



HAL
open science

PSIBIOM Technical Report LIS CNRS UTLN - 2024-11

Philémon Prevot, Lisa Ferre, Tristan Villepreux, Stéphane Chavin, Hervé
Glotin

► **To cite this version:**

Philémon Prevot, Lisa Ferre, Tristan Villepreux, Stéphane Chavin, Hervé Glotin. PSIBIOM Technical Report LIS CNRS UTLN - 2024-11. Université de Toulon. 2024. hal-04918712

HAL Id: hal-04918712

<https://hal.science/hal-04918712v1>

Submitted on 29 Jan 2025

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.

Rapport d'avancement et Livrables



Auteur.es :

Philémon PRÉVOT, Lisa FERRE, Tristan VILLEPREUX, Stéphane CHAVIN, Hervé GLOTIN

*Université de Toulon, Aix Marseille Univ, CNRS, LIS, DYNI, Marseille, France
Centre Int. d'IA en Acoustique Naturelle, Toulon, France, <https://cian.univ-tln.fr>*

SOMMAIRE.....	1
Vision planning + Replanifié	2
LOT 1 Module vision.....	2
<i>Livrable 1.1 Développement de l'Intelligence Artificielle offline vision</i>	2
<i>Livrable 1.2 Développement de l'Intelligence Artificielle embarquée vision</i>	2
LOT 2. Module écoute	2
<i>Livrable 2.1 Calcul des indicateurs éco-acoustiques.....</i>	3
<i>Livrable 2.2 Développement de l'Intelligence Artificielle offline acoustique</i>	3
<i>Livrable 2.3 Développement de l'Intelligence Artificielle embarquée acoustique.....</i>	3
LOT 3. Base de données d'entraînement des IA	3
LOT 4. Conception des produits vision et écoute	3
<i>Livrable 4.2 Architecture et design général du système autonome d'écoute.....</i>	4
<i>Livrable 4.3 Conception du système autonome Vision</i>	3
<i>Livrable 4.4 Conception du système autonome d'écoute.....</i>	3
<i>Livrable 4.6 Développement et validation des prototypes du système autonome Écoute</i>	4
Travail réalisé et principaux résultats	4
I. Analyse vision IA online + IA offline	4
II. Analyse acoustique.....	5
<i>A. Tâche classification taxons Orthoptère, Anoures, Oiseaux, Hémiptères.....</i>	6
<i>B. Tâche de classification de familles de Chiroptères.....</i>	7
<i>C. Tâche de classification IA de 10 espèces d'oiseaux d'intérêt.....</i>	8
III. Implémentation des analyses acoustiques sur l'ensemble QHB	11
Conclusion et perspectives	12

Vision Planning + Replanifié

LOT 1. Module vision

Livrable 1.1 Développement de l'Intelligence Artificielle offline du module vision

Les catégories détectées par l'IA sont Cerfs, Chevreuils, Sangliers, et Autres.

Un modèle de CNN basé sur l'architecture SSDLite320 a d'abord été développé avant de se tourner vers l'architecture YOLOv5 une fois la carte de SIConsult reçue. En effet, grâce aux composants permettant l'utilisation d'un modèle plus lourd, l'architecture YOLO plus fiable a pu être mise en oeuvre ce qui n'était pas envisageable dès le départ. Deux modèles ont ainsi été développés : YOLOv5 320x320 et YOLOv5 224x224

Les résultats obtenus grâce à ces deux modèles n'atteignent pas les taux de 5% de faux positifs et 5% de faux négatifs attendus. En effet, selon les classes le taux moyen de Faux Positifs (mFPR) se situe entre 31% et 38% quand le taux moyen de Faux Négatif (mFNR) se situe entre 12% et 20%. Ces résultats couplés à des analyses du dataset mettent en évidence la grande hétérogénéité du dataset en ce qui concerne les instances de chaque classe. Les prochains développements chercheront à équilibrer le dataset ainsi qu'à trouver des classes plus représentatives des données. Par exemple en séparant une classe Humains de la classe Autres, et en supprimant la classe Chamois dont le nombre d'instances trop bas diminue les performances des modèles.

Livrable 1.2 Développement de l'intelligence Artificielle embarquée du module vision

Le portage des modèles précédemment décrits sur la carte Pollux a été effectué, ainsi qu'une étude comparative des résultats de ces modèles par rapport aux résultats off-line. Sur les échantillons utilisés pour l'étude les détections démontrent l'efficacité du modèle, cependant deux problèmes de post-traitement des résultats de classification subsistent :

- La création de Bounding Boxes approximatives
- Le plafonnement du score de classification à une valeur arbitraire

Ces problèmes ne sont toutefois pas bloquants pour le déploiement d'une première série de prototypes car ils n'affectent pas la performance du modèle, mais sont dus à la manière dont les sorties de classification sont interprétées.

LOT 2. Module écoute

Livrable 2.1 Calcul des indicateurs éco-acoustiques

Deux indices éco-acoustiques ont été sélectionnés pour leur pertinence en contexte agricole métropolitain, l'ADI (Acoustic Diversity Index) et l'ACI (Acoustic Complexity Index) qui renseignent en priorité sur la biodiversité aviaire qui est la plus chantante.

- L'ADI tel que décrit par Villanueva-Rivera et al., 2011, reflète la diversité des sons d'un environnement. Il est donc déjà largement utilisé comme proxy de la diversité des espèces chantantes présentes dans le milieu. Son utilisation est donc à faire en prenant en compte l'ensemble des bruits ambiants et en choisissant les moments avec un minimum de bruit anthropophonique ou géophonique afin de cibler la biophonie visée. Une alternative pourrait être de le remplacer par une version modifiée décrite par Zhi-yong Xu et al. 2023, dépendant des bandes de fréquence de l'enregistrement afin d'être plus résistant aux bruits n'appartenant pas à la biophonie.
- L'ACI renseigne sur la richesse de sons produits par une communauté spécifique, donc peut être utilisée comme proxy de l'état de santé d'une communauté précise. Il est résistant aux bruits continus tels que les avions mais peut-être trompé par les bruits ponctuels comme le vent, il convient donc d'analyser la météo et de filtrer les moments venteux ou orageux afin de limiter ces erreurs. Ces indices ont pu être testés par la laboratoire LIS dans de précédents projets et sont intégrés dans les ensembles QHB + stéréo embarqués et calculés à chaque temps d'analyse.

Pieretti, Nadia & Farina, Almo & Morri, D.. (2010). A new methodology to infer the singing activity of an avian community: The Acoustic Complexity Index (ACI). Ecological Indicators. 11. 10.1016/j.ecolind.2010.11.005.

Livrable 2.2 Développement de l'Intelligence Artificielle offline acoustique

Les grands groupes taxonomiques détectés par notre modèle IA sont : Oiseaux, Anoures, Orthoptère et Hémiptère. L'IA utilisée est un CNN de petite architecture avec un modèle/poids par espèce/taxons détectés.

En ce qui concerne les chiroptères, leur activité tant nocturne, le choix fait est de lancer la détection sur une période temporelle équivalente à la nuit +/- 1h. La tâche de classification des Chiroptères ne consiste désormais plus en la stratégie de développer une chaîne de traitement sur la base du niveau de fréquence / puissance du signal comme acté dans le précédent COPIL mais plutôt le même CNN que pour les groupes taxonomiques avec la détection de 5 taxons majeurs : Murin (*Myotis*), Pipistrelle (*Pipistrellus*), Rhinolophe (*Rhinolophus*), Oreillard (*Plecotus*) et Noctule (*Nyctalus*) sans pour autant tomber dans le surcoût d'énergie. Les métriques d'évaluation sont propres à chaque modèle et donc à chaque espèce/taxons (cf Tableau 1, 2 et 3). L'objectif d'une fiabilité de 80% d'aire sous la courbe ROC est donc atteint pour les Oiseaux, Anoures, Orthoptère, Hémiptère et les différents groupes de chiroptères, mais seulement pour une espèce d'oiseaux *Emberiza calandra*.

Livrable 2.3 Développement de l'Intelligence Artificielle embarquée acoustique

Un des ensembles QHB + micro stéréo est déjà déposé par TerrOïko, les 9 autres seront envoyés d'ici fin novembre. Ces ensembles admettent un envoi fréquent de leur état afin de connaître les statuts d'analyse ou d'écoute et ainsi décider ou non d'un éventuel arrêt. Un même fichier de configuration est présent sur les 9 ensembles QHB + stéréo permettant une écoute de 24h. L'écoute se fait pendant 5 minutes toutes les 30 minutes et l'analyse est exécutée pendant les temps OFF. Les classifications se font toutes par IA embarqués chaque espèce/familles/taxons ayant son propre modèle avec ses poids.

L'écoute permet la génération de spectrogrammes échantillonnée à 256kHz et analysée sur fenêtre de 10 secondes (2 x 5s). La tâche de classification des 4 taxons Orthoptère, Anoures, Oiseaux, Hémiptères est exécutée à chaque temps d'analyse et se fait à partir d'un Mel spectrogramme sur une plage de fréquence entre 50 Hz et 32000 Hz. La classification des 5 familles de Chiroptère se fait aussi sur Mel spectrogramme sur une plage de fréquence entre 10kHz et 128KHz et l'analyse des Chiroptères est uniquement lancée de 18h à 7h du matin. Si le modèle de détection d'oiseaux a une sortie positive, les modèles de détection des 10 espèces d'oiseaux sont exécutés alléant ainsi le temps d'inférence et la mémoire. Les 10 espèces d'oiseaux détectées sont les suivantes : *Emberiza calandra*, *Anthus pratensis*, *Emberiza hortulana*, *Coturnix coturnix*, *Curruca communis*, *Lullula arborea*, *Saxicola rubetra*, *Alauda arvensis*, *Emberiza cirrus*, *Emberiza citrinella*. Cette classification se fait sur le même Mel spectrogramme que les oiseaux. Un rapport contenant les indices éco-acoustiques ainsi que les détections des taxons/espèces est ensuite stocké dans la carte micro-sd, il peut être ainsi récupéré par la suite..

LOT 3. Base de données d'entraînement des IA

Livrable 3.1 Plateforme d'annotation des données acoustiques d'espèces => en attente

Workshop prévu sur la question en mi novembre

LOT 4. Conception des produits vision et écoute

L'ensembles des tâches attendues pour les livrables du lot 4 et effectuées se trouvent ci-dessous:

Livrable 4.2 Architecture et design général du système autonome d'écoute

- Définition et choix de l'architecture globale du système autonome Écoute
- Rédaction des spécifications fonctionnelles détaillées
- Etude du design industriel des enveloppes mécaniques pour répondre aux fonctions contraintes de résistances à l'environnement, de maintien en position et des contraintes de prise de son

Livrable 4.4 Conception du système autonome d'écoute

- Conception électronique embarquée
- Conception logiciel embarqué gestion de la mission

Rendu: Dossier de définition permettant la fabrication des prototypes (2-3 versions), Descriptif du protocole de communication, Descriptif du Profil de mission ?

Livrable 4.6 Développement et validation des prototypes du système autonome Écoute

- Fabrication dont intégration sous-ensemble audio (10 à 15 protos)
- Développement logiciel embarqué gestion mission
- Test unitaires
- Tests globaux

Rendu : Plan et résultats de validation système autonome Écoute, Version stable logiciel embarqué système autonome Écoute

Travail réalisé et principaux résultats

Depuis le dernier rapport de recherche, plusieurs avancements ont eu lieu, notamment le développement de l'implémentation des analyses acoustiques et la configuration des cartes sons, ainsi que la construction du rapport stocké sur carte sd avec les résultats des analyses. En parallèle de ce travail, pour les tâches de classification, des modèles aux architectures embarquables sur terrain ont été entraînés et optimisés.

I. Analyse vision IA online + IA offline

Sur la matrice de confusion du modèle on observe un fort taux de bonnes détections pour les espèces Cerf, Sanglier, Chevreuil et Autres. Cependant la classe Chamois, qui est sous-représentée, présente de moins bons résultats et augmente les taux de Faux Positifs des autres classes. Ainsi on a un taux moyen de vrais positifs à 0.8 en prenant en compte la classe Chamois, et à 0.88 si on ne la prend pas en compte. Ainsi on atteint le *taux de Vrais Positifs* > 0.8 dans les deux cas, mais les *taux de Faux Positifs* et *Faux Négatifs* sont à améliorer.

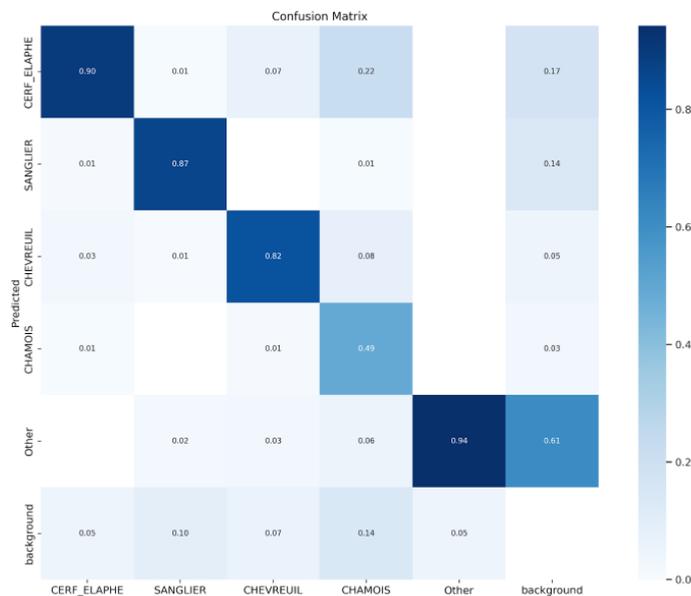


Figure 1. Matrice de confusion du modèle YOLOv5 320

De même, on observe sur les courbes suivantes (cf Fig 2.) que les résultats du modèle sont élevés pour toutes les classes sauf la classe Chamois, mettant encore une fois en lumière la nécessité d'un rééquilibrage du dataset ainsi que d'une meilleure sélection de classes.

A. Tâche classification taxons Orthoptère, Anoures, Oiseaux, Hémiptères

Le nombre d'annotation par groupes taxonomiques est assez faible (cf *Tableau 1*) pour autant les modèles de détection restent performants (cf *Figure 1*) avec les objectifs de stabilité atteints ($ROC > 0.8$).

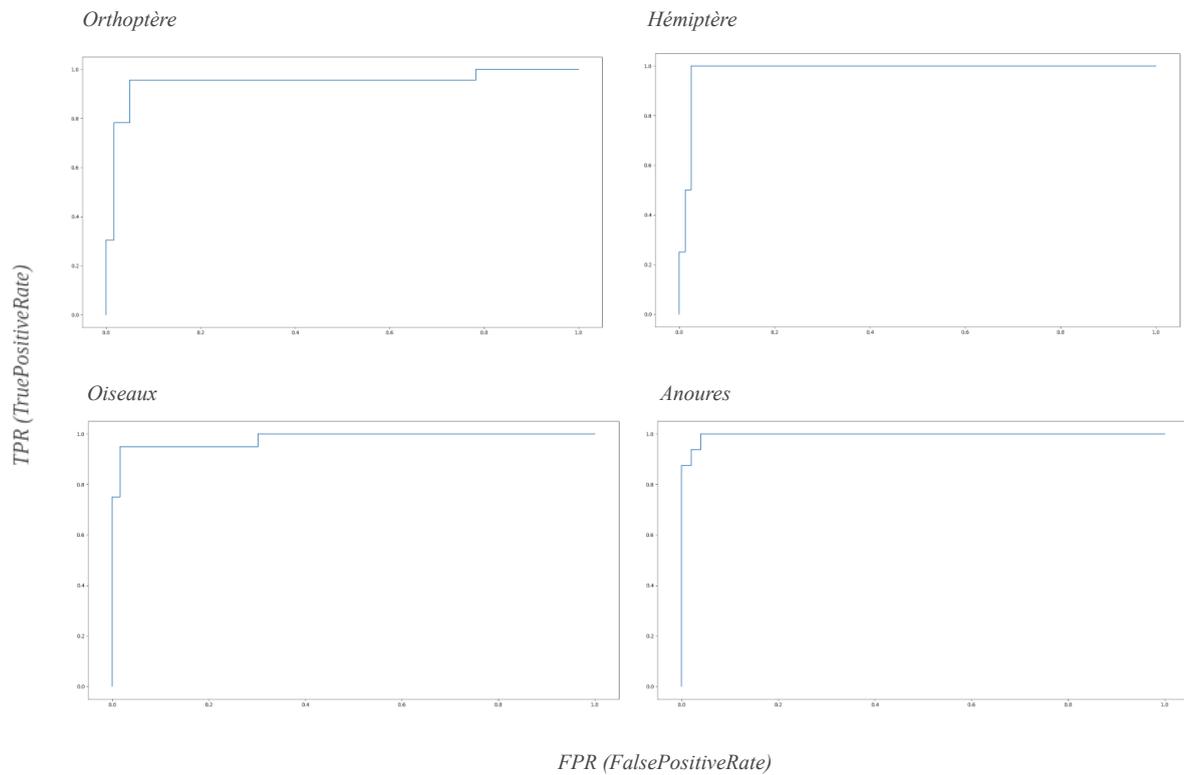


Figure 1. Courbe ROC des modèles lors du test pour les taxons Orthoptère, Hémiptère, Anoure et Oiseaux

Sélection d'un seuil de confiance

Les performances des 4 modèles Orthoptère, Hémiptère, Anoures et Oiseaux étant très bonnes, l'influence du seuil de confiance reste peu impactante. Le seuil de confiance par défaut (0.5) est donc choisi pour l'ensemble des taxons à l'exception des Hémiptères ou un seuil de confiance à 0.25 apportent de meilleurs scores (cf *Figure 1*).

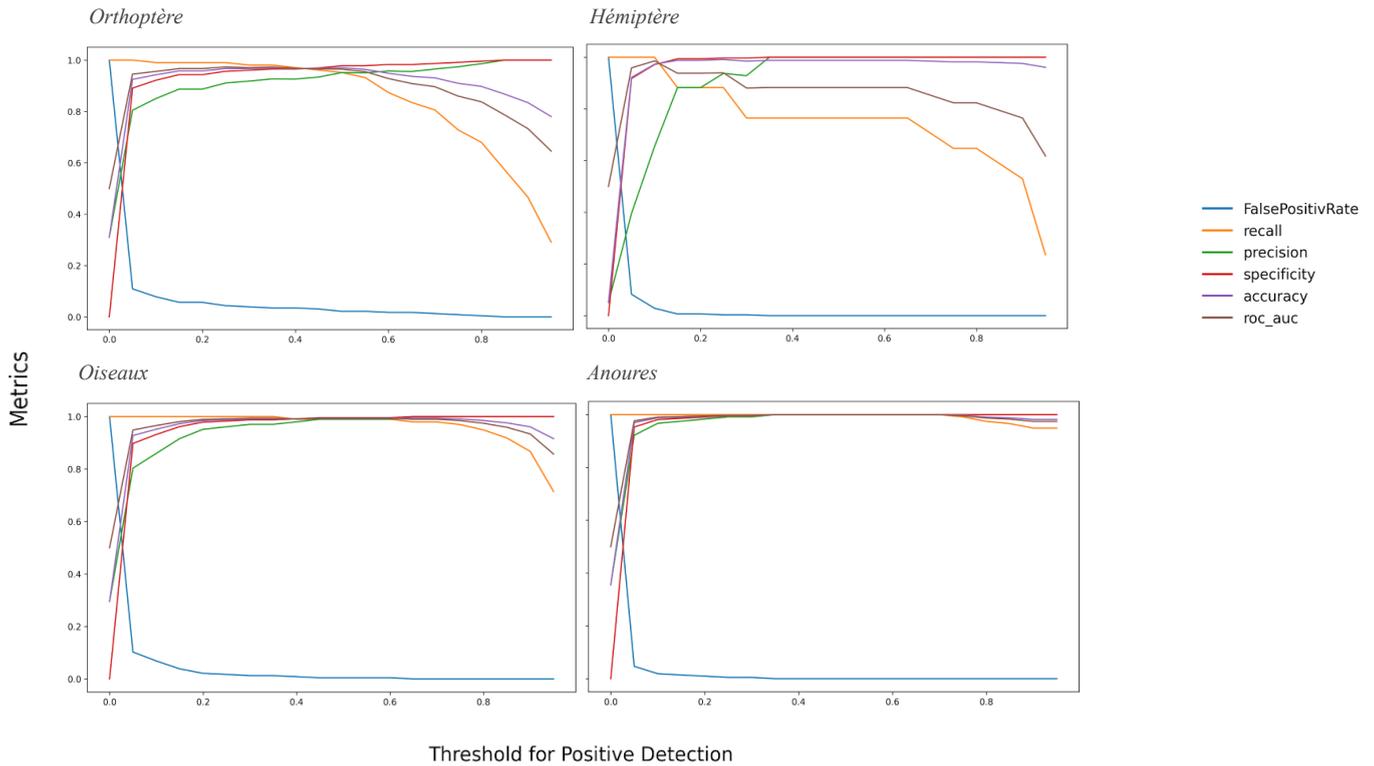


Figure 2. Evolution des métriques d'évaluation des modèles lors du test en fonction du seuil de confiance pour les taxons Orthoptère, Hémiptère, Anoures et Oiseaux

Evaluation des modèles

Les modèles CNN de détection de taxons Orthoptère, Hémiptère, Anoures et Oiseaux sont sur le dataset actuel très performants avec des scores de recall proche de 1, et une précision qui suit cette même tendance. Néanmoins ces modèles sont entraînés sur des datasets très petits, ici sur le set de test les Hémiptères ont seulement 17 images annotées (5 secondes/image). Les Anoures possèdent le plus gros dataset avec 118 images annotées. En conclusion les performances actuelles des modèles ne garantissent pas une bonne détection sur terrain si les enregistrements diffèrent ne serait que légèrement. Afin de remédier à cette limite de nouveaux enregistrements annotés sont nécessaires d'où la nécessité de récupérer les futurs enregistrements terrains avec la V1 des ensembles QHB + stéréo analyse intégrée.

Tableau 1. Récapitulatif des métriques d'évaluation des 4 modèles de détection de taxons (ajouter autres métriques)

species	total positive label	FalsePositivRate	recall	precision	specificity	accuracy	mAP	roc_auc	threshold positive pred
Anoures	118	0.000	1.00	1.00	1.00	1.00	1.00	1.00	0.5
Bird	98	0.004	0.99	0.99	1.00	0.99	0.98	0.99	0.5
Hémiptère	17	0.003	0.88	0.94	1.00	0.99	0.83	0.94	0.25
Orthoptère	103	0.022	0.95	0.95	0.98	0.97	0.92	0.96	0.5

B. Tâche de classification de familles de Chiroptères

Sélection d'un seuil de confiance

Les performances des modèles sont évaluées pour chaque famille de chiroptères

- Myotis 0,33
- Pipistrellus 0,6
- Rhinolophus 0,6
- Plecotus 0,3
- Nyctalus 0,3

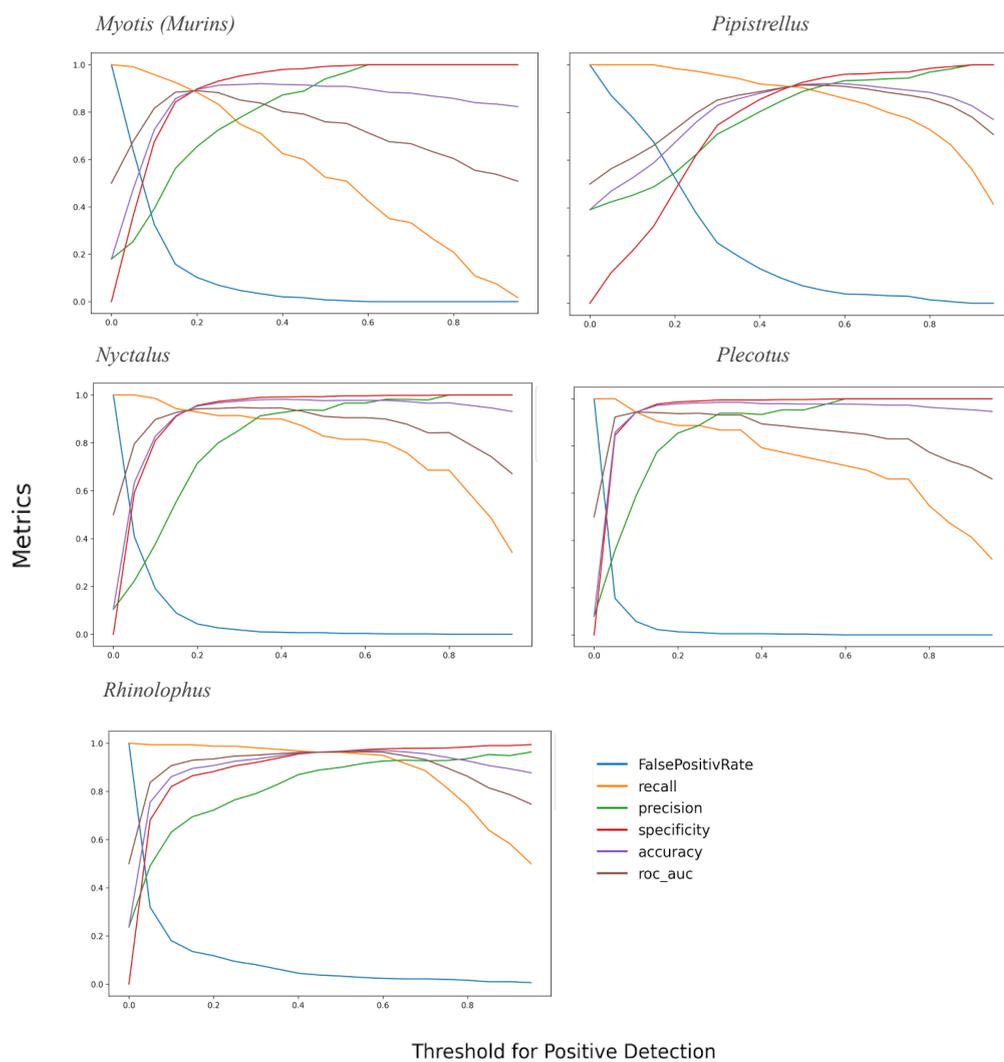


Figure 3 Evolution des métriques d'évaluation des modèles lors du test en fonction du seuil de confiance

Evaluation des modèles

Les valeurs de seuils choisies permettent de remplir les conditions suivantes :

- Taux de faux Positifs < 0,05
- Précision > 0,8
- Recall > 0,75

C.Tâche de classification IA de 10 espèces d'oiseaux d'intérêt

Comme le montre les courbes ROC ci-dessous (cf Figure 5) la détection des 10 espèces d'oiseaux indicatrices montre ses limites.

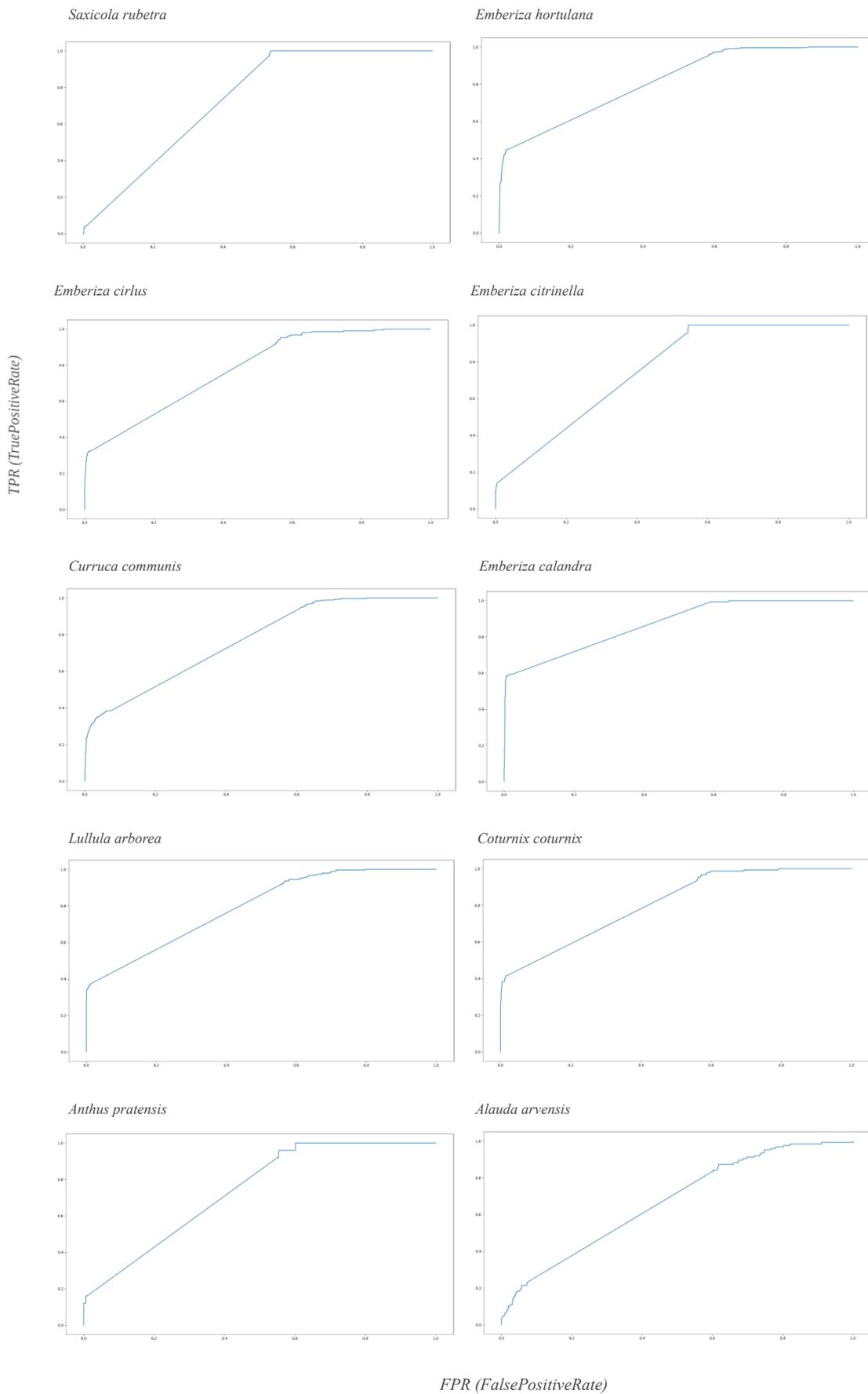
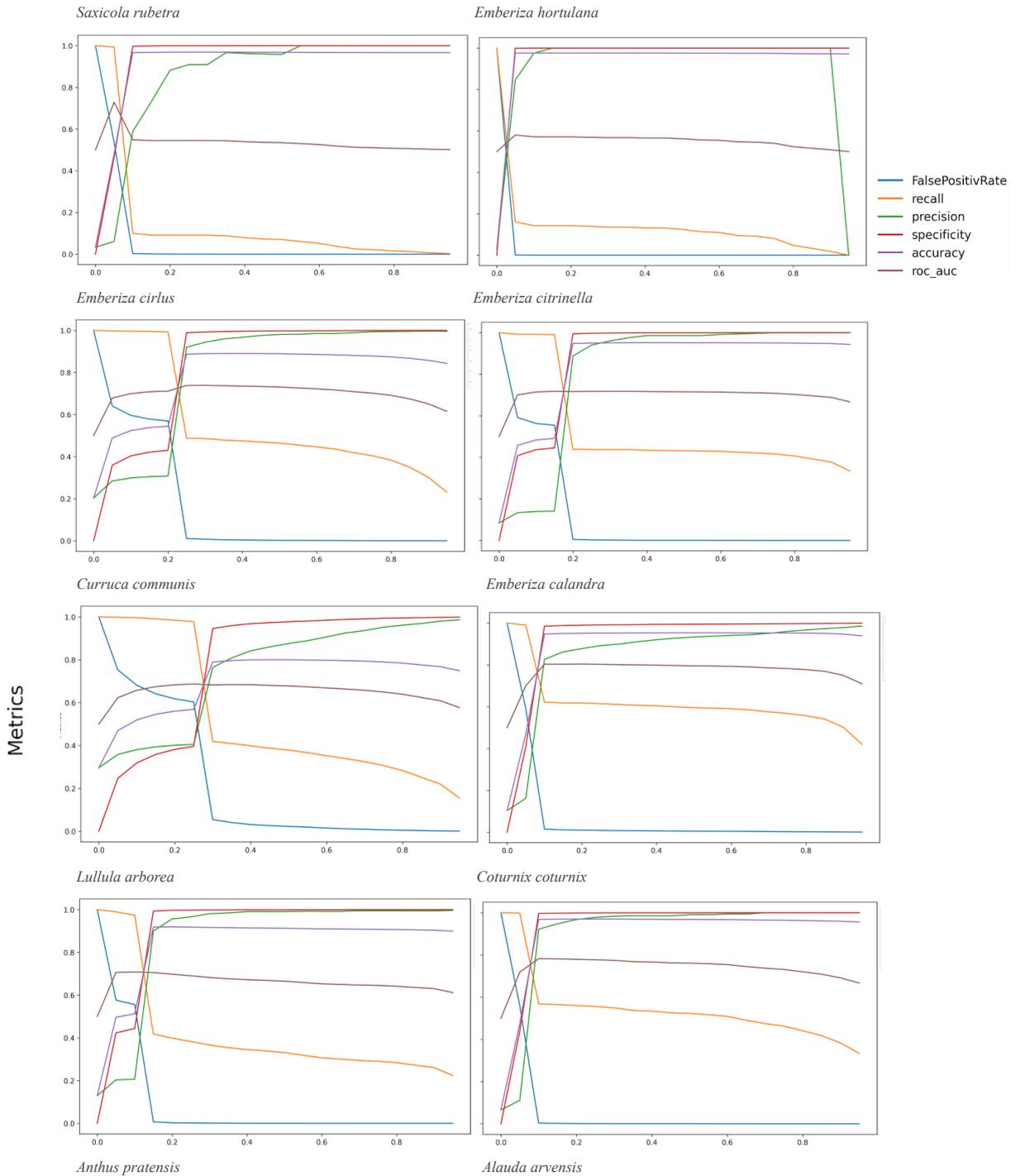


Figure 4. Courbe ROC des modèles lors du test pour les 10 espèces d'oiseaux d'intérêt

Sélection d'un seuil de confiance

Les modèles actuels ont des difficultés à détecter les espèces d'oiseaux malgré leur facilité à détecter des oiseaux de façon générale. Par conséquent, la sélection d'un seuil de confiance porte toute son importance ici. Les seuils de confiance ont été sélectionnés pour une valeur de *FalsePositivRate* d'environ 0, et un maximum de recall pour un valeur de précision à minima de 0.8 (cf Tableau 3).



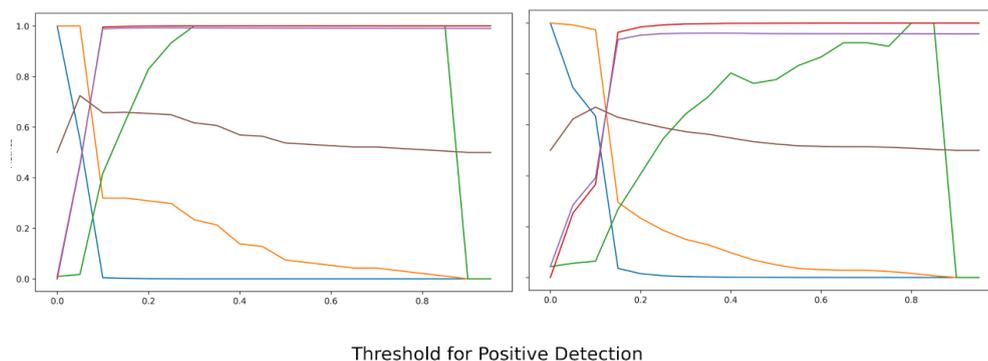


Figure 5. Evolution des métriques d'évaluation des modèles lors du test en fonction du seuil de confiance pour les 10 espèces d'oiseaux d'intérêt

Evaluation des modèles

Les 10 modèles CNN de détection d'espèce d'oiseaux sont pour le moment faibles en recall mais très bons en précision. Avec un *False Positive Rate* proche de 0, quasiment aucune détection n'est fautive néanmoins les modèles ne détectent pas suffisamment de vocalises d'oiseaux.

Tableau 2. Récapitulatif des métriques d'évaluation des 10 modèles de détection d'espèces d'oiseaux d'intérêt

species	total positive label	FalsePositivRate	recall	precision	specificity	accuracy	mAP	roc auc
Emberiza_calandra	1014	0.006	0.60	0.92	0.99	0.95	0.60	0.80
Anthus_pratensis	94	0.001	0.31	0.83	1.00	0.99	0.26	0.65
Emberiza_hortulana	273	0.000	0.14	1.00	1.00	0.98	0.17	0.57
Coturnix_coturnix	644	0.001	0.56	0.97	1.00	0.97	0.57	0.78
Curruca_communis	2905	0.023	0.38	0.88	0.98	0.80	0.52	0.68
Lullula_arborea	1271	0.007	0.42	0.90	0.99	0.92	0.45	0.71
Saxicola_rubetra	329	0.002	0.10	0.59	1.00	0.97	0.09	0.55
Alauda_arvensis	420	0.001	0.10	0.80	1.00	0.96	0.12	0.55
Emberiza_cirius	1994	0.007	0.49	0.94	0.99	0.89	0.56	0.74
Emberiza_citrinella	827	0.001	0.43	0.99	1.00	0.95	0.48	0.72

La détection des 10 espèces d'oiseaux d'intérêt reste à ce jour limitée par deux facteurs, les limites mémoire et énergétique d'une carte son embarquée et la taille du dataset annoté actuel. En effet, afin de porter cette détection sur terrain, les architectures des modèles ont été sensiblement réduites. Or la détection d'espèces d'oiseaux consiste à la différenciation de vocalises très similaires, tâche complexe ne pouvant pas être facilement accomplie par un CNN à peu de couches de convolution. Là où intervient le deuxième facteur, les annotations qui sont ici de 94 images contenant l'espèce désirée à 2905 pour son maximum. Augmenter la dataset serait la première clé dans le but d'obtenir des performances plus satisfaisantes, pour cela les données terrains qui seront obtenues par TerrOiko en automne 2024 seront essentielles et devront être envoyées au laboratoire LIS.

III. Implémentation des analyses acoustiques et configuration des ensembles QHB + stéréo

L'implémentation du calcul des Indices Acoustiques sélectionnés (*ACI* et *ADI*) a pu être faite et validée. Nous obtenons ainsi les mêmes résultats en off-line et en on-line sur des enregistrements égaux.

Une fois les modèles développés ils ont pu être portés sur les cartes QHB et validés par des tests en laboratoire. Ces tests ont mis en évidence les résultats suivants :

Temps d'inférence des modèles

Le temps d'inférence diffère selon les modèles, mais deux moyennes peuvent être associées aux modèles plus lourds (*Architecture 1*) et aux plus légers (*Architecture 2*). Ainsi nous avons pour les modèles lourds un temps d'inférence moyen de 7 secondes et pour les modèles légers un temps d'inférence moyen de 3.5 sec pour des fenêtres audio de 10 secondes.

Dans les “meilleures conditions” actuelles (au sens du programme embarqué, c.-à-d. lorsque le moins de modèles sont lancés) le temps d’inférence total en prenant en compte la création des mel-spectrogrammes se situe autour de 2 minutes. Dans le pire des cas, le temps d’inférence total approche les 5 minutes.

Protocole d’enregistrement

Les résultats de temps d’inférence ci-dessus nous obligent à penser un protocole optimisé afin de minimiser le nombre de CNNs utilisés pour une analyse. Des discussions avec les équipes d’écologues sont donc en cours pour sélectionner des plages horaires d’activité optimales et concentrer les écoutes et analyses sur ces plages horaires. Un exemple de règle pour un tel protocole serait l’analyse de présence des chiroptères seulement dans les heures proches des couchés et levés de soleil, heures où elles sont actives. Ce qui évite de lancer 5 CNNs en pleine journée.

Remontée GSM des données

Un certain nombre de données sont remontées par GSM à chaque fois qu’une analyse est réalisée sur une plage de 10 secondes :

- Les scores des détections sur la plage données
- Les instants des détections dans les fichiers audio
- Les indices acoustiques de l’enregistrement
- La date/heure de l’analyse

Ces données seront traitées en off-line par Terroiko afin d’en tirer des conclusions naturalistes.

De plus, les instants de détections permettent de retrouver les fichiers audio correspondants aux détections afin de réaliser une étude plus approfondie (par exemple la durée des chants).

Conclusion et perspectives

De façon générale, les modèles Vision et Acoustique pourront être améliorés grâce aux retours des prototypes terrain. Une seconde version V2 devrait montrer de meilleures performances et une meilleure adaptation aux outils hardware. Le recueil des prochaines données terrain pourrait permettre la vérification des modèles et amélioration de ces derniers, ainsi que viser une optimisation d’heure d’écoute selon les présences. Une fois la dernière version des ensembles Vision et Écoute, la prochaine étape sera le LOT3, l’intégration des analyses sur la plateforme OCAPI, donc travailler sur la transmission des rapports d’analyses des ensembles acoustiques et vision.