

# タイムスパン木に基づくメロディモーフィング法

浜中雅俊<sup>†1</sup> 平田圭二<sup>†2</sup> 東条敏<sup>†3</sup>

<sup>†1</sup>筑波大学, 科学技術振興機構 さきがけ

<sup>†2</sup>NTT コミュニケーション科学基礎研究所

<sup>†3</sup>北陸先端科学技術大学院大学

hamanaka@iit.tsukuba.ac.jp

本稿では, あるメロディと別のメロディの間にある複数のメロディをある尺度のもとで順序付けて生成する, メロディのモーフィング手法について述べる. 従来, 市販の楽譜エディタやシーケンサが操作できる対象は, 音符, 休符など表層的な構造に限定されていた. 音楽理論 GTTM による楽曲分析の結果得られるタイムスパン木は, 曲中の各音の構造的な重要度を階層的に表示したもので, 音楽の深層構造の分析を可能とする. 本研究では, メロディの部分簡約法を提案し, タイムスパン木に対して最小上界, 最大下界という演算を行うことでメロディのモーフィングを可能にした.

## Melody Morphing Method based on Time-span Tree

Masatoshi Hamanaka<sup>†1</sup> Keiji Hirata<sup>†2</sup> Satoshi Tojo<sup>†3</sup>

<sup>†1</sup>University of Tsukuba / PRESTO, Japan Science and Technology Agency,

<sup>†2</sup>NTT Communication Science Laboratories,

<sup>†3</sup>Japan Advanced Institute of Science and Technology

1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8573 Japan

This report describes a melody morphing method which generates an intermediate melody between a melody and another melody. Commercial music sequence software today only operates on the surface structure of music, such as the notes, and rests. The time-span tree, which acquired from the music surface by using the music theory GTTM, enables us to analyze the deeper structure. Our method makes it possible one melody morph to the other melody by proposing melody divisional reduction method with applying least upper bound and greatest lower bound to the time-span trees.

### 1. はじめに

音楽というメディアの認識や表現は曖昧なため, 音楽知識が乏しいユーザが思い通りに計算機に作曲させたり演奏させたりすることは一般に困難である. 本研究の最終的な目標は, 音楽知識の乏しいユーザを支援し, メロディ, リズム, 和声といった高次の音楽的構造を適切に操作できる音楽システムを実現することである.

音楽知識が乏しいユーザが操作可能な音楽システムを実現する上で重要なのは, 1) 音楽をいかにして操作するのか, 2) ユーザの意図をいかにして反映するのか, の2つであると我々は考えている. その際注意すべきなのは, 操作対象の抽

象度を上げると操作は容易になる反面, ユーザの意図を反映しにくくなる可能性がある点である. たとえば, 市販の楽譜エディタやシーケンサ[1]が操作できる対象は, 音符, 休符, 和音名などあいまい性の低い表層的な構造に限定されている. したがって, 音楽知識が乏しいユーザがそれらの構造を適切に扱うことは困難である. 一方, Grageband[2]などは, システムがあらかじめ多くのループ素材を用意することで, それを組み合わせるといったシンプルな操作のみで作曲を行うことができるが, 作った曲のメロディの一部を修正したいと考えた場合には, 手動で音符や休符など表層的な構造を操作する必要があるため, 音楽知識が乏しいユーザがその意図を反映することは困難である.

本研究では、「音楽をいかにして操作するのか」について、音楽理論を積極的に援用することで、音楽知識の乏しいユーザでも操作可能なシステムを構築する。その際、音楽理論としては Generative Theory of Tonal Music (GTTM) [3]を採用する。我々はこれまで、GTTM の計算機上への実装を目指し FATTA を構築してきた[4-6]。FATTA は、GTTM に基づく楽曲分析の結果得られるタイムスパン木を自動で獲得することができる。

音楽理論 GTTM の特徴は、音楽が備える多様な側面を包括的に表象しているという点である。音楽知識の乏しいユーザを支援し音楽的な構造を適切に操作するという我々の目標と照らし合わせると、音楽の持つメロディ、リズム、和声という3つの側面に関して一貫性のある操作を実現する必要があると考える。たとえば、楽曲を2つに分割するという単純な操作を考えたとき、着目する音楽的な構造によってその操作の実現は異なってくるが、装飾が付いた楽曲とそうでない楽曲に対して2つに分割する箇所は本質的に同じであることが望ましい。GTTM では、メロディの区切りを表現するグルーピング構造とリズムや韻律を表現する拍節構造をもとに、メロディや和声を本質的な部分と装飾的な部分に区別するタイムスパン木を抽出する手順が提案されている。GTTM に従えば、メロディ、リズム、和声という3つの側面に関して一貫性のある操作の実現が期待できよう。

一方、「ユーザの意図をいかにして反映するのか」に関して、本稿ではメロディを生成する場合を取り上げる。たとえば、ユーザがメロディ A の一部を修正し何らかのニュアンスを付加したい時、ユーザはそのようなニュアンスを持つメロディ B を知っているとする。この時、ユーザがシステムに対して「メロディ A にメロディ B のニュアンスを付加せよ」と指示できれば、ユーザの意図を簡易かつ的確にシステムに伝達することができよう。ここで、メロディ A にメロディ B のニュアンスを付加して少しずつメロディ B に近づけていく操作はモーフィングと呼ばれる。モーフィングには、上記の簡易かつ的確という利点の他に、システムの入力と出力の因果関係の理解が比較的容易、システム操作が簡便という利点もある。我々は、メロディモーフィング手法を提案することによって、ユーザの意図を反映したメロディ生成を実現する。

従来多くの音楽システム[7-9]は、システム固有の限定的な楽曲分析の方法を採用しており、そのような方法では音楽的な構造や意味を十分考慮することができず、ユーザの意図を反映した操

作を実現することは困難であった。これに対し、文献[10,11]では音楽理論 GTTM に基づく多声音楽の表現手法と基本演算を定義し、編曲のアルゴリズムが構築できる可能性が示された。しかし、それらの手法で実際に実現できたのは、類似度の計算[14]および不完全な模倣[15]のみであった。一方、文献[12,13]では、相対擬補元を用いた編曲手法を提案した。しかし、任意のメロディ A,B に対する相対擬補元を効率的に求める手法は明らかではなく、実現には至っていない。

本研究では、メロディの簡約が可能という GTTM のタイムスパン木の特徴に注目し、モノフォニー(和音を含まない単旋律)A と B を入力し、A と B それぞれの特徴を反映させる度合いを変化することで A と B の間にある複数個のモノフォニー C をある尺度のもとで順序付けて生成するメロディモーフィングアルゴリズムを設計した。その際、メロディ A に対して、メロディ A と B の共通部分以外の枝を簡約するメロディ部分簡約法を提案する。

## 2. Generative Theory of Tonal Music

本研究では、音楽理論 GTTM に基づく楽曲分析の結果得られるタイムスパン木を用いてメロディのモーフィングを実現する。本節では、その予備知識として GTTM の概略を述べる。

GTTM は、音楽に関して専門知識のある聴取者の直観を形式的に記述するための理論として Fred Lerdahl と Ray Jackendoff により提唱された。この理論は、グルーピング構造分析、拍節構造分析、タイムスパン簡約、プロロンゲーション簡約という4つのサブ理論から構成されており、楽譜に分析を加えることで、楽譜に内在する様々な階層構造を深層構造として顕在化させる。タイムスパン簡約は、あるメロディを簡約化することによって、そのメロディの装飾的な部分が削ぎ落とされ、本質的なメロディが抽出されるという直観を表したもので、構造的に重要な音が幹になるような2分木(タイムスパン木)を求める分析である(図1)。

タイムスパン木は、以下のようにボトムアップに構成される。まず、グルーピング・拍節構造分析の結果を用いて、楽曲を階層的なタイムスパンに分割する。次に、各タイムスパンにおいて重要な音(head と呼ぶ)がそのタイムスパンを代表する。

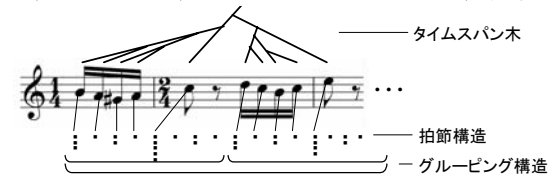


図1: グルーピング構造, 拍節構造, タイムスパン木

## 2.1 メロディの簡約

図2は、タイムスパン木を用いたメロディの簡約の例である。図のメロディ A の上にある木構造は、メロディ A をタイムスパン簡約した結果得られたタイムスパン木である。タイムスパン木のレベル B より下にある枝の音符を省略するとメロディ B になる。さらに、レベル C より下にある枝の音符を省略するとメロディ C になる。このとき、メロディ B はメロディ A と C の間のメロディであることから、メロディの簡約もメロディのモーフィングの一種と考えることができる。

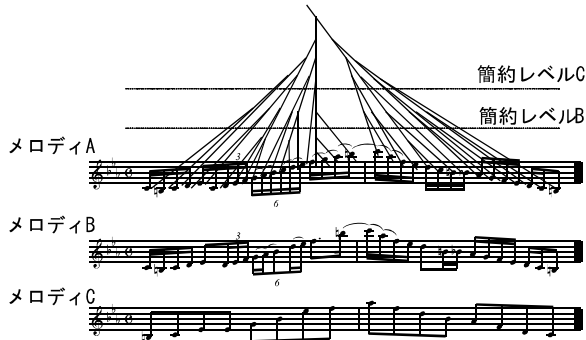


図2: メロディの簡約

## 2.2 タイムスパン木に基づく基本演算

メロディのモーフィングを実現するため、文献[10-13]で定義されている演算である、包摂関係  $\sqsubseteq$ , meet (最大下界)  $\sqcap$  と join (最小上界)  $\sqcup$  を用いる。包摂関係  $\sqsubseteq$  は、F1 を下位の構造、F2 を上位の構造 (下位の構造を含んでそれ以上の構造を持つ) としたとき、 $F1 \sqsubseteq F2$  と表記し、F2 は F1 を包摂すると言う。たとえば、図2のメロディ A, B, C のタイムスパン木 (簡約されたタイムスパン木)、 $T_A, T_B, T_C$  の包摂関係は、以下のように表せる。

$$T_C \sqsubseteq T_B \sqsubseteq T_A$$

meet (最大下界) は、 $T_A, T_B$  の共通部分のタイムスパン木  $T_A \sqcap T_B$  である。join (最小上界) は、メロディ A, B のタイムスパン木  $T_A, T_B$  が矛盾を起こさない限り統合したタイムスパン木  $T_A \sqcup T_B$  である (図3)。

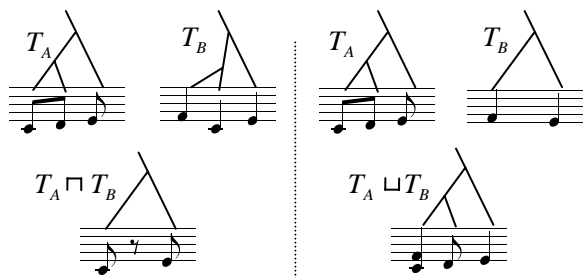


図3: meet  $\sqcap$  と join  $\sqcup$  の例

## 3. GTTM に基づくメロディモーフィング

本節では、モノフォニー (和音を含まない単旋律) A と B を入力し、A と B それぞれの特徴を反映させる度合いを変化することで、A と B の間にある複数個のメロディ C をある尺度のもとで順序付けて生成するメロディモーフィング手法を提案する。このとき、メロディ A, B, C は以下の条件を満たすものとする。以下、条件 1 と 2 は出力されるモノフォニー C に関する条件であり、条件 3 と 4 は本モーフィング手法に関する条件である。

1. A と B より A と C のほうが類似している、かつ、A と B より B と C のほうが類似している。
2. B が A と同じ場合、C も A となる。
3. A と B それぞれの特徴を反映させる度合いに応じて複数の C を出力する。
4. モノフォニー A と B を入力とし、モノフォニー C を出力する。

### 3.1 メロディモーフィングの概要

モーフィングという用語は、通常、二つの画像がある場合に、片方の画像からもう一つの画像へ滑らかに変化していくよう、その間を補うための画像を作成することをいう。2次元の顔画像のモーフィングの場合、たとえば以下のような操作で中間的な画像の生成が実現できる。

- 1) 目や鼻など2つの画像の特徴点の対応づけ (図4a)。
- 2) 各画像の形状 (位置) や色の重み付け。
- 3) 2つの画像の重ね合わせ。

一方、本研究で提案するメロディモーフィングでは、以下のような操作で中間的なメロディの生成を実現する。

- 1) 2つのメロディの共通部分の対応づけ (図4b)。
- 2) 各メロディについてメロディの部分簡約。
- 3) 両方のメロディの重ね合わせ。

以下本節では、メロディモーフィングの処理について説明する (図5)。

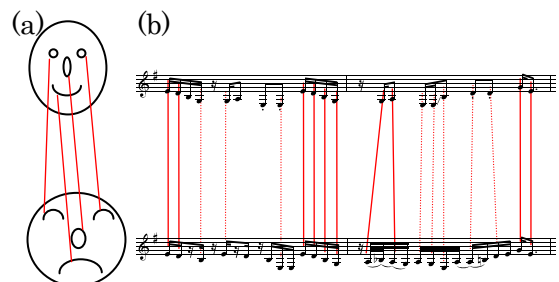


図4: 2つの画像/メロディの対応づけの例

### 3.2 メロディの共通部分の対応づけ

2つのメロディ A, B のタイムスパン木  $T_A, T_B$  を求め、その共通部分(最大下界)  $T_A \sqcap T_B$  を求める。これにより、タイムスパン木  $T_A, T_B$  は、それぞれ共通部分と非共通部分に分けることができる。

本研究では、FATTA [4]を用いて、メロディからそのタイムスパン木の自動獲得を行う。FATTA は、分析の対象をモノフォニー(和音を含まない単旋律)に限定しているため、本研究もモノフォニーを対象とする。以下、本稿ではメロディという語はモノフォニーを表すものとする。

$T_A \sqcap T_B$  は、タイムスパン木  $T_A, T_B$  をトップダウンに見て、最も大きく共通する部分を取り出すが、その際、オクターブの異なる2音(たとえば C4 と C3)を異なる音とみなす場合と、同じ音とみなす場合で結果が異なる。異なる音とみなす場合には、 $C4 \sqcap C3$  の解は空 $\perp$ となる。一方、同じ音とみなす場合には、オクターブの情報が捨象され解は C となる。オクターブ情報が未定義の場合、3.3節以降の処理が困難となることから、オクターブの異なる2音は、異なる音として扱うことにした。

### 3.3 メロディの部分簡約

3.2節で求めたメロディ A と B のタイムスパン木  $T_A, T_B$  の非共通部分には、それぞれ相手のメロディにはない特徴が表われていると考えられる。したがって、メロディのモーフィングを実現するためには、それら非共通部分の特徴をなめらかに増減させ、中間的なメロディを生成するような手法が必要となる。

本研究では 2.1 節で述べたメロディの簡約を応用し、メロディの非共通部分についてのみ簡約を実行するメロディ部分簡約法を提案する。

メロディ部分簡約法では、メロディ A のタイムスパン木  $T_A$  と、メロディ A, B のタイムスパン木の共通部分  $T_A \sqcap T_B$  から、次のアルゴリズムでメロディ  $C_m$  ( $m=1,2,\dots,n$ ) を生成する。 $C_m$  の添え字  $m$  は、 $C_m$  に含まれるが、 $T_A \sqcap T_B$  には含まれない部分(メロディの非共通部分)に表われる音符の数である。

Step1: 簡約レベルの指定

簡約レベル  $L$  をユーザが指定する。 $L$  は 1 以上、 $T_A$  に含まれるが  $T_A \sqcap T_B$  には含まれない部分(メ

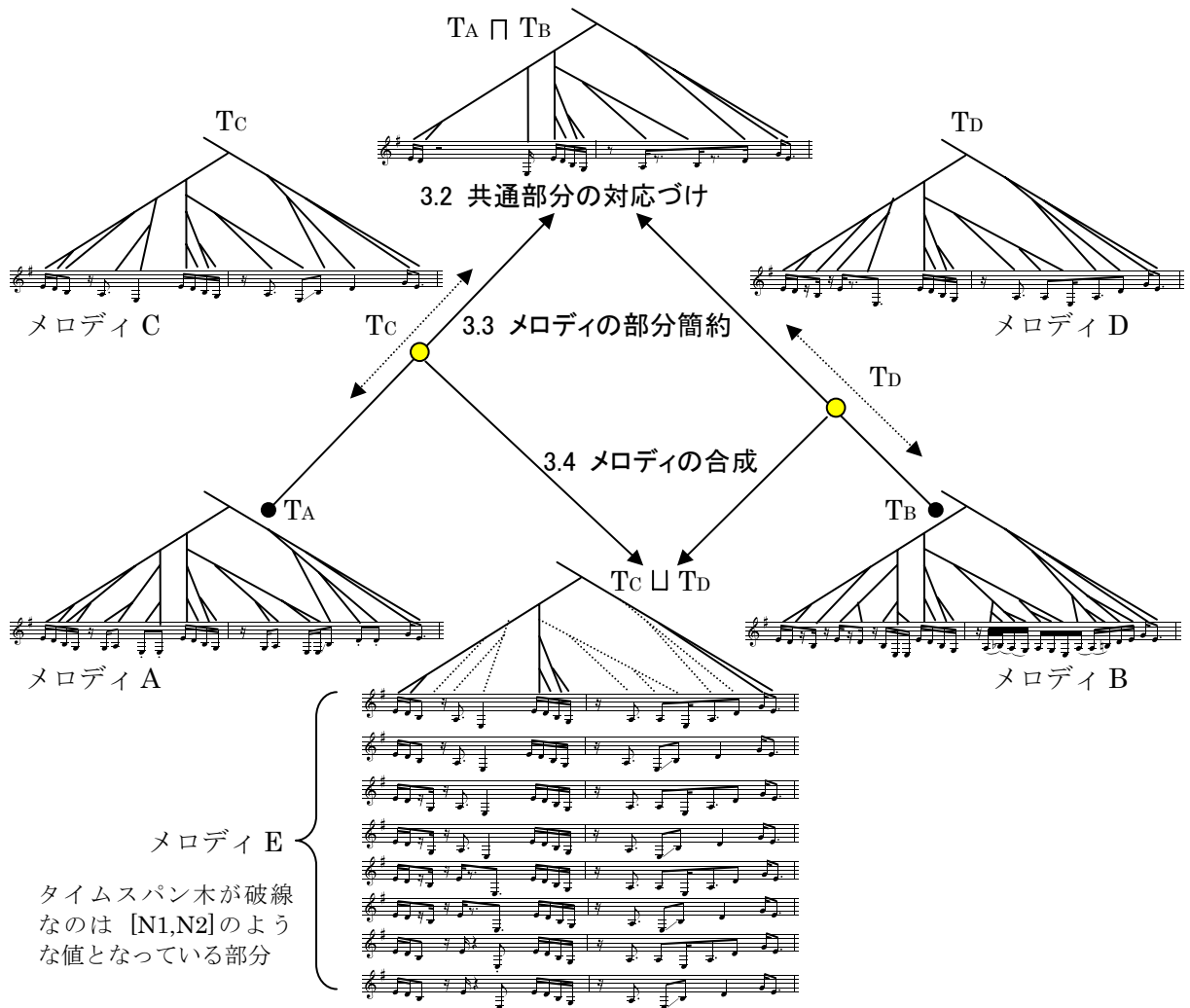


図5: メロディモーフィングの概要

ロディの非共通部分) 現れる音符の数.

**Step2:** 非共通部分の簡約

非共通部分のタイムスパンに含まれる拍点の数が最小のものを選び, その head(枝の音符)を簡約する. 拍点は GTTM の拍節構造分析により求まる. 拍点の数が最小のものが複数あった場合には, 楽曲の先頭に近いほうの head を簡約する.

**Step 3:** 繰り返し

Step2 の操作を  $L$  回繰り返す.

このアルゴリズムで生成される複数のメロディ  $C_m$  ( $m=1,2,\dots,n$ ) のタイムスパン木  $T_{C_m}$  は, 複数の  $T_{C_m}$  同士の間で包摂関係が成立する. したがって, 次式が成立する.

$$T_A \sqsupset T_B \sqsupset T_{C_1} \sqsupset T_{C_2} \sqsupset \dots \sqsupset T_{C_{n-1}} \sqsupset T_{C_n} \sqsupset T_A$$

図 5 の場合では,  $T_A$  には,  $T_A \sqsupset T_B$  にはない音符が 9 個含まれているため,  $n$  の値は 8 となり,  $T_A$  と  $T_A \sqsupset T_B$  の間のメロディが 8 種類得られることになる.

上記のようにして求めたメロディ C は, メロディ B にはないメロディ A のみが持つ特徴の一部を減衰させたものと考えることができる. 同様に,  $T_B$  と  $T_A \sqsupset T_B$  から下記を満たすメロディ D を生成する.

$$T_A \sqsupset T_B \sqsupset T_D \sqsupset T_B$$

**3.4 メロディの合成**

メロディ A と B のそれぞれのタイムスパン木の一部を簡約したメロディ C とメロディ D を統合 (最小上界) し, 合成したメロディ E を生成する.

2 つのタイムスパン木  $T_C$  と  $T_D$  の統合を行う際, 図 3a に示したような join (最小上界) を素朴に実装するだけでは  $T_C, T_D$  の両方がモノフォニーであっても,  $T_C \sqcup T_D$  がモノフォニーとなるとは限らない. つまり,  $T_C$  と  $T_D$  のタイムスパン木を重ね合わせる際, タイムスパン木の枝は重なるが (時間構造は一致するが), 音高が異なるような場合には, 解に和音が含まれることになってしまう. すなわち 3 節冒頭の条件 4 に反する.

そこで, 本研究では異なる 2 音を N1, N2 としたとき [N1, N2] のような「N1 または N2」を意味する特殊な値を導入し,  $N1 \sqcup N2$  の解を, [N1, N2] とする. すると  $T_C \sqcup T_D$  の解には, [N1, N2] のような値が複数含まれることになる. そして, それらすべての組み合わせ, すなわち複数のモノフォニーを  $T_C \sqcup T_D$  の解とする.

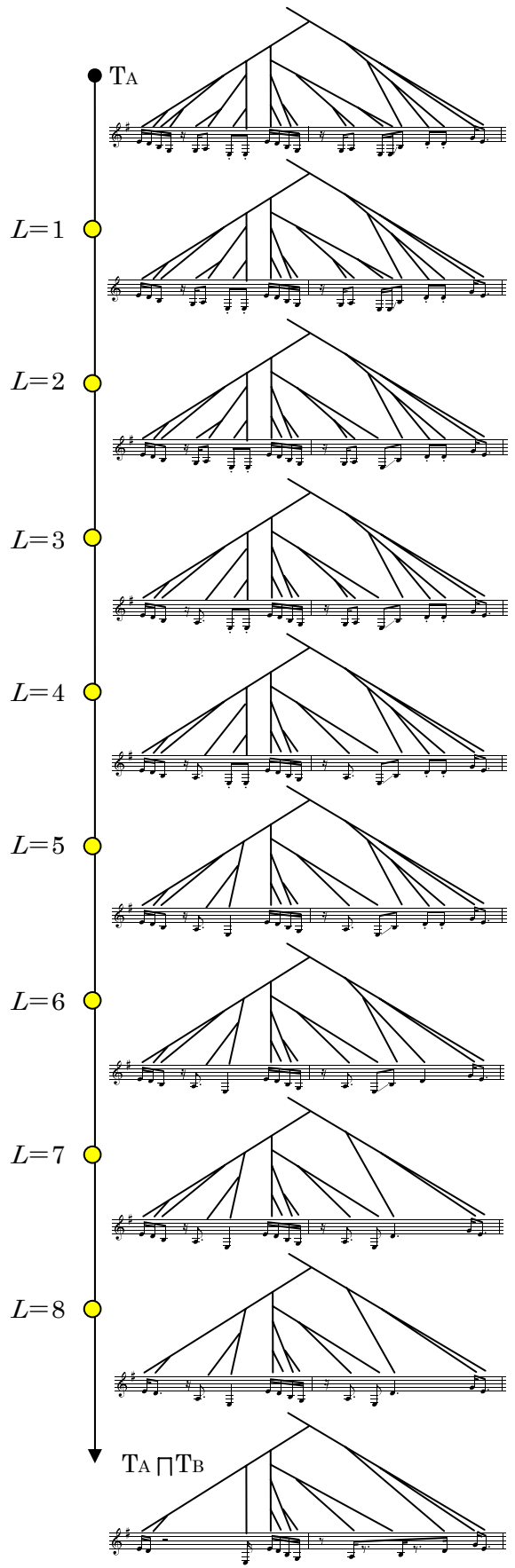


図 6: メロディの部分簡約

## 5. 議論

本稿で提案したメロディモーフィングの手法では、ユーザは2つのパラメータ(2つのメロディそれぞれの特徴を反映させる度合い)を操作する必要がある。また、メロディの合成の際にも複数のメロディを出力するため、ユーザはそれらを確認しなくてはならない。文献[16]で提案する電気モーフは、ユーザの操作性を上げるため、操作するパラメータを1つとするとともに、出力されるメロディも1つになるよう工夫したものである。以下の2点についても検討を行っていく。

### 5.1 モーフィング用データベースについて

本稿で提案したモーフィング手法では、2つのメロディAとBのmeet(最大下界)を算出した、ある程度の大きさを持つ楽曲同士のmeetを計算すると、一般には、かなり類似した楽曲同士でない限り、meetの結果は空集合あるいはそれに近い値になってしまう、モーフィングすることが困難になる。

今後、あるメロディAを与えたときに、Aとモーフィング可能なメロディを自動で検索し提示するようなモーフィング用データベースの構築を検討していく。

### 5.2 主観的類似度について

本稿で提案したモーフィング手法では、メロディAとBのモーフィングによって生成されたメロディEに、AとBのタイムスパン木のmeetが含まれている。したがって、文献[14]で定義されているような類似度を用いれば、EがAとBの間に位置することが確認できる。

今後、主観的にもAとBの間に感じられるか、実験によって確認したいと思っている。その際問題となるのは、モーフィングの結果、区別がすることが困難なほど非常に似通ったメロディが複数出力され、被験者への負荷が非常に高くなる可能性があることである。被験者実験の方法についても検討していく。

## 6. まとめと今後の課題

本稿では、あるメロディと別のメロディの間にある複数個のメロディをある尺度のもとで順序付けて生成するメロディモーフィング手法について述べた。そして、2つのメロディそれぞれの特徴を反映させる度合いを変化させることで、間のメロディが複数生成されることを確認した。

今回は入力をモノフォニーに限定して、システムの構築を行ったが、今後、フォモフォニー(和音を含む単旋律)や、ポリフォニー(複旋律)の場合にも適用可能であるか検討していく予定である。

## 参考文献

- [1] 佐藤 天平, コンピュータ・ミュージックスーパー・ビギナーズマニュアル, ソフトバンククリエイティブ, 1997.
- [2] <http://www.apple.com/jp/ilife/garageband/>
- [3] F. Lerdahl, and R. Jackendoff. A Generative Theory of Tonal Music. Cambridge, Massachusetts: MIT Press, 1983.
- [4] Matstoshi Hamanaka, Keiji Hirata, and Satoshi Tojo. Implementing 'A Generative Theory of Tonal Music'. Journal of New Music Research, 35:4, 249-277, 2006.
- [5] Matstoshi Hamanaka, Keiji Hirata, and Satoshi Tojo. FATTA: Full Automatic Time-span Tree Analyzer, Proceedings of the 2007 International Computer Music conference, Vol. 1, pp. 153-156, 2007.
- [6] 浜中雅俊, 平田圭二, 東条敏: 音楽理論 GTTM に基づくグルーピング構造獲得システム, 情報処理学会論文誌, Vol. 48, No. 1, pp. 284-299, 2007.
- [7] Mira Balaban. The Music Structures Approach to Knowledge Representation for Music Processing, Computer Music Journal, Vol. 30, No.2, pp. 96-111, 1996.
- [8] David Cope. Experiments in Musical Intelligence, A-R Editions, Inc. 1996.
- [9] Roger Dannenberg. Machine Tongues XIX: Nyquist, a Language for Composition and Sound Synthesis, Computer Music Journal, Vol. 21, No. 3, pp. 50-60, 1997.
- [10] 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, 情報処理学会論文誌 Vol.43, No.2, 2002.
- [11] 平田圭二, 平賀 譲: GTTM に基づく音楽表現手法再考, 情報処理学会研究報告 2002-MUS-45, pp.1-7, 2002.
- [12] 平田圭二, 東条 敏: 相対擬補元を用いたメディアデザイン操作の形式化について, 第19回 人工知能学会 全国大会, 2B3-08, 2005.
- [13] 平田圭二, 東条 敏: 楽曲構造束とその上の演算系, 第20回 人工知能学会 全国大会, 1D2-4, 2006.
- [14] K.Hirata, and S. Matsuda. "Interactive Music Summarization based on Generative Theory of Tonal Music." Journal of New Music Research, 32:2, 165-177, 2003.
- [15] 平田圭二, 青柳龍也: バービーブン: 音符レベルでユーザ意図を把握して編曲を行う事例ベースシステム, 情報処理学会研究報告 2000-MUS-37, pp.17-23, 2000.
- [16] 浜中雅俊: 電気モーフ, 情報処理学会研究報告 2008-MUS-74, 2008.