音楽基盤モデルは音高情報を螺旋構造に埋め込むか？

八木 颯斗[1,a)]　高道 慎之介[1,2,b)]

概要：本研究では，音楽基盤モデルの中間表現を解析し，音楽の基本概念である音高が基盤モデルにおいてどのように表現されるかを調査した結果を報告する．学習済み音楽基盤モデルに単音を入力して中間表現を抽出し，その主成分を分析をすることで，音高のオクターブ周期性を反映した螺旋構造を持つことを明らかにする．さらに，基盤モデルとモデルサイズによっては，この螺旋構造が局所的にのみ出現することを示す．本手法は，音楽基盤モデルの内部メカニズム解明への新たなアプローチを提示する．

1. はじめに

自然言語処理や画像処理の分野で大きな成功を収めた基盤モデルは，音楽情報処理にも大きな影響を与え，その応用が急速に進んでいる．特に，大量の音楽データを用いて自己教師あり学習（self-supervised learning; SSL）を行うことで，従来の手法を凌駕する性能を持つ音楽特化モデルが次々と登場し，これらは音楽基盤モデル（music foundation models）と総称されている [1], [2], [3], [4], [5]．音楽基盤モデルは，音楽生成だけでなく，楽曲の分類や分析といった音楽理解においても高い汎用性を示し，新たな創作や体験の可能性を拓くものとして大きな注目を集めている [5], [6].

音楽基盤モデルは高度な生成能力を示す一方で，その内部の計算過程は依然として不透明であり，モデルがどのようにして音楽的知識を獲得し利用しているのかは解明されていない．このブラックボックス問題に対し，モデルが学習した中間表現を分析するアプローチが注目されている．音楽基盤モデルは，人間が作曲した音楽データセットから音楽の構造的・統計的法則を学習する．そのため，モデルの中間表現には，人間が音楽を理解・構成するために用いる音高や和声といった音楽理論に基づく知識が，何らかの構造として潜在的に符号化されている可能性が考えられる．自然言語処理の分野では，大規模言語モデルが周期性を持つ概念を幾何学的構造として表現していることが確認されつつある [7], [8] が，音楽基盤モデルに対する研究では，幾何学的構造に踏み込んだ解析はほとんど行われていないのが現状である．

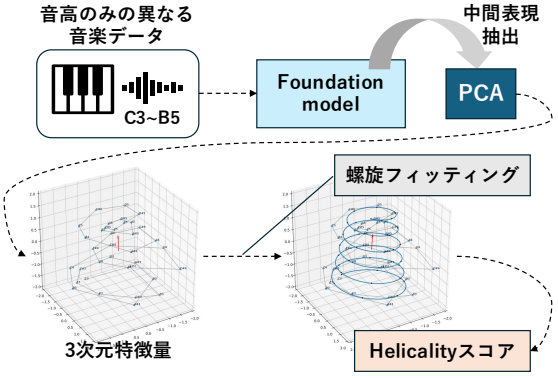

図 1　本研究の概要

本研究では，音楽基盤モデルの内部表現に内在する構造の解明に向けた第一歩として，音楽の最も基本的な要素である音高に着目する．音楽心理学の分野において，音高が持つ高さの連続性とオクターブごとの周期性は，古くから音高の螺旋構造（pitch helix）として知覚的にモデル化されてきた [9]．この人間にとって根源的な音楽構造が，音楽基盤モデルの内部でもデータ駆動的に獲得されているのではないかという仮説に基づき，本研究はモデルの中間表現に音高の螺旋構造が埋め込まれているかを検証する．そのために，音楽基盤モデルから音高に対応する中間表現を抽出しに，螺旋構造らしさを測る指標を用いて定量的に評価する（図 1）．この試みは，モデルの

[1] 慶應義塾大学
Keio University
[2] 東京大学
University of Tokyo
[a)] hayatobuti523@keio.jp
[b)] shinnosuke_takamichi@keio.jp

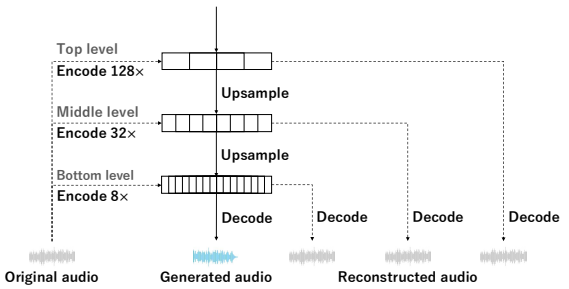

図 2 Jukebox モデル図 [1].

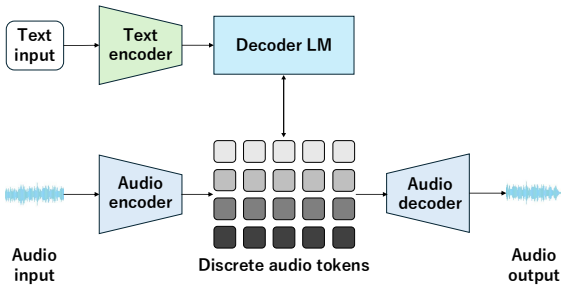

図 3 MusicGen モデル図 [3].

動作原理に関する新たな知見をもたらし，将来的にはその透明性と制御性の向上に貢献することを目指すものである．

2. 関連研究

2.1 音楽基盤モデル

音楽基盤モデルは，言語や画像における基盤モデルと同様に，大規模な音楽データセットを用いて SSL によって事前学習された，汎用性の高い深層学習モデルを指す．このモデルは，単一の特定タスクのために訓練されるのではなく，音楽理解から生成に至るまで，幅広い下流タスクの基盤として機能することを目的としている．代表的な音楽基盤モデルには，音響波形を直接生成する Jukebox [1]，テキスト記述から音楽を生成する MusicLM [2] や MusicGen [3]，音楽理解タスクに特化した MERT [4]，そして音楽の理解と生成の両タスクに応用可能な SONIDO [5] などがある．ここでは，後述する実験的評価で使用する Jukebox と MusicGen についてその構造を詳述する．

- **Jukebox**：Jukebox は，VQ-VAE（vector quantized variational autoencoder）[10] によって音波形を離散的なコード列に符号化し，そのコードを Transformer [11] ベースの階層的なデコーダ言語モデルによって生成するモデルであるこの階層構造は，音楽の全体的な構造などの最も抽象的な情報を捉える Top レベルと，その情報を元に段階的に音の解像度を上げていく Middle 及び Bottom レベルの役割の異なる 3 つのモデルで構成されている（図 2）．
- **MusicGen**：MusicGen は，事前学習済みのニューラルオーディオコーデック（EnCodec [12]），T5 テキストエンコーダ [13]，および Transformer デコーダから構成される，非階層的な生成モデルである（図 3）．

2.2 基盤モデルの内部解析

基盤モデルの高性能さの背景にある中間表現の構造や意味の解明は，特に自然言語処理において近年注目を集めており，モデル内部の表現や計算を人間が理解できるレベルまで抽象化する解析と，それらを言語や世界の知識と結びつけて解釈するアプローチが活発に行われている [7], [8], [14]．例えば，GPT-2 [15] のような言語基盤モデルが，曜日や月といった周期的な概念を，その内部表現空間において円環構造として保持していることが確認されている [7]．このような研究は，モデルが単に統計的なパターンを記憶しているだけでなく，人間が持つ概念構造と類似した形で知識を体系化している可能性を示唆する．

一方，音楽情報処理領域においても同様の試みは進められており，特にプロービング [16] と呼ばれる手法が中心となっている．これは，モデルの各層から抽出した中間表現を，線形分類器のような単純なプローブに入力し，特定の音楽的属性をどの程度予測できるかを評価することで，中間表現が当該属性を符号化しているかを間接的に探るアプローチである．

実際に，音楽基盤モデルの内部解析に関する研究は，複数の方向から行われている．MusicGen や Jukebox といったモデルが，音高，音程，和音，テンポといった基本的な音楽理論の概念をどの程度学習しているかを検証している研究がある [17]．また，MERT や MusicGen を対象に，音高や和音のルート音といった音楽の内容に関する概念が，層が深くなるにつれてより識別的に表現されることが示されている [18]．同様に，SONIDO [5] を用いた研究でも，中間表現に楽器やジャンルといった多様な音楽的特徴の情報が含まれていることが示されている．

しかし，これらの研究は主に，プローブの分類性能を通して中間表現にどのような情報が含まれているかを検証している．そのため，自然言語処理の分野で見られるような，中間表現が形成する幾何学的構造そのものを分析し，音楽理論上の概念がモデル内部でどのように表現されているかに直接的に踏み込んだ研究は，依然として限定的である．

2.3 音高の螺旋構造

音高は，その絶対的な高さを示すピッチハイト（pitch height） [19] と，オクターブを無視した際の音名に対応するピッチクラス（pitch class） [20] という，2 つの知覚的次元からなる構造を持つことが，音楽心理学の分野で広く知られている．この構造は，周波数が 2 倍異なる 2 音を類似した音として知覚するオクターブ等価性（octave equivalence） [21], [22] を反映しており，pitch class が周期的な性質を持つ根拠となっている．この直線的な pitch height と円環的な pitch class の関係性を統合したモデルとして，音高の螺旋構造（pitch helix）が提案されている [9]（図 4）．このモデルでは，音高は 3 次元空間上の螺旋として表現され，その軸方向が pitch height，回転角が pitch class に対応することで，オクターブごとに 1 周して同じ位置に戻るという知覚的特性を幾何学的に捉えている．

近年，この音高の螺旋構造は，ラベルなしの音響データから教師なしに発見できることが報告されている [23]．この手法では，まず音楽信号を定 Q 変換スペクトルに変換し，各周波数帯間のピアソン相関を計算することで類似度グラフを構築する．次に，多様体学習アルゴリズムである Isomap [24] を用いてこのグラフ構造を 3 次元空間に埋め込むことで，螺旋状の構造を可視化する．

しかしながら，このアプローチには，可視化による定性的な評価に留まるという課題がある．この問題を解決するため，3 次元空間上の点群がどの程度理想的な螺旋形状に近いかを定量化する指標として，Helicality が提案された [25]．Helicality は，埋め込まれた点群に対して理想的な螺旋モデルをフィッティングし，その適合度を評価することで，音響データに含まれるオクターブ等価性の強さを一つの数値として捉えることを可能にしている．本研究でも，音高に焦点を当て，音楽信号そのものを分析対象としてきた Helicality の枠組みを音楽基盤モデルの内部解析に応用する．

3. 提案手法

本研究では，音楽基盤モデルの内部表現に，音楽の基本要素である音高が，人間の知覚構造である螺旋構造として埋め込まれているかを検証する．以下に，本研究で用いる手法の詳細を述べる．

3.1 音高に依存する 3 次元特徴量の獲得

本研究では，図 1 に示すように，まず音楽基盤モデルに，ある特定の音高を持つ単一の音楽信号を入力し，各音高に対応する中間表現ベクトルを獲得する．具体的には，先行研究 [17] を参考に，モデル内部の各 Transformer [11] 層から，入力された単音に対応する時系列の中間表現を抽出する．そして，この時系列の中間表現を時間方向に平均化することで，入力された単音全体を代表する単一の中間表現ベクトルを得る．入力には，あらかじめ定めた音高範囲に含まれる N_{key} 個の単一音高データを使用する．この操作を，異なる音高の単音およびモデルの各層に対して繰り返し行うことで，音高と層に関する中間表現ベクトルの集合を構築する．

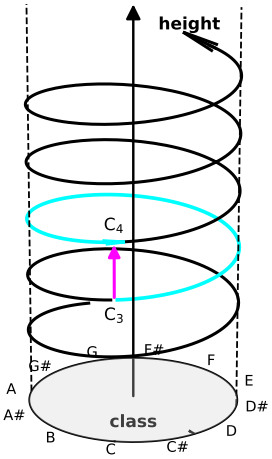

図 4 Shepard の音高螺旋構造 [9].

次に，得られた中間表現ベクトルの集合から，音高の構造的特徴を捉える低次元特徴量を獲得する．具体的には，まず特定の層を一つ固定し，その層から得られた複数音高にわたる中間表現ベクトルの集合に対して，主成分分析（PCA）を適用する．本研究では，単一の 3 次元空間への射影に限定せず，まず第 5 主成分までを算出する．そして，得られた 5 つの主成分から 3 つを選択する全ての組み合わせを考え，各音高に対して ${}_5C_3 = 10$ 組の 3 次元特徴量ベクトルを生成する．これにより，ある特定の層に対して 10 通りの 3 次元特徴量集合が得られる．これらは，その層が保持する音高情報を異なる射影方向から捉えたものであり，後述する螺旋構造らしさの定量的評価に用いられる．

3.2 螺旋構造らしさの定量的評価

3.2.1 音高螺旋モデル

本研究では，音高に関する 3 次元特徴量に対して，パラメトリックな螺旋関数を仮定し，定式化を行う．具体的には，3 次元空間における螺旋構造を表現するため，以下のパラメータを用いて螺旋関数を定義する：

- 初期高さ：h_0
- 高さ変化係数：h_{pitch}
- 初期半径：r_0
- 半径変化係数：r_{slope}
- 角周波数：ω_{chroma}
- 回転の初期位置：t_0

これらを用いて，音高インデックス t に対する螺旋関数 $\boldsymbol{y}(t)$ は次式で与えられる：

$$\boldsymbol{y}(t) = h(t)\cdot\boldsymbol{c} + r(t)\left(\cos\theta(t)\cdot\boldsymbol{u} + \sin\theta(t)\cdot\boldsymbol{v}\right)$$
$$\text{where}\begin{cases} h(t) = h_{\text{pitch}}\cdot t + h_0 \\ r(t) = r_{\text{slope}}\cdot t + r_0 \\ \theta(t) = \omega_{\text{chroma}}\cdot(t - t_0) \end{cases} \quad (1)$$

$\boldsymbol{c}, \boldsymbol{u}, \boldsymbol{v}$ は $\mathbb{R}^3$ の正規直交基底である．$\boldsymbol{c}$ は螺旋の中心軸を表し，$\boldsymbol{u}, \boldsymbol{v}$ によって回転平面が張られる．

$h(t), r(t), \theta(t)$ はそれぞれ，音高インデックス t に依存した高さ，半径，位相であり，t に関する一次式として記述される．$h(t)$ と $\theta(t)$ が t に関して線形に増加することは．図 4 に示す Shepard の音高螺旋構造と整合する．一方で，同図において定数としている半径を，上式では一次式としている．これは，後述する実験的評価において，半径が変化する円錐構造が観察されたためである（4.4.3 節）．各パラメータにおける $\boldsymbol{y}(t)$ の形状の変化については，付録 A.1 を参照されたい．

3 次元特徴量に対しモデルとの 2 乗誤差を最小とするようにパラメータを最適化する．

3.2.2 螺旋構造らしさのスコア

本研究では，螺旋構造と 3 次元特徴量群とのフィッティング度合いを定量評価するために，Helicality スコア [25] を導入する．このスコアは，特徴量と式 (1) との 2 乗平均誤差（mean squared error; MSE）の逆数として定義され，螺旋構造に近いほど高い値を取る．具体的には，3 次元特徴量を $\{\boldsymbol{x}_t\}_{t=1}^{N_{\text{key}}}$ とすると，Helicality スコア は以下の式で定義される：

$$\text{Helicality} = \left(\frac{1}{N_{\text{key}}}\sum_{t=1}^{N_{\text{key}}}\left|\boldsymbol{x}_t - \boldsymbol{y}(t)\right|^2\right)^{-1} \quad (2)$$

3.1 節に述べた ${}_5C_3$ の組み合わせのそれぞれについて，このスコアを計算する．そのなかで最も高いスコアを，当該音楽基盤モデル・当該 Transformer 層のスコアとする．

4. 実験的評価

4.1 実験条件

音楽基盤モデルが音高に関する螺旋構造を有するかを実験的に調査した．

4.1.1 モデル条件

音楽基盤モデルとして学習済みの Jukebox [1] と MusicGen [3] を対象とし，それぞれのデコーダ言語モデルから中間表現を抽出した．Jukebox において，本研究ではこの階層構造全体に着目し，top-level decoder（5B，1B_lyrics），middle-level decoder（1B），bottom-level decoder（1B）の 4 つのモデルすべてを解析対象とする．これらのモデルは，`jukemirlib` [6] を通じて，公式配布元[*1] より取得した．デコーダの層数は，全て 72 層である．中間表現の抽出に際しては，各階層のモデルに音高データを入力し，デコーダ各層から得られる中間表現を時間軸方向に平均プーリングして，それぞれ 4800，1920，1920 の次元を持つベクトルとして得られるようにした．

MusicGen において，本研究では Hugging Face の `transformers` [26] ライブラリを通じて公開されている公式の学習済みモデルを用い，モデルサイズの異なる small（300M）[*2]，medium（1.5B）[*3]，large（3.3B）[*4] の 3 種類を解析対象とした．テキストエンコーダは使用せず，凍結された EnCodec に音楽信号を入力し，デコーダ言語モデルから中間表現を抽出した．デコーダの層数はモデルサイズにより異なり，small は 24 層，medium および large は 48 層である．各層から得られた中間表現を時間軸方向に平均プーリングすることで，最終的にそれぞれから 1024，1536，2048 の次元を持つベクトルとして抽出し，解析に用いた．

4.1.2 データ条件

入力する単一音高データには，先行研究 [17] で提案された合成音楽理論データセット SynTheory を，本研究の目的に合わせて調整したものを使用する．本実験で使用した入力データの作成条件は以下の通りである．

- **基本データセット**：先行研究 [17] で提案された SynTheory データセットのうち，**Notes**（単音）のデータセットを使用した．
- **音楽条件**：
  - **音高**：西洋音楽の平均律における全 12 の pitch class を対象とした．
  - **オクターブ**：C3 から B5 までの 3 オクターブの範囲に限定した（131 Hz から 988 Hz）．

[*1] `https://openaipublic.azureedge.net/jukebox/models/` の末尾に，取得したいファイル名 `5b/vqvae.pth.tar` や `1b_lyrics/prior_level_2.pth.tar` を指定する．
[*2] `https://huggingface.co/facebook/musicgen-small`
[*3] `https://huggingface.co/facebook/musicgen-medium`
[*4] `https://huggingface.co/facebook/musicgen-large`

– **音色**：元データセットに含まれる 92 種類の楽器の中から，`Acoustic Grand Piano` の音源のみを使用した．音源の生成には `TimGM6mb.sf2` [27] サウンドフォントが用いられている．
– **リズム**：テンポ 120 BPM の二分音符で演奏された．これは 1 秒間に 1 音再生されるペースに相当し，強拍・弱拍の区別はない．なお，先行研究 [17] の論文内では四分音符と記載されているが，提供されたコードおよび実際の音高データを確認した結果，実際には二分音符であったため，本稿では観測された仕様に基づいて記述する．

- **オーディオ条件**：
  – **サンプリング周波数**：元のサンプリング周波数は 44.1 kHz である．Jukebox への入力時にはこのまま使用し，MusicGen への入力時には 32 kHz にリサンプリングした．
  – **チャネル数**：元のオーディオはステレオであり，全てのオーディオはモノラルに変換した．
  – **長さ**：MusicGen には 4 秒間の音高データを入力した．Jukebox については，各階層で扱う Context Length（8192 トークン）は共通であるが，1 トークンが表現する時間解像度は top-level から順に 128，32，8 サンプルと異なる．このため，44.1 kHz の音高データにおける入力長を，top-level decoder には約 24 秒，middle-level decoder には約 6 秒，bottom-level decoder には約 1.5 秒となるよう調整した．

4.1.3 フィッティング条件

式 (1) のパラメータの最適化には，ベイズ最適化フレームワークである Optuna [28] を用いた．各パラメータの探索範囲は，予備実験を通じて式 (2) の Helicality スコアが高くなる傾向が見られた領域に基づき，モデル別に手動で設定した．各条件（基盤モデル，モデルサイズなど）において，試行回数 1000 回のパラメータ最適化を，異なる乱数シードを用いて 3 回繰り返し行った．その結果，最も良い Helicality スコアを記録したものを当該条件のスコアとした．この試行回数と繰り返し回数の影響，およびパラメータ探索範囲については，付録 A.2 を参照されたい．

4.2 Jukebox に関する分析

4.2.1 階層に関する分析

まず，Jukebox の各階層の内部表現が音高の螺旋構造を持つかを調査した．具体的には，単一音高データを Jukebox の Top，Middle，Bottom 階層の各デコーダに入力し，それぞれの Transformer 層から中間表現を抽出した．そして，これを主成分分析によって 3 次元に圧縮し，螺旋構造らしさを定量化する Helicality スコアを算出して評価を行った．図 5 はその結果である．

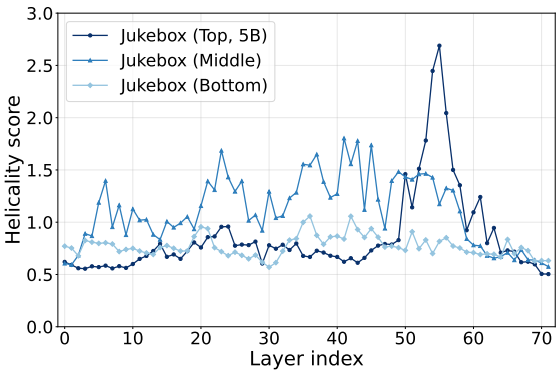

図 5 Jukebox の階層別 Helicality スコアの比較．

表 1 各モデルにおける 1 トークンあたりの受容野

モデル	受容野
Jukebox（Top）	2.9 ms
Jukebox（Middle）	0.73 ms
Jukebox（Bottom）	0.18 ms
MusicGen	20 ms

この図が示すように，Jukebox の階層ごとに音高螺旋構造の表現のされ方に明確な差異が見られる．Top 階層は全体で最大の Helicality スコア 2.69 を記録し，特に第 55 層付近で極めて鋭いピークを形成している．Middle 階層は Top 階層ほどの最大値は見られないものの，平均スコアが 1.04 と 3 つの階層の中で最も高く，多くの層にわたって安定的に螺旋構造を成すことがわかる．対照的に，Bottom 階層のスコアは総じて低く，明確な螺旋構造は観測されない．

階層による差異を分析するために，表 1 に各階層の受容野を示す．Top 階層は，その広い受容野により信号を粗く圧縮し，深い Transformer 層による自己注意を経ることで，音高の階層構造を獲得したと考えられる．一方で Bottom 階層は，その受容野の狭さから音楽波形の微細な再現に重きを置く．そのため，抽象的な階層構造を獲得しないと考えられる．

以上のことから，Jukebox の階層構造は，階層によって音高に関する機能分化を生じさせ，広い受容野の階層ほど階層構造を獲得すると結論付けられる．

4.2.2 モデルサイズに関する分析

次に，モデルサイズの影響を調査するため，Jukebox の Top 階層において，5B モデルと 1B_lyrics モデルの Helicality スコアを比較した．図 6 はその結果である．

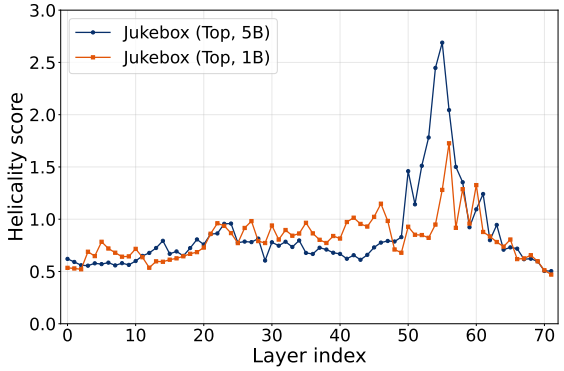

図 6 Jukebox のモデルサイズ別 Helicality スコアの比較．Jukebox (Top, 5B) の線は，図 5 と同じである．

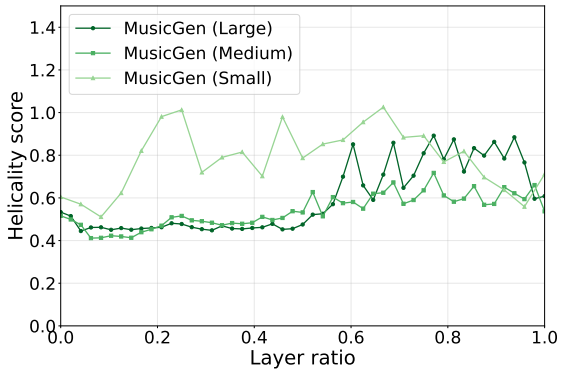

図 7 MusicGen のモデルサイズ別 Helicality スコアの比較．モデルサイズによって層数が異なるため，層インデックスを層総数で正規化した値を横軸としている．縦軸の最大値は，図 5，6 よりも小さいことに注意する．

両モデルとも層が深くなるにつれてスコアが上昇し，第 55 層付近でピークに達するという類似した分布を示している．しかし，そのピーク値には顕著な差が見られ，5B モデルのほうが顕著に高いスコアを示している．この結果は，モデルサイズと機能分化の間に強い関連があることを示唆している．具体的には，モデルサイズが大きいほど，音高幾何に特化する層が生まれることを示唆する．

深い層ほど音高表現を獲得する傾向は，先行研究 [17], [18] のプロービング実験の結果と一致する．これらの研究でも，様々な音楽基盤モデルにおいて深い層ほど音高やルート音といった情報が符号化されることが示されている．本研究で観測された，深い層ほど螺旋構造が明確になるという出現パターンも，この一般的な傾向と同様のパターンを示している．このことは，本研究の発見が単一モデルの特殊な事例ではなく，音楽基盤モデルの普遍的な特性の側面を捉えたものである可能性を示唆している．

4.3 MusicGen に関する分析

続いて，MusicGen においてサイズの異なる Large, Medium, Small モデルの Helicality スコアを比較した．図 7 はその結果である．

いずれのモデルサイズにおいても Helicality スコアは深い層にピークを形成する傾向が見られる．特に，モデルサイズが大きくなるほどピークが明確に出現する傾向がある．

この傾向は，Jukebox において述べた「モデルサイズが大きくなるほど，音高幾何に特化した層が生まれ，その構造が局所化する」という仮説を，アーキテクチャの異なる MusicGen においても支持するものである．また，Jukebox とは対照的に，最もモデルサイズが小さい Small モデルが最も高いピークスコアを示すという興味深い結果が得られている．

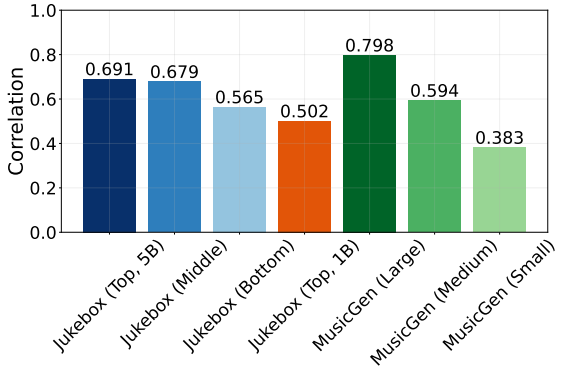

図 8 r_0 と Helicality スコアの相関係数

4.4 パラメータの分析

4.4.1 初期半径 r_0 と Helicality スコアの関係

まず，初期半径 r_0 と Helicality スコアの関係を調べた．具体的には，各モデルについて，全ての層の r_0 と Helicality スコアを用いて両者の相関値を計算した．図 8 はその結果である．相関値の計算にあたる各層のデータについては付録 A.3 を参照されたい．

この図が示すように，ほとんどのモデルにおいて r_0 と Helicality スコアに中程度以上の正の相関がみられる．r_0 が大きいほど特徴量空間において音高同士が互いに離れて配置されることを踏まえると，この一貫した相関は r_0 が音高表現の識別性に貢献し，その識別性と階層構造が相関することを意味すると言える．

4.4.2 角周波数 ω_{chroma} の結果と考察

次に，螺旋の回転速度を決定する角周波数 ω_{chroma} の値について調べた．表 2 は，各モデルで最も高い Helicality スコアを示した層におけるパラメータの一覧である．この表を見ると，ω_{chroma} が $\pi/3 \approx 1.047$ あるいは $\pi/6 \approx 0.524$ に収束する傾向を確認できる．

$\omega_{\text{chroma}} = \pi/6$ が 12 音で 1 周する螺旋であるのに

表 2 各モデルの最高スコア層におけるフィッティングパラメーター一覧

モデル	層	スコア	r_0	$\boldsymbol{c}$	h_{pitch}	r_{slope}	ω_{chroma}	h_0	t_0
Jukebox（Top, 5B）	55/72	2.689	1.771	(0.999, −0.046, −0.028)	0.095	−0.025	1.008	−1.588	0.238
Jukebox（Middle）	41/72	1.803	2.026	(−0.976, 0.180, 0.124)	0.094	−0.050	−1.023	−1.545	−0.103
Jukebox（Bottom）	36/72	1.058	1.147	(1.000, −0.015, 0.007)	0.102	−0.006	1.058	−1.866	−4.313
Jukebox（Top, 1B）	56/72	1.726	1.599	(0.992, 0.039, 0.119)	0.083	−0.025	0.996	−1.515	−3.059
MusicGen（Large）	37/48	0.892	1.401	(−0.982, −0.002, −0.189)	0.088	−0.039	−0.567	−1.557	5.430
MusicGen（Medium）	37/48	0.717	1.424	(0.620, −0.728, −0.293)	0.061	−0.036	0.525	−1.102	−0.367
MusicGen（Small）	16/24	1.025	1.532	(0.998, 0.049, −0.035)	0.088	−0.026	−1.041	−1.482	3.998

表 3 $\boldsymbol{c}$ へ射影した値と音高の順序の相関係数に関する統計量

モデル	平均値	標準偏差	最小値	最大値
Jukebox（Top, 5B）	0.942	0.039	0.757	0.974
Jukebox（Middle）	0.965	0.016	0.900	0.988
Jukebox（Bottom）	0.977	0.010	0.951	0.992
Jukebox（Top, 1B）	0.944	0.079	0.338	0.986
MusicGen（Large）	0.762	0.182	0.241	0.963
MusicGen（Medium）	0.440	0.308	0.013	0.924
MusicGen（Small）	0.923	0.050	0.757	0.961

対し，$\omega_{\text{chroma}} \approx \pi/3$ は 12 音で螺旋が 2 周する構造を意味する．すなわち，$\omega_{\text{chroma}} \approx \pi/3$ の場合は，1 オクターブのちょうど半分であるトライトーンの関係にある 2 音（例：ドとファ♯）は，螺旋上で円の対極に位置するのではなく，円の同じ位置にあることになる．

この特異な構造は，音楽理論におけるトライトーンの特殊な役割をモデルが捉えた結果と解釈できる．トライトーンは，五度圏における対称軸を形成するだけでなく，ジャズ理論などでは代理コード（裏コード）として機能的に等価な役割を果たすことが知られている [29]．モデルが獲得したこの構造は，単なる音響的類似性や音の高さの順序だけでなく，より抽象的な音楽理論を捉えている可能性を示唆している．

4.4.3 半径変化係数 r_{slope} の結果と考察

続いて，螺旋の半径が音高の高さに応じて変化する度合いを示す r_{slope} について分析を行った．表 2 から，Helicality スコアが最も高い層では，全てのモデルにおいて r_{slope} が一貫して負の値を取るという傾向が読み取れる．

$r_{\text{slope}} < 0$ は，螺旋構造が図 4 のような円筒ではなく，高音域になるにつれて半径がわずかに狭まる円錐形であることを示している．この結果は，4.4.1 節の r_0 の分析で示された，大きな半径による識別性の確保という性質が，すべての音域で一様に適用されるわけではないことを示唆している．

4.4.4 中心軸 $\boldsymbol{c}$ と pitch height の関係

最後に，中心軸 $\boldsymbol{c}$ が，音高の高さ，すなわち pitch height をどの程度線形に表現しているかを検証した．具体的には，各層の特徴量を中心軸 $\boldsymbol{c}$ へ射影した値と，実際の音高の順序（pitch order）との間の相関係数を算出した．表 3 は，各モデルの全層について中心軸 $\boldsymbol{c}$ へ射影した値と，実際の音高の順序との相関係数を統計量としてまとめたものである．図 9 はモデル別の相関係数の推移を示している．

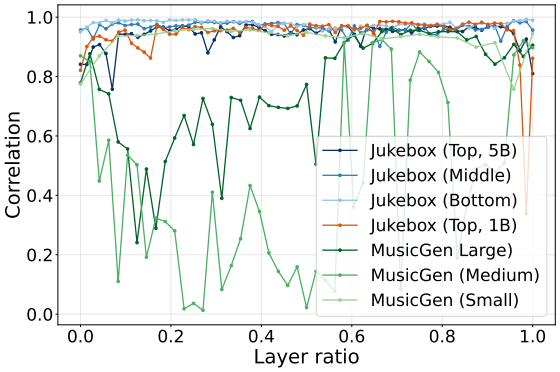

図 9 モデル別の $\boldsymbol{c}$ へ射影した値と音高の順序の相関係数の推移．

表 4 $\boldsymbol{c}$ と PC1 の内積の絶対値に関する統計量

モデル	平均値	標準偏差	最小値	最大値
Jukebox (Top, 5B)	0.968	0.051	0.695	0.999
Jukebox (Middle)	0.979	0.020	0.883	0.999
Jukebox (Bottom)	0.987	0.010	0.962	0.999
Jukebox (Top, 1B)	0.956	0.134	0.035	0.999
MusicGen (Large)	0.906	0.143	0.120	1.000
MusicGen (Medium)	0.330	0.261	0.012	0.996
MusicGen (Small)	0.985	0.021	0.913	1.000

特に Jukebox は，すべての階層において極めて高く安定した相関を示している．さらに，表 3 が示すように，Bottom 階層ほど pitch height との相関が強く，安定している．これは，周波数を忠実に捉える能力が，モデルの時間解像度に直接的に依存していることを示唆している．実際に，図 10 に示すように，高い時間解像度を持つ Bottom 階層では相関係数が $r = 0.992$ という線形関係に達しており，入力された周波数を潜在空間内の線形的な位置として正確にマッピングできていることがわかる．

対照的に，MusicGen の結果は，モデルサイズによって全く異なる振る舞いを示している．表 3 が示すように，Small は Jukebox に匹敵する高く安定し

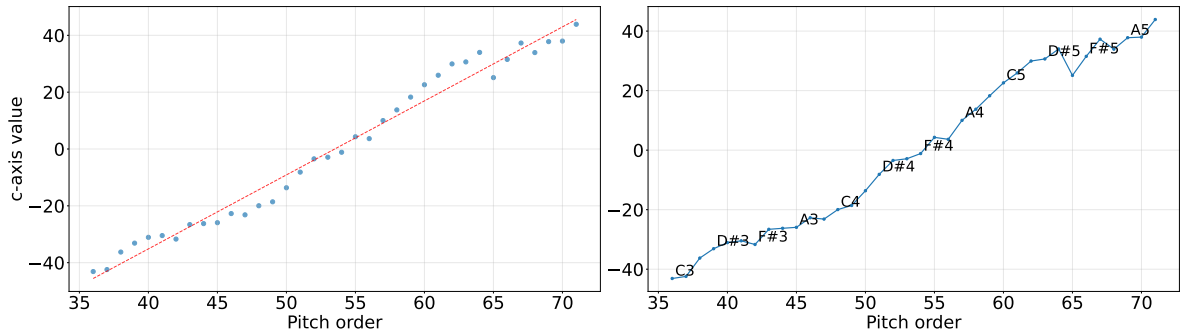

図 10 中心軸 $\boldsymbol{c}$ への射影と音高の関係図．Jukebox bottom の 72 層目の中間表現に対してのデータである．$r = 0.992$ の強い相関が見られる．

た相関を示ししている．これは，限られたモデル容量の中で，音楽の最も基本的な構造である pitch height の線形スケールを優先的に学習した結果であると推察される．一方で，Large および Medium は平均相関が大幅に低い．ただし，最大の相関値は Small と比肩している．これは，4.3 節に示したように，モデルサイズが大きくなるほど機能分化が発生することに対応している．

さらに，この中心軸 $\boldsymbol{c}$ が，第一主成分（PC1）とどの程度一致するかを調査した．表 4 は，各モデルの全層について $\boldsymbol{c}$ と PC1 方向との内積の絶対値を統計量としてまとめたものである．ここでも，表 4 が示すように，Jukebox はすべての階層で高い内積の平均値を示している．これらの結果から，本研究でフィッティングされた螺旋の中心軸 $\boldsymbol{c}$ は，データ内在的な音高の高さという最も主要な軸を正確に捉えたものであると言える．

5. まとめ

本研究では，音楽基盤モデルが音高を，人間の知覚構造と類似した音高の螺旋構造として内部に埋め込むかという問いを検証した．そのために，代表的な音楽基盤モデルである Jukebox と MusicGen から単一音高に対応する中間表現を抽出し，その 3 次元的な幾何構造を可視化するとともに，フィッティングによって螺旋らしさを測る Helicality スコアを用いて定量評価を行った．

実験の結果，特に Jukebox において，音高の幾何学的表現が明瞭に学習されていることが確認された．この構造の明瞭さは，モデルサイズや階層的アーキテクチャに起因すると考えられる．さらに，先行研究 [17], [18] と同様に，音高のような抽象的な概念はモデルの深い層でより明確に表現される傾向が見られた．

しかし，本研究にはいくつかの懸念点も残されている．第一に，分析がピアノ音源と特定の螺旋モデルに限定されている点である．人間の音高知覚は必ずしも単一の螺旋構造ではないという研究 [30] や，楽器の倍音構成が構造形成に影響を与えることが報告されている [23], [25]．そのため，他の楽器の音色や倍音を含まない純音，そして螺旋以外の幾何学的構造も視野に入れた検証が必要であると考える．さらに，本研究で用いたデータセット [17] は合成された単音であり，実際の複雑な楽曲における表現とは異なる可能性がある．

本研究は，言語基盤モデルにおける幾何構造の内部解釈に対応する試みを音楽モデルに応用した初期的な事例である．今後は，本手法をリズムや和声といった他の音楽的概念へ拡張することや，今回特定された幾何構造に対して介入実験を行い，生成音楽の制御性向上に応用することが期待される．

謝辞： 本研究は，JST 創発的研究支援事業 JPMJFR226V，JSPS 科研費 23K28108 の支援を受けて実施した．

参考文献

[1] P. Dhariwal, H. Jun, C. Payne, J. W. Kim, A. Radford, and I. Sutskever, “Jukebox: A generative model for music,” *arXiv preprint arXiv:2005.00341*, 2020.

[2] A. Agostinelli, T. I. Denk, Z. Borsos, J. Engel, M. Verzetti, A. Caillon, Q. Huang, A. Jansen, A. Roberts, M. Tagliasacchi, M. Sharifi, N. Zeghidour, and C. Frank, “MusicLM: Generating music from text,” *arXiv arXiv:2301.11325*, 2023.

[3] J. Copet, F. Kreuk, I. Gat, T. Remez, D. Kant, G. Synnaeve, Y. Adi, and A. Défossez, “Simple and controllable music generation,” in *Advances in Neural Information Processing Systems*, 2023.

[4] Y. Li, R. Yuan, G. Zhang, Y. Ma, X. Chen, H. Yin, C. Xiao, C. Lin, A. Ragni, E. Benetos *et al.*, “MERT: Acoustic music understanding model with large-scale self-supervised training,” *arXiv preprint arXiv:2306.00107*, 2023.

[5] W.-H. Liao, Y. Takida, Y. Ikemiya, Z. Zhong,

C.-H. Lai, G. Fabbro, K. Shimada, K. Toyama, K. W. Cheuk, M. A. Martínez-Ramírez, S. Takahashi, S. Uhlich, T. Akama, W. Choi, Y. Koyama, and Y. Mitsufuji, "Music foundation model as generic booster for music downstream tasks," *TMLR*, 2025. [Online]. Available: https://openreview.net/forum?id=kHl4JzyNzF

[6] R. Castellon, C. Donahue, and P. Liang, "Codified audio language modeling learns useful representations for music information retrieval," in *ISMIR*, 2021.

[7] J. Engels, E. J. Michaud, I. Liao, W. Gurnee, and M. Tegmark, "Not all language model features are one-dimensionally linear," in *ICLR*, 2025.

[8] Z. Liu, O. Kitouni, N. Nolte, E. J. Michaud, M. Tegmark, and M. Williams, "Towards understanding grokking: An effective theory of representation learning," in *NeurIPS*, A. H. Oh, A. Agarwal, D. Belgrave, and K. Cho, Eds., 2022. [Online]. Available: https://openreview.net/forum?id=6at6rB3IZm

[9] R. N. Shepard, "Geometrical approximations to the structure of musical pitch," *Psychological Review*, vol. 89, no. 4, pp. 305–333, 1982.

[10] A. Van Den Oord, O. Vinyals *et al.*, "Neural discrete representation learning," *Advances in neural information processing systems*, vol. 30, 2017.

[11] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, "Attention is all you need," in *NeurIPS*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

[12] A. Défossez, J. Copet, G. Synnaeve, and Y. Adi, "High Fidelity Neural Audio Compression," in *ICLR*, 2024.

[13] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer," *Journal of machine learning research*, vol. 21, no. 140, pp. 1–67, 2020.

[14] B. Heinzerling and K. Inui, "Monotonic representation of numeric attributes in language models," in *Proceedings of the 62nd ACL (Volume 2: Short Papers)*, L.-W. Ku, A. Martins, and V. Srikumar, Eds. Bangkok, Thailand: Association for Computational Linguistics, Aug. 2024, pp. 175–195. [Online]. Available: https://aclanthology.org/2024.acl-short.18/

[15] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever *et al.*, "Language models are unsupervised multitask learners," *OpenAI blog*, vol. 1, no. 8, p. 9, 2019.

[16] G. Alain and Y. Bengio, "Understanding intermediate layers using linear classifier probes," in *ICLR*, 2017.

[17] M. Wei, M. Freeman, C. Donahue, and C. Sun, "Do music generation models encode music theory?" in *ISMIR*, 2024.

[18] W. Ma, X. Li, and G. Xia, "Do music LLMs learn symbolic concepts? a pilot study using probing and intervention," in *Audio Imagination: NeurIPS 2024 Workshop AI-Driven Speech, Music, and Sound Generation*, 2024. [Online]. Available: https://openreview.net/forum?id=uvzw0gS0Nn

[19] American National Standards Institute, *Acoustical Terminology SI.1-1960*, American Standards Association, 1960.

[20] R. N. Shepard, "Circularity in judgments of relative pitch," *The Journal of the Acoustical Society of America*, vol. 36, no. 12, pp. 2346–2353, 1964.

[21] G. d'Arezzo, *Micrologus*, J. Smits van Waesberghe, Ed. Rome: American Institute of Musicology, 1955, an edition of the original 11th-century treatise.

[22] D. Deutsch, "Octave generalization of specific interference effects in memory for tonal pitch," *Perception & Psychophysics*, vol. 13, no. 2, pp. 271–275, 1973.

[23] V. Lostanlen, S. Sridhar, B. McFee, A. Farnsworth, and J. P. Bello, "Learning the helix topology of musical pitch," in *IEEE ICASSP*, 2020, pp. 11–15.

[24] J. B. Tenenbaum, V. d. Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[25] S. Sridhar and V. Lostanlen, "Helicality: An isomap-based measure of octave equivalence in audio data," in *ISMIR*, 2020.

[26] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. von Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. Le Scao, S. Gugger, M. Drame, Q. Lhoest, and A. M. Rush, "Transformers: State-of-the-art natural language processing," in *EMNLP*. Association for Computational Linguistics, 2020, pp. 38–45. [Online]. Available: https://www.aclweb.org/anthology/2020.emnlp-demos.6

[27] T. Brechbill, "Timidity++," https://timbrechbill.com/saxguru/Timidity.php, 2004.

[28] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2019, pp. 2623–2631.

[29] SoundQuest, "トライトーン代理," https://soundquest.jp/quest/chord/chord-mv4/tritone-substitution/, 2025年6月8日.

[30] R. Marjieh, T. L. Griffiths, and N. Jacoby, "Pitch is not (always) a helix: Probing the structure of musical pitch across tasks and experience," *bioRxiv*, 2024. [Online]. Available: https://www.biorxiv.org/content/10.1101/2023.06.13.544763v3

[31] S. Geman, E. Bienenstock, and R. Doursat, "Neural networks and the bias/variance dilemma," *Neural Computation*, vol. 4, pp. 1–58, 01 1992.

付　　録

A.1　パラメータと螺旋モデルの関係

式 (1) で表される音高螺旋モデルにおける各パラメータが螺旋構造の形状に与える影響について，3 次元可視化と 2 次元投影図を用いて説明する．図 A·1, A·2, A·3, A·4 は以下の 3 つの視点から螺旋構造を示している：

- 3 次元表示
- X-Z 平面への投影
- X-Y 平面への投影

A.1.1　初期半径 r_0 の効果

図 A·1 は，初期半径 r_0 を変化させた際の螺旋構造の変化を示している．r_0 は音高インデックス $t = 0$ における螺旋の半径を表すパラメータである．

$r_0 = 1.0$ の場合（上段），螺旋は中心軸に近い位置から開始される小半径の螺旋構造となる．一方，$r_0 = 1.5$ の場合（下段）では，螺旋が中心軸から離れた位置から開始される大半径の螺旋構造となる．

X-Y 投影図からは，r_0 が大きいほど，螺旋の開始位置が中心から離れ，全体的に大きな円形軌道を描くことが観察される．これは，式 (1) における $r(t) = r_{\text{slope}} \cdot t + r_0$ の初期値の影響を反映している．

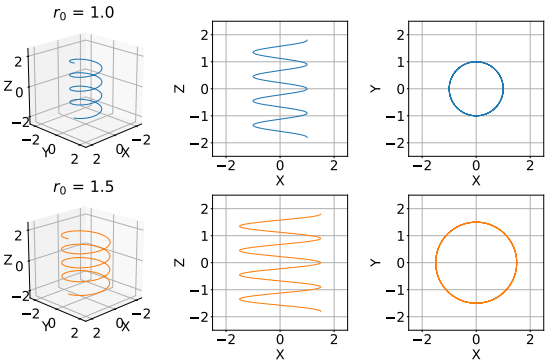

図 A·1　初期半径 r_0 を変化させた際の螺旋構造の可視化．（上段）$r_0 = 1.0$，（下段）$r_0 = 1.5$.

A.1.2　高さ変化係数 h_{pitch} の効果

図 A·2 は，高さ変化係数 h_{pitch} を変化させた際の螺旋構造の変化を示している．h_{pitch} は音高インデックス t に対する高さの変化率を表すパラメータである．

$h_{\text{pitch}} = 0.1$ の場合（上段），螺旋は密に巻かれた構造となり，音高の変化に対して高さの変化が緩やかである．一方，$h_{\text{pitch}} = 0.2$ の場合（下段）では，螺旋が疎に巻かれた構造となり，音高の変化に対して高さが急に変化する．これは，式 (1) における $h(t) = h_{\text{pitch}} \cdot t + h_0$ の線形関係を反映している．

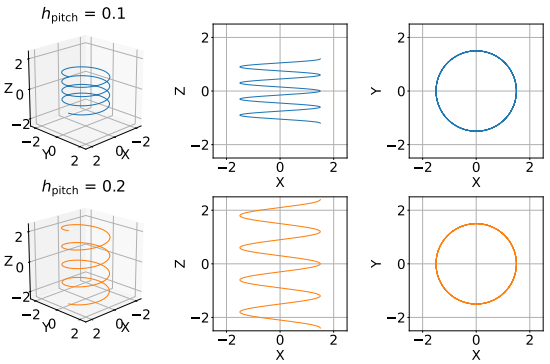

図 A·2　高さ変化係数 h_{pitch} を変化させた際の螺旋構造の可視化．（上段）$h_{\text{pitch}} = 0.1$，（下段）$h_{\text{pitch}} = 0.2$.

A.1.3　半径変化係数 r_{slope} の効果

図 A·3 は，半径変化係数 r_{slope} を変化させた際の螺旋構造の変化を示している．r_{slope} は音高インデックス t に対する半径の変化率を表すパラメータである．

$r_{\text{slope}} = 0.1$ の場合（上段），螺旋の半径が音高の増加に伴って拡大する螺旋構造となる．$r_{\text{slope}} = 0.0$ の場合（中段）では，半径が一定の円柱螺旋構造となる．$r_{\text{slope}} = -0.1$ の場合（下段）では，螺旋の半径が音高の増加に伴って縮小する螺旋構造となる．

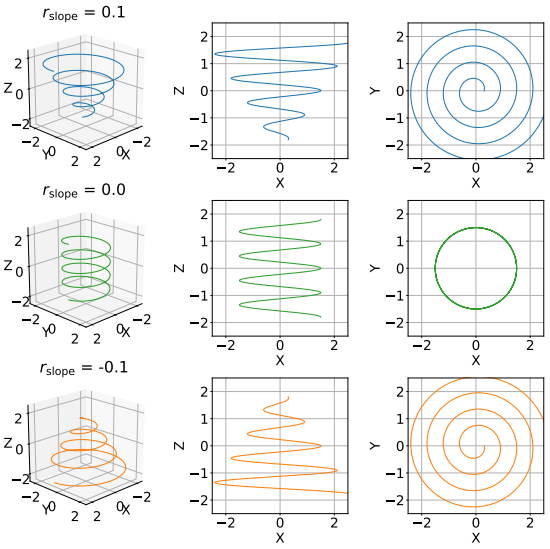

図 A·3　半径変化係数 r_{slope} を変化させた際の螺旋構造の可視化．（上段）$r_{\text{slope}} = 0.1$（中段）$r_{\text{slope}} = 0.0$（下段）$r_{\text{slope}} = -0.1$．r_{slope} の符号と大きさにより，螺旋の半径が音高に応じて拡大，一定，縮小する．

X-Y 投影図からは，各音高における螺旋の断面

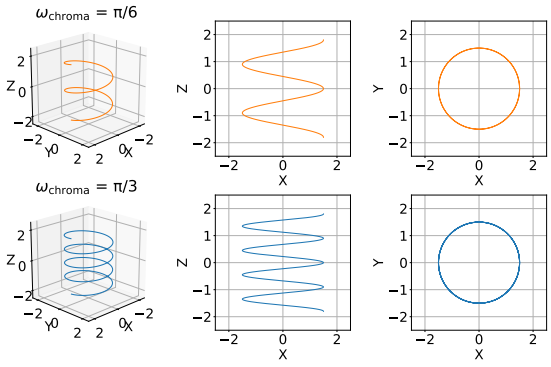

図 A·4 角周波数 ω_{chroma} を変化させた際の螺旋構造の可視化．（上段）$\omega_{\text{chroma}} = \pi/6$（下段）$\omega_{\text{chroma}} = \pi/3$.

が円形を保ちながら，その半径が音高に応じて変化することが観察される．これは，式 (1) における $r(t) = r_{\text{slope}} \cdot t + r_0$ の線形関係を反映している．

A.1.4 角周波数 ω_{chroma} の効果

図 A·4 は，角周波数 ω_{chroma} を変化させた際の螺旋構造の変化を示している．ω_{chroma} は音高インデックス t に対する位相の変化率を表すパラメータである．

$\omega_{\text{chroma}} = \pi/6$ の場合（上段），螺旋は緩やかに回転する転螺旋構造となる．一方，$\omega_{\text{chroma}} = \pi/3$ の場合（下段）では，螺旋が急激に回転する螺旋構造となる．

X-Y 投影図からは，ω_{chroma} が大きいほど，同じ音高範囲において螺旋がより多くの回転を行うことが観察される．これは，式 (1) における $\theta(t) = \omega_{\text{chroma}} \cdot (t - t_0)$ の関係を反映している．

A.2 Optuna による最適化条件の詳細

A.2.1 パラメータ探索範囲と詳細説明

表 A·1 に，各基盤モデルおよびモデルサイズに対して用いたパラメータ探索範囲を示す．

また，表 A·1 に示すパラメータのうち，θ および ϕ は螺旋の中心軸方向を決定する球座標パラメータである．これらのパラメータは，3 次元空間における単位ベクトル $\boldsymbol{c}$ を以下の式で表現する：

$$\boldsymbol{c} = \begin{bmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{bmatrix} \tag{A.1}$$

ここで，$\theta \in [0, \pi]$ は天頂角を表し，$\phi \in [-\pi, \pi]$ は方位角を表す．

A.2.2 最適な試行回数と繰り返し回数の検討

本研究では，Optuna による螺旋パラメータの最適化において，適切な試行回数（Max trials）と繰り返し回数（Repeats）の設定を検討した．これらのハイパーパラメータは，最適化の精度と計算コストのトレードオフを決定する重要な要素である．特に，異なる層における螺旋構造の複雑さに応じて，最適な設定が異なる可能性があるため，複数の層での検証が必要である．

最適化の品質を評価するために，式 (A.2) で定義される合成指標（Performance）[31] を導入した：

$$\text{Performance} = \sqrt{\left(\frac{1}{\text{Mean}}\right)^2 + \text{StdDev}^2} \tag{A.2}$$

この指標は，各組み合わせで 5 回実験行った Helicality スコアの平均値（Mean）の逆数と標準偏差（StdDev）を組み合わせたものであり，値が小さいほど良好な最適化結果を示す．試行回数は 200 から 1600 まで，繰り返し回数は 1 から 5 まで変化させて，各組み合わせにおける合成指標を計算した．

図 A·5 に示すように，3 つの層（Layer8，42，56）において，試行回数と繰り返し回数の増加に伴い合成指標が改善される傾向が観察できる．特に，試行回数 800 以上，繰り返し回数 3 以上では，指標の改善が緩やかになり，収束傾向が見られる．Layer8 では試行回数 1000，繰り返し回数 3 で十分な精度が得られた．Layer42 および 56 ではより多くの試行回数が必要な場合もあるが，全体的に試行回数 1000，繰り返し回数 3 の設定で安定した結果が得られることが確認された．

以上の結果から，本研究では試行回数 1000，繰り返し回数 3 を標準設定として採用した．この設定により，計算コストを抑えながら，十分な精度と安定性を確保できることが示された．また，異なる層においても一貫した設定で良好な結果が得られることから，本研究全体での統一的な最適化パラメータとして適していると判断した．

A.3 r_0 と Helicality スコアの関係

図 A·6 は，各モデルにおける初期半径 r_0 と Helicality スコアの関係を示す散布図である．各点は 1 つの層を表し，7 つのモデルについて分析を行った．

全モデルにおいて r_0 と Helicality スコアの間に正の相関が観察される．Jukebox では比較的強い相関を示し，MusicGen では相関にばらつきが見られる．

表 A·1 各モデルにおけるパラメータの探索範囲

モデル	θ	ϕ	h_{pitch}	r_0	r_{slope}	ω_{chroma}	t_0	h_0
Jukebox（Top, 5B）	$[0, \pi]$	$[-\pi, \pi]$	$[0.05, 0.17]$	$[1.0, 2.0]$	$[-0.03, 0.03]$	$[-\frac{\pi}{2}, -\frac{\pi}{6}] \cup [\frac{\pi}{6}, \frac{\pi}{2}]$	$[-6.0, 6.0]$	$[-2.3, 2.0]$
Jukebox（Middle）	$[0, \pi]$	$[-\pi, \pi]$	$[0.05, 0.17]$	$[1.4, 2.3]$	$[-0.05, 0.05]$	$[-\frac{\pi}{2}, -\frac{\pi}{6}] \cup [\frac{\pi}{6}, \frac{\pi}{2}]$	$[-6.0, 6.0]$	$[-2.5, 2.0]$
Jukebox（Bottom）	$[0, \pi]$	$[-\pi, \pi]$	$[0.05, 0.17]$	$[1.0, 2.0]$	$[-0.03, 0.03]$	$[-\frac{\pi}{2}, -\frac{\pi}{6}] \cup [\frac{\pi}{6}, \frac{\pi}{2}]$	$[-6.0, 6.0]$	$[-2.3, 2.0]$
Jukebox（Top, 1B）	$[0, \pi]$	$[-\pi, \pi]$	$[0.05, 0.17]$	$[1.0, 2.0]$	$[-0.03, 0.03]$	$[-\frac{\pi}{2}, -\frac{\pi}{6}] \cup [\frac{\pi}{6}, \frac{\pi}{2}]$	$[-6.0, 6.0]$	$[-2.3, 2.0]$
MusicGen（Large）	$[0, \pi]$	$[-\pi, \pi]$	$[0.04, 0.16]$	$[0.9, 1.9]$	$[-0.04, 0.04]$	$[-\frac{\pi}{2}, -\frac{\pi}{6}] \cup [\frac{\pi}{6}, \frac{\pi}{2}]$	$[-6.0, 6.0]$	$[-2.7, 2.0]$
MusicGen（Medium）	$[0, \pi]$	$[-\pi, \pi]$	$[0.04, 0.15]$	$[0.9, 2.0]$	$[-0.04, 0.04]$	$[-\frac{\pi}{2}, -\frac{\pi}{6}] \cup [\frac{\pi}{6}, \frac{\pi}{2}]$	$[-6.0, 6.0]$	$[-2.6, 2.0]$
MusicGen（Small）	$[0, \pi]$	$[-\pi, \pi]$	$[0.04, 0.15]$	$[0.9, 2.1]$	$[-0.04, 0.04]$	$[-\frac{\pi}{2}, -\frac{\pi}{6}] \cup [\frac{\pi}{6}, \frac{\pi}{2}]$	$[-6.0, 6.0]$	$[-2.6, 2.0]$

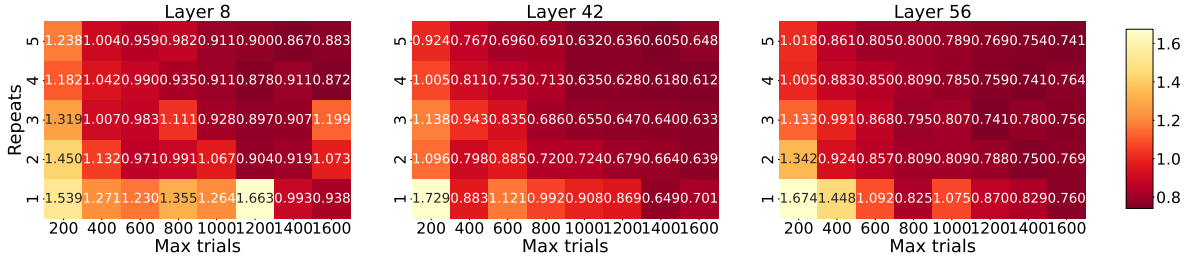

図 A·5 Jukebox Middle 階層における試行回数と繰り返し回数別の合成指標のヒートマップ．濃淡は式 (A.2) で計算される合成指標によって決まる．Helicality スコアの異なる Layer8，42，56 で比較を行った．

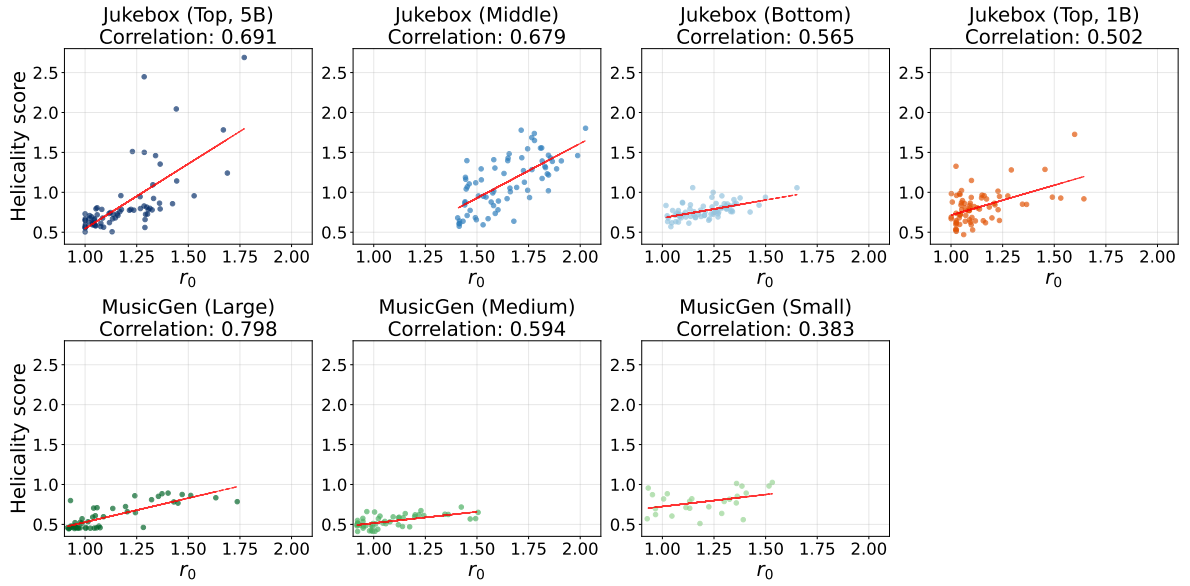

図 A·6 r_0 と Helicality スコアの散布図