JAIST Repository
https://dspace.jaist.ac.jp/

Title	音響特徴を用いた音楽推薦システムに関する調査研究 [課題研究報告書]
Author(s)	中川，洋志
Citation	
Issue Date	2014-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/12055
Rights	
Description	Supervisor:赤木正人，情報科学研究科，修士

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

第1章　序論

1.1　音楽推薦システムとは

音楽推薦システムとは, ユーザの嗜好を分析・予測し, 提示するシステムのことである. ユーザの嗜好分析は多くの場合, そのユーザが過去に聴いてきた楽曲や, そうした楽曲につけた評価などから行なわれる.

インターネットの発達と音楽のデジタル化を初めとした技術の発展により, ユーザが触れることのできる音楽情報量は著しく増大している. これが原因で, ユーザが楽曲の探索に時間をとられてしまうという事態が発生している. こうした状況の改善のために, 音楽推薦システムの研究が求められている.

以下では音楽推薦システム研究の背景, および現行の音楽推薦システムを取り巻く状況について述べていく.

1.1.1　音楽推薦システム研究の背景

音楽推薦システムが研究されるようになったのは, インターネットの発達と音楽のデジタル化により, 我々が得られる音楽情報が膨大かつ多様になったからである.

例えば, あるユーザがなにかクラシック音楽を聴きたくなったとしよう. そのユーザがまず直面するのは,「どの作曲家の音楽を聴くか」である. ベートーベンやモーツァルトを初めとして, クラシック音楽の作曲家は数多い. どの作曲家の音楽を聴きたいかが決まっていたとしても, 次には「どの演奏様式の音楽を聴きたいか」を選ばなくてはならない. クラシックには管弦楽曲や弦楽四重奏, ピアノの独奏曲などのように種々の演奏様式が存在しているからである. どの様式が聴きたいかが決まったとしても, 次に「どの演奏者を選ぶか」という問題がある. たとえ同じ楽曲であったとしても, 演奏者によって表現方法は変わってくる. さらに演奏者によっては同じ楽曲を何回か演奏していることもあり, そうした場合には「いつ演奏されたのか」といった要素も考慮する必要がある.

演奏者や演奏された年代などはある程度クラシックに精通した人間でない限り気にしないかもしれない. しかし, それを考慮したとしても, ユーザが選択できる楽曲情報が相当に多くなっているというのは疑うべくもない. もしジャンルの選択から「好きな音楽」の探索を行おうとするならば, そもそもどこから手をつけていいかわからないといった状況に陥ってしまうことも考えられる.

ある楽曲が嗜好に合うかどうかを判断するには, 少なからず時間がかかる. 音楽を聴いて判断しようと考えるならば当然のことだ. 数え切れないほど多くの楽曲の中からお気に入りの楽曲を探し出すという作業は, 非常に時間と労力のかかるものとなっている.

音楽推薦システムは, こうした状況を打破することができる. ユーザの嗜好を分析し, それに合う音楽を推薦することにより, ユーザがお気に入りの音楽を探索する時間を短縮することができるからだ. そのために, 音楽推薦システムの研究が必要とされている.

1.1.2 現行の音楽推薦システム

現在でもすでにいくつかの推薦システムが実用化され, 商業的に利用されている. 音楽のみに関わらず, 広い範囲での推薦を行っている例としては Amazon などが挙げられ, 音楽のみのものとしては iTunes Store などで用いられているものがある.

これらの推薦システムは, ユーザの購入履歴や, ユーザがあるアイテムにつけた評価などを基に推薦を行っている. こうしたシステムはユーザと似た嗜好を持った他ユーザが高く評価したアイテム、いわば「口コミ」のものを推薦対象にすることができる. しかしながら, 評価がなされていないアイテムは推薦対象とすることができず, また人気やアーティストといった面で推薦結果に偏りが生じてしまうという問題がある.

基本的な音楽推薦システムには, 以下の三種が存在する。ユーザによって付与された主観的評価を使用して推薦を行なう協調フィルタリングを用いた音楽推薦システム, 付与されたジャンルやアーティスト名などのような第三者から付与されたテキスト情報を用いたメタデータ使用型の音楽推薦システム, そして楽曲の内容を表す音響特徴量を用いて楽曲の推薦を行なうコンテンツベースのフィルタリングである. 現在では協調フィルタリングおよびメタデータ使用型の音楽推薦システムが主流となっており, 音響特徴量を用いたコンテンツベースの音楽推薦システムは数が少ない.

1.2 音楽推薦システムに求められる能力

音楽推薦システムをより良いものへとしていくためには, どのような能力がシステムに求められているかを念頭に置き, 研究を進めていく必要がある.

過去の研究においても様々な目標設定がなされてきた. その中でも音楽推薦システムに対する根本的な要求を示しているのが Uitdenbogerd らである [41]. Uitdenbogerd らは自身の論文の中で, 「ユーザが興味を抱くだろう楽曲を推薦できる」音楽推薦システムこそが理想であると主張している. また同様に, 吉井らも音楽推薦システムに求められる能力として, 「思いがけない発見を可能にする能力（Serendipity）」が特に重要であると分析している [44]. この Serendipity とは, 元々は Herlocker らによって提案された推薦システムの評価指標の一つである [21]. 人の力だけでは予測・発見することが難しいようなアイテムを推薦できる能力を表すものであり, この能力を検出するためには推薦されたアイテムがユーザをどの程度引き付け, 驚きを与えたかを測定する必要があるとされている [53].

もともと楽しみのために聴かれる音楽を推薦すると考えると, この能力は音楽推薦システムに必須のものであるといえる. 推薦システム全般を見た Serendipity を研究している例としては村上らや

一般的に, ある程度音楽に興味があり, かつ音楽を好む人々は能動的に音楽の探索を行う場合が多い. 例えば, ある音楽が気に入った場合, その音楽を演奏しているアーティストの他の作品を探してみたり, 一歩進んでそのアーティスト周辺のアーティストに手を伸ばしてみたり, といった具合にである. あるいは, ロックにおけるブルースやカントリーなどのように, 気に入ったジャンルの源流となるジャンルに手を伸ばす人々もいるかもしれない.

こうした音楽愛好家たちをユーザとして想定した場合, 既存の音楽推薦システムで彼らを満足させる音楽を推薦できるかは疑問である. 既存のものに用いられている技術では, 推薦できる音楽が人気やジャンルなどで偏ってしまうことが多いからだ. ユーザに新鮮なアイテムを提供することを考えた場合, 新しく作成された楽曲や世間的にマイナーな楽曲を推薦対象に取り上げる能力が求められる.

1.3 本研究の目的

前項において、音楽推薦システムに求められる能力として思いがけない発見をもたらす能力, すなわち Serendipity が重要となることを述べた. しかしながら, 現在においてこの能力を持った音楽推薦システムの実現は困難とされている. なぜならば, 先述したとおり Serendipity という能力はあるユーザがどれだけ興味を持ったか, あるいは驚いたかを表す主観的な指標であり, 客観的なものではないからである [53].

本研究では, この Serendipity を持った音楽推薦システムの実現に向け, 音響特徴を用いた音楽推薦システムに関する研究を初めとした関連研究の調査を行う. それを基に音楽推薦システムの発展のために解決しなければならない課題の発見, 改善策の検討を行っていく.

1.4 関連研究の要求項目

本項では, 関連研究の評価を行うために必要な評価項目について検討する. 「思いがけない発見」というものをシステムに直接判断させるのは現時点では難しい. そのため, まずはどのようなアイテムがユーザにとって「思いがけない発見」として捉えられるかについて考える.

まず一番に驚きを与えられるものとして考えられるのが, 新しいアイテム, すなわち作られて間もない楽曲である. 製作されたばかりの段階ではそれらを耳にするユーザが少なく, どのような内容であるかを把握する人間は少ない. 楽曲の製作者が有名なアーティストであれば, そのファンなどには予測もできるかもしれない. しかし, アーティストが新人

であった場合などはそれこそ内容が予測不可能であり, すぐに手を出すことは難しいだろう. こうした音楽をユーザの嗜好にそって推薦することができれば, それはユーザにとって好ましいものとなると考えられる.

第二に考えられるのが「知名度の低い音楽」である. クラシックの例を挙げるのであれば, ベートーベンやモーツァルトの楽曲群のような知名度の高い音楽は, ユーザにとって耳にする機会も多く, また調査も容易である. しかし, そうでない作曲家の楽曲は内容の予測はおろか発見すらも難しい. インターネットの発達によって, こうしたマイナーな楽曲と触れる機会が増えると考えられる. こうしたマイナーな楽曲群の推薦も, ユーザの興味を引くと思われる.

第三に, 畑違いの音楽が挙げられる. 音響的特徴は似通っているのに, 別のジャンルに分類されているケースである. 例えばロックを例に挙げると, ロックは時代を経るにしたがい様々な他ジャンルの影響を受け, プログレッシブロックやカントリーロックのような数々のサブジャンルを生み出している. そうしたサブジャンルの中には, 本来ロックを聴かないユーザにとっても好ましい楽曲が含まれている可能性がある. そうした楽曲を推薦することができれば, ユーザに新鮮味のある楽曲を提供できるだけではなく, そのユーザに新たな興味を抱かせることが可能となる.

こうした楽曲群を推薦するために評価すべき項目は, 以下のようなものが挙げられると考えられる.

1.4.1 音響信号への対応

音楽推薦システムには, 楽曲が音響信号のみの状態で渡された場合においても即座にそれを分析し, 推薦対象とする能力が必要となる.

新しい楽曲が製作されたとする. この時, 新楽曲は誰からも聴かれることがないために評価が付けられず, またメタデータも付与されていない. しかし新しい楽曲はユーザの興味を引きやすいものであるため, 素早い推薦対象への取り入れが求められる.

また, こうして作られた楽曲の多くは純粋な音響信号である. そのため Musical Instrument Digital Interface(MIDI) データなどのような記譜情報は作られるのに時間がかかってしまう. 現在において, 音響特徴量の抽出は音響信号・記譜情報のどちらからも可能となっている. しかし, 取り込みの速度を考えた場合, 記譜情報は音響信号に比べ取り込みに時間がかかるとここでは考える.

1.4.2 未評価楽曲への対応

音楽推薦システムには, 取り扱う楽曲データベース内に未評価の楽曲があったとしても, それがユーザの嗜好に合うものであれば推薦する能力が必要となる. またそれと同時に, 楽曲に付与された評価データの多少によって推薦結果に偏りが生じないようにしなければならない.

新しい楽曲の取り込みとは別に, すでに取り扱う楽曲データベースに評価されていない, あるいは評価の少ない楽曲がある場合においても, 音楽推薦システムはそれらの楽曲を評価されている楽曲と同等に扱い, 推薦することが求められる. 他人からは不評であっても, そのユーザにとっては嗜好に合う楽曲である可能性がある.

評価を受けていないということは, その楽曲は他の楽曲に比べて「マイナー」である可能性が高い. すなわち, ユーザ独自の調査では発見が困難な楽曲である可能性がある. こうした楽曲を提供するためにも, 未評価楽曲の推薦対象取り入れは重要である.

1.4.3 音楽的特徴を踏まえた音響特徴量の選択

音響特徴量を使用すると仮定した時、音楽推薦システムは音楽の特徴に関連する音響特徴量を選択し、推薦のために利用する必要がある。

音楽は、メロディ・リズム・ハーモニーなどのような音楽的特徴を持っている。聴き手はこうした音楽的特徴から音楽への印象を形成し、かつ音楽的嗜好の形成を行っている。現在の音楽推薦システムでは、こうした音楽的特徴を音響信号から直接分析することは難しい。そのため、音響信号を自動解析することによって得られる低次音響特徴量を用いている。

そうした低次音響特徴量を無意味に利用するのではなく、音楽的特徴と関わりのあるものを選択し、使用できているかは音楽推薦システムにとって重要な位置を占める。

1.4.4 推薦精度以外の評価軸の有無

音楽推薦システムにとって, ユーザの嗜好にあった楽曲を推薦することは一番の課題となる. しかしそれだけではなく, 推薦した楽曲の多様性のような推薦精度以外の観点から推薦を行う能力が音楽推薦システムにとって重要となる.

多くの音楽推薦システムの研究において, 推薦精度は「推薦した楽曲群の中にユーザが気に入った楽曲がいくつあるか」という計算で求められる [53]. 確かに推薦システムである以上, ユーザの嗜好にあった楽曲を推薦しなければならない. しかし, ただ「好みに合う」というだけでは, そのユーザの好きなアーティストの曲ばかりを集めたり, 演奏者違いの同じ楽曲ばかりを集めたりするという結果になる恐れがある. こうした推薦結果から「予期せぬ出会い」を期待することは難しい.

吉井らは自身の調査の中で, 推薦精度がある一定以上に達すると, システムを利用するユーザが違いに気づかなくなるとの報告を行っている [44]. このような報告を念頭に置くと, 音楽推薦システムにおいて推薦精度ばかりを追求するのは効果が薄い. そのため, 効果的な推薦を行うために, 例えば推薦結果内のアーティスト, または音響的特徴の多様性などのような精度以外の評価軸にしたがって推薦を行う能力が必要となる.

1.5 本論文の構成

本項では, 本論文の構成について述べる.

第二章では, 一般的な音楽推薦システムの形式について説明を行う. 基本的な手法である協調フィルタリングとコンテンツベースのフィルタリングについて触れた後に, 現在の研究において主流に扱われているハイブリッド型音楽推薦システムと, 音楽推薦に特有である二つの関連研究について説明を行う.

第三章では音響特徴を用いた音楽推薦システムに関する研究をいくつか例に上げ, それらについて内容の説明と評価を行っていく. 例示した研究は大きく分けて, 音響特徴を用いて楽曲のグルーピングを試みたもの, 音響特徴量のみを使用して音楽推薦をおこなったもの, そしてハイブリッド型音楽推薦システムに関するものの三種類である.

第四章では第三章での評価を基に現在の音楽推薦システム研究での不足点はどこか, また, どのような課題を設定するべきかを議論していく.

第五章では本論文全体の統括を行う.

第2章　各種音楽推薦システムの説明

本章では, 基本的な音楽推薦システムの方式について説明を行う.

まず, 音楽推薦システムのみならず一般的な推薦システムの基本形でもある協調フィルタリング使用型とコンテンツベースのフィルタリングを使用したものについて述べた後に, その二つを組み合わせたハイブリッド型音楽推薦システム, また音楽推薦に特有である二つの推薦形式についても触れていく.

2.1　協調フィルタリング

協調フィルタリングは, 最初期の音楽推薦システム [36] に導入されていた形式である. Amazon や iTunes など, 多くの商業用音楽推薦システムで利用されている.

この形式の最大の特徴は, 取り扱うアイテムの中身ではなく, それにつけられたユーザの評価を基として推薦を行うという点にある. 具体例でいうならば、Amazon などでいう五段階の星評価がそれにあたる. このため, アイテムの分析が不要であり, かつどのようなアイテムに対しても適用可能, という利点がある [53].

しかしながら, 採用される推薦システムが多いために研究でも比較対象とされることが多く, いくつかの問題点も指摘されている. 代表的なものとしては、ユーザから評価されていない楽曲が推薦の対象とならない、という問題がある [45] [32]. それに加え, 多数のユーザの評価が得られないと, 効果的な推薦を行うことができないという指摘もなされている [10] [35]. これらのように、楽曲に対する評価が不足していると質のいい推薦ができなくなってしまう問題をコールドスタート問題と呼ぶ [53] [44]. それとは別に, 楽曲の人気や他ユーザのアーティストに対する好みの偏重などが原因で, 推薦結果に偏りが生じてしまい, 実際のユーザの嗜好とかけ離れたバラエティに乏しい推薦結果が出てしまうといった問題も挙げられている [10] [45].

この協調フィルタリングを取り扱った研究として多く比較対象となるのは, 最初期に発表された音楽推薦システムの一つである Shardanand らの Ringo [36] や, 多くの推薦サービスで用いられる Saewar らのアイテムベースの協調フィルタリング [34] である. また, ユーザの評価だけでなく Web ページの情報も利用した Cohen らによる研究もある [16].

2.2 コンテンツベースのフィルタリング

コンテンツベースのフィルタリングは, 協調フィルタリングと並び基礎的な推薦システムの一つである. その名のとおり, アイテムの中身を参照した上でユーザに推薦を行う.

音楽を対象とした場合, 使用できる「中身」は大きく分けて二つである. 楽曲に付与されたテキスト情報であるメタデータと, 音響信号を解析して得られる音響特徴量である [44].

以下でそれぞれを使用した推薦システムの特徴を述べる.

2.2.1 メタデータ使用型

楽曲に付与されたメタデータには様々な種類がある. 楽曲のタイトル, アーティスト, あるいはジャンルといった音楽の内容を直接的に表すものであったり, あるユーザがその楽曲に対して抱いた印象であったりする. 後者の場合は不特定多数のユーザによって付与されるソーシャルタグである場合が多い [44]. メタデータを使用している推薦システムの例としては音楽配信サービスである Pandora や Last.fm といったものが挙げられる. 前者は専門家が楽曲を分析した上でアノテーションを行っており, 後者ではソーシャルタグを利用している。

こうしたメタデータは音楽検索のクエリとして用いられることもあるように, ユーザがある程度の音楽的知識を持ち, また自身の嗜好を把握していれば効果的である. ジャンルなどを基にそのユーザの嗜好に合う音楽を推薦することが可能となる. また, 曲から受ける印象などを頼りに, 自力の調査だけでは探し出せない楽曲を発見できる可能性もある.

しかしながら問題も多い. まず、一般にユーザが自身の音楽的嗜好をはっきりと自覚することが難しいという点がある [45]. 特にこれから音楽を調べてみようというような使い始めのユーザにとっては, ジャンルに対するイメージが沸きにくかったり, 印象語が表す音楽的内容が把握できない, ということが充分に考えられる. また, かつて嫌いだった音楽が, 年齢を重ねるにつれ音楽的知識・経験が深まることにより, 受容できるようになるケースも想定される. また別の問題として, こうしたメタデータは基本的に人力で付与されるという前提からくるものが考えられる. 例えば, 実際に定義されたジャンル情報がユーザの嗜好を表すのに適切でない場合が考えられたり [25] [1], メタデータの数が多いために取得に時間がかかってしまったり [35], といった問題である. そもそもにしてメタデータを付与する評価者が音楽を聴かないかぎりメタデータは付与されないため、メタデータのない楽曲の存在は充分にありうる。こうした楽曲が推薦対象にあがることがないのも問題となる。

メタデータを使用した推薦システムを扱った研究としては, Celma ら [14] や Basu ら [5] の研究がある. また, メタデータの実用性を示した研究として Whitman らのものがある.

2.2.2 音響特徴利用型

音響特徴量とは楽曲の内容を計算機に取り扱える形式にしたものである. 高速フーリエ変換 (Fast Fourier Transform) の登場, そしてハードウェアの発達などによって音響信号の分析が素早く行われるようになり, 音響特徴量を利用した音楽推薦システムの研究は盛んになった. 使用される音響特徴量は様々であり, 詳しくは付録にて説明を加える. この特徴量の類似度をコサイン距離やピアソンの相関係数のような尺度に基づいて計算することがこの形式の推薦システムの基本となる [44]. この仕組みを用いた商業用の推薦システムは, 協調フィルタリングやメタデータ使用型に比べて数が少ない.

これを利用した音楽推薦システムの特徴としては, 解析さえ行えば楽曲を推薦対象にできる点が挙げられる. 先に紹介した二つの音楽推薦システム, すなわち協調フィルタリング使用型やメタデータ使用型は, その仕組み上対象とするデータが存在しないと楽曲を推薦対象にすることができない. 言い換えると、新しい楽曲が追加された場合でもデータが付与されないかぎり推薦対象に挙げられず, ユーザに新しい楽曲を提供するのが送れてしまうという問題が生じてしまう. それに対して音響特徴を使用した音楽推薦システムはすぐに新しい音楽を提供することができる. 推薦結果が偏りがちな協調フィルタリングに対して, 偏りのない推薦結果を出すことができるという利点もある [44].

この形式の推薦システムで指摘されている課題は以下のとおりである. まず, ユーザの嗜好との関連付けが難しいために, 嗜好分析能力に難があるとみなされていること [45] [35] が挙げられる. それに加え, ユーザが持っている文化的背景を初めとした, ユーザの嗜好を形成する要因を反映させることが難しいという問題も指摘されている [44].

音響特徴のみを利用した音楽推薦システムとしては, 音響信号を利用した Cano らの研究 [10] に Logan の研究 [30], Hoashi らの研究 [23]、そして Bogdanov らの研究 [6] などがある.MIDI からの特徴量を用いたものであれば竹川らの研究 [51] があり, 特定のプラットフォーム用の研究としては Gasser らのものがある [20].

2.3 ハイブリッド型

ハイブリッド型音楽推薦システムとは, 前述した協調フィルタリングとコンテンツベースのフィルタリングを組み合わせたものである. 一般的には協調フィルタリングと音響特徴を利用した推薦システムの組み合わせが多い.

協調フィルタリングと音響特徴を利用したフィルタリングは, それぞれ推薦システムとして重要な問題を抱えている. 前者であれば評価を付与されていない楽曲が推薦対象になることがなく, 後者であればユーザの嗜好分析を行いにくい. そしてこれらの問題は, それぞれの形式だけに頼ると改善が難しかった. そのため現在では, この二つの形式を組み合わせることにより, 互いの欠点を打ち消そうとする試みがなされている.

この形式を用いた音楽推薦システムの代表的な研究としては, 2006 年に吉井らが行った研究 [45] や tiemann らによる研究 [38] などがある. これらを皮切りに、音楽推薦システ

ムの研究ではハイブリッド型の推薦システムを取り扱ったものが主流となった [28] [35]. また, 協調フィルタリングと音響特徴だけではなく, メタデータなどを取り入れた研究もある [32] [19].

2.4 状況に応じた音楽推薦

状況に応じた音楽推薦とは、ユーザの心情やユーザが置かれている状況からその場に最適な音楽を推薦する仕組みのことである。「うれしい」「かなしい」などのような印象語が用いられることが多い.

このような推薦の仕組みは音楽推薦に特有なものであるとされている [44]. 技術の発展により, ポータブル音楽プレーヤーの性能・容量が向上し, どこにでも多くの楽曲データを持ち運べるようになったために研究が盛んになった分野といえる. たとえプレーヤー内の楽曲データをユーザが把握していたとしても, その時々によって聴きたい楽曲というのは変化する. 例えば, 陰鬱な気分のときに明るいロックの楽曲を聴いたとしても普段とは違って楽しめないだろうし, 車の行き交う道路のそばでクラシック音楽を楽しもうとしても騒音で邪魔されてしまうことも考えられるだろう. このような状況に対応できる音楽推薦システムがあれば, 楽曲を選択する手間を省くことができる.

この形式を扱った研究としては以下のものが挙げられる. 小田川らは、自動車に搭載する音楽推薦システムを念頭に置き, 「明るい」「ノリがいい」「静かな」といった楽曲の印象を示すことでそのイメージに応じた楽曲を再生するシステムを考案した [50]. なおこのシステムでは, 応用として時間帯や天気、走行状況などを考慮した上での推薦も可能となっている. 仲村らは, 楽曲およびその歌詞から聞き手が想起する色彩を調査することにより, 楽曲・歌詞と色彩の結びつきに着目した楽曲推薦の実現可能性を論じている [52].

2.5 プレイリストの自動生成

プレイリストとは, 「何らかの意図をもって楽曲を並べたもの」 [44] である. 前項で例に出したポータブル音楽プレイヤーなどでは, ユーザがいくつかの楽曲を編集し, このようなプレイリストを作成することが可能となっている. また、ロックやポップスなどのアルバムという形態も, このプレイリストといえる. こうしたプレイリストは、多くの場合に初めの楽曲から順に再生して鑑賞することが想定されている. 特にロックアーティストなどが意図やメッセージなどを込めて作成するコンセプトアルバムなどではそれが顕著となる.

プレイリストの自動生成は文字通り, 類似度や順序などを考慮した上で楽曲の推薦を行うシステムのことを指す. 多くの場合, ユーザからプレイリストの元となる楽曲 (これを seed song という [44]) を指定してもらい, それを参考にデータベース中から適切な楽曲を調達する仕組みである. こうしたプレイリストの自動生成システムの例としては iTunes

で使用される Genius が挙げられる．これは協調フィルタリングによって類似した楽曲を探索したシステムであり, その有用性は Barrington らによって示されている [4]．しかし協調フィルタリングを用いているために, 協調フィルタリング特有の問題を抱えている．

このシステムを取り上げた研究としては, MFCC を用いて類似楽曲を集めたプレイリストを作成した Logan らの研究 [31] や歌詞から得られたアノテーションを利用した梶らの研究 [43], Aucouturier ら [2] や Vignoli ら [42] の研究などがある．

第3章　関連研究調査のまとめ

3.1　関連研究の選択基準

本章では, 調査した関連研究の中から音響特徴量を使用しているものをいくつか抜粋し, 紹介と評価を行っていく.

紹介する関連研究は, いずれも何らかの形で音響特徴を用いたものである. ここで協調フィルタリングやメタデータ使用型の音楽推薦システムを選択しなかったのは, これら二つの形式の音楽推薦システムが, 評価基準となる「音響信号への対応」と「未評価楽曲への対応」の二つを満たせないと判断できたからである. また, プレイリストの自動生成や状況に基づいた音楽推薦は発展的な研究であると考え, 本項では取り上げていない.

本章で紹介していく関連研究は以下のとおりとなる. まず, 音響特徴を用いて楽曲をグルーピングし, 推薦を補助しようとした Chen ら [15] の研究と Kuo ら [25] らの研究について触れる. 次に, 音響信号から抽出した音響特徴のみを利用した、純粋な音響特徴利用型の音楽推薦システムに関する,Logan [30], さらに音楽的な内容も考慮した Cano ら [10] や Bogdanov [6] らの研究について述べていく. 最後に, 現在の音楽推薦システム研究の主流であるハイブリッド型推薦システムを取り扱った, 吉井ら [45], Li ら [28], Lu ら [32], Shao ら [35], そして Domingues ら [19] について言及していく.

なお, Cano らや Li らは上記で挙げたもの以外にも音楽推薦システムの研究を行っている [11] [27]. しかしこれらの研究は前述したそれぞれの研究と内容が似通っているためここでは触れない.

3.2　音響特徴を用いたグルーピング

音響特徴を用いたグルーピングは, 音響特徴を用いた音楽推薦システムの中では初期に行われた研究である. 取り扱う楽曲をそれぞれの音響特徴に基づいてグループ分けすることにより, ユーザの嗜好にあった楽曲推薦を行いやすくするものである.

この時期では音響信号からの音響特徴量抽出が難しかったこともあってか, 両者とも MIDI ファイルから特徴量の抽出を行っている. そのため, 音響信号への対応は低くなっている.

3.2.1 Chen and Chen(2001)

Chen らの研究の概要を述べる. 彼らはユーザから要求したキーワードから検索を行うシステムには限界があると主張し, テキスト情報ではなく, ピッチや音量のような音響特徴を用いた音楽推薦の方法を提案した.

彼らは MIDI ファイルのメロディパートを選択したあと、そこから音響特徴の抽出を行っている。まず、メロディパートの選択にはトラック中の音高を MIDI ファイル中の音高で割ることによって得られる音高密度 (Pitch Density) を用いている。その後、求められたトラックから以下の六つの音響特徴量を抽出している。音価の平均 (Mean) と標準偏差 (Standard Deviation), 音高密度、音高エントロピー、テンポ、そして音量である。これらの特徴量の中から二つ用いて楽曲のグルーピングを行う仕組みとなっている。この研究中、Chen らは分類実験においてエラー率 (Error Rate) の低かったピッチエントロピー、テンポと音量の三つの音響特徴量を選び出している。

グループ分けを行った後, ユーザがいつ, どのグループの楽曲にアクセスしたかを示すアクセスパターンを元として推薦を行う. 推薦手法として彼らは三つの手法を考案している. ユーザが過去にアクセスしたグループの中から楽曲を推薦する CB Method, アクセス履歴からユーザのグルーピングを行った後, そのグループ内で好まれている楽曲グループから推薦を行うという COL Method, そして推薦システム中で長い間アクセスされている楽曲と、短期間で多くアクセスされた楽曲を半々で推薦する STA Method である. COL Method と STA Method は, あるユーザが本来アクセスしない楽曲を推薦するための仕組みとなっている.

これらの推薦手法の評価基準として, Chen らは推薦された楽曲のうち, ユーザがどれだけその楽曲へアクセスを行ったかという推薦精度を用いている. その結果, CB Method がもっとも推薦精度が高く, 次いで STA Method, COL Method の順で精度が下がっている. これは CB Method がユーザがアクセスした楽曲グループの中からしか楽曲を推薦していないからではないか, と Chen らは主張している.

音響特徴によって楽曲をグルーピングするという発想, およびユーザに「予期せぬ発見」をさせるために複数の手法を考案したのは評価できる. しかしながら, 抽出した音響特徴が音楽的内容を反映しているかが不透明であり, 推薦に使用する音響特徴の選択を, 分類実験における Error Rate で選出するのが適切であるかどうかには議論の余地がある. また, COL Method の提案において予期せぬ発見を目的としているのにも関わらず, それが実際に効果を発揮しているかどうかを判断する評価基準を設けなかったのは不十分である.

3.2.2 Kuo and Shan(2002)

Shao らは楽曲をメロディ形式によって分類し, ユーザがアクセスした楽曲のメロディパターンからそのユーザの嗜好を分析する音楽推薦システムを提案していた. ユーザのアクセス履歴を学習することによって楽曲のグルーピングが変化する仕組みになっている. これは当時の音楽推薦システムが第三者によって定義されたジャンルにしたがってユーザの

嗜好を推定していたため, 定義されたジャンルの定義がユーザにそぐわない場合を想定してのものである.

彼らはまず MIDI ファイルからメロディパートを各パートの音高の密度から抜粋し, そこから独自のアルゴリズムにしたがってコード進行を推定, それを楽曲にマイニングすることによって各楽曲のメロディ形式を決定している. この形式にしたがって楽曲の分類を行っている. これらの効果を実証するため, Kuo らは中国民謡・エンヤ・ショパン・バッハの楽曲を用いて分類実験を行っている. 結果, 中国民謡とバッハ間で正しく分類された割合が約 70 %, エンヤとショパン間での正答率が約 60 %, それ以外の組み合わせがいずれも 80 %以上という結果となっている.

また, 彼らはこれにしたがって分類された楽曲を推薦させるために三つの手法を考案している. ユーザが選択したカテゴリのみから推薦を行う Boolean Recommendation(BR), ユーザが過去に選択してきたカテゴリにランク付けを行い、そのランクが高いものから推薦を行う Total Rank Recommendation(TR), そして選んだ楽曲群のランクに加え, 埋め込まれたコード進行の順番も散光に推薦を行う Total Confidence Recommendation(TC) である. これら三つの手法の検証を行うために, Kuo らは中国民謡, アメリカのフォークソング, エルトン・ジョン, ビートルズ, ショパン, バッハというカテゴリーを事前に準備して実験を行った. 評価の指標は推薦楽曲中でいくつユーザが楽曲をダウンロードしたかで求められる推薦精度である. 結果、BR が他二つの手法に比べ高い精度を示した. そして TC はユーザが使い込むにつれ精度が上昇するという結果を残している. しかしながら、手法間における精度の差はほとんど見られない.

この年代において、メロディという音楽的特徴に着目し、グルーピングを行おうとした意図は評価できる。しかしながら、それに基づいて行われた推薦結果にほとんど違いが見受けられないために、コード進行の抜粋がうまく行われていたのか、音楽的特徴を本当に反映できているかという点には疑問が残る。

3.3 音響特徴を利用した音楽推薦システム

2000 年代に入ると, 大量の楽曲群を機能的に管理・検索するための研究分野である Music Information Retreival(MIR) が立ち上げられ, 活発に研究される様になった. この背景には音響信号処理とハードウェア, 両面での進歩があったからだと考えられている [46].

MIR の発展に伴い, 音響信号から直接音響特徴量を抽出し, それを基として楽曲の推薦を行う推薦システムの研究が行われるようになった.

以下で代表的な研究について触れていく.

3.3.1 Logan (2004)

Logan は, ある楽曲に関連のある楽曲群から抽出した音響特徴量を使用し, 楽曲の推薦を行う音楽推薦システムを提案した. 著者は公共で発表されたプレイリストや意見, ある

いは専門家によって付与されたメタデータを参照して推薦を行う音楽推薦システムでは, 聴き手にとって新しいアーティストの発見が難しいのでは, と考えている. そして, ユーザの求める楽曲の代表となるような楽曲から音響特徴を用いることによって効果的な音楽推薦が可能になるのでは, という仮説を立て, 研究を開始している.

Loganが研究中で用いている特徴量は, 音響信号から抽出された19次元のMel Frequency Cepstral Coefficients(MFCCs) である. これは楽曲の楽器編成や音色に関わりを持つ特徴量となる. これをK-means法によってクラスタリングしモデル化したあと, それぞれのクラスタの平均や共分散, ウエート (Weight) を学習させる. それぞれのモデルの比較はEarth-Mover's distance [33] によって行う. これは, 混合ガウス分布 (Gausian Mixture Model) 間の距離尺度をとる場合に用いられる距離尺度で, あるGMMの分布をもう片方のGMMに変化させるための最小コストを表したものである [44].

これを用いた推薦手法として, LoganはEMDの平均距離, Median距離, 最小距離の三つを用意しシステムの評価を行った, 使用した楽曲データベースは, 1523枚のアルバムから得られた18647曲の楽曲である, ここでのアルバムは、楽曲群として扱われるものである. システムへの学習には, データベースに使用された1523枚のアルバム中から無作為に一曲ずつ抽出された1523曲の楽曲を用いている. 評価尺度としては, あるアルバムの楽曲を選択したとき, 推薦システムが学習用の楽曲から同じアルバムに収録されている楽曲を推薦できたか, または違うアルバムだが同じアーティストによって演奏されている楽曲を推薦できたかという尺度を用いている. 著者らは前者を厳密に正しい楽曲推薦、後者を広義的に正しい楽曲推薦として扱っている。結果, 三つの手法の中では最小距離が最も正確に推薦を行うことができる, ということが判明した.

今研究は音響特徴を用いた音楽推薦を取り扱った研究では最初期にあたるものであり, 以降の研究の比較対象となる機会が多い. ユーザが求める音を示す楽曲群を用意することによって適切な推薦を行うことができる, という考えは, 音響特徴を利用した音楽推薦の基礎となる考え方である. しかしながら, 今回はその楽曲群にアルバムを使用しているために, その思考が本当に推薦の結果を反映しているかを判別することが難しくなっている.

3.3.2 Cano et al.(2005)

Canoらは, 楽曲のリズムやハーモニーといった音楽的特徴を用いることにより, メタデータが付けられていない楽曲でも推薦を行えるシステムを考案した. 著者らは比較対象としてメタデータ使用型および協調フィルタリング使用型の音楽推薦システム, そして楽曲の断片から該当する楽曲を推定するオーディオフィンガープリンティング [8] を比較対象としている. メタデータ型はユーザに知識がないと有効活用することができず, 協調フィルタリングは評価データを大量に必要とする上に楽曲の人気度によって結果に偏りが生じる. そしてオーディオフィンガープリンティングでは音響特徴のみを取り扱うために, 実際に発見された楽曲がどのようにユーザに受け止められるかを考慮していないと彼らは主張した.

彼らが推薦システムに使用した音楽特徴は多岐に渡る。まず彼らは聴き手が音楽を判断する際にはリズムや拍子という Temporal Organization と、音色や調性、メロディといった Pitch Organization の二つの概念から理解していると考えた。それを元に、テンポ、拍子、リズムパターン、リズムの揺れ (Swing) をリズムの構成要素とし、楽曲中の楽器の構成や調性、主となる調、その調が長調であるか短調であるか、などを取り扱うことを決定している。また、楽曲のスペクトル形状から音色を、音響信号から音量をそれぞれ抽出し、使用している。また、類似性の尺度としてジャンルも扱っている。

これらの音楽的特徴を音楽的な次元として扱い, その中で全体の類似性を求めることによって推薦を行うと Cano らは論文中で説明している. システムの評価においては ISMIR2004 [9] で用いられたフレームを使用し, その中からアーティスト特定能力を用いている. 結果として, 11257 人のアーティストによって作成された 273751 個の楽曲という大きな楽曲群を取り扱ったとしても, ISMIR2004 において評価されたアーティスト推定精度と同様の 24 %を達成することができた, と彼らは主張している.

単純な音響特徴ではなく, 音楽の内容を表す音楽的特徴を用いて音楽推薦を行ったという点は多く評価できる. しかしながら, その音楽特徴の抽出方法や推薦に使用した距離尺度について論文中で深く触れていないために, その結果の正当性が判断しにくいものとなっている. また, 評価尺度として用いたアーティスト推定精度が音楽推薦システムの評価にふさわしいかが疑わしい.

3.3.3 Bogdanov et al.(2010)

Bogdanov らは楽曲間の関連性と新規性を持ったコンテンツベースの音楽推薦を行うために, ユーザから提供された楽曲群から抽出した音楽的特徴, または音響的特徴を用いた音楽推薦システムを提案した.

使用した特徴量としては, 音響特徴と音楽的特徴の二種類がある. 音響特徴は MFCCs のみである. 音楽的特徴としては, ユーザからの楽曲群にサポートベクターマシンを使用することによって得られた, ジャンル, 文化, ムード, 楽器編成, ハーモニー, 表現, 歌詞, 楽曲の製作された年, 楽曲の長さなどの十四種の情報を用いている.

推薦手法としては五種類の提案を行っている. 大きく分類すると, 音楽的特徴間の距離をピアソンの相関係数を用いて計算した推薦手法 (Semantic Distance) と MFCCs のみを用いて推薦を行う手法 (Timbral Distance) である. 前者ではそれぞれユーザの評価の平均との距離 (Mean), すべての楽曲間の距離 (All), GMM モデルの比較 (GMM) を行い, 後者では距離と GMM を使用した手法を考案している. また, Bogdanov らはこれらの比較対象として, メタデータを使用した音楽推薦システムである Last.fm からの推薦を用意した.

システムの評価を行うにあたり, 彼らはジャンル, スタイル, アレンジ, 地域性など特徴的に幅広いアーティスト 4700 人による, 10000 個の楽曲群を用意した. これらの楽曲群から 30 秒間の抜粋を行い, それを推薦対象として扱っている. 推薦結果の評価として、ユーザがその楽曲に慣れ親しんでいるかを表す「親密性 (Familiarity)」と楽曲が好みであった

かを表す「好み (Liking)」という二つの指標を定義し, それぞれを 12 人のユーザに 0 から 4 の五段階で評価させた. 0 が最低値であり, 4 になるにしたがって高評価となる. そして、親密性が 2 未満かつ好みの評価が 2 より上であった場合を hit, 好みの評価が 3 未満であった場合を fail, 親密性が 1 より上かつ好みの評価が 2 より上であった場合を trust と定義し, これに当てはまる楽曲をいくつ推薦できたかで各手法の評価を行った. 結果, fail 率は Last.fm の推薦が最も低く, 次いで Semantic 系, Timbre 系の順で結果が悪くなっていった. hit 率は高い順に Last.fm, Semantic 系, Timbre 系となった. trust 率は Last.fm が群を抜いて高く, 他の手法の推薦結果はほぼ横並びであった.

この研究において高く評価すべき点として, Logan の考えを拡張したコンセプトや音楽的特徴の利用が挙げられる. Last.fm の推薦結果には劣っていたものの, MFCCs のみを用いた音楽推薦との比較では無視できない差が見られ, 音楽的特徴の効果を示す結果となっている. しかしながら, 新奇性を求めた楽曲推薦を目標に掲げていたにも関わらず, それに寄与するであろう推薦結果の楽曲の多様性に対する考察が欠けている. そのため, 新規性を獲得できたかどうかが不明確である.

3.4 ハイブリッド型音楽推薦システム

音楽推薦システムの研究は, 協調フィルタリングを用いたアプローチと音響特徴を用いたアプローチの二つが主流だった. しかしながら, その仕組み上, どちらか片方だけを選択すると回避できない問題があるということが明らかになっていった. 協調フィルタリングであれば楽曲に付けられた評価からくるコールドスタート問題であり, 音響特徴使用型ではユーザの嗜好分析が行いにくい問題がそれにあたる.

そのため, これら二つを組み合わせることによって互いの欠点を補おうとする, ハイブリッド型の音楽推薦システムの研究が行われるようになった.

現在ではこのようなハイブリッド型が音楽推薦システム研究の主流となっている. 以下では代表的な研究を挙げ, それらについて説明を加えていく.

3.4.1 吉井ら (2006)

吉井らは, 推薦の正確さ, アーティストのバラエティの豊富さ, 未評価楽曲の発見能力という三つの要件を満たした音楽推薦システムを実現するために, ユーザによる楽曲評価と楽曲の音響特徴を 3 方向アスペクトモデルと呼ばれるベイジアンネットワークを用いることによって確率的に統合したハイブリッド型音楽推薦システムを提案した.

音響特徴として, 彼らは Aucouturier らによって提案された多重音色 (Polyphonic Timbre) という概念を利用している [3]. これは楽曲中の個々の楽器の音色ではなく, 楽曲全体の「聴こえ」を表すものである. 混合音中の楽器音分離という困難な作業を必要としないために, さまざまな音響信号に適用することができる. この研究では 13 次元の MFCCs とエ

ネルギー, それらのデルタ成分という 28 次元の特徴量ベクトルから GMM 推定を行うことでこの多重音色を求めている.

推薦手法として吉井らが用いたのは, 3 方向アスペクトモデルと呼ばれるベイジアンネットワークである. これは本来, 文書推薦用に設計されたモデルであり、直接観測できない本質的なユーザの嗜好に対応する隠れ変数を備えているという特徴がある. 吉井らはこれに多重音色とユーザの評価, ジャンル, 楽曲, 楽曲の多重音色という四つの要素を確率的に統合するによって独自のモデルである Bag-of-Timbres モデルを作り上げた. システムの評価は協調フィルタリング型, および MFCCs を用いた音響特徴利用型の音楽推薦システムと比較して行われた. 使用した楽曲群, およびそれに付与された評価データは Amazon から収集したものを使用している. また, 今回の実験で扱われているジャンルはポップスのみである. 収集した楽曲を選定した結果, ユーザ数は 316 人, 楽曲数は 358 曲となった. 評価方法としては, 推薦システム上の評価とユーザとの評価が合致していたかで示される推薦精度とそれぞれのユーザごとのアーティストのバラエティである. アーティストのバラエティには, 推薦ランク上位中のアーティスト数と, そのユーザから評価を当てられていないアーティスト数という二通りの評価指標が用意されている. 結果として, 推薦精度では二つの手法を上回っていることが判明した. また, ランク上位中のアーティストバラエティにおいては音響特徴利用型に僅かに劣ったものの, 評価されていないアーティストの推薦ではやや高い値を示した. 未評価楽曲に関しては, システム上で評価を行えた楽曲の数が音響特徴利用型に対して少なかったものの, 適切に推薦が行えていると著者らは判断している.

本研究はハイブリッド型の音楽推薦システムに関する研究の中では先駆的な存在となっている. 推薦楽曲の多様性にもつながるアーティストのバラエティや未評価楽曲の発見能力など, ハイブリッド型に求められる要件を設定しているのは高く評価されるべきである. しかし, 今回実験に用いた楽曲がポップスのものだけである. それが楽曲の多様性にどれほどつながっているかは判断しにくい. また, アーティストのバラエティに関しては音響特徴を利用した音楽推薦システムの結果とほとんど違いが見られなかったり, 未評価楽曲の発見が本当に行えているか曖昧であったりと, システムの能力に関しては疑問が残る.

3.4.2 Li et al.(2006)

Li らは楽曲を音響特徴と付与された評価でグルーピングを行った上で, ユーザ評価をガウシアン分布ととらえ, それを元として推薦を行う手法を考案した. なお, この研究で提案された音楽推薦システムは携帯電話の着信メロディを推薦するためのものである. 研究時の問題意識として彼らは協調フィルタリングのために特徴量をどのように最適化するかを議論した研究が少ないことを挙げている.

使用する音響特徴として彼らは Tzanetakis らの研究 [40] から特徴量を選出している. これらの特徴量の詳細な説明は付録で行う. 音色に関連する特徴量としては音響信号を短時間フーリエ変換 (Short-Time Fourie Transform) にかけることによって得られるスペクト

ラルセントロイド, スペクトラルロールオフ, スペクトラルフラックス, 楽曲全体の音量分布 (原語ではSum of scale factor), そしてMFCCsである. リズムに関連する特徴量はビートヒストグラムから得られるA0,A1,RA, P1, P2 SUMを使用している. 最後に音高に関係する特徴量としては, ピッチヒストグラムから得られるピーク, 傑出したピーク二つの間の音高的距離, そしてヒストグラムの総和となる. Liらはこれらの音響特徴から高次的な音楽的特徴であるAggregate Featuresを導き出している. この特徴とユーザによる評価を元にアイテムコミュニティを作成し, そこから推薦を行っている. システムの評価実験のために, 二種類の楽曲データベースが用意された. 一つは328人のユーザ, 760曲の楽曲, そしてそれらに付与された4340個の評価データを格納したデータベースである. もう片方はEachMovieというサイトから抽出されたデータセットで, 1623本の映画, 61625人のユーザ, そしてそのユーザたちによって付与された二千八百万の評価データを格納したものである. これらのデータベースを使用し, 協調フィルタリングの代表的な研究であるSarwerの手法 [34] と比較実験を行った. 評価指標としては、推薦システムの評価とユーザの実際の評価から求められる推薦の正確さを表したMean Absolute Error(MAE) と, ユーザがどのランクまで参照したかを表すRank Scoreによって行われた. 結果として, どちらのデータベースにおいてもSarwarの手法よりMAE, Rank Scoreともに高い結果を残した. また, 彼らは使用する音響特徴に関しても実験を行い, その効果を検証している. その結果、特徴量単体ではMFCCs, スペクトラルロールオフ, そしてリズムの順でMAEによい結果を残している. また, それらを組み合わせることによってよりMAEを改善することが正銘されている.

この研究でもっとも評価すべきは, 特徴量の組み合わせに関して実験を行っている点である. 複数の音響特徴を使用した研究は多いものの, そのうちどの特徴量が良い影響を及ぼしているかを説明したものは少ない. しかしながら, ジャンルの多様性などを考慮していないという点ではまだ改善の余地があると思われる.

3.4.3 Lu et al.(2009)

Luらは、協調フィルタリングと音響特徴に加え, 楽曲の印象に関する情報を組み合わせたハイブリッド型音楽推薦システムを提案した. 彼らは商業的に多く用いられている協調フィルタリングを比較対象と捉え, 未評価楽曲の提案, 「不評だった楽曲」の再推薦防止, よりユーザが関心をもつ楽曲の推薦といった要件の達成を目標としている.

他のハイブリッド型推薦システムと異なる点として, Luらの研究では音響信号ではなくMucisXMLファイルから音響特徴の抽出を行っていることが挙げられる. このファイルは, スコアのある楽曲をXMLの形に直したものである.

抽出する音響特徴は以下のとおりである。まず音高に関係する特徴量として、Maximum Interval, Maximum Pitch, Minimum Pitch, Average Relative Pitch, Average Absolute Pitchがある。これらの特徴量は音高の分散を計算することで得られる。次に、拍数を楽曲の長さで割ることで得られるAverage Tempo, ファイル中の音符の最大数を楽曲の長さ

で割ることで得られる Rhythmic Speed, そして楽曲の調性, コード数, 調性の変化, 拍子の変化である.

また, 楽曲の印象情報は, ポジティブなものとして Hapiness, Love, Relief, Pride, Gratitude, Satisfaction の六種類と、ネガティブなものとして Sadness, Fear, Pity, Anger, Hatred, Disappointment, Shame, Loneliness, Anxiety の九種類の、計 15 種の印象語から決定される. データベースに格納される時点でそれぞれの楽曲は専門家によってアノテーションをつけられる. それ以降はユーザによる印象情報の付与が行われ, その都度フィードバックを行う.

推薦手法は, 協調フィルタリング, 音響特徴, そして印象情報それぞれで推薦アイテムのランク付けを行った後, それらを統合することで推薦結果を提示するという方法をとっている. 楽曲の評価は Favorite, Disfavorite, 無評価の三つで行われ, Disfavorite の評価が付けられた楽曲は推薦の対象から避けられるようになっている. 評価実験では 680 個の MIDI ファイルを元にデータベースの構築を行い, 27 名のユーザに実際に使用してもらい, 評価を行っている. 評価指標としては推薦楽曲中いくつの楽曲がユーザの嗜好に合ったか, から計算される推薦精度と, それぞれの推薦形式 (協調フィルタリング、音響特徴利用型、印象情報利用型) の推薦の重要性を用いている. 結果として, 提案された推薦システムは利用回数が増えるごとに推薦精度が上昇することがわかった. また, 二回目の推薦においては 90 %以上の精度で推薦を行えると著者らは主張している. 推薦の重要性では, 音響特徴よりも印象情報のほうが重要だと捉えられているという結果が得られた.

ユーザにとって不評だった楽曲の推薦防止や印象情報の取り入れなどはコンセプトとして評価できる. しかしながら, 音響特徴の抽出元が記述形式のファイルであるために, 新しい楽曲が導入されたときの対応が難しくなることが予想される. また, 印象情報として扱われた印象語に曖昧な定義のものが多いために信憑性が薄く, かつ数が多いために処理に時間がかかってしまう可能性がある. またユーザが関心を持てる楽曲を提供できたかを示す評価指標が用意されていない.

3.4.4 Shao et al.(2009)

Shao らは, ユーザのアクセスパターンと音響特徴を結びつけた音楽推薦システムを提案した. これは協調フィルタリングとコンテンツベースのフィルタリング両者の欠点を打ち消し, かつ推薦精度の高さと楽曲の新奇性を両立した音楽推薦を行うことを目的としたものである. なお, ここで言う新奇性とは, ユーザの嗜好の範囲で, 音楽的内容に多様性のある推薦が行える能力のことをいう.

彼らは, 違う音響特徴に対しても直接重み付けができることと個々のユーザによって音響特徴の捉え方が異なることを指摘した上で, 音響特徴の選択を行った. その結果, スペクトラルのセントロイド, ロールオフ, フラックスにゼロ交差率, ロウエネルギー (Low Energy), そして MCFFs という音色に関する音響特徴量に加え, Daubechies Wavelet Coefficient Histograms(DWCH) [29] と呼ばれる独自の音響特徴を採用している. DWCH は, もとも

と画像検索に使用されていたDaubechis Wavelet Filterから得られるものである.

推薦手法としては. ユーザのアクセスパターンと音響特徴を参考に, アクセスされていない楽曲への重み付けを行うSong Graphを生成する方法を考案している. 推薦の際は種となる楽曲をユーザに提示してもらい、Song Graphと照らし合わせて楽曲の推薦を行う. 評価実験にはhttp://www.newwisdom.netの利用者のプレイリストを元に, 様々なジャンル・アーティストの楽曲10000曲を用意している. そして, 同じアーティストのみを推薦する手法, アクセスパターンを使用した手法, 協調フィルタリングと音響特徴を組み合わせた手法の三つを比較対象とし, 提案手法の評価を行った. 評価指標はアーティストおよび音響特徴のバラエティ, そして各手法によって生成されたプレイリストがユーザの嗜好を予測できていたか、の三つである. 結果, アーティストと音響特徴のバラエティに関しては提案手法はハイブリッドと同等の良好な結果を残し, ユーザからの評価に関してはハイブリッドの結果を上回っていた.

新奇性やユーザ個々の嗜好を考慮した上でシステムの設計を行っているのは評価できる。評価実験も, ある程度の楽曲評価データが必要となる協調フィルタリングと音響特徴を組み合わせたハイブリッド型推薦システムに, 評価データではなくアクセスパターンという暗黙的な手法を用いた上でほぼ同等の楽曲の多様性を実現させている. しかし, アクセスされていない楽曲がどの程度推薦されているかが不明瞭である. また, ほぼ独自の特徴量といっていいDWCHがどのように音楽的内容に寄与しているかが不明となっている.

3.4.5 Domingues et al.(2013)

Dominguesらは、楽曲につけられた評価と, 音響特徴およびメタデータを組み合わせたハイブリッド型推薦システムを提案した. 協調フィルタリングとコンテンツベースのフィルタリング, 両方の欠点を補った上で, それぞれ独立したシステムよりも優れた推薦を行うことを目的としている. なお, この推薦システムはLast.fmなどでは取扱いの少ない楽曲を専門に扱ったサイトに実装され, 検証を行っている. そのため, ロングテール効果 [13] において人気度の低いアイテムを取り扱うことになる.

使用する音響特徴は, スペクトラルのセントロイド, ロールオフ, フラックス, そしてMFCCsという音色に関するもののみである. これらを使った推薦においてはユークリッド距離を用いて楽曲間の類似度を計測している.

提案された手法は以下の通りである。アイテム間の評価の類似性を用いた協調フィルタリングと、付与されたタグにLatent Semantic Analysis [17] を行い、それをコサイン距離を使って類似性を求めるメタデータ使用型に加え、先述した音響特徴利用型にそれぞれの推薦ランクを付与させる。そしてその三つのランクの平均をとり、それを推薦結果として扱う. また, ユーザの主観により不要な楽曲の推薦を除去することもできる.

評価実験には、実際にオンラインに実装して行われた。14777人のアーティストによる楽曲61000曲がデータベースとして用いられている。サイトに登録されたユーザは、先述したアーティストも含め約76000人だった。楽曲に付与するソーシャルタグはジャンル、

楽曲中の楽器、影響を受けたアーティストの三種に分類され、最低でも一つのタグが楽曲に付与されている。比較対象として用いられたのは楽曲に付与された評価を用いた協調フィルタリングと、音響特徴とメタデータを合わせたコンテンツベースのフィルタリングである。評価基準として、推薦システムを利用した回数Sとプレイリストに追加された楽曲数P、そしてブラックリストに追加された曲数Bを用いて以下のようなものが用意された。PとBの和をSで割って得られるActivity Rate, PをSで割って得られるAbsolute Acceptance Rate, そしてPをPとBの和で割って得られるRelative Acceptance Rateである。これらのサンプルは十一回に渡って採取された.

結果として, 提案手法はRelative Acceptance Rateでは他二つの手法とほとんど違いが見られなかった. しかしながら, Activity RateやAbsolute Acceptance Rateでは他二つに比べ突出した結果を残している. この結果から, 著者らは提案法がユーザに関心を抱かせる楽曲の推薦を実践できていると主張している.

この研究は音楽推薦システムに関するものの中ではほぼ最新のものとなっている. メタデータを合わせたコンテンツと協調フィルタリングの組み合わせは, 音響特徴だけでは把握できない音楽の背景も踏まえた推薦ができる可能性がある. しかし, 提案法がユーザの関心を向けるだけの推薦結果を提示できたかは、コンテンツの多様性といった観点から議論の余地があると考えられる. また, 今回はロングテール効果における人気度の低い部分のみを取り扱っていた. 本来であれば人気のある楽曲, ない楽曲とを混合した上で評価を行わないと, マイナーな楽曲の推薦が行えているかが判断できないと思われる.

第4章　考察

4.1　参考論文調査結果の統括

本章では, 前章で行ってきた参考論文の評価を元に, 現在の音響特徴を利用した音楽推薦システムでの課題の発見と改善策の検討を行う.

まず, これまでに参考にしてきた論文から, 現在の音楽推薦システムの研究における傾向を把握する. 前章では, 音響特徴を用いたグルーピングに関する研究から二つ, 音響特徴のみを利用した音楽推薦システムに関する研究から三つ, そしてハイブリッド型の音楽推薦システムに関する研究から五つ, 計十個の関連研究の概論を述べ, それぞれに評価を加えてきた.

研究のテーマから見るならば、現在の音楽推薦システムの主流は音響信号のみではなく, それらに他の形式の推薦システムを組み合わせたハイブリッド型の音楽推薦が現在における音楽推薦システム研究の中心となっていることがわかる. 音響特徴のみ, あるいは協調フィルタリングのみを取り扱って研究を進めても, その仕組みの根本からくる問題を解決することが難しいという共通認識が広がっていると推測できる. また, 音響信号を取り扱った研究ではほぼ例外なく音色に関連する音響特徴が用いられている (詳しくは付録参照). その中でも MFCCs は, 今回参考として挙げたそうした研究においてすべて使われている. 抽出のしやすさや関連性の明確さによるものだと間が得られる.

各研究分野について触れていく. 音楽のグルーピングは、例えば Li らによって手法の一部として組み込まれたり、Shao らによってアクセスパターンを利用する手法が採用されたりするように、他の研究の基礎となっていることはあるものの、それ自身に関する研究はほぼ見られなくなっている.

それに対して, 音響特徴のみを用いた音楽推薦システムに関する研究は細々と続けられている. 例えば, Bogdanov らの研究がハイブリッド型が研究の主流となっていた 2010 年に発表されている. このような音響特徴利用型の研究特有の動きとして, 低次の音響特徴ばかりでなく、高次の音楽的内容を扱って, 推薦結果の質を大きく上げようとする動きが上げられる. Cano らはリズムや音高に関連する音楽的内容を抽出しようと試みているし, Bogdanov らは楽曲の文化や年代などのような, 音楽の内容を裏打ちする背景的知識を利用している. このような音楽的内容を扱う研究がハイブリッド型に関するものにないのは, Dominugues らが言うように, 実践で扱うには処理に時間がかかりすぎてしまうからだと考えられる [19].

ハイブリッド型の研究では, 協調フィルタリングと音響特徴を組み合わせるだけではな

く、そこにメタデータを加えた研究が行われるようになっている [32] [19]. これによって音響特徴とユーザからの評価だけでは参照できない, 楽曲の雰囲気やアーティストの情報などを取り込もうという発想から来ているものだと思われる. しかしながら両者の研究を見るに, メタデータを加えたハイブリッド型音楽推薦システムの研究では、三つの評価をどのようにして統合するかという問題がある.

まとめると、音楽推薦システムの研究全体での主流はハイブリッド型となっている. ハイブリッド型の中では協調フィルタリングのみならず, メタデータも含めたコンテンツベースのフィルタリングとの組み合わせが考えられ始めている. 一方で、音楽的内容を利用する音響特徴利用型の研究がもう一つの流れを作っているように思われる. また使用される音響特徴は音色に関するものがほとんどであり, 特に MFCCs は必須ともいえる.

4.2 Serendipity 向上のための課題発見

ここでは, 一章で提示した予期せぬ発見を行う能力に寄与すると考えられる四つの要求項目について調査した研究を評価し, 解決すべき課題を検討する. 要求項目の内訳は, 音響信号への対応, 未評価楽曲への対応, 音楽的特徴を踏まえた音響特徴量の選択, そして推薦精度以外の評価軸の有無となっている.

4.2.1 音響信号への対応能力

まず第一に音響信号への対応能力である. 何も付属していない音響信号を取り込むことができるかを表す. これは新しい楽曲を素早く取扱いの対象にできるか, を示す能力となる. これはその名の通り, 音響信号を取り扱う前提の音楽推薦システムであれば満たすことのできる能力といえる. そのため, MIDI ファイルをしようした Chen らや Kuo らの研究, XML ファイルを用いた Lu らの研究は, 新しい楽曲情報からそれぞれの形式のファイルを作成する手間がかかるため, この能力を満たさない.

4.2.2 未評価楽曲への対応能力

第二に, 未評価楽曲への対応能力である. これはデータベース内に他ユーザからの評価を付与されていない楽曲がある場合でも, 推薦が可能であるかどうかを示すものである. 楽曲の人気度合いに左右されずにマイナーな楽曲を推薦する能力に対応している. これは抽出元のデータ形式を問わず, 音響特徴を用いて推薦を行う音楽推薦システムであれば満たす能力である. そのため, 今回調査した関連研究に関して言えば, 全ての提案手法がこの能力を満たしている. しかしながら, マイナーな楽曲を推薦する, という観点から考えた場合, 協調フィルタリングを導入しているハイブリッド型音楽推薦システムでは, 推薦対象となる楽曲の評価の付与度合いによって推薦結果に偏りが出る疑いがある. 例外とし

てはShaoらの提案手法で、これは明示的な評価データではなく暗示的なアクセスパターンを使用しているため、協調フィルタリングを使用したハイブリッド型の推薦システムと比較して人気度合いによる偏りが避けられるのではないかと予想される. 以上より, この能力に関しては音響特徴を使用した全ての提案手法が未評価楽曲の対応が可能であると考えられる. とはいえ, 協調フィルタリングを使用したハイブリッド型の提案手法に関しては, 楽曲の評価付与の度合いによって結果に偏りが現れる恐れがある.

4.2.3 音楽的特徴を踏まえた音響特徴の選択

第三に, 音楽的特徴を踏まえた音響特徴の選択が行えているかを判断していく. これは無作為に選ばれた音響特徴ではなく, 音楽的な内容を踏まえた上で選出された音響特徴が使われているかを判断するための項目である. ジャンル違いであっても, 共通した音響特徴を持った楽曲の推薦能力に対応している. 前述した通り, 音響信号から抽出した音響特徴を使用している音楽推薦システムはほとんどがMFCCsを使用している. MFCCsは楽曲の音色に対応した音響特徴であるため, これを用いた提案手法はこの要件を満たしている. 例外としては, Shaoらの研究で用いられていたDWCHがある. この特徴量は今回調査した研究の中でもここでしか用いられていない, ある意味独自性の高いものである. DWCHを提案したLiと荻原 [29] は、これは音楽信号の局所的・全般の情報を含んでいると主張しているものの, 明確にどの音楽的要素に関わるかを述べていない. 自動ジャンル分類において効果を上げていると示しているものの, 現在のままではかつてのMFCCsのように「理由はわからないけれど使うとうまくいく」特徴量として捉えられてしまう. 対応の確認・検討が望まれる. また、MFCCsを用いていない、すなわちChenらやLuらのような記述式のファイルから特徴量抽出を行っている手法に関しては、音高やテンポなどを組み合わせ方に正当性があるかを確認するのが好ましい。また特筆すべき点として、用いられている音響特徴がMFCCsを元とした音色に関する特徴量のみであることが挙げられる。音色に関連する特徴量は用いるものの、それ以外、たとえばTzanetakisら [40] が提案したリズム・音高に関する音響特徴群を用いた研究は少ない. すべての特徴量群を用いた例としてはLiらの研究がある. 彼らはそれぞれの音響特徴量を使って推薦を行った場合の推薦精度を試験しており, 結果として音色に関する音響特徴群だけでなく, リズムもまた推薦制度に寄与していることを示している. さらに, 彼らは音色に関する特徴量群とリズムを組み合わせることによって, 単体で使用した時よりもさらに精度を高めることができると発表している. 音色以外の音響特徴を導入し, それらが推薦に与える効果を検証することが必要である. 以上をまとめると, 音色という観点から見ると多くの提案手法が音楽的特徴を踏まえた音響特徴の選択ができている. しかし, 独自の特徴量の対応関係や記述形式から抽出した音響特徴の組み合わせ方, そして音色以外に関連する音響特徴の使用検討といった課題が見受けられる.

4.2.4 推薦精度以外の評価軸の有無

最後に, 推薦精度以外の評価軸の有無について評価する. これは推薦結果を推薦精度のみでなく, 楽曲の多様性のような評価軸から判断できているかを表したものである. 精度のみを優先し, 同じアーティストの楽曲ばかり, または違うアーティストによる同じ楽曲ばかりといった推薦結果の偏りを避けるために設けた項目である. 2000 年代の後半に入ると、研究者の多くは推薦精度だけでなく「ユーザが興味を持つような」楽曲の推薦を要件として掲げた。これは音楽が本質的に娯楽のために存在しており、正確さなどよりもどれだけ新鮮味のあるアイテムを推薦できるかがユーザの期待する結果だからと考えられる. しかし, 興味を持たせるために行った推薦を客観的に判断した例は少ない. 吉井らのアーティストバラエティや Shao らの音響特徴上のバラエティなどは, その例として挙げられる. しかし, 同じアーティストでも年代を経て曲想が変化するケースや, 同じアーティストだがデュオの相手が違うケースなどを考えると, アーティストバラエティは純粋に楽曲の多様性を示すとはいえない. そのため, Shao らの音響特徴上のバラエティのみが今回の要件を満たすものとする. もちろん、音響特徴上のバラエティが高かったからといってそれが Serendipity に直結するわけではない. 例えばそれだけを重視するのであれば, 無作為に楽曲を選んでユーザに提出すればすむ話なのである. それが効果的でないのは,Shao らが研究中で述べているように, 個々のユーザにとって音響特徴の許容範囲があるからに他ならない. 同様に, 何の根拠もなく楽曲を推薦したとしてもユーザはその結果を信頼しないだろう. 他のユーザの評価やジャンルを元としたメタデータ, あるいは音響特徴のような礎がないと, ユーザに推薦結果を受け入れてもらうことは難しい. まとめると、この条件を満たすのは推薦結果中の音響特徴のバラエティを評価基準とした Shao らの提案手法だけであった。しかしながら、この評価基準だけが必ずしも Serendipity の向上につながるわけではないということは念頭に置くべきである.

4.3 考えられる改善策

ここまでの関連研究の評価から, 音楽推薦システムの改善に向けた課題を発見する.

音響信号への対応という観点から見ると, 種類によっては音楽的特徴の特徴量化に前情報が必要となるため, 新しい音響信号が登場しても対応できない恐れがある. これを解決するためには, こうした楽曲に対応できる音楽的特徴量の調査が課題となる. もしくは音響情報からソーシャルタグを付ける諸研究 [37] [26] [39] を適用し, 音楽的特徴を特徴量化するために必要なメタデータを自動的に入手する, という手法も考えられる. Bogdanov らの研究が示すように, 音楽的特徴はうまく使えば単純に音響特徴を使う場合よりも高い精度を得ることができる. 推薦に使いやすい音楽的特徴の検討を行うべきである.

未評価楽曲への対応に関しては, 協調フィルタリングを導入したハイブリッド型音楽推薦において, 人気度による推薦結果の偏りが現れるかをまず確かめるべきである. 偏りが現れるということは, マイナーな楽曲推薦が推薦されにくくなることのつながる. こうし

た場合はShaoら [35] のようにアクセスパターンを導入を検討すべきかもしれない.

音楽的特徴を踏まえた音響特徴量の選択にという観点から見ると, 音色に関連するもの以外の音響特徴量が推薦結果に与える影響を調査する必要がある. また、それらを組み合わせることによってどのような影響が出るかも検討するべきである.

推薦精度以外の評価軸としては現時点では推薦結果中の楽曲の音響特徴面から見たバラエティがもっとも Serendipity に寄与するものだと考えられる。そのため、「ユーザに関心を持たせる」推薦を志すのであればまずはこの評価軸も推薦精度と並べて採用するべきである.

最後に、ユーザの個人情報や位置情報の推薦システムへの導入を提案する. Uitdenbogerdらは年齢や音楽的教育経験などが音楽推薦に役立つと述べている [41]. 同様に小田川らも自動車の走行状況や周囲の天候といった情報を音楽推薦の参考とする方法を提案している [50]. このように、ユーザ自身やその周囲の情報を推薦システムに学習させることで, システムがユーザの嗜好分析を行いやすくなる可能性がある.

第5章　結論

本研究では, 音楽推薦に求められる能力として思いがけない発見をもたらす能力 (Serendipity) を備えた音楽推薦システムの実現に向け, 関連研究の調査を行い, それを元に解決しなければならない課題の発見や改善策の検討を行った. まず商業用で多く用いられている協調フィルタリングを初め, メタデータ使用型, 音響特徴利用型, 協調フィルタリングと音響特徴利用型を組み合わせたハイブリッド型, そして発展的な研究である状況に基づいた音楽推薦とプレイリストの自動生成というそれぞれの概要を述べた. その後, 音響特徴を用いた音楽推薦システムの研究のため, 音響特徴量によるグルーピングを扱った研究, 音響特徴のみを利用した音楽推薦システムの研究, そしてハイブリッド型音楽推薦システムの研究という三つのテーマから十個の研究を選択肢、それぞれの調査と評価を行なった. それを元として, Serendipity の高い音楽推薦の実現に必要であると仮定した四つの評価項目から課題の発見と改善策の検討を行なった.

その結果, 音楽的特徴を特徴量として扱うために, 音響信号から直接推定できる音楽的特徴の検討, または自動タグ付けの研究の適用がまず挙げられた. 他にも協調フィルタリングを組み合わせたハイブリッド型音楽推薦において, 楽曲の人気によって結果が偏ることがあるかどうかの考察や音色に関する音響特徴以外の音響特徴使用の検討, そして評価軸として推薦結果中の楽曲の音響特徴量バラエティの導入, などが課題として発見された.

また、ユーザの年齢や音楽教育の有無といった個人情報や, ユーザの位置情報を音楽推薦システムに学習させ, ユーザの嗜好分析の補助に使用するという提案を行なった. これを実現させることができれば, ユーザが選んだ音楽のみでなく, ユーザ自身の情報やユーザを取り巻く状況から最適な音楽を選択して推薦するという, 人間に近い推薦を音楽推薦システムに行なわせることが可能となると考えられる.

付 録A　音響特徴量の説明

音響特徴量とは、音響信号を自動解析することによって得られるデータのことである [44]. 音響信号から音響特徴量を抽出することにより, 音響信号を計算機に処理させることが可能となる.

これは楽曲を取り扱う場合でも同様となる. 楽曲の内容を楽曲から音響特徴量を抽出することによって, 計算機に楽曲同士の比較を行なわせることができるようになるのである. 音楽推薦システムにコンテンツベースのフィルタリングを導入しようとした場合, 音響特徴量は必要不可欠なものとなる.

ここでは, 人間が楽曲を聴いたときに内容を判断するための要素を「音楽的」特徴, 楽曲を自動解析することによって得られる音響特徴量を低次音響特徴量と定義し, それぞれについて説明を加えていく.

A.1　「音楽的」特徴

「音楽的」特徴とは, 音楽の内容を描写する概念のことである [12]. 英語では High-Level Music Content Description [12] や Semantic [22] といわれることが多い. 具体的な例を挙げると、リズムやハーモニー、メロディといった要素である。Casey らはこれらの要素に加え、音色や調、楽曲の構造なども「音楽的」特徴に含まれると説明している [12].

「音楽的」特徴を取り扱った研究で挙げられるのがハミング検索である [24] [48] [47] [49]. ユーザに検索したい楽曲のメロディをハミングしてもらい、それをキーに検索を行うというものである。しかしながら、コンピュータ自身にポリフォニー様式楽曲からメロディラインを抽出させるのは未だに難しいといわれている [12]. メロディのみならず, 音楽的特徴の多くはコンピュータに直接判断させることが難しい. そのために, 現在では楽曲から抽出したスペクトルレベルの特徴量を用いることが主流となっている.

しかしながら, 低次の音響特徴では, 楽曲の音響的な特徴を捉えることはできても, リズムの構造やコード進行といった音楽内容を考慮することができない. そのため, 今後の音楽推薦システムの発展のためには, このような音楽的内容を特徴量化する技術に関する研究が不可欠だと主張する研究者もいる [44]. こうした高次の音楽的特徴を特徴量化する研究を行っているものとしては, Bogdanov ら [7] や Herrera ら [22], Dittmar [18] らなどがいる.

A.2 低次音響特徴量

先述した「音楽的」内容を高次の特徴として取り扱うとき、音響信号を解析して得られた音響特徴量は低次音響特徴量と呼ばれる。こうした音響特徴は高速フーリエ変換 (FFT) によって抽出されたスペクトラルから求められるものが多い [12].

音響特徴量は楽曲のメロディ・リズムといった音楽の一側面の情報を含んでいる。MIR 分野の研究が始まってから現在にいたるまで、このような音響特徴量の抽出に関する研究は枚挙に暇がない。これらの研究を取りまとめて紹介したものとしては、Casey らの研究がある [12].

現在の音楽推薦システムにおいては、Mel Frequency Cepstral Coefficients(MFCCs) と Tzanetakis らの研究 [40] で提唱された音響特徴量が多く用いられている. 以下ではその二つの音響特徴について説明を加えていく.

A.2.1 MFCCs: Mel Frequency Cepstral Coefficients

MFCC は、音声信号処理のために使われていた, 短時間フーリエ変換によって求められる音響特徴量である。1990 年代後半から楽曲の分析にも用いられるようになったものの、なぜ MFCCs が音楽情報処理に有効であるかは言及されていなかった。2000 年になって、Logan らが初めて MFCCs の音楽情報処理に対する有用性について言及した.

音響信号から MFCCs を抽出するには以下の工程が必要となる [31]。まず第一に音響信号を短時間フレームで分割し, 窓関数を利用して角を除去する. 次に, それぞれのフレームを離散的フーリエ変換 (Discrete Fourier Transform) し, 振幅スペクトラム (Amplitude Spectrum) を入手する. その振幅スペクトラムのログをとった後, スペクトラムを平滑化し, ピッチ, または周波数に対する主観尺度であるメル尺度に直す. そしてこのメルスペクトラムを GMM によってモデル化しベクトルを得た後, 離散的コサイン変換 (Discrete Cosine Transform) することによって 13 個, もしくはそれ以上の数の MFCCs を入手することができる.

MFCCs はスペクトラルの形状を表す特徴量であるために, ピッチやテンポといった音響特徴よりも音響信号の音色と関連が深い. 音響信号からの抽出しやすさもあり, 多くの音響特徴を用いた音楽推薦システムに使用されている.

A.2.2 Tzanetakis et al. の提案した音響特徴量

Tzanetakis らが行っていた研究は, 元々は自動ジャンル推定に関するものである [40]. 彼らはここで, 当時研究の進んでいなかったリズム・音高を含めた三つの特徴量群を定義し, それによって効率的な自動ジャンル分類を行っていた.

彼らが定めた特徴量群は, 音色に関するもの, リズムに関するもの, そして音高に関するものの三つである. 以下でそれぞれの特徴量群についての説明を述べていく.

音色に関する特徴量群は, 音響信号を短時間フーリエ変換して求められるものである. まずスペクトラルセントロイド (Spectral Centroid) は短時間フーリエ変換によって得られたマグニチュードスペクトラム (Magnitude Spectrum) の重心の位置である. これは音響信号の明るさに対応している. 次にスペクトラルロールオフ (Spectrul Rolloff) がある. これはスペクトラムの形状を表す指標の一つで, マグニチュード分布における 85 %より下の周波数を指す. スペクトラルフラックス (Spectrul Flux) は連続的なスペクトラル分布から得た, 標準化されたマグニチュード間の二乗距離である. 局所的なスペクトラムの変化を表す. そして, フレームごとの時間領域で何度 0 の値をとったかを示すゼロ交差率 (Zero Crossings) がある. これは音響信号がどれだけうるさいかを表す指標となっている。また, 長いスパンでの音響特徴の性質を把握するために, 短時間フーリエ変換で使われるフレームより大きな窓関数であるテクスチャー窓 (Texture Window) が使用されている. そのテキスチャー窓の平均的なエネルギーより, 小さなエネルギーを持った分析窓の割合から求められるロウエネルギー (Low-Energy Feature) も, 音色に関係する音響特徴として用いられる. Tzanetakis らは先述した MFCCs を含めたこれらの音響特徴量を音色に関わる特徴量群と定義している.

リズムに関する音響特徴群は, ビートヒストグラムから抽出される. ビートヒストグラムは音響信号をオクターブの周波数分に離散的ウェーブレット変換 (Discrete Wavelet Transform) し, 帯域ごとに時間領域での周波数エンベロープを求め, それらを統合することで得られる. そこから抽出できる音響特徴は以下の通りである. 得られたビートヒストグラムの一番目のピークの振幅 A0 と, 二番目のピークから得られる振幅 A1. A1 を A0 で割ることで得られる Ratio of the Amplitude(RA) に, 第一ピーク・第二ピーク区間の Beats Per Minutes(BPM) である P1・P2 に, ヒストグラムの総和である SUM である.

最後に、音高に関する音響特徴群について述べる. これらの音響特徴はピッチヒストグラム (Pitch Histgram) から抽出される. ピッチヒストグラムには調性のある楽曲に対応した Folded Pitch Histgram(FPH) と, そうした対応を考えない Unfolded Pitch Histgram の二種類がある. これらのピッチヒストグラムから抽出できる音響特徴は以下の通りである. FPH の中の最高ピークの振幅 FA0. これは楽曲のメロディに対応した音響特徴でもある. そして FPH・UPH の最高ピークの期間である FP0・UP0. そして FPH 中で, 他のピークよりも突出している二つのピーク間の間隔 IPO1. 最後に, それぞれのヒストグラムの総和である SUM となる.

Tzanetakis らが提唱したこれらの音響特徴群は, 楽曲への関連性がわかりやすいためにしばしば音響特徴を用いた音楽推薦システムに関する研究で取り上げられる. 三つの音響特徴群をすべて使用した例としては Li らの研究 [28] があり, 音色に関連する音響特徴を用いた例としては Shao ら [35] や Domingues ら [19] の研究がある. 一般に, 音色に関連した音響特徴が特に多く音楽推薦システムに関する研究で利用されている.

参考文献

[1] J-J. Aucouturier and F. Pachet. Representing musical genre: a state of the art. *Journal of New Music Research*, Vol. 32, No. 1, pp. 83–93, 2003.

[2] J-J Aucouturier and Francois Pachet. Scaling up music playlist generation. In *Multimedia and Expo, 2002. ICME'02. Proceedings. 2002 IEEE International Conference on*, Vol. 1, pp. 105–108. IEEE, 2002.

[3] Jean-Julien Aucouturier, François Pachet, and Mark Sandler. The way it sounds": timbre models for analysis and retrieval of music signals. *IEEE Transactions on Multimedia*, Vol. 7, No. 6, pp. 1028–1035, 2005.

[4] Luke Barrington, Reid Oda, and Gert RG Lanckriet. Smarter than genius? human evaluation of music recommender systems. In *ISMIR*, Vol. 9, pp. 357–362, 2009.

[5] Chumki Basu, Haym Hirsh, William Cohen, et al. Recommendation as classification: Using social and content-based information in recommendation. In *AAAI/IAAI*, pp. 714–720, 1998.

[6] Dmitry Bogdanov, Martín Haro, Ferdinand Fuhrmann, Emilia Gómez, and Perfecto Herrera. Content-based music recommendation based on user preference examples. In *1st Workshop On Music Recommendation And Discovery (WOMRAD), ACM RecSys, 2010, Barcelona, Spain*, 2010.

[7] Dmitry Bogdanov, Joan Serrà, Nicolas Wack, Perfecto Herrera, and Xavier Serra. Unifying low-level and high-level music similarity measures. *Multimedia, IEEE Transactions on*, Vol. 13, No. 4, pp. 687–701, 2011.

[8] Pedro Cano, E Batle, Ton Kalker, and Jaap Haitsma. A review of algorithms for audio fingerprinting. In *Multimedia Signal Processing, 2002 IEEE Workshop on*, pp. 169–173. IEEE, 2002.

[9] Pedro Cano, Emilia Gómez, Fabien Gouyon, Perfecto Herrera, Markus Koppenberger, Beesuan Ong, Xavier Serra, Sebastian Streich, and Nicolas Wack. Ismir 2004 audio description contest. *Music Technology Group of the Universitat Pompeu Fabra, Tech. Rep*, 2006.

[10] Pedro Cano, Markus Koppenberger, and Nicolas Wack. Content-based music audio recommendation. In *Proceedings of the 13th annual ACM international conference on Multimedia*, pp. 211–212. ACM, 2005.

[11] Pedro Cano, Markus Koppenberger, and Nicolas Wack. An industrial-strength content-based music recommendation system. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 673–673. ACM, 2005.

[12] Michael A Casey, Remco Veltkamp, Masataka Goto, Marc Leman, Christophe Rhodes, and Malcolm Slaney. Content-based music information retrieval: Current directions and future challenges. *Proceedings of the IEEE*, Vol. 96, No. 4, pp. 668–696, 2008.

[13] Òscar Celma and Pedro Cano. From hits to niches?: or how popular artists can bias music recommendation and discovery. In *Proceedings of the 2nd KDD Workshop on Large-Scale Recommender Systems and the Netflix Prize Competition*, p. 5. ACM, 2008.

[14] Òscar Celma, Miquel Ramírez, and Perfecto Herrera. Foafing the music: A music recommendation system based on rss feeds and user preferences. In *in ISMIR*. Citeseer, 2005.

[15] Hung-Chen Chen and Arbee LP Chen. A music recommendation system based on music data grouping and user interests. In *CIKM*, Vol. 1, pp. 231–238, 2001.

[16] William W Cohen and Wei Fan. Web-collaborative filtering: Recommending music by crawling the web. *Computer Networks*, Vol. 33, No. 1, pp. 685–698, 2000.

[17] Scott C. Deerwester, Susan T Dumais, Thomas K. Landauer, George W. Furnas, and Richard A. Harshman. Indexing by latent semantic analysis. *JASIS*, Vol. 41, No. 6, pp. 391–407, 1990.

[18] Christian Dittmar, Christoph Bastuck, and Matthias Gruhne. Novel mid-level audio features for music similarity. In *International Conference on Music Communication Science*, pp. 38–41, 2007.

[19] Marcos Aurélio Domingues, Fabien Gouyon, Alípio Mário Jorge, José Paulo Leal, João Vinagre, Luís Lemos, and Mohamed Sordo. Combining usage and content in an online recommendation system for music in the long tail. *International Journal of Multimedia Information Retrieval*, Vol. 2, No. 1, pp. 3–13, 2013.

[20] Martin Gasser and Arthur Flexer. Fm4 soundpark: Audio-based music recommendation in everyday use. In *Proc. of the 6th Sound and Music Computing Conference (SMC 2009), Porto, Portugal*, 2009.

[21] Jonathan L Herlocker, Joseph A Konstan, Loren G Terveen, and John T Riedl. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, Vol. 22, No. 1, pp. 5–53, 2004.

[22] Perfecto Herrera, Juan Bello, Gerhard Widmer, Mark Sandler, Òscar Celma, Fabio Vignoli, Elias Pampalk, Pedro Cano, Steffen Pauws, and Xavier Serra. Simac: Semantic interaction with music audio contents. In *Integration of Knowledge, Semantics and Digital Media Technology, 2005. EWIMT 2005. The 2nd European Workshop on the (Ref. No. 2005/11099)*, pp. 399–406. IET, 2005.

[23] Keiichiro Hoashi, Kazunori Matsumoto, and Naomi Inoue. Personalization of user profiles for content-based music retrieval based on relevance feedback. In *Proceedings of the eleventh ACM international conference on Multimedia*, pp. 110–119. ACM, 2003.

[24] Jyh-Shing Roger Jang and Hong-Ru Lee. Hierarchical filtering method for content-based music retrieval via acoustic input. In *Proceedings of the ninth ACM international conference on Multimedia*, pp. 401–410. ACM, 2001.

[25] Fang-Fei Kuo and Man-Kwan Shan. A personalized music filtering system based on melody style classification. In *Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on*, pp. 649–652. IEEE, 2002.

[26] Mark Levy and Mark Sandler. Music information retrieval using social tags and audio. *Multimedia, IEEE Transactions on*, Vol. 11, No. 3, pp. 383–395, 2009.

[27] Qing Li, Byeong Man Kim, Dong Hai Guan, et al. A music recommender based on audio features. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 532–533. ACM, 2004.

[28] Qing Li, Sung Hyon Myaeng, and Byeong Man Kim. A probabilistic music recommender considering user opinions and audio features. *Information processing & management*, Vol. 43, No. 2, pp. 473–487, 2007.

[29] Tao Li, Chengliang Zhang, and Mitsunori Ogihara. A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression. *Bioinformatics*, Vol. 20, No. 15, pp. 2429–2437, 2004.

[30] Beth Logan. Music recommendation from song sets. In *ISMIR*. Citeseer, 2004.

[31] Beth Logan and Ariel Salomon. A music similarity function based on signal analysis. In *ICME*, 2001.

[32] Cheng-Che Lu and Vincent S Tseng. A novel method for personalized music recommendation. *Expert Systems with Applications*, Vol. 36, No. 6, pp. 10035–10044, 2009.

[33] Yossi Rubner, Carlo Tomasi, and Leonidas J Guibas. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, Vol. 40, No. 2, pp. 99–121, 2000.

[34] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, pp. 285–295. ACM, 2001.

[35] Bo Shao, Mitsunori Ogihara, Dingding Wang, and Tao Li. Music recommendation based on acoustic features and user access patterns. *Audio, Speech, and Language Processing, IEEE Transactions on*, Vol. 17, No. 8, pp. 1602–1611, 2009.

[36] Upendra Shardanand and Pattie Maes. Social information filtering: algorithms for automating " word of mouth ". In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 210–217. ACM Press/Addison-Wesley Publishing Co., 1995.

[37] Mohamed Sordo, Cyril Laurier, and Oscar Celma. Annotating music collections: How content-based similarity helps to propagate labels. In *ISMIR*, pp. 531–534, 2007.

[38] Marco Tiemann, Steffen Pauws, and Fabio Vignoli. Ensemble learning for hybrid music recommendation. In *ISMIR*, pp. 179–180, 2007.

[39] Derek Tingle, Youngmoo E Kim, and Douglas Turnbull. Exploring automatic music annotation with acoustically-objective tags. In *Proceedings of the international conference on Multimedia information retrieval*, pp. 55–62. ACM, 2010.

[40] George Tzanetakis and Perry Cook. Musical genre classification of audio signals. *Speech and Audio Processing, IEEE transactions on*, Vol. 10, No. 5, pp. 293–302, 2002.

[41] Alexandra L Uitdenbogerd and Ron G Schyndel, van Schyndel. A review of factors affecting music recommender success. *ISMIR*, Vol. 2, pp. 204–2008, 2002.

[42] Fabio Vignoli and Steffen Pauws. A music retrieval system based on user driven similarity and its evaluation. In *ISMIR*, pp. 272–279. Citeseer, 2005.

[43] 梶克彦, 平田圭二, 長尾確. 状況と嗜好に関するアノテーションに基づくオンライン楽曲推薦システム. 情報処理学会研究報告, Vol. 127, pp. 33–38, 2004.

[44] 吉井和佳, 後藤真孝. 音楽推薦システム. 情報処理, Vol. 50, No. 8, pp. 751–755, 2009.

[45] 吉井和佳, 後藤真孝, 駒谷和範, 尾形哲也, 奥乃博. ユーザの評価と音響的特徴との確率的統合に基づくハイブリッド型楽曲推薦システム. 情報処理学会 音楽情報科学研究会 研究報告, pp. 45–52, 2006.

[46] 後藤真孝, 平田圭二. 音楽情報処理の最近の研究. 日本音響学会誌, Vol. 60, No. 11, pp. 675–681, 2004.

[47] 獅々堀正幹, 大西泰代, 柘植覚, 北研二. Earth mover's distance を用いたハミングによる類似音楽検索手法. 情報処理学会論文誌, Vol. 48, No. 1, pp. 300–311, 2007.

[48] 小杉尚子, 櫻井保志, 山室雅司, 串間和彦. Soundcompass: ハミングによる音楽検索システム. 情報処理学会論文誌, Vol. 45, No. 1, pp. 333–345, 2004.

[49] 小杉優, 伊藤彰則, 伊藤仁, 牧野正三. 複数 f0 候補を用いた音楽音響信号からのハミング楽曲検索. 情報処理学会研究報告.[音楽情報科学], pp. 1–6, 2010.

[50] 小田川智, 児玉泰輝, 莪山真一, 鈴木康悟, 松下文雄, 塩田岳彦. 楽曲レコメンドシステム. 電子情報通信学会技術研究報告. EA, 応用音響, Vol. 106, No. 160, pp. 49–53, 2006.

[51] 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.

[52] 仲村哲明, 川西紘平, 坂本真樹. 歌詞と色彩に基づいた楽曲推薦の可能性 (ヒューマンコミュニケーション,¡ 小特集¿ 電子情報通信における萌芽的研究論文). 電子情報通信学会論文誌. A, 基礎・境界, Vol. 94, No. 2, pp. 85–94, 2011.

[53] 土方嘉徳. 嗜好抽出と情報推薦技術. 情報処理, Vol. 48, No. 9, pp. 957–965, 2007.