

# GTTMに基づくグルーピング構造及び拍節構造の自動分析の試み

井田 健太郎  
北陸先端科学技術  
大学院大学  
情報科学研究科

平田 圭二  
NTT コミュニケーション科学  
基礎研究所

東条 敏  
北陸先端科学技術  
大学院大学  
情報科学研究科

## 要旨

本研究報告では Generative Theory of Tonal Music (GTTM) という音楽理論のグルーピング構造と拍節構造の自動分析を計算機上に実装する際の問題点とその解決法について述べる。楽曲に対する処理の合理化を考えると、自然言語処理のように前処理として構造分析することが考えられる。GTTMは楽曲分析の有望な理論の1つであるが、幾つかの点がその実装困難にしている。それは、ホモフォニーへの制限、用語定義の曖昧性、規則適用の非一貫性である。これらを解決するために、ボイスリーディング分析、第一声部進行分析を導入し、グルーピングの認識率改善について議論する。

## The attempt of the automatic analysis of the grouping structure and metrical structure based on GTTM.

Kentarou Ida  
Japan Advanced Inst. of  
Science and Technology

Keiji Hirata  
NTT Communication  
Science Laboratories

Satoshi Tojo  
Japan Advanced Inst. of  
Science and Technology

## Abstract

In this paper, we describe the problems of the automatic analysis of grouping structures and metrical structures in Generative Theory of Tonal Music (GTTM), and then propose methods to solve them. For the efficient processing of a musical piece, we need to analyze its structure in advance, similarly to natural language processing. Although GTTM offers such framework, the theory has problems in several respects that make the implementation on computers difficult; that is, the restriction to homophonic music, the ambiguity of terminological definitions, and the inconsistency of rule application. We propose to apply the voice leading and the first voicepart progress to the theory, and discuss how we can improve the recognition rates of grouping.

## 1 はじめに

現在、計算機上で音楽を扱う様々な研究が行われている。自動伴奏や楽曲の検索などがその例である。しかし、音楽的に質の高い結果を得るのは大変困難である。その主な原因としては、楽曲のある一部分の表層的な情報だけから出力を計算していることが考えられる。よって、楽譜に明示されていない構文的な情報や、音符間の関係を考慮に入れて処理を進めることが必須となる。我々は楽曲に対する様々な処理は、楽曲を分析し、その結果得られた構造に基づいて処理を進めるのが最も合理的であると考えた [2]。このような楽曲

の構造を分析する理論の1つに、Fred LerdahlとRay Jackendoffによって提唱された Generative Theory of Tonal Music(GTTM)[3]がある。

しかし現在、この理論はコンピュータ上のプログラム化することを前提としておらず、用語や規則が厳密に定義されていない。従って、コンピュータ上への実装は困難であると考えられる。

本研究報告では GTTM 自動化への第一歩として、グルーピング構造と拍節構造の解析における問題点を明らかにし、その解決法について述べ、その結果を検討する。

## 2 Generative Theory of Tonal Music

GTTMは以下の2つの基本構造分析と2つの簡約理論から構成されている。

- 1 グルーピング構造分析
- 2 拍節構造分析
- 3 タイムスパン簡約理論
- 4 延長的簡約理論

グルーピング構造は、楽曲をより小さいまとまり(グループ)に分ける構造である。その個々のグループは、あるまとまりを持って聞こえる単位である。例えば歌を歌う時、一息で歌う範囲の様なものである[5]。さらにそのグループを重ね、より大きなグループを作っていく、最終的には楽曲全体が1つのグループにまとまる。図1のスラーのような括弧はグルーピング構造の一例である。この1つの括弧が1つのグループを表す。

拍節構造は、指揮者が指揮棒を振ったり、聞き手が足でリズムをとることを表現している[5]。この構造は楽曲中の最も短い音符から、より長い音符のレベルすべてに対し、強拍と弱拍を同定する分析である。図1の角ばった括弧は拍節構造の一例である。1つの括弧が1つの拍節を表す。

タイムスパン簡約は、あるメロディを抽象化、簡約化し、より短いメロディを得るという考えを表すものである[5]。つまりある楽曲は、その楽曲の中で最も重要と思われる音からの派生で構成されているという考えである。グルーピング構造分析と拍節構造分析の結果からタイムスパン簡約を求める。2つの連続する音符のうち曲のイメージを構成するために、より重要と思われる音符を選択する(簡約する)。これを繰り返してい、最後に残った1音が、タイムスパン的にその楽曲中で最も重要な音符と見なされる(図1)。我々は、このタイムスパン簡約が、楽曲のもつ意味の一面を表していると考え。

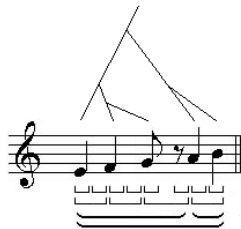


図 1: タイムスパン簡約の例

延長的簡約は、タイムスパン簡約では表現できない、以下の2つの構造を表すものである。1つは、グループの終わりの音と同じ音が次のグループの始まりである場合。つまりグループにまたがる音の持続である。も

う1つはメロディの進行における緊張と弛緩の構造である。この構造は和声の進行に関する構造分析である。延長的簡約はタイムスパン簡約の結果を用いてトップダウンに行われる[6]。

上記の各々の分析や理論は以下の2種類の規則によって定義されている。

- 1 構成規則 (Well-formedness rule)

構造を生成するためのルール。

- 2 選好規則 (Preference rule)

複数の構造が構成規則を満たす場合、好ましい構造を示すためのルール

構成規則は、構造を与えたり、生成された構造に制約を加えるための必要条件である。優先規則は、複数の構造が構成規則を満たす場合において、構造間に優先順位をつけるための規則である。本研究報告では文献[3]にならい、各規則を基本構造分析の各頭文字と規則名の頭文字で省略表示する(例えば GPR,MPR 等)。

グルーピング構造における各々の規則は、グループの境界となる場所の候補を特定するものである。GPR1-3はグルーピングを行う際、最も基本となる選好規則である。GPR4-7は、より高いレベルにおけるグルーピングを行う際、使用される選好規則である。拍節構造分析において、MPR2-5は最も低いレベルにおける拍節構造を特定するための基本となる選好規則である。MPR1,6-10は、より高いレベルにおける拍節構造分析を行うための選好規則である。

### 3 グルーピング構造分析と拍節構造分析における問題点

前述したように GTTM のコンピュータ上への実装は困難であるとされている。また我々は、実用の視点から Standard MIDI File(SMF)[4]を入力として GTTM の自動化を目指す。従って、スラーや休符などの情報が得られない。このため、少ない情報に対する GTTM の規則適用を考えなければならない。

本研究報告では、この GTTM 自動化への第1歩としてグルーピング構造と拍節構造における問題点に着目する。

#### 3.1 ポリフォニーへの拡張

GTTM では、全ての楽曲を本質的にホモフォニー(homophony)として扱っている。そのため本来ポリフォニー(polyphony)の楽曲でもホモフォニーとして分析されてしまう(図2)。ポリフォニーはホモフォニーが時間的に重畳していると見なして、各旋律ごとに解析する必要があると思われる。

#### 3.2 用語定義の曖昧性

GTTM の規則中、定義の曖昧な用語がいくつか現れる。拍節構造を例にとってみると、MPR5(length)



図 2: GTTM によるホモフォニー的分析の例

では、相対的に長い音、相対的に長く続く一定の音量、のような「より長い」という記述がある。ここで相対的に長いと書いてあるが、その比較対象が定義されていない。そのため比較する範囲には、この定義を適用する分析者もしくは実装者の主観が入ってしまう。

同様に曖昧な定義が MPR6(バス音が拍節的に安定した拍節構造を優先する), MPR7(カデンツでは拍節的に安定した構造を優先する)にも見られる。ここでは拍節的に安定したという表現が使用されているが、どのような状態が拍節的に安定しているのかということが定義されていない。そのため、ここでも主観が入ってきてしまう。

### 3.3 規則適用の非一貫性

基本的にグルーピング構造の各規則が適用された音符の所がグループ境界となる。しかし、GTTMのグルーピング例を見てみると同様の条件でも切れている所と切れていない所が存在する。以下の図3は人手によって GTTM の規則を適用した結果である。

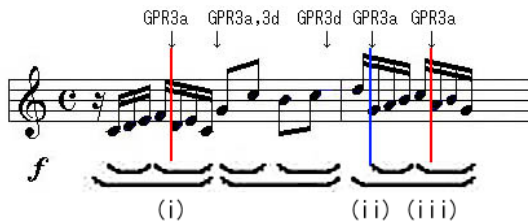


図 3: グルーピング構造における曖昧性の例

この例では、(ii)の GPR3a ではグループは切れている。しかし、同様に GPR3a が適用されている (i),(iii) ではグループは切れていない。これは GPR5(Symmetry) と GPR6(Parallelism) が競合したためである。また他の暗黙的な理由により、GTTM の規則が全く適用されていない場所でグループが切れる場合がある。従って、規則どうしの競合解消や暗黙的な規則の補完を行わなければならない。

## 4 本手法における問題点の解決

上述の問題点に対しプログラムの主観をできるだけ入れずに、音楽的に妥当であると考えられる規則のみで実装を行う必要がある。本手法では、以下のような対応で問題点を解決した。

表 1: 本手法における問題点への対応

ポリフォニーへの拡張	→ ボイスリーディング分析
用語定義の曖昧性	→ ヒューリスティックスの導入, → 暗黙的パラメータの明示化
規則適用の非一貫性	→ ボイスリーディング分析, 第一声部進行分析

### 4.1 改良した理論における楽曲の扱い

本手法では、ポリフォニーをホモフォニーの時間的重畳として扱う。即ち、まずホモフォニーレベルでのグルーピングを行い、次にホモフォニー間のグルーピングを行う。最終的に一つのグループに集約する。(図4に例を示す)。



図 4: 本手法におけるポリフォニー的分析の例

これにより、異なる旋律間での音符の並行性(同じ様な音列)を検出すること等が期待できる。

### 4.2 ヒューリスティックス

曖昧な用語の定義を適当に補填してしまうと、GTTM の文献や人手によって分析された結果に即した出力が得られない。そのため、本システムでは用語の定義の曖昧性を解消するために、ヒューリスティックスを導入した。その典型的な例を以下に示す(表2)。表中、MPR5(Length)の相対性を解消するために文献[3]から例題を20個用いて、ヒューリスティックスとして導入したパラメータの値を調整した。

表 2: ヒューリスティックスの例

SearchRangeMPR5a	=	2;
・音長を比較する音符は前後 2 個		
RelativelyLongMPR5a	=	1.1;
・比較する音符の音長が 1.1 倍以上 になっていれば規則を適用		

### 4.3 ボイスリーディング

GTTM の規則は全てグループやフレーズの切れ目を見出す規則である。従ってポリフォニーに GTTM の規則を適用すると、重畳したホモフォニー間に多数の切れ目を見出ししてしまう。そこで GTTM の規則適用を適切な範囲に制限するために、グループやフレーズの連続性を見出す規則を付加する。また、3.3 節で述べた問題は、グルーピングに関する情報が不足しているものと考え、それを補うためにやはり連続性に関する情報を検出する。それがボイスリーディングと次節で述べる第一声部進行である。

ボイスリーディングは音の進行を表したものである(図 5)。



図 5: ボイスリーディングの例

図中 1 音目と 2 音目, 4 音目と 5 音目は音程が 2 半音以下になっているので、連続性があると考えボイスリーディングが成立する。3 音目の前後においては 3 半音以上の音程が見られるので、ここには連続性がないと考える。

このボイスリーディングを求めることにより、GTTM の規則に、ボイスリーディング中の規則とそうでない規則とに強弱をつけることができる。また、規則が存在しない場所でもボイスリーディングが切れていれば、そこはグループやフレーズの切れ目の候補とすることができる。本手法では、ヒューリスティックとして、ある音と次の音との音程が 2 半音以下のものをボイスリーディングと定義している。

### 4.4 第一声部進行

第一声部進行は主旋律の抽出を意図している。主旋律は、楽曲のイメージの中でも最も強い音の進行であると考えられる。今回は、この主旋律を捉えるために

楽曲中の各音符につき第一声部 (最も高い音) を抽出した (図 6 の太線)。ボイスリーディングは音程が 2 半音以上離れている音に関しては連続性がないとして扱うのに対し、第一声部進行ではそのような音にも連続性があると考えられる。



図 6: 第一声部進行

## 5 テストデータに対するグルーピング実験

### 5.1 実装

我々は、提案手法を実験するためのプロトタイプシステムを作成した。本システムでは入力として標準 MIDI ファイル (SMF) を用いる。今後、様々な音楽アプリケーションシステムへの導入や発展を考え、本システムの入力は標準的かつインターネット等からも容易に入手できるファイル形式が好ましい。また、本システムが利用される環境を考えると、互換性の高いプログラム言語と出力ファイル形式を用いることが望まれる。そこで本システムは実装に JAVA を使い、出力ファイル形式に Extensible Markup Language (XML) [1] を用いた。図 7 は本システムのプロトタイプの構成である。

プロトタイプシステムは 3 つのサブシステムから構成されている。まず始めのサブシステムにおいて、入力された SMF を XML ファイルに変換する。SMF はバイナリ形式なので、プログラムで扱いやすいようにテキスト形式に変換する。次のサブシステムで、変換された XML ファイルを入力として、note on/off の対応をとる。これにより MIDI 形式で書かれていた音符情報を、1 つの音符単位にまとめてタグを作成し、XML ファイルで出力する。3 つ目のサブシステムは、音符単位の情報表現する XML ファイルを入力として、GTTM に基づき構造解析する。今回の実験ではグルーピング構造と拍節構造の分析を行い、その結果求められた構造を XML ファイルで出力する。

本システムは、現在 GPR2-3, MPR2-5 のみを実装している。また、最小単位のグループを 1 音としても分析結果の正当性を損なわないので、GPR1 は実装しなかった。

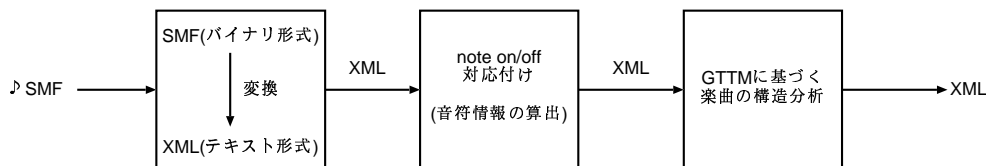


図 7: プロトタイプシステムの構成

## 5.2 実験

本手法で導入したヒューリスティクス，ボイスリーディング，第一声部進行の有効性を確認するために，以下のテストデータを入力した。

テスト曲 1 W.A.Mozart / G Minor Symphony

テスト曲 2 J.S.Bach / Invention No.1

テスト曲 1 は文献 [3] より採用し，テスト曲 2 は良く知られたポリフォニーとして採用した。

各テストデータ毎に，規則の適用結果とシステムの出力を示す(図 8, 図 9)。文献 [3] もしくは専門家によるグルーピングを正解とし，各レベルにおいて，あるグループが含むサブグループもしくは音符が異なっていると誤りとした。本システムでは GPR1 を省いたため，1 グループに 1 音しか含まれないグループは，グループとして数えていない。また，システムが出力しなかった正解グループも数えた(表 3, 表 4)。以下の式により再現率と適合率を求めた。

$$\text{再現率} = \frac{\text{システムが出力した正しいグループ数}}{\text{全ての求めたいグループ数}}$$

$$\text{適合率} = \frac{\text{システムが出力した正しいグループ数}}{\text{システムが出力した全てのグループ数}}$$

### (1) W.A.Mozart/G Minor Symphony

表 3: グルーピングの結果と再現率，適合率

正しいグループ	9
誤ったグループ	4
得られなかったグループ	3
再現率	$9/(9+3) = 0.75$
適合率	$9/(9+4) \approx 0.69$

図 8 において，○が正しいグループ，×が誤ったグループである。本システムでは，サブグループをまとめてより大きなグルーピングを行う手順として，曲の頭から 4 分音符レベルを 2 個づつまとめている。このテストデータでは，3 小節目の休符を境に誤ったグルーピ

ングが行われてしまっている。この場合，正解のグルーピングは，3 小節目の休符を境に曲の前後で 2 つの大きなグループを形成するものである。グルーピングを誤った理由は，現在のシステムでは GPR4(Intensification) を実装していないため，休符の位置で複数の規則が適用されていてもグループの切れる強さは 1 つの規則と変わらない。

また，本システムは入力として SMF を用いているため休符の情報を得られず，このため 3 小節目の休符が強いグループの境界になっていることを認識できていない。

### (2) J.S.Bach/invention No.1

表 4: グルーピングの結果と再現率，適合率

正しいグループ	67
誤ったグループ	0
得られなかったグループ	0
再現率	$67/(67+0) = 1.00$
適合率	$67/(67+0) = 1.00$

図 9 の a レベルでは，各 GPR の適用，ボイスリーディングを参照しながら，曲中の最長音符(4 小節目 4 拍目の 4 分音符 C5)を基準に拍節構造の 4 分音符レベル単位でグルーピングを行った。本手法では GPR4-7 の各規則を適用する最小単位を 4 分音符レベルと設定したため，4 分音符レベルでグルーピングした。b レベルは，各拍(4 分音符)ごとに a レベルのグループをまとめた。c レベルではさらに 2 拍分づつグループをまとめた。

図 9 では，第 1 小節 2 拍目の F5, D5, E5, C5 というパターンに対しシステムは a レベルで (F5)(D5, E5)(C5) というグルーピング結果を出力した。他のグルーピング候補として (F5, D5)(E5, C5) も考えられるが，本システムは (D5, E5) のボイスリーディングを優先してこれを正解とした。

このテストデータでは，ポリフォニーのホモフォニーへの分離がうまく行われ，理想的な結果が得られている。

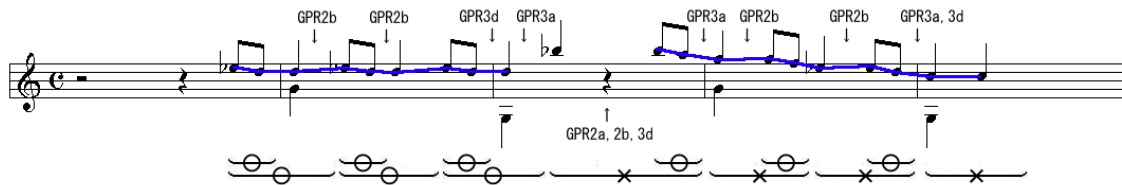


図 8: G Minor Symphony(規則の適用結果とグルーピングの結果)

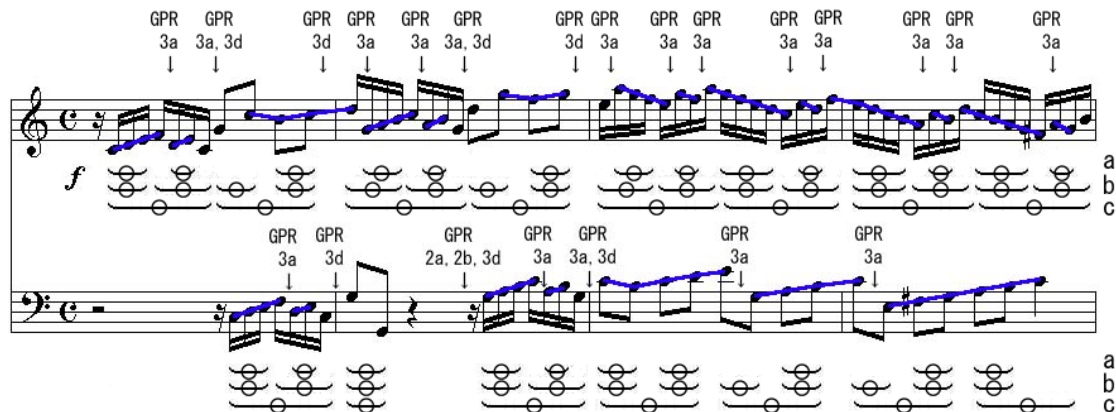


図 9: Invention(規則の適用結果とグルーピングの結果)

## 6 おわりに

本手法では、ボイスリーディングと第一声部進行を導入し、いくつかの拡張を行った。その結果、グルーピングを行う際の情報が増え、GTTMの規則が不足している所ではそれを補い、過剰な所では候補を絞り込むことができるようになった。またテストデータ実験によりその効果を確認できたが、いくつかの問題点も発見された。

現在 GPR4-7 の実装が遅れている原因は、GPR4-6 は互いに競合する規則であり、その優先順位を適切に定義できないためである。GPR7 はタイムスパン簡約の結果をフィードバックしてくるため、現時点では実装不可能である。これらに対し、現在次のような対策を考えている。

GPR4-6 が競合する問題には、各々の規則に効用を持つエージェントを導入し、全ての解候補を並列に計算していく方法を考えている。GPR7 のタイムスパン簡約からのフィードバックには、マルチエージェントにより求めた全ての解から、最も GPR7 に適合する解を選ぶ方法を考えている。

このため今後の課題点として、休符の認識や小節線の認識、現在未実装である GPR4-6 (Parallelism, Symmetry, Intensification) の実装が挙げられる。

## 参考文献

- [1] Gerd Castan, Michael Good, and Perry Roland, "Extensible Markup Language(XML) for Music Applications: An Introduction", Computing in Musicology 12, pp.95-102 (1999-2000).
- [2] 平田圭二, 青柳龍也, "パーピープン: ジャズ和音を生成する創作支援ツール", 情報処理学会論文誌, Vol.42, No3, pp.633-641 (2001).
- [3] Fred Lerdahl and Ray Jackendoff, "A Generative Theory of Tonal Music", The MIT Press (1983).
- [4] Curtis Roads, "コンピュータ音楽 -歴史・テクノロジーアート-", 東京電気大学出版局 (2001).
- [5] Alan Rittenberg, "Review and Discussion of A Generative Theory of Tonal Music", <http://alanr.www.media.mit.edu/people/alanr/Jackendoff&LerdahlFinal.html> (1994).
- [6] 竹内好宏, "音楽の構造解析とその応用", bit 別冊 コンピュータと音楽の世界, 共立出版, pp.224-240 (1998).