

創作ノート

RAVEを用いたモジュラーシンセサイザの制作 Modular synthesizer using realtime audio variational autoencoder

公文 太一, 小坂 直敏, 高橋 時市郎
Taichi KUMON, Naotoshi OSAKA, Tokiichiro TAKAHASHI
東京電機大学
Tokyo Denki University

概要

本研究では機械学習による音響生成を導入することによるモジュラーシンセサイザの拡張を目的としたモジュールの制作を行った。RAVEはVAE型の機械学習モデルであり実時間で音声信号を合成することが可能である。機械学習モデルを用いることにより従来ではサンプリング等でしか表現できなかった音色を合成することを可能にしている。ここではモデルの潜在空間を4chのCV(Control Voltage)やノブを利用して探索することにより音響を合成する。

1. はじめに

1.1. 電子音楽における録音と合成

電子音楽において用いられる音響を作成する手法として録音と合成がある。録音を用いた音楽はSchaefferを中心としたミュージックコンクレートに代表される。彼らの実験により再生速度の変更、逆再生など今日につながる技法が考案され、また創作の手段としての録音の可能性が提示された。合成においては19世紀末に登場したテルハーモウム以降多くの電子楽器が登場してきた。特にmoogによる電圧制御方式のシンセサイザの登場以降、多数の音楽家によって実験や制作が行われてきた。またポピュラー音楽にも広く利用されている。デジタル信号処理の発展以降これらの境界は以前よりもあいまいになったと考えられる。例えば大量の音の粒子を操作し音響を合成する手法であるグラニュー合成分がある。XenakisはこのアイデアをAnalogique A-B(Xenakis 1959)においてテープの切り貼りを用いて実現した。その後Roadsによりコンピュータへ実装され作曲に用いられた。さらにTruaxによって実時間動作が可能に改良された。これは録音をもとにしながらも、パラメータの操作により録音に対して動的な変化を加

えながら音響を合成できる点において録音と合成の間とも考えられる。

1.2. 機械学習技術による合成

近年では機械学習、特に生成モデルの発展が著しくまた社会的にも注目されている。それに伴いそれらを活用した作品も多く制作されている。様々なドメインのデータを生成可能なモデルが開発され続けているが今回は音響合成に着目する。音響の合成が可能な機械学習モデルとしてDDSP(Differentiable Digital Signal Processing)や今回用いたRAVE(Realtime audio variational autoencoder)(Caillon 2021)などがある。機械学習を用いた音響の合成は、録音されたデータに基づいている、一方でその出力は録音データの再生ではなく計算に基づく合成であり学習データと異なるデータを出力できるという点においてシンセサイザや録音芸術を拡張しうると考えられる。Scurtoらは機械学習技術による新たな音楽体験を目的として学習済みRAVEモデルの潜在空間と対応したVR空間を歩くことにより音響を合成するシステムを制作した(Scurto 2023)。またIntelligent Instrument LabのArmitageはライブコーディング言語のTidleCycles用のRAVE実行プログラムであるtidal-raveを実装しそれを用いたライブコーディングパフォーマンスを行った(Armitage 2022)。これらの例のように機械学習モデルによる音響合成が様々なプラットフォームにおいて試みられている。

1.3. モジュラーシンセサイザ

近年は特にユーロラック規格のモジュラーシンセサイザが注目されている。ユーロラック規格はDoepfer社が規格を整備しその仕様を公開したことで広まった。統一された規格によりさまざまなメーカーあるいは個人が制作したモジュールを自由に組み合わせてシステムを

構築することが可能となった。モジュラーシンセサイザの演奏においてはパッチケーブルによってモジュール同士を接続し CV(Control Voltage) と呼ばれる電圧を受け渡すことにより合成を行う。音響生成はオシレータやフィルタといった基本的な信号処理をこの接続によって組み合わせることによる合成が主体である。録音に着目したモジュールも存在しており例えば、テープレコーダのエミュレーションモジュールである Make Noise Morphogen(makenoise 2021) はミュージックコンクレートを意識したモジュールであることがその説明書などから確認できる。モジュラーシンセサイザと機械学習の組み合わせについては発展途上である。これは組み込みデバイスに用いられる計算機の性能が不足していることが理由として考えられる。一方近年ではエッジ AI と呼ばれる組み込みデバイスで機械学習モデルの推論を行う技術が登場している。この技術を用いて研究レベルにおいては Ircam-ACIDs による機械学習によりインパクト合成を行うモジュールである Neurorack(ACIDs 2022) や、筆者が制作したステップシーケンサによるキックのパターン入力に対してスネアやハイハットを生成するドラムシーケンサがある(公文 2023)。

以上の点に基づいて、モジュラーシンセサイザの拡張を目的として機械学習を音響合成に用いるモジュールの作成を行った。

2. RAVE について

2.1. RAVE の概要

RAVE は Caillon らによって提案された機械学習モデルであり、リアルタイムで音響を生成することが可能である。RAVE は波形を生成する VAE(variational autoencoder) である。このモデルのエンコーダは入力信号に対するマルチバンド分解と畳み込みニューラルネットワークで構成されている。デコーダは逆畳み込み層と残差スタックを用いた構成となっている。デコーダの出力は波形生成、ラウドネス生成、ノイズシンセサイザのフィルタの特性を生成するサブネットワークに与えられそれらを用いて最終的な出力を生成している。また通常の VAE としての学習を行った後、識別器を用いてデコーダ部分を敵対的に調整することにより品質を向上させている。このモデルを用いることにより音響合成や音色変換を行うことができる。また RAVE には Raspberry Pi 上での動作が可能な構成のモデルも存在している。これはパラメータ数の削減やノイズシンセサイザの省略により軽量化したモデルである。

2.2. nn~を用いた Pure data 内での利用

Max や Pure Data に RAVE を読み込み推論をするための nn~というエクスターナルオブジェクトが提供されている。これを用いることでオーディオを入力とした音色変換や、デコーダにパラメータを入力することによる音響合成を行うシステムをグラフィカルプログラミング環境を用いて構築することができる。

3. モジュールのコンセプト

3.1. コンセプト設定背景

モジュラーシンセサイザによる音の合成には、オシレータ、フィルタ、ディレイなど伝統的で比較的単純な信号処理機能を持つモジュール同士をパッチケーブルによって接続することで行われる。近年の市販モジュールはアナログ、デジタルのモジュールが混在しているがデジタルモジュールであってもその内部で行われる処理は古典的な信号処理の組み合わせであることが多い。また筆者らによる議論の中で「シンセサイザによる合成音は情報量が少ないのか」という議題が挙がった。これはシンセサイザによる合成は関数に基づいて数学的に表現可能であり、その性質から声や自然音等と比べてシンセサイザによる合成音は単純なのではないかという問いであった。これらに基づいて機械学習を用いることで従来のモジュールの組み合わせでは表現が難しかったドメインのサウンドを機械学習を用いて合成することによりシンセサイザを拡張するという基本方針が定まった。また筆者は機械学習を利用したインタラクティブな音楽について研究、制作を行ってきた。今回制作したモジュールもリアルタイムに動作する機械学習モデルを利用しインタラクティブに演奏できるものである。

3.2. コンセプト

前項の内容や筆者の研究・作品に通底している機械学習とのインタラクションという点を中心に本器のコンセプトとして「機械学習による合成を導入することによるモジュラーシンセサイザ拡張」と設定した。

4. 潜在空間の探索による生成

VAE は潜在変数が正規分布に従うと仮定しデータの特長を潜在空間に圧縮する。また潜在空間の点をサンプリングすることによりデータを生成することができる。図 1 は生成方法のイメージ図である。図は 3 次元の潜在空間の場合の例であり、3 つの入力数値により潜在空間上の点を指定してその値をデコーダへ入力することにより音響を生成する例である。図では 3 次元

の潜在空間の例として示したが、今回実際に用いるモデルは 8 次元の潜在空間を持ち 8 次元の入力に基づいて生成を行っている。今回制作したモジュールにおいては潜在空間上の点を CV の入力値とノブによってコントロール可能な LFO を用いて潜在空間を探索することによって音響の生成を行う。VAE は事前分布に標準正規分布 (図 2) を仮定して学習を行うため潜在空間上の点の値の大部分は -3 から 3 の範囲に収まる。この点を考慮し CV と LFO のマッピングを行っている。

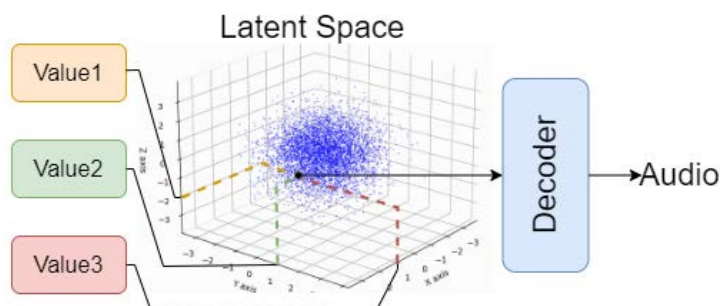


図 1: 潜在空間からの生成のイメージ

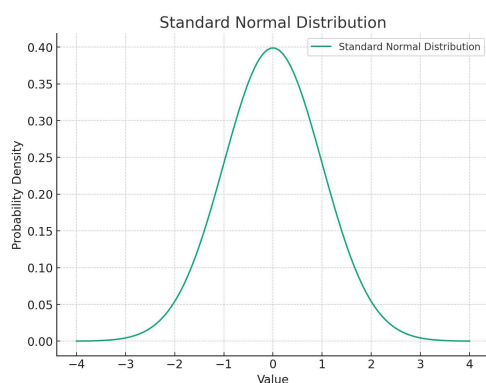


図 2: 標準正規分布

5. デバイスの実装

ここではモジュラーシンセサイザの CV とノブによって RAVE の潜在空間を探索するモジュールを設計しユーロラック規格で実装した。

5.1. ハードウェア構成

モジュールのハードウェア構成を図 3 に示す。デバイスの本体は Raspberry Pi4 Model B 8GB である。これに 2 基の ADC と DAC が搭載された基板を接続している (図 5)。モジュールのインターフェースを図 4 に示す。パネル上に入力として 4 つのノブと、4ch の CV 用

の 3.5mm ジャックがある。電源は一般的なユーロラック電源に対して Raspberry Pi の消費電力が大きいため別途 USB type-C ケーブルとアダプタを利用して供給することとした。

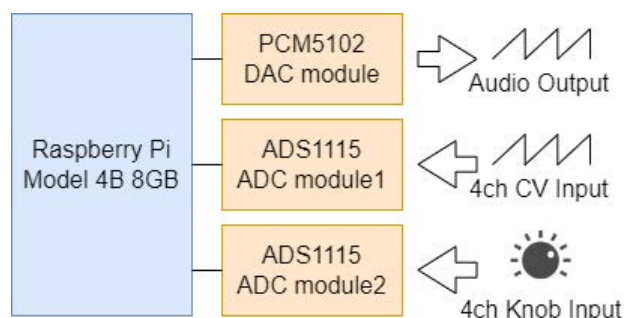


図 3: ハードウェアの接続図

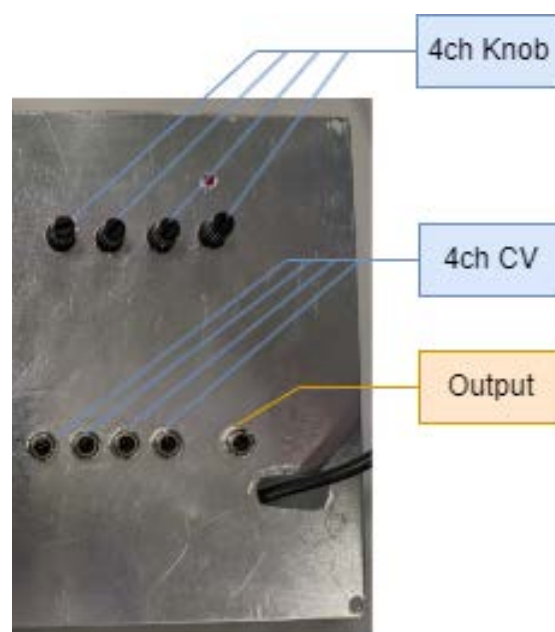


図 4: モジュールのインターフェース

5.1.1. CV, ノブによる入力

CV とノブの入力用 ADC として Texas Instruments ADS1115 をモジュール化した製品を利用した。この ADC は 4ch, 16bit の入力を持つ。ADS1115 は差動入力に対応しているが今回はシングルエンド入力として利用している。そのため実質的には 15bit で電圧の読み取りを行っている。接続インターフェースは I2C (Inter Integrated Circuit) となっていてこれを用いて Raspberry Pi と通信を行う。この ADC を 4ch の CV 用と 4 つのノブ用の 2 系統用意した。また Raspberry Pi の動作電圧は 3.3V であるが ADC は 5V にて動作させるため I2C

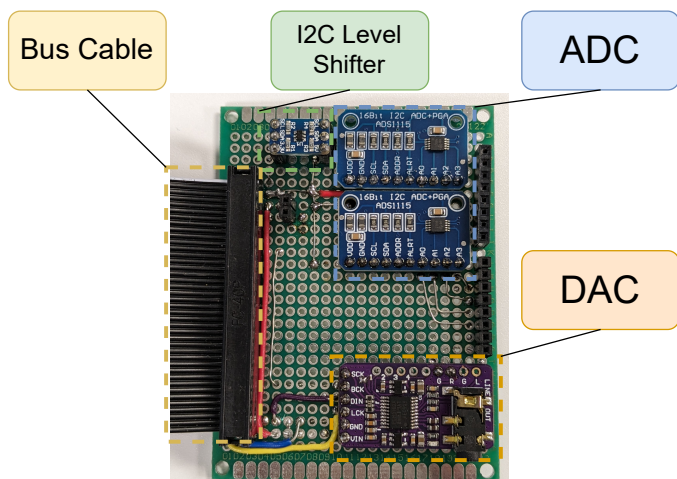


図 5: ADC,DAC を実装したモジュールの基板

のレベル変換モジュールを用いてこの変換を行った。図 5 は CV の入力回路である。ジャックに入力された CV はショットキーバリアダイオードを用いて電圧を 0V-5V にクリップする保護回路を通して ADC に入力される。本器はこの入力回路を 4 つ持ち ADS1115 のそれぞれの入力に接続されている。

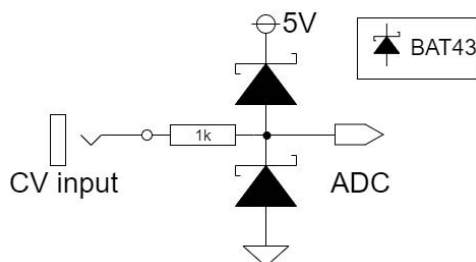


図 6: CV 入力回路図

図 6 はノブの入力回路である。20K オームの可変抵抗器によって分圧された電圧を ADC で読みとっている。

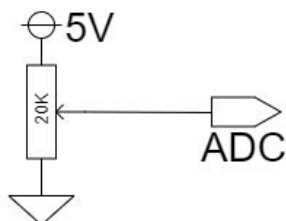


図 7: ノブによる入力回路図

5.1.2. DAC による出力

本器の音響出力用 DAC には Texas Instruments PCM5102 をモジュール化したものを利用した。I2S(Inter-IC Sound) を用いて Raspberry Pi と接続することにより 1ch,44.1kHz,16bit の音響を出力している。こちらは Raspberry Pi にドライバが用意されているためその設定を行い接続を行うことで出力が可能になる。

5.2. ソフトウェア構成

本器のソフトウェアは ADC と OSC(Open Sound Control) を管理する Python プログラムと、前述の nn~ が中心の Pure data パッチから構成されている。Python プログラムは I2C と OSC のホストとなり ADC が読み取った値を OSC のプロトコルに従い Pure data へ送信する。Python プログラムはモジュールの起動と同時に実行され以下の動作を行う。

1. I2C による ADC との通信を開始
2. Pure data を起動
3. OSC のホストとして通信を開始
4. 2 基 ADC の入力値を読み取り
5. 読み取ったデータの OSC メッセージ化

Pure data のプログラムは nn~と LFO を中心とした構成となっておりノブの入力と CV の 1ch は LFO の周波数に接続されている。残りの 3ch の CV は nn~に直接接続されている。LFO の周波数は 0V から 5V の入力電圧によって 0Hz から 20Hz または 0 から 5Hz の範囲にコントロールされる。また LFO の振幅は前述の潜在空間の特性に基づいて 2.9 である(式 1)。また nn~に直接接続されている CV は 0V から 5V の入力電圧を-3 から 3 の範囲に変換している。

$$\text{Freq} = r \times \max \left(0, \min \left(\frac{\text{Input}}{32767}, 1 \right) \right), r = 20 \text{ or } 5 \text{ (式 1)}$$

nn~から出力された音声は DAC~オブジェクトより出力される。Pure data は Python から起動されたのち以下の動作を行う

1. DAC の起動
2. OSC メッセージを受け取り解読
3. 受け取った各値を nn~または LFO のパラメータとして入力
4. nn~による音響合成
5. DAC からの出力

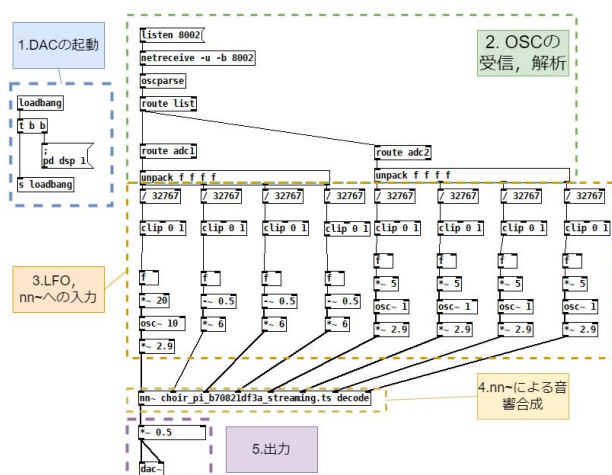


図 8: 作成した Pure data パッチ

5.3. RAVE モデルの学習

本モジュールに搭載するためのモデルの学習を行った。データセットとして約1時間の歌声や話し声のデータを Raspberry Pi 用の RAVE モデルに学習させた。学習は Raspberry Pi では行わず一般的なコンピュータを用いている。CPU に AMD Ryzen7 5800X, GPU に Nvidia Geforce RTX3080 を搭載した windows マシンにより約 60 時間の学習を行った。

6. デバイスの動作

制作したモジュールをユーロラックモジュラーシステムに組み込んだ様子を示す。制作したモジュールを市販のモジュールと組み合わせることにより、モジュラーシンセサイザの演奏に機械学習モデルによる合成を取り入れることができた。一方でノブや CV によって狙い通りに出音をコントロールすることの難易度は高かった。これは機械学習モデルの潜在空間が人間にとって理解可能な形式でないことが理由であると考えられる。潜在空間の次元ごとに出音に与える影響の大きさ等が異なるため、インターフェースの設計や数値のマッピング方法検討する必要がある。

7. まとめ

モジュラーシンセサイザの演奏における音響の録音と合成の拡張を目的として RAVE を音源としたモジュールの制作を行った。本モジュールでは CV による入力とノブによる操作が可能で LFO を用いてモデルの潜在空間を探索することにより音響の合成を行う。Raspberry Pi4 を中心として ADC, DAC などのハードウェア, Python プログラムと nn~ を中心とした Pure data パッチからなるソフトウェアによってユーロラックモ

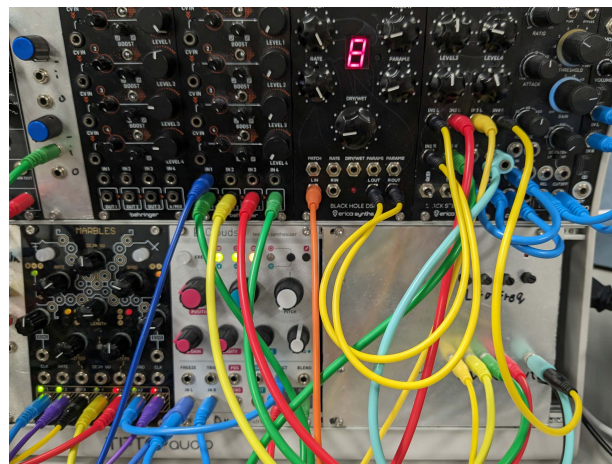


図 9: 作成したモジュール (右下) をケースに組み込んだ様子

ジュールとしての実装を行った。これにより機械学習を音源としたモジュールによるモジュラーシンセサイザの拡張と、CV やノブによる潜在空間を探索による合成を実現した。一方で潜在空間は人間にとって意味のある空間ではないためその間を埋めるインターフェースや数値のマッピングについて検討する必要がある。

8. 謝辞

JSSA53 回研究会と併設して開催される ICSAF2024 の参加に際しまして、東京電機大学理工学部、柴山拓郎先生にご協力いただきました。この場を借りて感謝申し上げます。

9. 参考文献

- Curtis Roads, *The Computer Music Tutorial*, 1996. 青柳, 小坂, 平田, 堀内訳・監修 後藤, 引地, 平野, 松島 訳. 『コンピュータ音楽 歴史・テクノロジー・アート』(東京電機大学出版局, 2001.)
- Antoine Caillon, Philippe Esling. 2021. RAVE: A variational autoencoder for fast and high-quality neural audio synthesis (<https://arxiv.org/abs/2111.05011>) 2024 年 1 月 29 日閲覧.
- Hugo Scurto, Ludmila Postel. 2023. *Soundwalking Deep Latent Spaces* nime, Mexico City, Mexico (<https://arxiv.org/abs/2111.05011>) 2024 年 1 月 29 日閲覧.
- Intelligent Instruments Lab. *Release of RAVE Models Repository* (<https://iil.is/news/ravemodels>) 2024 年 1 月 29 日閲覧.

Acids - Ircam. 「Neurorack」(<https://acids.ircam.fr/neurorack/>) 2024 年 1 月 29 日閲覧.

公文太一, 小坂直敏, 高橋時市郎. 2023. 「機械学習によるパターン生成を用いたモジュラーシンセサイザ向けドラムシーケンサ」『先端芸術音楽創作学会会報』 Vol.15, No.2, pp.1-5

makenoise 『MORPHAGENE』 2024 年 1 月 29 日閲覧. (<https://www.makenoisemusic.com/content/manuals/morphagene-manual-jp.pdf>) 2024 年 1 月 29 日閲覧.

10. 参考作品

Iannis Xenakis.1997. 『Analogique A-B』.

Jack Armitage. 2022. 『live coding speech RAVE models』.

11. 著者プロフィール

公文 太一 (Taichi KUMON)

2000 年生まれ, 2023 年より東京電機大学未来科学研究科情報メディア学専攻, ビジュアルコンピューティング研究室所属.



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/> をご覧頂るか、Creative Commons, PO Box 1866, Mountain View, CA 94042, USA までお手紙をお送りください。