

# ロボット聴覚に基づくマルチスケール野外鳥類観測とこれから

## Multi-scale wild bird observation based on robot audition and future prospects

鈴木麗壘<sup>1\*</sup>

<sup>1</sup> 名古屋大学, Nagoya University

**Abstract:** 本稿では、ロボット聴覚技術である HARK と、鳥類鳴き声観測のために開発したインターフェイスである HARKBird を活用し、行動観測のスケールを、詳細な空間的行動傾向や複数個体の行動傾向、社会的関係、森林のサウンドスケープ（仰角・方位角空間での音源定位、多種間相互作用、生物集団）、自然と人工物（生成 AI）のように拡げて理解する試みについて概要を紹介する。

### 1 はじめに

ある種の鳥類（鳴禽類）の鳴き声には、繁殖期における近隣個体へのなわばりの主張や異性に対するアピールのための比較的長い鳴き声である“歌（もしくはさえずり）”と、それ以外の威嚇などの特定の伝達内容や機能をもった短い鳴き声である“地鳴き”がある [1]。“鳥は環境のバロメータ”ともいわれ、鳴き声の録音を用いた鳥類の自動行動観測は、自律録音装置の普及や AI 技術の進展に伴い、長期・広範囲の行動観測や環境評価、自然保護などへの応用が期待されている [2]。また、生態音響学 [3] では、自然の音風景の主要な構成要素である鳥類の鳴き声は重要な研究・調査対象であり、例えば音響ニッチ仮説 [4] に基づく鳴き声の時間・周波数領域の棲み分けなどが知られている。

我々は、ロボット聴覚技術を活用した、自然の音風景における動的な行動調整過程に基づく複雑系としての歌う鳥の集団の理解 [5] に取り組んでいる。ロボット聴覚技術は、多数のチャンネルを持ったマイクであるマイクロホンアレイの利用を軸にロボットが周囲の音環境を詳細に把握するための“耳”を実現する一連の技術の総称である。筆者らは、ロボット我々オープンソースソフトウェアである HARK (Honda Research Institute Japan Audition for Robots with Kyoto University) [6, 7] に基づく音源定位・分離技術を活用した、野外フィールドにおける鳥類の鳴き声観測ツール HARKBird [8] を用いた観測と理解を試行している。マイクアレイの生態観測への活用は従来から期待されている [9, 10] が、ソフト・ハード両面の制約等により限定的であった。そこに、ロボティクスで培われた実環境での動作において頑健な枠組みを持ち込むことで、どんな新しいことが

わかるか、定量化できるかを探ってきた。

本稿では、著者が代表を務めた、科研費国際共同研究加速基金 (国際共同研究強化 (B)) “鳥類の鳴き声を題材としたロボット聴覚技術のマルチスケール行動生態観測への応用”における、主に著者が直接かかわる取り組みについて概観する。本課題は、ロボット聴覚による鳥類観測技術を携えて、国外の共同研究先においてフィールドワークを行うことが一つの特色であったが、当初からコロナ禍の影響で渡航自体が難しくなり、その影響も含め共同研究先等の状況にも少なからず影響があった。そのため、従来の国内・海外調査データを活用した分析の洗練や成果としてのとりまとめ、将来を見据えた国内での調査や技術開発を進める状況が続いた。厳しい状況であったが、幸いにも後半においては、従来から音源定位技術に事前の経験や知識のある共同研究先からコロナ禍中でもサンプルデータを受け取り、後の渡航調査につなげることができた。また、最新の生態音響分析技術についても、その先端に行く研究者との交流や支援を得ることができた。一方、いずれも萌芽的段階であり、継続した共同研究への展開が望まれる状況にある。

本課題のもう一つのポイントはマルチスケールである。当初の目標からはやや意味合いが異なる面もあるが、前述の状況の中でありながらも、研究分担者の先生方や HARKBird に興味を持ち研究に協力して下さった方々、多くの研究室の学生と共に、行動観測のスケールを、詳細な空間的行動傾向や複数個体の行動傾向、社会的関係、森林のサウンドスケープ（仰角・方位角空間での音源定位、多種間相互作用、生物集団）、自然と人工物（生成 AI）のように少しずつ拡げ、応用の可能性を検討したいいくつかの取り組みについて概説する。それぞれの研究の詳細については各論文を参照されたい。

\*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

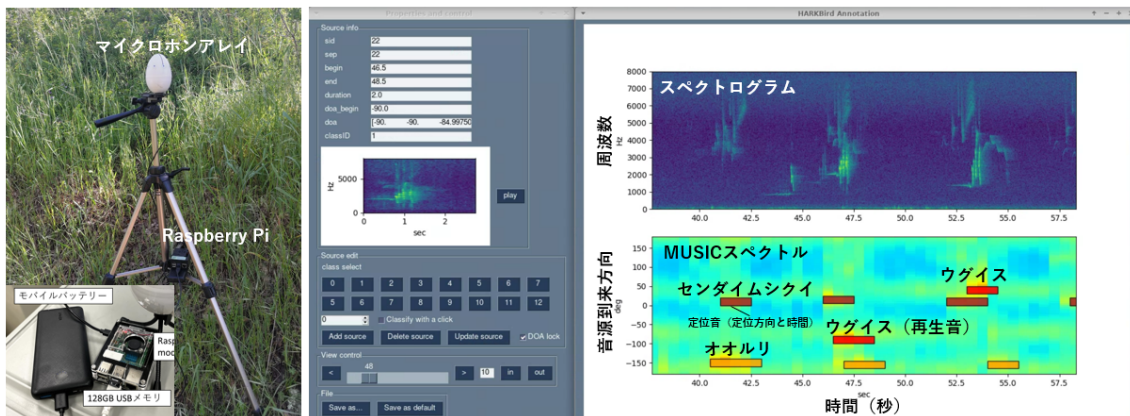


図 1: 左) 録音ノード. 右) HARKBird での音源定位結果の可視化・アノテーション例.

## 2 HARKBird とその拡張

HARK[6] は、マイクアレイを搭載したロボットが周囲のどの方向から音源がやってきているか（音源定位）、また、それがどんな音であるか（音源分離）をリアルタイムに抽出するためのソフトウェアである。多様な目的のリアルタイム音声処理を実現可能な汎用的な枠組みであり、最近では Python のパッケージとして利用可能になるなど発展を続けている。一方、森にマイクアレイをしばらく設置して音構造を可視化したり分析したりするなど、生態観測に応用するには別途枠組みが必要である。HARKBird[8] は HARK から得られる音源定位・分離情報を可視化・可聴化し、必要に応じて編集（アノテーション）することができる Python スクリプト集である<sup>1</sup>。

図 1 は、筆者らが使用する森林での録音ノード（左）と、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での録音の分析例（右）である。ノードは市販の USB 接続の 8 チャンネルマイクロホンアレイ（システムインフロンティア社製、TAMAGO-03）、Raspberry Pi、モバイルバッテリー、三脚からなる [11]。当初はノート PC をマイクアレイに直接接続し、そのうえで HARKBird も動かす形態をとっていたが、可搬で複数のマイクアレイでの録音を容易にするために、ノード自体が WiFi のアクセスポイントとなり、スマートフォンを用いて Web インターフェイスを通して時刻の自動設定や録音スケジュールの設定ができるように改良した。結果的に、この拡張は海外共同研究先での円滑な録音サンプルの取得に大きく貢献した。

右図は、HARKBird3.0 のスナップショットである。従来の HARKBird と比べて動作が安定していることに

加え、アノテーションのしやすさや、2 つのマイクでの音源定位結果とマイクの位置・方位情報をもとに三点測量で 2 次元音源定位する機能、次元圧縮手法である UMAP を用いたクラス分け機能を試行的に備えるなどの改良が加えられている。上はスペクトログラム、いわゆる鳥の“声紋”であり、横軸は時間（秒）、縦軸は周波数（Hz）であり、いくつかの異なるパターンが存在するのがわかる。右図下は HARK の要の一つであり、各方向の音源の到来強度を示す MUSIC (Multiple Signal Classification) [12] スペクトルである。HARK はこの値が高まった部分を追跡することで音源定位し、その情報をもとに GHDSS 法 [13] に基づいて音源を分離する。

MUSIC スペクトル上の矩形は HARKBird でその結果（音源定位開始方向と持続時間）を示しており、クリックすると図左のウィンドウに分離音のスペクトログラム等の情報が音再生とともに表示される。同図から、ウグイス、オオルリ、センダイムシクイがマイク周辺の異なる方位でさえざっていることがわかる。屋外での録音には分析の対象外の音源が無数に含まれるため、このインターフェイスを用いて生態分析に必要な音声のみを取り出してその後の分析に取り組むことになる。

## 3 詳細な個体の行動傾向の抽出

鳴き声を持つ役割をフィールドで調べる方法に、同種の鳴き声をスピーカで再生して反応を調べるプレイバック実験がある。本課題開始までに、野生のウグイスに対して同種の歌を再生し、再生歌の種類や頻度などが野生個体の単一のマイクから見た鳴き返しの方向や頻度に与える影響を調べてきた [14]。また、揖斐川河口のオオヨシキリのソングポストでさえざる個体を複数のマイクアレイで定位することも試みてきた [15]。

<sup>1</sup><https://sites.google.com/view/alcorsuzuki/home/harkbird>

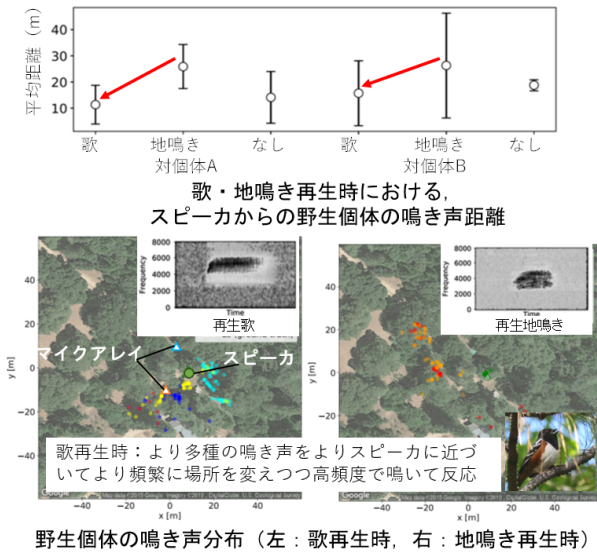


図 2: ホシワキアカトウヒチョウの再生鳴き声 (同種歌・地鳴き) に対する反応 [16].

本課題では、これらの取り組みを拡張し、共同研究者である UCLA Martin Cody 名誉教授, Charles Taylor 名誉教授の主要な調査地である米国カリフォルニア州アマドール郡の森林において、ホシワキアカトウヒチョウ (Spotted Towhee) に対して同種の歌・地鳴きをスピーカ再生した際の個体の平面的な移動傾向を複数のマイクアレイで観測し、定位音源の空間情報を人手で、もしくは、音源定位タイミングと次元圧縮法である t-SNE による類似度を組み合わせた手法で抽出することを試みた [16] (図 2)。課題は複数マイクアレイからの同一音源情報のすり合わせ (一方のマイクでの定位音源が他方のどの定位音源に対応させるか) であり、上記の手法では必ずしも十分ではなかったが、手作業によるアノテーション結果からは、歌をスピーカ再生すると、地鳴き再生時と比べて、より多種の鳴き声をよりスピーカに近づいてより頻繁に場所を変えつつ高頻度で鳴いて反応することが示唆された。

UC Berkeley の Zac Harlow 博士の協力を得て、コロナ禍で渡航が制限される中でも、同大学の自然保護区である Olue Oak Ranch Reserve に生息する複数のホシワキアカトウヒチョウの複数マイクノードによる録音サンプルを収集することができた。分析からは、同種は同森林内になわばりをもって数多く生息し、いくつかの種類の 2 秒弱の短い鳴き声を比較的大きな声で繰り返しさえずるため、音源定位に基づく観測に適した種であることが判明した。

そこで、現地において複数のホシワキアカトウヒチョウに対する同種歌のプレイバック実験を行った [17]。図 3 は道沿いに東西に分かれたなわばりの境界付近に同

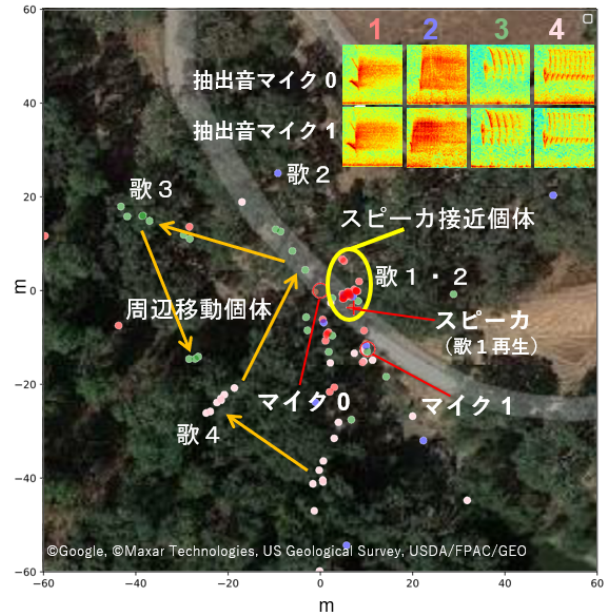


図 3: 鳴き声プレイバックに対する複数のホシワキアカトウヒチョウの反応 [17].

種の鳴き声をスピーカで再生した時に、2 体の同種個体がさえずった場所を可視化したものである。2 つのマイクアレイからの情報のすり合わせについては、分離音源を変分オートエンコーダーで表現学習した特徴空間を利用して、あらかじめ取り出した 4 種の典型的な歌に近い音源を取り出し、半自動的に対応付ける方法を採用した。ノイズが含まれるが全体の挙動を把握するには十分であったと考える。

実験では、東側の道沿いにスピーカを設置し、数日前に収録したそのなわばり個体の鳴き声を数回再生し、そのまま数分程度観測を続けた。その結果、右側のなわばりの個体は終始スピーカのすぐ近くに接近し、急に現れた自身と同じ鳴き声持つ侵入者に対して激しい威嚇の反応を示し続けたのに対し、左側のなわばりの個体はその様子を認識して近寄ってはいるものの、数十メートルの一定の距離を置いてさえずっている様子が観測された。なわばりに対するスピーカの設置位置によって異なる反応が生じたことが示唆される。

より実践的な生態観測として、本課題の研究分担者の松林准教授らが中心となって、オオジシギ [18] やサンカノゴイ [19] 等の国内の希少種の行動傾向のマイクアレイを用いた観測に取り組んでいる。

## 4 社会的関係の抽出

以上の事例はスピーカによる鳴き声再生で仮定の侵入者を想定したものであったが、より自然な環境で多



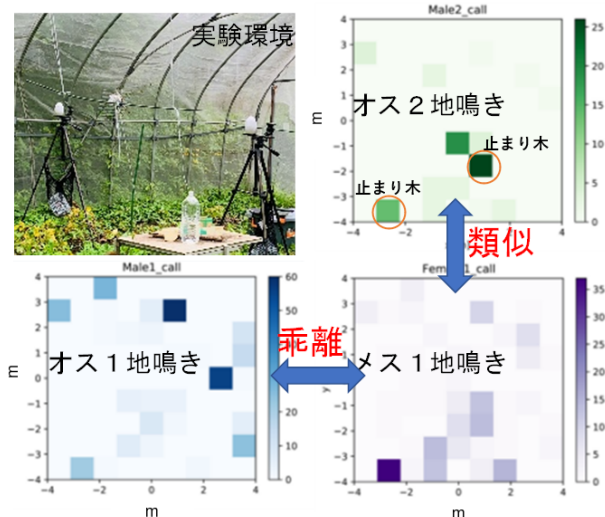


図 4: 屋外テント内での3羽のキンカチョウ（オス2羽，メス1羽）の鳴き声分布 [20].

くの個体間の社会関係を取り出したいという期待がある。一方、実生態において都合よく個体間相互作用を録音観測するのはそう簡単でないというジレンマも存在する。そこで、北海道大学和多和宏准教授の研究協力のもと、同大学の圃場にテントを張りその中で数羽のキンカチョウ個体を放つことで自由に飛翔可能な環境を構築（図4左上写真）し、複数のマイクアレイを設置して精度良い鳴き声位置と種類（個体と歌・地鳴き）の推定を試みた [20].

その結果、次のような面白い予備的知見が得られた。オス1羽のみを放鳥して数時間観測した場合、えさ場や止まり木で頻繁に鳴く様子が観察された。同種は社会性の高い種であり、今回注目した地鳴きは distant call と呼ばれる社会的結合の維持の役割があるとされるため、仲間を探していたことが考えられる。次に、オス2羽の場合、鳴き声頻度は小さくなったが、両個体が近い場所で一緒に鳴きがちであることが、鳴き声場所の時系列データから推測される個体間距離から明らかになった。近所で仲間とたまに鳴き合って落ち着いているのかもしれない。ところが、そのオス2羽に加え新たにメス1羽を加えた場合、オス同士は距離をより保って鳴く一方、メスはオスのいずれかの近くにどまりがちになる傾向があった（図4）。つまり、第三者の存在が二者の関係を改変する様子を定量化できたといえる。

キンカチョウは発声学習等のモデル生物であり、マイクが付いた小さなリュックサックを個体に装着して相互作用を定量化する試みなど様々な手法が開発 [21] されているが、この結果はテント内ではあるものの非侵襲的な方法で繊細な社会的関係の変化を観測できる

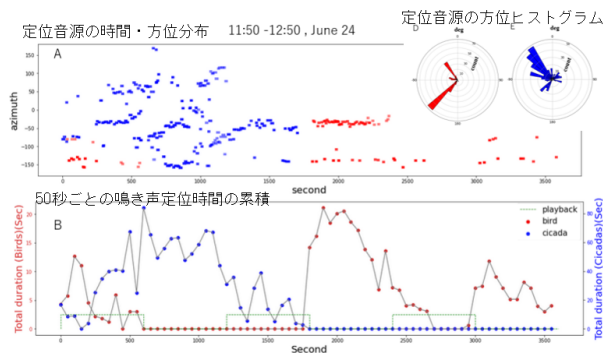


図 5: 8ch マイクアレイ TAMAGO を用いたセミ（エゾハルゼミ）と鳥類（主にコルリ）の鳴き声定位. 赤：鳥類，青：セミ [22].

可能性を示したと考えている。

## 5 森林のサウンドスケープの抽出

これまで、2次元空間でのプレイバック時の複数個体の挙動や限定された状況での社会的関係の抽出の可能性が示された。一方、自然界の音環境は、サウンドスケープと呼ばれるように、より高次元の空間においてより多くの音源から全体が構成される。そのようなより複雑な鳥類に関わる音環境の抽出の試みのいくつかをまとめる。

名古屋大学稲武フィールドでは、6月初めごろからエゾハルゼミというセミの鳴き声が2週間から1か月ほどの間あちこちでよく聞こえる。このセミは、カエルと間違えそうな短い鳴き声の繰り返しと、それに続くヒグラシのような長い鳴き声で構成され、一部の鳥類の鳴き声と周波数が重複するために、鳥類の鳴き声の観測にとって厄介である。また、セミと鳥類の鳴き声の間で時間的な重複を避ける場合があることも指摘されており [23]、鳥類の行動への影響も少なからず存在することが考えられる。

一方、これを異種間の相互作用や複数種が創るサウンドスケープの典型例であると考え、俄然興味の対象となる。そこで、セミの鳴き声が鳥類行動に及ぼす影響を議論する知見が得られるかを検討するために、セミとトリの両種の鳴き声を音源定位の対象とした観測実験を実施した [22]。図5はTAMAGOを用いてエゾハルゼミとコルリを主とする鳥類がさえずる録音を音源定位し、音源をセミ（青）とトリ（赤）に分類した一例である。上図は時間経過に伴ってマイク周辺で多くのセミがバーストのように鳴きだし、それが落ち着くと2個体ほどの鳥類個体が鳴きだす様子を示している。下図は50秒ごとにセミ、トリそれぞれの定位音

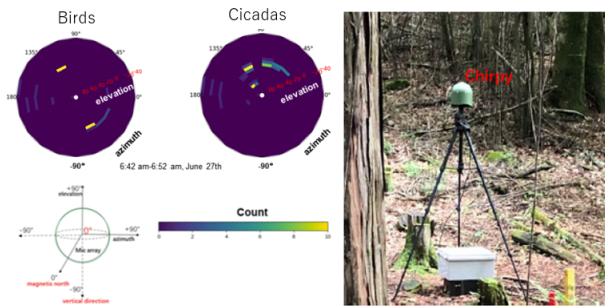


図 6: 右: 16ch マイクアレイ CHIRPY を用いた観測風景. 左: 10 分間の鳥類・セミの鳴き声の仰角・方位角分布. 手作業によるアノテーション結果に基づく [24].

源の長さを積算して、発声の活動量として定量化したものである。同図からも、セミとトリが鳴く時間を分けている様子が見える。なお、図中緑色の矩形のグラフは、近隣でエゾハルゼミが連続して鳴く録音を 10 分おきに再生した時刻を示している。両種の活動量はセミ録音再生時にはやや低下したが、この録音では有意な影響はなかった。異なる状況での観測と分析については論文を参照されたい。

次に、研究分担者の中臺教授が中心となって開発された、屋外において単体で長期バッテリー動作し、音源の仰角・方位角が取得可能なマイクアレイ CHIRPY (図 6 右) を用いて両種の音源の分布の抽出を試みたところ、鳥類音声はマイクから見て類似の高さの一定の方向で定位され、セミの鳴き声は仰角・方位角に関してより幅を持って定位されることが示された (図 6 左) [24].

この時分析で問題になったのは、次元の増加の影響等で定位音源の中に誤検出と思われる偽の音源や、風や反射の影響と思われる音源が少なからず含まれ、単純な生態音響特徴量 (例えば音響複雑性 [25]) に基づく分類が容易でないことであった。そこで、機械学習を用いた鳥類音声や生態音響の識別 [26, 27, 28] に見識の深い Tilburg University の Dan Stowell 准教授らとの研究議論も踏まえ、研究分担者の小島講師が公開している UCLA の Charles Taylor 名誉教授らによるカリフォルニアの鳥類鳴き声のアノテーションデータベースである Bird-DB [29] を利用して学習した wav2vec2 や、一般的な人間の会話を利用して学習した wav2vec2、また、鳥類学のメッカの一つである Cornell 大が開発公開する BirdNET [30] による、分離音の埋め込み表現と SVM を用いた分類の学習を試行している [31]。現在のところ、埋め込み表現はいずれも生態音響特徴量を用いた分類より成績が良いが、BirdNET, wav2vec2 (鳥類音声) の成績が人間の会話に基づく wav2vec2 (人間会話) よりも良いようである。

筆者の知る限り、セミや鳥類の個体レベルの鳴き声

の分布を可視化したり活動を定量化した例はなく、自然のサウンドスケープを可視化・定量化する新しい方法論であると考えている。

さらには、研究分担者の中臺教授らは CHIRPY を活用したサウンドスケープの定量的解析に取り組んでおり [32] いる。また、トリルラボの大坂氏と共同で 16 チャンネルマイクアレイ DACHO の音源定位性能の評価を実フィールドで行い、マイクの上方に存在するスピーカ再生音の仰角・方位角に関する音源定位精度等について分析を進め、成果をまとめている [33].

## 6 生成音が鳥類生態に及ぼす影響

最後に、やや創造力を働かせて自然と人工システムとの相互作用にまでスケールアップした初期的取り組みについて述べる。近年の AI 技術の大きな発展要因の一つは、前述のような高次元データの特徴を低次元で表現する埋め込み手法の進展によるところが大きい。特に鳥類音声を利用して学習した埋め込みモデルは、他種の音声の埋め込みにも有用であることなどが示され、自動分類等への活用が検討されている [35].

同時に、これらの手法の多くは、従来のデータにない埋め込み表現からリアルでありつつも新奇なデータを生み出す生成モデルとしても利用可能であり、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したり [36], 学習範囲外の生成音の特徴を分析して動物音声に潜在する役割を調べる新たな手法が提案されている [37].

本課題では、これを一歩推し進め、生成音を用いたプレイバック実験を試行した。具体的には、ホシワキアカトウヒチョウの歌のスペクトログラム画像をデータセットとして VAE で 2 次元の潜在空間を作成し、その中心から離れ、歌の構造が不明瞭になるに従って、野生個体の縄張り内でスピーカ再生した場合のような影響を及ぼしうるかを予備的に実験した [34]. その結果、歌の構造が明瞭に表現される潜在空間中心付近の歌は周辺の複数個体の接近と鳴き返しを引き起こした (図 7a 左下) が、大きく構造が崩れた鳴き声であっても 1 個体が離れた場所から様子をうかがうような挙動も観測された (図 7a 右下)。つまり、人にはノイズのように聞こえる音源であっても、気づかないうちに何かしらの生態的な影響が及ぶことが示された。

また、筆者は長らく人工生命研究において特にエージェントベース進化モデルを用いた生物の適応進化に関する研究に従事しており、大規模言語モデルをはじめとする生成モデルを用いて遺伝・形質表現をよりリアルなものに拡張する試みに取り組んでいる [38]. そこで、鳥類の歌の適応進化に関わる性選択モデルを題材に、フィールドで収集した鳥類鳴き声に基づく生成



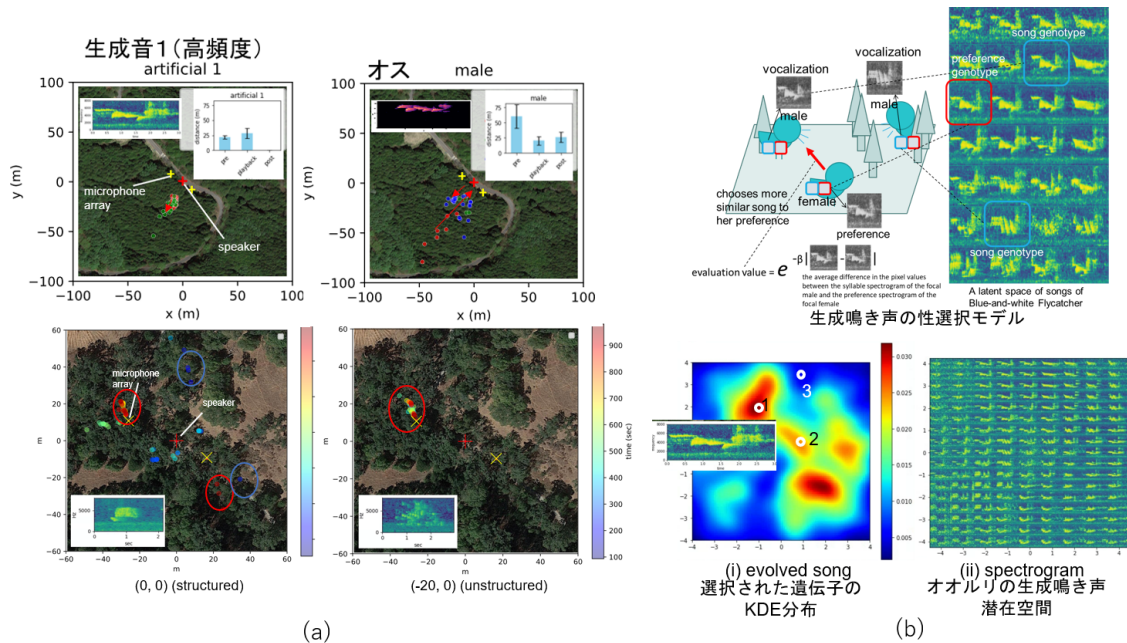


図 7: a) 生成鳴き声のプレイバック実験. 上段: 性選択モデルで頻繁に選択されたオオルリの生成鳴き声 (左) と野生のオスの鳴き声 (右) 再生時のなわばり個体のさえずり分布. 下段: ホシワキアカトウヒチョウの生成鳴き声 (左: 明瞭な構造, 右: 不明瞭) 再生時のなわばり個体のさえずり分布. b) オオルリの生成鳴き声の性選択モデル (上段) と進化実験で頻繁に選択された鳴き声遺伝子の分布 (左下), オオルリの鳴き声潜在空間上の鳴き声分布 (右下) [34].

モデルにおける埋め込み表現を遺伝子, 生成音を表現型として用いることで, 実際に聞くことのできる鳴き声の進化モデルを構築した. 具体的には, フィールドで収集した鳥類の鳴き声スペクトログラムを畳み込み変分オートエンコーダで表現学習し, 鳴き声の2次元潜在空間を作成した (図7b 右下). 空間上の2つの座標を各個体が遺伝子として持ち, そこから生成される鳴き声をそれぞれオスとしての鳴き声とメスとしての鳴き声に対する選好性 (好む歌の鑄型) とみなす (図7b 上). 形質がスペクトログラムであるため, 実際に再生して聞くことができる.

このモデルで頻繁に選択され生き残った鳴き声 (図7b 左下) を野生個体に再生し, 生成音や進化実験で生じる音声の生態への繊細な影響を HARKBird で調べる枠組みを提案している [39, 34]. つまり, モデルで進化する鳴き声とはどんなものか, また, それは実際の鳥類をビックリさせられるかを調べる試みである. 例えば, 名古屋大学稲武フィールドのオオルリを対象にこの枠組みを適用した予備的実験では, モデルで頻繁に選択された歌は明瞭な適度に複雑な構造を持ち, それを野生個体のなわばり内で再生するとその場でさえずる個体が立ち去る (図7a 左上) など, 実生態でも個体に対する繊細な影響がありうることを期待させる結果を得たが, 信頼性の高い実験と分析が必要である.

いずれも予備的な試みだが, 生成 AI が注目を浴びる中生成された人工物が生態に与える影響や自然と人工物のあり方を考える一つの方法であると考えている.

## 7 まとめとこれから

本稿では, ロボット聴覚技術である HARK と, 鳥類鳴き声観測のために開発したインターフェイスである HARKBird を軸に, 行動観測のスケールを, 詳細な空間的行動傾向や複数個体の行動傾向, 社会的関係, 森林のサウンドスケープ (仰角・方位角空間での音源定位, 多種間相互作用, 生物集団), 自然と人工物 (生成 AI) のように広げて理解する試みについて概説した.

これまでに述べたいくつかの取り組みにも含まれるように, 特に埋め込み表現を活用したデータの分析手法の進展は著しく, 最近では録音から自然言語で指定した種類の音源を取り出す仕組みまで提案されている [40]. これは, 単に特定の種類の音源を分類するのみならず, 録音の中の文脈全体を自由に取り出せることを示唆している. このような近年の技術の急速な進展に合わせた積極的な活用が期待される. 例えば, これまでの HARK の活用方法では定位した各音源を分類しているが, これらの手法でまず取り出した音源について, 定位方向を HARK で取り出す方法もありうる.

音源定位に関しては、本研究では単一のユニットに複数のマイクが搭載されたマイクロホンアレイを採用しているが、単一マイクの録音ユニットを多数広範囲に配置し、音源の到達時間差で定位する方法 (TDOA) も、ユニットの低価格化と長期録音の容易さ等から広まりつつある。一方で、プレイバック実験や希少種の観測のように、観測場所があらかじめ絞りにくい場合や、移動を繰り返しながら観測する場合等では、可搬性や精度面も含めて本課題で採用しているマイクアレイユニットの方式が有用であると考えられる。

また、HARK の最も重要な特徴の一つでありながらこれまで十分活かしきれていないものに、リアルタイム処理がある。前節のように、生成 AI は人工物に自然と同レベル・同文脈の複雑さを生成可能にしている。近年、スマートフォンなどで手軽に鳴き声を再生することが及ぼす鳥類生態への影響が心配される中、前述のような生成音声は自然に与える影響はより複雑なものになりうる。しかし、今後より重要となるのは、人工物と生態との相互作用であると考えている。我々の試みの中では、野生のウグイスの鳴き声を定位した後数秒後にシステムが同種の鳴き声を鳴き返す実験がある [14]。HARK のリアルタイム処理と、生成 AI、前述の進化計算・エージェントベースモデルなどの適応システムを組み合わせ、自然と人工システム間の音を介した新しい関係を模索することが可能であり、将来の方向性の一つであると考えている。

## 謝辞

本研究の一部は JSPS 科研費 JP17H06383(#4903), JP19KK0260, JP21K12058, JP20H00475 の助成を受けた。

## 参考文献

- [1] C. K. Catchpole and P. J. B. Slater. *Bird Song: Biological Themes and Variations*. Cambridge University Press, 2008.
- [2] J. Shonfield and E. M. Bayne. Autonomous recording units in avian ecological research: Current use and future applications. *Avian Conservation and Ecology*, Vol. 12, 14, 2017.
- [3] A. Farina and S. H. Gage. *Ecoacoustics: The Ecological Role of Sounds*. John Wiley and Sons, 2017.
- [4] B. L Krause, et al. The niche hypothesis: a virtual symphony of animal sounds, the origins of musical expression and the health of habitats. *The Soundscape Newsletter*, Vol. 6, pp. 6–10, 1993.
- [5] R. Suzuki and M. L. Cody. Complex systems approaches to temporal soundscape partitioning in bird communities as a self-organizing phenomenon based on behavioral plasticity. *Artificial Life and Robotics*, Vol. 24, No. 4, pp. 439–444, 2019.
- [6] K. Nakadai, H. G. Okuno, and T. Mizumoto. Development, Deployment and Applications of Robot Audition Open Source Software HARK. *Journal of Robotics and Mechatronics*, Vol. 27, pp. 16–25, 2017.
- [7] K. Nakadai, T. Lourens, H. Okuno, and H. Kitano. Active audition for humanoid. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence*, pp. 832–839, 2000.
- [8] R. Suzuki, S. Matsubayashi, R. W. Hedley, K. Nakadai, and H.G. Okuno. HARKBird: Exploring acoustic interactions in bird communities using a microphone array. *Journal of Robotics and Mechatronics*, Vol. 27, pp. 213–223, 2017.
- [9] D. Blumstein, et al. Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus. *Journal of Applied Ecology*, Vol. 48, pp. 758–767, 2011.
- [10] T. A. Rhinehart, L. M. Chronister, T. Devlin, and J. Kitzes. Acoustic localization of terrestrial wildlife: Current practices and future opportunities. *Ecology and Evolution*, Vol. 10, No. 13, pp. 6794–6818, 2020.
- [11] 炭谷晋司, 大和祐介, 鈴木麗璽, 小島諒介, 有田隆也, 中臺一博, 奥乃博. 野外での鳥類鳴き声観測のための web ベース録音ユニットと可視化ツールの試作. 第 39 回日本ロボット学会学術講演会予稿集, 2D4-03, 2021.
- [12] R. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, Vol. 34, pp. 276–280, 1986.
- [13] H. Nakaajima, K. Nakadai, Y. Hasegawa, and H. Tsujino. Blind source separation with parameter-free adaptive step-size method for robot audition. *IEEE Transactions on Audio, Speech and Language Processing*, Vol. 18, pp. 1476–1485, 2010.
- [14] R. Suzuki, S. Sumitani, Naren, S. Matsubayashi, T. Arita, K. Nakadai, and H. G. Okuno. Field observations of ecoacoustic dynamics of a japanese bush warbler using an open-source software for robot audition hark. *Journal of Ecoacoustics*, Vol. 2, EYAJ46, 2018.
- [15] R. Suzuki, S. Matsubayashi, F. Saito, T. Murate, T. Masuda, K. Yamamoto, R. Kojima, K. Nakadai, and H. G. Okuno. A spatiotemporal analysis of acoustic interactions between great reed warblers (*acrocephalus arundinaceus*) using microphone arrays and robot audition software hark. *Ecology and Evolution*, Vol. 8, pp. 812–825, 2018.
- [16] S. Sumitani, R. Suzuki, S. Matsubayashi, T. Arita, K. Nakadai, and H. G. Okuno. Fine - scale observations of spatio - spectro - temporal dynamics of bird vocalizations using robot audition techniques. *Remote Sensing in Ecology and Conservation*, Vol. 7, pp. 18–35, 2020.

- [17] Reiji Suzuki, Shinji Sumitani, Zachary Harlow, Shiho Matsubayashi, Takaya Arita, Kazuhiro Nakadai, and Hiroshi G. Okuno. Extracting bird vocalizations from a complex natural soundscape in forests using robot audition techniques. In *Proceedings of 2023 IEEE/SICE International Symposium on System Integrations (SII2023)*, pp. 728–733, 2023.
- [18] S. Matsubayashi, K. Nakadai, R. Suzuki, T. Ura, M. Hasebe, and H. G. Okuno. Monitoring the courtship flight trajectory of latham’s snipe (*gallinago hardwickii*) using microphone arrays. *Frontiers in Robotics and AI*, Vol. 9, 854572, 2022.
- [19] S. Matsubayashi, H. Osaka, R. Suzuki, K. Nakadai, and H. G. Okuno. Monitoring the courtship flight trajectory of latham’s snipe (*gallinago hardwickii*) using microphone arrays. *Ecology and Evolution*, Vol. 13, e9938, 2023.
- [20] 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.
- [21] L. F. Gill, W. Goymann, A. Ter Maat, and M. Gahr. Patterns of call communication between group-housed zebra finches change during the breeding cycle. *Elife*, Vol. 4, e07770, 2015.
- [22] H. Zhao, R. Suzuki, S. Sumitani, S. Matsubayashi, T. Arita, K. Nakadai, and H. G. Okuno. Visualization and quantification of activities of animal vocalizations in forest species using robot audition techniques. *Journal of Ecoacoustics*, Vol. 7, No. 1, 2, 2023.
- [23] P. J. Hart, R. Hall, W. Ray, A. Beck, and J. Zook. Cicadas impact bird communication in a noisy tropical rainforest. *Behavioral Ecology*, Vol. 26, pp. 839–842, 2015.
- [24] H. Zhao, R. Suzuki, R. Kojima, T. Arita, and K. Nakadai. A soundscape analysis of bird and cicada vocalizations based on azimuth and elevation localization using robot audition and machine learning techniques. In *Proceedings of the 28th International Symposium on Artificial Life and Robotics (AROB 2023)*, pp. 449–454, 2023.
- [25] 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*, Vol. 11, pp. 868–873, 2011.
- [26] D. Stowell and M. D. Wood and H. Pamula and Y. Stylianou and H. Glotin. Automatic acoustic detection of birds through deep learning: The first bird audio detection challenge. *Methods in Ecology and Evolution*, Vol. 10, No. 3, pp. 368–380, 2019.
- [27] D. Stowell. Computational bioacoustics with deep learning: a review and roadmap. *PeerJ*, Vol. 10, e13152, 2022.
- [28] D. Priebe, B. Ghani, and D. Stowell. Efficient speech detection in environmental audio using acoustic recognition and knowledge distillation. *Sensors*, Vol. 2024, No. 24, 2046, 2024.
- [29] J. G. Arriaga, M. L. Cody, E. E. Vallejo, and C. E. Taylor. Bird-db: A database for annotated bird song sequences. *Ecological Informatics*, Vol. 27, pp. 21–25, 2015.
- [30] S. Kahl, C. M Wood, M. Eibl, and H. Klinck. Birdnet: A deep learning solution for avian diversity monitoring. *Ecological Informatics*, Vol. 61, 101236, 2021.
- [31] H. Zhao, R. Suzuki, R. Kojima, T. Arita, and K. Nakadai. Utilizing embedding methods for soundscape analysis of forest animal vocalization based on azimuth and elevation localization. 第64回人工知能学会 AI チャレンジ研究会資料, in press, 2024.
- [32] 山本遼, 西田健次, 糸山克寿, 松林志保, 鈴木麗璽, 中臺一博. ロボット聴覚用音響処理ソフトウェア hark を用いたサウンドスケープの解析. 日本鳥学会 2022 年度大会講演予稿集, p. 109, 2022.
- [33] R. Suzuki, K. Hayashi, H. Osaka, S. Matsubayashi, T. Arita, K. Nakadai, and H. G. Okuno. Estimating the soundscape structure and dynamics of forest bird vocalizations in an azimuth-elevation space using a microphone array. *Applied Sciences*, Vol. 13, No. 6, 3607, 2023.
- [34] 鈴木麗璽, 古山諒, Zachary Harlow, 中臺一博, 有田隆也. 生成モデルに基づく形質表現を利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. 第63回 AI チャレンジ研究会予稿集 (SIG-Challenge-063-07), pp. 31–37, 2023.
- [35] B. Ghani, T. Denton, S. Kahl, et al. Global bird-song embeddings enable superior transfer learning for bioacoustic classification. *Scientific Reports*, Vol. 13, 22876, 2023.
- [36] T. Sainburg and T. Q. Genter. Toward a computational neuroethology of vocal communication: From bioacoustics to neurophysiology, emerging tools and future directions. *Frontiers in Behavioral Neuroscience*, Vol. 15, 811737, 2021.
- [37] G. Beguš and A. Leban and S. Gero. Approaching an unknown communication system by latent space exploration and causal inference. *arXiv e-print*, arXiv:2303.10931, 2023.
- [38] R. Suzuki and T. Arita. An evolutionary model of personality traits related to cooperative behavior using a large language model. *Scientific Reports*, Vol. 14, 5989, 2024.
- [39] R. Suzuki, S. Sumitani, C. Ikeda, and T. Arita. A modeling and experimental framework for understanding evolutionary and ecological roles of acoustic behavior using a generative model. In *Proceedings of ALIFE 2022*, No. isal.a.00542, 58, 2022.
- [40] X. Liu, Q. Kong, Y. Zhao, H. Liu, Y. Yuan, Y. Liu, R. Xia, Y. Wang, M. D. Plumbley, and W. Wang. Separate anything you describe, 2023. *arXiv e-print*, arXiv:2308.05037, 2023.