Vol.2016-MUS-112 No.19
2016/8/1

自動合奏のための演奏タイミング結合モデル

前澤 陽[1,a)]

概要：本稿では，人間の演奏に合わせて機械が伴奏するような合奏システムにおける，伴奏パートの発音タイミングについて述べる．合奏中の演奏者は，(1) 自分自身のパートに対して，目標となる演奏タイミングを持ち，(2) 相手の演奏を予測しながら聴き，(3) 互いの主従関係を踏まえながら演奏タイミングを補正すると考えられる．また，これらの三要素は，楽曲中のコンテキストやリハーサルや対話を通じて，独立して学習もしくは制御されると考えられる．そこで本研究では，演奏者の演奏タイミング予測，伴奏パート自体の生成過程，演奏者・伴奏パート間の主従関係の 3 要素を，独立に学習もしくは制御が可能な，発音タイミングの数理モデルを提案する．

Coupled Timing Model for Automatic Music Accompaniment

Akira Maezawa[1,a)]

1. はじめに

自動合奏システムとは，人間の演奏に対し，機械が合わせて伴奏を生成するシステムである．本稿では，クラシック音楽のように，合奏システムと人間それぞれが弾くべき楽譜表現が与えられている合奏システムについて論じる．このような自動合奏システムは，音楽演奏の練習支援や，演奏者に合わせてエレクトロニクスを駆動するような音楽の拡張表現など，幅広い応用がある．なお，以下では，合奏エンジンが演奏するパートのことを「伴奏パート」と呼ぶ．

音楽的に整合した合奏を行うためには，伴奏パートの演奏タイミングを適切に制御することが必要である．適切なタイミング制御には，以下 4 つの要求がある：

要求 1　原則として，自動合奏システムは，人間の奏者が弾いている場所を弾く必要がある．したがって，自動合奏システムは，再生する楽曲の位置を，人間の演奏者に合わせる必要がある．特にクラシック音楽では，演奏速度 (テンポ) の抑揚が音楽表現上重要であるため，演奏者のテンポ変化を追従する必要がある．また，より精度が高い追従を行うために，演奏者の練習 (リハーサル) を解析することで，演奏者のクセを獲得することが好ましい．

要求 2　合奏システムは，音楽的に整合した演奏を生成すること．つまり，伴奏パートの音楽性が保たれる範囲内で人間の演奏を追従する必要がある．

要求 3　楽曲のコンテキストに応じて，伴奏パートが演奏者に合わせる度合い（主従関係）を変えることが可能であること．楽曲中には，音楽性を多少損なってでも人に合わせるべき場所や，追従性を損なっても伴奏パートの音楽性を保持すべき場所がある．従って，要件 1 と要件 2 でそれぞれ述べた「追従性」と「音楽性」のバランスは楽曲のコンテキストにより変わる．たとえば，リズムが不明瞭なパートは，リズムをよりはっきり刻むパートに追従する傾向がある．

要求 4　演奏者の指示によって，即座に主従関係を変えることが可能であること．追従性と合奏システムの音楽性のトレードオフは，リハーサル中に人間同士が対話を通じて調整することが多い．また，このような調整を行った場合，調整を行った箇所を弾き直すことで，調整結果を確認する．したがって，リハーサル中に追従性の挙動を設定できる合奏システムが必要である．

これらの要求を同時に満たすためには，演奏者が演奏している位置を追従した上で，音楽的に破綻しないように伴

[1] ヤマハ株式会社
Yamaha Corporation, Iwata, Shizuoka 438–0942, Japan
[a)] akira.maezawa@music.yamaha.com

奏パートを生成する必要がある．これらを実現するためには，合奏システムは，(1) 演奏者の位置を予測するモデル，(2) 音楽的な伴奏パートを生成するためのタイミング生成モデル，(3) 主従関係を踏まえ，演奏タイミングを補正するモデル，の三要素が必要となる．また，これらの要素は独立して操作もしくは学習できる必要がある．しかし，従来はこれらの要素を独立に扱うことが難しかった．

そこで，本稿では，(1) 演奏者の演奏タイミング生成過程，(2) 合奏システムが音楽的に演奏できる範囲を表現した演奏タイミング生成過程，(3) 合奏システムが主従関係を持ちながら演奏者に合わせるための，合奏システムと演奏者の演奏タイミングを結合する過程，これら三要素を独立にモデル化し，統合することを考える．独立に表現することにより，個々の要素を独立に学習したり，操作することが可能になる．システム使用時には，演奏者のタイミング生成過程を推論しながら，合奏システムが演奏できるタイミングの範囲を推論し，合奏と演奏者のタイミングを協調させるように伴奏パートを再生する．これにより，合奏システムは音楽的に破綻しない合奏を，人間に合わせながら演奏することが可能になる．

2. 関連研究

従来の合奏システムでは，楽譜追従を用いることで演奏者の演奏タイミングを推定する [1]．その上で，合奏エンジンと人間を協調させるため，大きく分けて二つのアプローチが用いられる．

第一に，多数のリハーサルを通じて演奏者と合奏エンジンの演奏タイミングに対する関係性を回帰することで，楽曲における平均的な挙動 [14]，もしくは時々刻々と変化する挙動 [15]，を獲得することが提案されている．このようなアプローチでは，合奏の結果自体を回帰するため，結果的に伴奏パートの音楽性と，伴奏パートの追従性を同時に獲得できる．一方，演奏者のタイミング予測，合奏エンジンの生成過程と，合わせる度合いを切り分けて表現することが難しいため，リハーサル中に追従性や音楽性を独立に操作することは難しいと考えらえる．また，音楽的な追従性を獲得するためには，人間同士の合奏データを別途解析する必要があるため，コンテンツ整備にコストがかかる．

第二に，少ないパラメータで記述される動的システムを用いることでテンポ軌跡に対して制約を設けるアプローチがある．このアプローチでは，テンポの連続性といった事前情報を設けた上で，リハーサルを通じて演奏者のテンポ軌跡などを学習する．また，伴奏パートは伴奏パートの発音タイミングを別途学習できる．これらは少ないパラメータでテンポ軌跡を記述するため，リハーサル中に伴奏パートや人間の「癖」を容易に手動で上書きできる．しかし，追従性を独立に操作することは難しく，追従性は演奏者と合奏エンジンそれぞれが独立に演奏した時における発音タイミングのばらつきから間接的 [12] に得られていた．

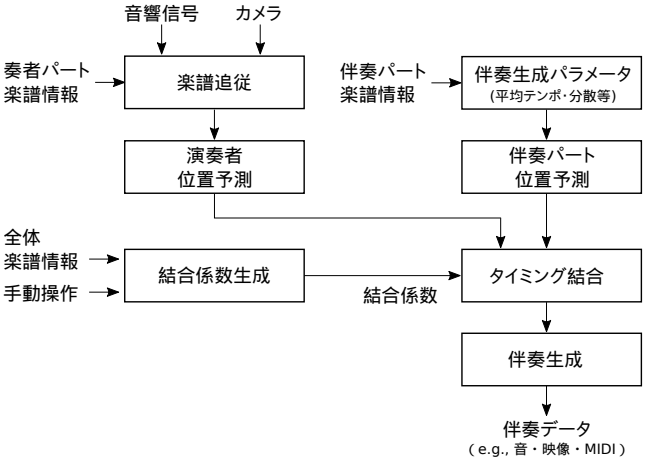

図 1 本システムの構成.

リハーサル中における瞬発力を高めるためには，合奏システムによる学習と，合奏システムと演奏者との対話を交互に行うこと [18] が有効と考えられる．そこで，追従性を独立に操作するため，合奏再生ロジック自体を調整する方法が提案されている [2]．本手法では，このようなアイディアに基づき，「合わせ方」「伴奏パートの演奏タイミング」「演奏者の演奏タイミング」を独立かつ対話的に制御できるような数理モデルを考える．

3. 手法の概要

本合奏システムの構成を図 1 に示す．本手法では，演奏者の位置を追従するために，音響信号とカメラ映像に基づき楽譜追従を行う．また，楽譜追従の事後分布から得られた統計情報を元に，演奏者の演奏している位置の生成過程に基づき，演奏者の位置を予測する．伴奏パートの発音タイミングを決定するためには，演奏者のタイミングを予測モデルと，伴奏パートが取りうるタイミングの生成過程を結合することで，伴奏パートのタイミングを生成する．

4. 楽譜追従

演奏者が現在弾いている楽曲中の位置を推定するために，楽譜追従を用いる．本システムの楽譜追従手法では，楽譜の位置と演奏されているテンポを同時に表現する離散的な状態空間モデルを考える．観測音を状態空間上の隠れマルコフ過程 (hidden Markov model; HMM) としてモデル化し，状態空間の事後分布を delayed-decision 型の forward-backward アルゴリズムで逐次推定する [*1]．事後分布の MAP 値が楽譜上でオンセットとみなされる位置を通過した時点で，事後分布のラプラス近似を出力する．

状態空間の構造に関して述べる．まず，楽曲を R 個の

[*1] Delayed-decision 型の foward-backward アルゴリズムとは, forward アルゴリズムを逐次実行し，現在の時刻がデータの終端と見なし backward アルゴリズムを走らせることで，現在の時刻より数フレーム前の状態に対する事後分布を算出することを言う．

区間に分け，それぞれの区間を一つの状態とする．r 番目の区間では，その区間を通過するのに必要なフレーム数 n と，それぞれの n に対し，現在の経過フレーム $0 \leq l < n$ を状態変数として持つ．つまり，n はある区間のテンポに相当し，r と l を組み合わせたものが楽譜上の位置に相当する．このような状態空間上の遷移を，次のようなマルコフ過程として表現する：

(1) (r, n, l) から自分自身：p

(2) $(r, n, l < n)$ から $(r, n, l+1)$：$1-p$

(3) $(r, n, n-1)$ から $(r+1, n', 0)$：$(1-p)\frac{1}{2\lambda^{(T)}} e^{-\lambda^{(T)}|n'-n|}$.

このようなモデルは，explicit-duration HMM [16] と left-to-right HMM 両者の特長を兼ね持つ．すなわち，n の選択により，区間内の継続長を大まかに決めつつも，区間内における微小なテンポ変動を自己遷移確率 p で吸収できる．区間の長さや自己遷移確率は，楽譜データを解析して求める．具体的には，テンポ指令や，フェルマータといったアノテーション情報を活用する．

次に，このようなモデルの観測尤度を定義する．それぞれの状態 (r, n, l) には，ある楽曲中の位置 $\tilde{s}(r, n, l)$ が対応している．また，楽曲中における任意の位置 s に対して，観測される定 Q 変換 (CQT) と ΔCQT の平均値 $\bar{\mathbf{c}}_{\mathbf{s}} \in \mathbb{R}^F$ と $\boldsymbol{\Delta}\bar{\mathbf{c}}_{\mathbf{s}} \in \mathbb{R}^{+F}$ に加え，精度 $\kappa_s^{(c)}$ と $\kappa_s^{(\Delta c)}$ がそれぞれ割り当てられる．これらに基づき，時刻 t ににおいて，CQT，$\mathbf{c}_t \in \mathbb{R}^F$，と ΔCQT，$\Delta\mathbf{c}_t \in \mathbb{R}^{+F}$，を観測したとき，状態 (r_t, n_t, l_t) に対応する観測尤度を以下のように定義する：

$$
\begin{aligned}
&p(c_t, \Delta c_t | (r_t, n_t, l_t), \lambda, \{\bar{c}_s\}_{s=1}^S, \{\Delta\bar{c}_s\}_{s=1}^S) \\
&\quad = \mathrm{vMF}(c_t | \bar{c}_{\tilde{s}(r_t,n_t,l_t)}, \kappa^{(c)}_{\tilde{s}(r_t,n_t,l_t)}) \\
&\quad \times \mathrm{vMF}(\Delta c_t | \Delta\bar{c}_{\tilde{s}(r_t,n_t,l_t)}, \kappa^{(c)}_{\tilde{s}(r_t,n_t,l_t)}). \quad (1)
\end{aligned}
$$

ここで，$\mathrm{vMF}(\mathbf{x}|\mu, \kappa)$ とは von Mises-Fisher 分布を指す [*2].

$\bar{c}$ や $\Delta\bar{c}$ を決める際には，楽譜表現のピアノロールと，各音から想定される CQT のモデルを用いる．まず楽譜上に存在する音高と楽器名のペアに対して固有のインデックス i を割り当てる．また，i 番目の音に対して，平均的な観測 CQT w_{if} を割り当てる．楽譜上の位置 s において，i 番目の音の強度を h_{si} を置くと，$\bar{c}_{s,f}$ は次のように与えられる：

$$
\bar{c}_{s,f} = \sum_i h_{s,i} w_{i,f}. \quad (2)
$$

$\Delta\bar{c}$ は，$\bar{c}_{s,f}$ に対して s 方向に一次差分を取り，半波整流することで得られる．また，付録に記すように，このようなモデルから得られたデータを元に，演奏者のテンポ軌跡や音色を学習することもできる．

無音の状態から楽曲を開始する際には，視覚情報がより重要になるため [5]，そこで，本システムでは，図 2 に例示するように，演奏者の前に配置されたカメラから検出されたキューを活用する．本手法では，合奏システムをトップダウンに制御するアプローチ [8] とは異なり，観測尤度に直接キューの有無を反映させることで，音響信号とキュー情報を統一的に扱う．そこで，まず楽譜情報にキューが必要とされる箇所 $\{\hat{q}_i\}$ を抽出する．$\hat{q}_i$ には，楽曲の開始地点やフェルマータの位置が含まれる．楽譜追従を実行中にキューを検出した場合，楽譜上の位置 $\cup[\hat{q}_i - \tau, \hat{q}_i]$ に対応する状態の観測尤度を 0 にすることで，キューの位置以降に事後分布を誘導する．

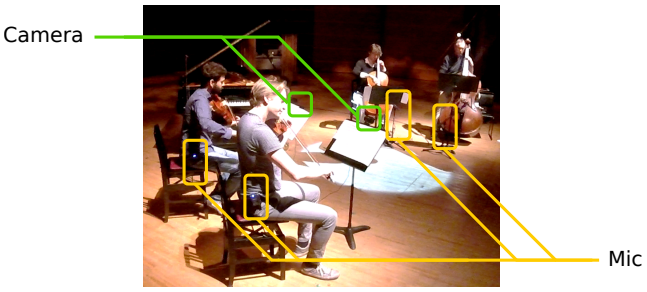

図 2 カメラの配置例.

楽譜追従により，合奏エンジンは，楽譜上で音が切り替わった位置から数フレーム後に，現在推定される位置やテンポの分布を正規分布として近似したものを受け取る．すなわち，楽譜追従エンジンは，楽譜データ上に存在する n 番目の音の切り替わり (「オンセットイベント」と呼ぶ) を検出したら，そのオンセットイベントが検出された時刻のタイムスタンプ t_n と，推定された楽譜上の平均位置 μ_n とその分散 σ_n^2 を合奏タイミング生成部に通知する．なお，delayed decision 型の推定を行うため，通知自体には 100ms の遅延が生じる．

5. 演奏タイミング結合モデル

合奏エンジンは, 楽譜追従から通知された情報 (t_n, μ_n, σ_n^2) を元に，適切な合奏エンジンの再生位置を計算する．合奏エンジンが演奏者に合わせるためには，(1) 演奏者が演奏するタイミングの生成過程 (2) 伴奏パートが演奏するタイミングの生成過程 (3) 演奏者を聞きながら伴奏パートが演奏する過程の三つを独立にモデル化することが好ましい．このようなモデルを使い，伴奏パート生成したい演奏タイミングと，演奏者の予測位置を加味しながら，最終的な伴奏パートのタイミングを生成する．

5.1 演奏者の演奏タイミング生成過程

演奏者の演奏タイミングを表現するため，演奏者が，t_n と t_{n+1} の間で楽譜上の位置を，速度 $v_n^{(p)}$ で直線運動していると仮定する．すなわち，$x_n^{(p)}$ を t_n での演奏者が弾いている楽譜上の位置とし，$\boldsymbol{\epsilon}_n^{(p)} \in \mathbb{R}^2$ を速度や楽譜上の位置に対するノイズとし，次のような生成過程を考える：

[*2] $\mathbf{x} \in S^D$, $\mu \in S^D (S^D : D-1$ 次元単位球面) となるよう正規化し $\mathrm{vMF}(\mathbf{x}|\mu, \kappa) \propto \frac{\kappa^{D/2-1}}{I_{D/2-1}(\kappa)} \exp(\kappa\mu'\mathbf{x})$ として定義する.

$$x_n^{(p)} = x_{n-1}^{(p)} + \Delta T_{n,n-1} v_{n-1}^{(p)} + \epsilon_{n,0}^{(p)} \quad (3)$$
$$v_n^{(p)} = v_{n-1}^{(p)} + \epsilon_{n,1}^{(p)} \quad (4)$$

ただし $\Delta T_{m,n} = t_m - t_n$ とする．

ノイズ $\boldsymbol{\epsilon}_n^{(p)}$ は，テンポの変化に加え，アゴーギクや発音タイミング誤差が含まれる．前者を表すためには，テンポ変化に応じて発音タイミングも変わる [13] ことを踏まえ，t_n と t_{n-1} の間を，分散 ψ^2 の正規分布から生成された加速度で遷移するモデルを考える．すると，$\epsilon_n^{(p)}$ の共分散行列は，$\mathbf{h} = [\frac{\Delta T_{n,n-1}^2}{2}, \Delta T_{n,n-1}]$ とすると $\Sigma_n^{(p)} = \psi^2 \mathbf{h}'\mathbf{h}$ と与えられ，テンポ変化と発音タイミング変化が相関するようになる．また，後者を表すため，標準偏差 $\sigma_n^{(p)}$ の白色雑音を考え，$\sigma_n^{(p)}$ を $\Sigma_{n,0,0}^{(p)}$ に加算する．したがって，$\sigma_n^{(p)}$ を $\Sigma_{n,0,0}^{(p)}$ に加算した行列を $\hat{\Sigma}_n^{(p)}$ とすると，$\boldsymbol{\epsilon}_{\mathbf{n}}^{(\mathbf{p})} \sim \mathcal{N}(0, \hat{\Sigma}_n^{(p)})$ と与えられる *3．

次に，楽譜追従システムが報告する，ユーザの演奏タイミングの履歴 $\boldsymbol{\mu}_n = [\mu_n, \mu_{n-1}, \cdots \mu_{n-I_n}]$ と $\boldsymbol{\sigma}_n^2 = [\sigma_n^2, \sigma_{n-1}^2, \cdots \sigma_{n-I_n}^2]$ を，式 (3) や式 (4) と結びつけることを考える．　ここで，I_n は，考慮する履歴の長さであり，t_n よりも 1 拍前のイベントまでを含むように設定される．このような $\boldsymbol{\mu}_n$ や $\boldsymbol{\sigma}_n^2$ の生成過程を次のように定める：

$$\boldsymbol{\mu}_n \sim \mathcal{N}\left(\mathbf{W}_n [x_n^{(p)} v_n^{(p)}], \mathrm{diag}(\boldsymbol{\sigma}_n^2)\right). \quad (5)$$

ここで $\mathbf{W}_n$ は，$x_n^{(p)}$ と $v_n^{(p)}$] から観測 $\boldsymbol{\mu}_n$ を予測するための回帰係数である．本稿では，$\mathbf{W}_n$ を以下のように定義する：

$$\mathbf{W}_n^T = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \Delta T_{n,n} & \Delta T_{n,n-1} & \cdots & \Delta T_{n,n-I_n+1} \end{pmatrix}. \quad (6)$$

従来のように，観測値としてに直近の μ_n を使う [11] のではなく，それ以前の履歴も用いることにより，楽譜追従が一部で失敗しても動作が破綻しにくくなると考えられる．また，W_n をリハーサルを通じて獲得することも可能であると考えられ，テンポの増減のパターンといった，長時間の傾向に依存する演奏法にも追従ができるようになると考えられる．このようなモデルは，テンポと楽譜上の位置変化の関係性を明記するという意味では，トラジェクトリ HMM [17] のコンセプトを連続状態空間に適用したものに相当する．

5.2　伴奏パートの演奏タイミング生成過程

前述したような，演奏者のタイミングモデルを使うことで，演奏者の内部状態 $[x_n^{(p)}, v_n^{(p)}]$ を，楽譜追従が報告した位置の履歴から推論することができる．合奏システムは，このような推論と，伴奏パートがどのように「弾きたいか」というクセを協調させながら，最終的な発音タイミングを

*3　$\mathcal{N}(x|\mu, \Sigma) = \frac{1}{2\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right)$

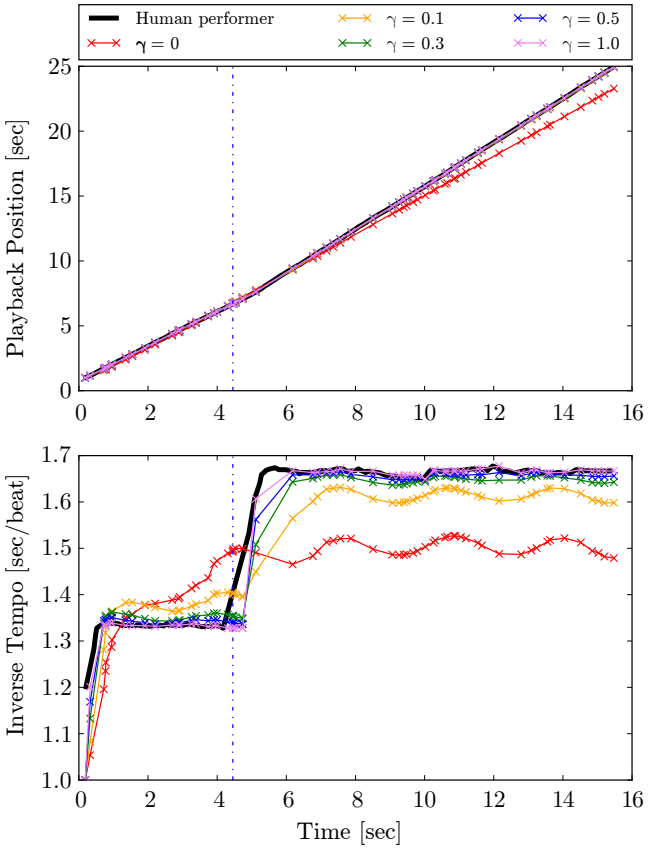

図 3　$\beta = 0.9$，$\gamma = [0, 0.1, 0.3, 0.5, 1.0]$ で結合タイミングモデルをシミュレーションしたときの伴奏パート再生位置 (上) とテンポ軌跡 (下)．ユーザ (Human performer) は 4 秒過ぎ (縦破線) でテンポをステップ型に切り替え，合奏システムは定数テンポを正弦波で振幅変調させた．γ を変えることで，伴奏パートのテンポ軌跡と伴奏パートのテンポ軌跡の貢献が変わることが分かる．また，$\gamma = 0$ と $\gamma = 1$ の間で挙動が滑らかに変化することも分かる．

推論する．そこで，ここでは伴奏パートがどのように「弾きたいか」という，伴奏パートにおける演奏タイミングの生成過程について考える．

伴奏パートの演奏タイミングでは，与えられたテンポ軌跡から一定の範囲内のテンポ軌跡で演奏される過程を考える．与えられるテンポ軌跡とは，演奏表情付けシステム [10] や人間の演奏データを使うことが考えられる．合奏システムが n 番目のオンセットイベントを受け取ったときに，楽曲上のどの位置を弾いているかの予測値 $\hat{x}_n^{(a)}$ とその相対速度 $\hat{v}_n^{(a)}$ を次のように表現する：

$$\hat{x}_n^{(a)} = x_{n-1}^{(a)} + \Delta T_{n,n-1} v_{n-1}^{(a)} + \epsilon_{n,0}^{(a)} \quad (7)$$
$$\hat{v}_n^{(a)} = \beta v_{n-1}^{(a)} + (1-\beta)\bar{v}_n^{(a)} + \epsilon_{n,1}^{(a)}. \quad (8)$$

ここで，$\bar{v}_n^{(a)}$ とは時刻 t_n で報告された楽譜上の位置 n において事前に与えたテンポであり，事前に与えたテンポ軌跡を代入する．また，$\epsilon^{(a)}$ は，事前に与えたテンポ軌跡から生成された演奏タイミングに対して許容される逸脱の範囲を定める．このようなパラメータにより，伴奏パートと

して音楽的に自然な演奏の範囲を定める．$\beta \in [0,1]$ とは事前に与えたテンポにどれだけ強く引き戻そうとするかを表す項であり，テンポ軌跡を $\bar{v}_n^{(a)}$ に引き戻そうとする効果がある．このようなモデルはオーディオアラインメントにおいて一定の効果があるため，同一楽曲を演奏するタイミングの生成過程として妥当性があると示唆される [9]．なお，このような制約がない場合 ($\beta = 1$)，$\hat{v}$ はウィナー過程に従うため，テンポが発散し，極端に速かったり遅い演奏が生成されうる．

5.3 演奏者と伴奏パートの演奏タイミング結合過程

ここまでは，演奏者の発音タイミングと，伴奏パートの発音タイミングをそれぞれ独立にモデル化した．ここでは，これらの生成過程を踏まえた上で，演奏者を聞きながら，伴奏パートが「合わせる」過程について述べる．

そこで，伴奏パートが人に合わせる際，伴奏パートが現在弾こうとする位置の予測値と，演奏者の現在位置の予測値の誤差を徐々に補正するような挙動を記述することを考える．以下では，このような，誤差を補正する程度を記述した変数を「結合係数」と呼ぶ．結合係数は，伴奏パートと演奏者の主従関係に影響される．例えば，演奏者が伴奏パートよりも明瞭なリズムを刻んでいる場合，伴奏パートは演奏者に強めに合わせること多い．また，リハーサル中に主従関係を演奏者から指示された場合は，指示されたように合わせ方を変える必要がある．つまり，結合係数は，楽曲のコンテキストや演奏者との対話に応じて変わる．そこで，t_n を受け取った際の楽譜位置における結合係数 $\gamma_n \in [0,1]$ が与えられたとき，伴奏パートが演奏者に合わせる過程を以下のように記述する：

$$x_n^{(a)} = \hat{x}_n^{(a)} + \gamma_n(x_n^{(p)} - \hat{x}_n^{(a)}) \tag{9}$$

$$v_n^{(a)} = \hat{v}_n^{(a)} + \gamma_n(v_n^{(p)} - \hat{v}_n^{(a)}). \tag{10}$$

このモデルでは，γ_n の大小に応じて，追従度合いが変わる．例えば，$\gamma_n = 0$ の時は，伴奏パートは演奏者に一切合わせず，$\gamma_n = 1$ の時は，伴奏パートは演奏者に完璧に合わせようとする．

このようなモデルでは，伴奏パートが演奏しうる演奏 $\hat{x}_n^{(a)}$ の分散と，演奏者の演奏タイミング $x_n^{(p)}$ における予測誤差も結合係数によって重み付けられる．そのため，$x^{(a)}$ や $v^{(a)}$ の分散は演奏者の演奏タイミング確率過程自体と，伴奏パートの演奏タイミング確率過程自体が協調されたものになる．そのため，演奏者と合奏システム，両者が「生成したい」テンポ軌跡を自然に統合できていることがわかる．

$\beta = 0.9$ における，本モデルのシミュレーションを図 3 に示す．このように γ を変えることで，伴奏パートのテンポ軌跡 (正弦波) と，演奏者のテンポ軌跡 (ステップ関数) の間を補完できることが分かる．また，β の影響により，生成されたテンポ軌跡は，演奏者のテンポ軌跡よりも伴奏パートの目標とするテンポ軌跡に近づけるようになっていることが分かる．つまり，$\bar{v}^{(a)}$ よりも演奏者が速い場合は演奏者を「引っ張り」，遅い場合は演奏者を「急かす」ような効果があると考えられる．

5.3.1 結合係数 γ の算出方法

結合係数 γ_n に表すような演奏者同士の同期度合いは，いくつかの要因により設定される [4]．まず，楽曲中のコンテキストに主従関係が影響される．例えば，合奏をリードするのは，分かりやすいリズムを刻むパートであることが多い [7]．また，対話を通じて主従関係を変えることもある．

楽曲中のコンテキストから主従関係を設定するため，楽譜情報から，音の密度 $\phi_{\mathbf{n}} = [$ **伴奏パートに対する音符密度の移動平均，演奏者パートに対する音符密度の移動平均** $]$ を算出する．音の数が多いパートの方が，テンポ軌跡を決めやすいため，このような特徴量を使うことで近似的に結合係数を抽出できると考えられる．このとき，伴奏パートが演奏を行っていない場合 ($\phi_{n,0} = 0$)，合奏の位置予測は演奏者に完全に支配され，また，演奏者が演奏を行わない箇所 ($\phi_{n,1} = 0$) では，合奏の位置予測は演奏者を完全に無視するような挙動が望ましい．そこで，次のように γ_n を決定する：

$$\gamma_n = \frac{\phi_{n,1} + \epsilon}{\phi_{n,1} + \phi_{n,0} + 2\epsilon}. \tag{11}$$

ただし，$\epsilon > 0$ は十分に小さい値とする．人間同士の合奏では，完全に一方的な主従関係 ($\gamma_n = 0$ や $\gamma_n = 1$) は発生しにくい [5] のと同様に，上式のようなヒューリスティックは，演奏者と伴奏パートどちらも演奏している場合は完全に一方的な主従関係にはならない．完全に一方的な主従関係は，演奏者・合奏エンジンどちらかがしばらく無音である場合のみ起こるが，このような挙動はむしろ望ましい．

また，γ_n はリハーサル中など，必要に応じて，演奏者やオペレータが上書きすることができる．γ_n の定義域が有限であり，かつその境界条件での挙動が自明であることや，γ_n の変動に対し挙動が連続的に変化することは，リハーサル中に適切な値を人間が上書きする上で望ましい特性であると考えられる．

5.4 オンライン推論

合奏システムの運用時は，(t_n, μ_n, σ_n^2) を受け取ったタイミングで，前述の演奏タイミングモデルの事後分布を更新する．提案手法はカルマンフィルタを用いて効率的に推論することができる [6]．(t_n, μ_n, σ_n^2) が通知された時点でカルマンフィルタの predict と update ステップを実行し，時刻 t において伴奏パートが演奏すべき位置を以下のように予測する：

$$x_n^{(a)} + (\tau^{(s)} + t - t_n)v_n^{(a)}. \tag{12}$$

ここで $\tau^{(s)}$ とは，合奏システムにおける入出力遅延である．

なお，本システムでは，[11] と同様，伴奏パート発音時

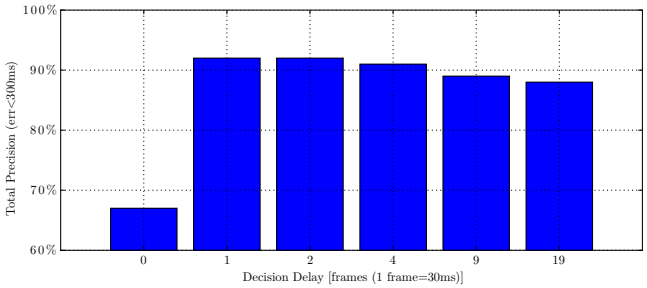

図 4 Delayed decison forward-backward アルゴリズムの遅延量に対する total precision.

にも状態変数を更新する．つまり，前述したように，楽譜追従結果に応じて predict/update ステップを実行することに加え，伴奏パートが発音した時点で，predict ステップのみを行い，得られた予測値を状態変数に代入する．

6. 評価実験

本システムを評価するため，まず演奏者の位置推定精度を評価する．合奏のタイミング生成に関しては，合奏のテンポを規定値に引き戻そうとする項である β や，伴奏パートを演奏者にどれだけ合わせるかの指標である γ の有用性を，演奏者へのヒアリングを行うことで評価する．

6.1 楽譜追従の評価

楽譜追従精度の評価を行うため，Bergmuller のエチュードに対する追従精度を評価した．評価データとして，Bergmuller のエチュード (Op. 100) のうち，14 曲 (1 番, 4 番-10 番, 14 番, 15 番, 19 番, 20 番, 22 番, 23 番) をピアニストが演奏したデータを収録したものを使い，譜面追従精度を評価した．なお，この実験ではカメラの入力は使用しなかった．評価尺度には MIREX [3] に倣い，Total precision を評価した．Total precision とは，アラインメントの誤差がある閾値 τ に収まる場合を正解とした場合の，コーパス全体に対する精度を示す．

まず，delayed decision 型の推論に関する有用性を検証するため, delayed decision forward backward アルゴリズムにおける遅延フレーム量に対する total precision ($\tau = 300$ms) を評価した．結果を図 4 に示す．数フレーム前の結果の事後分布を活用することで精度が上がることが分かる．また，遅延量が 2 フレームを超えると精度は徐々に下がることも分かる．また，遅延量 2 フレームの場合 $\tau = 100$ ms で total precision=82%, $\tau = 50$ ms で 64%であった．

6.2 演奏タイミング結合モデルの検証

演奏タイミング結合モデルの検証は，演奏者へのヒアリングを通じて行った．本モデルの特徴としては，合奏エンジンが想定テンポに引き戻そうとする β と，結合係数 γ の存在であり，これら両者についての有効性を検証した．

まず，結合係数の影響を外すため，式 (4) を $v_n^{(p)} = \beta v_{n-1}^{(p)} + (1-\beta)\bar{v}_n^{(a)}$ とし，$x_n^{(a)} = x_n^{(p)}$，$v_n^{(a)} = v_n^{(p)}$ としたシステムを用意した．つまり，テンポの期待値が $\bar{v}$ にあり，その分散が β により制御されるようなダイナミクスを仮定しながら，楽譜追従の結果をフィルタリングした結果を直接伴奏の演奏タイミング生成に使うような合奏エンジンを考えた．まず $\beta = 0$ に設定した場合の合奏システムを，ピアニスト 6 名に一日間利用してもらったあと，使用感に関してヒアリングを行った．対象曲はクラシック・ロマン派・ポピュラーなど幅広いジャンルの曲から選曲した．ヒアリングでは，合奏に人間が合わせようとすると，伴奏パートも人間に合わせようとし，テンポが極端に遅くなったり速くなるという不満が支配的であった．このような現象は，式 (12) における $\tau^{(s)}$ が不適切に設定されていることにより，システムの応答が演奏者と微妙に合わない場合に発生する．例えば，システムの応答が想定よりも少し早い場合，ユーザは少し早めに返されるシステムに合わせようとするため，テンポを上げる．その結果，そのテンポに追従するシステムが更に早めに応答を返すことで，テンポが加速し続ける．

次に，$\beta = 0.1$ で同じ曲目を使って別のピアニスト 5 名と，$\beta = 0$ の実験にも参加したピアニスト 1 名で実験を行った．$\beta = 0$ の場合と同じ質問内容でヒアリングを行ったが，テンポが発散する問題は聞かれなかった．また，$\beta = 0$ でも実験に協力したピアニストからも追従性が改善しているというコメントがあった．ただし，演奏者がある曲に対して想定しているテンポと，システムが引き戻そうとするテンポに大きな齟齬がある場合，システムがもたつく・急かす，といったコメントが聞かれた．この傾向は特に未知の曲を弾く場合，つまり演奏者が「常識的な」テンポを知らない場合，において見られた．このことから，システムが一定のテンポに引き込もうとする効果により，テンポの発散を未然に防ぐ一方で，伴奏パートとテンポに関する解釈が極端に異なる場合，伴奏パートに煽られるような印象を受けることが示唆された．また，追従性に関しては，楽曲のコンテキストに応じて変えたほうがよいことも示唆された．なぜならば，楽曲の特性よって「引っ張ってもらったほうがいい」「もっと合わせて欲しい」といった，合わせ方の度合いに関する意見がほぼ一貫したためである．

最後に，プロの弦カルテットに $\gamma = 0$ に固定したシステムと，演奏のコンテキストに応じて γ を調整したシステムを使ってもらったところ，後者の方が挙動が良いというコメントがあり，その有用性が示唆された．ただし，この検証では後者のシステムが改善後のシステムであることを被験者が知っていたため，今後 AB 法などを使い追加検証する予定である．また，リハーサル中の対話に応じて γ を変更する局面がいくつか存在したため，結合係数をリハーサル中で変更することが有用であると示唆された．

7. 実証実験

本手法の実証実験のため，東京藝術大学とベルリンフィル・シャルーンアンサンブル協力のもと，弦カルテットと，自動合奏システムが制御するピアノパートの共演をコンサートとして実施した．この実証実験では，シューベルトのピアノ五重奏「鱒」を，弦はシャルーンアンサンブル，ピアノはプレイヤーピアノが演奏した．以下では，この実証実験から得れれた，自動合奏技術の要求仕様に関する知見を報告する．

7.1 リハーサルでの運用を通して

リハーサルでは，(1) 少数データから演奏を学習することと，(2) 追従エンジンを対話的に操作し，操作結果が即座に反映されることが要求される．なぜならば，リハーサルのほとんどは，曲中の一部の演奏方法を，演奏と対話を通じてすり合わせることに使われるからである．つまり，実際に楽曲全体を通す回数は限られており，従来の学習ベースによるリハーサルで必要とされている，10 回程度の通し練習 [11, 15] を収録することは困難である．従って，少ない数の演奏データから演奏のクセを抽出できるメカニズムが必要である．また，対話を通じたすり合わせでは，口頭で演奏に関する議論をを行った直後に，該当箇所を演奏しなおし，合意が形成されたことを確認する．従って，リハーサルでは演奏者の口頭指示などを即座に合奏エンジンに反映させるための瞬発力が求められ，対話と学習の二つを統合すること [18] が重要になる．

7.2 コンサートでの運用を通して

演奏の本番中には，オペレータが介入できるシステムが必須である．なぜならば，合奏システムの誤推定に対する救済措置が必要だからだ．従って，合奏システムでは，システムのパラメータをトップダウンかつ予測可能な形で変更できることが必要である．

また，合奏エンジンが破綻してもシーケンスが最後まで破綻せず再生するフェイルセーフが望ましい．本システムでは，合奏エンジンの状態変数とタイムスタンプを，合奏エンジンとは別のプロセスで動いているシーケンサに Open Sound Control(OSC) メッセージとして UDP 上で送信し，シーケンサは受信した OSC を元にデータを再生した．これにより，仮にエンジンが破綻し自動合奏プロセスを終了しても，シーケンサ自体は直近の状態で再生し続けることができる．この場合，合奏システムの状態変数は，テンポと再生タイミングのオフセットといった，通常のシーケンサとして表記しやすいものであることが好ましい．

8. まとめ

本稿では，既知の曲を，人間の演奏者に合わせながら合奏するための自動合奏エンジンにおける，アンサンブルのタイミングモデルについて論じた．演奏者の演奏タイミングと，合奏パートの演奏タイミングと，合奏パートが演奏者に合わせる過程を分離して表現することにより，これら三要素を独立に学習・操作できるようになる．今後の課題としては，本手法のより詳細な定量評価，実世界での運用を通じたシステムの実用性評価，少数データからの学習，結合係数の予測を行うためのモデル化，伴奏パートのタイミング生成過程の緻密化などが挙げられる．

謝辞

実証実験にあたっては東京藝術大学「センター・オブ・イノベーション (COI) プログラム」の協力を受けた．また，システムの評価においては，ヤマハ (株) 楽器開発統括部の中村吉就氏にご協力いただき，演奏のキュー検出にはヤマハ (株) 研究開発の山本和彦氏にご協力いただいた．

参考文献

[1] Cont, A.: A Coupled Duration-Focused Architecture for Real-Time Music-to-Score Alignment, *IEEE PAMI*, Vol. 32, No. 6, pp. 974–987 (2010).
[2] Cont, A., Echeveste, J., Giavitto, J. and Jacquemard, F.: Correct automatic accompaniment despite machine listening or human errors in Antescofo, *Proc. ICMC* (2012).
[3] Cont, A., Schwarz, D., Schnell, N. and Raphael, C.: Evaluation of Real-Time Audio-to-Score Alignment, *Proc. ISMIR*, Vienna, Austria, pp. 315–316 (2007).
[4] Fabian, D., Timmers, R. and Schubert, E.(eds.): *Expressiveness in music performance*, Oxford University Press (2014).
[5] Goebl, W. and Palmer, C.: Synchronization of timing and motion among performing musicians, *Music Perception: An Interdisciplinary Journal*, Vol. 26, No. 5, pp. 427–438 (2009).
[6] Kalman, R. E.: A new approach to linear filtering and prediction problems, *Journal of basic Engineering*, Vol. 82, No. 1, pp. 35–45 (1960).
[7] Keller, P. E.: Attentional resource allocation in musical ensemble performance, *Psychology of Music*, Vol. 29, No. 1, pp. 20–38 (2001).
[8] Lim, A., Mizumoto, T., Cahier, L.-K., Otsuka, T., Takahashi, T., Komatani, K., Ogata, T. and Okuno, H. G.: Robot musical accompaniment: integrating audio and visual cues for real-time synchronization with a human flutist, *Proc. IROS*, IEEE, pp. 1964–1969 (2010).
[9] Maezawa, A., Itoyama, K., Yoshii, K. and Okuno, H. G.: Unified inter- and intra-recording duration model for multiple music audio alignment, *Proc. WASPAA*, pp. 1–5 (2015).
[10] Okumura, K., Sako, S. and Kitamura, T.: Laminae: A stochastic modeling-based autonomous performance rendering system that elucidates performer characteristics, *Proc. ICMC* (2014).
[11] Raphael, C.: A Bayesian Network for Real-Time Musical Accompaniment, *Proc. NIPS*, pp. 1433–1439 (2001).
[12] Raphael, C.: Music Plus One and Machine Learning, *Proc. ICML*, pp. 21–28 (2010).

[13] Schulze, H. H., Cordes, A. and Vorberg, D.: Keeping Synchrony While Tempo Changes: Accelerando and Ritardando, *Music Perception*, Vol. 22, No. 3, pp. 461–477 (2005).
[14] Wada, S., Horiuchi, Y. and Kuroiwa, S.: Temo Prediction Model for Accompaniment System, *Proc. ICMC*, pp. 1298–1303 (2014).
[15] Xia, G., Wang, Y., Dannenberg, R. B. and Gordon, G.: Spectral Learning for Expressive Interactive Ensemble Music Performance, *Proc. ISMIR*, pp. 816–822 (2015).
[16] Yu, S.-Z. and Kobayashi, H.: An efficient forward-backward algorithm for an explicit-duration hidden Markov model, *IEEE signal processing letters*, Vol. 10, No. 1, pp. 11–14 (2003).
[17] Zen, H., Tokuda, K. and Kitamura, T.: Reformulating the HMM as a trajectory model by imposing explicit relationships between static and dynamic feature vector sequences, *Computer Speech & Language*, Vol. 21, No. 1, pp. 153–173 (2007).
[18] 堀内靖雄，奥井　学，鈴木泰山，田中穂積：伴奏システムのためのリハーサル，情処研報音楽情報科学研究会 1994-MUS-8-10，No. 103, pp. 51–56 (1994).

付　　録

A.1 リハーサルからの学習

演奏者の「癖」を獲得するため，楽譜追従から算出された時刻 t での MAP 状態 $\hat{s}_t$ と，その入力特徴系列 $\{\mathbf{c}_t\}_{t=1}^{T}$ を元に，h_{si} と w_{if} およびテンポ軌跡を推定する．ここでは，これらの推定方法について簡単に述べる．h_{si} と w_{if} の推定においては，次のような Poisson-Gamma 系の Informed NMF モデルを考え，事後分布を推定する：

$$c_{t,f} \sim \mathrm{Poisson}(\sum_{i}^{I} h_{\hat{s}_t,i} w_{i,f}) \tag{A.1}$$

$$h_{s,i} \sim \mathrm{Gamma}(a_0^{(h)}, b_{0,s,i}^{(h)}) \tag{A.2}$$

$$w_{i,f} \sim \mathrm{Gamma}(a_{i,f}^{(w)}, b_{i,f}^{(w)}). \tag{A.3}$$

ここで現れる超パラメータは楽器音データベースや楽譜表現のピアノロールから適当に算出する．

事後分布は，変分ベイズ法で近似的に推定する．具体的には，事後分布 $p(h, w|c)$ を $q(h)q(w)$ という形で近似し，事後分布と $q(h)q(w)$ の間の KL 距離を，補助変数を導入しながら最小化する．このようにして推定された事後分布から，楽器音の音色に相当するパラメータ w の MAP 推定を保存し，以降のシステム運用で使う．ピアノロールの強さに相当する h を使うこともできるが，本システムでは未使用である．

続いて，演奏者がそれぞれの楽曲上の区間を演奏する長さ (すなわちテンポ軌跡) を推定する．テンポ軌跡を推定すると演奏者特有のテンポ表現を復元できるため，演奏者の位置予測が改善される．一方，リハーサルの回数が少ない場合は推定誤差などによりテンポ軌跡の推定が誤り，位置予測の精度がむしろ悪化する可能性もある．そこで，テンポ軌跡を変更する際には，テンポ軌跡に関する事前情報をまず持たせ，演奏者のテンポ軌跡が事前情報から一貫して逸脱している場所のテンポのみを変えることを考える．

まず，演奏者のテンポがどれだけばらつくかを計算する．ばらつき度合いの推定値自体もリハーサルの回数が少ないと不安定になるため，演奏者のテンポ軌跡の分布自体にも事前分布を持たせる．演奏者が楽曲中の位置 s におけるテンポの平均 $\mu_s^{(p)}$ と分散 $\lambda_s^{(p)^{-1}}$ が $\mathcal{N}(\mu_s^{(p)}|m_0, b_0\lambda_s^{(p)^{-1}})\mathrm{Gamma}(\lambda_s^{(p)^{-1}}|a_0^{\lambda}, b_0^{\lambda})$ に従うとする．すると，M 回のリハーサルから得られたテンポの平均が $\mu_s^{(R)}$, 精度が $\lambda_s^{(R)^{-1}}$ であったとすると，テンポの事後分布は以下のように与えられる:

$$\begin{aligned} q(\mu_s^{(P)}, \lambda_s^{(P)^{-1}}) &= p(\mu_s^{(P)}, \lambda_s^{(P)^{-1}}|M, \mu_s^{(R)}, \lambda_s^{(R)}) \\ &= \mathcal{N}\Big(\mu_s^{(p)}|\frac{b_0 m_0 + M\mu_s^{(R)}}{b_0 + M}, (b_0 + M)\lambda_s^{(p)^{-1}}\Big) \\ &\quad \times \mathrm{Gamma}\Bigg(\lambda_s^{(p)}|a_0^{\lambda} + \frac{M}{2}, \\ &\quad b_0^{\lambda} + \frac{1}{2}\Big(M\lambda_s^{(R)^{-1}} + \frac{M b_0(\mu_s^{(R)} - m_0)^2}{M + b_0}\Big)\Bigg) \end{aligned} \tag{A.4}$$

このようにして得られた事後分布を，楽曲中の位置 s で取りうるテンポの分布 $\mathcal{N}(\mu_s^S, \lambda_s^{S^{-1}})$ から生成された分布とみなした場合の事後分布を求めると，その平均値は以下のように与えられる：

$$\begin{aligned} &\langle \mu_s^{(S)} \rangle_{p(\mu_s^{(S)}|\mu_s^{(P)}, \lambda_s^{(P)}, M)} = \\ &\qquad \frac{\langle \lambda_s^{(P)} \rangle \mu_s^{(S)} + \lambda_s^{(S)} \langle \mu_s^{(P)} \rangle}{\lambda_s^{(S)} + \langle \lambda_s^{(P)} \rangle} \end{aligned} \tag{A.5}$$

このようにして算出されたテンポを元に，式 (3) や式 (4) で用いられる $\boldsymbol{\epsilon}$ の平均値を更新する．