

渡り鳥の夜のフライトコール(NFC:Nocturnal Flight Call)

録音調査に向けた自動判別技術の有効性

Effectiveness of automatic detection for nocturnal flight calls (NFC) recording survey of migratory birds

大坂英樹¹ 加藤義清²

Hideki Osaka¹ and Yoshikiyo Kato²

¹トリルラボ*

¹ toriR Lab., Japan

² 認定NPO法人バードリサーチ 会員

² Member of Japan Bird Research Association, Japan

Abstract: To protect migratory birds, whose populations are significantly declining, it is necessary to conserve their habitats, including the ones along their migration routes. Because many species fly with nocturnal flight calls (NFC) and they can indicate their migration routes, we initiated a project of NFC recording survey in Japan. The challenge is to identify NFCs from a long recording. Our approach to addressing this challenge involves two steps: firstly, the detection of NFCs, and secondly, the precise identification of species. This study is concerned with the first step of detecting NFCs in field recordings. We investigated whether embeddings computed by neural models pre-trained on bird vocalization data are suitable for the purpose of detecting NFC candidates from field recordings. We compared two publicly available pre-trained models: Google Perch and BirdNET, both of which were trained on large scale bird vocalization datasets. Embeddings from pre-trained models were used as input features to a classifier, where we compared two classification methods: logistic regression and support vector machines. We prepared a dataset that consists of 335 call and 249 noise labels, which was used to train and evaluate classifiers. Among different configurations that were studied, the best performing one was the combination of BirdNET and SVM, which achieved PR-AUC of 0.925. Furthermore, it was found that the utilizing transfer learning model enables effective utilization of automatic detection even with small-size training data, resulting in expedited development of tools.

1. はじめに

個体数を大きく減らす渡り鳥の保護には飛翔ルートを含む全行動領域の保全が必要である[1]. ホオジロ属(*Emberiza*), ツグミ属(*Turdus*)を初め多くの種が夜にフライトコール(Nocturnal Flight Call)と呼ばれる声を発しながら飛ぶことが知られており[2], 北米ではレーダによる計測に加え, レーダではわからない種名をNFC録音で特定している[3, 4]. 欧州においてもNFC録音による渡りの研究は20年以上の歴史があり[5], オーストラリアでは90サイトで音響天文台と呼ぶ音響センサネットワークが設置さ

れている[6]. これらの知見をもとに生態解明の他, 海上・地上建造物との衝突や都市光害の影響などが研究されている[7-11]. 残念ながら日本ではヤイロチョウのNFCによる渡り調査の事例があるが[12], 全国的な調査はされていないようだ.

夜に小鳥の地鳴きが聞かれたことをきっかけに市民科学で全国でNFC録音調査を思い立ちプロジェクトを開始した[13]. 目的は夜の渡りの解明であり, (1)どこをどんな鳥がどんなタイミングで飛んでいるかの把握, (2)渡りのルート解明, (3)保全の貢献, である. 応用として生態解明の他, 風力発電所

* 連絡先: 〒915-0242 福井県越前市栗田部町34-1-9, E-mail:hideki.osaka@torir.net

など大型建造物のアセスメント、渡り予報によるブレード停止による衝突回避、不要照明の消灯などが考えられる。

NFC録音調査の課題は長大な録音データの中からNFCを特定することである。NFCは多種間で似ており[14]、人でも声からの種判別が難しく、NFCと種を紐づけるデータセットも日本では未整備である。そこで自動判別ステップをNFC切り出しと、その種同定の2段階に分ける方針とした。本報告では調査方法と判別アルゴリズムのNFC切り出しの有効性について述べる。

2. 方法

2.1 NFC録音調査方法と処理プロセス

調査は市民科学で進め、SNS呼びかけで13名から応募があった。調査地は北海道（江別、霧多布）から、新潟、石川、福井、埼玉、神奈川、大阪、福岡、熊本、沖縄（宮古、八重山）までと日本をカバーし、調査期間は2024年4月から3年で、春と秋の3ヶ月ずつ各地で録音される。データ処理のプロセスは、ネットワーク帯域が必ずしも高くなく、データ容量が年間最大1TBになること、処理分担の観点から、参加者によるNFCの自動切り出しとノイズ削除し、結果のみをサーバにアップロードする方法とした。2024年はNFCを収集し、2025年以降、種同定を試みる。

2.2 NFCの自動検出

従来のNFC録音調査は人手による録音と分析が必要で、時間と労力がかかっていた。そこで本研究では、録音調査で得られる長時間の音声データから効率的にNFCを検出するための手法の開発を目指した。精度の高い自動検出が可能になれば、NFCの特定作業に要する手間を大幅に削減でき、広域かつ大規模な録音調査が現実的なものになる。

近年は大量のデータで学習された深層学習モデルを、少数のデータで新たなタスクに適用させる転移学習が成功を取めている（例えば[14]）。事前学習モデルは事前学習の過程で訓練データに内在する重要な特徴を学習しており、そのことにより事前学習とは異なるタスクにも少数の訓練データで対応でき（few-shot learning）、時には訓練データなしでも対応が可能（zero-shot learning）である。音声処理の分

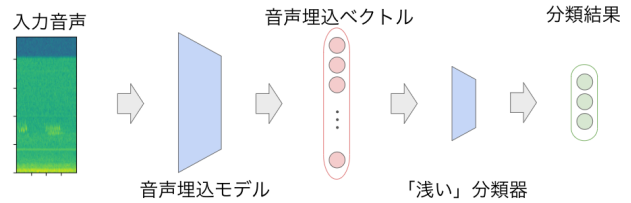


Fig.1. 事前学習モデルが出力する埋め込みを入力として利用する分類器。

野においても、鳥の鳴き声で学習された深層学習モデルで転移学習が有効であると報告されており[15]、プロジェクト開始から時間をかけずにツール実装できるモデルを検討候補とした。

2.3 埋め込みに基づくNFCの判別

本研究では事前学習モデルで埋め込みを計算し、それを入力とする分類器を構築した(Fig. 1)。

鳥の鳴き声で学習された事前学習モデルは2023年時点でいくつか公開されている。本研究では[15]で比較対象として取り上げられたもののうち、Google Perch [16] および BirdNET [17] を検討対象とした。Perchはパラメータ数7.8MのEfficientNet-B1 [18] のネットワークアーキテクチャを採用する。野生生物の音声共有サイトであるXeno-canto [19] で公開されている鳥の鳴き声のデータを訓練データとして、約10000種を対象とした種同定タスクで事前学習されている。入力はサンプリング周波数32kHzで5秒間のフレームを受け付け、1280次元の埋め込みベクトルを出力する。本研究ではGoogle Perch Version 4を使用した。

BirdNETはいくつかバージョンがあり、本研究ではV2.4を利用した。BirdNET V2.4ではEfficientNet-B0を基本とするアーキテクチャを採用しており、パラメータ数は5.8Mである。学習データにはXeno-cantoの他、Macaulay Library [20]から鳥の鳴き声の音声を利用する。それに加えて AudioSet [21]、Freefield1010 [22]、Warblr [23] などのデータセットから鳥の鳴き声を含まない環境音などを学習データとして利用している。サンプリング周波数48kHzで3秒間のフレームを入力として受け付け、1024次元の埋め込みベクトルを出力する。

NFCの判別は事前学習モデルから得られる埋め込みに対して分類器を適用して行う。本研究ではロジスティック回帰とSVMの各手法を検討した。各

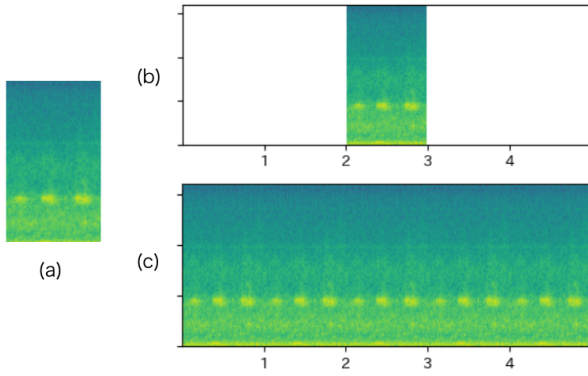


Fig.2. 2つのパディング処理方法：(a)1秒の入力データ，(b)入力データを5秒のフレームの2秒後に挿入し，残りをゼロパディング方法(offset 2)の例，(c)5秒のフレーム内で入力データを繰り返す方法。

手法ともscikit-learn [24] の実装を利用した。

3. 実験

3.1 データセット

データセットには，筆者らが収集し，ラベルを付与した音声データを用いた。音声データは国内の鹿児島県と福井県の合計7地点で録音され，収録時間は合計で約24時間となる。ラベルは，NFCを音響特徴に基づき分類した8つの正例ラベルと，カエルや昆虫など調査対象外種からの音声や環境音などのノイズに対して付与した負例ラベルを定義して利用した。ラベル付けには音声アノテーションのWEBサービスtoriR（トリルラボ）[25]を用いた。

本発表ではNFC検出を2値分類問題として扱い，これらのラベルを正例および負例の2値として利用している。ラベル数は正例で335件，負例で249件である。アノテーションデータに加えて，音声データをランダムサンプルして負例として追加した（追加データ）。追加データはNFCではないと保証されないが，録音時間に対してNFCの発生は疎であることから影響は少ないと判断した。追加データ数は正例の2倍の670件とした。元のデータと追加データを合わせたうえで，8:2で訓練データと評価データに分割し，実験に使用した。

3.2 データ処理

NFCの持続時間は典型的には50 msecから250 msecであり，それを踏まえてNFCを対象とした分類

モデルのNighthawk [26] では入力フレーム長を1 secとしている。本研究でも，Nighthawkにならってフレーム長を1 secとして扱う。本研究で利用したアノテーションにはNFCの開始時刻が含まれており，アノテーションに対応する音声データは開始時刻の前後500 msecから切り出した。

2.3節で述べたように，事前学習モデルはいずれも入力フレーム長が1 secより大きいのでパディング処理が必要となる。パディング処理として2つの方法を試した。1つ目は，入力データを事前学習モデルの入力フレームの特定の位置に挿入したうえで，入力データ以外の部分をゼロパディングする方法である（オフセット・ゼロパディング）(Fig. 2b)。2つ目は入力データを繰り返すことにより，入力フレームにゼロパディングが生じない（繰り返し）(Fig. 2c)。

4. 結果

4.1 結果

実験の結果をFig. 3, Fig. 4, およびTable 1に示す。全体ではBirdNETとSVMの組み合わせが最も高い性能を示し，PR-AUC（Area Under the Precision-Recall Curve）は0.925であった。Perchはロジスティック回帰との組み合わせで最も高い性能を示し，PR-AUCは0.795であった。パディング処理の影響を見ると，Perch, BirdNETでも共にオフセット・パディング法はロジスティック回帰との組み合わせで性能が高くなる傾向がある一方，繰り返し法はSVMとの組み合わせで性能が高くなることが分かった。

4.2 考察

BirdNETの結果がよかった要因の一つとして，学習データの特徴が挙げられる。BirdNETは鳥の鳴き声そのものだけでなく，環境音も学習データとして利用している。今回利用したデータセットは虫の鳴き声など鳥の鳴き声以外の音が混在しており，その中からNFCを判別するのに環境音も含めて学習しているBirdNETが有利だった可能性がある。

パディング処理は，判別性能に影響を与えることが分かった。今後，異なる分類手法の利用を検討する際には，分類手法ごとにパディング処理について

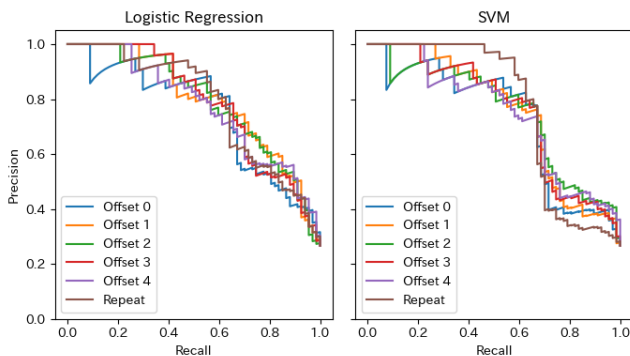


Fig. 3. Perchの埋め込みを利用した分類器のパディング処理ごととPR曲線.

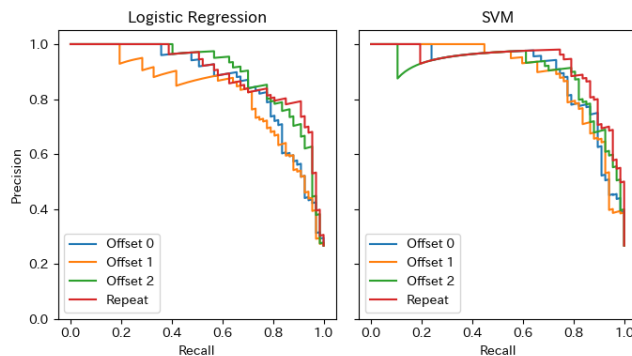


Fig. 4. BirdNETの埋め込みを利用した分類器のパディング処理ごとのPR曲線.

Table 1. 事前学習モデル, 分類手法, パディング処理の組み合わせごとのPR-AUC. 事前学習モデルと分類手法の組み合わせごとに最高値を太字で示した.

Embedding model	Classification method	Padding method					
		Offset 0 sec	Offset 1 sec	Offset 2 sec	Offset 3 sec	Offset 4 sec	Repeat
Perch	Logistic Regression	0.747	0.795	0.790	0.785	0.772	0.782
Perch	SVM	0.726	0.754	0.751	0.753	0.745	0.769
BirdNet	Logistic Regression	0.862	0.819	0.897	-	-	0.895
BirdNet	SVM	0.895	0.893	0.895	-	-	0.925

評価が必要となる. パディング処理についてはNFCの前後の音声そのまま利用したり, ガウスノイズでパディングしたり[26], 他の方法も考えられるが, その検討は今後の課題である.

また, 転移学習モデルを用いることで少ない学習データでも自動判別技術を有効に使えることがわかり, ツールの短期開発に目処がついた.

5. おわりに

日本で夜のフライトコール (NFC: Nocturnal Flight Call) 録音調査を行う市民科学プロジェクトを立ち上げた. 短期ツール実装を目指し, 課題の長時間録音データからNFCの自動判別を鳥の鳴き声で事前学習された深層学習モデルを活用した. 対象はモデルとしてGoogle PerchとBirdNETであり, 分類器としてロジスティック回帰とSVMを取り上げ, これらの組み合わせ評価から以下の結論を得た.

1. BirdNETとSVMを組み合わせた性能が高かった (PR-AUC=0.925).
2. パディング処理 (ゼロパディング, 繰り返し) で分類性能が変わることから, 使用する分類器により最適なパディング処理が異なる.

3. 事前学習モデルの利用で短期開発が可能.

今後, 本研究成果をベースに一般参加者が処理できるツールを開発し, 2024年はNFCデータを収集する. また2025年以降, NFCからの種同定を試みる.

謝辞

本研究の一部はNPO法人バードリサーチが主催する2023年調査研究支援の助成による. バードリサーチと支援者各位に感謝する. また, 録音データを提供頂いた中道暁美氏 (鹿児島市) に感謝する. 事前学習モデルの使い方について助言を頂いたTom Denton氏に感謝する.

参考文献

1. スコット ワイデンソール, 樋口広芳 (監修): 世界を翔ける翼, 化学同人, (2023).
2. Michael Lanzone, Emma Deleon, Lewis Grove and Andrew Farnsworth: Revealing Undocumented or Poorly Known Flight Calls of Warblers (Parulidae) Using a Novel Method of Recording Birds in Captivity, *The Auk*, Vol. 126, No. 3, pp. 511-519, (2009).
3. Farnsworth A. Flight Calls and Their Value for Future Ornithological Studies and Conservation Research. *The*

- Auk.; Vol. 122, No. 3, pp. 733–746, (2005).
4. Nadja Weisshaupt, Juha Saari and Jarmo Koistinen: Evaluating the potential of bioacoustics in avian migration research by citizen science and weather radar observations, *PLoS ONE* 19(3), (2024).
 5. Michael HEISS: Nocturnal bird migration at Besh Barmag bottleneck in Azerbaijan as revealed by means of acoustic monitoring, *Bird Conservation International*, Vol. 28, No.4, pp. 593-605, (2018).
 6. Paul Roe, Philip Eichinski, Richard A. Fuller, Paul G. McDonald, Lin Schwarzkopf, Michael Towsey, Anthony Truskinger, David Tucker and David M. Watson: The Australian Acoustic Observatory, *Ecology and Evolution*, Vol. 12, No. 10, pp. 1802-1808, (2021).
 7. Simon Gillings and Chris Scott: Nocturnal flight calling behaviour of thrushes in relation to artificial light at night, *IBIS*, Vol. 163, No. 4, pp.1379-1393, (2021).
 8. Adam D. Smith, Peter W. C. Paton and Scott R. McWilliams: Using Nocturnal Flight Calls to Assess the Fall Migration of Warblers and Sparrows along a Coastal Ecological Barrier, *PLoS One*, Vol. 9, No. 3, e92218, (2014).
 9. Richardson, W. John: Bird migration and wind turbines: migration timing, flight behavior, and collision risk, *Proceedings of the National Avian-wind Power Planning Meeting III*, San Diego, California. (1998).
 10. Chen, K., Kross, S. M., Parkins, K., Seewagen, C., Farnsworth, A., & Van Doren, B. M.: Heavy migration traffic and bad weather are a dangerous combination: Bird collisions in New York City. *Journal of Applied Ecology*, Vol. 00, pp.1-13, (2024).
 11. Bianca Michalik, Vera Brust and Ommo Hüppop: Are movements of daytime and nighttime passerine migrants as different as day and night?, *Ecology and Evolution*, Vol.10, No. 20, pp.10913-11846, (2020).
 12. 植松永至, 佐野清貴, 峯光一, 鳥飼久裕, 小倉豪, 溝口文男, 森下功啓, 中村豊, 中原聡, 中原亨, 岡部海都, 山口典之, 大槻恒介, 井上拓海, 上田浩一, 貞光隆志, 西剛, 古田慎一, 楠木憲一, 岩本忠義, 平田智法, 中村滝男, 森茂晃, 星野由美子, 黒田治男, 齊木孝, 牛込祐司, 松宮裕秋, 今森達也, 増川勝二: フライトコールによるヤイロチョウの渡り調査 2～5年間の渡来時期の傾向～, 日本鳥学会大会プログラム, (2023).
 13. 大坂英樹, 参加者募集 ～みんなで夜のフライトコール録音(NFC録音)調査～, <https://tinyurl.com/ylqjot8z>
 14. Snell, J., Swersky, K. and Zemel, R. S.: Prototypical networks for few-shot learning. In *Proc. of NIPS'17*, pp. 4080-4090, (2017).
 15. Ghani, B., Denton, T., Kahl, S. and Klinck, H.: Global birdsong embeddings enable superior transfer learning for bioacoustic classification. *Scientific Reports*, Vol. 13, 22876 (2023).
 16. Google Perch, <https://www.kaggle.com/models/google/bird-vocalization-classifier/frameworks/tensorFlow2/variants/bird-vocalization-classifier>
 17. Kahl, S., Wood, C. M., Eibl, M. And Klinck, H.: BirdNET: A deep learning solution for avian diversity monitoring. *Ecological Informatics*, Vol. 61, 101236, (2021).
 18. Tan, M. And Le, Q.: EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *PMLR*, Vol. 97, pp. 6105-6114, (2019).
 19. Xeno-canto, <https://xeno-canto.org/>
 20. Macaulay Library, <https://www.macaulaylibrary.org/>
 21. J. F. Gemmeke *et al.*: Audio Set: An ontology and human-labeled dataset for audio events, 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), p. 776-780, (2017).
 22. Stowell, D., Plumbley, M.D.: An open dataset for research on audio field recording archives: freefield1010, *arXiv Preprint*, arXiv:1309.5275, (2013).
 23. Mesaros, A. *et al.*: Detection and classification of acoustic scenes and events: outcome of the DCASE 2016 challenge, *IEEE/ACM Trans. Audio Speech Lang. Process*, Vol 26, No. 2, pp. 379-393, (2018).
 24. Pedregosa, F. *et al.*: Scikit-learn: Machine Learning in Python, *Journal of Machine Learning Research*, Vol 12, pp. 2825-2830, (2011).
 25. トリルラボ, <https://www.torir.net>
 26. Van Doren, B.M. *et al.*: Nighthawk: Acoustic monitoring of nocturnal bird migration in the Americas, *Methods in Ecology and Evolution*, Vol. 15, No. 2, pp. 329-344, (2024).