

# 音楽理論 GTTM に基づく映画の構造化手法

竹内 星子<sup>†1</sup> 浜中 雅俊<sup>†2</sup> 星野 准一<sup>†1</sup>

本稿では、音楽理論 Generative Theory of Tonal Music ( GTTM ) に基づく映画要約システムを提案する。本システムでは、人間の映画に対する認知的な分析を、場面のグルーピングと映画全体の木構造 (タイムスパン木) 化によってルール化した。獲得した映画のタイムスパン木によって映画を要約する手法を提案し、評価実験により内容の理解に重要な場面を保存しながら要約映像が作成できるか検証した。

## Method of structuring a film based on the Generative Theory of Tonal Music

SEIKO TAKEUCHI<sup>†1</sup> MASATOSHI HAMANAKA<sup>†2</sup>  
JUNICHI HOSHINO<sup>†1</sup>

In this paper, we propose a movie summarization system based on the Generative Theory of Tonal Music ( GTTM ). In this system, we proposed a rule which reproduction human's cognitive analysis by grouping scenes and get a tree structure ( time-span tree ) of the entire movie. We propose a method to summarize the movie by time-span tree of the movie, and verified this system can be created a summary of movie while preserving the important scene to the understanding of the contents.

### 1. はじめに

近年、人々は気軽に WEB 上から音楽や映像などのメディアデータをダウンロードできるようになった。また HDD の大容量化によって、大量のメディアデータが保持できるようになり、それらを消費するための効率的な視聴方法が求められている。例えば、視聴に時間のかかる映画や楽曲などの時系列メディアデータを、もとの時間長よりも短い時間長で内容を理解できれば、メディアデータ選択の効率化が期待できる。この中の一つに映画の要約が挙げられる。出口ら[12]は映画を制作する上で共通して用いられる撮影技法や編集技法の規則集を用いて、重要な場面を自動的に検出し、要約映像を生成する方法を提案した。しかし、自動認識できる編集技法に焦点をあてた分析であったため、映画の種類や雰囲気、物語内容の複雑さによっては、生成した要約映像を観た視聴者の内容理解に関する評価が下がるという問題があった。堀内ら[15]は、あらかじめ用意した物語の定石の記述とユーザによる内容記述を用いて、実際の映像をあてはめて場面同士の因果関係を構造化する手法を提案した。しかし専門的な知識を持たない視聴者にとって物語内容の記述という作業は負担が大きく正確性も低いという問題があった。

このように映画を含む時系列メディアデータの要約に関する研究では、人がメディアデータを視聴するときの目的に着目したさまざまなアプローチがあるが、映画の物語内容を要約するためには、視聴者が物語内容の理解におい

て、映像のどの部分を重要視しているかを明らかにする必要はある。そして映画の物語内容を要約する際に、次の条件を満たす必要があると考えた。

- 1) . 一般的な要約が可能: 特定の個人の関心でなく、多くの人にとって共通性の高い要約を実現
- 2) . 専門的な知識を要さない: 因果関係の分析などの高い読解力が必要とされず、単純なルール群によって体系的な分析を実現
- 3) . 人間の分析に近い: 鑑賞者の認知を考慮に入れることで、人間が行う分析に近い要約を実現

映画と同様に一般的な要約が困難とされている時系列メディアデータの一つである音楽分野においては、音楽理論 Generative Theory of Tonal Music ( 以下 GTTM ) [18] による楽曲の簡約がある。GTTM は楽曲に対する人間の認知的分析に着目し、それをルール化した。楽曲全体を一つの木構造 (以下タイムスパン木) で表し、人間が音楽的に重要と感じる音を優先的に残す楽曲の簡約を可能にした。GTTM の概念は、三浦ら[17]により、会議中の議事録を GTTM のように構造化し、議論タイムスパン木から重要な発言のみに要約できる手法など、ほかのメディアデータにも応用されている。音楽を対象として、一般的な要約が可能であり、単純なルール群の適用による分析が可能で、人の認知を利用しているという点で、映画の分析においても利用できると考えられる。また、タイムスパン木によって映画の要約の度合いを簡単に調節できるなどの利点も考えられる。

本稿では、人間の認知に基づく音楽理論 GTTM を用いて映画の構造化手法を提案する。我々は、人間が映画の物語内容を理解する過程で、映画の持つ音楽や映像などの要素

<sup>†1</sup> 筑波大学  
Tsukuba University.

<sup>†2</sup> 京都大学  
Kyoto University

に対して、どのような認知や分析を行っているかに着目し、上記の3つの条件を満たす構造化ルールを設計した。また構造化ルールによって獲得したタイムスパン木から要約映像を生成し、視聴者の内容理解度を評価した。実験の結果、物語内容に焦点を当てた映画の構造化によって、要約の度合いに対して理解度の高い映画の要約が可能であることを確認した。

本稿では、2章で映画に関する従来研究について、3章でGTTMを応用した映画の構造化方法について、4章でタイムスパン木の構造と、要約映像の生成方法について説明する。5章で映画の構造化を利用した要約システムの提案、6章で映画の構造化に基づく要約映像についての評価実験について述べる。

## 2. 関連研究

映画を構成する音楽や映像、文章などの芸術分野においてさまざまな研究が行われている。たとえば映画制作や映画の歴史や読み解き方について言及した映画理論、映画における物語の構造、映画を含む時系列メディアデータの要約に関する研究やメディアデータの構造化に関する研究がある。

### 2.1 映画に関する研究

映画は音楽、映像、物語、文章などの複数の芸術分野によって創造され、総合芸術とも言われている。映画に関する研究には、映画を構成する複数の芸術分野に対して、言語学[16]、精神分析学[4][6]、イデオロギー理論[4]などの分野から研究アプローチがとられてきた。

映画学に関する研究は、大きく分けて、制作分野と理論分野に分けられる。制作分野においては、撮影・編集技法、映画監督の役割、脚本の制作方法、映写方法などが含まれる。理論分野においては、映画理論、映画史、映画評論などが含まれる。

映画制作に関する研究では、「映画の文法」[11]がある。これは映画に関する基礎的な知識から、映画を読み解くために必要な、制作者側が用いる共通の規則をまとめたものである。

物語論では、ロシアのすべての魔法昔話の内容を31の機能分類によって構造化できること示した研究[3]、意味素の矛盾・反対・前提といった関係から出発して物語を深層のレベル、媒介のレベル、表層のレベルの重層構造ととらえ、その生成過程を明らかにした研究[6]がある。

### 2.2 要約に関する研究

映画の物語内容に焦点を当てた要約を目指す研究には、出口ら[12]の、映画の文法を用いて編集上強調された場面を検出したものや、栗原ら[8]の、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がある。

これらの手法では、映画に対して、個人的な嗜好が反映されない一般的な要約映像の生成を目的としている。しかし、これらの方法では、要約の自動化を重要視しているため、物語内容に対する人間の認知的な分析が考慮されていないという問題があった。

映画以外のメディアデータの要約に関する研究には、議事録から重要な発言のみを検出するもの[13]や、GTTMに基づく楽曲の簡約[18]がある。三浦ら[17]はGTTMに基づく楽曲の構造化を応用して会議の議事録を構造化し、議論タイムスパン木を作成し、評価した。これらの手法には、要約の自動化ではなく人間の認知に焦点を当てた要約を行っているため人間が手作業で行う要約に近いという特徴がある。

### 2.3 メディアデータの構造化に関する研究

人間はある物事を理解するために、その物事を整理してまとめる作業を行うことがある。このような、理解に必要な部分がどこであるか、部分同士の因果関係の分析を、本稿では構造化と呼ぶ。音楽、文章、声や映像などのメディアデータの内容理解[6][14]、要約[8][12][13][17]、検索[15]、自動生成[5]を目指す研究分野において、構造化によるアプローチがある。

映画に関する構造化では、堀内ら[15]が、あらかじめ物語の定石が記述されたプロトタイプとユーザの内容記述を用いて、実際の映像をあてはめて、場面同士の因果関係を構造化した。しかし、専門的な知識を持たない視聴者にとって物語内容を記述する作業は負担が大きいという問題が考えられる。

## 3. GTTMに基づく映画の構造化

本稿では、音楽理論GTTMを応用した映画の構造化を行い、物語内容が理解できる、映画の要約手法について述べる。本章では、GTTMに基づく楽曲の簡約手順と、映画と音楽の持つ共通点からGTTMの映画への応用可能性について述べる。

### 3.1 音楽理論GTTMについて

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。楽曲を構造化するためには、旋律やリズム、和声などの高次の音楽的な構造を分析する必要があり、GTTMにはそれらを分析するためのルールが存在する。

GTTMの大きな特徴は、以下の2点である。

- 楽曲を構造化するためのルール
- 人間の楽曲に対する認知を表すタイムスパン木

GTTMは、楽曲の分析をルール化することで、計算機実装や、音楽知識の無い人でも楽曲を体系的に分析することができる。

GTTM では、人間が楽曲に対して行う認知は2分木で表現できるとしている。例えば、ある曲 A とそれを編曲して作った曲 B からは共通してみられるメロディや、リズムがあるが、GTTM ではそれを本質的な音とし、装飾的な音と区分するタイムスパン木を獲得する。

楽曲のタイムスパン木の獲得は大きく分けて以下の2段階の分析を行う。

- グルーピング構造
- タイムスパン簡約

グルーピング構造は音符を音楽的なまとまりのあるグループに分割したものである。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図 1(a)のような階層的なグルーピング構造を表現できる。

図 1 はタイムスパン木を用いた簡約の例である。図 1 の(a) をタイムスパン簡約した結果、上にある木構造が得られる。図 1 (b) の Level1 と Level2 のように、簡約の度合いである深さを設定し、Level1 より下にある枝の音符を省略すると簡約メロディ 1 が得られ、Level2 より下にある枝の音符を省略すると簡約メロディ 2 を得られる。

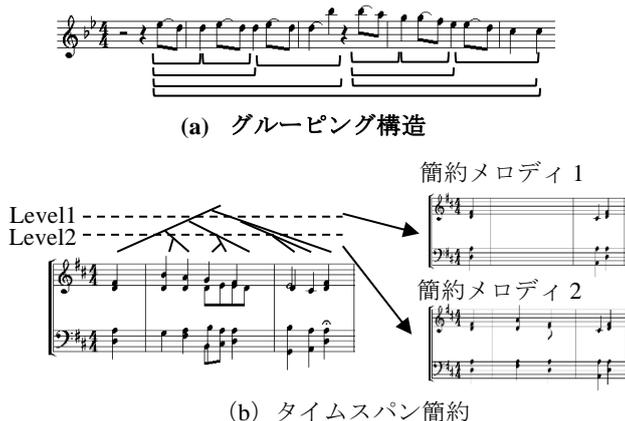


図 1 タイムスパン木に基づく楽曲の簡約  
 Figure 1 Reduction of music based on the time-span tree.

### 3.2 映画と楽曲の構成要素

GTTM を映画に応用する際、我々は表 1 のように、楽曲における音符、フレーズ、メロディを映画のショット、シーン、シークエンスに対応づけた。GTTM では、人間が、いくつかの連続した音符をフレーズというグループ、いくつかの連続したフレーズをメロディというグループと認知し、音楽的内容を階層的にグルーピングして聞き取っていると考える。映画においても、いくつかの連続したショットからシーンというグループを形成し、いくつかの連続したシーンからシークエンスを形成するといわれており [6]、その類似した階層構造の形成過程から GTTM におけるグルーピング構成ルールを応用する。

表 1. 映画と楽曲の構成要素

映画	楽曲
ショット：映画を構成する映像の最小単位	音符：音を表す最小単位
シーン：「場面」を指す、数ショットから成る場合が多い	フレーズ：数音符から成る階層的まとまり
シークエンス：数シーンから成る短いエピソードを指す	メロディ：数フレーズから成り、音楽的内容を持つ

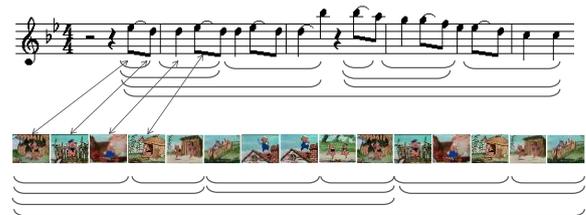


図 2 映画と楽曲の構成要素

Figure 2 Components of movies and music

### 3.3 物語内容が理解できる映画の構造化

本稿においての物語内容とは、E.M.フォースターの小説の諸相 [2] におけるプロットとする。フォースターはストーリーを時間の進行に従って事件や出来事を語ったもの、プロットを事件や出来事の因果関係に重点を置いて定義した。たとえば、ある物語において、「王様が市に、そして悲しみのために王妃が死んだ」がプロットで、「王様が死に、それから王妃が死んだ」がストーリーであるとした。また、フォースターは「プロットにおいては、行動も言葉もすべて重要な意味が無くてはならない」、「プロットは無駄で無駄のないもの」とし、物語内容を理解するために必要のない場面が存在しているとした。これらのことから我々は、プロットに関係する場面を本質的とし、それを補う場面を区分するタイムスパン木を作成する構造化ルールを設計する。

### 4. 映画の構造化ルール

GTTM に基づく楽曲の構造化を映画に応用するにあたり、グルーピング構造分析と、タイムスパン簡約の2段階に分け、それぞれの構造化ルールを提案する。映画の構造化ルールは、主に GTTM に基づいて設計した。しかし、映画のもつ画像情報、音楽情報、台詞情報など、さまざまな芸術要素に対しての分析を行うためには、映画特有の構造化ルールが必要である。

映画では、映像の撮影や編集の技法や、BGM の使い方、脚本によって、内容や人物の心情を暗喩的に表現することがある。このような映画を制作する上で、共通して用いられる規則をまとめた文献 [11] [16] がある。映画に対して、人間がどのように映画の内容を認識しているかを文献 [10]

によって検証し、ルールとして提案した。

以下に、それぞれの構造化ルールとその設計方法について述べる。

#### (1) グルーピング構造分析

楽曲におけるグルーピング構造は音楽的まとまりのあるグループに分割して得るが、映画においては、一連の意味的な内容を映像群にグルーピングする。映画のタイムスパイン木を作成する上で、必ず満たすべきルールであるグルーピング構成ルール (Grouping Well Formedness Rules. ; GWFR) と、視聴者によって導き出されるグルーピング選好ルール (Grouping Preference Rules ; GPR) を定義した。以下に 2 種類の構造化サブルールとルールの適用例を示す。

##### <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 シーンはショットを 1 つ以上含む
- GWFR3 構成要素が連続している場合のみグループを作成できる
- GWFR4 一つの作品は一つのシーンである

##### <GPR> (Grouping Preference Rules.)

- GPR1 BGM が連続している
- GPR2 台詞が連続している
- GPR3 人物が継続して登場している
- GPR4 場所の移動がない
- GPR5 大幅な時間経過がない
- GPR6 短いショットが連続する
- GPR7 質問と回答が成立する
- GPR8 類似したショットが連続して現れる
- GPR9 特定の人物が交互に現れる
- GPR10 特定の場所が交互に現れる
- GPR11 以下が見られる場合、その前後に境界がある可能性が高い
  - a. インサート、空抜け
  - b. フェードイン・アウト、ワイプ、トランジション効果
  - c. オーバーラップ
  - d. 黒コマ、白コマのインサート
  - e. 曲調が変化
  - f. サブタイトル
  - g. 小道具 (ロウソク、タバコ、暖炉、キャンプファイヤー、時計、カレンダー、ドア、ブラインド、窓の日よけ、カーテン、室内照明、絵画、新聞、朝日、夕日、風景、壁、人気のない建物、海)
  - h. 完全な静止
  - i. ジャンプカット
  - j. ロングショット
  - k. 動きのないショット
  - l. 焦点のぼけたショット

以下に、GPR の設計過程を説明する。

GPR1 : BGM は映像に印象を付加する役割を果たす。たとえば、別れなどの悲しい場面には落ち着いた雰囲気のある BGM を挿入する。BGM の雰囲気が変わると同時に映像の雰囲気の変化、つまり、場面が変わることを視聴者は認識する。

GPR2,3 : 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 [11]。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、人が話している途中で場面転換がされることはまれで、特定の人物が継続している場合には、場所の移動や、時間の経過が起こることが少ない。

GPR4 : 場所の移動がない場合、人は一連の場面と認識する可能性が高い。

GPR5,7 : 被験者実験による共通の分析である。

GPR6 : 人間は、提示されたある画像を理解するために、ある程度の時間が必要である。短いショットが並んでいる場合には、それらを連結させて意味を読み取ろうとする。

GPR8 : 視聴者が一連の場面を認識するためには、その中にさらに細かいグルーピングを行っていると考えられている [6]。たとえば疑問形の発言があったあとには、その答えである発言が続くと予測しながら視聴しており、発言に対する回答が確認できると一つのグループを作成する。

GPR9,10 : 特定の人物や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や、共通に流れる時間を描いている場合が考えられる。それが続く間は同一のシーンと分析する。

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。これらの撮影技法が見られた場合、自然言語学という「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 [11]。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。

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、制作者は映像に映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 [14]。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。

GPR11 : 映画の文法において、シーンの区切りに使われる。

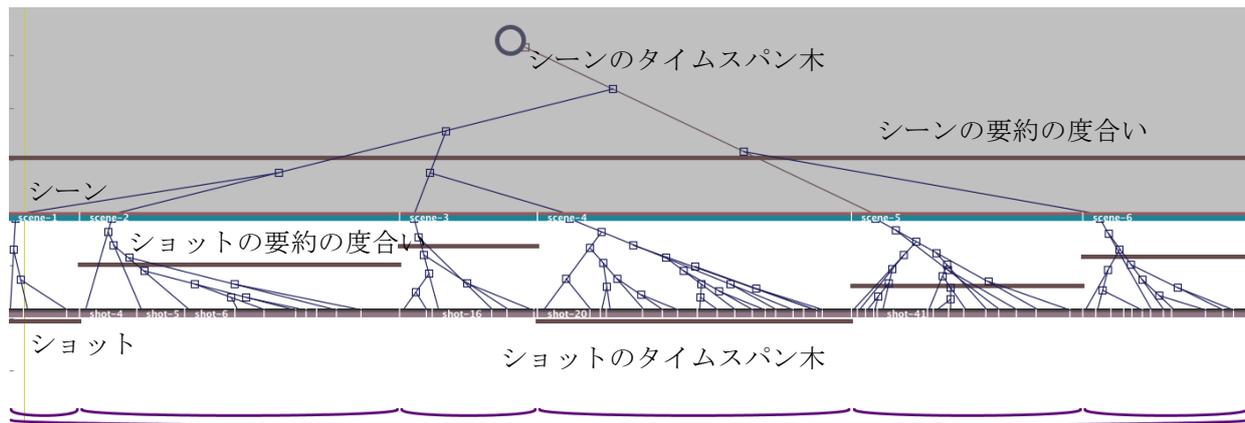


図 3 タイムスパン木に基づく映画の要約

Figure 3 Summary of the movie based on the time-span tree

## (2) タイムスパン簡約

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本質的な構造を抽出することを指す。タイムスパン簡約では、音楽的に本質的な音と、装飾的な音を区分するタイムスパン木を獲得する。

映画タイムスパン木を獲得するためのルールとして、グルーピングと同様に、タイムスパン簡約構成ルール (Time-Span Reduction Well Fomedness Rules ; TSRWFR) とタイムスパン簡約選好ルール (Time-Span Reduction Preference Rules ; TSRPR) がある。

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

TSRWFR1 より重要度の高い場面が優先的に幹となる

TSRWFR2 各シーンはそれぞれ最も重要なショットを持つ

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

TSRPR1 シーンの導入は重要である

TSRPR2 ショット長、シーン長が長いほうが重要である

TSRPR3 シーンの開始部分より終止部分の方が重要である

TSRPR4 登場時間の長い人物の場面は重要である

TSRPR5 情報量が多い場面は重要である

TSRPR6 BGM のサビ部分は重要である

TSRPR7 登場人物の性格や特徴がわかるシーン

TSPWFR8 全体の登場人物数における場面の登場人物数の割合が高い

以下に、各 TSPWFR についての設計過程を説明する。

TSPWFR1,3 : 認知的にそう分析されやすい[18].

TSPWFR2 : 含みを持つ、重要な意味を持つショットや、

シーンはそれだけ長く編集される可能性が高い。

TSPWFR4: 映画において主人公とされる人物が登場する場合、物語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物語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。

TSPWFR5 から TSPWFR8 : 映画のタイムスパン木獲得の実験[10]において、物語のあらすじを語る場面や、文字による説明、また、登場人物の説明は、情報量の多いという理由から、重要な場面選ばれた。また、BGM においては、グルーピング構造分析においても、着目されていたが、曲の盛り上がる部分は、映画の内容にも関連しているという共通の認識がある。

## 5. 映画の要約システム

本章では、3 章および 4 章で述べた映画の構造化手法に基づいて、映画を要約することができるシステムについて述べる。

### 5.1 映画の要約システムの構成

我々が提案する映画の要約システムでは、大局的構造と局所的構造を表す図 3 に示すようなタイムスパン木を獲得する方法[9]を採用している。人間が物語内容を理解するためには、意味的な内容を持つ映像群が必要である。例えばあるショットにおいて返事のみ発言があった場合、前後のショットの発言と統合することではじめて意味的な内容を確認できると考えられる。そのため数秒で構成されるショットのような短い映像のみの従属関係のみで作成された

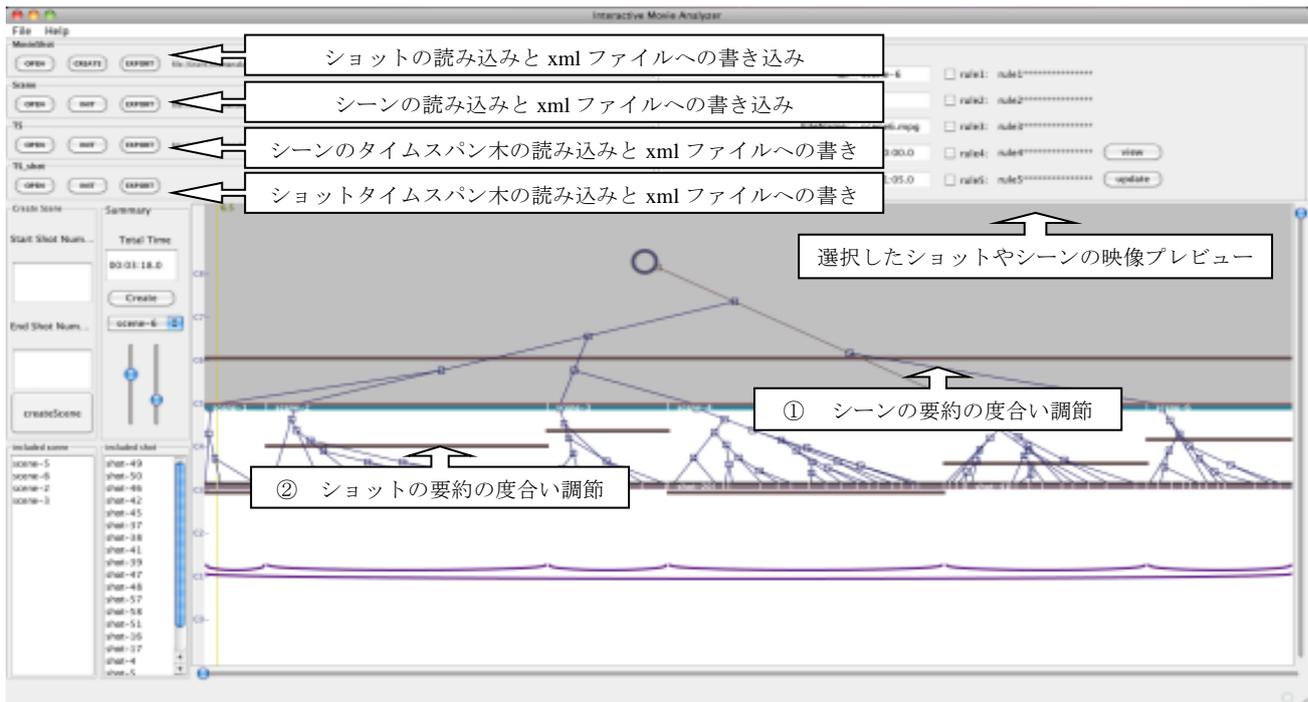


図4 映画の要約システム  
 Figure 4 Summarization system of movie

タイムスパン木では内容を理解することは難しい。また、人間は映像を観るとき、大局的な視点と局所的な視点を持って物語を分析しているとされている [14] ことから、タイムスパン木を映像の単位ごとに獲得する手法を提案する。

映画において映像の最小単位の映像とされているショットでは局所的構造を表すタイムスパン木を、関連する連続したショット群であるシーンでは、大局的構造を表すタイムスパン木を獲得する。シーンのタイムスパン木によって意味的な内容を持つレベルでの要約を行い、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短い映像を用いた細かい調節を行う。

映画の要約はシーンのタイムスパン木、ショットのタイムスパン木のそれぞれに、楽曲の簡約と同じ作業を行う。図4に示すように①シーンのタイムスパン木の要約の度合いを設定し、意味的な内容を保ったレベルでシーンの選択を行う。そして、選択された各シーン内の②ショットの要約の度合いを調節し、視聴者の求める要約映像の長さに収めることが可能になる。

## 5.2 映画の要約システムの構成

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グルーピング構造分析や、タイムスパン簡約を補助する役割を持つ。エディタを用いて操作可能なこととして以下がある。

- ① ショットの読み込み
- ② シーンの作成とシーンの読み込み

- ③ シーンタイムスパン木の作成とシーンタイムスパン木の読み込み
- ④ ショットタイムスパン木の作成とショットタイムスパン木の読み込み
- ⑤ 要約レベルの調節

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ファイルを用いる。xmlファイルを読み込むと、時系列順に並んだショットが表示される。それをグルーピングし、xmlファイルに書き込む。またショット、シーンそれぞれのタイムスパン木を作成しxmlファイルに書き込む。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、置換の作業も容易に行える。要約の度合いをシーンやショットそれぞれに対して設定し、要約映像を生成する。ショット分割された動画ファイルがあれば、ショット、シーンごとに映像を視聴確認することや、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、xmlファイルの読み込みのみで、ユーザは自分の観たい映画の目的の要約が可能になることも期待できる。

## 6. 評価実験

映画の構造化ルールに基づいて筆者が分析を行い獲得した映画のタイムスパン木から、物語内容が理解できるよ

うな要約が可能であるかを実験によって検証した。

### 6.1 実験概要

約8分の短編映画「三匹の子ぶた」(ウォルト・ディズニー, 1933年)から, タイムスパン木の深さ別に, 2分, 4分, 6分の要約映像を作成した。そして20代の男性5人と女性1人が, 3種類の要約映像と元の8分の映像を短い順に視聴し, 設問に回答した。本稿ではその得点によって理解度を評価した。以下に評価基準と具体例について説明する。

- (I) 表層的な内容の読み取りができていますか
- (II) 簡単な因果関係を理解しているか
- (III) 複雑な因果関係を理解しているか

(I) の設問の一つである「3匹の子ぶたがつくった家はそれぞれ何か?」は, 映像から何かを推測する能力は必要なく, 観たものをそのまま回答するだけでよい。(II) の設問の一つである「3匹の子ぶたが作った家ごとの子ぶたの性格は?」は, 視聴者の経験や読解力から, 映像上では名言されていない設定を読み取ることができているかを確認できる。(III) の設問の一つである「子ぶたが恐れている動物はなぜ2匹の子ぶたを追ってきたか?」は, 台詞での説明がないが, 直前の描写から推測することで回答できる。

### 6.2 実験結果

タイムスパン木から作成した要約映像において, 要約映

像が長くなるにつれて得点も増加することを確認した。図5は各評価基準の得点である。の(d)から, 8分の映像を半分(4分)に要約した結果, 総得点の平均は18点となり, 元の8分の映像を視聴した際の平均得点23.5点と比較し, 7割以上を理解できていた。

図5の(a), (b), (c)のグラフは, 設問の評価基準別の得点である。(a)について, 4分の要約映像ですでに平均の回答率は収束しており, 内容理解に必要なと考えられる表層的な内容は十分にされているといえる。(b)については, 表層的な映像から簡単な因果関係を読み取ることができていれば得点率は上がるが, 4分の映像でも平均得点は5割以下となった。(c)について, A, C, Dの被験者は, 2分の要約映像から8分の元の映像のすべてにおいて一定の得点となった。

### 6.3 考察

実験結果から(I)については, 内容に関わる場面を優先的に残す要約ができた。(II)に関しては, 要約の度合いに対して正答率が低い。このことからタイムスパン木によって場面同士の従属関係が適当に表現されていない可能性がある。現在の構造化ルールでは, 場面に対して重要であるかどうかの判断基準はあるが, 各場面がそれぞれ前後の場面のどちらに從属するかの判断基準が不十分であり,

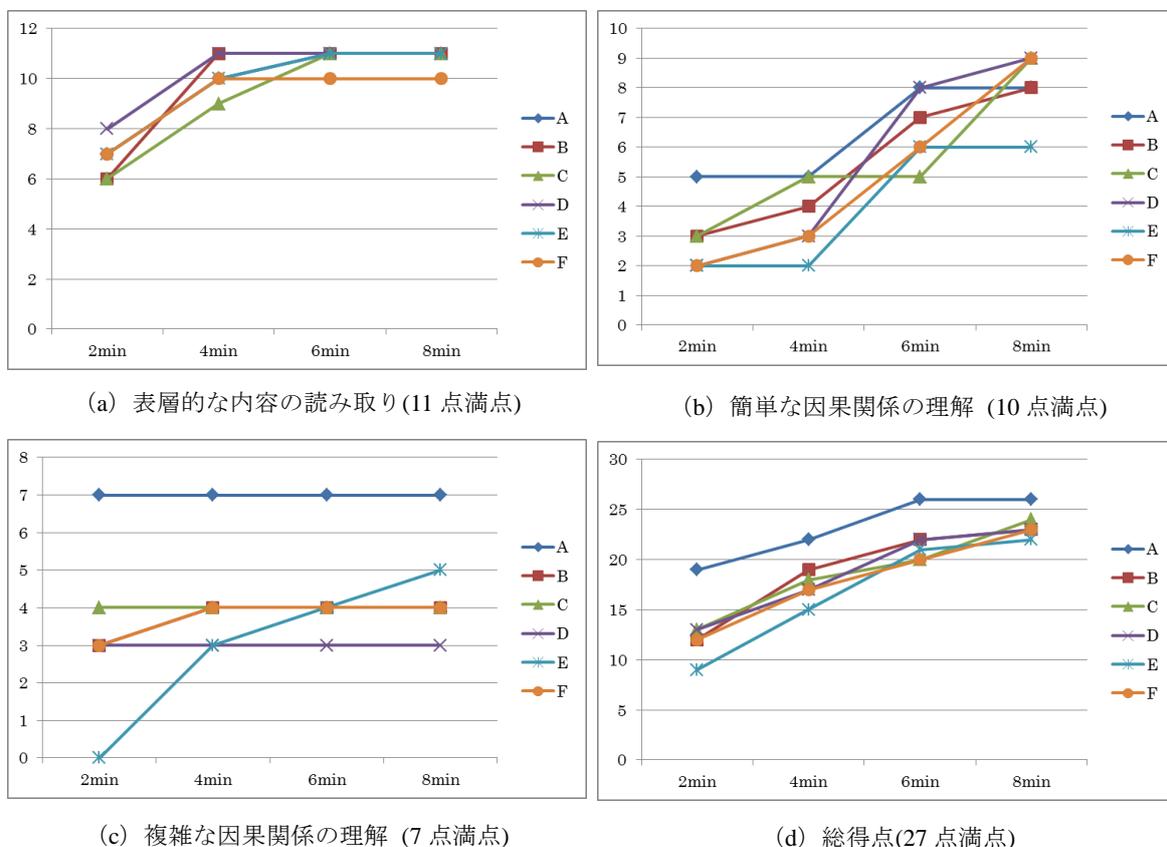


図5 実験結果

Figure 5 Experimental result

新たなルールの追加, 設計が必要である。(III)について, 図5より見ると他の評価基準と比較して(c)は, 個人差が大きいことがわかる。これは, 映像から得た情報から内容を読み取る能力や, 推察力には個人差があり, 読解能力に影響されやすい設問であったためと考えられる。しかし, A, C, Dの被験者は, すべての映像に対して, 正答数が一定であることから, 元の映像から読み取ることができる内容を2分の要約映像からも被験者たちは読み取っているといえる。

## 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の設計と映画のタイムスパン木を提案した。また我々は, この手法を用いた映画の要約システムによって, これまで難しいとされてきた, (1) 一般的な要約が可能, (2) 専門的な知識を要さない, (3) 人間の分析に近い, の3つを目的とした。以上の3つを解決するために, 映画に関する研究や, 実験によって確認した映画に対する人間の認知的な分析を基に構造化ルールを設計した。映画の物語内容に関する設問に回答する評価実験において, 要約の度合いに対して得点が高い要約が可能であると確認できた。

今後は, 本手法による映画の要約手法を, 様々なジャンルの映画に適用し, その有効性を検証する必要がある。また, ジャンルごとに必要となる新たな構造化ルールの追加も検討していく。

## 参考文献

- 1) Algirdas Julien Greimas, 田島 宏, 鳥居 正文, “構造意味論”, 紀伊國屋書店, 1988.
- 2) E.M.フォースター, 中野 健司, “小説の諸相”, みすず書房, 1994.
- 3) ウラジーミル.R.プロップ, 北岡 誠司, 福田 美智代, “昔話の形態学”, 1987.
- 4) 岩本 憲児, 波多野 哲朗(編集), “映画理論 集成”, フィルムアート社, 1982.
- 5) 小方 孝, 堀 浩一, 大須賀 節雄, “物語生成システムのための物語構造の分析と物語生成過程の検討”, 認知科学 Vol.3, No.1P1\_72-1\_109, 1996.
- 6) 川崎 恵里子, “知識の構造と文章理解”, 風間書房, 2008.
- 7) クリスチャ・メッツ(著), 鹿島 茂(翻訳), “映画と精神分析—想像的シニフィアン”, 白水叢書:白水社, 1981.
- 8) Kazutaka Kurihara. "CinemaGazer: a System for Watching Videos at Very High Speed.", In Proc. Of AVI 12., pp.108-115, 2012.
- 9) 竹内 星子, 浜中 雅俊, “音楽理論に基づく映画の構造化”, 人工知能学会全国大会 2014, 1K5-OS-07b-4, 2014.
- 10) 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 5ZD-7, 2014.
- 11) ダニエル アリホン, 岩本 憲児, 出口 丈人訳:映画の文法十作品にみる撮影と編集の技法, 紀伊国屋書店, 1980.
- 12) 出口 嘉紀, 吉高 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.
- 13) Nagao, Katashi, Kaji Katsuhiko, Yamamoto Daisuke, and Hironori

Tomobe, “Discussion Mining: Annotation-Based Knowledge Discovery from real World Activities.”, Proc. Of Pacific-Rim Conference on Multimedia ( PCM 2004 ), pp.522-531, 2004.

- 14) 中島 義明, “映像の心理学—マルチメディアの基礎—”, サイエンス社, 1996.
- 15) 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.
- 16) マルセル・マルタン(著), “映画言語”金子 敏 男(翻訳), みすず書房, 1957.
- 17) 三浦 寛也, 森 理美, 長尾 確, 平田 圭二, “音楽理論 GTTM に基づく議論タイムスパン木の生成方式とその評価”, 情報処理学会研究報告, Vol.2013-MUS-100, No.2, 2013.
- 18) Lerdah, F. and Jackendoff, R. “A Generative Theory of Tonal Music”, the MIT Press, Cambridge, 1983.
- 19) 浜中 雅俊, 平田 圭二, 東条 敏, “タイムスパン木獲得システムの完全自動化”, 情報処理学会 音楽情報科学研究会 研究報告, 2007-MUS—71-16, Vol.2007, No.81, pp93-98, 2007.