

生成モデルによる形質表現を活用した 鳥類の鳴き声に関する進化モデルとフィールド実験の融合の試み

An approach to integrating evolutionary models and field experiments on avian vocalization using trait representations based on generative models

鈴木麗瑩^{1*} 古山諒¹ Zachary Harlow² 中臺一博³ 有田隆也¹
Reiji Suzuki¹ Ryo Furuyama¹ Zachary Harlow² Kazuhiro Nakadai³ Takaya Arita¹

¹ 名古屋大学, Nagoya University

² University of California, Berkeley

³ 東京工業大学, Tokyo Institute of Technology

Abstract: 本研究は鳥類の鳴き声を題材とし、生成モデルによる形質表現を活用した進化モデルとフィールド実験の融合を目的とした予備的試みについて、オオルリ、ホシワキアカトウヒチョウを対象とする最近の事例を報告する。オス・メスの遺伝子を VAE の潜在空間上のベクトル、それから生成される鳥類音声のスペクトログラム画像を形質・選好性とみなした性選択モデルを構築した。実験の結果、生成鳴き声の中でも明瞭であり単純すぎないものが選択されがちであることが示唆された。さらに、生成モデルに基づく再生音が野生個体にどのような影響を及ぼすかについて、プレイバック実験による予備的検討を行った。野生のオオルリに対する実験では、進化実験でよく選択された鳴き声について対象個体の鳴き返しを抑制する傾向がありうるということが示唆された。ホシワキアカトウヒチョウに対する実験では、人の耳にはノイズのように聞こえる生成音でも、なんらかの生態的な影響がありうるということが示唆された。

1 はじめに

生物や社会集団の進化や相互作用を理解するための計算論的アプローチの一つに、エージェントベースモデルがある。これは、生物・社会集団における主体間の相互作用や進化のルールを記述し計算機上で動かすことで、集団全体に生じる複雑な構造や進化のダイナミクスを理解するものである。その中でも、生命・社会現象の定性的な理解を志向するエージェントベースモデルの多くは、単純な構成要素やルールを設定することが一般的であり、そこから生じる複雑な過程は抽象レベルの理解や知見をもたらす。しかし、同時に、これらのモデルは実世界や社会の独特な複雑さと比べ大きな差異があり、直接的に比較するのが難しい場合もある。

一方、近年目覚ましい発展を遂げている ChatGPT や Stable Diffusion といった生成モデルは、インターネット上などの大量データを利用した表現学習を通じて、言語や画像に含まれる構造を潜在空間として抽象

化する。これにより、空間上の任意の特徴ベクトルから、元のデータには存在しない言わば架空のデータを生成することが可能となる。この技術によって、現実には即しつつも新奇で複雑な表現を計算機上で生み出せるようになった。これらの生成物のリアルさ、複雑性、新奇性は人々を驚かせ、特に大規模言語モデルの急速な普及は人間社会に大きな影響を与えている。

エージェントベース進化モデルに対して生成モデルがもたらすメリットは少なくとも2つあると考えられる [1] (図1)。一つは、従来よりも現実的、新奇で複雑な形質表現をモデルに盛り込むことができる点である。本研究で注目する鳥類の鳴き声に関連する生物・生態音響学において、多数の動物音声のスペクトログラムの表現学習や次元圧縮手法の利活用が盛んに検討されている。Sainburg らは鳥類をはじめとする様々な動物の鳴き声の録音に対し、深層学習ネットワークによる表現学習や種々の次元圧縮法を用いることで、その潜在空間上の構造を分類や関係の分析に利用するための Python パッケージを公開している [2]。また、Thomas らは近年汎用的な次元圧縮手法として多方面で利用される UMAP [3] に基づく動物音声の潜在空間作成につ

*連絡先：名古屋大学大学院情報学研究所
〒464-8601 愛知県名古屋市千種区不老町
E-mail: reiji@nagoya-u.jp

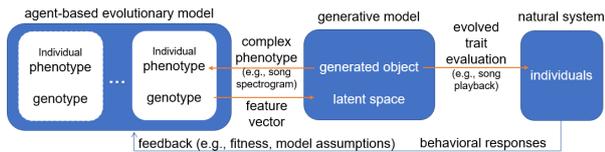


図 1: 生成モデルに基づく形質表現によるエージェントベース進化モデルの拡張

いて詳細に検討している [4]. Best らは、変分オートエンコーダ (VAE) に基づく鳥類や海棲哺乳類の鳴き声のクラスタリングの精度について詳細に分析している [5]. これらは既存の収録音の自動識別や特徴の分析のために低次元空間での分布を利用する試みである。

同時に、生成モデルはその低次元空間の任意の位置から当初の高次元データを生成することが可能である。Sainburg らは、VAE などの表現学習に基づく潜在空間表現で複雑な非線形な音声特徴を操作しつつ生成できたり、特徴を線形的に表現できることを指摘している [6]. 近年でも、マッコウクジラの音声を表現学習し、学習範囲外の生成音の特徴を分析してシグナルに潜在する役割を調べる手法も提案されている [7]. 前述の UMAP も最近潜在空間からの高次元データの生成も可能であり、生成モデルとしての役割も期待される。このような潜在空間をエージェントベース進化モデルの遺伝子空間とし、各遺伝子から生成される音声を形質とすることで、実生態の形質と同等な複雑さを持ったリアルな音声の進化モデルを構築することが可能である。

もう一つのメリットは、進化・創発したリアルな形質を直接フィールド実験で利用したり、その影響を観測したりできることである。しかし、このような生成音の影響を実生態においてプレイバック実験等で調べる試みは知る限りまだわずかである。さらには、両メリットを組み合わせることで、進化モデルと野外生物との直接の相互作用を検討することが考えられる。このような取り組みは音声を介した動物とコンピュータやエージェント間の相互作用 [8] に関連するといえるが、生成モデルを介したエージェントベース進化モデルと実フィールドとの接続の観点に基づく試みは知る限りない。一方、我々は、ロボット聴覚技術を用いた鳥類生態観測システム HARKBird [9] を構築し、プレイバック実験に基づく野外鳥類の行動観測を試行する中で、HARK で分離した音源のクラスタリング等に表現学習や次元圧縮手法を以前から活用しており [10, 11, 12], また、生成音のプレイバックによる影響についても一部調査を行ってきた [13].

以上を踏まえ、本研究は鳥類の鳴き声を題材とし、生成モデルによる形質表現を活用した進化モデルとフィールド実験の融合を目的とした最近の予備的試みについて 2 つの事例を報告する。具体的には、起点となる進

化モデルとして、Higashi らのオスの形質とメスのえり好みに関する数理モデル [14] に着想を得た性選択モデルを構築する。Higashi らのモデルでは、えり好みとその対象となるオスの形質がそれぞれ正負の整数値で表され、メスは自身のえり好みと評価対象のオスの形質を掛け合わせた値が大きいほどそのオスを選びやすい。ある好みを持つメスが好みのオスを選ぶことで個体内の遺伝子間の相関が高まることが繰り返されて生じるランナウェイ過程が正負の方向に同時に働き、同所的種分化が生じることが報告されている。本研究では、オス・メスの遺伝子を生成モデルの潜在空間上のベクトル、それから生成される鳥類音声のスペクトrogram画像を形質・選好性とみなし、メスは自身の選好性を表すスペクトrogram画像により近い形質を表すスペクトrogram画像を持つオス個体を繁殖相手としてより頻繁に選択する。本モデルを用いて、リアルで複雑な音声が発現しえり好みの対象となる性選択モデルではどのような進化ダイナミクスが創発し、どのような歌が選択されがちかを検討する。これをオオルリ、ホシワキアカトウヒチョウの二種に関する生成モデルを用いて行い、進化の一般的傾向について論ずる。

次に、生成モデルに基づく再生音が野生個体にどのような影響を及ぼすかについていくつかの実験の事例を紹介する。まず、オオルリの鳴き声に関する性選択モデルにおいて頻繁に選択された歌のいくつかを、同種オスの野生個体にプレイバックすることで、野外の鳥類に対してどのような影響を与えるかを検討する。次に、ホシワキアカトウヒチョウを対象にして、原点周辺に存在する潜在空間上の実録音の写像位置から順に遠ざかるように音声を生成し、明瞭なものから構造が崩れるにつれ、どれほどシグナルとしての役割を保持しうるかを検討する。

2 生成モデルによる鳴き声の潜在空間の作成

本研究ではオオルリ (*Cyanoptila cyanomelana*) とホシワキアカトウヒチョウ (*Pipilo maculatus*) に関する 2 種の生成モデルを変分オートエンコーダ (VAE) を利用して作成した。オオルリについては、2018 年 5 月に名古屋大学大学院生命農学研究所附属フィールド科学教育研究センター稲武フィールドにおいて、16 チャネルマイクアレイ (DACHO; System in Frontier 社製) で録音された野外録音データを採用した。マイクは針葉樹が生い茂り近くに小川が流れる林道中に設置されており、数個体のオオルリが近隣になわばりをもっていたと考えられる。録音からオオルリのさえずりを Praat を用いて歌いだしから 3 秒間を切り出し、496 × 128 ピクセルの音声スペクトrogram画像に変換した。全 30

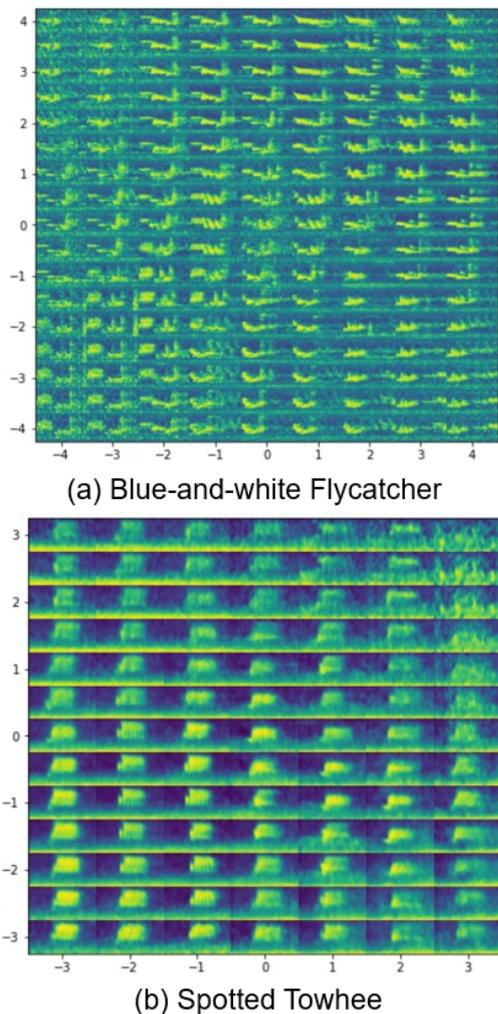


図 2: VAE で作成した 2 次元潜在空間上の鳴き声分布

種類からなる計 304 個のスペクトログラム画像をデータセットとして畳み込み変分オートエンコーダ VAE (8 畳み込み層と 3 完全接続層を持ち、2 次元にまで圧縮するエンコーダと、エンコーダと対称なデコーダ) の表現学習を行った。ネットワークの構成は Sainburg らの VAE[2] を参考にした。

図 2(a) は、生成モデルの 2 次元潜在空間において、各座標位置をデコーダに入力して生成されたスペクトログラムを規則的に並べたものである。同図から、座標の原点付近を中心に異なる種類の歌が集まって分布しており、外側に広がるに従いノイズが大きくなっていることがわかる。明瞭に生成された鳴き声には、オオルリの歌の特徴であるいくつかの音素に続く高周波の「ジジッ」まで再現されているものもある。実録音に対応する生成音を比較すると、生成音は対応する音声をおおむね再現しているが、細かな周波数や強弱の振動等はぼける傾向があった。

図 2(b) は、ホシワキアカトウヒチョウの鳴き声分布である。2023 年 5 月 14 日、米国カリフォルニア州の UC Berkeley の自然保護区である Blue Oak Ranch Reserve の林道沿いに生息する、複数のホシワキアカトウヒチョウのなわばり周辺において、8 チャンネルマイクアレイ (TAMAGO-03; System in Frontier 社製) で録音を行った。野外鳥類音源定位分離ソフトウェア HARKBird[9] を用いて音源定位・分離し、同種の歌 (2 秒) を約 1000 個取り出した。これらを上記と同様の方法で VAE で学習した。同種は個体ごとに短いさえずりをいくつか持っており、同図からそれらが少しずつ音響特徴を変えながら中心付近に分布していることがわかる。また、中心付近から離れるに従い、音響構造が明瞭でなくノイズが多かった状況であることもわかる。

以上から、HARK の音源分離による鳴き声スペクトログラムを潜在空間上にマップすることができることが分かった。一方、各音源の種類ごとに間隔をあけてクラスタが形成されるまでには至らず、これはサンプルサイズが影響している可能性がある。

3 生成スペクトログラムを用いた鳴き声と選好性の進化

3.1 モデル

前述の Higashi らの性選択による同所的種分化の数理モデルをもとに、前節で作成した 2 種の鳴き声生成モデルそれぞれについて、2 次元の潜在空間上の座標を遺伝子とし、その座標から生成されるスペクトログラム画像をオスの鳴き声、メスの選好性とした性選択モデルを構築した (図 3)。各 N 個体からなるオス集団とメス集団を考える。各個体は潜在空間上の位置 $((x, y)$ 座標のペア) を 2 つ実数値の遺伝子型として持ち、それぞれ自身がオスの際に発現する歌、メスの際に発現する歌に対する選好性に関する遺伝子とする。オス個体は、自身の歌の遺伝子型を入力したデコーダネットワークから生成されるスペクトログラム画像が表す歌をさえずるとする。また、各メス個体は、オスがさえずる歌に対する好みとして、好みの遺伝子型から生成されるスペクトログラムを生得的な選好性の鋳型として配偶者選択に用いる。

各メスは 1 個体のオスを選択して繁殖する。各メスはすべてのオスを $\exp^{-\beta \times x}$ で評価する。ここで、 x は注目するオスの歌スペクトログラムと自身の選好性スペクトログラムの各画素値の差の平均、 β は係数である。すべてのオス個体から、評価値に比例した確率で 1 個体のオスを選択する。同式は、歌のスペクトログラムと選好性が近いオスほど選ばれやすいことを示して

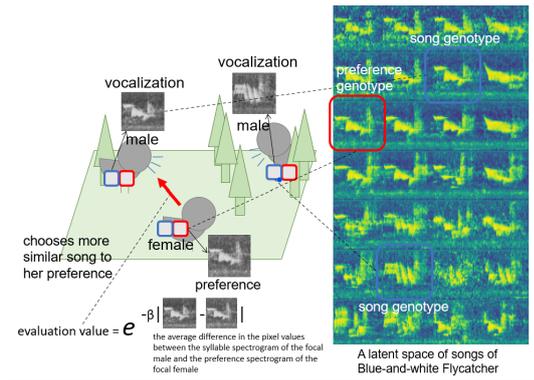


図 3: 生成スペクトログラムを用いた形質表現に基づく性選択モデル

いる。自身の遺伝子と選択した個体の遺伝子で2体の子個体のペアが作られる。この際、確率 p_c で BLX- α 交叉 [20] が生じ、加えて確率 p_m で各遺伝子に突然変異（平均 0, 標準偏差 σ の正規乱数の値を加算）が生じる。ペアのうち一方をオス、もう一方をメスとランダムに決定する。1 試行は初期集団の各遺伝子の値を $[-W, W]$ の一様乱数から決定するものとして、 T 世代にわたって行った。

3.2 進化実験

本研究では、 $N=100$, $W=5.0$, $\beta=0.3$, $\alpha=1.1$, $T=100$, $p_c=0.5$, $p_m=0.15$, $\sigma=0.2$ を採用した。両種の生成モデルを用いた実験に共通することとして次のことが分かった。数試行の分析から、図 4 に示されるように、各試行では初期集団から鳴き声遺伝子と選好性遺伝子が相関し合いつついくつかの集団に分化する傾向が見られた。これは、鳴き声と選好性のスペクトログラムの代わりに遺伝子の座標そのものを採用し、選好性は座標間の距離の逆数で計算した場合において、原点付近に急速に収束する結果と異なった。つまり、スペクトログラムの複雑な特徴が、Higashi らが示したような複数同時に生じるランナウェイ過程（オスのある特徴を好む遺伝的特性を持つメスが何らかの要因（遺伝的浮動やその他の淘汰圧、外界の影響など）で増加すると、メスはその特徴をより持ったオスを配偶者として選んで子個体をつくるため、次世代でその鳴き声遺伝子を持つ個体は同時にそれを好む遺伝子を持つ傾向が強まる。これは、メスがその特徴を持ったオスを選ぶことが自身の好み遺伝子も間接的に選ぶことにつながる。これが繰り返されてよりその特徴を持った鳴き声と好みと同時に集団に広まっていく過程）のような形で同所的なすみわけをもたらしたと考えられる。

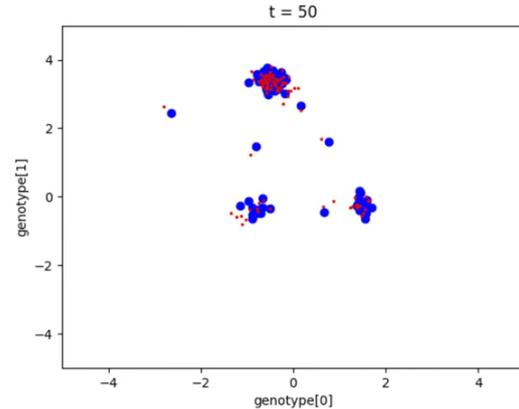


図 4: 進化実験におけるオスの遺伝子分布の例。青：鳴き声遺伝子，赤：選好性遺伝子。50 世代目のスナップショット（最終世代では上と右のクラスタのみ生存）。

一方、分化した結果収束する遺伝子は試行毎に大きく異なった。そこで、多数実験を行い全体的な傾向を抽出することにした。図 5(i) は、2 種それぞれ (a, b) に関して、全 (a)500 試行または (b)2000 試行における最終世代のオスの鳴き声遺伝子の頻度分布を KDE (kernel density estimation) 分布で示したものであり、色が青・緑・黄・赤となるに従い頻度が高いことを示している。(2) は比較のための図 2 の再掲、(3) は潜在空間上のスペクトログラムに対して音響複雑性指標 (Acoustic Complexity Index, ACI) [15] を計測したものであり、色が青・緑・黄・赤となるに従い ACI が高いことを示している。

まず、オオルリに関して、鳴き声遺伝子分布 (i) から、選択された鳴き声は全体として原点の周囲の広い範囲に分布することがわかる。これは、スペクトログラム分布 (ii) と比較すると、ノイズが少なく鳴き声が比較的明瞭に生成されている範囲に分布しがちであることがわかる。例えば、最も頻繁に選択された鳴き声遺伝子分布 (i) 中の 1 の位置の鳴き声は、典型的なオオルリのさえずりが明瞭に再生されるものであった。ACI 分布 (iii) と比較すると、ACI が低めの黄緑からやや明るい青色付近に分布していることがわかる。ACI は環境ノイズを取り除いた録音内に存在する生物音声を定量的に抽出する指標であり、各周波数ビンごとのパワーの時間変化が大きいほど大きい。同図では、ノイズが大きいスペクトログラムに対してそのまま適用しているため、ノイズの大きい領域では赤く極端に高い値になっているが、進化した鳴き声遺伝子はそこを避けて分布しており、ノイズが少なく明瞭な鳴き声を選択されがちであることがわかる。また、最も値の低い領域、例えばオオルリでは図中上付近も避けられているように見え、同時に単純すぎても選択されにくいこともわ

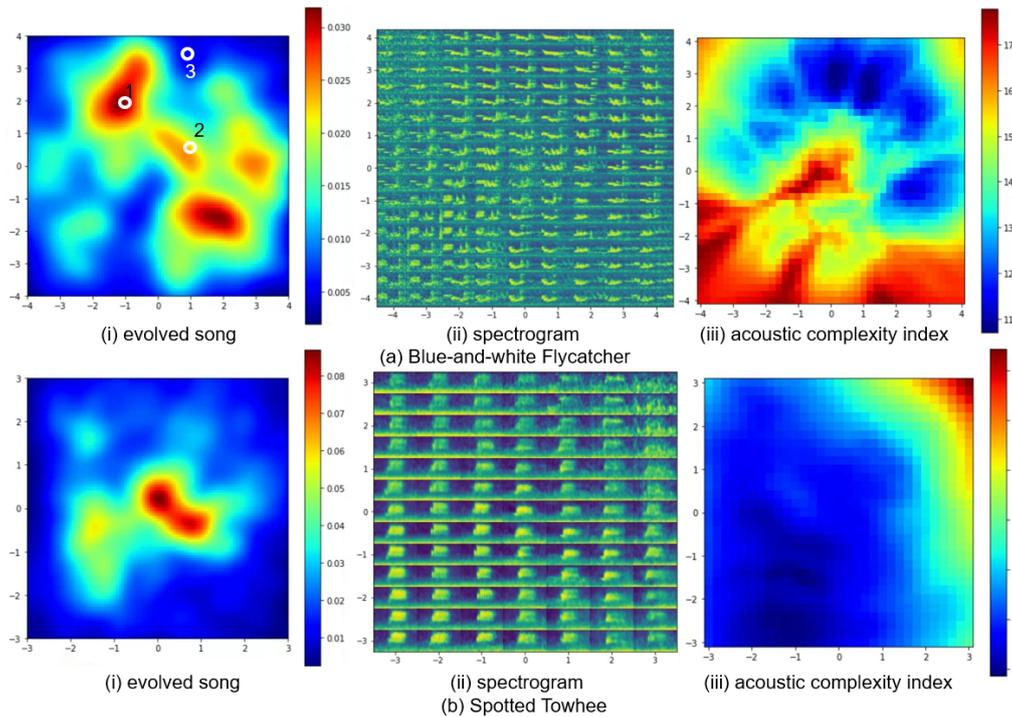


図 5: 進化実験結果. (a) オオルリ, (b) ホシワキアカトウヒチョウ. (i) オスの鳴き声遺伝子の分布, (ii) スペクトログラムの分布, (iii) 音響複雑性 (ACI) 指標の分布.

かる. ホシワキアカトウヒチョウについても, 中心付近だがやや上方と右下にずれた位置のピークと, 左下に中頻度で選択されたピークがあり, 明確ではないものの上記の傾向を反映しているといえる. これらから, 本進化モデルでは明瞭であり単純すぎない鳴き声を選択されがちであることが示唆された.

4 生成音を用いたプレイバック実験

前節での進化モデルで選択された歌や生成音が実際の生態でどのような特徴を持つか検討するため, オオルリ, ホシワキアカトウヒチョウそれぞれについて生成音を用いたプレイバック実験を試行した. 状況設定が大きく異なるため, 両種についてそれぞれ説明する.

4.1 オオルリに対する実験

進化モデルで選択された歌が実際の個体に与える影響の予備的検討のため, 2023年6月27日に名古屋大学大学院生命農学研究科附属フィールド科学教育研究センター稲武フィールドにて, オオルリのオス1羽に対して前節の生成音を用いたプレイバック実験を行った. 実験場所は, 周囲を針葉樹で囲まれた林道で, Raspberry Pi に接続され長時間録音が可能な USB マイクロホンア

レイ (TAMAGO-03, システムインフロンティア社製) を複数台設置し, 各マイクで同時に録音した. 現地を縄張りとしているとみられるオス個体に対して, 図 5(i) 中の 1~3 の位置に対応する生成音と, 同フィールドの別の場所で録音されたオスの鳴き声, “CD 鳴き声ガイド日本の野鳥” [16] に収録されたメスの鳴き声を加えた計 5 種である. 実際には, より多くの種類の鳴き声も用いて 26 日にも実験を行ったが, エゾハルゼミのや多種の鳴き声が大きく鳥類の行動に影響する可能性があったり, 音源定位が難しくなるため, 27 日の状況の良い録音のみを分析に用いた. それぞれについて 6.5 秒の間隔を置いて 10 回再生するファイルを作成し, 3 回繰り返し再生した.

実験間に 30 分以上の間隔を挟み, 実験場所付近にオスの個体が接近した際に実験を開始し, プレイバック前, プレイバック中, プレイバック後の約 5 分ずつの鳴き声を調査対象とした. HARKBird を用いて 2 つのマイクで到来方向に関する音源定位を行い, 同時刻に両マイクから定位されたオオルリの鳴き声の方向を用いて 3 点測量の方法で音源の位置を計算した. HARKBird の設定はオオルリの鳴き声をできるだけ抽出するように設定したが, 音源との距離やその他の音の影響で抽出できない場合があり, すべての音源の位置情報を取り出すことはできなかった. また, 基本的には同じ場所でさえずりを繰り返すため, 60m を超える移動は外

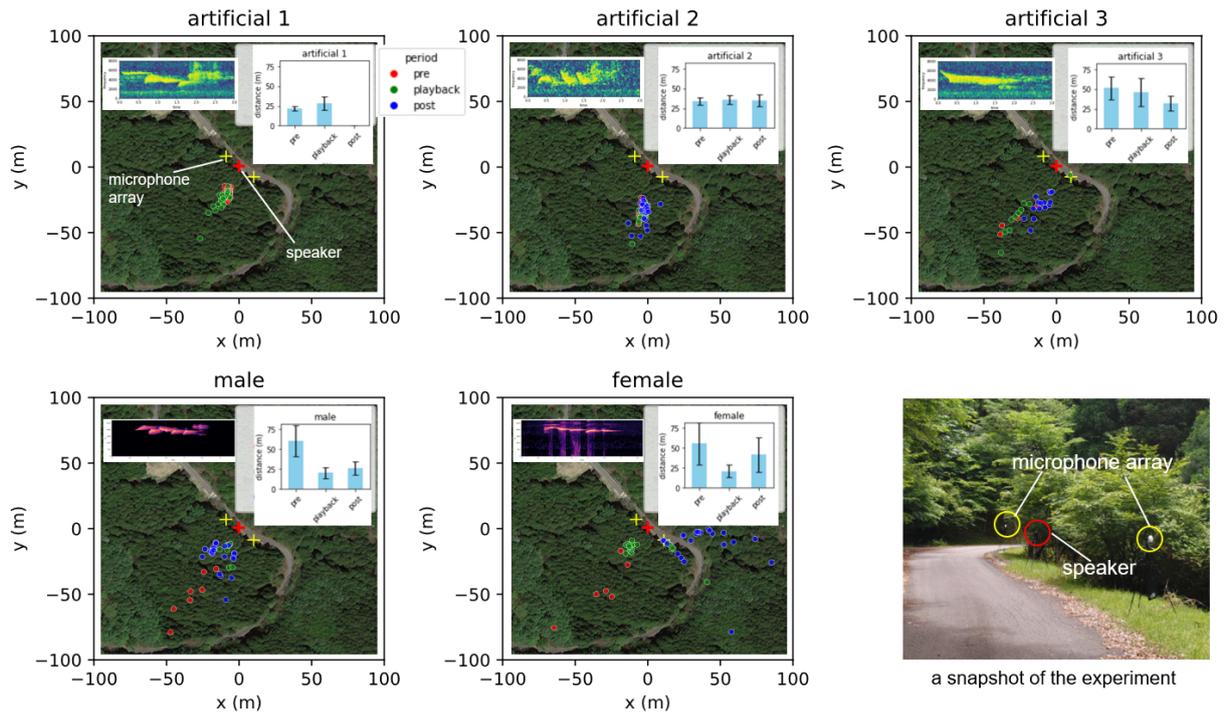


図 6: オオルリに対するプレイバック実験結果.

れ値として除外した. 以上のように予備的な分析であるが, 大まかな分布を知るには有益であると考えた.

図 6 は各再生音ごとの, プレイバック前・中・後で定位置された音源の二次元分布と, スピーカからの距離 (平均, 標準偏差) を示したものである. 対象の個体は, 林道沿いに設置したスピーカ・マイクから見て南側 (図中下側) の林道に囲まれた森の針葉樹の上でさえぎっていたことがわかる. 図中下段のオス・メスの鳴き声の場合, やや遠い位置でのさえぎりからプレイバック中に 20m を下回る極めて近い位置にまで接近してさえぎっており, プレイバック音に対する積極的な鳴き返しであると推測される. なお, オオルリはメスもさえぎる場合がある北半球において稀な種として知られており, かつて実施した実験においてもメスの鳴き声のプレイバック音に対して反応があったことを報告している [17]. 森の中でさえぎる様子はわかっていても, その中の詳細な位置関係を現地で観測するのは容易ではないため, このようなデータは繊細な生成音の影響の分析に役立つことが示唆される.

次に, 上段の 3 つの生成音を用いた実験に注目する. これらの生成音 1~3 は, それぞれ, 進化実験において, 高頻度・中頻度・低頻度 (もしくはほぼ選択されず) で選択された鳴き声であり, これらがもたらす影響を検討する目的で実施した. まず, 生成音 2, 3 の中頻度・低頻度で選択された鳴き声の場合, 実験期間を通して明確なさえぎり位置やスピーカからの距離の変化

は見られなかったが, 同様な位置でのさえぎりが維持された. 一方, 高頻度で選択された生成音 1 の場合は, 当初から近い位置でさえぎっていたのが, プレイバック中にやや離れたのちにその場から去る結果になった. これが, 進化実験の結果よく選択された鳴き声が明瞭で特徴的であり強力な生成音であるため鳴き返しをあきらめたものか, 単にプレイバック音に対する興味を失ったものか判断はつかないが, 同図の実験以外の設定において個体が去ってしまう状況はこれ以外に観測されておらず, 特徴的な行動であったことが考えられる. また, 実験を通して, 対象個体の当初のさえぎり頻度が低い場合はプレイバック期間に頻度が増加する一方, 当初の頻度が高い場合は逆にプレイバック期間に頻度が下がる傾向もやや見られ, 影響の可能性が示唆されるが, 検討が必要である. また, 実験時期が繁殖期の後期であり, より活発な時期における調査では異なる知見が得られる可能性がある.

4.2 ホシワキアカトウヒチョウに対する実験

VAE で生成した 2 次元潜在空間上では, 原点の周囲に鳴き声の特徴がよく表現された音が生成され, そこからさらに広がるにつれて構造が崩れノイズの多い音声生成された. この傾向において, 原点から遠ざかるに従ってもなおどれほど実際の個体に対してシグナ

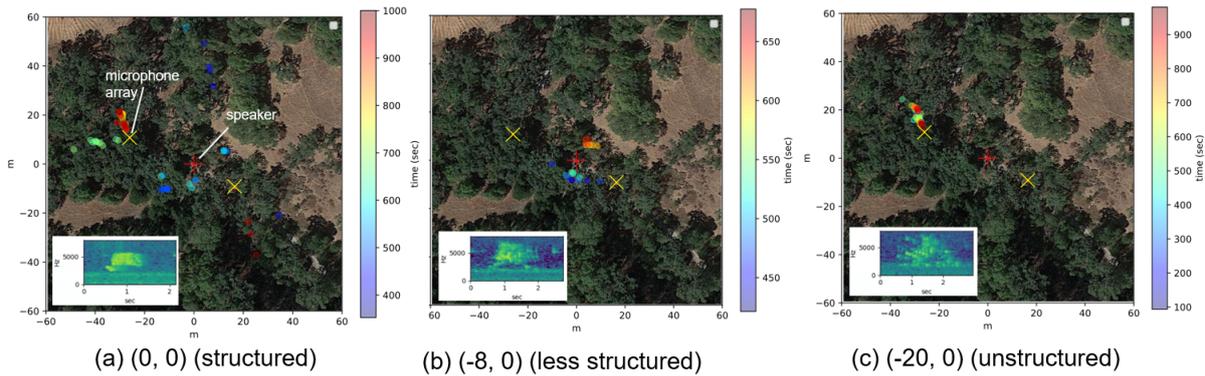


図 7: ホシワキアカトウヒチョウに対するプレイバック実験結果.

ルとしての影響がありうるかを検討するため、ホシワキアカトウヒチョウの鳴き声潜在空間において原点からの距離の異なる 3 つの生成音を用いてプレイバック実験を実施した。

2023 年 5 月 17 日、米国カリフォルニア州 UC Berkeley の自然保護区である Blue Oak Ranch Reserve の林道において実験を実施した。同フィールドの林道沿いには複数のホシワキアカトウヒチョウの縄張りが点々と存在しており、複数個体の反応が得られる可能性の高い場所にスピーカ、および、2 台のマイクアレイ (TAMAGO-03, システムインフロンティア社製) を 40m ほど離して設置し、各実験 30 分以上の間隔をあけて実験を行った。録音は午前中に行われ、20 分毎のファイルに保存された。プレイバック期間を含む各 20 分の録音におけるプレイバック対象個体の歌 (地鳴きは含まず) を HARKBird で音源定位し、三点測量の方法で 2 次元定位した。この時、到来方向は各音源が定位された時間を通した値の平均値を用いた。なお、いずれの場合もプレイバックをきっかけに単一、複数の個体がスピーカ付近に飛来したことを目視で確認している。

図 7(a) は、3 種のうち原点 (0, 0) の位置から生成された音声を再生した場合の結果である。この場合、図左下にあるようにホシワキアカトウヒチョウの典型的な鳴き声が明瞭に生成されている。同図から、時間変化に伴ってスピーカ周辺で複数の鳴き声が定位されていることがわかる。また、類似の時刻を表す同系色の音源が複数個所に分布しており、2, 3 個体が飛来し鳴き返していることを示している。これは、生成音が周辺の複数個体による強い反応を引き起こし、相互に影響し合ったことを示していると考えられる。

図 7(b) は、原点から離れて音響構造が崩れた生成音 (-8, 0) を再生した場合であり、1 個体が飛来し、スピーカ周辺にかなり接近し林道をまたぎつつ鳴き返す様子が観測された。明確に積極的な威嚇の行動であり、構造が明確でなく人工的でノイズを含んだ音声であって

も反応を引き起こしたのは興味深いといえる。

図 7(c) は、さらに原点から離れた (-20, 0) の位置からの生成音を再生した場合である。この時、西側 (図中左側) になわばりをもつ個体とその境界と思われる林道の端まで来て、スピーカからは遠い位置を維持しつつ鳴き返し、プレイバック終了後まもなく去っていった。実際の鳴き声ほどではないにせよ、何らかの興味を惹いたと考えられる。金属的なノイズのように聞こえる再生音でもなお、個体の行動に影響したことは興味深いと同時に、生成音には明示的でない形で生態に影響する特徴が含まれうることも示唆していると考えられる。

総じて、VAE の潜在空間において、よく鳴き声の特徴が反映される原点周辺から遠ざかるに従い、鳴き返し等の反応が弱くなるが、相当離れたほぼノイズのような鳴き声でも反応が得られる傾向が観察された。なお、同フィールドの別の場所で同様の実験を行った際には、(0, 0) を再生した場合にのみ個体の鳴き返しの反応が観測された。その際の近隣は当初の場所よりも静かで、全体的に活発でない様子であったが、部分的に上記の傾向を支持する結果であるといえる。

5 おわりに

本研究は鳥類の鳴き声を題材とし、生成モデルによる形質表現を活用した進化モデルとフィールド実験の融合を目的とした最近の予備的試みについて、オオルリ、ホシワキアカトウヒチョウを対象とする事例を報告した。オス・メスの遺伝子を VAE の潜在空間上のベクトル、それから生成される鳥類音声のスペクトログラム画像を形質・選好性とみなした性選択モデルを構築した。実験の結果、生成鳴き声の中でも明瞭であり単純すぎないものが選択されがちであることが示唆された。

次に、生成モデルに基づく再生音が野生個体にどのような影響を及ぼすかについて、プレイバック実験による予備的検討を行ったところ、野生のオオルリに対する実験では、対象個体のさえずりが維持される場合や、進化実験でよく選択された鳴き声が対象個体のさえずりの抑制や忌避をもたらす場合がありうるということが示唆された。また、ホシワキアカトウヒチョウに対する実験では、明瞭な生成音は野生個体の積極的な反応を招き、人の耳にはノイズのように聞こえる生成音であっても生態的な影響がありうるということが示唆された。

ChatGPT や Stable Diffusion のような生成 AI が社会に浸透し、人を驚かせたり、仕事を効率化したり、楽しませたりすると同時に、社会が生成されたコンテンツであふれたり知らないうちに人の行動にバイアスがかかったりするなどの課題がある。上記の知見は類似の課題が自然と AI 社会との接点においても生じうることを示唆していると考えられる。一方、これらの影響をうまく利用すれば、生成されたり進化する音を利用した環境保全など、エージェント進化モデル・人工システムと自然・生態との新しい接点やあり方を考えることができるといえる。今後はこのような可能性も考えつつ、ロボット聴覚技術を活用した進化モデルとフィールド実験の融合の試みを進めていきたいと考えている。

謝辞

本論文の研究の一部は JSPS 科研費 JP17H06383 (#4903), JP19KK0260, JP21K12058, JP20H00475 の助成を受けた。また、カリフォルニアの録音のアノテーションでは Hao Zhao 氏の協力を得た。ここに謝意を表する。

参考文献

- [1] Reiji Suzuki, Shinji Sumitani, Chihiro Ikeda, and Takaya Arita. A modeling and experimental framework for understanding evolutionary and ecological roles of acoustic behavior using a generative model. In *Proceedings of ALIFE 2022*, isal.a.00542, 2022.
- [2] Tim Sainburg, Marvin Thielk, and Timothy Q. Genter. Latent space visualization, characterization, and generation of diverse vocal communication signals. *bioRxiv*, 870311, 2020.
- [3] Leland McInnes, John Healy, and James Melville. UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv e-prints*, arXiv:1802.03426, 2018.
- [4] Mara Thomas, Frants H. Jensen, Baptiste Averly, Vlad Demartsev, Marta B. Manser, Tim Sainburg, Marie A. Roch, Ariana Strandburg-Peshkin. A practical guide for generating unsupervised, spectrogram-based latent space representations of animal vocalizations. *Journal of Animal Ecology*, 91(8), 1567–1581, 2022.
- [5] Paul Best, Sébastien Paris, Hervé Glotin and Ricard Marxer. Deep audio embeddings for vocalisation clustering. *PLoS ONE*, 18(7), e0283396, 2023.
- [6] Tim Sainburg and Timothy Q. Genter. Toward a computational neuroethology of vocal communication: From bioacoustics to neurophysiology, emerging tools and future directions. *Frontiers in Behavioral Neuroscience*, 15, 811737, 2021.
- [7] Gašper Beguš, Andrej Leban and Shane Gero. Approaching an unknown communication system by latent space exploration and causal inference. *arXiv e-prints*, 10.3389/fnbeh.2021.811737, 2023.
- [8] Roger K. Moore, Ricard Marxer, and Serge Thill. Vocal interactivity in-and-between humans, animals, and robots. *Frontiers in Robotics and AI*, 3, 61, 2016.
- [9] Reiji Suzuki, Shiho Matsubayashi, Richard W. Hedley, Kazuhiro Nakadai, and Hiroshi G. Okuno. HARKBird: Exploring acoustic interactions in bird communities using a microphone array. *Journal of Robotics and Mechatronics*, 27, 213–223, 2017.
- [10] Shinji Sumitani, Reiji Suzuki, Shiho Matsubayashi, Takaya Arita, Kazuhiro Nakadai, and Hiroshi G. Okuno. An integrated framework for field recording, localization, classification and annotation of bird-songs using robot audition techniques - Harkbird 2.0. In *Proceedings of ICASSP 2019*, pp. 8246–8250, 2019.
- [11] Shinji Sumitani, Reiji Suzuki, Takaya Arita, Kazuhiro Nakadai, and Hiroshi G. Okuno. Non-invasive monitoring of the spatio-temporal dynamics of vocalizations among songbirds in a semi free-flight environment using robot audition techniques. *Birds*, 2, 158–172, 2021.
- [12] 炭谷晋司, 鈴木麗璽, 有田隆也, 和多和宏, 松林志保, 中臺一博, 奥乃博. 複数マイクアレイを用いたキンカチョウの時空間的発声パターンに基づく個体間相互作用の調査. 第 58 回人工知能学会 AI チャレンジ研究会資料, pp. 12–20, 2021.
- [13] 炭谷晋司, 松林志保, 鈴木麗璽, 有田隆也, 中臺一博, 奥乃博. 生成モデルに基づく鳴き声を用いた鳥類に対するプレイバック実験の試行. 第 55 回 AI チャレンジ研究会資料, pp. 6–11, 2019.
- [14] Masahiko Higashi, Gaku Takimoto, and Norio Yamamura. Sympatric speciation by sexual selection. *Nature*, 402 (6761), 523–526, 1999.
- [15] N. Pieretti, A. Farina, and D. Morri. A new methodology to infer the singing activity of an avian community: The Acoustic Complexity Index (ACI). *Ecological Indicators*, 11, 868–873, 2011.
- [16] 松田道生. CD 鳴き声ガイド日本の野鳥. 日本野鳥の会, 東京, 2016.
- [17] 古山諒, 鈴木麗璽, 炭谷晋司, 有田隆也. 鳴禽類のメスのさえずりの役割の理解に向けた音源定位手法の活用に関する一検討. 第 58 回人工知能学会 AI チャレンジ研究会資料, pp. 6–11, 2021.