

マイクロフォンを用いた 鳥類の判別と音源位置特定

1250053 川村基貴

高知工科大学 システム工学群 建築・都市デザイン専攻

近年、里山や農地に生息する鳥類の減少が報告されている。一般的な生息数を観測する手法は観測者の経験に依存し、客観性や継続性に課題がある。本研究では、マイクロフォンから得た音声を用いて鳥の鳴き声を判別し、複数地点の音声データを用いて音源の位置を特定する手法を構築した。鳴き声の判別では周波数解析を用い種判定を行った。判定精度は全体で約 62%であり誤検出や見逃しも多かった。一方位置特定では、時間差を利用し最大で約 3.1m 誤差が生じた。より精度を向上させるために、周波数範囲と閾値の調整やパワースペクトルを活用した位置特定が課題である。

Key Words: 鳥類, 音声分類, 時刻同期, パワースペクトル, テンプレートマッチング

1. はじめに

環境省自然環境局生物多様性センターの調査によると、2009 年から 2020 年までの間に里山や農地・草地に生息する鳥類が減少していることが明らかになっている[1]。主な要因として、土地開発や里地・里山の整備不足による生息地の減少が挙げられる。鳥類が減少すると、植物の種子の散布が減少し、多様性の喪失や害虫の増加に伴う農作物の食害が増加することが経験則や推測によって議論されている。この問題を科学的に検証するためには、鳴き声の発生点の観測を自動的に多地点で行うことが不可欠である。

一般的な鳥類の観測手法にはラインセンサス法とスポットセンサス法がある。ラインセンサス法は決められたルートを歩きながら観測し、スポットセンサス法は定点で観測する。しかし、これらは観測者の労力がかかり連続的な観測としては客観性に課題があるため課題がある。

近年では、画像や音声を用いた観測手法が注目されている。画像を用いる方法では、対象が鳴かなくても観測できる利点があるが、観測範囲が限られる。一方、音声を用いる方法では、観測範囲が広い。鳴き声を発する生物しか検出できないという制約があるが連続的な観測が可能である。

そこで本研究では、音声を用いた連続的かつ客観的な観測を目指す。音声データから鳥の鳴き声を検出し、その位置を特定する手法の構築を目標

とする。先行研究では、深層学習を用いて鳥類を判別するモデルが提案されているが、判別手法がブラックボックス化するという問題がある[1]。また、音圧の差を用いた位置推定手法もあるが、精度が低いという課題が指摘されている[2]。これらの問題を解決するため、本研究では鳥類の判別においては周波数特性を利用し、位置特定には到達時間差を利用する手法を採用して試みた。

2. 鳥類の判別

2.1 観測フィールド

音声データは、高知県香美市土佐山田町後入に位置する佐岡研究フィールド母屋で取得した。本学からの位置と母屋周辺環境を図1に示す。

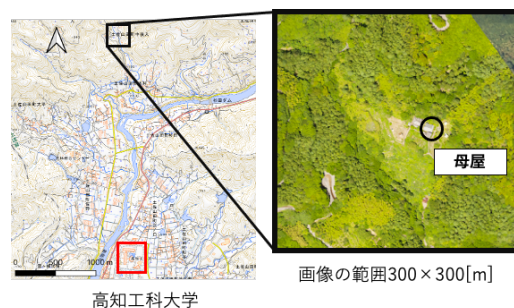


図1 対象地域

2.2 使用機材と録音状況

図2のようにUbuntuOSがインストールされたノートPC（Dynabook R73/NB24E）にマイクロフォン（ECM-PCV80U）を接続し、自作のPythonプログラムを実行し、録音を行った。録音設定は、サンプリングレートを44.1kHzとした。サンプリングレートとは、音声の1秒間に含まれるデータの個数を示すものである。

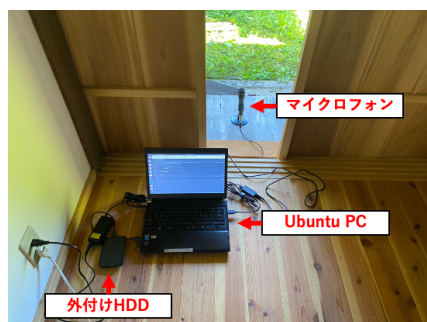


図2 録音環境と使用機材

表1に録音期間を示した。59分30秒のデータを1ファイルとして保存している。

表1 録音期間

2024年	5月	6月	7月	8月	9月
録音データ					
	10月	11月	12月	1月	2月

図3は録音した音声データから1分間を抽出したものである。音声データには、風音や、川の水音、雨音などの背景音が含まれる。赤い丸で囲っている部分がウグイスの鳴き声の波形である。

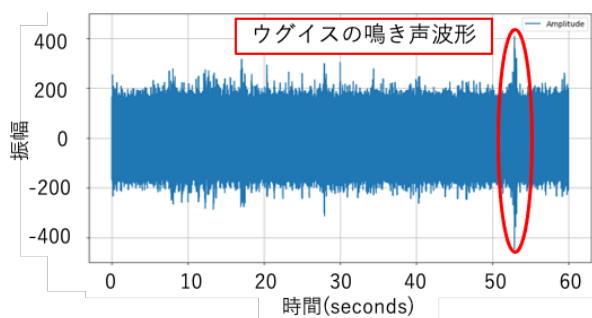


図3 1分間の音声データ

2.3 周波数解析

音声データを読み込み、時間領域の信号を周波数領域の信号に変換するために高速フーリエ変換を実施した。さらに、周波数ごとの振幅を取得し、パワースペクトルに変換して解析を行った。

図4はアカショウビンと生物の鳴き声が入っていない背景音のパワースペクトルをグラフ化したものである。

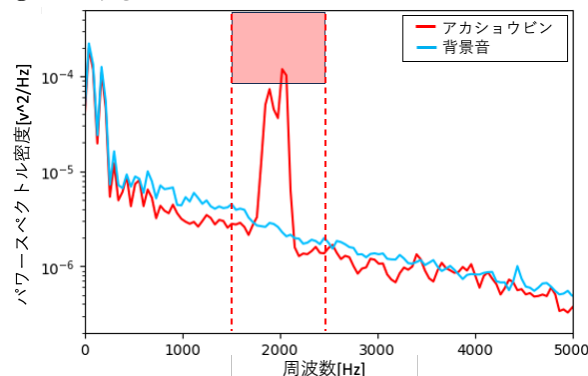


図4 パワースペクトル

背景音と比較し特徴的な周波数範囲内のパワースペクトル密度のピークの平均値を閾値として設定する。図4にアカショウビンの周波数範囲と閾値を網掛けで表示している。ピーク値が周波数範囲で閾値を超えた場合、その鳥類の鳴き声と判定する。1秒ずつ抽出した音声データのパワースペクトル密度からピーク値を求める。これを0.5秒ずつシフトさせて抽出を行った。

2.4 結果

2024年5月19日5:00～6:00の音声データを解析した結果、表2のような検出結果が得られた。教師データは聴覚を頼りに作成した。ウグイスとアカショウビンの周波数範囲が一部重なっており、ウグイスの閾値が低いため、アカショウビンの鳴き声がウグイスとして誤検出されるケースが多かった。また、カラスはウグイスやアカショウビンと比べ鳴き声が大きいため、場所によって音圧の差が大きくなる。音圧が低下するとパワースペクトルの形状は変わらないものの、値が小さくなるため、設定した閾値を超えず見逃しが多く発生した。

表2 検出結果

判別精度 約62%	鳴き声の 総回数	正しい 検出	鳴き声 見逃し	誤検出
ウグイス	88回	85回	3回	125回
アカショウビン	49回	43回	6回	3回
カラス	209回	78回	131回	0回

3. 音源を用いた位置特定

3.1 実験フィールド

実験フィールドは、高知工科大学地域連携棟屋上(約 16m)および北側駐車場である。(図 5)

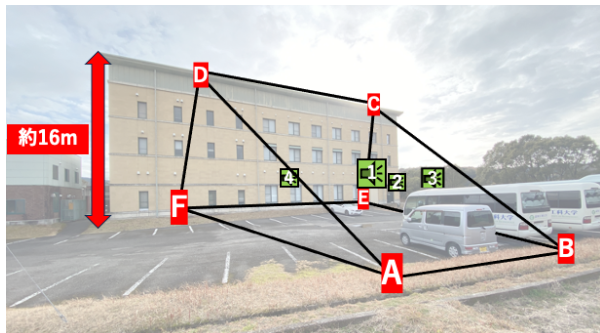


図 5 実験フィールド

3.2 使用機材と録音手法

図 6 のスピーカー (SONY SRS-XB10)、マイクロフォン (TASCAM DR-05) を用いて実験を行った。あらかじめトータルステーションを用いて絶対座標を求めた。図 5 の ABCDEF の位置にマイクロフォンを配置した。

サンプリングレートは実験で使用するマイクロフォンの最大値である 96kHz に設定をした。よって録音データは $1/96\text{k} \approx 1.0 \times 10^{-5}\text{s}$ ごとに記録されており、音速を 340m/s とした時、距離に換算すると約 3mm となる。

録音開始時に、図 7 のように 6 個のマイクロフォンをスピーカーから一定の距離に配置し、時刻同期のための音源を録音する。次に、6 個のマイクロフォンを結んで形成される立体の範囲内において、スピーカーを標尺にくくりつけ、図 5 の 1～4 の位置で音源を鳴らした。最後に、再度時刻同期のための音源を録音した。



図 6 使用機材



図 7 時刻同期のための
マイク配置

3.3 時刻同期

時刻同期の際は、図 8 のようにマイク A から音源のテンプレート波形 10ms を抽出し、それを他のマイクロフォンのデータの中で同じ音源の波形が

含まれる 1s の音源を設定し照合して最も相関が高くなる箇所を一致させる。その時、テンプレート波形は $1/96\text{k}$ ずつシフトさせながら一致箇所を照合した。

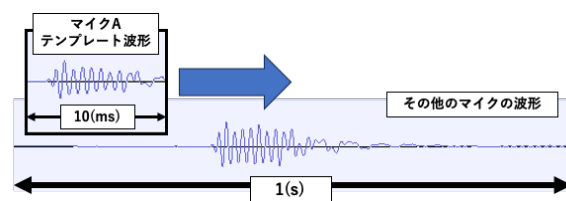


図 8 テンプレートマッチング

図 9 は、開始時の波形をマッチングした後の開始時と終了時の波形である。開始から約 9 分経過すると、終了時の波形が最大で約 10ms ずれていることが確認できた。これは、マイクロフォンの時計の個体差によるものと考えられ、そのままでは約 3.4m の誤差が生じる。そこで解析の前に各データに補正係数を適用した。



図 9 マイクロフォンの時刻ずれ

3.4 音源位置特定手法

図 10 はマイクロフォン 2 点の距離差のイメージ図である。実験で得た音声データを用い、マイクロフォン 2 点の音声の到達時間差を求め、それを距離差 (D_i) に変換する。一方、フィールド内の任意の三次元範囲に 1m 間隔でグリッド線を引き、ある格子点 j からマイクロフォン 2 点の距離差 (d_i) を求める。したがって、 $D_i - d_i$ が誤差をとる。マイクロフォン 2 点の組み合わせは 15 通りとなり、それぞれの誤差の 2 乗和を求めることにより $RMSE_j$ を求めた。求めた $RMSE_j$ の値が最も小さくなった格子点を音源の推定位置とした。

$$RMSE_j = \sqrt{\frac{1}{N} \sum_{i=1}^N (D_i - d_i)^2} \quad - (1)$$

N = マイク 2 点の組み合わせ数(15 通り)

D_i = 音源の到達時間の差から求めたマイクの距離差(m)

d_i = 対象座標から求めた距離差(m)

i = マイク組み合わせ j = 格子点番号

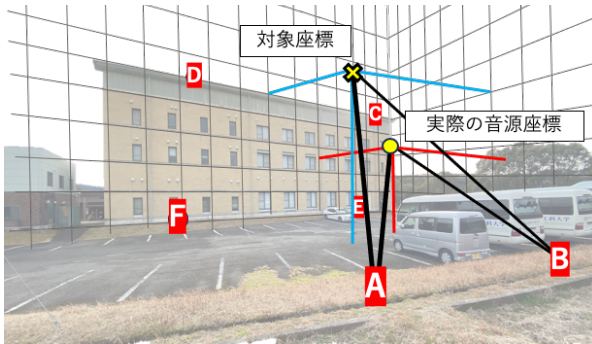


図 10 2 点の距離差計算

3.5 実験結果

図 1 1 は, グリッドの交点座標ごとに求めた $RMSE_j$ の値を 0~8m の範囲で 3D 積み上げヒートマップとして表示したものである. なお, 高さ方向の z 座標については, 一部の面を抜粋して表示している.

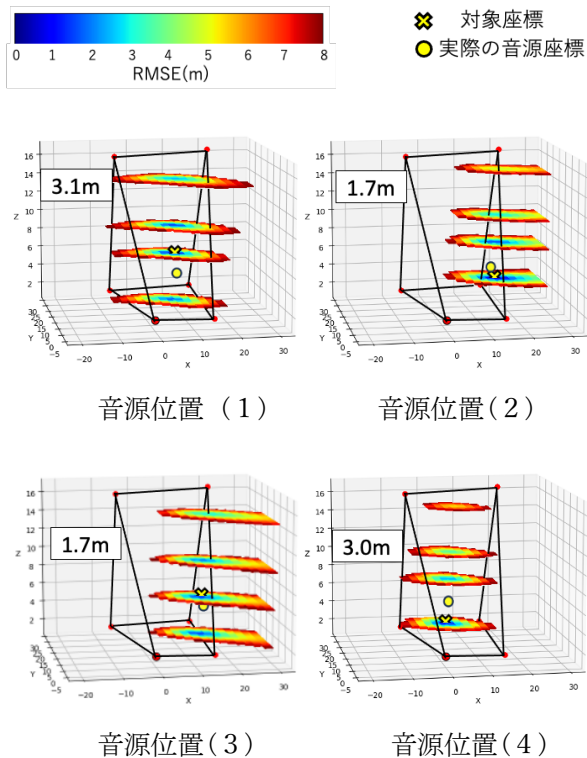


図 1 1 RMSE ヒートマップ

実際の音源座標と推定された座標の誤差は, 最大で 3.1m となった. 誤差が大きくなった要因として, マッチング精度の低さが挙げられる. 図 1 2 は, 音源位置(1)で鳴らした際のマイク A およびマイク B の波形を示している. 音源とマイクロフォンの位置関係が変化すると, 時刻同期時のように類似した波形が見られない. その結果, マッチング時の相関係数が低くなり, 誤差の発生につながったと考えられる.

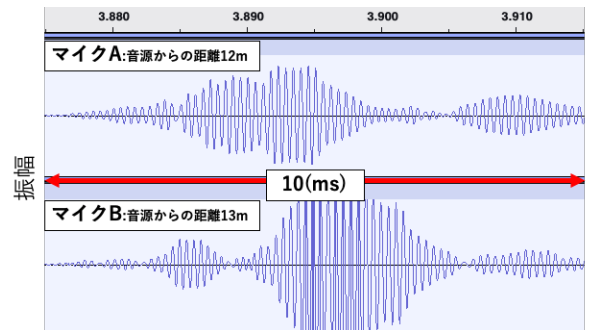


図 1 2 音源からの距離が異なる
マイク A と B の波形の違い

4. おわりに

本研究では音声データを周波数解析することで鳥類の鳴き声を判別する手法と, 複数のマイクロフォンを用いて音源到達の時刻差より位置を特定する手法の構築を行った.

鳥類の判別においては, より多くの鳴き声のデータを用いて周波数解析を行い, 周波数範囲や閾値を細かく調整しながら検討を進める必要がある.

位置特定に関しては, 誤差を低減するために, 鳥類の判別と同様に波形をパワースペクトルに変換してマッチングを行い, 精度を向上させる必要がある.

最終的に実際のフィールドで利用するには, これら 2 つの手法を組み合わせ, 長期的な鳥類観測を実現する手法の構築が必要である.

参考文献

- [1] モニタリングサイト 1000 里地調査 2005-2022 年度とりまとめ報告書, 2024
- [2] 佐藤 匠ら, 深層学習を用いた鳴き声による鳥類の種判別システムの開発と今後の展望, 2023
- [3] 岩見 聡ら, 猛禽類の鳴き声による位置推定システム: 音声レーダーの開発と検証, 2022