

楽譜に基づく鍵盤楽器演奏の統計的モデル化手法

奥村 健太^{1,a)} 酒向 慎司^{1,b)} 北村 正^{1,c)}

受付日 2012年7月2日, 採録日 2013年1月11日

概要: 楽器の演奏には演奏者の楽曲解釈によって発生する固有の特徴があり, それは演奏時に参照された楽譜からの芸術的意図に基づく逸脱として観測される. そのような逸脱は楽譜の情報に含まれる文脈に類するものに依存して発生していると考えられる. 本稿では音符ごとの逸脱に対して楽譜から得られる詳細な演奏指示を関連付けることで, 演奏者の楽曲解釈と楽器操作の傾向を楽譜の情報を制約に用いて分類する手法について論じる. 分類に有効な制約が楽譜と演奏の内容から一意に決定されることが本手法の利点であり, これによって演奏者以外の恣意の混入を抑止した公正な分類が可能である. 実際の演奏から取得した逸脱を用いた検証から, 楽器演奏の特徴を説明するための制約として楽譜の情報をを用いることの有効性を示した. また, 楽器演奏に関する既知の知見について, モデルによって視覚化された演奏はそれを裏付ける傾向を示した. このことから, 本手法はそのような知見の客観的な証明にも有用であることが分かった.

キーワード: 音楽分析, 音楽情報の表現, 楽譜処理, 芸術的逸脱

Stochastic Modeling of the Expressive Musical Performance of the Keyboard Instrument Using Information of the Score

KENTA OKUMURA^{1,a)} SHINJI SAKO^{1,b)} TADASHI KITAMURA^{1,c)}

Received: July 2, 2012, Accepted: January 11, 2013

Abstract: This paper presents a method for describing the characteristics of human musical performance. We consider the problem of building models that express the ways in which deviations from a strict interpretation of the score occurs in the performance, and that cluster these deviations automatically. The clustering process is performed using expressive representations unambiguously notated on the musical score, without any arbitrariness by the human observer. The result of clustering is obtained as hierarchical tree structures for each deviational factor that occurred during the operation of the instrument. This structure represents an approximation of the performer's interpretation with information notated on the score they used during the performance. Through validations of applying the method to the data measured from real performances, we show that the use of information regarding expressive representation on the musical score enables the efficient estimation of generative-model for the musical performance. In addition, this method is also useful for objective proof of the existing knowledge about the musical performance by information to support such a knowledge having been shown from our model.

Keywords: music analysis, music information expression, music score processing, artistic deviation

1. はじめに

楽器演奏は専門的な知識と高度な技巧の両立によって実

現される文化的活動である. 楽譜に従ってなされた楽器演奏(以降, 実演奏と称する)には固有の特徴があり, 同一の楽譜を用いても演奏の結果は演奏者によって異なったものになりうる. ある程度以上の技巧をとまなう演奏においては, 相対的にそのような相違を決定付けるものとして主要なものは演奏者が楽曲に対して持つ解釈であろうと考えられる. しかしながら, そのような専門的知識といえる観点や尺度は主観的かつ抽象的なものであり, 共通の表現や

¹ 名古屋工業大学
Nagoya Institute of Technology, Nagoya, Aichi 466-8555, Japan

^{a)} k09@mmssp.nitech.ac.jp

^{b)} sako@mmssp.nitech.ac.jp

^{c)} kitamura@mmssp.nitech.ac.jp

基準を用いて説明を行うことは困難であるという問題がある。実演奏に固有の特徴は、厳密に仮定した楽譜上の演奏指示からの芸術的な意図に基づく逸脱 [1] として観測される。本稿ではこの逸脱の生成過程を統計的にとらえ、その特徴的な傾向と、それに至る制約を演奏者が参照した楽譜の情報によって説明する手法を提案する。提案法では、任意の楽器演奏において楽曲や演奏者が持つ特性を具体的な指標をともなって系統的に示すことで、その特異性や相違などの検証が可能になる。本研究の最終的な目的は既存の演奏から得られる知見を楽器演奏に関わる人々に還元することであるが、本稿ではその基底となる技術的な部分を扱う。

このような仕組みを確立するためには、演奏者が楽曲をどのように解釈して、どのように演奏に反映しているかを一般性のある表現で説明する必要がある。その説明の結果は楽譜とその実演奏によって一意に決定するものであることと、実演奏を行った演奏者以外の恣意の混入を可能な限り抑止した公正なものであることが重要である。そのような要求を満たして逸脱の生成過程をとらえるには、それを統計モデルによって学習的に獲得することが有効である [2], [3], [4], [5]。しかしその傾向をただとらえるだけでは不十分であり、その傾向を制御する要因が具体的に示されるべきである。その解決には、実演奏の事例を大量に収集して、それらに共通するルールを抽出することが有効である [6], [7], [8]。しかしそれらのルールもただ並べ立てるだけでは不十分であり、それらがどのような比重でどのように関係し、実演奏の逸脱傾向に対してどのように作用しているのかが系統的に説明されるべきである。よって逸脱の傾向とその理由を同時に扱い、それらの関係を系統的に表現できる仕組みが必要である。演奏すべき内容について演奏者が得ている確かな情報は楽譜であり、そこから得られる一般的な情報のみを説明に用いることでこの問題を解決できると考える。

入力とする実演奏には、その一般的な出力形態である音響信号をそのまま使用できることが望ましい。しかしその場合は逸脱の観測以前に、任意の入力情報が楽譜上ではどの演奏指示に相当するのかを確実に一致させることが非常に困難である。実演奏の特徴の可視化を試みた先行研究 [9], [10], [11] では音響信号を入力としていたが、その結果からはその実演奏に固有の特徴について具体的な説明は得られない。これは音響信号を入力とした場合に得られる情報の種類や量が限られることが主な要因と考えられる。本稿では情報の多彩さや精度を優先し、MIDI や MusicXML で記述されたシンボリックなデータを入力とする。現在そのようなデータは主に手作業で作成されているため、多様な演奏者や楽曲の組合せを対象とすることは困難である。しかし近年は音響信号からそのようなデータの自動的な獲得を可能にする研究が進展している [12], [13]。

このような技術は現在の手作業を完全に置換しうるものではないが、将来的にはその効率化に貢献するものである。

以降、2 章では楽器演奏という行動の特性からどのようにそれをとらえるべきかを考察し、それをモデル化するための手法について論じる。3 章では本手法の有効性を複数の観点からの評価によって検証する。4 章では以上から得られた結果をまとめ、本手法の今後の展望を述べる。

2. モデルの提案

本章では、人間の演奏者による鍵盤楽器の演奏がどのようにして行われたのかを一般的に説明できるモデル化手法の提案を行う。楽器演奏という行動についての理解を通じて、それをどのようにとらえることが適切であるかを考える。また、実際にモデル学習を行うために各種の情報をどのように用いるかについてもここで述べる。

2.1 楽器演奏における演奏者の役割

楽器演奏において入力は楽譜、出力はそれが可聴化された演奏であり、この入出力の変換を行うものとして両者の間に位置するのが演奏者である。その変換は、楽譜を読み解き楽曲を解釈することと、解釈に基づき楽器を操作することの 2 段階からなると考えられる。ここで要点といえるのは、同一の楽譜を用いても演奏の結果は演奏者によって異なったものになりうることである。しかし同一の楽譜を用いて演奏を行う限り、演奏者は皆楽譜から同一に得られる情報に基づいて楽器を操作しているといえ、その結果である実演奏も同一の楽譜からの逸脱として観測が可能である。楽譜と実演奏からモデル化を行ううえで、本稿では任意の演奏指示に対する楽器の操作傾向に独自性をもたらす最大の要因は演奏者の楽曲解釈であると考え、演奏者自身の身体的制約などによる影響は無視できるものと仮定する。この場合、任意の楽器演奏の特徴を一般性をもってとらえるには、演奏者に関して以下の 2 点を説明できればよい。

- 楽譜の演奏指示をどのように解釈したのか。
- その解釈がどのような演奏表現（楽器の操作）に帰結するのか。

2.2 鍵盤楽器演奏における芸術的逸脱の発生

完結した演奏の中で何らかの一貫した個性が保たれているからには、その逸脱傾向もある程度一貫したものであると考えられる。楽器には様々な種類があるが、本稿では計算機との親和性を重視し、対象を鍵盤で操作できる楽器に限定して話を進めることにする。その場合、楽器の操作はいずれかの鍵を任意の強さで押さえ、任意の時間の後でそれを離すという一連の動作が 1 単位となる。この動作が楽譜が演奏を指示する 1 つ 1 つの音符について行われているほか、各音符は楽曲の時間的な流れを大まかに制御するテンポの変動の中で発音が指示されていると考え、逸脱

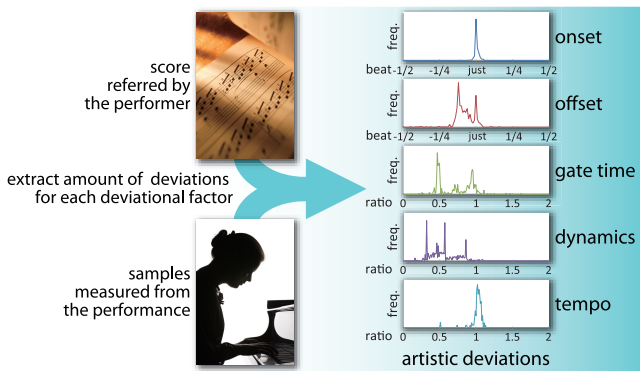


図 1 実演奏における芸術的逸脱の発生傾向

Fig. 1 Occurrence tendency of artistic deviations in the actual performance.

は打鍵 (onset), 離鍵 (offset), 押鍵長 (gate time), 打鍵強度 (dynamics), そしてテンポ (tempo) の 5 種類の要素について観測できる。

この 5 種類の要素について, ある楽曲の楽譜が指示する音符列についてなされた実演奏の音符列 \mathbf{o} が持つ逸脱傾向を図 1 に例示する. ここから, 逸脱傾向は楽譜が示す厳密な基準にある程度近い位置に偏り, いくつかの傾向を持った分布となっていることが分かる. これらが何らかの確率に基づいて生成されていると考え, 各傾向はガウス分布として近似的に表現でき, 全体はそれらを複数結合した混合ガウス分布 (Gaussian Mixture Model) と考えることができる. M 個のガウス分布を線形結合した GMM λ は, 各混合重み c_m , 平均ベクトル μ_m , 共分散行列 σ_m により, 以下の式 (1) で表される.

$$\lambda = \{c_m, \mu_m, \sigma_m | m = 1, 2, \dots, M\} \quad (1)$$

ここで c_m は以下の式 (2), (3) を満たす.

$$\sum_{m=1}^M c_m = 1 \quad (2)$$

$$0 \leq c_m \leq 1 \quad (3)$$

そして, 任意の \mathbf{o} に対するモデル λ の出力確率 \mathcal{P} は以下の式 (4) で表される.

$$\mathcal{P}(\mathbf{o}|\lambda) = \sum_{m=1}^M c_m \mathcal{N}(\mathbf{o}|\mu_m, \sigma_m) \quad (4)$$

ここで $\mathcal{N}(\mathbf{o}|\mu_m, \sigma_m)$ は m 番目の混合要素となるガウス関数である. \mathbf{o} は D 種類の逸脱要素について観測され, o_d は d 番目の逸脱要素における任意の演奏すべき音符列 \mathbf{o} の 1 音であり, μ_{md} は平均ベクトル μ_m の d 番目の逸脱要素の係数, σ は対角共分散行列の d 番目の対角要素であるとする, 任意のガウス基底密度 m に関する確率密度関数 \mathcal{N} は以下の式 (5) で表される.

$$\begin{aligned} \mathcal{N}(\mathbf{o}_m|\mu_m, \sigma_m) &= \prod_{d=1}^D \mathcal{N}(o_{md}|\mu_{md}, \sigma_{md}) \\ &= \frac{\exp\left(-\sum_{d=1}^D \frac{(o_{md}-\mu_{md})^2}{2\sigma_{md}}\right)}{\sqrt{(2\pi)^D \prod_{d=1}^D |\sigma_{md}|}} \end{aligned} \quad (5)$$

GMM によって逸脱傾向を学習した場合, 結果的には複数の分布によって異なる逸脱傾向が表現されることになり, 背景にはそのように分類される何らかのルールが存在していると考えられる. しかし, 逸脱量の統計的性質のみで分類するだけでは, そのようなルールに相当するものを明確にできない点が問題である. そのようなルールを獲得するために理論的に可能な方法として, 分類された各混合要素に含まれる学習データそれぞれに対して楽譜の情報を関連付けたうえで, それらのデータをマージしていく方法があげられる. 単に逸脱傾向を分類するという観点では, この方法によって分類の効率化などが可能になるとは考えにくい. しかし, 楽譜の情報が逸脱傾向の分岐を制御しているという仮定の下でそのようなルールを具現化するためには, 楽譜の情報を制約にして逸脱の傾向を分類することが妥当なアプローチであると考えられる.

2.3 演奏表現における楽曲解釈の説明

逸脱の発生やその傾向の分岐には複数の要因に基づく判断基準が複雑に関係しており, 各演奏者によるその基準の相違が演奏の相違の要因と考えられる. 行われた演奏表現における楽曲解釈について演奏者自身による説明が得られれば, それは演奏の特徴を知るために有益な情報となりうる. しかし, そこには思考や言語などに依存する曖昧さや表現のばらつきがあり, それらを用いて演奏の特徴を一般的に説明するのは困難である.

先述のとおり演奏者は楽曲の演奏に際し, 楽譜から等しく演奏指示を得ている. 演奏者はそれらの指示を独自に解釈して楽器の操作に反映することで, 各々の演奏表現を構築しているといえる. 本稿では演奏者の楽曲解釈そのものを扱うのではなく, 楽譜から一意に得られる情報が楽曲解釈に影響を与えていると考え, それを実演奏の逸脱と関連付けることで間接的に演奏表現の説明を行う.

2.3.1 楽曲解釈を制御する情報の導入

音符ごとに観測される実演奏の逸脱は, それぞれ楽譜上の一意な情報の組合せによって発音が指示されている. そのような情報について, 図 2 をもとに具体例を説明する. ある実演奏中の音符 \mathbf{o}' に対応する楽譜中の音符 \mathbf{o} は, 小節内の位置は 3 拍目であり, 音名は E であり, 音価は 16 分音符であるなどの情報を持つ. また, この音符と前後の音符との関係を見ると, 音名は $C\sharp, E, D$ の順で推移している. 本稿では, これらの情報が形成する「文脈」に類するものが実演奏における楽曲解釈に寄与していると仮

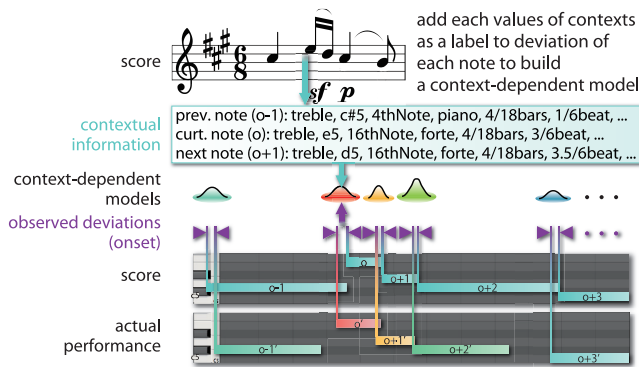


図 2 楽譜からの文脈情報の抽出と文脈依存モデルの構築

Fig. 2 Extraction of contextual information from the score for the construction of context-dependent models.

定する。仮に逸脱がまったくない実演奏を考えると、その実演奏にはそのような文脈の類がいったい関与していないといえる。しかし現実の実演奏でそのような場合はほぼありえないため、逸脱がどのような傾向を持つかはその文脈の内容に依存しているといえる。両者を関連付けたモデルによって、逸脱傾向の分岐条件を一般性の保たれた表現で説明することが可能になる。このようなモデルを音符ごとに独立して定義し、文脈依存モデル (context-dependent model) と称する。

楽譜上では一見類似した文脈情報を持つ音符でも、逸脱の傾向が類似するとは限らない。また、任意の音符がその前後の音符とどのような関係であるかがその傾向の分類に重要な基準となる場合も考えられる。このモデルでは、一般的な鍵盤楽器の楽譜から一意に得られる範囲で可能な限り多様な情報を使用するほか、その音符のみでなく前後の音符についても同様に多様な情報を考慮する。それによって実演奏において観測されるすべての音符を個別のモデルと見なし、あらゆる文脈要素について分類基準としての有効性を検討できるようになる。

2.3.2 文脈依存モデルの分類

より詳細な文脈情報を楽譜から抽出するには、多様な文脈要素の組合せを扱う必要がある。しかし詳細な情報を扱いモデルを細分化することによって、それぞれのモデルを学習するためのデータが不足する。本稿では Tree-based クラスタリング [14] を導入し、音符の数だけ存在する文脈依存モデルを文脈情報に関する制約によって分類することで、モデル間で類似する統計量を共有する。この手法の利点といえるのは、分類に有効な制約となる文脈情報が一意に選択されることである。それによって、先述した効率の問題が解決されるだけでなく、演奏者以外の恣意の混入を抑止した公正な分類を行える。

この手法による分類結果は、 S_1 から S_M までの M 個のリーフを持つ二分木 $U(S_1, \dots, S_M)$ として表現される。文脈情報に関する二択の質問が分類のための制約となり、

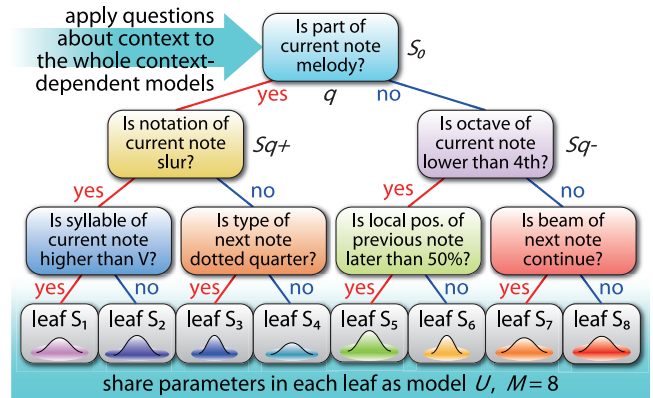


図 3 Tree-based クラスタリングによる木構造の構築

Fig. 3 Construction of the tree by the Tree-based clustering.

リーフに至る各ノードにそれぞれ保持される。ノードが持つこれらの制約をたどることで、リーフが持つ逸脱の各傾向を文脈情報の組合せによって説明することが可能になる。図 3 に例示する二分木では、逸脱傾向が文脈情報の組合せによって 8 種類に分類されている。

Tree-based クラスタリングの計算手順は以下のとおりである。

- (1) まずすべての文脈依存モデルはツリーのルートノードに集められ、すべてのモデルの共有を仮定した対数尤度が求められる。
- (2) その後、学習データの対数尤度を最大化するように文脈情報に関する質問を適用していく。
- (3) 文脈情報に関する質問はモデルの構造を決定付けるものであり、分類を行うごとにあらかじめ定義されたもののすべてをあてはめていく。
- (4) 各質問に従ってノードの分割を行った場合を仮定してモデルのガウス基底関数のパラメータを推定していき、学習データの尤度が最大になる質問が分割の基準として選択される。
- (5) 親ノードはその分類によって対数尤度が最大となる質問によって子ノードに二分される。

以上の計算手順は分割後の各子ノードにおいても対数尤度の最大値が増加する限り繰り返される。

学習データとして実演奏の音符列 $\mathbf{o} = (\mathbf{o}_1, \dots, \mathbf{o}_T)$ が与えられた場合、ノード S_m の対数尤度 $L(S_m)$ は以下の式 (6) で表される。

$$\begin{aligned}
 L(S_m) &= -\frac{1}{2} \sum_{t=1}^T \Gamma_m \left(D \log(2\pi) \log |\Sigma_m| \right. \\
 &\quad \left. + (\mathbf{o}_t - \mu_m)^t \sum_{m=1}^{-1} (\mathbf{o}_t - \mu_m) \right) \\
 &= -\frac{1}{2} \Gamma_m (D + D \log(2\pi) L \log |\Sigma_m|) \quad (6)
 \end{aligned}$$

ここで $\Gamma_m = \sum_{t=1}^T \gamma_m$ は S_m における学習データの総量、 μ_m と Σ_m は S_m における平均ベクトルとガウス分布

である。

なお、図 1 から分かるように、逸脱要素によってその傾向の特徴は大きく異なる。そのため、各種の逸脱の発生に関わる理由はそれぞれ個別の分類結果によって説明される必要がある。本稿では逸脱要素ごとにモデルを分けることによって学習データを分割し、各逸脱要素のデータに対して個別に Tree-based クラスタリングを行うことで、それぞれに独立した木構造を獲得する。

2.3.3 分類の規模の決定

図 1 で示した逸脱の傾向からは、実質的にユニークといえるパターンはそれほど多くないことも分かる。よって木構造におけるリーフ数もその程度の規模とすることが望ましいが、ノードの分割をどこで終えるかが問題となる。本稿では最小記述長 (Minimum Description Length) 基準 [15], [16] を導入し、分割の実行または停止を自動的に決定する。これはデータ列 $\mathbf{x} = (x_1, \dots, x_N)$ に対して記述長が最小となるモデルを最適とする基準である。ここで任意のモデルの最大尤度を L とすると、その記述長 $\text{MDL}(\mathbf{x})$ は一般に以下の式 (7) で表される。

$$\text{MDL}(\mathbf{x}) = -\log L(\mathbf{x}) + \frac{D}{2} \log N \quad (7)$$

モデルがより複雑になれば学習データに対する尤度は大きくなるため、式 (7) における第 1 項の値は減少する一方、記述長の増加に従って第 2 項の値は増加する。両者の間にはトレードオフの関係があり、記述長 MDL はデータ \mathbf{x} の表現に適当な複雑さのモデルで最小値をとる。また、この第 2 項は単調増加の一次関数であり、この傾きが急であればより早い段階で分割が停止することになる。第 2 項の傾きは Tree-based クラスタリングで分割を停止するための閾値として使用でき、これを設定することで決定木の規模をある程度制御することが可能である。大規模な木構造とすれば、演奏に含まれる特異な特徴でも平滑化することなく精緻にとらえることができる。一方、小規模な段階で分割を停止した場合、木構造が持つのは分類に際して高効率な制約のみとなるため、実演奏データの中でより本質的といえる逸脱傾向に焦点を当てることができる。

この基準を用いてモデル U のノード S_m を質問 q によって S_{mq+} と S_{mq-} に分割する場合を考える。分割したモデル U' の記述長は以下の式 (8) で表される。

$$\begin{aligned} I(U') = & \sum_{m'=1, m' \neq m}^M \frac{1}{2} \Gamma_{m'} \left(D + D \log(2\pi) + \log |\Sigma_{m'}| \right) \\ & + \frac{1}{2} \Gamma_{mq+} \left(D + D \log(2\pi) + \log |\Sigma_{mq+}| \right) \\ & + \frac{1}{2} \Gamma_{mq-} \left(D + D \log(2\pi) + \log |\Sigma_{mq-}| \right) \\ & + D(M+1) \log W + C \end{aligned} \quad (8)$$

ここでモデル U' のノード数は $M+1$ であり、 Γ_{mq+} , Γ_{mq-} はそれぞれノード S_{mq+} , S_{mq-} の占有率、 $W = \sum_{m=1}^M \Gamma_m$

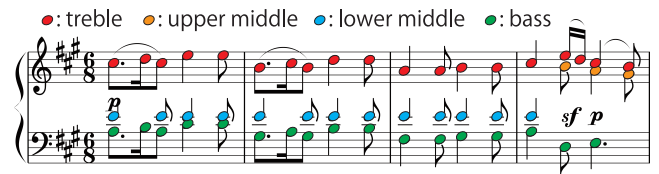


図 4 音符が属する声部の抽出

Fig. 4 Extraction of voice part that each note belongs to.

と C はモデルの選択に必要な符号長 (定数) である。また、分割前後のモデルの記述長の差分 $\Delta_m(q)$ は以下の式 (9) で表される。

$$\begin{aligned} \Delta_m(q) = & l(U') - l(U) \\ = & \frac{1}{2} \left(\Gamma_{mq+} \log |\Sigma_{mq+}| + \Gamma_{mq-} \log |\Sigma_{mq-}| \right. \\ & \left. - \Gamma_m \log |\Sigma_m| \right) + D \log \sum_{m=1}^M \Gamma_m \end{aligned} \quad (9)$$

分割を行う際、まず $\Delta_0(q')$ を最小化する質問 q' を決定し、それがルートノードとなる。 $\Delta_0(q') < 0$ において、ノード S_0 は $S_{q'+}$ と $S_{q'-}$ に分割される。その 2 個のノードにおいても同様であり、この計算手順は分割すべきノードがなくなるまで繰り返される。すなわち、 $\Delta_0(q') > 0$ において分割は行われない。

2.4 モデルの設計

ここでは楽譜と実演奏から抽出する文脈や逸脱など、どのような情報を用いてモデルの学習を行うのかを説明する。

2.4.1 音符が所属する声部と和音

先述のとおり、1 つの音符についての文脈情報はその前後の音符との関係も含めて記述される。和音ではない並行な声部が同時に存在するような場合、その連続性の考慮は同一の声部における音符列において行われるべきである。よって、和音ではない並行な声部が同時に発音されている場合、音符の連続性の考慮において各声部に属する音符は明確に分離して扱われる必要がある。本稿の評価で使用する楽譜内の音符は、すべて以下の声部のいずれかに分類する。

treble 大譜表の上段にあり、主に旋律を担う音符列

bass 大譜表の下段にあり、主に低音を担う音符列

middle 上下各段において上記の音符列とは別に並走する音符列

分類例を図 4 に示す。一般的な音域の性質から、上段において観測される音符列の声部を treble、下段のそれを bass と称する。同一の段においてある音符が発音中に付尾の方向がそれと異なる音符が存在する場合、第 2 の音符列があると見なすことができる。各段において平均的に中央音域寄りの音符列の声部を middle とし、上段のそれを upper middle、下段のそれを lower middle と称する。一般的な楽曲では treble が主な旋律、bass が低音であり、

middle はそれ以外の修飾的な役割を担う場合が多いと考えられる。しかしすべての楽曲がそれに従うとは限らないため、ここでは単純に声部の区別のみを行い、各声部の比重などの差は設けない。

一方、発音タイミングと付尾の方向、音価が同一の音符が複数あれば、それらは同一声部内の和音の構成音と見なすことができる。同一声部内において、和音を構成する音符群は1つの大きな音符であると見なすと、各音符の前後関係があるのはその和音の前後に発音された音符となると考えられる。ここで和音の前後が単音の場合、和音の構成音符はその単音に対して等しく前後関係があると見なす。ここで前後の音符も和音である場合、その中で独自の声部のような関係が構成されていると考えられるため、その場合の前後関係は各和音中での位置関係と音高が最も近傍となる音符どうしで構築する。また、和音中の各音符には主従関係が想定され、主たる音符が和音中のどの位置のものかは声部など他の文脈要素に影響すると考えられる。最高音または最低音以外となる音符の数は和音の構成音符数によって変動しうるため、和音となる音符については3種類（最高音、最低音、それら以外の音）に区別する。なお、全音符の場合は付尾がないことを考慮し、同時に発音される音価が同一であることを条件として和音の判定を行う。

2.4.2 文脈要素

一見類似した特徴を持つ音符であっても逸脱の傾向が類似するとは限らないため、各音符の文脈情報には、単独では同一に見える音符でもすべて個別のものと見なせる程度の詳細さが必要である。そのような情報には、一般的な鍵盤楽器の楽譜の様式に従ったもので、なおかつ音符単位で一意に得られるものが相応しい。ここでは主に以下の12種類の要素についての情報を音符ごとに抽出する。

syllable その音符の音名と楽曲の主音との音程を度数（短三度や完全五度など）で示す。主音は楽曲の途中でも可変としており、楽譜上で転調などが行われる可能性にも対応できる。

step C から B までの12音でその音符の音名を示す。

accidental 臨時記号の有無と種類を示す。

octave その音符の属する大まかな音域を示す。一般的なピアノの88鍵盤をA0からC8までとし、CからBを1組とするオクターブの領域を0から8の値域で表す。

chord その音符が和音の一部であるかどうかを示す。判断基準および記述内容は2.4.1項で定めた表記に従う。

type その音符の音価を示す。

staff その音符が大譜表の上下段どちらに書かれているかを示す。音符単位で判断するため上下段をまたぐ連符の場合は途中で情報が変化するが、そのような場合の音符の連続性はbeamの情報から判断できる。

beam その音符の符尾の形状を示す。

local その音符の小節内での時間的な位置を示す。例と

して4/4の楽曲で3拍目に打鍵すべき音符の情報は50%となる。

global その音符の楽曲全体での時間的な位置を示す。楽曲の開始位置を0%、終了位置を100%とする。

voice その音符が所属する声部を示す。判断基準および記述内容は2.4.1項で定めた表記に従う。

notations スタッカートやトリルなど、その音符に個別に付加された指示がある場合、それらを示す。

実演奏から抽出した音符列には楽譜の音符と完全に一致しない音符が含まれる場合があるが、それらの大部分は楽譜上にあるトリルなどの指示に付随して発生するものであると考えられる。それらについては打鍵のタイミングと音高が最も近い音符をそのトリガと見なし、文脈情報を共有する。また、記譜されていない音符であることを明記する情報(not written in the score)を追加することで、それらをすべて個別の音符と見なす。

2.4.3 逸脱要素

実演奏と楽譜を対照し、音符ごとに5種類の要素についての逸脱量を以下の手順で抽出する。

Onset beat 打鍵タイミングの逸脱を抽出し、拍単位で表記する。算出は音符単位で行われるため、同時刻に発音する音符であっても実演奏では必ずしも同時刻とはならない。例として、任意の音符の打鍵が楽譜の指示より1/4拍遅く行われた場合の逸脱量は0.25となる。

Offset beat 離鍵タイミングの逸脱を抽出し、拍単位で表記する。表記方法はOnset beatと同様で、楽譜の指示する音価に従って押鍵を続けるべきタイミングよりも1/2拍早く離鍵した場合の逸脱量は-0.5となる。

Gate time ratio タイミングとは関係なく、楽譜が指示する音価に対して押鍵（打鍵から離鍵まで）されていた時間的な長さの比率を示す。押鍵長の変化を指示する条件（スタカートやテヌートなど）は文脈情報としてとらえているため、逸脱量は一律に元の音価を基準にして算出する。例として、任意の音符についての押鍵長が音価の3/2であった場合の逸脱量は1.5となる。

Dynamics ratio 打鍵強度を示し、音符の発音のためMIDI音源を駆動するペロシティとして観測される127段階の値から算出される。楽譜中に強弱の指示があれば、記述箇所以降から次の指示までの音符については ppp , pp , p , mp , mf , f , ff , fff の8段階に従って127の値域を等分した値を基準とし、それと実演奏データにおけるペロシティ値の差を逸脱量とする。楽曲全体を通して強弱記号がない楽譜の場合、基準値は値域の中央とする。

Tempo ratio 楽曲の演奏開始時のBPM (initial BPM of the performance) に対する瞬時BPM (BPM of current beat) の比率で、以下の式(10)から算出される。

表 1 実演奏データの詳細
Table 1 Details of the actual performance data.

| データセット | データベース | 楽譜（楽曲，作曲者，版，頁） | 楽曲数 | 音符数 | 演奏者数 | クラス |
|-----------|--------------------|--|-----|-------|------|----------|
| JSB-846 | C.M.PEDB version 2 | “Well-Tempered Clavier BWV846 Prelude,” J.S. Bach, Wiener Urtext Ed., pp.2–3 | 1 | 549 | 3 | Experts |
| FFC-N02 | C.M.PEDB version 2 | “Nocturne Op.9, No.2,” F.F. Chopin, Paderewski Ed., pp.15–17 | 1 | 709 | 3 | Experts |
| FFC-N10 | C.M.PEDB version 2 | “Nocturne Op.9, No.10,” F.F. Chopin, Paderewski Ed., pp.54–55 | 1 | 555 | 3 | Experts |
| WAM-545 | C.M.PEDB version 2 | “Piano Sonata K.545 the 1st Mov.,” W.A. Mozart, G. Henle Verlag, pp.266–269 | 1 | 1,278 | 2 | Experts |
| WAM-331T2 | C.M.PEDB version 2 | “Piano Sonata K.331 Theme of the 1st Mov.,” W.A. Mozart, G. Henle Verlag, p.160 | 1 | 239 | 3 | Experts |
| WAM-331T3 | C.M.PEDB version 3 | “Piano Sonata K.331 Theme of the 1st Mov.,” W.A. Mozart, G. Henle Verlag, p.160 | 1 | 239 | 8 | Students |
| JSB-TPI | Original | “Two-Part Inventions BWV772–786,” J.S. Bach, G. Henle Verlag, pp.2–31 | 15 | 8,807 | 5 | Students |
| WAM-265 | Original | “Twelve Variations K.265,” W.A. Mozart, Wiener Urtext Ed., pp.51–61 | 13 | 6,123 | 5 | Students |
| WAM-331 | Original | “Piano Sonata K.331 the 1st Mov.,” W.A. Mozart, Wiener Urtext Ed., pp.18–27 | 7 | 6,151 | 5 | Students |

$$\text{Tempo ratio} = \frac{\text{BPM of current beat}}{\text{initial BPM of the performance}} \quad (10)$$

瞬時 BPM の値は拍ごとに更新されるため，楽譜上で同一の拍に含まれる音符の逸脱量はすべて同じ値となる。

2.4.4 実演奏データ

次章の評価で使用するすべての実演奏データを表 1 にまとめて示す。評価の特性に合ったデータを使用するため，データセットは以下の 3 種類のデータベースから取得する。

CrestMusePEDB version 2 ピアノ演奏の波形を特定の MIDI 音源において聴覚上で模倣する MIDI データを人間（耳の確かな音楽専門家）による手作業で作成し，楽譜からの逸脱を音符ごとに抽出したデータベース [17], [18] である。楽譜からの逸脱として提供されているデータは，音符ごとには打鍵および離鍵の拍単位での逸脱と指定 MIDI 音源を駆動するペロシティ（音符ごとの音量），拍ごとには演奏開始時の BPM を基準とした瞬時 BPM 変動である。逸脱の抽出に用いられた音源は一般に名演とされるものであり，データの品質は相当に高いと考えられる。しかし演奏者や楽曲のラインナップが様でなく，後述の Original で意図したような対照比較が可能な選択肢は非常に少ない。また，ペダルの操作による離鍵タイミングなどへの影響を波形から特定することは困難であるため，演奏表現はすべて鍵盤操作に集約されていると考えることになる。

CrestMusePEDB version 3 芸術大学に在籍，または

卒業後演奏活動が続けている演奏者による MIDI ピアノ演奏の逸脱データベースである。逸脱は前述の CrestMusePEDB version 2 と同一の楽譜から抽出されているが，ペダル操作などの情報を含むため，Offset beat および Gate time ratio の直接比較はできない。

Original 異なる演奏者が同一の楽曲演奏した事例を多数獲得するため独自に収録した，アマチュアピアノ演奏者による MIDI ピアノの実演奏データである。演奏者の特性，演奏の収録方法および瞬時 BPM などの各種逸脱量の記述方法は前述の CrestMusePEDB version 3 と同様である。しかし両者間では収録環境や楽譜の版などが異なるため，モデルの学習や比較などに際して混用はできない。

なお，Original での逸脱データの作成手順は CrestMusePEDB と共通している。楽譜を厳密に変換した MIDI データと逸脱を含む実演奏の MIDI データとの間で音符ごとの発音を対照し，MIDI データ上の時間軸で各拍の頭となる位置で演奏開始時の BPM を基準とした瞬時 BPM の変動を求める。音符ごとの各種逸脱量はその BPM 変動の下で算出している。

2.4.5 文脈依存モデルの学習

文脈依存モデルではすべての音符に個別のモデルが対応しており，個々のモデルにおける逸脱の統計量の記述は単純化が可能と考えられる。よって式 (1)，(5) において単一ガウス分布 ($M = 1$)，5 種類の逸脱要素 ($D = 5$) のモデルとし，以下の 2 段階の手順でそれらの学習を行っている。(1) 学習データの大まかな傾向を把握するため，音符単独の音高のみを文脈情報とした初期モデルを作成し，逸

脱の統計量をそれぞれ求める。

- (2) 音符ごとの文脈情報を前後音符を含むすべての組合せに拡張して、すべての音符が個別となるような文脈依存モデルを作成する。当該音の音高についての文脈情報が一致する初期モデルを基に、それぞれの文脈依存モデルにおいて逸脱の統計量を更新する。

また、各文脈要素についての情報は、先述のとおり当該音符だけでなくその前後の音符についても考慮している。よって、任意の音符の逸脱傾向は、上記 12 種類の文脈要素を用いた総数 36 種類のパラメータによって説明されるといえる。

3. 評価実験

提案法について、本章では手法としての有効性を示し、実用的な用途のための可能性を探る。まずは演奏者識別実験によって提案法の有効性を確認する。次に、得られた決定木からモデルが演奏者の解釈をどのようにとらえているかを検証する。最後に、既存の調査研究で示唆されている実演奏に関する知見の証明を試み、提案法が演奏の分析に有用であるかを検討する。

3.1 Tree-based クラスタリングの有効性の検証

Tree-based クラスタリングは手法としては既存のものであるが、楽器演奏にこれを適用した場合の効果は未知である。その有効性を示すには、付加的な情報を用いずボトムアップに逸脱傾向のクラスタリングを行ったモデルと比較して評価を行う。ここでは同一の楽曲を複数の演奏者が演奏している場合を想定し、その演奏者の識別を試みる。

評価は以下の 2 手法の比較によって行う。

Tree-based clustering 楽譜の情報を使用して実演奏に含まれる逸脱の傾向を M 個のリーフを持つ木構造に分類するモデル（提案法）。

Bottom-up clustering 逸脱を数値の分布のみから複数の傾向に分類するモデル。ここでは上述のモデルの木構造が持つリーフ数に相当する混合数 M の GMM を使用する（以降、ベースラインと称する）。

提案法のモデルは図 5 の要領で獲得した。演奏者の間に共通の見解があると仮定して、あらかじめ全員の逸脱量を用いて共通の木構造を獲得しており、それらが持つ分類基準に従った場合の逸脱傾向のみが演奏者独自のものとなっている。また、ここでは演奏者間での演奏内容は音符単位で統一しており、楽譜に音符単位で具体的な指示がない箇所（装飾音など）の演奏は学習データに含まれていない。

今回の条件では純粋に逸脱傾向の違いのみで識別を行うことになるため、一般的に考えられる演奏者の識別よりも困難な課題となることが予想された。その難易度を示す指標を獲得するため、主観評価実験によって人間はこれをどの程度識別できるかを調査した。主観評価実験では、複数

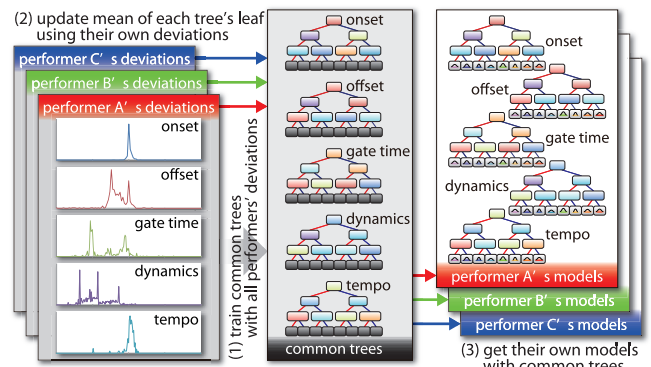


図 5 演奏者識別のためのモデル学習（演奏者 3 名の場合）

Fig. 5 Model training for identification of three performers.

の演奏者による同一な楽譜の演奏を選択候補として用意し、異なる楽譜を演奏したりファレンスと同じ演奏者による演奏と思われるものを選択させている。被験者（Human listeners）は 20 名とし、先述の各モデル化手法での実験と同じ実演奏データで MIDI 音源を鳴らしたものをリファレンスおよび選択候補に用いている。選択候補の数は先述したモデル化手法の実験と同人数としている。また、参考として無作為に候補を選択した場合（Random choice）についても併記する。

3.1.1 中規模な実演奏データでの検証

まずはなるべく多数の類似した楽曲の実演奏データでの検証を行うため、表 1 の JSB-TPI, WAM-265, WAM-331 を使用した評価を行った。1 章でも述べたように、大規模な実演奏データの収集は困難であるため、このデータセットは中規模程度のものであると考えることにする。モデル学習用と識別用のデータは楽曲単位で分割している。おおむね全体の 1/4 程度の音符数となる楽曲の組合せでモデルの学習を行い、残りの楽曲のいずれかを識別用に使用した。

結果を図 6 (a) に示す。これらのデータセットは作曲方針の共通する楽曲集や同一の主題を基にした変奏曲であり、セット内の楽曲は互いに類似したものと見なせる。しかし被験者の識別率は無作為選択（20%）に近い程度となっており、異なる楽曲間での演奏者の識別は非常に困難な課題であったといえる。そのような条件下でも提案法は被験者に対して倍以上の精度で識別が可能であったことから、提案法は単独の楽曲に依存しない演奏者固有の特徴を捕捉できていたと考えられる。また、どのセットでも提案法はベースラインに対して 1 割程度高い識別率を示している。以上から逸脱の傾向を統計的に捕捉することに加え、その傾向が生じる理由を文脈情報によって説明することが有効であるといえる。

3.1.2 小規模な実演奏データでの検証

文脈情報の使用が少量のデータに対してでも有効であることを確認するため、単独の楽曲内での演奏者識別も試みた。ここでは CrestMusePEDB version 2 から、3 曲

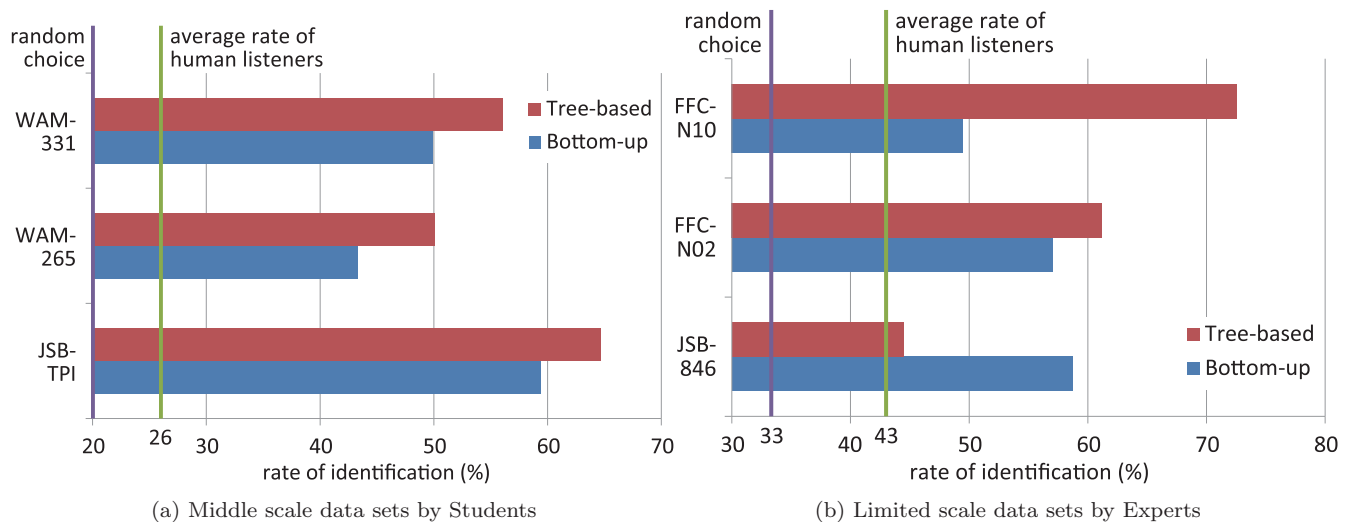


図 6 同一楽曲を演奏した異なる演奏者の識別率
Fig. 6 Identification rate of different performers who played the same music.

各 3 名の演奏者による実演奏データ (表 1 の JSB-846, FFC-N02, FFC-N10) を使用した. 実演奏データの量は先の条件 (中規模) に対しておおむね 1/10 以下となる. 学習データの量を可能な限り確保しつつオープンな評価を行うため, 楽曲の構成を考慮して 1/2 程度に分割したデータをモデルの学習に, 残りのデータをさらにフレーズの長さ程度に分割したものを単一の識別候補に用いた.

結果を図 6(b) に示す. 無作為選択 (33%) に対する被験者の識別率は先の条件よりは良くなっているが, 選択肢が 5 人から 3 人となっていることを考慮すると, こちらも容易な課題であったとは考えにくい. モデル間では, データセット JSB-846 において提案法の識別率がベースラインを下回る結果となった. これは元々ダイナミクスの表現に制約があるチェンバロのための楽曲であり, 楽譜から得られる文脈情報の種類が他の楽曲に比べて限られている. また, 楽譜上で特徴的な点として複数の音符が一組となったパターンが続くことがあげられる. 文脈要素の決定方法は無限に存在するが, 今回設計した文脈依存モデルで考慮されているのは当該音符とその直前直後の音符のみである. つまり局所的な文脈に依存したモデルであるといえるため, このような楽曲で逸脱傾向の分岐に寄与する制約を獲得するには不適であった可能性がある. なお, 制約として選択された文脈要素が何であったかについては, 3.2.1 項で個別に検証を行うことにする. 一方, 他の 2 セットはピアノのための楽曲であり, 楽譜からは JSB-846 よりも多様かつ詳細な文脈情報を獲得できる. どちらの楽曲でもベースラインを明らかに上回る結果となっており, 文脈情報を活用できる提案法がこのように少量の学習データにおいても有効に機能していると考えられる.

3.2 楽曲解釈に貢献する文脈要素の分析

3.1.2 項での検証結果において, 提案法の有効性に楽曲

の特性による影響がみられた. ここでは CrestMusePEDB version 2 から楽曲または演奏者について共通項のある実演奏データを組み合わせて学習したモデルを使用し, 実際に得られたモデルの木構造を通して楽譜から得られる情報の種類や量の差異がどれほどのものであったかを確認する.

ここで任意の文脈要素 Q の貢献度 C_Q を考える. M 個のリーフを持つ木構造において, m 番目のリーフは N_m 個の文脈依存モデルを共有しているとする. ルートノードからそのリーフに至るまでに通過するノードのいずれかは Q に関する質問を持つ可能性がある. ここで R_Q をルートノードからすべてのリーフまでの各経路において通過した Q に関する質問を持つノードの総数として, 以下の式 (11) から C_Q を評価する.

$$C_Q = \sum_{m=1}^M \left(\frac{N_m}{N_{all}} \times R_Q \right) \quad (11)$$

この式 (11) おいて, 任意のノードの文脈要素がより大きな貢献度を持つと見なされるのは以下の場合である.

- ルートノードに近い位置にあること
- 共有する文脈依存モデルの量が多いリーフへの経路に含まれること

3.2.1 演奏表現を制御する楽曲の特性

楽曲を共通項とする 3 種類のデータセット (JSB-846, FFC-N10, WAM-545) に式 (11) を適用した. その結果を図 7(a) に示す. 異なる楽曲の間でも貢献度の高い文脈要素の上位 5 種類 (e.g., C_{step} , C_{octave} , C_{type} , C_{local} , $C_{syllable}$) は同じものであった. 異なるのはその順位と比率で, 特に JSB-846 は C_{step} が突出して高いほか, C_{chord} や $C_{notations}$ がほほないなど, 他の楽曲と傾向が異なる. JSB-846 については 3.1.2 項で得られた結果でも他の楽曲と大きく異なる傾向となっており, その原因として今回設計したモデルの文脈情報では楽譜の特徴をうまくとらえら

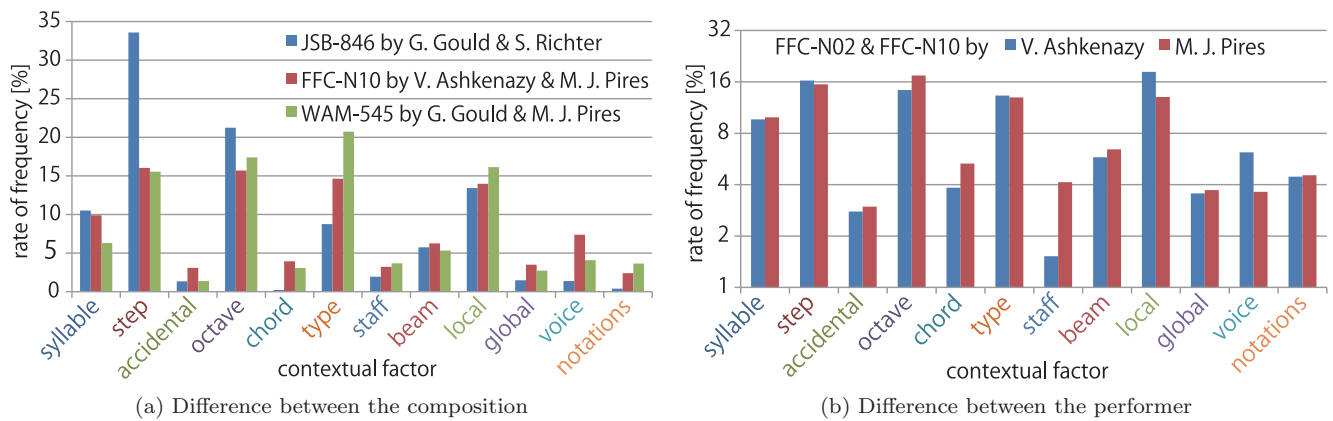


図 7 各文脈要素の木構造への貢献度

Fig. 7 Contribution to the tree structure of each contextual factor.

れていない可能性をあげていた。音高に関連する文脈要素である C_{syllable} も C_{octave} の比率も C_{step} とまとめて考慮すると、JSB-846 ではそれらだけで全体の 80% 近い貢献度を占めており、非常に偏りが大きい。このような楽曲のモデル化には使用する文脈要素の定義を見直しが必要と考えられる。他の楽曲では先にあげた上位の文脈要素のうち、 C_{syllable} 以外の 4 種類がともに 15% 前後の貢献度を持っている。その他の要素も数% 程度ではあるが同等に貢献しており、定義した文脈要素はもれなく活用されたといえる。

3.2.2 演奏者に固有の特徴の推定

次に、単純に演奏者のみの違いを見ることができるかを検証する。式 (11) を 2 種類のデータセット (FFC-N02, FFC-N10) に共通する演奏者 2 名の演奏について適用した。その結果を図 7 (b) に示す。各文脈要素の間には先の比較ほど大きな相違はなく、逸脱の傾向への影響力は一般に演奏者よりも楽曲の方が強いことが分かる。一方、 C_{staff} や C_{voice} など、比較的両者の間で明確な相違が見られる要素もある。 C_{staff} は大譜表において音符が表記された段の違いを表し、主にその音符が左右どちらの手で演奏されたかを意味する。それに対して C_{voice} は単純に声部を意味し、楽曲の音楽的構造に直結する情報である。二者の演奏の違いを表すならば、V. Ashkenazy の演奏では前者の情報に、M. J. Pires の演奏では後者の情報によって、より高い頻度でその逸脱傾向が制御されている。任意の演奏者に関して明確な共通項を持った実演奏データを用い、このような点に着目して木構造の分析を行うことで、その演奏者が一般的に楽譜上のどの点を重視して演奏しているかを知る手がかりとなると考えられる。

3.3 提案法による実演奏の分析

最後の評価として、提案法が実演奏の分析に有用であるかを検討する。ここでは、著名な演奏家と学生の演奏者が同一の楽曲を演奏した場合の相違に焦点を当てる。その話題を扱った既存の演奏分析研究 [19] において、両者の実演

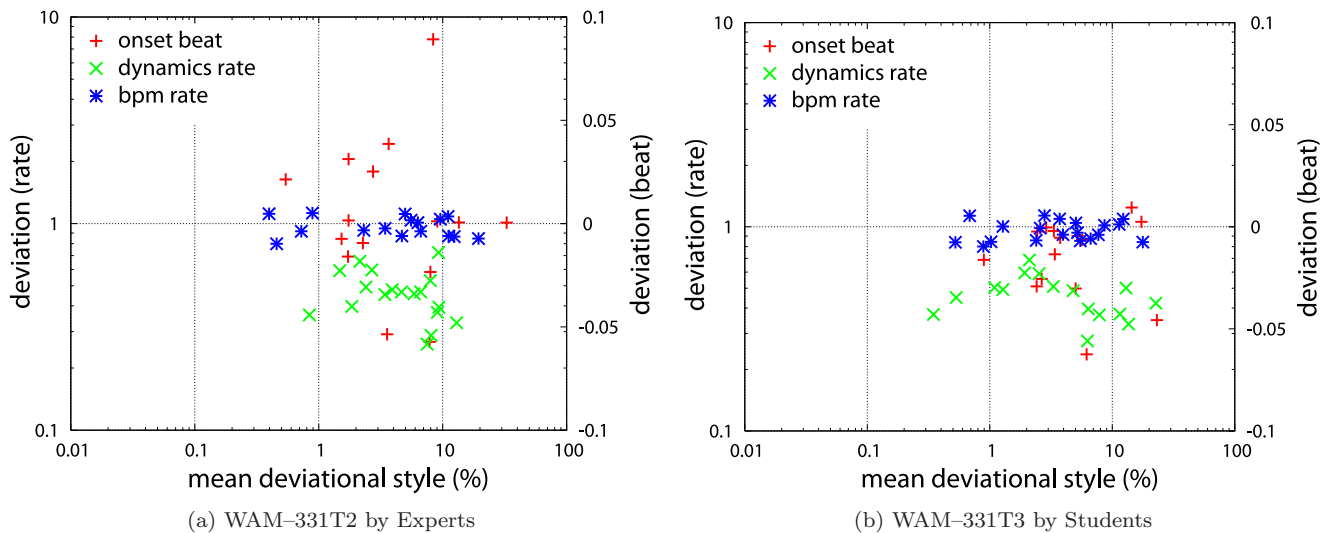
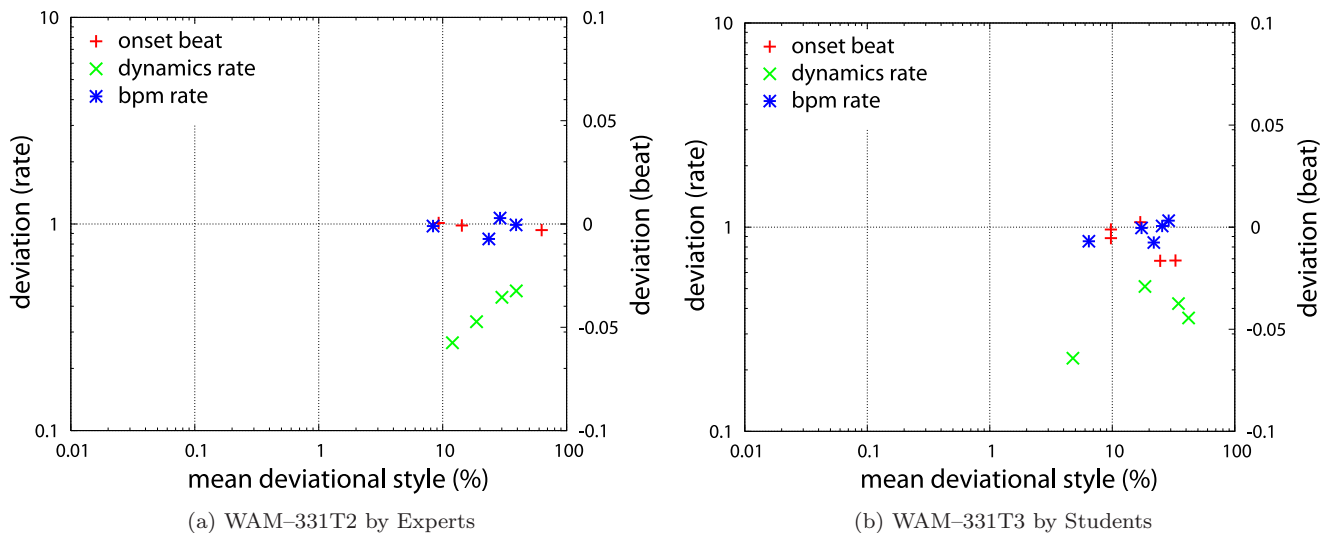
奏の比較から以下のような知見が得られている。

- (1) 熟練した演奏者 (Experts クラス) の演奏は一般の演奏者 (Students クラス) の演奏よりも強い個性を持つこと
- (2) 後者の演奏も本質的な傾向では前者のそれと極端な違いはなく、両者の基本的な演奏技術に大きな差はないと考えられること

同一の楽曲の実演奏を異なるクラスの演奏者で学習したモデルを比較することで、これらの知見の裏付けを行う。また、両クラスの演奏にどのような特徴の相違があり、それがどのような理由に基づくのかを検証する。

モデル学習には実演奏データとして表 1 に示すデータセット WAM-331T2 (Experts クラス)、WAM-331T3 (Students クラス) を使用し、分割停止基準を操作した Tree-based クラスタリングによって規模を変化させた木構造を獲得した。先述のとおり、クラス間で比較可能な逸脱要素は Onset beat, Dynamics ratio, Tempo ratio の 3 種類である。それらの木構造によって分類された逸脱の傾向を図 8 (中規模, $M \simeq 16$)、図 9 (小規模, $M \simeq 4$) に示す。図の縦軸は木構造ごとに各リーフが持つ逸脱量を表し、中央に近いほど楽譜どおりの演奏となる。図の横軸は M 種類の逸脱の特徴それぞれへの分岐頻度を表し、右端に近いほど頻度が高くなる。逸脱要素ごとの木構造について各リーフが共有する文脈依存モデルの数を考慮し、総計が 100% の頻度となるように正規化している。

知見 (1) に関しては、中規模な木構造の比較 (図 8) から、両者の相違は逸脱傾向の変動幅の違いに表れていることが分かる。Students の Onset beat は楽譜どおりから約 1/20 拍早い程度までの範囲で分布しているのに対し、Experts の Onset beat はそれに加えて 1/10 拍遅い程度までの広範囲の分布となっている。このことから、Experts の逸脱は Students のそれよりも強い特徴を持つ傾向があるといえる。一般に名演と称される Experts の演奏が決して出鱈目ではないことから、楽曲を音楽的に破綻させるこ

図 8 中規模な分類で見る逸脱の傾向 ($M \approx 16$)Fig. 8 Tendency of artistic deviation in medium-sized tree ($M \approx 16$).図 9 小規模な分類で見る逸脱の傾向 ($M \approx 4$)Fig. 9 Tendency of artistic deviation in small-sized tree ($M \approx 4$).

となく演奏に自らの個性を投影できるかが Students との違いであると考えられる。知見 (2) に関しては, Experts の木構造をきわめて単純なものにした比較 (図 9) において両者の逸脱が分布する領域に大きな相違はない。このことから, 演奏表現のための楽曲解釈について, 両者間で本質的な部分での違いは少ないことが推測される。先述の研究例とは異なる楽曲であるが, それらの知見を裏付ける傾向はここでも再現されている。このような傾向を数値的に得ることは, ボトムアップなクラスタリングのみでも実現が可能である。しかし, それらの発生に寄与する制約が何であるのかを知るためには, 提案法によって得られたモデルの木構造の分析が必要である。

両者のモデルの木構造を, 逸脱要素ごとに図 10, 図 11, 図 12 に例示する。ノードはその文脈要素によって図 7 と同様に塗り分けてあり, リーフには分類されたモデルが持つ逸脱の統計量が示されている。また, リーフに至る分岐

には, そのような逸脱への分岐頻度 (図 8, 図 9 と同様) を併記する。

Tempo ratio (図 12) においては両者の木構造に類出する文脈要素がほぼすべて一致しているほか, Dynamics ratio (図 11) でもそれぞれ同様の文脈要素が重要な判断に参与している。一方, Onset beat では両者の木構造に相違が大きい。ルートノードの付近に位置し, なおかつ比較的高い頻度で分岐するリーフに直結するノードの文脈要素は, 図 10(a) に示す Experts の木構造では C_{step} や C_{local} となっている。それに対し, 図 10(b) に示す Students の木構造で同等の位置にあるのはそれらの要素ではなく, $C_{syllable}$ や C_{beam} となっている。また, 前者のリーフ側では比較的少数ながら他とは傾向の相違が大きい逸脱量を分類するための構造が築かれている。それに対し, 後者では全体の 1/5 以上の音符の文脈依存モデルがルートノードのみを基準とする単一の傾向に分類されている。このことから,

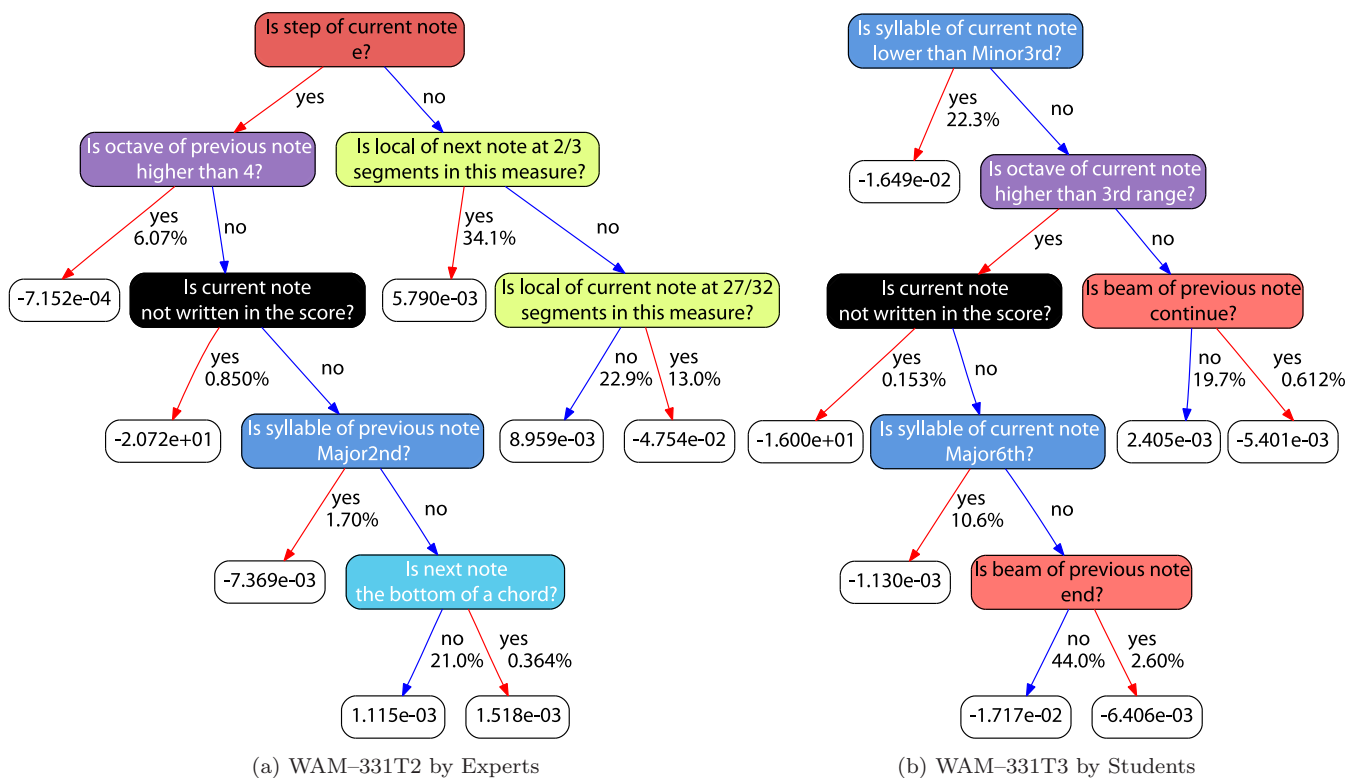


図 10 Onset beat の木構造
Fig. 10 Trees of onset beat.

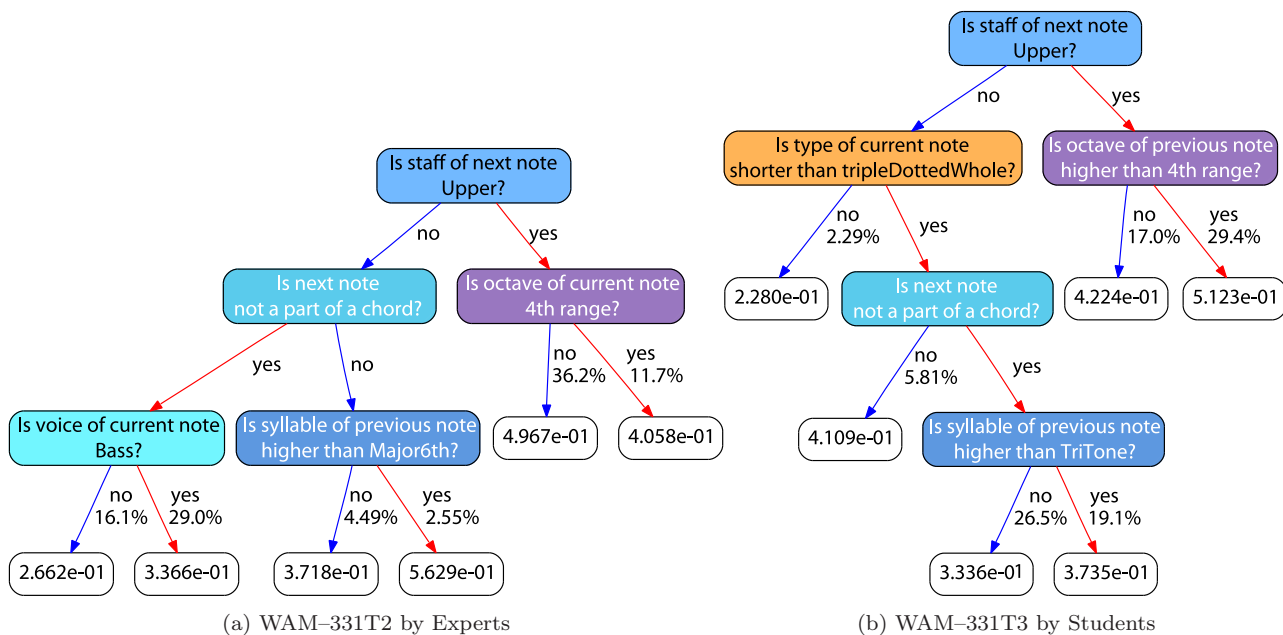


図 11 Dynamics ratio の木構造
Fig. 11 Trees of dynamics ratio.

Students の逸脱傾向は Experts のそれよりも比較的単純な説明で記述が可能であることが分かる。

本手法を用いることで、Experts と Students が楽譜上の指示をどのように演奏しているかを比較することが可能となった。これにより、楽譜を解釈して楽曲を演奏するための基本的な知識に関して、Students は Experts に近いレベルに到達していると推測できた。

4. おわりに

本稿では人間の鍵盤楽器の演奏における特徴を、楽譜から得られる情報を用いて説明する手法について述べた。提案法の利点はモデルの木構造の記述に必要な知識を実演奏データから直接かつ一意に学習可能なことと、その手順において演奏者以外の恣意の混入を抑止可能なことである。

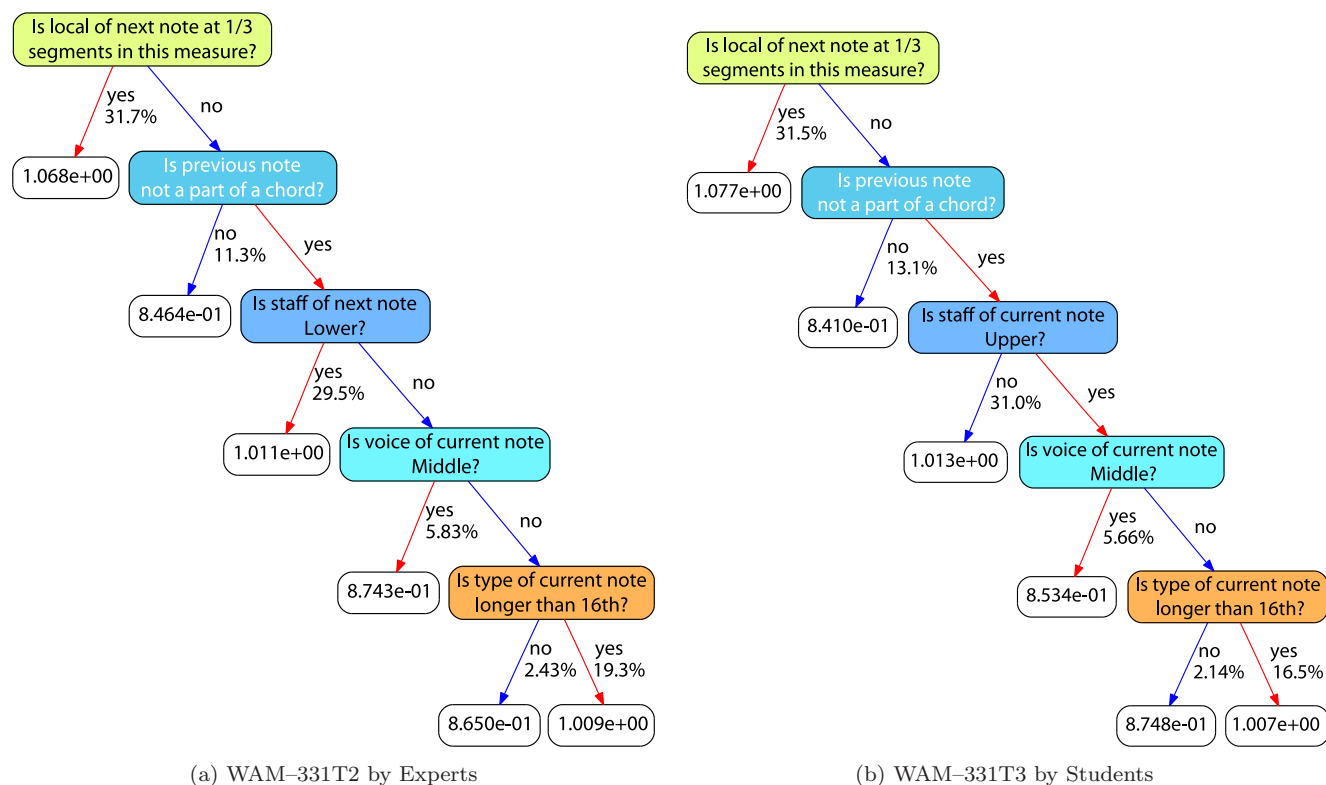


図 12 Tempo ratio の木構造

Fig. 12 Trees of tempo ratio.

演奏者識別の評価によって、逸脱のモデル化に楽譜から得られる文脈情報を使うことの有効性が示されたほか、提案法によるモデルは人間の受聴者には判別できない微妙な演奏の違いもとらえられることも分かった。また、得られたモデルを文脈要素の視点から分析することで、任意の実演奏においてどのような文脈が逸脱の傾向をより強く制御するのかを知ることができた。先述の演奏者識別で高評価とならなかった楽曲については、文脈要素として使用する情報の選択次第でその実演奏の特徴を効率良くとらえることが可能になると考えられる。しかしここで文脈要素といえるものは楽譜だけでなく、それ以外からでも際限なく定義することができるため、究極的な解は存在しない。モデル化に際して何を文脈要素として用いるべきかは提案法の本質的な課題であり、今後最も具体的に検討すべき事項の1つといえる。そして、提案法によるモデルには楽器演奏に関する既存の知見を裏付ける情報が含まれていることも分かった。これは本手法が実演奏を分析する用途において、そのような知見の客観的な証明に有用となる可能性を示している。しかし、本稿で検証したのはその一端であり、その有用性について確証を得るには今後多数の追試を要する。

今後の展望として、モデルが持つ情報を音楽的に自然な形態で演奏者に還元するための検討を進めている。また、既存の演奏の特徴を用いることについて本手法が関連する研究分野として、自動演奏生成 [20], [21], [22] がある。こ

れは未知の演奏内容に自然な逸脱を付加していく問題と考えられ、提案法もそのような用途への応用が可能である。

参考文献

- [1] Seashore, C.E.: *Psychology of Music*, Dover books on music, music history, Courier Dover Publications (1938).
- [2] Dixon, S., Goebel, W. and Widmer, G.: The performance worm: Real time visualisation of expression based on Langrer's tempo-loudness animation, *Proc. International Computer Music Conference (ICMC)*, Goteborg, Sweden, pp.361–364 (2002).
- [3] Grindlay, G. and Helmbold, D.: Modeling, Analyzing and Synthesizing Expressive Piano Performance with Graphical Models, *Machine Learning Journal*, Vol.65, No.2-3, pp.361–387 (2006).
- [4] Dorard, L., Hardoon, D.R. and Shawe-Taylor, J.: Can Style be Learned? A Machine Learning Approach towards 'Performing' as Famous Pianists, *Proc. Music, Brain and Cognition Workshop in The Neural Information Processing Systems* (2007).
- [5] Flossman, S., Grachten, M. and Widmer, G.: Expressive Performance Rendering: Introducing Performance Context, *Proc. Sound and Music Computing (SMC) Conference*, pp.155–160 (2009).
- [6] Widmer, G.: Large-Scale Induction of Expressive Performance Rules: First quantitative Results, *Proc. International Computer Music Conference (ICMC)*, pp.344–347 (2000).
- [7] Widmer, G.: Machine Discoveries: A Few Simple, Robust Local Expression Principles, *Journal of New Music Research*, Vol.31, pp.37–50 (2002).
- [8] Widmer, G.: Discovering Simple Rules in Complex Data:

- A Meta-Learning Algorithm and Some Surprising Musical Discoveries, *Artificial Intelligence*, Vol.146, pp.129–148 (2003).
- [9] Sapp, C.S.: Harmonic Visualizations of Tonal Music, *Proc. International Computer Music Conference (ICMC)*, Havana, Cuba, pp.423–430 (2001).
 - [10] Sapp, C.S.: Comparative analysis of multiple musical performances, *Proc. 8th International Society for Music Information Retrieval (ISMIR) Conference*, pp.497–500 (2007).
 - [11] CHARM Department of Music Royal Holloway: Mazurka Project, AHRC Research Centre for the History and Analysis of Recorded Music (online), available from <http://www.mazurka.org.uk> (accessed 2012-07-02).
 - [12] Saito, S., Kameoka, H., Takahashi, K., Nishimoto, T. and Sagayama, S.: SpecMurt Analysis of Polyphonic Music Signals, *IEEE Trans. Audio, Speech, and Language Processing*, Vol.16, No.3, pp.639–650 (2008).
 - [13] Müller, M., Ellis, D.P.W., Klapuri, A. and Richard, G.: Signal Processing for Music Analysis, Selected Topics in Signal Processing, *Journal of IEEE*, Vol.5, No.6, pp.1088–1110 (2011).
 - [14] Odell, J.J.: The Use of Context in Large Vocabulary Speech Recognition, Ph.D. Thesis, Cambridge University (1995).
 - [15] Rissanen, J.: Universal coding, information, prediction and estimation, *IEEE Trans. Information Theory*, Vol.30, No.4, pp.629–636 (1984).
 - [16] Shinoda, K. and Watanabe, T.: MDL-Based context-dependent subword modeling for speech recognition, *A. Acoustical Society Japan (E)*, Vol.21, No.1, pp.70–86 (2000).
 - [17] CrestMuse Project: CrestMuse PEDB, JST / CREST, JAPAN (online), available from <http://www.crestmuse.jp/pedb> (accessed 2012-07-02).
 - [18] Hashida, M., Matsui, T. and Katayose, H.: A New Music Database Describing Deviation Information of Performance Expressions, *Proc. 9th International Society for Music Information Retrieval (ISMIR) Conference*, pp.489–494 (2008).
 - [19] Repp, B.H.: Expressive timing in a debussy prelude: A comparison of student and expert pianists, *Musicae Scientiae*, Vol.1, No.2, pp.257–268 (1997).
 - [20] de Mantaras, R.L. and Arcos, J.L.: AI and music: From composition to expressive performances, *The AI Magazine*, Vol.23, No.3, pp.43–57 (2002).
 - [21] Kirke, A. and Miranda, E.R.: A survey of computer systems for expressive music performance, *ACM Computing Surveys (CSUR)*, Vol.42, No.1, pp.1–41 (2009).
 - [22] Widmer, G. and Goebel, W.: Computational Models of Expressive Music Performance: The State of the Art, *Journal of New Music Research*, Vol.33, No.3, pp.203–216 (2004).



奥村 健太 (学生会員)

2007 年名古屋工業大学工学部システムマネジメント工学科卒業。2009 年同大学大学院工学研究科情報工学専攻博士前期課程修了。現在、同専攻博士後期課程に在学中。音楽情報処理の研究に従事。電子情報通信学会、日本音

響学会、人工知能学会各会員。



酒向 慎司 (正会員)

1999 年名古屋工業大学知能情報システム学科卒業。2004 年同大学大学院電気情報工学専攻博士後期課程修了。同年東京大学大学院情報理工学研究科特任助手。2007 年名古屋工業大学大学院情報工学専攻助教。博士 (工学)。

音楽情報処理、音声情報処理、手話認識の研究に従事。2009 年度日本情報処理学会山下記念研究賞、2011 年度電子情報通信学会ヒューマンコミュニケーション賞受賞。電子情報通信学会、日本音響学会、人工知能学会、IEEE 各会員。



北村 正 (正会員)

1973 年名古屋工業大学工学部電子工学科卒業。1978 年東京工業大学大学院博士課程修了。同年東京工業大学精密工学研究所助手。1983 年名古屋工業大学工学部電子工学科講師、1984 年同助教。1993～1994 年ウエルズ

大学 (スウォンジー大学) 客員研究員。1995 年名古屋工業大学工学部知能情報システム学科教授。2003 年名古屋工業大学大学院工学研究科教授。博士 (工学)。マルチメディア情報処理、音楽情報処理、感性情報処理、手話認識等の研究に従事。電子情報通信学会、日本音響学会、映像情報メディア学会、IEEE、ISCA 各会員。