

## 創作ノート

# Raveによるリアルタイム音色変換を用いた ライブエレクトロニクス作品の制作 Creation of Live Electronics Music using Timber Transfer with Rave

公文 太一, 小坂 直敏  
Taichi KUMON, Naotoshi OSAKA  
東京電機大学  
Tokyo Denki University

## 概要

バイオリンの入力音に基づいた水の音をエフェクトとするライブエレクトロニクス作品を制作した。エフェクトには Realtime audio variational autoencoder (RAVE) という機械学習モデルを利用した。

このモデルでは、事前に学習させた目的音と、入力音に基づき実時間での音色変換を行うことが可能である。本作では、事前に水の音を学習させたモデルに、バイオリンの音を入力し変換している。

本作では、RAVEを中心とし、各種エフェクトを Max8 により実装し、それを Ableton Live11 上でコントロールすることにより実現した。本作において試みた、水の音を使用したモデルの学習や、その実行のための max8 の実装、システムの構築について紹介する。

## 1. はじめに

2022年12月16日スミダトリフォニー小ホールにて、当大学未来科学部主催の Media project vol.16 が開催された。そこで初演を行ったライブエレクトロニクス作品 Flow において活用した技術や制作について紹介する。

### 1.1. 制作背景

本研究室では、デジタル表現ならではの新たなエフェクトとしてのモーフィングや音色変換合成を行っている。作品の構想初期段階から、音色変換を用いた作品の制作を考えていた。同様に機械学習を用いたコンピュータ音楽の制作についても研究を行っており、その両方の視点から作品を制作したいと考えていた。

また、実時間の処理に利用できステージパフォーマンスに活用できる技術を調査していた。

実時間動作可能な、RAVE は音色変換、機械学習両方の視点から興味深いモデルであった。そのため、これを利用したライブエレクトロニクス作品を制作することとした。

### 1.2. 関連研究

前述の通り本研究室では、電子音楽制作に用いられる、サウンドエフェクトの研究を行っている。その中で、エフェクトの知覚的な分類をするべきという考えから構造的音色の分類を作成した(小坂, 2021)。その中で、合成音による構造的音色の分類として、サウンドコラージュ、サウンドハイブリッド、サウンドモーフィングなどをあげた。サウンドハイブリッドは、複数の音から、それらの要素を組み合わせるといったものとなっている。また、ミクストミュージックにおいて演奏者の演奏音に対してこれをもとにしたサウンドコラージュ、サウンドハイブリッドが追従する疑似エコーというエフェクトを提唱している。本作のエフェクトはサウンドハイブリッドに Rave を用いた疑似エコーと位置付けられる。

## 2. システムの構成

本作は、バイオリンのメロディを入力音として、コンピュータによりリアルタイムに加工するライブエレクトロニクス作品である。システムはバイオリンと2台のコンピュータを用いて演奏させる。バイオリンは電子バイオリンを使用した。各システム間の接続図を示す(図1)。

### 2.1. ハードウェア構成

本作における音響処理は一台のコンピュータによって行われる。オーディオインターフェースには MOTU-

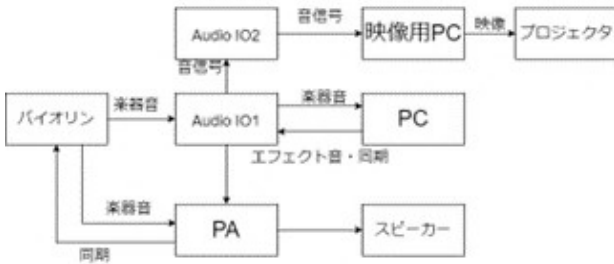


図 1: システムの構成

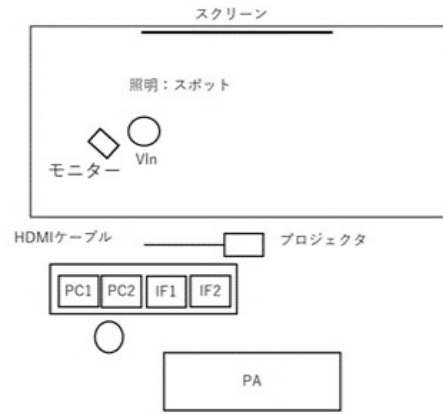


図 2: ステージ配置

M4(図1 Audio IO1)を使用した。こちらはバイオリンからの入力、ミキシングコンソールへの音声出力(2ch)、後述の映像用PCへの音声出力、演奏者への同期の出力に用いられる。

映像生成もコンピュータを一台使用している。音に反応する映像用のプログラムを実行するための入力用オーディオインターフェースには、Roland-UA55(図1 Audio IO2)を使用し、音響用のコンピュータから送られる信号を入力としている。映像はプロジェクタを使い会場のスクリーンに投影する。

## 2.2. ソフトウェア構成

エフェクトの実装には、Cycling74-Max8を使用している。RAVEによるエフェクトをはじめとして、リバーブディレイや、ドローン生成器などを実装した。

各エフェクトのパラメータコントロールや切り替えの利便性、システムの誤作動防止などの観点から各エフェクトをMax for Live(以下 M4L) デバイスとして実装し、Ableton-Live11 から実行する方法をとった。

また映像用のプログラムの実装にはDerivative-TouchDesignを使用した。

## 2.3. ステージ構成

ステージ構成は図2の通りである、演奏者は、ステージ下手で演奏を行い、その背面のスクリーン上に映像を投影する。オペレータはステージ下にて映像と音響のコントロールを行う。

## 3. RAVE について

### 3.1. RAVE とは

RAVE は Ircam - Antoine Caillon、Philippe Esling によって開発された音色変換を行う機械学習モデルである。ここでは、変換の目的となる音を目的音、変換の入力となる音を要素音と呼ぶ。Rave は、目的音となる音色を事前に学習することで、要素音から目的音への音色変換を実時間で行うことができるモデルである。

今回は、要素音が入力となるバイオリンのメロディ、目的音は水の音とした。図3は音色変換を示す図である。スペクトログラムの包絡を見ると要素音包絡が、出力に転写されていることが確認できる。

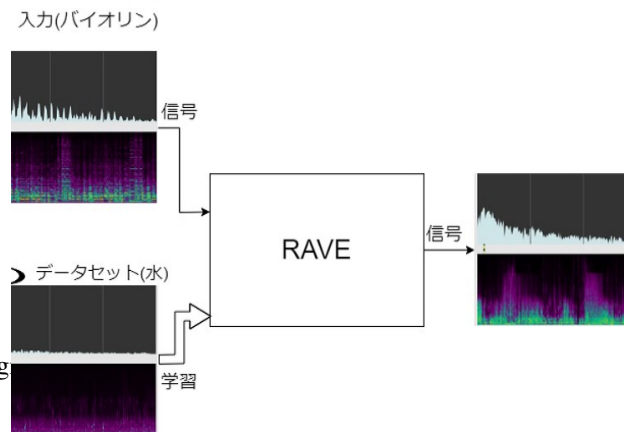


図 3: 音色変化のイメージ

### 3.2. RAVE の学習

本作では、RAVE に水の音を学習させたモデルを2つ制作した。一つ目のモデルは、フィールドレコーディングされた自然界の水の音 144 種類をデータセットで

ある。このデータセットは類似したサンプルが多く音色のバリエーションが少なかった。

2つ目のモデルは、食器やペットボトルなどを用いて、実際に録音した水音などを使用したデータセットで学習を行った。

学習の実行には Google-Colabratory(以下 colab)を使用した。Colab はブラウザ上にて、Python の記述や実行ができる環境であり機械学習などの用途に使用できる。Colab ではサーバの GPU を使用することができる。今回は主に Tesla-T4 を用いて学習を進めた。上記の環境を用いて各モデル 2.8Mstep、約 160 時間の学習を行った。

Rave は二段階の学習を行う。1 段階目では通常の VAE として学習され、次に敵対的調整が行われる。

図 4 が生成音に対する、評価の変動を示したグラフである。値が小さいほど高品質な生成が行えていることを示している。途中で評価が悪化しているようにも見えるが、学習方法の切り替わりのためで正常な挙動である。図 5 は 2 段階目の敵対的調整の損失のグラフであり、こちらは学習の進捗により、減少していることが確認できる。

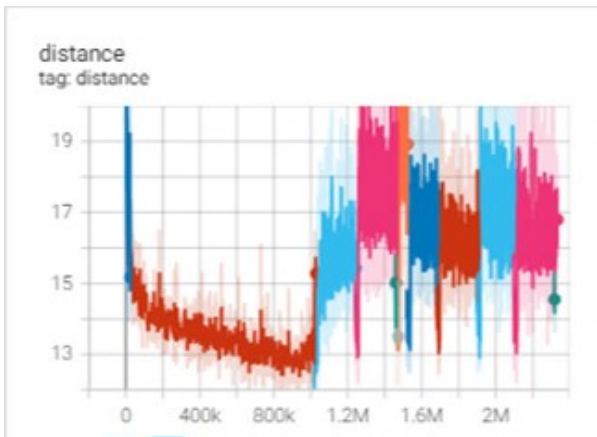


図 4: 生成物に対する評価

### 3.3. M4L としての RAVE の実行

RAVE は nn~ という Max8/puredata 用のオブジェクトが提供されている。Windows においてはビルドが必要である。Libtorch, visualStudio と C++tool, cmake を用いてビルドを行った。

nn~ を用いて Live11 内で Rave の推論を行うことができるデバイスを開発した。nn~ を利用しエンコーダ、デコーダとして読み込みこんでいる。エンコーダでは入力信号が潜在信号に変換する。デコーダでは、それを複合し音色変化が行われる。今回の実装では、デコーダへの入力の 1 つを LFO にしている。これは、事前の

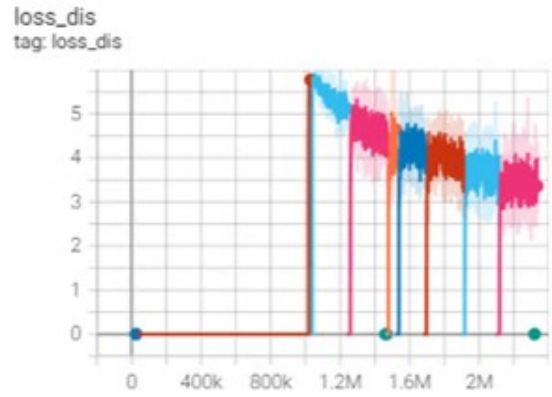


図 5: 敵対的調整の評価

実験により求める音がより出やすいように調整した結果である。

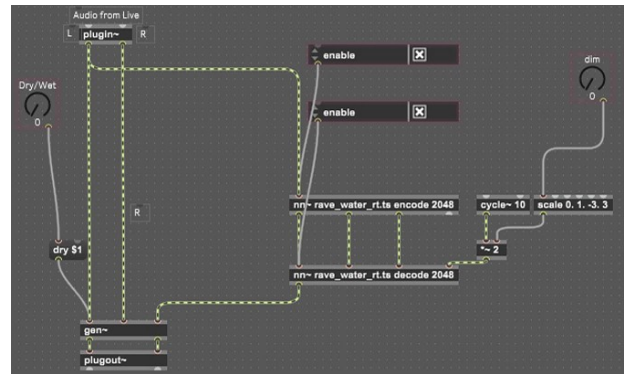


図 6: Rave を実行する M4L パッチ

## 4. その他のエフェクト

本作では、Rave を用いたエフェクトのほかにもいくつかのエフェクトを実装している。各エフェクトの詳細やパラメータについて示すが、コンサート当日の環境や奏者との打ち合わせにより調整を行う可能性があり、執筆時点から変化する可能性がある。

### 4.1. リバースディレイ

信号の入力をバッファに書き込み、それを逆から読み込むディレイである。代表的な製品として EARTH-QUAKER DEVICES-Avalanche Run Stereo Delay & Reverb がある。今回は、これを参考に Max8 によって実装した。

表 1: リバースディレイのパラメータ

パラメータ	値
右バッファの長さ	1小節のサンプル数 +0~500 sample
左バッファLの長さ	上に同じ
ディレイタイム	0~2000ms
ディレイの減衰率	0~1.0

このデバイスは、独立した 2ch のバッファとディレイによって構成されに 2ch での表現が可能になっている。

パラメータは Dry/Wet、ディレイの減衰率、ディレイタイム、左右バッファの長さとなっている。(表 1)。図 8 はエフェクトのダイアグラムとなっている。

左右のバッファの長さは読み込み時に、Live11 プロジェクトの 1 小節分となる。左右のバッファは 0~500 サンプル、長さを伸ばすことができる。バッファ長の差によって読み込まれる音にずれが生じる。これによりミニマルミュージックにおけるフェイズシフトのような効果を生み出すことができる。

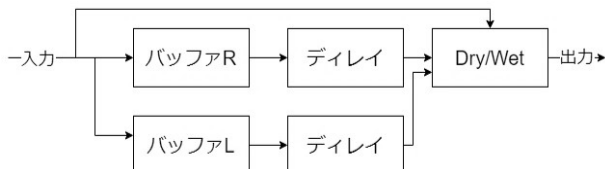


図 7: リバースディレイのダイアグラム

#### 4.2. ドローン生成

入力音の高さに従った低音のドローンを生成する。入力の音量に反応して FM 合成による音色の変化を実装した。エフェクトのダイアグラムは図に示す。パラ

表 2: ドローン生成のパラメータ

パラメータ	値
オシレータ周波数	バイオリンの入力に基づき、20~400Hz
FMの深さ	入力音量(線形)* (0~1000)
FM変調周波数	0~3000Hz
ポルタメント	0~500ms

メータはオシレータ周波数、FMの深さ、FM変調周波数、ポルタメントとなっている。(表 2)。

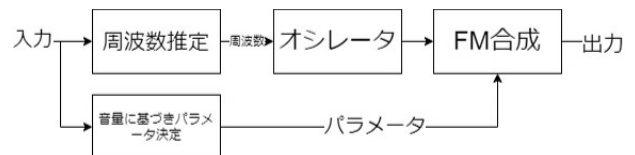


図 8: ドローン生成のダイアグラム

#### 4.3. 各エフェクトの接続

各エフェクトは Live11 内において平行に適用される。その後、各エフェクトをミックスして、コンプレッサ、リバーブなどからなるマスターエフェクトを適用して出力される。

### 5. 映像生成

本作は映像も伴う作品として制作した。短い映像を複数用意しその切り替えや、再生速度、輝度などを変更することによって映像表現を行っている。また一部 3DCG も取り入れた。

素材となる映像は Deform Diffusion を利用し生成した。これは画像生成モデルである Stable Diffusion を利用した映像生成を行うことができる。

## 7. 参考文献

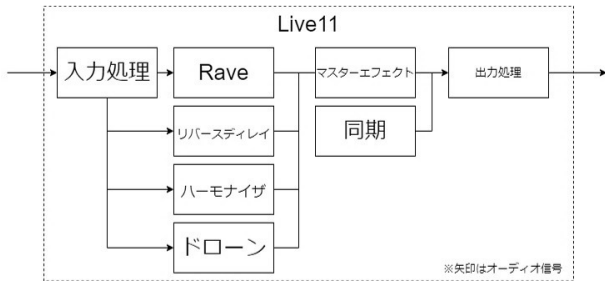


図 9: エフェクトの接続図

映像の再生管理、CG 合成を行うシステムは TouchDesigner を用いて実装した。

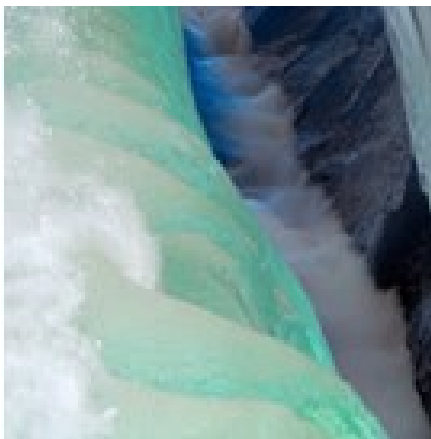


図 10: 生成映像の一部

## 6. まとめ

本作では、研究室のテーマでもある音色変換や、新たなエフェクトから、Rave による音色変換を用いて、バイオリンのためのライブエレクトロニクスを制作した。水の音を学習させた Rave を用いることでリアルタイムにバイオリンの音色から、水の音へ変換するエフェクトを実装し組み込んだ。

Max により各種エフェクトを実装し、Live11 でマニピュレーションを行う方法を採用した。

また拡散モデルによる映像生成を、利用したビジュアルも作成した。

Antoine Caillon, Philippe Esling (2021) “RAVE: A variational autoencoder for fast and high-quality neural audio synthesis” <https://arxiv.org/abs/2111.05011>(Accessed December 03, 2022).

acids-ircam, RAVE <https://github.com/acids-ircam/RAVE> (Accessed December 04, 2022).

acids-ircam, nn\_tilde [https://github.com/acids-ircam/nn\\_tilde](https://github.com/acids-ircam/nn_tilde) (Accessed December 04, 2022).

小坂直敏 (2021) “サウンドエフェクトと電子音楽制作” 音響学会春季研究発表会 1-9-13

Aaron Myles Pereira (2019) ” Reverse Delay Tutorial Max/MSP and gen~” <https://www.youtube.com/watch?v=S1cIKySow6Y> (Accessed December 04, 2022).

Deforum, Deforum Stable Diffusion <https://github.com/deforum-art/deforum-stable-diffusion>(Accessed December 03, 2022).

## 8. 著者プロフィール

### 公文 太一 (Taichi KUMON)

2000 年栃木県生まれ、2019 年東京電機大学未来科学部情報メディア学科入学。現在、同学科音メディア表現研究室に所属



この作品は、クリエイティブ・コモンズの表示 - 非営利 - 改変禁止 4.0 国際 ライセンスで提供されています。ライセンスの写しをご覧になるには、<http://creativecommons.org/licenses/by-nc-nd/4.0/> をご覧頂くか、Creative Commons, PO Box 1866, Mountain View, CA 94042, USA までお手紙をお送りください。