

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

A playback experiment on songbirds using simulated vocalizations based on a generative model

炭谷晋司^{1*} 鈴木麗瑩¹ 有田隆也¹ 和多和宏² 松林志保³ 中臺一博^{4,5} 奥乃博⁶
Shinji Sumitani¹ Reiji Suzuki¹ Takaya Arita¹ Kazuhiro Wada²
Shiho Matsubayashi³ Kazuhiro Nakadai^{3,4} Hiroshi G. Okuno⁵

¹ 名古屋大学, Nagoya University

² 北海道大学, Hokkaido University

³ 大阪大学, Osaka University

⁴ 東京工業大学, Tokyo Institute of Technology

⁵ (株) ホンダ・リサーチ・インスティテュート・ジャパン, Honda Research Institute Japan

⁶ 京都大学, Kyoto University

Abstract: 本稿では、鳥類の音声コミュニケーションにおける個体間相互作用の観測・理解を目的として行ったマイクアレイを用いた音源定位実験について報告する。具体的には、空間的に限定された屋外環境としてテントを用い、その中にキンカチョウ (*Taeniopygia guttata*) を放ち、複数のマイクアレイを用いて録音を行い、ロボット聴覚オープンソースソフトウェア HARK で音源の定位・分離を行うことでキンカチョウが鳴き合う様子の観測を試みた。生成モデルの1つである VAE (Variational Autoencoder) と SVM (Support Vector Machine) を組み合わせた鳴き声の個体・種類の分類に関する取り組みと、分類結果と鳴き声の定位結果に基づく個体間相互作用に関する調査結果から確認された個体数の変化によって生じる歌・地鳴きの発声パターンの違いについて報告する。

1 はじめに

鳥類にとって、鳴き声は重要な意思伝達手段である。様々な発声を用いて他の個体と多様なコミュニケーションをとることが知られているが、その際に用いられる鳴き声の種類は大きく分けて「歌」と「地鳴き」の2つに分類される。歌は比較的長く複雑な発声で、縄張りの防衛や雌への求愛などに用いられ [1]、地鳴きは比較的短く単純な鳴き声で、捕食者への警告や社会的結合を形成するための信号など、より具体的な情報を交換するために用いられる [2]。鳥類の発声行動には、鳴くタイミングに関する個体間の時間的相互作用をはじめとして、個体ごとのなわばりの関係などの空間的相互作用、種間・種内それぞれに異なる歌の種類・周波数などの性質を持ち、互いに作用しあう音響的相互作用といった様々な次元での相互作用を持つ。

鳥類の鳴き声における相互作用の理解は、様々な鳥類で調査されているが、キンカチョウ (*Taeniopygia gut-*

tata) も研究が盛んな種の1つである。キンカチョウは、歌の音声学習におけるモデル生物 [3] として調査されたり、地鳴きは集団内における社会的結合の形成や維持に関する調査対象 [4] として扱われるなど、鳴き声に関する研究事例は多くある。ここでキンカチョウの個体間相互作用に関する研究事例についていくつか例を挙げる。Gill らは、キンカチョウを対象として、ペアの形成から孵化までの繁殖期間における個体間相互作用について、実験個体に背負させた小型のマイクを用いて録音を行い調査した [5]。録音の分析結果からは、繁殖段階に応じてペア間で用いる地鳴きの種類が変化することが示唆された。また、Ikebuchi らは、ペアの形成に関してオス、メスを2羽ずつケージ内に入れ、録音とビデオ撮影による観測によって個体間の接近、発声の分析を行った [6]。その結果、オス個体は、ケージに入れられた直後に他のオス個体へ攻撃的な姿勢を取り、これによって集団内でのオスのヒエラルキーが形成されることが確認された。また、メスはオスの歌の質よりもこの社会的優位性によってオスを選好する傾向があることが示唆されている。

*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:sumitani.shinji@h.mbox.nagoya-u.ac.jp

のケージ内で飼育された個体であり、実験中以外はオスとメスを分けて別のケージで飼育した。オスとメスはそれぞれの発声は聞こえる状態ではあるが、視覚的に遮断された状態で飼育した。9時30分から13時30分の4時間の録音に関して分析を行い、比較した結果について報告する。なお、本動物実験は、北海道大学動物実験委員会のガイドラインと承認(18-0053)を得て実施している。

2.3 2次元定位

2次元定位は、三角測量の要領で、2台のそれぞれのマイクアレイから音源の定位方向へ直線を伸ばして交点ができた場合に定位位置とする方法で行うが、本実験では、先行研究結果[12]に基づいて場所に応じて適したマイクアレイのペアを選定した上で定位を行った。これは、音源の位置がペアのマイクアレイを結ぶ直線上に近いほど2次元定位は困難になることから、そのような位置での2次元定位を可能な限り避けることを実現している。図2の分割された場所ごと示されている数字のペアがその場所での2次元定位に利用したマイクアレイのペアである。2次元定位を行う前に、HARKBirdを用いてキンカチョウ個体の鳴き声をうまく定位するように定位のパラメータを適宜調整し、音源の定位・分離を行った。短時間フーリエ変換によって得られた各チャンネルのスペクトログラムからMUSIC法[15]を用いて音源定位を行い、その定位結果に基づいてGHDSS法[16]を用いて対応する音源を抽出した。また、2次元定位は以下の方法で行った。

1. 定位時間の重複するすべての各音源ペアに関して、HARKBirdで抽出した音源定位情報を用いて三角測量に基づく2次元定位を行う。
2. 2次元定位がなされた(交点が作られた)場合、交点ができた時間分だけ定位音源に対応する分離音を切り出す。
3. 切り出した音をグレースケール画像(128×128)に変換し、UMAP[17]を用いて2次元にデータの次元を削減、UMAPの特徴空間上で切り出した音源のペア間の距離が $d < 3$ である場合、その2次元定位情報を採用する。
4. 上記の処理を全てのマイクアレイのペアで行い、それぞれのペアが担当する定位範囲の2次元定位結果のみを抽出し、それらを結合する。これを最終的な2次元定位結果とする。

今回は、個体識別を別で行うため特徴空間上での音源ペア間の距離は大きく確保し、個体鳴き声の定位漏れ

が生じないようにした。また、採用する定位結果は中央のマイクアレイの位置を座標平面の原点として、縦軸、横軸でそれぞれ $[-4, 4]$ で限定した。

2.4 個体識別と歌・地鳴きの判別

複数個体を放鳥した場合の鳴き声および個体の識別は、VAE (Variational Autoencoder)[13]とSVM (Support Vector Machine)[14]を用いる。判別の流れを図3に示す。まず、1個体放鳥時の録音の定位結果を用いて、手作業で歌と地鳴きとそれ以外に分類する。今回、地鳴きはdistance callと呼ばれる個体で比較的特徴のある鳴き声のみを分類した。distance callは、飛行時や警戒時、求愛など様々な場面で発することが報告されており、ペアとなったオスとメスが結合を維持するためのものだと考えられている[18]。また、ペア形成の前や直後でdistance callを用いた相互作用が多いことも報告されており[5]、個体数条件の違いで発声に違いがみられることが期待される地鳴きである。分類したオス3羽、メス3羽の鳴き声の音源データをデータセットとして、VAEで学習する。各個体の鳴き声の分離音を時間幅1.0秒、シフト幅が地鳴きは0.1秒、歌は0.2秒で分割して複数の音声データを生成し、この音声データをPythonライブラリのmatplotlibを用いて周波数[1500,8000]の領域を128×128の画像データに変換する。VAEは、潜在変数の次元が32次元で、エンコーダ側が8層の畳み込み層と3層の全結合層となり、デコーダ側がエンコードの逆関数となるモデルを機械学習ライブラリであるPyTorchを用いて構築した。データセットの前処理としてtransformによるグレースケール化を施し、学習は、実験に用いたオス個体3羽の歌・地鳴き、メス個体3羽の地鳴きそれぞれ3000個の分割された画像データを用い、5000回の学習を行った。

画像データの分類は、学習後のVAEに未知データを入力して得られた32次元の潜在変数をSVM (Support Vector Machine)に入力することで行う。SVMの学習・適用はPythonライブラリのscikit-learnを用いて行い、パラメータはデフォルト値を使用した。まず、分類を行う実験の録音から得た定位・分離音源に対して学習データと同様の音声分割と画像データへの変換処理を行い、生成した未知の画像データと録音実験で用いた個体の鳴き声の学習データとを共にVAEに適用し、それぞれの潜在変数を得る。次に、ラベルが既知である学習データのみを教師データとしてSVMで学習し、学習したSVMモデルに残りの未知データを適用して分類結果を得る。各音源は画像データに変換する際に複数個のデータに分割されているため、それぞれの分類結果をもとに多数決をとり、もっとも多い分類結果を

音源の分類とした。また、未知データにはキンカチョウの鳴き声以外の音が多く含まれており、分類に大きく影響する。そのため、これらノイズの大部分をだまかに排除する方法として以下を行った。各未知データに対してその潜在変数からコサイン距離で直近の50個の学習データを取り出し、この距離の平均を未知データのノイズ度と定義する。ノイズ度が全未知データの平均値より大きい未知データをノイズとみなし、分類時の多数決には含めないようにした。また、分類結果は手直しによって誤ったラベルを正しいラベルになおし、これを最終的な結果とした。

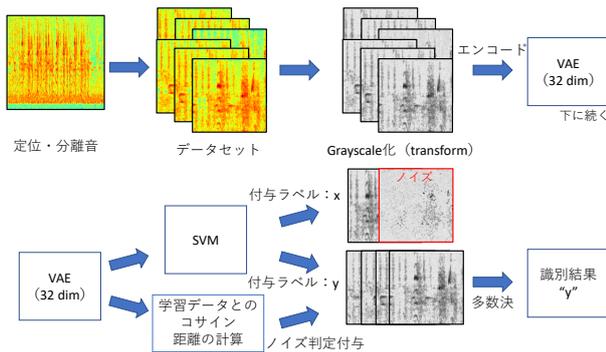
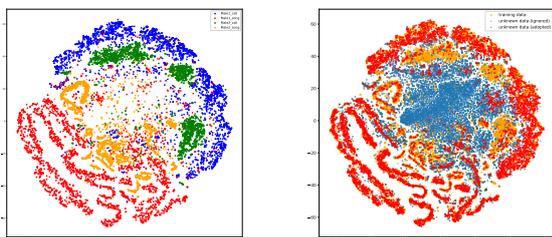


図 3: 鳴き声判別の流れ。

図 4 は、2羽のオス個体の鳴き声を VAE に適用後、t-SNE[19] によって 2 次元に次元を削減して可視化した結果を示す。図 4 (a) からは各個体の歌・地鳴きがうまく分かれて分布していることがわかる。また、図 4 (b) では、中央付近に学習データでは確認できない分布が存在している。これらのほとんどが歌・地鳴き (distance call) 以外の音であるが、提案手法により適切に除去されていることがわかる。



(a) 学習データ (b) ノイズ除去結果。

図 4: t-SNE による 2 次元での VAE 潜在空間の可視化。左図の各色は 2 羽のオス個体の歌と地鳴きを示す。右図の各色はそれぞれ橙: 学習データ, 赤: 未知データ (採用), 青: 未知データ (除外) を示す。

モデルの評価として、学習データには用いていない各ラベル 1000 個ずつのテストデータを SVM で分類を試みた結果、正しいラベルに分類されたものは最小で 861/1000 個 (Female2 の地鳴き)、最大で 963/1000 個 (Male1 の地鳴き) であった。この分類結果に基づいて多数決を行うことを考慮すると、高い精度での個体識別、歌・地鳴きの判別が可能であるといえる。

3 分析結果

3.1 1 羽放鳥実験

図 5 はオス 1 羽 (Male1) のみ放った場合の 2 次元定位結果をもとに各地点での歌・地鳴きの頻度分布を示したものである。この頻度分布は、手法で示した定位結果の採用範囲 $[-4, 4]$ を各軸で 10 分割している。歌と地鳴きの分布はある程度一致しているが、歌に関しては餌場 (地点 1) で鳴く割合が高いことが確認できる。また、鳴き声の頻度の高い地点は餌場と図中左下の止まり木 (地点 2) と巣 (地点 3) の 3 か所に集中している。

図 6 は実験 1 とは異なるオス個体 (Male2) を放った場合の 2 次元定位の頻度分布を示す。歌・地鳴きの分布は比較的類似しており、また歌の頻度が餌場 (地点 1) で高い。これらの傾向は Male1 と一致する。また、鳴き声の頻度の高い場所は餌場と図中左下の止まり木 (地点 2) の 2 か所に集中しており、これは Male1 と類似する。この結果から、オス個体のキンカチョウは 1 羽の場合、ある程度決まった場所で鳴き声を発する傾向があることが推測される。

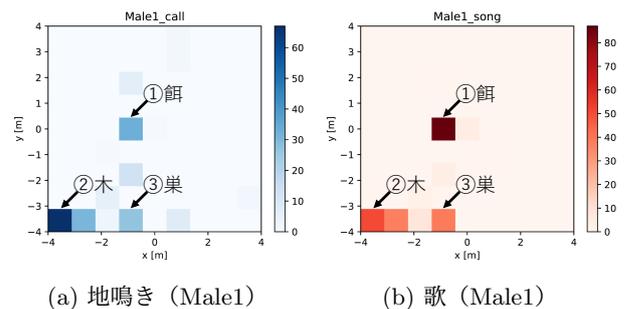


図 5: 鳴き声の頻度分布 (実験 1)。

3.2 2 羽放鳥実験

図 7 は、Male1 と Male2 の 2 羽をテント内に放鳥したときの 2 次元定位結果から得た鳴き声の頻度分布を示す。この結果からは、双方の個体で歌・地鳴きの分

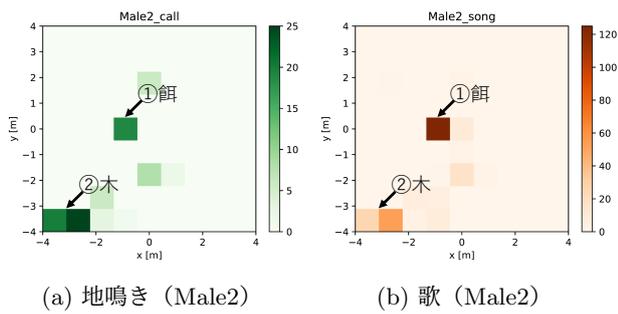


図 6: 鳴き声の頻度分布 (実験 2) .

布が比較的一致していることが確認できる。これは、1羽条件と同様の傾向にある。一方で、餌場（地点1）で鳴く傾向は2羽ともに減少している。また、2羽の分布は一方の個体のみが鳴いている地点もあるが、分布が重なる箇所もあり、特に図中中央下にある止まり木（地点2）においてその傾向が確認できる。このことから、2羽の個体は止まり木などの場所を共有しながら鳴いていることが示唆される。

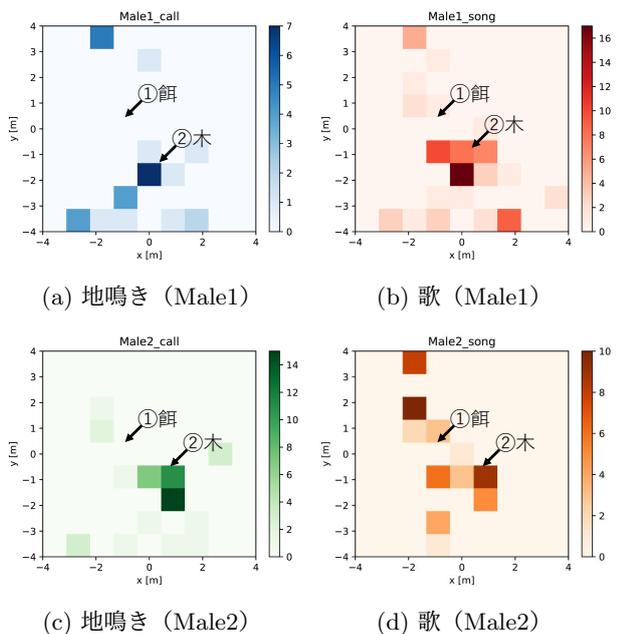


図 7: 鳴き声の頻度分布 (実験 3) .

3.3 3羽放鳥実験

図 8 は、オス 2 羽 (Male1, Male2), メス 1 羽 (Female1) をテント内に放鳥したときの 2次元定位結果から得られた鳴き声の頻度分布を示す。オス 2 羽の条件

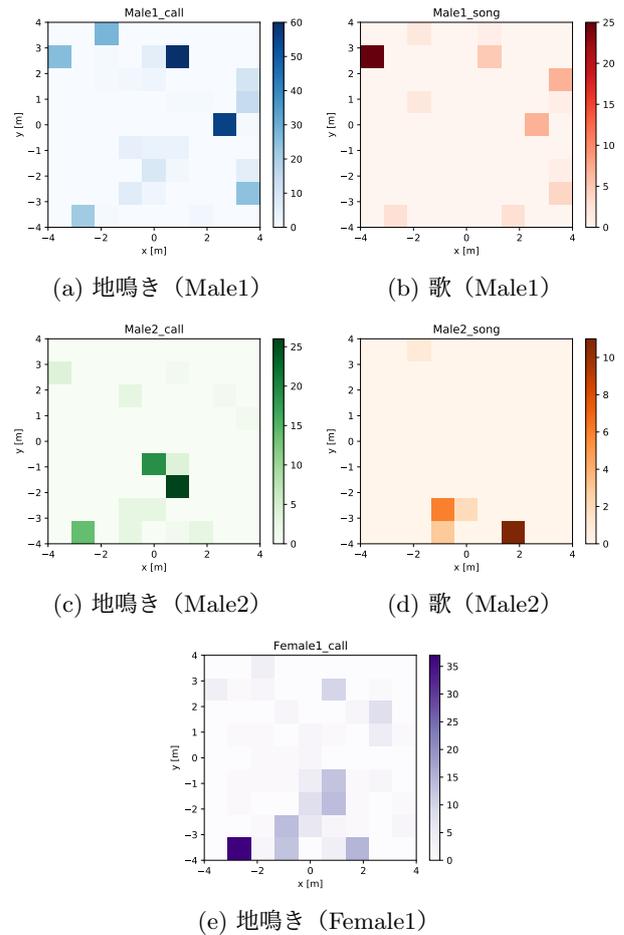


図 8: 鳴き声の頻度分布 (実験 4) .

と比較すると、Male1 の鳴き声の分布はテント全体に広がっており、Male2 の鳴き声はテントの図中下半分で多く確認できる。さらに、頻度の高い地点だけ注目してみると、Male1, Male2 でそれぞれ図中の上半分・下半分で明確に分かれており、メス個体の居ない 2羽の条件とは傾向が大きく変化していることが確認できる。また、Female1 の地鳴きはテント全体に散らばっているが、概ね下半分に集中しており、これらの多くは Male2 の鳴き声の分布と重なっている。

3.4 発声頻度の比較

次に、条件の違いによるオス 2 羽の鳴き声の発声頻度の変化を調査した。表 1 は、各条件でのオス 2 羽の歌・地鳴きの定位回数をまとめたものである。それぞれの条件を比較してみると、地鳴きは個体で回数に差はあるものの、1羽条件と3羽条件で多く、2羽条件で少ない傾向があった。キンカチョウは、野生環境では群れで生活する動物であり、実験個体も複数個体と同

じケージで飼育されている。このことから、1羽条件は普段と異なる状況であり、自身の存在の主張あるいは他個体への呼びかけとして多くの地鳴きを発声した状況であったことが推測される。また、2羽条件の傾向に関しては2つの考察ができる。1つは、2羽条件は親しみのあるオス個体と一緒にいる状況であり、1羽の状況から脱して多くの呼びかけの必要が無くなったことにより地鳴きが減少したことである。もう1つは、distance callにはペア間での確認に用いられているという報告や、ペアの形成前や直後で多く用いられそれは繁殖段階が進むにつれ減少していく報告がある。オス同士においても同様の意味があると仮定すれば、結合の強いオス同士が時折地鳴きを用いて結合の確認を行っている状況だと考察できる。一方で、3羽条件ではMale1で地鳴きが非常に多い。そこで、それぞれの個体の鳴き声に関して他個体の発声がどの程度重複しているか調査を行った。その結果、Male2の歌に対するMale1の地鳴きの重複が多くみられ、これはMale1がMale2の歌の発声に対して地鳴きによる阻害を行っているようにも見える。Distance callは様々な状況で用いられているとすれば、そのような利用方法も十分に考えられる。

一方で、歌は1羽条件で最も多く用いられ、2羽条件、3羽条件という順で回数は減る傾向があった。キンカチョウは、メスに向けた歌(directed song)とそれ以外の歌(undirected song)を区別して用いることが報告されている[20]。また自分の歌を維持するために、聴覚的なフィードバックが必要であることが報告されている[21]。以上のことから、1羽条件で歌が多いのは、undirected songを用いた自身の歌の維持によるものであると考えられる。また、2羽条件で1羽条件よりも歌の回数が減ったのは、聴覚的なフィードバックが必要という点で他個体が鳴いていないタイミングを計って発声するためであることが要因の一つだったと考えられる。最後に、3羽条件での歌が最も頻度が少ない理由が2つ考えられる。1つには、2羽条件と同様の理由であり、メス個体の導入によって2羽条件よりも発声のタイミングが失われたということである。もう1つは、キンカチョウのペア形成においては、質の高い歌が好まれる傾向があることも報告されており[22]、発声回数を増やすよりも質の高い歌を歌うために自ずと回数が減ったということが考えられる。

3.5 発声と個体間距離の関係

2次元定位の結果から、個体間で共通する発声分布があるなど個体の位置と発声に関係があることが示唆されたので、それらの関係についてより詳細な分析を行った。図9は、それぞれの個体の組み合わせについて、

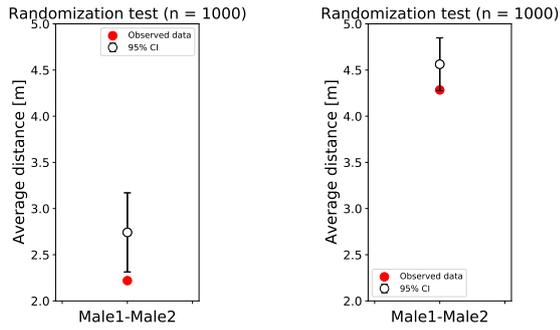
表 1: 各条件の Male1, Male2 の歌・地鳴きの定位回数.

		1羽条件	2羽条件	3羽条件
Male1	地鳴き	203	29	302
	歌	240	79	61
Male2	地鳴き	90	48	86
	歌	269	52	23

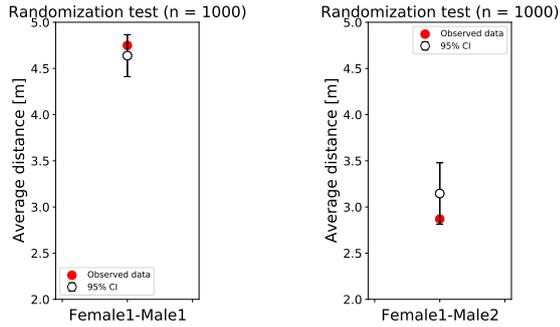
一方が発声した場合にもう一方が前回の発声場所にいると仮定したとき、それぞれの定位位置間の距離を求めて平均を出したものである。赤点は観測値で、95%信頼区間は各個体に関して定位位置をランダムに入れ替えて平均距離を求めたランダムイゼーションテスト($n = 1000$)の信頼区間である。Male1とMale2との距離間は、実験3で約2.3m、実験4で約4.3mと大きく異なっている。また、実験3では観測値が信頼区間より下回っている。この要因として、一方の鳴き声が他方を呼び寄せる役割があることが推測される。また、個体間の距離を結合の強さと仮定すれば、2羽実験においてはオス個体同士は強い結合を持っていたが、メスの投入によってその結合が弱まったということが考えられる。実験4に関して、2次元定位の頻度分布の結果からも確認できたように、Male1はMale2、Female1どちらも距離を取って鳴いている傾向が確認できる。Male2とFemale1は、観測値こそランダム化した場合の信頼区間内ではあるが、その他の組み合わせと比較するとFemale2との距離は短く、比較的近くで鳴き合っていることが示され、このペアはMale1とFemale2のペアよりも結合の強い、親密な関係であることが示唆される。

次に、オス個体の発声と個体間の距離に大きく差が見られたので、歌と地鳴きで違いがあるかを調査した。図10は、それぞれのオス個体に注目し歌と地鳴きそれぞれで鳴き声を発した場合の他方のオス個体との距離の関係を示したものである。歌、地鳴きともに実験4は実験3と比較して平均距離が大きくなっていることが確認できる。特に、歌はオス2羽の条件では近くでも多く歌われているが、メスを入れた実験では他方のオスの近くで歌うことが稀になっている。キンカチョウのオス個体は、ペアを形成する際に個体間のヒエラルキーを確立し、噛みつくなどの攻撃的な行動を行うことが報告されている[6]。これを考慮すると、観測された状況はメス個体をめぐって敵対し合う他のオスの近辺では容易に歌が歌えない状況であったことが1つの仮定として考えられる。

地鳴きに関しては平均距離は大きくなったものの、依然として近い距離でも用いていることが確認できる。特に、Male2に関しては最もMale1と近い距離で用いた地鳴きの頻度が最も高い。地鳴きが社会的結合の確認



(a) Male1-Male2 (実験 3) (b) Male1-Male2 (実験 4)



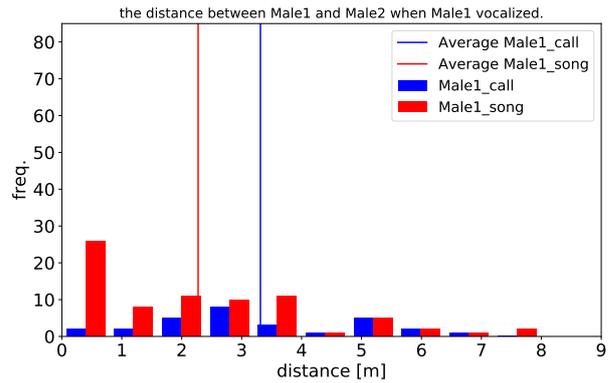
(c) Female1-Male1 (実験 4) (d) Female1-Male2 (実験 4)

図 9: 発声時の個体間の平均距離とランダムイゼーションテスト ($n = 1000$) による 95%信頼区間。

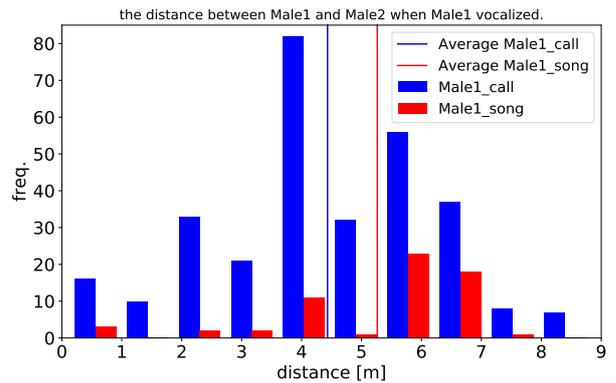
の際に用いられることを考慮すれば、敵対する状況においてもそれぞれの結合を確認し合うためにある程度の近距離においても地鳴きは発せられている状況ではないかと考えられる。あるいは、distance call は様々な用途があるということから、警戒の意味として用いたという可能性もある。一方で、この結果はオス間の関係を示したものであり、メス個体の状況が考慮されていない。そのため、これらの結果はメス個体の影響を受けた結果である可能性もある。

4 おわりに

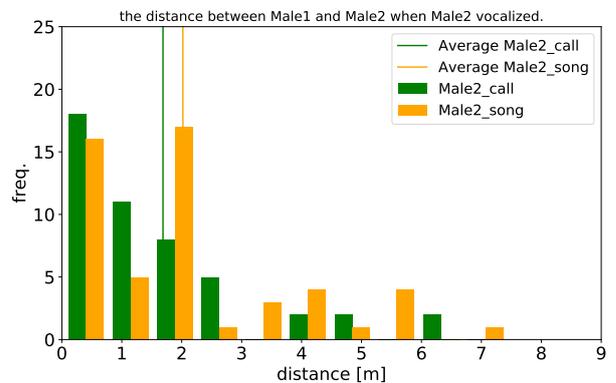
鳥類の鳴き声に基づく個体間相互作用の観測・理解のためにマイクアレイを用いてキンカチョウが鳴く様子を様々な条件で観測を試みた。提案した個体識別手法は、個体の鳴き声の特徴をうまく学習し、実際の分析にも適用ができた。また、分析結果は、状況に応じて挙動を変えている様子の観測に成功した。具体的には、キンカチョウ個体がオス 1 羽では比較的定位置で鳴く傾向があること、オス 2 羽では互いに止まり木などを共有しながら接近して鳴く傾向があること、メス 1 羽を入れるとオス個体が互いに距離を取りつつ鳴き合う傾向があり、歌と地鳴きで分布に差が生じること



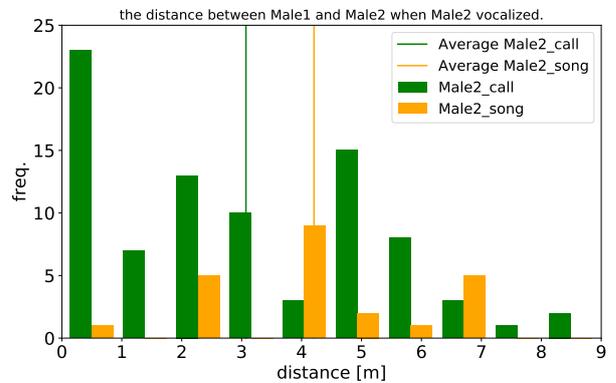
(a) Male1 (実験 3)



(b) Male1 (実験 4)



(c) Male2 (実験 3)



(d) Male2 (実験 4)

図 10: オス個体発声時の他方オス個体との距離の分布。

が確認できた。実際、実験を目視で確認した際に、2羽条件では一方の移動後にはもう一方が続いて移動したり、3羽条件では一方のオス個体が近づいた際にオス同士が威嚇をして鳴き合っている状況は多くあった。これは、ロボット聴覚技術によって実際の個体間で生じた相互作用が定量的に抽出できたことを意味する。

一方で、本実験は多くの課題が残る。例えば、地鳴きは distance call のみを抽出しており、その他の地鳴きに関する調査が不十分である。キンカチョウは、様々な地鳴きを介して社会的結合の確認を行うことが知られており [5, 23], 他の地鳴きのみで生じる相互作用の存在も考えられる。この意味で、今回の手法ではキンカチョウの個体間相互作用の理解には限界がある。歌に関しても directed song と undirected song の区別をしていない。条件の違いによる個体の挙動の変化をより詳細に調査するには、それぞれの条件でどちらの歌がどのように歌われたかを調査する必要がある。

今後の課題としては、上記の精査に加えて、個体数を増やしたより複雑な状況での観測、実際の自然環境における提案観測手法の適用可能性の検討が挙げられる。

謝辞

柴田ゆき野氏 (北海道大学) の実験協力で謝意を表す。本研究の一部は JSPS 科研費 JP21K12058, JP20J13695, JP20H00475, JP19KK0260, JP17H06383 (#4903) 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] P. Marler. Chapter 5 - bird calls: a cornucopia for communication. In Peter Marler and Hans Slabbekoorn, editors, *Nature's Music*, pp. 132 – 177. Academic Press, San Diego, 2004.
- [3] L. A. Eales. Song learning in zebra finches: some effects of song model availability on what is learnt and when. *Animal Behaviour*, Vol. 33, No. 4, pp. 1293 – 1300, 1985.
- [4] D. Stowell, L. Gill, and D. Clayton. Detailed temporal structure of communication networks in groups of songbirds. *Journal of The Royal Society Interface*, Vol. 13, No. 119, 2016. Art. no. 20160296.
- [5] 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, , 2015. Art. no. e07770.
- [6] Maki Ikebuchi and Kazuo Okanoya. Growth of pair bonding in zebra finches: physical and social factors. *Ornithological Science*, Vol. 5, pp. 65–75, 2006.
- [7] 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.
- [8] R. Suzuki, S. Matsubayashi, 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] S. Sumitani, R. Suzuki, S. Matsubayashi, K. Arita, T. Nakadai, and H. G. Okuno. An integrated framework for field recording, localization, classification and annotation of birdsongs using robot audition techniques - harkbird 2.0. In *Proceedings of ICASSP 2019*, pp. 8246–8250, 2019.
- [10] R. Suzuki, S. Matsubayashi, F. Saito, T. Murate, T. Masuda, Y. 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.
- [11] 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, , 2018. Art. no. EYAJ46.
- [12] Shinji Sumitani, Reiji Suzuki, Takaya Arita, Kazuhiro Nakadai, and Hiroshi Okuno. Non-invasive monitoring of the spatio-temporal dynamics of vocalizations among songbirds in a semi free-flight environment using robot audition techniques. *Birds*, Vol. 2, pp. 158–172, 2021.
- [13] D. P. Kingma and M. Welling. Auto-encoding variational bayes. *arXiv preprint*, 2013. Art. no. arXiv:1312.6114.
- [14] C. Cortes and V. Vapnik. Support vector networks. *Machine Learning*, Vol. 20, pp. 273–297, 1995.
- [15] R. Schmidt. Bayesian nonparametrics for microphone array processing. *IEEE Transactions on Antennas and Propagation (TAP)*, Vol. 34, No. 3, pp. 276–280, 1986.
- [16] H. Nakajima, 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.
- [17] L. McInnes and J. Healy. UMAP: Uniform manifold approximation and projection for dimension reduction. ArXiv e-prints 1802.03426, 2018.
- [18] Richard Zann. Structural variation in the zebra finch distance call. *Zeitschrift für Tierpsychologie*, Vol. 66, No. 4, pp. 328–345, 1984.
- [19] L. van der Maaten and G. Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, Vol. 9, pp. 2579–2605, 2008.
- [20] Richard A Zann. *The zebra finch: a synthesis of field and laboratory studies*, Vol. 5. Oxford University Press, 1996.
- [21] Gerald E. Hough and Susan F. Volman. Short-term and long-term effects of vocal distortion on song maintenance in zebra finches. *Journal of Neuroscience*, Vol. 22, No. 3, pp. 1177–1186, 2002.

- [22] Michelle L. Tomaszycki and Elizabeth Adkins-Regan. Experimental alteration of male song quality and output affects female mate choice and pair bond formation in zebra finches. *Animal Behaviour*, Vol. 70, No. 4, pp. 785–794, 2005.
- [23] Pietro B.D’Amelio, Lisa Trost, Andries ter Maat. Vocal exchanges during pair formation and maintenance in the zebra finch (*taeniopygia guttata*). *Frontiers in Zoology*, Vol. 14, No. 13, 2017.