

# 標題音楽アノテーションのための階層的物語タグの設計

西山 正紘      北原 鉄朗      駒谷 和範      尾形 哲也      奥乃 博  
 京都大学大学院 情報学研究科 知能情報学専攻

## 1. はじめに

ドラマや映画のような映像コンテンツでは、音楽がBGMとして効果的に用いられ、その物語を大いに盛り上げる役割を果たしている。BGMとして用いられる音楽は、映像コンテンツ内で描写される情景に合う必要があるため、コンテンツ毎にプロによって新たに作曲される場合が多い。一方、初心者が映像コンテンツを制作する場合、ある情景に合う曲を作曲するための知識やノウハウを持っていないため、自身のコレクションや市販のBGM集から適当な音楽を選定する場合が多く、作曲する場合に比べて表現の自由度は低くなる。ここで、その知識やノウハウ、つまり情景を表現する音楽要素とは何かを理解されれば、それを利用してBGMの作曲・編曲や効率的な検索を支援することができ、初心者のコンテンツ制作に大いに貢献できると考える。

「情景を表現する音楽要素とは何か」という命題は、音楽理解の重要な側面に直結するにも関わらず、十分な検討はなされてこなかった。その1つの理由に、従来の研究アプローチではこの命題には立ち向かえなかったことがある。自動採譜や和声解析、音楽構造解析といったこれまでの音楽認識・理解研究は、音楽を分析し解釈するのが主であり、人間の音楽理解モデルの獲得を指向していなかった。後藤は「しろうとの音楽理解」という観点から「何が得られれば音楽を理解したといえるか」を問い直し、独自の問題設定で音楽理解研究に取り組んだ [1] が、上の命題に解を与えるまでには至っていない。

本研究では、物語を表現した音楽である標題音楽を対象とし、情景要素と音楽要素のマッピングモデルの獲得を目指す、新しい研究アプローチを提案する。対象を標題音楽に限定することで、作曲を物語から音楽へのマッピング、聴取・理解をその逆問題ととらえることができる。標題音楽では作曲者の意図が物語の形で入手可能なので、情景要素と音楽要素の関連をより客観的な形で分析・議論することが可能である。

## 2. 標題音楽を用いた音楽理解研究構想

標題音楽とは、物語を音楽で表現しようとする音楽のことである。一般に、物語はいくつかの場面（シーン）の流れによって構成され、各シーンではそれぞれの情景が描写される。また、シーンを意味内容を基にいくつかまとめたものを幕（アクト）と呼ぶ。つまり物語は図1のような階層構造を持つものと考えることができる。

物語を表現する標題音楽にもこのような階層構造を考えることができる。アクトやシーンに対応する構成要素を図1ではそれぞれ楽章、楽節と呼ぶ。しかし、この呼び方はあくまで便宜的なものであり、必ずしも音楽の持つ楽章、楽節の構造と一致するとは限らない。

シーン単位での物語と音楽の対応の例を図2に示す。このように標題音楽とは、各シーンで描写される情景の流れに沿って、その情景を音楽で表現したものである。

### 2.1 標題音楽の作曲・聴取過程のモデル化

本研究では、作曲家が標題音楽を作曲する際の過程を以下に述べるように仮定する。まず表現したい情景から、

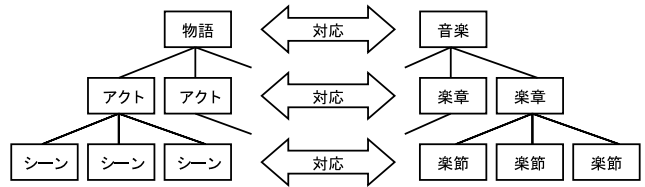


図1: 物語構造と標題音楽構造の対応。

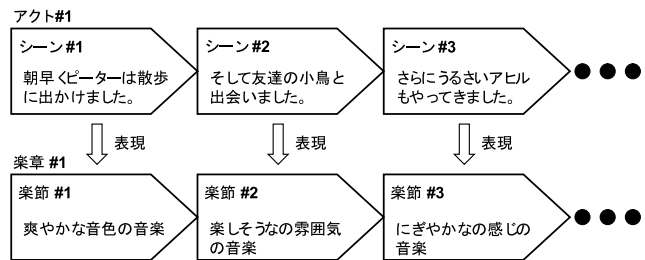


図2: シーン単位での物語と音楽の対応例。

その情景を特徴付けるような構成要素（情景要素）を抽出する。例えば「アヒルが狼に追いかけている」というような情景を音楽で表現したい時、主要人物はアヒルであるとか、その気分は不快であるとか、追いかける行動から動きのある情景であるというような情景要素を元の情景から抽出する。そして、これらの情景要素をどのような音楽要素でうまく表現できるかを考える。例えば、アヒルの鳴き声に似せてメロディの楽器はオーボエにしようとか、動きのある情景なのでテンポは早くしようという具合にである。この際に、作曲家は情景要素と音楽要素を結びつけるマッピングモデルを用いて、情景要素を音楽要素に変換していると考えられる。最後にその音楽要素を組み合わせて音楽を作曲する。このように、作曲家は図3で示されるような流れで情景を基に標題音楽を作曲する。

このとき、聴取理解の過程は、作曲の過程の逆問題として図3で示されるような流れで説明できる。まず聴取者は実際に演奏されている音楽を聴き、それから音楽要素を抽出する。そして音楽要素をマッピングモデルを用いて情景要素に変換する。最後に得られた情景要素を解釈して、音楽が描写している情景を理解するとする。

両者のマッピングモデルに共通項が多ければ、聴取者は音楽を聴いた時に作曲家が表現した情景を理解することができるが、少なければ、その情景を十分に理解することができない。マッピングモデルは、個々人にとってユニークなものであり、この共通項が普遍的な情景要素と音楽要素間のマッピング関係を表現し、異なる項は個人の個性を表現しているものと考えられる。

本研究ではこの共通項、すなわち普遍的な情景要素と音楽要素のマッピングモデルの獲得を目指す。

### 2.2 マッピングモデル獲得における技術的課題

前述のように、標題音楽の作曲では表現したい情景からそれを特徴付ける情景要素を抽出し、それを音楽要素によって表現する。情景音楽では、元となる情景は自然言語によって記述され、作曲の結果は楽譜によって記述されるので入手可能である。しかし、情景要素や音楽要

Hierarchical Narrative Tag Design for Program Music Annotation: Masahiro Nishiyama, Tetsuro Kitahara, Kazunori Komatani, Tetsuya Ogata, and Hiroshi G. Okuno (Kyoto Univ.)

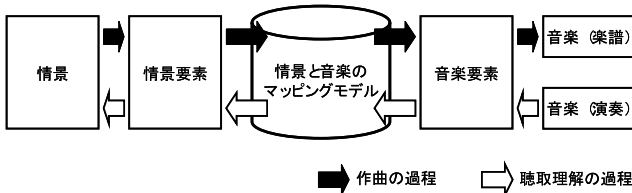


図 3: 標題音楽の作曲および聴取理解の過程。

素は明示されていない。例えば、前述の例では、アヒルの気分が不快であることや、動きのある情景であることは、情景を記述した文から解釈する必要がある。また、情景要素と音楽要素の関係を分析する上では、分析の候補を予め決めておく必要がある。以上の考察から、本研究における技術的課題は次の3つにまとめられる。

1. 情景要素の設計: 一般に情景は自然言語で表現されるので、本研究ではこれから情景要素を抽出する。具体的にどのような情景要素を設計するかについては次章で述べる。
2. 音楽要素の設計: 作曲家は音楽を楽譜という客観的な形で作曲するので、音楽要素は楽譜という客観的なレベルで規定されていると考えられる。客観的なレベルの音楽データとしては、標準 MIDI ファイル (SMF) で記述された音楽データが使用できる。よって、本研究では、SMF から音楽要素を抽出する。
3. 両要素の関係の分析: 本研究では、両要素の関係の分析を統計的手法に基づいて行う。現在、非説明変数を情景要素、説明変数を音楽要素とする判別分析を行うことで、ある情景要素を表現するために関係のある音楽要素の抽出の可能性を検討している。

### 3. 階層的物語タグの設計と音楽への付与

物語と情景音楽は、図1で示したようなそれぞれ対応する階層構造を持つ。それぞれの対応を明示するために、本研究では音楽データ (SMF) に階層的物語タグを付与する。ここで、階層的物語タグとは、図1の物語の階層構造で示されるような構造を表したものと定義され、物語、アクト、シーンの各階層に対応するタグをそれぞれ narrative タグ (ルート要素)、act タグ、scene タグとする。また scene タグは情景要素を子要素として持つものとする。表1に階層的物語タグを示す。

ここで、シーンで記述される情景要素をどのように定義するかが重要である。大石らは、物語からイメージされる情景の構成に必要な要素として、「登場人物」、「行動」、「感情」を挙げている [2]。これを参考に考えると、作曲者が情景を基に音楽を作曲する際には、主要人物があるライトモチーフで表現したり、その感情を調性で表現したりすることがあると考えられる。一方、行動をそのままの要素として音楽で表現することはなく、むしろ行動から受ける印象を音楽で表現しているように考えられる。例えば「川が激しく流れる」情景と「川が緩やかに流れる」情景では、行動要素はどちらも「流れる」である。しかし、これらの情景を音楽で表現する場合は、行動要素から受ける「激しさ」という印象を基に表現されると考察される。

本研究ではこの考察を基に、情景要素を表現するタグ (情景要素タグ) として「主要人物に関するもの (character タグ)」と「行動を含んだ情景全体の印象に関するもの (impression タグ)」の2つを定義する。それぞれの情景要素タグは表1に示されるような属性を持つものとする。

以上のように設計したタグを標題音楽の SMF に付与する。タグの付与の例として、標題音楽である交響的物

表 1: 階層的物語タグ。

タグ名	属性名	属性値集合	属性の意味
<narrative>	—	—	—
<act>	id	整数値	id番号
<scene>	id	整数値	id番号
	location	SMFのパス	パス
	onset	整数値	オンセット
	offset	整数値	オフセット
<character>	name	name1,name2...	主要人物名
	emotion	comfort,neutral,discomfort	主要人物の感情
<impression>	intensity	intensive,normal,moderate	情景の激しさ
	liveliness	lively,medium,quiet	情景のにぎやかさ
	velocity	fast,medium,slow	情景の速さ

<アクト#1、シーン #1>  
ある朝早く、ピーターは庭の木戸を開けて、森の広い草原へ散歩に出かけました。

タグ付け

```
<act id="01">
  <scene id="01" location="/01.mid" onset="0" offset="37740">
    <character name="peter" emotion="comfort" />
    <impression intensity="normal" liveliness="medium" velocity="medium" />
  </scene>
</act>
```

図 4: 情景文に基づいた SMF へのタグ付けの例。

語「ピーターと狼」に階層的物語タグを付与したものの一例を図4に示す。このように、タグ付与は情景を表す文章から情景要素を手動で抽出することで行う。

### 4. 考察

本研究の特徴、意義は「標題音楽を対象とし、音楽から情景を理解するという人間の音楽理解モデルの獲得を目指している」点にある。関連研究として、印象語と音楽要素との関連を分析した研究が存在するが、これらの研究では、音楽に対して印象語を人間が付与することにより評価が行われている [3]。それゆえに、ある曲に対して付与される印象語の個人差が大きいという問題点が存在する。これは、音楽から受ける印象が主観的なものであるためである。それに対して、標題音楽は音楽それ自体が物語を表現しているため、客観的に情景と音楽要素との関連を分析できると考えられる。

### 5. おわりに

本研究では、情景要素と音楽要素のマッピングモデルの獲得を目指し、階層的物語タグを設計した。今後、様々な標題音楽にタグを付与し、マッピングモデルの獲得のための分析を行う予定である。

謝辞: 本研究の一部は、日本学術振興会科学研究費補助金、21世紀 COE プログラムの支援を受けた。

### 参考文献

- [1] 後藤 真孝: "リアルタイム音楽情景記述システム: 全体構想と音高推定手法の拡張", 情報処理学会 音楽情報科学研究会 研究報告 2000-MUS-37-2, Vol.2000, No.94, pp.9-16, 2000.
- [2] 大石誠, 鬼沢武久: "物語テキストからイメージされるシーン構成に必要な要素抽出", 第 24 回ファジィ・ワークショップ講演論文集, pp.5-8, 2005.
- [3] 池添剛, 梶川嘉延, 野村康雄: "音楽感性空間を用いた感性語による音楽データベース検索システム", 情報処理学会論文誌, Vol.42, No.12, pp.3201-3212, 2001.