

## 共感覚に基づく視覚化によって直感的な音作りを可能にするシンセサイザー — 音色が「見える」シンセサイザーの開発 —

### 1. 背景

シンセサイザーは豊かな音色を奏でられる一方、操作によって目的の音色を実現する「音作り」は非直感的でユーザの負担が大きいという課題がある。

従来の音作りの方法の一つは、シンセサイザーに内蔵されている音色、いわゆるプリセットから選択するという方法である。しかし、プリセット名からどのような音色かを想像することは困難なため、試聴作業を延々と繰り返すことになり、時間を奪われる。

もう一つは、音色のパラメータを直接編集するという方法である。しかし、パラメータ名の意味を理解するためにはシンセサイザーの発音方式に関する専門的な知識が必要とされる。その上、パラメータと音色の関係は複雑であるため、どの値のときにどのような音色になるのかを瞬時に想像することは熟練なくしては困難である。

### 2. 目的

本プロジェクトでは、音作りのユーザインタフェースの直感性を向上するためのアプローチとして、音色の視覚化に焦点を当てる。これは、音という実時間メディアを、画像という一覧性・俯瞰性の高いメディアに変換することに他ならない。音色が一目で把握できる新たなシンセサイザーを製作することによって、音作りにおけるユーザの負担を軽減し、より楽しく音作りができるようになることを目指す。

### 3. 開発の内容

本プロジェクトで開発したシステムを一言で表すと、音色が「見える」シンセサイザーである(図 1)。本システムを構成するソフトウェア、ハードウェア、画像推定モデルの製作に取り組んだ。なお、発音部分はオープンソースの FM シンセサイザーエミュレータを利用した。



図 1: 開発したシンセサイザーの演奏風景

ユーザは最初に、音色に画像をアノテーションすることを求められる(図 2)。30 個の音色が提示されるので、それぞれを聴きながら自身の共感覚(感覚間協応)に基づいて合致すると思う画像を回答する。本システムでは音色は球に突起を生やした画像で表現され、突起の数や長さといった形状に関するパラメータと、金属感と光沢感という材質に関するパラメータの、計 7 つのパラメータを指定することができる。

プリセット選択画面では、ユーザの回答結果に基づいて各プリセットの音色を視覚化した結果が表示される(図 3)。時間変化する音色については、先頭 1 秒間のパラメータ変化から推定された一連の画像をアニメーションとしてループ再生する。また、音量の時間変化は拡大・縮小、音高の時間変化は回転によって表される。

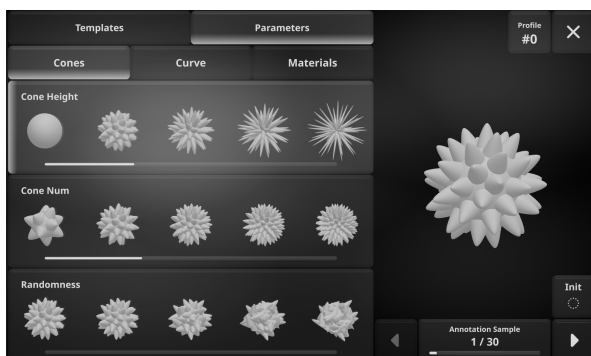


図 2: アノテーション画面

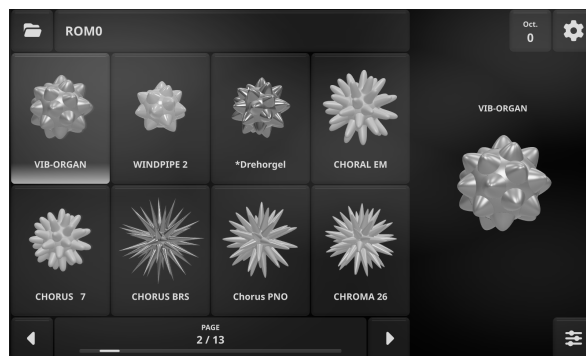


図 3: プリセット選択画面

パラメータ編集画面では、パラメータの値を指定するスライダーの各位置に、その値に対応する音色を視覚化した結果が表示される(図 4)。各パラメータの効果は他のパラメータの値との関係によって変わりうるため、視覚化の結果は他のパラメータの値の変更に応じてリアルタイムに更新される。

また、音色パラメータの生成モデルによるバリエーション生成機能も利用できる(図 5)。画面上のスライダーは、中心が元の音色に対応しており、左右に動かすことで音色の連続的なバリエーションが得られる。ここでも、それぞれのバリエーションが視覚的に提示される。

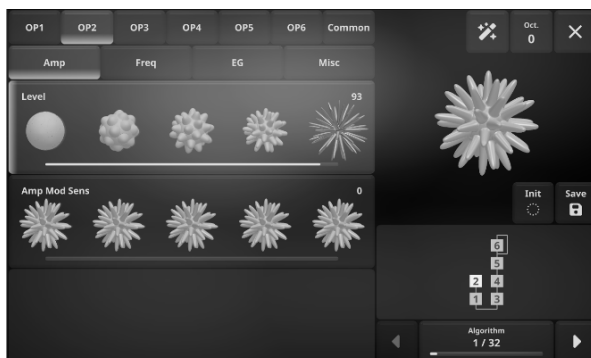


図 4: パラメータ編集画面

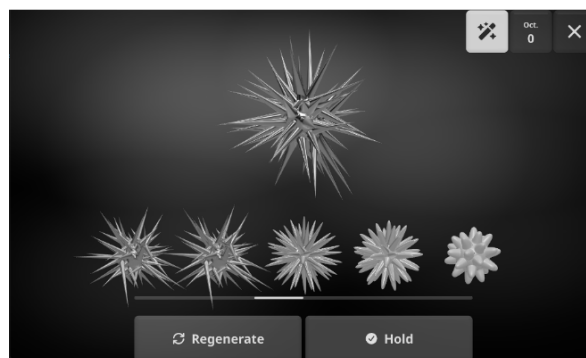


図 5: バリエーション生成画面

以上のユーザインタフェースにおける音色の視覚化は、深層学習モデルによる画像パラメータの推定によって実現している。大規模事前学習モデル CLAP に基づく音色の聴覚的

類似度の予測を利用し、聴覚的に類似する音色は類似する画像パラメータで表現されるという仮定のもとで画像パラメータを推定するアルゴリズムを開発した。加えて、推論パイプラインを Transformer のエンコーダ構造に基づく小規模なモデルで蒸留することによって高速化を行い、小型のコンピュータでも実用上支障のない速度での推論を可能にした。

また、音色の視覚表現の推定や表示を前提とした新たなシンセサイザーのハードウェアの設計・組み立てを行った。主要な処理を行うシングルボードコンピュータとして Jetson Orin Nano Super を搭載している。これは画像の 3D レンダリングや深層学習モデルの推論を GPU 上で高速に実行するためである。インターフェースとしては、5 インチタッチスクリーン、鍵盤ユニット (Roland K-25m)、数値入力用ノブ、スピーカ、オーディオ出力端子等を備えている。筐体の造形には 3D プリンタを利用した。

#### 4. 従来の技術(または機能)との相違

本シンセサイザーのユーザインターフェースは、従来のユーザインターフェースと比べて、音作りにおけるユーザの負担を軽減する面で優位性がある。

プリセットを選択するユーザインターフェースでは、プリセット名だけが表示される従来のユーザインターフェースとは異なり、各プリセットに対応する音色の視覚化の結果がアイコンとして表示されることで、どのような音色かを試聴前に一覧することができる。これによって、視覚的に把握した音色の中で気になったものだけを試聴することが可能になる。

また、音色のパラメータを編集するユーザインターフェースでは、視覚化の結果がスライダー上の対応する位置に表示されることで、パラメータと音色の関係を直感的に把握することができる。これによって、各パラメータの名称と効果を覚えていなければいけない従来のユーザインターフェースとは異なり、パラメータに対する深い知識がなくても音色に予測可能な編集を加えることが可能になる。

#### 5. 期待される効果

楽曲制作やサウンドデザインの領域で、これまで十分に使いこなされていなかったプリセットやパラメータの活用を促進することが期待される。これにより、単に作業効率が向上するのではなく、多様な音色と出合うきっかけが与えられることで音楽表現そのものの幅を広げることが可能であると考えられる。

さらに、本システムはシンセサイザーの学習用途でも有用性が期待される。プリセット名の非直感性やパラメータの難解さといったハードルを下げ、音作りに入門しやすくなることによって、より多くの人々とシンセサイザーの楽しさを共有することが可能になると考える。

#### 6. 普及(または活用)の見通し

本システムのターゲットは、シンセサイザー奏者やサウンドクリエイター、アーティストといったプロから、シンセサイザー入門者や趣味でシンセサイザーを楽しむ者まで多岐にわたる。また、本システムの核となる技術は、既存の鍵盤あるいは PC 上の DAW と接続して使用できるモジュール型デバイスや、DAW 上のプラグインなど、より普及しやすい形式にも展開が可能であると考えられる。本システムに対するフィードバックを継続して収集すると同時に、このような普及しやすい形式での提供も検討しながら、長期的な視点でシンセサイザー文

化に影響を与えることを目指す。

7. クリエータ名(所属)

真家 彩人(東京大学 工学部 システム創成学科 知能社会システムコース)