

音楽検索のための楽曲印象値の自動付与手法

三好真人^{†1} 柘植 覚^{†2} Choge Kipsang Hillary^{†1}
尾山匡浩^{†3} 伊藤桃代^{†1} 福見 稔^{†1}

本稿では、楽曲に適切な印象値を自動付与する手法を提案する。提案手法は特徴抽出手法及び印象値付与手法から構成される。特徴抽出手法では、楽曲の印象値付与に有効と考えられる音量、音色、リズム、和音に関する特徴量を抽出する。印象値付与手法では、抽出された特徴量を用いてニューラルネットワークにより楽曲に印象値を付与する。401個の楽曲パターンを用いて印象値付与実験及び主観評価実験を行い、提案手法の有効性を検証した。

Automatic Mood Score Detection Method for Music Retrieval

MASATO MIYOSHI,^{†1} SATORU TSUGE,^{†2}
HILLARY KIPSANG CHOGE,^{†1} TADAHIRO OYAMA,^{†3}
MOMOYO ITO^{†1} and MINORU FUKUMI^{†1}

In this paper, we propose an automatic mood score detection method for music retrieval. This method consists of the feature extraction method and the mood score detection method. In the feature extraction method, feature parameters which are related to intensity, timbre, rhythm, and harmony are extracted from audio signals. In the mood score detection method, a score of each music mood is detected using Multi-Layer Neural Networks (MLNNs). In order to evaluate the proposed method, we conducted the mood score detection experiment and the subjective evaluation experiment using 401 music patterns.

^{†1} 徳島大学
The University of Tokushima

^{†2} 大同大学
Daido University

^{†3} 神戸市立工業高等専門学校
Kobe City College of Technology

1. はじめに

近年、インターネットによる音楽配信や大容量のデジタルオーディオプレイヤー・PCなどにより、大量の音楽を個人で保存可能となり、容易に利用することが可能である。これらはデータが大量となると人手による検索や分類が困難になる。そのため、有効な音楽検索・分類手法が要求されている。音楽検索・分類には主に楽曲のジャンル^{(1)–(3)}、アーティスト⁽⁴⁾、印象^{(5)–(11)}が利用されている。その中でも音楽の印象を利用した音楽検索・分類手法は、ユーザが直観的に音楽を検索することが可能であること、また楽曲に関する情報(ジャンル、アーティストなど)を知らなくても楽曲を検索することが可能であるという利点がある。印象による音楽検索・分類を行うためには、楽曲に対して適切な印象または印象値の付与が必要となる。

現在までに音楽の印象分類・印象値付与に関しては多くの研究がなされており、文献5)–7)において印象分類手法、文献8)–11)において印象値付与手法が提案されてきた。しかし、有力な音響特徴量が発見されていないことや印象分類・印象値付与結果に対する主観的な評価が行われないなどの問題があり、精度向上や実環境への応用のためには改善の余地があるといえる。

そこで本稿では、楽曲の適切な印象値を自動付与する手法を提案する。提案手法は特徴抽出手法及び印象値付与手法から構成される。また、提案手法の有効性を検証するために印象値付与実験及び主観評価実験を行った。

本稿は以下の通りに構成される。2節では、提案手法である特徴抽出手法及び印象値付与手法について述べる。3節では、提案手法の有効性を評価するために行った印象値付与実験及び主観評価実験について述べる。4節では、本稿のまとめと今後の課題について述べる。

2. 楽曲の印象値付与手法

本節では楽曲の印象値付与手法について述べる。図1に提案手法の概要を示す。提案手法は特徴抽出手法及び印象値付与手法から構成される。特徴抽出手法及び印象値付与手法について次節以降で説明する。

2.1 特徴抽出手法

特徴抽出手法では、音響信号から音量、音色、リズム、和音に関する特徴量を抽出する。音響特徴量はスライドウィンドウ方式によってフレーム化された音響信号中から抽出される。各音響特徴量に関して以下で説明する。

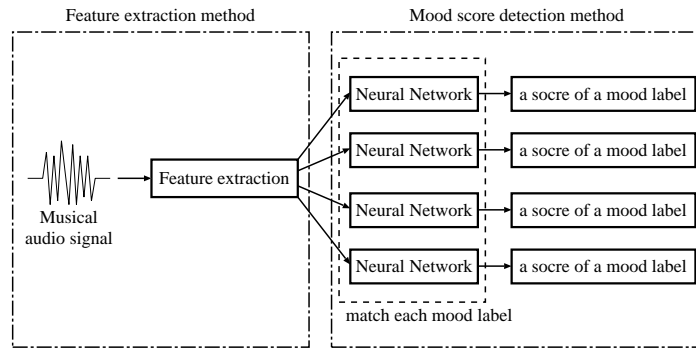


図 1 楽曲印象値付与手法の概要
Fig.1 Overview of the proposed method

2.1.1 音量特徴量

音量を示す特徴量として、以下の 4 種類を用いる。

- フレーム内エネルギー⁸⁾
フレーム内エネルギーは、フレーム化された音響信号の振幅値の二乗を総和し対数化したものであり、賑やかさといった楽曲の印象を特徴づける特徴量である。
- 対数スペクトル和及びその一次回帰係数⁹⁾
対数スペクトル和は、高速フーリエ変換によって得られた対数パワースペクトルの総和であり、周波数領域における音響信号の音量を表現している。また、その一次回帰係数は音響信号の変動を表現している。
- 低エネルギーフレーム数⁹⁾
低エネルギーフレーム数は、フレーム内エネルギーが閾値以下となるフレーム数であり、フレーム内エネルギーが小さいフレームが多い楽曲は穏やかな楽曲であると考えられる。

2.1.2 音色特徴量

音色を示す特徴量として、以下の 6 種類を用いる。

- セントロイド¹⁾
セントロイドは各フレームの周波数を振幅スペクトルで重み付けしたパラメータであり、楽曲の明暗を示す特徴量として用いる。
- 周波数帯域幅⁸⁾
周波数帯域幅はセントロイドと各フレームの周波数の差の二乗をパワースペクトルで重

み付けしたパラメータであり、楽曲のエネルギー分布を表現した特徴量である。楽曲で演奏される楽器数が多ければエネルギー分布が広がり賑やかに感じると考えられる。

- スペクトラルロールオフ¹⁾
スペクトラルロールオフは楽曲のスペクトル分布を表現した特徴量である。スペクトルの分布が低周波域に偏っている楽曲は暗く感じ、高周波域に偏っている楽曲は明るく感じると考えられる。
- 音色の変動特徴
楽曲の音色の変化を表現するためスペクトラルフラックス¹⁾ 及びスペクトルのコサイン類似度⁹⁾ を用いる。
- ゼロクロス¹⁾
ゼロクロスは楽曲中の基本周波数に関係し、音高などの特徴が表現できる。

2.1.3 リズム特徴量

バスドラム、ハイハット等の打楽器音は音響信号中においてインパルス形状の波形として出現し、周波数領域では高周波帯域まで周波数成分が存在する。そこでパワースペクトルの時間変化が楽曲のリズムを表現していると考え、パワースペクトルのピークを検出し、そのピーク数をリズム特徴量として用いる。パワースペクトルのピーク数⁹⁾ は帯域分割、移動平均フィルタリング、差分信号計算及びピーク検出によって求められる。

2.1.4 和音特徴量

楽曲の和音に関する特徴量を抽出するためにクロマベクトル¹²⁾ を求める。クロマベクトルは音響信号における 12 音階の半音ごとのスペクトル成分を表現したものである。求めたクロマベクトルから以下の 3 種類の特徴量を求める。

- クロマベクトルフラックス⁹⁾
音階の変動を特徴量として抽出するために各フレームのクロマベクトル間のユークリッド距離であるクロマベクトルフラックスを用いる。
- メジャーコード成分及びマイナーコード成分⁹⁾
メジャーコードが用いられている楽曲は明るく感じ、マイナーコードが用いられている楽曲は暗く感じる傾向があることから、コードは楽曲の雰囲気を特徴づける要因の 1 つである。そこで 3 和音を対象として、楽曲のメジャーコード及びマイナーコード成分を求め、和音特徴量として用いる。

2.2 印象値付与手法

人間の感性特性は非線形であると考えられることから、非線形問題に関して高い判別能力

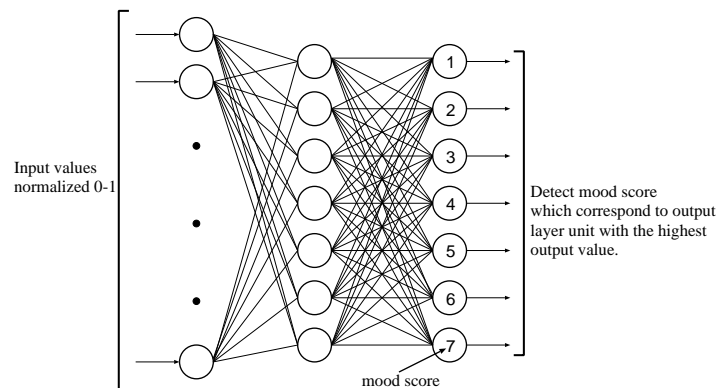


図2 ニューラルネットワークによる印象値付与
Fig.2 Mood score detection using neural networks

を持つニューラルネットワークを印象値付与に用いる。提案手法において、楽曲の印象値はニューラルネットワークの出力層の各ユニットに対応している。これを図2に示す。出力層ユニットの中で最も大きな数値を出力したユニットに対応する印象値を楽曲の印象値として付与する。また、入力値は各特徴軸で0~1の範囲の数値になるように正規化する。ニューラルネットワークは各印象ラベルに対する印象値付与ごとに構築する。

3. 実験

本節では提案手法を評価するために行った印象値付与実験及び主観評価実験について述べる。3.1節で本実験で用いるデータセットについて述べる。3.2節、3.3節で各実験について述べる。

3.1 データセット

本実験では、RWC研究用音楽データベース¹³⁾の「クラシック音楽」、「ジャズ音楽」、「ポップス音楽」及び「著作権切れ音楽」の楽曲225曲から切り出した401個の15秒の楽曲パターンを用いた。各楽曲パターンは、ステレオ、サンプリング周波数44.1kHz、16bit量子化である。印象ラベルにはSD法による楽曲の印象測定に関する先行研究¹⁴⁾において、印象測定に用いられている形容詞の中から「明るさ」、「賑やかさ」、「アップテンポさ」、「軽快さ」の4つを選択した。本実験で使用する楽曲パターンに対する印象値の測定は、被験者が各楽曲パターンを聴取し、それぞれに対して各印象ラベルの印象値を付与することで行っ

た。被験者数は6名、楽曲パターンの印象値は1~7の7段階評価とした。

3.2 印象値付与実験

本節では、提案手法を評価するために行った印象値付与実験について述べる。3.2.1節では、特徴量抽出条件について述べる。3.2.2節、3.2.3節では実験条件及び実験結果について述べる。

3.2.1 特徴量抽出条件

音量特徴量、音色特徴量及びリズム特徴量はフレーム長23.2ms、フレームシフト長11.6msで抽出される。和音特徴量はフレーム長185.8ms、フレームシフト長80msで抽出される。窓関数にはBlackmann窓を用いる。音量特徴量(低エネルギーフレーム数を除く)と音色特徴量(周波数帯域幅は平均値のみ)は音響信号の5秒ごとに平均値と標準偏差、和音特徴量(クロマベクトルフラックスのみ)は音響信号の1秒ごとに平均値を計算し、特徴量として用いる。リズム特徴量は音響信号の5秒ごとに抽出される。特徴ベクトルの次元数は72である。

3.2.2 実験条件

提案手法の学習データには特定ユーザに依存しないようにするために各被験者が付与した印象値のメディアンを使用し、評価データには各被験者が付与した印象値を使用する。実験はクロスバリデーション法により行った。全401個の楽曲パターンの内、1曲から切り出された楽曲パターン^{*1}を評価データ、残りを学習データとし、楽曲数だけ実験を繰り返した。

ニューラルネットワークの入力層、中間層及び出力層ユニット数はそれぞれ72, 7, 7である。最大学習回数は50,000回、学習係数は0.1、慣性項は0.7である。

提案手法の評価尺度には、提案手法によって付与された印象値(付与印象値)と聴取によって測定された印象値(真値)との差を用いる。“Error 0 Rate”を付与印象値と真値との差が0であった楽曲パターンの全楽曲パターン数に対する割合、“Error 1 Rate”を付与印象値と真値との差が±1以内であった楽曲パターンの全楽曲パターン数に対する割合とする。

3.2.3 実験結果

被験者ごとの全印象ラベルの平均精度を表1に示す。実験結果より、全被験者の平均精度として“Error 0 Rate”で29.2%、“Error 1 Rate”で71.9%が得られた。また、各被験者ごとで精度が大きく異なっていることが分かる。例えば、被験者Fの場合に最も高い精度となっており、“Error 0 Rate”で38.2%、“Error 1 Rate”で84.3%となっている。“Error 0

*1 1曲から複数の楽曲パターンを切り出し使用する。

表 1 実験結果

Table 1 Mood score detection experimental results

被験者	A	B	C	D	E	F	平均
Acc.							
Error 0 Rate	29.0%	31.9%	28.5%	34.3%	13.5%	38.2%	29.2%
Error 1 Rate	73.9%	73.7%	73.2%	77.4%	48.9%	84.3%	71.9%

Rate”は38.2%と低いが，“Error 1 Rate”で80%以上の精度が得られていることから，提案手法は適切に印象値を付与できていると考えられる．しかし，被験者 E の場合に最も精度が低く，“Error 0 Rate”で13.5%，“Error 1 Rate”で48.9%となっている．被験者 E において著しく精度が低下した原因として，学習データに用いた楽曲パターンの印象値の分布と被験者 E が付与した楽曲パターンの印象値の分布が大きく異なっていると思われる．被験者 E, F 及び学習データとして用いた楽曲パターンの印象値分布を図 3, 図 4 及び図 5 にそれぞれ示す．これらの図より，被験者 F と学習データの印象値分布は類似しているが，被験者 E と学習データの印象値分布が大きく異なっていることが分かる．従って，提案手法は学習データと類似した印象値分布を持つユーザに対しては有効であるが，異なる印象値分布を持つユーザに対しては適切に印象値を付与できないといえる．

3.3 主観評価実験

本節では，提案手法を実用面から評価するために行った主観評価実験について述べる．3.3.1 節，3.3.2 節で実験条件及び実験結果について述べる．

3.3.1 実験条件

本実験では，図 6 に示す音楽検索システムを用いる．まず本システムにおいて，被験者は各印象ラベルの印象値を入力する．本システムは入力された印象値を持つ楽曲パターンを検索し，3つの楽曲パターンを提示する．3つの楽曲パターンは異なる印象値付与手法によって印象値を付与されている．各楽曲パターンは，

- 提案手法による印象値付与 (“Proposed”)
- ランダム選択による印象値付与 (“Random”)
- 聴取による印象値付与^{*1} (“Ideal”)

によってそれぞれ印象値が付与されている．被験者は各楽曲パターンを聴取し，それぞれに対して1~5の5段階で満足度を決定する．本実験における5段階の満足度はそれぞれ，5:

*1 各被験者が付与した印象値のメディアンを使用することを意味する．

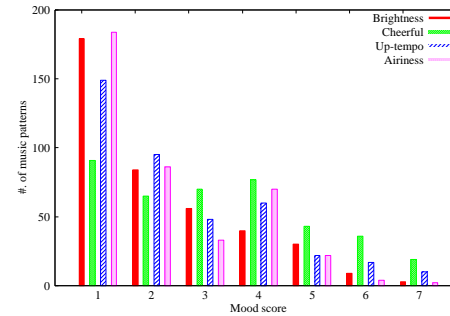


図 3 被験者 E の印象値分布

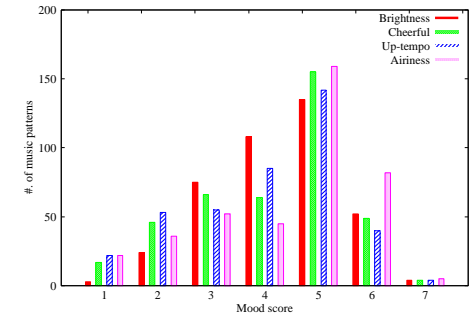


図 4 被験者 F の印象値分布

Fig. 3 Distribution of mood scores in test subject E Fig. 4 Distribution of mood scores in test subject F

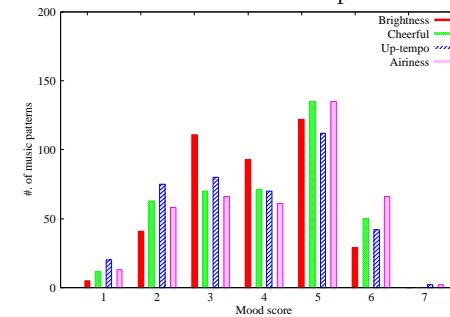


図 5 学習データの印象値分布

Fig. 5 Distribution of mood scores in training data

満足，4:やや満足，3:どちらでもない，2:やや不満足，1:不満足，である．

被験者数は5名であり，各被験者がシステムに入力する印象値の組み合わせパターンは表 2 の12パターンである．表 2 内の“×”に対応する印象ラベルは検索には使用しないことを示している．楽曲パターンはランダムに提示される．

3.3.2 実験結果

各被験者ごとの平均満足度を表 3 に示す．表 3 より，提案手法に対する満足度はランダム選択に対する満足度よりも全ての被験者において高いことが分かる．この結果は，提案手法によって印象付与された楽曲パターンを被験者が要求する楽曲パターンとして検索システムが提示できていることを示していると考えられる．従って，提案手法は有効な印象付与手

主観評価実験

印象値入力							
感性語	弱い←→強い						キャンセル
	1	2	3	4	5	6	
明るさ	○	○	○	○	○	○	○
賑やかさ	○	○	○	○	○	○	○
アップテンポさ	○	○	○	○	○	○	○
軽快さ	○	○	○	○	○	○	○

入力印象値決定

図 6 音楽検索システム
Fig. 6 Retrieval system

表 2 主観評価実験で用いる検索パターン
Table 2 Retrieval patterns in the subjective evaluation experiments

No.	明るさ	賑やかさ	アップテンポさ	軽快さ
1	6	×	×	×
2	2	×	×	×
3	×	6	×	×
4	×	2	×	×
5	×	×	6	×
6	×	×	2	×
7	×	×	×	6
8	×	×	×	2
9	2	2	2	2
10	5	5	5	5
11	3	4	2	2
12	4	5	6	6

表 3 実験結果

Table 3 Subjective evaluation experimental results

Method	被験者					平均
	A	B	C	D	E	
Proposed	3.58	4.00	3.75	4.08	3.75	3.83
Random	2.67	3.08	2.17	2.83	2.50	2.65
Ideal	4.42	4.42	3.83	4.50	4.42	4.32

表 4 各印象値付与手法における満足度分布

Table 4 User satisfaction level distribution against each mood score detection method

Method	満足度				
	5	4	3	2	1
Proposed	28	15	2	9	6
Random	6	14	8	16	16
Ideal	37	12	5	5	1

法であるといえる。また、提案手法に対する満足度が全被験者において3以上であることから、提案手法による印象付与に対して満足な傾向を示しているといえる。しかし、提案手法に対する満足度は聴取による印象値付与に対する満足度よりも全ての被験者において低いことが分かる。この原因として、提案手法が不適切な印象値を付与した楽曲パターンが存在しているためであると考えられる。

この点に関して詳細を調べるために、各印象値付与手法における満足度の分布を表4に示す。表4より、満足度が2以下(不満足であると感じる)の楽曲パターンの割合は、提案手法で25.0%、聴取による印象付与で10.0%、ランダム選択による印象付与で53.3%である。提案手法はランダム選択による印象付与よりは低い割合を示しているが、聴取による印象付与よりは高い割合となっている。つまり、提案手法では不適切な印象値も付与しているといえる。この原因としては、3.2節内の表1で示した通り、“Error 1 Rate”は高いが“Error 0 Rate”が低いために、不適切な印象値を付与していると考えられる。従って、“Error 0 Rate”を改善することが提案手法に対する満足度向上のために必要であるといえる。一方表4より、満足度が4以上(満足であると感じる)の楽曲パターンの割合は、提案手法で71.7%、聴取による付与で81.7%、ランダム選択で33.3%である。不適切な印象値付与のため、提案手法に対する満足度は聴取による印象値付与に対する満足度より低い、70%以上の楽曲パターンを満足であると被験者が評価していることから、有効な印象値付与手法であると考えられる。

3.4 考察

3.2節、3.3節において、提案手法の有効性を評価するために印象値付与実験及び主観評価実験を行った。実験結果より、提案手法に対する満足度を向上させるためには“Error 0 Rate”を改善する必要があるといえる。改善する一つとして、提案手法を個人に適應させることが考えられる。前実験では、提案手法を特定個人に依存しないようにして実験を行ったが、提案手法を特定個人に適應させることで“Error 0 Rate”が向上すると考えられる。本節では、各被験者が付与した印象値を学習データに用いることで提案手法を個人適應させた場合の有効性に関して検討する。

各被験者が付与した印象値のメディアンを学習データに用いるのではなく、各被験者が付与した印象値を学習データに用いて、提案手法を各被験者ごとに適應させる。その他の実験条件に関しては、3.2.2節と同様とした。実験結果を表5に示す。実験結果より、全被験者における平均精度は、“Error 0 Rate”で37.9%、“Error 1 Rate”で78.5%である。表1と比較すると、平均精度において、“Error 0 Rate”で8.7%、“Error 1 Rate”で6.6%向上した。表1で特に精度が悪い被験者Eにおいては、“Error 0 Rate”で28.9%、“Error 1 Rate”で23.1%と大きく向上していることが分かる。

実験結果より、提案手法において被験者が付与した印象値を学習データとして用いることで、精度が向上することを確認した。特に、提案手法を個人非依存とするために用いた学習データと大きく異なる印象値を付与する被験者に対しては有効である。しかし、学習データ

表 5 実験結果
Table 5 Adaptation experimental results

被験者 Acc.	A	B	C	D	E	F	平均
Error 0 Rate	39.5%	36.8%	31.4%	36.3%	42.4%	41.3%	37.9%
Error 1 Rate	83.7%	74.6%	72.3%	81.7%	72.0%	86.5%	78.5%

に被験者が付与する印象値を用いることは、実環境においてユーザが多数の楽曲パターンを聴取し、印象値を付与する必要があることを意味する。これはユーザにとって大きな負担になる。そのため、ユーザが付与した印象値を少数だけ用いて提案手法を個人に適應する手法が望ましいといえる。

4. おわりに

本稿では、楽曲に適切な印象値を自動付与する手法を提案した。提案手法は、楽曲の印象値付与に有効と考えられる音量、音色、リズム、和音特徴量を音響信号から抽出し、抽出された特徴量を用いてニューラルネットワークにより楽曲に印象値を付与する。

提案手法を評価するために印象値付与実験及び主観評価実験を行った。印象値付与実験結果より、全被験者間の平均精度は“Error 0 Rate”で29.2%、“Error 1 Rate”で71.9%であり、被験者 F においては“Error 0 Rate”で38.2%、“Error 1 Rate”で84.3%の最高精度が得られた。さらに主観評価実験結果より、提案手法に対する平均満足度はランダム選択による印象値付与よりも高く、全被験者の場合でも同様の結果が得られた。また、提案手法が印象値を付与した楽曲パターンの内、70%以上の楽曲パターンに対して被験者が満足していることを示した。しかし、聴取による印象値付与よりは低い満足度であり、これは“Error 0 Rate”が低いために提案手法が不適切な印象値を楽曲パターンに付与してしまうためである。そこで提案手法の精度を改善するために個人適應に関して検討を行い、被験者が付与した印象値を学習データとすることで印象値付与精度の向上を確認した。

今後の課題としては、3.4 節において述べたように、ユーザが付与した少数の印象値を用いた適應手法に関して検討していく予定である。また提案手法の精度を向上させるため、印象値付与に有効な音響特徴量に関する検討していく予定である。

謝辞 本研究の一部は文部科学省科学研究費、若手研究 (B)19700172 の補助を受け行った。最後に、楽曲の聴取実験及び主観評価実験にご協力いただいた皆様に感謝致します。

参考文献

- 1) Tzanetakis, G. and Cook, P.: Musical Genre Classification of Audio Signals, *IEEE Transaction on Speech and Audio Processing*, Vol.10, No.5, pp.293-302 (2002).
- 2) Jiang, D.-N., Lu, L., Zhang, H.-J., Tao, J.-H. and Cai, L.-H.: Music Type Classification by Spectral Contrast Feature, *Proc. of ICME2002* (2002).
- 3) McKinney, M.F. and Breebaart, J.: Features for Audio and Music Classification, *Proc. of ISMIR2003*, pp.325-330 (2003).
- 4) 藤原弘将, 北原鉄朗, 後藤真孝, 駒谷和範, 尾形哲也, 奥乃博: 伴奏音抑制と高信頼度フレーム選択に基づく楽曲の歌手名同定手法, *情報処理学会論文誌*, Vol.47, No.6, pp.1831-1843 (2006).
- 5) Trohidis, K., Tsoumakas, G., Kalliris, G. and Vlahavas, I.: Multi-Label Classification of Music Into Emotions, *Proc. of ISMIR2008*, pp.325-330 (2008).
- 6) Liu, D., Lu, L. and Zhang, H.-J.: Automatic Mood Detection from Acoustic Music Data, *Proc. of ISMIR2003* (2003).
- 7) Pohle, T., Pampalk, E. and Widmer, G.: Evaluation of Frequently Used Audio Features for Classification of Music Into Perceptual Categories, *Proc. of ISMIR2005* (2005).
- 8) 平江遼, 西隆司: 感性に基づくクラシック音楽の分類, *日本音響学会誌*, Vol.64, No.10, pp.607-615 (2008).
- 9) Miyoshi, M., Tsuge, S., Choge, H.K., Oyama, T., Ito, M. and Fukumi, M.: Music Impression Detection Method for User Independent Music Retrieval System, *Proc. of KES2010*, pp.612-621 (2010).
- 10) 田上直樹, 西隆司: 時間・周波数特性の画像パターンを用いた音楽の感性分類, *日本音響学会講演論文集*, pp.885-886 (2009).
- 11) 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検索システムに関する研究, *電子情報通信学会 データ工学ワークショップ* (2003).
- 12) 後藤真孝: リアルタイム音楽情景記述システム: サビ区間検出手法, *情報処理学会音楽情報科学研究会 研究報告*, Vol.2002, No.100, pp.27-34 (2002).
- 13) 後藤真孝, 橋口博樹, 西村拓一, 岡隆一: RWC 研究用音楽データベース: 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース, *情報処理学会論文誌*, Vol.45, No.3, pp.728-738 (2004).
- 14) 杉原太郎, 森本一成, 黒川隆夫: SD 法を通してみた音楽に対する感性の基本特性, *電子情報通信学会技術研究報告*, pp.57-63 (2001).