

GTTM に基づく楽曲構造分析の実装: グルーピング構造と拍節構造の獲得

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、楽曲のグルーピング構造および拍節構造を自動で獲得するシステムについて述べる。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、ルール適用の優先順位やアルゴリズムが決められていないため、ルールの競合が起こるという問題がある。この問題の解決法として本研究では、ルールの優先順位を決めるためのパラメータを導入する。そして、ボトムアップに求めた局所的なグルーピング境界の強さおよび拍点の強さを用いて、トップダウンに階層的な構造を獲得する手法を提案する。実験の結果、パラメータの調節により、性能が向上することが確認できた。

An Automatic Music Analyzing System based on GTTM: Acquisition of Grouping Structures and Metrical Structures

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 music analyzing system, which acquire grouping structures and metrical structures, based on the Generative Theory of Tonal Music (GTTM). 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. Our system based on the GTTM makes it possible to construct hierarchical grouping and metrical structures in a top-down process using bottom-up detection of local boundary strength and local metrical strength. The experimental results show that our method outperformed the baseline performance by tuning the parameters.

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、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、その自動生成は、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おり、タイムスパン木はタイムスパン簡約の結果として求まる。本稿では、タイムスパン簡約を行うために必要となる、グルーピング構造分析と拍節構造分析を計算機上へ実装する方法について述べる。

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。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。

拍節構造分析は、4分音符/2分音符/1小節/2小節/4小節などそれぞれの拍節レベルにおける強拍と弱拍を同定するもので、聴取者が曲に合わせて手拍子を打つタイミングや指揮者がタクトを振るタイミングを求めるような分析である。従来のビートトラッキングに関する手法[11][12][13][14]は、階層的な拍節構造の獲得を実現していたが、そこで扱われたのは小節レベルまでの拍節構造で、本研究で扱うような、それより高次の拍節構造については考慮していなかった。

従来、GTTMのグルーピング構造分析と拍節構造分析の実装が試みられたが[15][16]、それらの手法ではルール競合の問題が解消できず、階層的なグルーピング・拍節構造を獲得することは困難であった。

これに対し我々は、ルールの優先順位を制御するためのパラメータを導入し、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していた[17]。しかし、そこで用いた入力データには、ダイナミクスおよびアーティキュレーションの情報が含まれていなかったため、それらに関するルールが実装できず、グルーピング性能が悪かった。

本研究では、新たに拍節構造分析を実装するとともに、ダイナミクスおよびアーティキュレーションに関するグルーピングルールを実装した。実験の結果、ルールを追加することでグルーピング性能が向上すること、ユーザがパラメータを適切に設定することによって拍節構造分析の性能が向上することが確認できた。

2. GTTMに基づく楽曲構造分析

GTTMは、Fred LerdahlとRay Jackendoffによって提唱された、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である。図1に、システムの全体像を示す。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

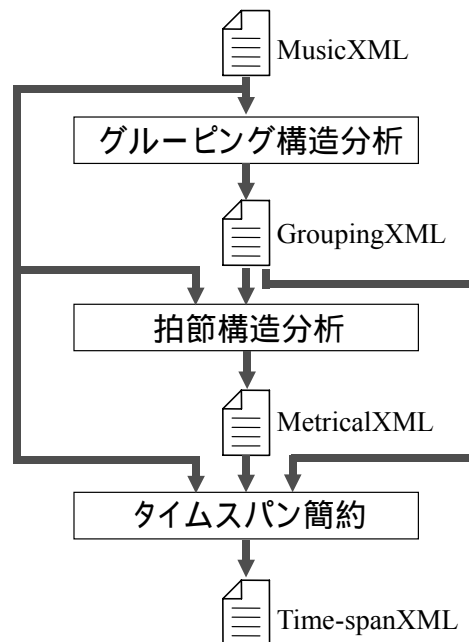


図1: システムの全体像

容易な MusicXML[18]を採用した。本研究では、グルーピング構造分析、拍節構造分析、タイムスパン簡約の結果のそれぞれの出力形式として、GroupingXML、MetricalXML、Time-spanXMLを提案する(図2)。MusicXML、GroupingXML、MetricalXML、Time-spanXML上にあるノートエレメントは、Xpointer[19]とXlink[20]を用いてそれぞれお互いにリンクしている。

2.1 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。

実験で用いたすべての入力データは、文献[17]で用いた MusicXML に、一人の音楽家が手作業で強弱記号及びアーティキュレーション記号を追加したものである。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[☆]。

[☆] 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。

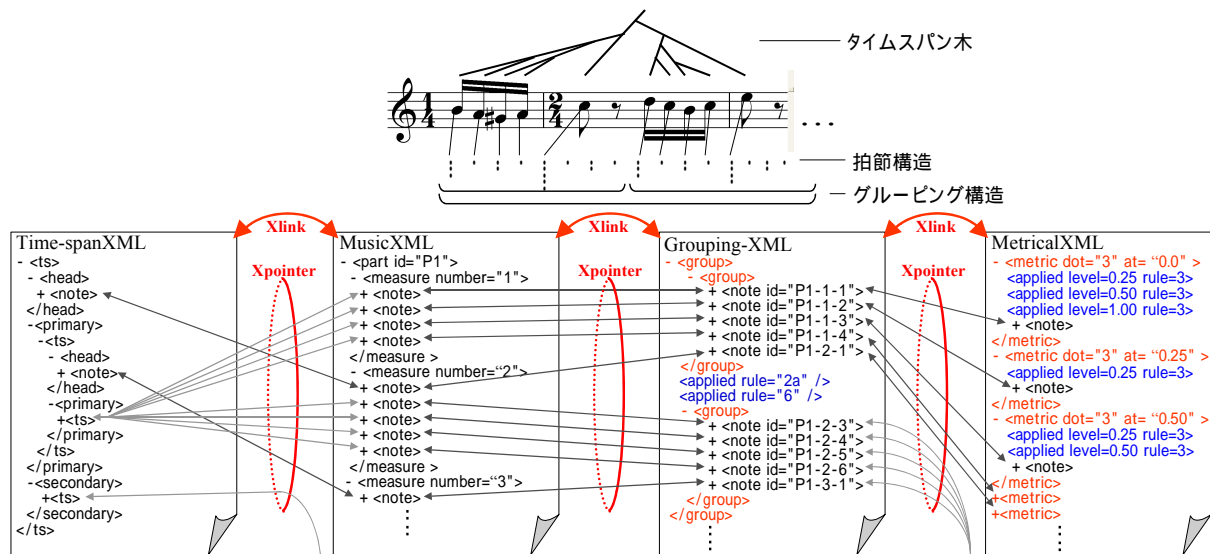


図 2: MusicXML, GroupingXML, MetricalXML, Time-spanXML

2.2 GroupingXML

階層的なグルーピング構造の出力形式として **GroupingXML** を提案する。 **GroupingXML** は、グループエレメント、ノートエレメント、アプライドエレメントからなる。すべてのノートエレメントは発音時刻順に並んでおり、階層的なグループエレメントの内部に存在する。アプ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。

2.3 MetricalXML

拍節構造の出力形式として **MetricalXML** を提案する。 **MetricalXML** は、メトリックエレメントとメトリックエレメントの内部にあるアプライドエレメント、ノートエレメントからなる。

メトリックエレメントは、曲中の拍の強さを最小拍節レベルの拍ごとに表したもので、アプライドエレメントは各拍節レベルに適用されるルールを表す。

2.4 Time-spanXML

Time-spanXML をタイムスパン木の出力形式として提案する。 **Time-spanXML** は、タイムスパンエレメント、ヘッドエレメント、プライマリーエレメント、セカンダリーエレメントおよびノートエレメントからなる。タイムスパンエレメントの中には、ヘッドエレメント、プライマリーエレメント、セカンダリーエレメントがそれぞれ一つずつある。ノートエレメントはヘッドエレメントの中に現れる。そして、プライマリーエレメントとセカンダリーエレメントの下にはタイムスパンエレメントが再起的に現れる。タイムスパン木の末端では、プライマリーエレメントとセカンダリーエレメントを含まないヘッドエレメントが現れる。

3. ルール適用における問題点

グルーピング構造分析と拍節構造分析は、それぞれ構成ルールと選好ルールの 2 種類によって定義されている。そして、2 種類のルールのうち計算機上に実装する際に問題となるのは選好ルールのほうである。構成ルールは、構造が成立するために必要な条件や制約で、選好ルールは、複数の構造が構成ルールを満たす場合、どれが好ましいかを示すルールである。

本節では、 **GTTM** のグルーピング選好ルール (**Grouping Preference Rule**, 以下 **GPR**) と拍節選好ルール (**Metrical Preference Rule**, 以下 **MPR**) を計算機上に実装する上での問題点について述べる。

3.1 ルールの競合

GPR や **MPR** 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。

本研究では、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、この問題を解決する。

3.2 定義の曖昧性

GPR4, **6** や **MPR5**, **6**, **7** 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あるかは明かになっていない。

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。

4. グルーピング構造分析の実装

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 (図 3)。入力は MusicXML, 出力は GroupingXML である。

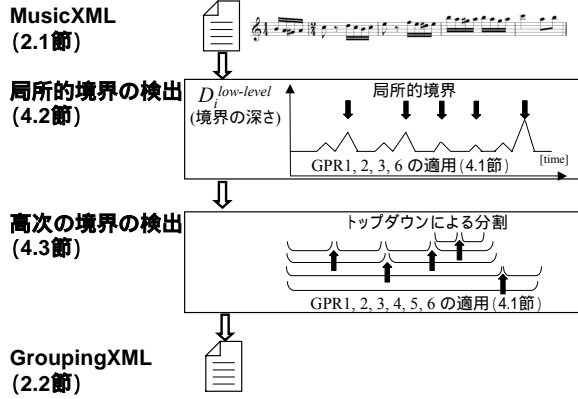


図 3: グルーピング構造分析

4.1 グルーピング選好ルールの適用

文献[17]では, GPR1, 2a, 2b, 3a, 3d, 4, 5, 6 の適用を行っていたが, 本研究では新たに GPR3b, 3c の適用を行う。図 4 は, GPR の適用とパラメータの関係を表したものである。

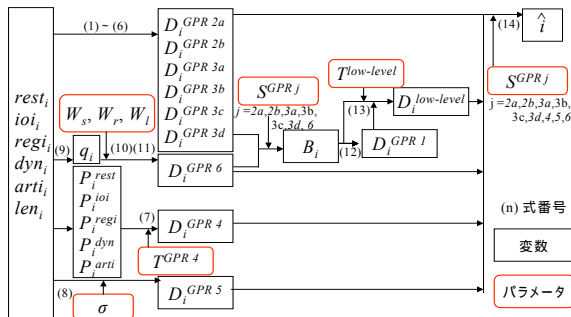


図 4: GPR の適用とパラメータの関係

- **基本変数の算出.** MusicXML から 6 つの基本変数を算出する。6 つの変数はそれぞれ, 消音時刻から次の発音時刻までの間隔 $rest_i$, 発音時刻間隔 ioi_i , 音高の差 $regi_i$, ダイナミクスの差 dyn_i , 楽譜上の音符の長さとして実際に演奏された音の長さの比の差 $arti_i^*$, 音価の差 len_i である。各基本変数について添え字の i が何番目の遷移であるかを表している。ただし, 第 1 音から第 2 音への変化を遷移 1, 第 2 音から第 3 音への変化を遷移 2 とする。

* GTTM では, アーティキュレーションの意味について明確な定義がない。ここでは, 文献[1]での適用例から, 楽譜上の音符の長さとして実際に演奏された音の長さの比という意味であると判断した。

- **GPR2,3,4 の適用.** $D_i^{GPR j}(j=2a, 2b, 3a, 3b, 3c, 3d, 4)$ は, GPR のルールが成立する度合いを表す変数で, 遷移 i でルールが成立する場合には 1, 成立しない場合には 0 を示す関数である。そして, GPR2,3 の 6 つのルールは各基本変数の変化の多いところで成立する (式 1-6)。また, GPR4 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。 $P_i^{rest}, P_i^{ioi}, P_i^{regi}, P_i^{dyn}, P_i^{arti}$ は, それぞれ, GPR2a, 2b, 3a, 3b, 3c, 3d が明白に成立するほど大きな値を示す変数で, $T^{GPR4} (0 \leq T^{GPR4} \leq 1)$ 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 (式 7)。

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

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

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

$$D_i^{GPR3b} = \begin{cases} 1 & dyn_{i-1} = 0 \text{ and } dyn_i \neq 0 \text{ and } dyn_{i+1} = 0 \\ 0 & dyn_{i-1} \neq 0 \text{ or } dyn_i = 0 \text{ or } dyn_{i+1} \neq 0 \end{cases} \quad (4)$$

$$D_i^{GPR3c} = \begin{cases} 1 & arti_{i-1} = 0 \text{ and } arti_i \neq 0 \text{ and } arti_{i+1} = 0 \\ 0 & arti_{i-1} \neq 0 \text{ or } arti_i = 0 \text{ or } arti_{i+1} \neq 0 \end{cases} \quad (5)$$

$$D_i^{GPR3d} = \begin{cases} 1 & len_{i-1} = 0 \text{ and } len_i \neq 0 \text{ and } len_{i+1} = 0 \\ 0 & len_{i-1} \neq 0 \text{ or } len_i = 0 \text{ or } len_{i+1} \neq 0 \end{cases} \quad (6)$$

$$D_i^{GPR4} = \begin{cases} 1 & \max(P_i^{rest}, P_i^{ioi}, P_i^{regi}, P_i^{dyn}, P_i^{arti}) > T^{GPR4} \\ 0 & \text{else} \end{cases} \quad (7)$$

ただし,

$$P_i^{rest} = rest_i / \sum_{j=i-1}^{i+1} rest_j, \quad P_i^{ioi} = ioi_i / \sum_{j=i-1}^{i+1} ioi_j,$$

$$P_i^{regi} = \begin{cases} regi_i / \sum_{j=i-1}^{i+1} regi_j & \left(\sum_{j=i-1}^{i+1} regi_j > 0 \right) \\ 0 & \text{else} \end{cases},$$

$$P_i^{dyn} = dyn_i / \sum_{j=i-1}^{i+1} dyn_j, \quad P_i^{arti} = arti_i / \sum_{j=i-1}^{i+1} arti_j$$

$$(0 \leq P_i^{rest}, P_i^{ioi}, P_i^{regi}, P_i^{dyn}, P_i^{arti} \leq 1)$$

- **GPR5 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_i^{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換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_i^{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 (式 8)。

$$D_i^{GPR5} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\left(\sum_{j=start}^i ioi_j - \sum_{j=start}^{end} ioi_j / 2\right)^2}{2\sigma^2}} \quad (8)$$

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

- **GPR6 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_r, W_s, W_l を導入する ($0 \leq W_r, W_s, W_l \leq 1$)。 W_r 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を決めるパラメータ， W_s 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 (式 9)。

$$q_i = \left\lfloor \frac{\sum_{k=1}^i ioi_k}{div} \right\rfloor \quad \left(\begin{array}{l} [] \text{はガウス記号. } [n] \text{は} n \text{を} \\ \text{越えない最大の整数を表す.} \end{array} \right) \quad (9)$$

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_j+r 拍目までの区間と q_j 拍目から q_i+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 (式 10)。

$$G_{ijr}^{start} = \frac{z_{q_i} q_j r}{y_{q_i} q_j r} \times W_r' \times r W_l + \frac{y_{q_i} q_j r}{x_{q_i-r} + x_{q_j r}} \times W_r \times r W_l \quad (10)$$

$$G_{ijr}^{end} = \frac{z_{q_i-r} q_j-r r}{y_{q_i-r} q_j-r r} \times W_r' \times r W_l + \frac{y_{q_i-r} q_j-r r}{x_{q_i-r} + x_{q_j-r}} \times W_r \times r W_l$$

ただし，

$$W_r' = 1 - W_r, \quad W_s' = 1 - W_s, \quad W_l' = 1 + W_l$$

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

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

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

そして，平行な度合い D_i^{GPR6} はその両者の重み付け和によって求まる (式 11)。

$$D_i^{GPR6} = \sum_j \sum_r \begin{cases} G_{ijr}^{start} \times W_s' & m_{ij} = b \\ G_{ijr}^{end} \times W_s & m_{ij} = e \\ G_{ijr}^{start} \times W_s' + G_{ijr}^{end} \times W_s & m_{ij} = t \\ 0 & m_{ij} = s \end{cases} \quad (11)$$

ただし，

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

- **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 ($D_i^{GPR2} = 0$) 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． S^{GPRj} ($0 \leq S^{GPRj} \leq 1$) は各ルールの強さを決めるパラメータである．

$$D_i^{GPR1} = \begin{cases} 1 & B_{i-1} \leq B_i \text{ and } B_i \geq B_{i+1} \\ 0 & B_{i-1} > B_i \text{ or } B_i < B_{i+1} \end{cases} \quad (12)$$

ただし，

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

4.2 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^{GPR3b} , D_i^{GPR3c} , 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 ($D_i^{low-level} = 0$) を決めるための閾値である． B_i は，式(12)と同じである．

$$D_i^{low-level \ boundary} = \begin{cases} 1 & B_i > T^{low-level} \text{ and } D_i^{GPR1} = 1 \\ 0 & B_i \leq T^{low-level} \text{ or } D_i^{GPR1} \neq 1 \end{cases} \quad (13)$$

4.3 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．

$$\hat{i} = \underset{i}{\operatorname{argmax}} (D_i^{low-level \ boundary} \times \sum_j D_i^{GPRj} \times S^{GPRj}) \quad (14)$$

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

5. 拍節構造分析の実装

階層的な拍節構造は、局所的な拍点の強さの算出と、次の階層の拍節構造の選択を再起的に繰り返すことにより獲得する(図5)。入力はMusicXMLとGroupingXML, 出力はMetricalXMLである。

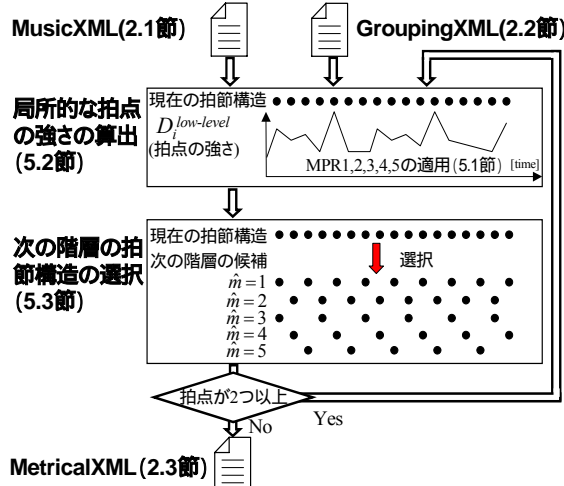


図5：拍節構造分析

5.1 拍節選択ルールの適用

図6は、MPRの適用とパラメータの関係を表したものである。 D_{ik}^{MPR1} および D_i^{MPRj} ($j=2, 3, 4, 5a, 5b, 5c, 5d, 5e$)は各MPRが成立する度合いを表す関数で、0から1の値で示される。 S^{MPRj} ($0 \leq S^{MPRj} \leq 1$)は、各ルールの強さを決めるパラメータである。

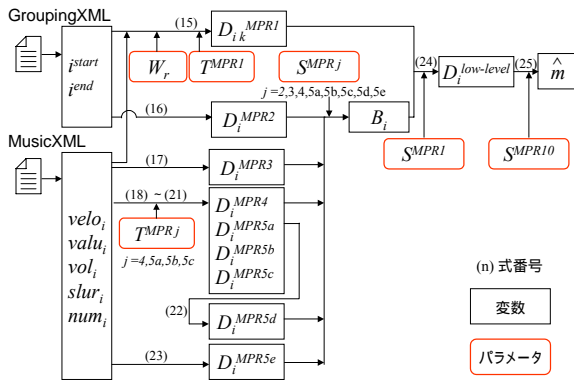


図6：MPRの適用とパラメータの関係

- **基本変数の算出.** MusicXML から5つの基本変数を算出する。5つの変数はそれぞれ、拍点から始まる音のベロシティ $velo_i$, 音価 $valu_i$, 連続する音量^{*}の長さ vol_i , スラーの長さ $slur_i$, 音高 num_i である。各基本変数について添え字の i は、現在の拍節レベルで何番目の拍点であるかを表している。5つの基本変数の平均値をそれぞれ μ_{velo} , μ_{valu} , μ_{vol} , μ_{slur} , μ_{num} とする。

^{*}本研究では、連続する音量の長さを、拍点から開始して次の拍点まで休符を挟まずに鳴り続けている長さとして定義した。

- **MPR1の適用.** MPR1では、並列的なグループは並列的な拍節構造を優先する。ここでは、調整可能な2つのパラメータ W_r, T^{MPR1} を導入する ($0 \leq W_r, T^{MPR1} \leq 1$)。 W_r 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、 T^{MPR1} は、拍点 i と k が並列的であるか ($D_{ik}^{MPR1}=1$)、そうでないか ($D_{ik}^{MPR1}=0$)を決める閾値である。 i^{start} と i^{end} を求めるためのグループは、拍点 i を含み、拍点が2つ以上ある最も小さなグループである。

$$D_{ik}^{MPR1} = \begin{cases} 1 & \frac{y_{ik} \times W_r + z_{ik} \times (1 - W_r)}{x_{ik} \times y_{ik}} > T^{MPR1} \\ 0 & \text{else} \end{cases} \quad (15)$$

ただし、

$$x_{ik} = \begin{cases} 1 & \text{if } velo_i > 0 \\ 0 & \text{if } velo_i = 0 \end{cases} + \begin{cases} 1 & \text{if } velo_k > 0 \\ 0 & \text{if } velo_k = 0 \end{cases}$$

$$y_{ik} = \begin{cases} 1 & \text{if } velo_i > 0 \text{ and } velo_{k+i} > 0 \\ 0 & \text{else} \end{cases}$$

$$z_{ik} = \begin{cases} 1 & \text{if } velo_i > 0 \text{ and } num_{i-1} = num_i \text{ and } num_{k+i-1} = num_{k+i} \\ 0 & \text{else} \end{cases}$$

i^{start} : グループの先頭の拍点の i

i^{end} : グループの終わりの拍点の i

- **MPR2の適用.** MPR2では、最も強い拍がグループの中で比較的早く出る拍節構造を優先する。 D_i^{MPR2} は、拍点がグループの先頭に近いほど大きな値を示す関数である。

$$D_i^{MPR2} = (i^{end} - i) / (i^{end} - i^{start}) \quad (16)$$

- **MPR3の適用.** MPR3では、拍点に音符がある拍節構造を優先する。

$$D_i^{MPR3} = \begin{cases} 1 & \text{if } velo_i > 0 \\ 0 & \text{if } velo_i = 0 \end{cases} \quad (17)$$

- **MPR4, 5a, 5b, 5cの適用.** MPR4では、強く弾いた拍が強拍となる拍節構造を優先する。

MPR5a, 5b, 5cでは、相対的に長い音、長い音量、長いスラーが強拍となる拍節構造を優先する。GTTMでは、具体的にどのような場合が強く、どのような場合が長いのが定義されていない。そこで本研究では、それらを判定するための閾値として、調節可能なパラメータ T^{MPRj} ($0 \leq T^{MPRj(j=4,5a,5b,5c)} \leq 1$)を導入する。

$$D_i^{MPR4} = \begin{cases} 1 & \text{if } velo_i > 2 \times \mu_{velo} \times T^{MPR4} \\ 0 & \text{if } velo_i \leq 2 \times \mu_{velo} \times T^{MPR4} \end{cases} \quad (18)$$

$$D_i^{MPR5a} = \begin{cases} 1 & \text{if } valu_i > 2 \times \mu_{valu} \times T^{MPR5a} \\ 0 & \text{if } valu_i \leq 2 \times \mu_{valu} \times T^{MPR5a} \end{cases} \quad (19)$$

$$D_i^{MPR5b} = \begin{cases} 1 & \text{if } vol_i > 2 \times \mu_{vol} \times T^{MPR5b} \\ 0 & \text{if } vol_i \leq 2 \times \mu_{vol} \times T^{MPR5b} \end{cases} \quad (20)$$

$$D_i^{MPR5c} = \begin{cases} 1 & \text{if } slur_i > 2 \times \mu_{slur} \times T^{MPR5c} \\ 0 & \text{if } slur_i \leq 2 \times \mu_{slur} \times T^{MPR5c} \end{cases} \quad (21)$$

- **MPR5dの適用.** MPR5dでは、相対的に長いアーティキュレーションパターンの繰り返しが強拍となる拍節構造を優先する。GTTMでは、アーティキュレーションパターンの繰り返しに関する具体的な定義がない。ここでは文献[1]での適用例から、MPR5aが連続で適用されている場所だと判断した。

$$D_i^{MPR5d} = \begin{cases} 1 & D_i^{MPR5a} = 1 \text{ and } D_{i+1}^{MPR5a} = 1 \\ 0 & D_i^{MPR5a} \neq 1 \text{ or } D_{i+1}^{MPR5a} \neq 1 \end{cases} \quad (22)$$

- **MPR5eの適用.** MPR5eでは、同一音高が連続している場合に強拍となる拍節構造を優先する。

$$D_i^{MPR5e} = \begin{cases} 1 & num_i = num_{i+1} \\ 0 & num_i \neq num_{i+1} \end{cases} \quad (23)$$

5.2 局所的な拍点の強さの算出

D_i^{MPRj} ($j=2, 3, 4, 5a, 5b, 5c, 5d, 5e$)の重み付け和により、各拍点の局所的な強さを算出する。その際MPR1を反映させる。

$$D_i^{low-level\ metrical} = B_i + \sum_k \begin{cases} B_k \times S^{MPR1} & D_{ik}^{MPR1} = 1 \\ 0 & D_{ik}^{MPR1} = 0 \end{cases} \quad (24)$$

ただし、

$$B_i = \sum_{j=(2,3,4,5a,5b,5c,5d,5e)} D_i^{MPRj} \times S^{MPRj}$$

5.3 階層的な拍節構造の獲得

局所的な拍の強さから、次の階層の拍節構造を獲得する。その際、MPR10を反映させる。次の階層の拍節構造 \hat{m} は、図5のように5通りが考えられる。現在の拍節レベルに拍点が2つ以上ある場合、次式によって次の階層の拍節構造が求まる。

$$\hat{m} = \underset{m=(1,2,3,4,5)}{\operatorname{argmax}} \sum_i \begin{cases} D_i^{low-level\ metrical} & (i-m) \bmod 2 = 0 \\ D_i^{low-level\ metrical} \times S^{MPR10} & (i-m) \bmod 3 = 2 \\ 0 & \text{else} \end{cases} \quad (25)$$

6. 実験結果

本節では、グルーピング構造分析と拍節構造分析の性能の評価を、適合率 P (precision)と再現率 R (recall)を組み合わせた F 値で評価する。 F 値は、適合率と再現率の値が高いほど、高くなる。グルーピングではグループが所属する階層に関係なく、システムの入力と正解データの両方に同じグループがあった場合を適合とした。一方、拍節構造分析では、システムの入力と正解データの同じ拍節レベルで同じ位置にある拍点を適合とした。

$$F_{\text{値}} = 2 \times \frac{P \times R}{P + R} \quad (26)$$

ただし、

P : 適合率 (正解データのグルーピング (拍点) と同じグループ (拍点) が、システムの入力に含まれている割合)

R : 再現率 (システムが入力したグルーピング (拍点) と同じグループ (拍点) が、正解データに含まれている割合)

このような F 値による評価を行うためにはグルーピング構造分析および拍節構造分析の正解データを用意する必要がある。本研究では、文献[17]でグルーピングの正解データに用いた、クラシック曲から切り出した8小節の長さの100個のメロディに対して拍節構造分析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作成したものを、3人のGTTMの専門家がクロスチェックしたものである。

グルーピング構造、拍節構造は、パラメータの調整によって変化する。パラメータの初期値は、 S^{GPRj} ($j=2a, 2b, 3a, 3d, 4, 5, 6$)= 0.5 , $\sigma=0.05$, $W_s=0.5$, $W_r=0.5$, $W_f=0.5$, $T^{GPR4}=0.5$, $T^{low-level}=0.5$, S^{MPRj} ($j=1, 2, 3, 4, 5a, 5b, 5c$)= 0.5 , および T^{MPRj} ($j=4, 5a, 5b, 5c$)= 0.5 である。そして、手作業で1曲につき10分間でパラメータの調整を行った。

- **GPR3b, 3cによるグルーピング性能の向上.** 文献[17]でのグルーピング性能(ベースライン)と本システムのパラメータ調整後のグルーピング性能を比較した(表1)。ダイナミクスに

表1: グルーピング構造分析・拍節構造分析の F 値

メロディ	グルーピング構造分析		拍節構造分析	
	ベースライン	GPR3b,3c追加後	ベースライン	パラメータ調整後
1. 紡ぎ歌	1.00	1.00	0.88	1.00
2. 小犬のワルツ	0.57	0.58	0.86	0.92
3. ソルヴェイグの歌	0.21	0.64	0.86	0.86
4. アルルの女	0.88	0.88	0.92	0.98
5. タランテラ	0.62	0.80	0.63	0.69
6. モルダウの流れ	0.33	0.63	0.73	1.00
7. 別れの曲	0.65	0.82	0.88	1.00
8. ダニェーブ川の漣	0.50	0.83	0.72	0.97
9. ホフマンの舟歌	0.11	0.54	0.72	0.79
10. ジムノペディ	0.31	0.68	0.66	0.82
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮
平均(100個のメロディ)	0.67	0.77	0.84	0.90

関するルール GPR3b とアーティキュレーションに関するルール GPR3c を追加したことにより、F 値が 0.10 向上した。

- パラメータ調節による拍節構造分析の性能向上。システムにグルーピングの正解データを入力し、パラメータ調節前の拍節構造分析の性能（ベースライン）とパラメータ調節後の性能を比較した(表 2)。その結果、パラメータの調節により F 値が 0.06 向上した。

本研究では、グルーピング構造および拍節構造を表示するため、GTTM ビューアを作成した(図 7)。GTTM 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、その下の棒グラフが拍節構造を表す。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 7 のタイムスパン木は、システムの出力するグルーピング構造分析と、拍節構造分析の結果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簡約の実装によって今後自動化できる予定である。

7. 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と拍節構造分析を実装した。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階層的なグルーピング・拍節構造の獲得を実現した。実験の結果、GPR3b,3c を適用することによりグルーピングの性能が F 値で 0.1 向上すること、パラメータの調節により拍節構造分析の性能が 0.06 向上することを確認した。

現在、GTTM の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。

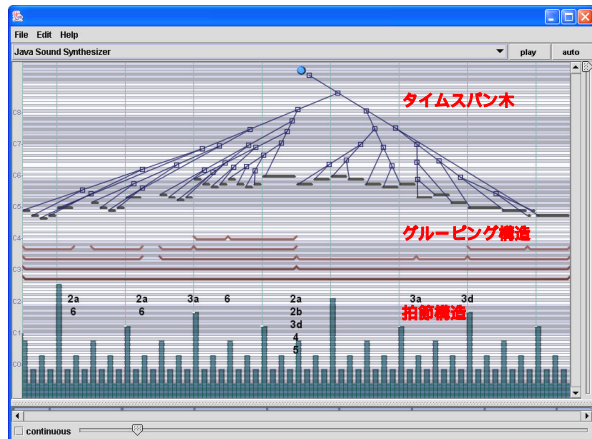


図 7: GTTM ビューア

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, パピプーーン: 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] D. Rosenthal. "Emulation of human rhythm perception." Computer Music Journal 16(1), 64-76, 1992.
- [12] D. Rosenthal. "Intelligent rithm tracking." In proceedings of the International Computer Music Conference, pp. 227-230, 1992.
- [13] Masataka Goto. "An Audio-based Real-time Beat Tracking System for Music With or Without Drum-sounds." Journal of New Music Research, 30:2, 159-171, 2001.
- [14] 後藤 真孝, 村岡 洋一, 音響信号を対象としたリアルタイムビートトラッキングシステム - コード変化検出による打楽器音を含まない音楽への対応 -, 電子情報通信学会論文誌, Vol.J81-D-II, No.2, pp.227-237, 1998.
- [15] 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.
- [16] 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.
- [17] 浜中雅俊, 平田圭二, 東条敏, 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, 情報処理学会研究報告 2004-MUS-55, pp. 1-8, 2004.
- [18] Recordare LLC. "MusicXML 1.0 Tutorial." <http://www.recordare.com/xml/musicxml-tutorial.pdf>, 2004.
- [19] W3C. "XML Pointer Language (XPointer)." <http://www.w3.org/TR/xptr/>, 2002.
- [20] W3C. "XML Linking Language (XLink) Version 1.0." <http://www.w3.org/TR/xlink/>, 2001.