

自動型と対話型の遺伝的プログラミングを用いた 自動表情付けシステム

丹 治 信^{†1} 伊 庭 斉 志^{†1}

本稿では遺伝的プログラミングを用いた音楽の自動表情付けシステム ConBreO(Conductor Breeding Optimization) について述べる。ConBreO では個々の音符に対して、遺伝的プログラミングで学習された4種類 (Velocity, Tempo, Attack, Release) のルールを適用することで演奏表情データを得る。学習段階では人間の弾いた演奏データとの誤差を評価値とする自動型進化と、ユーザが演奏を聴いて評価する対話型進化 (IEC) を組み合わせて用いる。IEC ではユーザの疲労が問題になるが、本研究ではユーザサポートのためにユーザの評価値を予測するガウシアンプロセスモデルを用いて、無駄だと思われる個体を提示しない“選択提示”を導入する。また、多人数での評価を可能にするため、WEB アプリケーションによるインタフェースを公開した。

A Hybrid System of Interactive and Automated Genetic Programming for Expressive Music Rendering

MAKOTO TANJI ^{†1} and HITOSHI IBA^{†1}

This paper presents a system for creating expressive music performance by means of GP (Genetic Programming). The system applies expression rules (velocity, tempo, attack and release) into each music note. In the learning stage, the system concatenates an automated evolution system which uses expressed data performed by human player and an interactive evolution system which uses users evaluation. To reduce users fatigue, the system introduces “selective presentation” which learns users evaluation using Gaussian Process. Using the interface of web application, more than one people can evaluate the system.

^{†1} 東京大学大学院工学系研究科
Engineering, University of Tokyo
{tanji, iba}@iba.t.u-tokyo.ac.jp

1. はじめに

本研究では、遺伝的プログラミングを用いた音楽の自動表情付けについて述べる。楽曲の楽譜が与えられたときに機械的に演奏することは、計算機上で実現されている。しかし、人間の演奏者は楽譜を機械的に演奏せず、個々の音に強弱を付けたり、テンポを揺らして演奏する。これによって聴衆はその楽曲のもつキャラクタをより深く感じることができる。演奏表情を付与する過程は、個人の楽曲に対する解釈や感情に依存すると考えられ、十分に理解されてはいない。

自動表情付けとは、上で述べたような演奏表情を人間の手を借りずに生成することである。工学的には、まだ演奏されていない未知の楽曲や新しく作曲した楽曲に対して、ある程度人間らしい演奏を生成することは有用であると言える。また、人間がどのように表情豊かな音楽を演奏するかと言う問題に対し、作って理解するという構成論的アプローチとも捉えられる。

自動表情付けの研究は、人間が与えた規則に基づくシステム³⁾、演奏データベース中からの類似断片の演奏表情を付加するもの⁶⁾、確率モデルの学習を行うもの^{7),10)} がある。

我々は遺伝的プログラミング⁴⁾を用いた音楽の自動表情付けシステムについて研究している。遺伝的プログラミングは進化計算の一種で、集団による確率的な最適化手法である。その中でも上で紹介したように、人間の演奏データを使って最適化を行う自動型進化とユーザと対話的に最適化を行う“対話型進化計算 (Interactive Evolutionary Computation, 以降 IEC と記述)”⁹⁾ と呼ばれる手法を組み合わせて用いる。自動型進化とは、ここでは対話型進化計算と対比させ人間の評価を介さない通常の進化計算を指す。一方、対話型進化計算 (Interactive Evolutionary Computation, IEC) は評価関数を定義することが難しい問題や、そもそも個人の好みに適合した解を得るために評価関数が人間の頭の中にしかない問題に対して、有効に働くことが期待されている。

IEC ではユーザに評価値を求める特性上、現実に行える評価の回数に制限がかかる。数多くの評価を必要とする進化計算においては、複雑な問題には何からの計算機によるサポートが必要であると考えられる。本稿ではデータベースからの自動型進化に IEC によって得られた好み関数を加えることでユーザの少ない評価回数で自動型進化を進めるシステムを提案する。また、選択提示と名づける手法を導入し、“高価”なユーザの評価を有効に活用することを目標とする。また、WEB アプリケーションとして実装することで、多くのユーザに使ってもらえるシステムとして拡張を行う。

二章で遺伝的プログラミングによる自動表情付けシステムの詳細を述べる．三章では，対話型進化計算においてユーザの評価は非常に高価であるという考えにより，自動型進化と対話型を有効に組み合わせるための手法について考察し，提案する混合手法を使ったシステム“ConBreO”について述べる．WEB アプリケーションとして拡張したインタフェースを四章で示し，最後にまとめとする．

2. ConBreO (Conductor Breeding Optimization)

ConBreO では表題の通り，自動型と対話型の遺伝的プログラミングを用いて演奏表情を生成する．遺伝的プログラミング (Genetic Programming, 以降 GP) は，進化計算の一種であり，LISP の S 式のような木構造データを進化させることで，関数やプログラムの最適化を行う手法である．

2.1 GP による演奏表情の生成

GP では問題を解くのに必要であろう変数や，それらの演算記号などを定義する．通常の GP の処理の流れは以下の通りである．

- Step1** [初期化] ランダムに作った個体（木構造を持つ）を P 個作る．これを集団と呼ぶ．
- Step2** [評価] 集団中の個体に評価値を与える．終了条件が満たされていれば終了する．
- Step3** [選択] 評価値に基づき，優良個体を選択する．
- Step4** [組み換え] 交叉・突然変異と言った遺伝オペレータを施し，優良解の組み換えを行う．
- Step5** [世代交代] 古い世代と組み換えにより生成された集団と交換する．Step2 に戻る．

楽譜は各音符のシーケンス $\{notet_i\}$ として表現される．それぞれの音符 $notet_i$ は，発音の時間，音価 (四分音符などの音の長さ)，音高 (MIDI ノート番号)，ダイナミクス (f, p などの強弱記号)，スラー，スタッカートなどの演奏記号などの情報を持つ．

本研究では，上で述べた要素を引数として，演奏表情の値を返す関数を GP で作成する．表 1 に ConBreO で使用する GP の非終端記号と終端記号を示す．個々の音符に関する表 1 の下に示すパラメータをまとめて T と表記する．演奏表情は，音の強さ *Velocity*，発音時間 *Attack*，消音時間 *Release*，テンポ *Tempo* の 4 つの値を対象とし，それぞれ別の GP を進化させて最後に 4 つの演奏表情ルールを未知の楽曲に適応することで最終的な演奏表情付きの演奏が得る．

演奏表情の値は機械的な演奏を 1.0 としたときの割合で表現する．例えば，*Velocity* が平均 60 の場合に，ある音が 90 で演奏された場合は *Velocity* の演奏表情の値は 1.5 となる．GP で目指す関数は以下の形になり，一つの音符の終端記号の情報から，1.0 を中心とする実数

表 1 GP で使う非終端記号と終端記号
非終端記号

+	2	加算
−	2	減算
*	2	乗算
/	2	除算
<i>iflt</i>	4	引数 1 より引数 2 が小さければ引数 3，それ以外は引数 4 を返す
終端記号 (一つの音符に関する情報)		
R		乱数
POS		曲中の位置
MPOS		一小節中の位置
CP		音高
PP		前の音符の音高
CD		音価
LP		和音中の位置
DY		ダイナミクス
PV		前の音符の演奏表情の値
CR		クレッシェンド
SL		スラー
AC		アクセント
ST		スタッカート

を出力する関数となる．

$$f : T \mapsto \mathbb{R}$$

図 1 に演奏表情付加のイメージを示す．

上で定義した演奏表情の値は，通常 1.0 付近の値を取り，2 以上の値や 0.0 近くの値を取ることは経験的に稀であることがわかっている．GP の出力をそのまま演奏表情とすると，そのような値の範囲に収まる保証が無い．そのため，全ての実数値を 0 から S までの領域にマッピングするような以下のシグモイド関数により，変換を施す．

$$sig(x) = \frac{S}{1 + \exp(-x)} + 1 - \frac{S}{2}$$

ここで， x は GP による出力であり，実際に使用する演奏表情の値は $sig(x)$ である．

2.2 自動型評価と対話型評価

IEC によって理想的には人間が弾いたデータベースを利用せずに対話型進化計算を用いることで演奏表情ルールが生成できると考えられるが，IEC ではユーザに評価値を求めるた

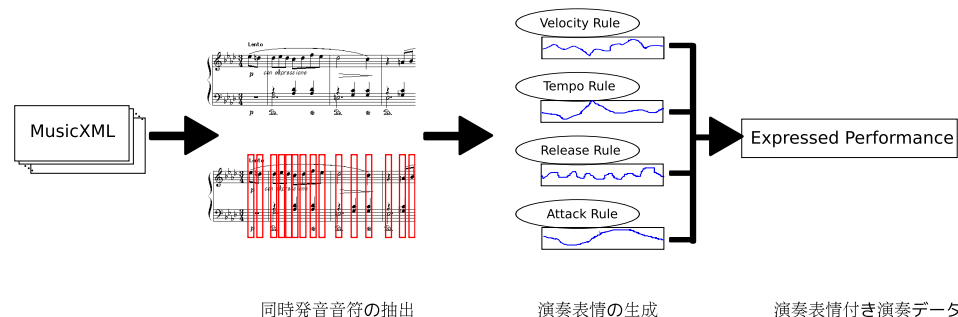


図1 演奏表情付加の模式図．楽譜ファイルから同時発音グループに分解した後，Velocity, Tempo, Attack, Release の4つのルールから生成された演奏表情を統合して表情付き演奏データを得る．

め，大量の評価回数をユーザがこなさなければならない．特に GP のような探索空間の広い問題設定においては，数千個体 × 数十世代程度の評価回数が望ましいが，ユーザにとって 100 ~ 1000 回を越える評価は疲労ややる気の問題から現実的ではない．

以上の理由から，本システムでは自動型と対話型の両方を用いて進化を進める．

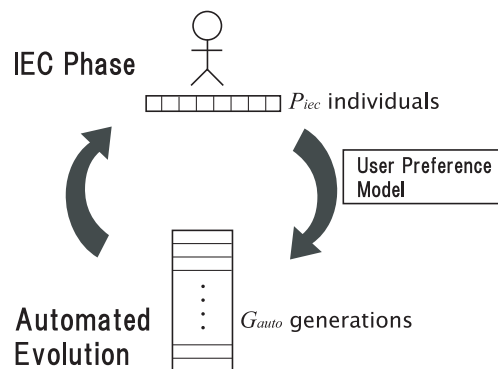


図2 自動型と対話型の GP システムの模式図．

対話型では，ユーザの評価値をもとにユーザの評価値予測モデルを構築する．これは本研究ではガウシアンプロセスを用いる．自動型では基本的にデータベースからの誤差をもとに評価値を与えるが，ユーザの好みを反映させるために確率 p_{DB} でデータベースとの誤差を

用い，それ以外のときにガウシアンプロセスにより予想された評価値を用いる．

$$f(t) = \begin{cases} f_{DB}(t) & \text{with probability } p_{DB} \\ f_{IEC}(t) & \text{with probability } 1 - p_{DB} \end{cases}$$

2.2.1 自動型

自動型進化では，人間の弾いた演奏データを使用して評価を行う．評価関数は以下のような単純に誤差を取ったものを使用する．

$$f_{DB}(t) = \frac{1}{N} \sum_{n=1}^N |t(n) - exp_n^{DB}|$$

ここで $t(n)$ を n 番目の音符に対する GP の出力， exp_n^{DB} を人間が弾いた演奏データ中の n 番目の演奏表情とする．本システムでは人間の演奏データとして CrestMusePEDB⁸⁾ を使用する．自動型進化の特徴として以下があげられる．

- 演奏データベースがあれば，自動で進化を進めることができる．
- データベースに無い楽曲に関しては学習できない．
- 細かな調整・変更などできない．

2.2.2 対話型

IEC ではユーザが評価値を与える．ユーザインタフェースについては4章で詳しく記述する．

$$f_{IEC}(t) = UserEvaluation$$

IEC を使うことの特徴として以下があげられる．

- ユーザの好みにあった演奏表情ルールを最適化できることが期待できる．
- 演奏データを用意する必要がない．
- 個体の演奏を評価するために多くの時間が必要．

IEC において問題点として共通するのは，評価回数の増加に伴ってユーザの負担が増加し，実現できる評価回数が非常に制限されてしまう点である．進化計算は集団ベースの確率的探索手法であり，本来，膨大な解空間を探索するためには十分な集団数と世代数が必要である．この問題を回避するために考えられるのは，探索範囲を予め限定する，ユーザの評価値を学習するような仕組みを導入することである．

3. IEC の選択提示によるユーザサポート

前節で述べたように IEC の最大の問題点は，大量の評価回数からくるユーザの疲労であ

る．本システムでは，ユーザの好みを学習し，評価回数を減らすための機構として，ガウシアンプロセスを使用する．既存研究では NN(Neural Network) やファジー推論などが使われてきた．本研究でガウシアンプロセスを使用する理由は，非線形関数を扱えること，以下に述べるように分散や密度関数が解析的に求めることが可能なためである．

ガウシアンプロセスはパラメータの分布及び出力の誤差として正規分布を仮定した回帰モデルの一種であり，カーネル法の一つとして解釈することができる^{1),5)}．IEC で評価する個体 X の特徴量を θ_X とする．ここでは特徴量として，出力された演奏表情の値の標準偏差，出力された演奏表情の値と音高との相関係数など合計 7 特徴量を使用した．

ガウシアンプロセスでは， N 個の学習データを $\{\theta_n, y_n\}_{n=1,2,\dots,N}$ とすると，新しい入力 θ に対する MAP(Maximum a Posteriori) 推定値は以下で計算される．

$$\hat{y}(\theta) = \sum_{n=1}^N w_n k(\theta_n, \theta)$$

$$\mathbf{w} = (\mathbf{G} + \beta \mathbf{I})^{-1} \mathbf{y}$$

ここで， $\mathbf{G}$ は要素 (i, j) を $k(\theta_i, \theta_j)$ とする Gram 行列である．また，以下のカーネル関数を使用した．

$$k(\mathbf{x}_A, \mathbf{x}_B) = \exp\left(-\frac{|\mathbf{x}_A - \mathbf{x}_B|^2}{2\sigma^2}\right)$$

3.0.3 選択選択

IEC フェーズでユーザに提示する個体を以下の要領で選択する．Marcus らの研究²⁾によって提案されたガウシアンプロセスの期待向上値を判断基準とする．これは新しい入力 θ に対するガウシアンプロセスの推定値が向上する期待値を与えるもので，MAP 推定値が大きく，かつ分散が大きい領域が大きな値を持つ．個体 X の特徴量 θ_X に関する，期待向上値 (Expected Improvement) $EI(\theta_X)$ を以下の式に従い計算する．

$$EI(\theta_X) = s(\theta_X) \left(u \Phi(u) + \phi(u) \right)$$

$$u = \frac{\hat{y}(X) - f_{best}}{s(\theta_X)}$$

$$\Phi(u) = \frac{1}{2} \operatorname{erf}\left(\frac{u}{\sqrt{2}}\right) + \frac{1}{2}, \quad \phi(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)$$

$\Phi(u)$, $\phi(u)$ はそれぞれ標準正規分布の確率密度分布と累積確率分布である． f_{best} は基準点であり，本システムでは個体の評価値の平均値を用いた．

IEC フェーズではなるべく少ない個体で，有用な好みの分布を得たいため，期待向上値が高い個体をユーザに評価させるべきである．本システムでは 10 個の個体について， $EI(X)$ を評価し，その中で $EI(X)$ を最大にする個体 X をユーザに提示する．これによって少な

い IEC の個体評価で有効なユーザの好み関数を獲得することを目指す．

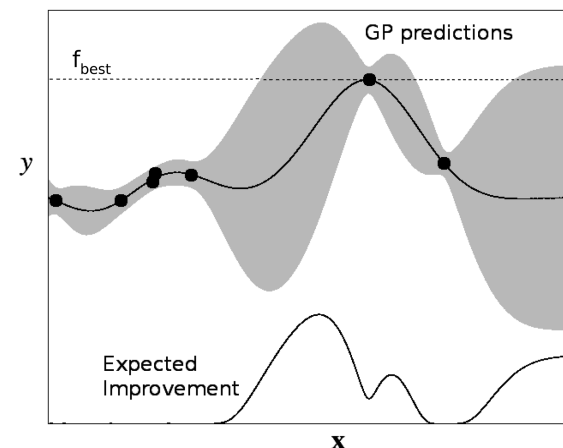


図 3 ガウシアンプロセスの予測曲線と分散と期待向上値．グレーの領域が分散の大きさを，下の曲線が期待向上値を示す．(文献²⁾より抜粋)

- 例えば (-CP CP) のような演奏表情を一様に出力する関数は，ユーザに低い点数を付けられると，再び提示されにくくなった．
- 学習したガウシアンプロセス上での分散が大きい個体が提示されやすいため，様々な個体が提示されるようになった．

4. WEB アプリケーションによる多人数型 IEC

ConBreO を多人数で使うために WEB アプリケーションとして拡張を行った．ConBreO WEB アプリケーション版と呼ぶ．5 に WEB ブラウザ上でのインターフェースのスクリーンショットを載せる．本システムは以下の URL でアクセスが可能である (<http://www.iba.t.u-tokyo.ac.jp/ConBreO/>) ．

4.1 インタフェース

複数の楽曲で評価して良いと評価された個体の方がロバストに演奏表情を生成できると考えられる．そのため，複数の楽曲の中からユーザが選択可能としている．図中上部の楽譜が並んでいる部分を選択することで，評価する楽曲を選ぶことができる．現在は選択可能な

楽譜は Crest Muse PEDB の MusicXML ファイルの 90 曲程度である。楽譜の選択は色々な曲を聴くことでユーザが飽きるのを防ぐ狙いがある。

図中下部では、演奏の聴取、視覚的な確認、評価値の設定、評価を完了される”Next Individual” ボタンがある。まず、ユーザは個体の生成した演奏を聴く。視覚的にも確認でき、ピアノロールを発展させたような画像で Velocity, Tempo, Attack, Release の演奏表情が付加された楽曲を視ることができる。これは、例えばまったく演奏表情に乏しい個体などが一目でわかり、聴かなくても評価ができるといったメリットと、個体の特徴を視覚的に覚えておくことに役立つ。

ユーザは生成された楽曲を聴いた後、星印により 5 段階の評価を付ける。その後”Next Individual” ボタンを押すことで、評価値を確定させサーバに情報を送る。

4.2 世代交代モデル

多人数で行う IEC の場合、以下の点で通常の IEC と違いがある。

- ・ユーザが評価する回数が予測できない。一個体だけ評価して終わるユーザもいる。
- ・平行して評価が行われるため、GP の世代交代をすると、前の世代の個体を評価しているユーザがいる可能性がある。

以上の理由により、ConBreO WEB アプリケーション版では、Steady State 型に似た GP を用いる。

- ・ユーザが評価を行った i 番目の個体は集団中の j 番目の悪個体と置き換えられる。ここでは通常の選択の逆の尺度で悪個体を選ぶ。
- ・2 個体を選択し、交叉・突然変異を行って生成された個体を i 番目の個体とする。

これによって、集団が一度に世代交代を行うことは無くなり、かつユーザのの評価値が共有される。

5. ま と め

本稿では、遺伝的プログラミングを用いた自動型と対話型の自動表情付けシステム ConBreO について述べた。ConBreO では個々の音符に対して 4 種のルールを適用することで演奏表情を得る。データベースからの自動型と対話型システムを組み合わせるシステムを導入した。また、IEC におけるユーザの評価回数を減らして、疲労を軽減するために”選択提示”と呼ばれるユーザサポート手法を導入した。ConBreO は WEB アプリケーションとして公開しており (<http://www.iba.t.u-tokyo.ac.jp/ConBreO/>)、多人数での評価が可能である。今回は定量的な評価を行っていないため、アンケート調査や WEB アプリケーション上でのユー

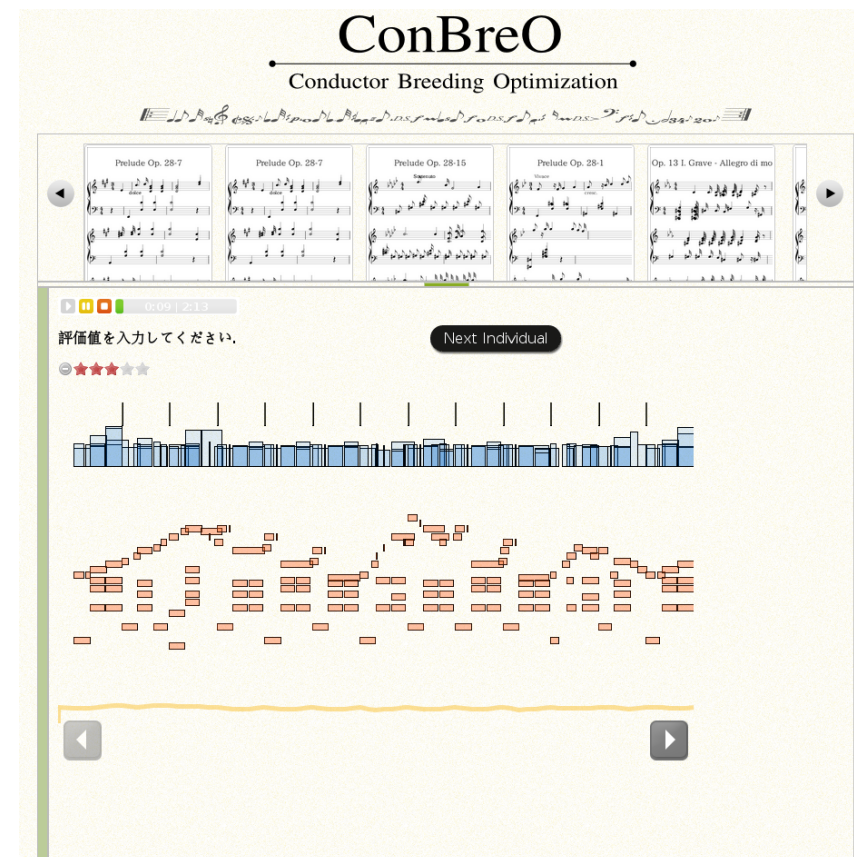


図 4 WEB ブラウザ上でのインタフェース。上の楽譜で演奏を評価する楽曲を選ぶ。下のパネルで演奏を聴き、また視覚的に確認できる。Next Individual ボタンを押すことで評価が完了し、次の個体に置き換わる。

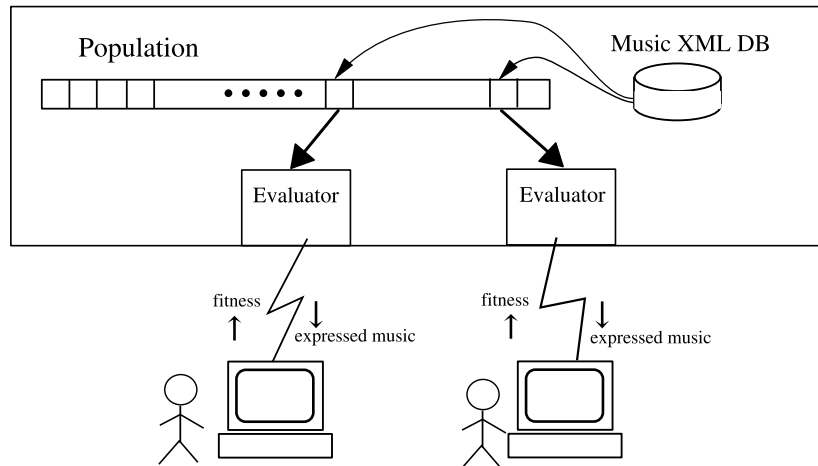


図 5 サーバ・クライアントモデルの模式図。ユーザが評価するのはサーバ上での集団の一個体。評価が完了すると別な個体と置き換わる。

ザの評価値などをもとに、定量的な評価を行う必要がある。

参 考 文 献

- 1) Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, 2006.
- 2) Marcus Frean and Phillip Boyle. *Lecture Notes in Computer Science, Vol. 5360*, chapter Using Gaussian Processes to Optimize Expensive Functions, pages 258–267. Springer Berlin / Heidelberg, 2008.
- 3) Anders Friberg. Generative rules for music performance: A formal description of a rule system. *Computer Music Journal*, 15, No. 2:56–71, 1991.
- 4) J.R. Koza. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, 1992.
- 5) D.J.C. MacKay. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, 2003.
- 6) Taizan Suzuki. The second phase development of case based performance rendering system “kagurame”. In *IJCAI*, 2003.
- 7) Keiko Teramura and Hideharu Okuma. Gaussian process regression for rendering music performance. In *International Conference on Music Perception and Cognition (ICMPC)*, 2008.

- 8) 橋田 光代, 松井 淑恵, 北原 鉄朗, 酒造 祐介, and 片寄 晴弘. 楽演奏表情データベース crestmusedb の計画と公開について. 音楽知覚認知学会, 2007.
- 9) 高木英行, 畝見達夫, and Tatsuo Unemi 寺野隆雄. 対話型進化計算法の研究動向. 人工知能学会誌, vol.13, no.5:pp.692–703, 1998.
- 10) 金泰憲, 深山 覚, 西本 卓也, and 嵯峨山 茂樹. 単旋律と和音の確率モデルの組み合わせによるピアノ曲演奏の自動表情付け. In 情報処理学会研究報告 MUS85, 2010.