

GTTM グルーピング構造分析の実装： ルールを制御するパラメータの導入

浜中雅俊^{†1, †2} 平田圭二^{†3} 東条敏^{†4}

^{†1} 日本学術振興会特別研究員 PD, ^{†2} 産業技術総合研究所

^{†3} NTT コミュニケーション科学基礎研究所

^{†4} 北陸先端科学技術大学院大学

m.hamanaka@aist.go.jp

本研究報告では、音楽理論 Generative Theory of Tonal Music (GTTM) に基づき、曲をフレーズ、モチーフなどに自動でグルーピングするシステムについて述べる。従来のメロディ分割手法は、メロディの局所的な境界を求めるのに主眼を置いていたのに対して、本手法では、ボトムアップ処理により求めた局所的境界を用いて、トップダウンに階層的なグルーピング構造を獲得することを可能とする。GTTM は計算機上への実装が期待される音楽理論の 1 つであるが、ルールの優先順位が決まっていなかったために、ルール間の競合が起きるといった問題がある。この問題を解決するため本研究では、ルールの優先順位を決めるためのパラメータを導入した。実験の結果、パラメータの調節によりグルーピング性能が 3 割向上することが確認できた。さらに本研究では、グルーピング結果と MusicXML に含まれる拍節構造を用いて生成したタイムスパン木を示す。

An Implementation of Grouping Rules of the GTTM: Introducing of Parameters for Controlling Rules

Masatoshi Hamanaka^{†1, †2} Keiji Hirata^{†3} Satoshi Tojo^{†4}

^{†1} Research Fellow of the Japan Society for the Promotion of Science,

^{†2} National Institute of Advanced Industrial Science and Technology (AIST),

^{†3} NTT Communication Science Laboratories, ^{†4} Japan Advanced Institute of Science and Technology

Mbox 0604, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568 Japan

This report describes an automatic grouping system, which segments the music into units such as phrases or motives, based on the Generative Theory of Tonal Music (GTTM). Previous segmentation methods have been unable to construct hierarchical grouping structures because they have focused on detecting the local boundaries of the melody. Our system of segmentation based on the GTTM makes it possible to construct hierarchical grouping structures in a top-down process using bottom-up detection of local boundaries. The GTTM is considered to be one of the most promising theories of music in regard to computer implementation; however, no order in applying those rules is given and thus, more often than not, may result in conflict among them. To solve this problem, we introduce adjustable parameters, which enable us to give priority among rules. We show the experimental results that our method outperformed the baseline performance by over thirty percent, tuning the parameters. In addition, we show that the system displays the time-span tree based on these grouping rules together with metric information given by input MusicXML.

1. はじめに

本研究の目的は、音楽理論 Generative Theory of Tonal Music (GTTM)[1]に基づき、曲中の各音の構造的な重要度を階層的に表示したタイムス

パン木を、自動で獲得することである。タイムスパン木の自動生成は、音楽の深層構造の分析 [2][3]を可能とするだけでなく、音楽情報検索システムの出力結果を提示する際などに有用な、音楽要約[4][5]を実現する。

GTTM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いる。タイムスパン簡約を行うためには、グルーピング構造分析の結果と拍節構造分析の結果が必要である。グルーピング構造分析は、連続したメロディをフレーズやモチーフなどに階層的に分割するもので、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ある。本研究では、タイムスパン木を自動生成するための第1歩として、グルーピング構造分析を計算機上へ実装する。

従来のメロディ分割手法は、メロディの局所的な境界を求めることに主眼を置いていたため、階層的なグルーピング構造を得ることはできなかった[6][7][8][9]。文献[10]では、ピアノロール上に描いたボロノイ線図を用いて、多声音楽の階層的なグルーピングを実現していたが、音楽的な根拠に乏しく、曲によっては、うまくグルーピングできない場合があった。一方、文献[11][12]では、GTTM グルーピング構造分析の計算機上への実装が試みられたが、それらの手法では、ルールの競合の問題が解決できていなかった。

本研究では、GTTMに基づき、ボトムアップ処理により求めた局所的境界を用いて、トップダウンに階層的なグルーピング構造を獲得することを可能とする。そして、ルールの優先順位を制御するためのパラメータを導入することで、ルールの競合の問題を解決した。階層的なグルーピング構造はユーザによるパラメータの変更に伴い変化する。システムの入力には MusicXML を採用し、得られたグルーピング結果の出力には、新たに提案する GroupingXML を用いることにした。実験の結果、ユーザがパラメータを適切に設定することによって、グルーピング性能が3割向上することが確認できた。さらに本研究では、グルーピング結果と MusicXML に含まれる拍節構造を用いて生成したタイムスパン木を示す。

2. ルール適用における問題点

グルーピング構造分析は、曲はグループの集合からなり、グループはサブグループを含んでいるという直観を形式化しようとしたものである。グルーピング結果は、楽譜の下に階層的に並んだ円弧で描かれる。GTTMには、グルーピング構成ルール (Grouping Well-Formedness Rules, 以下 GWFR) とグルーピング選好ルール (Grouping Preference Rules, 以下 GPR) という2種類のグルーピングに関するルールがある。GWFRは、グルーピング構造が成立するために必要な条件や制約であり、GPRは、GWFRが成り立つグルーピング構造が複数存在する場合、どれが好まし

いかを示すルールである。GPRは、GPR1 (alternative form), GPR2 (proximity), GPR3 (change), GPR4 (intensification), GPR5 (symmetry), GPR6 (parallelism), GPR7 (time-span and prolongational stability) という7つのルールからなる。7つのルールは、GPR1から3までの局所的な境界を求めるためのルールと、GPR4からGPR7までの高次の境界を求めるためのルールに2分できる。さらに、GPR2は、(a) (slur/rest)と(b) (attack-point)の2通りに分かれ、GPR3は(a) (register), (b) (dynamics), (c) (articulation) および(d) (length)の4通りに分かれる。

本節では、GPRを計算機上へ実装する際の問題点について述べる。

2.1 ルールの競合

GPRを適用する際、ルールの適用順序が決まっていなかったため、ルールの競合がしばしば起きる。図1は、GPR2b (attack-point) と GPR3a (register)が競合している例である。図1では、GPR2bが第3音-4音間に適用され、GPR3aが第2音-3音間に適用されているが、GPR1 (alternative form)は第3音が単音のグループになることを嫌うため、両者を境界とすることはできない。

本研究では、ルールの優先順位を制御するための、調節可能なパラメータを導入することで、この問題を解決する。



図1: ルールの競合の例

2.2 定義の曖昧性

GPR4, 6の定義には、抽象的で曖昧な部分が多く含まれているため実装が困難である。例えば、GPR4(intensification)では、GPR2,3で示される効果が比較的明白なところがグループの境界になりやすいと定義しているが、比較対照が示されていないため、どのような場合が比較的明白で、どのような場合が比較的明白でないかは明かになっていない。また、GPR6(parallelism)では、2つ以上の並列的な部分があった場合には、並列性のあるグルーピングを行うことが定義されているが、どのような場合が並列的であるか明らかになっていない。

本研究では、ルールが適用できるかどうかを決める基準を定式化することにより、この問題を解決する。

3. GTTM に基づく自動グルーピング

図 2 に、システムの構成を示す。システムの入力形式には、楽譜作成や、分析、検索ツールが普及しており、フォーマットの相互変換が容易な MusicXML[13]を採用した。階層的なグルーピング構造は、ボトムアップ処理により求めた局所的境界を用いて、トップダウンに獲得する。獲得したグルーピング結果の出力形式としては、GroupingXML を新たに制定した。GTTM ではすべての楽曲をホモフォニー (homophony) として扱っているが、本研究では、システムの性能を正しく評価するため、分析の対象をモノフォニー (monophony) に限定する。

3.1 MusicXML

MusicXML は XML (extensible mark-up language) に基づく楽譜表記の方法で、アトリビュートエレメントとノートエレメントからなる。アトリビュートエレメントには、調記号、拍子記号および音部記号が記述され、ノートエレメントには、音高、音価およびノーテーションエレメントが記述される。ノーテーションエレメントには、タイ、スラー、フェルマータ、アルペジオ、強弱記号、装飾音、アーティキュレーション、などが記述される。

文献[11][12]では、システムの入力として SMF (Standard MIDI File) を採用していたため、シーケンスソフトで入力した SMF だけでなく、MIDI 楽器などを用いて、人間が演奏した SMF

をシステムに入力することができるという利点があった。しかし、実際に手弾きの SMF を用いる場合には、クオンタイズなどの前処理が必要であった。

これに対して、MusicXML は SMF からの変換が容易であるばかりでなく、SMF では記述することが困難な休符やスラーに関する情報を扱うことができるという特長がある。

実験で用いたすべての入力データは、楽譜作成ソフトウェア *Finale*TM[14]を用いて手作業で入力し、MusicXML 出力用のプラグイン *Doler*TM でエクスポートして作成したもので、強弱記号およびアーティキュレーションは含まれていない。

3.2 グルーピング選好ルールの適用

本節では、GPR1, GPR2a, GPR2b, GPR3a, GPR3d, GPR4, GPR5, および GPR6 の適用について述べる。入力データに強弱記号およびアーティキュレーションが含まれていないため、GPR3b (articulation) と GPR3c (dynamics) については適用しない。また、タイムスパン簡約やプロロングーション簡約の結果が必要となる GPR7 (time-span and prolongational stability) も適用しない。各ルールにより生じるグループの境界の深さは、 $D_i^{GPRj} = (1, 2a, 2b, 3a, 3d, 4, 5, 6) \cdot 0 \leq D_i^{GPRj} \leq 1$ で表される。

本研究では次の 13 個の調節可能なパラメータを導入する (表 1)。

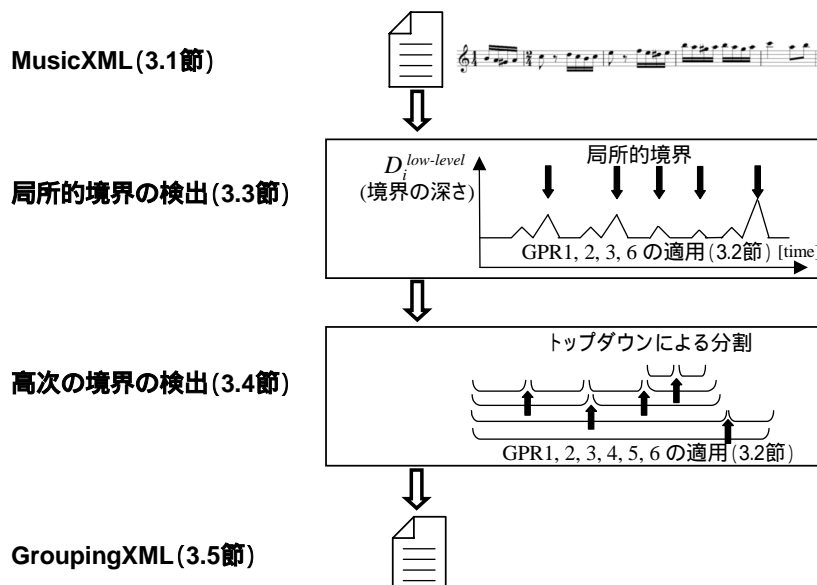


図 2: システムの構成

ホモフォニーは和音を含む単旋律，モノフォニーは和音を含まない単旋律である。

表 1: 13 個の調節可能なパラメータ

パラメータ	説明
S^{GPRj}	各ルールの強さを表すパラメータ。値が大きいほど、ルールの強さが強くなる。 $j=(2a,2b,3a,3d,4,5,6)$ 式(15),(16)で使用。
W_s	GPR5 で用いる、平均をグループの中心とする正規分布の分散。値が大きくなるほど正規分布の裾野が広がる。式(6)で使用。
W_r	GPR6 で、リズム方向のずれと音高方向のずれのどちらを重視するかを決めるパラメータ。値が大きいほど、音高を重視する。式(12), (13), (14)で使用。
W_l	GPR6 で、パラレル区間の長さをどのくらい重視するかを決めるパラメータ。値が大きいほど、長いパラレル区間を重視する。式(12), (13), (14)で使用。
T^{GPR4}	GPR4 で、GPR 2,3 の効果が明白であるかどうかを決める閾値。値が小さいほど、GPR4 が成立しやすくなる。式(5)で使用。
$T^{low-level}$	低レベルの境界であるかどうかを決める閾値。値が小さいほど、境界と認識されやすくなる。式(16)で使用。

- **GPR2, 3, および 4 の適用。** GPR2, 3, および 4 は連続する 4 音 ($n1, n2, n3, n4$) に関するルールである。 i 番目の遷移 (i 音目と $i+1$ 音目の間) が各ルールにより生じる境界の深さは、境界に成り得る ($D_i^{GPRj} = 1$) かそうでない ($D_i^{GPRj} = 0$) かで表される。

GPR2a (slur/rest) では、 $n2$ の終わりから $n3$ の始まりまでの時間間隔が $n1$ の終わりから $n2$ の始まりまでの時間間隔および $n3$ の終わりから $n4$ の始まりまでの時間間隔よりも長い場合、グループの境界と認識される。GPR2a は次式のように定式化できる。

$$D_i^{GPR2a} = \begin{cases} 1 & rest_{i-1} < rest_i \text{ and } rest_i > rest_{i+1} \\ 0 & rest_{i-1} \geq rest_i \text{ or } rest_i \leq rest_{i+1} \end{cases} \quad (1)$$

ただし、

$rest_i$: 消音時刻から次の発音時刻までの間隔

GPR2b (attack-point) では、 $n2$ の始まりから $n3$ の始まりまでの時間間隔が $n1$ の始まりから $n2$ の始まりまでの時間間隔および、 $n3$ の始まりから $n4$ の始まりまでの時間間隔よりも長い場合、グループの境界と認識される。GPR2b は次式のように定式化できる。

$$D_i^{GPR2b} = \begin{cases} 1 & ioi_{i-1} < ioi_i \text{ and } ioi_i > ioi_{i+1} \\ 0 & ioi_{i-1} \geq ioi_i \text{ or } ioi_i \leq ioi_{i+1} \end{cases} \quad (2)$$

ただし、

ioi_i : 発音時刻間隔 (inter onset intervals)

GPR3a (register) では、 $n2, n3$ の音高差が $n1, n2$ の音高差および $n3, n4$ の音高差よりも大きい場合、グループの境界と認識される。GPR3a は次式のように定式化できる。

$$D_i^{GPR3a} = \begin{cases} 1 & regi_{i-1} < regi_i \text{ and } regi_i > regi_{i+1} \\ 0 & regi_{i-1} \geq regi_i \text{ or } regi_i \leq regi_{i+1} \end{cases} \quad (3)$$

ただし、

$regi_i$: 音程

GPR3d (length) では、 $n2$ と $n3$ が異なった音長をもち、 $n1$ と $n2$ および $n3$ と $n4$ が同じ音長の場合、グループの境界と認識される。GPR3d は次式のように定式化できる。

$$D_i^{GPR3d} = \begin{cases} 1 & len_{i-1} = 0 \text{ and } len_i \neq 0 \text{ and } len_{i+1} = 0 \\ 0 & len_{i-1} \neq 0 \text{ or } len_i = 0 \text{ or } len_{i+1} \neq 0 \end{cases} \quad (4)$$

ただし、

len_i : 音価の差

GPR4 (intensification) では、GPR2,3 で示される効果が比較的明白な場合、グループの境界と認識される。GPR4 は次式のように定式化できる。 $P_i^{rest}, P_i^{ioi}, P_i^{regist}$ は、それぞれ、GPR2a, GPR2b, GPR3a が明白に成立するほど大きな値を示す。 T^{GPR4} ($0 < T^{GPR4} < 1$) は、GPR2,3 の効果が明白であるかどうかを決める閾値である。

$$D_i^{GPR4} = \begin{cases} 1 & P_i^{rest} > T^{GPR4} \\ & \text{or } P_i^{ioi} > T^{GPR4} \text{ or } P_i^{regist} > T^{GPR4} \\ 0 & P_i^{rest} \leq T^{GPR4} \\ & \text{and } P_i^{ioi} \leq T^{GPR4} \text{ and } P_i^{regist} \leq T^{GPR4} \end{cases} \quad (5)$$

ただし、

$$P_i^{rest} = \frac{rest_i}{\sum_{j=i-1}^{i+1} rest_j} \quad (0 \leq P_i^{rest} \leq 1),$$

$$P_i^{ioi} = \frac{ioi_i}{\sum_{j=i-1}^{i+1} ioi_j} \quad (0 \leq P_i^{ioi} \leq 1),$$

$$P_i^{regist} = \begin{cases} \frac{regi_i}{\sum_{j=i-1}^{i+1} regi_j} & regi_{i-1} \neq regi_i \text{ or } regi_i \neq regi_{i+1} \\ 0 & regi_{i-1} = regi_i \text{ and } regi_i = regi_{i+1} \end{cases} \quad (0 \leq P_i^{regist} \leq 1)$$

- **GPR5 (symmetry)の適用.** GPR5 では、グループの分割が長さの等しい2つの部分からなるようグルーピングすることを優先する．このようなことを実現するため、本研究では、分割されたグループの長さが等しいほど高い値を示す関数 D_{GPR5} を定義する．そのような関数は様々考えられるが、ここでは、平均をグループの中心、分散を σ とする正規分布を用いる．正規分布は、パラメータを調節することにより裾野の広さを自由に変えられるという特長がある．分散 σ は、調節可能なパラメータで、0 に近いほど正規分布の裾野が狭まり、グループの中心付近での D_{GPR5} の値が大きくなる．したがって、より中心に近いほどグループの境界になりやすくなる．逆に σ が大きくなった場合、中心付近以外の位置でもグループの境界になりやすくなる．

$$D_i^{GPR5} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\left(\frac{\sum_{j=start}^i ioi_j - \sum_{j=start}^{end} ioi_j}{2}\right)^2}{2\sigma^2}} \quad (6)$$

ただし、

- $start$: グループの始まりの遷移
- end : グループの終わりの遷移

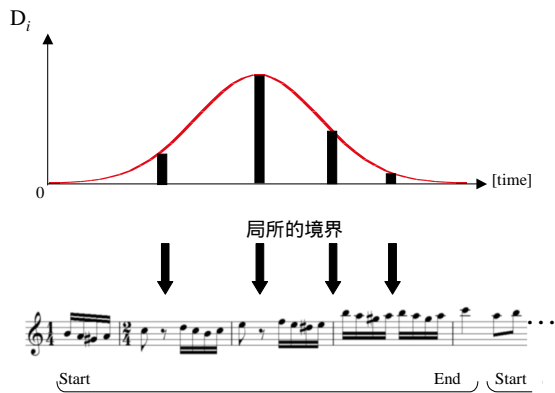


図 3 : GPR5 による境界の深さ

- **GPR6 (parallelism)の適用.** GPR6 では、並列性のある部分は並列性のあるグルーピングを優先する．ここでは、調整可能な3つのパラメータ W_s, W_r, W_l を導入する ($0 < W_s, W_r, W_l < 1$) . W_s は、リズム方向のずれと音高のずれのどちらを重視するか決めるパラメータ、 W_r は、平行な区間の始まりと終わりでどちらを重視するかを決めるパラメータ、 W_l は、平行区間の長さをどのくらい重視するかを決めるパラメータである．平行な部分の長さ r は、1 拍の長さの整数倍とする．したがって、

平行区間の始まりと終わりは、拍の境目になる． q_i は i 番目の遷移が曲の頭から数えて何拍目であるかを表す．

$$q_i = \left\lfloor \frac{\sum_{k=1}^i ioi_k}{div} \right\rfloor \quad \left(\lfloor \cdot \rfloor \text{はガウス記号} \cdot [n] \text{は} n \text{を越えない最大の整数を表す} \right) \quad (7)$$

ただし、

- div : 4 分音符の音価
- ioi_i : 発音時刻間隔 (inter onset intervals)

m_{ij} は、遷移 i と遷移 j が平行区間の開始や終わりになり得るかどうかを表す．遷移 i が平行区間の開始となるためには、遷移 $i-1$ に拍線がある、すなわち、 $q_i - q_{i-1} = 1$ が成立する必要がある．一方、遷移 i が平行区間の終わりとなるためには、遷移 $i+1$ に拍線がある、すなわち、 $q_i - q_{i+1} = -1$ が成立する必要がある． m_{ij} は、 i および j が平行区間の開始となり得る場合には $b(=begin)$ 、平行区間の終わりとなり得る場合には $e(=end)$ 、平行区間の開始と終了の両方になり得る場合には $t(terminal)$ 、開始にも終了にもなり得ない場合は $s(=else)$ とする．

$$m_{ij} = \begin{cases} b & q_i \neq q_{i-1} \text{ and } q_j \neq q_{j-1} \text{ and } q_i = q_{i+1} \text{ and } q_j = q_{j+1} \\ e & q_i = q_{i-1} \text{ and } q_j = q_{j-1} \text{ and } q_i \neq q_{i+1} \text{ and } q_j \neq q_{j+1} \\ t & q_i \neq q_{i-1} \text{ and } q_j \neq q_{j-1} \text{ and } q_i = q_{i+1} \text{ and } q_j = q_{j+1} \\ s & \text{else} \end{cases} \quad (8)$$

$x_{q_i r}$ は、 q_i 拍目から $q_i + r$ 拍目の間に、何個の音があるかを表す． $y_{q_i q_j r}$ は、 q_i 拍目から $q_i + r$ 拍目までの区間と q_j 拍目から $q_j + r$ 拍目までの区間を比較して、発音のタイミングが一致している音が何音あるかを表す． $z_{q_i q_j r}$ は、 q_i 拍目から $q_i + r$ 拍目までの区間と q_j 拍目から $q_j + r$ 拍目までの区間で発音のタイミングが一致している音を比較して、前の音からの音高変化 (音程) が一致する音の数を表す．

$$x_{q_i r} = \sum_j \begin{cases} 1 & q_i \leq q_j \text{ and } q_j \leq q_i + r \\ 0 & q_i > q_j \text{ or } q_j > q_i + r \end{cases} \quad (9)$$

$$y_{q_i q_j r} = \sum_k \sum_l \begin{cases} 1 & (q_i - q_j) \times div = \sum_{g=1}^k ioi_g - \sum_{g=1}^l ioi_g \\ 0 & (q_i - q_j) \times div \neq \sum_{g=1}^k ioi_g - \sum_{g=1}^l ioi_g \end{cases} \quad (10)$$

$$z_{q_i q_j r} = \sum_k \sum_l \begin{cases} 1 & (q_i - q_j) \times div = \sum_{g=1}^k ioi_g - \sum_{g=1}^l ioi_g \text{ and } regi_i = regi_j \\ 0 & (q_i - q_j) \times div \neq \sum_{g=1}^k ioi_g - \sum_{g=1}^l ioi_g \text{ or } regi_i \neq regi_j \end{cases} \quad (11)$$

平行な部分の長さが r のとき遷移 i と遷移 j の GPR6 による境界の深さは G_{ijr}^{start} と

および G_{ijr}^{end} で表される。 G_{ijr}^{start} は、 q_i 拍目から q_i+r 拍目までの区間と q_j 拍目から q_j+r 拍目までの区間が平行な場合に高い値になり、 G_{ijr}^{end} は、 q_i-r 拍目から q_i 拍目までの区間と q_j-r 拍目から q_j 拍目までの区間が平行な場合に高い値となる（図4）。

$$G_{ijr}^{start} = \frac{z_{q_i q_j r}}{x_{q_i r} + x_{q_j r}} \times W_r' \times r W_i + \frac{y_{q_i q_j r}}{z_{q_i q_j r}} \times W_r \times r W_i \quad (12)$$

$$G_{ijr}^{end} = \frac{z_{q_i-r q_j-r r}}{x_{q_i-r r} + x_{q_j-r r}} \times W_r' \times r W_i + \frac{y_{q_i-r q_j-r r}}{z_{q_i-r q_j-r r}} \times W_r \times r W_i \quad (13)$$

平行な度合い D_i^{GPR6} は次式で表される。

$$D_i^{GPR6} = \sum_j \begin{cases} G_{ijr}^{start} \times W_s' & m_{ij} = b \\ G_{ijr}^{end} \times W_s & m_{ij} = e \\ G_{ijr}^{start} \times W_s' + G_{ijr}^{end} \times W_s & m_{ij} = t \\ 0 & m_{ij} = s \end{cases} \quad (14)$$

ただし、

$$\begin{aligned} W_s' &= 1 - W_s \\ W_r' &= 1 - W_r \\ W_i' &= 1 + W_i \end{aligned}$$

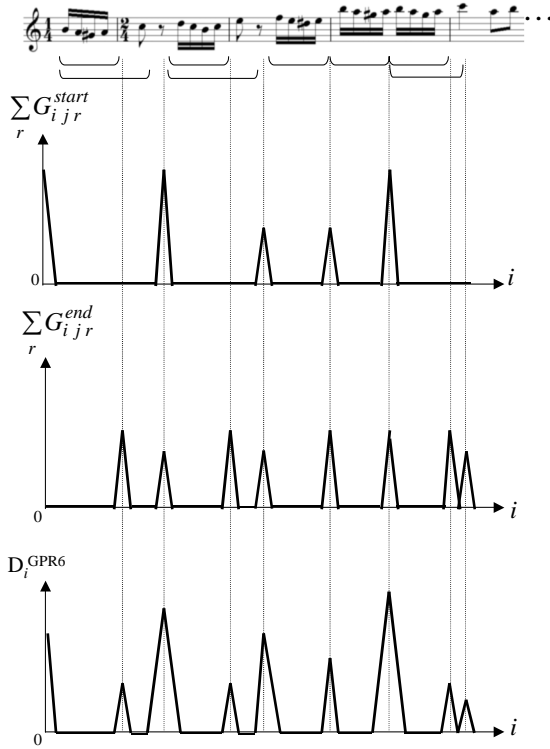


図4：GPR6による境界の深さ

- **GPR1の適用**。GPR1では、非常に小さいグループ、特に単音からなるグループへの分割は避ける。したがって、グループの境界となるため

には、局所的な境界の深さが前後の遷移より深い必要がある。GPR1による境界の深さは、境界に成り得る ($D_i^{GPR1} = 1$) かそうでない ($D_i^{GPR1} = 0$) かで表される。 B_i は、局所的な境界の深さを0から1の実数で表したものである。 B_i の値が大きいほど、境界が深いことを表す。

$$D_i^{GPR1} = \begin{cases} 1 & B_{i-1} \leq B_i \text{ and } B_i \geq B_{i+1} \\ 0 & B_{i-1} > B_i \text{ or } B_i < B_{i+1} \end{cases} \quad (15)$$

ただし、

$$B_i = \frac{\sum_{j=(2a,2b,3a,3d,6)} D_i^{GPRj} \times S^{GPRj}}{\max_{i'} \left(\sum_{j=(2a,2b,3a,3d,6)} D_i^{GPRj} \times S^{GPRj} \right)}$$

3.3 局所的境界の検出

局所的な境界は、 D_i^{GPR1} 、 D_i^{GPR2a} 、 D_i^{GPR2b} 、 D_i^{GPR3a} 、 D_i^{GPR3d} 、および D_i^{GPR6} を用いて検出する。 $T^{low-level}$ は、遷移 i がグループの境界となる ($D_i^{low-level} = 1$) かそうでない ($D_i^{low-level} = 0$) かを決めるための閾値である。 B_i は、式(15)と同じである。

$$D_i^{low-level} = \begin{cases} 1 & B_i > T^{low-level} \text{ and } D_i^{GPR1} = 1 \\ 0 & B_i \leq T^{low-level} \text{ or } D_i^{GPR1} \neq 1 \end{cases} \quad (16)$$

3.4 階層的なグルーピング構造の検出

階層的なグルーピング構造は、ボトムアップ処理により求めた局所的境界を用いて、トップダウンに獲得する。グループがその内部に局所的境界を含んでいる場合、次式によって、次の階層の境界が再帰的に求まる。

$$\hat{i} = \underset{i}{\operatorname{argmax}} (D_i^{low-level} \times \sum_j D_i^{GPRj} \times S^{GPRj}) \quad (17)$$


ただし、

$$j = (2a, 2b, 3a, 3d, 4, 5, 6)$$

i はグループに含まれるすべての遷移

3.5 GroupingXML

本研究では、階層的なグルーピング構造の出力形式として Grouping-XML を提案する（図5）。Grouping-XML は、グループエレメント、ノートエレメント、アブライドエレメントからなる。ノートエレメントは、発音時刻順に並んでおり Xpointer[15]と Xlink[16]を用いて MusicXML 上の各音とリンクしている。すべてのノートエレメントは、階層的なグループエレメントの内部に存在する。アブライドエレメントは、グループの終了タグと次のグループの開始タグの間に位置し、GPRの適用位置を表す。本研究では、グルーピ



```

-<group>
-<group>
+<note id="P1-1-1"/>
+<note id="P1-1-2"/>
+<note id="P1-1-3"/>
+<note id="P1-1-4"/>
+<note id="P1-2-1"/>
</group>
<applied rule="2a"/>
<applied rule="6"/>
-<group>
+<note id="P1-2-3"/>
+<note id="P1-2-4"/>
+<note id="P1-2-5"/>
+<note id="P1-2-6"/>
+<note id="P1-3-1"/>
</group>
</group>

```

図 5 : GroupingXML の例

ング構造を表示するため、GroupingXML ビューアを作成した(図 6) .GroupingXML ビューアは、音符をピアノロール形式で表示し、音符の下に階層的に並んだ円弧がグルーピング構造を表す。音符の上につながっている木構造は、タイムスパン木である。図 6 のタイムスパン木は、システムの出力するグルーピング結果と MusicXML に含まれる拍節構造を忠実に反映して手作業で作成したものであるが、この描画はタイムスパン木作成規則の実装によって今後自動化できる予定である。

4. 実験結果

本節では、グルーピングの性能の評価を、再現率(recall)と適合率(precision)を組み合わせた F

値で評価する。F 値は、再現率と適合率が高いほど、すなわち、グルーピングの性能が高いほど高くなる。

$$F_{\text{値}} = 2 \times \frac{P \times R}{P + R} \quad (18)$$

ただし、

- R: 再現率 (システムが出力したグルーピングと同じグループが、正解データに含まれている割合)
- P: 適合率 (正解データのグルーピングと同じグループが、システムの出力に含まれている割合)

F 値の計算は、グループが所属する階層に関係なく行う。したがって、局所的な境界が正しく求まっていない位置では、すべての階層で誤ったグルーピングとなり、その結果、F 値が大きく落ちる。

このような評価を行うためにはグルーピングの正解データを用意する必要がある。文献[7]や文献[17]では、グルーピングの正解データとして、エッセンフォークソング集[18]を用いて評価を行っている。エッセンフォークソング集は、6000 個のモノフォニーのメロディを EsAC フォーマット[19]でテキスト出力したものに、フレーズの切れ目を記述したものである。エッセンフォークソング集には、局所的な境界しか記述されていないため、それを用いて階層的なグルーピング構造を評価することは困難である。

そこで、本研究では、新たにグルーピングの正解データを作成した。正解データは、GTTM をよく理解している一人の音楽家がクラシック曲から切り出した 8 小節の長さの 100 個のメロディについて手作業でグルーピング構造分析をしたものである。正しいグルーピング構造となってい

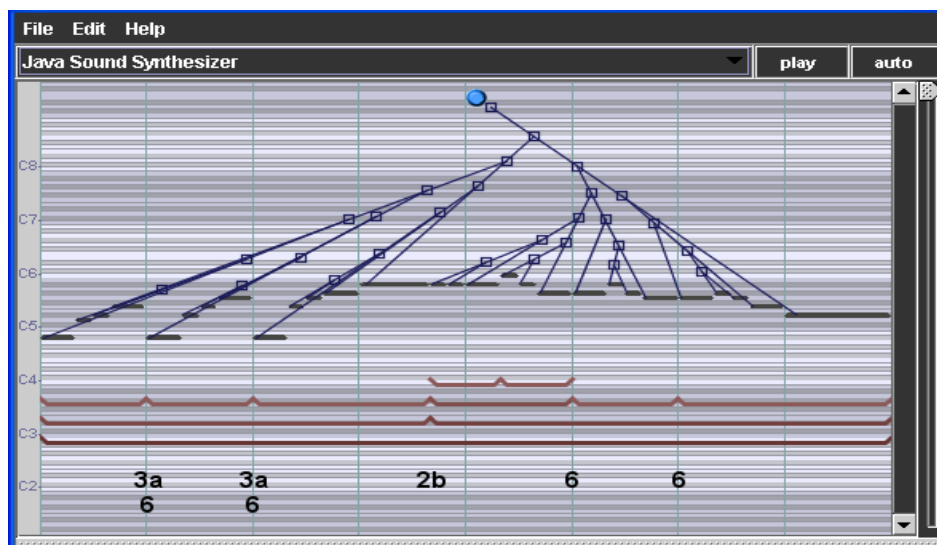


図 6 : GroupingXML ビューア

表 2:パラメータ調節前と後の F 値

メロディ	ベースライン	調整後
1. トルコ行進曲	0.09	0.95
2. 子守唄	0.41	1.00
3. 「椿姫」乾杯の歌	0.03	0.90
4. 「ジャンニ・スキッキ」私のお父様	0.03	0.11
5. 「クルミ割り人形」行進曲	0.01	0.05
6. 歌の翼に乗せて	0.42	0.62
7. ヨモレスク	0.34	0.76
8. 楽しい農夫	0.28	0.96
9. 花の歌	0.08	0.14
10. グリーンスリーブス	0.42	0.79
⋮	⋮	⋮
平均(100 個のメロディ)	0.32	0.67

るか確認するため、3 人の GTTM の専門家がグル
クrosチェックを行った。

グルーピング構造は、パラメータの調整によ
って変化する。そこでまず、パラメータ調節前
(ベースライン)の性能を求める。パラメータ
の初期値は、 $SGPR_j (= 2a, 2b, 3a, 3d, 4, 5, 6) = 0.5$,
 $= 0.05$, $W_s = 0.5$, $W_r = 0.5$, $W_f = 0.5$, $TGPR_A = 0.5$,
および $Tlow-level = 0.5$ である。次に、手作業で
パラメータを調節した。パラメータの調節は、
局所的な境界に関するパラメータから先に調
節し、そのあと高次の境界に関するパラメータ
を調節した。1 曲につき 10 分間でパラメータ
の調整を行った結果、ベースラインの性能より
F 値が 0.3 以上向上した(表 2)。

5. まとめ

本研究では、GTTM に基づき計算機上にグルー
ピング構造分析を実装した。グルーピング構造分
析 GPR の実装は、ルールの競合や、定義の曖昧
性などの問題で、これまで困難であったが、本研
究では、ルールが適用できる基準を定式化する
とともに、ルールの優先順位を決めるパラメータ
を導入することでグルーピングを実現した。そし
て、パラメータを調整することによりグルーピン
グの性能が 3 割向上することを確認した。さら
に、本研究では、得られたグルーピング結果の
出力形式として GroupingXML を提案し、グルー
ピング結果を表示するための専用のビューアを
作成した。

現在、GTTM の拍節構造分析およびタイムス
パン簡約を実装中であり、完成すればタイムス
パン木を自動導出した初めての例となる。今
後、曲ごとに最適なパラメータの値を自動で
設定する方法について考えていく。

参考文献

- [1] Lerdahl, F., and R. Jackendoff. A Generative Theory of Tonal Music. Cambridge, Massachusetts: MIT Press, 1983.
- [2] 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.
- [3] K. Hirata, and T. Aoyagi. "Computational Music
Representation on the Generative Theory of Tonal
Music and the Deductive Object-Oriented Database."
Computer Music Journal 27(3), 73–89, 2003.
- [4] 平田圭二, 松田周, パピブーン: GTTM に基づく
音楽要約システム, 情報処理学会研究報告
2002-MUS-46, pp.29-36, 2002.
- [5] 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.
- [6] D. R. Stammen, B. Pennycook. "Real-time
Segmentation of Music using an Adaptation of Lerdahl
and Jackendoff's Grouping Principles." In proceedings
of the International Conference on Music Perception
and Cognition, pp. 269-270, 1994.
- [7] D. Temperley. The Cognition of Basic Musical
Structures. Cambridge, Massachusetts: MIT Press,
2001.
- [8] E. Cambouropoulos. "The Local Boundary Detection
Model (LBDM) and its application in the study of
expressive timing." In Proceedings of the International
Computer Music Conference, pp. 290–293. Havana,
Cuba: International Computer Music Association, 2001.
- [9] M. Ferrand, P. Nelson, and G. Wiggins. "Memory and
Melodic Density: A Model for Melody Segmentation."
In Proceedings of the XIV Colloquium on Musical
Informatics (XIV CIM 2003), pp. 95–98. Firenze, Italy,
2003.
- [10] M. Hamanaka, K. Hirata: Applying Voronoi Diagrams
in the Automatic Grouping of Polyphony, Information
Technology Letters, Vol.1, No.1, pp. 101-102, 2002.
- [11] 井田健太郎, 平田圭二, 東条敏, GTTM に基づくグ
ルーピング構造および拍節構造の自動分析の試み,
情報処理学会研究報告 2001-MUS-42, pp.49-54,
2001.
- [12] 東洋武士, 平田圭二, 東条敏, 佐藤健, グルーピン
グ規則適用を改良した GTTM の実装, 情報処理学
会研究報告 2002-MUS-47, pp.121-126, 2002.
- [13] Recordare LLC. "MusicXML 1.0 Tutorial."
<http://www.recordare.com/xml/musicxml-tutorial.pdf>,
2004.
- [14] Recordare LLC. "MusicXML Software."
<http://www.recordare.com/software.html>, 2004.
- [15] W3C. "XML Pointer Language (XPointer)."
<http://www.w3.org/TR/xptr/>, 2002.
- [16] W3C. "XML Linking Language (XLink) Version 1.0."
<http://www.w3.org/TR/xlink/>, 2001.
- [17] B. Thom, C. Spevak, and K. Hothker. "Melodic
segmentation: evaluating the performance of algorithms
and musical expert." In Proceedings of the International
Computer Music Conference, pp. 65–72. Goteborg,
Sweden: International Computer Music Association,
2002.
- [18] E. Dahlig. "Essen folk song collection."
<http://www.esac-data.org/>, 1999.
- [19] E. Dahlig. "EsAC."
<http://www.cs.uu.nl/events/dech1999/dahlig/>, 1999.