CONVENTION DE PARTENARIAT DE RECHERCHE PNG – CNRS (CNRS N 074099)

ANNEXE 1

Mise au point d'algorithmes de détection et d'identification automatique de vocalisations animales

Maxime Sainlot [1] — maxime.sainlot@insa-lyon.fr
Thierry Aubin [2] — thierry.aubin@u-psud.fr

Rapport de stage 4BiM

[1]*Bio-Informatique et Modélisation, INSA de Lyon, Villeurbanne, France*
[2]*Maitre de Stage, Directeur de Recherche CNRS, CNPS - Equipe Communication Acoustique, Orsay, France*

INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON — UNIVERSITÉ PARIS-SUD 11 — cnrs — Parc national de la Guadeloupe

1. Introduction

1.1 Présentation du sujet de stage

Deux populations de grenouilles ont été enregistrées en différents endroits du parc National de la Guadeloupe . L'objectif était de détecter automatiquement les cris de ces espèces de grenouilles afin de pouvoir les quantifier et estimer leur densité.

1.2 Intérêt global dans le contexte actuel

Aujourd'hui, en bioacoustique, les différentes étapes de traitement des données enregistrées (une fois de retour du terrain) sont généralement faites manuellement (voir 2.7 page 5 pour plus de détails sur les étapes). Ces enregistrement réalisés sur plusieurs jours sont souvent conséquents (plusieurs dizaines d'heures) et leurs traitements peut s'avérer laborieux. Depuis quelques années se développent des techniques informatiques de traitement automatique de ces enregistrements appelés BigData. Cependant il n'y a pas de réponse simple à ce problème qui est actuellement à l'état de recherche. C'est dans ce contexte que s'inscrivent les problématiques du stage : la confrontation aux difficultés rencontrées lors du traitement automatique de données audio et la recherche, sans pré-connaissance, de solutions à ce problème très actuel.

2. Matériels & Méthodes

Cette section présente des notions de base en traitement de signal, indispensables pour l'étude des communications acoustiques et donc pour la compréhension des problématiques et de leurs résultats.

2.1 Caractéristiques du son

Le son est une onde qui se propage longitudinalement par compression et décompression du milieu. Cette onde nécessite un milieu pour se propager (fluides ou milieux solides) et ne peut donc se déplacer dans le vide.

On peut caractériser le son notamment par deux grandeurs :

La fréquence correspond à un nombre d'oscillations par seconde. Cette grandeur, notée f, est mesurée en Hertz (Hz). Elle permet de définir la hauteur du son. Plus la fréquence est faible, plus le son est grave et inversement, plus la fréquence est élevée, plus le son est aigu.

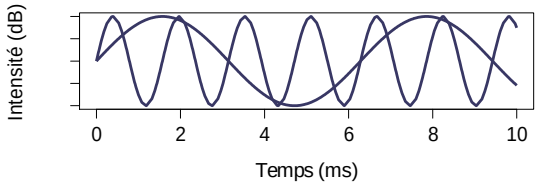


FIGURE 1. Deux courbes de mêmes amplitudes mais de fréquences différentes. Courbe sombre de fréquence faible (son grave) et courbe claire de fréquence élevée (son aigu).

On définit la période T comme la durée d'une oscillation complète. Ces deux grandeurs sont liées par la relation $f = \frac{1}{T}$.

Tout comme l'œil avec les couleurs, l'oreille humaine présente certaines limites. Certains sons ne peuvent être entendus. Habituellement, l'oreille humaine est capable de percevoir des sons allant de $20Hz$ à $20kHz$. Les bornes inférieures et supérieures de cet intervalle définissent respectivement les plages des infra- et des ultra-sons.

L'intensité sonore, l'amplitude, le niveau sonore, l'énergie, ou encore la pression acoustique sont autant de grandeurs aux significations différentes mais évoluant toutes dans le même sens pour décrire la puissance d'un son. On exprime généralement cette intensité en décibel (dB).

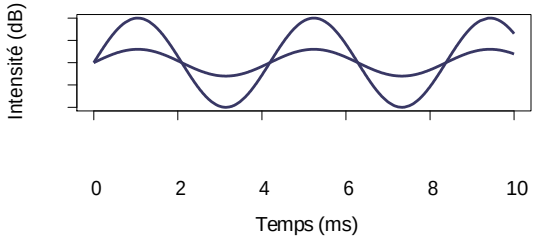


FIGURE 2. Deux courbes de mêmes fréquences mais d'amplitudes différentes. Courbe sombre d'amplitude élevée (son fort) et courbe claire d'amplitude faible (son faible).

FIGURE 3. Echelle des dB.

2.2 La transformée de Fourier

On définit comme son pur, un signal périodique sinusoïdal composé d'une fréquence unique. Un son complexe consiste en une somme de sons purs de fréquences et d'amplitude s différentes. Un son complexe reste néanmoins périodique.

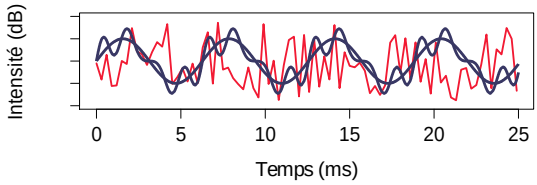


FIGURE 4. Son pur (bleu foncé), son complexe (bleu clair), bruit (rouge)

La théorie des séries de Fourier expose la possibilité de décomposer n'importe quel signal périodique S en une somme de sinus et de cosinus.

$$S = \sum_{n=0}^{+\infty} a_n cos(nt) + b_n sin(nt)$$

La transformée de Fourier est une généralisation des séries de Fourier aux fonctions non-périodiques. Elle permet d'associer à une fonction intégrable (notre son), une nouvelle fonction (appelée transformée de Fourier) dont la variable n'est plus le temps mais la fréquence. Elle s'écrit de la sorte :

$$F(f) : \nu \mapsto \hat{f}(\nu) = \int_{-\infty}^{+\infty} f(t)\, e^{-i2\pi\nu t}\, dt$$

avec t en secondes ν la fréquence (en Hz) [1].

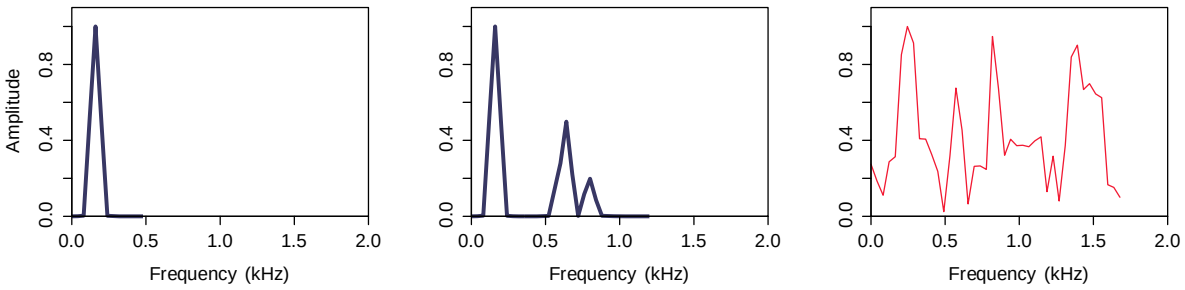


FIGURE 5. Spectre fréquentiel obtenu par la transformée de Fourier des signaux de la figure 4. Le signal bleu foncé est un son pur constitué d'une sinusoïde de fréquence de 150Hz, le signal bleu clair est un son complexe constitué de 3 sinusoïdes de fréquences et d'amplitudes différentes et que le bruit est une combinaison de nombreuses sinusoïdes.

2.3 Numérisation/Enregistrement

Dans la nature, le son est analogique et donc continu. Enregistrer un son revient à le numériser et cela nécessite de discrétiser le signal en mesurant des valeurs séparées par un intervalle de temps. On parle d'échantillonnage. Afin d'obtenir un enregistrement de bonne qualité, l'échantillonnage doit se conformer à certaines règles. En effet, seules sont conservées les valeurs mesurées à chaque intervalle de temps. Ainsi, si l'intervalle de mesure est trop important, une part de l'information sera perdue et l'enregistrement sera de mauvaise qualité. En revanche si l'intervalle est trop faible, cela risque de générer un fichier de taille inutilement importante.

Comment bien choisir sa fréquence d'échantillonnage ? Le théorème de Nyquist stipule que la fréquence d'échantillonnage doit être 2 fois supérieure à la composante fréquentielle la plus élevée du signal mesuré sans quoi l'ensemble des composantes hautes fréquences risque de se replier à une fréquence comprise dans le spectre d'intérêt (bande passante) [2].

1. www.fr.wikipedia.org/wiki/Transformation_de_Fourier
2. www.ni.com

La composante fréquentielle audible la plus élevée étant aux alentours de $20kHz$, la fréquence d'échantillonnage conseillée serait donc de $40kHz$ (40 000 mesures par seconde). En règle générale et par sécurité, les échantillonneurs utilisent une fréquence d'échantillonnage de $44,1kHz$.

2.4 Représentation/Visualisation

2.4.1 Oscillogramme

C'est la représentation la plus connue (cf. figure 6). On visualise l'amplitude du son en fonction du temps. C'est le signal brut (tel qu'il est enregistré). Cette représentation ne fournit pas directement d'information sur la fréquence.

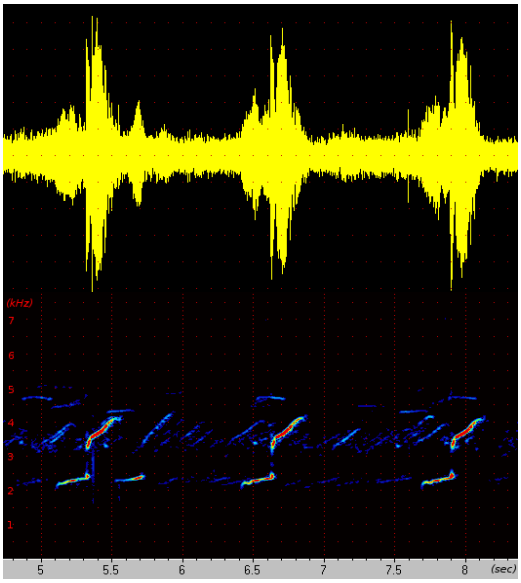


FIGURE 7.

2.4.2 Sonagramme

Cette représentation est la plus utilisée en bioacoustique (cf. figure 7). Elle a l'avantage de fournir simultanément les informations relatives au temps, à la fréquence et à l'amplitude du son. Ces trois grandeurs permettent à elles seules de décrire un son. Sur cette représentation le temps est en abscisses et les fréquences sont en ordonnée (hauteur de la note). L'échelle d'intensité des couleurs des pixels donne une indication sur l'intensité sonore (pour une fréquence donnée à un temps donné). Cette visualisation est finalement beaucoup plus intuitive que la précédente et apporte rapidement un nombre important d'informations (en 3 dimensions).

Cette seconde représentation est obtenue en effectuant la transformée de Fourier de l'oscillogramme à l'aide d'une fenêtre temporelle que l'on fait glisser le long du signal. La fenêtre doit être suffisamment grande pour pouvoir estimer les fréquences présentes (et leurs amplitudes respectives) dans la portion de signal, mais suffisamment étroite pour conserver une précision temporelle. Cela implique une impossibilité d'être précis en fréquence et en temps simultanément. Pour augmenter toutefois cette précision, on peut choisir de chevaucher la fenêtre avec son emplacement précédent lors de son décalage. Une fois la transformée effectuée sur la première fenêtre temporelle, on décale cette fenêtre puis on recommence jusqu'à la fin du signal (transformées de Fourier glissante). Il s'agit alors de trouver un compromis entre précision fréquentielle, temporelle, et temps de calcul.

2.5 Spécificité de la propagation sonore

Les obstacles (feuilles, arbres, bâtiments . . .) altèrent la propagation du son et jouent le rôle de filtre. Les sons graves se propagent mieux et plus loin au travers des obstacles. Toutefois, ils sont difficilement localisables, c'est-à-dire qu'il n'est pas facile de localiser la source d'émission d'un son grave par opposition aux sons aigus qui se propagent moins loin (vite stoppés par les obstacles) mais qui permettent une localisation spatiale plus aisée.

2.6 Les expériences basiques

2.6.1 Expérience de playback

C'est une expérience où l'on diffuse un signal acoustique à l'espèce étudiée. Sa réaction face à cette diffusion permet de déduire la fonction de ce signal. Un enregistrement modifié peut permettre d'identifier la partie du son qui signe pour la réaction attendue (détection d'un intrus, signature d'un cris de détresse. . .).

2.6.2 Expérience de propagation

C'est une expérience permettant d'évaluer l'impact de l'environnement sur la transmission du signal sonore. Une fois émit, le signal subit les modifications de l'environnement. Il est soumis à différentes contraintes qui vont l'atténuer et le bruiter. C'est notamment la nature de l'environnement qui va générer ces contraintes. L'expérience

de propagation va avoir pour but d'évaluer la proportion et la nature du bruit additionné en fonction de la distance à la source et du milieu traversé. Concrètement, après avoir réalisé l'enregistrement «focus» (à un mètre de l'individu), on va le diffuser et le réenregistrer à différentes distances et dans différents environnements. Ainsi, on est capable de déterminer un espace actif (*active space)* qui correspond à la région entourant l'individu et pour laquelle le signal est audible et porteur d'information.

2.7 Les étapes basiques du traitement automatique

2.7.1 La segmentation

C'est une étape qui consiste à identifier précisément le début et la fin de chaque unité d'intérêt dans un enregistrement. Une segmentation peut avoir plusieurs échelles. C'est-à-dire que l'unité d'intérêt peut être le chant en lui-même (dans sa totalité), la phrase (portion de chant), la syllabe (succession de quelques notes), la note (unité élémentaire)…

2.7.2 La labellisation

Cette étape intervient nécessairement après la segmentation. Il s'agit d'attribuer un identifiant à chaque segment permettant ainsi de grouper les unités d'intérêt identiques.

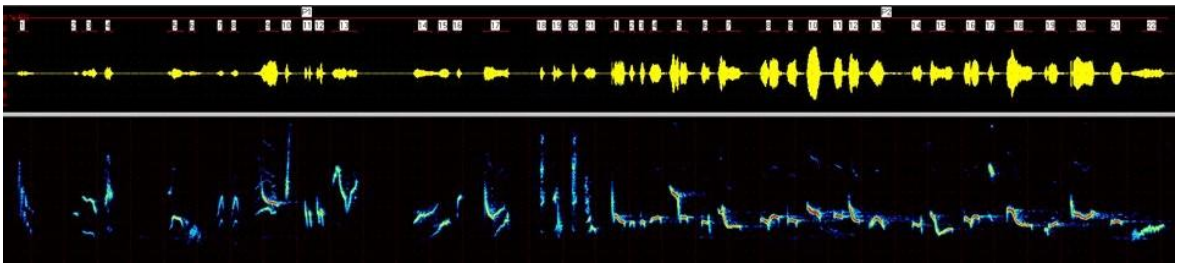


FIGURE 8. Un sonagramme de Fauvette à tête noire segmenté par phrase et par syllabes.

2.8 Les outils informatiques d'analyses et de traitements

2.8.1 Avisoft[3]

Avisoft est un outil très complet d'analyse bioacoustique. Il inclut de nombreuses possibilités pour traiter les sons. Il permet notamment de visualiser le son (oscillogramme et sonagramme), de le segmenter et de le labelliser, de réaliser différentes transformations (Fourier) et d'effectuer des mesures. Les figures 7, 6, 8, 11d et 11c on été réalisées sous Avisoft.

2.8.2 Matlab[4]

Désignant à la fois l'environnement de programmation et le language, Matlab est un outil puissant et polyvalent permettant la création d'algorithmes de calcul et de visualisation dans des domaines extrèmement variés par le biais de *toolboxes* spécifiques. Il est notamment tout à fait adapté au traitement du signal et à ses applications en bioacoustiques.

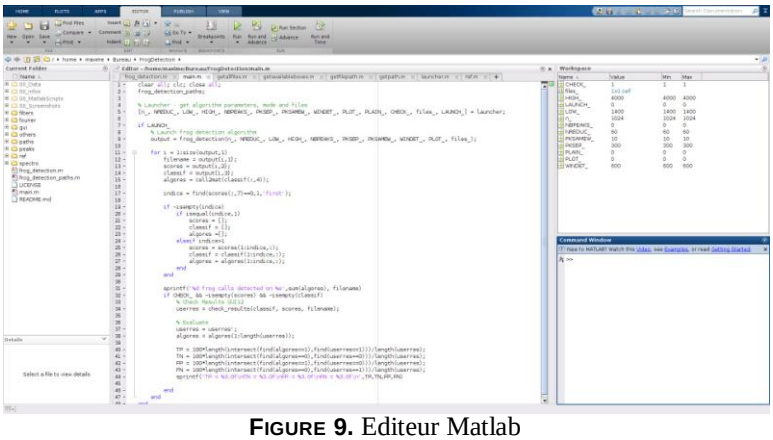


FIGURE 9. Editeur Matlab

3. http ://www.avisoft.com/soundanalysis.htm
4. http ://www.mathworks.fr/products/matlab/

Détection, comptage et discrimination automatique de 2 espèces de grenouilles

1. Introduction

1.1 Présentation des espèces

FIGURE 10. La Soufrière

Sur les pentes du volcan de la Soufrière en Guadeloupe (cf figure 10) existent deux espèces endémiques de grenouilles que l'on retrouve uniquement dans les Antilles :

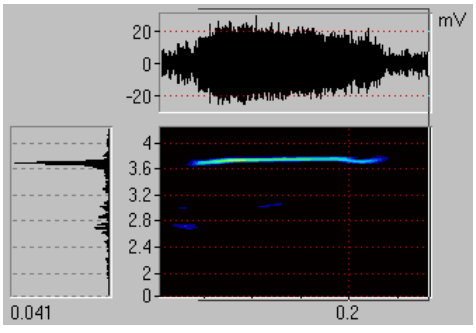
- *Eleutherodactylus pinchoni* ou Hylode de Pinchon (cf. figure 11a)
- *Eleutherodactylus martinicensis* ou Hylode de Martinique (cf. figure 11b)

Les Hylodes sont proches des rainettes mais se différencient de celles-ci par l'absence de palmures aux pattes postérieures. Elles font entre 20 et 50 mm et sont de couleurs ternes. Ces deux espèces cohabitent.

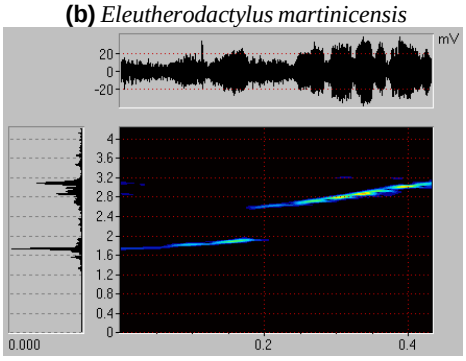
L'Hylode de Jonhnstone (*Eleutherodactylus johnstonei*) est, quant à elle, une grenouille colonisatrice. Depuis sa découverte, cette grenouille est suspectée d'être compétitrice de l'Hylode de Pinchon et potentiellement de l'Hylode de Martinique si elle venait à s'introduire dans la forêt humide. Ces espèces sont donc respectivement «en danger» et «quasi-menacée» [5] (cf. figures 11a, 11b et 12).

(a) *Eleutherodactylus pinchoni*

(b) *Eleutherodactylus martinicensis*



(c) Sonagramme (centre), spectre fréquenciel (gauche) et oscillogramme (haut) de *Pinchoni*



(d) Sonagramme (centre), spectre fréquenciel (gauche), et oscillogramme (haut) de *Martinicensis*.

FIGURE 11. Illustration des grenouilles et de leur profil vocal respectif

Les cris de ces grenouilles, visibles en figures 11c et 11d, sont peu modulés en fréquence. Ils varient peu d'un individu à l'autre mais restent relativement identiques entre les individus d'une même espèce. Ces cris sont produit principalement la nuit ou durant une pluie. Ils peuvent être [1] :

5. D'après la liste rouge de l'IUCN : http ://www.iucnredlist.org

- des avertissements territoriaux à destination des autres mâles pour avertir de la présence d'un « propriétaire » et de sa propension à défendre le territoire.
- des appels sexuels des mâles à destination des femelles.

1.2 Pourquoi étudier ces espèces ?

Sur les pentes de la Soufrière, on distingue schématiquement deux types d'environnement : la forêt, à la végétation dense située sur la partie basse du volcan, et la savane, à la végétation rase voire inexistante sur la partie haute. Les deux espèces de grenouilles se retrouvent dans chacun des deux environnements. Toutefois, elles présentent des différences morphométriques et de hauteur des cris d'un environnement à l'autre notables et significatives.

Lors d'une expérience de playback, la diffusion de l'enregistrement d'un individu de la même espèce et provenant du même environnement suscite des réactions de défense de la part de l'individu présent sur le territoire (attaque du haut parleur). En revanche, la diffusion de l'enregistrement d'un individu de la même espèce mais ne provenant pas du même environnement (forêt / savane) suscite peu de réaction chez le défenseur du territoire, ce qui laisse penser que ces espèces sont en voie de spéciation. Ce fait est d'autant plus troublant qu'en 1976 la Soufrière est rentrée en éruption éradiquant en totalité les populations présentent sur la partie haute du volcan (savane). Nous serions alors en train d'assister à un phénomène de spéciation rapide dont nous pourrions dater l'origine.

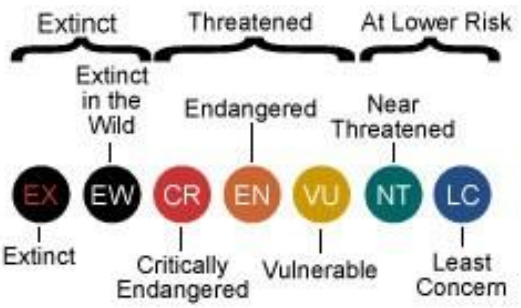


FIGURE 12. Status IUCN

D'autre part, ces espèces étant menacées, le parc National de Guadeloupe a décidé de mener une étude pour pouvoir évaluer l'évolution de ces populations au fil des saisons et des années. L'approche acoustique a donc été envisagée comme une possibilité d'estimation de l'évolution des densités de populations.

1.3 Plan d'expérience et conditions d'acquisition des données

Thierry Aubin et Renaud Boistel se sont rendu à plusieurs reprises sur le terrain avec un prototype d'enregistreur automatique réalisant 1 minute d'enregistrement continu toutes les heures et ce durant plusieurs jours à différentes périodes de l'année. Plusieurs de ces dispositifs on été répartis à différents endroits du parc et ont enregistré les sons de la nature environnante (voir rapport).

1.4 Objectifs

L'objectif a été de créer un algorithme capable d'automatiquement détecter les vocalisations des espèces de grenouilles. Idéalement, il serait intéressant qu'il puisse discriminer les deux espèces, c'est-à-dire les compter séparément.

1.5 Difficultés rencontrées

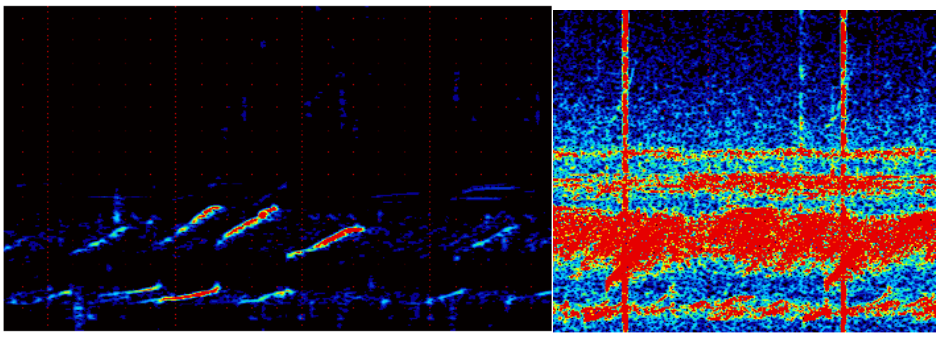
1.5.1 Le rapport signal-bruit

La difficulté première de ce jeu de données vient du rapport signal-bruit (SNR pour *signal-noise ratio*) très faible. La présence d'autres espèces dans la forêt (cigales, criquets et oiseaux principalement) émettant des vocalisations dans les mêmes gammes de fréquences va avoir pour effet d'augmenter le bruit de fond et donc de diminuer le SNR. D'autre part, les distances des individus au microphone vont avoir un impact sur l'intensité du signal.

1.5.2 La limite signal-bruit

Les notions de signal et de bruit sont importantes. Elles sont différentes pour chaque jeu de données et peuvent varier suivant la problématique que l'on se pose. Lorsque l'on souhaite détecter des vocalisations précises, le signal correspond à ces vocalisations et le bruit correspond « au reste ». Seulement, en regardant la figure 13a on aperçoit 3 signaux qui émergent bien du bruit de fond par leur intensité. On peut, en descendant notre seuil visuel d'intensité, en compter 2 de plus. Puis, en étant vraiment très tolérant, on peut en trouver encore d'autres (car il y en a d'autres !). La question se pose alors de savoir où se trouve la limite entre ce qui correspond au signal et ce qui correspond à du bruit. L'atténuation progressive du son en fonction de la distance au micro ne créé pas de limite franche entre le signal et le bruit de fond. À une certaine distance, des chants de grenouilles seront sous le bruit de fond ou en limite de bruit de fond. Même quelques dB en dessous, des signaux restent audiblement discriminables. Dès lors, il devient très compliqué de les détecter.

C'est là toute la difficulté de la détection. Nous savons que nous ne pourrons pas tout détecter, mais jusqu'où se rapprocher du bruit de fond ? Cela revient à trouver un compromis entre sensibilité et mauvaise détection.



(a) Signal acceptable **(b)** Signal bruité

FIGURE 13. Qualité du signal

1.5.3 La qualité du signal

Il arrive fréquemment que l'on obtienne un signal encore plus difficilement exploitable. C'est le cas sur la figure 13b. Le vent, la pluie, la présence d'autres espèces ou même la trop forte densité d'individus peut être à l'origine de ce signal bruité. Le vent et la pluie augmentent le bruit de fond sur l'ensemble des fréquences diminuant ainsi le rapport signal-bruit. Des sons ponctuels peuvent également détériorer le signal, c'est le cas de deux «clacs» visibles sur cette même figure. Finalement, une forte densité d'individus augmente le risque de chevauchement (*overlap*) de leurs vocalisations. Il en résulte une coloration «grenouille» visuellement évidente grâce aux bandes continues de fréquence nous informant de la présence d'individus. Toutefois, leurs chants n'étant pas distincts, il devient difficile voire impossible de les compter.

1.5.4 La variabilité biologique

Quand il s'agit de biologie, il n'y a pas de constance mais des variabilités plus ou moins importantes. Les vocalisations que nous tentons de détecter, même si elles suivent à peu près le même schéma de modulation de fréquence au cours du temps au sein d'une espèce, conservent une certaine variabilité intra spécifique. Il n'est alors pas évident d'identifier les paramètres pertinents précis pour détecter certains sons spécifiques.

2. Méthode

Il convient de préciser que l'algorithme produit a été conçu en se focalisant exclusivement sur la détection des *martinicensis*. **Le travail pour la détection spécifique de *pinchoni* n'a pas été réalisé dans le cadre de ce stage, mais suit en tout point la méthodologie utilisée pout l'autre espèce.**

2.1 L'algorithme

Les premières étapes sont triviales :

- Importation du fichier audio à l'aide d'`audioread` (fonction Matlab déjà implémentée) et récupération des paramètres d'enregistrement (fréquence d'échantillonnage).
- Passage de la source audio en mono si nécessaire.
- Définition des variables et récupération des constantes nécessaires au traitement (passées par l'utilisateur).

On rentre alors dans le cœur de l'algorithme qui se compose de trois étapes :

2.1.1 La segmentation

Il s'agit de détecter les cris de grenouilles. C'est ce qu'on appelle de la segmentation. Nous avons dès le départ abandonné l'idée de détecter ces chants en dessous du bruit de fond. L'algorithme va tenter alors de déterminer un seuil d'intensité définissant une limite entre le bruit de fond et le signal sur l'oscillogramme. Ce seuil permet d'exclure les portions de silence (au sens d'absence de signal) de la recherche. Cela n'est cependant pas suffisant. Sur les portions d'enregistrement restant, on va rechercher les pics d'intensités.

Pour se faire, on recherche les maxima d'amplitudes. Ces pics ont toutefois des contraintes pour ne pas en trouver une infinité. Chaque pic doit être séparé d'une certaine valeur (en temps) du pic suivant. L'utilisateur est libre de modifier cette valeur (voir `Peak separation` sur la figure 14) qui est par défaut de 300 ms soit la durée moyenne des cris de grenouilles (après mesures manuelles). Ainsi, on impose une limite à la détection. Si deux individus vocalisent à un intervalle de temps très faible, l'algorithme n'en détectera qu'une. Ce choix se justifie afin de ne pas faire l'erreur inverse qui serait de trouver 2 cris (en début et en fin de vocalisation) là où il n'y en a qu'un. On se rend dès à présent compte que certains choix, même s'ils sont nécessaires, détérioreront la qualité de notre détection.

On obtient ainsi une première segmentation peu contraignante qui est sensée nous indiquer les probables instants de vocalisations. Toutefois, une grande majorité de ces points ne sont que bruit de fond.

2.1.2 Le calcul de scores

L'étape suivante est la mise en place d'une méthode de *scoring*. Pour chaque pic trouvé l'algorithme se focalise alors sur une portion de signal (une fenêtre temporelle) autour du pic. Cette fenêtre mesure par défaut 2 fois la taille moyenne des vocalisations soit 600 ms (voir `Window size` sur la figure 14). De cette façon, où que soit trouvé le pic d'intensité sur le signal, l'analyse inclura forcément la totalité de la vocalisation. Afin de décider si le pic d'intensité identifié correspond effectivement à une vocalisation, 4 scores sont calculés. Ils sont basés sur :

- **L'évolution de l'intensité par filtre passe-bande :** Un filtre passe-bande consiste à laisser passer uniquement une plage de fréquences donnée. Ces fréquences sont choisies pour conserver notre signal et éliminer assez grossièrement du bruit de fond. Par défaut, pour ces espèces de grenouilles, l'algorithme propose une plage $[1.4 - 4]kHz$ avec la possibilité pour l'utilisateur de la modifier (voir `Bandpass filter`, figure 14). Lors d'un tel traitement, la valeur globale de l'amplitude (intensité) est diminuée, seulement, elle diminue d'autant plus que la portion de signal possédait des valeurs d'intensité importantes pour des fréquences en dehors de la plage conservée relativement à celles à l'intérieure de cette plage.

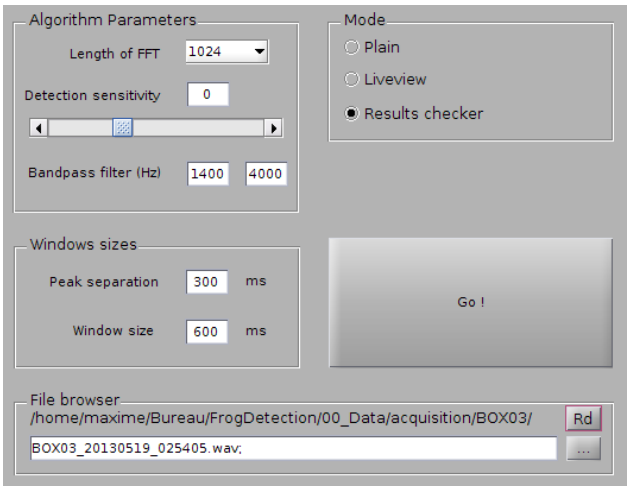


FIGURE 14. Lanceur d'algorithme

Ainsi, un «clac» (cf figure 13b), détecté comme potentielle vocalisation lors de la segmentation verra (sur l'oscillogramme) son intensité réduite de manière plus importante que celle d'une portion de vocalisation (dont la majorité de l'intensité se trouve dans la plage de fréquence). Cet indice de réduction par application d'un filtre passe-bande permet ainsi d'écarter ces «clac» de la détection.

- **L'adéquation avec la bande de fréquence recherchée :** En travaillant sur le spectre fréquentiel du signal brut (sans filtre passe-bande), l'algorithme va favoriser la présence de fortes intensités dans la plage de fréquence du filtre passe-bande en attribuant des scores élevés proportionnellement aux intensités.

- **La corrélation croisée `xcorr` par rapport à une référence :**Le spectre fréquentiel de *martinicensis* est caractéristique (voir figure 11d partie gauche), il est constitué de 2 pics : le premier à $\sim 1.8kHz$ et le second à $\sim 3kHz$. Un spectre fréquentiel de référence est obtenu par moyenne d'une vingtaine de spectres fréquentiels provenant d'individus différents. L'algorithme va alors tenter de superposer cette référence au spectre du pic détecté courant en les faisant glisser l'un par rapport à l'autre et en calculant, pour chaque pas, une corrélation de Pearson. La valeur de la corrélation maximale et le décalage entre courbe de référence et pic courant pour la valeur de corrélation maximale servent à constituer le 3ème score. Ce score permet de s'assurer que le pic détecté corresponde bien à une *martinicensis*.

- **L'inverse de la somme des carrées des écarts par rapport à une référence :** Cette mesure nous fourni une information sur la qualité de la détection. Ce score peut être vu comme un indice de confiance car un signal qui émergerait tout juste du bruit de fond aurait une somme des carrées des écarts plus importante qu'un signal ayant un bon rapport signal/bruit. Ce score pondère donc les scores précédents.

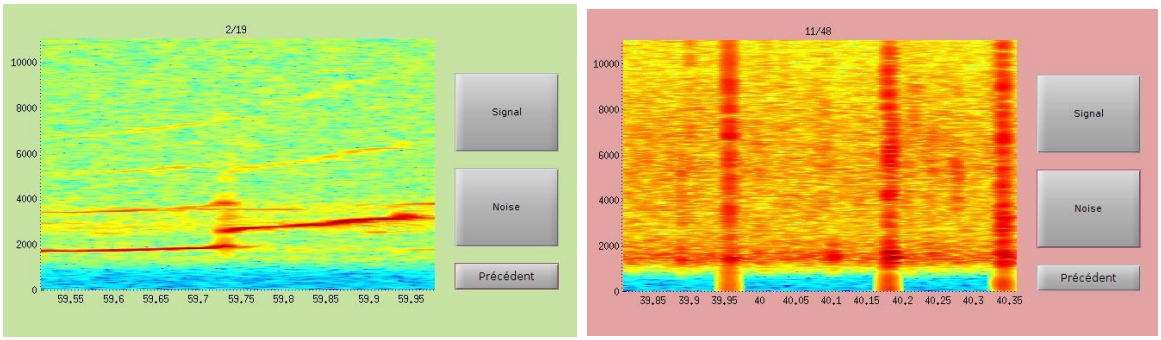
2.1.3 La prise de décision

Une fois les scores attribués à chaque pic détecté, l'algorithme va normaliser chaque score entre -1 et 1. La somme de chacun de ces scores donne un score global. Une valeur seuil modifiable (voir `Detection sensitivity` sur la figure 14) permet de moduler la sensibilité de la détection et de trancher définitivement sur la présence ou non d'une vocalisation. On constate en effet que pour un même enregistrement, si la sensibilité augmente, l'algorithme identifiera plus de pics d'amplitude comme étant des vocalisations et détectera de cette façon un plus grand nombre de cris (vrai positifs). En contrepartie, cette baisse de la sensibilité aura pour effet d'augmenter également le nombre de fausses détections (faux positifs). A l'inverse, une baisse de la sensibilité entraînera moins de détections, mais celles-ci seront plus fiables. Ceci permettra d'affirmer avec une plus grande certitude le fait que les signaux détectés sont réellement des grenouilles, en étant toutefois conscient que plusieurs n'ont pas été détectées.

2.2 Visalisation graphique et choix des paramètres

Au lancement de l'algorithme, il est proposé à l'utilisateur, via une interface graphique, de modifier un certain nombre de paramètres (voir figure 14), d'ouvrir un ou plusieurs fichier(s) (manuellement ou aléatoirement) et de choisir le mode. Il y en a trois :

- `Plain` : C'est le mode d'utilisation basique de l'algorithme. Une fois lancé, l'algorithme tourne et fournit le nombre de grenouille détectées dans le(s) fichier(s), sans demander quoi que ce soit à l'utilisateur.
  Ex de réponse:`18 frog calls detected on BOX03_20130520_021642.wav`
- `Liveview` : Permet de contrôler pas à pas les agissements de l'algorithme. Simultanément aux calculs, l'algorithme propose une fenêtre avec de nombreux graphiques pour que l'utilisateur puisse évaluer comment le programme se comporte suivant la portion de signal à analyser. C'est plus un mode de développement permettant d'être plus proche de ce que fait réellement le programme.
- `Results checker` : C'est un mode qui propose une suite au mode `Plain`. Après avoir effectué tout les calculs, le programme présente à l'utilisateur ses décisions et l'utilisateur peut alors fournir les résultats réels (que le programme aurait du trouver) à l'aide d'une interface graphique légère (voir figure 15). La décision de l'algorithme est présentée à l'aide d'un fond coloré (vert ou rouge). L'utilisateur peut alors indépendamment de la couleur de fond choisir si le sonagramme présenté (qui correspond à la fenêtre temporelle autour du pic détecté) est une grenouille `[Signal]` ou du bruit `[Noise]`. En comparant les résultats trouvé par l'algorithme à ceux rentrés par l'utilisateur, il est ensuite possible d'avoir accès au tableau des bonnes et des mauvaises détections et ainsi de pouvoir évaluer la qualité du programme (voir figure 16).



(a) Vrai positif **(b)** Vrai négatif

FIGURE 15. Vérification des détections obtenues par l'algorithme

3. Résultats et Discussion

3.1 Évaluation des résultats

Le mode `results checker` à été développé pour pouvoir évaluer les résultats du programme, ceci afin d'évaluer la fiabilité du programme. Finalement, le programme offre un pourcentage de bonnes détections variant entre 60% et 85%, dépendant des data examinées. Cette approximation est basée sur une comparaison entre les résultats obtenus par calcul (`Plain`) et par exploration visuelle des fichiers (`Checker`). Il existe aujourd'hui d'autres programmes qui ont le même but. On considère qu'ils deviennent intéressant au delà de 75 % de bonnes détections. Ce programme ne permet donc pas de fournir une estimation totalement fiable du nombre de vocalisations dans un enregistrement. D'autant plus que le pourcentage annoncé donne le pourcentage de bonnes détections parmi les pics d'amplitudes détectés. Cela signifie que lorsque le programme `Plain` annonce 85.3% de bonnes détections, le programme n'a naturellement pas pris en compte les pics d'amplitude inférieurs à un seuil et qui peuvent correspondre à des grenouilles. Il convient également de préciser que l'algorithme estime un nombre de cris et non pas un nombre d'individus. Or, un individu chante généralement par séquences d'une dizaine de cris.

Feuille2

	Positive	Negative	Total	
True	**18** *(52.9%)*	**11** *(32.4%)*	**29** *(85.3%)*	*Bonnes Détections*
False	**0** *(0%)*	**5** *(14.7%)*	**5** *(14.7%)*	*Mauvaises Détections*
Total	**18** *(52.9%)*	**16** *(47.1%)*	**34** *(100%)*	
	Détecté comme vrai	*Détecté comme faux*		

FIGURE 16. Exemple de tableau de détection de *martinicensis* établit à partir d'une minute d'enregistrement effectuée à 19h en Mai 2013 en forêt des Bains Jaunes.

Toutefois, **ce programme à l'avantage d'être robuste en évaluation relative.** C'est à dire que si le nombre de cris de grenouilles détecté par l'algorithme n'est pas fiable en lui-même, sa comparaison avec les résultats d'autres simulations respecte les fluctuations des densités de cris. L'objectif initial étant de pouvoir évaluer l'évolution des densités de populations, une réponse pourrait être apportée par cette évaluation relative.

3.2 Retour sur la méthode

Rétrospectivement, l'ensemble de la méthode est à revoir. L'utilisation de pics d'amplitude pour la segmentation est une technique simple mais elle fournit des résultats médiocres dès que l'enregistrement comprend des éléments non maîtrisés de forte amplitude, comme c'est souvent le cas dans la nature (vent, pluie, autres espèces...). D'autre part, la mise en place d'un certain nombre de constantes (seuils) contraint l'algorithme, lui faisant perdre de sa souplesse. C'est le cas notamment lors de la prise de décision finale (§2.1.3). Il existe plusieurs méthodes bien théorisées pour ce genre d'ajustement. Il serait alors intéressant de les utiliser et de pouvoir ainsi évaluer la pertinence des différentes méthodes de scoring.

Conclusion

Les algorithmes produits durant ce stage sont spécifiques à l'espèce étudiée. Le défi prochain serait d'enlever ou du moins de réduire cette spécificité et de pousser la fiabilité de détection au-delà des 85% quelque soient les conditions environnementales. L'intérêt des problématiques abordées au cours de ce stage est d'autant plus grand que celles-ci sont très actuelles et que de nombreux chercheurs étudiant la biodiversité se penchent actuellement sur ces questions de détection acoustique automatique de populations, d'espèces ou d'individus.

Références

Références des articles et ouvrages

[1] Jean-François Maillard. *Faune des Antilles : Guide des principales espèces soumises à réglementation*. Broché, 2009.

[2] D. Mennill L. Fitzsimmons, N. Barker. Individual variation and lek-based vocal distinctiveness in songs of the screaming piha (lipaugus vociferans),a suboscine songbird. *The Auk*, 125(4) :908–914, 2008.

[3] D. Mennill L. Fitzsimmons, N. Barker. Further analysis supports the conclusion that the songs of screaming pihas are individually distinctive and bear a lek signature. *The Auk*, 128(4) :790–792, 2011.

[4] Zhixin Chen and Robert C. Maher. Semi-automatic classification of bird vocalizations using spectral peak tracks. *J Acoust Soc Am*, 120(5 Pt 1) :2974–2984, Nov 2006.

[5] E. Maurin D. de Reyer JF. Sciabica, L. Daudet. Détection automatique des larves xylophages dans le bois. 10ème Congrès Français d'Acoustique, Avril 2010.

[6] Placer and Slobodchikoff. A fuzzy-neural system for identification of species-specific alarm calls of gunnison's prairie dogs. *Behav Processes*, 52(1) :1–9, Oct 2000.

[7] J. Placer and C. N. Slobodchikoff. A method for identifying sounds used in the classification of alarm calls. *Behav Processes*, 67(1) :87–98, Jul 2004.

[8] John Placer, C. N. Slobodchikoff, Jason Burns, Jeffrey Placer, and Ryan Middleton. Using self-organizing maps to recognize acoustic units associated with information content in animal vocalizations. *J Acoust Soc Am*, 119(5 Pt 1) :3140–3146, May 2006.

[9] C. N. Slobodchikoff and J. Placer. Acoustic structures in the alarm calls of gunnison's prairie dogs. *J Acoust Soc Am*, 119(5 Pt 1) :3153–3160, May 2006.

[10] Wollerman. Acoustic interference limits call detection in a neotropical frog hyla ebraccata. *Anim Behav*, 57(3) :529–536, Mar 1999.

[11] A. Bonneau D. Fohr Y. Laprie, S. Jarifi. Détection automatique de sons bien réalisés. LORIA CNRS.

Références des sites internet

[12] Matlab *http ://www.mathworks.fr*

[13] StackExchange *http ://www.stackexchange.com*

[14] Avisoft *http ://www.avisoft.com*

[15] IUCN *http ://www.iucnredlist.org/*

[16] Neotropical Birds *http ://neotropical.birds.cornell.edu/*

Documentaires

[17] Antonio Fischetti. Le chant du capitaine de la forêt. *http ://videotheque.cnrs.fr/doc=2846*, 2011.

[18] Antonio Fischetti. Bonjour les morses. *http ://videotheque.cnrs.fr/doc=2019*, 2009.

[19] Antonio Fischetti. Crocodile melody. *http ://videotheque.cnrs.fr/doc=4168*, 2013.

Conférences

- Dan Stowell. Machine listening for birds automatic recognition of bird species and beyond. *CNPS, Orsay*, 2014.
- Leonida Fusani. Proximate and ultimate factors behind the elaborate courtship of *Golden collared manakins*. *CNPS, Orsay*, 2014.