様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ２７ 年　 ６ 月 １５ 日現在

機関番号：１４３０１
研究種目：若手研究(B)
研究期間：2012～2014
課題番号：２４７００１６８

研究課題名（和文）統計的機械学習による音楽情景分析と音楽的要素のディレクションの研究

研究課題名（英文）A Study of Musical Scene Analysis and Direction of Musical Elements based on Statistical Machine Learning

研究代表者

糸山　克寿（Katsutoshi, Itoyama）

京都大学・情報学研究科・助教

研究者番号：６０６１４４５１

交付決定額（研究期間全体）：（直接経費）　　3,400,000円

研究成果の概要（和文）：本研究では，以下を達成した．(1)ノンパラメトリックベイズ法に基づく音楽音響信号の分析手法，(2) ベイズ推定に基づく和音の認識，(3)音楽音響信号からのバイオリン運指推定，(4)仮想楽器音源パラメータを推定，(5)ギター演奏者の習熟度に応じたタブ譜自動生成，(6)歌い方の特徴を抽出し歌手の歌い方のライブラリを作成，(7) 歌声と伴奏を分離し，歌声にビブラートやこぶしなどの歌唱表現を付与する音楽編集システムを開発，(8) 音響信号に対する残響抑圧，(9) 反復的な和音・音高推定方法の開発．

研究成果の概要（英文）：(1) Development of a musical audio signal analysis method based on nonparametric Bayesian manner, (2) development of an automatic chord recognition method based on Bayesian estimation, (3) development of a violin fingering estimation method from musical audio signals, (4) development of a method that estimates parameters of virtual instrument sound synthesizers (5) development of an automatic guitar tablature transcription method based on guitar player's proficiency, (6) constructing a singing style library from professional singers' singing voices, (7) development of a method that separates musical audio signals into singing voices and accompaniment signals, and edits singing styles of the separated singing voice, (8) development of an automatic dereverberation method, (9) development of a repetitive chord and pitch estimation method.

研究分野：音楽情報処理

キーワード：能動的音楽鑑賞　音源分離　歌声分離　自動採譜　和音認識　残響抑圧　ノンパラメトリックベイズ

１．研究開始当初の背景

(1) 能動的音楽鑑賞 近年，音楽鑑賞スタイルが受動的なものから能動的なものへと変化している．受動的な音楽鑑賞とは良い音を聴くことであり，高品質スピーカ，5.1ch などの高臨場感再生環境，アクティブノイズキャンセルなどで実現されてきた．従って，受動的な音楽鑑賞は技術革新で一般化が進められてきた．一方，能動的な音楽鑑賞は好みの音を聴くことであり，作曲・編曲・楽器演奏などの創作活動で実現されてきた．最近では，user-generated content (UGC) や consumer-generated media (CGM) 等に代表されるウェブサービス上で能動的な音楽鑑賞が楽しまれている．しかし，作曲などの創作活動には知識・経験・道具などが必要であるため，一部の人々のみに限定されていた．そこで，近年では能動的音楽鑑賞インタフェース[1]と呼ばれる，音楽情報処理技術を応用して音楽を能動的に楽しむための研究がとりくまれている．

(2) 音楽情景分析 機械学習に基づく手法で音楽音響信号を分析し，高度な検索や鑑賞に生かす音楽情景分析は近年広く取り組まれている．音楽情報検索に関する学会である ISMIR (International Society of Music Information Retrieval) が設立され，検索の基礎技術としての音楽分析は重要な分野となっている．また，音楽の分析能力を競う MIREX (Music Information Retrieval EXchange) では，毎年設定される様々な音楽分析タスクに対して世界中の研究者が最新の研究成果を持ち寄っている．国内でも，情報処理学会音楽情報科学研究会で「機械学習特別セッション」が組まれるなど，その重要性が広く認識されている．

２．研究の目的

本研究では，統計的機械学習による単音・楽器・和音・調の階層的音楽情景分析，および分析結果に基づく音楽的要素のディレクションの実現を目標に，個別の音楽的要素の統計的な表現とその推定，およびそれらの組み合わせによる複数の音楽的要素の推定，楽器音の合成までの総合的な技術の確立を目指す．

３．研究の方法

(1) 汎用性・頑健性の高い多重音の分析：楽器音モデルを用いて多重音を分析し，楽曲を構成する音楽的要素の抽出手法を研究する．

1 音楽的要素のモデル化：楽器音モデルを用いて多重音を分析し，楽曲を構成する音楽的要素（単独音・楽器の音色・和音と調）を抽出する．単独音のスペクトルのゆらぎ，楽器の音色のばらつきは，楽器音モデルのパラメータの分布をベイズ推定することで表現する．和音と調の曖昧性は，和音の構成音やその進行（遷移）から和音名や調の名前を間接的に推定することで，等価な和音をひとまとめにした上で，和音遷移の文脈を用いて再度具体化する．本手法による多重基本周波数の推定精度，楽器数の推定精度，和音進行の推定精度などの向上を検証する．

2 複数要素の認識を統合：上記の音楽的要素を同時に推定することで，独立な推定よりも高精度に推定する．ベイジアンノンパラメトリックを用いて，楽器の数や和音の種類が未知の条件下での推定にも取り組む．推定アルゴリズムの統合による計算量の増大を抑制するため，モデルの変分近似と周辺化や，単純なモデルで推定を収束させた後に複雑なモデルで推定する手法を開発する．モデル統合により，パラメータ推定には 10GB 以上の大きなメモリが必要となるため，大容量メモリを備えた計算機で実験を行う．

3 頑健性向上：残響を推定するため楽器音モデルを畳み込みモデルに拡張する．楽器音モデルと残響の推定には異なる型の目的関数が用いられるため，そのままでは同時推定は不可能である．補助関数法を用いて目的関数を変形することでこれらを同時推定する手法を開発する．また，楽曲の分析には，事前情報として楽譜を用いることができるようにする．楽譜に不備がある場合（音符の挿入置換誤り，一部の楽器パートのみ，音高・テンポのずれ）が想定されるため，音響信号と楽譜との同期処理を応用して楽譜の信頼できる部分を推定してその部分だけから楽曲構成要素を抽出する．

(2) 音楽的要素のディレクション：楽器演奏や作曲に精通しているユーザは楽器音の音色，演奏者の特徴といった，本システムが直接扱う特徴に着目して楽曲を鑑賞することに慣れているが，しろうとのユーザはこのような細かい部分に着目して楽曲を聴くことは少なく，曲名，ジャンル，演奏者といった，様々な要素を含む事例と関連づけて楽曲を鑑賞することが多い．そこで，しろうとのユーザでも簡単に楽曲をディレクションできるシステムには，楽曲を構成する要素を直接操作するだけではなく，事例を通じてこれらの要素を操作することが必要となる．特定の要素だけを操作する場合でも，他の要素との関係性を考慮して，音楽的に自然な楽曲となるようなタッチアップを実現する．例えば，和音進行を操作した場合には，旋律もそれに応じて操作する．

1 楽器音合成：楽器音の合成方法に関して，基礎的な技術を比較検討する．正弦波重畳モデルによる手法，パワースペクトルから位相を推定し時間領域信号を復元する手法，ソースフィルタモデルによる手法，既存の楽曲やその断片をつなぎ合わせる手法など，様々な手法を対象とする．また，以上画像検出手法やミッシングフィーチャー理論を応用し，楽器音の歪み（＝異常な箇所）を検出することで歪みを最小化する手法を開発し，2～10dB の歪み軽減を目標とする．

2 統計的学習に基づく音楽的要素の操作：統計的学習に基づいて混合音を操作・合成する手法を開発する．上記の音色などの他に，デ

ィレクションに適した要素を発見するため，ノンパラメトリックベイズに基づく無限関係モデルなどを用いて事例のクラスタリングで新たな共通要素を発見する．
3 事例に基づく音楽的要素の操作：事例となるコンテンツを基に混合音を操作・合成する手法を開発する．既存楽曲に対して，その楽曲を構成する要素の一部を事例から抽出した要素で置き換えることで，楽曲のタッチアップを実現する．複数の事例を用いて要素を置き換える場合には，各事例に対して要素ごとの重みパラメータなどを設定し，要素のパラメータを計算する．被験者実験を通じて，事例に基づく楽曲ディレクションの有効性と課題を検証する．

４．研究成果

(1)ノンパラメトリックベイズ法に基づく音楽音響信号の分析手法：多重奏の音楽音響信号に対して，従来法と同等以上の性能で数十倍高速な処理を行う多重基本周波数推定法を開発した．多重基本周波数を正しく推定するためには，調波構造の倍音強度比を正しく推定することが不可欠である．誤った倍音強度比が推定されることを防ぐため，楽器音シンセサイザで多種多様な楽器音の倍音強度比を分析し，その範囲内の倍音強度比のみが推定されるような手法を開発した．

(2)ベイズ推定に基づく和音の認識：音響特徴・ベース音高・和音遷移の手がかりを確率的に統合した和音認識手法を開発した．複数の手がかりの相互作用を考慮することで，73.7%の和音認識率を達成した．

(3)音楽音響信号からのバイオリン運指推定：手の形状やその変化速度などを考慮し，バイオリン演奏の音響信号から運指を自動推定する手法を開発した．バイオリンには弦が 4 本あり，音がどの弦から発せられているかを推定するために調波構造の倍音強度比を手がかりとして，倍音強度比と手の変形・移動コストを確率的にモデル化した．評価実験により，79.3%の推定精度を達成した．

(4) 音源分離などに起因する雑音や歪みを含む楽器音に対して，それらを含まないクリーンな楽器音を得るための仮想楽器音源のパラメータを推定する．多数の楽器音をランダムに生成し，楽器音からフレームベースの音響特徴量とその統計量を計算する．重回帰分析を用いて音源パラメータと音響特徴量との関係を学習し，未知楽器音のパラメータをその関係性を用いて推定する．

(5) ギター演奏者の演奏支援をするために実際のギター演奏音から演奏者の習熟度に応じたタブ譜を自動生成する手法について述べる．具体的には，初級者向けには音符の欠落などを許容してでも演奏が容易なタブ譜を，上級者向けには音高を正確に再現するタブ譜を，それぞれ生成する．推定される運指の難易度は，音響再現度と運指容易度の相対的な重みをユーザー側で調整することによって変更可能である．

(6) 市販楽曲らビブラート，こぶしやグリッサンドといった歌い方に関係する特徴を歌唱表現として抽出することで、歌手の歌い方のライブラリを作成する手法について述べる．これらの特徴は，歌唱 F0 軌跡中の特徴的な変動として現れる．本手法ではまず，時間周波数領域での最適経路探索問題を定式化することにより高周波数分解能，高精度な歌唱 F0 推定を行う．推定 F0 軌跡からパターンマッチングにより各歌唱表現を同定，パラメータ表現する．

(7) 歌声と伴奏を分離し，歌声にビブラートやこぶしなどの歌唱表現を付与する音楽編集システムを開発した．歌声分離は 2 つのステップからなる．第 1 ステップでは入力混合音を伴奏成分と歌声成分に粗く分離する．第 2 ステップでは分離された歌声からその音高を推定し，音高に基づく調波構造マスクにより歌声を際立たせる．本歌声分離手法は，音楽情報処理アルゴリズムの世界的なコンテストである MIREX2014 の歌声分離トラックにおいて最高性能を達成した．

(8) 音響信号からその残響を取り除く手法を構築した．残響は音高推定や和音認識などの各種音楽情報処理アルゴリズムの精度を大きく低下させるため，これらのアルゴリズムを頑健に動作させるための前処理として残響抑圧技術は有用である．

(9) 音楽音響信号に対する和音推定と音高推定の精度を高めるため，これらを反復的に実行する手法を実現した．入力楽曲に対して音高を粗く推定し，推定された音高から和音認識の特徴量を計算する．そこから和音を認識し，その結果を事前分布として用いることで再度音高推定を行う．

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）

〔雑誌論文〕（計 7 件） 全て査読有

[1] Akira Maezawa, Katsutoshi Itoyama, Kazunori Komatani, Tetsuya Ogata, Hiroshi G. Okuno, “Automated Violin Fingering Transcription through Analysis of an Audio”, Computer Music Journal, Vol.36, No.3, pp.57-72, 2012
DOI: 10.1162/COMJ_a_00129

[2] Daichi Sakaue, Katsutoshi Itoyama, Tetsuya Ogata, Hiroshi G. Okuno, “Robust Multipitch Analyzer against Initialization based on Latent Harmonic Allocation using Overtone Corpus”, Journal of Information Processing, Vol.21, No.2, pp. 246-255, 2013
DOI: 10.2197/ipsjjip.21.246

[3] 神田 直之, 糸山 克寿, 奥乃 博, “音声中の任意検索語検出のための未知語区間推定に基づく選択的インデクス統合法”, 情報処理学会論文誌, Vol.55, No.3, pp.1201-1211, 2014
http://id.nii.ac.jp/1001/00099473/

[4] **平山 直樹**, **吉野 幸一郎**, **糸山 克寿**, **奥乃 博**, "**擬似生成した複数方言言語モデル混合による混合方言音声認識**", **情報処理学会論文誌**, Vol.55, No.7, pp.1681-1694, 2014
http://id.nii.ac.jp/1001/00102165/
[5] Akira Maezawa, Katsutoshi Itoyama, Hiroshi G. Okuno, "Nonparametric Bayesian Dereverberation of Power Spectrograms based on Infinite-order Autoregressive Processes and Interpretation", IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol.22, Issue 12, pp.1918-1930, 2014
DOI: 10.1109/TASLP.2014.2355772
[6] Yoshiaki Bando, Takuma Otsuka, Kazuhiro Nakadai, Satoshi Tadokoro, Masashi Konyo, Katsutoshi Itoyama, Hiroshi G. Okuno, "Posture Estimation of Hose-shaped Robot by using Active Microphone Array", Advanced Robotics, Vol.29, Issue 1, pp.35-49, 2015
DOI: 10.1080/01691864.2014.981291
[7] Naoki Hirayama, Koichiro Yoshino, Katsutoshi Itoyama, Shunsuke Mori, Hiroshi G. Okuno, "Automatic Speech Recognition for Mixed Dialect Utterances by Mixing Dialect Language Models", IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol.23, Issue 2, pp.373-382, 2015
DOI: 10.1109/TASLP.2014.2387414

〔学会発表〕(計 22 件)
[1] Katsutoshi Itoyama, Tetsuya Ogata, Hiroshi G. Okuno, "Automatic Chord Sequence Recognition based on Probabilistic Integration of Acoustic Features and Chord Transition", The 25th International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems (IEA/AIE 2012), 2012/6/9-2012/6/12, **大連 (中国)**
[2] Daichi Sakaue, Katsutoshi Itoyama, Tetsuya Ogata, Hiroshi G. Okuno, "Bayesian Nonnegative Harmonic-temporal Factorization and Its Application to Multipitch Analysis", The 13th International Society for Music Information Retrieval Conference (ISMIR 2012), 2012/10/8-2012/10/12, **ポルト (ポルトガル)**
[3] Kazuki Yazawa, Daichi Sakaue, Kohei Nagira, Katsutoshi Itoyama, Hiroshi G. Okuno, "Audio-based Guitar Tablature Transcription using Multipitch Analysis and Playability Constraints", The 2013 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2013), 2013/5/26-2013/5/31, **バンクーバー (カナダ)**
[4] Daichi Sakaue, Takuma Otsuka, Katsutoshi Itoyama, Hiroshi G. Okuno, "Initialization-robust Bayesian Multipitch Analyzer based on Psychoacoustical and Musical Criteria", The 2013 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2013), 2013/5/26-2013/5/31, **バンクーバー (カナダ)**
[5] Naoyuki Kanda, Katsutoshi Itoyama, Hiroshi G. Okuno, "Multiple Index Combination for Japanese Spoken Term Detection with Optimum Index Selection based on OOV-region Classifier", The 2013 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2013), 2013/5/26-2013/5/31, **バンクーバー (カナダ)**
[6] Naoki Hirayama, Koichiro Yoshino, Katsutoshi Itoyama, Shunsuke Mori, Hiroshi G. Okuno, "Automatic Estimation of Dialect Mixing Ratio for Dialect Speech Recognition", The 14th Annual Conference of the International Speech Communication Association (INTERSPEECH 2013), 2013/8/25-2013/8/29, **リヨン (フランス)**
[7] Yoshiaki Bando, Takeshi Mizumoto, Katsutoshi Itoyama, Kazuhiro Nakadai, Hiroshi G. Okuno, "Posture Estimation of Horse-shaped Robot using Microphone Array Localization", The 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-2013), 2013/11/3-2013/11/7, **東京ビックサイト (東京)**
[8] Koutarou Furukawa, Keita Okutani, Takuma Otsuka, Katsutoshi Itoyama, Kazuhiro Nakadai, Hiroshi G. Okuno, "Noise Correlation Matrix Estimation for Improving Sound Source Localization by Mulrirotor UAV", The 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-2013), 2013/11/3-2013/11/7, **東京ビックサイト (東京)**
[9] Kazuki Yazawa, Katsutoshi Itoyama, Hiroshi G. Okuno, "Automatic Transcription of Guitar Tablature from Audio Signals in Accordance with Player's Proficiency", The 2014 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014), 2014/5/4-2014/5/9, **フィレンツェ (イタリア)**
[10] Yukara Ikemiya, Katsutoshi Itoyama, Hiroshi G. Okuno, "Transcribing Vocal Expression from Polyphonic Music", The 2014 IEEE International Conference on

Acoustics, Speech, and Signal Processing (ICASSP 2014), 2014/5/4-2014/5/9, **フィレンツェ（イタリア）**

[11] Akira Maruyama, Motoko S. Fujita, Katsutoshi Itoyama, Hiroshi G. Okuno, Mamoru Kanzaki, "Development of Automatic Bird-species Recognition System from Birdsongs in Tropical Forests", The 26th International Ornithological Congress (IOC2014), 2014/8/18-2014/8/24, **立教大学（東京）**

[12] Takahiro Iyama, Osamu Sugiyama, Takuma Otsuka, Katsutoshi Itoyama, Hiroshi G. Okuno, "Visualization of Auditory Awareness based on Sound Source Positions Estimated by Depth Sensor and Microphone Array", The 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-2014), 2014/9/14-2014/9/18, **シカゴ（イリノイ，米国）**

[13] Katsutoshi Itoyama, Hiroshi G. Okuno, "Parameter Estimation of Virtual Musical Instrumental Synthesizers", The 2014 Joint Conference on the 40th International Computer Music Conference (ICMC) and the 11th Sound and Music Computing Conference (SMC) (ICMC|SMC 2014), 2014/9/14-2014/9/20, **アテネ（ギリシャ）**

[14] Osamu Sugiyama, Katsutoshi Itoyama, Kazuhiro Nakadai, Hiroshi G. Okuno, "Sound Annotation Tool for Multidirectional Sounds based on Spatial Information Extracted by HARK Robot Audition Software", The 2014 IEEE International Conference on Systems ,Man, and Cybernetics (SMC 2014), 2014/10/5-2014/10/8, **サンディエゴ（カリフォルニア，米国）**

[15] Yoshiaki Bando, Katsutoshi Itoyama, Satoshi Tadokoro, Masashi Konyo, Kazuhiro Nakadai, Kazuyoshi Yoshii, Hiroshi G. Okuno, "A Sound-based Online Method for Estimating the Time-varying Posture of a Hose-shaped Robot", The 12th International Symposium on Safety, Security, and Rescue Robotics (SSRR-2014), 2014/10/1-2014/10/6, **洞爺湖文化センター（北海道）**

[16] Akira Maezawa, Katsutoshi Itoyama, Kazuyoshi Yoshii, Hiroshi G. Okuno, "Bayesian Audio Alignment based on a Unified Generative Model of Music Comosition and Performance", The 2014 International Society on Music Information Retrieval Conference (ISMIR 2014), 2014/10/27-2014/10/31, **台北（台湾）**

[17] Izaya Nishimuta, Naoki Hirayama, Kazuyoshi Yoshii, Katsutoshi Itoyama, Hiroshi G. Okuno, "A Robot Quizmaster that can Localize, Separate, and Recognize Simultaneous Utterances for a Fastest-voice-first Quiz Game", The 2014 IEEE-RAS International Conference on Humanoid Robots (Humanoids 2014), 2014/11/18-2014/11/20, **マドリード（スペイン）**

[18] Izaya Nishimuta, Kazuyoshi Yoshii, Katsutoshi Itoyama, Hiroshi G. Okuno, "Development of a Robot Quizmaster with Auditory Functions for Speech-based Multiparty Interaction", The 2014 IEEE/SICE International Symposium on System Integration (SII 2014), 2014/12/12-2014/12/13, **中央大学（東京）**

[19] Yoshiaki Bando, Takuma Otsuka, Ikkyu Aihara, Hiromitsu Awano, Katsutoshi Itoyama, Kazuyoshi Yoshii, Hiroshi G. Okuno, "Recognition of In-field Frog Chorusing using Bayesian Nonparametric Microphone Array Processing", AAAI-2015 Workshop on Computational Sustainability, 2015/1/25-2015/1/30, **オースティン（テキサス，米国）**

[20] Yukara Ikemiya, Katsutoshi Itoyama, Kazuyoshi Yoshii, "Singing Voice Analysis and Editing based on Mutually Dependent F0 Estimation and Source Separation", The 2015 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2015), 2015/4/19-2015/4/24, **ブリスベン（オーストラリア）**

[21] Satoshi Maruo, Kazuyoshi Yoshii, Katsutoshi Itoyama, Matthias Mauch, Masataka Goto, "A Feedback Framework for Improved Chord Recognition based on NMF-based Approximate Note Transcription", The 2015 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2015), 2015/4/19-2015/4/24, **ブリスベン（オーストラリア）**

[22] Yoshiaki Bando, Takuma Otsuka, Katsutoshi Itoyama, Kazuyoshi Yoshii, Yoko Sasaki, Satoshi Kagami, Hiroshi G. Okuno, "Challenges in Deploying a Microphone Array to Localize and Separate Sound Sources in Real Auditory Scenes", The 2015 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2015), 2015/4/19-2015/4/24, **ブリスベン（オーストラリア）**

〔その他〕

ホームページ等

http://winnie.kuis.kyoto-u.ac.jp/

６．研究組織

(1)研究代表者

糸山 克寿（ITOYAMA, Katsutoshi）
京都大学・大学院情報学研究科・助教
研究者番号：60614451