

音楽のパーツ表現

吉井 和佳^{1,a)}

概要：本講演では、音楽音響信号や楽譜を構成要素(パーツ)の集合として表現する技術について紹介する。音楽音響信号に対する自動採譜や音源分離は、音楽情報処理分野における中心的な課題のひとつであり、二十年以上に渡って盛んに研究されている。これらの課題に取り組むうえでは、混合音は一見複雑に見えても、高々有限個の楽器音が重なりあってできているという性質が重要である。したがって、混合音の冗長性を除去するための技術が必要であり、最近是非負値行列因子分解 (nonnegative matrix factorization, NMF) を用いて、混合音のスペクトログラムを基底行列とアクティベーション行列との積で低ランク近似することが一般的である [1]。このとき、基底行列は基底スペクトルを並べたものであり、各基底スペクトルはある楽器音のある音高の平均的なスペクトルを表していることが期待される。しかし、この方法では、音楽音響信号を楽器パート(音色)ごとに分離することが困難で、位相情報を取り扱うことができないという問題があった。本講演では、これらの問題を解決するための、音楽音響信号を複数の音高と音色に分解することができる複合自己回帰モデル (composite autoregressive model, CAR) [2,3] や重畳離散全極モデル (superimposed discrete all-pole model, SDAP) [4] と、位相情報(周波数ビン間の相関)をモデル化することで高品質な音源分離を行うことができる半正定値テンソル分解 (positive semidefinite tensor factorization, PSDTF) [5,6] について紹介する。さらに、自動採譜における最近の試みとして、楽譜に基づいて音響信号の生成過程を表現する NMF のような音響モデルだけではなく、楽譜自体の生成過程を表現する言語モデルを考え、統一的な階層ベイズモデルとして統合する手法 [7,8] を紹介する。これにより、自動採譜を行うと同時に、音符の階層構造や和音構造のような一種の音楽文法(メタなパーツ)を音響信号のみから自動推論することができる。これらの手法はすべて確率的に定式化されているため、ノンパラメトリックベイズモデルへの拡張が可能で、音高数・音色数・音源数・文法数・和音数などを自動的に推定することができる。従来は、音響信号を楽譜に変換することが自動採譜の目的であったが、さらに一歩進んで、楽譜(音符配置)の背後に存在するはずの少数の音楽規則を自律的に見出すことができるようになりつつあり、計算機による音楽理解に向けて着実に進んでいる。

参考文献

- [1] P. Smaragdis and J. C. Brown. Non-negative matrix factorization for polyphonic music transcription. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 177–180, 2003.
- [2] H. Kameoka and K. Kashino. Composite autoregressive system for sparse source-filter representation of speech. In *IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 2477–2480, 2009.
- [3] K. Yoshii and M. Goto. Infinite composite autoregressive models for music signal analysis. In *International Society for Music Information Retrieval Conference (ISMIR)*, pages 79–84, 2012.
- [4] K. Yoshii, K. Itoyama, and M. Goto. Infinite superimposed discrete all-pole modeling for source-filter decomposition of wavelet spectrograms. In *International Society for Music Information Retrieval Conference (ISMIR)*, pages 86–92, 2015.
- [5] K. Yoshii, R. Tomioka, D. Mochihashi, and M. Goto. Infinite positive semidefinite tensor factorization for source separation of mixture signals. In *International Conference on Machine Learning (ICML)*, pages 576–584, 2013.
- [6] K. Yoshii, R. Tomioka, D. Mochihashi, and M. Goto. Beyond NMF: Time-domain audio source separation without phase reconstruction. In *International Society for Music Information Retrieval Conference (ISMIR)*, pages 369–374, 2013.
- [7] H. Kameoka, K. Ochiai, M. Nakano, M. Tsuchiya, and S. Sagayama. Context-free 2D tree structure model of musical notes for Bayesian modeling of polyphonic spectrograms. In *International Society for Music Information Retrieval Conference (ISMIR)*, pages 307–312, 2012.
- [8] Y. Ojima, E. Nakamura, K. Itoyama, and K. Yoshii. A hierarchical Bayesian model of chords, pitches, and spectrograms for multipitch analysis. In *International Society for Music Information Retrieval Conference (ISMIR)*, pages 309–315, 2016.

¹ 京都大学 大学院情報学専攻 知能情報学専攻
Yoshida-honmachi, Sakyo, Kyoto, Kyoto 606-8501, Japan
^{a)} yoshii(at)kuis.kyoto-u.ac.jp