原著論文

統計モデルと楽譜情報に基づく指標を用いた演奏表情の比較分析

奥村 健太, 酒向 慎司, 北村 正

名古屋工業大学

Comparative Analysis of Performance Expression using Similarity Metrics based on Statistical Model and Musical Score Information

Kenta OKUMURA, Shinji SAKO and Tadashi KITAMURA

Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya, Aichi, 466-8555, Japan

Abstract : There is a complex phenomenon in the performance expression of Western music. It is difficult even for performers themselves to describe the causality of “score directions” and “instrument operation” clearly and objectively. In this paper, we propose a method that can help to elucidate the factors contributing to performance expression. It is based on a technique to obtain the systematized structure of the statistical models. The model associates the feature quantities of performance expression and the combination of simple musical directions that can be surely retrieved from the score. In this method, degrees to which those directions contribute to the performance expression are also defined. It allows the application of standard multivariate analysis techniques for comparative analysis of the expressions between performances. Evaluation showed the utility of the proposed method in contributing to an objective understanding of the factors that performers emphasized for control of expression in their performances.

Keywords : Performance expression, Statistical model, Multivariate analysis

1. は じ め に

一般的な楽器演奏において、演奏者は楽譜の指示から楽曲の意図を解釈し、相応しい表情を付与して楽器を操作する[1]。楽曲が同一の場合は演奏者への指示も共通であり、付与される表情の特徴には複数の演奏者間で類似性がある[2]。しかし演奏の過程で指示をどのように解釈して楽器の操作に反映するかは演奏者の専門知識や感性に依存しており、実際の演奏に付与される表情には、演奏者に依存した部分的な相違もある[3, 4]。解釈に寄与した要因には演奏者の個人性のほか、楽曲の影響も含まれている。そのため複数の演奏間では、演奏者や楽曲が同一であっても、解釈は必ずしも同一ではない[5]。その要因を明確かつ客観的に説明することは、演奏者自身にも困難である。本稿では、複数の演奏において解釈に寄与した要因を明確に説明し、演奏間の表情の類似性を客観的に比較する問題を考える。その解決のため、楽譜から一意に得られる情報の組合せを用い、演奏間で表情の比較分析を行うための手法を提案する。これにより、楽器演奏に限らず、特定の個人らしさを含む行動と要因との関係を明確かつ客観的に説明することに貢献できる。ここで必要となるのは、演奏者が付与した表情の理由を表す規則に類するものと、規則ごとの表情への寄与の度合いを示す指標である。

筆者らは表情の特徴量を近似した生成モデルに楽譜から得られる情報を関連付け、各特徴量の生成に寄与した規則に類するものの体系を楽譜の指示に基づいて構造化する手法を提案してきた[6]。目的は、演奏者の解釈の構成要素を一般性

の高い指標によって代替的に表現することにある。この手法は特定の演奏者が持つ表情の特徴を備えた演奏の自動生成[7]にも応用され、これに基づく自動演奏表情付けシステムは、大規模合同評価実験において他のシステムを上回る良好な性能を示している[8]。しかしここで得られる演奏の表情は、既存の演奏に含まれる特徴を反映したものである。利用者の好みを反映した表情の付与を考えると、演奏における表情と指示の関係を分析できることは、自動生成される演奏の品質に貢献する知見の獲得につながる。先行研究では特定の演奏者や楽曲の特徴に依存した体系構造の調査を行っており、楽器演奏に関する経験的な知見を客観的に説明できる可能性が示されている[9]。しかしこの手法で得られる体系構造は演奏の特性に依存して異なる。各演奏に基づく体系構造の直接的な比較などからでは、複数の演奏間で表情に寄与する指示の相違を定量的に分析することは困難である。本稿ではこの体系構造の構築過程に着目し、任意の演奏において特定の指示が持つ表情への寄与の度合いを定める。これにより、表情の付与における各指示の関係の定量的な明示と、複数の演奏間におけるそれらの比較分析を実現する。

演奏に付与された表情を直観的に図示できる枠組みとして、テンポや強弱に関する特徴量の変動の軌跡を描画する手法[10-12]が提案されている。また、音響信号から抽出される特徴量の傾向を類似性に基づき可視化する手法[13]では、演奏間での傾向の比較分析も既に行われている[14]。しかしこれらの手法では特徴量と指示との関係が説明されないため、特定の表情の付与に至る要因の分析が困難である。提案手法では個々の観測において抽出された特徴量とその

指示に関する情報を関連付け、一意な識別を可能にする。この点で類似する枠組みとして、特定の表情に寄与した規則を演奏から自動的に抽出する手法も提案されており、楽譜の情報に基づく分析が行われている[15, 16]。しかしこれらの手法では、分析対象となる演奏の楽曲構造を予め手動で分析して入力する必要がある。提案手法ではそのような情報を用いず、分析の過程で演奏者以外の影響の混入を抑止する。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るため、抽出された各規則について、表情の各傾向にそれぞれどのように寄与するのか説明される必要がある[17]。提案手法ではそのような規則を抽出するほか、関係性も説明する。

以降、第2章では本稿で用いる表情の構成要素を定め、それらの特徴量を演奏から取得する手続きを示す。また、各特徴量の生成過程に寄与した規則を関連付けた統計モデルを定義する。これらは先行研究として最新の応用例[8]に基づく。第3章では表情に寄与する規則の体系を楽譜の指示に基づき構造化する手続きを示す。また、その体系構造を用いることで各規則の関係性を楽譜の指示に基づき分析する手法を提案する。第4章では提案手法の有効性を評価する。第5章では以上から得られた成果をまとめ、今後の課題を述べる。

2. 表情の構成要素の定義と演奏の生成過程のモデル化

2.1 表情の特徴量の統計モデリング

演奏に付与された表情の特徴量は、表情が一切付与されていない演奏との比較から抽出できる[18]。本提案では演奏者や独奏楽曲の豊富さなどの利点から対象を鍵盤楽器に限定し、楽譜の内容をMusicXML[注1]で記述可能な指示（以後「指示」と称する）の系列、それに対する演奏者の演奏の内容をStandard MIDI FormatのMIDIイベントで記述可能な事例（以後「事例」と称する）の系列と考える（音符にはノートイベント、拍にはメタイイベントのセットテンポを使用）。鍵盤楽器の操作において各事例の表情の特徴を構成する個別の要素を f とし、各要素と特徴量の概要を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 (Beats Per Minute, 音楽で演奏のテンポを表す単位) の比率。

ここでは f の集合を F とし、次式に従い、音符の F_{note} には3個、拍の F_{beat} には1個の f 进行分类する：

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

$$(1b)$$

表情が付与された演奏の事例の系列 $(y_1, \dots, y_N)$ において、任意の音符または拍の事例を y_n 、その指示に正確な演奏をした場合を x_n とし、 f についての特徴量 $o_{n,f}$ を次式で表す：

$$o_{n,f} = \begin{cases} y_{n,f} - x_{n,f}, & (f \in F = \{\text{AT}\}), \\ y_{n,f} / x_{n,f}, & (f \in F = \{\text{GR}, \text{DR}, \text{BR}\}). \end{cases} \quad (2a)$$

$$(2b)$$

先行研究[9]から、特徴量には楽譜の指示に比較的忠実な値を中心とした正規分布に近い形状として観測される傾向が示されている。本提案ではこの傾向に基づき、特徴量は正規分布に従うと仮定する。同様の仮定に基づく設計は、自動演奏表情付けに関する既存手法[19, 20]にもみられる。これにより、 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：

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

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 o_n は F_{note} の場合 $o_n = [o_{n,\text{AT}}, o_{n,\text{GR}}, o_{n,\text{DR}}]$ 、 F_{beat} の場合 $o_n = [o_{n,\text{BR}}]$ のベクトル、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 表情の特徴量と楽譜の指示の関連付け

各事例に付与された表情には、単一の音符や拍の周囲を含む、局所的な文脈に類する指示の系列の寄与が考えられる。本提案では式(3)に従って特徴量を近似した統計モデルに対してそのような指示の文脈に関する情報の組合せ（以後「文脈情報」と称する）を関連付けることで、楽曲解釈との関係を説明することを考える。そこで先行研究[8]に倣い、「文脈依存モデル」を定義する。図1に示すように、 y_n の特徴量に (x_{n-1}, x_n, x_{n+1}) の指示に関する文脈情報を関連付ける。これによって事例と指示との一対一の対応関係を説明するほか、他の全事例との区別を可能にする。事例の表情に寄与した規則を一般性の高い情報で記述するため、文脈情報の構成要素（以後「文脈要素」と称する）には、鍵盤楽器の楽譜の

図1 音符と拍の事例における特徴量と文脈情報の関連付け

統計モデルと楽譜情報に基づく指標を用いた演奏表情の比較分析

指示として、楽曲を問わず得られるものが望ましい。音符の場合は主に和声に関するもの、拍の場合は主にリズムに関するもの、両者に共通するものとしては小節・楽曲内での時間的位置などが挙げられる。ここでは文脈要素の集合を T とし、音符の T_{note} には12個、拍の T_{beat} には4個の文脈要素を定義する。それらの内訳を表1に示す。各文脈要素についての記述に用いる情報の定義は、MusicXMLの仕様に従う。

事例の文脈依存モデルに関連付けられた文脈情報は、その演奏に与えられた指示の内容を符号化したものとなる。特定の事例の表情にどのような楽譜の指示が寄与したのかを文脈情報によって説明することを考えた場合、有効な手法の一つに、決定木に基づくコンテキストクラスタリング [21] がある。この手法によって、文脈依存モデルの集合に対して文脈情報に関する分類基準を用いたトップダウンなクラスタリングを行うことで、図2に示すような木構造を構築できる。この木構造に着目すると、リーフノード群 $S = \{s_1, \dots, s_M\}$ を除く全てのノード群 $B = \{b_1, \dots, b_K\}$ には、文脈依存モデルの集合を分割する際に選択された分類基準がそれぞれ保持されている。各事例の文脈依存モデルはいずれかのリーフノードに分類されており、ルートノードからそこに至るまでの各ノードに含まれる分類基準を辿ることで、事例ごとの表情に寄与した規則に類するものを一意に記述できる。本提案では先行研究 [8] に従い、個々の分類基準を「音価は四分音符

よりも長いのか?」や「拍の位置は一拍目であるのか?」など、特定の文脈要素に関する単純な二択の質問とし、それらの組合せによって特定の表情に寄与する規則を表現することを考える。これらの質問は各文脈要素と不可分な関係にあり、設計は各要素の定義の細かさ (表1参照) に依存する。音符に関する12要素については合計3060個、拍に関する4要素については合計1513個の質問を定義できる。

3. 表情に寄与する楽譜の指示の評価指標

2.2節で述べた手法に基づく木構造の構築過程に着目すると、各ノードに保持される質問は、それによって文脈依存モデルの集合を二分した場合における分割前後の対数尤度の向上量を評価値に用い、あらゆる候補の中から最も高い評価値を得たものが選択されている。木構造上でルートノードに比較的近い位置のノードで選択される質問は、文脈依存モデルの集合をより明確に分割し、比較的高い評価値を持つ。そのような質問に基づく基準は、木構造上でそのノード以下に分類された事例の表情にもより強く寄与したものと考えられる。また、文脈依存モデルをすべて分割した木構造中には、同一の文脈要素に関連し、基準の相違がわずかな質問が選択されたノードが複数存在する可能性がある。楽曲の解釈に寄与した指示の傾向を表現する上で、それらを全て個別に扱うことは現実的ではないため、任意の文脈要素に関する質問群をそれぞれある程度類似した楽譜の指示の系統とみなすことを考える。以上から、ここでは各文脈要素に関する質問群が持つ表情への貢献度に類するものを、質問が選択された際の評価値に基づいて定義する。また、それらの貢献度を用いることで、複数の演奏における表情の類似性を各文脈要素の関係性に基づいて分析する手法を提案する。

木構造の構築に際し、まず $O = (o_{1,F}, \dots, o_{N,F})$ を学習データとした文脈依存モデル群が与えられた場合を考える。全ての文脈依存モデル群の共有を仮定したノードを s_m 、その学習データの総量を $\sum_{n \in N} \gamma_{s_m}(o_{n,F})$ 、平均、分散を μ_m, σ_m とすると、 s_m の対数尤度 $L(s_m)$ は次式で定義される：

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

ここで $|F|$ は特徴量の要素数を示し、学習データの事例が音符の場合は $|F_{\text{note}}| = 3$ 、拍の場合は $|F_{\text{beat}}| = 1$ である。

次に、 s_m に含まれる文脈依存モデル群を $s_{m,\text{no}}$ と $s_{m,\text{yes}}$ に分割する場合を考える。任意の文脈要素 t に関する質問の数を U_t とすると、これは $\sum_{i \in T} U_i$ 個のあらゆる質問の中から t に関する質問 $q_{t,u}$ を用いて s_m を分割し、ノード b_m を生成する手続きとなる。この際に文脈依存モデル群の分類基準として $q_{t,u}$ を選択することの有効性は、次式に示す分割前後の対数尤度の向上量 $\Delta L_{q_{t,u}}(b_m)$ で評価される：

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

表1 文脈依存モデルに関連付ける楽譜の指示

文脈要素 t	概要 (定義)	関連付け
syllable	階名 (12 種類)	T_{note}
step	音名 (7 種類)	T_{note}
accidental	変化記号 (5 種類)	T_{note}
octave	音域 (オクターヴ単位)	T_{note}
chord	和音構成音 (上下関係) または単音	T_{note}
type	音価 (64 種類)	T_{note}
staff	大譜表の段 (上下)	T_{note}
beam	付尾の形状と向き (6 種類)	T_{note}
local	小節内での時間的位置 (100 分割)	$T_{\text{note}}, T_{\text{beat}}$
global	楽曲内での時間的位置 (100 分割)	$T_{\text{note}}, T_{\text{beat}}$
voice	声部 (4 種類)	T_{note}
notations	特殊記号 (47 種類)	T_{note}
density	拍内発音数 (音符数依存)	T_{beat}
pattern	拍内発音パターン (声部別, 24 分割)	T_{beat}

図2 文脈依存モデルの分類による木構造の構築

ノードの分類基準は、予め定義されたあらゆる質問について式(5)に従い尤度向上量を算出し、それらの中から $\Delta L_{q,t,u} \rightarrow \max$ となるものを選択する。分割後の $s_{m,no}$ と $s_{m,yes}$ においても、それらに含まれる文脈依存モデル群を対象に、あらゆる質問の中から最適な分類基準を選択する手続きを反復する。以上が木構造の構築過程となる。

ここで O の表情における t の貢献度を c_t とする。 O に基づく木構造に含まれる B の中で t のみに関する質問を持つ K_t 個のノード群を $B_t = \{b_{t,1}, \dots, b_{t,K_t}\}$ とし、 B_t に含まれるあらゆるノードに対して式(5)を適用した結果を $\Delta L_t = \{\Delta L_{t,1}, \dots, \Delta L_{t,K_t}\}$ とすると、 c_t は次式で得られる：

$$c_t = \sum_{k=1}^{K_t} \Delta L_{t,k}. \quad (6)$$

以上から得られた各文脈要素の貢献度 $[c_1, \dots, c_T]$ の値により、その木構造における各文脈要素の表情への寄与の度合いを評価できる。任意の演奏を単位として木構造を構築し、これらの貢献度を求めることで、複数の演奏間における各文脈要素の表情への寄与にそれぞれどのような相違があるかを比較するための指標を得られると考えられる。なお、各演奏の木構造の規模は学習データ量に依存して異なるため、第4章の各評価実験では、すべての文脈要素の貢献度の合計値を $\sum_{t \in T} c_t = 1$ として正規化している。

各文脈要素の貢献度を用いた演奏の分析には、典型的な多変量解析手法を適用できる。ここでは主成分分析(principal component analysis, PCA)や、多次元尺度構成法(multi-dimensional scaling, MDS)など、付与された表情の類似性に基づく複数の演奏の関係を低次元の空間に写像できる手法を用いる。これらの適用によって得られる空間内での各演奏の配置は、他の演奏との表情の類似性を表す尺度としても有用となりうる。

4. 評価実験

演奏間における表情の比較分析を行うデータセットとして、作曲者ごとに類似性の高い楽曲群について数名程度の演奏者による演奏の組合せをデータセットとして定義した。内訳を表2に示す。M(2)はM(1)の演奏者による異なる楽曲、C(2)はC(1)の楽曲についての異なる演奏者の演奏であり、これらを組み合わせたセットをM(1+2)、C(1+2)とする。各演奏における表情の特徴量はCrestMuse PEDB

表2 演奏データセット

Dataset	Compositions (abbreviation)	Performers	Composer
M(1)	Sonata K.279 Mv.1-3 (S279).	G. Gould (GG),	W. Mozart
M(2)	Sonata K.310 Mv.1 (S310), Sonata K.331 Mv.1 (S331), Sonata K.545 Mv.1-3 (S545).	M. Pires (MP).	
B(1)	Sonata No.8 Mv.2, 3 (S008).	C. Arrau (CA), V. Ashkenazy (VA), W. Backhaus (WB).	L. Beethoven
C(1)	Nocturne No.2 (N002), Nocturne No.10 (N010), Prelude No.7 (P007), Prelude No.20 (P020).	V. Ashkenazy (VA), M. Pires (MP).	F. Chopin
C(2)	Prelude No.7 (P007), Prelude No.20 (P020).	M. Argerich (MA), C. Arrau (CA), R. Blechacz (RB), J. Bolet (JB), A. Cortot (AC), T. Dang (TD), I. Pogorelich (IP). M. Argerich (MA), A. Cortot (AC).	

Version 2 [注2] から取得し、文脈依存モデルの学習に用いた。このデータベースには、著名な演奏者による名演奏の音響信号から抽出された音符ごとの表情やBPMの変動量が記述されている。楽曲の長さや内容はこのデータベースに従い、各演奏の学習にはその記述内容をすべて用いた。なお、各多変量解析手法の適用にはR言語を用いた。

4.1 多変量解析の適用における文脈要素の貢献度の有効性

第3章で、各文脈要素の貢献度を指標とした演奏の分析に際して多変量解析手法を適用し、それらの関係を低次元の空間に縮約できることを述べた。本節ではPCAの適用に際し、分析対象として低次元への縮約が有効な演奏の組合せの規模を検証する。表2の各データセットで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、データセットごとにPCAを適用した。各データセットで得られた F_{note} 、 F_{beat} の累積寄与率について、低次の第一主成分(PC1)から第四主成分(PC4)までを図3に示す。

M(1)、B(1)、C(1)の各特徴量では、PC2までで0.8程度の累積寄与率が得られている。一方、M(1+2)とC(1+2)で同等の累積寄与率を得るにはPC3までの考慮を要する。この結果から、M(1)、B(1)、C(1)ではPCAの適用結果を平面に写像した場合でも演奏間の関係を概ね良好に表現でき

図3 主成分の累積寄与率

統計モデルと楽譜情報に基づく指標を用いた演奏表情の比較分析

と考えられる。演奏間における表情の類似性や文脈要素の関係性を明確に説明する上で、より低次元な空間に写像を行えることが望ましい。そこで、以降はこれらのデータセットを用いた分析を通じて提案手法の有効性の評価を行うものとする。

F_{beat} の特性として、 $M(1+2)$ では $M(1)$ に対して PC1 での累積寄与率の低下が目立つ。ここから、演奏間における F_{beat} の関係性は楽曲の増加によって表現が困難になっているといえ、拍に関する特徴量 BR には特定の楽曲への依存が考えられる。一方、 F_{note} については $C(1)$ や $C(1+2)$ での累積寄与率の低さが目立つ。それらの演奏間では、演奏者が F_{note} に関する表情を制御する際に寄与する各文脈要素の傾向に複雑な相違が考えられる。次節では各演奏を単位とした比較を行い、そのような相違を生じた要因を分析する。

4.2 文脈要素の寄与の度合いと関係性の比較分析

$M(1)$, $B(1)$, $C(1)$ の各演奏を単位として木構造を構築し、データセットごとに PCA を適用した。PC1, PC2 における各文脈要素の貢献度と関係性、それらに基づく各演奏の比較分析の結果を平面に写像し、図4に示す。ここでは PC1 を横軸、PC2 を縦軸とした。また、いずれかの軸で 0.01 を超える因子負荷量を示した主要な文脈要素のベクトルを赤色の矢印で表す。なお、因子負荷量の値は各図の上と右の軸に示す。

(1) 文脈要素の因子負荷量による各演奏の傾向の比較分析

まず、提案手法を多様な演奏の組合せに適用した場合の一般的な傾向を分析する。ここでは図4の各平面において、各データセット間である程度の共通性が認められる点に着目する。

F_{note} , F_{beat} の双方において、横軸 (PC1) 方向には文脈要素 **global** が比較的高い因子負荷量を有する場合が多い。ここから、表情の変動に楽曲内の時間的位置への比較的強い依存が示唆される。よって、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、演奏者が着目した楽譜の指示の傾向や、その時間的位置に応じた変動の範囲などを示すことができると考えられる。

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} では **pattern**, **density** が寄与する傾向が共通している。それらの関係性はデータセットに依存するが、比較的強い寄与が認められる文脈要素はおおむね共通している。 F_{note} において比較的強い寄与が認められる文脈要素としては、 $M(1)$ では **type**, $B(1)$ では **staff**, **voice**, $C(1)$ では **voice**, **type** が挙げられる。個々の関係性はデータセットに依存して異なるが、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しうる点で共通している。ゆえに、これらは **global** と比較して高い局所性を有する文脈要素といえる。

図4 演奏の分布と各文脈要素の表情への寄与傾向 (F_{beat} , F_{note})

以上から、図4の各平面を構成する二軸における一般的な傾向として、それぞれ主に楽譜の大域的指示と局所的指示に関する文脈要素の寄与が示された。なお、第三主成分以降における各文脈要素の因子負荷量の傾向については、第二主成分までで示されたよりも大きな値は得られていないほか、各データセット間で共通性も認められない。したがって、PCAの適用結果を平面に写像した場合に演奏間の関係を有効に分析できるデータセットは、4.1節で示した累積寄与率を尺度とすることである程度適切に選別できていると考えられる。

(2) 同一楽曲における演奏者の傾向の比較分析

次に、特定のデータセットに依存した特殊な傾向に着目し、それぞれについてより詳細な分析を行う。

M(1)に注目すると、図4(a)に示す拍の場合では、演奏者ごとの演奏(演奏者GGは●▲■、演奏者MPは○△□で示す)が他のデータセットに比べて明確に分離している。これらの演奏に寄与した文脈要素の傾向は、主に演奏者に依存していると考えられる。この傾向は、図4(d)に示す音符の場合でも同様に認められる。GGによる各演奏ではglobalやtype、MPによる各演奏ではstaffやsyllableの寄与が示されている。各ベクトルの向きから、これらの文脈要素は各主成分に対する寄与の傾向がそれぞれ異なることがわかる。そのような相違の上で、これらの演奏はいずれも名演の一つとして成立している。以上から、M(1)に含まれる各楽曲には、演奏者の比較的自由的な解釈を許容する余地が考えられる。

一方、C(1)に注目すると、比較的多数の演奏が楽曲ごとに近い位置に分布しており、図4(c)に示す拍、図4(f)に示す音符の各平面上では演奏者に依存した相違が比較的少ない。後者の場合では楽曲P007のPC1での距離が比較的大きいが、PC1に寄与する文脈要素が主にglobalのみであることを考慮すると、他の多くの要素については類似が考えられる。よってC(1)の各演奏における表情には、演奏者の意図よりも楽譜の指示による制約が強く寄与していると考えられる。

残るB(1)に注目すると、図4(b)に示す F_{beat} については、図4(a)に示すM(1)との間で各文脈要素の因子負荷量に比較的近い値と互いの関係性が認められる。一方、図4(e)に示す F_{note} では比較的多数の文脈要素が寄与しており、その傾向はM(1)よりも、図4(f)に示すC(1)に近い。楽曲が作曲された年代に従うと、各データセットはM(1)、B(1)、C(1)の順となる。それぞれが含む楽曲において時代的な隔たりがあり、B(1)はM(1)とC(1)の中間に位置する。ここで得られた結果から、B(1)は表情においてもM(1)とC(1)の双方の特徴を備えており、両者の中間に位置することが示唆される。

4.3 演奏間における表情の類似性の比較分析

4.2節で示したように、提案手法に多変量解析手法を適用することで得られた分析結果において、各演奏間の距離はそれらの類似性を表すものと考えられる。そこで演奏間の類似度を主観評価実験によって測定し、その結果との比較から、

提案手法によって示された複数の演奏の関係性が、人間の主観的な類似性の判断に準じるものであるかを検証する。

M(1)、B(1)、C(1)のそれぞれに含まれる演奏間の表情について、類似度を問う主観評価を実施した。被験者には、任意の演奏対における表情の傾向について、5段階の評価軸に基づく類似・非類似の判定を依頼した。その被験者として、鍵盤楽器に関する高度な演奏技術やそれを専門とした活動経験を持たない、20名の協力を得た。以上により得られた演奏間の評価値を、類似・非類似の最大値がそれぞれ0、1となるように集計し、データセットごとにMDSを適用した。各演奏の分析結果を平面に配置したものを図5に示す。また、提案手法でも同様にM(1)、B(1)、C(1)にMDSを適用した。予備検討をふまえ、ここでは F_{note} についての各演奏の分析結果を平面に配置したものを図6に示す。各図の対照から、主観評価と提案手法の分析結果とを比較する。

図5(a)に示すM(1)の主観評価については、楽曲S279-2における類似性と、各演奏者における楽曲間の関係性が、図6(a)に示す提案手法の分析結果によって再現されている。また、図5(b)および図6(b)に示すB(1)では、各楽曲が横軸の値の正負によって分類される傾向が双方で共通して認められる。そして、演奏者CAとWBの類似性、そしてこの2名とWAの非類似性が再現されている。以上から、これらのデータセットにおいて、提案手法の分析結果は人間の主観的な判断に準じた演奏間の関係性の図示ができていると考えられる。

一方、図5(c)および図6(c)に示すC(1)では、他のデータセットにあるような類似点を見出すのが比較的困難となっている。作曲者に依存した傾向として、C(1)の楽曲にはより複雑な演奏指示が多数含まれる点が、M(1)、B(1)とは異なる。今回被験者に要求した判断には、ある程度以上の楽器演奏に関する経験に基づく知見を要するといえる。主観評価に際し、C(1)の楽曲における表情の類似性はM(1)、B(1)に比べて判定が困難であった可能性が考えられる。その要因の解明には、分析対象を単一の演奏内での特定部分に絞るなど、より詳細な分析に基づく考察を要する。

4.4 楽曲内における声部間の表情の比較分析

単一の演奏における表情の傾向についてより詳細な分析を行うため、声部に依存した音符の表情を比較する。ここでは楽曲の楽譜が示す楽節単位に従って演奏を声部別に4分割し、trebleとbassの各声部に属する音符の事例群を学習データに用いた木構造をそれぞれ構築した。演奏としてM(1)からS279-1GGとS279-1MP、B(1)からS008-2VAとS008-3VA、C(1)からN002VAとN002MPを用い、それぞれにPCAを適用した。その結果として図7に示す各平面を、所属するデータセット別に比較する。

図7(a)、(b)、(c)、(d)に示すM(1)およびB(1)の演奏に共通する傾向として、bass(△で示す)の演奏の多くがglobalのベクトルに従って分布している。4.2節の分析結果では多数の演奏においてglobalの比較的強い寄与が認められ

統計モデルと楽譜情報に基づく指標を用いた演奏表情の比較分析

図5 主観評価とMDSに基づく演奏間の表情の類似度

図6 提案手法とMDSに基づく演奏間の表情の類似度

図7 演奏内で声部別に分割した表情の比較分析

たが、この傾向に該当する演奏は主にbassのものであったと考えられる。これに対し、treble(●で示す)の各演奏には、global以外の比較的多数の文脈要素に分散して分布する傾向が共通している。その組合せは演奏に依存して異なるが、M(1)に比べて、B(1)やC(1)に属する演奏においてより複雑な傾向が認められる。この点も4.2節の考察を裏付けるものといえる。なお、trebleの演奏における寄与の度合いが比較的強い文脈要素としては、図7のすべての場合に共通してtype、特に図7(a), (c)ではlocal、図7(b), (e)ではsyllable、図7(d)ではchordなどが認められる。これらはいずれもglobalに比べて局所性の高い文脈要素である。

一方、図7(e), (f)に示すC(1)の演奏では比較的多数の文脈要素がbassの演奏に寄与している。この傾向は他のデータセットでは認められない。よって、C(1)の演奏における表情の類似性の判定には、他のデータセットの演奏よりも多様な要素についての注意を同時に要するといえる。また、このような傾向は、4.3節の評価で得られた分析結果や考察の裏付けとなると考えられる。

5. お わ り に

本稿では複数の演奏間における解釈の構成要素と相違の傾向を特定の言語に依存せず分析することを問題として、楽譜から一意に得られる情報の組合せに基づく指標を用い、演奏間でそれらの比較分析を行う手法を提案した。提案手法では文脈依存モデルと各文脈要素の貢献度を定義することで、統計モデリングと多変量解析による解決を試みた。

第4章の各評価で用いた演奏の組合せは限定的であるが、それらの結果から、特定の演奏者や楽曲の表情に寄与する要素が楽譜の指示の組合せに基づいて明示された。提案手法は特定の演奏者や楽曲が持つ特性を、明確かつ客観的に説明することに貢献しうる。その際取り組むべき分析の対象として、特定の演奏者や楽曲についてのより多様な演奏の組合せが挙げられる。

なお、第3章で提案した分析手法では、木構造中で選択された質問を文脈要素ごとに扱っている。本提案は表情の類似性の要因を分析する上で、個々の演奏における演奏者と楽曲の関係性を明確に分離することを視野に入れたものではない。そのため本稿では演奏間で共通かつ一意な記述が可能な分類基準として楽譜の指示を用い、指示の種類ごとにまとめて評価軸とした。しかしその組合せは分析を行う演奏によって最適化の余地があるほか、楽譜の指示以外にも検討すべき比較軸が考えられる。それらの検討は今後の課題であり、より多様な演奏を用いた分析を要する。

謝 辞

本研究の一部は、科学研究費補助金(課題番号:26730182)および科学技術振興機構(JST)研究成果最適展開支援プログラムA-STEP(課題番号:AS262Z00622H)の支援を受けたものです。

注

- [注1] MakeMusic : MusicXML. <http://www.musicxml.com/> (2015/4/30 閲覧)
- [注2] CrestMuse Project : CrestMuse PEDB. <http://www.crestmuse.jp/pedb/> (2015/4/30 閲覧)

参 考 文 献

- [1] Palmer, C.: Music performance, Annual review of psychology, Vol.48, No.1, pp.115-138, 1997.
- [2] Repp, B. H.: Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's "Träumerei", Journal of the Acoustical Society of America, Vol.92, No.5, pp.2546-2568, 1992.
- [3] Gabrielsson, A. and Juslin, P. N.: Emotional expression in music performance: Between the performer's intention and the listener's experience, Psychology of music, Vol.24, No.1, pp.68-91, 1996.
- [4] Clarke, E. F.: Generative principles in music performance, Generative processes in music. The psychology of performance, improvisation, and composition (Sloboda, J.A., ed.), Clarendon/Oxford Press, chapter 1, pp.1-26, 1988.
- [5] Gabrielsson, A.: Music performance research at the millennium, Psychology of music, Vol.31, No.3, pp.221-272, 2003.
- [6] Okumura, K., Sako, S. and Kitamura, T.: Stochastic Modeling of a Musical Performance with Expressive Representations from the Musical Score, in Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR), pp.531-536, 2011.
- [7] Kirke, A. and Miranda, E. R.: Guide to Computing for Expressive Music Performance, Springer Science & Business Media, 2012.
- [8] Okumura, K., Sako, S. and Kitamura, T.: Laminæ: a stochastic modeling-based autonomous performance rendering system that elucidates performer characteristics, in Proceedings of ICMC-SMC-2014, pp.1271-1276, 2014.
- [9] 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.
- [10] Dixon, S., Goebl, W. and Widmer, G.: The performance worm: Real time visualisation of expression based on Langrer's tempo-loudness animation, in Proceedings of International Computer Music Conference (ICMC), pp.361-364, 2002.
- [11] Grachten, M., Goebl, W., Flossmann, S. and Widmer, G.: Phase-plane representation and visualization of gestural structure in expressive timing, Journal of New Music Research, Vol.38, No.2, pp.183-195, 2009.

- [12] Grachten, M., and Widmer, G.: Linear basis models for prediction and analysis of musical expression, *Journal of New Music Research*, Vol.41, No.4, pp.311-322, 2012.
- [13] Sapp, C. S.: Harmonic Visualizations of Tonal Music, in *Proceedings of International Computer Music Conference (ICMC)*, pp.423-430, 2001.
- [14] Sapp, C. S.: Comparative analysis of multiple musical performances, in *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR)*, pp.497-500, 2007.
- [15] Widmer, G.: Machine Discoveries: A Few Simple, Robust Local Expression Principles, *Journal of New Music Research*, Vol.31, pp.37-50, 2002.
- [16] Tobudic, A. and Widmer, G.: Relational IBL in Music with a New Structural Similarity Measure, in *Proceedings of the 13th International Conference on Inductive Logic Programming*, pp.365-382, 2003.
- [17] 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.
- [18] Seashore, C. E.: *Psychology of Music*, Courier Dover Publications, 1938.
- [19] 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.
- [20] Flossmann, S., Grachten, M. and Widmer, G.: Expressive performance rendering: Introducing performance context, in *Proceedings of the 6th Sound and Music Computing Conference (SMC)*, pp.155-160, 2009.
- [21] Odell, J. J.: *The Use of Context in Large Vocabulary Speech Recognition*, PhD Thesis, Cambridge University, 1995.

奥村 健太 (非会員)

2007年 名古屋工業大学工学部卒業. 2009年 名古屋工業大学大学院博士前期課程修了. 現在 名古屋工業大学大学院工学研究科在学. 修士(工学). 音楽情報処理の研究に従事.

2012年度情報処理学会東海支部学生論文奨励賞受賞. 情報処理学会, 日本音響学会, 人工知能学会各会員.

酒向 慎司 (正会員)

1999年 名古屋工業大学工学部卒業. 2004年 名古屋工業大学大学院博士後期課程修了. 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. 2012年 ミュンヘン工科大学客

員研究員. 2014年 AGH科学技術大学客員研究員. 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, 情報処理学会, 日本音響学会, 人工知能学会, 日本感性工学会, IEEE各会員.

北村 正 (非会員)

1973年 名古屋工業大学工学部卒業. 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年 ウェールズ大学(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). 感性情報処理, 音楽情報処理, デジタル信号処理, 手話認識等の研究に従事. 電子情報通信学会, 情報処理学会, 日本音響学会, 映像情報メディア学会, IEEE, ISCA各会員.