

テクニカルノート

# 音符を単位としない旋律編集のための旋律概形抽出手法

土屋 裕一<sup>1,a)</sup> 北原 鉄朗<sup>2,b)</sup>

受付日 2012年6月26日, 採録日 2012年9月10日

**概要:** 本論文では, 音楽の非専門家が旋律を編集するための, 音符によらない旋律表現を提案する. これまで様々な自動作曲システムが開発されてきたが, その結果を非専門家が編集するのは困難である. 我々は, 旋律を旋律概形と呼ばれる曲線に変換し, 旋律概形上での旋律編集を提案する. 旋律概形は, 旋律の音高の時系列に対してフーリエ変換を行い, 低次のフーリエ係数を抽出することで得ることができる. 実験の結果から, 提案手法により得られた旋律概形は, 旋律の大まかな特徴を反映していることが分かった.

**キーワード:** 旋律編集, フーリエ変換, 旋律表現, 旋律概形, 隠れマルコフモデル

## Melodic Outline Extraction Method for Non-note-level Melody Editing

YUICHI TSUCHIYA<sup>1,a)</sup> TETSURO KITAHARA<sup>2,b)</sup>

Received: June 26, 2012, Accepted: September 10, 2012

**Abstract:** In this paper, we propose a non-notewise melody representation to enable untrained users to edit melodies. Although various automatic music composition systems have been developed, few systems enable untrained users to edit the generated melodies. We propose a method for transforming a melody to a curve called a melodic outline to achieve melody editing based on the outline. The melodic outline is obtained by extracting low-order Fourier coefficients after applying the Fourier transform to the pitch trajectory of the melody. Experimental results show that the obtained melodic outline reflected the coarse characteristics of the original melody.

**Keywords:** melody editing, Fourier transform, melody representation, melodic outline, hidden Markov model

### 1. はじめに

自動作曲システム [1], [2], [3], [4], [5] は, 作曲スキルのない人が手軽にオリジナル曲を得る手段として, 近年, 注目を集めつつある. これらのシステムは, ユーザが入力した歌詞などに基づき, 自動的に旋律を生成する. しかし, 実際には, ユーザがシステムに楽曲への要望を詳細に伝えることは容易ではなく, ユーザが十分に満足いく楽曲が生成されるのは, むしろまれである. このような場合, 通常

は, 市販の MIDI シーケンサなどを使って手動で生成結果を編集して楽曲を仕上げていくことになるが, 作曲スキルのないユーザには, 困難な作業である.

本研究のゴールは, 市販の MIDI シーケンサを使えないようなユーザ (以下, 非専門家という) でも, 自動作曲システムが生成した旋律に対して修正を試行錯誤し, 自分なりに満足のいく旋律を探求する環境を実現することである. 非専門家が MIDI シーケンサを使うのが難しい理由は, 2つあると考えられる. 1つは, 非専門家は, 旋律を音符列として聴取しているとは限らないということである. 旋律の認知や記憶には, 旋律全体の輪郭が重要であるといわれている [6]. そのため, 編集のための旋律表現も, 旋律全体の輪郭を表すような表現であれば, 非専門家でも試行錯誤がしやすくなると予想される. もう1つは, 非専門家には, スケール外の音や伴奏と不協和を生じさせうる音を避

<sup>1</sup> 日本大学大学院総合基礎科学研究科  
Graduate School of Integrated Basic Sciences, Nihon University, Setagaya, Tokyo 156-8550, Japan

<sup>2</sup> 日本大学文理学部  
College of Humanities and Sciences, Nihon University, Setagaya, Tokyo 156-8550, Japan

a) tsuchiya@kthrlab.jp

b) kitahara@chs.nihon-u.ac.jp

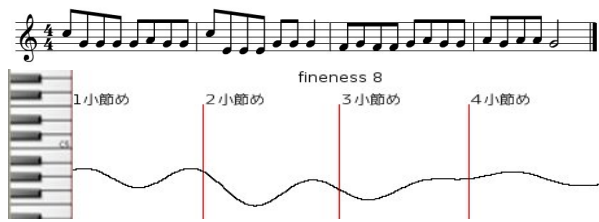


図 1 旋律概形の例  
Fig. 1 Example of melodic outline.

けることが難しいということである。そのため、非専門家が試行錯誤しやすくするには、このような音を計算機が自動的に避ける機構が有効である。

本論文では、旋律全体の輪郭を 1 本の曲線で表した「旋律概形」と呼ばれる旋律表現を提案する。この表現は、旋律の音高の時系列をフーリエ変換し、低次の係数のみを残して逆フーリエ変換することで得られる。旋律概形が得られたら、ユーザはその概形を自由に描き直すことができる。その後、その概形のフーリエ係数と元の旋律の高次のフーリエ係数を組み合わせて逆フーリエ変換することで音高の時系列を再生成し、隠れマルコフモデルを用いてスケールに合った旋律を推定する。このように、旋律の音高や音価を陽に表現せず、音楽の詳細をあえて示さないことで、音楽の非専門家が旋律の全体像の編集に注力できるようになり、非専門家にとってのユーザビリティが向上すると期待できる。

## 2. 旋律概形およびそれを用いた旋律編集

「旋律概形」とは、旋律の大まかな形が 1 本の曲線で表されたものである。旋律概形の例を図 1 に示す。これは、MIDI シーケンスのような音符列から自動的に得られ、ユーザがこの曲線を描き直した後、あらためて音符列に逆変換されることを想定している。そのため、単に音高の時系列を平滑化するのではなく、編集後に音符列に戻せる仕組みが必要である。我々は、旋律概形は次の 4 つの要件を満たすべきと考える。

- (1) 各音符の音高や音価が陽には表現されない。
- (2) 音符表現を旋律概形に変換し、編集せずに音符表現に再変換すると、元の音符表現に戻る。
- (3) 旋律概形上で編集を行って音符表現に変換した場合、音楽的に不適切な音は避けられる。
- (4) 旋律概形に旋律のどの程度細かな動きが表現されるかは、ユーザが制御できる。

これまで、この 4 つの要件を満たす旋律概形表現は、提案されていなかった。旋律を粗い表現に変換する手法は、様々なものが提案されていた（たとえば文献 [7]）が、音楽検索における旋律の照合を目的としており、音符列に逆変換することは指向していなかった。ユーザが描いた曲線から旋律を生成する作曲支援システムとして Lazy Composer [8]

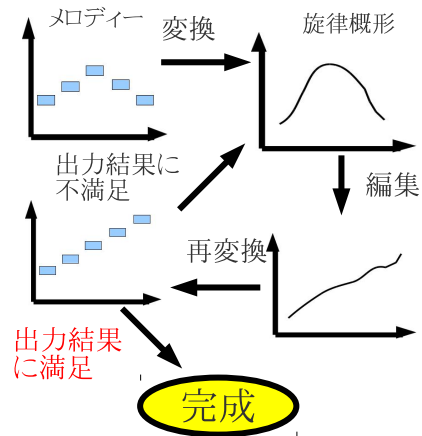


図 2 楽曲編集の流れ  
Fig. 2 Flow of melody editing.

が提案されている。曲線から音符列を生成する点は本研究と共通しているが、音符列表現との相互変換を指向しているものではなく、また、旋律概形の細かさを制御することはできない。OrpheusBB [9] は、本研究と同様の視点に立ち、自動生成された旋律などを編集する機能が提供されている。確率モデルを用いて主旋律と伴奏とが不協和にならないように調整する工夫は行っているが、編集対象は音符である。

本研究で提案する「旋律概形」を用いた楽曲編集の流れを図 2 に示す。ユーザは、初めに何らかの自動作曲システムを用いて旋律を生成することを想定している。その後、その旋律が、3.1 節で述べる手法により旋律概形に変換される。ユーザは、その旋律概形を自由に描き直すことができる。そして、3.2 節で述べる手法を用いることで、旋律概形が音符列に逆変換される。このようにして得られた音符列を試聴し、もしも気に入らなかった場合、改めて旋律概形に変換して描き直す。これを何度も繰り返すことでユーザが納得できる旋律に近づけていく。また、試行錯誤のサイクルの中で、旋律概形の「細かさ」は自由に変えることができるので、必要に応じて細かさを変えて編集することができる。たとえば、最初は細かさのパラメータを低い値にし、旋律の大まかな特徴を作った後、このパラメータの値を高くして細部を作り込んでいく、といった使い方ができる。

## 3. 音符列と旋律概形の相互変換手法

### 3.1 音符列からの旋律概形抽出手法

音符列から旋律概形を抽出する手法の概要を図 3 に示す。まず、入力された MIDI シーケンス (図 3(a)) は、音高の時系列 (図 3(b)) に変換される。ここでいう音高の時系列とは、旋律の音の高さを時間の流れに沿って並べた系列のことである。音高が対数で表現され、その値は中央の C の音が 60.0、半音が 1.0 (ノートナンバの表現とは異なり、非整数値でもかまわない) となる。休符については、直前の音が継続しているものとして扱う。

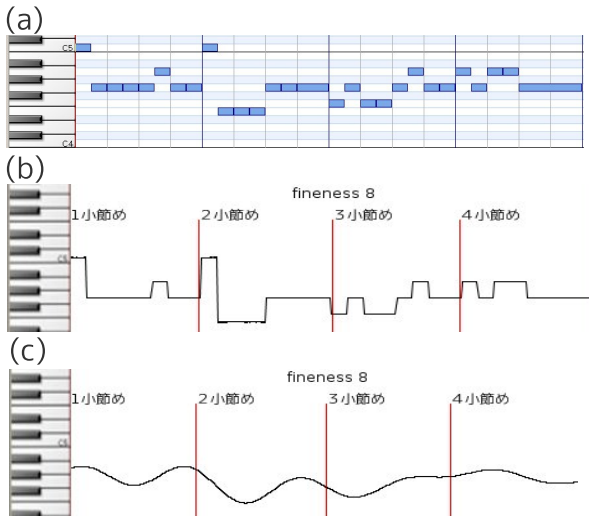


図 3 音符列から旋律概形を抽出する手法の概要. (a) 旋律の MIDI シーケンス, (b) 音高の時系列, (c) 旋律概形

Fig. 3 Overview method of extracting note sequence to melodic outline. (a) MIDI sequence of melody, (b) Pitch trajectory, (c) Melodic outline.

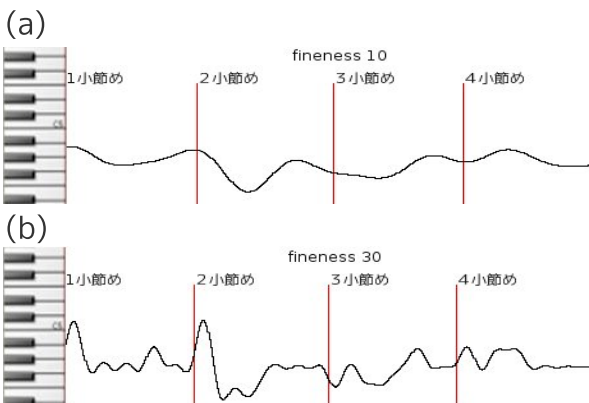


図 4 低次の次数の範囲で表現される旋律概形の粗さ. (a) 低次の次数が 10 次までで表現される図, (b) 低次の次数が 30 次までで表現される図

Fig. 4 Fineness in melodic outlines. (a) Melodic outline with fineness of 10, (b) Melodic outline with fineness of 30.

次に、この音高の時系列を周期信号と見なしてこの信号全体に対して離散フーリエ変換を行う。このとき、高次のフーリエ係数は旋律の細かい特徴を表現するのに対して、低次のフーリエ係数は旋律の大まかな動きを表すものである。そこで、フーリエ変換で得られた係数のうち、低次のもののみを取り出して逆フーリエ変換を行う。こうすることで、音高の時系列から旋律の大まかな成分のみを取り出すことができる (図 3(c))。

このように、フーリエ係数の低次のもののみを用いることで、2章で議論した要件 (1) を満たすことができる。また、抽出するフーリエ係数の次数の上限を変えることで、旋律概形が表す旋律の細かさを制御することができ (図 4)、要件 (4) を満たす。

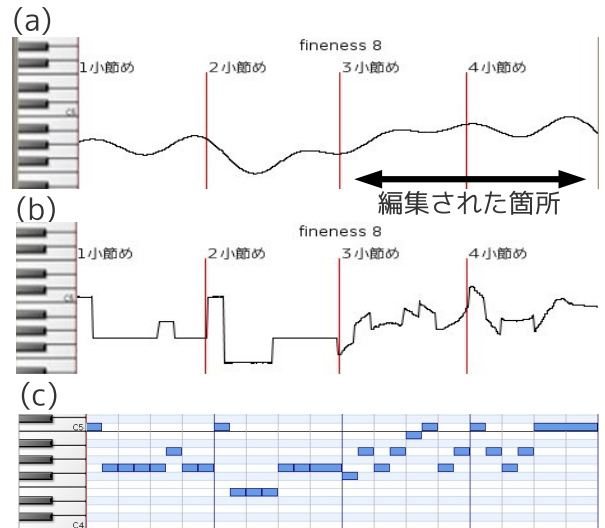


図 5 編集後の旋律概形から音符列を再生成する手法の概要. (a) 編集後の旋律概形, (b) 編集後の音高の時系列, (c) 編集後の旋律

Fig. 5 Overview of transforming melodic outline to note sequence. (a) Edited melodic outline, (b) Generated pitch trajectory, (c) Generated melody.

### 3.2 旋律概形からの音符列生成手法

編集後の旋律概形から音符列を再生成する手法の概要を図 5 に示す。編集後の旋律概形 (図 5(a)) をフーリエ変換して得られたフーリエ係数と、旋律概形抽出時に使用しなかった高次のフーリエ係数を結合して逆フーリエ変換を行う。これにより、旋律の大まかな流れにユーザによる操作を反映させながら、旋律の細かい動きに元の旋律の特徴を残した音高の時系列が得られる (図 5(b))。また、まったく編集せずにこの操作を行った場合は、元の音高の時系列と同じものが得られる (2章の要件 (2) を満たしたことになる)。

音高の時系列が得られたら、それを音符列に変換する。簡単のため編集対象を音高のみとし、各音符の発音時刻や消音時刻は元の旋律と等しいものとする。その場合、決定すべきは各音符のノートナンバのみなので、各音符の発音区間 (発音から消音までの区間) における音高の時系列の代表値 (中央値など) とすればよい。しかし、編集した箇所を音高の時系列に変換すると、図 5(b) のように歪みが生じる。そこで、隠れマルコフモデル (HMM) を用いて、この歪みのある音高の時系列から、スケールに即した音符列を推定する。

使用する HMM の概要を図 6 に示す。歪んだ音高の時系列は、音楽的に正しいノートナンバの系列にノイズが加わって観測されたものであると考える。各状態が各ノートナンバに対応し、正規分布に従ったノイズが付加されて音高の値が出力されると考える。たとえば、ノートナンバ 60 に対応する状態は、60.0 の平均と  $\sigma^2$  の分散を持つ正規分布に従って音高の値を出力する。分散  $\sigma^2$  の値はすべての状態で共通であり、実験的に決定される。遷移確率は、生

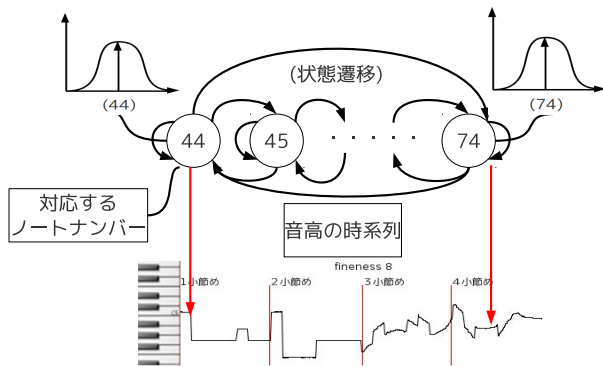


図 6 音高の時系列から音符列を推定する HMM の概要  
 Fig. 6 Overview of HMM for estimating note sequence.

成される旋律の主にスケール的な特徴および隣り合う音符どうしの遷移の特徴を表す. たとえば, メジャースケールの音に対応する状態への遷移確率が高く設定されている場合は, 推定される音符はメジャースケールのものになりやすくなる. 適切に状態遷移確率を設定することで, スケールに即していない音高が出力される可能性を抑えることができる (2章の要件 (3) を満たしたことになる).

現状では簡単のため, 楽曲全体がハ長調である (転調はしない) と仮定し, ハ長調のダイアトニックスケールに即した音符列が出力されるようにする. 具体的には, 同スケールに含まれる音への状態遷移確率には高い値を, そうでない音への状態遷移確率には低い値を手で設定している.

この HMM に対してビタビアルゴリズムを適用することで, 歪んだ音高の時系列 (図 5(b)) からスケールに即したノートナンバ列を推定することができる. ただし, ここでは旋律の音符数やリズムは編集前のものと等しくするため, 音符と音符の変わり目でのみ状態遷移するものとする. このようにしてノートナンバ列が得られたら, 音符列 (実際には MIDI シーケンス) として出力する (図 5(c)).

## 4. 提案手法の試用

### 4.1 提案手法を用いた旋律編集の例

前章までで提案した手法に基づき, 旋律から旋律概形を抽出して編集を行う過程の一例を示す. 元となる旋律は, 歌詞に基づく自動作曲システム Orpheus [5] を用いて作成した. Orpheus に詩『夢みたものは…』(立原道造) [10] の一節「夢みたものはひとつの幸福/ねがったものはひとつの愛」を入力して得られた旋律の冒頭 4 小節 (図 7(a)) を用いた. これを旋律概形に変換したものが図 7(b) である. 1 小節めと 2 小節めの冒頭に跳躍進行があること, 3 小節めから 4 小節めにかけて音高が徐々に上がっていくこと, 4 小節めの後半で音高が少し下がって旋律が終了することを読み取ることができ, 旋律の大まかな動きが適切に表現されているといえる.

図 7(c) は, 旋律概形を一部描き直したものである. こ

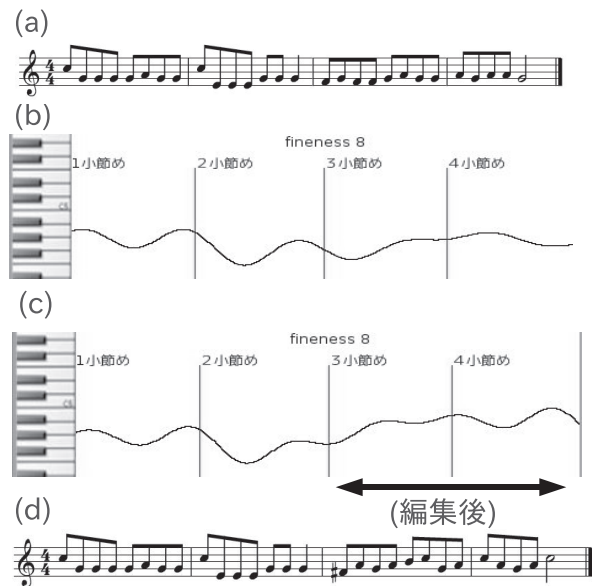


図 7 旋律編集の例. (a) 入力された旋律, (b) (a) の旋律概形, (c) (b) を編集した旋律概形, (d) (c) を音符表現に変換した旋律

Fig. 7 Example of melody editing. (a) Input melody, (b) Melodic outline of (a), (c) Edited melodic outline, (d) Note representation of generated melody.

では, 3~4 小節めの音高が徐々に上がっていく箇所に対して, 上昇率を上げ, 現状よりも高い音まで達するようにする, という意図を持って描き直した. もし, まったく編集を行わなかった場合, 2章で述べたように元の旋律と同様な旋律が生成される. この描き直した旋律概形から生成された旋律を図 7(d) に示す. 編集時の意図どおり, 編集前よりも高い音に達する旋律を得ることができた. 生成された旋律を見ると, ユーザが描いた旋律概形よりも音高が上下していることが分かる. これは, 旋律が単調になることを避けるために, HMM の状態遷移確率において同じ状態 (ノートナンバ) に停留する確率を低めに設定しているからである. 2~3 度上または下に移行する確率が高めに設定されているから, このような旋律が生成されたと考えられる. また, 編集した箇所が一番最初の音がファ♯になった. ノンダイアトニックノートは, ダイアトニックノートに比べて出現する確率は低く設定されているが, 0 にはなっていない. 今後は, 和声も考慮してノンダイアトニックノートの出やすさを制御できるようにすることが望まれる.

このように, 生成結果に一部ノンダイアトニックノートが出てくる場合があるが, それ以外については, 音符列から旋律概形の抽出, 旋律概形の編集, 旋律概形から音符列の再生成は正常に動作したと考える.

### 4.2 被験者による試用

被験者に同様の旋律編集を試してもらった結果について述べる. 前節と同様に, Orpheus を用いて元となる旋律を作成した. 『クローバーの原っぱで…』(高橋順子) [10] の



図 8 被験者実験の出力結果  
Fig. 8 Melodies created by subjects.

表 1 質問紙の結果 (指示付き編集)  
Table 1 Questionnaire results (instructed editing).

	A	B	C	D	E	F	平均
Q1	6	5	7	6	7	7	6.3
Q2	6	7	5	6	7	6	6.1
Q3	5	6	6	6	6	6	5.8

一節「お酒を呑んでもいい/それが楽しいことだったら」を入力した。これにより生成された旋律を図 8(a) に示す。旋律編集は各被験者に 2 回ずつ行ってもらった。1 回めは、「4 小節めの音を少し低くするように」という指示をし、その指示に従って編集してもらった。2 回めは、被験者自身の自由な発想で編集してもらった。それぞれの編集後には質問紙により次の項目について 7 段階で回答してもらった。

- Q1 出力結果に満足したか。
  - Q2 編集操作は難しくなかったか。
  - Q3 思いどおりに旋律を編集することができたか。
- 被験者は大学生 6 名とした。

指示付きでの編集に対する質問紙の結果を表 1 に示す。おおむね良好な結果を得ることができた。編集結果のうち、被験者 C と被験者 F によって生成された旋律を図 8(b), (c) に示す。(b) については、指示どおりに元の旋律よりも 4 小節めが低い音になっており、かつ音楽的に違和感のない旋律であった。(c) については、一部元の旋律よりも高い音は存在するが、こちらも音楽的に違和感のない旋律であった。

自由な発想による編集に対する質問紙の結果を表 2 に示す。こちらもおおむね良好な結果を得ることができた。編集結果のうち、被験者 A と被験者 E によって生成された旋律を図 8(d), (e) に示す。どちらの旋律も、音楽的におおむね違和感のない旋律となった。(e) の 3 小節めは最初の音がラリとされており、聴取者によっては違和感を覚

表 2 質問紙の結果 (自由な発想による編集)

Table 2 Questionnaire results (free editing).

	A	B	C	D	E	F	平均
Q1	6	6	6	5	6	5	5.6
Q2	6	7	7	3	6	6	5.8
Q3	6	3	6	3	7	6	5.1

える可能性のある音となっているが、当該被験者の Q3 の回答が 7 であることをふまえると、本人としては問題のない旋律ととらえていると解釈できる。また、Q3 に 3 を回答した被験者が 2 名いた。これは進行の都合上、試行錯誤をする時間が十分にとれなかったからと考えられる。今後は、1 試行に対する時間を長くとり、試行錯誤の様子を詳細に記録して、より詳細な分析を行いたい。

## 5. おわりに

本論文では、旋律概形と呼ばれる旋律の表現法を提案し、これを旋律からフーリエ変換によって抽出する手法を提案した。試用による結果、フーリエ変換によって抽出した旋律概形が、適切に旋律の大まかな形を表していることが分かった。また、旋律概形から音符列を逆生成する手法も提案した。これは、旋律概形抽出時に使わなかった高次のフーリエ係数を旋律概形のフーリエ係数と結合させて逆フーリエ変換することによって得られる。音楽的に不適切な音を避けるため、逆フーリエ変換の結果から、HMM を使って音楽的に適切な音符列を推定する。これにより、音楽的におおむね問題のない旋律が生成された。ただし、まれにノンダイアトニックノートが選ばれることがあり、課題が残った。

今後は、編集後の旋律概形から音符列を生成する処理について改良を進めていく。まず、現在の手法では和声を考慮していないので、和音に応じて HMM の状態遷移確率を切り替える仕組みを導入する。また、状態遷移確率を曲調ごとに最適化されたものを用意し、ユーザが切り替えることで、生成される旋律の曲調を変化させられる仕組みも導入したい。その他、状態遷移確率を実楽曲のコーパスから学習すること、直前の音符のみではなく、より長い履歴を元に状態遷移先を決定するようにすることなどがあげられる。

謝辞 有益なご意見をくださった松原正樹氏 (筑波大学) に感謝する。

## 参考文献

- [1] Hiller, L. and Lsaacson, L.: Musical composition with a high-speed digital computer, *Journal of Audio Engineering Society* (1958).
- [2] Ames, C. and Domino, M.: Cybernetic composer: An overview, *Understanding Music with AI*, AAAI Press, pp.186-205 (1992).
- [3] Cope, D.: *Computers and Musical Style*, Oxford Uni-

- versity Press (1991).
- [4] 安藤大地, Dahlstedt, P., Nordahl, M.G., 伊庭齊志: 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, Vol.4, No.2, pp.77–86 (2005).
  - [5] 深山 覚, 中妻 啓, 米林雄一郎, 酒向慎司, 西本卓也, 小野順貴, 嵯峨山茂樹: Orpheus: 歌詞の韻律に基づいた自動作曲システム, 情報処理学会研究報告, 2008-MUS-76-30, pp.179–184 (2008).
  - [6] ハーグリーブス, D.J.: 音楽の発達心理学, 田研出版株式会社 (1993).
  - [7] Marolt, M.: A mid-level representation for melody-based retrieval in audio collections, *IEEE Trans. Multimedia*, Vol.10, No.8, pp.1617–1625 (2008).
  - [8] 池田輝政, 菱田隆彰: 劇伴の作成を支援する楽曲作成システム「Lazy Composer」, 情報処理学会研究報告, 2011-MUS-91-10, pp.1–4 (2011).
  - [9] 北原鉄朗, 深山 覚, 片寄晴弘, 嵯峨山茂樹, 長田典子: OrpheusBB: Human-in-the-loop 型の自動作曲システム, インタラクシオン 2011, pp.57–64 (2011).
  - [10] 川口晴美: 名詩の絵本, ナツメ社 (2010).



土屋 裕一

2012 年日本大学文理学部情報システム解析学科卒業。同年同大学大学院総合基礎科学研究科入学。音楽情報処理に興味を持つ。日本音響学会学生会員。



北原 鉄朗 (正会員)

2002 年東京理科大学工学部卒業。2007 年京都大学大学院情報学研究科博士後期課程修了。博士(情報学)。日本学術振興会特別研究員(DC2), 科学技術振興機構 CREST 研究員を経て, 現在, 日本大学文理学部専任講師。音楽・音声等の音メディアの情報処理全般に興味を持つ。京都大学第 2 回総長賞等受賞。電子情報通信学会, 人工知能学会, 日本音響学会各会員。