



**HAL**  
open science

# Complexité vocale et contrôle cognitif chez le corbeau freux (*Corvus frugilegus*)

Killian Martin

► **To cite this version:**

Killian Martin. Complexité vocale et contrôle cognitif chez le corbeau freux (*Corvus frugilegus*). Sciences cognitives. Université de Tours, 2022. Français. NNT: . tel-04185669

**HAL Id: tel-04185669**

**<https://hal.science/tel-04185669v1>**

Submitted on 23 Aug 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

# UNIVERSIT  DE TOURS

* COLE DOCTORALE : Sant , Sciences Biologiques et Chimie du Vivant*

**UMR 7247 Physiologie de la Reproduction et des Comportements**

## TH SE pr sent e par : **Killian MARTIN**

soutenue le : 13 d cembre 2022

pour obtenir le grade de : **Docteur de l'universit  de Tours**  
Discipline/ Sp cialit  : Biologie

### **Complexit  vocale et contr le cognitif chez le corbeau freux (*Corvus frugilegus*)**

**TH SE dirig e par :**

**Mme DUFOUR Val rie**

Charg e de recherche, HDR, CNRS, INRAE Nouzilly

**CO-ENCADR E par :**

**M. OBIN Nicolas**

**M. ADAM Olivier**

Ma tre de conf rences, IRCAM, Sorbonne Universit 

Professeur des universit s, CNRS, Sorbonne Universit 

**RAPPORTEURS :**

**M. GLOTIN Herv **

**M. LEBOUCHER G rard**

Professeur des universit s, universit  de Toulon

Professeur des universit s, universit  Paris-Nanterre

**PR SIDENT du jury :**

**M. VENTURINI Gilles**

Professeur des universit s, universit  de Tours

---

**JURY :**

**Mme DUFOUR Val rie**

**M. OBIN Nicolas**

**M. ADAM Olivier**

**M. GLOTIN Herv **

**M. LEBOUCHER G rard**

**M. VENTURINI Gilles**

**Mme GEBERZAHN Nicole**

Charg e de recherche, HDR, CNRS, INRAE Nouzilly

Ma tre de conf rences, IRCAM, Sorbonne Universit 

Professeur des universit s, CNRS, Sorbonne Universit 

Professeur des universit s, universit  de Toulon

Professeur des universit s, universit  Paris-Nanterre

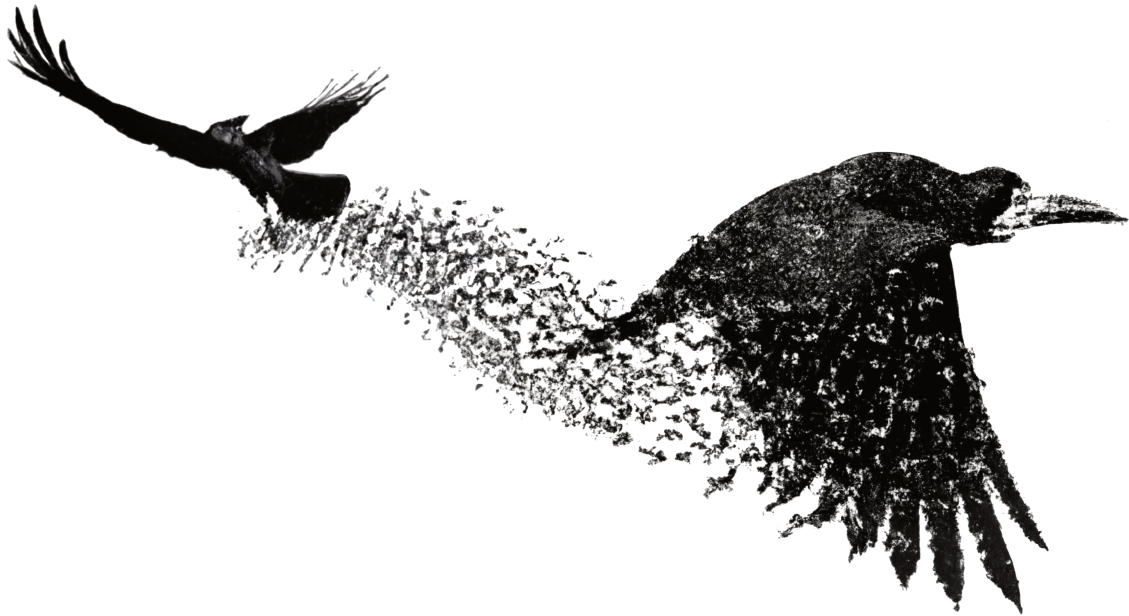
Professeur des universit s, universit  de Tours

Ma tresse de conf rences, universit  Paris-Nanterre





# Complexité vocale et contrôle cognitif chez le corbeau freux (*Corvus frugilegus*)





# Remerciements

Je remercie d'abord les membres du jury, Hervé Glotin, Gérard Leboucher, Nicole Geberzahn, et Gilles Venturini pour avoir accepté d'évaluer mon travail de thèse.

Je remercie bien entendu mes encadrants de thèse, Valérie Dufour, Nicolas Obin, et Olivier Adam. Valérie, pour m'avoir guidé tout au long de la thèse malgré toutes les difficultés, m'avoir donné bien plus d'idées que je ne pourrais en faire la liste ici, et m'avoir permis de me perdre (peut-être trop, mais c'est à son crédit) dans les considérations de machine learning... et m'avoir parfois rappelé que je faisais une thèse de biologie quand je m'y perdais trop. Si elle y a réussi, j'en laisserai juger le lecteur. Nicolas, très largement pour les mêmes choses, venant de l'autre côté de la barrière du machine learning. Ta réactivité et tes idées ont guidé en grande partie le développement de Rookognise. Olivier, merci pour ta mise à l'étrier en bioacoustique, pour tes idées, même celles que nous n'avons jamais pu mettre en œuvre par manque de temps. À tous les trois, merci pour vos conseils et votre guidance.

Je remercie également l'Eurométropole de Strasbourg, pour nous avoir accordé le terrain sur lequel la volière a pu être construite, et la Fondation Michelin, pour leur mécénat qui a permis la réalisation de cette thèse.

Je remercie tout le monde au labo à Strasbourg, permanents et non permanents. Merci à Bernard pour des discussions, parfois éclairantes, parfois... moins directement. Merci à Odile, pour son accueil dès mon premier stage, et bien trop de petites choses depuis (mais en particulier toutes les photos de l'album de la honte...!). Merci à Thomas et Nancy, les meilleurs camarades de bureau que j'aurais pu espérer, même si le Covid a coupé court à beaucoup trop de choses. Nancy, merci beaucoup pour toutes nos discussions de statistiques, de maths... et pour avoir été la meilleure Normande du labo! Merci à Anna et Laurie, même si nous n'avons pas eu beaucoup de temps ensemble. À toutes les deux, je vous souhaite beaucoup de chance pour le reste de la thèse.

An enormous thank you to Nicola S. Clayton, Francesca Cornero, and Emily Danby from the Cambridge lab. Thank you for having me, and thank you for all of your help with the birds. It was only a month and it was all too late, but it was a truly wonderful time. Congratulations on keeping the lab open against all odds. Francesca, I hope Leo keeps treating you well!

Je remercie énormément Simon Burgun pour avoir accepté de m'héberger au début de cette thèse, et pour tous ses bricolages à la volière, sans qui cette thèse aurait été bien différente (et aurait sûrement été faite à pied).

Un travail de thèse ne se fait jamais seul, et celle-ci a été particulièrement un travail de groupe. Je remercie tous les stagiaires que j'ai eu l'occasion d'encadrer pendant ces trois ans de travail : Agnès, Émeline, Julie, Louise, Hélène, Virgile, Agathe, Aurore, Laly, Lilou, Lucille, Soline, Ameline, Baptistin, Éloi, Marin, Marine, Nicolas, Pierre-Alexandre, Tom. Merci pour votre aide dans ce projet, que votre travail soit inclus dans ce manuscrit ou non. Je remercie tout particulièrement Agnès et Émeline pour leurs talents de photographes, et Agnès pour ses talents de dessinatrice (voir page suivante). Certaines de leurs photos ont été utilisées pour illustrer cette thèse, à côté des miennes. La distinction est laissée en exercice, certainement trivial, au lecteur.

Je remercie ma famille, malgré la distance. À mes parents, à mes sœurs, et à mes petites nièces. Merci pour votre soutien, merci pour votre présence (façon de parler, hélas) dans les moments les plus

difficiles.

Enfin, une petite pensée pour les corbeaux freux dans ce manuscrit : Balbo (malgré tout le sucre que j'ai pu casser sur son dos), Bashir, Braad, Brain, Elie, Feisty, Gigi, Jonas, Jolene, Kafka, Merlin, Osiris, Pomme, Siobhan, Tom à Strasbourg, et Aristotle, Bussell, Cassandra, Connelly, Fry, Huxley, et Leo à Cambridge. Merci à Reva, Rubis, Willow, et Zéphir, et à celles et ceux qui n'ont jamais eu de nom. Et à la mémoire de ceux qui manquent désormais. J'espère que ce travail aura été à leur hauteur.



# Table des matières

<b>1</b>	<b>Introduction générale</b>	<b>3</b>
1.1	Communication et complexité . . . . .	3
1.2	La complexité vocale dans l’histoire de vie d’une espèce . . . . .	13
1.3	La complexité socio-communicative chez les oiseaux . . . . .	26
1.4	Objectifs de la thèse . . . . .	33
<b>2</b>	<b>Méthodes générales</b>	<b>35</b>
2.1	Sujets d’études . . . . .	35
2.2	Enregistrements des données . . . . .	37
2.3	Analyses acoustiques . . . . .	40
<b>3</b>	<b>Rookognise : détection et identification automatique des vocalisations de corbeaux freux en milieu naturel par un réseau de neurones multi-tâches</b>	<b>53</b>
3.1	Introduction . . . . .	54
3.2	Material and methods . . . . .	57
3.3	Results . . . . .	65
3.4	Discussion . . . . .	71
3.5	Conclusions . . . . .	73
3.6	Acknowledgements . . . . .	74
3.7	Annexe 1 . . . . .	75
3.8	Annexe 2 . . . . .	77
<b>4</b>	<b>Complexité vocale chez le corbeau freux : commonalités et idiosyncrasies</b>	<b>81</b>
4.1	Introduction . . . . .	82
4.2	Matériel et Méthodes . . . . .	85
4.3	Résultats . . . . .	90
4.4	Discussion . . . . .	99
4.5	Annexe . . . . .	104
<b>5</b>	<b>Signature vocale chez le corbeau freux, de l’individu à la colonie</b>	<b>107</b>
5.1	Introduction . . . . .	107
5.2	Matériel et Méthodes . . . . .	111
5.3	Résultats . . . . .	114

5.4	Discussion . . . . .	118
<b>6</b>	<b>Les corbeaux freux peuvent-il se synchroniser à un stimulus rythmique ?</b>	<b>123</b>
6.1	Introduction . . . . .	123
6.2	Matériel et Méthodes . . . . .	126
6.3	Résultats . . . . .	129
6.4	Discussion . . . . .	134
<b>7</b>	<b>Discussion générale</b>	<b>139</b>
7.1	Suivi acoustique individuel avec un réseau de neurones profond . . . . .	139
7.2	Complexité vocale des corbeaux freux . . . . .	143
7.3	Le chant des corbeaux freux . . . . .	147
7.4	Conclusion . . . . .	152
	<b>Bibliographie</b>	<b>153</b>

# Chapitre 1

## Introduction générale

Pourquoi certaines espèces produisent-elles une grande richesse de vocalisations pour communiquer, alors que d'autres n'utilisent qu'un répertoire restreint à quelques vocalisations différentes ? Comment les espèces avec un répertoire large utilisent-elles ce répertoire dans leur comportement ?

### 1.1 Communication et complexité

#### 1.1.1 Communication : du signal au sens

##### 1.1.1.1 Définir la communication

La communication entre (au moins) deux individus, dans son sens le plus large, peut être considérée comme un transfert d'information d'un émetteur à un receveur (J. W. SMITH, 1965 ; WILEY, 2012). Pratiquement par définition, la communication est donc un comportement social (LIEBAL et al., 2012), puisqu'elle implique le transfert d'information entre plusieurs individus. Ce transfert d'information est médié par un signal sensoriel, qui peut emprunter un ou plusieurs canaux : visuel, acoustique, chimique, mécanique, électrique (BRADBURY et VEHRENCAMP, 1998). La propagation du signal est conditionnée par ses propriétés physiques. En milieu fermé, comme en forêt, où la visibilité est limitée, les signaux acoustiques (notamment les vocalisations) sont privilégiés car ils ne dépendent pas de la ligne de vue comme un signal visuel, et peuvent se propager à des distances bien supérieures (MARLER, 1967 ; WASER et BROWN, 1986). En outre, la production vocale est peu coûteuse en énergie par rapport à d'autres mouvements corporels (OBERWEGER et GOLLER, 2001 ; WARD et al., 2003). Par conséquent, la production de vocalisations à plusieurs avantages qui peuvent expliquer leur prévalence dans les systèmes de communication animale.

Le signal, propagé à travers le milieu approprié, est perçu par le receveur, qui l'interprète et ajuste son comportement en conséquence (CHERRY, 1955 ; SLATER, 1983 ; J. W. SMITH, 1965). Cet ajustement comportemental, et l'information qui est donc implicitement extraite par le receveur, correspond au sens du signal du point de vue de celui-ci (MARLER, 1961 ; SEYFARTH et CHENEY, 2017 ; J. W. SMITH, 1965). Il faut noter que l'information émise et l'information extraite ne sont pas nécessairement identiques. En particulier, un signal peut être honnête ou malhonnête, selon que l'information transmise bénéficie à tous les individus de façon équivalente ou bénéficie plus à l'émetteur qu'à receveur (KREBS et DAWKINS, 1984). Cependant, une communication malhonnête, si elle est répétée comme dans les espèces sociales où des individus peuvent interagir régulièrement avec les mêmes partenaires sociaux, peut permettre au receveur d'ignorer un émetteur qui cherche à le tromper (CHENEY et SEYFARTH, 2018). Par conséquent, il est probable que les signaux malhonnêtes sont défavorisés



chez les espèces sociales, comme les primates ou certains oiseaux. En effet, la production d'un signal a toujours un coût, métabolique certes mais aussi social et écologique : un cri signale la position de l'individu, ce qui peut donc l'exposer à un prédateur. Par ailleurs, un individu malhonnête ignoré par le groupe peut perdre accès aux bénéfices de la vie en groupe, comme la protection face aux prédateurs ou aux conspécifiques rivaux, l'accès aux ressources ou à la reproduction (LIGON et MCGRAW, 2016; NOWICKI et SEARCY, 2005). Ce cas particulier illustre le fait que l'information contenue dans une vocalisation (ou tout autre signal) va être perçue de manière contextualisée, en fonction de l'environnement (e.g., position relative de l'émetteur), du congénère qui la perçoit, (e.g. relation sociale avec l'émetteur), ou d'un acteur tiers comme par exemple un prédateur (J. W. SMITH, 1965).

### 1.1.1.2 Encodage d'informations dans les vocalisations

Les différentes vocalisations produites par des individus ou une espèce constituent leur répertoire vocal respectif (SEARCY, 1992). Ce répertoire est donc composé de vocalisations de structures acoustiques variées (BRADBURY et VEHRENCAMP, 1998), qui peuvent correspondre à différents types d'information. Ces informations sont utilisées dans les nombreuses situations dont les individus peuvent faire l'expérience : maintien d'un contact entre individus distants, alerte à la présence de prédateurs, signalement d'une source de nourriture, quémandage, agression, soumission, défense d'un territoire, recherche d'un partenaire sexuel (BRADBURY et VEHRENCAMP, 1998; CHEN et WIENS, 2020; MARLER et SLABBEKOORN, 2004). L'information contenue dans les vocalisations peut être encodée par des variations dans la structure des vocalisations. Par exemple, les différents types de cris produits par les primates et autres mammifères terrestres, ainsi que par les oiseaux, ont des structures différentes qui reflètent des fonctions différentes (MARLER, 2004a). Ainsi, les cris d'alarme signalent la présence de prédateurs (STEPHAN et ZUBERBÜHLER, 2014; SUZUKI, 2014), les cris alimentaires signalent la présence d'une source de nourriture (PENDERGRAFT et MARZLUFF, 2019), les cris de quémandage indiquent la sollicitation d'un nourrissage auprès d'un parent (notamment chez les juvéniles) ou d'un partenaire (LEONARD et HORN, 2005), les cris de contacts servent à coordonner des activités de groupe (KONDO et WATANABE, 2009)... Ces vocalisations conditionnent le type d'interaction avec les congénères pour certaines fonctions majeures et véhiculent une information très précise. Par exemple, certaines espèces produisent des cris d'alarmes spécifiques à certains prédateurs (primates : SEYFARTH et al., 1980; STEPHAN et ZUBERBÜHLER, 2014, oiseaux : MARLER et EVANS, 1996; YORZINSKI et VEHRENCAMP, 2009).

Certaines variations n'informent pas sur la fonction de la vocalisation, mais peuvent dépendre de caractéristiques internes à l'individu, comme l'anatomie ou l'état émotionnel. L'anatomie de l'individu comme sa taille ou sa masse corporelle, et par extension du tractus vocal, influe sur les caractéristiques acoustiques de ses vocalisations (EY et al., 2007; FAVARO et al., 2017; FITCH et HAUSER, 1995; TAYLOR et REBY, 2010). Par exemple, les mâles de grande taille produisent fréquemment des vocalisations plus graves, chez plusieurs espèces de manchots (FAVARO et al., 2017; MIYAZAKI et WAAS, 2003), ainsi que chez le cerf élaphe (*Cervus elaphus*, REBY et al., 2005). L'état émotionnel de l'individu peut également induire une variation dans ses vocalisations (BRIEFER, 2012; MORTON, 1977). Les vocalisations agonistiques émises par un agresseur peuvent être plus inharmo-

niques et graves, alors que les vocalisations affiliatives ou de soumissions sont plus harmoniques et aiguës, une relation retrouvée chez un certain nombre de mammifères et d'oiseaux (MORTON, 1977). Les cris d'alarme de certaines espèces peuvent également varier pour exprimer une urgence liée à un danger proche (CLAY et al., 2012). La corneille d'Amérique (*Corvus brachyrhynchos*) et le geai bleu (*Cyanocitta cristata*) produisent les mêmes vocalisations, mais plus longues et à intervalles plus courts proportionnellement à l'intensité de la réaction à un prédateur (DAHL, 2015; YORZINSKI et VEHRENCAMP, 2009), alors que la mésange à tête noire (*Poecile atricapillus*), l'écureuil terrestre (*Spermophilus beldingi*), et le vari (*Varecia variegata*) utilisent des cris d'alarmes différents en fonction de la proximité ou de la taille d'un prédateur (MACEDONIA et EVANS, 1993; TEMPLETON et al., 2005).

### 1.1.2 Variabilité dans la communication vocale

Différents mécanismes sont à l'origine des variations dans les vocalisations des animaux, depuis les contraintes anatomiques aux mécanismes d'apprentissage du répertoire, ainsi qu'au contrôle exercé par l'émetteur même de la vocalisation.

#### 1.1.2.1 Information statique : morphologie et signature vocale

Les vocalisations sont produites par un appareil vocal composé de l'appareil respiratoire de l'animal et d'une variété de muscles adjacents à cet appareil, organisées selon ce qui a été qualifié de modèle source-filtre (TAYLOR et al., 2016). La source consiste en plis vocaux vibrant au passage de l'air expiré par les poumons. La fréquence de vibration de ces plis vocaux conditionne la fréquence sonore de la vocalisation. L'anatomie exacte de la source conditionne les sons qui peuvent être produits : le larynx des mammifères contient un jeu unique de plis, les cordes vocales, situé en haut de la trachée, alors que le syrinx des oiseaux consiste essentiellement en un double jeu de cordes vocales, situé en haut de chaque branche pulmonaire des bronches (BRACKENBURY, 1980, 1982, 1989). Chaque moitié du syrinx est contrôlée par un ensemble de muscles, chacun innervé par un nerf crânien différent (LARSEN et GOLLER, 2002; NOWICKI, 1987). Cette organisation permet la production de sons selon des mécanismes extrêmement complexes à deux voix, un phénomène dit de biphonation (NOWICKI, 1987; WILDEN et al., 1998). La biphonation permet de générer deux flux sonores qui peuvent interagir dans la trachée, où les fréquences des vibration de chaque flux d'air peuvent résonner de façon constructive (i.e. amplifiée) ou destructive (i.e. atténuée voire annulée), permettant à l'individu la production de vocalisations avec des dynamiques acoustiques complexes, comme par exemple la production de vocalisations à deux voix (SUTHERS et ZOLLINGER, 2004), à structure chaotique (FLETCHER, 2000), ou avec des structures non-linéaires comme des saut de fréquences brusques (TANIMOTO et al., 2017; ZOLLINGER et al., 2008). La biphonation n'est pas exclusive aux oiseaux, mais existe aussi chez certains mammifères (canidés WILDEN et al., 1998; cétacés CAZAU et al., 2016), mais n'a pas été décrite chez les primates.

Outre les différences anatomiques qui existent entre les espèces, un degré de polymorphisme existe entre les individus d'une même espèce qui influe la production vocale. En particulier, des para-

mètres anatomiques comme la taille ou la masse corporelle sont inversement corrélées à la fréquence des vocalisations, i.e. des individus de grande taille produisent fréquemment des vocalisations plus graves (EY et al., 2007 ; FAVARO et al., 2017 ; FITCH et HAUSER, 1995 ; MIYAZAKI et WAAS, 2003 ; TAYLOR et REBY, 2010). Ces paramètres peuvent être utilisés par les animaux pour signaler une information : par exemple, chez les cerfs élaphe mâles, la repasse de vocalisations graves (simulant donc un mâle concurrent de grande taille) augmentait les comportements de vigilance et la production de vocalisations plus graves en réponse (REBY et al., 2005). Il est notable que cet exemple illustre également la notion de signal malhonnête, ainsi que de contrôle par l'individu de sa production vocale : les cerfs produisent ainsi des vocalisations plus graves en réponse à la repasse que leurs vocalisations spontanées. Cela peut être avantageux pour éviter des compétitions coûteuses comme pour l'accès à la reproduction, ou bien favoriser un mâle qui serait désavantagé si une telle compétition devait arriver (NOWICKI et SEARCY, 2005). Ces variations anatomiques encodent une information qui peut être considérée comme statique : toutes les vocalisations d'un individu donné sont influencées de la même façon par sa morphologie, ce qui peut servir à les identifier (FAVARO et al., 2017 ; REBY et al., 2005). Il faut noter que ces considérations ne semblent valables qu'en comparant les individus d'une même espèce. Les comparaisons entre espèces ne donnent pas nécessairement la même corrélation entre morphologie et paramètres des vocalisations : ainsi, LAIOLO et ROLANDO (2003) ont montré chez les corvidés que les vocalisations des espèces de grande taille sont plus aiguës que celles des espèces de petite taille.

Plus généralement, l'encodage vocal de l'identité, aussi appelé signature vocale, est un phénomène extrêmement répandu dans le règne animal (BEECHER, 1989 ; BERGER et LIGON, 1977 ; LINHART et al., 2019 ; oiseaux : BENTI et al., 2019 ; BERG et al., 2012 ; BOECKLE et al., 2012 ; CHARRIER et al., 2001 ; ELIE et THEUNISSEN, 2018 ; MATES et al., 2015 ; VIGNAL et al., 2008 ; YORZINSKI et al., 2006 ; primates : BOUCHET et al., 2013 ; CANDIOTTI et al., 2012 ; KEENAN et al., 2020 ; PRICE, 2015 ; SPILLMANN et al., 2017 ; mangouste : JANSEN et al., 2012 ; cétacés : KERSHENBAUM et al., 2013 ; KREMERS et al., 2012 ; QUICK et JANIK, 2012), sans doute par l'avantage que la signature vocale confère pour des espèces sociales : en effet, la formation d'une relation sociale nécessite la capacité de reconnaître des individus particuliers au cours d'interactions répétées. Au vu de l'importance fonctionnelle de la communication vocale, encoder l'identité dans les vocalisations offre un moyen rapide de reconnaître des partenaires sociaux. Et en effet, la capacité à reconnaître des individus particuliers par leurs vocalisations a été démontrée chez plusieurs espèces (BEER, 1971 ; BRECHT et NIEDER, 2020 ; CANDIOTTI et al., 2013 ; ELIE et THEUNISSEN, 2018 ; KONDO et al., 2010 ; RØSKAFT et ESPMARK, 1984), et est sans doute présente chez toutes les espèces où des relations sociales existent. Certaines variations sont communes à plusieurs individus, comme des indicateurs du sexe et de l'âge (BOUCHET et al., 2010 ; GARCIA DE LA CHICA et al., 2020 ; LAIOLO et al., 2000 ; MATES et al., 2015 ; VOLODIN et al., 2015), qui pourraient être liées à des polymorphismes plus systématiques entre ces différentes classes d'individus.

Ces informations statiques peuvent néanmoins entraîner un conflit dans l'encodage d'information. En particulier, la transmission des autres types d'informations (comme, pour ne prendre qu'un seul exemple, les cris d'alarme décrits dans la section précédente) peut être gênée si les vocalisa-

tions sont trop individualisées (CALCARI et al., 2021 ; KEENAN et al., 2020). Ce conflit entre les deux types d'information peut être résolu de plusieurs façons : 1) une ségrégation de l'information, par exemple avec des vocalisations individualisées et d'autres porteuses d'autres informations, 2) une signature vocale constante à travers le répertoire entier de l'individu, et 3) un processus de mémorisation de l'identité de l'individu séparé pour chaque vocalisation. Relativement peu d'études ont testé ces hypothèses, car la plupart des études jusqu'à récemment se sont concentrées sur un seul type de vocalisations commun à tout le groupe d'étude (BENTI et al., 2019 ; BERG et al., 2012 ; BOECKLE et al., 2012 ; BOUCHET et al., 2013 ; CANDIOTTI et al., 2012 ; CHARRIER et al., 2001 ; KERSHENBAUM et al., 2013 ; KREMERS et al., 2012 ; PRICE, 2015 ; QUICK et JANIK, 2012 ; SPILLMANN et al., 2017 ; STOWELL et al., 2016a ; VIGNAL et al., 2008 ; YORZINSKI et al., 2006). Un nombre comparativement très réduit d'études ont cherché à identifier des individus à travers plusieurs types de productions vocales (CHENG et al., 2012 ; ELIE et THEUNISSEN, 2018 ; E. FOX et al., 2006 ; E. J. FOX et al., 2008 ; JANSEN et al., 2012 ; KEENAN et al., 2020 ; KIRSCHER et al., 2009 ; MATES et al., 2015 ; PISANSKI et al., 2020 ). De ces études se dégagent un mélange de mécanismes derrière les signatures vocales : des propriétés acoustiques communes entre différentes vocalisations peuvent exister (CHENG, 2002 ; E. J. FOX et al., 2008 ; KIRSCHER et al., 2009), parce qu'elles sont liées à des contraintes anatomiques (MATES et al., 2015 ; PISANSKI et al., 2020). Cependant, les paramètres permettant d'identifier les individus diffèrent fréquemment en fonction des vocalisations (ELIE et THEUNISSEN, 2018 ; KEENAN et al., 2020). Dans ce cas, posséder une vocalisation dont l'unique fonction est de signaler l'identité d'un individu pourrait faciliter la communication entre individus (JANSEN et al., 2012).

### 1.1.2.2 Environnement physique

L'habitat d'une espèce fait partie intégrante de son histoire de vie. Un habitat ouvert ou fermé, la pression de prédation, la compétition inter- et intra-spécifique, sont autant de facteurs conditionnant l'écologie, le comportement, et la communication entre individus.

Le degré d'ouverture de l'habitat conditionne la facilité avec laquelle les individus peuvent communiquer. Un milieu fermé, où les individus ne peuvent pas se voir, favorise la communication vocale en général (MARLER, 1967 ; WASER et BROWN, 1986) et des vocalisations plus individuellement distinctes (CANDIOTTI et al., 2012). La séparation par un obstacle visuel favorise également la communication vocale : des partenaires sociaux séparés produisent ainsi leurs cris de contact plus fréquemment (diamant mandarin *Taeniopygia guttata* PEREZ et al., 2015 ; corbeaux freux *Corvus frugilegus* MUNTEANU et al., 2009). Des obstacles non visuels peuvent également gêner la communication, mais dans ce cas la simple production vocale ne suffit pas : par exemple, les interférences dues aux vocalisations d'autres individus, à celles d'autres espèces, ou à des bruits abiotiques, peuvent interférer avec la communication. Communiquer malgré ces interférences implique plusieurs stratégies (SLABBEKOORN, 2004). Une stratégie simple est l'effet Lombard (ZOLLINGER et BRUMM, 2011), qui consiste à augmenter le volume sonore des vocalisations. Cet effet est retrouvé chez tous les oiseaux et mammifères chez qui il a été testé (perruche ondulée : MANABE et al., 1998 ; caille japonaise : POTASH, 1972 ; diamant mandarin : CYNX et al., 1998 ; rossignol : BRUMM et TODT, 2002 ; chauve-souris : JIANG et al., 2019 ; être humain : ZOLLINGER et BRUMM, 2011). Une autre stratégie

est de répéter la vocalisation. Cette hypothèse pourrait expliquer la répétition des chants chez plusieurs espèces d'oiseaux (DEONIZIAK et OSIEJUK, 2019 ; PRICE, 2013) : en répétant l'information, l'individu augmente la probabilité qu'elle soit transmise, même en présence d'interférences. Enfin, l'individu peut modifier la structure même de ses vocalisations, par exemple en augmentant la fréquence sonore (SLABBEKOORN et PEET, 2003). Toutes ces stratégies modifient la production vocale pour en faciliter la transmission dans l'habitat.

L'environnement social et la pression de prédation peuvent également entraîner des variations dans la production vocale. Il est avantageux de transmettre un signal, ce qui favorise des vocalisations puissantes, et parfois d'être localisable, comme pour attirer un partenaire sexuel. Cependant, ce même signal peut entraîner un risque s'il est détectable par un prédateur. Pour limiter ce risque, certaines espèces utilisent des signaux différents, plus difficiles à localiser pour les prédateurs (hypothèse de la dissimulation acoustique, RUXTON, 2009). Les stratégies de dissimulation varient selon les espèces et les situations, comme un "évitement acoustique", c'est-à-dire la non-production de vocalisations en situation risquée (CURIO, 2012 ; RUXTON, 2009), à des vocalisations adaptées pour compliquer la localisation de l'émetteur, comme des vocalisations aiguës ou à large bande de fréquence (MARLER, 1955 ; REDONDO et DE REYNA, 1988 ; RUXTON, 2009). Cependant, d'autres stratégies impliquent une modification active de la production vocale de l'individu : chez les baleines, les mères et leurs veaux produisent des vocalisations qui se diffusent moins loin par rapport aux autres vocalisations de l'espèce (PARKS et al., 2019). Certaines espèces de grenouilles (*Physalaemus pustulosus*) adaptent leurs vocalisations en fonction de l'équilibre entre compétition sexuelle et risque de prédation : quand plusieurs mâles vocalisent en même temps des vocalisations élaborées sont privilégiées, au risque d'attirer les prédateurs, tandis que les mâles isolés produisent des vocalisations plus simples (RYAN et al., 1982). Les cris d'alarme sont un autre exemple de variation acoustique, où plusieurs espèces adaptent leur production vocale au danger posé par le prédateur (primates : SEYFARTH et al., 1980 ; STEPHAN et ZUBERBÜHLER, 2014, oiseaux : FURRER et MANSER, 2009 ; YORZINSKI et VEHRENCAMP, 2009). Ainsi, le répertoire vocal ne suit pas une relation structure-fonction stricte, mais peut être modifié en fonction du contexte et de l'état interne de l'individu.

### 1.1.2.3 Plasticité vocale : apprentissage, imitation, dialectes

Le répertoire vocal est acquis à un stade précoce dans la vie de pratiquement toutes les espèces qui vocalisent (BEECHER et BRENOWITZ, 2005 ; CHEN et al., 2016 ; KUHL, 2012). Chez de nombreuses espèces, cet apprentissage est médié par une forte phase de plasticité où le juvénile apprend par l'imitation des vocalisations de conspécifiques adultes, particulièrement pour l'acquisition du chant chez les oiseaux (ADRET-HAUSBERGER, 1982 ; KROODSMA, 2004 ; MARLER et PETERS, 1987). Chez la plupart des oiseaux chanteurs (estimée à 80% des espèces, BEECHER et BRENOWITZ, 2005 ; MARLER, 1997 ; MARLER et PETERS, 1987) et la grande majorité des primates (HAMMERSCHMIDT et FISCHER, 2008 ; TEN CATE, 2021) cette phase de plasticité est finie (une phase dite de cristallisation chez les oiseaux), et l'individu ne modifie plus son répertoire ensuite (sans nécessairement empêcher une certaine flexibilité, comme nous verrons dans la section suivante). Seules quelques espèces maintiennent une capacité d'apprentissage à l'âge adulte (BEECHER et BRENOWITZ, 2005 ;

VERPOOTEN, 2021), principalement certains oiseaux (BRENOWITZ et al., 1997; KONDO, 2021; WEST et KING, 1990) et cétacés (REISS et MCCOWAN, 1993). Ce maintien de la plasticité est très variable : certaines espèces, notamment sociales, maintiennent un apprentissage par imitation qui permet une convergence vocale entre individus interagissant régulièrement (correspondant à une modification des vocalisations de chaque individu jusqu'à atteindre une structure intermédiaire commune, CANDIOTTI et al., 2012; MONTEIRO et al., 2021; SEWALL, 2009), tandis que d'autres, comme les oiseaux dits "à chant ouvert" (BEECHER et BRENOWITZ, 2005; ROBINSON et al., 2019)), peuvent apprendre de nouvelles vocalisations tout au long de leur vie (SCARL et BRADBURY, 2009; perroquets THOMSEN et al., 2019, corvidés E. D. BROWN, 1985, étourneau sansonnet *Sturnus vulgaris* EENS et al., 1992). Certaines espèces possèdent des capacités de mimétisme particulièrement développées, leur permettant notamment d'imiter des sons hétérospécifiques (perroquets THOMSEN et al., 2019, corvidés E. D. BROWN, 1985, étourneau sansonnet EENS et al., 1992, oiseau lyre *Menura alberti* BACKHOUSE et al., 2022). Un cas particulier est Le drongo brillant (*Dicrurus adsimilis*), qui peut imiter les cris d'alarmes d'autres espèces comme les suricates pour les éloigner d'une source de nourriture, et est de plus capable de varier ces imitations pour éviter une habituation de l'espèce imitée (FLOWER et al., 2014).

Par le même mécanisme d'apprentissage, de nombreuses espèces sociales développent des variations géographiques dans leurs vocalisations, le phénomène donnant lieu à des dialectes vocaux différents (BAKER et al., 2000; HUGHES et al., 2007; MUNDINGER, 1982; SEWALL, 2009; WRIGHT et DEMEURE, 2021; WRIGHT et WILKINSON, 2001). Ces dialectes ne sont pas liés à des relations génétiques (MUNDINGER, 1982), mais bien à l'apprentissage par les juvéniles qui intègrent des groupes stables et en imitent les vocalisations (SALINAS-MELGOZA et WRIGHT, 2012). Ce partage vocal est également présent dans certaines espèces territoriales, où des individus maintenant des territoires adjacents partagent une partie de leur chants (BEECHER et CAMPBELL, 2005; GRIESSMANN et NAGUIB, 2002; HUGHES et al., 2007). En cas de conflit, ces chants communs peuvent être utilisés préférentiellement dans un phénomène de contre-chant (*matched countersinging*), où un individu répondra au chant de l'autre par le même chant. Ce contre-chant sert de signal agonistique et a pour conséquence d'amplifier le conflit (BEECHER et CAMPBELL, 2005; SHACKLETON et RATCLIFFE, 1994). Dans d'autres cas, l'individu peut atténuer le conflit en chantant un chant différent, en particulier non partagé avec le rival (SHACKLETON et RATCLIFFE, 1994).

### 1.1.3 Complexité vocale

#### 1.1.3.1 Complexité : un phénomène intrinsèquement multidimensionnel

De nombreux mécanismes introduisent de la variabilité dans le répertoire vocal d'une espèce. Leur combinaison permet l'encodage d'une vaste quantité d'information dans un système de communication complexe. L'étude de la complexité est devenue une question centrale dans la communication animale, en particulier dans le chant des oiseaux, en raison de parallèles possibles dans son organisation, son développement et les circuits neuronaux impliqués dans sa production (BOLHUIS et al., 2010; DOUPE et KUHL, 1999; JARVIS, 2004; MIYAGAWA et al., 2014; PETKOV et JARVIS, 2012;

SAINBURG et GENTNER, 2021 ; SAINBURG et al., 2019 ; WILBRECHT et NOTTEBOHM, 2003). Cependant, étudier la complexité dans un contexte biologique requiert une définition opérationnelle, tant pour fixer la question biologique d'une étude donnée, que pour assurer la comparabilité avec d'autres études. En effet, la définition de la complexité conditionne en partie les mesures utilisées pour la quantifier, ce qui peut donner des résultats discordants selon les mesures ou les études (BENEDICT et NAJAR, 2019 ; KERSHENBAUM et al., 2018).

La définition qui s'est dégagée de la littérature s'accorde sur les propriétés attendues d'un système complexe, notamment la capacité à encoder une grande quantité d'informations (FREEBERG et al., 2012 ; PECKRE et al., 2019). En particulier, FREEBERG et al. (2012) ont défini la complexité, dans le contexte d'un système de communication, comme "de nombreux éléments structurellement et fonctionnellement distincts (e.g. avec une grande taille du répertoire des signaux) ou qui possèdent une grande quantité de fragments d'information"<sup>1</sup>. Quantifier l'information contenue dans un système est un vaste problème, qui est le sujet du champ de la théorie de l'information (SHANNON, 1948). Cependant, un système peut devenir complexe par plusieurs mécanismes, qui ont été reflétés par les nombreuses études sur différents pans de la communication chez les animaux. Ces mécanismes peuvent être répartis en trois axes indépendants : la diversité, la flexibilité et la combinabilité du système (REBOUT et al., 2021). Ces trois axes correspondent à différents mécanismes par lesquels les signaux qui composent un système de communication peuvent encoder l'information. Il est important de noter qu'ils peuvent être mesurés à des échelles différentes, de l'espèce entière à un unique individu.

### 1.1.3.2 Diversité vocale : quantité de signaux différents et leur distribution

La diversité d'un répertoire vocal correspond à trois éléments distincts : le nombre de signaux différents dans le répertoire, à quel point ces signaux sont distincts, et leur répartition dans la production vocale.

Augmenter le nombre de signaux différents dans un répertoire vocal (la taille du répertoire, ou sa richesse) est une façon simple d'encoder une plus grande quantité d'informations différentes (MANSER et al., 2014 ; MARLER et SLABBEKOORN, 2004). Il est important de noter que cela ne signifie pas simplement la quantité de vocalisations, mais aussi la quantité de vocalisations à fonction équivalente. Pour reprendre l'exemple des cris d'alarmes cités plus haut, une espèce comme les cerco-pithèques qui possède trois cris d'alarmes, chacun adapté à un prédateur différent (comme décrit chez plusieurs espèces, e.g. FURRER et MANSER, 2009 ; STEPHAN et ZUBERBÜHLER, 2014 ; YORZINSKI et VEHCAMP, 2009), est plus complexe qu'une espèce qui n'en posséderait qu'un seul produit quel que soit le prédateur : au-delà de l'information "prédateur", les différents cris d'alarme peuvent ajouter une information permettant aux receveurs du signal d'ajuster leur comportement de façon fine, comme par exemple de surveiller un prédateur aérien ou terrestre en regardant en hauteur ou au sol (ZUBERBÜHLER et al., 1997). La richesse du répertoire a longtemps été la mesure principale de la complexité vocale, en partie par sa simplicité, mais elle n'est qu'un seul aspect de cette com-

---

1. "Complex communicative systems are those that contain a large number of structurally and functionally distinct elements (e.g. large display repertoire sizes) or possess a high amount of bits of information." FREEBERG et al. (2012)

plexité (PECKRE et al., 2019). En particulier, les différentes vocalisations constituant le répertoire ne sont pas équivalentes, et peuvent notamment être produites dans des proportions différentes suivant les contextes (MERCADO et PERAZIO, 2022). Si ces vocalisations peuvent être considérées comme fonctionnellement équivalentes, la diversité n'est pas forcément équivalente à la richesse. Prenons un exemple : un individu possède deux cris de contact et les produit dans deux contextes différents. Dans le contexte A, un de cris de contact est produit presque exclusivement, l'autre est produit occasionnellement. Dans le contexte B, les deux cris sont produits en proportion équivalente. La richesse du répertoire est alors de 2 dans les deux contextes. Cependant, il est plus facile de prédire quel cri sera produit dans le contexte A que dans le contexte B. La diversité du répertoire devrait donc être plus élevée dans le contexte B que dans le contexte A. La prise en compte de la distribution des différentes vocalisations est donc utile à la quantification de la diversité, notamment dans des approches comparatives (REBOUT et al., 2021).

Enfin, la notion de quantité de signaux est compliquée par l'existence de signaux dits gradés, où des vocalisations à structure et sens différents sont séparées par un continuum de vocalisations de structure intermédiaire (HAUSER, 1997 ; MARLER, 1967), sans séparation claire dans la structure ou un changement de sens marqué (RANGE et FISCHER, 2004). La gradation permet la communication d'informations subtiles (MANSER et al., 2014), mais peut compliquer l'interprétation des signaux en affaiblissant le lien entre structure de la vocalisation et sens du signal (Green 1975). La gradation a été très documentée chez les primates (primates FISCHER et al., 2017 ; HAMMERSCHMIDT et FISCHER, 1998 ; KEENAN et al., 2013 ; KEENAN et al., 2020 ; OUATTARA et al., 2009 ; WADEWITZ et al., 2015), mais seulement chez quelques autres mammifères (canidés YIN et MCCOWAN, 2004, suricates MANSER et al., 2014), et encore moins chez les oiseaux (MILLER, 1979 ; SUZUKI, 2014). L'évolution de signaux gradés pourrait être favorisée chez les espèces vivant en milieu ouvert ou quand les congénères maintiennent un contact visuel. Par exemple, les cercopithèques dianes produisent des vocalisations moins gradés quand les individus ne peuvent pas se voir (CANDIOTTI et al., 2012). De nombreuses espèces d'oiseaux vivent en milieu fermé (par exemple forestier) ou interagissent à des distances qui interdisent le contact visuel (CATCHPOLE et SLATER, 2008). Il est donc possible que la gradation ne soit simplement pas avantageuse dans ces situations et ne se soit donc pas développée chez ces espèces.

### 1.1.3.3 Flexibilité vocale : signaux variables, gradation et découplage structure-fonction

La flexibilité vocale correspond à toutes les modifications réversibles qu'un individu peut appliquer à son répertoire. Deux classes de modifications, et donc de flexibilité, sont possibles : structurale et fonctionnelle.

La flexibilité structurale recouvre toutes les modifications de la structure acoustique des vocalisations que peuvent imprimer les individus dans un contexte donné, et inclut tant les effets des émotions que des influences sociales et l'imitation. La gradation évoquée dans la section précédente peut alternativement être considérée comme de la flexibilité structurale. L'état émotionnel de l'individu peut conduire à des modifications de la structure des vocalisations, mais souvent relativement



simples (BRIEFER, 2012). Dans les signaux d'urgence, comme pour signaler la proximité d'un prédateur, cela peut conduire à un allongement des vocalisations ou à une augmentation de la fréquence de vocalisation (WHEATCROFT, 2015 ; YORZINSKI et VEHCAMP, 2009). La flexibilité structurale peut également se manifester par les ajustements vocaux que font certaines espèces lors d'interactions sociales, notamment la convergence vocale au sein de groupes sociaux stables (BALSBY et SCARL, 2008 ; MONTEIRO et al., 2021 ; SEWALL, 2009). Chez les espèces imitatrices comme certains perroquets, cette convergence peut être suffisamment rapide pour se produire en un seul échange vocal (SCARL et BRADBURY, 2009).

La flexibilité fonctionnelle correspond à un découplage structure-fonction : certaines espèces utilisent les mêmes vocalisations dans des contextes différents (OLLER et al., 2013). Cette capacité sous-entend un contrôle de la production vocale par l'individu qui n'a pas été démontré chez de nombreuses espèces. Les principales études ont été réalisées chez les primates (voir FISCHER et al., 2017 ; TAYLOR et REBY, 2010 pour des revues détaillées), mais quelques espèces d'oiseaux ont également démontré un découplage des vocalisations de leur contextes usuels (perroquets : MONTES-MEDINA et al., 2016 ; corvidés : BRECHT et al., 2019). Par exemple, les corneilles noires (*Corvus corone*) peuvent être conditionnées à produire des vocalisations en réponse à des stimuli arbitraires, en-dehors de leur fonction normale (BRECHT et al., 2019). Les observations conduites par ROSKAFI et ESPMARK (1982) sur le corbeau freux (*Corvus frugilegus*) ont relevé une association du croassement le plus fréquent produit avec sept contextes différents. Cependant, peu d'études formelles sur la flexibilité vocale fonctionnelle ont été faites chez les oiseaux.

#### 1.1.3.4 Combinabilité vocale : signaux composés, séquences, syntaxe, et chant

La combinabilité concerne tous les signaux composés à partir de plusieurs vocalisations arrangées en séquences. Le nombre de signaux possibles dans un répertoire est limité par les contraintes anatomiques de l'espèce (ARNOLD et ZUBERBÜHLER, 2006), et probablement également par des contraintes cognitives. Combiner plusieurs vocalisations en un nouveau signal est donc un moyen simple de générer une immense variabilité : en ignorant toutes contraintes, le nombre de signaux possibles augmente de façon linéaire avec le nombre de vocalisations, mais géométrique avec la longueur de la combinaison (suivant la formule  $N^L$ , avec  $L$  la longueur de la combinaison et  $N$  le nombre de vocalisations).

L'utilisation de combinaisons de vocalisations semble très répandue parmi les animaux (TOWNSEND et al., 2018 ; primates : ARNOLD et ZUBERBÜHLER, 2006 ; OUATTARA et al., 2009, mangoustes : MANSER et al., 2014 ; oiseaux ; CATCHPOLE et SLATER, 2008 ; SUZUKI, 2021). Les combinaisons les plus simples sont la répétition d'une vocalisation donnée, comme pour signaler l'approche d'un prédateur (SUZUKI, 2014 ; WHEATCROFT, 2015). Cependant, il n'est pas clair s'il s'agit d'un véritable cas de combinabilité ou un simple indice de l'état émotionnel de l'individu, et ces signaux pourraient davantage correspondre à un message d'urgence qu'à une nouvelle information proprement dite (FURRER et MANSER, 2009). Des combinaisons hétérogènes de vocalisations peuvent encoder des informations différentes, soit par simple juxtaposition de deux vocalisations avec leurs propres sens, soit en combinant des vocalisations en un signal à sens séparé (SUZUKI et al., 2019). Les deux

possibilités ont été mises au jour chez les primates (ARNOLD et ZUBERBÜHLER, 2006), les mangoustes (JANSEN et al., 2012) et les oiseaux (FICKEN et al., 1987 ; SUZUKI, 2014). Par exemple, les mangoustes encodent l'identité de l'individu émetteur dans une vocalisation et le signal à proprement parler dans une autre (JANSEN et al., 2012). Cependant, il n'est pas encore clair si les signaux combinés peuvent faire l'objet d'une syntaxe, où le sens du signal dépend de celui des signaux qui le composent ainsi que de leur ordre (SUZUKI et al., 2019), ce qui serait le premier élément pour commencer à les considérer comme un langage (BEECHER, 2021).

Sans aucun doute, le cas de combinabilité vocale le plus répandue dans la littérature est le chant ouvert trouvé chez une partie des oiseaux. Presque toutes les espèces d'oiseaux produisent un chant composé de vocalisations hétérogènes arrangées en séquence (BEECHER et BRENOWITZ, 2005 ; CATCHPOLE et SLATER, 2008). Selon les espèces, le répertoire de chant peut consister en une seule séquence par individu (MARLER et TAMURA, 1964) jusqu'à plusieurs centaines (KROODSMA et PARKER, 1977). Cependant, chez l'immense majorité des espèces (environ 80% des espèces étudiées, BEECHER et BRENOWITZ, 2005) le chant n'est pas un véritable exemple de combinabilité. D'abord, parce que les différents chants d'une espèce sont fonctionnellement équivalents (BEECHER, 2021 ; NELSON et al., 2022), i.e. deux chants différents encodent la même information : attirer un partenaire sexuel, ou repousser un rival du territoire de l'individu (CATCHPOLE et SLATER, 2008 ; de KORT et al., 2009). Ensuite, le chant d'un individu ne varie pas au cours de sa vie chez la majorité des oiseaux, sauf chez les espèces à chant ouvert (BEECHER et BRENOWITZ, 2005 ; ROBINSON et al., 2019). Même chez ces espèces, le chant peut être en grande partie partagé entre individus proches socialement et au sein d'une saison de reproduction, comme chez l'étourneau (ADRET-HAUSBERGER, 1982 ; ADRET-HAUSBERGER et GÜTTINGER, 1984 ; EENS et al., 1992 ; SAINBURG et al., 2020). Quelques espèces, comme le rossignol (*Luscinia megarhynchos*, KIPPER et al., 2004) et le moqueur roux (*Toxostoma rufum*, BOUGHEY et THOMPSON, 1981), possèdent un répertoire de chant de très grande taille, de plusieurs centaines de chants différents à plus d'un millier (BOUGHEY et THOMPSON, 1981 ; KIPPER et al., 2004 ; KROODSMA et PARKER, 1977). Mais seule une fraction des espèces possède un chant qui peut véritablement être considéré comme complexe, avec un grand répertoire d'éléments et une production variable à une échelle de temps courte.

## 1.2 La complexité vocale dans l'histoire de vie d'une espèce

### 1.2.1 Communiquer dans un environnement social complexe

Le but de la communication est de partager des informations avec d'autres individus. L'ajustement comportemental en fonction des informations transmises au cours des communications est vital à la survie de l'individu (ENDLER et BASOLO, 1998 ; STEGMANN, 2013 ). Les espèces sociales ont un niveau supplémentaire d'information à interpréter : la relation entre les individus impliqués (J. W. SMITH, 1965), qui dépend de la complexité sociale de l'espèce.

### 1.2.1.1 Complexité sociale chez les animaux

La notion de complexité sociale, comme la complexité vocale précédemment, peut être basée sur un ensemble hétérogène d'éléments interagissant entre eux de façon répétées et difficilement prévisible, correspondant à une grande quantité d'information dans le système (FREEBERG et al., 2012). De nombreuses études (résumées dans PECKRE et al., 2019) ont considéré que le nombre d'individus conditionne la complexité sociale, et donc que les espèces socialement complexes étaient les espèces vivant en groupe. Cependant, une organisation sociale peut inclure de nombreux individus sans être pour autant complexe, si les interactions entre individus sont fonctionnellement équivalentes (BLUMSTEIN et ARMITAGE, 1997). Par exemple, un essaim de criquets, un vol d'oiseaux, un banc de poissons, peuvent être décrit par des règles simples : 1) rester à proximité du groupe, 2) éviter les collisions, 3) se déplacer dans la même direction que le groupe (COUZIN et al., 2002 ; PARRISH et al., 2002). Pour considérer une organisation sociale comme complexe, la seule taille du groupe ne suffit pas, la nature des relations entre les individus doit également être considérée.

Une organisation sociale est définie par la structure démographique, la structure sociale, et le système de reproduction (KAPPELER et VAN SCHAIK, 2002), et chacun de ces facteurs contribue à la complexité sociale. La structure démographique correspond à la taille et à la composition de l'unité sociale de base : on distingue ainsi la structure solitaire (où l'unité sociale est l'individu), la structure en paires (où l'unité sociale est un couple reproducteur et leur progéniture), et la structure en groupe (qui regroupe toutes les structures où plusieurs individus vivent ensemble, sans être nécessairement apparentés). La structure sociale est définie à plusieurs échelles : l'interaction entre individus, la relation sociale (composée d'interactions répétées entre individus données), et le réseau social (composé par les relations entre les individus d'un groupe) (HINDE, 1976). Le système de reproduction dépend de l'exclusivité de la relation entre des individus reproducteurs : la monogamie n'inclut qu'un seul mâle et une seule femelle, la polyandrie et la polygynie inclut un seul individu (femelle ou mâle respectivement) s'accouplant avec plusieurs individus du sexe opposé (FUENTES, 2007).

La complexité sociale est basée sur des relations entre individus qui interagissent fréquemment entre eux. Cela résulte en une influence des interactions passées sur le résultat d'une interaction présente. En particulier, différentes formes de coopération peuvent se former entre des individus spécifiques ou entre l'ensemble des individus du groupe. La vie en groupe apporte en effet de nombreux bénéfices à un individu dans la protection contre les prédateurs, l'accès aux ressources, ou l'accès à la reproduction (KRAUSE et RUXTON, 2002), mais apporte également le risque de conflits liés à la compétition intra-spécifiques pour ces mêmes ressources, ou simplement pour l'espace vital (CLUTTON-BROCK et HUCHARD, 2013 ; EMLÉN, 1997 ; RUBENSTEIN, 1978 ; SCHRADIN et al., 2010 ; STERCK et al., 1997 ; WEST-EBERHARD, 1979). Les relations sociales, notamment affiliatives (i.e. positives, ALEXANDER, 1974), favorisent la coopération entre les membres du groupe, augmentant ces bénéfices et réduisant les conflits liés à la vie en groupe. Par exemple, le groupe peut coopérer face aux prédateurs par une réponse de *mobbing* (ARROYO et al., 2001 ; BERGER, 1979), coopérer pour l'élevage des juvéniles (BEISNER et al., 2011 ; M.-C. LORETTO et al., 2012 ; SMITH et al., 2010), coopérer pour sécuriser l'accès à la nourriture (C. R. BROWN, 1986 ; DANCHIN et WAGNER, 1997 ; RATCLIFFE et TER HOFSTEDE, 2005 ; RICHNER et HEEB, 1995 ; WARD et ZAHAVI, 1973). De façon encore plus

spécifique, des relations entre deux individus uniquement (ou au moins entre un sous-ensemble des individus du groupe) peuvent influencer la résolution de conflits au sein du groupe (DE VRIES et al., 2006). Cette coopération est fréquemment favorisée chez les individus apparentés (BAGLIONE et al., 2003; EBERLE et KAPPELER, 2006; HAMILTON, 1964; KOMDEUR, 1994a; KOMDEUR, 1994b), mais pas uniquement : des individus non apparentés sont aussi fréquemment affiliés (comme dans les interactions de toilettage mutual communes chez les primates et les oiseaux, BARRETT et al., 1999; MORALES PICARD et al., 2020). Remarquablement, les interactions affiliatives dans un contexte peuvent être généralisées à des interactions affiliatives dans un autre contexte, comme chez les vervets, où des individus qui se toilettent se soutiennent plus lors de conflits (SEYFARTH et CHENEY, 1984); chez les oiseaux coloniaux, des individus voisins de nids peuvent se soutenir lors du *mobbing* face aux prédateurs (KRAMS et FREEBERG, 2006; WHEATCROFT et PRICE, 2008), ou même défendre le nid d'un voisin face à un conspécifique étranger (COOMBS, 1960). La vie en groupe apporte donc des avantages forts, et la complexité sociale permet de les amplifier en favorisant la coopération (FREEBERG et KRAMS, 2015), ce qui a sans doute favorisé sa sélection dans l'évolution.

La complexité sociale a également un coût au-delà de ceux associés à la vie en groupe. En particulier, elle nécessite de garder trace des relations entre les individus pour réagir correctement à leurs comportements. L'environnement social d'un individu peut également changer au cours de la vie, réversiblement ou non, ce qui augmente encore la complexité sociale et donc les coûts cognitifs qui y sont associés. Dans les systèmes de fission-fusion communs à de nombreuses espèces de mammifères et oiseaux (AMICI et al., 2008; ARCHIE et al., 2006; AURELI et al., 2008; COUZIN et al., 2002; SILK et al., 2014; WILLIS et BRIGHAM, 2004), les individus vivant dans un groupe social peuvent se séparer ou, au contraire, fusionner en groupes sociaux de taille variable. Chez certains oiseaux, cela peut également inclure des groupes migrateurs rejoignant des groupes sédentaires à certaines périodes (PATTERSON et al., 1971). Ces systèmes de fission-fusion sont généralement considérés comme extrêmement complexes socialement (FREEBERG et al., 2012; GRUETER et al., 2020; PAPAGEORGIOU et FARINE, 2021), d'autant plus que certaines espèces ajoutent d'autres niveaux de complexité sociale, comme des couples monogames au sein des groupes et d'autres relations entre certains individus (BOUCHERIE et al., 2016).

### 1.2.1.2 Communication dans un environnement social complexe

La plupart des avantages d'une organisation sociale complexe passe par les interactions entre les individus, médiées par des signaux de communication comme les vocalisations (FREEBERG et al., 2012). Les vocalisations permettent à l'individu de s'identifier par sa signature vocale particulière (LINHART et al., 2019), en plus des autres informations encodées (e.g. alarme, nourriture, agression). Ces deux informations peuvent donner lieu à un conflit, puisque le besoin d'identifier un individu devrait favoriser une divergence vocale, tandis que toutes les autres informations devraient favoriser des vocalisations similaires entre les individus pour être intelligibles. Pour éviter ce conflit, certaines espèces produisent des vocalisations spécialisées encodant la signature vocale séparément de celles encodant les autres informations, comme chez les mangoustes (JANSEN et al., 2012; MANSER et al., 2014), les perroquets (CORTOPASSI et BRADBURY, 2006), ou les cétacés (QUICK et JANIK, 2012;

VERGARA et MIKUS, 2019). D'autres espèces pourraient séparer l'information en gardant certaines caractéristiques acoustiques communes entre toutes les vocalisations d'un individu, qui serviraient alors de signature vocale. Il n'est pas clair que cette deuxième option soit utilisée chez les animaux, principalement par manque d'études sur plus d'un type de vocalisation (BENTI et al., 2019; BERG et al., 2012; BOECKLE et al., 2012; BOUCHET et al., 2013; CANDIOTTI et al., 2012; CHARRIER et al., 2001; KEENAN et al., 2020; MATES et al., 2015; PRICE, 2015; SPILLMANN et al., 2017; VIGNAL et al., 2008; YORZINSKI et al., 2006). Les quelques études sur le sujet ont entraîné des algorithmes de classification d'individu sur un type de vocalisation et testé cet algorithme sur un autre type de vocalisations. Cette procédure échoue chez le diamant mandarin (ELIE et THEUNISSEN, 2018), mais réussit chez la corneille d'Amérique (MATES et al., 2015), chez le rhipidure hochequeue (*Rhipidura leucophrys*, E. FOX et al., 2006), le méliphage chanteur (*Gavicalis virescens*, E. FOX et al., 2006) le canari (*Serinus canaria*, E. FOX et al., 2006), bulbul de Chine (*Pycnonotus sinensis*, CHENG et al., 2012), le pouillot de Hume (*Phylloscopus humei*, CHENG et al., 2012), et le pouillot du Sichuan (*Phylloscopus yunnanensis*, CHENG et al., 2012). Cependant, le manque de comparabilité entre ces études, tant du point de vue des espèces que des méthodes utilisées, ne permet pas de tirer de conclusions générales. Chez les primates, le répertoire vocal est fréquemment gradé et commun entre individus, mais certaines vocalisations peuvent être très individualisées, notamment les cris de contact (KEENAN et al., 2020; LEMASSON et al., 2011a), tandis que d'autres cris le sont beaucoup moins, comme les cris d'alarme (KEENAN et al., 2020). Cependant, les cris des primates sont aussi fréquemment utilisés de façon flexible, tant dans la structure (CANDIOTTI et al., 2013) que dans le contexte (REBOUT et al., 2021).

Ce conflit entre signature vocale et intelligibilité est d'autant plus marqué dans les espèces socialement complexes. La signature vocale tend à être plus marquée dans les espèces vivant en groupe social de grande taille (POLLARD et BLUMSTEIN, 2011). Notamment, cette individualité peut être variable en fonction de la population locale ; par exemple, une étude chez la perriche veuve (*Myiopsitta monachus*) a montré que des populations invasives, à la complexité sociale plus faible que ce qui est normalement observé pour l'espèce, produisaient des signatures vocales moins élaborées que des populations natives (SMITH-VIDAURRE et al., 2021). Il est donc possible, en particulier pour les espèces qui apprennent leurs vocalisations, que la signature vocale soit soumise aux besoins liés à la complexité sociale du groupe dans lequel évoluent les individus. La signature vocale permet aux individus de s'identifier entre eux (ELIE et THEUNISSEN, 2018; KONDO et al., 2010). Une conséquence majeure dans les espèces sociales est la formation des relations sociales, mais aussi la favorisation des signaux honnêtes entre conspécifiques. En effet, un individu dont les signaux sont perçus comme mal-honnête par les congénères peut se retrouver isolé du réseau social ou ses signaux ignorés (CHENEY et SEYFARTH, 2018).

La complexité vocale des espèces sociales permet d'encoder des informations subtiles. En particulier, certaines espèces, comme les macaques, utilisent leurs vocalisations de façon flexible en relâchant le couplage structure-contexte (REBOUT et al., 2021). D'autres espèces utilisent leurs vocalisations pour spécifier certaines formes d'information : des systèmes de vocalisations gradées et des combinaisons permettent de moduler l'information transmise (ENGESSER et al., 2018; KEENAN

et al., 2020). Cependant, la complexité de l'information à transmettre augmente également la complexité de son interprétation, comme dans le conflit identité-information dans certaines vocalisations (KEENAN et al., 2020). La complexité vocale permet également de résoudre ce conflit, en autorisant plus de possibilités à l'espèce ; certaines espèces l'utilisent ainsi pour séparer les vocalisations à signature vocale forte et d'autres à contenu informatif fort (JANSEN et al., 2012 ; KEENAN et al., 2020 ; MANSER et al., 2014). La complexité vocale et la complexité sociale semblent ainsi fortement liées dans l'histoire de vie d'une espèce.

### 1.2.1.3 Hypothèse de la complexité socio-communicative

La notion d'une relation entre complexité sociale et complexité vocale (ou plus généralement communicative) a été suggérée dès LAMARCK (1809) et DARWIN (1897). Au vu de la complexité sociale vue dans la section précédente, il paraît logique qu'une espèce socialement complexe a besoin de communiquer des informations subtiles pour que les individus puissent s'échanger des informations mais aussi s'identifier entre eux (KRAMS et al., 2012). Ce "besoin", traduit par l'hypothèse de la complexité socio-communicative (BLUMSTEIN et ARMITAGE, 1997 ; FREEBERG et al., 2012 ; PECKRE et al., 2019), est que la complexité sociale et la complexité vocale auraient été co-sélectionnées au cours de l'évolution. On peut également noter que ces deux complexités pourraient également être liées à des contraintes cognitives fortes ; notamment, le lien entre complexité sociale et cognition a été formalisé par l'hypothèse de l'intelligence sociale (DUNBAR, 1998).

L'hypothèse de la complexité socio-communicative a été corroborée par des très nombreuses études chez diverses espèces (PECKRE et al., 2019). Les primates représentent la famille chez qui ces études ont été de loin les plus détaillées. Une comparaison de 42 espèces de primates a démontré une corrélation positive entre taille du groupe social et taille du répertoire vocal (MCCOMB et SEMPLE, 2005 ; POLLARD et BLUMSTEIN, 2011). D'autres études, sur différentes espèces, ont trouvé des corrélations positives entre différentes mesures de ces complexités sociales et vocales. De façon très générale les espèces vivant en larges groupes multi-mâles et multi-femelles ou en système de fission-fusion, possèdent des répertoires vocaux de plus grande taille (GUSTISON et al., 2012), une structure plus variable des vocalisations (BOUCHET et al., 2013), ou une distribution hétérogène des cris (MACIEJ et al., 2013). En particulier, une étude comparative chez les macaques a montré que les espèces plus tolérantes (macaques de Tonkean *Macaca tonkeana*, macaques noirs *Macaca nigra*), sont vocalement plus complexes que des espèces moins tolérantes (macaques rhésus *Macaca mulatta*, macaques japonais *Macaca fuscata*) : les espèces plus tolérantes ont ainsi un répertoire de vocalisation plus large et l'utilisent de manière plus flexible en fonction du contexte que les espèces moins tolérantes (REBOUT et al., 2021). Les études chez les autres espèces ont également été nombreuses, mais beaucoup, en particulier chez les oiseaux, se sont limitées à mesurer la taille du répertoire vocal et la taille du groupe social ou l'existence de reproduction communautaires comme mesures de complexité (KNÖRNSCHILD et al., 2020 ; KRAMS et al., 2012 ; PECKRE et al., 2019). Bien que ces mesures correspondent à des aspects complexes, elles ne correspondent qu'à une partie des possibilités de complexité.

## 1.2.2 Compétences cognitives dans la complexité vocale

Le comportement d'un animal est éminemment variable au cours de la vie. D'abord, le juvénile doit apprendre à exprimer les comportements typiques de son espèce, puis les affiner au cours de son développement, jusqu'à l'âge adulte. L'étendue de l'apprentissage est extrêmement variable en fonction des espèces.

### 1.2.2.1 Complexité sociale et compétences cognitives

Les contraintes liées à un environnement social complexe nécessitent des compétences cognitives développées (DUNBAR et SHULTZ, 2010; SHULTZ et DUNBAR, 2010). En effet, un individu dans un tel environnement social doit non seulement faire face aux contraintes liées à l'environnement physique, mais aussi s'adapter à celles de l'environnement social : mémoriser les individus partenaires, les interactions passées (et donc les relations sociales associées à chaque individu), les relations de ces partenaires sociaux avec leurs propres partenaires sociaux dans le groupe, etc. L'hypothèse de l'intelligence sociale (DUNBAR, 1998; FREEBERG et al., 2012) propose que ces contraintes ont été à l'origine de l'évolution de compétences cognitives développées chez les espèces sociales. Cette hypothèse a été bien soutenue par des études chez les mammifères (primates : DUNBAR, 1998; hyènes : HOLEKAMP et al., 2007; cétacés : CONNOR, 2007), notamment en lien avec la taille du cortex cérébral relativement à la taille du corps. Il a été proposé que la taille du groupe social est un bon prédicteur du développement des compétences cognitives chez les mammifères, cependant, ce n'est pas le cas chez les oiseaux (BEAUCHAMP et FERNÁNDEZ-JURICIC, 2005; EMERY et al., 2007; SHULTZ et DUNBAR, 2010B). Chez les oiseaux, les compétences cognitives, estimées tant avec des approches comportementales (EMERY et al., 2007) que neuroanatomiques (BURISH et al., 2004), semblent mieux corrélées avec d'autres traits de la complexité sociale, notamment liées à une diversité de relations inter-individuelles : affiliations, coopération, la formation de colonies pour la reproduction, l'apprentissage social, et jeu social (BERGMAN et BEEHNER, 2015; BURISH et al., 2004; EMERY et al., 2007; LAMBERT et al., 2019; WEY et al., 2008).

Ces corrélations positives à l'échelle inter-espèces entre compétence cognitive et complexité sociale d'une part, et complexité sociale et complexité vocale d'autre part, pourraient laisser envisager une même corrélation positive entre compétence cognitive et complexité vocale. Cependant, aucune étude ne semble avoir démontré cette corrélation, en tout cas dans le cas du chant des oiseaux, pourtant l'exemple le plus étudié de complexité vocale chez les animaux non humains (résumé dans SEARCY et NOWICKI, 2019). Aucune espèce ne montre d'association entre compétence cognitive et complexité vocale (BOOGERT et al., 2011; DUBOIS et al., 2018; TEMPLETON et al., 2014), même les espèces à chant ouvert (FARRELL et al., 2012; KEAGY et al., 2011). Seulement chez l'étourneau, les individus avec une meilleure performance sur les tests cognitifs produisent des chants plus longs (FARRELL et al., 2012), ce qui pourrait être considéré comme un chant plus complexe chez cette espèce où le chant est particulièrement variable (EENS et al., 1992), mais cela reste peu clair. Au niveau anatomique, les noyaux cérébraux contrôlant la production et l'apprentissage du chant semble plus développés chez les individus à chant complexe chez certaines espèces, en particulier l'étourneau (DEVOOGD

et al., 1993; DEVOOGD, 2004), mais cela ne semble pas le cas chez d'autres espèces (LEITNER et CATCHPOLE, 2004). En résumé, si la complexité vocale est potentiellement liée à certaines capacités cognitives (puisque'il y a nécessairement un coût cognitif à mémoriser un chant, proportionnellement à sa complexité), cela ne semble pas être un indicateur général de la capacité cognitive. En d'autres termes, des individus cognitivement plus compétents ne sont pas nécessairement des chanteurs plus complexes. En revanche, la compétence cognitive peut influencer, sinon le chant final, du moins son apprentissage.

### 1.2.2.2 Apprentissage des vocalisations

Les modalités d'acquisition du répertoire vocal semblent particulièrement variées en fonction des espèces (NIEDER et MOONEY, 2020). Chez de nombreux mammifères, les vocalisations semblent au moins partiellement innées, comme démontré par des expériences d'adoptions croisées chez les primates (BROCKELMAN et SCHILLING, 1984; OWREN et al., 1992; WINTER et al., 1973) et les souris (KIKUSUI et al., 2011), où un répertoire typique est néanmoins produit même par un animal élevé par une autre espèce. En revanche, les vocalisations des oiseaux semblent presque exclusivement apprises plutôt qu'innées : des expériences d'assourdissement (KONISHI, 2004), d'adoption croisées (CLAYTON, 1990; ERIKSEN et al., 2009; HANSEN et al., 2010; SLAGSVOLD et al., 2002), et d'élevage hors de la présence de conspécifiques (JANIK et SLATER, 2000) ont montré que l'acquisition des vocalisations typiques de l'espèce nécessite la présence de conspécifiques adulte (JANIK et SLATER, 2000). Les vocalisations apprises initialement sont instables, avant que l'individu ne les affine au cours de son développement jusqu'à acquérir son répertoire adulte (CAROUSO-PECK et GOLDSTEIN, 2019; CHEN et al., 2016). Cette acquisition dépend fréquemment d'un renforcement social par imitation des congénères adultes (CHEN et al., 2016). Cette phase d'apprentissage est clôturée, notamment chez les oiseaux, par une phase de cristallisation, au-delà de laquelle le répertoire vocal devient fixe (BEECHER et BRENOWITZ, 2005; HAMMERSCHMIDT et FISCHER, 2008; MARLER, 1997; MARLER et PETERS, 1987; TEN CATE, 2021).

L'apprentissage vocal semble peu fréquent chez les animaux (TYACK, 2020), n'ayant été retrouvé que dans quelques taxons : les oiseaux chanteurs (CATCHPOLE et SLATER, 2008; CHEN et al., 2016; CLAYTON, 1990; MARLER et TAMURA, 1964; THORPE, 1958; VERNES et al., 2021), les perroquets (BRADBURY et BALSBY, 2016; CRUICKSHANK et al., 1993; GRAMZA, 1970; PEPPERBERG, 2006), les colibris (TEN CATE, 2021), les pinnipèdes (REICHMUTH et CASEY, 2014), les cétacés (JANIK, 2014), les éléphants (STOEGER et MANGER, 2014), les chauves-souris (KNÖRNSCHILD, 2014), et les humains (TYACK, 2020). Le maintien de cet apprentissage au-delà de la phase initiale est encore plus rare, restreint à un sous-ensemble des espèces des mêmes groupes (NOWICKI et SEARCY, 2014; TEN CATE, 2021; TYACK, 2020; VERNES et al., 2021; VERPOOTEN, 2021). Deux degrés de complexité de l'apprentissage peuvent être distingués (TYACK, 2020) : l'apprentissage limité, correspondant à des ajustements de vocalisations pré-existantes, et un apprentissage complexe, où l'individu acquiert des vocalisations entièrement nouvelles (TYACK, 2020). L'apprentissage limité permet la convergence vocale chez les espèces sociales tant aviaires que simiennes, qui permet aux individus de signaler leur appartenance à un groupe social (MONTEIRO et al., 2021; MUNDINGER, 1982; NOWICKI, 1989;



WRIGHT et DAHLIN, 2017 ; ZÜRCHER et al., 2021), ou à des couples monogames de rester en contact à distance (LUEF et al., 2017 ; SEWALL, 2009 ; SNOWDON et ELOWSON, 1999). L'apprentissage vocal complexe est fréquemment médié par l'imitation de sons divers, notamment hétérospécifiques, voire abiotiques (MARSHALL, 1950). Cette capacité est notamment développée chez les oiseaux, comme les différentes espèces d'oiseaux lyres (BACKHOUSE et al., 2022 ; DALZIELL et MAGRATH, 2012), l'étourneau (EENS et al., 1992), les coucous (MUNDY, 1973), la majorité des espèces de perroquets (BRADBURY et BALSBY, 2016 ; CRUICKSHANK et al., 1993 ; PEPPERBERG, 2006), ou les corvidés (E. D. BROWN, 1985 ; BROWN et FARABAUGH, 1997), mais a également été documentée chez les dauphins (HOOPER et al., 2006 ; REISS et MCCOWAN, 1993).

Une distribution aussi hétérogène des capacités d'apprentissage suggère une convergence évolutive plutôt qu'une origine commune (TYACK, 2020). Par exemple, parmi les primates seuls les humains paraissent capables d'apprentissage vocal complexe (VERNES et al., 2021). Cette convergence suggère un fort avantage évolutif lié à l'apprentissage vocal, et plusieurs hypothèses ont été proposées pour expliquer cette évolution (NOWICKI et SEARCY, 2014 ; SEWALL, 2009). Apprendre les vocalisations des partenaires sociaux, comme dans les phénomènes de dialectes et de convergence vocale, permet aux individus de s'identifier comme faisant partie du même groupe, favorisant un isolement reproducteur entre dialectes différents (BAKER, 1982 ; SLABBEKOORN et SMITH, 2002 ; D. WANG et al., 2022). Alternativement, l'apprentissage de nouvelles vocalisations peut conduire à une meilleure individualisation vocale, permettant aux individus de se reconnaître plus facilement, notamment dans des environnements sociaux complexes (MCCOMB et SEMPLE, 2005 ; POLLARD et BLUMSTEIN, 2011). Une étude chez le Cacatoès ros-albin (*Eolophus roseicapilla*) a montré un cas extrême à cet effet : cette espèce converge vocalement très rapidement, ce qui est utilisé par les individus qui interagissent pour s'adresser spécifiquement à leur interlocuteur (SCARL et BRADBURY, 2009). Une troisième possibilité est l'adaptation des vocalisations à l'environnement pour améliorer leur transmission en cas d'interférences (DEONIZIAK et OSIEJUK, 2019 ; JIANG et al., 2019 ; SLABBEKOORN, 2004). Enfin, l'apprentissage de nouvelles vocalisations peut aussi permettre d'augmenter la taille du répertoire, facilitant la transmission d'informations plus variées, mais aussi augmentant l'attractivité de l'individu pour la reproduction (CATCHPOLE et SLATER, 2008 ; NOWICKI et SEARCY, 2005, 2014 ; ROBINSON et al., 2019). Cependant, aucune hypothèse seule ne semble s'appliquer à la majorité des espèces connues (LACHLAN et SLATER, 1999). Une autre possibilité est que la sélection de l'apprentissage par imitation est un épiphénomène. Un répertoire complexe est avantageux pour l'accès à la reproduction (NOWICKI et SEARCY, 2005, 2014 ; ROBINSON et al., 2019) sans qu'il soit nécessairement avantageux dans l'histoire de vie de l'animal (LACHLAN et SLATER, 1999 ; WINTERS, 2019). Par exemple, les différents chants utilisés dans la reproduction, quand l'individu en possède plusieurs, sont fonctionnellement équivalents (NELSON et al., 2022), et sont coûteuses en énergie et en détectabilité par les prédateurs (RYAN et al., 1982 ; ZAHAVI, 1975), mais un répertoire complexe est néanmoins privilégié pour la reproduction, car *a priori* plus attractif pour les femelles.

Enfin, l'apprentissage vocal par imitation est par définition un phénomène social, que l'individu imité soit un congénère ou d'une autre espèce. La propagation d'une vocalisation, d'un dialecte ou d'un chant à travers une, voire plusieurs populations, passe par des interactions vocales (ENGGIST-

DUEBLIN et PFISTER, 2002 ; KONDO, 2021 ; MARLER et TAMURA, 1964 ; WILLIAMS et LACHLAN, 2022 ; WRIGHT et DAHLIN, 2017). Ces interactions, et donc la transmission de ces productions vocales, ont lieu de façon largement indépendante de la sélection naturelle. Il s'agit plutôt d'un cas de transmission culturelle, permettant l'établissement de traditions vocales si les vocalisations se maintiennent dans la population (LOGUE et LECA, 2020 ; MARLER et TAMURA, 1964 ; WILLIAMS et LACHLAN, 2022 ; WRIGHT et DAHLIN, 2017). Ce phénomène culturel peut bien entendu mener à des barrières reproductrices qui permettent l'action de la sélection naturelle. Ainsi, plusieurs espèces d'oiseaux s'accouplent préférentiellement avec des partenaires produisant le même dialecte (MILLIGAN et VERNER, 1971 ; NELSON et POESEL, 2007 ; WRIGHT et DAHLIN, 2017). En cas d'adoption croisée, cela peut conduire le juvénile, ayant appris les chants de son espèce d'adoption, à la préférer à son espèce d'origine lors de la reproduction (CLAYTON, 1990). Certaines espèces imitent également, au cours de leur histoire de vie, les vocalisations des autres espèces présentes dans leur habitat (MARSHALL, 1950 ; RIEGERT et JŮZLOVÁ, 2018). Ces vocalisations peuvent être incorporées au chant de l'espèce, perdant donc leur fonctionnalité d'origine (DALZIELL et MAGRATH, 2012 ; RIEGERT et JŮZLOVÁ, 2018). Cependant, cela n'est pas toujours le cas : le drongo brillant imite ainsi les cris d'alarme d'autres espèces pour les éloigner d'une source de nourriture, et peut varier le cri utilisé pour éviter les effets d'habituation (FLOWER et al., 2014). Cette utilisation implique en outre une compréhension de la vocalisation (au moins par l'effet que cette vocalisation a sur l'espèce d'origine) et un certain contrôle par l'individu lui-même.

### 1.2.2.3 Contrôle volontaire de la production vocale

La production vocale est influencée par l'état interne de l'individu (BRIEFER, 2012 ; CYNX et al., 1990) et par son environnement (comme l'effet Lombard, CYNX et al., 1998 ; JIANG et al., 2019 ; MANABE et al., 1998 ; ZOLLINGER et BRUMM, 2011). Certaines vocalisations en sont indissociables (TOMASELLO, s. d.), comme le rire ou le pleur chez le nouveau-né humain (OLLER et al., 2013). Cependant, une partie des vocalisations de certaines espèces peuvent être produites de manière flexible, et notamment dans une variété de contextes (CLAY et al., 2015 ; MARLER, 1997). Cette flexibilité implique un découplage entre la vocalisation et son contexte, et donc un choix possible par l'individu. Cela peut se manifester par un choix dans la production d'une vocalisation, ou du moment où une vocalisation est produite (ADRET, 1993 ; BRECHT et al., 2019 ; HAGE et al., 2013 ; JANIK et SLATER, 2000 ; VEIT et al., 2021). En conditions naturelles, il n'est pour l'instant pas possible de déterminer avec certitude si une vocalisation est produite de manière volontaire ou non. En laboratoire cependant, cette capacité peut être démontrée par conditionnement opérant, et a donc été démontrée chez plusieurs primates (AITKEN et WILSON, 1979 ; GHAZANFAR et al., 2019 ; HAGE et al., 2013 ; KODA et al., 2007), les chauve-souris (LATTENKAMP et al., 2018), et un corvidé, la corneille d'Amérique (BRECHT et al., 2019). Les perroquets sont une autre famille où la production des vocalisations peut être considérée comme volontaire et fortement associée aux capacités cognitives et d'apprentissage. Le perroquet Alex, un gris du Gabon (*Psittacus erithacus*), a été capable d'apprendre un vocabulaire de quelques centaines de mots anglais et de les utiliser pour répondre à certaines questions (PEPPERBERG, 2006 ; PEPPERBERG et CAREY, 2012). D'autres espèces pro-

duisent leurs vocalisations avec une grande flexibilité fonctionnelle, mais ces démonstrations restent limitées aux études chez les macaques (REBOUT et al., 2020) et les perroquets (MONTES-MEDINA et al., 2016). La capacité à choisir quelle vocalisation produire ou quand la produire semble donc être effectivement présente, comme nous l'aurions attendu, chez des espèces cognitivement douées et socialement complexes. Chez les macaques, les espèces moins socialement complexes semblent être limitées dans leur flexibilité (REBOUT et al., 2020), mais des études chez les oiseaux manquent pour conclure sur l'impact de la complexité sociale sur le contrôle cognitif des vocalisations.

Outre la flexibilité fonctionnelle, un individu peut choisir quand produire ou non une vocalisation. L'exemple le plus frappant de ce contrôle vocal sont les duos et prises de paroles à tour de rôle (*turn-taking*, ADRET, 2022 ; BANERJEE et VALLENTIN, 2022 ; PAYNE et PAYNE, 2010 ; PIKA et al., 2018 ; RAVIGNANI et al., 2019 ; ROGERS et al., 2006 ; THORPE, 1963). Chez les oiseaux, ces duos correspondent le plus souvent à des véritables échanges antiphoniques (un individu vocalise, le deuxième lui répond, le premier vocalise à nouveau, le deuxième lui répond, etc.), moins fréquemment à des formes plus lâches (DAHLIN et BENEDICT, 2014). Ces comportements sont principalement exprimés entre membres d'un même couple (DAHLIN et BENEDICT, 2014 ; ELIE et al., 2010 ; HALL, 2009), où ils ont souvent une fonction affiliative entre les membres du couple (ELIE et al., 2010), et plus rarement de protection du couple face à des rivaux (DAHLIN et BENEDICT, 2014). La précision temporelle du duo est notamment liée au contact social, augmentant lorsque le couple est séparé (PEREZ et al., 2015). La précision nécessaire à cette coordination est acquise au cours du développement du juvénile (CHOW et al., 2015 ; LEMASSON et al., 2011a ; RIVERA-CÁCERES et al., 2018), tout comme le reste du répertoire vocal, impliquant donc un processus d'apprentissage qui impliquerait un meilleur contrôle chez les adultes.

Un individu peut également s'interrompre pendant sa production vocale si nécessaire, jusqu'à une certaine limite. Plusieurs expériences d'interruptibilité, où un stimulus, comme un flash lumineux, un souffle d'air, ou un son sont dirigés vers un animal en cours de vocalisation, permettent de mettre au jour la résolution temporelle à laquelle l'individu peut s'arrêter si besoin. Les premières expériences sur les chants de plusieurs espèces d'oiseaux (diamant mandarin : CYNX et al., 1990 ; rossignol : RIEBELL et TODT, 1997) ; pigeon : TEN CATE et BALLINTIEN, 1996) et de primates (tamarins : C. T. MILLER et al., 2003 ; marmouset : POMBERGER et al., 2018) ont permis de déterminer les unités élémentaires (syllabes) dans le répertoire vocal de ces espèces : ce sont les unités que l'individu ne peut pas interrompre. Il est notable que dans aucun cas, l'individu ne peut s'interrompre entièrement librement : les syllabes initiées doivent être achevées. Cela démontre que même si un contrôle sur la production des vocalisations existe, celui-ci reste soumis à certaines contraintes (C. T. MILLER et al., 2003 ; TEN CATE et BALLINTIEN, 1996).

### 1.2.3 Innovations dans la production vocale

Le répertoire vocal offre un grand potentiel de variation, entre l'apprentissage des vocalisations, leur flexibilité, et les combinaisons possibles de signaux. Le besoin d'intelligibilité impose des contraintes sur cette variation (THOMSEN et al., 2019 ; ZÜRCHER et al., 2021). Cependant, les chants d'oiseaux, vocalisations particulièrement complexes du répertoire, sont fonctionnellement équivalents (NELSON

et al., 2022), et l'accès à la reproduction est favorisé par un répertoire de chant complexe (BEECHER et BRENOWITZ, 2005 ; CATCHPOLE et SLATER, 2008 ; MARLER et SLABBEKOORN, 2004). Cela devrait favoriser des stratégies conduisant à la maximisation de la complexité du chant à partir du répertoire dont dispose un individu. Malgré cela, la majorité des espèces produit largement les mêmes chants tout au long de la vie, ou au minimum au sein d'une saison de reproduction (ADRET-HAUSBERGER et GÜTTINGER, 1984 ; BEECHER et BRENOWITZ, 2005), plutôt que d'explorer les possibilités offertes par leurs capacités vocales. Quelles pourraient être les mécanismes expliquant cette constance, et en quelles circonstances peuvent-ils être dépassés pour permettre aux animaux d'innover vocalement ?

### 1.2.3.1 Maintien du répertoire de chant

Malgré la variation potentielle permise, d'une part par l'apprentissage vocal, et d'autre part par la combinaison de plusieurs vocalisations dans un chant (CATCHPOLE et SLATER, 2008 ; MARLER et SLABBEKOORN, 2004), la plupart des oiseaux produisent toujours les mêmes chants au cours de leur vie (BEECHER et BRENOWITZ, 2005), et même les espèces à chant ouverts produisent généralement les mêmes chants au cours du temps (ADRET-HAUSBERGER et GÜTTINGER, 1984 ; SAINBURG et al., 2019). L'acquisition du chant dépend largement de la présence de conspécifiques adultes que le juvénile peut imiter (BEECHER, 2021 ; CAROUSO-PECK et GOLDSTEIN, 2019). En l'absence de ces conspécifiques, ou si pour toute autre raison l'individu ne peut apprendre un chant typique, son succès reproducteur peut s'en voir diminué (NOWICKI et al., 2002a). Chez le bruant chanteur par exemple, des mâles juvéniles non exposés à des mâles adultes chanteurs, et qui sont donc contraints d'inventer une partie de leur chant pendant l'apprentissage, sont moins choisis par les femelles pour la reproduction par rapport à des juvéniles qui ont appris la chanson naturelle de leurs conspécifiques (NOWICKI et al., 2002a). Il y a donc une sélection sexuelle pour des chants typiques, notamment ceux correspondant à un dialecte commun aux deux partenaires (MARLER et TAMURA, 1962 ; MCGREGOR, 1980 ; SEARCY, 1992). Cette préférence favorise un certain conformisme culturel dans l'apprentissage vocal (R. F. LACHLAN et al., 2018). Cette préférence pour le conformisme est elle-même conditionnée par l'apprentissage : ainsi, les mêmes femelles diamant mandarin ne préfèrent ces chants typiques que si elles y ont été exposés en tant que juvéniles (LAUAY et al., 2004). Il y a donc une certaine inertie culturelle dans la production vocale à cause de cette sélection sexuelle réciproque.

Le maintien d'une structure acoustique constante du chant est également opérée par l'individu lui-même par une correction active des erreurs de production. Cette correction active a principalement été étudiée chez le diamant mandarin, par deux mécanismes possibles : un mécanisme de compensation basé sur l'audition, et un mécanisme potentiel basé sur l'activité cérébrale pendant le sommeil. Ainsi, SOBER et BRAINARD (2009) ont montré que la perception auditive du chant influençait directement la production des chants suivants : en exposant des individus à leurs propres chants, mais décalés en timbre pour être plus graves ou plus aigus, les chants suivants étaient alors produits de façon à compenser le décalage (i.e. plus aigu pour une exposition à un chant décalé vers les graves, plus grave pour une exposition à un chant décalé vers les aigus). Une autre étude sur l'activité cérébrale (YOUNG et al., 2017) a montré que les régions cérébrales associées à la production du chant étaient

actives pendant le sommeil, mais avec une variation beaucoup plus élevée que pendant la véritable production de chants. La fonction de cette activité cérébrale pourrait correspondre à un deuxième mécanisme de correction d'erreurs, mais cela reste peu clair.

### 1.2.3.2 Différentes motivations du chant

Si la qualité de l'apprentissage du chant est soumise à la sélection sexuelle, alors la question inverse se pose : pourquoi un chant complexe serait-il avantageux ? La réponse pourrait être dans les différentes motivations derrière la production du chant.

Le chant est généralement considéré comme uniquement associé avec la défense du territoire face aux rivaux, ou l'attraction d'un partenaire sexuel (CATCHPOLE et SLATER, 2008 ; CHEN et WIENS, 2020). Ces chants sont généralement produits selon un certain schéma, différent selon le contexte. Les chants territoriaux peuvent être adaptés au cours d'interactions entre des individus rivaux par le contre-chant, soit en favorisant les réponses similaires, soit en les évitant. Dans ce cas, un répertoire de chant complexe, notamment par sa taille, permettrait à un individu de mieux interagir avec ses voisins : si le contre-chant permet de déplacer la rivalité vers une interaction purement vocale (plutôt qu'un combat physique où les individus risquent des blessures, voire la mort), il est avantageux de pouvoir s'adapter au rival (BEECHER et CAMPBELL, 2005 ; SHACKLETON et RATCLIFFE, 1994). Quant aux chants produits pour attirer des partenaires, ils sont fréquemment répétés pour favoriser leur détection par un partenaire sexuel (PRICE, 2013). Les mécanismes qui favoriseraient leur variation restent cependant discutés : l'hypothèse la plus simple est la sélection sexuelle pour un répertoire complexe, qui servirait d'indicateur de qualité de l'individu chanteur (NOWICKI et al., 1998 ; NOWICKI et SEARCY, 2005, 2014 ; ROBINSON et al., 2019). Cependant, la variation a également été proposée comme mécanisme pour éviter une habituation chez les receveurs qui diminuerait l'effet attractif du chant (HARTSHORNE, 1956 ; INCE et SLATER, 1985 ; KROODSMA, 1978), éviter la fatigue chez l'émetteur qui aboutirait à une dérive progressive dans la structure du chant (LAMBRECHTS et DHONDT, 1987 ; POESEL et KEMPENAERS, 2000), ou une perte de motivation à chanter le même chant de façon répétée (POESEL et KEMPENAERS, 2000 ; WEARY et al., 1988). Cependant, ces hypothèses sont encore très discutées dans la littérature (CATCHPOLE et SLATER, 2008).

Une autre possibilité, qui a été soulevée plusieurs fois dans la littérature, mais semble gagner en popularité récemment, serait que certains chants ne sont pas produits comme un moyen vers une fin, mais comme une fin en soi. Certaines espèces produisent en effet des chants hors de ces contextes classiques, dits non-orientés (*undirected songs*, DUNN et ZANN, 1996 ; KNEUTGEN, 1969 ; RITERS et al., 2022 ; RITERS et al., 2019 ; SOSSINKA et BÖHNER, 1980). En particulier, chez les étourneaux, ces chants, produits dans des contextes sociaux, sont liés à la voie de récompense dopaminergique (STEVENSON et al., 2020) : la production de ces chants active la voie, et réciproquement, des injections de dopamine induisent la production de ces chants. Remarquablement, la voie dopaminergique semble être nécessaire à un bon apprentissage vocal : des moineaux du Japon (*Lonchura striata*) chez qui cette voie est rendue inopérante par des lésions cérébrales, l'apprentissage vocal est complètement inhibé, sans toutefois modifier la production des chants déjà appris (HOFFMANN et al., 2016). Les chants non-orientés ont également une structure différente : ils sont fréquemment plus

lents (SOSSINKA et BÖHNER, 1980), et leur structure est moins stéréotypée même chez les espèces à chant autrement fixe, comme le diamant mandarin (DUNN et ZANN, 1996). En revanche, ces chants ressemblent au “sous-chant” (*subsong*), i.e. des séries de vocalisations produites par les juvéniles pendant la phase d'apprentissage (KNEUTGEN, 1969). Chez le shama à croupion blanc (*Kittacincla macrura*) les différents types de chants sont associés à des postures différentes : les chants territoriaux à une posture dressée avec les ailes écartées et orientées vers un individu rival, les chants reproducteurs à une posture dressée, les ailes repliées et queue également dressée, mais les chants non-orientés sont, eux, associés à une posture beaucoup plus proche du repos, assis ou semi-assis avec les pattes repliées sous le corps (KNEUTGEN, 1969). Notamment, ces derniers sont produits indépendamment de tout contexte social, que l'oiseau soit seul en chambre anéchoïque ou avec son groupe social. Les chants non-orientés correspondent donc sans doute à des états émotionnels beaucoup plus calmes, potentiellement à valence positive comme suggéré par le lien avec la voie de récompense dopaminergique chez les étourneaux. Ceci, associé à la structure beaucoup plus libre qui semble commune à ces chants non-orientés chez les espèces qui les produisent, pourrait faire de ces chants un excellent terrain d'innovation vocale pour l'oiseau.

### 1.2.3.3 Variations, idiosyncrasies, créativité et innovation dans le chant

L'apprentissage vocal est généralement considéré comme fidèle et précis, sauf pour des erreurs de copies, généralement considérées comme délétères (NOWICKI et al., 2002b ; SLATER et LACHLAN, 2003). Cependant, l'existence des chants non-orientés suggère que ce n'est pas toujours le cas. Si les variations vocales ne sont pas délétères (sans nécessairement être avantageuses), l'introduction de variations dans le répertoire, par apprentissage de nouvelles vocalisations, modifications de vocalisations pré-existantes, ou variations dans la vitesse de production du chant, devrait donner lieu à des idiosyncrasies dans le répertoire d'une espèce, i.e. des différences inter-individuelles. Ces idiosyncrasies pourraient être la marque d'innovations vocales et d'une certaine créativité de l'espèce.

L'innovation dans le chant a été observée uniquement chez quelques espèces d'oiseaux (GARLAND et MCGREGOR, 2020 ; MORI et al., 2018 ; NOWICKI et al., 2001), ainsi que chez certaines baleines (GARLAND et MCGREGOR, 2020). Chez les oiseaux testés, ces innovations ont principalement été permises par des privations sensorielles, i.e. des juvéniles élevés en l'absence de conspécifiques auprès de qui apprendre. Chez le bruant chanteur, ces variations sont délétères : l'innovation vocale défavorise l'accès à la reproduction en raison des préférences des femelles de l'espèce (NOWICKI et al., 2001 ; NOWICKI et al., 2002b). Cela pourrait expliquer que cette espèce est à chant fixe : la sélection sexuelle est conservatrice des structures vocales apprises. Chez le canari, une espèce à chant ouvert, des individus assourdis récapitulent leur répertoire d'une année sur l'autre mais inventent également de nouvelles vocalisations en proportion non négligeable (environ 40%, MORI et al., 2018).

Etudier les idiosyncrasies, et plus généralement l'innovation dans le comportement animal, est particulièrement difficile : comment identifier une innovation ? Les différences d'organisation du chant retrouvées entre espèces, même phylogénétiquement proches, impliquent que ces innovations existent et ont pu se transmettre, probablement par le même mécanisme d'apprentissage vocal que pour le reste du répertoire (LOGUE et LECA, 2020 ; NOWICKI et al., 2001), mais observer ces inno-

vations est, par leur définition même, difficile : comment confirmer qu'un comportement jusque-là non observé est bien nouveau, dans le sens qu'il ne dérive pas d'un apprentissage social auprès d'un autre individu (R. L. BROWN, 2014 ; RAMSEY et al., 2007a, 2007b) ? Une définition possible a été proposée par RAMSEY et al. (2007a) : un comportement peut être considéré comme une innovation si : 1) il est apparu chez un individu à un instant donné, 2) ce comportement a une fonction écologique ou sociale, 3) ce comportement peut être transmis à d'autres individus, et 4) ce comportement peut se répandre dans une population. Cependant, cette définition, en particulier le critère 4, est trop limitant : par définition, une idiosyncrasie nouvelle ne serait alors pas une innovation ? et dans le cas du chant, puisque chaque chant est fonctionnellement équivalent (NELSON et al., 2022), peuvent-ils être considérés comme des innovations ?

La première condition est sans doute un potentiel pour une forte flexibilité comportementale dans l'espèce d'étude. Ensuite, cette espèce devrait être vocalement active (pour étudier les idiosyncrasies vocales), et produire des vocalisations suffisamment complexes, mais aussi stables, pour pouvoir observer d'éventuelles innovations vocales. Chez les oiseaux, peu d'espèces ont été suffisamment étudiées pour répondre à ces critères. Parmi elles, seuls les corvidés et les perroquets semblent suffisamment flexibles dans leur comportement.

### 1.3 La complexité socio-communicative chez les oiseaux

L'hypothèse de la complexité socio-communicative a été testée chez de très nombreuses espèces. Cependant, à part chez les primates, les mesures de complexité sociale et vocale chez les autres espèces ont dans l'ensemble été concentrées sur des mesures de complexité partielles, probablement par manque de définitions opérationnelles (PECKRE et al., 2019). En particulier, la complexité sociale des oiseaux a majoritairement été mesurée par la taille du groupe social et par le fait que l'espèce se reproduise en communauté ou non. Ces mesures ne sont pas nécessairement les plus appropriées, mais l'étude de la complexité sociale chez les oiseaux a également été compliquée par le caractère monogame de l'immense majorité des espèces (GREENWOOD, 1980), qui pourrait réduire le potentiel d'émergence d'une société complexe. Néanmoins, de nombreuses espèces d'oiseaux forment des groupes sociaux multi-échelles avec des relations inter-individuelles fortes (GRUETER et al., 2020 ; PAPAGEORGIU et FARINE, 2021), de façon très similaire aux sociétés observées chez les primates sociaux (REBOUT et al., 2020). Malgré cela, relativement peu d'espèces ont été étudiées pour leur complexité sociale, notamment comparé aux primates. Parmi ces espèces, les corvidés et les perroquets se détachent, non seulement par leur complexité sociale, mais aussi par leurs compétences cognitives élaborées, qui leur ont permis de coloniser de très nombreuses niches écologiques. Ces espèces ne sont pas les seules parmi les oiseaux à posséder l'une ou l'autre des compétences cognitives étudiées chez les primates, mais elles sont potentiellement les seules à les mettre en œuvre d'une manière qui est souvent comparée à celle des primates. Cependant, la complexité vocale de ces espèces est mal connue.

### 1.3.1 Socio-écologie

#### 1.3.1.1 Socio-écologie générale

Les corvidés et les perroquets ont de nombreuses caractéristiques communes dans leurs histoires de vie. La plupart des espèces des deux familles ont une espérance de vie longue (15 à 20 ans pour les corvidés et les petits perroquets comme les perruches, jusqu'à 80 pour les plus grandes espèces de perroquets) permettant la formation de liens inter-individuels forts. Les perroquets comme les corvidés sont des espèces généralistes, occupant de nombreuses niches écologiques dans des environnements extrêmement différents tout autour du globe, notamment grâce à un régime alimentaire largement omnivore et opportuniste (BRADBURY et BALSBY, 2016; EMERY et CLAYTON, 2004; EMERY et al., 2007; GOODWIN, 1986) mais aussi à leur forte flexibilité comportementale (EMERY et al., 2007; SOL et al., 2005). La reproduction est nidicole, avec une longue période de soin aux jeunes. Ceux-ci restent dépendants des parents pendant une longue période. Le soin aux jeunes est fréquemment partagé au moins entre les deux parents (EMERY et al., 2007), avec l'assistance d'autres individus chez certaines espèces, comme les jeunes d'une saison de reproduction précédente comme chez la corneille noire (BAGLIONE et al., 2002), ou d'autres membres du groupe social comme chez certaines perruches (THEUERKAUF et al., 2009) et le geai à gorge blanche (*Aphelocoma coerulescens*, SCHOECH et al., 1991). Selon les espèces, les juvéniles, une fois indépendants, se dispersent et forment des groupes sociaux de juvéniles ou peuvent intégrer directement un groupe social comprenant des adultes.

Une deuxième observation socio-écologique non négligeable est l'importance du jeu dans ces espèces (O'HARA et AUERSPERG, 2017). Tant les perroquets que les corvidés pratiquent le jeu, avec des objets mais aussi en coordination avec leurs congénères, et ce tout au long de leur vie (DIAMOND et BOND, 2003; EMERY et CLAYTON, 2015; FICKEN, 1977; OSVATH et SIMA, 2014). Chez les corvidés, le jeu d'objet en particulier semble néanmoins diminuer avec l'âge (AUERSPERG et al., 2015; AUERSPERG et al., 2014; KENWARD et al., 2011; O'HARA et AUERSPERG, 2017), peut-être en corrélation avec leur tendance plus prononcée à la néophobie, l'évitement des objets et situations inconnus (METTKE-HOFMANN et al., 2002; R. MILLER et al., 2022). Les corvidés sont ainsi plus prudents dans leur exploration de l'inconnu, mais cette prudence est largement atténuée par la présence de congénères (R. MILLER et al., 2022), surtout lorsque ceux-ci sont familiers avec la situation en question : les corbeaux freux, par exemple, suivent l'exemple de leurs congénères lorsqu'ils sont exposés à une nourriture inconnue d'eux mais familière au congénère (DALLY et al., 2008). Tant les corvidés que les perroquets démontrent une forte intelligence sociale, en plus de leur flexibilité comportementale.

Malgré leurs similarités socio-écologiques, les perroquets et les corvidés présentent des différences majeures. En particulier, l'écologie des corvidés est en grande partie tournée vers l'accumulation de nourriture, qu'ils dissimulent en différents endroits (de KORT et CLAYTON, 2006; GRODZINSKI et CLAYTON, 2010). Une grande partie de leur comportement concerne ainsi ces caches, tant pour leur dissimulation, la remémoration de leurs emplacements, et le pillage des caches faites par d'autres individus et les conflits qui en découlent. Un exemple frappant de ce comportement est le cassenoix d'Amérique (*Nucifraga columbiana*) : cette espèce cache sa nourriture principale, des graines de pins,



pendant l'été, en quantités allant jusqu'à plusieurs dizaines de milliers de graines réparties en centaines d'emplacement, et les retrouve en hiver avec une grande précision, malgré les modifications de l'environnement liées au changement de saison (BALDA et KAMIL, 1992; BALDA et TUREK, 1984; VANDER WALL, 1982). De plus, les corvidés semblent doués de compétences impressionnantes en termes de théorie de l'esprit (BUGNYAR et al., 2016), en particulier la capacité à mémoriser l'identité des individus pillant leurs caches (EMERY, 2004) et à dissimuler leurs comportements de cache aux yeux d'observateurs potentiels (BUGNYAR et al., 2016). Leur monde social potentiellement plus compétitif que chez les perroquets en fait une espèce d'intérêt majeur pour l'étude de leurs capacités socio-communicatives. Enfin, pour des raisons pratiques d'accès à l'espèce en extérieur, mon travail s'est focalisé sur les corvidés, même si la forte flexibilité comportementale des perroquets en ferait également un groupe à fort intérêt pour l'étude de l'évolution de la complexité vocale.

### 1.3.1.2 Organisation sociale : du couple à la colonie

L'organisation sociale de pratiquement toutes les espèces de corvidés est, à son niveau le plus simple, basée sur la relation de couple (comme une majorité des espèces d'oiseaux, GREENWOOD, 1980). Cependant, ces relations de couple ont pour particularité leur constance : les membres d'une paire restent ensemble tout au long de l'année, pas seulement lors de la reproduction, et à très long terme, potentiellement toute la durée de vie des partenaires (CLAYTON et EMERY, 2007; de KORT et al., 2006; FRASER et BUGNYAR, 2010; HOLZHAIDER et al., 2011). Chez les corvidés, cette relation de couple est extrêmement forte comparée aux autres relations sociales (BOUCHERIE et al., 2016; de KORT et al., 2009; EMERY et al., 2007) : les partenaires d'un couple dirigent la majorité de leurs interactions affiliatives vers l'autre, et chez les espèces cachant de la nourriture, un individu ne tolère le pillage de ses caches que par le partenaire (DALLY et al., 2006; de KORT et al., 2006).

Au-delà de cette relation de couple, l'organisation sociale est extrêmement variable en fonction des espèces, des conditions du milieu, ou du stade de l'histoire de vie (BRAUN et BUGNYAR, 2012; BROWN et FARABAUGH, 1997; CLAYTON et EMERY, 2007). Certaines espèces forment des paires territoriales, d'autres des groupes sociaux mais néanmoins territoriaux envers les autres groupes, et d'autres peuvent former des systèmes de fission-fusion particulièrement complexes, formant des groupes sociaux stables non-territoriaux qui fusionnent en colonie, pour dormir ou se reproduire (BRAUN et BUGNYAR, 2012; BROWN et FARABAUGH, 1997; CLAYTON et EMERY, 2007). Même au sein d'une espèce, cette organisation sociale peut être très flexible. Par exemple, chez la corneille noire, les juvéniles indépendants des parents et les adultes non appariés suivent une dynamique de fission-fusion (BRAUN et BUGNYAR, 2012; M. C. LORETTO et al., 2017), mais les adultes appariés forment le plus souvent des paires territoriales (HEINRICH, 1999). Cependant, si les ressources du milieu sont insuffisantes pour permettre l'établissement de ces territoires, l'espèce peut aussi former des communautés de reproduction (BAGLIONE et al., 2002; FRASER et BUGNYAR, 2010).

Entre relations de couple, colonies et variabilité, l'organisation sociale de nombreux corvidés est complexe, avec plusieurs échelles d'organisation spatiales et temporelles (BOUCHERIE et al., 2016; SEED et al., 2007). Notamment, des réseaux d'alliance se forment entre certains individus. Fréquemment formées entre individus apparentés ou appariés (EMERY et al., 2007), ces relations ont également

été décrites entre individus qui ne le sont pas (grand corbeau : FRASER et BUGNYAR, 2010; corbeau freux : BOUCHERIE et al., 2016). Ces relations impliquent des comportements affiliatifs comme le toilettage ou le contact physique (MORALES PICARD et al., 2020), mais aussi des coalitions dans les conflits inter-individuels (CLAYTON et EMERY, 2007; FRASER et BUGNYAR, 2012). La différence majeure avec les primates est dans la résolution des conflits. Contrairement aux primates (de WALL et van ROOSMALEN, 1979), les corvidés ne démontrent pas de réconciliation après les conflits (FRASER et BUGNYAR, 2010; MOHAMMED BENKADA et al., 2020). Certaines espèces montrent des interactions de consolation entre la victime du conflit et un individu extérieur (grand corbeau, FRASER et BUGNYAR, 2012), mais d'autres résolvent les conflits par évitement entre la victime et l'agresseur (corbeau freux, MOHAMMED BENKADA et al., 2020; SEED et al., 2007). La réconciliation permet d'éviter des conflits répétés entre les individus qui peuvent conduire à des effets délétères sur le groupe, comme la séparation de certains membres, ou les blessures. Cependant, chez les corvidés, le couple est la relation sociale principale et est très rarement conflictuelle (MOHAMMED BENKADA et al., 2020; SEED et al., 2007). Par conséquent, la réconciliation n'est pas avantageuse. L'environnement social des corvidés est donc particulièrement complexe du point de vue de la variété des relations entre les congénères.

### 1.3.1.3 Socio-écologie du corbeau freux

Le corbeau freux fait partie des espèces de corvidés les plus sociables. Les couples d'individus intègrent des groupes sociaux stables de quelques dizaines d'individus vivant ensemble toute l'année (CLAYTON et EMERY, 2007). Ces groupes sociaux fonctionnent en système de fission-fusion (BOUCHERIE et al., 2016; CLAYTON et EMERY, 2007; PATTERSON et al., 1971; RÖELL et BOSSEMA, 1982), formant régulièrement des colonies pour dormir ou se reproduire, qui peuvent compter plusieurs dizaines de milliers d'individus (CLAYTON et EMERY, 2007). La taille de la colonie dépend principalement des ressources disponibles et de la population locale de corbeaux freux, avec plusieurs colonies entrant en compétition pour les ressources, y compris pour les sites de nidification (GRIFFIN, 1998; GRIFFIN et THOMAS, 2000). Les groupes sociaux restent associés au sein de la colonie, s'en séparant régulièrement pour chercher de la nourriture pendant la journée (HATTORI et al., 2022; PATTERSON et al., 1971). Les colonies de corbeaux freux sont fréquemment associées à d'autres espèces de corvidés, principalement les choucas (COOMBS, 1961a, 1961b).

L'établissement des couples se fait dans la première année de vie, lorsque les juvéniles devenus indépendants forment des agrégations composées essentiellement de juvéniles, avec seulement quelques adultes (DUNNET et al., 1969; EAST, 1988; HENDERSON et HART, 1991). Les juvéniles forment principalement des relations affiliatives avec leurs frères et sœurs avant la maturité sexuelle (CLAYTON et EMERY, 2007; SCHEID et al., 2008), basées sur le partage de nourriture, le toilettage et la proximité physique (EMERY et al., 2007; SCHEID et al., 2008). Au fur et à mesure de la maturation, les affiliations sont dirigées progressivement vers d'autres individus du sexe opposé, jusqu'à la formation du couple (EMERY et al., 2007; RØSKAFT et ESPMARK, 1984). Une fois adulte, les comportements affiliatifs se concentrent sur la relation de couple (EMERY et al., 2007; MOHAMMED BENKADA et al., 2020), quoique des relations peuvent subsister avec d'autres individus (BOUCHERIE

et al., 2016 ; SEED et al., 2007). Les nouveaux couples semblent rester fréquemment dans la même colonie que leurs parents au moins au début de leur vie (GRIFFIN, 1999). Plus généralement, les couples semblent rester dans le même groupe social et la même colonie d'une saison de reproduction à l'autre et pourraient même conserver le même emplacement de nid dans la colonie (GOODWIN, 1955), permettant l'établissement de liens de familiarité entre voisins de nid. Le site de nidification d'un couple devient un mini-territoire au sein de la colonie, qu'ils défendent contre les autres individus (COOMBS, 1960). Notamment, des couples voisins peuvent ainsi avoir des mini-territoires qui se recouvrent, et peuvent donc être défendus mutuellement contre les intrus (COOMBS, 1960).

L'environnement social du corbeau freux est rendu particulièrement complexe par la conjonction des relations sociales entre couple, groupe social, et colonie, mais aussi par une exposition régulière à de nouveaux individus par les processus de migration. L'espèce est en effet semi-migrante : certaines populations restent sédentaires, tandis que d'autres forment des flux migratoires pouvant traverser plusieurs centaines de kilomètres autour de la période de reproduction (BUSSE, 1965 ; BUSSE, 1969 ; MCKILLIGAN, 1980 ; PATTERSON et al., 1971). Ces populations migrantes rejoignent les populations sédentaires locales à leur colonie de nidification (MCKILLIGAN, 1980 ; PATTERSON et al., 1971). Si les nids sont effectivement réutilisés d'une année à l'autre, un corbeau freux peut donc être ainsi exposé aux mêmes individus d'une année à l'autre (sans tenir compte des variations dues à la natalité et à la mortalité), mais uniquement à certaines périodes. Cela n'empêche pas nécessairement la formation de relations sociales (certains corvidés peuvent en effet se souvenir d'individus particuliers, y compris hétérospécifiques, sur plusieurs années (BOECKLE et BUGNYAR, 2012 ; TAYLOR, 2014), mais cela complexifie fortement l'environnement social.

L'hypothèse de l'intelligence sociale prédit que les espèces socialement complexes devraient aussi être cognitivement douées. Les corvidés n'y font pas exception, possédant de très nombreuses aptitudes cognitives, qu'ils utilisent tant pour appréhender le monde que leurs relations sociales.

### 1.3.2 Compétences cognitives des corvidés

Les nombreuses et impressionnantes compétences cognitives des corvidés ont été mises au jour depuis des décennies (EMERY, 2004 ; EMERY et CLAYTON, 2004 ; EMERY et al., 2007). Ces compétences cognitives sont soutenues par un cerveau de grande taille relativement à leur tronc cérébral, une mesure du potentiel cognitif (EMERY, 2004 ; LEFEBVRE et al., 2002 ; PORTMANN, 1947). En outre, le cerveau des oiseaux compte environ deux fois plus de neurones que le cerveau des grands singes, permettant un nombre de connexions synaptiques comparable malgré la taille physique réduite du cerveau (OLKOWICZ et al., 2016). Les compétences cognitives permises par cette architecture cérébrale permettent à ces oiseaux d'appréhender le monde avec une grande finesse et de négocier leur environnement social complexe (voir e.g. BACIADONNA et al., 2021 ; LAMBERT et al., 2019 ; TAYLOR, 2014). Dans ce texte, je me concentrerai uniquement sur les compétences cognitives en lien avec la complexité sociale et vocale : compétences en cognition sociale, flexibilité comportementale, capacités d'innovation et de contrôle volontaire du comportement.

La complexité sociale des corvidés est sous-tendue par des compétences cognitives élaborées (LAMBERT et al., 2019). Les relations inter-individuelles sont permises par la capacité à reconnaître

divers individus, tant grâce à leurs vocalisations (KONDO et al., 2010 ; KONDO et al., 2012 ; RØSKAFT et ESPMARK, 1984 ; SZIPL et al., 2015 ; WASCHER et al., 2012b) que par leur apparence physique (BRECHT et al., 2017 ; KONDO et al., 2012), y compris pour des individus d'une autre espèce comme les humains (BRECHT et al., 2017). Leurs compétences sociales s'étendent également à la transmission de connaissances entre congénères. L'apprentissage vocal a été démontré chez certaines espèces (corbeau à gros bec *Corvus macrorhynchos osai* KONDO, 2021, geai à gorge blanche *Aphelocoma coerulescens* WEBBER et STEFANI, 2010), et l'apprentissage social est omniprésent (BUGNYAR et KOTRSCHAL, 2002 ; FEDERSPIEL, 2010). Par exemple, des humains associés à une expérience négative par certains individus sont reconnus et harcelés par des congénères qui n'ont pas eu cette expérience négative, un flux d'information qui dépasse même les limites de la colonie (LEE et al., 2011 ; MARZLUFF et al., 2010 ; SWIFT et MARZLUFF, 2015).

Une des caractéristiques comportementales des corvidés est leur grande flexibilité. Cette flexibilité est visible dans le jeu, déjà évoqué plus haut, mais également dans d'autres domaines, comme dans des tâches de généralisation d'apprentissage, et notamment l'inversion d'une tâche apprise (e.g. associer un stimulus A à une récompense et un autre stimulus B à une absence de récompense, puis inverser l'association et mesurer le temps que met un individu à résoudre la nouvelle tâche). De nombreuses expériences chez les corvidés ont montré cette capacité d'inversion chez toutes les espèces de cette famille (BEBUS et al., 2016 ; BOND et al., 2007 ; GOSSETTE et al., 1966 ; van HORIK et EMERY, 2018), contrairement à des espèces d'oiseaux d'autres familles (GOSSETTE et al., 1966). Cette flexibilité comportementale se manifeste également par une forte propension à l'innovation, similaire à celle observée chez les grands singes, en particulier dans l'utilisation d'outils (LAMBERT et al., 2017 ; LEFEBVRE, 2013 ; LEFEBVRE et al., 2004 ; SOL et al., 2002), ainsi que dans leur manufacture (BIRD et EMERY, 2009 ; JELBERT et al., 2018 ; LAMBERT et al., 2017), même chez des espèces qui n'ont jamais été observées en utiliser en milieu naturel (TEBBICH et al., 2007).

Enfin, les corvidés ont démontré des capacités de contrôle de leur propre comportement, par exemple dans des tâches de détour et de délai de gratification (DUFOUR et al., 2012 ; R. MILLER et al., 2019), et directement de leur production vocale (BRECHT et al., 2019). Les tâches de détour nécessitent une capacité à s'éloigner temporairement d'une récompense pour pouvoir ensuite l'atteindre ; par exemple, le test du cylindre, où un individu apprend à récupérer une récompense dans un cylindre opaque ouvert sur les côtés. Ensuite, ce cylindre est remplacé par un cylindre transparent ; l'individu doit alors comprendre qu'il doit quand même passer par les côtés ouverts, plutôt que d'essayer d'atteindre directement la récompense à travers le cylindre (MACLEAN et al., 2014). Remarquablement, seuls les corvidés du genre *Corvus* réussissent cette tâche (KABADAYI et OSVATH, 2017), pas les autres espèces (KABADAYI et OSVATH, 2017 ; MACLEAN et al., 2014 ; STOW et al., 2018 ; VERNOUILLET et al., 2016). Les espèces du genre *Corvus* acquièrent ce contrôle de soi très tôt dans leur développement (10 semaines après éclosion chez le grand corbeau KABADAYI et OSVATH, 2017). Ce contrôle de soi est prononcé, puisque ces espèces peuvent attendre plusieurs minutes dans une tâche de délai de gratification (DUFOUR et al., 2012 ; HILLEMANN et al., 2014), en cachant la récompense et en la récupérant au moment de l'échange (DUFOUR et al., 2012 ; HILLEMANN et al., 2014). De plus, les cas d'échec de la tâche ont lieu tôt dans les expériences, suggérant moins un

manque de contrôle qu'une décision véritablement volontaire, que le délai ne vaut pas la gratification supplémentaire (DUFOUR et al., 2012; HILLEMANN et al., 2014; WASCHER et al., 2012a)). Les corvidés, notamment les espèces du genre *Corvus*, possèdent donc non seulement une forte capacité d'innovation comportementale, mais aussi un contrôle de soi, qui pourraient leur permettre une certaine créativité dans leur comportement, notamment vocal.

### 1.3.3 Le potentiel pour une complexité vocale chez les corvidés, et en particulier le corbeau freux

Malgré les décennies d'études sur leurs compétences sociales et cognitives, les capacités vocales des corvidés restent largement inconnues, tout particulièrement en comparaison des capacités vocales des autres oiseaux et des primates.

Les corvidés sont des animaux particulièrement vocaux, utilisant des répertoires de nombreuses vocalisations différentes dans leurs interactions sociales (BROWN et FARABAUGH, 1997; CHAMBERLAIN et CORNWELL, 1971; ELLIS, 2008; MATES et al., 2015; ROSKAFT et ESPMARK, 1982; RØSKAFT et ESPMARK, 1984). Les vocalisations sont caractérisées par la production fréquente de sons particulièrement inharmoniques par rapport aux autres oiseaux (AMADON, 1944; FLETCHER, 2000; STOWELL et al., 2016a). La majorité des études sur les vocalisations ont mis au jour des variations inter-individuelles fortes, permettant l'établissement de signatures vocales individuelles (e.g. BENTI et al., 2019; BERG et al., 2012; BOECKLE et al., 2012, 2018; MATES et al., 2015; STOWELL et al., 2016a; YORZINSKI et al., 2006), que les individus utilisent pour se reconnaître entre eux (KONDO et al., 2010; RØSKAFT et ESPMARK, 1984).

La variabilité vocale de ces espèces se manifeste également dans leurs capacités d'imitation et d'apprentissage vocal complexe. Les corvidés sont capables d'incorporer les vocalisations d'autres espèces (E. D. BROWN, 1985; BROWN et FARABAUGH, 1997; KAPLAN, 1999), ainsi que celles de leurs conspécifiques (E. D. BROWN, 1985; KONDO, 2021; WEBBER et STEFANI, 2010), et notamment leurs partenaires sociaux (comme le grand corbeau, où les membres d'un couple partagent leurs cris de contact à longue distance, LUEF et al., 2017). Cette transmission des vocalisations au sein de l'espèce permet l'établissement d'une culture vocale (ENGGIST-DUEBLIN et PFISTER, 2002) par laquelle des partenaires sociaux partagent leur répertoire vocal. Cette flexibilité vocale pourrait représenter un handicap dans la reconnaissance sociale, puisqu'elle nécessite un apprentissage constant des signatures vocales des partenaires sociaux (THOMSEN et al., 2019). En contrepartie, elle permet un contrôle de la production vocale, comme l'utilisation choisie de certaines vocalisations en fonction du contexte, comme l'accès à la nourriture (BUGNYAR et al., 2001), la présence de prédateurs particuliers (YORZINSKI et VEHCAMP, 2009), ou même de façon arbitraire dans des expériences de conditionnement opérant (BRECHT et al., 2019).

La flexibilité vocale des corvidés se manifeste également dans une autre production vocale, similaire au chant des autres oiseaux. Ce chant des corvidés est cependant encore moins bien connu que le reste de leurs vocalisations. Jusqu'à très récemment, les corvidés étaient considérés comme des "oiseaux chanteurs sans chant" (BUGNYAR et al., 2016). Quelques mentions semi-anecdotiques dans la littérature ont cependant noté une production vocale longue, semblable au chant des autres oiseaux,

chez certains corvidés sociaux (E. D. BROWN, 1985; BROWN et FARABAUGH, 1997; COOMBS, 1960; WITHERBY, 1943). La description la plus détaillée nous vient de COOMBS (1960) sur le corbeau freux. Lors du chant, un corbeau freux se perche à l'écart de la colonie et produit une série de vocalisations, entièrement distinctes du répertoire vocal usuel. COOMBS (1960) note en particulier que ce chant n'est en aucun cas utilisé dans des contextes territoriaux ou reproducteurs. E. D. BROWN (1985) a fait des observations similaires chez la corneille d'Amérique. Ces chants pourraient donc correspondre aux chants non orientés produits par certaines espèces (DUNN et ZANN, 1996; KNEUTGEN, 1969; RITERS et al., 2019; SOSSINKA et BÖHNER, 1980). Ceux-ci sont produits plus lentement et plus variablement que les chants classiques dans les espèces qui les produisent, et pourraient donc être une piste prometteuse pour étudier les innovations vocales chez les corvidés.

## 1.4 Objectifs de la thèse

L'objectif principal de ma thèse était de tester l'hypothèse de la complexité socio-communicative chez les oiseaux, en prenant l'exemple d'une espèce extrêmement complexe socialement, le corbeau freux. Comme de nombreux corvidés, le corbeau freux possède des capacités cognitives fortes en sus de son organisation sociale complexe, mais ses capacités vocales sont relativement mal connues. Seules quelques études existent à ce sujet, mais elles suggèrent le potentiel pour une forte complexité vocale : le répertoire vocal est de grande taille avec une forte variabilité acoustique tant inter- qu'intra-individuelle (ROSKAFT et ESPMARK, 1982), et les individus peuvent au minimum discriminer des individus familiers et non-familiers (RØSKAFT et ESPMARK, 1984) grâce à des cris individualisés (BENTI et al., 2019; MUNTEANU et al., 2009). En outre, le corbeau freux (ainsi que certains autres corvidés sociaux) semble produire des vocalisations similaires aux chants des autres passeriformes, décrits décrits de façon anecdotique (BROWN et FARABAUGH, 1997; COOMBS, 1960; WITHERBY, 1943) et utilisés hors des contextes typiques des chants d'oiseaux (CATCHPOLE et SLATER, 2008). Cependant, la difficulté d'analyser les vocalisations des corvidés, en raison de leur structure chaotique (FLETCHER, 2000; STOWELL et al., 2016a), a limité ces études.

Le développement récent d'outils statistiques, analytiques et informatiques beaucoup plus puissants pourrait résoudre ces problèmes. En particulier, l'étude des répertoires vocaux en général est limitée avec les méthodes classiques à des vocalisations suffisamment similaires pour mesurer un ensemble de paramètres acoustiques communs. Ces paramètres acoustiques doivent en outre être choisis soigneusement pour ne pas donner des résultats biaisés ou incomplets, ce qui est difficile sans une connaissance experte du répertoire vocal d'une espèce. Cette connaissance experte est rarement disponible, même chez les espèces modèles classiques. Des techniques développées dans les dernières années permettent cependant d'utiliser le spectrogramme entier, sans chercher à singulariser les paramètres acoustiques. Cela peut permettre d'analyser les vocalisations même d'espèces aussi variables vocalement que les corvidés, mais aussi la totalité des espèces où les vocalisations sont, relativement parlant, plus simples. L'explosion de l'usage des réseaux de neurones profonds (*deep learning*) permet de traiter les données de façon extrêmement flexible et robuste. Une des applications possibles de ces réseaux de neurones à la recherche scientifique est de s'affranchir de la majeure partie des coûts

d'acquisition des données, comme je le développerai ensuite. Enfin, de nouveaux outils informatiques permettent d'accélérer immensément le traitement des données, de façon à traiter les jeux de données massifs modernes dans des temps raisonnables.

1. Pour étudier les capacités vocales du corbeau freux, il nous fallait d'abord collecter une quantité de données suffisante pour examiner leur potentiel en termes de complexité vocale. Pour ce faire, j'ai tout d'abord développé un système basé sur un réseau de neurones profond pour extraire les vocalisations de corbeaux freux à partir d'enregistrements longs en volière extérieure. Ce réseau de neurones a été entraîné sur des enregistrements de vocalisations de corbeaux freux, annotées pour identifier les corbeaux freux individuellement. Le système, une fois entraîné, devait permettre d'extraire très rapidement des vocalisations des enregistrements hors présence humaine, tout en permettant un suivi individuel des productions vocales. Ce système a été développé et testé sur un premier groupe à Strasbourg, ce qui a donné lieu à une publication dans *Ecological Informatics*. Le système a ensuite été étendu à un deuxième groupe captif au sous-département d'étude du comportement animal de l'Université de Cambridge, Royaume-Uni.
2. En utilisant les données collectées dans la première partie, y compris des données collectées hors présence humaine et extraites avec le réseau de neurones, j'ai étudié la complexité vocale dans les répertoires individuels de 22 corbeaux freux, regroupant deux groupes captifs (à Strasbourg et à Madingley, Royaume-Uni). La complexité a été mesurée formellement par la diversité, la gradation, et la similarité des répertoires vocaux.
3. J'examine la présence d'une signature vocale dans le cri au nid des femelles, une des seules vocalisations des corbeaux freux où un individu peut être identifié sans nécessiter un marquage invasif, car la femelle en reproduction reste au nid pendant plusieurs semaines. Il s'agit de vérifier l'existence d'une signature vocale individuelle dans cet autre type de vocalisations, mais aussi et surtout de tester si les femelles d'une colonie se ressemblent entre elles ou non. Les résultats de cette étude pourraient conditionner la mise au point d'un système de suivi des populations sauvages de corbeaux freux sans avoir besoin de les capturer.
4. Enfin, j'étudie le potentiel de contrôle vocal des corbeaux freux en étudiant leur capacité d'entraînement à un stimulus rythmique externe. Le protocole est basé sur une exposition non récompensée et non conditionnée pour s'assurer que les individus, s'ils s'ajustent, le font librement. Si le stimulus entraîne la production de chant, et que des stimuli à propriétés rythmiques différentes entraînent une modification dans le rythme des chants, cela permettrait de penser que les corbeaux freux contrôlent volontairement leur chant.

# Chapitre 2

## Méthodes générales

Au cours de cette thèse, j'ai développé plusieurs méthodes pour analyser la complexité vocale des corbeaux freux. Cette section décrit les sujets d'expérience, les conditions d'acquisition des données, et les méthodes utilisées dans les chapitres suivants pour les analyses (autres que les tests statistiques, qui sont détaillés dans chaque chapitre).

### 2.1 Sujets d'études

#### 2.1.1 Colonies captives

##### 2.1.1.1 Colonie captive de Strasbourg (France)

Cette colonie est composée de 15 individus au total au cours des trois années d'étude, mais la composition du groupe a été modifiée plusieurs fois (Table 2.1). Les premiers individus ont été capturés en 2006 en tant que juvéniles, puis d'autres individus ont été capturés et ajoutés au groupe en 2013, 2016 et 2019, en tant que juvéniles ou déjà adultes. Pour 2013 et 2016, les nouveaux individus correspondent à des captures effectuées dans des colonies sauvages autour de Strasbourg. Un individu (Balbo) a été introduit en 2019, reçu d'une association de réhabilitation de corvidés. La parenté entre les différents individus n'est pas certaine, hormis deux individus (Jonas et Elie) prélevés dans le même nid. Au moment de l'étude, les individus sont majoritairement en couple ou en triade, mais dans certains cas perturbés par les modifications de composition du groupe. Chaque individu est identifié par des bagues de couleur attachées aux pattes.

La colonie a été hébergée en volière extérieure depuis son établissement, et au début de l'étude a été transférée dans une volière au parc Albert-Schweitzer à Strasbourg. Cette volière, de 26 m de long, 8 m de large, et 2.50 à 3.70 m de haut, est constituée d'un compartiment unique ouvert (à l'exception d'un sas d'entrée), recouvert d'un filet souple. Des perchoirs ont été mis à disposition des corbeaux à plusieurs emplacements, ainsi que divers enrichissements (jouets, pelouse artificielle). Une clôture électrique a été installée pour protéger la colonie et la volière des animaux sauvages. La nourriture de la colonie est composée d'une pâtée de céréales et de boulettes de fruits, supplémentées de granulés d'insectes avec, tous les mois, un traitement complémentaire de vitamines et de minéraux pour éviter les carences. Nourriture et eau sont distribuées quotidiennement et disponibles *ad libitum* à un emplacement fixe. Pendant la reproduction, des œufs et de la viande rouge hachée complètent la nourriture, et des branches sont distribuées pour permettre la construction des nids. Des friandises (croquettes pour chaton, friandises pour chien à la viande de poulet, emmental) étaient distribuées ponctuellement, indépendamment de la réalisation d'une expérience.



TABLE 2.1 – Informations générales sur les corbeaux freux des deux groupes captifs étudiés : nom, code (uniquement pour le groupe de Strasbourg, et uniquement pour Rookognise (Chapitre 3), sexe (M : mâle, F : femelle), origine (chaque lettre correspond à une colonie d'origine différente), couple ou triade, année d'introduction dans le groupe respectif, âge à l'arrivée dans le groupe et parenté avec les autres individus quand cette information est disponible. Ces parentés n'ont cependant jamais été vérifiés génétiquement.

Nom	Code	Sexe	Origine	Couple	Entrée	Âge à l'entrée	Notes
Balbo	H	M			2019	3 ans	Reçu d'une association de réhabilitation
Bashir	A	M	A	Gigi	2013	Inconnu (juvénile)	
Braad	R	F	B		2017	2 mois	
Brain	B	M	B	Kafka, Pomme, Siobhan*	2006	2 mois	
Elie	E	M	B	Feisty, Tom	2006	4 mois	Prélevé dans le même nid que Jonas
Feisty	F	F	C	Elie, Tom	2016	4 mois	
Gigi	G	F	A	Bashir	2013	Inconnu (adulte)	
Jolene	L	F	C	Osiris	2016	2 mois	
Jonas	J	F	B	Merlin	2006	4 mois	Prélevé dans le même nid que Elie
Kafka	K	M	B	Brain, Pomme	2006	2 mois	
Merlin	M	M	B	Jonas	2006	4 mois	
Osiris	O	M	B	Jolene	2006	4 mois	
Pomme	P	F	A	Brain, Kafka	2013	Inconnu (adulte)	
Siobhan	S	F	A	Noah, Brain*	2013	Inconnu (adulte)	
Tom	T	M	B	Elie, Feisty	2006	2 mois	
Aristotle		M	E		2003	< 1 mois	
Bussell		F	F	Fry	2004	< 1 mois	
Cassandra		F	G	Leo	2007	< 1 mois	Prélevée dans le même nid que Leo
Connelly		M	F		2004	< 1 mois	
Fry		F	F	Bussell	2004	< 1 mois	
Huxley		F	E		2003	< 1 mois	
Leo		M	G	Leo	2007	< 1 mois	Prélevé dans le même nid que Cassandra

### 2.1.1.2 Colonie captive de Madingley (Royaume-Uni)

Cette colonie est composée de 7 individus au total (Table 2.1), capturés en tant que juvéniles dans deux colonies sauvages dans la région de Cambridge, entre 2002 et 2006. Au moment de l'étude, deux couples sont présents dans le groupe, avec 3 oiseaux célibataires. Chaque individu est identifié par des bagues de couleur attachées aux pattes.

Comme la colonie de Strasbourg, cette colonie a été hébergée en volière extérieure depuis sa capture, au sous-département d'étude du comportement animal de l'Université de Cambridge à Madingley, Royaume-Uni. Cette volière, d'approximativement 10 m de long, 4 m de large, et 4 m de haut, est divisée en un compartiment ouvert principal, et cinq sous-compartiments utilisés pour les expériences, séparées par des trappes ouvertes en général mais pouvant être fermées pour isoler certains individus (ce qui n'a pas été fait pendant cette étude). Des perchoirs ont été mis à disposition des corbeaux ainsi que divers enrichissements (jouets, cordes, galets au sol). L'eau et une nourriture composée d'un mélange varié (fruit, viande, graines pour oiseaux, pain, pâtes, et biscuits pour chien) sont distribuées quotidiennement et disponibles *ad libitum*. Pour la reproduction, des branches sont mises à disposition pour permettre la construction des nids. Contrairement à la colonie de Strasbourg, cette volière est également désinfectée quotidiennement par les soigneurs du département.

### 2.1.2 Colonies sauvages

Outre les deux colonies captives, trois colonies de corbeaux freux sauvages ont été enregistrées au cours de l'année 2022, deux autour de Strasbourg (une sur le campus CNRS de Cronenbourg,



FIGURE 2.1 – Photographies des deux volières. Rangée du haut, volière du groupe de Strasbourg : gauche, vue extérieure ; droite : vue intérieure partielle, montrant l'emplacement de l'observateur et du caméscope utilisé pour filmer le groupe. Rangée du bas, volière du groupe de Madingley, vues extérieures montrant toutes deux l'emplacement de l'observateur pendant les enregistrements.

une autre au parc public des Contades) et la troisième près de Madingley au Royaume-Uni. Les trois colonies comprennent plusieurs dizaines de nids et correspondent à des sites de reproduction utilisés chaque année. Elles ont été enregistrées spécifiquement à la saison de reproduction, la seule période où il a été possible d'identifier certains individus (spécifiquement, les femelles au nid) sans nécessiter de marquage. Cependant, pour la colonie sauvage de Madingley, les identités précises des femelles n'ont pas pu être déterminées sans ambiguïtés, en raison d'une coïncidence entre le début des observations et la pousse des feuilles.

## 2.2 Enregistrements des données

### 2.2.1 Matériel d'enregistrement

Les colonies captives et sauvages ont été enregistrées passivement par des enregistreurs autonomes (Song Meter 4, Wildlife Acoustics, USA), connectés à des microphones omnidirectionnels (SMM-A2,



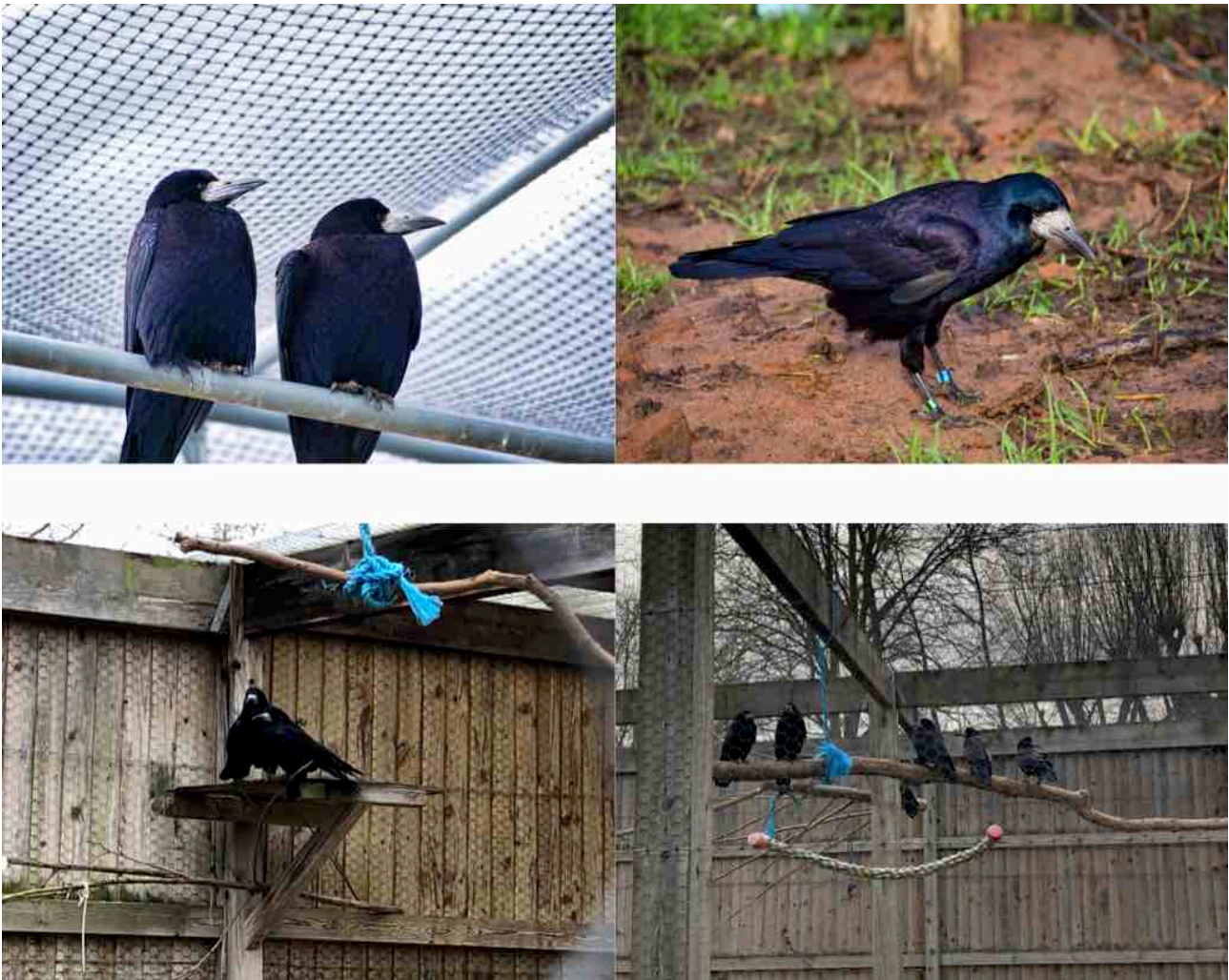


FIGURE 2.2 – Photographies des corbeaux des deux groupes captifs. Rangée du haut, à Strasbourg : à gauche, Elie et Tom au repos en contact, un comportement considéré comme affiliatif ; à droite, Merlin au sol, avec les bagues de couleur bien visibles (crédit : E. Braccini). Rangée du bas, à Madingley : à gauche, Leo et Cassandra au début de la construction des nids ; à droite, le groupe perché à proximité les uns des autres.

Wildlife Acoustic, USA). Des câbles de longueurs variables ont permis de placer les microphones à proximité des colonies pour les enregistrements.

Ces enregistreurs fonctionnent sur batteries et peuvent être programmés pour enregistrer suivant un programme fixe, adapté en fonction des études. La colonie captive de Strasbourg a été enregistrée quotidiennement en cours de matinée et en début d’après-midi, de 8h à 10h de janvier à mars 2020 puis de 8h à 14h à partir de mai 2020. Nous avons enregistré cette colonie jusqu’en août 2022. La colonie captive de Madingley a été enregistrée entre 8h et 17h quotidiennement entre février et mai 2022. Enfin, les colonies sauvages ont été enregistrées de façon plus opportuniste, ajustée tant à la présence des corbeaux freux qu’à celle des observateurs, qui ont parfois observé plusieurs colonies à des lieux différents dans la même journée.

Tous les enregistrements ont été effectués à une fréquence d’échantillonnage de 48 kHz avec un débit de 16 bit, et stockés en tant que fichiers non compressés au format WAV. Le gain de chaque microphone a été ajusté empiriquement en fonction des conditions saisonnières et des distances à la

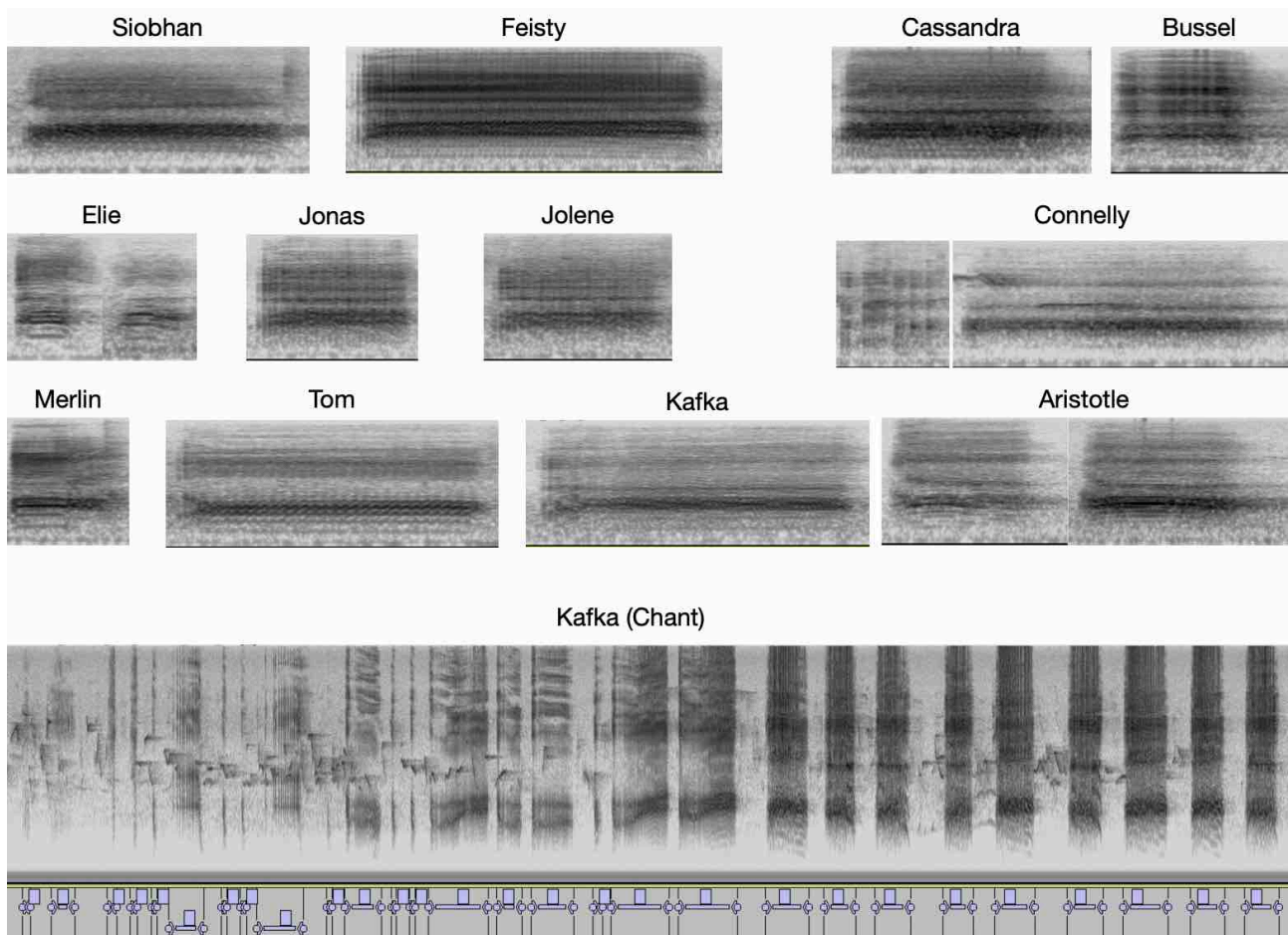


FIGURE 2.3 – Quelques exemples de vocalisations représentatives de corbeaux freux, sous forme de Mel-spectrogrammes, obtenus avec Audacity. L'individu émetteur est indiqué au-dessus. Toutes les vocalisations autres que le chant de Kafka en bas de la figure sont les cris les plus fréquents, déterminés par les approches HDBSCAN Chapitre 4, produits dans des contextes rarement discernables. La même échelle de temps et de fréquence a été utilisée pour prendre ces images. Pour le chant de Kafka, la piste en bas de l'image correspond aux délimitations des vocalisations

colonie enregistrée. Pour les colonies captives, chaque enregistreur portant deux microphones a pu être programmé avec deux gains : un gain faible pour enregistrer les vocalisations puissantes sans saturer, et un gain élevé pour capter les vocalisations plus faibles.

## 2.2.2 Protocole d'enregistrement sur les colonies captives

Les observations sur les colonies captives ont employé un protocole commun. Pendant une session d'enregistrement, un observateur se poste à proximité des corbeaux freux, et note toutes les vocalisations du groupe à l'aide d'un script Python. Ces notes incluent le temps où la vocalisation est émise, l'identité de l'oiseau, et le contexte de la vocalisation quand celui-ci est identifiable. La colonie de Strasbourg a également été filmée avec un caméscope pour couvrir la majorité du groupe. Ces vidéos ont servi à la vérification des notes pendant l'annotation des données vocales.

Une partie des enregistrements ainsi annotés ont permis la construction de la première base de données, composée d'environ 15h d'enregistrements sur la colonie de Strasbourg. Le jeu de données

est composé des fichiers sons enregistrés et synchronisés (quand plusieurs enregistreurs ont été utilisés), et des annotations réalisées principalement dans les logiciels AudioSculpt (BOGAARDS et al., 2004) puis Audacity (v.3.1.3). Les annotations correspondent aux vocalisations, délimitées individuellement et attribuées à l'individu émetteur. Les vocalisations impossibles à attribuer à un individu ont été attribuées une classe "Inconnu", et les vocalisations de plusieurs individus qu'il a été impossible de séparer ont été attribuées à une classe "Multiple"..

Environ 35h d'enregistrement supplémentaires ont été annotées à l'aide d'un réseau de neurones convolutionnel entraîné sur ce jeu de données (voir Chapitre 3). Sur ces 35h, 15 correspondaient aux enregistrements de la colonie captive de Madingley, et 20 correspondaient à des enregistrements supplémentaires de la colonie captive de Strasbourg. Environ 3h sur chaque colonie ont été prises sur les sessions enregistrées en l'absence de présence humaine pour tester la qualité du réseau.

### 2.2.3 Protocole d'enregistrement sur les colonies sauvages

Un protocole sensiblement similaire a été employé pour les colonies sauvages. Pour les colonies sauvages autour de Strasbourg, un observateur se poste à hauteur des nids de chaque colonie sur un bâtiment proche, et observe les femelles au nid. Pour la colonie sauvage de Madingley, en l'absence de bâtiment de hauteur suffisante, les observations ont été conduites depuis le sol. Pour chacune de ces trois colonies, un protocole d'échantillonnage focal a été utilisé autant que possible pour s'assurer de l'identité des individus : quand une femelle commence à vocaliser, chaque vocalisation est notée à l'aide d'un script Python (le même que pour les colonies captives), et la même femelle est ensuite suivie tant qu'elle vocalise ou jusqu'à ce que 15 vocalisations environ aient été notées. L'identité de cette femelle est alors notée par la position de son nid dans la colonie, afin de limiter les prises de données multiples sur les mêmes individus. Les vocalisations sont ensuite extraites dans Audacity en synchronisant les notes, vidéos et enregistrements sonores pour s'assurer des identités. Seules les vocalisations de bonne qualité (i.e. suffisamment fortes et non couvertes par d'autres sons) ont été utilisées. Cependant, les individus de la colonie sauvage de Madingley n'ont pas pu être identifiés précisément à cause de la couverture foliaire au moment des observations. D'autres vocalisations ont été extraites sans chercher à identifier l'identité de la femelle émettrice pour compléter les données.

## 2.3 Analyses acoustiques

### 2.3.1 Représentations du signal : de l'audio au Mel-spectrogramme

Les fichiers WAV enregistrent la représentation brute, en domaine temporel, du son. À ce stade, je considère les fichiers sons des différents microphones concaténés en parallèle et synchronisés entre eux, il s'agit donc de fichiers multi-canaux (un canal étant égal à un microphone). Cette représentation en domaine temporel est cependant peu pratique pour comparer les vocalisations. Elle est donc convertie en spectrogramme, représentation intermédiaire entre les domaines temporel et fréquentiel qui permet d'observer la trajectoire du son au cours du temps.



Le spectrogramme est obtenu en appliquant la transformée de Fourier à court terme (*Short-Time Fourier Transform*, STFT) aux fichiers sons. La STFT est conditionnée par plusieurs paramètres dont le choix précis peut être considéré dans l'ensemble comme très arbitraire. Dans tout ce manuscrit, j'ai utilisé les mêmes paramètres, choisis en fonction des valeurs qui donnaient des spectrogrammes avec la meilleure résolution temporelle et fréquentielle possible (mais il y a nécessairement un compromis entre les deux : ces résolutions sont reliées par une relation inverse, il n'est donc pas possible de les maximiser toutes les deux). Pour toutes les analyses sauf Rookognise (Chapitre 3), le spectrogramme a été défini avec une fenêtre de Hamming de 10 ms, avec 80% de recouvrement entre fenêtres successives. Les fenêtres correspondent à des fragments du son d'origine, auxquels sont appliqués la transformée de Fourier pour le spectre en fréquence du fragment sonore. Ces spectres en fréquences sont ensuite concaténés les uns à la suite des autres dans l'ordre chronologique pour former le spectrogramme. Pour Rookognise uniquement, les spectrogrammes ont été définis avec une fenêtre de Hamming de 50 ms avec 75% de recouvrement, pour des raisons pratiques liées au matériel informatique utilisé (voir Chapitre 3).

Enfin, le spectrogramme est converti d'une échelle de fréquences linéaire à l'échelle de fréquence Mel, permettant d'obtenir une représentation appelée Mel-spectrogramme dans la suite. L'échelle Mel est une échelle basée historiquement sur la psychoacoustique humaine (**Stevens1937**). Celle-ci a été très utilisée pour les algorithmes basés sur l'analyse de la parole humaine, car elle reproduit en théorie la perception auditive humaine, et a donc largement été utilisée dans la littérature bioacoustique (). Plus précisément, l'échelle Mel est quasi-linéaire dans les basses fréquences et devient logarithmique dans les hautes fréquences, avec le point limite entre les deux domaines généralement fixé à 1000 Hz. Pour les applications algorithmiques, cela permet notamment de réduire la taille de l'axe des fréquences des spectrogrammes, réduisant ainsi les calculs nécessaires dans les algorithmes, tout en préservant plus d'information dans les basses fréquences. Cela permet donc de concentrer les algorithmes que nous utilisons pour les analyses acoustiques sur la plage fréquentielle la plus utilisée par les corbeaux freux, tout en conservant les hautes fréquences, qui peuvent inclure des informations supplémentaires des vocalisations des freux mais aussi d'autres caractéristiques locales des enregistrements utilisables par les algorithmes. Une échelle alternative (notamment qui correspondrait davantage aux perceptions auditives des corbeaux freux) pourrait être plus adaptée en inspirant l'algorithme de traitement des données du traitement fait par les corbeaux freux eux-mêmes, mais peu d'études ont été faites sur ces perceptions auditives, et en particulier une telle échelle n'a pas jusqu'ici été mise au point. À ma connaissance, une seule étude sur la corneille mantelée (*Corvus corone cornix*) a été faite, et a trouvé une sensibilité auditive similaire à l'homme chez cette espèce, avec en particulier une sensibilité plus forte aux basses fréquences par rapport aux autres oiseaux chanteurs (**Jensen2006**). Par conséquent, l'échelle Mel représente sans doute un compromis acceptable en l'absence d'échelle véritablement adaptée.

### 2.3.2 Pré-traitement des vocalisations : débruitage, sélection du meilleur canal

Les vocalisations dans les enregistrements bruts sont typiquement enregistrées avec divers bruits de fond : bruits de fonctionnement des microphones, bruits biotiques (autres oiseaux notamment, y

compris les autres individus dans chaque colonie d'étude), ou bruits abiotiques (e.g. véhicules, vent, pluie, voix humaines). Pour limiter l'impact de ces bruits, nous avons appliqué plusieurs étapes de pré-traitement aux vocalisations, sauf pour Rookognise (Chapitre 3), qui utilise les Mel-spectrogrammes bruts avec tous les canaux. Ce pré-traitement correspondait à une pré-sélection des vocalisations, un débruitage dynamique, et la sélection du meilleur canal. Cette dernière étape a été réalisée manuellement uniquement pour les petits jeux de données au Chapitre 5, et dans les autres cas par les algorithmes décrits ci-dessous.

La pré-sélection des vocalisations consiste uniquement à retirer les vocalisations superposées à d'autres vocalisations de freux ou à des bruits trop forts (par exemple, des bruits de train).

Le débruitage dynamique correspond à un algorithme de seuillage, où un échantillon de bruit de fond est utilisé pour définir le seuil minimal. Cet algorithme était basé sur le code de SAINBURG et al. (2019), lui-même inspiré par le programme Luscinia (LACHLAN, 2007), et suit la procédure suivante :

- Extraire une vocalisation et calculer le spectrogramme multi-canal,  $S$
- Extraire un échantillon de bruit de fond (par exemple, les intervalles sans vocalisation annotée jusqu'à 10 secondes avant le début et jusqu'à 10 secondes après la fin de la vocalisation) et calculer le spectrogramme multi-canal,  $N$
- Calculer le seuil minimal à partir du spectrogramme du bruit de fond (par exemple, le 95ème centile supérieur, calculé séparément pour chaque bande de fréquence et chaque canal, sur l'axe temporel du spectrogramme),  $N'$
- Calculer un masque binaire correspondant aux emplacements où  $S > N'$ ,  $M$
- Lisser ce masque par deux opérations de fermeture et ouverture morphologiques (pour retirer les petits artéfacts) puis un filtre Gaussien
- Appliquer le masque au spectrogramme pour récupérer le spectrogramme débruité.

Après débruitage, les spectrogrammes sont convertis en échelle Mel, puis les silences en début et fin de vocalisation sont découpés. Ce découpage est contrôlé par un seuil (défini manuellement) et l'enveloppe d'amplitude du spectrogramme calculée avec la formule :

$$E_t = \sqrt{\frac{1}{F} \sum_{f=1}^F S_{t,f} \cdot \max_f(S(t, f))},$$

où  $E_t$  est l'enveloppe d'enveloppe au pas de temps  $t$ ,  $S_{t,f}$  est la valeur du spectrogramme au pas de temps  $t$  et au pas de fréquence  $f$ , et  $\max_f$  est la fonction retournant le maximum sur l'axe des fréquences. Les vocalisations où l'enveloppe d'amplitude n'atteignent pas le seuil minimal sont retirées de l'analyse.

Enfin, le canal avec le meilleur rapport signal sur bruit (SNR) est sélectionné. Le SNR est défini par :

$$SNR = \frac{\frac{1}{N} \sum_i x_i^2}{\frac{1}{M} \sum_i n_i^2},$$

où  $x$  est la représentation en domaine temporel de la vocalisation, et  $n$  est la même représentation de l'échantillon de bruit de fond, toutes deux débruitées avec l'algorithme ci-dessus et le même échantillon de bruit de fond. Le canal avec le SNR le plus élevé est sélectionné pour la suite des analyses.

### 2.3.3 Distance entre spectrogrammes : DFTW (Déformation Dynamique Spectro-Temporelle ou *Dynamic Frequency-Time Warping*)

Les signaux acoustiques sont en général des objets à grand nombre de dimensions, que ce soit la représentation temporelle (où chaque échantillon correspond à une dimension) ou le spectrogramme (où chaque valeur du spectrogramme correspond à une dimension différente), même après passage à l'échelle Mel. Cette haute dimensionnalité rend difficile leur comparaison directe, car deux vocalisations ont rarement la même durée par exemple, et donc les spectrogrammes n'ont pas la même longueur.

Les comparaisons des vocalisations ont longtemps reposé sur des examens des spectrogrammes et des sons, mais cette approche est peu adaptée à la taille massive des jeux de données modernes, et repose par ailleurs sur une part importante de subjectivité. Une autre approche classique est basée sur la mesure de paramètres acoustiques (e.g. durée, fréquence fondamentale, bande de fréquence, pente de la fréquence fondamentale). Ces approches peuvent être automatisées dans une certaine mesure, permettant de ramener toutes les vocalisations à un nombre fixe de dimensions (le nombre de paramètres acoustiques choisis), et ont l'avantage d'être facilement interprétables. Cependant, elles supposent un choix préliminaire des paramètres acoustiques qui peut être subjectif, même avec une connaissance experte de l'espèce (SAINBURG et GENTNER, 2021). Cette connaissance n'est pas toujours disponible, même chez les espèces modèles classiques comme le diamant mandarin (ELIE et THEUNISSEN, 2016). Par ailleurs, même dans le meilleur des cas, choisir ces paramètres peut retirer une partie importante de l'information contenue dans la structure du spectrogramme. Enfin, les paramètres acoustiques choisis sont fréquemment hautement colinéaires, ce qui peut fausser les résultats des analyses (GOFFINET et al., 2021).

D'autres possibilités ont émergé récemment, basées sur la comparaison des spectrogrammes entiers. La plus simple est la distance euclidienne entre paires de spectrogrammes. Cependant, cette distance simple nécessite que les spectrogrammes aient les mêmes dimensions, ou que le spectrogramme le plus petit soit complété par des zéros, ce qui peut également introduire un biais. Par ailleurs, si les deux vocalisations sont légèrement décalées, la distance euclidienne entre les deux sera très élevée, même si les deux vocalisations sont par ailleurs identiques (par exemple, si une vocalisation est précédée d'un petit élément). La distance euclidienne est donc très fragile (KEOGH et RATANAMAHATANA, 2005), sans doute trop pour des analyses bioacoustiques. Ce problème peut être résolu par l'algorithme de Déformation Temporelle Dynamique (Dynamic Time Warping, DTW), qui recherche l'alignement optimal entre les spectrogrammes en autorisant les décalages locaux. La distance associée aux spectrogrammes ainsi alignés peut être utilisée comme estimation de la dissimilarité entre les spectrogrammes. La distance de DTW est calculée par l'algorithme suivant :

- À partir des spectrogrammes A et B de formes respectives  $(N_A, F)$  et  $(N_B, F)$ , calculer la matrice de distance  $D$  entre chaque paire de pas de temps de A et B (par exemple avec la distance euclidienne), de forme  $(N_A, N_B)$
- La DTW est calculée avec la matrice  $M$ , initialisée par :

$$M[0, 0] = 0,$$



$$\forall i \in [1 \dots N_A], M[i, 0] = \sum_{k=1}^i D[k, 0],$$

$$\forall j \in [1 \dots N_B], M[0, j] = \sum_{k=1}^j D[0, k].$$

— Puis  $M$  est complétée par l'équation :

$$\forall (i \in [1 \dots N_A], j \in [1 \dots N_B]), M[i, j] = D[i, j] + \min \begin{cases} M[i-1, j] \\ M[i-1, j-1] \\ M[i, j-1] \end{cases} .$$

— Finalement, la distance associée à la DTW entre  $A$  et  $B$  est donnée par  $M[N_A, N_B]$ .

La DTW permet de s'affranchir des problèmes liés à des dimensions temporelles différentes ou à des différences dans la vitesse des vocalisations. Cependant, des vocalisations avec des spectrogrammes identiques, mais décalés en fréquence (par exemple, une vocalisation plus aiguë que l'autre mais autrement identique) pourraient être considérées comme plus similaires que deux vocalisations avec la même distribution en fréquence, mais avec des spectrogrammes autrement différents. La DTW seule ne permet pas de garantir ce résultat (pour une illustration voir Fig. 2.4). Dans ce manuscrit, j'utilise donc une extension basée sur un algorithme proposé récemment (SOMERVUO, 2019), la Déformation Dynamique Spectro-Temporelle (*Dynamic Frequency-Time Warping*, DFTW). Cet algorithme est une généralisation de la DTW qui permet de déformer les spectrogrammes non seulement le long de l'axe temporel, mais aussi le long de l'axe des fréquences, pour prendre en compte ces différences spectrales. L'algorithme précis utilisé dans ce manuscrit est légèrement différent de celui proposé par SOMERVUO (2019) : ce dernier propose, en somme, de remplacer la distance  $d$  dans l'algorithme de la DTW ci-dessus par la DTW elle-même. Cependant, cette procédure revient à calculer la DTW un très grand nombre de fois pour chaque paire de spectrogrammes comparés, ce qui allongeaient le temps de calcul de façon déraisonnable sur le matériel que j'ai utilisé. En effet, la DTW revient à calculer la distance entre toutes les paires d'éléments des séquences comparés (dans ce cas, un élément étant un pas de temps d'un spectrogramme, incluant tout l'axe des fréquences), une opération proportionnelle en temps au produit des longueurs des séquences ( $O(MN)$  avec les notations précédentes). La DFTW proposée par SOMERVUO (2019) revient alors à calculer la DTW, non seulement entre chaque paire d'éléments des spectrogramme, mais aussi entre chaque paire de valeurs de ces éléments (i.e. chaque valeur à une fréquence donnée des pas de temps respectifs), ce qui augmente le temps de calcul de la DFTW d'un facteur élevé ( $O(F^2)$  avec les notations précédentes, comparé au temps de calcul de la DTW normale). Pour réduire le temps de calcul, j'ai ajouté une contrainte à l'algorithme, que toutes les déformations se font le long de l'axe entier, temporel ou fréquentiel. Cela correspond à l'algorithme suivant :

- Calculer la matrice de DTW entre les spectrogrammes  $A$  et  $B$  avec l'algorithme précédent
- Trouver l'alignement optimal entre  $A$  et  $B$ , et les déformer le long de l'axe temporel suivant cet alignement afin d'obtenir  $A'$  et  $B'$
- Transposer  $A'$  et  $B'$  pour obtenir  ${}^T A'$  et  ${}^T B'$
- Calculer la matrice de DTW entre  ${}^T A'$  et  ${}^T B'$

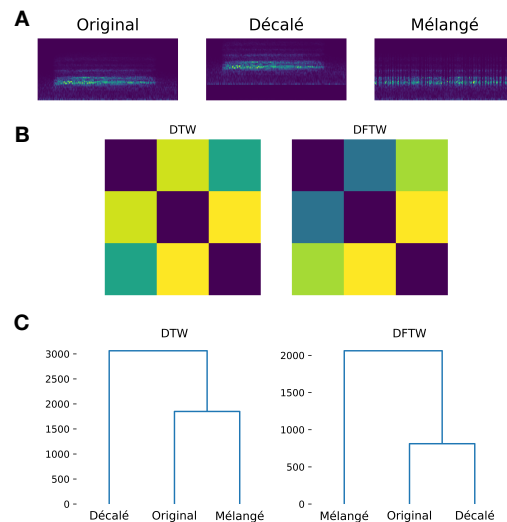


FIGURE 2.4 – Comparaison entre les résultats de la DTW et de la DFTW. A) Un cri d’une des femelles est utilisé comme exemple (gauche, "Original"), soumis à deux transformations : un roulement vers le haut simule un cri plus aigu mais autrement identique (milieu, "Décalé"), et un mélange aléatoire des pas du spectrogramme le long de l’axe temporel (droite, "Mélangé"), simulant une vocalisation de structure très différente mais avec la même distribution en fréquence. B) Matrices de distance entre les trois spectrogrammes pour la DTW (gauche) et la DFTW (droite). Pour chaque matrice, plus la distance est élevée entre deux spectrogrammes, plus la cellule correspondante est claire. C) Dendrogrammes obtenus par clustering sur les matrices de distance. Noter l’inversion de l’association d’Original avec chacun des autres spectrogrammes entre DTW et DFTW : DTW privilégie la ressemblance fréquentielle et place donc Original plus proche de Mélangé. La DFTW, elle, privilégie une forme acoustique commune, même avec des décalages en fréquence.

- Trouver l’alignement optimal entre  ${}^T A'$  et  ${}^T B'$  (ce qui correspond à la déformation le long de l’axe fréquentiel)
- Appliquer cet alignement aux spectrogrammes d’origine  $A$  et  $B$ , pour obtenir  $A''$  et  $B''$
- Calculer la distance de DTW entre  $A''$  et  $B''$ , qui correspond donc à la distance finale de DFTW

Cette procédure revient à imposer la même déformation aux spectrogrammes le long de chaque axe, temporel ou fréquentiel, et à calculer la DTW trois fois par paire de spectrogrammes comparés. Concrètement, la DFTW revient donc à considérer deux vocalisations comme similaires si leurs spectrogrammes peuvent être étirés, soit le long de l’axe temporel, soit le long de l’axe fréquentiel, pour être identiques.

Pour que les distances DFTW obtenues soient comparables entre toutes les paires de spectrogrammes, deux étapes de normalisation supplémentaires sont ajoutées (DANIEL MELIZA et al., 2013 ; R. . F. LACHLAN et al., 2010). Tout d’abord, les spectrogrammes  $A$  et  $B$  sont centrés et réduits conjointement pour éliminer les effets d’échelles potentiels, et ne comparer que les formes relatives des spectrogrammes plutôt que leurs valeurs absolues. Cette normalisation est appliquée par les équations :

$$\mu_{AB} = \frac{\sum_{i=1}^{N_A} \sum_{f=1}^F A_{i,f} + \sum_{j=1}^{N_B} \sum_{f=1}^F B_{j,f}}{(N_A + N_B) \cdot F},$$

$$\sigma_{AB} = \sqrt{\frac{\sum_{i=1}^{N_A} \sum_{f=1}^F (A_{i,f} - \mu_{AB})^2 + \sum_{j=1}^{N_B} \sum_{f=1}^F (B_{j,f} - \mu_{AB})^2}{(N_A + N_B) \cdot F - 1}},$$

$$A_{norm} = \frac{A - \mu_{AB}}{\sigma_{AB}},$$

$$B_{norm} = \frac{B - \mu_{AB}}{\sigma_{AB}}.$$

La deuxième normalisation est appliquée à la distance finale obtenue, qui est divisée par la somme des longueurs des deux spectrogrammes, afin que les vocalisations longues ne soient pas automatiquement plus distantes que les vocalisations courtes (car les algorithmes de DTW et DFTW n'incluent que des sommes de nombres positifs) :

$$dist_{DFTW}^{finale} = \frac{dist_{DFTW}}{N_A + N_B}.$$

### 2.3.4 Projection en espace à faible dimensionnalité : UMAP (*Uniform Manifold Approximation and Projection*)

La matrice de distance obtenue à l'étape précédente est encore à très haute dimension ; cela nuit à la performance des algorithmes de clustering comme l'HDBSCAN à cause de la "malédiction de la dimensionnalité" (qui correspond au fait que plus le nombre de dimensions est élevé, plus les données sont éparées dans le volume possible défini par ces dimensions, et donc plus il est difficile d'y trouver des clusters, BELLMAN, 1957). J'ai donc appliqué un algorithme de réduction de dimensions à la matrice de distance, ce qui permet aussi de la visualiser en deux dimensions.

J'ai choisi l'algorithme UMAP (MCINNES et al., 2018), un algorithme qui est devenu le standard dans la littérature récente (GOFFINET et al., 2021 ; PAGLIARINI et al., 2021 ; SAINBURG et GENTNER, 2021 ; SAINBURG et al., 2019 ; SAINBURG et al., 2020), servant d'alternative aux algorithmes précédents comme l'Analyse en Composantes Principales (ACP) ou le t-SNE (*t-distributed Stochastic Neighbour Embedding*). L'UMAP est un algorithme de réduction de dimensions non linéaire basé sur la construction d'un graphe reliant les points dans l'espace d'origine à leurs plus proches voisins (Fig. 2.5). Les arêtes de ce graphe relient les points voisins et reçoivent un poids basé sur la distance entre les points, modifiée par la densité locale. Cette notion de densité locale permet à l'UMAP de

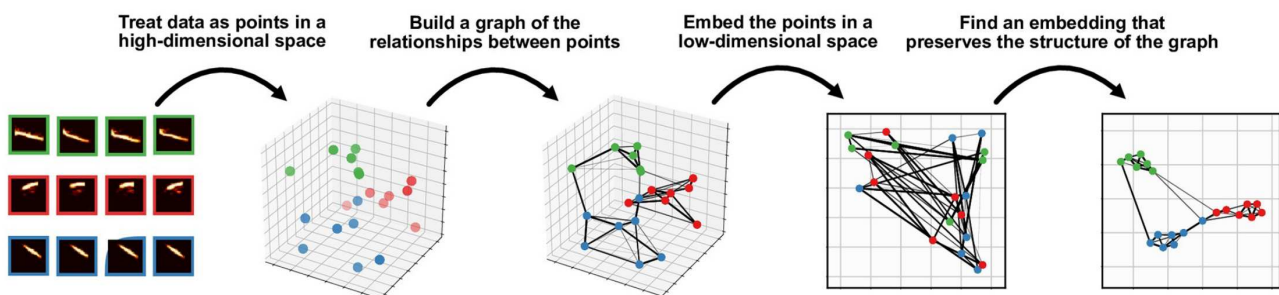


FIGURE 2.5 – Procédure de l'algorithme d'UMAP, d'un nuage de point à haute dimensionnalité à une projection en deux dimensions (extrait de SAINBURG et al., 2020)

rapprocher les points dans les régions de l'espace où de nombreux points sont présents, et d'éloigner les points séparés par des régions vides. Le graphe pondéré est ensuite projeté dans l'espace à faible dimension de façon à préserver au maximum les voisinages entre les points, mais pas forcément les distances. Concrètement, cela signifie que les points voisins dans l'espace d'origine restent voisins après la projection, mais l'UMAP tend à rapprocher les points voisins et à éloigner les points distants.

Contrairement à l'ACP, l'UMAP est non linéaire, et les dimensions finales de la projection ne sont pas interprétables directement (là où les dimensions d'une ACP sont des combinaisons linéaires des variables d'origine). En contrepartie, l'UMAP permet de mieux distinguer les clusters par rapport à l'ACP, en regroupant les points voisins. Le t-SNE est un algorithme similaire à l'UMAP. Sans rentrer dans les détails mathématiques, les deux algorithmes utilisent globalement la même procédure basée sur la construction de graphes, mais le t-SNE utilise des formules différentes à certains points de l'algorithme par rapport à l'UMAP (WANG et al., 2021). Ces différences permettent à l'UMAP une meilleure préservation de la structure globale des données par rapport au t-SNE, ce qui signifie globalement qu'une projection UMAP donne une meilleure idée de l'arrangement des différents clusters<sup>1</sup>. Par ailleurs, de façon très pragmatique, l'UMAP fonctionne beaucoup plus rapidement que le t-SNE, surtout quand les jeux de données deviennent massifs et à haute dimensionnalité<sup>2</sup>.

L'UMAP, comme n'importe quel algorithme de réduction de dimensionnalité, possède un certain nombre de paramètres. Dans ce cas, les paramètres principaux sont : le nombre de voisins  $k$  (qui dépend de l'application et que j'ai donc varié en fonction des études, entre 15 et 30), le nombre de dimensions finales (toujours fixé à 2 dans ce manuscrit, pour permettre la visualisation des résultats), et la distance minimale des points dans la projection finale (fixée à 0, i.e. des points identiques peuvent se superposer dans l'UMAP, pour permettre à l'HDBSCAN ci-dessous de fonctionner correctement). L'utilisateur peut également définir la fonction de distance utilisée par l'algorithme, mais dans les cas décrits dans ce manuscrit j'utilise systématiquement la matrice de distance calculée par la DFTW ci-dessus.

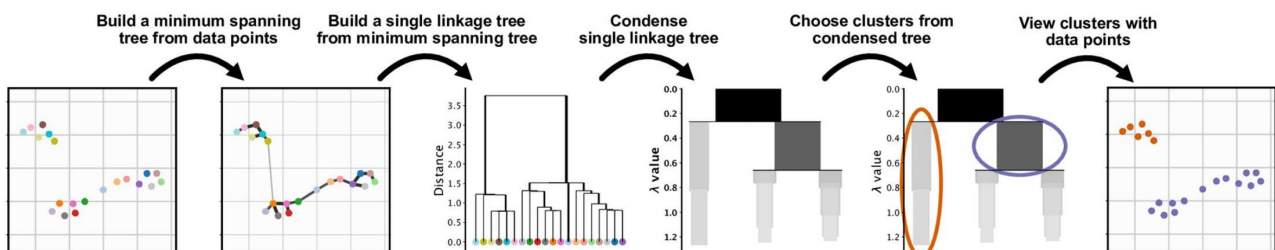


FIGURE 2.6 – Procédure de l'algorithme d'HDBSCAN, d'un nuage de point à un arbre de clustering, à l'extraction des clusters (extrait de SAINBURG et al., 2020).

### 2.3.5 Clustering hiérarchique : HDBSCAN (Hierarchical Density-Based Clustering for Applications with Noise)

Après la projection par l'UMAP, les points correspondants aux vocalisations peuvent être regroupés en cluster. J'ai choisi l'algorithme d'HDBSCAN (CAMPELLO et al., 2015; MCINNES et al., 2017), un algorithme basé sur la densité des points de données dans l'espace. Le principe d'HDBSCAN est basé sur un clustering hiérarchique, où chaque point commence dans son propre cluster, puis les points les plus proches sont agglomérés progressivement, permettant la construction d'un arbre hiérarchique, jusqu'à ce que tous les points soient agglomérés dans un seul cluster (2.6). Ensuite, cet arbre est condensé de façon à trouver les clusters les plus stables, c'est-à-dire ceux où des ensembles de points restent le plus longtemps dans un même cluster en remontant l'arbre hiérarchique. Ces clusters peuvent ensuite être extraits, mais l'HDBSCAN permet aussi de réaliser un clustering diffus : au lieu d'affecter les points à un cluster (ou à aucun, voir ci-dessous), il est possible de d'obtenir pour chaque point un vecteur de probabilités d'appartenir à chaque cluster de la solution finale (WADEWITZ et al., 2015). Cela permet d'incorporer naturellement la gradation des vocalisations dans les analyses, et par conséquent j'utiliserai systématiquement les résultats de clustering diffus pour les analyses statistiques.

Un avantage majeur de l'HDBSCAN est sa robustesse à la forme et à la distribution exactes des clusters. L'algorithme est basé sur la densité des points dans l'espace plutôt que leur distance à un centroïde de cluster (cette importance de la densité des points est la raison du choix du paramètre de distance minimale dans l'UMAP ci-dessus : la fixer à 0 permet aux points de se concentrer en groupes denses, plutôt que de réduire artificiellement la densité en imposant une distance minimale aux points dans la projection). En conséquence, l'algorithme peut détecter des clusters de points quelles que soient leur taille, leur forme, et dans une certaine mesure leur densité absolue, tant que le cluster est plus dense que l'espace qui l'entoure. C'est en contraste majeur avec des algorithmes plus classiques comme le k-means (ou son équivalent diffus, le c-means), qui nécessitent de spécifier le nombre de clusters *a priori* et font l'assomption que les clusters sont des sphéroïdes de tailles similaires. Par ailleurs, HDBSCAN autorise des points à ne pas être attribués à un cluster s'ils sont dans des régions peu denses, ce qui lui permet d'être robuste à la présence de bruit dans les données (encore une fois, contrairement au k-means qui répartit tous les points dans les clusters, au risque de déformer le clustering final).

### 2.3.6 Analyses de complexité : indices de Hill et de Morisita-Horn

Les résultats de clustering diffus obtenus par l'HDBSCAN permettent d'analyser la complexité vocale des corbeaux freux, sans faire d'hypothèses fortes sur la taille du répertoire des individus ou la gradation des vocalisations. Par exemple, le nombre de clusters trouvés par l'HDBSCAN ne correspond qu'à une valeur maximum de la taille du répertoire d'un individu (si tous les individus possèdent le même répertoire).

---

1. <https://pair-code.github.io/understanding-umap/>

2. <https://umap-learn.readthedocs.io/en/latest/benchmarking.html>

Pour quantifier les différents aspects de la complexité vocale définis dans l'Introduction, je me base sur la théorie de l'information et les mesures d'entropie (SHANNON, 1948). Pour étudier la diversité et la gradation des vocalisations, j'utilise notamment les indices de Hill (HILL, 1973), qui quantifient la distribution d'échantillons entre plusieurs classes (comme par exemple des individus répartis entre plusieurs espèces dans un écosystème, ou les vocalisations réparties entre les différents clusters de l'HDBSCAN). Les indices de Hill sont une famille d'indice définis par la formule :

$${}^pD = \left( \sum_{i=1}^C f_i^p \right)^{\frac{1}{1-p}},$$

où  $p$  est un entier positif définissant l'ordre de l'indice,  $C$  est le nombre de classes, et  $f_i$  est la proportion d'échantillons de classe  $i$  dans les données. Noter que le cas limite  $p = 1$  est indéfini avec cette formule. Dans ce cas, l'exponentielle de l'entropie de Shannon est utilisée (avec les mêmes notations que dans la formule précédente) :

$${}^1D = \sum_{i=1}^C f_i \ln f_i.$$

Les indices de Hill sont en particulier une généralisation de l'entropie de Shannon. Ils sont couramment utilisés en écologie pour quantifier la diversité des écosystèmes (CHAO et al., 2014; PRESLEY et al., 2014), et ont été récemment proposés pour des approches comparatives d'étude de la complexité vocale (REBOUT et al., 2021; ZSEBŐK et al., 2021). Selon l'ordre  $p$  choisi, l'indice de Hill est équivalent à des indices courants pour quantifier la diversité : la richesse ( $p = 0$ , i.e. le nombre de classes dans un système), la diversité de Shannon ( $p = 1$ ), ou la diversité de Simpson ( $p = 2$ ). Dans les études de ce manuscrit, j'utilise principalement  ${}^1D$ , la diversité de Shannon, et  ${}^2D$ , la diversité de Simpson. Ces deux diversités représentent une quantification du nombre de types de vocalisations, avec une emphase croissante sur les vocalisations les plus abondantes. Je les ai également choisies car elles sont interprétables comme à un "nombre effectif" de catégories (i.e. le nombre de catégories uniformément distribuées qui aboutirait à la même diversité observée).

J'utilise ces deux indices de diversité pour quantifier la diversité vocale, mais aussi la gradation dans les vocalisations des corbeaux freux. Pour ce faire, je considère les données d'HDBSCAN de deux façons différentes (Fig 2.7 : pour calculer la diversité vocale, je considère chaque individu comme la somme de ses vocalisations, et je calcule la diversité de cette distribution. Pour calculer la gradation, je calcule la diversité de chaque vocalisation, puis j'en étudie la distribution par individu.

Un autre aspect de la complexité vocale est le partage des vocalisations entre individus. Pour estimer ce partage de vocalisations, j'ai calculé la similarité entre les individus avec l'indice de Morisita-Horn (CHAO et al., 2006). Cet indice est interprétable comme la probabilité que deux échantillons, tirés chacun au hasard de deux distributions différentes, appartiennent à une même classe. L'indice de Morisita-Horn classique est défini par la formule :

$$MH = 2 \frac{\sum_i x_i y_i}{\left( \frac{1}{X} \sum_i x_i^2 + \frac{1}{Y} \sum_i y_i^2 \right) XY},$$

où  $x_i$  et  $y_i$  représentent les effectifs de la classe  $i$  dans les deux distributions, et  $X$  et  $Y$  sont les effectifs totaux dans chaque distribution. Pour appliquer l'indice aux résultats de clustering diffus, j'ai

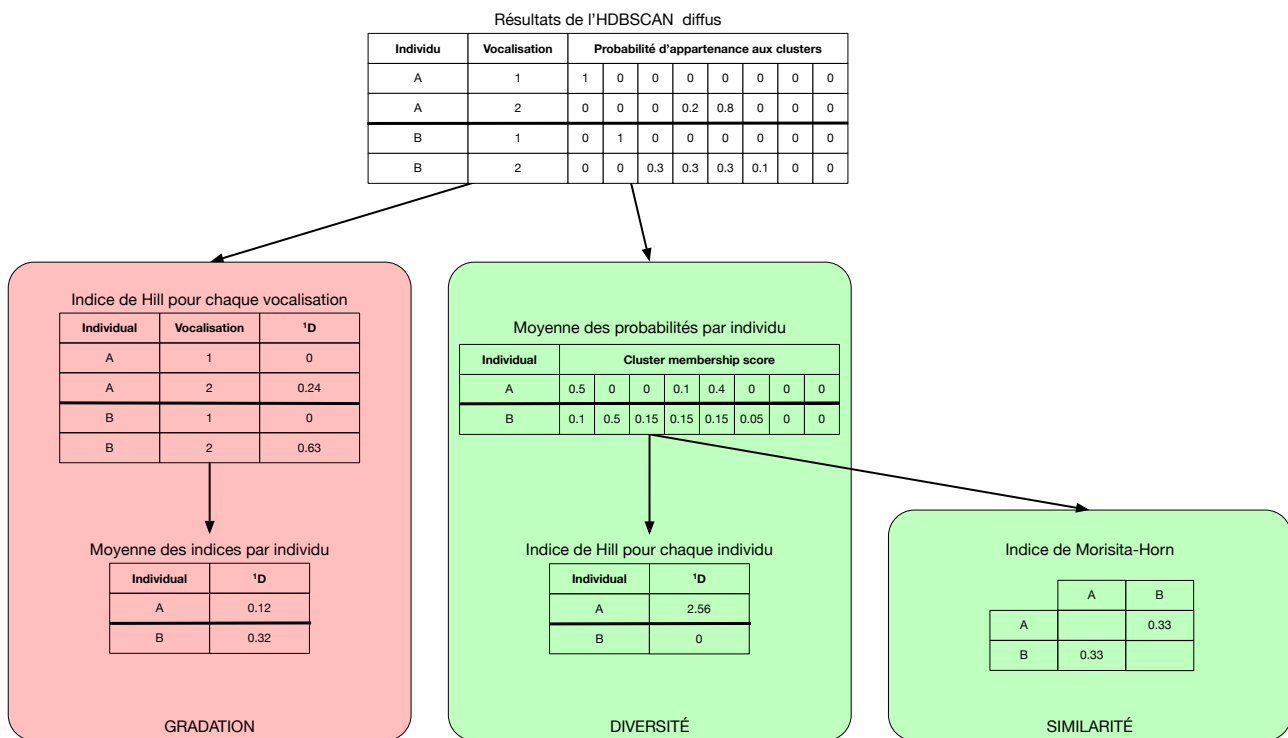


FIGURE 2.7 – Utilisation des résultats d’HDBSCAN diffus pour calculer les indices de complexité dans cette étude. En partant des résultats, le calcul de l’indice de Hill pour la gradation (à gauche) passe par un calcul pour chaque vocalisation, rendant compte du nombre de clusters dans lesquels se trouve chaque vocalisation. Ensuite, l’indice final pour un individu est obtenu en faisant la moyenne des indices des vocalisations de cet individu. Pour le calcul de l’indice de Hill pour la diversité, l’approche inverse est utilisée, en faisant d’abord la moyenne des vocalisations pour un individu, avant de passer par la formule de l’indice. Enfin, le calcul de l’indice de Morisita-Horn pour la similarité (à droite) commence comme pour la diversité, mais diverge ensuite pour appliquer la formule de l’indice aux données.

légèrement modifié la formule : à la place d’un effectif de classe, je considère  $x_i$  comme la somme des probabilités d’appartenir au cluster  $i$  pour l’ensemble des vocalisations de l’individu  $x$ .  $X$  dans la formule précédente est alors le nombre de vocalisations de cet individu. Cet indice, noté  $MH_{fuzzy}$  pour le distinguer de l’indice classique, peut alors être utilisé pour calculer la similarité vocale entre deux individus. Les deux indices  $MH$  et  $MH_{fuzzy}$  sont équivalents dans le cas d’un clustering classique (où les vocalisations sont attribuées à un seul cluster avec probabilité 1, et tous les autres avec probabilité 0), et valent 1 si la distribution des vocalisations des deux individus sont identiques. Je convertis les valeurs de l’indice de Morisita-Horn en pourcentage dans la suite de ce manuscrit, simplement en les multipliant par 100.

### 2.3.7 Apprentissage profond : réseau de neurones et automatisation de la prise de données

Avant toute analyse de la complexité vocale, une grande quantité de données doit être collectée. L’enregistrement des colonies est simple avec le matériel moderne (BLUMSTEIN et al., 2011; DARRAS et al., 2019; FRISTRUP et MENNITT, 2012; POTAMITIS et al., 2014; SHONFIELD et BAYNE,

2017), mais l'extraction manuelle des vocalisations de ces enregistrements est un processus extrêmement long (FRISTRUP et MENNITT, 2012), amplifié par la nécessité d'identifier les individus sur les vidéos en cas d'incertitude. En outre, la présence humaine peut influencer le comportement des oiseaux (CIBULSKI et al., 2014). L'automatisation de la collecte de données permet de grandement réduire ces problèmes. Les développements récents de l'apprentissage profond (*deep learning*) ont conduit à de très nombreuses applications en bioacoustique et plus généralement en écologie, de la détection générale de vocalisation (FANILOUDAKIS et POTAMITIS, 2017; GRILL et SCHLÜTER, 2017; KONG et al., 2017a; KONG et al., 2017b; LIAQAT et al., 2018; SEVILLA et GLOTIN, 2017; STOWELL et al., 2019b) à l'identification d'espèces par leurs vocalisations (JOLY et al., 2021; KAHL et al., 2021a; KAHL et al., 2021b).

La puissance de l'apprentissage profond est liée à l'immense capacité des modèles à apprendre des règles de classification implicites, en apprenant directement des données fournies au modèle (CARUANA, 1997). Un modèle (ou réseau de neurones, utilisé de façon interchangeable dans la suite) est construit à partir de couches successives de neurones. Une couche de neurones est essentiellement un ensemble de poids affectés à chaque entrée de la couche, qui déterminent la valeur de sortie (GOODFELLOW et al., 2015). Ces poids sont appris au cours de l'entraînement par l'algorithme de rétropropagation, qui correspond à un calcul de la différence entre la valeur de sortie du réseau et la valeur de sortie désirée (par exemple, un label particulier associé à une image), propagée ensuite à chaque couche du réseau dans l'ordre inverse des couches, de la couche de sortie à celle d'entrée (d'où le terme de rétropropagation). L'opération effectuée par une couche de neurones est une combinaison linéaire des valeurs d'entrée pondérée par les poids de la couche. Les couches de neurones successives sont séparées par des fonctions d'activation non-linéaires, qui permettent au réseau d'apprendre des relations plus complexes. En l'absence de ces fonctions, un réseau de neurones revient à un modèle linéaire classique, quels que soient ses autres paramètres. En particulier, un réseau à  $N$  couches successives est équivalent à un réseau à une seule couche : sans non-linéarité, le réseau revient à des combinaisons linéaires successives. Hors, une combinaison linéaire de combinaison(s) linéaire(s) reste une combinaison linéaire, limitant le nombre effectif de paramètres appris par le réseau, et donc sa capacité d'apprentissage.

Les opérations effectuées par les couches successives permettent d'intégrer des informations sur des parties de plus en plus importantes des données d'origine. Par exemple, une succession de couches convolutionnelles permet de détecter des motifs particuliers dans les images. Une couche convolutionnelle est définie par un nombre de filtres, correspondant chacun à une matrice de petite taille (par exemple  $3 \times 3$ ) dont les coefficients sont appris au cours de l'entraînement. Chaque filtre est appliqué à l'image reçue par la couche par une opération de convolution : chaque pixel de l'image de sortie de la couche correspond à la corrélation entre les valeurs du filtre et les valeurs des pixels dans une zone de la taille du filtre centrée sur le pixel correspondant. Cela permet de détecter un motif simple (par exemple, une ligne horizontale), qui peut lui-même faire partie d'un motif plus complexe pour la couche convolutionnelle suivante, et ainsi de suite. Les couches convolutionnelles sont paramétrées par ces filtres : nombre de filtres, leur taille, leur pas (e.g. pour un pas de 1 le filtre est appliqué à chaque pixel, pour un pas de 2 le filtre est appliqué tous les 2 pixels, etc.) et leur dilatation (e.g. un



filtre de taille 3 x 3 avec une dilatation de 2 correspond à éclater le filtre sur les coins d'un carré de taille 5 x 5, avec des pixels vides intermédiaires). Ces paramètres conditionnent le nombre de paramètres de la couche. Un plus grand nombre de paramètres donne plus de liberté au réseau d'apprendre les relations nécessaires pour résoudre la tâche sur laquelle il est entraîné. Cependant, cela augmente la taille en mémoire et diminue la vitesse du réseau, puisque plus de calculs sont nécessaires pour traiter les données.

D'autres types d'opérations ont également été utilisées dans la construction du réseau. Les couches de regroupement (*pooling*) permettent de réduire les dimensions de l'image de façon drastique, en combinant plusieurs valeurs de l'image d'entrée de la couche en une seule valeur de sortie. Plusieurs opérations sont possibles, comme le maximum, la moyenne, ou une couche convolutionnelle paramétrée correctement (par exemple avec un pas supérieur à 1). Les couches récurrentes sont utilisées pour traiter des séquences, où une valeur en un point donné dépend des valeurs précédentes. Ces couches suivent des opérations plus complexes et plusieurs variantes existent, mais elles permettent au réseau d'intégrer une certaine "mémoire" des valeurs observées plus tôt dans une séquence. Cela permet d'intégrer des informations temporelles : par exemple, dans notre cas d'analyse de sons, la probabilité de présence d'un son à un instant donné n'est pas indépendante de la probabilité de présence d'un son juste avant ou juste après, et l'opération de récurrence permet d'intégrer cette information.

Dans le cadre de cette thèse, un objectif majeur était de construire un système capable d'analyser les enregistrements sonores pour en extraire des vocalisations attribuées aux individus qui les produisent. Des approches d'apprentissage automatiques passées (non basées sur l'apprentissage profond) suggéraient la faisabilité de cette tâche dans des conditions plus simples (CHENG et al., 2012; E. J. FOX et al., 2008; KIRSCHER et al., 2009). Cependant, pour être applicable à des enregistrements bruts, le système doit être capable non seulement d'identifier les individus, mais de détecter les vocalisations dans un environnement naturel, donc notamment comprenant une grande diversité de bruits d'origines diverses (bruits de véhicules, autres oiseaux, vent, etc.), et ce malgré une grande diversité potentielle dans les vocalisations, tant entre les individus qu'entre les différentes vocalisations d'un individu.

Une limite majeure des systèmes d'apprentissage profond est leur nature de boîte noire, où il est pratiquement impossible de retracer les décisions prises par le réseau une fois que celui-ci est entraîné. Pour essayer de limiter cet aspect, j'ai choisi d'employer une approche en multi-tâche (CARUANA, 1997; LIEBEL et KÖRNER, 2018; RUDER, 2017), en divisant la tâche principale en plusieurs tâches intermédiaires : détection générale des vocalisations de corbeaux freux, sexage acoustique des individus, identification des individus indépendamment du temps, et enfin attribution des vocalisations aux différents individus. Cette division a deux buts principaux, d'abord permettre au réseau de séparer les sous-tâches, ce qui devrait en faciliter l'apprentissage, et ensuite permettre d'accéder à un niveau plus fin dans les décisions finales. Un utilisateur devrait pouvoir choisir, par exemple, à quel point être stringent sur la détection des vocalisations ou sur la précision de leur attribution aux différents individus.

## Chapitre 3

# Rookognise : détection et identification automatique des vocalisations de corbeaux freux en milieu naturel par un réseau de neurones multi-tâches

Killian Martin, Olivier Adam, Nicolas Obin, Valérie Dufour

Publié le 30 Septembre 2022 dans *Ecological Informatics* sous le titre

*Rookognise : Acoustic detection and identification of individual rooks in field recordings using multi-task neural networks*

(DOI : 10.1016/j.ecoinf.2022.101818)

### Résumé

Le suivi d'animaux individuels est essentiel dans de nombreuses études comportementales et bioacoustiques. La collecte et l'annotation de ces données sont des étapes coûteuses en d'effort humain, mais nécessaires avant de procéder à l'analyse. De nombreuses études sur les vocalisations des oiseaux impliquent également la manipulation des animaux ou la présence humaine pendant les observations, ce qui peut biaiser la production vocale. Des unités d'enregistrement autonomes peuvent être utilisées pour collecter de grandes quantités de données sans supervision humaine, éliminant en grande partie ces sources de biais. L'apprentissage profond peut en outre faciliter l'annotation de grandes quantités de données, par exemple pour détecter les vocalisations, identifier l'espèce ou reconnaître les types de vocalisation dans les enregistrements. L'identification acoustique des individus est toutefois restée jusqu'à présent largement limitée à un seul type de vocalisation pour une espèce donnée. Cela a limité l'utilisation de ces techniques pour la collecte automatique de données sur enregistrements bruts, où de nombreux individus peuvent produire des vocalisations de complexité variable, se chevauchant potentiellement les unes les autres, avec la présence supplémentaire de bruits de fond inconnus et variables. Cet article vise à combler cette lacune en développant un système permettant d'identifier des animaux individuels dans ces conditions difficiles. Notre système s'appuie sur une combinaison d'intégration d'information multi-échelle, d'audio multi-canal et d'apprentissage multi-tâche. Le paradigme de l'apprentissage multi-tâche est basé sur la division de la tâche globale en quatre sous-tâches, dont trois sont des tâches auxiliaires : la détection et la segmentation des vocalisations par rapport à d'autres bruits, la classification des individus qui vocalisent à tout moment pendant un échantillon, et le sexage des vocalisations détectées. La quatrième tâche est la tâche globale d'identification des individus. Pour tester notre approche, nous avons enregistré un groupe captif de corbeaux freux, une espèce de corvidé social d'Eurasie, qui compose un répertoire vocal riche. Nous avons utilisé

un réseau de multi-microphones et recueilli un ensemble de données à grande échelle de vocalisations enregistrées, délimitées en temps et attribuées aux individus émetteurs, et avons constaté que le système fonctionnait de manière fiable pour les tâches définies. À notre connaissance, ce système est le premier à identifier des individus indépendamment de la vocalisation produite. Notre système facilitera la collecte de données et le suivi individuel d'animaux en toutes conditions, même sur de longues périodes, et ce quelle que soit la complexité vocale de l'espèce.

### Abstract

Individual-level monitoring is essential in many behavioural and bioacoustics studies. Collecting and annotating those data is costly in terms of human effort, but necessary prior to conducting analysis. In particular, many studies on bird vocalisations also involve manipulating the animals or human presence during observations, which may bias vocal production. Autonomous recording units can be used to collect large amounts of data without human supervision, largely removing those sources of bias. Deep learning can further facilitate the annotation of large amounts of data, for instance to detect vocalisations, identify the species, or recognise the vocalisation types in recordings. Acoustic individual identification, however, has so far largely remained limited to a single vocalisation type for a given species. This has limited the use of those techniques for automated data collection on raw recordings, where many individuals can produce vocalisations of varying complexity, potentially overlapping one another, with the additional presence of unknown and varying background noise. This paper aims at bridging this gap by developing a system to identify individual animals in those difficult conditions. Our system leverages a combination of multi-scale information integration, multi-channel audio and multi-task learning. The multi-task learning paradigm is based the overall task into four sub-tasks, three of which are auxiliary tasks: the detection and segmentation of vocalisations against other noises, the classification of individuals vocalising at any point during a sample, and the sexing of detected vocalisations. The fourth task is the overall identification of individuals. To test our approach, we recorded a captive group of rooks, a Eurasian social corvid with a diverse vocal repertoire. We used a multi-microphone array and collected a large scale dataset of time-stamped and identified vocalisations recorded, and found the system to work reliably for the defined tasks. To our knowledge, the system is the first to acoustically identify individuals regardless of the vocalisation produced. Our system can readily assist data collection and individual monitoring of groups of animals in both outdoor and indoor settings, even across long periods of time, and regardless of a species' vocal complexity. All data and code used in this article is available online.

Keywords: deep learning, bird recognition, outdoor setting, vocalisation-independent, corvid, multi-channel audio

## 3.1 Introduction

Deep learning systems have become invaluable tools in the ecological sciences in recent years. They allow the automation of many monitoring tasks, enabling much wider spatio-temporal coverage than previous methods (Christin et al., 2019; Weinstein, 2019). Deep learning system have been applied to tasks as diverse as detecting species in complex environments (e.g Conrady et al., 2022; Dufourq et al., 2022; Fu et al., 2022; She et al., 2022; van Klink et al., 2022), censusing populations (Adi et al., 2010), surveying breeding success of potentially endangered species (Teixeira et al., 2022), or tracking invasive species both for plants and animals (Campos et al., 2022; Li et al., 2021;

Takimoto et al., 2021). Beyond the species level, observing individual animals remains a challenge for researchers (Ferreira et al., 2020) despite its essential role in ecological and behavioural studies (Clutton-Brock and Sheldon, 2010; Terry et al., 2005). Extensive research has evidenced the possibility of identifying individual animals by their vocalisations (see e.g. Beecher, 1989; Bradbury and Vehrencamp, 1998; Linhart et al., 2019), opening the possibility of applying passive acoustic monitoring techniques (Schneider and Mercado, 2019) to individual monitoring. Bird vocalisations have been studied in the ecological and evolutionary sciences due to their pervasive importance to the life history of birds (see e.g. Catchpole and Slater, 2008; Marler and Slabbekoorn, 2004 for reviews), but also because of their potential analogies with human language (e.g. Sainburg et al., 2019). Current research questions investigate, for example, the interplay between vocal production and sociality in the form of conversations, dialects and vocal signatures in many bird species.

The vocal signature is the set of bioacoustic features that are individually distinctive within a given call type. Such signatures have been found throughout the animal kingdom (Jansen et al., 2012; Kershenbaum et al., 2013; Linhart et al., 2019; McCordic et al., 2016). However, most studies in vocal signature have been limited to a single vocalisation type (e.g. Benti et al., 2019; Boeckle et al., 2012, 2018; Laiolo et al., 2000; Mates et al., 2015; Stowell et al., 2016a; Stowell et al., 2016b; Yorzinski et al., 2006). When encompassing multiple different vocalisation types, the concept of vocal signature becomes a lot more nebulous, as different vocalisation types can bear different signatures (Elie and Theunissen, 2018; Keenan et al., 2020). This plurality of signatures obviously complicates the identification of individuals in species with a rich repertoire, at least with classical methods (Catchpole and Slater, 2008; Elie and Theunissen, 2018). Despite this, there have been a few cases of successful individual identification with multiple vocalisation types (Cheng et al., 2012; E. J. Fox et al., 2008; Kirschel et al., 2009), most successfully with pre-deep learning era neural networks. Still, these approaches were based on supervised algorithms trained on vocalisations manually extracted from the recordings, limiting their applicability to field recordings (Kershenbaum et al., 2014; Thompson et al., 1994).

Neural networks appear as the most appropriate tool to acoustically classify birds down to the individual level. Recent breakthroughs in general machine learning and specifically deep learning (Stowell, 2022) have brought state-of-the-art results just in bird vocalisation detection (Fanioudakis and Potamitis, 2017; Grill and Schlüter, 2017; Kong et al., 2017a; Liaqat et al., 2018; LOSTANLEN et al., 2019; Sevilla and Glotin, 2017; Stowell et al., 2019b), species classification (Kahl et al., 2019; Kahl et al., 2021b), and segmentation of song with classification of vocalisation types (Cohen et al., 2022). But the deep learning community has so far only rarely investigated the problem of individual identification. One possible explanation is that *in situ* recording remains challenging, in particular recording individual birds (Folliot et al., 2022; Stowell et al., 2019a), due to interferences from other sound sources, including conspecifics, and simple wariness preventing the observer from approaching enough to obtain many recordings of good quality. As such, many studies relying on individual recordings simply record the bird alone in a separate chamber (e.g. Sainburg et al., 2019); however, this experimental setting deprives birds of their natural and social environment and can bias vocal production. In addition, acquiring manually individually annotated data is time consuming and gen-

erally costly (Stowell et al., 2019a). Autonomous recording units (ARUs) can be used to record audio in a more ecologically valid environment, i.e. outdoor, in a social group, and without the need for human presence. They have been increasingly used in ecological surveys, since they allow wide coverage with little human effort (Blumstein et al., 2011; Darras et al., 2019; Frstrup and Mennitt, 2012; Potamitis et al., 2014; Schneider and Mercado, 2019; Shonfield and Bayne, 2017).

Deep learning is adequately suited to process the vast amount of complex data that can be collected with ARUs thanks to the representational power offered by these networks. Furthermore, the network's architecture can be adapted to treat the various tasks at (in our case, detection of vocalisation and individualisation). For instance, multiple tasks can be tackled using a multi-task learning approach, i.e. by learning the tasks jointly within a single network. This approach enables the network to learn a shared representation, which has been suggested to effectively bias the learning process towards representations that benefit all tasks (Caruana, 1997; Liebel and Körner, 2018; Ruder, 2017). This shared representation can not only enhance training, but also the generalisation ability of the network. This is usually implemented by sharing early parts of the network, before splitting into task-specific heads. For instance, one head may learn to detect objects over the background, and another may learn to classify the detected objects (Morfi and Stowell, 2018; Pankajakshan et al., 2019). In our approach, we define four such sub-tasks: the *detection* sub-task whose objective is to detect for each audio frame whether a bird is vocalising or not (versus any other sound), the *sexing* sub-task whose objective is to further classify whether the individual producing the vocalisation is a male or a female, the *presence* sub-task whose objective is to identify the presence of a vocalisation of a particular individual within a large audio clip regardless to its precise time position, and the *identification* sub-task, which returns time-stamped vocalisations with the identity of the individual producing each vocalisation.

We then tested the performance of the system on rooks (*Corvus frugilegus*), held in a social group and in an outdoor setting. Rooks are a European social corvid with a rich vocal repertoire (Roskaft and Espmark, 1982). Corvids are known for producing a large number of different vocalisations with a chaotic structure and a wide frequency range (E. D. Brown, 1985; Marzluff and Angell, 2005; Roskaft and Espmark, 1982), particularly difficult to parametrise using classical acoustic measures (Fagerlund and Härmä, 2005; Fletcher, 2000). Furthermore, individuals only share some of their repertoire with their conspecifics (Boeckle et al., 2012; Kondo et al., 2010; Mates et al., 2015), further limiting the use of classical methods for individual identification. Rooks, in particular, may possess an even richer repertoire than previously thought, as they produce series of vocalisations similar to the song of many other oscines (Coombs, 1960), that have never been described in detail. While further description is beyond the scope of this paper, rooks appear to be of particular interest to develop a network capable of vocalisation-independent individual identification.

To summarise, in this paper we present the following contributions, to overcome the aforementioned limitations:

1. We present the first system able to acoustically identify individual birds whatever their vocalisation types, which detects vocalisations from raw recordings and attributes them to the individual emitter. Furthermore, the system was developed using data sampled in an outdoors aviary and

was trained to learn its tasks in an environment with multiple sources of noise. The system is trained in a fully supervised manner, on a dataset of identified rook vocalisations. The code for this system is available online.

2. The architecture is based on a multi-task learning approach. This architecture is used in order to learn information useful for multiple related tasks, and to increase the generalisation ability of the system compared to multiple single-task networks. The multi-task approach exploits information from a shared representation learned by all tasks, followed by task-specific heads. This formulation also lets us fuse the predictions of several sub-tasks together as an attention-like mechanism (e.g. where the detection head returns a low score, there is no need for the identification head to learn to identify an individual). Furthermore, a multi-scale module is added to allow the network to exploit information from very local cues in the spectrogram up to vocalisation-wide patterns to find individually-distinctive information.
3. The proposed multi-task network operates on multi-channel audio signals, as recorded from a multi-microphone array. This allows the system to exploit spatial information to enhance the detection of rook vocalisations as well as the subsequent tasks. The multi-channel recording also adds redundancy in the case of degraded, noisy or weak signals (e.g. saturation due to a bird vocalising close to the microphone, or strong non-vocal noises). To our knowledge, this is the first implementation of a neural network that exploits spatial information in the domain acoustic individual identification.
4. We collected the first large scale annotated dataset of rook vocalisations. This dataset includes vocalisations from 15 individual rooks collected over the course of one year, with each vocalisation time-stamped and attributed to the individual producing it. The data collection involved a multi-microphone array spread throughout the aviary to record each bird from as close as possible. As such, the recordings are composed of multi-channel audio, unlike most approaches in the literature. The entire annotated dataset is freely available online under a Creative Common (CC/BY 4.0) Licence<sup>2</sup>.

## 3.2 Material and methods

### 3.2.1 Data collection

We recorded the vocalisations produced by a captive colony of 15 adult rooks (8 males, 7 females) housed in an outdoor aviary in Strasbourg, France. All birds were identified by coloured leg rings and had been housed together since they had been caught as fledglings, except one adult male added shortly before the start of the recordings. We recorded the group using a multi-microphone array constituted of up to three ARUs, each with two microphones on 3 m cables (Song Meter 4 and SMM-A2 microphones, Wildlife Acoustics, USA). The ARUs were programmed to record for several consecutive hours each morning from January 2020 onwards. The microphones were spaced across

---

1. <https://gitlab.com/kimartin/rook-vocalisation-detection>

2. <https://doi.org/10.5281/zenodo.6091940>

the aviary so the distance between microphone and bird was at most approximately 10 m. The resulting multi-channel recordings were digitised at 48 kHz with a 16-bit resolution. A schematic layout of the aviary, with microphone placements and the usual spots the birds stayed at during observations, can be found in the supplementary material (Fig. A.1).

Among the daily recordings, eleven long sessions and a number of shorter recordings were conducted in the presence of an expert observer (KM). During these sessions, a custom Python program allowed the observer to note the identity of the bird that emitted each vocalisation (thanks to their leg rings). The program also recorded the time of each annotation, which later allowed synchronisation with the audio recordings. These on-site annotations were completed with video recordings also synchronisable to both the annotation and audio files. These videos were used to help resolve potential ambiguities (e.g. to identify which one of two birds vocalised if both were close to each other). These synchronised files, obtained throughout 2020 and 2021 (see Fig. A.2 for a timeline) were used to build an annotated dataset by the same expert observer with the AudioSculpt software (Bogaards et al., 2004). Each vocalisation was annotated with the start time, end time, emitter identity, and whether it was part of a sequence. Due to the multi-microphone array and sound propagation speed, the start and end times were defined as the time when a vocalisation was perceived on any audio channel for the first and last time, respectively. A vocalisation was defined as either an uninterrupted sound or a short succession of small, similar elements that were always produced together. Despite both notes and video assistance, ambiguities sometimes remained regarding the identity of emitters. When this was the case, the emitter was noted as either "Unknown" (single emitter could not be identified), or "Multiple" (multiple emitters vocalised simultaneously). The final annotated dataset included 17,662 vocalisations in 17.4h of audio recordings. The vocalisations were not equally distributed among individuals; one male in particular vocalised much more often than all the other birds (Fig. 3.1). Rooks, like many birds, vocalise following two general modes, calls and song-like sequences. Since our objective is to identify individuals regardless of what vocalisations they produce, we do not separate the two classes in this study.

All EU ethical guidelines (Directive 2010/63/EU) were followed for the care of the rooks throughout the study. No experimental procedure was necessary for the data collection.

### 3.2.2 Audio representation

From the multi-channel raw audio recordings, we extracted Mel-spectrograms corresponding to 10 s of audio, using all microphones available (up to 6). The length of the clips was based on Grill and Schlüter (2017), itself based on the lengths of most clips in the Bird Audio Detection Challenge dataset (Stowell et al., 2019b). This length was the largest value that fit in computer memory with the batch size chosen for the networks. We extracted 10 s clips, applied a pre-emphasis filter with the following equation:  $y[t] = x[t] - x[t - 1]$ , where  $x[t]$  is the original audio at time  $t$  and  $y[t]$  is the corresponding filter output. From the output of this filter, we computed the Short Time Fourier Transform of each clip using a 50 ms Hamming window with 75% overlap, then extracted the squared magnitude of the output. The spectrograms were then downsampled to 80 Mel-scale frequency coefficients to emphasise lower frequencies. We chose these parameters empirically based on the values that pro-

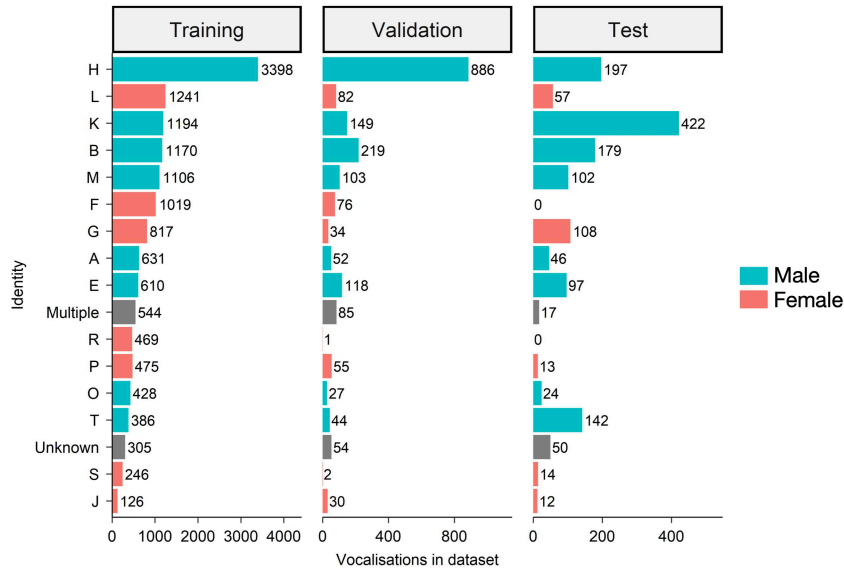


Figure 3.1 – Number of vocalisations per bird and dataset, in decreasing order based on the training dataset. The numbers next to the bars represent the exact number of vocalisations produced by each bird in this dataset. Bar colours represent the sex of the individual. Multiple and Unknown represent cases where vocalisations could not be attributed to an individual bird.

duced spectrograms with good time and frequency resolution during the annotations. This resulted in Mel-spectrograms with dimensions  $(F, T, C)$ , with  $F$  the number of Mel-coefficients,  $T$  the number of spectrogram time frames (800 in our implementation), and  $C$  the number of microphones (6 in our implementation). Smaller clips were zero-padded at the end of each dimension.

### 3.2.3 Network architecture

**Shared front-end.** The front-end (illustrated in Fig. 3.2B) encodes information from the input Mel-spectrograms through successive convolution and pooling operations. The first layer is a Per-Channel Energy Normalisation layer (PCEN, Wang et al., 2017) implemented so that all parameters of the PCEN operation are learned during the training process, with each parameter encoded as a vector of size 80, so that each frequency band has its own value for each parameter of the PCEN equation. The rest of its architecture is based on the sparrow network proposed by Grill and Schlüter (2017), variants of which were used in Schlüter (2018) and Kahl et al. (2021b) to classify bird species with excellent results. Following Grill and Schlüter (2017), we standardise each frequency band and audio channel separately with a Batch Normalisation layer between the PCEN layer and the rest of the front-end. Formally, the front-end corresponds to the following equation:

$$h_E = E(X),$$

where  $E$  is the front-end neural encoder,  $X$  is the input Mel-spectrogram, of dimensions  $(80, 800, 6)$ , and  $h_E$  is the corresponding latent encoding of  $X$  by  $E$ , of dimensions  $(1, 800, 512)$ .  $E$  has an overall receptive field size of  $(80, 39, 6)$ , corresponding to the entire frequency axis, slightly under 0.5 s of time, and all audio channels.



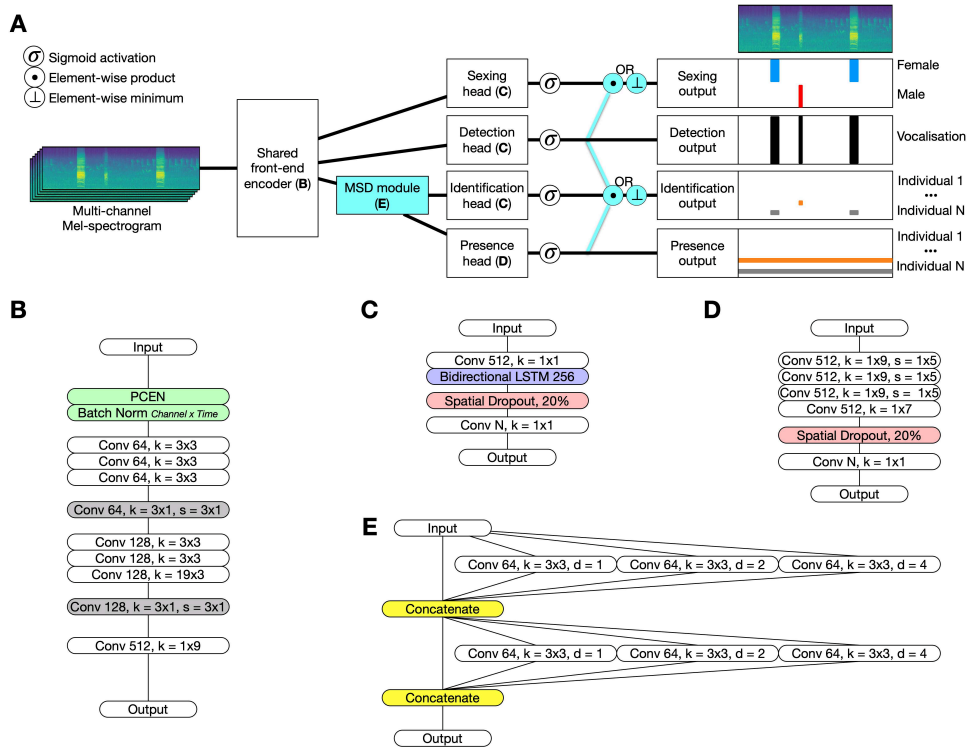


Figure 3.2 – The Rookognise system. A) Overview of the system. Elements in light blue (the MSD module and the two fusion operation, element-wise product and element-wise minimum) are optionally included at training time. On the right: example clip containing three vocalisations, two from a female and one from a male. Under the Mel-spectrogram are the corresponding ground truth labels (in the case of the presence output, stretched in time to match the shape of the other outputs). B) Front-end encoder detail. C) Head detail for the detection, identification and sexing tasks. D) Head detail for the presence task. E) MSD module detail. B-E: boxes represent one layer, with activation layers and Batch Renormalisation layers omitted for readability. Convolution layers (Conv) used the parameters in each box in the following manner, taking for instance the first grey Conv box in B: Conv 64,  $k=3 \times 1$ ,  $s=3 \times 1$  means a convolution with 64 filters of size  $3 \times 1$  (3 along the frequency axis, 1 along the time axis) and stride  $3 \times 1$  (likewise). Boxes coloured according to operation type (green: normalisation, white: convolution, grey: pooling, blue: LSTM, red: dropout, yellow: concatenation). N: number of classes for a particular task (detection: 1, sexing: 2, presence and identification: 15).

**Multi-Scale Densely-connected (MSD) module.** Salient information for individual identification may be present at multiple scales from local to vocalisation-wide. Several mechanisms exist to integrate multi-scale information, but we chose a module proposed by Pelt and Sethian (2017) for its simple design, which the authors note make it applicable to a wide variety of tasks aside from their original task in medical imaging. This module is built by stacking  $M$  successive blocks, each with  $L$  parallel dilated convolutional layers with identical parameters except for the dilation rate  $d$ . We build the module with  $M = 2$  blocks and  $L = 3$  convolutions per block, using dilation rates  $d$  of 1, 2, and 4 (Fig. 3.2E), although we did not extensively experiment with these settings. All convolutions use the same parameters except for this dilation rate: 64 filters,  $1 \times 3$  filter size. The module is placed directly after the front-end. Formally, this corresponds to the equation:

$$h_M = MSD(h_E),$$

where  $MSD$  is the module and  $h_M$  is the latent encoding of  $h_E$  by  $MSD$ , of dimensions (1, 800, 512).

**Task-specific heads.** We implemented the multi-task learning objective by defining heads with dedicated architecture for each classification task. All heads received the output of either the front-end or the MSD module. The detection, sexing and identification head shared the same architecture (Fig. 3.2C) including a pointwise convolution layer, a spatial dropout layer, a bidirectional LSTM layer, and a final pointwise convolution. All three heads thus preserved time resolution. The presence head summarised over the time axis using three convolutions with kernel size 1x9 and stride 1x5, followed by a convolution summarising along the remaining time axis information, and finally a pointwise convolution. All heads end with a sigmoid activation for the specific classification task. Formally, this corresponds to:

$$\hat{y}_{task} = H_{task}(h_X), task \in \{det, sex, pres, id\},$$

such that  $\hat{y}_{det}$  is the output of the detection head  $H_{det}$  of dimensions (1, 800, 1),  $\hat{y}_{sex}$  is the output of the sexing head  $H_{sex}$  of dimensions (1, 800, 2),  $\hat{y}_{pres}$  is the output of the presence head  $H_{pres}$  of dimensions (1, 1, 15) (15 corresponding to the number of individuals in the study group), and  $\hat{y}_{id}$  is the output of the identification head  $H_{id}$  of dimensions (1, 800, 15).  $h_X$  can be the encoded output of either the front-end  $h_E$  or the MSD module  $h_M$ .

**Fusing decisions.** We finally investigated whether explicitly combining the outputs of different heads could facilitate learning. For instance, combining detection and identification may allow the latter to only attend to frames containing vocalisations, as frames with low detection outputs are unlikely to contain vocalisations. This combination operation was always added after the sigmoid at the end of each head. We tested two operations: element-wise minimum and element-wise product. In our experiments, we only fused outputs with shapes naturally compatible with element-wise operations: 1) detection, presence and identification, and 2) detection and sexing (Fig. 3.2A).

### 3.2.4 Implementation details

In this section we detail the implementation of the neural network. We first describe non-standard elements used in the architecture, followed by the regularisation strategies used to prevent overfitting, and finally the loss function and the optimizer algorithm used to actually train the network.

**Non-standard architecture elements.** The field of deep learning has been extremely active in recent years, with both new applications and new architecture elements regularly proposed to improve previous results. We selected some of these elements for their theoretical benefits on performance, briefly described below, and included them in the final system after experimentally validating these benefits. As these elements are not part of our contributions, we do not focus on them further; however, in Table A.2 we show how using these non-standard elements improved the system compared to their standard counterparts. First, we used the Mish activation function instead of the ReLU family of activations, for its continuous derivability and self-regularising property, that prevent gradient vanishing and thereby improve training (Misra, 2019). Second, we used Batch Renormalisation (Ioffe, 2017) instead of Batch Normalisation. Batch Normalisation improves training efficiency and adds robustness to specific initialisation quirks (Ioffe and Szegedy, 2015), but suffers when batches are not independent and identically distributed, which Batch Renormalisation aims to correct. Third, we used strided convolutional pooling (Pankajakshan et al., 2018) instead of max pooling. Max pool-

ing essentially discards the majority of the data when downsampling since only the highest value for each pooling operation is kept. We expected this may learning, as individual identity may be encoded by fine-grained cues. Therefore, we chose strided convolutional pooling to avoid discarding any information during training.

**Regularisation.** Regularisation prevents the network from overfitting the training data, which risks losing generalisation ability to unseen data. We used several of these techniques in our experiments. First, we used early stopping with a waiting period of 15 epochs, to directly stop training when validation data performance no longer improved. The waiting period was based on the learning rate cycle (see below), to allow at least two minima of the learning rate to occur before stopping. Second, we used spatial dropout with a rate of 20% before the final layer of each head. Spatial dropout prevents overfitting by randomly dropping entire feature maps at each iteration of training, which prevents feature co-adaptation that could lead to overfitting (Srivastava et al., 2014). Third, we applied L2-norm weight decay with a strength of  $10^{-4}$ . This penalises large absolute values of the weights that can destabilise training (L. N. Smith, 2018). Fourth, we used label smoothing so that the actual training labels used to minimise the loss functions were not 0 and 1 but 0.05 and 0.95, respectively. Label smoothing similarly encourages lower weights for stable training, particularly in the last layer of the network (Szegedy et al., 2016).

**Objective function and optimizer.** All networks were trained to minimise a focal binary cross-entropy loss (FL, Lin et al., 2020) with the formula:

$$FL(y, \hat{y}) = -y \cdot (1 - \hat{y})^\gamma \cdot \log(\hat{y}) - (1 - y) \cdot \hat{y}^\gamma \cdot \log(1 - \hat{y}),$$

where  $y$  is a ground truth label,  $\hat{y}$  is the corresponding network prediction, and  $\gamma$  is a focusing parameter, which reduces the contribution of correctly classified samples to the loss. When  $\gamma = 0$ , FL reduces to the standard binary cross-entropy. We chose  $\gamma = 2$  in our experiments as it improved over the standard binary cross-entropy (see Table A.2 for details). The mean of the FL over all samples was used for each head. In the multi-task configurations, we used the unweighted sum of the loss corresponding to all four tasks:

$$FL_{multi-task} = \sum_{task \in \{det, sex, pres, id\}} FL_{task}(y_{task}, \hat{y}_{task}),$$

We used the Ranger21 optimiser (Wright and Demeure, 2021), with weight decay  $10^{-4}$ . Ranger21 serves to solve some issues with more usual optimisers such as Adam, which can be unstable at the beginning of the training process and so derail the training procedure, along with integrating a number of heuristics which improve various aspects of training (see Table A.2 for a comparison between Ranger21 and Adam). We chose a cyclical learning rate schedule (L. N. Smith, 2018), linearly interpolating between  $10^{-6}$  and  $10^{-3}$  with cycles of 6 epochs, and halved the maximum learning rate after each cycle. Cyclical learning rates can greatly accelerate convergence by allowing the network to update much faster when the learning rate grows, and converge to more precise minima of the loss when the learning rate decreases. Finally, we used mixed precision training (Narang et al., 2018) to accelerate training and decrease memory footprint.

### 3.2.5 Training procedure

The networks were trained for a maximum of 50 epochs to minimise the FL on the Mel-spectrograms of the multichannel recordings. We semi-randomly distributed the annotated sessions at the start of the experiments. We adjusted the distribution in the training data so all individuals were represented as equally as possible, although this also means not all individuals could be represented in the validation and test data (Fig. 3.1). The annotation files were not split during either training or evaluation, to avoid data leakage due to the way vocalisations were sampled during training.

For the training data, clips were randomly sampled around each vocalisation so as to include this vocalisation in its entirety (although other vocalisations in the same clip could be cut off). This meant the network was unlikely to ever receive the same spectrogram twice, encouraging it to learn the underlying structure of the data. Additional clips were randomly sampled from background-only parts of the recordings to increase the variety of background noises. For the validation and the test data, the recordings were simply split into adjacent, non-overlapping clips and sent in order to the network.

Vocalisations from the Unknown or Multiple classes were excluded from training by assigning them a weight of 0 during loss computation. The detection task was exempt from this, as identity is irrelevant to whether or not a vocalisation occurs.

Finally, we used the same random seed to initialise all runs for fair comparison. This meant all networks started from the same initial weights, and the training data was always fed to the network in the same order at a given epoch, so all differences between networks should be due to differences in architecture.

### 3.2.6 Experiments

We investigated and compared several configurations to build the final network, following the description in section 3.2.3. For clarity, all configurations were named using the nomenclature in Table 3.1. First, single-task networks were trained and evaluated to form a baseline. For each sub-task, one network was trained without and with the MSD module. Then, we trained the multi-task network, using the best configurations of each sub-task from the single-task experiment. Following this, we included either of the two fusion operations described above in the multi-task network. As a final test, we investigated the impact of using multi-channel audio compared to using only one channel. We hypothesised that multi-channel audio is more efficient due to including additional information that the network may access, such as delays between channels or simple differences in distance: while with all six microphones the maximum distance between bird and microphone was only about 10 m, with one channel distance to microphone it might increase up to about 26 m. The configuration corresponding to this case of using a single audio channel was named the Rookognise\_RC (Random Channel) configuration.

Table 3.1 – Description and nomenclature of the configurations in the experiments. From left to right: Configuration is the name used in the text for each network, the Task learning column denotes whether the network learned a single task or used the multi-task approach, the Task column corresponds to the task(s) learned in a given configuration, the MSD module column corresponds to whether the MSD module is included or not in a given configuration, and the Combination column refers to the element-wise operation used for combining decisions in the configuration, if any.

Configuration	Task learning	Task	MSD module	Combination
ST_Det	Single	Detection	No	
ST_Det_MSD	Single	Detection	Yes	
ST_Sex	Single	Sexing	No	
ST_Sex_MSD	Single	Sexing	Yes	
ST_Pres	Single	Presence	No	
ST_Pres_MSD	Single	Presence	Yes	
ST_Id	Single	Identification	No	
ST_Id_MSD	Single	Identification	Yes	
MT_None	Multi	All	Yes	
MT_Product	Multi	All	Yes	Product
MT_Min	Multi	All	Yes	Minimum

### 3.2.7 Evaluation

We ranked our models using the Area Under the Receiver Operating Curve (AUROC) and the Area Under the Precision-Recall Curve (AUPRC). AUROC and AUPRC both summarise the performance of a network over its entire output range. AUROC indicates the probability a classifier will score a random positive sample higher than a random negative sample. No matter the dataset, random guessing leads to 50% AUROC and perfect classification to 100% AUROC (Fawcett, 2006). However, AUROC can over-emphasise the correct rejection of negative samples, especially if negative samples outnumber positive samples, in which case AUPRC can be more representative (Saito and Rehmsmeier, 2015). AUPRC is based on precision (the ratio of true positives over predicted positives) and recall (the ratio of true positives over labelled positives), and evaluates the ability of the classifier to correctly classify positive samples without false positives or negatives, and ignores true negatives (i.e. correct rejections). Unlike AUROC, AUPRC is not independent of the proportion of positive samples in the dataset, and so cannot be compared between different datasets. During our experiments, we saved the models at the epoch with the highest AUPRC, as computed on the validation dataset. In the multi-task configurations, only the identification task validation AUPRC was used. The reported values of AUROC and AUPRC were the average of the AUROC and AUPRC associated with each class. The average was computed from the classes present in a given dataset (i.e., only 13 of the 15 birds were present in the test dataset, so the corresponding results were averaged over 13 classes instead of 15).

The ROC and PRC curves represent the performance of the network at various decision thresholds. The AUROC and AUPRC metrics integrate over all possible threshold values; as such, while they are convenient to evaluate network performance, they are not sufficient when we are interested in selecting a threshold when we apply the network to actual unseen data. To do so, we used the curves

themselves. Since our dataset is heavily imbalanced, we used the AUPRC curve, and we chose to directly plot precision and recall as a function of the threshold value for clearer visualisation.

### 3.3 Results

We first built the Rookognise system by comparing single-task networks with and without the MSD module, before combining the best configurations into a single multi-task network. We then compared the multi-task learning-only approach to the approaches incorporating output fusions, to arrive at the final system. We briefly investigated the benefit of using the multi-channel audio dataset compared to a mono-channel version of the dataset. Finally, we illustrated the use of the Rookognise system on unseen data from the test dataset.

#### 3.3.1 Integrating multi-scale information with the MSD module

In this section we investigated the benefits of multi-scale information integrated by the MSD module. We established the baseline performance of the system by training single-task networks dedicated to each of the four sub-tasks, with and without the MSD module (Table 3.2). In the following, we report results on the test data: positive numbers where the MSD module improved performance, and negative numbers otherwise. For the detection sub-task, ST\_Det outperformed ST\_Det\_MSD (-7.34% AUROC, -3.23% AUPRC). For the sexing sub-task, ST\_Sex outperformed ST\_Sex\_MSD (-0.07% AUROC, -0.01% AUPRC). For the presence sub-task, ST\_Pres\_MSD outperformed ST\_Pres (+2.88% AUROC, +1.28% AUPRC). Finally, for the identification sub-task, ST\_Id\_MSD outperformed ST\_Id (+0.94% AUROC, +3.22% AUPRC). Overall, the MSD module improved performance for the tasks involving more complex classification (in this case, with higher number of possible outputs). Though the improvement was not systematic for all the sub-tasks, the integration of multi-scale information in the MSD module likely improved the ability of the network to disentangle important information at both local and global scales to identify specific individuals. Overall, the MSD module efficiently captured multi-scale time-frequency information in audio signals for complex tasks. On the other hand, the front-end encoder alone was sufficient to detect rook vocalisations against all other sounds, and to classify rook vocalisations by sex, likely simpler tasks to learn.

#### 3.3.2 Multi-task learning and fusing sub-task decisions

In this section, we investigated the impact of multi-task learning. We compare the best single-task networks determined in section 3.3.1 with the multi-task configurations MT\_None, MT\_Product, and MT\_Minimum (Table 3.2, middle part). In this case, we report positive numbers where the MT configurations outperformed the ST networks, and negative numbers otherwise. The MT\_None configuration already outperformed the single-task configurations on the presence (+1.47% AUROC, +8.34% AUPRC) and identification (+0.69% AUROC, +2.36% AUPRC) sub-tasks, though not on the detection sub-task (-4.57% AUROC, -4.69% AUPRC). Adding either fusion operation maintained or even amplified the improved performance on presence and identification, and also improved de-

Table 3.2 – Test data performance of the networks, at the best epoch determined by validation data AUPRC. For each column, the best performance is emphasised in bold (separately for single-task and multi-task configurations). Underlined values represent where the multi-task configurations outperformed the corresponding best single-task network. The last line corresponds to the mono-channel configuration, obtained by training the network using only one of the microphones randomised for each audio clip.

Configuration	Detection		Sexing		Presence		Identification	
	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC
ST_Det	<b>94.03</b>	<b>81.04</b>						
ST_Det_MSD	86.69	77.81						
ST_Sex			<b>98.88</b>	<b>89.72</b>				
ST_Sex_MSD			98.81	89.71				
ST_Pres					83.62	53.11		
ST_Pres_MSD					<b>86.50</b>	<b>54.39</b>		
ST_Id							92.87	55.28
ST_Id_MSD							<b>93.81</b>	<b>58.50</b>
MT_None	89.46	76.35	97.83	87.26	<u>87.97</u>	<u>62.73</u>	<u>94.50</u>	<u>60.86</u>
MT_Min	<u>94.69</u>	<u>81.72</u>	97.64	87.47	<b>89.94</b>	<b>64.65</b>	93.47	<u>61.30</u>
MT_Product (Rookognise)	<b>96.95</b>	<b>89.31</b>	<b>98.21</b>	<b>88.95</b>	<u>88.03</u>	<u>63.64</u>	<b>96.11</b>	<b>63.06</b>
Rookognise_RC	92.35	81.78	95.04	81.65	84.15	58.70	93.70	59.29

tection performance. In particular, MT\_Product obtained greatly improved performance on all three sub-tasks compared to any other configuration. Interestingly, none of the MT configurations outperformed the single-task networks on the sexing sub-task, although MT\_Product was only slightly worse (-0.67% AUROC, -0.77% AUPRC). Overall, the multi-task approach greatly improved learning ability on the more complex tasks. Fusing the decisions of multiple sub-tasks further boosted performance, especially in the detection sub-task which did not benefit from the multi-task approach alone. In particular, using the element-wise product resulted in the strongest overall performance. Following these experiments, we selected the MT\_Product configuration as the final system, which hereafter corresponds to the "Rookognise" system.

### 3.3.3 The impact of multi-channel audio

Using multi-channel audio includes a lot more information than using only one audio channel. As just one example, vocalisations clearly audible on one channel may not be audible on another (Fig. 3.3). In a final experiment, we investigated to what extent the Rookognise system may benefit from the multi-channel aspect of the training data. To do so, we simply re-trained the system in the Rookognise configuration, using a single-channel version of the dataset obtained by randomly sampling one audio channel for each sample, resulting in the Rookognise\_RC configuration (Table 3.2, last line). This configuration obtained lower performance than the Rookognise system on all sub-tasks (detection: -4.59% AUROC, -7.53% AUPRC; sexing: -2.94% AUROC, -7.30% AUPRC; presence: -3.89% AUROC, -4.94% AUPRC; identification: -2.41% AUROC, -3.7% AUPRC). This shows the ability of the network to efficiently capture additional (e.g. spatial) information as provided by the multi-channel microphone array, in particular for the detection of rook vocalisations. Nevertheless,

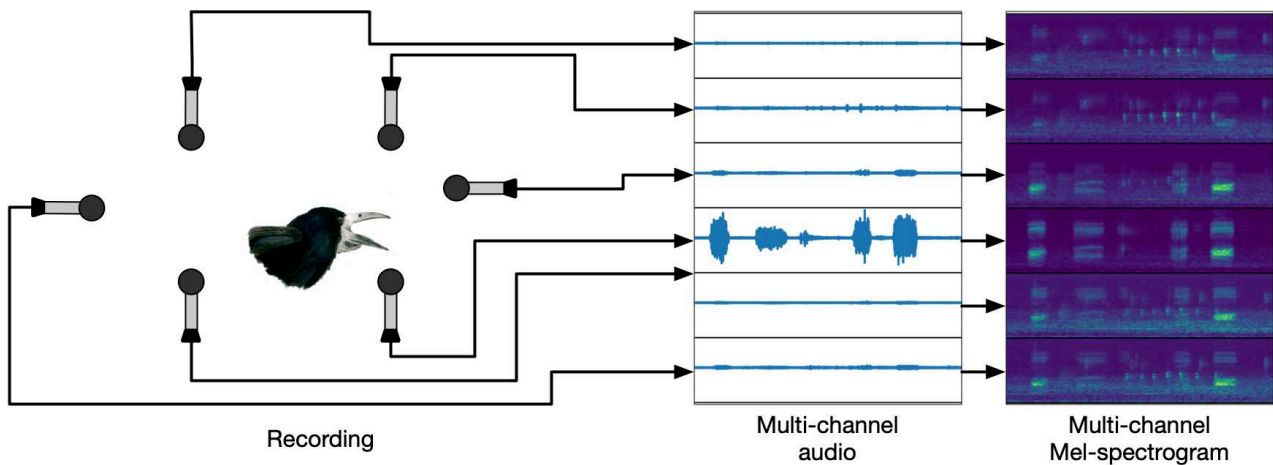


Figure 3.3 – Schematic illustration of the multi-channel setup. Here, a rook is vocalising closest to the rightmost microphone (left), corresponding to the fourth channel from the top in the audio recording (middle) and in the corresponding Mel-spectrograms (right). On the other hand, the leftmost microphone, furthest from the rook, barely records anything over the background noise

Rookognise\_RC still obtained reasonable performance despite using a single audio channel, making it usable if storage space for the audio recordings is a concern.

### 3.3.4 Detecting vocalisations with the Rookognise system

We examined in further detail the performance of the Rookognise system on unseen data, starting with the detection of rook vocalisations from raw recordings. We computed the precision (ratio of true positives to predicted positives) and recall (ratio of true positives to labelled positives) metrics, as well as the proportion of the recording data retained as a function of threshold value (between 0 and 1, corresponding to the output range of the system) for each sub-task (Fig. 3.4). For the detection sub-task (Fig. 3.4, upper left), a low threshold of 0.05 was already enough to discard approximately 75% of the data, i.e. most background noise could be easily eliminated. This low threshold missed very few vocalisations (95.0% recall) but also falsely classified some noise as rook vocalisations (35.9% precision). Conversely, a higher threshold of 0.49, computed to optimise the F1-score (the harmonic mean of precision and recall), had much better precision (90.1%) and retained only 8.5% of the data, at the cost of lower recall (79.2%). A good threshold value would of course depend on the user, but optimal thresholds may lie between those two values to balance human effort reduction (with a high threshold) and false detections (with a low threshold).

### 3.3.5 Identifying emitters with the Rookognise system

We finally directly evaluated how well the network identified individual rooks. For the sake of this evaluation, we used only predictions corresponding to vocalisations from identified individuals in the test labels (i.e. we discarded all Unknown or Multiple class vocalisations, as well as overlapping vocalisations). Each vocalisation frame was then attributed to the individual with the highest predicted score. We computed the confusion matrix from the resulting predictions, as well as the corresponding



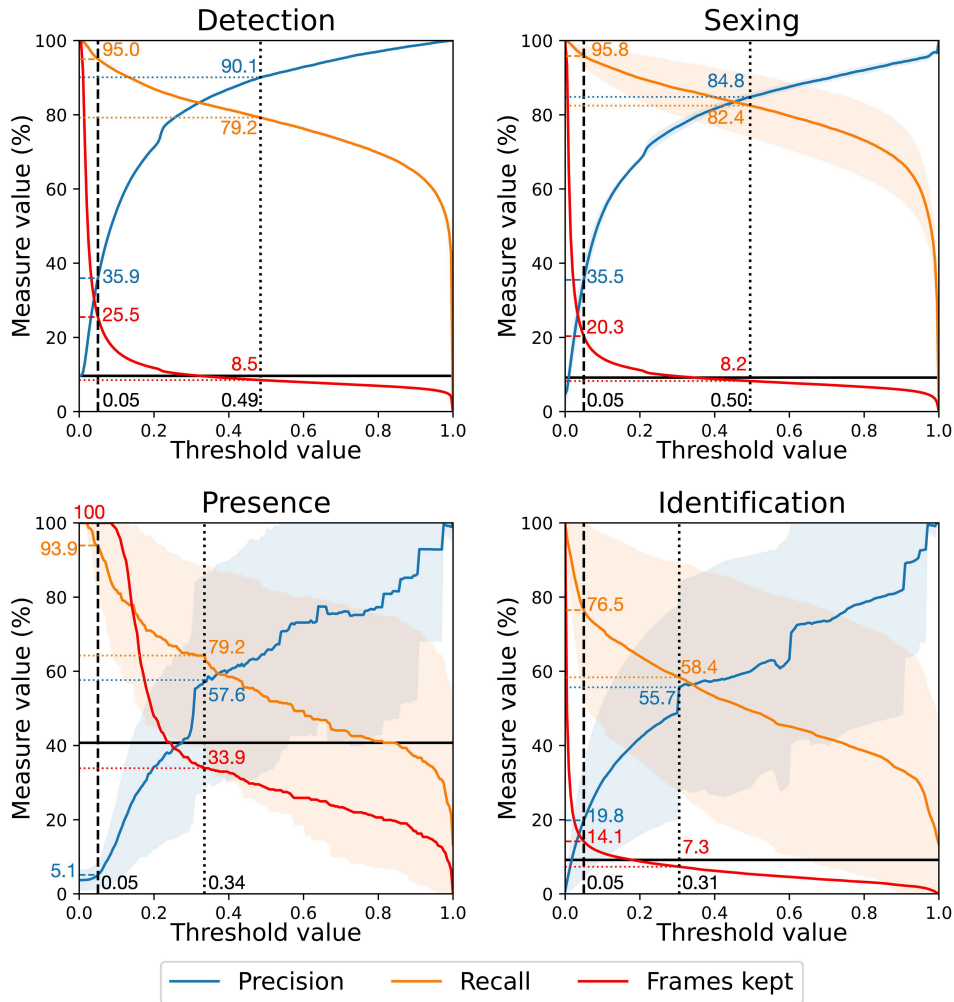


Figure 3.4 – Test performance of the Rookognise system for each sub-task. Performance is measured using precision, recall and proportion of spectrogram frames kept (as a proxy of human effort reduction). For the sexing, presence, and identification sub-tasks the precision and recall curves are averaged over classes and the shaded area represents one standard deviation away from this average. In all plots, the black vertical lines corresponds to specific threshold values: dashed lines correspond to the conservative threshold at 0.05, dotted lines to the threshold optimising F1-score (the harmonic mean of recall and precision). Coloured horizontal lines are the values of all three metrics at the corresponding thresholds, with the associated values in the corresponding colours.

Table 3.3 – Confusion matrix computed from the test data predictions. The predictions were restricted to labelled vocalisations, with overlapping and non-attributed vocalisations removed. Birds are coloured by sex (males in blue, females in red). Columns correspond to labels, rows to predictions. For instance: for the cell in the first row (A), second column (B) the network predicted 2 frames as produced by A that were actually produced by B. Diagonal cells (in grey) correspond to correct identifications. For each bird, the second-to-last column indicates predicted frames, the last column indicates precision, the second-to-last row indicates labelled frames, and the last row indicates recall. The four bottom-right cells contain: total frames (top left), average recall (bottom left), average precision (top right) and overall accuracy (bottom right). Individuals that were absent from the test dataset (F and R) had their recall and precision left blank (note that recall is undefined and precision is 0 in the absence of positive samples) and did not contribute to the averages.

	A	B	E	F	G	H	J	K	L	M	O	P	R	S	T	Predicted	Precision
A	1009	2	0	0	0	8	0	130	0	71	0	1	0	0	0	1221	82.6
B	14	3943	111	0	0	123	0	8	0	12	92	0	0	0	918	5221	75.5
E	0	8	437	0	0	0	0	81	4	204	0	0	0	0	124	858	50.9
F	0	0	13	0	18	0	0	0	53	0	0	0	0	0	0	84	-
G	1	22	41	0	2729	5	0	748	45	0	26	0	27	0	0	3644	74.9
H	4	0	106	0	0	3957	0	2078	6	0	0	0	0	0	20	6171	64.1
J	0	0	0	0	0	0	251	0	45	0	0	0	0	2	0	298	84.2
K	3	13	159	0	0	35	0	5541	3	8	5	0	0	0	934	6701	82.7
L	0	9	210	0	11	0	76	4	1583	0	196	0	15	0	0	2104	75.2
M	0	3	0	0	9	0	0	213	0	1783	372	0	0	0	65	2445	72.9
O	0	0	64	0	0	0	0	676	1	73	33	0	0	0	0	847	3.9
P	0	0	73	0	0	0	0	31	0	0	0	81	0	0	0	185	43.8
R	0	0	0	0	0	0	0	0	21	0	0	16	0	0	0	37	-
S	0	0	0	0	21	0	0	0	0	0	0	0	0	179	0	200	89.5
T	0	7	214	0	0	402	0	24	0	0	0	0	0	4	439	1090	40.3
Labelled	1031	4007	1428	0	2788	4530	327	9534	1761	2151	502	320	0	227	2500	31106	64.7
Recall	97.9	98.4	30.6	-	97.9	87.4	76.8	58.1	89.9	82.9	6.6	25.3	-	78.9	17.6	65.2	70.6

precision and recall associated with each individual (Table 3.3). The Rookognise system had strong overall performance identifying individuals, attributing 70.6% of vocalisation frames to the correct individual, with an average precision of 65.2% and average recall of 64.7%. Seven individuals in particular were identified with both precision and recall over 75%. Several individuals (O, P, and E) were more poorly identified. While these individuals were among the less represented in the data (e.g. O has 428 vocalisations in the training dataset), other individuals (e.g. S with 239 vocalisations and J with 125 vocalisations) were much better identified (S: 89.5% precision and 78.5% recall; J: 84.2% precision and 76.8% recall). In general, we found no correlation between identification performance (whether measured as precision or recall) and number of vocalisations in either the training or test data (four Pearson correlation tests: either precision or recall, correlated with number of vocalisations in either training or test data, all  $p > 0.05$ ). We further examined two noteworthy cases.

The first case was the misattribution of some of K’s vocalisations to H (Table 3.3, 6th row, 8th column), which was almost exclusively due to the misidentification of one long song-like bout produced by K. These two males, especially H, were also the ones who produce the longest song-like bouts.

The second case was the misattribution of almost all of O’s vocalisations to M (Table 3.3, 9th row, 10th column), resulting in O being by far the worst-identified individual in the group at only 3.9% precision and 6.6% recall. These two males shared several vocalisation types produced by not other individual, and to our ear, they sounded very similar in general. However, this was also the case for a pair of females, J and L, but these females were successfully identified by the network (J: 84.2% precision, 76.8% recall; L: 75.2% precision, 89.9% recall).

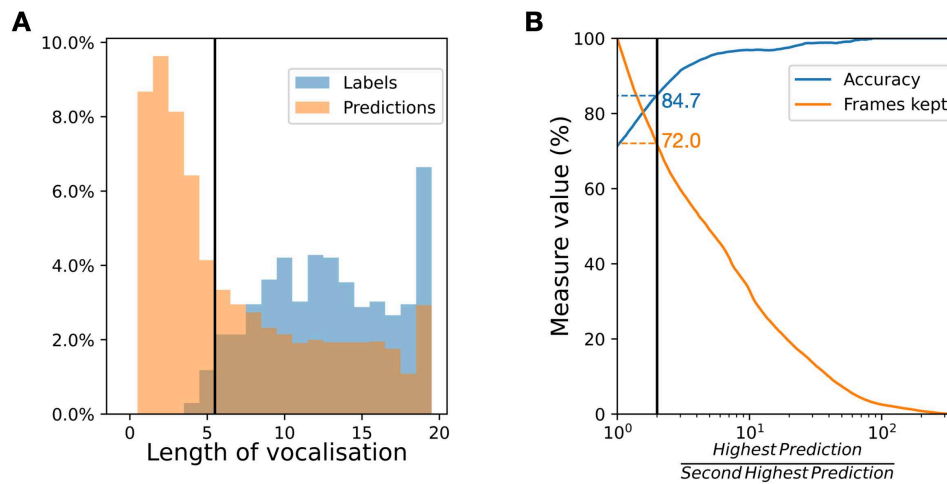


Figure 3.5 – Post-processing of the system predictions. A) Distribution of the Rookognise detection sub-task predictions (orange bars) and of the labelled vocalisations (blue bars) by length in frames. The black line represents the threshold, chosen at 5 consecutive frames. Given the distribution of lengths in the labels, Rookognise predictions shorter than 5 frames are likely false positives and can be safely removed. Only detections (for both labelled and predicted vocalisations) up to 20 frames long are included to avoid crowding the plot. B) Accuracy and proportion of frames retained as a function of the ratio between the highest and second-highest prediction for the identification task. Both measures are computed based on the frames corresponding to labelled vocalisations, as in Table 3.3. The black line represents the point where the ratio of highest prediction to second-highest prediction is 2. Identification predictions are only retained if the highest prediction sufficiently higher, meaning the network is confident in identifying one individual.

### 3.3.6 Post-processing decisions: an example on the test data

In this final section we combined the decisions determined in the previous sections, along with some simple post-processing steps which can improve the accuracy of the Rookognise predictions. We then separately compared their effects on the test data labels. First, we used the detection threshold of 0.05. Second, following Cohen et al. (2022), we discard short predictions that are likely false positives; we set the threshold at 5 successive frames, corresponding to 0.0625 s at our time resolution (Fig. 3.5A). The 5 frames threshold effectively discarded 37% of false positive detections, at the cost of removing 1.5% of the labelled vocalisations. This had relatively minor effect on the detection precision (37.8% instead of 35.9% without discarding short detections) and recall (94.7% instead of 94.9% without discarding short detections), since we remove relatively few frames in total. Third, we discarded predictions for the identification if the highest predicted score is not at least twice as high as the second highest score (Fig. 3.5B). This effectively discarded predictions where the network was not sufficiently confident in attributing the vocalisation to an individual, and increased accuracy to 84.7% (against 70.6% without this step) while retaining 72.0% of vocalisation frames. We show two examples of system outputs after applying these steps in Fig. 3.6. In Fig. 3.6A, four calls from the female L and one call from the male A were correctly detected, sexed, and attributed to the correct individual. In Fig. 3.6B, we illustrated a clip containing a long song-like bout produced by H. Again, the network correctly detected, sexed and attributed these vocalisations, including separating most

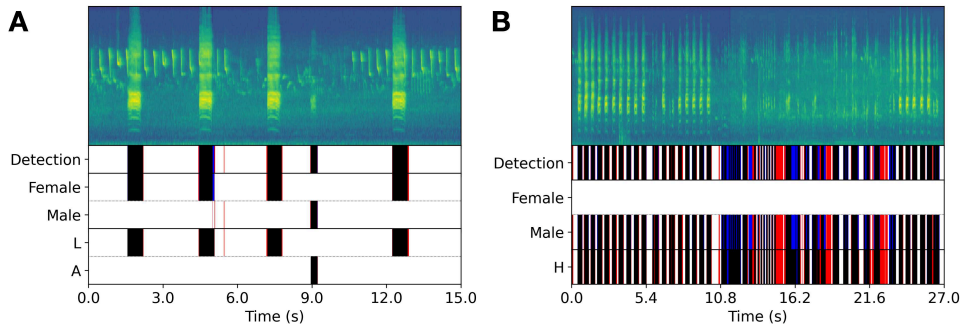


Figure 3.6 – Two example Mel-spectrograms from the test dataset (only one audio channel shown), with corresponding network outputs for detection, sexing, and identification, shown after the post-processing detailed in the text. A) Calls produced by one female (L) and one male (A). B) Song-like bout produced by one male (H). True negatives are in white, true positives in black, false negatives in blue, and false positives in red.

vocalisations even if they occur in quick succession.

### 3.4 Discussion

We proposed Rookognise, a deep learning system for the acoustic detection and vocalisation-independent identification of rooks in an outdoor soundscape. To our knowledge, the system is the first approach that can acoustically identify individuals regardless of the vocalisation produced. We approach the task by combining three different mechanisms: the multi-scale densely-connected MSD module, the multi-task approach where we divide the complex overall task into a cascade of increasingly complex tasks, and the explicit combination of outputs during training. All three mechanisms had substantial benefits to generalisation performance on unseen data. Multi-task learning has been proposed as acting as a regularisation mechanism (Liebel and Körner, 2018), although naive multi-task approaches may impede learning due to conflicting objectives. This may have caused performance loss on the detection and sexing sub-tasks for the  $MT_{None}$  configuration. In addition to the regularising effect of multi-task learning, the auxiliary tasks allow users to separately determine task-specific thresholds for more direct control over the precision and recall trade-off. Furthermore, conditioning the outputs of more complex tasks on the outputs of simple tasks may help the learning process, possibly by acting as an attention or a hierarchical classification (Nolasco and Stowell, 2022) mechanism: in this case the detection and presence tasks allow early rejection of, respectively, non-vocalisations or individuals absent from the audio clip.

The final system had strong overall performance according to all metrics used. Unfortunately, there are as yet few comparable approaches to acoustically identify individuals irrespective of vocalisation types (although there have been approaches to identify individuals by image data in multiple species, including birds). Stowell et al. (2019a) reported highly successful individual identification in three bird species, the little owl (*Athene noctua*), the chiffchaff (*Phylloscopus collybita*), and the tree pitpit (*Anthus trivialis*). They used a Random Forest classifier attributing an entire audio file to one individual, with high AUROC (85-94%, slightly lower than Rookognise). However, their metric

was computed as one decision per file and their dataset was by design restricted to one focal individual per file, so the comparison with our system is by necessity rather limited. Comparing different systems is generally an issue in deep learning, as many tasks lack a standardised benchmark with a well-defined dataset and associated metrics, to allow different approaches to be compared rigorously. For individual acoustic identification, this issue is even more prevalent due to the current absence of similar studies (Stowell, 2022). Nevertheless, Rookognise appears well suited to monitoring a social group in less controlled conditions, i.e. outdoors with multiple individuals in proximity.

The strong performance of the system was diminished in a few cases. From examining predictions on the test data, these errors consisted mostly of: 1) boundary errors, most often due to the network slightly overestimating the times spanned by vocalisations; 2) some false positives due to the acoustic properties of specific percussive noises, which share similar structure to some very short vocalisations produced by some of the females, and 3) identification errors, mostly confusions between specific individual pairs (e.g. K with H, O with M). In both cases, we could identify the source of the misattribution. In the K/H case, a long vocalisation bout produced by K was attributed to H. The system may have learned to preferentially attribute long song-like sequences to H, because he is the one with longest bouts. Other cases of identification errors may come from vocal similarity between the individuals in the study group. In the O/M case, for example, O and M shared their most common vocalisation, which accounted for a majority of O's vocalisation in the test data. Note that the system did not otherwise identify individuals proportionally to their dataset presence (i.e. individuals that vocalised less were not less correctly identified).

In many bird species, social proximity and vocal similarity are positively correlated (E. D. Brown, 1985; Grießmann and Naguib, 2002; Hausberger et al., 1995), but in our case the misidentified pairs did not correspond to a particularly close social bonds (Boucherie et al., 2016), while individuals involved in other bonds were not confused by the network (Table A.1). Individuals from similar genetic or cultural origins (same colonies) may also exhibit vocal similarities (Lemasson et al., 2011b). In our group, two individuals (E and J) were genetically related, but they were not substantially confused by the network (Table A.1). Even if some birds may share vocal similarities due to these factors (social, genetic or cultural proximity), the system still correctly classified identities.

At present, the Rookognise system necessitates knowing the individuals present in the study group to build the network and actively identify each individual. Unknown individuals cannot be identified by the system, but the system is designed to filter these unknown individuals out if desired: vocalisations from unknown individual rooks can be detected by the system (i.e. receive high scores from the detection sub-task) but without being attributed to an individual (i.e. receive low scores from the identification sub-task). As such, the Rookognise system can already be used as is to detect vocalisations from another group of rooks, which can greatly facilitate its deployment to other groups, i.e. only requiring to update the individual identification sub-task while other sub-tasks are already trained. More generally, we believe that the network can be deployed to monitor groups of any other species: birds, primates, cetaceans, provided that the number and identity of individuals in the group are known, and, of course, that the species produces vocalisations as part of its communication. More than a proof of concept, the system should be adaptable to any species no matter how harmonic their vocalisations

are. Indeed, the system is already successful on rooks, birds that are characterised by their harsh, inharmonic calls like many corvids (Fletcher, 2000). Neural networks often generalise well, although the risk of overfitting to a particular problem is always present. For instance, Stowell et al. (2019a) pointed out that a classifier trained to recognise one particular individual associated with a particular background noise, may fail to recognise the same individual in a different background noise. Our system was designed to limit this by explicitly training on background noise samples. In addition, the dataset itself, due to the complexity of the background noises recorded in an outdoors setting, likely limits the potential for overfitting on this type of data. Finally, none of the decisions involved in the design of the system were specifically tailored to the species or the group. The main limit to deploying the system to the study of another species is that said species must vocalise frequently enough to collect a sufficient amount of vocalisations in a reasonable time.

Extending the system to conditions where the number of individuals is not known *a priori* requires reformulating the identification task as an open-set recognition problem, a much more technically difficult task. Previous approaches to this problem have largely been in the form of speaker verification (Ntalampiras and Potamitis, 2021; Ptacek et al., 2016; Stowell et al., 2019a) or clustering approaches including deep learning-based ones (Goffinet et al., 2021; Kirschel et al., 2009; Pagliarini et al., 2021) to determine whether vocalisations are produced by a previously-known individual or by a potentially new one. However, neither these approaches, nor our system, can discriminate 1) between two unknown individuals, or 2) between two different vocalisations produced by the same unknown individual.

Finally, the multi-task design makes the system easily extendable to other tasks. Any behavioural information that can be classified can be incorporated as an additional sub-task; for instance, the acoustic type of the vocalisation (Bermant et al., 2019; Cohen et al., 2022; Oikarinen et al., 2019), its behavioural context, or its emotional valence (Briefer, 2012; Briefer et al., 2022; Marler and Evans, 1996. Briefer et al. (2022) recently proposed a system that could classify the emotional state of pigs by their vocalisations. The Rookognise system should be able to process these other informations as additional sub-tasks, which could have great implications, for instance for the assessment of welfare and management of captive animals (Laurijs et al., 2021). Finally, the system could be deployed in the wild to study the behaviours of any stable social group, so long as their identities are established.

### 3.5 Conclusions

In this paper, we proposed the Rookognise system as the first vocalisation-independent acoustic individual identification approach applicable to raw recordings, along with a large annotated dataset of rook vocalisation. The Rookognise system is designed to identify individual animals regardless of the vocal diversity exhibited at any level from the individual to the species. We leveraged several different mechanisms to help ensure good robustness for the problem at hand of automatic acoustic individual identification as well as great flexibility for the system to be deployed in other cases, including with additional or different tasks. The system leverages multi-task learning to improve generalisation performance on unseen data. This has two other consequences for users: first, the multiple outputs can

allow separate interpretation and facilitate the extraction of particular vocalisations if desired. Second, the system is designed to be easily extendable to other tasks by simply adding or replacing heads corresponding to the task at hand. Our implementation also uses multi-channel audio to capture additional information to improve performance, although single-channel audio still provides acceptable performance at much lower data storage costs. Future work will focus on extending the system toward the open-set recognition task, and on studying the internal representations learned by the network; for instance, to test whether the network identifies individuals based on vocalisation-independent features or if different signatures exist for different vocalisation types, as has been suggested in zebra finches finches (Elie and Theunissen, 2018).

### 3.6 Acknowledgements

This project was funded by a Michelin Foundation donation to V.D. (N°349-PE-3176). We are grateful to the Eurometropole of Strasbourg for granting us the use of the terrain for housing the animals. We also thank IRCAM for granting us access to the AudioSculpt software.



### 3.7 Annexe 1

Cette annexe correspond à l'annexe d'origine publiée avec l'article. Les figures sont reproduites telles qu'elles apparaissent dans l'article.

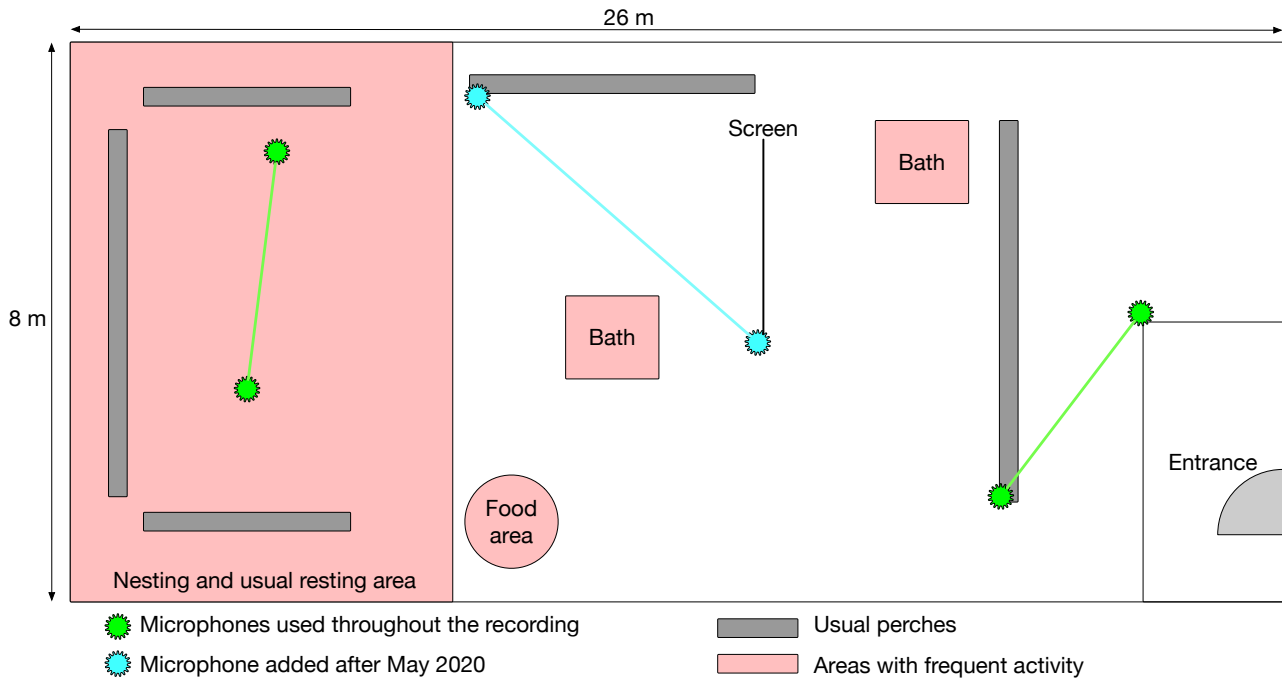


Figure S1.1 – Top-down schematic of the aviary, including microphone placement and areas most often used by the rooks. The aviary was entirely open with no separation between areas, except for the entrance area. Microphone locations are shown as the blue and green points. Observations were interrupted between mid-March and mid-May 2020 due to the COVID lockdown. An additional microphone was placed in the middle of the aviary after observations resumed.

Table S1.1 – Details of the birds used in this study. The rooks were identified by coloured leg rings. All were captured from wild colonies. Different labels indicate different colonies of origin. All but H were caught near Strasbourg (France), while H was caught in Southern France.

Individual	Sex	Mate	Introduction to group	Age at introduction	Colony of origin	Note
A	M	G	2013	Unknown (juvenile)	A	
B	M	P, K, S <sup>a</sup>	2006	2 months	B	
E	M	F, T	2006	2 months	B	Sibling with J
F	F	E, T	2016	2 months	C	Died during the recording period
G	F	A	2013	Unknown (adult)	A	
H	M		2019	2.5 years	Southern France	Hand-reared by an animal rescue association
J	F	M	2006	2 months	B	Sibling with E
K	M	B, P	2006	2 months	B	
L	F	O	2016	2 months	C	
M	M	J	2006	4 months	B	
O	M	L	2006	4 months	B	
P	F	B, K	2013	Unknown (adult)	A	
R	F		2017	2 months	D	Died during the recording period
S	F	B <sup>a</sup>	2013	Unknown (adult)	A	
T	M	E, F	2006	2 months	A	

<sup>a</sup> B and S formed a pair in 2021 during the recording period. B was previously in a triadic bond with K and P for several years, and S was a single bird after losing her mate in 2019



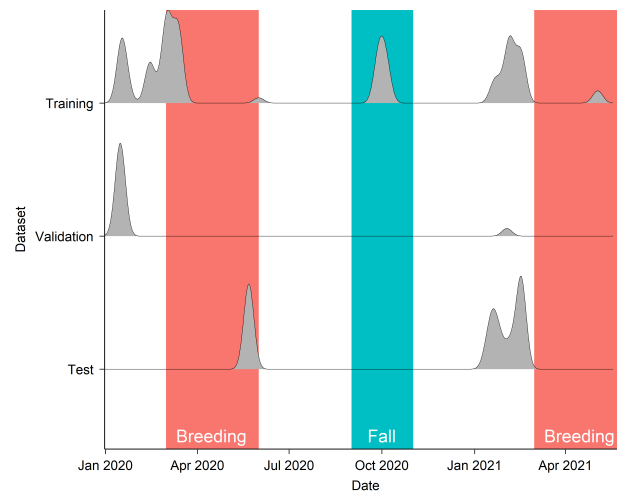


Figure S1.2 – Timeline of the data collection and distribution of the annotated data by period and dataset. Each line corresponds to the training, validation, or test dataset. The density lines correspond to the distribution of vocalisations over the recording period, normalized to the same maximum height. Shaded areas are periods of the year where the vocal behaviour of the rooks changed: during the breeding season (red), the males produced overall fewer vocal sequences while virtually all pairs were building nests; later in the season females stayed in the nest and produced loud begging calls. During fall (blue), the same behaviours occurred at a lower intensity and without nest building.

Table S1.2 – The non-standard architecture design elements included in Rookognise increase its performance. We report the effects of incorporating one of each of the non-standard architecture design elements described in Section 3.2.4. Each element is tested by comparing the full Rookognise system with all the non-standard elements included, a baseline with the same architecture but all the elements replaced by their standard counterparts, and a network with the same architecture, but with one non-standard element. The training and evaluation of all the networks reported here followed the same protocol as in the main text. The baseline elements were all based on the original *sparrow* network proposed by Grill and Schlüter (2017): LeakyReLU activation function with slope coefficient  $\alpha = 0.01$  (replaced by Mish), Batch Normalisation (replaced by Batch Renormalisation, Renorm), Max pooling (replaced by strided convolutional pooling, Conv), Binary cross-entropy loss (replaced by the Focal loss, with  $\gamma = 2$ ). Note that the binary cross-entropy can be obtained by setting  $\gamma = 0$  in the Focal loss formula), and Adam optimiser (replaced by Ranger21). Each respective network is denoted by Baseline + X, where X is the non-standard element included in that network. Values in bold are above the baseline, while underlined values are above the full Rookognise system. Note that the Ranger21 optimiser and the Focal loss yielded the largest improvements of the single elements. Nonetheless, the Rookognise generally performed the best in general, demonstrating that the non standard elements had non-redundant effects on performance.

	Detection		Sexing		Presence		Identification	
	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC	AUROC	AUPRC
Baseline	95.90	90.01	96.80	84.56	89.49	67.26	89.93	60.46
Baseline + Mish	95.43	<u>89.35</u>	96.55	83.78	<b>89.62</b>	66.54	<b>90.79</b>	<b>61.76</b>
Baseline + Renorm	95.77	89.81	96.79	83.83	88.72	<b>68.15</b>	<b>90.21</b>	59.69
Baseline + Conv	95.17	89.01	96.43	82.05	<b>90.29</b>	62.95	87.95	57.53
Baseline + Focal	<b>97.26</b>	<b>91.51</b>	<b>98.08</b>	<b>88.29</b>	86.08	61.85	<b>94.01</b>	<b>62.99</b>
Baseline + Ranger21	<b>96.44</b>	<b>90.30</b>	<b>97.30</b>	<b>84.83</b>	<b>90.06</b>	<b>70.30</b>	<b>93.05</b>	<b>64.73</b>
Rookognise	96.95	89.31	98.21	88.95	88.03	63.64	96.11	63.06

### 3.8 Annexe 2

Après la publication de l'article principal, le réseau a été ré-entraîné en incluant les données du groupe de Madingley. Les résultats obtenus avec ce nouveau réseau sont présentés rapidement ici pour comparaison avec l'entraînement sur le seul groupe de Strasbourg. Cependant, une légère différence dans l'architecture du réseau ici est que les quatre têtes du réseau sont connectées au module MSD, et non uniquement les têtes Présence et Identification. Ceci a été modifié après la publication de l'article, où l'utilisation du module MSD pour les quatre têtes améliore la performance d'identification des individus.

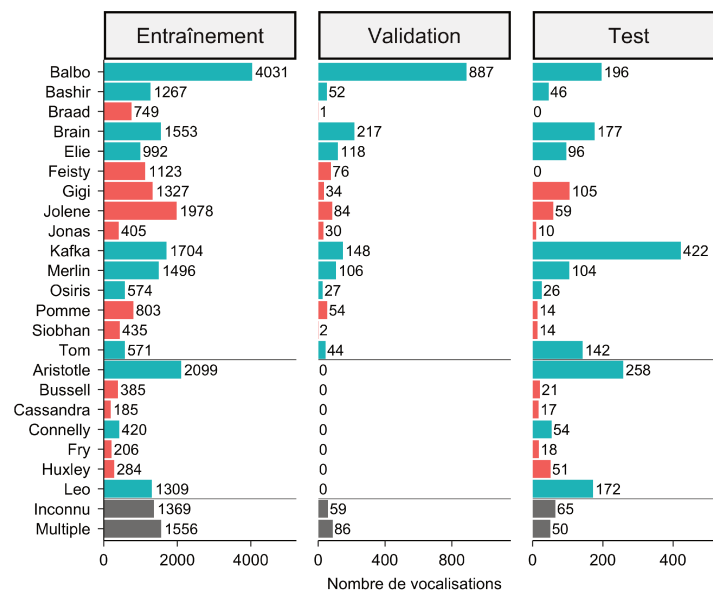


Figure S2.1 – Nombre de vocalisations des individus des deux groupes, par jeu de données. Les deux groupes sont séparés par les lignes noires horizontales. Les classes Inconnu et Multiples sont communes aux deux colonies, et sont donc à part. Les couleurs des barres correspondent au sexe de l'individu (bleu : mâle, rouge : femelle).

Table S2.1 – Matrice de confusion, calculée comme dans l'article principal mais à partir du jeu de données combinant les deux colonies captives. Mâles et femelles sont indiqués par les couleurs des cellules, et les cases en bas à droite représentent les statistiques globales (dans l'ordre horaire depuis la cellule supérieure gauche : nombre de classifications correctes, précision moyenne sur chaque individu, précision globale, rappel moyen pour chaque individu.

	Balbc	Bashi	Braac	Brain	Elie	Feistj	Gigi	Jolen	Jonas	Kafk:	Merli	Osin:	Pomr	Siobl	Tom	Arist	Buss	Cass	Comm	Fry	Huxl	Leo	Prédit	Précision
Balbo	3916	0	0	0	330	0	0	0	45	991	0	0	0	0	0	0	0	115	26	0	12	0	5435	72.05
Bashir	0	1039	0	0	2	0	55	0	0	51	0	0	0	0	8	0	0	0	0	0	26	0	1181	87.98
Braad	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	4	-
Brain	141	0	0	4046	192	0	0	0	0	103	7	87	0	0	568	0	0	0	0	0	0	0	5144	78.65
Elie	0	0	0	103	415	0	0	1	0	36	0	0	0	0	822	0	0	0	0	0	0	0	1377	30.14
Feisty	0	0	0	0	9	0	0	76	0	0	0	0	0	0	0	0	0	0	0	0	0	0	85	-
Gigi	0	42	0	0	94	0	2698	102	0	98	10	18	203	45	0	0	0	0	0	0	0	0	3310	81.51
Jolene	0	9	0	10	16	0	104	1711	125	80	0	0	10	13	19	8	0	0	0	0	0	0	2105	81.28
Jonas	0	0	0	0	0	0	0	0	121	0	1	0	0	0	0	0	74	0	0	0	0	0	196	61.73
Kafka	0	16	0	0	257	0	0	0	6	7757	0	0	0	0	249	0	0	0	0	0	0	0	8299	93.47
Merlin	0	0	0	0	0	0	0	0	0	36	2023	350	0	0	30	0	0	0	0	0	0	0	2439	82.94
Osiris	7	0	0	1	16	0	0	9	0	65	9	34	1	5	0	0	0	0	0	0	0	0	147	23.13
Pomme	0	0	0	0	0	0	6	0	0	0	0	0	93	0	0	0	0	0	0	0	0	0	99	93.94
Siobhan	0	0	0	0	0	0	105	0	0	0	0	0	0	246	0	0	0	0	0	0	0	0	351	70.09
Tom	323	0	0	73	104	0	0	0	20	6	0	0	0	0	1035	0	0	0	0	0	0	0	1561	66.30
Aristotle	0	0	0	0	27	0	0	0	0	0	0	0	0	0	0	9233	15	21	11	0	81	205	9593	96.25
Bussell	0	0	0	0	0	0	0	0	0	145	0	0	0	0	0	77	521	5	0	13	0	0	761	68.46
Cassandra	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	8	550	4	87	0	124	792	69.44
Connelly	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	5	0	2412	0	0	113	2538	95.04
Fry	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	545	0	0	545	100.00
Huxley	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	63	67	0	21	191	1052	18	1412	74.50
Leo	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	212	27	105	143	5	51	5836	6390	91.33
Annoté	4387	1106	0	4233	1462	0	2968	1913	317	9379	2050	489	307	309	2731	9620	717	796	2617	841	1226	6296	53764	87.15
Rappel	89.26	93.94	-	95.58	28.39	-	90.90	89.44	38.17	82.71	98.68	6.95	30.29	79.61	37.90	95.98	72.66	69.10	92.17	64.80	85.81	92.69	84.52	84.23

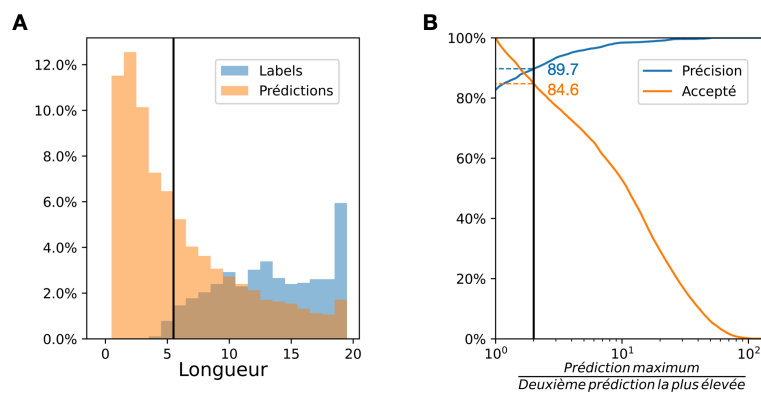


Figure S2.2 – Nettoyage des prédictions. Comme dans l'article principal: A) Distribution des prédictions (orange) et des labels (bleu) dans la tâche détection. La ligne noire représente un seuil où presque toutes les vocalisations sont détectées, mais une grande partie des faux positifs courts est éliminée. Les prédictions et les vocalisations de longueur supérieure à 20 pas de spectrogrammes sont omises pour ne pas encombrer le graphique. B) Précision dans la tâche d'identification et part des enregistrements acceptés en tant que vocalisation, en fonction de la confiance du réseau, estimée par le ratio entre la prédiction avec la valeur la plus élevée et la seconde plus élevée. La ligne noire verticale correspond à un ratio de 2.

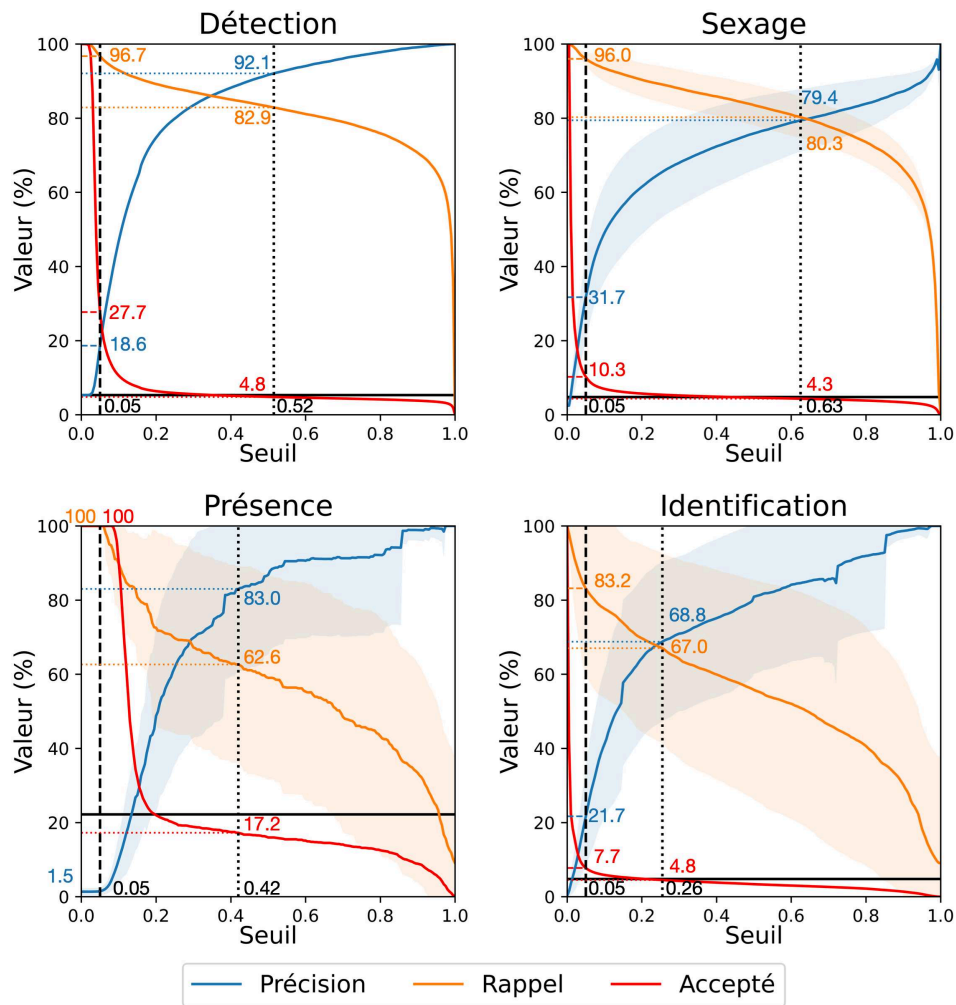


Figure S2.3 – Performance de Rookognise sur chaque sous-tâche. Comme dans l'article principal, la performance est mesurée avec la précision, le rappel et la proportion de prédictions acceptées, calculés sur le jeu de données de test. Les courbes pour les sous-tâches de sexage, présence, et identification correspondent à la moyenne des courbes de chaque classe, et l'aire colorée autour de chaque courbe correspond à la moyenne  $\pm$  écart-type. Les lignes verticales noires correspondent aux seuils, définis de la même manière que dans l'article : un seuil conservateur à 0.05, un autre seuil calculé de façon à maximiser le F1-score. Les lignes horizontales de chaque couleur correspondent à la valeur de chacune des trois courbes au seuil correspondant, indiqué à proximité.

# Chapitre 4

## Complexité vocale chez le corbeau freux : complexités et idiosyncrasies

Killian Martin, Francesca Cornero, Nicola S. Clayton, Olivier Adam, Nicolas Obin, Valérie Dufour

En préparation pour soumission à *Proceedings of the Royal Society B* sous le titre  
*Babel signalling in rooks*

### Résumé

La communication vocale est très répandue chez les animaux, avec divers niveaux de complexité dans le répertoire vocal. L'hypothèse de la complexité socio-communicative prédit que les espèces ayant des organisations sociales complexes (par exemple, les systèmes de fission-fusion, qui incluent des relations inter-individuelles variées) possèdent une plus grande complexité vocale. La complexité du répertoire est une combinaison d'un ou plusieurs des éléments suivants : diversité (la quantité de signaux différents dans le répertoire), gradation (si les vocalisations changent progressivement de structure et éventuellement de signification entre les différents types) et la combinabilité (l'arrangement des vocalisations en séquences, où l'ordre peut avoir une signification particulière). Parmi les oiseaux chanteurs, les corvidés sont parmi les espèces les plus complexes socialement. En particulier, les corbeaux freux vivent en système de fission-fusion, où certains individus vivent en groupe social toute l'année qui rejoignent d'autres groupes pour se reproduire, et certaines populations migrent alors que d'autres restent sédentaires. Les corvidés sont en outre bien connus pour leurs capacités cognitives développées, ce qui conforte l'hypothèse de l'intelligence sociale liant complexité sociale et capacité cognitive. Cependant, leurs capacités vocales restent peu étudiées. Les quelques études sur le répertoire vocal de ces espèces suggèrent des vocalisations très diverses tant intra- que inter-individuellement. Nous avons quantifié la complexité vocale dans deux groupes captifs de corbeaux freux, en mesurant la diversité et la gradation de leur répertoire, ainsi que la similarité inter-individuelle au niveau des vocalisations, et la complexité intra-individuelle au niveau du chant. Nous avons trouvé des différences entre les sexes, à savoir que les femelles produisaient des cris plus diversifiés et gradués, tandis que les mâles produisaient des vocalisations de chant plus diversifiées et graduées. Cependant, la similarité inter-individuelle était particulièrement surprenante : les corbeaux mâles produisaient des répertoires de cris presque complètement différents les uns des autres, même chez des partenaires sociaux proches, alors que les femelles produisaient des répertoires de cris pratiquement identiques, même entre deux groupes différents. La similarité des chants était intermédiaire en termes d'unités vocales, sans effet marqué du sexe ou de la proximité sociale. Enfin, chaque individu produisait des chants à partir de leur répertoire de vocalisations, de variabilité elle-même fonction de la diversité de leur répertoire. Les individus avec une plus forte diversité dans le répertoire pouvaient varier plus fortement l'organisation séquentielle de leurs chants.

## 4.1 Introduction

La vie dans un système social complexe impose des contraintes sur la communication entre individus : de l'échange d'une grande variété d'informations à la discrimination entre congénères (CHEN et WIENS, 2020 ; MARLER, 2004a). L'hypothèse de complexité sociale pour la complexité communicative prédit une relation entre la complexité sociale et la complexité communicative (FREEBERG et al., 2012 ; PECKRE et al., 2019). Une autre hypothèse, l'hypothèse de l'intelligence sociale (DUNBAR, 1998), postule également que chez de nombreuses espèces, des compétences cognitives avancées ont évolué en réponse aux exigences de leur socialité. En suivant ces deux hypothèses, nous pouvons prédire que les espèces sociales ayant des aptitudes cognitives avancées devraient également présenter une forte complexité vocale. La complexité vocale peut être abordée en examinant plusieurs caractéristiques du répertoire d'une espèce, comme sa diversité (par exemple, la quantité d'unités vocales dans le répertoire) et sa flexibilité (REBOUT et al., 2021). La flexibilité vocale peut être soit structurelle (par exemple avec une gradation, où des vocalisations de différents types existent à l'extrémité d'un continuum où les vocalisations changent progressivement de structure d'un type à l'autre, FISCHER et al., 2017 ; REBOUT et al., 2021 ; WADEWITZ et al., 2015), soit fonctionnelle (par exemple lorsqu'un individu peut produire une vocalisation acoustiquement identique dans plusieurs contextes différents, CLAY et al., 2015 ; MONTES-MEDINA et al., 2016). Le concept clé de la flexibilité est que les changements qu'elle implique doivent être réversibles, contrairement au concept de plasticité. La plasticité vocale se retrouve le plus souvent chez les juvéniles lorsqu'ils acquièrent leur répertoire vocal (MARLER, 1997 ; NELSON et MARLER, 1994), qui devient fixe après une phase de cristallisation (BEECHER et BRENOWITZ, 2005). La flexibilité vocale est un concept intrigant. D'une part, elle devrait fournir des avantages communicatifs, car elle permet potentiellement la transmission de signaux plus complexes, avec plus de nuances et d'expressivité (REBOUT et al., 2021). D'autre part, la vocalisation a pour principale fonction communicative de transmettre un message aux congénères. C'est en général le cas des cris du répertoire, qui servent des fonctions aussi diverses que l'alerte aux prédateurs, la localisation de la nourriture, les interactions sociales ou la coordination des activités sociales (CHEN et WIENS, 2020 ; MARLER, 2004a). Pour rester intelligibles aux auditeurs, les cris devraient donc rester un tant soit peu standardisés tant sur le plan acoustique que fonctionnel.

Les chants offrent un autre terrain potentiel de flexibilité vocale comparés aux cris chez les oiseaux adultes. Contrairement à d'autres vocalisations, des chants particuliers semblent fonctionnellement équivalents (CATCHPOLE et SLATER, 2008 ; NELSON et al., 2022), de sorte que différents chants sont fréquemment produits dans le même contexte. La grande majorité des espèces d'oiseaux (estimée à environ 80%, du moins parmi celles qui ont été étudiées, BEECHER et BRENOWITZ, 2005) acquièrent leur chant à l'âge juvénile et chantent ensuite les mêmes quelques chants stéréotypés tout au long de leur vie, sans varier l'organisation structurelle de chaque morceau de chant. À l'extrême, un individu peut ne produire qu'un seul chant stéréotypé (diamant mandarin *Taeniopygia guttata* : BOOGERT et al., 2011 ; TEMPLETON et al., 2014, bruant à couronne blanche *Zonotrichia leucophrys* : MARLER et TAMURA, 1962). Seules quelques espèces, dites à chant ouvert, sont capables de modifier leur chant au-delà de la phase d'acquisition juvénile (BEECHER et BRENOWITZ, 2005 ;

BRENOWITZ et al., 1997; MARLER, 1997). Certaines de ces espèces sont capables d'incorporer de nouvelles vocalisations par mimétisme, y compris de conspécifiques (étourneau sansonnet *Sturnus vulgaris*, EENS et al., 1992). Parmi ces espèces se trouvent des chanteurs extrêmement variés, avec plusieurs dizaines de chants à leur répertoire, comme les différentes espèces du genre *Troglodytidae* (KROODSMA et VERNER, 1978; VERNER, 1975), l'étourneau sansonnet (EENS et al., 1992), la grive solitaire (*Catharus guttatus*, DOOLITTLE, 2020), et l'exemple le plus extrême du moqueur roux (*Toxostoma rufum*) où plus de 1000 chants par individu ont été recensés (BOUGHEY et THOMPSON, 1981; KROODSMA et PARKER, 1977). Néanmoins, il n'est pas toujours clair si ces chants ouverts sont bien des exemples de flexibilité vocale. Malgré leurs compétences vocales, les étourneaux répètent souvent le même motif de base, ne modifiant vraiment leur chants que d'une saison de reproduction à l'autre (ADRET-HAUSBERGER et GÜTTINGER, 1984; BRENOWITZ et al., 1997), une capacité largement liée aux variations du niveau hormonal. La grande taille du répertoire de chants du moqueur et de la grive solitaire ne permet cependant pas de juger si une flexibilité vocale ou une réelle improvisation se produit. La flexibilité vocale pourrait plutôt correspondre à des variations répétées dans la production entre chants successifs.

Outre la diversité et la flexibilité du répertoire, d'autres variations vocales interindividuelles peuvent être des marqueurs de la complexité sociale. Un exemple de variation inter-individuelle correspond aux signatures vocales (BEECHER, 1989; BENTI et al., 2019; BRADBURY et VEHCAMP, 1998; ELIE et THEUNISSEN, 2018; KERSHENBAUM et al., 2013; LINHART et al., 2019; PRŮCHOVÁ et al., 2017; TERRY et al., 2005). Ces signatures vocales peuvent servir à identifier des individus (BEECHER, 1989; LINHART et al., 2019; oiseaux : BENTI et al., 2019; BERG et al., 2012; BERGER et LIGON, 1977; BOECKLE et al., 2012; CHARRIER et al., 2001; ELIE et THEUNISSEN, 2018; MATES et al., 2015; VIGNAL et al., 2008; YORZINSKI et al., 2006; primates : BOUCHET et al., 2013; CANDIOTTI et al., 2012; KEENAN et al., 2020; PRICE, 2015; SPILLMANN et al., 2017; mangouste : JANSEN et al., 2012; cétacés : KERSHENBAUM et al., 2013; KREMERS et al., 2012; QUICK et JANIK, 2012; phoque moine CHARRIER et al., 2017) ou signalent l'appartenance à des groupes sociaux (KREMERS et al., 2012; MONTEIRO et al., 2021). Dans la plupart des cas, les signatures vocales peuvent ajouter des variations individuelles aux cris mais ne modifient pas les paramètres acoustiques portant la fonction du signal. Dans les grands groupes sociaux, les individus peuvent converger vers des vocalisations similaires (KREMERS et al., 2012), formant parfois des dialectes partagés par des populations entières (MUNDINGER, 1982). Ces dialectes sont des vocalisations qui présentent des variations macro- ou micro-géographiques (CATCHPOLE et SLATER, 2008; MUNDINGER, 1982), et peuvent être causés par des adaptations locales à l'environnement acoustique. Par exemple, les perroquets juvéniles quittent le groupe social parental tôt dans leur vie pour rejoindre d'autres groupes d'adultes, et incorporent ensuite les cris de leur nouveau groupe dans leur répertoire (SALINAS-MELGOZA et WRIGHT, 2012). Les dialectes peuvent être un sous-produit de l'apprentissage vocal, car de nombreuses espèces acquièrent leurs vocalisations en apprenant de leurs partenaires sociaux (ANDREW, 1962; BITTERBAUM et BAPTISTA, 1979; CATCHPOLE et SLATER, 2008; SALINAS-MELGOZA et WRIGHT, 2012; WIENS, 1982), tels que les parents et les individus affiliés. Par conséquent, si les dialectes portent une information sur l'origine géographique des individus, ils peuvent également



fournir des informations sur l'appartenance présente à une colonie ou à un groupe (HAUSBERGER et al., 1995; MONTEIRO et al., 2021; SALINAS-MELGOZA et WRIGHT, 2012). Néanmoins, étant donné leur fonction communicative, les vocalisations des différents individus ou groupes ne devraient pas trop dévier dans leur structure acoustique de ce qui est typique dans le répertoire de l'espèce, sous peine d'entraver la transmission d'information aux récepteurs qui entendent la vocalisation.

Les oiseaux, avec leurs modèles d'agrégation variés (PAPAGEORGIOU et FARINE, 2021), sont des espèces idéales pour étudier le lien entre les systèmes sociaux et l'établissement de plusieurs couches de signatures vocales à différents niveaux sociaux. Parmi les oiseaux, la famille des corvidés comprend certaines des espèces les plus complexes socialement (CLAYTON et EMERY, 2007). En particulier, les corbeaux freux (*Corvus frugilegus*) forment des couples monogames stables, eux-mêmes associés en groupe sociaux stables tout au long de l'année, et forment régulièrement de très grandes colonies dans une dynamique de fission-fusion (AMICI et al., 2008; ARCHIE et al., 2006; BRADBURY et BALSBY, 2016; BRADBURY et al., 2001; BRAUN et BUGNYAR, 2012; CLAYTON et EMERY, 2007; CONNOR et al., 2000; EMERY, 2004; KELLEY et al., 2011; WILLIS et BRIGHAM, 2004), où de multiples groupes sociaux fusionnent de manière réversible en grandes colonies pour dormir ou se reproduire (COOMBS, 1960). En outre, différentes populations ont des habitudes migratoires différentes; par exemple, les corbeaux freux d'Europe occidentale quittent le site colonial après la reproduction et migrent sur de longues distances (BUSSE, 1969), alors que les corbeaux freux du sud du Royaume-Uni restent généralement à quelques kilomètres de leur site colonial (PATTERSON et al., 1971). En outre, les colonies de reproduction résidentes peuvent être rejointes par des oiseaux migrateurs (BURNS, 1957; BUSSE, 1965; MCKILLIGAN, 1980). Enfin, les couples pourraient revenir sur le même site de reproduction pendant plusieurs saisons de reproduction, mais cela n'a pas été formellement démontré (GOODWIN, 1955). L'organisation sociale des corbeaux freux est donc extrêmement complexe, car les individus sont impliqués dans de multiples couches de connexions sociales. Par conséquent, ils pourraient avoir régulièrement besoin de signaler non seulement leur propre identité, mais aussi à quel groupe social ou colonie ils appartiennent, dans un environnement social en constante évolution. Comme les autres corvidés, les corbeaux freux font en outre preuve d'excellentes capacités d'apprentissage, de remémoration, et de planification (BUGNYAR et al., 2016; EMERY, 2004; EMERY et al., 2007; GÜNTÜRKÜN et BUGNYAR, 2016; RABY et al., 2007). Par conséquent, nous pouvons nous attendre à ce que leur répertoire présente des niveaux très élevés de diversité ou de flexibilité pour permettre cette signalisation d'identité.

Les capacités vocales des corbeaux freux sont encore très peu étudiées, comparé à celles des autres oiseaux. Au niveau du groupe, les corbeaux freux possèdent un répertoire de cris de taille moyenne à grande, assez typique des corvidés (BAYENS, 1981; CHAMBERLAIN et CORNWELL, 1971; HA et al., 2003; HEINRICH, 1988; MUNTEANU et al., 2009; ROSKRAFT et ESPMARK, 1982), et une signature vocale individuelle a été trouvée dans le cri le plus commun des mâles (BENTI et al., 2019). Ils ont également démontré l'usage des vocalisations pour reconnaître d'autres individus (ROSKRAFT et ESPMARK, 1982). L'espèce possède également des capacités d'imitation, y compris de la voix humaine dans certains cas (Dufour, observations personnelles), des aptitudes qui ont été déjà mentionnées chez d'autres corvidés, mais aussi chez les perroquets (BRADBURY et BALSBY, 2016;

E. D. BROWN, 1985; KELLEY et al., 2011). Les corbeaux freux ont également la capacité de former des duos vocaux (KONDO et al., 2010; SEED et al., 2007). Ils semblent également posséder le même système neuronal de chant que les autres oscines (R. WANG et al., 2009), et produisent effectivement des "chants", comme quelques autres corvidés sociaux (E. D. BROWN, 1985; COOMBS, 1960; GOODWIN, 1986). Ces "chants" sont des séquences d'unités vocales organisées en phrases, qui peuvent durer de quelques secondes à plusieurs minutes. Ils ne sont pas produits lors de conflits territoriaux ou de parades, mais semblent plus similaires aux chants non-orientés qui ont été décrits chez plusieurs espèces d'oiseaux (DUNN et ZANN, 1996; KNEUTGEN, 1969; RITERS et al., 2019; SOSSINKA et BÖHNER, 1980). La structure des chants des corbeaux freux n'a cependant pas été décrite en détail dans la littérature. Cette espèce atypique vocalement et complexe socialement semble être un bon modèle pour évaluer la complexité vocale en relation avec la complexité cognitive et sociale.

Pour évaluer la diversité, la flexibilité et les signatures sociales dans la production des cris et des chants des corvidés, nous avons enregistré les vocalisations de deux groupes captifs, à Strasbourg (France) et à Madingley (Royaume-Uni). Des études précédentes ont montré que les paramètres acoustiques standards peuvent suffire à décrire la variabilité interindividuelle de types de vocalisations uniques (BENTI et al., 2019; BOECKLE et al., 2012; LAIOLO et al., 2000; STOWELL et al., 2016a; YORZINSKI et al., 2006). Cependant, nous avons choisi de construire chaque répertoire vocal individuel en faisant appel à une approche récente, qui a démontré une plus grande fidélité dans la préservation de la variation structurelle des vocalisations de plusieurs espèces (ELIE et THEUNISSEN, 2016; GOFFINET et al., 2021; SAINBURG et GENTNER, 2021; SAINBURG et al., 2020) par rapport aux études basées sur les paramètres acoustiques. Cette approche repose sur le calcul de représentations complètes de la vocalisation (i.e. le spectrogramme entier), la compression de ces représentations via des méthodes de réduction de la dimensionnalité (telles que t-SNE, UMAP, ou des auto-encodeurs variationnels), et enfin le regroupement de la projection à faible dimension pour déterminer des catégories de vocalisations, si elles existent (SAINBURG et GENTNER, 2021).

Nous avons prédit que les cris communs du répertoire ne devraient pas trop différer d'une colonie à l'autre, mais pourraient se regrouper en fonction du type de vocalisation et de la couche sociale : individu, couple, colonie. Cet effet pourrait être plus fort pour la population britannique, à l'origine non migrante. Alternativement, il se peut que chaque colonie ait acquis son propre répertoire de cris avec peu ou pas de chevauchement entre les répertoires. Nous avons également prédit que la flexibilité devrait se trouver à la fois dans les cris (avec un niveau élevé de vocalisations graduelles) et dans les chants (avec un grand nombre de modèles de chants).

## 4.2 Matériel et Méthodes

### 4.2.1 Sujets d'étude

Deux groupes de corbeaux freux captifs ont été enregistrés pour les besoins de cette étude sur deux sites différents, l'un à Strasbourg (France) et l'autre à Madingley (Royaume-Uni). Chaque groupe a été hébergé dans des volières extérieures depuis leur capture. Le groupe de Strasbourg était constitué

15 corbeaux pendant la période d'enregistrement (7 femelles et 8 mâles), et le groupe de Madingley de 7 corbeaux (4 femelles et 3 mâles). Tous les oiseaux étaient identifiés par des bagues de couleur attachées aux pattes.

#### 4.2.2 Protocole d'enregistrement

Les deux groupes ont été enregistrés et observés selon le même protocole. Toutes les vocalisations du groupe ont été enregistrées pendant plusieurs heures chaque jour à l'aide d'unités d'enregistrement autonomes (ARU) Song Meter 4 (Wildlife Acoustics, USA). Entre un et trois ARU ont été placées dans la volière et connectées à deux microphones sur des câbles de 3 mètres. Les microphones étaient espacés dans la volière afin de minimiser la distance entre le microphone et les oiseaux (jusqu'à un maximum d'environ 10 m). Les enregistrements ont ensuite été stockés sous forme de fichiers non compressés à 48 kHz avec une résolution de 16 bits. Le groupe de Strasbourg a été enregistré à partir de janvier 2020, et le groupe de Madingley a été enregistré entre février et mai 2022. Un observateur humain était présent pour certaines sessions (KM ou FC) afin de noter, à l'aide d'un programme Python, chaque vocalisation produite, ainsi que l'individu émetteur et les informations sur le contexte comportemental lorsque celui-ci pouvait être identifié.

#### 4.2.3 Extraction des vocalisations

Le premier ensemble de données a été collecté à partir de sessions d'enregistrement en présence de l'observateur. Les sessions de Strasbourg ont également été enregistrées sur vidéo pour faciliter le processus d'annotation. Les enregistrements audio, les notes et les vidéos ont été synchronisés afin que chaque vocalisation des sessions puisse être horodatée avec le début et la fin, et attribuée à l'individu émetteur. Ces annotations ont été effectuées par KM à l'aide des logiciels AudioSculpt (v3.4.6, BOGAARDS et al., 2004) et Audacity (v3.1.3). Les enregistrements vidéo ont été utilisés pour clarifier les ambiguïtés éventuelles sur l'identité des émetteurs. Cet ensemble de données est partiellement disponible en ligne (MARTIN et al., 2022). Une vocalisation a été définie comme un événement sonore qui se produisait toujours de la même manière. Cela incluait une grande majorité de sons ininterrompus séparés par des intervalles de silence, mais aussi d'autres vocalisations composées d'éléments extrêmement courts qui ne se produisaient jamais en dehors de ces vocalisations, et étaient donc considérées comme une seule vocalisation.

Cet ensemble de données a été utilisé pour entraîner un réseau de neurones profond (MARTIN et al., 2022, Chapitre 3). En bref, Rookognise reçoit des Mel-spectrogrammes bruts d'enregistrements audio et en extrait les vocalisations des corbeaux tout en les attribuant à des individus, ce qui automatise essentiellement le processus d'annotation. Rookognise a ensuite été appliqué à plusieurs sessions, y compris certaines où aucun humain n'était présent. Les annotations automatiques résultantes ont ensuite été validées manuellement par KM avant d'être incorporées aux données pour l'analyse.

Le jeu de données final comprend environ 48h de données (29,1h pour le groupe de Strasbourg et 18,8h pour le groupe de Madingley), dont 6h d'enregistrements non observés (3h pour chaque groupe), pour un total de 29 081 vocalisations, contenant des cris et des chants. Les chants ont été

définis comme des séries de plusieurs vocalisations produites en succession rapide, la limite entre deux chants étant fixée à 10 s sur la base de nos observations. Toutes les vocalisations non chantées ont été considérées comme des cris.

#### 4.2.4 Représentation des données sonores : mel-spectrogrammes

Les spectrogrammes de toutes les vocalisations ont été calculés par la Transformée de Fourier à Court Terme (STFT) avec une fenêtre de Hamming de 10 ms avec un chevauchement de 80%. L'amplitude de la STFT résultante a ensuite été mise à l'échelle Mel pour une dimension finale de 80 coefficients, pris au-dessus de 100 Hz pour éliminer les bruits basse fréquence.

Les vocalisations ont ensuite été débruitées par le biais d'un algorithme de seuillage spectral (SAINBURG et al., 2020). Le spectrogramme d'un échantillon de bruit à proximité de la vocalisation est utilisé en tant que seuil, en-dessous duquel le spectrogramme de la vocalisation est mis à zéro. Le seuil a été choisi au 95ème centile du spectrogramme de bruit de fond, calculé le long de l'axe temporel, et séparément pour chaque bande de fréquence et canal du spectrogramme.

Les spectrogrammes débruités ont ensuite été découpés pour éliminer les silences en début et fin de vocalisation, sur la base d'une enveloppe d'amplitude calculée à partir du spectrogramme débruité. Enfin, le canal présentant le rapport signal sur bruit le plus élevé a été sélectionné.

Enfin, les vocalisations n'ont été incluses dans l'analyse que si elles étaient produites par un individu identifié, sans chevauchement avec d'autres vocalisations, et si elles étaient d'assez bonne qualité (estimée manuellement pendant les annotations et vérifiée par le calcul du rapport signal-sur-bruit, qui devait être au-dessus d'une valeur minimale).

#### 4.2.5 Distances acoustiques entre paires de vocalisations : DFTW

Nous avons calculé la distance entre toutes les paires de spectrogrammes dans les données avec la distance de déformation dynamique spectro-temporelle (DFTW). La DFTW est une variante de la déformation temporelle dynamique (DTW) qui recherche l'alignement optimal entre deux spectrogrammes non seulement le long de l'axe temporel comme le fait le DTW, mais aussi le long de l'axe fréquentiel. La distance DTW est très efficace pour sa simplicité de mise en œuvre (BAGNALL et al., 2017; DANIEL MELIZA et al., 2013; R. . F. LACHLAN et al., 2010), mais ne prend pas en compte la distribution des fréquences dans un spectrogramme. Nous avons modifié l'algorithme DTW afin de trouver également la distorsion optimale le long de l'axe des fréquences du spectrogramme, de sorte que les vocalisations puissent être considérées comme similaires tant qu'elles peuvent être mises en correspondance par n'importe quel étirement en temps ou en fréquence. L'algorithme exact utilisé dans cette étude est détaillé dans les Méthodes générales.

#### 4.2.6 Réduction de dimensionnalité : UMAP

La matrice de distance DFTW correspond à une vue des données à très haute dimension. Les algorithmes de clustering souffrent généralement de la "malédiction de la dimensionnalité" (BELLMAN,

1957), selon laquelle les clusters deviennent de plus en plus difficiles à trouver à mesure que la dimensionnalité des données augmente. Pour contourner ce problème, nous avons utilisé l’algorithme UMAP (MCINNES et al., 2018) pour projeter la matrice de distance en 2D (nous avons imposé cette dimensionnalité afin de pouvoir visualiser la projection). Toutes les vocalisations ont été projetées dans le même espace 2D afin que les distances relatives entre les vocalisations soient comparables entre les différents individus. Les paramètres utilisés pour la projection étaient les suivants (en utilisant les notations du package, voir section 4.2.11) :  $n\_neighbors = 30$ ,  $min\_dist = 0$ ,  $n\_components = 2$ , et nous avons fixé une graine aléatoire pour la reproductibilité. Le paramètre  $n\_neighbors$  a été déterminé en comparant les mesures de fiabilité et de continuité (VENNA et KASKI, 2005), pour s’assurer que les points voisins dans la distance DFTW originale étaient également voisins dans l’espace 2D projeté. Les autres paramètres ont été choisis pour la visualisation et le clustering ( $n\_components$  étant le nombre de dimensions de la projection, et  $min\_dist$  permet aux points correspondants à chaque vocalisation de se superposer dans la projection s’ils sont suffisamment proches).

#### 4.2.7 Clustering : HDBSCAN

L’algorithme de clustering HDBSCAN (CAMPELLO et al., 2015; MCINNES et al., 2017) a ensuite été appliqué à la projection. HDBSCAN est un algorithme de clustering hiérarchique basé sur la densité, ce qui permet à l’algorithme de trouver des clusters indépendamment de leur forme (alors que des algorithmes plus classiques comme le k-means suppose que les clusters ont une forme globulaire comme celle résultant d’une distribution normale multivariée). HDBSCAN extrait automatiquement les clusters les plus stables, évitant de chercher le meilleur nombre de clusters en évaluant un grand nombre de solutions possibles. En outre, HDBSCAN incorpore la possibilité de faire un clustering diffus, où les vocalisations sont affectées à chaque cluster de la solution avec des probabilités variables, ce qui permet de l’intégrer naturellement à des mesures de la gradation vocale. Enfin, HDBSCAN tient également compte des vocalisations atypiques (c’est-à-dire des vocalisations trop atypiques pour former des clusters) en leur permettant de ne pas être regroupées si elles ne sont pas proches de suffisamment de vocalisations similaires pour former un cluster (nous n’avons cependant pas utilisé cette propriété dans cette étude, car l’utilisation du clustering diffus ne permettait pas de considérer certaines vocalisations comme extérieures au clustering). Nous avons imposé une taille minimale aux clusters de 30 vocalisations (dans l’implémentation du package :  $min\_cluster\_size = 30$  et  $min\_samples = 30$ ) afin de s’assurer que les clusters trouvés ne soient pas dus au hasard. Les résultats de HDBSCAN ont été évalués par l’indice de validité (MOULAVI et al., 2014).

#### 4.2.8 Estimation de la diversité de la gradation : indices de Hill

Nous avons estimé la gradation et la diversité du répertoire en utilisant des mesures de la théorie de l’information : les indices de Hill (CHAO et al., 2014; HILL, 1973; ZSEBŐK et al., 2021). La famille des indices de Hill est définie par la formule suivante :

$${}^qD = \left( \sum_{i=1}^C p_i^q \right)^{\frac{1}{1-q}}$$

où  $p_i$  est la proportion de la classe  $i$  dans l'échantillon,  $C$  est le nombre de classes dans l'échantillon, et  $q$  est un entier positif (généralement 0, 1 ou 2) qui définit l'indice spécifique utilisé. Puisque nous avons utilisé les scores d'appartenance de HDBSCAN,  $p_i$  était le score d'appartenance pour le cluster  $i$ , et  $C$  était le nombre de clusters dans la solution de clustering. La gradation a été estimée pour chaque vocalisation en calculant les valeurs respectives de  ${}^qD$  puis en les comparant entre individus, sexes et sites. La diversité a été estimée par oiseau individuel, en faisant d'abord la moyenne des probabilités d'appartenance à chaque cluster par individu, puis en calculant  ${}^qD$  pour chaque individu. Nous avons utilisé les indices définis par  $q = 1$  et  $q = 2$ , qui correspondent à la diversité de Shannon et à la diversité de Simpson respectivement (CHAO et al., 2014).

#### 4.2.9 Estimation de la similarité des répertoires entre individus : indice de Morisita-Horn

Nous avons estimé le degré de similarité du répertoire entre tous les individus des groupes d'étude, en nous basant également sur les résultats du clustering diffus. La similarité du répertoire peut être estimée par le degré auquel les individus partagent les mêmes vocalisations. Nous avons utilisé une variante de l'indice de Morisita-Horn (CHAO et al., 2006), adaptée au clustering diffus (voir Méthodes générales) :

$$MH_{fuzzy} = 2 \frac{\sum_{i=1}^C x_i \cdot y_i}{\left( \frac{\sum_{i=1}^C x_i^2}{X^2} + \frac{\sum_{i=1}^C y_i^2}{Y^2} \right) XY}$$

où  $x_i$  correspond à la somme des probabilités d'appartenir au cluster  $i$  sur toutes les vocalisations de l'individu  $x$  (et  $y_j$  suit une définition analogue), et  $X$  et  $Y$  sont le nombre de vocalisations des individus  $x$  et  $y$  respectivement. Cet indice est compris entre 0 et 1, 1 représentant un chevauchement parfait (i.e. les mêmes classes dans les mêmes proportions), et 0 représentant une absence totale de chevauchement.

#### 4.2.10 Similarité dans les chants

Pour estimer la similarité entre les chants de chaque individu, les points de la projection UMAP correspondant aux dix chants les plus longs des chanteurs les plus prolifiques ont été extraits. Ces chants étaient uniquement inclus si toutes les vocalisations qui les composaient étaient de bonne qualité (i.e. conservées dans le jeu de données final). Les transitions entre vocalisations successives ont ensuite été représentées pour comparaison visuelle.

#### 4.2.11 Implémentation et analyses statistiques

Le traitement audio, les calculs DFTW et les calculs des indices de Morisita-Horn ont été effectués en Python (v3.8, VAN ROSSUM et DRAKE, 2009). Le code de débruitage était basé sur SAINBURG et al. (2020). L'UMAP a été réalisé à l'aide du package umap-learn (v0.5.3, MCINNES et al., 2018), et le HDBSCAN a été réalisé à l'aide du package hdbscan (v0.8.28, MCINNES et al., 2017). L'ensemble du code Python utilisé pour ces analyses est disponible en ligne.

Les analyses statistiques de la distribution de la gradation, de la diversité et de la similarité ont été réalisées dans R (v4.1.1, R CORE TEAM, 2021) à l'aide de modèles linéaires généralisés (GLM) ou de modèles linéaires mixtes généralisés (GLMM) ajustés avec le package *lm4* (v1.1-30, BATES et al., 2015). Chacun de ces modèles utilisait un des trois indices précédents (diversité de Shannon, diversité de Simpson, indice de Morisita-Horn) comme variable réponse et le sexe, le type de vocalisation et le site comme variables indépendantes. Nous avons également inclus les interactions de second ordre entre les trois variables indépendantes. Les modèles impliquant les indices de Hill ont été entraînés en utilisant des GLMM Gamma et une fonction de lien log, avec l'identité de l'individu en facteur aléatoire. Pour le modèle de similarité, pour chaque variable indépendante, nous avons considéré la similarité au sein des valeurs prises par cette variable ainsi qu'entre ces valeurs (par exemple, pour le sexe, nous avons comparé la similarité entre mâles, entre femelles, et entre mâles et femelles). Ce modèle a été ajusté en utilisant un GLM binomial avec une fonction de lien logit.

Pour chaque modèle, nous avons ensuite ajusté tous les sous-modèles possibles en utilisant la fonction *dredge* du package *MuMIn* (v1.43.17, BARTON, 2020). Ces modèles ont été classés selon le critère AICc : le meilleur modèle est celui dont l'AICc est le plus faible. Les modèles situés à moins de 2 unités AICc au-dessus du meilleur modèle peuvent être considérés comme équivalents (ANDERSON, 2004), tandis que les modèles situés à plus de 2 unités AICc sont considérés comme significativement moins bons pour prédire la variable dépendante. Lorsque plusieurs modèles étaient équivalents selon cette définition, nous avons considéré le plus parcimonieux comme le meilleur modèle. Nous avons ensuite utilisé ce modèle pour des tests par paires afin de déterminer quelles différences étaient effectivement statistiquement significatives. Ces tests par paires ont été effectués à l'aide du package *emmeans* (v1.8.0, LENTH, 2022), les p-valeurs étant corrigées pour les tests multiples par la méthode de Tukey.

## 4.3 Résultats

### 4.3.1 Distribution des données

Un total de 23 514 vocalisations provenant des deux groupes d'étude a été inclus dans le jeu de données (Table 4.1). Comme tous les individus n'ont pas pu être enregistrés pendant la même durée, les comparaisons étaient basées sur la fréquence des vocalisations (vocalisations produites par heure enregistrée) de chaque oiseau. En regroupant les vocalisations par cris et par chants, les femelles ont vocalisé moins fréquemment que les mâles sur les deux sites ; le groupe de Madingley a également moins vocalisé en général que le groupe de Strasbourg. Une conséquence de ceci est que certaines femelles ont été échantillonnées beaucoup moins que le reste du groupe (en particulier, une femelle du groupe de Madingley a été enregistrée moins de 200 fois). Les vocalisations étaient réparties approximativement à parts égales entre cris (46,9%) et chants (53,1%). La majorité des individus des deux groupes a produit des chants pendant l'enregistrement (9 mâles sur 12, 6 femelles sur 10). Parmi ces individus, les mâles ont chanté plus fréquemment que les femelles, produisant en moyenne des chants plus nombreux et plus longs. La distribution des cris était plus hétérogène, puisque les femelles du site de Strasbourg ont produit des cris plus fréquemment que les mâles, alors que l'inverse était

TABLE 4.1 – Distribution des vocalisations dans les données, par site, sexe, et type de vocalisation, par la fréquence de vocalisation (par heure, moyenne  $\pm$  écart-type par individu) et le nombre d'individus inclus dans chaque catégorie.

Site	Sexe	Type de vocalisation	Vocalisations par heure (moyenne $\pm$ écart-type)	Nombre d'individus
Strasbourg	Male	Cri	18.1 $\pm$ 13.6	8
		Chant	40.8 $\pm$ 40.5	7
	Female	Cri	27.6 $\pm$ 22.1	7
		Chant	17.2 $\pm$ 13.8	4
Madingley	Male	Cri	24.6 $\pm$ 5.7	3
		Chant	60.8 $\pm$ 38.1	2
	Female	Cri	9.3 $\pm$ 1.9	4
		Chant	7.6 $\pm$ 0.9	2

vrai pour les individus du site de Madingley. Un contexte comportemental n'a pu être déterminé que pour 33% de toutes les vocalisations. Parmi les vocalisations qui ont pu être contextualisées, 48% ont été produites par des femelles, et 52% par les mâles. Parmi les vocalisations des femelles : 27% des vocalisations étaient associées à un comportement de *bow and tail fanning* (où l'individu s'incline vers l'avant, redresse sa queue et la déploie en éventail), 26% à la construction du nid, 20% à la réception de nourriture, 11% à des cris de quémandage, et 10% à des cris au nid, et les cris restants (environ 6%) étaient répartis entre des contextes agonistiques, affiliatifs, d'exploration, et de recherche de nourriture au sol. Parmi les vocalisations des mâles : 58% étaient associées à un *bow and tail fanning*, 21% à l'exploration d'objets, 12% à la recherche de nourriture, 3% à des comportements de parade, 3% à des comportement de nourrissage, et les vocalisations restantes étaient répartis entre des montes, des vocalisations agonistiques, des vocalisations associées à la construction des nids, et des cris de quémandage. Les chants représentent 53% des vocalisations totales (non comptés dans les vocalisations contextualisées précédemment), et ont été produits à 85% par les mâles. Parmi ces vocalisations de chant, 15% ont été émises dans un contexte reconnaissable : 9% étaient associées à un *bow and tail fanning*, 7% à l'exploration d'objets, 1% à un comportement de parade, et moins de 0,1% aux montes, à la recherche de nourriture et au nourrissage d'un autre individu.

#### 4.3.2 Projection et clustering

Toutes les vocalisations ont été projetées dans un espace 2D avec UMAP (Fig. 4.1) avant d'appliquer HDBSCAN. La projection était associée à une fiabilité (0,91) et une continuité (0,96) très élevées, indiquant une excellente préservation des voisinages entre les points. La projection montre un certain nombre de clusters bien définis en périphérie, globalement bien séparés par individu et par type de vocalisation (cri ou chant). Cependant, le centre de la projection est occupé par un ensemble de vocalisations plus diffus, sans cluster clairement défini, bien que certaines régions semblent plus denses que d'autres.

HDBSCAN a trouvé 68 clusters, représentant 71,5% des vocalisations, et les 28,5% restants étant



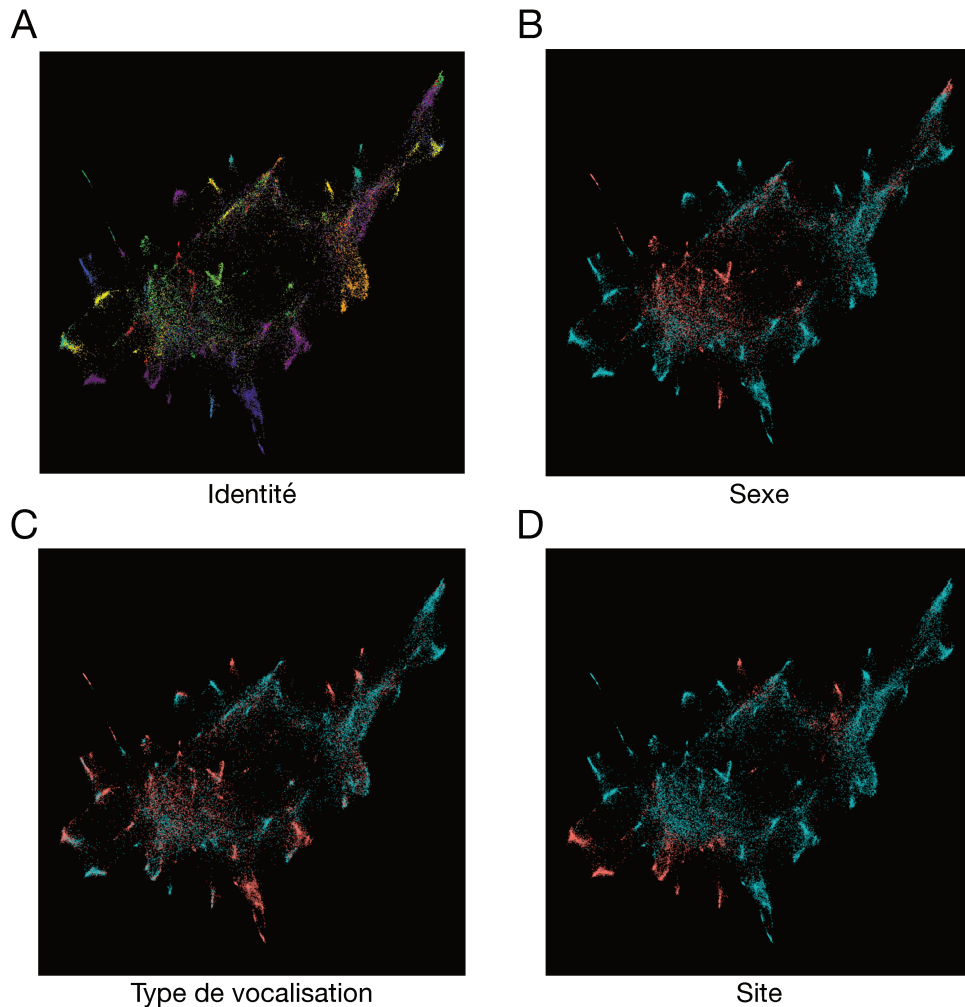


FIGURE 4.1 – Projection UMAP des vocalisations, colorée par A) identité individuelle de l'émetteur, B) sexe (femelles en rouge, mâles en bleu), C) type de vocalisation (cris en rouge, chants en bleu), et D) site (Madingley en rouge, Strasbourg en bleu)

trop épars pour être regroupés selon les paramètres choisis. Cependant, l'indice de validité associé à cette solution de clustering était très faible (0,09). Pour cette raison, et parce que la projection UMAP semble suggérer une forte gradation dans les vocalisations, nous avons choisi d'utiliser uniquement les résultats du clustering diffus pour le reste de l'analyse.

### 4.3.3 Gradation

La gradation du répertoire a été estimée à partir des deuxième et troisième indices de Hill calculés sur les valeurs d'appartenance provenant du regroupement (Fig. 4.2). Les deux indices ont donné des résultats sensiblement identiques ; par conséquent, seule la diversité de Shannon  $^1D$  sera discutée ici, (cf Annexe 1 pour les résultats avec la diversité de Simpson  $^2D$ ). Le meilleur modèle n'incluait que le sexe et le type de vocalisation, ainsi que leur interaction. Le site et ses interactions n'ont pas augmenté le pouvoir prédictif du modèle (Table 4.2). Le modèle nul était également significativement moins bon pour prédire la gradation (125.7 AICc au-dessus du meilleur modèle).

Les tests par paires ont révélé que les femelles produisaient des cris plus gradés que leurs chants

TABLE 4.2 – Sélection de modèle pour la gradation. Les trois premières colonnes correspondent aux trois variables incluses dans les modèles (sexe, type de vocalisation, site), et les trois colonnes suivantes correspondent à leurs interactions respectives. AICc : critère AICc, delta : nombre d'AICc au-dessus du meilleur modèle, poids : poids du modèle dans l'ensemble des modèles calculés. Seuls les modèles statistiquement indistinguables du meilleur modèle ( $\text{delta} < 2$ ), ainsi que les modèles nul et complet, sont inclus.

Sexe	Type	Site	Sexe :Type	Type :Site	Sexe :Site	AICc	delta	poids
+	+		+			197127.4	0	0.45
+	+	+	+			197129	1.6	0.2
+	+	+	+	+		197129	1.62	0.2
+	+	+	+	+	+	197130.8	3.52	0.08
						197253.1	125.69	0

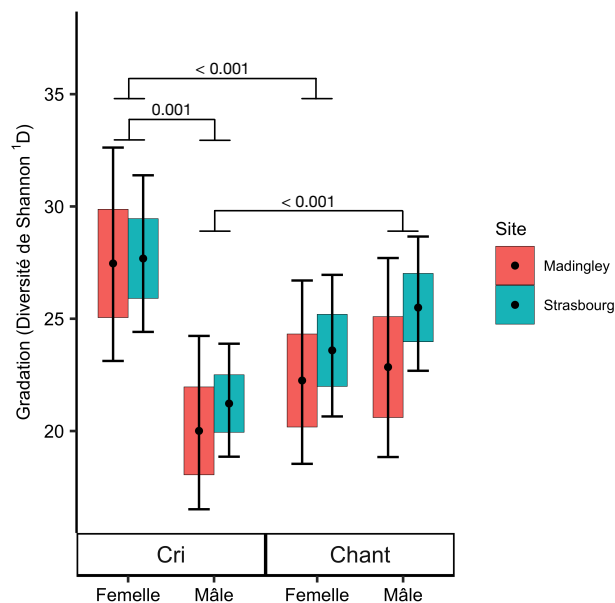


FIGURE 4.2 – Gradation dans le répertoire vocal par sexe, type de vocalisation, et site. Le site n'avait aucun effet statistique et n'était pas inclus dans les tests statistiques, mais est inclus dans cette visualisation. Pour chaque boxplot, le point correspond au paramètre estimé par le meilleur modèle, la boîte correspond à la moyenne  $\pm$  erreur standard, et les segments correspondent à l'intervalle de confiance à 95%. Les tests significatifs sont représentés par les lignes horizontales, avec la p-valeur indiquée au-dessus de la ligne.

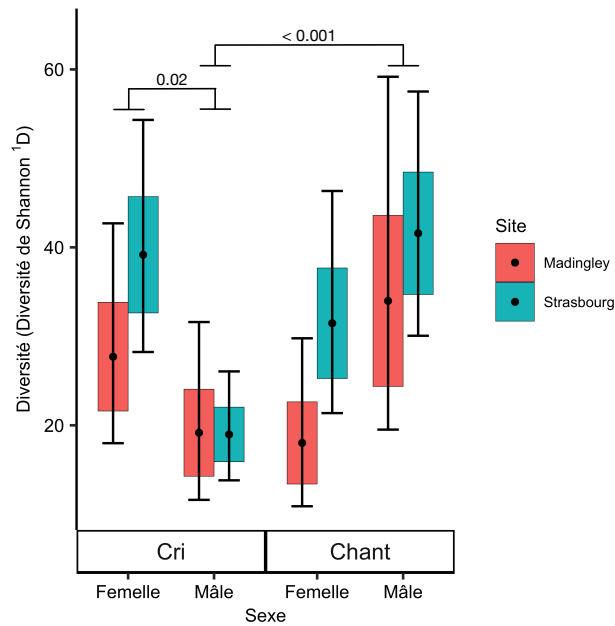


FIGURE 4.3 – Diversité dans le répertoire vocal par sexe, type de vocalisation, et site. Le site n’avait aucun effet statistique et n’était pas inclus dans les tests statistiques, mais est inclus dans cette visualisation. Pour chaque boxplot, le point correspond au paramètre estimé par le meilleur modèle, la boîte correspond à la moyenne  $\pm$  erreur standard, et les segments correspondent à l’intervalle de confiance à 95%. Les tests significatifs sont représentés par les lignes horizontales, avec la p-valeur indiquée au-dessus de la ligne.

(ratio cris/chants  $1,18 \pm 0,04$ ,  $z = 5,61$ ,  $p < 0,001$ ). Les mâles, au contraire, produisaient des chants plus gradés que leurs cris (ratio cris/chants  $0,84 \pm 0,01$ ,  $z = -9,99$ ,  $p < 0,001$ ). Enfin, pour les cris, les femelles étaient plus gradées que les mâles (ratio femelles/mâles  $1,32 \pm 0,10$ ,  $z = 3,74$ ,  $p = 0,001$ ), alors qu’il n’y avait pas de différence significative entre les sexes pour les chants (ratio femelles/mâles  $0,94 \pm 0,07$ ,  $z = -0,84$ ,  $p = 0,84$ ).

#### 4.3.4 Diversité

TABLE 4.3 – Sélection de modèle pour la diversité. Les trois premières colonnes correspondent aux trois variables incluses dans les modèles (sexe, type de vocalisation, site), et les trois colonnes suivantes correspondent à leurs interactions respectives. AICc : critère AICc, delta : nombre d’AICc au-dessus du meilleur modèle, poids : poids du modèle dans l’ensemble des modèles calculés. Seuls les modèles statistiquement indistinguables du meilleur modèle ( $\text{delta} < 2$ ), ainsi que les modèles nul et complet, sont inclus.

Sexe	Type	Site	Sexe :Type	Type :Site	Sexe :Site	AICc	delta	poids
+	+		+			288.9	0	0.57
+	+	+	+			290.5	1.61	0.25
+	+	+	+	+		295.5	6.62	0.02
					+	305.4	16.57	0

Comme pour la gradation, la diversité du répertoire de chaque individu a été estimée avec les deuxième et troisième indices de Hill (Fig. 4.3). Encore une fois, les deux indices ont donné des résultats sensiblement similaires et seule la diversité de Shannon  $^1D$  sera discutée ici (cf Annexe 1 pour la même analyse avec la diversité de Simpson  $^2D$ ). Comme pour la gradation, le meilleur modèle ne comprenait que le sexe, le type de vocalisation et leur interaction, et l'ajout du site comme variable ou dans les interactions n'a pas amélioré le pouvoir prédictif du modèle (Table 4.3). Le modèle nul était également significativement moins bon pour prédire la diversité du répertoire (16,57 AICc au-dessus du meilleur modèle).

Les tests par paire ont révélé que les femelles produisaient des cris plus divers que les mâles (ratio femelles/mâles :  $1,80 \pm 0,37$ ,  $z = 2,87$ ,  $p = 0,02$ ), et que les mâles produisaient des chants plus divers que leurs cris (ratio cris/chants :  $0,48 \pm 0,05$ ,  $z = -6,01$ ,  $p < 0,001$ ). Aucune différence dans la diversité des chants n'a pu être détectée entre les mâles et les femelles (ratio femelles/mâles :  $0,65 \pm 0,15$ ,  $z = -1,89$ ,  $p = 0,24$ ), ni dans la diversité entre les cris et les chants chez les femelles (ratio cris/chants :  $1,34 \pm 0,19$ ,  $z = 1,99$ ,  $p = 0,19$ ).

#### 4.3.5 Similarité

TABLE 4.4 – Sélection de modèle pour l'indice de similarité. Les trois premières colonnes correspondent aux trois variables incluses dans les modèles (sexe, type de vocalisation, site), et les trois colonnes suivantes correspondent à leurs interactions respectives. AICc : critère AICc, delta : nombre d'AICc au-dessus du meilleur modèle, poids : poids du modèle dans l'ensemble des modèles calculés. Seuls les modèles statistiquement indistinguishable du meilleur modèle ( $\text{delta} < 2$ ), ainsi que les modèles nul et complet, sont inclus.

Sexe	Type	Site	Sexe :Type	Type :Site	Sexe :Site	AICc	delta	poids
+	+	+	+			601.7	0	0.8
+	+	+	+	+	+	610.2	8.59	0.01
						658.9	57.3	0

Nous avons évalué le degré de similarité entre chaque paire d'individus à l'aide de l'indice de Morisita-Horn adapté au clustering diffus (Fig. 4.4). Les valeurs obtenues ont alors été comparées en fonction du sexe, du type de vocalisation et du site, en séparant dans les trois cas les valeurs de la variable correspondante (par exemple, pour l'effet sexe : mâles avec mâles, femelles avec femelles et femelles avec mâles). Le meilleur modèle indique des effets du sexe, du type de vocalisation et de leur interaction, mais aussi du site indépendamment des autres effets (Table 4.4). Les tests par paires sur l'effet du site ont révélé qu'en général, les oiseaux de Strasbourg sont plus similaires entre eux qu'aux oiseaux de Madingley (ratio entre sites / Strasbourg :  $0,64 \pm 0,12$ ,  $z = -2,37$ ,  $p = 0,047$ ), sans autre effet significatif cependant (ratio entre sites / Madingley  $1,05 \pm 0,37$ ,  $z = 0,13$ ,  $p = 0,99$ ; ratio Madingley / Strasbourg :  $0,61 \pm 0,22$ ,  $z = -1,39$ ,  $p = 0,34$ ).

Les effets du sexe et du type de vocalisation ont révélé deux dynamiques différentes dans la similarité inter-individuelle. Concernant les cris, les femelles sont très similaires entre elles, significativement plus que les mâles ne le sont entre eux (ratio mâle / femelle :  $6,92 \pm 3,61$ ,  $z = 3,70$ ,  $p = 0,007$ ).

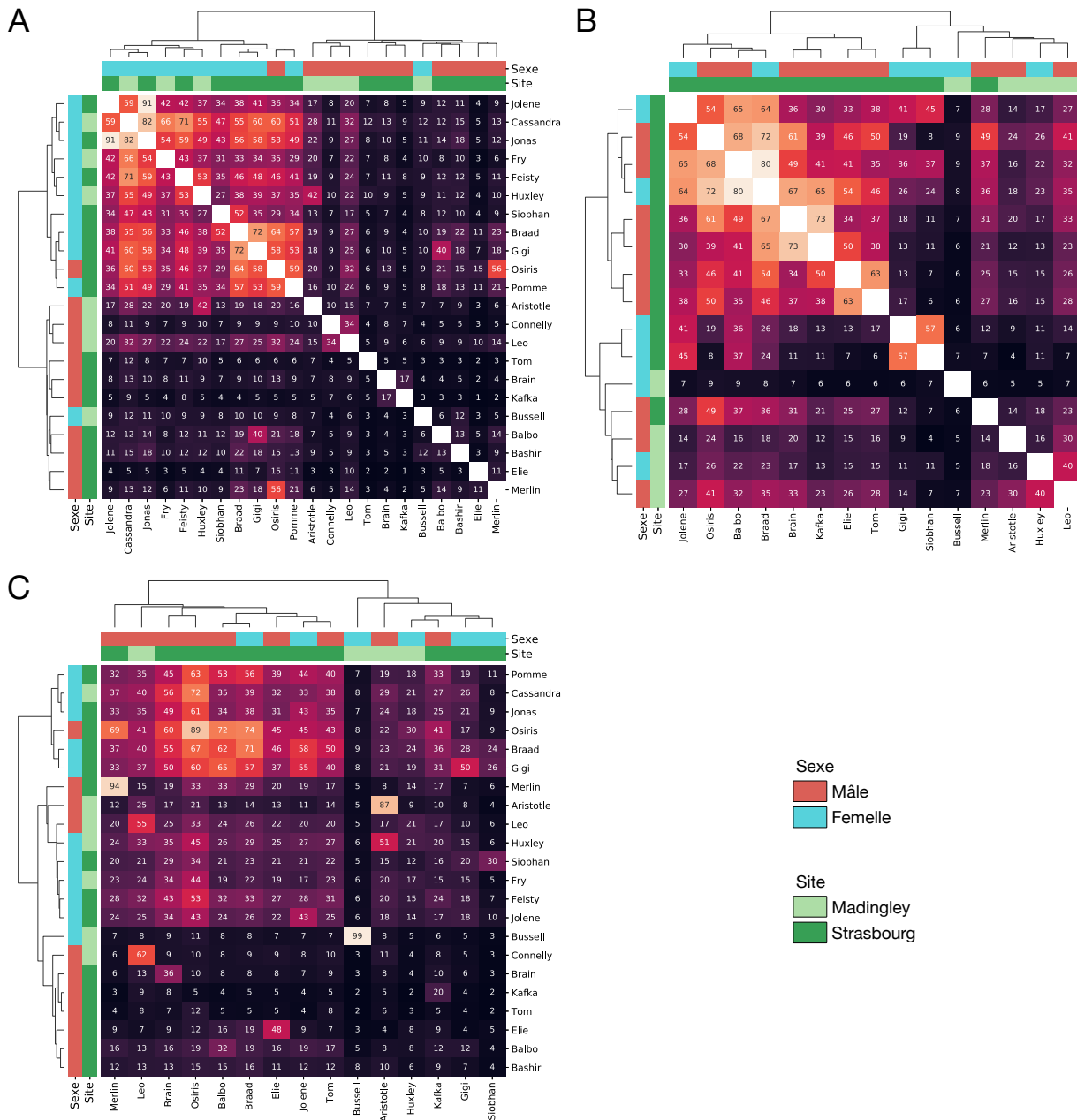


FIGURE 4.4 – Similarité entre les répertoires de chaque paire d'individus. A) Dans les cris. B) Dans les chants C) Entre cris (lignes) et chants (colonnes). Chaque ligne et colonne correspond à un individu. Le sexe et le site de chacun sont indiqués par les cellules colorées autour de chaque matrice. Chaque cellule des matrices contient la valeur de l'indice de Morisita-Horn correspondant, en pourcentage, avec le code couleur correspondant. Chaque matrice a été réarrangée avec la méthode de clustering hiérarchique de Ward pour regrouper les individus avec la même distribution dans leurs similarités vis-à-vis des autres individus.

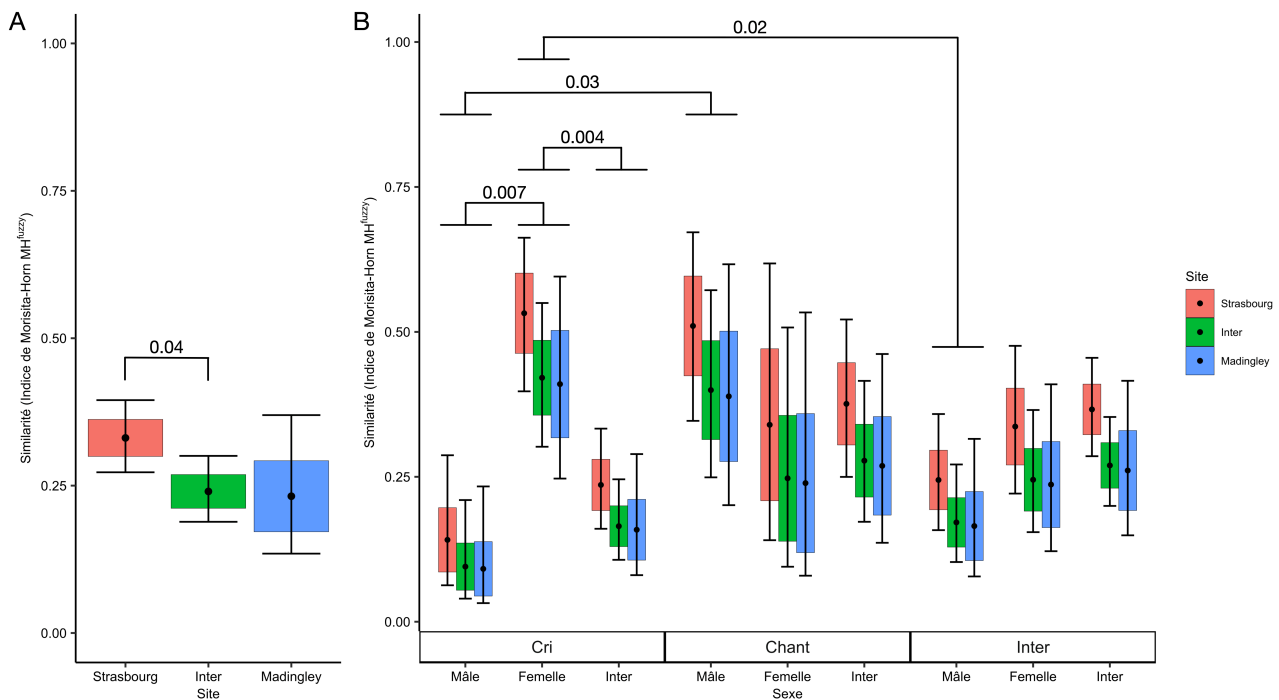


FIGURE 4.5 – Similarité entre les répertoires des individus par sexe, type de vocalisation et site. A) Le site a eu un effet statistique sans interaction avec les autres variables, et est donc représenté à part. B) Effets du sexe, du type de vocalisation et de leur interaction. Il est à noter que contrairement à la diversité et à la gradation, un niveau de comparaison supplémentaire (entre les deux sexes, entre les deux sites) est présent pour les deux variables et l'interaction. Pour chaque boxplot, le point correspond au paramètre estimé par le meilleur modèle, la boîte correspond à la moyenne  $\pm$  erreur standard, et les segments correspondent à l'intervalle de confiance à 95%. Les tests significatifs sont représentés par les lignes horizontales, avec la p-valeur indiquée au-dessus de la ligne.

La similarité entre sexes était indistinguable de celle des mâles entre eux (ratio entre sexes / mâles :  $1.88 \pm 0.96$ ,  $z = 1.24$ ,  $p = 0.95$ ), et significativement plus basse que celle des femelles entre elles (ratio entre sexes / femelles :  $0.27 \pm 0.093$ ,  $z = -3.82$ ,  $p = 0.004$ ). Concernant les chants, aucune différence n'existe dans la similarité entre individus d'un même sexe ou de sexes différents (e.g. ratio femelles / mâles :  $0.49 \pm 0.33$ ,  $z = -1.06$ ,  $p = 0.99$ ; ratio entre sexes / femelles :  $1.17 \pm 0.76$ ,  $z = 0.244$ ,  $p = 1$ ). D'autre part, les mâles sont significativement moins similaires entre eux dans leurs cris que dans leurs chants (ratio cris / chants :  $0.16 \pm 0.09$ ,  $z = -3.27$ ,  $p = 0.03$ ), alors que chez les femelles la différence n'est pas significative (ratio cris / chants :  $2.21 \pm 1.39$ ,  $z = 1.26$ ,  $p = 0.94$ ).

Malgré de forts effets statistiques, quelques exceptions existaient aux tendances suggérées par ces résultats. En particulier, certains mâles étaient néanmoins similaires dans leurs cris, mais uniquement deux à deux : par exemple, Merlin et Osiris à Strasbourg (56% de similarité dans leurs cris), et inversement Bussell à Madingley est hautement unique dans toutes ses vocalisations. Certaines paires d'individus sont plus similaires entre elles qu'avec les autres (e.g. Leo et Connelly à Madingley, 34% de similarité dans les cris, Balbo et Gigi à Strasbourg, 40% de similarité dans les cris également). Quant aux chants, la plupart des individus partagent une partie de leurs répertoires. Siobhan et Gigi semblent un cas particulier, étant similaires entre elles (57%) beaucoup plus qu'à tous les autres individus.

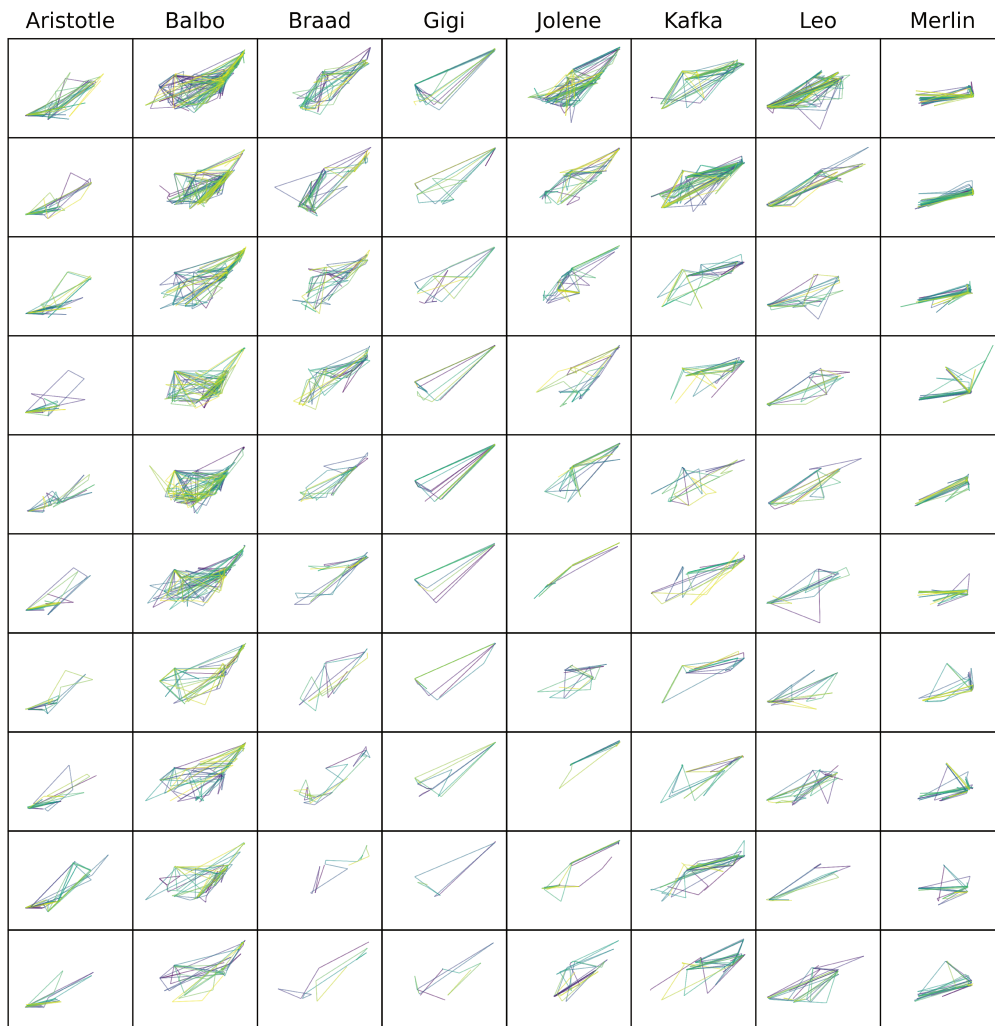


FIGURE 4.6 – Flexibilité dans les chants des individus les plus prolifiques. Les dix chants les plus longs de ces individus ont été extraits, puis les vocalisations correspondantes sont représentées après projection UMAP (celle du jeu de données complet pour que les positions des points soient comparables). Les transitions entre vocalisations successives sont représentées par des lignes entre chaque point, dont la couleur correspond à la position dans le chant (lignes sombres près du début, lignes claires près de la fin).

Nous n'avons pas inclus l'auto-similarité dans les analyses (i.e. la similarité entre les cris et les chants d'un même individu) afin de ne pas biaiser les résultats. La plupart des individus partagent une partie de leur répertoire entre leurs cris et leurs chants (e.g. Aristotle 87%, Bussell 99%, Merlin 94%), mais certains utilisent des répertoires en majorité différents (e.g. Kafka 20%, Tom 8%).

### 4.3.6 Flexibilité des chants

Enfin, nous avons examiné la potentielle flexibilité dans l'organisation séquentielle des chants (Fig. 4.6). Des différences inter-individuelles étaient apparentes dans la composition des chants : par exemple, entre Balbo et Merlin, Balbo occupait une région beaucoup plus étendue de l'espace que Merlin. Un individu donné tendait à occuper les mêmes régions de l'espace entre ses différents chants, illustrant une stabilité dans le répertoire vocal. Cependant, les transitions entre vocalisations successives variaient fortement entre les différents chants d'un individu (e.g. le premier chant de Balbo comparé à son deuxième : les vocalisations dans la partie supérieure gauche apparaissaient plus tôt et plus fréquemment dans le premier chant que dans le deuxième). Cependant, l'analyse n'a pas pu être poursuivie plus avant, en l'absence d'une méthode de description adéquate des transitions entre vocalisations qui permettrait de les comparer quantitativement.

## 4.4 Discussion

Pour évaluer la complexité vocale dans le répertoire d'une espèce complexe tant socialement que cognitivement, comme le corbeau freux, nous avons d'abord étudié la diversité et la gradation de leur répertoire. Les femelles ont un répertoire de cris plus divers et gradé que celui des mâles, mais étaient aussi pratiquement identiques dans leur répertoire de cris, y compris entre sites différents. Les mâles, eux, étaient presque uniques dans leurs cris, mais partageaient une partie non négligeable de leurs vocalisations de chants. Des vocalisations communes existent cependant entre mâles, bien qu'elles ne concernent généralement pas plus de deux individus à la fois. Enfin, l'examen des chants, bien que seulement descriptif, montre une grande variabilité inter-individuelle dans son arrangement. Certains individus, qui utilisent un nombre limité de vocalisations différentes dans leurs chants, tendent à toujours chanter de façon similaire. D'autres individus avec un répertoire de plus grande taille peuvent arranger leurs vocalisations de façon plus libre tant au sein d'un chant qu'entre deux chants.

Ces résultats sont frappants, en particulier l'extrême idiosyncrasie des cris des mâles. La communication entre les individus nécessite *a priori* des vocalisations communes, afin que les informations puissent être transmises et comprises. Cela favorise des répertoires vocaux communs dans les groupes sociaux, mais pas nécessairement des vocalisations communes à toute l'espèce (R. . F. LACHLAN et al., 2010). Ce répertoire vocal commun résulte fréquemment de l'imitation par les juvéniles de leurs partenaires sociaux lorsqu'ils acquièrent leurs vocalisations au cours de leur développement (ADRET-HAUSBERGER, 1982; KROODSMA, 2004; SALINAS-MELGOZA et WRIGHT, 2012), ou de la convergence vocale entre les partenaires sociaux adultes (CANDIOTTI et al., 2012; LUEF et al., 2017; SEWALL, 2009; WALLØE et al., 2015). Les groupes sociaux devraient partager au moins certaines de leurs vocalisations pour assurer l'intelligibilité de la communication, en particulier pour les signaux fonctionnels tels que les cris (MARLER, 2004b). Cependant, les corvidés semblent produire des vocalisations avec un haut degré de variabilité inter- et intra-individuelle (BENTI et al., 2019; MATES et al., 2015; ROSKRAFT et ESPMARK, 1982). Dans le répertoire vocal des corbeaux publié par ROSKRAFT et ESPMARK (1982), la plupart des cris étaient regroupés par fonction ou association contextuelle (e.g., duo, cri de nourrissage, cri de quémandage, croassement unique, croassement mul-



tuple), mais pas nécessairement par structure acoustique. Par exemple, même le "croassement unique", le cri le plus fréquemment émis, pouvait varier entre 0,25 et 0,65 s entre les individus, tout en étant stable pour un individu donné. Deux autres cris, dénommés "cri de la mouette" et "cri du squal", ont été décrits lors de disputes entre mâles dans les nids, mais leur modulation de fréquence était très variable, ce qui, selon les auteurs, pourrait indiquer un état d'excitation. MUNTEANU et al. (2009) ont également étudié le répertoire vocal des corbeaux freux et ont rapporté 9 cris structurellement différents et plusieurs variantes. Le répertoire vocal des corbeaux freux semble donc résister aux descriptions classiques.

Nous avons montré que les femelles avaient des répertoires de cris pratiquement identiques, y compris entre les deux sites. Il est intéressant de noter que la plupart des cris des femelles qui ont pu être contextualisés correspondaient à des cris associés soit au nourrissage, soit à un contexte de reproduction, en particulier les cris au nid. Ces cris sont probablement les plus fortement fonctionnels chez les femelles, car les mâles les nourrissent pendant la majeure partie de la saison de reproduction (COOMBS, 1961a ; ROSKAFT, 1983). Pendant la couvaison des œufs et des poussins jusqu'à leur autonomie thermique (environ 21 jours après l'éclosion), les femelles sont complètement dépendantes de leur compagnon, tant pour leur propre alimentation que pour celle des juvéniles (COOMBS, 1961a ; ROSKAFT, 1983). L'importance fonctionnelle de ces cris peut expliquer leur plus grande homogénéité entre individus. Ces fonctions communicatives essentielles peuvent aussi expliquer pourquoi les cris des femelles sont plus diversifiés et plus gradés que ceux des mâles. En admettant que les femelles partagent un répertoire vocal, le manque d'homogénéité présent dans les cris des mâles, tant entre eux qu'avec les femelles, a rarement été décrit chez d'autres espèces. Seules des études sur les cris de contact des perroquets (BERG et al., 2012 ; CORTOPASSI et BRADBURY, 2006) et les sifflements de signature des grands dauphins (QUICK et JANIK, 2012) ont trouvé une hétérogénéité inter-individuelle similaire, mais seulement dans ces vocalisations particulières. Comme les corbeaux, ces espèces suivent une organisation sociale complexe basée sur la fission-fusion (BRADBURY et al., 2001 ; CONNOR et al., 2000). Les perroquets, mais pas les dauphins ni les corvidés, ont démontré leur capacité à imiter vocalement des congénères particuliers pour s'adresser à eux lors d'événements sociaux (BALSBY et al., 2012). De plus, toutes ces espèces sont parmi les plus cognitivement aptes de toutes les espèces animales connues (EMERY, 2004 ; LAMBERT et al., 2019 ; MARINO, 2004) ; ils pourraient donc avoir la capacité de simplement mémoriser des cris particuliers en association avec des individus particuliers. Pourtant, à notre connaissance, chez les espèces autres que le corbeau freux, cette individualité était limitée aux cris de contact, les autres cris étant soit peu clairs, soit partagés entre les individus, d'une manière compatible avec ce que nous attendrions pour la communication. Au contraire, chez les corbeaux freux, chaque mâle utilise largement ses propres cris. Cela remet également en question le concept de signature individuelle chez cette espèce, où l'information portant la fonction d'un cri est censée rester standardisée afin d'être comprise par les autres. Une signature vocale avait été mise en évidence dans un des cris de plusieurs mâles du groupe de Strasbourg (BENTI et al., 2019). Ce cri, le plus fréquent émis par ces mâles, était produit sans contexte apparent, ou en plusieurs contextes différents (soumis cependant à interprétation humaine), et différait entre mâles selon certains paramètres acoustiques. Cependant, ce cri n'était produit que par certains mâles,

et n'avait été considéré comme un même cri que parce qu'il s'agissait du cri le plus commun, émis dans la même absence ou multitude de contextes. Il n'est donc pas forcément surprenant d'y trouver une signature individuelle claire, au vu de nos résultats. Nos résultats sont cependant différents de la découverte d'une simple signature vocale : la distance DFTW tient compte de toutes les différences de structure acoustique que l'on peut obtenir en étirant une vocalisation selon l'axe temporel ou l'axe fréquentiel. Ici, la plupart des cris des mâles avaient des structures acoustiques différentes sur la base du spectrogramme entier même après avoir effacé ces variations, ce qui devrait correspondre à des différences majeures de structure acoustique et devrait fortement entraver la communication d'informations entre congénères. Cependant, les relations sociales chez les corbeaux freux correspondent principalement en la relation de couple, avec des relations relativement ponctuelles entre les autres individus d'un group social stable (BOUCHERIE et al., 2016). Un mâle donné pourrait alors n'avoir besoin de reconnaître qu'un nombre restreint d'individus (le ou la partenaire de couple, et quelques partenaires sociaux), et simplement apprendre et mémoriser la façon de vocaliser de chacun de ces individus pour un contexte donné. L'observation du même phénomène dans les deux groupes suggère que ces différences dans les cris des mâles pourrait être généralisable au niveau de l'espèce, et n'est donc probablement pas un artéfact de la captivité.

Nos résultats sont d'autant plus surprenants que certains de ces individus avaient été en contact étroit pendant toute leur vie (tous les oiseaux de Cambridge d'une part, et la moitié des oiseaux de Strasbourg d'autre part), et que si nous détectons une certaine proximité vocale entre certains individus, elle ne peut être expliquée par une proximité sociale ou génétique (les seuls individus potentiellement apparentés étant Elie et Jonas à Strasbourg, et Cassandra et Leo à Madingley). Nous n'avons pas non plus trouvé de preuve que les cris ou les chants puissent porter des signatures vocales de couple ou de groupe, car même au sein des couples la similarité vocale est très faible (y compris pour les deux couples homosexuels, Elie et Tom à Strasbourg, Bussell et Fry à Madingley).

Nous avons montré que le répertoire des corbeaux peut être extrêmement riche, ce qui suggère un système de communication complexe. Les répertoires des différents individus étaient si divers qu'il devient difficile de conclure sur la façon dont les corbeaux freux communiquent, au moins les mâles. Nous suggérons que cette diversité peut être liée à une plus grande flexibilité dans la production vocale chez cette espèce par rapport aux autres oiseaux chanteurs. Une gradation importante dans les cris a été observée tant chez les mâles que chez les femelles, constituant un marqueur fort de flexibilité vocale. Il est important de noter que si tous les individus ne chantent pas, ceux qui le font présentent également des différences notables dans la composition des chants d'un chant à l'autre (dans l'arrangement séquentiel des unités vocales) et entre les individus (unités vocales différentes et quantité différente d'unités vocales). Certains mâles utilisaient également principalement des vocalisations différentes dans leur chant comparé à leur propre répertoire de cris. La flexibilité du chant, avec un arrangement vocal changeant d'un chant à l'autre, n'est pas une compétence courante chez les oiseaux chanteurs, même chez les espèces à chant ouvert (BEECHER et BRENOWITZ, 2005). L'étourneau sansonnet, l'une des espèces à apprentissage ouvert les mieux étudiées, peut modifier certaines parties des chants produits, mais le schéma général individuel est généralement maintenu d'un chant à l'autre (ADRET-HAUSBERGER et GÜTTINGER, 1984; SAINBURG et al., 2020). La flexibilité vocale

constatée chez les corbeaux rappelle ce qui a été décrit dans les cris d'autres espèces comme le mainate (*Gracula religiosa*, BERTRAM, 1970) et surtout les perroquets (MONTES-MEDINA et al., 2016), et les chants de certains chanteurs comme le rossignol (*Luscinia megarhynchos*, KIPPER et al., 2004), le moqueur roux (BOUGHEY et THOMPSON, 1981), ou l'oiseau lyre (*Menura alberti*, BACKHOUSE et al., 2022). Plusieurs de ces espèces, ainsi que les corvidés (E. D. BROWN, 1985), sont capables d'imiter les sons de l'environnement (BACKHOUSE et al., 2022 ; BRADBURY et BALSBY, 2016), et ils pourraient avoir suffisamment de flexibilité vocale pour innover fréquemment en vocalisant. Un autre type de flexibilité vocale est la flexibilité fonctionnelle, c'est-à-dire la capacité à détacher un cri d'un contexte. Certaines espèces peuvent utiliser un cri pour plusieurs contextes, et il a été suggéré qu'il s'agissait d'une étape importante dans l'évolution du langage humain (OLLER et al., 2013 ; OLLER et al., 2016). Bien qu'une grande partie des vocalisations des corbeaux freux n'ait pas pu être reliée à un contexte fonctionnel dans cette étude, des travaux antérieurs ont montré qu'un individu donné pouvait produire le même cri dans une variété de contextes (BENTI et al., 2019), ce qui pourrait satisfaire le critère de flexibilité vocale, mais nécessiterait une étude plus approfondie.

Une question restante pourrait être la fonction de cette grande diversité vocale et de cette flexibilité des chants chez les corbeaux mâles. Les chants chez les corbeaux freux ont très rarement été observés dans le contexte de la reproduction, et étant donné l'absence de territorialité chez cette espèce sociale, les chants ont très probablement une autre fonction. D'autres corvidés ont été observés produisant des vocalisations similaires avec la même absence de contexte (E. D. BROWN, 1985 ; BROWN et FARABAUGH, 1997 ; COOMBS, 1960). Le chant a été lié à la voie de récompense dopaminergique chez l'étourneau sansonnet (RITERS et al., 2022 ; RITERS et al., 2019 ; STEVENSON et al., 2020), et a été suggéré comme une forme de jeu auto-récompensé (RITERS et al., 2019). Ainsi, le chant pourrait également avoir une fonction hédonique, i.e. un oiseau pourrait chanter pour le plaisir de chanter. Certains oiseaux produisent encore des comportements de jeu à l'âge adulte (DIAMOND et BOND, 2003 ; FICKEN, 1977), et les corvidés ont été décrits comme parmi les espèces avec le jeu le plus marqué et complexe (EMERY et CLAYTON, 2015 ; FICKEN, 1977 ; O'HARA et AUERSPERG, 2017). Le chant pourrait donc être une autre manifestation du jeu chez les corbeaux freux. Une autre possibilité est que les corbeaux peuvent rechercher activement la singularité dans leurs vocalisations, comme un moyen de maintenir l'individualité pour la communication dans les grandes colonies de reproduction. Les chants pourraient être un moyen de pratiquer, d'essayer et d'innover de nouvelles vocalisations qui pourraient ou non intégrer leur répertoire vocal (SCHRAFT et al., 2017). Il pourrait en effet s'agir du seul moyen de se faire reconnaître lorsque les petits groupes sociaux stables fusionnent en immenses colonies à la reproduction, où non seulement des groupes sociaux de la même région, mais potentiellement des populations migrantes, se rassemblent dans un espace relativement restreint (COOMBS, 1960 ; PATTERSON et al., 1971). Les oiseaux individuels peuvent avoir besoin d'être très distinctifs les uns des autres. La divergence vocale, lorsqu'elle se produit dans des populations sympatriques, est généralement considérée comme un signe précoce de spéciation en raison de l'influence de la vocalisation sur le choix du partenaire (becs-croisés rouges *Loxia curvirostra* EDELAAR et BENKMAN, 2006, moineau friquet à oreilles blanches *Zonotrichia leucophrys* DALISIO et al., 2015 ; SANDOVAL et al., 2017). Cependant, cela ne semble pas être le cas ici car la plupart des individus diffèrent les uns

des autres, et pas seulement entre les sous-groupes.

Sur un plan plus technique, comparé à d'autres études dans lesquelles les vocalisations sont étudiées grâce à divers paramètres acoustiques, nous avons utilisé une distance entre spectrogrammes entiers, basée sur la DFTW. Cette distance nous a permis de comparer les vocalisations malgré des différences non seulement dans la vitesse de restitution, mais aussi dans la distribution des fréquences. La projection et le clustering de ces distances nous ont permis de quantifier la diversité, la gradation et la similarité vocale de chaque individu dans les groupes étudiés. La gradation et la diversité étaient toutes deux élevées dans les données. La similarité montre un dimorphisme sexuel important entre des femelles très similaires et des mâles très différents, avec néanmoins un effet atténué dans les chants, où un partage de vocalisation existe dans une certaine mesure y compris entre sexes. Outre leur complexité sociale, les corbeaux freux possèdent donc un grand potentiel de complexité vocale. L'approche décrite ici, moins basée sur des jugements perceptuels humains que les approches classiques, pourrait être mieux adaptée à l'étude de cette complexité vocale.

## 4.5 Annexe

Dans l'article principal je décrivais les calculs avec la diversité de Simpson  $^2D$  comme donnant des résultats très similaires à ceux présentés dans l'article. Dans cette annexe je les présenterai très rapidement.

TABLE S1 – Sélection de modèle pour la gradation et la diversité du répertoire, en utilisant  $^2D$  comme estimateur.

Modèle	Type	Sexe	Site	Type :Sexe	Type :Site	Sex :Site	AICc	delta	pois
Gradation	+	+	+	+	+		164132.4	0	0.66
	+	+	+	+	+	+	164134.2	1.8	0.27
							164369.7	237.28	0
Diversité	+	+		+			269.6	0	0.41
	+	+	+	+			270.5	0.81	0.28
	+	+	+	+	+		271.5	1.88	0.16
						+	282.9	13.23	0.001

Tout d'abord, la sélection de modèle donne exactement le même résultat pour la diversité que dans l'article principal : le meilleur modèle est celui qui inclut uniquement le type de vocalisation, le sexe, et leur interaction, sans effet du site. Deux autres modèles sont équivalents, mais sont moins parcimonieux, je ne les ai donc pas considérés plus avant. Pour la gradation cependant, le meilleur modèle inclut tous les termes, sauf l'interaction entre sexe et site. Le modèle complet est même équivalent à ce meilleur modèle. Dans le deux cas, le modèle nul prédit significativement moins bien la variable réponse que les meilleurs modèles.

Les résultats sur la diversité avec l'indice de diversité de Simpson  $^2D$  sont sensiblement similaires à ceux observés avec l'indice de diversité de Shannon  $^1D$ . Quant à la gradation, l'indice  $^2D$  révèle un statistiquement significatif du site, contrairement à l'indice  $^1D$ . Cependant, les tailles d'effets ne sont pas qualitativement différentes, et les conclusions restent les mêmes concernant la complexité vocale des mâles et des femelles, de cris et des chants.

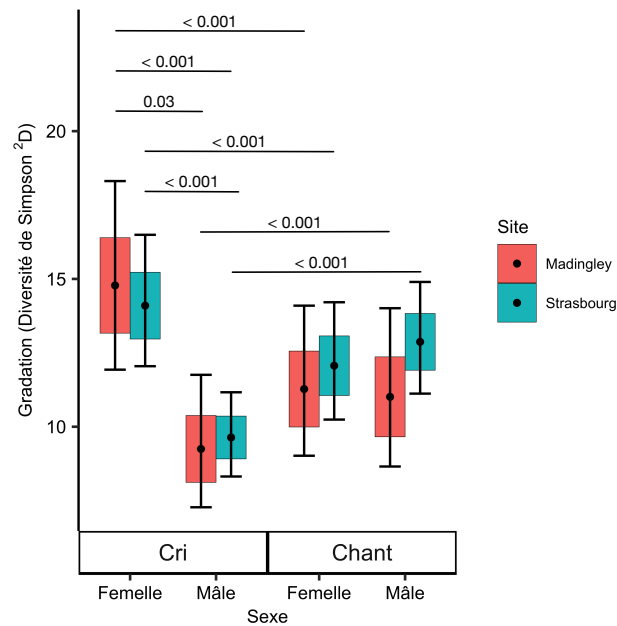


FIGURE S1 – Gradation dans le répertoire vocal par sexe, type de vocalisation, et site, estimée par l'indice  $^2D$ . Le site n'avait aucun effet statistique et n'était pas inclus dans les tests statistiques, mais est inclus dans cette visualisation. Pour chaque boxplot, le point correspond au paramètre estimé par le meilleur modèle, la boîte correspond à la moyenne  $\pm$  erreur standard, et les segments correspondent à l'intervalle de confiance à 95%. Les tests significatifs sont représentés par les lignes horizontales, avec la p-valeur indiquée au-dessus de la ligne.

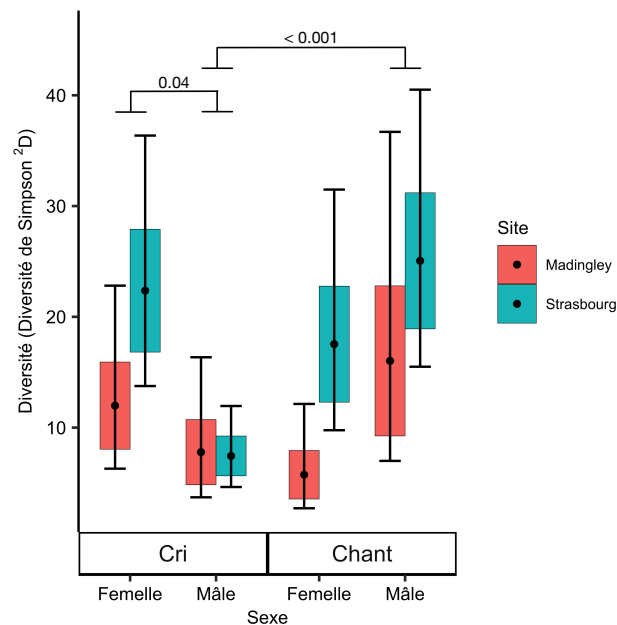


FIGURE S2 – Diversité dans le répertoire vocal par sexe, type de vocalisation, et site, estimée par l'indice  $^2D$ . Le site n'avait aucun effet statistique et n'était pas inclus dans les tests statistiques, mais est inclus dans cette visualisation. Pour chaque boxplot, le point correspond au paramètre estimé par le meilleur modèle, la boîte correspond à la moyenne  $\pm$  erreur standard, et les segments correspondent à l'intervalle de confiance à 95%. Les tests significatifs sont représentés par les lignes horizontales, avec la p-valeur indiquée au-dessus de la ligne.



# Chapitre 5

## Signature vocale chez le corbeau freux, de l'individu à la colonie

Killian Martin, Louise Nervet, Emily Danby, Francesca Cornero, Virgile Daunay, Émeline Braccini, Olivier Adam, Nicolas Obin, Nicola S. Clayton, Valérie Dufour

En préparation pour soumission à *Ethology* sous le titre

*Can we trace the colonial affiliation of rooks through their vocal signature ?*

### Résumé

La reconnaissance interindividuelle est une condition nécessaire à la formation de relations sociales. La plupart des oiseaux vocalisent de manière individuelle et distinctive ; certaines espèces ont démontré expérimentalement une reconnaissance médiée par la vocalisation. Cependant, de nombreuses espèces sont socialement plus complexes, et peuvent donc avoir besoin de reconnaître non seulement les individus, mais aussi le groupe social. Une signature vocale à l'échelle du groupe peut donc permettre la reconnaissance des partenaires sociaux, comme dans les dialectes que l'on observe dans les chants de nombreux oiseaux chanteurs. En outre, différents niveaux d'organisation sociale, et donc de reconnaissance, pourraient être encodés dans différentes signatures vocales. Nous avons testé cette hypothèse dans les cris de nidification des corbeaux freux, produits exclusivement pendant la saison de reproduction lorsque les femelles couvent dans leur nid. Ce cri est l'une des rares vocalisations qui peut être attribuée sans erreur à une femelle par simple observation, sans nécessiter de capture ou marquage préalable. Nous avons trouvé une augmentation légère mais significative de la distance acoustique en fonction de la distance physique, jusqu'à un maximum pour les comparaisons entre colonies différentes. Une signature vocale individuelle marquée a pu être trouvée, mais nous n'avons pas mis en évidence de signature vocale de groupe ou de colonies. Le cri au nid est une vocalisation puissante, utilisée notamment par la femelle pour appeler le mâle du couple pendant la reproduction, pendant laquelle ce dernier est le seul à lui apporter de la nourriture. Il ne serait donc pas avantageux pour des femelles voisines d'être trop similaires. Le cri au nid semble cependant être un bon outil de suivi individuel, étant donné ses avantages : la puissance du cri en facilite l'enregistrement, la position fixe de la femelle permet de l'identifier initialement, puis le cri peut être utilisé pour la suivre potentiellement à long terme. Des études futures devront vérifier la stabilité de la signature vocale dans ce cri.

### 5.1 Introduction

La capacité à reconnaître différents individus au cours de multiples interactions est cruciale pour la formation de relations interindividuelles (BRADBURY et VEHCAMP, 1998). Chez les oiseaux,



la reconnaissance est souvent médiée par les vocalisations (BEER, 1971 ; BRECHT et NIEDER, 2020 ; KONISHI, 1985 ; MARLER et SLABBEKOORN, 2004), en particulier dans les environnements socialement encombrés ou lorsque les informations visuelles ne sont pas toujours disponibles (CURÉ et al., 2009). La signature vocale, l'ensemble des caractéristiques acoustiques qui identifient un individu, dépend de l'association de certaines caractéristiques vocales à certains individus et permet de les distinguer des congénères (BEECHER, 1989 ; BENTI et al., 2019 ; BRADBURY et VEHRENCAMP, 1998 ; ELIE et THEUNISSEN, 2018 ; KERSHENBAUM et al., 2014 ; LINHART et al., 2019 ; PRŮCHOVÁ et al., 2017 ; TERRY et al., 2005). La capacité à utiliser ces informations vocales pour discriminer différents individus a été démontrée expérimentalement chez plusieurs espèces (BRECHT et NIEDER, 2020 ; ELIE et THEUNISSEN, 2018 ; KONDO et al., 2010 ; KONDO et WATANABE, 2009 ; ROSKAFI et ESPMARK, 1982 ; VIGNAL et al., 2008 ; WEARY et KREBS, 1992 ; WOODS et al., 2018). Dans les grands groupes sociaux, les individus acquièrent parfois des vocalisations similaires par imitation ('partage vocal', BAKER et al., 2000 ; KREMERS et al., 2012), ou modifient leurs vocalisations pour converger vers une structure similaire ('convergence vocale', CANDIOTTI et al., 2012 ; MONTEIRO et al., 2021 ; SEWALL, 2009), dans certains cas même à l'échelle d'une seule interaction comme chez les perroquets (SCARL et BRADBURY, 2009). Lorsque ces vocalisations communes suivent des variations macro- ou micro-géographiques, elles donnent lieu à des dialectes, partagés par l'ensemble d'une population sympatrique (CATCHPOLE et SLATER, 2008 ; MARLER et TAMURA, 1962 ; MCGREGOR, 1980 ; MUNDINGER, 1982 ; SEARCY, 1992 ; WRIGHT et DAHLIN, 2017). La similarité acoustique peut résulter de plusieurs facteurs. Par exemple, selon l'hypothèse de l'adaptation acoustique (HANSEN, 1979), les environnements fermés et les obstacles peuvent conduire à des fréquences sonores plus basses dans les vocalisations afin de compenser l'atténuation du son qui a lieu dans ce genre d'environnement. Inversement, les oiseaux vivant dans des environnements urbains produisent des vocalisations à plus haute fréquence, afin de compenser les bruits associés à cet environnement (DEONIZIAK et OSIEJUK, 2019 ; SLABBEKOORN et PEET, 2003). Les vocalisations et leurs paramètres acoustiques pourraient alors être assez similaires car adaptés au milieu de vie des individus. La convergence vocale peut également reposer sur la proximité génétique entre individus apparentés, ou être un sous-produit de l'apprentissage vocal, quand celui-ci est influencé par le modèle représenté par les partenaires sociaux (ANDREW, 1962 ; BITTERBAUM et BAPTISTA, 1979 ; BOUGHMAN, 1998 ; BROWN et FARABAUGH, 1997 ; MONTEIRO et al., 2021 ; SALINAS-MELGOZA et WRIGHT, 2012 ; WIENS, 1982), souvent les parents et les individus affiliés (MARLER et SLABBEKOORN, 2004 ; SHARP et al., 2005). L'hypothèse d'adaptation sociale propose que les dialectes augmentent la tolérance entre congénères à la manière d'un "mot de passe vocal" (BROWN et FARABAUGH, 1997 ; FEEKES, 1977). Chez les espèces territoriales, par exemple, ce "mot de passe" permet de reconnaître les individus des territoires voisins (BEECHER et CAMPBELL, 2005 ; CATCHPOLE et SLATER, 2008 ; SHACKLETON et RATCLIFFE, 1994). Ainsi, si les dialectes sont des indicateurs de l'origine géographique ou micro-géographique des individus, ils peuvent également porter une information sur l'origine sociale et la familiarité entre congénères (HAUSBERGER et al., 2008 ; SALINAS-MELGOZA et WRIGHT, 2012).

La complexité sociale est fréquemment associée à des systèmes de communication complexes, une

corrélation prédite dans le cadre de l'hypothèse de la complexité socio-communicative (BLUMSTEIN et ARMITAGE, 1997; FREEBERG et al., 2012; PECKRE et al., 2019). Dans les systèmes de fission-fusion, où de multiples groupes sociaux fusionnent de manière réversible en de grands groupes (AMICI et al., 2008; ARCHIE et al., 2006; BRADBURY et BALSBY, 2016; BRADBURY et al., 2001; CONNOR et al., 2000; KELLEY et al., 2011; SILK et al., 2014; WILLIS et BRIGHAM, 2004) les individus peuvent avoir besoin de signaler leur identité individuelle mais aussi les différentes couches sociales dans lesquelles ils sont impliqués : couple, famille, matriligne, groupe, voire colonie. Une signature individuelle peut alors permettre la reconnaissance entre des individus spécifiques, même à différentes couches d'organisation sociale (KONDO et WATANABE, 2009; VIGNAL et al., 2008).

Cependant, la plupart des espèces produisent un répertoire de plusieurs vocalisations différentes ayant une fonction donnée (SEARCY, 1992). Le codage d'informations supplémentaires sur l'identité et l'appartenance à une paire, un groupe ou une colonie introduit un potentiel de codage d'informations conflictuelles (MARLER, 2004a). Une solution possible à ce conflit est la segmentation temporelle par la production de vocalisations différentes, soit individuellement distinctives (c'est-à-dire des vocalisations qui peuvent servir à identifier un individu ou à signaler l'appartenance sociale à un groupe), soit fonctionnelles (c'est-à-dire des vocalisations qui communiquent d'autres informations). Cette séparation temporelle a été trouvée chez certains oiseaux (CHARRIER et al., 2001; ELIE et THEUNISSEN, 2018) et mammifères (BOUCHET et al., 2012; JANSEN et al., 2012; KEENAN et al., 2020). Une solution est une signature vocale cohérente, un ensemble de paramètres acoustiques gardés constants à travers tout le répertoire de l'individu. Cependant, cette possibilité n'a pas été soutenue par les quelques études sur cette question, d'une part par des variations dans la fiabilité de la signature vocale à travers le répertoire (BOUCHET et al., 2012; CHARRIER et al., 2001; KEENAN et al., 2020; PRICE et al., 2009; VANNONI et MCELLIGOTT, 2007), et par des démonstrations directes que des paramètres différents sous-tendent la signature vocale dans des vocalisations différentes (ELIE et THEUNISSEN, 2018; KEENAN et al., 2020), y compris par un manque de stabilité vocale des individus au cours du temps (ZDENEK et al., 2018).

Les oiseaux, avec leurs divers modèles d'agrégation, sont des espèces idéales pour étudier le lien entre les systèmes sociaux et l'établissement signatures vocales à différents niveaux sociaux (GRUETER et al., 2020). Pour les espèces migratrices ayant une dynamique de fission-fusion, les signatures vocales peuvent aider les individus à se regrouper avec des membres familiers du groupe ou de la colonie après plusieurs semaines ou mois de séparation. Elles peuvent aider à retrouver des partenaires, des parents, ou une progéniture parmi une multitude d'individus (BENTI et al., 2019; SAUVÉ et al., 2015). Compte tenu de son organisation sociale complexe souvent qualifiée comme suivant une dynamique de fission-fusion (BOUCHERIE et al., 2016; CLAYTON et EMERY, 2007; PATTERSON et al., 1971; RÖELL et BOSSEMA, 1982), le corbeau freux est un modèle particulièrement intéressant. En effet, il s'agit une espèce monogame mais aussi extrêmement sociale, vivant toute l'année en groupes sociaux, qui se regroupent en colonies de mars à juin pour la reproduction. Ces colonies peuvent compter de quelques couples à plusieurs dizaines de milliers (CLAYTON et EMERY, 2007; MARSHALL et COOMBS, 1957; PATTERSON et al., 1971; SWINGLAND, 1977; SWINGLAND, 1973). Après la saison de reproduction, les colonies se divisent en petits groupes d'individus et restent à

proximité de leur site colonial, se rassemblant régulièrement sur ce dernier avant de partir en plusieurs groupes à la recherche de nourriture (COOMBS, 1960 ; PATTERSON et al., 1971). En outre, différentes populations de corbeaux freux diffèrent dans leurs schémas de migration. En Angleterre, et dans les pays d'Europe du Nord, la population de corbeaux freux a été décrite comme étant principalement sédentaire (BUSSE, 1969 ; SWINGLAND, 1977), avec moins de 12% des oiseaux migrant chaque année. Cette organisation suggère que la plupart des oiseaux locaux se rencontrent fréquemment au cours de l'année, et pourraient se connaître, au niveau du groupe et de la colonie. En Europe de l'Ouest (France, Allemagne de l'Est), la population de corbeaux freux est composée de 28% à 63% d'oiseaux migrateurs qui peuvent parcourir de très grandes distances (parfois plus de 1000 km d'une année à l'autre, BUSSE, 1969). Selon BUSSE (1969), les oiseaux plus âgés ont tendance à moins migrer que les plus jeunes. Il n'est pas clair dans quelle mesure les individus de populations différentes, mais qui se regroupent dans les mêmes colonies, se reconnaissent entre eux. Cependant, l'hypothèse a été émise que les couples reviennent au même nid d'une année à l'autre (COOMBS, 1960 ; GOODWIN, 1955 ; RICHARDSON et al., 1979)). Au moins dans les populations britanniques, des couples voisins s'engagent dans une défense réciproque de leurs nids contre les oiseaux étrangers (COOMBS, 1960 ; RÖELL et BOSSEMA, 1982). Les voisins pourraient donc être familiers entre eux, et peut-être même faire partie du même groupe social tout au long de l'année.

Les corvidés sont généralement des oiseaux très vocaux et utilisent un répertoire vocal riche dans leurs interactions sociales (BAYENS, 1981 ; CHAMBERLAIN et CORNWELL, 1971 ; HA et al., 2003 ; HEINRICH, 1988 ; MATES et al., 2015). Leurs vocalisations sont souvent très inharmoniques, résultat de modulations rapides de fréquence et de non-linéarités permises par l'appareil vocal qu'ils partagent avec les autres oiseaux, appelé syrinx (BOCK, 1978). Néanmoins, certaines de leurs vocalisations sont individuellement distinctives (BENTI et al., 2019 ; BOECKLE et al., 2012 ; LAIOLO et al., 2000 ; MATES et al., 2015 ; STOWELL et al., 2016a ; TANIMOTO et al., 2017 ; YORZINSKI et al., 2006), et plusieurs espèces peuvent également reconnaître des individus particuliers par leurs vocalisations (HOPP et al., 2001 ; KONDO et al., 2010 ; WASCHER et al., 2012b), y compris le corbeau freux (BENTI et al., 2019 ; ROSKRAFT et ESPMARK, 1982). Chez les corbeaux freux, le répertoire a été décrit comme comptant une vingtaine de cris environ, mais cet décompte représente le répertoire vocal au niveau du groupe social (MUNTEANU et al., 2009 ; ROSKRAFT et ESPMARK, 1982, mais voir également Martin et al, in prep, Chapitre 4). Parmi ces cris, le plus commun présente une signature individuelle chez les corbeaux freux mâles (BENTI et al., 2019). Les femelles produisent également des cris puissants, exclusivement associés à la reproduction, en particulier lorsqu'elles restent dans le nid (ci-après le cri au nid). Cette vocalisation est puissante et facilement détectable même à travers un environnement aussi encombré vocalement que la colonie, ce qui pourrait faciliter l'identification des femelles par leur partenaire. Cependant, au-delà d'une signature individuelle, ce cri pourrait permettre d'étudier une signature vocale de groupe chez les corbeaux freux. Ici, nous avons cherché à savoir si ce cri pouvait être utilisé pour déduire l'appartenance des femelles à un groupe de nids ou à une colonie. En utilisant des approches modernes, nous avons évalué si ce cri porte des éléments vocaux de signature individuelle, de groupe et coloniale à la fois en France et en Angleterre. Nous attendons au moins une signature vocale individuelle dans le cri au nid de ces femelles. Par ailleurs, deux colonies captives,

une dans chaque pays, ont été enregistrées. Chacune de ces colonies compte quelques femelles qui ont vécu ensemble pendant plusieurs années, et nichent à proximité les unes des autres chaque année. Nous pouvons donc attendre que si des signatures vocales existent au niveau du groupe ou de la colonie, nous devrions les trouver au moins dans ces deux colonies captives.

## 5.2 Matériel et Méthodes

### 5.2.1 Colonies d'études

Nous avons enregistré des corbeaux freux appartenant à trois colonies sauvages et deux colonies captives. Deux colonies sauvages et une captive ont été enregistrées autour de Strasbourg (France), une sauvage et une captive ont été enregistrées près de Madingley (UK). Par souci de simplicité, nous désignons ci-après ces colonies : "CNRS" pour une colonie sauvage établie sur le campus CNRS de Cronembourg à Strasbourg (48° 36' 29" N ; 7° 42' 57" E), "Contades" pour la seconde colonie sauvage de Strasbourg établie dans le parc public urbain des Contades (48°35'29.303" N ; 7°45'28.564" E), "Strasbourg\_Captive" pour la colonie captive hébergée dans une volière extérieure dans le parc public urbain Albert-Schweitzer (48° 34' 21.496" N ; 7° 42' 34.936" E), "Cambridge\_Sauvage" pour la colonie sauvage enregistrée près de Madingley (52° 13' 27.88" N ; 0° 2' 30.15" E) et "Cambridge\_Captive" pour la colonie captive hébergée dans une volière extérieure au Sous-Département d'étude du Comportement Animal de l'Université de Cambridge, à Madingley (52° 13' 27.88" N ; 0° 2' 30.15" E).

Les deux colonies captives comprenaient des oiseaux capturés en tant que juvéniles dans des colonies sauvages proches de leurs sites respectifs. Tous les oiseaux étaient hébergés ensemble dans des volières extérieures communes depuis plusieurs années au moment de l'étude. Certains des oiseaux de la colonie Strasbourg\_Captive avaient été capturés dans la colonie du CNRS en 2006, et tous avaient été hébergés dans une volière extérieure sur le campus du CNRS de Cronembourg (à moins de 200 m de la colonie sauvage du CRNS) jusqu'en juillet 2018.

### 5.2.2 Protocole d'enregistrement

Nous avons enregistré les cris au nid des femelles pendant la saison de reproduction. Les sessions d'enregistrement ont été réalisées en 2020, 2021 et 2022 entre la mi-mars, juste avant la saison de reproduction, et la mi-avril, avant que les nids soient complètement cachés à la vue par les feuilles. Différentes plages horaires ont été enregistrées en fonction de la colonie : les colonies captives étaient enregistrées plusieurs heures consécutives chaque jour dans le cadre d'une autre étude. Les colonies sauvages étaient généralement enregistrées en début d'après-midi quand les conditions météorologiques le permettaient. Les enregistrements ont été réalisés par Song Meter 4 (Wildlife Acoustics, USA), avec des microphones externes montés sur des câbles de longueur variable (3 m pour les colonies captives, 20 m pour la colonie sauvage du CNRS) ou les microphones internes (pour les colonies Contades et Cambridge\_Wild). Les gains des microphones ont été ajustés empiriquement pour éviter la saturation, et les enregistrements ont été numérisés à 48 kHz avec un débit de 16 bits, et stockés sous forme de fichiers WAV non compressés.

Toutes les femelles captives portaient des anneaux de pattes colorés, et étaient habituées à la présence humaine pendant les observations. L'observateur restait cependant à distance des nids pour ne pas stresser les animaux. Les femelles sauvages ont été identifiées uniquement par la position du nid. Pour ce faire, les observateurs étaient postés en hauteur dans des bâtiments voisins, approximativement à la même hauteur que les nids, et aussi près que possible de la colonie. Les femelles qui vocalisaient ont été suivies de manière opportuniste (i.e. en suivant chaque femelle quand elle commençait à vocaliser, notant chaque vocalisation, jusqu'à environ 15 vocalisations consécutives). Des cartes des colonies sauvages avec les positions de chaque nid ont été dessinées pour éviter de répéter les observations, et des enregistrements vidéo ont été réalisés pour aider à lever d'éventuelles ambiguïtés. En outre, puisque les corbeaux freux nichent en groupes au sein de la colonie, nous avons également défini des groupes avec chaque colonie sauvage. Chaque groupe correspondait à un groupe de nids très proches (généralement dans le même arbre avec 4 m ou moins entre les nids, et à la même hauteur). Cependant, la colonie Cambridge\_Wild n'a pas pu être observée de cette manière. Les vocalisations ont donc été enregistrées mais non attribuées à des femelles individuelles ou à des groupes au sein de la colonie. Pour limiter la probabilité d'échantillonner la même femelle trop fréquemment, un maximum de 5 vocalisations ont été extraites par séquence de vocalisations (i.e. séparées par moins de 10 secondes environ).

Les cris au nid produits par les femelles identifiées ont été extraits à l'aide du logiciel Audacity (v3.1.3). Seuls les cris qui ne chevauchaient pas d'autres sons ont été utilisés. D'autres cris au nid, de bonne qualité mais qui ne pouvaient pas être attribués à des individus, ont également été extraits. Ceci a permis d'obtenir 3351 cris au total (Tableau 5.1).

TABLE 5.1 – Distribution des vocalisations par colonie. Pour chaque colonie : "Total" est le nombre de cris collectés au total des oiseaux de cette colonie, "Dont inconnues" décompte le nombre de vocalisations non attribuées, puis les colonnes suivantes correspondent à diverses statistiques : moyenne, écart-type, minimum et maximum, et enfin le nombre d'individus dans la colonie. Les vocalisations non attribuées à des oiseaux identifiés ne sont pas comptées dans les statistiques.

Colonie	Total	Dont inconnues	Moyenne $\pm$ Ecart-type	Min - Max	Individus
Strasbourg_Captive	480	0	80 $\pm$ 61.4	9 - 194	6
CNRS	1387	431	21.7 $\pm$ 20.6	2 - 94	44
Contades	327	38	14.4 $\pm$ 8.64	2 - 35	20
Cambridge_Captive	224	0	56 $\pm$ 4.40	51 - 61	4
Cambridge_Sauvage	933	933	-	-	-

### 5.2.3 Représentation des cris : Mel-spectrogrammes

Pour chaque cri, nous avons extrait l'audio et calculé le spectrogramme correspondant. Ce spectrogramme a ensuite été débruité en utilisant un algorithme de seuillage spectral basé sur des échantillons de bruit prélevés dans les enregistrements immédiatement avant et après chaque cri (voir Méthodes générales pour l'algorithme détaillé). Les spectrogrammes ont été calculés avec une fenêtre de Hamming de 10 ms avec un chevauchement de 80 %, puis converti en Mel-spectrogrammes avec 80 coefficients Mel. Cela a permis de réduire les dimensions des spectrogrammes et de mettre l'accent sur

les basses fréquences. où se concentre la majeure partie de l'énergie des vocalisations des corbeaux freux (BENTI et al., 2019 ; ROSKAFI et ESPMARK, 1982). La conversion à l'échelle Mel a été limitée aux fréquences supérieures à 100 Hz pour éliminer les bruits basse fréquence.

#### 5.2.4 Distance entre les cris : DFTW

Chaque paire de spectrogrammes a ensuite été comparée par déformation dynamique spectro-temporelle (DFTW). La distance DFTW consiste à trouver l'alignement entre les deux spectrogrammes qui minimise la distance, d'abord le long de l'axe temporel, puis le long de l'axe fréquentiel. Cela permet de tenir compte des différences de vitesse de restitution, ainsi que des différences de fréquence systématiques (par exemple, si un cri est décalé vers des fréquences aiguës par rapport à un autre). La matrice de distance par paire qui en résulte a ensuite été utilisée pour évaluer la similarité entre les femelles individuelles, les groupes et les colonies.

#### 5.2.5 Projection UMAP

À des fins de visualisation, nous avons projeté la matrice de distance DFTW en deux dimensions avec UMAP (MCINNES et al., 2018). UMAP vise à préserver les voisinages entre les points (i.e. les points initialement proches selon la distance DFTW devraient rester proches dans la projection), mais peut modifier la distance en rapprochant les points similaires et en éloignant les points dissimilaires. Nous avons utilisé UMAP avec les paramètres par défaut du package *umap-learn* (v0.5.3), à l'exception du réglage de la distance minimale entre les points à 0 (permettant ainsi aux points de se chevaucher dans la projection s'ils sont suffisamment similaires) et n'avons pas fait varier les paramètres plus avant.

#### 5.2.6 Analyses statistiques

Nous avons testé si des signatures vocales pouvaient être détectées dans les cris au nid en deux étapes : d'abord en vérifiant si les distances acoustiques entre vocalisations étaient corrélées avec la distance physique entre leurs sources respectives, puis en évaluant la précision d'un algorithme de classification sur les données.

Nous avons comparé les distances DFTW en utilisant un modèle linéaire mixte généralisé (GLMM) pour tenir compte des pseudo-répliques dues à l'échantillonnage de plusieurs cris pour chaque individu, groupe et colonie, avec une famille Gamma et une fonction de lien log. En raison de la manière dont les données ont été collectées, nous avons combiné les variables d'identité individuelle, d'identité de groupe et d'identité de colonie pour chaque cri en une variable "Condition" avec six niveaux possibles : 1) distance entre les vocalisations au sein d'une femelle individuelle, 2) entre les individus, mais au sein du même groupe, 3) entre les groupes, mais au sein de la même colonie, 4) entre colonies géographiquement proches (i.e. en comparant les colonies françaises entre elles et les colonies anglaises entre elles), 5) entre des colonies géographiquement éloignées (c'est-à-dire en comparant chaque colonie française avec chaque colonie britannique ; ces colonies auraient très peu de chances

d'entrer en contact) et 6) entre vocalisations de femelles non-identifiées, mais au sein de leurs colonies respectives.

Le GLMM a été entraîné avec la distance DFTW en variable réponse, Condition en variable indépendante et la dyade en facteur aléatoire (i.e. la combinaison spécifique d'individus, de groupes et de colonies comparés dans chaque cas), en utilisant le package *lme4* (v1.1-30, BATES et al., 2015) de R (v4.1.1, R CORE TEAM, 2021). Nous avons d'abord comparé le modèle à un modèle nul (avec uniquement la dyade comme facteur aléatoire), puis nous avons effectué des comparaisons par paires entre les niveaux de Condition en utilisant le package *emmeans* (v1.8.0, LENTH, 2022), en appliquant une correction de Tukey pour les tests multiples aux p-valeurs obtenues.

Nous avons effectué la classification en utilisant l'algorithme de k-Nearest Neighbours (*kNN*) du package *scikit-learn* (v1.1.2) de Python (v3.8, VAN ROSSUM et DRAKE, 2009). La précision de la classification des individus, des groupes, et des colonies a été quantifiée et comparée à une distribution aléatoire des identités en utilisant le kappa de Cohen, pour différentes valeurs de *k* et pour les colonies, les groupes par colonie et les individus par colonie. Pour éviter de biaiser les résultats, les vocalisations non attribuées à des individus identifiés ont été exclues de cette classification.

Toutes les analyses ont également été effectuées avec la distance euclidienne entre les points projetés avec UMAP. Pour la classification, les cris des femelles non identifiées ont été exclus avant l'UMAP.

### 5.3 Résultats

Nous avons comparé les distances acoustiques par paire entre les niveaux de distance physique. Le modèle incluant la distance physique est significativement meilleur que le modèle nul (likelihood ratio test,  $\chi^2(df=5) = 328,18$ ,  $p < 0.001$ ). Les comparaisons par paire ont révélé des effets fortement significatifs pour la plupart des comparaisons (Fig.5.1, Table 5.2), correspondant à une augmentation de la distance acoustique avec la distance physique. En particulier : les distances intra-individuelles sont significativement plus faibles que toutes les autres distances (e.g. comparées aux distances inter-individuelles mais intra-groupe : ratio  $0,89 \pm 0,01$ ,  $t = -8,86$ ,  $p < 0,001$ ), les distances intra-groupe sont plus faibles que les distances inter-colonies (e.g. comparées aux distances entre colonies géographiquement proches : ratio  $0,96 \pm 0,01$ ,  $t = -4,61$ ,  $p < 0,001$ ), et les distances inter-groupes intra-colonie sont également plus faibles que les distances inter-colonies (e.g. par rapport aux colonies géographiquement proches : ratio  $0,95 \pm 0,01$ ,  $t = -12,86$ ,  $p < 0,001$ ). Au sein des colonies, les distances intra-et inter-groupes ne diffèrent pas (ratio  $1,01 \pm 0,01$ ,  $t = 1,77$ ,  $p = 0,49$ ). Entre colonies, les distances entre colonies proches et éloignées ne diffèrent pas non plus (ratio  $0,98 \pm 0,01$ ,  $t = -2,39$ ,  $p = 0,16$ ). Il convient toutefois de noter que les tailles d'effet sont faibles, à l'exception de celles associées avec les distances intra-individuelles. Les autres comparaisons ne diffèrent que peu en taille d'effet, qu'elles soient significatives ou non. Lorsque l'identité de l'individu est inconnue, les résultats sont intermédiaires entre les distances inter-individuelles intra-groupe et les distances inter-groupes intra-colonie (Fig. 5.1).

Les distances dans la projection UMAP suivent la même distribution générale, mais exacerbée par

TABLE 5.2 – Contrastes de distances acoustiques entre les différents niveaux de distance physique, en fonction de la distance utilisée (DFTW : distance acoustique entre les spectrogrammes calculée par DFTW, UMAP : distance euclidienne dans la projection UMAP). Chaque ligne correspond à une comparaison : par exemple, la première ligne compare les distances intra-individuelles (Intra-individu, Intra-groupe, Intra-colonie) aux distances inter-individuelles au sein d'un même groupe d'une colonie (Inter-individu, Intra-groupe, Intra-colonie). Les quatre colonnes à droite représentent les contrastes estimés : ratio (entre les deux niveaux de distance comparés), SE (erreur standard sur le ratio), t (statistique de test), p (p-valeur après correction de Tukey).

Distance	Comparaison						Ratio	SE	t	p
	Individu	Groupe	Colonie	Individu	Groupe	Colonie				
DFTW	Intra	Intra	Intra	Inter	Intra	Intra	0.89	0.01	-8.86	< 0.001
	Intra	Intra	Intra	Inter	Inter	Intra	0.91	< 0.01	-9.31	< 0.001
	Intra	Intra	Intra	Inter	Inter	Proche	0.86	< 0.01	-14.44	< 0.001
	Intra	Intra	Intra	Inter	Inter	Distante	0.84	0.01	-12.89	< 0.001
	Intra	Intra	Intra	Inconnu	Inconnu	Intra	0.88	0.02	-5.20	< 0.001
	Inter	Intra	Intra	Inter	Inter	Intra	1.01	< 0.01	1.77	0.49
	Inter	Intra	Intra	Inter	Inter	Proche	0.96	< 0.01	-4.61	< 0.001
	Inter	Intra	Intra	Inter	Inter	Distante	0.94	0.01	-5.12	< 0.001
	Inter	Intra	Intra	Inconnu	Inconnu	Intra	0.98	0.02	-0.66	0.99
	Inter	Inter	Intra	Inter	Inter	Proche	0.95	< 0.01	-12.68	< 0.001
	Inter	Inter	Intra	Inter	Inter	Distante	0.93	< 0.01	-8.00	< 0.001
	Inter	Inter	Intra	Inconnu	Inconnu	Intra	0.97	0.02	-1.35	0.76
	Inter	Inter	Proche	Inter	Inter	Distante	0.98	< 0.01	-2.39	0.16
	Inter	Inter	Proche	Inconnu	Inconnu	Intra	1.02	0.02	0.99	0.92
Inter	Inter	Distante	Inconnu	Inconnu	Intra	1.05	0.03	1.85	0.44	
UMAP	Intra	Intra	Intra	Inter	Intra	Intra	0.53	0.02	-16.73	< 0.001
	Intra	Intra	Intra	Inter	Inter	Intra	0.53	0.02	-20.57	< 0.001
	Intra	Intra	Intra	Inter	Inter	Proche	0.42	0.01	-28.67	< 0.001
	Intra	Intra	Intra	Inter	Inter	Distante	0.33	0.01	-32.03	< 0.001
	Intra	Intra	Intra	Inconnu	Inconnu	Intra	0.46	0.03	-13.99	< 0.001
	Inter	Intra	Intra	Inter	Inter	Intra	1.01	0.03	0.44	0.99
	Inter	Intra	Intra	Inter	Inter	Proche	0.81	0.02	-7.3	< 0.001
	Inter	Intra	Intra	Inter	Inter	Distante	0.63	0.02	-13.45	< 0.001
	Inter	Intra	Intra	Inconnu	Inconnu	Intra	0.88	0.05	-2.37	0.17
	Inter	Inter	Intra	Inter	Inter	Proche	0.79	0.01	-14.91	< 0.001
	Inter	Inter	Intra	Inter	Inter	Distante	0.62	0.01	-20.36	< 0.001
	Inter	Inter	Intra	Inconnu	Inconnu	Intra	0.87	0.04	-2.95	0.04
	Inter	Inter	Proche	Inter	Inter	Distante	0.78	0.02	-10.93	< 0.001
	Inter	Inter	Proche	Inconnu	Inconnu	Intra	1.09	0.05	1.88	0.42
Inter	Inter	Distante	Inconnu	Inconnu	Intra	1.40	0.07	6.69	< 0.001	



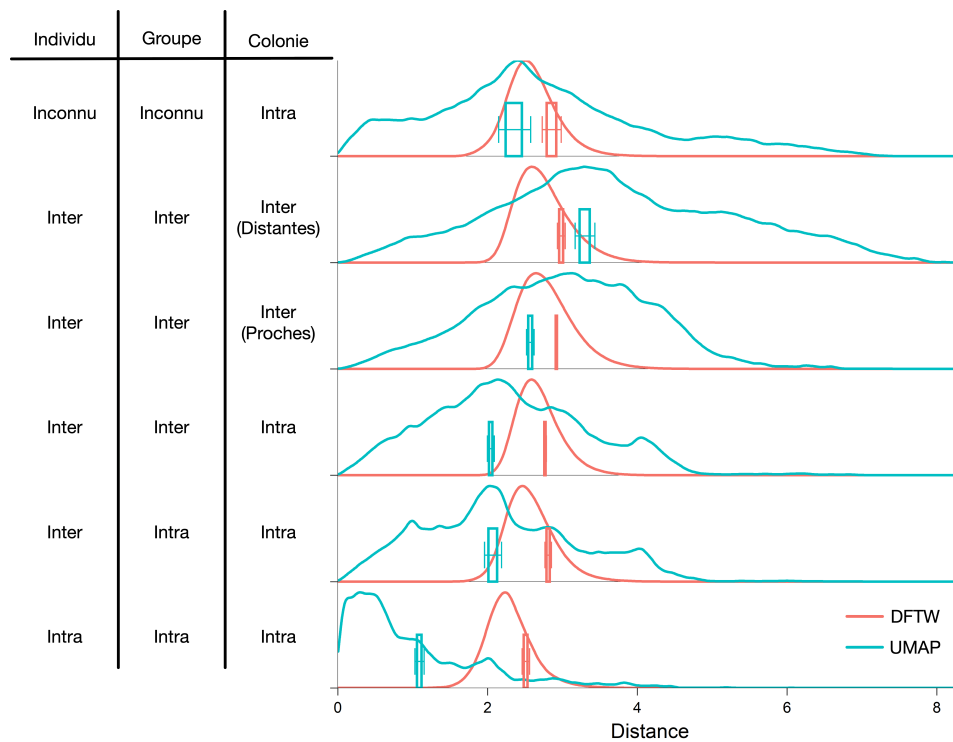


FIGURE 5.1 – Distribution des distances acoustiques par niveau de distance physique. Chaque courbe représente la densité des distances acoustiques, dans la distance DFTW d’origine (rouge) ou dans l’UMAP (bleu), normalisée à la même valeur maximum. Les boîtes correspondent aux paramètres estimés par le GLMM, pour chaque distance : le point correspond à la valeur du paramètre, la boîte correspond à la moyenne  $\pm$  erreur standard, et les lignes correspondent à l’intervalle de confiance à 95%.

l’algorithme (i.e. les vocalisations d’une même femelle sont rapprochées, les vocalisations des colonies distantes sont éloignées). Le modèle est significativement meilleur que le modèle nul, comme pour le modèle avec la DFTW (likelihood ratio test,  $\chi^2(df=5) = 727.38$ ,  $p < 0.001$ ). Les comparaisons par paires ont donné sensiblement les mêmes résultats que pour la DFTW, mais avec des tailles d’effets supérieures dues à l’UMAP (Table 5.2). Cette augmentation des tailles d’effets a cependant suffi à rendre significative la comparaison entre colonies proches et éloignées (ratio proche/distante :  $0,78 \pm 0,02$ ,  $t = -10,93$ ,  $p < 0,001$ ).

L’algorithme  $kNN$  a prédit avec précision l’identité des individus, des groupes, et des colonies (Fig. 5.2), pour de faibles valeurs de  $k$  (i.e. en ne prenant en compte que les cris les plus similaires), pour les deux distances utilisées (Fig. 5.2A). Cette précision chute cependant très rapidement à des valeurs de  $k$  plus élevées, d’abord pour les individus, puis pour les groupes, mais les colonies restent mieux classifiées que le hasard pour les valeurs de  $k$  considérées. Cette chute s’explique principalement par l’inclusion progressive de cris de plus en plus distants. Par exemple, pour la classification individuelle, certaines femelles ne représentent qu’une petite partie du jeu de données, expliquant la forte chute au-delà de  $k = 20$  environ : pour ces femelles avec moins de  $k$  cris, cela implique nécessairement d’inclure des cris d’autres femelles, même si ces femelles sont distantes. Le même raisonnement est valable pour les groupes, mais aucune colonie n’incluait moins de 250 cris (la valeur maximum de  $k$  considérée), d’où un maintien de la précision de l’algorithme.

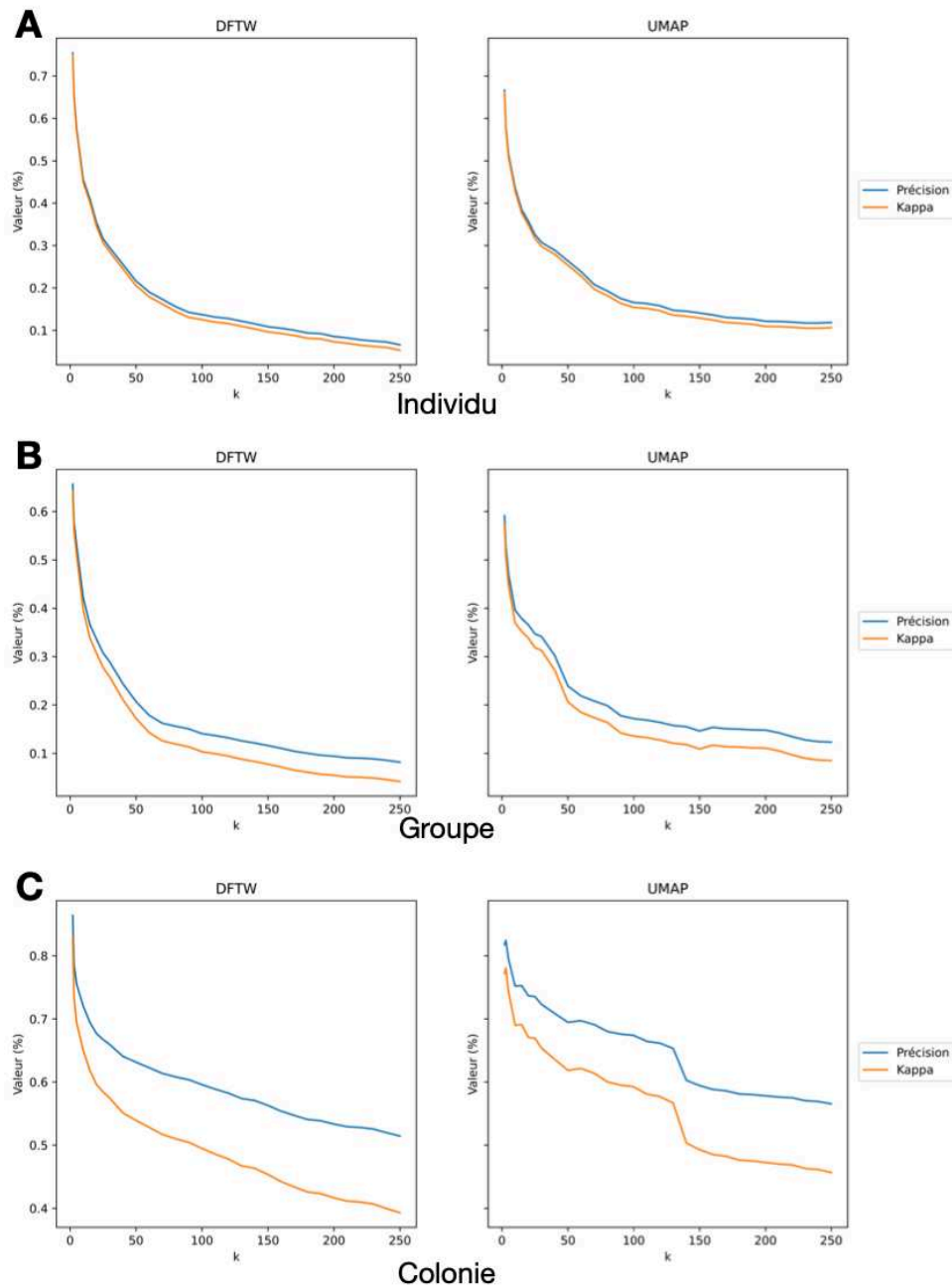


FIGURE 5.2 – Classification par kNN des cris sur la DFTW (gauche) ou la distance dans la projection UMAP, par A) Individu, B) Groupe et C) Colonie. Chaque courbe représente la précision de l’algorithme, pondérée par le nombre de cris de chaque individu, groupe, ou colonie respectivement (pour éviter une influence importante des individus avec un grand nombre de cris). Le kappa de Cohen (courbe orange) représente à quel point la précision obtenue est meilleure que celle correspondant au hasard.

Cependant, la visualisation de la projection UMAP montre que les colonies se recouvrent dans une large mesure (Fig. 5.3A), tout comme les groupes au sein de chaque colonie (Fig. 5.3B). Il ne semble donc pas y avoir de signature vocale de groupe ou de colonie, malgré les résultats du *kNN*. Les individus forment en effet pour la plupart des clusters séparés (Fig. 5.3C). Ce regroupement des individus a probablement induit une inertie dans la classification qui a permis à l'algorithme *kNN* de retrouver les groupes et les colonies, sans que cela corresponde à une similarité dans les vocalisations à ces niveaux.

## 5.4 Discussion

Nous avons montré que des signatures individuelles existent dans les cris au nid des corbeaux freux femelles. La classification *kNN* a correctement identifié les individus, les groupes et les colonies bien au-delà du niveau de chance. Cependant, la projection UMAP a révélé que cela était probablement dû au fait que les cris d'une même femelle étaient groupés ensemble à l'écart de ceux des autres femelles. Les groupes et les colonies se recouvraient, au lieu d'occuper des régions distinctes de la projection UMAP. Par conséquent, il n'y avait aucune preuve de signature de groupe ou de colonie. Les distances intra- et inter-colonies étaient très similaires, indiquant une absence de similarité vocale liée à la géographie. Plusieurs éléments de discussion découlent directement de ces résultats, concernant la technique utilisée, les contraintes potentielles associées à la collecte de données en extérieur, et la signification de ces résultats pour la compréhension de la socialité des corbeaux freux.

Nous avons observé qu'il était possible de détecter des signatures individuelles dans le cri au nid des femelles. La signature individuelle avait déjà été détectée dans le cri le plus fréquent de certains mâles (BENTI et al., 2019), bien qu'une analyse ultérieure (Chapitre 4) ait indiqué qu'au-delà d'une signature vocale individuelle, les mâles produisent des cris extrêmement variés entre individus, au point de remettre en question la notion de répertoire vocal de l'espèce (R. F. LACHLAN et al., 2010). Cependant, la même étude a montré que les femelles partagent plusieurs types de cris, indépendamment de leur colonie d'origine, dont le cri au nid. Celui-ci est émis uniquement par les femelles lorsqu'elles restent au nid plusieurs semaines pendant la reproduction, et dépendent entièrement du mâle pour les nourrir et nourrir les juvéniles (COOMBS, 1960, 1961b)). Étant donné son association forte avec la reproduction, ce cri a probablement une forte fonctionnalité sociale, et il n'est donc pas surprenant d'y trouver une signature vocale individuelle. Les corbeaux freux peuvent vivre en colonies allant de quelques nids à des centaines voire des milliers de couples reproducteurs (CLAYTON et EMERY, 2007; MARSHALL et COOMBS, 1957; PATTERSON et al., 1971; SWINGLAND, 1977; SWINGLAND, 1973), et une reconnaissance acoustique rapide et précise peut être cruciale à cette période de l'année. Ce cri est également l'un des plus puissants émis par les femelles adultes, ce qui permet probablement la communication entre partenaires sur de longues distances et dans des conditions sonores aussi bruyantes qu'au sein de la colonie, où toutes les femelles peuvent potentiellement vocaliser en même temps. Cette puissance facilite également grandement les enregistrements, tandis que l'immobilité des femelles au nid permet de les identifier malgré le nombre d'individus dans la colonie, et sans nécessiter de procédures invasives pour les marquer. Ce dernier point, couplé à la pré-

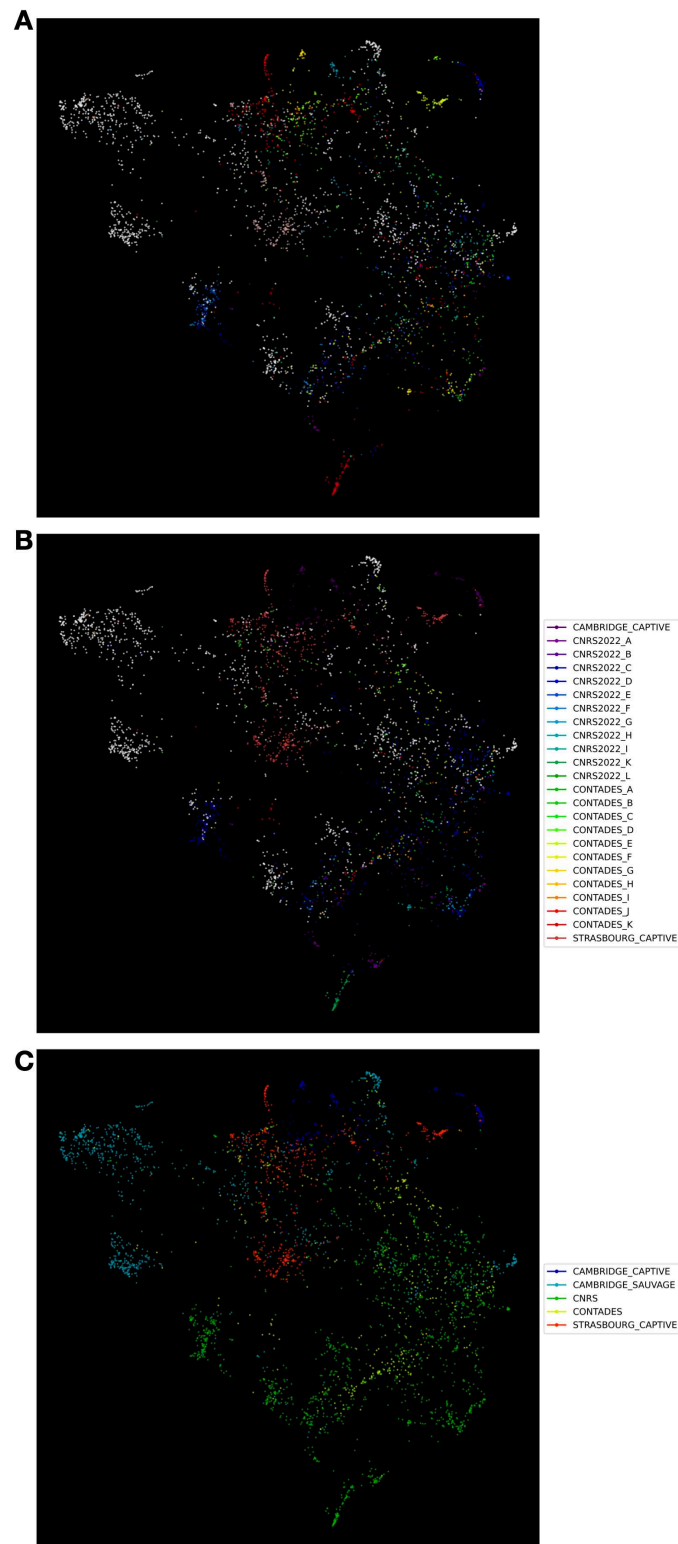


FIGURE 5.3 – Projection UMAP des cris au nid. A) Colorée par individu. B) Même projection, colorée par groupe. C) Même projection, colorée par colonie. La légende n’indique pas les individus pour ne pas encombrer l’espace. Les individus inconnus, et les groupes correspondants, dans le cas de Cambridge\_Sauvage, sont représentés par les points blancs.

sence à des périodes très prévisibles de l'année, permet d'installer des dispositifs de suivi à proximité des nids sans risque de stresser les animaux, simplement en plaçant les microphones avant l'arrivée des individus après l'hiver (BUSSE, 1965).

La prévisibilité de la présence des corbeaux freux sur un site donné est également une excellente opportunité pour étudier la fidélité des individus vis-à-vis de la colonie et de l'emplacement du nid, la stabilité de la composition du groupe, et potentiellement la stabilité de la colonie au cours du temps. L'hypothèse a été proposée d'une stabilité des colonies au cours du temps, avec une réutilisation des nids par les mêmes couples (GOODWIN, 1955; RICHARDSON et al., 1979). Nous nous attendions également à trouver une signature coloniale de plus ou moins prononcée en fonction de la sédentarité des populations, comme pour la population britannique (PATTERSON et al., 1971). Cependant, aucune signature coloniale n'a pu être trouvée dans cette étude. Cela peut néanmoins être dû à plusieurs facteurs, liés à l'écologie de l'espèce, ou à des facteurs extérieurs. D'une part, l'étude décrivant la population britannique comme sédentaire a été faite il y a environ 50 ans, et tant les populations que les conditions du milieu ont pu évoluer depuis. Les corvidés font preuve d'une forte flexibilité comportementale et sociale en fonction des conditions (BAGLIONE et al., 2002), et cela pourrait conduire des populations sédentaires à migrer ou des populations migrantes à devenir sédentaires. D'autre part, des interventions humaines ont également pu impacter les populations. Les corbeaux freux, comme beaucoup d'autres corvidés, ont longtemps été considérés comme une espèce nuisible, en raison de dommages sur les cultures agricoles causées par leur régime alimentaire en partie basé sur les graines (FEARE et al., 1974; FEARE, 1978; PORTER et al., 2008), et ont donc été la cible de nombreuses campagnes de destruction, en particulier en Europe occidentale continentale, avec environ 4 millions d'oiseaux tués chaque année (JIGUET, 2020). Enfin, si le cri au nid permet à la femelle d'appeler le mâle plutôt que d'interagir avec les autres individus de la colonie, il n'y a pas forcément de mécanisme favorisant la convergence vocale. Au contraire, une ressemblance entre plusieurs femelles pourrait conduire le mâle à nourrir la mauvaise femelle en cas d'erreur (ROSKAFT, 1983).

Les dialectes, la convergence vocale, ou le partage vocal sont connus pour avoir plusieurs origines, comme l'environnement géographique et social des oiseaux. Dès l'éclosion, et peut-être même dans l'œuf, les corbeaux freux juvéniles sont déjà exposés à l'environnement vocal de leur colonie. Si l'acquisition des cris est influencée par cet environnement social, comme dans les autres cas d'apprentissage vocal des juvéniles (ADRET-HAUSBERGER, 1982; BEECHER et BRENOWITZ, 2005; HAMMERSCHMIDT et FISCHER, 2008; KROODSMA, 2004; MARLER, 1997; MARLER et PETERS, 1987; TEN CATE, 2021), alors les individus devraient se ressembler vocalement, simplement parce qu'ils apprennent des mêmes sources. De la même façon, si les juvéniles rejoignent des groupes sociaux ensuite, ils pourraient également apprendre les mêmes vocalisations par imitation des congénères du groupe (SALINAS-MELGOZA et WRIGHT, 2012). Et en particulier, si les corbeaux freux reviennent effectivement chaque année sur le même site de reproduction, ces similitudes devraient être renforcées par exposition répétée. Cependant, les corbeaux freux se dispersent souvent après la saison de reproduction, et sont exposés même dans les colonies à des individus, soit de la même région, soit des individus migrants (BUSSE, 1969; PATTERSON et al., 1971). Étant donné que les corbeaux

freux, comme de nombreux autres corvidés, sont capables de mimétisme vocal même à l'âge adulte (E. D. BROWN, 1985 ; BROWN et FARABAUGH, 1997 ; KAPLAN, 1999)), la convergence vocale à l'âge adulte ne peut être exclue *a priori*. Cependant, il s'agit ici de considérations très générales. Selon la fonction exacte d'une vocalisation, le mimétisme peut être impliqué ou non. Par exemple, si la vocalisation est liée à l'identification individuelle, un maintien des différences individuelles pourrait être attendu.

Un partage vocal faible ou absent entre partenaires sociaux a été décrit chez plusieurs espèces territoriales, comme le pinson des arbres (*Fringilla coelebs*, SLATER et INCE, 1982) et le viréo aux yeux blancs (*Vireo griseus*, BORROR, 1987 ; BRADLEY, 1981). Le pinson des arbres ne montre aucune tendance particulière au partage entre voisins par rapport aux individus éloignés, tandis que le viréo aux yeux blancs montre un évitement vocal, les individus éloignés étant plus susceptibles de partager des vocalisations par rapport aux voisins. Dans les deux cas, cela peut s'expliquer par la dispersion des juvéniles suivie de la formation d'un territoire, où le pinson des arbres trouve un territoire au hasard et où le viréo aux yeux blancs trouve un territoire en évitant les oiseaux partageant ses vocalisations. Cet évitement pourrait avoir des fonctions territoriales, mais aussi d'évitement de la consanguinité (CATCHPOLE et SLATER, 2008). Cependant, ces espèces sont territoriales, et non des reproducteurs coloniaux comme le corbeau freux.

Nous pensons utile de revenir sur le fait que les colonies n'ont pas été séparées les unes des autres par l'UMAP dans cette étude. En effet, la séparation des colonies, si elle avait été observée, aurait pu avoir deux causes : une véritable signature vocale coloniale, ou des différences d'environnement sonore. En effet, les colonies sauvages françaises étaient dans des environnements urbains, alors que les colonies captives et la colonie sauvage anglaise étaient dans des environnements plus naturels. Par ailleurs, si les microphones ont pu être placés très proches des colonies captives et de la colonie sauvage du campus CNRS de Cronenbourg à Strasbourg, cela n'a pas été possible pour les colonies sauvages du parc des Contades, ainsi que pour la colonie sauvage de Madingley, qui ont été enregistrées de beaucoup plus loin en raison d'enregistrements plus tardifs dans la saison de reproduction. Ces environnements sonores différents auraient pu influencer les résultats. Cependant, le pré-traitement des données (en particulier le débruitage des spectrogrammes), l'utilisation de la DFTW, et dans une certaine mesure de l'UMAP, tendent à éliminer les variations aléatoires dues au bruit dans les données (SAINBURG et GENTNER, 2021). Il s'agit d'une première étape dans la validation d'une procédure de suivi individuel chez cette espèce.

La distance DFTW était explicitement conçue pour compenser les différences de distribution de fréquence qui sont à la base des paramètres acoustiques utilisées dans la plupart des études sur la signature vocale (BENTI et al., 2019 ; BOECKLE et al., 2012 ; ELIE et THEUNISSEN, 2018 ; KEENAN et al., 2020 ; LINHART et al., 2019 ; MATES et al., 2015 ; YORZINSKI et al., 2006). Les vocalisations des différents individus ont néanmoins pu être séparées, ce qui implique que des variations dans la forme spectrale des vocalisations existent. La distribution des fréquences (en particulier, celle de la fréquence fondamentale d'une vocalisation) est souvent liée à des contraintes physiologiques, comme la taille ou la masse corporelle (EY et al., 2007 ; FAVARO et al., 2017 ; FITCH et HAUSER, 1995 ; MIYAZAKI et WAAS, 2003 ; TAYLOR et REBY, 2010). Cependant, une signature vocale plus complexe

pourrait être nécessaire dans un environnement social complexe (SMITH-VIDAURRE et al., 2021), et des variations purement physiologiques pourraient ne simplement pas offrir suffisamment de libertés pour reconnaître des individus dans un environnement aussi complexe que la colonie de reproduction.

De nouveaux outils d'analyse bioacoustique apparaissent dans la littérature, avec des approches puissantes basées sur des algorithmes d'apprentissage automatique (SAINBURG et GENTNER, 2021). Ces nouvelles approches, avec des paramètres plus flexibles et une plus grande généralité d'application, comme la prise en compte de spectrogrammes entiers plutôt que de mesures acoustiques spécifiques, peuvent permettre des études beaucoup plus fines de la similarité et de l'apprentissage vocaux dans la nature. Les approches plus classiques, reposant sur les mesures acoustiques (e.g. durée du cri, fréquence fondamentale, distribution de l'énergie en temps ou en fréquence) et des modèles relativement simples comme l'analyse de fonctions discriminantes (BENTI et al., 2019; KEENAN et al., 2020), sont difficiles à appliquer de façon comparative. En effet, les paramètres acoustiques ne sont pas toujours accessibles selon les espèces, et sont intrinsèquement biaisés par la perception humaine (SAINBURG et GENTNER, 2021). Les corbeaux freux, comme beaucoup d'autres corvidés, produisent des vocalisations chaotiques (FLETCHER, 2000; STOWELL et al., 2016a), dans lesquelles des mesures communes comme la fréquence fondamentale ou toutes mesures basées sur la distribution harmonique des fréquences sont extrêmement difficiles à mesurer, même dans les meilleures conditions d'enregistrement (BENTI et al., 2019; STOWELL et al., 2016a). Les approches basées sur le spectrogramme, comme la nôtre, peuvent contourner cette limitation et ne nécessitent qu'une bonne qualité d'enregistrement pour extraire des informations. En tant que telles, elles pourraient être déployées dans des scénarios de surveillance de la faune ou pour suivre la dynamique des populations, en complément des campagnes actuelles de suivi GPS (FLACK et al., 2018; PFEIFFER et MEYBURG, 2015), ou en remplacement des techniques de capture-recapture. Le suivi acoustique évite en effet les manipulations invasives, et avec les nouvelles technologies, peut être effectué à des coûts très réduits (DARRAS et al., 2019). L'enregistrement des colonies et le suivi des individus par leur signature vocale pourraient permettre de répondre à des questions sur le comportement colonial et la dynamique sociale chez les oiseaux, et plus généralement cela devrait soutenir les investigations sur le lien entre socialité et comportement vocal chez des espèces socialement complexes comme le corbeau freux.

# Chapitre 6

## Les corbeaux freux peuvent-ils se synchroniser à un stimulus rythmique ?

Killian Martin, Maëlan Tomasek, Agnès Hivet, Andrea Ravignani, Olivier Adam, Nicolas Obin, Valérie Dufour

En préparation pour soumission à *Biology Letters* sous le titre

Can rooks (*Corvus frugilegus*) spontaneously entrain to rhythmic stimuli ?

### Résumé

La capacité à percevoir un rythme et à s'y synchroniser semble rare chez les animaux non-humains. De très nombreuses espèces sont capables de se synchroniser avec leurs congénères pour produire des duos ou des chœurs vocaux, où chaque individu ne vocalise que certains moments précis. Pourtant, seules quelques espèces ont démontré la capacité à percevoir des rythmes extérieurs et à s'y synchroniser (phénomène d'entraînement), et seulement par des mouvements corporels. Par ailleurs, ces études ont nécessité un renforcement positif pour que les individus parviennent à cette synchronisation. Cette étude examine la capacité des corbeaux freux à joindre leur chant à un stimulus rythmique, et à le synchroniser aux rythmes proposés. Nous montrons que les corbeaux freux commencent parfois spontanément à chanter lorsqu'ils entendent un stimulus externe rythmé, mais qu'ils n'ajustent pas précisément le rythme de leurs vocalisations au rythme entendu. Néanmoins, certains oiseaux ont émis des vocalisations plus courtes lorsque le tempo augmentait, ce qui suggère une tentative d'ajustement. Bien que cet ajustement rythmique soit imprécis, les corbeaux semblent posséder un certain contrôle sur le timing de leur production vocale. Cela indique que les corbeaux sont à la fois cognitivement et vocalement flexibles, et qu'ils représentent une espèce intéressante pour de futures études sur la perception rythmique. Ces prochains travaux pourraient contribuer à faire la lumière sur les origines des capacités musicales de l'espèce humaine.

### 6.1 Introduction

Les animaux sociaux et grégaires adaptent souvent leur comportement à celui de leurs congénères en synchronisant leurs mouvements, leur alimentation ou leurs activités de repos. Ces comportements collectifs sont cruciaux car ils contribuent à maintenir la cohésion sociale du groupe, avec des bénéfices sur la protection contre les prédateurs, l'accès aux ressources du milieu, et la reproduction (KREBS et al., 1978). Il est intéressant de noter que la façon dont les individus se synchronisent aux autres est très diverse, et peut aller d'une co-occurrence aléatoire à une synchronisation précise (RAVIGNANI et NORTON, 2017). Par exemple, alors que les chimpanzés produisent des chœurs



dans lesquels le groupe vocalise ensemble, souvent de manière peu coordonnée (GHIGLIERI, 1984; MERKER et al., 2009), les humains peuvent maîtriser une synchronisation précise lorsqu'ils jouent dans un orchestre. L'homme n'est pas la seule espèce capable d'entraîner à un rythme (dans ce cas particulier, "entraîner" signifie "synchroniser des mouvements corporels ou des sons à un stimulus externe", à distinguer dans la suite du "conditionnement", où un animal apprend à effectuer une action à l'aide d'un renforcement positif), et certaines autres espèces font preuve d'une synchronisation tout aussi précise dans leur comportement. Par exemple, les lucioles peuvent clignoter ensemble avec une très grande précision (BUCK, 1988), les grenouilles forment des chœurs synchronisés et adaptent leurs cris à ceux des autres individus (WELLS, 1977; pour une revue générale des comportements de chœurs, voir RAVIGNANI et al., 2014), les dauphins peuvent percer la surface à quelques millisecondes près des autres membres du groupe (CONNOR et al., 2006; MCCUE et al., 2020), les crabes violonistes finissent souvent par agiter leurs pinces en synchronisation (BACKWELL et al., 1998). Une proposition récente est que la capacité à entraîner pourrait trouver son origine dans les oscillations neuronales, et pourrait donc être basée sur un circuit neuronal de base conservé tout au long de l'évolution animale (ROUSE et al., 2016; WILSON et COOK, 2016). Comme le soulignent WILSON et COOK (2016), la question n'est donc pas "qui ou quelles espèces peuvent entraîner", mais plutôt "pourquoi certaines espèces n'entraînent pas toujours".

Une possibilité est qu'à mesure que la complexité cognitive d'une espèce augmente, la capacité à entraîner à un stimulus externe dépend moins d'une réponse automatique au stimulus sensoriel et plus d'une flexibilité comportementale tout au long de la synchronisation. Cette flexibilité pourrait être médiée par un contrôle volontaire, qui implique la capacité de prêter attention à un stimulus plutôt qu'à un autre, de commencer et d'arrêter volontairement un comportement, et d'entraîner même à un stimulus non biologiquement pertinent (RAVIGNANI et al., 2014; WILSON et COOK, 2016). La compétence cognitive nécessaire pour entraîner volontairement pourrait n'être présente que chez quelques espèces, ce qui pourrait expliquer pourquoi les preuves expérimentales d'un entraînement précis à des stimuli non biologiquement pertinents restent rares (WILSON et COOK, 2016). Parmi les espèces qui peuvent entraîner, certaines sont capables d'extraire le rythme de la musique et d'adapter leurs mouvements à son tempo, comme cela a été rapporté chez certaines espèces de perroquets (Snowball, PATEL et al., 2009), chez les perruches (HASEGAWA et al., 2011; SEKI et TOMYTA, 2019) et chez l'otarie (Ronan, COOK et al., 2013; ROUSE et al., 2016). Bien que les macaques et les chimpanzés puissent également posséder certaines des capacités nécessaires à l'entraînement rythmique à des stimuli non biologiquement pertinents, ils semblent être plus sensibles aux stimuli rythmiques visuels (SELEZNEVA et al., 2013; TAKEYA et al., 2017; ZARCO et al., 2009). Cependant, dans la grande majorité des cas où un entraînement rythmique a été trouvé, les sujets avaient été conditionnés à le faire, i.e. leur comportement n'était pas naturel et spontané. La seule exception à cette règle est Snowball, un cacatoès à crête soufrée qui n'a pas été systématiquement conditionné, mais qui a néanmoins probablement été encouragé et récompensé par ses propriétaires les premières fois qu'il a hoché la tête en synchronisation avec un rythme musical (PATEL et al., 2009). Par conséquent, pour l'instant, toutes les occurrences connues d'entraînement rythmique chez les non-humains ont été au moins renforcées. Il est à noter que les cacatoès et les otaries pourraient réussir à entraîner par des prédispositions natu-

relles. En effet, les perroquets ont de fortes capacités d'imitation vocale (SCHACHNER, 2010), et les otaries sont de proches parents des éléphants de mer, dont on a récemment montré qu'ils percevaient et se souvenaient des spécificités rythmiques vocales des mâles voisins de leur colonie (MATHEVON et al., 2017).

Les performances de l'otarie Ronan montrent que l'apprentissage vocal et le mimétisme vocal ne sont peut-être pas une condition préalable nécessaire à la synchronisation avec d'autres stimuli non écologiques. Néanmoins, l'un ou l'autre pourraient être des facteurs favorables importants (SCHACHNER, 2010; SCHACHNER et al., 2009), surtout si l'on en croit la multitude de vidéos amateurs de perroquets domestiques sur Internet, qui pourraient être les seuls véritables cas d'entraînement chez des animaux non-humains (FITCH, 2009). Cependant, les preuves expérimentales d'entraînement sont souvent plus gestuelles que vocales, même chez les espèces capables de mimétisme vocal comme les perroquets. En plus du mimétisme vocal, de bonnes compétences en matière de cognition sociale pourraient jouer un rôle dans les compétences d'entraînement, comme cela a été observé dans la cognition rythmique humaine (RAVIGNANI et al., 2014; WILSON et COOK, 2016). Les corvidés sont des espèces à fortes compétences sociales et cognitives, notamment en ce qui concerne la théorie de l'esprit, c'est-à-dire la prise de perspective (BUGNYAR et al., 2016), et pourraient donc être de bonnes espèces modèles pour tester des hypothèses comparatives sur la synchronie et les origines évolutives de l'entraînement. En effet, bien que peu d'études décrivent le comportement vocal des corvidés à ce jour, ces espèces peuvent imiter les voix humaines de la même manière que les perroquets (rapporté de manière anecdotique chez des corvidés élevés par l'être humain, Dufour, communication personnelle). Les corvidés peuvent également produire des comportements de duo (KONDO et al., 2010; SEED et al., 2007) montrant qu'ils peuvent coordonner leur production vocale avec celle de leurs congénères. Les corbeaux freux produisent des chants, des séries de vocalisations variées, dans lesquels ils font preuve d'une flexibilité vocale prononcée en variant également l'organisation séquentielle des vocalisations d'un chant à l'autre (Martin et al. en prép., voir chapitre 4). Dans une étude précédente (Tomasek et al., sous presse), deux corbeaux freux ont vocalisé spontanément en entendant des bruits d'eau, commençant et arrêtant fréquemment leur chant dans les 3 secondes suivant le début et la fin du bruit, respectivement. Bien que ce résultat démontre qu'une espèce complexe socialement et cognitivement comme le corbeau freux est capable d'une synchronisation approximative, il ne s'agit pas d'une preuve convaincante d'entraînement au rythme, ni même d'un maintien du rythme.

La présente étude examine plus en détail les aptitudes des corbeaux freux à se joindre vocalement et à entraîner à un stimulus rythmique non biologiquement pertinent. Les oiseaux ont été exposés individuellement à des stimuli avec différents tempos et structures métriques. Chaque fois que cette exposition a conduit l'oiseau à chanter, nous avons enregistré ces chants et évalué si les oiseaux ont adapté leurs chants au tempo et/ou à la structure en cours. Nous faisons l'hypothèse que si les corbeaux freux peuvent entraîner à un rythme, ils devraient être capables de co-occurrence vocale. Ils devaient également prêter une attention particulière aux tempos et aux structures et entraîner au stimulus.

## 6.2 Matériel et Méthodes

### 6.2.1 Sujets d'études

Nous avons étudié un groupe de 11 corbeaux freux adultes captifs (8 mâles, 3 femelles), gardés dans une volière extérieure à Strasbourg (France). Tous les oiseaux ont été capturés à l'âge juvénile dans des colonies sauvages de la région de Strasbourg, et ont été gardés ensemble dans une même volière depuis leur capture. L'eau et la nourriture étaient fournies quotidiennement et accessibles *ad libitum*, y compris pendant les expériences. L'étude a été réalisée à deux périodes différentes : en septembre et octobre 2021 par KM, et entre février et juin 2022 par AH et les étudiants.

### 6.2.2 Création des stimuli rythmique

Les stimuli auditifs ont été créés à l'aide d'un script R. Pour éviter les effets possibles de l'habituation sur la motivation des corbeaux à chanter, nous avons créé une variété de stimuli de structures différentes. Chaque stimulus était un enregistrement de 2 min composé de six séquences rythmiques de 15 s, séparées par 5 s de silence (Fig. 6.1). Chaque séquence était composée d'une alternance de pulsations sonores et d'intervalles de silence. Les séquences pouvaient avoir un des quatre tempos différents (36, 60, 84 ou 120 bpm) et une des trois structures métriques différentes (isochrone, binaire, ternaire) différenciées par le rapport son/silence. La structure isochrone était composée de pulsations de durée égale séparées par des intervalles de silence selon un schéma 80-20 (par exemple, pour le tempo de 60 bpm, les pulsations duraient 0,8 s et les silences 0,2 s), la structure binaire alternait entre des pulsations longues (identiques au schéma isochrone) et courtes (selon un schéma 50-50), et la structure ternaire était composée de pulsations selon un schéma long-court-court (voir Fig. 6.1). Les séquences composant un stimulus étaient tirées de manière semi-aléatoire parmi les 12 combinaisons tempo-rythmiques possibles, avec la contrainte que les séquences successives devaient partager soit le tempo, soit la structure rythmique (par exemple, une séquence "36 bpm isochrone" pouvait être suivie d'une séquence "36 bpm binaire" ou d'une séquence "120 bpm isochrone", mais pas d'une séquence "120 bpm binaire", ni d'une autre séquence "36 bpm isochrone"). Les pulsations ont été extraites de deux courts clips de deux types de bruit différents : un robinet d'eau courante et un mini-drone. Chaque pulsation a été extraite en coupant le clip à la longueur appropriée, sans accélération ou ralentissement. Par ailleurs, les deux clips ont été normalisés avant extraction pour avoir la même amplitude maximale.

### 6.2.3 Protocole d'exposition au stimuli

Afin d'exposer sélectivement les individus au stimulus et d'éviter les interférences dues à l'exposition d'autres individus au même stimulus, six appareils de "douche sonore" ont été construits. Chaque appareil était composé d'un hémisphère rigide rembourré de mousse insonorisante. À l'intérieur de l'hémisphère se trouvait une structure conçue pour contenir un haut-parleur (Sony XRS-XB12 ou XRS-XB13), connecté à un ordinateur pour diffuser les stimuli. Le haut-parleur était orienté de manière à faire face au fond de l'hémisphère, afin que le son se réverbère dans une zone limitée sous

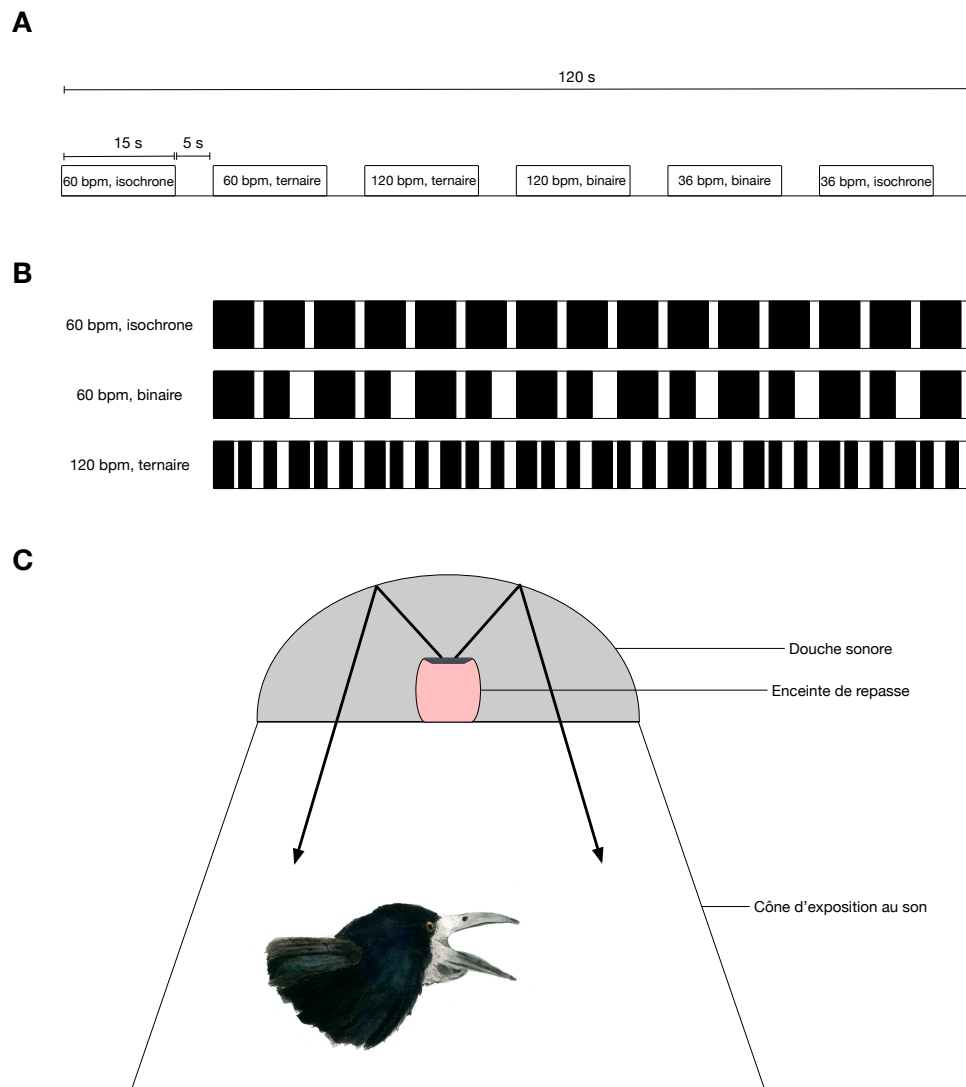


FIGURE 6.1 – Représentation schématique du protocole et des stimuli utilisés. A) Exemple de stimulus, montrant les séquences rythmiques et les durées respectives des séquences, silences et du stimulus entier. Les séquences (boîtes blanches) suivent les règles définies dans le texte : deux séquences successives peuvent différer soit en tempo, soit en structure métrique, mais pas les deux. B) Trois exemples de séquences de structures différentes. Les pulsations suivent la même échelle relative et les règles décrites dans le texte. Chaque boîte noire représente une pulsation sonore, séparée des autres pulsations par les intervalles de silence en blanc. C) Un corbeau freux sous la douche sonore est exposé au son. Le dispositif permet de restreindre le son à un cône au-delà duquel le son n'est pas entendu, notamment par les autres individus.

la douche, et ne soit pas audible par les personnes situées en dehors de cette zone limitée. Cet appareil permettait également d'enregistrer le son des oiseaux qui vocalisaient sans que le son de leurs vocalisations soit couvert par le son du stimulus. Chaque douche sonore était suspendue dans la volière à des endroits où les corbeaux se perchaient souvent. Pour permettre l'habituation des corbeaux freux au dispositif, les douches sonores ont été installées deux semaines avant la première partie de l'expérience et ont été laissées dans la volière par la suite jusqu'à la fin des expériences.

Les corbeaux ont pu venir et quitter librement les douches sonores sans intervention de l'expérimentateur (en particulier, sans récompense avant ou après une expérience). Lorsqu'un corbeau se trouvait sous la douche sonore, un stimulus choisi au hasard était diffusé et le comportement de l'oiseau était observé jusqu'à son départ (y compris si le stimulus se terminait avant le départ du corbeau). Un microphone directionnel (RØDE NTG8, connecté à un ordinateur avec une carte son Focusrite Scarlett 2i2) a été utilisé pour enregistrer les vocalisations de l'oiseau focal. Le stimulus et l'enregistrement étaient synchronisés au moment de l'expérience par l'utilisation de deux pistes Audacity pour la diffusion et l'enregistrement : une piste diffusait le stimulus, une autre enregistrait les sons reçus par le microphone.

#### 6.2.4 Acquisition des données

Les enregistrements correspondant aux réponses correctes (c'est-à-dire lorsqu'un individu vocalisait pendant que le stimulus était diffusé) ont été annotés manuellement dans Audacity par AH. Le début et la fin de chaque vocalisation ont été notés sur la base du spectrogramme, ce qui a permis de calculer les intervalles entre les vocalisations (IOI, défini comme l'intervalle de temps entre le début de deux vocalisations successives, RAVIGNANI et NORTON, 2017) et la durée de la vocalisation. Le délai entre le début de chaque vocalisation et le début de la dernière pulsation dans le stimulus a également été calculé.

#### 6.2.5 Analyses statistiques

Toutes les analyses ont été effectuées avec R (v4.1.1, R CORE TEAM, 2021), et étaient basées sur des modèles linéaires mixtes généralisés (GLMM), avec un modèle par oiseau individuel car nous nous attendions à ce que les différents oiseaux soient influencés par les stimuli de différentes manières. Les modèles ont été entraînés avec le package *lme4* (v1.1-30, BATES et al., 2015) avec une famille Gamma et une fonction de lien log. Un GLMM différent a été défini pour chacune des trois variables mesurées (délai, IOI, durée des vocalisations), avec la structure de la séquence (tempo, structure métrique, et interaction entre les deux) en variables indépendantes, et le stimulus entier en facteur aléatoire. À partir de ces modèles, nous avons calculé tous les sous-modèles possibles à l'aide de la fonction *dredge* du package *MuMIn* (v1.43.17, BARTON, 2020), et nous avons effectué une sélection des modèles sur la base du critère AICc : le meilleur modèle avait l'AICc le plus faible, et les modèles situés moins de 2 AICc au-dessus de ce modèle étaient considérés comme statistiquement équivalents (ANDERSON, 2004). Lorsque plusieurs sous-modèles étaient équivalents, nous avons considéré le plus parcimonieux. Lorsque celui-ci ne correspondait pas au modèle nul, nous avons effectué des tests

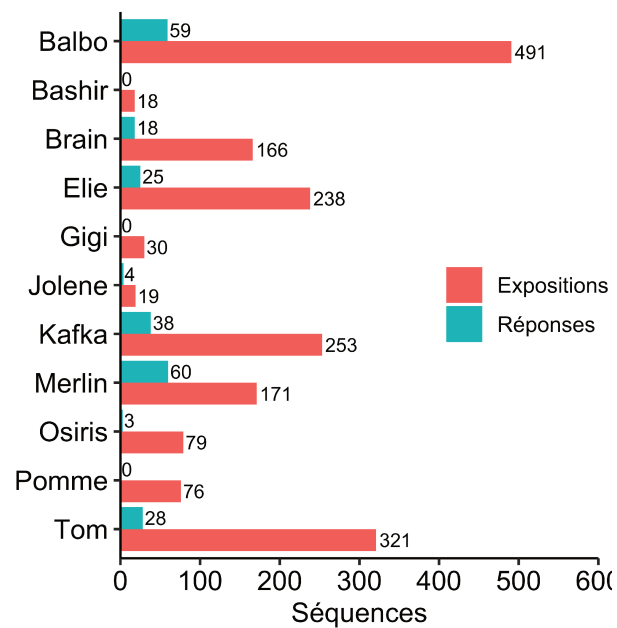


FIGURE 6.2 – Participation des individus à l'étude, par nombre de séquences auxquelles ils ont été exposés et ont répondu.

post-hoc par paire pour déterminer quels tempos et structures rythmiques étaient significativement différents. Ces tests post-hoc ont été réalisés à l'aide du package *emmeans* (v1.8.0, LENTH, 2022), avec p-valeurs ajustées par la méthode de Tukey pour les tests multiples.

## 6.3 Résultats

### 6.3.1 Nombre d'expositions et de réponses aux stimuli des individus

Chacun des onze corbeaux du groupe a été exposé au moins une fois à un stimulus, et huit ont chanté au moins une fois en réponse (Fig. 6.2). Compte tenu du nombre total de séquences auxquelles ils ont été exposés, la proportion de réponses à ces séquences (c'est-à-dire la production d'au moins deux vocalisations au cours d'une séquence) était de  $14,6 \pm 9,6\%$  par individu (moyenne  $\pm$  écart-type, plage 3,8 - 35,1%). Parmi les individus ayant répondu, deux (Osiris et Jolene) ont répondu très rarement aux stimuli (4 sur 73 séquences et 3 sur 19 séquences, respectivement), et ont donc été exclus de l'analyse.

### 6.3.2 Réponses préférentielles à certains stimuli

Nous avons évalué si certains individus sont plus susceptibles de répondre à des tempos ou des structures métriques particuliers (Fig. 6.3). Les proportions de réponses à chaque stimulus suggèrent que certains individus répondent préférentiellement à des séquences particulières (par exemple, Elie a répondu plus fréquemment au tempo de 120 bpm, indépendamment de la structure métrique, qu'à tout autre stimulus ; Merlin a répondu plus fréquemment aux séquences ternaires à 84 bpm). Cependant, ces préférences ne sont pas reflétées d'un point de vue statistique (test binomial

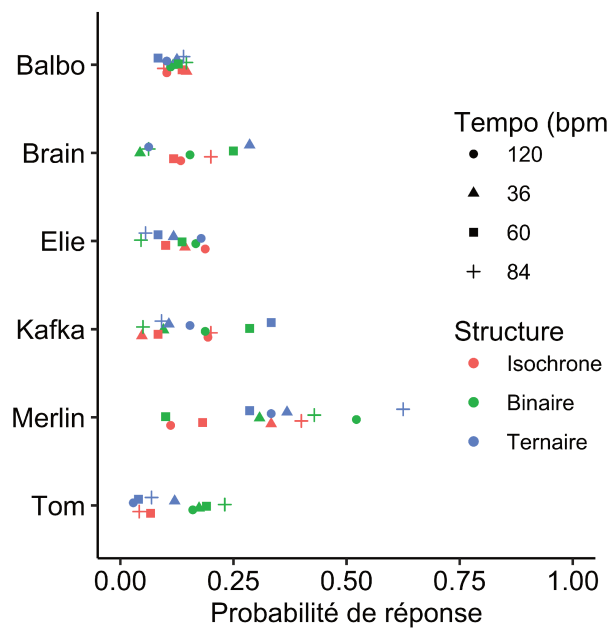


FIGURE 6.3 – Probabilité de répondre à une séquence de structure donnée après y avoir été exposé, pour chaque individu inclus dans l'étude. Une réponse est considérée comme la production d'au moins deux vocalisations pendant la séquence.

pour chaque individu, tous les  $p > 0,05$ ). Notamment, Balbo a répondu de manière approximativement égale à tous les stimuli.

Pour la suite des analyses temporelles, nous avons utilisé les réponses comptant au moins 5 vocalisations durant la séquence.

### 6.3.3 Les corbeaux peuvent-ils synchroniser leur chant aux stimuli ?

Nous avons d'abord évalué la capacité des corbeaux à faire correspondre la structure de leur chant à la structure des séquences qu'ils entendaient. Pour ce faire, nous avons étudié le délai entre le début de chaque vocalisation et celui de la dernière pulsation, normalisé par le tempo de la séquence (Fig. 6.4). Le meilleur sous-modèle était le modèle nul pour tous les oiseaux (Table 6.1, haut). Par conséquent, la structure de la séquence (qu'il s'agisse du tempo ou de la structure métrique) n'avait pas d'effet sur la synchronisation des vocalisations à la séquence. Au vu de la distribution des délais (Fig. 6.4A), nos résultats étaient plus cohérents avec une occurrence aléatoire du début de la vocalisation par rapport aux pulsations. Les corbeaux n'ont donc pas synchronisé leurs réponses vocales aux stimuli.

### 6.3.4 Influence des stimuli sur le rythme du chant

Nous avons ensuite cherché à savoir si la structure métrique et le tempo des séquences influençaient les intervalles entre vocalisations (IOI) et la durée des vocalisations du chant.

Pour l'IOI, les graphiques de distribution n'ont montré que de légères différences entre les stimuli pour la plupart des individus (Fig. 6.5). Chez quatre individus, le modèle nul était soit le meilleur, soit

TABLE 6.1 – Sélection de modèle pour chaque variable réponse et chaque individu. Les variables réponses testées étaient le délai entre le début de la vocalisation et la dernière pulsation du stimulus (Délai), l'intervalle entre vocalisations successives (IOI), et la durée de la vocalisation (Durée). Seuls les modèles moins de 2 AICc au-dessus du meilleur modèle sont montrés. Pour chaque modèle, les termes inclus sont indiqués par un +. df : degrés de libertés, logLik : logarithme de la fonction de vraisemblance, AICc : critère AICc, delta : nombre d'AICc au-dessus du meilleur modèle, et poids : poids du modèle.

Variable réponse	Individu	Rythme	Tempo	Interaction	df	logLik	AICc	delta	poids
Délai	Balbo				3	-228.43	462.89	0	0.78
	Brain				3	-20.14	46.54	0	0.74
	Elie				3	-10.93	28.36	0	0.7
	Kafka				3	-69.59	145.24	0	0.73
	Merlin				3	-151.23	308.5	0	0.68
	Tom				3	-25.66	57.54	0	0.74
IOI	Balbo	+	+	+	14	-450.11	928.61	0	1
	Brain				3	-111.57	229.41	0	0.85
	Elie				3	-67.08	140.68	0	0.35
	Elie	+	+		6	-64.02	141.99	1.31	0.18
	Elie	+	+	+	6	-64.02	141.99	1.31	0.18
	Elie	+			5	-65.53	142.41	1.74	0.15
	Elie		+		5	-65.53	142.42	1.74	0.15
	Kafka				3	-116.19	238.46	0	0.66
	Kafka	+			5	-115.11	240.39	1.93	0.25
	Merlin	+			5	-472.6	955.28	0	0.82
	Tom		+		6	-69.91	152.72	0	0.52
Tom				3	-73.66	153.57	0.85	0.34	
Durée	Balbo	+	+	+	14	726.59	-1424.8	0	1
	Brain				3	52	-97.75	0	0.85
	Elie	+	+		6	48.27	-82.62	0	0.47
	Elie	+	+	+	6	48.27	-82.62	0	0.47
	Kafka	+	+	+	12	279.07	-533.24	0	0.63
	Kafka	+			5	271.04	-531.91	1.33	0.33
	Merlin				3	910.6	-1815.2	0	0.66
	Merlin	+			5	911.69	-1813.3	1.87	0.26
	Tom		+		6	95.5	-178.2	0	0.41
	Tom				3	91.7	-177.17	1.03	0.25
Tom	+	+		8	97.28	-177.15	1.04	0.24	



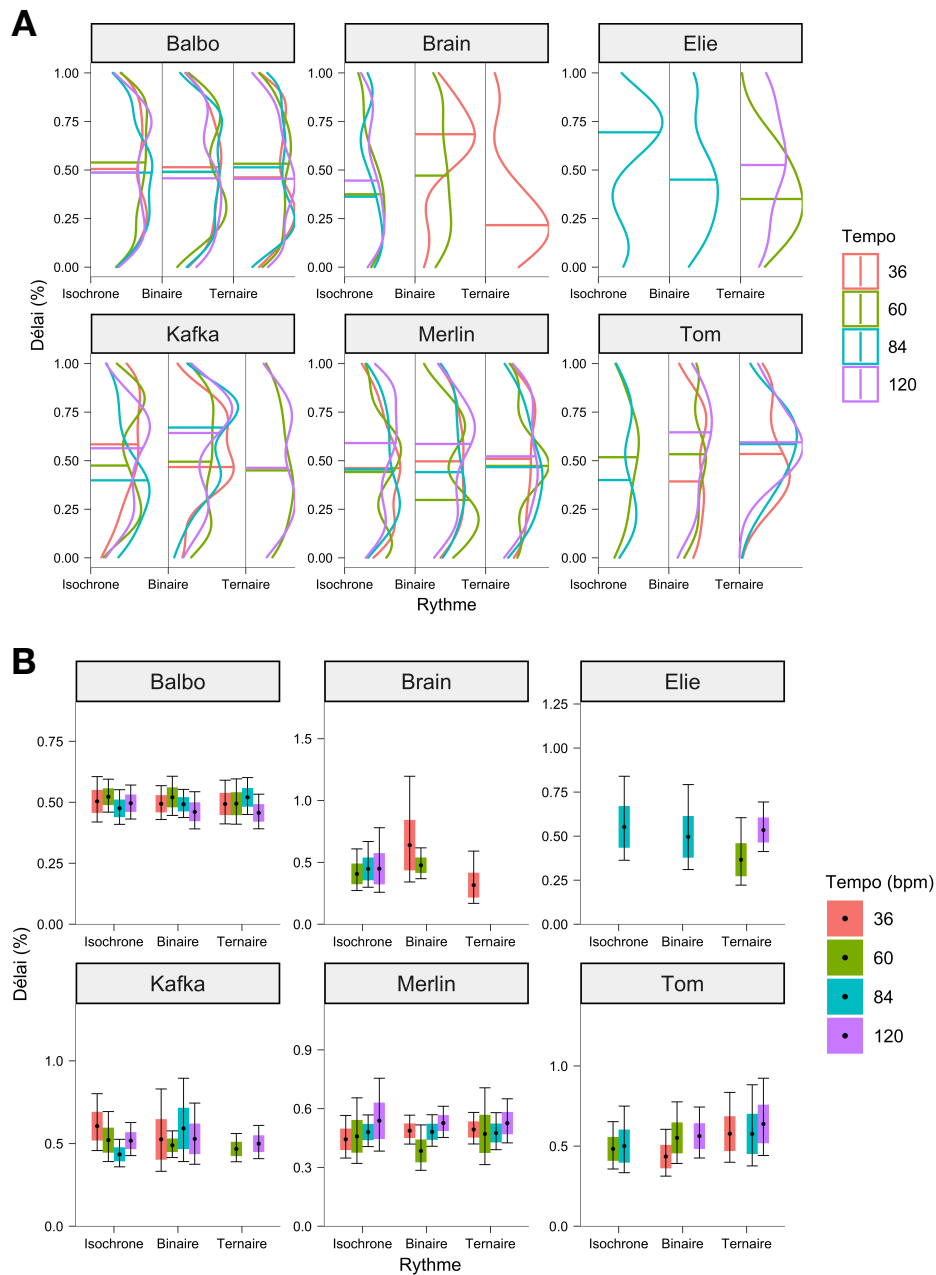


FIGURE 6.4 – Effets de la structure des stimuli sur le délai entre le début d’une vocalisation et celui de la dernière pulsation précédente. A) Distribution des délais, normalisés par le tempo, par individu, tempo, et structure métrique. B) Paramètres estimés par les GLMM. Les points correspondent aux valeurs moyennes estimées, les boîtes aux valeurs  $\pm 1$  erreur standard, et les moustaches aux intervalles de confiance à 95%. Les lignes horizontales correspondent aux tests significatifs, avec la p-valeur associée au-dessus.

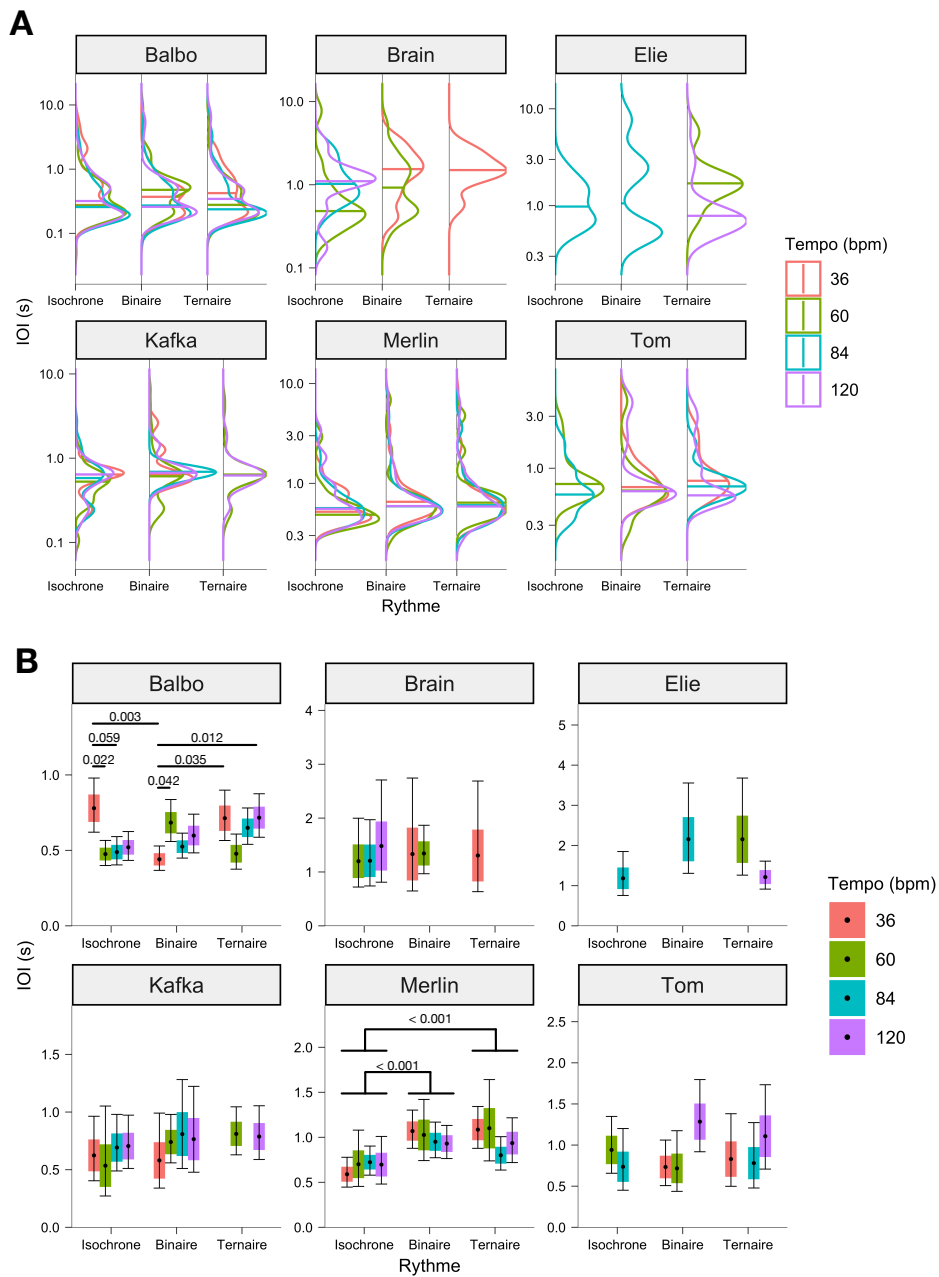


FIGURE 6.5 – Effets de la structure des stimuli sur les IOI A) Distribution des IOI par individu, tempo, et structure métrique. B) Paramètres estimés par les GLMM. Les points correspondent aux valeurs moyennes estimées, les boîtes aux valeurs  $\pm 1$  erreur standard, et les moustaches aux intervalles de confiance à 95%. Les lignes horizontales correspondent aux tests significatifs, avec la p-valeur associée au-dessus.

équivalent au meilleur sous-modèle (Table 6.1, milieu). Cependant, pour Balbo et Merlin, la procédure de sélection des modèles a indiqué que le meilleur modèle est un modèle non nul, et significativement meilleur que ce dernier. Nous n'avons donc effectué des tests par paires que pour ces deux individus. Pour Balbo (Fig. 6.5A et B, en haut à gauche), les IOI étaient significativement plus longs pour le tempo de 36 bpm et la structure isochrone que pour tout autre tempo à cette structure métrique, bien qu'aucune autre différence entre tempos n'ait émergé. De même, les IOI pour le tempo de 36 bpm dans la structure binaire ont conduit à des IOI significativement plus courts que les IOI du tempo de 36 bpm dans la structure isochrone et ternaire. Pour Merlin (Fig. 6.5A et B, milieu bas), le tempo n'a pas eu d'effet sur les IOI. Cependant, la structure métrique a eu un effet significatif. En effet, les séquences isochrones étaient associées à des IOI significativement plus courts que les séquences binaires (ratio binaire / isochrone :  $1,43 \pm 0,11$ ,  $z = 4,80$ ,  $p < 0,001$ ) ou ternaires (ratio ternaire / isochrone :  $1,39 \pm 0,06$ ,  $z = 4,24$ ,  $p < 0,001$ ). Les IOI associés avec les séquences binaires et ternaires, en revanche, ne différaient pas significativement (ratio binaire / ternaire :  $1,03 \pm 0,06$ ,  $z = 0,53$ ,  $p = 0,86$ ).

Pour la durée de la vocalisation, le meilleur modèle était soit le modèle nul (Brain, Merlin), soit équivalent au modèle nul (Tom, avec un AICc associé au modèle nul seulement 1.03 plus haut que le meilleur modèle). Pour les trois individus restants, un modèle non nul était significativement meilleur (Table 6.1, bas). Les tests par paires sur ces trois derniers individus ont révélé que les effets suggérés par la sélection de modèles n'atteignaient pas la significativité pour Kafka, Balbo suivait le même schéma global que pour les IOI, et pour Elie plusieurs effets étaient significatifs (mais Elie n'a aussi répondu qu'à certaines structures de séquences). Concernant Balbo, comme pour l'IOI, les vocalisations les plus longues ont été produites en réponse aux séquences isochrones à 36 bpm. Cependant, contrairement à l'IOI, il ne s'agissait pas simplement d'un allongement des vocalisations produites : des vocalisations courtes restaient présentes dans le chant (Fig. 6.6B, en haut à gauche). Pour Elie, nous avons observé des effets significatifs de la structure rythmique et du tempo : les vocalisations les plus courtes correspondaient au stimulus à 84 bpm de tempo avec une structure rythmique binaire. Les vocalisations en réponse aux stimuli isochrones de 84 et ternaires de 120 bpm étaient également significativement plus rapides que les vocalisations émises en réponse à 60 bpm.

## 6.4 Discussion

Nous avons montré que tous les membres du groupe pouvaient être exposés individuellement à des stimuli rythmiques et qu'une majorité de sujets pouvait répondre vocalement à ces stimuli. Certains oiseaux ont répondu préférentiellement à certains tempos ou structures métriques par rapport à d'autres. Nous n'avons pas trouvé d'ajustements précis entre les pulsations du stimulus et le début de la vocalisation suivante, ce qui montre que les oiseaux ne se sont pas synchronisés à la structure rythmique ou au tempo qu'ils entendaient. Néanmoins, certains oiseaux ont modifié la façon dont ils vocalisaient en réponse à certaines structures rythmiques ou à certains tempos, en ralentissant lorsqu'ils étaient exposés à un tempo plus lent ou en produisant des vocalisations plus courtes à des tempos plus élevés.

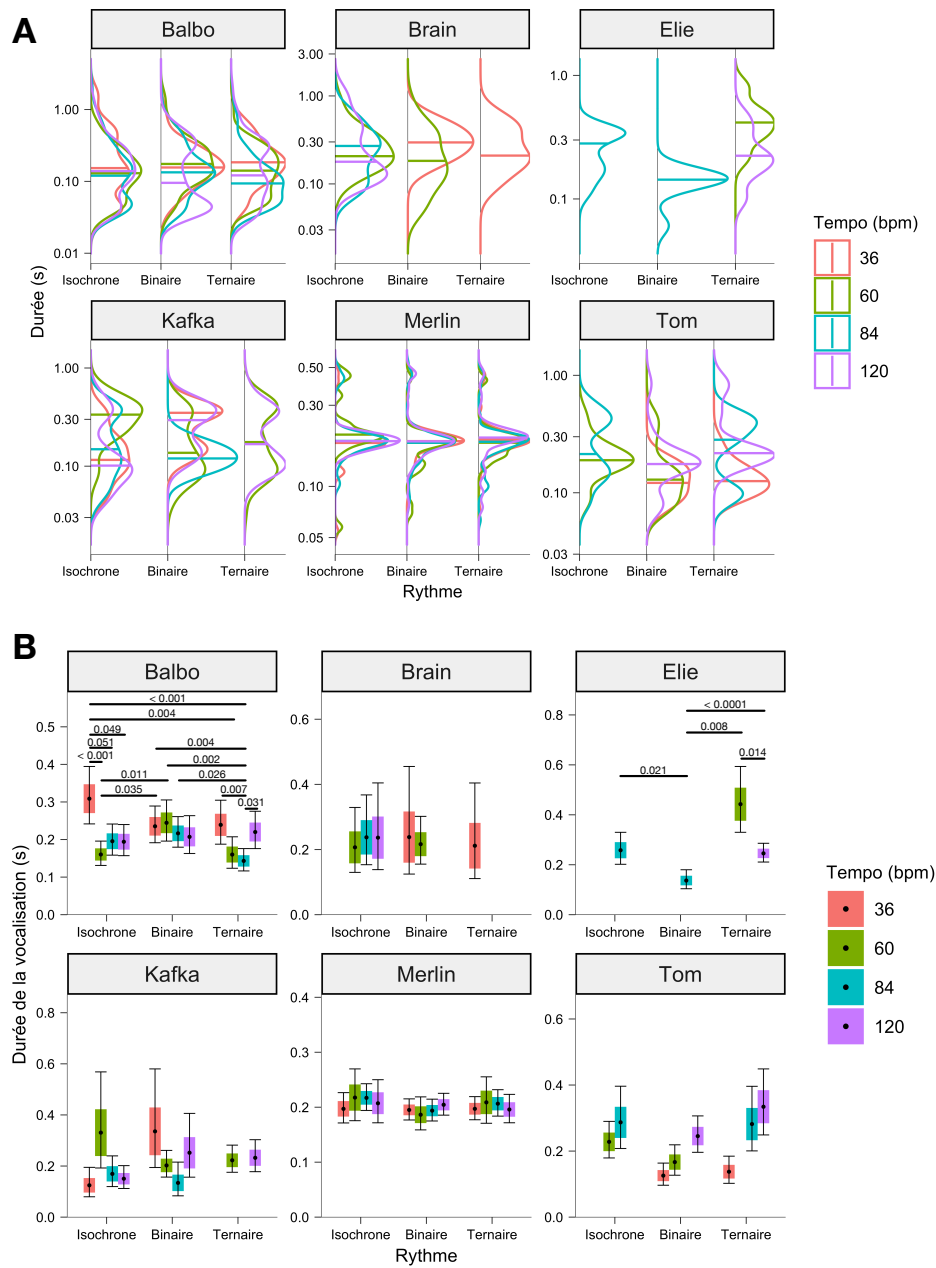


FIGURE 6.6 – Effets de la structure des stimuli sur la durée des vocalisations. A) Distribution des durées par individu, tempo, et structure métrique. B) Paramètres estimés par les GLMM. Les points correspondent aux valeurs moyennes estimées, les boîtes aux valeurs  $\pm 1$  erreur standard, et les moustaches aux intervalles de confiance à 95%. Les lignes horizontales correspondent aux tests significatifs, avec la p-valeur associée au-dessus.

Bien que le stimulus ait clairement suscité un intérêt, il n'a pas systématiquement conduit à une réponse. Nous ne sommes donc pas en présence d'une réponse automatique ou d'un conditionnement. Il est à noter qu'aucune récompense n'a jamais été donnée lors de cette expérience. Certains oiseaux ont également répondu à un nombre limité de structures temporelles ou rythmiques, tandis que d'autres n'ont pas montré de préférence. Il peut y avoir une variabilité interindividuelle considérable dans la façon dont le stimulus a été perçu, dans l'attraction qu'il a générée et dans la motivation ultérieure à répondre. L'attraction pour la musique ou les sons rythmiques n'est pas une caractéristique universelle chez les animaux ou même chez les vertébrés. Chez les primates, les chimpanzés peuvent préférer la musique au silence (MINGLE et al., 2014), tandis que les tamarins à tête blanche préfèrent les tempos plus lents et évitent la musique lorsque cela est possible (MCDERMOTT et HAUSER, 2007). Les humains (COOK et FUJISAWA, 2006) et certains oiseaux (poussins nouveau-nés de *Gallus gallus*, CHIANDETTI et VALLORTIGARA, 2011) préfèrent la musique consonante à la musique dissonante, contrairement aux primates non-humains (MCDERMOTT et HAUSER, 2007). Ici, nous avons utilisé deux types de stimuli : un son chaotique proche d'un bruit blanc (eau courante), et un son légèrement plus harmonique (drone). Nous n'avons pas constaté de différence nette dans le taux de réponse à l'un ou l'autre de ces deux stimuli. En outre, certains individus ont répondu davantage à certains tempos (e.g. Elie a répondu le plus souvent au tempo de 120 bpm) ou à des structures rythmiques particulières (e.g. Merlin a répondu le plus souvent aux séquences ternaires à 84 bpm), bien que nous n'ayons aucune explication quant à la raison pour laquelle ces séquences peuvent être plus attrayantes que d'autres pour ces individus. Il est intéressant de noter qu'aucun des sujets n'a montré une nette préférence pour les séquences isochrones, contrairement aux autres oiseaux (ROESKE et al., 2020). Cette réponse est surprenante car l'isochronie permet de prédire plus facilement le moment de la prochaine pulsation (FRAISSE, 1982 ; MERKER et al., 2009), et nous aurions donc pu nous attendre à ce que les séquences isochrones soient plus faciles pour l'entraînement, et donc plus stimulantes. Lors de la conception de cette expérience, nous avons ajouté des structures métriques binaires ou ternaires pour augmenter la diversité des stimuli et limiter le risque d'accoutumance des corbeaux aux stimuli, ce qui pourrait diminuer leur motivation à s'entraîner. L'utilisation de plusieurs types de structures métriques était également un moyen d'étudier comment les oiseaux réagissent à des motifs acoustiques d'ordre supérieur, et s'ils peuvent percevoir des groupes d'intervalles. Les humains peuvent regrouper des séquences rythmiques variant en accentuation (groupement iambique) ou en durée (groupement trochaïque) et assembler des séries de pulsations dans des groupes correspondants (de la MORA et al., 2013 ; FITCH, 2013). Bien que les capacités rythmiques des animaux non humains restent mal connues, certaines espèces présentent la capacité de percevoir des rythmes simples, notamment les oiseaux (ten CATE et SPIERINGS, 2019), les rats (CELMA-MIRALLES et TORO, 2020) et les primates (MERCHANT et HONING, 2014). Chez ces espèces, la perception du rythme semble être largement trochaïque, mais pas iambique. Dans cette étude, les stimuli suivaient un schéma trochaïque car les seules différences entre pulsations étaient leur durée. Par conséquent, les corbeaux doivent au moins également avoir la capacité de percevoir ces schémas, même si nous ne pouvons pas conclure plus avant sur la perception de ces rythmes chez cette espèce.

La perception des sons par les animaux reste un sujet de recherche actif, en particulier en ce qui

concerne l'origine de la musicalité, telle qu'on la trouve dans les langues humaines. Certaines espèces peuvent distinguer différents styles de sons, comme différents compositeurs (pigeons, *Columba livia*, PORTER et NEURINGER, 1984) ou genres musicaux (carpe, *Cyprinus carpio*, CHASE, 2001). Cependant, la manière dont la discrimination est réalisée reste largement obscure (BOUWER et al., 2021 ; KRIENGWATANA et al., 2022). Les perceptions des oiseaux, par exemple, semblent indiquer une haute résolution temporelle, plus axée sur les oscillations temporelles et la trajectoire du son dans le temps (e.g. une montée dans les aigus ou une descente dans les graves) que sa fréquence absolue (BREGMAN et al., 2016 ; CYNX et al., 1990 ; DOOLING et PRIOR, 2017 ; ten CATE et al., 2016). La perception du rythme, cependant, est encore peu claire. Certaines espèces sont capables de distinguer des rythmes différents (cf KRIENGWATANA et al., 2022 pour une revue), mais uniquement entre des rythmes isochrones et non isochrones, et non entre différents rythmes non isochrones (CELMA-MIRALLES et TORO, 2020 ; HAGMANN et COOK, 2010). Cependant, les étourneaux sansonnets (*Sturnus vulgaris*) pourraient être plus performants dans ce type de tâche (HUMPAL et CYNX, 1984). La perception du rythme a été reliée à l'apprentissage vocal et au mimétisme vocal chez les oiseaux (DOOLING et al., 2002), mais cette hypothèse reste débattue : certaines espèces qui n'apprennent pas de vocalisations, comme les chimpanzés, peuvent néanmoins produire des rythmes réguliers, y compris en structure métrique binaire (Barney, DUFOUR et al., 2017). Cela suggère une représentation interne du rythme, et pourrait être une autre forme d'entraînement. Chez les oiseaux, certains auteurs suggèrent que l'apprentissage vocal et le mimétisme vocal sont deux facteurs qui augmentent la probabilité qu'une espèce soit capable d'entraînement (WILSON et MOEHLIS, 2014). Les corvidés sont connus pour imiter les voix humaines, mais l'importance de l'apprentissage vocal dans l'acquisition de leur répertoire n'est pas bien décrite (mais voir KONDO, 2021). Chez les corbeaux freux, nous avons montré dans une étude précédente que les mâles ne possèdent pratiquement aucun répertoire vocal commun, et que les quelques cris partagés entre les individus ne l'étaient qu'entre des paires spécifiques (Martin et al. en préparation, Chapitre 4). Lorsque les individus partageaient un cri, ils ne partageaient pas un lien social particulier. Nous avons montré que les corbeaux étaient flexibles dans leur façon de vocaliser, avec des cris variés et gradués, et une façon irrégulière de chanter, sans schéma évident de vocalisations pour certains oiseaux. Ainsi, bien que la phase de développement de l'acquisition de leur vocalisation ne soit pas bien connue, leurs aptitudes au mimétisme vocal peuvent soutenir leurs tentatives d'adaptation au stimulus.

L'entraînement rythmique est la capacité de percevoir une pulsation qui marque des points réguliers dans un stimulus sonore, puis d'aligner les actions motrices sur cette pulsation ou ce battement (MERCHANT et HONING, 2014). Ajuster le chant à une pulsation externe n'est pas un exemple standard d'entraînement moteur, mais peut néanmoins être considéré comme un mouvement sonore ("*sonorant movement*", CHAUVIGNÉ et al., 2014). Pendant qu'ils chantaient, les corbeaux se mouvaient également pour produire la vocalisation (des mouvements également décrits par COOMBS, 1960 et observés lors des chants produits hors de cette expérience). Aucun mouvement rythmique n'était produit alors qu'ils étaient silencieux (i.e. pas de hochement de tête, de tapement de pied, etc. comme observés chez Snowball PATEL et al., 2009). Cette capacité d'entraînement vocal est peut-être rare chez les animaux non humains. À ce jour, chez les vertébrés, les perroquets sont peut-être la seule

autre espèce capable d'entraînement vocal, mais aucun cas spontané d'entraînement vocal n'a été documenté, ni chez eux, ni chez les autres espèces.

Étant donné l'absence de correspondance exacte des pulsations dans la synchronisation des vocalisations, nous pouvons éventuellement discuter de la notion d'entraînement ici. Notre étude a porté sur les réponses spontanées des oiseaux qui n'ont été ni conditionnés, ni encouragés à faire correspondre le moment du début de leur vocalisation à celui des pulsations du stimulus. Cependant, nos résultats suggèrent que les corbeaux pourraient être un excellent candidat pour évaluer s'ils peuvent être conditionnés à entraîner vocalement. Alors que la littérature sur l'entraînement et sur les origines évolutives de la musicalité est montée en flèche ces dernières années, seule une poignée de ces publications présente de nouvelles données. Il reste à voir avec quelle facilité les corbeaux peuvent être ainsi conditionnés, à la fois en l'absence de renforcement positif (e.g. nourriture), et en raison de la néophobie commune à de nombreux corvidés (GREENBERG et METTKE-HOFMANN, 2001 ; GREGGOR et al., 2016 ; R. MILLER et al., 2022). Cette néophobie complique les expériences avec les corvidés, chez qui la familiarité avec les protocoles et les expérimentateurs influencent fréquemment la performance dans des tests cognitifs (CIBULSKI et al., 2014). Cependant, les indices de contrôle vocal chez ces espèces comme dans la possibilité d'apprendre à associer des vocalisations particulières, ou l'inhibition de leur production, à des stimuli arbitraires (BRECHT et al., 2019) méritent un examen plus poussé des capacités vocales des corvidés.

Si les corbeaux freux peuvent percevoir des rythmes et ajuster leur production vocale en conséquence, pourraient-ils également avoir une représentation interne du rythme ? Si tel est le cas, ils pourraient être en mesure de prédire les pulsations futures (i.e. timing prédictif) au lieu de simplement réagir aux pulsations passées (i.e. timing réactif), ce qui devrait être reflété dans un maintien du rythme même après que le stimulus se soit arrêté (RAVIGNANI et al., 2014). La seule espèce où un timing prédictif a pu être trouvé est le chimpanzé (DUFOR et al., 2017), une autre espèce complexe socialement et flexible sur le plan comportemental, qui ne montre pourtant aucun indice d'apprentissage vocal, contrairement aux humains et aux oiseaux. La découverte d'un timing prédictif chez une espèce aussi éloignée phylogénétiquement que le corbeau freux pourrait donner un aperçu des capacités ou des conditions nécessaires à l'apparition de la musicalité chez les humains.

# Chapitre 7

## Discussion générale

Au cours de cette thèse, j'ai étudié plusieurs aspects de la complexité vocale chez le corbeau freux. Un aspect majeur de ces études a été le développement d'outils basés sur les développements récents de l'apprentissage automatique et de la bioacoustique. Au chapitre 3, j'ai développé Rookognise, un réseau de neurones profond pour extraire les vocalisations de corbeaux freux et identifier les individus vocalisant dans des enregistrements de terrain. Au chapitre 4, j'ai adapté des mesures de similarité entre spectrogrammes pour quantifier la diversité, gradation et similarité dans le répertoire de l'espèce. Au chapitre 5, j'ai étudié l'existence d'une signature vocale à différents niveaux sociaux dans le cri au nid de corbeaux freux femelles en période de reproduction. Enfin, au chapitre 6, j'ai étudié l'aptitude des corbeaux freux à joindre spontanément leur chant à un rythme extérieur.

### 7.1 Suivi acoustique individuel avec un réseau de neurones profond

Les signatures vocales individuelles ne sont pas un développement récent ni dans leur détection acoustique (BEECHER, 1989; BRADBURY et VEHCENCAMP, 1998; LINHART et al., 2019), ni dans leur découverte dans différentes espèces (e.g. BEECHER, 1989; BENTI et al., 2019; BERG et al., 2012; BOECKLE et al., 2012, 2018; BRADBURY et VEHCENCAMP, 1998; ELIE et THEUNISSEN, 2018; KEENAN et al., 2020; KONDO et al., 2010; KONDO et al., 2012; LINHART et al., 2019; MANSER et al., 2014; MATES et al., 2015; QUICK et JANIK, 2012; RIBOT et al., 2012; SEWALL, 2009; STOWELL et al., 2016a; SUEUR et al., 2021; THOMSEN et al., 2019; TYACK, 2000; WASCHER et al., 2012b; WRIGHT et DAHLIN, 2017; YORZINSKI et al., 2006; ZDENEK et al., 2018), ni dans leur utilisation pour reconnaître et communiquer entre partenaires sociaux (RØSKAFT et ESPMARK, 1984; SMITH-VIDAURRE et al., 2020). Cependant, leur exploitation a jusqu'à récemment été limitée par les méthodes classiques. Ces méthodes requièrent souvent l'extraction préalable des vocalisations ou la mesure de paramètres acoustiques, tous deux fondamentalement incompatibles avec une analyse automatique, notamment sur un répertoire vocal riche et des vocalisations complexes comme celles des corvidés, où certains paramètres acoustiques ne sont tout simplement pas accessibles dans la majorité des vocalisations (BENTI et al., 2019; FLETCHER, 2000; STOWELL et al., 2016b). Par exemple, la fréquence fondamentale d'une vocalisation aviaire peut ne pas être détectable ou ne pas être unique, si la biphonation est marquée dans cette vocalisation (NOWICKI, 1987; WILDEN et al., 1998). Même chez les espèces où les signatures vocales individuelles ont pu être décrites avec ces paramètres acoustiques, ces approches sont limitées car la signature vocale individuelle est plus ou moins fiable en fonction des vocalisations (CANDIOTTI et al., 2013; CHARRIER et al., 2001; JANSEN et al., 2012; KEENAN et al., 2020; MANSER et al., 2014; PRICE et al., 2009; VANNONI et MCELLIGOTT, 2007)



et varie souvent au sein du répertoire (ELIE et THEUNISSEN, 2018; JANSEN et al., 2012; KEENAN et al., 2020; SALMI et al., 2014).

Pour ces raisons, j’ai choisi d’employer l’apprentissage profond. Le réseau Rookognise développé au Chapitre 3 repose sur l’analyse de Mel-spectrogrammes extraits directement des enregistrements, sans autre pré-traitement tel qu’un débruitage. Le réseau peut séparément détecter les vocalisations de corbeaux freux, les sexer, et identifier les individus vocalisant. Cela a plusieurs implications pratiques. Outre le fait d’accélérer l’extraction de données sur le groupe d’étude originel, y compris sur des enregistrements pris hors présence humaine, les différentes tâches correspondent à une hiérarchie dans leur spécificité. L’identification individuelle ne fonctionne correctement que sur le groupe sur lequel le réseau a été entraîné, à cause de la nature même de l’entraînement supervisé. Cependant, les tâches de détection de vocalisation et de sexage vocal sont plus générales. J’ai ainsi pu utiliser le réseau entraîné uniquement sur la colonie captive de Strasbourg pour faciliter l’acquisition de données sur la colonie captive de Madingley, car le réseau détecte extrêmement bien les vocalisations de ces autres individus, et, dans une moindre mesure, arrive également à les sexer (Table 7.1). Le réseau est beaucoup plus efficace après avoir été ré-entraîné sur la colonie de Madingley en plus de celle de Strasbourg, mais reste très bon même sans cet entraînement. Les réseaux de neurones profonds généralisent bien à des données similaires aux données utilisées pour les entraîner, tant que certaines précautions sont prises pour éviter le surapprentissage (GOODFELLOW et al., 2015; KUKAČKA et al., 2017; L. N. SMITH, 2018), comme les méthodes de régularisation que j’ai utilisées pour Rookognise au Chapitre 3.

TABLE 7.1 – Rookognise détecte et sexe les vocalisations des corbeaux freux en généralisant à d’autres colonies. Les métriques d’AUROC et d’AUPRC utilisées au Chapitre 3 pour évaluer le réseau ont été calculées à partir de données collectées sur la colonie de Madingley. La première colonie correspond aux données utilisées pour entraîner le réseau, la deuxième colonne correspond aux données utilisées pour l’évaluation.

Colonie utilisée pour entraîner	Colonie testée	Détection		Sexage	
		AUROC	AUPRC	AUROC	AUPRC
Strasbourg	Strasbourg	92.35	81.78	95.05	81.65
Strasbourg	Madingley	95.04	65.75	91.73	32.30
Strasbourg + Madingley	Strasbourg	96.84	90.53	98.41	98.09
Strasbourg + Madingley	Madingley	98.91	93.15	99.09	75.95

Un autre résultat issu de Rookognise, plus surprenant *a priori* au vu de la littérature sur les signatures vocales, est issu de la représentation apprise par le réseau avant la classification finale. En général, les réseaux de neurones sont des boîtes noires, en cela que le cheminement entre les données d’origine et le résultat final n’est pas accessible à un utilisateur, sauf certaines approches indirectes comme la carte d’activation de classe (GradCAM) qui permet de visualiser sur l’image d’origine les régions qui contribuent le plus à une décision (SELVARAJU et al., 2016; SIMONYAN et al., 2014). Cependant, sans chercher à déterminer ce cheminement, il est possible d’extraire une représentation intermédiaire (par exemple, le résultat de l’avant dernière couche de la tête d’identification du réseau, juste avant la classification finale identifiant l’individu vocalisant) et de la visualiser avec l’UMAP

(Fig. 7.1). Les différents individus forment dans l'ensemble des clusters bien séparés. Certains individus correspondent à un cluster unique (Bashir en bleu en bas à droite, Merlin en orange à gauche, Bussell en bleu au milieu gauche, entre Merlin et les autres) : Rookognise peut alors représenter toutes les vocalisations de ces individus d'une seule façon pour les identifier. D'autres individus cependant forment de multiples clusters (Kafka en jaune au milieu, Balbo en violet en haut à droite, Aristotle en violet en bas), même si ces clusters peuvent être eux-mêmes bien séparés des autres : Rookognise n'a pas réussi à apprendre une représentation commune des vocalisations de cet individu. Rookognise avait pourtant accès pendant l'entraînement aux labels des vocalisations, ce qui aurait dû faciliter l'apprentissage d'une représentation commune, si une telle représentation existe. Ce résultat implique l'absence d'une signature vocale individuelle commune à tout le répertoire de l'individu pour cette espèce, comme pour d'autres espèces étudiées auparavant (ELIE et THEUNISSEN, 2018; JANSEN et al., 2012; KEENAN et al., 2020; SALMI et al., 2014).

Le développement de Rookognise au cours de cette thèse rejoint les applications d'apprentissage profond qui ont déjà été développées dans plusieurs autres domaines écologiques (STOWELL, 2022 : détection acoustique de vocalisations d'oiseaux (FANIOUDAKIS et POTAMITIS, 2017; GRILL et SCHLÜTER, 2017; KONG et al., 2017a; LIAQAT et al., 2018; LOSTANLEN et al., 2019; SEVILLA et GLOTIN, 2017; STOWELL et al., 2019b), de primates (HEINICKE et al., 2015), ou d'autres sons animaux (FOLLIOT et al., 2022), classification d'espèces (JOLY et al., 2021; KAHL et al., 2021b), divers suivis écologiques (CONRADY et al., 2022; DUFOURQ et al., 2022; FU et al., 2022; GRAY et al., 2019; SHE et al., 2022; TEIXEIRA et al., 2022; van KLINK et al., 2022; WEINSTEIN, 2018), et identification individuelle, mais sur image (FERREIRA et al., 2020). Pour le coût de quelques heures d'enregistrement ou quelques milliers de vocalisations annotées à la main, le réseau peut en théorie immensément faciliter la collecte de données, potentiellement y compris sur des espèces difficiles à observer, ou sans perturbations humaines. Une étude préliminaire que j'ai supervisée a tenté d'appliquer Rookognise à une autre espèce complexe socialement et vocalement, le macaque de Tonkean (REBOUT et al., 2020; THIERRY, 2007). Cependant, cette espèce est aussi beaucoup moins vocale que le corbeau freux, ce qui allonge considérablement le temps d'acquisition d'une base de données pour entraîner le réseau à identifier les différents individus. Sur une espèce comme le corbeau freux cependant, où les vocalisations sont extrêmement fréquentes (environ 800 par heure dans les données annotées pour cette thèse), Rookognise peut être entraîné facilement avec seulement quelques heures de données annotées. Une étude de la quantité minimale de données d'entraînement nécessaire pour approcher la performance obtenue permettra de raffiner les conditions nécessaires à un bon déploiement de Rookognise à d'autres espèces. Une fois entraîné, Rookognise peut être appliqué à de très vastes quantités de données, comme celles obtenues avec les enregistreurs autonomes que j'ai utilisés dans cette thèse. À titre d'illustration, en comptant environ 300 heures d'enregistrements sur les 2.5 années de suivi depuis le début de l'étude (de janvier 2020 à août 2022), en considérant une fréquence de 600 vocalisations par heure (environ 30000 vocalisations sur 50 heures annotées) cela reviendrait à collecter et identifier environ 1.8 millions de vocalisations automatiquement. Rookognise peut en outre être réglé manuellement pour équilibrer qualité ou quantité dans les vocalisations extraites, en particulier du point de vue de la réduction de l'effort humain nécessaire pour valider les

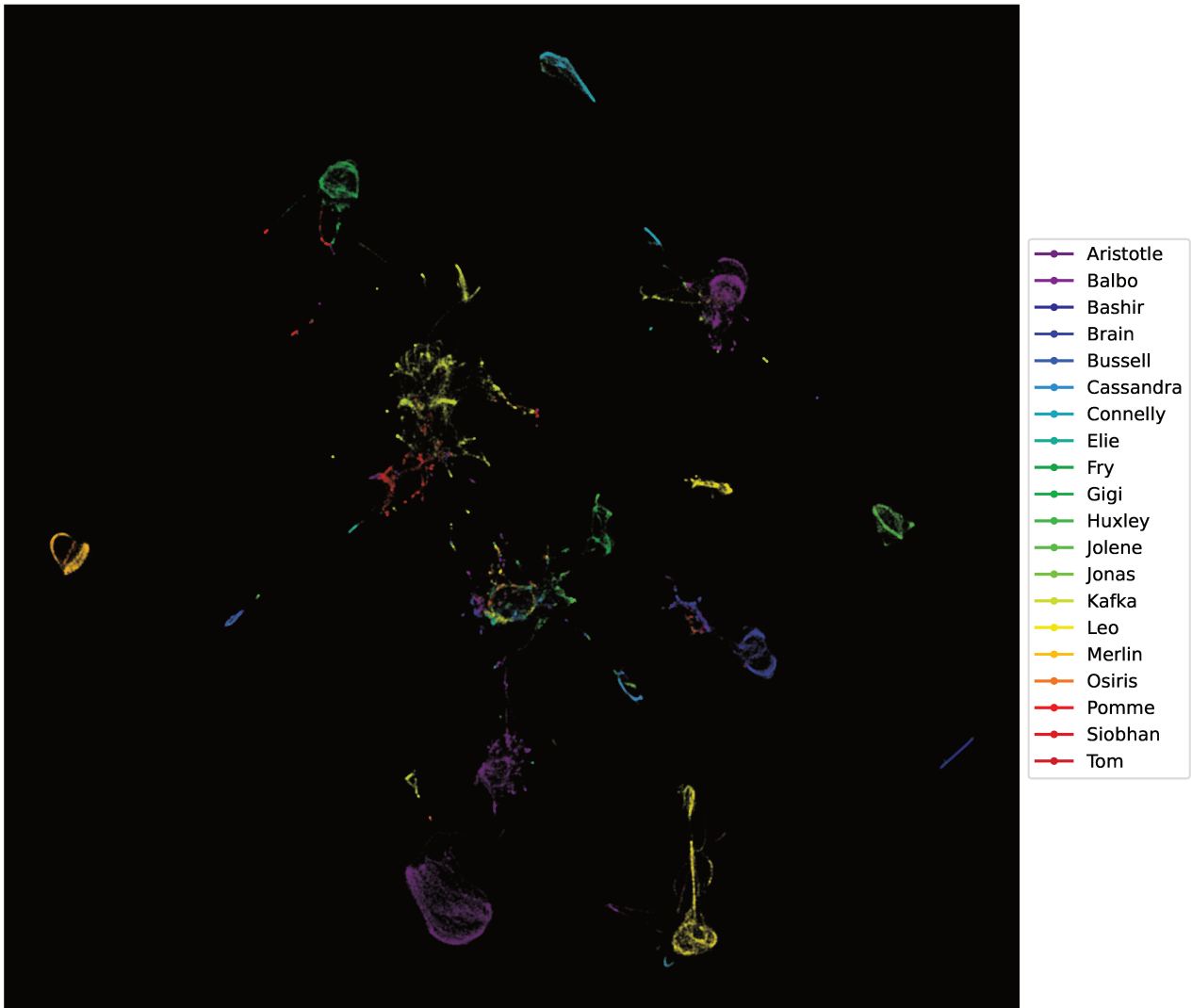


FIGURE 7.1 – Projection UMAP à partir d’une représentation intermédiaire des vocalisations des corbeaux freux des deux colonies captives. Un système Rookognise est entraîné sur les données des deux colonies captives, puis testé sur d’autres vocalisations (non vues pendant l’entraînement) des deux colonies. La représentation intermédiaire juste avant l’identification finale des individus (i.e. la sortie de l’avant-dernière couche de la tête d’identification) est extraite, les non-vocalisations sont rejetées, et les données restantes sont passées dans UMAP pour obtenir cette projection.

résultats. En particulier, un seuil de choix d'accepter ou de rejeter des vocalisations, et un autre pour accepter ou rejeter l'attribution de la vocalisation à un individu, permettent de contrôler l'équilibre entre faux positifs et faux négatifs. Des seuils élevés permettent d'extraire des vocalisations avec une grande certitude que les décisions du réseau sont correctes, mais manquent une partie des vocalisations (faux négatifs). À l'inverse, des seuils bas peuvent extraire presque toutes les vocalisations, mais acceptent aussi des non-vocalisations et des attributions erronées (faux positifs), nécessitant un effort de nettoyage plus élevé. En combinant la capacité d'analyse de Rookognise à cette quantité massive de données, il sera possible d'extraire des vocalisations particulières, comme les chants, pour poursuivre les études bioacoustiques sur le corbeau freux, mais aussi sur bien d'autres espèces. Par ailleurs, comme j'en discutais au Chapitre 3, l'architecture en multi-tâche du réseau le rend facilement applicable à d'autres tâches, en plus ou à la place de certaines des tâches présentées dans ce travail, comme la classification des vocalisations si le répertoire de l'espèce est disponible (COHEN et al., 2022; VERSTEEGH et al., 2016), ou l'état émotionnel de l'individu (BRIEFER et al., 2022), ce qui pourrait par exemple être utile dans des protocoles de suivi du bien-être animal (LAURIJS et al., 2021).

## 7.2 Complexité vocale des corbeaux freux

J'ai étudié une partie des aspects de la complexité vocale dans le répertoire du corbeau freux, plus précisément, la diversité des vocalisations, la gradation dans leur structure, et la similarité des répertoires des différents individus, en comparant entre les sexes et entre les deux colonies captives suivies. J'ai choisi une approche très générale, basée sur le même raisonnement que pour Rookognise, d'éviter les paramètres acoustiques discrets. Les Mel-spectrogrammes des vocalisations ont été utilisés, avec la DFTW pour prendre en compte les décalages en temps ou en fréquence qui pourraient être dus à des variations non fonctionnelles (par exemple, cela permet d'effacer les effets purement physiologiques sur le timbre des vocalisations). Il s'agit sans doute d'une assomption assez stricte, et l'application de ces méthodes à des espèces où le répertoire vocal est mieux connu, par exemple le diamant mandarin (ELIE et THEUNISSEN, 2016), l'étourneau (SAINBURG et al., 2019), ou les macaques (REBOUT et al., 2021), permettront de vérifier si des variations potentiellement importantes ne sont pas effacées. Un autre point important à noter est que la DFTW est une distance coûteuse à calculer. En particulier, le temps de calcul croît proportionnelle au carré de la taille du jeu de données; par exemple, le calcul de la matrice de distance entre les 3351 vocalisations au Chapitre 5 a pris environ 20 minutes sur le matériel informatique utilisé, alors que le calcul équivalent sur les 23 514 vocalisations au Chapitre 4 prend environ 10 heures (soit une augmentation de temps de calcul de 30 fois pour une augmentation de taille du jeu de données de seulement 7 fois). Cette considération pourra limiter l'usage de la DFTW dans des jeux de données extrêmement massifs, ou nécessiter l'adaptation du matériel informatique pour l'analyse.

L'UMAP a permis ensuite de réduire les dimensions de la matrice de distance DFTW, pour que l'algorithme de clustering HDBSCAN accède plus facilement aux différents clusters de vocalisations. Ces clusters, que j'ai gardé diffus pour éviter d'imposer des conditions supplémentaires à la procé-

ture, m'ont permis de mesurer la complexité vocale du répertoire. Je me suis basé sur la théorie de l'information et les mesures d'incertitude dans un système (CHAO et al., 2014; REBOUT et al., 2021) pour conserver une approche aussi objective et applicable à d'autres cas d'études que possible. Les indices de Hill (CHAO et al., 2014; HILL, 1973) permettent de quantifier la quantité de classes différentes dans un système, intégrant ainsi leur nombre et leur distribution dans une seule mesure. En appliquant deux indices de Hill, correspondant aux diversités de Shannon et de Simpson (Chapitre 4, j'ai ainsi pu quantifier de façon objective la diversité vocale et la gradation dans le répertoire des corbeaux freux, en fonction notamment du sexe et du type de vocalisation (chant / cri). Les femelles corbeaux freux possèdent un répertoire de cris plus divers et plus gradés que les mâles, tandis que les chants suivent un schéma inverse.

Plus surprenant cependant est le résultat sur les similarités vocales. J'ai utilisé l'indice de recouvrement de Morisita-Horn (CHAO et al., 2006), basé sur les analyses de biodiversité. Cet indice devrait donc permettre de mettre à jour la quantité de clusters communs aux différentes paires d'individus. En comparant encore une fois entre sexes, type de vocalisation, et colonie, j'ai découvert que les cris des mâles étaient très fortement idiosyncratiques, même au sein d'une même colonie, alors que les cris des femelles étaient hautement similaires, même entre individus de colonies différentes. Ce résultat a été trouvé malgré l'effet de la DFTW sur les différences en temps ou en fréquence, indiquant des structures de cris très différentes entre mâles. Cette dissimilarité vocale inter-individuelle n'a à ma connaissance jamais été observée chez d'autres espèces, *a fortiori* dans les cris qui ont en général des fonctions précises (MARLER et SLABBEKOORN, 2004). La communication requiert une interprétation par le receveur (CHERRY, 1955; SLATER, 1983; J. W. SMITH, 1965). Si cette interprétation doit être faite non seulement sur la structure de la vocalisation, mais aussi sur l'identité de l'émetteur, la compréhension devient beaucoup plus difficile. Plusieurs possibilités existent pour expliquer ce phénomène, certaines particulières à l'espèce, d'autres plus générales et qui ont trait aux fonctions potentielles des cris dans les deux sexes, à la ségrégation de la signature vocale et des autres fonctions de communication, et aux capacités d'apprentissage vocal chez les corvidés.

L'association entre structure et fonction des vocalisations chez les corvidés est particulièrement difficile à étudier, en raison notamment de leurs vocalisations très variables et gradées (E. D. BROWN, 1985; MATES et al., 2015; ROSKRAFT et ESPMARK, 1982; TANIMOTO et al., 2017). Plusieurs études ont noté que les individus produisent des vocalisations acoustiquement similaires dans plusieurs contextes différents (BROWN et FARABAUGH, 1997; ROSKRAFT et ESPMARK, 1982), mais aussi des vocalisations acoustiquement différentes dans des contextes similaires (BROWN et FARABAUGH, 1997; TANIMOTO et al., 2017). Il est donc possible que les différents cris des mâles soient en partie le résultat de cette association peu stricte entre structure et fonction. Chez les femelles, une proportion non négligeable des vocalisations a pu être associée à des contextes fonctionnels : quémante, nourrissage par le partenaire, construction de nids, cris au nid. Cependant, la plupart des vocalisations des mâles n'ont pu être associées qu'au contexte de chant, alors que les cris n'étaient majoritairement associables à aucun contexte. La seule exception notable était le comportement de *bow and tail fanning*, qui est lui-même très ambivalent, étant produit dans divers contextes : affiliation, alarme, agression, pré-copulation, ou chant (COOMBS, 1960; SEED et al., 2007). Les autres cris des mâles, s'ils sont

véritablement non fonctionnels outre leur caractère idiosyncratique, pourraient alors correspondre à des cris de contact, et d'autres cris plus fonctionnels pourraient moins différer entre les différents individus. Cependant, les observations et les données collectées ne permettent pas strictement de conclure sur cette question. Chez les autres espèces sociales, les cris de contact sont parmi ceux qui diffèrent le plus entre individus, ce qui semble logique si leur fonction unique est de porter une signature vocale (CORTOPASSI et BRADBURY, 2006; KEENAN et al., 2020; KONDO et WATANABE, 2009; KREMERS et al., 2012). Chez les conures à front orange (*Eupsittula canicularis*), les cris de contact ont des structures différentes entre individus à l'échelle locale, mais globalement sont constitués des mêmes éléments, simplement arrangés différemment (CORTOPASSI et BRADBURY, 2006). Cela ne semble pas être le cas chez les corbeaux freux où les cris sont chaotiques plutôt que composés de plusieurs éléments différents (Chapitre 4, BENTI et al., 2019; ROSKAFI et ESPMARK, 1982). Cependant, appliquer la méthode utilisée ici aux cris de conures pourrait permettre de retrouver des résultats similaires. En effet, la DFTW est une mesure de dissimilarité globale : si deux vocalisations sont composées de mélange d'éléments identiques, mais dans des ordres différents, elles seront considérées comme très dissimilaires.

Les idiosyncrasies vocales des mâles remettent en question la notion de communication dans cette espèce. Pour se comprendre, les individus ont *a priori* besoin d'un code de communication commun, où des vocalisations données devraient avoir un sens donné (BEECHER, 2021). Ce code est généralement présent à différents niveaux des communications vocales d'autres espèces. À son niveau le plus évident, il s'agit du concept de répertoire plus usuel, où les individus d'une espèce tendent globalement à partager les mêmes vocalisations au sein des groupes sociaux et qui permet de définir la fonction de certains cris (MARLER et al., 1992; SEYFARTH et al., 1980; SUZUKI, 2014, 2016), même si tous les individus d'une espèce ne partagent pas toutes les vocalisations du répertoire (R. . F. LACHLAN et al., 2010). L'avantage de posséder des vocalisations communes servant à une fonction commune est que le sens d'une vocalisation ne dépend plus de l'individu émetteur, une condition *a priori* nécessaire pour devenir des signaux référentiels (MARLER et al., 1992; SUZUKI, 2016), analogues aux mots du langage humain. Le partage de vocalisations peut avoir lieu à des échelles sociales plus petites et pour des fonctionnalités différentes : à l'échelle d'une population, il s'agit des dialectes, qui servent à identifier un individu comme faisant partie du groupe (BAKER et al., 2000; HUGHES et al., 2007; MUNDINGER, 1982; SEWALL, 2009; WRIGHT et DEMEURE, 2021; WRIGHT et WILKINSON, 2001). À l'échelle sociale la plus réduite, une paire d'individu peut partager un code commun, par exemple pour maintenir le contact à distance comme chez les couples de grands corbeaux (LUEF et al., 2017), ou pour s'adresser spécifiquement à un individu dans un groupe social comme chez les conures à front orange (BALSBY et al., 2012). Cependant, chez les corbeaux freux, ce code commun semble absent des cris des mâles, même au sein des couples. Même parmi les deux couples homosexuels (Elie et Tom, un couple de mâles à Strasbourg, et Bussell et Fry, un couple de femelles à Madingley), cette ressemblance n'est pas retrouvée. Par ailleurs, Bussell est même la seule exception à la similarité vocale des femelles (Chapitre 4, Fig. 4.5). Le potentiel de similarité est présent, comme montré par les cris des femelles (Chapitres 4 et 5) et dans une moindre mesure les chants des deux sexes (Chapitres 4). Par ailleurs, les deux colonies captives ont vécu plusieurs

années ensemble, certains individus ayant même été élevés ensemble depuis leur capture, et ont donc été exposées aux mêmes environnements sonores. Ils auraient donc pu apprendre les vocalisations des mêmes sources, et un partage vocal aurait pu être observé si l'acquisition du répertoire dépendait uniquement d'un apprentissage par imitation. Les deux colonies n'ont en outre jamais été en contact. Par conséquent, il est peu probable que les phénomènes observés d'idiosyncrasies soient le seul fait d'un artefact de la captivité, car il n'y aurait *a priori* aucune raison que le même phénomène se manifeste indépendamment dans les deux colonies.

Ces mêmes capacités d'apprentissage, particulièrement développées chez les corvidés, pourraient également leur permettre de se libérer des contraintes liées à un code de communication commun (FEDERSPIEL, 2010 ; LAMBERT et al., 2019). Le corbeau freux est une espèce particulière complexe socialement, avec un système de fission-fusion des colonies (CLAYTON et EMERY, 2007 ; COOMBS, 1960), mais aussi la force de la relation de couple (BOUCHERIE et al., 2016 ; MOHAMMED BENKADA et al., 2020) comparativement aux autres relations sociales (BOUCHERIE et al., 2019 ; BOUCHERIE et al., 2016). La complexité sociale, notamment dans les mesures de taille des groupes sociaux, est corrélée avec une plus forte individualité dans les vocalisations (POLLARD et BLUMSTEIN, 2011 ; SMITH-VIDAURRE et al., 2021). Chez le corbeau freux, la complexité sociale est forte, car un individu est régulièrement exposé à un grand nombre de congénères, y compris de nouveaux individus intégrant les groupes sociaux pré-existants. Les idiosyncrasies vocales des mâles pourraient alors être un mécanisme permettant à ceux-ci de rester identifiables. Cependant, le dimorphisme sexuel, où les femelles utilisent un répertoire de cris communs (même si elles restent individualisées, comme montré par l'étude des cris au nid et la performance de Rookognise), reste difficile à expliquer. En général, les vocalisations des femelles chez les oiseaux restent moins bien connues que celles des mâles (AUSTIN et al., 2021) Des études supplémentaires sur les fonctions des cris, mais aussi sur leur acquisition chez les juvéniles, seront nécessaires pour déterminer les mécanismes qui sous-tendent ces idiosyncrasies et le dimorphisme sexuel observé.

La signature vocale trouvée chez les femelles, couplée à la forte similarité inter-individuelle de leur répertoire, offre un outil prometteur de suivi des populations chez le corbeau freux, mais aussi d'autres espèces où certains individus restent à des emplacements fixes à des périodes prévisibles. La bioacoustique offre un excellent outil non invasif d'identification des individus (E. J. FOX et al., 2008 ; TERRY et al., 2005), qui pourrait être facilement déployé dans toute espèce où certains individus restent à un emplacement donné pendant au moins certaines périodes de l'année. Par exemple, les colonies de reproduction des corbeaux freux (COOMBS, 1960, 1961a), ou les territoires des rouges-gorges ou des bruants (HUGHES et al., 2007), pourraient être enregistrés continuellement en posant des microphones avant que les animaux n'arrivent. En conjonction avec des observations et un outil de type Rookognise, cela permettrait de faciliter très fortement les études bioacoustiques. Des approches de clustering comme HDSBCAN (Chapitre 4), ou des méthodes basées sur l'apprentissage automatique comme les cartes auto-organisatrices (BLUMSTEIN et al., 2011 ; PRIYADARSHANI et al., 2018) pourraient également être employées pour, par exemple, opérer des comptages rapides et automatiques des populations.

Enfin, je n'ai pas pu étudier au cours de cette thèse deux autres aspects de la complexité vo-

cale précédemment évoqué : la flexibilité contextuelle des vocalisations, et leur combinabilité. Les corvidés ont une importante complexité dans la production de leurs vocalisations (Chapitre 4, E. D. BROWN, 1985 ; BROWN et FARABAUGH, 1997 ; KONDO, 2021 ; MATES et al., 2015 ; ROSKAFT et ESPMARK, 1982). La difficulté à définir un répertoire vocal en associant vocalisations et contexte (ELLIS, 2008 ; ROSKAFT et ESPMARK, 1982 ; TANIMOTO et al., 2017) est liée au fait que la grande majorité des vocalisations n'a pas pu être associée à un contexte particulier, hormis les chants (qui pourraient eux-mêmes correspondre à un contexte particulier). La combinabilité des cris, dans le sens où certains cris pourraient être combinés en signaux dont la fonction est distincte de leurs composants, n'a pas pu être étudiée pour la même raison d'impossibilité d'associer des contextes aux productions vocales. Chez d'autres corvidés, certains cris sont combinés, principalement par répétition d'un cri unitaire (ROSKAFT et ESPMARK, 1982 ; THOMPSON, 1982 ; YORZINSKI et VEHCENCAMP, 2009). Cependant, à part chez la corneille d'Amérique où la répétition signale une urgence (YORZINSKI et VEHCENCAMP, 2009), les différences fonctionnelles entre ces répétitions et le cri unitaire qui les composent restent inconnues. Je distingue ici la combinabilité des cris à celle des chants, car les chants, s'ils sont composés d'une combinaison variée de vocalisations, sont fonctionnellement équivalents chez les oiseaux (BEECHER, 2021 ; NELSON et al., 2022), et ne rentrent donc pas dans la définition de la combinabilité utilisée ici.

### 7.3 Le chant des corbeaux freux

J'ai étudié pour la première fois le chant d'un corvidé au cours de cette thèse. Les corvidés ont longtemps été considérés comme des oiseaux chanteurs sans chant (BUGNYAR et al., 2016), malgré certaines mentions passagères dans la littérature depuis au moins un siècle (E. D. BROWN, 1985 ; BROWN et FARABAUGH, 1997 ; COOMBS, 1960 ; GOODWIN, 1986 ; WITHERBY, 1943). En effet, les chants des corvidés sont souvent décrits comme des vocalisations douces, très différentes des cris pour lesquels ils sont mieux connus, et les mentions qui en sont faites s'accordent à dire qu'ils ne sont pas produits dans les contextes usuels des chants d'oiseaux (CATCHPOLE et SLATER, 2008 ; CHEN et WIENS, 2020). Par exemple, les corbeaux freux s'éloignent activement d'une colonie, et donc de leur nid, pour chanter sur un perchoir à proximité (COOMBS, 1960). Ces chants calmes et isolés correspondent à la quasi-totalité des chants que j'ai pu observer chez les deux colonies captives, où les oiseaux qui chantaient tendent à se mettre à l'écart du reste du groupe pour chanter. Par ailleurs, tant mâles que femelles ont chanté au cours de l'étude, quoique les femelles chantaient moins fréquemment que les mâles en moyenne. Les unités vocales utilisées dans les chants d'un individu sont fréquemment différentes de ses cris (à l'exception de Merlin, Osiris, Aristotle et Bus-sel, tous très similaires entre leurs cris et leurs chants, Chapitre 4 Fig. 4.5C). Les individus d'une même colonie, même mâles, sont remarquablement similaires dans les vocalisations composant leurs chants, y compris entre sexes. La complexité du chant diffère drastiquement entre individus, certains produisant des chants de composition très simple (e.g. Merlin), alors que d'autres utilisent un répertoire élaboré et avec de nombreuses transitions différentes (e.g. Balbo). Encore une fois, je n'ai pas constaté de différence marquée dans la complexité des chants entre mâles et femelles. Cette variété



structurale apparente, l'absence de dimorphisme sexuel dans la production de chant, leur absence de fonction classique, et la production de chant hors contexte social est très similaire aux chants non-orientés retrouvés chez certaines espèces (KNEUTGEN, 1969; RITERS et al., 2022; RITERS et al., 2019; SOSSINKA et BÖHNER, 1980).

Les chants non-orientés sont souvent considérés comme étant potentiellement associées à des états émotionnels positifs. Plusieurs espèces chez qui ces chants non-orientés les produisent dans des situations calmes, sociales ou non (KNEUTGEN, 1969; SOSSINKA et BÖHNER, 1980). Chez l'étourneau, un lien étroit a été trouvé entre ces chants et la voie dopaminergique de la récompense (RITERS et al., 2022; RITERS et al., 2019; STEVENSON et al., 2020). Ces chants pourraient constituer un "entraînement" de l'oiseau pour se préparer à chanter ses chants plus classiques (SCHRAFT et al., 2017), mais ils pourraient également représenter un but en soi, voire posséder une fonction hédonique, i.e. une sorte de jeu vocal (EMERY et CLAYTON, 2015; FICKEN, 1977; STEVENSON et al., 2020). Les chants non orientés ont en effet été décrits plus variables que les autres chants du répertoire d'un individu (KNEUTGEN, 1969; RITERS et al., 2019), ce qui pourrait laisser entendre que l'individu explore plus librement, et potentiellement exerce un contrôle plus souple, sur sa production vocale dans ces chants. Une espèce comme le corbeau freux, où les chants ne correspondent qu'à ces chants non-orientés, pourrait permettre d'élucider ces questions de la fonction hédonique du chant. Cependant, l'étude des chants des freux au cours de ce travail n'a pas permis de conclure, en partie pour des raisons techniques. En terme d'organisation des vocalisations, soit la forte gradation du répertoire, soit les limites de l'approche en clustering diffus, n'ont pas permis de déterminer des catégories de vocalisations pour en étudier les séquences avec les méthodes courantes comme les alignements avec la distance de Levenshtein ou les modèles de Markov (GARLAND et al., 2017; KERSHENBAUM et al., 2014; SAINBURG et al., 2020), qui requièrent une abstraction des vocalisations en catégories (par exemple par différents types de structures acoustiques) pour être appliquées. Des techniques d'analyses futures devraient chercher à intégrer le caractère gradé des vocalisations d'espèces comme les corvidés ou les primates. Du point de vue de l'organisation temporelle, j'ai étudié les dépendances des intervalles de temps entre vocalisations successives avec les modèles ARMA ou les analyses de récurrences (RAVIGNANI et NORTON, 2017), sans pouvoir mettre au jour d'organisation particulière dans les chants des corbeaux freux. En l'état, le chant des corbeaux freux semble trop complexe pour être décrit avec les méthodes classiques.

Un aspect de la complexité vocale en lien avec la flexibilité comportementale est le contrôle de la production vocale, comme par exemple la capacité de s'ajuster à un rythme externe (KOTZ et al., 2018). Nous avons donc cherché à tester ce contrôle vocal associé au chant en diffusant des stimuli rythmiques aux corbeaux freux et en observant leur réaction, d'abord de se mettre à chanter, et ensuite d'ajuster leur chant au stimulus. Les résultats démontrent que les corvidés cherchent parfois à s'ajuster à la structure métrique ou au tempo qu'ils entendent, même si la synchronisation n'est pas du tout parfaite. Ces résultats sont très prometteurs, et représentent le premier cas connu d'entraînement (ou du moins de tentative d'entraînement) spontané chez un animal non-humain. Cette capacité d'entraînement spontané implique également l'existence d'une capacité à détecter les différentes structures rythmiques. Au cours de cette expérience, nous avons appliqué un grand nombre de combinaisons

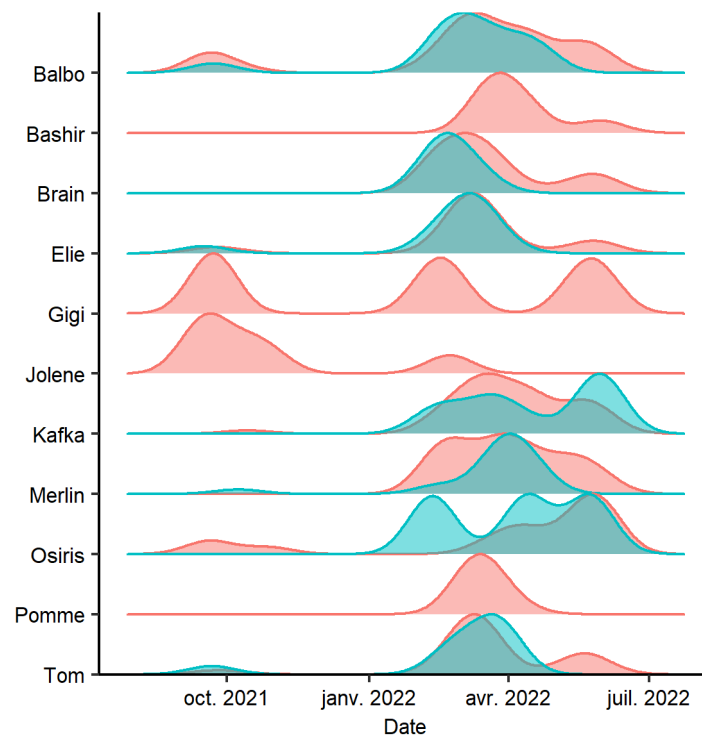


FIGURE 7.2 – Fréquence d'exposition des individus aux stimuli au fil de l'expérience. Chaque courbe correspond à la fréquence relative (i.e. le point le plus haut correspond au point où l'individu vient le plus fréquemment, mais n'est pas indicateur de la fréquence absolue), en rouge les expositions sans réponse, en bleu les expositions avec réponse vocale. En particulier, les mâles sont venus beaucoup plus fréquemment et répondaient également plus souvent au printemps après la reproduction.

tempos et de structures pour tenter de limiter l'habituation aux stimuli utilisés. Cependant, au vu du nombre total de sessions qui ont déclenché une réponse vocale, il n'est pas clair si l'habituation et la perte possible de motivation qui en découlerait seraient véritablement un problème dans les délais impartis. Un autre aspect remarquable de cette expérience est la variabilité des réponses entre les saisons (Fig. 7.2) : en particulier, l'immense majorité des réponses a été obtenue après la fin de la période de reproduction au printemps, et aucune pendant l'automne. La néophobie des corbeaux freux est influencée par les changements saisonniers des taux d'hormones, et en particulier la période de reproduction est marquée par une baisse de la néophobie (GREGGOR et al., 2016). Il serait intéressant de vérifier s'il existe un lien entre changements hormonaux, néophobie et capacités d'innovation.

Avant de conclure, je mentionnerai une dernière étude, qui n'a pas pu être achevée dans le cadre de la thèse et que j'ai donc choisi de ne pas présenter de manière détaillée. Pour résumer, cette étude consistait à tester si le chant est bien associé à un contexte émotionnel positif. Dans les conditions des colonies, nous ne pouvons pas opérer de manipulations trop invasives comme des mesures hormonales (STEVENSON et al., 2020), sous peine de stresser les oiseaux. Le stress semble inhiber la production de chant, ce qui aurait rendu impossible une mesure d'association entre chants et émotions positives, et tenter de classifier les vocalisations par valence émotionnelle (BRIEFER et al., 2022) aurait sans doute conduit à un raisonnement circulaire en l'absence de répertoire connu. La position des plumes, notamment au niveau du crâne, est un autre indicateur de l'état émotionnel chez les oiseaux (BERTIN et al., 2020 ; BERTIN et al., 2018 ; MORRIS, 1955), qui peut être observé sans manipulations

invasives, simplement sur photographies au vidéo. Grâce à une caméra très haute résolution, nous avons donc cherché à comparer les positions des plumes dans différents contextes. Pour les contextes à valence négative, nous avons réalisé des expériences de monopolisation alimentaire susceptible d'occasionner des interactions agonistiques au niveau de la nourriture donc plus faciles à filmer que dans le reste de la volière. Pour les contextes à valence neutre, nous avons observé les individus au repos ou en train de manger seuls la nourriture disponible *ad libitum* dans la volière. Pour les contextes à valence positive, nous avons observé des interactions affiliatives, notamment les toilettages. Enfin, nous avons observé les positions des plumes lors des chants. Les résultats préliminaires montrent que les chants entraîneraient une disposition des plumes plutôt intermédiaire pour les chants, comparée à la disposition des plumes au repos d'un côté, ou lors d'interactions agonistiques de l'autre (Fig. 7.3). Cependant, les mouvements corporels pendant les chants ont compliqué les analyses, et notamment la collecte de données en quantité suffisante pour conclure définitivement, car pour être comparable, la position des plumes ne pouvait être mesurée que sur une image nette et de profil des oiseaux. En outre, tous les contextes à valence négative et positive correspondaient à des interactions sociales. Il est possible que les mouvements des plumes dans ces contextes correspondent à une communication visuelle. Or, le chant des corbeaux freux semble non-social, au vu de leur éloignement des autres membres du groupe et de la faible intensité des vocalisations. Les mouvements des plumes dans ce contexte pourraient donc être également non-sociaux, et donc différer de ces contextes. Une étude plus approfondie, impliquant notamment des contextes positifs ou négatifs moins dirigés vers des interactions sociales, permettra de conclure sur la question, et de déterminer si les chants correspondent davantage à un état émotionnel positif, ou s'ils constituent un état complètement différent des autres interactions des corbeaux freux avec leurs congénères. Le chant des corbeaux freux, comme le reste de leurs vocalisations, semble donc être particulièrement atypique parmi les oiseaux. Il reste à déterminer si les corbeaux freux sont eux-mêmes atypiques ou non, parmi les autres corvidés et les autres espèces complexes socialement.

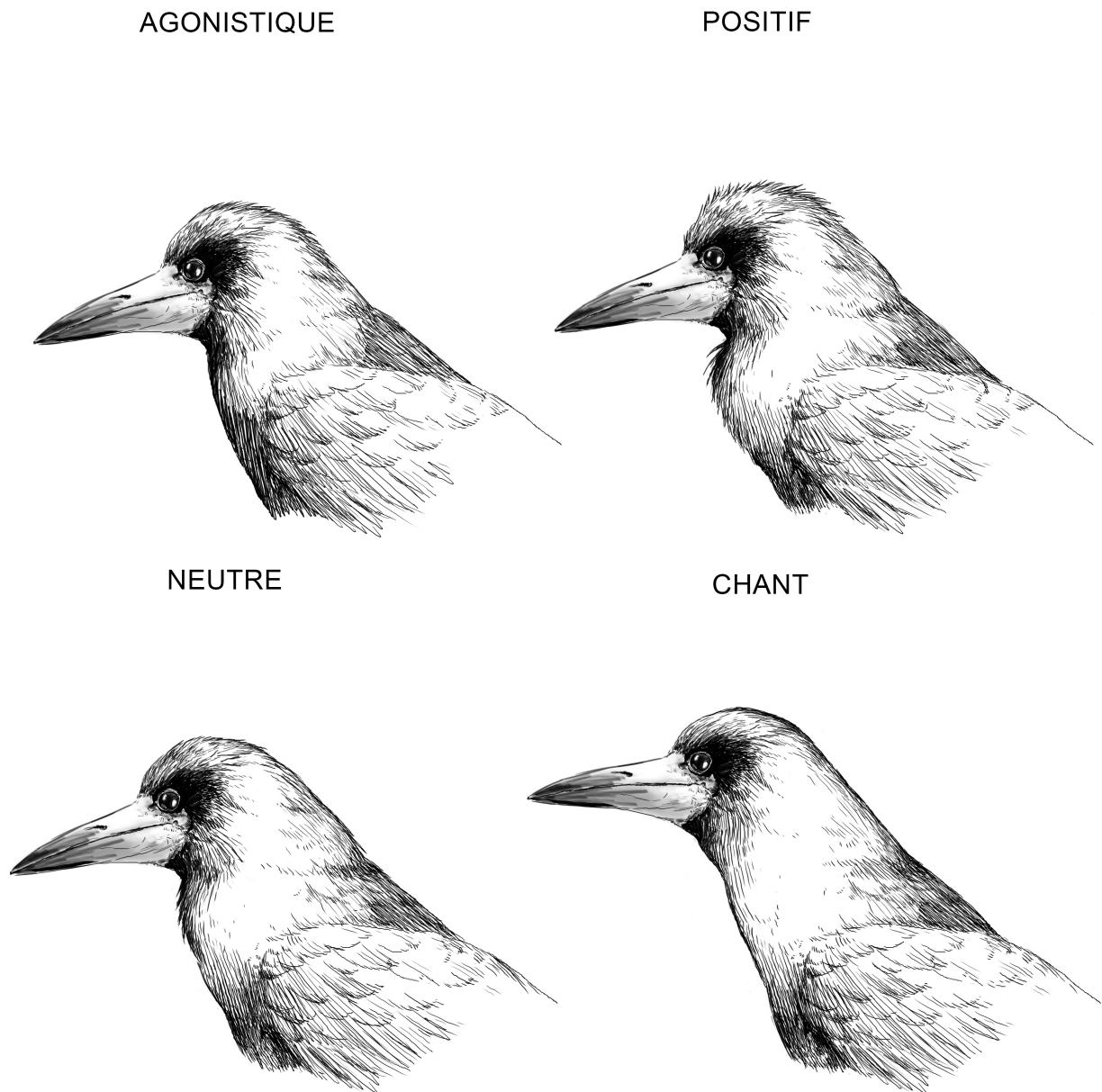


FIGURE 7.3 – Position des plumes du crâne chez le corbeau freux en fonction du de la valence du contexte comportemental. Chaque dessin correspond à une image représentative de la position des plumes. Le contexte négatif correspond à des conflits induits par une expérience de monopolisation alimentaire, le contexte positif correspond à des interactions affiliatives (e.g. toilettage), le contexte neutre correspond à la position de repos au perchoir et la prise de nourriture (nourriture habituelle, hors récompense ou conflit), et le dernier contexte est le chant. Pour ce dernier, la représentation n'est qu'une des positions observées cependant, en raison des mouvements de la tête liés à la production vocale. Crédit : A. Serda, M. Serda.

## 7.4 Conclusion

Mon travail de thèse sur le corbeau freux a démontré une forte complexité vocale dans plusieurs aspects du répertoire de l'espèce, ou peut-être de façon plus précise, des différents répertoires des individus. Ces résultats soutiennent l'hypothèse de la complexité socio-communicative chez une espèce aviaire aussi complexe socialement. Cependant, si ce travail de thèse, sur une seule espèce peu étudiée dans le monde, peut paraître limité au premier abord, de nombreuses perspectives ont été ouvertes du point de vue technique. J'ai ainsi développé plusieurs outils pour faciliter la collecte de données bio-acoustiques sur des individus connus en limitant la présence humaine, qui peut être source de biais. J'ai développé une approche générale de quantification de la complexité vocale structurale, en l'absence de répertoire défini dans une espèce. Les fortes idiosyncrasies trouvées chez les mâles corbeaux freux représentent un résultat particulièrement surprenant dans le cadre de l'étude de la communication animale. Il reste cependant à vérifier si ces idiosyncrasies sont retrouvées dans d'autres espèces. En particulier, si elles représentent une conséquence de la forte flexibilité comportementale des corvidés, ainsi que de leurs capacités d'apprentissage vocal par imitation, elles devraient être retrouvées chez les perroquets et les autres corvidés. Cela nécessiterait une étude comparative chez ces diverses espèces, en prenant en compte leurs différences d'histoire de vie, d'écologie, de complexité sociale, ou de compétences cognitives. En conjonction avec les méthodes développées dans cette thèse, cela ouvre de nouvelles voies d'étude de la communication chez les animaux, à différentes échelles spatiales et temporelles.

# Bibliographie

- ADI, K., JOHNSON, M. T., & OSIEJUK, T. S. (2010). Acoustic censusing using automatic vocalization classification and identity recognition. *The Journal of the Acoustical Society of America*, 127(2), 874-883. <https://doi.org/10.1121/1.3273887>
- ADRET, P. (1993). Vocal learning induced with operant techniques : an overview. *Netherlands Journal of Zoology*, 43(1-2), 125-142.
- ADRET, P. (2022). Developmental Plasticity in Primate Coordinated Song : Parallels and Divergences With Duetting Songbirds. *Frontiers in Ecology and Evolution*, 10(March), 1-9. <https://doi.org/10.3389/fevo.2022.862196>
- ADRET-HAUSBERGER, M. (1982). Social influences on the whistled songs of starlings. *Behavioral Ecology and Sociobiology*, 11(4), 241-246. <https://doi.org/10.1007/BF00299300>
- ADRET-HAUSBERGER, M., & GÜTTINGER, H. R. (1984). Constancy of Basic Pattern in the Songs of two Populations of Starlings (*Sturnus vulgaris*). A Comparison of Song Variation between Sedentary and Migratory Populations. *Zeitschrift für Tierpsychologie*, 66(4), 309-327. <https://doi.org/10.1111/j.1439-0310.1984.tb01371.x>
- AITKEN, P. G., & WILSON, W. A. (1979). Discriminative vocal conditioning in rhesus monkeys : Evidence for volitional control ? *Brain and Language*, 8(2), 227-240. [https://doi.org/10.1016/0093-934X\(79\)90051-8](https://doi.org/10.1016/0093-934X(79)90051-8)
- ALEXANDER, R. D. (1974). The Evolution of Social Behaviour. *Annual review of ecology and systematics*, 325-383. <https://doi.org/10.1111/eth.13212>
- AMADON, D. (1944). The genera of Corvidae and their relationships. *American Museum Novitates*, 1251(1251), 1-21.
- AMICI, F., AURELI, F., & CALL, J. (2008). Fission-Fusion Dynamics, Behavioral Flexibility, and Inhibitory Control in Primates. *Current Biology*, 18(18), 1415-1419. <https://doi.org/10.1016/j.cub.2008.08.020>
- ANDERSON, D. R. (2004). Multimodel Inference Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, 33(2), 261-304. <https://doi.org/10.1177/0049124104268644>
- ANDREW, R. J. (1962). Evolution of intelligence and vocal mimicking. *Science*, 137(3530), 585-589. <https://doi.org/10.1126/science.137.3530.585>
- ARCHIE, E. A., MOSS, C. J., & ALBERTS, S. C. (2006). The ties that bind : Genetic relatedness predicts the fission and fusion of social groups in wild African elephants. *Proceedings of the Royal Society B : Biological Sciences*, 273(1586), 513-522. <https://doi.org/10.1098/rspb.2005.3361>
- ARNOLD, K., & ZUBERBÜHLER, K. (2006). The alarm-calling system of adult male putty-nosed monkeys, *Cercopithecus nictitans martini*. *Animal Behaviour*, 72(3), 643-653. <https://doi.org/10.1016/j.anbehav.2005.11.017>
- ARROYO, B., MOUGEOT, F., & BRETAGNOLLE, V. (2001). Colonial breeding and nest defence in Montagu's harrier (*Circus pygargus*). *Behavioral Ecology and Sociobiology*, 50(2), 109-115. <https://doi.org/10.1007/s002650100342>
- AUERSPERG, A. M., OSWALD, N., DOMANEGG, M., GAJDON, G. K., & BUGNYAR, T. (2015). Unrewarded Object Combinations in Captive Parrots. *Animal Behavior and Cognition*, 1(4), 470-488. <https://doi.org/10.12966/abc.11.05.2014>

- AUERSPERG, A. M., van HORIK, J. O., BUGNYAR, T., KACELNIK, A., EMERY, N. J., & von BAYERN, A. M. P. (2014). Combinatory Actions During Object Play in Parrots (*Psittacus erithacus*) and Corvids (*Corvus*). *Journal of Comparative Psychology*, *1*, 62-71. <https://doi.org/10.1037/a0038314>
- AURELI, F., SCHAFFNER, C. M., BOESCH, C., BEARDER, S. K., CALL, J., CHAPMAN, C. A., CONNOR, R., DI FIORE, A., DUNBAR, R. I., PETER HENZI, S., HOLEKAMP, K., KORSTJENS, A. H., LAYTON, R., LEE, P., LEHMANN, J., MANSON, J. H., RAMOS-FERNANDEZ, G., STRIER, K. B., & VAN SCHAİK, C. P. (2008). Fission-fusion dynamics new research frameworks. *Current Anthropology*, *49*(4), 627-654. <https://doi.org/10.1086/586708>
- AUSTIN, V. I., DALZIELL, A. H., LANGMORE, N. E., & WELBERGEN, J. A. (2021). Avian vocalisations : the female perspective. *Biological Reviews*, *96*(4), 1484-1503. <https://doi.org/10.1111/brv.12713>
- BACIADONNA, L., CORNERO, F. M., EMERY, N. J., & CLAYTON, N. S. (2021). Convergent evolution of complex cognition : Insights from the field of avian cognition into the study of self-awareness. *Learning and Behavior*, *49*(1), 9-22. <https://doi.org/10.3758/s13420-020-00434-5>
- BACKHOUSE, F., DALZIELL, A. H., MAGRATH, R. D., & WELBERGEN, J. A. (2022). Higher-order sequences of vocal mimicry performed by male Albert's lyrebirds are socially transmitted and enhance acoustic contrast. *Proceedings of the Royal Society B : Biological Sciences*, *289*(1970). <https://doi.org/10.1098/rspb.2021.2498>
- BACKWELL, P., JENNIONS, M., PASSMORE, N., & CHRISTY, J. (1998). Synchronized courtship in fiddler crabs. *Nature*, *391*(6662), 31-32. <https://doi.org/10.1038/34076>
- BAGLIONE, V., CANESTRARI, D., MARCOS, J. M., & EKMAN, J. (2003). Kin selection in cooperative alliances of carrion crows. *Science*, *300*(5627), 1947-1949. <https://doi.org/10.1126/science.1082429>
- BAGLIONE, V., MARCOS, J. M., & CANESTRARI, D. (2002). Cooperatively breeding groups of Carrion Crow (*Corvus corone corone*) in northern Spain. *Auk*, *119*(3), 790-799. <https://doi.org/10.2307/4089974>
- BAGNALL, A., LINES, J., BOSTROM, A., LARGE, J., & KEOGH, E. (2017). The great time series classification bake off : a review and experimental evaluation of recent algorithmic advances. *Data Mining and Knowledge Discovery*, *31*(3), 606-660. <https://doi.org/10.1007/s10618-016-0483-9>
- BAKER, M. C., HOWARD, T. M., & SWEET, P. W. (2000). Microgeographic variation and sharing of the gargle vocalization and its component syllables in black-capped chickadee (*Aves*, Paridae, *Poecile atricapillus*) populations. *Ethology*, *106*(9), 819-838. <https://doi.org/10.1046/j.1439-0310.2000.00602.x>
- BAKER, M. C. (1982). Vocal dialect recognition and population genetic consequences. *Integrative and Comparative Biology*, *22*(3), 561-569. <https://doi.org/10.1093/icb/22.3.561>
- BALDA, R. P., & KAMIL, A. C. (1992). Long-term spatial memory in clark's nutcracker, *Nucifraga columbiana*. *Animal Behaviour*, *44*(4), 761-769. [https://doi.org/10.1016/S0003-3472\(05\)80302-1](https://doi.org/10.1016/S0003-3472(05)80302-1)
- BALDA, R. P., & TUREK, R. J. (1984). The cache-recovery system as an example of memory capabilities in Clark's nutcracker. *Animal Cognition*, 513-532.
- BALSBY, T. J., MOMBERG, J. V., & DABELSTEEN, T. (2012). Vocal Imitation in Parrots Allows Addressing of Specific Individuals in a Dynamic Communication Network. *PLoS ONE*, *7*(11). <https://doi.org/10.1371/journal.pone.0049747>
- BALSBY, T. J., & SCARL, J. C. (2008). Sex-specific responses to vocal convergence and divergence of contact calls in orange-fronted conures (*Aratinga canicularis*). *Proceedings of the Royal Society B : Biological Sciences*, *275*(1647), 2147-2154. <https://doi.org/10.1098/rspb.2008.0517>

- BANERJEE, A., & VALLENTIN, D. (2022). Convergent behavioral strategies and neural computations during vocal turn-taking across diverse species. *Current Opinion in Neurobiology*, 73, 102529. <https://doi.org/10.1016/j.conb.2022.102529>
- BARRETT, L., HENZI, S. P., WEINGRILL, T., LYCETT, J. E., & HILL, R. A. (1999). Market forces predict grooming reciprocity in female baboons. *Proceedings of the Royal Society B : Biological Sciences*, 266(1420), 665-670. <https://doi.org/10.1098/rspb.1999.0687>
- BARTON, K. (2020). MuMIn : Multi-Model Inference. <https://cran.r-project.org/package=MuMIn>
- BATES, D., MÄCHLER, M., BOLKER, B., & WALKER, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. <https://doi.org/10.18637/jss.v067.i01>
- BAYENS, G. (1981). The role of the sexes in territory defence in the Magpie ( *Pica pica* ). *Ardea*, 69(1), 69-82. [http://ardea.nou.nu/ardea%7B%5C\\_%7Dshow%7B%5C\\_%7Dabstract.php?lang=uk%7B%5C%7Dnr=763](http://ardea.nou.nu/ardea%7B%5C_%7Dshow%7B%5C_%7Dabstract.php?lang=uk%7B%5C%7Dnr=763)
- BEAUCHAMP, G., & FERNÁNDEZ-JURICIC, E. (2005). The group-size paradox : Effects of learning and patch departure rules. *Behavioral Ecology*, 16(2), 352-357. <https://doi.org/10.1093/beheco/arh169>
- BEBUS, S. E., SMALL, T. W., JONES, B. C., ELDERBROCK, E. K., & SCHOECH, S. J. (2016). Associative learning is inversely related to reversal learning and varies with nestling corticosterone exposure. *Animal Behaviour*, 111, 251-260. <https://doi.org/10.1016/j.anbehav.2015.10.027>
- BEECHER, M. D. (1989). Signalling systems for individual recognition : an information theory approach. *Animal Behaviour*, 38, 248-261.
- BEECHER, M. D. (2021). Why Are No Animal Communication Systems Simple Languages? *Frontiers in Psychology*, 12(March), 1-16. <https://doi.org/10.3389/fpsyg.2021.602635>
- BEECHER, M. D., & BRENOWITZ, E. A. (2005). Functional aspects of song learning in songbirds. *Trends in Ecology and Evolution*, 20(3), 143-149. <https://doi.org/10.1016/j.tree.2005.01.004>
- BEECHER, M. D., & CAMPBELL, S. E. (2005). The role of unshared songs in singing interactions between neighbouring song sparrows. *Animal Behaviour*, 70(6), 1297-1304. <https://doi.org/10.1016/j.anbehav.2005.03.008>
- BEER, C. G. (1971). Individual Recognition of Voice in the Social Behavior of Birds. *Advances in the Study of Behavior*, 3(100), 27-74. [https://doi.org/10.1016/S0065-3454\(08\)60154-0](https://doi.org/10.1016/S0065-3454(08)60154-0)
- BEISNER, B., JACKSON, M., CAMERON, A., & MCCOWAN, B. (2011). Effects of natal male alliances on aggression and power dynamics in rhesus macaques. *American Journal of Primatology*, 73(8), 790-801. <https://doi.org/10.1002/ajp.20907>
- BELLMAN, R. (1957). *Dynamic programming*. Princeton University Press.
- BENEDICT, L., & NAJAR, N. A. (2019). Are commonly used metrics of bird song complexity concordant? *Auk*, 136(1), 1-11. <https://doi.org/10.1093/auk/uky008>
- BENTI, B., CURÉ, C., & DUFOUR, V. (2019). Individual signature in the most common and context-independent call of the Rook (*Corvus frugilegus*). *The Wilson Journal of Ornithology*, 131(2), 373. <https://doi.org/10.1676/18-41>
- BERG, K. S., DELGADO, S., CORTOPASSI, K. A., BEISSINGER, S. R., & BRADBURY, J. W. (2012). Vertical transmission of learned signatures in a wild parrot. *Proceedings of the Royal Society B : Biological Sciences*, 279(1728), 585-591. <https://doi.org/10.1098/rspb.2011.0932>
- BERGER, J. (1979). "Predator Harassment" as a Defensive Strategy in Ungulates. *American Midland Naturalist*, 197-199.



- BERGER, L. R., & LIGON, J. D. (1977). Vocal communication and individual recognition in the pinon jay, *Gymnorhinus cyanocephalus*. *Animal Behaviour*, 25(PART 3), 567-584. [https://doi.org/10.1016/0003-3472\(77\)90107-5](https://doi.org/10.1016/0003-3472(77)90107-5)
- BERGMAN, T. J., & BEEHNER, J. C. (2015). Measuring social complexity. *Animal Behaviour*, 103, 203-209. <https://doi.org/10.1016/j.anbehav.2015.02.018>
- BERMANT, P. C., BRONSTEIN, M. M., WOOD, R. J., GERO, S., & GRUBER, D. F. (2019). Deep Machine Learning Techniques for the Detection and Classification of Sperm Whale Bioacoustics. *Scientific Reports*, 9(1), 1-10. <https://doi.org/10.1038/s41598-019-48909-4>
- BERTIN, A., BERAUD, A., LANSADE, L., MULOT, B., & ARNOULD, C. (2020). Bill covering and nape feather ruffling as indicators of calm states in the Sulphur-crested cockatoo (*Cacatua galerita*). *Behavioural Processes*, 178(June), 104188. <https://doi.org/10.1016/j.beproc.2020.104188>
- BERTIN, A., CORNILLEAU, F., LEMARCHAND, J., BOISSY, A., LETERRIER, C., NOWAK, R., CALANDREAU, L., BLACHE, M. C., BOIVIN, X., ARNOULD, C., & LANSADE, L. (2018). Are there facial indicators of positive emotions in birds? A first exploration in Japanese quail. *Behavioural Processes*, 157, 470-473. <https://doi.org/10.1016/j.beproc.2018.06.015>
- BERTRAM, B. (1970). The Vocal Behaviour of the Indian Hill Mynah, *Gracula religiosa*. *Animal Behaviour Monographs*, 3, 79-IN1. [https://doi.org/10.1016/s0066-1856\(70\)80003-6](https://doi.org/10.1016/s0066-1856(70)80003-6)
- BIRD, C. D., & EMERY, N. J. (2009). Insightful problem solving and creative tool modification by captive nontool-using rooks. *Proceedings of the National Academy of Sciences of the United States of America*, 106(25), 10370-10375. <https://doi.org/10.1073/pnas.0901008106>
- BITTERBAUM, E., & BAPTISTA, L. F. (1979). Geographical variation in songs of California house finches. *The Auk*, 96(3), 462-474.
- BLUMSTEIN, D. T., & ARMITAGE, K. B. (1997). Does sociality drive the evolution of communicative complexity? A comparative test with ground-dwelling sciurid alarm calls. *American Naturalist*, 150(2), 179-200. <https://doi.org/10.1086/286062>
- 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., & KIRSCHER, A. N. G. (2011). Acoustic monitoring in terrestrial environments using microphone arrays : applications , technological considerations and prospectus. *Journal of Applied Ecology*, (48), 758-767. <https://doi.org/10.1111/j.1365-2664.2011.01993.x>
- BOCK, W. (1978). Morphology of the larynx of *Corvus brachyrhynchos* (Passeriformes : Corvidae). *The Wilson Bulletin*, 90(4), 553-565. <http://www.jstor.org/stable/10.2307/4161125>
- BOECKLE, M., & BUGNYAR, T. (2012). Long-term memory for affiliates in ravens. *Current Biology*, 22(9), 801-806. <https://doi.org/10.1016/j.cub.2012.03.023>
- BOECKLE, M., SZIPL, G., & BUGNYAR, T. (2012). Who wants food? Individual characteristics in raven yells. *Animal Behaviour*, 84(5), 1123-1130. <https://doi.org/10.1016/j.anbehav.2012.08.011>
- BOECKLE, M., SZIPL, G., & BUGNYAR, T. (2018). Raven food calls indicate sender's age and sex. *Frontiers in Zoology*, 15(1), 1-9. <https://doi.org/10.1186/s12983-018-0255-z>
- BOGAARDS, N., RÖBEL, A., & RODET, X. (2004). Sound Analysis and Processing with AudioSculpt 2. *Proc. Int. Computer Music Conference (ICMC)*, 2-5.
- BOLHUIS, J. J., OKANOYA, K., & SCHARFF, C. (2010). Twitter evolution : Converging mechanisms in birdsong and human speech. *Nature Reviews Neuroscience*, 11(11), 747-759. <https://doi.org/10.1038/nrn2931>

- BOND, A. B., KAMIL, A. C., & BALDA, R. P. (2007). Serial Reversal Learning and the Evolution of Behavioral Flexibility in Three Species of North American Corvids (*Gymnorhinus cyanocephalus*, *Nucifraga columbiana*, *Aphelocoma californica*). *Journal of Comparative Psychology*, *121*(4), 372-379. <https://doi.org/10.1037/0735-7036.121.4.372>
- BOOGERT, N. J., ANDERSON, R. C., PETERS, S., SEARCY, W. A., & NOWICKI, S. (2011). Song repertoire size in male song sparrows correlates with detour reaching, but not with other cognitive measures. *Animal Behaviour*, *81*(6), 1209-1216. <https://doi.org/10.1016/j.anbehav.2011.03.004>
- BORROR, D. J. (1987). Song in the white-eyed vireo. *Wilson Bull.*, *99*(3), 377-397.
- BOUCHERIE, P. H., LORETTO, M. C., MASSEN, J. J., & BUGNYAR, T. (2019). What constitutes “social complexity” and “social intelligence” in birds? Lessons from ravens. *Behavioral Ecology and Sociobiology*, *73*(1). <https://doi.org/10.1007/s00265-018-2607-2>
- BOUCHERIE, P. H., MARIETTE, M. M., BRET, C., & DUFOUR, V. (2016). Bonding beyond the pair in a monogamous bird : Impact on social structure in adult rooks (*Corvus frugilegus*). *Behaviour*, *153*(8), 897-925. <https://doi.org/10.1163/1568539X-00003372>
- BOUCHET, H., BLOIS-HEULIN, C., & LEMASSON, A. (2013). Social complexity parallels vocal complexity : A comparison of three non-human primate species. *Frontiers in Psychology*, *4*(JUL), 1-15. <https://doi.org/10.3389/fpsyg.2013.00390>
- BOUCHET, H., BLOIS-HEULIN, C., PELLIER, A. S., ZUBERBÜHLER, K., & LEMASSON, A. (2012). Acoustic variability and individual distinctiveness in the vocal repertoire of red-capped mangabeys (*Cercocebus torquatus*). *Journal of Comparative Psychology*, *126*(1), 45-56. <https://doi.org/10.1037/a0025018>
- BOUCHET, H., PELLIER, A. S., BLOIS-HEULIN, C., & LEMASSON, A. (2010). Sex differences in the vocal repertoire of adult red-capped mangabeys (*Cercocebus torquatus*) : A multi-level acoustic analysis. *American Journal of Primatology*, *72*(4), 360-375. <https://doi.org/10.1002/ajp.20791>
- BOUGHEY, M. J., & THOMPSON, N. S. (1981). Song Variety in the Brown Thrasher (*Toxostoma rufum*). *Zeitschrift für Tierpsychologie*, *56*(1), 47-58. <https://doi.org/10.1111/j.1439-0310.1981.tb01283.x>
- BOUGHMAN, J. W. (1998). Vocal learning by greater spear-nosed bats. *Proceedings : Biological Sciences*, *265*(1392), 227-233. <https://www.jstor.org/stable/50807>
- BOUWER, F. L., NITYANANDA, V., ROUSE, A. A., & TEN CATE, C. (2021). Rhythmic abilities in humans and non-human animals : A review and recommendations from a methodological perspective. *Philosophical Transactions of the Royal Society B : Biological Sciences*, *376*(1835). <https://doi.org/10.1098/rstb.2020.0335>
- BRACKENBURY, J. H. (1980). Control of Sound Production in the Syrinx of the Fowl *Gallus Gallus*. *Journal of Experimental Biology*, *85*(1), 239-251. <https://doi.org/10.1242/jeb.85.1.239>
- BRACKENBURY, J. H. (1982). The structural basis of voice production and its relationship to sound characteristics. In *Acoustic communication in birds* (p. 53-73). Elsevier.
- BRACKENBURY, J. H. (1989). Functions of the syrinx and the control of sound production. In *Form and function in birds* (p. 193-220). Academic Press London.
- BRADBURY, J. W., & BALSBY, T. J. (2016). The functions of vocal learning in parrots. *Behavioral Ecology and Sociobiology*, *70*(3), 293-312. <https://doi.org/10.1007/s00265-016-2068-4>
- BRADBURY, J. W., CORTOPASSI, K. A., & CLEMMONS, J. R. (2001). Geographical variation in the contact calls of orange-fronted parakeets. *The Auk*, *118*(4), 958-972. <https://doi.org/10.1093/auk/118.4.958>
- BRADBURY, J. W., & VEHCAMP, S. L. (1998). *Principles of Animal Communication* (2nd Editio). Sinauer Associates, Inc.,

- BRADLEY, A. (1981). Song variation within a population of white-eyed vireos (*Vireo griseus*). *The Auk*, 98(January), 80-87.
- BRAUN, A., & BUGNYAR, T. (2012). Social bonds and rank acquisition in raven nonbreeder aggregations. *Animal Behaviour*, 84(6), 1507-1515. <https://doi.org/10.1016/j.anbehav.2012.09.024>
- BRECHT, K. F., HAGE, S. R., GAVRILOV, N., & NIEDER, A. (2019). Volitional control of vocalizations in corvid songbirds. *PLOS Biology*, 17(8), e3000375. <https://doi.org/10.1371/journal.pbio.3000375>
- BRECHT, K. F., & NIEDER, A. (2020). Parting self from others : Individual and self-recognition in birds. *Neuroscience and Biobehavioral Reviews*, 116, 99-108. <https://doi.org/10.1016/j.neubiorev.2020.06.012>
- BRECHT, K. F., WAGENER, L., OSTOJIĆ, L., CLAYTON, N. S., & NIEDER, A. (2017). Comparing the face inversion effect in crows and humans. *Journal of Comparative Physiology A : Neuroethology, Sensory, Neural, and Behavioral Physiology*, 203(12), 1017-1027. <https://doi.org/10.1007/s00359-017-1211-7>
- BREGMAN, M. R., PATEL, A. D., & GENTNER, T. Q. (2016). Songbirds use spectral shape, not pitch, for sound pattern recognition. *Proceedings of the National Academy of Sciences of the United States of America*, 113(6), 1666-1671. <https://doi.org/10.1073/pnas.1515380113>
- BRENOWITZ, E. A., MARGOLIASH, D., & NORDEEN, K. W. (1997). An introduction to birdsong and the avian song system. *Journal of Neurobiology*, 33(5), 495-500. [https://doi.org/https://doi.org/10.1002/\(SICI\)1097-4695\(19971105\)33:5<495::AID-NEU1>3.0.CO;2-{\%}23](https://doi.org/https://doi.org/10.1002/(SICI)1097-4695(19971105)33:5<495::AID-NEU1>3.0.CO;2-{\%}23)
- BRIEFER, E. F. (2012). Vocal expression of emotions in mammals : Mechanisms of production and evidence. *Journal of Zoology*, 288(1), 1-20. <https://doi.org/10.1111/j.1469-7998.2012.00920.x>
- BRIEFER, E. F., SYPHERD, C. C., LINHART, P., LELIVELD, L. M., PADILLA DE LA TORRE, M., READ, E. R., GUÉRIN, C., DEISS, V., MONESTIER, C., RASMUSSEN, J. H., ŠPINKA, M., DÜPJAN, S., BOISSY, A., JANCZAK, A. M., HILLMANN, E., & TALLET, C. (2022). Classification of pig calls produced from birth to slaughter according to their emotional valence and context of production. *Scientific Reports*, 12(1), 1-10. <https://doi.org/10.1038/s41598-022-07174-8>
- BROCKELMAN, W. Y., & SCHILLING, D. (1984). Inheritance of stereotyped gibbon calls. *Nature*, 312(5995), 634-636. <https://doi.org/10.1038/312634a0>
- BROWN, C. R. (1986). Cliff swallow colonies as information centers. *Science*, 234(4772), 83-85. <https://doi.org/10.1126/science.234.4772.83>
- BROWN, E. D. (1985). The Role of Song and Vocal Imitation among Common Crows (<i>Corvus brachyrhynchos</i>). *Zeitschrift für Tierpsychologie*, 68(2), 115-136. <https://doi.org/10.1111/j.1439-0310.1985.tb00119.x>
- BROWN, E. D., & FARABAUGH, S. M. (1997). What birds with complex social relationships can tell us about vocal learning : Vocal sharing in avian groups. In C. T. SNOWDON & M. HAUSBERGER (Éd.), *Social Influences on Vocal Development* (p. 98-127). Cambridge University Press. <https://doi.org/10.1017/cbo9780511758843.007>
- BROWN, R. L. (2014). Identifying Behavioral Novelty. *Biological Theory*, 9(2), 135-148. <https://doi.org/10.1007/s13752-013-0150-y>
- BRUMM, H., & TODT, D. (2002). Noise-dependent song amplitude regulation in a territorial songbird. *Animal Behaviour*, 63(5), 891-897. <https://doi.org/10.1006/anbe.2001.1968>
- BUCK, J. (1988). Synchronous flashing of fireflies. *The Quarterly Review of Biology*, 63(3), 265-289. <https://doi.org/10.1086/415929>

- BUGNYAR, T., KIJNE, M., & KOTRSCHAL, K. (2001). Food calling in ravens : are yells referential signals? *Animal Behaviour*, *61*, 949-958. <https://doi.org/10.1006/anbe.2000.1668>
- BUGNYAR, T., & KOTRSCHAL, K. (2002). Observational learning and the raiding of food caches in ravens, *Corvus corax* : Is it 'tactical' deception? *Animal Behaviour*, *64*(2), 185-195. <https://doi.org/10.1006/anbe.2002.3056>
- BUGNYAR, T., REBER, S. A., & BUCKNER, C. (2016). Ravens attribute visual access to unseen competitors. *Nature Communications*, *7*, 1-6. <https://doi.org/10.1038/ncomms10506>
- BURISH, M. J., KUEH, H. Y., & WANG, S. S. (2004). Brain Architecture and Social Complexity in Modern and Ancient Birds. *Brain, Behavior and Evolution*, *63*(2), 107-124. <https://doi.org/10.1159/000075674>
- BURNS, P. S. (1957). Rook and jackdaw roosts around bishop's stortford. *Bird Study*, *4*(2), 62-71. <https://doi.org/10.1080/00063655709475872>
- BUSSE, P. (1965). Nest building dynamics of a breeding colony of rook (*Corvus frugilegus* L.) *Ekologia Polska. Seria A* *13*, 25, 491-514.
- BUSSE, P. (1969). Results of ringing of European Corvidae. *Acta Ornithologica*, *11*, 236-328.
- CALCARI, C., PILENGA, C., BACIADONNA, L., GAMBA, M., & FAVARO, L. (2021). Long-term stability of vocal individuality cues in a territorial and monogamous seabird. *Animal Cognition*, *24*(6), 1165-1169. <https://doi.org/10.1007/s10071-021-01518-z>
- CAMPELLO, R. J., MOULAVI, D., ZIMEK, A., & SANDER, J. (2015). Hierarchical density estimates for data clustering, visualization, and outlier detection. *ACM Transactions on Knowledge Discovery from Data*, *10*(1), 1-51. <https://doi.org/10.1145/2733381>
- CAMPOS, I. B., FEWSTER, R., LANDERS, T., TRUSKINGER, A., TOWSEY, M., ROE, P., LEE, W., & GASKETT, A. (2022). Acoustic region workflow for efficient comparison of soundscapes under different invasive mammals' management regimes. *Ecological Informatics*, *68*, 101554. <https://doi.org/https://doi.org/10.1016/j.ecoinf.2022.101554>
- CANDIOTTI, A., ZUBERBÜHLER, K., & LEMASSON, A. (2012). Convergence and divergence in Diana monkey vocalizations. *Biology Letters*, *8*(3), 382-385. <https://doi.org/10.1098/rsbl.2011.1182>
- CANDIOTTI, A., ZUBERBÜHLER, K., & LEMASSON, A. (2013). Voice discrimination in four primates. *Behavioural Processes*, *99*, 67-72. <https://doi.org/10.1016/j.beproc.2013.06.010>
- CAROUSO-PECK, S., & GOLDSTEIN, M. H. (2019). Female Social Feedback Reveals Non-imitative Mechanisms of Vocal Learning in Zebra Finches. *Current Biology*, *29*(4), 631-636.e3. <https://doi.org/10.1016/j.cub.2018.12.026>
- CARUANA, R. (1997). *Multitask Learning* (thèse de doct.). Carnegie Mellon University. [https://doi.org/10.1007/978-1-4899-7687-1\\_100322](https://doi.org/10.1007/978-1-4899-7687-1_100322)
- CATCHPOLE, C. K., & SLATER, P. J. (2008). *Bird song : Biological themes and variations, second edition*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511754791>
- CAZAU, D., ADAM, O., AUBIN, T., LAITMAN, J. T., & REIDENBERG, J. S. (2016). A study of vocal nonlinearities in humpback whale songs : From production mechanisms to acoustic analysis. *Scientific Reports*, *6*(June), 1-12. <https://doi.org/10.1038/srep31660>
- CELMA-MIRALLES, A., & TORO, J. M. (2020). Discrimination of temporal regularity in rats (*Rattus norvegicus*) and humans (*Homo sapiens*). *Journal of Comparative Psychology*, *134*(1), 3-10. <https://doi.org/10.1037/com0000202>
- CHAMBERLAIN, D. R., & CORNWELL, G. W. (1971). Selected vocalizations of the common crow. *The Auk*, *88*(3), 613-634.

- CHAO, A., CHAZDON, R. L., COLWELL, R. K., & SHEN, T. J. (2006). Abundance-based similarity indices and their estimation when there are unseen species in samples. *Biometrics*, 62(2), 361-371. <https://doi.org/10.1111/j.1541-0420.2005.00489.x>
- CHAO, A., CHIU, C. H., & JOST, L. (2014). Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through hill numbers. *Annual Review of Ecology, Evolution, and Systematics*, 45, 297-324. <https://doi.org/10.1146/annurev-ecolsys-120213-091540>
- CHARRIER, I., JOUVENTIN, P., MATHEVON, N., & AUBIN, T. (2001). Individual identity coding depends on call type in the South Polar skua *Catharacta maccormicki*. *Polar Biology*, 24(5), 378-382. <https://doi.org/10.1007/s003000100231>
- CHARRIER, I., MARCHESSEAU, S., DENDRINOS, P., TOUNTA, E., & KARAMANLIDIS, A. A. (2017). Individual signatures in the vocal repertoire of the endangered Mediterranean monk seal : New perspectives for population monitoring. *Endangered Species Research*, 32(1), 459-470. <https://doi.org/10.3354/esr00829>
- CHASE, A. R. (2001). Music discriminations by carp (*Cyprinus carpio*). *Animal Learning and Behavior*, 29(4), 336-353. <https://doi.org/10.3758/BF03192900>
- CHAUVIGNÉ, L. A., GITAU, K. M., & BROWN, S. (2014). The neural basis of audiomotor entrainment : An ALE meta-analysis. *Frontiers in Human Neuroscience*, 8(SEP), 1-18. <https://doi.org/10.3389/fnhum.2014.00776>
- CHEN, Y., MATHESON, L. E., & SAKATA, J. T. (2016). Mechanisms underlying the social enhancement of vocal learning in songbirds. *Proceedings of the National Academy of Sciences of the United States of America*, 113(24), 6641-6646. <https://doi.org/10.1073/pnas.1522306113>
- CHEN, Z., & WIENS, J. J. (2020). The origins of acoustic communication in vertebrates. *Nature Communications*, 11(1), 1-8. <https://doi.org/10.1038/s41467-020-14356-3>
- CHENEY, D. L., & SEYFARTH, R. M. (2018). Flexible usage and social function in primate vocalizations. *Proceedings of the National Academy of Sciences of the United States of America*, 115(9), 1974-1979. <https://doi.org/10.1073/pnas.1717572115>
- CHENG, J., XIE, B., LIN, C., & JI, L. (2012). A comparative study in birds : Call-type-independent species and individual recognition using four machine-learning methods and two acoustic features. *Bioacoustics*, 21(2), 157-171. <https://doi.org/10.1080/09524622.2012.669664>
- CHENG, K. (2002). Generalisation : mechanistic and functional explanations. *Animal cognition*, 5(1), 33-40. <https://doi.org/10.1007/s10071-001-0122-7>
- CHERRY, C. (1955). Communication Theory and Human Behavior. *Studies in communication*, 45-67.
- CHIANDETTI, C., & VALLORTIGARA, G. (2011). Chicks like consonant music. *Psychological Science*, 22(10), 1270-1273. <https://doi.org/10.1177/0956797611418244>
- CHOW, C. P., MITCHELL, J. F., & MILLER, C. T. (2015). Vocal turn-taking in a non-human primate is learned during ontogeny. *Proceedings of the Royal Society B : Biological Sciences*, 282(1807). <https://doi.org/10.1098/rspb.2015.0069>
- CHRISTIN, S., HERVET, É., & LECOMTE, N. (2019). Applications for deep learning in ecology. *Methods in Ecology and Evolution*, 10(April), 1632-1644. <https://doi.org/10.1111/2041-210X.13256>
- CIBULSKI, L., WASCHER, C. A., WEISS, B. M., & KOTRSCHAL, K. (2014). Familiarity with the experimenter influences the performance of Common ravens (*Corvus corax*) and Carrion crows (*Corvus corone co-*

- rone) in cognitive tasks. *Behavioural Processes*, 103, 129-137. <https://doi.org/10.1016/j.beproc.2013.11.013>
- CLAY, Z., ARCHBOLD, J., & ZUBERBÜHLER, K. (2015). Functional flexibility in wild bonobo vocal behaviour. *PeerJ*, 2015(8), 1-20. <https://doi.org/10.7717/peerj.1124>
- CLAY, Z., SMITH, C. L., & BLUMSTEIN, D. T. (2012). Food-associated vocalizations in mammals and birds : What do these calls really mean ? *Animal Behaviour*, 83(2), 323-330. <https://doi.org/10.1016/j.anbehav.2011.12.008>
- CLAYTON, N. S. (1990). The effects of cross-fostering on assortative mating between zebra finch subspecies. *Animal Behaviour*, 40(6), 1102-1110. [https://doi.org/10.1016/S0003-3472\(05\)80176-9](https://doi.org/10.1016/S0003-3472(05)80176-9)
- CLAYTON, N. S., & EMERY, N. J. (2007). The social life of corvids. *Current Biology*, 17(16), 652-656.
- CLUTTON-BROCK, T. H., & HUCHARD, E. (2013). Social competition and selection in males and females. *Philosophical Transactions of the Royal Society B : Biological Sciences*, 368(1631). <https://doi.org/10.1098/rstb.2013.0074>
- CLUTTON-BROCK, T., & SHELDON, B. C. (2010). Individuals and populations : The role of long-term, individual-based studies of animals in ecology and evolutionary biology. *Trends in Ecology and Evolution*, 25(10), 562-573. <https://doi.org/10.1016/j.tree.2010.08.002>
- COHEN, Y., NICHOLSON, D. A., SANCHIONI, A., MALLABER, E. K., SKIDANOVA, V., & GARDNER, T. J. (2022). Automated annotation of birdsong with a neural network that segments spectrograms. *eLife*, 11, 1-32. <https://doi.org/10.7554/eLife.63853>
- CONNOR, R. C., WELLS, R. S., MANN, J., & READ, A. J. (2000). Social Relationships in a Fission-Fusion Society. *Cetacean Societies : Field Studies of Dolphins and Whales*, 91-126.
- CONNOR, R. C. (2007). Dolphin social intelligence : Complex alliance relationships in bottlenose dolphins and a consideration of selective environments for extreme brain size evolution in mammals. *Philosophical Transactions of the Royal Society B : Biological Sciences*, 362(1480), 587-602. <https://doi.org/10.1098/rstb.2006.1997>
- CONNOR, R. C., SMOLKER, R., & BEJDER, L. (2006). Synchrony, social behaviour and alliance affiliation in Indian Ocean bottlenose dolphins, *Tursiops aduncus*. *Animal Behaviour*, 72(6), 1371-1378. <https://doi.org/10.1016/j.anbehav.2006.03.014>
- CONRADY, C. R., ER, Ş., ATTWOOD, C. G., ROBERSON, L. A., & de VOS, L. (2022). Automated detection and classification of southern African Roman seabream using mask R-CNN. *Ecological Informatics*, 69, 101593. <https://doi.org/https://doi.org/10.1016/j.ecoinf.2022.101593>
- COOK, N. D., & FUJISAWA, T. X. (2006). The psychophysics of harmony perception : harmony is a three-tone phenomenon. *Empirical Musicology Review*, 1(2), 106-126. <https://doi.org/10.18061/1811/24149>
- COOK, P., ROUSE, A., WILSON, M., & REICHMUTH, C. (2013). A California sea lion (*Zalophus californianus*) can keep the beat : Motor entrainment to rhythmic auditory stimuli in a non vocal mimic. *Journal of Comparative Psychology*, 127(4), 412-427. <https://doi.org/10.1037/a0032345>
- COOMBS, C. J. F. (1960). Observations on the Rook *Corvus Frugilegus* in Southwest Cornwall. *Ibis*, 102(3), 394-419. <https://doi.org/10.1111/j.1474-919X.1960.tb08417.x>
- COOMBS, C. J. F. (1961a). Rookeries and roosts of the rook and jackdaw in south-west cornwall. *Bird Study*, 8(1), 32-37. <https://doi.org/10.1080/00063656109475986>
- COOMBS, C. J. F. (1961b). Rookeries and Roosts of the Rook and Jackdaw in South-West Cornwall. Part II. Roosting. *Bird Study*, 8(2), 55-70. <https://doi.org/10.1080/00063656109475989>

- CORTOPASSI, K. A., & BRADBURY, J. W. (2006). Contact call diversity in wild orange-fronted parakeet pairs, *Aratinga canicularis*. *Animal Behaviour*, 71(5), 1141-1154. <https://doi.org/10.1016/j.anbehav.2005.09.011>
- COUZIN, I. D., KRAUSE, J., JAMES, R., RUXTON, G. D., & FRANKS, N. R. (2002). Collective memory and spatial sorting in animal groups. *Journal of Theoretical Biology*, 218(1), 1-11. <https://doi.org/10.1006/jtbi.2002.3065>
- CRUICKSHANK, A. J., GAUTIER, J.-P., & CHAPPUIS, C. (1993). Vocal mimicry in wild African Grey Parrots *Psittacus erithacus*. *Ibis*, 135(3), 293-299. <https://doi.org/10.1111/j.1474-919X.1993.tb02846.x>
- CURÉ, C., AUBIN, T., & MATHEVON, N. (2009). Acoustic convergence and divergence in two sympatric burrowing nocturnal seabirds. *Biological Journal of the Linnean Society*, 96(1), 115-134. <https://doi.org/10.1111/j.1095-8312.2008.01104.x>
- CURIO, E. (2012). *The ethology of predation*. Springer Science & Business Media.
- CYNX, J., WILLIAMS, H., & NOTTEBOHM, F. (1990). Timbre discrimination in zebra finch (*Taeniopygia guttata*) song syllables. *Journal of Comparative Psychology*, 104(4), 303-308. <https://doi.org/10.1037/0735-7036.104.4.303>
- CYNX, J., LEWIS, R., TAVEL, B., & TSE, H. (1998). Amplitude regulation of vocalizations in noise by a songbird, *Taeniopygia guttata*. *Animal Behaviour*, 56(1), 107-113. <https://doi.org/10.1006/anbe.1998.0746>
- DAHL, J. A. (2015). *Responses of blue jays (Cyanocitta cristata) to raptors that differ in predatory threat* (thèse de doct.). <https://doi.org/10.3184/175815618X15263796305056>
- DAHLIN, C. R., & BENEDICT, L. (2014). Angry birds need not apply : A perspective on the flexible form and multifunctionality of avian vocal duets. *Ethology*, 120(1), 1-10. <https://doi.org/10.1111/eth.12182>
- DALISIO, A. C., JENSEN, W. E., & PARKER, T. H. (2015). Divergence of vocal culture among isolated alpine habitats is inconsistent among three Oscine species. *Journal of Ornithology*, 156(1), 165-178. <https://doi.org/10.1007/s10336-014-1099-8>
- DALLY, J. M., CLAYTON, N. S., & EMERY, N. J. (2006). The behaviour and evolution of cache protection and pilferage. *Animal Behaviour*, 72(1), 13-23. <https://doi.org/10.1016/j.anbehav.2005.08.020>
- DALLY, J. M., CLAYTON, N. S., & EMERY, N. J. (2008). Social influences on foraging by rooks (*Corvus frugilegus*). *Behaviour*, 145(8), 1101-1124. <https://doi.org/10.1163/156853908784474470>
- DALZIELL, A. H., & MAGRATH, R. D. (2012). Fooling the experts : Accurate vocal mimicry in the song of the superb lyrebird, *Menura novaehollandiae*. *Animal Behaviour*, 83(6), 1401-1410. <https://doi.org/10.1016/j.anbehav.2012.03.009>
- DANCHIN, E., & WAGNER, R. H. (1997). The evolution of coloniality : The emergence of new perspectives. *Trends in Ecology and Evolution*, 12(9), 342-347. [https://doi.org/10.1016/S0169-5347\(97\)01124-5](https://doi.org/10.1016/S0169-5347(97)01124-5)
- DANIEL MELIZA, C., KEEN, S. C., & RUBENSTEIN, D. R. (2013). Pitch- and spectral-based dynamic time warping methods for comparing field recordings of harmonic avian vocalizations. *The Journal of the Acoustical Society of America*, 134(2), 1407-1415. <https://doi.org/10.1121/1.4812269>
- DARRAS, K., BATÁRY, P., FURNAS, B. J., GRASS, I., MULYANI, Y. A., & TSCHARNTKE, T. (2019). Autonomous sound recording outperforms human observation for sampling birds : a systematic map and user guide. *Ecological Applications*, 29(6). <https://doi.org/10.1002/eap.1954>
- DARWIN, C. (1897). *The Expression of the Emotions in Man and Animals*. [http://darwin-online.org.uk/converted/pdf/1897%7B%5C\\_%7DExpression%7B%5C\\_%7DF1152.pdf](http://darwin-online.org.uk/converted/pdf/1897%7B%5C_%7DExpression%7B%5C_%7DF1152.pdf)

- de la MORA, D. M., NESPOR, M., & TORO, J. M. (2013). Do humans and nonhuman animals share the grouping principles of the iambic-trochaic law? *Attention, Perception, and Psychophysics*, 75, 92-100. <https://doi.org/10.3758/s13414-012-0371-3>
- DE VRIES, H., STEVENS, J. M., & VERVAECKE, H. (2006). Measuring and testing the steepness of dominance hierarchies. *Animal Behaviour*, 71(3), 585-592. <https://doi.org/10.1016/j.anbehav.2005.05.015>
- de KORT, S. R., & CLAYTON, N. S. (2006). An evolutionary perspective on caching by corvids. *Proceedings of the Royal Society B : Biological Sciences*, 273, 417-423. <https://doi.org/10.1098/rspb.2005.3350>
- de KORT, S. R., ELDERMIRE, E. R., CRAMER, E. R., & VEHRENCAMP, S. L. (2009). The deterrent effect of bird song in territory defense. *Behavioral Ecology*, 20(October), 200-206. <https://doi.org/10.1093/beheco/arn135>
- de KORT, S. R., EMERY, N. J., & CLAYTON, N. S. (2006). Food sharing in jackdaws, *Corvus monedula* : what, why and with whom? *Animal Behaviour*, 72, 297-304. <https://doi.org/10.1016/j.anbehav.2005.10.016>
- DEONIZIAK, K., & OSIEJUK, T. S. (2019). Habitat-related differences in song structure and complexity in a songbird with a large repertoire. *BMC Ecology*, 19(1), 1-11. <https://doi.org/10.1186/s12898-019-0255-7>
- DEVOOGD, T. J., KREBS, J. R., HEALY, S. D., & PURVIS, A. (1993). Relations between song repertoire size and the volume of brain nuclei related to song : Comparative evolutionary analyses amongst oscine birds. *Proceedings of the Royal Society B : Biological Sciences*, 254(1340), 75-82. <https://doi.org/10.1098/rspb.1993.0129>
- DEVOOGD, T. J. (2004). Neural constraints on the complexity of avian song. *Brain, Behavior and Evolution*, 63(4), 221-232. <https://doi.org/10.1159/000076783>
- de WALL, F. B. M., & van ROOSMALEN, A. (1979). Reconciliation and consolation among chimpanzees. *Behavioral Ecology and Sociobiology*, 5, 55-66.
- DIAMOND, J., & BOND, A. B. (2003). A comparative analysis of social play in birds. *Behaviour*, 140(8-9), 1091-1115. <https://doi.org/10.1163/156853903322589650>
- DOOLING, R. J., LEEK, M. R., GLEICH, O., & DENT, M. L. (2002). Auditory temporal resolution in birds : Discrimination of harmonic complexes. *The Journal of the Acoustical Society of America*, 112(2), 748-759. <https://doi.org/10.1121/1.1494447>
- DOOLING, R. J., & PRIOR, N. H. (2017). Do we hear what birds hear in birdsong? *Animal Behaviour*, 124, 283-289. <https://doi.org/10.1016/j.anbehav.2016.10.012>
- DOOLITTLE, E. L. (2020). "Hearken to the Hermit-Thrush"<sup>1</sup> : A Case Study in Interdisciplinary Listening. *Frontiers in Psychology*, 11(December). <https://doi.org/10.3389/fpsyg.2020.613510>
- DOUPE, A. J., & KUHL, P. K. (1999). Birdsong and human speech : Common themes and mechanisms. *Annual Review of Neuroscience*, 22, 567-631. <https://doi.org/10.1146/annurev.neuro.22.1.567>
- DUBOIS, A. L., NOWICKI, S., PETERS, S., RIVERA-CÁCERES, K. D., & SEARCY, W. A. (2018). Song is not a reliable signal of general cognitive ability in a songbird. *Animal Behaviour*, 137, 205-213. <https://doi.org/10.1016/j.anbehav.2018.01.020>
- DUFOUR, V., PASQUARETTA, C., GAYET, P., & STERCK, E. H. (2017). The extraordinary nature of Barney's drumming : A complementary study of ordinary noise making in chimpanzees. *Frontiers in Neuroscience*, 11(JAN). <https://doi.org/10.3389/fnins.2017.00002>
- DUFOUR, V., WASCHER, C. A., BRAUN, A., MILLER, R., & BUGNYAR, T. (2012). Corvids can decide if a future exchange is worth waiting for. *Biology Letters*, 8(September), 201-204. <https://doi.org/10.1098/rsbl.2011.0726>



- DUFOURQ, E., BATIST, C., FOQUET, R., & DURBACH, I. (2022). Passive acoustic monitoring of animal populations with transfer learning. *Ecological Informatics*, 70, 101688. <https://doi.org/https://doi.org/10.1016/j.ecoinf.2022.101688>
- DUNBAR, R. I. (1998). The Social Brain Hypothesis. *Evolutionary Anthropology*.
- DUNBAR, R. I., & SHULTZ, S. (2010). Bondedness and sociality. *Behaviour*, 147(7), 775-803. <https://doi.org/10.1163/000579510X501151>
- DUNN, A. M., & ZANN, R. A. (1996). Undirected song in wild zebra finch flocks : Contexts and effects of mate removal. *Ethology*, 102(4), 529-539. <https://doi.org/10.1111/j.1439-0310.1996.tb01145.x>
- DUNNET, G. M., FORDHAM, R. A., & PATTERSON, I. J. (1969). Ecological studies of the rook (*Corvus frugilegus* L.) in North-East Scotland : proportion and distribution of young in the population. *Journal of Applied Ecology*, 6(9), 459-473.
- EAST, M. (1988). Crop selection, feeding skills and risks taken by adult and juvenile Rooks *Corvus frugilegus*. *Ibis*, 130(2), 294-299. <https://doi.org/10.1111/j.1474-919X.1988.tb00981.x>
- EBERLE, M., & KAPPELER, P. M. (2006). Family insurance : Kin selection and cooperative breeding in a solitary primate (*Microcebus murinus*). *Behavioral Ecology and Sociobiology*, 60(4), 582-588. <https://doi.org/10.1007/s00265-006-0203-3>
- EDELAAR, P., & BENKMAN, C. W. (2006). Replicated population divergence caused by localized coevolution ? A test of three hypotheses in the red crossbill-lodgepole pine system. *Journal of Evolutionary Biology*, 19(5), 1651-1659. <https://doi.org/10.1111/j.1420-9101.2006.01113.x>
- EENS, M., PINXTEN, R., & VERHEYEN, R. F. (1992). Song learning in captive European starlings, *Sturnus vulgaris*. *Animal Behaviour*, 44(6), 1131-1143. [https://doi.org/10.1016/S0003-3472\(05\)80325-2](https://doi.org/10.1016/S0003-3472(05)80325-2)
- ELIE, J. E., MARIETTE, M. M., SOULA, H. A., GRIFFITH, S. C., MATHEVON, N., & VIGNAL, C. (2010). Vocal communication at the nest between mates in wild zebra finches : A private vocal duet ? *Animal Behaviour*, 80(4), 597-605. <https://doi.org/10.1016/j.anbehav.2010.06.003>
- ELIE, J. E., & THEUNISSEN, F. E. (2016). The vocal repertoire of the domesticated zebra finch : a data driven approach to decipher the information-bearing acoustic features of communication signals. *Animal Cognition*, 19(2), 285-315. <https://doi.org/10.1007/s10071-015-0933-6>
- ELIE, J. E., & THEUNISSEN, F. E. (2018). Zebra finches identify individuals using vocal signatures unique to each call type. *Nature Communications*, 9(1). <https://doi.org/10.1038/s41467-018-06394-9>
- ELLIS, J. M. (2008). *The Vocal Repertoire of the White-Throated Magpie-Jay (<i>Calocitta Formosa</i>)* (thèse de doct. May). Cornell University.
- EMERY, N. J. (2004). Are corvids 'feathered apes' ? Cognitive evolution in crows, rooks and jackdaws. *Comparative Analysis of Mind*, 181-213.
- EMERY, N. J., & CLAYTON, N. S. (2004). The mentality of crows : Convergent evolution of intelligence in corvids and apes. *Science*, 306(5703), 1903-1907. <https://doi.org/10.1126/science.1098410>
- EMERY, N. J., & CLAYTON, N. S. (2015). Do birds have the capacity for fun ? *Current Biology*, 25(1), R16-R20. <https://doi.org/10.1016/j.cub.2014.09.020>
- EMERY, N. J., SEED, A. M., VON BAYERN, A. M., & CLAYTON, N. S. (2007). Cognitive adaptations of social bonding in birds. *Philosophical Transactions of the Royal Society B : Biological Sciences*, 362(1480), 489-505. <https://doi.org/10.1098/rstb.2006.1991>
- EMLÉN, S. T. (1997). Predicting family dynamics in social vertebrates. In J. R. KREBS & N. B. DAVIES (Éd.), *Behavioural ecology : an evolutionary approach* (4th, p. 228-253). Blackwell.

- ENDLER, J. A., & BASOLO, A. L. (1998). Sensory ecology, receiver biases and sexual selection. *Trends in Ecology and Evolution*, 13(10), 415-420.
- ENGESSER, S., RIDLEY, A. R., MANSER, M. B., MANSER, A., & TOWNSEND, S. W. (2018). Internal acoustic structuring in pied babbler recruitment cries specifies the form of recruitment. *Behavioral Ecology*, 29(5), 1021-1030. <https://doi.org/10.1093/beheco/ary088>
- ENGGIST-DUEBLIN, P., & PFISTER, U. (2002). Cultural transmission of vocalizations in ravens, *Corvus corax*. *Animal Behaviour*, 64(6), 831-841. <https://doi.org/10.1006/anbe.2002.2016>
- ERIKSEN, A., LAMPE, H. M., & SLAGSVOLD, T. (2009). Interspecific cross-fostering affects song acquisition but not mate choice in pied flycatchers, *Ficedula hypoleuca*. *Animal Behaviour*, 78(4), 857-863. <https://doi.org/10.1016/j.anbehav.2009.07.005>
- EY, E., PFEFFERLE, D., & FISCHER, J. (2007). Do age- and sex-related variations reliably reflect body size in non-human primate vocalizations? A review. *Primates*, 48(4), 253-267. <https://doi.org/10.1007/s10329-006-0033-y>
- FAGERLUND, S., & HÄRMÄ, A. (2005). Parametrization of inharmonic bird sounds for automatic recognition. *13th European Signal Processing Conference, EUSIPCO 2005*, (June), 1039-1042.
- FANIOUDAKIS, L., & POTAMITIS, I. (2017). Deep Networks tag the location of bird vocalisations on audio spectrograms. <http://arxiv.org/abs/1711.04347>
- FARRELL, T. M., WEAVER, K., AN, Y. S., & MACDOUGALL-SHACKLETON, S. A. (2012). Song bout length is indicative of spatial learning in European starlings. *Behavioral Ecology*, 23(1), 101-111. <https://doi.org/10.1093/beheco/arr162>
- FAVARO, L., GAMBA, M., GILI, C., & PESSANI, D. (2017). Acoustic correlates of body size and individual identity in banded penguins. *PLoS ONE*, 12(2). <https://doi.org/10.1371/journal.pone.0170001>
- FAWCETT, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861-874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- FEARE, C. J., DUNNET, G. M., & PATTERSON, I. J. (1974). Ecological Studies of the Rook (*Corvus frugilegus* L.) in North-East Scotland : Food Intake and Feeding Behaviour. *The Journal of Applied Ecology*, 11(3), 867. <https://doi.org/10.2307/2401751>
- FEARE, C. J. (1978). The ecology of damage by rooks (*Corvusfrugilegus*). *Annals of Applied Biology*, 88(2), 329-334. <https://doi.org/10.1111/j.1744-7348.1978.tb00714.x>
- FEDERSPIEL, I. G. (2010). *Sociality , Social Learning and Individual Differences in Rooks , Jackdaws and Eurasian Jays* (thèse de doct.).
- FEEKES, F. (1977). Colony specific song in *Cacicus cela* (Icteridae, Aves) : the password hypothesis. *Ardea*, 65, 197-202.
- FERREIRA, A. C., SILVA, L. R., RENNA, F., BRANDL, H. B., RENOULT, J. P., FARINE, D. R., COVAS, R., & DOUTRELANT, C. (2020). Deep learning-based methods for individual recognition in small birds. *Methods in Ecology and Evolution*, 11(9), 1072-1085. <https://doi.org/10.1111/2041-210X.13436>
- FICKEN, M. S. (1977). Avian play. *The Auk*, 94(July), 573-582.
- FICKEN, M. S., WEISE, C. M., & REINARTZ, J. A. (1987). A Complex Vocalization of the Black-Capped Chickadee. II. Repertoires, Dominance and Dialects. *The Condor*, 89(3), 500. <https://doi.org/10.2307/1368640>
- FISCHER, J., WADEWITZ, P., & HAMMERSCHMIDT, K. (2017). Structural variability and communicative complexity in acoustic communication. *Animal Behaviour*, 134, 229-237. <https://doi.org/10.1016/j.anbehav.2016.06.012>

- FITCH, W. T. (2009). Biology of Music : Another One Bites the Dust. *Current Biology*, 19(10), R403-R404. <https://doi.org/10.1016/j.cub.2009.04.004>
- FITCH, W. T. (2013). Rhythmic cognition in humans and animals : distinguishing meter and pulse perception. *Frontiers in Systems Neuroscience*, 7(October), 1-16. <https://doi.org/10.3389/fnsys.2013.00068>
- FITCH, W. T., & HAUSER, M. D. (1995). Vocal production in nonhuman primates : Acoustics, physiology, and functional constraints on “honest” advertisement. *American Journal of Primatology*, 37(3), 191-219. <https://doi.org/10.1002/ajp.1350370303>
- FLACK, A., NAGY, M., FIEDLER, W., COUZIN, I. D., & WIKELSKI, M. (2018). From local collective behavior to global migratory patterns in white storks. *Science*, 360(May), 911-914.
- FLETCHER, N. H. (2000). A class of chaotic bird calls? *The Journal of the Acoustical Society of America*, 108(2), 821-826. <https://doi.org/10.1121/1.429615>
- FLOWER, T. P., GRIBBLE, M., & RIDLEY, A. R. (2014). Deception by flexible alarm mimicry in an African bird. *Science*, 344(6183), 513. <https://doi.org/10.1126/science.1249723>
- FOLLIOU, A., HAUPERT, S., DUCRETTET, 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. <https://doi.org/10.1016/j.scitotenv.2022.155883>
- FOX, E., ROBERTS, J. D., & BENNAMOUN, M. (2006). Text-independent Speaker Identification in Birds. *Interspeech*, (January), 2122-2125. <https://doi.org/10.21437/Interspeech.2006-196>
- FOX, E. J., ROBERTS, J. D., & BENNAMOUN, M. (2008). Call-independent individual identification in birds. *Bioacoustics*, 18(1), 51-67. <https://doi.org/10.1080/09524622.2008.9753590>
- FRAISSE, P. (1982). Rhythm and tempo. In *The psychology of music* (p. 149-180).
- FRASER, O. N., & BUGNYAR, T. (2010). Do ravens show consolation? responses to distressed others. *PLoS ONE*, 5(5). <https://doi.org/10.1371/journal.pone.0010605>
- FRASER, O. N., & BUGNYAR, T. (2012). Reciprocity of agonistic support in ravens. *Animal Behaviour*, 83(1), 171-177. <https://doi.org/10.1016/j.anbehav.2011.10.023>
- FREEBERG, T. M., DUNBAR, R. I., & ORD, T. J. (2012). Social complexity as a proximate and ultimate factor in communicative complexity. *Philosophical Transactions of the Royal Society B : Biological Sciences*, 367(1597), 1785-1801. <https://doi.org/10.1098/rstb.2011.0213>
- FREEBERG, T. M., & KRAMS, I. (2015). Does social complexity link vocal complexity and cooperation? *Journal of Ornithology*, 156, 125-132. <https://doi.org/10.1007/s10336-015-1233-2>
- FRISTRUP, K. M., & MENNITT, D. (2012). Biacoustical monitoring in terrestrial environments. *Acoustics Today*, 8(3), 16-24.
- FU, X., LIU, Y., & LIU, Y. (2022). A case study of utilizing YOLOT based quantitative detection algorithm for marine benthos. *Ecological Informatics*, 70, 101603. <https://doi.org/https://doi.org/10.1016/j.ecoinf.2022.101603>
- FUENTES, A. (2007). Social organization : social systems and the complexities in understanding the evolution of primate behavior. In C. CAMPBELL, A. FUENTES, K. MCKINNON & M. PANGER (Éd.), *Primates in perspective* (p. 609-621). Oxford University Press.
- FURRER, R. D., & MANSER, M. B. (2009). The evolution of urgency-based and functionally referential alarm calls in ground-dwelling species. *American Naturalist*, 173(3), 400-410. <https://doi.org/10.1086/596541>

- GARCIA DE LA CHICA, A., HUCK, M., DEPEINE, C., ROTUNDO, M., ADRET, P., & FERNANDEZ-DUQUE, E. (2020). Sexual dimorphism in the loud calls of Azara's owl monkeys (*Aotus azarae*) : evidence of sexual selection? *Primates*, *61*(2), 309-319. <https://doi.org/10.1007/s10329-019-00773-6>
- GARLAND, E. C., & MCGREGOR, P. K. (2020). Cultural Transmission, Evolution, and Revolution in Vocal Displays : Insights From Bird and Whale Song. *Frontiers in Psychology*, *11*(September). <https://doi.org/10.3389/fpsyg.2020.544929>
- GARLAND, E. C., RENDELL, L., LILLEY, M. S., POOLE, M. M., ALLEN, J., & NOAD, M. J. (2017). The devil is in the detail : Quantifying vocal variation in a complex, multi-levelled, and rapidly evolving display. *The Journal of the Acoustical Society of America*, *142*(1), 460-472. <https://doi.org/10.1121/1.4991320>
- GHAZANFAR, A. A., LIAO, D. A., & TAKAHASHI, D. Y. (2019). Volition and learning in primate vocal behaviour. *Animal Behaviour*, *151*, 239-247. <https://doi.org/10.1016/j.anbehav.2019.01.021>
- GHIGLIERI, M. P. (1984). Feeding Ecology and Sociality of Chimpanzees in Kibale Forest, Uganda. In P. S. RODMAN & J. G. H. CANT (Éd.), *Contributions to an Organismal Biology of Prosimians, Monkeys, and Apes* (p. 161-194). Columbia University Press. <https://doi.org/doi:10.7312/rodm90184-008>
- GOFFINET, J., BRUDNER, S., MOONEY, R., & PEARSON, J. (2021). Low-dimensional learned feature spaces quantify individual and group differences in vocal repertoires. *eLife*, *10*, 1-23. <https://doi.org/10.7554/eLife.67855>
- GOODFELLOW, I. J., VINYALS, O., & SAXE, A. M. (2015). Qualitatively characterizing neural network optimization problems. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*.
- GOODWIN, D. (1986). *Crows of the World*. Emdunsbury Press, Suffolk, United Kingdom.
- GOODWIN, D. (1955). Some observations on the reproductive behaviour of rooks. *British Birds*, *48*, 97-107.
- GOSSETTE, R. L., GOSSETTE, M. F., & RIDDELL, W. (1966). Comparisons of successive discrimination reversal performances among closely and remotely related avian species. *Animal Behaviour*, *14*(4), 560-564. [https://doi.org/10.1016/S0003-3472\(66\)80060-X](https://doi.org/10.1016/S0003-3472(66)80060-X)
- GRAMZA, A. F. (1970). Vocal Mimicry in Captive Budgerigars (*Melopsittacus undulatus*). *Zeitschrift für Tierpsychologie*, *27*(8), 971-983. <https://doi.org/10.1111/j.1439-0310.1970.tb01913.x>
- GRAY, P. C., FLEISHMAN, A. B., KLEIN, D. J., MCKOWN, M. W., BÉZY, V. S., LOHMANN, K. J., JOHNSTON, D. W., CAROLINA, N., & GRAY, P. C. (2019). A convolutional neural network for detecting sea turtles in drone imagery. *Methods in Ecology and Evolution*, *10*(August 2018), 345-355. <https://doi.org/10.1111/2041-210X.13132>
- GREENBERG, R., & METTKE-HOFMANN, C. (2001). Ecological Aspects of Neophobia and neophilia in birds. *Current Ornithology, Volume 16, 16*, 119-178. [https://doi.org/10.1007/978-1-4615-1211-0\\_3](https://doi.org/10.1007/978-1-4615-1211-0_3)
- GREENWOOD, P. J. (1980). Mating systems, philopatry and dispersal in birds and mammals. *Animal Behaviour*, *28*(4), 1140-1162. [https://doi.org/10.1016/S0003-3472\(80\)80103-5](https://doi.org/10.1016/S0003-3472(80)80103-5)
- GREGGOR, A. L., JOLLES, J. W., THORNTON, A., & CLAYTON, N. S. (2016). Seasonal changes in neophobia and its consistency in rooks : the effect of novelty type and dominance position. *Animal Behaviour*, *121*, 11-20. <https://doi.org/10.1016/j.anbehav.2016.08.010>
- GRIESSMANN, B., & NAGUIB, M. (2002). Song Sharing in Neighboring and Non-Neighboring Thrush Nighthingales (*Luscinia luscinia*) and its Implications for Communication. *Ethology*, *14*, 377-387. <https://doi.org/10.1046/j.1439-0310.2002.00781.x>
- GRIFFIN, L. R. (1998). The distribution and abundance of the rook *Corvus frugilegus* L. as influenced by habitat suitability and competitive interactions., 175.

- GRIFFIN, L. R., & THOMAS, C. J. (2000). The spatial distribution and size of rook (*Corvus frugilegus*) breeding colonies is affected by both the distribution of foraging habitat and by intercolony competition. *Proceedings of the Royal Society B : Biological Sciences*, 267(1451), 1463-1467. <https://doi.org/10.1098/rspb.2000.1165>
- GRIFFIN, L. R. (1999). Colonization patterns at rook *corvus frugilegus* colonies : Implications for survey strategies. *Bird Study*, 46(2), 170-173. <https://doi.org/10.1080/00063659909461128>
- GRILL, T., & SCHLÜTER, J. (2017). Two convolutional neural networks for bird detection in audio signals. *25th European Signal Processing Conference, EUSIPCO 2017, 2017-Janua*, 1764-1768. <https://doi.org/10.23919/EUSIPCO.2017.8081512>
- GRODZINSKI, U., & CLAYTON, N. S. (2010). Problems faced by food-caching corvids and the evolution of cognitive solutions. *Philosophical Transactions of the Royal Society B : Biological Sciences*, 365(1542), 977-987. <https://doi.org/10.1098/rstb.2009.0210>
- GRUETER, C. C., QI, X., ZINNER, D., BERGMAN, T., LI, M., XIANG, Z., ZHU, P., MIGLIANO, A. B., MILLER, A., KRÜTZEN, M., FISCHER, J., RUBENSTEIN, D. I., VIDYA, T. N., LI, B., CANTOR, M., & SWEDDELL, L. (2020). Multilevel Organisation of Animal Sociality. *Trends in Ecology and Evolution*, 35(9), 834-847. <https://doi.org/10.1016/j.tree.2020.05.003>
- GÜNTÜRKÜN, O., & BUGNYAR, T. (2016). Cognition without Cortex. *Trends in Cognitive Sciences*, 20(4), 291-303. <https://doi.org/10.1016/j.tics.2016.02.001>
- GUSTISON, M. L., le ROUX, A., & BERGMAN, T. J. (2012). Derived vocalizations of geladas (*Theropithecus gelada*) and the evolution of vocal complexity in primates. *Philosophical Transactions of the Royal Society B : Biological Sciences*, 367(1597), 1847-1859. <https://doi.org/10.1098/rstb.2011.0218>
- HA, R. R., BENTZEN, P., MARSH, J., & HA, J. C. (2003). Kinship and association in social foraging northwestern crows (*Corvus caurinus*). *Bird Behavior*, 15(2), 65-75. <http://www.ingentaconnect.com/content/cog/bb/2003/00000015/00000002/bb063>
- HAGE, S. R., GAVRILOV, N., & NIEDER, A. (2013). Cognitive control of distinct vocalizations in rhesus monkeys. *Journal of Cognitive Neuroscience*, 25(10), 1692-1701. [https://doi.org/10.1162/jocn\\_a\\_00428](https://doi.org/10.1162/jocn_a_00428)
- HAGMANN, C. E., & COOK, R. G. (2010). Testing meter, rhythm, and tempo discriminations in pigeons. *Behavioural Processes*, 85(2), 99-110. <https://doi.org/10.1016/j.beproc.2010.06.015>
- HALL, M. L. (2009). Chapter 3 A Review of Vocal Duetting in Birds. Academic Press. [https://doi.org/https://doi.org/10.1016/S0065-3454\(09\)40003-2](https://doi.org/https://doi.org/10.1016/S0065-3454(09)40003-2)
- HAMILTON, W. (1964). The Genetical Evolution of Social Behaviour. II. *Journal of Theoretical Biology*, 7, 17-52.
- HAMMERSCHMIDT, K., & FISCHER, J. (1998). The Vocal Repertoire of Barbary Macaques : A Quantitative Analysis of a Graded Signal System. *Ethology*, 104, 204-216.
- HAMMERSCHMIDT, K., & FISCHER, J. (2008). Constraints in primate vocal production. In D. K. OLLER & U. GRIEBEL (Éd.), *Evolution of communicative flexibility : complexity, creativity, and adaptability in human and animal communication* (p. 92-119). TMIT Press.
- HANSEN, B. T., JOHANNESSEN, L. E., & SLAGSVOLD, T. (2010). Interspecific cross-fostering of great tits (*Parus major*) by blue tits (*Cyanistes caeruleus*) affects inter- and intraspecific communication. *Behaviour*, 147(3), 413-424. <https://doi.org/10.1163/000579509X12578482434926>
- HANSEN, P. (1979). Vocal learning : its role in adapting sound structures to long-distance propagation, and a hypothesis to its evolution. *Animal Behaviour*, 27(4), 1270-1271.

- HARTSHORNE. (1956). The monotony-threshold in singing birds. *The Auk*, 73, 176-192.
- HASEGAWA, A., OKANOYA, K., HASEGAWA, T., & SEKI, Y. (2011). Rhythmic synchronization tapping to an audio-visual metronome in budgerigars. *Scientific Reports*, 1, 1-8. <https://doi.org/10.1038/srep00120>
- HATTORI, M., SOBAGAKI, T., MATSUDA, H., NAKAMURA, S., & TOKUDA, M. (2022). The wintering ecology of the Rook *Corvus frugilegus* in Northern Kyushu, Japan. *Urban Ecosystems*. <https://doi.org/10.1007/s11252-022-01273-0>
- HAUSBERGER, M., RICHARD-YRIS, M. A., HENRY, L., LEPAGE, L., & SCHMIDT, I. (1995). Song Sharing Reflects the Social Organization in a Captive Group of European Starlings (*Sturnus vulgaris*). *Journal of Comparative Psychology*, 109(3), 222-241. <https://doi.org/10.1037/0735-7036.109.3.222>
- HAUSBERGER, M., BIGOT, E., & CLERGEAU, P. (2008). Dialect use in large assemblies : A study in European starling *Sturnus vulgaris* roosts. *Journal of Avian Biology*, 39(6), 672-682. <https://doi.org/10.1111/j.1600-048X.2008.04307.x>
- HAUSER, M. D. (1997). *The evolution of communication*. MIT Press.
- HEINICKE, S., KALAN, A. K., WAGNER, O. J., MUNDRY, R., LUKASHEVICH, H., & KÜHL, H. S. (2015). Assessing the performance of a semi-automated acoustic monitoring system for primates. *Methods in Ecology and Evolution*, 6(7), 753-763. <https://doi.org/10.1111/2041-210X.12384>
- HEINRICH, B. (1988). Winter foraging at carcasses by three sympatric corvids, with emphasis on recruitment by the raven, *Corvus corax*. *Behavioral Ecology and Sociobiology*, 23(3), 141-156. <https://doi.org/10.1007/BF00300349>
- HEINRICH, B. (1999). *Mind of the Raven*. Cliff Street Books.
- HENDERSON, I. G., & HART, P. J. (1991). Age-specific differences in the winter foraging strategies of rooks *Corvus frugilegus*. *Oecologia*, 85(4), 492-497. <https://doi.org/10.1007/BF00323760>
- HILL, M. O. (1973). Diversity and Evenness : A Unifying Notation and Its Consequences. *Ecology*, 54(2), 427-432. <http://www.jstor.org/stable/1934352>
- HILLEMANN, F., BUGNYAR, T., KOTRSCHAL, K., & WASCHER, C. A. (2014). Waiting for better, not for more : Corvids respond to quality in two delay maintenance tasks. *Animal Behaviour*, 90, 1-10. <https://doi.org/10.1016/j.anbehav.2014.01.007>
- HINDE, R. (1976). Interactions, relationships and social structure. *Man*, 72(4), 1242-1252.
- HOFFMANN, L. A., SARAVANAN, V., WOOD, A. N., HE, L., & SOBER, S. J. (2016). Dopaminergic contributions to vocal learning. *Journal of Neuroscience*, 36(7), 2176-2189. <https://doi.org/10.1523/JNEUROSCI.3883-15.2016>
- HOLEKAMP, K. E., SAKAI, S. T., & LUNDRIGAN, B. L. (2007). Social intelligence in the spotted hyena (*Crocuta crocuta*). *Philosophical Transactions of the Royal Society B : Biological Sciences*, 362(1480), 523-538. <https://doi.org/10.1098/rstb.2006.1993>
- HOLZHAIDER, J. C., SIBLEY, M. D., TAYLOR, A. H., SINGH, P. J., GRAY, R. D., & HUNT, G. R. (2011). The social structure of New Caledonian crows. *Animal Behaviour*, 81(1), 83-92. <https://doi.org/10.1016/j.anbehav.2010.09.015>
- HOOPER, S., REISS, D., CARTER, M., & MCCOWAN, B. (2006). Importance of Contextual Saliency on Vocal Imitation by Bottlenose Dolphins Six Flags Marine World , U . S . A . Wildlife Conservation Society , U . S . A . 19(1), 116-128.
- HOPP, S. L., JABLONSKI, P., & BROWN, J. L. (2001). Recognition of group membership by voice in Mexican jays, *Aphelocoma ultramarina*. *Animal Behaviour*, 62(2), 297-303. <https://doi.org/10.1006/anbe.2001.1745>

- HUGHES, M., ANDERSON, R. C., SEARCY, W. A., BOTTENSEK, L. M., & NOWICKI, S. (2007). Song type sharing and territory tenure in eastern song sparrows : implications for the evolution of song repertoires. *Animal Behaviour*, 73(4), 701-710. <https://doi.org/10.1016/j.anbehav.2006.09.013>
- HUMPAL, J., & CYNX, J. A. (1984). Discrimination of Temporal Components of Acoustic Patterns by Birds. *Annals of the New York Academy of Sciences*, 423(1), 600-602. <https://doi.org/10.1111/j.1749-6632.1984.tb23466.x>
- INCE, S. A., & SLATER, P. J. (1985). Versatility and continuity in the songs of thrushes *Turdus* spp. *Ibis*, 127(3), 355-364. <https://doi.org/10.1111/j.1474-919X.1985.tb05077.x>
- IOFFE, S. (2017). Batch Renormalization : Towards reducing minibatch dependence in batch-normalized models. *Advances in Neural Information Processing Systems, 2017-Decem*, 1946-1954.
- IOFFE, S., & SZEGEDY, C. (2015). Batch normalization : Accelerating deep network training by reducing internal covariate shift. *32nd International Conference on Machine Learning, ICML 2015, 1*, 448-456.
- JANIK, V. M. (2014). Cetacean vocal learning and communication. *Current Opinion in Neurobiology*, 28, 60-65. <https://doi.org/10.1016/j.conb.2014.06.010>
- JANIK, V. M., & SLATER, P. J. (2000). The different roles of social learning in vocal communication. *Animal Behaviour*, 60(1), 1-11. <https://doi.org/10.1006/anbe.2000.1410>
- JANSEN, D. A., CANT, M. A., & MANSER, M. B. (2012). Segmental concatenation of individual signatures and context cues in banded mongoose (*Mungos mungo*) close calls. *BMC Biology*, 10(1), 97. <https://doi.org/10.1186/1741-7007-10-97>
- JARVIS, E. D. (2004). Learned birdsong and the neurobiology of human language. *Annals of the New York Academy of Sciences*, 1016, 749-777. <https://doi.org/10.1196/annals.1298.038>
- JELBERT, S. A., HOSKING, R. J., TAYLOR, A. H., & GRAY, R. D. (2018). Mental template matching is a potential cultural transmission mechanism for New Caledonian crow tool manufacturing traditions. *Scientific Reports*, 8(1), 1-8. <https://doi.org/10.1038/s41598-018-27405-1>
- JIANG, T., GUO, X., LIN, A., WU, H., SUN, C., FENG, J., & KANWAL, J. S. (2019). Bats increase vocal amplitude and decrease vocal complexity to mitigate noise interference during social communication. *Animal Cognition*, 22(2), 199-212. <https://doi.org/10.1007/s10071-018-01235-0>
- JIGUET, F. (2020). The Fox and the Crow. A need to update pest control strategies. *Biological Conservation*, 248(July), 108693. <https://doi.org/10.1016/j.biocon.2020.108693>
- JOLY, A., GOËAU, H., KAHL, S., PICEK, L., LORIEUL, T., COLE, E., DENEU, B., SERVAJEAN, M., DURSO, A., BOLON, I., GLOTIN, H., PLANQUÉ, R., de CASTAÑEDA, R. R., VELLINGA, W. P., KLINCK, H., DENTON, T., EGGEL, I., BONNET, P., & MÜLLER, H. (2021). Overview of LifeCLEF 2021 : An Evaluation of Machine-Learning Based Species Identification and Species Distribution Prediction. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12880 LNCS, 371-393. [https://doi.org/10.1007/978-3-030-85251-1\\_24](https://doi.org/10.1007/978-3-030-85251-1_24)
- KABADAYI, C., & OSVATH, M. (2017). Ravens parallel great apes in flexible planning for tool-use and bartending. *Science*, 357(6347), 202-204. <https://doi.org/10.1126/science.aam8138>
- KAHL, S., DENTON, T., KLINCK, H., GLOTIN, H., GOËAU, H., VELLINGA, W. P., PLANQUÉ, R., & JOLY, A. (2021a). Overview of BirdCLEF 2021 : Bird call identification in soundscape recordings. *CEUR Workshop Proceedings*, 2936, 1437-1450.
- KAHL, S., STÖTER, F. R., GOËAU, H., GLOTIN, H., PLANQUÉ, R., VELLINGA, W. P., & JOLY, A. (2019). Overview of BIRDCLEF 2019 : Large-scale bird recognition in soundscapes. *CEUR Workshop Proceedings*, 2380, 9-12.

- KAHL, S., WOOD, C. M., EIBL, M., & KLINCK, H. (2021b). BirdNET : A deep learning solution for avian diversity monitoring. *Ecological Informatics*, 61(January), 101236. <https://doi.org/10.1016/j.ecoinf.2021.101236>
- KAPLAN, G. (1999). Song structure and function of mimicry in the Australian magpie (*Gymnorhina tibicen*) : Compared to lyrebird (*Menura ssp.*) *International Journal of Comparative Psychology*.
- KAPPELER, P. M., & VAN SCHAIK, C. P. (2002). Evolution of primate social systems. *International Journal of Primatology*, 23(4), 707-740. <https://doi.org/10.1023/A:1015520830318>
- KEAGY, J., SAVARD, J. F., & BORGIA, G. (2011). Complex relationship between multiple measures of cognitive ability and male mating success in satin bowerbirds, *Ptilonorhynchus violaceus*. *Animal Behaviour*, 81(5), 1063-1070. <https://doi.org/10.1016/j.anbehav.2011.02.018>
- KEENAN, S., LEMASSON, A., & ZUBERBÜHLER, K. (2013). Graded or discrete? A quantitative analysis of Campbell's monkey alarm calls. *Animal Behaviour*, 85(1), 109-118. <https://doi.org/10.1016/j.anbehav.2012.10.014>
- KEENAN, S., MATHEVON, N., STEVENS, J. M., NICOLÈ, F., ZUBERBÜHLER, K., GUÉRY, J. P., & LEVRÉRO, F. (2020). The reliability of individual vocal signature varies across the bonobo's graded repertoire. *Animal Behaviour*, 169, 9-21. <https://doi.org/10.1016/j.anbehav.2020.08.024>
- KELLEY, J. L., MORRELL, L. J., INSKIP, C., KRAUSE, J., & CROFT, D. P. (2011). Predation risk shapes social networks in fission-fusion populations. *PLoS ONE*, 6(8). <https://doi.org/10.1371/journal.pone.0024280>
- KENWARD, B., SCHLOEGL, C., RUTZ, C., WEIR, A. A., BUGNYAR, T., & KACELNIK, A. (2011). On the evolutionary and ontogenetic origins of tool-oriented behaviour in New Caledonian crows (*Corvus moneduloides*). *Biological Journal of the Linnean Society*, 102(4), 870-877. <https://doi.org/10.1111/j.1095-8312.2011.01613.x>
- KEOGH, E., & RATANAMAHATANA, C. A. (2005). Exact indexing of dynamic time warping. *Knowledge and Information Systems*, 7(3), 358-386. <https://doi.org/10.1007/s10115-004-0154-9>
- KERSHENBAUM, A., BLUMSTEIN, D. T., ROCH, M. A., AKÇAY, Ç., BACKUS, G., BEE, M. A., BOHN, K., CAO, Y., CARTER, G., CÄSAR, C., COEN, M., DERUITER, S. L., DOYLE, L., EDELMAN, S., FERRER-I-CANCHO, R., FREEBERG, T. M., GARLAND, E. C., GUSTISON, M., HARLEY, H. E., ... ZAMORA-GUTIERREZ, V. (2014). Acoustic sequences in non-human animals : A tutorial review and prospectus. *Biological Reviews*, 91(1), 000-000. <https://doi.org/10.1111/brv.12160>
- KERSHENBAUM, A., DÉAUX, É. C., HABIB, B., MITCHELL, B., PALACIOS, V., ROOT-GUTTERIDGE, H., & WALLER, S. (2018). Measuring acoustic complexity in continuously varying signals : how complex is a wolf howl? *Bioacoustics*, 27(3), 215-229. <https://doi.org/10.1080/09524622.2017.1317287>
- KERSHENBAUM, A., SAYIGH, L. S., & JANIK, V. M. (2013). The Encoding of Individual Identity in Dolphin Signature Whistles : How Much Information Is Needed? *PLoS ONE*, 8(10), 1-7. <https://doi.org/10.1371/journal.pone.0077671>
- KIKUSUI, T., NAKANISHI, K., NAKAGAWA, R., NAGASAWA, M., MOGI, K., & OKANOYA, K. (2011). Cross fostering experiments suggest that mice songs are innate. *PLoS ONE*, 6(3), 1-10. <https://doi.org/10.1371/journal.pone.0017721>
- KIPPER, S., MUNDY, R., HULTSCH, H., & TODT, D. (2004). Long-term persistence of song performance rules in nightingales (*Luscinia megarhynchos*) : A longitudinal field study on repertoire size and composition. *Behaviour*, 141(3), 371-390. <https://doi.org/10.1163/156853904322981914>



- KIRSCHER, A. N., EARL, D. A., YAO, Y., ESCOBAR, I. A., VILCHES, E., VALLEJO, E. E., & TAYLOR, C. E. (2009). Using songs to identify individual Mexican antthrush : Comparison of four classification methods. *Bioacoustics*, *19*(1-2), 1-20. <https://doi.org/10.1080/09524622.2009.9753612>
- KNEUTGEN, J. (1969). "Musikalische" Formen im Gesang der Schamadrossel (*Kittacincla macroura* Gm.) und ihre Funktionen. *Journal of Ornithology*, *110*(3), 245-285. <https://doi.org/10.1007/BF01671063>
- KNÖRNSCHILD, M. (2014). Vocal production learning in bats. *Current Opinion in Neurobiology*, *28*, 80-85. <https://doi.org/10.1016/j.conb.2014.06.014>
- KNÖRNSCHILD, M., FERNANDEZ, A. A., & NAGY, M. (2020). Vocal information and the navigation of social decisions in bats : Is social complexity linked to vocal complexity? *Functional Ecology*, *34*(2), 322-331. <https://doi.org/10.1111/1365-2435.13407>
- KODA, H., OYAKAWA, C., KATO, A., & MASATAKA, N. (2007). Experimental evidence for the volitional control of vocal production in an immature gibbon. *Behaviour*, *144*(6), 681-692. <https://doi.org/10.1163/156853907781347817>
- KOMDEUR, J. (1994a). The effect of kinship on helping in the cooperative breeding Seychelles warbler (*Acrocephalus sechellensis*). *Proceedings of the Royal Society B : Biological Sciences*, *256*(1345), 47-52. <https://doi.org/10.1098/rspb.1994.0047>
- KOMDEUR, J. (1994b). Experimental evidence for helping and hindering by previous offspring in the cooperative-breeding Seychelles warbler *Acrocephalus sechellensis*. *Behavioral Ecology and Sociobiology*, *34*(3), 175-186. <https://doi.org/10.1007/BF00167742>
- KONDO, N., IZAWA, E. I., & WATANABE, S. (2010). Perceptual mechanism for vocal individual recognition in jungle crows (*Corvus macrorhynchos*) : Contact call signature and discrimination. *Behaviour*, *147*(8), 1051-1072. <https://doi.org/10.1163/000579510X505427>
- KONDO, N. (2021). A report on a characteristic vocalization in *Corvus macrorhynchos osai* with an indication of vocal learning. *bioRxiv*.
- KONDO, N., IZAWA, E. I., & WATANABE, S. (2012). Crows cross-modally recognize group members but not non-group members. *Proceedings of the Royal Society B : Biological Sciences*, *279*(1735), 1937-1942. <https://doi.org/10.1098/rspb.2011.2419>
- KONDO, N., & WATANABE, S. (2009). Contact calls : Information and social function. *Japanese Psychological Research*, *51*(3), 197-208. <https://doi.org/10.1111/j.1468-5884.2009.00399.x>
- KONG, Q., XU, Y., & PLUMBLEY, M. D. (2017a). Joint detection and classification convolutional neural network on weakly labelled bird audio detection. *25th European Signal Processing Conference, EUSIPCO 2017, 2017-Janua*, 1749-1753. <https://doi.org/10.23919/EUSIPCO.2017.8081509>
- KONG, Q., XU, Y., WANG, W., & PLUMBLEY, M. D. (2017b). A joint detection-classification model for audio tagging of weakly labelled data. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, (Figure 1), 641-645. <https://doi.org/10.1109/ICASSP.2017.7952234>
- KONISHI, M. (1985). Birdsong : From behavior to neuron. <https://doi.org/10.1146/annurev.neuro.8.1.125>
- KONISHI, M. (2004). The role of auditory feedback in birdsong. *Annals New York Academy of Science*, (1016), 463-475. <https://doi.org/10.1196/annals.1298.010>
- KOTZ, S. A., RAVIGNANI, A., & FITCH, W. T. (2018). The Evolution of Rhythm Processing. *Trends in Cognitive Sciences*, *22*(10), 896-910. <https://doi.org/10.1016/j.tics.2018.08.002>
- KRAMS, I., & FREEBERG, T. (2006). Mobbing behaviour : reciprocity- based co-operation in breeding Pied Flycatchers *Ficedula hypoleuca*. *Ibis*, *148*, 50-54.

- KRAMS, I., KRAMA, T., FREEBERG, T. M., KULLBERG, C., & LUCAS, J. R. (2012). Linking social complexity and vocal complexity : A parid perspective. *Philosophical Transactions of the Royal Society B : Biological Sciences*, 367(1597), 1879-1891. <https://doi.org/10.1098/rstb.2011.0222>
- KRAUSE, J., & RUXTON, G. D. (2002). *Living in groups*. Oxford University Press.
- KREBS, J. R., & DAWKINS, R. (1984). Animal singals : mind-reading and manipulation. In J. KREBS & N. DAVIES (Éd.), *Behavioural ecology : an evolutionary approach* (2nd, p. 380-402). Blackwell.
- KREBS, J. R., KACELNIK, A., & TAYLOR, P. (1978). Test of optimal sampling by foraging great tits. *Nature*, 275(5675), 27-31. <https://doi.org/10.1038/275027a0>
- KREMERS, D., LEMASSON, A., ALMUNIA, J., & WANKER, R. (2012). Vocal sharing and individual acoustic distinctiveness within a group of captive orcas (*Orcinus orca*). *Journal of Comparative Psychology*, 126(4), 433-445. <https://doi.org/10.1037/a0028858>
- KRIENGWATANA, B. P., MOTT, R., & ten CATE, C. (2022). Music for animal welfare : A critical review & conceptual framework. *Applied Animal Behaviour Science*, 251(March), 105641. <https://doi.org/10.1016/j.applanim.2022.105641>
- KROODSMA, D. E. (1978). Continuity and versatility in bird song : support for the monotony-threshold hypothesis.
- KROODSMA, D. E. (2004). The diversity and plasticity of birdsong. *Nature's Music : The Science of Birdsong*, 108-131. <https://doi.org/10.1016/B978-012473070-0/50007-4>
- KROODSMA, D. E., & PARKER, L. D. (1977). Vocal virtuosity in the brown thrasher. *The Auk*, 94(4), 783-785. <https://doi.org/10.2307/4085282>
- KROODSMA, D. E., & VERNER, J. (1978). Complex singing behaviors among Cistothorus wrens. *The Auk*, 95(October), 703-716. <https://doi.org/10.1093/acprof:oso/9780199278145.003.0007>
- KUHL, P. K. (2012). Social Mechanisms in Early Language Acquisition : Understanding Integrated Brain Systems Supporting Language. *The Oxford handbook of social neuroscience*, (June 2018), 649-667.
- KUKAČKA, J., GOLKOV, V., & CREMERS, D. (2017). Regularization for Deep Learning : A Taxonomy, 1-23. <http://arxiv.org/abs/1710.10686>
- LACHLAN, R. F. (2007). Luscinia : a bioacoustics analysis computer program.
- LACHLAN, R. F., RATMANN, O., & NOWICKI, S. (2018). Cultural conformity generates extremely stable traditions in bird song. *Nature Communications*, 9(1). <https://doi.org/10.1038/s41467-018-04728-1>
- LACHLAN, R. F., & SLATER, P. J. (1999). The maintenance of vocal learning by gene-culture interaction : The cultural trap hypothesis. *Proceedings of the Royal Society B : Biological Sciences*, 266(1420), 701-706. <https://doi.org/10.1098/rspb.1999.0692>
- LACHLAN, R. . F., VERHAGEN, L., PETERS, S., & ten CATE, C. (2010). Are There Species-Universal Categories in Bird Song Phonology and Syntax ? A Comparative Study of Chaffinches (*Fringilla coelebs*), Zebra Finches (*Taenopygia guttata*), and Swamp Sparrows (*Melospiza georgiana*). *Journal of Comparative Psychology*, 124(1), 92-108. <https://doi.org/10.1037/a0016996>
- LAIOLO, P., PALESTRINI, C., & ROLANDO, A. (2000). A study of Choughs' vocal repertoire : Variability related to individuals, sexes and ages. *Journal fur Ornithologie*, 141(2), 168-179. <https://doi.org/10.1046/j.1439-0361.2000.00074.x>
- LAIOLO, P., & ROLANDO, A. (2003). The evolution of vocalisations in the genus *Corvus* : Effects of phylogeny, morphology and habitat. *Evolutionary Ecology*, 17(2), 111-123. <https://doi.org/10.1023/A:1023003110969>
- LAMARCK, J.-B. (1809). *Philosophie zoologique*.

- LAMBERT, M. L., JACOBS, I., OSVATH, M., & VON BAYERN, A. M. (2019). Birds of a feather? Parrot and corvid cognition compared. *Behaviour*, 156(5-8), 505-594. <https://doi.org/10.1163/1568539X-00003527>
- LAMBERT, M. L., SCHIESTL, M., SCHWING, R., TAYLOR, A. H., GAJDON, G. K., SLOCOMBE, K. E., & SEED, A. M. (2017). Function and flexibility of object exploration in kea and new caledonian crows. *Royal Society Open Science*, 4(9). <https://doi.org/10.1098/rsos.170652>
- LAMBRECHTS, M., & DHONDT, A. A. (1987). Differences in singing performance between male great tits marcel lambrechts & andre a. dhondt. *Ardea*, 75, 43-52.
- LARSEN, O. N., & GOLLER, F. (2002). Direct observation of syringeal muscle function in songbirds and a parrot. *Journal of Experimental Biology*, 205(1), 25-35. <https://doi.org/10.1242/jeb.205.1.25>
- LATTENKAMP, E. Z., VERNES, S. C., & WIEGREBE, L. (2018). Volitional control of social vocalisations and vocal usage learning in bats. *Journal of Experimental Biology*, 221(14), 1-8. <https://doi.org/10.1242/JEB.180729>
- LAUAY, C., GERLACH, N. M., ADKINS-REGAN, E., & DEVOOGD, T. J. (2004). Female zebra finches require early song exposure to prefer high-quality song as adults. *Animal Behaviour*, 68(6), 1249-1255. <https://doi.org/10.1016/j.anbehav.2003.12.025>
- LAURIJS, K. A., BRIEFER, E. F., REIMERT, I., & WEBB, L. E. (2021). Vocalisations in farm animals : A step towards positive welfare assessment. *Applied Animal Behaviour Science*, 236, 105264. <https://doi.org/10.1016/j.applanim.2021.105264>
- LEE, W. Y., im LEE, S., CHOE, J. C., & JABLONSKI, P. G. (2011). Wild birds recognize individual humans : Experiments on magpies, *Pica pica*. *Animal Cognition*, 14(6), 817-825. <https://doi.org/10.1007/s10071-011-0415-4>
- LEFEBVRE, L. (2013). Brains, innovations, tools and cultural transmission in birds, non-human primates and fossil hominins. *Frontiers in Human Neuroscience*, 7(MAY), 1-10. <https://doi.org/10.3389/fnhum.2013.00245>
- LEFEBVRE, L., NICOLAKAKIS, N., & BOIRE, D. (2002). Tools and brains in birds. *Behaviour*, 139(7), 939-973. <https://doi.org/10.1163/156853902320387918>
- LEFEBVRE, L., READER, S. M., & SOL, D. (2004). Brains, innovations and evolution in birds and primates. *Brain, Behavior and Evolution*, 63(4), 233-246. <https://doi.org/10.1159/000076784>
- LEITNER, S., & CATCHPOLE, C. K. (2004). Syllable repertoire and the size of the song control system in captive canaries (*Serinus canaria*). *Journal of Neurobiology*, 60(1), 21-27. <https://doi.org/10.1002/neu.10331>
- LEMASSON, A., GLAS, L., BARBU, S., LACROIX, A., GUILLOUX, M., REMEUF, K., & KODA, H. (2011a). Youngsters do not pay attention to conversational rules : Is this so for nonhuman primates? *Scientific Reports*, 1, 12-15. <https://doi.org/10.1038/srep00022>
- LEMASSON, A., OUATTARA, K., PETIT, E. J., & ZUBERBÜHLER, K. (2011b). Social learning of vocal structure in a nonhuman primate? *BMC Evolutionary Biology*, (362).
- LENTH, R. V. (2022). emmeans : Estimated Marginal Means, aka Least-Squares Means. <https://cran.r-project.org/package=emmeans>
- LEONARD, M. L., & HORN, A. G. (2005). Ambient noise and the design of begging signals. *Proceedings of the Royal Society B : Biological Sciences*, 272(1563), 651-656. <https://doi.org/10.1098/rspb.2004.3021>
- LI, J., HASSANI, A., WALTON, S., & SHI, H. (2021). ConvMLP : Hierarchical Convolutional MLPs for Vision. <http://arxiv.org/abs/2109.04454>

- LIAQAT, S., BOZORG, N., JOSE, N., CONREY, P., TAMASI, A., & JOHNSON, M. T. (2018). Domain Tuning Methods For Bird Audio Detection. <https://github.com/UKYSpeechLab/ukybirddet>
- LIEBAL, K., WALLER, B., BURROWS, A., & SLOCOMBE, K. (2012). *Primate communication : a multimodal approach*. Cambridge University Press.
- LIEBEL, L., & KÖRNER, M. (2018). Auxiliary Tasks in Multi-task Learning, 1-8. <http://arxiv.org/abs/1805.06334>
- LIGON, R. A., & MCGRAW, K. J. (2016). Social costs enforce honesty of a dynamic signal of motivation. *Proceedings of the Royal Society B : Biological Sciences*, 283(1841). <https://doi.org/10.1098/rspb.2016.1873>
- LIN, T. Y., GOYAL, P., GIRSHICK, R., HE, K., & DOLLAR, P. (2020). Focal Loss for Dense Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2), 318-327. <https://doi.org/10.1109/TPAMI.2018.2858826>
- LINHART, P., OSIEJUK, T. S., BUDKA, M., ŠÁLEK, M., ŠPINKA, M., POLICHT, R., SYROVÁ, M., & BLUMSTEIN, D. T. (2019). Measuring individual identity information in animal signals : Overview and performance of available identity metrics. *Methods in Ecology and Evolution*, 10(9), 1558-1570. <https://doi.org/10.1111/2041-210X.13238>
- 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>
- LORETTO, M. C., SCHUSTER, R., ITTY, C., MARCHAND, P., GENERO, F., & BUGNYAR, T. (2017). Fission-fusion dynamics over large distances in raven non-breeders. *Scientific Reports*, 7(1), 1-9. <https://doi.org/10.1038/s41598-017-00404-4>
- LORETTO, M.-C., FRASER, O. N., & BUGNYAR, T. (2012). Ontogeny of Social Relations and Coalition Formation in Common Ravens (*Corvus corax*). *International journal of comparative psychology*, 25(3), 180-194. <http://www.ncbi.nlm.nih.gov/pubmed/25892846%7B%5C%%7D0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4398861>
- LOSTANLEN, V., SALAMON, J., FARNSWORTH, A., KELLING, S., & BELLO, J. P. (2019). Robust sound event detection in bioacoustic sensor networks. *PLoS ONE*, 14(10), 1-31. <https://doi.org/10.1371/journal.pone.0214168>
- LUEF, E. M., TER MAAT, A., & PIKA, S. (2017). Vocal similarity in long-distance and short-distance vocalizations in raven pairs (*Corvus corax*) in captivity. *Behavioural Processes*, 142(May 2016), 1-7. <https://doi.org/10.1016/j.beproc.2017.05.013>
- MACEDONIA, J. M., & EVANS, C. S. (1993). Essay on Contemporary Issues in Ethology : Variation among Mammalian Alarm Call Systems and the Problem of Meaning in Animal Signals. *Ethology*, 93(3), 177-197. <https://doi.org/10.1111/j.1439-0310.1993.tb00988.x>
- MACIEJ, P., NDAO, I., HAMMERSCHMIDT, K., & FISCHER, J. (2013). Vocal communication in a complex multi-level society : constrained acoustic structure and flexible call usage in Guinea babblers. *Frontiers in Zoology*, 10(52), 1-15.
- MACLEAN, E. L., HARE, B., NUNN, C. L., ADDESSI, E., AMICI, F., ANDERSON, R. C., AURELI, F., BAKER, J. M., BANIA, A. E., BARNARD, A. M., BOOGERT, N. J., BRANNON, E. M., BRAY, E. E., BRAY, J., BRENT, L. J., BURKART, J. M., CALL, J., CANTLON, J. F., CHEKE, L. G., ... ZHAO, Y. (2014). The evolution of self-control. *Proceedings of the National Academy of Sciences of the United States of America*, 111(20). <https://doi.org/10.1073/pnas.1323533111>

- MANABE, K., SADR, E. I., & DOOLING, R. J. (1998). Control of vocal intensity in budgerigars ( *Melopsittacus undulatus* ) : Differential reinforcement of vocal intensity and the Lombard effect. *The Journal of the Acoustical Society of America*, 103(2), 1190-1198. <https://doi.org/10.1121/1.421227>
- MANSEY, M. B., JANSEN, D. A., GRAW, B., HOLLÉN, L. I., BOUSQUET, C. A., FURRER, R. D., & le ROUX, A. (2014). Vocal Complexity in Meerkats and Other Mongoose Species. *Advances in the Study of Behavior*, 46, 281-310. <https://doi.org/10.1016/B978-0-12-800286-5.00006-7>
- MARINO, L. (2004). Dolphin cognition. *Current biology*, 14(21), 910-911. <https://doi.org/10.1016/j.cub.2004.10.010>
- MARLER, P. (1955). Characteristics of some animal calls. *Nature*, 176(4470), 6-8.
- MARLER, P. (1961). The logical analysis of animal communication. *Journal of Theoretical Biology*, 1(3), 295-317. [https://doi.org/10.1016/0022-5193\(61\)90032-7](https://doi.org/10.1016/0022-5193(61)90032-7)
- MARLER, P. (1967). Animal communication signals. *Science*, 157(3790), 769-774. <https://doi.org/10.1126/science.157.3790.769>
- MARLER, P. (1997). Three models of song learning : Evidence from behavior. *Journal of Neurobiology*, 33(5), 501-516. [https://doi.org/10.1002/\(SICI\)1097-4695\(19971105\)33:5<501::AID-NEU2>3.0.CO;2-8](https://doi.org/10.1002/(SICI)1097-4695(19971105)33:5<501::AID-NEU2>3.0.CO;2-8)
- MARLER, P. (2004a). Bird calls : a cornucopia for communication. In *Nature's music* (p. 132-177). Elsevier.
- MARLER, P. (2004b). Bird calls : Their potential for behavioral neurobiology. *Annals of the New York Academy of Sciences*, 1016, 31-44. <https://doi.org/10.1196/annals.1298.034>
- MARLER, P., & EVANS, C. (1996). Bird calls : Just emotional displays or something more ? *Ibis*, 138(1), 26-33. <https://doi.org/10.1111/j.1474-919x.1996.tb04310.x>
- MARLER, P., EVANS, C. S., & HAUSER, M. D. (1992). Animal signals : motivational, referential, or both ? In H. PAPOUSEK, U. JÜRGENS & M. PAPOUSEK (Éd.), *Nonverbal vocal communication : comparative and developmental approaches*. Cambridge University Press.
- MARLER, P., & PETERS, S. (1987). A sensitive period for song acquisition in the song sparrow, *Melospiza melodia* : a case of age-limited learning. *Ethology*, 79, 89-100. <https://doi.org/10.1111/j.1439-0310.1987.tb00675.x>
- MARLER, P., & SLABBEKOORN, H. (2004). *Nature's Music : The Science of Birdsong* (P. MARLER & H. SLABBEKOORN, Éd.). Elsevier.
- MARLER, P., & TAMURA, M. (1962). Song "dialects" in three populations of white-crowned sparrows. *The Condor*, 64, 368-377.
- MARLER, P., & TAMURA, M. (1964). Culturally Transmitted Patterns of Vocal Behavior in Sparrows. *Science*, 146(3650), 1483-1486. <https://www.jstor.org/stable/1714851>
- MARSHALL, A. J. (1950). The function of vocal mimicry in birds. *Emu*, 50(1), 5-16. <https://doi.org/10.1071/MU950005>
- MARSHALL, A. J., & COOMBS, C. J. (1957). The Interaction of Environmental, Internal and Behavioural Factors in the Rook, *Corvus F. Frugilegus* Linnaeus. *Proceedings of the Zoological Society of London*, 128(4), 545-588. <https://doi.org/10.1111/j.1096-3642.1957.tb00275.x>
- MARTIN, K., ADAM, O., OBIN, N., & DUFOUR, V. (2022). Rookognise : Acoustic detection and identification of individual rooks in field recordings using multi-task neural networks. *Ecological Informatics*, 72(September). <https://doi.org/10.1016/j.ecoinf.2022.101818>
- MARZLUFF, J., & ANGELL, T. (2005). Cultural Coevolution : How the Human Bond with Crows and Ravens Extends Theory and Raises New Questions. *Journal of Ecological Anthropology*, 9(1), 69-75. <https://doi.org/10.5038/2162-4593.9.1.5>

- MARZLUFF, J. M., WALLS, J., CORNELL, H. N., WITHEY, J. C., & CRAIG, D. P. (2010). Lasting recognition of threatening people by wild American crows. *Animal Behaviour*, 79(3), 699-707. <https://doi.org/10.1016/j.anbehav.2009.12.022>
- MATES, E. A., TARTER, R. R., HA, J. C., CLARK, A. B., & MCGOWAN, K. J. (2015). Acoustic profiling in a complexly social species, the American crow : Caws encode information on caller sex, identity and behavioural context. *Bioacoustics*, 24(1), 63-80. <https://doi.org/10.1080/09524622.2014.933446>
- MATHEVON, N., CASEY, C., REICHMUTH, C., & CHARRIER, I. (2017). Northern Elephant Seals Memorize the Rhythm and Timbre of Their Rivals' Voices. *Current Biology*, 27(15), 2352-2356.e2. <https://doi.org/10.1016/j.cub.2017.06.035>
- MCCOMB, K., & SEMPLE, S. (2005). Coevolution of vocal communication and sociality in primates. *Biology Letters*, 1(4), 381-385. <https://doi.org/10.1098/rsbl.2005.0366>
- MCCORDIC, J. A., ROOT-GUTTERIDGE, H., CUSANO, D. A., DENES, S. L., & PARKS, S. E. (2016). Calls of North Atlantic right whales *Eubalaena glacialis* contain information on individual identity and age class. *Endangered Species Research*, 30(1), 157-169. <https://doi.org/10.3354/esr00735>
- MCCUE, L. M., CIOFFI, W. R., HEITHAUS, M. R., BARRÈ, L., & CONNOR, R. C. (2020). Synchrony, leadership, and association in male Indo-pacific bottlenose dolphins (*Tursiops aduncus*). *Ethology*, 126(7), 741-750. <https://doi.org/10.1111/eth.13025>
- MCDERMOTT, J., & HAUSER, M. D. (2007). Nonhuman primates prefer slow tempos but dislike music overall. *Cognition*, 104(3), 654-668. <https://doi.org/10.1016/j.cognition.2006.07.011>
- MCGREGOR, P. K. (1980). Song Dialects in the Corn Bunting (*Emberiza calandra*). *Zeitschrift für Tierpsychologie*, 54(3), 285-297. <https://doi.org/10.1111/j.1439-0310.1980.tb01246.x>
- MCINNES, L., HEALY, J., & ASTELS, S. (2017). hdbscan : Hierarchical density based clustering. *The Journal of Open Source Software*, 2(11), 205. <https://doi.org/10.21105/joss.00205>
- MCINNES, L., HEALY, J., & MELVILLE, J. (2018). UMAP : Uniform manifold approximation and projection for dimension reduction. *arXiv*.
- MCKILLIGAN, N. G. (1980). The winter exodus of the rook from a Scottish Highland valley. *Bird Study*, 27(2), 93-100. <https://doi.org/10.1080/00063658009476663>
- MERCADO, E., & PERAZIO, C. E. (2022). All units are equal in humpback whale songs, but some are more equal than others. *Animal Cognition*, 25(1), 149-177. <https://doi.org/10.1007/s10071-021-01539-8>
- MERCHANT, H., & HONING, H. (2014). Are non-human primates capable of rhythmic entrainment ? Evidence for the gradual audiomotor evolution hypothesis. *Frontiers in Neuroscience*, 7(JAN), 1-8. <https://doi.org/10.3389/fnins.2013.00274>
- MERKER, B. H., MADISON, G. S., & ECKERDAL, P. (2009). On the role and origin of isochrony in human rhythmic entrainment. *Cortex*, 45(1), 4-17. <https://doi.org/10.1016/j.cortex.2008.06.011>
- METTKE-HOFMANN, C., WINKLER, H., & LEISLER, B. (2002). The significance of ecological factors for exploration and neophobia in parrots. *Ethology*, 108(3), 249-272. <https://doi.org/10.1046/j.1439-0310.2002.00773.x>
- MILLER, C. T., FLUSBERG, S., & HAUSER, M. D. (2003). Interruptibility of long call production in tamarins : Implications for vocal control. *Journal of Experimental Biology*, 206(15), 2629-2639. <https://doi.org/10.1242/jeb.00458>
- MILLER, E. H. (1979). An approach to the analysis of graded vocalizations of birds. *Behavioral and Neural Biology*, 27(1), 25-38. [https://doi.org/10.1016/S0163-1047\(79\)92732-8](https://doi.org/10.1016/S0163-1047(79)92732-8)

- MILLER, R., BOECKLE, M., JELBERT, S. A., FROHNWIESER, A., WASCHER, C. A., & CLAYTON, N. S. (2019). Self-control in crows, parrots and nonhuman primates. *Wiley Interdisciplinary Reviews : Cognitive Science*, 10(6), 1-17. <https://doi.org/10.1002/wcs.1504>
- MILLER, R., LAMBERT, M. L., FROHNWIESER, A., BRECHT, K. F., BUGNYAR, T., CRAMPTON, I., GARCIA-PELEGRIN, E., GOULD, K., GREGGOR, A. L., IZAWA, E. I., KELLY, D. M., LI, Z., LUO, Y., LUONG, L. B., MASSEN, J. J., NIEDER, A., REBER, S. A., SCHIESTL, M., SEGUCHI, A., ... CLAYTON, N. S. (2022). Socio-ecological correlates of neophobia in corvids. *Current Biology*, 32(1), 74-85.e4. <https://doi.org/10.1016/j.cub.2021.10.045>
- MILLIGAN, M. M., & VERNER, J. (1971). Inter-Populational Song Dialect Discrimination in the White-Crowned Sparrow. *The Condor*, 73(2), 208-213. <https://doi.org/10.2307/1365840>
- MINGLE, M. E., EPPLEY, T. E., CAMPBELL, M. W., HALL, K., HORNER, V., & de WALL, F. B. M. (2014). Chimpanzees prefer African and Indian music over silence. *Journal of Experimental Psychology : Animal Learning and Cognition*, 40(4), 502-505. <https://doi.org/10.1037/xan0000032>
- MISRA, D. (2019). Mish : A Self Regularized Non-Monotonic Activation Function. <http://arxiv.org/abs/1908.08681>
- MIYAGAWA, S., OJIMA, S., BERWICK, R. C., & OKANOYA, K. (2014). The integration hypothesis of human language evolution and the nature of contemporary languages. *Frontiers in Psychology*, 5(June), 1-6. <https://doi.org/10.3389/fpsyg.2014.00564>
- MIYAZAKI, M., & WAAS, J. R. (2003). Correlations between body size, defensive behaviour and reproductive success in male Little Blue Penguins *Eudyptula minor* : Implications for female choice. *Ibis*, 145(1), 98-105. <https://doi.org/10.1046/j.1474-919X.2003.00126.x>
- MOHAMMED BENKADA, A., PONTIER, F., & DUFOUR, V. (2020). Conflict management in rooks (*Corvus frugilegus*) : Victims do not display post-conflict affiliation but avoid their former aggressor. *Behavioural Processes*, 179. <https://doi.org/10.1016/j.beproc.2020.104198>
- MONTEIRO, R. d. A., FERREIRA, C. D., & PERBICHE-NEVES, G. (2021). Vocal repertoire and group-specific signature in the smooth-billed ani, *crotophaga ani linnaeus*, 1758 (Cuculiformes, aves). *Papeis Avulsos de Zoologia*, 61, 1-9. <https://doi.org/10.11606/1807-0205/2021.61.59>
- MONTES-MEDINA, A. C., SALINAS-MELGOZA, A., & RENTON, K. (2016). Contextual flexibility in the vocal repertoire of an Amazon parrot. *Frontiers in Zoology*, 13(1), 1-13. <https://doi.org/10.1186/s12983-016-0169-6>
- MORALES PICARD, A., MUNDY, R., AUERSPERG, A. M., BOEVING, E. R., BOUCHERIE, P. H., BUGNYAR, T., DUFOUR, V., EMERY, N. J., FEDERSPIEL, I. G., GAJDON, G. K., GUÉRY, J. P., HEGEDIČ, M., HORN, L., KAVANAGH, E., LAMBERT, M. L., MASSEN, J. J., RODRIGUES, M. A., SCHIESTL, M., SCHWING, R., ... SLOCOMBE, K. E. (2020). Why preen others ? Predictors of allopreening in parrots and corvids and comparisons to grooming in great apes. *Ethology*, 126(2), 207-228. <https://doi.org/10.1111/eth.12999>
- MORFI, V., & STOWELL, D. (2018). Deep learning for audio event detection and tagging on low-resource datasets. *Applied Sciences (Switzerland)*, 8(8). <https://doi.org/10.3390/app8081397>
- MORI, C., LIU, W. C., & WADA, K. (2018). Recurrent development of song idiosyncrasy without auditory inputs in the canary, an open-ended vocal learner. *Scientific Reports*, 8(1), 1-10. <https://doi.org/10.1038/s41598-018-27046-4>
- MORRIS, D. (1955). The Feather Postures of Birds and the Problem of the Origin of Social Signals, 75-112.

- MORTON, E. S. (1977). On the Occurrence and Significance of Motivation-Structural Rules in Some Bird and Mammal Sounds. *The American Naturalist*, 111(981), 855-869. <https://doi.org/10.1086/283219>
- MOULAVI, D., JASKOWIAK, P. A., CAMPELLO, R. J., ZIMEK, A., & SANDER, J. (2014). Density-based clustering validation. *SIAM International Conference on Data Mining 2014, SDM 2014*, 2(1), 839-847. <https://doi.org/10.1137/1.9781611973440.96>
- MUNDINGER, P. C. (1982). Microgeographic and macrogeographic variation in the acquired vocalizations of birds. *Acoustic communication in birds*.
- MUNDY, P. J. (1973). Vocal Mimicry of Their Hosts By Nestlings of the Great Spotted Cuckoo and Striped Crested Cuckoo. *Ibis*, 115(4), 602-604. <https://doi.org/10.1111/j.1474-919X.1973.tb02000.x>
- MUNTEANU, A., PAPUC, I., FEDERSPIEL, I., CLAYTON, N. S., & EMERY, N. (2009). The Social Context of Contact Calls By Rooks (*Corvus Frugilegus*). *Cluj Veterinary Journal*, 15(1), 33-43. <https://doi.org/10.52331/cvj.v15i1.7>
- NARANG, S., DIAMOS, G., ELSER, E., MICIKEVICIUS, P., ALBEN, J., GARCIA, D., GINSBURG, B., HOUSTON, M., KUCHARIEV, O., VENKATESH, G., & WU, H. (2018). Mixed precision training. *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, 1-12.
- NELSON, C. M. V., SHERWIN, W. B., & ORD, T. J. (2022). Why does the complexity of functionally equivalent signals vary across closely related species? *Behavioral Ecology*, (June). <https://doi.org/10.1093/beheco/amac050>
- NELSON, D. A., & MARLER, P. (1994). Selection-based learning in bird song development. *Proceedings of the National Academy of Sciences of the United States of America*, 91(22), 10498-10501. <https://doi.org/10.1073/pnas.91.22.10498>
- NELSON, D. A., & POESEL, A. (2007). Segregation of information in a complex acoustic signal : individual and dialect identity in white-crowned sparrow song. *Animal Behaviour*, 74(4), 1073-1084. <https://doi.org/10.1016/j.anbehav.2007.01.018>
- NIEDER, A., & MOONEY, R. (2020). The neurobiology of innate, volitional and learned vocalizations in mammals and birds. *Philosophical Transactions of the Royal Society B : Biological Sciences*, 375(1789). <https://doi.org/10.1098/rstb.2019.0054>
- NOLASCO, I., & STOWELL, D. (2022). Rank-Based Loss for Learning Hierarchical Representations. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2022-May*, 3623-3627. <https://doi.org/10.1109/ICASSP43922.2022.9746907>
- NOWICKI, S., SEARCY, W. A., & PETERS, S. (2002a). Brain development, song learning and mate choice in birds : A review and experimental test of the "nutritional stress hypothesis". *Journal of Comparative Physiology A : Neuroethology, Sensory, Neural, and Behavioral Physiology*, 188(11-12), 1003-1014. <https://doi.org/10.1007/s00359-002-0361-3>
- NOWICKI, S. (1987). Vocal tract resonances in oscine bird sound production : Evidence from birdsongs in a helium atmosphere. *Nature*, 325(6099), 53-55. <https://doi.org/10.1038/325053a0>
- NOWICKI, S. (1989). Vocal plasticity in captive black-capped chickadees : the acoustic basis and rate of call convergence. *Animal Behaviour*, 37(PART 1), 74-89. [https://doi.org/10.1016/0003-3472\(89\)90008-0](https://doi.org/10.1016/0003-3472(89)90008-0)
- NOWICKI, S., PETERS, S., & PODOS, J. (1998). Song learning, early nutrition and sexual selection in Songbirds. *American Zoologist*, 38(1), 179-190. <https://doi.org/10.1093/icb/38.1.179>
- NOWICKI, S., & SEARCY, W. A. (2005). Song and mate choice in birds : How the development of behavior helps us understand function. *The Auk*, 122(1), 1-14. [https://doi.org/10.1642/0004-8038\(2005\)122\[0001:SAMCIB\]2.0.CO;2](https://doi.org/10.1642/0004-8038(2005)122[0001:SAMCIB]2.0.CO;2)



- NOWICKI, S., & SEARCY, W. A. (2014). The evolution of vocal learning. *Current Opinion in Neurobiology*, 28, 48-53. <https://doi.org/10.1016/j.conb.2014.06.007>
- NOWICKI, S., SEARCY, W. A., HUGHES, M., & PODOS, J. (2001). The evolution of bird song : Male and female response to song innovation in swamp sparrows. *Animal Behaviour*, 62(6), 1189-1195. <https://doi.org/10.1006/anbe.2001.1854>
- NOWICKI, S., SEARCY, W. A., & PETERS, S. (2002b). Quality of song learning affects female response to male bird song. *Proceedings of the Royal Society B : Biological Sciences*, 269(1503), 1949-1954. <https://doi.org/10.1098/rspb.2002.2124>
- NTALAMPIRAS, S., & POTAMITIS, I. (2021). Acoustic detection of unknown bird species and individuals. *CAAI Transactions on Intelligence Technology*, 6(3), 291-300. <https://doi.org/10.1049/cit2.12007>
- OBERWEGER, K., & GOLLER, F. (2001). The metabolic cost of birdsong production. *Journal of Experimental Biology*, 204(19), 3379-3388.
- O'HARA, M., & AUERSPERG, A. M. (2017). Object play in parrots and corvids. *Current Opinion in Behavioral Sciences*, 16, 119-125. <https://doi.org/10.1016/j.cobeha.2017.05.008>
- OIKARINEN, T., SRINIVASAN, K., MEISNER, O., HYMAN, J. B., PARMAR, S., FANUCCI-KISS, A., DESIMONE, R., LANDMAN, R., & FENG, G. (2019). Deep convolutional network for animal sound classification and source attribution using dual audio recordings. *The Journal of the Acoustical Society of America*, 145(2), 654-662. <https://doi.org/10.1121/1.5087827>
- OLKOWICZ, S., KOCOUREK, M., LUÈAN, R. K., PORTEŠ, M., FITCH, W. T., HERCULANO-HOUZEL, S., & NEMEC, P. (2016). Birds have primate-like numbers of neurons in the forebrain. *Proceedings of the National Academy of Sciences of the United States of America*, 113(26), 7255-7260. <https://doi.org/10.1073/pnas.1517131113>
- OLLER, D. K., BUDER, E. H., RAMSDELL, H. L., WARLAUMONT, A. S., CHORNA, L., & BAKEMAN, R. (2013). Functional flexibility of infant vocalization and the emergence of language. *Proceedings of the National Academy of Sciences of the United States of America*, 110(16), 6318-6323. <https://doi.org/10.1073/pnas.1300337110>
- OLLER, D. K., GRIEBEL, U., & WARLAUMONT, A. S. (2016). Vocal Development as a Guide to Modeling the Evolution of Language. *Topics in Cognitive Science*, 8(2), 382-392. <https://doi.org/10.1111/tops.12198>
- OSVATH, M., & SIMA, M. (2014). Sub-adult Ravens Synchronize their Play : A Case of Emotional Contagion ? *Animal Behavior and Cognition*, 2(2), 197. <https://doi.org/10.12966/abc.05.09.2014>
- OUATTARA, K., LEMASSON, A., & ZUBERBÜHLER, K. (2009). Campbell's monkeys concatenate vocalizations into context-specific call sequences. *Proceedings of the National Academy of Sciences of the United States of America*, 106(51), 22026-22031. <https://doi.org/10.1073/pnas.0908118106>
- OWREN, M. J., DIETER, J. A., SEYFARTH, R. M., & CHENEY, D. L. (1992). 'Food' calls produced by adult female rhesus (*Macaca mulatta*) and Japanese (*M. fuscata*) macaques, their normally-raised offspring, and offspring cross-fostered between species. *Behaviour*, 120(3-4), 218-231. <https://doi.org/10.1163/156853992X00615>
- PAGLIARINI, S., TROUVAIN, N., LEBLOIS, A., HINAUT, X., PAGLIARINI, S., TROUVAIN, N., LEBLOIS, A., HINAUT, X., APPLIED, L.-d. G. A. N., PAGLIARINI, S., TROUVAIN, N., LEBLOIS, A., & HINAUT, X. (2021). What does the Canary Say ? Low-Dimensional GAN Applied to Birdsong. <https://hal.inria.fr/hal-03244723v1>
- PANKAJAKSHAN, A., BEAR, H. L., & BENETOS, E. (2019). Polyphonic sound event and sound activity detection : A multi-task approach. *arXiv*, 1-5.

- PANKAJAKSHAN, A., THAKUR, A., THAPAR, D., RAJAN, P., & NIGAM, A. (2018). All-conv net for bird activity detection : Significance of learned pooling. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2018-Septe*, 2122-2126. <https://doi.org/10.21437/Interspeech.2018-1522>
- PAPAGEORGIOU, D., & FARINE, D. R. (2021). Multilevel Societies in Birds. *Trends in Ecology and Evolution*, 36(1), 15-17. <https://doi.org/10.1016/j.tree.2020.10.008>
- PARKS, S. E., CUSANO, D. A., VAN PARIJS, S. M., & NOWACEK, D. P. (2019). Acoustic crypsis in communication by North Atlantic right whale mother–calf pairs on the calving grounds. *Biology Letters*, 15(10). <https://doi.org/10.1098/rsbl.2019.0485>
- PARRISH, J. K., VISCIDO, S. V., & GRÜNBAUM, D. (2002). Self-organized fish schools : An examination of emergent properties. *Biological Bulletin*, 202(3), 296-305. <https://doi.org/10.2307/1543482>
- PATEL, A. D., IVERSEN, J. R., BREGMAN, M. R., & SCHULZ, I. (2009). Experimental Evidence for Synchronization to a Musical Beat in a Nonhuman Animal. *Current Biology*, 19(10), 827-830. <https://doi.org/10.1016/j.cub.2009.03.038>
- PATTERSON, I. J., DUNNET, G. M., & FORDHAM, R. A. (1971). Ecological Studies of the Rook, *Corvus frugilegus* L., in North-East Scotland : Dispersion. *The Journal of Applied Ecology*, 8(3), 815. <https://doi.org/10.2307/2402685>
- PAYNE, R. B., & PAYNE, L. L. (2010). Field observations, experimental design, and the time and place of learning bird songs. In C. T. SNOWDON & M. HAUSBERGER (Éd.), *Social Influences on Vocal Development* (p. 57-84). Cambridge University Press. <https://doi.org/10.1017/cbo9780511758843.005>
- PECKRE, L., KAPPELER, P. M., & FICHTEL, C. (2019). Clarifying and expanding the social complexity hypothesis for communicative complexity. *Behavioral Ecology and Sociobiology*, 73(1). <https://doi.org/10.1007/s00265-018-2605-4>
- PELT, D. M., & SETHIAN, J. A. (2017). A mixed-scale dense convolutional neural network for image analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 115(2), 254-259. <https://doi.org/10.1073/pnas.1715832114>
- PENDERGRAFT, L. J. T., & MARZLUFF, J. M. (2019). Fussing over food : factors affecting the vocalizations American crows utter around food. *Animal Behaviour*, 150, 39-57. <https://doi.org/10.1016/j.anbehav.2019.01.024>
- PEPPERBERG, I. M. (2006). Cognitive and communicative abilities of Grey parrots. *Applied Animal Behaviour Science*, 100(1-2), 77-86. <https://doi.org/10.1016/j.applanim.2006.04.005>
- PEPPERBERG, I. M., & CAREY, S. (2012). Grey Parrot Number Acquisition : the Inference of Cardinal Value from Ordinal Position on the Numeral List. *Cognition*, 125(2), 219-232. <https://doi.org/10.1016/j.cognition.2012.07.003>
- PEREZ, E. C., FERNANDEZ, M. S., GRIFFITH, S. C., VIGNAL, C., & SOULA, H. A. (2015). Impact of visual contact on vocal interaction dynamics of pair-bonded birds. *Animal Behaviour*, 107, 125-137. <https://doi.org/10.1016/j.anbehav.2015.05.019>
- PETKOV, C. I., & JARVIS, E. D. (2012). Birds, primates, and spoken language origins : Behavioral phenotypes and neurobiological substrates. *Frontiers in Evolutionary Neuroscience*, 4(AUG), 1-24. <https://doi.org/10.3389/fnevo.2012.00012>
- PFEIFFER, T., & MEYBURG, B. U. (2015). GPS tracking of Red Kites (*Milvus milvus*) reveals fledgling number is negatively correlated with home range size. *Journal of Ornithology*, 156(4), 963-975. <https://doi.org/10.1007/s10336-015-1230-5>

- PIKA, S., WILKINSON, R., KENDRICK, K. H., & VERNES, S. C. (2018). Taking turns : Bridging the gap between human and animal communication. *Proceedings of the Royal Society B : Biological Sciences*, 285(1880). <https://doi.org/10.1098/rspb.2018.0598>
- PISANSKI, K., RAINE, J., & REBY, D. (2020). Individual differences in human voice pitch are preserved from speech to screams, roars and pain cries. *Royal Society Open Science*, 7(2). <https://doi.org/10.1098/rsos.191642>
- POESEL, A., & KEMPENAERS, B. (2000). When a bird is tired from singing : A study of drift during the dawn chorus. *Etologia*, 8(November), 1-7.
- POLLARD, K. A., & BLUMSTEIN, D. T. (2011). Social group size predicts the evolution of individuality. *Current Biology*, 21(5), 413-417. <https://doi.org/10.1016/j.cub.2011.01.051>
- POMBERGER, T., RISUENO-SEGOVIA, C., LÖSCHNER, J., & HAGE, S. R. (2018). Precise Motor Control Enables Rapid Flexibility in Vocal Behavior of Marmoset Monkeys. *Current Biology*, 28(5), 788-794.e3. <https://doi.org/10.1016/j.cub.2018.01.070>
- PORTER, D., & NEURINGER, A. (1984). Music discriminations by pigeons. *Journal of Experimental Psychology : Animal Behavior Processes*, 10(2), 138-148. <https://doi.org/10.1037/0097-7403.10.2.138>
- PORTER, R. E., CLAPPERTON, B. K., & COLEMAN, J. D. (2008). Distribution, abundance and control of the rook (*corvus frugilegus* L.) in hawke's bay, new zealand, 1969–2006. *Journal of the Royal Society of New Zealand*, 38(1), 25-36. <https://doi.org/10.1080/03014220809510544>
- PORTMANN, A. (1947). Biologische Fragmente zu einer Lehre vom Menschen. *Zeitschrift für Philosophische Forschung*, 1(4).
- POTAMITIS, I., NTALAMPIRAS, S., JAHN, O., & RIEDE, K. (2014). Automatic bird sound detection in long real-field recordings : Applications and tools. *Applied Acoustics*, 80, 1-9. <https://doi.org/10.1016/j.apacoust.2014.01.001>
- POTASH, L. M. (1972). Noise-induced changes in calls of the Japanese quail. *Psychonomic Science*, 26(5), 252-254. <https://doi.org/10.3758/BF03328608>
- PRESLEY, S. J., SCHEINER, S. M., & WILLIG, M. R. (2014). Evaluation of an integrated framework for biodiversity with a new metric for functional dispersion. *PLoS ONE*, 9(8). <https://doi.org/10.1371/journal.pone.0105818>
- PRICE, J. J. (2013). Why is birdsong so repetitive ? Signal detection and the evolution of avian singing modes. 150, 995-1013. <https://doi.org/10.1163/1568539X-00003051>
- PRICE, J. J. (2015). Rethinking our assumptions about the evolution of bird song and other sexually dimorphic signals. *Frontiers in Ecology and Evolution*, 3(APR), 1-6. <https://doi.org/10.3389/fevo.2015.00040>
- PRICE, T., ARNOLD, K., ZUBERBÜHLER, K., & SEMPLE, S. (2009). Pyow but not hack calls of the male putty-nosed monkey (*Cercopithecus nictitans*) convey information about caller identity. *Behaviour*, 146(7), 871-888. <https://doi.org/10.1163/156853908X396610>
- PRIYADARSHANI, N., MARSLAND, S., & CASTRO, I. (2018). Automated birdsong recognition in complex acoustic environments : a review. *Journal of Avian Biology*, 49(5), 1-27. <https://doi.org/10.1111/jav.01447>
- PRŮCHOVÁ, A., JAŠKA, P., & LINHART, P. (2017). Cues to individual identity in songs of songbirds : testing general song characteristics in Chiffchaffs *Phylloscopus collybita*. *Journal of Ornithology*, 158(4), 911-924. <https://doi.org/10.1007/s10336-017-1455-6>

- PTACEK, L., MACHLICA, L., LINHART, P., JASKA, P., & MULLER, L. (2016). Automatic recognition of bird individuals on an open set using as-is recordings. *Bioacoustics*, 25(1), 55-73. <https://doi.org/10.1080/09524622.2015.1089524>
- QUICK, N. J., & JANIK, V. M. (2012). Bottlenose dolphins exchange signature whistles when meeting at sea. *Proceedings of the Royal Society B : Biological Sciences*, 279(1738), 2539-2545. <https://doi.org/10.1098/rspb.2011.2537>
- R CORE TEAM. (2021). R : A Language and Environment for Statistical Computing. <https://www.r-project.org/>
- RABY, C. R., ALEXIS, D. M., DICKINSON, A., & CLAYTON, N. S. (2007). Planning for the future by western scrub-jays. *Nature*, 445(7130), 919-921. <https://doi.org/10.1038/nature05575>
- RAMSEY, G., BASTIAN, M. L., & VAN SCHAİK, C. (2007a). Animal innovation defined and operationalized. *Behavioral and Brain Sciences*, 30(4), 393-437. <https://doi.org/10.1017/S0140525X07002373>
- RAMSEY, G., BASTIAN, M. L., & VAN SCHAİK, C. (2007b). On the concept of animal innovation and the challenge of studying innovation in the wild. *Behavioral and Brain Sciences*, 30(4), 425-432. <https://doi.org/10.1017/S0140525X07002567>
- RANGE, F., & FISCHER, J. (2004). Vocal Repertoire of Sooty Mangabeys (*Cercocebus torquatus atys*) in the Taï National Park. *Ethology*, 110(4), 301-321. <https://doi.org/10.1111/j.1439-0310.2004.00973.x>
- RATCLIFFE, J. M., & TER HOFSTEDE, H. M. (2005). Roosts as information centres : Social learning of food preferences in bats. *Biology Letters*, 1(1), 72-74. <https://doi.org/10.1098/rsbl.2004.0252>
- RAVIGNANI, A., BOWLING, D., & FITCH, W. T. (2014). Chorusing, synchrony and the evolutionary functions of rhythm. *Frontiers in Psychology*, 5(SEP), 1-15. <https://doi.org/10.3389/fpsyg.2014.01118>
- RAVIGNANI, A., & NORTON, P. (2017). Measuring rhythmic complexity : A primer to quantify and compare temporal structure in speech, movement, and animal vocalizations. *Journal of Language Evolution*, 2(1), 4-19. <https://doi.org/10.1093/jole/lzx002>
- RAVIGNANI, A., VERGA, L., & GREENFIELD, M. D. (2019). Interactive rhythms across species : the evolutionary biology of animal chorusing and turn-taking. *Annals of the New York Academy of Sciences*, 1453(1), 12-21. <https://doi.org/10.1111/nyas.14230>
- REBOUT, N., DE MARCO, A., LONE, J. C., SANNA, A., COZZOLINO, R., MICHELETTA, J., STERCK, E. H., LANGERMANS, J. A., LEMASSON, A., & THIERRY, B. (2020). Tolerant and intolerant macaques show different levels of structural complexity in their vocal communication. *Proceedings of the Royal Society B : Biological Sciences*, 287(1928). <https://doi.org/10.1098/rspb.2020.0439>
- REBOUT, N., LONE, J. C., DE MARCO, A., COZZOLINO, R., LEMASSON, A., & THIERRY, B. (2021). Measuring complexity in organisms and organizations. *Royal Society Open Science*, 8(3). <https://doi.org/10.1098/rsos.200895>
- REBY, D., MCCOMB, K., CARGNELUTTI, B., DARWIN, C., FITCH, W. T., & CLUTTON-BROCK, T. (2005). Red deer stags use formants as assessment cues during intrasexual agonistic interactions. *Proceedings of the Royal Society B : Biological Sciences*, 272(1566), 941-947. <https://doi.org/10.1098/rspb.2004.2954>
- REDONDO, T., & DE REYNA, L. A. (1988). Locatability of begging calls in nestling altricial birds. *Animal Behaviour*, 36(3), 653-661. [https://doi.org/10.1016/S0003-3472\(88\)80147-7](https://doi.org/10.1016/S0003-3472(88)80147-7)
- REICHMUTH, C., & CASEY, C. (2014). Vocal learning in seals, sea lions, and walruses. *Current Opinion in Neurobiology*, 28, 66-71. <https://doi.org/10.1016/j.conb.2014.06.011>

- REISS, D., & MCCOWAN, B. (1993). Spontaneous vocal mimicry and production by bottlenose dolphins (*Tursiops truncatus*) : evidence for vocal learning. *Journal of comparative psychology (Washington, D.C. : 1983)*, 107(3), 301-312. <https://doi.org/10.1037/0735-7036.107.3.301>
- RIBOT, R. F., BUCHANAN, K. L., ENDLER, J. A., JOSEPH, L., BENNETT, A. T., & BERG, M. L. (2012). Learned Vocal Variation Is Associated with Abrupt Cryptic Genetic Change in a Parrot Species Complex. *PLoS ONE*, 7(12), 1-9. <https://doi.org/10.1371/journal.pone.0050484>
- RICHARDSON, S. C., PATTERSON, I. J., & DUNNET, G. M. (1979). Fluctuations in Colony Size in the Rook, *Corvus frugilegus*. *Journal of Animal Ecology*, 48(1), 103-110. <http://www.jstor.org/stable/4103>
- RICHNER, H., & HEEB, P. (1995). Communal life : honest signaling and the recruitment center hypothesis. *Behavioral Ecology*, 7(1), 115-117.
- RIEBELL, K., & TODT, D. (1997). Light flash stimulation alters the nightingale's singing style : implications for song control mechanism. *Behaviour*, 134, 789-808.
- RIEGERT, J., & JÚZLOVÁ, Z. (2018). Vocal mimicry in the song of Icterine warblers (*Hippolais icterina*) : possible functions and sources of variability. *Ethology Ecology and Evolution*, 30(5), 430-446. <https://doi.org/10.1080/03949370.2017.1412356>
- RITERS, L. V., POLZIN, B. J., MAKSIMOSKI, A. N., STEVENSON, S. A., & ALGER, S. J. (2022). Birdsong and the Neural Regulation of Positive Emotion. *Frontiers in Psychology*, 13(June), 1-17. <https://doi.org/10.3389/fpsyg.2022.903857>
- RITERS, L. V., SPOOL, J. A., MERULLO, D. P., & HAHN, A. H. (2019). Song practice as a rewarding form of play in songbirds. *Behavioral Processes*, 163, 91\*98. <https://doi.org/10.1016/j.beproc.2017.10.002>.
- RIVERA-CÁCERES, K. D., QUIRÓS-GUERRERO, E., ARAYA-SALAS, M., TEMPLETON, C. N., & SEARCY, W. A. (2018). Early development of vocal interaction rules in a duetting songbird. *Royal Society Open Science*, 5(2). <https://doi.org/10.1098/rsos.171791>
- ROBINSON, C. M., SNYDER, K. T., & CREANZA, N. (2019). Correlated evolution between repertoire size and song plasticity predicts that sexual selection on song promotes open-ended learning. *eLife*, 8, 1-26. <https://doi.org/10.7554/eLife.44454>
- RÖELL, A., & BOSSEMA, I. (1982). A comparison of nest defence by Jackdaws, rooks, magpies and crows. *Behavioral Ecology and Sociobiology*, 11(1), 1-6. <https://doi.org/10.1007/BF00297658>
- ROESKE, T. C., TCHERNICHOVSKI, O., POEPPPEL, D., & JACOBY, N. (2020). Categorical Rhythms Are Shared between Songbirds and Humans. *Current Biology*, 30(18), 3544-3555.e6. <https://doi.org/10.1016/j.cub.2020.06.072>
- ROGERS, A. C., MULDER, R. A., & LANGMORE, N. E. (2006). Duet duels : sex differences in song matching in duetting eastern whipbirds. *Animal Behaviour*, 72(1), 53-61. <https://doi.org/10.1016/j.anbehav.2005.08.019>
- ROSKAFT, E. (1983). Sex-role partitioning and parental care by the rook *Corvus frugilegus*. *Ornis Scandinavica*, 14(3), 180-187. <https://doi.org/10.2307/3676151>
- ROSKAFT, E., & ESPMARK, Y. (1982). Vocal communication by the rook *Corvus frugilegus* during the breeding season. *Ornis Scandinavica*, 13(1), 38-46. <https://doi.org/10.2307/3675971>
- RØSKAFT, E., & ESPMARK, Y. (1984). Sibling recognition in the rook (*Corvus frugilegus*). *Behavioural Processes*, 9(2-3), 223-230. [https://doi.org/10.1016/0376-6357\(84\)90042-1](https://doi.org/10.1016/0376-6357(84)90042-1)
- ROUSE, A. A., COOK, P. F., LARGE, E. W., & REICHMUTH, C. (2016). Beat keeping in a sea lion as coupled oscillation : Implications for comparative understanding of human rhythm. *Frontiers in Neuroscience*, 10(JUN), 1-12. <https://doi.org/10.3389/fnins.2016.00257>

- RUBENSTEIN, D. I. (1978). On Predation, Competition, and the Advantages of Group Living, 205-231. [https://doi.org/10.1007/978-1-4684-2901-5\\_9](https://doi.org/10.1007/978-1-4684-2901-5_9)
- RUDER, S. (2017). An Overview of Multi-Task Learning in Deep Neural Networks. (May). <http://arxiv.org/abs/1706.05098>
- RUXTON, G. D. (2009). Non-visual crypsis : A review of the empirical evidence for camouflage to senses other than vision. *Philosophical Transactions of the Royal Society B : Biological Sciences*, 364(1516), 549-557. <https://doi.org/10.1098/rstb.2008.0228>
- RYAN, M. J., TUTTLE, M. D., & RAND, A. S. (1982). Bat Predation and Sexual Advertisement in a Neotropical Anuran. <https://doi.org/10.1086/283899>
- SAINBURG, T., & GENTNER, T. Q. (2021). Toward a Computational Neuroethology of Vocal Communication : From Bioacoustics to Neurophysiology , Emerging Tools and Future Directions. 15(December), 1-24. <https://doi.org/10.3389/fnbeh.2021.811737>
- SAINBURG, T., THEILMAN, B., THIELK, M., & GENTNER, T. Q. (2019). Parallels in the sequential organization of birdsong and human speech. *Nature Communications*, 10(1), 1-11. <https://doi.org/10.1038/s41467-019-11605-y>
- SAINBURG, T., THIELK, M., & GENTNER, T. Q. (2020). Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires. *PLoS Computational Biology*, 16(10), 1-48. <https://doi.org/10.1371/journal.pcbi.1008228>
- SAITO, T., & REHMSMEIER, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE*, 10(3), 1-21. <https://doi.org/10.1371/journal.pone.0118432>
- SALINAS-MELGOZA, A., & WRIGHT, T. F. (2012). Evidence for vocal learning and limited dispersal as dual mechanisms for dialect maintenance in a parrot. *PloS one*, 7(11). <https://doi.org/10.1371/journal.pone.0048667>
- SALMI, R., HAMMERSCHMIDT, K., & DORAN-SHEEHY, D. M. (2014). Individual distinctiveness in call types of wild western female gorillas. *PLoS ONE*, 9(7). <https://doi.org/10.1371/journal.pone.0101940>
- SANDOVAL, L., EPPERLY, K. L., KLUCKA, J., & MENNILL, D. J. (2017). The biogeographic and evolutionary history of an endemic clade of Middle American sparrows : Melozone and Aimophila (Aves : Passerellidae). *Molecular Phylogenetics and Evolution*, 110, 50-59. <https://doi.org/10.1016/j.ympev.2017.03.008>
- SAUVÉ, C. C., BEAUPLET, G., HAMMILL, M. O., & CHARRIER, I. (2015). Mother-pup vocal recognition in harbour seals : influence of maternal behaviour, pup voice and habitat sound properties. *Animal Behaviour*, 105, 109-120. <https://doi.org/10.1016/j.anbehav.2015.04.011>
- SCARL, J. C., & BRADBURY, J. W. (2009). Rapid vocal convergence in an Australian cockatoo, the galah *Eolophus roseicapillus*. *Animal Behaviour*, 77(5), 1019-1026. <https://doi.org/10.1016/j.anbehav.2008.11.024>
- SCHACHNER, A. (2010). Auditory-motor entrainment in vocal mimicking species : Additional ontogenetic and phylogenetic factors. *Communicative and Integrative Biology*, 3(3), 290-293. <https://doi.org/10.4161/cib.3.3.11708>
- SCHACHNER, A., BRADY, T. F., PEPPERBERG, I. M., & HAUSER, M. D. (2009). Spontaneous Motor Entrainment to Music in Multiple Vocal Mimicking Species. *Current Biology*, 19(10), 831-836. <https://doi.org/10.1016/j.cub.2009.03.061>

- SCHEID, C., SCHMIDT, J., & NOË, R. (2008). Distinct patterns of food offering and co-feeding in rooks. *Animal Behaviour*, 76(5), 1701-1707. <https://doi.org/10.1016/j.anbehav.2008.07.023>
- SCHLÜTER, J. (2018). Bird identification from timestamped, geotagged audio recordings. *CEUR Workshop Proceedings*, 2125(1).
- SCHNEIDER, J. N., & MERCADO, E. (2019). Characterizing the rhythm and tempo of sound production by singing whales. *Bioacoustics*, 28(3), 239-256. <https://doi.org/10.1080/09524622.2018.1428827>
- SCHOECH, S. J., MUMME, R. L., & MOORE, M. C. (1991). Reproductive Endocrinology and Mechanisms of Breeding Inhibition in Cooperatively Breeding Florida Scrub Jays (*Aphelocoma c. coerulescens*). *The Condor*, 93(2), 354-364. <https://doi.org/10.2307/1368951>
- SCHRADIN, C., KÖNIG, B., & PILLAY, N. (2010). Reproductive competition favours solitary living while ecological constraints impose group-living in African striped mice. *Journal of Animal Ecology*, 79(3), 515-521. <https://doi.org/10.1111/j.1365-2656.2009.01651.x>
- SCHRAFT, H. A., MEDINA, O. J., MCCLURE, J., PEREIRA, D. A., & LOGUE, D. M. (2017). Within-day improvement in a behavioural display : wild birds 'warm up'. *Animal Behaviour*, 124, 167-174. <https://doi.org/10.1016/j.anbehav.2016.12.026>
- SEARCY, W. A. (1992). Song repertoire and mate choice in birds. *Integrative and Comparative Biology*, 32(1), 71-80. <https://doi.org/10.1093/icb/32.1.71>
- SEARCY, W. A., & NOWICKI, S. (2019). Birdsong learning, avian cognition and the evolution of language. *Animal Behaviour*, 151, 217-227. <https://doi.org/10.1016/j.anbehav.2019.01.015>
- SEED, A. M., CLAYTON, N. S., & EMERY, N. J. (2007). Postconflict Third-Party Affiliation in Rooks, *Corvus frugilegus*. *Current Biology*, 17(2), 152-158. <https://doi.org/10.1016/j.cub.2006.11.025>
- SEKI, Y., & TOMYTA, K. (2019). Effects of metronomic sounds on a self-paced tapping task in budgerigars and humans. *Current Zoology*, 65(1), 121-128. <https://doi.org/10.1093/cz/zoy075>
- SELEZNEVA, E., DEIKE, S., KNYAZEVA, S., SCHEICH, H., BRECHMANN, A., & BROSCHE, M. (2013). Rhythm sensitivity in macaque monkeys. *Frontiers in Systems Neuroscience*, 7(SEP), 1-9. <https://doi.org/10.3389/fnsys.2013.00049>
- SELVARAJU, R. R., COGSWELL, M., DAS, A., VEDANTAM, R., PARIKH, D., & BATRA, D. (2016). Grad-cam : Why did you say that ? visual explanations from deep networks via gradient-based localization. *Revista do Hospital das Clínicas*, 17, 331-336. <http://arxiv.org/abs/1610.02391>
- SEVILLA, A., & GLOTIN, H. (2017). Audio bird classification with inception-v4 extended with time and time-frequency attention mechanisms. *CEUR Workshop Proceedings*, 1866.
- SEWALL, K. B. (2009). Limited adult vocal learning maintains call dialects but permits pair-distinctive calls in red crossbills. *Animal Behaviour*, 77(5), 1303-1311. <https://doi.org/10.1016/j.anbehav.2009.01.033>
- SEYFARTH, R. M., & CHENEY, D. L. (1984). Grooming, alliances and reciprocal altruism in vervet monkeys. *Nature*, 308(5959), 541-543. <https://doi.org/10.1038/308541a0>
- SEYFARTH, R. M., & CHENEY, D. L. (2017). The origin of meaning in animal signals. *Animal Behaviour*, 124, 339-346. <https://doi.org/10.1016/j.anbehav.2016.05.020>
- SEYFARTH, R. M., CHENEY, D. L., & MARLER, P. (1980). Monkey responses to three different alarm calls : Evidence of predator classification and semantic communication. *Science*, 210(4471), 801-803. <https://doi.org/10.1126/science.7433999>
- SHACKLETON, S. A., & RATCLIFFE, L. (1994). Matched Counter-singing Signals Escalation of Aggression in Black-capped Chickadees (*Parus atricapillus*). *Ethology*, 97(4), 310-316. <https://doi.org/10.1111/j.1439-0310.1994.tb01049.x>

- SHANNON, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3), 379-423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- SHARP, S. P., MCGOWAN, A., WOOD, M. J., & HATCHWELL, B. J. (2005). Learned kin recognition cues in a social bird. *Nature*, 434(7037), 1127-1130. <https://doi.org/10.1038/nature03522>
- SHE, J., ZHAN, W., HONG, S., MIN, C., DONG, T., HUANG, H., & HE, Z. (2022). A method for automatic real-time detection and counting of fruit fly pests in orchards by trap bottles via convolutional neural network with attention mechanism added. *Ecological Informatics*, 70, 101690. <https://doi.org/https://doi.org/10.1016/j.ecoinf.2022.101690>
- SHONFIELD, J., & BAYNE, E. M. (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>
- SHULTZ, S., & DUNBAR, R. (2010). Encephalization is not a universal macroevolutionary phenomenon in mammals but is associated with sociality. *Proceedings of the National Academy of Sciences of the United States of America*, 107(50), 21582-21586. <https://doi.org/10.1073/pnas.1005246107>
- SILK, M. J., CROFT, D. P., TREGENZA, T., & BEARHOP, S. (2014). The importance of fission-fusion social group dynamics in birds. *Ibis*, 156, 701-715.
- SIMONYAN, K., VEDALDI, A., & ZISSERMAN, A. (2014). Deep inside convolutional networks : Visualising image classification models and saliency maps. *2nd International Conference on Learning Representations, ICLR 2014 - Workshop Track Proceedings*, 1-8.
- SLABBEKOORN, H. (2004). Singing in the wild : the ecology of birdsong. In *Nature's music* (p. 178-205). Elsevier.
- SLABBEKOORN, H., & PEET, M. (2003). Birds sing at higher pitch in urban noise. *Nature*, 424(July), 267.
- SLABBEKOORN, H., & SMITH, T. B. (2002). Bird song, ecology and speciation. *Philosophical Transactions of the Royal Society B : Biological Sciences*, 357(1420), 493-503. <https://doi.org/10.1098/rstb.2001.1056>
- SLAGSVOLD, T., HANSEN, B. T., JOHANNESSEN, L. E., & LIFJELD, J. T. (2002). Mate choice and imprinting in birds studied by cross-fostering in the wild. *Proceedings of the Royal Society B : Biological Sciences*, 269(1499), 1449-1455. <https://doi.org/10.1098/rspb.2002.2045>
- SLATER, P. J. (1983). The study of communication. *Animal Behaviour*, 2, 9-42.
- SLATER, P. J., & INCE, S. A. (1982). Song Development in Chffinches : What Is Learnt and When ? *Ibis*, 124(1), 21-26. <https://doi.org/10.1111/j.1474-919X.1982.tb03737.x>
- SLATER, P. J., & LACHLAN, R. (2003). Is innovation in bird song adaptive ? In *Animal Innovation*. <https://doi.org/10.1093/acprof:oso/9780198526223.003.0005>
- SMITH, J. E., VAN HORN, R. C., POWNING, K. S., COLE, A. R., GRAHAM, K. E., MEMENIS, S. K., & HOLEKAMP, K. E. (2010). Evolutionary forces favoring intragroup coalitions among spotted hyenas and other animals. *Behavioral Ecology*, 21(2), 284-303. <https://doi.org/10.1093/beheco/arp181>
- SMITH, J. W. (1965). Message, meaning, and context in ethology. *The American Naturalist*, (4), 18-19.
- SMITH, L. N. (2018). A Disciplined Approach To Neural Network Hyper-Parameters : Part 1 – Learning Rate, Batch Size, Momentum, and Weight Decay. *arXiv*, 1-21.
- SMITH-VIDAURRE, G., ARAYA-SALAS, M., & WRIGHT, T. F. (2020). Individual signatures outweigh social group identity in contact calls of a communally nesting parrot. *Behavioral Ecology*, 31(2), 448-458. <https://doi.org/10.1093/BEHECO/ARZ202>



- SMITH-VIDAURRE, G., PEREZ-MARRUFO, V., & WRIGHT, T. F. (2021). Individual vocal signatures show reduced complexity following invasion. *Animal Behaviour*, *179*, 15-39. <https://doi.org/10.1016/j.anbehav.2021.06.020>
- SNOWDON, C. T., & ELOWSON, A. M. (1999). Pygmy marmosets modify call structure when paired. *Ethology*, *105*(10), 893-908. <https://doi.org/10.1046/j.1439-0310.1999.00483.x>
- SOBER, S. J., & BRAINARD, M. S. (2009). Adult birdsong is actively maintained by error correction. *Nature Neuroscience*, *12*(7), 927-931. <https://doi.org/10.1038/nn.2336>
- SOL, D., DUNCAN, R. P., BLACKBURN, T. M., CASSEY, P., & LEFEBVRE, L. (2005). Big brains, enhanced cognition, and response of birds to novel environments. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(15), 5460-5465. <https://doi.org/10.1073/pnas.0408145102>
- SOL, D., TIMMERMANS, S., & LEFEBVRE, L. (2002). Behavioural flexibility and invasion success in birds. *Animal Behaviour*, *63*(3), 495-502. <https://doi.org/10.1006/anbe.2001.1953>
- SOMERVUO, P. (2019). Time–frequency warping of spectrograms applied to bird sound analyses. *Bioacoustics*, *28*(3), 257-268. <https://doi.org/10.1080/09524622.2018.1431958>
- SOSSINKA, R., & BÖHNER, J. (1980). Song Types in the Zebra Finch *Poephila guttata castanotis*. *Zeitschrift für Tierpsychologie*, *53*(2), 123-132. <https://doi.org/10.1111/j.1439-0310.1980.tb01044.x>
- SPILLMANN, B., van SCHAIK, C. P., SETIA, T. M., & SADJADI, S. O. (2017). Who shall I say is calling? Validation of a caller recognition procedure in Bornean flanged male orangutan (*Pongo pygmaeus wurmbii*) long calls. *Bioacoustics*, *26*(2), 109-120. <https://doi.org/10.1080/09524622.2016.1216802>
- SRIVASTAVA, N., HINTON, G., KRIZHEVSKY, A., SUTSKEVER, I., & SALAKHUTDINOV, R. (2014). Dropout : A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, *15*, 1929-1958.
- STEGMANN, U. (2013). *Animal communication theory : information and influence*. Cambridge University Press.
- STEPHAN, C., & ZUBERBÜHLER, K. (2014). Predation affects alarm call usage in female Diana monkeys (*Cercopithecus diana diana*), 321-331. <https://doi.org/10.1007/s00265-013-1647-x>
- STERCK, E. H., WATTS, D. P., & VAN SCHAIK, C. P. (1997). The evolution of female social relationships in nonhuman primates. *Behavioral Ecology and Sociobiology*, *41*(5), 291-309. <https://doi.org/10.1007/s002650050390>
- STEVENSON, S. A., PIEPENBURG, A., SPOOL, J. A., ANGYAL, C. S., HAHN, A. H., ZHAO, C., & RITERS, L. V. (2020). Endogenous opioids facilitate intrinsically-rewarded birdsong. *Scientific Reports*, *10*(1), 1-12. <https://doi.org/10.1038/s41598-020-67684-1>
- STOEGER, A. S., & MANGER, P. (2014). Vocal learning in elephants : Neural bases and adaptive context. *Current Opinion in Neurobiology*, *28*, 101-107. <https://doi.org/10.1016/j.conb.2014.07.001>
- STOW, M. K., VERNOUILLET, A., & KELLY, D. M. (2018). Neophobia does not account for motoric self-regulation performance as measured during the detour-reaching cylinder task. *Animal Cognition*, *21*(4), 565-574. <https://doi.org/10.1007/s10071-018-1189-8>
- STOWELL, D. (2022). Computational bioacoustics with deep learning : a review and roadmap. *PeerJ*, *10*, e13152. <https://doi.org/10.7717/peerj.13152>
- STOWELL, D., MORFI, V., & GILL, L. F. (2016a). Individual identity in songbirds : Signal representations and metric learning for locating the information in complex corvid calls. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 08-12-Sept*, 2607-2611. <https://doi.org/10.21437/Interspeech.2016-465>

- STOWELL, D., PETRUSKOVÁ, T., ŠÁLEK, M., & LINHART, P. (2019a). Automatic acoustic identification of individuals in multiple species : Improving identification across recording conditions. *Journal of the Royal Society Interface*, 16(153). <https://doi.org/10.1098/rsif.2018.0940>
- STOWELL, D., WOOD, M. D., PAMUŁA, H., STYLIANOU, Y., & GLOTIN, H. (2019b). Automatic acoustic detection of birds through deep learning : The first Bird Audio Detection challenge. *Methods in Ecology and Evolution*, 10(3), 368-380. <https://doi.org/10.1111/2041-210X.13103>
- STOWELL, D., WOOD, M., STYLIANOU, Y., & GLOTIN, H. (2016b). Bird detection in audio : A survey and a challenge. *IEEE International Workshop on Machine Learning for Signal Processing, MLSP, 2016-Novem*(September). <https://doi.org/10.1109/MLSP.2016.7738875>
- SUEUR, J., KRAUSE, B., & FARINA, A. (2021). Acoustic biodiversity. *Current Biology*, 31(19), R1172-R1173. <https://doi.org/10.1016/j.cub.2021.08.063>
- SUTHERS, R. A., & ZOLLINGER, S. A. (2004). Producing song : The vocal apparatus. *Annals New York Academy of Science*, 1016, 109-129. <https://doi.org/10.1196/annals.1298.041>
- SUZUKI, T. N. (2014). Communication about predator type by a bird using discrete, graded and combinatorial variation in alarm calls. *Animal Behaviour*, 87(100), 59-65. <https://doi.org/10.1016/j.anbehav.2013.10.009>
- SUZUKI, T. N. (2016). Semantic communication in birds : evidence from field research over the past two decades. *Ecological Research*, 31(3), 307-319. <https://doi.org/10.1007/s11284-016-1339-x>
- SUZUKI, T. N. (2021). Animal linguistics : Exploring referentiality and compositionality in bird calls. *Ecological Research*, 36(2), 221-231. <https://doi.org/10.1111/1440-1703.12200>
- SUZUKI, T. N., WHEATCROFT, D., & GRIESSER, M. (2019). The syntax – semantics interface in animal vocal communication. *Philosophical Transactions of the Royal Society B : Biological Sciences*, 375. <https://doi.org/10.1098/rstb.2018.0405>
- SWIFT, K. N., & MARZLUFF, J. M. (2015). Wild American crows gather around their dead to learn about danger. *Animal Behaviour*, 109, 187-197. <https://doi.org/10.1016/j.anbehav.2015.08.021>
- SWINGLAND, I. R. (1977). The social and spatial organization of winter communal roosting in Rooks (*Corvus fmgilegus*). *Journal of Zoology*, 182(4), 509-528. <https://doi.org/10.1111/j.1469-7998.1977.tb04167.x>
- SWINGLAND, I. R. (1973). The social organisation of roosting in rooks and jackdaws. *B.Sc. (Lond)*.
- SZEGEDY, C., VANHOUCKE, V., IOFFE, S., SHLENS, J., & WOJNA, Z. (2016). Rethinking the Inception Architecture for Computer Vision. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016-Decem*, 2818-2826. <https://doi.org/10.1109/CVPR.2016.308>
- SZIPL, G., BOECKLE, M., WASCHER, C. A., SPREAFICO, M., & BUGNYAR, T. (2015). With whom to dine ? Ravens' responses to food-associated calls depend on individual characteristics of the caller. *Animal Behaviour*, 99, 33-42. <https://doi.org/10.1016/j.anbehav.2014.10.015>
- TAKEYA, R., KAMEDA, M., PATEL, A. D., & TANAKA, M. (2017). Predictive and tempo-flexible synchronization to a visual metronome in monkeys. *Scientific Reports*, 7(1), 1-12. <https://doi.org/10.1038/s41598-017-06417-3>
- TAKIMOTO, H., SATO, Y., NAGANO, A. J., SHIMIZU, K. K., & KANAGAWA, A. (2021). Using a two-stage convolutional neural network to rapidly identify tiny herbivorous beetles in the field. *Ecological Informatics*, 66, 101466. <https://doi.org/https://doi.org/10.1016/j.ecoinf.2021.101466>
- TANIMOTO, A. M., HART, P. J., PACK, A. A., & SWITZER, R. (2017). Vocal repertoire and signal characteristics of 'Alalā, the Hawaiian Crow (*Corvus hawaiiensis*). *Wilson Journal of Ornithology*, 129(1), 25-35. <https://doi.org/10.1676/1559-4491-129.1.25>

- TAYLOR, A. M., & REBY, D. (2010). The contribution of source-filter theory to mammal vocal communication research. *Journal of Zoology*, 280(3), 221-236. <https://doi.org/10.1111/j.1469-7998.2009.00661.x>
- TAYLOR, A. H. (2014). Corvid cognition. *Wiley Interdisciplinary Reviews : Cognitive Science*, 5(3), 361-372. <https://doi.org/10.1002/wcs.1286>
- TAYLOR, A. M., CHARLTON, B. D., & REBY, D. (2016). *Vocal Production by Terrestrial Mammals : Source, Filter, and Function*. [https://doi.org/10.1007/978-3-319-27721-9\\_8](https://doi.org/10.1007/978-3-319-27721-9_8)
- TEBBICH, S., SEED, A. M., EMERY, N. J., & CLAYTON, N. S. (2007). Non-tool-using rooks, *Corvus frugilegus*, solve the trap-tube problem. *Animal Cognition*, 10(2), 225-231. <https://doi.org/10.1007/s10071-006-0061-4>
- TEIXEIRA, D., LINKE, S., HILL, R., MARON, M., & van RENSBURG, B. J. (2022). Fledge or fail : Nest monitoring of endangered black-cockatoos using bioacoustics and open-source call recognition. *Ecological Informatics*, 69, 101656. <https://doi.org/https://doi.org/10.1016/j.ecoinf.2022.101656>
- TEMPLETON, C. N., GREENE, E., & DAVIS, K. (2005). Behavior : Allometry of alarm calls : Black-capped chickadees encode information about predator size. *Science*, 308(5730), 1934-1937. <https://doi.org/10.1126/science.1108841>
- TEMPLETON, C. N., LALAND, K. N., & BOOGERT, N. J. (2014). Does song complexity correlate with problem-solving performance inflocks of zebra finches? *Animal Behaviour*, 92, 63-71. <https://doi.org/10.1016/j.anbehav.2014.03.019>
- TEN CATE, C. (2021). Re-evaluating vocal production learning in non-oscine birds. *Philosophical Transactions of the Royal Society B : Biological Sciences*, 376(1836). <https://doi.org/10.1098/rstb.2020.0249>
- TEN CATE, C., & BALLINTIJN, M. R. (1996). Dove Coos and Flashed Lights : Interruptibility of "Song" in a Nonsongbird. *Journal of Comparative Psychology*, 110(3), 267-275. <https://doi.org/10.1037/0735-7036.110.3.267>
- ten CATE, C., & SPIERINGS, M. (2019). Rules, rhythm and grouping : auditory pattern perception by birds. *Animal Behaviour*, 151, 249-257. <https://doi.org/10.1016/j.anbehav.2018.11.010>
- ten CATE, C., SPIERINGS, M., HUBERT, J., & HONING, H. (2016). Can birds perceive rhythmic patterns? A review and experiments on a songbird and a parrot species. *Frontiers in Psychology*, 7(MAY). <https://doi.org/10.3389/fpsyg.2016.00730>
- TERRY, A. M., PEAKE, T. M., & MCGREGOR, P. K. (2005). The role of vocal individuality in conservation. *Frontiers in Zoology*, 2, 1-16. <https://doi.org/10.1186/1742-9994-2-10>
- THEUERKAUF, J., ROUYS, S., MÉRIOT, J. M., GULA, R., & KUEHN, R. (2009). Cooperative breeding, mate guarding, and nest sharing in two parrot species of New Caledonia. *Journal of Ornithology*, 150(4), 791-797. <https://doi.org/10.1007/s10336-009-0400-8>
- THIERRY, B. (2007). Unity in diversity : Lessons from macaque societies. *Evolutionary Anthropology*, 16(6), 224-238. <https://doi.org/10.1002/evan.20147>
- THOMPSON, N. S. (1982). A comparison of cawing in the European carrion crow (<i>Corvus corone</i>) and the American common crow (<i>Corvus Brachyrhynchos</i>). *Behaviour*, 80(1/2), 106-117. <https://doi.org/10.1017/CBO9781107415324.004>
- THOMPSON, N. S., LEDOUX, K., & MOODY, K. (1994). A system for describing bird song units. *Bioacoustics*, 5, 267-279.
- THOMSEN, H. M., BALSBY, T. J., & DABELSTEEN, T. (2019). The imitation dilemma : Can parrots maintain their vocal individuality when imitating conspecifics? *Behaviour*, 156(5-8), 787-814. <https://doi.org/10.1163/1568539X-00003548>

- THORPE, W. H. (1958). The learning of song patterns by birds, with especial reference to the song of the chaffinch *Fringilla coelebs*. *Ibis*, *100*(4), 535-570. <https://doi.org/10.1111/j.1474-919X.1958.tb07960.x>
- THORPE, W. H. (1963). Antiphonal singing in birds as evidence for avian auditory reaction time. *Nature*, *197*(4869), 774-776. <https://doi.org/10.1038/197774a0>
- TOMASELLO, M. (s. d.). *Origins of Human Communication*. MIT Press. [https://doi.org/10.9793/elsj.27.2\\_567](https://doi.org/10.9793/elsj.27.2_567)
- TOWNSEND, S. W., ENGESESSER, S., STOLL, S., ZUBERBÜHLER, K., & BICKEL, B. (2018). Compositionality in animals and humans. *PLoS Biology*, *16*(8), 1-7. <https://doi.org/10.1371/journal.pbio.2006425>
- TYACK, P. L. (2000). Dolphins Whistle a Signature Tune. *Science*, *289*(5483), 1310-1311. <https://doi.org/10.1126/science.289.5483.1310>
- TYACK, P. L. (2020). A taxonomy for vocal learning. *Philosophical Transactions of the Royal Society B : Biological Sciences*, *375*(1789). <https://doi.org/10.1098/rstb.2018.0406>
- VAN ROSSUM, G., & DRAKE, F. L. (2009). *Python 3 Reference Manual*. CreateSpace. <https://doi.org/10.5555/1593511>
- VANDER WALL, S. B. (1982). An experimental analysis of cache recovery in Clark's nutcracker. *Animal Behaviour*, *30*(1), 84-94. [https://doi.org/10.1016/S0003-3472\(82\)80240-6](https://doi.org/10.1016/S0003-3472(82)80240-6)
- van HORIK, J. O., & EMERY, N. J. (2018). Serial reversal learning and cognitive flexibility in two species of Neotropical parrots (*Diopsittaca nobilis* and *Pionites melanocephala*). *Behavioural Processes*, *157*, 664-672. <https://doi.org/10.1016/j.beproc.2018.04.002>
- van KLINK, R., AUGUST, T., BAS, Y., BODESHEIM, P., BONN, A., FOSSØY, F., HØYE, T. T., JONGEJANS, E., MENZ, M. H., MIRALDO, A., ROSLIN, T., ROY, H. E., RUCZYŃSKI, I., SCHIGEL, D., SCHÄFFLER, L., SHEARD, J. K., SVENNINGSSEN, C., TSCHAN, G. F., WÄLDCHEN, J., ... BOWLER, D. E. (2022). Emerging technologies revolutionise insect ecology and monitoring. *Trends in Ecology and Evolution*, *20*(20), 1-14. <https://doi.org/10.1016/j.tree.2022.06.001>
- VANNONI, E., & MCELLIGOTT, A. G. (2007). Individual acoustic variation in fallow deer (*Dama dama*) common and harsh groans : A source-filter theory perspective. *Ethology*, *113*(3), 223-234. <https://doi.org/10.1111/j.1439-0310.2006.01323.x>
- VEIT, L., TIAN, L. Y., HERNANDEZ, C. J., & BRAINARD, M. S. (2021). Songbirds can learn flexible contextual control over syllable sequencing. *eLife*, *10*, 1-19. <https://doi.org/10.7554/eLife.61610>
- VENNA, J., & KASKI, S. (2005). Local multidimensional scaling with controlled tradeoff between trustworthiness and continuity. *WSOM 2005 - 5th Workshop on Self-Organizing Maps*, 695-702.
- VERGARA, V., & MIKUS, M. A. (2019). Contact call diversity in natural beluga entrapments in an Arctic estuary : Preliminary evidence of vocal signatures in wild belugas. *Marine Mammal Science*, *35*(2), 434-465. <https://doi.org/10.1111/mms.12538>
- VERNER, J. (1975). Interspecific Aggression between Yellow-Headed Blackbirds and Long-Billed Marsh Wrens. *The Condor*, *77*(3), 328-331. <https://doi.org/10.2307/1366229>
- VERNES, S. C., JANIK, V. M., FITCH, W. T., & SLATER, P. J. (2021). Vocal learning in animals and humans. *Philosophical Transactions of the Royal Society B : Biological Sciences*, *376*(1836). <https://doi.org/10.1098/rstb.2020.0234>
- VERNOUILLET, A., ANDERSON, J., CLARY, D., & KELLY, D. M. (2016). Inhibition in Clark's nutcrackers (*Nucifraga columbiana*) : results of a detour-reaching test. *Animal Cognition*, *19*(3), 661-665. <https://doi.org/10.1007/s10071-016-0952-y>
- VERPOOTEN, J. (2021). *Complex vocal learning and three-dimensional mating environments* (T. 36). Springer Netherlands. <https://doi.org/10.1007/s10539-021-09786-2>

- VERSTEEGH, M., KUHN, J., SYNNAEVE, G., RAVAUX, L., CHEMLA, E., CÄSAR, C., FULLER, J., MURPHY, D., SCHEL, A., & DUNBAR, E. (2016). Classification and automatic transcription of primate calls. *The Journal of the Acoustical Society of America*, *140*(1), EL26-EL30. <https://doi.org/10.1121/1.4954887>
- VIGNAL, C., MATHEVON, N., & MOTTIN, S. (2008). Mate recognition by female zebra finch : Analysis of individuality in male call and first investigations on female decoding process. *Behavioural Processes*, *77*, 191-198. <https://doi.org/10.1016/j.beproc.2007.09.003>
- VOLODIN, I. A., VOLODINA, E. V., KLENOVA, A. V., & MATROSOVA, V. A. (2015). Gender identification using acoustic analysis in birds without external sexual dimorphism. *Avian Research*, *6*(1). <https://doi.org/10.1186/s40657-015-0033-y>
- WADEWITZ, P., HAMMERSCHMIDT, K., BATTAGLIA, D., WITT, A., WOLF, F., & FISCHER, J. (2015). Characterizing vocal repertoires - Hard vs. Soft classification approaches. *PLoS ONE*, *10*(4), 1-16. <https://doi.org/10.1371/journal.pone.0125785>
- WALLØE, S., THOMSEN, H., BALSBY, T. J., & DABELSTEEN, T. (2015). Differences in short-term vocal learning in parrots, a comparative study. *Behaviour*, *152*(11), 1433-1461. <https://doi.org/10.1163/1568539X-00003286>
- 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), 1-12. <https://doi.org/10.1038/s41467-022-28881-w>
- WANG, R., SUN, Y., ZHANG, X., ZENG, S., XIE, W., YU, Y., ZHANG, X., & ZUO, M. (2009). Song control nuclei in male and female large-billed crows (*Corvus macrorhynchos*). *Zoological Science*, *26*(11), 771-777. <https://doi.org/10.2108/zsj.26.771>
- WANG, Y., HUANG, H., RUDIN, C., & SHAPOSHNIK, Y. (2021). Understanding how dimension reduction tools work : An empirical approach to deciphering T-SNE, UMAP, TriMap, and PaCMAP for data visualization. *Journal of Machine Learning Research*, *22*, 1-73.
- WANG, Y., GETREUER, P., HUGHES, T., LYON, R. F., & SAUROUS, R. A. (2017). Trainable frontend for robust and far-field keyword spotting. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 5670-5674. <https://doi.org/10.1109/ICASSP.2017.7953242>
- WARD, P., & ZAHAVI, A. (1973). The Importance of Certain Assemblages of Birds As “Information-Centres” for Food-Finding. *Ibis*, *115*(4), 517-534. <https://doi.org/10.1111/j.1474-919X.1973.tb01990.x>
- WARD, S., SPEAKMAN, J. R., & SLATER, P. J. (2003). The energy cost of song in the canary, *Serinus canaria*. *Animal Behaviour*, *66*(5), 893-902. <https://doi.org/10.1006/anbe.2003.2250>
- WASCHER, C. A., DUFOUR, V., & BUGNYAR, T. (2012a). Carrion crows cannot overcome impulsive choice in a quantitative exchange task. *Frontiers in Psychology*, *3*(APR), 1-6. <https://doi.org/10.3389/fpsyg.2012.00118>
- WASCHER, C. A., SZIPL, G., BOECKLE, M., & WILKINSON, A. (2012b). You sound familiar : Carrion crows can differentiate between the calls of known and unknown heterospecifics. *Animal Cognition*, *15*(5), 1015-1019. <https://doi.org/10.1007/s10071-012-0508-8>
- WASER, P. M., & BROWN, C. H. (1986). Habitat acoustics and primate communication. *American Journal of Primatology*, *10*(2), 135-154. <https://doi.org/10.1002/ajp.1350100205>
- WEARY, D. M., & KREBS, J. R. (1992). Great tits classify songs by individual voice characteristics. *Animal Behaviour*, *43*(2), 283-287. [https://doi.org/10.1016/S0003-3472\(05\)80223-4](https://doi.org/10.1016/S0003-3472(05)80223-4)

- WEARY, D. M., KREBS, J. R., EDDYSHAW, R., MCGREGOR, P. K., & HORN, A. (1988). Decline in song output by great tits : exhaustion or motivation. *Animal Behaviour*, 36(3), 1242-1244.
- WEBBER, T., & STEFANI, R. A. (2010). Evidence for Vocal Learning by a Scrub Jay. *The Auk*, 107(1), 202-204. <https://www.jstor.org/stable/4087823>
- WEINSTEIN, B. G. (2018). Scene-specific convolutional neural networks for video-based biodiversity detection. *Methods in Ecology and Evolution*, 9(6), 1435-1441. <https://doi.org/10.1111/2041-210X.13011>
- WEINSTEIN, B. G. (2019). A computer vision for animal ecology. *Journal of Animal Ecology*, (87), 533-545. <https://doi.org/10.1111/1365-2656.12780>
- WELLS, K. D. (1977). The social behaviour of anuran amphibians. *Animal Behaviour*, 25(PART 3), 666-693. [https://doi.org/10.1016/0003-3472\(77\)90118-X](https://doi.org/10.1016/0003-3472(77)90118-X)
- WEST, M. J., & KING, A. P. (1990). Mozart's Starling.
- WEST-EBERHARD, M. J. (1979). Sexual selection, social competition, and evolution. *Proceedings of the American Philosophical Society*, 123(4), 222-234.
- WEY, T., BLUMSTEIN, D. T., SHEN, W., & JORDÁN, F. (2008). Social network analysis of animal behaviour : a promising tool for the study of sociality. *Animal Behaviour*, 75(2), 333-344. <https://doi.org/10.1016/j.anbehav.2007.06.020>
- WHEATCROFT, D. (2015). Repetition rate of calls used in multiple contexts communicates presence of predators to nestlings and adult birds. *Animal Behaviour*, 103, 35-44. <https://doi.org/10.1016/j.anbehav.2015.02.009>
- WHEATCROFT, D. J., & PRICE, T. D. (2008). Reciprocal cooperation in avian mobbing : playing nice pays. *Trends in Ecology and Evolution*, 23(8), 414-416. <https://doi.org/10.1016/j.tree.2008.05.003>
- WIENS, J. A. (1982). Song pattern variation in the sage sparrow (*Amphispiza belli*) : dialects or epiphenomena? *The Auk*, 99(April), 208-229. <https://doi.org/10.2173/tbna.326.p>
- WILBRECHT, L., & NOTTEBOHM, F. (2003). Vocal learning in birds and humans. *Mental Retardation and Developmental Disabilities Research Reviews*, 9(3), 135-148. <https://doi.org/10.1002/mrdd.10073>
- WILDEN, I., TEMBROCK, G., HERZEL, H., & PETERS, G. (1998). Subharmonics, biphonation, and deterministic chaos in mammal vocalization. *Bioacoustics*, 9(3), 171-196. <https://doi.org/10.1080/09524622.1998.9753394>
- WILEY, R. H. (2012). Communication as a transfer of information : measurement, mechanism and meaning. In U. STEGMANN (Éd.), *Animal communication theory : information and influence* (p. 113-129). Cambridge University Press.
- WILLIAMS, H., & LACHLAN, R. F. (2022). Evidence for cumulative cultural evolution in bird song. *Philosophical Transactions of the Royal Society B : Biological Sciences*, 377(1843). <https://doi.org/10.1098/rstb.2020.0322>
- WILLIS, C. K., & BRIGHAM, R. M. (2004). Roost switching, roost sharing and social cohesion : Forest-dwelling big brown bats, *Eptesicus fuscus*, conform to the fission-fusion model. *Animal Behaviour*, 68(3), 495-505. <https://doi.org/10.1016/j.anbehav.2003.08.028>
- WILSON, D., & MOEHLIS, J. (2014). An energy-optimal approach for entrainment of uncertain circadian oscillators. *Biophysical Journal*, 107(7), 1744-1755. <https://doi.org/10.1016/j.bpj.2014.08.013>
- WILSON, M., & COOK, P. F. (2016). Rhythmic entrainment : Why humans want to, fireflies can't help it, pet birds try, and sea lions have to be bribed. *Psychonomic Bulletin and Review*, 23(6), 1647-1659. <https://doi.org/10.3758/s13423-016-1013-x>

- WINTER, P., HANDLEY, P., PLOOG, D., & SCHOOT, D. (1973). Ontogeny of squirrel monkey calls under normal conditions and under acoustic isolation. *Behaviour*, 47(3-4), 230-239. <https://doi.org/https://www.jstor.org/stable/4533549>
- WINTERS, J. (2019). Escaping optimization traps : the role of cultural adaptation and cultural exaptation in facilitating open-ended cumulative dynamics. *Palgrave Communications*, 5(1), 1-13. <https://doi.org/10.1057/s41599-019-0361-3>
- WITHERBY, H. F. (1943). *Handbook of British birds, Volume 1 : crows to firecrest* (H.F and G.).
- WOODS, R. D., KINGS, M., MCIVOR, G. E., & THORNTON, A. (2018). Caller characteristics influence recruitment to collective anti-predator events in jackdaws. *Scientific Reports*, 8(1), 1-8. <https://doi.org/10.1038/s41598-018-25793-y>
- WRIGHT, L., & DEMEURE, N. (2021). Ranger21 : a synergistic deep learning optimizer. <http://arxiv.org/abs/2106.13731>
- WRIGHT, T. F., & WILKINSON, G. S. (2001). Population genetic structure and vocal dialects in an amazon parrot. *Proceedings of the Royal Society B : Biological Sciences*, 268(1467), 609-616. <https://doi.org/10.1098/rspb.2000.1403>
- WRIGHT, T. F., & DAHLIN, C. R. (2017). Vocal dialects in parrots : Patterns and processes of cultural evolution. *Emu*, 118(1), 50-66. <https://doi.org/10.1080/01584197.2017.1379356>
- YIN, S., & MCCOWAN, B. (2004). Barking in domestic dogs : Context specificity and individual identification. *Animal Behaviour*, 68(2), 343-355. <https://doi.org/10.1016/j.anbehav.2003.07.016>
- YORZINSKI, J. L., & VEHCAMP, S. L. (2009). The effect of predator type and danger level on the mob calls of the American Crow. *Condor*, 111(1), 159-168. <https://doi.org/10.1525/cond.2009.080057>
- YORZINSKI, J. L., VEHCAMP, S. L., MCGOWAN, K. J., & CLARK, A. B. (2006). The Inflected Alarm Caw of the American Crow : Differences in Acoustic Structure Among Individuals and Sexes. *The Condor*, 108(3), 518-529. <https://doi.org/10.1093/condor/108.3.518>
- YOUNG, B. K., MINDLIN, G. B., ARNEODO, E., & GOLLER, F. (2017). Adult zebra finches rehearse highly variable song patterns during sleep. *PeerJ*, 2017(11), 1-18. <https://doi.org/10.7717/peerj.4052>
- ZAHAVI, A. (1975). Mate selection-A selection for a handicap. *Journal of Theoretical Biology*, 53(1), 205-214. [https://doi.org/10.1016/0022-5193\(75\)90111-3](https://doi.org/10.1016/0022-5193(75)90111-3)
- ZARCO, W., MERCHANT, H., PRADO, L., & MENDEZ, J. C. (2009). Subsecond timing in primates : Comparison of interval production between human subjects and rhesus monkeys. *Journal of Neurophysiology*, 102(6), 3191-3202. <https://doi.org/10.1152/jn.00066.2009>
- ZDENEK, C. N., HEINSOHN, R., & LANGMORE, N. E. (2018). Vocal individuality, but not stability, in wild palm cockatoos (*Probosciger aterrimus*). *Bioacoustics*, 27(1), 27-42. <https://doi.org/10.1080/09524622.2016.1272004>
- ZOLLINGER, S. A., & BRUMM, H. (2011). The Lombard effect. *Current Biology*, 21(16), R614-R615. <https://doi.org/10.1016/j.cub.2011.06.003>
- ZOLLINGER, S. A., RIEDE, T., & SLITHERS, R. A. (2008). Two-voice complexity from a single side of the syrinx in northern mockingbird *Mimus polyglottos* vocalizations. *Journal of Experimental Biology*, 211(12), 1978-1991. <https://doi.org/10.1242/jeb.014092>
- ZSEBŐK, S., SCHMERA, D., LACZI, M., NAGY, G., VASKUTI, É., TÖRÖK, J., & ZSOLT GARAMSZEGI, L. (2021). A practical approach to measuring the acoustic diversity by community ecology methods. *Methods in Ecology and Evolution*, 12(5), 874-884. <https://doi.org/10.1111/2041-210X.13558>

- ZUBERBÜHLER, K., NOË, R., & SEYFARTH, R. M. (1997). Diana monkey long-distance calls : Messages for conspecifics and predators. *Animal Behaviour*, *53*(3), 589-604. <https://doi.org/10.1006/anbe.1996.0334>
- ZÜRCHER, Y., WILLEMS, E. P., & BURKART, J. M. (2021). Trade-offs between vocal accommodation and individual recognisability in common marmoset vocalizations. *Scientific Reports*, *11*(1), 1-10. <https://doi.org/10.1038/s41598-021-95101-8>



## In Memoriam

En janvier 2023, après ma soutenance de thèse, la volière de Strasbourg a été attaquée par des prédateurs, et les 14 corbeaux présents ont été tués ou se sont échappés. Nous ne les avons jamais revus.

En leur mémoire, je laisse cette petite trace d'eux et des souvenirs que j'ai pu me constituer pendant mes années de travail auprès d'eux. J'ai essayé de trouver des photos d'autant des 19 corbeaux (les 15 adultes et les 4 bébés, qui n'ont jamais pu intégrer l'étude) que possible, mais certaines femelles ne se sont jamais suffisamment approché pour. Merci beaucoup à Émeline et Agnès, sans qui je n'aurais même pas ces photos.

Au revoir, mes cocos. J'aurais aimé pouvoir vous souhaiter une bonne vie en partant, et peut-être à une prochaine fois. Vous me manquez énormément...





## Killian MARTIN

Complexité vocale et contrôle cognitif chez le corbeau freux (*Corvus frugilegus*)

Pourquoi certains oiseaux produisent-ils un chant complexe ? L'hypothèse de la complexité socio-communicative prédit une corrélation positive entre complexité de l'organisation sociale et complexité du système de communication, et pourrait fournir une explication. Parmi les oiseaux, les corvidés sont parmi les mieux connus pour leur complexité sociale, ainsi que leurs fortes compétences cognitives. Cependant, leurs capacités vocales restent globalement mal connues, hormis leurs capacités d'imitation. J'ai développé des outils d'évaluation de la complexité vocale, d'un système d'extraction et d'attribution de vocalisations à l'individu émetteur, à un protocole de mesure de la complexité vocale directement à partir des spectrogrammes des vocalisations. Ces outils ont été développés et testés par l'étude de la complexité vocale chez le corbeau freux. Cette espèce présente une forte complexité vocale, basée notamment sur des idiosyncrasies marquées dans les vocalisations des mâles. Les femelles sont au contraire très similaires, ce qui m'a permis d'étudier la signature vocale dans une vocalisation de quémandage au nid dans l'optique de développer un outil de suivi non invasif des populations. J'ai également mis au jour une capacité d'entraînement vocal spontané des corbeaux freux à un stimulus rythmique, indicatrice d'un contrôle sur la vitesse de production du chant. Ces résultats soutiennent l'hypothèse de la complexité socio-communicative, mais posent également la question du code de communication entre différents individus chez cette espèce. Des signaux hautement individualisés pourraient nécessiter un apprentissage vocal complexe pour être compris par d'autres individus.

Mots clés : apprentissage automatique, corvidé, diversité, gradation, idiosyncrasies, différences inter-individuelles

Why do some birds produce a complex song? The social complexity hypothesis for communicative complexity predicts a positive correlation between complexity in species social organisation and complexity in species communication system, and may provide an explanation. Corvids are amongst the birds better known for their social complexity and high cognitive skills. However, their vocal abilities remain globally poorly known, apart from their mimicking abilities. I developed tools for assessing vocal complexity, from a system for extracting and attributing vocalisations to the individual emitter, to a protocol for measuring vocal complexity directly from the spectrograms. These tools were developed and tested by studying vocal complexity in rooks. This species exhibits a high degree of vocal complexity, based in particular on pronounced idiosyncrasies in the vocalisations of males. Females, on the other hand, are very similar, which allowed me to study the vocal signature in a nest begging vocalisation with the purpose of developing a non-invasive tool that might assist population monitoring. I also uncovered a capacity for spontaneous vocal entrainment to a rhythmic stimulus in this species, suggesting control over the rate of song production. These results support the social complexity hypothesis for communicative complexity, but also raise the question of the communication code between different individuals in this species. Highly individualised signals may require complex vocal learning to be understood by other individuals.

Keywords : machine learning, corvid, diversity, gradation, idiosyncrasies, inter-individual differences