

演奏意図関数に基づく表現力を反映させた 音響信号の伸縮修正

小泉 悠馬^{1,a)} 伊藤 克亘²

概要: 奏者の意図表現情報を時間の関数として表現する“演奏意図関数”を、音響信号から推定する手法を提案する。また、演奏意図関数から出力される意図表現情報を用いて、演奏者の意図した演奏に修正する手法を提案する。本稿では、演奏意図関数の推定対象は発音時刻に絞る。演奏意図関数は多項式回帰モデルとしてモデル化する。聴覚特性を考慮した新しい発音検出技術により発音位置を検出し、楽譜情報と発音時刻から、リッジ回帰を用いて演奏意図関数を推定する。推定した演奏意図関数を用いて、観測演奏音のパワースペクトログラムをノートごとに伸縮修正することにより、演奏者の意図した演奏へと修正する。聴取実験において、提案法による修正により意図表現情報が明確になったかを MOS で評価した結果、全ての楽器の評価において、評定に有意差の認められる上昇が確認された。

キーワード: 音楽意図表現, 発音時刻検出, 回帰分析, 楽音修正

Acoustic Signal Adjustment by Considering Musical Expressive Intention Using a “Performance Intension Function”

YUMA KOIZUMI^{1,a)} KATUNOBU ITOU²

Abstract: We propose an estimation method for a “performance intention function”, which is a function of fluctuation of a musician’s intent shaped as a smooth curve, from the acoustic signal. In addition, we propose a method to remediate acoustic signals based on a musician’s intent by using intention information obtained by performance intention functions. In this paper, discussion of estimation of the function is focused on onset time. The function is modeled as a polynomial regression with respect to time. Onset times are detected by a new onset detection method that considers aural characteristics, and the function is estimated by ridge regression by using onset times and score information. Power spectrogram of observed musical signal is remediated by using the function with respect to each note and musical images, and intentions of the player are clarified. In addition, in terms of closeness of rhythm, MOS of remediated sounds by using performance intention function were higher than the original recorded sound, and significant differences in all instruments were observed.

Keywords: Musical expressive intentions, Onset detection, Regression analysis, Sound remediation

1. まえがき

楽器の奏者は、その曲想や個性を活かすために、楽譜情報を元に楽音をイメージし、それを奏法に変換する。よって、実際の演奏は楽譜に対して正確ではなく、意図表現によって楽譜から逸脱した動的な変動成分が含まれる (e.g.,

¹ 法政大学 大学院 情報科学研究科
Graduate School of Computer and Information Sciences, Hosei University

² 法政大学 情報科学部
Faculty of Computer and Information Sciences, Hosei University

^{a)} 12t0005@cis.k.hosei.ac.jp

ビブラート*1やアゴーギク*2)。しかしながら、演奏の習熟度が低い奏者は、自分の意図通りに楽器をコントロールすることが出来ず、音響信号には、意図表現による楽譜からの逸脱の他に、楽器の制御の誤りによる逸脱成分が含まれる。よって、このような演奏から奏者の意図表現を明確にするためには、奏法誤差を取り除く修正を行う手法が必要である。

現在、音響信号を修正・制御する手法は様々なものが検討されている[1-4]。しかしながら、これらの手法の多くは楽譜情報に基づいて修正を行っており、奏者の意図表現に基づいて奏法誤差を除去する音響信号の修正法はまだ十分に検討されていない。

奏者の意図表現に基づく音響信号の修正を行うためには、音響信号から奏者の意図表現を推定する必要がある。大石らは歌声の音高軌跡において、人間の発声機構に基づいた意図表現のモデル化を行っている[5]。また我々は擦弦楽器の演奏において、意図表現による音色の変化をモデル化する手法を提案してきた[6]。演奏の習熟度の評価手法では、奏者の意図表現をなめらかな曲線でモデル化する“演奏傾向曲線”を用いた手法がある[7]。演奏傾向曲線は習熟度の評価において、専門家の評価とほぼ同等の評価性能を示しているため、人間の意図表現のモデル化において有効な指標であると考えられる。しかしながら演奏傾向曲線に用いられる発音位置の推定などには、MIDI キーボードなどのデバイスが用いられており、音響信号からの推定は行われてこなかった。

本稿では、奏者の意図表現情報を時間の関数として表現する“演奏意図関数”を、音響信号から推定する手法を提案する。また、演奏意図関数から出力される意図表現情報を用いて、演奏者の意図した演奏に修正する手法を提案する。演奏傾向曲線は、ノートの強度や持続時間長に対しても推定されるが、本稿では演奏傾向曲線の推定対象を発音時刻に絞って議論を進める。また、演奏意図関数の推定対象は、擦弦楽器、撥弦楽器の独奏とする。

2. 楽譜からの逸脱成分のモデル化と修正

音響信号から発音時刻と、その楽譜からの逸脱量を推定し、それらを元に演奏意図関数を推定する。推定された演奏意図関数を元にノート長を修正することにより、奏法誤差の影響を除去し、奏者の意図表現を明確なものとする。

2.1 演奏意図関数

音響信号中の n 個のノートの発音時刻ベクトル $y_n = (y[1], y[2], \dots, y[n])^T$ を以下のように定義する。

$$y_n = t_n + x_n + e_n \quad (1)$$

*1 意図的な、絶対音高からの基本周波数の揺らぎ。

*2 意図的なテンポやリズムの揺らぎ。

ここで、 t_n は楽譜に記載されている発音時刻、 x_n は意図表現による発音時刻からの逸脱量、 e_n は奏法誤差による発音時刻からの逸脱量である。

しかしながら、音楽は時間軸方向に連続なものであり、奏者の意図表現も、時間変化に伴い連続的に変化するものと考えられる。よって、音符の発音時刻以外や休符時刻などの任意の時刻にも、意図表現は存在すると仮定する。すると、意図表現による発音時刻からのずれ x_n は、奏者の意図表現の時間変化の関数 $f(t)$ から、楽譜に記載されている発音時刻 t_n によってサンプリングされたものと見なせ、以下のように書くことが出来る。

$$x_n = f(t_n) \quad (2)$$

ここで我々は、意図表現の時間変化の関数 $f(t)$ を“演奏意図関数”と定義する。

2.2 音響信号の修正

ノートの持続時間の定義は楽器の系統によって様々であるが、本稿では問題の一般化のために、ノートの持続時間を、対象とするノートの発音時刻から次のノートの発音時刻までと定義する。つまり、観測音響信号の i 番目のノートの持続時間は $y[i+1] - y[i]$ として求められる。

ノート長の伸縮は、ノートの持続時間を、以下の式で定義される伸縮率 α によって伸縮することによって実現される。

$$\alpha[i] = \frac{(t[i+1] + x[i+1]) - (t[i] + x[i])}{y[i+1] - y[i]} \quad (3)$$

3. 演奏意図関数の推定

本章では演奏意図関数の推定手順について説明する。3.1 節で、観測音響信号から発音時刻ベクトル y_n および楽譜に記載されている発音時刻 t_n の推定法を説明し、3.2 節で、演奏特性関数の推定法を説明する。

3.1 発音時刻の推定

音響信号からの発音時刻検出 (Onset Detection) は多く行われている。検出のための特徴量には、位相 [8] や複素スペクトル [9] の変化などが存在する。これらの特徴量は、楽器の種類や奏法によって有効なものが異なるといわれている [10]。

本研究では、演奏意図関数を多様な種類の楽器の演奏に対して推定するために、全ての系統の楽器に対して有効な特徴量を使用する必要がある。そこで本稿では、聴覚特性を考慮したサブバンドメルスペクトルの変化量の特徴量として用いて発音時刻の候補点を検出し、さらにスコアに記載されている相対音長を援用して、発音時刻を検出する。

3.1.1 発音時刻の候補の生成

ノートの変化は、どのような意図表現であっても、聴衆が知覚できるように演奏されることが考えられる。そこで発音

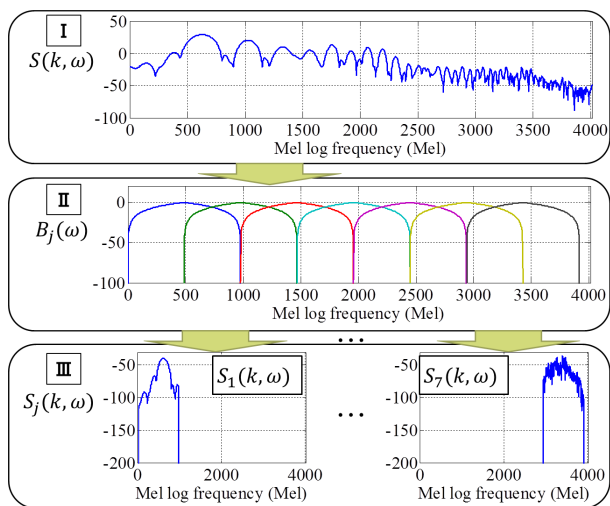


図 1 サブバンド正規化メルスペクトルの生成 (図中の各スペクトルの y 軸は対数スケール [dB])

Fig. 1 Procedure of generating sub-band mel-frequency normalized spectrum. (The y -axis of all figures are logarithmic scale [dB].)

時刻の検出には、聴覚特性を考慮した特徴量を用いる。本研究では、観測スペクトログラムをメルサブバンドで分割し、スペクトルの急激な時間変化の知覚をモデル化するために、微小時間 τ 前のフレームとの Kullback-Leibler Divergence(KLD) の各バンド総和を特徴量として用いる。

まず、音響信号から KLD を求める。音響信号をフーリエ変換して得られる振幅スペクトログラムをメル軸に変換したものを、 $S(k, \omega)$ とする (図 1-I)。ここで k は時間のインデックス、 ω はメル対数周波数である。また、本稿では、発音時刻を細かい粒度で検出するために、短時間フーリエ変換 (STFT) のシフト幅を 1ms とし、STFT に用いるデータ点数は 10ms とする。さらに、得られた $S(k, \omega)$ をサブバンド分割する。バンド数は事前実験により $J = 7$ とし、分割は窓長の半分のオーバーラップを持つ三角窓 $B_j(\omega)$ により行う (図 1-II)。KLD の計算のために、バンドごとにパワーを正規化する。 k 番目のフレームにおける j バンド目のサブバンド正規化メルスペクトルは以下のように定義する (図 1-III)。

$$S_j(k, \omega) = \frac{S(k, \omega)B_j(k, \omega)}{\sum_{\omega} S(k, \omega)B_j(\omega)} \quad (4)$$

そして、 k 番目のフレームにおけるスペクトルの変化の特徴量 $D(k)$ を、KLD として以下のように求める。

$$D(k) = \sum_{j=1}^J \sum_{\omega} S_j(k - \tau, \omega) \log \left(\frac{S_j(k - \tau, \omega)}{S_j(k, \omega)} \right) \quad (5)$$

本稿では、STFT に用いるデータ点数の長さから、 $\tau = 10$ (i.e., 10ms) とした。

次に、 $D(k)$ からピーク値を検出することにより、発音時刻候補集合 \mathcal{Y} を生成する。インデックス k を中心とし

た区間 $[k - T/2, k + T/2]$ で切り出した KLD 時系列 d_k を以下のように定義する。

$$d_k = \left(D \left(k - \frac{T}{2} \right), \dots, D \left(k + \frac{T}{2} \right) \right)^T \quad (6)$$

本稿では、後述する σ_d およびメディアン計算のための十分なデータ点を確保するために、 $T = 100$ (i.e., 100ms) とした。

ピーク検出の動的な閾値 $\delta(k)$ は、文献 [10] を参考に以下のように定め、閾値以上の KLD のピークを持つ時刻の集合を \mathcal{Y} とする。

$$\delta(k) = \lambda(\sigma_d + \text{Median}(d_k)) + \frac{\text{Median}(D)}{2} \quad (7)$$

ここで σ_d は d_k の標準偏差である。また、 λ の値は $|\mathcal{Y}| \geq n$ となるように設定する。我々の事前実験では、擦弦楽器と撥弦楽器の発音検出において、 $\lambda = 0.9$ または 1 とすることにより十分な検出性能を得た。

3.1.2 発音時刻候補からの発音時刻の決定

生成した発音時刻の集合 \mathcal{Y} から発音時刻のベクトル y_n を決定する。ここで、 y_n は楽譜に記載されている発音時刻 t_n から KLD の重み付き距離が最小であるものと仮定し、以下のように求める。

$$y[i] = \underset{i}{\operatorname{argmin}} \frac{|y[i] - t[i]|}{D(k_i)} \quad (8)$$

ここで k_i は $y[i]$ に対応する時間のインデックスを表す。

しかし、 y_n を求めるために用いる t_n は未知の情報であり、音響信号から正確に推定する必要がある。そこで、 t_n を推定するための制約条件としてスコア情報を援用する。ここで、 t_n を以下のようにモデル化する。

$$t_n = \xi h_n + \psi \mathbf{1}_n \quad (9)$$

ここで、 h_n は 4 分音符を 1 とした累積相対音符長、 ξ は一定のテンポにより決定する 4 分音符の持続時間、 ψ は音響信号の開始時刻を 0 としたときの、スコア上の第 1 音目の発音時刻、 $\mathbf{1}_n$ は要素が全て 1 の n 次元ベクトルである。

式 (9) より、推定すべき変数は ξ と ψ である。これは、音響信号から BPM を推定する問題ともとらえることが出来る。そこで、スコアによる正確な発音時刻 t_n も、真の発音時刻のベクトル y_n との誤差を最小化するものと仮定し、以下の目的関数を最小化するパラメータの探索問題と置き換える。

$$G(y_n, t_n) = \sum_{i=1}^n (y[i] - t[i])^2 \quad (10)$$

しかしながら、式 (8) と (10) を同時に最適化することは難しいため、以下の反復解法を用いて解を探索する。

ステップ 1: 初期値の設定

ξ と ψ の初期値を以下のように与える。

$$\xi = \frac{L_s}{f_s \times h[n]}, \quad \psi = 0 \quad (11)$$

ここで f_s はサンプリング周波数, L_s は音響信号のデータ点数を表す.

ステップ 2: 音響信号の発音時刻ベクトルの更新

ξ と ψ を用いて, 式 (9) により t_n を更新し, 式 (8) を用いて y_n を更新する.

ステップ 3: スコアの発音時刻ベクトルの更新

式 (10) を最小化する ξ, ψ を次式のように更新する. 更新が収束しなければ, ステップ 2 へ戻る.

$$\xi = \frac{\mathbf{y}^T \mathbf{h}_n - \psi \mathbf{1}_n^T \mathbf{h}_n}{\mathbf{h}_n^T \mathbf{h}_n} \quad (12)$$

$$\psi = \frac{\mathbf{y}^T \mathbf{1}_n - \xi \mathbf{1}_n^T \mathbf{h}_n}{n} \quad (13)$$

収束時点での y_n と t_n を, 音響信号の発音時刻ベクトルと楽譜に記載されている発音時刻ベクトルとする.

3.2 演奏意図関数の推定

演奏意図関数を曲線として推定するために, 演奏意図関数を多項式回帰モデルとして以下のように定義する.

$$f(t_n) = \begin{pmatrix} 1 & t[1] & \dots & t[1]^M \\ 1 & t[2] & \dots & t[2]^M \\ \vdots & \vdots & \ddots & \vdots \\ 1 & t[n] & \dots & t[n]^M \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_M \end{pmatrix} = \mathbf{B}w \quad (14)$$

よって, 式 (1) は以下のように書き換えられる.

$$\mathbf{y}_n = \mathbf{t}_n + \mathbf{B}w + \mathbf{e}_n \quad (15)$$

ここで, $\mathbf{z}_n = \mathbf{y}_n - \mathbf{t}_n$ とすることにより, 式 (15) は以下のように書き換えられる.

$$\mathbf{z}_n = \mathbf{B}w + \mathbf{e}_n. \quad (16)$$

よって, 演奏意図関数の推定問題は, 式 (14) の回帰係数ベクトル w の最適化問題となる. 式 (16) の回帰係数ベクトル w は, 最小二乗法により解析的に求めることが可能だが, 最小二乗法は, しばし観測に過適合を起こし, 真の演奏意図関数が求まらない. そこで, 観測への過適合を防ぎ, また曲線に十分な柔軟性を持たせるために, 本稿では回帰係数ベクトル w を, 誤差の二乗和に切片を除く回帰係数ベクトル w_1 の L_2 ノルム正則化項を付与した以下の式の最小化であるリッジ回帰 [11] で推定する.

$$R_\gamma(w) = (\mathbf{z}_n - \mathbf{B}w)^T (\mathbf{z}_n - \mathbf{B}w) + \gamma w_1^T w_1 \quad (17)$$

本稿では, 事前実験により, 多項式回帰の次元を $M = 10$, リッジ回帰の正則化係数を $\gamma = 0.1$ とする. 最適化された回帰係数 w を用いて, 演奏意図関数を構成する (図 2).

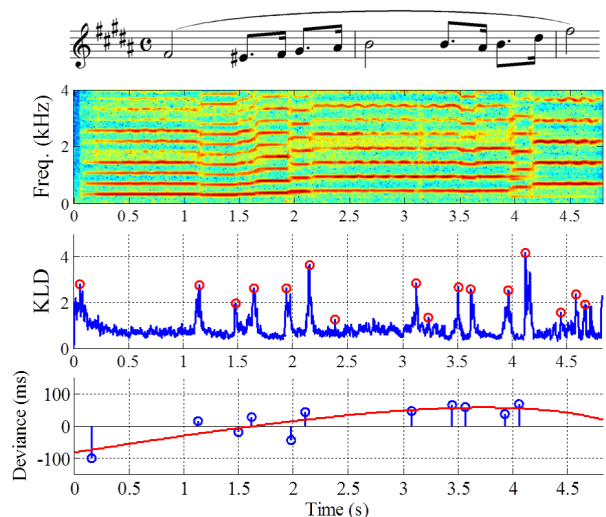


図 2 演奏意図関数の推定結果. 図は上から, 演奏に用いられた楽譜, 観測演奏音のスペクトログラム, KLD (青線) および発音時刻の候補集合 (赤丸), 逸脱量 z_n (青丸) および推定された演奏意図関数 (赤線) を示す.

Fig. 2 A estimation result of a performance intention function. Top figure shows musical score, second figure shows spectrogram of performance sound (linear frequency), third figure shows KLD (blue line) and candidate set of onset times (red circle), and bottom figure shows deviances z_n (blue circle) and an estimated performance intention function (red line).

4. 演奏意図関数を用いた音響信号の伸縮修正実験

4.1 演奏意図関数を用いた音響信号の伸縮修正

推定された演奏意図関数を用いて, 各ノートの音長を伸縮する. 伸縮率 α は式 (3) により決定する. 伸縮には, パワースペクトログラムの時間方向の伸縮による速度変換手法 [12] を用いる. フレームシフト幅を変化させることにより発生する位相の不整合は, Griffin らの位相再構成法 [13] を用いて除去する.

本稿では, 各ノートについて求められた $\alpha[i]$ を用いて, 各ノートごとのシフト幅を $\alpha[i]$ 倍することによって, 音響信号の伸縮を実現する.

4.2 評価実験

提案法の, 発音時刻の検出精度, および奏者の意図表現情報が推定精度を評価する. また提案法による修正により意図表現が明確になったかを評価する. 実験では, 楽器の演奏を 3 年以上経験している大学生が演奏したデータを用いた. 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. 比較実験において, 本手法により明確にする奏者の意図表現は未知の情報である. そこで本実験では, 目標とする意図表現として, プロの奏者が演奏した録音を用いた. 奏者は, プロの演奏を聴き,

表 1 収録条件

Table 1 Recording conditions.

演奏者	演奏経験を 3 年以上持つ大学生
収録機材	Roland, EDIROL R-09 Roland, UA-25EX
収録条件	48kHz, 16bit
収録部屋	空調を切った防音室

表 2 使用楽曲

Table 2 Used phrases in this evaluations.

バイオリン	A. Dvorak, "Symphony No. 8" 1. 1 楽章 244-250 小節目 (1st Violin) R.Wagner, "Tannhauser Act.II ~Grand March~" 2. 40-44 小節目 (1st Violin) 3. 64-68 小節目 (1st Violin)
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

30 分間の練習を行った上で、その意図表現を模倣するように、メトロノームを用いずに演奏した。奏者は、各楽器に対して 2 名ずつとし、収録する楽曲は各楽器に対して 3 曲ずつとした。バイオリンとチェロの演奏音は、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、標本化周波数 48kHz、量子化 bit 数 16bit とした。詳細な収録条件、および使用楽曲を表 1、2 に示す。

4.2.1 発音時刻の検出実験

検出された発音時刻と、人出でラベリングした発音時刻との平均二乗誤差 (RMSE) を評価した。比較は特徴量に対して行う。用いた特徴量は、提案法である KLD (KLD) と位相 (PHA) [8]、および複素スペクトル (COM) [9] である。従来法は、提案法のように候補集合から発音時刻を選択するものではないが、比較のために、発音時刻の選択は 3.1.2 節の手法で行った。各特徴量の発音時刻の候補集合生成のためのピーク検出パラメータは事前に調節した。

図 3 に、各特徴量における RMSE を示す。今回評価した全ての種類の楽器において、提案法である KLD を用いた発音検出の RMSE が最も小さいことが見て取れる。また、KLD の RMSE は、全ての種類の楽器において、先行研究 [10] の発音検出の許容誤差である 50ms 以内に収まっている。さらに、先行研究 [10] と同様に、位相を用いた検出手法は、複素スペクトルの変化を用いた検出法と比べて、擦弦楽器の発音検出精度が良いことも見て取れる。また、ギ

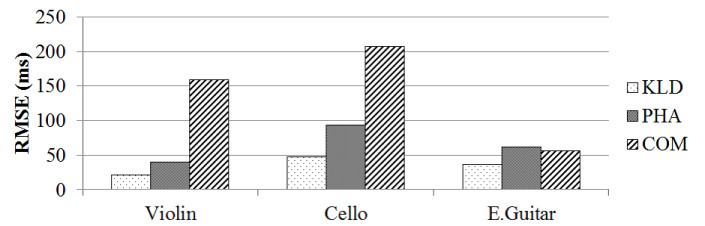


図 3 発音検出実験における、各特徴量の RMSE

Fig. 3 RMSE of onset detection.

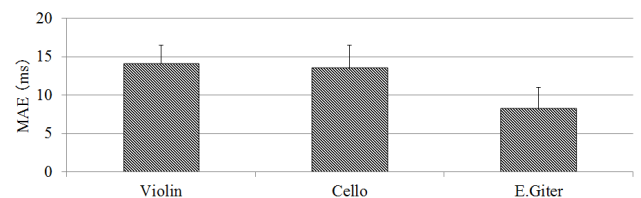


図 4 目標演奏と実演奏音の演奏意図関数の MAE

Fig. 4 MAE of values of performance intention functions.

ターなどの "Pitched-Percussive" と呼ばれる種類の楽器における発音検出精度が、位相と複素スペクトルで差が小さいという結果も、先行研究 [10] と同様の結果である。

以上の結果より、提案法は、3.1.2 節の発音検出手法を用いた擦弦楽器と撥弦楽器の発音検出において、従来法よりも高精度で発音検出を行えると考えられる。しかし、今回評価を行ったフレーズは、各楽器に対して 3 つと少ないため、今後さらに多くのフレーズに対して発音検出実験を行う必要がある。

4.2.2 演奏意図関数の推定実験

提案法により、奏者の意図表現情報が推定されているかを評価する。録音した実演奏音は、プロ奏者の演奏表現を模倣しているため、実演奏音の演奏意図関数による逸脱は、プロ奏者の演奏意図関数と近い値を取るはずである。そこで、目標演奏から推定される演奏意図関数と、実演奏音から推定される演奏意図関数の乖離を評価する。

乖離度の評価には、目標演奏と実演奏音の、各ノートの発音時刻におけるの演奏意図関数値の平均絶対誤差 (MAE) を用いる。各楽器において、逸脱量の平均絶対誤差を、以下の式のように奏者と曲において平均を取る。

$$MAE = \frac{1}{6} \sum_{s=1}^3 \sum_{q=1}^2 \frac{1}{n_s} \sum_{i=1}^{n_s} |x_s^p[i] - x_{s,q}^a[i]| \quad (18)$$

ここで s は曲番号、 q は奏者番号、 n_s は s 番目の曲の総ノート数、 $x_s^p[i]$ は目標演奏の i 番目のノートの演奏意図関数による逸脱量、 $x_{s,q}^a[i]$ は q 番目の奏者の実演奏の i 番目のノートの演奏意図関数による逸脱量を示す。

図 4 に、各楽器の MAE の平均値と標準誤差を示す。全ての楽器において、演奏意図関数の MAE は 15ms 以内に抑えられていることが確認できる。表現によるビートのゆらぎを含んだ演奏からのビートトラッキングシステムの許

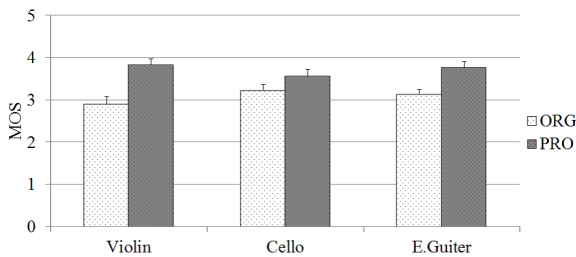


図 5 聴取実験の結果

Fig. 5 Result of the subjective evaluation.

容誤差 70ms[14] よりも誤差が小さいことから、提案法による演奏意図の推定誤差は非常に小さなものと考えられる。このことから、提案法により、奏者の意図表現情報が推定できていると考えられる。

4.2.3 聴取実験

演奏意図関数を用いた楽音修正により、意図表現が明確になったかを評価する。聴取実験の被験者は、5年以上の音楽経験を持つ5名とした。被験者は、目標演奏、実演奏音(ORG)、提案法による修正合成音(PRO)を比較した。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。評価は、実演奏音と修正合成音のリズムの、目標演奏音との近さを評価した。評価には mean opinion score (MOS) が用いられた。各評定は1が非常に遠い、5が非常に近いを表す。

MOSにより算出された、各音の平均値と標準誤差を、各楽器ごとに図5に示す。提案法による修正合成音の評定は、全ての楽器において実演奏の評定よりも上昇していることが確認できる。Dunnnettの多重比較検定[15]により有意差を検定した結果、全ての楽器の評価において、危険率5%で有意差が認められた。奏者は目標演奏を演奏意図として演奏しており、提案法を用いた修正により、修正音が目標演奏に有意に近づいたことから、提案法により、リズムにおける奏者の意図表現が明確になったといえる。

5. あとがき

本稿では、奏者の意図表現情報を時間の関数として表現する“演奏意図関数”を、音響信号から推定する手法を提案した。また、演奏意図関数から出力される意図表現情報を用いて、音響信号から奏法誤差の成分を除去する手法を提案した。実演奏音からの、奏者の意図表現情報の推定精度を評価した結果、目標演奏との演奏意図関数の平均絶対誤差は15ms以下であった。また楽音修正により意図表現情報が明確になったかをMOSで評価した結果、全ての楽器の評価において、評定に有意差の認められる上昇が確認された。このことから提案法は、リズムにおける奏者の意図表現情報が推定可能であると言える。

本稿では、演奏意図関数を発音時刻に絞って推定したが、意図表現の変化はノート強度や音高、音色の変化におい

ても発生する。今後、演奏意図関数をこれらの意図表現情報についても推定できるよう改良していく。

また、本稿では独奏のみの演奏意図関数の推定法について提案したが、実際の演奏では多重音から演奏意図関数を推定する必要もある。今後、多重音の音源分離などを利用して発音時刻を求め、修正する手法についても検討していく。

参考文献

- [1] Hirokazu Kameoka, et.al., “Complex NMF: A New Sparse Representation for Acoustic Signals,” In Proc. ICASSP 2009, pp. 3437-3440, Apr. 2009.
- [2] Katsutoshi Itoyama, et.al., “Integration and Adaptation of Harmonic and Inharmonic Models for Separating Polyphonic Musical Signals,” in Proc. ICASSP 2007, pp. 57-60, April 2007.
- [3] 安良岡 直希, 他, “フレーズ置換のための調波非調波 GMM・NMF に基づく音源分離・演奏合成,” 情報処理学会論文誌, Vol. 52, pp. 3839-3852, 2011.
- [4] 安部 武宏, 他, “音色の音高依存性を考慮した楽器音の音高操作手法,” 情報処理学会論文誌, Vol. 50, pp. 1054-1066, 2009.
- [5] Yasunori Ohishi, et.al., “A Stochastic Model of Singing Voice F0 Contours for Characterizing Expressive Dynamic Components,” In Proc. INTERSPEECH 2012, Sep. 2012.
- [6] Yuma Koizumi, et.al., “Performance expression synthesis for bowed-string instruments using “Expression Mark Functions” ” Proceedings of Meetings on Acoustics (POMA). Vol. 15, pp. 035003, Nov. 2012.
- [7] 三浦 雅展, 他, “ピアノによる1オクターブの上下行長音階演奏に対する熟達度の自動評価” 日本音響学会誌 66(5), pp. 203-212, 2010.
- [8] Bello, J.P., et.al., “Phase-based note onset detection for music signals,” In Proc. ICASSP 2003, vol.5, 6-10 Apr. 2003.
- [9] Bello, J.P., et.al., “On the use of phase and energy for musical onset detection in the complex domain,” Signal Processing Letters, IEEE, vol.11, no.6, pp. 553- 556, June 2004
- [10] Bello, J.P., et.al., “A Tutorial on Onset Detection in Music Signals,” Speech and Audio Processing, IEEE Transactions on, vol.13, no.5, pp.1035-1047, Sept. 2005
- [11] Arthur E. Hoerl, et.al., “Ridge Regression: Biased Estimation for Nonorthogonal Problems”, Technometrics, Vol.12, No.1., pp.55-67, Feb., 1970
- [12] 水野 優, 他, “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,” 日本音響学会聴覚研究会資料, vol.39, no.6, pp.447-452, Oct. 2009.
- [13] D.W.Griffin, et.al., “Signal estimation from modified short-time Fourier transform” In Proc. ICASSP 1984, vol.32, no.2, pp.236-243, Apr. 1984.
- [14] Simon Dixon, “Automatic extraction of tempo and beat from expressive performances,” Journal of New Music Research, vol. 30, pp. 39-58, 2001.
- [15] C.W.Dunnnett, “New Tables for Multiple Comparisons with a Control”, Biometrics, Vol.20, No.3, pp.482-491, 1964.