MusCat:
楽曲の印象表現に基づいた一覧表示の一手法

草間かおり[†]　　伊藤貴之[††]

近年のマルチメディア技術の発達により，今日では個人が PC やオーディオプレイヤーに楽曲を保有することが主流となった．保有する楽曲数の増加に伴い，ユーザが聴きたい楽曲を見つけることが困難となる．そこで，曲名やアーティスト名等のメタデータに依存せず，旋律や調性などの楽曲特徴や印象に基づいて楽曲を検索するツールを提案する．楽曲の印象を瞬時に直感的に認識するために，各楽曲データから特徴を検出し，楽曲をユーザのニーズに合わせて階層型にクラスタリングする．その後，その特徴量に基づいて印象画像を自働生成し，印象画像を用いて一覧表示する．

MusCat:
A Music Browser Featuring Abstract Pictures and Zooming User Interface

Kusama Kaori[†]　and Takayuki Itoh[††]

Today many people store music media files in personal computers or portable audio players, thanks to recent evolution of multimedia technologies. The more music media files these devices store, the more difficult it is to search for tunes that users want to listen to. We propose MusCat, a music browser to interactively search for the tunes according to features, not according to metadata (e.g. title, artist name). The technique firstly calculates features of tunes, and then hierarchically clusters the tunes according to the features. It then automatically generates abstract pictures, so that users can recognize characteristics of tunes more instantly and intuitionally. It finally visualizes the tunes by using abstract pictures. The technique enables intuitive music selection with the zooming user interface.

1. 研究背景

音楽の再生機材としてPCやポータブル音楽プレイヤーが主流になり，その内蔵記憶装置（ハードディスクや半導体メモリ）の記憶容量の増大に伴い，個人が保有する楽曲数が膨大化している．これにより，ユーザが聴きたい楽曲を見つけるのが困難になることが多くなると考えられる．多くの場合においてユーザは，曲名やアーティスト名などのメタデータをもとに楽曲を検索する．しかし，これらのメタデータに頼らすに，旋律や調性などの楽曲特徴や印象に基づいて楽曲を検索する技術には，まだ検討の余地があると考えられる．例えばカフェやバーのBGMの選曲のように，楽曲のタイトルやアーティストにこだわらず，場所，時間，状況によるムードに基づいて楽曲を選曲したい場面があるとする．あるいはユーザが，アーティストや作曲者に関わらず，ある楽曲について似た曲調の楽曲を見つけたい場合があるとする．このような場合には，楽曲を印象に基づいて分類し，特定の印象を有する楽曲群を簡単に提示できるツールがあると便利である，と考えられる．

以上の背景に基づいて本報告では，膨大な楽曲をメタデータに依存することなく，楽曲そのものの特徴から，印象に基づいて整理，分類し，わかりやすく一覧表示するツールを提案する．本手法では楽曲の一覧表示に，CAT[1]という大量画像の一覧可視化手法を用いる．CAT は階層型にクラスタリングした画像を表示し，ズームインおよびズームアウト操作により，高階層クラスタから低階層クラスタまで順に表示することが可能である． 視覚的に認識することの出来ない楽曲を一覧表示する場合，各楽曲を画像におきかえて表示する方法は有益である．そこで我々は楽曲の印象画像を媒体として一覧表示する．CAT を楽曲再生ソフトウェアとして用いているという意味から，本報告では提案手法を MusCat (Music CAT)と呼ぶ．

2. 関連研究

2.1 音楽と感性

感性語による音楽データベース探索システムに関する研究[2]では，楽曲データのサンプリング個数や周波数強度の値に対して，感性語との相関性をそれぞれ求め，楽曲の特徴と感性語の関係を，ニューラルネットワークを用いて検出している．

コレスポンデンス分析による楽曲の特徴認識[3]では，楽曲に対する印象を表わす形容詞を評価することによって，音楽の特徴認識には【重い―軽やかさ】,【スピードのある―ゆったりしている】,【power のある―power のない】という 3 軸で認識すること

[†] お茶の水女子大学大学院
Ochanomizu University
[††] お茶の水女子大学大学院
Ochanomizu University

を述べている．本研究では，これらのように楽曲の特徴を感性語で表すはなく，より直感的に画像で表現することを目的としている．

2.2 色と感性

色空間と感性の反映方法[4]では，単色刺激から受ける感性情報の因子として Evaluation（評価性）， Potency（力量性あるいは潜在性），Activity（活動性あるいは躍動性）があること，また 2 色配色の刺激から受ける感性情報の因子として Harmony（調和性）が存在することを述べている．このことから，色を用いて感性を表現することが可能であると考えられる．

カラーシステム[5]では，色を warm-cool 軸，soft-hard 軸からなる 2 次元の感性空間に配置することによって，印象を表現している．その際感性空間における単色配色の分布には偏りがみられ．感性空間に色が存在しない領域が存在する．一方多色配色では，感性領域全体を補うことができる．しかし，多色配色の場合，色の組み合わせにより組み合わせが多数存在し，色相配色やトーン配色の違い，清色配色や濁色配色の違いにより，様々な効果が生まれる．このことから単色配色よりも多色配色の方が，色の選択に十分な検討が必要だが的確な印象を表現するができる．

2.3 楽曲の印象と画像

印象の合う画像で楽曲を表現する手法に MIST[6]がある. MIST ではサンプルとなる楽曲や画像に対して，その感性語適合度をユーザに答えさせ，ニューラルネットワークを用いてその相関性を学習させる．それを基に任意の楽曲や画像の感性語適合度を自動算出し，楽曲の印象に適した画像をアイコンとして表示する．またこれに似た手法として Music icons[7]では，ニューラルネットワークを用いて適した幾何学模様の画像を選択している．

2.4 楽曲分析

Lie ら[8]は，楽曲を音量，リズム，音質に基づいて階層的にクラスタリング手法を提案している．楽曲を，音量によって高階層クラスタを生成した後，リズム，音質で低階層クラスタを生成することで，非階層的な分類よりも的確に分類している．このことにより，大量の楽曲から雰囲気に基づいて目的の楽曲を見つける際に，階層的に楽曲を分類することができる．ただしこの手法は，的確な楽曲分類の実現に特化しており，その表示方法については論じていない．

2.5 ユーザインタフェース

Musicream[9]では，マウス操作によって，特定の楽曲と類似した楽曲をユーザの意志に従って選曲する機能がある．Search inside the music [10]では，楽曲の多次元特徴ベクトルを 3 次元に変換し，それを 3 次元空間上に配置させて表示するインタフェースを提案している．また，楽曲のアルバムアートワークを表示し，その画像をクリックすることによって楽曲を再生したり，また再生している楽曲アートワークの近くに類似曲のアルバムアートワークを表示することで，わかりやすい表示を実現している．

3. MusCat の処理手順

MusCat では楽曲データとして CD に採用されている音響データ書式(wav ファイル)を用いて，(1)楽曲の特徴量の検出，(2)クラスタリング，(3)印象画像の生成，(4)一覧表示の 4 段階で処理を施す．

3.1 特徴検出

近年さまざまな楽曲特徴検出手法が提案されており，その中には提供特徴検出ツールとしてオープンソースで配布されているものもある．我々は本研究において，MIR toolbox[11]を用いて，楽曲から表 1 に示す特徴量を検出している．

表 1 楽曲特徴

特徴量	説明
RMS energy	音量
Low energy	弱音の割合
Tempo	テンポ
Zero crossing	波形が 0 値をとる回数
Roll off	85%を占める低音域の割合
Brightness	1500Hz 以上の音域の割合
Roughness	不協和音の多さを示す値
Spectral irregularity	音質の変化の大きさ
Inharmonicity	根音に従っていない音の量
Key	主に使われている音
Mode	major と minor の音量の差

本手法ではこれら 11 個の特徴量を同等に扱うために，特徴量 f を正規化した f’を用いる．ここで $f' = (f - f_{\min})/(f_{\max} - f_{\min})$ であり，f_{min} と f_{max} は特徴量の最大値と最小値である．

楽曲には，次第に曲の印象が変わるものや，一つの楽曲に複数の印象を合わせ持った曲があり，楽曲全体の特徴量を一つに定めるのは困難である．そこで，一つの画像を生成するには一定区間ごとに特徴量を検出する．次第に楽曲が変化する場合は各区間の特徴量の平均値をその楽曲全体の特徴量とし，特定の時刻から明確に印象が変わ

る楽曲に対しては，印象が変化する前と後に分けて複数の特徴量を一つの楽曲の特徴量とする．現段階では，楽曲の中間部のある 5 秒間から得た特徴量を楽曲全体の特徴量として暫定的に定めている．

3.2 クラスタリング

続いて，特徴量に基づいて階層的にクラスタリングする．楽曲は様々な特徴すべてを複合してひとつの楽曲であるので，各楽曲に対しすべての特徴を加味して分析する多変量クラスタ分析法を用いる．ここで，多変量解析にもいくつかの方法（最短距離法，最長距離法，群平均法，重心法，メディアン法，ウォード法等）がある．**エラー! 参照元が見つかりません。**にクラスタ分析結果の例をあげると，鎖状にクラスタが連結する，あるいは類似度距離が高階層と低階層で逆転する場合がある．そこで我々は，分類感度が高く，クラスタの要素数が比較的均一なウォード法を採用している．

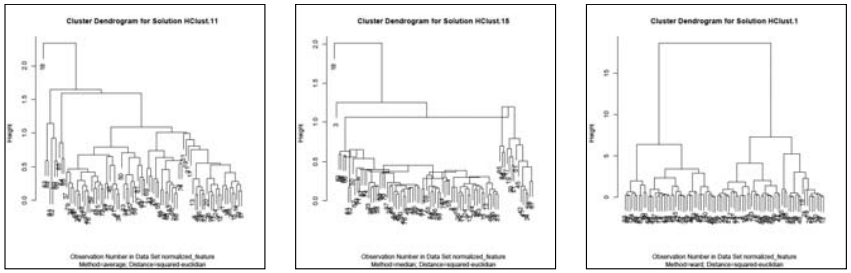

図 1　クラスタ分析の比較（左）群平均法（中央）メディアン法（右）ウォード法

3.3 画像生成

本手法では，ユーザが特定の印象を有する楽曲を直感的に発見するために，抽象的な印象画像を提示する．旧来から芸術分野では，目で見た風景や様子を楽曲で表現，曲を聴いた印象を絵画で表現する[12]ことがなされてきた．このことから，楽曲の印象を画像で表現することは有効であると考えられる．また人間の感覚には，音楽や音を聞いて色を感じる‘色聴’という知覚があること[13]から，音と色の印象は密接な関係にあるといえる．さらに，音と色の印象はしばしば同一の形容詞で表現される．以上のことから，視覚的に表現され得ない楽曲の印象を可視化する際には，色印象に重きを置いた抽象画像を媒体として表示することにより，楽曲の背景知識や歌詞の意味よりも全体的な印象を重視した視覚的表現ができると考えられる．

3.3.1 色の選択

まず，印象画像を生成する色を決定する．我々の実装では，カラーイメージスケール[5]（図 2 参照）をもとに印象画像の色を選択する．カラーイメージスケールは，warm-cool 軸と soft-hard 軸の 2 次元の感性空間に色を分布させたものである．本手法では楽曲を感性空間に配置し，感性空間上で最も距離の近い色を楽曲の色とする．

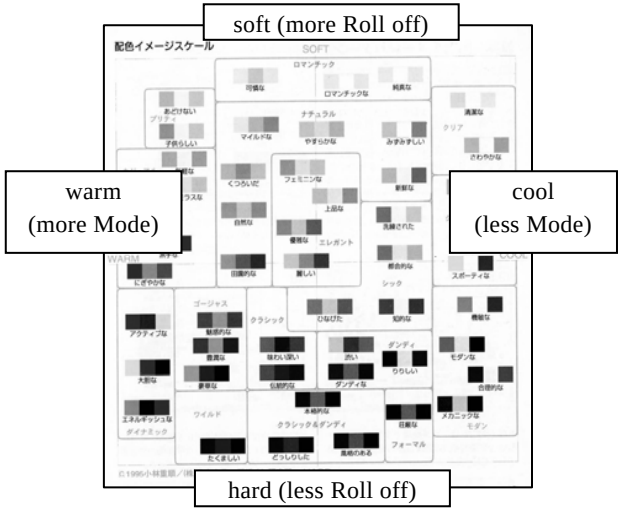

図 2　カラーイメージスケールの感性空間

続いて，warm-cool 軸，soft-hard 軸に対応する楽曲特徴量を考える．一般にメジャーコードに明るい・暖かいといったポジティブな印象を表わすコードであり，マイナーコードは冷たい・悲しいといったネガティブな印象を表わすコードである．このことから本手法では，マイナーコードとメジャーコードの音量差を示す特徴量 mode を warm-cool 軸に割り当てる．また一般的に，低音が少なければ軽い・やわらかい音，低音が多ければ固い・重い音と表現することが多い．このことから本手法では，低音域の音量を示す特徴量 roll off を soft-hard 軸に割り当てる．以上により，mode, roll off の 2 特徴量に基づいて，楽曲を色の感性空間に配置する．

その後，楽曲の座標値 (m_{wc}, m_{sh}) と最も近い座標の色をその楽曲の色とする．

$$color = \min \sqrt{(m_{wc} - c_{wc})^2 + (m_{sh} - c_{sh})^2}$$

3.3.2 デザインの選択

続いて，楽曲の特徴量に対応したデザインを生成する．本報告では，デザインの文法[13]に基づいたデザインの一例を示す．このデザインでは，RMS energy を背景のグラデーション, Tempo を円の個数, Spectral irregularity を円の配置バランス，Roughness を各円のサイズのばらつき，Brightness を星の個数に割り当てる（図 3 参照）．

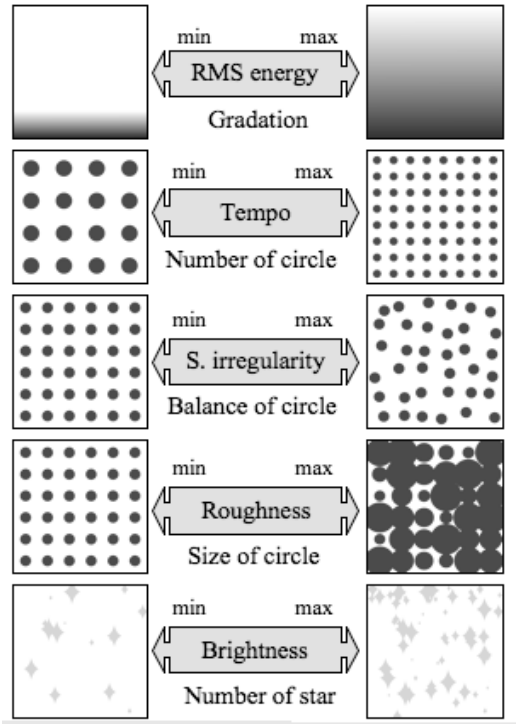

図 3　楽曲特徴に基づくデザイン生成

下からのグラデーションは重みと広がりを感じさせる構図となり，楽曲の重厚感を表すことができると考えられる．そこで我々は，音量を表す楽曲特徴量である RMS energy を，背景となるグラデーションに割り当てている．

Tempo を表す「速い—遅い」という形容詞から，「細かい—粗い」という形容詞も連想しうると考えられる．そこで我々は，デザインの細かさをオブジェクトの数で表現し，Tempo を割り当てている．

Spectral irregularity は音質の変化の大きさを表す値であり，不安定さを表しているとも考えられる．それをデザインで表現するために我々は，オブジェクトの配置にランダムさを与え，Spectral irregularity を割り当てている．

Roughness は不協和音の量を示す値であり，この値が大きいほど，音の調和が取れていないと考えられる．そこで均衡を崩したデザインを表現するために我々は，オブジェクトのサイズで不均一さを表現し，Roughness を割り当てている．

また Brightness は高音域の割合を示している．我々は高音域の割合を示すために，きらめき・輝きを示す星形のオブジェクトを扱い，Brightness の値に比例して星形のオブジェクト数を増やして表現する．

これらのデザインをそれぞれ生成した後，3 枚の画像を重ね合わせることによって，印象画像となる一枚の画像を生成する（図 4 参照）．

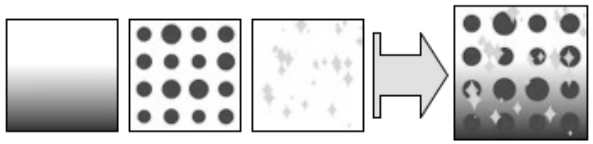

図 4　画像生成

なお，本報告で示す印象画像のデザインは，あくまでも文献に基づいて主観的に決めたデザインであり，まだ検討の余地が大きい．楽曲の印象と印象画像が対応するよう今後検証し，被験者からの意見をもとに，楽曲の印象に合うデザインを考案したい．我々は本手法における現段階の印象画像のデザインを，複数の楽曲の印象を示す標識的な役割を担ったものと考えている．

3.4　一覧表示

本手法では，画像ブラウザ CAT[1]に楽曲再生機能を設けたブラウザ MusCat を用いて，印象画像を一覧表示する．

CAT では，前処理として大量画像を階層型にクラスタリングし，それを互いに重ならずに等しいサイズで一覧表示する．CAT では各画像をサムネイル表示し，サムネイルを長方形の枠で囲うことでクラスタを表現する．さらに CAT では，ズーム率に合わせた詳細度制御を設けている．ズームイン時は，低階層クラスタの各々の画像サムネイルを表示する．そしてズームアウト操作に伴って，低階層クラスタを示す長方形領域を，各低階層クラスタの代表画像で置き換えて表示する．さらにズームアウト操作を続けると，高階層クラスタを示す長方形領域を各高階層クラスタの代表画像で置き換えて表示する．このように，CAT は階層化された画像群に対するズーム操作によって，直感的に画像を絞り込みながら閲覧できる．MusCat は，以上のような CAT の諸機能を継承し，かつ印象画像に対するマウス操作での楽曲再生機能を有する．

MusCat による印象画像の表示結果を，図 5 に示す．図 5（左）はズームアウト時，（右）は左の画像をズームイン操作することによって得られた画像である．左下のクラスタには，縁のオブジェクトの数が多く，その円のサイズにばらつきがある．つまりテンポが速く不協和音が多い楽曲のクラスタであることがわかる．また，右上のクラスタの画像は暖色配色のものが多く，円のサイズが均一である．このことから，長調が多く協和音の多いクラスタであることがわかる．このように MusCat を使うことで，楽曲を聴く前に楽曲の特徴を理解することが可能となる．

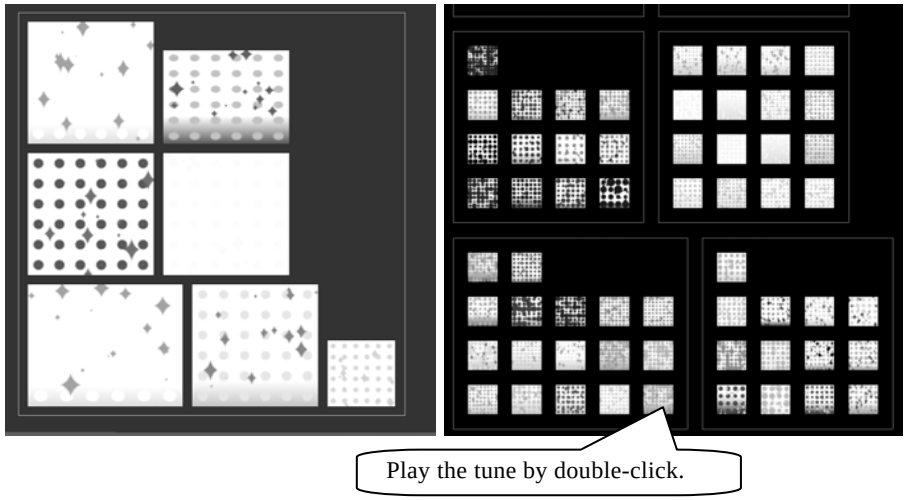

図 5　MusCat での表示結果

4. 実験結果

我々は実験として，88 曲 11 ジャンル(ポップス，ロック，ダンス，ジャズ，ラテン，クラシック，行進曲，ワールド，声楽，邦楽，アカペラ)で 31 のタグデータ（例えばラテンではボサノバ，サンバ，レゲエ，タンゴ）を持つ楽曲について，本手法を施して印象画像を生成し，MusCat で一覧表示した．そして，(1) 楽曲が印象によって妥当性をもって分類されているか， (2) 楽曲の印象に適した画像が生成されているかについて，15 人の被験者に対して以下の実験を行った．

4.1 クラスタリング

まず，被験者に楽曲 2 曲を視聴させ，その 2 曲が似ているか 5 段階で評価させた．本実験では，同一クラスタ内のサンプル楽曲を 4 組，異なるクラスタに含まれるサンプル楽曲を 4 組用意した．被験者は 8 組のサンプル楽曲をランダムに試聴し，似ている場合を 5，似ていない場合を 1，として 5 段階評価を記録した．どの組が同一クラスタの組であるかは，被験者には知らされていない．同一クラスタに含まれている楽曲に関する評価を表 2，異なるクラスタに含まれる楽曲に関する評価を表 3 に示す．

同一クラスタに含まれる組の楽曲について，Pair 1 は同一ジャンルの楽曲である．同じジャンルに含まれる楽曲は，計算上類似性が高いと判断されており，高い評価が得られた．それと比較して，Pair 2～4 を見てみると楽曲が類似しているかについては被験者の意見がわかれ，全体的に広がった分布の結果を得た．一方，異なるクラスタに含まれる楽曲では，評価値が低く良い結果が得られた．

同一クラスタにおける類似度の認識は，必ずしも非常に良い結果であるとは言い難い．しかし，同一クラスタの楽曲の組であるか否か被験者には知らせない状態での実験であるにも関わらず，全体的に同一クラスタの類似度は異なるクラスタの類似度に比べ高いという結果を得ることができた．

表 2　同一クラスタの楽曲における類似度

	5(similar)	4	3	2	1(different)
Pair 1	13	2	0	0	0
Pair 2	0	5	3	3	4
Pair 3	3	2	3	5	2
Pair 4	1	3	5	5	1

表 3　異なるクラスタの楽曲における類似度

	5(similar)	4	3	2	1(different)
Pair 5	0	1	3	4	6
Pair 6	0	0	0	11	4
Pair 7	0	1	1	5	8
Pair 8	0	3	2	3	7

4.2 楽曲と画像の印象適合度

続いて被験者に楽曲を視聴させて，その楽曲の印象画像を見て，楽曲と画像の印象が適しているか 5 段階で評価させた．その結果を表 4 に示す．いくつかの印象画像(Tune 1, 6, 7)については良い評価を得たが，その一方では低い評価値を得た印象画像(Tune 3, 4, 8, 9, 12)もある．その検証については，次節でコメントも踏まえて論じる．

4.3 検証結果

ユーザテストのフリーコメントを紹介する．4.1 節と 4.2 節で紹介した結果それぞれについて，被験者から自由に意見を述べてもらった．以下に代表的なものを紹介する．

- 印象画像の印象は色で大きく左右される．
- グラデーションに用いる色が薄い色だと，グラデーションを認識しにくい．
- 被験者によっては，楽曲の色を社会的背景色で関連付けしている人がいる．（例えば，ロックであれば白黒等）

これらの意見は研究を進める上で重要なポイントである．今後色の印象や配色バランスについて検討を進めていく．

また，今後どのようなユーザインタフェースとして発展することを期待するかと被

験者に求めたところ，雰囲気に従って楽曲を検索しても，興味のある楽曲が見つかった時や実際に再生する際には，画像の示す楽曲のメタデータを表示する機能があると便利との回答を得た. 今後これらをポップアップ表示することを検討中である. 他に，拡大表示するとどの部分を拡大したのか分かりにくくなるという意見も寄せられた. そこで CAT に対して，表示領域全体を小さく示し，そのうちの表示部分を枠で囲む，といったナビゲーションウィンドウの機能を追加することで, この問題も解決したい.

表 4　楽曲と画像の適合度

	1(suitable)	2	3	4	5(unsuitable)
Tune 1	1	11	2	1	0
Tune 2	1	3	7	4	0
Tune 3	0	0	4	7	4
Tune 4	2	3	2	7	1
Tune 5	3	5	6	1	0
Tune 6	9	4	2	0	0
Tune 7	4	9	0	2	0
Tune 8	0	3	4	2	6
Tune 9	0	6	4	3	2
Tune 10	1	4	8	2	0
Tune 11	3	4	5	3	0
Tune 12	1	4	4	5	1

5. まとめと今後の課題

本報告では，楽曲を印象に基づいて直感的に認識するために，楽曲の特徴量から印象画像を自動生成して，クラスタリングした楽曲に対して一覧表示する手法 MusCat を提案した．以下に，本研究の今後の課題を示す.

楽曲特徴量抽出の改善： 現時点では，ランダムに採取した 5 秒間の特徴その楽曲の特徴量としている．今後は，一つの楽曲全体を 5 秒ごとに分割し，分割したそれぞれから特徴量を検出し，最も適した特徴量をその楽曲の特徴量として扱いたい.

クラスタリングの改善： 本手法は 11 個の特徴量を全て同等に扱ってクラスタリングしている．その結果として，ユーザ自身の主観による類似性に基づく分類結果とは差異が生じる．この差異を埋めるために今後は，楽曲の主観的分類への各特徴量の寄与を評価し，その結果に基づいて特徴量を重み付けしてクラスタリングに活用したい.

抽象画像の再デザイン： 現時点で我々は，楽曲の特徴を表す標識的役割を担ったデザインを提案している．そして経験的に我々は，デザインよりも色の方を先に認識しやすく，また 3 色の割り当てによっても印象が大きく異なる傾向を感じた．今後はそれらを更に深く検討し，より楽曲の印象を表すためのデザインを提案したい．また，より楽曲の印象を表すためのデザインを提案したい.

ユーザインタフェースの拡張： 現時点の我々の MusCat の実装では，画像のクリック操作によって楽曲を再生することが可能である．今後はこれに加えて，代表画像をダブルクリックすることでそのクラスタに含まれる楽曲をループ再生する機能や，再生中の楽曲についてメタデータをポップアップ表示する機能を付与したい.

謝辞　ユーザテストに協力頂いた諸氏に，謹んで感謝の意を表する．本報告の実験には，RWC 研究用音楽データベースに収録された楽曲を使用させていただいた.

参考文献

1)　Gomi, A. Miyazaki, R. Itoh, T. and Li, L.: CAT: A Hierarchical Image Browser Using a Rectangle Packing Technique, 12th International Conference on Information Visualization, pp.82-87 (2008).
2)　大塚玲朗，梶川嘉延，野村康雄: PCM データに対応した感性語による音楽データベース探索システムに関する研究，第 14 回電子情報通信学会データ工学ワークショップ(DEWS2003), 8-p-5 (2003).
3)　山脇一宏，椎塚久雄: コレスポンデンス分析による楽曲の特徴認識，感性工学研究論文集, vol.7, no.4,pp.659-663 (2008).
4)　富岡弘志，三木光範，廣安知之: 色空間と感性の反映方法: ISDL Report, no.20040621002 (2004).
5)　小林重順: カラーシステム，日本カラーデザイン研究所（編）， 講談社(2001).
6)　小田瑞穂，伊藤貴之: MIST: 音楽アイコンの自動選択の一手法，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(WISS)，pp.115-116 (2007).
7)　Kolhoff, P. Preub, J. and Loviscach, J.: Music Icons: Procedural Glyphs for Audio Files, IEEE SIBGRAPI, pp. 289-296 (2006).
8)　L. Lie, D. Liu, and H. Zhang: Automatic mood detection and tracking of music audio signals, IEEE Transactions on Audio, Speech, and Language Processing, Vol.14, pp. 5-18 (2006).
9)　後藤孝行，後藤真孝: Musicream: 楽曲を流してくっつけて並べることのできる新たな音楽再生インタフェース，第 12 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 (WISS2004)論文集, pp.53-58 (2004).
10)　P. Lamere, D. Eck: Using 3D Visualizations to explore and discover music, Proceedings of the 8th International Society for Music Information Retrieval, pp.173-174 (2007).
11)　O. Lartillot: “MIRtoolbox,” http://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox
12)　本江邦夫: すぐわかる画家別抽象絵画の見かた，東京美術, pp.24-33(2005).
13)　Harrison, J.: 共感覚―もっとも奇妙な知覚世界，新曜社 (2006).
14)　Leborg, C.: Visual Grammar: デザインの文法，ビー・エヌ・エヌ新社 (2007).