

音楽音響信号理解に基づく能動的音楽鑑賞インタフェース

後藤 真孝

産業技術総合研究所

m.goto [at] aist.go.jp

あらまし 本稿では、音楽音響信号の様々な要素を理解できる技術を応用することで、エンドユーザの音楽の聴き方をどのように豊かにできるかを探求する我々の研究アプローチを紹介する。我々はその研究事例として、7種類の能動的音楽鑑賞インタフェースを実現してきた。能動的音楽鑑賞とは、通常の受動的な鑑賞とは異なり、能動的にインタラクションしながら音楽を鑑賞する方法である。例えば、楽曲構造を見ながら楽曲内で興味のない区間を簡単に飛ばしたり、楽曲中の楽器の音色を変更したり、音楽コレクションをブラウジングして興味のある楽曲やアーティストに出会ったりするためのインタフェースを開発した。これらを通じて、音楽理解技術の重要性とそれが実際にエンドユーザの役に立つことを示す。

Active Music Listening Interfaces Based on Understanding of Musical Audio Signals

Masataka Goto

National Institute of Advanced Industrial Science and Technology (AIST)

Abstract This paper introduces our research approach aimed at enriching end-users' music listening experiences by applying technologies for understanding musical audio signals. Toward this goal, we have built seven *active music listening interfaces*. Active music listening is a way of listening to music through active interactions. For example, we have developed interfaces for skipping sections of no interest within a musical piece while viewing a graphical overview of the entire song structure, for changing the timbre of instrument sounds during music playback, and for browsing a large music collection to encounter interesting musical pieces or artists. These interfaces demonstrate the importance of music-understanding technologies and the benefit they offer to end users.

1 はじめに

「音楽との能動的なインタラクション」というと、通常は、作曲家や歌手、演奏者のような音楽家（音楽の作り手）だけが可能な楽しみだと考えられがちである。これは、「能動的」という言葉が、音楽の創作を暗示するからである。一方、音楽家でない普通の人達は、音楽を鑑賞したり、BGMとして音楽を聞いたり、ラジオやテレビで音楽番組を聴いたりするだけで、音楽を受動的に聴いて楽しむことが多い。しかし、必ずしも完全に受動的とは限らない。例えば、音楽音響信号の記録が可能となる以前は、聴き手は生演奏された楽曲をその場で聴くことができなかったが、拍手や手拍子、歓声等によって演奏者へ能動的に働きかけることがあった。その後、カセットテープやコンパクトディスク(CD)のような各種オーディオ用記録媒体へ自分で好きなように録音できるようになると、再生曲順を自分で決定したり、高音質（ハイファイ）を追求したり、グラフィックイコライザや音

質調整つまみで周波数特性を調整したりする、新たな能動的な楽しみ方も生まれてきた。

近年は、技術のさらなる進歩が音楽鑑賞の形態を変えつつある。計算機の高性能化やインターネットの普及、安価な音響入出力デバイスの浸透により、CDから楽曲をリッピングしたり音楽配信サービスを利用したりして、個人の計算機上で多数の楽曲を容易に扱えるようになった。さらに、数万曲を携帯型音楽プレーヤに入れて、いつでもどこでも手軽に聴けるようになった。一方、こうしたエンドユーザの環境の変化に加え、音楽音響信号理解技術も進歩してきており^{1)~4)}、様々な楽器音や歌声を含む実世界の複雑な混合音が扱えるようになってきた。こうした両者の技術の進歩は、音楽家でない一般のエンドユーザ（聴き手）の音楽との関わり方に大きな影響を与え、仮に音楽を創作できなくても、様々な方法で音楽と能動的にインタラクションすることが可能になりつつある。

そこで我々は2003年に、従来の受動的な鑑賞とは違う、能動的な音楽鑑賞を可能にするエンドユーザ向

け音楽インタフェースを実現する研究アプローチを提唱した⁵⁾。そして、これを「**能動的音楽鑑賞インタフェース**」(active music listening interface)と名付けて⁶⁾、音楽音響信号理解に基づく様々な事例を実現する研究を進めている。ここでの「能動的」という言葉は、音楽の創作は意味せず、音楽鑑賞を楽しむ上でのあらゆる能動的なインタラクション(音楽再生中の新たな支援や加工、楽曲の検索やブラウジング等)を意味する。例えば、能動的音楽鑑賞インタフェースにより、楽曲中の興味のない区間を容易に飛ばしながら好きな部分だけ楽しんだり、楽曲中の楽器の音色を再生中に別の楽器に置き換えたり、音楽コレクションをブラウジングしながら未知の楽曲と出会ったりすることが可能になる。

本稿では以下、「音楽音響信号理解研究が音楽の聴き方をどのように豊かにできるか」を、能動的音楽鑑賞インタフェースの具体例を示しながら議論していく。

2 能動的音楽鑑賞

エンドユーザによる音楽鑑賞には様々な側面があるため、「能動的音楽鑑賞」の研究で能動的なインタラクションを実現していく際にも、様々な方向性が考えられる。ここでは、音楽再生、音楽加工(タッチアップ)、音楽検索・ブラウジング(閲覧)の三つの代表的な観点から、音楽音響信号理解に基づく能動的音楽鑑賞の潜在的な可能性を、具体例を挙げて議論する。

2.1 音楽再生

ユーザが音楽を再生する際には、再生位置を変更したり、それに伴う映像を見たり、歌詞に着目したりすることがある。そこで、これらの現状を踏まえた上で、音楽音響信号理解によって、音楽再生中にどのような能動的なインタラクションが可能になるのかを考える。

- 再生位置の変更

音楽再生中に可能な最も単純なインタラクションの一つは、音楽プレーヤの「次の曲へ頭出し」のボタンを押して、興味のない曲を飛ばす操作である。音楽の録音が可能となる以前は、興味のない曲を飛ばすだけでも困難であったが、現在はCDプレーヤやメディアプレーヤ上で容易に操作できる。しかし、楽曲内部で興味のない箇所を的確に飛ばすことは難しかった。

そこで、音楽CDから楽曲の構造を自動推定すれば、構造上の区間単位でそうした興味のない箇所

を自由に飛ばすような、能動的なインタラクションが可能となる。

- 楽曲再生に伴う映像

MTV (Music Television) 等によって音楽のプロモーションビデオが普及したように、楽曲の再生に映像も加わると、より音楽に没入したり楽しんだりできる。そのため、近年の計算機上のメディアプレーヤでは、音楽に同期したアニメーションを表示できる視覚エフェクト(ビジュアライザ)機能が提供されることが多い。これは、楽曲再生中にその波形や周波数帯域ごとのパワー等に基づいて、幾何学図形のようなアニメーションを表示する機能である。場合によっては、その表示内容をキーボード等を押してリアルタイムに変化させることもできて、一種の能動的なインタラクションとなっている。

その際、単なる波形やパワーでなく、より高次の音楽的要素を推定すれば、楽曲の内容を的確に反映したアニメーションをユーザは楽しむことができる。例えば、音楽音響信号中のビートを自動推定することで、そのビートに同期したアニメーションが表示可能となる。

- 歌詞

歌詞は歌を伴う楽曲において重要な役割を果たしており、楽曲のメロディーやアレンジより重視する人も多く、能動的なインタラクションを考える上で欠かすことができない。単に音楽を聴くだけでなく、印刷あるいは表示された歌詞を見ながら音楽を聴くと、聞き取りにくい歌詞の一部がわかったり、その歌詞が伝えるメッセージをより深く理解できることがある。さらに、より能動的な楽しみ方として、楽曲再生に合わせて自分で歌詞を歌いながら楽しむこともできる。しかし、現在歌詞のどこが再生されているかは、通常は自分で目で追わなければわからなかった。

そこで歌詞を楽曲と自動的に同期させることで、原曲を聴いている場合であっても、カラオケ用装置等を使っているときのように、再生に合わせて歌詞の色を変えることができる。仮に携帯電話のような小さな画面でも、スクロールしながら表示すればそうした歌詞表示は可能である。逆に、画面上に表示された歌詞を見ながら、ある単語をユーザがクリックすることで、その単語から即座に再生するようなこともできる。現在の技術では難しいが、仮にもし将来、歌詞が他の言語に自動的に翻訳できるようになれば、外国語の楽曲の再生に

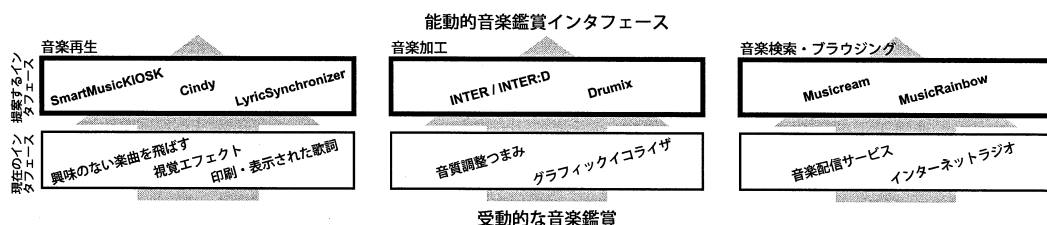


図1: 本研究アプローチの位置付け: 音楽音響信号理解に基づく7種類の能動的音楽鑑賞インタフェース

同期して、翻訳された母国語の歌詞も表示可能となるかも知れない。

2.2 音楽加工（タッチアップ）

音楽を聴きながら、もしこの歌が他の歌手によって歌われていたらどうなるか、この楽器が他の演奏者によって演奏されていたらどうなるか、このパートが違うアレンジで演奏されていたらどうなるかを聴いてみたいという願望を抱くかも知れない。そうした願望の一部は、アーティストによって、カバー曲やリミックスバージョンとして提供されて実現することがある。しかし、既存の楽曲に対して、ミックスダウン前の個々のトラック（歌声や楽器音）の信号が提供されるような特殊なケースを除いて、通常はエンドユーザがそうした願望を実現することはできない。

そこで、メロディーの歌声やドラム音のトラックが音楽CDから個別に推定できれば、ユーザがそれらの音量を能動的に変更して楽しむことが可能となる。本研究では、エンドユーザによるそうした能動的な変更を、音楽加工あるいは音楽タッチアップ (music touch-up) と呼ぶ。こうした軽微で個別な変更を可能にする機能は、音楽のパーソナライズ（個人化）機能あるいはカスタマイズ機能と捉えることもできる。従来は、グラフィックイコライザや高域と低域の音質調整つまみで周波数特性を調整するぐらいしか、エンドユーザが楽曲再生中に加工する手段はなかった。

2.3 音楽検索・ブラウジング（閲覧）

近年のダウンロード型の音楽配信サービスや聴き放題の定額制音楽配信サービスで提供されているような数百万曲の楽曲をユーザがすべて把握することは、不可能に近い。そこで、楽曲やアーティストを検索したり、そうした大規模な音楽コレクションをブラウジングしたりすることで、ユーザが聴きたい楽曲を能動的に見つけて聴くことができるインタフェースが必要となる。一般に音楽配信サービスでは、曲名やアーティスト名、ジャンル等の書誌情報を指定して検索をしたり、似ているアーティストをリストアッ

プしたり、個人の好みに応じたインターネットラジオを聴いたりすることができる。そうした既存の商用サービスの多くは、テキストの書誌情報（メタデータ）や、人手で作成したアノテーション（楽曲に対するラベル付け）、ユーザの行動（購買）履歴等に基づいており、音楽音響信号理解に基づいているものはほとんどない。

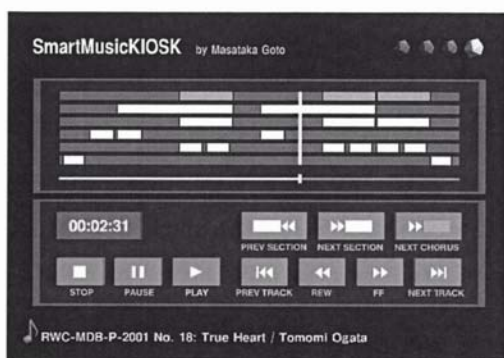
そこで、楽曲の音響信号から特徴量を抽出することで、アーティスト間や楽曲間の類似度が計算可能になると、アノテーションや履歴が乏しいような、新人あるいは無名なアーティストやその楽曲、最近発売された楽曲を扱うことができるようになる。こうした楽曲の音響的特徴に基づく処理は人気や流行とは無関係であり、音楽コレクションに対してユーザが能動的にインタラクションする際に、音響的特徴に基づかない既存のサービスの手法と相補的に活用できる。

3 音楽音響信号理解に基づく能動的音楽鑑賞インタフェース

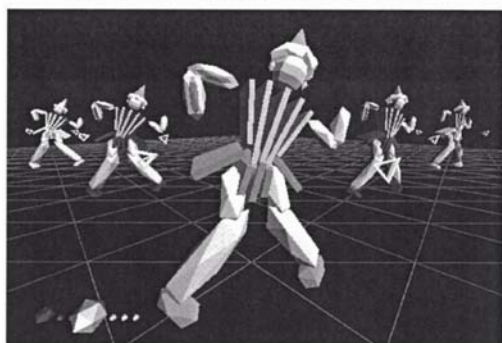
2章で述べたように、能動的音楽鑑賞のための様々なインタフェースを実現する上で、音響信号処理に基づく音楽理解技術が重要となる。様々な研究者の取り組みにより、既に、複数の楽器音が混在した実世界の音楽音響信号から、メロディー、ベース、ドラムス、ビート、サビ等の音楽情景記述 (music scene description)^{1)~4)} が推定できるようになってきた。そして、これらの技術が実際にエンドユーザの役に立つことを示すために、我々は図1に示す7種類の能動的音楽鑑賞インタフェースを実現してきた。以下では、これらのインタフェースによって、エンドユーザの音楽の聴き方がどう豊かになるか、どのような能動的なインタラクションが可能になるかを紹介する。

3.1 SmartMusicKIOSK: サビ出し機能付き音楽試聴機

SmartMusicKIOSK^{5),7)} は、楽曲構造を利用することで、ポピュラー音楽の楽曲内部のブラウジングや



(a) SmartMusicKIOSK: 視覚化された楽曲の繰り返し構造（上側の「音楽地図」ウィンドウ）を見ながら、楽曲の様々な箇所を自在に行き来して能動的に聴くことができる。



(b) Cindy: 音楽に同期して踊る仮想のダンサーの動作系列を、能動的に切り替えて音楽再生を楽しむことができる。

図 2: 音楽再生のための能動的音楽鑑賞インタフェース

楽曲の試聴を容易にする音楽再生インタフェースである。図 2(a) の下側のボタン群の中の「サビ出しボタン (NEXT CHORUS)」を押すことで、次のサビの先頭へジャンプ（瞬時に早送り）することができ、自分でサビを探す煩わしい作業から解放される。また、図 2(a) の上側には楽曲中の繰り返し構造を視覚化した「音楽地図」が表示されており、ユーザはこれを見ながら興味のない区間を容易に飛ばすことができる。「音楽地図」の横軸は時間軸で楽曲全体を表示しており、最上段がサビ区間の繰り返し構造を、その下の5段が他の様々な長さの区間の繰り返し構造を表している（各段で着色された区間同士が類似していて繰り返しを表す）。従来の音楽再生インタフェースでは、楽曲単位でしか興味のない音楽を飛ばせなかったのに対し、SmartMusicKIOSKによって初めて、楽曲内部の興味のない箇所も容易に飛ばすことが可能になった。

ここで必要なサビ区間や繰り返し区間は、我々のサビ区間検出手法 RefraiD⁷⁾ によって自動推定した。RefraiDは、ポピュラー音楽の音響信号を対象に、様々

な繰り返し区間の相互関係を調べることで、楽曲中で繰り返されるすべてのサビ区間を網羅的に検出しようとする手法である。

3.2 Cindy: ビートトラッキングによって踊る仮想のダンサー

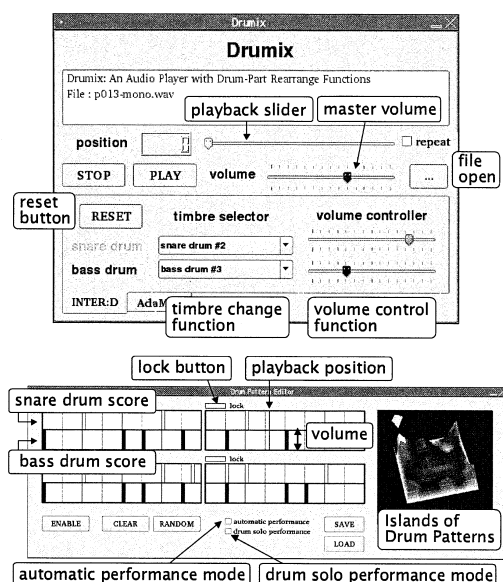
Cindy⁸⁾は、音楽のビートに同期してリアルタイムに踊る（動作や位置が変化する）仮想のダンサー（図 2(b)）を、コンピュータグラフィックスでアニメーション表示するシステムである。本システムでは、それぞれ異なった雰囲気を持つ踊りの動作系列（1小節単位で繰り返す踊り）が数種類用意されている。ユーザは、楽曲の再生に合わせて任意のタイミングでボタンを押して、それらの動作系列を次々と切り替えることができる。実際の動作系列の切り替わりのタイミングや系列内の各動作のタイミングは、ビートに自動的に同期するので、ユーザは気楽に切り替えるだけで音楽に合った踊りの振り付けができ、能動的なインタラクションを楽しむことができる。

ここで必要な階層的なビート構造（四分音符に相当するビートと小節の先頭の系列）は、我々のビートトラッキング手法⁸⁾によって自動推定した。本手法は、ドラム音の有無を問わず、テンポがほぼ一定な4/4拍子のポピュラー音楽を扱うことができる。こうした音楽音響信号中のビートに合わせてリアルタイムに踊る仮想ダンサーのアニメーションは、1994年に初めて実現された⁹⁾。

3.3 LyricSynchronizer: 歌詞と楽曲との自動的な時刻同期

LyricSynchronizer は、カラオケのように楽曲の再生と同期して色が変わる歌詞を表示するシステムである。歌詞のテキストを楽曲と自動的に同期させて表示することで、楽曲中のどこが再生されているかが歌詞上で容易にわかるようになる。さらに、画面に表示された歌詞上の任意の単語をクリックすることで、その単語の位置へジャンプして再生することも可能である。

この歌詞と歌声との時間的な対応付けを可能にするために、まず、我々のメロディー音高推定手法 PreFEst¹⁰⁾によって、様々な楽器音が含まれる混合音から、メロディーの歌声を含む最も優勢な音高を推定してその音を分離した。次に、間奏等を除いた実際に歌っている区間だけをHMMを用いて検出し、その区間の分離歌声に対して、Viterbiアラインメント（強制アラインメント）手法によって歌詞の各音韻



(a) Drumix: 音楽再生中にドラムスの音色やドラムパターンを能動的に変更できる。上側のウィンドウはINTER:Dと同等である。

図 3: 音楽加工（タッチアップ）のための能動的音楽鑑賞インタフェース

（母音）の位置を推定した¹¹⁾。これにより、歌詞中の各単語がどの時刻に歌われているかがわかる。

3.4 INTER: 音楽音響信号に対する楽器イコライザ

INTER¹²⁾は、楽曲中の複数の楽器パート（トラック）のそれぞれの音量を変化させることで、楽器単位のイコライジングをして音量バランスを変更（リミキシング）できる音楽再生インタフェースである。一般に、音楽CDのような複雑な音楽音響信号から任意の楽器パートを抜き出すのは非常に困難であり、文献12)で提案した完全なINTERを実現するには、文献13)のようなさらなる研究が必要である。そこで、その第一段階として、我々はドラムイコライズシステムINTER:D¹²⁾を実現した。ユーザはINTER:Dを用いることで、楽曲再生中にそのバスドラムとスネアドラムの音量や音色を、能動的に変更することができる。

ここで必要なバスドラムとスネアドラムの発音時刻と音色は、我々のドラム音認識手法¹⁴⁾によって自動推定した。本手法は、ドラム音の音色のテンプレートを利用して、それを対象楽曲へ適応させながら同時に楽曲中から発音時刻を検出することができ、楽曲中のドラム音の音色が事前にわからなくても適切に推定できる利点がある。

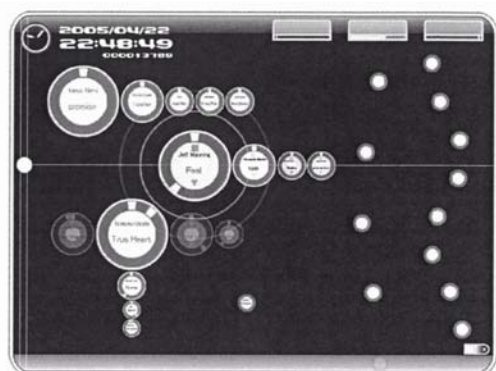
3.5 Drumix: ドラムパートのリアルタイム編集機能付き音楽再生インタフェース

Drumix¹⁵⁾は、あたかも他のドラマーが違うドラムパターンで演奏しているかのように、楽曲を再生中にリアルタイムにドラムパートを加工できる音楽再生インタフェースである。図3(a)の上側のウィンドウでは、INTER:Dと同様にバスドラムとスネアドラムの音量や音色を変更できる。さらに、下側のウィンドウを使うと、ドラムパターン（ここではバスドラムとスネアドラムの演奏パターン）を容易に編集したり引用したりすることが可能である。このようにユーザはDrumixを用いることで、楽曲の再生中に気ままにドラムスの音色やパターンを変える能動的なインタラクションを楽しむことができる。

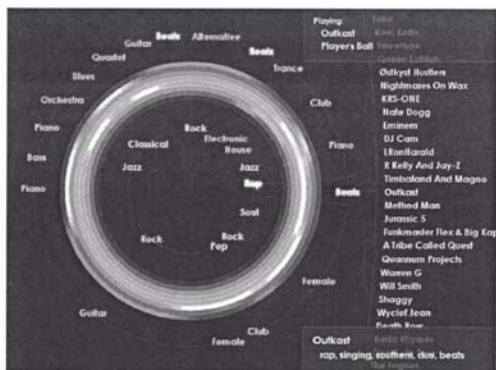
これは、INTER:Dを発展させたインタフェースに位置付けられ、同じドラム音認識手法¹⁴⁾を用いて実現した。また、楽曲中のドラムパターンを小節単位で編集できるようにするために、我々のビートトラックキング手法⁸⁾も用いた。

3.6 Musicream: 楽曲との能動的で多様な出会い方を可能にする総合的な音楽再生インタフェース

Musicream¹⁶⁾は、エンドユーザが抱く「何か聴きたい」「思い通りに聴きたい」という欲求を満たすことができる音楽検索・再生インタフェースである。これは、従来の音楽再生インタフェースで典型的であった「曲名やアーティスト名を列挙する」という固定概念から脱却し、能動的で自由度の高い音楽再生を実現することを目指したものである。まず、「何か聴きたい」という欲求を満たすために、様々な楽曲に容易に出会うことができる機能を提供する。図4(a)の画面上の右側を楽曲群（円盤）が上から下へ次々と流れ、ユーザが楽曲を何気なく手に取って聞くことで、未知の楽曲に能動的に出会うことができる。その楽曲が気に入ったら、流れている他の楽曲にそれを接触させることにより、雰囲気のような楽曲だけを選択的に次々とくっつけられる（制御可能な類似楽曲検索に相当する）。次に、「思い通りに聴きたい」という欲求を満たすために、複数のプレイリスト間の再生順を指定するような自由度の高いプレイリスト編集を可能にする。そして、Musicream上のすべての操作を記録することで、あたかもタイムマシンを使ったかのように、過去の画面を呼び出したり、そこから新たな未来を作ったりできる。



(a) Musiccream: 音楽コレクションをブラウジングして未知の楽曲に能動的に出会うことができる。



(b) MusicRainbow: 音楽コレクションをブラウジングして未知のアーティストに能動的に出会うことができる。

図 4: 音楽検索・ブラウジングのための能動的音楽鑑賞インタフェース

Musiccreamは、楽曲間の類似度を求める任意の手法と組み合わせることができるが、現在の実装では、MARSYAS¹⁷⁾で求めた30次元の楽曲特徴量同士の類似度を用いた。楽曲からの特徴抽出に基づいたインタフェースには他にも様々な研究事例があり、例えば、GenreSpaceやGenreGram¹⁸⁾、Islands of Music¹⁹⁾等が提案されている。

3.7 MusicRainbow: 音響信号の類似度とWebによる自動ラベリングを用いたアーティスト発見インタフェース

MusicRainbow²⁰⁾は、未知の音楽コレクションに含まれる多数のアーティストの中から、好みのアーティストを発見するためのインタフェースである。図4(b)の左側の虹色の円周上に、似たアーティスト同士が近くなるように全アーティストが配置され、アーティストを表す言葉(属性)がその周辺に表示される。ユーザは、ダイヤル型のデバイスを回して虹色

の円周を回転させることで、アーティストを表す言葉(属性)を見ながらアーティスト間を渡り歩くことができ、デバイスを押すことで、図4(b)の右側中央のアーティストの楽曲断片を聴くことができる。これにより、例えば好みのアーティストの周辺でそれに似た他の未知のアーティストを発見したりと、能動的にアーティストを探索することを可能にした。

MusicRainbowでは、楽曲の音響信号から特徴量を抽出し、楽曲特徴量間の類似度に基づいてアーティスト間の距離を求めることで、アーティストを自動配置した。そして、Web検索エンジンを用いてアーティストに関連したWebページを収集し、それらを解析することで各アーティストの属性を自動ラベリングした。

4 議論

本研究では、「音楽音響信号理解研究によって、エンドユーザの音楽の聴き方をどのように豊かにできるか」という研究課題に対する我々の一つの回答として、3章の7種類の能動的音楽鑑賞インタフェースを紹介してきた。ここで豊かにする対象である「音楽の聴き方」という言葉が意味する範囲は広く、例えば以下のような様々な聴き方が考えられる。

- 多聴/精聴

少しでも多くの楽曲を聴きたい人もいれば、逆に少数の楽曲を繰り返しじっくり聴きたい人もいる。

- 歌詞

同じ歌を聴いても、その歌詞をきちんと聴く人もいれば、歌詞をほとんど聴かない人もいる。

- 楽器パート

楽曲中の個々の楽器パートを個別に追いかけて聴く人もいれば、楽器パートをほとんど意識せずに全体の雰囲気で聴く人もいる。

- 好み

ジャンルやアーティスト、楽曲、音色等に対する好みは人それぞれで、聴くときに重視すべきポイントもそれぞれで違う。

- 聴き方の変化

楽曲を聴くときの年齢、時期、状況等によって、同じ曲でも聴き方やその印象が大きく異なることがある。また、しばしば聴き方の変化は不可逆で、例えば楽曲中のある要素がわかるようになった後では、それがわからなかった頃の過去の聴き方を思い出すことは難しくなる。

このように「音楽の聴き方」が多面的である以上、それを豊かにする方法も様々な側面から考えなければ

ならない。つまり、ある一つの究極的な能動的音楽鑑賞インタフェースがあると考えられるのではなく、多様な能動的音楽鑑賞インタフェースを検討していく必要がある。

ここで、「音楽の聴き方を豊かにする」ということは、その土台となる「音楽を聴く力」を暗黙の前提としている。しかし、音楽家の能力として、「音楽を創作する力」や「音楽を演奏する力」と併せて議論することはしばしばあるが、音楽家でないエンドユーザの能力としての「音楽を聴く力」を議論することは難しい。他人が楽曲中の何の要素をどう聴いているかを直接知ることではできないため、間接的に知るしかないが、必ずしも言語化ができないからである。そもそも、「自分が音楽を聴いているときに何がわかっていないのか」がわからないことが多い。そのため、ユーザが「音楽を聴く力」を向上させたい、音楽をより深く理解したいと思っても、それを可能にする方法は自明ではない。

そこで、今後の重要な方向性の一つとして、「音楽の理解力を拡張してくれるインタフェース（音楽理解力拡張インタフェース）」を技術的に実現することで、音楽をより深く理解できるようにする研究アプローチを提案する。今回提案した能動的音楽鑑賞インタフェースの中で、以下は、音楽理解力拡張インタフェースとしても位置付けられる。

- SmartMusicKIOSK

楽曲構造や音楽家の意図を理解する力を拡張してくれる。例えば、普段、楽曲の構造を意識せずに音楽を聴いている人が、それを意識して聴くきっかけとなりえる。また、ユーザがサビだけを続けて聴くことで、サビの繰り返しごとに歌詞やアレンジが変化していく様子（そこに込められた音楽家の意図）を、よりの確に把握できる可能性もある。

- LyricSynchronizer

歌詞を理解する力を拡張してくれる。例えば、普段は歌詞を意識して聴かない人も、再生と同期して色に変化する歌詞を見たり、興味のある歌詞の一部をクリックしてそこから聴いたりすることで、その歌詞に注目して聴き、歌詞が伝えるメッセージをよりの確に把握できる可能性がある。

- INTER, Drumix

個々の楽器パートを聞き分けて理解する力を拡張してくれる。例えば、普段はドラム音の音色やドラムパターンを意識して聴かない人も、Drumixを使ってそれらを加工することで、バスドラムとスネアドラムを聞き分け、ドラムパターンが楽曲の

印象にいかにか大きな影響を与えているかを把握できるようになる可能性がある。

我々は、上記が提供している

- 楽曲が内在している情報（音楽的要素）の視覚化
- 音楽再生と同期した情報提示
- 音楽再生や音楽的要素を制御できるインタラクティブなインタフェース

の一連の機能が、音楽理解力拡張インタフェースを実現する上で大切だと考えている。そして、上記で歌詞を聴くきっかけや、ドラムパートを聴くきっかけを与えたように、ある音楽的要素に気づかせるきっかけを与えて、ユーザが見過ごしがちな点をサポートするアプローチは効果的である。また、能動的でインタラクティブであることで没入感が得られ、より深くその楽曲と向かい合える状況を作り出すことも効果的だと考えられる。

これは「音楽を聴く力の支援」として位置付けられるが、その先にあるのは、「音楽を聴く力自体の向上・トレーニング」を可能にする研究である。外国語を聴く力を向上したければ語学学校や語学教材という手段があるが、音楽を聴く力を向上したくても、それに類する手段は整備されていない。音楽に関する学校や教材の多くは音楽家（音楽の作り手）がターゲットであり、ここでいうエンドユーザ（音楽の聴き手）が利用できるものは少ない。また、聴音トレーニングや絶対音感の習得も、「音楽を聴く力」を向上させて音楽をより深く理解するための手段とは違う。このような「音楽理解力向上インタフェース」を実現することも、今後の重要な方向性の一つだと言える。

5 おわりに

本稿では、音楽家でないエンドユーザが、通常の受動的な鑑賞とは異なり、より能動的にインタラクションしながら音楽を楽しむことができる能動的音楽鑑賞インタフェースについて述べた。そして、既に実現した7種類のインタフェースにより、音楽音響信号理解研究が音楽の聴き方をどのように豊かにできるかを例示した。その一部は音楽理解力拡張インタフェースとしても位置付けられ、音楽をより楽しめるようにするだけでなく、音楽をより深く理解できるきっかけを与えてくれる可能性がある。しかし、人間の音楽理解能力と比べれば、現在の音楽音響信号理解の技術にはまだまだ制約も多く、性能も不十分である。今後、より多様で高機能なインタフェースを実現するには、自動理解結果に誤りが含まれていても役に立つような応用上の工夫や、音楽音響信号理解技術の

さらなる発展が必要である。

なお、能動的であることの重要性は既に他の分野では知られており、話者の言葉を聴き手が反復することでより良いコミュニケーションを可能にする能動的聴取 (active listening) や、下線を引いたりコメントを書き込むことでより良い文章理解を可能にする能動的読書 (active reading) が過去に提案されている。さらに、電子的に書き込めるようにすることで能動的読書を支援するインタフェース²¹⁾も開発されている。しかし、それは能動的読書という既存の活動を支援するアプローチであり、新しい技術がなければ実現できないような能動的音楽鑑賞の実現を目指す本研究のアプローチとは、異なるものであった。本研究によって、今までとは違う角度から、音楽をより深く豊かに理解でき、より音楽に没入して楽しむことが実現できればと考えている。

2章で能動的音楽鑑賞の潜在的な可能性を議論した具体例の中で、まだ実現されていないものがある。例えば、外国語の歌詞から自動的に翻訳された歌詞を音楽に同期して表示できる音楽再生インタフェースや、楽曲再生中にその歌声を他の歌手の歌声に変更できる音楽加工インタフェースの実現は、今後の課題である。今後、多くの研究者によって様々な能動的音楽鑑賞インタフェースが実現され、それらが普及することで、日常生活での音楽の聴き方がより多様で豊かに変化していくことを期待したい。

謝 辞

3章で紹介した能動的音楽鑑賞インタフェースを共同で実現した奥乃博氏、後藤孝行氏、Elias Pampalk氏、藤原弘将氏、吉井和佳氏 (五十音順) に感謝する。本研究は、科学技術振興機構 CrestMuseプロジェクトによる支援を受けた。

参 考 文 献

- [1] Klapuri, A. and Davy, M. (eds.): *Signal Processing Methods for Music Transcription*, Springer (2006).
- [2] Goto, M.: Music Scene Description Project: Toward Audio-based Real-time Music Understanding, *Proc. of ISMIR 2003*, 231-232 (2003).
- [3] 後藤真孝, 平田圭二: 解説 “音楽情報処理の最近の研究”, 日本音響学会誌, **60**, 11, 675-681 (2004).
- [4] Goto, M. and Hirata, K.: Invited review: Recent studies on music information processing, *Acoustical Science and Technology (edited by the Acoustical Society of Japan)*, **25**, 6, 419-425 (2004).
- [5] 後藤真孝: SmartMusicKIOSK: サビ出し機能付き音楽試聴機, 情処学論, **44**, 11, 2737-2747 (2003).
- [6] 後藤真孝: 招待講演「音楽情報処理が実世界と結びつく: あなたも使える音楽情報処理」, 日本音響学会音楽音響研究会資料 MA2005-39, 97-100 (2005).
- [7] Goto, M.: A Chorus-Section Detection Method for Musical Audio Signals and Its Application to a Music Listening Station, *IEEE Trans. on Audio, Speech, and Language Processing*, **14**, 5, 1783-1794 (2006).
- [8] Goto, M.: An Audio-based Real-time Beat Tracking System for Music With or Without Drumsounds, *Journal of New Music Research*, **30**, 2, 159-171 (2001).
- [9] Goto, M. and Muraoka, Y.: A Beat Tracking System for Acoustic Signals of Music, *Proc. of ACM Multimedia 94*, 365-372 (1994).
- [10] 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*, **43**, 4, 311-329 (2004).
- [11] Fujihara, H., Goto, M., Ogata, J., Komatani, K., Ogata, T. and Okuno, H. G.: Automatic Synchronization Between Lyrics and Music CD Recordings Based on Viterbi Alignment of Segregated Vocal Signals, *Proc. of ISM 2006*, 257-264 (2006).
- [12] Yoshii, K., Goto, M. and Okuno, H. G.: INTER:D: A Drum Sound Equalizer for Controlling Volume and Timbre of Drums, *Proc. of EWIMT 2005*, 205-212 (2005).
- [13] Itoyama, K., Goto, M., Komatani, K., Ogata, T. and Okuno, H. G.: Integration and Adaptation of Harmonic and Inharmonic Models for Separating Polyphonic Musical Signals, *Proc. of ICASSP 2007* (2007).
- [14] Yoshii, K., Goto, M. and Okuno, H. G.: Drum Sound Recognition for Polyphonic Audio Signals by Adaptation and Matching of Spectrogram Templates with Harmonic Structure Suppression, *IEEE Trans. on Audio, Speech, and Language Processing*, **15**, 1, 333-345 (2007).
- [15] Yoshii, K., Goto, M., Komatani, K., Ogata, T. and Okuno, H. G.: Drumix: An Audio Player with Real-time Drum-part Rearrangement Functions for Active Music Listening, 情処学論, **48**, 3, 1229-1239 (2007).
- [16] Goto, M. and Goto, T.: Musicream: New Music Playback Interface for Streaming, Sticking, Sorting, and Recalling Musical Pieces, *Proc. of ISMIR 2005*, 404-411 (2005).
- [17] Tzanetakis, G. and Cook, P.: MARSYAS: A Framework for Audio Analysis, *Organized Sound*, **4**, 3, 169-175 (2000).
- [18] Tzanetakis, G., Essl, G. and Cook, P.: Automatic Musical Genre Classification of Audio Signals, *Proc. of ISMIR 2001*, 205-210 (2001).
- [19] Pampalk, E., Dixon, S. and Widmer, G.: Exploring Music Collections by Browsing Different Views, *Proc. of ISMIR 2003*, 201-208 (2003).
- [20] Pampalk, E. and Goto, M.: MusicRainbow: A New User Interface to Discover Artists Using Audio-based Similarity and Web-based Labeling, *Proc. of ISMIR 2006*, 367-370 (2006).
- [21] Schilit, B. N., Golovchinsky, G. and Price, M. N.: Beyond paper: supporting active reading with free form digital ink annotations, *Proc. of CHI '98*, 249-256 (1998).