社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2008－MUS－76 (11)
2008／8／7

MusicThumbnailer: 音響的特徴に基づく楽曲のサムネイル画像生成手法

吉井和佳　後藤真孝

産業技術総合研究所
{k.yoshii,m.goto}@aist.go.jp

本稿では，音楽音響信号から抽出した特徴量に基づき，楽曲をサムネイル画像に変換する手法MusicThumbnailerについて述べる．生成されたサムネイル画像は音響信号を試聴することなくその音楽内容を推測するのに役立つ．我々は，対象となる楽曲群の性質に合わせてサムネイル画像群を自動的に最適化することを試みた．そのため，アドホックな変換規則を設計する代わりに，サムネイル画像の記憶性，表現力，区別性の三つの評価基準に基づくコストが最小化となるサムネイル画像群を生成する手法を開発した．実験の結果，得られたサムネイル画像は楽曲の音楽内容に関する有用なヒントを与えることができることが分かった．

MusicThumbnailer: A Visualization Method for Generating Thumbnail Images of Musical Pieces based on Acoustic Features

Kazuyoshi Yoshii and Masataka Goto

National Institute of Advanced Industrial Science and Technology (AIST)

This paper presents a novel method called *MusicThumbnailer* to transform musical pieces into visual thumbnail images based on acoustic features extracted from their audio signals. These thumbnails can help users guess the musical contents of audio signals without trial listening. We aimed to automatically optimize thumbnails according to the characteristics of a target music collection. To achieve this, our method generates a set of thumbnails that minimizes cost based on three criteria regarding memorability, representation capability, and distinguishability of thumbnails, instead of designing ad hoc transformation rules. Experimental results indicate that generated thumbnails can provide users with useful hints as to the musical contents of musical pieces.

1. はじめに

オンライン音楽配信サービスにおいて，音楽推薦システムはユーザが大規模音楽データベースから好みに合う楽曲を発見するのに役立っている．例えば，協調フィルタリングに基づく推薦システムは，他のユーザがどのような評価を行ったかを参考にして楽曲を推薦する[1),2)]．内容に基づくフィルタリングシステムでは，音楽内容（音響的特徴量）の点で，ユーザの好みの楽曲と類似した楽曲を推薦する[3),4)]．最近では，それら二つの技術を統合するようなハイブリッド型フィルタリングシステムもいくつか提案されている[5),6)]．

このような推薦・検索システムに共通する本質的な問題は，楽曲が提示されてもユーザはそれらの音楽内容をただちに把握することができないことである．しかし，この問題はこれまで扱われてこなかった．ユーザは好みに合う楽曲を聴き続けようと思っても，提示された楽曲を逐一試聴しなければ，それらが聴くに値するかどうか判定できない．そのため，しばしば好みに合わない楽曲も聴くことになる．さらに悪いことに，音響信号は時間的なメディアであるので試聴には時間がかかる．一方，画像は空間的なメディアであり，簡単に全体を見渡すことができる．

この問題を解決するため，本稿では個々の楽曲の音響信号に対応するコンパクトなサムネイル画像を生成する手法MusicThumbnailerを提案する．本手法は，ユーザが楽曲を視聴せずにその音楽内容を推測するための手助けをする．例えば，**図1**に示すように推薦システムと組み合わせると効果的である．最初のうちは，ユーザは推薦された楽曲を実際に聴く際に，それらに付与されたコンパクトなサムネイル画像を流し見するだけである．そのような経験が蓄積されるうち，ユーザは無意識的にサムネイル画像の視覚的特徴と好みの楽曲とを結びつける．この結果，サムネイル画像が暗示す

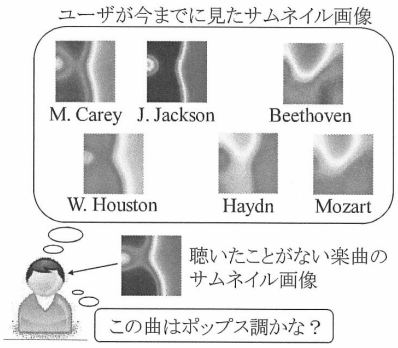

図 1 想定されるシナリオ: ユーザは時間のかかる試聴を行わなくても未知楽曲の音楽内容をサムネイル画像から推測できる.

る音楽的な意味を理解できるようになる．最後には，ユーザは耳ではなく目を使って，好みに合う音楽音響信号を選べるようになることが期待できる．

我々の手法の利点は，与えられた楽曲群に対してサムネイル画像の視覚的特徴（色やパターン）を自動的に最適化できることである．これを達成しようとすると，音響的特徴と視覚的特徴とを恣意的に結びつけるアドホックな変換規則を設計することはできない．なぜならば，そのような規則は，異なる性質の音楽群に対して一貫した正当性を持ちえないからである．本稿では，ユーザビリティの観点から，生成されたサムネイル画像に対して複数のトップダウンな評価基準を定義する．すなわち，これらの基準は音楽群の性質とは独立である．次に，これらの基準を統合し，ひとつのコスト関数として数学的に表現する．そのコスト関数を最小化することで，音響的特徴と視覚的特徴の関連付けが自己組織化できる．

本稿の構成は以下の通りである．まず，2 章で関連研究と本研究の位置づけについて述べる．3 章で提案手法 MusicThumbnailer を説明する．4 章で RWC 音楽データベース[7] を用いた評価実験について報告する．5 章でまとめとする．

2. 関 連 研 究

大量の楽曲群（音楽コレクション）の分布表現，構造化，ブラウジングなどを目的として，これまで多くの可視化手法が提案されてきた[8)-16)]．それらの手法は典型的には，二次元か三次元空間内に，類似した音楽内容（音響的特徴）をもつ楽曲が互いに近くになるよう配置するものである．これにより，音響的特徴の類似性が空間的な距離として視覚的に観測できるため，ユーザは容易に楽曲間の関係を把握することができる．数学的な観点からは，この種の可視化はなんらかの評価基準に従って高次元特徴ベクトルを情報圧縮して低次元に埋め込むことであると解釈できる．例えば，自己組織化マップ (SOM) はこの目的でよく利用され，“Islands of Music”[8] などの研究例がある．

一方，音楽再生インタフェース Musicream[17] では，アドホックな規則を用いて個々の楽曲を色つきの円盤として可視化を行っていた．円盤の色相 (H) と彩度 (S) は，HSV 色環における中心からの角度と距離として定まる．各楽曲の特徴量ベクトルを主成分分析で二次元ベクトルに圧縮し，そのベクトルを極座標形式で見たときの偏角と動径を色相と彩度に割り当てていた．Web ラジオサービス Musicovery[18] では，特定のジャンルと色とが恣意的な規則で関連付けられている．

本研究では，アドホックな規則を用いずに，音楽群ではなく個々の楽曲をコンパクトなサムネイル画像として可視化することを目的とする．一般に，画像は大量の画素のカラー値を保持する高次元ベクトルとして表現される．そのため，我々の目的は，低次元の音響空間（数十次元）を高次元の視覚空間（数千次元）へ適切に変換することである．この問題の難しさは，そのような写像を設計する際の自由度が極めて高いことにある．このような不良設定問題を最適化手法を用いて解くためには，マッピングの適切さに関するいくつかの評価基準を導入する必要がある．

3. 提 案 手 法

本章では，音響的特徴に基づいて楽曲のサムネイル画像を生成する手法について述べる．

3.1 問 題 設 定

楽曲の集合が入力として与えられたとき，それらのもつ音響的特徴を反映するようなサムネイル画像を出力したい．我々はまず，生成されるサムネイル画像の適切さを評価するためいくつかの評価基準を準備する（後述）．本稿では，将来的にフルカラーのサムネイル画像を生成するための第一段階として，グレースケールのサムネイル画像を生成することを目的とする．すなわち，各サムネイル画像に含まれる画素の明度（ブライトネス）のみを扱えばよいことを意味する．以降，単純に画素値と呼ぶことにする．

まず，本稿で用いる定数や変数の記号を**表 1** に定義しておく．N は楽曲数であり，$n\,(1 \le n \le N)$ を楽曲のインデックスとする．S は音響的特徴量の次元数であり，T は生成されるサムネイル画像の画素数である．ここで，W, H をそれぞれサムネイル画像の横と縦の画素数とすると，$T = WH$ が成り立つ．

入力データは $\boldsymbol{X} = [\boldsymbol{x}_1 \boldsymbol{x}_2 \cdots \boldsymbol{x}_N]$ で与えられ，全楽曲の特徴量ベクトルを並べたものである．楽曲 n の特徴量ベクトルは $\boldsymbol{x}_n = (x_{n,1}, \cdots, x_{n,S})^T$ と表現でき，$s\,(1 \le s \le S)$ は特徴量のインデックスである．出力データは $\boldsymbol{Y} = [\boldsymbol{y}_1 \boldsymbol{y}_2 \cdots \boldsymbol{y}_N]$ で表わされ，生成されるサムネイル画像の画素値ベクトルを並べたものである．楽曲 n のサムネイル画像は $\boldsymbol{y}_n = (y_{n,1,1}, \cdots, y_{n,1,H}, \cdots, y_{n,W,1}, \cdots, y_{n,W,H})^T$ ち表現でき，$w, h\,(1 \le w \le W, 1 \le h \le H)$ は横と縦のインデックスである．$y_{n,w,h}$ は明度を表す画素値であるので，0 から 1 の範囲を取るものとする．我々は，$\boldsymbol{X}$ を $\boldsymbol{Y}$ に変換する問題に取り組む．

表 1 本稿で用いる記号一覧

N	楽曲数
S	音響的特徴量の次元数
T	サムネイル画像の画素数（$T = WH$）
W, H	サムネイル画像の横と縦の画素数
$\boldsymbol{x}_n$	楽曲 n の音響信号から得られた特徴量ベクトル $\boldsymbol{x}_n = (x_{n,1}, \cdots, x_{n,S})^T$
$\boldsymbol{y}_n$	楽曲 n のサムネイル画像の画素値ベクトル $\boldsymbol{y}_n = (y_{n,1,1}, \cdots, y_{n,1,H}, \cdots, y_{n,W,1}, \cdots, y_{n,W,H})^T$
X	特徴量ベクトルの集合（入力） $X = [\boldsymbol{x}_1 \boldsymbol{x}_2 \cdots \boldsymbol{x}_N]$
Y	画素値ベクトルの集合（出力） $Y = [\boldsymbol{y}_1 \boldsymbol{y}_2 \cdots \boldsymbol{y}_N]$

3.2 トップダウンな評価基準

生成されるサムネイル画像の適切さを評価するために，我々はユーザビリティの観点から以下の三つの評価基準を導入する．

(1) 記憶性：各サムネイル画像はユーザが簡単に記憶できるものでなければならない．**図 2** に示すように，視覚的なパターンは記憶のしやすさに影響を与える重要な要因である．これは，ユーザが記憶に基づいて視覚的パターンから音楽的な意味を推測できるかどうかと密接な関係がある．我々は，グラデーション画像がこの目的に適していると考える．

(2) 表現力：各サムネイル画像はできるだけ多くの情報をユーザに伝達できることが望ましい．例えば，**図 3** のように，画素値の分散が大きくすることが考えられる．これにより，ユーザは音楽内容に関する詳細な情報を得ることができる．

(3) 区別性：異なる楽曲のサムネイル画像は容易に区別できなければならない．これを達成するには，**図 4** に示すように，各サムネイル画像がその音楽内容を反映する特徴的な視覚的パターンを持っている必要がある．これにより，ユーザは効率的にお目当ての楽曲を探すことができる．

上記の評価基準は，音楽の特定の性質にもサムネイル画像の特定の色にも言及していないことに注意されたい．

このような評価基準は，サムネイル画像をアドホックではないやり方でデザインするために利用できる．一般的には，システムデザイナが音響信号の音楽内容と画像の特定の色とを直接関連付けるアドホックな規則を定義する．例えば，ロックは赤，ジャズは緑，クラシックは青といった具合である．しかし，このような恣意的な変換規則の正当性は保証できない．対照的に，我々が提案する評価基準は，変換の適切さを評価できる．サムネイル画像の実際の色やパターンは，評価基準をもっともよく満たすように自己組織的に最適化される．したがって，我々のアプローチには方法論的かつ数学的な妥当性がある．

以降，これらの評価基準の数学的定式化について述べる．ただし本稿で提案するのはシンプルな実装であり，さまざまな改良の余地が残されている．

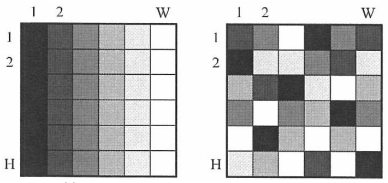

図 2 サムネイル画像の記憶のしやすさの違い（二つのサムネイル画像の画素値ヒストグラムは同一）

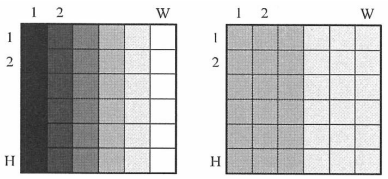

図 3 サムネイル画像の表現力の違い

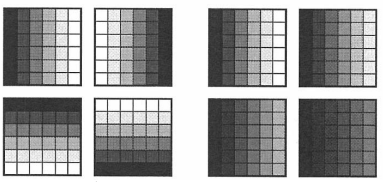

図 4 異なる楽曲のサムネイル画像の区別のしやすさの違い

3.3 数学的定式化

我々の目的を数学的な観点から捉え直すと，S 次元空間を T 次元空間に変換する写像モデルを仮定し，三つの評価基準に基づくコスト関数を最小化するように最適なパラメータを求めることである．このような変換を設計する上では極めて高い自由度がある．なぜなら，より高次の空間はもとの空間の完全な情報を容易に保存できてしまうからである．変換の自由度を効果的に拘束するためのひとつの方法は，写像モデルを線形写像に基づいて設計することである．これにより，音響空間で互いに近くにある楽曲は，変換後の空間でも近い位置に配置されることも期待できる．これは，図 1 に示したシナリオを実現するのに適している．

本稿では，写像モデルを以下で定義する．

$$Y = Sig(AX), \tag{1}$$

ここで，A は $T \times S$ の変換行列である．

$$\begin{bmatrix} \boldsymbol{A}_{1,1}^T \\ \vdots \\ \boldsymbol{A}_{1,H}^T \\ \vdots \\ \boldsymbol{A}_{W,1}^T \\ \vdots \\ \boldsymbol{A}_{W,H}^T \end{bmatrix} = \begin{bmatrix} A_{1,1,1} & A_{1,1,2} & \cdots & A_{1,1,S} \\ \vdots & & & \vdots \\ A_{1,H,1} & A_{1,H,2} & \cdots & A_{1,H,S} \\ \vdots & & & \vdots \\ A_{W,1,1} & A_{W,1,2} & \cdots & A_{W,1,S} \\ \vdots & & & \vdots \\ A_{W,H,1} & A_{W,H,2} & \cdots & A_{W,H,S} \end{bmatrix} \tag{2}$$

また，Sig はシグモイド関数である．

$$Sig(x) = \frac{1}{1+e^{-x}} \quad (-\infty < x < \infty) \tag{3}$$

式 (1) において，出力 Y は画素値そのものであり，0 から 1 の範囲でなければならないため，線形写像の結果 AX に対してシグモイド関数を適用している．ここで，シグモイド関数の導関数は以下で与えられる．

$$Sig'(x) = Sig(x)\,(1 - Sig(x)) \tag{4}$$

次に，写像モデルの適切さを評価するため，コスト関数 C を以下の通り定義する．

$$C = C_s + \alpha_w C_w + \alpha_b C_b, \tag{5}$$

ここで，C_s，C_w および C_b は 3.2 節で述べた三つの評価基準にそれぞれ対応するコストである．α_w と α_b は重み係数であり，どれを重視するかを考慮して事前に与えるものとする．本研究では，トータルコストが各コストの重み付き線形和で表現できると仮定した．

3.3.1 隣り合う画素値の差の最小化

グラデーション画像を生成するため，隣り合う画素の画素値が互いに近くなければならないという必要条件に着目する．これを数学的に実装する方法のひとつは，各サムネイル画像について隣り合う画素の画素値の差を最小化するようにすることである．しかし，この計算はすべてのサムネイル画像に対して行なわなければいけないため効率的ではない．また，変換の自由度をより強く拘束するため，我々は変換行列 A に対して直接コスト関数 C_S を定義する．

$$C_s = \frac{1}{WHS} \sum_{w,h,s} D_{w,h,s} \tag{6}$$

ここで，$D_{w,h,s}$ は s を固定した時の $A_{w,h,s}$ を包囲する八つの係数との距離の平均である．

$$D_{w,h,s} = \frac{1}{8} \sum_{i,j=\pm 1} \left(A_{w,h,s} - A_{w+i,h+j,s}\right)^2 \tag{7}$$

$$+ \frac{1}{8} \sum_{i=\pm 1} \left(A_{w,h,s} - A_{w+i,h,s}\right)^2 \tag{8}$$

$$+ \frac{1}{8} \sum_{j=\pm 1} \left(A_{w,h,s} - A_{w,h+j,s}\right)^2 \tag{9}$$

3.3.2 画素値のサムネイル内分散の最大化

各サムネイル画像の表現力を増加させるには，サムネイル内における画素値の分散を最大化させればよい．この条件を各サムネイルに対して定式化し，全サムネイル画像に対する平均を符号反転したものをコスト関数 C_w とする．

$$C_w = -\frac{1}{N} \sum_n \frac{1}{WH} \sum_{w,h} (y_{n,w,h} - \bar{y}_n)^2 \tag{10}$$

ここで，$\bar{y}_n$ は楽曲 n のサムネイル画像における画素値の平均であり，以下で計算できる．

$$\bar{y}_n = \frac{1}{WH} \sum_{w,h} y_{n,w,h} \tag{11}$$

3.3.3 画素値のサムネイル間分散の最大化

生成されたサムネイル画像が明確に区別できるためには，それらの画素値ベクトルが互いによく分離していなければならない．この条件を満たす単純な方法は，全サムネイル画像間の画素値の分散を最大化することである．ここで，分散は画素内の位置 (w,h) ごとに計算し，それらを画像平面全体で平均して符号反転したものをコスト関数 C_b とする．

$$C_b = -\frac{1}{WH} \sum_{w,h} \frac{1}{N} \sum_n (y_{n,w,h} - \bar{y}_{w,h})^2 \tag{12}$$

ここで，$\bar{y}_{w,h}$ は，位置 (w,h) にある画素値の平均であり，以下で計算できる．

$$\bar{y}_{w,h} = \frac{1}{N} \sum_n y_{n,w,h} \tag{13}$$

3.4 パラメータの最適化

コスト関数 C を最小化するため，我々は最急降下法を利用して反復的にパラメータを更新していくことにする．このとき，更新式は以下で与えられる．

$$A_{w,h,s} \leftarrow A_{w,h,s} - \eta \frac{\partial C}{\partial A_{w,h,s}} \tag{14}$$

ここで，$\eta\ (0 < \eta < 1)$ は学習係数であり，$\frac{\partial C}{\partial A_{w,h,s}}$ は以下の三つの項に分解できる，

$$\frac{\partial C}{\partial A_{w,h,s}} = \frac{\partial C_s}{\partial A_{w,h,s}} + \alpha_w \frac{\partial C_w}{\partial A_{w,h,s}} + \alpha_b \frac{\partial C_b}{\partial A_{w,h,s}} \tag{15}$$

3.4.1 更新式の導出

本節では，更新式の導出について説明する．式 (15) の第一項は以下の通り計算できる．

$$\frac{\partial C_s}{\partial A_{w,h,s}} = \frac{2}{WHS} \left(A_{w,h,s} - \bar{A}_{w,h,s}\right) \tag{16}$$

ここで，$\bar{A}_{w,h,s}$ は $A_{w,h,s}$ の近傍の値の平均である．

$$\bar{A}_{w,h,s} = \frac{A_{w,h\pm 1,s} + A_{w\pm 1,h,s} + A_{w\pm 1,h\pm 1,s}}{8} \tag{17}$$

第二項および第三項は，さらに以下の通り分解できる．

$$\frac{\partial C_w}{\partial A_{w,h,s}} = \frac{\partial C_w}{\partial y_{n,w,h}} \cdot \frac{\partial y_{n,w,h}}{\partial A_{w,h,s}} \tag{18}$$

$$\frac{\partial C_b}{\partial A_{w,h,s}} = \frac{\partial C_b}{\partial y_{n,w,h}} \cdot \frac{\partial y_{n,w,h}}{\partial A_{w,h,s}} \tag{19}$$

ここで，各項は以下の通り求めることができる．

$$\frac{\partial C_w}{\partial y_{n,w,h}} = -\frac{2}{NWH} (y_{n,w,h} - \bar{y}_n) \tag{20}$$

$$\frac{\partial C_b}{\partial y_{n,w,h}} = -\frac{2}{NWH} (y_{n,w,h} - \bar{y}_{w,h}) \tag{21}$$

$$\frac{\partial y_{n,w,h}}{\partial A_{w,h,s}} = Sig'(\boldsymbol{A}_{h,w}^T \boldsymbol{x}_n) x_{n,s} \tag{22}$$

$$= Sig(\boldsymbol{A}_{h,w}^T \boldsymbol{x}_n) \left(1 - Sig(\boldsymbol{A}_{h,w}^T \boldsymbol{x}_n)\right) x_{n,s} \tag{23}$$

3.4.2 パラメータ更新による視覚的効果

本節では，式 (15) に含まれる三つの項が，一回のパラメータ更新でサムネイル画像にどのような視覚的効果をもたらすのかについて考察する．式 (16) は，変換行列をそれ自身を平滑化した行列に近づける働きをする．式 (17) では，平滑化した行列を計算するため，周

表 2 実験用楽曲のジャンルとサブジャンル（括弧内は楽曲数）

ポップス	ポップス (3)，バラード (3)
ロック	ロック (3)，ヘヴィメタル (3)
ダンス	ラップ・ヒップホップ (3)，ハウス (3)，テクノ (3)，ファンク (3)，ソウル・R&B(3)
ジャズ	ビッグバンド (3)，モダンジャズ (3)，フュージョン (3)
ラテン	ボサノヴァ(3)，サンバ (3)，レゲエ (3)，タンゴ (3)
クラシック	バロック（管弦楽）(1)，古典派（管弦楽）(1)，ロマン派（管弦楽）(1)，近代（管弦楽）(2)
行進曲	ブラスバンド (3)
クラシック	バロック（器楽）(2)，古典派（器楽・室内楽）(2)，ロマン派（器楽・室内楽）(2)，近代（器楽）(1)
ワールド	ブルース (3)，フォーク (3)，カントリー (3)，ゴスペル (3)，アフリカン (3)，インディアン (3)，フラメンコ (3)
声楽	シャンソン (3)，カンツォーネ (3)
邦楽	演歌 (3)，民謡 (3)，雅楽 (3)
ア・カペラ	ア・カペラ (1)

辺画素値の畳み込みに基づく画像ぼかしアルゴリズムと同様の手法が利用されている．式 (20) と式 (21) は，各画素の画素値をサムネイル内平均およびサムネイル間平均から遠ざける働きをする．これは，各サムネイル画像のダイナミックレンジとサムネイル画像のバラエティを増加させる．このような視覚的効果は我々の直感によく合致している．

4. 評価実験

本章では，MusicThumbnailer の有用性を評価するために行ったサムネイル画像生成実験について報告する．

4.1 実験条件

入力となる楽曲群として，“RWC Music Database: Music Genre” (RWC-MDB-G-2001)[7] に収録されている 100 曲を用いた（$N = 100$）．**表 2** に示す通り，99 曲は 10 のジャンルに分類され，さらに 33 のサブジャンルに細分される．残りの 1 曲はア・カペラである．

音響的特徴量を抽出するため，我々は MARSYAS[19] を利用した．まず，各楽曲から 42 次元の特徴量ベクトルを求めた．その内訳はスペクトル特徴量（セントロイド，ロールオフ，フラックス）およびゼロクロス率の曲全体に渡る平均と分散で 8 次元，13 次元メル周波数ケプストラム係数の平均と分散で 26 次元，ビートの周期性に関するリズム特徴量で 8 次元である．次に，主成分分析を用いて，寄与率 95%で打ち切り 20 次元に圧縮した（$S = 20$）．

サムネイル画像のサイズは横・縦ともに 50 とした（$W = H = 50, T = 2500$）．この実験では美的な観点から，**表 3** で示したカラースキームに従って，グレースケール画像をフルカラー画像に変換した．これは便宜的な処置であり，本来このようなアドホックな規則は利用するのは望ましくない．学習係数 η と重み係数 α_w および α_b は 0.1，0.4，0.4 と実験的に定めた．

表 3 フルカラー画像に変換するためのカラースキーム

画素値	0.00	0.33	0.66	1.00
RGB 値	(0,0,1)	(0,1,1)	(1,1,0)	(1,0,0)

4.2 実験結果

図 5 の実験結果が示すように，生成されたサムネイル画像は音楽内容を推測するための手がかりになることが示せた．ジャンルレベルでは，ポピュラー・ロック，クラシック（管弦楽），行進曲におけるサムネイル画像は，それぞれ互いに類似したものとなった．ダンスカテゴリーでは，各サブジャンルにおいて類似したサムネイル画像が得られた．ここで，ファンクのサムネイル画像は，レゲエやアフリカンのそれと類似している．これは，ファンクがレゲエやアフリカンの特性を取り込みながら発展してきたことを示唆している．ラテン，ワールドの一部（アフリカン，インディアンなど），邦楽においては，各サブジャンルごとに類似したサムネイル画像が得られた．声楽カテゴリーでも同様であったが，データベース中で特に特徴的な模様が見られた．一方，ジャズカテゴリーでは，各サブジャンルにおいて比較的大きなバリエーションがあった．フュージョンでは，サムネイル画像から様々なジャンルの音楽の特性を反映していることが分かる．

5. おわりに

本稿では，楽曲の音楽内容を反映するようなサムネイル画像を自動生成できる，音を画像に変換する手法 MusicThumbnailer について述べた．アドホックな規則を用いないで変換を行うため，我々は記憶性，表現力，区別性に関する三つの評価基準を提案した．これらの評価基準は，ユーザビリティの観点から，生成されたサムネイル画像の適切さを評価するのに利用した．これを数学的な観点で捉え，評価基準に基づくコスト関数の最小化問題として定式化した．

実験結果は手法の有効性を示しているものの，いまだ多くの問題が残されている．まず，音響空間における楽曲間の位置的関係を視覚空間でも保存できるような評価基準を導入する予定である．また，局所解に陥りにくい最適化手法の適用を検討することが望ましい．合わせて正則化手法の導入も必要であると考えられる．このような改良は，被験者実験を通じて検証しながら進めていくことが重要である．

本手法の応用としては，一曲に含まれる音楽内容を動的に表現する視覚効果（画像の時系列）を生成することが考えられる．この場合には，音楽群から得られる特徴量の代わりに，ある楽曲の時系列特徴量を入力とすればよい．このようなビジュアライザは楽曲の構造を把握するために役立つであろう．

謝辞 本研究において有益な議論をして頂いた麻生英樹氏と松坂要佐氏（産業技術総合研究所）に感謝する．

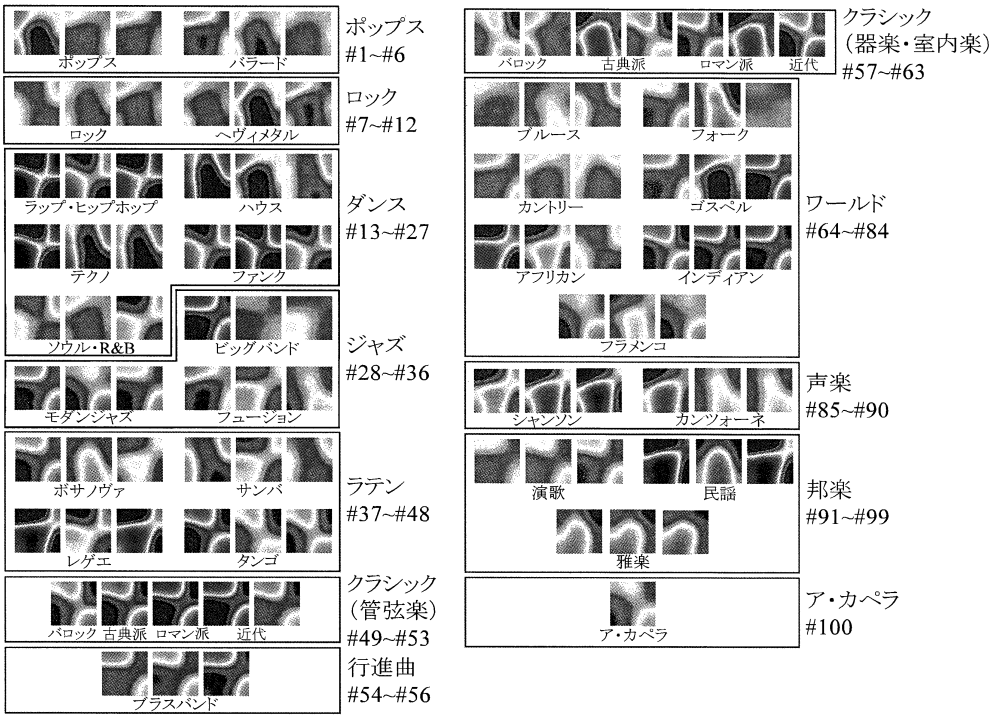

図 5 RWC-MDB-G-2001 に収録された 100 曲を入力として MusicThumbnailer が出力したサムネイル画像

参 考 文 献

1) Shardanand,U. and Maes,P., "Social Information Filtering: Algorithms for Automating "Word of Mouth"," *ACM Conf. on Human Factors in Computer Systems.*, 1995, pp. 210–217.
2) Cohen,W. and Fan,W., "Web-Collaborative Filtering: Recommending Music by Crawling the Web," *Computer Networks*, Vol. 33, No. 1–6, pp. 685–698, 2000.
3) Hoashi, K., Matsumoto, K., and Inoue, N., "Personalization of User Profiles for Content-based Music Retrieval based on Relevance Feedback," *ACM Multimedia*, 2003, pp. 110–119.
4) Logan, B., "Music Recommendation from Song Sets," *ISMIR*, 2004, pp. 425–428.
5) Celma, O., Ramirez, M., and Herrera, P., "Foafing the Music: A Music Recommendation System based on RSS Feeds and User Preferences," *ISMIR*, 2005, pp. 464–457.
6) Yoshii, K., Goto, M., Komatani, K., Ogata, T., and Okuno, H. G., "An Efficient Hybrid Music Recommender System Using an Incrementally-trainable Probabilistic Generative Model," *IEEE Trans. on Audio, Speech and Language Processing*, Vol.16, No.2, pp.435–447, 2008.
7) Goto, M., Hashiguchi, H., Nishimura, T., and Oka, R., "RWC Music Database: Music Genre Database and Musical Instrument Sound Database," *ISMIR*, 2005, pp.229-230.
8) Pampalk, E., Dixon, S., and Widmer, G., "Exploring Music Collections by Browsing Different Views," *ISMIR*, Vol. 28, Mo. 2, pp. 49–62, 2004.
9) Torrens, M., Hertzog, P., and Arcos, J., "Visualizing and Exploring Personal Music Libraries," *ISMIR*, 2004, pp. 421–424.
10) Mörchen, F., Ultsch, A., Nöcker, M., and Stamm, C., "Databionic Visualization of Music Collections According to Perceptual Distances," *ISMIR*, 2005, pp. 396–403.
11) Mayer, R., Dittenbach, M., and Rauber, A., "PlaySOM and PocketSOMPlayer: Alternative Interfaces to Large Music Collections," *ISMIR*, 2005, pp. 618–623.
12) Mayer, R., Lidy, T., Rauber, A., "The Map of Mozart," *ISMIR*, 2006, pp. 351–352.
13) Knees, P., Schedl, M., Pohle, T., and Widmer. G., "An Innovative Three-dimensional User Interface for Exploring Music Collections Enriched with Meta-Information from the Web," *ACM Multimedia*, 2006, pp. 17–24.
14) Leitich, S. and Topf, M., "Globe of Music: Music Library Visualization Using GEOSOM," *ISMIR*, 2007, pp.167–170.
15) Donaldson, J. and Knopke, I., "Music Recommendation Mapping and Interface based on Structual Network Entropy," *ISMIR*, 2007, pp. 181–182.
16) Lamere, P. and Eck, D., "Using 3D Visualizations to Explore and Discover Music," *ISMIR*, 2007, pp. 173–174.
17) Goto, M. and Goto, T., "Musicream: New Music Playback Interface for Streaming, Sticking, Sorting, and Recalling Musical Pieces," *ISMIR*, 2005, pp. 404–411.
18) Musicovery: http://musicovery.com/.
19) Tzanetakis, G. and Cook, P., "MARSYAS: A Framework for Audio Analysis," *Organized Sound*, No. 4, Vol. 3, pp. 169–175, 2000.