

# 事例に基づく演奏表情の生成に関する研究

鈴木泰山, 木暮雅樹, 徳永健伸, 田中穂積

東京工業大学 情報理工学研究科 計算工学専攻

{taizan, kigure, take, tanaka}@cs.titech.ac.jp

我々は事例に基づく手法を用いた演奏表情の自動生成の研究を行なっている。この手法では、対象曲と与えたい演奏表情の特徴を入力すると、あらかじめ用意した演奏データ集の中から対象曲に類似した曲の演奏事例を検索し、それらに見られる演奏表情を対象曲に適用することで対象曲の演奏データを生成する。一般に事例に基づく手法では事例の不足が大きな問題となる。本論文では、事例に基づく演奏表情の生成手法において事例の不足が生じにくくするための技術として、楽曲を断片に分割し、断片ごとに類似した事例を手法について述べる。またその際に事例に見られる演奏表情を適切に分析し対象曲に適用する手法についても述べる。

## A Case-Based Approach to Generating Musical Expression

Taizan Suzuki, Kigure Masaki, Takenobu Tokunaga, Hozumi Tanaka

Department of Computer Science

Graduate School of Information Science and Engineering

Tokyo Institute of Technology

We have been developing a case-based method to generate musical expression. This method generates musical expression for the target piece through retrieval and analysis of pre-compiled performance data for similar musical pieces. Generally speaking, in the case-based method, data sparseness leads to lackluster results. We introduce a technique to reduce the effects of data sparseness within our case-based approach. This technique divides pieces into segments of various granularity, and retrieves examples for each segment. Special description of musical expression is required, and we introduce a descriptive schema and present analysis and calculation methods based thereupon.

## 1 はじめに

人間の音楽の演奏はテンポや音の強さが曲中で様々に変化している。楽曲の演奏中にみられるこれらのテンポや音の強さの芸術的逸脱のことを演奏表情と呼ぶ。演奏表情は表情豊かで聴者を魅了するような音楽演奏に必要な不可欠な要素である。しかしながらその与え方や変化量などは楽譜には殆んど記載されておらず事実上全てが演奏者の裁量に委ねられているといつてよい。演奏表情の与え方は非常に自由であり、演奏者はその知識や経験、趣向に基づいて各々が独自の演奏表情を与えている。従って演奏

表情は演奏者や演奏者を取り巻く環境によって様々に変化する。このような楽曲以外の演奏表情に影響を及ぼす要素を演奏状況と呼ぶ。

我々は与えられた楽曲に対して演奏表情を生成する手法として事例に基づく演奏表情の生成手法の提案を行なった [1, 2, 3]。この手法は、生成の過程で演奏状況を容易に考慮することが可能であり、これに従って単一の楽曲に対して演奏状況に応じた特徴のことなる様々な演奏表情を生成することが可能である。

本論文では、我々の演奏表情生成手法について概説した上で、本手法で用いている演奏表情の分析お

よび合成の手法について述べる。

## 2 事例に基づく演奏表情生成

本章では事例に基づく演奏表情の生成手法の概要および機構について説明する。

事例に基づく演奏表情の生成手法の概要を図1に示す。本手法では演奏表情の知識として演奏データ集を利用する。演奏データ集は人間が演奏したデータを集めたもので、この中の個々のデータは1) 演奏曲の楽譜情報、2) 人間の演奏データ、3) 演奏時の演奏状況からなる。演奏データ集はあらかじめ用意しておく必要があるが、必要に応じて随時新しいデータを追加することも可能である。

演奏表情を生成する際には、1) 対象曲情報と、2) 演奏状況設定、を入力として与える。対象曲情報は主に対象曲の楽譜情報からなる。演奏状況設定は生成される演奏表情の特徴や雰囲気を指定するパラメータである。

入力を与えられると、演奏データ集から対象曲に類似した楽曲を検索する。類似した楽曲の検索は、対象曲全体だけではなく対象曲を構成するフレーズや小節などの旋律断片に対しても行なう。これによって得られる対象曲に類似した旋律断片の集合が参考事例集である。

次に、参考事例集の中の事例ごとの重要度を評価する。重要度は、対象曲の演奏表情を生成する際にその事例がどの程度参考になるかを表すスコアである。事例の重要度は「入力された演奏状況と事例の演奏状況の類似性」と「対象曲と事例との旋律断片の類似性」の2つの要素から決定する。いずれも対象曲と事例とが類似しているほど重要度は高くなる。旋律断片の事例の重要度は、対象となる断片同士の類似性だけではなく、それぞれの断片の前後に連続する断片同士の類似性なども考慮する。

また、重要度の評価と同時に個々の事例の演奏表情の分析を行なう。それぞれの事例中の局所的なテンポや音の強さの数値を計算し、大局的な数値と比較することで、演奏表情の相対的な変化量を求める。

分析によって得られた個々の演奏事例の演奏表情を、事例の重要度で加重平均して合成し、対象曲全体の演奏表情を生成する。最後に、この対象曲の演奏表情を対象曲の楽譜に適用し、演奏表情の与えられた対象曲の演奏データを出力する。

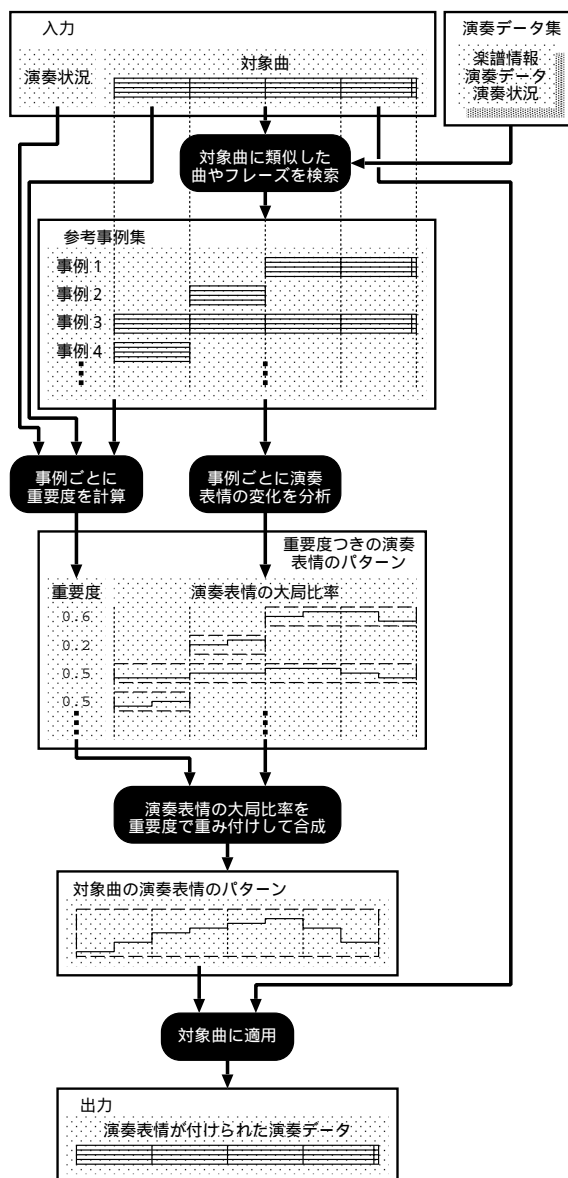


図1: 事例に基づく演奏表情の生成手法の概要

この手法では、重要度の評価の際に、対象曲と参考事例との楽曲の類似性だけではなく入力された演奏状況と参考事例の演奏状況の類似性も考慮しており、同じ曲でも演奏状況の与え方によって事例ごとの重要度が変化する。これによって、最終的に生成される演奏表情も変化するため、演奏状況の設定に応じた多様な演奏表情の生成が可能となる。

また、演奏データ集からルールなどの知識情報への変換を行なう必要がないため、新しい知識(演奏データ)の追加の際に生じるコストが非常に小さいという利点がある。

さらに、本手法では楽曲構造に関する認知的な知識が不要である。従来、演奏表情生成では認知的

楽曲構造が必要不可欠と言われている。多くのシステムでは演奏表情の生成に楽曲構造の分析に関する知識を導入しており、この知識によって演奏表情は大きく左右される傾向にある。本手法では、演奏事例を検索し分析する過程で、楽曲構造と演奏表情との関連づけの知識だけではなく、対象曲と楽曲構造との関連づけの知識も暗に獲得されているため、認知的楽曲構造に関する深い知識を与える必要がない。

一方で、ルールなどの一般化した知識を利用する手法に比べ、本手法は演奏表情の生成のメカニズムが不透明なため、本手法は人間の演奏表情の生成の分析などを目的とした分析的研究には適さないなどの問題点もある。

### 3 演奏表情の分析と合成

本節では事例に基づく演奏表情生成手法での演奏表情の取り扱いについて述べる。

#### 3.1 事例検索の効率化

本手法に限らず、事例に基づく手法が潜在的に抱える問題として事例の不足がある。事例に基づく手法では、データ集の中から対象に適用可能な事例を検索し、それらを分析して対象に適用する。従って検索の結果の事例が少ないと有効な結果が得られない。

事例不足の問題の可能性を減らす最も有効な手法は、対象曲を分割して検索することである。対象曲を分割して類似旋律を検索する手法としては、一音ごとに周囲の音符との認知的構造の類似性などを考慮して事例を検索する手法がある [4]。しかし、演奏表情はその音の前後に連続する音符やフレーズによって影響を受けるため、一音単位の検索では対象となる音と周囲との情報があまり反映できず、結果として適切な事例が得られない可能性がある。従って、一音ごとだけではなく、ある程度の長さの連続した旋律についての事例の検索が必要である。

本手法では、対象曲や演奏データ集のデータを様々な長さの旋律断片に分割し、旋律断片を対象として参考事例の検索を行なう。これにより事例の総数が増化する (図 2) 上に、個々の検索単位の長さが短くなるため類似旋律の検索に成功する可能性が

向上 (図 3) する。

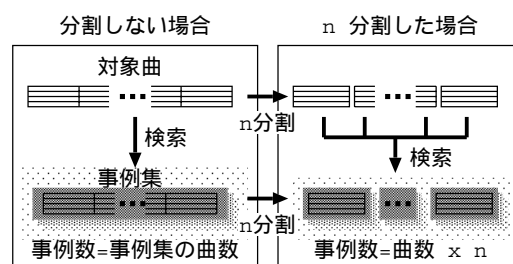


図 2: 旋律の分割による効用 1 (事例数の増化)

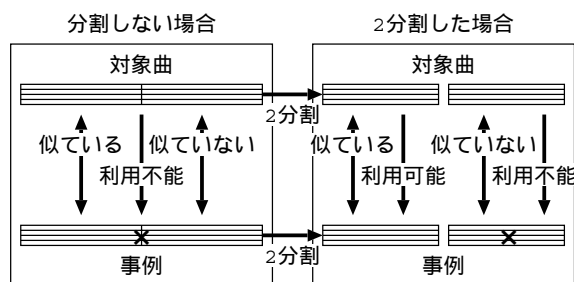


図 3: 旋律の分割による効用 2 (旋律が類似する確率の向上)

曲全体について網羅的に参考事例を検索するためには、様々な分割方法を用いて分割を行ないそれぞれについて検索する方がよい。理想としては、楽曲の可能な分割を全て試すのがよいが、そうすると検索する断片の数が、対象曲中の音符の数の冪乗になってしまうので、計算量の観点から現実的ではない。そこで現状では、小節、フレーズ、繰り返し、などの楽曲構造で楽曲を分割し、それぞれの断片について類似した旋律の検索を行なっている。楽曲構造は、楽曲の分割方法として楽譜などに記載されていて比較的容易に利用できる点や、分割の手法にある程度の一貫性が期待できるためランダムな分割などに比べ類似旋律の検索の際により類似した事例が見つかる可能性が高い点などで優れている。

後述の演奏表情の記述法を利用するためには、粒度の異なる分割の間に階層性があり、且つ粒度の小さい分割の境界が粒度の大きい分割の境界の部分集合である必要がある。楽曲構造による分割はこの点からも好都合である。

#### 3.2 演奏表情の分析

旋律断片に分割して事例を検索すると、粒度の小さい断片については事例の総数の増化などにより参考事例が見つかる可能性の向上が期待できるが、粒

度が大きくなるほど分割による効用が小さくなってしまふ。粒度の小さい参考事例のみを組み合わせることによっても演奏表情の生成は可能であるが、より大きい範囲で見た際の演奏表情のつながりや全体的なバランスなどが適切には得られないため、粒度の大きな断片に対する事例もある程度必要である。

粒度の大きな事例の場合、分割などによって手軽に事例の数自体を増やすことが困難なので、事例の類似性の評価基準を緩くすることで事例が見つかる可能性を向上させる。本手法では、事例を検索する際に、粒度の大小に関わらず断片内の全体的な楽曲の特徴の類似性のみ注目して比較している。このようにすることで粒度の大小に関わらず旋律同士が類似している確率を一様にするのが可能である。単純に旋律などの比較を行なった場合、旋律が長くなるほど旋律同士が類似している確率が減少することから、全体的な特徴のみによる比較は、とくに粒度の大きい旋律断片の類似事例を検索する際に有効である。

しかし、全体的な特徴が類似しているだけの事例では、局所的な部分の類似性が保証されていないので、局所的な部分の演奏表情を参考にするのは適切ではない。従って、全体的な演奏表情の変化と局所的な変化を分離し、全体的な演奏表情のみを取り出す必要がある。



図 4: 全体的な演奏表情と局所的な演奏表情の分離

この全体的な演奏表情のみの獲得をするために、演奏表情の大局比率という概念を導入している。これについては次章で述べる。

## 4 演奏表情の大局比率

### 4.1 演奏表情の大局比率の分析

演奏事例集から演奏表情を獲得し対象曲に適用するためにはテンポや音の強さなどの演奏表情の絶対

的な数値 (演奏表情値) を何らかの形式でデータとして記述する必要がある。演奏表情の記述には、一音ごとの演奏表情値をそのまま用いる手法や、演奏表情値の変化量を曲線近似して記述する方法 [5] などが用いられることが多い。しかし、一音ごとの演奏表情値を記録する手法では前章で述べたような全体的な演奏表情と局所的な演奏表情の分解が行えない。また、曲線で近似する手法は、全体的な演奏表情と局所的な演奏表情との分解は比較的容易に行なえるが、近似によって実際の数値との微妙なずれが生じる点や演奏表情値を連続的な数値として扱うこと自体の問題 [6] などがあまり適切ではない。

そこで、本手法では演奏表情の変化を表すために演奏表情の大局比率を用いた再帰的表記法を提案する。この表記法を用いるためには、対象曲について以下の条件を満たすような粒度の異なる楽曲の境界を定義する必要がある (図 5)。

1. 粒度の大きさごとに階層的に何段階かの分割が定義できる
2. 粒度の小さい分割の境界の集合は粒度の大きい分割の境界の部分集合である

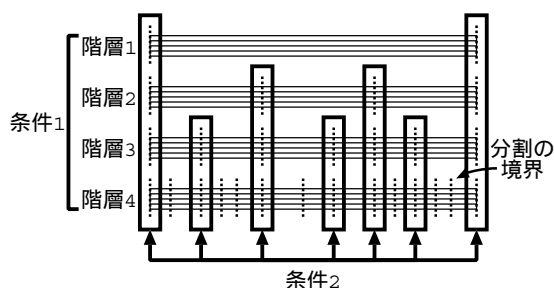


図 5: 分割の境界の条件

演奏表情の大局比率は、注目している旋律断片での演奏表情値の平均値を、それよりひとまわり大きい旋律断片 (一階層上の分割で注目している旋律断片を内包する旋律断片) での演奏表情値の平均値との比率で表現したものである。旋律断片  $s_{i,j}$  の演奏表情の大局比率  $P_{exp}$  を (1) に示す。

$$P_{exp}(s_{i,j}) = \frac{\exp(s_{i,j})}{\exp(s_{i-1,k})} \quad (1)$$

$$\{s_{i,j} \subset s_{i-1,k}\}$$

$s_{i,j}$  は上から  $i$  番目の階層の分割の先頭から  $j$  番目の旋律断片である。 $s_{i-1,k}$  は  $s_{i,j}$  のよりひとつ上の階層の断片で、 $s_{i,j}$  は  $s_{i-1,k}$  の一部である。 $\exp(s)$  は断片  $s$  での演奏表情値の平均値である。

例として図6に4小節からなるフレーズの小節ごとの演奏表情の大局比率の計算例を示す。このフレーズは全体の平均テンポが120(0.5秒/四分音符)で、それぞれの小節の平均のテンポが順番に115(0.52秒/四分音符), 133(0.45秒/四分音符), 150(0.4秒/四分音符), 95(0.63秒/四分音符)で演奏されたとする。また、各小節と4小節からなるフレーズ全体が旋律断片であったとする。

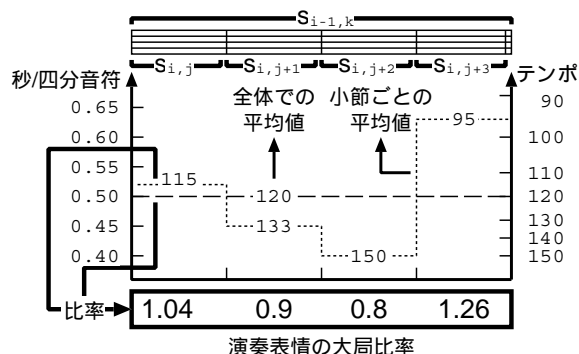


図6: テンポの演奏表情の大局比率の計算例

テンポは演奏時間の逆数であるため、平均などの計算を行なうと実際の値と一致しなくなってしまう(上記の例でも115, 133, 150, 95の平均は120にはならない)。よって、テンポに限っては演奏表情の大局比率の計算にテンポの逆数である「四分音符の演奏長(秒/四分音符)」を利用する。

演奏表情の大局比率は注目する旋律断片の演奏表情値とその親の演奏表情値との比率である。この例では各小節の演奏表情の大局比率は個々の小節での四分音符の演奏長と全体の四分音符の演奏長との比率になる。よって、この例の各小節のテンポの演奏表情の大局比率は順に1.04, 0.9, 0.8, 1.26, となる。

演奏表情の大局比率から個々の旋律断片  $s_{n,j}$  の演奏表情の平均値  $exp(s_{n,j})$  を計算する際には、(2)に示すように曲全体での平均値  $exp(s_{1,1})$  に音符自身の大局比率と音符の全ての祖先の大局比率  $P_{exp}$  を掛け合わせればよい。

$$exp(s_{n,j}) = exp(s_{1,1}) \cdot \prod_{1 < i \leq n} P_{exp}(s_{i,j_i}) \quad (2)$$

$$\{s_{n,j} \subset s_{i,j_i} \text{ for each } i\}$$

図7にこの計算の一例を示す。半小節の演奏表情の大局比率は小節の演奏表情値と半小節の演奏表

情値の比率である。また、小節の演奏表情の大局比率はフレーズの演奏表情値と小節の演奏表情値の比率である。従って、半小節の演奏表情の大局比率と小節の演奏表情の大局比率を掛け合わせることで、フレーズの演奏表情値と半小節の演奏表情値の比率が求まる。これにフレーズ全体の演奏表情値を掛けると、半小節の演奏表情値が得られる。この計算は可逆計算であり、もともとの演奏表情値の値は完全に復元することが可能である。

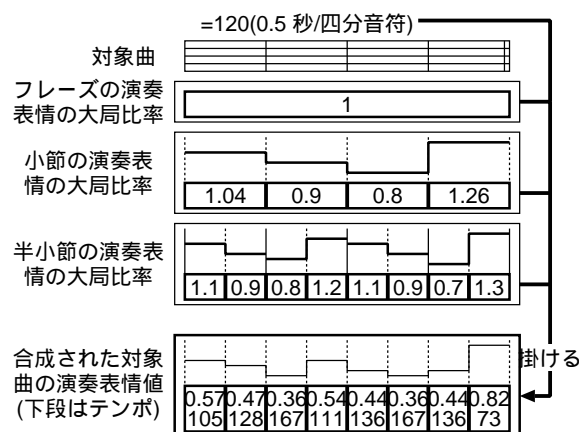


図7: 演奏表情の大局比率の合成の例

## 4.2 演奏表情の大局比率の合成

演奏表情生成の最終段階では、個々の参考事例の演奏表情の大局比率を合成し、対象曲の演奏表情を作成する。本節では演奏表情の大局比率の合成について述べる。

対象曲の演奏表情の大局比率の合成は次の手順で行なう。

1. 対象曲の旋律断片ごとに参考事例の演奏表情の大局比率を合成
2. 各旋律断片の演奏表情の大局比率を組み合わせる

演奏表情を合成する段階では、対象曲に対する全ての参考事例について、演奏表情の大局比率と事例の重要度が得られている。

最初の段階では対象曲の旋律断片ごとに、その断片に対する参考事例の演奏表情の大局比率を重要度で加重平均し、対象曲の旋律断片に対する演奏表情の大局比率を計算する(図8)。対象曲の旋律断片  $s_{i,j}$  の演奏表情の大局比率  $P_{exp}(s_{i,j})$  の計算を

(3) に示す． $E_{i,j}$  は  $s_{i,j}$  に対する参考事例の集合， $W(s)$  は参考事例  $s$  の重要度を表す．

$$P_{exp}(s_{i,j}) = \frac{\sum_{s \in E_{i,j}} W(s) \cdot P_{exp}(s)}{\sum_{s \in E_{i,j}} W(s)} \quad (3)$$

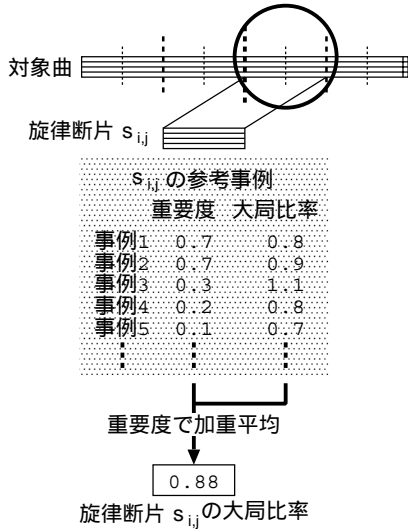


図 8: 演奏表情の合成，第一段階

曲全体の演奏表情の合成は，個々の旋律断片の演奏表情の大局比率を掛け合わせるによって算出する．図 7 に示すように，楽曲の全ての旋律断片の演奏表情の大局比率が求まっている場合，個々の音符の演奏表情値は演奏表情の大局比率から (2) によって計算することができる．よって，対象曲の一拍ごとの演奏表情値は，その音自身とその全ての先祖の旋律断片の大局比率の積に曲全体の演奏表情を掛けることで得られる．対象曲の  $m$  番目の音符  $s_{n,m}$  の演奏表情値  $exp(n_m)$  の計算式を 4 に示す． $n$  は旋律断片の階層の総数である．

$$exp(s_{n,m}) = exp(s_{1,1}) \cdot \prod_{1 < i \leq n} P_{exp}(s_{i,j}) \quad (4)$$

$$\{s_{n,m} \in s_{i,j_i} \text{ for each } i\}$$

旋律断片  $s_{i,j}$  の演奏表情の大局比率  $P_{exp}(s_{i,j})$  は (3) で求めたものである．曲全体の演奏表情  $exp(s_{1,1})$  は曲全体に対する参考事例から獲得する．

## 5 おわりに

本論文では事例に基づく演奏表情生成手法の要素技術として，楽曲の分割による事例の有効利用と，演奏表情の表記，分析，計算手法について述べた．

対象曲と演奏データを様々な粒度の旋律断片に分割し，断片ごとに類似した事例を検索することでデータ不足の問題が発生しにくくすることが可能である．また，参考事例の検索を行なう際に，全体的な特徴の類似のみに注目することで，粒度の大きな断片に対しても事例が見つかる可能性を高め，結果として全体的な演奏表情のバランスなどの情報の取得が可能となる．

ここで述べた事例の有効利用のための手法を実装するためには，演奏表情の全体的な流れと局所的な変化を分離できるような演奏表情の分析手法が必要となる．ここでは部分的な演奏表情を比率を用いて階層的に記述する手法を示すとともに，事例に基づく演奏表情生成へのこの記述法の導入について述べた．

本論文で述べた演奏表情の記述法は時間方向の変化のみに注目しており，和音の一拍ごとの強さの違いを扱うことはできない．今後，この表現方法を拡張し時間方向だけでなく周波数方向（和音など）の変化も適切に表現できるようにする必要がある．

## 参考文献

- [1] 鈴木泰山, 徳永健伸, and 田中穂積. 事例ベースの演奏表情生成手法に関する研究. 情報処理学会研究報告 97-MUS-21, 97(67):7-12, 1997.
- [2] 鈴木泰山, 宮本朋範, 西田深志, 徳永健伸, and 田中穂積. Kagurame phase-i —事例ベースの演奏表情生成システム—. 情報処理学会音楽情報科学研究会研究報告, 98(14):61-68, 1998.
- [3] T. Suzuki, T. Tokunaga, and H. Tanaka. A case based approach to the generation of musical expression. In *Proceedings of the sixteenth International Joint Conference on Artificial Intelligence*, pages 642-648. IJCAI, 1999.
- [4] J. L. Arcos, R. L. de Mántaras, and X. Serra. SaxEx : a case-based reasoning system for generating expressive musical performances. In *Proceedings of the 1997 International Computer Music Conference*, pages 329-336. ICMA, 1997.
- [5] 田口友康. 自動演奏と演奏モデル:演奏の芸術性を目指して. *bit 別冊, コンピュータと音楽*, pages 42-53, 1987.
- [6] P. Desain and H. Honing. Tempo curves considered harmful. a critical review of the representation of timing in computer music. In *Proceedings of the 1991 International Computer Music Conference*, pages 143-149. ICMA, 1991.