

鳥類の音声コミュニケーションに関する  
時空間的ダイナミクスの理解に対する  
ロボット聴覚アプローチ

炭谷 晋司

# 概要

本論文は、ロボットの環境認識のために、多チャンネルのマイクロホンで構成されるマイクロホンアレイを用いた音源の方向推定（定位）や抽出（分離）等の音声信号処理を実時間で行う「ロボット聴覚技術」を活用した、鳥類の鳴き声及び鳥類集団内の音声コミュニケーションの観測と分析に関する取り組みをまとめたものであり、全6章からなる。

序論となる1章では、まず背景として鳥類における音声コミュニケーションの役割について述べた後、相互作用する鳥類集団を複雑系として捉える考え方や、生物由来の音と環境との関わりの理解を目指す生態音響学と鳥類音声との関係について概説する。次に、主に個体間相互作用理解の際に生じる従来の鳥類音声観測手法の抱える問題として、空間的情報や複数の音情報の抽出が困難であることを挙げる。これらを踏まえ、本研究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理解に焦点を合わせた、ロボット聴覚技術を用いた観測・分析手法の実現を目的とする。特に、鳥類の積極的な相互作用が観測可能な実験環境の構築、実験条件に応じた分析手法の構築と生態データの検討、抽出音源の分類・統合手法の確立に注目したシステムの構築と3つの観測実験について報告する。

2章では、利用するロボット聴覚オープンソースソフトウェアであるHARKの基本的な仕組みを述べた上で、提案する観測手法（音源定位・音源分離等）について説明する。その後、HARKを利用し鳥類音声観測用に音声信号処理を容易に利用可能にしたソフトウェアであるHARKBirdの機能や用途について述べる。

3章では、野生の鳥類個体が積極的に動き、鳴く様子を観測できるように同種の鳴き声を再生するプレイバックによって疑似的な鳴き合う状況を構築できる実験の枠組みを構築した。その評価として、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ウグイス (*Horornis diphone*) に対してプレイバック実験を行い、その影響を単一マイクロホンアレイによる音源の方向情報を用いた定量的分析によって調査した。その結果、プレイバックの条件によって対象個体に与える影響が異なることを定量的に示した。具体的には、再生頻度が高いほど対象個体はプレイバックに対して強い警戒を示す傾向がある一方、対象個

体の発声後にプレイバックを行うインタラクティブ実験では再生頻度が低いにもかかわらず強い影響を与える傾向が観測された。これにより、提案手法によるプレイバック・インタラクティブ実験の生態観測への意義が示された。さらに、定位方向に基づいた発声と移動に関する調査では、遠方への移動と見なせる音源定位方向の大きな変化があった際の発声には警戒を示す歌が用いられる傾向も観測された。

4章では、3章のウグイスに対する実験の発展として、より詳細な挙動の変化の抽出と提案システムのより多様な環境での適用を目指し、米国加州の森林に生息するホシワキアカトウヒチョウ (*Pipilo maculatus*) に対するプレイバック実験を題材とした2台のマイクロホンアレイによる音源の2次元定位と音源の分類に向けた検討を行った。本実験では、次元削減アルゴリズム (t-SNE) を用いた定位音源の分類支援手法と半自動的な音源定位手法を提案し、鳥類生態理解と音風景理解という異なるスケールでの鳴き声の観測を試みた。その結果、提案した分類支援手法は音源の分類における時間的コストの削減に貢献し、さらに音風景の概観にもうまく適用できた。プレイバック実験の2次元定位結果からは、対象個体がプレイバック音の種類によって発声頻度や種類、スピーカへの接近傾向などを変えている状況を詳細に把握でき、生態学的に妥当な時空間的ダイナミクスの抽出が可能であることも示された。

5章では、多数個体が鳴き合うより複雑な状況での生態理解のための観測手法の提案として、北海道大学札幌キャンパス圃場の屋外テント内に構築した鳥類が自由に飛び回れる環境での複数のキンカチョウ (*Taeniopygia guttata*) 個体の鳴き声の観測を試みた。多数個体集団内の発声の様子を抽出を目的とした実験では、動く複数の同種鳥類の発声が複数マイクロホンアレイに基づく観測手法によってどの程度詳細に観測可能かを検証した。本研究で提案した、マイクロホンアレイごとに定位担当範囲を割り当てる2次元定位手法は、止まり木や巣などのキンカチョウが鳴く位置の詳細な把握に貢献した。また、定位結果と定位音源の音響特性情報の組み合わせからはキンカチョウ集団の鳴き声に関する時空間的ダイナミクスの抽出に成功した。社会的関係に焦点を合わせた実験では、VAEとSVMに基づく鳴き声の分類手法の提案と雌雄の数を変えた録音実験を行うことで、社会状況の変化によって生じる個体間関係の変化の抽出を試みた。提案した分類手法は、対象個体の鳴き声の特徴を把握して、ノイズの除去と鳴き声に基づく個体識別に貢献した。また、観測された結果からは、個体の組み合わせによって個体間距離や発声頻度に差が生じ、状況に応じて結合の度合いを変化させるオス個体の挙動の観測に成功した。

6章では、結論として上記の実験で得られた知見をまとめ、一連の提案の枠組みは個体の積極的な相互作用を観測可能であること、マイクロホンアレイの使用台数・音源定位情報の利用方法に応じた粒度の異なる観測データの取得が可能であること、これらが観測目標・観測環境に応じた分類手法を適切に選ぶことで実現できることを示した。以上から、本研究が鳥類の音声コミュニケーションにおけるロボット聴覚技術を用いた観測・分析手法の発展に貢献したと結論づけた。

# 目次

<b>第1章</b>	<b>序論</b>	<b>11</b>
1.1	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	11
1.2	鳥類音声観測技術	12
1.3	ロボット聴覚技術	14
1.4	研究目的	15
1.5	論文の構成	17
<b>第2章</b>	<b>HARK・HARKBird</b>	<b>19</b>
2.1	HARK	19
2.1.1	概要	19
2.1.2	音源定位	19
2.1.3	音源分離	21
2.2	HARKBird	21
<b>第3章</b>	<b>ウグイスに対するプレイバック実験における鳴き声の方向分布分析</b>	<b>26</b>
3.1	背景と目的	26
3.2	実験手法	27
3.2.1	ウグイス ( <i>Horornis diphone</i> )	27
3.2.2	HARKBird のリアルタイム録音・プレイバック (インタラクティブ実験) 機能の拡張	28
3.2.3	プレイバック実験とインタラクティブ実験	29
3.3	分析結果	32
3.3.1	歌の方向分布分析	32
3.3.2	プレイバックの影響の定量的分析	35
3.3.3	統計分析	38
3.4	議論	41

第4章	ホシワキアカトウヒチョウに対するプレイバック実験における鳴き声の時空間的分析	43
4.1	背景と目的	43
4.2	実験手法	45
4.2.1	ホシワキアカトウヒチョウ ( <i>Pipilo maculatus</i> )	45
4.2.2	実験環境	45
4.2.3	プレイバック実験	50
4.2.4	鳴き声の2次元定位	50
4.2.5	分析	56
4.3	分析結果	57
4.3.1	t-SNEの低次元特徴量空間を用いた分類手法の評価	57
4.3.2	鳴き声に関する時空間ダイナミクスの分析	59
4.3.3	音源の類似度に基づいた2次元定位による音風景の観測	60
4.4	議論	64
第5章	複数のマイクロホンアレイを用いたキンカチョウ集団における個体間相互作用の観測	68
5.1	背景と目的	68
5.2	キンカチョウ	70
5.2.1	生態及び鳴き声	70
5.2.2	先行研究	71
5.3	小規模鳥類集団における鳴き声の時空間ダイナミクスの抽出と分析	72
5.3.1	実験手法	72
5.3.2	結果	76
5.4	個体間の社会的関係に焦点を合わせた録音実験	86
5.4.1	実験手法	86
5.5	分析結果	91
5.5.1	1羽放鳥実験	91
5.5.2	2羽放鳥実験	94
5.5.3	3羽放鳥実験	98
5.5.4	発声頻度の比較	98
5.5.5	発声と個体間距離の関係	101

5.6 議論 . . . . .	104
第 6 章 結論	107
謝辞	111
参考文献	111

# 目次

2.1	HARK の音声信号処理フロー.	20
2.2	HARKBird の構成.	22
2.3	HARKBird : Recording タブ.	22
2.4	HARKBird : Localization タブ.	23
2.5	HARKBird : Analysis タブ.	24
2.6	HARKBird : アノテーション画面. 図に示される 2 つのグラフは, 録音のスペクトログラム (上) と音源定位結果の時間的分布を MUSIC スペクトル上に示したもの (下) が示されている.	25
3.1	ウグイス ( <i>Horornis diphone</i> ).	28
3.2	ウグイスの歌のスペクトログラム画像. (a) : H 型, (b) : L 型. このスペクトログラム画像の歌は, 実際にプレイバック実験に用いた歌 (SH, SL) である.	29
3.3	実験環境とシステムの配置図.	30
3.4	拡張された HARKBird システム. 本システムは録音を行いながらスピーカから一定間隔, あるいは音源定位後にプレイバックを行う.	31
3.5	音源定位結果に基づくウグイスの歌の方向分布 (プレイバックなし条件). 周囲の円は H 型, L 型それぞれの歌に関する 30 度ずつの定位方向のヒストグラムを示す. 中央の矢印は, 歌を定位の時系列順に繋いだものであり, 簡易的な移動の推移を示す.	32
3.6	音源定位結果に基づくウグイスの歌の方向分布 (実験個体自身の歌再生時).	33
3.7	音源定位結果に基づくウグイスの歌の方向分布 (近隣の同種他個体の歌再生時).	34



3.8	各実験条件におけるウグイス個体のH型の発声割合（横軸）と歌の 定位方向の変化（縦軸）．それぞれのエラーバーは，ブートストラッ プ法による10万回のリサンプリングにより求められた95%信頼区間 を示す．	36
3.9	歌の定位方向の変化量に関するヒストグラム（赤：H型，青：L型）	37
3.10	各条件における大きな定位方向の変化後にL型を歌う割合．赤い点は 実際の観測値，エラーバーは10万回のランダムイゼーションテスト によって得られた信頼区間を示す．	38
3.11	定位方向の変化量と後に続く歌の種類に関する移動エントロピー．赤 い点は実際の観測値で求めた移動エントロピー，エラーバーは10万 回ブートストラップ法によってリサンプリングした時の95%信頼区間 を示す．	40
4.1	ホシワキアカトウヒチョウ ( <i>Pipilo maculatus</i> )	46
4.2	2台のマイクロホンアレイに基づくHARKBirdシステム．母艦であ るノートPCには，2台のマイクロホンアレイ（もう1台は写真外） がUSB接続されており，プレイバック用のスピーカはBluetoothで 接続されている．	47
4.3	録音・プレイバック用HARKBirdシステムの配置図．図中の緑色の 丸はマイクロホンアレイ，黄色の丸はスピーカを示す．各実験では， 直線距離で約150mほど離れた2個体のホシワキアカトウヒチョウの 縄張り内にシステムを配置した．	48
4.4	2台のマイクロホンアレイに基づく2次元定位．画像全体の着色は MUSICスペクトルの強度を示す．青色は弱い強度を示し，赤色は強 い強度を示す．2次元定位は三角測量に基づく手法で行われ，各マイ クロホンアレイから定位方向へ半直線を伸ばし，作られた交点を定位 結果とする．この2次元定位は実時間処理でも実行可能で，その際 にはパラメータで指定したPERIODに基づく時間ステップで定位状況 が更新される．	49
4.5	実験中に2個体の実験個体から観測されたホシワキアカトウヒチョウ の歌と地鳴き．プレイバック音には（a）の歌と（k）の地鳴きを用い ている．	51

4.6	2次元定位の説明図. (a) : 水平方向における三角測量に基づく2次元定位手法. (b) : 音響特性の特徴量空間を利用した音源の手動分類に基づく2次元定位の音源ペアの選定. (c) : 特徴量空間上の距離(類似度)に基づく2次元定位の音源ペアの選定. . . . .	52
4.7	特徴量空間に基づく分類ツール. 各点は音源を示し, クリックすることによって対応するスペクトログラム画像の表示と音源の再生ができる. ラベル付けの際には, 多角形で囲んだ範囲で一度に多くのラベル付けが行える. . . . .	55
4.8	各実験条件における分離音源の分類結果. 各点はHARKBirdを用いて定位した音源を示す. また, 色付けされた音源は実験個体の鳴き声, あるいはプレイバック音を示す. 灰色の音源は鳴き声やプレイバック音ではないノイズなどの音である. . . . .	58
4.9	個体Aの鳴き声に関する空間的・音響的・時間的ダイナミクスの抽出結果. (a) : 鳴き声の空間的分布. (b) : スピーカからみた定位した鳴き声の距離と方向. (c) : 定位した鳴き声の距離と方向のヒストグラム. . . . .	61
4.10	個体Bの鳴き声に関する空間的・音響的・時間的ダイナミクスの抽出結果. . . . .	62
4.11	各プレイバック条件における実験個体の鳴き声に関する分析結果. (a) : 測定値に基づくプレイバック音の2次元定位の定位誤差. (b) : 鳴き声の頻度. (c) : スピーカからみた鳴き声の定位方向の変化量. (d) : スピーカから鳴き声の定位位置までの距離. (a), (c) および (d) のエラーバーは, 標準偏差を示す. . . . .	63
4.12	音響特性の類似度に基づく2次元定位手法によって得られた2次元定位結果. (a) : すべての可能なペアで2次元定位を行った場合の結果. (b) : 提案手法 ( $L=2$ ) による2次元定位結果. (c) : 特徴量空間上での音源の分布. (d) : (b) の2次元定位結果に用いられた配色. 2次元定位には音源ペアの特徴量空間上での座標の midpoint の色が採用されている. . . . .	65
5.1	キンカチョウ ( <i>Taeniopygia guttata</i> ). . . . .	71
5.2	屋外のメッシュテント内に構築した実験環境. . . . .	73
5.3	実験環境の配置図. . . . .	74
5.4	中央のマイクロホンアレイの音源定位結果. . . . .	75

5.5	音源定位に用いるマイクロホンアレイのペアごとの担当箇所を示す図。各色で示された場所では、中央のマイクロホンアレイと、その色に対応するマイクロホンアレイによる2次元定位が行われる。 . . . . .	77
5.6	観測環境内における定位音源の時空間的分布。点は定位された音源を示し、色は定位された時間を示す。灰色の線は定位順に音源を繋いだもの。 . . . . .	78
5.7	中央のマイクロホンアレイの定位音源に関する t-SNE を用いて得た特徴量空間上での分布。 . . . . .	79
5.8	キンカチョウ集団の鳴き声に関する時空間的ダイナミクス。定位された音源が中央のマイクロホンアレイからどの程度離れて（上）、どの方向（下）で定位されたかの時間的推移を示す。各音源の色は図5.7に対応する。 . . . . .	81
5.9	図5.8の40～70秒に着目した結果。 . . . . .	82
5.10	図5.8の440～480秒に着目した結果。 . . . . .	82
5.11	歌と地鳴きの分類結果に基づく鳴き声の時空間的ダイナミクス（0～600秒）。 . . . . .	84
5.12	歌と地鳴きの分類結果に基づく鳴き声の時空間的ダイナミクス（40～70秒）。 . . . . .	84
5.13	歌と地鳴きの分類結果に基づく鳴き声の時空間的ダイナミクス（440～480秒）。 . . . . .	85
5.14	テント内環境。 . . . . .	86
5.15	配置図。 . . . . .	87
5.16	鳴き声判別の流れ。 . . . . .	90
5.17	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。 . . . . .	91
5.18	鳴き声の頻度分布（実験1）。 . . . . .	92
5.19	鳴き声の頻度分布（実験2）。 . . . . .	93
5.20	鳴き声の頻度分布（実験3）。 . . . . .	93
5.21	鳴き声の頻度分布（実験4）。 . . . . .	95
5.22	鳴き声の頻度分布（実験5）。 . . . . .	96
5.23	鳴き声の頻度分布（実験6）。 . . . . .	97

5.24	鳴き声の頻度分布（実験7）	99
5.25	鳴き声の頻度分布（実験8）	100
5.26	オス2羽放鳥実験における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	102
5.27	実験7における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	103
5.28	実験8における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	103

# 表 目 次

1.1	実施した3つの実験内容. . . . .	16
4.1	音響特性の類似度に基づいた2次元定位手法による実験個体の鳴き声の定位成功割合. 手動分類によって得られた結果を正解データとして, その音源の中で定位できている割合を求めている. . . . .	66
5.1	各条件の Male1, Male2 の歌・地鳴きの定位回数. . . . .	101
5.2	各条件の Male1, Male3 の歌・地鳴きの定位回数. . . . .	101
5.3	各条件の Male2, Male3 の歌・地鳴きの定位回数. . . . .	101

# List of Algorithms

1	手動分類に基づく2次元定位. . . . .	54
2	音響特性の類似度に基づく2次元定位. . . . .	56

# 第1章 序論

## 1.1 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

鳥類にとって、鳴き声は重要なコミュニケーション手段である [1]。様々な鳴き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鳴き声で、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伝達するために用いられる [2]。鳥類の鳴き声に基づく相互作用には、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・周波数などの性質に基づく音響的相互作用といった様々な次元がある。

例えば、鳥類集団の鳴き声の相互作用によって生じる現象として、時間的重複回避行動 [3,4] がある。これは、集団内の鳥類は好き勝手自由に発声するわけではなく、他の鳥の鳴き声と重ならないように鳴くという行動であるが、音空間を時間や周波数帯域等で制限のある資源と捉えれば、音響ニッチ仮説 [5] に基づいた適応的な戦略行動によって生じた現象だといえる。この鳴き声に基づく時間的重複回避行動はこれまでもいくつかの調査がある [6-10]。Brummらは、サヨナキドリ (*Luscinia megarhynchos*) に対して、他種の歌の録音のプレイバック実験を行い、実験個体がプレイバックの歌を避けるか調査した。その結果、実験個体は優先的にプレイバック音のない時間帯に鳴き始める傾向が見られ、歌の時間的な発声パターンは他種の歌の影響も受けることが示唆されている。鈴木らはこのような複雑システムとしての鳥類集団に興味を持ち、個体間で生じるさまざまな創発現象の理解・解明を目指している [9,10]。例えば、音響ニッチ仮説に基づく歌うタイミングの重複回避に焦点を合わせ、進化モデルによる行動多様性の創発分析と野外の鳥類の行動観測を併用した分析を行った結果、より長い鳴き声を持つ種が相互作用の時間的ダイナミクスを支配する傾向があることが示唆されている。このような鳴き声の複雑な相互作用

用のダイナミクスや、生態学的機能や文脈を深く理解するためには、相互作用し合う鳥類集団がいつ・どこで・どんな鳴き声を発したかを知る必要がある。

また、鳥類の鳴き声は、生物由来の音と周囲の自然・社会環境との関係を空間・時間などのスケールで調査・理解することを目指した生態音響学 [11] の分野において、重要な主題の1つとして注目されている [12,13]。その理由には、鳥類の音声コミュニケーション自体が複雑なコミュニケーションシステムとして研究対象となること、環境モニタリングのためのツールとして環境に関する有用な情報を有していること [14,15]、生態音響の観測技術確立のための題材となることが挙げられる。

## 1.2 鳥類音声観測技術

一般に、鳥類音声に関しての調査を行う場合、録音を用いた観測手法が主であり、これまでに様々な鳥類音声の観測手法が提案されている。近年では、録音機材の進歩や低価格化に伴って録音に基づく行動観測データが容易に得られるようになっており、特に、自律型録音ユニット (autonomous recording unit; ARU) の開発が進み、鳥類を含む生態音響研究に有用である長期的な音響データの取得が可能になっている [16]。例えば、人手による観測と ARU による観測を比較した実験が多くあり [17-20]、これらの実験結果からは、ARU による観測は、人手による観測よりも時間的にも空間的にも高い観測能力を有し、詳細な鳥類生態観測を行うことができると報告されている。さらに、ARU によって得られた録音の活用として、ディープニューラルネットワークに基づくシングルチャネルの音の分離 [21] や、音圧レベルを用いた音源の距離推定 [22] など様々な試みが行われている。しかし、シングルチャネルのマイクロホンを用いて、複数の音の情報や音環境の構造を知ることは多くの音が混在する自然環境では容易ではなく、多数の個体間で生じる動的な相互作用の分析にも問題が生じるという課題が存在する。

個々の個体情報を抽出する有効な手法として、個体ごとに発信機器や録音機器を取り付ける方法がある [23-26]。例えば、Farine らは鳥類個体に RFID タグを取り付け、巣箱に取り付けた RFID リーダーで個体の在否をみることで、採餌パッチに関する情報伝達に種間と種内で差があることを示している [23]。また、Stowell らは、鳥類個体に小型のマイクロホンを背負わせて録音を行うことにより、録音から個体ごとの鳴き声や飛行などの音を伴う行動の抽出 [25] や抽出した発声に基づいた社会的結合に関する構造の時間的変化について議論している [26]。



このような鳥類に機器を取り付ける観測方法は、個体同士の接触や相互作用に関するデータの取得には有効な手段である。一方で、このような手法は侵襲的であることが懸念されるうえに、鳴き声の発声場所など正確で詳細な空間的情報の取得が困難である。特に後者に関して、個体間の相互作用は縄張りや植生などの空間的な要因に依存する可能性もあると指摘されている [27,28].

一方、複数のチャンネルのマイクホンで構成されるマイクロホンアレイを用いた音源定位技術は、音の方向や位置などの空間的情報を非侵襲的に把握することが出来ることから有望なアプローチとして検討されている [29-31]. 鳥類の音声コミュニケーションに関する理解を深めるには、植生などの環境的要因や縄張り関係などの個体間の位置関係といった空間的情報は重要な情報であるため [32], マイクロホンアレイを用いた鳥類音声の定位に関する研究は観測技術確立のための実証的な研究から実際の生態観測への応用研究など多種多様に行われている [33-39]. Gayk と Mennill は、8 個のマイクロホンをそれぞれ直方体（縦横約 25 m, 高さ約 7 m）の頂点となる配置で並べ、飛行中の渡り鳥の発声位置を三次元三角測量によって推定する試みを行っており、システムの上を通過する渡り鳥の鳴き声をスピーカ音の定位実験と比較しても大差ないほどの誤差で定位できることを示し、音源定位技術が渡り鳥の個体数の定量化や行動監視に活用可能であると論じている [40]. また、Mennill らは、4 台のステレオ録音機を 25 m, 50 m 四方の頂点に 1 台ずつ設置したマイクロホンアレイを構築し、2 次元定位の精度の調査のために様々な動物の鳴き声を用いたプレイバック音による音源定位テストを行った [34]. その結果、音源定位の精度は音源の位置や、環境によって変動する傾向はあったが、2 m 程度の誤差での定位を実現した。Hedley らは、4 チャンネルのマイクロホンアレイで鳥類の鳴き声を用いたスピーカーテストを行い、定位精度を検証した。その結果、95%以上のプレイバック音が水平方向、垂直方向共に 10 度以内の誤差で定位され、マイクロホンアレイを用いた音源の到来方向（direction of arrival; DOA）の推定が、生物多様性調査における個体生息数の評価能力を向上させる可能性があるとして論じている [41].

このように、音源定位技術の生態観測への活用は非常に活発に検討されている。しかし、現状ではマイクロホンアレイによる音源定位技術は生態学者など現場の研究者には広く利用されていない。それは、ソフトウェアやハードウェアの利用が困難であるからだと考えられる。例えば、Mennill らの調査 [34] で用いたデバイスは Wildlife Acoustics Song Meters (model: SM2-GPS; Wildlife Acoustics Inc.) というステレオ録音機であるが、1 台が 1 kg 弱の重さがあるため持ち運びにコストがかか

る。また、録音後には各チャンネルでの録音データの統合処理、音源の注釈付け、音源定位処理をそれぞれ別のソフトウェアで行わなければならない、一貫した処理を行うことが出来ない。その他研究者の提案システムに関しても、マイクロホンアレイの構造が独自に構成したもので一般入手が困難であるなどの理由から広く普及していないのが現状である。

### 1.3 ロボット聴覚技術

音源定位技術に関連して、ロボット聴覚技術がある。これは、実環境下でロボットがマイクロホンアレイを用いて音源の定位・分離・認識などの音声信号処理を統合的に行い、周囲環境の認識や同時発話などの雑音のある音声を認識するための、ロボットにとっての聴覚機能を構築し、活用することを目的とした研究分野である [42,43]。ロボット聴覚技術の特徴から、鳥類音声の観測への活用、特に我々が目指す複数個体が鳴き合う状況における音環境の理解への活用が可能であることは明らかである。例えば、従来手法である目視やビデオ撮影を用いて観測を行う場合、特に木々が生い茂る場所では死角が生じるため鳥類が鳴いた場所を把握することは困難であり、巣など特定の場所に焦点を当てた場所で観測が行われる [44]。一方、ロボット聴覚技術に基づく手法は、鳥が鳴いている場合のみ位置を推測できるというデメリットはあるものの、上記の観測範囲の制約がないために様々な場所・状況を想定した行動観測・生態理解への応用が期待される。

研究に利用可能なオープンソースのロボット聴覚ソフトウェアもいくつか公開されているが [45–47]、ロボット聴覚技術の生態観測への活用はこれまでにない。

鈴木らは、このロボット聴覚技術の鳥類観測への活用を目指し、ロボット聴覚オープンソースソフトウェア HARK (Honda Research Institute Japan Audition for Robots with Kyoto University) [45] の音声信号処理機能を活用した、ノート PC と市販のマイクロホンアレイで構成される鳥類観測のための録音分析システム HARKBird を開発し、その応用や発展の可能性について検討している [48–50]。HARKBird は、マイクロホンアレイによる録音の実行や、録音からの発生音の開始・終了のタイミング及び空間情報（マイクロホンアレイからの位置や方向など）の抽出（音源定位）、音源の分離、定位結果などの可視化を行うことができる GUI ベースのソフトウェアである。これまでに、森林性鳥類が複数鳴き合う様子の観測 [48] やヨシ原で鳴く複数のオオヨシキリ (*Acrocephalus orientalis*) の鳴き声の観測 [51,52] などに活用さ

れている。一方で、これらの研究は基本的にはソングスポットなど一定の場所で歌う個体をターゲットとした観測であり、移動を伴う発声パターンのような詳細な行動に基づく個体間相互作用の観測への適応可能性の検討は未着手である。理由の1つとしては、自然環境において個体同士が積極的に相互作用し合う状況を狙って観測するのは容易ではないことが挙げられる。

## 1.4 研究目的

本論文では、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理解に焦点を合わせた、ロボット聴覚技術を用いた観測・分析手法の実現を目的とする。ロボット聴覚技術は、複数の音源が同時に発生する状況においてそれぞれの音源の到来方向を知ることができ、またそれらの音源の抽出をすることが出来る。特に、これらの特徴を最大限に活かせるのは、複数個体が同時に鳴き合う状況であり、発声のタイミング、鳴き声の種類、発声位置を抽出することで様々な次元で個体間関係の議論が出来る。これは、鳥類の個体間相互作用の理解に大いに貢献すると考える。一方で、その実現には観測の際に生じる問題の解決や観測・分析手法の検討を行う必要がある。

1つには、前述で述べた、自然環境で個体同士が鳴き合う状況を狙って観測することは容易ではないという課題がある。例えば、繁殖期にはオス個体間同士の縄張り争いによって積極的な発声や移動を伴う相互作用が生じるが、このような相互作用が生じる状況は限られており、観測データの蓄積は困難である。また、移動を伴う状況を考える場合、鳥類個体が移動した際にマイクロホンアレイが録音可能な範囲から離れてしまい、十分に鳴き声が取れなくなる可能性がある。これら問題を解決するために、積極的に個体が発声し、かつマイクロホンアレイから実験個体が離れない状況を人工的に作り上げて観測する手法を提案する。具体的には、鳥類の鳴き声の録音をスピーカから再生するプレイバック実験を用いた方法と、自由に飛行可能でかつ空間的に制限した観測環境（フリーフライト環境）で鳥類集団の個体間相互作用を観測する方法の2つの観測手法を提案し、各調査手法の有用性を議論する。

2つ目は、観測手法に関する検討である。例えば、一定の位置で鳴く個体の発声タイミングを抽出したい場合には単一のマイクロホンアレイによる録音で十分抽出可能であるが、移動を伴う個体の発声状況（場所、タイミング）を抽出したい場合には、複数のマイクロホンアレイによる観測が必要である。このように、観測目標

に応じて適切な観測手法が選択されるべきである。本稿では、この課題に関して使用するマイクロホンアレイの台数に応じた観測手法をそれぞれ提案し、各手法で得られる鳥類生態に関する観測データの粒度を調査することで、ロボット聴覚技術に基づく鳥類生態観測の汎用性について議論する。

最後に、抽出した音源の分類に関する課題である。録音には、鳥類の鳴き声以外に瀬音や葉擦れ音など様々な音源が含まれるが、分析の際には目標鳥類の鳴き声を抽出する必要がある。従来、識別は音源定位結果をを時系列順に逐次確認して行っていたが、これは時間的にコストがかかる。また、これまでの実験では鳥類個体が比較的動かない状況での観測であったため定位方向情報からある程度音源の識別が出来たが、今後の目標とする、複数鳥類個体が動きながら鳴き合う状況での観測では定位方向から識別するのは困難である。

本研究では、この課題解決の取り組みとして機械学習技術の活用を考え、比較的単純な機械学習の活用から深層学習を用いた大規模データに基づく分類まで試行し、その適用可能性を調査する。

表 1.1: 実施した3つの実験内容.

実験番号	実験手法	主な検討内容	実験対象種	マイクロホンアレイ 使用台数
1	プレイバック実験	HARKによるプレイバック実験が可能なシステムの構築	ウグイス ( <i>Horornis diphone</i> )	1
2	プレイバック実験	2次元定位手法、個体の時空間的発声ダイナミクスの抽出及び鳴き声の分類支援	ホシワキアカトウヒチョウ ( <i>Pipilo maculatus</i> )	2
3	フリーフライト環境での録音観測実験	構築環境での詳細な個体間相互作用の抽出と鳴き声の分類および個体識別	キンカチョウ ( <i>Taeniopygia guttata</i> )	5

上記に向けた取り組みとして、表 1.1 に示す3つの実験 [53–55] を行う。1つ目の実験 [53] では、HARKの実時間処理と音声信号処理フロー内でのPythonスクリプトの実行を組み合わせることで録音と同時にプレイバック実験が実施可能な枠組みの構築を行う。構築したシステムを用いてウグイス (*Horornis diphone*) に対してプレイバック実験を実施し、その様子を1台のマイクロホンアレイを用いて観測することで音源の方向情報からプレイバックが実験個体に与える影響が観測可能かを調査する。

2つ目の実験 [54] では、2台のマイクロホンアレイによる鳴き声の観測が1台のマイクロホンアレイによる方向分布の分析と比較してどの程度具体的に個体の挙動

が観測できるかを調査する。具体的には、2羽のホシワキアカトウヒチョウ (*Pipilo maculatus*) に対して役割の異なる鳴き声のプレイバックを行い、プレイバック音の違いによる個体の時空間的発声ダイナミクスの変化を2台のマイクロホンアレイの録音から調査する。また、機械学習を用いた分類支援の提案及び試行とその応用を検討する。

3つ目の実験 [55] では、フリーフライト環境での録音実験の枠組みを構築し、構築した環境内に複数のキンカチョウ (*Taeniopygia guttata*) を放ち、その状況で生じる個体間相互作用の観測を行う。観測には複数台のマイクロホンアレイを用い、1台あるいは2台のマイクロホンアレイによる録音と比較してより高い解像度での発声ダイナミクスの抽出を目指す。さらに、機械学習による鳴き声の分類および個体識別を検討する。

## 1.5 論文の構成

まず、序論となる1章では、本研究の背景と目的、概要を述べる。

2章では、本研究で利用するソフトウェアであるロボット聴覚オープンソースソフトウェア HARK と HARK の音声信号処理機能を鳥類音声観測用に容易に扱えるよう組み込んだシステムである HARKBird の基本的な仕組みを述べた上で、提案する観測手法（音源定位・音源分離等）の原理について説明する。

3章では、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内のウグイスに対して行ったプレイバック実験に関して、単一マイクロホンアレイを用いて録音・分析した結果並びに音源の方向情報を用いた定量的分析について報告する。

4章では、3章の取り組みの発展として、ホシワキアカトウヒチョウに対して行ったプレイバック実験に関して、2台のマイクロホンアレイを用いた2次元定位手法の検討や、実験個体に対するプレイバック音の違いによる影響についての調査結果を報告する。

5章では、複数個体が鳴き合う状況での相互作用の観測・調査として、キンカチョウを屋外テントに構築した観測環境に放鳥し、その様子を複数台のマイクロホンアレイを用いて録音・分析した調査結果について、鳥類集団の鳴き声の抽出と複数台のマイクロホンアレイによる鳴き声の観測手法の検討を目的とした予備的調査と、実際の社会状況を想定した模擬環境での録音調査に分けて報告する。

6章では，結論として，上記3つの調査実験についてまとめ，本手法の鳥類生態観測・理解への有用性について論じる．

## 第2章 HARK・HARKBird

本章では，本研究で扱う鳥類音声観測手法の土台となるソフトウェア HARK と，HARK の音声信号処理機能を鳥類観測用に容易に扱えるように組み込んだシステムである HARKBird の基本的な仕組みについて述べる。

### 2.1 HARK

#### 2.1.1 概要

HARK は，音源の定位・分離・認識などの音声信号処理を統合的に行うことのできるオープンソースロボット聴覚ソフトウェアである．本ソフトウェアは，様々な環境においてロボットが音を聞き分けて適切な処理を行えるようになることを前提として開発されており，そのような処理に対応するための様々な信号処理に関するモジュール群を提供している．モジュールには，音源定位・音源分離に関するものや，音源の特徴量の抽出に関するものなどがあり，これらの処理をオフラインあるいはリアルタイムで実行できる．さらに，Python スクリプトを信号処理中に実行させるためのモジュールもあり，拡張性は非常に高い．

これらのモジュール群に基づく音声信号処理は，HARK designer という，ブラウザ上で動作する GUI ベースのエディタによって直感的に構築できる．図 2.1 に，HARK designer で構築した音声信号処理の例を示す．この例は，実際に録音の音源定位と分離を行う際に用いた処理フローであるが，この処理フローには，マイクロホンアレイに基づく録音信号を入力として受け取って音源定位と音源分離を行う処理，定位に関する情報をファイルに書き込むための処理が実装されている．

#### 2.1.2 音源定位

HARK による音源定位について説明する．音源定位は，MUSIC (MUltiple Signal Classification) [56] によって行われる．MUSIC 法は，マイクロホンアレイから得た

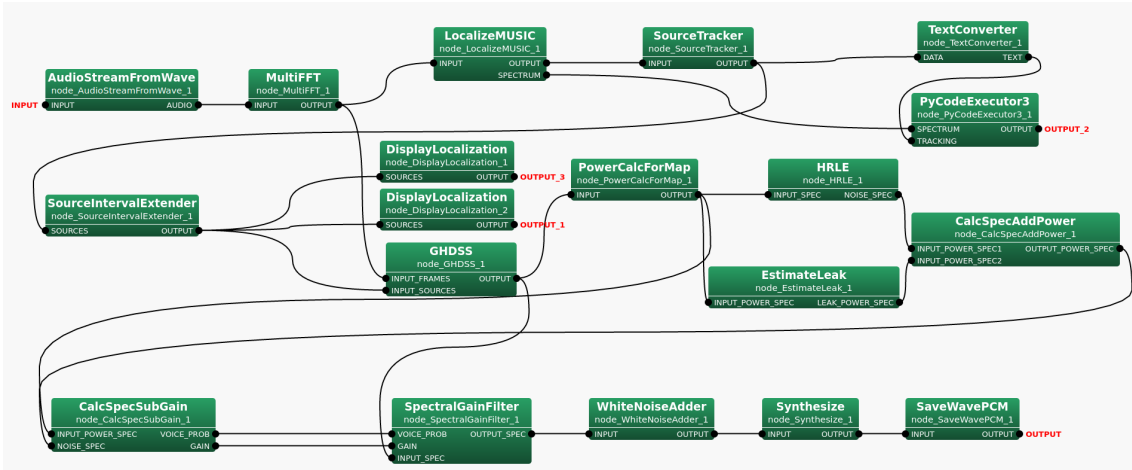


図 2.1: HARK の音声信号処理フロー。

入力信号のチャンネル間の相関行列を固有値分解することによって音源の定位方向を推定する手法である。手順としては、まずマイクロホンアレイの入力信号を短時間フーリエ変換することによって得られる各チャンネルの周波数領域の信号ベクトルを求め、この信号ベクトルからチャンネル間の相関行列を得る。得られた相関行列に関して、固有値分解を行って固有ベクトルと固有値を得る。このとき、得られた固有値は音源のパワーと相関があり、つまり得られた固有ベクトルに対応する固有値が高いものは音源、低いものは雑音とみなせる。しかし、このベクトル自体には方向的な意味がないのでただちにこのベクトルから音源の定位方向を求めることができない。そこで、計測あるいは幾何学的に計算された、ある方向に存在する音源から各マイクロホンに到達するまでに生じる振幅や位相の変化等の入出力特性を記述した伝達関数を用いて、伝達関数における到来方向に対応するベクトルと雑音に対応する固有ベクトルとの内積に基づく計算から、目標信号に対応するベクトルを探す。この処理によって、MUSIC スペクトルという伝達関数の方向分解能に基づく音声信号のパワーが得られる。この MUSIC スペクトルのピークが音源の到来方向となる。図 2.1 の処理フローでは、各チャンネルの周波数領域の信号ベクトルの取得を MultiFFT ノードで行い、その値を MUSIC 法の処理に基づく LocalizeMUSIC ノードに入力し、音源定位の処理を行うことで音源定位結果を得ている。なお、LocalizeMUSIC ノードからは音源定位結果が出力されるが、その出力を入力とする SourceTracker ノードでは得られた定位音源の MUSIC パワーがある閾値より高い場合のみ音源定位結果として採用する処理が行われている。本論文の実験では、8 チャンネルのマイク



ロホンアレイを用いたが、基準となるマイクロホンが磁北方向に向かうように設置することで実際の環境内での方向と得られた定位方向との対応をとっている。

### 2.1.3 音源分離

次に、音源分離に関する説明を行う。音源分離は、GHDSS (Geometric High-order Dicorrelation-based Source Separation) アルゴリズム [57] によって行われる。GHDSS は、観測時の空間の伝達関数と観測ノイズの影響を受けたマイクロホンアレイが観測する音源に基づく信号から、本来の信号に近い分離音が抽出できるような分離行列を推定し、分離音を抽出する手法である。本手法では、音源信号間 (分離音源間) の高次相関行列の対角成分以外が 0 (高次無相関化) になる制約と直接成分が歪みなく分離される幾何的制約のもとで分離行列を推定していく。図 2.1 の処理フローに示される GHDSS ノードでは、入力に混合音のマルチチャンネル複素スペクトル (各チャンネルの周波数領域の信号ベクトル) と定位された音源の方向情報が入力され、この情報に基づいて音源分離の処理が行われる。

## 2.2 HARKBird

HARKBird は、フィールドにおいて安価で容易に利用可能なシステムを目指した、野鳥の歌行動観測のためのノート PC と USB マイクロホンアレイを用いた観測・分析システムである (図 2.2)。このシステムは、HARK, HARK-Python, PySide などがインストールされた Ubuntu Linux 上で動作し、マイクロホンアレイでの録音の開始、終了をはじめ、HARK の機能を用いた音源定位と分離や、分析結果の可視化・注釈付けといった機能が「Recording」・「Localization」・「Analysis」の 3 つのタブで構成される GUI 上での操作によって利用できる。

Recording タブ (図 2.3) では、マイクロホンアレイを使った録音の実行や、保存した録音の再生や分割を行うことができる。録音は、図 2.3 右側に示すような設定画面でマイクロホンアレイのチャンネル数や録音時間などのパラメータを指定後、そのパラメータに基づいてシェルコマンドである arecord によって録音が実行される。ノート PC に複数のマイクロホンアレイを接続すれば、同時に複数のマイクロホンアレイでの録音も可能である。

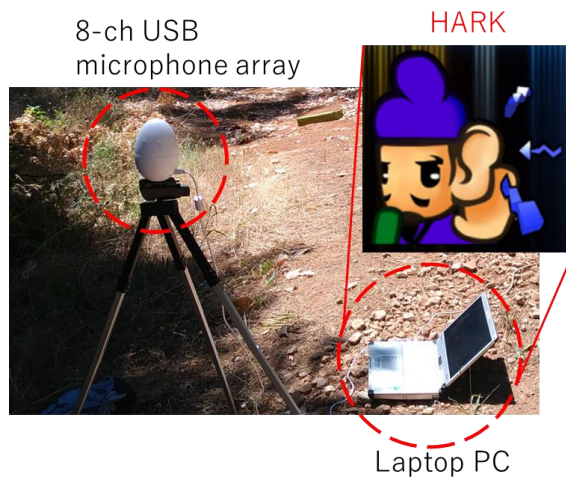


図 2.2: HARKBird の構成.

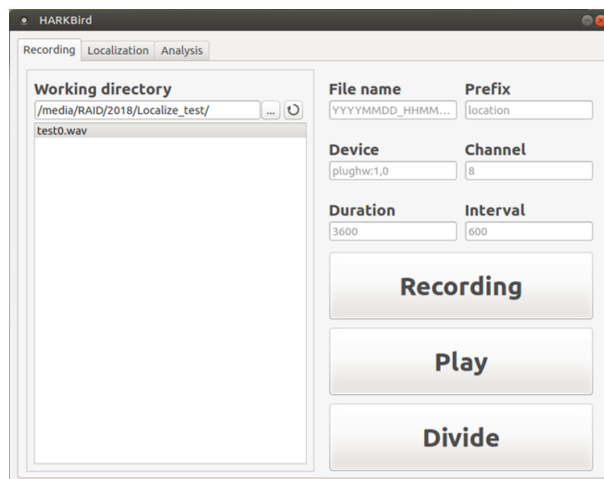


図 2.3: HARKBird : Recording タブ.

Localization タブ（図 2.4）では，選択した録音に対して図 2.1 に示す HARK の音源信号処理フローを実行し，音源定位・分離の結果を任意のフォルダに出力する．音源定位と音源分離に必要なパラメータは，図 2.4 右側の Parameter から設定できる．デフォルトで設定された各パラメータの値は，経験的に鳥類の鳴き声が定位できるようなパラメータが設定してある．代表的なパラメータを以下に説明する．

- NUM\_SOURCE：音源定位のピーク検出で強調すべき目的音源数．
- LOWER\_BOUND\_FREQUENCY：音源定位のピーク検出時に考慮する周波数帯域の下限．
- UPPER\_BOUND\_FREQUENCY：音源定位のピーク検出時に考慮する周波数帯域の上限．
- THRETH：音源定位結果を無視すべきかを判定するための MUSIC スペクトルの閾値．
- PERIOD：音源定位の演算周期（PERIOD/100 sec.）．

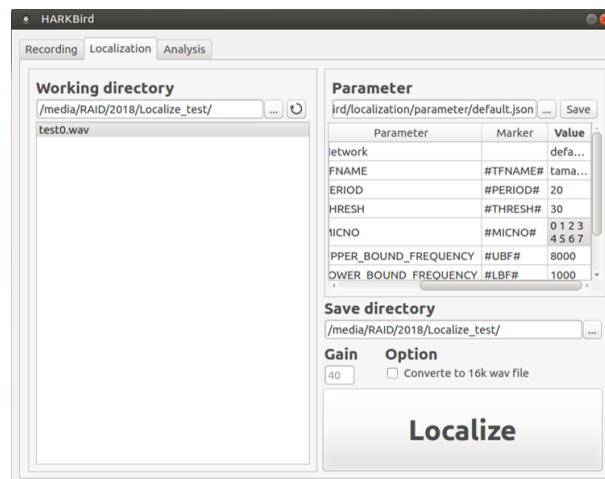


図 2.4: HARKBird : Localization タブ．

Analysis タブ（図 2.5）では，音源の方向結果の出力や可視化，分析に関する出力や，それらの結果に基づいたインタラクティブなデータの可視化や注釈付けなどを行うことができる．音源定位・分離の結果が入ったフォルダを指定して Export file ボタンを押すと，録音のスペクトログラムと音源定位結果の時間的分布が可視化さ

れたファイルが出力される。また、Annotation ボタンを押すと、出力されるファイルと同様の結果が可視化されたウィンドウ（図 2.6）が開かれる。このウィンドウでは、Export file ボタンの処理によって出力された結果に関して、インタラクティブに定位音源の再生や、スペクトログラムの確認、定位結果の修正や削除などの編集が行える。

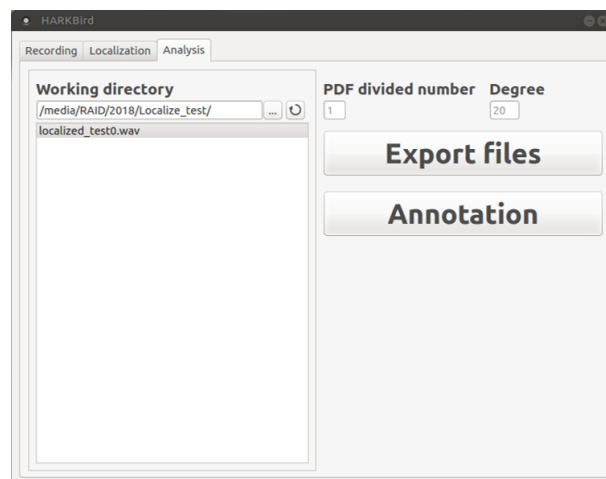


図 2.5: HARKBird : Analysis タブ.

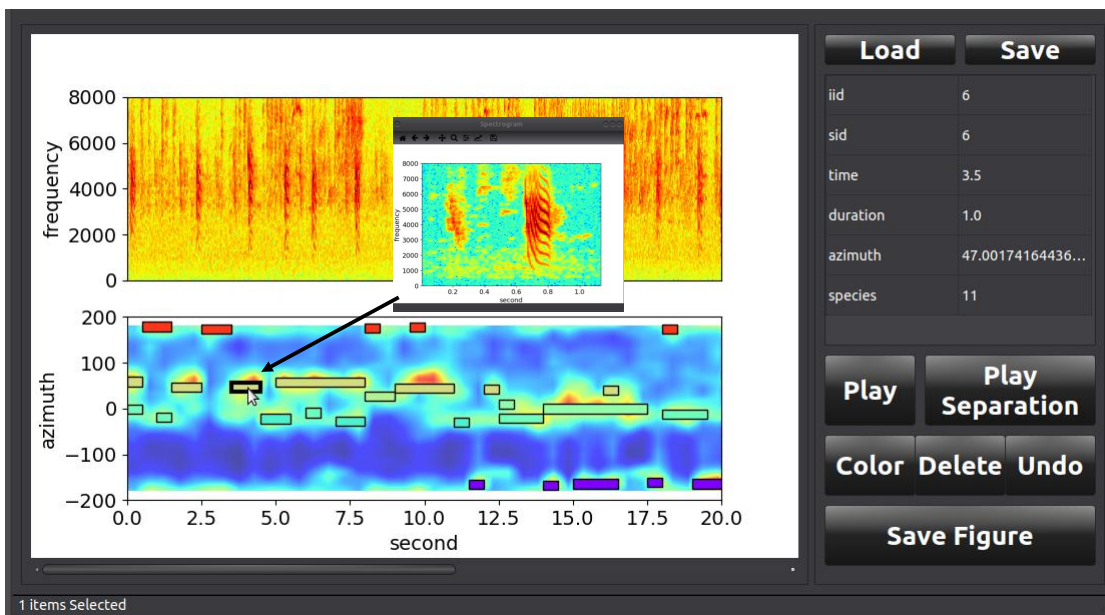


図 2.6: HARKBird : アノテーション画面. 図に示される 2つのグラフは, 録音のスペクトログラム (上) と音源定位結果の時間的分布を MUSIC スペクトル上に示したもの (下) が示されている.

# 第3章 ウグイスに対するプレイバック 実験における鳴き声の方向分布 分析

## 3.1 背景と目的

鳥類の生態観測には様々な手法があるが、その1つとしてスピーカーから鳥の鳴き声を再生し、対象となる個体の反応を調べるプレイバック実験という手法がある。また、プレイバック実験の中には対象となる個体の行動に合わせてスピーカから鳴き声を再生するインタラクティブ実験と呼ばれる枠組みもあり、この手法によって対象となる個体の行動と鳴き声の種類・タイミングとの間に生じる相互作用を詳細に分析することができる。野鳥を対象として行われたプレイバック実験として、Dabelsteenの実験を紹介する [58]。この実験では、クロウタドリ (*Turdus merula*) の縄張り内で、攻撃性が昇順で強いとされる LI, HI, SS の3種のさえずりを周期的に3つの方法で再生し、その時個体が行う攻撃的行動の観察を行う実験を行った。3つの方法とは、同じ歌を再生、個体とスピーカとの距離に応じて再生音を変える（近いほど攻撃性の高い歌を再生する）インタラクティブな再生、ランダムに再生音を変えて再生する方法である。その結果、同じ歌を再生、あるいはランダムに再生した場合には、HI と SS で行動に差が出なかったものの、インタラクティブな実験でのみ差を確認することができた。これにより、相手の距離に応じて鳴き声を変えるインタラクティブな性質が野鳥のコミュニケーションにとって重要であるということを示した。このように、鳥類の鳴き声の適応的意義や意味を理解するためには歌う鳥と周囲の環境や他個体との関係を知ることが非常に重要であるといえる。しかし、このような実験では再生実験を行いながら個体がどのように動き、どのように鳴くのかを調べる必要があるためその追跡は大変な作業となる。一方で、2章で説明した通り、HARK の機能には Python スクリプトによる簡単な拡張や、リアルタイム処

理などプレイバック実験に活用できる機能が多くある。例えば、これらの機能を使うことによってプレイバック実験における音声再生の自動化や、音源定位を組み合わせることによる定位時のみのプレイバック、つまりインタラクティブ実験などに活用可能となる。

本研究は、HARKの様々な機能を用いることで、1台のマイクロホンアレイでリアルタイムに録音を行い音源の到来方向を特定しながら、自動的にプレイバック実験やインタラクティブ実験を行うことができる枠組みを構築し、その適用可能性を検討することを目的とする。その構築システムの評価として、ウグイス (*Horornis diphone*) を対象としてプレイバック実験を行った。具体的には、ウグイス個体の縄張り内に、スピーカから鳴き声を再生できるシステムを設置し、ウグイスの歌を様々な時間間隔で再生する実験を行った。その後、プレイバックが対象個体の動きや発声に与える影響をマイクロホンアレイによる音源の定位方向情報を用いて定量的に調査した。

## 3.2 実験手法

### 3.2.1 ウグイス (*Horornis diphone*)

ウグイス (*Horornis diphone*) は日本の代表的な鳴禽類であり、日本国内に広く分布する種である (図 3.1)。ウグイスのオスは、似ているようで微妙に異なる2種類のH型、L型と呼ばれる歌を歌う [59]。H型の歌は「ホーホケキョ」と一般的に聞かれる歌であり、L型と比較すると周波数が高い歌である (図 3.2a)。一方、L型の歌は「ホー・ホ・ホケキョ」と歌の導入部分が断続的になる特徴があり、H型と比較すると全体的に周波数は低い歌である (図 3.2b)。L型の歌は、オスのウグイスが縄張り内の外側付近でよく鳴くことから、近隣個体への威嚇の意味がある歌といわれる。オス個体に対して同種の個体の歌のプレイバックを行った結果、対象個体は周囲を飛び回って他の個体を探す挙動を示しながら、H型の歌の頻度を減らし、L型の歌を頻繁に歌う傾向があったという報告もある [59, 60]。また、最近ではウグイスの歌構造の地域差、選択圧の影響調査を行った研究もある [61, 62]。浜尾は、日本の本州、三宅島・新島などの島々、およびアメリカ合衆国・ハワイのオアフ島でウグイスのさえずりを録音し、比較を行った。その結果、日本の本州のウグイスのさえずりは、他の地域のさえずりに比べて一回のさえずり時間が長く、音の数も

多いことを示し、本州以外の島々ではさえずりはよりシンプルなものになっていることを観測した。本州のウグイスは、繁殖期になると縄張りを形成し繁殖を行うが、この縄張りは一定の場所ではなく移り変わっていく。一方、島々ではオス同士の競争も緩く、縄張りを移動する必要がないためさえずりの単調化が進んだ可能性を指摘した。その他にも、周囲の環境によって音の伝達特性が変化することなども理由として挙げた。HARK を用いた歌の動きと歌の種類の時間的な関係を詳細に把握可能な分析手法は、このようなウグイスの歌の生態的な役割についての理解をさらに深めることが期待される。



図 3.1: ウグイス (*Horornis diphone*).

### 3.2.2 HARKBird のリアルタイム録音・プレイバック（インタラクティブ実験）機能の拡張

本実験を行う当時、HARKBird はプレイバックを自動で行う機能やインタラクティブ実験を行うために用いられるリアルタイム録音に対応していなかった。そこで、Bluetooth 接続したスピーカを用いて、指定した録音ファイルの自動再生や、実時間での録音・定位を行いながらその定位に応じてプレイバックを行う機能をこれまで用いてきた HARKBird で用いられる HARK の処理フローの一部を変更して追加した。具体的には、入力を受け付けるノードである `AudioStreamFromWave` を `AudioStreamFromMic` に変更してマイクロホンアレイの入力を受け付ける。録音中



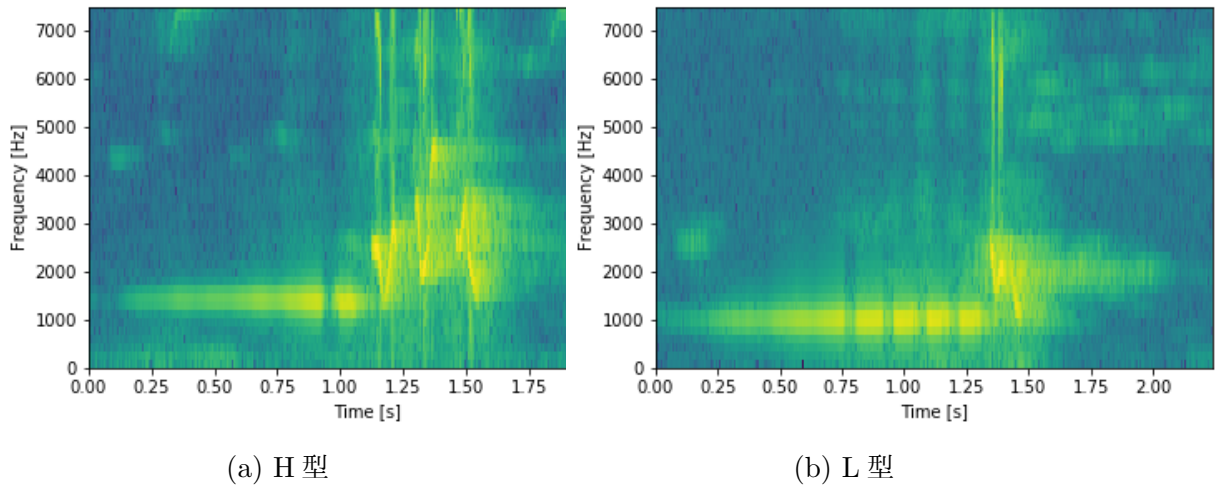


図 3.2: ウグイスの歌のスペクトログラム画像. (a) : H 型, (b) : L 型. このスペクトログラム画像の歌は, 実際にプレイバック実験に用いた歌 (SH, SL) である.

一定間隔でプレイバック, または音源が定位された際にのみプレイバックが行えるよう HARK の Python による処理が行えるノード (PycodexExecutor) 内で実行されるスクリプトにプレイバックの処理を追加する. なお, 音源定位に用いるパラメータは, ウグイス個体の歌が十分に定位でき, かつ水の音や葉音などの他の音の定位をできるだけ抑制するパラメータを用いた (NUM\_SOURCE=3, LOWER\_BOUND\_FREQUENCY = 2200 and THRESH = 28.5).

実験に用いる HARKBird システムには, ノート PC (TOUGHBOOK CF-C2, パナソニック株式会社), 三脚に設置した 8 チャンネルマイクロホンアレイ (TAMAGO, 株式会社システムインフロンティア), プレイバック用スピーカ (MM-SPBTBK; サンワサプライ株式会社) を用いた. ノート PC は, HARKBird による音源定位処理を実行するために, Ubuntu Linux 12.04 の OS 環境に HARK と HARK-Python をインストールしたものを採用している. また, 実験で利用したマイクロホンアレイである TAMAGO は, ノート PC に USB 接続し実験中は 24 ビット 16kHz での録音を行った.

### 3.2.3 プレイバック実験とインタラクティブ実験

実験は, 2016 年 5 月 21 日に愛知県豊田市にある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の森林内 (35°21'N, 137°57'E)

で行った。本フィールドは主にスギ・ヒノキ・アカマツからなる針葉樹人工林で、コナラ・シデ・カエデ等の小さな広葉樹パッチが点在する地域であり、繁殖期の真っ最中のため多くの鳥が活発に鳴いている状況であった。

具体的な実験設定について説明する。実験は、フィールド内の周囲を針葉樹・広葉樹に囲まれている開けた舗装の無い駐車場(図 3.3)で行い、その中央にシステムを配置した(図 3.4)。スピーカおよび PC は、マイクロホンアレイを原点として真東の方角に 5m 離れた地点に配置した。このとき、スピーカは指向性を考慮して上向きに設置した。この場所は、実験の対象個体としたウグイスが頻繁に鳴いている場所であり、このウグイスの縄張りである可能性が高い。また、対象個体が歌うソングポストは、マイクロホンアレイ周辺から大きな差のない距離の木々にあり、方向を計測することで個体の移動パターンをおおよそ把握できる。そのため、単一のマイクロホンアレイでの実験には適した場所であった。

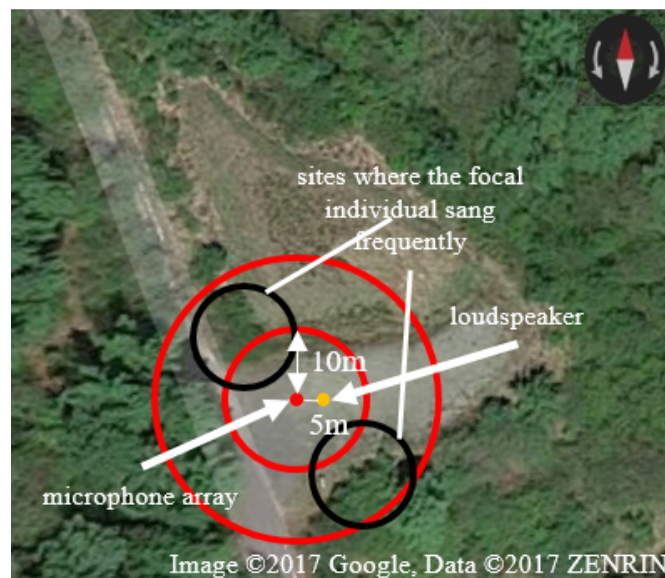


図 3.3: 実験環境とシステムの配置図.

実験には 4 種の異なるウグイスの歌を用いた。そのうち 2 つは事前に録音した実験の対象個体の H 型の歌 (SH) と L 型の歌 (SL)，もう 2 つには実験実施場所から 200m ほど離れた地点で鳴いていたウグイス個体の H 型 (OH) と L 型 (OL) の歌を用いた。それぞれの音源には、デジタルオーディオエディタである Audacity<sup>1</sup>を用いて録音に含まれていたノイズの除去、正規化の処理を行った。それぞれの歌の

<sup>1</sup><https://www.audacityteam.org/>



図 3.4: 拡張された HARKBird システム. 本システムは録音を行いながらスピーカから一定間隔, あるいは音源定位後にプレイバックを行う.

録音を以下の3つの条件, 計12条件でプレイバックを行った: 1) 8秒間隔での再生 (F8), 2) 12秒間隔での再生 (F12), 3) HARKの機能を使った, 音源定位後3秒後にプレイバックを行う実験 (インタラクティブ実験; I3). また, 上記のプレイバック実験に加えて, プレイバックを行わない自然の状態での録音を2回行った (N1, N2).

それぞれの条件で10分以上のプレイバック実験を行い, そのうち対象ウグイス個体がシステム近隣で積極的に鳴いていた500秒に注目し, 分析を行った. OH/F12 (他個体のH型を12秒間隔で再生)の実験条件に関しては, 対象個体が実験中に飛び去ってしまう状況が生じ, 連続的に鳴いていた録音を確保できなかった. そのため, この条件だけは750秒間の録音を分析して, 個体が鳴いていた時間が合計で500秒になるようにした. 実験中は, 注目個体以外のウグイスは近隣で鳴いていないことを確認した.

実験後, HARKBirdを用いて音源定位を行い, 実験個体の歌をH型とL型に目視で確認しながら分類した. これにより, プレイバック実験中のウグイスが鳴いた歌の種類とタイミング, 方向を得た. 以降の結果はそのデータに基づくものである.

### 3.3 分析結果

#### 3.3.1 歌の方向分布分析

図 3.5-3.7 は，行った録音実験のすべての条件における，対象ウグイス個体が鳴いた歌の定位結果に基づく方向分布を示したものである．それぞれの図について，赤色と青色の丸は H 型あるいは L 型のさえずりをマイクから見てその方向で歌ったことを意味し，方向は 0 度が北（磁北）を指す．さえずり同士を繋ぐ矢印は，定位したさえずりを定位した順に繋いだものであり，さえずり前後での移動を簡易的に示したものである．ヒストグラムは，方向を 30 度ごとに区切った各範囲内での H 型，L 型の各歌の回数を示す．図 3.5 は，プレイバックを行っていない条件の結果であるが，ウグイス個体は，H 型と L 型の歌をどちらも利用して鳴いていることがわかる．また，移動の推移を示す矢印が少ないことから，一定の位置で留まりながら鳴いている状況であると推測される．一方で，図 3.6，図 3.7 のプレイバックのある条件では，概して H 型の歌が減った一方で L 型の歌はあまり減っていない傾向があった．

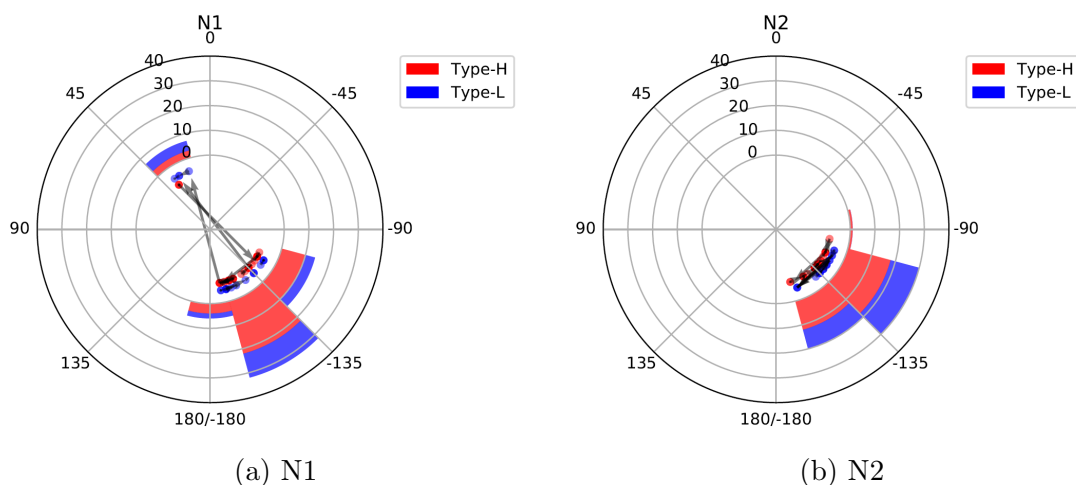


図 3.5: 音源定位結果に基づくウグイスの歌の方向分布（プレイバックなし条件）．周囲の円は H 型，L 型それぞれの歌に関する 30 度ずつの定位方向のヒストグラムを示す．中央の矢印は，歌を定位の時系列順に繋いだものであり，簡易的な移動の推移を示す．

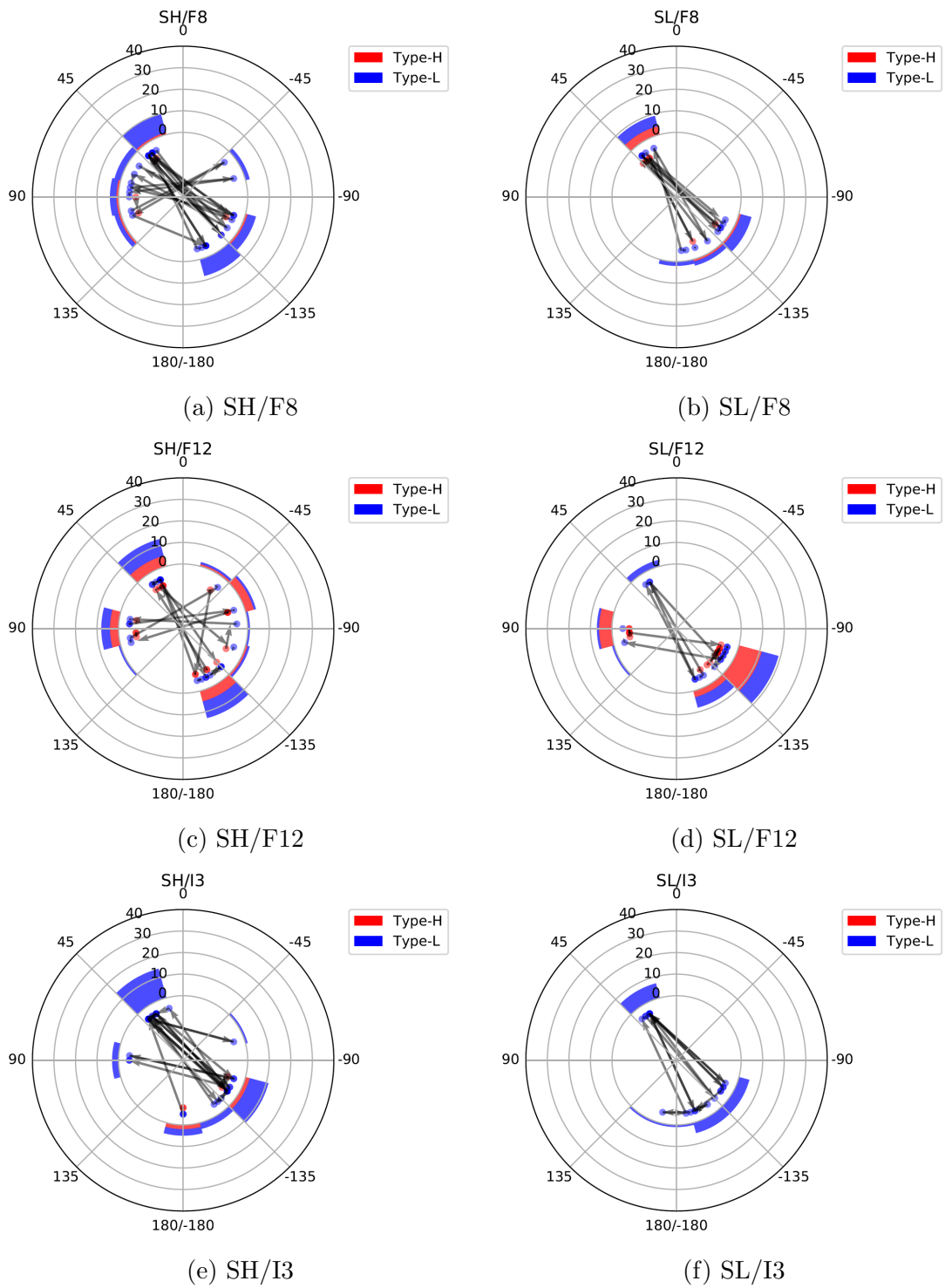
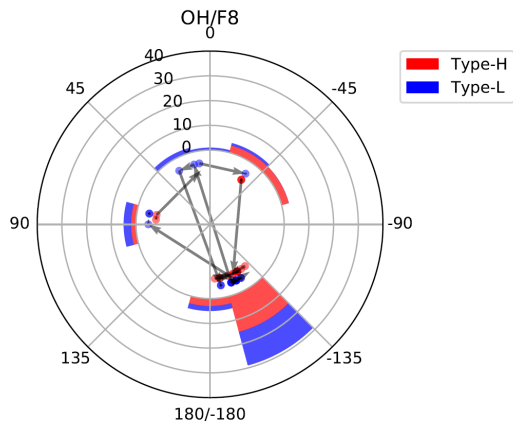
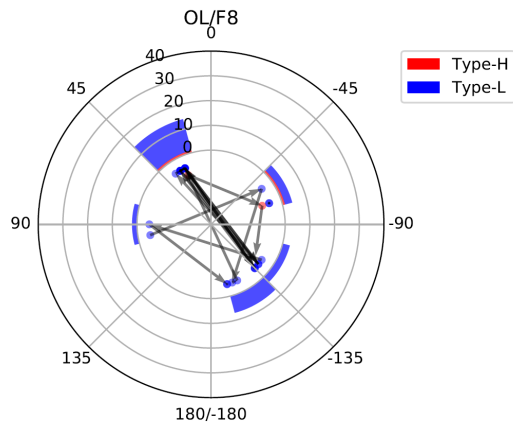


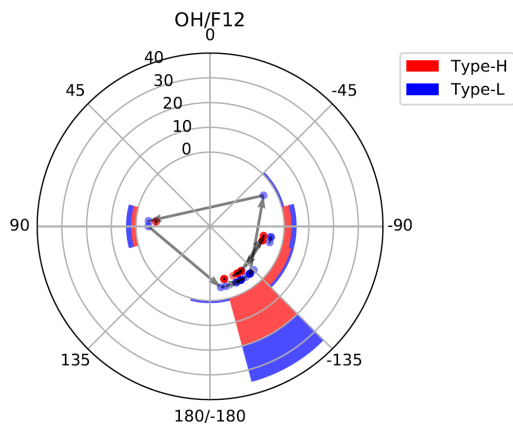
図 3.6: 音源定位結果に基づくウグイスの歌の方向分布 (実験個体自身の歌再生時).



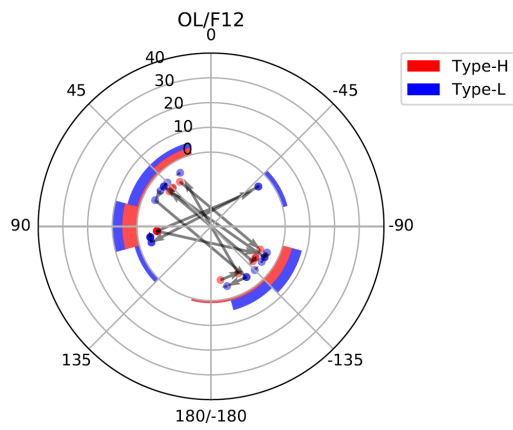
(a) OH/F8



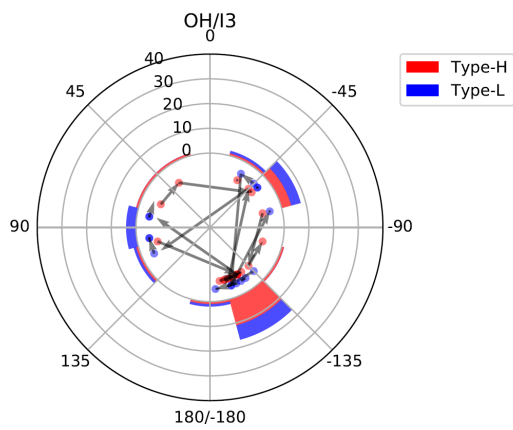
(b) OL/F8



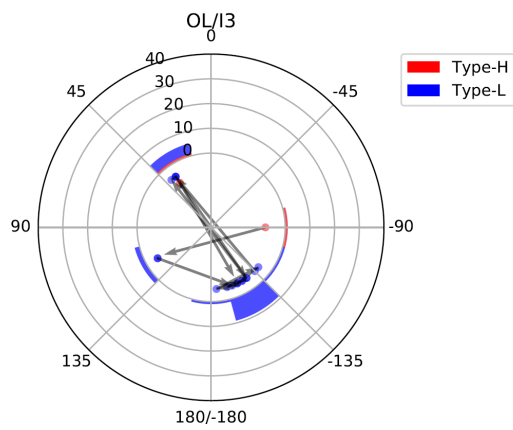
(c) OH/F12



(d) OL/F12



(e) OH/I3



(f) OL/I3

図 3.7: 音源定位結果に基づくウグイスの歌の方向分布 (近隣の同種他個体の歌再生時) .

### 3.3.2 プレイバックの影響の定量的分析

プレイバックを行うと、通常の様子とは異なる傾向を示すことが確認された。そこで、次は歌の種類に注目して歌の種類と定位方向の変化、つまり、歌と移動の関係に着目して分析を行った。図3.8は、横軸に対象個体が鳴いた歌のうちのH型の割合、縦軸にある歌の定位方向とその次の歌の定位方向の差の平均を各条件で求めたものである。結果からは、プレイバックなしの条件ではプレイバックがある条件と比較してH型の割合が高く、方向の変化が小さいことが確認できる。一方で、プレイバックのある条件では、H型の割合は低下しており、また定位方向の変化量は大きくなっている。条件別にみると、この傾向は、プレイバックの間隔が12秒（F12）の場合よりも8秒（F8）の場合に顕著にみられる。このことから、再生頻度が高いほどプレイバック音が対象個体の行動に強い影響を与えていることが示唆される。インタラクティブ実験（I3）に注目すると、12秒間隔のプレイバック（F12）よりも強い影響を受けているようである。I3の条件におけるプレイバックの平均再生間隔は19.2秒であり、これは12秒間隔（F12）よりもプレイバック頻度は低いことから、インタラクティブ実験でのプレイバックは、周期的なプレイバックとは異なる、より強い影響を対象個体に与えた可能性がある。また、他個体の歌（O）を再生した場合には、L型を再生した時（OL）はH型を再生した場合（OH）と比較して定位方向の変化量とL型の頻度が共に高い。L型の歌が他個体への警戒であるということを考えれば、自身に向けられたL型の歌に対して強い警戒を持ったということが推測される。一方、自身の歌（S）の再生した実験に関しては、F8・F12条件ではH型再生時、I3条件ではL型再生時にL型の頻度と定位方向の変化量が高くなった。自身の歌が聞こえる状況は不自然な状況であることから、条件によって傾向が変化したと推測される。

次に、方向変化量（移動量）と歌の種類の影響を調査した。図3.9は各条件における、次の歌（H or L）が前の歌からどの程度定位方向が変化して歌われたかをヒストグラムで示したものである。この結果から、変化量の小さな箇所（0-30度）と、変化量の大きな箇所（主に120度以上）の2か所で歌頻度のピークがあることが確認できる。

定位方向の小さな変化は、同じ木々、あるいは隣の木など比較的近くの移動を示すと考えられ、その際には実験個体はH型・L型どちらも歌っている。一方、定位方向の大きな変化は離れた遠くの木への移動を示すと考えられるが、このとき実験

個体はL型を多く歌う傾向があった。このことから、実験個体の発声ダイナミクスにおいて、歌の種類と移動に何らかの関係があることが示唆される。

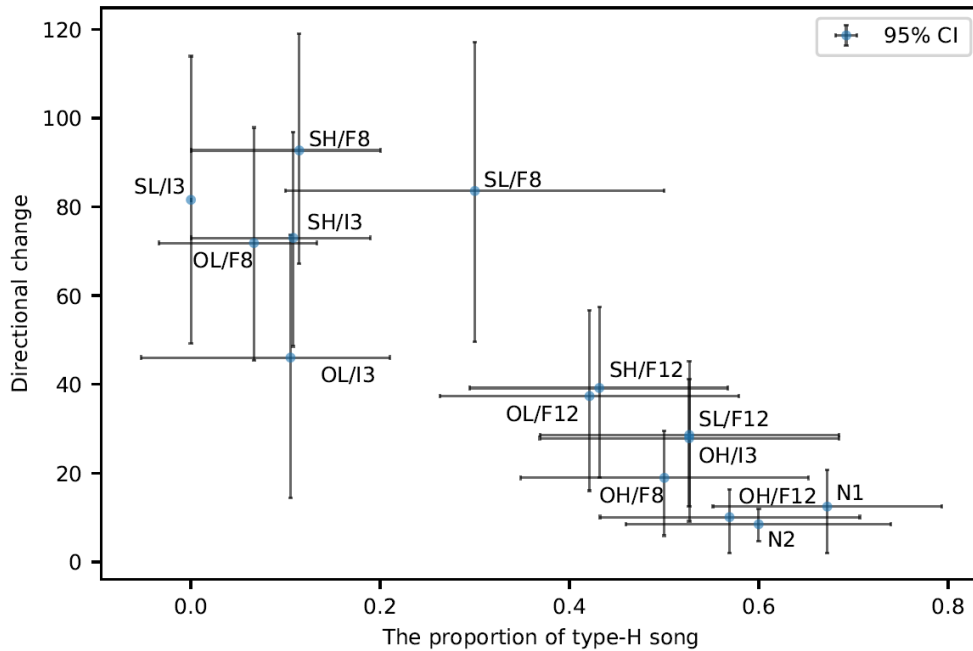


図 3.8: 各実験条件におけるウグイス個体の H 型の発声割合（横軸）と歌の定位方向の変化（縦軸）。それぞれのエラーバーは、ブートストラップ法による 10 万回のリサンプリングにより求められた 95%信頼区間を示す。



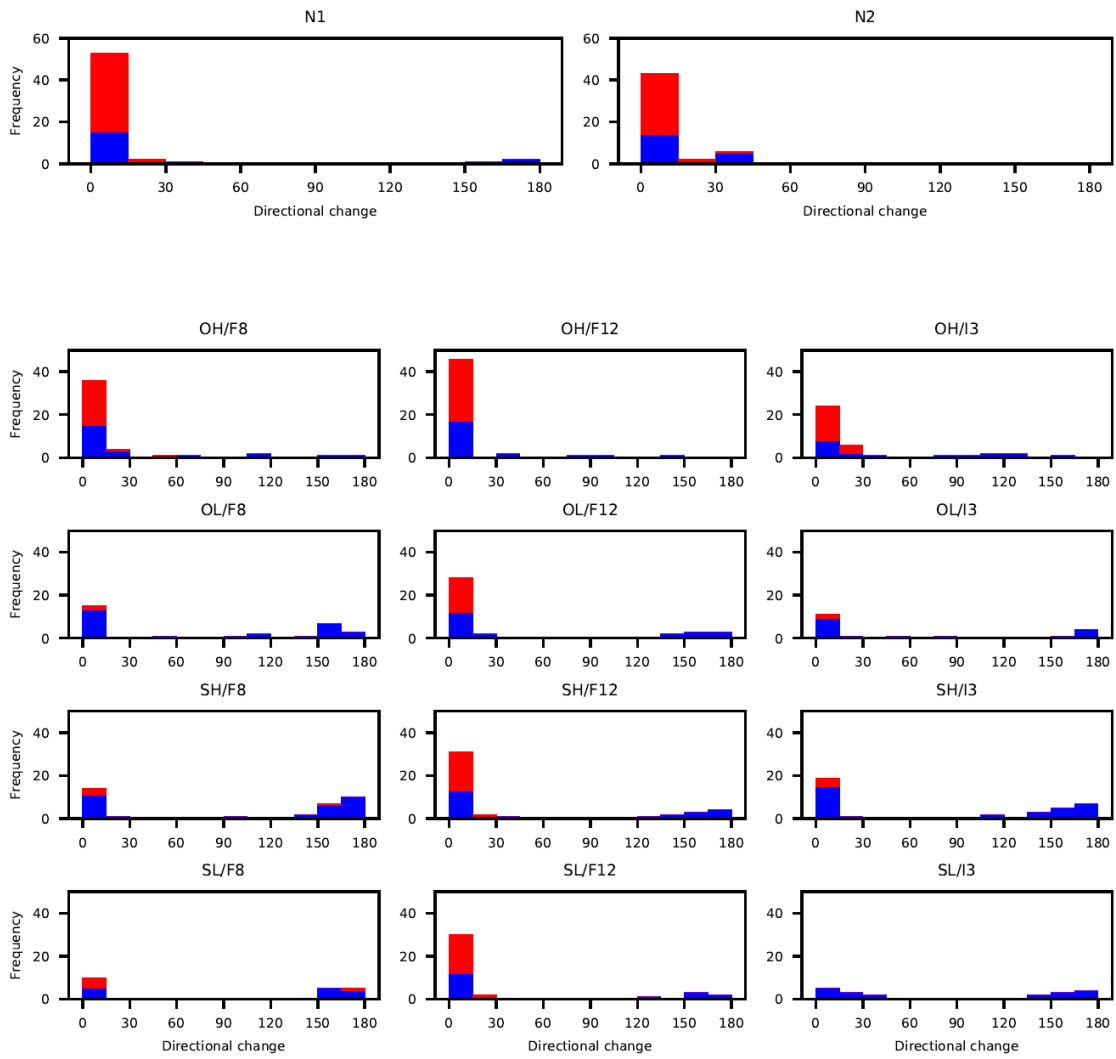


図 3.9: 歌の定位方向の変化量に関するヒストグラム (赤 : H 型, 青 : L 型) .

### 3.3.3 統計分析

前項で確認された傾向をより深く議論するために、定位方向と鳴き声の種類に関するランダムマイゼーションテストを実施した。具体的には、前項の方向変化量に関して、30度より大きい値と30度以下の値で分ける。前者を大きな移動、後者を小さな移動と仮定して、大きな移動の後にL型の歌がどの程度の割合で歌われたかを調べた。ランダムマイゼーションテストに関しては、定位方向と歌の種類の時系列データに関して、データの順番をランダムに入れ替え、その際の大きな移動後に歌われたL型の割合を求めた。これを10万回行い、95%の信頼区間を出した。観測値と信頼区間を図3.10に示す。観測値では、ほぼ全ての条件で大きな移動後にL型を歌う割合が10割を示した。一方で、ランダムマイゼーションテストでは、多くの条件でその観測値よりも低い値を示し、統計的に対象個体が移動後にはL型を歌う傾向があるということが確認された。

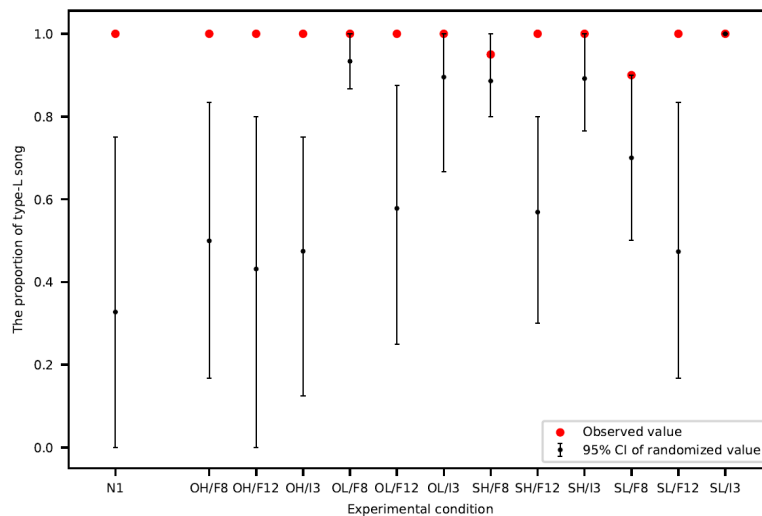
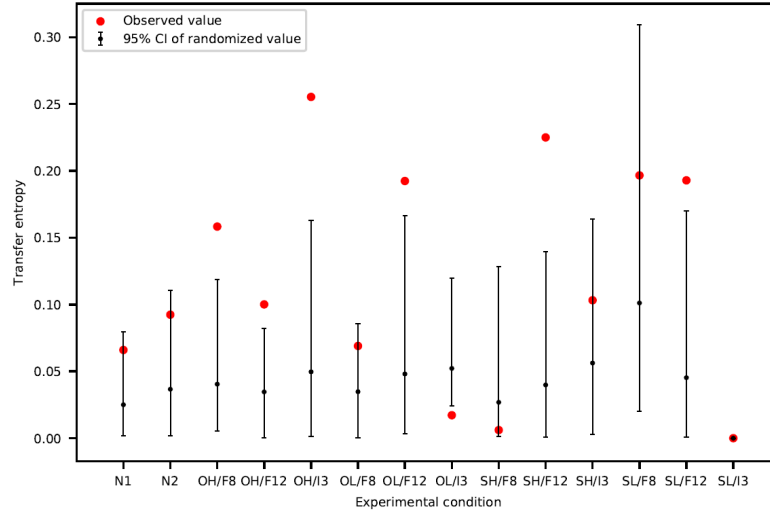


図 3.10: 各条件における大きな定位方向の変化後にL型を歌う割合。赤い点は実際の観測値，エラーバーは10万回のランダムマイゼーションテストによって得られた信頼区間を示す。

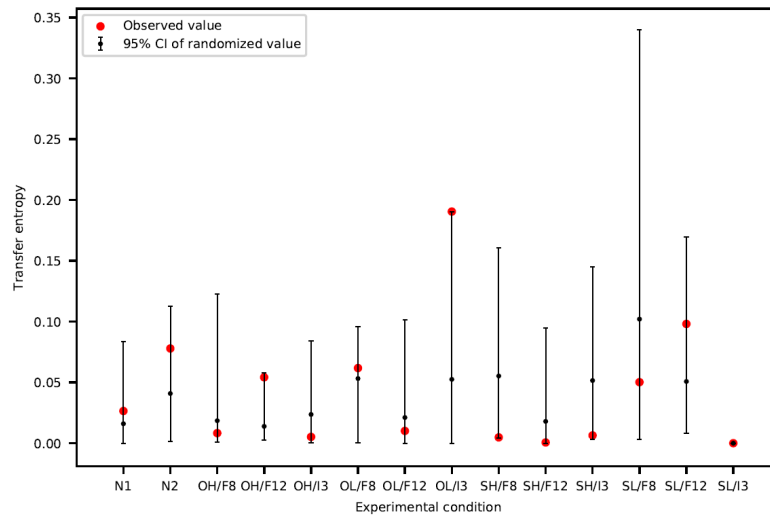
次に、移動とL型の関係に関して、どのような因果関係があるかを調査するために、値の大小で離散化した方向の変化量とその後の歌の種類との間の移動エントロピーを求めた。移動エントロピーは、ある時系列から別の時系列への方向性のある情報の流れの期待値を定量化したものであり、離散的な時系列  $Y = \{y : t\} : t = 1, 2, \dots$  から別の離散的な時系列  $X = \{x : t\} : t = 1, 2, \dots$  への移動エントロピー  $TE : Y \rightarrow X$  を意味する。過去の  $X$  の値が与えられたとき、過去の  $Y$  の値を知ることによって、将来の  $X$  の値に関する不確実性がどれだけ減少する（予測できる）かを以下の式で求める。

$$TE : Y \rightarrow X = \sum_{x_{t+1}, x_t, y_t} p(x_{t+1}, x_t, y_t) \log\left(\frac{p(x_{t+1}|x_t, y_t)}{p(x_{t+1}|x_t)}\right) \quad (3.1)$$

本実験では、 $X$  と  $Y$  は定位方向の変化量（大小で1と0）とその後の歌の種類（H型：0，L型：1）の時系列データに対応する。統計的に有意かどうかを調査するために、観測値における移動エントロピーと、上記で用いた10万回のランダム化テストにおける移動エントロピーを比較した。図3.11に求めた移動エントロピーを示す。(a)は定位方向の変化量を式3.1の $X$ 、その後の歌の種類を $Y$ としたものの、(b)はその逆の結果である。(a)においては、6条件において、統計的に有意な差がある結果が示された。一方で、(b)においては統計的に有意な差のある結果はなかった。このことより、歌う前の移動が移動後の歌の種類に影響を与えているということが示唆される。



(a) 歌の定位方向の変化量とその次に歌われる歌の種類に関する移動エントロピー.



(b) 歌の種類とその次に定位される歌の定位方向に関する移動エントロピー.

図 3.11: 定位方向の変化量と後に続く歌の種類に関する移動エントロピー. 赤い点は実際の観測値で求めた移動エントロピー, エラーバーは10万回ブートストラップ法によってリサンプリングした時の95%信頼区間を示す.

### 3.4 議論

本章では、HARKのリアルタイム処理と、Pythonプログラムの追加機能を用いた現行のHARKBirdの拡張を行うことによって、プレイバックおよびインタラクティブ実験が可能な枠組みを構築し、その枠組みの評価として、様々な再生条件における歌の種類と移動との間のダイナミクスを理解するためにウグイスに対するプレイバック実験を行った。その結果、プレイバックが実験ウグイス個体の行動パターンに影響を与えることの観測に成功した。

実験対象のウグイス個体は、プレイバックの影響を受けて、歌の頻度を減らして定位方向が大きく変わる、つまり移動を頻繁に行う挙動を示した。この傾向は、先行研究における目視による観測結果と一致する [60]。今回の結果は、1個体だけの観測による結果であるが、プレイバック音と再生頻度の組み合わせを変えて実験を行った結果、再生頻度を変えた実験では、再生頻度が高いほどプレイバックの影響が強くなるという一般的な傾向が見られた。一方で、インタラクティブにプレイバックを行う実験では、プレイバックの平均再生間隔が19.2秒であり、これは他の条件よりも長い間隔であるにもかかわらずより強い影響を与えた。

Maynardの行った実験 [63] では、オナガセアオマイコドリ (*Chiroxiphia linearis*) のオスが、近隣の個体と歌い合う場合は歌の重複を避けるが、プレイバックによって行った疑似的な他個体との鳴き合いの時には重複を避けないことが示されている。この結果は、一定間隔のプレイバックとインタラクティブなプレイバックの影響の違いを示唆していると考えられ、これは今回行った研究における傾向と一致している。また、本システムは、プレイバック処理を担うPythonのスクリプトを変更することで条件の設定を容易に変更できる。そのため、インタラクティブな再生でのプレイバック実験を行うことのできるHARKに基づくシステムは、鳴き合う個体同士の生態観測、特に様々な再生条件の違いによる詳細な行動変化の分析において貢献しうる。

また、再生音には自身の歌と近隣個体の歌の録音を用いたが、それぞれのプレイバック実験では異なる影響があることが確認された。この結果からは、実験個体は少なくともそれぞれのプレイバックを異なる個体の歌として認識していることが推測される。一方で、実験個体は自身の歌をどのように認識したかどうかは議論の余地がある。考えられる可能性としては、自身の歌としての認識、自身の親や子個体の歌としての認識、それ以外の他個体としての認識などがあげられる。そのため、今

回のプレイバック音に対する影響をより深く議論するには更なる調査が必要であると考える。

さらに、ランダマイゼーションテストを行うことによって、マイクロホンアレイからみたウグイス個体の歌の定位方向が大きく変化した際にはL型の歌を歌う傾向があるという、移動と歌の種類の関係を明らかにすることができた。この傾向を移動エントロピーを用いて分析したところ、移動から次の鳴き声への情報の流れについて有意な差があることが確認された。この傾向は、L型の歌が近隣個体に対する威嚇として用いられることを考慮すると以下の2つが推測される。1つには、未知の場所へ移動する際の予備的警戒としての発声である。実験個体の移動先は未知の場所であり、他個体が居る可能性が高く、そのための発声である可能性がある。もう1つは、実験個体がプレイバック個体を探していたという可能性である。プレイバックはスピーカから再生しており、実験個体は個体の実態を認識できない状態にある。そのため、警戒という状況と、飛び回って探すという状況が組み合わさった結果移動後にL型を歌うという状況が出来上がったのではないかと考えられる。

このような歌と移動（定位方向）の情報を得るためには、手作業で歌以外の不要な音源を定位結果から取り除き、対象個体の歌をH型とL型に分類する必要があった。しかし、それぞれの歌の音源定位情報は提案システムから得られたものであり、これは人手による観測を行った場合には簡単には得られない重要なデータであり、その意味で価値のある観測手法であるといえる。

# 第4章 ホシワキアカトウヒチョウに対するプレイバック実験における鳴き声の時空間的分析

## 4.1 背景と目的

前章では、ウグイスに対して行ったプレイバック・インタラクティブ実験の様子を単一のマイクロホンアレイで録音し、音源定位を行うことで得た音源の定位方向情報を分析することにより、プレイバックの条件に応じてウグイス個体の挙動が変化することを定量的に示した。一方で、観測環境は単一のマイクロホンアレイによる観測で、個体の居場所がある程度把握できるような場所を選んでいる。実際、鳥類生態の観測は森林内で行われることが多く、その状況では単一のマイクロホンアレイによる観測から個体の場所が推測できるような場所は限定される。

鳥類個体の鳴き声の発声場所を方向的にではなく、空間的に把握する方法として、複数台のマイクロホンアレイを用いた三角測量に基づく手法がある。Suzukiらは、ヨシ原に生息する複数のオオヨシキリ (*Acrocephalus orientalis*) が鳴いている状況を、3台のマイクロホンアレイを用いて録音を行い、個体が歌うソングポストの空間的位置の推定を2次元定位によって試みた。2次元定位は、マイクロホンアレイからみて、定位した音源方向へ半直線を伸ばした時、他のマイクロホンアレイから伸ばした半直線と交点が発生したときにこれを音源の空間的な位置とみなす手法である。この手法により得られたオオヨシキリのソングポストの定位結果は、マイクロホンアレイ同士の距離が30~70 mと比較的離れていたにもかかわらず、見通しの良い場所からの目視による観測で記録した個体の位置と比較しても  $5.5 \pm 4.5$  m 程度の誤差で観測され、比較的高い精度で音源の位置を定位できたことを示している。また、鳴いていた2個体のオオヨシキリについて、一方の歌が他方の歌にどれだけ影響を与えるかを移動エントロピーにより調査した結果、一方の歌がもう一方に影

響を与えている傾向があることが示唆された。

以上のように、複数のマイクロホンアレイによる2次元定位は、空間的な鳴き声の位置を把握でき、かつ個体間相互作用に関する理解に貢献することができる。しかし、このオオヨシキリに対する実験では、観測個体はほとんど一定の位置で鳴いているという状況であったため、2次元定位による鳥類音声の観測が、縄張り構造、縄張り防衛などの生態学的条件に応じた鳥類行動に関する空間的ダイナミクスの理解に貢献するかどうかの検討は不十分である。その検討には、観測個体の積極的な動きや発声を促すためのプレイバックを行う実験での観測が必要であると考えられる。また、2次元定位手法に関して、各マイクロホンアレイの音源同士で交点を作る際に、同じ時間に複数の音源があった場合、音源が同じ音源でなく他の音源とも交点を作ってしまう場合があり、その際にはゴーストという本来音のない場所で音が定位されてしまう。それを避けるために、上記のオオヨシキリの実験では前章のウグイスに対する実験と同様に手作業による確認と注釈付けによってオオヨシキリの鳴き声の抽出を行っているが、これは録音を時系列順に逐次確認しながら行っていくコストのかかる作業である。

本章では、ロボット聴覚技術による2次元定位が鳥類発声行動に基づく時空間的ダイナミクスの理解に貢献しうるかをより具体的な状況で検討すること、上記の定位音源の分類タスクの改善を目的とする。その題材として、米国カリフォルニア州の森林内でホシワキアカトウヒチョウ (*Pipilo maculatus*) に対してプレイバック実験を行った。歌と地鳴きの、用途がそれぞれ異なる鳴き声を用いてプレイバック実験を行うことにより、プレイバック音の違いによる挙動の変化を観測した。一般に、歌は縄張りの主張や求愛、地鳴きは社会的結合の維持や警戒を表す意味があるとされているため、歌と地鳴きでは個体の行動に差が生じることが予測される。例えば、歌を再生した場合にはより攻撃的な姿勢を持つことが考えられる。また、個体の攻撃的な姿勢の強さを測る際には個体間の距離を測る場合もある [64, 65]。そのため、今回の実験ではスピーカに対して実験個体がどのような挙動を示したかについて注目する。プレイバックの様子は2台のマイクロホンアレイで録音し、2次元定位を行うことで音源の空間的分布からその影響を調査した。分類タスクの改善に関しては、次元削減アルゴリズム t-SNE [66] に基づく低次元特徴量空間上でのアノテーション機能を HARKBird に追加することにより実現した。また、この特徴量空間上での距離を音源の類似度とみなすことで、半自動的に2次元定位を行う手法の検討も行った。音響生態学において、録音から音風景を理解する試みは、環境全体の生物



多様性の把握や、環境のダイナミクスを知るという意味で注目されている。例えば、PierettiらはAcoustic Complexity Index (ACI)という周波数領域ごとの音強度の時間的変動に基づく指標を用いて、鳥類の鳴き声に基づく音風景とACIの関係について調査した [67]。その結果、この指標と録音中の鳥の鳴き声の数には相関関係があり、ACIが音空間の豊かさを反映したことを示している。ロボット聴覚技術に基づく観測は音源ごとに情報の抽出が行えるという点で、上記で提案される指標よりも詳細な環境の情報を持つ可能性がある。そのため、今回提案する半自動的な2次元定位手法によって容易に音風景の様子を概観することができれば、ロボット聴覚技術の生態音響学への貢献が大いに期待される。

## 4.2 実験手法

### 4.2.1 ホシワキアカトウヒチョウ (*Pipilo maculatus*)

ホシワキアカトウヒチョウ (*Pipilo maculatus*, 図4.1)は、北アメリカの西海岸から中西部にかけて広く生息する種であり、林縁部、生垣、農業地域の藪など、密集した低木林や雑木林に営巣する。外形的特徴としては黒い頭部、赤い目を持ち、背中と翼には和名のもとになっている白い斑点がある。歌は、短い導入音とトリルと呼ばれる2つの音を連続して往復(振動)させて出す音から構成されるもの、トリルのみで構成されるものなど、構造は比較的単純だが個体で様々な歌を持つ。また、猫の鳴き声ように聞こえる地鳴きを持つが、採餌中、木に留まっている時など様々な場面で用いられる [68]。

広い範囲に生息する種であることから、河岸付近の植林による生態系の回復経過を観察するための指標として営巣状況が観測された例もある [69]。その一方で、鳴き声に関する生態調査は多くなく、今回のプレイバック実験ではこれまでに調査されていない移動と鳴き声の関係などに関する知見の獲得が期待される。

### 4.2.2 実験環境

実験は、2018年5月に米国カリフォルニア州ボルケーノ近郊の針葉樹と広葉樹の混交林が生い茂る森林内(38°48'N, 120°63'W)で行った。実験を実施した時期は、本実験フィールドに生息する鳥類にとっては繁殖期であり、様々な鳥が鳴いている状



図 4.1: ホシワキアカトウヒチョウ (*Pipilo maculatus*) .

況であった。実験対象種としたホシワキアカトウヒチョウの鳴き声はフィールドに生息する他の種の鳴き声と比較しても音量が大きくはっきりとしている。そのため、様々な鳥類が鳴き合う状況でも実験個体の鳴き声の定位・抽出が容易である。

実験中のプレイバックおよび録音は、HARKBirdに基づくシステムを用いて行う。システムは、ノート PC (TOUGHBOOK CF-C2; Panasonic) に 2 台の 8 チャンネルマイクロホンアレイ (TAMAGO; 株式会社システムインフロンティア) を接続し、プレイバック用のスピーカ (A3143; ANKER) を Bluetooth 接続した構成となっており、このシステムをホシワキアカトウヒチョウが頻繁に鳴いている、個体の縄張り内に配置した (図 4.2)。ノート PC の環境には、音源定位を行うための HARK と HARK-Python がインストールされた Ubuntu Linux 16.04 を採用した。マイクロホンアレイは三脚の上部に取り付け、スピーカはシステム近くにある木に音の指向性を考慮して上向きに設置した。

実験は、2 個体のホシワキアカトウヒチョウ (個体 A, 個体 B とする) に対して行った。これらの個体は、お互いに約 150 m ほど離れた場所に生息しており、縄張りを隣接してもつ個体であると考えた。図 4.3 に示すように、それぞれの個体が頻繁に鳴く場所にシステムを配置した。個体 A に対する実験では、2 台のマイクロホンアレイを約 23.2 m の距離を離して設置した。また、スピーカをそれぞれのマイクロホンアレイから 15.4 m, 15.5 m ほど離れた場所に設置した。個体 B に対する実験

では、2台のマイクロホンアレイを約13.9 mの距離を離して設置し、スピーカを各マイクロホンアレイから8 mほど離れた位置に設置した。マイクロホンアレイ及びスピーカの設置位置は、プレイバック実験の事前に実験個体およびスピーカの音が十分に定位できることを確認した上で決定した場所である。これは、ウグイスの実験で用いた実時間処理の機能を活用して、2次元定位の様子をフィールドで確認できるような拡張を行うことで実現している [70]。これによってフィールドにおいても図4.4のような音源定位の状況が確認できるようになり、本機能はシステムの配置場所の検討に貢献した。

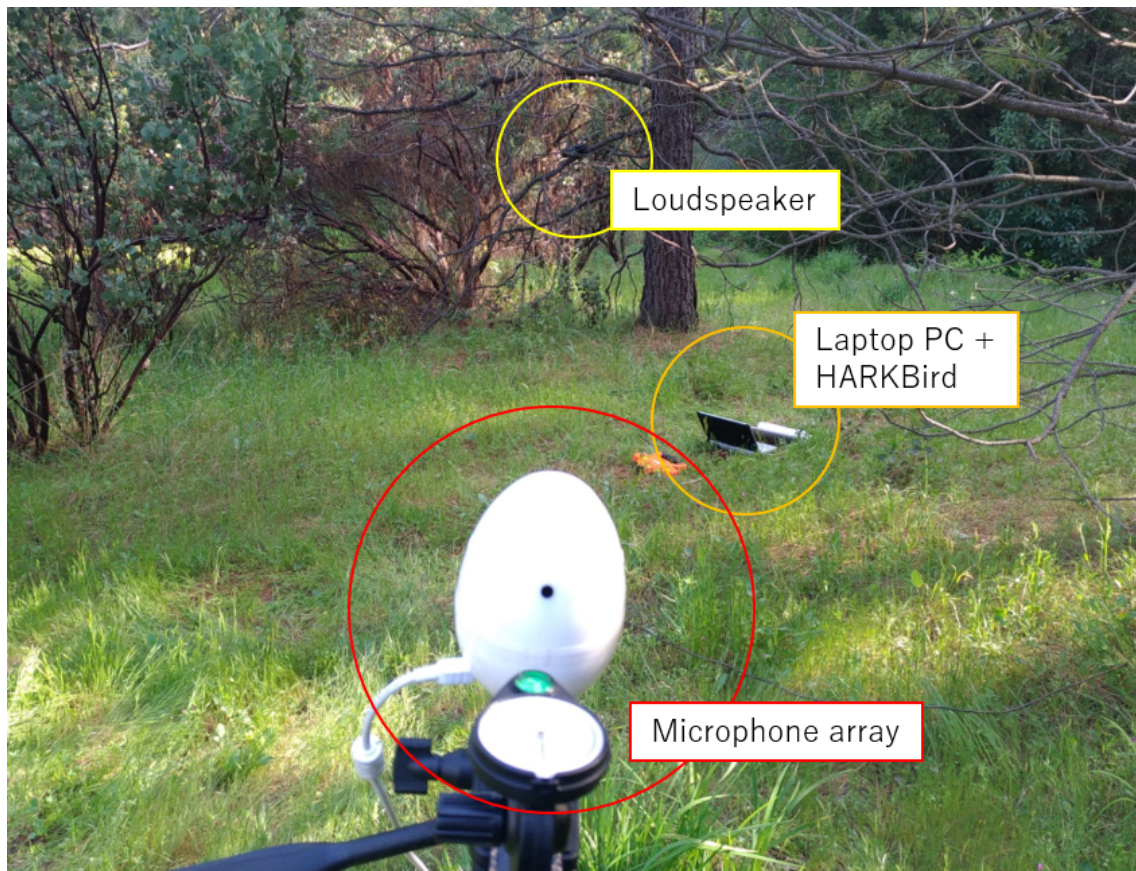


図 4.2: 2台のマイクロホンアレイに基づく HARKBird システム。母艦であるノート PC には、2台のマイクロホンアレイ（もう1台は写真外）が USB 接続されており、プレイバック用のスピーカは Bluetooth で接続されている。



図 4.3: 録音・プレイバック用 HARKBird システムの配置図. 図中の緑色の丸はマイクロホンアレイ, 黄色の丸はスピーカを示す. 各実験では, 直線距離で約 150m ほど離れた 2 個体のホシワキアカトウヒチョウの縄張り内にシステムを配置した.

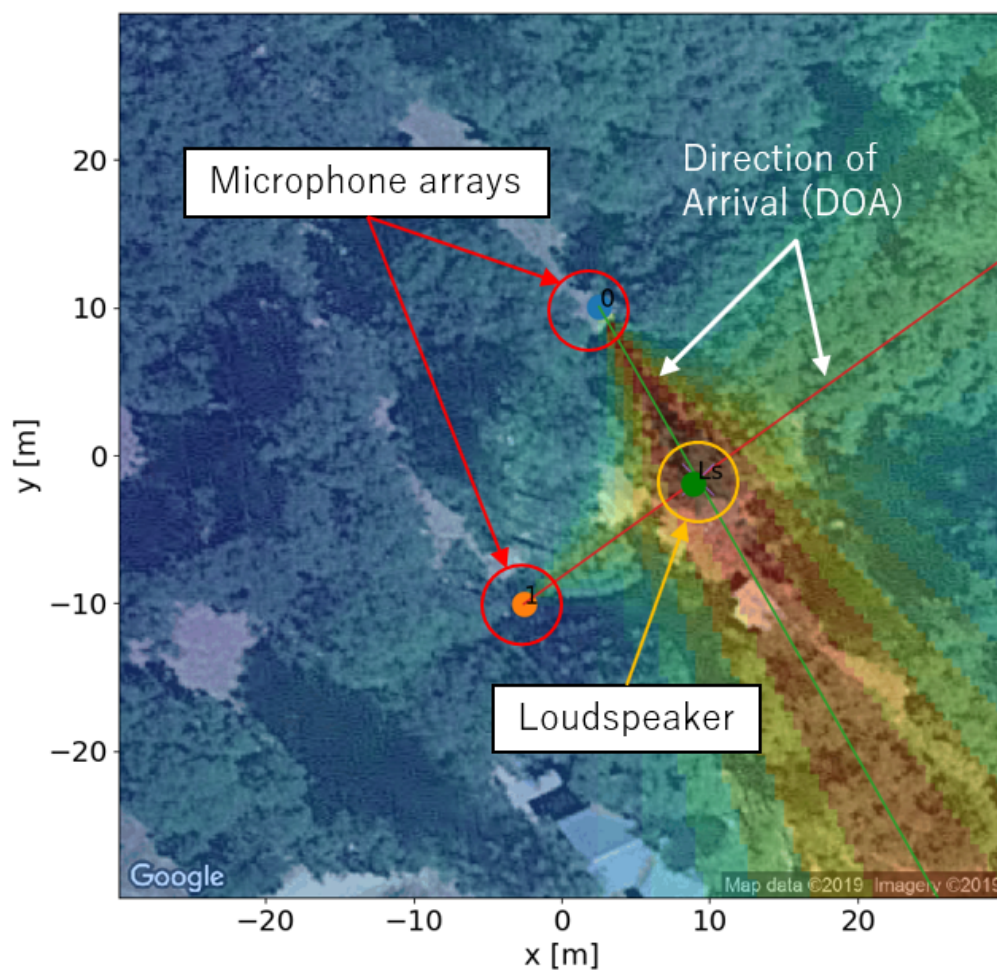


図 4.4: 2 台のマイクロホンアレイに基づく 2 次元定位. 画像全体の着色は MUSIC スペクトルの強度を示す. 青色は弱い強度を示し, 赤色は強い強度を示す. 2 次元定位は三角測量に基づく手法で行われ, 各マイクロホンアレイから定位方向へ半直線を伸ばし, 作られた交点を定位結果とする. この 2 次元定位は実時間処理でも実行可能で, その際にはパラメータで指定した PERIOD に基づく時間ステップで定位状況が更新される.

### 4.2.3 プレイバック実験

プレイバック実験には、ホシワキアカトウヒチョウの歌と地鳴きを用いた。この鳴き声は、実験の事前に録音を行った個体 A の歌 (図 4.5 (a)) と地鳴き (図 4.5 (k)) を採用しており、それぞれ Audacity により録音中に含まれるノイズの除去と正規化の処理を行った。処理後のそれぞれの録音の音圧を実験環境で測定すると、どちらも 1m 離れた場所から約 83.5dB ほどであった。この音圧は実際のホシワキアカトウヒチョウの鳴き声よりも大きいですが、プレイバック音を実験個体にできるだけ聞こえるようにするという意味ではプレイバック実験において適した音量であると考えられる。

これらの再生音を 2 個体の実験個体に対して 2 分ごとに 2 回再生することでプレイバック実験を行った。この再生条件は、本実験フィールドの鳥類生態を熟知した共同研究者の Martin Cody, Charles Taylor 両名誉教授 (University of California, Los Angeles) による提案を受けて採用した。このプレイバック実験に加え、プレイバックを行わない条件で録音を行い、合計で以下の 6 条件で録音を行った: 1) 個体 A への歌のプレイバック実験 (SA), 2) 個体 A への地鳴きのプレイバック実験 (CA), 3) 個体 A のプレイバック無しでの録音 (NA), 4) 個体 B への歌のプレイバック実験 (SB), 5) 個体 B への地鳴きプレイバック実験 (CB), 6) 個体 B のプレイバック無しでの録音 (NB)。それぞれの条件に関して、60 分の録音を行ったが、NA と NB に関しては録音中にシステムの不具合が発生したため、それぞれ 25 分、45 分だけの録音のみ得られた。そのため、プレイバックなしの条件では取得できた録音時間だけの分析を行った。

### 4.2.4 鳴き声の 2 次元定位

各実験条件で得られたそれぞれのマイクロホンアレイの録音から、HARKBird を用いて音源定位・分離処理を行った。その際に用いたパラメータは、以下のとおりである: THRESH = 25.0, NUM\_SOURCE = 1, UPPER\_BOUND\_FREQUENCY = 8,000, LOWER\_BOUND\_FREQUENCY = 2,200。このパラメータは、実験対象個体の鳴き声を漏れなく定位でき、かつ葉音や他の鳥類の鳴き声を出来るだけ定位しないように調整した。音源定位は、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を用いる MUSIC 法 [56] を採用し、その定位結果に基づいて GHSS 法 [57] によって対応する音源の抽出を行った。

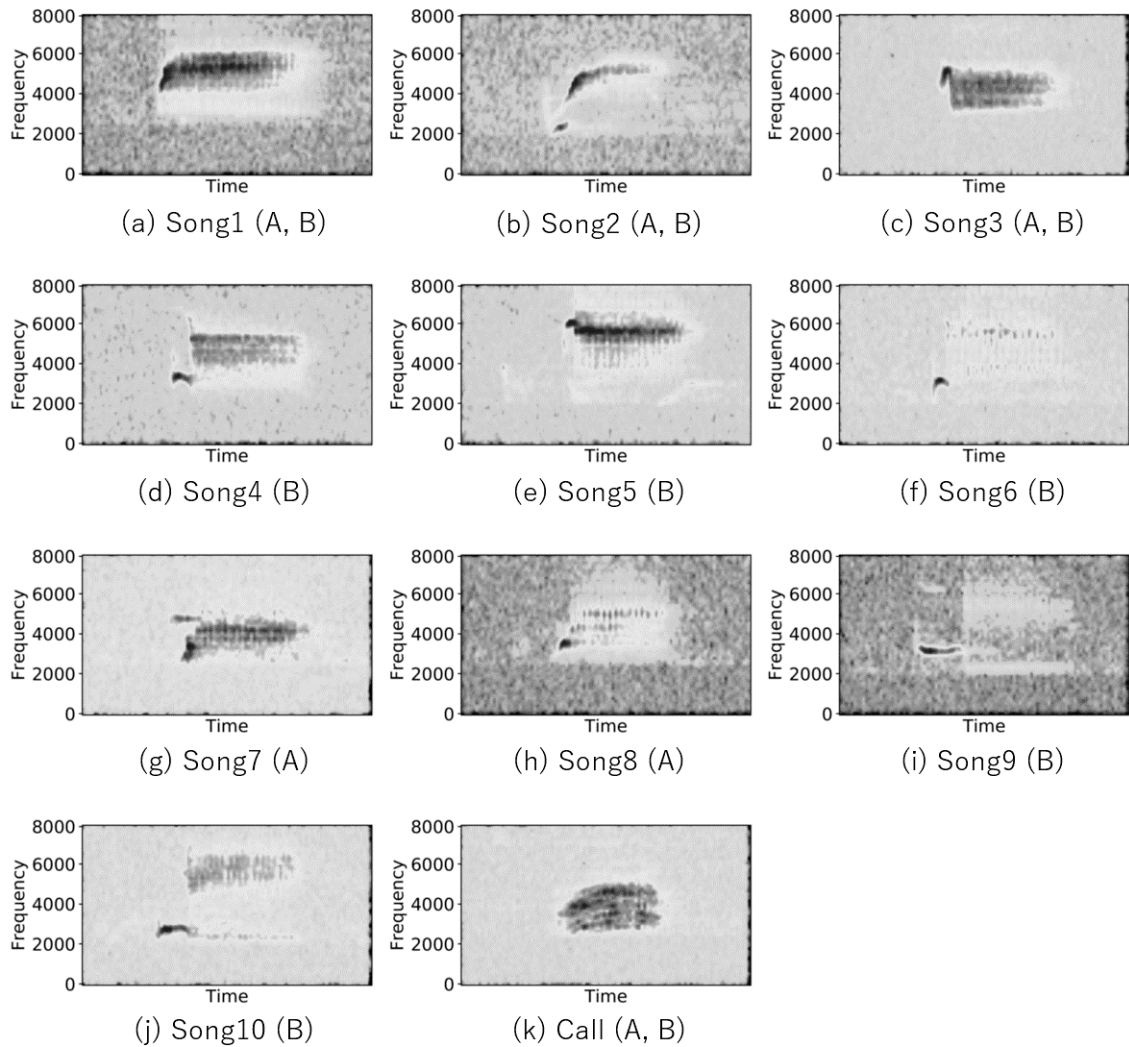


図 4.5: 実験中に 2 個体の実験個体から観測されたホシワキアカトウヒチョウの歌と地鳴き. プレイバック音には (a) の歌と (k) の地鳴きを用いている.

2次元定位は、最もシンプルな方法である三角測量の要領で行った。図4.6 (a)に、提案する2次元定位の模式図を示す。2次元定位は、各マイクロホンアレイの位置から定位した方向へ半直線を伸ばし、音源同士で交点が生じた場合にそれを2次元定位の結果とする(図4.6 (a)上)。それぞれの定位された音源は、個々に持続時間を持ち、その時間ステップごとに定位方向情報を持っており、2次元定位の結果は随時変化する情報であるため、今回は音源のペアが交点を作る時間を半分にする時点での音源定位結果を採用した(図4.6 (a)下)。

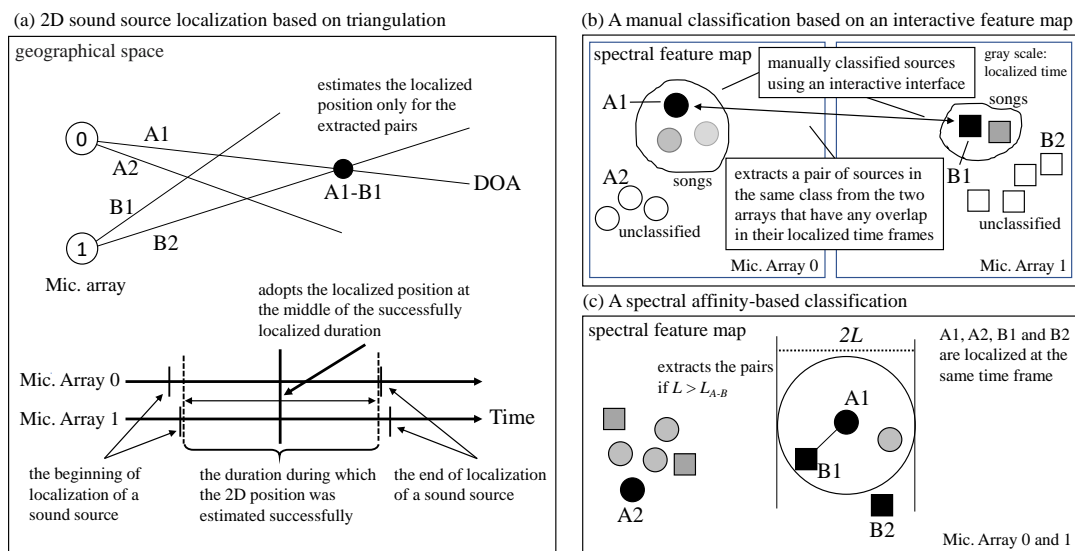


図 4.6: 2次元定位の説明図. (a) : 水平方向における三角測量に基づく2次元定位手法. (b) : 音響特性の特徴量空間を利用した音源の手動分類に基づく2次元定位の音源ペアの選定. (c) : 特徴量空間上の距離(類似度)に基づく2次元定位の音源ペアの選定.

実験背景で述べたように、2次元定位を行う際は音源のペアの選定が重要であり、各マイクロホンアレイからの音源は同じ音源でなければ本来存在しない音源を定位結果として扱ってしまう。そのため、音源の分類を行う必要があるが、従来の手法では鳥類の鳴き声以外のノイズも存在する時系列データを見ていく必要があったため、コストがかかることが問題であった。ここでは、この問題の解決として、t-SNE (t-distributed Stochastic Neighbor Embedding) [66] の低次元の特徴量空間を用いた2つの手法を提案する。



## 特徴量空間上での音源の手動分類に基づく 2次元定位

1つ目の方法は、各マイクロホンアレイから得られた定位音源を手動で分類する手法である。アルゴリズム 1 は、本手法による 2次元定位の手順について示したものである。最初に、各マイクロホンアレイの録音を用いて音源定位・音源分離を行い、分離音源を抽出する。その後、抽出した分離音源それぞれを  $100 \times 64$  のグレースケールスペクトログラム画像に変換する。それぞれの画像データのピクセルの濃淡をデータとして、t-SNE [66] を用いることで 2次元までデータを圧縮する。t-SNE は、元データの高次元空間と次元削減後の低次元空間それぞれで、データ同士の類似度を確率を用いて表現し、低次元に変換する際には高次元での関係性を保つ変換を行う。類似度の表現に関しては、高次元側では正規分布、低次元側では t 分布に基づくため、低次元側では高次元側よりも類似度の高いデータほど近傍に配置されるよう学習される。そのため、鳥類の鳴き声のような特徴のある音声は低次元空間上では近くに集まることが期待される。t-SNE の実行には、Python ライブラリの scikit-learn を用いた。パラメータに関しては、低次元の分布に影響を与えるパラメータをグリッドサーチで逐次変換しながら様々なパラメータで次元削減を実施し、2次元の分布の見た目上良く音源が分かれている結果を採用した。なお、特徴量空間の取得は各マイクロホンアレイのデータごとに行った。これは、特徴量空間上に多くの音源があると視覚的に分類が困難になる場合があり、それを避けるためである。

音源の分類は、図 4.7 に示すツールを作成することで低次元空間上で行えるようにした [70]。本ツールは、千葉らが開発したプロトタイプのカテゴリ分類ツール [49] に対して、上記のグリッドサーチの機能や、ラベル付けの際の配色機能等を拡張したものである。このツールでは、簡単なキーボード・マウス操作によって、音源のラベル付けを行うことができる。例えば、データの点群は各定位音源を示しているが、それぞれをクリックすることでデータに対応するスペクトログラム画像の表示と音源の再生が行われる。また、ラベル付けを行う際にはクリックした座標に基づいて多角形を作り、その多角形に囲まれる音源に対して一斉に同じラベルを付与することができる。これによって、従来手法よりも短時間で音源の分類が行えるようになった。本実験では、このツールを用いてそれぞれの音源を個体の鳴き声（ラベル名：Song 1 から Song 10 と Call）とプレイバック音（ラベル名：Playback）、その他の音（ラベル名：None）に分類した。その結果、実験中に 2 個体が鳴いた鳴き声は図 4.5 に示す 10 種類の歌と 1 種類の地鳴きに分類された。低次元空間上での分類後、2 台

---

**Algorithm 1** 手動分類に基づく 2次元定位.

---

```
for それぞれのマイクロホンアレイの録音について do
    音源定位を 0.2 秒ごとに行い, それぞれの定位音源の定位方向と定位時間を得る (MUSIC 法).
    分離音源を得る (GHDSS アルゴリズム).
end for
for それぞれのマイクロホンアレイについて do
    得られた分離音源を 100×64 のグレースケールのスペクトログラム画像に変換する.
    t-SNE で変換した分離音源の画像を 2次元に落とし込む.
    得られた 2次元特徴量空間の分布を参考に, 分離音源にラベル付けを行う.
end for
for 時間的重複のある異なるマイクロホンアレイで定位された音源に関して do
    if 同じラベルであれば then
        ペアとなる音源の定位方向を用いて 2次元定位を実行する.
        交点が出来ている時間の中間時点における交点を 2次元定位位置として採用する.
    end if
end for
    定位されたすべての 2次元定位結果に基づいて, 音源の空間的分布を作成.
```

---

のマイクロホンアレイそれぞれの分類結果を組み合わせる 2次元定位を行った (図 4.6 (b)). 2次元定位を行ったペアには, 同じラベルで分類された音源で, かつ定位時間に重複があるもの全てを採用した.

#### t-SNE の低次元特徴量を音源間の類似度と見なした 2次元定位

2つ目の手法は, t-SNE の特徴量空間上での距離を分離音源の類似度と見なし, その類似度に基づいて音源定位を行うものである. 図 4.6 (c) に, 2次元定位に用いるペアの選定方法を, アルゴリズム 2 に本手法での 2次元定位の流れを示す. 特徴量空間を作る以前の処理は, 分類ツールを用いて行った分類結果に基づく 2次元定位と同じである. 本手法では, 特徴量空間を作る際に 2台のマイクロホンアレイのデータを合わせて同じ低次元特徴量空間に落とし込んだ. 2次元定位に用いる音源のペアの選定は, 低次元特徴量空間上の距離を用いる. まず, それぞれの定位音源に関して, 定位時間が重複している音源を探し, 時間重複のある音源があった場合には注目音源と時間重複のある音源との距離を求める. これらの距離が任意の距離範囲  $L$  以内にある場合, このペアを用いて 2次元定位を行う. 上記の処理は, 音源を定位した時系列順に行うため, 既に音源定位のペアとして採用された音源は, 以降の処理でペアとして採用されたとしても 2次元定位は行わないようにした. これは, 同じ音源を重複して 2次元定位に利用しないようにするためである.

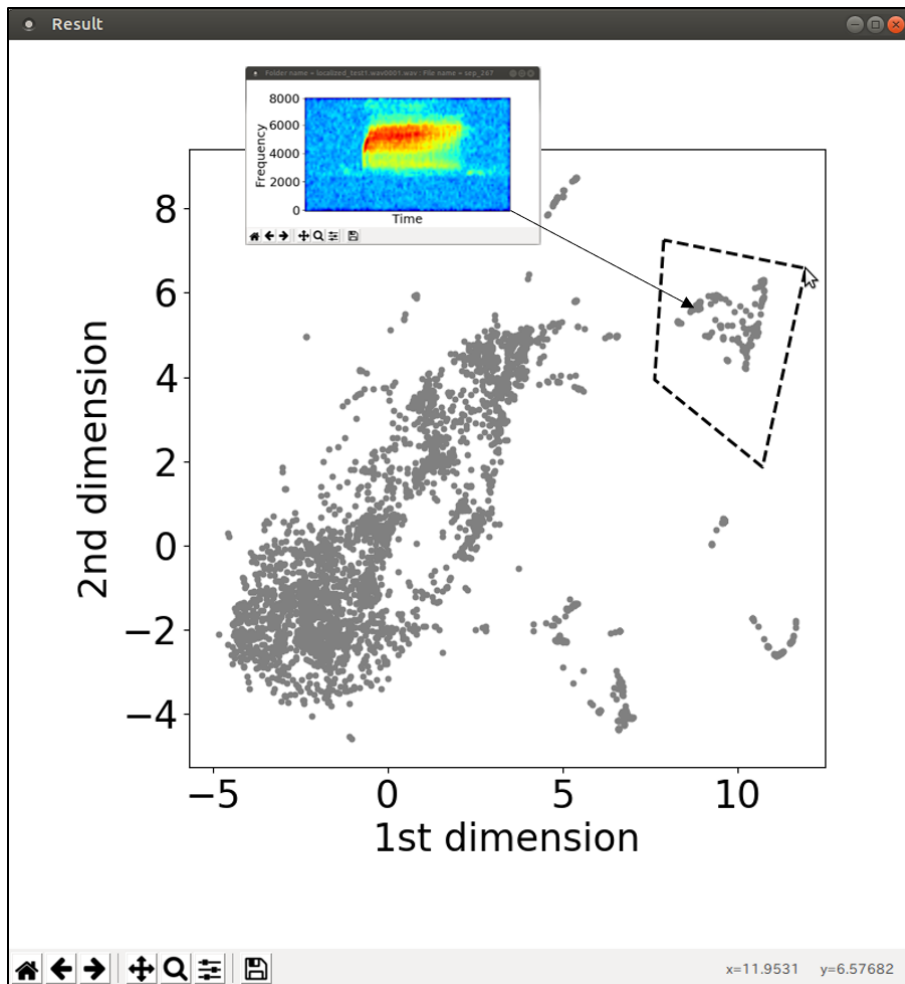


図 4.7: 特徴量空間に基づく分類ツール. 各点は音源を示し, クリックすることによって対応するスペクトログラム画像の表示と音源の再生ができる. ラベル付けの際には, 多角形で囲んだ範囲で一度に多くのラベル付けが行える.

---

**Algorithm 2** 音響特性の類似度に基づく 2次元定位.

---

```
for それぞれのマイクロホンアレイについて do
    音源定位を 0.2 秒ごとに行い, それぞれの定位音源の定位方向と定位時間を得る (MUSIC 法)
    分離音源を得る (GHDSS アルゴリズム).
end for
2 台のマイクロホンアレイで得られた全ての分離音源を 100×64 のグレースケールのスペクトログラム画像に変換する.
t-SNE で変換した分離音源の画像を 2 次元に落とし込む.
for 異なるマイクロホンアレイで得られた, 時間的重複のある音源ペアについて (時系列順にみていく). do
    特徴量空間上での距離を計算.
    if 求められた距離が  $L \in \{1, 2 \text{ or } 3\}$  であり, かつどちらの音源も 2 次元定位に用いられてない then
        2 つの音源の定位方向情報を用いて 2 次元定位を行う.
        交点が出来ている時間の中間時点における交点を 2 次元定位位置として採用する.
    end if
end for
定位されたすべての 2 次元定位結果に基づいて, 音源の空間的分布を作成.
```

---

t-SNE の特徴から, 音響的特徴のないノイズは特徴量空間上で広く分布し, 音源間の距離が大きくなる傾向がある. そのため, 本手法は不要な雑音の定位を妨げ, 鳥類の鳴き声などの特徴のある音声のみを定位できることが期待される.

#### 4.2.5 分析

分析は, 特徴量空間を用いた手動分類に基づく 2次元定位結果と, 音源の類似度に基づいた 2次元定位結果でそれぞれ分けて行う. 前者は, 実験個体の鳴き声のみが抽出された実験結果である. そのため, 結果を用いて実験の最中の個体の発声に関する時空間ダイナミクスの調査を行った. 具体的には, 定位された鳴き声の空間的分布や頻度, スピーカとの距離や移動頻度などを調べ, 再生条件間での行動パターンを比較することによってプレイバックの影響を調査した. 後者では, 本手法によって得られた 2次元定位結果と前者の定位結果を比較することで, 定位結果が音環境の状況をどの程度把握しているかを調査した. これは, 実験の最中は実験個体が最も鳴いていたため, 提案手法によって音環境の把握ができるのであれば, 2次元定位の結果は実験個体の歌を抽出した前者手法の定位結果を反映したものとなるという仮定のもと行った. 発声回数などを定量的に比較することによって詳細にその精度を議論した.

## 4.3 分析結果

### 4.3.1 t-SNE の低次元特徴量空間を用いた分類手法の評価

図 4.8 は、用いた 2 台のうちの 1 台のマイクロホンアレイによって得られた定位音源に関して、特徴量空間を作り、分類ツールによって手動分類を行った結果である。ほとんどの実験条件において、実験個体の鳴き声 (Song 1 から Song 10 と Call) とプレイバック音 (Playback) がそれぞれ小さなクラスタを形成していることが確認できる。None で示される対象個体の鳴き声とプレイバック音以外の定位音源は、大きなクラスタを形成している。この結果は、t-SNE を用いて作成した特徴量空間がノイズの除去や注目個体の鳴き声の抽出 (分類) に貢献することを示している。

一方で、個体 B の実験の歌再生条件では、黄色で示される Call (図 4.8 Song playback Individual B, 右上付近) が、プレイバックのなしの条件では、Song 1 や Call (図 4.8 Normal condition Individual B, 右下付近) がノイズのクラスタに混ざるように分布している。これらの分離音源のスペクトログラム画像を確認してみると、鳴き声の波形が弱く、背景ノイズが強く出ているような画像であった。このことから、個体がマイクロホンアレイから離れて鳴いた鳴き声が定位はされたものの、鳴き声の特徴が低次元特徴量に反映されなかったために生じたものだと考えられる。また、各条件を比較してみると、概して注目個体の発声が多いほど、鳴き声のクラスタがはっきりと構成されていることがわかる。そのため、本分類ツールは実験個体の発声が多く、ノイズが少ないほどうまく機能することが示唆される。

他方、None のクラスタに関して、このクラスタには葉音などのノイズの他に、チャバライカル (*Pheucticus melanocephalus*) という鳥類種の歌が含まれていることが確認された。鳥類の歌とノイズが同じクラスタに含まれたのは、この鳥がマイクロホンアレイから遠い場所で鳴いていたために分離音にはノイズ成分が多く現れたこと、あるいは音源定位・分離のパラメータ設定がホシワキアカトウヒチョウの鳴き声のみをうまく抽出出来るような設定であり、抽出の際にホシワキアカトウヒチョウ以外の鳴き声はうまく抽出出来なかったなどの要因が挙げられる。以上のことから、本ツールを鳥類観測に用いる際には適切なパラメータ設定が必要であるといえる。

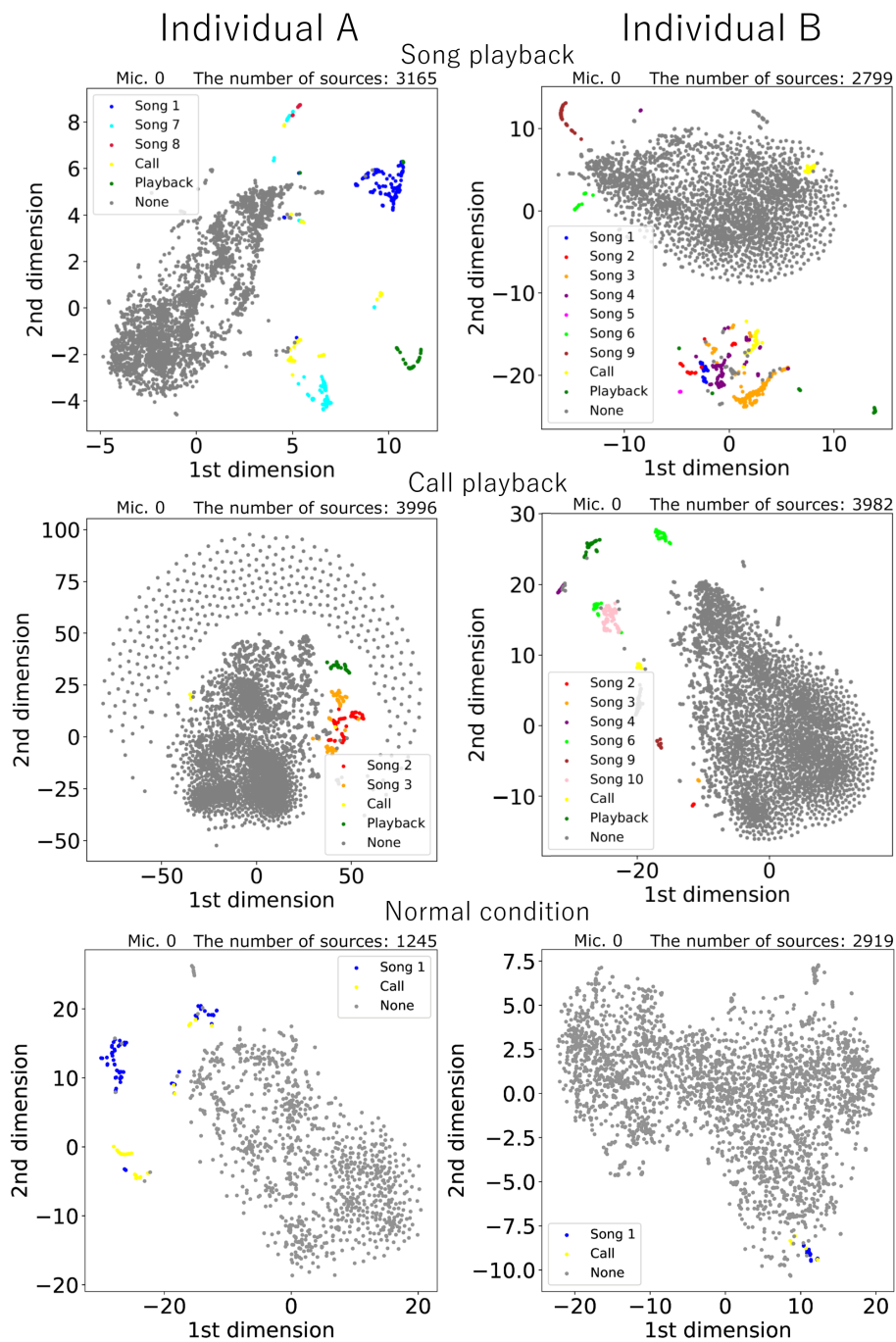


図 4.8: 各実験条件における分離音源の分類結果. 各点は HARKBird を用いて定位した音源を示す. また, 色付けされた音源は実験個体の鳴き声, あるいはプレイバック音を示す. 灰色の音源は鳴き声やプレイバック音ではないノイズなどの音である.

### 4.3.2 鳴き声に関する時空間ダイナミクスの分析

前項の手動による分類結果に基づいて、2 個体の実験個体の発声に関する時空間ダイナミクスの抽出を行った結果を図 4.9-4.11 に示す。まず、全体的な結果について述べていく。図 4.9 (a) 及び図 4.10 (a) は、各条件での実験個体の鳴き声とプレイバック音の 2 次元定位を航空写真上に示したものである。この結果から、実験個体がどこで、何の種類の鳴き声を鳴いたかおおよそ把握できていることがわかる。この定位結果に着目してみると、実験個体は概してスピーカの周りで鳴いている傾向があることがわかる。この傾向について、定位した鳴き声に関してスピーカからみてどの程度の距離で鳴いたかの平均距離を出すと、各条件でそれぞれ以下の通りであった：SA : 13.5 m, CA : 27.8 m, NA : 17.4 m, SB : 15.7 m, CB : 26.1 m, NB : 18.8 m。さらに、発声の時系列パターンについて着目すると、すべての条件において、対象個体はある鳴き声を一定回数連続して鳴く傾向があることが確認できる。これは、本実験対象種の発声パターンの傾向が抽出できたことを意味する。

次に、プレイバックの影響に関して述べる。図 4.11 (b) - (d) は実験中の実験個体の発声に関しての定量的な分析結果である。この結果からは、2 個体の実験個体はプレイバックの条件に応じて以下のように行動を変えていることが示唆された：1) 歌を再生した条件は、地鳴きを再生した条件よりも多くの種類の鳴き声を頻繁に鳴く。2) 歌を再生した条件は地鳴きを再生した条件よりもスピーカの近くで鳴く。3) 歌を再生した条件では、地鳴きを再生した条件よりもスピーカからみた個体の発声ごとの定位方向の変化量が大きい。スピーカからみた距離、方向変化量には、観測データをもとに 10 万回のブートストラップ法によって求めた標準偏差を示している。それぞれ各条件での統計的な有意差は見られなかったが、平均値や信頼区間の幅から考慮してもプレイバックの影響に差がある傾向は確認できる。プレイバックを行っていない条件では、分析した時間が短いということを考慮しても発声回数は少ない。このことから、地鳴きを再生した条件においてもプレイバックの影響を受けていることが示唆される。

また、個体 B は個体 A よりも発声の頻度や種類が多く、再生音の影響を強く受けていることが示唆される。この行動の違いは、実験に用いた鳴き声が個体 A の鳴き声であるということに起因すると考えられる。なぜなら、プレイバックの状況は個体 B にとっては自分の縄張りに近隣個体が侵入してきた状況であり、頻繁に動き回りながら多くの鳴き声を発することで警戒していた状況であることが推測されるか

らである。

次に、2次元定位に関する結果について述べていく。図4.11 (a) は、実験で用いたプレイバック音が実際の測定値の位置と比較してどの程度離れていたかを求め、得られたデータの標準偏差を示したものである。結果はそれぞれ、SA:  $0.867 \pm 0.616$  m, CA:  $1.613 \pm 0.494$  m, SB:  $1.315 \pm 0.813$  m, CB:  $1.510 \pm 0.26$  m という値を示した。すべての条件でプレイバック音の定位誤差は2.0 m 以内に収まっており、実験個体の鳴き声も同様の誤差で定位できているとすれば、得られた傾向は実際の傾向を反映しているといえる。一方で、図4.10の地鳴きを再生した条件に関して、スピーカからの距離と時間の関係を示した図4.10 (b) を見てみると、2300~2600秒付近で不自然な短時間での定位距離の変動が見られる。この変動は、2台のマイクロホンアレイを繋ぐ直線上付近での2次元定位が各マイクロホンアレイの定位角度の変動の影響を強く受けることに起因する。

### 4.3.3 音源の類似度に基づいた2次元定位による音風景の観測

図4.12に、音源の類似度に基づいて2次元定位のペアを決定する手法によって行った2次元定位の結果を示す。図4.12 (a) は、時間重複のある全てのペアに2次元定位を行った結果、図4.12 (b) は特徴量空間上でのペアとして採用する距離制限を $L=2$ として提案手法を用いて行った2次元定位結果、図4.12 (c) はt-SNEで得た分離音源の2次元特徴量空間上の分布である。これらの結果は、プレイバック実験において最も頻繁に実験個体が鳴き、移動していたSB条件での結果である。図4.12 (b) の定位音源の色には、図4.12 (c) の特徴量空間上で採用されたペアの中点の座標に対応する図4.12 (d) の色が採用されている。ペアの選定を行わない場合、非常に多くの音源が定位されており、これは実際の音風景を反映しているようには見えない。一方で、提案手法による結果は、多少ばらけた音源はあるものの、図4.10のSB条件での2次元定位結果の実験個体の発声分布を反映していることがわかる。また、音源定位の色を見ると、定位場所ごとに似た色が集中しており、これは定位された鳴き声の種類ごとに異なる低次元特徴量空間上の座標位置が反映された結果である。

この定位手法の評価のために、手動分類によって抽出した実験個体の発声を実際の音環境の状況であると仮定して、特徴量空間上でのペア採用のための距離制限を変えて、提案手法による実験個体の発声の抽出精度を求めた(表4.1)。 $L$ は定位の



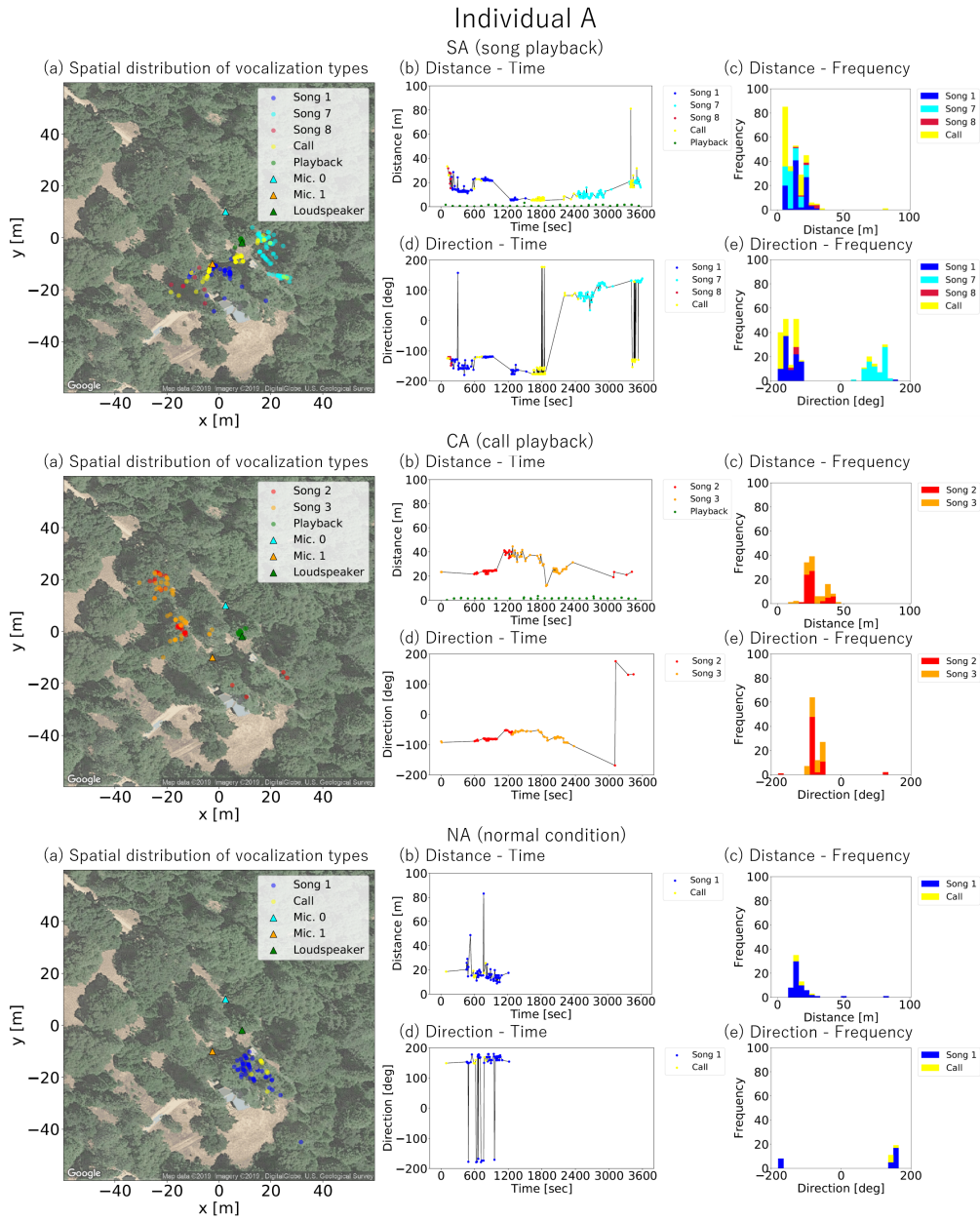


図 4.9: 個体 A の鳴き声に関する空間的・音響的・時間的ダイナミクスの抽出結果。  
 (a): 鳴き声の空間的分布. (b): スピーカからみた定位した鳴き声の距離と方向. (c):  
 定位した鳴き声の距離と方向のヒストグラム.

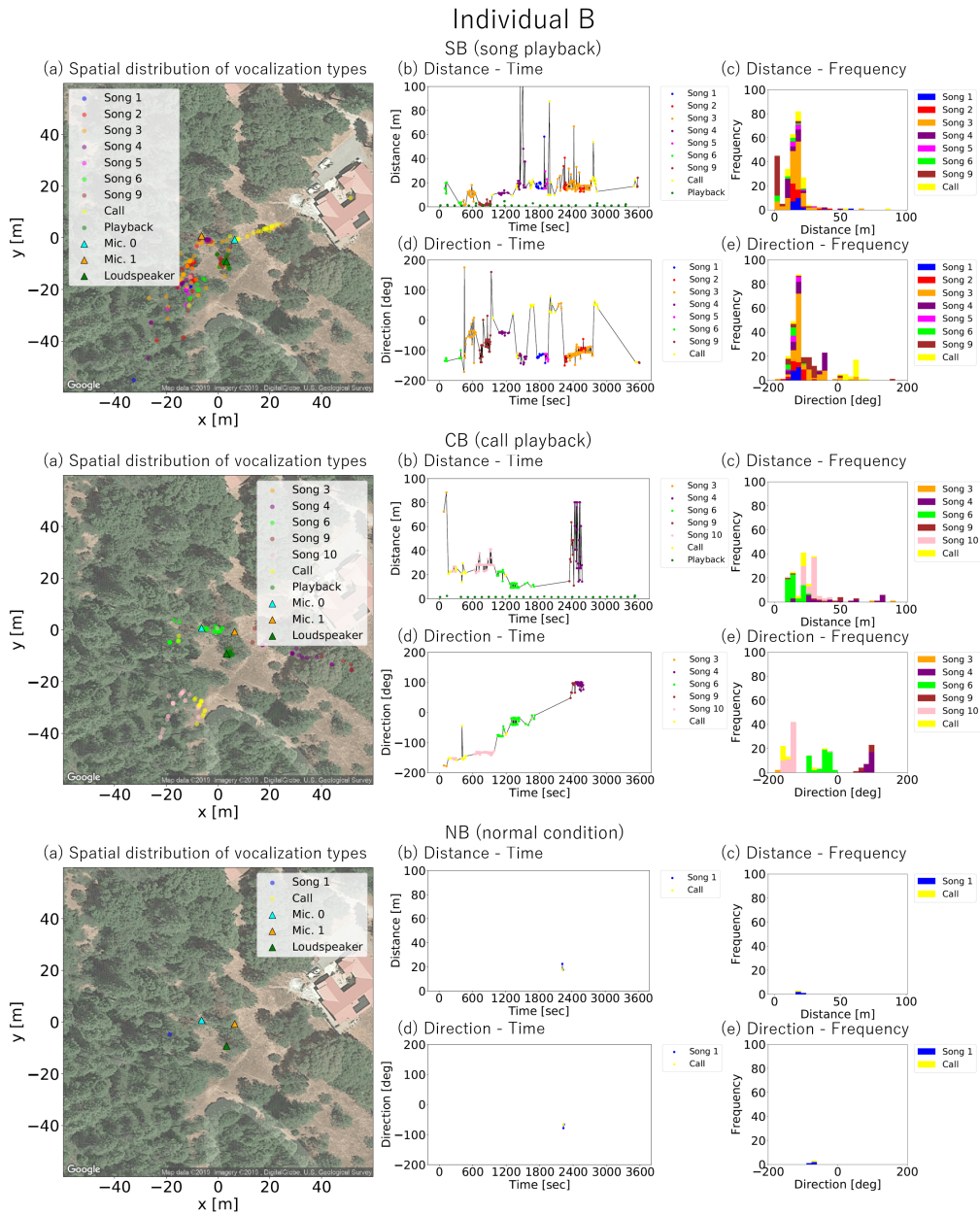


図 4.10: 個体 B の鳴き声に関する空間的・音響的・時間的ダイナミクスの抽出結果。

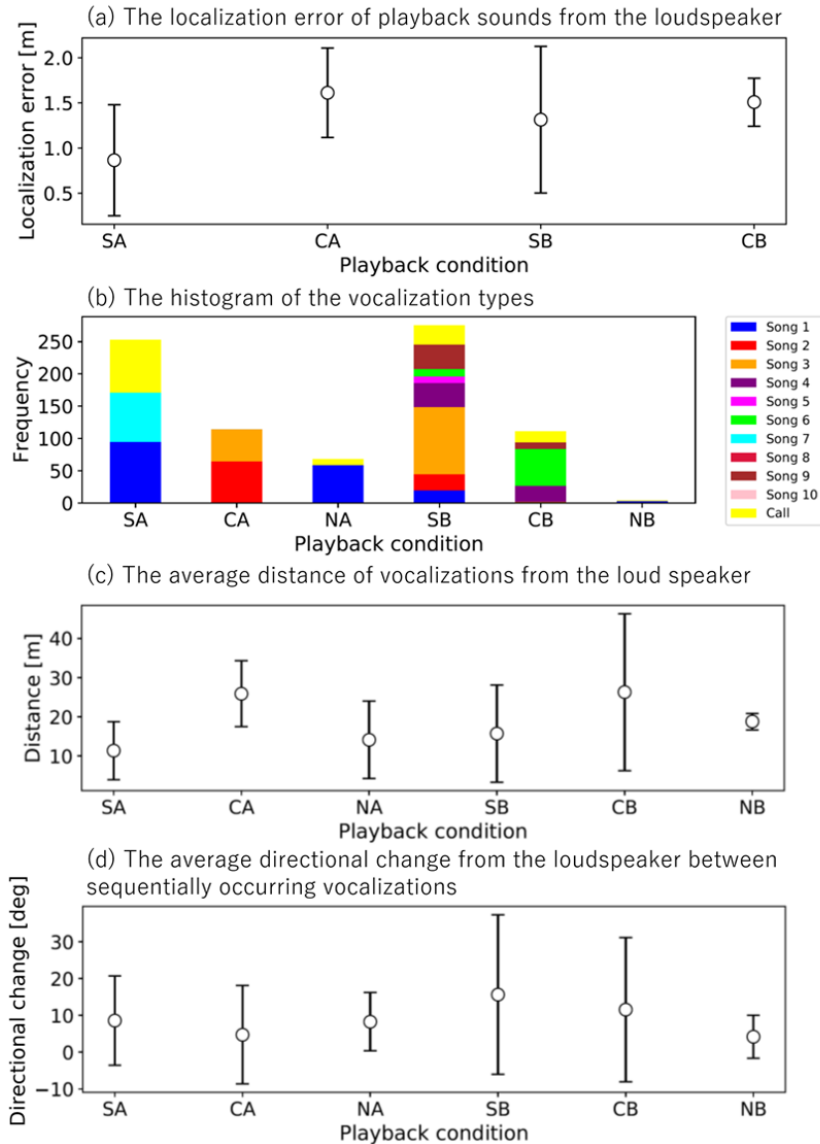


図 4.11: 各プレイバック条件における実験個体の鳴き声に関する分析結果. (a): 測定値に基づくプレイバック音の2次元定位の定位誤差. (b): 鳴き声の頻度. (c): スピーカからみた鳴き声の定位方向の変化量. (d): スピーカから鳴き声の定位位置までの距離. (a), (c) および (d) のエラーバーは, 標準偏差を示す.

状況を確認しながら変更し、条件ごとに適したパラメータで実施している。結果は、ペア採用に関する閾値  $L$  に大きく依存することがわかる。例えば、CB の条件に着目すると、 $L=3.0$  の場合は 85.1% と多くの実験個体の鳴き声の抽出に成功している一方で、 $L=2.0$ ,  $L=1.0$  では 53.1%, 5.1% と大きく変動している。しかしながら、 $L$  を高くするほど音風景を詳細に把握できるというわけではない。それは、ノイズの定位も  $L$  の増加に従って増えるためである。例えば、図 4.12 (b) で示した  $L=2.0$  の場合の SB の結果は目視によって定位されたノイズが少なく、かつ実験個体の鳴き声が比較的よく抽出された結果を挙げているが、抽出できた鳴き声は  $L=3.0$  の場合よりも少ない。そのため、ノイズの除去と、実際の音風景の把握はトレードオフの関係にあることが考察される。しかし、各条件において、各マイクロホンアレイで定位した音源の数が 3000~4000 ほどあったことを考えると、本手法によって不要な音源の定位を大幅に抑制しているといえる。また、NB の条件のように対象個体の発声が少ない場合は鳴き声の抽出は全くできておらず、本手法が十分に機能しないことが示唆された。

#### 4.4 議論

本章では、ロボット聴覚技術による 2 次元定位において必要な音源の分類タスクのコスト削減の提案と、2 次元定位が鳥類発声行動に基づく時空間的ダイナミクスの理解に貢献しうるかの検討を目的として、ホシワキアカトウヒチョウに対するプレイバック実験を題材に提案手法の評価、プレイバックの影響の定量的分析を行った。

2 次元定位に関する分類手法として次元削減アルゴリズム t-SNE に基づく低次元特徴量空間上での分類手法の提案と、低次元特徴量上での距離を音源の類似度と見なした音源定位手法の提案を行った。前者では、音源の特徴を反映した特徴量空間上での音源の手動分類を実現し、鳴き声の種類ごとに形成されたクラスごとにラベル付けが容易にできるようになった。これは、時系列データを追って行っていた従来手法と比較してもコストのかからない手法である。後者は、特徴量空間上での距離が近い音源同士で 2 次元定位を実施する手法であるが、これによって得られた定位結果は、手動分類の結果をもとに得た実験個体の鳴き声の分布をうまく反映した。この結果から、コストのかかる分類を行わずとも音風景の理解が可能であることが示唆された。双方の手法において、t-SNE に基づく特徴量空間は、ノイズの排除と意味のある音源（ここでは対象個体の鳴き声）の抽出に貢献している。これは

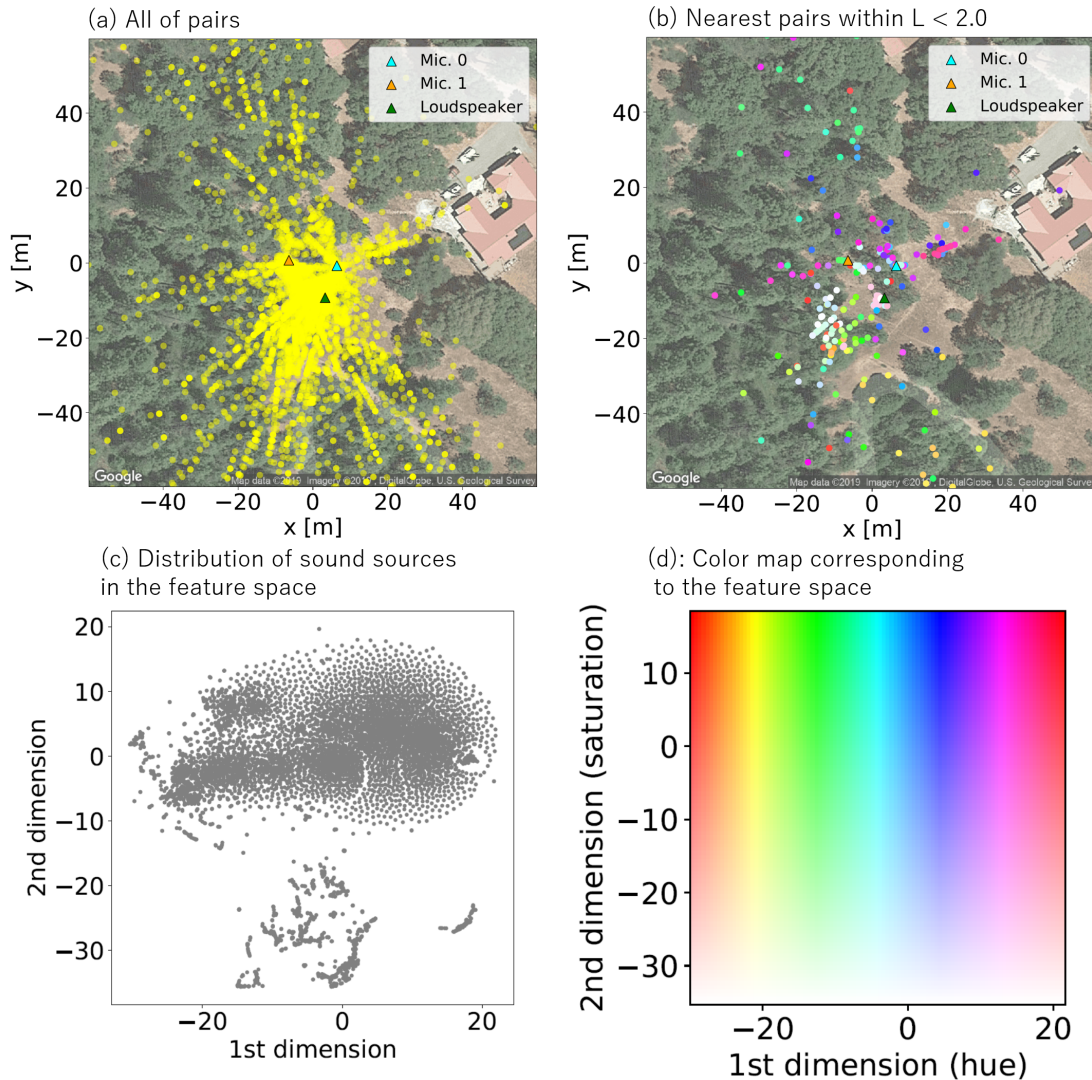


図 4.12: 音響特性の類似度に基づく 2次元定位手法によって得られた 2次元定位結果. (a): すべての可能なペアで 2次元定位を行った場合の結果. (b): 提案手法 ( $L=2$ ) による 2次元定位結果. (c): 特徴量空間上での音源の分布. (d): (b) の 2次元定位結果に用いられた配色. 2次元定位には音源ペアの特徴量空間上での座標の中点の色が採用されている.

表 4.1: 音響特性の類似度に基づいた 2 次元定位手法による実験個体の鳴き声の定位成功割合. 手動分類によって得られた結果を正解データとして, その音源の中で定位できている割合を求めている.

playback condition	$L$	localized vocalizations (affinity-based)	localized vocalizations (manual annotation)	accuracy (%)
SA	1.0	21	261	8.0
	2.0	72		27.6
	3.0	118		45.2
CA	0.5	40	114	35.1
	0.75	54		47.4
	1.0	59		51.8
NA	0.5	20	68	29.4
	0.75	26		38.2
	1.0	32		47.1
SB	1.0	100	275	36.4
	2.0	135		49.1
	3.0	165		60.0
CB	1.0	9	175	5.1
	2.0	93		53.1
	3.0	149		85.1
NB	0.5	0	4	0
	1.0	0		0
	1.5	0		0

t-SNE の類似するものが近くに配置するという特徴によって, 特徴のある鳴き声の音源はそれぞれが密になって集まるが, ノイズなどの特徴のない音源は広く散らばりがちになったことによつてうまく実現した結果である.

生物音響学の分野では, 今回採用した t-SNE のような次元削減アルゴリズムや, 潜在空間上での動物音声の可視化などの応用が注目されている. 例えば, Sainburg は上記の目的のために Animal vocalization generative network (AVGN) と呼ばれるパッケージを提案している [71]. このパッケージには, 動物の発声録音を音節ごとに分けて, これを生成モデルなどの機械学習を用いて学習し, 学習した潜在空間内で可視化したり, 新しい音節を生成する機能などがある. その他にも, 教師なし学習を用いた音節のクラスタリングや, 音節の順序を可視化するための機能などもあ

り，特に鳥類の音声分析において有用に利用できるという例を示している．今回提案した音源の分類手法も，AVGN で利用されている同様の技術に基づくものではあるが，利用方法に関しては，複数のマイクロホンアレイを用いた録音の際に生じる問題を解決するものであり，いわば技術の異なる活用方法を提案している．この提案は，これまでの関連研究では議論されておらず，そういった意味でも，機械学習の新しい分野への適用可能性を示すことができたといえる．

プレイバック実験の分析結果からは，実験個体は，再生条件に応じて挙動を変えている様子を観測することができた．それは，地鳴きを再生した条件よりも歌を再生した条件の方がより攻撃的な反応を示すという，生態学的にも妥当性のある結果であった．これらの傾向は，発声の頻度や種類，スピーカから実験個体までの距離や方向といった情報で定量的に示されたことであり，ロボット聴覚技術を用いた2次元定位による鳥類観測が，鳥類の発声行動に基づく時空間的ダイナミクスの理解に貢献することを示唆する．実験個体が鳴く場所によってはうまく音源定位ができないという問題が発生したものの，本研究で提案したノート PC と2台のマイクロホンアレイという最小限のシステムで，鳴き声に関する時空間的ダイナミクスの抽出が行えたのは注目すべき点であるといえる．

# 第5章 複数のマイクロホンアレイを用いたキンカチヨウ集団における 個体間相互作用の観測

## 5.1 背景と目的

前章では、ロボット聴覚技術に基づく音源の2次元定位が鳥類発声行動に基づく時空間的ダイナミクスの理解に貢献しうるかどうかの検討として、2個体のホシワキアカトウヒチヨウに対してプレイバック実験を行った。分析結果からは、プレイバックの状況による実験個体の行動の違いが定量的に示され、生態学的条件に応じた行動ダイナミクスの観測に成功した。また、半自動的な手法によって音環境の概観も実現し、ロボット聴覚技術の幅広い適用可能性が示された。

これまでの実験における課題を整理する。発声行動の観測という意味で大きな課題としては2点挙げられる。1つには、マイクロホンアレイから遠く離れた鳴き声は、音源分離がうまくできず、鳴き声の特徴が十分に抽出できないという点である。その場合、前章で提案した次元削減アルゴリズムに基づく分類手法はうまく機能せず、ノイズと鳴き声を分けることができなかった。また、これまで行ってきたプレイバック実験は、個体の縄張り内という比較的实验個体が遠くへ行かない状況のもとで行われた実験であるが、実際の状況を考えると、縄張り内に他個体が侵入して鳴く状況はかなり限定された状況である。より広い文脈での鳴き声に関する時空間的ダイナミクスの理解にはより広い範囲での観測が必要となるが、その場合上記の問題がより顕著に生じる可能性がある。もう1つの課題としては、提案した2次元定位手法では、音源の場所によって本来の音源位置から定位位置が大きく変化してしまうという課題がある。これは、2台のマイクロホンアレイを結ぶ直線上付近で生じる問題である。上記2点の課題を解決する方法としては、より多くのマイクロホンアレイを用いて録音を行うことが挙げられる。例えば、一定間隔でマイクロホ



ンアレイを配置し、それぞれのマイクロホンアレイごとに定位範囲を定めれば、広範囲での音源定位が可能となる。また、その際にマイクロホンアレイを繋ぐ直線上での定位を避けるようにすれば2つ目の課題も解決できる。

上記の背景を踏まえ、本章では複数台のマイクロホンアレイを用いた音源定位手法の検討と、これまでの観測の発展としてより具体的、複雑な状況での鳥類の発声行動の観測・理解を目的として、本来の興味である複数鳥類が鳴き合う状況での個体間相互作用の理解に向けた調査を行った。具体的には、空間的に限定された屋外テント内に、鳥類が自由に飛び回れるほどの観測環境（フリーフライト環境）を構築し、その中に複数個体のキンカチョウ（*Taeniopygia guttata*）を放ち、鳴き合う様子を複数台のマイクロホンアレイで録音・観測する実験を行った。

本実験は、小規模鳥類集団における鳴き声に関する時空間的ダイナミクスの抽出を検討した実験、社会関係に焦点を当てた個体間相互作用の理解に関する実験の2つに分かれる。前者では、5個体のキンカチョウを観測環境に放ち、鳴き合う様子を複数台のマイクロホンアレイで録音・観測した。本章の実験は、これまでの自然環境で行ってきた実験と比較しても観測範囲は狭く、空間的に制限された環境での観測であり、これは初めての試みである。そのため、このような状況でもキンカチョウ個体が鳴く様子を抽出可能か、また複数のマイクロホンアレイによる観測に関する提案手法がうまく機能するかについて検討した。さらに、得られた定位結果から示唆されるキンカチョウの発声行動における傾向の議論と、簡単な手法による歌と地鳴きの分類手法の提案を行った。その結果、複数マイクロホンアレイによって詳細な音源位置の定位の実現ならびに雑音の定位を低減、低コストな手法で歌・地鳴きの区別ができること示し、マイクロホンアレイによる非侵襲的な観測手法によってキンカチョウ集団が鳴き合う様子を観測できることを確認した。

後者では、実際の社会状況を想定した模擬環境での録音調査を行った。具体的には、テント内のキンカチョウのオス・メスの個体数を变化させた様々な条件で録音実験を行い、キンカチョウの発声・移動傾向がどのように変化するか調査した。前者の実験では、上述の成果は得られたものの、個体識別という個体間相互作用の観測には欠かせない課題が残されていた。そこで、本研究では個体識別と歌・地鳴きの判別課題の解決として、変分オートエンコーダ（Variational Autoencoder, VAE）[72]と、サポートベクターマシン（Support Vector Machine, SVM）[73]を組み合わせた分類手法を提案した。その後、提案した分類手法の結果に基づいて、オス1羽放鳥時、オス2羽放鳥時、オス2羽・メス1羽放鳥時の録音に着目して分析を行った。

その結果、条件によって発声の空間的分布や頻度に違いがあり、特にオス個体間の関係に変化が生じることが観測された。以降の節からは、キンカチョウに関する知見を述べた後、上記の2つの実験に関してそれぞれ分けて報告を行う。なお、本章の観測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）を得て実施した。

## 5.2 キンカチョウ

### 5.2.1 生態及び鳴き声

キンカチョウ (*Taeniopygia guttata*) (図5.1) は、ペットとしてもよく飼育されるオーストラリアを原産とする鳴禽類である。野生では群れを成して生活し（集団営巣性鳥類）、一夫一妻制でペアを形成する種である。キンカチョウは、個体特有の特徴を持った歌（オスのみ）と、様々な種類の地鳴き（雌雄）を発声することが知られている。歌に関しては、発声学習をすることによって自身の歌を確立することが知られており [74–76]、発声学習のモデル生物として神経行動学、動物行動学、言語学など様々な分野で研究が行われている [77–80]。歌は、メスに向けたホッピングなどのようなダンスを行いながら歌う directed song と、単独で行う、ダンスなどの行動なしで歌う undirected song があり、この2種の歌は音響的にも異なることが報告されている [81]。また、オスの鳴く歌はペアの形成において重要であるが、ペアの絆が十分に形成されると意味を失うことも示唆されている [82]。

地鳴きは、個体で特徴のあるものや、個体間で比較的差のないものなど様々な種類があり、これら多数の地鳴きを使い分けながら発声することによって個体間の結合の確認や、警戒などの意思を示すことが知られ [83,84]、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として扱われる [24,26,85,86]。Elie らの実施したキンカチョウの鳴き声の分類によると [84]、地鳴きは、ヒナが鳴く地鳴き (Begging call; 餌の要求, Long Tonal call; 親個体や兄弟個体との確認)、結合の確認や、営巣の際などに用いられる個体間のコミュニケーションにおいて用いられる地鳴き (Tet call および Stack call; 近隣個体との確認, Distance call; 遠方個体との確認, Nest call および Whine call; 営巣時)、同種への攻撃の際、あるいは攻撃された際などに用いられる敵対性を意味する地鳴き (Wsst call; 攻撃時, Distress call; 被攻撃時, Alarm call; 警戒時) に分けられる。これらの地鳴きの中で、distance call は特

に個体間で鳴き声の構造に差がある鳴き声であり [87], この鳴き声によって個体認識が行われているという報告がある [88,89]. 一方で, 飛行時や軽い警戒時, 求愛など様々な場面で発することが報告されていることから, 特定の場面で用いるような鳴き声ではないようである.



図 5.1: キンカチョウ (*Taeniopygia guttata*).

## 5.2.2 先行研究

次に, キンカチョウの個体間相互作用に関する実験についての先行研究を紹介する. 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, 発声状況を分析し, ペアの形成とその他個体間の関係の調査を行った [86]. 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, 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. また, 形成されたペアは時間が経過するにつれて接近する頻度が多くなった一方で, オスが鳴くメスへ向けた歌頻度が減っていく傾向が確認された. また, Gill らや D'Amelio らの研究では, キンカチョウのペアの形成過程における発声行動やペア間の接近などの接触行動の変化を鳥類個体に取り付けた小型録音機による録音とビデオ撮影によって調査している [24,85]. 前者の研究では, 繁殖段階に応じてペア間の相互作用で用いる鳴き声に変化していくことが示唆されている. また, 後者の研究では, 近距離での

確認の際に用いられる stack call がペアの相互作用に重要な役割を果たしており，確立したペアでは，初対面の状況から新しく形成されたペアよりオスとメスとで回数に差のない地鳴きで相互作用し合い，より多くの接触行動を行う傾向があると報告している．

上記の研究のように，キンカチョウの相互作用に関する知見は様々な手法によって多く報告されているが，多くの場合はケージ内といった実験室的環境で行われたもので，相互作用において重要となる空間的相互作用の抽出にも限界がある．また，1章で述べたように，マイクロホンを個体に背負わせて行う録音手法では，空間的な相互作用の抽出ができず，個体への侵襲性も懸念される．一方，マイクロホンアレイを用いた観測によって従来手法で生じる課題の解決が可能であると考えられる．また，キンカチョウは，野生では集団で生活する種（集団営巣性鳥類）であり，歌といくつかの地鳴きを用いて個体間で様々な社会関係を構築する．そのため，テント内という制限された空間内においても，比較的野生に近い状況での個体間相互作用や社会関係の変化の観測が期待されるため，本実験の対象種として理想的である．

### 5.3 小規模鳥類集団における鳴き声の時空間ダイナミクスの抽出と分析

#### 5.3.1 実験手法

##### 環境構築と録音実験

録音実験は，北海道大学札幌キャンパス内にある圃場（43°4'18.3"N, 141°20'28.4"E）に，約7m四方のメッシュテントを設営して行った（図5.2）．この環境は，キンカチョウが自由に飛び回るには十分な広さである．図5.3に構築した観測環境の配置図を示す．藁で作られた人工的な巣5個と，止まり木4本を発声位置が音源定位位置から把握できるようにそれぞれ離して配置した．また，テントの中央付近には水と飼料を置いた餌場を配置した．録音を行うためのマイクロホンアレイ（TAMAGO-03, 株式会社システムインフロンティア）は，テントの中央に1台，テントの角付近にそれぞれ1台ずつの計5台を配置し，1台のノートPC（Toughbook CF-C2, Panasonic 株式会社）にUSB接続した．巣，止まり木，マイクロホンアレイは，高さの差による影響が定位結果に大きな影響を生じさせないように地上から約1.5mの高さにな

るように調整した。

録音実験は、2019年6月3日に実施した。実験環境に、オスのキンカチョウを5羽放鳥し、その様子を5台のマイクロホンアレイで録音した。本実験では、キンカチョウの放鳥後1時間以内に開始した10分間の録音に注目し、2次元定位や分析の検討を行った。実験に用いたキンカチョウは、全て飼育された個体であり、実験以前はお互いに馴染みはなかった。そのため、録音中にはオスのキンカチョウ同士が行う、活発な初期段階の社会的な結合の構築過程の観測が期待できる。また、飼育下にいるキンカチョウは、止まり木や巣の近くで発声する傾向があり、提案した観測環境において発声状況がうまく観測出来れば、その発声は、構築環境の止まり木や巣を反映した結果になることが期待される。



図 5.2: 屋外のメッシュテント内に構築した実験環境。

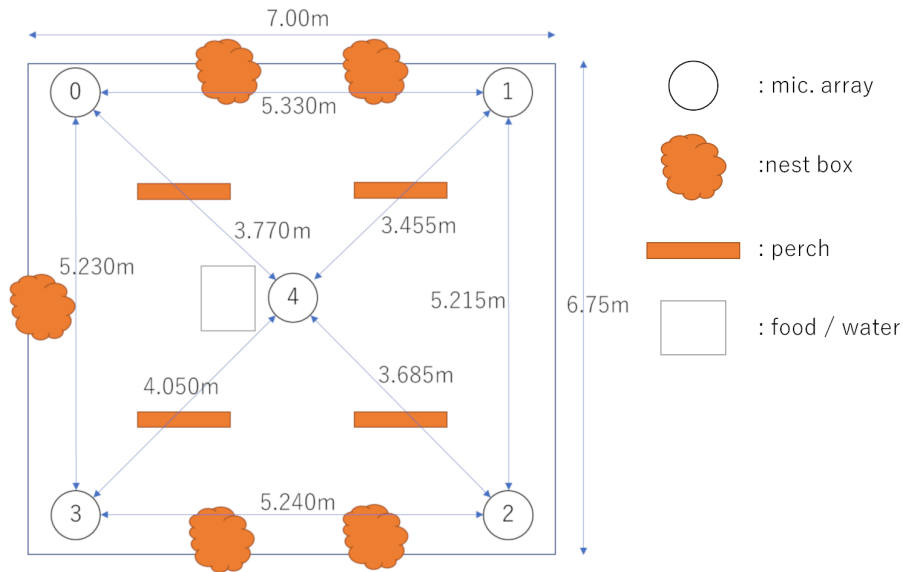


図 5.3: 実験環境の配置図.

### 複数マイクロホンアレイを用いた 2 次元定位

最初に、各マイクロホンアレイの録音それぞれを用いて HARK による音源定位・音源分離を行った。音源定位・分離には、2 章で紹介したアルゴリズムを用い、音源定位には MUSIC 法、音源分離には、GHDSS アルゴリズムを利用した。音源定位と分離を行う際に必要な伝達関数は、伝達関数の生成・可視化ツールとして利用される、HARKTOOL5 を用いて幾何的に計算したものを採用した。また、パラメータはキンカチョウの鳴き声をうまく定位できるように以下のように設定した: PERIOD = 5, THRESH = 28, UPPER\_BOUND\_FREQUENCY = 8,000, LOWER\_BOUND\_FREQUENCY = 3,000, NUM\_SOURCE = 2. このパラメータは、歌を構成する音素ごとに定位が行える設定であり、後述の 2 次元定位手法をより正確に行うための処理である。図 5.4 は、観測環境中央に配置したマイクロホンアレイの録音に対して、音源定位処理を行った結果である。上の図はスペクトログラムを示し、下の図の背景色は MUSIC スペクトルの強度を示し、黒い線が音源定位の結果を示す。連続的に定位されている音源はキンカチョウの鳴き声であり、適切にキンカチョウの鳴き声を定位できたことが確認できた。

次に、それぞれのマイクロホンアレイの音源定位結果に基づいて 2 次元定位を行う。図 5.5 には、2 次元定位の例を示す。2 次元定位の方法には、前章で紹介した方法（アルゴリズム 2）と同様の三角測量に基づいた方法を採用している。また、2 次

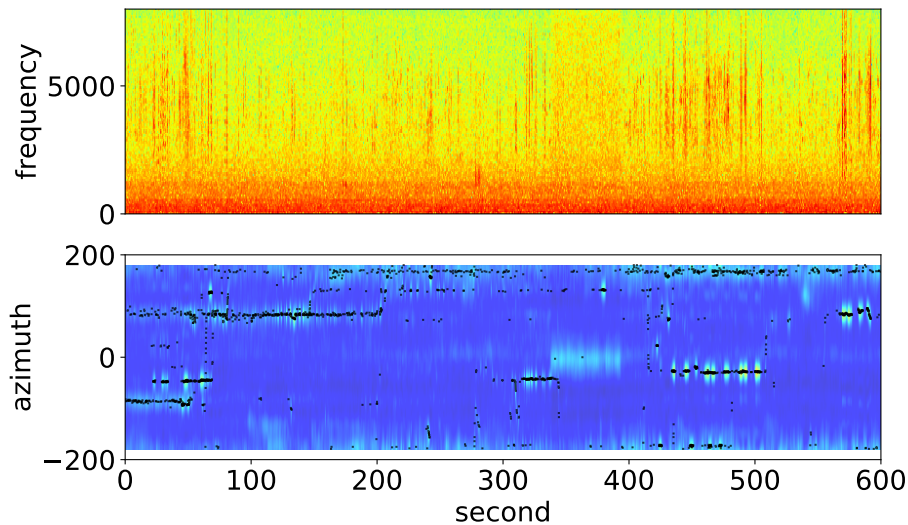


図 5.4: 中央のマイクロホンアレイの音源定位結果.

元定位に用いる音源のペアを採用する際に用いるアルゴリズムも、前章と同様の手法を採用している (図 5.5 右). ただし, ペアとして採用する距離は  $L < 10$  とした. 図 5.5 左の色分けは, テント内における, 2次元定位を行うマイクロホンアレイのペアごとの担当箇所を示す. 例えば, 桃色の範囲は, 中央のマイクロホンアレイと桃色で示されるマイクロホンアレイの音源定位結果を用いて 2次元定位を行う. この配置では, 前章で 2次元定位の課題として述べた, 2台のマイクロホンアレイを繋ぐ直線上での音源定位を行わずに済む. そのため, 安定した定位結果が得られることが期待される. 担当箇所に関して, テント外側の定位範囲はテントの中央から  $[-4, 4]$  で制限した. これにより, テント外の定位音を結果に含めないようにした. 最終的に各ペアで得られた, 担当箇所での音源定位結果を統合し, 全体の音源定位結果とした.

### 音響特性の抽出

さらに得られた 2次元定位結果に関して, 音響特性に関する分析を行った. 音響特性の抽出は次の 2つの方法で行った. なお, それぞれの方法は, 上記の 2次元定位手法によって得られた結果に基づいており, テント内で定位された音源はすべてキンカチョウの鳴き声であると見なして行っている.

1つ目は, t-SNE の低次元特徴量空間に基づく方法である. 本手法では, 中央の

マイクロホンアレイからの定位音源を上記の音源の類似度計算手法と同様の処理で t-SNE [66] による次元削減を行い、得られた特徴量空間における分布を、定位された音源の音響的特性とみなす。これを 2次元定位結果と組み合わせることによって、複数個体のキンカチョウの発声に関する時空間的ダイナミクスを統合的に得た。

2つ目は歌と地鳴きの区別によって行う。5.2 で述べたように、キンカチョウには歌と地鳴きの2種の鳴き声をうまく使い分けながら相互作用していることが報告されている。そのため、これらの区別を行うことによって詳細な相互作用を観測できると考えた。本手法も上記の手法と同様に、中央のマイクロホンアレイで得た分離音源を用いて行う。キンカチョウの歌と地鳴きには、発声時間と音素のレパートリー（＝複雑さ）に差があり、一般的に歌の発声時間は長く、様々な音素の組み合わせで構成された複雑な鳴き声であるのに対し、地鳴きの発声時間は短く、単一の音で構成された鳴き声である。このように大きく特徴に差があるため、今回はそれらの音響特性に基づいた以下の分類手法を考え、検討した。最初に次のパラメータを調整し直して音源定位を行う：PERIOD=10, THRESH =28.5。先程のパラメータは、歌を音素ごとに抽出するものであったが、今回は1つの歌全体を1つの音源としてつながった状態で定位できるパラメータである。なお、記述していないパラメータの変更はない。次に、抽出した音源に関して音量のピークの数（抑揚の数）と定位時間を算出し、定位時間が0.8秒以上の音源で、かつ音量のピークが1秒間に4回以上あった場合を歌とし、両条件を満たさない分離音源は地鳴きと分類した。この分類結果に関しても、2次元定位結果と組み合わせることで、キンカチョウの発声に関する時空間的ダイナミクスを得た。

### 5.3.2 結果

#### 2次元定位結果

図5.6は、実験中のキンカチョウの発声に関する2次元定位結果を示す。各点が定位された音源を示し、その色は図の右に示すカラーマップに対応した定位時刻を表す。 $[x,y] = [1 \text{ m}, 1.5 \text{ m}]$ の付近に注目すると、橙色と寒色系の音源が確認できるが、この位置には止まり木があり、これらの音源は止まり木で鳴いていたキンカチョウの鳴き声を定位したものだと考えられる。これは、定位位置が観測環境の配置を反映した結果であることを示している。また、同じ止まり木だとしても左と右で定位



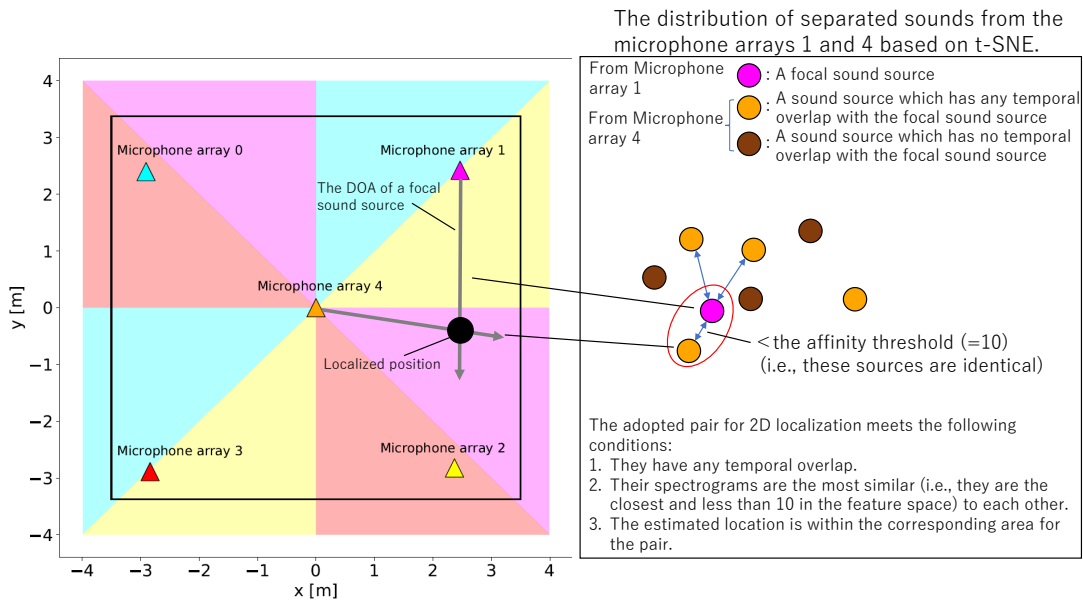


図 5.5: 音源定位に用いるマイクロホンアレイのペアごとの担当箇所を示す図. 各色で示された場所では, 中央のマイクロホンアレイと, その色に対応するマイクロホンアレイによる2次元定位が行われる.

場所が分かれており, 本手法によってキンカチョウの詳細な発声位置の観測ができたといえる.

また, 発声場所が時間的に変化している状況から, キンカチョウが移動しながら鳴いている, あるいは他個体が順番に鳴いているような傾向が示唆される. 図 5.6 の定位音源を繋ぐ線は, 定位された時系列順に音源を繋いだものであるが, 100 秒付近 (青色), 500 秒付近 (橙色) で複数の箇所を行き来している. この結果からは, 複数のキンカチョウの中には個体間で発声によって相互作用を行っていた個体があったことが示唆される. 以上から, 総じて, 本手法によって複数のキンカチョウの発声に関する時空間的ダイナミクスの抽出に成功したといえる.

一方で, 図 5.6 左下に見られるように, テント外で定位されている音源も存在した. これは, マイクロホンアレイを配置した際, あるいは配置した位置の計測の際に生じた位置や方向のわずかな誤差に起因するものと考えられる.

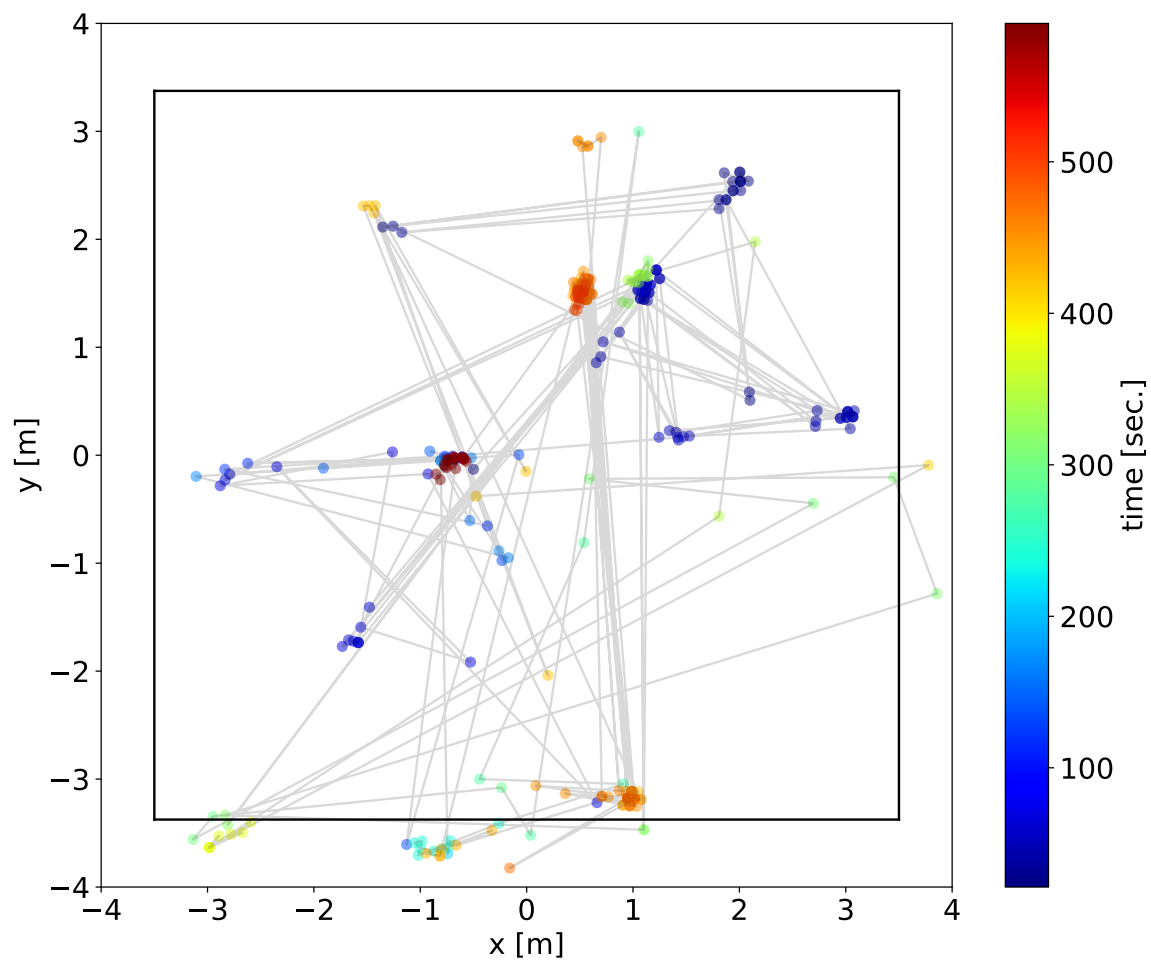


図 5.6: 観測環境内における定位音源の時空間的分布. 点は定位された音源を示し, 色は定位された時間を示す. 灰色の線は定位順に音源を繋いだもの.

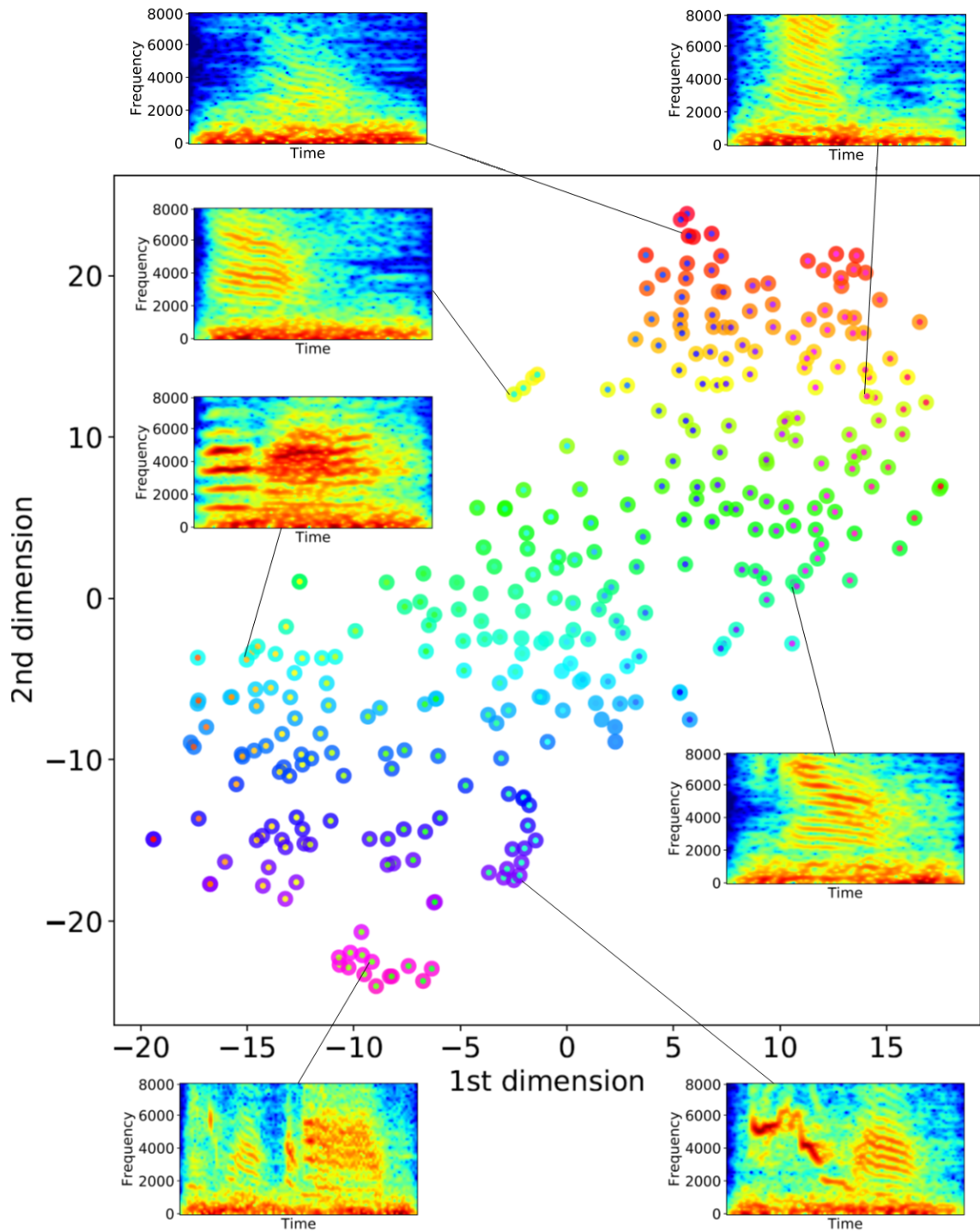


図 5.7: 中央のマイクロホンアレイの定位音源に関する t-SNE を用いて得た特徴量空間上での分布.

## 音響的特性の抽出

図 5.7 は、音源定位結果として得られた音源に対応する中央のマイクロホンアレイの分離音源を、スペクトログラム画像に変換して t-SNE によって 2 次元に次元削減することで得た音響特性の低次元特徴量空間における分布を示す。音源の色に関して、内部の色は第一次元の値を HSV 色空間の色相に対応させて示したもので、枠線の色は第二次元の値を matplotlib のカラーマップの `gist_rainbow` に対応させて示したものである。図からは、いくつかの小さなクラスタを形成していることがわかる。実際に音源を見て、聞いて確認してみると、これらのクラスタはスペクトログラム画像が類似したもので、つまり音響特性が似た音源同士で構成されていた。これは、特徴量空間上の分布が音源の音響特性を反映したものであることを意味する。また、これらの音源にはつばみ音などが含まれていたが、ほとんどがキンカチョウ個体の発声した鳴き声であった。以上から、この結果は、キンカチョウ個体の発声に関する音響特性や文脈、あるいはそれらの関係についての理解に貢献することが期待できる。

## 鳴き声に関する時空間的ダイナミクスの抽出

次に、2次元定位によって抽出したキンカチョウ集団における鳴き声の時空間的ダイナミクス（図 5.6）と、先ほどの音響特性（図 5.7）を組み合わせ、総合的な多次元に跨るダイナミクスの抽出を行った。図 5.8 はそれらの統合結果であり、横軸に時間（時間的ダイナミクス）、縦軸に中央のマイクロホンアレイからみた定位位置の距離（上図）と方向（下図）（空間的ダイナミクス）、音源の色に、図 5.6 の音響特性を反映した結果を示している。下のグラフの定位方向に関する結果は、図 5.4 の下のグラフに示されたテント中央に配置したマイクロホンアレイによる音源定位結果と比較しても、音源の定位時間と分布がほとんど一致している。これは、2次元定位がキンカチョウの集団の発声を漏らすことなく定位できたことを示している。

また、結果からはキンカチョウの集団レベルの行動の変化を観測できていることが示唆される。例えば、0~100 秒あるいは 300~500 秒の時間帯では、複数の方向・距離で音源が同時に定位されていることが確認でき、これは個体同士で鳴き合っている状況であると推測される。一方で、100~200 秒、500~600 秒の時間帯では 1 つの方向・距離のみで音源定位がされており、これは同じ場所で単一の個体、あるいは

複数の個体が鳴いていた状況であると考えられる。さらに、それぞれの時間における定位音源の音響特性に着目してみると、例えば0~100秒では緑色系が多く、400~500秒では青色系が多いように、時間帯ごとに類似した音響特性を持つ音源が定位されている。このことから、キンカチョウ集団が同時に発声を行う際には、少なくとも部分的に共有する発声を用いて鳴き合う傾向があることが示唆された。

詳細な個体間の相互作用を確認するために、より短い時間に焦点を当てて調査した。図5.9,5.10は、図5.8において複数音源が同時刻に定位されていた時間帯である40~70秒、440~480秒の結果をそれぞれ抽出したものである。同時に定位されている音源は、非常に短い時間で様々な方向・距離で定位されていることから、それぞれの音源は単一の個体が素早く移動しながら鳴いていた状況ではなく、異なる個体が異なる場所で発声した結果であるといえる。どちらの時間帯に関しても、ある方向で音源定位が開始してからすぐに、別の方向での定位が始まっている。これは、ある個体の最初の発声が、他の発声のきっかけとなる可能性があることを示している。また、先程述べたように、それぞれの時間帯において、音源の音響特性は比較的類似している。これは、キンカチョウ個体が特定の発声を用いて相互作用していることを示しており、それぞれの時間帯で生じた社会的状況を反映している可能性がある。

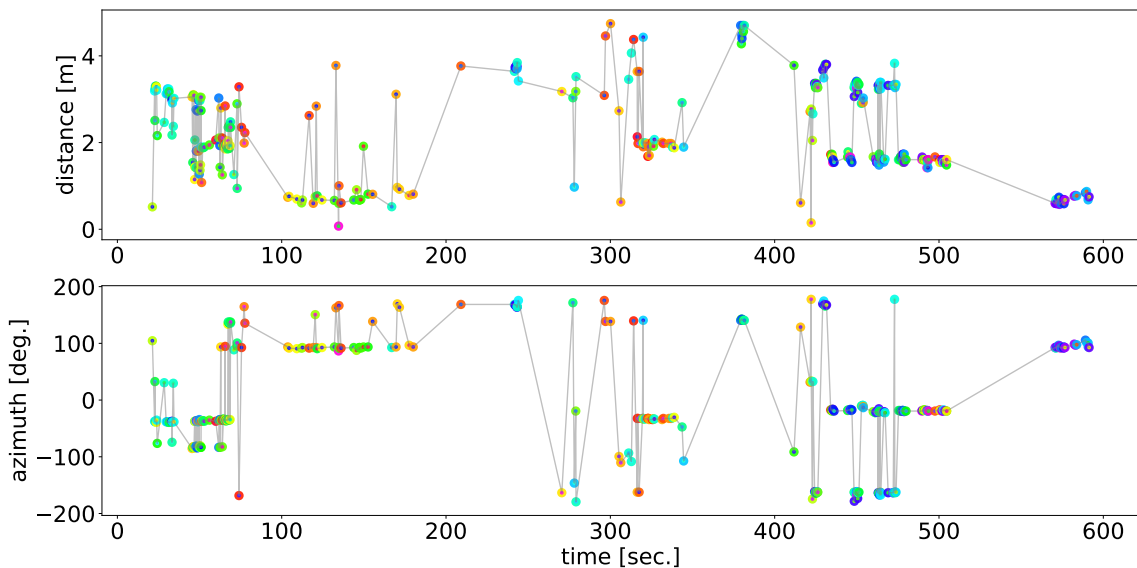


図 5.8: キンカチョウ集団の鳴き声に関する時空間的ダイナミクス。定位された音源が中央のマイクロホンアレイからどの程度離れて（上）、どの方向（下）で定位されたかの時間的推移を示す。各音源の色は図5.7に対応する。

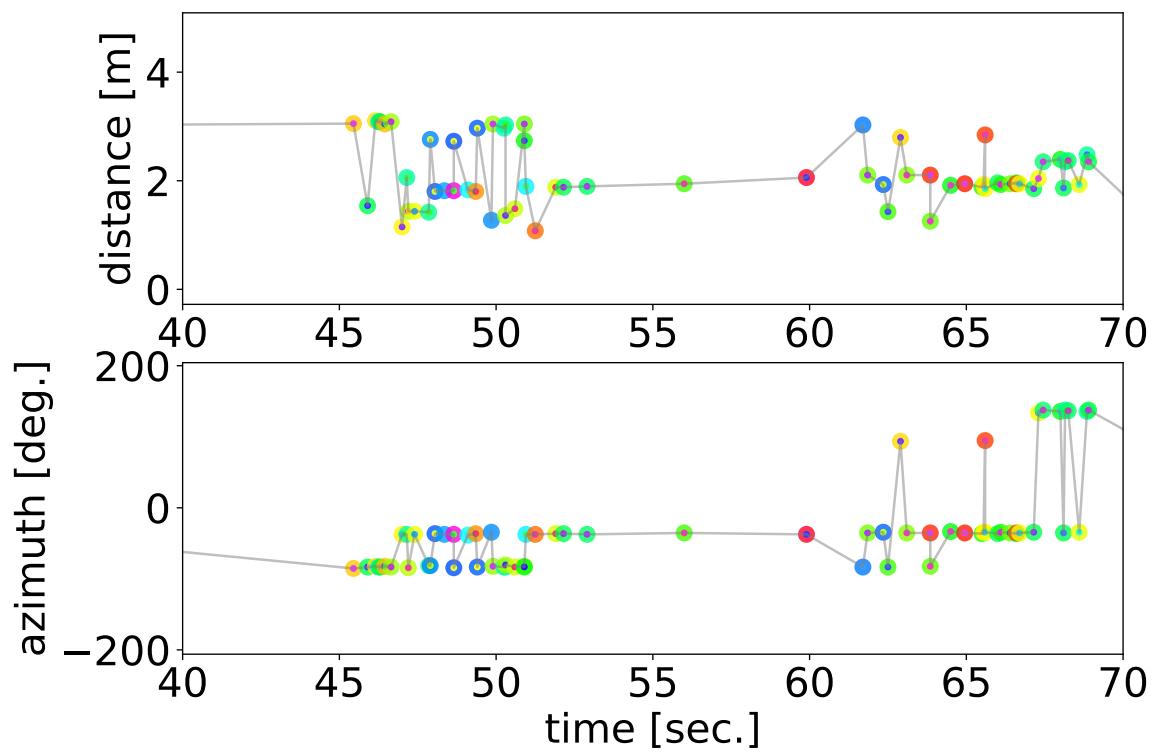


図 5.9: 図 5.8 の 40~70 秒に着目した結果.

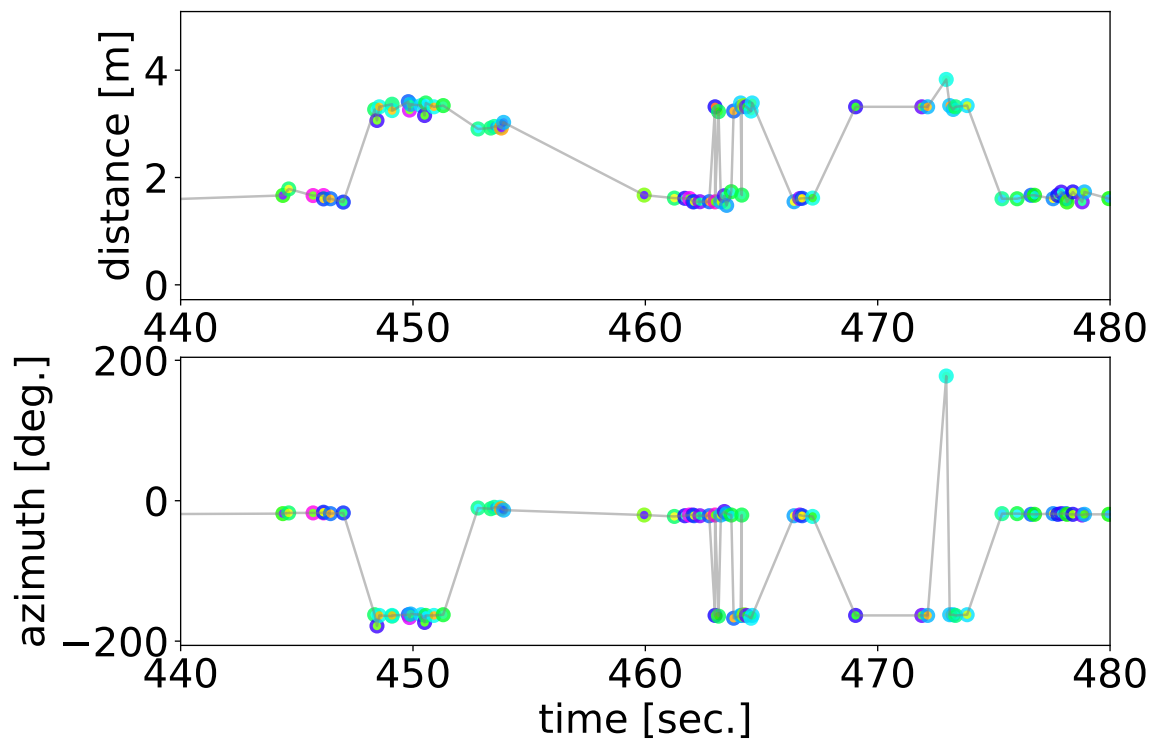


図 5.10: 図 5.8 の 440~480 秒に着目した結果.

## 歌・地鳴きの分類に基づく発声ダイナミクスの分析

図 5.11–5.13 は、提案手法によって分類された、歌と地鳴きの中央のマイクロホンアレイからの定位方向と距離の時間的推移を示す。これらの図は、それぞれ図 5.8–5.10 の時間軸に対応した結果である。10 分間の録音全体では、全部で 99 個の音源が定位され、そのうち 44 個が歌、45 個が地鳴きに分類された。この分類結果に関して手作業で確認を行ったところ、歌が歌として正しく分類されたのは 91% (40/44 個) であった。また、その中で歌ではなかったものは、短時間の中に連続して繰り返された地鳴きであることを確認した。一方、地鳴きが地鳴きとして正しく分類されたのは 73% (33/45 個) であった。地鳴きとして分類された分離音の中には、歌は含まれていなかったものの、歌の直前に発せられる導入部分の発声 (introductory notes) や、餌のついでみ音が含まれていることを確認した。以上から、本手法によって歌と地鳴きの区別の大まかな分類は可能であるといえる。

観測された結果では、キンカチョウ集団が、時間帯によって歌と地鳴きを使い分けている様子がわかる。例えば、図 5.11 の 100~300 秒付近では、地鳴きが多く使われている一方で、400 秒以降は歌が主に使われている。また、先ほど注目した 40~70 秒、440~480 秒の複数個体が鳴き合っている状況では、どちらの時間帯においても歌を用いており、その歌は複数の距離・方向で交互に定位している傾向がある。これより、注目した時間帯はキンカチョウが鳴き交わしをしていた状況であると推測される。

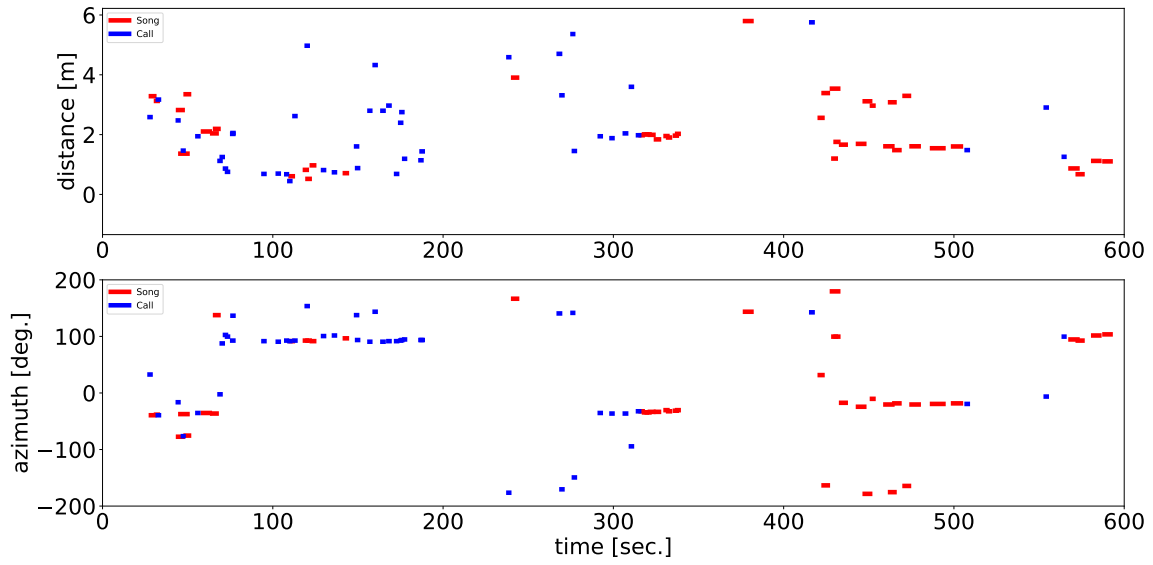


図 5.11: 歌と地鳴きの分類結果に基づく鳴き声の時空間的ダイナミクス (0~600秒).

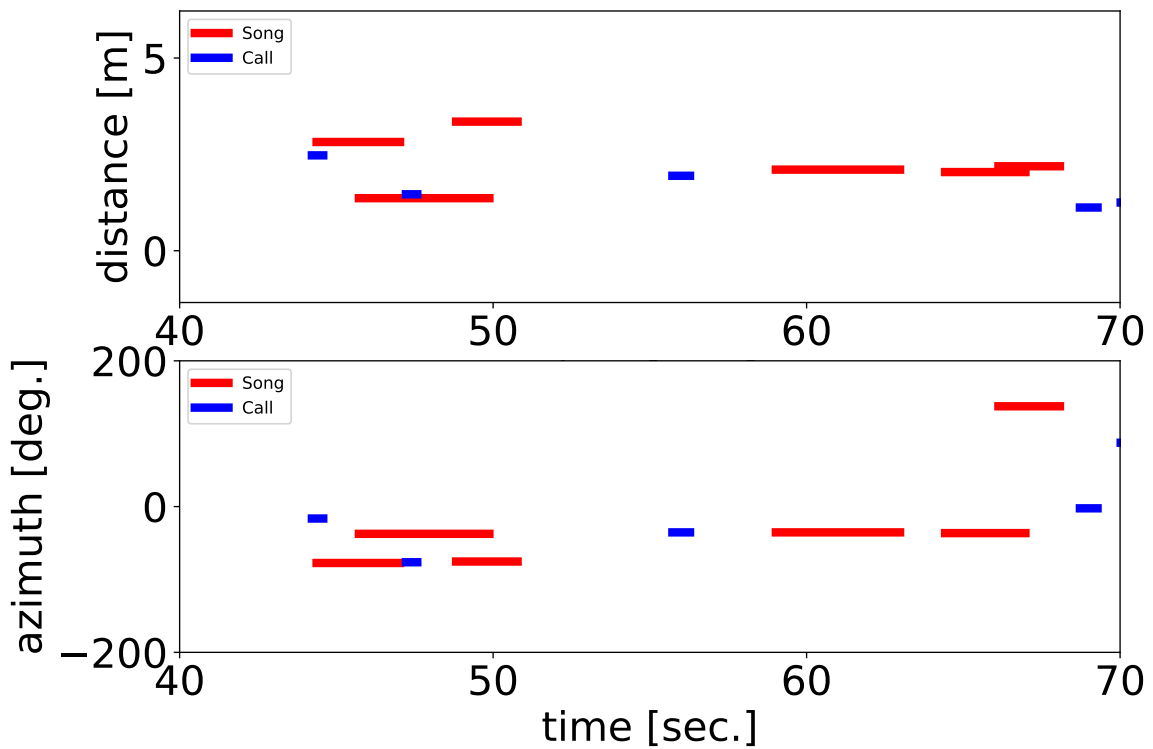


図 5.12: 歌と地鳴きの分類結果に基づく鳴き声の時空間的ダイナミクス (40~70秒).



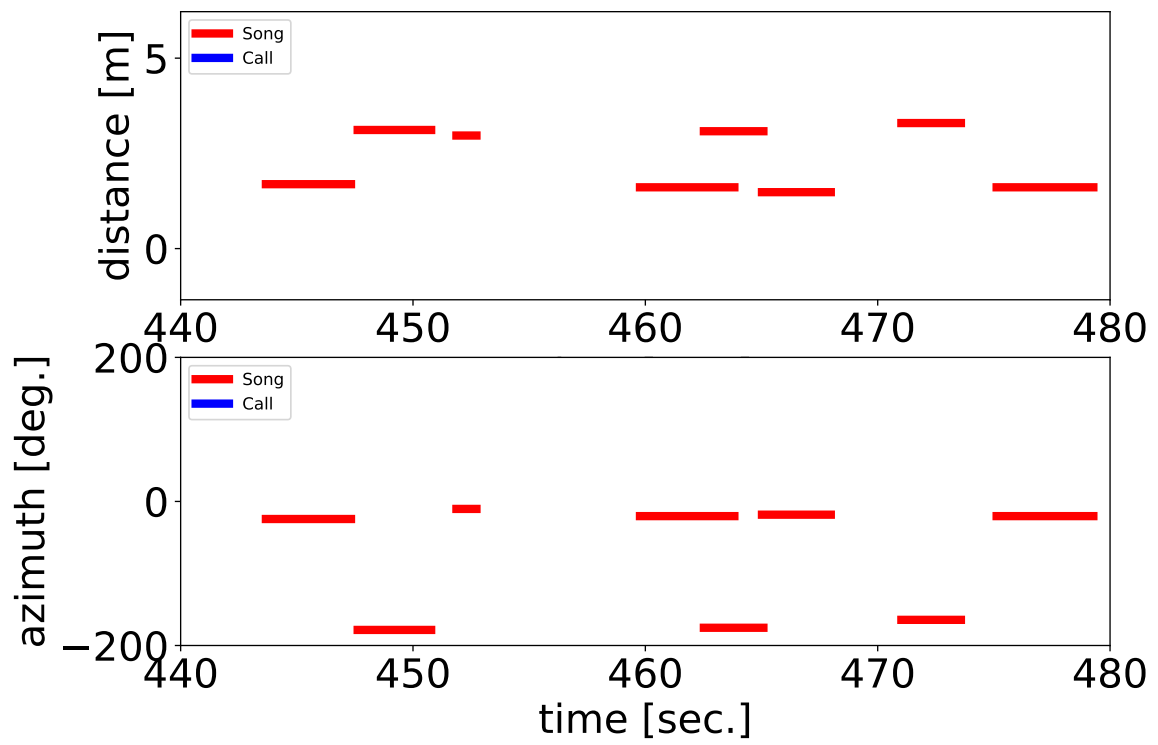


図 5.13: 歌と地鳴きの分類結果に基づく鳴き声の時空間的ダイナミクス (440~480 秒)。

## 5.4 個体間の社会的関係に焦点を合わせた録音実験

### 5.4.1 実験手法

#### 環境構築と録音実験

録音実験は、2020年の8月から9月前半にかけて、2019年に実施した実験と同様の場所である、北海道大学札幌キャンパス内にある圃場内の屋外テント内で実施した。観測環境には、止まり木と巣、餌場と録音用のマイクロホンアレイを設置したが、前回の録音実験から配置を変更し、図5.15のような配置とした。図中に示される2つの数字の組み合わせは、数字が記述されたエリアの音源定位を担当するマイクロホンアレイのペアを示す。この配置および範囲は、音源定位を行うマイクロホンアレイのペア同士で音源からの距離がある程度同じになるようにした。これは、抽出された分離音源の質（背景ノイズの強さ）が音源からのマイクロホンアレイまでの距離に関係するためである。また、2次元定位のペアの決定は音源のスペクトログラム画像の類似度に基づいて行われるが、この類似度は背景ノイズの強さの影響を受けるため、ペアとなる音源が同程度の背景ノイズを含めばその影響が小さくなると考え、この配置とした。



図 5.14: テント内環境.

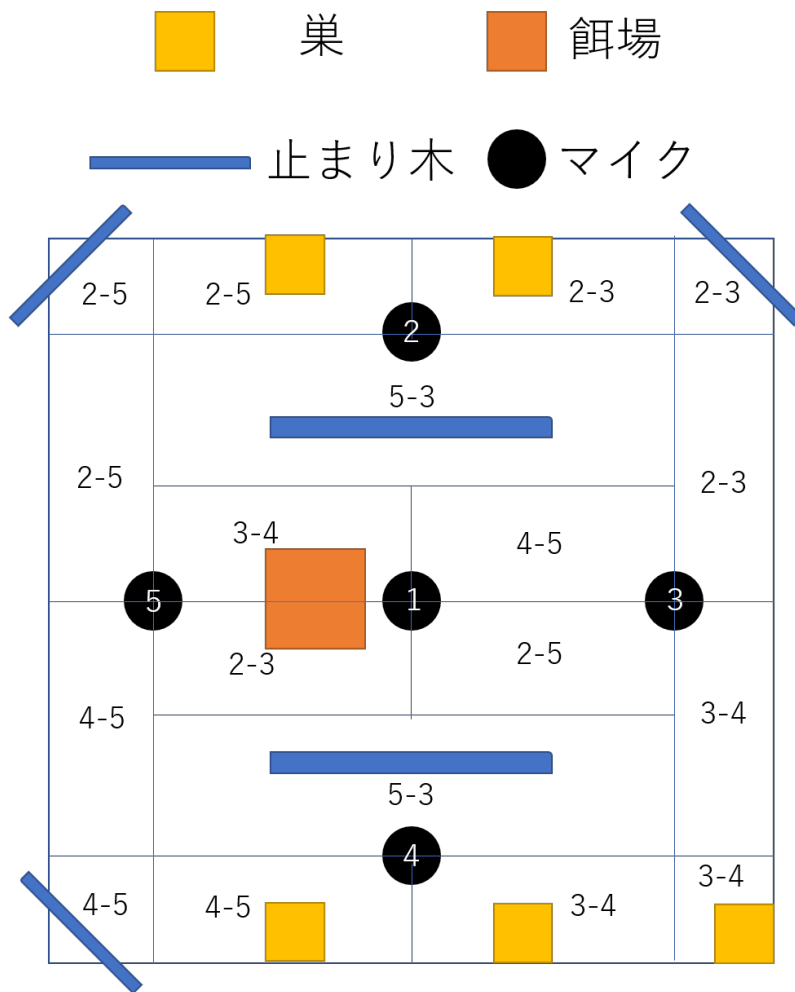


図 5.15: 配置図.

録音実験は、1 個体のみ放鳥した条件、個体が放鳥されている環境に新たな個体を放鳥する条件の下で行った。各実験では、前日の 18 時以降に翌日の実験個体となるキンカチョウを実験環境に放鳥し、その翌日早朝 5 時から 18 時まで録音を行った。複数個体の実験では、9 時から 9 時 30 分間に追加個体を放鳥し、13 時 30 分から 14 時の間に追加個体を再び捕獲した。録音実験は、異なるオスを 1 羽ずつ放った 3 条件（実験 1-実験 3）とオス 1 羽のいる環境に他のオス 1 羽を追加した 3 条件（実験 4-実験 6）、オス 2 羽のいる環境にメス 1 羽を追加した 2 条件（実験 7-実験 8）で行った。実験に用いた個体は、実験室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された。各条件に関して、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。

## 2 次元定位

2 次元定位は、前回の実験と同様に前章で紹介した方法（アルゴリズム 2）に基づいて行うが、処理速度の高速化、より安定した類似度の取得を目指して処理の一部を変更した。具体的な処理方法は以下のとおりである。なお、大きな変更点は下線部で示している。

1. 類似度の計算よりも先に 2 次元定位を行い、すべての組み合わせに関する音源定位結果を得る。
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. 切り出された分離音源を  $128 \times 128$  のグレースケール画像に変換し、UMAP (Uniform Manifold Approximation and Projection) [90] を用いて画像データの次元を 2 次元に削減、UMAP の特徴空間上で切り出した音源のペア間の距離が 3 以下である場合その 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ロホン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、それらを統合する。これを最終的な 2 次元定位結果とする。

本実験では、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、個体の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ロホン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。

## 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) [72] と SVM (Support Vector Machine) [73] を用いる。判別の流れを図 5.16 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 [87]。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 [24]、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  $128 \times 128$  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 層の全結合層からなり、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大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．

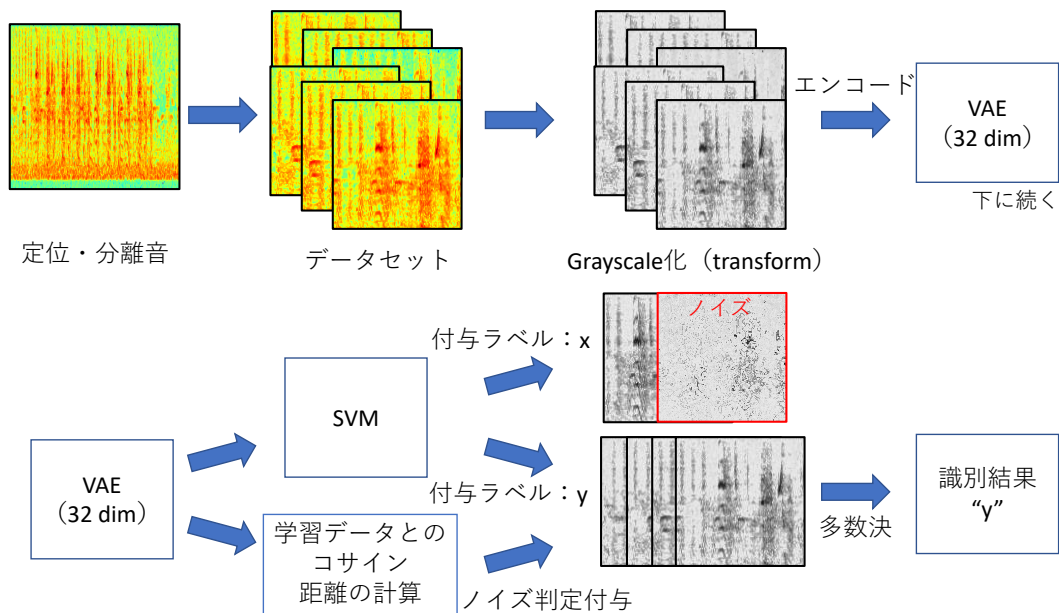
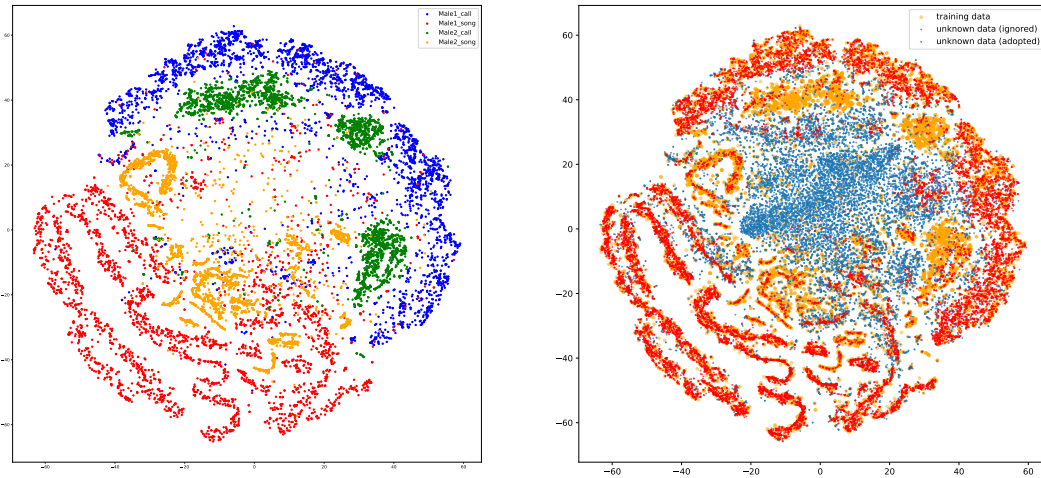


図 5.16: 鳴き声判別の流れ.

ノイズ判定の適用結果の例として，図 5.17 に実験 1 に用いた 2 羽のオス個体の鳴き声を VAE に適用後，t-SNE [66] 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 5.17 (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 5.17 (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．

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



(a) 学習データ

(b) ノイズ除去結果.

図 5.17: 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.

データを SVM に適用した結果, 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.

## 5.5 分析結果

### 5.5.1 1 羽放鳥実験

図 5.18-5.20 はオス 1 羽を放った条件について, 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 (以降, それぞれの実験で扱った個体を Male1, Male2, Male3 とする). 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. オス 1 羽の条件では, それぞれの条件について以下の共通する傾向が確認された. 1 つ目は, 歌と地鳴きの分布がある程度一致している傾向があるという点である. 頻度自体には差はあるものの, 色の出ている (鳴いている) 地点は, 歌と地鳴きで大きな差はないことが確認できる. また, 餌場

(図 5.18-5.20 の地点 1) での発声については、地鳴きよりも歌の発声割合が高い。2 つ目は、個体はある程度決まった場所で鳴くという点である。定位位置について着目すると、実験個体は様々な場所で鳴いているが、発声頻度の高い場所は餌場や止まり木 (図 5.18-5.20 の地点 2), 巣 (図 5.18, 5.20 の地点) などの 2~3 か所に絞られている。以上のことから、オス個体のキンカチョウは 1 羽で放鳥された (近隣に他個体がない) 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。

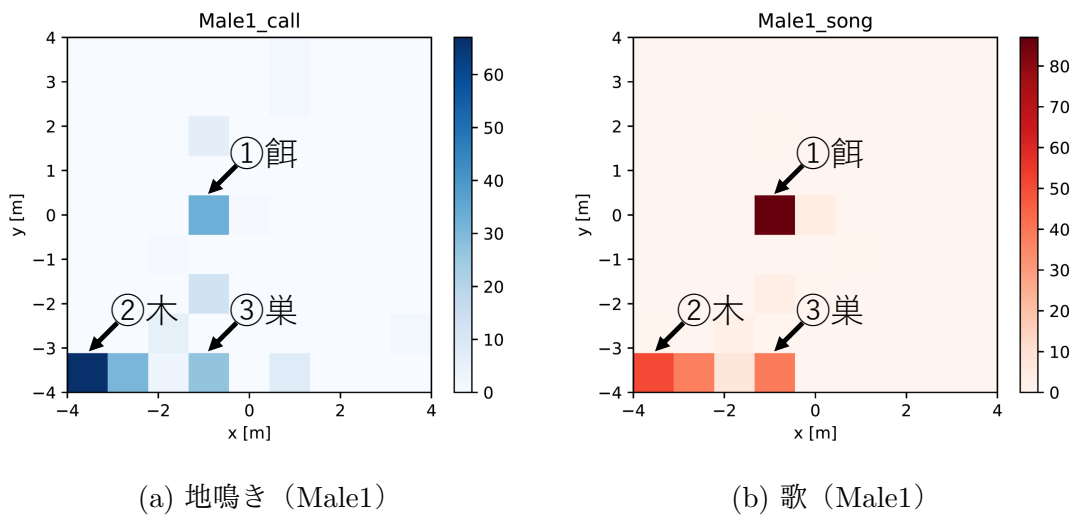


図 5.18: 鳴き声の頻度分布 (実験 1)。



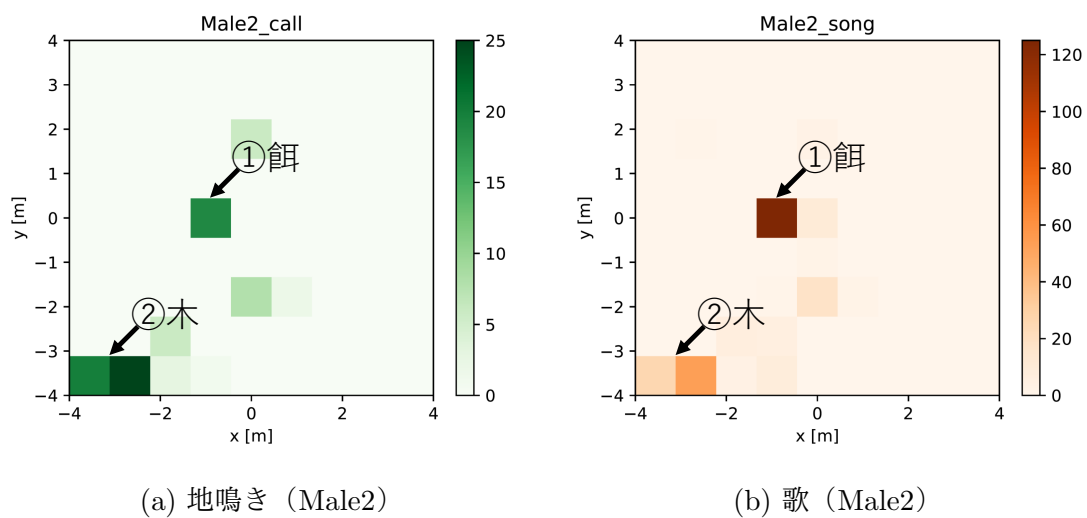


図 5.19: 鳴き声の頻度分布 (実験 2) .

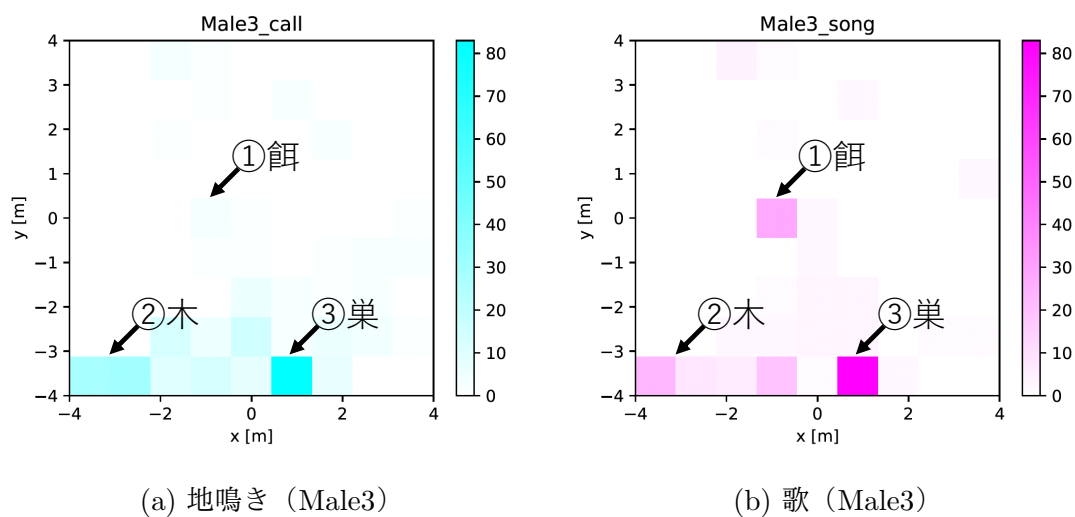
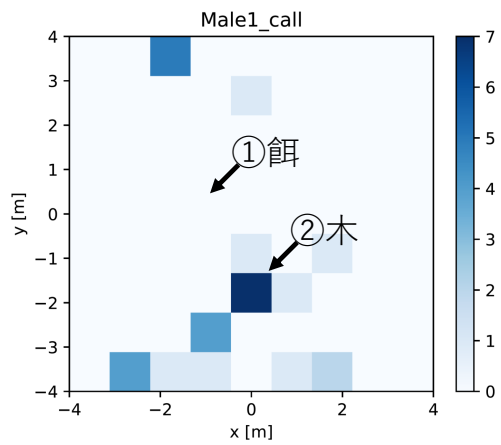


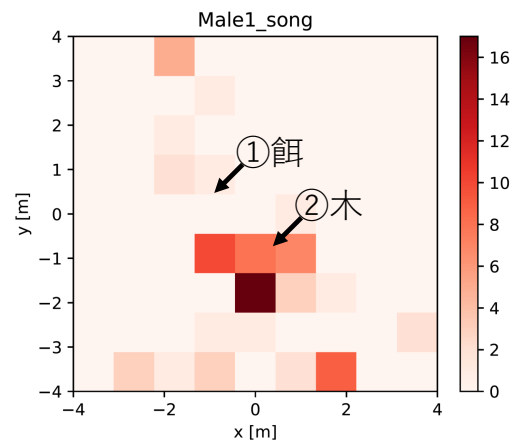
図 5.20: 鳴き声の頻度分布 (実験 3) .

### 5.5.2 2羽放鳥実験

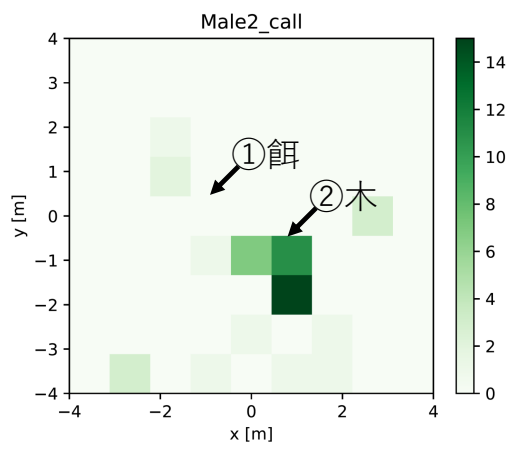
図 5.21-5.23 は、それぞれ Male1 と Male2, Male1 と Male3, Male2 と Male3 の 2羽のオス個体ペアをテント内に放鳥したときの 2次元定位結果から得た鳴き声の頻度分布を示す。オス個体を 2羽入れた状況においても、各条件で共通する傾向が確認された。まず、それぞれの個体に関して歌・地鳴きの分布が比較的一致している傾向である。これは、1羽条件においても確認された傾向であるが、1羽条件ほど顕著ではない。また、1羽条件の傾向に反して、餌場（図 5.21-5.23 の地点 1）での発声は減少した。個体間で分布を比較すると、単一の個体のみが発声した場所も存在するが、各図中の地点 2 や地点 3 で示される止まり木や巣の場所において発声位置が重なっている場所が多くみられる。これは、2羽の個体が場所を共有して鳴き合っていた状況であるといえる。



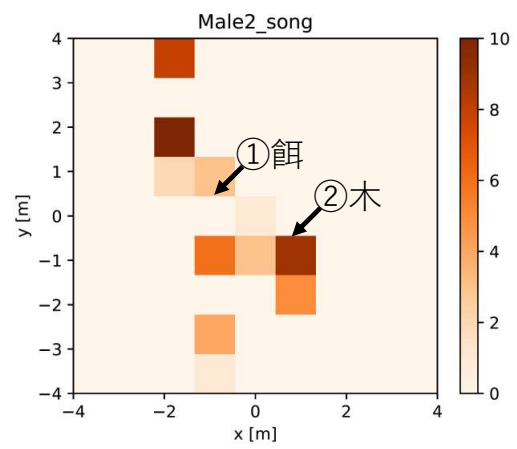
(a) 地鳴き (Male1)



(b) 歌 (Male1)

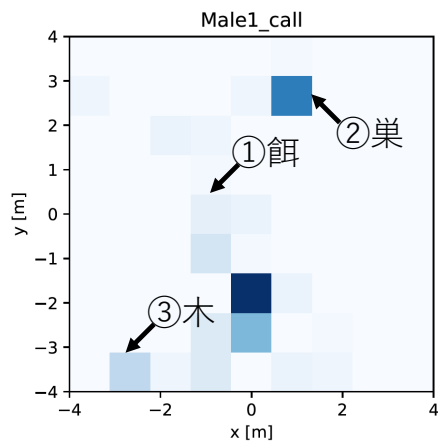


(c) 地鳴き (Male2)

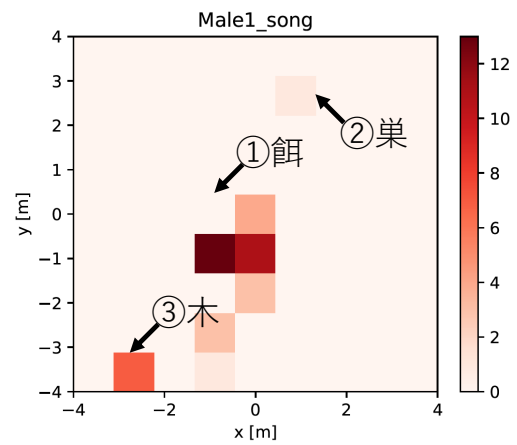


(d) 歌 (Male2)

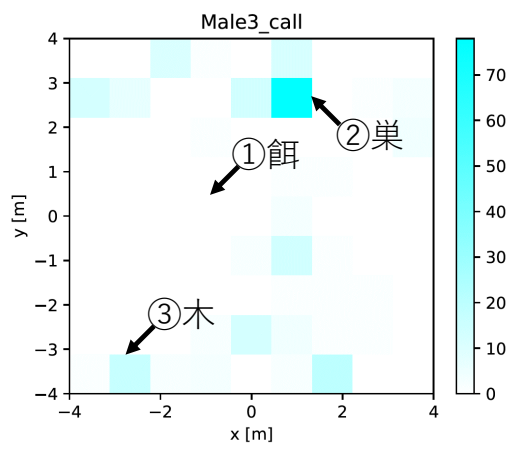
図 5.21: 鳴き声の頻度分布 (実験4) .



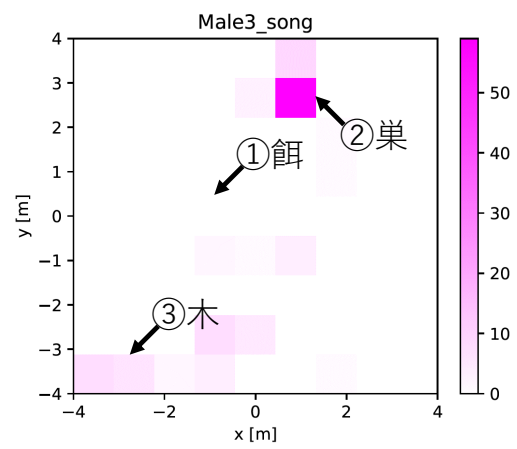
(a) 地鳴き (Male1)



(b) 歌 (Male1)

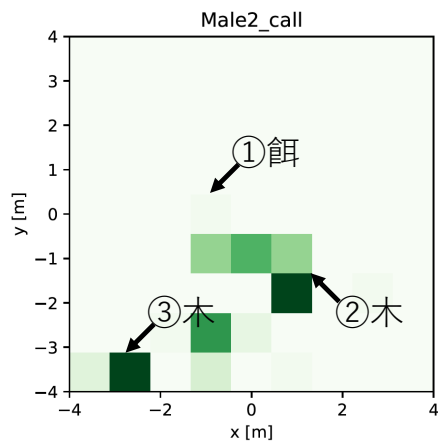


(c) 地鳴き (Male3)

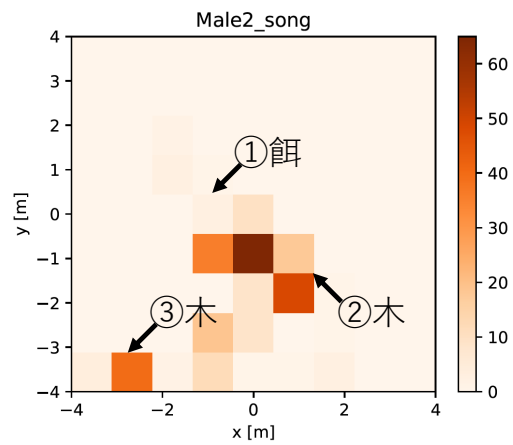


(d) 歌 (Male3)

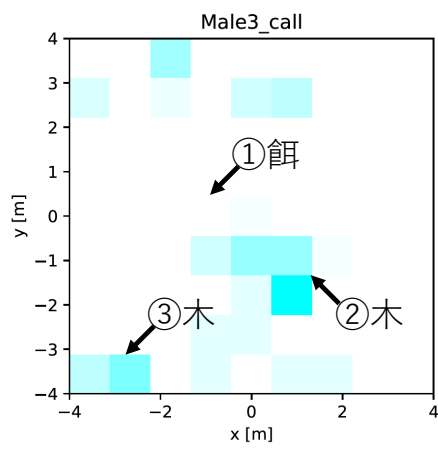
図 5.22: 鳴き声の頻度分布 (実験5).



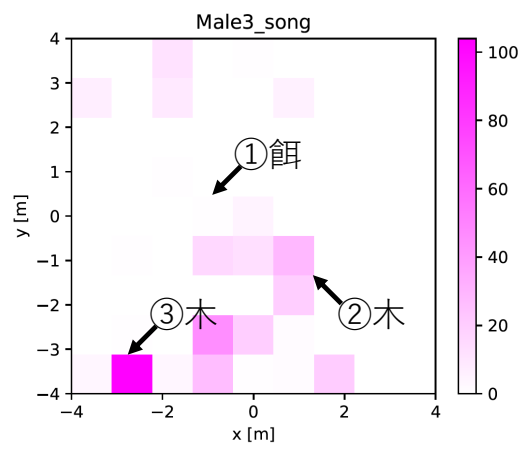
(a) 地鳴き (Male2)



(b) 歌 (Male2)



(c) 地鳴き (Male3)



(d) 歌 (Male3)

図 5.23: 鳴き声の頻度分布 (実験6).

### 5.5.3 3羽放鳥実験

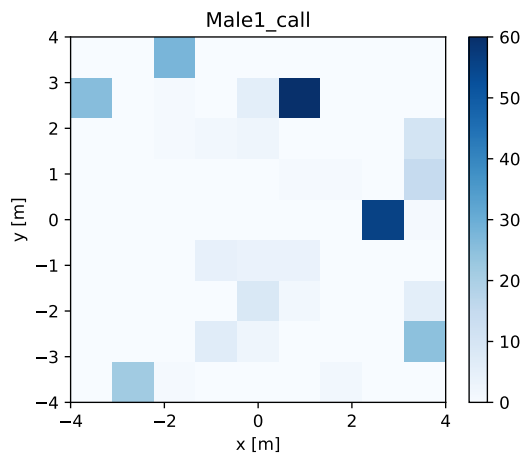
図 5.24, 5.25 は, オス 2 羽, メス 1 羽をテント内に放鳥した際の 2 次元定位結果から得られた鳴き声の頻度分布を示す. 実験に用いた個体は, 実験 7 が Male1, Male2, Female1, 実験 8 が Male1, Male3, Female2 である. それぞれの条件に関して, オス 2 羽を放鳥した条件と比較すると, 2 羽の実験では, オス個体同士の分布は重なる場所があった一方で, 3 羽の実験では, 重なる場所が減少した. 特に, 濃い色で示された頻繁に発声する場所は各オス個体で異なる地点に存在している. 例えば, 実験 7 では Male1 の鳴き声は図の上半分が多く, Male2 は下半分が多い. 実験 8 では, Male1 の地鳴きは図の右下の巣, 歌は図右下の巣と図中央の下にある止まり木に集中している一方で, Male3 の歌と地鳴きは図の左上の巣と左下の巣で多くみられる. これらの結果から, メス個体が居ないオス個体 2 羽の条件とは傾向が大きく変化していることが確認され, 状況に応じてオス個体が挙動を変化させていることが示唆される.

### 5.5.4 発声頻度の比較

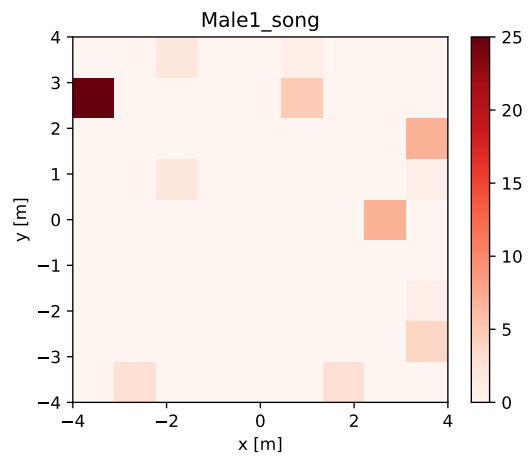
次に, 条件の違いによるオスの鳴き声の発声頻度の変化を調査した. 表 5.1-5.3 は, 各条件でのオス 2 羽の歌・地鳴きの定位回数をまとめたものである. この結果から, オス個体のキンカチョウが歌・地鳴きともに条件に応じて異なる発声頻度で鳴く傾向があることが示唆された.

まず地鳴きについて, 1 羽条件では, 個体によって差はあるが概して多く鳴いていることが確認できる. 2 羽条件では, 個体の組み合わせによって定位回数に差が見られるが, 一方が少ないともう一方は少なく (Male1-Male2), 逆に一方が多い場合はもう一方も多くなる (Male1-Male3) 傾向がある. 3 羽条件においては, オス 2 羽間の地鳴きの定位回数に大きな差が生じている.

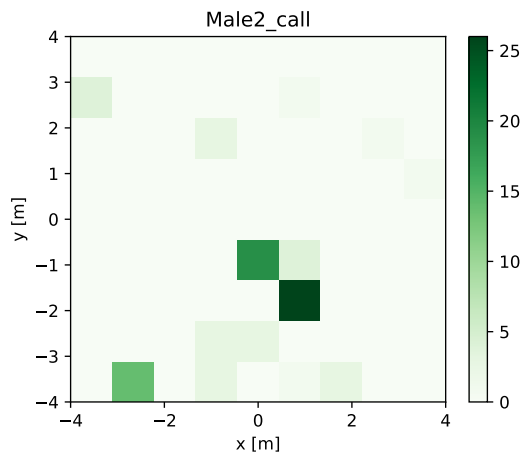
歌に関しては, 1 羽条件では地鳴きと同様に比較的多い歌が定位されている. 2 羽の条件では, ペアの組み合わせによって違いがみられ, Male1 と Male2, Male1 と Male3 の 2 条件では双方の鳴き声は比較的小さいが, Male2-Male3 は他の条件と比較してより多くの歌が定位されている. また, 3 羽条件では, 地鳴きの傾向と同様に一方が多く, もう一方が少ないという傾向があった.



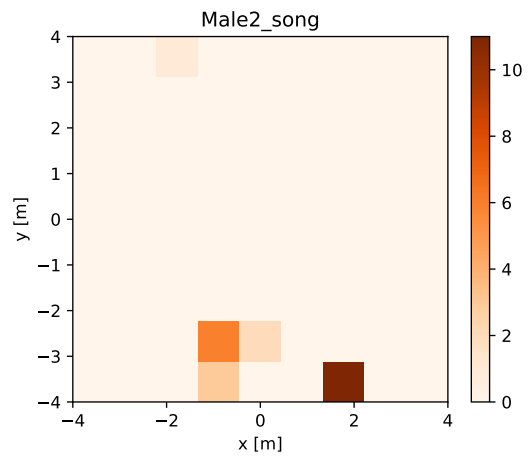
(a) 地鳴き (Male1)



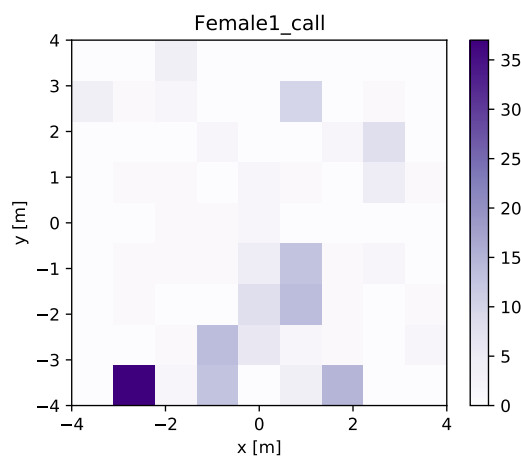
(b) 歌 (Male1)



(c) 地鳴き (Male2)

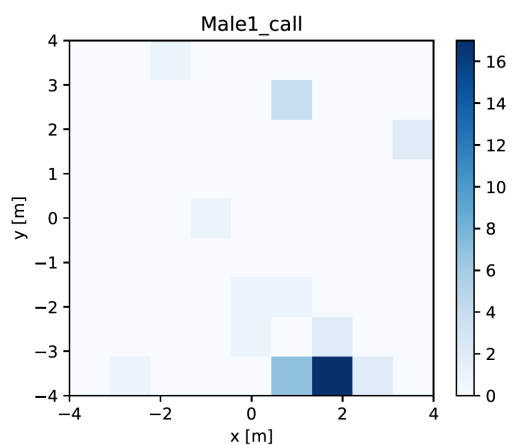


(d) 歌 (Male2)

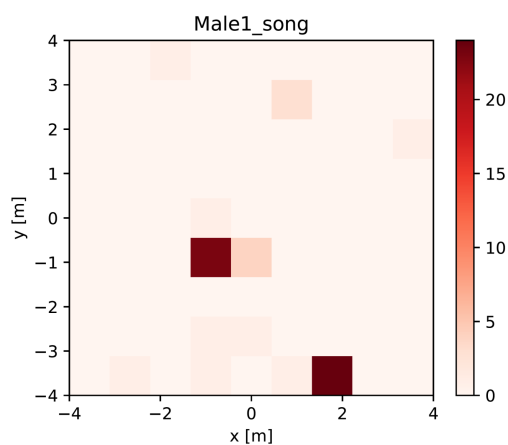


(e) 地鳴き (Female1)

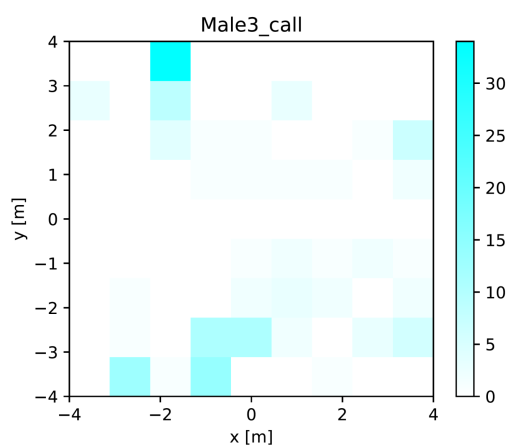
図 5.24: 鳴き声の頻度分布 (実験 7).



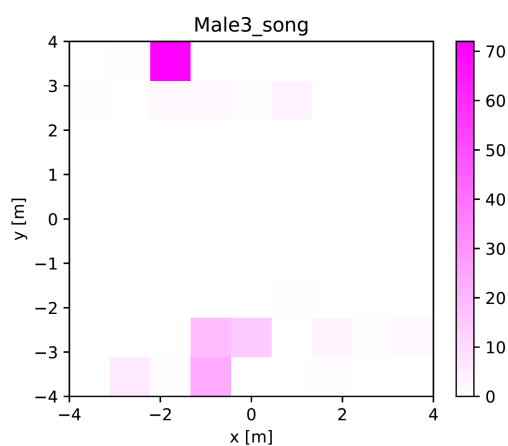
(a) 地鳴き (Male1)



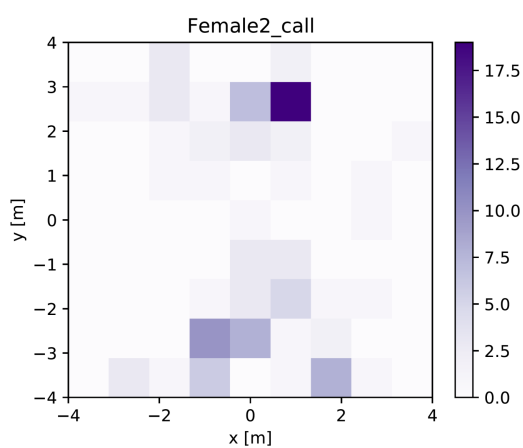
(b) 歌 (Male1)



(c) 地鳴き (Male3)



(d) 歌 (Male3)



(e) 地鳴き (Female2)

図 5.25: 鳴き声の頻度分布 (実験 8).



表 5.1: 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

表 5.2: 各条件の Male1, Male3 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	157	40
	歌	240	43	62
Male3	地鳴き	258	236	149
	歌	203	114	156

表 5.3: 各条件の Male2, Male3 の歌・地鳴きの定位回数.

		1羽条件	2羽条件	3羽条件
Male2	地鳴き	90	127	n/a
	歌	269	289	n/a
Male3	地鳴き	258	123	n/a
	歌	203	353	n/a

### 5.5.5 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ため、より詳細な分析を行った。図 5.26-5.28 は、2羽以上のそれぞれ条件・個体の組み合わせについて、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。結果からは、2羽条件では、個体は3羽条件と比較してより近くで鳴く傾向が示された。例えば、図 5.26 (a) の Male1, Male2 の2羽条件に関して、この実験と対応する3羽条件の実験は図 5.27 の (a) であるが、観測値は、2羽条件で 2.3 m 付近、3羽条件で 4.3 m 付近と大きく異なる。また、Male1, Male3 の組み合わせに関しても、2羽条件では 3.5 m 付近である一方、3羽条件では、4.2 m 付近にあり2羽条件の方が近い。また、全ての2羽放鳥条件では、観測値は95%の信頼区間を下回っている。以上のことから、オス個体同士は2羽条件においては、3羽条件と比較して友好的関係を持っていることが示唆され、さらに一方の個体の鳴き声他方を呼び寄せる役割があることが推測される。

また、3羽の条件においては、発声時の個体間距離は個体の組み合わせによって違いがあり、実験 7 (図 5.27) では、Female1 と Male2 の発声時の個体間距離は近い一方、Male1 と他の個体の組み合わせでは個体間の距離は遠い。実験 8 (図 5.28) では、観測値に関していえば、Female2 と 2羽のオスとの個体間の距離は比較的近いが、オス同士の個体間距離は Female2 との距離よりも離れている。メス個体の導入によってオス個体間の距離が離れるということから、この結果はキンカチョウ同士の関係性を反映したものであると推測される。

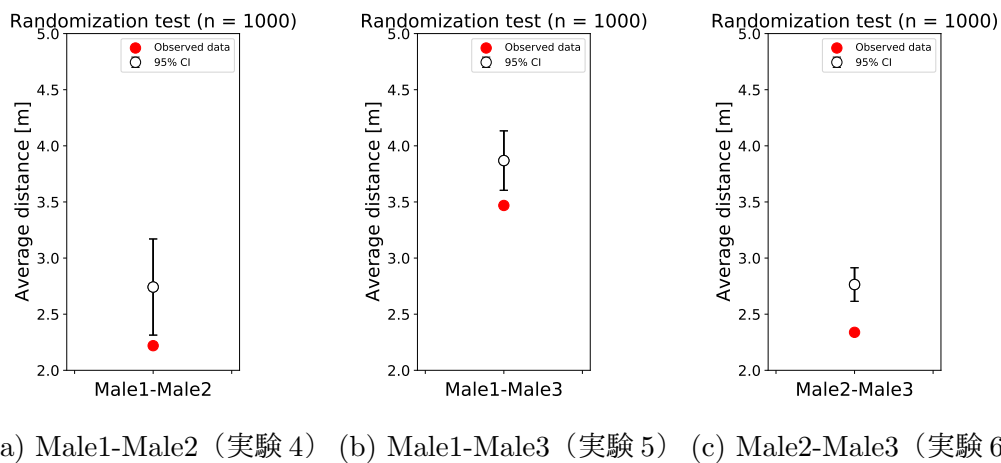


図 5.26: オス 2羽放鳥実験における発声時の個体間の平均距離とランダムイゼーションテスト ( $n = 1000$ ) による 95%信頼区間。

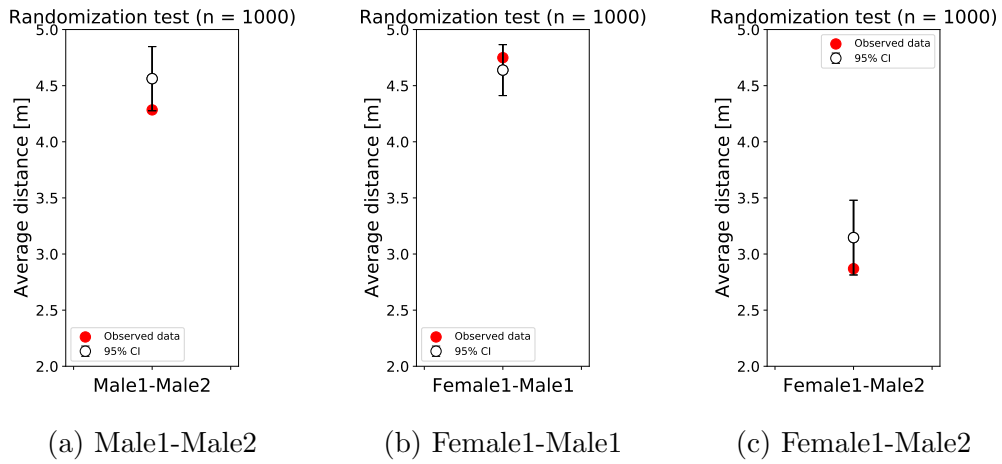


図 5.27: 実験 7 における発声時の個体間の平均距離とランダム化テスト ( $n = 1000$ ) による 95%信頼区間.

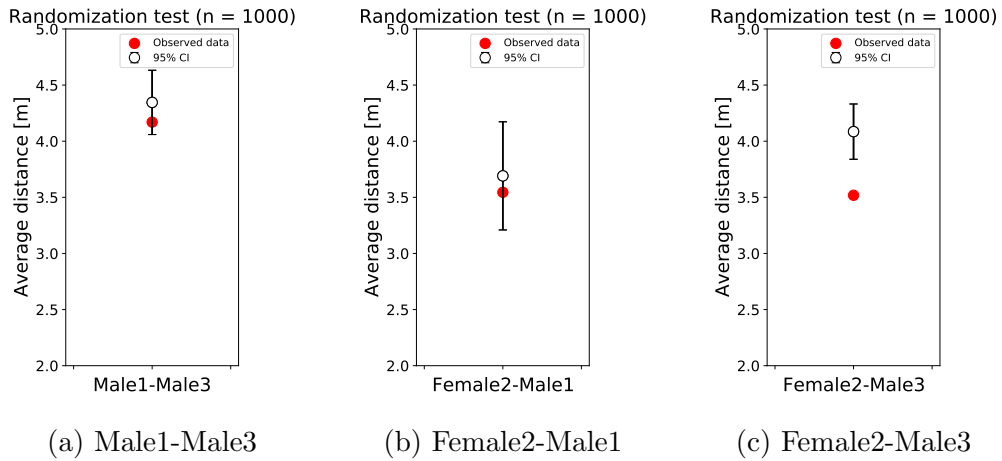


図 5.28: 実験 8 における発声時の個体間の平均距離とランダム化テスト ( $n = 1000$ ) による 95%信頼区間.

## 5.6 議論

本章では、屋外テント内に構築した観測環境での複数のマイクロホンアレイを用いた鳥類集団の相互作用を観測する枠組みを提案し、その検証としてキンカチョウを題材として2種の録音実験を行った。1つ目は、小規模集団における鳴き声の時空間ダイナミクスの抽出と分析に関する実験である。この実験では、比較的狭い場所で鳴き合う複数の鳥類の発声の状況が提案手法でどの程度詳細に観測可能かを検証した。複数のマイクロホンアレイに基づく定位手法は、止まり木や巣などのキンカチョウが鳴く位置を詳細に把握することができた。得られた定位結果と、次元削減アルゴリズムに基づいて抽出した音響特性情報を組み合わせることによって得られた結果からは、キンカチョウ集団が特定の発声を用いて相互作用する鳴き声に関する時空間的ダイナミクスの抽出に成功した。また、音響特性に基づいた歌と地鳴きの分類手法によって大まかな歌と地鳴きが分類に成功した。

2つ目の実験では、個体間の社会関係に注目して個体数・雌雄の変化によって個体の挙動がどのように変化するかを観測した。この実験では、VAEとSVMに基づく鳴き声の識別手法を提案し、これを実施した録音に適用した結果、個体の鳴き声が高い精度で識別され、実際の鳥類観測における分類課題にも適用可能であることが示された。また、この分類結果と2次元定位の結果を組み合わせることで、実験中の個体の発声状況を定量的に示すことに成功し、結果からは、実験個体が状況に応じて挙動を変化させている傾向が確認された。例えば、2羽のオスを観測環境に放った条件では、止まり木や巣を共有し合って鳴くことが観測され、また一方のオス個体の鳴き声他方呼び寄せの可能性があることが有意に示された。その一方で、メス個体を入れた条件ではオス個体同士の間隔の共有が減り、互いの距離が離れるようになった。この結果は、提案手法によってキンカチョウ個体の社会的状況を考慮した発声の時空間的ダイナミクスが抽出できたことを意味する。Vignalらの実験 [88] では、キンカチョウのオス個体がケージに入れられた個体との社会関係の違いに応じて、仲間個体の鳴き声に対する反応を変えることが報告されており、これは本実験で得られた結果と一致する。今回の分析では先行研究で得られた以上の知見を得るまでには至っていないが、個体間の距離関係に関してより詳細に調査を行うことで、コミュニケーションにおける社会関係・距離関係の優先度が議論可能であると考えられる。

発声頻度に関しても、社会的な状況によって差が生じた。例えば、1羽でいる場

合には歌・地鳴きともに発声回数が多い傾向にあった。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た。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。歌に関しては、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ことから [91], 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。2羽条件の場合には、地鳴きは各個体がそれぞれ比較的似た頻度で鳴いている結果が観測された。頻度が似ているという状況は、地鳴きとして抽出した distance call が個体間で特徴があり、個体認識の際に用いられる地鳴きである [88,89] ことを考慮すれば、一方の地鳴きに対して他方の個体が呼応して地鳴きを鳴き返した状況であることが推測される。歌に関しては、個体の組み合わせによって頻度が大きく変化している。Dunn らの報告によれば、飼育下のキンカチョウの歌頻度は、ペアの形成の有無や同種個体の存在の有無の影響を受けるという [92]。この際、結合の強い個体と同じケージに入れると歌の頻度は少なく、一方で馴染みのない個体と一緒に同じケージに入れると多くなる。今回の実験は実験用に借用した個体であり、実験前の飼育状況などが把握しきれておらず、事前の個体間関係を知ることが出来なかった。そのため、この歌頻度の結果は事前の飼育状況を反映した可能性がある。メス個体を投入した3羽条件では、歌・地鳴きの頻度は個体ごとに差があり、歌頻度が高い（低い）個体は地鳴き頻度も高い（低い）傾向にあった。特に、Male1 と Male2 を用いた実験では、Male2 がメス個体と近くで鳴く傾向がある一方で、Male1 は Male2, Female1 どちらからも遠い場所で鳴くというように、個体間の位置関係も大きく違っていた。Male2 の歌は極端に少ないが、ペアの絆が形成されると歌の意味を失うという報告もあり [82], これを考慮すれば、Male2 と Female1 は既にペア形成の段階に至っていた可能性がある。また、2つの3羽条件での実験に関して、それぞれの個体の鳴き声に対して他個体の鳴き声がどの程度重複があったかを分析したところ、発声が多い個体の地鳴きが他方の歌に重複する傾向が確認された。このことから、キンカチョウの地鳴き（distance call）は歌を阻害するものとして用いられる可能性があることが推測される。

今回用いた音源定位手法は、前章で述べた半自動的な音環境の概観の検討の際に提案した方法に基づくものである。自然環境での適用時には鳥類の鳴き声を漏らさずに定位する場合、その他のノイズを多く含めてしまうという問題があったが、今

回の構築環境での適用では、定位された音源のほとんどがキンカチョウの鳴き声であった。このように、実際の自然環境では調整の必要な観測手法が、空間的に制限された観測環境においてはうまく機能するという知見が得られたのは今後の枠組みの発展を検討する上で非常に大きな成果であるといえる。

両実験共に実験のサンプル数は少ないが、観測された個体の挙動の中には妥当といえる傾向もあり、十分に本提案手法が個体間相互作用の観測・理解に有用な手段であることは示されたと考える。特に、従来の録音・観測手法では十分に得られなかった個体間の距離といった空間的情報が得られるのは大きな利点であるといえる。

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が [24,85]、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、今回の観測・分析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。

## 第6章 結論

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理解に焦点を合わせた、ロボット聴覚技術を用いた観測・分析手法の実現を目的として、複数の人工的環境に基づいた実験・分析及び分類手法の枠組みの構築とその検証を行った。

まず最初に、HARKBird をロボット聴覚ソフトウェア HARK のリアルタイム処理に基づいてプレイバック・インタラクティブ実験が実施可能なシステムに拡張し、1台のマイクロホンアレイによって得られる方向情報から、ウグイス個体に対するプレイバックの影響が抽出可能か調査した。その結果、概してプレイバックにウグイス個体は反応を示し、プレイバックの頻度によって与える影響が異なるという傾向を定量的に示すことに成功した。また、再生間隔は他の一定間隔でのプレイバック実験よりも長いインタラクティブ実験において、より強い影響をウグイス個体と与えることが確認され、提案手法によるプレイバック・インタラクティブ実験の生態観測への適用可能性が示された。さらに、定位方向に基づいた発声と移動に関する調査では、遠方への移動と見なせる音源定位方向の大きな変異があった際の発声には警戒を示すL型が用いられる傾向も観測された。この研究結果からは、1台のマイクロホンアレイという、最小限のシステム構成による観測においても生態観測が十分に可能であることが示された。

次に、より詳細な挙動の変化の抽出と提案システムのより多様な環境での適用を目指し、ホシワキアカトウヒチョウに対するプレイバック実験を題材とした2台のマイクロホンアレイによる音源の2次元定位と音源の分類課題に向けた検討を行った。本実験では、次元削減アルゴリズムを用いた定位音源の分類手法と半自動的な音源定位手法の提案を行い、鳥類生態理解と音風景理解というスケールの異なる状況において鳴き声の観測を試みた。その結果、提案手法は音源の分類における時間的コストの削減に貢献し、音風景の概観にも成功した。また、2次元定位結果からは、ホシワキアカトウヒチョウが、プレイバックの種類によって発声頻度や種類、スピーカへの接近傾向などを変えている状況を詳細に把握し、時空間的ダイナミクスの抽出に成功した。本実験によってロボット聴覚技術による鳥類生態観測の枠組み

は拡張された。

最後に、多数個体が鳴き合う、より複雑な状況での生態理解のための観測手法の提案として、屋外テント内に構築したフリーフライト観測環境での複数キンカチョウ個体の鳴き声の観測を試みた。多数個体で構成された集団内の発声の様子の抽出を目的とした1つ目の実験では、自然環境と比較して狭い観測環境で鳴き合い、動く複数の鳥類の発声が複数のマイクロホンアレイに基づく観測手法によってどの程度詳細に観測可能かを検証した。提案手法による2次元定位は、止まり木や巣などのキンカチョウが鳴く位置を詳細に把握することができた。また、この定位結果と定位音源の音響特性情報を組み合わせることで、キンカチョウ集団の鳴き声に関する時空間的ダイナミクスの抽出に成功した。個体間の社会的関係に焦点を合わせた2つ目の実験では、VAEとSVMに基づく鳴き声の分類手法の提案と雌雄の数を変更させた録音実験を行うことで、社会状況の変化によって生じる個体間関係の変化の抽出を試みた。提案分類手法は、実験個体の鳴き声の特徴をうまく把握して、ノイズの除去と鳴き声の識別に貢献した。また、観測された結果からは、状況によって結合の度合いが変化するオス個体の挙動の変化の観測に成功した。これは、観測環境を含めた本提案手法が詳細な社会的個体間相互作用の理解に貢献していることを示唆する。

改めて序論で述べた本研究での取り組んだ3つの課題を整理する：(1) 鳥類の積極的な相互作用が観測可能な実験環境の構築、(2) マイクロホンアレイの使用台数に基づいた分析手法の構築と得られる生態データの検討、(3) 分離音源の分類手法の構築と検討。

(1) に関して、本研究ではプレイバック実験およびフリーフライト環境での観測実験の枠組みを提案した。前者では、プレイバックあり、なしの条件を比較すると実験個体の挙動は大きく変化する傾向が観測された。また、後者では複数個体が同時に鳴く状況や、観測環境内の個体数の変化に伴って鳥類個体の挙動が変化する様子が観測された。以上の結果から、提案した2つの手法によって積極的に鳥類個体が鳴く状況を作り出すことに成功したといえる。しかし、これらの実験は最小限の取り組みであり、まだ十分な検討とは言えない。例えば、プレイバック実験では1 vs 1を想定した、かつスピーカの位置を固定した最小限の実験である。相互作用において移動が生じないのは不自然な状況であり、発展としてはスピーカを複数台用いた複数個体あるいは単一個体の移動を想定したプレイバック実験などが考えられる。また、今後のフリーフライト環境のより幅広い適用を考える場合には、様々



な広さ、種での検討が必要であると考える。

(2) に関して、本研究では1台、2台、5台のマイクロホンアレイを用いた観測を実施した。1台のマイクロホンアレイを用いた観測では、実験個体の移動状況が方向情報からある程度推測可能な環境での実験ではあるものの、プレイバック条件によって異なる挙動の観測と移動と歌の種類に関する新奇的発見に成功した。マイクロホンアレイ2台を用いた2次元定位による観測では、1台で得られる情報に加えて距離に関する情報が得られたため、スピーカへの接近傾向などのより詳細な実験個体の挙動の抽出に成功した。5台のマイクロホンアレイを用いた観測では、観測する場所に応じて用いるマイクロホンアレイを変更させる手法を提案した。その結果、同じ止まり木で定位された音源がはっきりと左右で分かれるなどの細かい解像度での鳴き声位置の抽出や、組み合わせによって異なる詳細な個体間距離の抽出に成功した。以上の結果から、多数のマイクロホンアレイを用いる場合、定位結果の統合にはコストが掛かるが定位結果の解像度の向上に貢献することが示された。観測したい対象に応じて、観測手法を適宜選ぶことが鳥類観測には重要であるといえる。この課題に関して検討はまだ不十分であり、今後取り組むべきことはある。例えば、複数マイクロホンアレイの自然環境での適用などが挙げられる。提案手法による複数マイクロホンアレイの利用は音源定位担当範囲のみに基づいており、台数などに制限はない。そのため、より多数のマイクロホンアレイをより広範囲に展開することによって群集レベルでの相互作用・音環境の把握が可能であると考えられる。これは、生態系保護の観点から実施される環境調査といった領域にも展開できると考える。

(3) に関して、3つの実験では時系列データに基づく手動分類（ウグイスの実験）・音響特性に基づく分類支援ツールを用いた手動分類（ホシワキアカトウヒチョウの実験）、機械学習による自動分類（キンカチョウの実験）をそれぞれ行った。ウグイスの実験では、実験中は実験個体以外の鳥類は頻繁に鳴いておらず、定位された音源の多くが実験個体の鳴き声であった。また、分析時間が比較的短かったために鳴き声は容易に識別できた。しかし、多くの鳥類が鳴く状況での録音から実験個体の鳴き声のみを抽出する場合、より長時間の録音の分析を行う場合には、困難な作業となるだろう。ホシワキアカトウヒチョウの実験では、分離音源のスペクトログラム画像を次元削減アルゴリズムで学習し、特徴量空間を生成した。その特徴量空間上では、鳴き声は他の音と異なるクラスターを形成し、容易に鳴き声が抽出できた。一方、注目個体の鳴き声データが少ない場合はその特徴をうまく学習すること

ができず、利用には制約があることが示された。キンカチョウの実験で用いたVAEとSVMに基づく分類は、学習データに用いた鳴き声の特徴をうまく学習し、ノイズの除去と高い精度の鳴き声の識別を実現した。また、本手法は鳴き声データが少ない場合にもうまく機能する手法である。学習データを準備する必要はあるため、そのコストを考えれば長期的な録音を分析する際に有効な手法であるといえる。一方で、今回の分類手法では個体で差のある歌および地鳴き（distance call）のみ抽出しており、その他の地鳴きは考慮していない。これは、単調でシンプルな地鳴きの分類が提案手法では困難であったためである。今後は小さな鳴き声の特徴の差も検出可能な他の分類手法を検討する必要がある。さらに、そのような個体識別が実現された場合、HARKBirdとの組み合わせによって更に柔軟な実験の枠組みが構築可能であると考えられる。例えば、プレイバック機能に関して、特定の個体の鳴き声のみが定位された場合のみプレイバックが行えれば、周囲で多くの鳥が鳴く環境においても自動的なインタラクティブ実験が可能となる。

以上のように、今回提案した実験および分析手法は、録音に用いるマイクロホンアレイの台数をはじめとして、実験手法・実験環境などが変化した場合でも鳥類の発声に基づく生態観測を様々な状況・スケールでの生態観測を可能にした。これは、本フレームワークが観測したい対象に応じて適切に手法を切り替えられることを意味し、高い適用可能性を示しているといえる。今後は、上記の大規模スケールでの観測や、事例のない鳥類種の生態観測などへの適用、システムの拡張を検討し、さらなる知見を増やしていくことが重要である。これは、様々な状況での利用を考慮したロボット聴覚技術だからこそ可能であると考えられる。

# 謝辞

本論文の研究の一部は JSPS 科研費 JP21K12058, JP20J13695, JP20H00475, JP19KK0260, JP18K11467, JP17H06841, JP17H06383 (#4903), JP16K00294, 15K00335, 24220006 の助成を受けて行われました。本論文をまとめるにあたり、名古屋大学大学院情報学研究科鈴木麗璽准教授、有田隆也教授には丁寧な指導と励ましを承りました。厚く御礼申し上げます。また、各実験は様々な方々の協力のおかげで実施することができました。ウグイスの実験では、肘井直樹教授、梶村恒准教授、高部直紀氏、山口法雄氏（名古屋大学）のフィールド調査への協力（実施にあたっての相談、送迎、環境管理等）、ホシワキアカトウヒチョウの実験では、Martin Cody 名誉教授、Charles Taylor 名誉教授（University of California, Los Angeles）らの実験環境の提供及び実験手法に関する提案や助言、キンカチョウの実験では、柴田ゆき野氏（北海道大学）の実験補助、和多和宏准教授（北海道大学）の実験環境・実験個体の提供及び実験手法に関する提案や助言を頂きました。さらに、中臺一博教授（東京工業大学、ホンダ・リサーチ・インスティテュート・ジャパン）、奥乃博名誉教授（京都大学）にはロボット聴覚技術に関する相談および助言、松林志保特任准教授（大阪大学）には各実験において、生態学的側面からの助言を頂きました。厚く御礼申し上げます。

最後に、有田・鈴木研究室のメンバーには常に刺激的な議論を頂き、精神的にも支えられました。本当にありがとうございました。特に、博士前期課程から共に研究活動を行ってきた大和祐介氏の存在は研究を進める上で大きな存在となりました。感謝の意を表します。

## 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 P. Marler and H. Slabbekoorn, editors, *Nature's Music*, pp. 132–177. Academic Press, San Diego, 2004.
- [3] M. L. Cody and J. H. Brown. Song asynchrony in neighbouring bird species. *Nature*, Vol. 222, pp. 778–780, 1969.
- [4] Robert W. F., Millicent S. F., and Jack P. H. Temporal pattern shifts to avoid acoustic interference in singing birds. *Science*, Vol. 183, No. 4126, pp. 762–763, 1974.
- [5] B. Krause. Bioacoustics: Habitat ambience and ecological balance. *Whole earth review*, pp. 14–18, 1987.
- [6] R. W. Ficken, M. S. Ficken, and J. P. Hailman. Temporal pattern shifts to avoid acoustic interference in singing birds. *Science*, Vol. 183, No. 4126, pp. 762–763, 1974.
- [7] H. Brumm. Signalling through acoustic windows: nightingales avoid interspecific competition by short-term adjustment of song timing. *Journal of Comparative Physiology A: Neuroethology*, Vol. 192, pp. 1279–1285, 2006.
- [8] R. Planqué and H. Slabbekoorn. Spectral overlap in songs and temporal overlap in a peruvian bird assemblage. *Ethology*, Vol. 114, pp. 262–271, 2008.
- [9] R. Suzuki, C. E. Taylor, and M. L. Cody. Soundscape partitioning to increase communication efficiency in bird communities. *Artificial Life and Robotics*, Vol. 17, No. 1, pp. 30–34, 2012.

- [10] R. Suzuki and M. L. Cody. Complex systems approaches to temporal soundspace 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.
- [11] A. Farina and S. H. Gage. *Ecoacoustics: The Ecological Role of Sounds*. John Wiley and Sons, 2017.
- [12] A. Gasc, D. Francomano, J. B. Dunning, and B. C. Pijanowski. Future directions for soundscape ecology: The importance of ornithological contributions. *The Auk*, Vol. 134, No. 1, pp. 215–228, 2016.
- [13] D. Stowell. *Computational Bioacoustic Scene Analysis*, chapter 11, pp. 303–333. 2018.
- [14] Pertti Koskimies. Birds as a tool in environmental monitoring. *Annales Zoologici Fennici*, Vol. 26, No. 3, pp. 153–166, 1989.
- [15] R. Bardeli, D. Wolff, F. Kurth, M. Koch, K.-H. Tauchert, and K.-H. Frommolt. Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring. *Pattern Recognition Letters*, Vol. 31, No. 12, pp. 1524–1534, 2010.
- [16] J. Shonfield and E. M. Bayne. Autonomous recording units in avian ecological research: Current use and future applications. *Avian Conservation and Ecology*, Vol. 12, No. 14, 2017.
- [17] A. Digby, M. Towsey, B. D. Bell, and P. D. Teal. A practical comparison of manual and autonomous methods for acoustic monitoring. *Methods in Ecology and Evolution*, Vol. 4, pp. 675–683, 2013.
- [18] S. Vold, C. M. Handel, and L. B. McNew. Comparison of acoustic recorders and field observers for monitoring tundra bird communities: Acoustic monitoring of tundra-breeding birds. *Wildlife Society Bulletin*, Vol. 41, pp. 566–576, 2017.

- [19] C. Abrahams. Comparison between lek counts and bioacoustic recording for monitoring western capercaillie (*tetrao urogallus* l.). *Journal of Ornithology*, Vol. 160, No. 3, pp. 685–697, 2019.
- [20] K. Darras, P. Batáry, B. J. Furnas, I. Grass, Y. A. Mulyani, and T. Tscharnke. Autonomous sound recording outperforms human observation for sampling birds: a systematic map and user guide. *Ecological Applications*, Vol. 29, No. 6, 2019. Art. no. e01954.
- [21] Q. Kong, Y. Xu, W. Wang, P. J. B. Jackson, and M. D. Plumbley. Single-channel signal separation and deconvolution with generative adversarial networks. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI2019)*, 2019.
- [22] D. A. Yip, E. C. Knight, E. H. Audet, S. J. Wilson, C. Charchuk, C. D. Scott, P. Sólymos, and E. M. Bayne. Sound level measurements from audio recordings provide objective distance estimates for distance sampling wildlife populations. *Remote Sensing in Ecology and Conservation*, Vol. 6, pp. 301–315, 2019.
- [23] D. R. Farine, L. M. Aplin, B. C. Sheldon, and W. Hoppitt. Interspecific social networks promote information transmission in wild songbirds. *Proceedings of the Royal Society B: Biological Sciences*, Vol. 282, No. 1803, 2015. Art. no. 20142804.
- [24] 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.
- [25] D. Stowell, E. Benetos, and L. Gill. On-bird sound recordings: Automatic acoustic recognition of activities and contexts. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 25, pp. 1193–1206, 2016.
- [26] 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, , 2016. Art. no. 20160296.

- [27] D. Todt and M. Naguib. Vocal interactions in birds: The use of song as a model in communication. *Advances in the Study of Behavior*, Vol. 29, pp. 247–296, 2000.
- [28] M. Naguib. *Singing interactions in songbirds: implications for social relations and territorial settlement*, pp. 300–319. Cambridge University Press, 2005.
- [29] D. Blumstein, D. J. Mennill, P. Clemens, L. Girod, K. Yao, G. Patricelli, J. L. Deppe, A. H. Krakauer, C. Clark, K. A. Cortopassi, S. F. Hanser, B. McCowan, A. M. Ali, and A. N. G. Kirshel. Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus. *Journal of Applied Ecology*, Vol. 48, pp. 758–767, 2011.
- [30] H. Goëau, Glotim H., W. P. Vellinga, R. Planqué, and A. Joly. LifeCLEF Bird Identification Task 2016: The arrival of Deep learning. *LifeCLEF 2016 working notes*, 2016.
- [31] 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.
- [32] D. G. Richards and R. H. Wiley. Reverberations and amplitude fluctuations in the propagation of sound in a forest: Implications for animal communication. *The American Naturalist*, Vol. 115, pp. 381–399, 1980.
- [33] D. J. Mennill, J. M. Burt, K. M. Fristrup, and S. L. Vehrencamp. Accuracy of an acoustic location system for monitoring the position of duetting songbirds in tropical forest. *The Journal of the Acoustical Society of America*, Vol. 119, No. 5, pp. 2832–2839, 2006.
- [34] D. J. Mennill, M. Battiston, and D. R. Wilson. Field test of an affordable, portable, wireless microphone array for spatial monitoring of animal ecology and behaviour. *Methods in Ecology and Evolution*, Vol. 3, pp. 704–712, 2012.
- [35] T. C. Collier, A. N. G. Kirschel, and C. E. Taylor. Acoustic localization of antbirds in a Mexican rainforest using a wireless sensor network. *The Journal of the Acoustical Society of America*, Vol. 128, pp. 182–189, 2010.

- [36] Z. Harlow, T. Collier, V. Burkholder, and C. E. Taylor. Acoustic 3d localization of a tropical songbird. In *IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP)*, pp. 220–224, 2013.
- [37] P.M Stepanian, Hille D.C. Horton, K. G., C.E Wainwright, P.B. Chilson, and J.F. Kelly. Extending bioacoustic monitoring of birds aloft through flight call localization with a three-dimensional microphone array. *Ecology and Evolution*, Vol. 6, pp. 7039–7046, 2016.
- [38] M. Araya-Salas, K. Wojczulanis-Jakubas, E. M. Phillips, D. J. Mennill, and T. F. Wright. To overlap or not to overlap: context-dependent coordinated singing in lekking long-billed hermits. *Animal Behaviour*, Vol. 124, pp. 57–64, 2017.
- [39] J. P. Ethier and D. R. Wilson. Using microphone arrays to investigate micro-habitat selection by declining breeding birds. *Ibis*, Vol. 162, No. 3, pp. 873–884, 2020.
- [40] Z. Gayk and D. J. Mennill. Pinpointing the position of flying songbirds with a wireless microphone array: three-dimensional triangulation of warblers on the wing. *Bioacoustics*, Vol. 29, pp. 1–12, 2019.
- [41] R. W. Hedley, Y. Huang, and K. Yao. Direction-of-arrival estimation of animal vocalizations for monitoring animal behavior and improving estimates of abundance. *Avian Conservation and Ecology*, Vol. 12, pp. 116–128, 2017.
- [42] 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.
- [43] H. G. Okuno and K. Nakadai. Robot audition: Its rise and perspectives. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5610–5614, 2015.



- [44] A. J. Pierce and K. Pobprasert. A portable system for continuous monitoring of bird nests using digital video recorders. *Journal of Field Ornithology*, Vol. 78, No. 3, pp. 322–328, 2007.
- [45] 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.
- [46] X. Anguera, C. Wooters, and J. Hernando. Acoustic beamforming for speaker diarization of meetings. *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 15, No. 7, pp. 2011–2021, 2007.
- [47] F. Grondin and F. Michaud. Lightweight and optimized sound source localization and tracking methods for open and closed microphone array configurations. *Robotics and Autonomous Systems*, Vol. 113, pp. 63–80, 2019.
- [48] 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.
- [49] 松林志保 鈴木麗璽 有田隆也 中臺一博 奥乃 博 千葉尚彬. ロボット聴覚技術を活用した野鳥の歌行動観測・分析ツール harkbird の機能強化. 第35回ロボット学会学術講演会講演概要集, 2017. Art. no. 3A3-03.
- [50] 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.
- [51] S. Matsubayashi, R. Suzuki, F. Saito, T. Murate, T. Masuda, K. Yamamoto, R. Kojima, K. Nakadai, and H. G. Okuno. Acoustic monitoring of the great reed warbler using multiple microphone arrays and robot audition. *Journal of Robotics and Mechatronics*, Vol. 27, pp. 224–235, 2017.
- [52] 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.
- [53] 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.
- [54] 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.
- [55] 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.
- [56] R. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, Vol. 34, pp. 276–280, 1986.
- [57] 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.
- [58] T. Dabelsteen and S. B. Pedersen. Song and information about aggressive responses of blackbirds, *turdus merula*: evidence from interactive playback experiments with territory owners. *Animal Behaviour*, Vol. 40, No. 6, pp. 1158–1168, 1990.
- [59] S. Hamao. Japanese bush warbler. bird research news, 2007. [http://www.bird-research.jp/1\\_shiryō/seitai/uguisu.pdf](http://www.bird-research.jp/1_shiryō/seitai/uguisu.pdf).
- [60] H. Momose. *Mechanism of maintaining territories by acoustic communication*. Reproductive strategies of birds (S. Yamagishi, Editor). Toukaidaigaku-shuppankai, 1986.

- [61] S. Hamao. Acoustic structure of songs in island populations of the Japanese bush warbler, cettia diphone, in relation to sexual selection. Vol. 31, pp. 9–15, 2012.
- [62] S. Hamao. Differences in vocalizations of japanese bush warblers on chichijima and hahajima in the ogasawara islands. *Bulletin of the National Museum of Nature and Science, Series A (Zoology)*, Vol. 44, No. 4, pp. 125–132, 2018.
- [63] D. F. Maynard. The role of vocalizations in mate attraction and male-male interactions. Master’s thesis, 2012.
- [64] P. K. McGregor. Quantifying responses to playback: One, many, or composite multivariate measures? In P. K. McGregor, editor, *Playback and Studies of Animal Communication*, chapter 6, pp. 79–96. Springer, 1992.
- [65] C. K. Catchpole and P. J. B. Slater. *Bird Song: Biological Themes and Variations*. Cambridge University Press, 2008.
- [66] L. van der Maaten and G. Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, Vol. 9, pp. 2579–2605, 2008.
- [67] 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.
- [68] Spotted Towhee Identification, All About Birds. Cornell Lab of Ornithology, 2020. <http://www.allaboutbirds.org>.
- [69] S. L. Small, F. R. III Thompson, G. R. Geupel, and J. Faaborg. Spotted Towhee Population Dynamics in a Riparian Restoration Context. *The Condor*, Vol. 109, No. 4, pp. 721–733, 2007.
- [70] S. Sumitani, R. Suzuki, S. Matsubayashi, K. Arita, T. Nakadai, and H. G. Okuno. Extracting the relationship between the spatial distribution and types of bird vocalizations using robot audition system HARK. In *Proceedings of 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2485–2490, 2018.

- [71] T. Sainburg, M. Thielk, and T. Q. Gentner. Latent space visualization, characterization, and generation of diverse vocal communication signals. *bioRxiv*, 2019.
- [72] D. P. Kingma and M. Welling. Auto-Encoding Variational Bayes. In *2nd International Conference on Learning Representations, ICLR 2014*, 2014.
- [73] C. Cortes and V. Vapnik. Support vector networks. *Machine Learning*, Vol. 20, pp. 273–297, 1995.
- [74] 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.
- [75] P. J.B. Slater, L. A. Eales, and N.S. Clayton. Song learning in zebra finches (*taeniopygia guttata*): Progress and prospects. Vol. 18 of *Advances in the Study of Behavior*, pp. 1–34. Academic Press, 1988.
- [76] P. J.B. Slater. Bird song learning: causes and consequences. *Ethology Ecology & Evolution*, Vol. 1, No. 1, pp. 19–46, 1989.
- [77] J. Böhner. Song learning in the zebra finch (*taeniopygia guttata*): Selectivity in the choice of a tutor and accuracy of song copies. *Animal Behaviour*, Vol. 31, No. 1, pp. 231–237, 1983.
- [78] H. Williams. Models for song learning in the zebra finch: fathers or others? *Animal Behaviour*, Vol. 39, No. 4, pp. 745–757, 1990.
- [79] N. J. Boogert, L. Giraldeau, and L. Lefebvre. Song complexity correlates with learning ability in zebra finch males. *Animal Behaviour*, Vol. 76, No. 5, pp. 1735–1741, 2008.
- [80] O. Fehér, H. Wang, S. Saar, P. Mitra, and O. Tchernichovski. pp. 564–568, 2009.
- [81] F. Jesse and K. Riebel. Social facilitation of male song by male and female conspecifics in the zebra finch, *taeniopygia guttata*. *Behavioural Processes*, Vol. 91, No. 3, pp. 262–266, 2012.

- [82] M. Tomaszycki and E. Adkins-Regan. Is male song quality important in maintaining pair bonds? *Behaviour*, Vol. 143, pp. 549–567, 2006.
- [83] R. A. Zann and M. Bamford. *The Zebra Finch: A Synthesis of Field and Laboratory Studies*. Oxford ornithology series. Oxford University Press, 1996.
- [84] J. Elie and F. Theunissen. The vocal repertoire of the domesticated zebra finch: a data-driven approach to decipher the information-bearing acoustic features of communication signals. *Animal cognition*, Vol. 19, pp. 285–315, 2016.
- [85] P. D’Amelio, L. Trost, and A. Maat. Vocal exchanges during pair formation and maintenance in the zebra finch (*taeniopygia guttata*). *Frontiers in Zoology*, Vol. 14, No. 13, 2017.
- [86] M. Ikebuchi and K. Okanoya. Growth of pair bonding in zebra finches: physical and social factors. *Ornithological Science*, Vol. 5, pp. 65–75, 2006.
- [87] R. Zann. Structural variation in the zebra finch distance call. *Zeitschrift für Tierpsychologie*, Vol. 66, No. 4, pp. 328–345, 1984.
- [88] C. Vignal, N. Mathevon, and S. Mottin. Audience drives male songbird response to partner’s voice. *Nature*, Vol. 430, pp. 448–451, 2004.
- [89] C. Vignal, N. Mathevon, and S. Mottin. Mate recognition by female zebra finch: Analysis of individuality in male call and first investigations on female decoding process. *Behavioural Processes*, Vol. 77, No. 2, pp. 191–198, 2008.
- [90] L. McInnes and J. Healy. UMAP: Uniform manifold approximation and projection for dimension reduction. ArXiv e-prints 1802.03426, 2018.
- [91] G. E. Hough and S. 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.
- [92] A. M. Dunn and R. Zann. Effects of pair bond and presence of conspecifics on singing in captive zebra finches. *Behaviour*, Vol. 134, No. 1/2, pp. 127–142, 1997.

# 研究業績一覧

## 主論文に関連する研究業績

### 主著者論文等

- 査読付き学術論文

- [1] 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 (2021/04).
- [2] Shinji Sumitani, Reiji Suzuki, Shiho Matsubayashi, Takaya Arita, Kazuhiro Nakadai and Hiroshi G. Okuno: Fine-scale observations of spatio-spectro-temporal dynamics of bird vocalizations using robot audition techniques, *Remote Sensing in Ecology and Conservation*, 7 (1), 10.1002/rse2.152 (2020/03).
- [3] Reiji Suzuki, Shinji Sumitani, Naren, Shiho Matsubayashi, Takaya Arita, Kazuhiro Nakadai and Hiroshi G. Okuno: Field observations of ecoacoustics dynamics of a Japanese bush warbler using an open-source software for robot audition HARK, *Journal of Ecoacoustics*, 2 (3), EYAJ46 (2018/06).

- 査読付き学術雑誌論文

- [4] Shinji Sumitani, Reiji Suzuki, Kazuhiro Wada, Takaya Arita, Kazuhiro Nakadai, Hiroshi G. Okuno: A robot audition approach toward understanding social interactions among songbirds in a semi-free flight environment, *Proceedings of the 25th International Symposium on Artificial Life and Robots (AROB)*, pp. 986–981 (2020/01).

- [5] Shinji Sumitani, Reiji Suzuki, Takemi Morimatsu, Shiho Matsubayasi, Takaya Arita, Kazuhiro Nakadai and Hiroshi G. Okuno: Soundscape Analysis of Bird Songs in Forests Using Microphone Arrays, *Proceedings of 2020 IEEE/SICE International Symposium on System Integration (SII)*, pp. 634–639 (2020/01).
- [6] Shinji Sumitani, Reiji Suzuki, Naoaki Chiba, Shiho Matsubayashi, Takaya Arita, Kazuhiro Nakadai and Hiroshi G. Okuno: An Integrated Framework for Field Recording, Localization, Classification and Annotation of Birdsongs Using Robot Audition Techniques — Harkbird 2.0, *Proceedings of 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8246–8250 (2019/05).
- [7] Shinji Sumitani, Reiji Suzuki, Shiho Matsubayashi, Takaya Arita, Kazuhiro Nakadai, Hiroshi G. Okuno: Extracting the Relationship between the Spatial Distribution and Types of Bird Vocalizations Using Robot Audition System HARK, *Proceedings of 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2485–2490 (2018/10).
- 国内会議予稿
    - [8] 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, pp. 12-20 (2021/11).
    - [9] 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 (2021/09).
    - [10] 炭谷 晋司, 鈴木 麗璽, 和多 和宏, 有田 隆也, 中臺 一博, 奥乃 博: ロボット聴覚技術を用いた鳥類小集団の相互作用観測環境の構築と予備調査, 第37回日本ロボット学会学術講演会予稿集, 2I1-03 (2020/09).
    - [11] 炭谷 晋司, 鈴木 麗璽, 松林 志保, 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 回人工知能学会 AI チャレンジ研究会, pp. 6–11 (2019/11).

- [12] 炭谷 晋司, 松林 志保, 鈴木 麗璽: ウグイスに対するプレイバック実験におけるマイクロホンアレイを用いたさえずりの方向分布分析, 第 46 回人工知能学会 AI チャレンジ研究会, pp. 18–24, (2016/11).

- 国際会議での口頭発表

[4]–[7] をそれぞれ会議で口頭発表

- [13] Shinji Sumitani, Reiji Suzuki, Shiho Matsubayashi, Takaya Arita, Kazuhiro Nakadai, Hiroshi G. Okuno: Understanding relationships between spatial movements and bird song-types using a robot audition system HARK with microphone arrays, *27th International Ornithological Congress*, O39.06 (2018/08).

- 国際会議でのポスター発表

- [14] Shinji Sumitani, Takemi Morimatsu, Reiji Suzuki, Shiho Matsubayashi, Takaya Arita, Kazuhiro Nakadai and Hiroshi G. Okuno: Robot Audition Approaches to Field Observation of Bird Songs, *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2019)*, WePS2.33, (2019/11).
- [15] Shinji Sumitani, Reiji Suzuki, Shiho Matsubayashi, Takaya Arita, Kazuhiro Nakadai and Hiroshi G. Okuno: Spatial localization of vocalizations of Spotted Towhee (*Pipilo maculatus*) in playback experiments using robot audition techniques, *10th International Conference on Ecological Informatics (ICEI2018)*, pp. 265 (2018/10).
- [16] Shinji Sumitani, Reiji Suzuki, Takaya Arita, Naren, Shiho Matsubayashi, Kazuhiro Nakadai and Hiroshi G. Okuno: Field observations and Virtual Experiences of Bird Songs in the Soundspace Using an Open-Source Software for Robot Audition HARK, *4th International Symposium on Acoustic Communication by Animals*, pp. 116–117 (2017/07).

- 国内会議での口頭発表



[8] – [11] をそれぞれの会議で発表.

- [17] 炭谷 晋司, 鈴木 麗璽, 松林 志保, 有田 隆也, 中臺 一博, 奥乃 博: ロボット聴覚技術に基づく鳥類音声の録音・分析手法の開発, バードリサーチ鳥類学大会 2020 Online, pp. 10 (2020/12).

● 国内会議でのポスター発表

- [18] 炭谷 晋司, 鈴木 麗璽, 有田 隆也, 和多 和宏: フリーフライト環境における鳥類集団の歌行動の観測, 新学術領域「共創的コミュニケーションのための言語進化学」第4回領域全体会議 (2020/04).
- [19] 炭谷 晋司, 鈴木 麗璽, 和多 和宏, 有田 隆也, 中臺 一博, 奥乃 博: 鳥類集団の音声コミュニケーション理解のための半野外音源定位環境の構築と予備的調査, 日本鳥学会 2019 年度大会予稿集, P80, (2019/09).
- [20] 炭谷 晋司, 鈴木 麗璽, 松林 志保, 有田 隆也, 中臺 一博, 奥乃 博: ロボット聴覚技術に基づく鳥類に対するプレイバック実験における歌行動の観測, 第66回日本生態学会要旨集, P1-068, (2019/03).
- [21] 炭谷 晋司, 鈴木 麗璽, 松林 志保, 有田 隆也, 中臺 一博, 奥乃 博: ロボット聴覚技術に基づく鳥類の歌行動の二次元定位精度改善と次元圧縮に基づく分類支援, 日本鳥学会 2018 年度大会予稿集, P008, (2018/09).
- [22] 炭谷 晋司, 鈴木 麗璽, 有田隆也, 松林 志保, 中臺 一博, 奥乃 博: マイクロホンアレイを利用したウグイスの歌行動の時空間分析, 日本鳥学会 2017 年度大会予稿集, P022, (2017/09).
- [23] 炭谷 晋司, 松林 志保, 鈴木 麗璽: マイクロホンアレイを用いたプレイバック実験に基づくウグイスのさえずりの方向分布分析, 日本鳥学会 2016 年度大会予稿集, P014, (2016/09).

## 主著者以外の論文等

- [24] Reiji Suzuki, Hao Zhao, Shinji Sumitani, Shiho Matsubayashi, Takaya Arita, Kazuhiro Nakadai, Hiroshi G. Okuno: Visualizing Directional Soundscapes of Bird Vocalizations Using Robot Audition Techniques,

*Proceedings of 2021 IEEE/SICE International Symposium on System Integrations (SII)*, pp. 487-492 (2021/01).

- [25] Reiji Suzuki, Shinji Sumitani, Shiho Matsubayashi, Takaya Arita, Kazuhiro Nakadai and Hiroshi G. Okuno: Fine-scale observations of spatiotemporal dynamics and vocalization type of birdsongs using microphone arrays and unsupervised feature mapping, *Proceedings of the 10th International Conference on Ecological Informatics*, pp. 72–73 (2018/09).
- [26] Reiji Suzuki, Shinji Sumitani, Shiho Matsubayashi, Takaya Arita, Kazuhiro Nakadai and Hiroshi G. Okuno: Understanding ecoacoustic interactions among songbirds as complex systems using robot audition techniques, *Abstract Booklet of EVOSLACE: Workshop on the emergence and evolution of social learning, communication, language and culture in natural and artificial agents in ALIFE2018*, pp. 22 (2018/07).
- [27] Hao Zhao, Reiji Suzuki, Shinji Sumitani, Shiho Matsubayashi, Takaya Arita, Kazuhiro Nakadai, Hiroshi G. Okuno: Visualizing Soundscape of Animal Vocalizations in Forests Using Robot Audition Techniques, 第57回 AI チャレンジ研究会予稿集, pp. 15–20 (2020/11).
- [28] 鈴木 麗璽, Hao Zhao, 炭谷 晋司, 有田隆也, 中臺 一博, 奥乃 博: ロボット聴覚技術の活用による鳥類音声の到来方向に基づく音風景の可視化の検討, 第38回ロボット学会学術講演会予稿集, 1D3-01 (2020/10).
- [29] 森松 健充, 炭谷 晋司, 鈴木 麗璽, 松林 志保, 有田 隆也, 中臺 一博, 奥乃 博: 複数のマイクロホンアレイをネットワーク制御可能な鳥類の歌行動観測システムの構築, 第36回日本ロボット学会学術講演会予稿集, 2J2-03 (2018/08).
- [30] 千葉 尚彬, 炭谷 晋司, 松林 志保, 鈴木 麗璽, 有田 隆也, 中臺 一博, 奥乃 博: ロボット聴覚技術を活用した野鳥の歌行動観測・分析ツール HARKBirdの機能強化, 第35回日本ロボット学会学術講演会予稿集, 3A3-03 (2017/09).
- [31] 鈴木 麗璽, 炭谷 晋司, 松林 志保, 有田 隆也, 中臺 一博, 奥乃 博: 鳥類の鳴き声観測に対するロボット聴覚アプローチ, 日本生態学会第68回大会講演要旨集, P2-061 (2021/03).

- [32] 森松 健充, 炭谷 晋司, 鈴木 麗璽, 松林 志保, 有田 隆也, 中臺 一博, 奥乃 博: 複数のマイクロホンアレイの遠隔制御に基づく鳥類の歌行動の二次元定位, 日本鳥学会 2018 年度大会予稿集, P006, (2018/09).
- [33] 鈴木 麗璽, 炭谷 晋司, 奥乃 博, 中臺 一博: ロボット聴覚技術を用いた鳥類の歌行動分析の試みー複数のマイクロホンアレイを用いた二次元リアルタイム歌定位ー, 計測自動制御学会システムインテグレーション部門講演会 (SI2017), pp. 1124-1126, (2017/12).

## 受賞

- [34] [11] の発表に対して, 研究会優秀賞を受賞.
- [35] [18] の発表に対して, 優秀賞を受賞.
- [36] [33] の発表に対して, 優秀講演賞を受賞.