

# 空間情報を用いた音声生成に関する研究動向

岩淵 勇樹

## 1 はじめに

アート・エンターテインメントや楽器インターフェースなどの分野において、近年は音声情報を用いた視覚表現の研究が盛んである [1]。

一方、センサ情報を用いて音声を生成する研究も同分野で盛んに行なわれている。本稿では、その中でも空間情報（2次元 / 3次元）を利用したもの、あるいは画像情報を用いたものについて概観する。

## 2 平面情報の時間-ピッチへの変換

Iannis Xenakis が発明した *UPIC*[2] はいわば楽譜のデジタル的な概念をアナログに置き換えたようなもので、描画した線形について横軸が時間、縦軸が音高として音楽が生成される。

音の可視化技術としてスペクトログラムがあるが、これも横軸が時間、縦軸が音高（周波数）に対応するという点では *UPIC* と同様である。実際、可視化の逆として、スペクトルグラムをラスタ画像として編集して音声信号を得る *MetaSynth* などのソフトウェアが開発されている。エイフェックス・ツインなどの音楽家が実際にこれらのソフトウェアを用い、幾何学図形や写真に基づいた楽曲<sup>1</sup>を発表している [5]。

また、動画投稿サイトでは最近 *The Music Animation Machine*[6] などを用いた“MIDI アニメ”と呼ばれる作品が多く投稿されており、*UPIC* と楽譜（ピアノロール）の中間的存在ともいえる。

*TENORI-ON*[4] も横軸を時間、縦軸をピッチとしているので同じ系統であるといえる（ただしモードによる）。

## 3 平面情報を用いたその他の音声信号化

前節のように平面図形や画像の横軸を時間、縦軸をピッチ（周波数）とみなす方法は直感的である。しかしながら、これらの方法では入力図形が回転した場合に出力が異なるなど、図形的特徴を反映したものではない。

Levin による *Manual Input Sessions*[7] では、シルエット画像から輪郭線を抽出し、面積や周囲長などの特徴量を MIDI 信号に変換しており、図形的特徴を反映している。

また、著者の研究は、入力図形の拡大・縮小や回転に対して割り当てられる音色が同じであり、図形的特徴を音声信号と結びつけることに重きを置いている [8]。

ラスタ画像の輝度値を音声信号に変換する研究もいくつかあるが、音楽に用いる以外にシーン情報の伝達手段として研究される<sup>2</sup>例もある [9]。

## 4 汎用的な視聴覚情報処理システム

アート・エンターテインメント分野で画像を用いた音声処理を行なう環境としては Max がもっとも有名であり、これを用いたさまざまな作品 [13][15] が作られているが、この他にも特定の画像処理を中心とした汎用的な環境が開発されている。

*BigEye* は、画像情報を MIDI 信号に変換するためのプログラミング環境であり、検出した対象物の色情報や動きをどうやって MIDI 信号に変換するかを自由に記述できる。 [12]

<sup>1</sup> 画像情報を用いる場合、高周波成分も多く含まれることになり、金属的な音になりやすい。

<sup>2</sup> ちなみに、鳥の鳴き声研究に携わってる知人によると、鳥の声と脳波を照らし合わせるため、脳波を音として聞いているそうです。

*EyesWeb* は Max に似た環境であり、主にダンスと音楽の融合のために作られている [11]。

## 5 ジェスチャー入力

そもそも、電子音楽の祖であるテルミンも一種の空間情報を用いた楽器インターフェイスであり、更に遡れば管弦楽器などもピッチを（3次元空間上での1次元の）空間情報を用いて制御している。空間情報を用いた音色生成という観点からすると、任意の画像や図形から音色を生成するものよりは、人間のジェスチャーを解析して入力インターフェイスとして用いるものの方が圧倒的に多い [10]。

ジェスチャーを解析する方法は、上記のような画像処理によるものと、*Wii Music* のように加速度センサーを用いるものに大別される。

楽器の模擬としてのインターフェイスも数多く研究されているが、*Very Nervous System*[14] や *Mouthesizer*[10]、*SoFA*[15]、などのように、体の動きそのものを楽器インターフェイスとして用いるものも増えている。

## 6 視聴覚融合インターフェイスおよび作品

楽器インターフェイスは近年 *KAOSSILATOR*[3] や *TENORI-ON*[4] など、作曲やライブパフォーマンスにおいて実用的な製品が多く開発されており、*Audiopad*[2] などのタンジブルインターフェイスを用いたものもある。

## 7 まとめ

本稿では、空間情報を音声情報に変換する研究を概観した。今回紹介した作品・システムの一部などを以下の表にまとめる。

画像処理やセンサ情報処理に特化した環境が充実する中、インターフェイスはアイデア次第で次々と

新しいものが作られているが、それをどのような音に変換するかはまた別問題であり、その処理で変わり映えするものは少ない。特に、音色を自由に作り出す方法についてはあまり検討されておらず、これから研究を進めたいところである。

## 参考文献

- [1] 江渡 浩一郎, “アート・エンターテインメントにおける音インタフェース,” 情報処理学会研究報告. SLP, 音声言語情報処理, Vol. 2004, No. 74, pp. 53-58, 2004.
- [2] H. Solís, “Improvisatory music and painting interface,” Masters thesis, Massachusetts Institute of Technology, 2004.
- [3] KORG “KAOSSILATOR”  
<http://www.korg.co.jp/Product/Dance/kaossilator/>
- [4] Yu Nishibori, Toshio Iwai, “TENORI-ON”, Proceedings of NIME '06.
- [5] bastwood.com “The Aphex Face”  
<http://www.bastwood.com/aphex.php>
- [6] S. Malinowski “The Music Animation Machine” <http://www.musanim.com/>
- [7] G. Levin and Z. Lieberman, “Sounds from Shapes: Audiovisual Performance with Hand Silhouette Contours in The Manual Input Sessions,” Proceedings of NIME '05, pp. 26-28.
- [8] 岩淵 勇樹, 秋田 純一, 北川 章夫, “閉曲線図形に基づいた音色生成方法の検討,” エンタテインメントコンピューティング研究会, 2008
- [9] 港隆史, 関戸智史, 石黒浩, 河原英紀, “全方位視覚の特性を利用した画像から音信号への変換,” 日本ロボット学会第 20 回学術講演会予稿集, 2002.

作品・システム	ソース種別	判断材料	動作
解析信号曲線 (著者) UPIC TENORI-ON KAOSSILATOR Mouthesizer Manual Input Sessions Very Nervous System SoFA BigEye EyesWeb 音の種 (仮)	ペンタタブレット等 (曲線) ペンタタブレット (曲線) LED マトリクス タッチパッド リアルタイム画像 リアルタイム画像 リアルタイム画像 リアルタイム画像 リアルタイム画像 リアルタイム画像 各種物理センサー	解析信号にパラメータ変換した閉曲線 曲線の座標情報 座標情報 XY 各軸の座標値 口の形状 シルエットの特徴量 対象物のパーツ変位等 表情 (顔パーツのオプティカルフロー) 対象物の RGB 情報・大きさ 形状情報 生波形の周期化・周波数変換など (?)	オシレーター (VCO) ピッチ ピッチ ピッチ・エフェクト等 エフェクタ 音色生成・MIDI 音生成 汎用音声処理 エンベロープ等 汎用音声処理 (MIDI) 汎用音声処理 音色生成 (?)

- [10] 長嶋 洋一, “音楽/芸術表現のための新インターフェース,” 静岡文化芸術大学研究紀要, Vol.4, pp.137–151, 2004.
- [11] A. Camurri, M. Ricchetti and R. Trocca, “EyesWeb-toward gesture and affect recognition in dance/music interactive systems,” IEEE International Conference on Multimedia Computing and Systems, Vol.1, pp. 643-648, 1999
- [12] STEIM “BigEye”  
<http://www.steim.org/steim/bigeye.html>
- [13] J. M. Pelletier, “A Shape-Based Approach to Computer Vision Musical Performance Systems,” Proceedings of NIME '04, pp. 197–198.
- [14] D. Rokeby, “Very Nervous System,” <http://homepage.mac.com/davidrokeby/vns.html>
- [15] M. Funk, K. Kuwabara and M. J. Lyons, “Sonification of Facial Actions for Musical Expression” Proceedings of NIME '03, pp. 127–131.