

特集「AIと美学・芸術」

# 人工知能と音楽

—音／音楽自動生成システム：サウンドロイドの開発を通して—

## Artificial Intelligence and Music: Development of a Sound/Music Generation System “Soundroid”

古川 聖 東京藝術大学  
Kiyoshi Furukawa Tokyo University of the Arts.  
furukawa@fa.geidai.ac.jp, <http://furukawalab.org>

大村 英史 東京理科大学  
Hidefumi Ohmura Tokyo University of Science.  
hidefumi.ohmura@gmail.com, <https://sites.google.com/site/hidefumiohmura/>

濱野 峻行 東京藝術大学  
Takayuki Hamano Tokyo University of the Arts.  
hamano.takayuki@fa.geidai.ac.jp, <https://takayukihamano.net>

川村 剛 TAJISOFT, RF ルーカス  
Tsuayoshi Kawamura TAJISOFT./RF Locus Inc.  
kawamura@tajisoft.jp, <https://tajisoft.jp>

柴山 拓郎 東京電機大学  
Takuro Shibayama Tokyo Denki University.  
takuro@mail.dendai.ac.jp, <https://www.facebook.com/takuro.shibayama.14>

**Keywords:** artificial intelligence, deep learning, computer music, generative music, music emotion, music cognition, ambient music.

### 1. はじめに

本稿では私達が開発している人工知能技術を応用した音／音楽の自動生成システム「サウンドロイド」を中心に、その内容やコンセプト、開発において出合った諸問題の検討を通して、音楽と人工知能について書いてみたい。この開発プロジェクトは三井不動産と共同でビジネスに適した音環境を提供するという明確な目的のもと、東京藝術大学古川 聖を中心としたさまざまな芸術家や研究者が集まった研究グループで行われてきた。音楽の自動生成の研究と音楽と情動に関するいくつかのアイデアが結び付きスタートされ、2018年11月からβ版の一般公開を始めた (<https://soundroid.com/>)。音／音楽を自動生成するためには、生成された音／音楽とそれに対する人間側の反応・評価を関係付けるデータ

が不可欠で、三井不動産との共同作業は重要な意味をもつ。現時点でサウンドロイドは仕事の能率化という限定された目的をもつが、このプロジェクト自体はもっと広い射程をもち、私達を取り巻くさまざまな状況に対応する音／音楽の自動生成システムを志向している。2019年の4月に正式版のリリースを行う予定である。

### 2. サウンドロイドの概要

「音／音楽の自動生成」のように音と音楽が並べて書かれているのは、サウンドロイドが生成し出力する音が、いわゆる音楽と環境・生活音（街の音、雨の音、風の音、etc.）、この両者の中間にあるようなタイプのものをイメージしているからである。それはポップソングや古典的クラシック音楽のように構成された音楽作品ではなく、また上記の環境・生活音のように私達の日常を取り

巻く音でもない。比較的、環境音楽やアンビエントと呼ばれる音楽に似ているが、これらには何らかの表現者、作者が存在する。これらは録音され固定されているか、またリアルタイムに生成されるにしても、そこに作者側の意図や表現が含まれ、ある一定の形をしていることが多い。一方で、サウンドロイドが生成する音/音楽は、ユーザ主体で生成され変化していく不定形なものである。サウンドロイドの特徴を以下のようにまとめることができる\*1。

- 1) ユーザからのデータをもとに音/音楽をリアルタイムで自動生成するアプリケーションである（現在はPC上で動作、スマートフォン版も準備中）
- 2) サウンドロイドの現在のバージョンには仕事に集中するための音環境をつくるという明確な目的が設定されている。
- 3) サウンドロイドはユーザが自分の目的のために、システムとともに自分で音環境をデザインしていく仕組み自体であり、100人100様に音/音楽がパーソナライズされる。
- 4) サウンドロイドは音/音楽の自動生成のためのパラメータをユーザが設定している。また、サウンドロイドを利用時における人間の反応や評価をサウンドロイドはデータとして取得している。「徐々に成長していくシステム」のような長期的なイメージのもと、プロジェクトは進められている。β版公開運用以降にデータ収集を行い、段階的にデータ分析や機械学習の手法を応用していく予定である。

### 3. 自動作曲と人工知能

音楽史において音楽の生成の形式化は主に作曲技法という形で常に行われてきたが、その延長線上で現在、コンピュータ音楽の中に自動作曲、アルゴリズム作曲という領域が形成されている。サウンドロイドが実際に行っていることは、自動作曲、アルゴリズム作曲のことで、その歴史の流れを見ながらサウンドロイドの音楽史の中の位置を確認し、特質を明らかにしたい。

#### §1 コンピュータ以前

音楽をつくることは人間的な行為であり、作曲者がい

ることもあるし民謡のように作者不詳でも音楽はほとんどの場合、人間によって生み出されてきた。その一方、コンピュータが現れるずっと前から、音楽の一部を形式化（自動化、機械化、手続き化）することは行われていた。ヨーロッパ中世、アルスノバ（例えばG. マシヨー）のイソリズムックモテットの技法 [Bent 01] や、モーツァルトの「音楽のサイコロ遊び」 [Nierhaus 09] などはその例である。近代になればB. バルトークの黄金分割（音程と時間軸の構成の比率に応用）、そしてA. シューンベルク（12音技法）から始まりP. ブレーズ（トータルセリー）に至るセリエルミュージックなど、枚挙にいとまがない。

#### §2 コンピュータ以後

コンピュータが現れるとまず、L. ヒラーはいくつかの習作の後、1957年にイリノイ大学のILIACコンピュータを使って計算された、弦楽四重奏のための「イリアック組曲」（主にマルコフ連鎖と乱数生成による）を作曲した [Hiller 59]。ほぼ同時にI. クセナキスの名作、オーケストラのための「メタスタシス」（確率論により音楽を形式化、楽譜データを出力、この場合人間が手で計算）がつくられ、その後の彼の連年のコンピュータによる作品の制作につながっていく。コンピュータはまずは研究機関に導入され、しだいに個人にまで普及し手軽に使えるツールとなり、現在までに、商用、芸術用を問わず、さまざまなモデルに基づくアルゴリズム作曲、自動作曲が行われてきた。この辺に関してはC. ローズのComputer Music Tutorialの18～19章に詳しい [Roads 96]。本稿の著者の一人である、古川も1990年代にフラクタルなどの非線形構造の自己組織化のプロセスを音構造にマッピングした作品群 [古川 07] をつくっている。

#### §3 コンピュータと人工知能

音楽を形式化する、つまりシンボルや数を使って、人間からは独立した音楽の知識表現を行うような自動作曲、アルゴリズム作曲という行為自体、人工知能を志向しているともいえる。数的な構造のように全く音楽と関係ない構造をモデルとして使用する以外、現在または過去の音楽の構造（メロディー、対位法、和音、音階、モード、リズム、分節化、階層性 etc.）を形式化、音楽を知識表現化するようなアルゴリズム作曲はすべて、人工知能と関係しているといえる。人工知能研究の変遷、発展と並行し、それらの成果を生かした音楽研究、音楽制作は常に継続的に行われてきた。ニューラルネットワーク、制約プログラミング、パターン照合検索、フラクタル、1/f、ベイズ統計学、形式文法（生成文法を含む）など、さまざまな試みが行われている。過去においてはやモーツァルトやショパンなどの旋律法、和声法、対位法、楽曲形式などを知識表現し、人間が事細かにルールを入力しモーツァルトやショパンのような音楽を生成するエキスパートシステムもあり、D. コープのものが有名だ [Cope 05]。270ものルールを入れ込んだ、K. Ebcioğluの1980

\*1 類似の音楽自動生成アプリケーションはすでに多くあり、Brain.fm (<https://www1.brain.fm/about>) が概念的にはSoundroidに近いが、ユーザの状態、環境情報を取らないで一方的に音楽を提供している点がSoundroidとは異なる。Jukedeck (<https://www.jukedeck.com/>) は人工知能がもたらす最先端技術をベースにしており、ニューラルネットワークを訓練し、創造性を助ける制作ツールを構築している。ほかにマルコフモデルを使用するFlowMachines/Sony CSL (<http://www.flow-machines.com/>)、Ampermusicβ版 (<https://www.ampermusic.com/>)、GoogleのリサーチプロジェクトでTensorflowを用いた自動音楽生成アプリケーションMagenta (<https://github.com/tensorflow/magenta>)、そのほか、Computoser、Bozhidar Bozhanov (<http://computoser.com/>)などをあげておく。

年代の研究である J. S. バッハのコラールの和声付けシステム [Ebcioğlu 86] を体験したことがあり、その完成度に驚かされたが、ドイツの音楽大学で J. S. バッハの和声法を学習、研究した筆者（古川）にはオリジナルとの差異はすぐにわかる程度のものであった。過去に行われ、現在は私達の生活の中に人工知能として入り込んできているエキスパートシステムと同様、この音楽用のエキスパートシステムは自分で知識を獲得できないことが弱点で成果は限定的であり、現在、音楽においてこの方向での研究はあまり大きな展開は見られない。270 程度のルールでは一人の作曲家のスタイルのモデルとしては張りぼてのようなもので、フレーム問題は解決されておらず、応用のきかない“何か”に止まっていると思う。

そして現在、再び音楽と人工知能というトピックスが活気付いている。その理由は、人工知能の応用の諸分野が活気付いているのとはほぼ同じもので、機械学習技術の発展によりビッグデータを扱う研究領域における新たな展開にある。さらに音楽に関しては特に脳科学の進展により、音楽の脳内認知のプロセスがより詳しくわかってきたこと [Koelsch 12]、それに相まって、音楽の重要な内実である音楽情動のメカニズム [Juslin 11] の解明が大きく進んできたことにもよる\*2。サウンドロイドはこのような人工知能と脳科学の新たな展開の中、始められた。

#### 4. 人工知能による音楽生成がもつ問題

##### §1 音楽の全体性

録音技術が生まれるまでは音楽の向こう側には必ず、それを演奏したり歌ったりする人間がいて、それは人間を含む一つのイベントであり、全体性をもったコミュニケーションの場でもあった。録音技術、音楽の複製技術がこのようなイベントから、音だけを引きはがしてやることを可能とした。私達はこの全体性のことを忘れがちで、この引きはがされた音の部分だけを指して音楽と呼んだりしている。ジャスティン・ビーバーでも AKB 48 でも、彼らの歌唱パフォーマンスにおいて、歌詞、メロディ、リズム、ハーモニーなどの曲自体は、そのパフォーマンスの一部であり、メロディー（曲）は彼らの生き方、メッセージ、ファッション、容姿、アイドル性への情動、共感を乗せるメディアであって、ヒットの直接の原因がメロディーや音楽構造にはあるわけではない。曲、つまり音楽構造は客観的に見ればどれもこれも 4 小節構造をもち、ほぼ予想どおりの場所で半終止、全終止を繰り返す典型的なものである。逆さまから見れば、アイドル性を乗せるメディアはできるだけありきたりで明快で特別なことはせず、典型的で違和感のないものが良いといえるだろう。

「人工知能で音楽」というときにこの音楽の全体性が忘れられ、音楽＝曲自体という意味で言葉が使われ議論が錯綜している。1000 個のヒット曲のメロディを機械学習にかけたら究極のヒット曲ができるだろう、人工知能が無限に適切な音楽を生成するだろう、などという、音楽のもつ全体性を考慮しないトンチンカンな話 [ケリー 16] を読んだり聞いたりするが、そもそも論理にも飛躍がある。良いものを 1000 個集め機械学習にかけたら最高のものができるだろうか、人気ラーメン店 100 店のレシピから究極のラーメンがつくれるだろうか、現在のところヒット曲を生み出すためには作詞家、作曲家ではなく有能なプロデューサーの感性、そしてその人が作り出すアイドルとして十分な魅力のある歌手をめぐるストーリーの組立てが一番重要なファクタとなるだろう。

##### §2 目的の設定

さて、ここから上記の音楽の全体性から切り離した、引きはがした、音／音楽の構造の話に移る。前に書いたように音楽用のエキスパートシステムでは、「ショパンのような」などの目標があり、ショパンの楽曲をデータとして使えば、仮の評価関数はつくれるので、時間をかけて音楽構造の生成のルールを人間が入力すれば、それなりの結果を得ることができるだろう。何はともあれ人工知能で音楽を生成するためには「音楽」という漠然とした言い方ではなく、まずは「……のような音楽」という具体的・限定的な言い方目標の設定が必要で、それも「深遠な」音楽というより「明るい」、「楽しくなる」というような、より情動的な設定が必要だろう。とはいえ実際には、どのような音楽の諸要素の組合せが「明るい」音楽、や「楽しい」音楽に対応するのかに関しては、ほとんど初期的な研究 [Juslin 11, Stern 85] しかないの、現在のところ満足な目的関数を書くことができない。つまり、システムに十分な人間の情動と音楽の構造のルールの関係すらシステムは使うことができないといえる。サウンドロイドは当面「ビジネスに役に立つ」音や音楽を生成するという具体的な目標をもつが、そのための評価関数をもっているわけではない。サウンドロイドはもう少し別のストラテジー（後述）をもつ。

##### §3 人間の認知と事象をつなぐデータの欠如

「明るい」、「楽しくする」というような感情は人間の外側にあるのではなく、人間の認知、認識であり脳中で生起する。またモーツァルトの音楽も楽譜の中にあるのではなく CD でもなく、私達の脳内に生起する。現在においてこの脳内現象である「明るい」、「楽しくする」などの人間が行う認知と音楽自体の構造との相関データは、人間から直接得るしかなく、「人工知能技術で音楽」をつくらうとしたときに本格的に取り組むべきはこの音楽構造、事象に対応する人間からのデータ、ある音楽構造、事象が私達人間の認知機能、感情にどのように働きかけるかというデータである。このデータを多数集めることによって、ようやく分析や学習が可能になると

\*2 本稿の著者である古川 聖、大村英史、濱野峻行、柴山拓郎は 2008～14 年に JST ERATO 岡ノ谷情動情報プロジェクト内において音楽サブグループを形成し、音楽情動の研究を行った。

私達は考えている。サウンドロイドが「徐々に成長していくシステム」であるというのは、この人間認知と音楽構造の相関データを多数集め、それによってシステムの質を向上させていくという意味である。確かに、人工知能と音楽というキーワードのもとに、音楽構造（メロディーや音色など）を多数集めてそれらを学習器にかけている様々な結果を得ている研究（例えば、<https://experiments.withgoogle.com/collection/ai>）があるが、人間からの直接のデータのフィードバックはほとんどなく、「人間のための音楽を生成する」というトピックスからは距離があるように私達には感じられる。

## 5. サウンドロイドの実装

### 5.1 サウンドロイドが志向するもの、コンセプトのようなもの

#### §1 失われつつある音環境

サウンドロイドは音／音楽を生成するのだが、そもそもこの音／音楽とは何を志向しているのだろうか。私達はさまざまな音に取り囲まれている。私達の生活の中において、物理的な現象や私達のアクションには必ず音が不可分に付随していて、聴覚を含む五感を通して、環境から生存に必要な情報を得ている。また、それだけではなく街の音、話し声、川の音、雨音、滝の音、鳥の鳴き声、その他などの環境音は私達の生活を取り巻き包んでいる。私達はほんの少し前までは、それらの音を感じながら仕事をし生活してきたはずなのだが、現在の私達の住環境、仕事環境はどうであろうか。環境から断絶した人工的な環境は過酷なものかもしれない。私達が滝の音などに深い共感を覚えるのは失われつつある音環境への憧憬かもしれない。サウンドロイドの背景にはこのような音環境に対する問題意識がある。

#### §2 音楽の解体、再構築と環境化

音楽とは大変認知的負荷の大きい脳活動で、特に芸術音楽に至っては認知の複雑性、負荷の大きさは相当なものになる。S. ピンカーの聴覚のチーズケーキという表現にうなずける部分もある [ピンカー 03]。サウンドロイドは複雑性（5.2節参照）というキーワードのもとにこの音楽聴取を解体し、認知的負荷の大きさを環境音のレベルにまで下げる。つまり音楽を聴いたときに自動的に脳内で起こる構造的認知、シンタックスの形成<sup>\*3</sup>を抑え、

\*3 私達が普通に聴いている音楽は、音の構造認知という側面から見て大まかにいうと、個別の音のグルーピング化に始まる階層性をもつ。つまり個別の音が集まり、関係付けられフレーズがつけられ、フレーズが集まり、小楽節をつくり、小楽節が集まり大楽節をつくり…、といった具合で、それらは記憶されるのだが、すべての音を記憶するのではなく、各レベルごとに抽象化され、そこに含まれる音程や特徴的なリズム、階層が上になれば、グループ間の関係性（類似、対比、その他）が構造的に記憶される。これらは半ば自動的に起こり、認知的リソースを消費する。

多数のパラメータに解体、そのパラメータをコントロールし、音環境を再構築する試みである。仕事などの作業に使われる認知的リソースと競合しないように、認知的負荷を落とすまま、その境界領域において揺らぎを与え、そこに湧き上がってくる心的状態、音楽感情、クオリアなどを含めた総体を人間の目的設定に沿って、各パラメータの最適値探索を通して捉えることをサウンドロイドは目指している。

### §3 新しい音環境の再デザイン

最終的にサウンドロイドが目標とするものは音／音楽の境界線上、意識と無意識の中で私達の生活、仕事環境に適切な形で溶け込むような、私達がひょっとしたら失いつつあるかもしれない音環境につながっていくような、新しい音環境の快適性を再デザインすること、またはそのデザインを可能とする仕組みをつくることである。

### 5.2 サウンドロイドがもつ仮説

サウンドロイドは音／音楽を生成するのだが、音／音楽の生成にあたって、私達が採用したのはバーラインによる最適複雑性モデル [Berlyne 70] として知られる複雑性の知覚に関する研究である。このモデルを端的に説明すると、横軸に複雑性としての興奮 (arousal)、縦軸に情動としての快楽 (hedonic) を設定すると、両者の関係は逆U字で描かれる [大村 15]、つまり認知において、単純すぎる音楽でも複雑すぎる音楽でもなく、適度に複雑な音楽に人間は快を感じるというものである。サウンドロイドの音生成のルーチンには以下のような12の音楽に関連するパラメータが設定され、各パラメータに0.0～1.0までの複雑度がアサインされるようになっている。生成ルーチンはこれらのパラメータによりコントロールされ、主にこの12種類のパラメータの複雑度の値の相関の中にユーザが求めているものが表現されるといふ仮説のもとにつくられている。

実はこの最適複雑性モデルの関数がどのような形であるのか（例えば、極大値はどこなのか、またパラメータの選択、個々の複雑度の設定が適切かどうか）は、それほどの問題ではないといえる。なぜなら、将来において

表1 サウンドロイドの音生成パラメータ

カテゴリー	パラメータ
ピッチ	多様さ 音階の長短 音階の明るさ 音域の広さ 全体的な音の高さ なめらかさ
リズム	店舗 多様さ 規則正しさ
音	楽器 (音色) 音量 残響

ユーザから膨大なデータを集め、サウンドロイドが分析・学習することによって調整されるからである。さらにこれらの過程を経ることにより、このような仮説ではなくもっと有効なパラメータやコントロールの方法が見つかるかもしれない。このような仮説の上にスタートするのは哲学者の千葉雅也の言葉を使えば、この生成システムの射程を「有限化」し、このプロジェクトを完遂可能なものにするためである。いつかディープラーニングということになれば、その原則からして音の生成方法は乱数でもよいはずなのだが、現在のような少ないデータしかもたない状況からの入口としては、まずは私達が現在知っている音や音楽の構造に近いものが生成されやすいものがよい。なぜなら、サウンドロイドのユーザに全くのゼロ、つまり乱数から発生させられたバラバラな音や音楽を直接に調整、評価するようなスタートを強いるアプリケーションは提供できないからだ。

### 5.3 サウンドロイドの機能モデル

ここでサウンドロイドがどのように機能するのかもう少し詳しく見てみることにする。まず、断っておかなければならないのは、これから説明するサウンドロイドの機能モデルは、実践のためのイメージのようなものであり、制作の進行を優先する現場の仮定であり、時として研究としての厳密性をもたない場合もある。

#### §1 複雑性と認知的リソースと逆U字関数

10個の音が、順次出力された状況を考える。音の認知に関してはゲシュタルト原理が働くので、聴取者は個々の音を順次、記憶しながら、自動的にその10個の音の間の関係性の吟味を始めるのだが、ここで多くの関係性が認知されたほうが複雑度の高い出力だと考える。

例えば、doの繰返し (do, do, do, do, do, do, do, do, do, do) より、3音ずつ順次上向 (do, re, mi / re, mi, fa / mi, fa, sol / fa) のほうが複雑度の高い出力である。つまりサウンドロイドがいう複雑度とは認知的リソースの消費量である。しかし、複雑度とは認知的リソース負荷は常に正比例関係になっているとは限らず、複雑度が上がりすぎると認知が放棄され、逆U字形に認知的負荷は下がってくる (5.2節参照)。

#### §2 複数の認知的リソースの競合

また、音楽を聴きながら (非言語的作業) 文章を書いたり (言語的作業) することは私達もよく行っていることだが、電車の中で本を読んだり (言語的作業)、文章を作成 (言語的作業) したりしているときには、周囲の話声 (言語的情報) が邪魔になる、つまり同種の作業のためにリソースは競合することもある (この種のものとしてストループ効果がよく知られているが、ストループ効果は文字の意味と文字の色の発話の競合)。認知的リソースは複数ありサウンドロイドにおいて設定されたパラメータに対応する音楽認知 (例えば、ピッチ、音量、音色) は相互に少しずつ異なる認知的リソースを使って

いると仮定し、それらは競合することもあるし、しないこともあると考える。

#### §3 複数の認知的リソースの補完

人間の脳活動、認知活動は止むことがなく、多数の認知的リソースを働かせており、またそれらのリソースは常に入力を待っているとも考えられる。ある作業に集中するためには、その作業に必要な認知的リソース以外の消費を排除するために、無音の環境が良いとはいえない。というのも、実際には無音は耐え難く、それは作業に使われない多数の認知的リソースが入力を待ってウズウズしているような状況になっているからであると考えられる。サウンドロイドは作業に必要な認知的リソースにも適度の刺激、認知的負荷を与えることにより、音環境の全体性を確保し、作業場所によっては作業環境にある不必要な音刺激に対しそれを中和する、音のバリアのように機能することも考えられる。

#### §4 複数の認知的リソースの組合せのバランス

サウンドロイドがある作業に最適な音環境を実現するために、複雑度を通してコントロールにしようと思っているのは、複数の認知的リソースの消費量とその組合せのバランスである。より詳しくいえば、ユーザの作業に使われている認知的リソースの消費量も含んだ、全体のバランスである。その意味でサウンドロイドは音生成アプリとして現実世界にも開かれている必要があり、そのためにサウンドロイドは音を生成する前に、ユーザから得る、ユーザの気分、天気、表情、作業場所、作業内容も重要な情報として記録する。特に作業内容として選択する、休息・ルーチンワーク・判断力を使う・思考力を使う、は音生成に大きく影響すると考えられる。なぜなら、既存データのインプットタスク (ルーチンワーク) に必要な集中力と、何か特別のことを思いつこう (思考力を使う) とするような、よりクリエイティブな思考に集中しているときに使われる、認知的リソースとその組合せは随分違ったものであると考えられるからだ。現在、私達には私達が設定したパラメータの認知的負荷がどのように競合し、作業においてどのように作用するか明らかになっていないが、それを収集するデータを通して明らかにし、サウンドロイドの機能へとフィードバックし、サウンドロイドを成長させていきたいと考えている。

#### §5 パラメータと複雑性

ここで、サウンドロイドで音／音楽の生成に実際に使われている音楽パラメータと複雑度・認知的負荷の関係について解説する。注意すべきは複雑度と認知的負荷は逆U字関数で示されるような関係にあり、しかも、認知的負荷がプラスからマイナス方向へ転じるその中央値、分岐点は複数のパラメータ間の全体のバランスで変動するし、個人個人でも異なっているであろうと推測されることである (この複雑な相対的な関係性を得ることはこのプロジェクトの主要な目的の一つである)。そして、運用においてはこれらの認知的負荷は (音楽的な面

白さとははるか遠くに離れ)とても低い状態に抑えなくてはならない。サウンドロイドが生成するのはある作業の認知リソースに干渉しない、適切な聞き流せる程度の少量の認知的負荷を与える音/音楽である。

1) ピッチに関するパラメータ (多様さ, 音階の長短, 音階の明るさ, 音域の広さ, 全体的な音の高さ, なめらかさ)

現れる音のピッチが多様であるほど複雑度は上がり, 音階の明るさはそこで使われる音程の協和度のパラメータで完全5度だけのもののほうが, 完全5度に長短3度が混ざったものより, 響きが透明で明るく, 複雑度は低いと考える。そして前にも書いたように音の認知に関してはゲシュタルト原理が働くので音域が狭い=音と音のピッチが近い(近接の原理)または, 音の並びがなめらか=音と音のピッチが近い(近接の原理)ほうが音のグルーピングが起りやすく, 複雑度は中央値に近づくと考えられる。音階は長調より短調のほうがいくらか複雑度は高いと考えられる。

2) リズムに関するパラメータ (テンポ, 多様さ, 規則正しさ)

リズムはその時間的な規則性, 拍動の規則性の認知度において複雑度を設定している。規則正しく等間隔に刻まれる拍動の複雑度は低く, 全く秩序の感じられない音の連なりは中央値を超え, 複雑度が非常に高いと考えられる。またテンポが早くなると単位時間当たりの情報量が増えるので複雑度は増し, 音は時間的に近接しゲシュタルト原理が働き, 音のグルーピングが起りやすく複雑度は高くなると考えられる。多様なリズムが現れると複雑度は高くなる。

3) 音色に関するパラメータ (楽器=音色, 音量, 残響)

楽器の音色にも複雑度を大まかに設定することができる。音色の複雑度はそれに含まれる整数・非整数倍音の成分の強さとそのエンベロープ(時間的な変化)によって特徴付けることができる。基音に対して倍音の成分が多く, 強いほど, また高次倍音が多いほど, またエンベロープの変化の激しいほど音色は複雑になると考えられる。音量は大きいほど情報量は多くなり複雑度は高くなると考えられる。残響は音色に空間の情報加わるので適度な量までは複雑度は高くなると考えられる。

#### 5.4 サウンドロイドアプリケーションの開発

これまで述べてきた機能モデルの実装として, Webアプリケーション soundroid.comの開発を行った。ユーザはWebブラウザからアクセスして利用することができる。まずユーザがアプリケーションを利用する流れについて説明し, 次にアプリケーションを構成する各機構の開発について解説する。

#### §1 アプリケーション利用の流れ

アプリケーションを利用する基本的な流れは, ユーザがデータを入力するとそれに応じて音/音楽の合成が行われ, それに対してユーザが評価を行うというものである。

詳細なステップは以下の四つからなる。

- 1) ユーザアカウントの開設と基本情報の登録 (年齢, 性別など)。
- 2) 毎回利用時の情報入力と設定 (ユーザの気分, 顔写真の撮影 (表情分析に用いる)\*4 (図1), 作業内容 (図2), 作業場所, 作業予定時間, 周囲の音環境, 天気など)。
- 3) セッションの開始: 2) で入力した内容に基づき, セッション中は自動で継続的に音生成が行われる。
  - オプション1: ユーザによる音生成に関するパラメータの調整 (図3)。
  - オプション2: 上記で調整したパラメータをひとまとめにしたプリセットの切替え。
  - オプション3: ビジュアルライザの表示切替え (ビジュ



図1 Microsoft Emotion API を利用した感情検出



図2 作業内容の入力

\*4 毎回利用時のユーザの状態を把握するものとして, Webカメラにより顔写真を撮影し表情分析を行っている。この分析には Microsoft Emotion API を利用している。 <https://azure.microsoft.com/ja-jp/services/cognitive-services/emotion/>

アライザはリアルタイムに音が生成される様子をアニメーションで可視化したもので、あくまでもエンタテイメントとしての付加的機能である(図4)。

4) セッションの終了。

- セッションに対するユーザの評価(5段階評価および自由記述)。
- SNSによる音生成プリセットの共有: ユーザの設定による音生成を他のユーザも体験できるようにできる。



図3 サウンドロイドの複雑性をコントロールするパラメータの入力

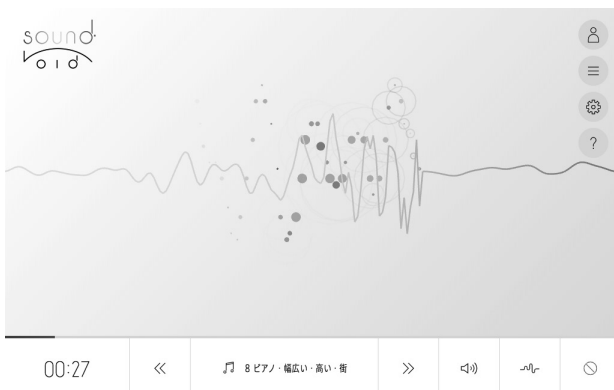


図4 ビジュアライザによる生成音のリアルタイムアニメーション(エンタテイメントとしての付加的機能)

基本情報や音生成のパラメータは随時サーバに記録され、次回以降のログインに引き継がれる。現段階の音生成は、仮実験のデータをもとにユーザの情報との簡易的な類似度の算出を行って音生成パラメータに反映させている。今後の運用を通してデータがある程度集まった際には、パラメータ推薦モデルのアップデートを随時行う。セッション中は音生成のパラメータのカスタマイズを随時行うことができる(図3)。これを行うことによって、自分の状況に適した音環境を徐々に整えていくことができる。「設定の送信」を押して調整したパラメータをサーバに送信すると、次回以降はその履歴が音生成に反映される。さらにパラメータはSNSにより共有することもできる。

## §2 サウンドロイドのデータの活用

サウンドロイドでは、音楽と人間の感覚の関係を定量的に取得するために複雑性のパラメータを設定している。この関係は、バーラインの最適複雑モデルと同等になるであろうという仮説は立てられるが、実際に運用してみないと正確なところはわからない。現在の目下の目的としては人間の感覚と音楽の構造の定量的な関係付けを得ることにある。また、ユーザ自身が入力するデータだけでなく、Webカメラを用いた顔表情分析から得られる情動的なデータも取得しているため、情動と音楽とタスクの関係の定量化も行っていく。

前述のようにサウンドロイドは「育てるようなシステム」というような展望をもつ。つまり現在のプリセットデータは、作曲家の手によるチューニングによって行われているが、実際のユーザによる運用から得られるデータの分析によって特定のパターンの発見が見込まれる。これらのパターンのフィードバックによりサウンドロイド自体のバージョンアップが可能になり「育てるようなシステム」が実現する。また、膨大なデータに対応すべく学習機能(情報推薦技術や機械学習技術)を組み込むことにより、特定のユーザだけでなく、ユーザの特徴量に合致した出力ができるような、オンラインで学習を行うシステムに拡張中である。

## §3 音生成機構の開発

音生成は、各楽器音のサンプルを取めたサウンドスプライトを事前に作成し、ブラウザ上でそれを用いてリアルタイム音生成を行っている。現在は7種類の楽器音が利用可能である。またサンプルはリバーブあり・なしの2種類を用意し、生成時にミックスすることでリバーブ効果の度合いを調整できる。

## 6. サウンドロイドのもつ問題点

現在、サウンドロイドはβ版のリリースを行ったが、まだ発展途上な点もあり現在次の版に向けて改良中である。ユーザインタフェースに関しては、使用にあたって不必要に繰り返されるプロセスがあり、ユーザが望むページへすぐに行けるように、全体の整合性を確保したうえで変更を行う必要があるだろう。また、音／音楽におけるパラメータとは別に、ユーザの評価の履歴などを視覚化して、モチベーションを高める心理的な強化学習のような機能も追加すべきかもしれない。音／音楽生成に関してはリズム楽器のようなものがあればいいかもしれないし、現在は一つの楽器でのみ演奏されているが、複数の楽器を使うことも考えられる。また、和音などのハーモニーを奏するパートがあってもいいかもしれない。しかしこのような既成の音楽の方向への延長は元々のサウンドロイドのコンセプトに照らし合わせながら注意深く進める必要がある。そしてスマートフォンなどへのプラットフォームの拡張は急務である。本当にサウン

ドロイドを多くの人に使うためには速やかに行わなければならない。

長期的にはサウンドロイドの心臓部であるデータ収集と分析・学習手法の選定の問題がある。現在のユーザインタフェースでは十分に効率良くデータが集められているとはいえ、この部分を急いで再検討する必要がある。また、近い将来に始められるオンラインからのデータを使った機械学習の導入にあたって、現在のような収集データの構成がオプティマルなのかどうか検討の必要がある。

## 7. 展 望

現在のサウンドロイドはその目的を仕事の効率化にフォーカスしているが、目的の設定とデータ収集を工夫すれば、原理的には他の目的、例えばリラックス、睡眠、気分誘導、商店などのBGM(販売促進)、医療(音楽療法、不安軽減)、レストランやホテルなど雰囲気づくり、その他にも応用できるだろう。また、現在は個人のパソコン上で動作し個人向けになっているが、将来的にはAIスピーカのようなものに入り込み、家族のような複数の人間が共有する音環境において、家族全員のデータから帰結される音環境が生成されるようなことになるかもしれない。サウンドロイドはデータによって音を生成し、そこに著作権は生じないので、これまでBGMとして流れている既存の楽曲とは違った使われ方が生まれてくるかもしれない。

仮にたくさんのユーザからのデータが集まれば、そのデータに価値が生じるが、それを中途半端に商用に囲い込むことはしたくない。データが十分に得られ「育てられたサウンドロイド」になったときに、このシステムがどのように使われて、またどのような音/音楽が生まれるか、それは胸の踊るような話ではないか、そのためにもデータを公共財として皆で管理、運用できるような方法も徐々に検討したいと思う。

最後にこの文章を読んでくださった方に、このプロジェクト、サウンドロイドにユーザとして参加して下さることをお願いしたいと思う。<https://soundroid.com/>へアクセスし体験してください。

## 謝 辞

本稿の執筆にあたって、以下の方の援助をいただいた。

長谷川丈 (三井不動産株式会社)

尾崎德行, 廣川翔太, 宮本佳奈 (株式会社博報堂)

松尾謙二郎, 岩田裕大 (インビジュアル・デザインズ・ラボ)

大谷智子 (東京藝術大学)

心より御礼を申し上げます。

## ◇ 参 考 文 献 ◇

- [Bent 01] Bent, M.: *Isorhythm, The New Grove Dictionary of Music and Musicians*, second edition, London (2001)
- [Berlyne 70] Berlyne, D. E.: Novelty, complexity, and hedonic value, *Perception and Psychophysics*, Vol. 8, Issue 5, pp. 279-286 (1970)
- [Cope 05] Cope, D.: *Computer Models of Musical Creativity*, the MIT Press Cambridge, Massachusetts, England (2005) (日本語の翻訳が近く出版される予定あり)
- [Ebcioğlu 86] Ebcioğlu, K.: An expert system for chorale harmonization, *Proc. AAAI-86* (1986)
- [古川 07] 古川 聖: 音楽作品集「数による音楽」(CD), フォンテック (2007)
- [Hiller 59] Hiller, L. and Isaacson, L.: *Experimental Music*, McGraw-Hill, Massachusetts, England (1959)
- [ケリー 16] ケリー .K. 著, 服部 桂 翻訳: 〈インターネット〉の次に来るもの, NHK 出版 (2016)
- [Koelsch 12] Koelsch, S.: *Brain and Music*, Wiley-Blackwell (2012), ケルシュ, S. 著, 佐藤正之 監修・翻訳, 橋 亮輔, 田部 井賢一, 小野健太郎, 森 数馬 翻訳: 音楽と脳科学: 音楽の脳内過程の理解をめざして, 北大路書房 (2016)
- [Juslin 11] Juslin, P. N. and Sloboda, J., A.: *Handbook of Music and Emotion: Theory, Research, Applications*, Oxford University Press (2011)
- [Nierhaus 09] Nierhaus, G.: *Algorithmic Composition: Paradigms of Automated Music Generation*, pp. 36 & 38, Springer Science & Business Media (2009)
- [大村 15] 大村英史, 柴山拓郎, 片上大輔: 人工的雰囲気生成のための分布関数のデザイン, *31st Fuzzy System Symposium* (2015)
- [ピンカー 03] ピンカー, S. 著, 棕田直子 翻訳: 心の仕組み—人間関係にどう関わるか, NHK ブックス (2003)
- [Roads 96] Roads, C.: *The Computer Music Tutorial*, the MIT Press (2001), Roads C. 著, 青柳龍也, 小坂直敏, 平田圭二, 堀内靖雄 翻訳: コンピュータ音楽—歴史・テクノロジー, 東京電機大学出版局 (2001)
- [Stern 85] Stern, D. N.: *The Interpersonal World of The Infant A View from Psychoanalysis and Developmental Psychology*, Basic Books, New York (1985)

2018年9月2日 受理

## 著 者 紹 介



### 古川 聖

高校卒業後渡独、ベルリン芸術大学、ハンブルク音楽演劇大学にてイサン・ユン、ジェルジ・リゲティのもとで作曲を学ぶ。スタンフォード大学で客員作曲家、ハンブルク音楽大学で助手、講師を経てドイツのカールスルーエのZKMでアーティスト研究員。作品は、新しいメディアや科学と音楽の接点において成立するものが多く、1997年のZKMの新館のオープニングでは委嘱を受けて、マルチメディアオペラ『まだ生まれぬ神々へ』を制作・作曲。多くの受賞歴がある。東京藝術大学先端芸術表現科教授、芸術情報センター長兼任。



### 大村 英史 (正会員)

2002年東京農工大学工学部機械システム工学科卒業。2009年東京工業大学大学院総合理工学研究科知能システム科学専攻博士課程修了(新田克己研究室)。博士(工学)。科学技術振興機構ERATO岡ノ谷情動情報プロジェクト研究員、国立精神・神経医療研究センター精神保健研究所流動研究員などを経て、2015年より東京理科大学理工学部情報科学科助教。人間の情動に関わる音楽の構造分析や、音楽構造の定量化とその構築に関する研究に従事。また、雰囲気量の定量化・応用に関する雰囲気工学の研究に従事。音響学会、情報処理学会などの各会員。



**瀨野 峻行**

国立音楽大学音楽文化デザイン学科卒業。オランダ王立音楽院ソロジー研究科修士課程修了。東京藝術大学大学院美術研究科後期博士課程修了。元科学技術振興機構 ERATO 岡ノ谷情動情報プロジェクト研究員。現在、東京藝術大学および国立音楽大学にて非常勤講師を務める。メディアアートに関わるテクノロジーと ICT 教育を専門とする。

**川村 剛**

東京工科大学卒業後、同大学院にて工学修士課程を修了。SEとしてIT会社に勤務。JSTで客員技術者、母校で非常勤講師を務め、2014年末に自営業として独立。さまざまなシステム開発をする傍ら、ArduPilotというドローン制御ソフトウェアに魅せられ、ドローンを活用した救助コンペ出場やセミナー講師、物流分野のスタートアップ企業初期メンバーとして参画し、RFIDおよびAI関連のシステム開発に従事。

**柴山 拓郎**

東京音楽大学および同大学院、東京藝術大学大学院美術研究科(先端芸術表現領域)修了。修士(音楽)、博士(美術)取得。作曲を西村朗、池邊晋一郎、古川聖の各氏に師事。2015年電子音響ビープルプロジェクトを立ち上げ、コンピュータ音楽を多様な人々と共創するフィールドデザインに取り組んでいる。2017年ZKM招聘芸術家・研究員。2018年にはZKMで講演・個展・展示を行った。現在、Saitama Muse Forum運営委員、東京電機大学理工学部情報システムデザイン学系准教授。