研究報告

対話型機械学習の実践的な作曲への応用
APPLYING INTERACTIVE MACHINE-LEARNING INTO ACTUAL MUSICAL COMPOSITION

安藤 大地
Daichi ANDO
首都大学東京システムデザイン学部インダストリアルアートコース
Div. of Industrial Art, Tokyo Metropolitan University

概要

近年機械学習を芸術作品制作へ応用する事例が多く報告されている．しかし，基本的には既存の画像を「学習」してそれに近い画像を生成する形であり，独創性を重視する創作手法とは言い難い．著者は，1. 独創性を担保するため，対話型学習プロシージャによって人間が曖昧な評価関数として振る舞うこと，2. ロジックの記述性を求めるため，アート作品の生成モデルを知識として与えること，の 2 つを実現するために，遺伝的プログラミングを対話型にして実際の作曲に機械学習を用いている．本稿では，対話型遺伝的プログラミングに到るまでの機械学習の検討について論じた上で，近年の作品で用いた音楽作品のための対話型学習プロシージャと，そこで用いた記号的な作曲モデルと音響生成モデルについて報告する．

Applying Deep-Learning, is typical of current machine-learning researches, into art creation field has been made progress. However, object of almost the research in this the field are making just copies or immitations of the training data. Although it is thought that “Creativity” in the art creation field is nessesary element, almost the research has been not noted that.

In this paper, the author discusses the right and wrong of that applying machine-learning into the art creation field, then expresses adopting “Interactive Genetic Programming” to realize 2 objects: human playing as objective function in interactive machine-learning procedure to give creativity to the system, including knowledge of logic of the art creation as model into the system. Then the author reports his actural composition by means of Interactive Genetic Programming for both musical elements, notation and sound synthesis.

1. はじめに

音楽情報科学分野でベイズ的手法が一般的になって以降，コンピュータ音楽の作曲家の間に音楽情報科学や人工知能の諸研究を理解しようとする雰囲気がなくなり，深層学習 (Deep-Learning) が研究の主流になってからその傾向が強くなっているのではないか，と著者は感じでいる．

一方で Google が音楽のための深層学習の研究群を公開する (Megenta [5]) など，人工知能の研究者からは「画像処理の次は音楽だ」とばかりに音楽分野に熱い視線が投げかけられているが，こちらも汎用アルゴリズムである深層学習にモデルの獲得を任せてしまい「音楽的知識」の重要性が蔑ろにされている，と作曲家は感じているのではないだろうか．

本稿ではコンピュータ音楽の作曲のために機械学習を利用することを前提として，機械学習の基礎知識と独創性について議論した上で，その後に著者自身の作曲への機械学習の取り込みについて報告する．

2. 機械学習の概要と問題点

一般的に機械学習と呼ばれる手法には，教師データ (トレーニングデータ) に合致するように学習する「教師あり学習」と，与えた特定のルールで得られる点数を最大化するように学習する「強化学習」の 2 通りがある．さらにその前段階としてランダム探索について述べる．

2.1. ランダム探索

作曲においては，特定の生成ルールに乱数 (もしくは何らかのアート的根拠を与えられた数列) を突っ込み，大量の旋律や音響を生成し，それを片っ端から人

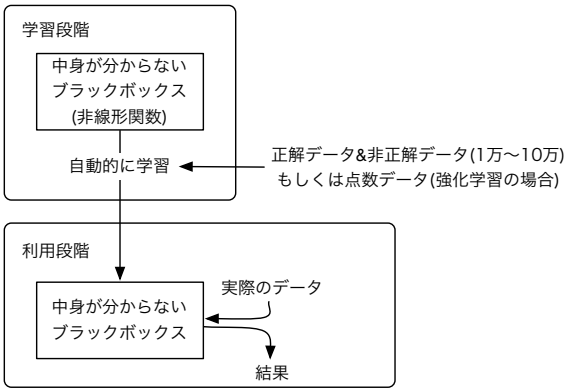


図 1. 通常の教師あり学習や強化学習の模式図

間が聞いて，良いものをピックアップする，という手法がランダム探索と言える．

ランダム探索は，生成ルールで表現される探索領域内では「独創性」を担保されている．しかし狭義での探索すなわち「学習」を行っているわけではなく，どれだけ繰り返しても次に出てくるものが良くなっている保証はない．無限に時間があるならば良いものを手に入れられる可能性は高くなるが，現実的な時間でランダム探索で良いものを探そうとするなら，乱数を使わずに直接固定したルールを書き下してしまった方が効率的である．ただしそれは「予想外の良い結果」という「独創性」を生み出さなくなってしまう．

「予想外の良い結果」を得る効率を上げるために，コンピュータに何らかの学習をさせて学習させて良いものだけを取り出そうとするのが「機械学習」である．

2.2. 教師あり学習

教師あり学習とは, 何らかの学習する対象のパラメータやその組み合わせを自動的に学習するものである．一番わかりやすいのは分類問題や回帰問題である．

回帰問題で一番簡単な例は単純線形回帰分析である．x と y の 2 座標からなるデータが大量にある時，すべてのデータは

$$y = ax + b \tag{1}$$

のように直線上にある (誤差があるので綺麗に直線上に乗らなくてもよい) と仮定して，その式の a と b を求めることが単純線形回帰分析である．これが曲線や x の数が複数になることも考えられる．

深層学習とは，乱暴に言ってしまえば，この式 1 が非常に複雑な，とある式 (深層ニューラルネット，Deep Neural-Network) で表される時の学習手法のことである．この場合学習された中身を人間が把握することは非常に難しいので，学習された式は「ブラックボックス」化する．

分類問題とは多くのデータをその特性に応じて定められた数に分類する問題であるが，結局は回帰分析と同じく，データを分類できる直線もしくは曲線を引くという，回帰問題と同じものとして捉えることができる．

つまり教師あり学習とは，データに「フィット」する a や b の値の組み合わせを探すことである．これを「フィッティング」とも言う．

実際にどのぐらい「フィット」しているかは, 各データと，仮に与えた a と b で作られる直線の誤差の合計で判断される．この誤差の合計が最小になるように「フィッティング」を進めるため，理想は誤差の合計が 0 になることである．(もちろん「データは最初から誤差を含んでいる」ことが前提なので，0 になることはない)

これを「目的関数 (誤差を算出する関数) を最小化する」と呼ぶ.

2.3. 強化学習

教師あり学習が「持っているデータとの誤差 0 を目指す」のに対して，強化学習とは，持っているデータがなく「試行錯誤した結果が一番良い」状態を目指すものである．「試行錯誤」に対して点数をつける関数のことを「価値関数」と呼ぶ．一定回数試行錯誤を行い，徐々に試行錯誤の内容を多くの点数が得られる方向に変えていき，この価値関数の出力が最大のものが最終的な結果となる．

こちらは教師あり学習に対して「目的関数 (評価値を算出する価値関数) を最大化する」と言うことができる．

強化学習は，教師あり学習のように既知のデータがなくとも，試行錯誤する環境と価値関数を定めればよい．ただし環境と価値関数は「明確なルール」である必要がある．つまり「何が良くて何が悪い」のかが，予めわかっていなければならない．

2.4. 独創性の問題

作曲に応用して考えた場合，2.2 項で述べた教師あり学習は生成物つまり旋律や音響が教師データに限りなく近づくことを目的としており，2.3 項で述べた強化学習では旋律や音響が何かしら定めた明快なルールにより算出された点数を多くすることを目的としている．いずれも「目的関数」が明確である必要がある．

さて我々が「独創性」や「創造性」と呼んでいるものは何であろうか？ 一般的にいわゆるシリアス音楽の作曲家は，既存の作品を分析しつつ，分析結果をベースに試行錯誤しながら作ることを繰り返し，その過程で「作品を生成する新たなロジック」を導出する．ルール

ベースのアルゴリズム作曲を行う際は 2.1 項で述べたランダム探索になり，「様々なルールや入力をランダムに試しながら自分が一番良いと思うものを探す」と言うことができる．

ところが「自分が一番良いと思うもの」は，当然ながら最初から既知ではない．試行錯誤の過程の中で「揺れ動き」，最終的に「とりあえず今回はこれでいこう」と (妥協しつつ) 定めるものである．

作品制作の当初から既知ではないということは，目的関数を定めることができないということであり，機械学習を利用することができない．逆に言えば予め「これが良い」とわかっている場合は利用することができるため「模倣」には最適である．

しかし，音楽のように (直線 ± 誤差で表現されるよりはるかに) 複雑な生成物の場合は，学習されたものは非常に複雑でありブラックボックス化している．そのため学習により「模倣」したものから「人間が理解できるロジック」を取り出すことができなくなってしまう．これは音楽の場合は「アナリーゼした結果が，わけのわからない式で，なぜこの曲がその式でできているのか未だにわからない」，つまり「実質的にアナリーゼを行っていない」状態とも言える．

つまりアナリーゼによる知識の再利用ができないため「新たなロジック」つまり「独創性」を得ることはできない．

独創性の創発に必要なのは，

1. 試行錯誤中に揺れ動く価値観を取り込む
2. 音楽の生成ロジックを人間が理解できる形で獲得できる (手動修正可能)

機械学習システムである (図 2).

2.5. Creative Adversarial Network

Elgammal が提案した CAN(Creative Adversal Networks) という画像を生成する深層ニューラルネットワークの学習プロシージャ [3] は「いかに独創性を学ぶか」をテーマにしている．CAN のベースとなった GAN(Generative Adversal Networks) という深層学習のプロシージャでは，

1. 入力した正解データを模倣した画像を生成する深層ニューラルネット
2. 与えられた画像が，1. のネットワークが生成した偽物なのか，それとも本物の写真や人間が描いた絵なのか，を判別する深層ニューラルネット

の 2 つの深層ニューラルネットを同時に学習することで，極めて写実的，もしくは極めてアニメ絵的な画像を生成可能であることが報告されている．

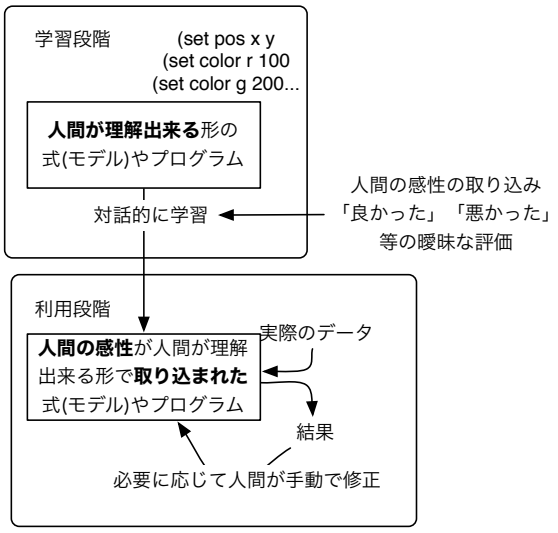


図 2. デザイン・アート向けに「行いたい」対話型機会学習機会学習の模式図

ただし GAN は正解データを用いて学習する教師あり学習プロシージャであり，「正解データと判断される」つまり「模倣した，独創性のない」画像しか生成することができない．そこで Elgammal の手法では，GAN の評価関数を一部反転させ「既存の絵画スタイルに対する当てはまりにくさ」を独創性の価値基準として「アート作品っぽさ」を学習させている．

しかしながら，ここで学習できた「アート作品っぽさ」は「既存の作品に当てはまりにくい」だけであり，また「それがどのようなものなのか」は深層ニューラルネットの特性としてブラックボックス化されており，音楽で重要視される「ロジックとしての正当性を持った独自性」を獲得することはできない．

2.6. 対話型進化計算と対話型遺伝的プログラミング

そこで著者は遺伝的アルゴリズム (Genetic Algorithm, GA)[4] などの機械学習システムを「対話型」にした，対話型進化論的計算 (Interactive Evoluationary Computation, IEC) を用いている．さらにクラシック音楽の生成モデルとして LISP 系統の言語が頻繁に用いられることに注目し，LISP のプログラムを生成する遺伝的プログラミング (Genetic Programming, GP)[6] を用いている．

対話型進化論的計算では，強化学習において明確に決めなければならなかった「環境」と「価値関数」を人間が (音楽の場合は) 対話的に直接聴いて評価をつける．これにより強化学習の弱点であった「予め，何が良いかを明確に決めておかなければならない」制約がなくなる．さらに，複数の次元 (軸) の評価があった時，

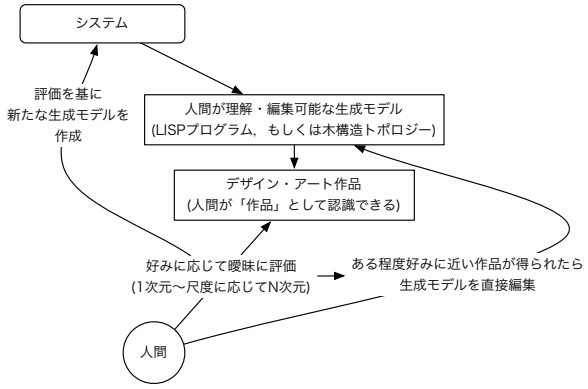


図 3. 対話型遺伝的プログラミングの模式図

すなわち複数の目的関数が存在する場合，通常の機械学習ではパレート解を用いた多目的最適化を行わなければならないのに対して，対話型進化論的計算は厳密にパレート解的な多目的最適化を行う必要がない.[1]

遺伝的プログラミングにより「音楽を生成する LISP プログラムを探索する」機械学習では，数理的な関数に支配されない楽理的な観点のロジック表現可能な音楽を生成するプログラムを音楽をそのまま探索することができる.

これにより

- 試行錯誤中に価値が揺れ動く
- ロジックが明確になる
- ランダムネスから生まれる独創性
- 単なるランダム探索ではなく収束する

対話型学習を行うことが可能になる (図 3).

3. 制作の実例

3.1. ノーテーションの生成

著者はこれまでも著者自身が構築した，クラシック音楽的なを生成ロジックを表現したプログラムを進化させる対話型遺伝的プログラミングシステム CACIE[2, 12] や，写真の画素情報を入力に取りクラブミュージック的なループ音楽をリアルタイムに生成する CACIE P 型 [1, 10] により，ノーテーションの生成を行なってきた.

[1] 本稿では詳細は省くが，アート・デザインを対象とした対話型進化論的計算では近似解が得られた時点で探索を打ち切ることができるため，群の中で異なる評価軸・具体的には異なる目的関数にそれぞれ特化した個体 (パレートではない解) が同居できる，もしくは Genome Storage という手法 [2, 12] により強制的に群内に同居させる・抜き出して保存するなどの操作が対話的に可能になるため，多目的最適化でもパレート解に縛られない．非対話型進化論的計算による多目的最適化を進め収束させるとパレート解の集合が得られる傾向がある.

本稿では，2014 年 9 月に JSSA 先端芸術音楽創作学会 Sonic Arts Project Vol.4, No.1 コンサートにて初演された著者のテナーサキソフォンのための曲 *Whitehearted* for tenor-saxophone and computer のために構築した，既存のルールに依らずにバス課題を解き，対応するスケールを生成するシステムについて述べる.

3.1.1. バス課題を解くプログラムの対話的探索

これまで著者はサキソフォンのためのライブインタラクティブ曲として「明らかな和声」を用いた曲を作曲してこなかった．しかし，このコンサートの時には演奏家と本番の日までリハの時間を取ることができない状況にあった．これはライブインタラクティブな曲としては致命的な状況である．そこで，クリック音とある程度の音響からなるフィックスドメディアによるバックトラックを予め演奏家に渡しておいてサックスパートの演奏を練習してもらい，そこに本番当日のリハでライブインタラクティブに生成した音響を載せるという方法を取った．そのような形態の練習，上演を行うためには，バックトラックに電子音響の (一般の演奏家が認知可能な) ハーモニーパートがある方が，練習する演奏家としては自然に感じる，と著者は考えた．そのため，著者の作風である「綺麗な調性の曲っぽいけど，なんか違う」和声進行を生成する必要が生じた.

一般的にバス課題を解いて和声進行を生成することはトップダウン的なルール，例えば機能和声やピッチクラス集合論に基づく和声進行を記述・実行することで行われる．しかし著者オリジナルのルールを作ろうと思った時，トップダウン的にルールを記述すると，細部では手動修正を行う必要がある，トップダウン的に記述したルールに満足できなくなる，などの問題が生じていた.

そこで著者は「バス課題を入力に取り和声進行を出力するプログラム」を対話型遺伝的プログラムにより生成した．その LISP プログラムで使用する関数や端末ノード表 1 に示す.

図 4 で与えたバス課題を入力に取り，図 5 の和声進行を出力するプログラムを生成した.

表 1 で示した端末ノードには「音楽の知識」が含まれている．遺伝的プログラムは LISP のプログラムをそのまま生成するため数理的に記述することができない「問題に対する知識」を取り込むことができる．NFL 定理 [8] が示すとおり，汎用機械学習アルゴリズムを問題に適用する場合には「その問題に対する知識」が「適切なレベルで含まれている」ことが必要となる．このことからも遺伝的プログラミングの有効性を示すことができる

表 1. 和声進行生成に用いた LISP プログラムのノード

関数ノード	機能
prog2	引数 2: 与えられた引数 (サブツリー) を順番に実行する
addNote	引数 1: 音高を追加する.
+, -, *, /	引数 2: 四則演算
IfLessThenElse	引数 4: 第 1 引数が第 2 引数より小さければ第 3 引数を実行し，それ以外は第 4 引数を実行する.
IfMoreThenElese	IfLessThenElse の反対.
端末ノード	**機能**
addEmpty	和声に空の音高を一つ加える.
currentBassPit	現在の場所の与えられた音高 (図 1 で表される音高) を取得.
currentHarmPit1〜5	現在の場所の和声に既に与えられた音高を取得. (現在の和声構成音数が 5 に満たない場合は和声転回的に既に追加されているものを取得)
previousHarmPit1〜5	一つ前に自分が作成した和声の個別の音高を取得.
定数 1〜11	1〜11 の定数
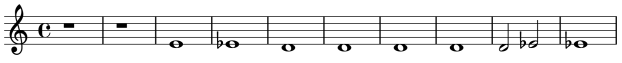


図 4. 与えた無調ピッチ配列

3.1.2. 対応するスケールの対話的生成

さらに別のプログラムノード群を用いて，ジャズで言えばアヴェイラブルノートスケールに相当する「和声進行に対応する人工音階」を生成し，それを予めインターバル配列として持っていたモチーフに適用して，図 6 の旋律を得た.

述べてきたように，この曲のノーテーション部分は，1. バス課題を解くプログラムの生成, 2. 対応するスケールの生成，という 2 段階生成からなる．この手法を用いた結果，著者の元々の作風をボトムアップ的に獲得することができた.

3.2. 音響の生成に適用するアイデアと実装

遺伝的プログラミングを用いた音響生成は可能である．LISP プログラム内で sin 関数ノードを使って表現すればよい．遺伝的プログラミングでは関数同定 (Symbol Regression) がキラーアプリケーションとなっており [6]，周期的関数を組み合わせた複雑な時系列関数の合成問題が，株価・通貨取引の予測や自動売買プログラムの自動合成に取り込まれている [11].

しかし遺伝的プログラムで音響生成を行おうとする場合，遺伝的プログラムのプログラムは可変長でなければならない，かつプログラム 1 つ (1 個体) をサンプル数の回数分だけ評価しなければならないため，計算コストが非常に高くなってしまうという問題がある.
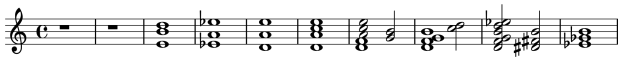


図 5. 図 4 のピッチ配列から生成された無調の和声と進行
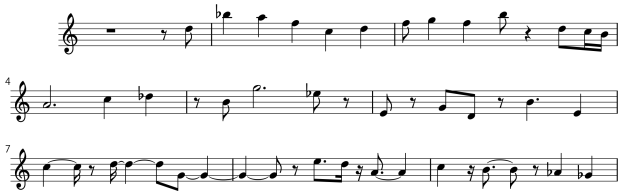


図 6. 図 5 の和声進行に対して生成したスケール (インターバル配列) を適用したモチーフ．与えたモチーフが 5/4 拍子のため図 5 の進行を部分拡大して適用している．Daichi Ando: *Whitehearted* for tenor-saxophone and computer 冒頭部.

CPPN(Compositional Pattern Producing Networks) という，ニューラルネットの発火関数を周期関数に置き換える手法が提案されている (図 7)[7]．これを再帰ニューラルネットに適用することで，固定長の表現で複雑な波形を生成することが可能となる．また固定長であることは実装においても非常に有効であり，例えば GPGPU などを適用することができる.

CPPN の対話的探索にはニューラルネットワークのトポロジーと重み付けを遺伝的アルゴリズムや類似の手法で探索するニューロエボリューション [9] を応用して実装することができる.

著者による *Thimbletail* for alto-saxophone and electro-acoustic(Electro-Acoustic Music Study 2017 付属コンサートで初演) では，CPPN を利用した対話型ニューロエボリューションを GPGPU で実装して，音響を生成している.

4. おわりに

本稿では，対話型機械学習について，その独創性やロジックの記述性などについて議論を行った上で，対話型遺伝的プログラミングの作曲への応用の有効性を，著者自身の曲での実例を挙げながら示した.

現在深層学習のアート・デザイン制作への応用の研究が進んでいるが，研究が高度化していくにつれ，人工知能の研究者とアート作家・デザイナーの間の溝は深くなっており，アート・デザイン固有の問題である独創性やロジックの記述性などの問題はこれから表出してくると考えられる.

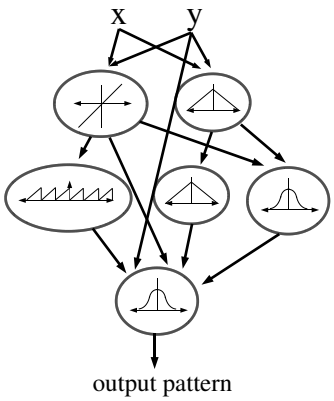


図 **7**. CPPN の模式図．[7] より．

コンピュータ音楽の作曲者・研究者の両方の立場からの議論が深まっていくことを期待する．

5. 参考文献

[1] Daichi Ando. Real-time breeding composition system by means of genetic programming and breeding procedure. In *Proceedings of International Computer Music Conference 2014 and Sound and Music Computing 2014, Athene, Greece*, Sept. 2014.

[2] Daichi Ando, Palle Dahlsted, Matts G. Nordahl, and Hitoshi Iba. Computer aided composition by means of interactive gp. In *Proceedings of International Computer Music Conference 2006, New Orleans, USA*, pp. 254–257. ICMA, October 2006.

[3] Ahmed Elgammal, Bingchen Liu, Mohamed Elhoseiny, and Marian Mazzone. Can: Creative adversarial networks, generating “art” by learning about styles and deviating from style norms. arXiv https://arxiv.org/abs/1706.07068, 21, Jun, 2017.

[4] D. E. Goldberg. *Genetic Algorithms in Search, Optimization & Machine Learning*. Addison-Wesley, 89.

[5] Google. magenta: Make music and art using machine learning. https://magenta.tensorflow.org, Accessed on 1, Sept. 2017.

[6] John Koza. *Genetic Programming: On the Programming of Computer by Means of Natural Selection*. MIT Press, 1992.

[7] Kenneth O. Stanley. Compositional pattern producing networks: A novel abstraction of development. In *Genetic Programming and Evolvable Machines Special Issue on Developmental Systems*. Springer, NY, 2007.

[8] David H. Wolpert and William G. Macready. No free lunch theorems for search. Technical Report SFI-TR-95-02-010, Santa Fe Institute, 1995.

[9] Yao Xin. Evolving artificial neural networks. *Proceedings of the IEEE*, Vol. 87, No. 9, pp. 1423–1447, Sept. 1999.

[10] 安藤大地. クラブ系ループ音楽パフォーマンスのためのリアルタイムに音楽を breeding するシステム. 情報処理学会論文誌, Vol. 53, No. 3, pp. 1030–1040, March 2012.

[11] 佐々木崇, 伊庭斉志, 石塚満. 遺伝的プログラミングを用いた株価指数予測. 情報処理学会 全国大会講演論文集 第 58 回 (人工知能と認知科学), pp. 189–190. 情報処理学会, Mar. 1999.

[12] 安藤大地, P. Dahlstedt, M. G. Nordahl, 伊庭斉志. 対話型 gp を用いたクラシック音楽のための作曲支援システム. 芸術科学会論文誌, Vol. 4, No. 2, pp. 77–86, 2005.

6. 著者プロフィール

安藤 大地 (Daichi ANDO)

国立音楽大学音楽学部声楽学科を経て音楽デザイン学科卒業．Sweden, Chalmers University of Technology より MSc. 授与．東京大学大学院新領域創成科学研究科基盤情報学専攻博士課程修了．博士 (科学)

声楽を松本進に，作曲とコンピュータ音楽を Palle Dahlstedt と菜孝之，Cort Lippe に，サキソフォンを宍戸陽子にそれぞれ師事．

2009 年より首都大学東京システムデザイン学部インダストリアルアートコース助教．情報処理学会，先端芸術音楽創作学会それぞれ正会員．

対話型進化システムや人工生命，群知能などの音楽創作分析への応用を中心とした研究を行っている．また自らの研究成果を応用した音楽作品，メディアアート作品の創作を行っている．