2006－MUS－66（6）
2006／8／7

SOMを用いたベースラインからの音楽ジャンル解析

土　橋　佑　亮†　片　寄　晴　弘††

音楽ジャンルは Web 上での楽曲検索において有力な指標となる. これまで音響信号を用いての様々な音楽ジャンル解析の研究がなされてきたが, そのほとんどは様々なパートが混成する音楽を対象していた. 本稿では複音からの音源分離が比較的容易なベースパートに注目したジャンル推定を取り扱う. まずスケール, リズム, 音色の等の特徴量の設定と有効性の考察をし, それらを用いてマハラノビス距離,F 値最大境界による実験を行う. 更に Music Island を利用し, ジャンルの可視化と島の変化を調べる. マハラノビス距離による音楽ジャンル解析において,Metal/Punk では 73 %,Jazz/Blues では 80 %の認識率を得た.Music Island においては, 注目する特徴量に応じて島が変化することを確認した.

Music Genre Classification from Base-Part using SOM

Yusuke Tsuchihashi† and Haruhiro Katayose††

Musical genre helps us to search for songs on the web. Most of the previous works have focused on audio signal analysis for the pieces composing of various instruments. This paper presents an approach to music genre classification focusing on base part,the fundamental frequency of which is comparatively easy to be estimated. First, the paper describes features regarding scale, rhythm, timber, and examine those validity. Next, this paper describes about twe experiments based on maharanobis distance and one song to multi-genre correspondence. Finally, we illustrated music genre visualization with Music Island based on SOM. Experimental results by using mahalanobis distance show success rates of 78 % for Metal/Punk, and 80 % for Jazz/Blues. And we confirmed transformation of Music Island depending on each features.

1. はじめに

人々の音楽の楽しみ方はここ数年で大きく変化してきている. 我々は CD や MD などの媒体を用いて音楽を再生するだけでなく, ネットから配信された楽曲を mp3 や wma 等の圧縮形式としてハードディスクに大量に保存して聴くようになった. それに伴い web 上やジュークボックス内で楽曲を検索する機会も増えてきている.Yahoo や Google を用いた web 上での情報検索は今や情報収集の基本手段となっているが, 曲名やアーティスト名, 作曲者など楽曲の固有情報を対象としたキーワード検索が主流である. これに対し, 近年音響信号を対象とした情報検索技術に関する研究が積極的に行われている. 最近では街角で流れる音楽を携帯電話に録音し曲名や歌手名を検索する「あて!?メロ」のように, 実用化がなされ話題となったシステムもある.

† 関西学院大学理工学部研究科
Kwansei Gakuin University
†† 関西学院大学理工学部
Kwansei Gakuin University

Pop や Jazz などの「音楽ジャンル」は音楽を分類する上でよく用いられる. また音楽ジャンルは個人の音楽の好み, 感性に大きく関わるものでもあり, 楽曲検索における 1 つの尺度として大きな役割を果たす. 音楽ジャンル分けは国際会議 ISMIR(International Symposium on Music Information Retrieval) においても主要テーマとして扱われており, 近年の研究報告としては Roberto Basili[1] の機械学習を用いた調査や Cory McKay[2] らのニューラルネットを用いた手法がある.

ところで, ほぼ全てのジャンルの楽曲においてベースパートが存在する. ベースパートの特性としては

- 低域から和声構造を支え, 楽曲全体の印象を形成する, いわば骨組みとしての役割を担う
- 楽曲中の大部分で演奏が継続し, 無音区間の割合が極めて低い
- 装飾的な役割を果たすギターやキーボードとは異なり, 基本的に一曲中に一パートしか存在せず, かつほぼ単音での演奏である

等があげられる. このようにベースパートは楽器編成において重要でかつジャンルに関わる情報を含んでい

る. また後藤の PreFest[3] でボーカルパートとベースパートの音源分離が報告されているように,CD 等の混成音楽からの分離ができる. 現在のインターネットのブロードバンド化を考えると, 今後トラック別にデータが配信される可能性も大いにあり, ジャンル推定の切り口として適切であると考える.

CD や mp3 を始めとする世の中に存在する音楽のほとんどは音響信号のデータである. 本研究では音響信号としてのベースパートに着目したジャンル推定を課題とする. まず第 2 章では判別に用いる特徴量抽出について述べる. 第 3 章においてはマハラノビス距離と F 値最大距離での判定を用いた実験を通して「音高・音階」「リズム」に関する特徴量の有効性を確認する. 続いて第 4 章では Elias Pampalk によって提唱された Music island[4] を利用したジャンル情報の可視化を行い,「音高・音階」「リズム」「音色・奏法」の各特徴への重み付けによる島の変化について考察する.

2. ベースパートにおける特徴量抽出

2.1 特徴量の抽出

ジャンルを判別するためにはジャンル固有の特性をよく表した特徴量を利用する必要がある. ベースから得られる特徴量は大きく**「音高・音階に関する特徴量」「リズムに関する特徴量」「音色・奏法に関する特徴量」**の 3 つに分類される. これらの特徴量はベースパートの音響信号から図 1 に示すような一連の処理によって抽出される. 図 2 にパワースペクトルのピークから基本周波数を推定した例を示す. このデータをさらに分析することにより, 音高推移やリズムに関する特徴量を得る. 音色に関する特徴量としては効果に実績のある MFCC を利用する.MFCC 以外の特徴量の有効性については次章で検討する. なお, 本研究では音響データを対象に議論を進めるが,MIDI データによるベースパートのジャンル判定も同様の処理によって実現可能である.

2.2 ジャンル判別の対象

ジャンル認識の実験を行う際に判別ジャンルを設定する必要がある. ここでは ISMIR 2004 でのジャンル分けコンテストで採用された分類を基準とし, ベース音のない Classic と, データの収集が困難な World を削除する. 代わりに一般的に普及率の高い R & B/Funk を追加し, 以下の 5 ジャンルでの判別実験を行う.

[Pop/Rock] [Metal/Punk] [Electronic/Dance]
[R & B/Funk] [Jazz/Blues]

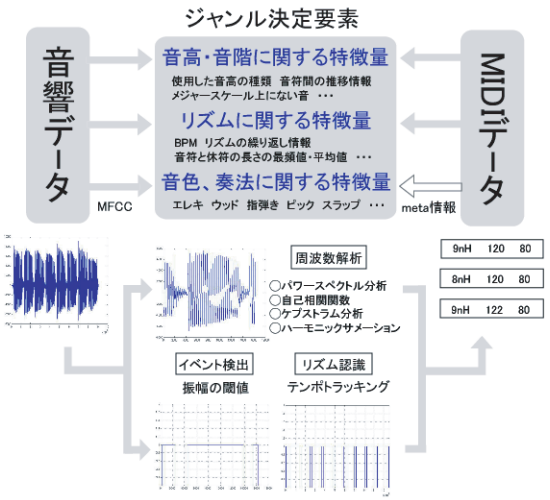

図 1 ジャンル情報と抽出関係

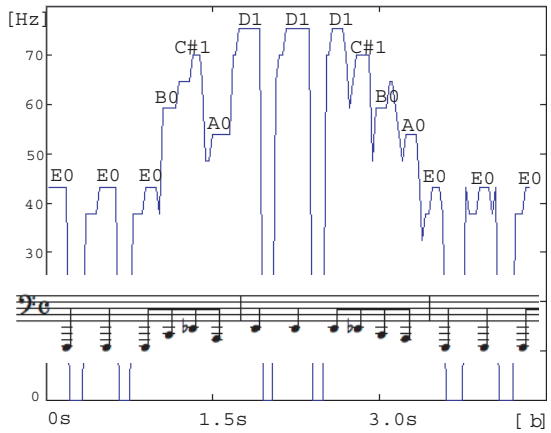

図 2 ベース音響信号からの自動採譜 (Satisfaction/The Rolling Stones)

なお判別の実験を行うにあたり, 訓練用と実験用に扱うデータとして音響信号に加え, 実験データ収集のためネット上のフリーサイト[5][6] より各ジャンルの smf の収集を行う. 各ジャンル 30 個, 合計 150 個のデータを収集した.

3. 音高・音階, リズムに関する特徴量の設定とジャンル分け実験

本章ではまず「音高・音階」「リズム」に関する特徴量候補 40 種を用意する. その後独立性を確認し, 判別に有効な特徴量の絞り込みを行う.

3.1 特徴量候補

3.1.1 音高・音階に関する特徴量

音高に関する情報と, その更に 12 との剰余により得られた音名の情報を用いて設定される以下 31 個の特

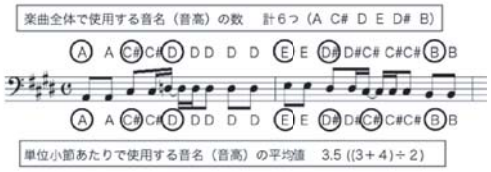

図 3　楽曲全体で使用した音高の数と単位小節あたりの音高の数

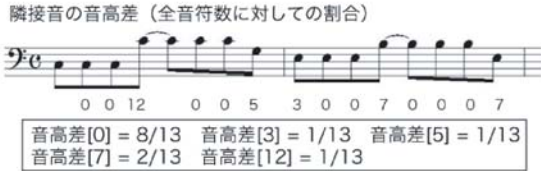

図 4　隣接音の音高差の絶対値

徴量について述べる.[ ] 内は要素数を表している.

- 楽曲全体で使用する音高と音名の数 (図 3 参照)[2]
- 単位小節あたりで使用する音高と音名の数の平均値 (図 3 参照)[2]
- 楽曲中の全入力における音高の平均値 [1]
- メジャースケール上にない音の割合 [1]
- Key を基準とした各音名の使用頻度 [12]
- 隣接音の音高差の絶対値 (図 4 参照)[13]

スケールの情報としてメジャースケールのみに注目したのは, 楽曲において最も使用頻度の高いスケール上にない割合を求めることでジャンルの差を獲得したいという理由による. 隣接音の音高差の絶対値とは, 音符と次の音符との pitch の差をとり, 何度推移したのかを 1 オクターブ変化までを考慮した全 13 個の配列に回数として保存し, 全音符数に対しての割合を特徴量としたものである.

3.1.2　リズムに関する特徴量

ここでは「音符と休符の長さ」として *IOO*(inter onset offset) を用いる. また「小節数」はドラムパートのビートの繰り返し周期を計算することにより獲得する. 音符や休符の長さや単位小節あたりの数に関する情報として以下の 9 つの特徴量を設定する.

- BPM(1)
- 楽曲全体での音符と休符の長さの最頻値 (2)
- 楽曲全体での音符と休符の長さの平均値 (2)
- 単位小節あたりの音符と休符の数の平均値 (2)
- リズムの繰り返し ratio(2)

リズムの繰り返し ratio とは小節単位での音符の時間配置を考えた時, 前の小節と同じ配置をとる割合を

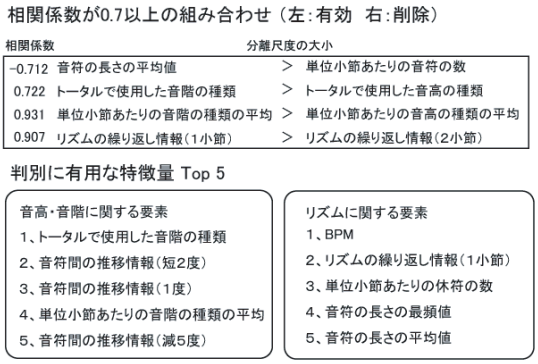

図 5　特徴量検討

意味する.1 小節単位の一致と 2 小節単位での一致の割合という意味で 2 つの特徴量として設定した.

3.2　有効な特徴量設定

特徴量候補 40 種について正規化を行った後, クラス間分散とクラス内分散によるジャンルの分離尺度を測定する. また相関係数による類似度によって判別に用いる特徴量を絞り込みを行う. 一般的に相関の強いとされる 0.7 を上回る組み合わせにおいて分離尺度の高いものを使用する特徴量として選択した. この結果を図 5 に示す.

3.3　single genre 対応実験

single genre 対応実験では全ての標本がただ一つのジャンルへと認識される. ここでのジャンル判別は各ジャンルの母集団へのマハラノビス距離の比較により行い, 最も距離の近いジャンルを第 1 推定とする. この実験において訓練データと実験データを分ける手法としてジャックナイフ法を用いた.

ここでは各ジャンル 24 個の合計 120 のベースパートで訓練を行い, 残り各ジャンル 6 個の合計 30 のベースパートにおいて判別を行う. これをデータを変換しながら 5 回繰り返した. この実験では前節で絞り込んだ特徴量を用いて主成分分析を実施し, 上位 4 つの成分を用いて判別を行う. 4 次元という設定は訓練データに 24 個に対して信頼のおける軸の数の限界に相当するものである. 図 6 で表す結果はジャックナイフ法による全 5 回の平均をとったものである. 各ジャンルの楽曲がどのジャンルとして判定されたのかを図 7 に示す.

図 7 からは,Metal/Punk と Jazz/Blues の判別においての識別率が高いことがわかる.Pop/Rock の第 1 推定のよる認識率の低さは R & B/Funk を始めとする他ジャンル全般への誤認識率が高いことが要因であると考えられる. また Electronic/Dance のデータに

再現率	第1推定での正解		第2推定までの正解	
	正解個数	再現率	正解個数	再現率
Pop/Rock	9/30	0.30	17/30	0.57
Metal/Punk	22/30	0.73	26/30	0.87
Electronic/Dance	17/30	0.57	21/30	0.70
R&B/Funk	17/30	0.57	25/30	0.83
Jazz/Blues	24/30	0.80	27/30	0.90
Total	89/150	0.59	116/150	0.77

図 6 認識結果（single genre 認識実験）

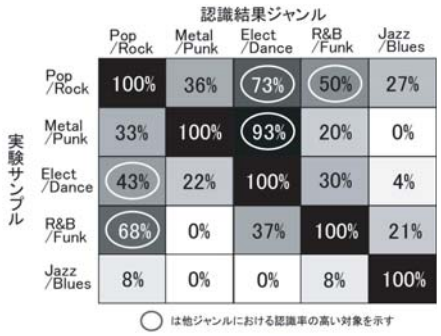

図 7 single genre 認識実験における第 1 推定ジャンル

再現率 F値	正解個数	再現率	重心から境界線までの距離平均	境界線における判定精度(F値)
Pop/Rock	22/30	0.73	1.968	0.513
Metal/Punk	15/30	0.50	1.480	0.640
Electronic/Dance	23/30	0.77	2.278	0.477
R&B/Funk	19/30	0.63	1.996	0.495
Jazz/Blues	24/30	0.80	2.398	0.725
Total	100/150	0.67		

図 8 認識結果 (multi genre 認識実験)

おいては R & B/Funk へ,R & B/Funk のデータにおいては Pop/Rock への誤認識が多いことから, これらのジャンルは非常に似通った概念空間に配置されるものと予想される.

3.4 multi genre 対応実験

本節で扱う multi genre 対応実験ではそれぞれの楽曲について全 5 ジャンルに対し, （属す◯）か（属さない×）の 2 択で判定していく. 方法としてはまず訓練データによって各ジャンルの要素として認める範囲の境界を事前に決定しておく. この境界は各ジャンルの集合の重心から広げていった時, 判別における F 値が最大となる距離において設定される. 判定は実験データが各ジャンルの判定範囲に入るかどうかに依存し, 重心からの距離が境界の距離より近ければそのジャンルに（属す◯）の判定をし, そうでなければ（属さない×）の判定をする.

multi genre 対応実験の結果を図 8 と図 9 に示す. 図 8 より single genre 対応実験と比べて認識率の向上が確認できる. 特に single genre 認識において認識率の低かった Pop/Rock の認識率は飛躍的に向上し, 続いて Electronic/Dance の認識率も上昇した.

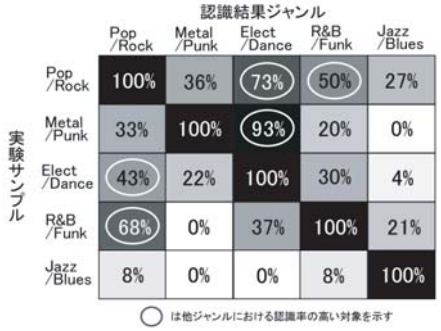

図 9 multi genre 対応実験において認識されたジャンル比率

また図 9 においては各ジャンルのデータが本来のジャンル以外にどのジャンルと認識されたのかを示している. このデータより以下のことが言える.

- Jazz/Blues のデータは他のジャンルとして判定されることが極めて低い. 一方 Pop/Rock と R & B/Funk については Jazz/Blues と判別される場合が多く, いずれも 20 %強である
- Metal/Punk における Electronic/Dance への認識率の高が高い. これは Metal/Punk の分布が Electronic/Dance に包括される形となっていることが原因であると予想される. これは本実験における Metal/Punk の認識率の低さの原因となっている.
- Pop/Rock,Electronic/Dance,R & B/Funk において相互間の認識率が高い. これは single genre 対応実験において認識率の低かったジャンルであり, これら 3 つのジャンルは非常に似通っていることがこの実験から確認された.

これら 2 つの実験結果を通じ, ここで取り上げた音高・音階、リズムに関する特徴量は特定のジャンルの判別には有効であるのに対し, 一方で判別が上手くいかないケースも確認された. この要因としては「1. ジャンル概念そのものの不完全性」と「2. 用意した特徴量の不完全性」が考えられる. 1 に関しては複数の判別ビューに基づく人間のインタラクションを前提としたジャンル判別が有効な解決法の一つと考えられる. 以下の章では Elias Pampalk によって提唱された Music Island を利用したジャンル分析について述べる.

4. Music Island を利用したジャンル分析

4.1 SOM アルゴリズムと Music Island

SOM[7] とは T.Kohonen により発表された教師なし学習ニューラルネットワークである.SOM は多次元のデータを圧縮して低次元のマップを描くことができ,

似た構造を持つデータは近くに, そうでないものは離れた位置に配置する.Music Island[4] とは Elias Pampalk によって提唱された SOM を用いた可視化システムであり, データが密集する場所で島を形成し, 更に密集度の高い場所では山, 低いところでは海を形成する.Music Island は概念空間を視覚化する方法として非常に優れている.

4.2 Music Island 適用に向けての設定

学習に用いる特徴量は「音高・音階に関する特徴量」「リズムに関する特徴量」「音色・奏法に関する特徴量」それぞれ主成分分析による上位 3 次元. さらにこれに配置をある程度矯正するための「ジャンル情報メタデータ」4 次元を加え, 系 13 次元のデータによる可視化を行う. マップは 10 × 10 のノードを設定し, 初期配置はランダムから始めるものとする. また実際の計算には Matlab の SOM Toolbox(version2)[8] と SDH Toolbox[9] を用いた.

「音高・音階」「リズム」に関する特徴量としては第 3 章で述べたものを用いる.「音色・奏法に関する特徴量」の設定についてはベースパートの冒頭 10 秒間のデータに対して MFCC を利用する☆. また実験におけるデータ数確保のため, 自負の演奏に加え Quick Time 音源から生成した音響データに対する MFCC を利用する. ただし Quick TIme 音源から生成した音響信号から得られる MFCC はほぼ同一座標に集中し現実的ではない. そこで正規乱数により分散させ以下の分析を実施した. 生成された音色は [Acoustic][Finger][Pick][Fretless][Slap(pull)][Slap(thumb)] [Synth1][Synth2] の全 8 種類である.

4.3 各特徴量重視による島の変化

我々は Music Island を用いて島の生成を行い,「音高・音階」「リズム」「音色・奏法」の 3 つに分類した特徴量の比重を自由に変化させることにより島の変化を調査した. 図 10 は各特徴量の比重を同一の値にした場合のマップである.

例として図 11〜図 13 において 3 分類した特徴量の比重を変化させた結果を示す.

図 11 は「音高・音階に関する特徴量」の比重を高くしていったときのマップである. マップの北東の方角には Jazz/Blues の島がそのまま残っており, それを囲うように R & B/Funk の島が点在している. これらは使用する音階のバリエーションが豊富でかつ Funk 特有の動きの激しいフレーズやウォーキングライン等により音高推移の激しいジャンルである. 南西の方角には Metal/Punk の大部分と Pop/Rock,Electronic/Dance の一部から成る島がある. これらは対照的にルート弾きが頻繁にあり, 音の推移に乏しいジャンルである.

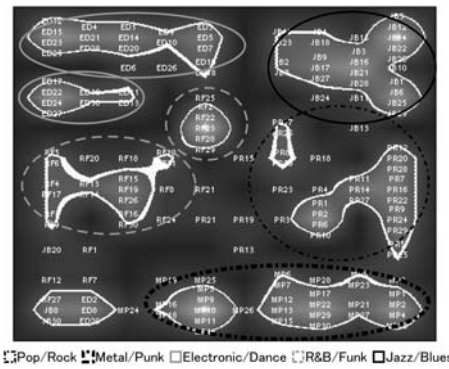

図 10 比重均等図 (メタ:ピッチ:リズム:音色 = 10 : 10 : 10 : 10)

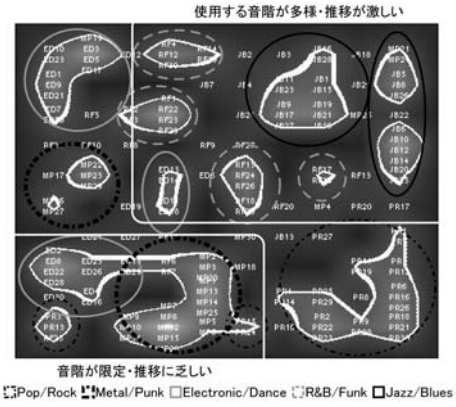

図 11 ピッチ重視図 (メタ:ピッチ:リズム:音色 = 10 : 25 : 3 : 3)

図 12 は「リズムに関する特徴量」の比重を高くしていったときのマップである.Metal/Punk を中心とした BPM の高いジャンルは南東に集まり, 対照的に BPM の低い Jazz/Blues やその他のジャンルにおける一部の楽曲が北東に島を作っている. また北西の方角にある様々なジャンルが混在する島では音符の長さが短い, もしくは休符が多い楽曲が集まった. これは 4 部音符でのウォーキングラインが特徴の Jazz/Blues と対局の位置にできることにおいてもよく納得できる.

最後に図 13 は「音色・奏法に関する特徴量」の比重を高くしていったときのマップである. このマップでは収集データの約半数を占めた指弾きの楽曲を中心に各音色, 奏法の島が生成されている. 注目したいの

☆ 具体的には Matlab の Auditory Toolbox[10] を用い, 低次 5 次元までの MFCC の平均と分散の計 10 個の特徴量を利用する.

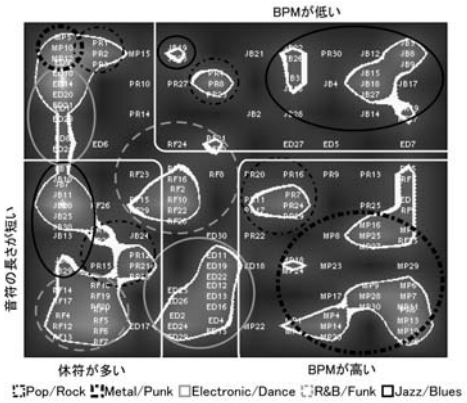

図 12 リズム重視図 (メタ:ピッチ:リズム:音色 = 10 : 3 : 25 : 3)

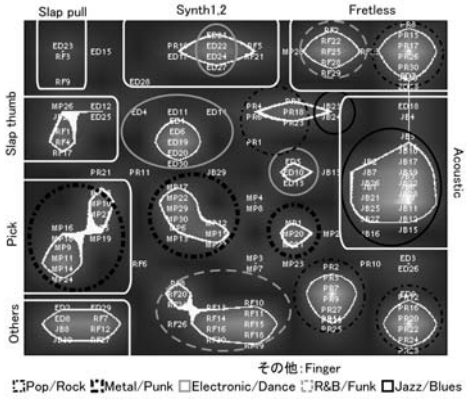

図 13 音色重視図 (メタ:ピッチ:リズム:音色 = 10 : 3 : 3 : 25)

は, 音色の類似度の高い Fretless と Acoustic, さらに Slap pull と Slap thumb がそれぞれ近隣に島を作っていることである.

5. おわりに

本稿では, ベースパートからの特徴量を用いての判別実験と Music Island を用いての可視化を行った. マハラノビス距離を用いた single genre 対応実験では Metal/Punk において 73 %,Jazz/Blues において 80 %の認識率を得た. また F 値を境界線として用いた multi genre 対応実験を通して Pop/Rock,Electronic/Dance,R & B/Funk の分布の類似性を確認した.Music Island を用いての可視化においては「音高・音階」「リズム」「音色・奏法」のそれぞれの特徴量に比重をかけることにより柔軟にマップを表示した. またそれぞれのマップにおいてジャンル間の配置の特徴を示した.

音楽の好みは人それぞれ異なる. それはメロディーを重視する人, リズムに依存する人, 特定のアーティストの歌声に惹かれる人, それぞれが実に多様な角度から音楽を楽しんでいるからである.Music Island を利用することでユーザのビューに合わせたジャンルの可視化が可能となった. 本研究ではベースパートに注目することにより, アンサンブル音響では利用が困難なピッチやリズムパターンを特徴量として利用することが可能となっている. ただし現段階では時系列メディアとしてのデータ獲得を十分に利用していない. 今後は Conklin[11] らのアプローチを取り入れ, ジャンル判別の基礎能力の向上についての検討を進めたい.

謝辞 この研究は,Elias Pampalk 博士によって提供された Music Island のシステムを利用しています.

参 考 文 献

1) Basili, R., Serafini, A. and Stellato, A.: Classification of musical genre: a machine learning approach., *ISMIR* (2004).
2) McKay, C. and Fujinaga, I.: Automatic Genre Classification Using Large High-Level Musical Feature Sets, *ISMIR*, pp. 13–20 (2004).
3) Goto, M.: A Real-time Music-scene-description System: Predominant-F0 Estimation for Detecting Melody and Bass Lines in Real-world Audio Signals, *Speech Communication (ISCA Journal)*, Vol. 43, No. 4, pp. 311–329 (2004).
4) Pampalk, E., Dixon, S. and Widmer, G.: Exploring music collections by browsing different views, *ISMIR*, pp. 201–208 (2003).
5) *MIDI DataBase*, http://www.mididb.com/.
6) *freemidi.org*, http://www.freemidi.org/.
7) *Self-Organizing Maps.Sprinter,3rd edition* (2004).
8) *SOM Toolbox: Laboratory of Information and Computer Science in the Helsinki University of Technology*, http://www.cis.hut.fi/projects/somtoolbox/.
9) Pampalk, E., Rauber, A. and Merkl, D.: *SDH Toolbox*, http://www.ofai.at/~elias.pampalk/sdh/.
10) Slaney, M.: *Auditory Toolbox Ver.2*, http://www.slaney.org/malcolm/pubs.html, technical report #1998-010, interval reseach corporation edition.
11) Conklin, D. and H.Witten, I.: Multiple Viewpoint Systems for Music Prediction., *Journal of New Music Reseach*, Vol. 24, No. 1, pp. 51–73 (1995).