

## 対話型作曲支援システムにおけるユーザ嗜好の獲得

稲田 雅彦 安藤 大地 丹治 信 伊庭 斉志

東京大学大学院新領域創成科学研究科基盤情報学専攻 〒277-8561 東京都文京区弥生 2-11-16

E-mail: {inada, dando, tanji, iba}@iba.k.u-tokyo.ac.jp

あらまし 対話型進化論的計算を用いた作曲支援システムに関する研究において、評価時のユーザ負担問題が問題点として挙げられる。フィルタリングやユーザインタフェースを改善することにより、これを軽減する研究が多くなされてきた。本稿では、個体特徴量およびユーザによる評価を、機械学習手法によりユーザの評価関数を同定するユーザ嗜好学習システムを提案する。ユーザ嗜好関数を獲得することで、ユーザ負担の軽減、感性を持った自動作曲システムへの応用が期待できる。

表現型の特徴量を訓練データとし、種々の機械学習手法を用いてシステムの評価を行った。実験結果によりユーザ嗜好学習の有効性を確認した。

キーワード 対話型進化論的計算, 作曲支援, ユーザ嗜好

## Acquisition of User Interests in a Composition-Aid IEC System

Masahiko INADA Daichi ANDO Makoto TANJI and Hitoshi IBA

Dept. Frontier Informatics, Graduate School of Frontier Sciences, University of Tokyo. 2-11-16, Yayoi, Bunkyo-ku, Tokyo 113-0032 Japan

E-mail: {inada, dando, tanji, iba}@iba.k.u-tokyo.ac.jp

**Abstract** We discuss the problem of user's burden in composition-aid IEC systems. Various filtering techniques and interfaces have been proposed to address this problem.

In this paper, we propose a new hybrid system which uses both IEC and machine learning. We aim to apply the system into automatic music composition system which has particular individual preference and reduce the user's burden to acquire user's interest.

We evaluate the correct ratios between interactive and automatic evaluation with various machine learning methods. We confirm our proposal that synthesized function produces user's interest and show its effect.

**Keyword** Interactive EC, Composition-Aid, User Interest

### 1. はじめに

本稿は、音楽創作支援のための対話型進化論的計算システムにおけるユーザ嗜好獲得アルゴリズムについて論じるものである。

これまで著者らは、対話型進化論的計算(Interactive Evolutionary Computation, IEC)を用いた音楽創作支援システムに関する研究を進めてきた[11][12] (図1)。

IECの問題点の一つとして挙げられるのが、ユーザが強いられる評価の負担である。この評価負担を軽減させるためにユーザの過去の評価の傾向を蓄積それに適合した個体のみを提示するフィルタを用いたシステム、個体の識別の問題や、ユーザの負担問題の解決を目的としたユーザインタフェースの提案などがされている[8]。

本稿では、提示個体の特徴量およびユーザの評価を強化学習を用いることで、ユーザ嗜好を学習するシ

テムを提案する。これによって、対話型進化論的計算手法の問題点であるユーザの負担問題を大幅に軽減させるとともに、ユーザ嗜好を備えた自動作曲システムに繋がるのが期待される。

本稿では、2章において、対話型進化論的計算の概要とユーザ負担の問題について述べる。3章で既存の研究の問題を解決すべく実装した、ユーザ嗜好学習システムについて述べる。4章では学習手法の分類精度などについて議論する。最後に5章でまとめを述べる。

### 2. IECを用いた作曲支援システム

#### 2.1. 対話型進化論的計算(IEC)

対話型進化論的計算は、生物の進化の過程をモチーフにした最適化手法である遺伝的アルゴリズム(Genetic Algorithm, GA)や遺伝的プログラミング(Genetic Programming, GP)などの進化アルゴリズム

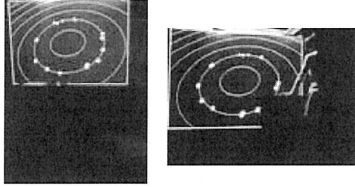


図 1: 対話型作曲支援システム CACIE

(Evolutionary Algorithm, EA) の評価関数を人間に置き換えたものである。

通常の EA において、進化した個体の評価は、システムが評価関数を用いて自動的に行うため、人間の関与は必要ない。しかし、現実世界の最適化問題を解こうとする時、評価関数を厳密に決められない問題がある。この場合、人間が個体に対して評価を与える、つまり人間が評価関数となることで最適を進めることができる。この手法が IEC である。

例えば、評価が個人や場合に依存してしまう場合がある。Takagi らは補聴器のデジタルフィルタの設計に対話型進化論的計算を用いた[7]。補聴器は、使用する患者の障害の度合いに応じたセッティングを用意する必要がある。Takagi らは、PDA を用いた携帯型評価システムを利用することで、対話型のフィルタ最適化システムを構築した。

### 2.2. ユーザ負担の問題

IEC の大きな問題点の一つとして、ユーザへの評価負担がある。ユーザの評価に対する負担問題としては、主に長時間の評価とユーザの疲労が挙げられる。

IEC を適用する場合は、前述の通り収束ポイントが最初から決まっているわけではないので、ユーザの判断の揺らぎや迷いが、判断に要する時間を大幅に増加させる。

ユーザの疲労に関しては、2D グラフィックの画像生成に IEC を用いた場合、訓練された人間が行った場合でも、連続して評価がまともにできる時間は 2, 3 時間であったと報告されている[2]。

以上により、IEC では評価回数を増やすことは出来ず、集団サイズの小型化と収束までに要する世代数の大なる縮小が IEC には必須の条件となる。また、評価を行うユーザの疲労を出来るだけ減らし、正常な判断をなるべく長く続けさせる工夫も必要となる。

### 3. ユーザ嗜好学習システム

本章では、提案するユーザ嗜好学習アルゴリズムについて述べる。対話型進化論的計算システムにおける個体の持つメロディから特徴量を抽出し、これとユーザによる評価値を訓練データとして機械学習手法を用いて学習を行う。本システムでは学習手法として遺伝

的プログラミングを主に用いる。以下において旋律の表現に必要な定義について述べた後、旋律特徴量、遺伝的プログラミング、および提案システムの説明を行う。

### 3.1. 旋律表現

複数の音から構成されるのが楽曲である[13]。ピッチ (Pitch) とは音の調子、音の高さを表す。これは周波数で規定され、記号で表される。ここで 440Hz の音は A (ラ) と表現され、倍の 880Hz の音は A の 1 オクターブ (Octave) 高い a を表す。また、ピッチは音 (Tone) とも呼ばれる。

音程 (Interval) はピッチ間の距離であり、1 オクターブ離れた音程を 12 等分した単位を半音 (Semitone) という。例えば A から a までは、A, A#, B, C, C#, D, D#, E, F, F#, G, G# と表す。G#, A#, C#, D#, F# はそれぞれ A ♭, B ♭, D ♭, E ♭, G ♭ とも表される。全音階的 (Diatonic) な音の並びとは、A, B, C, D, E, F, G, a または C, D, E, F, G, A, B, c のように、半音がそれぞれ 2, 1, 2, 2, 1, 2, 2 または 2, 2, 1, 2, 2, 2, 1 ずつ離れた配置を指す。前者を短調 (Minor)、後者を長調 (Major) という。また A, C をそれぞれの主音 (Key)、その他の 7 音を属音 (Dominant) という。全音階的ではない並びを無調的 (atonal) という。移調 (Transposition) とは、主音を短調、長調の並びを保ったまま相対的に移動することをいう。

各ピッチには長さ (Duration) があり、旋律 (Melody) とは長さを持つピッチの順次配置である。和音 (Chord) は複数のピッチの並列配置であり、これらは同時に演奏されることを想定している。音楽の意図する主要な旋律ではない部分を伴奏 (Accompaniment)、伴奏の付いた音楽を多声音楽 (Polyphony)、伴奏が無い音楽を単旋律 (Monophony) という。

五線譜 (Score) は、音の高さと長さを記入した記号 (音符) を順次的あるいは並列的に記述した表現である。旋律は音符の並びで、和音は音符の重なりで記述される。音符の長さの合計を拍 (Beat) という。小節 (Bar) とは同じ拍数を持つ五線譜の部分記述であり、拍子 (Time signature) によって小節あたりの拍数と拍子を表す。特にリズム (Rhythm) とは拍の規則的な並びを指す。

音楽は音楽として表現するために、多くの表情付けがなされている。これらは大きく 4 つに分類され、速度、強弱に関するもの、発想に基づくもの、奏法によるもの、装飾音に関するものがある。これらの表現は五線譜上に記載される。これらを計算機で表現するために、これまで、信号レベル (CD などでのアナログ的变化を表現)、操作レベル (MIDI など、時間のゆれを表現) および楽譜レベル (表現、解釈の多様性を表

現)で研究が行われてきた。

### 3.2. 旋律特徴量

旋律を分類するためには、当該音楽の有する特徴量の表現を考慮する必要がある、これらの条件は、不完全であることが多く、その大半は部分的である。

本システムでは、個体の有する旋律から特徴量を抽出し、ベクタ空間モデルを用いた特徴項目ベクタ空間の写像を行う。各次元は項目に対応し、どの旋律も空間内の点で表現できる。

#### 3.2.1. ピッチ輪郭

これまで提案されてきた旋律情報のための特徴量として、代表的なものに旋律輪郭 (Melody Contour) がある[3][4][10]。また、近年、楽譜化技法としてXMLに基づくものや五線譜利用による抽象化記述がある。旋律の輪郭情報のうち、ピッチ輪郭 (Pitch Contour) は多くの研究で用いられている。ピッチ輪郭表現は、単旋律において有効であり、直前の音と比べてピッチが相対的に高い、低い、同じという状態をそれぞれ、U, D, S で表し、輪郭を文字として記述する。この表現は、ピッチに対して相対的なので移調に強いが、雑音に反応して文字列の変化が起こりやすいといった問題が挙げられる。

ここで、図2を例にピッチ輪郭の説明を行う。図のピッチ輪郭は、-SUS USD DSUS DSDである。あるいは、長さを見無視すると、ピッチが同じ限り意味をなさないと考えられることから、記号 S を考えず、-UDDUDSD と表すこともできる。

ピッチ輪郭は様々な問題が挙げられており[9]、旋律記述に拍子、音符長を含まないため、判別のために多くの情報を必要とする。運命交響曲 (Beethoven)、雪山賛歌、Oh! My Darling, Clementine! (アメリカ民謡) は、ともに-SSD USSD と表され、判別を行うためにはより多くの情報が必要である。

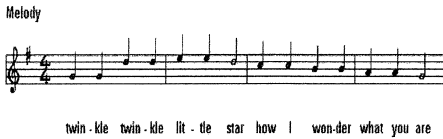


図2 メロディの例

#### 3.2.2. ピッチスペクトル

旋律が五線譜、楽譜で記述されるとき、小節単位にピッチスペクトル (Pitch Spectrum) を用いて特徴量として表現することができる[6]。ピッチスペクトルとは、小節内に生じる各音符を長さで集積し、そのヒストグラムをとることで得られる。ここでは、 $12 \text{音} \times 4 = 48$ 音である4オクターブの音を対象に、48次元ベクトル

を生成する。これを小節ごとにピッチスペクトルを算出することで得る。

図3に本システムから得られたピッチスペクトルを示す。ピッチスペクトルは各ピッチを長さで集積したヒストグラムであり、小節全体に和音を指定したものと異なる。ピッチ輪郭と異なり、小節単位の特徴を集約しているため、旋律の輪郭を表しておらず、むしろ、小節間の変化を記述しているため、旋律の時系列表現とも異なる。ピッチスペクトルは不完全な旋律に生じる問題を解消することができると考えられる。

ピッチスペクトルの問題としては、移調に変動しやすく、長調、短調の変化に反応しにくいことが挙げられる。前者の問題に対しては、旋律の主音から移調を行えばよいが、後者の問題に対しては、対応が困難であると考えられる。

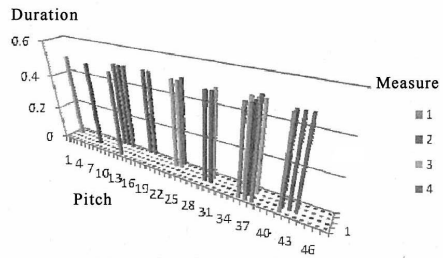


図3 ピッチスペクトル

### 3.3. 遺伝的プログラミング

遺伝的プログラミング (GP) は、遺伝的アルゴリズム (GA) の遺伝子型を木やグラフなどの構造的な表現が扱えるように拡張し、プログラムの生成や学習、推論、概念形成などに応用することを目的としており、GPの基本的思想はKozaらにより提案された[5]。GPの処理の流れは図4のようになる。

ここでGPの基本的なアルゴリズムは次のようになる。

#### (1) 初期設定

ランダムに初期世代の集団  $M(\phi)$  を生成する。

#### (2) 適合度計算

遺伝子プールの中から、ある選択基準で親を2個体決める。現在の集団  $M(t)$  内の各個体  $m$  に対して適合度  $u(m)$  を計算する。

#### (2) 親の選択

$u(m)$  に比例する確率分布を用いて、 $M(t)$  から個体  $m$  を選び出す。

#### (3) 遺伝的操作

選び出された個体にオペレータを作用させて、次の世代の集団  $M(t+1)$  を生成する。

(4) 終了処理

終了条件を調べて、終了していなければ、(2) から繰り返す。

以上が、標準的なアルゴリズムとなる。この処理自体は GA の処理と同様である。GP では遺伝子として木構造が用いられ、個体はプログラムを表す。木構造を用いると次式 (1) が図 5 のように表され、これらは、根 (ルート)、その枝葉となるノード、終端ノード、非終端ノードから構成される。

$$0.3y + x(0.1 + 0.34) \quad (1)$$

木に対する遺伝的オペレータとして、以下のオペレータが挙げられる。

- ・ 交叉

2 つの個体からノードをランダムに選択し、選択された部分木の交換を行うことで次世代の個体を生成する (図 6)。

- ・ 突然変異

終端ノードから非終端ノードへの新しい木の生成を行う突然変異やノードラベルの付け替えのみを行う突然変異、終端ノードから終端ノードへの突然変異などが挙げられる (図 7)。

- ・ 逆位

1 つの個体からノードをランダムに選択し、選択された部分木の交換を行う (図 8)。

個々の個体は評価関数によって評価され、評価関数は個体のプログラム自体に生成される出力情報から得られる。

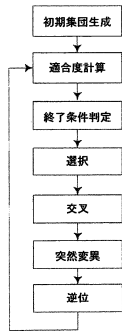


図 4 GP の処理の流れ

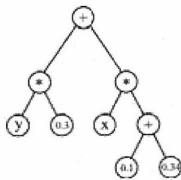


図 5 数式の木構造表現

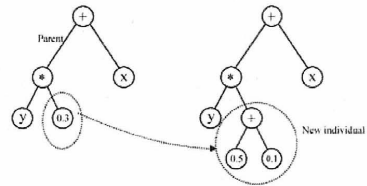
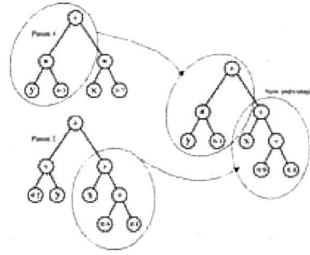


図 7 遺伝的的操作：突然変異

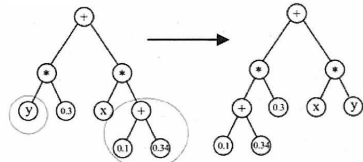


図 8 遺伝的的操作：逆位

### 3.4. GP によるユーザ嗜好の学習

ユーザが IEC において個体メロディを評価した履歴から、GP を用いたユーザ嗜好の学習を行う。得られた関数から IEC システムの評価値が生成される。

ユーザの感性情報を持った評価関数の GP による同定は、一般的なシステム同定問題に帰着できる。システム同定は、与えられた入出力データから未知のシステムの振舞いを予測する問題で、さまざまな分野で応用されている [1][14]。この問題は次のように形式的に定義される。 $m$  個の入力変数と 1 つの出力変数で規定される未知のシステムがあると仮定する。

$$y = \bar{f}(x_1, x_2, \dots, x_m) \quad (2)$$

ここで、入出力データの  $n$  個の実現値が与えられたとする。

INPUT	OUTPUT
$z_{11}, z_{12}, \dots, z_{1m}$	$y_1$
$z_{21}, z_{22}, \dots, z_{2m}$	$y_2$
...	...
$z_{N1}, z_{N2}, \dots, z_{Nm}$	$y_N$

このときシステム同定の目的は真の関数  $f$  の近似  $\bar{f}$  することである。近似関数を得ることである。近似関数  $\bar{f}$  を得ることにより、任意の入力  $(x_1, x_2, \dots, x_m)$  に対して出力の予測値  $\bar{y}$  が

$$\bar{y} = \bar{f}(x_1, x_2, \dots, x_m) \quad (3)$$

のように求まる。

### 3.4.1. 感性情報の学習

関数の終端記号は、個体メロディの旋律特徴量を用いる。旋律の特徴量に関して、ピッチ輪郭およびピッチスペクトルの各場合において適用方法を定義する。

ピッチ輪郭を用いる場合、旋律  $d$  に生じる上昇 (U)、同じ (S)、下降 (D) の発生個数が特徴量となる。一定数の小節 (4 小節) での出現数をその個体の特徴量とみなす。

ピッチスペクトルの場合、旋律  $d$  はベクトル列  $(w_1, \dots, w_k)$  であり、各ベクトル  $w_i$  は第  $i$  小節に対応している。各小節を 48 次元で考えるとき、特徴を先頭 4 小節で判断すると  $48 \times 4 = 192$  次元のベクトル空間で考えればよい。

非終端記号は数学的初等関数を用いた。GP による各々の数学的関数の生成には、各個体の特徴量およびユーザの評価値が読み込まれる。また、関数による評価値とユーザによる評価の比較から、得られた関数の妥当性を評価できる考えられる。

## 4. システムの評価実験

以下の実験ではユーザが IEC を用いてユーザがユーザ嗜好に基づく旋律を進化させ、その際に獲得したユーザ嗜好関数の評価を行う。

### 4.1. 評価方法

ユーザ嗜好学習システムの評価方法を以下に示す。

・16 個体のユーザによる個体評価を 15 世代分行い、Cross-validation によりユーザの個体評価に対してシステムがユーザの嗜好学習により推測した個体評価との正解率を算出する。なお、個体評価は 5 段階評価で行う。

・正解率は、

$$\frac{\text{正しく分類された個体数}}{\text{総個体数}} \times 100(\%)$$

で定義される。

表 1 GP のパラメータ

集団数	100
交叉確率	90%
突然変異確率	0.07%
終端ノード	{(0-191), (U), (D), (S)}
非終端ノード	{+, -, /, *, If, >, <, Pow, &,  , Max, Min, Exp, Log}
訓練データ数 (N)	15

・他の機械学習手法との比較を行うために、AdaBoost, ニューラルネット, 決定木の機械学習手法を適用し、同様に正解率の算出を行う。ここで本実験に用いた GP のパラメータを表 1 に示す。

・各手法に対して、特徴量として用いたピッチ輪郭, ピッチスペクトル, これら全てのデータセットを Cross-validation により正解率を算出する。

### 4.2. 実験結果

ピッチ輪郭およびピッチスペクトル, 両者の特徴量として適用した場合の手法ごとの正解率を図 9, 10, 11 に示す。ピッチ輪郭よりもピッチスペクトルの特徴量としての優位性が決定木を除く各学習手法においてみられた。また、GP, ニューラルネットでは特徴量を組み合わせることによって正解率の向上が確認できる。実験において GP によって合成された関数の例を以下に示す。

関数:

$$\text{Pow} / (-(\text{Co } 0.117, 35), \text{Log}(\text{Min}(179, \text{Co } 0.882))), \text{Log}(\text{+}(\text{Pow}(\text{Co } 0.117, S), \text{Pow}(179, S))))$$

また、各手法ともに世代が増加するに従って正解率が低下する現象がみられた。この原因としては、個体の多様性が低下し、個体差異が小さくなるためにユーザ嗜好の同定性能が低下するためであると考えられる。これに対して個体が収束するにつれて、個体特徴量を増やすことで学習性能の低下を防ぐことができると思われる。

### 4.3. 考察

システムによるユーザ嗜好学習は、表現型における基本的な特徴量からも高い学習性能が見られた。これをユーザ評価に支援することで IEC システムのユーザ負担問題の軽減が期待できる。

本研究では、特徴量としてピッチ輪郭, ピッチスペクトルのみを用いた。これらの特徴量には問題点も多く挙げられており、改良の加えられたものも多く提案されている。これらを考慮することによる学習精度の改善, また、表現型における特徴量だけでなく、遺伝型における個体特徴量を用いた学習を行うことできら



なる精度向上が期待できると思われる。

しかしながら、ユーザの評価にはばらつきが多く、評価をするにつれて評価が変化することも多くある。IECにはこれを吸収するロバスト性があるが、ユーザ嗜好学習システムには、ユーザによる評価のばらつきは訓練データのノイズになり、学習性能の低下に繋がる。今後は個体の提示方法の改善や提示数を動的に変更するなどの処理を行うことで問題を改善していきたい。

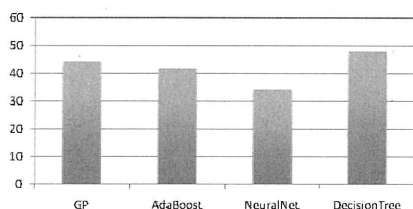


図9 正解率 (ピッチ輪郭)

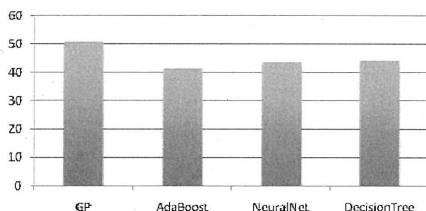


図10 正解率 (ピッチスペクトル)

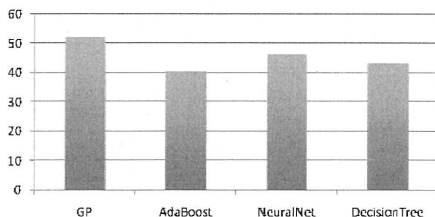


図11 正解率 (ピッチ輪郭およびピッチスペクトル)

## 5. おわりに

本稿では、音楽創作支援のための対話型進化論的計算システムにおけるユーザ嗜好獲得アルゴリズムを提案した。個体メロディの特徴量を数学的関数としてGPを用いて帰納的に獲得し、ユーザ嗜好獲得システムの有効性を検討した。また、システムの評価により、本システムの有効性を確認した。本システムで導入した機能は、ユーザが強いられる評価負担の軽減に有効だ

と考えられる。

本研究の今後の課題としては以下の項目が挙げられる。現在の特徴量には問題点も多く、特徴量の改善とともに表現型だけでなく、遺伝型の特徴量も取り入れ評価を行うこと、ユーザ支援を動的に行い、ユーザとの相互作用をより協調的に行い、システムの評価方法についても定量的な評価を検討すること、などについて取り組む予定である。

## 参考文献

- [1] Astrom, K. J. and Eykhoff, P.: System Identification, a survey, Automatica, Vol. 7, pp.123-162 (1971).
- [2] Cook, T. and Congdon, C. B.: Preliminary Results with GAUGUIN, an Evolutionary Computation Approach to Creating Art in the Suprematist Style", Proc. the IEEE Congress on Evolutionary Computation 2007(CEC2007), pp.4250-4258 (2007).
- [3] Dowling, W. J.: Scale and Contour - two components of a theory of memory for melodies, Psychological Reviews 85-4, pp.341-354 (1978).
- [4] Kim, Y. E. et al.: Analysis of A Contour-based Representation for Melody, Proc. International Symposium on Music Information Retrieval (2000).
- [5] Koza, J.: Genetic programming: A paradigm for genetically breeding populations of computer programs to solve problems, Report No. STAN-CS-90-1314, Dept. of Computer Science, Stanford Univ. (1990).
- [6] Miura, T. and Shioya, I.: Similarities among Melodies for Music Information Retrieval, ACM Conf. on Information and Knowledge Management(CIKM) (2003).
- [7] Takagi, H. and Ohsaki, M.: Interactive Evolutionary Computation-Based Hearing Aid Fitting, IEEE Transactions on Evolutionary Computation, 11, 3, pp.414-427 (2007).
- [8] Tokui, N. and Iba, H.: Music Composition with Interactive Evolutionary Computation, Proc. 3rd International Conference on Generative Art (GA2000) (2000).
- [9] Uitdenbogerd, A. L. et al.: Manipulation of Music For Melody Retrieval System, Intn'l Computer Music Conf. (ICMC)(1998).
- [10] Uitdenbogerd, A. L. et al.: Matching Techniques for Large Music Databases, ACM MultiMedia Conf. (1999).
- [11] 安藤大地, Dahlstedt, P., Nordahl, M. G., 伊庭斉志: 対話型GPを用いたクラシック音楽のための作曲支援システム, 芸術科学会論文誌, 4, 2, pp.77-86 (2005).
- [12] 稲田雅彦, 安藤大地, 丹治 信, 伊庭斉志: Sign Do Sound: 対話型作曲支援システムの能動的音楽聴取インタフェース, 映像情報メディア学会ヒューマンインフォメーション研究会 (2008).
- [13] 石桁真礼生他: 新装楽典-理論と学習, 音楽之友社, (2001).
- [14] 伊庭斉志, 佐藤泰介: システム同定アプローチに基づく遺伝的プログラミング, 人工知能学会誌, vol.10,no.4 (1995).