

2025 年 11 月 13 日
先端技術ラボ 大城武史

NeuroTech × Art

— ②音楽編 —

《本レポートの位置づけ》

- 本レポートは、NeuroTech × Art の新しい可能性を開拓するための手がかり提供を目指す技術レポートである
- Art を取り巻く国内外の社会環境もスコープに入れつつ、主に NeuroTech に焦点を当てながら、Art の新しい表現形態とそのビジネス応用可能性を模索するのが主目的である
- NeuroTech × Art はシリーズ刊行。第 2 回目は NeuroTech × 音楽に焦点を当てる
- [第1回目のレポート①視覚芸術編](#)を読むと、より分かりやすくなる

《要点》

- ◆ 音楽とヒトの関りは長く、その効果は「身体」と「社会」に、両方に刻まれている
- ◆ 音楽はデジタル・情報技術と結びつき、巨大ビジネスへ成長（296 億米ドル 2024 年）
- ◆ 日本の音楽市場規模は世界第 2 位（3070 億円）であり、物理メディア（例 CD やレコード）の売り上げが 66%を占める特異的な構造を持つ
- ◆ 本レポートでは音楽を「時間芸術」として捉え、音楽の「多時間スケール」に着目する
- ◆ 脳には音楽の多時間スケールに対応する仕組みが存在する
- ◆ 自動作曲分野における、音楽生成 AI によるブレイクスルーはまだ起こっていない
- ◆ 次世代型系列モデルアーキテクチャ: Mamba の音楽生成有用性に注目する
- ◆ 脳 × 生成 AI: ニューロ(ブレイン)ミュージックは、AI による自動作曲を補完する
- ◆ AI 音楽サービスは普及しつつあるが、著作権・知的財産権を巡った訴訟が相次ぐ

- 本件に関するご照会は、当社ホームページの [問い合わせフォーム](#) より、先端技術ラボ 大城 宛にお願いいたします。
- 日本総研・先端技術ラボについては、[特集ページ](#) をご覧ください。

本資料は、情報提供を目的に作成されたものであり、何らかの取引を誘引することを目的としたものではありません。作成日時で弊社が一般に信頼出来るとされる資料に基づいて作成されたものですが、本資料の情報の正確性・完全性を保証するものではありません。また、情報の内容は、経済情勢等の変化により変更されることがあります。本資料の情報に基づき起因してご閲覧者様及び第三者に損害が発生したとしても執筆者、執筆にあたっての取材先及び弊社は一切責任を負わないものとします。

はじめに:本レポートの構成について

本レポートでは、NeuroTech×AIによる自動作曲技術への応用可能性を主題とする。そのため、本レポートは下記のような構成を取る。まず、レポート前半では、様々な観点から音楽を俯瞰することで、音楽の社会的価値や効用性を明らかにする。第1節では人類史に見るヒトと音楽の強い繋がりと、神経生理学に基づく音楽的効用を紹介する。第2,3節では現在の音楽市場動向について述べる。その中で、音楽市場が今後も堅調な成長が見込まれていることを紹介する。

レポート中盤では、NeuroTech×AIによる自動作曲を目指した技術ポイントを模索するために、音楽の芸術的特異性を明らかにしていく。この点において、本レポートでは音楽を「時間芸術」として捉えることで、他の芸術とは異なる音楽の特性を明らかにする(第4節)。さらに、第5節では、**音楽の時間特性を処理する脳の仕組み**について概説する。第6節では、**自動作曲における音楽生成AIの現状の課題**(例 ①音楽の時間特性の学習や②生成された音楽の独創性や新規性の欠如)について述べる。①の課題を克服するため、次世代型系列モデルアーキテクチャ: Mamba (Linear-Time Sequence Modeling with Selective State Spaces)に注目する(第7節)。さらに、問題点②を克服するために、NeuroTechを用いた自動作曲技術を紹介する(第8節)。

レポート終盤では、現状の生成AIを用いた音楽サービスを紹介し、その使用リスク等について述べる(第9節)。

1. ヒトと音楽の関わり

ヒトは生まれながらに音楽に反応し、やがて奏で出す。この特性は文化を超えたヒトの普遍的な性質の1つである。親が乳幼児に歌いかける行為は世界中で見られる自然かつ原初的なコミュニケーション手段の1つであり、親の子守歌やハミングは乳幼児の注意を引きつけ、情緒を落ち着かせ、お互いの感情共有の手段として機能する¹。少なくとも、ヒトは約100万年前には現代人に近い発声器官(声帯・喉頭)を備え、音程を制御する基盤を持ち始めていたと考えられている。このことは、発声に基づく原始的な**音楽が人類史の早期から重要な社会的役割を担っていた可能性を示唆している**²。

音楽は発声に限らない。現存する最古の楽器(骨や象牙で作成された笛)は約3~4万年前に遡る(図1)³。この発見が示すように、ヒトは早くから楽器を創造し、音楽を拡張してきた。実際、音楽は様々な社会的な場面(例 宗教や共同体)で、**その機能性を発揮する**。音楽は単なる娯楽以上に、親と子のコミュニケーションから社会的な結びつきの形成まで、大きな役割を果た

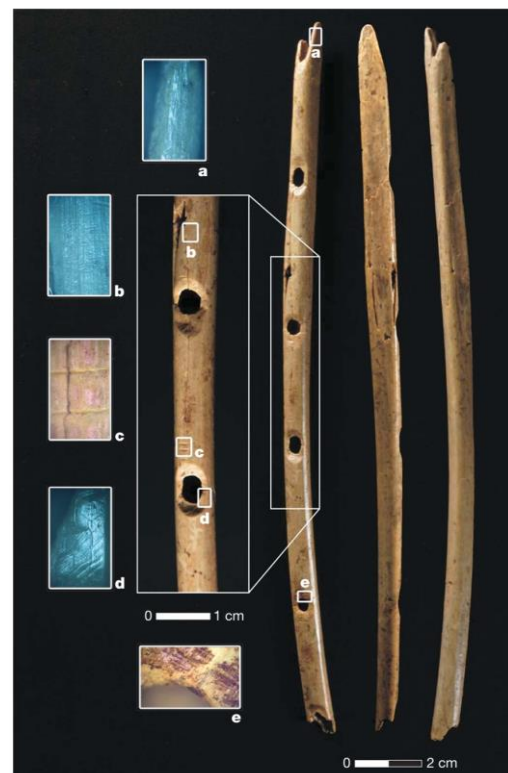


図1. ホーレ・フェルス遺跡で出土した現存する最古の骨製の笛. Conard et al. (2009)より引用。

してきた⁴。

そのような音楽の機能性を神経生理学的に理解する手段の1つとして、音楽と脳に関する研究は示唆に富む。音楽鑑賞は快感や動機づけに関わる脳領域（中脳辺縁系ドーパミン系）、気分の安定化や情動調整を担う脳領域（縫線核セロトニン系）、さらには社会的親和性や安心感を高める脳領域（視床下部オキシトシン系）を活性化させる⁵。このように、音楽の効用性は科学的にも裏付けられてきており、長い時間の中で、ヒトの身体及び社会に深く刻まれている。

2. 音楽市場（海外と日本の比較）とデジタル・情報技術

社会的結びつきと生理学的機構に働きかける音楽はデジタル技術や情報技術と結びつき、巨大ビジネスへ発展を遂げた。図2が示すように、グローバルな音楽市場（録音・ライブ演奏を含む）は拡大を続けており、今後も堅調な成長が見込まれている。過去30年間の録音音楽市場（例 CD の売り上げ）は、1999～2000年に約270億ドルで一度ピークを迎え、その後2000年代を通じて急激に縮小した（2005年207億ドル→2013年150億ドル）。この背景にはデジタル技術、とりわけMP3による音楽ファイルの普及とP2Pファイル共有（違法性のあるサービスも含む）の影響が大きかったと考えられている。その後、音楽マーケットはCD等のフィジカル販売から、楽曲ダウンロードによる新しいビジネスモデルへと移行した。代表例として2003年にApple社がiTunes Music Storeを開設。これによってユーザーは合法的に、楽曲単位でも、音楽の購入が可能となった。2000年代後半になると、定額で楽曲を聴き放題にする音楽サービス、ストリーミングサービスが登場する。最も代表的なストリーミングサービスは、スウェーデン発の企業であるSpotifyである（2008年公開、定額音楽配信で世界シェア約31.7%、世界第1位）。2025年現在、株価も公開当時から約4倍と大きな成長を果たしている。次いでApple Music（世界シェア約12.6%）、Amazon Music（世界シェア約11.1%）となっている。2024年末時点では、世界全体の有料音楽ストリーミング契約数は7億5200万件であ

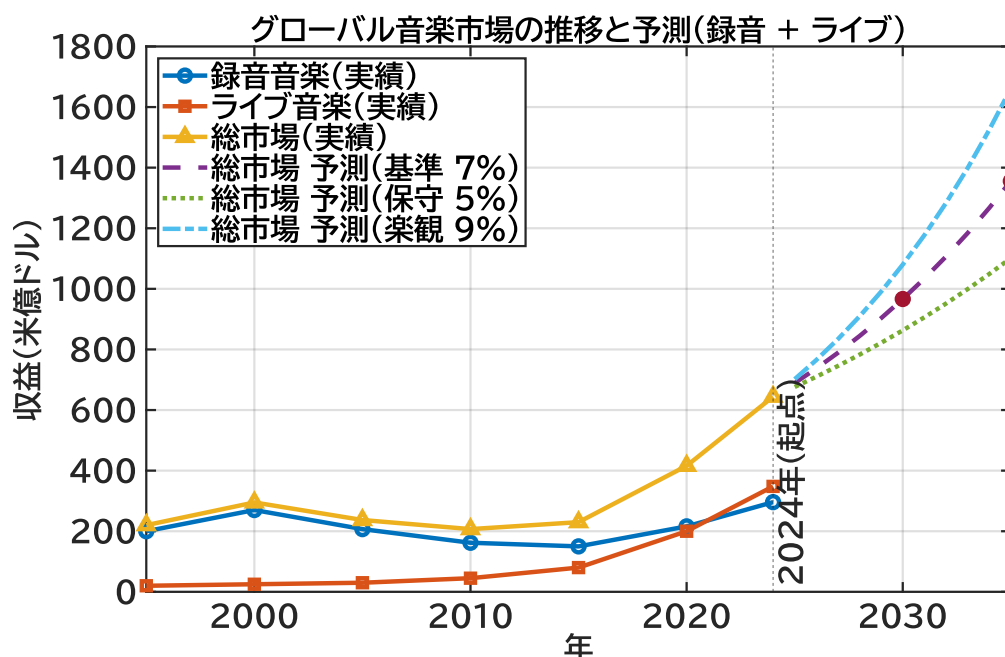


図2. グローバル音楽市場の収益推移(1995-2024年)と将来予測値。

資料出典: 国際レコード産業連盟『Global Music Report 2024 - State of the Industry』より作成。

る。音楽ストリーミングによる収益は今やグローバルな音楽市場収入の約 69%を占めている(図 3)⁶。結果として、音楽市場は 2014 年以降持続的な再成長を果たしている⁷。2023 年の世界音楽売上は前年から 10.2%増加し、286 億ドル、翌年 2024 年には過去最高の約 296 億ドルに到達している。この再成長の主な原動力(ストリーミングによる売り上げ増)に加えて、近年はアナログ・レコードやカセットテープの復活人気もあり、フィジカルの売上も 2 年連続で二桁成長(2023 年は+13.4%)するなど、デジタルとフィジカルが一部併存する状況となっている。

ライブ音楽市場はこの 30 年間一貫して成長しており、1995 年では約 20 億ドルであった収益が 2024 年には 348 億ドルへと大きく拡大している。その結果、グローバル音楽市場の 2024 年総収益(録音+ライブ)は約 644 億ドルと過去最高を記録した。図 2 では、この 2024 年の売り上げを起点とし、年平均成長率に基づく予測値を「保守シナリオ 5%成長」、「基準シナリオ 7%成長」、「楽観シナリオ 9%成長」で算出している。基準シナリオでは、2030 年には約 970 億ドル。2035 年には約 1360 億ドル相当の市場規模に到達する可能性がある。

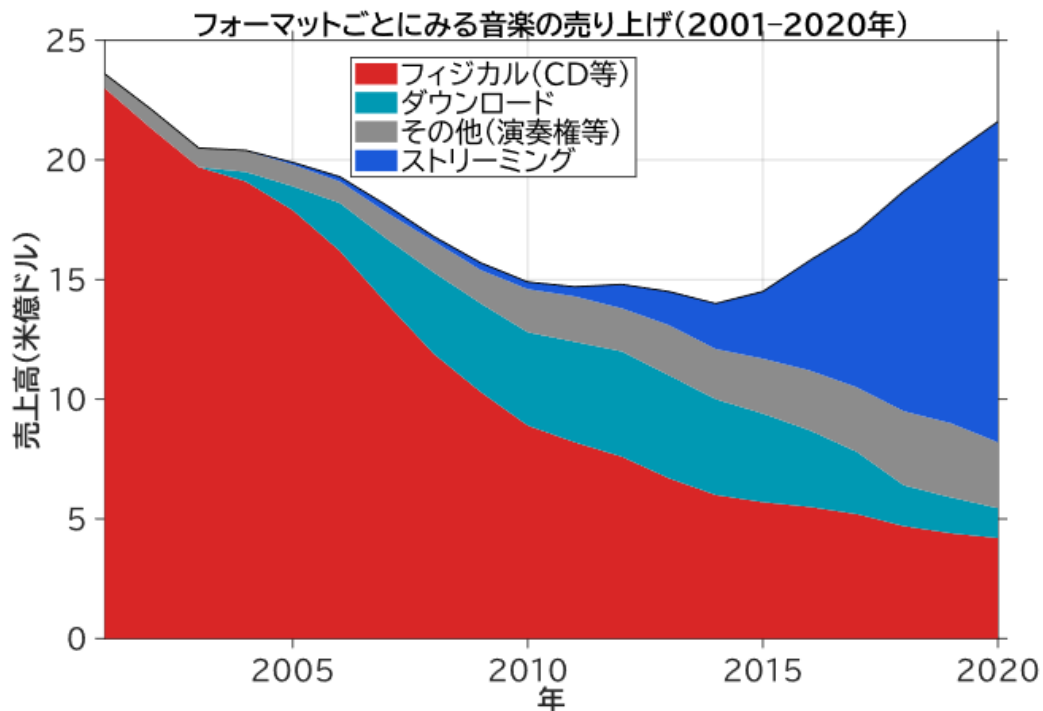


図 3. 音楽販売形式(フォーマット)ごとにみる収益推移. “Streaming Changes the Tide for the Global Music Industry”⁶をもとに再作成.

3. 海外市場と日本市場の比較

米国は世界全体の音楽市場の約 4 割を占めている。日本はその米国に次ぐ世界第 2 位の音楽市場であり、2022 年の売上は約 2.4 億ドル(3070 億円)である⁸。日本の音楽市場はデジタル音楽配信が前年比 +17%と大きく成長しているが、長期的に見ると日本の音楽市場規模はピークの 2007 年から大きく縮小しており、2022 年時点でも 2007 年比で 34%減少している。このため、今後他国の成長次第では、日本はランキングで現在の 2 位から順位を下げ、2030 年頃には 5 位前後に後退するとの予測もある⁸。

日本の音楽市場の特異性は音楽メディア消費の構造に見られる。米国では 2022 年時点で音楽売上の

84%がストリーミングなどデジタル収入で占められ、CD やレコード等のフィジカルは 11%に過ぎない。一方、日本では物理メディアが未だ市場の 66%を占め、デジタルは 34%程度に留まる⁸⁾。日本で CD 等のフィジカルが根強い理由としては、デジタルより馴染みのある CD を買い続ける層の存在と、日本独自のファン文化が主な要因と考えられている⁸⁾。しかしながら、コロナ以降はコンサートがオンライン化され、デジタル配信やストリーミングの利用が一気に拡大した。さらに音楽業界では NFT (Non-Fungible Token) やメタバース上での音楽体験など新技術の導入も進みつつある⁹⁾。今後、日本でもデジタルシフトが加速し、世界的なトレンドに追随すると予想されている。

4. 音楽の芸術的特異性：「時間デザイン」としての音楽

音楽は「時間芸術」と言われる¹⁰⁾。これは音楽を構成する主要素であるリズム(拍子)、テンポ(速さ)、メロディー、ハーモニー(和音)、音と音との間(休符、残響)、これら全てが時間によって形付けられることに由来する(図 4)。例えば、リズムは 2 つ以上の音があって初めて、その音の時間間隔から知覚できる。メロディーも同様に、2 つ以上の音が何らかの繋がり(例 反復や類似)があって初めて認識できる。このヒトの時間知覚特性に根差している音楽の性質こそが、音楽が他の芸術形式以上に「時間そのものを素材として扱い組織化する」¹⁰⁾ 芸術といわれる所以である。

音楽鑑賞において、音楽がもつ時間性は、大きく3つのスケール(速・中・遅)に分けて知覚される。1つ目は「音色(timbre)」と呼ばれる速い時間スケール特徴である。これは数ミリ秒～数十ミリ秒単位で変化する音の微細な波形構造等に由来する。音色は楽器や声の性質を決定する。同じ高さ、音量であったとしても、ピアノ、バイオリン、オーボエ、ヒトの声では全く異なる印象を持つが、それは音色が異なるためである。また、同じ楽器で奏でられた音であったとしても、その弾き方や演奏者によって音色は変わりうる。

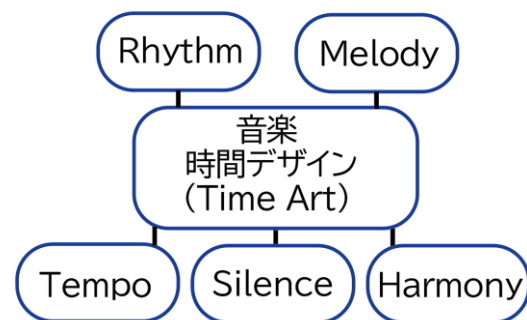


図 4. 時間デザインとしての音楽とその構成要素。

2つ目は、「リズム(拍子)」と呼ばれる中程度の時間スケールである。これは数百ミリ秒から数秒単位で繰り返される周期的なパターンであり、私たちが音楽を聴くときに自然に身体を揺らしたり、動きを取ったりする際の基盤となる。リズムは、音楽の「ノリ」や「グルーヴ感」を形成し、高揚感や安定感をもたらす。また、「ハーモニー(和声)」は同時に鳴る音の組み合わせと、その和音が時間の中で組み合わせを変えながら進行すること(時間変化)を意味し、中程度の時間スケールで知覚可能である。例えば、ヒトは協和音(例 C[ド]-E[ミ]-G[ソ])を安定的な音として知覚する一方で、不協和音(例 C[ド]-F[ファ]-G[ソ])はざわつき、緊張感、不安定感を伴って知覚する。音楽では、この不協和音を時間の流れの中で、巧みに用いることで曲に「推進力」や「彩り」を与えることが可能となっている。

3つ目は、「楽曲構造」や「フレーズ展開」に関わる遅い時間スケールでの特徴である。これは数十秒から数分単位で現れるもので、楽曲全体の物語性やダイナミクスの流れを形作る。例えば、広く認知されているフレデリック・フランソワ・ショパン(1810-1849)の前奏曲(Op.28-15 《通称 雨だれ》)の ABA' 三部形式などが、この遅い時間スケールに該当する。この前奏曲の A 部は変二長調で静かに始まり、左手の低音部

分では変二音(D♭)が途切れなく反復される(このリズムが「雨音」を連想させる)。しかし、B 部になると調性が嬰ハ短調(C# minor)に転調し、曲の雰囲気を変貌する。音色も和音の連打や強奏が増え(速い時間スケールの変化)、さながら「嵐の到来」を連想させる。一方、この転調後も左手は変二音(D♭)を奏で続けるため、曲の一貫性は損なわれない。A' では再び変二長調に転調し、楽曲冒頭の穏やかな雰囲気に回帰する。しかし、B 部分を経た前後では、同じ変二長調であっても A と A' の印象は大きく異なる。このように、音楽の時間性は「速(音色)」「中(リズム)」「遅(メロディーや楽曲構造)」という複数の階層で同時に知覚され、相互に影響を与えながら全体的な音楽体験を形成している。これらが、音楽が「時間芸術」と呼ばれる所以である。

5. 脳における多時間スケールと音楽

音楽の多時間スケールに対応するかのように、脳も異なった時間スケールで活動する。その典型例が脳波である。脳波には超低周波(～0.1 Hz)、デルタ波(約 1-4Hz)、シータ波(4-8Hz)、アルファ波(8-12Hz)、ベータ波(12-30Hz)、ガンマ波(30-80 Hz)、リップル(80-250 Hz)といった様々な時間スケール(周波数)で活動する特性¹¹があり、音楽の処理にも、この脳波特性(多時間スケール)が発揮される(図 5)¹²⁻¹⁵。例えば、音

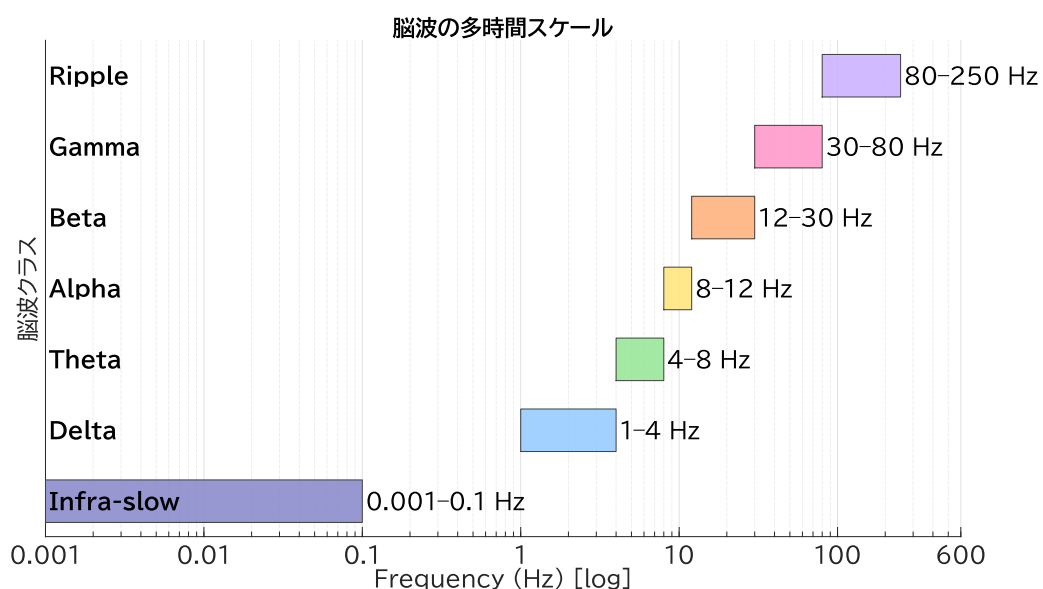


図 5. 脳の多時間スケール. 脳には異なった時間で活動する特性があり、音楽処理にもこの特性が利用される。Penttonen and Buzáki (2003)より再作成。

楽の早い時間スケール(音色の知覚)に対応する活動として、音を処理する脳領域のガンマ波が挙げられる¹⁶。先行研究では、楽器トレーニング未経験の子供でも、1年間のピアノレッスンで、ピアノ音色に特異的なガンマ波の増強が生じることが報告されている¹⁶。同様に、リズムや拍子・小節といった中程度の時間スケールにおいて現れる音楽特徴に対しては、ゆっくりとした脳波(8 < Hz)が同調することが知られている(音楽のリズムに合わせて、脳波と音楽が揃う)¹²。さらにより遅い脳波(～0.1 Hz)は音楽のフレーズ構造(遅い時間スケール)を安定的に追跡したり、時には先行(予測)することが明らかになっている¹⁷。このことは、脳は多時間スケール性を利用して、音楽構造を理解するだけでなく、脳内部に内的リズムやフレーズを生成し、予測する能動的プロセスを有していることを示唆している¹⁸。重要な点として、こうした多時間ス

ケールに基づく脳活動はパラレルに動作しながらも、場合によっては相互に連動し得るという点である。例えば、位相-振幅結合 (phase-amplitude coupling; PAC) という現象は、異なる時間スケールの脳波が相互に結びつき、さまざまな時系列情報を統合して処理するメカニズムとして知られている (図 6)^{19,20}。PAC が音楽に内在する様々な時間スケール (音色・拍・フレーズなど) の統合処理を支える神経基盤として機能する直接的な証拠はまだ呈示されていないが、音楽処理に関与する可能性は既に報告されている²¹。

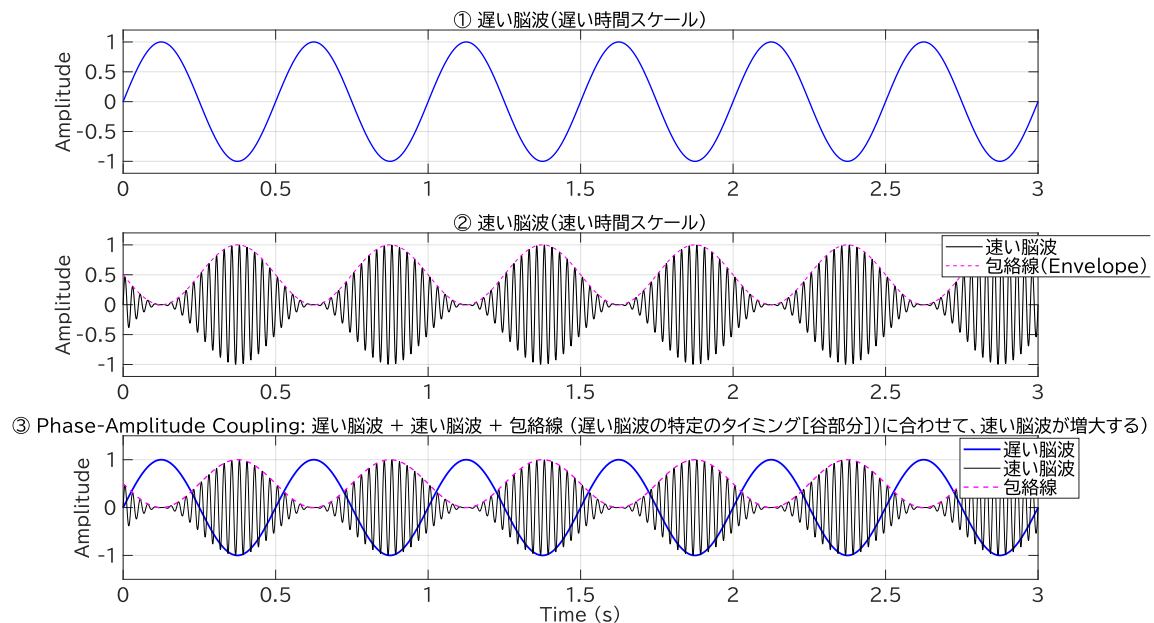


図 6. 位相-振幅結合. 遅い脳波と速い脳波の結びつき. 遅い脳波 (青線) に対して、早い脳波 (黒線 + マゼンタ) は、特定のタイミング (青線の谷部分) で大きく活動する。

6. 音楽生成 AI に「音楽」は生み出せるのか

音楽と脳の関係を探る研究が進む一方で、近年、新たな潮流として注目を集めているのが音楽生成 AI である。なかでも Transformer は自然言語処理 (NLP) 分野の発展を牽引した代表的なアルゴリズムであり、音楽生成 AI の分野においても、この構造を基盤とするモデルが数多く開発されている。同様に、画像生成において大きな成功を収めた Diffusion Model (参照: [NeuroTech × Art①視覚芸術編](#)) による音響生成の分野では、川のせせらぎや風の音などの「環境音生成」において、非常に高いクオリティに到達している。しかしながら、「音楽生成 (自動作曲)」においては、言語や画像分野で達成されたような決定的ブレイクスルーには至っていないのが現状である²²。

Transformer の利点は、自己注意機構により時系列データから構造やモチーフを自動学習できる点にある (例: MusicGen (Meta), MusicLM (Google), Jukebox (OpenAI), Music Transformer 等)。しかし、学習過程や獲得した表現の可視化は困難であり、生成される音楽は、必ずしもヒトが培ってきた音楽理論や感情の流れと整合していない²³。これは、音楽が複数の時間スケールに基づく多層的構造 (4節参照) をもつため、既存モデルではその多時間的階層関係を十分に扱えないことが主因と考えられる。Diffusion Model による環境音生成が成功しているのは、短いパターンの連続で表現できるためであるが、音楽生成では同様の時間的構造の課題が依然として残る²⁴。勿論、囲碁や将棋において AI が定石を超えた新しい棋譜を示

したように、必ずしも AI から生成される音楽を従来の音楽理論や構造に縛る必要はない。しかし、「勝ち負け」が自明ではない音楽生成において、現時点で以下 2 つの根本的問題が残されている：

- ① 長尺構造の脆弱性：曲全体のセクション設計やクライマックスの配置が散漫になりやすい
- ② 個性・新規性の不足：学習データ分布に収束しやすく、作家性や独創的変化に乏しい

これらを踏まえると、既存の生成 AI だけでは不十分であり、これらの弱点を補うハイブリッドあるいは拡張的なアプローチや、全く新規のアプローチが求められている。

7. 次世代型系列モデルアーキテクチャ: Mamba

上記の問題点①の解決策として、筆者が注目している 1 つ目の技術が、**Mamba (Linear -Time Sequence Modeling with Selective State Spaces)**と呼ばれる次世代型の系列モデルアーキテクチャ²⁵である。Mamba は従来の状態空間モデル (State-Space Model, SSM) を基盤としているが、従来の SSM は時間方向の計算ルールが固定的であり、入力の内容に応じて「どの時間的特徴を強調するか」を切り替えることができなかった。これに対して、Mamba は**選択機構 (Selection Mechanism)**を新たに導入することで、時間上でどの情報を強調すべきかを動的に選択可能になった(図 7)。

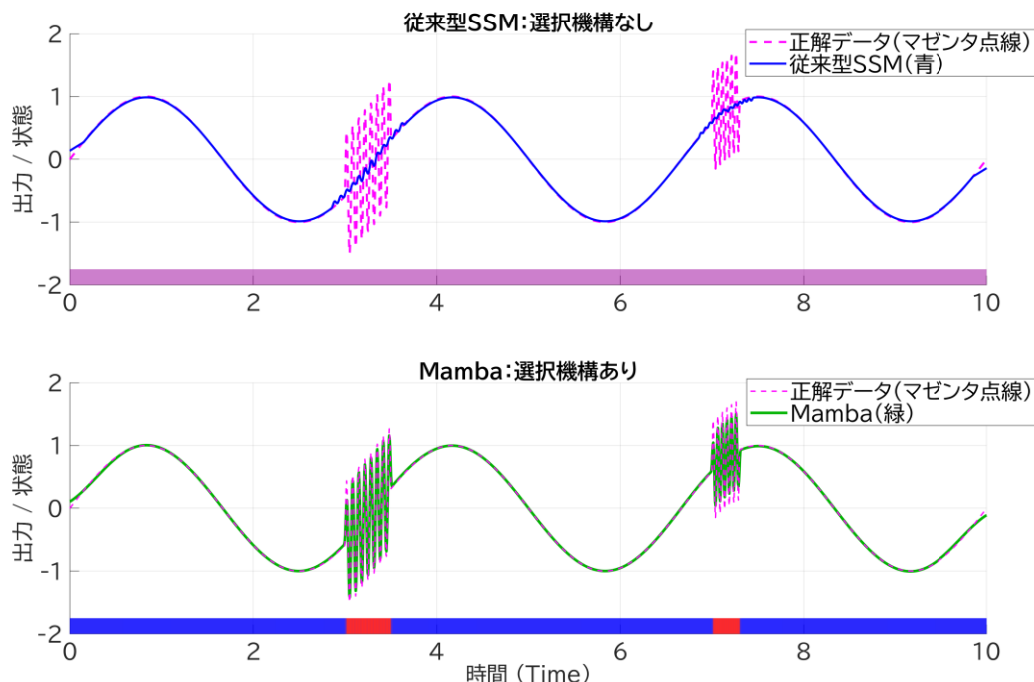


図 7. Mamba の概念図. 上段パネルは従来型の SSM 型。入力データに対して計算ルールが固定化されているため、高周波 (マゼンタのスパイク状の波形) を捉えられていない。下段パネルは Mamba の計算特性 (選択機構) を示したもの。データの変化に応じて、柔軟に対応できる。尚、両パネル下段の青～赤のグラデーションは、時間に応じた柔軟な情報選択を表している。

また、Mamba は、Transformer などを用いられる注意機構(Attention Mechanism)の最大のボトルネックである計算コストを大幅に改善し、より長尺の一貫性(例: 楽曲構造やフレーズの展開に関わる遅い時間スケール)や、局所的な重要情報(例: 音色やアタックなどの速い時間スケール)を効率的に表現できる点で注目されている。実際、Mamba の考案者である Gu と Dao の論文²⁵では、4 時間分のピアノソロ音楽データセットを用いた長文脈予測課題において、従来モデルを上回る性能を示した。特に、音楽データの時間長が長くなるほど予測精度が向上するなど、Mamba が長期的依存関係を効率的かつ効果的に保持できることが示されている。これらの結果は、Mamba が音楽のように、速い時間スケールから遅い時間スケールまで複雑に絡み合う系列データに対して極めて高い親和性を持つことを示唆しており、今後、より高度な音楽生成モデルへの応用が期待される。

8. 脳×音楽生成 AI: ニューロ(ブレイン)ミュージックの登場

前節で述べた問題点②(個性・新規性の不足)は、Mamba のようなモデルの改良だけでは根本的な解決が難しい。そこで、筆者は 2 つ目の鍵として、**脳波をはじめとするヒトの脳活動に着目する**。これは音楽に対応した脳活動パターンが明らかになりつつあること(5 節参照)に加えて、**脳波そのものが個人特性を強く反映することに基づく**²⁶。そこで本節では、脳波を音楽生成に応用しようとするこれまでの取り組みを概観する。

脳と音楽を直接的に組み合わせる試みは、大きく 3 つに分類できる。1 つ目は、脳の活動を「**可聴化(sonification)**」するアプローチである。この方法では、脳波などの神経活動を音響に変換することが主な目的である。つまり、「音楽を作る」のではなく、「脳の活動を聴く」ことに主眼が置かれている。この分野の最初期の試みとして、Alvin Lucier による実験音楽作品「[Music for Solo Performer](#)」(1965)が挙げられる(図 8)。Lucier 氏は自身の頭部に電極(脳波計)を装着し、測定されたアルファ波によって打楽器を駆動させ、音を奏でる仕組みを構築した²⁷。このアルファ波は自発脳波(参照:[NeuroTech × Art①視覚芸術編](#))

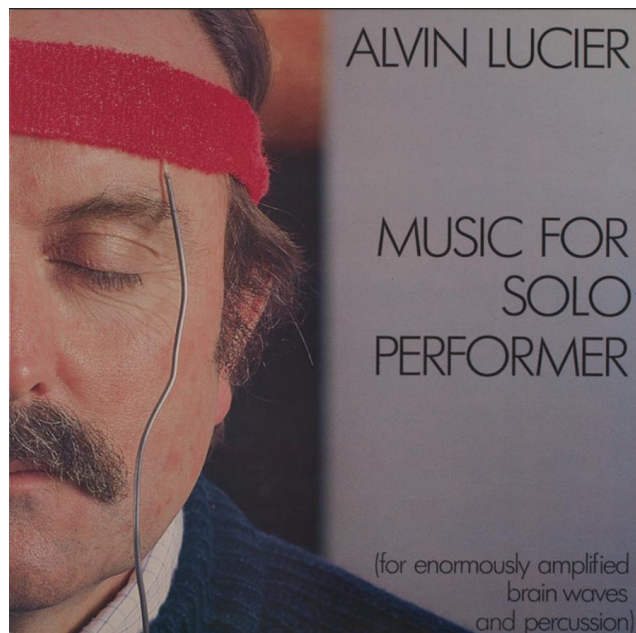


図 8. Alvin Lucier (1982) Music for Solo Performer (For Enormously Amplified Brain Waves and Percussion, 1982, Lovely Music, Ltd.)

の 1 種であり、Lucier の作品は意図に寄らない脳の自発的な活動を音に変換したニューロ(ブレイン)ミュージックの原型として捉えることが出来る。脳活動可聴化の研究は現在も行われている²⁸。

2 つ目は、**デジタル音楽装置(例 シンセ、DAW 等)を脳活動によって「操作・制御」することで、音楽を作成・演奏するアプローチ**である。これは Neuralink が手掛ける Brain Machine/Computer Interface の技術的枠組みに属し、例えば四肢麻痺の患者が脳活動を通して、PC のカーソルを動かすといった応用と共通している。実際の事例として、身体的に動作困難な音楽家が、脳活動を通じて演奏や作曲を行うプロジェク

トも既に存在する(例 [Aeon-Music of the Mind](#))。こうした試みは、身体的制約を超えて音楽創造性を発揮する新たな道を拓くものであり、Neurotech と音楽の融合が持つ社会的意義を示している。

3つ目は、本レポートが最も注目するアプローチである「脳×生成 AI」による自動作曲(脳波信号を用いた作曲支援)である。これは、脳波の可聴化や機器の操作といった従来の応用を越えて、作曲家の脳活動そのものを、生成 AI の「創作条件」として組み込む。すなわち、既存の AI が苦手とする「個性・新規性の不足」を、作曲家自身の脳活動に含まれる個性、多時間スケール、感情表現の多様性によって補完しようとするものである。このアプローチには大きく2つの方向性がある。

①条件付け型

このアプローチでは、脳活動を条件として与え、音楽理論から逸脱することなく新しい楽曲を生成する。Miranda (2006)は拡張遷移ネットワーク(Augmented Transition Networks [ATN])を用い、脳波の周波数帯域パワーや複雑性に応じて作曲ルールを動的に切り替える仕組みを提案した²⁹。例えば、アルファ波が優勢なときには緩やかな様式を、複雑性が高いときにはテンポを速めるなど、その時点での脳波特徴を楽曲の進行や構成に反映できる。このアプローチの最大の利点は、ATN に音楽的な文法や作曲家のスタイルを明示的に埋め込むため、人間が培ってきた音楽の構造や展開に忠実な生成が可能である。一方で、ルール設計のコストが高く、指定スタイルを超える創造性は限定的である。図 9 には、実際にこの手法で生成された音楽の一例を示す。



図 9. 脳波×ATN で生成された音楽。シューマンとベートヴェンの作風が反映されている。

Miranda (2006)より引用。

②感情駆動型

感情駆動型アプローチでは、リスナーに「快／不快」や「覚醒度」などの感情ラベルを付与した刺激(例 動画)を提示し、その際に記録された脳波を感情情報と対応づけて、音楽生成 AI への条件(conditioning signal)として利用する。例えば、Ran ら(2024)は、脳波から推定された感情と音楽構造的特徴の双方を入力条件として与える設計を採用している³⁰。検証の結果、Ran らのモデルは、和声の音域といった客観的特徴において既存モデルを上回り、さらに主観評価(感情表現の自然さ・豊かさ)でも高い評価を得ている。ただし、感情駆動型アプローチにも課題がある。第 1 に、感情駆動型アプローチでしばしば用いられる感情データセット(例:EMOPIA)が限られたカテゴリ(喜び・悲しみ・怒り・安らぎなど)に基づくため、生成される音楽が典型的な感情表現に収束しやすい。その結果、音楽本来がもつ多様で複雑な感情表現——たとえば、ワーグナー《トリスタンとイゾルデ》の冒頭に登場する「トリスタン和音」に象徴されるような、曖昧で多義的な感情——を十分に表現することは難しい。第 2 に、脳波と感情との対応関係は、AI 研究者が想定するよ

うな一対一的な厳密さを持たない。たとえば、アルファ波が安静時に出現することは知られているが、それが直ちに「アルファ波が穏やかな感情状態」を意味するわけではない。したがって、脳波に基づく感情推定には、神経科学的な裏付けがまだ十分ではない点に留意する必要がある。

9. 生成 AI による音楽生成サービスの市場動向

本節では、近年急速に拡大している生成 AI を用いた音楽生成サービス市場の全体像を概観する。方向性としては、①歌もの／楽曲生成サービス、②BGM 生成サービス、③研究・基盤型(R&D)モデルの3類型に大別できる(Table 1)。①の歌ものの生成系は、個人クリエイターや小規模制作チームがデモ制作や SNS 配信用コンテンツを迅速に生成する用途で拡大している。テキスト入力によりジャンルや歌詞を指定し、数分で完成曲を生成するのが一般的である。主要サービスである [Suno](#) や [Udio](#) は、英語話者を中心に急速に普及しており、個人契約の料金は概ね 10～30 米ドル/月、上位プランでは 100 米ドル前後で追加クレジット制を採用している。②の BGM 生成サービスは、映像・アプリ・店舗などでの常用 BGM 需要を背景に、法人契約を含めた継続課金モデルが主流である。[Mubert](#)、[SOUNDDRAW](#)、[Beatoven.ai](#) などが代表的で、API 連携による組み込み需要(SaaS／アプリ開発)も拡大している。価格帯は個人向け 10～40 米ドル/月、法人向けは 100～150 米ドル/月以上が多い。③の研究・基盤型モデルは、Google 社の MusicLM や OpenAI の Jukebox など、研究成果や PR 目的で開発・公開された例が多い。これらは商用サービス化されていないが、企業 R&D 部門や大学研究者による PoC(概念実証)や社内利用が進んでいる。近年は Open Source Software としてモデルや学習コードの一部が公開されている。欧米では著作権侵害訴訟(例:Suno, Udio)も進行中³¹のため、注意が必要。

区分	ユースケース	主なサービス例	入力方式	基盤モデル	料金(概算)	対象	契約形態
① 歌もの／楽曲生成	・SNS／配信コンテンツ制作 ・広告ジングルやデモ制作	Suno Udio	テキスト／歌詞／参照音源	Diffusion + Transformer (多モーダル音声生成)	個人：\$10-30/月 上位：グレード \$30-100/月	個人クリエイター／小規模制作会社／広告会社	サブスクリプション契約／商用ライセンス(法人契約あり)
② BGM生成	・動画・アプリ・店舗・ゲーム内 BGM ・長時間	Mubert SOUND DRAW Beatoven.ai	テキスト／キーワード／参照音源(任意)	Token-based LM Latent Diffusion	個人：\$10-40/月 法人：\$100-150/月	個人配信者／映像制作会社／店	サブスク契約／B2B ライセンス／API 連携

	BGM 配信					舗 運 営 法 人	
③ 研究・ 基盤型	・ R&D ・ 社 内 PoC ・ 生成モデ ル 研究 / 技術検証	Google MusicLM OpenAI Jukebox	テキス ト / 生 体信号 (EEG 等)	Hierarchical Transformer VAE Hybrid	市販なし 自前実装・ 計算資源費	研 究 機 関 / 大 学 / 企 業 R&D	OSS 利 用 / 自社環境 構築

Table 1. 生成 AI による音楽生成サービス一覧(用途・コスト感・主対象・契約形態). ※価格は公開情報に基づく概算であり、為替・年払い割引・利用量により変動します。商用利用条件・著作権処理は各社規約に従う必要があります。欧米では著作権侵害訴訟(例: Suno, Udio)も進行中³¹。

おわりに一まとめと展望、そしてリスク

本レポートでは、音楽を時間芸術と捉え、音楽と脳の多時間スケール性の対応関係、さらにそれを拡張する形での音楽生成(自動作曲)AI およびニューロミュージックの可能性について概観した。音楽は、人類史の初期から社会的・感情的コミュニケーションの核として機能しており、脳には音楽の多時間スケールを処理する機構が存在している。この機構を理解することは、より優れた音楽生成 AI の開発においても有意義であると思われる。

現行の Transformer や Diffusion モデルは、環境音などの音響生成は高精度である。一方で、音楽的時間(例 文脈的展開や感情の変化といった長期構造)を十分に考慮した生成は現状できていない。これを補う次世代技術として、Mamba のような選択的状態空間モデル(Selective SSM)や、脳活動を生成条件として組み込む「脳×生成 AI」による音楽生成が注目に値する。

一方で、音楽生成 AI の普及にはいくつかのリスクや課題も存在する。第 1 に、著作権・知的財産権の問題である³¹。AI が既存楽曲を学習して生成した結果が、どの程度「創作性」を持ち、どの範囲まで「引用」や「派生」とみなされるかは、各国で議論が分かれている。第 2 に、個人の心理データや脳波情報が利用されることによるプライバシー侵害や倫理的リスクが挙げられる。特に「脳波に基づく音楽生成」では、個人特性の抽出が可能であるため、データ管理や匿名化の厳格な枠組みが求められる。

参考文献

1. Nguyen, T. *et al.* Sing to me, baby: Infants show neural tracking and rhythmic movements to live and dynamic maternal singing. *Dev Cogn Neurosci* 64, 101313 (2023).
2. MacLarnon, A. M. & Hewitt, G. P. The evolution of human speech: The role of enhanced breathing control. *Am J Phys Anthropol* 109, 341–363 (1999).
3. Conard, N. J., Malina, M. & Münzel, S. C. New flutes document the earliest musical tradition in southwestern Germany. *Nature* 460, 737–740 (2009).

4. Montagu, J. How Music and Instruments Began: A Brief Overview of the Origin and Entire Development of Music, from Its Earliest Stages. *Frontiers in Sociology* 2, (2017).
5. Chanda, M. L. & Levitin, D. J. The neurochemistry of music. *Trends Cogn Sci* 17, 179–193 (2013).
6. Chart: Streaming Changes the Tide for the Global Music Industry | Statista.
https://www.statista.com/chart/4713/global-recordedmusicindustryrevenues/?utm_source=chatgpt.com.
7. *GLOBAL MUSIC REPORT 2023*. https://www.ifpi.org/wp-content/uploads/2020/03/Global_Music_Report_2023_State_of_the_Industry.pdf (2023).
8. Japan Music Market. <https://www.trade.gov/market-intelligence/japan-music-market>.
9. Peters, Z. & Cartwright, P. A Perspective on NFTs in the Arts-and-Music Industry. *International Journal of Music Business Research* 12, 57–77 (2023).
10. Langer, S. K., Scribner, C., Sons, S. & York, N. *FEELING AND FORM A THEORY OF ART Developed from Philosophy in a New Key*. (1929).
11. Buzsáki, G., Logothetis, N. & Singer, W. Scaling brain size, keeping timing: Evolutionary preservation of brain rhythms. *Neuron* 80, 751–764 (2013).
12. Doelling, K. B. & Poeppel, D. Cortical entrainment to music and its modulation by expertise. *Proc Natl Acad Sci U S A* 112, E6233–E6242 (2015).
13. Doelling, K. B., Florencio Assaneo, M., Bevilacqua, D., Pesaran, B. & Poeppel, D. An oscillator model better predicts cortical entrainment to music. *Proc Natl Acad Sci U S A* 116, 10113–10121 (2019).
14. Teng, X., Tian, X., Doelling, K. & Poeppel, D. Theta band oscillations reflect more than entrainment: behavioral and neural evidence demonstrate an active chunking process. *Eur J Neurosci* 48, 2770 (2017).
15. Nozaradan, S. Exploring how musical rhythm entrains brain activity with electroencephalogram frequency-tagging. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369, 20130393 (2014).
16. Shahin, A. J., Roberts, L. E., Chau, W., Trainor, L. J. & Miller, L. M. Music training leads to the development of timbre-specific gamma band activity. *Neuroimage* 41, 113–122 (2008).
17. Teng, X., Larrouy-Maestri, P. & Poeppel, D. Segmenting and Predicting Musical Phrase Structure Exploits Neural Gain Modulation and Phase Precession. *Journal of Neuroscience* 44, (2024).
18. Doelling, K. B., Florencio Assaneo, M., Bevilacqua, D., Pesaran, B. & Poeppel, D. An oscillator model better predicts cortical entrainment to music. *Proc Natl Acad Sci U S A* 116, 10113–10121 (2019).
19. Ohki, T. Measuring phase-amplitude coupling between neural oscillations of different frequencies via the Wasserstein distance. *J Neurosci Methods* 374, 109578 (2022).
20. Ohki, T. & Takei, Y. Neural mechanisms of mental schema: a triplet of delta, low beta/spindle and ripple oscillations. *European Journal of Neuroscience* 48, 2416–2430 (2018).
21. Tseng, Y. L. *et al.* Lingering Sound: Event-Related Phase-Amplitude Coupling and Phase-Locking in Fronto-Temporo-Parietal Functional Networks During Memory Retrieval of Music Melodies. *Front Hum Neurosci* 13, 150 (2019).

22. Wei, L., Yu, Y., Qin, Y. & Zhang, S. From Tools to Creators: A Review on the Development and Application of Artificial Intelligence Music Generation. *Information (Switzerland)* vol. 16 Preprint at <https://doi.org/10.3390/info16080656> (2025).
23. Briot, J.-P. & Pachet, F. Music Generation by Deep Learning – Challenges and Directions. <https://doi.org/10.1007/s00521-018-3813-6> (2018) doi:10.1007/s00521-018-3813-6.
24. Kim, S. *et al.* Enhancing Diffusion-Based Music Generation Performance with LoRA. *Applied Sciences* 2025, Vol. 15, Page 8646 15, 8646 (2025).
25. Gu, A. & Dao, T. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. <http://arxiv.org/abs/2312.00752> (2024).
26. van Pelt, S., Boomsma, D. I. & Fries, P. Magnetoencephalography in twins reveals a strong genetic determination of the peak frequency of visually induced gamma-band synchronization. *Journal of Neuroscience* 32, 3388–3392 (2012).
27. Colafoglio, T. *et al.* NeuralPMG: A Neural Polyphonic Music Generation System Based on Machine Learning Algorithms. *Cognit Comput* 16, 2779–2802 (2024).
28. Sanyal, S., Nag, S., Banerjee, A., Sengupta, R. & Ghosh, D. Music of brain and music on brain: a novel EEG sonification approach. *Cogn Neurodyn* 13, 13–31 (2019).
29. Mirrander, E. R. Brain-Computer music interface for composition and performance. *International Journal on Disability and Human Development* 5, 119–126 (2006).
30. Ran, S. *et al.* Mind to Music: An EEG Signal-Driven Real-Time Emotional Music Generation System. *International Journal of Intelligent Systems* 2024, 9618884 (2024).
31. 音楽生成 AI の Suno と Udio を全米レコード協会が著作権侵害で提訴 – ITmedia NEWS. <https://www.itmedia.co.jp/news/articles/2406/25/news101.html>.