

統計的に学習可能な自動ジャズセッションシステムのための 数理モデル・演奏特徴量・事例活用の検討

保利 武志^{1,a)} 中村 和幸¹ 嵯峨山 茂樹¹

概要: 従来のジャズセッションシステムではヒューリスティックや感性情報を必要とする手法でセッションを実現しているのに対し、本研究では楽譜と演奏データから統計的に学習可能な枠組みによる自動ジャズセッションシステムの実現に向けて、その数理モデルや演奏の解析手法と、これらに基づいたプロトタイプとしてのジャズ演奏合成システムを検討中である。これまで我々は、音楽演奏を音楽事象空間におけるトラジェクトリとしてモデル化し、それを実際に学習可能とするための空間を量子化する手法としてクラスタリングを選択した。ただしクラスタリング手法にも複数の選択肢があるため、もっとも良いクラスタリング手法を選択するために、各小節または各拍単位で演奏特徴量を抽出して k -means, GMM, NMF の 3 手法でクラスタリングしたクラスラベルに対して、trigram 及び HMM による演奏予測精度によってそれぞれ比較評価したところ、NMF によるクラスタリングが効果的であることを示した。本システムではこれらの解析結果を用いて、Piano のみの演奏データに対して、事前にシステムに組み込んだ事例データの中からもっとも適切であると思われる Bass 及び Drums パートの演奏を選択して変換・編集・出力する。

キーワード: ジャズ, 事例編集合成, NMF, MIDI

Statistically Trainable Model of Jazz Session: Computational Model, Music Rendering Features and Case Data Utilization

TAKESHI HORI^{1,a)} KAZUYUKI NAKAMURA¹ SHIGEKI SAGAYAMA¹

Abstract: In contrast to most previous jazz session systems that require heuristics to realize a session, we are developing a jazz session system based on concatenation of case data from real jazz session recordings as a prototype toward a statistically trainable mathematical model of jazz session. So far, we modeled music performance in session as a set of individual trajectories in a music feature space, also compared clustering methods of k -means, GMM and NMF in terms of prediction accuracy in trigram and HMM, and found NMF most effective for the purpose of quantizing the continuous vector space. In the present system utilizing the above results, human piano performance is automatically accompanied by synthetic bass and drums to form a piano trio music by selection and concatenation of case data from recorded trio data.

Keywords: Jazz, Concatenated synthesizer, NMF

1. はじめに

人間と計算機とが協調演奏を行うセッションシステムの研究はこれまでも数多く行われており、自動伴奏シス

テム [1] をはじめとして、4 小節程度の掛け合いによるインタラクティブなシステム [2] やジャズを対象としたシステム [3], [4], [5], [6], [7], [8] などが挙げられる。

我々はこれまで演奏誤りや弾き直し、テンポ変化等にもフレキシブルに対応する楽譜を既知とした幅広い自由度を有した自動伴奏システム Eurydice[9] を開発してきた。本研究で我々はその発展として、インプロビゼーションにも

¹ 明治大学先端数理科学研究科
Graduate school of Advanced Mathematical Sciences, Meiji University

^{a)} cs51003@meiji.ac.jp

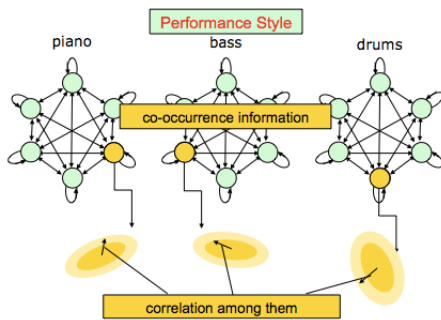


図 1 HMM によるピアノトリオの数理モデル

対応できる広義な自動伴奏システムという位置付けで、主としてジャズを対象とした、統計的に学習可能なジャズセッションシステムの開発に着手している。

従来のジャズを対象としたシステムでは、演奏の物理的特徴量から人間の演奏意図を推定し演奏を生成・出力するために、心理実験等の主観評価が必要であり、ヒューリスティックの削減、排除は課題の一つであった。そこで、我々はセッション中に起こる奏者間のインタラクションは実際の演奏データから抽出可能な演奏特徴量によっても追跡可能ではないかという観点から、演奏特徴量の変化や楽器間の対応関係を学習する枠組みを提案した [10]。特に統計的に学習可能な枠組みの中でセッションの数理モデルを定式化したことにより、適切なリード譜と演奏データが多いほど、システムはよりよいセッションを提供することが可能となると期待できる。

本研究では [10] のモデルに基づいたセッションシステムの開発を目的とし、その第一段階として、オフラインバッチ処理による、事例データを用いたジャズ編集合成システムを提案する。

2. モデリング

2.1 数理モデルとクラスタリング

我々はこれまでに音楽演奏は音楽事象空間 Ω 内のトラジェクトリ T として表現されると仮定し、これを学習可能な枠組みで扱うために、Hidden Markov Model(HMM) による確率的状態遷移モデルで定式化を行った (図 1)。HMM はマルコフ性を持って確率的に遷移する隠れ状態から、それぞれの確率分布に従って事象が生起されるとしたモデルであり、それぞれの状態は離散的に遷移する。トラジェクトリーモデルによる音楽演奏を HMM に適用すると、それぞれの隠れ状態から生起された演奏が時系列として観測されることとなる。各状態を別個の演奏スタイルと定義すると、これらが確率的に様々なスタイルへと変化することによって様々な演奏が生成されるとしたモデルであると言える。また同時に、音楽合奏では実に様々な楽器間でのインタラクションが行われるが、これらは各隠れ状態によって表される演奏スタイル間の共起関係によって起こりうる

と考えれば、楽器間の各状態にはそれぞれ共起確率を持つと捉え直すことができる。

統計的に学習可能とするための空間の量子化及びトラジェクトリの離散化手法に関しては、音楽的背景及びジャズ演奏家の意見を元に、リード譜と演奏データから抽出できる物理的特徴量から 68 次元のパラメータを選択し、これを「スタイルパラメータ」として定義することにより、スタイルパラメータを用いたクラスタリングによる量子化・離散化によって対応可能となった。

クラスタリングに関してはいくつかの手法が考えられるが、[10] では k -means、Gaussian Mixture Model (GMM)、Non-negative Matrix Factorization (NMF) の 3 手法に対して、それぞれ trigram 及び HMM の 2 パターンを用いて演奏予測を行い、その精度によって評価している。演奏予測精度は Yamaha Music Datashop のピアノトリオ 13 曲に対して交差検定法を用いて計算し、その結果、NMF がもっとも有効であることを示された。

2.2 スタイルパラメータ

リード譜と MIDI データから得られる、特にジャズセッションに関する特徴量として 68 次元のパラメータを定義した。抽出は小節単位及び拍単位に対してそれぞれ行われるが、本稿では小節単位のパラメータに限定した抽出を行っているため、以下に小節単位で抽出した 68 次元スタイルパラメータを挙げる

- ピアノに関する特徴量
 - ダイアトニックコード構成音の数、テンションノート数、アヴォイドノート数、ブルーノート数
 - 音域幅
- ベースに関する特徴量
 - 音域幅
- ドラムスに関する特徴量
 - ハイハットシンバル音数、スネアドラム音数、クラッシュシンバル音数、その他の音数
- 全楽器共通の特徴量
 - 音数、同時音数、平均ベロシティ
 - 上記特徴量について、前小節の数で除した値
 - 上記特徴量について、1 曲を通した合計数で除した値
 - 裏拍であれば 1、そうでない場合は 0

3. 共起関係のみを考慮したジャズピアノトリオ編集合成システム Ver.0

3.1 楽器間の共起関係

我々の提案するセッションの数理モデルでは、時間的な制約による確率的偏りに加え、楽器間のインタラクションによる共起関係を考慮している。実際人間同士のセッションでも楽器間でのインタラクションによって互いに音楽的ルールに基づいた制約や影響を受けつつバリエーションに

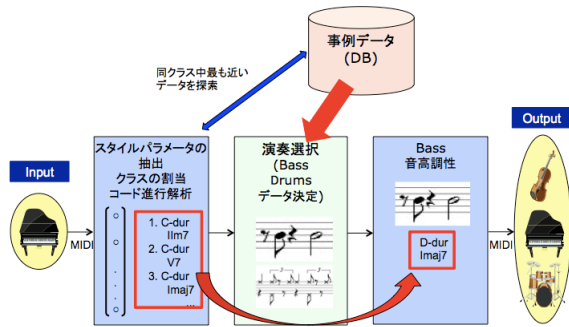


図 2 Jazz Piano trio 編集合成システムのアウトライン

富んだ演奏が生成される．本研究では特にこの共起関係に焦点を当て、音の生成とシステムへの実装を行った．

図 2 は我々が現在開発を進めているジャズピアノトリオ編集合成システムの全体的なアウトラインである．本稿で紹介する Ver.0 では、Piano だけの演奏データをインプットとして、オフライン上のバッチ処理でインプット演奏に適した Bass と Drums の演奏を事例データの中から探索し編集合成を行う．テンポや拍子といった情報はあらかじめ Piano による入力 MIDI 情報に含まれており、これらをそのまま利用する．ただし、今回の編集合成に際しては 1 小節ごとに Bass と Drums の演奏を合成するだけであっても、特徴量の相関関係が維持されることによって不自然ではない曲を生成できるという想定の下、共起関係のみを考慮したシステムへの実装を行った．なお今後のアップデート予定として、時間的制約の導入と演奏予測に基づいた編集合成、また事例データを利用するのではなく、作曲の観点から演奏を生成することも考えていく．

3.2 システムの概要

システムはオフライン上で Piano のみの演奏データをインプットとして受け取り、Bass と Drums の演奏が合成されたジャズピアノトリオの演奏データを出力する．編集合成用の Bass と Drums のデータは学習で使用した Yamaha Midi Datashop13 曲のデータをそのまま事例データとして利用した．

ある小節の Piano の演奏を見たとき、その演奏と共起関係の強い演奏を行っている Bass と Drums の演奏を探索して合成することができれば、少なくともその 1 小節を見たときのピアノトリオ演奏は自然な演奏として成り立っていることが期待される．また、これらを順次処理することによって、Piano を主としてそれに追従する Bass と Drums の演奏が合成された曲を生成することが可能である．ただし、時系列を考慮しないことである程度の大局的な流れを含有した演奏の生成は困難である点が本システム Ver.0 の根源的な限界であると言える．

以上より、本稿のジャズピアノトリオ編集合成システム Ver.0 とは、時系列的な制約がなくとも特徴量の相関関係

が維持されていれば、局所的に見たときにある程度は自然な演奏を生成することが可能であるとした、短フレーズの作曲等簡易的な作曲支援を想定した限定的なシステムである．

また入力として与えられた Piano のある 1 小節に対してスタイルパラメータを抽出した時、同様にして先に抽出しておいた全学習データの中からユークリッド距離においてもっとも近い特徴量ベクトルを持つ曲及び小節番号を探索し、その箇所の Bass と Drums の演奏を Input の Piano に合成することによって、特徴量の相関関係を一定の水準で保持することができる．同時に、学習時にクラスタリング (NMF) によって得られたパラメータを利用することでインプットのクラスを推定し、同じクラスという制限の下で探索することにより、計算量だけでなく、極端にはずれた演奏が選択される可能性を減じることが可能である．

本システムでは学習に用いたデータを事例データとして最大限活用するために、あらかじめ全事例データを小節単位に区切ってシステムに MIDI 情報を与えている．ただし事例データは元の曲における調性やコード等に従って音高やテンポが決定されたはずである．従って、特に Bass に関してはインプットに対してそのまま事例データを合成するわけにはいかず、インプットと事例間の調性を比較した上で同じキーとコードを持つ事例データの中から探索するか、その都度音高調整する必要がある．本稿では事例データ量を勘案して後者の手法をとった．

以下は実際の処理手順をまとめたものである．

- (1) 事前に学習データから小節単位で MIDI 情報を抽出し事例データベースとして管理．
- (2) 学習時に NMF によって基底行列を得る．
- (3) Piano のみの Input データを用意し、システムに入力として与える．
- (4) Input データのクラスタリング及び楽曲分析を行う．
- (5) Input データのある小節を見たとき、同じクラスを持つ事例データの中からユークリッド距離がもっとも小さくなる曲・小節番号を探索する．
- (6) Piano の調性やコードに合うように、事例データの Bass の音高を調整する．

3.3 Bass の音高調整

Piano によるインプットに対して単純に事例の Bass 演奏データをそのまま合成しても、ほとんどの場合において調性の不一致から不協和音が生じる．特にジャズにおける Bass のラインは主音のルート音やその構成音、ダイアトニックスケールを弾くことが多く、例えばインプットが Cdur の Cmaj7 であるのに対し、最近傍として得られた事例が Ddur の Dmaj7 であった場合、そのまま Ddur の Imaj7 を軸として生成されたであろう事例のフレーズを使用して合成しても調性の不一致が生じる場合が多い．



図 3 Autumn Leaves のコード進行一部

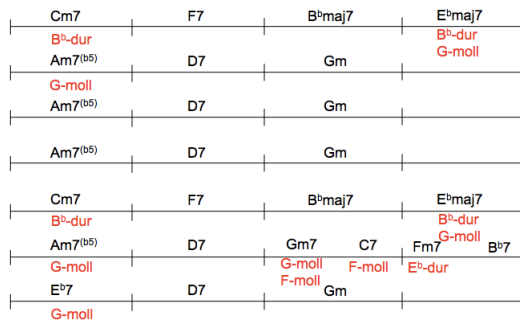


図 4 Autumn Leaves のコード進行と調推定

以上のようなケースでは一旦 Cdur に移調したデータを用いるほうが Piano との一致感は強く感じられるであろう。従って本システムでは「ダイアトニックスケール内の音を演奏すれば Piano の調性を共有可能である」という立場で音高の調整を行うことにした。

一般にジャズのスタンダード曲では、奏者がリード譜で与えられたコードに対して使用できるスケールを選択し、そのスケール上の音でメロディーラインを考える。そのため、ジャズにおいては曲全体を支配する調性と同時にノンダイアトニックコードによる部分的な転調などを明確にしておかなければならない。

ジャズにおいてもっともよく用いられるのは $IIIm7 - V7 - Imaj7(Im7)$ という、ツー・ファイヴ・ワンと呼ばれるコード進行である。V7 は裏コードである II^b7 を用いられることもあるが、いずれの場合においてもこれらを明確に把握することが非常に重要であると言われている。

図 3 は「Autumn Leaves」のコード進行の一部を抜粋したものである。主キーは Gmoll だが、最初の 3 小節のコード進行 $Cm7 - F7 - Bbmaj7$ はキー $Bbdur$ のツー・ファイヴ・ワンであり、一時的に主キーの平行調でドミナントモーションを行ったとみなすことができる。4 小節目では $Bbdur$ のダイアトニックコードである $IVmaj7$ であり、かつ $Gmoll$ のダイアトニックコードでもある VI^bmaj7 という Pivot Chord(共通のダイアトニックコード)を経て、5-7 小節目は $Gmoll$ の $IIIm7(b9) - V7 - Im7$ で解決している。他にも調性を判別するための手がかりとして、 $I - VIm7 - IIIm7 - V7$ のような一般に循環コードと呼ばれるコード進行パターンなどがあり、これらをシステムに組み込むことで調性の自動推定が可能となる。(図) は実際に「Autumn leaves」のコード進行を解析した実行例である。

インプット曲に対して得られた解析結果を用いて、選択された事例の Bass の音高を変化させる。本システムでは

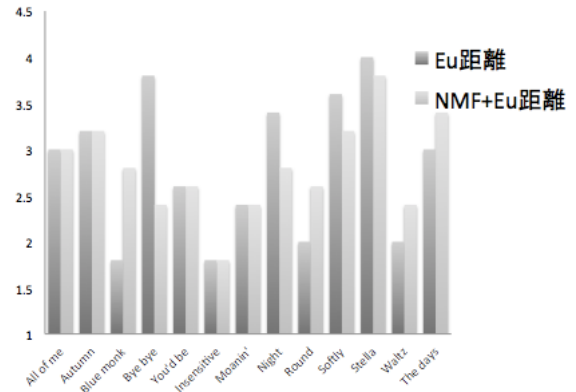


図 5 テンポ感の一致度

コードのルート音をインプット曲に合わせた後、その結果修正されたメロディーラインが推定した調のダイアトニックスケールに含まれていない場合に、さらに半音上下させる処理を行っている。

4. 評価実験

Yamaha Music Datashop のピアノトリオ MIDI データ 13 曲中、12 曲を学習及び事例生成のために利用し、残りの 1 曲の Piano パートのみを抽出してインプットデータとして用いて、Bass と Drums パートを編集合成し出力するという操作を 13 曲全てについて行った。評価は主として以下の 2 点を重視した。

(1) Piano 演奏に対する Bass と Drums のテンポや印象の一致感

(2) Bass の調性と全体的な一致感

これらについて、全ての事例データからユークリッド距離がもっとも近いものを探索した手法と、NMF によるクラスタリングによって同クラスからのみ探索する手法それぞれにおいて生成した MIDI データについて、それぞれ別に被験者 5 名による 5 段階評価を行った。テンポと印象の一致度に関しては皆楽器経験 10 年以上の音楽を専攻する大学生以上の学生に依頼し、Bass の調性と全体的な一致間については楽器経験 10 年以上が 1 名、未経験者 4 名に依頼した。

図は全て横軸が生成した曲名、縦軸が 5 段階評価による点数を表す。図 5、図 6 を見ると全体的にクラスタリングをせずにユークリッド距離だけで編集合成を行ったもののほうが高い傾向にあったが、生成された音楽自体を小節単位で個別にみれば、楽器未経験者を含めた総合評価では全体的に NMF を用いてクラスタリング結果を用いて生成したもののほうが高い評価を得た。

各曲別に見てみると、「Blue Monk」と「How Insensitive」は全体的に評価が低い。その理由として考えられるのは、まず前者はブルースであること、後者は曲全体を支配する調性感が薄く、コードトーンではなく部分ごとの調性が常

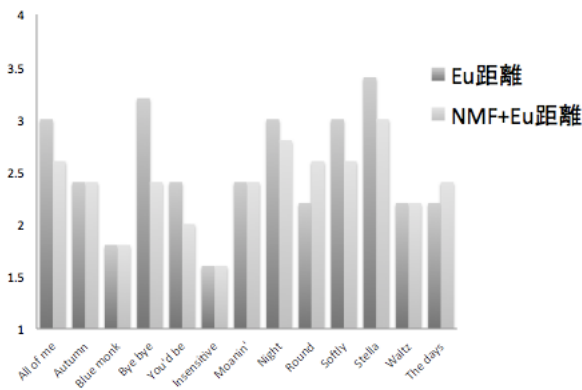


図 6 印象の一致度

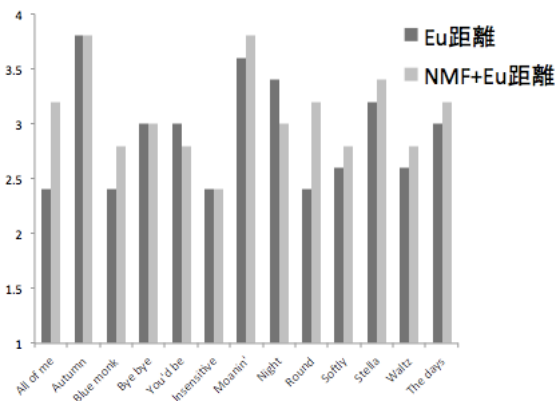


図 7 総合評価

に意識されながら演奏される曲であることが挙げられる。ブルースでは通常ドミナント V7 として扱われることが多いセブンス・コードがブルージーな響きをもったトニックとして使用されるため、他の曲と同様に特徴を捉えるのが困難な曲であったと言える。同様に「How Inesensitive」もまた調性を捉えるのが困難な曲であり、ある種の浮遊感を与える特徴的な曲である。本稿ではまだ十分な学習データを得られているとは言えず、事例データもまた同様に十分量あるとは言えないため、限定されたデータ内でこれらの曲を表現しようとした結果、他の曲よりも低い評価となってしまったものであると考えられる。

また今回は時間分解能として小節単位での編集合成を行ったが、スタイルパラメータの抽出範囲が大きくなると細かいリズムの変化を追跡することは難しい。その結果としてテンポや印象の一致性に対してクラスタリングが大きく寄与できない結果となった。

事例データについては、所々不協和な場所が出てくるが、実験後のヒヤリング等から自然な合成となっているものも多いという一定の評価を得ることができた。多くの場合においてはダイアトニックスケール内の音で対応できるが、常にそれでは単調な演奏にもなりかねない。アヴェイラブルノートスケールとテンションやアヴォイドノートといったものも分析対象に加え、Bass が許容できるフレーズの生

成についてより検討を深める必要がある。

5. おわりに

本稿では統計的に学習可能な枠組みで提案したセッションの数理モデル及び統計解析結果を用いて、実際に Piano のみの演奏データからピアノトリオのセッションデータを出力する作曲システムについて述べた。本システムでは時系列的な制約を考えずに、特徴量の相関関係が維持されていればある程度は自然なセッションデータを生成できるという仮定を元に、Piano 演奏との共起関係のみを考慮した演奏の編集合成を行った。

生成した曲を実際に楽器経験者・未経験者の双方から評価してもらった結果、調性やテンポの不一致など不満点もあるが、数小節程度のみを見れば自然な音となっているものも多く、また実際の演奏でも使用されたフレーズを持つ事例同士が編集合成され活用されていることで、思いもよらない高クオリティな演奏が生成されることもあり、自身の作曲活動にも生かすことができそうだという意見も得られた。

本稿で紹介したジャズピアノトリオ編集合成システムは、現状ではまだいくつかの特殊な仮定を前提としていることやデータ数の問題などまだこれからも解決すべき課題が数多く残っている。今後はまず時系列を考慮して自然な音のつながりになるように編集合成を行うと同時に、事例ではなく実際にインプロビゼーションすることでフレーズを生成するシステムを検討する。

参考文献

- [1] Dannenberg, R. B.: *An On-line Algorithm for Real-time Accompaniment*, Proc. ICMC, pp.193-198, (1984).
- [2] Nishijima, N. and Watanabe, K.: *Interactive Music Composer on Neural Networks*, Proc. ICMC, pp.53-56, (1992).
- [3] Rowe, R.: *Machine Listening and Composing with Cypher*, Computer Music Journal, Vol.16, No.1, pp.43-63, (1992).
- [4] Aono, Y. and Katayose, H. and Inokuchi, S.: *Development of Band-like Musical Assistant System*, IPSJ SIG Technical Reports, 94-MUS-8, Vol.94, No.103, pp.45-50, (1994).
- [5] Aono, Y. and Katayose, H. and Inokuchi, S.: *An Improvisational Accompaniment System Obseving Performer's Musical Gesture*, Proc. ICMC, pp.106-107, (1995).
- [6] Wake, S. and Kato, H. and Saiwaki, N. and Inokuchi, S.: *Cooperative Musical Partner System Using Tension-Parameter: JASPER (Jam Session Partner)*, Trans. IPS Japan, Vol.35, No.7, pp.1469-1481, (1994).
- [7] Goto, M and Hidaka, I. and Matsumoto, H. and Kuroda, Y. and Muraoka, Y.: *A Jazz Session System for Interplay among All Players - VirJa Session (Virtual Jazz Session System)*, Proc. ICMC, pp.346-349(1996).
- [8] Hamanaka, M. and Goto, M. and Asoh, H. and Otsu, N.: *Guitarist Simulator: A Jam Session System Statistically Learning Player's Reactions*, IPSJ Journal,

- Vol.45, No.3, pp.698-709, (2004).
- [9] Nakamura, E and Takeda, R and Yamamoto, R and Saito, Y. and Sako, S. and Sagayama, S.: *Score Following Handling Performances with Arbitrary Repeats and Skips and Automatic Accompaniment*, IPSJ Journal, Vol.54, No.4, pp.1338-1349, (2013).
- [10] 保利武志, 中村和幸, 嵯峨山茂樹: ジャムセッションシステムに向けた数理モデル化と演奏特徴量の統計解析, 情報処理学会研究報告, Vol.2016-MUS-111, No.44, (2016).
- [11] Yamaha,: *Yamaha Music Datashop*, 入手先 (<https://yamahamusicdata.jp/>) (参照 2015-10-14).