

A I（人工知能）が創り出す音楽に関する一考察

ーオルタ 3 や VOCALOID0:AI 美空ひばりがもたらしたものー

茨城県教育研修センター教科教育課 石津 美代子

A Consideration on Music That AI Creates:

What Alter3 Simulator and VOCALOID:AI *Misora Hibari* Have Brought

ISHITSU Miyoko

【要旨】

近年、人工知能（AI）は音楽の分野にも波及している。「人」を相手にオーケストラを指揮したり、限りなく「本人」に似た歌声で歌ったりする AI などが開発されている。さらに AI による「ベートーベン風」や「ショパン風」の音楽などが次々と創られている。しかし、AI が創る音楽に対して、賞賛よりも畏怖や怒りなど人は複雑な反応を示す。そこで、オルタ 3 や VOCALOID0:AI 美空ひばりの開発及びそれを目にした私たち「人」の反応を基に、これからの音楽に対する向き合い方について考察する。

キーワード：オルタ 3、VOCALOID0:AI、EMI、

1 はじめに

近年、人工知能の発展はすさまじいものがある。海外では 1996 年の DeepBlue（チェス）、2011 年の Watson（クイズ）、日本では 2017 年の AlphaGO（囲碁）など、ゲームで勝つ AI が次々開発された。特に Ponanza は、2013 年にプロ棋士佐藤慎一四段（当時）を 141 手で破り、コンピュータが初めてプロ棋士に平手で勝利した。そのわずか 4 年後、2017 年、佐藤天彦名人（当時）との対局で、第 1 局 71 手、第 2 局 94 手で勝利し、注目を浴びた。このような駒の動きが明確なチェスや将棋は、ディープランニングや強化学習によって膨大な情報を処理し、よりよい方法を導き出していく、人工知能の最も得意とする分野である。

では、芸術の分野ではどうだろう。高等学校の将来構想に関する参考資料（平成 29 年 5 月 29 日）では、「人工知能やロボット等による代替可能性が低い職業」の中に、「作詞家」、「作曲家」、「声楽家」、「ミュージシャン」が挙げられている。芸術などクリエイティブな仕事は想像力が必要であるため、人工知能等での代替えは難しい、とされている。しかし、それならば「VOCALOID0:AI」や「オルタ 3」の存在はどのような位置づけとなるのだろうか。

本研究では、「オルタ 3」や「VOCALOID0:AI」の開発及び社会に投げかけた影響などを述べていくとともに、これからの音楽との向き合い方について考察していく。

2 AI 指揮者「オルタ」

(1) 機械人間「オルタ」の誕生

2016 年 7 月、新作アンドロイド「機械人間オルタ」が発表された。開発を手がけたのは、テレビなどでも放送されたタレントのマツコ・デラックスにそっくりなアンドロイド「マツコロイド」の開発を監修した大阪大学大学院基礎工学研究科教授 石黒浩である。「マツコロイド」が人間そっくりなのに対し、「オルタ」は顔や手の一部分のみシリコンでカバーされているが、その他は機械が露出したむき出しで上半身と腕までしかない。一見すると「不気味」な存在である。石黒（2016）はこのオルタの外見について「機械人間オルタは、これまで僕が作ってきた人に似たロボット＝アンドロイドとは逆のアプローチで開発したものである。（中略）見た目の人間らしさを削ぎ落としても生命らしく見えるとしたら、それはどういうものなのかを追求したくて始めたものだ。」と述べている。そしてその外見は「オルタ 2」「オルタ 3」にバージョンアップされても変わらない。

「オルタ」の体の仕組みについて、共同研究者の東京大学大学院総合文化研究科教授 池上（2016）は

オルタの体には 42 個の動かせる場所（関節）があって、圧搾空気をコンピュータによって出し入れして動いている。このコンピュータがオルタの、ある意味で、脳にあたる。ただし、オルタの動きは事前にプリグラムして作った動きではなく、リアルタイムに生成されている。動きの生成にはセントラル・パターン・ジェネレータ（CPG）とニューラルネットワーク（人工の神経細胞ネットワーク：NN）という二つの方法を使っている。（中略）オルタにはセンサーで感知した外部からの刺激も取り込まれる。つまり、オルタは、内部のダイナミクスと外部情報をともに処理して動いているということである。結果、複雑な動きをリアルタイムに作り出し、なおかつ学習していくことを可能にしている。」（下線は筆者による）。つまり、だれもオルタの動きを予測することができない、ということである。「オルタ」は私たち人間が使う言語で話すことはできない。しかし、歌を歌う（音を出す）ことはできる。著書では「声そのものもオルタの身体運動に合わせて周波数を削ったりしている。口はその音量に合わせて開くようにした。その意味でこの音声はリアルタイムで生成されている。」

と説明している（下線は筆者による）。「オルタ」の声は YouTube 上にアップされているので、興味がある方は聴いてみていただきたい。

AI 指揮者として音楽を指揮したのは「オルタ 2」である。2017 年オーストラリ

ア、2018 年日本科学未来館でアンドロイド・オペラ「Scary Beauty」で「オルタ 2」はボーカルと指揮を担当している。オペラを作曲したのは渋谷慶一郎、オーケストラ演奏は国立音楽大学学生・卒業生有志オーケストラ 30 名である。朝日新聞（2018 年 8 月 27 日付）の記事の見出しは「指揮・アンドロイド生まれた音楽は『呼吸を通じ奏者と相互理解』」。オルタ取材した安部（2018）はその記事の中で

練習を始めた当初、奏者たちは、オルタ 2 の踊るような予測のつかない動きに合わせることができなかった。周期的な肩と腰の上下動を加えると、「呼吸」が感じられるようになった。拍を刻む音を補うイヤホンを次々に外し、指揮を見るだけで音を合わせるように。バイオリンの一人は『一緒に音楽をしているような感覚が生まれた。本番では、本当に彼が指揮をしていると感じた。』

と述べている。また、このプロジェクトでコンセプト・作曲・ディレクション・ピアノとして携わった渋谷（2018）は

アンドロイドによる指揮の開発とオーケストラの協働が困難を極めた。プログラムは何度も開発を繰り返し幾重にもレイヤーされ、修正するというプロセスは『人は何をリズムと感じるか？心拍数と感じるか？』という問題に辿り着いた。（中略）アンドロイド相手に無人のリハーサルを繰り返した。その結果、『アンドロイドの指揮の腕の動きだけに注力するのではなく、人間の呼吸のような肩、腰の上下運動の反復を与える』ことがオーケストラとアンドロイドによる指揮のコミュニケーションを格段に容易にするという結論に達した。実際、その後のリハーサルではオーケストラメンバーが『この指揮ならイヤホンでクリックを聴かなくてもいけるかも』と言いながらイヤホンを外してアンドロイドの指揮だけを見て格段と自由に演奏するメンバーが続出した。これは普遍的な問題でもある。つまり、人間がアンドロイドに対してできるのはプログラムしたり自立性を与えたりすることだけではない。人間のオーケストラはアンドロイドと一緒に演奏する過程で人間のそれとは確実に違う指揮から、いかに一緒に演奏することが可能かを探り、発見できることもできる。このことを僕たちは忘れやすい。（中略）最初にオーケストラメンバーにこのプロジェクトを話したときに若い団員から言われた言葉、『どうやって息をしていない指揮者とわたしたちは演奏したらいいんですか？』に一周回って帰結する。僕たちは呼吸をして影響し合う。

と述べている。オーケストラ奏者は、ただただ指揮の軌跡を見ているのではない。指揮者の表情、視線、体の動き。指揮者の動き一つ一つを全身で感じ取り、その空気を共有していく。指揮者が打点を打つ直前の呼吸に合わせて奏者も息をする。そうすることによって一気に緊張感が高まり、音となってあふれ出来る。しかし「オルタ2」は呼吸をしない。その代わりに呼吸をまねた動作を入れることによって、奏者がそれに「合わせる、合わせようとする」ことから、音楽が始まるといえる。

2019年、「オルタ」はさらに進化する。「オルタ3」は、両目のカメラ、口からの発声機能といった、人間により近いセンサーシステムや表現能力に加え、歌唱のために口周りの表現力を向上させた。また、オルタ2よりも出力が強化され、身体表現の即時性の向上やダイナミックな動きが可能となった。

渋谷（2019）はインタビューの中で

これまで人間は、人間が表現する歌や演奏、スポーツなどに感動してきた。アンドロイドやロボットなど、人間『のようなもの』に対して感動できる時代が来つつある。人類史的に大きな転換期になる

と述べている。

(2) 「オルタ」の開発から見えてくるもの

「オルタ」は全く「人間」にみえない。むしろ機械的（ロボットの）言っている。この点について、石黒の考えは前述している。さらに開発メンバーの一人である大阪大学基礎工学研究科講師の小川浩平（2019）は、「見る人の想像力をかき立てる姿にした。」とデザインの意図を語っている。年齢も性別も、見る人によって多様であり、この記事を書いた記者は「30歳ぐらいの女性に見えた。」と語っている。私は映像を見てオルタ3は20歳後半男性と見ていた。機械的だからこそ、表情、動きといったわずかな情報から、人間はイメージを広げていくのだろう。渋谷（2019）も

観客がオルタ3に自分自身を投影して見る。ある人には笑って、ある人には泣いて見える。鏡のようなものだ。自分の心理がそのまま、オルタ3の姿に投影される

と述べている。

では、音楽そのものはどうだろうか。オルタ3は思いや意図をもって指揮を振っているのではない。オルタ3は組み込まれた楽譜データに基づき、振っている。そしてその楽譜データを入力しているのは「人間」である。テンポをどうするか、強弱変化をどこで、どのように付けるか等は、「人間」が考え、オルタ3に渡している。オルタ3はそのデータを元に、これまでの経験で得たデータをと合わせて、リアルタイム

で表現を作り出し、その結果をさらにデータ化して学習する。そのように集めた膨大のデータの中から、最適なものを抽出し、動いているのである。従ってオルタ 3 の意思で行っているわけではない。このことから、その音楽がオルタ 3 の表現だった、とは言えない。しかし、オーケストラの奏者が「一緒に音楽をしているような感覚が生まれた。本番では、本当に彼が指揮をしていると感じた。」と言っているように、人間側に「合わせよう」とする意思が、共に音楽を奏でる一種の共同体としての思いを引き出していくのではないだろうか。

聴く側がオルタ 3 に自分自身を投影し、奏者と同じように「聴き取ろう、この時間を共有しよう」としていった先に、人間『のようなもの』が創り出すものに共感・感動する日もくるのかもしれない。

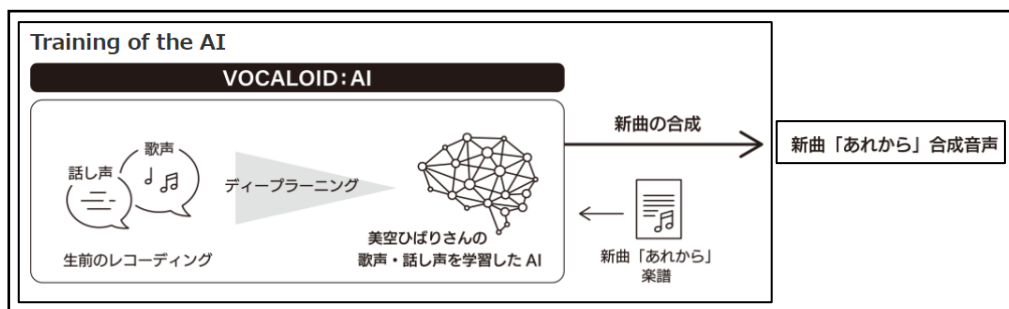
3 VOCALOID:AI「美空ひばり」

2019 年の大晦日、美空ひばりの新曲が披露された。画面に映し出された「AI 美空ひばり」は、遠目には本人と見間違えるほど、精巧にできていた。そしてその歌声。VOCALOID（ヤマハが開発した歌声合成技術）とは思えない、なめらかな発音そして生前の美空ひばりさんの癖を見事に再現し「おそらく本人だったらこんなふうに歌っただろう、語っただろう」と思わせるものだった。カメラが観客を映し出すと、じっと凝視する人、メロディーに合わせてうなずきながら聴く人、ハンカチで目頭を押さえている人。この曲を聴いて心を動かされて人がたくさんいたことは、映像を見ても明らかである。そしてその後の反応も、「すばらしい」「AI によってよみがえった」「死者への冒涇」等議論を呼んだ。ここでは、音楽の部分にスポットを当て、どのようにこの歌声がつくられていったのか、を開発者のインタビュー、資料を基に述べていく。

(1) VOCALOID:AI「美空ひばり」ができるまで

「VOCALOID」（ボーカロイド、略称ボカロ）は、ヤマハが開発した音声合成技術及びその応用製品の総称である。メロディーと歌詞を入力することでサンプリングされた人の声を元に、歌声を合成することができる。初音ミクをはじめとするバーチャルアイドルの歌声や作曲などに活用されている。教育の間でもこの「VOCALOID」を活用した授業実践が行われている。一方「VOCALOID:AI」はヤマハが開発した人工知能技術を搭載した歌声合成技術で、あらかじめ目標となる歌手の歌声を収集し、そこに含まれる音色や歌いまわしなどの特徴を深層学習技術（ディープラーニング：多層に積み重ねたニューラルネットワークを用いた学習システムのこと）により学習することで、その歌手独特の癖やニュアンスを含んだ歌声を、任意のメロディーと歌詞で作りに出すことを可能にした（資料 1）。

資料1 ディープラーニングの仕組み



この「VOCALOID」と「VOCALOID:AI」の違いを、第一研究開発部 AI グループの大道（2020）は

「VOCALOID:AI」の特長は、『この楽譜に文脈を与えたらどう歌うか』というのを想定して歌ってくれる点です。これは、歌手の実際の歌声を AI が学習すれば、ある程度その人らしい歌声を『自動的に』作り出せるということです。これは、すでに商品化されている VOCALOID との大きな違いです。VOCALOID の場合、最終的には人間であるユーザーが、ユーザーの頭の中にあるイメージを実現するため、歌い方を事細かに指示すると、VOCALOID が 100%従う形で歌います。

と述べている。そのため、「VOCALOID:AI」には歌手の実際の歌声、今回であれ奈美空ひばりの生前の音声データを大量に学習させる必要があるのだ。しかもそれは「ある程度その人らしい歌声を作り出せる」レベルであり、事細かな指示は「人間」が行う必要がある。この点について同グループ才野（2020）は『『どの時代のひばりさん』『どんなシチュエーションで語るひばりさん』の声を望むかという点は、われわれ人間の意思を込める必要があります。』と述べている。AI に「こんなふうに歌いたい」という意思はない。人間が「ひばりさんはステージに 30 年ぶりに立って、どんな気持ちになるのだろうか。」「この 30 年間の出来事（震災、コロナ禍など）を振り返ったとき、この曲にどんな思いを込めて歌うのだろうか。」といったことを考え、AI にリクエストを出す。AI はその情報を基に、何通りもの歌い方を返してくる。それが合っているかを判断するのは、人間である。大道はこの作業について

AI にリクエストを出しながらの音作りとは言え、返ってくる歌声も語りも、元はあくまでも純度 100%のひばりさんの声のみです。ただ、AI は時代や企画の意図を理解することはできないため、われわれが『今回のひばりさんならこの歌をこう歌うだろう』ということをイメージしつつ、

『いつの時代のどの曲みたいに歌ってほしい』という形でリクエストを出すのです。

と述べている。わずかなビブラートの違いや発声の違いによって、聴く側の受け取り方は違ってくる。思いや意図をもたない AI だからこそ、この作業はリクエストを出す人間の「感性」にゆだねられる部分が多い。

(2) 聴き手の反応

「歌声」に関して、番組や様々なサイトからどのように感じたか、視聴者の声を挙げてみる。

まず、美空ひばり (AI 歌唱)「あれから」(メモリアル映像)の中で、「あれから」を聴いた後、村上(2019)は「何か少し足りない感じもするんですけど。正直、この音源から。でもなかなか今生きていて普段出会わない、何か美空ひばりさんの声の暖かさというものは感じた。」と述べている。開発関係者を前にしての発言なので鵜呑みにはできないが、おそらくそれまであった「AI が歌う＝機械的な歌声」とは違う「何か」が感じられたのだろうと推察できる。

ツイッターなどの感想をみていくと、肯定的な意見としては「歌声の抑揚なども上手に表現されていたと思う。」「私は美空ひばりをリアルタイムで聴いたことがなかったが、AI 美空ひばりの新曲の声にボロボロ涙が出た。」等があった。一方で否定的な意見としては、倫理面(死者への冒瀆)や映像に違和感があるといった意見が多い中で、歌声に搾っていくと「合成音声っぽさが残っていて残念。」「子音などの発音があまい。」「技術的な面で、ここまで歌えることはすごいと思う。が、美空ひばりかといわれると、そうではない。」といった意見がみられた。

3 AI が創り出す音楽、人間が創り出す音楽

(1) 模倣から見えてくるもの

作曲家、音楽理論家、情報科学者という顔をもつ David Cope は、作曲システム Experiments in Musical Intelligence (EMI) を開発し、1981 年から 2003 年の間にバッハ、モーツァルト、ショパンといった 35 名の作曲家風音楽をつくりだした。その中には、ベートーベンの未完の交響曲第 10 番も含まれている。1980 年代と言えば、初の人工知能ビジネスが立ち上げられるとともに、機械学習の研究が本格化した時期である。しかし、1980 年代後半になると、知識獲得とメンテナンスのコスト高が開発の壁となり、それまで順調に拡張路線を歩んできた人工知能研究者が実世界問題の難しさを思い知らされ、初めて大きくつまづいた時期でもあった。このような時期に EMI によって創られた音楽は、現在も YouTube で聴くことができる。興味がある方は、ぜひ聴いて頂きたい。わたしは、「ショパン風ノクターン」と「モーツァルト風ソナタ 2-1」を聴いてみた。確かに、「ショパン風」であり「モーツァルト風」ではあったが、模倣の域を出ていない。従って最初は「確かに〇〇風だ。」と感

じたものの、途中から飽きてしまった。皆さんはどのように感じるだろうか。

「AI 美空ひばり」や実在の作曲家の作風をまねた「〇〇風」の音楽は、オリジナルを超えることはできない。聴衆である私たちは、無意識にオリジナルとそうでないものを比べ、「（似ているから）歌声に感動した」、「この音色や音の運びがオリジナルそっくりである」と承認する、あるいは「こんな歌い方をするはずがない」、「こんな構成は絶対にしないはずだ」と拒絶する、どちらかに分かれるだろう。

一般的な聴き手は「AI はどれほどの音楽を創り出すのか」という好奇心をもって聴く。AI 技術者は「AI 技術がどれほど『人間』が創るものに近づいたのか」に関心をもつ。そして音楽に携わる者（演奏者、作曲者、指揮者、評論家等）は、「AI に創造的な音楽を創れるはずがない。AI が創った音楽は「音楽」ではない」といった先入観をもって聴く。この点について、Cope（2005/2019）は次のように述べている。

聴衆はコンピュータが作った音楽を、創造的に制作された芸術とはみなせないがために、過小評価する。（中略）彼らにとってコンピュータが作った音楽とは美的な経験というよりは哲学的挑戦である。以下に示すように、聴衆の感想に現れるであろう賛辞の多くは、音楽を直視することを避けるための言い逃れである。

たとえば、私のプログラムの音楽は本当の作曲でないとか、創造のプロセスに光を当てていないなどと批判されている。（中略）批評家たちが人間の作った曲、特に歴史的な様式の音楽を美術的に評価するとき、恒常的に用いることばは型にはまった狭い意味で適宜されている。したがって、コンピュータが作った音楽を評価するために用いるのは、不可能ではないにしても困難である。しかしながら多くの聴衆は、その音楽がコンピュータ作曲と知らなければ、感情が深く反応するのを経験するはずである。そしてほとんどの人は深い疑問を抱かない。

Cope が開発した EMI は、20 年も前のものである。現在は AI 技術も進化し、人工知能はディープランニングによって「画像認識」及び「音声認識」を手に入れた。おそらく、EMI 以上に「〇〇風」な音楽を簡単に創り上げていくだろう。（実際に、そのような作曲アプリを私たちは簡単に手に入れることができる。）しかし、コープの言う「コンピュータ作曲と知らなければ、感情が深く反応するのを経験するはずである。」というのには、疑問を感じる。なぜなら、現時点でも AI が過去に創られたデータから音楽を再構成する、あるいは今流行している要素を付け加えて、それらしく創り出しているにすぎないからである。そして聴衆である私たちは、その音楽がコンピュータ作曲であるかどうかではなく、その創り出された音楽

が、自分にとってどういう価値があるのかによって判断し、感情が深く反応するのである。

したがって、人間が創り出す音楽であっても、パターン化された過去の模倣や、聴衆である私たちが価値を見いだすことが出来なければ、感情が深く反応することは決してないのである。

4 おわりに

Cope (2005/2019) はコンピュータの活用について、次のように述べている。

EMI は、単に私に時間と気力があるなら手で書くであろう曲を作曲しているだけである。EMI の作品が 1000 曲以上公開されたという事実は、私が自分の著作の中で主張してきたとおり、作曲のリバースエンジニアリングが可能であることを証明している。私がバッハに似たコラールを作曲した時は、これらの作曲原理に基づき手で作ってみたのだが、プログラムによって同じタスクを行うより 8 万倍も長い時間を要した。(私は 5.5 時間を要したが、コンピュータでは 0.25 秒である。)

同様に、岡田(2019)も次のように述べている。

人工智能に作曲らしきことができるのは、別に驚くようなことではない。とりわけパターン抽出と順列組み合わせの可能性の網羅という点で、それは大いに作曲家の時間節約の役に立ってくれるはずである。(中略)

しかし AI を有効にツールとして使えるためには、主体性はあくまで作曲家自身になくってはなるまい。それはつまり「AI におまかせ」にはしない、全自動化はしないということだ。単に売れ線のコード進行に売れ線のリズムパターンをつぎはぎして、次から次へダンス・ミュージックなどを作ってきた職業作曲家たちが、これまではいた(今もいる)だろう。別に批判しているわけではない。そもそもダンス・ミュージックや BGM の類は、誰もそれに注目して耳を傾けたりはしないから、それはそれで構わないのだ。しかし少なくとも確実に言えるのは、AI による自動作曲が進化すれば、パターンの寄せ集めだけで曲を作ってきた「主体性がない」人たちは、真っ先に失職するということだ。(中略) マニュアル対応を絶対にしない——まさにこれこそが偉大な音楽家たちの共通点である。大芸術家はパターンから入ってパターンを破る。だからこそ彼らは大芸術家なのだ。

奏者も同様であろう。ただただ譜面通りに演奏するのならば、自動演奏で十分である。譜面に示されていない部分、例えば休符にも満たない「間」や強弱の振れ幅、タッチの違いによる音色変化などの表現は、演奏者によって変わってくる。AI ならば、この部分もひょっとしたら出来そうである。ただし、データを必要とする点で、やはり模倣の域はでないのである。作曲者の意図を楽譜から読み取り、解釈し、さらに自身の音楽的感覚や感情、経験などを加え、二つとない音楽が創り出していくことこそが、人間だからできることであると考ええる。

では、聴き手はどうであろうか。

AI の技術が発達すれば、聴いただけではそれが「人間」が創り出したものなのか、AI が創り出したものなのか、判別は難しくなるだろう。ましてや、AI が創った曲を人が指揮し演奏したら、その違いはますます分からなくなるだろう。そのような音楽が溢れる中で、大切なのは、「自分にとって、この音楽はどう感じるか」ではないだろうか。同じ音楽を聴いても、「心地よい」と感じる人もいれば「不快だ」と感じる人もいる。自分はこの音楽のどこに「心地よい（あるいは不快）だと感じたのか」を考えることが必要なのではないか。「思考する」ことこそが「AI」と人間とを分ける大きな視点になると考える。

【引用文献】

将棋

人間と機械のあいだ 心はどこにあるのか 池上高志＋石黒浩 講談社
朝日新聞（2018 年 8 月 27 日付）

ATAK Android Opera Scary Beauty 2018

読売新聞オンライン（2019 年 3 月 22 日）

美空ひばり VOCALOID0:AI—AI に関する取り組み— ヤマハ株式会社
女性自身（2019 年 9 月 30 日）

東京新聞 TOKYO Web（2020 年 2 月 4 日）

人工知能が音楽を創る デイヴィッド・コープ 音楽之友社

音楽と出会う 21 世紀的つきあい方 岡田暁生 世界思想社

【参考資料】

人間と機械のあいだ 心はどこにあるのか 池上高志＋石黒浩 講談社
朝日新聞（2018 年 8 月 27 日付）

読売新聞オンライン（2019 年 3 月 22 日）

NHK スペシャル AI でよみがえる美空ひばり 新曲「あれから」 NHK

美空ひばり（AI 歌唱）「あれから」（メモリアル映像）

マンガでわかる人工知能 三宅陽一郎監修 池田書店

小学校学習指導要領（平成 29 年告示） 文部科学省

音楽と出会う 21 世紀的つきあい方 岡田暁生 世界思想社

本物の思考力を磨くための音楽学 泉谷閑示 ヤマハ

人工知能が音楽を創る デイヴィッド・コープ 音楽之友社