楽器音認識技術を用いた音楽の可視化

北原 鉄朗[*1 *2]　後藤 真孝[*1 *3]　奥乃 博[*1 *4]　片寄 晴弘[*1 *2]

Music Visualization Using Musical Instrument Recognition Technique

Tetsuro Kitahara,[*1*2]　Masataka Goto,[*1*3]　Hiroshi G. Okuno,[*1*4]　and Haruhiro Katayose[*1*2]

Abstract – **Music visualization is expected to play an important role in enhancing enjoyment of music. This paper surveys music visualization systems using musical audio signal processing and introduces our prototype system of a music player with a function of visualizing instrumentation.**

1.　はじめに

音楽は本来耳で楽しむメディアであるが，目でも楽しむことができれば，楽しみ方の幅は広がっていくであろう．たとえば，楽曲の中身をグラフィカルに表示しながら再生してくれるツールがあれば，自分が聴いている楽曲がどんな音の組み合わせで成り立っているのかを発見しながら聴くことができる．また，視覚の一覧性を利用して，楽曲の中身を一目で把握した上で，どこをどう聴きたいかを指示しながら能動的に音楽を鑑賞するといったこともできる [1]．検索の観点からは，ユーザが楽曲リストから望みの楽曲を探しだすという場面で，楽曲を簡潔に視覚表現として表すこと（音楽サムネイル）ができれば，楽曲リストの内容を一目で確認することができ，楽曲の選択が容易になる．

このように音楽（楽曲内部）の可視化は重要な課題であるため，これまで様々な研究者が様々な方法で取り組んできた．本稿では，これまでの音楽可視化研究をサーベイし，我々が開発している楽器音認識技術を用いた音楽可視化システムを紹介する．

2.　音楽の可視化

音楽の可視化表現において，最も古く，また現在でもおそらく最も広く用いられているものは，楽譜（五線譜）である．楽譜は，エンターテインメントのための可視化というより，楽曲の記録や伝達のためのものである．楽譜を用いることで，楽曲を「読む」ことができるようになり，音を聴かずに，時間の流れを越えて楽曲の中身を視覚的に把握できるようになった．そのため，音楽家同士の情報伝達手段として，現在でも広く普及している．しかし，読むには訓練が必要である，エンターテインメント向けとしては表現が繁雑である，といった欠点がある．また，計算機技術で音響信号から楽譜を自動的に生成する問題は極めて難しく，現状では数種類の楽器による多重奏に対しても十分な精度とは言い難い [2]．

デスクトップミュージックの分野で普及しているもう１つの可視化表現はピアノロールである．ピアノロールは楽譜に比べてより直感的で，読むための訓練はあまり必要ない．しかし，それでもエンターテインメント向けとしては表現が繁雑で，内容を一目で把握するのは困難である．また，楽譜同様，音響信号から自動生成するのは困難である．

一方，音響信号から容易に生成が可能な可視化表現としては，スペクトログラムがある．スペクトログラムは各時刻・各周波数における成分の強さを色の濃淡で表したものである．これは信号の周波数特性を知るには便利で，実際，多くの音楽プレイヤーには，楽曲再生に連動してスペクトルを時々刻々と棒グラフで表示する機能が搭載されている．しかし，そこから音楽的に意味のある情報を見出すのは容易ではない．

このような状況の下，計算機で自動的に生成可能で，かつ音楽的意味の明確な可視化表現を求め，以下のような研究がなされてきた．

*1: 科学技術振興機構戦略的創造研究推進事業 CrestMuse プロジェクト
*2: 関西学院大学理工学研究科
*3: 産業技術総合研究所
*4: 京都大学情報学研究科
*1: The CrestMuse Project, CREST, JST
*2: Graduate School of Science and Technology, Kwansei Gakuin University
*3: National Institute of Advanced Industrial Science and Technology (AIST)
*4: Graduate School of Informatics, Kyoto University

Specmurt

Specmurt [3] は，ピアノロールとスペクトログラムの中間ともいえる可視化表現である．楽曲の音響信号に対して典型的な調波構造パターンを逆畳み込みすることで倍音成分を抑制して，スペクトログラムと同様に可視化することで得られる．倍音成分を抑制することでスペクトログラムからピアノロールに近づくため，ピアノロール表現に慣れたユーザが楽曲の中身を確認するには有用である．ただし，楽器認識を行っていないため，本来のピアノロールのように楽器ごとに別ウィンドウに表示したり色を変えたりといったことはできない．

SmartMusicKIOSK

サビ出し機能つき音楽試聴機 SmartMusicKIOSK では，楽曲の繰り返し構造を可視化した「音楽地図」というものが用意されている [4]．これは音響信号から自動的に生成することができ，イントロ，A メロ，B メロ，サビといった楽曲構造の位置関係を一目で確認することができる．この音楽地図を直接クリックすることで，望みの箇所に瞬時に移動することができる．

GenreGram モニター

GenreGram モニターは，楽曲のジャンル解析結果を可視化したものである [5]．解析対象ジャンルごとに円柱が用意され，ジャンル解析結果の信頼度に合わせてリアルタイムに円柱が上がったり下がったりする．各円柱には対応するジャンルを代表するような画像がテクスチャとして貼り付けられている．ジャンル解析結果として単一のジャンル名を出力するではなく，このように可視化することで，1 つのジャンルに決めがたいような楽曲や曲調が楽曲内で変化するような楽曲に対しても有効に働くようになっている．

TimbreGram モニター

TimbreGram モニターは，楽曲間の音色の類似度を色の近さで表現したものである [5]．各楽曲は横長の長方形で表され，横軸が時間を表す．長方形は，横軸の時刻に対応する音色特徴量が色のついた縦線で表されることで，縦縞模様のようになっている．色は音色の類似度を反映するように決定され，楽曲の類似度を長方形の色の類似度を見ることで知ることができる．この可視化表現は，楽器の音色に着目した点で後述の Instrogram と共通だが，TimbreGram は音色の類似度だけが表現されて楽器の種類を知ることはできない．また，ここでの音色とは，複数の楽器が同時に演奏した音の全体的な特徴を指し，各楽器の個別の音の特徴に着目したものではない．

MIDI データに対する可視化

その他，演奏データを 3 次元仮想空間にマッピングして可視化するシステム [6] なども提案されているが，MIDI データを対象としており，音響信号に適用するのは困難である．

3. 楽器構成の確率表現 Instrogram を用いた音楽可視化システム

本章では，我々が開発している楽器構成に着目した音楽可視化システムについて述べる．我々は楽器構成（その楽曲がどんな楽器で演奏されているか）を，楽曲を特徴づける重要な要素と位置付けている．同じ楽曲を異なる楽器で演奏すると，聴いたときの印象がまったく異なるものになる場合がある．ここから示唆されるように，楽器構成はその楽曲の雰囲気と密接な関係にあると考えている．

音響信号から楽器構成を推定する問題は，同時に 1 つの楽器しか演奏しないという条件での楽器識別なら様々な研究がある（e.g. [7]）が，多重奏に対する研究は多くない．多重奏に対する楽器識別の難しさの 1 つに，前処理としての発音時刻や基本周波数（F0）の推定のエラーの悪影響があげられる．いくつかの楽器識別手法 [8], [9] では発音時刻や F0 の推定結果を利用して特徴抽出を行うので，これらの推定エラーが致命的になる場合も少なくない．我々は，こういった推定処理を行った上で決定論的に楽器識別をするのではなく，楽器構成を Instrogram と呼ばれる確率表現の形で表し，スペクトログラムのような視覚表現として可視化する．

3.1 Instrogram とは

Instrogram[1][10] は，スペクトログラムに似た楽器存在確率の視覚表現である．解析対象となる楽器ごとに 1 つの画像が存在し，画像の色の強さによってその楽器が存在する確率を表す．各画像は横軸が時刻，縦軸が F0 を表し，対象楽器 $\Omega = \{\omega_1, \cdots, \omega_m\}$ に対して，i 番目の画像の各ピクセル (t, f) の色の強さが，時刻 t において f を F0 とする楽器 ω_i の音が存在する確率 $p(\omega_i; t, f)$ を表す．図 1 に例を示す．これは，ピアノ，バイオリン，フルートによる「蛍の光」の三重奏を，ピアノ，バイオリン，クラリネット，フルートを対象に Instrogram を作成し

1: http://winnie.kuis.kyoto-u.ac.jp/~kitahara/instrogram/

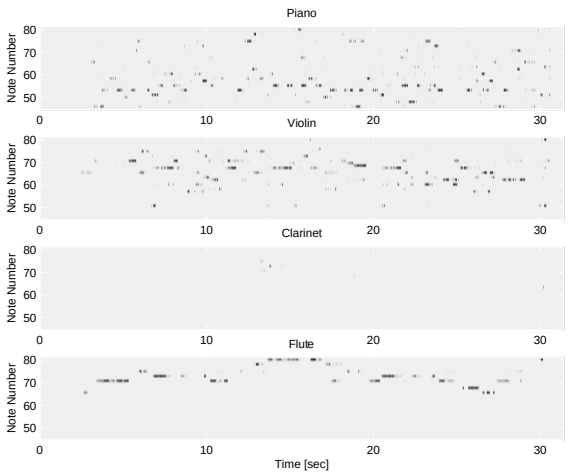

図 1　Instrogram の例（ピアノ，バイオリン，フルートによる「蛍の光」の三重奏）

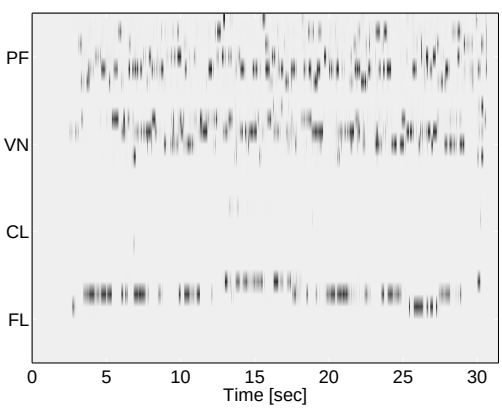

図 2　図 1 の簡略版（低周波数分解能版）

たものである．ここで，時間分解能は 10ms，周波数分解能は 100cent とした．

高い周波数分解能が要らない場合は，周波数軸をいくつかの区間に分割して区間内の値をマージすることで周波数分解能を粗くすることもできる．全周波数区間を N 個の区間 $I_1, \cdots, I_N$ に分割したとき，k 番目の周波数区間 I_k の楽器存在確率 $p(\omega_i; t, I_k)$ は $p(\omega_i; t, \bigcup_{f \in I_k} f)$ と定義する．簡略化された Instrogram を図 2 に示す．図 1 あるいは図 2 より，この楽曲は高音部はフルート，中音部はバイオリン，低音部はピアノによる演奏であることがわかる．

Instrogram は上述の TimbreGram と楽器の音色に着目した点で共通だが，各対象楽器の音が存在するかどうかを確率として表現している点で異なる．

3.2　Instrogram の作成方法

Instrogram 作成における中心的な課題は，楽器存在確率 $p(\omega_i; t, f)$ の計算である．いま，同時刻において F0 が同じ音が 2 つ以上鳴ることはない，すなわち，$\forall \omega_i, \omega_j \in \Omega\colon i \neq j \Longrightarrow p(\omega_i \cap \omega_j; t, f) = 0$ と仮定する．何らかの楽器の音が存在するという全対象楽器の和事象を $\mathrm{X} (= \omega_1 \cup \cdots \cup \omega_m)$ と書くこととすると $\omega_i \cap \mathrm{X} = \omega_i$ であるので，$p(\omega_i; t, f)$ は次の 2 つの確率の積で表すことができる：

$$p(\omega_i; t, f) = p(\mathrm{X}; t, f)\ p(\omega_i | \mathrm{X}; t, f).$$

ここで，$p(\mathrm{X}; t, f)$ は**不特定楽器存在確率**といい，時刻 t において f を F0 とする何らかの楽器の音が存在する確率を表し，$p(\omega_i | \mathrm{X}; t, f)$ は**条件付き楽器存在確率**といい，時刻 t において f を F0 とする何らかの楽器の音が存在するとすると，その楽器が ω_i である確率を表す．前者は PreFEst [11] を，後者は隠れマルコフモデルを用いることで計算することができる．詳細は [10] を参照されたい．

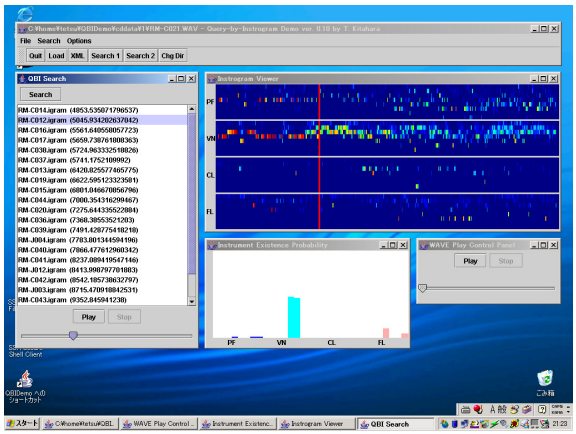

図 3　楽器構成可視化機能つき音楽プレイヤー

3.3　楽器構成可視化機能つき音楽プレイヤー

我々は，上記で述べた技術に基づき，楽器構成を可視化しながら楽曲を再生する音楽プレイヤー（図 3）を試作した．このプレイヤーは楽器構成を 2 種類の方法で可視化する．1 つは Instrogram をそのまま表示するもの（図 3 の上のウィンドウ）である．ウィンドウ内には解析対象楽器ごとに時間・周波数平面が用意され，楽器存在確率は色で表現される．再生位置は赤い縦線で表される．もう 1 つは楽器存在確率を棒グラフとして表示するもの（図 3 の下のウィンドウ）である．棒グラフの棒が長いほど楽器存在確率が高いことを表し，この棒グラフの長さが，楽曲の再生に合わせて時々刻々と変化していく．

このプレイヤーは，単に演奏を見ながら聴くだけでなく，たとえば「バイオリンが弾き始めるところから聴く」といった楽器構成の移り変わりに基

づく頭出しを，Instrogramを見ながら該当する時刻の点をクリックすることで容易に行える．その他，Instrogramの類似度（楽器構成の類似度）に基づいて楽曲検索を行う機能も実装している．

3.4 考察と今後の展望

現状のプロトタイプシステムでは，楽器構成はInstrogramと棒グラフの2種類のみによる可視化であるが，たとえば楽器のイラストの大きさで楽器存在確率を表したり，楽器存在確率に連動して変化するアニメーションなどで楽器構成を表現することも可能である．こうした多様な表現を導入することで，楽器構成の可視化は，主に次の3つの可能性があると考える．

見て楽しい

本システムを用いることで，どんな楽器が演奏しているのかを発見しながら音楽を聴くことができる．これに，イラストやアニメーションによる表現が加われば，楽器の発見をより直感的かつ楽しみながら行うことができる．これは，子どもに対する音楽教育においても有用と考えられる．

楽器に基づく頭出し

上でも述べたように，本システムを使うと「バイオリンが弾き始めるところから聴く」といった聴き方を容易に実現できる．この機能はSmartMusicKIOSK [4] で実現したサビ出し機能とも関連するが，SmartMusicKIOSKで実現されていた音楽地図は基本的に繰り返しの検出しか行っていないのに対し，本システムでは楽器構成を時々刻々と推定し，ユーザはそれに基づいて頭出しポイントを選ぶ，という点が異なる．この機能は，たとえば楽器練習者に対する支援として有用と思われる．

楽曲サムネイル

各楽曲に含まれる楽器のイラストをその楽曲のアイコンのように表示することで，楽器構成に着目した楽曲サムネイルとして活用できる．これを用いることで，効率的かつ楽しみながら自分が聴きたい楽曲を選択できるようになると期待される．

4. おわりに

本稿では，音楽をより楽しみながら聴くための一手段として音楽の可視化に着目し，特に音楽音響信号処理技術を用いて音楽を可視化する手法やシステムを紹介した．その後，我々が考案した楽器存在確率の視覚表現Instrogramを述べ，Instrogramを用いた楽器構成可視化機能つき音楽プレイヤーのプロトタイプを紹介した．

商用音楽にはいわゆるプロモーションビデオが製作され，カラオケに映像が付与されるなど，音楽を目でも楽しむという形態は確実に普及しつつあるように思われる．実際，PC上で動作するいくつかのメディアプレイヤーに「視覚エフェクト」の名で音楽の可視化機能が実装されている．しかし，視覚エフェクトが楽曲の内容を適切に表しているとは言い難く，十分な視覚効果が得られているとは言えない．一方，本稿で紹介した研究事例では，いずれも最新の音楽音響信号処理技術を用いて音響信号から音楽的に意味のある特徴を抽出して可視化に利用することで，より効果的な可視化を実現している．音楽の可視化は音楽音響信号処理技術の応用先としても有望であり，さらなる発展を期待したい．

参考文献

[1] Goto, M.: Active Music Listening Interfaces based on Signal Processing, *Proc. ICASSP*, Vol. IV, pp. 1441–1444 (2007).

[2] Klapuri, A. and Davy, M.(eds.): *Signal Processing Methods for Music Transcription*, Springer (2006).

[3] 亀岡弘和，斎藤翔一郎，西本卓也，嵯峨山茂樹：Specmurtにおける準最適共通調波構造パターンの反復推定による多声音楽信号の可視化とMIDI変換，情処研報，2003-MUS-56, pp. 41–48 (2004).

[4] 後藤真孝：SmartMusicKIOSK：サビ出し機能付き音楽試聴機，情処学論， Vol. 44, No. 11, pp. 2737–2747 (2003).

[5] Tzanetakis, G.: Manipulation, Analysis and Retrieval Systems for Audio Signals, PhD Thesis, Princeton University (2002).

[6] 宮崎麗子，藤代一成，平賀瑠美：comp-i: MIDIデータの視覚探索システム，情処学論，Vol. 45, No. 3, pp. 739–742 (2004).

[7] Martin, K. D.: Sound-Source Recognition: A Theory and Computational Model, PhD Thesis, MIT (1999).

[8] 柏野邦夫，村瀬　洋：適応型混合テンプレートを用いた音源同定，信学論，Vol. J81-D-II, No. 7, pp. 1510–1517 (1998).

[9] 北原鉄朗他：多重奏を対象とした音源同定：混合音テンプレートを用いた音の重なりに頑健な特徴量の重みづけおよび音楽的文脈の利用，信学論，Vol. J89-D, No. 12, pp. 2721–2733 (2006).

[10] Kitahara, T. et al.: Instrogram: Probabilistic Representation of Instrument Existence for Polyphonic Music, *IPSJ Journal*, Vol. 48, No. 1 (2007). (also published in IPSJ Digital Courier, Vol.3, p.1–13, 2007).

[11] Goto, M.: A Real-time Music-scene-description System: Predominant-F0 Estimation for Detecting Melody and Bass Lines in Real-world Audio Signals, *Speech Comm.*, Vol. 43, No. 4, pp. 311–329 (2004).