

音楽理論に基づく映画の構造化

Structure of the film based on the music theory.

竹内 星子^{*1}
Takeuchi Seiko

浜中 雅俊^{*1*2}
Hamanaka Masatoshi

^{*1} 筑波大学 ^{*2} 科学技術振興機構さきがけ
University of Tsukuba PRESTO, Japan, Science and Technology Agency

This document describes structure of a movie that based on the Generative Theory of Tonal Music (GTTM). Various research of summaries of a movie have been developed, it has not easy to representation a relationship of cause and effect between scenes. We propose the method to represent structure of a movie by using Time-Span Tree of GTTM, and we design the rules for analyzing the structure of a movie.

1. はじめに

本稿では、映画を要約するための映画の構造化と、それを実現するための構造化ルールについて述べる。メディアの要約技術は現在多くの場面で求められるようになった。例えば、会議中の発言から重要な発言のみに要約することや、楽曲のサビ部分の抽出、録画映像の盛り上がった場面の検出などがある。これまでも数多くの手法が提案されてきたが[オン 2010, 出口 2004, 堀内 1997], これらは特定の要約の目的を達成するシステムの実現である。一方、要約は個人により目的が異なり、大体的な内容が理解できればよい、自分の定めた制約時間長にしたい、自分の関心の強い部分だけをまとめて視聴したいなど様々な目的があるが、その個人差に着目した研究はこれまでに十分されてきてはいない。そこで、各目的に限定して提案されてきたこれまでの研究を効率的に統合させる技術が必要となる。

これまでに我々は、同じ時系列メディアである楽曲を構造化した音楽理論 GTTM[Lerdahl 1983]に着目し、映画のストーリーにおける重要な場面と、各場面の従属関係をタイムスパン木で表現する方法を提案した。具体的にはタイムスパン木を編集することで個人の目的に合わせた柔軟な要約が可能となる。GTTM の特徴は、人間が楽曲に対して行っている分析をルールとして記述している点である。GTTM によって獲得した楽曲のタイムスパン木を用いることによって、楽曲の編曲や、繰り返し部分を省略する簡約が可能となる。さらに GTTM は人間が楽曲に対して行う共通の分析をルール化している。映画においても、各目的に対する人間の分析をルール化できれば、映画を構造化し、楽曲の簡約のように映画を要約することが期待できる。

本研究では、GTTM による楽曲の構造化を応用し、人間の映画に対する分析に近づけるために多段階のタイムスパン木を提案し、映画の要約映像を生成することを目指す。そこで本稿では、映画の構造化方法とそのためルールを設計し、ルールによる映画のタイムスパン木の獲得を試みる。

2. 音楽理論 GTTM による楽曲の構造化

音楽理論 GTTM は楽曲を分析するための複数のサブ理論から構成され、それぞれにルールが記述されている。ここでは、楽曲と映画に共通してみられ、映画の構造化でも用いる分析を行うグルーピング構造と、タイムスパン簡約について述べる。グルーピング構造、タイムスパン簡約には、音楽的観点から、必ず

満たすべきルールである構成ルール (Well Formedness Rules ; WFR) と、聴き手によって導き出される選好ルール (Preference Rules ; PR) がある。

(1) グルーピング構造

グルーピング構造とは楽曲をまとまりのあるグループに分割したものである。また、各グループの中でさらにまとまりのあるグループに分割することで階層的なグルーピングができる。以下に 2 種類の構造化ルールと、ルールの適用例を図 1 に示す。

<GWFR> (Grouping Well Formedness Rules.)

GWFR1 構成要素が連続する場合のみでグループを形成

GWFR2 一つの曲は一つのグループである

GWFR3 グループはより小さなグループを内部に含んでもよい

GWFR4 グループはサブグループの一部だけを含めない

GWFR5 グループがサブグループを含むなら、グループ構造が交差しないサブグループ群によって内部を埋め尽くす

<GPR> (Grouping Preference Rules.)

GPR1 非常に小さいグループへの解析は避ける。特に単音をグループにすることは避ける

GPR2 4 つの音符 (n1, n2, n3, n4) が連続しているとする以下の条件のどれかが成立すれば、n2 と n3 の間グループの境界と認識される

a. 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及び n3 の終わりから n4 の始まりまでの時間間隔より長い

b. 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及び、n3 の始まりから n4 の始まりまでの時間間隔より長い

GPR3 4 つの音符が連続しているとき、以下の条件のどれかが成立すれば n2 と n3 の間グループの境界と認識される

a. n2-n3 間の音高さが n1-n2 間の音高さ及び n3-n4 間の音高さよりも大きい

b. n2-n3 間でダイナミクスの変化があり n1-n2 間、n3-n4 間でそれがない

c. n2-n3 間でアーティキュレーションパターンの変化があり、n1-n2 間、n3-n4 間ではそれがない

d. n2 と n3 が異なった音長を持ち、n1 と n2 もしくは n3 と n4 が同じ音長の場合

GPR4 GPR2, 3 で示される効果が比較的明白なところは大局的視点においてもグループの境界がそこで位置づけられる可能性が高い

GPR5 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にグルーピングすることを優先する

GPR6 グループ間で並行した部分を形成することができる2つもしくはそれ以上のグルーピングは、並行性のあるグルーピングを行う



図1.グルーピングルール適用

(2) タイムスパン簡約

ある曲を編曲した時、人間はそれらの曲から共通する何らかの音楽的要素を認知する。例えば、ある曲 A とそれを編曲して作った曲 B に対して共通してみられるメロディや、リズムがそうである。簡約とは、音の高さや長さ、音程の変化などの表層的な構造から、認知した音楽的要素のような抽象的な構造を抽出することを指す。また、GTTM では、「人間は一般化した構造をもって楽譜の中のすべての音を関連付け、各音は相対的な重要度という尺度で階層化される」という簡約仮説を提唱している [Lerdahl 1983]。それを木構造で表現したものがタイムスパン木であり、重要な音が幹となるような二分木となる

タイムスパン木を獲得するためのルールとして、グルーピングルールと同様に、獲得したタイムスパン木による楽曲の簡約を図2に示す。楽曲の簡約方法は、図2の level1, level2 のように、タイムスパン木の深さを設定し、深さととの交点より深い枝を省略し、簡約メロディ1, 2を得る。



図2.タイムスパン木を用いた楽曲の簡約

3. 音楽理論 GTTM に基づく映画の構造化

音楽理論 GTTM の楽曲を分析する手順は、人間が映画を観るときに自然と行っている分析と共通する部分が存在する。人間が、映画や楽曲に対して無意識に行っていると考えられる共通の分析を以下に示す。我々はこのような共通点から GTTM の映画への応用が可能であると考えた。

<映画における分析>

- (1) 短い映像をつなぎ合わせて意味を見出す
- (2) ストーリー理解のために場面同士を関連付ける

<楽曲における分析>

- (1) 音を意味のあるまとまりにグルーピングする
- (2) 音同士の従属関係を分析

音楽理論 GTTM による楽曲の構造化において、映画の分析と共通するルールを応用し、タイムスパン木を獲得する方法について述べる。共通している分析である以上の2点に関して、楽曲の音符や、フレーズなどに対応する映画の構成要素を表1に、映画の構成要素を用いて分析する。各構成要素に対して、以下のような構造化ルールを設計し、獲得したタイムスパン木を図3に示す。

(1) グルーピング構造

<GWFR> (Grouping Well Formedness Rules.)

- GWFR1 連続するショットの集まりをシーンとする
- GWFR2 連続するシーンの集まりをシークエンスとする

表1 映画と楽曲の構成要素

映画	楽曲
ショット:切れ目のない一連の映像	音符:音を表す
シーン:関連のある連続したショットの集合	フレーズ:いくつかの音符から成る階層的まとまり
シークエンス:作品を構成する一つの要素であるシーンの集合	メロディ:音楽的内容を持つ複数のフレーズの集合

- GWFR3 シーンはショットを1つ以上含むこととする
- GWFR4 連続している構成要素のみがグループを形成できる
- GWFR5 シークエンスはシーンを2つ以上含むこととする
- GWFR6 1つの作品は1つのシークエンスである
- GWFR7 シークエンスは序破急の3部構成または起承転結の4部構成になることが望ましい

<GPR> (Grouping Preference Rules.)

GPR1 以下が成立する場合、グループである可能性が高い

- a. BGMが継続している
- b. 台詞が継続している
- c. 人物が継続して登場している
- d. 場所の移動がない
- e. 大幅な時間経過がない
- f. 短いショットが連続する
- g. 類似しているショットが連続する

(2) タイムスパン簡約

<TSRWFR> (Time-Span Reduction Well Fomedness Rules.)

- TSRWFR1 タイムスパンは内部に最も重要な場面を持つ
- TSRWFR2 幹は枝よりも構造的である
- TSRWFR3 シーンとショットの2段階に分けられる
- TSRWFR4 重要である枝はより浅い位置に接続する

<TSRPR> (Time-Span Reduction Preference Rules.)

- TSRPR1 より重要度の高い場面が優先的に幹となる
- TSRPR2 導入シーンは重要である
- TSRPR3 ショット長が長いほうが重要である
- TSRPR4 シーンの開始部分は重要である
- TSRPR5 シーンの開始部より終止部が
- TSRPR6 並行的な部分は並行したヘッドとなる

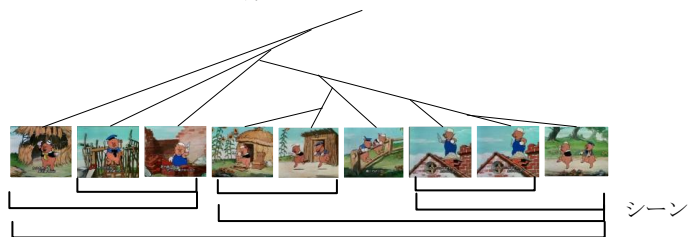


図3.タイムスパン木を用いた楽曲の簡約

4. 映画に特化した構造化ルールとタイムスパン木

音楽理論 GTTM に基づくルールでは、楽曲とは異なる映画独特の分析には対応していないため、映画を創るうえで共通に用いられる撮影技法[アリホン 1980] や、演出などを参考にルールの設計を行った。さらに本研究では、映画に対する人間の分析を再現することと、制約時間を満たすために場面選択の細かい調節を可能にすることを目的とし、タイムスパン木をショットとシークエンスの2段階で獲得する方法を提案した。

4.1 映画特有の構造化ルールの設計

新たに追加したルールとその設計方法を以下に述べる。映画の構造化ルールの設計では、[竹内 2014]で、要約の割合

に対する内容の理解度が高かった主観による構造化を参考にした。GTTM を応用した構造化ルールと、新たに設計したルールを用いて獲得したタイムスパン木と要約例を図 4 の示す。

<GPR>

GPR1 会話中の質問と回答で一つのグループが作成できる

GPR2 特定の人物が交互に現れる

GPR3 特定の場所が交互に現れる

ドキュメンタリーなど特定のジャンルを除くと、多くの映画では、会話によって物語が進行する。会話は質問と回答の連続であり、局所的に分析すると、その一回のやり取りは一つのグループになると考えられる。また特定の物体が交互に現れる場合、両物体に共通して起こっている事象があると考えられる。それが続く間は同一のシーンと分析する。

GPR4 以下の撮影技法が見られるとき境界がある場合が高い

- a. インサート、空抜け
- b. フェード・イン(始), フェード・アウト(終), ワイプ
- c. オーバーラップ(ディゾルブ)
- d. 黒コマ, 白コマのインサート
- e. ジャンプカット
- f. ロングショット

撮影技法は特定の情報を伝達する効果がある。例えば、a) インサート、空抜けには、時間経過、移動時間、動作やアクションの省略を視聴者に意識させる効果、b) フェードイン・アウト、ワイプには、場面転換を象徴する効果がある。具体的には、これらの撮影技法が見られた場合、自然言語学でいう「そして」や、「しかし」などの句読点や接続詞と認識することができる。したがって特定の撮影技法の出現はシーンの境界の出現と深いつながりがあると考えてよい。

GPR5 以下のショットが出現する場合、境界である場合が高い

- a. サブタイトル
- b. 曲調の変化
- c. 時間を表す小道具(ロウソク, たばこ, 暖炉, 時計, キャンプファイヤー, カレンダー, 日付の入った新聞の見出し)
- d. シーンの始まりに使われる小道具(ドア, ブラインド, 窓の日よけ, カーテン, 室内照明, 絵画, 写実的刺繍, 新聞に手ている写真)
- e. ひかりの変化(朝日と夕日)
- f. 言葉の繰り返し
- g. 完全な静止
- h. 動きのない映像
- i. 風景, 壁, 人気のない建物, 海
- j. 前後と全く関係のないショット, シーン
- k. 焦点のぼけた映像

映画中に出現する物体は何かを象徴する役割を持っているものが多く、それを活用することで、場面の境界を検出することが可能となる。視聴者は気づかないこともあるが、制作者は映像に移る一つ一つの物体を使ってあるメッセージを伝えようとする。例えば、c) の小道具は、時間によってその形状が変化することから、時間経過を暗示する。物体を記号として扱うことで、言葉による伝達とは違う効果が期待できる。また、視聴者は無意識にそれを認識し、分析している。

<TSRPR>

TSRPR1 登場時間の長い人物の場面は重要である

TSRPR2 特定人物の台詞が多い、又は長い場面は重要である

映画の中には主人公とされる人物が登場することがある。その場合、ストーリーは、主人公を中心に進行するため、主人公が登場する場面は重要である可能性が高くなる。しかし、多くの場

合主人公の定義はストーリー中に語られることはない。そこで、主人公を特定する基準の一つに、登場する時間と台詞の多さ、長さがある。登場時間や発言時間、回数が多いほど、視聴者はその人物に注目し、主人公であると認識する。

4.2 映画特有の 2 段階タイムスパン木

映画のタイムスパン木は、シークエンスのタイムスパン木とショットのタイムスパン木で構成される。映画では、時系列的に離れた場面同士に重要な因果関係が存在することがある。人間が内容を理解する為には、意味的な内容を持つ映像が必要である。数秒で構成されるショットのような短い映像のみでタイムスパン木を獲得することは、人間の分析とは異なり、困難な作業となる。

そこで我々は、シークエンスによって大局的構造を、ショットによって局所的構造を表す図 4 に示すようなタイムスパン木を採用する。人間は映画を観るとき、大局的な視点を持ってストーリーを分析していると我々は考え、シークエンスのタイムスパン木では、意味的な内容を持つレベルで要約を行い、ショットのタイムスパン木では、それだけでは意味を持たないような映像を用いて要約における細かな調節を行う方法を提案する。そして、映画のタイムスパン木から要約の度合いに対する内容の理解度が高い要約ができることを確認している[竹内 2014]。

映画の要約はシークエンスのタイムスパン木、ショットのタイムスパン木それぞれに、楽曲の簡約と同じ作業を行う。図 4 の level1 によってシークエンス 1 と 2 が選択される。Level2 よりシークエンス 1 内のショット①, ②, level3 よりシークエンス 2 内のショット⑤から⑧が選択され、つなぎ合わせたものを要約映像とする。

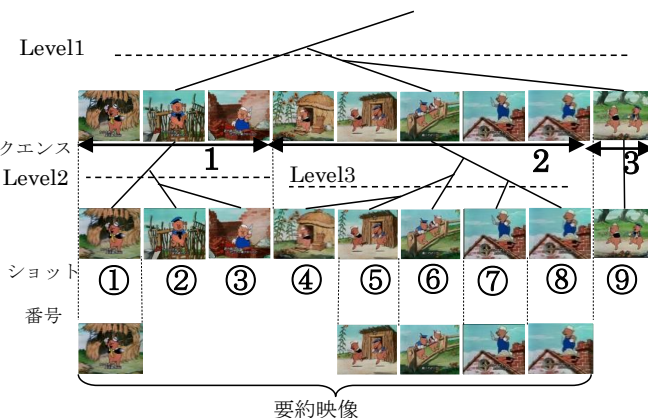


図 4.タイムスパン木を用いた映画の要約

5. 映画の要約システム

我々が開発した映画の要約システムは、エディタとして、映画の構造化における 1) グルーピング構造分析や、2) タイムスパン簡約分析を補助する役割を持つ。タイムスパン木の獲得には映画のショット情報:ショット番号, ショット時刻, ショット長が記述された xml ファイルがあればよい。xml ファイルを読み込むと、時系列順に並んだショットを、グルーピングし、シーンとシークエンス情報を xml ファイルに、シークエンス, ショットから枝を張り、タイムスパン木を作成し xml ファイルに保存できる。手書きでは負担となるタイムスパン木作成, 編集における枝の挿入, 削除, 置換などの作業も簡単に行える。要約の度合いをシークエンス, 各シークエンス内のショットに対して設定することができ、定めた要約の度合いによって選択されたショットが表示される。ショット分割された動画ファイルがあれば、ショット, シーン, シークエンスごとに映像を視聴することや、要約映像の出力も可能である。

ユーザによって要約の目的別の各 xml ファイルを共有することで、ユーザは自分の観たい映画の目的の要約が、xml ファイルの読み込みだけで可能になる。

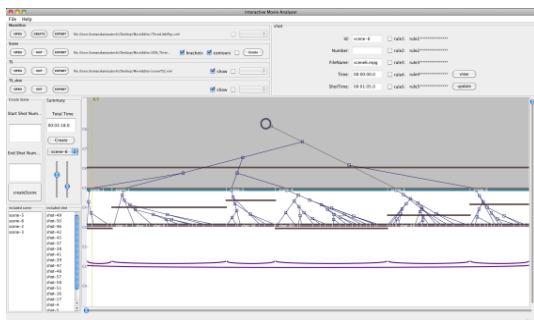


図 4.映画の要約システム

6. 実験

設計したルールの有効性を評価するために、ルールによる映画の構造化と、被験者が行う映画に対する分析の比較実験を行った。対象作品は約 8 分の短編映画である「3 匹の子ぶた」(ウォルト・ディズニー, 1933 年)とした。20 代の男性 5 人と女性 2 人の被験者の主観によって分析し、獲得したタイムスパン木と、ルールによって分析し、獲得したタイムスパン木を比較した。

まず、ショット数 62 の対象作品を視聴する。内容が理解できた時点で、ショット①から⑮までをシーンとシークエンスの階層にグルーピング、ショットとシークエンスそれぞれに対して、タイムスパン構造分析を行い、タイムスパン木を獲得する。そしてあらかじめルールによって分析し獲得したタイムスパン木のグループ構造とタイムスパン構造を比較する。

実験の結果、主観による分析と、ルールによる分析の共通点から、いくつかの有効なルールを確認することができた。主観によって獲得した各グルーピング構造、タイムスパン木で、完全に一致するものはなかったが、シークエンスでは 6 個、シーンでは境界が 8 個検出された。また、7 人の被験者のうちの検出人数の割合は表 2, 3 となった。シークエンス、シーン境界である 3-4 やシーン境界である 6-7 などの検出割合の高さから被験者の分析に共通する部分があることが分かった。

表 2. 検出されたシークエンス境界とその検出割合 (%)

境界	3-4	6-7	8-9	10-11	12-13	13-14
割合	100	85.7	14.2	28.5	57.1	28.5

表 3. 検出されたシーン境界とその検出割合 (%)

境界	3-4	4-5	5-6	6-7	8-9	9-10	10-11	12-13	13-14
割合	100	14.2	14.2	100	71.4	28.5	57.1	71.4	71.4

<グルーピング構造分析>

15 のショットのうち、ショット①から③, ショット④から⑥はルールによる分析、主観による分析の両方で共通の分析をしている。ここで用いたルールは GTTM を応用した GPR1g, 新たに設計した GPR5a であることから、上記の分析に使われたルールは被験者の分析を再現できているといえる。

また、主観による分析で 6 人中 3 人が検出したショット⑮とショット⑯のシークエンス境界は、ルールによる分析ではシークエンスとして検出できていないが、シーン階層では検出できている。このことから、シーンの作成からシークエンスを作成するまでの手順に個人差が問題であり、この部分にも新たにルールの設計が必要である可能性が高い。

次に、グルーピングによって得られたシーンの数を見ると、主観による分析よりもルールによる分析の方がシーンの数多く、人間の分析をルール数で補おうとしたために、分析が複雑化したと考えられる。この点に関しては被験者の分析を調査し、必要なルールを厳選する、またはルールの取捨選択基準を設定するなどの改善が必要である。

<タイムスパン構造分析>

シークエンスにおいては、ルールによる分析で得たタイムスパン木から枝を、削除、置換、挿入が最大 2 回の作業で、一種類のタイムスパン木に変形できた。このことから、木の形に大きな差が無いことが分かった。しかし、ショットにおいては、主観による分析で得たタイムスパン木同士を比較しても共通する分析が確認できないため、ルールによる分析で得たタイムスパン木が被験者の分析を再現しているかの検証はできていない。ショットのような短い映像に対する被験者の分析を厳密に調査する必要がある。

7. まとめ

音楽理論 GTTM に基づく楽曲の構造化を応用して、映画の分析に合わせた構造化ルールとタイムスパン木を提案した。映画と楽曲における、類似した分析は GTTM に記述されたルールを参考に、映画特有の分析においては、新たなルールの設計を行った。実験の結果、主観で作成したタイムス

パン木とルールに従って作成したタイムスパン木はシークエンスのタイムスパン木ではルールによって被験者の分析に近い分析が行えることが確認できた。しかし、ショットのタイムスパン木においては、被験者の分析において共通した特徴が見られず、ルールによる分析が被験者の分析を再現できるか確認することができなかった。

現在は、シークエンスとショットの 2 段階のタイムスパン木を獲得しているのに対して、ルールは一通りとしている。実験において、被験者の、シークエンスとショットに対する分析が異なるということが確認できたため、今後、ルールを段階別に設計する必要がある。また、被験者の分析プロセスを細分化し、厳密なルールの設計を行っていく。

参考文献

- [Lerdahl 1983] Lerdahl, F. and Jackendoff, R: A Generative Theory of Tonal Music, the MIT Press, Cambridge, 1983.
- [ONG 2010] Kok-Meng ONG, 大野 雄也, 亀山 渉: 瞳孔径・視線と心拍情報を用いた映像要約方法とその評価, 電子情報通信学会論文誌 A, Vol. J93-A, NO.11, pp.697-707, 2010.
- [出口 2004] 出口 嘉紀, 吉孝 淳夫: 映画の文法に基づく要約映像の生成, データベース・システム研究報告 DBS-132, pp33-40, 2004.
- [堀内 1997] 堀内 直明, 上原 邦明: ストーリーの内容記述に基づく映像の検索と要約, 電子情報通信学会技術研究報告 DE, データ工学 97(161), pp73-78, 1997.
- [栗原 2012] 栗原 一貴, 佐々木 洋子, 緒方 淳, 後藤 真孝: 音声区間自動検出技術を用いた変則再生方式による映像の高速鑑賞システムの検討, 情報処理学会研究報告 Vol.2012-HCI-149, NO.13, 2012
- [竹内 2014] 竹内 星子, 浜中 雅俊: 音楽理論に基づく映画の要約映像生成手法, 情報処理学会全国大会 2014 発表予定
- [Arijon 1980] Daniel Arijon: GRAMMAR OF THE FILM LANGUAGE, 1976.