

# Comparison of Music Features for Music Information Retrieval

○<sup>1</sup> 中山 達喜 ,                      <sup>2</sup> 吉田 真一  
○<sup>1</sup>Tatsuyoshi Nakayama,                  <sup>2</sup>Shinichi Yoshida  
<sup>1</sup>高知工科大学 大学院工学研究科 情報システム工学コース  
<sup>1</sup>Graduate School of Engineering , Kochi University of Technology  
<sup>2</sup>高知工科大学 情報学群  
<sup>2</sup>School of Information , Kochi University of Technology

**Abstract:** Music recommendation system using the content based music retrieval plays an important role of multimedia information retrieval. Clustering songs using music features is a basis of music information retrieval. However the parameter of a clustering algorithm and music features have not been studied yet. We investigate the various numbers of music features and compares with humans' sensitivity by subjective experiment for 200 MIDI music. The result will be the foundation for the future songs retrieval systems.

近年、コンピュータの性能の向上とともに、インターネット上にマルチメディア情報が増大している。それは音楽情報も例外ではない。これら多くの音楽の中から、好みの音楽を効率良く探し出すために音楽内容に基づいた楽曲推薦システムがある。本研究では音楽検索を目的として、類似楽曲のクラスタリングを行い、人間の感覚に合った分類となり得るかの検討を行う。これまでも、音楽分類を行う際にどのような特徴量を利用することが有効であるかについて、多くの研究が行われているが [1][2]、まだ体系的には明らかになっていない。そこで本稿では、クラスタリングを行う際にどのような特徴量を利用することがユーザの好みの音楽分類を行うことができるのかを調査する。本稿では 200 曲の MIDI 楽曲から抽出した特徴量を利用し K-means 法により楽曲のクラスタリングを行い、その結果と被験者が分類した結果を比較し、利用した特徴量が音楽分類に有効であるか、またどのような特徴量がクラスタリングにより大きな影響を与えているのかの検討を行う。

本稿では，MIDI データを対象として，文献 [3] の特徴を用いて K-means 法で 8 個のクラスに分類する．

特徴を抽出するデータとして MIDI データを利用する。MIDI データは音高や音長など音楽としての特徴がとりやすいという利点や、カラオケ用データやイン

ターネット上に公開されている個人 / 法人が作成したデータが多くあるなどの特徴がある。また、自動採譜システムなど、音響情報から MIDI データなどの音楽情報を再現する研究も進んでいることから MIDI データから有効な特徴を抽出することが期待されている。このため本研究では MIDI データを用いる。本研究で利用するデータは RWC 研究用データベース [7] の MIDI データを利用する。RWC 研究用データベースの中から、クラシックを 50 曲、ジャズを 50 曲、ポップスを 100 曲の計 200 曲の MIDI データを利用し特徴の抽出、クラスタリングを行う。

本稿で利用する特徴として、文献 [3] で記述している特徴を利用する。文献 [3] では音楽心理学者 Hevner が定義した音楽の構成要素を基に特徴を構成している。Hevner は音楽の構成要素として調（長調 / 短調）、旋律（上昇 / 下降）、音高（高 / 低）、和声（単純 / 複雑）、テンポ、リズム（固定 / 流動）を定義している [4][5][6]。これら Hevner が定義した音楽の構成要素を文献 [3] において表 1 に示す 41 項目で表現している。

また、主成分分析，クラスタ分析によって，音高・音長パターンの特徴量は曲を分類する上で有効である [5] ことが述べられている．そのため，本稿では文献 [1] において利用されている 41 項目の特徴量の中から音高・音長パターンに当たる 17 項目の特徴量をクラスタリングの際の特徴量とする．表 2 に本稿で利用する特徴量を示す．

表 1: 文献 [1] の特徴量

調	C, C+, D, D+, E, F, F+, G, G+, A, A+, B, 長調, 短調
和声	2 和音, 3 和音, 4 和音の各割合, 各音符長平均, 各音符長二乗平均
旋律	上昇音, 下降音, 水平音の各割合, 各音高平均, 各音高二乗平均
音高	音高平均, 音高二乗平均, 主音以上の高音, 主音未満の低音の各割合
リズム	2 分音符の割合, 付点 4 分音符の割合, 4 分音符の割合, 8 分音符の割合
テンポ	テンポ

表 2: 抽出特徴

旋律	上昇音, 下降音, 水平音各割合 上昇音高, 下降音高, 水平音高各平均 上昇音高, 下降音高, 水平音高各二乗平均
音高	音高平均 音高二乗平均 主音以上の高音の割合 主音未満の音の割合
リズム	2 分音符の割合 付点 4 分音符の割合 4 分音符の割合 8 分音符の割合

### 2.3 特徴量の抽出

前節で述べた特徴量の抽出方法について説明する．特徴量は旋律，音高，リズムに大分されるがそれぞれ複数の特徴量を持つ．それぞれの特徴量について式を示す．ただし，

- ・  $n$  番目のメロディの構成音の音高:  $Note(n)$
- ・  $n$  番目のメロディの構成音の音符長 (4 分音符の音符長を 1 とする):  $Len(n)$
- ・ メロディの音符数:  $N$
- ・ 主音の音高:  $Note(0)$

とする．また，式中において

$$\Delta Note(n) = Note(n+1) - Note(n)$$

$$\Delta Len(n) = Len(n+1) - Len(n)$$

$$H_0(k) = \begin{cases} 1 & (k \geq 0) \\ 0 & (k < 0) \end{cases}$$

$$U(k) = \begin{cases} 1 & (k > 0) \\ 0 & (k \leq 0) \end{cases}$$

$$\delta(k) = \begin{cases} 1 & (k = 0) \\ 0 & (k \neq 0) \end{cases}$$

であるとする．

#### 2.3.1 旋律

旋律内での音の変化を表す．

$$\text{上昇音割合} = \frac{\sum_{n=0}^{N-1} H_0(\Delta Note(n))}{N} \quad (1)$$

$$\text{下降音割合} = \frac{\sum_{n=0}^{N-1} (1 - U(\Delta Note(n)))}{N} \quad (2)$$

$$\text{水平音割合} = \frac{\sum_{n=0}^{N-1} \delta(\Delta Note(n))}{N} \quad (3)$$

$$\text{上昇音高平均} = \frac{\sum_{n=0}^{N-1} H_0(\Delta Note(n)) \frac{\Delta Note(n)}{\Delta Len(n)}}{\sum_{n=0}^{N-1} H_0(\Delta Note(n))} \quad (4)$$

$$\text{下降音高平均} = \frac{\sum_{n=0}^{N-1} (1 - U(\Delta Note(n))) \frac{-\Delta Note(n)}{\Delta Len(n)}}{\sum_{n=0}^{N-1} (1 - U(\Delta Note(n)))} \quad (5)$$

$$\text{水平音高平均} = \frac{\sum_{n=0}^{N-1} \delta(\Delta Note(n)) Note(n)}{\sum_{n=0}^{N-1} \delta(\Delta Note(n))} \quad (6)$$

$$\text{上昇音高二乗平均} = \frac{\sum_{n=0}^{N-1} H_0(\Delta Note(n)) \left( \frac{\Delta Note(n)}{\Delta Len(n)} \right)^2}{\sum_{n=0}^{N-1} H_0(\Delta Note(n))} \quad (7)$$

$$\text{下降音高二乗平均} = \frac{\sum_{n=0}^{N-1} (1 - U(\Delta Note(n))) \left( \frac{-\Delta Note(n)}{\Delta Len(n)} \right)^2}{\sum_{n=0}^{N-1} (1 - U(\Delta Note(n)))} \quad (8)$$

$$\text{水平音高二乗平均} = \frac{\sum_{n=0}^{N-1} \delta(\Delta Note(n)) Note(n)^2}{\sum_{n=0}^{N-1} \delta(\Delta Note(n))} \quad (9)$$

### 2.3.2 音高

楽曲全体の音の高さとその分布を表す．

$$\text{音高平均} = \frac{\sum_{n=0}^{N-1} \text{Note}(n)}{N} \quad (10)$$

$$\text{音高二乗平均} = \frac{\sum_{n=0}^{N-1} \text{Note}(n)^2}{N} \quad (11)$$

$$\text{主音以上の音符の割合} = \frac{\sum_{n=0}^{N-1} U(\text{Note}(n) - \text{Note}(0))}{N} \quad (12)$$

$$\text{主音未満の音符の割合} = \frac{\sum_{n=0}^{N-1} (1 - U(\text{Note}(n) - \text{Note}(0)))}{N} \quad (13)$$

### 2.3.3 リズム

楽曲の中に含まれる音符の長さの割合を表している．

$$2 \text{ 分音符の割合} = \frac{\sum_{n=0}^{N-1} \delta(\text{Len}(n) - 2.0)}{N} \quad (14)$$

$$\text{付点 4 分音符の割合} = \frac{\sum_{n=0}^{N-1} \delta(\text{Len}(n) - 1.5)}{N} \quad (15)$$

$$4 \text{ 分音符の割合} = \frac{\sum_{n=0}^{N-1} \delta(\text{Len}(n) - 1.0)}{N} \quad (16)$$

$$8 \text{ 分音符の割合} = \frac{\sum_{n=0}^{N-1} \delta(\text{Len}(n) - 0.5)}{N} \quad (17)$$

## 2.4 K-means 法によるクラスタリング

K-means 法によるクラスタリングと被験者による音楽の分類を行った．今回被験者による実験の結果として 8 クラスタに分類されたため，K-means 法ではクラスタ数を 8 に設定し，クラスタリングを行った．

## 3 楽曲クラスタリングの結果

### 3.1 クラスタリングの結果

抽出した特徴量による K-means 法のクラスタリング結果を示す．クラスタ番号と各楽曲の各特徴量の散布図を用いて，クラスタリング結果を検討する．各グラフの横軸は特徴量，縦軸はクラスタ番号（1～8）である．

K-means 法のクラスタリング結果の散布図からは，複数の特徴量において特徴的な分布が見られる．図 3 では水平音の割合とクラスタの散布図を表している．

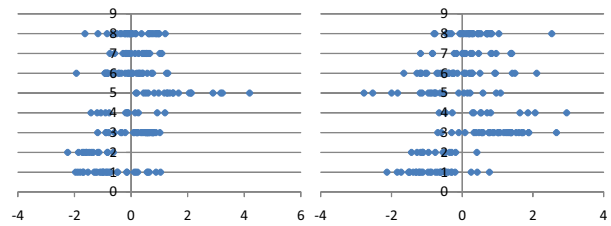


図 1: 上昇音の割合

図 2: 下降音の割合

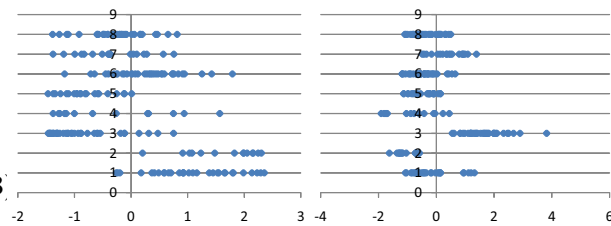


図 3: 水平音の割合

図 4: 上昇音高平均

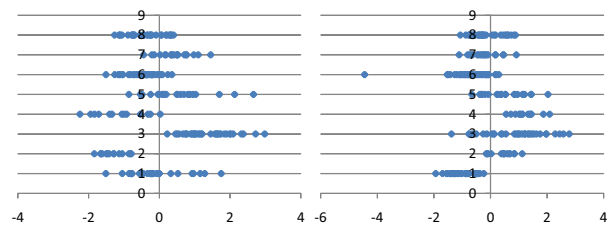


図 5: 下降音高平均

図 6: 水平音高平均

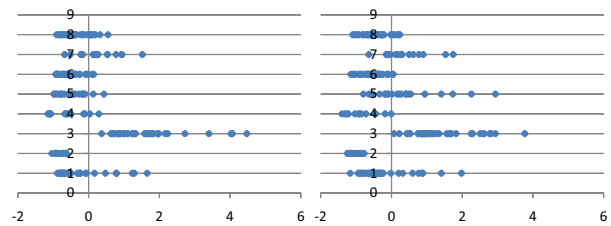


図 7: 上昇音高二乗平均

図 8: 下降音高二乗平均

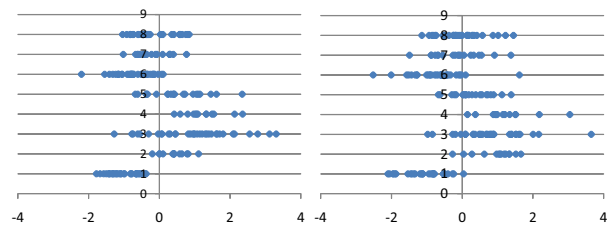


図 9: 水平音高二乗平均

図 10: 音高平均

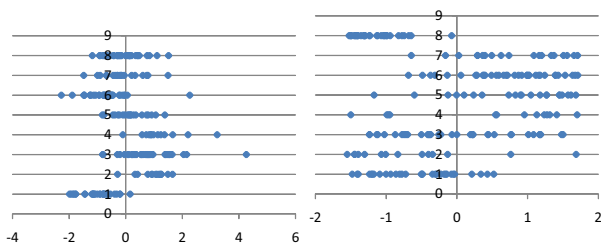


図 11: 音高二乗平均

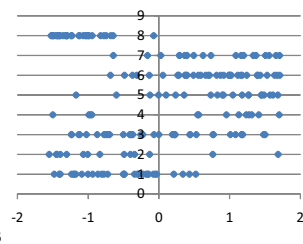


図 12: 主音以上の音符の割合

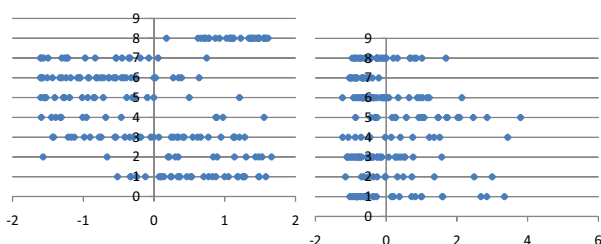


図 13: 主音未満の音符の割合

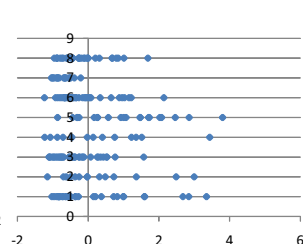


図 14: 2分音符の割合

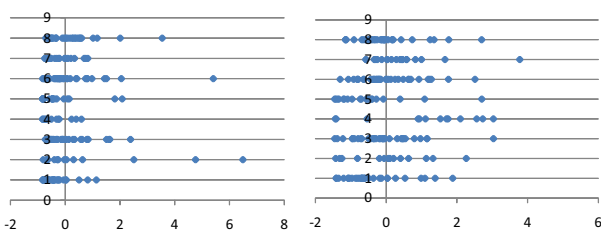


図 15: 符点4分音符の割合

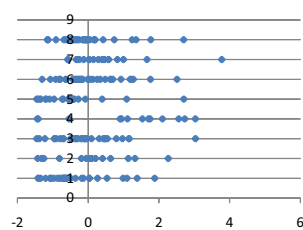


図 16: 4分音符の割合

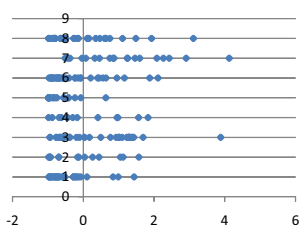


図 17: 8分音符の割合

この特徴量では第1, 第2クラスと第3, 第4, 第5クラスを分類していることが確認できる。図4, 5, 6, 7, 8では, それぞれ上昇音高平均, 下降音高平均, 水平音高平均, 上昇音高二乗平均, 下降音高二乗平均とクラスターの散布図を表している。これらの特徴量は第3クラスと他のクラスを分類する特徴量であることが確認できる。図9では水平音高二乗平均とクラスターの散布図を表している。この特徴量は第1, 第6, 第7, 第8クラスと第2, 第3, 第4, 第5クラスを分類していることが確認できる。これらの特徴で特に第3クラスを分類する特徴が多いため, K-means法によって得られた第3クラスを試聴した結果, 全体として穏やかな曲, 急激な変化を伴わない曲でクラスターが形成されていた。

### 3.2 被験者によるクラスタリング結果

被験者によるクラスタリング結果について示す。クラスター番号と各楽曲の各特徴量の散布図を用いて, クラスタリング結果を検討する。各グラフの横軸は特徴量, 縦軸はクラスター番号(1~8)である。

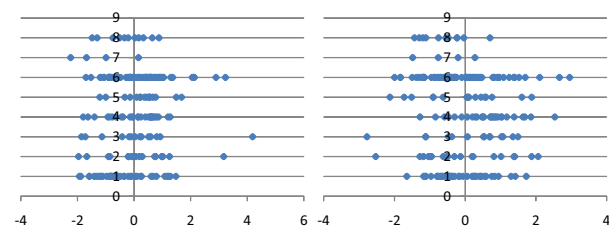


図 18: 上昇音の割合

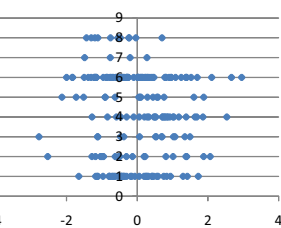


図 19: 下降音の割合

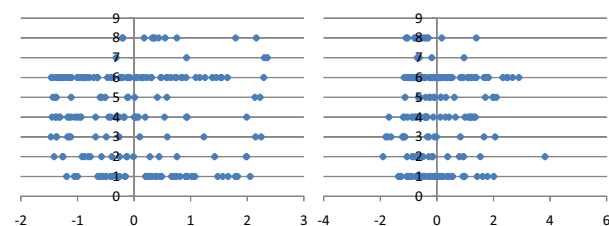


図 20: 水平音の割合

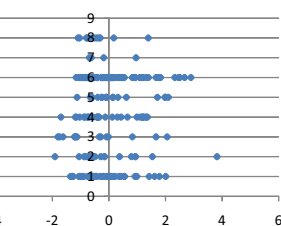


図 21: 上昇音高平均

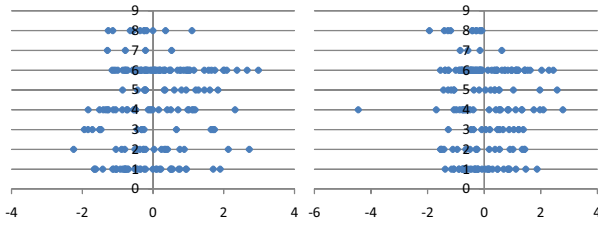


図 22: 下降音高平均

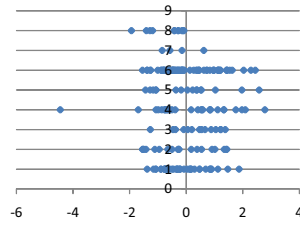


図 23: 水平音高平均

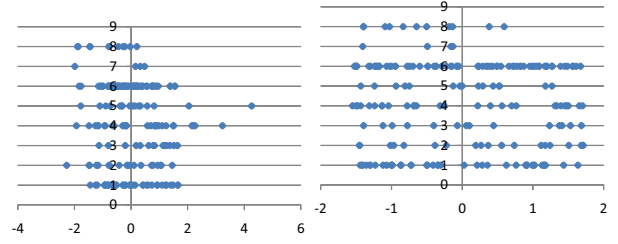


図 28: 音高二乗平均

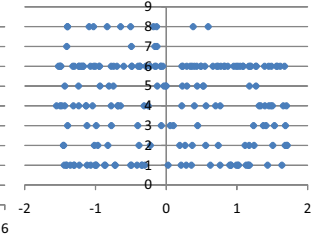


図 29: 主音以上の音符の割合

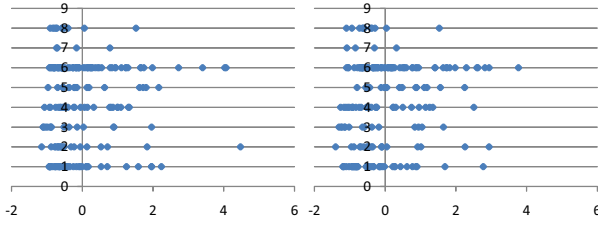


図 24: 上昇音高二乗平均

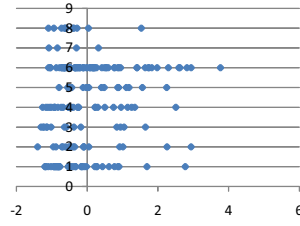


図 25: 下降音高二乗平均

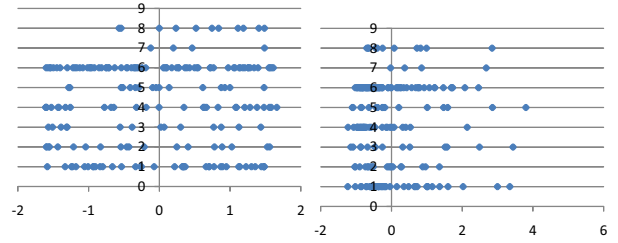


図 30: 主音未満の音符の割合

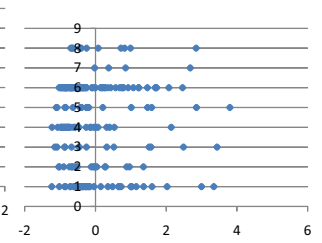


図 31: 2 分音符の割合

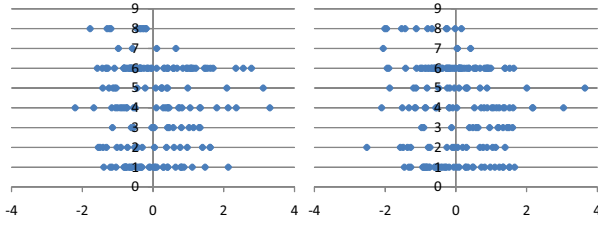


図 26: 水平音高二乗平均

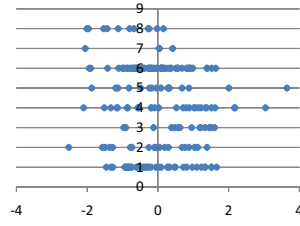


図 27: 音高平均

被験者によるクラスタリングにおいて、図からは各特徴がクラスタの分類に影響を与えていることが確認できない。そのため、今回使用した特徴量が分類に有効であるとはいえない。

### 3.3 クラスタリング結果の評価

K-means 法クラスタリングの結果の評価として F 値を求める。

$$\text{再現率: Recall}_{hk} = \frac{|A_h \cap C_k|}{|A_h|}$$

$$\text{適合率: Precision}_{hk} = \frac{|A_h \cap C_k|}{|C_k|}$$

$$F \text{ 値: } F_{hk} = \frac{2\text{Recall}_{hk}\text{Precision}_{hk}}{\text{Recall}_{hk} + \text{Precision}_{hk}}$$

ここで、 $A_h$  は被験者によるクラスタリングの結果から、 $C_k$  は K-means 法によるクラスタリングの結果からそれぞれクラスタを使用する ( $h=1 \sim 8$ ,  $k=1 \sim 8$ )。

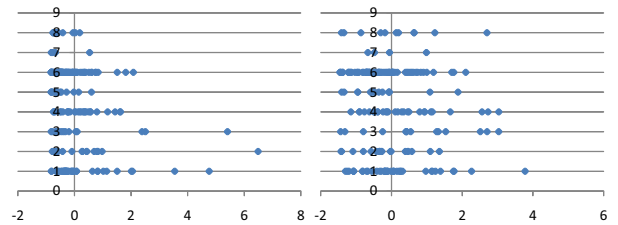


図 32: 符点 4 分音符の割合

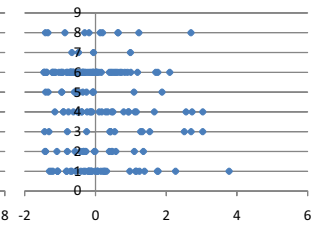


図 33: 4 分音符の割合

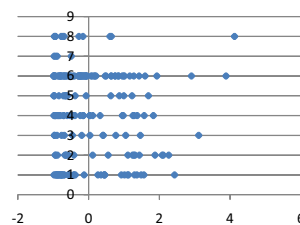


図 34: 8 分音符の割合

再現率は K-means 法によって分類されたクラスターの要素が被験者が分類したクラスターの要素をどの程度網羅しているのかを表す。また、適合率は K-means 法によって分類されたクラスターの要素と被験者が分類したクラスターの要素がどの程度合致しているのかを表す。クラスタリング全体の評価の F 値は次の式で求められる。

$$F = \sum_{h=1}^K \frac{|A_h|}{N} \max_k F_{hk}$$

ここで N は全体の要素数, K はクラスターの数である。また,  $\max_k F_{hk}$  は被験者が分類した一つのクラスターに対して最も高い F 値を算出する K-means 法によるクラスターとの間の F 値である。

今回のクラスタリング結果の平均再現率, 平均適合率, F 値は表 3 に示す。平均再現率, 平均適合率は F 値を算出する際に使用した再現率, 適合率の平均とした。

表 3: クラスタリング評価

平均再現率	0.291
平均適合率	0.263
F 値	0.246

### 3.4 考察

K-means 法でのクラスタリングでは, 特に音高に関連する特徴量がクラスタリングの際に重要な役割をしていることが, 図 3 ~ 図 9 から確認することができる。また, 複数の特徴量で分類が顕著に現れている第 3 クラスターを試聴した結果, 穏やかで急激な変化が見られない曲が集中していることが確認できた。しかし, 被験者によるクラスタリングでは, 音高に関連する特徴量とクラスターにおいて特徴的な分布が見られない。したがって, 人間に対して音高に関連する特徴量が適しているかは確認することができない。

クラスタリングの評価としてはほとんど人間の感覚通りのクラスタリングはできていないものと考えられる。これは, 曲全体の特徴を用いたため, 曲のイメージが大きく変化するような曲には特徴が有効ではなかったと考えられる。

## 4 まとめ

音楽内容検索を目的として音楽特徴による分類を行った。クラスタリングの結果, 特定の曲印象を持つ楽曲を分類する特徴はあると考えられるが, 人間の感覚に合った分類は出来ていなかった。今後の課題として, 曲

全体の特徴のみでなく, 曲の部分的な特徴を抽出するために, 曲中のフレーズを抜きだしフレーズごとの特徴を抽出し, 組み合わせること等が考えられる。

## 参考文献

- [1] 辻康博, 星守, 大森匡: 曲の局所パターン特徴量を用いた類似曲検索・感性語による検索, 信学技報, SP96-124, pp.17-24, 1997.
- [2] 竹川和毅, 土方嘉徳, 西田正吾: 内容に基づく音楽データの探索・推薦システム, 人工知能学会論文誌, Vol. 23, No. 5, pp.330-343, 2008.
- [3] 秋口俊輔: ソフトコンピューティング手法を用いた曲印象からの楽曲自動生成システムの構築, 知能と情報, Vol. 21, No. 5, pp. 782-791, 2009.
- [4] K.Hevner: Expression in Music: A Discussion of Experimental Studies and Theories, Psychological Review, Vol. 42, pp. 186-204, 1935.
- [5] K.Hevner: Experimental Studies of the Elements of Expression in Music, American J.Psychology, Vol. 48, pp. 246-168, 1936.
- [6] K.Hevner: The Affective value of pitch and tempo in music, American J.Psychology, Vol.49, pp. 621-630, 1937.
- [7] RWC 研究用音楽データベース, <http://staff.aist.go.jp/m.goto/RWC-MDB/index-j.html>.

## 連絡先

〒 782-8501 高知県香美市土佐山田町宮ノ口 185  
高知効果大学 大学院工学研究科 基盤工学専攻  
情報システム工学コース 吉田研究室  
中山達喜  
E-mail: 135074r@gs.kochi-tech.ac.jp  
電話: 0887-53-1020  
FAX: 0887-57-2220