

研究報告

ベートーヴェン ミーツ トムとジェリー
ー AI (人工知能) による視聴覚記憶のリミックス ー

Beethoven meets Tom & Jerry

- Remixing audiovisual memory with AI (Artificial Intelligence) -

古川 聖

Kiyoshi FURUKAWA

東京藝術大学

Tokyo University of the Arts

森本 洋太

Yota MORIMOTO

Coton Inc.

Coton Inc.

概要

私たちは昨年 (2020) の 12 月に「AI-Beethoven」という Beethoven の生誕 250 年の為のイベントにて、アニメーション「Tom & Jerry」(音付き映像) の 1 シーンを断片化し、音声を中心にその特徴量を取り出し、それらを Beethoven の「エリーゼのために」の演奏データ (音声) を教師データとして並べ替えるという、AI の手法を開発し、作品の制作を行った。それは「Tom & Jerry」の断片から成り立っているが、その中から「エリーゼのために」が聴こえてくるという、両義的な音楽、視覚体験をめざす作品である。これは既存の音楽や映像作品、いわば文化的遺産をマテリアルとして断片化、再構成し、新しい表現へと統合する行為であり、広い意味で、20 世紀の終わり、ポストモダン以降に現れた新しい音楽の表現方法、美意識であるリミックスであるといえる。私たちの日常を取り巻く、これらの文化的遺産はデジタル化、数値化されることにより、コンピュータコードレベルで、AI などのデータサイエンスとの自由な組み合わせが可能であり、とりわけメディア理論家の L. マノビッチが“トランスコーディング”(L. マノヴィッチ, 2013) とよぶ、音楽という一つの領域をこえた (たとえば映像、経済統計、気象、環境、etc.) データとの異種交配による、新しい全く未知、未体験の表現領域が現前している。本報告はわたしたちの研究、制作、実践を通して探索している、これらの新しい領域の持つ新しいアスペクトと私たちが開発した技術の概要を報告するものである。

Last December, at an event called “AI-Beethoven” to commemorate the 250th anniversary of Beethoven’s birth, we fragmented Tom & Jerry (video with sound), extracted features, and sorted them using the performance data (video

+ sound) of Beethoven’s “For Elise” as teacher data. In other words, like a portrait made up of fruits by the Italian Mannerist painter Artimboldo, it sounds like “For Elise,” but it is made up of fragments of Tom & Jerry. This is an act of fragmenting and reconstructing existing music and video works as materials and integrating them into a new form of expression. Nowadays, these cultural heritages and materials have been digitized and converted into numerical values, and computer codes have made it possible to combine them with data science such as AI. In particular, media theorist L. Manovich calls it “transcoding”, and a new, completely unknown, and inexperienced field of expression is emerging through the cross-pollination of data that transcends the single domain of music (e.g., video, economic statistics, weather, environment, etc.). This paper reports and discusses the aspect of these new fields through our research, production, and practice.

1. はじめに

私たちの研究プロジェクトは、1) 私たちを取り巻く、いわば文化的遺産として現前するデジタルデータと、2) AI などの新しいデータサイエンスなどの科学技術の両者が交差するところにあり、研究の目的はリミックスという音楽の表現方法、美意識に現在の AI などのデータサイエンスの手法を組み合わせ、実際に表現行為を行い、そこに現れる新しいアスペクトの広がりを探ることにある。リミックス (編集と再構成) は当初、音楽の領域でアナログな方法、レコード、テープ媒体などを使って行われてきたが、現在、それらのマテリアルの多くはデジタル化、数値化され、コンピュータコードによって一元的なデータとして扱われ

るという全くあたらしい曲面をむかえ、私たちはこれらのマテリアルをAIなどのデータサイエンスと組み合わせる手法を開発した。私たちが注目するのはメディア理論家のL. マノビッチがトランスコーディングと呼ぶ、領域をこえた、たとえば映像、経済、気象、環境、その他のデータの異種交配による、新しい手法による全く未知、未体験の表現領域である。私たちは自ら開発した、ソフトウェアを用い、この未体験の表現領域において、それらの新しい方法の結果を実際に目と耳と身体を通して検証行う。

2. 研究の背景と特徴

以下、この研究プロジェクトを特徴づける4つのアスペクトを説明する。

2.1. 過去の文化的遺産とAIの出会い、新しいリミックスのかたち

個人情報なども含め、私たちの生活は様々な情報、データに取り巻かれている。表現やアートに関しても同様で、WEBを初めとした情報空間には溢れるほどの様々なデータが堆積、日々累積しつづけている。これらの過去からの文化的遺産とも言えるデジタルデータをどのように現代、そして未来に向けて活用していくかは私たちの文化にとって重要かつ喫緊の問題でもある。創作とは新しいものをゼロから作るもの、創作、創造は歴史的コンテキストの中で一度だけ起こる、などなどいろいろな議論がありうるが、私たちが思い描く新しいリミックスのかたちが持つ方向性は、既存の表現を直接切り取って貼り付けるコラージュ的手法ではなく、既存の表現をAIなどのデータサイエンスを活用し、そこから直接の表現、表現の表層と共に、その表現を様々な角度から分析し、メタデータをも取り出しそれらを自由に組み合わせて活用することである。

2.2. 現代の音楽聴取の特殊性、認知と長期記憶

現代の音楽の聴き方、聴取形態は音楽の複製技術の発明、発達に従い、とりわけここ、30年くらい、音楽のデジタル化と情報通信網との結びつきにより、大きく変化している。ベートーベン自分の作品であるにもかかわらず、“運命”交響曲をおそらくは10回も聴いていないと思われるが、現代のベートーベンファンは複製音源を通して数百回聴くこともあり、いわばその音楽は長期記憶に深く入りこみ、結果、現代の私たちは“運命”を耳にこびりつくほどよく知っており、私たちはその断片を聴いただけでも“運命”全体が長期記憶から引っ張り出され、その全体のイメージが想起されるような状況にある。私たちの作品、Beethoven meets

Tom & Jerry も、このような私たちの脳に刻み込まれた長期記憶の存在を前提としている。つまりそれはTom & Jerryの断片から成り立っているがその中から「エリーゼのために」に聴こえ、染み出してくるという、二重の音楽体験をめざす作品のため、このような強固な記憶／認知の仕組みが利用された。（複製技術が現れる前、ベートーベンのような音楽は主にモティーフなどを使った特徴を手掛かりに、短期記憶を中心にグルーピング、階層化され聴取されていた。）

2.3. 自由な表現とコピーライト

この研究プロジェクトの実施にあたって、当然、まずコピーライトが問題となることは明白であろう。つまり、多くの過去の作品、表現などには著作権が存在するケースが多く、リミックスという方法に対して法律によってブロックされているという、リミックスという行為、つまりアートとか表現行為とは別レベルの事項がここにある。盗作、剽窃、改変、引用など、過去の文化的遺産をリミックスというコンテキストで扱う場合、これらのことに実際に向き合っていかななくてはならない。一見、これらの概念、コピーライトとのようなものは一定の合理性を持つと言えるが、これらは近代の概念であって、よく知られているようにバウハウスの時代には多くの技法や形式、和声法などは共有の財産とされていた（これがバロック音楽とか、ロマン派とか古典派などの様式をうみだしたともいえる）。現代において明確なルールの不在のまま知的財産保護の名目で行われている大企業や国レベルでの過剰ともいえる権利ビジネスの存在と、情報のグローバル化とポストモダンの文化的成熟の必然として現れてきたリミックスのような表現との関係をどうするか、過去からの遺産と自由な創作の関係について考え始める機会になった。（結局、私たちは文化・メディア関連の著作権に明るい東京芸術大学の顧問弁護士、福井 健策氏に複数回相談し、調査を依頼し、いくつもの候補のなかから、戦前の映画「オズの魔法使い」の中の音楽がメインに使われていない部分には著作権がないとの判断により、これを用いた。Tom & Jerryはグレーズーンにあり、今回はこれを回避した。残念！）

2.4. AIを使った表現、作品ではなく、新たな音楽体験

AIをアートや表現に用いることはこれまでも継続的に行われてきた。その多くは既成の音楽のスタイルを直接の目標としたり、既知の音色（例 美空ひばりの声とイントネーション）などを目標とする実用的なものであるが、本研究のアプローチはそれらとは異なり、AIを使い、既知の音楽とそこから取り出したメタデータを共に使い、音楽作品というより、新たな音楽体

験を作り出すものである（3.1で詳説）。そして、現在は、音楽データを中心に扱っているが、本研究はジャンルや領域の異なるデータと掛け合わせた（クロスコーディング）未知の表現の探査、開拓をも指向している。つまりデータを扱うソフトウェアにとって、オーディオデータも楽譜データも動画のデータも経済や環境の統計データも同じレベルの数値表現であり、これらをAIを使って縦横に関係付けることにより、視覚的認知と聴覚的認知から社会全体のあり方などに広がる未知の表現、体験が得られるかもしれない。

3. コンサート「AI-BEETHOVEN」

2020年12月16日、ベートーヴェンの誕生日にベートーヴェン生誕250周年を記念して、ベートーヴェン作品を題材とした多数の演奏会が東京藝術大学で行われた。私たちはコンサート『AI Beethoven』に本研究で報告する、“Beethoven Meets ‘The Wizard of Oz’”を出品した。演奏などは前もって録画され、コロナ禍のもと、オンライン配信の形で行われた。（古川2020）

3.1. 作品のプレゼンテーション

コンサートにおいては著作権の関係で、Beethoven meets Tom & Jerry を諦め、“Beethoven Meets ‘The Wizard of Oz’”を演奏したのだが、コンサートでは以下の4つの作品を順番に演奏した。

- A ベートーベン「エリーゼのために」（演奏：谷原佐智）
- B 「オズの魔法使い」より一つのシーン
- C 「オズ」による マッシュアップ「エリーゼ」
- D 「エリーゼ」による マッシュアップ「オズ」

コンサートではリミックス（マッシュアップ）の素材となったオリジナル作品とリミックスされたものを順に提示する。これらの対比から聴こえてくる、関係性こそがこの作品の内容であるからだ。Aはベートーベンの1810年の作品、Bは1939年のミュージカル映画「オズの魔法使い」からの一つの短いシーンである。CはBの「オズの魔法使い」を細かく切り刻んで「エリーゼのために」に聴こえるにAIを使って並べ替えたもの、Dは反対にAの「エリーゼのために」を細かく切り刻んで「オズの魔法使い」に聴こえるように並べ替えたもの。

3.2. 作品のコンセプト（アルチンボルドとマッシュアップ）

イタリア、マニエリスムの画家、アルチンボルド（図1）を知っているでしょうか。きっと、あの花や果物、

植物の組み合わせにより描かれた不思議な肖像画を目にしたことがあると思います。全体を見ると人の顔に見えるのに、細部を見ると花や果物が見えるという視覚の二重性の揺れを楽しむ絵画です。私たちがAIを使って行うプロジェクトは、考え方としてちょっとこれに似ているところがあります。

まず、ベートーヴェンのある音楽作品（今回は「エリーゼのために」）の録音データを数千の音の短い断片に切り刻みます。そしてAIを使いその断片を全く異なるほかの音楽作品の録音データ（今回は「オズの魔法使い」）と比べ、「オズの魔法使い」に聴こえるように並べ替えます。また反対に数千の音の短い断片になった「オズの魔法使い」を「エリーゼのために」に聴こえるように並べ替えます。（AI風の言い方だとベートーヴェンの「エリーゼのために」を教師データにして機械学習にかけるといいます。詳細は4.1で後述）このように音楽を切り刻み、短い断片にしてしまい、それから新たなものを作り出す行為はDJイベントなどで使われるリミックスの手法であるマッシュアップ（Mashup）によく似ているのでここではマッシュアップという言葉を使っています。



図1: アルチンボルド『四季』より『夏』 ©furukawa

3.3. 耳の冒険、あたらしい音楽体験のデザイン

このプロジェクトの目的は、直接に音楽作品と作るということではなく、音楽が「オズの魔法使い」でもあり「エリーゼのために」でもあるといった作品の二重性を楽しむという、あたらしい音楽体験、音楽の違った楽しみ方、耳の冒険を、AI を使いデザインすることです。私たちの脳は「エリーゼのために」という良く知っているメロディーに沿って、「オズの魔法使い」の短い断片の連なりの中に「エリーゼのために」を探し組み立て、意識は記憶のなかに広がるベートーヴェンと「オズの魔法使い」との間を行き来し、めまいのように脳全体を揺さぶるような、あたらしい聴き方を体験します。

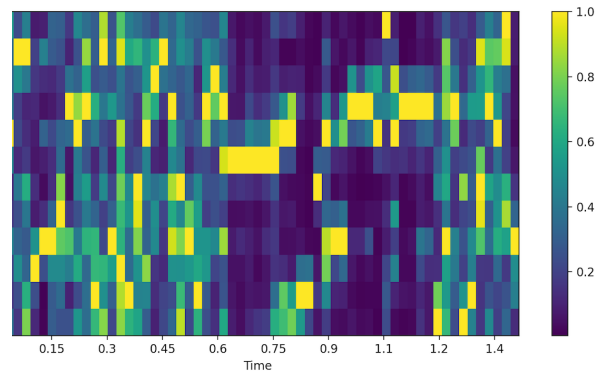


図 2: エリーゼのために

4. コンサート BEETHOVEN MEETS TOM & JERRY において用いられた技術

4.1. 機械学習を用いた mashup の手法・概念の徹底化と拡張

Beethoven meets Tom & Jerry においてビッグ・データを背景としたデータ探索や機械学習の手法と、従来のデジタル音声信号処理技術の知見を応用し、手作業では不可能な音声・映像の mashup を実現した。これは具体的には Beethoven の曲の音声ファイル X と Tom & Jerry のサウンド・トラック Y を機械学習によって分析し、非常に細かな時間単位（数ミリ～数百ミリ秒）での特徴量を抽出した後（図 2, エリーゼのために, 図 3, トムとジェリー）、X と Y の特徴量の関係性のマップ（相関マトリクス）を構成する手法をとる（図 4, 相関マトリクス）。特徴量には曲の「遺伝的情報」とも考えられるピッチクラス集合を採用した。この手法によって可能となる mashup は、例えば、Beethoven の曲の構造を Tom & Jerry のビデオに転写することであったり、Beethoven の録音素材を Tom & Jerry のサウンド・トラックの特徴量で再構成するようなことである。このような mashup は機械の圧倒的な速度やビッグ・データ処理能力によってのみ可能で、そのように拡張された mashup の手法による新しい音響と私たちの持つ Beethoven や Tom & Jerry の記憶との関係性から可能な表現を探索するものである。

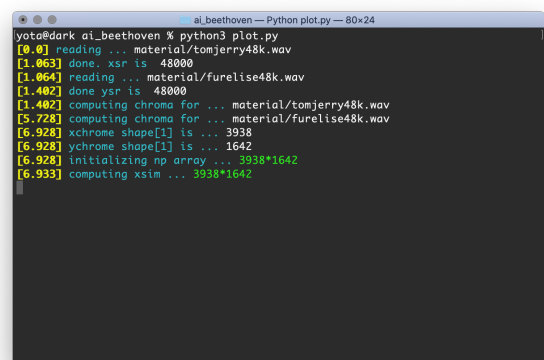


図 3: トムとジェリー

ル上で動作可能であり、コンテナ化や仮想化を視野に入れた構成となっている。Beethoven や Tom & Jerry のような素材を用いて、このシステムによって実現する作品は、それが音楽作品でも映像作品でもなく、私たちの耳の記憶と音楽／映像の間に成立するという点に於いて特異性を持ち、作曲行為は機械学習を含めた種々の自動化ソフトウェア・ツールのオーケストレーションに代替される。

5. 研究のまとめと展望

4.2. 自動化ソフトウェア・ツールのオーケストレーション

本プロジェクトを構成するシステムは、numpy、librosa、SuperCollider、ffmpeg などである。データサイエンスやコンピュータ音楽、映像編集のツールの混合から成っている点に於いてもこのプロジェクトの先鋭性が理解されるであろう。構成要素の全てが Linux シェ

本研究プロジェクトの意義はこれらの AI と音楽というまだ十分に開拓されていない領域で、新しい方向性を提案し、自ら開発した手法で、網羅的に実験を行い、その結果を公表し、その結果の意味するもの、その表現の可能性について私たちの身体を聴覚、視覚、感性、思考、その他を通し研究し、その未知の領域を開拓することである。

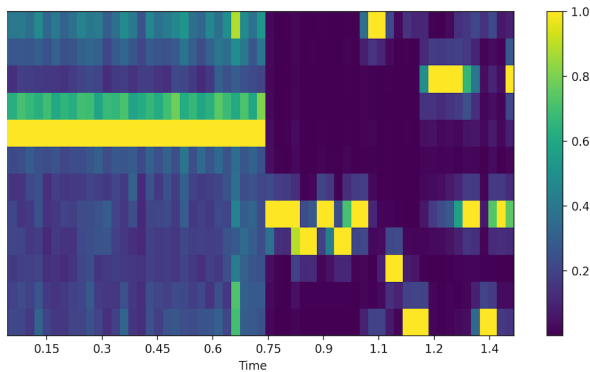


図 4: 相似マトリクス

5.1. 新しい聴覚の体験、そしてその先に...

本研究プロジェクトでは私たちの生きる現代の状況から必然的に起こる、現代の特殊な音楽聴取とその長期記憶に注目し、音楽自身ではなく、私たちの脳が音楽を長期記憶から取り出しつつ、比較し、音楽をつなぎ合わせるという、これまでに私たちが知らなかった音楽の聴き方が提示した。しかし、これは二つの音楽のみの組み合わせであり、これまで何度もふれた“クロスコーディング”の持つ可能性のことを考えれば、視覚、聴覚、以外にもいろいろな組み合わせが可能であり、どんなものがこれからこのプロジェクトを通して生まれてくるか今後の実験が楽しみである。おそらくは新しい方向性として、私たちの意識、無意識に強く働きかけてくるような音楽への方向性が開てくるような気がしている。

5.2. 過去の文化的遺産の活用と創作的行為、自由な創作のための問題提起

リミックスに限らず、過去の文化的遺産を単なる遺産としてではなく、未来へむけて、私たちの文化を豊かにするために利用することはどのようにして可能なのであろうか。現在、コピーライトなどに関し、合理的な取り決めがないことで自由な創作が妨げられているとも言える状況がある。この状況の法的な問題点を整理し、リミックスのように既成の音楽をあたらしいコンテキストに組み入れて創作をおこなうことが、ただの文化的な剽窃であるのか、それはどのような意味、意義、すばらしい可能性を開くのかを、単なる思考実験はなく、実際に行われた表現として体験可能な形にし発表し、これらに関する議論を新たに起こし、新たな社会的合意の形成にむけた問題提起を広く行う必要があるだろう。

5.3. “AI とアート” の方向性とクロスコーディング

2.4 でも書いたように AI を音楽に用いることはこれまでも行われてきた。その多くは既成の音楽のスタイルコピーを直接の目標としたり、既知の音色などを目標とする実用的なものが多かったのだが、それが創作というコンテキストではどのような意味、意義、価値を持つのかは十分は議論がなされてこなかった。私たちは研究プロジェクトを通して、AI と音楽、AI とアートという領域に新しい方向性を示すことができたかもしれない。つまりバッハのスタイルを再現するというようなことではなく、私たちの脳の認知機能、記憶に注目し、既存のデータ、文化的遺産と AI などの新しいデータサイエンス手法を組み合わせ、音楽を中心とした映像、その他の既存のデータのからクロスコーディングにより、実際に表現を生み出し、作品として定着し直接体験できるものとして公開した。私たちはこのプロジェクトの持つ多様なアスペクトをディスカッションの俎上にのせ、既存のアート、エンターテインメントの形式をこえた、過去の文化的遺産を現代に生かす、新しい活用法、アプリケーションの可能性をしめすことを継続してやっていきたいと思う。

6. 参考文献

レフ・マノヴィッチ (2013)、堀潤之訳：“ニューメディアの言語 デジタル時代のアート、デザイン、映画” みすず書房 p.69

7. 参考作品

古川 聖 (2020) *AI-Beethoven* <http://furukawalab.org/project/aibeethoven/> accessed on 6.30,2021

10

8. 著者プロフィール

古川 聖 (Kiyoshi FURUKAWA)

ベルリン芸術大学、ハンブルク音楽演劇大学にて I.YUN, G.Ligeti 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うけて、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。東京芸術大学・先端芸術表現科教授。

森本 洋太 (Yota MORIMOTO)

ハーグ王立音楽院修士、英国バーミンガム大学博士（作曲）。2006年に渡蘭、欧州を中心に活動。室内楽作品が各地のアンサンブルにより委嘱・演奏されている他、ハーグ市立美術館、Transmedialeなどでインスタレーション作品を発表。楽曲／サウンド・デザインをアムステルダム映像博物館、資生堂、本田技術研究所、NTTコミュニケーション科学基礎研究所などに提供。近年では公共空間の音環境デザインに取組み、2019年には「シンガポール空港のための音楽」を発表。



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。