

科学研究費助成事業（学術研究助成基金助成金）研究成果報告書

平成25年5月30日現在

機関番号：82626
 研究種目：若手研究(B)
 研究期間：2011～2012
 課題番号：23700184
 研究課題名（和文） ノンパラメトリックベイズ理論に基づく音楽音響信号の構造学習と音源分離
 研究課題名（英文） Structure Learning and Source Separation of Music Audio Signals based on Bayesian Nonparametrics
 研究代表者
 吉井 和佳 (YOSHII KAZUYOSHI)
 産業技術総合研究所 情報技術研究部門 主任研究員
 研究者番号：20510001

研究成果の概要（和文）：本研究の目的は、音楽音響信号データに対して、音源分離と構造学習を教師なしで同時に行うためのノンパラメトリックベイズモデルの提案である。まず、(1)「楽譜（記号データ）に対する構造学習」に関する研究を行った。和音とは本来、C major や D minor などのコードラベルが付与されているものに限定されているわけではなく、あらゆる音高の組み合わせが和音を形成しうると考えるのが自然である。このことから、語彙として可算無限種類の和音が与えられたときに、それらの生成確率と遷移確率とを統一的に記述することができる確率モデルを提案した。さらに、(2)「信号（音響データ）に対する音源分離」に関する研究を行った。具体的には、市販 CD のように様々な楽器の音色や音高から構成されるモノラルの混合音を、従来のように音高ごとではなく、楽器パート（音色）ごとに分離するため、ソース・フィルタ理論と非負値行列分解に基づく確率モデルを提案した。今後の課題は、各確率モデルを洗練させていくとともに、両者の様々な統合の方式についてさらに検討を進めていくことである。

研究成果の概要（英文）： This research aims to develop a nonparametric Bayesian model that can be used for jointly performing source separation and structure learning of music audio signals in an unsupervised manner. First, we have studied on structure learning of music scores (symbolic data). Since we need to deal with an infinite number of kinds of chords (pitch combinations) that do not always correspond to conventional chord labels such as C major and D minor, we have proposed a probabilistic model that can represent the generative probability of each chord and the transition probabilities between adjacent chords in a unified manner. Then, we have studied on source separation of mixture signals (audio data). To separate a given mixture signal consisting of various timbres and pitches into musical instrument parts (different timbres), we have proposed a probabilistic model based on the source-filter theory and nonnegative matrix factorization (NMF). Future work includes the refinement of each model and investigation of various ways to integrate both models.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
交付決定額	3,400,000	1,020,000	4,420,000

研究分野：総合領域

科研費の分科・細目：情報学・知能情報学

キーワード：音楽情報処理、機械学習

1. 研究開始当初の背景

「音源分離」とは歴史のある研究テーマで、いくつかの音源から発生した音響信号が混じり合った音響信号（混合音）が観測されたときに、音源ごとの音響信号を復元する問題である。これはそのままでは一意に解けない不良設定問題であるため、何らかの適切な仮定や制約を導入することで、それを満たすように音源分離する手法が提案されてきた。代表的なものに、個々の音源信号の独立性を仮定する独立成分分析（ICA: Independent Component Analysis）や、各音源の振幅スペクトルパターンと各時刻の励起パワーが非負値である性質を反映した非負値行列分解（NMF: Nonnegative Matrix Factorization）などが知られている。

音楽情報処理分野では、このような汎用的手法を音楽音響信号の音源分離に適用するだけではなく、音楽の持つ構造に着目することで音源分離精度を向上させようとする試みが行われてきた。人間の場合には、複数の楽器音が複雑に重なり合っていたり、高速なフレーズであったりして聴き取りにくい箇所については、和音の構成や前後の音楽的つながりを考慮することで採譜結果の信頼性を高めることができる。これを計算機上で実現するには、隠れマルコフモデルやNグラムなどの確率モデルを用いて音高の遷移パターンを別データからあらかじめ学習しておき、音源分離時に学習済みのモデルを適用することがしばしば行われる。十分な品質と分量の学習データを準備するには、大変な労力が必要であった。

2. 研究の目的

本研究の目的は、最新の統計的機械学習の枠組みであるノンパラメトリックベイズ理論に基づいて、音楽の「構造学習」と「音源分離」とを教師なしで行うための数理モデルを確立することである。すなわち、音楽音響信号を与えるだけで、混合音の生成過程や音高間の同時的・経時的な相互依存関係を解釈するのに最も適した音楽構造を自己組織的に獲得すると同時に、そのような音楽構造を制約として高精度な音源分離を行える計算機を実現するための枠組みを提案し、その具体的な実現方法について研究を行う。

3. 研究の方法

近年、データに内在する構造を教師なしで推定しようとする「構造学習」と呼ばれる新しい研究パラダイムが注目されている。代表的なものに、自然言語処理分野における文章データに対する確率的文脈自由文法（PCFG: Probabilistic Context Free Grammar）の学習が挙げられる。従来は、人手で記述した多

数の文法規則に対し、各規則の出現確率のみが推定可能であった。一方、近年では必要な文法規則自体を推定することが可能であり、異なる言語に対しても文章データだけから言語構造を自己組織的に獲得することができる。このような構造学習を支える強力な統計的機械学習の枠組みがノンパラメトリックベイズ理論であり、画像処理や自然言語処理などの諸分野において目覚ましい成果を上げている。

本研究では、音源分離と構造学習のための音楽特有の現象を記述可能な確率モデルを提案することに取り組む。これを達成するため、ノンパラメトリックベイズ理論に基づく音響データに対する音源分離モデル、楽譜データに対する構造学習モデルの定式化に分けて取り組むことにする。

4. 研究成果

(1) 楽譜データに対する構造学習

和音系列に対する統計的言語モデルとして、ノンパラメトリックベイズ理論に基づくNグラムモデルについての研究を行った。従来の経験的なスムージングに基づくNグラムモデルには、理論的な裏付けがなく改善が困難であること、Nの値を一意に指定しなければならないこと、考慮する和音の種類（語彙）は恣意的に決めざるを得ないことといった問題が存在した。

これらの問題を解決するため、我々は語彙フリー無限グラムモデルを提案した。このモデルは、あらゆる音の組合せを和音として許容するため語彙が不要で、和音系列中の各和音が異なるコンテキスト長（理論上は無限でもよい）を持つことを許容する。ある和音系列が与えられた上で、次の和音を予測するときには、Nの値を一意に決めることなくあらゆる可能性を考慮できる。また、これまで観測したことがない和音が出現したとしても、その和音のゼログラム確率（構成音の同時出現確率）をこれまで観測してきた和音の構成音に基づいて計算することで、適切にNグラム確率を推定することができる。実験の結果、従来のNグラムモデルよりも低いパープレキシティを達成することが分かった。

ビートルズの「Let It Be」における和音系列に対して、各和音のもつ履歴長Nの事後分布を推定した例を図1に示す。本研究で提案するNグラムモデルを学習する際には従来のコードラベルを利用する必要はないが、わかりやすさのためにコードラベルの系列に変換して表記していることに注意されたい。ポピュラー音楽において定番としてよく利用される和音進行の最後に位置する和音であれば、先行するよく多くの和音から導出されていると推定することができた。

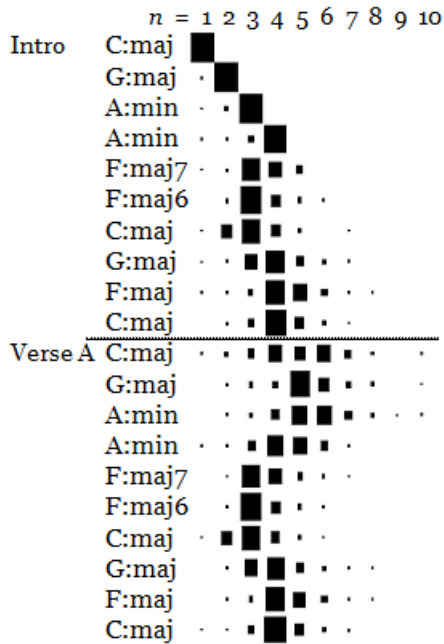


図1：各和音に対するNの事後分布推定結果

(2) 音響データに対する音源分離

音楽音響信号に含まれる複数の基本周波数(F0)を推定するための確率モデルについての研究を行った。従来、ソース・フィルタ理論に基づく非負値行列因子分解(NMF)では、非負値の観測スペクトログラムを行列とみなし、二種類のスペクトル(ソースのスペクトルおよびフィルタの周波数応答)と、それらの各組み合わせに対する時間方向のアクティベーション(音量変化)との積に分解することが行われていた。このとき、楽器の発音機構は自己回帰(AR)系でモデル化できると仮定すれば、F0をもつ周期信号(くし型スペクトル)や白色雑音(平坦スペクトル)などの音源により、楽器の反響特性(音色)を表現する全極型フィルタが駆動されることで音響信号が生成される過程を考えることができる。しかし、この種の複合自己回帰モデルにおいては、本来未知であるはずのソースやフィルタの個数を事前に指定しておく必要があるという問題があった。

この問題を解決するため、ガンマ過程を用いることで、上限のない個数のソースやフィルタを理論上許容可能なノンパラメトリックベイズモデルである無限複合自己回帰モデルを提案した。図2に確率モデルの概要を示す。観測データが与えられたときに、適切な個数のソースとフィルタを同時に推定することで、混合音中の複数の基本周波数を推定することが出来るだけでなく、音色の違いに基づいて楽器パートに分離することも

できる。モデルの学習のため、変分ベイズ法(VB)と乗法更新則(MU)を組み合わせた効率的な反復最適化アルゴリズムを導出し、スパースな学習を実現した。

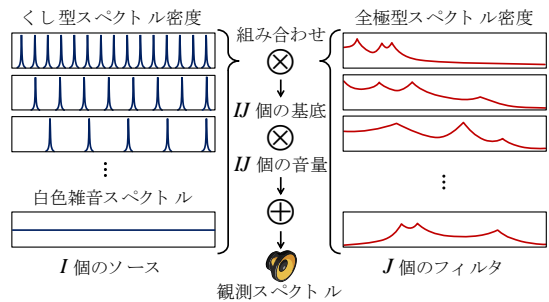


図2：無限複合自己回帰モデル

ピアノ音楽データベース MAPS を用いた混合音に対する音高推定実験で、フレームレベルで48.4%のF値を得た。この結果は関連研究で報告されている最新の成果には及ばないが、無限の複雑さを持つ空間においてスパースな学習を行うことができる性質から、言語モデルとの統合を行う上で、提案モデルは依然有望であると考えている。

(3) 音色と音高の同時推定

さらなる研究成果として、音響信号のスペクトル包絡と基本周波数とを同時に推定するための新しい線形予測分析法を開発することができた。本研究で開発した技術は、

(2)の無限複合自己回帰モデルを時間領域で定式化しなおし、音源分離の品質を改善するために利用可能である。さらに、非負値行列分解の本質的な拡張が可能になる。

従来、ソース・フィルタ理論に基づく線形予測分析法では、所与の観測信号(音声信号)はガウス性白色雑音を入力信号とする自己回帰系からの出力信号であると仮定して、全極型フィルタの係数を推定することが行われていた。しかし、声帯振動に起因する周期パルス(周波数領域では調波構造)が入力信号である場合には、推定されたスペクトル包絡(全極型フィルタの周波数応答)は、調波構造の倍音周波数で不必要に大きなピークをもつ問題があった。

この問題を解決するため、入力信号と出力信号との関係をノンパラメトリックベイズガウス過程回帰モデルで表現する無限カーネル線形予測分析法を提案した。本手法では、異なる基本周波数に対応する可算無限個のカーネルを考え、それらの凸結合で入力信号の周期性を表現する。ここで、無限個の非負の重みに対してガンマ過程事前分布を導入すると、スパースなマルチカーネル学習を行うことができる。すなわち、全極型フィルタ

の係数を推定すると同時に、基本周波数に対応する優勢なカーネルを同定できる。本手法を話声・歌声信号に対して適用し、基本周波数をもつ有声区間を同定しながら、基本周波数の影響を考慮したスペクトル包絡を推定できることを確かめた。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[学会発表] (計 10 件)

- ① Kazuyoshi Yoshii, Ryota Tomioka, Daichi Mochihashi, and Masataka Goto. Infinite Positive Semidefinite Tensor Factorization for Source Separation of Mixture Signals. The 30th International Conference on Machine Learning (ICML), 2013/06/18, Atlanta (USA).
- ② Kazuyoshi Yoshii and Masataka Goto. Infinite Kernel Linear Prediction for Joint Estimation of Spectral Envelope and Fundamental Frequency. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2013/05/29, Vancouver (Canada).
- ③ 吉井和佳, 後藤真孝. スペクトル包絡と基本周波数の同時推定のための無限カーネル線形予測分析法. 情報処理学会 第 99 回音楽情報科学研究会, 2013/05/11, お茶の水女子大学 (東京都).
- ④ 吉井和佳, 後藤真孝. 音楽音響信号解析のためのガンマ過程に基づく無限複合自己回帰モデル. 第 15 回情報論的学習理論ワークショップ, IBISML2012-51, Vol. 112, No. 279, pp. 121-128, 2012/11/7, 筑波大学東京キャンパス (東京都).
- ⑤ Kazuyoshi Yoshii and Masataka Goto. Infinite Composite Autoregressive Models for Music Signal Analysis. The 13th International Society for Music Information Retrieval Conference (ISMIR), pp. 79-84, 2012/10/09, Porto (Portugal).
- ⑥ 吉井和佳, 後藤真孝. 多重音基本周波数解析のための無限複合自己回帰モデル. 情報処理学会 第 96 回音楽情報科学研究会, 2012/08/10, 近江町交流プラザ (石川県).
- ⑦ Kazuyoshi Yoshii and Masataka Goto. Unsupervised Music Understanding Based on Nonparametric Bayesian Models. IEEE International Conference on Acoustics, Speech, and Signal

Processing (ICASSP), pp. 5353-5356, 2012/03/2, Kyoto International Conference Center (Kyoto).

- ⑧ Kazuyoshi Yoshii and Masataka Goto. A Unified Probabilistic Model of Note Combinations and Chord Progressions. The 4th International Workshop on Music and Machine Learning (MML), 2011/12/17, Granada (Spain).
- ⑨ Kazuyoshi Yoshii and Masataka Goto. A Vocabulary-Free Infinity-Gram Model for Nonparametric Bayesian Chord Progression Analysis. The 12th International Society for Music Information Retrieval Conference (ISMIR), pp. 645-650, 2011/10/27, Miami (USA).
- ⑩ 吉井和佳, 後藤真孝. 和音進行解析のための語彙フリー無限グラムモデル. 情報処理学会 第 91 回音楽情報科学研究会, 2011/07/27, 関西学院大学 (兵庫県).

[その他]

<http://staff.aist.go.jp/k.yoshii/>

6. 研究組織

(1) 研究代表者

吉井 和佳 (YOSHII KAZUYOSHI)
産業技術総合研究所
情報技術研究部門 主任研究員
研究者番号: 2050001