GRAPE : グラデーション画像による ポータブルビジュアルプレイリスト

魚田 知美[*1]　伊藤 貴之[*1]

GRAPE : Portable Visual Playlist featuring gradation images

Tomomi Uota[*1], Takayuki Itoh[*1]

Abstract – 実際に聞くまで中身を把握できない音楽を，短時間で直感的に把握する一手段として可視化が有用である．また音楽鑑賞の際，次に聞く曲を逐一選択する方法よりもプレイリストやアルバム単位で音楽を選択する方法のほうが一般的だと考える．そこで「1 画像で 1 プレイリスト」を表すプレイリスト単位の音楽可視化をポータブルミュージックプレイヤー上で実現する．処理手順は，3 ステップである．まず各楽曲から楽曲特徴量を抽出する．抽出した特徴量に基づいて各楽曲に色を割り当て，SOM（Self Organization Map : 自己組織化マップ）を用いて各楽曲を配置する．SOM を用いることで似ている曲は近くに集まるため，GRAPE では自動的にグラデーションのような画像が生成される．

Keywords : **music visualization, playlist, portable music player, Android**

1　はじめに

携帯型音楽プレイヤーの登場や音楽配信技術の発展で，音楽と触れ合う手段が増え，音楽は主要なコンテンツに成長したといえる．しかし「音楽」というメディアは，実際に聞いてみるまで中身を把握できないため，内容把握に時間がかかる．音楽の内容を短時間で直感的に把握する一手段として，画像で情報の全体像を表現する「可視化」が非常に有用であるといえる．

本報告では,「1 画像で 1 プレイリスト」を表すプレイリスト単位の音楽の可視化，GRAPE (GRadation Arranged Playlist Environment) を提案する．表 1 に示すアンケート結果からもわかるように，我々は音楽鑑賞の際，1 曲聞くごとに次に聞く曲を逐一選択する操作方法よりも，プレイリストやアルバムといった音楽のまとまりを選択する操作方法のほうが一般的だと考える．表 1 の選択肢 A，B，C はいずれも音楽のまとまりを選択していると言える内容であり，全体の 78 %を占めている．よって，プレイリストやアルバムの全体像を可視化することで，多くのユーザの選曲操作スタイルを変えることなく選曲操作を支援できると考える．現在の音楽プレイヤーでは一般的に，プレイリストはその名前のみで表示されるものであり，その名前を知らない場合には中身を推察することが難しい．また，そのプレイリストの中身を知らなければ，収録されている個々の曲がどのような印象であるかを推察することはできない．それに対して，本報告で提案する GRAPE は，色のついた正方形のタイルを並べてプレイリストを表現することで，プレイリスト全体を通してどんな印象を有するかという全体像と，個々の曲の印象と，両方を同時に表現する．

GRAPE は音楽再生機器上での選曲を容易にする他に，次のような効果も考えられる．まず，プレイリストや CD の中身に関する傾向や差異を知りたい時に，本手法によってまとまったオーディオ情報を概観できる．また，プレイリストを単純に画像として扱い，視覚的に楽しむ，ユーザ間で共有する，といったことが容易になる．可視化結果である画像が様々な機器やソフトウェアで扱いやすいメディアであることから，この他にも様々な活用ができると期待できる．

なお，本報告で言う「プレイリスト」とは，ユーザが作成したプレイリストや音楽プレイヤーによって自動作成されたプレイリストだけでなく，各アルバムや各アーティストまたは音楽コレクション全体などといった音楽のまとまりを総称するものとする．

本手法では SOM (Self Organization Map) を適用することで，グラデーション風の画像を自動生成する．グラデーションには，リズム感・メリハリを

[*1]お茶の水女子大学
[*1]Ochanomizu University

持つという表現効果がある．また，人の心を動かす心理効果がある．単色に比べてグラデーションの物には「触りたい」,「欲しい」といった要求が大きく作用する．この心理効果は GUI のボタン，その他の各種製品にも広く利用されている．このような効果は，可視化や情報を専門とする人だけではなく，音楽を聴く全ての人を対象とする音楽プレイヤーのデザインに有益と考えた．

表 1　アンケート結果
Table 1　Questionnaire results

設問：ポータブルプレイヤーで音楽を聴く際，どのような操作によって選曲することが多いか（回答数 261）

選択肢	%
アルバムやアーティストを選び，再生する	35
シャッフル機能を使い自分から曲を選ばない	22
プレイリストを選び，再生	21
1 曲ずつ操作して選ぶ	16
ポータブルプレイヤーを持っていない	5
アプリを使い，似た感じの曲を集めて再生	1

2　関連研究

楽曲特徴量から楽曲の可視化画像を生成する手法の例として，MusCat[1] や MusicThumbnailer[2] がある．MusCat は，数値解析ソフトウェア MATLAB の上に実装された楽曲特徴分析パッケージ MIRtoolbox[3] を用いて楽曲特徴量を抽出し，その特徴量に基づいて印象画像を生成する．以下に示す，図 1(左) からわかるように，各楽曲の特徴を推測できる可視化が実現されている．MusicThumbnailer は，MARSYAS[4] を用いて楽曲特徴量を抽出し，その特徴量に基づいてサムネイル画像を生成する．図 1(右) からわかるように，ジャンルが同じ曲は似たような画像が生成される．これより，各音楽のジャンルを推測できる可視化がされている．

以上にあげた手法はいずれも楽曲単位の視覚的表現を目指している．表示方法によって 1 画面にプレイリストを表示することはできるが，プレイリスト単位での音楽表現に特化しているとはいえない．一方，本研究 GRAPE では「1 画像で 1 プレイリスト」を目標とし，プレイリスト単位での音楽表現に特化していることから，ユーザの選曲操作スタイルに沿った可視化であると考えられる．しかし代わりに，以上の手法と比べると各楽曲の細かい内容の描写に欠ける面はある．

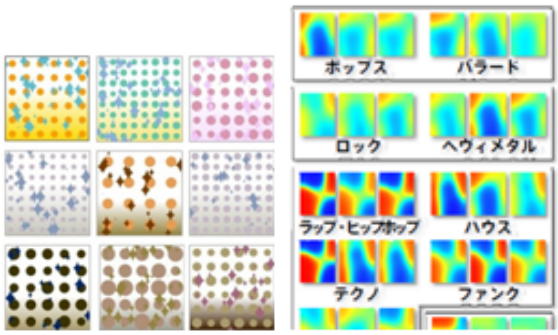

図 1　(左)MusCat[1] による可視化例
(右)MusicThumbnailer[2] による可視化例
Fig. 1　(left)Visualization result of MusCat[1]
(right)Visualization result of MusicThumbnailer[2]

3　提案手法

3.1　楽曲特徴量抽出

まず各楽曲から楽曲特徴量を抽出する．楽曲特徴量の抽出には，数値解析ソフトウェア MATLAB の上に実装された，楽曲特徴分析パッケージ MIRtoolbox を用いる．GRAPE では MIR toolbox で抽出可能な特徴量のうち，RMSenergy（音量平均値），Tempo，Brightness（高音域の割合）の 3 つの特徴量を，以下の式によって正規化して使用する．

$$f' = \frac{f - f_{min}}{f_{max} - f_{min}}$$

f_{max}　楽曲特徴量の最大値

f_{min}　楽曲特徴量の最小値

現在は，各楽曲の任意秒数を指定して特徴量を抽出しているが，多くの楽曲は曲の全体の中で特徴量が変化する．我々が実験に用いた楽曲の数ヶ所から試験的に特徴量を抽出してみた結果，楽曲中で値が比較的安定している上記 3 つの特徴量を代表として使用することとした．

3.2　YCbCr 色空間を利用した楽曲色

続いて，抽出した楽曲特徴量に基づいて各楽曲に色を割り当てる．色を割り当てる際に我々は，YCbCr 色空間を採用した．我々が実験に用いた楽曲を，試験的に RGB，HSB などの多種な色空間に配置した結果，YCbCr 空間に配置したときが最も色の分布がばらつく傾向にあり，GRAPE で採用したときの視覚的効果が高いと判断した．現時点では図 2 のように，Brightness を Y 軸， RMSenergy

を Cr 軸，Tempo を Cb 軸へとそれぞれ対応させて色を決定している．

Brightness は単純に各楽曲の輝かしさと表現することができる特徴量であるから，Y 軸である明度と対応させた．RMSenergy は楽曲の音量平均値を示し，アコースティックな曲では小さな値，エレクトリックな曲では大きな値になる傾向がある．これは赤の色が持つイメージ [5] である「活動的」や「エネルギッシュ」というキーワードに意味が近いと考えて対応させた．Tempo は値が大きいほど速い曲である．よって青の色が持つイメージ [5] である「開放感」や「爽やか」というキーワードに意味が近いと考えて対応させた．また，YCbCr 色空間では Cr 軸と Cb 軸の値が小さくなると緑に近づく．つまり RMSenergy と Tempo の値が小さい楽曲には緑が割り当てられるということである．これは緑の色が持つイメージ [5] である「穏やか」や「落ち着いた」というキーワードに意味が近いと考えた．以上の理由から YCbCr 色空間へ 3 つの特徴量を図 2 のように対応させた．

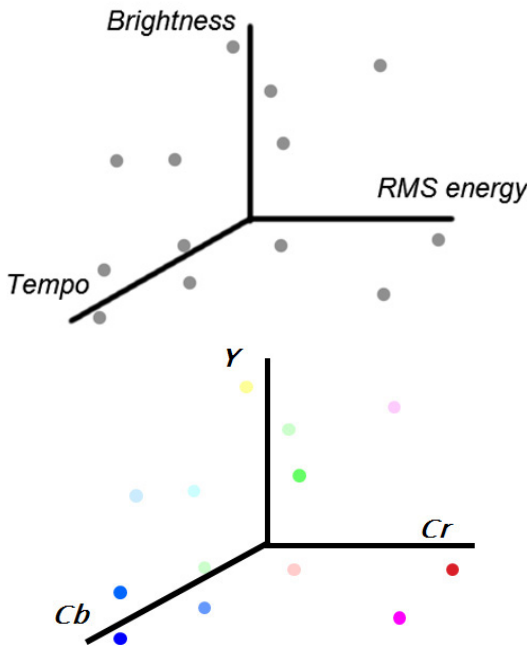

図 2　3 つの特徴量と YCbCr 色空間の対応
Fig. 2　Corresponding YCbCr color space and three features

3.3　SOM による楽曲配置

続いて，抽出した特徴量に基づいて各楽曲を配置する．各楽曲を配置する手法として SOM を用いる．SOM とは Kohonen が提案した人工ニューラルネットワークの 1 種であり，以下のステップで構成される．

1. 初期値としてランダムな値を持った 2 次元空間の中から，入力要素と 1 番近い点（勝者ユニット）を探しだす
2. 勝者ユニットとその近傍が入力要素に近づくように学習を行う
3. 全ての入力要素について 1. と 2. を適用し，これを反復する

SOM を用いることで，入力データ群の距離を保ったまま，任意の次元へと写像することができる． 1 楽曲を 1 要素として SOM を適用することにより，似ている楽曲は近い場所に集まり，あまり似ていない楽曲は遠い場所に配置される．これにより自動的にグラデーション風の画像が生成される．また今後，使用する特徴量を増やした際にも，多次元情報を持つ要素を 2 次元空間上に写像すること強みとする SOM であれは柔軟に対応できると考えられる．

GRAPE では各楽曲を正方形のタイルのように見立てプレイリストごとに配列して 1 枚の長方形を形成する．SOM を用いる以外の手段として，ボロノイ図などの領域分割手法を適用して各領域に楽曲を割り当てることも考えたが，各楽曲を表現する領域の大きさと形状が均一である方が，GUI 上での選曲操作を容易にできると考え，本報告では SOM を採用した．

また，正方形はディスプレイ上で小さく表示した時に最もつぶれにくい形状であることが知られている．よって今後，GRAPE を携帯電話やポータブルプレイヤーのような小さな画面を持つデバイス上に実装しても，容易にプレイリストの印象を感じ取ることができる上に，小さく表示した場合にもタッチ操作等で個々の曲を選ぶことができる．このことも GRAPE が採用したデザインの利点の一つといえる．

3.4　画像一覧表示

以上により取得した楽曲の位置情報と色情報をプレイリスト毎にまとめて画像とし，プレイリスト群を一覧表示させる．

3.4.1　PC 上での表示

我々が開発した GRAPE の GUI を図 3 に示す．この GUI には，画像一覧表示の他に，曲名表示や楽曲再生機能も備えている．この機能を利用することで，各楽曲と割り当てられた色との対応を確認することができる．また，従来の音楽プレーヤーの代替ソフトウェアとして利用できると考える．

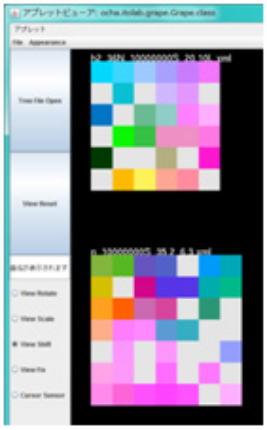

図 3　PC 上での表示
Fig. 3　Displaying on PC

3.4.2　Android アプリケーション上での表示

現在，一般的な音楽鑑賞デバイスとなったポータブルミュージックプレイヤーでの実装も合わせて行った．PC 上での表示の際に使用したデータを用い，同様の可視化結果を得ることができる．これより本研究を，ポータブルミュージックプレイヤーのようにどこへでも持ち歩けるプレイリスト，ポータブルビジュアルプレイリストと表現した．

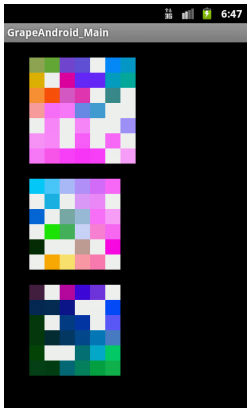

図 4　Android アプリケーションでの表示
Fig. 4　Displaying on Android Application

4　実行結果

本報告では，GRAPE にプレイリスト A，B，C について可視化を実行した．プレイリスト A は 36 曲，プレイリスト B は 28 曲，プレイリスト C は 29 曲の楽曲が入っている．可視化結果を図 5 を以下に示す．なお，可視化結果内で白くなっている部分は，楽曲が割り当てられていない領域である．

図 5　ポータブルビジュアルプレイリスト
(上) プレイリスト A　(左下)B　(右下)C
Fig. 5　Portable Visual Playlist (top)playlist A
(left bottom)B (right bottom)C

次に，プレイリスト別の各特徴量の分散値と平均値を図 6,7 に示す．図 5 に示す可視化画像から筆者が直感的に得られる視覚的印象と，図 6,7 に示す楽曲特徴量の数値的傾向は，概ね整合していた．例えばプレイリスト C は A,B と比べて Y 値と Cb 値の平均が極端に小さいが，画像を見てもプレイリスト C には明るい色や青みがかった色が少ない．また Y 値の分散はプレイリスト A よりも B,C のほうが大きいが，画像を見てもプレイリスト B,C のほうが明るい色も暗い色も混在している．さらに，プレイリスト A と B は特徴量の分散値・平均値が近いが Cr 値はプレイリスト A のほうが大きい．画像を見てもプレイリスト A の方が暖色が多く，よりエネルギッシュな楽曲が多いことを連想できる．

以上のように，GRAPE によって各プレイリストの特徴を視覚的に表現できたことがわかる．

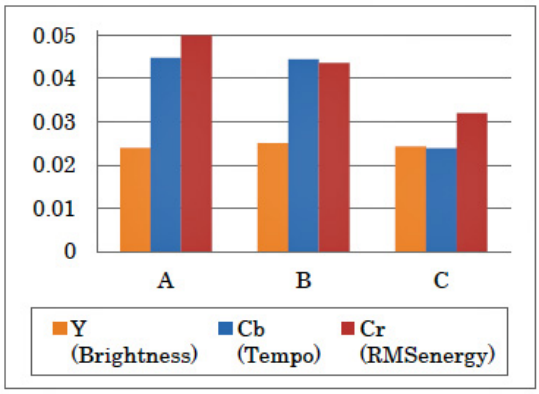

図 6　プレイリスト別の各特徴量の分散値
Fig. 6　Variance value of each feature

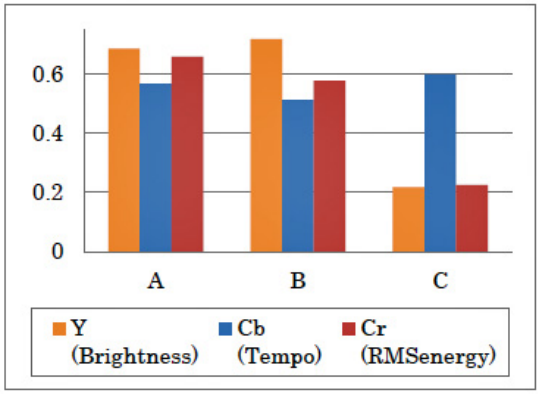

図 7　プレイリスト別の各特徴量の平均値
Fig. 7　Average value of each feature

5　ユーザテスト

我々は，被験者 (20 代女性) に図 5 の可視化結果画像 3 枚を観察してもらい，それぞれの画像について 3 つの問いをした．それぞれの設問内容と回答について以下に述べる．

ユーザテスト問 1

このプレイリストは他のプレイリストと比べてどんな楽曲が多いと想像されるか（14 キーワード複数選択可，14 名に実施）

問 1 の結果を表 2 にまとめた．プレイリスト A，C は特定のキーワードへ回答が集中し，各プレイリストへの認識・理解が多くの被験者間で共通していることが確認できた．

プレイリスト A に関して約 8 割の人が選んだキーワードとして「明るい」「にぎやか」「POPS」がある．「明るい」は Y 軸の Brightness と，「にぎやか」は Cr 軸の RMSenergy と対応していると考えられ，いずれも特徴量と整合している回答だと言える．また「ポップス」が選ばれていることから，J-POP と J-ROCK の楽曲のみを集めたプレイリスト A のジャンルを推定できていることが確認できた．

プレイリスト C については「クラシック」「バラード」というキーワードに過半数の回答が集まっている．プレイリスト C はクラシック楽曲のみを集めたプレイリストであるので，実際に含まれている楽曲ジャンルを概ね当てているといえる．また，9 割以上の人が「暗い」「遅い」というキーワードを選択した.「暗い」は Y 軸の Brightness と対応していると考えられ，特徴量と整合している回答である．しかし「遅い」は Cb 軸の Tempo の値と整合しているとは言い難い．クラシックなどのアコースティックな楽曲は楽曲中でテンポが変化しやすいが，本研究では任意秒から抽出した特徴量を楽曲の代表特徴量として使用している．このことから被験者の認識と特徴量との間に差が生まれてしまったと考えられる．

最後にプレイリスト B だが，プレイリスト A，C と比べて特定のキーワードへの回答の集中が見られなかった．これは，プレイリスト B が様々な雰囲気やジャンルの楽曲を含んでいることが原因だと考える．また反対に，このプレイリスト B の内容を良く反映した結果だとも言える．

表 2　ユーザテスト問 1 の結果 (%)
Table 2　Results of user test’s Question 1(%)

キーワード	A	B	C
明るい	86	57	0
暗い	0	0	100
速い	14	29	0
遅い	0	14	93
にぎやか	79	14	0
静か	0	21	21
壮大	14	0	64
繊細	7	43	7
ロック	36	14	0
ポップス	79	36	0
ジャズ	0	21	36
クラシック	0	21	57
バラード	7	21	64
RB	7	0	7
インスト	0	43	14
テクノ	7	21	0

ユーザテスト問 2

このプレイリストの中でクリックしてみたいと思った曲はどれか，またその理由は何か（複数選択可，14 名に実施）

次に，問 2 で得られた回答について，図 8 にまとめた．選択された楽曲の上に票数を表示している．図 8 内にまとめた以外にも，プレイリスト A では「ピンク色の楽曲全体」，プレイリスト C では「左端一列の楽曲全体」という風に，プレイリスト内で頻出な色の楽曲のまとまりを選んだ被験者もいた．

結果から，いずれのプレイリストにおいても，比較的プレイリスト内で少ない色の楽曲へ票が集まっていることが確認できた．これは，本来のプレイリストの傾向の読み取りを損なわずに，プレイリスト内で特異な存在の楽曲を即座に発見するという活動を支援できる．

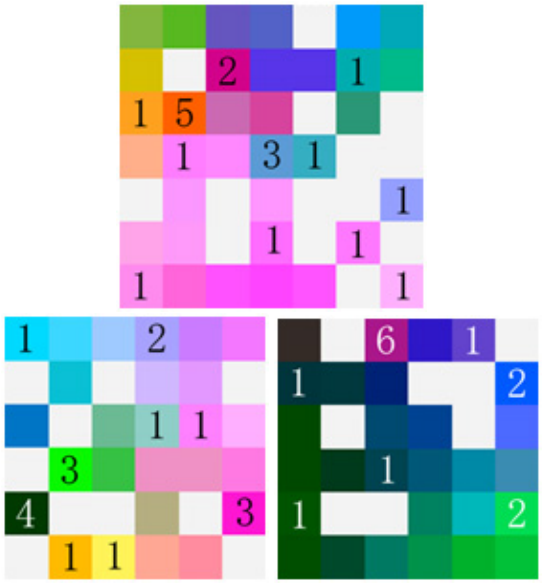

図 8　ユーザーテスト問 2 の結果 (票数)
Fig. 8　Results of user test's Question 2(Votes)

ユーザテスト問 3

実際に視聴して印象は合致したか，またその理由は何か（5 段階評価，27 名に実施）

最後に，問 3 で得られた回答について図 9 にまとめた．いずれのプレイリストにおいても 1,2 と評価した人より 4,5 と評価した人の方が多数であった．しかしプレイリスト A については，1,2 と評価した人が 3 割を超えた．プレイリスト A において 1,2 と評価した人の理由として，歌手か歌詞から得るイメージと楽曲色との差異が多く挙げられていた．例えば，女性歌手の楽曲や恋愛の楽曲はピンクといったイメージである．このように特徴量による音楽認識をしていないユーザにとっては特徴量による楽曲の把握には向いていない，また GRAPE では有益な効果を与え辛いことがわかった．今後ユーザテスト等を実施する際には，ユーザが普段どのように楽曲を認識しているのかを合わせて調査すべきだと考える．

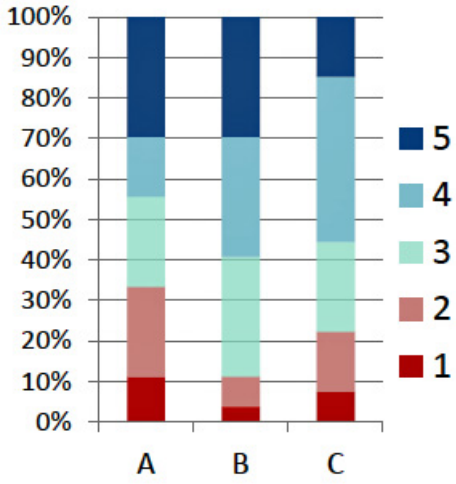

図 9　ユーザーテスト問 3 の結果
Fig. 9　Results of user test's Question 3

6　まとめと今後の課題

本報告では，SOM と YCbCr 色空間を利用して，特徴量に基づいたプレイリスト毎の画像を生成し，まとまった楽曲情報を表現する可視化手法 GRAPE を提案した．実行結果やユーザテストから，数値情報に頼らずに各プレイリストの特徴が表現でき，特定の傾向については読み取れることを示せた．

今後の課題として，各楽曲の色や位置が，人間が直感的に感じる色，それに伴った配置に近づくよう工夫していきたい．また人とのポータブルビジュアルプレイリスト共有を支援する等の拡張を考えている．最後に，現在 GRAPE では MIRtoolbox で抽出できる楽曲特徴量のみを使用しているが，今後はその他の情報（メタ情報，歌詞など）も組み合わせていくことで，より効果的な音楽可視化を実現していきたいと考えている．

参考文献

[1] K. Kusama, T. Itoh, MusCat: A Music Browser Featuring Abstract Pictures and Zooming User Interface, ACM Symposium on Applied Computing, Multimedia Visualization Track, pp. 1227-1233, 2011.

[2] K. Yoshii, M. Goto, Visualizing Musical Pieces in Thumbnail Images Based on Acoustic Features, 9th International Conference on Music Information Retrieval (ISMIR), pp. 211-216, 2007.

[3] O. Lartillot, MIRtoolbox, http://www.jyu.fi/hum/laitokset/musiikki/en /research/coe/materials/mirtoolbox.

[4] G. Tzanetakis et al, MARSYAS, http://marsyas.info/

[5] 芳原, 色彩の教科書, 洋泉社, 2011