RE-001

奏者の意図したテンポ変動の推定に基づく演奏録音の自動伸縮修正法*

小泉 悠馬 (法政大学大学院 情報科学研究科), 伊藤 克亘 (法政大学 情報科学部)

1 まえがき

情報処理技術の発展に伴い，アマチュアの楽器奏者が動画共有サイトなどを通して自身の演奏をインターネットを通して公開する user-generated content（UGC）が増加している．しかし，演奏の熟練度が低い奏者は楽器を意図通りに制御できず，テンポ，音量，音高，音色に，“奏者の意図しない逸脱” が含まれてしまう．よって多くの場合，奏者は演奏の公開前に，楽器の制御ミスによる逸脱の除去を，自身の手で行う必要がある．

音楽演奏の音響信号の修正は，リズムであればメトロノーム通りに，音高であれば平均律で定義される音高通りに，というように “機械的” には行われない．なぜなら，音楽の演奏にはアゴーギグやビブラートなどの，機械的な演奏[1]からの “意図した逸脱” が含まれるためである．これら音楽的な逸脱は，演奏の表現力や音楽性，また自然性に関係する．よって音楽信号の修正では，奏者が意図した楽譜からの逸脱を解析・理解し，それを反映させて修正する必要があり，音楽とコンピュータ操作の専門知識や技術，労力を要する．本稿では，音楽表現の知覚の中でも特に重要なテンポの変動 [1] から，奏者の意図しない逸脱を除去する楽音修正法について考える．

演奏録音からのテンポ変動の解析では，演奏表現と深くかかわるテンポの変動はテンポ曲線（tempo-curve）と呼ばれる．しかし従来のテンポ曲線解析 [2, 3] は，意図しない逸脱を含まない熟練した奏者の演奏からの音楽表現の解析を目的とし，意図しない逸脱を含んだ演奏からの演奏表現解析の研究 [4] は少ない．

本稿では，意図しない逸脱を含んだ独奏音から，奏者の意図した “真のテンポ曲線” を推定する手法を提案する．また，真のテンポ曲線を用いて，演奏録音のテンポ変動を自動修正する手法を提案する．さらに，様々な奏法の演奏からテンポ変動を高精度に解析するために，人間の聴覚特性を考慮した発音時刻検出法を提案する．ただし，本稿では意図的なテンポ変動は滑らかに変化すると仮定するため，スィングやウィンナワルツのような，滑らかに変動しない意図的なテンポ変動は扱わない．

2 奏法誤差成分によるテンポ変動と修正

実際の演奏では，テンポは一定ではない．熟練した奏者[2]は意図表現に基づき，フレーズ中にテンポを滑らかに変動させる（図 1 左）．これは，多くの先行研究のテンポ曲線である．一方，熟練度の低い奏者[3]のテンポは滑らかに変化せず，ばらつく（図 1 右）．本稿では，作曲家がテンポ変動を指定しないフレーズでは，アマチュア奏者も滑らかなテンポ変動を意図して演奏するが，楽器の制御ミスによりテンポがばらつくと仮定する．この意図しないテンポ変動を “奏法誤差成分” と定義する．

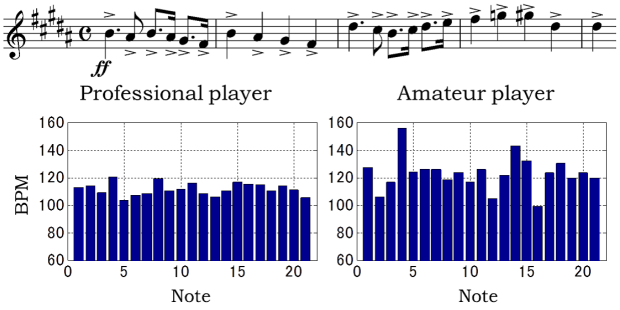

図 1: プロ奏者とアマチュア奏者のテンポ変動例

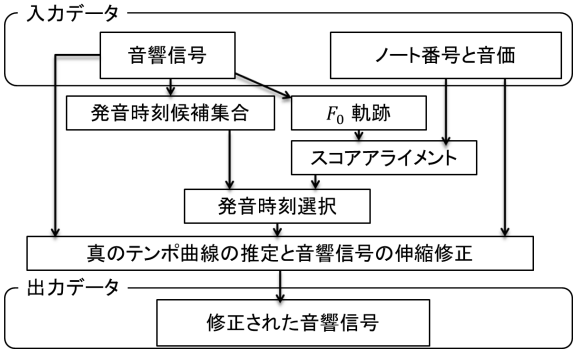

図 2: 提案法の処理の流れ

図 2 に提案法の概要を示す．まず，奏法誤差を含む音響信号と，楽譜情報としてノート番号[4]と音価[5]を入力する．次に，音響信号から発音時刻候補集合と基本周波数（F_0）を求める．その後，F_0 と楽譜情報のアライメントを行い発音時刻の初期値を求め，発音時刻候補集合から，初期値に近い時刻を発音時刻として選択する．次に，真のテンポ曲線の逆数を多項式でモデル化し，発音時刻から多項式回帰で推定する．多項式の次数は赤池情報量基準 [5] の最小化で決定する．最後に，推定した真のテンポ曲線に基づき音響信号を伸縮修正する．

---

* Title: An Automatic Musical Signal Adjustment Method Based on Estimation of Intended Tempo Fluctuation *Yuma Koizumi (Hosei Univ.) et al.*

[1]一定のテンポや，平均律で定義される絶対音高で演奏したもの

[2]以降，本来の定義とは異なるが，“プロ奏者” と呼ぶ．

[3]以降，本来の定義とは異なるが，“アマチュア奏者” と呼ぶ．

[4]“ノート番号” は各音高について割り振られた値である．本稿では “Middle C”（261.6 Hz）を 60，A3（440 Hz）を 69 とする．

[5]“音価” は楽譜上の音符の長さである．本稿では，4 分音符を 1，2 分音符を 2，8 分音符を 0.5 のように定義する．

Copyright © 2013 by Information Processing Society of Japan and
The Institute of Electronics, Information and Communication Engineers
All rights reserved.

3 発音時刻検出

本章では，奏法誤差を含む N 個の音符の発音時刻 $\boldsymbol{y} = (y[1], ..., y[N])^T$ を求める．発音時刻検出の音響特徴量には，位相の変化 [6]，複素スペクトルのユークリッド距離 [7]，スペクトルフラックス [8] など様々な特徴量が提案されている．これらの音響特徴量は楽器や奏法の種類によって有効なものが異なり [9, 10]，特に *legato* 奏法の演奏音からの発音時刻検出は難しい問題である．

一方ビートトラッキングでは，動的時間伸縮法（DTW）を用いて楽譜と観測 F_0 のアライメントを取り発音時刻を求める手法 [11, 12] も存在するが，発音時刻付近の F_0 は非調波成分の影響で正確に求まらないことが多く，正確な発音時刻の推定は難しい．

本稿では，様々な奏法の演奏の修正を行うために，多様な奏法で正確な発音時刻が求まる手法が必要である．また，音符の伸縮修正のために，発音時刻と楽譜情報のアライメントが取られている必要がある．

これらの要件を満たすために，人間の聴覚特性を考慮した，複素メルスペクトルに基づく音響特徴量を提案する．さらに，正確な発音時刻を求めるために，F_0 と楽譜情報の DTW 法を援用し，発音時刻を検出する．

3.1 発音時刻候補集合の生成

聴衆は音符の発音時刻を，音量や音高などの様々な聴覚的要素の違いを手掛かりに知覚する．そこで発音時刻検出の音響特徴量に，聴覚特性を考慮した複素メルスペクトルの KL 情報量（CMKLD）を提案する．CMKLD は，時刻 k で観測された複素メルスペクトル $\boldsymbol{S}_{\mu,k}$ と，微小時間 τms 前から予測される時刻 k の複素メルスペクトル $\hat{\boldsymbol{S}}_{\mu,k}$ の KL 情報量である．ここで μ はメル対数周波数軸を均等に分割した際の周波数ビンである．これは，従来法 [13] を複素メル周波数領域に拡張した，聴覚的な “驚き” をモデル化した尺度である．

以降では，$\boldsymbol{X}_{\omega,k}$ と $\phi_{\omega,k}$ をそれぞれ，音響信号を短時間フーリエ変換（STFT）して得られる振幅と位相スペクトルとする．ここで，ω は線形周波数ビン，k は離散時刻を表す．また，mel[·] は，線形周波数領域のスペクトルをメル周波数軸上で均等になるように各周波数ビンをリサンプリングして，メル対数周波数領域に変換する処理である．

各時刻 k での複素メルスペクトルを求め，CMKLD を計算する．まず，観測振幅スペクトル $\boldsymbol{X}_{\omega,k}$ をメル周波数領域に変換し，時刻 k の振幅メルスペクトル $|\boldsymbol{S}_{\mu,k}|$ を求める．

$$\boldsymbol{X}^{\mathrm{mel}}_{M,k} = \mathrm{mel}\left[\boldsymbol{X}_{\Omega,k}\right] \tag{1}$$

$$|\boldsymbol{S}_{\mu,k}| = \frac{\boldsymbol{X}^{\mathrm{mel}}_{\mu,k} + C}{\sum_{\mu} \boldsymbol{X}^{\mathrm{mel}}_{\mu,k} + C} \tag{2}$$

ここで C は STFT による白色雑音の振幅スペクトルの不確定性を抑える正の定数である．次に，$|\boldsymbol{S}_{\mu,k}|$ と対応する位相スペクトルを求める．まず連続する周波数ビンの位相のジャンプ量が π 以上のとき，その値を 2π の補数に変更し，位相スペクトルを単調増加に変換する処理 unwrap[·] を用いて，$\phi_{\omega,k}$ と極座標系で等価な位相スペクトル

$$\psi_{\omega,k} = \mathrm{unwrap}[\phi_{\omega,k}] \tag{3}$$

を求める．そして，先行研究 [6] の手法を用いて，予測位相スペクトル $\hat{\psi}_{\omega,k}$ を求める．

$$\hat{\psi}_{\omega,k} = (2\psi_{\omega,k-2\tau} - \psi_{\omega,k-\tau}) \tag{4}$$

最後に，各位相スペクトルをメル周波数領域に変換する．

$$\phi^{\mathrm{mel}}_{M,k} = \mathrm{princarg}\left[\mathrm{mel}[\psi_{\Omega,k}]\right] \tag{5}$$

$$\hat{\phi}^{\mathrm{mel}}_{M,k} = \mathrm{princarg}\left[\mathrm{mel}[\hat{\psi}_{\Omega,k}]\right] \tag{6}$$

ここで princarg[·] は，絶対値が π 以上の位相スペクトルを，2π の補数を用いて範囲 $[-\pi, \pi]$ に変換する関数である [6]．すると，各複素メルスペクトルは

$$\boldsymbol{S}_{\mu,k} = |\boldsymbol{S}_{\mu,k}| \exp(j\phi^{\mathrm{mel}}_{\mu,k}) \tag{7}$$

$$\hat{\boldsymbol{S}}_{\mu,k} = |\boldsymbol{S}_{\mu,k-\tau}| \exp(j\hat{\phi}^{\mathrm{mel}}_{\mu,k}) \tag{8}$$

と求められ，CMKLD は以下のようになる．

$$\mathrm{CMKLD}[k] = \sum_{\mu} \left| \boldsymbol{S}_{\mu,k} \log \frac{\boldsymbol{S}_{\mu,k}}{\hat{\boldsymbol{S}}_{\mu,k}} \right| \tag{9}$$

$$= \sum_{\mu} |\boldsymbol{S}_{\mu,k}| \sqrt{\left(\log \frac{|\boldsymbol{S}_{\mu,k}|}{|\hat{\boldsymbol{S}}_{\mu,k}|} \right)^2 + (\phi^{\mathrm{mel}}_{\mu,k} - \hat{\phi}^{\mathrm{mel}}_{\mu,k})^2} \tag{10}$$

しかし，式 (10) の平方根内の第二項 $(\phi^{\mathrm{mel}}_{\mu,k} - \hat{\phi}^{\mathrm{mel}}_{\mu,k})$ は，位相の周期性により，原点の選択に強く依存する．すなわち，$\phi^{\mathrm{mel}}_{\mu,k} = \pi - \epsilon$，$\hat{\phi}^{\mathrm{mel}}_{\mu,k} = -\pi + \epsilon$，$0 < \epsilon \ll \pi$ であるとき，極座標系での偏角の距離は 2ϵ であるが，式 (10) では $2\pi - 2\epsilon$ である．そこで実際の計算時には，$\Phi_{\mu,k} = \mathrm{princarg}[\phi^{\mathrm{mel}}_{\mu,k} - \hat{\phi}^{\mathrm{mel}}_{\mu,k}]$ とし，CMKLD を以下の近似式で求める．

$$\mathcal{D}[k] \approx \sum_{\mu} |\boldsymbol{S}_{\mu,k}| \sqrt{\left(\log \frac{|\boldsymbol{S}_{\mu,k}|}{|\hat{\boldsymbol{S}}_{\mu,k}|} \right)^2 + \Phi^2_{\mu,k}} \tag{11}$$

式 (11) より CMKLD は，観測した調波構造に対して大きな重みを与える係数 $|\boldsymbol{S}_{\mu,k}|$ を乗じて，振幅スペクトルと位相スペクトルの予測との乖離を同時に考慮する特徴量である．

次に $\mathcal{D}$ からピーク値を検出することにより，発音時刻の候補集合 $\mathcal{Y}$ を生成する．動的閾値の決定のために，$\boldsymbol{d}_k$ を時刻 k を中心とする長さ $\mathcal{T}$ の閉区間 $[k-\mathcal{T}/2, k+\mathcal{T}/2]$ で $\mathcal{D}$ を切り出したものと定義する．

$$\boldsymbol{d}_k = \left(\mathcal{D}\left[k - \frac{\mathcal{T}}{2}\right], \cdots, \mathcal{D}\left[k + \frac{\mathcal{T}}{2}\right] \right)^T \tag{12}$$

Copyright © 2013 by Information Processing Society of Japan and
The Institute of Electronics, Information and Communication Engineers
All rights reserved.

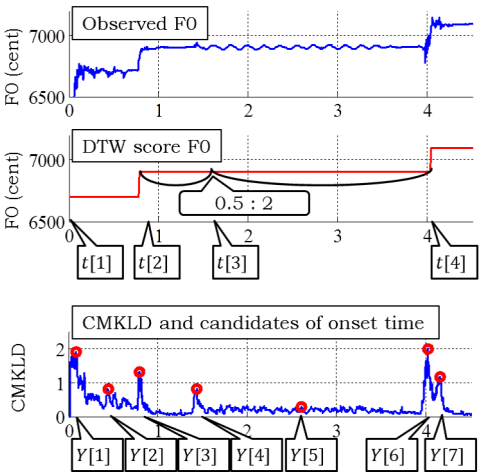

図 3: 音価 $=(0.5, 0.5, 2, 0.5, ...)$，ノート番号 $=(64, 69, 69, 71, ...)$ の発音時刻選択の例．x 軸は時間 (秒)．

そして $\mathcal{D}[k]$ から，動的閾値 $\delta[k]$ よりも大きなピーク値を選択し，その時刻を候補集合 $\mathcal{Y}$ とする．動的閾値は先行研究 [9] のものを拡張し以下のように求める．

$$\delta[k] = \lambda \mathrm{Median}(\boldsymbol{d}_k) + \frac{\mathrm{Median}(\mathcal{D})}{2} \tag{13}$$

ここで λ は，初期値 ξ から始め，$|\mathcal{Y}| \geq N$ とならなければ $\Delta\xi$ 減少させ再度ピーク検出を行う．

3.2 候補集合からの発音時刻の選択

発音時刻候補集合 $\mathcal{Y}$ から発音時刻を選択する．まず，F_0 軌跡を基本周波数推定法 YIN [14] で推定する．次に，F_0 軌跡と楽譜情報の DTW によるスコアアライメント [11] でスコア F_0 軌跡を生成する．そして，楽譜 F_0 軌跡の音高が変化する時刻を，発音時刻の初期値 $t[n]$ とする．但し，隣接する音符のノート番号が変化しない場合は，隣接するノートの音価の比率を用いて $t[n]$ を決定する．例として図 3 では，$t[3]$ はノート番号が変化しないため検出されない．そこで $t[2]$ と $t[4]$ を音価を用いて 0.5 : 2 で分割し $t[3]$ を決定する．最後に，発音時刻候補集合 $\mathcal{Y}$ の中から，以下の式で定義される CMKLD 重み付距離が最小の発音時刻候補 $Y[i]$ を選択し，$y[n]$ とする．

$$y[n] = \arg\min_{Y[i] \in \mathcal{Y}} \frac{|Y[i] - t[n]|}{\mathcal{D}[Y[i]]} \tag{14}$$

図 4 に，バイオリンの *legato* と *marcato* 奏法の独奏音からの発音検出結果を示す．従来難しいとされていた *legato* の発音時刻も，CMKLD により検出されていることが確認できる．また，*marcato* のフレーズに対しても，発音時刻を検出できていることが確認できる．

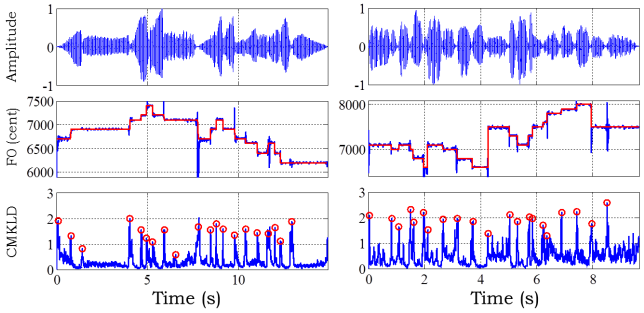

図 4: *legato*（左）と *marcato*（右）奏法の演奏からの発音時刻検出例．上図が時間波形，中央図が観測 F_0 軌跡（青線）と時間伸縮された楽譜 F_0 軌跡（赤線），下図が CMKLD（青線）と選択された発音時刻（赤丸）．

4 真のテンポ曲線の推定と音響信号の修正

本章では，奏者の意図したテンポ変動である真のテンポ曲線を，検出された発音時刻 $\boldsymbol{y}$ から推定する．さらに，真のテンポ曲線を用いて音響信号を伸縮修正する．

4.1 真のテンポ曲線の推定

音符の持続時間の定義は楽器の系統によって様々だが，本稿では一般化のために，対象とする音符の発音時刻から次の音符の発音時刻までとする．すなわち休符は考慮せず，8 分音符と 8 分休符を一つの 4 分音符として扱う．音価についても同様の定義を行う．すると，奏法誤差成分を含まない n 音目の発音時刻は，1 音目から $(n-1)$ 音目までの持続時間の和となり，音価とテンポ（beats/min）を用いて以下のように書ける．

$$\text{発音時刻}\ [n] = \sum_{m=1}^{n-1} \frac{60}{\text{テンポ}\ [m]} \times \text{音価}\ [m] \tag{15}$$

しかし実際の発音時刻には奏法誤差成分が含まれる．ここで奏法誤差成分は，真のテンポ曲線によって決まる発音時刻に対し加法的に作用すると仮定すると，n 音目の観測発音時刻 $y[n]$ は以下のように書ける．

$$y[n] = \sum_{m=1}^{n-1} \frac{60}{b[m]} h[m] + e[n] \tag{16}$$

ここで $h[m]$ は m 音目の音価，$b[m]$ は m 音目の真のテンポ曲線の値，$e[n]$ は n 音目の奏法誤差成分の値（秒）である．

さらに，テンポ変動を曲線として推定するために，武田らの曲線フィッティング [15] を参考に，真のテンポ曲線の逆数を以下の多項式で定義する．

$$b[n]^{-1} = \sum_{p=0}^{P} w_p g[n]^p, \quad g[n] = \sum_{m=1}^{n-1} h[m] \tag{17}$$

Copyright © 2013 by Information Processing Society of Japan and
The Institute of Electronics, Information and Communication Engineers
All rights reserved.

ここで P は多項式の次数である．よって，式 (16)(17) より，n 音目の持続時間 $\Delta y[n]$ は以下のようになる．

$$\Delta y[n] = y[n+1] - y[n] = \frac{60}{b[n]}h[n] + e[n+1] - e[n]$$
$$= 60\sum_{p=0}^{P} w_p g[n]^p h[n] + e[n+1] - e[n] \qquad (18)$$

ただし，音響信号中に存在しない $(N+1)$ 音目の発音時刻は $y[N+1] = L_x/f_s$ とする．ここで L_x は音響信号のデータ点数であり，f_s はサンプリングレートを表す．

ここで，$N \times (P+1)$ の説明変数行列を $\boldsymbol{G}_{n,p} = \{g[n]^{(p-1)}h[n]\}$ と置くことにより，音符の持続時間ベクトル $\Delta\boldsymbol{y} = (\Delta y[1], ..., \Delta y[N])^T$ は以下のように書ける．

$$\Delta\boldsymbol{y} = 60\boldsymbol{G}\boldsymbol{w} + \Delta\boldsymbol{e}, \qquad (19)$$

ここで $\boldsymbol{w}$ は回帰係数を並べたベクトル $\boldsymbol{w} = (w_0, ..., w_P)^T$ であり，$\Delta\boldsymbol{e}$ は奏法誤差成分のデルタベクトル $\Delta\boldsymbol{e} = (e[2]-e[1], e[3]-e[2], ..., -e[N])^T$ である．

ここで先刻研究 [4] を参考に，$e[n] \sim \mathcal{N}(0, \sigma^2)$ と仮定すると，正規分布の再生性より，$\Delta\boldsymbol{e}$ の各要素も正規分布に従う．よって，最小二乗法により回帰係数ベクトル $\boldsymbol{w}$ を求めることで，式 (17) よりテンポ曲線が求まる．多項式回帰の問題として，最適な多項式の次数 P の決定が挙げられるが，本稿では赤池情報量基準（AIC）[5] の最小化で次数 P を決定する．

$$\sigma^2_{\Delta e} = \frac{1}{N}\sum_{n=1}^{N}\left(\Delta y[n] - 60\sum_{p=0}^{P} w_p g[n]^p h[n]\right)^2 \qquad (20)$$

$$\mathrm{AIC} = N\log(2\pi\sigma^2_{\Delta e}) + N + 2(P+2) \qquad (21)$$

4.2 音響信号の伸縮修正

音響信号の修正は，“各音符の持続時間から奏法誤差による変動を除去すること” と定義できる．奏者の意図した音符の持続時間 $\hat{z}[n]$ は，真のテンポ曲線 $\boldsymbol{b}$ を用いて以下のように書ける．

$$\hat{z}[n] = \frac{60}{b[n]}h[n] \qquad (22)$$

また，観測された n 音目の持続時間は $y[n+1]-y[n]$ であるため，音響信号の修正は，n 音目の持続時間を以下の式で表される伸縮係数 $\alpha[n]$ 倍することとなる．

$$\alpha[n] = \frac{\hat{z}[n]}{y[n+1]-y[n]} \qquad (23)$$

音響信号の修正伸縮には，パワースペクトログラムの逆短時間フーリエ変換（IDFT）のシフト幅の伸縮による速度変換手法 [16] を用いる．本稿では，各音符ごとの IDFT のシフト幅を $\alpha[n]$ 倍して，音響信号を伸縮する．シフト幅の変化による位相の不整合は，Griffin らの位相再構成法 [17] で除去する．

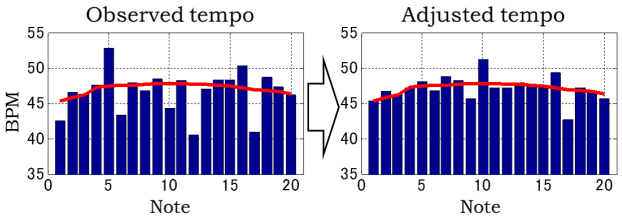

図 5: 観測音と修正音のテンポ変動例．左図のバーが観測音のテンポ変動，右図のバーが修正音のテンポ変動，両図の赤線が観測音から推定された真のテンポ曲線．

図 5 に修正結果の例を示す．左図は奏法誤差を含む観測テンポ変動を示し，右図が修正された音響信号から求めたテンポ変動を示す．提案法の修正により，テンポ変動が真のテンポ曲線に近づいていることが確認できる．修正後の一部の音符のテンポ変動が真のテンポ曲線に一致しないのは，修正前，または修正後の発音時刻検出で誤差が生じたためである．

5 評価実験

発音時刻検出と楽音修正の評価実験を行う．各パラメータは，発音時刻を精度よく検出するために，STFT 長は $0.01 \times f_s$ 点（i.e., 10 ms），シフト点数は $0.001 \times f_s$ 点（i.e. 1 ms）とした．また，スペクトル予測時間は STFT 長から $\tau = 10$ ms（i.e., $0.01 \times f_s$ 点）とした．$\boldsymbol{d}_k$ の切り出し区間の長さ $\mathcal{T}$ は，メディアンの計算のために 100 ms（i.e. $0.1 \times f_s$ 点）とした．式 (2) の定数と動的閾値決定のための変数 λ の初期値および変位は事前実験より，$C = 0.2$，$\xi = 1.1$，$\Delta\xi = 0.1$ とした．

5.1 発音時刻検出の評価実験

提案する発音時刻検出法の検出精度を，バイオリンやサキソフォンなどの連続励起振動楽器と，エレキギターやシロフォンなどの打・撥弦楽器の独奏音で評価した．本稿では楽譜と音響信号のアライメント法の先行研究 [19] と同様に，検出された発音時刻と，ラベリングされた時刻との平均絶対誤差（MAE）で評価した．

連続励起振動楽器の評価には，先行研究 [9] のデータセットからバイオリンの独奏を 1 フレーズ，先行研究 [18] のデータセットからバイオリン，サキソフォン，クラリネット，トランペットの独奏をそれぞれ 1 フレーズと，我々が収録したプロ奏者の独奏演奏 6 フレーズを用いた．我々のデータセットのフレーズは *legato*，*marcato*，*feroce*（荒々しく）などの，様々な演奏記号および奏法を含む．全ての演奏は 196kHz，24bit で録音し，48kHz にダウンサンプリングした．先行研究のデータセットと合わせた音符の総数は 413 音である．

打・撥弦楽器の評価には，先行研究 [9] のデータセッ

Copyright © 2013 by Information Processing Society of Japan and
The Institute of Electronics, Information and Communication Engineers
All rights reserved.

トからシロフォン，グロッケン，ハープシーコードの独奏をそれぞれ 1 フレーズと，我々が収録したギター経験が 6 年のアマチュア奏者のエレキギターの独奏 3 フレーズを用いた．我々のデータセットは，ライン入力で 48kHz，16bit で録音した．先行研究のデータセットと合わせた音符の総数は 172 音である．

実験の結果 MAE は，連続励起振動楽器は 12.1ms ，打・撥弦楽器は 22.4ms であった．連続励起振動楽器の MAE は，クラシックの分野で比較的速いテンポの速度記号 Allegro(BPM ≈ 120，速く，快活に．) の 4 分音符の 1/41 以下の長さである．またこの MAE が BPM=120 の 4 分音符のテンポ推定に与える影響は ±3 以下であり，十分な精度といえる．一方，打・撥弦楽器の MAE は，消音部分を発音時刻と誤検出した音符が存在したため，誤差が大きくなったが，今後有音区間判定を組み込むことにより，改善が可能である．

5.2 音響信号修正の動作実験

提案楽音修正法の動作実験と修正精度の評価をした．修正前および修正後の音響信号の観測テンポ変動と真のテンポ曲線の音価重み付平均絶対誤差（weighted-MAE）を評価した．提案法により修正音のテンポ変動が真のテンポ曲線に近づくならば，この値は減少する．

音響信号の観測テンポ変動 $\tilde{\boldsymbol{b}}$（beats/min）と weighted-MAE（beats/min）は以下の式で求める．

$$\tilde{b}[n] = \frac{60h[n]}{\Delta y[n]} \tag{24}$$

$$\text{weighted-MAE} = \frac{\sum_n h[n]\left|b[n] - \tilde{b}[n]\right|}{\sum_n h[n]} \tag{25}$$

評価には，バイオリンの独奏演奏 5 フレーズを用いた．全ての演奏は 48kHz，16bit で，IC レコーダーを用いて録音した．これらのフレーズの楽譜上の BPM は 50–150 であり，演奏時間は 10–30 秒である．

weighted-MAE は，修正前が 5.8677，修正後が 2.6834 であった．提案法により，奏法誤差による真のテンポ曲線からのずれが半分以下になっている．この結果から，提案法は真のテンポ曲線に合わせて音響信号を伸縮修正することにより，奏法誤差を減少させている．

5.3 主観評価実験

提案修正法により，音響信号が奏者の意図したテンポ変動に修正されているかを，聴取実験で評価した．対象とした楽器は，バイオリン，チェロ，エレキギター（エフェクトなし）とした．

本研究で推定する真のテンポ曲線は，奏者の意図したテンポ変動であり，正解データが存在しない．そこで本実験では，目標とするテンポ変動として，プロ奏者の演奏を用いた．楽器の演奏を 3 年以上経験しているアマチュア奏者が，プロ奏者の演奏を聴き，30 分間練習し，そのテンポ変動を模倣するように，メトロノームを用いずに演奏した．よって，正解データはプロ奏者の演奏のテンポ変動であり，修正が正しく行われているならば，修正後のテンポ変動はプロ奏者のものに近づく．

表 1: 使用楽曲

バイオリン （1st Violin）	A. Dvorak, “Symphony No. 8” 1. 1 楽章 244-250 R.Wagner, “Tannhauser” Act.II ∼Grand March∼” （1st Violin） 2. 40-44 小節目 （1st Violin） 3. 64-68 小節目
チェロ	A. Dvorak, “Symphony No. 8” 1. 1 楽章 1-6 小節目 2. 1 楽章 165-169 小節目 3. 4 楽章 26-33 小節目
ギター	1. LUNKHEAD “ENTRANCE” 5-12 小節目 2. MONKEY MAJIK “アイシテル” 52-56 小節目 3. 松本孝弘 “Thousand Dreams” 2-9 小節目

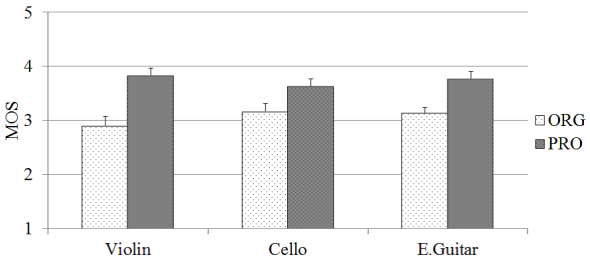

図 6: 主観評価結果

アマチュア奏者は各楽器 2 名ずつとし，楽曲は各楽器に対して 3 曲ずつとした（表 1）．これらのフレーズは，楽譜上の BPM は 60–180，平均音符数は 22 個，演奏時間は 9–16 秒である．擦弦楽器の演奏音は，IC レコーダーを用いて，防音室で録音した．ギターの演奏音は，オーディオインターフェースを用いて，ライン入力により録音した．収録条件は 48kHz，16bit とした．

聴取実験では，5 年以上の音楽経験を持つ，演奏者と別の 5 名が，実演奏音（ORG）と修正音（PRO）のテンポ変動の，プロ奏者の演奏との近さを評価した．評価には 5 段階の mean opinion score (MOS) を用いた．各評定は 1 が非常に遠い，5 が非常に近いを表す．音圧は，被験者の聴きやすいレベルとなるよう事前に調節した．

各楽器ごとの MOS と標準誤差を図 6 に示す．修正音の評定は，全ての楽器で実演奏の評定よりも上昇していることが確認できる．t-検定で有意差を検定した結果，全ての楽器の評価で，バイオリンとギターは危険率 1%で，チェロは危険率 5%で有意差のある上昇が認められた．アマチュア奏者はプロ奏者の演奏のテンポ変動を意図して演奏しており，提案法を用いた修正により，修正音がプロ奏者の演奏に有意に近づいたことから，提案

Copyright © 2013 by Information Processing Society of Japan and
The Institute of Electronics, Information and Communication Engineers
All rights reserved.

法は，奏者の意図したテンポ変動を推定し，その変動に合わせて音響信号を伸縮修正できるといえる．

6 おわりに

本稿では，奏法誤差成分を含んだ独奏音から，奏者の意図したテンポ変動である真のテンポ曲線を推定する手法を提案した．また，真のテンポ曲線を用いて，音響信号のテンポ変動を奏者の意図したものに自動修正する手法を提案した．さらに，真のテンポ曲線推定のために，人間の聴覚特性を考慮した発音時刻検出法を提案した．発音時刻検出法を評価した結果，連続励起振動楽器の独奏音の発音時刻を，標準絶対誤差が 12.1 ms で検出できることを示した．聴取実験では，修正音と目標演奏のテンポ変動の類似性が修正前と比べ向上した．従って提案法は，奏者の意図したテンポ変動を推定し，それに基づき楽音修正を行えるといえる．

本稿では，休符を明示的に扱っていないが，実際のテンポ変動は休符にも表れる．また，本稿の F_0 を用いた DTW による発音時刻検出では，トリルや過度なフェルマータを含む演奏には対応できない．今後，音符の “offset” を考慮することにより，発音時刻検出や真のテンポ曲線推定の高精度化を図る．

また一つの音符の中には，発音部，定常部，消音部と呼ばれる 3 つの状態が存在する．特に発音部の長さは，擦弦楽器の演奏表現の知覚に大きく影響を与える [20]．本稿では各音符の伸縮を，音符の全体を伸縮することで実現したが，今後は各音符の状態推定 [21] を行い，持続部のみを伸縮しなくてはならない [22]．

今後の展望として，本稿で推定した真のテンポ曲線は，奏法誤差成分を含む音響信号からの，意図表現の特徴抽出とみなせる．意図表現情報抽出技術は，奏者認識 [23]，合成音への表現力付与 [24] などに応用されている．本手法もこれらの分野への応用法を検討していく．

参考文献

[1] R. Parncutt: “A perceptual model of pulse salience and metrical accent in musical rhythms”, Music Perception, 11, pp. 409–464, 1994.

[2] E. D. Scheirer: “Tempo and beat analysis of acoustical musical signals” J. Acoust. Soc. Amer., 103, 1, pp. 588–601, 1998.

[3] D. P. W. Ellis: “Beat tracking by dynamic programming”, J. New Music Res., 36, 1, pp. 51–60, 2007.

[4] Cyril Joder and Slim Essid and Gaël Richard: “Hidden Discrete Tempo Model: A Tempo-Aware Timing Model for Audio-to-Score Alignment”, ICASSP-11, pp.397-400, 2011.

[5] H. Akaike: “Information theory and an extension of the maximum likelihood principle”, Proc. the 2nd Int. Sympo. on Information Theory, 1, pp. 267–281, 1973.

[6] J. P. Bello and M. Sandler: “Phase-based note onset detection for music signals”, Proc. ICASSP-03, pp.49–52, 2003.

[7] J. P. Bello, C. Duxbury, M. Davies and M. Sandler: “On the use of phase and energy for musical onset detection in the complex domain”, IEEE Signal Proces. Letters, 11, 6, pp. 553–556, 2004.

[8] P. Masri: “Computer modelling of sound for transformation and synthesis of musical signal”, Ph.D. dissertation, Univ. of Bristol, UK, 1996.

[9] J.P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies and M. Sandler: “A tutorial on onset detection in music signals”, IEEE Trans. Audio, Speech, & Lang. Process., vol.13, no.5, pp.1035–1047, 2005.

[10] S. Dixon: “Onset detection revisited”, Proc. DAFx-06, pp. 133–137, 2006.

[11] N. H. Adams, M. A. Bartsch, J. B. Shifrin and G. H. Wakefield: “Time series alignment for music information retrieval”, Proc. ISMIR-04, pp. 303–310, 2004.

[12] N. H. Adams, Mark A. Bartsch and Gregory H. Wakefield: “Note segmentation and quantization for music information retrieval”, IEEE Trans. Audio, Speech, & Lang. Process., 14, pp. 131-141, 2006.

[13] S. Hainsworth and M. Macleod: “Onset detection in musical audio signals”, Proc. ICMC-03, 2003.

[14] A. de Cheveigne and H. Kawahara: “Yin, a fundamental frequency estimator for speech and music”, J. Acoust. Soc. Am., 111, 4, pp. 1917–1930, 2002.

[15] H. Takeda, T. Nishimoto and S. Sagayama: “Rhythm and tempo recognition of music performance from a probabilistic approach”, Proc. ISMIR-04, pp. 357–364, 2004.

[16] 水野 優, 小野 順貴, 西本 卓也, 嵯峨山 茂樹: “パワースペクトログラムの伸縮に基づく多重音信号の再生速度と音高の実時間制御”, 聴覚研究会資料, 39, pp. 447–452, 2009.

[17] D. W. Griffin and J. S. Lim: “Signal estimation from modified short-time fourier transform”, IEEE Trans. Audio, Speech, & Lang. Process., 32, 2, pp. 236–243, 1984.

[18] P. Leveau, L. Daudet and G. Richard: “Methodology and tools for the evaluation of automatic onset detection algorithms in music”, Proc. ISMIR-04, 2004.

[19] N. Orio, D. Schwarz: “Alignment of Monophonic and Polyphonic Music to a Score” Proc. ICMC-01, 2001.

[20] K. Guettler and A. Askenfelt: “Acceptance limits for the duration of pre-helmholtz transients in bowed string attacks”, J. Acoust. Soc. Am., 101, 5, pp. 2903–2913, 1997.

[21] 小泉 悠馬, 伊藤 克亘: “連続励起振動楽器のためのパワーに基づく音符内状態推定”, 音講論 (秋)’13, 2013

[22] 安部 武宏, 糸山 克寿, 吉井 和佳, 駒谷 和範, 尾形 哲也, 奥乃 博: “音高による音色変化を考慮した楽器音の音高・音長操作手法” 情報処理学会, 音楽情報処理研究会研究報告, SIGMUS-76, 2008.

[23] R. Ramirez, E. Maestre and X. Serra: “Automatic performer identification in commercial monophonic jazz performances”, Pattern Recognition of Non-Speech Audio, 31, 12, pp. 1514–1523, 2010.

[24] T. Nakano and M. Goto: “Vocalistener: A singing-to-singing synthesis system based on iterative parameter estimation”, Proc. SMC-2009, pp. 343–348, 2009.

Copyright © 2013 by Information Processing Society of Japan and
The Institute of Electronics, Information and Communication Engineers
All rights reserved.