Instrogram: 多重奏中の楽器構成に関する確率的表現法

Instrogram: A Probabilistic Representation of Instrumentation in Polyphonic Music

北原鉄朗 [1,2] 後藤真孝 [1,3] 奥乃 博 [1,4] 片寄晴弘 [1,2]
Tetsuro Kitahara Masataka Goto Hiroshi G. Okuno Haruhiro Katayose

[1] 科学技術振興機構 CREST CrestMuse プロジェクト
CrestMuse Project, CREST, JST
[2] 関西学院大学理工学研究科
School of Science and Technology, Kwansei Gakuin University
[3] 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology
[4] 京都大学大学院情報学研究科
Graduate School of Informatics, Kyoto University

1 はじめに

近年，大容量携帯音楽プレイヤーやディジタル音楽配信などが普及し，音楽音響信号を対象とした情報検索技術へのニーズが高まりつつある．音響信号は情報検索を行うには抽象度が低いので，抽象度の高い何らかの表現形式に変換するのが一般的である．本稿では，これまで提案されてきた表現形式を概観し，我々が提案した楽器構成の確率的表現法 Instrogram について述べる．

2 音楽音響信号の表現

2.1 演奏の記号表現

音楽音響信号の表現として古くから用いられているものに楽譜がある．楽譜は典型的な音楽の記号表現の 1 つで，楽曲の記録や伝達手段として広く用いられている．そのため，音楽音響信号から楽譜表現を得る試みが，様々な研究者によってなされてきた [1]．しかし，多重奏の複雑な音響信号を精度よく楽譜表現に変換するには，信号処理技術だけでなく高度な知識処理を必要とし，いまだ難しい課題である．

2.2 演奏の非記号表現

楽譜表現は精度良く取得できれば非常に強力だが，楽譜のように完全に記号化された表現でなくても有用な表現を考えることができる．音響信号を記号化せずに表現したものにスペクトログラムがあるが，互いに重なりあう数多くの倍音成分がそのまま表示されるため繁雑で，使い勝手はよくない．そこで，以下のような方法で基本周波数成分のみを抽出・表示する方法が提案された．

- 調波構造を制約付き GMM でモデル化し，どの周波数にどの程度の強さの基本周波数成分があるかを時刻ごとに EM アルゴリズムで推定する方法 [2, 3].
- 調波構造のモデルを時間周波数平面に拡張したもの [4].
- 典型的な調波構造をスペクトルに逆畳み込みすることで倍音成分を抑制する方法 [5].
- Sparse coding や non-negative matrix factorization を用いて，スペクトログラムを，調波構造を表す基底行列と演奏情報を表す重み行列に分解する方法 e.g.[6].

これらは，自動採譜のための中間表現としてだけでなく，鼻歌検索や類似楽曲検索など様々な場面での利用が考えられる．しかし，演奏についての情報は得られるものの，「どんな楽器による演奏か」（楽器構成）については知ることはできない．

2.3 楽器構成の表現

楽器構成は，音楽情報検索において重要な要素であると考えている．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異なったものになることからも，楽器構成の重要性が示唆される．楽器構成の表現に関連する研究として，楽器音認識が行われてきた [7] が，多重奏に対する楽器音認識は多くなく，楽器構成をどういった形式で表現すべきかも十分に議論されていなかった．

楽器構成をグラフィカルに表現する方法として，Timbregram が提案されている [8]．これは，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．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る．これは，楽器構成の大まかな類似性を知るには便利だが，具体的な楽器構成や各楽器の大まかな演奏内容を知ることはできない．

3 Instrogram：楽器構成の確率的表現法

我々は，楽器構成や各楽器の大まかな演奏内容を確率的に表現するものとして Instrogram[9] を提案した．

3.1 Instrogram とは

Instrogram[1]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・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．

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

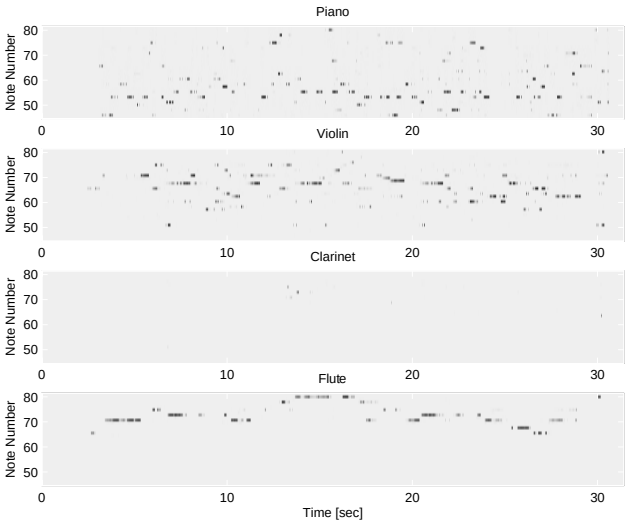

図 1　Instrogram の例（ピアノ，バイオリン，フルートによる「蛍の光」の三重奏）

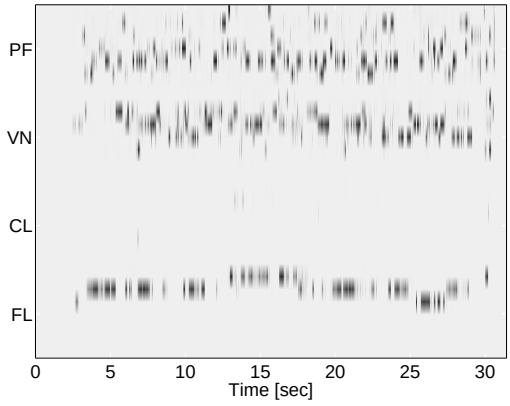

図 2　図 1 の簡略版（低周波数分解能版）

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，$\forall \omega_i, \omega_j \in \Omega$: $i \neq j \implies p(\omega_i \cap \omega_j; t, f) = 0$ 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，

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

と，2 種類の確率の積で表すことができる．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[2] 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 [9] を参照されたい．

3.3　Instrogram の評価と音楽情報検索への応用

Instrogram の生成を，楽器音データベースから計算機上で人工的に生成した三重奏の音響信号と，比較的小規模なクラシックおよびジャズの実演奏の両方に対して行い，良好な結果を得た．生成された Instrogram を元に楽器ラベルの自動タグ付けを行ったところ，前者に対しては平均 87.5%，後者に対しては平均 69.4%の適合率が得られた．また，Instrogram の類似度を用いて類似楽曲検索を試したところ，弦楽曲の類似楽曲ベスト 3 がすべて弦楽曲になるなど，楽器構成の類似度を適切に反映した結果が得られた．同じ実験を MFCC の類似度を用いて行ったところ，楽器構成の類似度を適切に反映しないことがあった．こちらも詳細は [9] を参照されたい．

3.4　議論

Instrogram は，多重奏の演奏内容を，楽器別に，非決定論的に表したものである．

楽器別の表現—— 2.2 節で述べた表現では，いずれも全楽器の音を同じ時間・周波数平面上で表していたが，楽器（音色）の情報は，聴覚的情景分析の観点からも応用の観点からも重要であり，楽器ごとに別々の時間・周波数平面で表現すべきと考えている．

非決定論的な表現—— 従来の楽器音認識 [7] では決定論的に楽器を同定することを目的としており，楽器構成の非決定論的な表現法については検討されてこなかった．人間は，たとえば「フルートのような楽器の音が高音域で鳴っている」というあいまいなまま音楽を理解することが少なくない．Instrogram はこのようなあいまいな音楽理解を模した表現とも言える．

4　おわりに

本稿では，音楽音響信号の様々な表現を概観した後，楽器構成の確率的表現法である Instrogram を紹介した．Instrogram は，楽器構成を記号化せずに表現するというこれまでなかった表現形式である．今後は，ハーモニーなどの他の音楽要素も取り入れ，よりリッチな音楽の非記号表現を目指していきたい．

参考文献

[1] Klapuri, A. and Davy, M.(eds.): *Signal Processing Methods for Music Transcription*, Springer (2006).
[2] 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).
[3] 亀岡弘和他：ハーモニック・クラスタリングによる多重奏信号音高抽出における音源数とオクターブ位置推定，情処研報，2003-MUS-51, pp. 29–34 (2003).
[4] 亀岡弘和他：調波時間構造化クラスタリング（HTC）による音楽音響特徴量の同時推定，情処研報，2005-MUS-61, pp. 71–78 (2005).
[5] 亀岡弘和他：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).
[6] Abdallah, S. A. and Plumbley, M. D.: Polyphonic Music Transcription by Non-negative Sparse Coding of Power Spectra, *Proc. ISMIR*, pp. 318–325 (2004).
[7] Herrera-Boyer, P., Klapuri, A. and Davy, M.: Automatic Classification of Pitched Instrument Sounds, *Signal Processing Methods for Music Transcription* (Klapuri, A. and Davy, M., eds.), Springer (2006).
[8] Tzanetakis, G.: Manipulation, Analysis and Retrieval Systems for Audio Signals, PhD Thesis, Princeton University (2002).
[9] 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).