人とロボットの合奏のための多人数合奏の主導権推定

○水本武志 †, 尾形哲也 ‡, 奥乃博 †
† 京都大学大学院情報学研究科, ‡ 早稲田大学基幹理工学部表現工学科

1. はじめに

複数の人々が一緒に演奏する**合奏**は, さまざまな文化で広く行われている活動である. 合奏の規模も 2 人のデュオ合奏から数十人のオーケストラまで多様に存在する. いずれの形態においても, 合奏の共演者たちは互いの演奏を同期して演奏している. 本研究では, 人間と同じように多人数と同期して合奏できるロボット (合奏ロボット) の実現を目的とする.

多人数の合奏におけるリーダーには2つの特徴がある.

1. **リーダーは複数存在する.** 楽譜中で複数の主メロディある場合や, ベースとドラムが互いにタイミングを合わせている場合などに, 複数の共演者が全体のリズムを制御する.
2. **リーダーは時間遷移する.** 主メロディが別パートに移ったときや, リーダーが演奏ミスをしたときなどに, 別の共演者が代わりにリズムを制御する.

したがって, 多人数合奏の共演者たちは, 演奏中に (1) リズムを合わせるリーダーの選択と, (2) リーダーの変化の検出を行う必要がある.

近年さかんに開発されている演奏を聞きながら人と合奏するロボットでは, リーダーを事前に定めたり, リーダーの人数を 1 人と仮定したりして本問題を避けていた. 例えば, ロボットとドラム奏者 [1], フルート奏者 [2], サックス奏者 [3] との合奏の研究は人がリーダーであると仮定している. また, Weinberg らはリーダーが遷移する 2 人-2 ロボットの即興合奏を実現している [4] が, リーダーは一人と仮定している.

そこで本稿では, このような時間変化する複数のリーダーをもつ合奏 (図 1) における各共演者の主導権推定手法とそれに基づくオンセット推定手法を開発する. キーアイデアは, 主導権の定量指標として**リーダー度**を定義することである. リーダー度に結合振動子系によるリズム予測モデル [1] を組み合わせ, 多人数合奏におけるリズムの状態空間モデルを構築する. そして, 本モデルに非線形システムが予測できる Unscented Kalman Filter (UKF) [5] を適用することで多人数合奏における主導権の遷移とオンセット時刻の予測を実現する.

2. 合奏の心理モデル

合奏は, 前節で述べた合奏ロボットに関する研究だけでなく, 様々な分野で研究されている. 例えば神経科学では時間制約の強い感覚運動同期タスクとして広く研究されており [6], 心理学では, 人がメトロノームに合わせて指を叩くタッピング課題を用いて, 人のリズム知覚が非線形振動子を用いてモデル化されている [7,8].

なかでも Keller は, 同期演奏する共演者の心理プロセスを次の 3 つの認知過程, Anticipatory Auditory Imagery (AAI, 予測聴覚像), Prioritized Integrative Attention (PIA, 優先度付き統合的注意), Timing Adaptation (TA, タイミング適応) で説明している [9]. ピアニストによる実験から, 各認知過程の能力と演奏中における体の動作の調和の関連を示唆している [9]. 本稿で設計する合奏の状態空間モデルは Keller のモデルと関連が深いので, まずその構成要素について述べる. 具体的な関連はモデルを定義した後の 3·6 節で議論する.

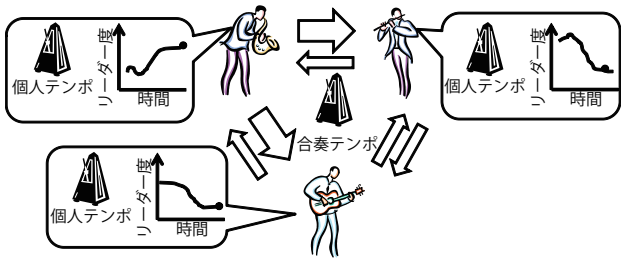

図 1 多人数合奏における主導権

AAI は共演者のもつ音楽の心的イメージで, 全員で共有される合奏の目標となる. Auditory Imagery は楽譜から音をイメージする能力に相当する. 共演者は AAI による他共演者の予測によって演奏動作を生成し, 同期ズレを用いて自らの AAI を修正する.

PIA は誰により注意を向けるかを決める認知過程である. 合奏では自分の演奏に加えて他者との関係の保持も重要なので, 他共演者へ注意を向けることは必須である. 容易に考えられるのは他者のうち注意を向ける一人を選ぶ戦略か, 全員に平等に注意を向ける戦略である. Keller は, 優先順位によってこれらを統合し, その優先順位を演奏に合わせて切り替える PIA が最適であると主張している [10].

TA はリズムを保持する認知過程であり, 次の 2 つの方法で共演者と合うように修正される. 一方タイミングを修正する位相更新で, 常に行われる. もう一方がテンポ更新で, こちらは明確な変化時にのみ行われる.

これらは次のように連携して働く. AAI は全員で共有される合奏の目標イメージ, TA はより詳細なテンポやオンセット時刻を表し, それは他共演者の演奏によって修正される. 誰にどれだけ影響を受けるかは PIA による優先順位によって決定される.

3. 多人数合奏の状態空間モデル

まず問題設定を述べ, 次にリーダー度を設計する. そして合奏の状態空間モデルの設計とその推定手法について述べる. 詳細は [11] を, 記号の定義は表 1 を参照.

3·1 問題設定

入力: 各共演者の直前のオンセット時刻とテンポ
出力: 各共演者の次の時刻における
リーダー度, オンセット時刻, テンポ

表 1 記号の定義

$t, \Delta t$	時間, 時間間隔
N	共演者数
M	記憶する過去のオンセット数
i	共演者インデックス ($i \in \{1, ..., N\}$)
$\omega_s(t)$	集約した合奏全体のテンポ
$\omega_i(t)$	第 i 共演者のテンポ
$\theta_i(t)$	第 i 共演者の位相 (2π の倍数の時刻で発音)
$l_i(t)$	第 i 共演者のリーダー度
$\mathbf{x}(t)$	状態ベクトル, ($\mathbf{x}(t) \in \mathbf{R}^{1+N(M+2)}$)
$\mathbf{z}(t)$	観測ベクトル, ($\mathbf{z}(t) \in \mathbf{R}^{2N}$)

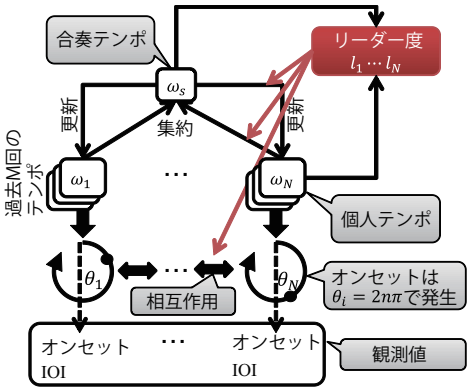

図 2 データの流れ

仮定: (1) 全参加者は四分音符の位置で演奏
(2) 互いにオンセット時刻を合わせる

本稿では音声を用いたリズムの同期に着目するので, 入力は, 各共演者のオンセット時刻とテンポとする. これらはビートトラッキング [12] を用いれば得られる. 出力は, 次のオンセット時刻とテンポ, それに加えて主導権の定量指標であるリーダー度とする (3·2 節参照). 仮定 1 はリズム構造の問題を避けるために定めた. ただし, 結合振動子系によるリズム構造表現 [13] を用いればこの仮定の緩和は可能である. 仮定 2 は共演者全員が協力的な合奏であれば自明である.

3·2 リーダー度の設計: 主導権の定量化

本稿では, リーダーはテンポ変化時にのみ現れ, テンポ安定時はいないと仮定する. なぜなら, テンポが安定して全員が同期していれば, 相互作用や競合は無くリーダーは不要だからである. 一方, テンポ変化時は共演者間のテンポ不一致による相互作用や競合が生じるため, リーダーとなって全体を収束させる必要がある.

リーダーのもつ主導権の定量指標として**リーダー度**を定義する. リーダー度の総和は一定で, テンポ安定時はリーダー度は各共演者に等分配され, 変化時はリーダーに多く分配されるとする. そのために, 時刻 t の第 i 共演者のリーダー度 $l_i(t)$ の総和を 1 とする. リーダー度を状態遷移の学習係数に用いて高いリーダー度をもつ共演者との誤差をより重視する設計を行うことで, 共演者が受ける影響力をリーダー度で制御できる.

具体的なリーダー度の定義は合奏テンポからの差とテンポの安定性との積とする. 前者は共演者のテンポ変化の意志を表すためである. 後者はテンポ変化のみではテンポの安定しない信頼性できない共演者がリーダーになりうるからである. 言い換えると, 「現在のテンポを変化させる強い意志のある共演者が高いリーダー度を持つ」という設計である.

合奏テンポからの差 $p_i(t)$ は, 値域が $[0, 1]$ となるように絶対誤差の sigmoid に似た関数で定義する.

$$p_i(t) = \frac{2}{3}\frac{1}{1+\exp\left(-|\omega_i(t)-\omega_s(t)|\right)} - \frac{1}{2}. \quad (1)$$

テンポの安定性 $s_i(t)$ は, 過去 M 回のテンポの標準偏差の指数関数で定義する.

$$s_i(t) = \exp\left(-Std[\omega_i(t), \cdots \omega_i(t-M-1)]\right) \quad (2)$$

ただし, $Std[x_1, ..., x_N]$ は標準偏差を表す. この定義も, 値域が $[0, 1]$ となるように定めた. もし過去の M 回のテンポが完全に同一なら, $s_i(t)$ は最大値 1 をとる.

リーダー度 $l_i(t)$ はこれらの積なので

$$l_i(t+1) = s_i(t)p_i(t). \quad (3)$$

ただし, $l_i(t+1)$ は総和が 1 となるように正規化する.

3·3 状態空間モデルの設計：状態遷移モデル

状態遷移モデルは結合振動子のテンポと位相, 前節で設計したリーダー度の更新式から構成する. したがって, 状態ベクトルは合奏全体のテンポが 1 次元, 各共演者のテンポが MN 次元 (リーダー度計算用の過去テンポを含む), 位相が N 次元, リーダー度が N 次元, 合計 $N(M+2)+1$ 次元である.

テンポは, まず合奏全体のテンポ $\omega_s(t)$ をリーダー度の高い共演者のテンポと近い値となるようにリーダー度による重み付き和で定義しする:

$$\omega_s(t+1) = \sum_{i=1}^{M} l_i(t)\omega_i(t). \quad (4)$$

そして, 各共演者のテンポは $\omega_s(t)$ に漸近するよう更新式を設計する.

$$\omega_i(t+1) = \omega_i(t) + (1-l_i(t))(\omega_s(t)-\omega_i(t)). \quad (5)$$

位相 $\theta_i(t)$ は, 結合振動子モデル [1] に従って更新式を定める. ただし, 高いリーダー度を持つ共演者の, 他共演者への強い影響力を表すため, [1] では定数だった結合強度をリーダー度に置き換える.

$$\theta_i(t+1) = \theta_i(t) + \omega_i(t)\Delta t + \sum_{j=1}^{N} l_j(t)\sin(\theta_j(t)-\theta_i(t)) \quad (6)$$

リーダー度 $l_i(t)$ は更新式は式 (3) を用いる.

3·4 状態空間モデルの設計：観測モデル

共演者は他者の位相とテンポを観測できるので, 次元は $2N$ である. ただし, 観測値はオンセット時刻でしか得られない. そこで, 第 i 共演者のオンセット時刻の観測値を $\theta_i(t)$ が 0, $\omega_i(t)$ を前回のオンセット時刻との差とし, それ以外では $\theta_i(t) = \omega_i(t) = \emptyset$ とする.

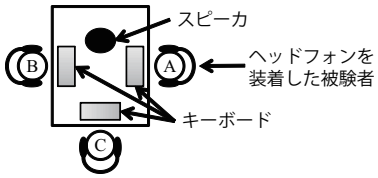

図 3 被験者と実験装置の配置

3·5 状態推定とロボット制御

本モデルは非線形な状態空間モデルなので, 状態ベクトルを UKF [5] で推定する. ただし, 観測ベクトルが $\emptyset$ の場合は観測値を用いた状態ベクトル推定値の更新をしない. 状態空間モデルの設計ではノイズに関する議論はしていないが, UKF 適用時に加法性のガウスノイズを考慮する.

状態ベクトルが推定できれば, ロボット (第 j 共演者とする) は $\omega_j(t)$ の値が $2n\pi$ の倍数となる時刻にオンセットを演奏するように制御すれば他の共演者と同期した合奏が可能となる.

3·6 Keller のモデルとの関係

Keller のモデルを定量的に定義し状態空間モデルで再構成したものが本モデルである. 具体的な対応関係は以下のとおりである. AAI は 共演者のテンポ $\omega_i(t)$ と合奏テンポ $\omega_s(t)$ に対応する. $\omega_i(t)$ は他共演者の予測として, $\omega_s(t)$ は合奏全体の目標値として用いられるからである. PIA はリーダー度 $l_i(t)$ による重み付けに対応している. なぜなら, リーダー度を優先度とみなせば, 連続的な重み付けによる更新 (式 (4)(5)(6)) は優先度付きの統合注意そのものだからである. TA は本モデルの結合振動子系に対応する更新式, 位相修正の式 (6) とテンポ修正の式 (6) に対応している.

4. 評価実験

リーダー度定義の妥当性と本モデルのオンセット予測性能を検証するため, 複数人によるタッピング実験を行い, 本手法で解析する.

4·1 被験者実験によるデータ収集

まず被験者 (男性 8 名, 女性 1 名, 合計 9 名. 21 歳から 38 歳まで) から復元抽出でランダムに 2 人組を 4 組, 3 人組を 3 組構成する. 被験者は, 各組ごとに 図 3 のように席につき, ヘッドフォンとキーボードが与えられる. 以降は, 被験者を座席ごとに A, B, C と呼ぶ (2 人組は A,B のみ). ヘッドフォンへはそれぞれ異なる刺激が与えられる. キーボードのキーを押すとそれぞれ異なる純音がスピーカから再生される (A, B, C の音高はそれぞれ 880, 440, 220Hz).

被験者への刺激は準備準備, 本刺激から構成される (表 2). 準備刺激では被験者ごとに異なる初期テンポが与えられ, 5 秒の無音の後に開始キューが全員同時に与えられる. 本刺激では順番にメトロノームによるテンポ指示が与えられ, 終了キューで刺激が終わる.

被験者に与えられる指示は次の 3 つである.

1. 初期テンポを覚え, 開始キュー直後にそのテンポでタッピングを開始する. これは被験者の同期能力の有無を調べるためである.
2. 目を閉じ, 音だけを聞いて他の被験者にキーを押すタイミングを合わせる. これは本稿では音にのみ着目しているからである.
3. メトロノーム音が聞こえたら, 周囲の音は無視してそのテンポに合わせる. これによって, リーダーは順番に遷移すると期待できる.

表 2 被験者への刺激: 上は二人組 (A, B), 下が三人組 (A, B, C) で使用した刺激. s5, s25 はそれぞれ 5, 25 秒の無音, cue は 100msec の 880Hz 純音を表す. 50, 60, 80 はメトロノームのテンポ (bpm) を表す.

	準備刺激			本刺激				
A	60	s5	cue	s25	s25	80	s25	cue
B	80	s5	cue	s25	50	s25	s25	cue

	準備刺激			本刺激					
A	50	s5	cue	s25	80	s25	s25	s25	cue
B	60	s5	cue	s25	s25	50	s25	s25	cue
C	80	s5	cue	s25	s25	s25	80	s25	cue

以上の実験を各被験者ごとに練習 1 試行, 本番 3 試行の合計 4 試行行い, キーの打撃時刻を記録した.

4·2 リーダー度定義の妥当性評価

まず, リーダー度定義の妥当性を評価するためデータの解析を行った. 結果を図 4 に示す. 個人差を排除するため, A, B, C ごとに平均値を求め, さらに窓幅 5 の移動平均によって平滑化して表示している.

まず IOI 軌跡 (図 4 上段) について議論する. 初期テンポは異なるので時間 0 における IOI は異なるが, すぐに軌跡が重なる. したがって, 被験者は相手とタイミングを同期する能力を持っていることが確認できる. 次に, テンポが与えられると, まずその被験者が指示どおりにテンポを修正し, 残りの被験者が追従している. (例えば図 4 左側 50 秒付近では, まず A が追従している.) したがって, 被験者はヘッドフォンによる指示に従っていることが確認できる.

次に, IOI の標準偏差 (図 4 下段) について議論する. テンポ指示の変化直後は被験者間の同期が崩れるために標準偏差が増加し, 時間が経つに連れて同期に成功するために減少していく. 左右の標準偏差を比べると 3 人組の方が標準偏差の減少が遅い. これは, 3 人組の方がより困難なタスクであることが原因である.

最後に, リーダー度の軌跡 (図 4 中段) について議論する. まず, 左右両方について両端の指示なし区間ではリーダー度の差が無くなっている. これは, 主導権をもつ被験者がおらず, 全員が互いに合わせていることを示す. 一方, 指示のある区間では, リーダー度に偏りが生じている. まず指示を受けた被験者のリーダー度が上がり, 時間が経つにつれてその偏りがなくなっていく. これは, 指示が与えられた直後はその被験者は現在のテンポ変化を牽引するので主導権の偏りが生じるが, 他の被験者が同期できれば牽引の必要はなくなるからである. このことから, リーダー度は設計の意図どおり, テンポの変化を牽引する時にのみ生じる主導権を定量的に捉えることができたといえる.

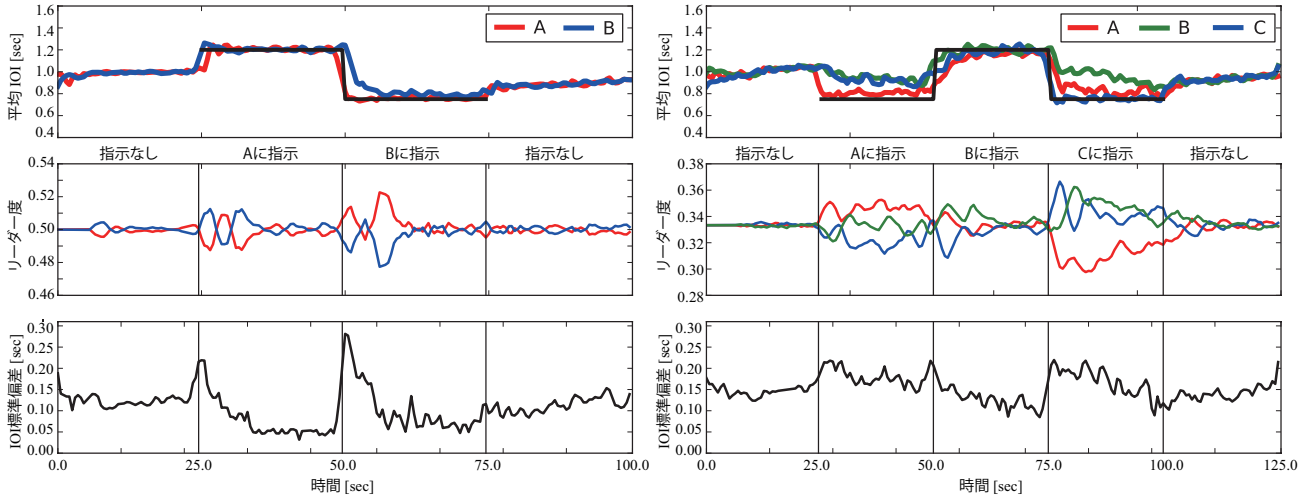

図 4 実験結果: 左が 2 人組, 右が 3 人組の結果. 両図の横軸は時間, 縦軸は上から順に, IOI (Inter-Onset Interval) の平均値, リーダー度, IOI の分散を表す. ただし時間は本刺激開始時刻を 0 としている. 上段の赤, 緑, 青の実線が被験者の IOI 軌跡の平均値, 黒の実線が刺激の IOI を表す. 中, 下段の縦線は刺激の変化時刻を表す.

表 3 オンセット誤差: 単位はすべて msec

	2 人組		3 人組	
	中央値	標準偏差	中央値	標準偏差
被験者	135	68	239	61
本手法	181	154	241	161

4·3 オンセット時刻予測性能評価

被験者実験得た被験者ごとのオンセット時刻の時系列を逐次的に UKF への入力として与えてテンポとオンセット時刻を予測する. 時間ステップ Δt は 5 msec, UKF の状態推定のパラメータである観測誤差相関行列と状態遷移誤差相関行列はそれぞれ対角成分が 0.05 の対角行列, 状態ベクトルの初期値は, テンポ $\omega_i(t)$ は刺激から求めた値, 位相は 0, リーダー度は $1/N$, 保持する過去テンポ数 M は 5 とした.

評価尺度はオンセット時刻の絶対誤差の中央値とし, 被験者と予測値, 被験者同士ををそれぞれ評価した. 中央値は, 誤差の外れ値の影響を軽減するために用いた. 評価結果を表 3 に示す. 被験者, 本手法のどちらも, 3 人組の方が誤差が大きい. これは人数が多い方がより困難なタスクであったことを示している. また, 2 人組, 3 人組共に誤差の中央値は被験者と似た値である. これより, 本手法は人と同等の予測性能を実現できたといえる. ただし, 標準偏差は本手法の方が大きいので, 予測精度の安定性は人の精度には到達していない.

5. おわりに

合奏ロボットのための主導権を用いた多人数合奏モデルについて報告した. キーアイデアは主導権を定量化したリーダー度を定義することであった. 我々はリーダー度と結合振動子系で非線形状態空間モデルを構築し, UKF でオンセット時刻とリーダー度遷移を予測した. 実験の結果, リーダー度定義の妥当性, 人と同等の予測精度を確認した. 今後は, ロボットへの実装, 視聴覚統合, 複雑なリズム構造への拡張を行う予定である.

謝辞 状態空間モデル構築に関する助言を頂いた中村佳祐氏に感謝する. 本研究はホンダ・リサーチ・インスティチュート・ジャパン, 科研費 (S, 新学術領域研究, 特別研究員奨励費) の支援を受けた.

参考文献

[1] T. Mizumoto *et al.* "Human-robot ensemble between robot thereminist and human percussionist using coupled oscillator model", in *IROS*, 2010, pp. 1957–1963.
[2] A. Lim *et al.* "Robot musical accompaniment: Integrating audio and visual cues for real-time synchronization with a human flutist", in *IROS*, 2010, pp. 1964–1969.
[3] K. Petersen *et al.* "Development of a aural real-time rhythmical and harmonic tracking to enable the musical interaction with the waseda flutist robot", in *IROS*, 2009, pp. 2303–2308.
[4] G. Weinberg *et al.* "The creation of a multi-human, multi-robot interactive jam session", in *NIME*, 2009, pp. 70–73.
[5] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation", *IEEE*, vol. 92, no. 3, pp. 401–422, 2004.
[6] R. J. Zatorre *et al.* "When the brain plays music: auditory-motor interactions in music perception and production", *Nature Rev Neurosci*, vol. 8, pp. 547–558, 2007.
[7] H. Haken *et al.* "A theoretical model of phase transitions in human hand movements", *Biological Cybernetics*, vol. 51, pp. 347–356, 1985.
[8] E. W. Large and M. R. Jones, "The dynamics of attending: How people track time-varying events", *Psychol Rev*, vol. 106, no. 1, pp. 119–159, 1999.
[9] P. E. Keller, "Joint action in music performance", in *Enacting Intersubjectivity: A Cognitive and Social Perspective on the Study of Interactions*, pp. 205–221. IOS Press, Amsterdam, 2008.
[10] P. E. Keller, "Attentional resource allocation in musical ensemble performance", *Psychology of Music*, vol. 29, pp. 20–38, 2001.
[11] T.Mizumoto *et al.* "Who is the leader in a multiperson ensemble? – multiperson human-robot ensemble model with leaderness–", in *IROS*, 2012, *to appear*.
[12] K. Murata *et al.* "A robot uses its own microphone to synchronize its steps to musical beats while scatting and singing", in *IROS*, 2008, pp. 2459–2464.
[13] E. W. Large and C. Palmer, "Perceiving temporal regularity in music", *Cognitive Sceince*, vol. 26, pp. 1–37, 2002.