

# 旋律を簡約・操作する一手法

## A Method for Melody Reduction and Manipulation

北原 鉄朗\*<sup>1</sup>      土屋 裕一\*<sup>1</sup>  
Tetsuro Kitahara      Yuichi Tsuchiya

\*<sup>1</sup> 日本大学文理学部

College of Humanities and Sciences, Nihon University

We have proposed a melody reduction and manipulation method by extracting a melodic outline using the Fourier transform. A melodic outline is a non-notewise representation of melodies, so it is easy for musically untrained people to manipulate. In this paper, we present an overview of this melody reduction and manipulation method and discuss an extension of the method using the wavelet transform.

### 1. はじめに

人はどのように音楽を聴いているか——この問いに対する答えを文法や法則などの形でまとめた理論を「認知的音楽理論」と呼ぶ。その中で最も有名なものの1つに Generative Theory of Tonal Music (GTTM)[Lerdahl 83]がある。GTTMでは、音符列をひとまとまりとして認知する条件を「グルーピングルール」としてまとめており、このルール群に従って旋律から二分木が生成される。また、各ノードが持つ2つの子ノード間に「重要度」を導入し、適当な深さ以降の重要度が低い方のノードを省略することで、「重要な音符」のみを残した旋律を得ることができる(簡約(reduction)と呼ばれる)。2つの旋律に対して共通の旋律が得られるまで簡約を行い、この2つの旋律の音符同士の対応付けを行うことで、旋律間のモーフィングを行う手法も提案されている[浜中 08]。

では、作曲などを専門としない音楽の非専門家にとって、「重要な音符を残し、そうでない音符を切り捨てる」という簡約の方法は本当に妥当なのだろうか。

認知的音楽理論を構築し、計算機上に実装する目的は様々であるが、その1つに音楽の非専門家に対する音楽創作支援があげられる。作曲者は旋律を何らかの意図を持って創作すると考えられるが、音楽の非専門家は「こんな旋律を創りたい」という意図があったとしても、聴く側が汲み取れるようにその意図を表出することができない。実際の旋律(音符列)とそれを聴いたときに人間が受け取る認知構造の関係を計算可能な形で理論化できれば、人間の音楽認知に即した形で作曲意図を表出し、計算機がその意図を旋律に変換することも可能になる。

この場合、認知的音楽理論のベースとなる旋律表現(melody representation)が、非専門家による音楽の聴き方に適合していることが重要である。では、非専門家が音楽をどのように聴いているのか。あくまで経験に基づく答えでしかないが、我々は、少なくとも音符表現(notewise representation)がベースではないと考えている。音楽の非専門家は、個々の音符を1つ1つとらえて旋律を聴いているというよりも、「高い音から始めて段々音が低くなる」「最初は段々音が高くなって後半は音が低くなる」などのように、旋律の全体的な概形をとらえて旋律を聴いているはずである。少なくとも、隣り合う音符間で重要度を考えるような聴き方はしていない。

我々は、このような観点から音符表現に立脚しない新たな認知的音楽理論の構築を目指して、旋律の新たな簡約手法を提案した[土屋 13a, 土屋 13b, 土屋 14]。この手法では、旋律は旋律概形と呼ばれる曲線に変換される。この旋律概形では、旋律中の音の上がり下がりに関する大まかな情報のみを表し、調性、和声、リズムなどの情報はあえて隠されている。また、旋律概形は音符列に再変換することができる。そのため、たとえば既存の旋律の旋律概形を書き換えて音符列に再変換することで、調性・和声・リズムを保存したまま音の上がり下がり进行操作することができる。本稿では、旋律概形に基づいた旋律の簡約・操作手法の概要を述べた後、その拡張について議論する。

### 2. 旋律概形に基づいた旋律の簡約・操作手法

#### 2.1 旋律概形とは

旋律概形とは、旋律の大まかな形を1本の曲線で表したものである。旋律概形の例を図1(c)に示す。旋律概形は旋律を簡約する一手法であるが、前述の通りGTTMのように「重要な音符」を抽出して旋律を簡約化するのではなく、旋律全体のいわば分解能を下げることで細かな音符の変化を見えなくする手法である。

この簡約方法を考える上で、我々は旋律概形が次の4つの要件を満たすべきであると考えた。

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、パラメータによって制御できる。

1. は上述の通り、音楽の非専門家による旋律の認知に即した表現を指向するためである。2., 3. は旋律概形を作曲支援に応用するためである。旋律概形を音楽の非専門家が自由に書き換えて、それを旋律に再変換することによって作曲を行うことを想定しており、そのためにこれらの要件が必要となる。4. は、旋律認知の分解能は聴取者の音楽の経験や旋律の複雑さによって異なると予想され、その分解能に合わせた旋律概形を得るために必要となる要件である。

連絡先: 北原 鉄朗, 日本大学文理学部, 東京都世田谷区桜上水  
3-25-40, kitahara@chs.nihon-u.ac.jp

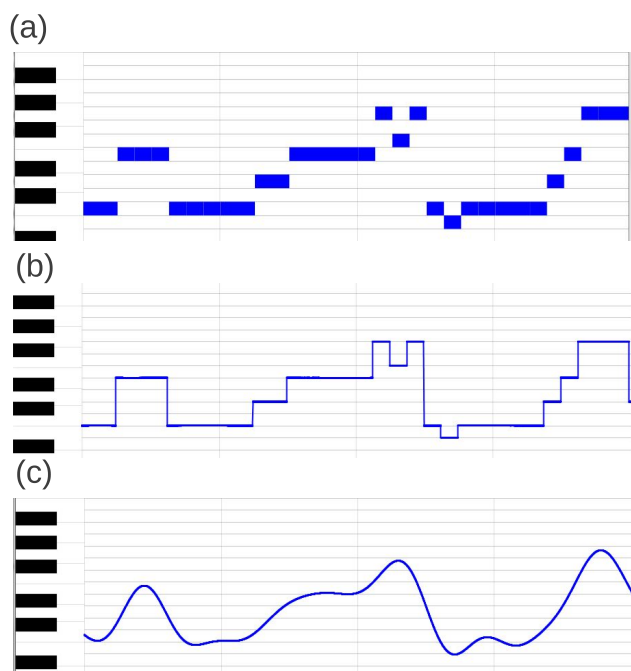


図 1: 音符列から旋律概形を抽出する手法の概要

我々は、この 4 つの要件を満たす旋律概形を旋律（音符列）から自動的に得る手法を提案した。この手法では、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のみを残して逆フーリエ変換を行う。これによって旋律の大まかな形のみを残した曲線を得ることができる。

## 2.2 旋律の簡約手法

音符列から旋律概形を抽出する手法の概要を図 1 に示す。

まず、旋律（図 1(a)）が、音高の時系列（図 1(b)）に変換される。ここで言う音高の時系列とは、旋律の音の高さを時間軸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 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。

次に、この音高の時系列を周期信号とみなしてこの信号全体に対して離散フーリエ変換を行う。この時、高次のフーリエ係数が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。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 1(c)）。使わなかった高次のフーリエ係数は旋律への再変換時に使用するの保持しておく。

この旋律の大まかな成分のみを表した曲線（図 1(c)）を旋律概形と呼ぶ。この概形は、図から明らかなように要件 (1) を満たす。また、この旋律概形に対して改めてフーリエ変換を行い、上で保持しておいた高次のフーリエ変換と足し合わせて再度逆フーリエ変換することで、元の旋律の音高の時系列が得られる（要件 (2) を満たす）。詳細は後述するが、音高の時系列から旋律（音符列）を得る際に、隠れマルコフモデルを用いて音楽的に不適切な音は避けるようにしている（要件 (3) を満たす）。さらに、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る（図 2、要件 (4) を満たす）。

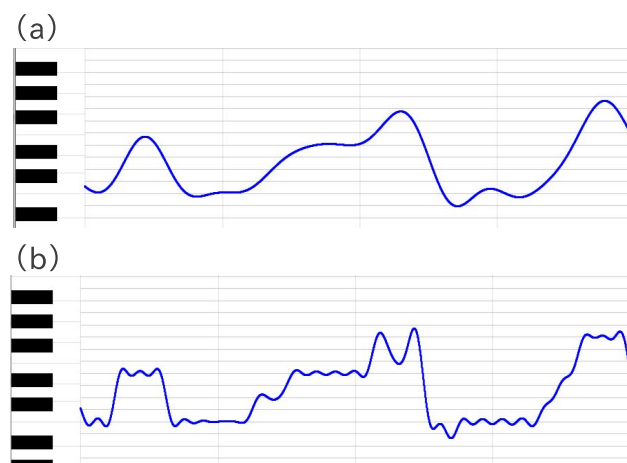


図 2: 低次の次数の範囲で表現される旋律概形の粗さ。

## 2.3 旋律の操作（編集・再生成）手法

旋律概形を用いると、音符レベルでの操作をせずに旋律を操作することができる。ここで、操作とは旋律に加工や編集を加えて別の旋律に作り替えることを意味する。旋律の操作は、旋律概形をユーザがマウスなどで書き換えることで行う。ユーザが旋律概形を書き換えると、その旋律概形に対して改めてフーリエ変換が行われる。それによって得られたフーリエ係数を、旋律概形抽出時に使わなかった高次のフーリエ係数と足し合わせて、再度逆フーリエ変換が行われる。これにより、ユーザによる操作を反映させた音高の時系列が得られる。この音高の時系列を音符列に変換することで、新たな旋律が得られる。音符列への変換においては隠れマルコフモデル（HMM）を用いる。この HMM では各状態が各ノートナンバーに対応する。求めるべき音符列を状態遷移の系列と見なし、音楽的に正しい音符列から正規分布に従ったノイズを含みながら音高の時系列が生成されたと考えることにより、音高の時系列から Viterbi 探索により音楽的に正しい音符列を求める。状態遷移確率は、旋律の調や和声に合わせて音楽的に不適切な音高は選ばれにくくなるように設定してあるものとする。

## 2.4 旋律概形に基づいた旋律の簡約・操作の実施例

旋律概形を用いて、旋律を簡約・操作する一例を示す。ここでは、[土屋 13b] で提案した、概形レベルによる旋律操作と音符レベルによる旋律操作を組み合わせることで旋律を編集する例を取り上げる。このプロトタイプシステムでは、旋律（図 3(a)）を与えると旋律概形が前面に濃い色で、音符レベルの情報が背後に薄い色で表示される（図 3(b)）。この状態では旋律概形に対して操作できるようになっており、旋律概形を描き直すと自動的に旋律が再生成され表示される（図 3(c)）。図では 2 小節めの終わりから 3 小節めの旋律概形が山形になるように書き直した。生成された旋律を試聴して気に入ると、気に入った音符を「固定」することができる。固定するとその後の再生成の処理で上書きがされなくなる。操作対象を旋律概形から音符に切り替え、特定の操作を行うと対象の音符を表すバーの色が変化し、「固定」される（図 3(d)）。再度操作対象を旋律概形に切り替えて旋律概形を描き直しても、固定された音符は変化されなくなる（図 3(e)）。このように、旋律概形による操作と音符の固定を繰り返してほぼ好みの旋律が出来上がったら、必要に応じて音符レベルで細かな編集を行い（図 3(f)）、旋律を完成させる（図 3(g)）。

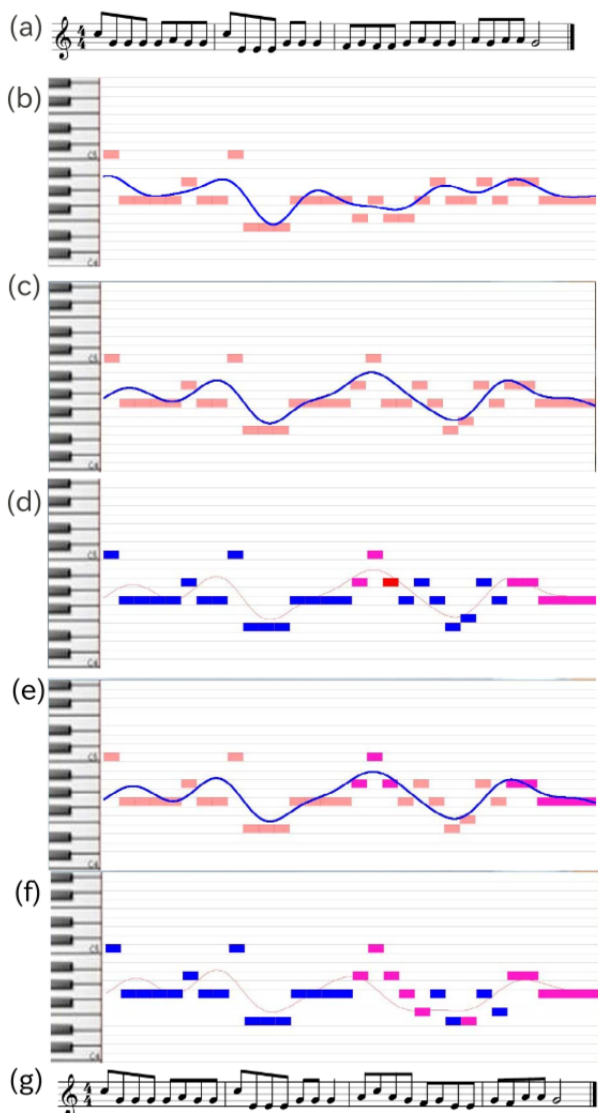


図 3: 旋律概形を用いた旋律の簡約・操作の実施例

### 3. ウェーブレット変換による旋律概形の拡張

2. で述べた手法では、音高の時系列に対してフーリエ変換を行うことで、旋律の大まかな流れから微細な動きまでを多段階的に表している。音高の時系列を  $\{x_t\}$ ,  $\{x_t\}$  に対してフーリエ変換した結果を  $\{y_n\}$  とすると、 $k$  次のフーリエ係数  $y_k$  は、 $k-1$  次までのフーリエ係数  $y_1, \dots, y_{k-1}$  で得られる旋律概形に対する元の旋律からの誤差のうち、最も大まかな成分を表していると考えることができる。フーリエ変換は対象とする信号が定常であることが前提であるので、この誤差も旋律の始めから終わりまで定常的な周期信号であることを仮定している。しかし、旋律は起承転結があるように創られるのが普通であり、旋律概形からの旋律の誤差が定常であることは考えにくい。

そこで、旋律概形とそこからの誤差を表すパラメータが木構造になるように、簡約手法を拡張する。旋律を木構造の形で表すことができれば、同様に木構造として旋律を表す GTTM など積極的に研究されている旋律の代数的操作 [平田 02] が適用できるようになり、旋律の操作手段の幅が広がると期待される。

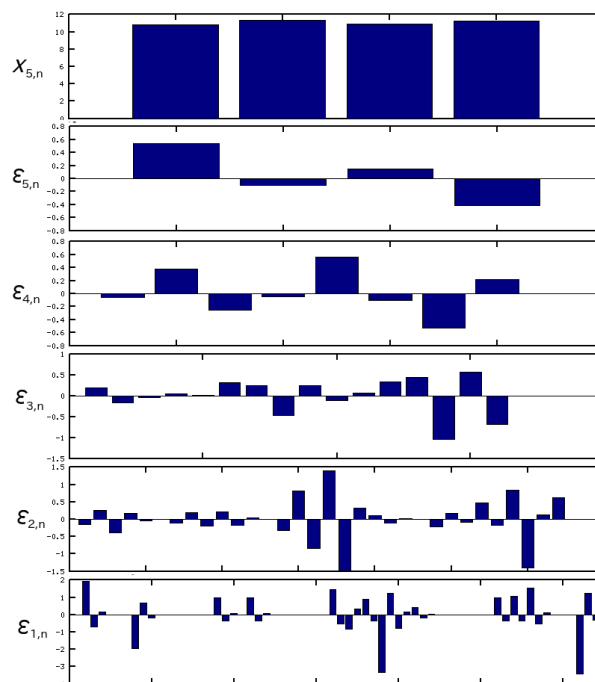


図 4: 旋律に対して離散ウェーブレット変換 (多重解像度解析) を行った例

#### 3.1 離散ウェーブレット変換を用いた旋律概形抽出手法

旋律を木構造の形で表すため、本研究では離散ウェーブレット変換を用いる。音高の時系列  $\{x_1, \dots, x_T\}$  に対して離散ウェーブレット変換を行うと、この時系列をスケーリング関数の一次結合で近似する。近似された音高の時系列を  $\{x_{1,1}, \dots, x_{1,T/2}\}$ , 近似誤差を  $\{\varepsilon_{1,1}, \dots, \varepsilon_{1,T/2}\}$  とする。次に、 $\{x_{1,1}, \dots, x_{1,T/2}\}$  に対して離散ウェーブレット変換を行って時間分解能が半分になった近似系列  $\{x_{2,1}, \dots, x_{2,T/4}\}$ , 近似誤差  $\{\varepsilon_{2,1}, \dots, \varepsilon_{2,T/4}\}$  を得る。これを再帰的に繰り返す (多重解像度解析と呼ばれる) ことで任意の分解能 (ただし系列の長さの  $1/2^k$ ) の近似系列を得ることができる。この近似系列に対する誤差が多段階的に得られ、二分木をなす)。

#### 3.2 旋律概形抽出の一例

離散ウェーブレット変換を用いて旋律に対して多重解像度解析を行い、その結果から旋律概形を抽出する処理を試行した。旋律には図 1 で用いたものと同じものを用いた。マザーウェーブレットには Daubechies 4 ウェーブレットを用いた。多重解像度解析の結果を図 4 に示す。誤差系列が下から上に行くにつれて分解能が  $1/2$  倍になっており、これらの値を二分木の形で記述できることが分かる。次に、この結果をもとに旋律概形を生成した結果を図 5 に示す。これは図 4 の  $\{x_{5,n}\}$ ,  $\{\varepsilon_{5,n}\}$ ,  $\{\varepsilon_{4,n}\}$  のみを用いて逆ウェーブレット変換を行って得たものである。フーリエ変換を用いた場合と同様に、旋律の大まかな形を抽出できていることが分かる。ただし、単に旋律概形を抽出するだけでは 2. のフーリエ変換を用いた手法で実現済みである。ウェーブレット変換を用いる利点は、旋律の旋律概形からの誤差が二分木の形で記述できることである。今後、そのメリットを活かした旋律操作手法を検討していく必要がある。

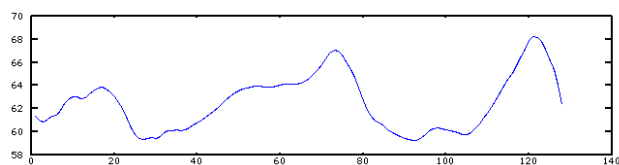


図 5: 離散ウェーブレット変換 (多重解像度解析) の結果から旋律概形を抽出した例

## 4. 議論

### 4.1 関連研究

旋律を簡約・操作する手法で最も有名なものが、先ほどから述べている GTTM である。GTTM は、旋律から二分木が生成されるという点で、3. で提案した手法と類似している。しかし、二分木の全ノードに何らかの音符が対応している点で 3. の手法と異なっている。3. の手法では、 $x_{k,n}$  は区間  $\left[\frac{n-1}{2^k}T, \frac{n}{2^k}T\right)$  全体に対するパラメータである。また、計算機上での実装が極めて容易なものも GTTM と異なっている。

音符表現に立脚せずに旋律の大まかな傾向を抽出しようとした研究に Itopul [伊藤 07] がある。Itopul は事例探索に基づいた演奏表情づけシステムで、類似事例の探索のために、旋律の大まかな概形を表すパラメータを抽出している。この手法では、旋律を 2 つの線分からなる区分的線形関数で近似する。2 つの区分の境界はロングノートや休符の存在などに基づいて決定されるが、ユーザによる修正も可能である。各区分に対して同じ処理を繰り返すことで再帰的に旋律を分割していくことができるので、GTTM や 3. の手法と同様に旋律を二分木として表すことができる。ただし、3. の手法では、系列の再帰的な分割は必ず二等分によってなされることとなっているため、3. の手法よりも Itopul の手法の方が音楽的な直感に適合した結果が得られる可能性がある。なお、Itopul の手法では、概形から旋律を再生成する手段は用意されていない。

また、Kagurame [日野 10] では、旋律間の類似度を計算するため、旋律を時間周波数表現 (ピアノロール画像に等価) に変換する。この変換後、この時間周波数表現を用いて旋律の類似度を計算している ([日野 10] では「画像比較による旋律類似性評価」と呼んでいる)。多少の音高の違いによって類似度が下がらないように、この時間周波数表現を縦方向 (周波数方向) にぼかすような処理を導入している。ただし、類似する旋律の検索に特化したものであり、旋律の認知との関連は議論されておらず、旋律の操作 (加工・編集) には利用できない。

この他にも、旋律の類似度計算のために旋律をより粗い表現 (mid-level representation と呼ばれる) に変換する試みは多数行われている (e.g. [Marolt 08])。これらの研究の多くは鼻歌検索のために用いることを想定しており、類似度の計算および類似旋律の検索には使えるが、上と同様、旋律の認知との関連は議論されておらず、旋律の操作には利用できない。

### 4.2 今後の展望

今後の展望として考えられるのは、まず [平田 02] で提案されている旋律の木構造表現に対する各種演算方式を 3. の手法に対しても導入することである。これにより、[浜中 08] とは違った旋律のモーフィングができるようになる可能性がある。

次に、認知的妥当性の検証があげられる。現状では、旋律の簡約・操作の方法を定義し、その一例を述べたのみである (2. で述べたフーリエ変換に基づく手法は、作曲支援システムとしての有用性について被験者実験を行っているが [土屋 14])。

たとえば、旋律概形に基づいて旋律の類似度尺度を定義して、人間による認知的な類似度と比較するなどして、認知的な妥当性を検証することは重要な課題である。

最後に、フレーズの導入について述べる。現状の方法では、与えられた旋律の全体を 1 本の曲線で表している。通常、楽曲はいくつかのフレーズに分割して認知する。そのため、旋律概形においても、別々のフレーズとして認知されるものは別の曲線で表すなど、フレーズ概念を導入し、フレーズに基づいた処理をすべきである。しかし、2. で述べた手法では、フーリエ変換を用いる以上 (たとえば 4 小節や 8 小節の) 旋律全体を何度もループする旋律の一部であるとみなすこととなり、フレーズごとに分割して処理を行うのは困難であった。3. で述べたウェーブレット変換を用いた手法をベースに、フレーズの導入方法について今後検討する必要がある。

## 5. おわりに

本稿では、音楽の非専門家の旋律認知を模すという観点から、音符表現によらない旋律の簡約・操作手法の重要性を指摘し、フーリエ変換およびウェーブレット変換を用いた旋律の簡約・操作手法を述べた。これまでの音楽理論はいずれも音符表現をベースとすることを暗黙の了解としており、1 個 1 個の音符をとらえずに旋律を聴くような音楽の非専門家による音楽認知の理論化はほとんど未開拓である。この研究が、音符表現によらない新たな音楽理論の構築の一助になれば幸いである。

## 参考文献

- [Lerdahl 83] Lerdahl, F. and Jackendoff, R.: *A Generative Theory of Tonal Music*, MIT Press (1983)
- [Marolt 08] Marolt, M.: A Mid-level Representation for Melody-based Retrieval in Audio Collections, *IEEE Trans. Multimedia*, Vol. 10, No. 8, pp. 1617–1625 (2008)
- [伊藤 07] 伊藤 洋介, 橋田 光代, 片寄 晴弘: 複数の生成プロセスが制御可能な演奏生成システム「Itopul」, 情処研報, 2007-MUS-73-10, pp. 45–50 (2007)
- [土屋 13a] 土屋 裕一, 北原 鉄朗: 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, 情報処理学会論文誌, Vol. 54, No. 4, pp. 1302–1307 (2013)
- [土屋 13b] 土屋 裕一, 北原 鉄朗: 旋律概形を用いた旋律編集: 概形レベルと音符レベルの編集をシームレスに行えるインターフェース, インタラクシオン 2013, 1EXB-51 (2013)
- [土屋 14] 土屋 裕一, 北原 鉄朗: 旋律概形を用いた作曲支援システム: ユーザビリティ実験の報告, 情処全大, 1R-2 (2014)
- [日野 10] 日野 達也, 鈴木 泰山, 野池 賢二, 徳永 幸生, 杉山 精: ピアノロール画像の比較による旋律類似性評価手法の検討, 情処研報, 2010-MUS-86-17 (2010)
- [浜中 08] 浜中 雅俊, 平田 圭二, 東条 敏: タイムスパン木に基づくメロディモーフィング法, 情処研報, 2008-MUS-74-19, pp. 107–112 (2008)
- [平田 02] 平田 圭二, 青柳 龍也: 音楽理論 GTTM に基づく多声音楽の表現手法と基本演算, 情処学論, Vol. 43, No. 2, pp. 277–286 (2002)