

動的時間伸縮調波時間構造化クラスタリング (TW-HTC) による 音楽音響信号と楽譜の自動整合*

武内昭典^{*1} 亀岡弘和^{*1} 齋藤大輔^{*1}
深山 覚^{*1} 嵯峨山茂樹^{*1}

[要旨] 本論文では、楽譜における各音符が音響信号においてどの成分に対応しているかを推定するオブジェクト整合の問題を扱う。従来は、演奏誤りがないことが前提となっていたのに対し、本研究では演奏誤りを含む音響信号を対象とし、楽譜における各音符と音響信号中の成分との整合と音響信号中の演奏誤り箇所の検出を同時解決するオブジェクト整合手法を提案する。評価実験において、一般の曲において詳細な加工を行えることを示した。また、誤り演奏を含む実演奏信号について実験を行い、誤り演奏を含む信号に対しても正しく演奏された信号箇所と音符の整合を行うことができ、誤り演奏位置を推定できることを示した。

キーワード 時間調波構造化クラスタリング (HTC), スパース制約, 誤り演奏検出

1. はじめに

近年、個人による音楽活動の普及に伴い、専門的な知識を持たないユーザーが簡単に音楽音響信号を操作できる技術に需要がある。

本研究では、音楽音響信号と楽譜を入力とし、楽譜上の各音符と音楽音響信号中の成分を対応付けることで、ユーザが楽譜上の簡単な操作で音響信号を音符単位で加工し、出力できるインターフェースを目指している。本稿では、その実現に向け、楽譜と音響信号との自動整合の枠組みについて論ずる。

楽譜における演奏位置が音響信号においてどの時刻に対応しているかを推定する問題はスコアアラインメントと呼ばれ、多くの研究が行われている (例えば [1-4] 等)。これらに対し本研究では、楽譜における各音符が音響信号においてどの成分に対応しているかを推定するより詳細な問題を扱う。以後、本稿では、この問題をスコアアラインメントと区別してオブジェクトアラインメントと定義する。近年、オブジェクトアラインメント手法 [5, 6] の検討も徐々に進められているが、そのほとんどにおいて、音響信号に演奏誤りが含まれないことが暗に想定されていた。しかしながら、一般ユーザを対象とした音響信号加工インターフェースを目指す場合、演奏誤りを含む音響信号にも適切に対応可能な整合手法が必要になる。

従来のスコアアラインメント手法の中には演奏誤り

に対応したもの [7] が提案されているが、本研究で目指すのは演奏誤りに対応したオブジェクトアラインメント手法の実現である。しかしながら、その実現は必ずしも容易ではない。なぜなら、演奏誤りを含む音楽音響信号に対し高い精度のオブジェクトアラインメントを行うには演奏誤り箇所または正しく演奏された箇所を正しく推定する必要がある。一方で、演奏誤り箇所を正しく検出するには、音楽音響信号と楽譜が正しく対応付けられている方が明らかに都合が良い。このように、オブジェクトアラインメントと演奏誤り検出の問題は相補的に依存しており、いわば鶏と卵の関係にあるからである。従って、両者の問題を同時解決するアプローチが有効であると考えられるが、それをどのように実現するかが本研究における課題である。

以上より、本稿では、

- 楽譜における各音符と音響信号中の成分との整合
- 音響信号中の演奏誤り箇所の検出

を同時解決するオブジェクトアラインメント手法を提案する。

2. 問題の定式化

以下、演奏誤りを含む音楽音響信号に対するオブジェクトアラインメント手法の問題を定式化する。

演奏誤りのない場合でのオブジェクトアラインメントの問題は、楽譜情報を援用した多重音解析の問題と捉えることができる。例えば、楽譜における各音符の音高と発音順序の情報をもとに、各音符に対応するスペクトログラムモデルを時間周波数平面上に配置し、これらを重畳したもの (以後、楽譜対応モデルと呼ぶ) で観測スペクトログラムにフィッティングすることで、各音符が対応するように観測信号を時間周波数成分に分解することができると考えられる。実際の演奏では、テンポは一定であるとは限らないし、各ノートの発音タ

* Time Warped-Harmonic Temporal Clustering for Score Alignment to Music Signal,
by Akinori Takeuchi, Hirokazu Kameoka, Daisuke Saito, Satoru Fukayama and Shigeki Sagayama.

^{*1} 東京大学大学院情報理工学系研究科
Graduate School of Information Science and Technology,
University of Tokyo
(問合先: 武内 昭典 東京都文京区本郷 7-3-1 工学部 6 号館 145 号室, Tel/Fax: 03-5841-6902, E-mail: takeuchi@hil.t.u-tokyo.ac.jp)

イミングと音高は所定の時刻とピッチからずれる場合が多い。また、各ノートに対応する信号のスペクトログラムの概形も楽器の種類や演奏方法によってさまざまな形をとりうる。従って、これらの不確定な変動要因を推論しながらモデルフィッティング問題を解決する必要がある。

ここでさらに、演奏誤りが含まれる場合には、演奏誤りの可能性も考慮に入れる必要がある。演奏誤りは、楽譜対応モデルと観測スペクトログラムとの間の上述の変動要因では説明できないミスマッチを生み、このミスマッチがモデルフィッティング全体に悪影響を及ぼす可能性がある。これを回避するため、演奏誤りに対応するスペクトログラムもモデル化し(これを以後、誤り演奏モデルと呼ぶ)、これと楽譜対応モデルを重畳したもので観測スペクトログラムにフィッティングすることで、演奏誤り箇所の検出ができるようになるだけでなく、演奏誤りの有無にあまり影響されずにオブジェクトアラインメントが適切に行えるようになると考えられる。以下では、上述の考え方にに基づき、演奏誤りを含む音響信号のスペクトログラムのモデル化を行い、観測スペクトログラムからそのモデルのパラメータを同時に推定するためのアルゴリズムについて述べる。

3. 動的時間伸縮調波時間構造化クラスタリング

3.1 楽譜から音響信号の生成過程

本節ではまず、楽譜から音楽音響信号が生成されるまでのプロセスについて議論する。

演奏者は楽譜に含まれる多様な情報に基づいて演奏するが、まず演奏者はどの楽器を用いるかをはじめに決め、曲全体をどのようなテンポで弾くかを計画し、意識的ないし無意識的に局所的にテンポを変化させたり音量に起伏を持たせたりピブラートやポルタメントなどのテクニックを駆使したりしながら演奏を行う。機械的な演奏とは異なり、このように実際の演奏には、

- テンポが(多くの場合、緩やかに)変化する
- 発音のタイミングが所定時刻からずれる場合がある
- ピッチが所定の音高からずれる場合がある
- 楽器の種類や演奏方法により音色が変化する
- 演奏誤りを含む場合がある

などの特徴があり、楽譜情報にこれらの特徴が加わることで人間味のある表情豊かな演奏となるわけである。以下では、実際の演奏におけるこれらの特徴をもとに楽譜対応モデルをモデル化する。

また、演奏中に意図しない演奏誤りが発生する可能性がある。演奏誤りには、

- 1) 楽譜に示されている音が鳴らなかった
- 2) 楽譜に示されている音とは異なる音が鳴った
- 3) 楽譜に示されていない音が余分に鳴った

などのケースが考えられよう。本稿では、これらを演奏誤りと定義して誤り演奏モデルをモデル化する。

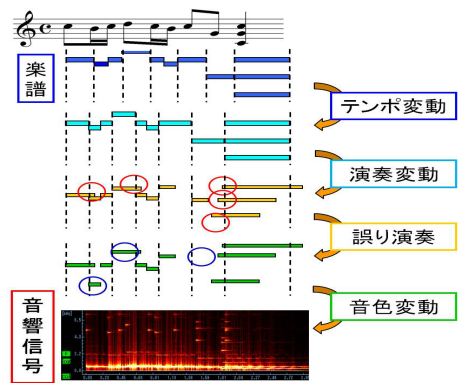


図-1 楽譜から音楽音響信号が生成される過程

3.2 スペクトログラムモデル

ここでは、本稿で基本となるスペクトログラムモデルについて述べる。 x を対数周波数、 t を時刻として、観測音響信号のスペクトログラムを $Y(x, t)$ と表す。一方、テンポを一定で演奏した場合の楽譜上での仮想的な時刻(以後これを実際の時刻 t と区別して正規化時刻と呼ぶ)を u とし、楽譜対応モデルを $F(x, u)$ と表す。ここで、各楽音のスペクトログラムを、調波時間構造化クラスタリング(Harmonic-Temporal Clustering; HTC) [8] と同様、周波数方向の調波構造と時間方向のパワーエンベロープをパラメトリックな関数で表現したスペクトログラムモデルを用いてモデル化する。楽譜上の k 番目の音符に対応するスペクトログラムモデルを $r_k(x, u)$ とすると、楽譜対応モデルはこれを K 個重畳したもの

$$F(x, u) = \sum_{k=1}^K r_k(x, u) \quad (1)$$

となる。 $r_k(x, u)$ は総エネルギーが w_k 、時刻 u において基本周波数が $\mu_k(u)$ の楽音スペクトログラムのモデルであり、時刻 u における周波数スペクトルが基本周波数 $\mu_k(u)$ とその高調波周波数 $\mu_k(u) + \log n$ ($n = 2, \dots, N$) を中心としたガウス分布関数の和で表される。時刻 u における各高調波のパワーを $U_{k,n}(u)$ とすると、 $r_k(x, u)$ は具体的に

$$r_k(x, u) = w_k \sum_{n=1}^N \frac{U_{kn}(u)}{\sqrt{2\pi}\sigma_k} e^{-\frac{(x - \mu_k(u) - \log n)^2}{2\sigma_k^2}} \quad (2)$$

と表される。ここでさらに、パワーエンベロープ $U_{kn}(u)$ を M 個のガウス分布関数の和

$$U_{kn}(t) = \sum_{m=1}^M \frac{\rho_{knm}}{\sqrt{2\pi}\phi_{kn}} e^{-\frac{(u - \tau_k - m\phi_{kn})^2}{2\phi_{kn}^2}} \quad (3)$$

で表す。ただし、

$$\sum_{n,m} \rho_{knm} = 1 \quad (4)$$

とする。各パラメータの物理的意味は、表 1 に示されているとおりである。これらのパラメータをまとめて

表-1 HTC モデルパラメータ Θ_F

表記	意味
k	楽譜における音符インデックス
n	調波成分の次数
m	パワーエンベロープモデルにおけるガウス分布関数のインデックス
σ_k	音符 k の調波構造モデル中のガウス分布関数の周波数方向の広がり
$\mu_k(u)$	音符 k , 時間 u における基本対数周波数
w_k	音符 k のエネルギー
$U_{kn}(u)$	音符 k の第 n 高調波成分のパワーエンベロープ関数
ρ_{knm}	音符 k における楽音スペクトログラムの形状パラメータ
τ_k	音符 k の発音時刻
ϕ_{kn}	音符 k , 第 n 高調波成分のパワーエンベロープ関数の時間長

集合 Θ_F で表すこととする .

演奏誤りモデルについても , 楽譜対応モデルと同様のモデルを用い , $E(x, u)$ と表す . 演奏誤りモデルのパラメータをまとめて集合 Θ_E で表す .

楽譜対応モデルと演奏誤りモデルの和 $G(x, u) = F(x, u) + E(x, u)$ は , テンポが一定と仮定した演奏のスペクトログラムモデルであるが , これに時間伸縮を施したもの $G(x, \eta_t)$ により実演奏のスペクトログラムをモデル化する . ここで , $u = \eta_t$ は実際の時刻 t を仮想時刻 u に対応付ける関数であり , 演奏のテンポ変動に該当する .

3.3 最適化問題としての定式化

今 , 解きたい問題は , $G(x, \eta_t)$ が $Y(x, t)$ と近くなるように η , Θ_F , Θ_E を求める最適化問題である . $G(x, \eta_t)$ と $Y(x, t)$ との間の乖離度は , HTC と同様 , I ダイバージェンス

$$\begin{aligned} \mathcal{I}(\eta, \Theta_F, \Theta_E) &= \sum_{x,t} \mathcal{D}(Y(x, t), G(x, \eta_t)) \quad (5) \\ &= \sum_{u,t} \mathbf{1}(\eta_t = u) \sum_x \mathcal{D}(Y(x, t), G(x, u)) \end{aligned}$$

により測ることとする . ただし ,

$$\mathcal{D}(a, b) = a \log \frac{a}{b} - a + b \quad (6)$$

である . $\mathbf{1}(\eta_t = u)$ は括弧内の条件式を満たす場合に 1 を , 満たさない場合に 0 を返す指示関数

$$\mathbf{1}(\eta_t = u) = \begin{cases} 1 & (\eta_t = u) \\ 0 & (\eta_t \neq u) \end{cases} \quad (7)$$

である . また , $u = \eta_t$ は , 実時刻と正規化時刻のペア $p_\ell = (t_\ell, u_\ell)$, ($\ell = 1, \dots, L$) の経路と見なすことができる . 例えば , テンポが一定の場合 , 経路は傾きが 1 の直線 , すなわち $u = t$ となる . テンポは緩やかに変化すると仮定し , 経路に図 2 に示すように局所的な傾斜制限

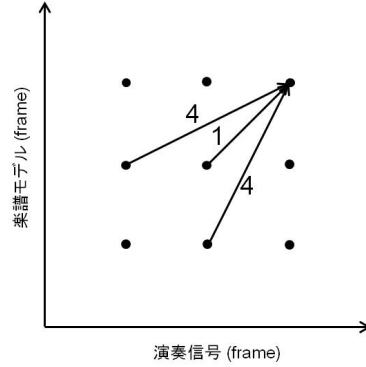


図-2 局所経路と経路コスト : ここでは楽譜を飛ばして演奏する場合は扱わず , 経路が設計される

$$\alpha(p_\ell, p_{\ell-1}) = \begin{cases} 4 & (p_\ell - p_{\ell-1} = (2, 1)) \\ 1 & (p_\ell - p_{\ell-1} = (1, 1)) \\ 4 & (p_\ell - p_{\ell-1} = (1, 2)) \\ \infty & (\text{otherwise}) \end{cases} \quad (8)$$

を与えることとする . 各局所コストは実験を通して決定した . 経路全体の累積コスト $\mathcal{J}_1(\eta)$ を ,

$$\mathcal{J}_1(\eta) = \beta_1 \sum_{\ell=2}^L \alpha(p_\ell, p_{\ell-1}) \quad (9)$$

と定義する . β_1 は正の定数である .

ここで , 式 (5) は ,

$$\mathcal{U}(\eta) = \{u^{\exists t} : u = \eta_t\} \quad (10)$$

$$\bar{Y}(x, u) = Y(x, \eta^{-1}[u]) \quad (11)$$

という記号を用いて ,

$$\mathcal{I}(\eta, \Theta_F, \Theta_E) = \sum_{u \in \mathcal{U}(\eta)} \sum_x \mathcal{D}(\bar{Y}(x, u), G(x, u)) \quad (12)$$

と書き直すことができる . ただし , $\eta^{-1}[u]$ は $u = \eta_t$ の逆関数を表す .

以下 , 発音時刻・音高のずれ , 音色変動に対するコスト設計を行う . まず , k 番目の音符の発音時刻 τ_k は ,

楽譜によって定まっている所定の発音時刻（つまり，楽譜（MIDI）に示されているテンポと音符に従って音響信号にした際の各音符の発音時刻） s_k からあまり大きく逸脱しないようにコストを設けたい．そこで， τ_k の s_k からの逸脱に対し，二乗誤差コスト

$$\mathcal{J}_2(\tau) = \beta_2 \sum_k (\tau_k - s_k)^2 \quad (13)$$

を設ける． β_2 は正の定数である．

続いて，基本周波数の所定ピッチからの逸脱にかかるコストについて述べる．弦楽器や木管楽器や管楽器などによる演奏では，ピブラートなどの演奏方法に応じて基本周波数が連続的に変化する．そこで，基本周波数 μ_t についても，楽譜によって決まるノート k の所定の基本周波数（つまり，楽譜に示されている音符の音高） g_k からの多少の逸脱を許し，先と同様に，二乗誤差コスト

$$\mathcal{J}_3(\mu) = \beta_3 \sum_k \sum_u (\mu_k(u) - g_k)^2 \quad (14)$$

を設ける． β_3 は正の定数である．

最後に，音色変動に対するコスト関数を設計する．楽器音は各楽器毎に類似したスペクトル包絡を持つ．推定精度向上のため，事前情報として各楽器のスペクトル形状を利用できる．音色変動は，オブジェクト k の音色を表すパラメータ ρ_{knm} が，実際に使用されている楽器音の代表的な形状 ξ_{nm} ($\sum_{n,m} \xi_{nm} = 1$) と近い形状を持つとし， ρ_{knm} と ξ_{nm} の Kullback-Leibler (KL) ダイバージェンス

$$\mathcal{J}_4(\rho) = \beta_4 \sum_k \sum_{n,m} \xi_{nm} \log \frac{\xi_{nm}}{\rho_{knm}} \quad (15)$$

をコスト関数とする． β_4 は任意の正の定数である．また，演奏誤りモデルにおける各楽音モデルの音色変動に対するコスト関数も楽譜対応モデルと同様に，オブジェクト l の音色を表すパラメータ ρ_{lnm}^E と実際に使用されている楽器音の代表的な形状 ξ_{nm} との間の KL ダイバージェンス

$$\mathcal{J}_5(\rho^E) = \beta_5 \sum_l \sum_{n,m} \xi_{nm} \log \frac{\xi_{nm}}{\rho_{lnm}^E} \quad (16)$$

とする． β_5 は正の重み係数である．

3.3.1 階層スパース制約

アプリケーションを使用するユーザは，楽譜通りに演奏しようとするが，時折間違えると想定される為，演奏信号中は楽譜通りの信号が誤り演奏信号よりも多く含まれている可能性が高いと考えられる．そこで，基本は楽譜モデルを使って演奏信号を表現し，出来ない場合は誤り演奏モデルを用いるという方針が考えられる．また，1つの音を表す際に，可能な限り1つのモデルで表したい．そこで，楽譜モデルで信号を表現しようとし，どうしても表現できない場合は可能な限り

少ない誤り演奏モデルを用いて表現しようとするよう，階層的なスパース制約

$$\mathcal{J}_6 = \beta_6 \sum_k |w_k|^{\gamma_1} \quad (17)$$

$$\mathcal{J}_7 = \beta_7 \sum_l |w_l^E|^{\gamma_2} \quad (18)$$

を設ける． w_k ， w_l^E は楽譜対応モデルおよび演奏誤りモデルにおける各オブジェクトのエネルギーを表すパラメータである． $\beta_6, \beta_7, \gamma_1, \gamma_2$ は正の定数である．また， $\beta_6 < \beta_7$ かつ $\gamma_1 > \gamma_2$ とする．

4. 最適化アルゴリズム

以上より，演奏誤りを含む音響信号のオブジェクトアラインメントの問題は，

$$\begin{aligned} \mathcal{H}(\Theta_F, \Theta_E, \eta) = & \mathcal{I}(\Theta_F, \Theta_E, \eta) + \mathcal{J}_1(\eta) \quad (19) \\ & + \mathcal{J}_2(\tau) + \mathcal{J}_3(\mu) + \mathcal{J}_4(\rho) \\ & + \mathcal{J}_5(\rho^E) + \mathcal{J}_6(w) + \mathcal{J}_7(w^E) \end{aligned}$$

を目的関数とした最適化問題に帰着する．この目的関数を最小化する Θ_F ， Θ_E ， η を解析的に求めることはできないが，局所最適解を効率的に探索する反復アルゴリズムを動的計画法と補助関数法に基づき導くことができる．

まず，式 (19) について， Θ_F ， Θ_E を固定し， η に関して $\mathcal{H}(\Theta_F, \Theta_E, \eta)$ の最小化問題を考えよう．この問題は，スペクトログラムモデル $G(x, \eta_t)$ と観測スペクトログラム $Y(x, t)$ との間の Dynamic Programming (DP) マッチングに帰着し，効率的に解くことができる．

次に，式 (19) について， Θ_E ， η を固定し， Θ_F についてダイバージェンス $\mathcal{H}(\Theta_F, \Theta_E, \eta)$ の最小化を行う．式 (19) から Θ_F に関する項を取り出すと，

$$\begin{aligned} \mathcal{H}_2(\Theta_F, \eta) = & \sum_{u \in \mathcal{U}(\eta), t} \sum_x \left(\bar{Y}(x, u) \log \frac{\bar{Y}(x, u)}{F(x, u)} \right. \\ & \left. - \bar{Y}(x, u) + F(x, u) \right) + \mathcal{J}_1(\eta) \\ & + \mathcal{J}_2(\tau) + \mathcal{J}_3(\mu) + \mathcal{J}_6(w) \quad (20) \end{aligned}$$

となる． $\mathcal{H}_2(\Theta_F, \eta)$ をパラメータ群 Θ_F について最小化を行うことで， $\mathcal{H}(\Theta_F, \Theta_E, \eta)$ は単調減少する．

式 (20) は u が不連続であるために，解析的にパラメータ群 Θ_F を求めることができない．そこで，補助関数法を用い，補助関数を設計する． $\mathcal{U}(\eta)$ の補集合である， $\mathcal{U}^c(\eta)$ に対し， $u \in \mathcal{U}^c(\eta)$ において定義される補助変数 $\bar{F}(x, u)$ を用いて，補助関数 $\mathcal{G}(\Theta_F, \eta, \bar{F})$ を設計すると，

$$\begin{aligned} \mathcal{G}(\Theta_F, \eta, \bar{F}) = & \mathcal{H}_2(\Theta_F, \eta) \\ & + \sum_{u \in \mathcal{U}^c(\eta)} \sum_x \left(\bar{F}(x, u) \log \frac{\bar{F}(x, u)}{F(x, u)} \right. \\ & \left. - \bar{F}(x, u) + F(x, u) \right) \quad (21) \end{aligned}$$

$$= \sum_{u,x} \left(\hat{Y}(x,u) \log \frac{\hat{Y}(x,u)}{\bar{F}(x,u)} - \hat{Y}(x,u) + F(x,u) \right) + \mathcal{J}_1(\eta) + \mathcal{J}_2(\tau) + \mathcal{J}_3(\mu) + \mathcal{J}_6(w) \quad (22)$$

となる．ここで，

$$\hat{Y}(x,u) = \begin{cases} Y(x,u), u \in \mathcal{U}(\eta) \\ \bar{F}(x,u), u \in \mathcal{U}^c(\eta) \end{cases} \quad (23)$$

である．

式 (22) は，まだ解析的にパラメータ群 Θ_F を求めることは困難である．式 (22) の右辺の \log の項 $\log \frac{1}{F(x,u)}$ に，HTC モデルの式 (1) を用いて展開すると，

$$\log \frac{1}{F(x,u)} = \log \frac{1}{\sum_k r_k(x,u)} \quad (24)$$

となり，対数の中に総和が入っているため，解析的に求めることが難しい．そこで，解析的にパラメータ群 Θ_F を求めるために，この項について補助関数を設計する．

$\log x$ は凸関数である為，以下の Jensen の不等式が成り立つ．

$$\log \sum_k r_k(x,u) \geq \sum_{k,n,m} \lambda_{knm}(x,u) \log \frac{r_k(x,u)}{\lambda_{knm}(x,u)} \quad (25)$$

ただし， $\lambda_{knm}(x,u)$ は， $\forall k, n, m, x, u : \lambda_{knm}(x,u) > 0, \forall x, u : \sum_{k,n,m} \lambda_{knm}(x,u) = 1$ を満たす補助変数であり，

$$\lambda_{knm}(x,u) = \frac{r_k(x,u)}{\sum_{nm} r_k(x,u)} \quad (26)$$

と表される．式 (22) の第 1 項の $\log \frac{1}{F(x,u)}$ を，この不等式の右辺で置き換えると，

$$\begin{aligned} \mathcal{G}(\Theta_F, \eta, \bar{F}) &\leq \sum_{u,x} -\hat{Y}(x,u) \sum_{k,n,m} \lambda_{knm}(x,u) \\ &\quad \left(\log \frac{w_k \rho_{knm}}{\sigma_k \phi_{kn}} - \frac{(x - \mu_k(u) - \log n)^2}{2\sigma_k^2} \right. \\ &\quad \left. - \frac{(u - \tau_k - m\phi_{kn})^2}{2\phi_{kn}^2} \right) + \sum_k w_k \\ &\quad + \mathcal{J}_1(\eta) + \mathcal{J}_2(\tau) + \mathcal{J}_3(\mu) + \mathcal{J}_6(w) \\ &\quad + C \end{aligned} \quad (27)$$

となる．ただし， C はパラメータ群 Θ_F によらない定数である．上式右辺を補助関数 $\mathcal{G}'(\Theta_F, \eta, \bar{F})$ とする．

補助関数法の原理により， $\mathcal{G}'(\Theta_F, \eta, \bar{F})$ を補助変数及びパラメータ群 Θ_F に対し交互に最小化することで，元の式 (20) の局所最適解を得ることができる．よって， $\mathcal{G}'(\Theta_F, \eta, \bar{F})$ に関して各パラメータの偏微分が 0 になるという方程式を解くことで，パラメータの更新式が導きだされる．ただし， ρ_{knm} の更新式を求める際は，式 (4) の制約の下で解く必要があり，ラグランジェの未

表-2 動的時間伸縮調波時間構造クラスタリングアルゴリズム

STEP 0.	Θ , Θ_0 に適当な初期値を与える
STEP 1.	Θ , Θ_0 を固定し，非線形時間伸縮関数 η を更新
STEP 2.	非線形時間伸縮関数 η , Θ_0 を固定し， Θ を更新
STEP 3.	非線形時間伸縮関数 η , Θ を固定し， Θ_0 を更新
STEP 4.	目的関数が収束した場合，終了． 収束しなかった場合，STEP 1.-Step 4. を繰り返す

表-3 実験条件

サンプリング周波数	16kHz
周波数分解能	10cent
時間分解能	8ms
定 Q フィルタバンクの最低中心周波数	55Hz

表-4 実験 1:目的関数の各種重み係数の値

重み係数	数値
β_1	10
β_2	10
β_3	10
β_4	10

定乗数項

$$\sum_k \gamma_k \left(\sum_{n,m} \rho_{knm} - 1 \right) \quad (28)$$

を考慮する． γ_k は未定乗数項である．また，誤り演奏モデルパラメータ Θ_E に関しても同様に求められる．紙面の都合上，各種更新式については割愛する．

アルゴリズムを表 2 に示す．本来ならば，何かしらの基準を定めて収束判定を行うのが良いが，本論文では一定回数の更新を行うことでこのアルゴリズムを用いる．

以上のオブジェクトアラインメント手法を動的時間伸縮調波時間構造化クラスタリング (Time-Warping Harmonic-Temporal Clustering; TW-HTC) と呼ぶ．

5. 実験 1:実演奏に対する整合精度の確認

提案手法がどの程度の整合精度を持つか検証するために，実演奏を用いた実験を行った．入力信号として，RWC データベースを用いた．既知の楽譜として与えられる MIDI に含まれる各ノートと，音楽音響信号にウェーブレット変換を行ったものとアライメントを行う．RWC データベースで提供される発音時刻データは人手で与えられているため，小さな誤差を含む場合があるが，正解データと扱う．更新回数は 20 回，各種パラメータの初期値は，表 4 の通りで，一律 10 とする．結果を表 5 に示す．

この実験結果から，BPM120 程度の曲の場合，32 分音符単位でオンセット位置を 95% 以上の精度で推定できることが示された．

表-5 発音時刻推定の正解率 (全発音数 52)

正解と見なす誤差閾値 (ms)	提案法
128(BPM 約 120 で 16 分音符)	98.1%
64(BPM 約 120 で 32 分音符)	96.2%
32(BPM 約 120 で 64 分音符)	63.5%

表-6 実験 2:目的関数の各種重み係数の値

重み係数	数値
β_1	60
β_2	10
β_3	10
β_4	10
β_5	10
β_6	1
β_7	10
γ_1	0.5
γ_2	0.001

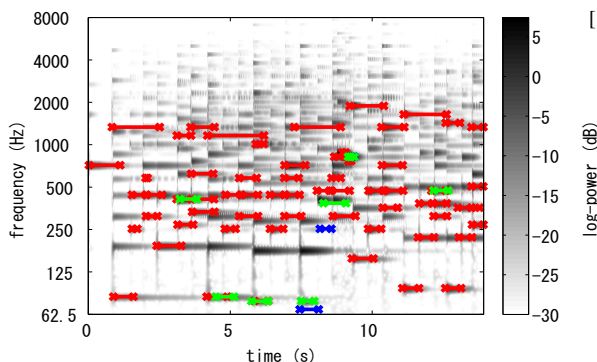


図-3 誤りを含む演奏信号の推定結果: 赤線が楽譜モデル推定, 青線が脱落誤り推定箇所, 緑線が挿入誤り推定箇所

6. 実験 2:誤り演奏を含む音楽音響信号に対する整合

本手法の有効性を示すため, RWC データベースの No.26,M05,Tr.05 を実際に演奏し, それをライン入力したのに対して本手法を用いて実験を行った. 実験条件は表 3, 各種パラメータは表 6 に示す.

実験結果を図 3 に示す. 誤り演奏箇所の提示を行えたことが観察できた. また正しく演奏された箇所についても推定可能であることを示した.

7. まとめと今後の展望

本研究では, 簡易に音楽音響信号の加工を行えるよう, 誤り演奏を含む音楽音響信号と楽譜との整合を提案し, 誤り演奏箇所と正しく演奏された箇所の同時推定問題を解いた. また実演奏を用いて実験を行った結果, 誤り演奏位置の検出ができ, 本手法の有効性が示された. 今後は, 多様な環境で本手法を適用可能にするために, 雑音や残響を考慮したモデル化や, GUI を

含めたアプリケーションの作成を予定している. また, 大規模データを用いて最適な目的関数のパラメータ設定を行う.

文 献

- [1] Nicola, O., *et al.*, "Score following: state of the art and new developments," in Proc. NIME2003, pp. 36–41, 2003.
- [2] Diemo, S., *et al.*, "Robust polyphonic midi score following with hidden markov models," in Proc. ICMC2004, pp. 1–4, 2004.
- [3] Raphael, C., "A hybrid graphical model for aligning polyphonic audio with musical scores," in Proc. ISMIR2004, pp. 387–394, 2004.
- [4] Peeling, P., *et al.*, "A probabilistic framework for matching music representations," in Proc. ISMIR2007, 2007, pp. 267–272.
- [5] 松本他, "楽譜からの音楽音響信号生成モデルに基づく楽譜と音響信号の詳細な整合," 音講論 (秋), pp. 847–848, 2007.
- [6] Maezawa, A., *et al.*, "Polyphonic audio-to-score alignment based on bayesian latent harmonic allocation Hidden markov model," in Proc. ICASSP2011, pp. 185–188, 2011.
- [7] Diemo, S., *et al.*, "Robust Polyphonic Midi Score Following with Hidden Markov Models," in Proc. ICMC2004, pp. 1–4, 2004.
- [8] Kameoka, H., *et al.*, "A Multipitch Analyzer Based on Harmonic Temporal Structured Clustering," IEEE Transactions on Audio, Speech, and Language Processing, pp. 982–994, 2007.