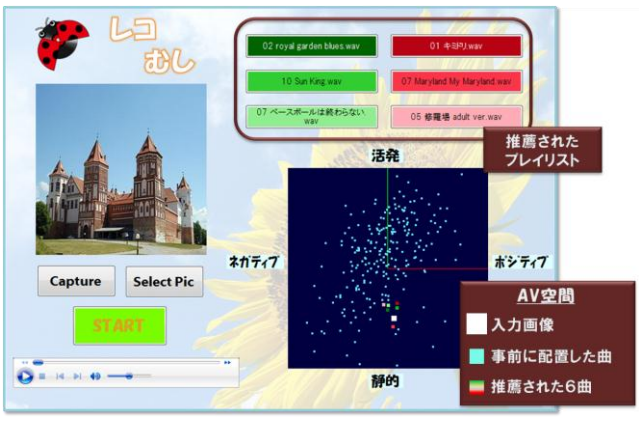

図 1. 本システムの GUI 図と解説

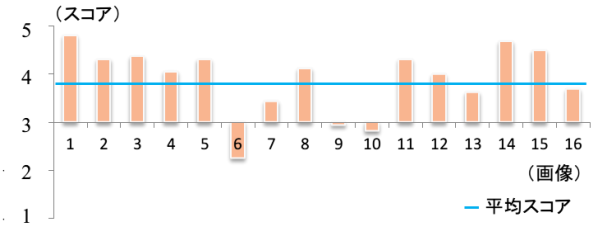

図 2. 各画像に対するスコア

「各画像の印象とどちらの選曲が妥当であるか」を 5 段階で評価した．今回，16 枚の画像に対して 20 代男女(男:14 人，女:2 人)を被験者とした．各画像に対するスコアを図 2 に示す．ここで，表示画像と推薦楽曲が調和している場合のスコアを 5 とした．平均のスコアは 3.89 であった．図 2 より，本システムによる推薦楽曲の印象が，入力した画像と感性的に対応が取れていることを確認することができた．しかし，AV 空間の中心に位置する画像や Arousal，Valence の値が共に高い画像においては，ランダム選曲の調和度の方が高い傾向があることもわかった．前者の場合は，AV 空間の中心が与える印象は強い特徴がないため評価が難しかったからではないかと考えられる．また後者では，今回のシステムは橙色が高い AV 値として観測されるため，夕日など落ち着いた場面の画像においてスコアが低くなったと考えられる．

5 おわりに

本稿では，AV 空間を用いて入力画像と楽曲の印象を対応付けることで，情景の印象に合った楽曲を推薦する新たな手法を提案した．

今後の課題として，新たな色の尺度を増やすことによる精度の向上が考えられる．

参考文献

[1] J. Russell, “A circumflex model of affect,” J. Personality Social Psychology, 1980.
[2] V. Patricia, et al., “Effects of color on emotions,” Journal of Experimental Psychology, 1994.
[3] P. Lang, et al., “IAPS: Affective ratings of pictures and instruction manual,” Gainesville, 2008.
[4] H. Tamura, et al. “Textural features corresponding to visual perception,” IEEE TSMC, June 1978.
[5] T. Eerola, et al., “Prediction of multidimension alemotional ratings in music from audio using multivariate regression models,” ISMIR, 2009.
[6] M. Goto, et al., “RWC music database: Music Genre Database and Musical Instrument Sound Database,” ISMIR 2003, pp. 229-230.
[7] O. Lartillot, et al., “MIR in matlab (II): A toolbox for musical feature extraction from audio,” ICMIR, 2007.