

楽譜と表情を関連付けた統計モデルに基づく楽器演奏の類似性指標

Similarity measure for musical performances based on the statistical model that associates expression and notation

奥村 健太 *¹ 酒向 慎司 *¹ 北村 正 *¹
Kenta Okumura Shinji Sako Tadashi Kitamura

*¹名古屋工業大学
Nagoya Institute of Technology

A method for quantitative elucidation of the factors that contribute to the expression of the musical performance is proposed. It is based on a technique to obtain the systematized structure of the generative models that associates the feature quantities of the expression and the combination of musical directions available from the score. It allows the objective analysis of the reasons of the expression in any performances with the different expressions that are rendered even from the score of the same composition. Evaluations showed a utility of the proposed method as a measure to contribute an objective understanding of the factors that were emphasized by the performer for control of expression in the rendering process of the performances.

1. はじめに

楽器演奏の演奏者は楽譜の指示から楽曲の意図を解釈し、楽器の操作に表情の付与を行う。その過程で指示をどのように解釈して操作に反映するかは演奏者に依存するため、同一の楽曲の演奏でも、付与される表情には相違が生じる [Seashore 38]。しかし特定の表情付けにどのような解釈が寄与するかを齟齬のない表現で説明することは、演奏者本人にも容易ではない [Gabrielsson 03]。本稿では、任意の演奏の表情付けにどのような解釈が寄与するかを客観的に分析する問題を考える。その解決に必要なものは、演奏者による表情付けの理由を説明する規則に類するものと、それらの表情付けへの寄与の度合いを示す定量的な指標である。

本提案では演奏が楽譜の指示に基づくことを前提として、表情を構成する特徴量を表す生成モデルに楽譜の指示から一意に得られる情報の組合せを関連付ける。それらの特徴を楽譜の情報に従って分類することで、演奏の表情付けに寄与した規則の組合せを体系化できる。この体系の構築過程に着目することで、任意の演奏において楽譜の各指示が持つ表情付けへの寄与の度合いを獲得するほか、他の指示との関係についての客観的な分析を可能にする。

任意の演奏における表情付けの傾向を分析可能な枠組みとしては、テンポと強弱に関する特徴量の変動の軌跡を平面に写像する手法 [Dixon 02] や、音響信号から観測可能な特徴量の傾向を階層的な図形として描画する手法 [Sapp 01] が提案されている。しかしこれらが扱うのは特徴量の変動のみであり、特定の表情付けがどのような規則に依存するかを説明することは困難である。本提案では一般的な楽譜の指示に基づき記述した規則を用い、その表情付けへの寄与の度合いや、他の規則との関係の分析を行う。この点で本提案と類似するものとして、任意の傾向の表情付けに寄与した規則を演奏から抽出する手法 [Widmer 02, Tobudic 03] も提案されている。しかしこれらの手法では分析対象とする演奏の楽曲構造を入力としており、演奏者の解釈に依存する部分への影響が考えられる。提案手法ではそのような情報の入力を求めることなく体系の構築を行うため、演奏者以外の意図が混入する可能性を排除できる。

また、実際の演奏では単一の楽曲の中でも複数の表情の使い分けが生じており [Repp 98]、各規則がその使い分けにどのように関係するのかについても説明される必要がある。提案手法によって得られる体系は表情付けに寄与した規則をノードに持つ木構造として構築されるため、規則間の関係性も説明できる。

2. 提案手法

2.1 演奏表情の特徴量の統計モデリング

本提案では楽譜の内容を MusicXML で記述可能な指示の系列、それに対する演奏者の演奏の内容を MIDI イベントで記述可能なシンボリックな事例の系列と考える。実際の演奏における任意の事例を y_n 、その生成に際して演奏者が従った指示に表情を付けず正確に演奏した場合の MIDI イベントを x_n とすると、表情として観測可能な任意の要素 f についての特徴量 $o_{n,f}$ は次式から得られる：

$$o_{n,f} = \begin{cases} y_{n,f} - x_{n,f} & (f \in F = \{\text{AT}\}), \\ y_{n,f} / x_{n,f} & (f \in F = \{\text{GR}, \text{DR}, \text{BR}\}). \end{cases} \quad (1a) \quad (1b)$$

任意の事例は一音符または一拍を観測単位とし、その表情の特徴を構成する要素群を F とする。ここでは鍵盤楽器の操作に関するものをその構成要素とし、以下のように定義する：

AT (Attack Timing) 拍長を基準とする、打鍵時刻の差。

GR (Gatetime Ratio) 音符楽譜上の音価に対する演奏から計測された、打鍵から離鍵までの継続長の比率。

DR (Dynamics Ratio) 記譜された dynamics の値と演奏された MIDI velocity の値との比率。

BR (Local BPM Ratio) 指示に対して拍単位で変動を算出した局所的な BPM の比率。

これらは音符または拍に依存し、 F_{note} , F_{beat} に分類される：

$$f \in \begin{cases} F_{\text{note}} = \{\text{AT}, \text{GR}, \text{BR}\}, & |F_{\text{note}}| = 3, \\ F_{\text{beat}} = \{\text{BR}\}, & |F_{\text{beat}}| = 1. \end{cases} \quad (2a) \quad (2b)$$

連絡先: 奥村健太, 名古屋工業大学, 名古屋市昭和区御器所町,
k09@mmsp.nitech.ac.jp

任意の演奏として生成された事例群の特徴量について、本提案ではその傾向が正規分布に従うと考える。それにより、 F_{note} または F_{beat} に関する各特徴量の確率密度関数は、次式に示す多変量正規分布として表せる：

$$P(\mathbf{o}_n | \boldsymbol{\mu}, \boldsymbol{\sigma}^2) = \prod_{f \in F} \left(\frac{1}{\sqrt{2\pi\sigma_f^2}} \exp\left(-\frac{(o_{n,f} - \mu_f)^2}{2\sigma_f^2}\right) \right). \quad (3)$$

ここで μ_f , σ_f^2 は、 f に関するモデルのパラメタ（平均、分散）である。また、左辺の \mathbf{o}_n は F_{note} の場合 $\mathbf{o}_n = [o_{n,\text{AT}}, o_{n,\text{GR}}, o_{n,\text{DR}}]$, F_{beat} の場合 $\mathbf{o}_n = [o_{n,\text{BR}}]$ のベクトルであり、右辺の $o_{n,f}$ は各場合のベクトルに含まれる任意の要素についてのスカラーを示す。

2.2 楽譜の指示に基づく表情付けの規則の体系化

楽譜の局所的な指示の系列を文脈に見立てると、それに従って演奏された事例の表情付けには、その周囲も含む局所的な指示の寄与が考えられる。本提案では y_n のモデルに (x_{n-1}, x_n, x_{n+1}) の指示に関する情報を関連付け、これを文脈依存モデルと呼称する。任意の事例とその指示を一意に対応させるため、各事例の文脈依存モデルには他の全事例との区別が可能な程度に精緻な情報の組合せを関連付ける。その文脈要素には、鍵盤楽器の楽譜から得られる情報として一般性が高いものが望ましい。本提案では先行研究 [奥村 13, Okumura 14] に倣い、音符の場合は主に和声関連、拍の場合は主にリズム関連、両者で共通して小節・楽曲内での時間的位置などの要素を用いる。各場合の文脈要素の数はそれぞれ $|T_{\text{note}}| = 12$, $|T_{\text{beat}}| = 4$ とする。

任意の表情が付与される条件は、その演奏を指示した楽譜から得られる規則の組合せに基づくと考えられる。本提案では文脈依存モデルを特徴量の傾向に従って分類した木構造を構築する手法 [Odell 95] を用いる。この手法では分類基準の候補に文脈要素に関する規則を用い、分類の結果を二分木 $S = \{s_1, \dots, s_M\}$ として表現する。そのルートノードから各リーフノードに至るまでのあらゆる中間ノード群 $B = \{b_1, \dots, b_K\}$ には、それぞれ分類基準となった規則が保持されている。この木構造を辿ることで、任意の表情付けに寄与する規則の組合せだけでなく、各規則の優先順位についても一意な記述が可能となる。

この木構造の構築に際し、まず $O = (o_{1,F}, \dots, o_{n,F}, \dots, o_{N,F})$ を全学習データとした文脈依存モデルが与えられた場合を考える。 $\sum_{n \in N} \gamma_{s_m}(\mathbf{o}_{n,F})$ を s_m の学習データの総量、 μ_m , σ_m を s_m の平均、分散とすると、これら全ての共有を仮定したノード s_m の対数尤度 $L(s_m)$ は次式で定義される：

$$L(s_m) = -\frac{1}{2} \left(\log\left((2\pi)^{|F|} \sigma(s_m)\right) + |F| \right) \sum_{n \in N} \gamma_{s_m}(\mathbf{o}_{n,F}). \quad (4)$$

次に、 s_m に含まれるモデル群を $s_{m,\text{no}}$ と $s_{m,\text{yes}}$ に分割する場合を考える。任意の文脈要素 $t \in T$ に関する規則群 $Q_t = \{q_{t,1}, \dots, q_{t,u}, \dots, q_{t,U_t}\}$ から $q_{t,u}$ を選択して s_m を分割し、中間ノード b_m を生成する。その選択の有効性は、次式に示す分割前後の対数尤度の向上量 $\Delta L_q(s_m)$ によって評価できる：

$$\Delta L_{q_{t,u}}(b_m) = L(s_{m,\text{no}}(q_{t,u})) + L(s_{m,\text{yes}}(q_{t,u})) - L(s_m). \quad (5)$$

この分割を全ての文脈要素のあらゆる規則 $\sum_{t \in T} Q_t$ の適用を仮定して試行し、 $\Delta L_{q_{t,u}} \rightarrow \max$ となる $q_{t,u}$ を選択する。対数尤度が増加する $q_{t,u}$ が存在する限り、 $s_{m,\text{yes}}$ および $s_{m,\text{no}}$ において更にノードの分割を繰り返す。

2.3 木構造に基づく文脈要素の評価

ノードの分割を繰り返すことで、任意の事例の文脈依存モデルは最終的に木構造の末端にあるリーフノードのいずれかに分類される。その状態の木構造は学習データの量に応じて極めて大規模なものとなるため、直接的な観測には適さない。しかしこの木構造の構築過程において、ルートノードに比較的近い段階のノード分割で選択され、あるいはより大きな対数尤度の向上量をもたらす規則は、演奏においても表情付けの傾向をより明確に分岐させたものと考えられる。そこで本提案では、任意の t が持つ貢献度に類するものを対数尤度の向上量に基づいて定義し、それらの値を各規則の表情付けへの寄与度合いの分析に用いる。

B の中で Q_t に含まれる質問 $q_{t,u}$ のいずれかを分類基準に持つ中間ノード群を $B_t = \{b_{t,1}, \dots, b_{t,K_t}\}$ とし、 O の表情付けにおける t の貢献度を c_t とする。以下に示すように、 c_t は B_t に含まれるあらゆる中間ノードに対して式 (5) を適用した総和として得られる：

$$c_t = \sum_{b_{t,k} \in B_t} \Delta L_{q_{t,u}}(b_{t,k}). \quad (6)$$

以上から得られた $[c_1, \dots, c_T]$ の値に主成分分析 (PCA) を適用することで、各文脈要素の表情付けへの寄与の度合いを定量的に評価できる。それによって得られた各文脈要素についての因子負荷量の傾向からは、文脈要素間の相対的な関係の分析も可能となる。また、演奏ごとに構築した木構造から各貢献度を求め、それらの主成分得点を低次元の空間に写像した結果は、複数の演奏間の客観的な距離尺度としても有用と考えられる。尚、木構造の規模は学習データの量に依存し、 c_t の値も木構造の規模によって変動する。そこで本稿では、各文脈要素についての評価値の合計 $\sum_{t \in T} c_t$ を正規化の上で分析を行う。

3. 評価実験

提案手法の評価に用いた演奏の組合せを表 1 に示す。各データセットは統一された作曲者による楽曲の組合せとし、表情の特徴量は全て同一のデータベース *1 から取得している。

3.1 主成分分析による文脈要素の寄与度の縮約

ここではデータセット M(1), C(1) の各演奏を単位として構築した木構造群から各文脈要素の貢献度を算出し、これらに PCA を適用した場合の有効性を検証する。各データセットで得られた累積寄与率の傾向を、第一主成分から第四主成分まで (PC1, ..., PC4) それぞれ図 1 に示す。

M(1) では F_{note} で 0.8 以上、 F_{beat} で 0.9 以上の累積寄与率を PC2 までで得られており、平面への写像でも各演奏の関係を概ね良好に説明できるといえる。一方、C(1) の累積寄与率は特に F_{note} において M(1) を下回っている。ここから、C(1) における F_{note} の表情付けには、M(1) に比較してより複雑な解釈の相違が生じていると考えられる。そのような相違を生じる主要因が演奏者と楽曲のどちらにあるかは、特徴量群別の分析からの考察を要する。

3.2 文脈要素の寄与度合いと関係の分析

M(1), C(1) の各演奏ごとに木構造を構築し、各セットについて PCA を適用した結果を図 2 に示す。ここでは PC1 を横軸、PC2 を縦軸とし、両軸における各文脈要素の因子負荷量のベクトルを赤色の矢印で表す。各演奏は曲名と演奏者名の略称で記し、同じ演奏者によるものを同一色で結線している。

*1 CrestMusePEDB Version 2, <http://www.crestmuse.jp/pedb/>

表 1: 演奏データセット

Dataset	Compositions (abbreviation)	Performers (abbreviation)	Composer
M(1)	Sonata K. 279 Mv. 1, 2, 3 (S279-1, 2, 3).	G. Gould (GG), M. Pires (MP).	W. Mozart
C(1)	Nocturne No. 2, 10 (N002, N010), Prelude No. 7, 20 (P007, P020).	V. Ashkenazy (VA), M. Pires (MP).	F. Chopin

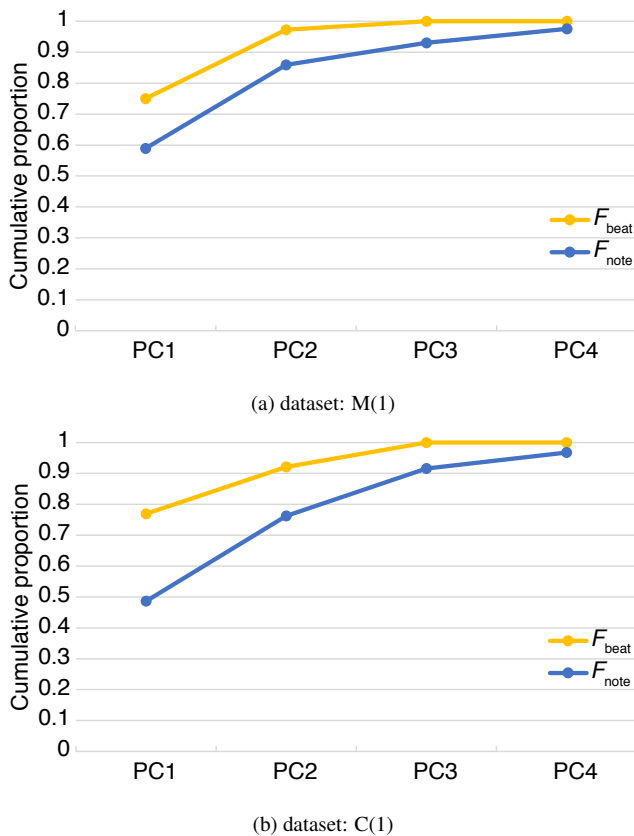


図 1: 主成分分析の累積寄与率

M(1) では各演奏への寄与の度合いの大きな文脈要素の傾向が演奏者に依存して明確に分離しており、演奏者 GG では文脈要素 global (楽曲内の時間的位置), type (音価), 演奏者 MP では staff (大譜表の上下段), syllable (音名) などの寄与が認められる。同一の楽曲の演奏は各演奏者がとりうる範囲内で比較的類似した位置に分布しているものの、それらの距離は必ずしも近くない。よって M(1) の楽曲には、演奏者の比較的自由な解釈を許容する余地があると考えられる。一方、C(1) では各楽曲が演奏者を問わず比較的近い距離に分布している。よって C(1) の各演奏における解釈では、楽曲による楽譜の指示が演奏者自身の意図に優先していると考えられる。

文脈要素ごとの傾向では、global が横軸 (PC1) 方向に比較的高い因子負荷量を有する場合が多い。ここから、表情付けの傾向の変動には楽曲内の時間的位置への比較的強い依存が示唆される。本稿では楽曲の演奏を単位として木構造の構築を行ったが、楽曲の時間的位置に合わせて分割した事例群に基づく木構造を比較することで、その演奏者が着目した楽譜の情報や、その時間的変化を詳細に説明できる可能性がある。一方、縦軸 (PC2) 方向に比較的高い因子負荷量を有する文脈要素は、 F_{beat}

の場合では pattern (発音パターン^{*2}), density (発音音符数) となっており、 F_{note} の場合では staff, voice (声部), type など挙げられる。これらの文脈要素に関する指示は単一の音符や拍を対象として完結し、同一の指示が楽曲中で複数回出現するため、global と比較して高い局所性を有する文脈要素といえる。以上から、図 2 の各平面を構成する二軸には、主に楽譜の大域的指示と局所的指示に関する文脈要素がそれぞれ寄与していると考えられる。

尚、両セットに共通して含まれる MP の傾向からは、各セットの間で大きく異なる文脈要素の寄与が認められる。このような傾向が MP に固有のものであるのか、あるいは多数の演奏者が一般的に同様の傾向を持つのかについて、本稿で用いたデータセットから確認することは困難である。この確認には、より多様かつ統一された楽曲と演奏者の組合せでの検証を要する。

4. おわりに

本稿では、楽器演奏の表情付けに寄与する各種規則の客観的な分析を目的とし、その寄与度合いや他の規則との関係、ならびに他の演奏との類似性を平面上に写像する手法を提案した。

評価実験からは、各特徴量の表情付けに寄与する文脈要素に、それぞれ演奏者や楽曲に依存して異なる傾向が示された。但し、本稿で用いた演奏の組合せや評価の観点は限定的である。任意の演奏者による多数の演奏や、任意の楽曲における多数の演奏者の組合せなど、特定の条件下での詳細な分析が今後の課題となる。表情付けに寄与する規則の構成要素は楽譜の指示以外にも考えられるが、個々の演奏者の経験に依存する部分が多い。その過程を説明する尺度として一般的に機能する要素についても、今後慎重に検討を進める。

謝辞

本研究は JSPS 科研費 (26730182) および財団法人電気通信普及財団の助成を受けた。

参考文献

- [Dixon 02] Dixon, S., Goebel, W., and Widmer, G.: The performance worm: Real time visualisation of expression based on Langrer's tempo-loudness animation, in *Proc. International Computer Music Conference (ICMC)*, pp. 361–364, Goteborg, Sweden (2002)
- [Gabrielsson 03] Gabrielsson, A.: Music performance research at the millennium, *Psychology of Music*, Vol. 31, No. 3, pp. 221–272 (2003)
- [Odell 95] Odell, J. J.: *The Use of Context in Large Vocabulary Speech Recognition*, PhD thesis, Cambridge University (1995)

*2 声部別の音符の有無を打鍵位置で記述したもの。

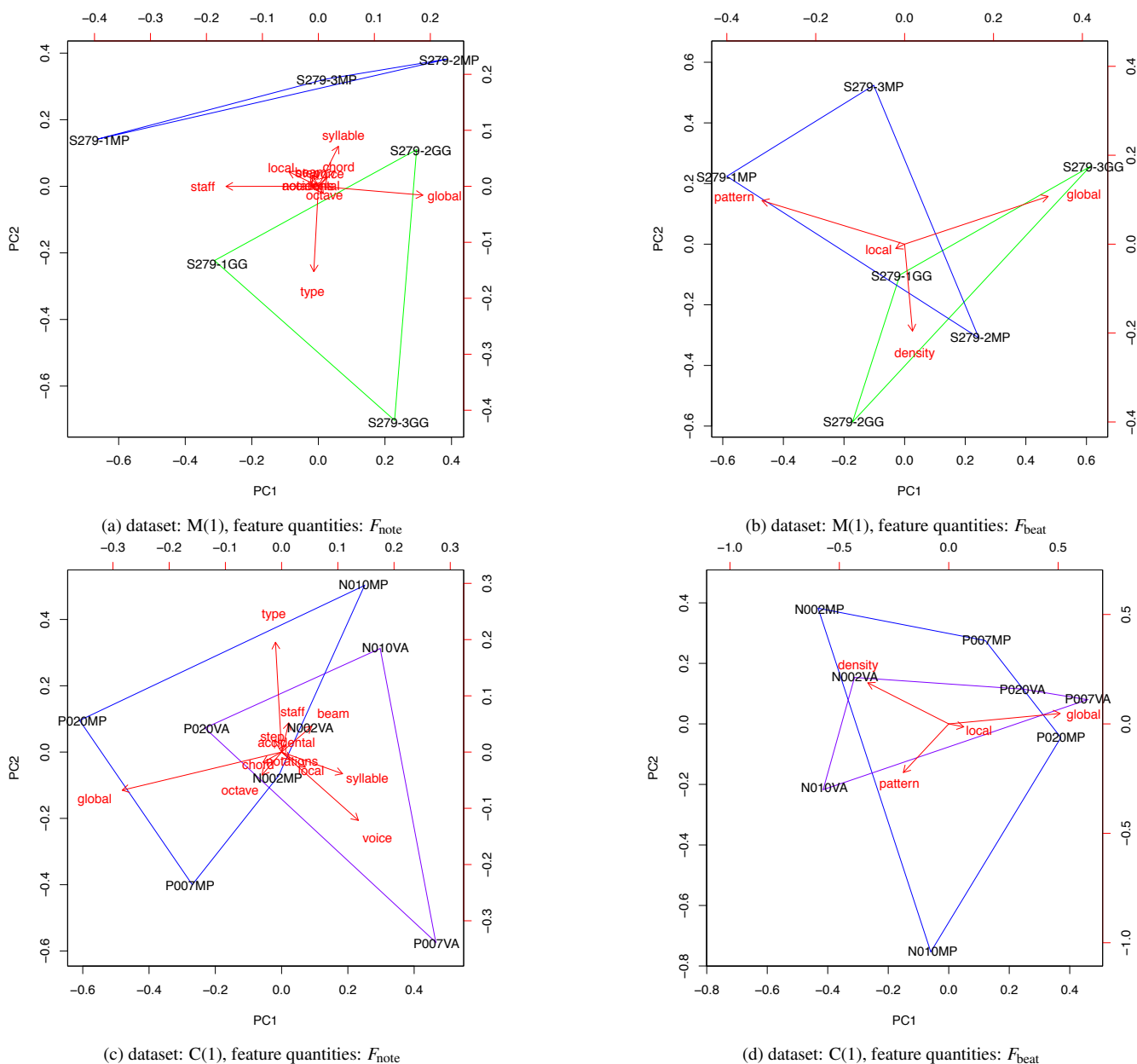


図 2: 演奏の分布と各文脈要素の表情付けへの寄与傾向

[Okumura 14] Okumura, K., Sako, S., and Kitamura, T.: Laminiae: A stochastic modeling-based autonomous performance rendering system that elucidates performer characteristics, in *Proc. Joint ICMC - SMC 2014*, pp. 1271–1276, Atehs, Greece (2014)

[Repp 98] Repp, B. H.: A microcosm of musical expression: I. Quantitative analysis of pianists’ timing in the initial measures of Chopin’s Etude in E major, *Journal of the Acoustical Society of America*, Vol. 104, No. 2, pp. 1085–1100 (1998)

[Sapp 01] Sapp, C. S.: Harmonic Visualizations of Tonal Music, in *Proc. International Computer Music Conference (ICMC)*, pp. 423–430, Havana, Cuba (2001)

[Seashore 38] Seashore, C. E.: *Psychology of Music*, Courier Dover Publications (1938)

[Tobudic 03] Tobudic, A. and Widmer, G.: Relational IBL in Music with a New Structural Similarity Measure, in *Proc. the 13th International Conference on Inductive Logic Programming*, pp. 365–382 (2003)

[Widmer 02] Widmer, G.: Machine Discoveries: A Few Simple, Robust Local Expression Principles, *Journal of New Music Research*, Vol. 31, pp. 37–50 (2002)

[奥村 13] 奥村 健太, 酒向 慎司, 北村 正: 楽譜に基づく鍵盤楽器演奏の統計的モデル化手法, *情報処理学会論文誌*, Vol. 54, No. 4, pp. 1288–1301 (2013)