
RAPPORT DE STAGE DE M1

Etude de la détection de ton dans le bruit par le système auditif humain

Stage réalisé du 2 mai au 31 juillet 2024

Azal LE BAGOUSSE

Master 1 de Mécanique, spécialité Acoustique - Sorbonne Université (2023-2024)

Stage encadré par M Léo Varnet, chercheur CNRS

Laboratoire des Systèmes Perceptifs, École Normale Supérieure, PSL (75005)

Sujet de stage & compétences

Sujet de stage

Ce stage a comme sujet la détection de ton pur dans le bruit par le système auditif humain et les indices utilisés pour cette détection. Pour cette étude, la Théorie de la Détection du Signal (*TDS*) et les méthodes de corrélation inverse psychoacoustique (*revcor*) ont été utilisées. Dans l'idée de suivre l'un des travaux majeurs de la TDS et fondateur de la revcor, on réplique l'étude *Time and frequency analyses of auditory signal detection* de Albert Ahumada et al. (1975) [1] sur la détection de ton dans le bruit. 9 participants ont passé l'expérience décrite dans l'article de 1975 puis les données ont été collectées avec la toolbox Matlab fastACI (Osses&Varnet, 2021) ont été traitées et analysées. Ces résultats ont été comparés à ceux obtenus avec le modèle énergétique de Green [2], le tout mis en parallèle avec les résultats de l'étude de 1975. En bonus, puisque le temps le permettait, un nouveau modèle du système auditif a été implémenté et ses résultats ont été comparés aux résultats précédents.

Compétences et outils mis en oeuvre

- Analyse et synthèse de la bibliographie scientifique associée au thème du stage pour construire un socle de connaissances solide et valider les choix méthodologiques de l'étude : maîtrise de la théorie de la détection du signal et de la corrélation inverse psychoacoustique
- Identification des paramètres utiles à l'étude et révision de leurs valeurs par le calcul et la passation d'expériences pilotes
- Compréhension et révision de scripts informatiques dans l'environnement Matlab R2021b, à la base de la toolbox fastACI
- Mise en place d'une expérience psychoacoustique sur des sujets humains : organisation et gestion des créneaux de passages et de la récolte de données
- Rédaction de multiples programmes informatiques (Python 3) pour le traitement des données acquises durant l'expérience
- Réalisation de modèles computationnels auditifs (Python 3) pour analyser la détection des tons dans le bruit : modèles énergétiques de Green, modèle Modulation Filterbank (MFB)
- Rédaction d'un rapport d'étude et composition d'une présentation orale : synthèse du projet, analyse critique, référencement...
- Gestion du temps (cf. [Diagramme de Gantt](#)) : alterner entre les tâches, coordonner divers événements : expérience sur public, lecture des ouvrages didactiques, réunions de laboratoire et présentations, mise en place des modèles, rédaction du rapport...

Table des matières

1 Résumé	1
2 Présentation du laboratoire d'accueil	2
3 Théorie de la détection du signal	3
3.1 Statistiques	3
3.2 Corrélation inverse & ACIs	5
3.3 Indices pour la détection	8
4 RéPLICATION - Méthodes	9
4.1 Expérience	9
4.2 Modèles auditifs	10
4.2.1 Modèle énergétique de Green	11
4.2.2 Modèle Modulation Filterbank (MFB)	12
5 Résultats & discussions	13
5.1 Système auditif humain	13
5.1.1 Paramètres de performance	13
5.1.2 ACIs	14
5.2 Modèles auditifs	16
5.2.1 Modèles de Green	17
5.2.2 Modèle MFB	18
6 Conclusions et perspectives	20
ANNEXE	

Chapitre 1

Résumé

Comme l'on peut s'en rendre compte quotidiennement, la compréhension de la parole, et donc la tenue d'une conversation, est plus facile dans le silence qu'en présence d'un bruit de fond. Il est donc important de comprendre par quels mécanismes le bruit perturbe notre perception des sons simples, à la base de la parole. Dans ce contexte, les chercheurs utilisent la théorie de la détection du signal, qui permet ici de relier les données obtenues à partir de tâches de détection de tons dans des bruits par le système auditif humain à l'analyse des indices acoustiques du bruit utiles à leur détection. Ils utilisent aussi la *revcor*, la corrélation inverse psychoacoustique, qui met en évidence la relation entre le stimulus auditif et la détection d'une cible dans le bruit.

Durant ce stage, la réPLICATION de l'étude d'un article fondateur de la théorie de la détection du signal et de la corrélation inverse psychoacoustique est mise en oeuvre : l'article *Time and frequency analyses of auditory signal detection* de Albert Ahumada Jr. et al., publié en 1975 [1]. On utilise la toolbox *fastACI*, une toolbox Matlab basée sur la *revcor*, développée par le laboratoire d'accueil (Osses & Varnet 2021) lors de la réPLICATION afin de pouvoir comparer les résultats obtenus avec cet outil aux résultats de l'article de 1975 et rédiger par la suite un article scientifique qui valide le fonctionnement de cette toolbox. Une expérience est mise en place sur un échantillon de 9 individus jeunes (20-25 ans) et normo-entendants pour évaluer leurs critères de discrimination de présence d'un ton dans un bruit, une réPLICATION de l'expérience d'Ahumada décrite dans son article. 3200 bruits de 0.5s sont joués, la moitié avec un ton de 500Hz en leur centre, de 0.2s à 0.3s, et l'autre sans. Les participant doivent déterminer quels bruits joués contiennent un ton et quels bruits n'en contiennent pas. On traite les résultats pour caractériser les performances des participants et obtenir des **images de classification auditives** (dites **ACIs**, Auditory Classification Images) [3], des représentations temps-fréquence obtenues par la corrélation inverse et permettant d'explorer les représentations mentales d'un ton dans le bruit. Les résultats permettent de déterminer quelles caractéristiques acoustiques du bruit influencent la perception à l'aide de modèles statistiques implémentés numériquement (Python 3). Les modélisations sont ensuite comparées avec celles obtenues par différents modèles computationnels du système auditif. Toutes ces analyses permettent de mieux comprendre comment fonctionnent l'appareil auditif humain et sa capacité à différencier des stimuli auditifs, et comment répliquer ce fonctionnement.



Time and frequency analyses of auditory signal detection

Al Ahumada Jr.

School of Social Sciences, University of California, Irvine, California

Richard Marken and Arthur Sandusky

Department of Psychology, University of California, Santa Barbara, California

(Received 24 May 1974; revised 4 November 1974)

FIGURE 1.1 – En-tête de l'article répliqué

Chapitre 2

Présentation du laboratoire d'accueil

Le laboratoire d'accueil pour ce stage est le **Laboratoire des Systèmes Perceptifs (LSP)**. Il fait partie du **Département d'Etudes Cognitives (DEC)** au sein de l'Ecole Nationale Supérieure (ENS) de Paris, celle-ci appartenant à l'université Paris Sciences et Lettres (PSL).

Le DEC est un département interdisciplinaire de l'ENS, aux interfaces des sciences humaines et sociales, des sciences du vivant et des sciences de l'ingénieur. Mis en place en 2005, il est composé de 5 unités de recherche (dont le LSP), 1 unité de service et 3 équipes transversales.

Le LSP est une Unité Mixte de Recherche de l'ENS Paris et du CNRS (UMR 8248), née en 2014 de la scission du Laboratoire de Psychologie de la Perception de l'université Paris Descartes. Son objectif est de mieux comprendre les mécanismes sous-jacents à la perception du monde qui nous entoure par le biais de la vision et l'audition en utilisant des outils de psychophysique comportementale, de neurosciences intégratives et modélisation computationnelle [4]. Rassemblant 45 membres permanents, dont 14 doctorants, il se divise en deux équipes : l'équipe **Audition**, gérée par Daniel Pressnitzer, et l'équipe **Vision**, gérée par Peter Neri. Certains membres de l'équipe Audition, équipe d'accueil, font aussi partie d'un groupe de recherche alliant des membres de différents laboratoires : le **Modulation Group**, auquel appartient Léo Varnet, maître de stage. Les équipes Audition et Vision se rassemblent lors de réunions de laboratoire hebdomadaires et le Modulation Group prévoit aussi fréquemment des réunions de groupe : ces séances permettent de rester à jour sur les recherches en cours de chaque secteur et chaque membre du laboratoire. Ces réunions permettent d'élargir les centres d'intérêt et d'ouvrir de nouvelles perspectives d'études, et elles apportent une forte cohésion entre tous les chercheurs du laboratoire.

Le LSP comporte un centre de recherche principal, des bureaux au deuxième étage du 29 rue d'Ulm à Paris, mais aussi une plate-forme expérimentale au sous-sol du bâtiment, partagée avec le Laboratoire de Sciences Cognitives et Psycholinguistique (LSCP/UMR 8554, DEC). Cette plate-forme comporte des cabines *Audition* et *Vision* ainsi tout le matériel informatique nécessaire aux expériences psychophysiques, utilisables à tout moment. C'est dans l'une de ces cabines *Audition* que l'expérience à la base de cette étude prendra place.

Chapitre 3

Théorie de la détection du signal

Dans l'optique de comprendre les mécanismes sous-jacents à la détection humaine de signaux spécifiques dans un environnement perturbateur, les chercheurs vont utiliser la théorie de la détection du signal en l'appliquant à la psychophysique¹ pour identifier les indices acoustiques utiles pour la détection de tons simples, sons purs, dans le bruit.

La théorie de la détection du signal (TDS) est développée dans les années 1950 par des mathématiciens et des ingénieurs américains (Harvard, MIT...) en se basant sur des travaux en statistiques et en électronique. Elle peut s'appliquer au domaine visuel et au domaine auditif. Dans le cadre auditif, elle permet de mesurer les performances d'un individu, qui définissent sa stratégie de discrimination, lors de sa prise de décision sur la présence de signaux auditifs précis dans des signaux perturbateurs comme le bruit [2]. Elle fournit un cadre pour caractériser la perception auditive de l'individu. Dans ce chapitre, la théorie et les fondements des différents outils utilisés liés à la détection du signal seront introduits, pour l'étude de la perception des tons purs dans le bruit.

3.1 Statistiques

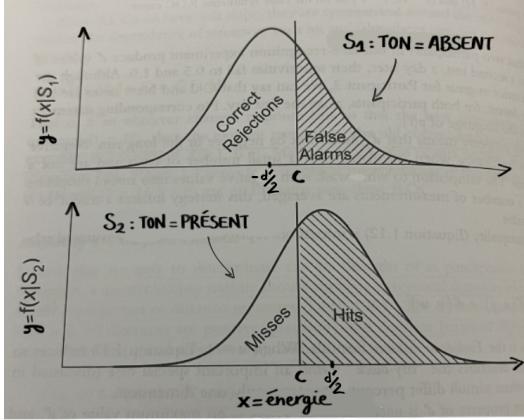
Les tâches de discrimination sont à la base des expériences en détection du signal : les participants doivent identifier un stimulus cible parmi des stimuli non-cible et ainsi le discriminer des autres stimuli qui lui sont présentés. Dans la présente étude, réPLICATION de l'étude de 1975 d'Ahumada, la tâche consiste à discriminer des bruits blancs contenant un ton en leur centre de ceux qui n'en contiennent pas (cf. [Expérience](#)). On se trouve dans une procédure Yes/No, la tâche de discrimination la plus simple, avec seulement 2 possibilités de réponse pour le participant : le ton est présent ou non. Les résultats obtenus après la réalisation d'une tâche Yes/No sont récapitulés en un unique tableau, à la base de la théorie de la détection du signal [2], qui présente 4 résultats possibles à chaque essai :

		SIGNAL	
		Présent	Absent
RESPONSE	Présent	Hit	False Alarm
	Absent	Miss	Correct Rejection

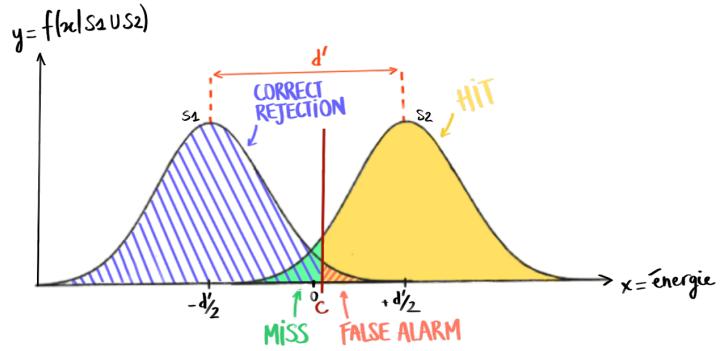
FIGURE 3.1 – Matrice selon la présence du signal (ton) et la réponse du participant

On peut statistiquement modéliser la variable de décision affiliée à cette tâche de discrimination en utilisant deux représentations gaussiennes correspondant aux densités de probabilité de l'énergie du signal. Ces distributions permettent ensuite de visualiser les probabilités d'un participant d'obtenir l'une des 4 possibilités [Fig.3.1] après un essai, soit qu'il perçoive ou non une variation d'énergie correspondant pour lui au ton dans le bruit, avec le ton présent ou non dans le stimulus. On appelle cette modélisation l'*espace de décision* du participant [5].

1. Psychophysique : Etude des rapports entre les phénomènes physiques (en particulier les stimuli nerveux) et les réactions, les sensations qu'ils provoquent. <https://www.cnrtl.fr/definition/psychophysique> (Fechner, 1878). → Psychoacoustique : application à l'audition, effet des stimuli auditifs sur la perception.



(a) Schéma biaxial [5, Fig.1.5 modif.]



(b) Schéma monoaxial

FIGURE 3.2 – Représentations d'un espace de décision arbitraire

La [Fig.3.2] est un exemple de représentation d'un espace de décision. Lorsque le ton est absent dans le signal perturbateur, l'énergie prend des valeurs appartenant à la gaussienne "S1", et lorsqu'il est présent, ses valeurs appartiennent à la gaussienne "S2". L'espace de décision est séparé en 4 par la valeur de seuil c . c et d' sont des paramètres de performance explicités dans la suite du rapport. Les 4 zones sur la [Fig.3.2](b) correspondent aux 4 possibilités du tableau des résultats [Fig.3.1] et leurs valeurs correspondent à l'énergie que doit atteindre le stimulus pour obtenir ces résultats.

On s'intéresse aux ratios de chacune des 4 possibilités (Hit **H**, Miss **M**, False Alarm **FA**, Correct Rejection **CR**) à chaque essai pour chaque participant afin de calculer différents paramètres statistiques propres à leur stratégie d'écoute. Ils se basent sur la présence du ton dans le stimulus émis et la réponse du participant à la tâche. En posant la variable "présent" = 1 et la variable "absent" = 0 :

$$\begin{array}{ll} \text{hit} : (\text{sig} = 1 \cup \text{rép} = 1) & \text{miss} : (\text{sig} = 1 \cup \text{rép} = 0) \\ \text{false alarm} : (\text{sig} = 0 \cup \text{rép} = 1) & \text{correct rejection} : (\text{sig} = 0 \cup \text{rép} = 0) \end{array} \quad (3.1)$$

On obtient alors les ratios en fonction des essais "signal présent" et des essais "signal absent" :

$$\boxed{\begin{array}{ll} \mathbf{H} = \frac{\sum \text{hit}}{\sum \text{hit} + \text{miss}} & \mathbf{M} = \frac{\sum \text{miss}}{\sum \text{hit} + \text{miss}} \\ \mathbf{FA} = \frac{\sum \text{false alarm}}{\sum \text{false alarm} + \text{correct rej.}} & \mathbf{CR} = \frac{\sum \text{correct rej.}}{\sum \text{false alarm} + \text{correct rej.}} \end{array}} \quad (3.2)$$

A partir de ces ratios, on calcule les paramètres propres à la détection individuelle du signal. Pour cela, il faut introduire $\Phi(t)$, la fonction de distribution cumulative normale standard ($\mu = 0, \sigma = 1$) et sa fonction inverse [5] :

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \xrightarrow{\sim} \Phi^{-1}(p) \quad (3.3)$$

Cette fonction inverse renvoie, pour une probabilité p donnée, la valeur $x = E$ telle qu'il y ait une probabilité p d'obtenir une énergie dans l'intervalle $[-\infty, E]$.

On s'intéresse alors aux paramètres de performance :

- le **critère de décision (c)** [Fig.3.2], seuil de séparation des gaussiennes de l'espace de décision du participant, qui lui permet de trancher sur la présence ou non du ton dans le bruit et

représente une mesure de biais² :

$$c = \Phi^{-1}(\mathbf{CR}) - \frac{d'}{2} \quad (3.4)$$

Avec d' la mesure de performance, décrite ci-dessous. $\Phi^{-1}(\mathbf{CR}) - \frac{d'}{2}$ renvoie le point en dessous duquel les valeurs appartiennent à l'aire CR sous la première gaussienne dans l'espace de décision [Fig.3.2]. Le biais est nul lorsque $\mathbf{FA} = \mathbf{M}$ et $\mathbf{CR} = \mathbf{H}$, caractéristique d'un profil impartial. Cela entraîne une valeur de critère nulle $c = 0$. Dans ce cas, le critère se trouve à l'intersection des deux gaussiennes de l'espace de décision du participant non biaisé.

- le **percent correct** ($\%_{cor}$), le pourcentage de réponses correctes du participant :

$$\%_{cor} = \frac{\sum \text{hit} + \sum \text{correct rej.}}{\text{nbr total d'essais}} * 100 \quad (3.5)$$

Cette mesure n'est que peu représentative de la sensibilité du participant puisqu'elle prend en compte le biais de ce dernier à travers un calcul dépendant des ratios qui dépendent directement du critère, mais reste indicative de sa performance générale.

- la **mesure de performance/sensibilité statistique (d')** [Fig.3.2], la distance entre les sommets des distributions gaussiennes de l'espace de détection du participant. Cette mesure représente la réelle mesure de sensibilité du participant puisqu'elle ne prend plus en compte le placement du critère dans l'espace de détection (inversement au $\%_{cor}$) en sommant les distances des sommets des 2 gaussiennes au critère. d' est donc la représentation de performance du participant la plus utilisée en détection du signal :

$$\begin{aligned} d' &= \Phi^{-1}(\mathbf{H}) - \Phi^{-1}(\mathbf{FA}) \\ \iff d' &= \Phi^{-1}(\mathbf{H}) + \Phi^{-1}(1 - \mathbf{FA}) \\ \iff d' &= \Phi^{-1}(\mathbf{H}) + \Phi^{-1}(\mathbf{CR}) \end{aligned} \quad (3.6)$$

3.2 Corrélation inverse & ACIs

La corrélation inverse (*revcor*) psychoacoustique est une approche apparue dans les années 1970 dans les travaux d'Albert Ahumada Jr. [6][7], chercheur à l'Université de Californie (USA). Cette méthode est fondée sur la mise en évidence d'une relation entre la structure spectro-temporelle du bruit de fond et la réponse du participant, détection ou non d'une cible sonore dans ce bruit. La *revcor* est un concept générique qui regroupe l'expérience psychophysique mise en place pour la collecte de données caractérisant la perception et les méthodes utilisées pour le traitement de ces données. Dans le cas de cette méthode appliquée à la psychoacoustique, on met en difficulté le système auditif en lui présentant des sons fortement bruités, et on examine les erreurs qu'il commet de façon systématique, qui sont le reflet des traitements qu'il effectue. On obtient suite à son utilisation des images de classification auditives (**ACIs**, Auditory Classification Images) [3]. Cette méthode de visualisation de la stratégie d'écoute d'un individu permet, dans le cas de ce projet, d'obtenir des "cartes" indiquant les régions du spectre sonore dans lesquelles la présence de bruit induit en erreur le système auditif sur la présence du ton dans les stimuli de bruits blancs. On a ici une représentation physique de la stratégie d'écoute de chaque participant sous forme d'une **ACI matricielle** :

2. Biais : tendance d'un participant à favoriser une certaine décision sur la présence du signal

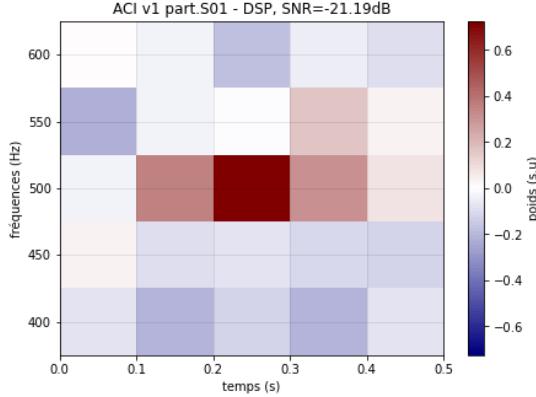


FIGURE 3.3 – ACI matricielle / S01

Cette image est une représentation spectro-temporelle de la stratégie d’écoute du participant S01 de l’étude résumée sous forme de matrice de pixels colorés. Conformément à l’étude originelle d’Ahumada, on choisit de se limiter à seulement 25 zones spectro-temporelles, avec un axe temporel de la longueur des stimuli de bruit et un axe fréquentiel centré autour de la fréquence de la cible à détecter dans le bruit (ici 500Hz). Pour l’étude, cette première représentation est utilisée puisque l’article répliqué présente une analyse équivalente sur 25 régions (cf. [ACI linéaire](#)), bien qu’aujourd’hui les ACIs sont plus précises grâce à une résolution bien plus élevée en pixels [Annexe Fig.7.1].

Les valeurs d’intensité des pixels des ACIs sont des valeurs sans unité que l’on appelle communément **poids perceptifs**, qui caractérisent l’importance de ces zones spectro-temporelles pour la détection d’un ton dans le bruit. Plus le poids est important en valeur absolue (pixel de couleur vive) plus la région est utilisée pour discriminer les deux réponses possibles. Un pixel rouge est en faveur du ton présent, un pixel bleu en faveur du ton absent. Plus le poids est faible en valeur absolue (pixel blanc) moins la région est utile à la détection. Les ACIs permettent ainsi d’analyser l’effet de l’amplitude spectrale et des modulations temporelles des stimuli sur la perception par système auditif à travers les poids perceptifs. On détermine alors la valeur de 2 indices énergétiques sur la détection d’un ton dans le bruit.

Une manière équivalente de représenter la stratégie du participant est l’**ACI linéaire** :

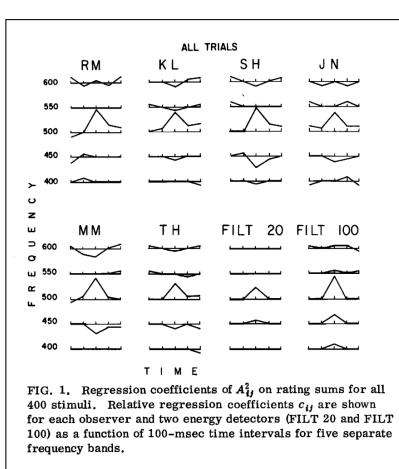


FIGURE 3.4 – Référence article : ACIs linéaires

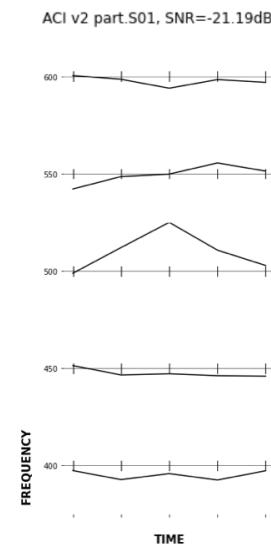


FIGURE 3.5 – RéPLICATION pour S01

Cette représentation celle est utilisée dans l’article d’Ahumada [1]. L’axe du temps et les unitées ne sont pas explicités pour suivre au mieux l’article. Cette ACI est strictement équivalente à l’ACI

matricielle : on peut observer la valeur des poids perceptifs sur les mêmes 25 régions spectro-temporelles. On voit sur [Fig.3.3] les mêmes tendances que sur [Fig.3.5], qui sont les 2 représentations de la même stratégie du participant S01 de l'étude. Lorsque le poids perceptif est important et que le pixel est rouge, le pic présent au niveau de la même région spectro-temporelle du deuxième modèle est plus prononcé vers le haut. Inversement, pour un pixel bleu de poids faible, le creux présent dans la même région du deuxième modèle est plus prononcé vers le bas. Pour les régions de poids neutre, les variations des courbes retrancrites sur les ACIs linéaires sont très faibles et les pics peu prononcés. Dans son article, Ahumada présente aussi une seconde version des ACIs linéaires en séparant les stimuli qui présentaient un ton (“SN”=signal+noise) de ceux qui n’en avaient pas (“N”=noise) durant son expérience. Ces représentations ont été calculées pour chaque participant de cette étude [Annexe Fig.7.3] mais ne sont pas discutées dans ce rapport au profit de l’analyse des figures principales.

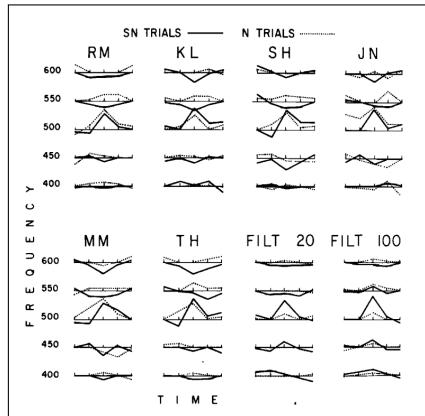


FIGURE 3.6 – Référence article : ACIs linéaires SN/N

De nombreuses méthodes mathématiques et numériques ont été développées pour le calcul d’une ACI (Murray, 2011) [8]. Dans cette étude, deux méthodes sont utilisées pour obtenir des ACIs de 25 régions spectro-temporelles : la Différence des moyennes et la Régression linéaire.

Différence moyennes

La méthode standard de calcul des ACIs.

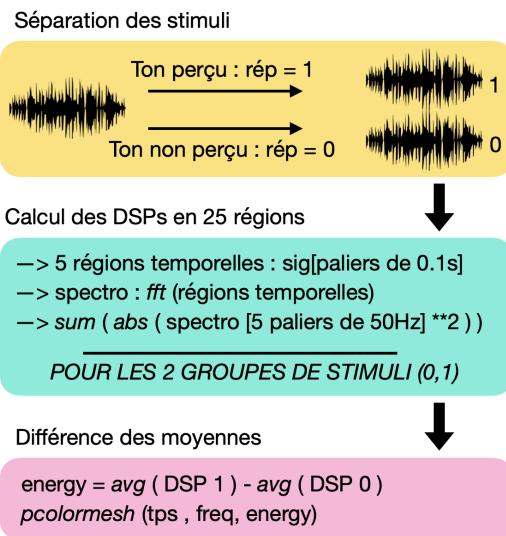


FIGURE 3.7 – Schéma méthode Diff.moy.

Dans le cas de cette étude, la deuxième méthode utilisée est en réalité la régression linéaire multiple, mais dans le cas où les prédicteurs (ici les stimuli) sont statistiquement indépendants elle se ramène

Régression linéaire

La méthode utilisée par A. Ahumada [1]

Calcul de la DSP en 25 régions

- > 5 régions temporelles : sig[paliers de 0.1s]
- > spectro : fft (régions temporelles)
- > sum (abs (spectro [5 paliers de 50Hz] **2))

POUR TOUS LES STIMULI

Calcul des coefficients d'énergie

```

X = DSP(stimuli) , y = réponses du participant
lin_reg = LinearRegression() / lin_reg.fit(X, y)
energy_coefs = lin_reg.coef
pcolormesh (tps , freq, energy_coefs)
  
```

FIGURE 3.8 – Schéma méthode Rég.Lin.

à l'équation de la régression linéaire simple.

Ces méthodes donnent des résultats identiques à un facteur négligeable près et peuvent être utilisées de manière interchangeable pour le calcul des ACIs. Elles font partie des diverses méthodes qui permettent de reproduire l'effet d'une corrélation entre les stimuli et les réponses des participants [8]. Les démonstrations des équivalences à un facteur près sont présentes en [annexe].

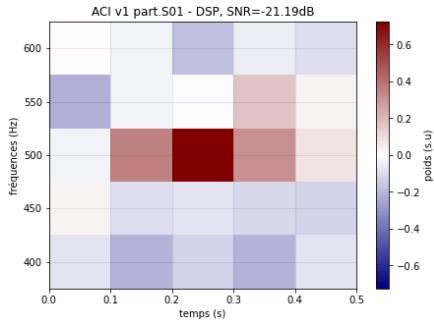


FIGURE 3.9 – ACI / Diff.moy.

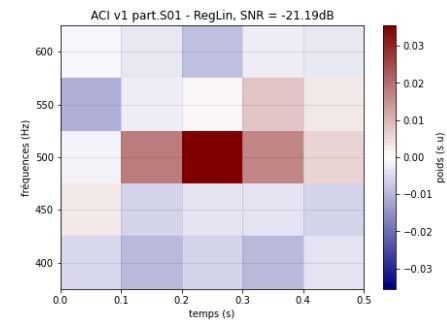


FIGURE 3.10 – ACI / Rég. lin.

Les ACIs obtenues avec ces méthodes vérifient donc bien une quasi-équivalence à un facteur près.

3.3 Indices pour la détection

Les indices étudiés sont les caractéristiques acoustiques du stimulus qui influencent la détection d'un ton présent ou non dans ce bruit. Dans les années 1960, l'étude de la détection d'un ton dans un bruit s'effectue à travers l'étude de l'effet des variations d'énergie du bruit sur la perception. En effet, les chercheurs supposent que lorsqu'un individu effectue une tâche de détection de ton dans le bruit, il utilise la quantité d'énergie dans certaines régions spectro-temporelles du stimulus pour décider si un ton y est présent ou absent. L'individu compare inconsciemment l'énergie perçue dans les différentes régions des stimuli à son critère de décision [3.4]. Il valide alors la présence du ton dans le bruit si l'énergie perçue dans la région où il s'attend à trouver le ton dépasse son critère. Inversement, si l'énergie perçue est plus faible que ce seuil, il estime le ton absent. L'énergie dans le bruit influence donc la décision du participant sur la présence du ton. Les chercheurs veulent alors étudier les indices énergétiques du bruit qui influencent la décision du participant : quelles caractéristiques amplifient le ton ou le masquent. Des modèles énergétiques du système auditif sont proposés (cf. [2] / [Modèle de Green](#)), mais leurs performances ne correspondent pas à celles des participants humains après une tâche de discrimination. La focalisation sur les indices énergétiques et le fonctionnement de ces modèles sont remis en cause dans les années 1970, entre autres par Ahumada [7][1] (cf. [Modèle de Green](#)). Par la suite, de nombreux psychoacousticiens se penchent sur l'étude des autres indices acoustiques qui influencent la détection du ton dans un bruit. V. Richards publie en 1991 un article sur la remise en question de l'effet des indices énergétiques à travers des expériences présentant des variations de niveau dans les stimuli [9]. Ces expériences prouvent que la présence de variations de niveau sonore, générées aléatoirement pour chaque stimulus, affectent aussi les décisions des participants. En effet, l'énergie varie dans ce cas à chaque essai mais les performances des participants demeurent stables et ne changent pas selon ces variations. Les stratégies des participants ne se basent donc pas seulement sur les indices énergétiques comme on le pensait précédemment. Au fur et à mesure des études, de nouveaux indices pour la détection émergent : les fluctuations rapides dans l'enveloppe des stimuli [10] ou encore les modulations dans l'enveloppe au niveau de l'onset et de l'offset du ton dans les stimuli. Ces indices ne seront cependant pas étudiés dans ce rapport qui reprend exclusivement la réPLICATION de l'article d'Ahumada (1975), soit l'étude des indices énergétiques seuls.

Chapitre 4

RéPLICATION - Méthodes

La réPLICATION de l'article au centre de cette étude [1] se divise en 2 parties : la mise en place d'une expérience de *revcor* sur des participants humains et le passage de cette même expérience par des modèles computationnels du système auditif humain (Modèle de Green, Modulation Filterbank model). Pour cette première partie de l'étude, l'expérience est générée par la toolbox *fastACI*, une toolbox Matlab développée par le Laboratoire des Systèmes Perceptifs (Osses & Varnet, 2021). Cette toolbox utilise les méthodes de *revcor* et permet de visualiser et caractériser les mécanismes auditifs impliqués dans la reconnaissance de sons spécifiques. La collecte des données se fait automatiquement avec cette même toolbox. Le traitement des données collectées est, quant à lui, effectué à travers différents programmes en Python conçus individuellement pour cette étude.

4.1 Expérience

Dans son article, Ahumada décrit l'organisation de l'expérience à répliquer : 6 membres de son département (Dpt of Psychology, UCSB) écoutent des stimuli de bruits blancs par un casque PDR-10 connecté à un magnétophone Sony TC 366. La moitié de ces stimuli joués contiennent un ton de 500Hz en leur centre et l'autre moitié non. Les participants doivent répondre "4" si le ton est assurément présent, "3" si le ton est probablement présent, "2" s'il est probablement absent, et "1" s'il est assurément absent. Pour le traitement de ces données, les réponses "1" et "2" sont regroupées (cas "ton absent"), ainsi que les réponses "3" et "4" (cas "ton présent"), revenant au cas d'une tâche de discrimination de type Yes/No. Certains détails de cette expérience sont explicités dans l'article :

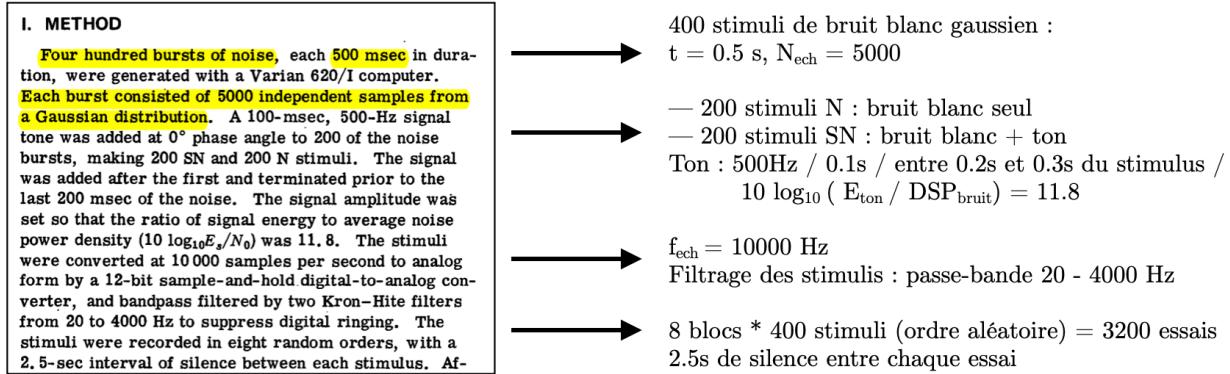


FIGURE 4.1 – Référence article : Méthode expérience

Cette expérience a pour objectif de caractériser les stratégies d'écoute et de discrimination auditive des 6 participants. A partir des données collectées, Ahumada et ses co-auteurs peuvent mesurer les paramètres de performance de ces individus et étudier les images de classification auditive de chacun. Dans le cas de la présente étude, cette expérience est répliquée en utilisant des techniques actuelles. Les réponses "1", "2", "3", "4" pour qualifier la présence du ton dans le bruit sont remplacées par "0" (ton absent) et "1" (ton présent). Les 400 stimuli (200 SN + 200 N) répétés dans 8 blocs d'ordre aléatoire sont remplacés par 3200 stimuli indépendants (1600 SN + 1600 N), regroupés en 8 blocs. En effet, le choix d'Ahumada de faire passer seulement 400 essais était purement pratique, lié aux contraintes techniques de l'époque. L'expérience passe entièrement par un ordinateur. Les filtres utilisés proviennent des librairies Matlab et Python. Le silence entre les essais n'est plus quantifié : le

prochain essai débute quand le participant donne sa réponse. On augmente le nombre de participants pour gagner en précision sur les comparaisons. L'expérience est chronophage et intense : 5h par participant, séparées en 8 blocs de 400 essais avec des pauses entre chaque bloc (au besoin des participants). L'expérience se fait en cabine Audition, sur la plate-forme expérimentale du laboratoire. Cette cabine insonorisée comporte un écran et un clavier sur lequel le participant peut renseigner sa réponse à chaque essai. L'écran est relié à un ordinateur externe à la cabine, sur lequel l'examinateur peut suivre les réponses du participant en temps réel. Pour chaque essai, l'expérience se déroule ainsi :



FIGURE 4.2 – Schéma de l'expérience répliquée

À ce stade, la seule variable non correctement définie est le Rapport Signal sur Bruit (RSB). Dans l'article d'Ahumada, le rapport est défini comme $10 \log_{10}(E_s/N_0) = 11.8$ (sans unité spécifiée) [1]. Ce rapport d'énergie du ton par rapport à la densité de puissance moyenne du bruit est fréquemment utilisé dans les travaux antérieurs sur la détection du signal [11][7]. L'article ne fournit pas de précisions supplémentaires sur le rapport utilisé ; ainsi, c'est à travers la documentation et de nombreux calculs que le RSB global équivalent est déterminé comme étant -21.19dB. Le principal défi de la réPLICATION est donc identifié : il faut palier l'absence de certaines informations méthodologiques sur la réalisation de l'expérience en 1975.

Les investigateurs de cette étude passent une première série d'essais et valident le fonctionnement de l'expérience. Les premiers passages avec des participants externes débutent. A ce stade, une difficulté est observée : les participants ne perçoivent jamais le ton dans le bruit, même lorsqu'il est présent. Malgré la mise en place d'une séance d'entraînement à l'expérience pour chaque participant, l'expérience est jugée trop difficile. Après passation d'expériences pilotes pour estimer un RSB qui permet d'obtenir le même $\%_{cor}$ que les participants d'Ahumada, le RSB est augmenté à -15dB. On a alors des données pour 2 groupes : 4 participants à -21.19dB et 9 participants à -15dB. Les 4 participants du premier groupe font aussi partie des 9 du second groupe.

Des programmes informatiques sont mis en place en parallèle pour visualiser les performances des participants. Un programme est écrit pour visualiser les paramètres de performance : $c, d, \%_{cor}$. Deux programmes sont écrits pour visualiser les ACIs (matricielles et linéaires) : un suivant la méthode Différence des moyennes et un suivant la méthode Régression linéaire. Pour les participants humains, on utilise le programme pour les ACIs suivant la première méthode afin de varier les approches.

4.2 Modèles auditifs

Dans la suite de son article, Ahumada mentionne les '*energy detector models*' [1]. Ce titre englobe les modèles computationnels du système auditif qui se basent sur les indices énergétiques pour détecter un signal dans un bruit. Le modèle énergétique le plus répandu est le modèle de Green, nommé d'après le chercheur David M. Green et explicité dans l'ouvrage *Signal detection theory and Psychophysics* (Green & Swets, 1966, ch.8) [2]. Ahumada mentionne dans son article 4 largeurs de bande de filtre et 3 temps d'intégration pour le modèle énergétique qu'il utilise : 12 modèles de Green présentant différentes combinaisons temps-fréquences sont alors mis en place dans cette étude. Ces modèles sont utilisés dans l'article original pour comparer leurs performances à celles des participants humains.

For each of the 400 stimuli, outputs of energy detectors of various bandwidths and integration times were computed to find which parameters of the energy detector model best predicted the observers' responses. Single-tuned digital filters with bandwidths of 20, 40, 100, and 250 Hz had their squared output summed over integration periods of 100, 300, and 500 msec. The integration periods were centered with respect to the signal interval.

- Trouver quel modèle énergétique (Green) est le meilleur pour prédire les réponses des participants
- Filtres passe-bande de largeur : 20/40/100/250 Hz
- Sorties au carré puis sommées
- Temps d'intégration : 100/300/500 msec autour du centre du stimulus

FIGURE 4.3 – Référence article : Méthode modèles de Green

Une fois la réPLICATION de l'article à la base de cette étude terminée, il était proposé d'implémenter un nouveau modèle du système auditif, plus récent que les modèles énergétiques. Le développement du Modulation Filterbank model (MFB), un modèle mis au point par le laboratoire (LSP), est alors initié pour ce projet. Le MFB prend en compte un nouvel indice pour la détection d'un ton dans le bruit : les modulations dans l'enveloppe des stimuli. Ainsi, on essaie de prédire les stratégies humaines d'écoute de manière plus précise qu'avec les modèles précédents. Cette partie additionnelle du projet permet de mieux comprendre la détection de ton dans le bruit par le système auditif et ses modèles.

4.2.1 Modèle énergétique de Green

L'article ne donne pas de précisions sur l'implémentation du modèle ni les filtres utilisés pour ce dernier. Pour cette étude, les choix effectués à ces niveaux sont arbitraires. Le modèle de Green est implémenté en Python. 12 modèles sont mis en place selon les combinaisons entre 4 largeurs de bande (20/40/100/250 Hz) centrées autour de 500 Hz, la fréquence du ton, et 3 temps d'intégration (100/300/500 msec) centrés autour de 250 msec, la moitié de la durée d'un stimulus.

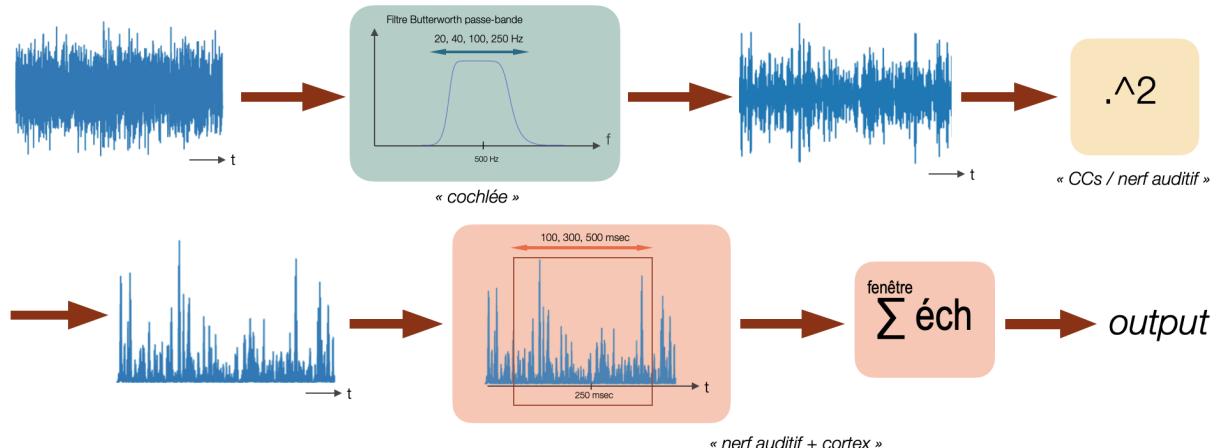


FIGURE 4.4 – Ossature d'un modèle de Green

Ce modèle reproduit le comportement de la cochlée, de la jonction entre les cellules ciliées et le nerf auditif, et du système auditif général. Il renvoie une valeur réelle pour chaque stimulus, simulant une "réponse" à chaque essai. L'expérience passée par les participants humains avec la toolbox *fastACI* est répliquée en Python afin de pouvoir la faire passer à ces 12 modèles. 3200 stimuli de bruits blancs, la moitié avec ton et l'autre sans, sont générés et passent par la structure du modèle pour donner les outputs. Le RSB est à 21.19dB pour répliquer exactement l'expérience de l'article. Avec le programme Python des ACIs par la méthode Régression linéaire, les outputs du modèle sont utilisés pour prédire les poids perceptifs de son image. Les paramètres de performances ne peuvent pas être définis dans ce cas car ils requièrent la mise en place d'un seuil de discrimination. Sans ce seuil, les outputs ne peuvent pas être regroupés en réponses "0" et "1" et les ratios ne peuvent pas être calculés. On utilise seulement la méthode Régression linéaire pour la même raison : la Différence des moyennes requiert des réponses binaires lors du traitement des données.

4.2.2 Modèle Modulation Filterbank (MFB)

Le MFB est un modèle du système auditif plus récent qui se base sur les caractéristiques énergétiques du stimulus mais aussi les modulations dans l'enveloppe de ce dernier pour détecter le ton. Mis en place par le laboratoire d'accueil de ce stage, il est inspiré par le travail de Torsten Dau (1997) et de Stefan Ewert (2000). Sa structure se présente ainsi :

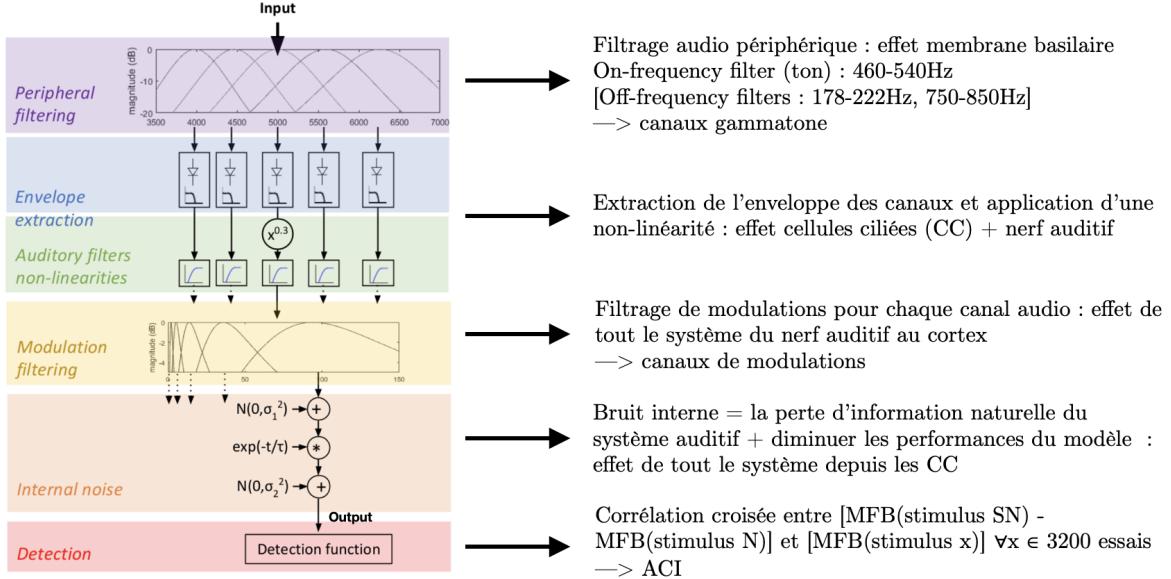


FIGURE 4.5 – Ossature du modèle MFB / depuis [12]

La structure du modèle est divisée en 5 parties, et une fonction de détection est mise en place pour traiter les outputs du modèle. On met en place 2 programmes Python : un pour l'ossature du MFB et un pour la fonction de décision. Le modèle prend en compte l'indice des modulations dans l'enveloppe autour du ton en instaurant une partie de filtrage dans les basses fréquences avec des filtres passe-bandes. Les matrices des signaux en sortie du modèle sont de taille ($L_{\text{gammatones}}$, $M_{\text{modulations}}$, $N_{\text{échantillons}}$). Dans le cas de cette étude, on utilise 3 canaux gammatone¹, avec 1 filtre on-frequency et 2 filtres off-frequency. Avec 10 canaux de modulations dont les filtres sont de fréquences centrales comprises entre 2Hz et 120Hz avec une largeur de bande logarithmique et les 5000 échantillons temporels, les matrices des signaux en sortie du modèle sont de taille (3,10,5000). Dans le cas de notre étude, on rajoute à cette structure une étape de *Lateral Inhibition Network* (LIN). Elle permet de rajouter un contraste qui est l'effet inhibiteur des cellules ciliées excitées sur les cellules ciliées adjacentes le long de la membrane basilaire. Dans ce cas, chaque canal gammatone reçoit une petite contribution négative des filtres adjacents. Le LIN peut se modéliser comme tel, avec $f \in [\text{filtres gammatones}]$ et $IR(t, f, mf)$ la représentation interne du modèle à ce stade : $IR_{\text{LIN}}(t, f, mf) = IR(t, f) - 0.3 * IR(t, f - 1) - 0.3 * IR(t, f + 1)$.

On fait passer l'expérience à RSB=-21.19dB au MFB pour obtenir 3200 outputs, des matrices de bruits filtrés. Pour le traitement des données et obtenir l'ACI du MFB, on met en place une fonction de détection. Celle-ci effectue la corrélation croisée entre un *template* de ton et les 3200 outputs individuellement. Le *template* de ton correspond à la matrice de la différence entre l'un des outputs du modèle pour un stimulus SN et l'un des outputs pour un stimulus N, ce qui donne une matrice représentant un ton. On a alors les “réponses” du modèle qui sont les valeurs des corrélations. En utilisant la même méthode de régression linéaire que pour les modèles de Green, on obtient les valeurs des poids perceptifs pour chaque zone spectro-temporelle souhaitée et donc une ACI correspondant à la simulation d'une stratégie d'écoute humaine selon le modèle MFB.

1. *Filtre gammatone* : filtre qui imite la réponse en fréquence des cellules ciliées dans la cochlée

Chapitre 5

Résultats & discussions

Après le passage de l'expérience par les participants humains et les modèles, les données sont traitées et les performances peuvent être analysées.

5.1 Système auditif humain

13 fichiers .mat sont obtenus à travers la toolbox Matlab après l'expérience sur participants. Ces fichiers sont traités comme indiqué dans le chapitre [RéPLICATION - Méthodes](#). Les résultats sont séparés en 2 groupes : un pour l'expérience à -15dB et un pour l'expérience à -21.19dB. La première partie de cette section consiste à discuter des paramètres de performance des participants et la seconde partie a comme sujet l'analyse des images de classification obtenues.

5.1.1 Paramètres de performance

Dans le cas de cette étude, il a été jugé pertinent de calculer le d' , le c et le $\%_{cor}$ des participants afin de visualiser leurs performances sous différents angles. Dans l'article de 1975, Ahumada mentionne seulement le d' de ses participants :

d'	
RM	2.36
KI	1.70
JN	1.86
SH	2.11
TH	1.71
MM	1.94

FIGURE 5.1 – Référence article : Tableau d'

Pour comparer les performances des participants de l'étude actuelle à celles des participants de 1975, le $\%_{cor}$ peut aussi être comparé malgré son absence dans l'article d'Ahumada. Dans ce cas, il faut utiliser les courbes du d' en fonction du $\%_{cor}$ dans le cas d'une tâche Yes/No, présentes dans le manuel *Detection Theory, A User's Guide* (2022, p.9-11)¹ [5]. En suivant ces courbes, le $\%_{cor}$ peut être approximé (à $c=0$) : $\%_{cor} \in [80; 88]\%$. Les paramètres sont alors calculés pour les participants de l'étude actuelle :

Observer :	SQ01	SQ02	SQ03	SQ04	SQ05	SQ06	SQ07	SQ08	SQ09	S01	S03	S05	S09
SNR (dB)	-15	-15	-15	-15	-15	-15	-15	-15	-15	-21.19	-21.19	-21.19	-21.19
Nb trials	3200	3200	3200	3200	3200	3200	3200	3200	3200	3200	3200	3200	3200
% correct	87,88	83,69	81,62	85,41	89,41	69,06	70,88	89,22	88,06	59,35	53,91	62,56	61,31
Hit rate	0,87	0,79	0,72	0,87	0,92	0,58	0,63	0,89	0,85	0,56	0,42	0,54	0,53
Miss rate	0,13	0,21	0,28	0,13	0,08	0,42	0,37	0,11	0,15	0,44	0,58	0,46	0,47
False Alarm rate	0,12	0,12	0,09	0,16	0,13	0,2	0,22	0,11	0,09	0,37	0,34	0,29	0,31
Correct Rejection rate	0,88	0,89	0,91	0,84	0,87	0,8	0,78	0,89	0,91	0,63	0,66	0,71	0,69
c	1,19	1,2	1,34	0,99	1,14	0,83	0,79	1,23	1,34	0,34	0,42	0,57	0,5
d'	2,34	2	1,93	2,11	2,51	1,04	1,13	2,48	2,38	0,48	0,21	0,66	0,59

FIGURE 5.2 – Paramètres des performances des participants (étude actuelle)

On nomme les 2 expériences selon leur RSB respectif pour fluidifier le rapport : “l'expérience -15dB” et “l'expérience -21.19dB”. Les valeurs diffèrent grandement entre les deux expériences. Comme prévu par les expériences pilotes lors de la recherche d'un RSB valable pour la réPLICATION, RSB=-15dB donne des valeurs de $\%_{cor}$ et de d' proches de celles des participants de 1975. La valeur moyenne est

1. Les auteurs présentent des courbes du d' selon divers critères dans le cas de différents paradigmes expérimentaux

de 82,8 % sur les 9 participants, avec la majorité des participants ayant un $\%_{cor} \in [80; 88]\%$. Pour le d', $d'_{moy1975} = 1.95$ et $d'_{moy2024} = 1.99$, avec $d'_{1975} \in [1.7; 2.36]$ et la plupart des d' de 2024 appartenant à cet écart. Les paramètres de performances à -15dB sont donc très similaires à ceux de 1975. Dans le cas de l'expérience -21.19dB, il y a un pourcentage de réponses correctes très faible pour chaque participant, sachant que $\%_{cor} = 50\%$ est le seuil de chance. Cela est expliqué par les difficultés de perception du ton dans le bruit à ce niveau faible de RSB. Ces pourcentages se sont améliorés avec les séances, passant d'un pourcentage de chance $\approx 50\%$ sur les premiers passages à un pourcentage de 10% plus élevé à la fin pour la majorité des participants. Cet effet est expliqué par la notion d'apprentissage : les participants parviennent à reconnaître le ton de manière plus efficace avec les essais malgré le caractère aléatoire de ces derniers. Par la lecture de la thèse d'Ahumada [7] ainsi que différents articles qui suivaient [9], on remarque que l'apprentissage prend une place conséquente dans ces expériences antérieures. On suppose par ces lectures que de multiples séances d'entraînement ont été mises en place pour les participants de l'étude de 1975. Ceci explique les valeurs des paramètres qui ne correspondent pas à celles du cas du RSB équivalent à celui de 1975, -21.19dB. De plus, seuls 400 stimuli différents étaient joués pour l'expérience d'Ahumada, dans un ordre aléatoire sur 8 blocs. Ainsi, l'hypothèse est que l'effet d'apprentissage au travers de nombreuses séances d'entraînement et des stimuli répétés permettent aux participants d'avoir de meilleures performances malgré un RSB faible. La difficulté de la réPLICATION provient à nouveau : aucun entraînement n'est mentionné dans l'article de 1975. Cette information cruciale à la réPLICATION est indisponible lors de la lecture et porte à confusion lors de l'analyse des résultats. Ainsi, l'expérience à -21.19dB est correcte mais les participants de l'étude d'Ahumada sont surentraînés par rapport aux participants "naïfs" de cette étude. Les valeurs faibles des performances au RSB correspondant à celui de 1975 sont donc expliquées.

5.1.2 ACIs

Les images de classification auditives sont générées pour chaque participant, pour les 2 groupes selon le RSB. Dans ce rapport, seules les ACIs des participants S01/SQ01, S03/SQ03, S05/SQ05 et S09/SQ09² sont étudiées : ces 4 participants ont passé les 2 expériences en parallèle et permettent une comparaison directe entre les images générées à -15dB et -21.19dB. Les ACIs des autres participants se trouvent en [annexe]. Pour cette étude, les représentations matricielles et linéaires sont utilisées.

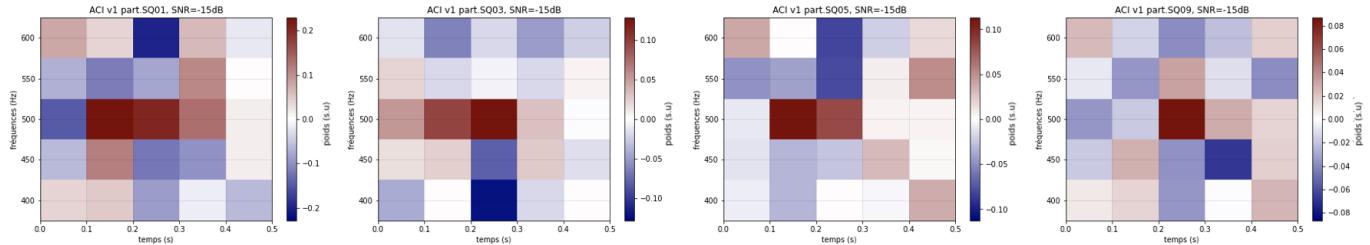


FIGURE 5.3 – ACIs matricielles / SQ01,SQ03,SQ05,SQ09 / RSB = -15dB

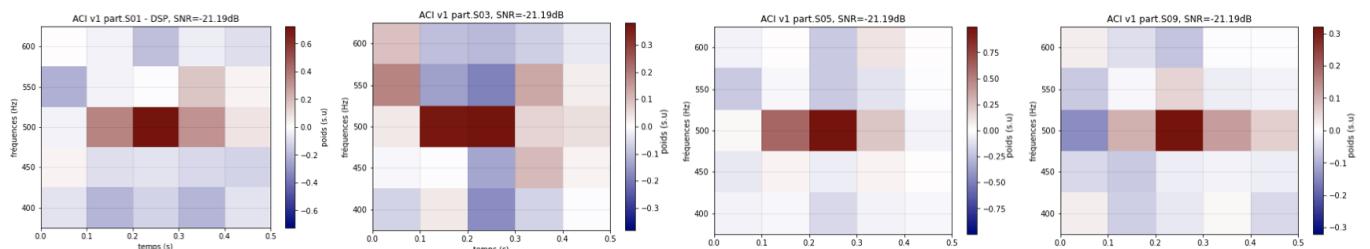


FIGURE 5.4 – ACIs matricielles / S01,S03,S05,S09 / RSB = -21.19dB

2. Ce sont les 4 mêmes participants : le préfixe "S" correspond à l'expérience à -21.19dB et "SQ" à -15dB

Les ACIs matricielles permettent une visualisation plus claire des stratégies de détection. L'élément le plus caractéristique de toutes les ACIs de cette étude est le pixel le plus rouge au centre, à 500Hz et entre 0.2 et 0.3s. Cette zone correspond à celle du ton, avec ses caractéristiques spectro-temporelles. Ce pixel présente toujours le poids perceptif le plus fort. Ainsi, la présence d'énergie dans le bruit au niveau de la région du ton suggère la réponse "ton présent" lors de l'écoute. Dans le cas où le ton est présent dans le stimulus, une forte énergie dans le bruit au niveau de cette région le met en exergue. Dans le cas où il est absent, l'énergie dans cette région induit le participant en erreur (cas "False Alarm"). Cette analyse est valable pour tous les participants de l'étude, pour les 2 RSB. Ainsi, la présence d'énergie dans certaines régions influence la décision du participant sur la présence du ton. En périphérie horizontale du pixel central, une à deux régions peuvent présenter des poids perceptifs élevés : la plupart des images comportent une barre de 3 pixels à 500Hz, entre 0.1 et 0.4s. Cette barre démontre une imprécision temporelle chez la plupart des participants. Ils se basent aussi sur la présence d'énergie dans le bruit juste avant et juste après le ton pour discriminer les stimuli. Le cas le plus commun est un poids perceptif plus élevé dans le premier des trois pixels que dans le dernier, ce qui décrit un effet d'anticipation du ton. Il y a aussi présence de multiple contrastes entre les régions, les plus importants étant ceux entre les pixels centraux et les pixels en périphérie verticale. Les pixels au dessus et au dessous du pixel central suggèrent une réponse "ton absent" au participant. C'est le phénomène de *masking* (masquage). Lorsque le ton est présent, si l'énergie du bruit blanc est plus forte dans les régions inférieures ou supérieures en fréquence, le ton est masqué pour le participant (cas "Miss"). Le système auditif humain peut donc détecter le ton à l'aide de ces contrastes. S09/SQ09 est un cas à part : sa stratégie diffère de celle des 9 participants. En effet, il est plus précis en temps : sa barre centrale est réduite au pixel central écrasant. De plus, il est le seul à présenter une imprécision en fréquence : il utilise aussi les caractéristiques du bruit à 550Hz pour discriminer les stimuli. Il serait possible de pousser l'analyse de ces différences sachant que ce participant est musicien (rythmique), tandis que les autres participants ne le sont pas.

Les ACIs linéaires présentent les mêmes tendances et permettent une comparaison directe avec les résultats d'Ahumada.

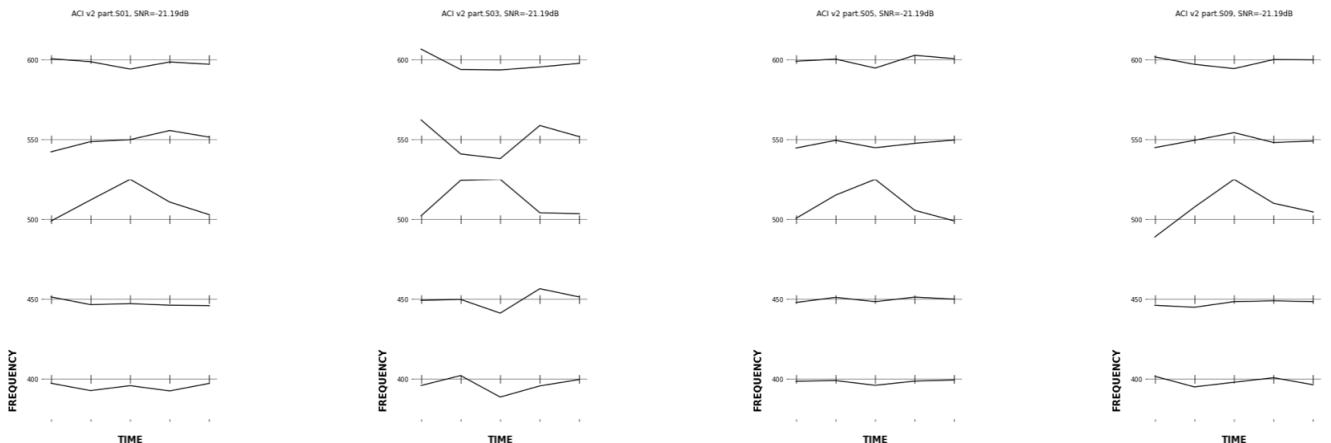


FIGURE 5.5 – ACIs linéaires / S01, S03, S05, S09 / RSB = -21.19dB

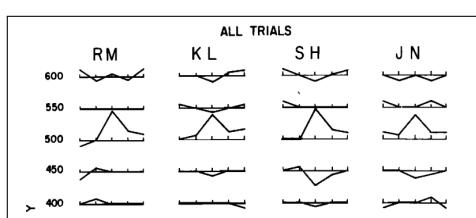


FIGURE 5.6 – Référence article (-21.19dB) : 4 ACIs linéaires

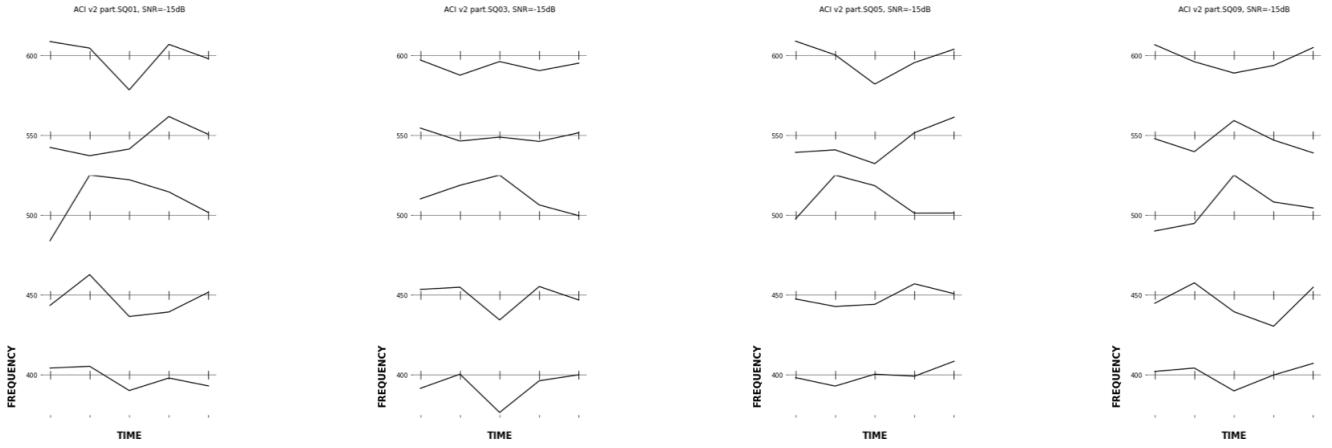


FIGURE 5.7 – ACIs linéaires / SQ01, SQ03, SQ05, SQ09 / RSB = -15dB

Les images linéaires sont équivalentes aux ACIs matricielles. Les mêmes analyses des ACIs peuvent être données. Entre les images de l'expérience -15dB et celle à -21.19dB, une différence est observée : les ACIs à RSB=-15dB sont beaucoup plus bruitées que les ACIs à -21.19dB. Les ACIs à -21.19dB présentent un pic central de manière évidente, avec des variations moins marquées sur la périphérie lointaine, peu utile à la détection. Ceci signifie que la stratégie des participants est précise à -21.19dB. Pour l'expérience à -15dB, l'ACI est moins ciblée et la stratégie plus aléatoire. Cela s'explique par la “facilité” de l'expérience : les participants obtiennent des très bonnes performances sans entraînement, leur stratégie n'est pas aussi spécifique. Ils entendent “trop” bien le ton lorsqu'il est présent, sans avoir à se baser autant sur les caractéristiques du bruit pour discriminer les stimuli. De plus, à -21.19dB, le stimulus contient plus de bruit donc les réponses du participant dépendent plus fortement du bruit. Dans ce cas, la méthode ACI est plus précise, tandis qu'à -15dB il y a plus d'incertitude dans l'estimation des poids. Après comparaison entre les ACIs de l'article [Fig.5.6] et les ACIs de cette étude [Figs.5.5,5.7], les ACIs à -21.19dB se rapprochent aussi plus des figures d'Ahumada. Ces ACIs présentent un pic central net qui s'étend sur 3 régions temporelles à 500Hz pour la plupart des participants. La notion d'imprécision temporelle est rappelée. Les contrastes sur la périphérie lointaine de la région centrale sont faibles. La périphérie lointaine des zones centrales n'influence pas le jugement. A l'inverse, on observe chez les participants d'Ahumada des variations de poids négatifs (creux) au niveau des zones supérieures et inférieures en fréquence (450/550 Hz), comme observé sur les ACIs matricielles. Dans le cas de l'expérience à -15dB [Fig.5.7], les images sont beaucoup plus disparates. Ces résultats appuient l'hypothèse du sur-entraînement des participants de 1975 : les performances à -15dB correspondent à celles des participants de l'article, mais les ACIs à -15dB ne corroborent pas celles d'Ahumada. L'expérience à -21.19dB est la bonne réplique.

Pour expliquer les contrastes visualisés sur les images, la notion de **STRF** est introduite. Les Spectro-Temporal Receptive Fields correspondent aux régions excitatrices et inhibitrices des neurones du système auditif. Ces régions sont à l'échelle du neurone ce que sont les régions spectro-temporelles des ACIs à l'échelle de l'humain. Les neurones auditifs permettent au système auditif humain de faire la différence entre la zone centrale et les zones périphériques et instaure les contrastes observés sur les ACIs.

5.2 Modèles auditifs

Dans le cas des modèles, les paramètres de performance ne peuvent pas être étudiés (cf. Section [4.2.1]). Les ACIs obtenues pour les modèles sont présentées et discutées.

5.2.1 Modèles de Green

Pour les modèles de Green, l'article ne présente que 2 modèles sur les 12 : le *FILT20* et le *FILT100* [1]. Ils correspondent aux modèles temps d'intégration (*IT*) = 100msec, pour des bandes passantes (*BW*) de 20Hz et 100Hz respectivement. Seules ces informations sont données dans l'article. L'étude est répliquée avec le RSB originel : -21.19dB. Dans ce cas, 3 modèles sont présentés : le $BW=20\text{Hz}/IT=100\text{msec}$, le $BW=100\text{Hz}/IT=100\text{msec}$ et le $BW=40\text{Hz}/IT=300\text{msec}$. 2 autres modèles pertinents se trouvent en [annexe].

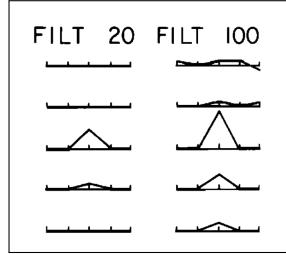


FIGURE 5.8 – Référence article : les 2 modèles énergétiques

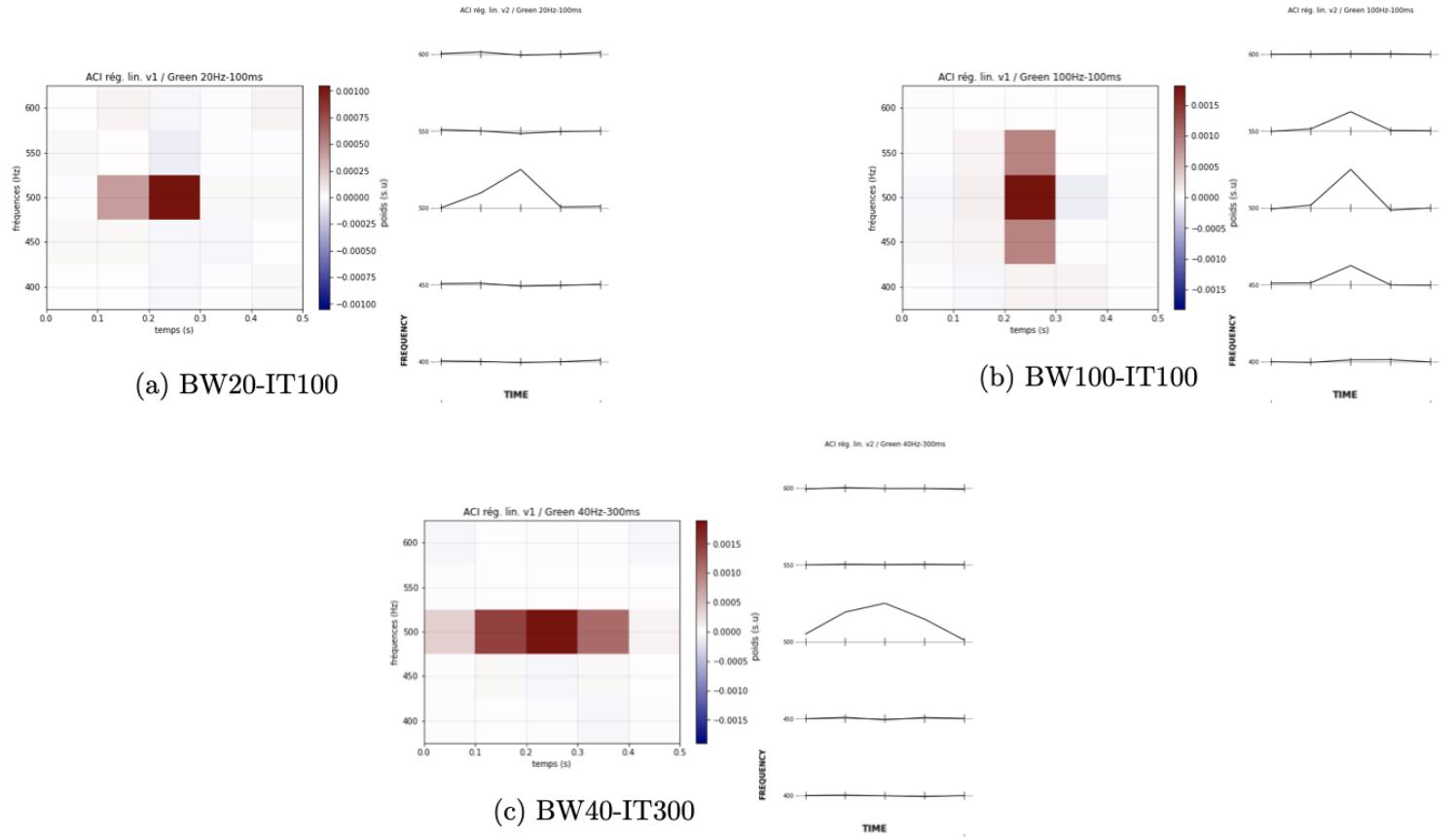


FIGURE 5.9 – 3 modèles de Green répliqués (RSB = 21.19dB)

Les ACIs linéaires des 2 premiers modèles en [Fig.5.9] sont comparées avec celles de l'article [Fig.5.8]. Les mêmes phénomènes sont observés : le pic principal est au centre, en région du ton, et le deuxième modèle [Fig.5.9](b) comporte aussi des pics aux fréquences 450Hz et 550Hz. Le modèle [Fig.5.9](a) est le plus étroit, en fréquence et en temps. Il ne laisse passer l'énergie que sur une bande de 20Hz autour de 500Hz, avec une fenêtre en temps de 0.1s au centre du stimulus. Seule l'énergie dans la région du ton est captée par ce modèle. Le pixel à gauche du pixel central sur l'ACI matricielle est

aussi légèrement rouge (poids de 0.00025) : c'est un effet de débordement, ici temporel, qui dépend du filtrage et du fenêtrage. Cet effet est aussi présent sur le FILT100 en [Fig.5.8], cette fois en fréquence. Hormis cet effet, l'ACI du modèle répliqué pour le FILT20 est équivalente à celle de l'article. L'image du modèle [Fig.5.9](b) est aussi équivalente à celle du FILT100 sans compter l'effet de débordement. L'image (b) est importante au niveau de sa périphérie : une barre de 3 pixels verticaux est présente entre 0.2 et 0.3s. Cette barre correspond à la plage de fréquences plus large permise par la bande passante de 100Hz pour ce modèle. Il n'y a donc pas de phénomène de contraste instauré par le modèle de Green, qui laisse passer toute l'énergie selon ses limites spectro-temporelles. Le dernier modèle, [Fig.5.9](c), est présenté car il est celui qui s'approche le plus du système auditif selon les ACIs obtenues. En effet, l'ACI (c) est celle qui ressemble le plus aux ACIs obtenues pour les participants de cette étude. Cependant, malgré la barre centrale de 3 pixels de poids positifs, il manque tout de même les contrastes en périphérie verticale. Sur les 3 modèles, on remarque malgré tout l'effet du pixel central sur la détection : le poids perceptif dans cette région est le plus fort, comme chez les participants humains.

Deux conclusions peuvent être formulées. En premier temps, le modèle de Green permet de reproduire une partie des résultats obtenus, notamment avec le pixel central. Cela signifie que l'énergie, le seul indice pris en compte par le modèle, a bien un rôle dans la discrimination des stimuli. Cependant, le modèle de Green ne reproduit pas tous les phénomènes observés sur les ACIs des participants. Les contrastes instaurés par le système auditif ne peuvent pas être retracés si l'on se base seulement sur les indices énergétiques. Ces indices ne sont donc pas les seuls à être pris en compte par le système auditif, inversement à ce qui était pensé dans les années 1960. Les régions négatives de l'ACI humaine suggèrent donc l'usage d'un ou plusieurs mécanismes supplémentaires par le système auditif humain pour la détection du ton.

5.2.2 Modèle MFB

Seule l'ACI matricielle est présentée puisque cette section sort du cas de la réPLICATION. L'expérience est passée par le modèle avec le RSB=−21.19dB.

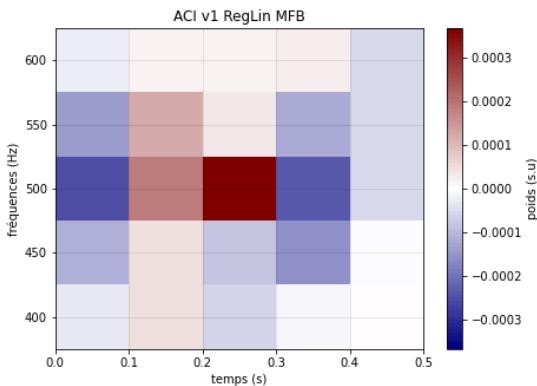


FIGURE 5.10 – ACI modèle MFB (RSB = 21.19dB)

Cette ACI a été obtenue après plusieurs tentatives empiriques d'ajustement des paramètres du modèle, pour finalement revenir aux valeurs par défaut. Elle correspond à l'image la plus proche des ACIs des participants de l'étude. Le pixel central est le plus important, à travers l'utilisation des indices énergétiques par le MFB. La zone à 500Hz entre 0.1 et 0.2 présente aussi un poids positif, retrançrant l'effet d'anticipation du ton. Les régions supérieures et inférieures à cette première zone sont aussi légèrement marquées, avec des poids de 0.00005 et 0.0001, ce qui traduit une imprécision fréquentielle croisée avec une imprécision temporelle, phénomène qui se retrouve chez certains participants. Cela signifie que ces participants n'ont pas une connaissance parfaite de la

position spectro-temporelle de la cible, si bien que le bruit dans des régions spectro-temporelles proches interfère aussi avec la décision. Le point principal de la figure [5.10] est les contrastes obtenus, notamment en périphérie verticale du pixel central. Les régions inhibitrices qui manquaient aux modèles de Green sont présentes avec le MFB. Ces zones apparaissent verticalement grâce à l'implémentation du LIN. On modélise simplement le fait que les cellules ciliées interagissent négativement avec les cellules ciliées adjacentes sur la membrane basilaire. Ce phénomène n'est pas retranscrit pour les modèles de Green.

Certains aspects de l'ACI [Fig.5.10] l'éloignent des ACIs humaines. La barre centrale horizontale n'est pas totale : le pixel à droite du pixel central représente un poids perceptif négatif. Pour ce modèle, l'énergie dans le bruit présente à 500Hz entre 0.3 et 0.4s n'est pas utile à la discrimination des stimuli. Ce résultat est contraire aux stratégies de certains participants, mais pas tous. Les régions inhibitrices horizontales sont expliquées par l'incertitude de phase et la réponse des filtres utilisés pour les canaux gammatoones et de modulations. Les filtres peuvent amener un déphasage qui empêche le signal d'être capté à 500Hz à certains moments du stimulus. Ils peuvent aussi causer une réponse en fréquence trop brutale ou trop précise. Pour se rapprocher au plus de l'audition humaine, de multiples techniques ont été utilisées : utilisation de filtres RII (Butterworth, Chebyshev, Elliptique, Bessel) puis de RIF (Hamming, Kaiser, FIR), implémentation d'un filtrage à phase nulle, évolution du coefficient dans le LIN et pour le bruit interne... Après comparaison de toutes les ACIs obtenues, le résultat le plus proche des images des participants était la [Fig.5.10]. Celle-ci est obtenue avec une configuration suivante : le filtrage est de type Butterworth, passe-bande, causal, le bruit blanc a un écart-type de 0.1 et le LIN dépend d'un coefficient 0.3 ($IR_{LIN}(t, f, mf) = IR(t, f) - 0.3 * IR(t, f - 1) - 0.3 * IR(t, f + 1)$). Le choix de ces paramètres permet un *ringing*³ mesuré. Si le *ringing* est trop présent, le contraste est trop important et seule l'énergie dans la zone du ton est utilisée par le modèle. S'il est trop peu présent, l'ACI ressemble aux ACIs des modèles de Green : aucun contraste n'est observé. Ainsi, un compromis est nécessaire et dépend de la configuration choisie du modèle.

Les différences superficielles observées entre les ACIs humaines et simulées par le modèle MFB ne doivent pas détourner l'attention de l'objectif premier de ces simulations. Le MFB est un modèle fonctionnel qui ne peut pas exactement reproduire les données humaines. Le point principal est qu'il conserve la structure générale des résultats, ce qui est le cas ici. Les contrastes verticaux et horizontaux sont reproduits pour la plupart, ce qui est le plus important dans le cas de la modélisation du système auditif. La réPLICATION exacte des données humaines nécessiterait un modèle phénoménologique ou biomécanique du système auditif humain, plus compliqués à mettre en place. Le MFB est donc un modèle optimal du système auditif à son niveau.

3. Ringing = phénomène oscillatoire (ici parasite). Référence aux oscillations indésirables qui apparaissent dans un signal après un filtrage, dues à la réponse impulsionnelle du filtre.

Chapitre 6

Conclusions et perspectives

Plusieurs objectifs ont été atteints à travers ce projet : comprendre les fondamentaux de la TDS, analyser le comportement du système auditif humain lors d'une tâche de discrimination et apprendre à répliquer une étude. L'utilisation de la toolbox *fastACI* a été maîtrisée, et les processus de conception et d'organisation d'une expérience sur des participants ont été assimilés. La compréhension de la méthode de corrélation inverse psychoacoustique a été approfondie. Cette étude a permis d'apprendre à évaluer les performances des participants humains ainsi que des modèles dans des tâches de discrimination, dans un contexte de la détection de ton dans le bruit. Une meilleure compréhension du fonctionnement du système auditif a été développée, de même pour les indices utilisés pour la détection des tons. Un contraste a été apporté entre les indices énergétiques et les modulations d'enveloppe pour la discrimination des stimuli.

Cette recherche ouvre la voie à des investigations sur la discrimination des différents phonèmes¹ dans le bruit ([13],[3]). Cela pourrait contribuer à une meilleure compréhension de la perception du discours et de ses mécanismes sous-jacents. Une meilleure maîtrise des processus impliqués dans la détection des tons dans des environnements bruyants peut éclairer les enjeux de la perception du langage.

Par ailleurs, les défis de la réPLICATION ont fréquemment été relevés tout au long de cette étude. L'expérience de l'article, qui utilisait des techniques obsolètes, se révélait particulièrement chronophage, répétitive et intense. Reproduire cette expérience avec un plus grand nombre de participants a ajouté aux difficultés. De plus, le manque d'informations dans l'article source a nécessité l'estimation de nombreux paramètres par le biais d'expériences pilotes et de programmes informatiques. Cela a impliqué de nombreux tests longs, parfois peu concluants. Cependant, passer par cette étape de réPLICATION est chose essentielle avant de commencer à publier et partager ses travaux. Cela permet de mettre en évidence les défauts de certains articles et de comprendre sur quels aspects il est nécessaire de se concentrer lors de la rédaction. Ce processus est particulièrement important dans le contexte de la crise de réPLICABILITÉ en sciences cognitives, où de nombreuses études antérieures manquent d'informations pour permettre à des équipes indépendantes de reproduire les résultats.

Enfin, ce stage a été une expérience riche pour le développement de compétences en traitement du signal, informatique et psychoacoustique.

1. Dans la linguistique, un phonème est un élément sonore du langage parlé, il permet de séparer et différencier les mots les uns des autres (<https://www.fondationpourlaudition.org/le-phoneme-720>)

ANNEXE

Time and frequency analyses of auditory signal detection

Al Ahumada Jr.

School of Social Sciences, University of California, Irvine, California

Richard Marken and Arthur Sandusky

Department of Psychology, University of California, Santa Barbara, California

(Received 24 May 1974; revised 4 November 1974)

Observers rated 500-msec bursts of wide-band Gaussian noise for presence or absence of a 100-msec, 500-Hz signal tone. The tone was present on half of the 400 trials and was centered in the noise bursts. Bandwidth and integration time estimates were found for each observer by correlating the observer ratings with the output of energy detectors of various bandwidths (20, 40, 100, 250 Hz) and integration times (50, 100, 300, 500 msec) to find the best correlating energy detector. The results, computed separately for signal (SN) and no signal (N) trials, indicate a bandwidth of 40 Hz and an integration time of 100 or 300 msec fit best on both SN and N trials. Energy variations in 25 50-Hz by 100-msec segments of the noise were correlated with the observer ratings and showed negative correlations in frequency and time intervals immediately surrounding the signal region on SN trials, but not on N trials. The results suggest observer monitor detectors sensitive to temporal and spectral changes in the energy of the noise bursts, not just the absolute level of the output of a simple filter-integrator energy detector. The N trial results are accounted for by assuming that the observer is looking for the same pattern of temporal and spectral changes as on SN trials, but is uncertain as to the exact location of the signal tone.

Subject Classification: 65.58, 65.35, 65.75.

FIGURE 6.1 – Introduction de l'article répliqué

- ACI moderne (2800 pixels) de la stratégie du participant S01. L'axe des fréquences est logarithmique pour imiter l'effet de la cochlée. Image générée avec la toolbox fastACI.

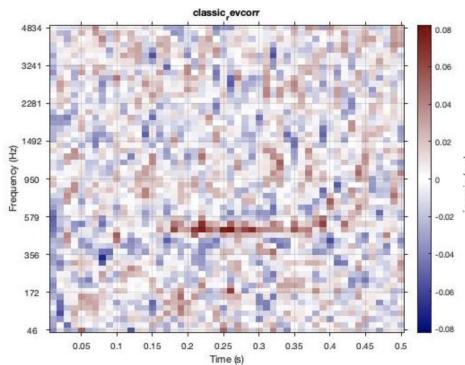


FIGURE 6.2 – ACI S01 - 2800 pixels

- Démonstrations des relations corrélation - Diff. des moyennes - Régression linéaire [8] :

Avec \mathbf{n} le champ des stimuli de bruit (avec ou sans ton), matrice de taille (3200,25), et r le vecteur des réponses du participant à la tâche de discrimination, on a :

$$\text{corr}[\mathbf{n}, r] = \frac{E[(\mathbf{n} - E[\mathbf{n}])(r - E[r])]}}{\sigma_{\mathbf{n}}\sigma_r} \quad (6.1)$$

Avec un bruit à moyenne nulle ($E[\mathbf{n}] = 0$, cas d'un bruit blanc gaussien) et un observateur non biaisé ($E[r] = 0.5$ car un observateur non biaisé donne les réponses 0 et 1 avec la même probabilité) :

$$\begin{aligned} \text{corr}[\mathbf{n}, r] &= \frac{E[\mathbf{n}(r - 0.5)]}{\sigma_{\mathbf{n}}\sigma_r} = \frac{E[\mathbf{n}(r - 0.5)|r = 1]P(r = 1) + E[\mathbf{n}(r - 0.5)|r = 0]P(r = 0)}{\sigma_{\mathbf{n}}\sigma_r} \\ &= \frac{E[\mathbf{n}|r = 1] - E[\mathbf{n}|r = 0]}{4\sigma_{\mathbf{n}}\sigma_r} \xrightarrow{*4\sigma_{\mathbf{n}}\sigma_r} E[\mathbf{n}|r = 1] - E[\mathbf{n}|r = 0] \end{aligned} \quad (6.2)$$

Il y a équivalence Corrélation - Diff. des moyennes à un facteur près.

La régression linéaire simple suit un modèle sous la forme $r = \sum_i a_i \mathbf{n}_i + b + \epsilon$, avec ici a_i les coefficients (poids perceptifs), \mathbf{n}_i les prédicteurs (stimuli), b le biais, r les réponses du participant et ϵ l'erreur. En implémentant ce modèle (considérant $\epsilon = 0$) :

$$\begin{aligned} \text{corr}[\mathbf{n}, r] &= \frac{E[(\mathbf{n} - E[\mathbf{n}])(a\mathbf{n} + b) - (aE[\mathbf{n}] + b)]}{\sigma_{\mathbf{n}}\sigma_r} = \frac{E[\mathbf{n} \cdot a\mathbf{n}]}{\sigma_{\mathbf{n}}\sigma_r} \\ &= \frac{aE[\mathbf{n}^2]}{\sigma_{\mathbf{n}}\sigma_r} = \frac{a\sigma_{\mathbf{n}}}{\sigma_r} \xrightarrow{\infty} C \end{aligned} \quad (6.3)$$

Il y a équivalence Corrélation - Régression linéaire à un facteur près.

- ACIs linéaires décomposées selon le type de stimuli (avec ou sans ton). Les 2 courbes SN-N sont moyennées pour donner l'image générale de l'ACI linéaire (tous essais) [Fig.3.5]. Ici seulement les ACIs SN/N de 4 participants : S01, S03, S05, S09.

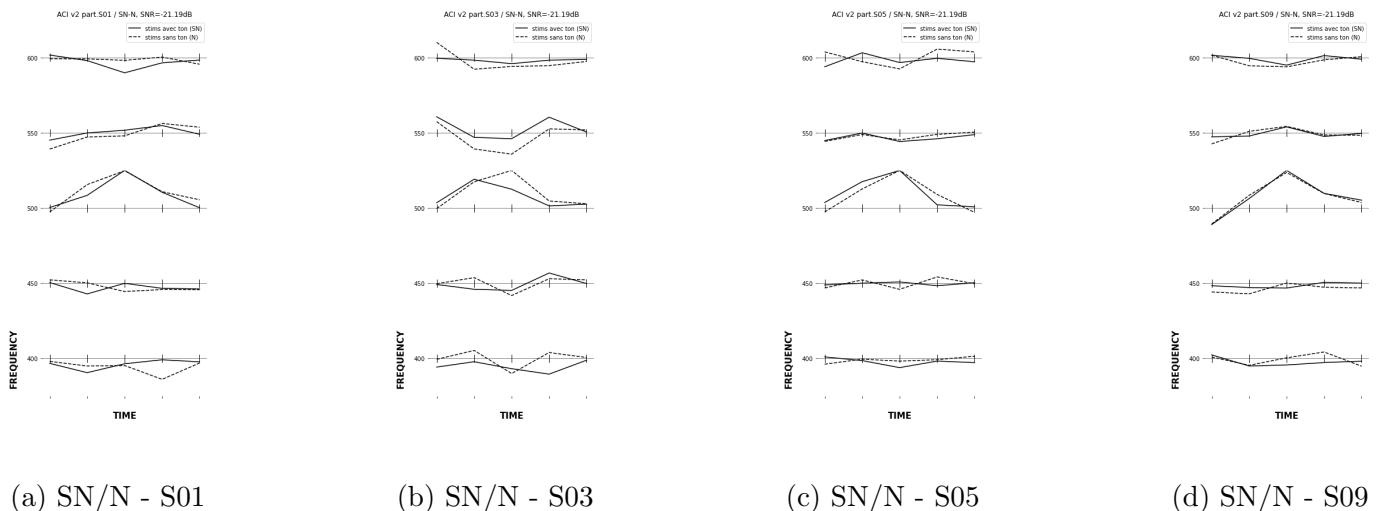


FIGURE 6.3 – ACIs linéaires SN/N pour 4 participants de l'étude

- ACIs matricielles pour les participants SQ02, SQ04, SQ06, SQ07, SQ08

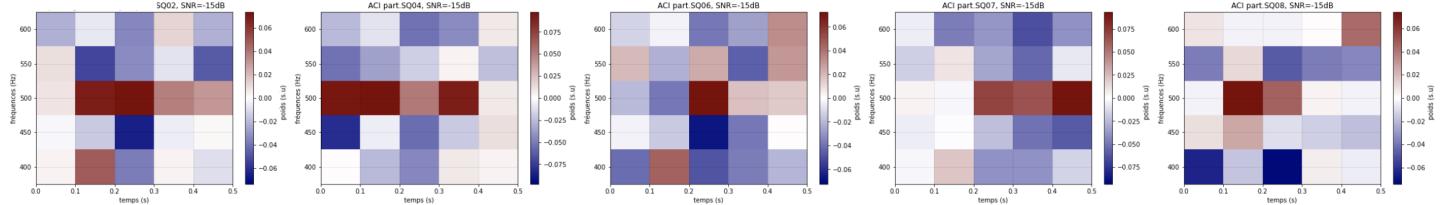


FIGURE 6.4 – ACIs : SQ02, SQ04, SQ06, SQ07, SQ08

- ACIs matricielles et pics pour les modèles de Green répliqués BW=40Hz / IT=500msec et BW=250Hz / IT=300msec. Les 7 autres modèles sont moins intéressants dans leur analyse donc pas présentés. Le modèle BW=40Hz / IT=500msec consiste en un modèle qui peut s'approcher de la stratégie de certains participants de manière plus proche que BW=40Hz / IT=300msec. Le modèle BW=250Hz / IT=300msec est un modèle large qui couvre une grande plage spectro-temporelle et renforce le manque conséquent de contrastes.

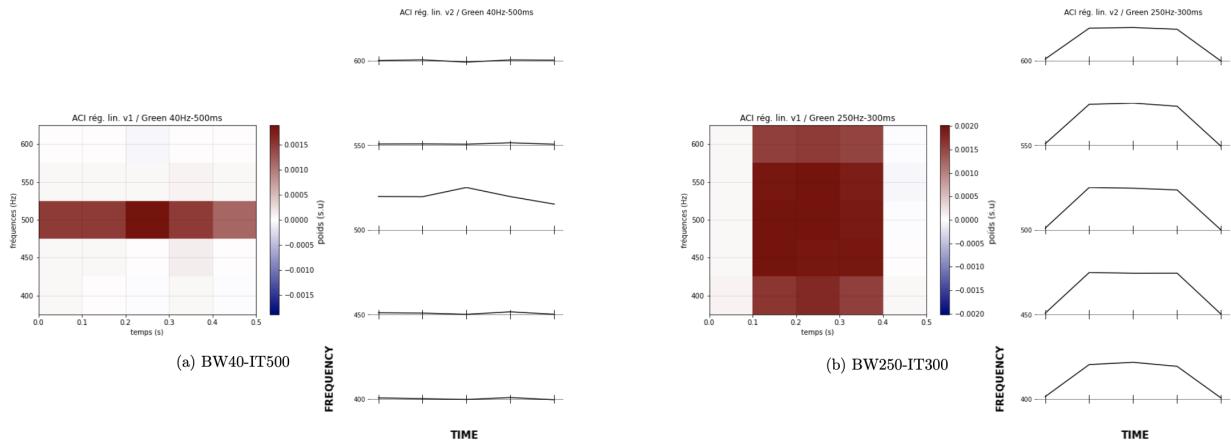


FIGURE 6.5 – ACIs de 2 autres modèles de Green

FIGURE 6.6 – Diagramme de Gantt - organisation du stage

Bibliographie

- [1] Arthur SANDUSKY Richard MARKEN, Al AHUMADA Jr. Time and frequency analyses of auditory signal detection. Technical report, University of California, Irvine/Santa Barbara, California, US, November 1974.
- [2] John A. SWETS David M. GREEN. *Signal Detection Theory and Psychophysics*. Peninsula Publishing (rev.2012), 1966.
- [3] Léo VARNET Alejandro OSSES. A microscopic investigation of the effect of random envelope fluctuations on phoneme-in-noise perception. In *The Journal of the Acoustical Society of America*. AIP Publishing, 2024.
- [4] Laboratoire des Systèmes Perceptifs,. Site du laboratoire :. <https://lsp.dec.ens.fr/fr>.
- [5] C. Douglas CREELMAN Michael J. HAUTUS, Neil A. MACMILLAN. *Detection Theory : A User's Guide (3rd ed.)*. Routledge, New York, US, 2022.
- [6] Albert AHUMADA Jr. *Detection of Tones Masked by Noise : A Comparison of Human Observers with Digital-Computer-Simulated Energy Detectors of Varying Bandwidths*. PhD thesis, Human Communication Laboratory, University of California, Los Angeles (US), November 1967.
- [7] John LOVELL Al AHUMADA Jr. Stimulus features in signal detection. Technical report, University of California, Irvine/Los Angeles, California, US, October 1970.
- [8] Richard F. MURRAY. Classification images : A review. In *Journal of vision (11)*, pages 1–25. Arvo Journals, 2011.
- [9] Virginia M. RICHARDS Laurie M. HELLER, David M. GREEN. The detection of a tone added to a narrow band of noise : The energy model revisited. In *The Quarterly Journal of Experimental Psychology Section A*, pages 481–501. Routledge, 1991.
- [10] F. A. WICHMANN V. H. SCHONFELDER. Identification of stimulus cues in narrow-band tone-in-noise detection using sparse observer models. In *The Journal of the Acoustical Society of America (p.447–463)*. AIP Publishing, 2013.
- [11] William M. HARTMANN. *Signals, Sound and Sensation*. Springer, New York, US, 2005.
- [12] Léo VARNET Christian LORENZI. Probing temporal modulation detection in white noise using intrinsic envelope fluctuations : A reverse-correlation study. In *The Journal of the Acoustical Society of America 151*. AIP Publishing, 2022.
- [13] Léo VARNET Géraldine CARRANANTE, al. Mapping the spectrotemporal regions influencing perception of french stop consonants in noise. In *bioRxiv (25.06.2024.600732)*. Cold Spring Harbor Laboratory, 2024.