

映像作品におけるシーンの印象別背景音楽選定のための 音響特徴量分析

Acoustic Feature Analysis for Selection of Background Music for Conceptual Scene Type of Audio-visual Content

西 香織

Kaori Nishi

法政大学情報科学部デジタルメディア学科

E-mail:kaori.nishi.8m@cis.hosei.ac.jp

Abstract

Visual media like movie and television are not composed only visual. It almost has sounds. In this study, to analyze relation between audio-visual content and back ground music (BGM), it is examined that acoustic feature of BGM used each situations of the story. First, variety audio-visual content are prepared, and scenes of each of content are classified five categories: “pleasure”, “sad”, “happy”, “fear”, “excitement” based on impression. And, acoustic feature of each category is analyzed by analyzing tempo, volume, and frequency of BGM. As a result, BGM used “pleasure” scene and “sad” scene was that the tempo was slow, the volume was low, and the band less than 500Hz had low power. BGM used “happy” scene was that the tempo was fast, the volume was middle and the band almost was average. And BGM used “fear” scene and “excitement” scene was that the tempo was chose fast tempo and slow tempo by the volume of action, and the volume was high. So, each category had acoustic feature. But it has a problem that there is no difference between “pleasure” scene and “sad” scene.

1.まえがき

一般に、映画やテレビのような映像メディアは、映像だけでは成り立たない。必ず音を伴っている。映像に加えられる音は、映像に表現された対象から発する音ばかりではない。映像作品をより印象的なものにするために、特殊な効果音や音楽が用いられる。それらは映像の一部と言ってもいいほど、映像表現の重要な側面を担う[1]。

文献[2]では、映像のみの場合と音加わった場合の映像作品の印象の違いを印象評定による主観評価実験により明らかにしている。実験材料として、「シェルブールの雨傘」や「スノーマン」などのミュージカル映画や音楽ビデオ、アニメーションなどのオーディオ・ビジュアル素材を利用している。また、文献[3]では印象評定による主観評価実験を行うことで、音楽の調性及びテンポと映像の速度および密度の組み合わせに意味的調和の存在を明らかにしている。この研究では、3Dグラフィックソフトで映像の密度や再生速度を調節できるアニメーシ

ョンと、シンセブラス、ベース、ドラムの3つのパートからなり、曲のテンポや調性を調節できる曲を用いている。このように、映像作品は背景音楽によって印象を高められていることが明らかになっている。

本研究では、映像作品において意図的に視聴者の感情をコントロールするように作られたシーンに使われている背景音楽にどのような音響の特徴があるかを解析する。つまり、視聴者が受ける印象のためにどのように背景音楽が使い分けられているのかを解析する。そのために、映像作品に使われている背景音楽をもとに、ストーリーの状況別に使われる背景音楽にどのような音響の特徴があるのかを検討する。これによって、簡単な映像作品を作る場合に初心者でもストーリーの状況に合った背景音楽を使うことを可能にすることや、映像作品の背景音楽の自動生成を可能にすることを目指す。

2.映像作品の印象とシーン

映像作品を見て、人は楽しいと感じたり、悲しいと感じたり、恐怖を感じたりする。そこで本研究では、人が楽しいと思うようなシーンに使われる背景音楽にはどのような特徴があり、悲しいと思うようなシーンにはどのような背景音楽が使われているかなどを分析する。

手法としては、まず映像作品の全てのシーンを割り出し、そのシーンを印象によってカテゴリに分類する。映像作品の印象は人間の感情に基づいて分類する。

2.1. 映像作品におけるシーン

映像作品におけるシーンとは一定の場所の中での動作の一区切りのことを言うので、これにしたがってシーンを割り出す。しかし、今回はストーリーの状況別に使われている背景音楽について解析するので、一定の場所の中でもカテゴリ（印象）が変わったら、その時点で別のシーンとして割り出す。

2.2. シーンの種類

人間の感情の代表的なものには、「喜怒哀楽」というものがある。しかし、映像作品には様々な分野があり、恐怖感を感じるものや興奮するものもある。そこで、「喜怒哀楽」に加えて「恐怖」と「興奮」という感情における背景音楽の特徴も分析する。ただし、「喜怒哀楽」の「怒り」というものは映像作品を見る上では興奮している状態なので「興奮」に含まれるものとする。よって、カテゴリは[喜び], [悲しい], [楽しい], [恐怖],

[興奮]の5つとする。[楽しい]シーンは明るくてにぎわっているようなシーンとし、おもしろいシーンなども含む。[悲しい]シーンは良くない出来事が起こったりする暗いシーンとし、切ないシーンやさみしいシーンも含む。また、[喜び]のシーンは涙を誘うようなシーンで、視聴者が良かったと思えるようなシーンとする。[恐怖]のシーンは恐怖感で驚くようなシーンとし、[興奮]のシーンは恐怖ではなく次の展開が気になるようなシーンとする。

3.実験

3.1.実験方法

まず、様々な分野の映像作品（ヒューマンドラマ、サスペンス、ホラーなど）を実験材料として用意する。そして、実験材料の映像作品を視聴し、映像作品のシーンを前述の5つのカテゴリに分類する。シーンをカテゴリごとに分類したら、テンポ、音量、周波数について音響的特徴を解析する。

3.2.実験材料

今回用いた映像作品はドラマと映画を合わせて10種類である。CMやオープニングテーマ、エンディングテーマを除いた映像作品全体の時間はドラマが約40分、映画が約1時間半である。カテゴリごとのシーン数を表1に示す。

表1: カテゴリごとのシーン数

| 喜び | 悲しい | 楽しい | 恐怖 | 興奮 |
|----|-----|-----|----|----|
| 10 | 67 | 37 | 33 | 30 |

3.3.テンポによる音響的特徴の解析

3.3.1.実験方法

それぞれのシーンの背景音楽のテンポをbeat counter[4]を使って手動で検出する。そして、カテゴリごとに背景音楽のテンポの特徴を分析する。

3.3.2.実験結果

実験結果を図1に示す。横軸はテンポをbpm（1分間に四分音符を何回刻むか）で示し、縦軸はそのカテゴリにおけるサンプル数の比率をパーセントで示した。点線は平均値を示し、実線の矢印は95%信頼区間、鎖線はカテゴリ全体の平均テンポ（100bpm）を示している。リズムやテンポが一定ではなく、不安定な音楽は「不規則」なテンポとした。

3.3.3.考察

それぞれのカテゴリごとに平均、95%信頼区間を求めた。その際、不規則なテンポは無視する。95%信頼区間は以下の式で求めた。

$$95\% \text{ 信頼区間} = \text{標本平均} \pm t \times \text{標本標準偏差} \quad (1)$$

このとき「t」は信頼区間で決められた分布の面積が95%になるような数値で、サンプル数が増えるとtは小さくなり、サンプル数が減るとtは大きくなる。また、標本標準偏差は標本平均の標準偏差のことである。

また、テンポの速さを相対的に比較するために、カテゴリ全体の平均テンポを求めると、100bpmだった。これによって以下のことが明らかになった。

[喜び]のシーンはテンポが遅めの背景音楽が使われやすい可能性がある(平均73bpm, 信頼区間22bpm~124bpm)。図を見ると狭い区間にサンプルが固まっているにもか

わらず、信頼区間の幅が広がったのはサンプル数が少なかったためだろう。[悲しい]シーンはテンポが遅い背景音楽が使われやすい(平均74bpm, 信頼区間56bpm~93bpm)。[楽しい]シーンはテンポが速い背景音楽が使われやすい(平均133bpm, 信頼区間85bpm~181bpm)。[恐怖]のシーンはテンポが速い曲も遅い曲も使われる(平均112bpm, 信頼区間55bpm~169bpm)。[興奮]のシーンはテンポが速い曲も遅い曲も使われる(平均108bpm, 信頼区間69bpm~147bpm)。

[恐怖]のシーンと[興奮]のシーンの背景音楽はテンポにあまり特徴がないように思えるが、闘ったり追いかけられたりするようなアクションが多いシーンには100bpm以上のテンポが速い背景音楽が使われており、何者かがゆっくり迫ってくるようなアクションが少ないシーンには100bpm未満のテンポが遅い背景音楽が使われていた。よって、背景音楽は感情だけではなく、動きの演出も考慮して選ばれていると言える。

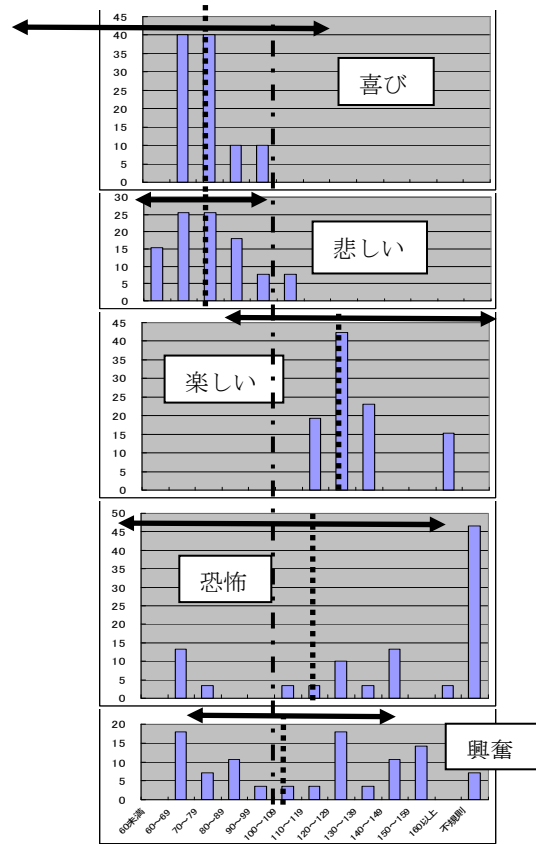


図1: カテゴリごとのテンポ

3.4.音量による音響的特徴の解析

3.4.1.実験方法

映像作品の映像データ (mpg) から PetitDecoco[5]により音データのみ (wav) を抽出する。そして、音のデータを Audacity[6]により、カテゴリごとに切り取る。このとき、声（せりふや溜息）や環境音（車のエンジン音、ドアの開閉音など）は音量が大きいので、声や環境音のない部分の背景音楽を切り取る。切り取った音のデータの

音量を MATLAB により分析する。単位は db (デシベル) とする。また、音データのサンプリング周波数は、48kHz である。

3.4.2. 音量の計測法

ある周波数で考えたとき、音圧レベル p の音量は、人間が聞こえる最小の音の音圧レベル p_{min} で割った比のデシベル値であらわされる。

$$20 \log_{10} \left| \frac{p}{p_{min}} \right| \quad (2)$$

ここでは、カテゴリごとに切り取ってきた音データを(2)式の p に代入し、背景音楽や声や環境音などが無い無音の部分の最小値を(2)式の p_{min} に代入する。そして、切り取ってきたデータごとに音量の平均を求め、カテゴリ別に特徴を分析する。

3.4.3. 実験結果

実験結果を図 2 に示す。図は(2)式により計算した値を四捨五入で整数値にして、ヒストグラムにしたものである。また、横軸は音量を db で示し、縦軸はそのカテゴリにおけるサンプル数の比率をパーセントで示したものとした。点線は平均値を示し、実線の矢印は 95% 信頼区間、鎖線はカテゴリ全体の平均音量 (22.31db) を示している。

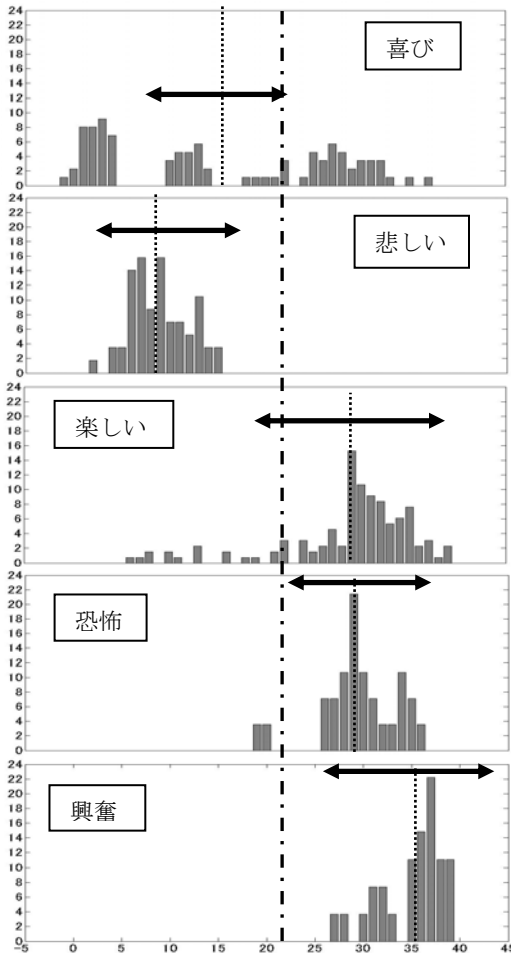


図 2: カテゴリごとの音量

3.4.4. 考察

それぞれのカテゴリごとに平均、95%信頼区間を求めた。95%信頼区間は(1)式を使って求めた。また、音量の大きさを相対的に比較するために、カテゴリ全体の平均音量を求めると、22.31db だった。このことから以下の事が明らかになった。

[喜び]のシーンは音量が小さい背景音楽が使われている(平均 15.14db, 信頼区間 6.89db~23.39db)。図を見ると使われている背景音楽の音量の幅が広く見えるが、これはセリフが入っている時に流れる背景音楽は音量が小さく感じたが、セリフが終わり背景音楽だけ流れるときに音量がだんだん大きくなるのが感じられたので、そのためだと考えられる。

[悲しい]シーンは音量の小さい背景音楽が使われやすい(平均 8.93db, 信頼区間 1.99db~15.87db)。

[楽しい]シーンは中間の音量の背景音楽が使われやすい(平均 28.70db, 信頼区間 19.48db~37.93db)。

[恐怖]のシーンは音量が大きめの背景音楽が使われやすい(平均 29.63db, 信頼区間 22.87db~36.39db)。

[興奮]のシーンは音量の大きい背景音楽が使われやすい(平均 35.57db, 信頼区間 25.98db~44.05db)。

3.5. 周波数による音響的特徴の解析

3.5.1. 実験方法

3.4.1 章で述べた方法と同様に背景音楽を切り取り、切り取った音のデータのパワースペクトルを MATLAB により分析する。分析の方法としては、ランニングスペクトルのフィルタバンク分析を行う。

3.5.2. ランニングスペクトルの分析

ランニングスペクトルとは、フレーム周期ごとに求められるスペクトルの時系列データのことを言う[7]。本研究では、サンプリング周波数 48kHz の音データに 4096 点 FFT を行い、その際の窓関数はハンニング窓を使用した。また、フレームのシフト幅は 2048 点である。

スペクトラムにフィルタバンクを用いると、多くの値からなるスペクトラム情報を数個の値にデータ量を減少でき、分析にかかる計算量が低減できる。フィルタバンクでは、得られた周波数スペクトラムを中心周波数と帯域幅をもつ複数の帯域に分割し、各帯域内において中心周波数の周りで加重平均を計算する[8]。本研究では、十二平均律を軸として、三角窓関数を用いてフィルタバンク分析を行った。軸の範囲はピアノの音階 (A0~C8) をもとに設定したので、軸の最小値は 55Hz, 最大値は 8400Hz とする。フレームのシフト幅は十二平均律の半音である。

3.5.3. 実験結果

実験結果は図 3 に示す。図の横軸は周波数を Hz で示し、縦軸はパワーの常用対数をとった。

3.5.4. 考察

それぞれのカテゴリの周波数分布に違いがあるのか評価するために、全てのシーンの周波数分布の平均と比較してカテゴリごとに特徴をみた。全体の平均との差が大きいほど特徴があると言える。まず、全てのシーンの周波数分布の平均を求めた。その全体の平均の周波数分布とそれぞれのカテゴリの周波数分布の各点ごとに差を求

め、プロットしたものが図 4 である。横軸は周波数を Hz で示し、縦軸はパワーを示した。また、4000Hz 以降はどのカテゴリも 0 に収束するので省いた。

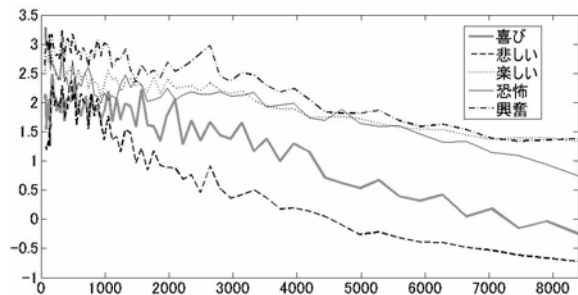


図 3: カテゴリごとの周波数分布

[喜び]のシーンと[悲しい]シーンは似たような分布になった。全ての周波数において平均より下回っていることから、ここでも音量が小さいということがわかる。また、500Hz 以下の低い音域は弱い。

[楽しい]シーンは、ほとんど 0 に近いことから、どの音域も平均的に使われていることがわかった。

[恐怖]のシーンは、100Hz 付近の低い音域が特に強く、ほかの音域はほぼ平均に近い。100Hz 付近の低い音域は人に恐怖感を与えらると思われ。

[興奮]のシーンは、200Hz 付近、500Hz 付近、1000Hz 付近、2600Hz 付近が強い。200Hz や 500Hz の低い音域と 1000Hz や 2600Hz の人に聞こえやすい音域を強くすることによって、より音量が大ききように感じるから人に不安感を与えていると思われ。また、全ての周波数において平均より上回っていることから、ここでも音量が大ききことがわかる。

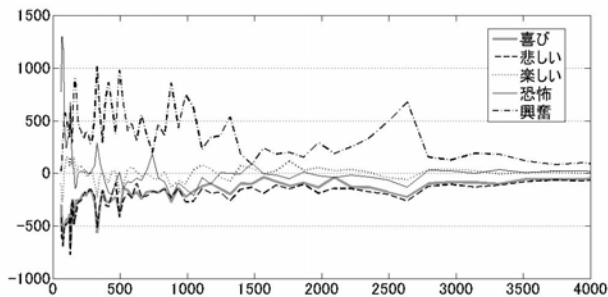


図 4: 全シーンとカテゴリごとの周波数分布の差

4.全体考察

[喜び]のシーンと[悲しい]シーンのように、視聴者をより感情移入させたいシーンの背景音楽はテンポが遅く音量が小さい。これによって、これらのシーンはセリフや役者の演出に視聴者の注意を引こうとしていると考えられる。

[楽しい]シーンの背景音楽は音量も周波数も平均的なものを使うことで、視聴者に背景音楽による強い刺激をあまり与えないようにしていると思われ。今回は研究の対象になっていないが、このシーンは背景音楽よりも環境音がよく聞こえる傾向があった。よって背景音楽と環

境音を組み合わせることで、にぎやかな雰囲気を演出していると考えられる。

[恐怖]のシーンと[興奮]のシーンはアクションが多いシーンと少ないシーンの 2 種類があり、これによっても背景音楽が使い分けられていることがわかった。また、この 2 つのシーンに共通して 200Hz 以下の低い音域のパワーが強かったため、この音域は人に恐怖感や不安感を与えると考えられる。

5.むすび

今回の研究により、それぞれのカテゴリごとに以下のような音響的特徴が明らかになった。

[喜び]のシーンと[悲しい]シーンの背景音楽はテンポが遅く、音量が小さく、500Hz 以下の低い音域はパワーが弱い。ただし、[喜び]のシーンよりも[悲しい]シーンの方がよりテンポが遅く音量が小さい。

[楽しい]シーンの背景音楽はテンポが速く、中間の音量で、周波数はどの音域も平均的に使われていた。

[恐怖]のシーンの背景音楽はアクションが多いところではテンポが速く、アクションが少ないところでは遅い。また、不規則なテンポの曲が多く使われていた。音量は大きめで 100Hz 付近の低い音域のパワーが強かった。

[興奮]のシーンの背景音楽はアクションが多いところではテンポが速く、アクションが少ないところではテンポが遅い。音量は大きく、低音や人に聞こえやすい音域のパワーが強くなっている。

しかし、今回の研究で[喜び]のシーンと[悲しい]シーンの背景音楽の違いをあまり明白にすることができなかった。ストーリーの状況別に背景音楽を使い分けするには、この二つのカテゴリの違いも明白にしなければならない。そのためには、楽器の数や種類を調べるなど別の方法で検討することで、この二つのカテゴリの違いを明らかにする必要がある。

文 献

- [1] 岩宮眞一郎, "映像作品における視聴覚コミュニケーション", 電子情報通信学会技術研究報告, vol.102, No.533, pp39-46, 2002
- [2] 岩宮眞一郎, "オーディオ・ヴィジュアル・メディアによる音楽聴取行動における視覚と聴覚の相互作用", 日本音響学会, vol.48, No.3, pp146-153, 1992
- [3] 岩宮眞一郎ほか, "音楽の調性及びテンポと映像の速度および密度が映像作品の印象に及ぼす影響", 音楽知覚認知研究, vol.8, No.2, pp53-64, 2002
- [4] <http://www.vector.co.jp/soft/win95/art/se107444.html>
- [5] <http://www.vector.co.jp/soft/win95/art/se185896.html>
- [6] <http://www.forest.impress.co.jp/lib/pic/music/soundedit/audacity.html>
- [7] 早坂昇ほか, "ランニングスペクトルフィルタを用いた雑音にロバストな音声認識", 信学技報, vol.103, No.146, pp31-36, 2003
- [8] 黒澤栄司ほか, "発音パラメータを用いた超低速音声通信の考察", 電子情報通信学会ソサイエティ大会講演論文集, pp473, 1996
- [9] 杉原雅ほか, "十二平均律に於ける音楽和音の音響学的研究(生活科学)", 京都府立大学学術報告, vol.13, No.1, pp75-82, 1959