

# らくがっきー：絵と音を融合したデジタルコンテンツの提案

後藤沙織<sup>†</sup> 水野慎士<sup>†</sup>

画像と音楽は密接な関係があり、音楽は画像の印象を強調したり変化させたりすることができる。そのため、画像から音楽を生成するような手法がいくつも研究されている。そこで本論文では、お絵描きに基づいてサウンドを生成することができる新しいインタラクティブメディアシステム"らくがっきー"を提案する。このシステムでは、ユーザはスクリーン内にお絵描きをすることで、システムがお絵描きに対応するサウンドを自動的に生成する。システムはユーザのお絵描き中に絵の解析を行い、その解析結果に基づいて対話的サウンドの生成や変更を行う。ユーザはサウンドを伴ったお絵描きを対話的に楽しむことができる。

## RAKUGACKY: A digital Content Harmonizing Drawing and Sound

SAORI GOTO<sup>†</sup> SHINJI MIZUNO<sup>†</sup>

Images and sounds are closely related, and sounds could emphasis or change impressions of images. Thus several methods to synthesize sounds from images are developed. In this paper, we propose a novel interactive media system "RAKUGACKY" that can make sound with drawing. In this system, a user draws a picture on a screen, and the system synthesizes sounds for the picture automatically. The system analyzes the picture during drawing, and synthesizes or changes sounds interactively according to the analyzing result. The user can enjoy drawing with sound interactively.

### 1. はじめに

コンピュータ技術の発展に伴い、デジタル技術を用いて映像やサウンドを提示するメディアアートや、人の操作などに対して対話的に反応したりするインタラクティブメディアシステムなど、数多くのデジタルコンテンツが開発されている。例えば、"GAYAIT"は人の動作に反応して、映像と音をコントロールしながら個と大衆の状態を対話的に切り替えながら作り出すことができるシステムであり、博物館の展示やライブの演出などに用いられている(図1(a)) [1]。また、"フキダシステム"は歩いている人を認識して、その人から吹き出しが出ているようにビデオ素材を表示する広告システムである(図1(b)) [2]。このように、デジタルコンテンツは広告やエンターテイメント、教育などの様々な分野で活用されており、今後の進展も大きく期待されている[3][4]。

デジタルコンテンツでは画像と音楽やサウンドを扱うものが多い。そして、画像から音楽を生成するような手法もいくつか研究されている。例えば、画像特徴量に基づいてメロディを自動生成する手法[5]や、画像の印象に合わせて自動的に音楽をアレンジする手法[6]などが提案されている。画像と音楽は密接な関係があり、音楽は画像の印象を強調したり変化させたりすることができる。

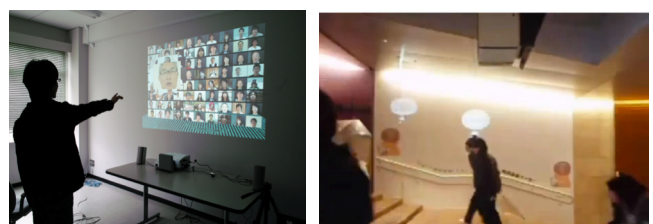
そこで、本研究では画像と音楽を組み合わせたデジタルコンテンツとして"らくがっきー"を提案する。"らくがっきー"では、画像としてお絵描きに着目したもので、お絵描きしながら絵の解析を行い、サウンド生成を行うことが

できるインタラクティブメディアシステムである。本研究と同様に、お絵描きに基づいてサウンドを生成するシステムとして"e-gakki"が提案されている[7]。これは、キャンパスの縦軸を音階、横軸を時間軸と考えて絵をメロディに変換するものであり、楽譜を別の形式で表現したものとと言える。それに対して、"らくがっきー"では、描かれた絵の中に含まれるオブジェクトの色、種類、座標、面積などに基づいて、空間に様々な音源を配置していくことで、絵に合わせたステレオサウンドを生成していく。そして、サウンド生成はお絵描き中に逐次行われるため、お絵描きによってサウンドが対話的に変化していく。"らくがっきー"ではサウンド生成によって拡張されたお絵描きを楽しんだり、環境BGMをデザインしたりすることができる。

### 2. システムの概要

本論文で提案する"らくがっきー"の外観を図2に示す。ユーザはPC画面内に表示されたキャンパスに、ペンタブレットを用いてパレットで色を選択しながら、一般的なペイントツールと同様に自由に絵を描いていく。

システムはキャンパス内の絵を解析することでサウンド



(a) GAYAIT. (b) フキダシステム.

図1 インタラクティブメディアシステムの例.

<sup>†</sup> 愛知工業大学 情報科学部  
Faculty of Information Science, Aichi Institute of Technology

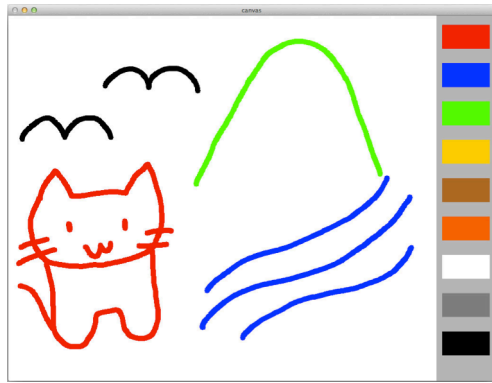


図2 "らくがっきー"の外観.

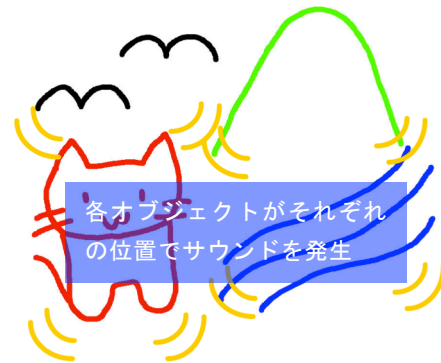


図3 "らくがっきー"によるサウンド生成の様子.

を生成する. まず色に基づいてオブジェクトに分解してから, 形状特徴量などに基づいてオブジェクトを分類する. そして分類結果に基づいて空間に音源を配置する. 音源の種類はオブジェクトの種類によって異なり, 音源のパラメータはオブジェクトの特徴量によって変化する. また, オブジェクトの位置によってステレオサウンドの左右バランスも変化する.

以上により, 描かれた絵に基づくステレオサウンドが合成される (図3). これらの処理はリアルタイムで行われるため, ユーザはお絵描きを通じて, その絵の雰囲気にあったサウンドを対話的に生成することができる.

### 3. 実現方法

#### 3.1 オブジェクトと音源

提案システムでは, 子供が描いた絵に基づいてあらかじめ約20種類のオブジェクトを想定している. オブジェクトは主に動物や自然に関するものである. そして, 各オブジェクトに対応する音源をサウンドファイルとして準備している. 図4に主なオブジェクトとその音源の種類を示す.

#### 3.2 お絵描きからオブジェクトへの分解

システムでは, ユーザはパレットでペンの色を選択して

から, キャンバスに絵を描いていく. 現在, ペンの色の種類は赤, 青, 緑, 黄, 茶, 橙, グレー, 黒, および消去用の白が用意されている.

システムはキャンバスに描かれた絵を色に基づいて領域に分解する. そして, 生成された各領域を描かれたオブジェクトとする.

#### 3.3 オブジェクトの種類の決定

配置する音源の種類を決定するためには, 各オブジェクトの種類を決定する必要がある. そのため, 各オブジェクトの形状特徴量などを計算してオブジェクトを解析する.

取得する特徴量は, オブジェクトの面積, 周囲長, 円形度, 左右端を結ぶ線分と水平線のなす角度などである. そして, これらの特徴量を組み合わせて用いることで, 各オブジェクトを分類していく. 使用する特徴量およびそのしきい値などは, 描画を想定するオブジェクトごとに発見的に決定している. 以下に青色オブジェクト, 緑色オブジェクト, および黒色オブジェクトの分類方法を示す.

<青色オブジェクト>

領域の面積  $S_B$ , 領域の円形度  $C_B$ , 領域の左右端点を結ぶベクトル  $V_B$  とキャンバス水平軸とのなす角度  $\theta_B$  を用いて,

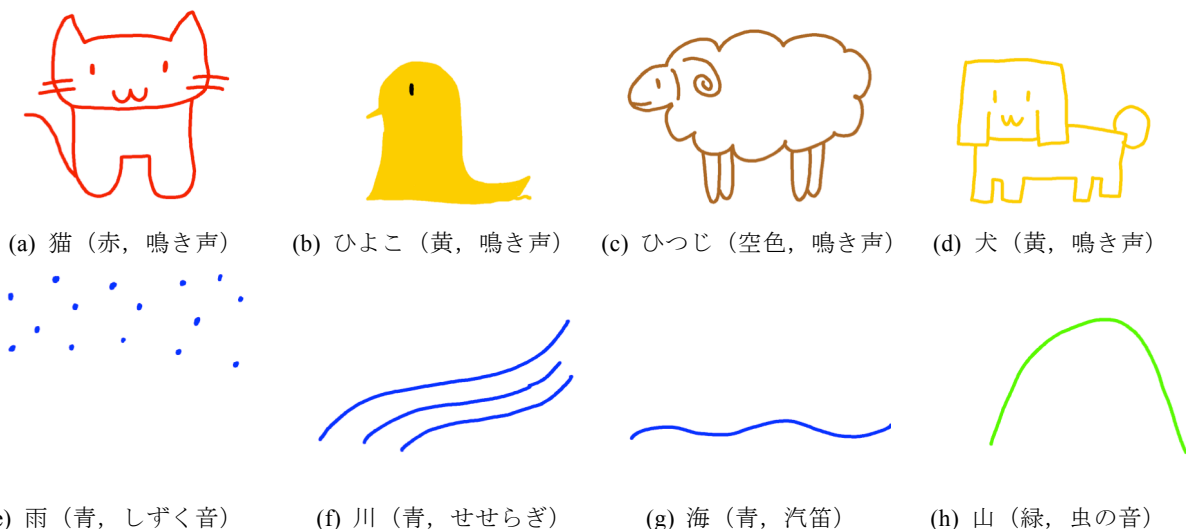


図4 主なオブジェクトの種類.

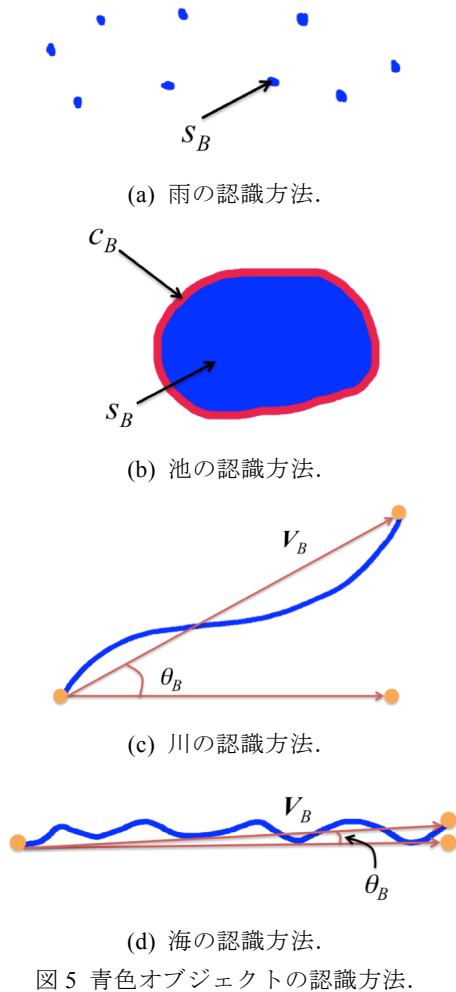


図5 青色オブジェクトの認識方法.

以下の手順で分類していく (図5).

1.  $S_B < S_{B1}$  → 雨粒
  2.  $C_B > C_{B1}$  → 池
  3.  $|V_B| > V_{B1}$  かつ  $|\theta_B| > \theta_{B1}$  → 川
  4.  $|V_B| > V_{B1}$  かつ  $|\theta_B| < \theta_{B1}$  → 海
  5. それ以外 → その他オブジェクト (音源なし)
- ( $S_{B1}$ ,  $C_{B1}$ ,  $V_{B1}$ ,  $\theta_{B1}$ : しきい値)

<緑色オブジェクト>

領域の上端点から左右端点への2つのベクトル  $V_{Ga}$ ,  $V_{Gb}$  とベクトル間の角度  $\theta_G$  を用いて, 以下の手順で分類していく (図6).

1.  $|V_{Ga}| / |V_{Gb}| < R_{G1}$  かつ  $\theta_{G1} < |\theta_G| < \theta_{Gh}$  → 山
  2. それ以外 → 草原
- ( $R_{G1}$ ,  $\theta_{G1}$ ,  $\theta_{Gh}$ : しきい値)

<黒色オブジェクト>

領域の左端点を起点に領域輪郭を8等分する点  $P_1 \sim P_8$  を求める. そしてベクトル  $P_1P_3$  と  $x$  軸のなす角度  $\theta_{Ka}$ , ベクトル  $P_3P_5$  と  $x$  軸のなす角度  $\theta_{Kb}$ , ベクトル  $P_2P_4$  とキャンバス水平軸のなす角度  $\theta_{Kc}$  を求めて, 以下の手順で分類していく (図7).

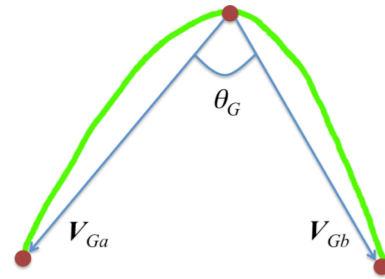


図6 緑色のオブジェクト (山) の認識方法.

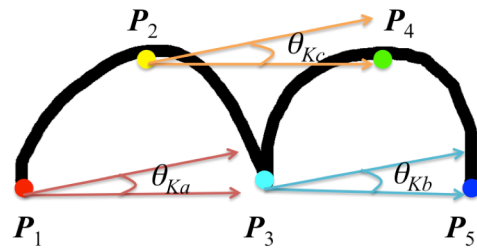


図7 黒色のオブジェクト (鳥) の認識方法.

1.  $P_2, P_4$  の2点の  $y$  座標が  $P_1, P_3, P_5$  のいずれの  $y$  座標よりも大きく, かつ  $|\theta_{Ka}| < \theta_{K1}$  かつ  $|\theta_{Kb}| < \theta_{K1}$  かつ  $|\theta_{Kc}| > \theta_{K1}$  → 鳥
  2. それ以外 → その他オブジェクト (音源なし)
- ( $\theta_{K1}$ : しきい値)

3.4 音源の配置とサウンド生成

前節の手法で各オブジェクトの種類が決定すると, 音源の種類も決定する. このとき, 音源の種類によっては, オブジェクトの面積に基づいて音源のピッチを変更する. ここでピッチとは音の速度と音程の両方を示す. 例えば, オブジェクトの面積が大きいほど, 音が速くて高音の音源設定になる.

そして音源の種類の設定が決定すると, 図8に示すように各オブジェクトの描画位置に基づいて, 各音源を空間中に配置する. このとき, 音源の位置によって生成するサウンドの左右音量を調整する. そして, 配置されたすべての音源がそれぞれの音を再生することで, 描画した絵に基づくステレオサウンドが合成される.

4. 実験

4.1 実装

提案したシステム"らくがっきー"を PC に実装して実験を行った. 使用した PC は MacBookPro (Core i7 2.5GHz, 8GB メモリ) で, 画像生成のために OpenGL, 描画オブジェクトの解析のために OpenCV, サウンド生成のために OpenAL を用いている. また, 各音源はモノラル WAV フ

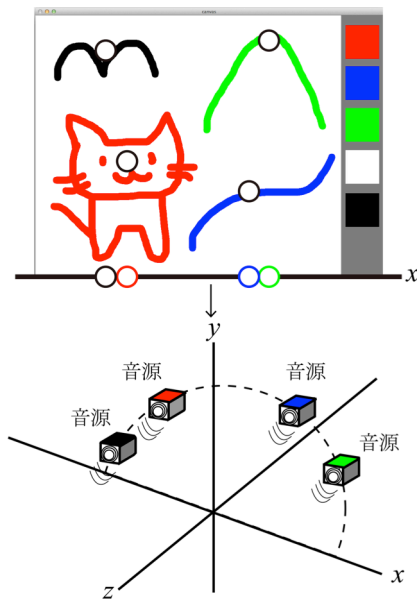


図8 オブジェクトの位置に基づく音源の空間配置.

ファイルで用意しており、それぞれ数秒の長さとなっている.

#### 4.2 実験結果

ユーザはペンタブレットを用いることでより、システムでもスケッチブックの感覚で絵を描くことができた. そして各色で絵を描いていくことで、様々なサウンドが対話的に追加されて合成されていくことを確認した.

赤色で猫を描けば猫の鳴き声が聞こえ (図 9(a)), 青色で雨の絵を描けば雨音が追加され (図 9(b)), 川を描けばせせらぎの音が聞こえ (図 9(c)), 緑色で山を描けば鈴虫の鳴き声 (図 9(d)), そして黒色で鳥を描けばカラスの鳴き声が聞こえてきて、最終的にはすべての音が合成された (図 9(e)).

音源は絵の解析結果に基づいて、種類やピッチを決定した上で空間に配置されてループ再生されるため、動物の鳴き声や水の音が、それぞれのピッチでそれぞれの場所から自然に聞こえてくる. そしてユーザがオブジェクトを描き加えれば、音源が追加されて、生成されるサウンドもリアルタイムで逐次変化していった.

また、絵の中に含まれるオブジェクトに応じて音源の種類や性質を変化させることもできる. 例えば、図 10(a)では猫だけが描かれており、猫は普通の鳴き方をしているが、図 10(b)のようにねずみを追加で描画すると、猫の鳴き声の音源およびピッチを変更することで、うなったような鳴き声に変化する.

現在のシステムでは、オブジェクトの分類は発見的手法に基づく. 従って、想定しない描き方をされた場合には、描かれたオブジェクトに合わない音源が配置されたり、音源が配置されなかったりする場合もあった.

#### 5. まとめ

本論文では、絵と音を融合したインタラクティブメディア

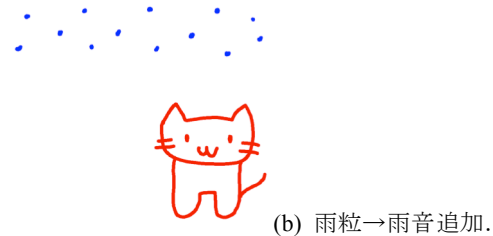
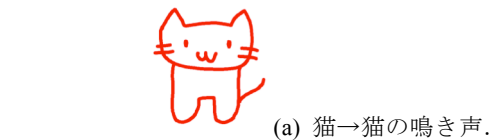


図 9 "らくがっきー"によるお絵描きとその絵に基づくサウンド生成の様子.



(a) 猫だけ→通常の鳴き声. (b) ねずみ追加→うなり声.  
 図 10 オブジェクトの組み合わせによりサウンドの変化.

システム"らくがっきー"を提案した. システムでは、絵の描画、絵の解析によるオブジェクトへの分解と分類、音源の配置、サウンド合成などの手法を開発して、実装と実

験を行った。"らくがっきー"では、お絵描きを進めるのにつれて聞こえてくるサウンドが逐次変わっていく特徴がある。そのため、様々な年代の人々を対象にしたインタラクティブメディアシステムであり、幼稚園や小学校低学年の図画において、お絵描きに対するモチベーション向上も期待できる。

現時点では、描画した絵の解析は色および発見的な特徴量を用いた単純な形状認識に基づくため、生成されるサウンドを細かく制御することは困難であり、ユーザの意図しないサウンドが生成される場合もある。描いた絵に一層適合したサウンドを生成するためには、より詳細なオブジェクト認識手法を開発する必要がある。また、スケッチブックに基づく三次元CG生成手法[8]との組み合わせも検討中である。そして、提案システムを学校などで子供たちに実際に使ってもらうことで、システムの改良を行う必要がある。

## 参考文献

- 1) Shinji Mizuno, Ryosuke Hirano, Yukihiko Tsutsumi, "GAYAIT: an Intetactive Video and Sound Art System handling a large number of Video Clips and its Applications", 芸術科学会論文誌, Vol. 11, No. 4, pp. 149-156, 2012.
- 2) フキダシステム: <http://www.fukidasystem.com>
- 3) 中村隆敏, 角和博, 穂屋下茂, 高崎光浩, 大谷誠, 藤井俊子, 古賀崇朗, 永溪晃二, 久家淳子, 時井由花, 河道威, 米満潔, 原口聡史, 本田一郎, 梅崎卓哉, "デジタル表現技術とメディアアートの関連性に関する一考察", 佐賀大学文化教育学部研究論文集, Vol. 15, No. 2, pp. 63-75, 2011.
- 4) 長嶋洋一, "フィジカル・コンピューティングとメディアアート", 情報処理学会研究報告, 音楽情報科学, Vol. 2008, No. 89, pp. 1-6, 2008.
- 5) 若林努, 小林孝浩, 平林 真実, "画像の特徴を用いたメロディ自動生成システムの提案", 情報処理学会インタラクション2011論文集, pp. 415-418, 2011.
- 6) 大山喜冴, 伊藤貴之, "DIVA: 画像の印象に合わせた音楽自動アレンジの一手法の提案", 芸術科学会論文誌, Vol. 6, No. 3, pp. 126-135, 2007.
- 7) 葛西信也, 後藤幸功, 村山優子, "e-gakki: 描画情報を音に変換するシステム", 情報処理学会全国大会講演論文集, pp. 4493-4494, 2007.
- 8) 近藤菜々子, 水野慎士, "スケッチブックでのお絵描きを拡張する映像ツールの提案と実現方法", 情報処理学会研究報告, デジタルコンテンツクリエーション, Vol. 2012-DCC-2, No. 20, 2012.