

# ジェスチャとグラフィックスを用いた音編集手法 の提案と評価

加藤 裕介<sup>†</sup> 林 貴宏<sup>‡</sup> 尾内理紀夫<sup>†‡</sup>

電気通信大学大学院電気通信学研究科<sup>†</sup>

電気通信大学電気通信学部情報工学科<sup>‡</sup>

## 1. はじめに

近年コンピュータの性能向上と一般への普及に伴って、コンピュータを利用して音編集が可能な環境ができてきている。特にムービーや Web ページの作成といった創作活動を個人で行うユーザが増加している現在、独自の効果音を作品に取り入れることで演出効果や自己表現をすることが考えられ、その目的を果たす手段として手軽な音編集は大きな意義を持つ。現在、音楽制御や音編集のシステムは数多く提案されているが、それらのシステムは音データを周波数などに数値化、または音符などに記号化して、ユーザに提示するものが多い。これらの数値や記号から音をイメージするには専門的な知識を必要とするため、音楽に馴染みのないユーザにとって扱いにくい面もある。

そこで本研究では、音の専門的な知識を持たない人でも直感的に理解しやすい音編集システムを目標とし、その第一歩として入力された音の特徴量をグラフィックスの形状や色といった特徴量に対応させ、ジェスチャを用いてグラフィックスを操作することで音を編集するシステムを試作し、評価する。以下 2 章では試作システムの概要、3 章では音とグラフィックスの対応関係、4 章ではジェスチャと音編集の対応関係について述べ、5 章で実験による本システムの評価について述べる。

## 2. 試作システム

### 2.1 システム全体構成

システムは DV (デジタルビデオ) カメラ、マイク、スピーカ、プロジェクタ、音響解析モジュール、画像描画モジュール、画像認識モジュールで構成されている (図 1)。

### 2.2 処理の流れ

システムの処理は以下のとおりである。音がマイクから入力されると、音の特徴量が音響解析モジュールで解析される。画像描画モジュールは解析結果に対応したグラフィックスを、プロジェクタを通してスクリーン上に描画する。次に描画されたグラフィックスに対してユーザがジェスチャを行う。ジェスチャは DV カメラで撮影され、画像認識モジュールで判別される。ジェスチャに応じて、画像描画モジュールはグラフィックスの色・形状を変化させ、音響解析モジュールは音を編集する。

### 2.3 各モジュールの動作

音響解析モジュールはマイクロホンから入力された音信号を周波数特性・音の長さ、音量に分けて解析する。また音信号を Wave データとして保存し、編集する。

Proposal and evaluation of a sound editing system using graphics and gesture

<sup>†</sup> Yusuke Kato, Graduate School of Computer Science, University of Electro-Communications

<sup>‡</sup> Takahiro Hayashi and Rikio Onai, Department of Computer Science, University of Electro-Communications

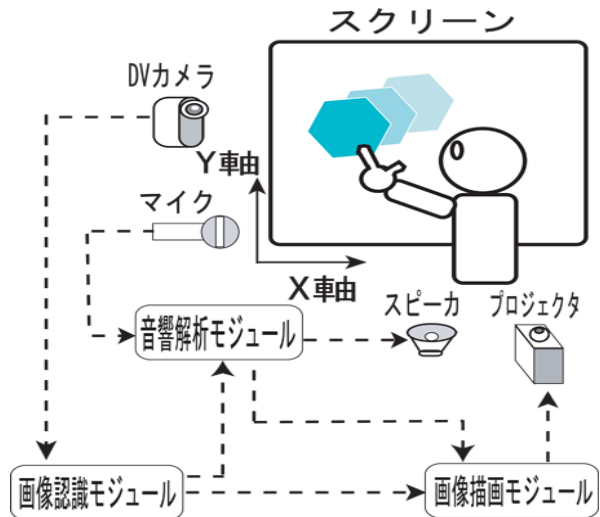


図 1 システム全体構成

画像描画モジュールは音響解析モジュール・画像認識モジュールから送られてきたデータに基づいてグラフィックスを描画する。

画像認識モジュールはユーザの行ったジェスチャを判定する。ジェスチャは手形状、手座標、手座標の移動の 3 つデータに基づいて判別される。例として図 2 に「引っ張る」というジェスチャの判別方法を示す。ユーザは手を「グー」の手形状のままグラフィックスの上に重ねる (図 2(a))。その後、「グー」の手形状を保ったまま手座標をグラフィックスの中心から外側に向かって移動させる (図 2(b)) とシステムは「引っ張る」ジェスチャが行われたと認識し、グラフィックスを拡大させる。グラフィックスの大きさが変化すると、対応する音データの音量も変化する。手形状が変化、または手座標の移動が一定値を超えるとジェスチャ認識はその時点で終了する。

### 3. 音とグラフィックスの対応関係

音とグラフィックスは「明るい」「暗い」といった同一の印象語によって表現される場合がある [1][2]。そこでこれら共通する印象語を決定する要因となる、音とグラフィックス各々の特徴量同士

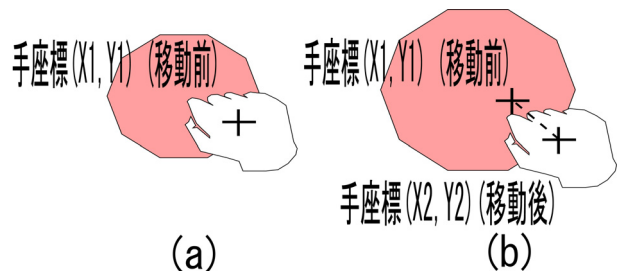


図 2 ジェスチャ判定方法

音の特徴量	グラフィックスの特徴量
音量	半径
音の長さ	横軸方向の長さ
音高	明度
倍音成分	形状

表 1 音とグラフィックスの対応関係

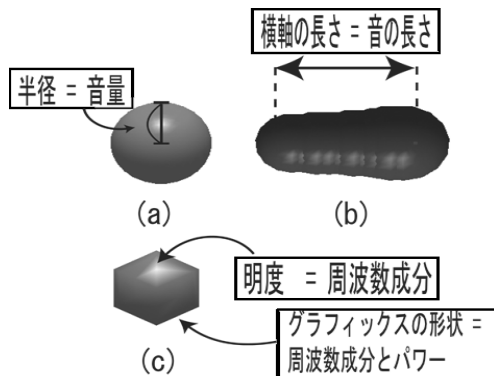


図 3 音とグラフィックスの対応関係

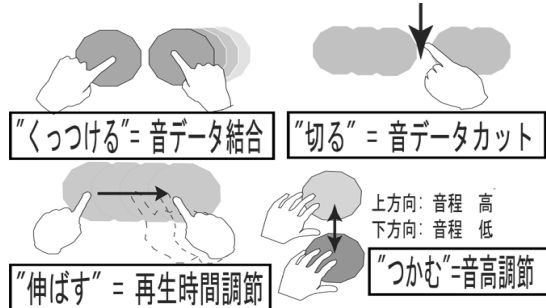


図 4 ジェスチャの種類

を対応づけることで、グラフィックスを見て音を直感的にイメージできるのではないかと考え、表 1 のような対応関係を定義した。音量はグラフィックスの半径と対応している(図 3(a))。音の長さは横軸方向の長さに対応している(図 3(b))。音の高さはグラフィックスの明度に対応している。音の周波数が高くなるほど、グラフィックスの明度が高くなる(図 3(c))。基本周波数の倍音成分は形状に対応している。高次の倍音成分に強いパワーがあれば、グラフィックスは鋭い形状へ近づいていく(図 3(c))。

#### 4. ジェスチャと音編集の対応関係

本システムでは「音量調節」「音高調節」「再生時間調節」「音データのカット」「音データの結合」の 5 つの基本的な編集機能と、それに対応した、それぞれ直感的な動作であると考えられるジェスチャを実装した。図 4 に図 2 で示した「引く張る(音量調節)」以外のジェスチャ動作を示す。

#### 5. 実験

本システムで定義したジェスチャとそれらジェスチャに対応したグラフィックスの特徴量変化が直感的であるか、という検証のためにアンケート調査を行った。被験者には実際にジェスチャによる音編集を行ってもらい、5 つのジェスチャそれぞれに対して以下の 2 つの質問について 5 段階評価で答えてもらった(5 点が最高、1 点が最低)。

**質問 1.** 本システムの 5 つのジェスチャはそれぞれ、グラフィックスの形状などを変化させる方法

ジェスチャのわかりやすさとグラフィックス変化のわかりやすさに関するアンケート	平均評価点
引く張る 大きさの変化	3.5
つかむ 位置と明度の変化	4.4
伸ばす・縮める 横方向の長さの変化	4.8
切る グラフィックスの分裂	4.5
くっつける グラフィックス結合	4.5

表 2 質問 1 に対するアンケート結果

グラフィックスの変化から音の変化がイメージできたかどうかのアンケート	平均評価点
大きさの変化 音量の変化	3.9
位置と明度の変化 周波数の変化	3.7
横方向の長さの変化 音の長さの変化	4.3
グラフィックスの分裂 音データの分裂	4.0
グラフィックスの結合 音データの結合	4.4

表 3 質問 2 に対するアンケート結果

として、直感的にわかりやすいものであったか。  
**質問 2.** グラフィックスの変化から、音の変化をイメージできたか。

表 2, 表 3 に質問 1, 質問 2 に対するアンケート結果を示す。平均評価点は有効数字 2 桁で表示している。

表 2 より「引く張る」以外のジェスチャは高い評価を得ており、これらジェスチャはユーザにとって違和感のないジェスチャであると考えられる。「引く張る」に対しては「片手より両手で引く張るほうが自然である」といった意見が聞かれた。また表 3 より「位置と明度の変化」以外に高い評価が得られた。「位置と明度の変化」の評価が他より低い理由として、「明度の変化率が音の高さの変化率とイメージ的に一致しない」といった意見が聞かれた。これは明度の変化量に対して人間が知覚する明るさの変化と、周波数の変化量に対して人間が知覚する音の高さの変化が比例関係にないためであると考えられる。こうした視覚心理、聴覚心理に関わる現象を定量的にモデル化することで、より一層グラフィックスと音の変化が直感的に結びつくと考えられる。

#### 6. おわりに

本研究ではユーザにとって直感的でわかりやすい音編集ができることを目的としたシステムを試作した。音とグラフィックスを対応づけ、グラフィックスを変化させるためのインターフェースとしてジェスチャを用いる手法が、わかりやすい音編集システムを実現する上で有効であることを評価実験により確認した。今後はジェスチャを用いることでキーボードやマウス操作に比べてユーザの負担が軽くなるような編集機能を実装する予定である。最後に画像処理にご指導いただいた電気通信大学小池英樹助教授、東京農工大学中西泰人助教授、ソニーコンピュータサイエンス研究所高田哲司アソシエイトリサーチャーに深謝する。

#### 参考文献

- [1] 岩井大輔, 長田典子, 津田学, 和氣早苗, 井口征士, 音と色のノンバーバルマッピング - 色聴保持者のマッピング抽出とその応用 -, 情報処理学会研究報告, MUS-47, pp. 97-104, 2000.
- [2] B.C.J.ムーア著, 大串健吾監訳, 聴覚心理学概論, 誠信書房, 1994.