

音声生成 LLM を用いたサウンドスケープの進化的探索手法 による生態音響指標の理解

Understanding Ecoacoustic Indices Through Evolutionary Exploration of Soundscapes Using Text-to-Audio LLMs

尾神 圭祐 有田 隆也 鈴木 麗璽

Keisuke Ogami Arita Takaya Reiji Suzuki

名古屋大学
Nagoya University

Abstract: 本研究は、音声生成 LLM と遺伝的アルゴリズムを組み合わせたサウンドスケープの進化的探索手法を提案し、生態音響指標の理解を深めることを目的とする。音声の時間方向の複雑さを抽出する ACI と周波数方向の複雑さを抽出する FACI、および両者の相乗平均を適応度として実験した。結果、ACI に基づく探索では生物音・非生物音を含む多様な自然音を抽出し、FACI は生物音を抽出し、両者の相乗平均は時間的・周波数的に複雑な構造を持つ生物音を抽出する傾向があることが示唆された。

1 はじめに

生態音響学 (Ecoacoustics) は、ある環境における音の総体であるサウンドスケープを基に、動物・自然・人間の関係を明らかにする研究領域である[1]。自然環境のサウンドスケープは、生物が発する音 (Biophony)、風や川の流れといった自然由来の音 (Geophony)、飛行機や道路の騒音など人工的に発せられる音 (Anthrophony) の3つの要素で複雑に構成されている。

自然のサウンドスケープがどのような特徴を持ち、どのような音源から構成されるかを定量的に分析するため、生態音響指標を用いた評価が行われている。生態音響指標は、サウンドスケープにおける周波数、時間、エネルギーの様々な特徴や変化を捉えるために広く利用されており、R の Seewave[2]や Python の scikit-maad[3]などのパッケージを利用して容易に計算できる。

生態音響指標には、Acoustic Complexity Index (ACI) (音響複雑性指数) という、主に鳥類の発声を定量化することを目的として考案された指標[4]や、Acoustic Diversity Index (ADI) (音響多様性指数) という、サウンドスケープの複雑さや多様性を定量的に評価する指標[5]などがある。近年、安価で長時間連続運用が可能な自動録音装置 (ARU) の普及により様々な場所で録音を大量に取得可能になったため、生態音響指標に基づくサウンドスケープの特徴やダ

イナミクスの抽出は重要さを増しているといえる。

本研究は、その代表例である ACI に注目する。ACI は、鳥のさえずりをはじめとする多くの生物音が短時間にわたり多様な強度を示す一方で、人工的な音は比較的一定の強度を持つという経験的な観察に基づき開発された[4]。その最も単純な計算手法では、スペクトログラムの単一の周波数ビン内で隣接する2つの時間フレーム間のエネルギーの絶対差を計算し、それを全時間フレームおよび周波数帯域で合計して求められる。ACI はエネルギーの変動性や複雑さを定量化し、生物多様性の代理指標として利用されてきた。しかし、必ずしも生物多様性のみを反映しないことも示唆されている。ACI は嵐や植物から落ちる雨粒の音、昆虫の鳴き声に対しても高い値を示す場合があることも指摘されている[6]。ACI と鳥類の多様性との関係は、研究対象や環境によって一貫性を欠くことが報告されており、その生態学的解釈の難しさは課題の一つとなっている[7]。

同時に、ACI は提案者である Farina らにより複数の発展版も提案されている[8]。例えば、当初の時間軸方向の複雑性を評価する ACI_t に加え、周波数軸方向の複雑性を評価する ACI_f 、および時間軸方向の音響強度の均等性を示す ACI_{te} 等がある。これらを組み合わせることで、複雑なサウンドスケープに含まれる構成要素 (野鳥の鳴声などの生物音、風や雨などの地理的物理音、航空機や交通音などの人工技術音) を自動的に検出可能にしている。例えば、 ACI_f と

ACIF_{te}からなるイベント空間上の位置に応じて、遠く
の合唱、豪雨、コマドリの地鳴きといった異なるイ
ベントを識別できる。最近では、これら3指標の強
度を組み合わせた音響コード分類手法も提案されて
おり[9]、ACIは生物音から地球物理音、技術音ま
でを包括するサウンドスケープ解析の重要なツール
としての役割が期待されている。

音声生成 AI (Text-to-Audio AI) は近年急速に発展
しており、深層学習技術の登場により、より複雑で
多様な音声を生成することが可能になった[10]。近
年の大規模言語モデル (Large Language Model, LLM)
の進展は、より柔軟で高度な音声生成を可能にし、
新たな可能性を開拓している。例えば、Stable
Audio[11]は、プロンプトを基に音楽制作や効果音生
成など多様な用途に対応できるモデルとして期待さ
れている。Stable Audioには、研究目的で利用が可能
なオープンソース版である、Stable Audio Open が存
在する。

大規模言語モデルは近年進化計算にも様々に応用
されている。Meyersonらは、LLMを用いて進化計算
における、Language Model Crossover (LMX)という
新しい手法を提案した[12]。この手法で、LLMによ
る交叉が、方程式、文、画像、Pythonコードの進化
で多様な解を生成できることが示された。Suzuki and
Aritaは、LLMを活用した協力行動に関する性格特
性の進化モデルを提案し、多様で複雑な性格特性の
表現に基づいた協力行動が進化することを示した
[13]。

上記で述べたように、生態音響指標はサウンドス
ケープの分析に広く用いられているものの、実デー
タには様々な要因や環境の特徴が含まれるため、そ
の値と、生物音および非生物音との対応関係の理解
は十分でない。その理由の一つとして、高いACIを
もたらすサウンドスケープにどのようなものがあり
うるかを十分把握しきれていないことが考えられる。
例えば、前述のACIF_tは定義としては音響の周波数方
向の複雑さを定量化するが、高い指標値をもたらす
自然音がどんなものであるかは必ずしも自明でない。
音声生成 AI が生み出す自然であり多様な音声を対
象にしてこれを検討することは、指標の生態的意義
づけの明確化に貢献しうると考えられる。

本研究は、音声生成 AI の豊かな表現力と、進化計
算の持つ任意の遺伝的記述を柔軟に探索可能な能力
を活用して、ありうるサウンドスケープを探索する
手法を構築し、生態音響指標の理解や新しい活用へ
の貢献の可能性を検討することを目的とする。具体
的には、音声生成 LLM と LLM を用いた進化計算を
利用し、特定の生態音響指標を最大化する人工的
なサウンドスケープのテキストによる記述を進化的に

探索する。特に、生存可能な個体の適応度に上限・
下限を設け、生成時に生じる突飛な個体の独占を抑
えることで探索を効率化する。分析では、探索過程
で生じたサウンドスケープを、その記述と生成音声
を特徴空間上の分布から分析し、対象とする指標が
反映するサウンドスケープの特徴を明らかにする。

本稿では、ACIに加え、ACIF_tに類する周波数軸方
向の変化を定量化するFACIを音響指標として定義、
採用し、両者、および、両者の相乗平均を適応度と
した実験を行い、提案手法の意義と可能性について
検討する。

2 音声生成 LLM を用いたサウンド スケープの進化的探索手法

本研究で構築した進化的探索手法のフローチャー
トを図1に示す。以下、大規模言語モデルを用いた
遺伝子作成や、音声生成 LLM を用いた遺伝子から
サウンドスケープの生成、生態音響指標による評価、
進化操作について説明する。

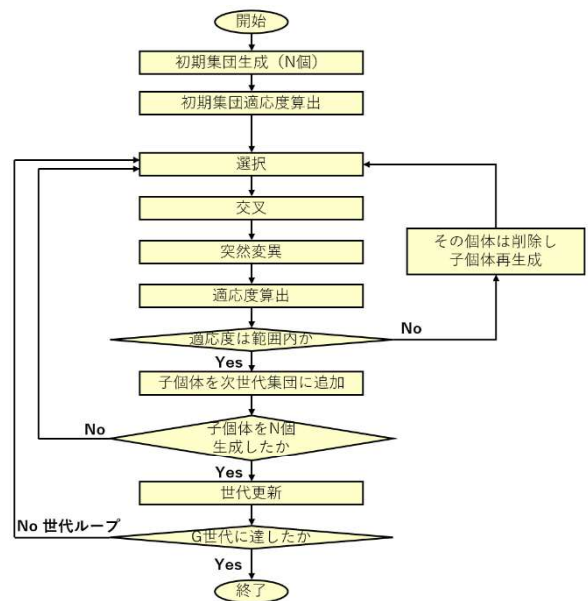


図1：進化的探索手法のフローチャート

2.1 初期集団の生成

大規模言語モデルの Llama-3.1 を使用する。初期集
団の生成は、Llama に対して以下の文章を入力し、
集団サイズ N 個分の「朝の森のサウンドスケープ」
を描写するプロンプトを生成した。ここで、 N は生
成するプロンプトの数を示す。

Create {N} one-sentence according to the theme presented below, following the rules.

Theme: "Soundscapes flowing through the forest in the morning"

Rules:

1. Create {N} descriptions.
2. There is no numbering and separate the sentences with slashes (/).
3. Each sentence must end with a period.

以下は実際に生成されたプロンプトの一部である.

The soft chirping of birds fills the air, a gentle melody.
The rustling of leaves beneath the gentle breeze is a soothing serenade.
The gentle babbling of a brook creates a peaceful ambiance.

このように、初期集団のプロンプトには、鳥のさえずりや葉のざわめき、小川のせせらぎなど、森に生息する生物や自然音について記された短い文を採用した。

2.2 プロンプトからサウンドスケープへの変換

プロンプトからサウンドスケープを表す遺伝子の表現型としての音声を生成するために、オープンソースの音声生成 AI である Stable Audio Open 1.0 を使用した[11]. このモデルは、テキストプロンプトに基づいて、最大 47 秒のステレオ音声を 44.1kHz のサンプリングレートで生成でき、テキストプロンプトから柔軟で高度な音声を生成する能力を有している。図 2 は、実際に「The soft chirping of birds fills the air, a gentle melody.」というプロンプトから生成された結果である。このサウンドスケープは、プロンプトの記述通り、優しく穏やかな数羽の鳥のさえずりで構成されていた。

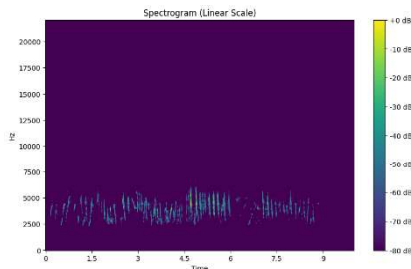


図 2 : 実際に生成されたサウンドスケープ
プロンプト : The soft chirping of birds fills the air, a gentle melody.

2.3 生態音響指標に基づく適応度評価

音声で表される表現型の適応度として、本研究では ACI と、周波数軸に関する強度の変化量を定量化する FACI の 2 種類の生態音響指標を採用する。なお、今回の ACI の定義はこれが初めて提案された文献[4]での定義に基づき、FACI の定義もこれを拡張したものを採用した。ACI と FACI の計算の概要を示す図を図 3 に示す。

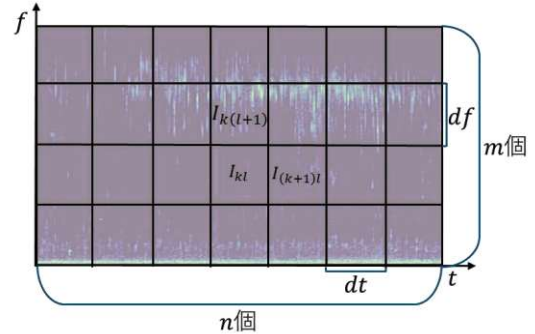


図 3 : ACI と FACI の計算の概要図

まず、音の時間方向の複雑性を定量的に評価できる ACI (音響複雑性指数) [4]を用いて、生物音の抽出を期待した。ACI は次のように計算される。音声ファイルに対して短時間フーリエ変換を適用し、その時間幅 dt と周波数幅 df を用いる。周波数帯域 l において、隣接する二つのタイムビンのエネルギーの絶対差 d_{kl} を求める。

$$d_{kl} = |I_{kl} - I_{(k+1)l}| \quad (1)$$

ここで I_{kl} は k 番目のタイムビン、 l 番目の周波数帯域のエネルギーを意味する。次に、その周波数帯域の d_{kl} の合計 D_l を求める。

$$D_l = \sum_{k=1}^n d_{kl} \quad (2)$$

D_l を、その周波数帯域内のエネルギー値の合計で割って正規化した値を ACI_l とする。

$$ACI_l = \frac{D_l}{\sum_{k=1}^n I_{kl}} \quad (3)$$

そして、全周波数帯域の ACI_l を合計した値を最終的な ACI とする。

$$ACI = \sum_{l=1}^m ACI_l \quad (4)$$

また、Frequency-based Acoustic Complexity Index (FACI) (周波数に基づく音響複雑性指数) も適応度のための指標として採用する。鳥類のさえずりや昆虫の鳴き声などの生物音は、複雑な倍音構造と周波

数変動を持ち、特定の周波数帯域にエネルギーを集中させるのに対し、非生物音は広範な周波数帯域に連続的にエネルギーが分布する。今回採用する FACI は、このような周波数方向の複雑性と帯域特性の違いに着目することで、生物音源をより効果的に検出できる指標として機能することを期待して、前述の ACI の定義を次のように改変して定義する。まず、音声ファイルに対して短時間フーリエ変換を適用する。そのスペクトログラムについて、時間幅 dt と周波数幅 df を与え、その幅で n 個のタイムビン、 m 個の周波数帯域に再分割する。タイムビン内において、隣接する二つの周波数帯域のエネルギーの絶対差 d'_{kl} を求める。

$$d'_{kl} = |I_{kl} - I_{k(l+1)}| \quad (5)$$

I_{kl} は、 k 番目のタイムビン、 l 番目の周波数帯域のエネルギーを意味する。そして、そのタイムビン内の d'_{kl} の合計 D_k を求める。

$$D_k = \sum_{l=1}^m d'_{kl} \quad (6)$$

D_k を、そのタイムビン内のエネルギー値の合計で割って正規化した値を $FACI_k$ とする。

$$FACI_k = \frac{D_k}{\sum_{l=1}^m I_{kl}} \quad (7)$$

そして、全タイムビンの $FACI_k$ を合計した値を $FACI$ とする。

$$FACI = \sum_{k=1}^n FACI_k \quad (8)$$

FACI は、周波数軸の振幅変動の複雑さを評価するため、複数の生物種が異なる周波数帯域で鳴いている、すなわち生物多様性の反映を期待する。

ACI と FACI を適応度として用いるにあたり、本研究では予め設定した最小値 ($ACI_{min}, FACI_{min}$) および最大値 ($ACI_{max}, FACI_{max}$) を用いて Min-Max 正規化を行う。音声生成 LLM を用いた探索において指標値を無制限に最大化すると、過剰に適合した不自然なノイズが生じる恐れがあるため、自然音として妥当な範囲を上限值として設けた。また、値の範囲が異なる各指標を同一のスケールに統一することで、最適化における偏りを防ぎ、対等な評価を可能にした。正規化された値である $ACI_{norm}, FACI_{norm}$ は以下の式で定義される。

$$ACI_{norm} = \frac{ACI - ACI_{min}}{ACI_{max} - ACI_{min}} \quad (9)$$

$$FACI_{norm} = \frac{FACI - FACI_{min}}{FACI_{max} - FACI_{min}} \quad (10)$$

本研究では、これらの正規化された値 ($ACI_{norm}, FACI_{norm}$) の相乗平均 GA_{AF} を以下の式で計算し、も

う一つの適応度として採用した。これにより、時間変動のみが激しい音や単調な持続音といった片方の指標のみが高い個体の評価を低く抑え、生物音源特有の「時間的な複雑性」と、「特定の周波数帯域へのエネルギー集中」の双方を兼ね備えた個体が高い適応度を得られるよう設計した。

$$GA_{AF} = \sqrt{ACI_{norm} \times FACI_{norm}} \quad (11)$$

2.4 大規模言語モデルを用いた進化操作

集団から次世代集団を生成する際には選択と交叉、突然変異を用いる。なお、図 1 において世代内での操作に適応度評価が最初に存在しないのは、新規個体生成時に生存可能個体の判断の際に適応度を算出するため、計算の必要がないためである。

選択にはルーレット選択を用いた。[0, 1] の範囲に正規化された適応度を用いてルーレット選択を行い、親個体 2 体を選択した。これらをそれぞれ親 1 と親 2 とする。

交叉・突然変異には大規模言語モデル「Llama-3.1」を用いた。交叉は、交叉確率 p_c で、Llama に対して以下の文章をプロンプト入力し、選択された親個体の遺伝子（サウンドスケープの言語記述）を組み合わせて、新たな子個体の遺伝子を生成するように指示した。プロンプト内の {parent1} および {parent2} には、親 1 および親 2 の遺伝子が入力される。もし親 1 と親 2 の遺伝子が完全に一致する場合、交叉は行わず、親 1 のプロンプトをそのまま子個体の遺伝子とする。交叉が適用されない場合、次の突然変異に進む。

Create a new prompt by combining the following two original prompts according to the following rules:

1. The new prompt must be a single, coherent, and concise sentence.
2. The new prompt must be clear and to the point.
3. The new prompt must be 30 words or less.

Original prompt 1: "{parent1}"

Original prompt 2: "{parent2}"

New prompt:

以下は、実際に行われた交叉の一つである。

親 1: A gentle hooting of an owl signals the start of a new day.

親 2: The soft chirping of a bird in flight creates a joyful melody.

子: A gentle hooting of an owl signals the start of a new day as a bird takes flight creating a joyful melody.

突然変異は、突然変異率 p_m で Llama に対して以下の文章を入力し、選択された親 1 の遺伝子（プロンプト）を書き換え、新たな子個体の遺伝子（プロンプト）を生成するよう指示した。なお、プロンプト内の {original_prompt} には、親 1 のプロンプトが入力

される。突然変異が適用されない場合、子個体のプロンプトは親 1 のプロンプトと同一となる。

Rewrite the original prompt into a new prompt according to the following rules:

1. The new prompt must keep the original meaning and context.
2. Paraphrase or incorporate a few creative variations.
3. The new prompt must be 30 words or less.

Original prompt: "{original_prompt}"

以下は実際に行われた突然変異の一つである。

元のプロンプト: The soft chirping of a bird in flight creates a joyful melody.
 突然変異後のプロンプト: Birdsong soars on gentle wings, filling the air with a sweet serenade of chirps and trills in perfect harmony.

本研究では、各指標 (ACI, FACI) の値がそれぞれの最大値および最小値 (ACI_{max} , ACI_{min} , $FACI_{max}$, $FACI_{min}$) の範囲外となるサウンドスケープは子個体として採用せず、新たな個体を生成しなおした。

また、計算コストを削減するため、同一のプロンプトに対応するサウンドスケープの生成を R 回まで再利用する手法を採用した。具体的には、全く同一のプロンプトが与えられた場合、 R 回まで再利用を行い、すでに R 回再利用されている場合は新たにサウンドスケープを生成する。

適応度評価と次世代集団の生成を G 世代繰り返すことにより、集団を進化させた。

3 異なる適応度評価指標に基づく

進化的探索の比較

本実験では、遺伝的アルゴリズム内の適応度評価において、様々な音響指標を適用し、比較分析を行う。

3.1 実験設定

表 1 に示すパラメータ設定を用いて、ACI, FACI, 両者の相乗平均を適応度とした条件で実験を行った。FACI の計算において、生物音源がもつ大局的な周波数構造を捉えるため、粗い解像度である $df = 1000\text{Hz}$ を採用した。一般に、風や雨などの非生物音は、広範な周波数帯域に渡ってエネルギーが連続的に分布しているが、鳥類などの生物音は種ごとに特定の周波数帯域にエネルギーを集中させる特性を持つ。したがって、ある程度広い帯域幅でエネルギーを比較し、特定の帯域に突出する生物音源の存在を効果的に検出することを期待する。

表 1 : 実験のパラメータ

1 世代あたりの個体数	$N = 20$
世代数	$G = 50$
交叉確率	$p_c = 0.2$
突然変異率	$p_m = 0.1$
サウンドスケープの最大再利用回数	$R = 2$
各条件での試行回数	10 回
ACI の最小値	$ACI_{min} = 250$
ACI の最大値	$ACI_{max} = 500$
FACI の計算における時間の分割幅	$ds = 0.1s$
FACI の計算における周波数の分割幅	$df = 1000\text{Hz}$
FACI の最小値	$FACI_{min} = 0.0$
FACI の最大値	$FACI_{max} = 200.0$

3.2 ACI を適応度とした場合

ACI を適応度として用いた場合の最大・平均・最小適応度の推移を図 4 (a) に示す。各値は全試行の平均であり、周囲のシェーディング領域は 10 試行間の標準偏差を示している。同図から、集団平均が世代の進行に伴い上昇しており、世代後半では安定して高い値を保っていることがわかる。これは、本手法が ACI 値に基づく適応度の高いサウンドスケープの生成を達成し、進化的探索手法として有効に機能していることを示している。

図 4 (b) は、全 10 試行の中から、収束した音響特性の傾向が異なる代表的な 4 試行を抽出し、その最終世代における代表個体のプロンプトとスペクトログラムを示したものである。なお、各試行の代表個体としては、適応度が中央値となる個体を選出した。これは、その世代の集団が獲得した音響的特徴を最も標準的に保持しているためである。生成された最終世代のサウンドスケープは、小川のせせらぎ (試行 2)、鳥のさえずり (試行 7, 9)、こおろぎの鳴き声 (試行 7)、枝の折れる音 (試行 10) など、生物音と非生物音が混在する多様な音源で構成されていた。スペクトログラムを確認すると、これらの音源はいずれも断続的で不規則なエネルギー変動を示しており、ACI の最大化という目的を達成されていることがわかる。例えば、小川のせせらぎは水流の不規則な変化として、鳥のさえずりやこおろぎの鳴き声は短時間のチャープとして、枝が折れる音は瞬間的なパルス音として現れている。

これらの結果は、ACI が鳥のさえずりだけでなく、時間的に複雑な非生物音に対しても反応するという

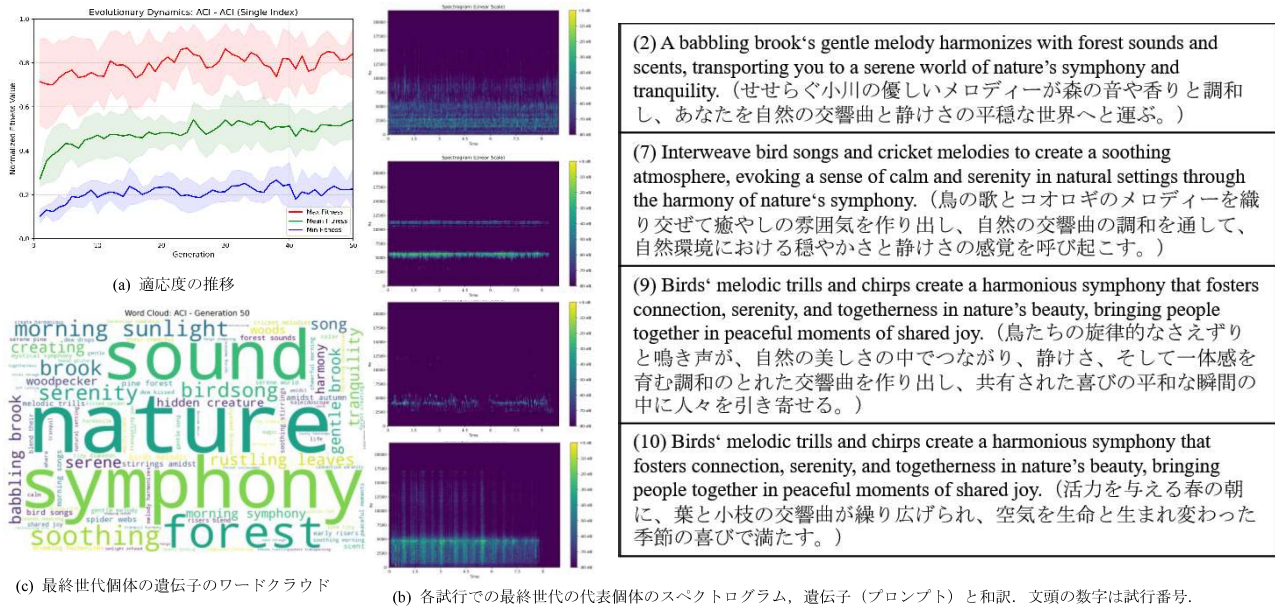


図4：ACIを適応度とした場合の実験結果

従来の知見を、進化的探索によって再現・実証したことを意味する。ACIはその計算ロジック上、時間方向の変動性を評価するため、結果として広範なサウンドスケープを生成するに至ったと考えられる。

この傾向は、図4(c)に示す最終世代個体の遺伝子（プロンプト）のワードクラウドからも見て取れる。出現頻度の高い単語には、「bird」や「birdsong」といった生物に関する語句だけでなく、「brook」や「leaves」といった自然環境や非生物に関する語句も多く含まれている。このことから、ACIの最大化は、特定の生物群に特化するのではなく、時間的に複雑なあらゆる自然音の生成を促進したことが示唆される。

3.3 FACIを適応度とした場合

次に、図4と同様にして、FACIを適応度として用いた場合の結果を図5に示す。図5(a)より、ACIの場合と同様に、集団平均が世代の進行に伴い上昇しており、世代後半では安定した高水準で収束している。この結果は、本手法がFACI値に基づく適応度の高いサウンドスケープの生成に成功し、進化的探索手法として有効に機能していることを示している。

図5(b)より、生成された最終世代のサウンドスケープは、鳥のさえずりや夜の動物の鳴き声（試行2）、朝の鳥の合唱（試行8）、鳥とこおろぎの鳴き声（試行10）といった生物音に特化しており、ACI最大化条件で多く見られた小川のせせらぎなどの非生物音は含まれていなかった。

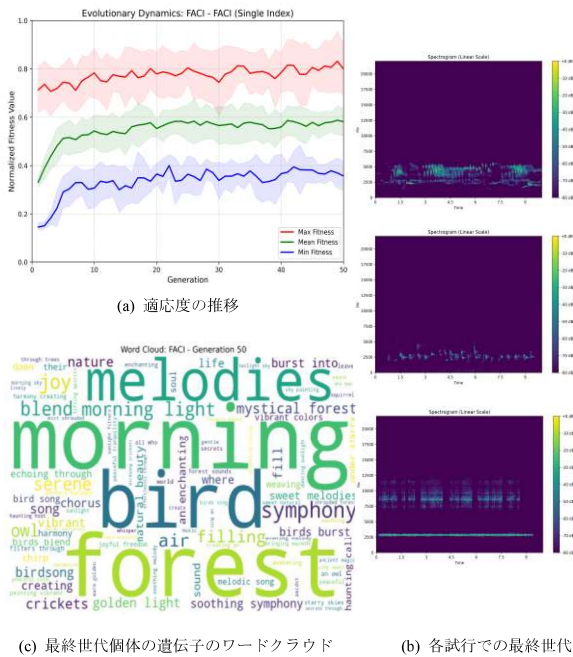
スペクトログラムを確認すると、ACIの解が全帯

域にエネルギーが広がる傾向にあったこととは異なり、FACIの解はエネルギーが特定の周波数帯域に集中し、周波数軸方向に対して顕著な不均一性を示していることが分かる。これは、FACIの計算において1000Hzという帯域幅でスペクトル構造を評価した設定が、広帯域に連続して分布する非生物音を抑制し、生物音源特有の特定の帯域へのエネルギー集中を効率的に抽出するフィルターとして機能したことを示唆している。つまり、FACIが周波数方向の構造に着目することで、時間方向の変動というACIと共通の特性を持つ非生物音を効果的に排除し、生物音源を選択的に抽出したことを裏付ける結果である。

この傾向は、図5(c)に示す最終世代の個体のプロンプトのワードクラウドにも表れている。「bird」や「melodies」と言った単語が圧倒的多数を占め、「crickets」といった生物名も見られる一方で、ACI条件で頻出した「brook」や「leaves」といった具体的な非生物音源を示す単語はほとんど見られない。このことから、FACIの最大化は、特定の生物群（特に鳥類やこおろぎ）の生成に特化するよう選択圧を与えたことが示唆される。

3.4 ACIとFACIの相乗平均を適応度とした場合

最後に、ACIとFACIの相乗平均を適応度とした場合の結果を図6に示す。図6(a)より、単一指標の条件と比較して集団平均の上昇速度が最も緩やかであることが読み取れる。これは、性質の異なる2つ



(2) Under starry skies, nocturnal creatures and morning birds blend in a mystical dance with nature's nocturnal serenade and morning's sweet melodies, weaving a spell of enchantment with leaves and shadows. (星空の下、夜行性の生き物と朝の鳥たちが、自然の夜のセレナーデと朝の甘いメロディーと共に神秘的なダンスで混ざり合い、葉と影で魅了する魔法を織り成す。)

(8) Awaken to a symphony of morning bird songs and golden light, where sunlight harmonizes and dances with the forest, infusing the world with serenity, joy, and vibrant peace. (朝の鳥の歌と黄金の光の交響曲で目覚めます。そこでは日光が森と調和して踊り、静けさ、喜び、そして活気に満ちた平和を世界に吹き込む。)

(10) Birds and insects blend melodies in harmony, cricketsong and forest sounds create a soothing symphony, echoing through the mist-shrouded forest, growing louder with golden light, as morning breaks. (鳥と昆虫が調和の中でメロディーを混ぜ合わせ、コオロギの歌と森の音が癒やしの交響曲を作り出し、朝が明けるにつれて黄金の光と共に大きくなりながら、霧に包まれた森に響き渡る。)

(a) 適応度の推移 (b) 各試行での最終世代の代表個体のスペクトログラム、遺伝子（プロンプト）と和訳。文頭の数字は試行番号。 (c) 最終世代個体の遺伝子のワードクラウド

図5：FACIを適応度とした場合の実験結果

の音響特性を同時に最大化する過程が探索の難度を高めていることを示唆するものの、最終的には高い適応度へ到達しており、本手法が指標の組み合わせのような複雑な評価設定においても有効に機能していることを示している。加えて、全条件において、世代が進行しても最大適応度と最低適応度の間に一定の隔たりが維持されている。これは、集団内の解の多様性が保たれていることを示しており、進化的探索が単一の局所解に過度に集中することなく、複数の試行において安定した探索を行っていることを裏付けている。

図6(b)より、生成された最終世代のサウンドスケープは、鳥のさえずりと森の情景（試行4）、鳥の合唱（試行5）、すずめの鳴き声（試行8）、こおろぎの鳴き声（試行9）といった生物音に特化しており、ACI最大化条件で多く見られた小川のせせらぎなどの非生物音は含まれていなかった。

スペクトログラムを確認すると、ACIとFACIの双方が指向する特徴が統合されていることが確認できる。具体的には、FACIの影響によりエネルギーが特定の周波数帯域に集中しており（広帯域ノイズの抑制）、かつACIの影響によりその帯域内で時間的に激しく変動する構造が見られる。これは、相乗平均をとることにより、ACIが求める時間的複雑性とFACIが求める周波数構造の偏在という二つの制約が同時に課され、その結果として、両方の条件を満たす複雑な変調を持つ生物音へと解が収束したことを示している。

この傾向は、図6(c)に示す最終世代の個体のプロンプトのワードクラウドにも表れている。「cricket」や「melodies」が大多数を占めるほか、「sparrow」、「bird」といった生物名も見られる。一方で、ACI単独条件で頻出した「brook」は見られず、ごく一部に「snapping twigs」が見られる程度であった。このことから、ACIとFACIの相乗平均の最大化は、非生物音を抑制しつつ、こおろぎや鳥類のような、時間的・周波数的に高度な構造を持つ生物種の生成に特化するよう選択圧を与えたことが示唆される。

3.5 進化過程の比較分析

本研究で提案した進化的探索手法により、プロンプト（遺伝子）とサウンドスケープ（表現型）がどのように進化したかを視覚的に理解するため、遺伝子空間と音響特徴空間における進化軌跡を比較した。プロンプトについては、各世代の各プロンプトをSentence Transformerによってベクトル化し、UMAPにより次元削減を行って二次元空間にマッピングした。本分析では、プロンプトと主要な単語を同一空間にプロットすることで、進化が特定の意味領域へどのように引き寄せられたかを視覚的に解釈する。サウンドスケープについては、生成された音声をスペクトログラムに変換し、AST（Audio Spectrogram Transformer）の特徴量として抽出し、UMAPにより二次元空間にマッピングした。なお、比較のため、次元削減は全ての条件での全試行のデータを利用し

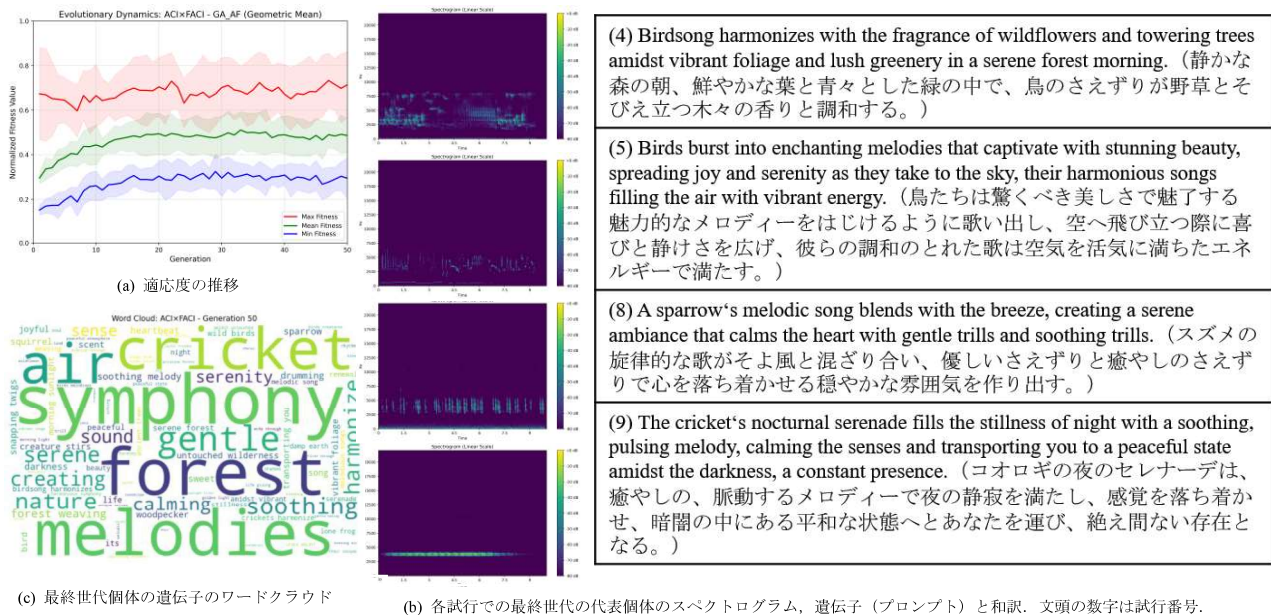


図6：ACIとFCIの相乗平均を適応度とした場合の実験結果

て行い、条件ごとにデータをプロットした。

図7に、条件ごとに色分けしたプロンプト空間とサウンドスケープ空間の重心の遷移を示す。同図に示すように、全ての適応度条件において、初期集団は空間の中央付近から探索を開始している。この初期状態から、各指標に基づき世代を経るごとに集団の重心が高適応度領域へと遷移しており、提案した進化的探索手法が有効に機能していることが確認できる。

遷移の傾向を見ると、ACIを適応度とした場合(b, f)が最も広い探索空間をカバーしているのに対し、FCIを適応度とした場合(c, g)が最も狭い空間に収束していることが示唆される。これは、ACIが評価する時間方向の複雑性が多様な領域に分散して存在している一方、FCIが評価する周波数方向の複雑性は特定の音に限定されるため、最適解の領域が狭いことを示唆している。

この差異は、適応度関数の選択が進化過程における選択圧として機能したことを意味している。ACIは幅広い自然音(生物音・非生物音の両者)に対して高い値を与えるため、解の探索空間への選択圧が弱く、拡散的になる。一方、FCIは周波数軸の構造に基づく強い選択圧により、生物音に特化した領域へと探索を集約させる。この現象は、単に高適応度解の特徴を反映するだけでなく、進化過程における選択圧が集団全体の進化方向を規定する進化ダイナミクスを示している。

ACIとFCIの相乗平均を適応度とした条件(d, h)では、両者の中間的な広さを示し、探索軌跡が安定していることから、両指標の制約が探索の方向性を

絞り込み、効率化に寄与していることが考えられる。この場合、2つの異なる選択圧が均衡することで、時間方向と周波数方向の複雑性を同時に追求する「バランスの取れた」解の領域が形成されることが示された。

4 おわりに

本研究では、音声生成LLMと、遺伝的アルゴリズムを組み合わせた、サウンドスケープの進化的探索手法を提案した。具体的には、LLMから生成されたプロンプトを遺伝子とし、そこから生成されるサウンドスケープを表現型とする進化的探索手法を構築した。この進化的探索手法を用い、ACI, FCI, ACIとFCIの相乗平均という3つの音響指標を適応度として用い、実験を行った。その結果、全ての適応度指標において集団の適応度が明確に向上し、その生態音響指標が目標とする音響特性へ収束するプロンプトとサウンドスケープを効率的に生成できることが示された。

実験結果の分析から、生態音響指標の理解においては、単なる指標値の特徴ではなく指標が持つ選別原理を認識することの重要性が示唆された。ACIは時間的変動性を普遍的にとらえる指標であるため、自然環境の複雑さ全般を反映する可能性がある一方、特定の生物類の検出に用いる場合には限界を持つ。FCIは周波数構造に焦点を当てることで、生物音への選択性を高めるが、その結果として評価対象の空間的な広さが制限される。両者の組み合わせで、より複雑な特徴を持つ生物音の抽出の可能性が示され

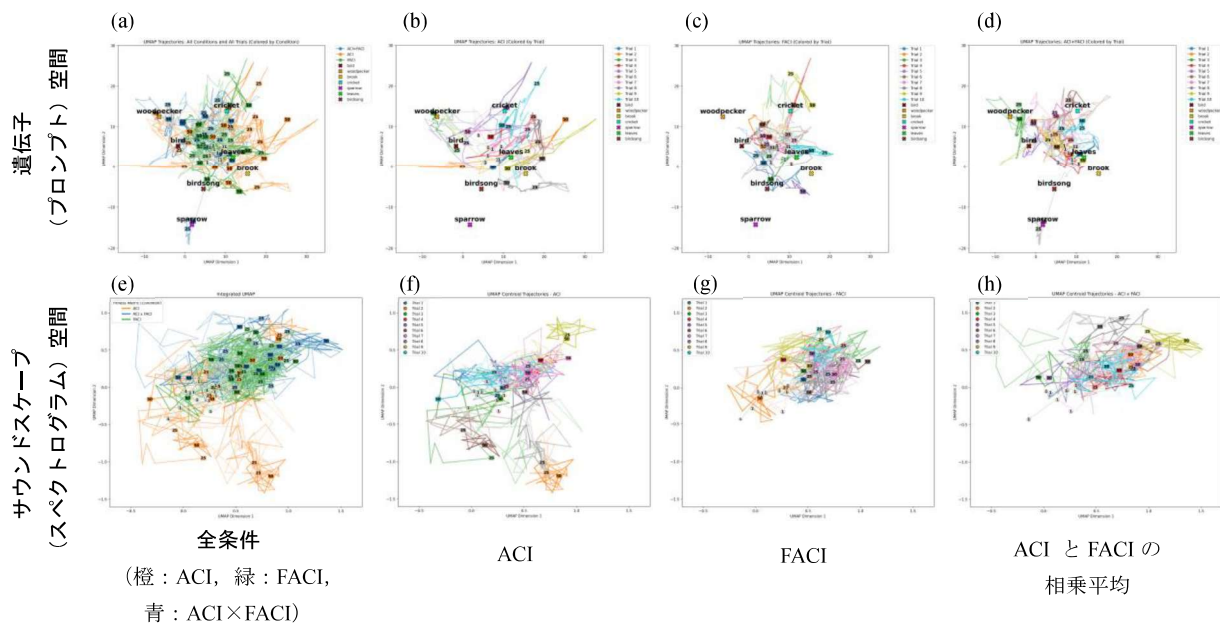


図 7：全・各適応度条件における遺伝子・サウンドスケープ空間上の重心の遷移

た。

今後、より異なる指標や、複数の指標の組み合わせにより、生態学的解釈の一貫性を高めながら、サウンドスケープの多次元的な特性を捉える手法の開発が期待される。

謝辞

本研究の一部は JSPS 科 研 費 JP24K15103 の助成による。

参考文献

[1] Farina, A., Gage, S. H. (2017). *Ecoacoustics: The ecological role of sounds*. John Wiley & Sons.

[2] Sueur, J., Aubin, T., and Simonis, C. (2008). Seewave, a free modular tool for sound analysis and synthesis. *Bioacoustics*, 18(2), 213–226.

[3] Ulloa, J. S., Hauptert, S., Latorre, J. F., Aubin, T., and Sueur, J. (2021). Scikit-maad: An open-source and modular toolbox for quantitative sound scape analysis in Python. *Methods in Ecology and Evolution*, 12(12), 2334–2340.

[4] Pieretti, N., Farina, A., Morri, D. (2011). A new methodology to infer the singing activity of an avian community: the Acoustic Complexity Index (ACI). *Ecological Indicators* 11, 868–873.

[5] Villanueva-Rivera, L. J., Pijanowski, B. C., Doucette, J., and Pekin, B. (2011). A primer of acoustic analysis for landscape ecologists. *Landscape ecology*, 26(9), 1233-

1246.

[6] Bradfer-Lawrence, T., Gardner, N., Bunnefeld, L., Bunnefeld, N., Willis, S. G., and Dent, D. H. (2019). Guidelines for the use of acoustic indices in environmental research. *Methods in Ecology and Evolution*, 10(10), 1796–1807.

[7] Bateman, J., and Uzal, A. (2021). The relationship between the Acoustic Complexity Index and avian species richness and diversity: a review. *Bioacoustics*, 31, 614–627.

[8] Farina, A., Pieretti, N., Salutari, P., Tognari, E., and Lombardi, A. (2016). The application of the acoustic complexity indices (ACI) to ecoacoustic event detection and identification (EEDI) modeling. *Biosemiotics*, 9, 227–246.

[9] Farina, A., Righini, R., Fuller, S., Li, P., and Pavan, G. (2021). Acoustic complexity indices reveal the acoustic communities of the old-growth Mediterranean forest of Sasso Fratino Integral Natural Reserve (Central Italy). *Ecological Indicators*, 120, 106953.

[10] Moysis, L., Iliadis, L. A., Sotiroidis, S. P., Boursianis, A. D., Papadopoulou, M. S., Kokkinidis, K.-I. D., Volos, C., Sarigiannidis, P., Nikolaidis, S., and Goudos, S. K. (2023). Music deep learning: deep learning methods for music signal processing—a review of the state-of-the-art. *IEEE Access, volume 11*, 17031–17052.

[11] Evans, Z., Parker, J. D., Carr, C., Zukowski, Z., Taylor, J., and Pons, J. (2024). Stable audio open. *arXiv preprint arXiv*, 2407.14358.

[12] Meyerson, E., Nelson, M. J., Bradly, H., Gaier, A., Moradi,

A., Hoover, A. K., Lehman, J. (2023). Language Model Crossover: Variation through few-shot prompting. *arXiv preprint arXiv*, 2302.12170.

- [13] Suzuki, R., Arita, T. (2024). An evolutionary model of personality traits related to cooperative behavior using a large language model. *Science Reports 14*, 5989.