

# 打楽器とロボットとの合奏のための 結合振動子モデルに基づく打撃時刻予測

水本武志<sup>†</sup> 中臺一博<sup>‡</sup> 大塚琢馬<sup>†</sup> 高橋徹<sup>†</sup> 尾形哲也<sup>†</sup> 奥乃博<sup>†</sup>

<sup>†</sup> 京都大学大学院情報学研究科

<sup>‡</sup> (株) ホンダ・リサーチ・インスティテュートジャパン, 東京工業大学大学院情報理工学研究科

## Onset Prediction using Coupled Oscillator Model for Human-Robot Ensemble

\*Takeshi Mizumoto<sup>†</sup> Kazuhiro Nakadai<sup>‡</sup> Takuma Otsuka<sup>†</sup>

Toru Takahashi<sup>†</sup> Tetsuya Ogata<sup>†</sup> Hiroshi G. Okuno<sup>†</sup>

<sup>†</sup> Graduate School of Informatics, Kyoto University <sup>‡</sup> Honda Research Institute Japan, Co., Ltd. and Graduate School of Information Science and Engineering, Tokyo Institute of Technology

**Abstract**— We present a novel beat-time prediction method for a human-robot ensemble using a coupled oscillator model. A prediction capability is essential for an ensemble to achieve synchronization between the human's and the robot's performances because the robot is required to play the instrument with the human on time. However, conventional ensemble studies have focused on adaptation of a playing speed, instead of prediction. To solve the problem, we develop a beat-time prediction method using the assumption that a human's drumming is modeled as an oscillator. Experimental results shows that our method reduces an onset error by 38% in the maximum.

**Key Words:** Ensemble, Theremin, Coupled Oscillator, Synchronization

### 1. はじめに

近年、楽器を演奏するロボットを用いて人と合奏することを目指した研究が始まっている [1-3] 従来の単独で楽器を演奏するロボット、例えばフルート演奏ロボットの WF シリーズ [4] などに対して、合奏は人がロボットと共に楽しめる参加型エンタテインメントを提供できる新たな応用として期待されている。特に音楽は言語や年齢への依存度が低いので、多くの人々を楽しませられるという点で有望なエンタテインメントといえる。

本稿では、合奏を異なるリズムや旋律を演奏している演奏者達がインタラクションを通して相互に同期していく演奏と定義する。このような合奏には 3 つの要素: (1) 楽器演奏ロボット, (2) 楽器演奏者, (3) 合奏中に演奏者同士で自律的に演奏を同期する機構が必要である。我々は、これまでに最も基本的な 2 人の合奏に焦点を当て、(1) にテルミン演奏ロボットを、(2) にドラム奏者を用いて合奏システムを開発している。(3) は、従来は実時間ビートトラッキングを用いた演奏速度の動的適応によって同期していた [5] が、その精度には限界があった。これは、ロボットが人と同時に演奏するために必須の予測機能が実装されていなかったからである。人の演奏を予測できないロボットは、人の演奏を認識した後にしか動作できないために遅れが避けられない。

そこで本稿では、生物のモデル化などに広く用いられている結合振動子モデル [6] に基づいてドラムの打撃時刻を予測する手法を開発し、合奏における同期の精度を向上させる。本手法は、演奏リズムは振動子でモデル化できると仮定し、結合振動子による引き込みで打撃時刻を予測する。本手法の利点は、人のリズム知覚モデル [7] と

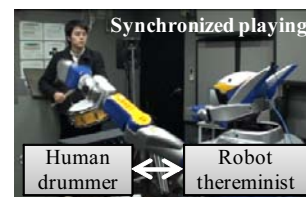


Fig.1 ドラムとテルミンの合奏

一貫した根拠のあるモデルである点と、結合する振動子の数を増やすことで 3 名以上の合奏への適用も理論的には可能である点の 2 点である。

以下では、まず従来の合奏研究を述べて本研究の位置づけを明かにし、次に結合振動子モデルと打撃時刻予測への応用について述べる。そして実験で従来のビート間隔のみを用いる手法に対する優位性を示す。

### 2. 従来の合奏研究

合奏に関する研究は、Dannenberg [8] の計算機と人の合奏から始まった。我々が扱う人とロボットとの合奏が計算機との合奏と異なるのは、演奏者が身体を持っている点である。演奏者の身体性は、演奏時の存在感を高めるだけでなく、共演者である人への影響という点でも重要である。例えば演奏者の体の動きと楽譜構造との相関に関する報告 [9] はこの主張を支持する結果である。

次に、3 つの主要な人とロボットの合奏の関連研究と本研究との違いについて述べる。Peterson らはフルート演奏ロボットとサクソ演奏者の合奏システムを開発した [2]。彼らは、演奏者が交互に演奏することを仮定しているため、すべての演奏者が同時に演奏する状況には対

応できない。Weinberg らは、人のドラム奏者とキーボード奏者、ロボットのドラム奏者と木琴奏者による 4 名の即興合奏システムを開発した [1]。ロボットは、人の演奏の模倣か類似の演奏を行うよう設計されている。従って、あるロボットが合奏に参加するにはそれと類似する楽器を演奏する人が必要である。我々は各演奏者が異なるメロディや旋律を演奏する合奏を目指すので、そのような形態の合奏には対応できない。Otsuka らは、テルミン演奏ロボットと人のドラム演奏の合奏システムを開発した [5]。我々と同様に各演奏者が同時に演奏することを仮定しているが、共演者の演奏の予測という視点を欠いているので、ビート時刻の誤差は大きい。

### 3. 結合振動子モデルに基づく打撃時刻予測

本節では、まず一般的な結合振動子モデルの概要を述べ、次に打撃時刻予測への応用について述べる。最後に本予測手法を用いた合奏システムの実装について述べる。

本手法のキーアイデアは、人の打撃タイミングとロボットが演奏する音符の遷移タイミングを共に振動子でモデル化する点である。本アイデアによって、結合振動子の引き込みで相手の演奏の予測が可能となる。

#### 3.1 結合振動子モデル

一般的な結合振動子モデルは、振動子とその相互作用から構成される。振動子とは、時間が経過するにつれて単調に増加する位相  $\phi(t)$  を独立変数にもつ自律システムであり、例えば振り子時計が振動子でモデル化できる。振動子の位相は次式で表される。

$$\phi(t) = (\phi_0 + 2\pi t/T_{osc}), \quad (1)$$

ただし、 $t$  は時間、 $T_{osc}$  は位相の変化速度、 $\phi_0$  は位相の初期値である。ここで、 $2\pi$  だけ異なる位相は振動子の同一の状態を表す。式 (1) を微分方程式で書き直すと、位相は次式でも表せる。

$$\frac{d\phi_1}{dt} = \omega_1, \quad (2)$$

ただし、 $\omega_1$  は振動子の角速度を表す。式 (2) から分かる通り、振動子は単独では位相が時間経過につれて速度  $\omega_1$  で単調に増加する系である。

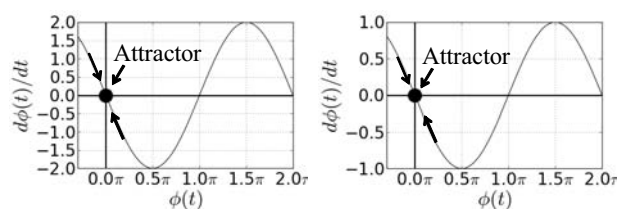
次に、2 つの振動子を結合する。一方の振動子の位相をもう一方の振動子の位相に反映するために、式 (2) に位相差を引数にとる周期  $2\pi$  の周期関数を相互作用項として加える。すると、二つの振動子それぞれの位相方程式は次のようになる

$$\frac{d\phi_1}{dt} = \omega_1 + K_1 Q(\phi_2 - \phi_1) \quad (3)$$

$$\frac{d\phi_2}{dt} = \omega_2 + K_2 Q(\phi_1 - \phi_2), \quad (4)$$

ただし、 $\phi_1, \phi_2$  は各振動子の位相、 $Q$  は周期  $2\pi$  の周期を持つ関数、 $K_1, K_2$  は正の結合強度、そして  $\omega_1, \omega_2$  は自然周期、すなわち結合が無いときの周期である。

関数  $Q$  に  $\sin$  関数を用いた蔵本モデル [10] は、最も基本的な結合振動子モデルとして知られている。位相方



(a)  $K_1 = K_2 = 1$                       (b)  $K_1 = 0, K_2 = 1$

Fig.2 蔵本モデルのアトラクタ

程式は以下の通りである

$$\frac{d\phi_1}{dt} = \omega_1 + K_1 \sin(\phi_2 - \phi_1) \quad (5)$$

$$\frac{d\phi_2}{dt} = \omega_2 + K_2 \sin(\phi_1 - \phi_2) \quad (6)$$

相互作用関数に  $\sin$  関数を用いることで、結合振動子の振る舞いを解析的に調べることが可能になる。以下では、振動子の位相差  $\phi = \phi_1 - \phi_2$  を用いて結合係数と位相差の関係性を議論する。 $\phi$  の式は式 (5) から式 (6) の両辺を引くことで求める：

$$\frac{d\phi}{dt} = \omega_1 - \omega_2 + K_1 \sin(-\phi) - K_2 \sin(\phi) \quad (7)$$

$$= (\omega_1 - \omega_2) - (K_1 + K_2) \sin(\phi) \quad (8)$$

それぞれの自然周波数  $\omega_1, \omega_2$  が等しいと仮定すると、上式をグラフに描画することで二つの振動子の振る舞いを調べられる。

位相差の振る舞いを図 2 に示す。縦軸は位相差の微分係数を、横軸は位相差を表す。図中の黒い丸は位相が変化しても時間経過に従って収束する安定停留点 (アトラクタ) である。黒い丸がアトラクタである理由は以下の通りである：位相がアトラクタから減少すると微分係数が正の値をとるので位相は増加し、増加すると微分係数は負の値をとるので位相は減少する。従って、時間が経過するにつれて位相差はアトラクタに収束する。図 2(a) は二つの振動子が等しい強度で影響を受ける状況、(b) は振動子 2 のみが影響をうける状況を表すが、どちらでも位相差 0 がアトラクタである。

#### 3.2 打撃時刻予測への応用

蔵本モデルを打撃時刻予測に用いるため、合奏に対して次の仮定をおく：

1. 演奏者は内部に振動子を持ち、振動子の位相が 0 のときにオンセットが出力されるような演奏を行う。例えば内部振動子の位相が 0 のときにドラム奏者はドラムを打ち、テルミン奏者は音符を遷移させる。
2. 演奏者は共演者の位相を、オンセットのみによって知る。例えばテルミン奏者はドラム打撃時刻にドラム奏者の内部振動子の位相が 0 であることを知る。

これらの仮定の下では、結合振動子によってロボットの演奏のオンセットを人の演奏に引き込むことで打撃時刻予測が可能になる。ただし、振動子モデルでは自然周期は定数と仮定しているが、人のドラム演奏速度は揺らぐのでその仮定は成立しない。そこで、蔵本モデルに自然周期の更新式を追加する。

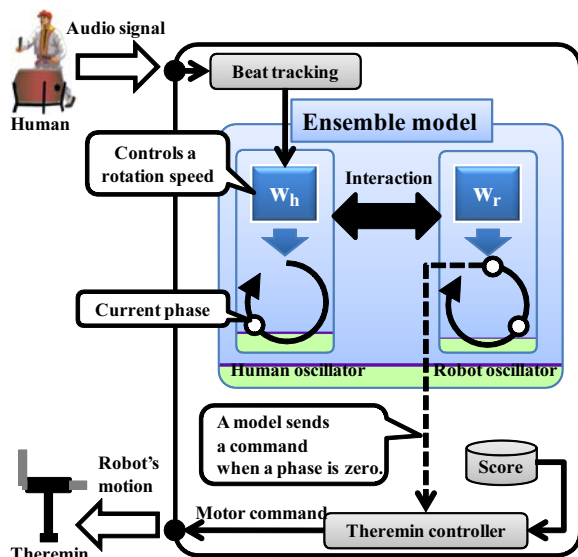


Fig.3 合奏システムのブロック図

本手法のアルゴリズムまとめると次のようになる。

2つの振動子の位相方程式:

$$\frac{d\phi_h}{dt} = \omega_h + K_h \sin(\phi_r - \phi_h) \quad (9)$$

$$\frac{d\phi_r}{dt} = \omega_r + K_r \sin(\phi_h - \phi_r). \quad (10)$$

ロボットの自然周期  $\omega_r$  の更新式:

$$\omega_r \leftarrow \omega_r + \mu(\omega_r - \omega_h), \quad (11)$$

ただし、 $\phi_r$  と  $\phi_h$  はロボットと人の位相、 $K_r$  と  $K_h$  はロボットと人の結合強度、 $\mu$  は学習係数を表す。

### 3.3 合奏システムへの実装

図3に本手法を用いた合奏システムの概要図を示す。本システムは、(1)リアルタイムビート予測、(2)テルミンの特性モデルに基づくロボット制御、(3)打撃時刻予測(Ensemble model)から構成される。本稿では概要のみを示すので、詳細は[3]を参照。

本システムは以下のように動作する。まずロボットが人の打撃音をロボット自身のマイクで収録する。その音から(1)によって、ビートが得られたタイミングでビート間隔を取得する。(1)から推定結果を得た時点で  $\omega_h$  を新しいビート間隔で書き直し、 $\phi_h$  を0にする。ロボットは位相  $\omega_r, \omega_h$  を更新しながら、 $\omega_r$  が0のとき次の音符を演奏するように動作制御を行う。

## 4. 実験

次の3条件におけるテルミンとドラムの合奏を行い、各オンセット時刻を比較することで本手法の予測精度を検証する。(1)相互作用も揺らぎも無い“完全な”ドラム奏者を想定した、メトロノームとのテルミンの合奏。(2)相互作用の無いドラム奏者を想定した、人工的に揺らぎを加えたメトロノームとの合奏。(3)人のドラム演奏とテルミンの合奏。比較対象には大塚らのビート間隔のみを



Fig.4 Aura Lee の楽譜

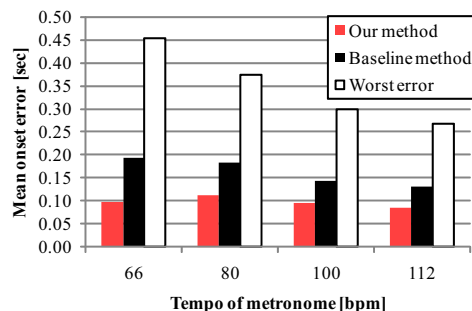


Fig.5 実験1: メトロノームとの合奏でのオンセット誤差

使ったドラムとテルミンの合奏システム[5]を用いる。

### 4.1 実験条件

ロボットには川田工業のHRP-2を、テルミンにはMoog Music社のEtherwave Thereminを用い、50cmの間隔を開けて配置した。曲にはアメリカの民謡“Aura Lee”を使用した(楽譜は図4)。振動子モデルのパラメータは、実験的に以下のように定めた:  $K_r = 0.4$ ,  $\omega_h = \omega_r = 2\pi/700$ ,  $\mu = 0.01$ 。位相と自然周期は50msec間隔で更新する。人がロボットに適応する強度を表すパラメータ  $K_h$  は、実験1,2ではメトロノームなので0、実験3ではロボットと同じ  $K_h = 0.4$  とした。

メトロノームのテンポは、ビートトラックが動作するテンポの範囲(60-120bpm)の全域を調べるため、次の4種類: 66, 80, 100, 112bpmを用いた。各テンポで3回ずつ試行を行い、テルミンとドラムのオンセット誤差を以下の式で評価した。

$$error = \frac{1}{N} \sum_{j=1}^N \min_{i=1, \dots, M} |onset_t(i) - onset_d(j)|, \quad (12)$$

ただし、 $N$  はドラムのオンセット数、 $M$  はテルミンのオンセット数、 $onset_d(j)$  と  $onset_t(i)$  はそれぞれドラムの  $j$  番目とテルミンの  $i$  番目のオンセット時刻を表す。

### 4.2 実験1: メトロノームとの合奏

図5に実験結果を示す。横軸はメトロノームのテンポを、縦軸はオンセット誤差を表す。赤いグラフは本手法の誤差を、黒いグラフはベースラインの誤差を、白いグラフは誤差の最大値、すなわちビート間隔の1/2を表す。誤差の最大値がテンポが上がると減少するのは、テンポが早くなるに従ってビート間隔が縮まるからである。結果より、全条件で本手法がオンセット誤差を改善した。また、平均で39%オンセット誤差を削減した。

加えて、誤差の下限として、ロボットの代わりに人がメトロノームに合わせてドラムを打った場合のオンセット誤差も求めた。その結果、平均誤差は1msecであった。

### 4.3 実験2: ゆらぎ付きメトロノームとの合奏

次に、メトロノームの代わりに、平均がビート間隔、標準偏差が平均の10%の正規分布に従う人工的な揺らぎ付きメトロノームとロボットとの合奏を行った。

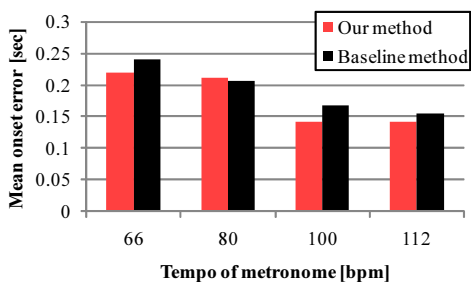


Fig.6 実験 2: 揺らぎ付きメトロノームとのオンセット誤差

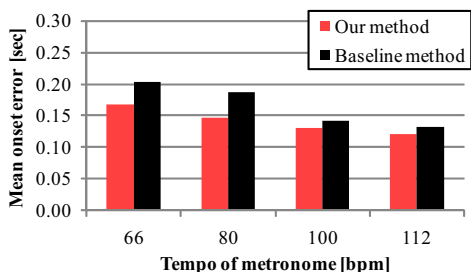


Fig.7 実験 3: 人とのオンセット誤差

図 6 に結果を示す。縦軸と横軸は図 5 と同様である。メトロノームに揺らぎが加えられた場合、80 bpm の場合を除いて本手法が誤差を改善しているものの、ベースラインとの差は減少している。これは、実験条件で自然周期の学習係数  $\mu$  を比較的大きくしたために揺らぎに敏感になったのが理由だと考えられる。 $\mu$  を小さくすると誤差に頑健になる一方でテンポ変化への追従性能は低下するので、トレードオフが存在する。

実験 1 と同様に人で実験を行った結果、全テンポの平均誤差は 120 msec であった。

#### 4.4 実験 3: 人のドラムとの合奏

最後に、本稿の目標である人とロボットの合奏を行った。実験 1, 2 と同様に初期テンポを与えるため、各試行前にドラム奏者にメトロノームを聞かせておき、ドラム演奏を開始するとそれを取り除くという操作を行った。

図 7 に結果を示す。縦軸と横軸は図 5 と同様である。本手法はベースラインに対して、全条件でオンセット誤差を改善しており、平均して 14% 削減した。誤差が実験 2 よりも小さいことから、本手法による予測が成功している。このことは、演奏者のテンポに相互引き込みが存在する可能性を示唆している。また、テンポが小さいときに誤差の削減量が多い。具体的には、テンポが 100, 112 bpm の場合の削減量が 8% であるのに対して、66, 80 bpm の場合の削減量は 20% である。これは、本手法が特にビート間隔が長い場合に有効である事を表している。

#### 4.5 考察

ロボットの代わりに人を用いた合奏のオンセット誤差について議論する。実験 1 では、誤差の平均は 1 msec であった。ロボットの平均誤差が 97 msec だった事と比較すると、人と同程度の同期精度を達成するにはまだ予測手法の改善の余地は残されている。ただし、ロボットは旋律楽器を用いているので、単に予測精度を向上するだけでなく、演奏表情付けによるアプローチも有効であろう。例えば、音高の場合なら、ピブラートを付加することで、人が感じる誤差の許容範囲を広げることができる。

一方、実験 2 における人の結果は平均 120 msec であ

る。この結果は、本手法とベースラインよりは良いものの、実験 1 と比較して 120 倍と、大幅に誤差が増加している。この誤差の増加はタスク自体の難しさを表している。実験 2 が困難なタスクだということは、人のリズム知覚が振動子で表現できると仮定すれば説明できる。実験 2 の刺激に用いた比較的大きな揺らぎのあるリズムは、単純な振動子では表現できない。そのため、知覚が困難、予測の誤差が増加したと考えられる。

## 5. 結論

本稿では、人のドラム奏者との合奏のための結合振動子モデルに基づく打撃時刻予測手法について報告した。結合振動子には蔵本モデルを用い、自然周期の更新則を加えることで人の演奏速度の変化への追従も可能にした。実験の結果、本予測手法はビート間隔のみを用いる手法に比べてオンセット誤差を最大 39% 削減できることが明らかになった。また、本稿では検証しなかったが、例えばアトラクタが  $\pi$  になるように係数を定めれば裏拍に合わせて合奏させることも可能である。

今後の課題は、まず人同士の合奏を解析し、本モデルがどの程度合奏をモデル化できているか調べる。このとき、オンセット誤差に加えて、ストロボスコープ法などの非線形時系列解析手法で同期の詳細な状態や適応速度などの解析も可能となる。また、スパイク状の波形を出力する弛緩振動子などの、より楽器に特化した振動子モデルを用いた同期精度の向上も行う予定である。

謝辞 振動子モデルに有益な助言を頂いた京都大学理学研究科の合原一究氏に感謝する。本研究の一部は、科研費 (No. 19100003, 22118502) と GCOE の援助を受けた。

#### 参考文献

- [1] G. Weinberg *et al.* The creation of a multi-human, multi-robot interactive jam session. *NIME*, pp. 70–73, 2009.
- [2] K. Petersen *et al.* Development of a aural real-time rhythmic and harmonic tracking to enable the musical interaction with the waseda flutist robot. *IROS*, pp. 2303–2308, 2009.
- [3] T. Mizumoto *et al.* Human-robot ensemble between robot thereminist and human percussionist using coupled oscillator model. *IROS*, 2010. *to appear*.
- [4] J. Solis *et al.* Development of Waseda flutist robot WF-4RIV: Implementation of auditory feedback system. *ICRA*, pp. 3654–3659, 2008.
- [5] T. Otsuka *et al.* Music-ensemble robot that is capable of playing the theremin while listening to the accompanied music. *IEA/AIE*, pp. 102–112, 2010.
- [6] S. H. Strogatz. *SYNC: The Emerging Science of Spontaneous Order*. Hyperion, 2003.
- [7] E. W. Large and M. R. Jones. The dynamics of attending: How people track time-varying events. *Psychological Review*, Vol. 106, No. 1, pp. 119–159, 1999.
- [8] R. B. Dannenberg. An on-line algorithm for real-time accompaniment. *ICMC*, pp. 193–198, 1984.
- [9] J. Mac Ritchie *et al.* Visualizing musical structure through performance gesture. *ISMIR*, pp. 237–242, 2009.
- [10] Y. Kuramoto. *Chemical Oscillations, Waves, and Turbulence*. Dover Publications, 2003.