

# Ryry: 多声楽器に対応可能な音響入力自動伴奏システム

山本 龍一<sup>1,a)</sup> 酒向 慎司<sup>1,b)</sup> 北村 正<sup>1,c)</sup>

概要: 協奏とは、複数人が同時に演奏を行う演奏形態である。人の演奏にはテンポの揺らぎ、演奏誤りなどの不確実性が含まれるが、人はそのような演奏を聴き、相手に合わせて演奏を行うことができる。本稿では、そのような人の働きを機械で実現する試みとして、楽器演奏者の演奏に自動的に追従して伴奏を再生させる自動伴奏システムを提案する。協奏において人の演奏に合わせてようとする働きは、計算機によって奏者が楽譜上のどの位置を演奏しているのかを瞬時に認識し、次の演奏位置を予測する問題と捉えられ、楽譜追跡と呼ばれる。本研究では、テンポの揺らぎ、演奏誤りなどを含む人の演奏を確率モデルとして記述することで、演奏の不確実性に頑健な楽譜追跡を実現し、多声楽器に対応可能な音響入力自動伴奏システムを構築した。

## Ryry: Automatic Accompaniment System Capable of Polyphonic Instruments

RYUICHI YAMAMOTO<sup>1,a)</sup> SHINJI SAKO<sup>1,b)</sup> TADASHI KITAMURA<sup>1,c)</sup>

**Abstract:** A concert is a live performance which consists of some performers. In a concert, each human performer can play the musical performance correspond to the other one, even if the performance includes uncertainty such as tempo fluctuation and mistakes. In this paper, we propose an automatic accompaniment system as an attempt to realize human-computer concerts, which follows the human performance and play the accompaniment correspond to the performance automatically. The capability of following a human performance is a problem to recognize and anticipate the score position in real-time by computer. It is called score following in the literature. In our work, we introduce a robust score following algorithm by modeling human performances which include uncertainty such as tempo fluctuation and mistakes and developed an automatic accompaniment system which is capable of polyphonic music signal.

### 1. はじめに

協奏とは、複数人が同時に演奏を行う楽器演奏の形態である。例としては、アンサンブル、吹奏楽、オーケストラなどがこれに当てはまる。人同士の協奏においては、複数の楽器演奏者（以下、奏者）が、互いにずれが発生しないように演奏することを期待されるが、人による演奏にはテンポの揺らぎ、時には演奏誤り（音符の置換、挿入、脱落）が含まれるため、一般に容易ではない。しかし、熟練の奏者であれば、相手の演奏を聴き、次にどのように演奏

されるかを瞬時に予測し、相手の演奏に柔軟に合わせるることができる。このような人の働きを機械で実現しようとする試みとして、奏者の演奏に追従して予め記録された伴奏を自動的に再生させる、自動伴奏の研究が盛んに行われている [1-7]。自動伴奏は、協奏の機会が多く得られない奏者に対して奏者単独での仮想的な協奏を実現するほか、伴奏付きの練習支援としての活用が期待される。

人に合わせて演奏するという行為は、相手が楽譜上のどの位置を演奏しているのかを認識し次の演奏位置を予測する働きと、物理的に演奏を行う働きが統合したものとして捉えられる。前者は人の音楽認知機能に関わる非常に重要な働きであり、計算機で実現しようとする試みは楽譜追跡と呼ばれる。本研究では、自動伴奏の練習支援としての活用を想定し、テンポの揺らぎ、演奏誤りや弾き直しを含む

<sup>1</sup> 名古屋工業大学  
Nagoya Institute of Technology, Japan  
a) ryuichi@mmsp.nitech.ac.jp  
b) sako@mmsp.nitech.ac.jp  
c) kitamura@mmsp.nitech.ac.jp

演奏に対しても頑健な楽譜追跡を実現することに重点を置く。既存研究では、MIDI 信号を入力とした演奏誤り、弾き直しにも頑健な楽譜追跡手法および自動伴奏システムは提案されている [4]。しかし、MIDI 出力を備えたごく限られた楽器にしか対応できないという本質的な問題が存在する。一方で、音響信号を入力とする自動伴奏システムも提案されている [5-7]。しかし、多声楽器で音響信号が複雑になりうる場合に頑健に追従可能で、かつ練習支援を想定した弾き直しを考慮したシステムは、楽譜追跡の難しさゆえにいまだに十分なもの存在していない。

そのような背景から、本研究ではテンポの揺らぎ、演奏誤り、弾き直しなどの演奏の不確実性を考慮しつつ、音響信号が複雑になりうる多声楽器の演奏に対しても頑健に楽譜追跡を実現する方法について検討してきた [8-10]。本稿では、[10] で提案した、演奏の確率モデルに基づく楽譜追跡手法を応用し、多声楽器にも対応可能な音響入力自動伴奏システムを提案する。

## 2. 自動伴奏システム Ryry の構成

人同士の協奏では、奏者は互いの演奏を聴き、次にどのように演奏されるかを予測し、相手の演奏に合うように演奏していると考えられる。この考えに基づき、自動伴奏の問題を、特徴抽出、楽譜追跡、伴奏再生の 3 つに分けて考える。図 1 に提案する自動伴奏システム Ryry (リリー) の構成を示す。システムは、上記の 3 つの問題に対応する要素から構成されており、入力演奏の音響信号、出力は同期された伴奏の MIDI 信号である。特徴抽出部では、音響信号から楽譜追跡の手がかりとなる音高、発音の情報が抽出される。具体的には、音響信号を一定のフレーム長とフレームシフトでフレーム分割し、ウェーブレット変換に基づくクロマベクトルと Spectral Flux に基づくオンセット特徴が算出される。それらの 2 つを合わせて特徴量として、楽譜追跡部ではある時刻での演奏位置と瞬時テンポを推定し、伴奏再生部では入力演奏に同期した伴奏を遅延なく再生する。なお、楽譜は既知とし、MIDI データとして与えられるとする。

以下、本システムの核となる楽譜追跡と伴奏再生について述べる。

## 3. 楽譜追跡

### 3.1 楽譜追跡の問題

楽譜追跡は、奏者の演奏に対して、楽譜の情報を元に楽譜上の演奏位置をリアルタイムで求める問題である。ここでは、伴奏再生のために演奏の速さを表すテンポの情報が必要となるため、楽譜追跡はある時刻の演奏位置と瞬時テンポを求める問題とする。自動伴奏においては、単旋律楽器のみならず、多声楽器においても頑健に追従可能であり、かつ練習の場合を想定して、演奏誤り、弾き直し、テンポ

の揺らぎを考慮することが望ましい。そのような楽譜追跡を実現するためには、以下の問題を解決する必要があると考えられる。

- (1) 演奏誤り、弾き直し、テンポの揺らぎなど、多くの演奏の不確実性が含まれる。
- (2) 特に多声音楽の場合に、追跡の手がかりとなる音高や発音時刻の不確実性が増加する。
- (3) 楽譜追跡は、リアルタイムという時間的な制約を満たし、かつ因果的なシステムでなければならない。

上記のうち、(1)、(2) に関しては、多くの研究で議論されてきた [11-16]。しかし、(3) については十分ではなく、著者は新たに (3) の制約の元で頑健な楽譜追跡手法を開発した [8, 10]。

次節では、本システムで用いる楽譜追跡手法について具体的に述べる。

### 3.2 確率モデルに基づく楽譜追跡

提案する楽譜追跡手法の基本的なアイデアは、演奏位置の遷移とテンポの揺らぎの依存関係を確率モデルで記述することにある。確率モデルで表すことで、演奏誤りや弾き直しを見通しよくモデル化することができ、テンポの揺らぎ、音高や発音時刻の不確実性を確率分布の揺らぎとして捉えることができる。また、楽譜追跡の問題は確率モデルの事後確率最大化として定式化することができるほか、確率モデルを記述するパラメータを、データから統計的に自動学習することも可能である。

提案法では、演奏位置の遷移はセミマルコフ条件付き確率場 (Segmental Conditional Random Fields; SCRFs)、テンポの揺らぎは線形動的システム (Linear Dynamical System; LDS) によってモデル化を行った [10]。演奏位置およびテンポの推定は、それぞれ Viterbi アルゴリズム、カルマンフィルタによって効率的に行うことができる。

リアルタイムという制約のため、従来法の多くでは、Viterbi アルゴリズムや動的計画法を逐次的に最適解を出力する形に近似していた [17-19]。しかし、特に信号が複雑な場合においてノイズに過敏に反応して誤推定を起こしてしまうことが考えられる。そこで、 $\alpha$  (遅延幅) 秒過去の演奏位置を Viterbi アルゴリズムによって推定し、現在位置はテンポにより予測する頑健な楽譜追跡方法を考案した。基本的なアルゴリズムは以下の通りである。

**Step 1** Viterbi アルゴリズムにより入力信号の演奏位置と発音時刻の系列を求める。

**Step 2** Step 1 の結果からカルマンフィルタによってテンポの推定を行う。

**Step 3** 遅延幅  $\alpha$  を一つ前の発音時刻までの時間と設定し、演奏位置の遷移の 1 次近似モデルにより現在の演奏位置を予測する。

上記のアルゴリズムを音響信号が入力される度に繰り返す

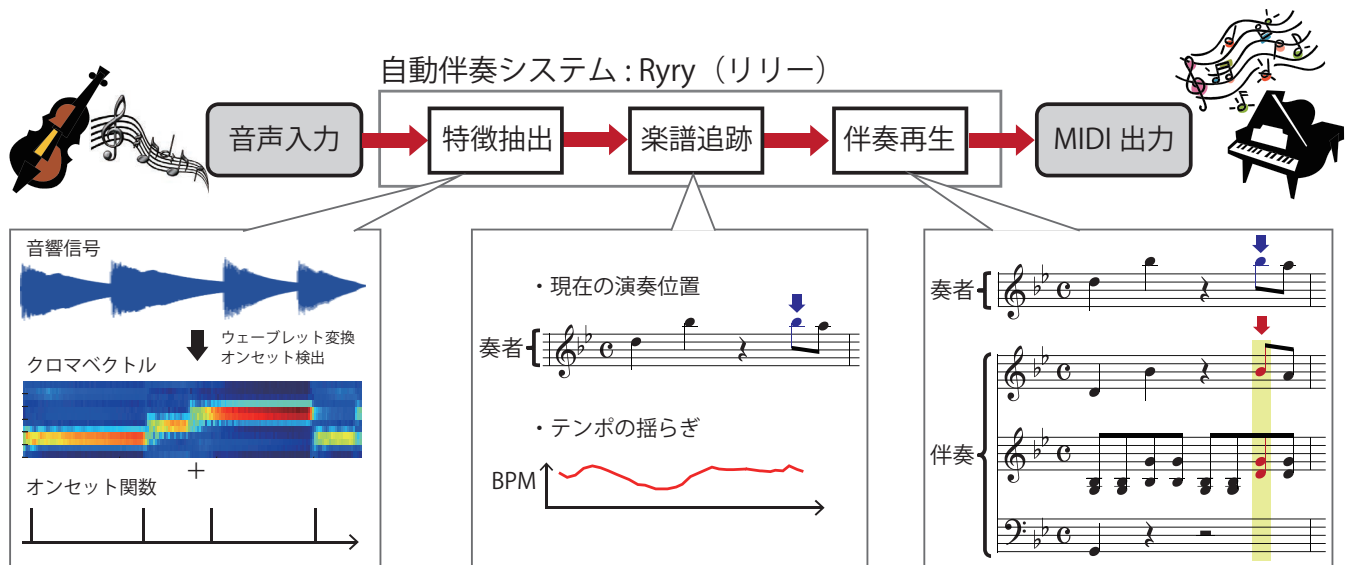


図 1 自動伴奏システム Ryry (リリー) の概要 .

Fig. 1 The outline of automatic accompaniment system “Ryry”

ことにより、リアルタイムで演奏位置とテンポを求めることができる。本システムでは、この楽譜追跡アルゴリズムを用いる。詳細は、[8, 10] を参照されたい。

#### 4. 伴奏再生

楽譜追跡の結果より得られる演奏位置とテンポから、伴奏の演奏位置とテンポを決定し、リアルタイムに同期された伴奏を再生する問題を伴奏再生とする。なお、伴奏の出力には MIDI 信号を用いる。伴奏に音響信号を用いる自動伴奏システム [6, 7] も存在し、それらは MIDI 信号では表現できない表情豊かな伴奏を可能にするほか、既存の音源をそのまま利用できる利点がある一方で、伴奏用の特定のパートだけの音響信号を予め用意しておく必要がある。これには、ある音源から特定のパートだけを除去するマイナスイオンと呼ばれる技術が必要であり、一般に困難である。MIDI であればそのような困難はなく、MIDI データを用意するだけで自動伴奏を実現することができる。

人同士が協奏を行う際には、協奏する相手の次の演奏のタイミングを予測して、自身の演奏を合わせようと考えられる。ここでは、このような予測に基づいて、奏者の演奏に追従する伴奏再生手法を考える。ある時点において、奏者と伴奏の演奏位置および瞬時テンポが一致し、かつ瞬時テンポが局所的に変化しないと仮定すれば、演奏位置の遷移の 1 次近似によって、伴奏の演奏のタイミングを予測することができると考えられる。しかし、奏者の演奏にはテンポの揺らぎが含まれるため、時間が経つにつれて奏者と伴奏の演奏位置にずれが生じることが予想される。したがって、本研究では、楽譜追跡において推定された奏者の演奏位置とテンポをリアルタイムで伴奏に同期することで、伴奏の演奏のタイミングを予測する方法を用いる。

計算機による MIDI 演奏であれば、伴奏の演奏をほぼ遅延なく再生することができるため、予測された位置に基づき伴奏の MIDI 信号を再生させることで、奏者に瞬時に追従する自動伴奏を実現することができる。弾き直しが行われた場合には、奏者と伴奏に演奏位置のずれが生じるが、そのような場合には伴奏を一時停止し、奏者の演奏位置とテンポを伴奏に同期することで対応可能である。また、実際には楽譜追跡は誤推定の可能性を考慮して、楽譜追跡が一定時間楽譜通りの演奏位置の遷移を出力していれば伴奏の MIDI 信号を再生させ、そうでなければ伴奏を一時停止することで、不自然な伴奏を回避することができる。

#### 5. Ryry の仕様、動作環境

Ryry はクロスプラットフォームアプリケーションを設計方針としており、Windows, Mac OS X, Linux などの多くの環境で動作する。使用に当たっては、最低限以下を揃える必要がある。

- (1) Ryry が動作する PC
- (2) 演奏する楽曲の MIDI ファイル
- (3) 音声入力デバイス (マイクロホン)
- (4) 伴奏用 MIDI 音源
- (5) 音声出力デバイス (スピーカー)

本システムでは、MIDI ファイルを用意することで任意の楽曲で自動伴奏を行うことができる。ユーザは、MIDI ファイルに含まれる複数のトラックから演奏するトラックと自動伴奏を行うトラックを選択し、自動伴奏の開始、停止を行うことができる。ただし、実際には上記に加えて演奏するパートの楽譜も必要である。なお、PC 付属のマイクロホンを含め、一般的なマイクロホンで動作することを確認しているが、楽譜追跡の誤推定を避けるため、周囲の

ノイズの影響を受けにくい、楽器に直接取り付け可能なピエゾタイプが望ましい。

本システムを用いて自動伴奏を行う様子の動画が\*1にて公開されている。

## 6. おわりに

本稿では、人と機械の協奏を実現する試みとして、多声楽器にも対応可能な音響入力自動伴奏システム Ryry を提案した。本研究では、協奏において人の演奏に合わせる働きを、奏者の演奏を聴く、楽譜上の演奏位置を認識・予測する、実際に演奏を行うという3つ分け、それらを計算機で解決することで自動伴奏を実現した。提案する自動伴奏システムは、演奏誤りや弾き直しなどの不確実性を含む人の演奏に対して頑健な楽譜追跡手法を応用することで、音響信号が複雑になりうる多声楽器であっても柔軟に追従することができる。

今後の課題としては、楽譜追跡アルゴリズムの計算量の削減がある。本稿で提案した楽譜追跡手法は、不確実性を含む多声楽器の演奏にも頑健な一方で、楽譜に含まれる和音数の増大に従ってリアルタイムでの動作が難しくなるため、今後この問題の解決に取り組みたいと考えている。

謝辞 本研究は、市原国際奨学財団および名古屋工業大学研究推進経費の支援を受けた。

## 参考文献

- [1] Dannenberg, R.: An on-line algorithm for real-time accompaniment, *Proc. ICMC*, pp. 193–198, (1984).
- [2] Vercoe, B.: The synthetic performer in the context of live performance, *Proc. ICMC*, pp. 199–200, (1984).
- [3] Bloch, J. and Dannenberg, R.: Real-time computer accompaniment of keyboard performances, *Proc. ICMC*, pp. 279–289, (1985).
- [4] 武田 晴人, 西本 卓也, 嵯峨山 茂樹.: HMM による MIDI 演奏の楽譜追跡と自動伴奏, 情報処理学会研究報告, vol.2006-MUS-66(90), pp. 109–116, (2006).
- [5] Cont, A.: ANTESCOFO: Anticipatory Synchronization and Control of Interactive Parameters in Computer Music, *Proc. ICMC*, (2008).
- [6] Raphael, C.: Music Plus One: A System for Expressive and Flexible Musical Accompaniment, *Proc. ICMC* (2001).
- [7] 中村 友彦, 水野 優, 鈴木 孝輔, 中村 栄太, 樋口 祐介, 深山 覚, 嵯峨山 茂樹.: 音楽演奏の誤りや反復に頑健な音響入力自動伴奏, 日本音響学会 2012 年秋季研究発表会, pp. 931–934, (2012).
- [8] 山本 龍一, 酒向 慎司, 北村 正.: 隠れセミマルコフモデルと線形動的システムを組み合わせた音楽音響信号と楽譜の実時間アライメント手法, 情報処理学会研究報告, vol.2012-MUS-96(13), pp.1–6, (2012).
- [9] 山本 龍一, 酒向 慎司, 北村 正.: セミマルコフ条件付き確率場を用いた音楽音響信号と楽譜のアライメント, 日本音響学会 2012 年秋季研究発表会, pp. 935–936, (2012).
- [10] Yamamoto, R., Sako, S., Kitamura, T.: Real-time Audio to Score Alignment Using Semi-Markov Conditional Random Fields and Linear Dynamical System, *MIREX submission*, (2012).
- [11] Dixon, S.: An On-line Time Warping Algorithm for Tracking Musical Performances, *Proc. IJCAI*, pp.1727–1728, (2005).
- [12] Dannenberg, R. and Hu, N.: Polyphonic audio matching for score following and intelligent audio editors, *Proc. ICMC*, pp. 27–34, (2003).
- [13] Raphael, C.: Aligning music audio with symbolic scores using a hybrid graphical model, *Machine Learning*, vol.65 no.2-3, pp389–409, (2006).
- [14] Cont, A.: A coupled duration-focused architecture for real-time music-to-score alignment, *IEEE, Trans. on Pattern Analysis and Machine Intelligence*, vol.32, no.6, pp974–987, (2010).
- [15] Otsuka, T., Nakadai, K., Takahashi, T., Komatani, K., Ogata, T. and Okuno, H. G.: Design and implementation of two-level synchronization for interactive music robot, *Proc. AAAI*, (2010).
- [16] Otsuka, T., Nakadai, K., Takahashi, T., Ogata, T. and Okuno, H. G.: Real-time audio-to-score alignment using particle filter for coplayer music robots, *EURASIP J. Adv. Signal Process*, (2011).
- [17] Cano P., Lascos A., Bonada J.: Score-Performance Matching Using HMMs, *Proc. ICMC*, pp. 441–444, (1999).
- [18] Orio, N., Dechelle, F.: Score Following Using Spectral Analysis and Hidden Markov Models, *Proc. ICMC* pp.151–154, (2001).
- [19] Suzuki, K., Ueda, Y., Raczynski, S. A., Ono, N. and Sagayama, S.: Real-time Audio to Score Alignment Using Locally-constrained Dynamic Time Warping of Chromagrams, *MIREX submission*, (2011).

\*1 <http://www.mmsp.nitech.ac.jp/~ryuichi/interaction2013/>