

# deepGTTM-I:ディープラーニングに基づく局所的グルーピング境界分析器

## deepGTTM-I: Local Grouping Boundary Analyzer based on Deep Learning

浜中 雅俊<sup>\*1</sup>  
Masatoshi Hamanaka

<sup>\*1</sup> 京都大学  
Kyoto University

This paper describes a method that enables us to detect the local boundaries of a generative theory of tonal music (GTTM). Although systems that enable us to automatically acquire local boundaries have been proposed such as a full automatic time-span tree analyzer (FATTA) or  $\sigma$ GTTM, musicologists have to correct the boundaries because of numerous errors. In light of this, we propose a novel method called deepGTTM-I for detecting the local boundaries of GTTM by using a deep learning technique. The experimental results demonstrated that deepGTTM-I outperformed the previous analyzers for GTTM in an F-measure of detecting local boundaries.

### 1. はじめに

本稿では、ディープラーニングに基づき音楽理論 Generative Theory of Tonal Music (GTTM) [Lerdahl 1985]における局所的グルーピング境界を自動で検出する手法について述べる。GTTM は、認知過程を含めた音楽の構造を抽出する音楽理論で、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の4つのサブ理論からなる。

局所的グルーピング境界の検出は、GTTM における最初の分析で、その分析結果は GTTM のすべての分析に影響を与えるため、精度高く分析することが極めて重要である。GTTM の局所的グルーピング境界の検出を可能としていた従来の分析器には ATTA [Hamanaka 他 2006], FATTA [Hamanaka 他 2007],  $\sigma$ GTTM [Miura 他 2009],  $\sigma$ GTTMII [Kanamori 他 2014]の4つがあったが、いずれも分析精度が低く、分析結果を音楽家が手動で修正する必要があった。

そこで本研究では、GTTM における局所的グルーピング境界の検出に、ディープラーニングを用いた手法 deepGTTM-I を構築し、検出精度の向上を目指す。ディープラーニングは多層化したニューラルネットワークで、大量のデータを用いてプレトレーニングと呼ばれる教師なし学習を繰り返し行うことで、ファインチューニングと呼ばれるラベル付きデータを用いた教師つき学習の効率を上げることを可能とする。本研究の最終的なゴールは、音楽家による分析結果をディープラーニングで学習させることで、人間と同じ分析結果を出力するネットワークを構築することである。

ディープラーニングを用いて音楽理論 GTTM に基づく分析器を構築するためには、以下の3つを考慮する必要がある。

#### ● 多タスクの識別問題であること

特徴ベクトルからその時のラベルを推定するのが、一般的な識別問題である。一方、局所的グルーピング境界は曲中の全ての音符間に発生しうる。そこで本研究では、ある音符と次の音符の間が局所的境界であるかどうかを判定する問題を1つのタスクとらえる。すると、楽曲から局所的グルーピング境界を

求める問題は、マルチタスクの識別問題と捉えることができる。3.3 節では、マルチタスクのディープラーニングを用いたモデル化について述べる。

#### ● 大量の学習データが必要となること

多層からなるニューラルネットワークを学習するためには大量のデータが必要である。ファインチューニングに用いる教師付データは、GTTM データベースに収められている300曲のラベル付きの曲をデータとして用いた[Hamanaka 他 2014]。一方、プレトレーニングに用いる教師なしデータは300曲では十分ではなく、もっと多くの曲が必要である。教師なしデータはラベルが不要であるため、GTTM データベースに収められている曲に加え、makemusic 社の MusicXML のページ [makemusic Inc. 2016]で紹介されているサイトからダウンロードした15,000曲のデータを用いることにした。?節では、学習セットの作成について述べる。

#### ● GTTM のルールが適用されていること

GTTM は複数のルールによって構成されており、局所的グルーピング境界の検出では、ルールが多く適用されている箇所がグルーピング境界になりやすい。GTTM データベースに収められている楽曲では、音楽家による分析結果として、局所的グルーピング境界の位置だけでなく、ルールの適用位置も収められている。これらルールの適用位置は、局所的グルーピング境界を検出するための大きな手がかりとなる。?節では、マルチタスクディープラーニングの学習に、適用されたルールを利用することについて述べる。

マルチタスクディープラーニングを用いて、局所的グルーピング境界の学習を行った結果、deepGTTM-I が従来の分析器に比べて高い性能であることを確認した。

### 2. 関連研究

これまで様々な音楽理論[Cooper and Meyer 1960, Narmour 1990, Temperley 2001] が提案されてきたが、GTTM は他の音楽理論と比べて比較的厳密なルールで記述されており、音楽知識を形式化する上で最も有望であると我々は考え、これまで10年以上にわたり分析システム、応用システムの両面から研究を進めてきた(図1)[Hamanaka 他 2016]。

連絡先: masatoshi@kuhp.kyoto-u.ac.jp

<http://gttm.jp/hamanaka/>

## 2.1 分析システム

図 1 のタイムラインより上は、我々がこれまで構築してきた分析システム・手法である。我々は、2004 年にグルーピング構造分析器および拍節構造分析器を構築し、それらを統合したタイムスパン木分析器 ATTA(Automatic Time-span Tree Analyzer) を 2005 年に構築した。ATTA は、46 個の調節可能なパラメータを持ち、それらを適切に調整すればグルーピング構造では 8 割弱、拍節構造では 9 割強の正解率、タイムスパン木では、6 割の正解率となった。ATTA は GTTM のルールの優先順序を適切に適用すれば、正解率の高いグルーピング構造や拍節構造を獲得できることを示した点で意義があるが、その一方で、パラメータを調整するためには、音楽的な知識が必要で、音楽家でなければ扱うことが困難であった。

2007 年に構築した FATTA は、タイムスパン木の安定性に関するルールに基づきタイムスパン木の安定性を定義し、その安定性が高くなるように ATTA のパラメータを自動調整するシステムであった。拍節構造分析の正解率は 9 割程度であったものの、グルーピング構造分析およびタイムスパン簡約の正解率は 5 割弱で、音楽家の手作業による修正を行わなければ分析結果を利用することは難しかった。

誤りのない完全な分析結果を出力する分析器を構築することが困難であったことから、2009 年に構築した Interactive GTTM analyzer は分析器による分析と手動による編集をシームレスに行うことを可能にしたものであった [Hamanaka 他 2009]。Interactive GTTM analyzer は現在でも分析データの蓄積のために使用されており無料でダウンロードすることができる [Hamanaka 2016]。

2008 年に構築した  $\sigma$ GTTM では、決定木を用いた統計的学習により自動で局所的グルーピング構造の検出を可能にしていた。FATTA より性能が高かったものの、パラメータ調整後の ATTA よりは低い性能であった。一方、 $\sigma$ GTTMII では、複数の学習済みの決定木を用意し、それらを手動で切り替えることで、

ATTA を超える性能を実現していた。 $\sigma$ GTTMII は、局所的グルーピング境界の検出では現在でも最も性能が高いが、適切な決定木の選択には、音楽知識が必要で、音楽家でなければ難しい作業であった。

一方、2015 年に構築した  $\sigma$ GTTMIII [Hamanaka 他 2015] では、音楽家による楽曲分析結果を確率文脈自由文法 (Probabilistic context-free grammar, PCFG) [Charniak 他 1996] に基づき統計的に学習することでタイムスパン木の自動分析を可能としていた。 $\sigma$ GTTMIII では、GTTM データベースに収められている 300 曲を教師データとして PCFG の生成規則と生成確率を学習し、タイムスパン木分析器としては最も高い性能を示していた。また、同じく PCFG に基づくタイムスパン木分析器である pGTTM [Nakamura 他 2016] では、教師なし学習を可能としていた。これら、 $\sigma$ GTTMIII および pGTTM の長は、タイムスパン木の先端付近では拍節の影響が強く、根の近くでは和声の影響が強くなるなど、階層によるコンテキストの違いを学習可能としている点である。

上記述べてきた分析システム・手法では、楽曲中で並列的なメロディを見つけたり、シンメトリーな構造を見つけるような基本的な処理が不足していたり、あるいはそれらの処理があった場合でも、それらの処理と他の処理とがうまく連携されていないという問題があった。本研究では、それらの一連の処理をディープラーニングで実現することを目指す。

## 2.2 応用システム

図 1 のタイムラインより下は、我々がこれまで構築してきた応用システムである。GTTM の分析の結果求まるタイムスパン木は、楽曲の要約 [Hirata and Matsuda 2002] や表情付け [Hirata and Hiraga 2003]、生成 [Hirata 他 2004]、メロディモーフィング [Hamanaka 他 2008]、作曲支援 [西田他 2008] に用いることができる。これらのシステムを有効に利用するためには、GTTM に基づく楽曲分析を精度高く自動で行うシステムの構築が必要である。

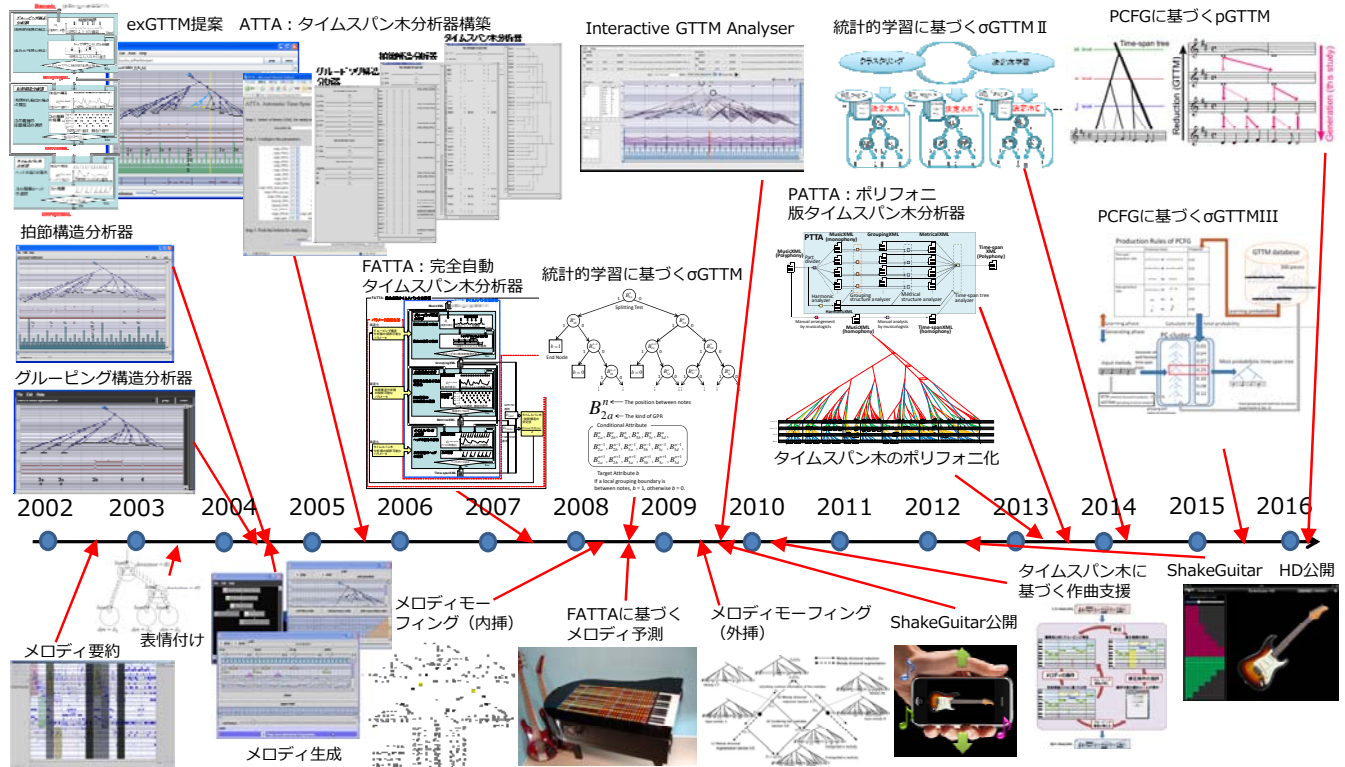


図 1: 音楽理論 GTTM に基づく分析システム・応用システム

### 3. deepGTTM-I: ディープラーニングに基づく局所的グルーピング境界の検出

GTTM のグルーピング構造分析は、長いメロディを歌うときにどこで息継ぎすべきかを見つけるような分析で、まず局所的なグルーピング境界を検出した後に、階層的な構造を分析する。本節では、楽譜情報を入力すると局所的グルーピング境界を出力するようなネットワークをディープラーニングで構成する方法について述べる。

#### 3.1 局所的グルーピング境界の分析における問題

GTTM では、局所的グルーピング境界を求めめるため、グルーピング選好ルールを定義している。しかし、グルーピング選好ルールのうち GPR4 (intensification), GPR5 (symmetry), GPR6 (parallelism)については、厳密な定義がなされておらず、分析の自動化が困難だった。また、選好ルールを適用する際には、複数のルール間での優先度が決まっていないため競合が起きることがある。図 2 は音程における跳躍点をグループの境界とするルール GPR3a とメロディの繰り返し時点を境界とするルール GPR6 が競合する場合である。この両方を境界とすると単音(第 4 音や 8 音)が単音のグループになってしまい、これは GPR1 で避けるべきとされている。したがって、GPR3a と GPR6 の一方しか適用されず、両者は競合関係にある。



図 2: ルールの競合

### 3.2 解決法: ディープラーニングを用いた学習

本研究では、ディープラーニングを用いて以下の 2 つを学習する。

#### 各ルールの適用/不適応の学習

楽譜およびグルーピング選好ルールの適用箇所を学習したネットワークをディープラーニングで構成し、新たな楽譜が入力されると、各ルールが適用されるか不適用であるかを出力できるようにする。

従来は、ルールごとの個別の処理をプログラマがコーディングしていたため、プログラマが音楽理論を正しく(深く)理解できていない場合には、すべての事例に対して正しい分析になることを保証することが難しかった。

これに対して、本研究では、ディープラーニングを用いて学習し、学習済みのネットワークを用いて統一的に処理を行うため、分析の正しさは学習データおよび学習するネットワークに依存する。

#### ルールの優先順序の学習

ルールの優先順序は、分析する楽曲のコンテキストに依存するため、 $\sigma$  GTTM や  $\sigma$  GTTMII のように適用されているルールだけを見てその優先順序を決定することは困難である。

これに対して本研究では、楽譜を入力として、どのルールを優先するかを、ネットワークの重みの形で統一的に学習する。

### 3.3 Deep Belief Network のマルチタスク学習

ディープラーニングの一手法である Deep Belief Networks (DBNs)を学習用のネットワークとして用いる。

入力は、楽譜の各音符のオンセットタイム、オフセットタイム、ピッチ、ベロシティである。出力はマルチタスク学習となっており、

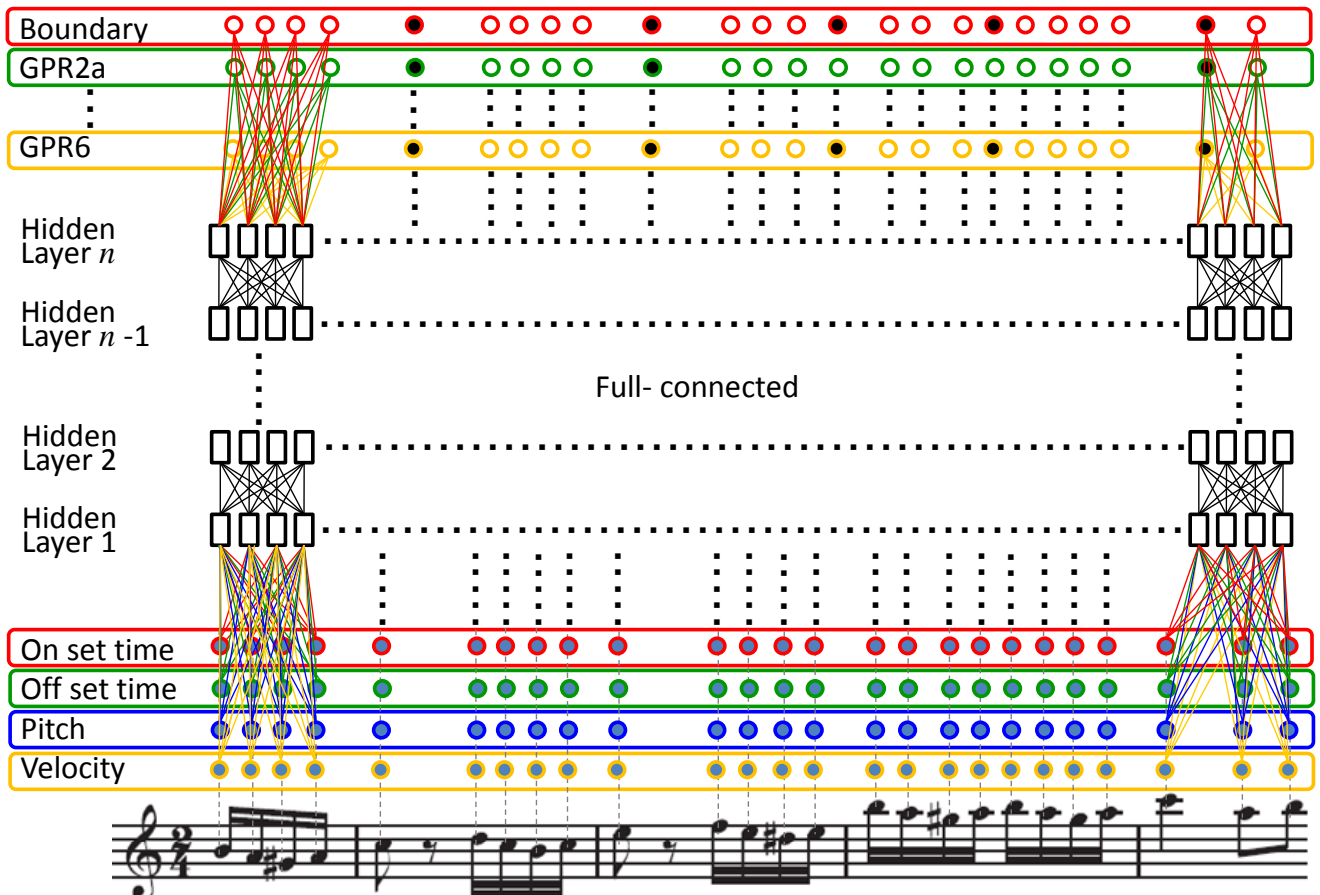


図 3: 局所的グルーピング境界を検出する Deep Belief Network.



GPR2a,2b,3a,3b,3c,4,5,6,7 の 10 種類のグルーピング選好ルールおよび、局所的境界で、1(ルール適用あるいは境界)または 0(ルール不適用あるいは境界でない)の 2 値である。Hidden Layer 1 から  $n$  までを共有しており、最終段はロジスティック回帰で構成した。

学習は、以下の 3 段階で行う。すべての学習において、曲順は 1 Epoch ごとにシャッフルした。

#### 教師なし学習

Make music 社の MusicXML のページ[Make music Inc. 2016]で紹介されているサイトからダウンロードした 15,000 曲のデータを用いて教師なし学習を行った。これは、ネットワークに「音楽」というものを教え込むことに相当する。教師なし学習は、100 Epochs 繰り返した。

#### ルールの教師付き学習

ルールの教師付き学習は二段階で行う。まず、教師なし学習で用いたデータに ATTA で、GPR2a,2b,3a,3b,3c,3d の正解ラベルを与え、教師付きマルチタスク学習を 100 Epochs 行った。次に、GTTM データベースに収められているグルーピング選好ルールが音楽家によりラベル付けされているデータを用いて、教師付きマルチタスク学習を 1000 Epoch 行った。

#### 境界の教師付き学習

GTTM データベースの局所的グルーピング境界のラベルを用いて教師付き学習を行った。200 曲の学習データを 1000 Epoch 繰り返した。

## 4. 実験結果

GTTM に収められている 300 曲のデータのうち、200 曲を教師付き学習に用い、残る 100 曲で評価を行った。表 1 は 3000 ユニットの 11 層のネットワークを学習し評価した結果である。

表 1: 評価結果

	再現率 P	適合率 R	F値
ATTA	0.737	0.441	0.690
$\sigma$ GTTM	0.467	0.736	0.571
$\sigma$ GTTM II	0.684	0.916	0.783
deepGTTM-I	0.784	0.814	0.799

ATTA および  $\sigma$  GTTMIII は性能が高くなるように、パラメータの調整や決定木の調整を行っている。一方、 $\sigma$  GTTM および deepGTTM-I は手動で調節が必要な部分がない。すなわち、deepGTTM-I は安定して高い性能が示せることが確認できた。

## 5. おわりに

本稿では、ディープラーニングに基づく楽曲構造分析器 deepGTTM-I を構築した。評価実験の結果、従来の分析器を上回る性能であることを確認した。今後、拍節構造分析やタイムスパン木など、GTTM の分析過程全体をディープラーニングを用いて実装してゆく。

## 参考文献

[Lerdahl 1985] Lerdahl, F., Jackendoff, R.: *A Generative Theory of Tonal Music*. MIT Press, 1985.  
 [Hamanaka 他 2006] Hamanaka, M., Hirata, K., Tojo, S.: Implementing 'a generative theory of tonal music', *Journal of New Music Research*, 35(4), 249–277, 2006.  
 [Hamanaka 他 2007] Hamanaka, M., Hirata, K., Tojo, S.: Fatta: Full automatic time-span tree analyzer, *In: Proceedings of the 2007 International Computer Music Conference (ICMC2007)*, pp. 153–156, 2007.  
 [Miura 他 2009] Miura, Y., Hamanaka, M., Hirata, K., Tojo, S.: Decision tree to detect gttm group boundaries, *In:*

*Proceedings of the 2009 International Computer Music Conference (ICMC2009)*, pp. 125–128, 2009.  
 [Kanamori 他 2014] Kanamori, K., and Hamanaka, M.: Method to Detect GTTM Local Grouping Boundaries based on Clustering and Statistical Learning, *In: Proceedings of the 2014 International Computer Music Conference (ICMC2014)*, pp. 125–128, 2014.  
 [Hamanaka 他 2014] Hamanaka, M., Hirata, K., Tojo, S.: Musical Structural Analysis Database Based on GTTM, *In: Proceeding of the 2014 International Society for Music Information Retrieval Conference (ISMIR2014)*, pp.325-330, 2014.  
 [Make music Inc. 2016] Make music Inc. : Music in MusicXML Format, url: <http://www.musicxml.com/music-in-musicxml/>, 参照 2016-2-28.  
 [Cooper and Meyer 1960] Cooper, G. and Meyer, L. B. *The Rhythmic Structure of Music*. The University of Chicago Press, 1960.  
 [Narmour 1990] Narmour, E. *The Analysis and Cognition of Basic Melodic Structure*. The University of Chicago Press, 1990.  
 [Temperley 2001] Temperley, D. *The Cognition of Basic Musical Structures*. MIT press, Cambridge, 2001.  
 [Hamanaka 他 2016] Masatoshi Hamanaka, Keiji Hirata, Satoshi Tojo: Implementing Methods for Analysing Music Based on Lerdahl and Jackendoff's Generative Theory of Tonal Music, *Computational Music Analysis* (pp. 221-249), Springer, 2016.  
 [Hamanaka 他 2009] Hamanaka, M., Hirata, K., Tojo, S.: Interactive GTTM Analyzer, *In: Proceedings of the 10th International Conference on Music Information Retrieval Conference (ISMIR2009)*, pp.291-296, 2009.  
 [Hamanaka 2016] Hamanaka, M.: Interctive GTTM Analyzer / GTTM Database, url <http://gttm.jp>, 参照 2016-2-28.  
 [Hamanaka 他 2015] Hamanaka, M., Hirata, K., Tojo, S.:  $\sigma$  GTTM III: Learning-based Time-span Tree Generator Based on PCFG, *In: Proceedings of the 11th International Symposium on Computer Music Multidisciplinary Research (CMMR 2015)*, pp.303-317, 2015.  
 [Nakamura 他 2016] Nakamura E., Hamanaka M., Hirata K., and Yoshii K.: Tree-Structured Probabilistic Model of Monophonic Written Music Based on the Generative Theory of Tonal Music, *In: proceedings of 41st IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2016)*, 2016.  
 [Charniak 他 1996] Charniak, E.: Tree-bank grammars. *In: Proceeding of Association for the Advancement of Artificial Intelligence (AAAI-96)*, pp. 1032–1036, 1996.  
 [Hirata and Matsuda 2002] Hirata K. and Matsuda S.: Interactive Music Summarization based on GTTM, *In: Proceeding of the 2002 International Society for Music Information Retrieval Conference (ISMIR2002)*, pp.86-93 , 2002.  
 [Hirata and Hiraga 2003] Hirata, K. and Hiraga R.: Ha-Hi-Hun plays Chopin's Etude, *In Working Notes of IJCAI-03 Workshop on methods for automatic music performance and their applications in a public rendering contest*, 2003.  
 [Hirata 他 2004] Hirata, K., Matsuda, S., Kaji K. and Nagao K.: Annotated Music for Retrieval, Reproduction, and Sharing, *In: Proceedings of the 2004 International Computer Music Conference (ICMC2004)*, pp.584-587, 2004.  
 [Hamanaka 他 2008] Hamanaka, M., Hirata, K., Tojo, S.: Melody morphing method based on gttm. *In: Proceedings of the 2008 International Computer Music Conference (ICMC2008)*, pp. 155–158, 2008.  
 [西田他 2008] 西田 智, 浜中 雅俊, 平田 圭二, 東条 敏: 類似した楽曲構造を持った旋律のインタラクティブな生成方式, 情報処理学会音楽情報科学研究会, 2010-MUS-84(4), pp. 1-6, 2010.