

Université de Montréal

Évaluer le potentiel et les défis de la variation intraspécifique pour les réseaux neuronaux
profonds de reconnaissance de chants d'oiseaux
*L'exemple des bruants des prés (*Passerculus sandwichensis*) de l'île Kent, Nouveau-Brunswick*

Par

Camille Rondeau Saint-Jean

Département de Sciences biologiques, Faculté des Arts et des Sciences

Mémoire présenté en vue de l'obtention du grade de

Maître ès Sciences (M.Sc.) en Sciences biologiques

Août 2023

© Camille Rondeau Saint-Jean, 2023

Université de Montréal

Département de Sciences biologiques. Faculté des Arts et des Sciences

Ce mémoire intitulé

**Évaluer le potentiel et les défis de la variation intraspécifique pour les réseaux neuronaux
profonds de reconnaissance de chants d'oiseaux**

L'exemple des bruants des prés (*Passerculus sandwichensis*) de l'île Kent, Nouveau-Brunswick

Présenté par

Camille Rondeau Saint-Jean

A été évalué par un jury composé des personnes suivantes

Frédérique Dubois
Président-rapporteur

Timothée Poisot
Directeur de recherche

Daniel J. Mennill
Codirecteur

Mikaël Jaffré
Membre du jury

Résumé

Les réseaux neuronaux profonds sont des outils prometteurs pour l'évaluation de la biodiversité aviaire, en particulier pour la détection des chants et la classification acoustique des espèces. Toutefois, on connaît mal l'étendue de leur capacité de généralisation face à la variation intraspécifique présente dans les chants d'oiseaux, ce qui pourrait mener à des biais.

Notre étude porte sur l'évaluation des performances de BirdNET, un réseau neuronal profond, pour le traitement d'un corpus d'enregistrements audio caractérisés par une variation intraspécifique significative, en utilisant l'exemple du chant du bruant des prés (*Passerculus sandwichensis*). Dans la population de l'île de Kent, au Nouveau-Brunswick, les individus sont suivis et enregistrés grâce à leurs bagues de couleur et la présence de microdialectes est solidement documentée. Nous avons recueilli et annoté 69 606 chants provenant de 52 individus et analysé ces données à l'aide d'une version récente de BirdNET.

Nos résultats révèlent que BirdNET démontre une précision globale suffisante, prédisant correctement 81,9 % des chants, ce qui dépasse les résultats rapportés par ses développeurs. Toutefois, nous avons observé une variation considérable dans les scores de confiance et les taux de prédiction exactes entre les individus, ce qui suggère des biais potentiels. Cependant, nos recherches n'ont pas mis en évidence de variation entre les résultats des différents microdialectes, ce qui souligne la relative robustesse de l'algorithme. Nous avançons que la variation observée entre les individus est due au fait que certains d'entre eux chantent systématiquement plus près des microphones, résultant en des chants plus clairs donc plus faciles à identifier.

Pour mieux comprendre le processus de prise de décision de BirdNET, nous avons tenté de produire des cartes d'activation de classe, qui constituent un outil précieux pour identifier les éléments d'un chant qui déterminent une prédiction. Cependant, il ne nous a pas été possible d'obtenir des cartes d'activation de classe d'après la version actuellement disponible du code de BirdNET sans avoir recours à des connaissances avancées en informatique. L'accès à des outils explicatifs adaptés aux innovations récentes dans les architectures de réseaux neuronaux

profonds serait crucial pour mieux interpréter les résultats et renforcer la confiance des utilisateurs.

Nos résultats soulignent la nécessité de poursuivre les recherches sur la capacité de généralisation des réseaux neuronaux profonds pour la bioacoustique en utilisant des ensembles de données monospécifiques portant sur de plus longues périodes ou des aires de répartition géographique plus vastes. En outre, l'extension de cette étude à des espèces ayant des répertoires plus importants ou des différences plus subtiles entre le chant des individus pourrait nous informer davantage sur les limites et le potentiel des algorithmes d'apprentissage profond pour la détection et la classification acoustiques des espèces.

En conclusion, notre étude démontre les performances prometteuses de BirdNET pour le traitement d'un large corpus de chants de bruants des prés, et confirme son potentiel en tant qu'outil précieux pour l'évaluation de la biodiversité aviaire. Les biais dus aux techniques d'enregistrement et la variation dans les taux de succès observés entre les individus méritent d'être étudiés plus en détail.

Mots-clés : Ornithologie, bruant des prés, chant d'oiseaux, culture animale, individualité animale, bioacoustique, algorithme d'apprentissage machine, apprentissage profond, réseau neuronal profond

Abstract

Machine learning, particularly deep neural networks, has gained prominence as a valuable tool in ecological studies and wildlife conservation planning. In the field of avian biodiversity assessment, deep neural networks have shown remarkable promise, particularly in acoustic species detection and classification. Despite their success, a critical knowledge gap exists concerning the generalization ability of these algorithms across intraspecific variation in bird song. This raises concerns about potential biases and misinterpretation of results.

This study focuses on evaluating the performance of BirdNET, a deep neural network, in processing audio recordings characterized by significant intraspecific variation in the Savannah Sparrow (*Passerculus sandwichensis*) song. Savannah Sparrows are an ideal candidate for this investigation, given their well-studied population on Kent Island, New Brunswick, Canada. Each male sings a unique, unchanging song throughout its life, and the population exhibits well-documented geographical microdialects.

We collected a large corpus of Savannah Sparrow songs using autonomous and focal recorders on Kent Island, yielding a total of 69,606 manually annotated songs from 52 different sparrows. We analyzed the audio data using BirdNET-Analyzer. The resulting confidence scores were used to assess the algorithm's performance across microdialects and individual birds.

Our results revealed that BirdNET exhibited considerable overall accuracy, correctly predicting 81.9% of the songs, which surpassed the results reported by the developers of BirdNET. We observed variations in BirdNET's confidence scores among individual birds, suggesting potential biases in its classifications. However, our investigation indicated no evidence of distinct biases towards specific microdialects, highlighting the algorithm's relative robustness across these groups. We suspect that the variation observed amongst individuals is caused by the fact that some were singing consistently closer to microphones, yielding clearer songs.

To gain insights into BirdNET's decision-making process, we sought to employ class activation maps, a valuable tool for identifying essential song elements contributing to species predictions. However, we were unable to produce class activation maps from the current version of BirdNET

without advanced computer science skills. Access to informative tools adapted to recent innovations in deep neural network architectures for bioacoustic applications is crucial for understanding and interpreting results better. Such tools would enhance user confidence and favour accountability for conservation decisions based on these predictions.

Our findings underscore the need for further research investigating the generalization capacity of deep neural networks in bioacoustics on single-species datasets with more extensive intraspecific variation and broader geographical ranges. Additionally, expanding this investigation to species with larger song repertoires or more subtle inter-individual song differences could provide valuable insights into the limits and potential of deep learning algorithms for acoustic species detection and classification.

In conclusion, our study demonstrates BirdNET's promising performance in processing a large corpus of Savannah Sparrow songs, highlighting its potential as a valuable tool for avian biodiversity assessment. Biases and variations in confidence scores observed across individual birds warrant further investigation.

Keywords: Ornithology, Savannah Sparrow, bird song, animal culture, animal individuality, bioacoustics, machine learning algorithm, deep learning, deep neural network

Table des matières

Résumé.....	3
Abstract	5
Table des matières	7
Liste des tableaux.....	9
Liste des figures.....	10
Remerciements	13
Chapitre 1 – Introduction générale.....	14
Techniques d’étude des chants d’oiseaux	15
Suivi acoustique passif	15
Reconnaissance automatisée.....	16
Les bruants des prés de l’île Kent.....	21
L’île Kent.....	21
Les bruants des prés.....	24
Objectifs de l’étude	29
Chapitre 2 – Evaluating the generalization ability of a deep neural network across intraspecific variation in Savannah Sparrow song.....	30
Contribution des co-auteurs	30
Abstract	32
Introduction.....	33
Methods	35
Results	37
Discussion.....	39

Chapitre 3 – Conclusion générale43

 Évaluation de la capacité de généralisation.....43

 Interprétabilité et transparence45

 Entraînement à la reconnaissance d’individus et de groupes culturels48

Références bibliographiques.....51

Annexes67

 Annexe 1.....67

 Annexe 2.....68

Liste des tableaux

Tableau 1	Sample sizes, proportion of correct predictions by BirdNET-Analyzer (Kahl et al. 2021) and average confidence score for 6 microdialects found in 2022 in Savannah Sparrows on Kent Island, New Brunswick. No analyses were performed on the songs of birds belonging to no microdialect.	42
Tableau 2	Number of annotated recordings for individual Savannah Sparrows from Kent Island, New Brunswick, with microdialect and recording type.....	67

Liste des figures

- Figure 1** Spectrogramme d'un chant de bruant des prés. Les quatre sections principales du chant sont indiquées suivant Williams et al. (2013).17
- Figure 2** Carte du complexe des Trois Îles, au Nouveau-Brunswick. Le système d'étude (ombré en jaune) et la grille sont représentés sur l'île Kent.22
- Figure 3** Bruant des prés (*Passerculus sandwichensis*) mâle nichant sur l'île Kent, au Nouveau-Brunswick. Les bagues de couleurs constituent la combinaison B.RW : noir (Black) et bague numérotée en aluminium à droite, rouge (Red) et blanc (White) à gauche. Photo: Daniel J. Mennill25
- Figure 4** Exemples de chacun des microdialectes observés chez les bruants des prés de l'île Kent, Nouveau-Brunswick, depuis 2016, et de chants n'appartenant à aucun groupe. Les cadres rouges indiquent la section centrale caractéristique du microdialecte.28
- Figure 5** Distribution of confidence scores attributed to Savannah Sparrow songs from the Kent Island population by BirdNET and rate of correct predictions, by microdialect and individual. Individuals are identified by letters representing the colour-bands on their legs.39
- Figure 6** Cartes d'activation de classe pour des chants de bruants à couronne blanche (*Zonotrichia leucophrys*), avec un score de confiance c . La zone en rouge met en évidence l'élément du chant qui est déterminante pour la classification de cette espèce. Extrait de Kahl (2020).45
- Figure 7** Distribution of confidence scores attributed to Savannah Sparrow songs from the Kent Island population by BirdNET, by microdialect and individual, according to the recording method used to collect songs.68
- Figure 7 shows that Song Meter recordings, which are fixed in the field following a 50 x 50 meter grid, often yield low confidence scores for birds singing far from the microphones (ie. L.GL, G.WG, NW.L and Y.YW) and consistently high scores for those often observed near or perching on recorder (ie. O.NL and RY.B). Focal recordings, which are performed on foot at varying distances from singing birds, do not reflect this pattern. There seems to be no link between the scores obtained through both recording methods within the songs of an individual. This leads us to think

that the results obtained when comparing average scores at the individual level were due to the bias in Song Meter recording distances, since there were not enough focal recordings for each bird to counteract this effect, and not to individual song characteristics.....68

À Patrice Bourgault, parce qu'un grand pédagogue peut faire toute la différence

Remerciements

Tim, Dan, I cannot thank you enough for your support. I could not have dreamed of a better team, and often wonder at the serendipity of us three coming together. No two ecologists and no two mentors could be more different in their approaches, but you both are exceptional examples of what a brilliant, passionate and, above all, kind scientist should be like. Thanks to your help, guidance and understanding, my scientific journey of the past two years was filled with joy. I hope to carry the spark of enthusiasm, generosity, and excellence I gained by working with both of you through the rest of my studies and the whole of my future career.

All my gratitude goes to Sarah and Mikayla for your passion and patience when showing me the ways of the Savannah Sparrows of Kent Island and for your hard work collecting many, many, *many* hours of focal recordings. I will never forget the month we spent together in the very middle of nowhere, nor your courage in dealing with my insular culinary experimentations.

Many thanks to all my colleagues from the Poisot and Mennill labs. I have learned a lot from our weekly discussions I am deeply grateful that me working online was never a barrier to feeling included and helping each other out. I wish you all the best in your future academic and professional pursuits and truly hope our paths will cross again!

Merci à mes ami.e.s pour leur soutien. Salix, tes questions et ton feedback m'ont donné le coup de pouce qu'il fallait pour sortir de plus d'un blocage intellectuel, et je ne sais pas ce que j'aurais fait sans ta précieuse solidarité dans les moments difficiles. Beu, j'ai du mal à croire qu'après toutes ces années, tu es encore pris à m'écouter parler d'un mémoire, et qu'en plus on remet ça bientôt avec une thèse. Jesse, absolutely none of this would have happened in the first place without your encouragements when I needed them the most.

N'oublions pas le trio gagnant qui a rendu les dernières semaines de rédaction bien plus agréables: Kaamelott, MIKA, et Pizza chez Zac.

Finally, no level of eloquence could adequately express my gratitude to Luke for his love and support through every single day of this journey. You are not only a brilliant scientist, an inspiration, and a great buddy with whom to explore the wonders and the weirdnesses of academia; I could never have gone through the various storms, lava pits and dragons that life threw at me in the past two years without the wings your presence give me. Onwards to new adventures, and bring on the beach'n'pizza weekends!

Chapitre 1 – Introduction générale

Reconnaître l'espèce d'un oiseau en entendant son chant est l'un des plus grands plaisirs de l'ornithologue professionnel ou amateur. En effet, avec un peu de pratique, on peut apprendre à distinguer les caractéristiques essentielles du chant de la paruline flamboyante, du bruant à gorge blanche, du troglodyte familier ou du moqueur chat, au-delà des particularités du chant de chaque individu. Mais d'où viennent ces variations, que l'on choisit souvent d'ignorer?

Les oiseaux chanteurs, également désignés sous le nom d'oscines ou passereaux, se distinguent par leur capacité à apprendre leur chant en écoutant et en imitant leurs congénères (Catchpole et Slater, 2008). Chez plusieurs espèces, les oisillons écoutent et mémorisent les chants des mâles adultes qui les entourent, et les jeunes mâles sélectionnent, pratiquent et raffinent ces chants lors de leur première saison de reproduction (Marler et Nelson, 1994). Ainsi, par un processus de transmission et d'imitation imparfaites, via des pressions environnementales (Brumm et Naguib, 2009) et à la suite de la sélection directionnelle ou non de certaines caractéristiques du chant (Aplin 2019, Williams 2021), un large éventail de variantes apparaît et persiste dans le chant des oiseaux d'une même espèce.

On dit donc que ces différences sont issues de phénomènes culturels, c'est-à-dire qu'ils découlent de procédés d'imitation et d'apprentissage au sein d'un groupe social (Bluff et al. 2010, Aplin 2019, Whiten 2019, Logue and Leca 2020, Williams 2021). Toutefois, des facteurs génétiques et environnementaux peuvent aussi contribuer à former les chants (Baker and Cunningham 1985, Derryberry 2009, Slade et al. 2017, Sebastianelli et al. 2022). Chez plusieurs espèces, certains de ces facteurs, en plus de patrons de dispersion et d'isolation, contribuent à créer des dialectes géographiquement dépendants (Borror and Gunn 1965, Otter et al. 2020, Rodríguez-Fuentes et al. 2021, Czocherová et al. 2022, Pandolfino and Douglas 2022). Ces dialectes sont modelés par différents mécanismes évolutifs, et peuvent se transformer avec le temps (Byers et al. 2010, Janes and Ryker 2013, Williams 2021, Jäckel et al. 2022). L'importance de la diversité culturelle pour la conservation est de plus en plus reconnue (McGregor et al. 1997, Terry et al. 2005, Ryan 2006, Keith and Bull 2017, Gruber et al. 2019, Crates et al. 2021), car la variation intraspécifique

transmise socialement peut jouer un rôle dans la spéciation (Slabbekoorn and Smith 2002) et participe à la résilience des populations aux perturbations (Crates et al. 2021, Brakes et al. 2021).

Bien que les différences entre le chant des individus d'une même espèce soient souvent trop subtiles ou trop rapides pour être bien saisies par l'oreille humaine, plusieurs études ont montré que les oiseaux peuvent reconnaître leurs conspécifiques de façon fiable (Marler 1960, Lind et al. 1996, Jouventin 1999, Wilson and Mennill 2010). De plus, des expériences ayant diffusé des enregistrements dans le territoire d'oiseaux sauvages ont montré que certains réagissent plus fortement aux chants de congénères issus de la même population qu'à ceux des étrangers, ce qui signifie qu'ils perçoivent probablement les différences entre groupes culturels (Nelson 1998, Nelson and Soha 2004, Williams et al. 2019). Il est toutefois aussi possible qu'ils réagissent simplement davantage aux chants qu'ils ont déjà entendus.

Le présent mémoire est consacré à l'exploration de l'impact que peut avoir la variation intraspécifique dans le chant des oiseaux sur les outils automatisés de reconnaissance acoustique basés sur les réseaux neuronaux. Cette introduction générale vise à brosser un portrait des différentes technologies utilisées pour le suivi acoustique et à présenter le système d'étude sur lequel reposent nos travaux.

Techniques d'étude des chants d'oiseaux

Suivi acoustique passif

Avec la préoccupation croissante des chercheurs et du public pour l'impact des changements climatiques et de la perte d'habitat sur la biodiversité (Parmesan 2006, Beyer and Manica 2020), l'intérêt est grand pour des outils qui permettent de mieux comprendre les réponses comportementales, phénologiques et démographiques des espèces aux changements rapides dans leur environnement. Les études acoustiques à grande échelle permettent le suivi à long terme des oiseaux sauvages, informant notre compréhension de l'écologie, de l'évolution, du comportement et de la conservation de ces espèces sur de vastes territoires (Krause and Farina 2016, Alcocer et al. 2022). En effet, la détection du chant de certaines espèces d'oiseaux est

souvent utilisée comme proxy pour faire le suivi de la biodiversité dans un habitat donné. Par exemple, dans le cadre du Suivi de la biodiversité du Québec, l'hétérogénéité acoustique est un bioindicateur utilisé pour inventorier la biodiversité associée aux milieux forestiers, humides et de toundra (Biodiversité Québec, 2023).

Les études acoustiques dépendent traditionnellement d'experts faisant des points d'écoute sur le terrain, ce qui est coûteux, long, et propice aux biais d'observateur (Emlen and Dejong 1992, Rosenstock et al. 2002, Hutto and Stutzman 2009). Le développement d'appareils d'enregistrement préprogrammés, ainsi que l'augmentation de la capacité de stockage de données au cours des dernières décennies, ont facilité la collecte de très grandes quantités de données audio (Zhang et al. 2013, Gibb et al. 2019, Sugai et al. 2019). Cette approche, connue sous le nom de suivi acoustique passif (en anglais *passive acoustic monitoring*, ou PAM), constitue une alternative économique ou un complément aux points d'écoute réalisés en personne (Digby et al 2013, Pankratz et al. 2017, Sugai et al. 2019).

Le suivi acoustique passif permet de collecter des données dans pratiquement n'importe quel habitat et laisse une trace permanente plutôt qu'une interprétation, ce qui est très précieux pour la confirmation des observations et pour l'utilisation des données pour des études à long terme (Blumstein et al. 2011, Shonfield and Bayne 2017). Toutefois, la difficulté des projets basés sur le suivi acoustique passif est qu'ils sont limités par le temps et l'expertise nécessaires pour isoler et classifier les sons d'intérêt au sein des quantités énormes de données audio récoltées (Priyadarshani et al. 2018, Gibb et al. 2019, Sugai et al. 2019).

Reconnaissance automatisée

Pour pallier ce problème, les chercheurs se sont tournés vers des outils d'apprentissage automatique qui accélèrent le traitement des enregistrements, dans plusieurs cas égalant ou surpassant en rapidité et en précision le travail d'annotateurs experts (Digby et al. 2013, Venier et al. 2017, Darras et al. 2019). Stowell (2022) a identifié 65 études utilisant l'apprentissage profond pour étudier les chants d'oiseaux, dans la plupart des cas pour détecter la présence de différentes espèces dans des enregistrements longs. De plus, les applications pour smartphone basées sur des algorithmes d'apprentissage automatique sont un outil pédagogique stimulant et

permettent la récolte de données écologiques par des membres du public sans qu'un haut niveau de connaissances ou des compétences spécialisées ne soient nécessaires (Jäckel et al. 2021, Wood et al. 2022, Clark et al. 2023).

Les réseaux neuronaux profonds

Mieux connus sous leur nom anglais de *deep neural networks*, les réseaux neuronaux profonds sont des algorithmes conçus pour traiter des données complexes et effectuer des tâches d'apprentissage automatique (*deep learning*). Un réseau neuronal est constitué de plusieurs couches d'unités interconnectées, appelées « neurones ». On distingue les réseaux neuronaux profonds par le fait qu'ils comportent un grand nombre de couches et présentent une architecture complexe, ce qui les rend très puissants mais demande aussi beaucoup de puissance de calcul. Pour une tâche de classification, l'apprentissage se fait en comparant les étiquettes associées à un jeu de données d'entraînement avec les prédictions, d'abord aléatoires, que le réseau fournit. Au fil des itérations, le modèle réajuste le poids de chaque neurone afin de minimiser la distance entre ses prédictions et les valeurs réelles associées aux données. Ainsi, s'il est exposé à un jeu de données représentatif et exhaustif, il apprend à se représenter les motifs et les relations complexes qui existent dans les données.

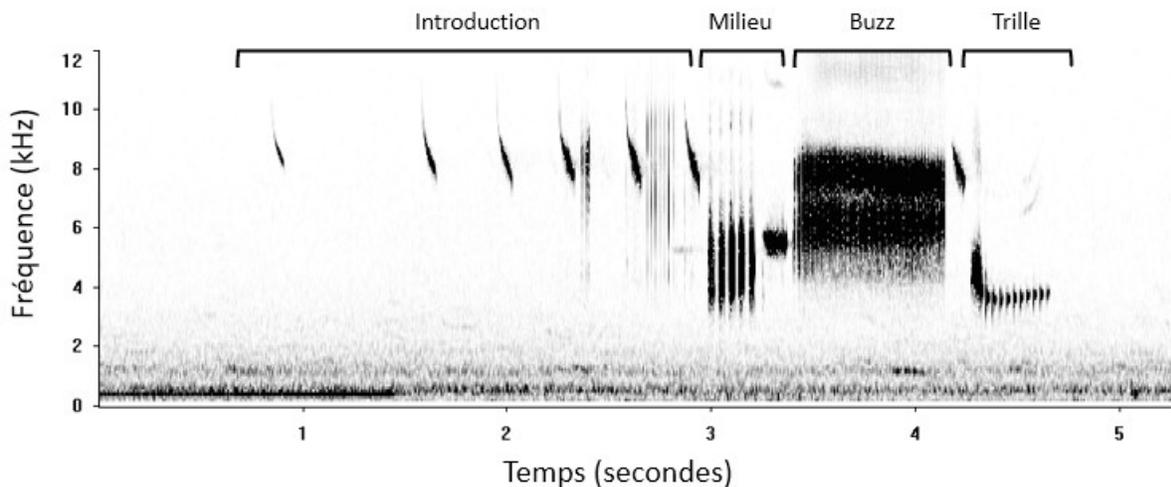


Figure 1 Spectrogramme d'un chant de bruant des prés. Les quatre sections principales du chant sont indiquées suivant Williams et al. (2013).

De nombreux modèles basés sur ce principe ont été développés pour reconnaître des groupes de pixels et détecter des patrons dans des images, réalisant ce qui est communément appelé « vision par ordinateur » (*computer vision*). Parmi ceux-ci, notons par exemple AlexNET (Krizhevsky et al. 2012), VGG (Simonyan and Zisserman 2014) et ResNet (He et al, 2016). Comme en bioacoustique, il est habituel de visualiser les sons à l'aide de spectrogrammes (Figure 1), il est possible de réutiliser l'architecture de ces modèles bien connus afin de classifier les sons. Stowell (2022) discute la possibilité de développer des modèles prenant en entrée la forme d'onde brute, ce qui implique certaines modifications architecturales, ou combinant cette méthode avec l'utilisation de spectrogrammes plus traditionnels.

Le principal défi lié à l'entraînement des réseaux neuronaux profonds est qu'ils requièrent des jeux de données d'entraînement souvent gigantesques, pouvant aller jusqu'à plusieurs milliers ou dizaines de milliers d'exemples par classe. Par ailleurs, afin que l'algorithme ne favorise pas une classe au détriment d'une autre et ne vienne fausser ses prédictions, il est nécessaire que les classes soient équilibrées, c'est-à-dire qu'elles soient représentées par un nombre équivalent d'échantillons. Il va sans dire que cela représente une difficulté considérable pour la récolte et l'étiquetage de données. L'écueil à éviter est le surentraînement, ce qui se produit quand le modèle « mémorise » de trop près les données qui lui sont présentées, et ne parvient pas à s'adapter lorsqu'on lui présente de nouveaux exemples. On dit alors qu'il n'arrive pas à « généraliser ». Contrairement à un humain, qui comprend la réalité matérielle de ce qu'est un chant d'oiseau et peut le reconnaître lorsqu'il l'entend dans un autre contexte acoustique ou à une plus grande distance, un algorithme de vision par ordinateur se fie sur des combinaisons de pixels, et doit être exposé à un très grand nombre de situations possibles associées à une classe avant d'arriver à généraliser sa capacité de reconnaissance.

Ainsi, l'enjeu majeur du développement d'un outil de reconnaissance de chants d'oiseaux basé sur un réseau neuronal profond est de trouver l'équilibre entre la capacité de généralisation à travers la variation présente au sein d'une classe, et la capacité de discernement entre les classes différentes. Ce problème peut être résolu en ajustant les paramètres de l'architecture du réseau, mais aussi en augmentant et en peaufinant le jeu de données utilisé pour son entraînement, car sa composition peut avoir un impact considérable sur la performance du modèle. Parmi les outils

automatiques de reconnaissance de chants d'oiseaux, BirdNET, un algorithme bien connu sur lequel est basée une populaire application pour téléphone intelligent, attire beaucoup d'intérêt tant chez le public que les chercheurs pour ses performances étonnantes.

BirdNET

À l'origine, BirdNET est un algorithme d'apprentissage profond capable d'identifier le chant de 984 espèces d'oiseaux d'Europe et d'Amérique du Nord (Kahl 2020, Kahl et al. 2021). Il a été entraîné avec 1.5 millions de clips audio extraits entre autres des collections de chants d'oiseaux Xeno-canto (Xeno-canto, 2023) et Macaulay (Macaulay, 2023). BirdNET-Analyzer, une version plus récente de BirdNET qui peut reconnaître jusqu'à 3300 espèces d'oiseaux et de batraciens à travers le monde, est aussi disponible pour les chercheurs sous la forme d'un script python utilisant la librairie d'apprentissage machine Tensorflow (Kahl et al. 2021). Cet algorithme constitue la base de projets communautaires comme HaikuBox (HaikuBox, 2023) et BirdWeather (BirdWeather, 2023).

BirdNET est utilisé dans divers projets étudiant par exemple la phénologie (Malamut 2022), la conservation (Bolcar 2020), ou les déplacements (Verreycken et al. 2021) des oiseaux. Plusieurs chercheurs le citent aussi comme étant une référence à laquelle ils comparent leur propres méthodes et algorithmes, sans nécessairement en faire usage (Folliot et al. 2022, Clark et al. 2023, Brunk et al. 2023). Répondant à l'appel de Digby et al. (2013) à faire davantage de comparaisons entre les modèles de classification automatisée et l'annotation manuelle, Pérez-Granados (2023a) compile plusieurs études utilisant BirdNET et évaluant sa performance. Il rapporte entre 72 et 85% de détections correctes avec cet algorithme. Parmi les études citées, Arif et al. (2020) ont testé BirdNET sur 225 enregistrements de 23 espèces, et relèvent une précision de 91.5%. Toenies et Rich (2021) rapportent aussi des résultats impressionnants, avec environ 4% de sons mal identifiés. Cole et al. (2022) ont trouvé que BirdNET détectait entre 65% et 90% des sons annotés par des experts, selon les seuils de confiance appliqués.

Reconnaissance automatisée et variation intraspécifique

Il reste à voir si des technologies de reconnaissance automatique innovantes comme BirdNET peuvent fournir des prédictions fiables quant aux espèces présentes dans un enregistrement,

malgré l'hétérogénéité des données cibles. En effet, nous ignorons largement à quel point la variation intraspécifique affecte la performance des algorithmes d'apprentissage profond destinés à la reconnaissance acoustique d'espèces. Alors que Wood et al. (2022) ont utilisé des enregistrements réalisés par le public grâce à l'application BirdNET pour étudier la variation intraspécifique chez les bruants à gorge blanche et chez les bruants jaunes, les dialectes d'intérêt ont été annotés manuellement et leur influence sur la qualité des prédictions n'a pas été évaluée. Ghani et al. (2023) ont exploré l'utilisation de méthodes de classification automatique entre des populations de troglodyte familial, mais n'ont pas non plus testé l'impact de ces groupes sur l'identification au niveau de l'espèce. Par ailleurs, Pérez-Granados (2023a) recommande davantage de travaux sur l'impact de la variation géographique sur la performance de BirdNET. De plus, Kahl (2020), le principal développeur de BirdNET, mentionne le besoin de considérer l'influence de la diversité intraspécifique et de la taille du répertoire des individus sur la quantité de données nécessaires pour l'entraînement et sur la performance de l'algorithme.

En général, les chercheurs devraient viser à entraîner des modèles qui maintiennent un équilibre entre une bonne capacité de généraliser à travers tous les types de variation au sein d'une espèce, et le maintien d'un bon discernement entre les espèces. Ainsi, s'assurer que le jeu de données d'entraînement couvre une assez grande part de la variation présente au sein d'une classe est crucial pour assurer une bonne performance malgré les différences génétiques, culturelles, géographiques et temporelles qui peuvent exister au sein des données ciblées. Pour étudier cette question, nous avons constitué un jeu de données acoustiques en récoltant des enregistrements dans la population de bruants des prés de l'île Kent, au Nouveau-Brunswick, où la variation intraspécifique est bien marquée et étudiée depuis plusieurs années.

Les bruants des prés de l'île Kent

L'île Kent

Géographie

L'île Kent (44°35'N 66°46'W), qui s'étend sur une centaine d'hectares et fait 2,5 km sur sa plus grande longueur, est située près de l'entrée de la baie de Fundy, à 8km au sud-est de l'île de Grand Manan (Figure 2). Son extrémité constitue le point le plus méridional du Nouveau-Brunswick, si l'on exclut le phare abandonné de Gannet Rock. Deux autres îles situées à proximité, Hay et Sheep, forment avec l'île Kent le complexe des Trois Îles. En raison des marées spectaculaires de la baie de Fundy, qui peuvent atteindre 5 ou 6 mètres à cet endroit, l'île Hay est accessible à pied sec depuis l'île Kent pendant quelques heures à chaque jour, à marée basse. La station scientifique Bowdoin, fondée en 1936 à la suite d'un don de la famille Rockefeller, est constituée d'une dizaine de bâtiments incluant deux laboratoires de biologie et un garage, en plus d'un quai adapté aux marées extrêmes et d'une station météo automatisée. Des scientifiques et étudiant.e.s universitaires y mènent des travaux de recherche chaque année d'avril à octobre depuis plusieurs décennies, à l'exception de l'année 2020.

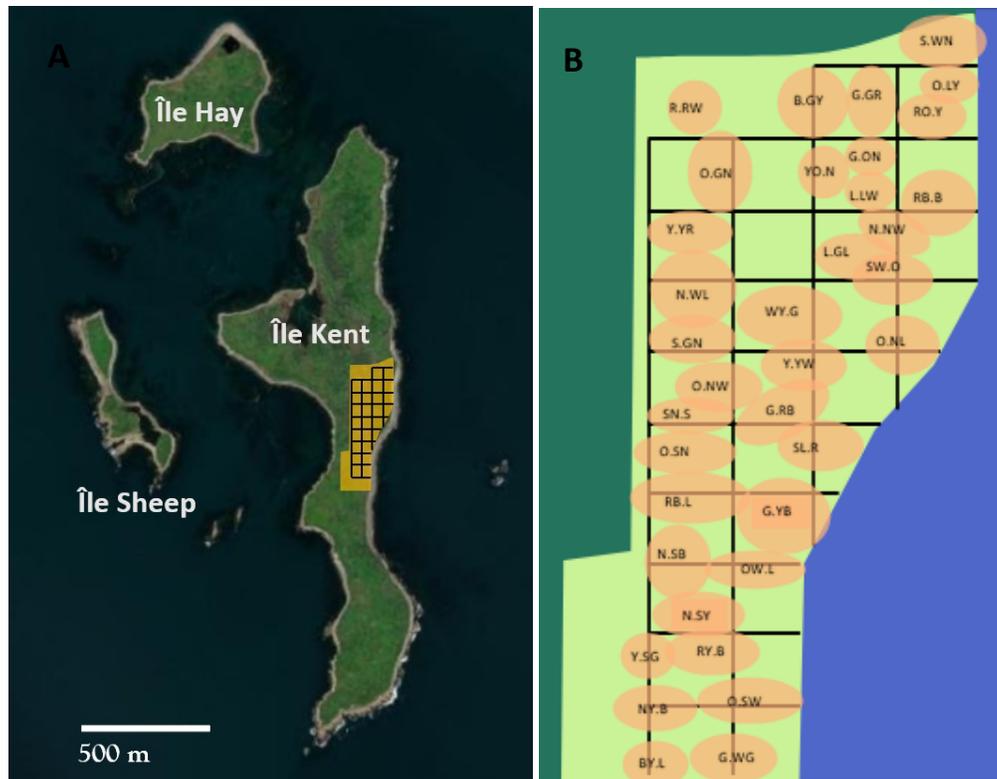


Figure 2 A. Carte du complexe des Trois îles, au Nouveau-Brunswick. Le système d'étude (ombré en jaune) et la grille sont représentés sur l'île Kent. B. Carte du système d'étude de l'île Kent. Les cercles rosés représentent le territoire des bruants des prés mâles présents au printemps 2022.

Biologie

Une pessière peu dense recouvre la partie nord et l'anse ouest de l'île Kent. Cependant, l'essentiel de la surface de l'île présente un milieu ouvert parsemé de quelques bouquets de sapins, d'aulnes ou de sorbiers, avec une couverture végétale majoritairement constituée d'herbacées et d'arbustes tels que les bleuetiers, ronces et groseillers. En termes de mammifères, l'île abrite de nombreux rats musqués et quelques loutres, mais aucun petit rongeur n'est présent. Les phoques gris sont abondants à marée basse dans la zone intertidale. La diversité aviaire est importante, avec plusieurs espèces de passereaux, d'oiseaux marins et d'oiseaux de rivage nicheurs et de passage en migration. Les goélands argentés sont très abondants, avec près de 10 000 couples nicheurs, et la population de bernaches du Canada est en nette augmentation. Les petits oiseaux subissent la prédation de rapaces, notamment les quelques pygargues qui nichent sur l'île Hay.

Les populations d'oiseaux les plus étudiées par les biologistes séjournant à la station de recherche sont les hirondelles bicolores, les bruants des prés et les océanites cul-blanc (par ex. Taylor et al. 2018, Mauck et al. 2023).

La grille

Afin de faciliter l'organisation des divers travaux sur le terrain, l'île est divisée en parcelles imaginaires de 50 x 50 mètres suivant un système de grille utilisant des lettres pour les lignes nord-sud et des nombres pour les lignes est-ouest. Dans le vaste milieu ouvert situé au centre et à l'est de l'île, qui constitue le principal système d'étude pour les travaux portant sur les hirondelles bicolores et les bruants des prés, la grille est indiquée par des marqueurs en métal reliés par des sentiers rectilignes, qui permettent d'éviter de piétiner la végétation et les nids qui pourraient s'y trouver (Figure 2). Ailleurs sur l'île, quelques sentiers permettent de se déplacer facilement et d'accéder aux plages, boisés et milieux ouverts, mais ils ne suivent pas la configuration de la grille.

Contexte acoustique

En été, le paysage sonore de l'île Kent est dominé de l'aube au crépuscule par les cris des goélands argentés, mais aussi par les vocalisations des corneilles, bruants, parulines, troglodytes et hirondelles. La nuit, ce sont les chants des océanites cul-blanc qui sont les plus proéminents. Les sons abiotiques comme le vent, la pluie et le bruit des vagues à marée haute sont omniprésents. Par ailleurs, la corne de brume de Long Point, sur l'île Whitehead, située à 5km au nord-est, est clairement audible à toute heure du jour ou de la nuit lorsqu'elle est activée. Les autres sons anthropogéniques entendus sur l'île sont principalement ceux des moteurs des bateaux de pêche, qui sont très nombreux à faible distance de la côte de l'île, ainsi que les voix, les pas et les outils de travail des chercheur.e.s et étudiant.e.s rattaché.e.s à la station scientifique et de leurs enfants.

Les bruants des prés

Écologie et comportement

Le bruant des prés est un petit oiseau au plumage rayé brun et beige, avec un bec bicolore, des pattes rose orangé et un sourcil jaune distinctif (Figure 3). Il est un peu plus petit qu'un moineau domestique ou un bruant chanteur, avec une queue courte, une envergure d'une vingtaine de centimètres et un poids d'une vingtaine de grammes. Comme son nom français l'indique, le bruant des prés habite les milieux ouverts naturels, les prairies, les bords de route, les marais salés ou la toundra, tous des milieux où la végétation est composée principalement d'herbacées et d'arbustes. Toutefois, son nom anglais de *Savannah Sparrow* n'évoque pas un habitat de savane, mais plutôt la ville de Savannah, dans l'État de la Géorgie, où l'espèce a été observée au XIX^e siècle, bien que le nom officiel de l'espèce dans plusieurs autres langues dont l'espagnol et le mandarin perpétuent une traduction erronée. Quant à lui, son nom latin *Passerculus sandwichensis* fait référence à Sandwich, dans les îles Aléoutiennes, en Alaska. En effet, les bruants des prés sont communs partout en Amérique du Nord à un moment ou à un autre de l'année. Si quelques populations résidentes se trouvent au centre du Mexique et sur la côte de la Californie, la vaste majorité des bruants des prés sont migrateurs. Ils hivernent au sud des États-Unis, à Cuba ou au Mexique, et se rendent au nord des États-Unis et au Canada pour passer la saison de reproduction. Quelques observations sont enregistrées aussi loin au nord que dans l'île Devon, en Arctique (eBird, 2023).



Figure 3 Bruant des prés (*Passerculus sandwichensis*) mâle nichant sur l'île Kent, au Nouveau-Brunswick. Les bagues de couleurs constituent la combinaison B.RW : noir (Black) et bague numérotée en aluminium à droite, rouge (Red) et blanc (White) à gauche. Photo: Daniel J. Mennill

À partir du moment où ils s'installent dans leur aire de reproduction, les mâles défendent un petit territoire d'environ 0.2 ha, ce qui équivaut approximativement à un cercle d'un diamètre de 50 mètres. Les territoires de mâles rivaux peuvent être situés très près les uns des autres. Les femelles construisent un nid au sol dans un tunnel de brindilles, et les deux parents participent à nourrir les oisillons d'une, deux, ou plus rarement, trois nichées par année avec des insectes ou des graines trouvés au sol.

On compte dix-sept sous-espèces de bruants des prés, parfois divisées en trois groupes, soit les bruants des prés à proprement parler, les bruants à bec large et les bruants d'Ipswich, bien que ce dernier groupe soit parfois considéré comme une espèce à part entière (Zink et al. 2005,

Zoonomen 2023). Le bruant des prés est considéré comme une espèce de préoccupation mineure par la liste rouge de l'Union Internationale pour la Conservation de la Nature (UICN), bien qu'elle subisse un déclin modéré (UICN, 2021). L'organisme Partners in Flight lui attribue une demi-vie de population de 52 ans, c'est-à-dire que c'est l'échéance à laquelle on prévoit que la population aura diminué de 50% si les tendances démographiques de la période allant de 2005 à 2016 se poursuivent (Partners in Flight, 2023).

Population de l'île Kent

Les bruants des prés de la sous-espèce *P. s. savanna*, ou bruant des prés de l'Est, sont présents en grand nombre sur les îles avoisinant Grand Manan, dans les milieux ouverts à couverture herbacée. Les îles Kent, Hay et Sheep abritent respectivement environ 225, 35 et 20 couples reproducteurs à chaque année. Woodworth et al. (2017) ont montré que les individus de cette population passent l'hiver sur la côte est des États-Unis, du Maryland à la Floride. Les mâles arrivent de leur migration printanière vers la fin du mois d'avril, et les femelles les rejoignent dans les premières semaines de mai. Les individus des deux sexes de cette population sont hautement philopatrics : malgré une migration de plusieurs milliers de kilomètres, ceux qui reviennent nicher sur l'île s'installent à une distance médiane de seulement 228.5m de leur lieu d'éclosion (Wheelwright & Mauck, 1998).

Les bruants des prés nichant dans le système d'étude situé au centre de l'île Kent sont étudiés depuis plusieurs décennies (Dixon 1978, Wheelwright et al. 1992, Williams et al. 2013, Woodworth et al. 2017, Mennill et al. 2018, Hensel et al. 2022a). Chaque année, tous les adultes non bagués observés dans ce secteur sont appelés avec des enregistrements de chants diffusés par des haut-parleurs et capturés avec des filets japonais installés dans les sentiers près de leur territoire. Ils sont bagués avec une bague en aluminium comportant un numéro unique, ainsi que trois bagues de couleur en plastique, pour un total de deux bagues très légères sur chaque patte (Figure 3). Ils sont ensuite sexés, mesurés et pesés, un échantillon de sang leur est prélevé et leur âge est estimé d'après les patrons de mue des plumes de leur queue. Toutes les manipulations sont réalisées par des personnes bien formées et suivant des protocoles stricts de bons soins aux animaux. De grands efforts sont faits pour que les oiseaux subissent un stress minimal, et ils sont la plupart du temps relâchés en moins de 15 minutes. Les nids sont repérés et protégés de la

prédation des corneilles grâce à des grillages, puis suivis quotidiennement jusqu'à l'envol des oisillons.

Depuis 2016, à l'exception de l'année 2020, des enregistreurs autonomes sont disposés dans tous les milieux ouverts des îles Kent et Hay pendant 24 à 48h en mai ou juin pour recueillir des enregistrements de chaque mâle (Hensel et al. 2022a). L'île Sheep, moins étendue et plus difficile d'accès, est couverte par ce protocole lors de certaines années. En 2022, 30 enregistreurs disposés à chaque intersection de la grille dans le système d'étude principal situé sur l'île Kent ont été programmés pour enregistrer à tous les jours du lever au coucher du soleil pendant le mois de mai, afin de récolter davantage de chants par individu et de dessiner un portrait plus complet de l'activité vocale des bruants des prés.

Chants et culture vocale

Plusieurs projets de recherche ont été consacrés au chant des bruants des prés mâles de l'île Kent, dont on sait que les fonctions principales sont la protection du territoire et l'attraction d'une partenaire (Moran et al. 2019). Certaines études visent à l'innovation méthodologique, comme celle de Hensel et al. (2022b), qui a notamment montré que le suivi acoustique passif peut servir à estimer la taille de la population et la longévité des individus. Toutefois, les écologistes sont particulièrement intéressé.e.s par le fait que les bruants des prés adultes produisent un seul chant individuellement distinctif, qu'il apprennent jeunes d'un tuteur social avec qui ils n'ont pas nécessairement de lien génétique, puis qui reste fixé toute leur vie durant (Wheelwright et al. 2008). Ce procédé se fait par attrition, ce qui signifie que les jeunes mâles apprennent plusieurs chants, mais n'en retiennent qu'un seul, qui est souvent le plus semblable à ceux de ses voisins (Thomas et al. 2021). Une expérience particulièrement intéressante a montré que les jeunes oiseaux de cette population peuvent apprendre des chants artificiels diffusés par haut-parleur, et transmettent à leur tour ces nouveaux chants aux générations suivantes (Mennill et al. 2018).

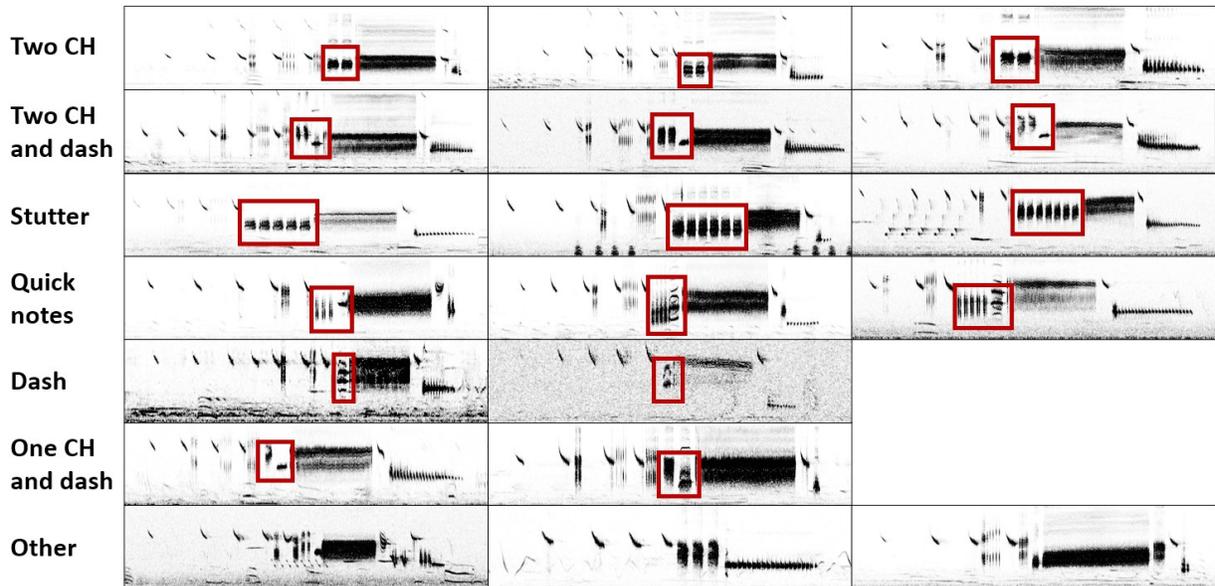


Figure 4 Exemples de chacun des microdialectes observés chez les bruants des prés de l'île Kent, Nouveau-Brunswick, depuis 2016, et de chants n'appartenant à aucun groupe. Les cadres rouges indiquent la section centrale caractéristique du microdialecte.

Il semble que différents mécanismes d'évolution culturelle et différents rythmes d'évolution affectent les quatre sections du chant des bruants des prés de l'île Kent (Williams et al. 2013). Notamment, le *buzz* est un marqueur populationnel stable, mais la section centrale, qui varie au sein même de la population, pourrait être le produit de la dérive culturelle (Williams et al. 2019, Williams 2021). Dans le cas du train de clics entendu dans la partie initiale du chant, on trouve la marque de l'évolution culturelle cumulative, où des changements successifs dans un comportement acquis résultent en une fonctionnalité améliorée, dans ce cas-ci un chant élicitant de plus fortes réactions de la part des congénères (Williams et al. 2022).

On connaît depuis longtemps la présence de dialectes à grande échelle géographique chez les bruants des prés (Bradley 1994, Sung et Handford 2006, Williams et al. 2019). Or, Hensel et al. (2022a) ont démontré la présence de microdialectes à très petite échelle géographique, qui sont le produit de la faible distance de dispersion et de la favorisation de chants similaires à ceux de leurs voisins chez les jeunes mâles. Depuis 2016, on observe dans la population de l'île Kent six de ces microdialectes caractérisés par une section centrale commune, en plus des chants issus

des expériences de tuteurs artificiels, qui comportent des syllabes encore rares dans la population, et les chants aberrants présents naturellement (Figure 4).

Objectifs de l'étude

Le but du présent mémoire est d'aider à déterminer dans quelle mesure BirdNET est sensible à la variation intraspécifique lorsqu'il attribue une prédiction de classe d'espèce à un chant, en utilisant des enregistrements de bruants des prés réalisés sur l'île Kent. Suivant les résultats rapportés par Pérez-Granados (2023a), nous nous attendons à ce que BirdNET soit à même d'identifier avec un haut score de confiance l'espèce à laquelle appartient un chant, malgré d'importantes variations intraspécifiques.

Nous présentons à l'occasion de cette étude un jeu de données unique constitué d'enregistrements focaux et omnidirectionnels récoltés au printemps 2022 et annotés manuellement avec un haut degré de confiance. Ce corpus détaillé compte près de 70 000 chants de bruants des prés attribués à une cinquantaine d'individus appartenant à six microdialectes.

En conclusion, nous discutons des questions encore en suspens quant à l'hétérogénéité intraspécifique, des enjeux de transparence et d'interprétabilité, et des avenues de recherche rendues possibles par le jeu de données dont nous disposons désormais.

Chapitre 2 – Evaluating the generalization ability of a deep neural network across intraspecific variation in Savannah Sparrow song

Contribution des co-auteurs

- Rondeau Saint-Jean, C. : Conceptualisation de l'étude, acquisition de fonds, récolte des données sur le terrain, annotation des données, analyse des données, écriture, visualisation
- Mennill, D. J. : Supervision, conceptualisation de l'étude, acquisition de fonds, analyse des données, aide logistique sur le terrain, méthodologie, révision du manuscrit
- Dobney, S. L. : Récolte de données et aide logistique sur le terrain,
- Bornais, M. : Récolte de données et aide logistique sur le terrain
- Norris, D. R. : Récolte de données
- Doucet, S. M. : Récolte de données
- Newman, A. E. M. : Récolte de données
- Williams, H. : Récolte de données
- Poisot, T. : Supervision, conceptualisation de l'étude, acquisition de fonds, analyse des données, méthodologie, révision du manuscrit

Evaluating the generalization ability of a deep neural network across intraspecific variation in Savannah Sparrow song

Camille Rondeau Saint-Jean^{1*}, Daniel J. Mennill², Sarah L. Dobney², Mikayla Bornais², D. Ryan Norris³, Stéphanie M. Doucet², Amy E. M. Newman³, Heather Williams⁴, Timothée Poisot¹

*Corresponding author (camille.rondeau.saint-jean@umontreal.ca)

¹Département de Sciences Biologiques, Université de Montréal, Montréal, QC, Canada

²Department of Integrative Biology, University of Windsor, Windsor, ON, Canada

³Department of Integrative Biology, University of Guelph, Guelph, ON, Canada

⁴Biology Department, Williams College, Williamstown, MA, U.S.A

Abstract

Machine learning, especially deep neural networks, has become a valuable tool for ecological and wildlife conservation studies. One such application is the acoustic detection and classification of bird species, exemplified by BirdNET, a deep convolutional neural network capable of identifying over 3000 bird species. However, the generalization capacity of deep learning algorithms in the face of intraspecific variation in bird songs remains poorly understood, raising concerns about potential misinterpretation of results. In this study, we evaluate the performance of BirdNET on a large corpus of Savannah Sparrow songs collected from a well-studied population on Kent Island, Canada. Savannah Sparrows exhibit significant intraspecific variation in song, including microdialects, making them an ideal test case. We investigate whether BirdNET exhibits biases in its classification across microdialects and individuals. Our results show acceptable prediction accuracy, with 81.9% of songs correctly assigned to the correct species. We find variation in BirdNET's confidence scores among individuals, indicating potential biases. However, we suspect that these discrepancies could be caused by the recording methods, since individuals consistently recorded far from the field recorders appear to have lower classification accuracy. Notably, we find no evidence for differences in confidence scores across microdialects. We emphasize the need for transparent tools, such as class activation maps, to understand deep neural network decisions and enable better interpretation of results.

Introduction

Machine learning is becoming an important tool for ecological studies (Valletta et al. 2017, Christin et al. 2019, Borowiec et al. 2022, Ghosh and Dasgupta 2022, Pichler and Hartig 2023) and for wildlife conservation planning (Humphries et al. 2018, Lapeyrolerie et al. 2022, Tuia et al. 2022). Methods based on deep learning are particularly well suited to the analysis of complex and large datasets, such as those collected through camera traps or passive acoustic monitoring projects. Indeed, deep neural networks aimed at the acoustic detection and classification of species considerably accelerate the processing of large audio datasets (Priyadarshani et al. 2018, Cole et al. 2022, Stowell 2022) and show excellent performance over a wide range of taxa and recording environments (Kumar et al. 2022, Maegawa et al. 2022, Stowell 2022). For example, BirdNET-Analyzer (hereafter “BirdNET”) is a recently developed deep convolutional neural network which can identify the song of over 3000 bird species (Kahl et al. 2021, Wood et al. 2022) with an accuracy between 72 and 85% across different datasets (Pérez-Granados 2023a).

Despite significant advances in automated bird species identification, we still know very little about the extent to which variation in song within species affects the performance of deep learning algorithms. Such a gap in our understanding of the generalization capacity of deep neural networks across intraspecific variation exposes the scientific community to misuse of these tools and misinterpretation of their results. Luccioni and Rolnick (2022) showed that for wildlife image recognition, a model trained on a biased, inaccurate, or incomplete dataset can fail to detect certain categories of targets and propagate the biases from the training dataset. Indeed, an acoustic species recognition model trained on a geographically biased subset of songs (Martin et al. 2012, Sugai et al. 2019, Hughes et al. 2021) could have an uneven performance in space if it turned out to detect some regional dialects less consistently than others. Furthermore, a model for species identification based on song will suffer from domain shift (Widmer & Kubat 1996, Stowell 2022) and become obsolete faster if it has not been trained with enough intraspecific variation to be robust against the cultural evolution that is likely to happen in many species in which songs are learned (Whiten 2017, Schuppli and Schaik 2019, Williams 2021).

Here we test the performance of BirdNET in processing audio recordings characterized by variation between individuals of the same species. We accomplish this by examining the output of BirdNET when presented with a large corpus of songs of a single species – Savannah Sparrows – labelled by experts and encompassing significant intraspecific variation. Our research focuses on a well-studied population of Savannah Sparrows on Kent Island, in the Bay of Fundy, New Brunswick, Canada. Birds in this study population have been color-banded and extensively recorded yearly since the 1980s (Wheelwright et al. 2008, Woodworth et al. 2017), therefore we have access to an exceptionally large set of audio data confidently labelled at the level of the individual. Each Savannah Sparrow male sings a single, short song that remains the same over his entire adult life (Wheelwright et al. 2008, Thomas et al. 2021). They can be identified acoustically with relative ease thanks to small, consistent variations in the characteristics of certain elements of their song (Williams et al. 2013, Mennill et al. 2018, Hensel et al. 2022b). Cultural evolution and geographical dialects have been well documented in this species (Bradley 1994, Sung and Handford 2006, Williams et al. 2013, Williams 2021, Williams et al. 2022). In particular, Hensel et al. (2022a) showed that six different variants in the middle section of the song of the Savannah Sparrows characterize the microdialects found within the Kent Island population.

We seek to determine how robust BirdNET is in the face of intraspecific variation. More precisely, we investigate if the algorithm is more confident in its classification of some microdialects or individuals as Savannah Sparrows than others. Since BirdNET has been trained on an abundant dataset sourced from participative platforms such as Macaulay Library (Macaulay, 2023) and Xeno-Canto (Xeno-Canto, 2023), with recordings originating from a variety of locations throughout the species' ranges (Kahl 2020), we predict that it will have a similar performance across microdialects and individuals, and that it will perform consistently with standards reported in Kahl (2020) when applied to a detailed study of individuals exhibiting microdialects in a well-recorded population.

Methods

Study site and population

We collected extensive recordings of Savannah Sparrows at the Bowdoin Scientific Station on Kent Island, New Brunswick, Canada (44.5818°N, 66.7547°W). This population has been studied for many decades (Dixon 1978, Wheelwright et al. 1992, Williams et al. 2013, Woodworth et al. 2017, Mennill et al. 2018). All adults breeding in the 10-ha study area in the center of the island are banded each year with unique colour combinations allowing individual identification (Woodworth et al. 2017). Between 28 and 48 males are recorded and observed annually in the central study area (Hensel et al. 2022b). As described in Woodworth et al. (2017) and Mennill et al. (2018), researchers also map territories, locate nests, and monitor nestlings. Territorial mapping, nest location, and nestling monitoring are carried out every year following the protocols described in Mennill et al. (2018). In recent years, 6 song types, or microdialects, have been described in the population based on the group of syllables constituting the distinctive middle section of the song (Williams et al. 2013, Williams 2021, Hensel et al. 2022a, Hensel et al. 2022b) (Figure 4).

Collection of recordings

We used autonomous recorders to collect extensive sampling of the songs of birds in the central area of the study population. Thirty automated recorders were arranged in a grid of 50 x 50 m (14 Wildlife Acoustics Song Meter SM2 and 16 Wildlife Acoustics Song Meter Mini Acoustic Recorder; 44.1 kHz sampling frequency; 16-bit accuracy; WAV format; Mennill et al. 2012, Hensel et al. 2022b). Recorders were programmed to sample birds' voices between 4 AM and 10 PM every day from April 29 to May 30, 2022. In addition to these automated recordings, we collected in-person focal recordings of the song of all adult males with field recorders (Marantz PMD661 digital recorders; Sennheiser ME66/K6 microphones mounted in Telinga parabolas; 44.1 kHz sampling frequency; 16-bit accuracy; WAV format; Mennill et al. 2018). When it became apparent that some microdialects were underrepresented in the population, we sought to record birds belonging to rarer song types breeding in other parts of Kent Island and on neighbouring Hay Island. We thus collected extensive focal and autonomous recordings for five birds with a

distinctive song outside the main study area. Overall, we obtained focal recordings for 51 individuals, and 36 of these individuals were also recorded with Song Meters recordings. One additional individual that was not targeted by focal recordings but well captured by a Song Meter was included in analyses. All procedures involving animals were reviewed and approved by the University of Windsor Animal Care Committee (permit AUPP-20-09).

Processing of recordings and sound annotation

We built a reference song library by gathering good quality examples of the song of 51 males from focal recordings where the colour bands of the bird could be confirmed visually by the recordist. Using Syrinx-PC sound analysis software (Burt 2000, v.2.6L), we browsed through all focal recordings and a fraction of Song Meter recordings during peak singing periods, mostly between 5 and 8 AM, from May 1st to May 15th (Moran et al. 2019), to label songs that could confidently be cross-referenced from the library based on individual acoustic characteristics (Williams et al. 2013, Williams 2021). We annotated a total of 69,606 songs. Overall, 24 males had more than 2250 songs annotated, and 6 microdialects were represented by at least 2 individuals and 5000 songs (See Annex 1). Four of the males could not be associated to a microdialect, either because their song was a natural outlier (i.e., they are a suspected immigrant to the population), or because it was derived from an artificial tutoring experiment performed from 2013 to 2018 in the population (Mennill et al. 2018). Partial revision of the annotations led us to estimate that the error rate for the manual attribution of songs to individuals is less than 1%. Using python library 'pydub' (Robert 2021, v.0.25.1), we clipped each annotated song into individual 3-second WAV files.

Computational work

We ran all song clips through BirdNET-Analyzer (Kahl 2022, v.2.2) with no temporal or geographical restrictions to the list of species, a detection sensitivity of 1.0, no overlap of segments and a minimum confidence threshold of 0.01. Of the 69,606 clips, 67,346 belonged to one of the six major microdialects, and 60,209 were attributed to an individual with more than 2250 songs. The output for each clip consisted of a list of all species for which the confidence score exceeded the specified threshold. We assigned a confidence score of 0 to the clips for which

the score attributed was below 0.01. A confidence score close to 1 denotes a high degree of certainty, whereas a lower score indicates more uncertainty. We compiled all results into a single file with the python library ‘pandas’ (The Pandas Development Team 2022, v.1.4.2).

Statistical analyses

We used R version (R Core Team 2023, v.4.3.1) and RStudio version (RStudio Team 2023, v.2023.06.1+524) for statistical computations. Data manipulation and visualization were performed using the ‘dplyr’ (Wickham et al. 2023, v.1.1.2) and ‘ggplot2’ (Wickham 2016, v.3.4.2) packages, respectively. We chose to visualize both confidence score, which is independent for each species, and prediction rank, which compares predicted species together. We used one-way ANOVAs and post-hoc Tukey tests to compare and group average confidence scores attributed to songs by BirdNET according to the individual and microdialect to which they belong. For the 24 individuals with more than 2250 songs, we fit a linear regression to confidence scores attributed to songs by BirdNET according to the individual and calculated estimated marginal means using Tukey’s adjustment. To account for the fact that individual males were sampled repeatedly, we averaged confidence scores for all birds and fit a linear regression to these averages according to the microdialect. We obtained letter-based groupings of the estimated marginal means, indicating which individuals or microdialects have significantly different means. Statistical tests required the packages ‘stats’ (R Core Team 2023, v.4.3.1), ‘emmeans’ (Lenth et al. 2023, v.1.8.7) and ‘multcomp’ (Hothorn et al. 2023, v.1.4-25).

Results

BirdNET performed well at recognizing Savannah Sparrow songs from our dataset. The proportion of manually labelled songs for which “Savannah Sparrow” was the class to which BirdNET attributed the highest score, and therefore correctly predicted the species, was 81.9%. 220 other species occurred as a top prediction for at least one clip. This included species that were present on Kent Island and may have been louder than Savannah Sparrows on recordings, such as Herring Gulls (10.2% of top predictions), Tree Swallows (1.3% of top predictions) and Canada Geese (0.2% of top predictions), as well as other gull species that could easily have been misattributed to loud

Herring Gull vocalizations (3.0% of top predictions). Most of the remaining 3.4% of top predictions were attributed to bird species that are not found on Kent Island, such as Superb Fairywren, Sulphur-crested Cockatoo, African Blue Tit, and Short-toed Snake-Eagle, or other categories of implausible sound events such as gunshot, fireworks, or coyote. For each clip, between 1 and 39 species were detected. The confidence scores BirdNET attributed to the species “Savannah Sparrow”, independently of its rank, ranged between 0 and 0.999 and were on average 0.52.

BirdNET showed varying scores and accuracy across microdialects and individual birds. For the different microdialects, the average confidence score for “Savannah Sparrow” ranged between 0.47 and 0.61, and the proportion of top predictions were between 78.9% and 85.7% (Tableau 1, Figure 5a and c). For both metrics, the ‘Stutter’ microdialect was the least successful, and the ‘Dash’ microdialect obtained best results. The average confidence score for the 24 individuals with more than 2250 songs ranged between 0.33 and 0.86 and the proportion of correct top predictions were between 63.5% and 98.3% (Figure 5b and d).

Confidence scores and success rates differed between clips recorded through Song Meters and manual focal recorders. Clips issued from focal recordings were correctly identified 89,3% of the time, and obtained an average confidence score of 0.62, whereas songs recorded with Song Meters had a 79,8% success rate and an average score of 0.50. Song Meter recordings represent 78,1% of the full dataset, between 71,8% and 99,2% of each microdialect, and between 66,6% and 100% of the clips of individuals, therefore they bear an important weight in the overall results of each category.

The analysis of variance (ANOVA) showed strong evidence that there were significant differences among the average scores attributed by BirdNET to different individuals, but not between microdialects. The results showed a significant effect of individuals on confidence scores ($F(23, 60185) = 609.4, p < 0.001$). The effect of microdialect on confidence scores was not significant ($F(5, 47) = 1.223, p = 0.313$). The calculated effect size was relatively small, suggesting that the variation explained by the microdialect was limited. Post-hoc comparisons conducted using Tukey’s adjustment revealed that the six microdialects formed one single group, whereas the 24 individuals formed 15 statistically different, overlapping groups (Figure 5a and b).

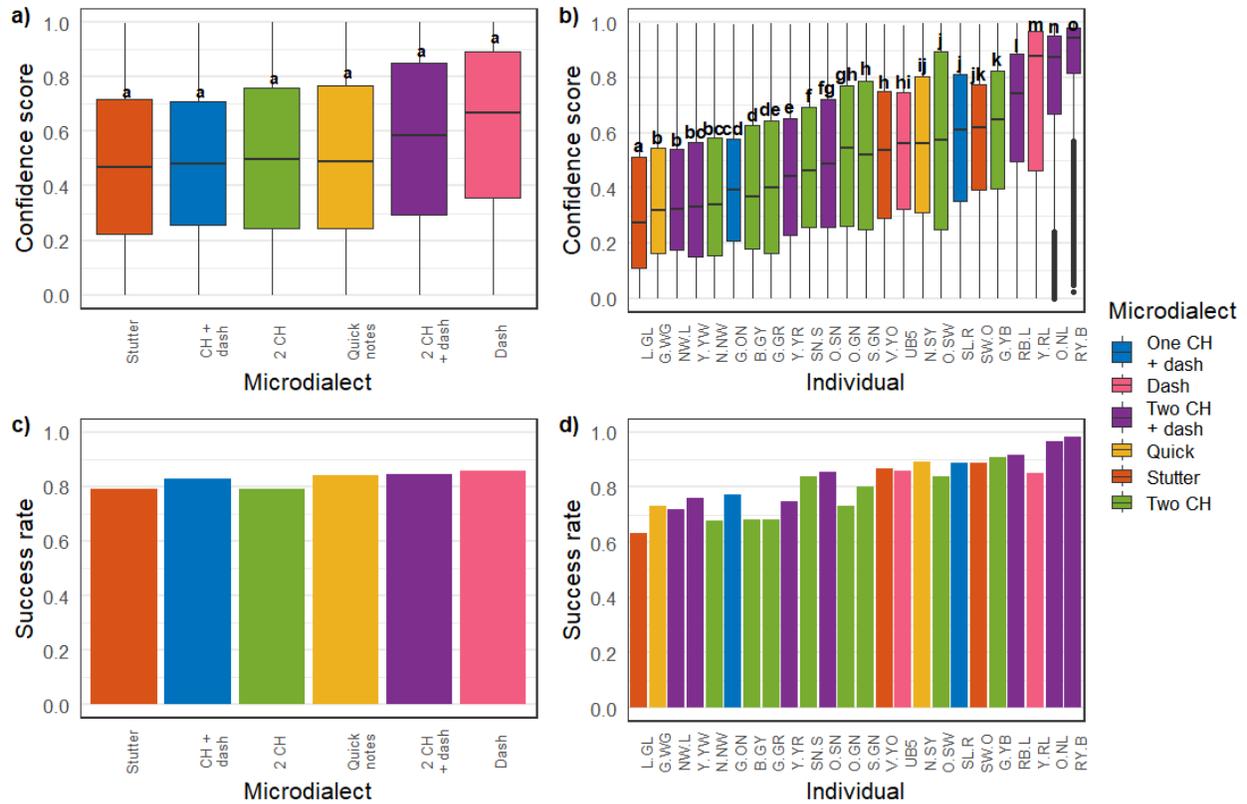


Figure 5 Distribution of confidence scores attributed to Savannah Sparrow songs from the Kent Island population by BirdNET and rate of correct predictions, by microdialect and individual. Individuals are identified by letters representing the colour-bands on their legs.

Discussion

To evaluate the performance of BirdNET, a deep neural network for acoustic species detection and classification (Kahl et al. 2021), we used a dataset of almost 70,000 manually annotated Savannah Sparrow songs collected from a long-term study population in eastern Canada. We compared the accuracy of the predictions and the average confidence scores across 6 microdialects and 24 extensively recorded individuals. We found that the prediction accuracy and the average confidence score were often higher than results presented by the developers of BirdNET. Indeed, we found that 81.9% of predictions on our dataset were correct, whilst Kahl (2020) found that 72.5% of Savannah Sparrow songs were correctly identified and Kahl et al. (2021) reported a 77.7% accuracy in predictions across all species, although our results differed

across microdialects and individuals. It is also important to note that we did not use the geographical and temporal filters that can narrow down the candidate classes from 3300 to a few hundred species that are realistically present at the time and place of recording, which would have potentially improved further the prediction accuracy of the algorithm. We conclude that the current version of BirdNET works effectively at recognizing a good proportion of Savannah Sparrow songs from the Kent Island population.

BirdNET's developers trained it with hundreds to thousands of samples for each species, including 2659 Savannah Sparrow clips (Kahl 2020). At first glance, our results seem to show that this was not sufficient for BirdNET to learn to generalize equally across various individuals, and that it recognized some of them more easily as Savannah Sparrows because their songs were closer to those present in the training data. However, there is no evidence that some microdialects were more recognizable than others, and there is no obvious pattern in the individuals results that could point towards what song characteristics would explain such discrepancies. One possibility is that the differences observed are due to bias in our data. Indeed, not all clips analyzed were of a high quality. Many songs were faint and distant, or overlapped with loud background noises. All these factors can negatively influence whether BirdNET can accurately and confidently identify the target species as a Savannah Sparrow, especially if sounds produced by other bird species are prominent in the recording. This would not pose a problem if lower quality recordings were equally present in the dataset of all individuals. However, Savannah Sparrows prioritize a small number of singing perches in their territory. Some will consistently sing from song perches far away from Song Meter recorders, and thus produce faint recordings with more background noise. Others will sing closer to a grid corner and their songs will regularly appear loud and clear on recordings, which leads to better recognition (Pérez-Granados 2023b). We suspect that a significant portion of the variation in average confidence scores and prediction success between individuals is due to this phenomenon (See Annexe 2).

The overall performance of BirdNET with our data was good, especially considering that our song clips are only 3 seconds long, and therefore offer much less opportunities to detect the species than the longer recordings often used in similar tests (Pérez-Granados 2023a). This could be attributed to the fact that we used a more recent version of BirdNET-Analyzer (v.2.2).

In the spirit of transparency, developing and giving access to tools allowing to better interpret the results of a model would increase the confidence of users and enable accountability for conservation decisions that rely on its predictions. For example, Kahl (2020) presents class activation maps, which consist in heatmaps overlaid on spectrograms showing which elements of a song are decisive in making a species prediction. In the present study, this would have been a valuable tool for attempting to determine if certain song characteristics make an individual or a microdialect more recognizable for BirdNET. However, we were not able to produce class activation maps from the newer versions of the model, and documentation to this effect was not available. We encourage the development of equally informative tools adapted to recent innovations in the architecture of deep neural networks used in bioacoustics.

In light of the results obtained in this study, future research efforts may focus on conducting similar comparisons of the performance of deep neural networks on single-species audio datasets encompassing significant intraspecific variation. It appears that this approach would benefit from the inclusion of more individuals and drawing from populations originating from larger geographical ranges, which would allow to detect more large-scale patterns. Furthermore, broadening this investigation to species with larger, multi-song repertoires or more subtle interindividual song differences or dialects could lead to a deeper understanding of the generalization potential and limits of deep neural networks.

Tableau 1 Sample sizes, proportion of correct predictions by BirdNET-Analyzer (Kahl et al. 2021) and average confidence score for 6 microdialects found in 2022 in Savannah Sparrows on Kent Island, New Brunswick. No analyses were performed on the songs of birds belonging to no microdialect.

Microdialect	Number of songs	Number of birds	Proportion of success	Average confidence score
Two “CH”	23,182	20	78.9%	0.50
Two “CH” and dash	19,006	17	84.7%	0.56
Stutter	8,391	3	79.1%	0.47
Quick notes	6,667	4	84.1%	0.50
Dash	5,067	2	85.7%	0.61
One “CH” and dash	5,033	2	82.7%	0.48
Other	2,260	4	NA	NA
Total	69,606	52		

Chapitre 3 – Conclusion générale

Nous avons montré que BirdNET affiche une très bonne performance, mais il ne nous est pas possible de confirmer qu'il n'est pas biaisé contre certains individus, bien qu'il semble reconnaître tous les microdialectes avec la même aisance. Il reste encore beaucoup à faire pour comprendre sur quoi s'appuient les réseaux neuronaux quand ils font une prédiction et pour discerner à quels biais ils sont susceptibles dans le domaine de l'écologie et de la bioacoustique. Pour conclure ce mémoire, voici quelques suggestions de pistes de recherches qui nous apparaissent nécessaires pour mieux appréhender les écueils qui risquent de se présenter aux utilisateurs de ces algorithmes, mais aussi pour mieux exploiter leur important potentiel.

Évaluation de la capacité de généralisation

Pour bien évaluer la performance de BirdNET à travers la variation présente dans les chants des bruants des prés, il serait nécessaire de faire appel à un corpus de chants issus d'une échelle géographique beaucoup plus importante. Il faudrait prendre exemple sur l'étendue couverte par des études portant sur la variation culturelle chez d'autres espèces ayant une vaste distribution (Otter et al. 2020, Wood et al. 2022, Ghani et al. 2023). En effet, les bruants des prés sont présents dans l'ensemble de l'Amérique du Nord, et l'existence de dialectes à grande échelle (Bradley 1994, Sung et Handford 2006, Williams et al. 2019) ainsi que d'un grand nombre de sous-espèces (Zink et al. 2005) est bien documentée. Il serait donc pertinent, pour déterminer si la performance de BirdNET est constante sur l'ensemble de l'aire de distribution de l'espèce, de comparer les scores de confiance et les taux de succès obtenus dans des populations éloignées les unes des autres. Par ailleurs, afin de vérifier si le modèle est robuste face à la dérive des données occasionnée par l'évolution culturelle (Williams 2019), il serait souhaitable d'établir des comparaisons entre des enregistrements réalisés à plusieurs décennies d'écart, idéalement avec des technologies audio équivalentes.

Il serait toutefois possible de mener des travaux plus approfondis même en se limitant à la population de l'île Kent. On a vu que l'utilisation d'un large corpus où peu d'individus sont représentés avec un très grand nombre de chants occasionnait des biais importants dus à leur territoire restreint et aux méthodes d'enregistrement passif. Afin de déterminer avec plus d'assurance si certains microdialectes sont défavorisés par l'algorithme, il serait souhaitable de tester plutôt un corpus incluant un petit nombre d'enregistrements de haute qualité pour tous les oiseaux des trois îles sur plusieurs années, en réutilisant par exemple les enregistrements analysés par Hensel et al. (2022a et 2022b).

Il reste aussi à déterminer si nos conclusions sur la capacité de généralisation de BirdNET face à la variation intraspécifique chez les bruants des prés s'appliquent à d'autres espèces. En effet, il faut considérer que le fait que les bruants des prés soient communs et répandus en Amérique du Nord implique qu'une grande quantité de données audio ont été disponibles pour entraîner le modèle sur des bases de données où les chercheurs et le public peuvent déposer des enregistrements, comme eBird (eBird, 2023), Xeno-Canto (Xeno-Canto, 2023), Macaulay (Macaulay, 2023) et AudioSet (AudioSet, 2023). Kahl (2020) révèle que la première version de BirdNET a été entraînée avec 2659 extraits audio de bruants des prés. À titre de comparaison, peu d'espèces dépassent la barre des 3000 exemples, alors qu'elles sont nombreuses à en compter moins de 1000 ou même moins de 500. Il faudrait tester si BirdNET est aussi robuste face à la variation intraspécifique chez les espèces pour lesquelles il a été entraîné avec un petit nombre d'exemples qu'il l'est chez les bruants des prés.

Il serait par ailleurs crucial d'étendre les études sur la capacité de généralisation de BirdNET à des espèces dont les patrons de variation intraspécifiques sont très différents de ceux des bruants des prés. Par exemple, il serait intéressant d'explorer cette question pour des espèces chez qui la variation entre les individus ou entre les dialectes est beaucoup plus marquée, ou pour celles où les individus exhibent un répertoire plus étendu. De plus, soulignons l'importance de mieux comprendre comment l'algorithme atteint l'équilibre entre le discernement entre espèces et la généralisation en dépit de l'hétérogénéité intraspécifique, dans les cas de familles d'espèces dont les chants peuvent facilement être confondus.

Interprétabilité et transparence

L'un des obstacles à l'adoption plus étendue des outils d'intelligence artificielle en écologie est la méfiance qu'ils peuvent susciter en raison du caractère de « boîte noire » de leur fonctionnement. En effet, comme ce domaine est relativement peu concerné par le secret commercial et par des problèmes éthiques liés à la vie privée, mais qu'il fournit des informations menant à des décisions gouvernementales, il fait face à un besoin de transparence et à des attentes élevées en termes d'interprétabilité. La Déclaration de Montréal (2018) va dans ce sens en énonçant le principe de responsabilité, selon lequel « seuls des êtres humains peuvent être tenus responsables de décisions issues de recommandations faites par des systèmes d'intelligence artificielle », ce qui implique que les recommandations en question doivent être compréhensibles et explicables. De plus, même les chercheurs intéressés par des questions plus fondamentales expriment le besoin de comprendre et de pouvoir interpréter les résultats que leur donnent leurs outils (Lucas 2020).

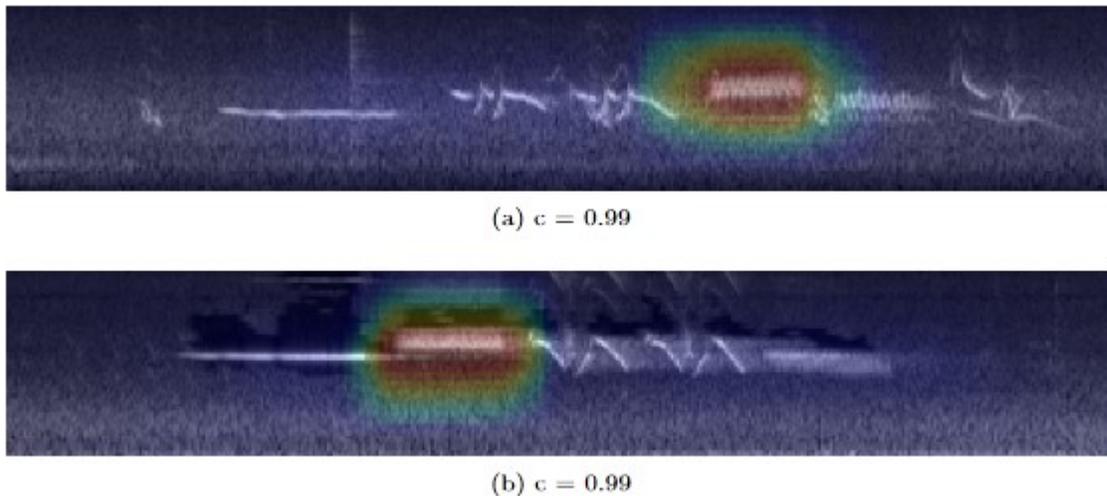


Figure 6 Cartes d'activation de classe pour des chants de bruants à couronne blanche (*Zonotrichia leucophrys*), avec un score de confiance c . La zone en rouge met en évidence l'élément du chant qui est déterminant pour la classification de cette espèce. Extrait de Kahl (2020).

Pour répondre à ce besoin et démontrer de façon convaincante la validité de certaines prédictions, Kahl (2020) présente des cartes d'activation de classe (Figure 6). Celles-ci indiquent par une carte thermique superposée sur un spectrogramme quels éléments d'un signal sonore sont les plus importants pour le modèle au moment de l'assigner à une classe (Zeiler and Fergus 2015, Zhou et al. 2016). Nous souhaitons reprendre cette méthode afin d'aider à expliquer la variation ou la cohérence dans les prédictions au sein de nos catégories de chants de bruants des prés ou entre elles. Nous espérons ainsi pouvoir montrer quand BirdNET s'appuie sur des éléments fiables du chant d'un bruant des prés, tels que les notes hautes initiales ou le buzz (Williams et al. 2019), pour prédire l'espèce, et quand il se laisse tromper par des bruits ou des éléments moins importants ou qui ne sont pas valides dans tous les contextes. Ce type d'outil représente une opportunité, trop rare en bioacoustique, de valider les prédictions et de détecter les biais, mais aussi, d'isoler les caractéristiques d'un chant qui le distinguent de toutes les autres catégories présentes dans le jeu d'entraînement du modèle dont il est issu. Une telle capacité d'interprétation permettrait par ailleurs des comparaisons inédites dans le domaine de la psychoacoustique, entre ce sur quoi se fient les oiseaux, les humains et les réseaux neuronaux profonds pour reconnaître et classifier le chant d'un oiseau.

Il faut noter ici que la littérature disponible sur le projet BirdNET, abondamment citée, concerne sa version initiale entraînée sur 984 espèces (Kahl 2020, Kahl et al. 2021). Cette version utilise la librairie Theano, aujourd'hui largement désuète, et est marquée comme obsolète sur le site de dépôt de code GitHub depuis avril 2022. C'est donc plutôt la version BirdNET-Analyzer, disponible en ligne depuis septembre 2021 et reconnaissant plus de 3000 espèces, que nous avons utilisée dans les analyses. Fait à noter, il est malheureusement souvent assez difficile de bien comprendre sur quelle version se basent les recherches et les discussions entourant BirdNET. Par exemple, Pérez-Granados (2023a) mêle des études menées avec plusieurs versions sans discuter cette question. Cela souligne l'importance de bien indiquer quelle version d'un algorithme est utilisée dans le cadre de travaux scientifiques.

BirdNET est décrit dans Kahl 2020 et Kahl et al. 2021 comme un algorithme qui applique aux fichiers audio un prétraitement qui en fait des spectrogrammes avant de les faire passer dans un réseau neuronal profond basé sur une architecture développée pour la reconnaissance d'images.

Nous avons donc envisagé d'utiliser Grad-CAM (Selvaraju et al. 2020), un algorithme fonctionnant avec la librairie python Tensorflow et conçu pour être greffé à des modèles de reconnaissance d'images pour produire des cartes d'activation de classe interprétant leurs prédictions. En effet, les représentations par spectrogramme décrites dans Kahl et al. (2021) se présentent sous la forme d'une matrice rectangulaire de 64 x 384 pixels, ce qui correspond à 64 « bins » de fréquences sur 3 secondes à 148 fenêtres par seconde. Or, plutôt que de correspondre à ce format, l'entrée des modèles récents consiste en un vecteur de 144 000 nombres, ce qui semble correspondre à la forme d'onde d'un enregistrement de 3 secondes effectué à une fréquence d'échantillonnage de 48kHz, la résolution habituellement utilisée pour les enregistrements scientifiques. Il ne nous a pas été possible d'identifier dans le code disponible et dans la documentation à quel moment et de quelle façon ce signal sonore est converti en spectrogramme, qui est ensuite traité comme une image par l'algorithme de vision par ordinateur. Ainsi, nous ne sommes pas parvenus à appliquer directement un algorithme comme Grad-CAM pour produire des cartes d'activation de classe, et il nous faudra pousser davantage nos recherches afin de comprendre quels outils pourront permettre une interprétabilité des prédictions aussi détaillée et immédiate.

Récemment, ce sont les représentations d'attributs (*feature embeddings*) qui attirent l'attention des écologistes cherchant à mieux comprendre les représentations internes et cachées qu'un réseau neuronal profond se fait d'un signal audio (Arora et al. 2019, Sainburg et al. 2020, Ghani et al. 2023, McGinn et al. 2023). Les représentations d'attributs consistent en des vecteurs décrivant les caractéristiques du fichier audio que l'algorithme a reçu en entrée, ce qui rejoint en partie la fonction des cartes d'activation de classe. McGinn et al. (2023) utilisent cette méthode pour différencier différents types de sons associés à une même espèce, ce qui pourrait avoir un intérêt certain pour étudier la variation intraspécifique dans les chants, mais ne remplace pas entièrement les cartes d'activation de classe pour l'interprétation détaillée et visuelle d'une prédiction individuelle.

Entraînement à la reconnaissance d'individus et de groupes culturels

On s'est ici longuement intéressés aux biais et aux difficultés qu'apporte la variation culturelle et individuelle au moment d'automatiser la classification acoustique d'espèces, mais elle peut aussi être mise au centre d'une démarche de recherche. Puisqu'on sait que la variation intraspécifique dans les chants est basée sur des caractéristiques physiques constantes et mesurables, la tâche de la reconnaissance d'individus et de sous-groupes au sein d'une même espèce est tout à fait désignée pour un réseau neuronal profond, étant simple mais extrêmement fastidieuse et répétitive pour les humains. Par ailleurs, les bruants des prés sont une espèce idéale avec laquelle explorer l'automatisation de l'identification acoustique individuelle en raison de leurs patrons très clairs d'individualité vocale et de leurs microdialectes bien définis.

Un réseau neuronal capable de distinguer des individus ou des groupes de la même espèce constituerait un outil très utile pour une variété d'études écologiques. On peut par exemple imaginer que des travaux portant sur la transmission culturelle et son importance pour différents aspects de la conservation d'une population (Brakes et al. 2021) puisse bénéficier d'outils permettant d'associer rapidement des données sonores au groupe culturel dont elles sont issues. De la même manière, une capacité de reconnaissance individuelle automatisée, associée à des grilles de microphones (Blumstein et al. 2011, Mennill et al. 2012, Ethier et Wilson 2020), pourrait faciliter des études comportementales détaillées (Kirschel et al. 2009, Valletta et al. 2017) et permettre le suivi non-invasif d'individus sur de longues périodes ou sur de vastes territoires (Terry et al. 2005, Blumstein et al. 2011, Mennill 2011). Finalement, la reconnaissance d'individus pourrait aider à raffiner et à accélérer les estimations démographiques basées sur des relevés acoustiques (Marques et al. 2013, Hensel et al. 2022b).

Très peu d'outils d'apprentissage profond existent pour faire le suivi de phénomènes culturels chez les animaux (Youngblood and Lahti 2022, Wang et al. 2022). Stowell (2018 et 2019) représente une exception, car très peu d'études sur les oiseaux ont tenté le défi de l'automatisation de la reconnaissance individuelle acoustique avec des algorithmes d'apprentissage profond, probablement en raison de la difficulté d'obtenir des jeux de données appropriés. En effet, la principale difficulté rencontrée lors du développement d'un réseau

neuronal capable de classifier la variation intra populationnelle dans les chants d'oiseaux est qu'une quantité considérable de données étiquetées par des experts est essentielle. Pour apprendre à identifier correctement des catégories de types vocaux, le modèle doit être exposé à un grand nombre d'exemples représentant la variation existante ainsi que les bruits de fond sur lesquels les sons d'intérêt seront superposés (Priyadarshani et al. 2018). De plus, il y a lieu de croire que plus les différences entre les classes sont subtiles, comme l'est souvent la variation existant entre les individus ou les groupes culturels, plus le jeu de données d'entraînement doit être important afin d'atteindre une performance satisfaisante, et plus la tâche d'annotation manuelle est délicate.

Or, le jeu de données considérable que nous avons recueilli auprès des bruants des prés de l'île Kent et annoté au niveau de l'individu dans le cadre de la présente étude est probablement suffisant en taille et en précision pour entraîner un modèle qui puisse relever le défi de la classification intraspécifique. Une étude particulièrement intéressante qui pourrait être menée serait d'entraîner avec notre jeu de données quatre algorithmes explicitement mis à la disposition de la communauté pour être adaptés à différents projets de recherche : BirdNET (BirdNET-Analyzer, 2023), Koogu (Koogu, 2023), OpenSoundscape (OpenSoundscape, 2023) et pykanto (Recalde 2023). Il serait instructif de comparer leur performance et leurs exigences en termes de puissance de calcul, afin d'informer et d'orienter les biologistes qui souhaitent en faire usage sur leurs propres jeux de données. Dans chaque cas, nous pourrions entraîner deux modèles : l'un qui classifie les chants au niveau du microdialecte, et l'autre destiné à reconnaître les 24 individus pour lesquels nous avons annoté plus de 2250 chants. Il serait par ailleurs crucial de tester la performance de chaque modèle avec différentes quantités de données d'entraînement par classe, afin d'établir des lignes directrices quant à la taille minimale des jeux de données nécessaires et ainsi de minimiser dans la mesure du possible la tâche d'annotation manuelle requise pour les travaux futurs.

Il sera fascinant d'explorer le potentiel des réseaux neuronaux pour la reconnaissance acoustique de variants culturels et d'individus et de développer des façons d'améliorer l'accessibilité de ces méthodes pour des projets de recherches variés. Bien que cela implique un processus d'étiquetage parfois fastidieux, le potentiel d'un modèle entraîné sur un jeu de données

substantiel peut être énorme et permettre des études détaillées et étendues. Il importe de rendre ces outils accessibles, abordables et crédibles pour l'ensemble de la communauté scientifique œuvrant en écologie et en conservation.

On comprendra donc que les réseaux neuronaux profonds ont un potentiel énorme, mais qu'ils amènent la possibilité de plusieurs formes de biais qui peuvent facilement passer inaperçus. Il faut souligner encore l'importance pour les chercheurs et les décideurs de pouvoir bien comprendre les outils qu'ils mettent en œuvre, ou du moins d'avoir la possibilité d'interpréter leurs résultats, afin d'obtenir des données scientifiques fiables et reproductibles et d'être en mesure de prendre des décisions véritablement éclairées quant à la conservation de la biodiversité. Espérons que nos travaux auront su favoriser une prise de conscience des limites et des possibilités des réseaux neuronaux profonds en bioacoustique, et qu'ils inspireront davantage de recherches visant à mieux les comprendre et mieux les utiliser.

Références bibliographiques

- Alcocer, I., Lima, H., Sugai, L. S. M., & Llusia, D. (2022). Acoustic indices as proxies for biodiversity: A meta-analysis. *Biological Reviews*, Early view(n/a). <https://doi.org/10.1111/brv.12890>
- Aplin, L. M. (2019). Culture and cultural evolution in birds: A review of the evidence. *Animal Behaviour*, 147, 179–187. <https://doi.org/10.1016/j.anbehav.2018.05.001>
- Arif, M., Hedley, R., & Bayne, E. (2020, July 31). *Testing the Accuracy of a birdNET, Automatic bird song Classifier*. ERA. <https://doi.org/10.7939/r3-6kxb-kz18>
- Arora, V., Sun, M., & Wang, C. (2019). Deep Embeddings for Rare Audio Event Detection with Imbalanced Data. *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 3297–3301. <https://doi.org/10.1109/ICASSP.2019.8682395>
- AudioSet. (n.d.). Retrieved August 2, 2023, from <https://research.google.com/audioset/>
- Baker, M. C., & Cunningham, M. A. (1985). The Biology of Bird-Song Dialects. *Behavioral and Brain Sciences*, 8(1), 85–100. <https://doi.org/10.1017/S0140525X00019750>
- Beyer, R. M., & Manica, A. (2020). Historical and projected future range sizes of the world's mammals, birds, and amphibians. *Nature Communications*, 11(1), Article 1. <https://doi.org/10.1038/s41467-020-19455-9>
- Biodiversité Québec | La boussole du suivi de la nature au Québec. (n.d.). Retrieved August 6, 2023, from <https://biodiversite-quebec.ca>
- BirdWeather. (n.d.). BirdWeather. Retrieved August 2, 2023, from <https://app.birdweather.com/>
- Bluff, L. A., Kacelnik, A., & Rutz, C. (2010). Vocal culture in New Caledonian crows *Corvus moneduloides*. *Biological Journal of the Linnean Society*, 101(4), 767–776. <https://doi.org/10.1111/j.1095-8312.2010.01527.x>
- Blumstein, D. T., Mennill, D. J., Clemins, P., Girod, L., Yao, K., Patricelli, G., Deppe, J. L., Krakauer, A. H., Clark, C., Cortopassi, K. A., Hanser, S. F., McCowan, B., Ali, A. M., & Kirschel, A. N. G.

- (2011). Acoustic monitoring in terrestrial environments using microphone arrays: Applications, technological considerations and prospectus. *Journal of Applied Ecology*, 48(3), 758–767. <https://doi.org/10.1111/j.1365-2664.2011.01993.x>
- Bolcar, M. (2020). *A Comparison of Avian Communities in Managed Early Successional Edge Habitat and Residential Edge Habitat in the Middle Patuxent Environmental Area*. <https://doi.org/10.13016/m2iviu-44e1>
- Borowiec, M. L., Dikow, R. B., Frandsen, P. B., McKeeken, A., Valentini, G., & White, A. E. (2022). Deep learning as a tool for ecology and evolution. *Methods in Ecology and Evolution*, 13(8), 1640–1660. <https://doi.org/10.1111/2041-210X.13901>
- Borrer, D. J., & Gunn, W. W. H. (1965). Variation in White-Throated Sparrow Songs. *The Auk*, 82(1), 26–47. <https://doi.org/10.2307/4082793>
- Bradley, R. (1994). Cultural change and geographic variation in the songs of the Belding's Savannah Sparrow (*Passerculus sandwichensis beldingi*). *Bulletin of the Southern California Academy of Sciences*, 93(3), 91–109.
- Brakes, P., Carroll, E. L., Dall, S. R. X., Keith, S. A., McGregor, P. K., Mesnick, S. L., Noad, M. J., Rendell, L., Robbins, M. M., Rutz, C., Thornton, A., Whiten, A., Whiting, M. J., Aplin, L. M., Bearhop, S., Ciucci, P., Fishlock, V., Ford, J. K. B., Notarbartolo di Sciara, G., ... Garland, E. C. (2021). A deepening understanding of animal culture suggests lessons for conservation. *Proceedings of the Royal Society B: Biological Sciences*, 288(1949), 20202718. <https://doi.org/10.1098/rspb.2020.2718>
- Brumm, H., & Naguib, M. (2009). Chapter 1 Environmental Acoustics and the Evolution of Bird Song. In *Advances in the Study of Behavior* (Vol. 40, pp. 1–33). Academic Press. [https://doi.org/10.1016/S0065-3454\(09\)40001-9](https://doi.org/10.1016/S0065-3454(09)40001-9)
- Brunk, K. M., Gutiérrez, R. J., Peery, M. Z., Cansler, C. A., Kahl, S., & Wood, C. M. (2023). Quail on fire: Changing fire regimes may benefit mountain quail in fire-adapted forests. *Fire Ecology*, 19(1), 19. <https://doi.org/10.1186/s42408-023-00180-9>

- Byers, B. E., Belinsky, K. L., & Bentley, R. A. (2010). Independent Cultural Evolution of Two Song Traditions in the Chestnut-Sided Warbler. *The American Naturalist*, 176(4), 476–489. <https://doi.org/10.1086/656268>
- Catchpole, C. K., & Slater, P. J. B. (2008). *Bird Song. Biological Themes and Variations* (2nd ed.). Cambridge University Press.
- Christin, S., Hervet, É., & Lecomte, N. (2019). Applications for deep learning in ecology. *Methods in Ecology and Evolution*, 10(10), 1632–1644. <https://doi.org/10.1111/2041-210X.13256>
- Clark, M. L., Salas, L., Baligar, S., Quinn, C. A., Snyder, R. L., Leland, D., Schackwitz, W., Goetz, S. J., & Newsam, S. (2023). The effect of soundscape composition on bird vocalization classification in a citizen science biodiversity monitoring project. *Ecological Informatics*, 75, 102065. <https://doi.org/10.1016/j.ecoinf.2023.102065>
- Cole, J. S., Michel, N. L., Emerson, S. A., & Siegel, R. B. (2022). Automated bird sound classifications of long-duration recordings produce occupancy model outputs similar to manually annotated data. *Ornithological Applications*, 124(2), duac003. <https://doi.org/10.1093/ornithapp/duac003>
- Crates, R., Langmore, N., Ranjard, L., Stojanovic, D., Rayner, L., Ingwersen, D., & Heinsohn, R. (2021). Loss of vocal culture and fitness costs in a critically endangered songbird. *Proceedings of the Royal Society B: Biological Sciences*, 288(1947), 20210225. <https://doi.org/10.1098/rspb.2021.0225>
- Czocherová, I., Rubáčová, L., Petrussek, A., & Petrusková, T. (2022). Contrasting patterns of geographical song variation in two closely related passerine species with a simple song. *Journal of Ornithology*, 163(1), 93–104. <https://doi.org/10.1007/s10336-021-01924-7>
- Darras, K., Batáry, P., Furnas, B. J., Grass, I., Mulyani, Y. A., & Tschardt, T. (2019). Autonomous sound recording outperforms human observation for sampling birds: A systematic map and user guide. *Ecological Applications*, 29(6), e01954. <https://doi.org/10.1002/eap.1954>
- Déclaration de Montréal IA responsable*. (2018). *Déclaration de Montréal IA responsable*. <https://declarationmontreal-iaresponsable.com/>

- Derryberry, E. P. (2009). Ecology Shapes Birdsong Evolution: Variation in Morphology and Habitat Explains Variation in White-Crowned Sparrow Song. *The American Naturalist*, 174(1), 24–33. <https://doi.org/10.1086/599298>
- Digby, A., Towsey, M., Bell, B. D., & Teal, P. D. (2013). A practical comparison of manual and autonomous methods for acoustic monitoring. *Methods in Ecology and Evolution*, 4(7), 675–683. <https://doi.org/10.1111/2041-210X.12060>
- Dixon, C. L. (1978). Breeding Biology of the Savannah Sparrow on Kent Island. *The Auk*, 95(2), 235–246. <https://doi.org/10.1093/auk/95.2.235>
- eBird—Discover a new world of birding... (n.d.). Retrieved August 2, 2023, from <https://ebird.org/home>
- Emlen, J. T., & DeJong, M. J. (1992). Counting Birds: The Problem of Variable Hearing Abilities (Contando Aves: El Problema de la Variabilidad en la Capacidad Auditiva). *Journal of Field Ornithology*, 63(1), 26–31.
- Ethier, J. P., & Wilson, D. R. (2020). Using microphone arrays to investigate microhabitat selection by declining breeding birds. *Ibis*, 162(3), 873–884. <https://doi.org/10.1111/ibi.12785>
- Folliot, A., Hauptert, S., Ducretet, M., Sèbe, F., & Sueur, J. (2022). Using acoustics and artificial intelligence to monitor pollination by insects and tree use by woodpeckers. *Science of The Total Environment*, 838, 155883. <https://doi.org/10.1016/j.scitotenv.2022.155883>
- Ghani, B., Klinck, H., & Hallerberg, S. (2023). Classification of group-specific variations in songs within House Wren species using machine learning models. *Ecological Informatics*, 74, 101946. <https://doi.org/10.1016/j.ecoinf.2022.101946>
- Ghosh, S., & Dasgupta, R. (2022). Study of Animal Behavior and Machine Learning. In S. Ghosh & R. Dasgupta (Eds.), *Machine Learning in Biological Sciences: Updates and Future Prospects* (pp. 231–237). Springer Nature. https://doi.org/10.1007/978-981-16-8881-2_27
- Gibb, R., Browning, E., Glover-Kapfer, P., & Jones, K. E. (2019). Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods in Ecology and Evolution*, 10(2), 169–185. <https://doi.org/10.1111/2041-210X.13101>

- Gruber, T., Luncz, L., Mörchen, J., Schuppli, C., Kendal, R. L., & Hockings, K. (2019). Cultural change in animals: A flexible behavioural adaptation to human disturbance. *Palgrave Communications*, 5(1), Article 1. <https://doi.org/10.1057/s41599-019-0271-4>
- HAIKUBOX. (n.d.). Retrieved August 2, 2023, from <https://haikubox.com/>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep Residual Learning for Image Recognition*. 770–778. https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html
- Hensel, A. L. J., Dobney, S. L., Doucet, S. M., Ryan Norris, D., Newman, A. E. M., Williams, H., & Mennill, D. J. (2022). Microgeographical variation in birdsong: Savannah Sparrows exhibit microdialects in an island population. *Animal Behaviour*, 188, 119–131. <https://doi.org/10.1016/j.anbehav.2022.04.008>
- Hensel, A. L. J., Dobney, S. L., Moran, I. G., Thomas, I. P., Burant, J. B., Woodworth, B. K., Doucet, S. M., Newman, A. E. M., Norris, D. R., Williams, H., & Mennill, D. J. (2022). Passive acoustic monitoring provides predictable and reliable underestimates of population size and longevity in wild Savannah Sparrows. *Ornithological Applications*, 124(3), duac018. <https://doi.org/10.1093/ornithapp/duac018>
- Hothorn, T., Bretz, F., Westfall, P., Heiberger, R. M., Schuetzenmeister, A., & Scheibe, S. (2023). *multcomp: Simultaneous Inference in General Parametric Models* (1.4-25) [Computer software]. <https://cran.r-project.org/web/packages/multcomp/index.html>
- Hughes, A. C., Orr, M. C., Ma, K., Costello, M. J., Waller, J., Provoost, P., Yang, Q., Zhu, C., & Qiao, H. (2021). Sampling biases shape our view of the natural world. *Ecography*, 44(9), 1259–1269. <https://doi.org/10.1111/ecog.05926>
- Humphries, G., Magness, D. R., & Huettmann, F. (Eds.). (2018). *Machine Learning for Ecology and Sustainable Natural Resource Management*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-96978-7>

- Hutto, R. L., & Stutzman, R. J. (2009). Humans versus autonomous recording units: A comparison of point-count results. *Journal of Field Ornithology*, 80(4), 387–398. <https://doi.org/10.1111/j.1557-9263.2009.00245.x>
- J Burt. (2000). *Syrinx-PC* (2.6L) [Computer software].
- Jäckel, D., Mortega, K. G., Brockmeyer, U., Lehmann, G. U. C., & Voigt-Heucke, S. L. (2022). Unravelling the Stability of Nightingale Song Over Time and Space Using Open, Citizen Science and Shared Data. *Frontiers in Ecology and Evolution*, 10. <https://doi.org/10.3389/fevo.2022.778610>
- Jäckel, D., Mortega, K. G., Sturm, U., Brockmeyer, U., Khorramshahi, O., & Voigt-Heucke, S. L. (2021). Opportunities and limitations: A comparative analysis of citizen science and expert recordings for bioacoustic research. *PLOS ONE*, 16(6), e0253763. <https://doi.org/10.1371/journal.pone.0253763>
- Janes, S. W., & Ryker, L. (2013). Rapid Change in a Type I Song Dialect of Hermit Warblers (Setophaga occidentalis). *The Auk*, 130(1), 30–35. <https://doi.org/10.1525/auk.2012.11273>
- Jouventin, P., Aubin, T., & Lengagne, T. (1999). Finding a parent in a king penguin colony: The acoustic system of individual recognition. *Animal Behaviour*, 57(6), 1175–1183. <https://doi.org/10.1006/anbe.1999.1086>
- Kahl, S. (2020). *Identifying birds by sound: Large-scale acoustic event recognition for avian activity monitoring*. Universitätsverlag Chemnitz.
- Kahl, S. (2023). *BirdNET-Analyzer* [Python]. <https://github.com/kahst/BirdNET-Analyzer> (Original work published 2021)
- Kahl, S., Wood, C. M., Eibl, M., & Klinck, H. (2021). BirdNET: A deep learning solution for avian diversity monitoring. *Ecological Informatics*, 61, 101236. <https://doi.org/10.1016/j.ecoinf.2021.101236>
- Keith, S. A., & Bull, J. W. (2017). Animal culture impacts species' capacity to realise climate-driven range shifts. *Ecography*, 40(2), 296–304. <https://doi.org/10.1111/ecog.02481>

- Kirschel, A. N. G., Earl, D. A., Yao, Y., Escobar, I. A., Vilches, E., Vallejo, E. E., & Taylor, C. E. (2009). Using Songs to Identify Individual Mexican Antthrush *Formicarius Moniliger*: Comparison of Four Classification Methods. *Bioacoustics*, *19*(1–2), 1–20. <https://doi.org/10.1080/09524622.2009.9753612>
- Krause, B., & Farina, A. (2016). Using ecoacoustic methods to survey the impacts of climate change on biodiversity. *Biological Conservation*, *195*, 245–254. <https://doi.org/10.1016/j.biocon.2016.01.013>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, *25*. <https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html>
- Kumar, Y., Gupta, S., & Singh, W. (2022). A novel deep transfer learning models for recognition of birds sounds in different environment. *Soft Computing*. <https://doi.org/10.1007/s00500-021-06640-1>
- Lapeyrolerie, M., Chapman, M. S., Norman, K. E. A., & Boettiger, C. (2022). Deep reinforcement learning for conservation decisions. *Methods in Ecology and Evolution*, *13*(11), 2649–2662. <https://doi.org/10.1111/2041-210X.13954>
- Lenth, R. V., Bolker, B., Buerkner, P., Giné-Vázquez, I., Herve, M., Jung, M., Love, J., Miguez, F., Riebl, H., & Singmann, H. (2023). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (1.8.7) [Computer software]. <https://cran.r-project.org/web/packages/emmeans/index.html>
- Lind, H., Dabelsteen, T., & McGREGOR, P. K. (1996). Female great tits can identify mates by song. *Animal Behaviour*, *52*(4), 667–671. <https://doi.org/10.1006/anbe.1996.0211>
- Logue, D. M., & Leca, J.-B. (2020). Animal Culture: How a New Birdsong Went Viral. *Current Biology*, *30*(16), R957–R959. <https://doi.org/10.1016/j.cub.2020.06.055>
- Luccioni, A. S., & Rolnick, D. (2022). *Bugs in the Data: How ImageNet Misrepresents Biodiversity* (arXiv:2208.11695). arXiv. <http://arxiv.org/abs/2208.11695>

- Macaulay Library. (2022, September 27). <https://www.macaulaylibrary.org/>
- Madhusudhana, S. (2023). *Koogu* [Python]. <https://github.com/shyamblast/Koogu> (Original work published 2020)
- Maegawa, Y., Ushigome, Y., Suzuki, M., Taguchi, K., Kobayashi, K., Haga, C., & Matsui, T. (2021). A new survey method using convolutional neural networks for automatic classification of bird calls. *Ecological Informatics*, *61*, 101164. <https://doi.org/10.1016/j.ecoinf.2020.101164>
- Malamut, E. J. (2022). *Using Autonomous Recording Units and Image Processing to Investigate Patterns in Avian Singing Activity and Nesting Phenology* [M.S.]. <https://www.proquest.com/docview/2676140480/abstract/1AA44629EBEA44F1PQ/1>
- Marler, P. (1960). Bird songs and mate selection. *Animal Sounds and Communication*, 348–367.
- Marques, T. A., Thomas, L., Martin, S. W., Mellinger, D. K., Ward, J. A., Moretti, D. J., Harris, D., & Tyack, P. L. (2013). Estimating animal population density using passive acoustics. *Biological Reviews*, *88*(2), 287–309. <https://doi.org/10.1111/brv.12001>
- Martin, L. J., Blossey, B., & Ellis, E. (2012). Mapping where ecologists work: Biases in the global distribution of terrestrial ecological observations. *Frontiers in Ecology and the Environment*, *10*(4), 195–201. <https://doi.org/10.1890/110154>
- Mauck, R. A., Pratte, I., Hedd, A., Pollet, I. I., Jones, P. L., Montevecchi, W. A., Ronconi, R. A., Gjerdrum, C., Adrianowycz, S., McMahon, C., Acker, H., Taylor, L. U., McMahon, J., Dearborn, D. C., Robertson, G. J., & McFarlane Tranquilla, L. A. (2023). Female and male Leach’s Storm Petrels (*Hydrobates leucorhous*) pursue different foraging strategies during the incubation period. *Ibis*, *165*(1), 161–178. <https://doi.org/10.1111/ibi.13112>
- McGinn, K., Kahl, S., Peery, M. Z., Klinck, H., & Wood, C. M. (2023). Feature embeddings from the BirdNET algorithm provide insights into avian ecology. *Ecological Informatics*, *74*, 101995. <https://doi.org/10.1016/j.ecoinf.2023.101995>

- McGregor, P. K., Holland, J., & Shepherd, M. (1997). The ecology of corn bunting *Miliaria calandra* song dialects and their potential use in conservation. *The Ecology and Conservation of Corn Buntings Miliaria Calandra*, 13, 76–87.
- Mennill, D. J. (2011). Individual distinctiveness in avian vocalizations and the spatial monitoring of behaviour: Spatial monitoring of vocal behaviour. *Ibis*, 153(2), 235–238. <https://doi.org/10.1111/j.1474-919X.2011.01119.x>
- Mennill, D. J., Battiston, M., Wilson, D. R., Foote, J. R., & Doucet, S. M. (2012). Field test of an affordable, portable, wireless microphone array for spatial monitoring of animal ecology and behaviour. *Methods in Ecology and Evolution*, 3(4), 704–712. <https://doi.org/10.1111/j.2041-210X.2012.00209.x>
- Mennill, D. J., Doucet, S. M., Newman, A. E. M., Williams, H., Moran, I. G., Thomas, I. P., Woodworth, B. K., & Norris, D. R. (2018). Wild Birds Learn Songs from Experimental Vocal Tutors. *Current Biology*, 28(20), 3273–3278.e4. <https://doi.org/10.1016/j.cub.2018.08.011>
- Moran, I. G., Lukianchuk, K. C., Doucet, Stéphanie. M., Newman, A. E. M., Williams, H., Norris, D. R., & Mennill, D. J. (2019). Diel and seasonal patterns of variation in the singing behaviour of Savannah Sparrows (*Passerculus sandwichensis*). *Avian Research*, 10(1), 26. <https://doi.org/10.1186/s40657-019-0165-6>
- Nelson, D. (1998). Geographic Variation in Song of Gambel's White-Crowned Sparrow. *Behaviour*, 135(3), 321–342. <https://doi.org/10.1163/156853998793066258>
- Nelson, D. A., & Marler, P. (1994). Selection-based learning in bird song development. *Proceedings of the National Academy of Sciences*, 91(22), 10498–10501. <https://doi.org/10.1073/pnas.91.22.10498>
- Nelson, D. A., & Soha, J. A. (2004). Perception of geographical variation in song by male Puget Sound white-crowned sparrows, *Zonotrichia leucophrys pugetensis*. *Animal Behaviour*, 68(2), 395–405. <https://doi.org/10.1016/j.anbehav.2003.08.027>
- OpenSoundscape—Opensoundscape 0.9.1 documentation*. (n.d.). Retrieved August 2, 2023, from <http://opensoundscape.org/en/latest/index.html>

- Otter, K. A., Mckenna, A., LaZerte, S. E., & Ramsay, S. M. (2020). Continent-wide Shifts in Song Dialects of White-Throated Sparrows. *Current Biology*, 30(16), 3231-3235.e3. <https://doi.org/10.1016/j.cub.2020.05.084>
- Pandolfino, E. R., & Douglas, L. A. (2022). Regional song dialects of the Ruby-crowned Kinglet. *Journal of Field Ornithology*, 93(2), art6. <https://doi.org/10.5751/JFO-00120-930206>
- Pankratz, R., Hache, S., Sólymos, P., & Bayne, E. (2017). Potential benefits of augmenting road-based breeding bird surveys with autonomous recordings. *Avian Conservation and Ecology*, 12(2). <https://doi.org/10.5751/ACE-01087-120218>
- Parmesan, C. (2006). Ecological and Evolutionary Responses to Recent Climate Change. *Annual Review of Ecology, Evolution, and Systematics*, 37, 637–669.
- Partners in Flight Databases – Avian Conservation Assessment and Population Estimates Databases*. (n.d.). Retrieved August 2, 2023, from <https://pif.birdconservancy.org/>
- Pérez-Granados, C. (2023a). A First Assessment of Birdnet Performance at Varying Distances: A Playback Experiment. *Ardeola*, 70(2), 221–233. <https://doi.org/10.13157/arla.70.2.2023.sc1>
- Pérez-Granados, C. (2023b). BirdNET: Applications, performance, pitfalls and future opportunities. *Ibis*, 165(3). <https://doi.org/10.1111/ibi.13193>
- Pichler, M., & Hartig, F. (2023). Machine learning and deep learning—A review for ecologists. *Methods in Ecology and Evolution*, n/a(n/a). <https://doi.org/10.1111/2041-210X.14061>
- Priyadarshani, N., Marsland, S., & Castro, I. (2018). Automated birdsong recognition in complex acoustic environments: A review. *Journal of Avian Biology*, 49(5), jav-01447. <https://doi.org/10.1111/jav.01447>
- R Core Team. (2023a). *R* (4.3.1) [Computer software].
- R Core Team. (2023b). *Stats-package {stats}* (4.3.1) [Computer software].
- Recalde, N. M. (2023). *pykanto: A python library to accelerate research on wild bird song*.

- Robert, J. (2021). *pydub: Manipulate audio with an simple and easy high level interface* (0.25.1) [Python; OS Independent]. <http://pydub.com>
- Rodríguez-Fuentes, J., Cadena, C. D., & Avendaño, J. E. (2021). Vocal behavior and microgeographic variation in song types in a Neotropical songbird. *Journal of Ornithology*. <https://doi.org/10.1007/s10336-021-01930-9>
- Rosenstock, S. S., Anderson, D. R., Giesen, K. M., Leukering, T., & Carter, M. F. (2002). Landbird Counting Techniques: Current Practices and an Alternative. *The Auk*, *119*(1), 46–53. <https://doi.org/10.1093/auk/119.1.46>
- RStudio Team. (2023). *RStudio* (2023.06.1+524) [Computer software].
- Ryan, S. J. (2006). The Role of Culture in Conservation Planning for Small or Endangered Populations. *Conservation Biology*, *20*(4), 1321–1324.
- Schuppli, C., & Schaik, C. P. van. (2019). Animal cultures: How we've only seen the tip of the iceberg. *Evolutionary Human Sciences*, *1*, e2. <https://doi.org/10.1017/ehs.2019.1>
- Sebastianelli, M., Lukhele, S. M., Nwankwo, E. C., Hadjioannou, L., & Kirschel, A. N. G. (2022). Continent-wide patterns of song variation predicted by classical rules of biogeography. *Ecology Letters*, *n/a*(*n/a*). <https://doi.org/10.1111/ele.14102>
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2020). Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *International Journal of Computer Vision*, *128*(2), 336–359. <https://doi.org/10.1007/s11263-019-01228-7>
- Shonfield, J., & Bayne, E. (2017). Autonomous recording units in avian ecological research: Current use and future applications. *Avian Conservation and Ecology*, *12*(1). <https://doi.org/10.5751/ACE-00974-120114>
- Simonyan, K., & Zisserman, A. (2015). *Very Deep Convolutional Networks for Large-Scale Image Recognition* (arXiv:1409.1556). arXiv. <https://doi.org/10.48550/arXiv.1409.1556>

- Slabbekoorn, H., & Smith, T. B. (2002). Bird song, ecology and speciation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 357(1420), 493–503. <https://doi.org/10.1098/rstb.2001.1056>
- Slade, J. W. G., Watson, M. J., & MacDougall-Shackleton, E. A. (2017). Birdsong signals individual diversity at the major histocompatibility complex. *Biology Letters*, 13(11), 20170430. <https://doi.org/10.1098/rsbl.2017.0430>
- Stowell, D. (2022). Computational bioacoustics with deep learning: A review and roadmap. *PeerJ*, 10, e13152. <https://doi.org/10.7717/peerj.13152>
- Stowell, D., Petrusková, T., Šálek, M., & Linhart, P. (2018). Automatic acoustic identification of individual animals: Improving generalisation across species and recording conditions. *ArXiv:1810.09273 [Cs, Eess]*. <http://arxiv.org/abs/1810.09273>
- Stowell, D., Petrusková, T., Šálek, M., & Linhart, P. (2019). Automatic acoustic identification of individuals in multiple species: Improving identification across recording conditions. *Journal of The Royal Society Interface*, 16(153), 20180940. <https://doi.org/10.1098/rsif.2018.0940>
- Sugai, L. S. M., Silva, T. S. F., Ribeiro, J. W., & Llusia, D. (2019). Terrestrial Passive Acoustic Monitoring: Review and Perspectives. *BioScience*, 69(1), 15–25. <https://doi.org/10.1093/biosci/biy147>
- Sung, H.-C. S.-C., & Handford, P. H. (2006). Songs of the Savannah Sparrow: Structure and geographic variation. *Canadian Journal of Zoology*, 84(11), 1637–1646. <https://doi.org/10.1139/z06-159>
- Taylor, L. U., Woodworth, B. K., Sandercock, B. K., & Wheelwright, N. T. (2018). Demographic drivers of collapse in an island population of Tree Swallows. *The Condor*, 120(4), 828–841. <https://doi.org/10.1650/CONDOR-18-75.1>
- Terry, A. M., Peake, T. M., & McGregor, P. K. (2005). The role of vocal individuality in conservation. *Frontiers in Zoology*, 2(1), 10. <https://doi.org/10.1186/1742-9994-2-10>

- The IUCN Red List of Threatened Species*. (n.d.). IUCN Red List of Threatened Species. Retrieved August 2, 2023, from <https://www.iucnredlist.org/en>
- The Pandas Development Team. (2022). *pandas: Powerful data structures for data analysis, time series, and statistics* (1.4.2) [Cython, Python; OS Independent].
- Thomas, I. P., Doucet, S. M., Norris, D. R., Newman, A. E. M., Williams, H., & Mennill, D. J. (2021). Vocal learning in Savannah Sparrows: Acoustic similarity to neighbours shapes song development and territorial aggression. *Animal Behaviour*, *176*, 77–86. <https://doi.org/10.1016/j.anbehav.2021.03.015>
- Toenies, M., & Rich, L. (2021). Advancing bird survey efforts through novel recorder technology and automated species identification. *California Fish and Wildlife Journal*, *107*(2), 56–70. <https://doi.org/10.51492/cfwj.107.5>
- Tuia, D., Kellenberger, B., Beery, S., Costelloe, B. R., Zuffi, S., Risse, B., Mathis, A., Mathis, M. W., van Langevelde, F., Burghardt, T., Kays, R., Klinck, H., Wikelski, M., Couzin, I. D., van Horn, G., Crofoot, M. C., Stewart, C. V., & Berger-Wolf, T. (2022). Perspectives in machine learning for wildlife conservation. *Nature Communications*, *13*(1), Article 1. <https://doi.org/10.1038/s41467-022-27980-y>
- Valletta, J. J., Torney, C., Kings, M., Thornton, A., & Madden, J. (2017). Applications of machine learning in animal behaviour studies. *Animal Behaviour*, *124*, 203–220. <https://doi.org/10.1016/j.anbehav.2016.12.005>
- Vellinga, W.-P. (2023). *Xeno-canto—Bird sounds from around the world* [dataset]. Xeno-canto Foundation for Nature Sounds. <https://doi.org/10.15468/QV0KSN>
- Venier, L., Mazerolle, M., Rodgers, A., McIlwrick, K., Holmes, S., & Thompson, D. (2017). Comparison of semiautomated bird song recognition with manual detection of recorded bird song samples. *Avian Conservation and Ecology*, *12*(2). <https://doi.org/10.5751/ACE-01029-120202>

- Verreycken, E., Simon, R., Quirk-Royal, B., Daems, W., Barber, J., & Steckel, J. (2021). Bio-acoustic tracking and localization using heterogeneous, scalable microphone arrays. *Communications Biology*, 4(1), Article 1. <https://doi.org/10.1038/s42003-021-02746-2>
- Wang, D., Forstmeier, W., Farine, D. R., Maldonado-Chaparro, A. A., Martin, K., Pei, Y., Alarcón-Nieto, G., Klarevas-Irby, J. A., Ma, S., Aplin, L. M., & Kempenaers, B. (2022). Machine learning reveals cryptic dialects that explain mate choice in a songbird. *Nature Communications*, 13(1), 1630. <https://doi.org/10.1038/s41467-022-28881-w>
- Wheelwright, N. T., & Mauck, R. A. (1998). Philopatry, Natal Dispersal, and Inbreeding Avoidance in an Island Population of Savannah Sparrows. *Ecology*, 79(3), 755–767. [https://doi.org/10.1890/0012-9658\(1998\)079\[0755:PNDAlA\]2.0.CO;2](https://doi.org/10.1890/0012-9658(1998)079[0755:PNDAlA]2.0.CO;2)
- Wheelwright, N. T., Schultz, C. B., & Hodum, P. J. (1992). Polygyny and male parental care in Savannah Sparrows: Effects on female fitness. *Behavioral Ecology and Sociobiology*, 31(4), 279–289. <https://doi.org/10.1007/BF00171683>
- Wheelwright, N. T., Swett, M. B., Levin, I. I., Kroodsma, D. E., Freeman-Gallant, C. R., & Williams, H. (2008). The influence of different tutor types on song learning in a natural bird population. *Animal Behaviour*, 75(4), 1479–1493. <https://doi.org/10.1016/j.anbehav.2007.08.030>
- Whiten, A. (2017). A second inheritance system: The extension of biology through culture. *Interface Focus*, 7(5), 20160142. <https://doi.org/10.1098/rsfs.2016.0142>
- Whiten, A. (2019). Cultural Evolution in Animals. *Annual Review of Ecology, Evolution, and Systematics*, 50(1), 27–48. <https://doi.org/10.1146/annurev-ecolsys-110218-025040>
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* (3.4.2) [Computer software].
- Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *dplyr: A Grammar of Data Manipulation* (1.1.2) [Computer software].
- Widmer, G., & Kubat, M. (1996). Learning in the presence of concept drift and hidden contexts. *Machine Learning*, 23(1), 69–101. <https://doi.org/10.1007/BF00116900>

- Williams, H. (2021). Mechanisms of Cultural Evolution in the Songs of Wild Bird Populations. *Frontiers in Psychology*, 12, 1047. <https://doi.org/10.3389/fpsyg.2021.643343>
- Williams, H., Levin, I. I., Norris, D. R., Newman, A. E. M., & Wheelwright, N. T. (2013). Three decades of cultural evolution in Savannah Sparrow songs. *Animal Behaviour*, 85(1), 213–223. <https://doi.org/10.1016/j.anbehav.2012.10.028>
- Williams, H., Robins, C. W., Norris, D. R., Newman, A. E. M., Freeman-Gallant, C. R., Wheelwright, N. T., & Mennill, D. J. (2019). The buzz segment of Savannah Sparrow song is a population marker. *Journal of Ornithology*, 160(1), 217–227. <https://doi.org/10.1007/s10336-018-1611-7>
- Williams, H., Scharf, A., Ryba, A. R., Ryan Norris, D., Mennill, D. J., Newman, A. E. M., Doucet, S. M., & Blackwood, J. C. (2022). Cumulative cultural evolution and mechanisms for cultural selection in wild bird songs. *Nature Communications*, 13(1), Article 1. <https://doi.org/10.1038/s41467-022-31621-9>
- Wilson, D. R., & Mennill, D. J. (2010). Black-capped chickadees, *Poecile atricapillus*, can use individually distinctive songs to discriminate among conspecifics. *Animal Behaviour*, 79(6), 1267–1275. <https://doi.org/10.1016/j.anbehav.2010.02.028>
- Wood, C. M., Kahl, S., Rahaman, A., & Klinck, H. (2022). The machine learning–powered BirdNET App reduces barriers to global bird research by enabling citizen science participation. *PLOS Biology*, 20(6), e3001670. <https://doi.org/10.1371/journal.pbio.3001670>
- Woodworth, B. K., Wheelwright, N. T., Newman, A. E., Schaub, M., & Norris, D. R. (2017). Winter temperatures limit population growth rate of a migratory songbird. *Nature Communications*, 8(1), Article 1. <https://doi.org/10.1038/ncomms14812>
- Youngblood, M., & Lahti, D. C. (2022). Content bias in the cultural evolution of house finch song. *Animal Behaviour*, 185, 37–48. <https://doi.org/10.1016/j.anbehav.2021.12.012>
- Zeiler, M. D., & Fergus, R. (2014). Visualizing and Understanding Convolutional Networks. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *Computer Vision – ECCV 2014* (pp. 818–833). Springer International Publishing. https://doi.org/10.1007/978-3-319-10590-1_53

- Zhang, J., Huang, K., Cottman-Fields, M., Trusking, A., Roe, P., Duan, S., Dong, X., Towsey, M., & Wimmer, J. (2013). Managing and Analysing Big Audio Data for Environmental Monitoring. *2013 IEEE 16th International Conference on Computational Science and Engineering*, 997–1004. <https://doi.org/10.1109/CSE.2013.146>
- Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). *Learning Deep Features for Discriminative Localization*. 9.
- Zink, R. M., Rising, J. D., Mockford, S., Horn, A. G., Wright, J. M., Leonard, M., & Westberg, M. C. (2005). Mitochondrial DNA Variation, Species Limits, and Rapid Evolution of Plumage Coloration and Size in the Savannah Sparrow. *The Condor*, *107*(1), 21–28. <https://doi.org/10.1093/condor/107.1.21>
- Zoonomen Nomenclature Resource Page*. (n.d.). Retrieved August 2, 2023, from <http://www.zoonomen.net/>

Annexes

Annexe 1

Individual	Focal recordings	Song Meter recordings	Total
Two CH			
B.GY	271	2,023	2,294
G.GR	388	2,318	2,706
G.YB	226	2,092	2,318
N.NW	445	2,237	2,682
O.GN	198	2,205	2,403
O.SW	50	2,253	2,303
S.GN	410	2,032	2,442
SN.S	513	2,088	2,601
Others	3,430	3	3,433
Total	5,931	17,251	23,182
Two CH and dash			
NW.L	244	2,047	2,291
O.NL	213	2,101	2,314
O.SN	343	2,165	2,508
RB.L	109	2,238	2,347
RY.B	292	2,067	2,359
Y.YR	227	2,237	2,464
Y.YW	380	2,049	2,429
Others	2,013	281	2,294
Total	3,821	15,185	19,006
Stutter			
L.GL	704	2,178	2,882
SW.O	261	2,078	2,339
V.YO	1,037	2,067	3,104
Others	66	0	66
Total	2,068	6,323	8,391

Individual	Focal recordings	Song Meter recordings	Total
Quick notes			
G.WG	235	2,156	2,391
N.SY	997	2,017	3,014
Others	645	617	1,262
Total	1,877	4,790	6,667
Dash			
UB5	0	2,756	2,756
Y.RL	43	2,268	2,311
Total	43	5,024	5,067
One CH and dash			
G.ON	313	2,204	2,517
SL.R	232	2,202	2,434
Others	74	8	82
Total	619	4,414	5,033
Others			
Total	878	1,382	2,260
Grand Total	15,237	54,369	69,606

Tableau 2 Number of annotated recordings for individual Savannah Sparrows from Kent Island, New Brunswick, with microdialect and recording type.

Annexe 2

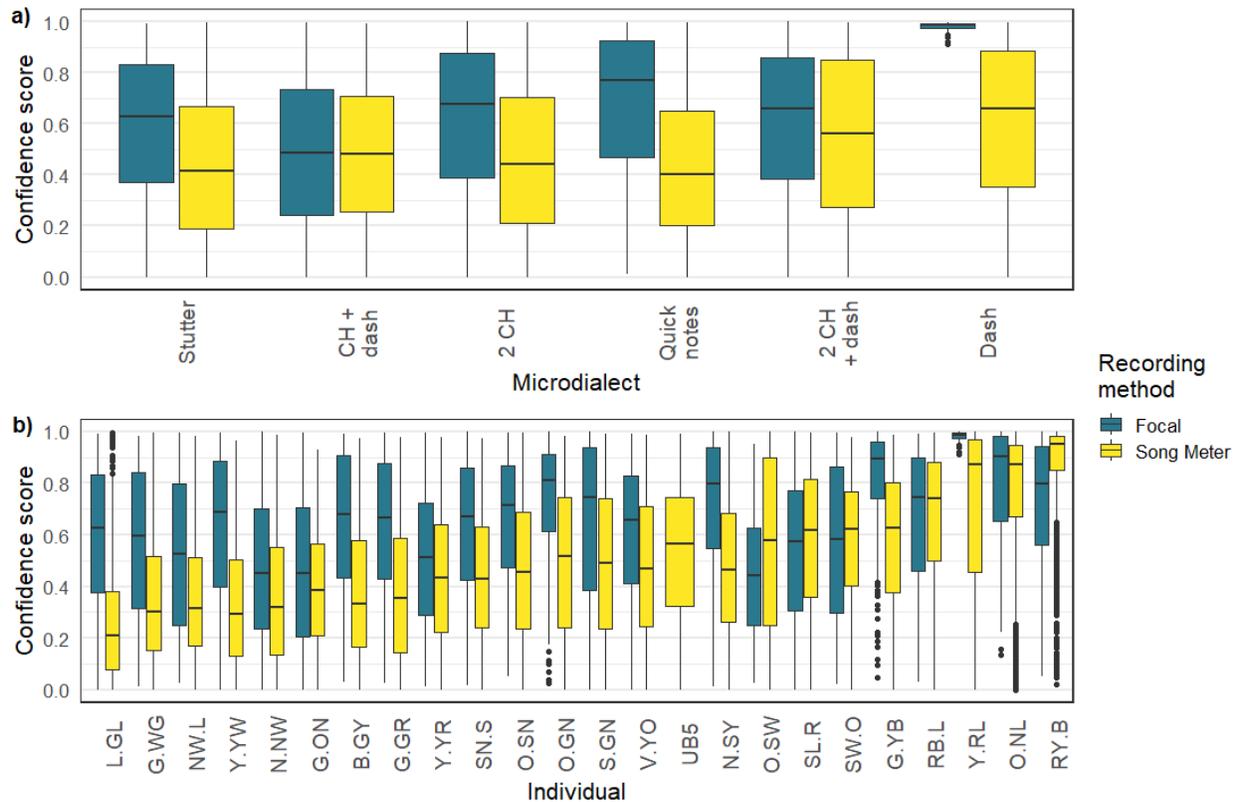


Figure 7 Distribution of confidence scores attributed to Savannah Sparrow songs from the Kent Island population by BirdNET, by microdialect and individual, according to the recording method used to collect songs.

Figure 7 shows that Song Meter recordings, which are fixed in the field following a 50 x 50 meter grid, often yield low confidence scores for birds singing far from the microphones (ie. L.GL, G.WG, NW.L and Y.YW) and consistently high scores for those often observed near or perching on recorder (ie. O.NL and RY.B). Focal recordings, which are performed on foot at varying distances from singing birds, do not reflect this pattern. There seems to be no link between the scores obtained through both recording methods within the songs of an individual. This leads us to think that the results obtained when comparing average scores at the individual level were due to the bias in Song Meter recording distances, since there were not enough focal recordings for each bird to counteract this effect, and not to individual song characteristics.

Note that the focal recordings in the “Dash” microdialect appear to show very high and concentrated scores because only 43 songs were recorded in that category, all of them from the bird “Y.RL”.