6X-9

# 楽曲イメージに合わせたグラフィックス自動生成に関する一検討

高柳 直　　林 貴宏　　尾内 理紀夫

電気通信大学

## 1. はじめに

音楽や映像はそれぞれ単体でも視聴者に明るい，悲しいといった多くの心理的影響を与える．しかし，映画の BGM や音楽コンサートの視覚的演出のように，音楽と映像を組み合わせることにより，視聴者の感情により多くの影響を与えることが可能となる．そのため，Fantasia[1]や Butterflies in the Rain[2]など，音楽と映像を融合させた作品が今まで数多く制作されてきた．さらに近年，マルチメディア環境の普及によって，個人でも音楽，映像作品を制作または編集することが可能になってきている．

しかし，音楽と映像を融合させるという作業は，誰もが容易に行えるものではない．たとえば悲しい音楽に合わせて悲しい映像を流すといったように，心理的影響を考慮して，効果的な対応付けを考えなければならないためである．

そこで本稿では，心理的影響を考慮して楽曲イメージに合わせたグラフィックスを自動生成するシステムの研究開発に向けた第一歩として，新たな手法の検討について報告する．以下，2 章で本研究で提案するグラフィックス生成手法，3 章で評価実験について述べ，4 章で関連研究について述べる．

## 2. グラフィックス生成手法

本手法では，楽曲を演奏するとともに，その演奏に合わせて，グラフィックスの描画を行っていく．グラフィックスの生成は次の手順で行う．まず，和音などの楽曲特徴量を抽出する．そしてその楽曲特徴量を明るい，悲しいといった感性レベルの情報である楽曲イメージに変換する．次に，そのイメージを明度などのグラフィックス特徴量に変換し，最後にそのグラフィックス特徴量をもとにしてグラフィックスを生成，描画する

本稿では，第一歩として制作した実験的システムについて述べる．実験的システムでは扱う楽曲イメージを明るさのみに限定している．以下の各節で，そのシステムの各処理について述べる．

### 2.1 楽曲特徴量の抽出

本システムで扱う楽曲データは拍子情報を含む標準 MIDI ファイル(SMF)とした．まず，この SMF から楽曲特徴量を抽出し，41 次元の特徴ベクトル($f_1, f_2, \cdots, f_{41}$)を構成する($f_n \in \{0,1\}$)．特徴量は大きく音色，音量，テンポ，音高ヒストグラム，和音，音階の 6 種類に分類できる．以下で各特徴量について述べる．

**・音色に関する特徴量:$f_1$-$f_{16}$**

MIDI における音色は楽器の種類で表現される．$f_1$ から $f_{16}$ はそれぞれピアノ系，オルガン系といった GM 音源の音色マップで定義される 16 種類の楽器カテゴリに対応し，カテゴリ $n(1 \leqq n \leqq 16)$ に属する楽器の音が鳴っているとき $f_n=1$，そうでないとき $f_n=0$ とする．

**・音量に関する特徴量:$f_{17}, f_{18}$**

MIDI のチャネル c における Velocity 値の総和を u(c)，Expression 値を e(c)，Volume 値を v(c)として音量 w を次のように計算する．

$$w = \ln\left(\sum_c (u(c)e(c)v(c))\right)$$

この音量 w に対して，$f_{17}$，$f_{18}$ を次のように定義する．

$(f_{17}, f_{18}) = (0,0)$　$(w=0)$
$(f_{17}, f_{18}) = (1,0)$　$(0<w<14)$
$(f_{17}, f_{18}) = (1,1)$　$(14 \leqq w < 16)$
$(f_{17}, f_{18}) = (0,1)$　$(16 \leqq w)$

**・テンポに関する特徴量:$f_{19}, f_{20}$**

MIDI からテンポを取得し，1 分間に演奏される四分音符の個数 t に換算する．t に対して，$f_{19}$，$f_{20}$ を次のように定義する．

$(f_{19}, f_{20}) = (0,0)$　$(t<70)$
$(f_{19}, f_{20}) = (1,0)$　$(70 \leqq t < 100)$
$(f_{19}, f_{20}) = (1,1)$　$(100 \leqq t < 130)$
$(f_{19}, f_{20}) = (0,1)$　$(130 \leqq t)$

**・音高ヒストグラムに関する特徴量:$f_{21}$-$f_{36}$**

音高ヒストグラムとは，現在鳴っている音に対して，MIDI の最低音から最高音までの音程を 8 等分し，各区間に含まれる音の数を示したものである．8 つの各区間に対し，$f_{2n+19}$，$f_{2n+20}$($1 \leqq n \leqq 8$)を以下のように定義する．

$(f_{2n+19}, f_{2n+20}) = (0,0)$　(音数が 0 個)
$(f_{2n+19}, f_{2n+20}) = (1,0)$　(音数が 1-2 個)
$(f_{2n+19}, f_{2n+20}) = (1,1)$　(音数が 3-5 個)
$(f_{2n+19}, f_{2n+20}) = (0,1)$　(音数が 6 個以上)

---

A Study on Method for Generating Graphics Matched with Impression of Music
Nao Takayanagi, Takahiro Hayashi, and Rikio Onai
University of Electro-Communications

・和音に関する特徴量:$f_{37}$-$f_{39}$

$f_{37}$, $f_{38}$, $f_{39}$は現在の和音が，長3度の響きを含むか否か，短3度の響きを含むか否か，増4度の響きを含むか否かを表し，含むとき 1，含まないとき 0 とする．これらの響きを含むか否かの判定は 1 小節中に鳴ったすべての音の音程を調べることで行う．

・音階に関する特徴量:$f_{40}$,$f_{41}$

$f_{40}$, $f_{41}$ はそれぞれ現在使用されている音階が，長音階であるか否か，短音階であるか否かを表し，その音階であるとき 1，そうでないとき 0 とする．なお，音階の判定は，過去 4 小節に鳴った音に対して，それぞれの音階に当てはまるか否かを調べることによって行う．

### 2.2 楽曲イメージへの変換

抽出した楽曲特徴量から楽曲が明るいか暗いかといった楽曲イメージへの変換を行う．楽曲イメージへの変換には，入力層，中間層，出力層のニューロンの数をそれぞれ 41，80，2 とした 3 層の階層型ニューラルネットワークを用いた．

入力層の 41 個のニューロンは，$f_1$ から $f_{41}$ までの楽曲特徴量と対応し，それぞれ対応する特徴量の値が入力される．

出力層は楽曲のイメージが明るいときに活性化するニューロン $N_1$ と暗いときに活性化するニューロン $N_2$ の 2 つで構成される．この 2 つのニューロンの出力をもとに，楽曲の明るさ度合いを次のように計算する．

(明るさ度合い)＝($N_1$の活性度)-($N_2$の活性度)

なお，ニューラルネットワークの学習は，あらかじめ数曲の楽曲のイメージを，明るい，中間，暗いの 3 段階で主観評価し，それを出力層の 2 つのニューロンに次のように教師信号として与えることで行った．

楽曲イメージの主観評価が明るいとき:

($N_1$の教師信号,$N_2$の教師信号)＝(1,0)

楽曲イメージの主観評価が中間のとき:

($N_1$の教師信号,$N_2$の教師信号)＝(1,1)

楽曲イメージの主観評価が暗いとき:

($N_1$の教師信号,$N_2$の教師信号)＝(0,1)

### 2.3 グラフィックス特徴量への変換

今回の実験的システムでは，グラフィックス特徴量としてグラフィックスの明度のみを扱う．明度は[0,100]の範囲の整数値とし，明るさ度合いが 0.01 より大きい間は値を増加させていき，明るさ度合いが-0.01 より小さい間は値を減少させていく．

### 2.4 グラフィックスの生成

最後に，グラフィックス特徴量を反映させて CG アニメーションを描く．たとえば明度が高ければ，背景色や照明などを明るくした CG を描く．

## 3. 評価実験

本稿で提案したグラフィックスの自動生成手法により実際に生成されたグラフィックスが楽曲のイメージに合っているかどうかを調べるため，アンケート調査による評価実験を行った．

実験にはクラシック音楽 4 曲を用いた．被験者には曲と生成したグラフィックスアニメーションを同時に視聴してもらい，曲の明るさという観点のみから，それぞれの曲についてイメージに合っているかどうかを答えてもらった．

アンケート結果を表 1 に示す．今回の結果では，よく合っている，やや合っているという回答が 50%を占めた一方で，あまり合っていない，まったく合っていないという回答も 30%あり，まだ不十分な部分も多いことがわかった．イメージに合わなかった原因としては，楽曲特徴量を抽出する際に過去に鳴った音を利用しているため，楽曲の演奏に対して使用する特徴量が遅れるといったことなどが考えられる．

表 1: アンケート結果

| | |
|---|---|
| イメージによく合っている | 19.4% |
| イメージにやや合っている | 30.6% |
| どちらでもない | 19.4% |
| イメージにあまり合っていない | 25.0% |
| イメージにまったく合っていない | 5.6% |

## 4. 関連研究

楽曲からのグラフィックス生成として，音楽情報の可視化に関する研究が多く行われている[3][4]．これらの研究は音楽の構造をわかりやすく提示することを目的としており，音楽や映像が視聴者に与える心理的影響を考慮して，楽曲のイメージに合わせたグラフィックス生成を行う本研究とはねらいが異なる．

## 5. おわりに

楽曲イメージに合わせたグラフィックスの自動生成手法を提案し，評価実験を行った．その結果，50%の精度でイメージに合ったグラフィックスを生成できたことを示した．今後，楽曲特徴量の抽出法などを改良するとともに，明るさ以外の楽曲イメージにも対応させ，よりイメージに適したグラフィックスの生成を目指し，検討を行っていく．

## 参考文献

[1] Walt Disney: “Fantasia”, 1940
[2] The Ohio State University: “Butterflies in the Rain”, SIGGRAPH 97, 1997
[3] Sean M. Smith, Glen N. Williams: “A Visualization of Music”, in Proceedings of IEEE Visualization, pp.499-503, 1997
[4] 宮崎 麗子，藤代 一成: “comp-i: MIDI データの 3 次元可視化”，情報処理学会研究報告，音楽情報科学，No.46-9，pp.55-60，2002