

音楽演奏の誤りや反復に頑健な音響入力自動伴奏

中村友彦, 水野優, 鈴木孝輔, 中村栄太 (東大院・情報理工),
樋口祐介 (東大・工), 深山覚, 嵯峨山茂樹 (東大院・情報理工)*

1 序論

本研究では, 楽譜をもとに人間の演奏に対して実時間で追従しつつ, 計算機に伴奏を同期して再生させる自動伴奏の実現を目指す. この技術は, 合奏曲の個人練習や一人でのオーケストラ伴奏を伴う曲の演奏を実現するために利用できる. Dannenberg や Vercoe の研究以来, 自動伴奏の研究は多い [1-3].

練習における演奏には, 演奏誤り (音符の音高の誤り), 演奏脱落 (音符の脱落), 弾き飛ばし, 弾き直しが含まれる. そのため, これらに適応可能な自動伴奏が望ましい. そのような自動伴奏として, MIDI 演奏入出力に対応した武田らによるものがある [4]. しかし, 依然音響信号を入力としてこのような演奏に対応することは容易ではない.

そこで, 本研究は演奏音響信号を入力とし, 演奏誤り, 演奏脱落, 弾き飛ばし・弾き直しを含む演奏に実時間で追従しつつ, 伴奏を同期再生する自動伴奏の実現を目的とする. 伴奏として音響信号を用いることで, 伴奏に MIDI では再現しづらい音が含まれる場合や伴奏音響データのみが入手可能な場合にも自動伴奏を行うことが可能である. 簡単のため, 単旋律楽器による演奏音響信号を入力とする.

2 自動伴奏アルゴリズムの構成

本研究の目的とする自動伴奏は, 演奏誤り, 演奏脱落, 弾き飛ばし・弾き直しを含む演奏の音響信号に対して, 伴奏を同期再生するものである (Fig. 1). これを実現するため,

1. 演奏音響信号から楽譜上の演奏箇所をいかに推定するか (楽譜追跡)
2. 伴奏の再生速度をどう推定すれば良いのか (演奏速度推定)
3. 録音された伴奏音響信号の再生速度をいかにして変換するか (再生速度変換)

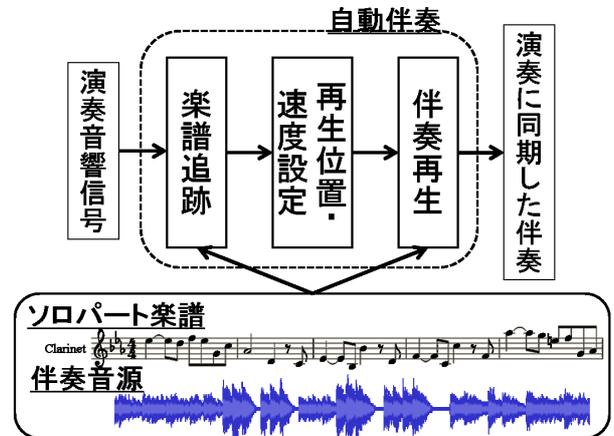


Fig. 1 提案する自動伴奏の構成

の3つの問題に取り組む必要がある. そのため次節から, 演奏誤り, 演奏脱落, 弾き飛ばし・弾き直しを含む演奏音響信号からの楽譜追跡, 伴奏の再生位置・速度の設定, 伴奏の再生速度変換について考える.

3 隠れマルコフモデルに基づく演奏音響信号の楽譜追跡

3.1 楽譜追跡の方針

本研究で考える演奏では, テンポ変動, 音長の変動に加え, 練習時に起こる弾き誤り, 弾き飛ばし・弾き直しによる音高の不一致や音符の脱落が存在する. また, 音響信号からの音高の正確な推定は容易ではなく, 楽譜追跡においてはこれも考慮する必要がある.

人間が伴奏する場合, 演奏者は楽譜を見て演奏し, 伴奏者は演奏者の演奏を聴いて楽譜上の演奏箇所を推定し, その箇所を伴奏する. すなわち, 伴奏者は演奏音響信号から演奏箇所を推定するという逆問題を逐次解いていると捉えられる. この解釈に基づけば, 演奏音響信号を入力とする楽譜追跡の方法は以下のように考えられる. 楽譜上の音符を状態と考えると, 楽譜に沿って演奏することは状態を遷移していく過程であり, この状態遷移に基づいて演奏の音響信号が生成

* Automatic accompaniment robust to errors and repetitions in acoustic musical performances by Tomohiko Nakamura, Yuu Mizuno, Kosuke Suzuki, Eita Nakamura, Yusuke Higuchi, Satoru Fukayama, Shigeki Sagayama (University of Tokyo)

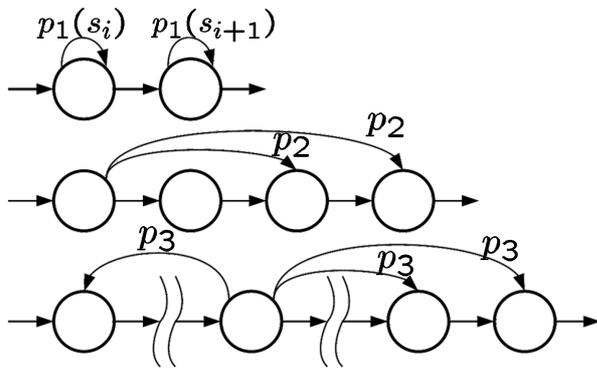


Fig. 2 遷移確率 p_1 , p_2 , p_3 を割り振った隠れマルコフモデル. p_1 : 音符 s_i の音長に対応. p_2 : 音符の脱落に対応. p_3 : 弾き飛ばし・弾き直しに対応.

されると捉えられる. 演奏誤り, 演奏脱落, 弾き飛ばし・弾き直しは確率的に起こると考えられ, 直前の音符のみに依存すると仮定すると演奏過程は一次マルコフモデルと捉えられる. さらに, 演奏音響信号は同一の楽譜箇所の演奏においても毎回変動するので, 音響信号の生成も確率的であると考えられる. このようなモデルは, 隠れマルコフモデル (HMM) と呼ばれ従来研究でも用いられている [5]. 演奏誤り, 演奏脱落, 弾き飛ばし・弾き直しを含む演奏過程も, HMM の出力確率や遷移確率によって表現できると考えられる.

3.2 演奏脱落, 弾き飛ばし・弾き直しに対応した HMM のトポロジー

音符が状態に対応するので, 演奏脱落は数个先への状態遷移により, 弾き飛ばし・弾き直しは自身以外への状態遷移によって表される (Fig. 2). また, 同一音符の演奏継続は自己遷移によって表現され, 自己遷移回数の期待値が音価に対応するように自己遷移確率を決定できる. 楽譜通りの演奏は 1 つ先の状態への遷移によって表現可能である.

3.3 演奏誤りに対応した HMM の出力確率の設定

同一音高であっても, 演奏音響信号は大きく異なる場合がある. そのため, 音高情報に感度が高く音色の変化に頑健である特徴量を音響信号から抽出する必要がある. クロマベクトルは, 各ピッチクラスの周波数のパワーを複数のオクターブに渡って加算した 12 次元のベクトルであり, 上記の条件を満たす [6]. また, クロマベク

トルのみでは捉えることは難しいオクターブの違いや同音連打については, オンセット時にスペクトルが大きく変動することに注目することにより, spectral flux を用いて次の状態への遷移確率を補正した [7]. 一般的に楽器ごとに倍音構造が異なるが, これは楽器ごとにクロマベクトルを学習することにより解決可能である [8].

次に, 弾き誤りを HMM の出力確率によりモデル化する. 各状態から音符のピッチクラスに対応したクロマベクトルが変動を伴い出力されると考えると, 弾き誤りは音符のピッチクラスとは異なるピッチクラスのクロマベクトルが出力されることにより表現される.

3.4 楽譜上の演奏箇所の実時間推定

自動伴奏においては, 演奏に合わせて実時間で伴奏を同期して再生しなくてはならないため, 楽譜上の演奏箇所を実時間で推定する必要がある. 任意箇所から任意箇所への弾き飛ばし・弾き直しの可能性がある演奏の楽譜追跡は, HMM 上で可能な状態遷移が前状態数の 2 乗オーダーになり, 状態系列の推定のために多くの場合使われる通常の Viterbi アルゴリズムでは, 状態数が多くなると実時間で状態推定を行うことが困難である. そのため, 高速 Viterbi アルゴリズムにより計算量を削減し, 逐次最尤状態系列から演奏箇所の推定を行った [9].

4 伴奏再生のための演奏速度推定

演奏入力に対して伴奏を同期再生するためには, 各時刻における伴奏音響信号の再生位置と再生速度の設定を行う必要がある. 楽譜追跡によって得られる情報は, 演奏者が楽譜上の演奏箇所の情報のみであるため, 演奏速度を推定する必要がある. また, 音響入力の楽譜追跡の結果には誤りが含まれている, という点も対応する必要がある.

武田ら [4] は, 人間の伴奏者がどのように再生位置・再生速度を判断しているかについての考察に基づき, 楽譜追跡結果において推定音符状態が変化した場合に, 過去の数回の音符状態の変化の履歴から再生位置・再生速度を設定する手法を提案した. 本研究では, これを楽譜追跡に誤りが含まれる場合に拡張し, 推定状態系列を直近数秒前の時刻まで traceback する手法を用いて, 楽譜追跡の誤りに対応した.

5 位相再構成法に基づく伴奏の実時間再生速度変換

伴奏の再生速度変換では、音高を保存したまま再生速度を実時間で変換する必要がある。また、多くの音楽は様々な楽器の混合音であるため、多重音にも適した変換や音質の劣化も少ないことが望ましい。

人間の聴覚系は、パワースペクトログラムに類似した特徴に依存して知覚していると考えられている [10]。そのため、所望の再生速度に応じたパワースペクトログラムを設計すると、聴感上自然な音響信号が得られると考えられる。そこで、パワースペクトログラム領域で速度変換を行い対応する信号波形を構成することで、伴奏の再生速度変換を実現できる。

パワースペクトログラムの時間方向の伸縮は、短時間フーリエ変換のフレームシフトを変化させることにより実現できる。例えば、再生速度を a 倍にするためには、フレームシフトを a 倍しパワースペクトログラム全体のフレーム数を $1/a$ 倍にすればよい (Fig. 3)。その後、適切な位相を付与し逆短時間フーリエ変換を行うことで、再生速度を変換した信号波形を合成できる [11]。

6 自動伴奏の評価

6.1 実験条件

本節では、提案する自動伴奏アルゴリズムの有効性を実験により確認する。評価指標として、弾き飛ばし・弾き直しへ追従するまでにかかる時間 (Following Time, FT) を用いた。FT が短ければ、弾き飛ばし・弾き直しに対する追従性能が高いことを示している。正解の演奏箇所から十分近い (時間 Δ 以内) 伴奏箇所が再生されたとき、伴奏再生が追従したとみなし、この条件が満たされないまま再度弾き飛ばし・弾き直しが起こった場合は追従できなかったとした。

対象楽曲として、RWC 音楽データベースを含むクラシック 3 曲、ポップス 2 曲、民謡・童謡 2 曲を用いた [12]。演奏データとしては、アマチュア演奏家 1 名によるクラリネットの演奏計 1305 s (平均 120 s 程度) を用いた。正解データは、人手により演奏箇所推定を行い用意した。演奏データでは 16 bit 量子化 16 kHz サンプル周波数を用い、クロマベクトルの解析フレーム長は 128 ms (2048 点)、フレームシフトは 32 ms (512 点) をとした。HMM の出力確率の学習のため、

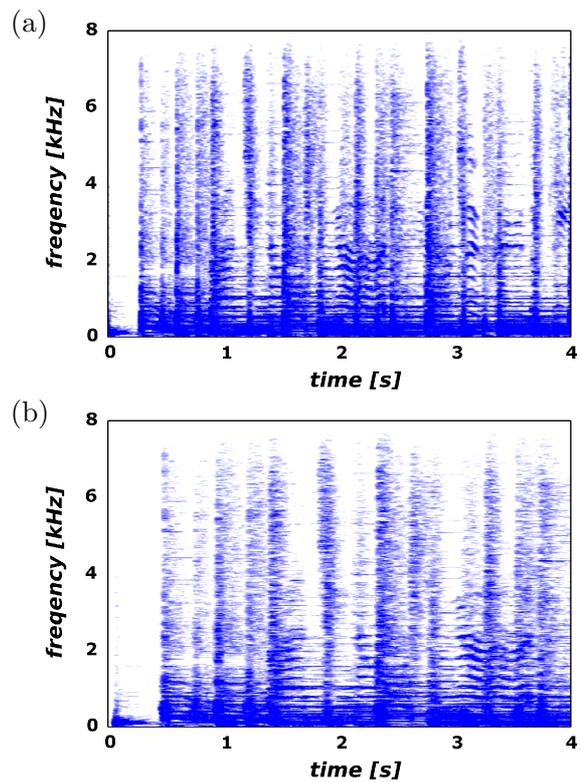


Fig. 3 スペクトログラムの伸縮による再生速度変換。(a) は変換前のパワースペクトログラムであり、(b) は変換後のパワースペクトログラムである。(b) は (a) を時間方向 (横軸) に伸ばし再生速度を遅くしたものである。

RWC 楽器音データベースからクラリネットの楽器音を用いた。

6.2 弾き飛ばし・弾き直しに対する追従性能の評価

実験に用いた演奏には、弾き飛ばし・弾き直しが 44 箇所含まれていた。 $\Delta = 1$ s の場合、弾き飛ばし・弾き直しに追従できた箇所は 27 箇所であり、その場合の FT の平均は約 4.4 s あった。0.5 s ごとに区切った FT のヒストグラム (Fig. 4) に見られるように、多くの箇所で 2 ~ 5 s 程度の遅れで正しく追従できていた。楽譜上の繰り返しや類似した音形に起因する追従誤りは不可避免の問題である。実際、追従できなかった箇所の多くは楽譜上の繰り返し・類似箇所に誤推定されていた。たとえば、ポピュラー音楽では 1 番の演奏が、2 番の演奏として推定された場合が存在した。このような繰り返し・類似箇所での誤推定は、HMM のトポロジーや遷移確率を適切に設定することにより改善が可能であると考えられる。

提案法による伴奏再生箇所と真の演奏箇所の

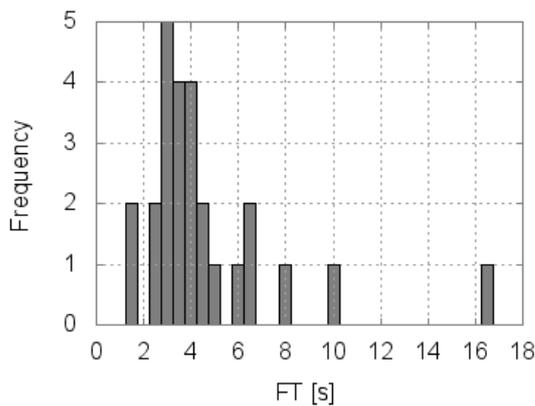


Fig. 4 弾き飛ばし・引き直しへ追従するまでの時間のヒストグラム

対応関係の例を, Fig. 5 に示す. 横軸が演奏者の演奏時刻を表し, 縦軸は楽譜通りに一定のテンポで演奏した場合の時刻を表す. Fig. 5 (a) は, 弾き直しが存在する場合でも 2 s 程度で追従できたことを示している. Fig. 5 (a), (b) 共に, 弾き飛ばし・弾き直し無く演奏している箇所では, 弾き飛ばし・弾き直しであると誤推定されている箇所が少ないことが分かる.

7 結論

本研究では, 演奏誤り, 演奏脱落, 弾き飛ばし, 弾き直しを含む演奏に対応した音響入力自動伴奏を実現するために, 楽譜追跡, 再生速度・再生位置設定, 音響信号の実時間再生速度変換の 3 つの問題について扱った. 楽譜追跡については, HMM に基づく演奏のモデル化を行い, 演奏誤り, 演奏脱落, 弾き飛ばし, 弾き直しを含む演奏音響信号に対しての楽譜追跡手法を提案した. また, 実験によって弾き飛ばし・弾き直しに対する有効性を確認した.

今後の展望としては, ギターなどの多重音楽器への対応や楽譜追跡手法の応用による伴奏信号と楽譜の自動インデックス付け, 楽器音分離手法を用いたマイナスイオン音源の自動生成により, より使い勝手の良い自動伴奏を実現することが挙げられる.

謝辞 本研究の一部は, 文部科学省/学術振興会科学研究補助費 課題番号 (23240021) から補助を受けて行われた. クラリネット演奏により評価に協力して下さった伊藤直哉氏に感謝する.

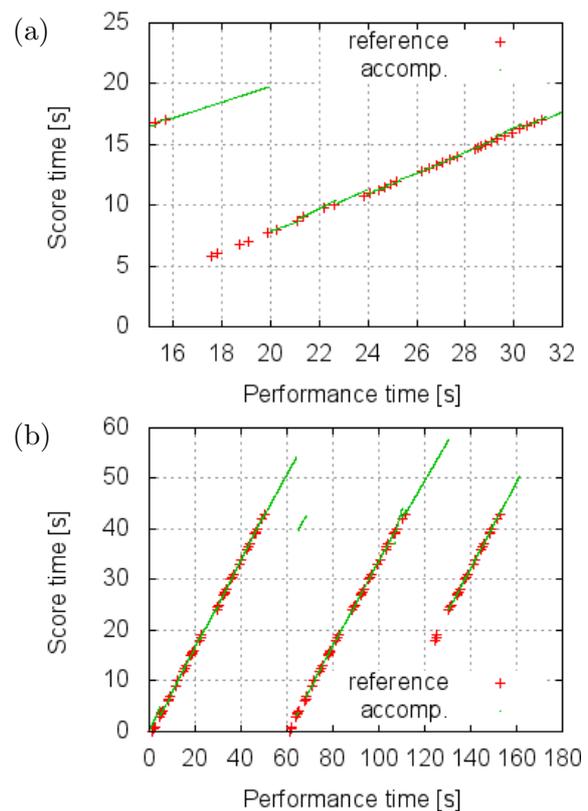


Fig. 5 提案法による伴奏再生箇所 (accomp.) と真の演奏箇所 (reference) の対応関係の例. (a) は Brahms の Clarinet Sonata 第 2 番第 1 楽章の冒頭に対する自動伴奏の結果である. (b) は, Amazing Grace に対する自動伴奏の結果である.

参考文献

- [1] R. B. Dannenberg, *Proc. ICMC*, pp. 193–198, 1984.
- [2] B. Vercoe, *Proc. ICMC*, pp. 199–200, 1984.
- [3] N. Orio *et al.*, *NIME*, pp. 36–41, 2003.
- [4] 武田他, 情処研報, 2006-MUS-66 (90), pp. 109–116, 2006.
- [5] A. Loscos *et al.*, *Proc. ICMC*, pp. 441–444, 1999.
- [6] T. Fujishima, *Proc. ICMC*, 464–467, 1999.
- [7] P. Masri, PhD thesis, University of Bristol, 1996.
- [8] 鈴木他, 情処研報, 2011-MUS-89, pp. 1–6, 2011.
- [9] 中村他, 情処論文誌 投稿中, 2012
- [10] B. C. J. ムーア, 大串健吾 (監訳), 聴覚心理学概論, 誠信書房
- [11] 水野他, 音講論, pp. 897–898, Sep, 2011.
- [12] 後藤他, 情処研報, 2001-MUS-42-6, pp. 35–42, 2001.