



HAL
open science

Codage neuronal de l'ordre des signaux acoustiques dans les chants des oiseaux

Aurore Cazala

► **To cite this version:**

Aurore Cazala. Codage neuronal de l'ordre des signaux acoustiques dans les chants des oiseaux. Neurobiologie. Université Paris Saclay (COMUE), 2019. Français. NNT : 2019SACLS481 . tel-02496165

HAL Id: tel-02496165

<https://theses.hal.science/tel-02496165v1>

Submitted on 2 Mar 2020

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

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

Codage neuronal de l'ordre des signaux acoustiques dans le chant des oiseaux chanteurs

Thèse de doctorat de l'Université Paris-Saclay
préparée à l'Institut des Neurosciences Paris-Saclay

École doctorale n°9197 Signalisations et réseaux intégratifs en
biologie (BIOSIGNE)
Spécialité de doctorat : Sciences de la vie et de la santé

Thèse présentée et soutenue à Orsay, le 10 décembre 2019, par

Aurore Cazala

Composition du Jury :

Bénédicte Poulin-Charronnat Chargée de recherche, Laboratoire d'étude de l'apprentissage et du développement	Rapporteuse
Jean François Léger Directeur de recherche, Ecole Normale Supérieure Paris	Rapporteur
Dalila Bovet Professeure, Laboratoire éthologie, cognition, développement Université Paris Nanterre	Examinatrice
Christophe Pallier Directeur de recherche, NeuroSpin	Examineur
Catherine Del Negro Maîtresse de conférences, Institut des Neurosciences Paris-Saclay	Directrice de thèse
Nicolas Giret Chargé de recherche, Institut des Neurosciences Paris-Saclay	Co-Directeur de thèse

Remerciements

Je remercie tout d'abord et sincèrement tous les membres du jury qui ont accepté d'évaluer mon travail de doctorat. Je remercie en particulier Bénédicte Poulin-Charronnat et Jean-François Léger d'avoir bien voulu être rapportrice et rapporteur de ce travail, ainsi que Dalila Bovet et Christophe Pallier de s'être intéressé.e.s à mes recherches et de les examiner.

Je tiens à remercier toutes les personnes qui, de près ou de loin ont pu contribuer à l'accomplissement de cette thèse. Merci à mes très chers collègues et amis que j'ai rencontrés tout au long de mon parcours universitaire, de Marseille et Paris, avec qui j'ai pu entretenir, avec joie, mon addiction à la caféine et au sucre. Je n'oublierai jamais tout le soutien et la motivation que vous avez pu me fournir. Je remercie tout particulièrement Antoine sans qui ma vie au laboratoire n'aurait pas été la même. Solal, MJ, Lucie, Nolwenn et Mehdi, merci de m'avoir prouvé qu'à n'importe quel âge, il est possible de se faire des amis extraordinaires.

Je remercie aussi mon compère de toujours, Yann, qui a toujours eu les mots pour me faire rire dans les moments les plus tristes. Je dirais juste pour immortaliser à jamais cette phrase de campagnard que nous sommes : « On est à Paris ! ».

Merci à ma famille, vous rendre fier a toujours été un de mes principaux objectifs, merci de m'avoir fait confiance et soutenue dans tous mes choix. Et je n'oublie pas ma nouvelle famille, pour votre soutien incroyable.

Maxime, je pense que je n'ai absolument aucun mot pour décrire avec justesse l'aventure qu'a été cette période de notre vie, te iubesc.

Je remercie également Xavier d'avoir essayé de traduire toutes les idées farfelues que j'ai pu avoir. Tu as presque réussi à me faire aimer les neurosciences computationnelles.

Enfin, je remercie mes deux « chefs ». Cette thèse n'aurait jamais pu exister sans vous. Je me rends compte de l'incroyable chance que j'ai eu de vous avoir. Merci d'avoir tenu bon et de m'avoir permis de ne pas me saborder comme j'ai toujours eu l'habitude de le faire. Et puisque je ne sais jamais faire les choses selon les conventions, je préfère directement m'adresser à vous à travers ces remerciements.

Catherine, tu as été un véritable modèle de femme brillante et indépendante. Travailler toutes ces années à tes côtés fut énormément inspirant. Ta vision d'entreprendre la science, ton honnêteté dans chacune de tes démarches, et ta gentillesse ne sont que

quelques-unes des qualités que j'admire chez toi. Merci d'avoir été aussi compréhensive. Merci d'avoir été à l'écoute et d'avoir toujours pris le temps de me comprendre (et on sait toutes les deux à quel point ça peut être rude). Merci pour les fous rires. Et surtout merci de m'avoir envoyé ce mail pour m'offrir cette opportunité.

Merci aussi à toi Nicolas, pour avoir toujours été disponible à chaque fois que j'avais un problème (c'est-à-dire très souvent). Tu es un formidable chercheur et une formidable personne. J'espère que je ne t'ai pas trop traumatisé pour ta première expérience en tant que co-directeur de thèse. Merci pour toutes ces discussions où tu as dû m'écouter raconter ma vie et mes peurs notamment sur le « après ». Je n'oublierai jamais tous les moments passés en ta compagnie et tous tes conseils. Je te souhaite beaucoup de gâteau à la fraise pour la suite !

Merci à vous deux pour tout. Je ne sais toujours pas comment vous avez survécu à ces années avec moi dans vos pattes. Et j'avoue que je suis un peu triste de finir cette thèse juste pour le fait que je ne pourrai plus partager mon quotidien avec vous. Je n'aurais jamais pu imaginer une telle équipe avec des discussions aussi passionnantes et drôles. Si je laissais exprimer un peu plus ma folie, je remplirais plus d'une page avec seulement des mercis et éloges.

Pour finir, je remercie aussi toutes les personnes qui vont lire ce travail, en espérant qu'ils l'apprécieront. Et à tous les thésards qui liront ces remerciements : Bonne chance !

Et puisqu'il est de bon ton de faire des citations, je vous laisse avec le texte « Ce que je suis » de Erik Satie :

Erik Satie

Ce que je suis

Mémoires d'un amnésique - 1912

Tout le monde vous dira que je ne suis pas un musicien. C'est juste.

Dès le début de ma carrière, je me suis, de suite, classé parmi les phonométrographes. Mes travaux sont de la pure phonométrie. Que l'on prenne le « Fils des Étoiles » ou les « Morceaux en forme de poire », « En habit de cheval » ou les « Sarabandes », on perçoit qu'aucune idée musicale n'a présidé à la création de ces œuvres. C'est la pensée scientifique qui domine.

Du reste, j'ai plus de plaisir à mesurer un son que je n'en ai à l'entendre. Le phonomètre à main, je travaille joyeusement & sûrement.

Que n'ai-je pesé ou mesuré ? Tout de Beethoven, tout de Verdi, etc. C'est très curieux.

La première fois que je me servis d'un phonoscope, j'examinai un si bémol de moyenne grosseur. Je n'ai, je vous assure, jamais vu chose plus répugnante. J'appelai mon domestique pour le lui faire voir.

Au phono-peseur un fa dièse ordinaire, très commun, atteint 93 kilogrammes. Il émanait d'un fort gros ténor dont je pris le poids.

Connaissez-vous le nettoyage des sons ? C'est assez sale. Le filage est plus propre ; savoir classer est très minutieux et demande une bonne vue. Ici, nous sommes dans la phonotechnique.

Quant aux explosions sonores, souvent si désagréable, le coton, fixé dans les oreilles, les atténue, pour soi, convenablement. Ici, nous sommes dans la pyrophonie.

Pour écrire mes « Pièces froides », je me suis servi d'un caléidophone-enregistreur. Cela prit sept minutes. J'appelai mon domestique pour les lui faire entendre.

Je crois pouvoir dire que la phonologie est supérieure à la musique. C'est plus varié. Le rendement pécuniaire est plus grand. Je lui dois ma fortune.

En tout cas, au monodynamophone, un phonométreur médiocrement exercé peut, facilement, noter plus de sons que ne le fera le plus habile musicien, dans le même temps, avec le même effort. C'est grâce à cela que j'ai tant écrit.

L'avenir est donc à la philophonie.

Revue musicale de la Société Internationale de Musique, 15 avril 1912.

Table des matières

Remerciements.....	1
Introduction générale	11
I. La dimension temporelle : un rôle essentiel dans la caractérisation des stimuli sonores.....	11
II. La dimension temporelle dans la compréhension du langage.....	12
A. L'importance de l'ordre dans la compréhension des mots	12
B. L'importance de l'ordre dans la construction des phrases	14
C. L'ordre et le contexte	16
III. Apprentissage du langage.....	17
A. Apprentissage social.....	17
B. Analyse computationnelle chez l'adulte	19
C. Analyse des « <i>chunks</i> ».....	21
D. Autres mécanismes impliqués dans le traitement de l'ordre.....	22
IV. Les oiseaux chanteurs comme modèle d'étude des bases neurobiologiques du traitement des séquences de signaux acoustiques.....	26
A. Le chant des oiseaux chanteurs : séquence de signaux acoustiques organisés temporellement.....	28
B. Apprentissage du chant des oiseaux.....	30
C. L'ordre est porteur d'information.....	35
V. Bases neurobiologiques du chant chez les Oiseaux chanteurs	38
A. Un noyau sensori-moteur, le HVC	40
B. Codage des probabilités de transition dans le HVC.....	44
C. Codage moteur de l'organisation temporelle du chant.....	46
VI. Le système auditif des oiseaux chanteurs	50
A. Aire auditive primaire.....	51
B. Aire auditive secondaire	51

VII. Objectifs de l'étude.....	57
A. Axe 1 : Étude du traitement de l'ordre des signaux acoustiques au sein d'une aire auditive, le NCM, chez le diamant mandarin.....	57
B. Axe 2 : Étude de la sensibilité à l'ordre des signaux acoustiques qui composent le propre chant de l'oiseau, au sein d'un noyau sensori-moteur, le HVC, chez le canari.....	58
Partie I : Traitement de l'ordre des signaux acoustiques dans une aire auditive secondaire.....	61
Article: Neuronal Encoding in a High-Level Auditory Area: From Sequential Order of Elements to Grammatical Structure.....	61
Introduction.....	61
Résultats.....	62
Conclusion.....	63
.....	77
Partie II : Structure globale du chant de canari : Organisation en séquences récurrentes (<i>chunks</i>) et importance de la position.....	79
Introduction.....	79
C. La structure du chant chez le canari : plusieurs niveaux d'organisation.....	81
D. Enchaînement probabiliste des phrases dans les chants de canaris.....	84
E. But de l'étude.....	87
Matériels et méthodes.....	89
F. Sujets et acquisition des chants.....	89
G. Analyse des chants.....	89
Résultats.....	92
H. Occurrence des phrases et séquences de phrases.....	92
I. Probabilités de transition.....	95
J. Dépendances à long terme.....	98
K. Importance de la position des phrases.....	101

L. Organisation des chants de canari selon les distances entre les phrases dans les chants.....	106
.....	110
Discussion.....	111
Partie III : Etude du codage neuronal de l'organisation temporelle du propre chant de l'oiseau dans une aire sensori-motrice	117
Introduction.....	117
Matériels et méthodes	122
A. Sujets	122
B. Stimuli auditifs.....	122
C. Création des différentes versions modifiées du BOS.....	126
D. Enregistrements électrophysiologiques.....	129
E. Analyse des données électrophysiologiques.....	130
F. Statistiques	131
Résultats.....	132
A. Réponses aux phrases du BOS.....	132
B. Importance du type de phrases qui précèdent	141
C. Réponses aux autres types de chant	144
D. Impact des changements globaux et locaux sur les réponses aux phrases...	153
E. Étude des réponses à la phrase finale	157
.....	158
F. Étude des réponses lors d'une modification du contexte à partir d'une seule phrase.....	159
G. Réponses en fonction du type de neurones.....	164
Discussion.....	167
A. Réponses des neurones et probabilités de transition	167
B. Réponses des neurones et position des phrases dans le chant	169

C. Réponses des neurones et <i>chunks</i>	171
D. Traitements des informations relatives à l'agencement des phrases par les deux populations neuronales.....	172
E. Intégration à long-terme du type de phrases qui précèdent	174
F. Conclusion.....	175
Discussion générale.....	178
Aspects méthodologiques.....	179
Contribution des études à la littérature.....	181
Encodage de l'ordre des syllabes dans le chant au sein du NCM et du HVC.	184
Questions soulevées par nos études.....	185
Conclusion	186
Références bibliographiques	191
Annexe	205
Table des figures.....	205
Table des tableaux :	208

INTRODUCTION GÉNÉRALE

Introduction générale

La communication acoustique, qu'elle se fasse à travers la parole comme chez l'humain, ou le chant comme chez les oiseaux chanteurs, implique de la part du récepteur une fine reconnaissance des signaux sonores produits par l'émetteur et de la part de l'émetteur, un contrôle précis de ce qu'il produit. La parole et le chant ne sont pas de simples enchaînements de sons ayant chacun une signification indépendamment les uns des autres. Le langage humain possède différents niveaux d'organisation (sons autrement appelés phonèmes, mots, phrases) qui s'étendent sur différentes échelles de temps (dizaines de millisecondes pour les phonèmes, centaines de millisecondes et plus pour les mots, plusieurs secondes pour les phrases). Au sein de chacun de ces niveaux d'organisation, les informations transmises dépendent de l'ordre dans lequel les signaux sont produits. La compréhension de la parole nécessite alors une intégration temporelle des informations acoustiques. La question au centre de ce travail est de déterminer les mécanismes cérébraux impliqués dans le traitement de l'ordre temporel des signaux acoustiques en prenant pour modèle d'étude les oiseaux chanteurs.

I. La dimension temporelle : un rôle essentiel dans la caractérisation des stimuli sonores

Qu'ils s'agissent de stimuli sonores simples comme un son de klaxon dans la rue ou de stimuli complexes tels que les mots que l'on peut prononcer au cours d'une conversation, il est nécessaire non seulement de pouvoir les détecter, mais aussi et surtout de les identifier. Dès que l'on cherche à décrire les stimuli sonores, pour la plupart, on fait appel non seulement à des paramètres physiques tels que la fréquence et l'intensité, mais aussi au temps. Le décours temporel des fréquences et de l'intensité caractérise les sons qui ont une signification dans notre environnement sonore. Ainsi, diffuser un mot à l'envers, même très court parce que constitué d'une seule syllabe (ex : le, un, des, nous, etc.) conduit à ce que l'on ne reconnaisse pas ce mot (Saffran et al. 2007). Pourtant, *la gamme de fréquences qui caractérise la syllabe n'aura pas changé*. En revanche, l'ordre dans lequel les fréquences de cette gamme arrivent sera différent, les

modulations de fréquences étant inversées. De la même façon, les modulations d'intensité auront été inversées.

Au-delà des mots monosyllabiques, l'ordre des syllabes dans un mot est important pour lui donner une identité. De plus, la question de l'ordre s'étend bien au-delà de la simple diffusion des sons mis bout à bout. Il faut prendre en compte que cet ordre est soumis à de nombreuses règles. Cependant, si pour d'autres processus sériels, comme le fait d'attraper un objet, les paramètres temporeux se définissent principalement par l'ordre de chaque événement, pour le langage, il faut tenir compte d'autres aspects temporels. La vitesse de production des sons, les silences, mais surtout, les nombreuses règles, syntaxiques et grammaticales, vont régir ce paramètre (Saffran et al. 2008; Ding et al. 2017; Chomsky 2017; Jusczyk 1999; Langus et al. 2017).

II. La dimension temporelle dans la compréhension du langage

A. L'importance de l'ordre dans la compréhension des mots

Les mots et les phrases du langage humain ne sont pas de simples séries de sons où chacun aurait acquis sa signification. En français, par exemple, le répertoire est limité à 37 phonèmes, mais à partir de ce répertoire plus de 35 000 mots ont été créés, chacun ayant une signification qui lui est propre, ces mots pouvant eux-mêmes être combinés de multiples manières dans les phrases. Le sens d'un mot ou d'une phrase ne dépendra pas seulement des phonèmes ou des lettres qui le composent, mais aussi de sa relation avec les autres mots, autrement dit de l'agencement de ces derniers. Ainsi depuis l'antiquité, la construction d'anagrammes, c'est-à-dire la construction de différents mots à partir des mêmes lettres, est un divertissement pour le plus grand nombre. À travers le temps, de nombreuses personnalités telles que Paul Verlaine (Pauvre Lélian), ont utilisé des anagrammes pour se créer des pseudonymes, mais aussi détourner le sens des mots ou même pour en cacher le véritable sens et transmettre discrètement des informations comme Galilée qui a annoncé certaines de ses découvertes de 1610 en astronomie par des anagrammes. Si l'on peut jouer avec les lettres, on peut aussi le faire avec les sons. La contrepèterie, dont on attribue la naissance à Rabelais dans son ouvrage Pantagruel, est l'art de jouer avec les mots en inversant certains phonèmes ou syllabes dans une phrase pour en créer une nouvelle, au sens souvent plus tendancieux. Dans ce jeu, seul le son

compte, l'orthographe du mot créé peut être fautive tant que la correspondance phonétique reste parfaite, et le public se doit de deviner seul la solution. On retrouve notamment à cet exercice Victor Hugo avec par exemple « Le vaincu de son cœur » ou « Goutez-moi cette farce » d'Estienne Tabourot.

Ces deux jeux mettent en avant l'importance de l'ordre dans lesquels les lettres et phonèmes sont agencés pour délivrer une information donnée. Si celui-ci est altéré, l'information contenue dans le mot en est modifiée.

S'il est important de respecter l'ordre des phonèmes au sein des mots, il l'est tout autant de respecter l'ordre des mots dans les phrases. La compréhension d'une langue implique d'en connaître (implicitement ou non) les règles. Ces règles qualifiées de syntaxiques se définissent comme les relations qui existent entre les unités linguistiques, c'est-à-dire les mots. Avant de développer ce point, selon le point de vue d'auteurs qui s'intéressent au langage humain et aux chants des oiseaux, on note deux types de syntaxe dans le langage, une syntaxe phonologique et une syntaxe lexicale (Berwick et al. 2011; Collier et al. 2014). La syntaxe phonologique regroupe les règles qui permettent de combiner des unités sonores simples comme des phonèmes pour créer un mot porteur d'un sens (Hurford, 2011). Par exemple, les sons [ti] et [tr] n'ont pas de sens isolés en français, mais lorsqu'ils sont associés en un seul mot en prennent un : « titre ». La syntaxe lexicale, présentée par les linguistes comme exclusivement humaine (Chomsky, 2017), lie la syntaxe et sémantique. Elle correspond à deux éléments sonores qui, indépendamment, apportent un sens et qui, combinés, en amènent un nouveau. Par exemple, [ra] et [vi], ont chacun un sens si on les écoute isolés l'un de l'autre : « rat » et « vie ». Ils acquièrent un nouveau sens si ces deux sons sont associés en une seule séquence formant « ravie ».

Pour revenir au fait que les mots et les suites de mots ne sont pas de simples séries de sons où chacun aurait acquis sa signification, par exemple, la signification de « comporte » n'est pas la même que celle de « qu'on porte » alors que dans les deux cas l'on entend le son ou autrement appelé phonème, [kon] suivi de [por] et [te]. Pour différencier ces deux séries de sons identiques, identifier la construction de la phrase est nécessaire. La compréhension d'une langue implique donc d'en connaître les règles et de les prendre en compte à un degré plus haut que celui de la simple écoute ou lecture sérielle.

B. L'importance de l'ordre dans la construction des phrases

L'agencement des mots au sein des phrases est régi par des règles grammaticales. La grammaire est l'ensemble des règles qui régissent la construction des phrases, on y retrouve notamment la syntaxe (composition des mots et phrases), les règles d'accord, et la conjugaison. La grammaire varie selon les langues, mais l'on retrouve des similitudes entre elles comme la présence de sujet, verbe et complément dans une phrase. Seulement l'ordre dans lequel ils sont organisés peut différer, voire être opposés, d'une langue à une autre comme l'allemand où le sujet peut être à la fin au contraire du français. Pour maîtriser une langue, il est donc important d'intégrer toutes ses règles, disposer d'un large vocabulaire ne suffit pas. Si l'on ne maîtrise pas ces règles, l'information contenue dans la phrase peut en être totalement altérée même si le vocabulaire employé est juste. Par exemple, si l'on inverse le sujet et le complément, la phrase pourra être grammaticalement correcte, mais l'information transmise sera différente.

Émilie aime le chocolat ne signifie pas que le chocolat aime Émilie.

Une distinction souvent présente dans la littérature est celle faite entre deux types de grammaires : des grammaires dites régulières ou à état fini (FSAs pour automate à état fini) et les grammaires dites libres du contexte (CFGs) (Fitch and Hauser 2004; Hauser 2002). Les grammaires régulières (FSGs) sont présentées comme les plus simples. Elles comprennent un nombre limité de règles pouvant être données à une machine (un automate) afin de générer un nombre fini d'états bien définis. En ce qui concerne ces grammaires régulières, la succession des états (des syllabes par exemple) est déterminée : après la syllabe A vient la syllabe B, ou dépendante des probabilités : après la syllabe A il peut y avoir B dans 30% des cas et C dans 70% des cas. Les grammaires libres du contexte serviraient de base pour distinguer le langage humain des systèmes de communication animale. Elles permettraient un nombre infini de phrases grâce à la récursivité. Si l'on représente les noms par As et les verbes par Bs, les grammaires FSGs permettent de générer des phrases suivant la structure « $(AB)^n$ » (**Figure 1A**) alors que les grammaires CFGs permettraient de générer des phrases suivant la structure « $A^n B^n$ » (**Figure 1B**). La récursivité consiste à construire des phrases qui s'emboîtent dans d'autres phrases de telle façon que l'on peut avoir par exemple pour $n=2$, $A_1 [A_2 [A_3 B_3] B_2] B_1$ pour la phrase (en anglais "the man the woman heard laughed"). En français,

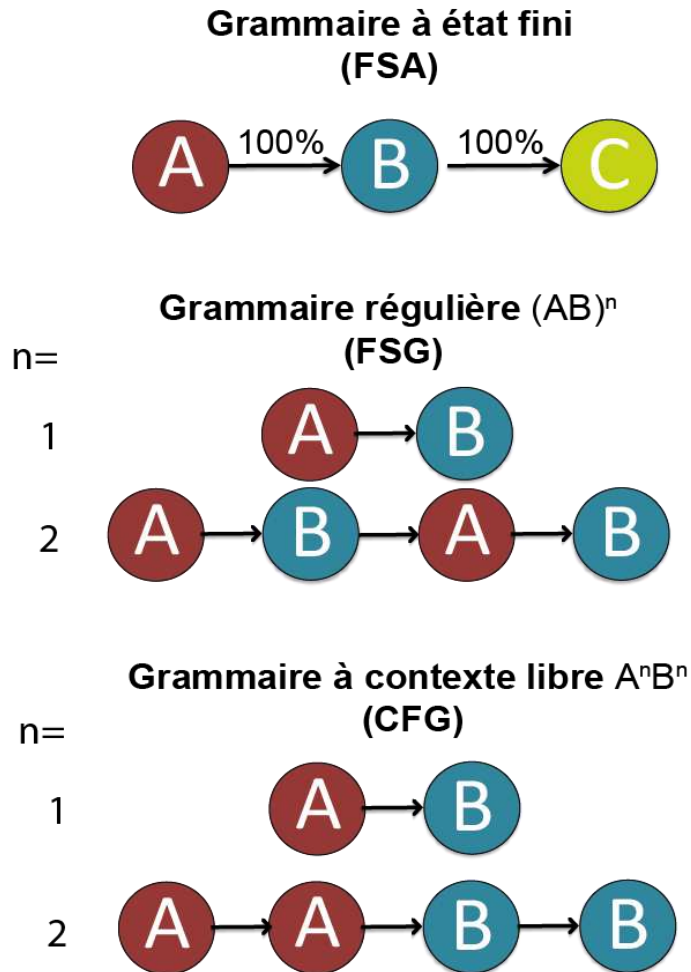


Figure 1 : Schématisation des trois types de grammaires : à état fini, régulière et libre du contexte

Chaque lettre présente un élément ou une catégorie d'élément et les flèches les transitions entre ces éléments.

(Haut) Grammaire à état fini ou finite state où chaque transition est déterminée par des probabilités fixes (100%)

(Milieu) Structure similaire à la grammaire à état fini dont les probabilités de transition peuvent varier entre les éléments (FSG) $(AB)^n$

(Bas) Grammaire libre du contexte (CFG) A^nB^n .

Ces deux grammaires (FSG et CFG) décrivent des schémas de séquences d'éléments de la catégorie « a » ou « b » où n correspond au nombre de répétition de chaque élément. Si l'on augmente le n, la séquence résultant pour la FSA apparaît comme une alternance entre les deux éléments (A) tandis que pour la CFG, il s'agit d'une structure récursive avec une symétrie au milieu entre les éléments a et b.

on ajoute à ce schéma la conjonction « que ». Dans l'exemple : « le chat que la souris voit mange ; la phrase « la souris voit » est emboîtée dans la phrase « le chat mange ». Dans cet exemple, nous comprenons qu'il s'agit du chat qui mange et non pas la souris. S'il est facile d'analyser des phrases basées sur une grammaire à état fini, la récursivité rend la structure des phrases plus complexe. Elle va nécessiter un haut degré d'analyse laissant entendre un encodage gardant en mémoire chaque élément sur un relativement long terme, bien plus long que celui nécessaire lors de l'utilisation d'une grammaire à état fini. En effet, il faut associer dans l'exemple précédent A_1 avec B_1 , A_2 avec B_2 et A_3 avec B_3 et ne pas mélanger les groupes au risque d'en changer la signification.

C. L'ordre et le contexte

Ainsi la grammaire est une composante importante pour définir la signification d'un mot, mais pas seulement. Le contexte dans lequel est un mot n'est pas seulement régi par les règles grammaticales et syntaxiques, mais aussi par les relations à plus ou moins long terme entre chaque mot. Selon le linguiste Claude Hadège, l'une des difficultés de la maîtrise d'une langue se retrouve, notamment, dans la sémantique évolutive de certains mots en fonction du contexte auquel ils sont associés. Par exemple, en anglais, « to call someone out » après une analyse mot à mot, la traduction supposée est « dire quelqu'un hors de », mais signifie en réalité « mettre quelqu'un dehors ». Cependant, si dans cet exemple, il reste possible d'en déduire le sens approximatif, ce n'est pas toujours le cas. Un autre exemple serait « to do someone in » ou « to do someone on », littéralement « faire quelqu'un dedans » ou « faire quelqu'un sur », tandis que cette séquence de mots se traduit par le verbe « tuer ». Dans ces cas-là, le sens n'est absolument pas déductible de l'analyse des éléments constituants. De même, seul le contexte dans lequel se situe "Short breathing exercises" peut déterminer sa traduction entre "brefs exercices de respiration" ou "exercices de respiration brève".

L'ordre entre chaque mot à long et court terme est donc un paramètre important dans la compréhension d'une langue et laisse entendre des mécanismes complexes qui sous-tendent cet apprentissage.

III. Apprentissage du langage

A. Apprentissage social

Le traitement de l'ordre des sons que l'on entend afin de reconnaître les mots et de comprendre les phrases n'existe pas d'emblée à la naissance. Il va résulter d'un apprentissage débutant très tôt au cours du développement.

Pendant les premiers mois de leur vie, les bébés humains sont sensibles aux sons et aux rythmes de toutes les langues du monde (Eimas, 1975; Eimas et al., 1971), capacité qui se perd aux alentours de la première année (Werker and Lalonde, 1988.; Kuhl 2004; 2010). Le bébé, écoutant ses parents parler, se constitue une « base de données » des sons entendus, dans un premier temps des syllabes comme [pa] [bi] [mo], *etc.* La perception des syllabes isolées est essentielle dans le traitement du langage oral que ce soit dans le silence ou dans le bruit, mais une autre capacité est fondamentale pour l'acquisition du langage. Il s'agit de la capacité à extraire et segmenter les mots du flot de sons de la parole. En effet, le signal de parole ne contient pas systématiquement de silence entre les mots ou d'indices acoustiques qui indiquent la fin ou le début d'un mot. Selon certains auteurs, notre capacité de segmentation pourrait reposer, notamment, sur une sensibilité aux structures statistiques des séquences de syllabes présentes dans notre langage (Saffran et al 1996; Santolin and Saffran 2018). Par exemple, en supprimant les espaces représentant les silences entre chaque mot, dans la phrase « l'apprentissage **ge** du langage est complexe », la segmentation innée que nous faisons entre les mots est possible grâce à, notamment, une intégration inconsciente des probabilités variables entre chaque syllabe. Les probabilités de transition entre « **ge** » et « **du** » sont plus faibles qu'entre « sa » et « **ge** », permettant de considérer la transition « ge » / « du » comme une transition entre deux mots. En effet, les transitions possibles entre chaque phonème ne sont en effet pas équivalentes : certains phonèmes ne seront jamais associés tandis que d'autres le seront fréquemment. En français par exemple aucun son [tr] ne peut suivre le phonème [br]. En étant exposé un nombre important de fois aux différents mots et phrases, l'idée serait que l'enfant apprend petit à petit à détecter les régularités de sa langue maternelle. C'est la répétition fréquente des paires de sons composant « papa » et « maman » qui ferait que l'enfant les détecte et les considère en tant qu'ensembles uniques, en d'autres termes comme une entité, c'est-à-dire un mot. Cet apprentissage dit

statistique (Saffran et al 1996; Goldin-Meadow and Yang 2017) (ou computationnel) existerait grâce à notre répertoire limité dont chaque phonème se retrouve au sein de plusieurs mots. Cet apprentissage qui s'opèrerait dès le début de l'apprentissage du langage permettrait de déduire implicitement les règles de la langue (Saffran et al 1996; Kuhl 2004).

Le langage oral étant notre principal système de communication, il est important que la détection des règles qui le définissent soit possible dès le plus jeune âge. Ainsi le bébé peut être capable d'apprendre des simili-mots sur la base de probabilité de transition avant même d'y avoir associé un sens (Saffran et al. 1996). Pour démontrer cette capacité, il a été diffusé à des enfants de huit mois, pendant deux minutes, une séquence composée de phonèmes synthétisés à partir d'un ordinateur. Cette séquence est construite de manière à ce qu'elle contienne quatre pseudo-mots du type :

« **pabiku** », « **tibudo** », « **golatu** » et « **daropi** »

Au sein de la séquence, il n'y a ni pause, ni accentuation, ni intonation sur certaines syllabes ou silences. En résumé, aucun indice ne permet de distinguer une syllabe d'une autre. Une séquence est par exemple :

...**tibudopabikugolatudaropipabiku**...

L'idée est que la seule stratégie que le nourrisson peut adopter est de se baser sur les probabilités de transition. Au sein des mots, les probabilités de transition entre chaque syllabe sont de 100%, par exemple dans le mot « **tibudo** », « **ti** » a 100% de chance d'être suivi de « **bu** » qui lui-même a 100% de chance d'être suivi de « **do** ». En revanche, entre les pseudo-mots, les probabilités de transition sont réduites à 33%, autrement dit la probabilité qu'avant « **ti** » il y ait « **ku** » est de 33%, « **ti** » faisant partie du mot « **tibudo** » et « **ku** » du mot « **pabiku** ». Il faut donc détecter les transitions faibles pour identifier où commence et finit le pseudo-mot. Après exposition, les enfants sont testés selon leur préférence d'écoute (à travers la mesure du temps d'écoute) entre deux mots originaux, par exemple « **tibudo** » et « **pabiku** » et deux non-mots, par exemple « **budopa** » et « **bikuti** ». La nouveauté est un paramètre plus stimulant pour les enfants en termes d'attention. Un temps d'écoute plus long pour les non-mots indique que l'écoute est considérée comme nouvelle par rapport celle des mots originaux auxquels ils ont été préalablement exposés. La préférence pour la liste contenant les deux non-mots met en avant une détection des irrégularités statistiques présentes dans les mots originaux. Plus tard, il a été montré que cette préférence n'est pas due aux fréquences d'apparitions des

syllabes (Saffran et al 1996b). Le choix n'est pas fait par rapport aux phonèmes les plus entendus, mais bien par rapport aux probabilités de transition entre ces phonèmes. Deux minutes d'exposition à une séquence suffisent à l'enfant de huit mois pour détecter les pseudo-mots. Ce serait à partir de cette capacité que l'on apprend à segmenter les phrases en une série de mots.

B. Analyse computationnelle chez l'adulte

Au cours du temps et des multiples expériences auditives, nous intégrerions donc implicitement des probabilités variables qui caractérisent les transitions d'une syllabe à une autre, permettant d'avoir une détection efficace de l'organisation syntaxique des mots et des phrases (Saffran et al 1996b). Cette organisation suivrait notamment une chaîne de Markov ou markovienne (Markov, 1913). La chaîne de Markov se caractérise par le fait que l'état au temps t du processus ne dépend que de son état au temps $t-x$ où x varie en fonction de l'état considéré. Cette règle peut s'appliquer à tout événement inclus dans une séquence, mais a été tout d'abord mise en avant dans le langage.

On observe plusieurs ordres dans une chaîne de Markov et le langage humain peut être régi selon des ordres supérieurs ou égaux à 1 (Markov, 1913; Petruszewycz, 1983). L'ordre 1 de cette chaîne correspond à une analyse probabiliste linéaire (Markov, 1913). Prenons l'exemple de l'écriture des mots de la langue française : après la lettre « R » peuvent suivre de nombreuses lettres dont les lettres « D » ou « E », avec leur probabilité d'apparition propre. Mais ces probabilités de transition d'ordre 1 ne suffisent pas à décrire l'enchaînement des lettres ou phonèmes qui composent les mots de la langue française, il faut également prendre en compte les probabilités d'ordre 2 (ou plus) (**Figure 2B**). Si avant la lettre « R », il y a la lettre « P », alors la lettre « E » peut suivre pour former l'enchaînement « PRE », mais il ne peut plus y avoir la lettre « D » puisque « PRD » n'existe pas dans notre langue. Notons que cette organisation en chaîne de Markov s'applique aussi bien aux mots, qu'aux phonèmes et aux phrases. Dans un sens plus large, les règles syntaxiques peuvent être considérées comme suivant des chaînes de Markov. En effet, la construction d'un mot peut dépendre de ce qui précède. Par exemple, la phrase « je la mange » est grammaticalement juste tandis que « je prends la mange » est fausse. Il est attendu après le deuxième « la » un nom, au contraire du premier où il est attendu un verbe.

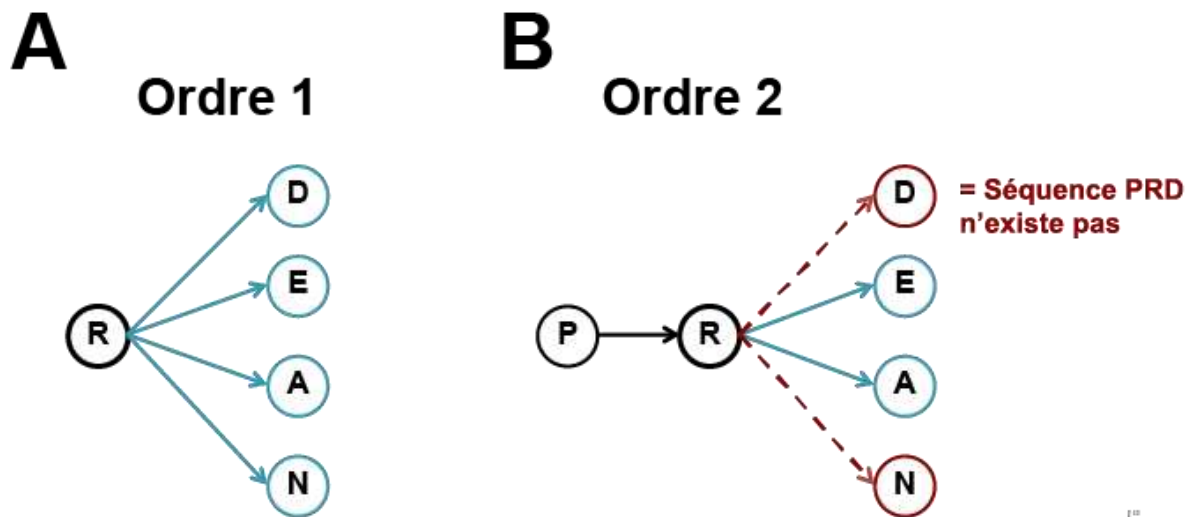


Figure 2 : Exemple de chaînes de Markov d'ordre 1 et 2

Les flèches bleues représentent les transitions possibles entre les lettres et les flèches rouges pointillées représentent les transitions inexistantes.

- (A) Calculer les probabilités de transitions d'ordre 1 consiste à calculer les probabilités pour chacune des transitions entre un élément (ici la lettre R) et celui qui suit (D, E, A ou N)
- (B) Les probabilités d'ordre supérieur et ici d'ordre 2, calculent les probabilités de transition pour un élément, R, et ce qui suit, tout comme l'ordre 1. Cependant, ces probabilités se doivent de considérer l'élément n-1 (P). Dans cet exemple, les séquences P-R-D et P-R-N n'existant pas, les probabilités de transition sont de zéro pour les transitions R-D et R-N (flèche rouge en pointillé).

L'exposition répétée aux mots, tout particulièrement dans un contexte social (Kuhl 2014; Kuhl et al. 2003), pourrait donc être importante pour détecter ces chaînes markoviennes (Fisher et al. 2006; Christophe et al. 2008; Bernal et al. 2010). Il est nécessaire de pouvoir entendre un grand nombre de conversations pour déterminer les différentes probabilités de transition qui régissent la création des mots et des phrases. Cela est d'autant plus remarquable que ce mécanisme s'effectue sans association d'une signification, c'est-à-dire d'un traitement sémantique du mot (Silva-Pereyra et al. 2007; Saffran et al. 2008). La détection des relations statistiques entre chaque son, de leur place dans la séquence, à court et long terme, serait donc essentielle dans l'apprentissage et la perception du langage.

C. Analyse des « *chunks* »

D'autres auteurs reviennent sur cette idée que l'apprentissage du langage résulterait de l'acquisition des régularités statistiques présentes dans ce que l'enfant entend. Ils suggèrent que l'apprentissage du langage ne passerait pas seulement par cette acquisition, mais aussi par la mémorisation de ce qui est appelé des « *chunks* » (Perruchet and Pacton, 2006; Perruchet and Vinter, 1998).

La notion de « *chunk* » renvoie à nos capacités de mémoire à court terme. Nous mémoriserons bien mieux la série 178920013456 si on la découpe en trois groupes ou « *chunks* » : 1789-2001-3456. Ainsi au lieu de retenir 12 chiffres, on doit seulement se souvenir de 3 nombres. Le terme de « *chunking* » est également utilisé pour indiquer le fait qu'une série d'éléments en étant regroupée forme alors une entité unique, un seul item. Ce terme caractériserait un mécanisme par lequel nous regrouperions des informations qui apparaissent de manière contingente et fréquente dans l'environnement.

L'idée serait que le mécanisme de *chunking* serait impliqué dans l'acquisition du langage (Hay et al., 2011; Perruchet et al., 2014). Il serait impliqué plus particulièrement dans la segmentation des mots et par là même contribuerait à l'acquisition du lexique (ensemble des mots de la langue).

La mise en avant du mécanisme de *chunking*, par rapport à celui du traitement statistique repose notamment sur le fait que l'apprentissage d'une grammaire artificielle, au cours duquel les sujets apprennent les règles qui régissent l'agencement d'éléments

(sons, lettres, et images) dans des séquences, donne de meilleures performances si les sujets segmentent les séquences en petits *chunks* revenant fréquemment qu'en *chunks* plus longs moins fréquents. Pourtant, les probabilités au sein de ces *chunks* sont les mêmes, c'est-à-dire que les différents éléments de ces *chunks* se succèdent toujours (Perruchet et al., 2002; Perruchet and Pacton, 2006). Il peut être ajouté qu'une étude a montré que, si des enfants de 17 mois sont préalablement entraînés dans la tâche de Saffran et collaborateur (1996) dans laquelle des pseudo-mots (de type « timay, dobu, piga, mano) sont présentés de façon ininterrompue en variant l'ordre de succession de ces pseudo-mots, mais sans jamais casser ces mots, les enfants apprennent bien plus vite à associer ces pseudo-mots à un objet et donc à détecter des mots avec une sémantique, que des enfants qui n'ont pas été soumis à cet apprentissage préalable (Estes et al 2007).

D. Autres mécanismes impliqués dans le traitement de l'ordre

Outre le mécanisme de *chunking*, d'autres ont été proposés pour comprendre comment les séquences de multiples éléments sont apprises et reconnues, comment les règles qui les régissent sont acquises et permettent d'apprendre à reconnaître les séquences ayant la même structure. Une revue récente qui se base sur des études rapportant des corrélats cérébraux des traitements étudiés propose cinq mécanismes (dont le *chunking*) qui seraient impliqués dans le traitement de l'ordre dans les séquences (Dehaene et al., 2015). Il faut dire que de nombreuses études d'imagerie à résonance magnétique (IRM), magnétoencéphalographie (MEG) et électroencéphalographie (EEG) se sont intéressées à comment sont traitées les séquences au niveau cérébral, cherchant particulièrement à déterminer quelles étaient les régions cérébrales mises en jeu. D'après les auteurs, selon le niveau d'abstraction requis, ces mécanismes sont (**Figure 3**) :

- 1) L'acquisition des transitions au sein des séquences par la mémorisation des temps entre les évènements ou stimuli qui se succèdent.

La survenue d'un stimulus donné serait à l'origine de l'attente d'un autre stimulus à un moment donné. Ce mécanisme est illustré par les résultats d'étude faisant appel au paradigme « d'oddball ». Ce paradigme consiste en la diffusion de séquence sous le schéma « A^n » (où n est le nombre de répétitions du stimulus A , présenté à intervalle régulier) tel que

AAAAAA...

Puis après un certain nombre de diffusions à intervalle régulier, il est diffusé un nouveau stimulus « **B** »,

AAAAAA...**AB**

Lorsque ce nouveau stimulus, **B**, remplace de façon inattendue le stimulus **A** attendu, une augmentation de l'activité au niveau de certaines régions cérébrales est enregistrée (Garrido et al., 2009). Cette activation est appelée une réponse de Mismatch (réponse MMR). On observe également cette réponse si le son A n'est pas diffusé à un instant *t*, soit la séquence « **A** » suivi d'un silence, notifié ici par un « _ »

AAAAAA...**A_**

Si deux stimuli différents A et B sont alternés et que le B n'est pas émis, soit

ABAB...**ABA_**

la réponse MMR apparaît suggérant qu'un mécanisme neuronal de prédiction a été mis en jeu. Il faut dire que la répétition d'un même stimulus peut s'accompagner dans certaines régions cérébrales, et plus particulièrement au sein des régions sensorielles, d'une diminution de l'activité des neurones, pouvant être suivie d'une augmentation si le stimulus change (ce phénomène est appelé adaptation spécifique du stimulus, SSA). Mais dans le cas où le stimulus n'arrive pas, un tel mécanisme neuronal ne peut expliquer la survenue de la réponse MMR. Un tel résultat suggère donc qu'il pourrait y avoir, d'un point de vue temporel, une attente du stimulus (Friston, 2005).

2) Le *chunking*.

D'autres arguments, que ceux présentés plus haut, provenant d'études de neurosciences, suggèrent l'existence d'un tel processus. Pour revenir au protocole qui vient d'être décrit, dans certaines études, au lieu de présenter de façon continue les stimuli, ils sont présentés en séquences

AAA**AB** AAA**AB**

A la MMR s'ajoute une autre réponse plus tardive qui disparaît au fur et à mesure des répétitions, si le stimulus B est remplacé par A, soit

AAA**AB** AAAAA

Cette réponse tardive réapparaît suggérant qu'une mémorisation de la séquence en tant qu'entité a eu lieu. Cela a été observé chez l'adulte et l'enfant de 3 mois.

3) Le traitement de la position.

On mémoriserait quel est le stimulus qui vient en premier dans la séquence, puis quel est le 2^{ème}, ainsi de suite, et notamment du dernier. Une influence de l'effet de primauté et récence suggère que ce processus intervient dans l'acquisition de l'organisation des séquences.

4) L'extraction de la structure séquentielle.

Par exemple, un enfant de 7 mois écoutant des séquences de syllabes telles que, par exemple « totobu », « mimiba », « paparo » acquiert, au bout de quelques minutes d'écoute, le fait que toutes ces séquences obéissent à la même structure « AAB ». Une extraction de l'organisation de la séquence, indépendamment de l'identité acoustique des syllabes pourrait donc avoir lieu.

5) L'extraction de la structure arborescente imbriquée.

Une idée importante est celle selon laquelle les mécanismes précédents ne suffiraient pas, à eux seuls, à expliquer l'acquisition du langage et notamment le fait que l'on puisse comprendre des phrases faisant appel à la récursivité.

Comme indiqué plus haut, une distinction souvent présente dans la littérature est celle faite entre deux types de grammaires : la grammaire à état fini (FSG) suivant la règle « $(AB)^n$ » et la grammaire structure phrasé ou *structure phrase* (FSP) « A^nB^n » (Fitch and Hauser 2004; Hauser et al. 2002). Cette distinction est faite par la possibilité d'avoir des dépendances adjacentes dans « A^nB^n » et non-adjacente « $(AB)^n$ », autrement dit on retrouve une structure en branche dans le FSP et dans la FSG, celle-ci est plutôt linéaire.

Maintenant, à partir de là, si l'on s'intéresse aux bases neurobiologiques du langage et plus particulièrement à celle de la syntaxe, de nombreuses questions se posent. On peut se demander quelles sont les régions impliquées dans chaque type de mécanismes

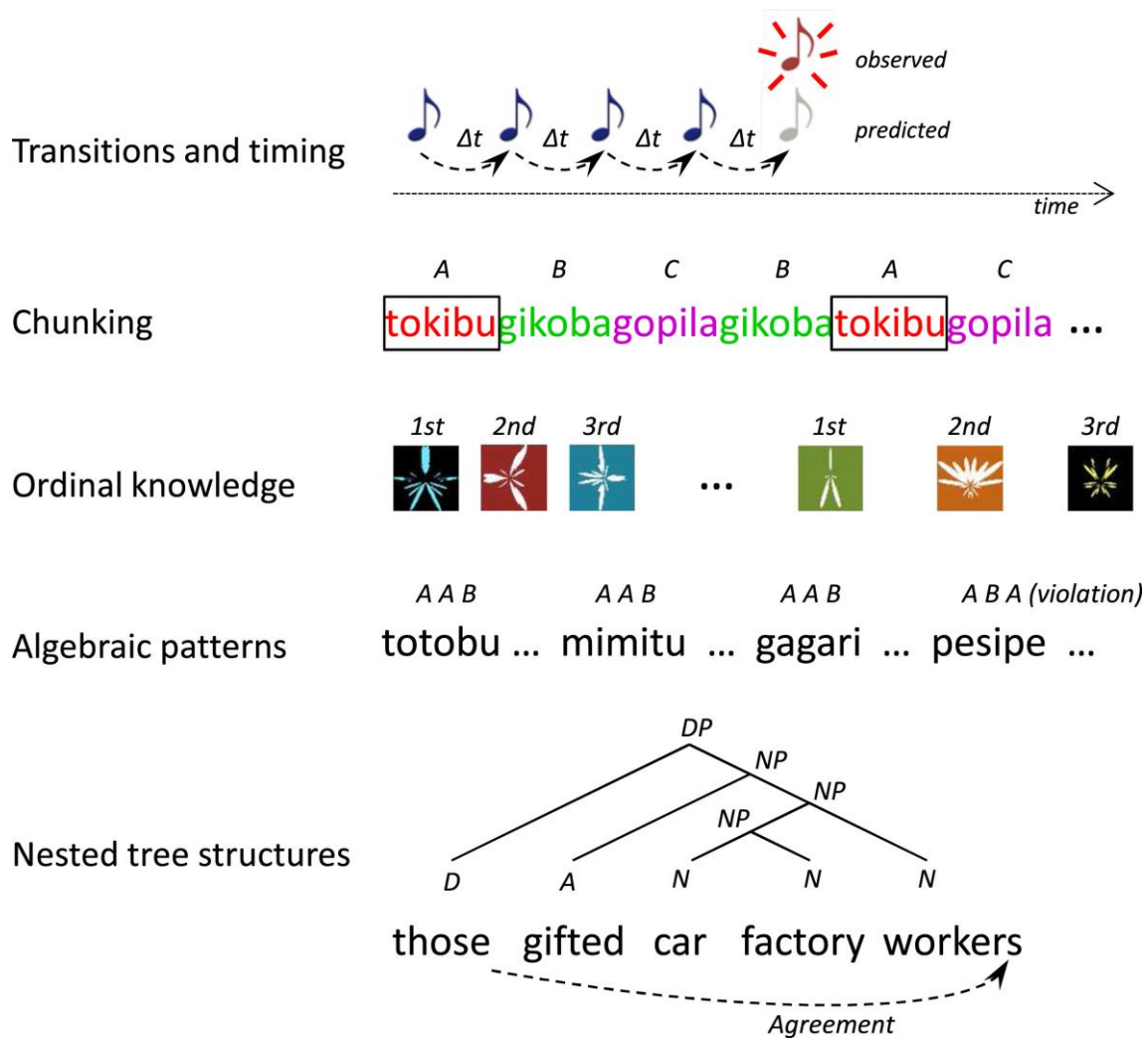


Figure 3 : Illustration des mécanismes de traitement de l'ordre.

Les séquences pourraient être traitées selon cinq niveaux possibles :

- (1) Les transitions entre chaque son (symbolisé par une note) sont espacées entre eux par un délai fixe (Δt).
- (2) Les séquences de *chunk* par exemple des mots dans une série de syllabes.
- (3) Une liste ordonnée avec une connaissance de l'ordre de chaque élément, qui est en premier, second, etc.
- (4) Un motif algébrique tel que AAB, indiquant que les deux premiers éléments sont identiques et le troisième différent.
- (5) Une structure arborescente, dont chaque séquence d'éléments est imbriquée dans d'autres séquences d'éléments comme observé dans la récursivité.

Figure extraite de Dehaene et al 2015

permettant un traitement de l'ordre des signaux vocaux, quels sont les rôles de ces différentes régions et quelles sont les propriétés des neurones au sein de ces régions. De nombreuses études ont abordé ces questions, en faisant appel à des techniques d'imagerie cérébrale (voir Dehaene et al. 2015), mais peu d'études en revanche se sont portées sur les propriétés des neurones impliqués dans ces traitements ou, en d'autres termes comment est codé l'ordre au niveau neuronal. Les modèles animaux permettent de soulever ce type de question, mais la question suivante qui se pose alors est quelle espèce animale choisir comme modèle d'étude et pour quelles raisons. Au cours des dernières décennies, l'étude du chant des oiseaux chanteurs ainsi que des régions cérébrales impliquées, qui sont pour certaines des analogues de régions cérébrales des mammifères, ont peu à peu amené à considérer les oiseaux chanteurs comme un remarquable modèle animal d'étude des mécanismes susceptibles d'être impliqués dans le langage.

IV. Les oiseaux chanteurs comme modèle d'étude des bases neurobiologiques du traitement des séquences de signaux acoustiques

Afin d'étudier le traitement des signaux acoustiques dans les différentes aires cérébrales, de nombreuses études ont été menées chez différentes espèces de mammifères. Cependant, le fait que la plupart de ces espèces n'organisent pas naturellement leurs vocalisations de façon à constituer des séquences peut être considéré comme une des limites de ces modèles animaux, si l'on s'intéresse au traitement de l'ordre dans les séquences. Parmi les primates non humains, les macaques sont capables de combiner leurs vocalisations de façon à constituer des séquences. La particularité de ces séquences serait d'avoir une signification (Arnold and Zuberbühler 2006). Des études se sont intéressées aux capacités de discrimination parmi des séquences organisées selon des règles différentes chez différentes espèces de primates non-humains ainsi qu'aux bases neurobiologiques de ces capacités à travers notamment des paradigmes identiques à ceux mis en place chez l'humain (Arnold and Zuberbühler 2006; Wilson et al. 2015; Wilson et al. 2017; Wilson et al. 2011).

Chez les vertébrés, outre les mammifères, les oiseaux chanteurs sont capables de produire des séries de signaux acoustiques complexes dont la particularité est d'être organisées selon des règles, qui sont propres à chaque espèce. Depuis plus d'une trentaine

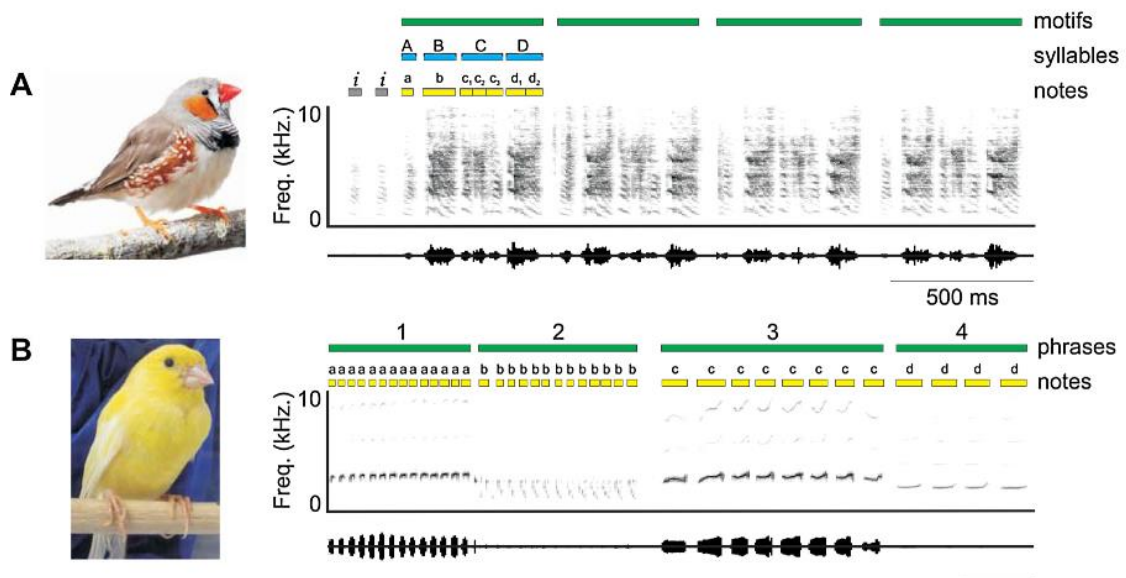


Figure 4: Chants de deux espèces d'oiseaux chanteurs.

Les chants des différentes espèces d'oiseaux chanteurs se différencient par les caractéristiques spectro-temporelles des notes, mais aussi par leur organisation. Les éléments des chants appelés des notes (en jaune) peuvent constituer des syllabes (en bleu). Les notes ou syllabes s'enchaînent pour former des séquences, appelées des motifs ou phrases (en vert). Le chant est composé d'une série de motifs ou de phrases. Ici sont présentons les chants de deux espèces (spectrogramme et oscillogramme), le diamant mandarin (haut) et le canari (bas).

(A) Le chant du diamant mandarin débute par des notes introductives, symbolisées par la lettre *i*, suivies d'un motif composé de différentes syllabes toujours produites selon le même ordre (ici notées A à D).

(B) Le chant des canaris se caractérise par une succession de phrases composées de la répétition d'une même syllabe, un nombre variable de fois. Chaque individu a un répertoire de phrases qui lui est propre et compose ses chants de façon variable à partir de ce répertoire.

Figure modifiée de Prather et al. 2017

d'années, des études de neurosciences cherchent à comprendre comment les chants comme séquences de différents signaux acoustiques, sont reconnus.

A. Le chant des oiseaux chanteurs : séquence de signaux acoustiques organisés temporellement

Les oiseaux chanteurs ou Oscines appartiennent au sous-ordre des Passériformes et regroupent plus de 4 000 espèces. Ils se distinguent des autres espèces d'oiseaux par le fait qu'ils produisent non seulement des cris (note isolée ou répétée sur une courte durée), mais aussi des chants.

Le chant est un enchaînement de signaux sonores plus ou moins complexes, appelés des syllabes (ou notes). La syllabe est l'unité de base du chant (Cynx, 1990) et se définit comme étant constitué d'une ou plusieurs notes, modulées ou non en fréquence, composées ou non d'harmoniques. Elles sont produites lors de l'expiration, rarement lors de l'inspiration (Goller and Suthers, 1996). Le type de syllabes en termes de paramètres acoustiques et la structure des chants varient d'une espèce à une autre (Prather et al. 2017) (**Figure 4**).

L'enchaînement des syllabes suit une organisation temporelle plus ou moins complexe suivant les espèces. Chez le diamant mandarin (*Taeniopygia guttata*) les syllabes s'enchaînent selon un ordre fixe et forment un motif, propre à l'individu, se répétant un nombre variable de fois (**Figure 4 haut**). Le canari dispose d'un large répertoire de syllabes à partir duquel il va créer différents chants. Une même syllabe se répète un nombre variable de fois pour former une phrase et c'est l'enchaînement des phrases qui constitue le chant (**Figure 4 bas**).

Les chants sont des signaux de communication jouant un rôle majeur dans la vie sociale des individus. Ces derniers peuvent intervenir dans des contextes de sélection du partenaire sexuel, essentiel à la survie de l'espèce, mais aussi à la défense du territoire, au comportement parental ou encore de coordination d'activités de groupe (Aubin and Jouventin 2002; Lengagne et al. 2000; Beecher 1991; Kroodsma and Byers, 1991; Nowicki and Searcy 2004).

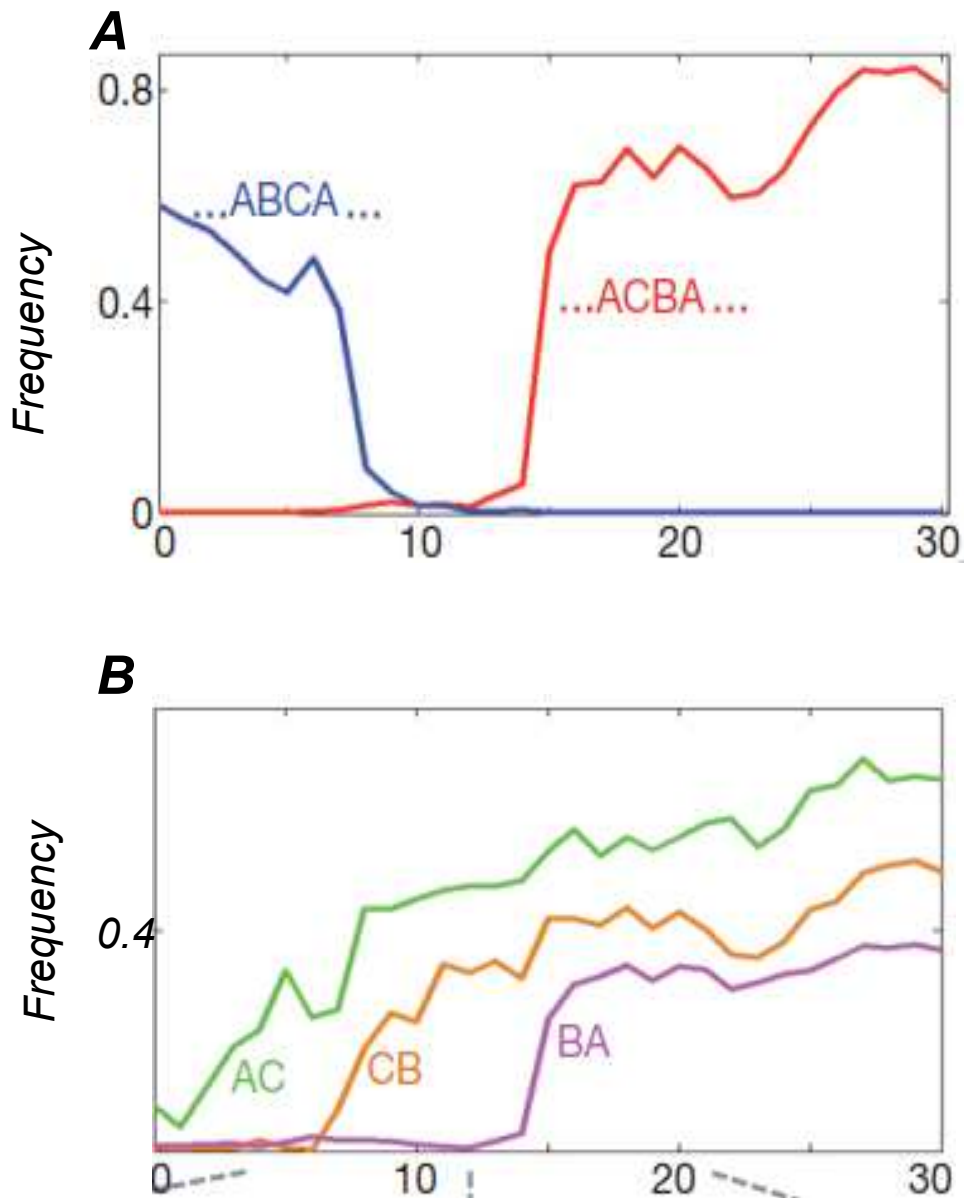


Figure 5 : Effet du changement de l'ordre des syllabes dans le chant du tuteur au cours de la période d'apprentissage.

Évolution au cours du temps (jours) de la fréquence de production a) des chants ABC et CBA et b) des trois enchaînements possibles AC, CB et BA du chant CBA.

Au cours de la période d'apprentissage, l'ordre des syllabes dans le chant du tuteur est modifié : le chant ABC (en bleu ; une lettre représente une syllabe) est remplacé par le chant CBA (en rouge). Les changements observés dans le chant des juvéniles montrent que les trois enchaînements possibles AC (en vert), CB (en orange) et BA (en violet) apparaissent par étape.

(Figure modifiée à partir de Lipkind et al 2013)

De par son répertoire de syllabes et l'organisation temporelle de celles-ci dans le chant, le chant permet d'identifier à la fois l'espèce, mais aussi la colonie et l'individu (Beecher 1994; Berwick et al. 2011; Hausberger 1997). Tout comme le langage, le chant nécessite un apprentissage social par imitation.

B. Apprentissage du chant des oiseaux

L'apprentissage du chant s'effectue selon deux phases, une phase sensorielle et une phase sensori-motrice qui peuvent ou non se chevaucher et se répéter tout au long de la vie en fonction de l'espèce considérée.

L'hypothèse la plus communément admise est que, lors de la phase sensorielle, le juvénile mémorise sous la forme d'un modèle de chant les chants entendus et produits par le ou les oiseaux de son entourage, qualifié(s) de tuteur(s). Ce modèle de chant sert de référence lors de la phase sensori-motrice. Durant cette seconde phase, le juvénile pratique son chant qui, d'une structure variable dans les premiers jours, se stabilise progressivement au cours du développement. À la fin de cette phase sensori-motrice atteinte à la maturité sexuelle, l'oiseau produit un chant dit cristallisé, typique de son espèce et similaire à celui de son tuteur. Après leur maturité sexuelle, on distingue deux catégories d'oiseau, *open-ended learner* et *closed-ended learner*. Les *open-ended learner* peuvent modifier une partie plus ou moins importante de leur chant et/ou répertoire chaque année. Au contraire, le répertoire des *closed-ended learner* ne montre plus de modifications ou très peu après l'âge adulte, il conserve alors le même chant toute leur vie. Mais, en fait, il existe un continuum entre ces deux catégories (Brenowitz and Beecher, 2005).

Le chant d'un juvénile est le résultat de l'imitation de celui de son tuteur. Son chant présente alors de nombreuses similitudes avec celui de son tuteur et de sa fratrie que ce soit sur le répertoire de syllabes, mais aussi sur l'organisation séquentielle de celles-ci (Doupe and Kuhl, 1999; Hausberger 1997; Griessmann and Naguib 2002).

Plusieurs études se sont penchées sur la question de l'apprentissage de l'organisation séquentielle présente dans les chants des individus d'une espèce donnée. Parmi ces espèces, le diamant mandarin est le plus fréquemment utilisé du fait de son unique chant. Plusieurs approches ont été suivies.

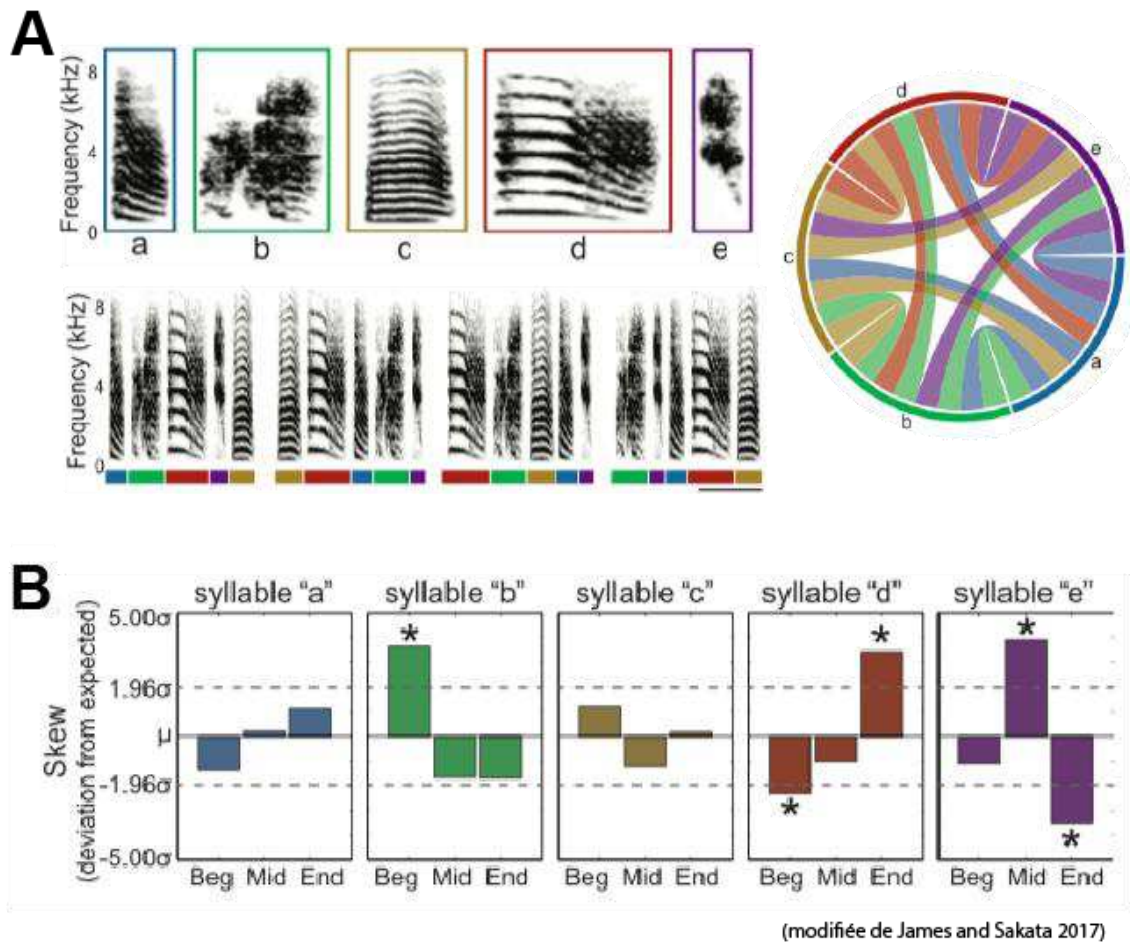


Figure 6: Position des différents types de syllabes dans les chants.

(A) Lors de l'apprentissage, les chants diffusés étaient composés à partir de 5 syllabes. La position des cinq syllabes variait d'un chant à l'autre. Exemples de chants (représentés par leur spectrogramme) composés à partir des 5 syllabes (a, b, c, d et e ; spectrogramme au-dessus). À droite : schéma pour indiquer que les 5 syllabes ont été agencées pour former les 120 motifs possibles. La barre noire représente 250ms, et chaque barre de couleur correspond à un type de syllabe.

(B) Position des 5 types de syllabes au début (Beg), au milieu (Mid) ou à la fin (End) des chants produits par les juvéniles soumis à l'apprentissage. La position a été évaluée par rapport à la position attendue : chaque syllabe aurait dû se trouver autant de fois au début, au milieu ou à la fin des chants. Les lignes pointillées indiquent la valeur +/- l'écart type ($p=0.05$) de la position attendue. Les oiseaux juvéniles placent plus souvent la syllabe « b » au début, la syllabe « e » au milieu, et la « d » à la fin des chants. De même, ils placent moins souvent la syllabe « d » au début et la syllabe « e » à la fin.

Lipkind et ses collaborateurs (Lipkind et al., 2013) ont abordé cette question en cherchant à déterminer si le changement de l'ordre des syllabes dans le chant qui sert de modèle d'apprentissage, suite à un premier apprentissage, s'accompagne de changements dans de trois syllabes, ABC, préalablement apprises par des diamants mandarins juvéniles en permutant les syllabes BC devenant ACB. Ils montrent que les oiseaux sont capables de modifier par étapes leur chant, introduisant tout d'abord la transition AC, puis CB et enfin plus tard la transition BA (*Figure 5*).

Une autre approche a consisté à se demander, si, soumis à n'importe quel agencement de syllabes, chaque oiseau produira un des agencements possibles. James et Sakata ont exposé des juvéniles à un chant de tuteur (noté ABCDE), dont les syllabes composant le chant ont été agencées selon toutes les manières possibles et présentées de manière équivalente (James and Sakata, 2017) (*Figure 6A*). Chacune des syllabes pouvant être à n'importe quelle place dans le chant, on pourrait supposer qu'au sein de la population, les juvéniles soumis à ces différentes écoutes auraient des chants cristallisés avec des agencements syllabiques très variables. Les résultats montrent au contraire un schéma similaire dans les chants appris par ces juvéniles. Malgré toutes les possibilités présentées, certains chants ne sont pas reproduits et encore plus surprenant certaines syllabes, A, C et E, se retrouvent significativement à la même place entre les différents chants, respectivement, début, milieu et fin du chant (*Figure 6B*). Il existe donc des règles intrinsèques à l'espèce.

Dans une autre étude, des diamants mandarin juvéniles ont été soumis à l'exposition de chants issus de deux tuteurs (Cate and Slater, 1991) notant que l'oiseau a tendance à copier des séquences de syllabes, autrement appelées *chunk*. Quand une syllabe est spécifique à un des deux tuteurs, alors la probabilité que la syllabe suivante soit aussi du même tuteur est élevée. Ainsi, les résultats indiquent que les diamants mandarins ont tendance à apprendre des *chunks* plutôt que des éléments isolés (Cate and Slater, 1991; Williams and Staples, 1992). L'ordre des syllabes est donc un paramètre pris en considération lors de l'apprentissage du chant.

Pour déterminer si le chant résulte d'un apprentissage de l'organisation séquentielle des syllabes, Rose et collaborateurs, (Rose et al., 2004) ont pris le problème autrement. Contrairement à l'étude de Lipkind et ses collaborateurs (2013) cherchant à

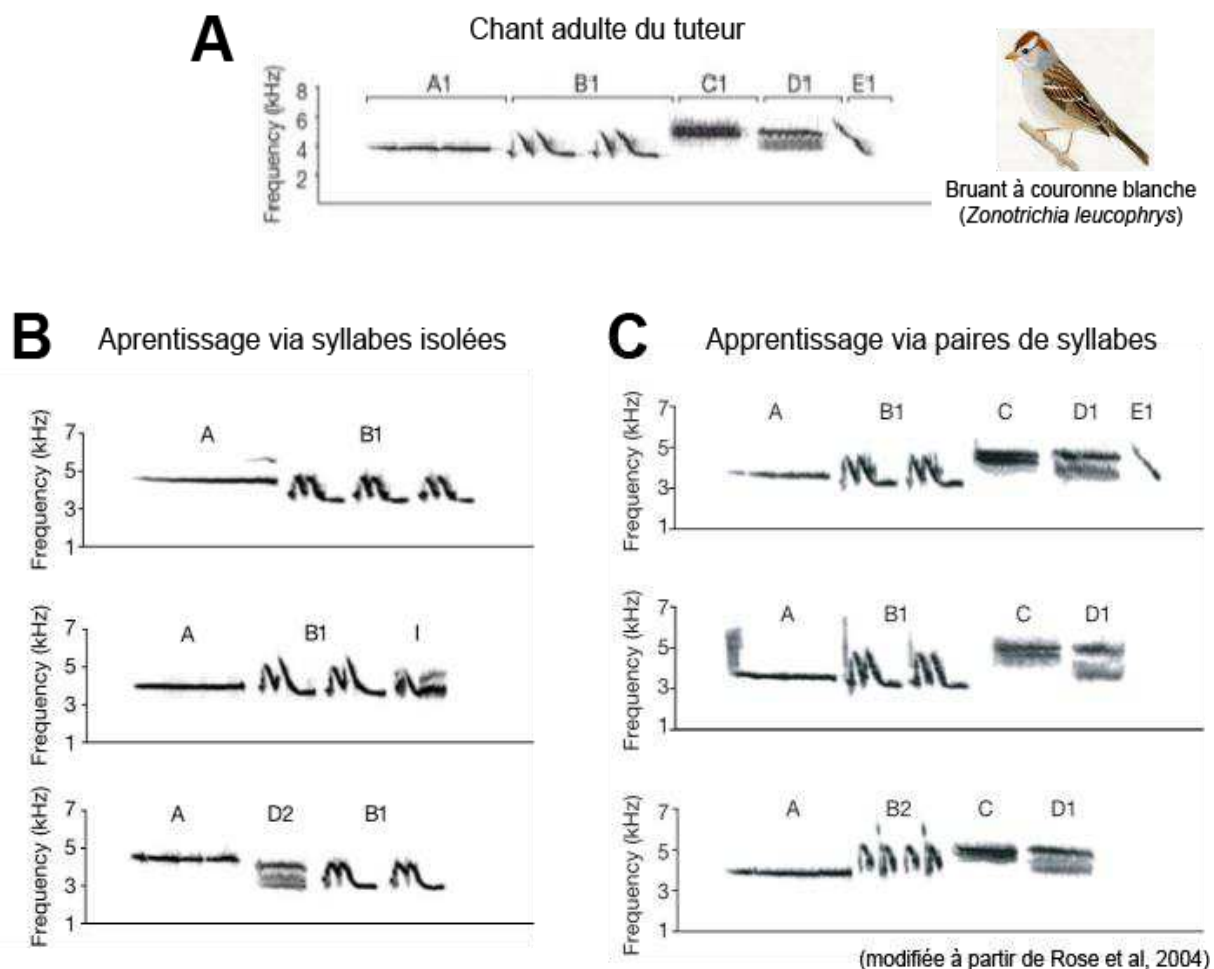


Figure 7 : Au cours de la période d'apprentissage, présenter des syllabes isolées ou des paires de syllabes influe sur la structure des chants.

Des jeunes bruants à couronne blanche sont soumis à un apprentissage au cours duquel il est présenté soit les syllabes du chant d'un adulte, de manière isolée soit les paires de syllabes de ce même chant.

(A) exemple d'un chant de l'adulte tuteur

(B) chants produits par les oiseaux soumis à l'écoute de syllabes isolées

(C) chants produits par les oiseaux soumis à l'écoute des paires de syllabes.

Seuls les juvéniles ayant entendu les paires de syllabes ont été capables de produire un chant ayant une structure représentative de l'espèce.

La syllabe I est considérée comme une improvisation et donc différente de ce que l'oiseau aurait dû produire.

observer une modification de l'agencement syllabique au sein du chant du juvénile, cette équipe a cherché à déterminer s'il est nécessaire que le juvénile soit confronté à des séquences de syllabes pour apprendre un chant. Cette étude a été conduite chez les bruants à couronne blanche (*Zonotrichia leucophrys*) dont le chant est composé de syllabes aisément catégorisables selon leurs caractéristiques spectro-temporelles (fréquence en fonction du temps) (**Figure 7A**). Leur chant se présente sous la forme A-B-C-D-E où chaque lettre correspond à un type de syllabe. Les chants nécessaires à l'apprentissage sont sous la forme de deux types de stimuli. Un groupe de juvéniles est soumis à la diffusion de syllabes espacées entre elles par de longs silences. L'autre groupe est soumis à la diffusion de paires de syllabes (par exemple AB, BC, CD, et DE), ces paires de syllabes étant séparées entre elles par de longs silences. Après apprentissage, les oiseaux du premier groupe, exposés à des syllabes isolées, ne sont pas capables de reproduire la structure ABCDE caractéristique du chant de leur espèce (**Figure 7B**). Les juvéniles du deuxième groupe, exposés à l'écoute de paires de syllabes, sont en capacité de produire un chant stéréotypé composé de la suite de syllabes ABCD. Il est intéressant de noter que l'apprentissage par paire amène une autre difficulté pour l'oiseau qui au lieu de produire cinq syllabes, se limite à la production des quatre premières syllabes, autrement dit deux paires, et donc omettra de produire la dernière (**Figure 7C**). Cela montre que l'oiseau doit être exposé à des séquences de syllabes pour apprendre correctement son chant et non à des syllabes isolées et que l'ordre influence l'apprentissage du chant. De plus, l'oiseau peut avoir des difficultés à produire des syllabes dites complexes, c'est-à-dire composées de deux éléments, comme les syllabes du groupe B, ces syllabes étant remplacées par des improvisations (notées I), autrement dit que l'on ne retrouve pas dans le répertoire du chant utilisé comme modèle. Ainsi, l'ordre impacte au-delà même de l'apprentissage de la structure globale du chant, mais aussi au sein même de syllabes complexes qui disposent d'une organisation temporelle fine de leur élément.

Le propre chant de l'oiseau (*Bird's own song* ou BOS) est intrinsèquement soumis à des règles syntaxiques imposées et apprises, laissant penser que l'agencement des syllabes est un marqueur important pour considérer le chant comme caractéristique de l'espèce. L'oiseau doit être confronté à une organisation temporelle pour produire un chant organisé en séquence typique de son espèce. L'ordre est un critère primordial dans

la mémorisation du chant modèle et *in fine* la production du propre chant de l'oiseau. Or il est reconnu que les oiseaux utilisent principalement les chants à but de reconnaissance interindividuelle (Mathevon and Aubin, 1997), on peut alors supposer que l'ordre des syllabes d'un chant est aussi un vecteur d'information.

C. L'ordre est porteur d'information

Le chant peut apporter de multiples informations sur l'espèce de manière générale, mais aussi sur l'individu au sein de celle-ci. Il peut révéler le sexe, l'âge, le statut reproducteur de l'individu tout comme son appartenance à une population locale (Hausberger 1997; Griessmann and Naguib 2002; Berwick et al. 2011; Briefer et al. 2009). Sachant cela, est-ce que la stratégie développée par l'oiseau pour détecter chacune de ces informations peut se baser sur la reconnaissance d'un congénère à partir de l'agencement de ses syllabes au sein du chant ?

C'est ce qui a été mis en avant notamment chez l'étourneau sansonnet (*Sturnus vulgaris*). Les étourneaux disposent d'un répertoire d'environ une dizaine de syllabes et peuvent créer différents chants. Gentner et Hulse ont entraîné des étourneaux à associer différents chants diffusés à leur émetteur (Gentner and Hulse, 2000). D'un chant à l'autre pour un même émetteur, les syllabes constitutives peuvent être ou non différentes. Une fois cet apprentissage fait, des chants des mêmes émetteurs, mais encore inconnus à l'oiseau ont été diffusés. Si l'ordre naturel des syllabes dans ces nouveaux chants est altéré, les performances d'association de l'oiseau entre le chant et l'émetteur sont significativement altérées. La reconnaissance de l'agencement des syllabes est donc une stratégie adoptée par l'étourneau pour reconnaître quel individu a produit quoi. De même chez le diamant mandarin mâle, *via* un test comportemental, Lawson et collaborateurs (Lawson et al., 2018) ont pu mettre en avant une discrimination entre les congénères familiers à l'oiseau sur la base de l'ordre entre le chant initial et ce même chant dont l'ordre a été modifié aléatoirement. La femelle du Vacher à tête brune (*Molothrus ater*) émet moins de réponses comportementales au chant de mâle si l'ordre a été modifié, supposant alors qu'elle dispose d'une représentation interne de l'ordre des phrases des chants conspécifiques (Ratcliffe and Weisman, 1986)

Une autre étude a visé à déterminer quel paramètre est le plus nécessaire à la discrimination des chants des congénères, les syllabes ou leur agencement (Holland et al.

2000). Lorsque les oiseaux ont recours naturellement au chant, l'environnement peut être bruyant et recouvrir une partie du chant. La question de Holland et collaborateurs est donc de déterminer sur quel paramètre, acoustique ou temporel, l'oiseau se base principalement pour initier sa réponse comportementale. Pour y répondre, différentes variantes, soit acoustiques (suppression des basses ou hautes fréquences), soit de l'ordre (organisation aléatoire des syllabes dans le chant), d'un même chant ont été diffusées. Lorsque l'un de ces deux paramètres est altéré, le comportement de l'oiseau est différent de celui adopté lors de la diffusion du chant naturel. De plus, on observe une distinction de comportement plus importante s'il y a modification de l'ordre, par rapport à celui lors d'une modification acoustique, le filtre passe-bas ayant la répercussion la plus minime d'entre tous sur le comportement des oiseaux. Au travers de ces études, on peut conclure que l'ordre est un critère important à la reconnaissance interindividuelle.

Les oiseaux chanteurs utilisent aussi leurs cris dans de nombreux contextes sociaux et il a été mis en évidence chez la mésange Japonaise (*Parus minor*) que l'information transmise par un individu varie suivant l'agencement des cris. Ces mésanges produisent un cri d'alerte, composé de trois notes (« ABC ») (**Figure 8A gauche**), signalant la présence d'un prédateur, les congénères scannent alors les environs pour repérer le prédateur et y échapper (Suzuki, 2014). Elles émettent aussi un cri de recrutement, composé d'une seule note (« D ») permettant le rassemblement des congénères autour de l'oiseau émetteur en vue d'aller collecter, par exemple, de la nourriture (**Figure 8A droite**) (Suzuki, 2014).

L'oiseau peut aussi combiner ces deux cri « ABC » + « D » afin d'associer ces deux comportements, le scan des environs et le regroupement, dans le but de déloger un prédateur (**Figure 8B gauche**). En inversant l'ordre « D » + « ABC », les réponses comportementales (nombre de scan (**Figure 8C gauche**) et approche du haut-parleur (**Figure 8C droite**)) sont réduites, démontrant ainsi l'importance de l'ordre pour transmettre une information décodable par les congénères (Suzuki et al. 2016). L'ordre

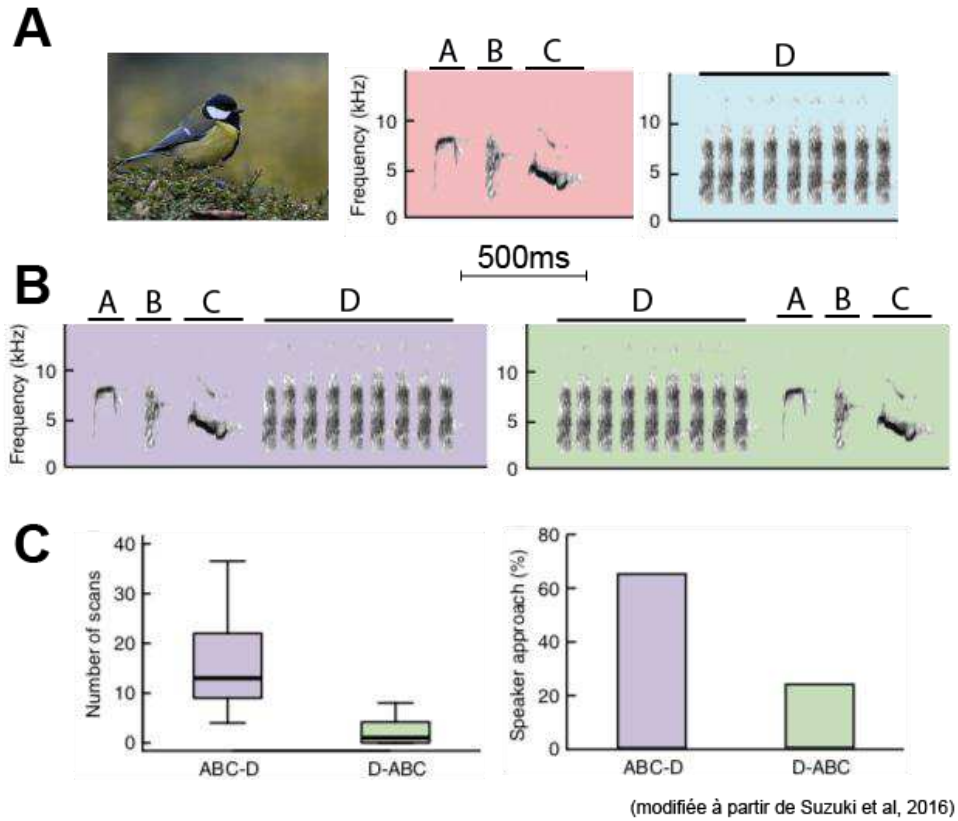


Figure 8: L'ordre est porteur d'information.

- (A) Spectrogrammes des deux cris ABC et D de la mésange (photo à gauche)
- (B) Il est possible de combiner ces deux cris en un seul ABC-D. Afin de déterminer si l'ordre dans cette séquence importe, ces deux cris ont été inversés dans une seconde séquence.
- (C) Présenter ces deux séquences n'induit pas les mêmes réponses comportementales. Suite à la diffusion de la séquence ABC-D, l'oiseau a un nombre de scans (balayages du regard dans son environnement) plus important (gauche) et s'approche plus souvent dans les deux mètres où se situe l'enceinte (droite) que lors de la diffusion de la séquence D-ABC. Il y a une discrimination comportementale de ces deux séquences sur la base de l'ordre des cris dans la séquence.

des signaux acoustiques joue ainsi un rôle d'importance pour la survie des espèces.

Lorsqu'un oiseau chante, le destinataire n'est pas toujours à côté de lui, le bruit environnant peut perturber l'écoute de chaque syllabe du chant dans les meilleures conditions. Or si une syllabe n'est pas entendue, la suivante sera toujours x secondes plus tard. Un des objectifs est alors de chercher à déterminer les mécanismes neurobiologiques mis en jeu dans le traitement de l'ordre des éléments constitutifs des chants chez les oiseaux chanteurs. Avant le début de ma thèse, seules quelques études s'étaient penchées sur cette question alors même que les oiseaux chanteurs représentent un modèle pertinent pour mener ce type d'études. Comme nous avons pu le voir, l'agencement des syllabes dans leur chant joue un rôle primordial pour la transmission d'informations. De plus, les oiseaux chanteurs présentent un réseau de structures cérébrales interconnectées et spécialisées pour la production, la perception et l'apprentissage du chant (Mello and Ribeiro, 1998; Petkov and Jarvis, 2012; Sutter and Margoliash, 1994).

V. Bases neurobiologiques du chant chez les Oiseaux chanteurs

Contrairement aux espèces d'oiseaux qui ne chantent pas et à un grand nombre d'espèces de mammifères, les oiseaux chanteurs possèdent au niveau du télencéphale un ensemble de régions spécialisées dans la perception, la production et l'apprentissage des signaux vocaux de communication (Margoliash et al. 1994 ; Petkov and Jarvis 2012 ; Feenders et al. 2008 ; Mello et al. 1998) (**Figure 9**).

On observe deux grands systèmes, le système auditif et le système des noyaux du chant. En ce qui concerne, le système auditif, comme chez les mammifères, il existe des connexions directes et indirectes reliant le noyau cochléaire aux noyaux du tronc cérébral (Theunissen and Shaevitz, 2006). Les informations convergent alors vers le noyau dorso latéral du mésencéphale, un analogue du Colliculus inférieur, qui projette sur le noyau ovoïdalis du thalamus, lui-même projetant sur le champ L, qui est considéré comme l'analogue du cortex auditif primaire. Le champ L se subdivise en plusieurs sous-régions (L1, L2a, L2b et L3). Les régions L1 et L3 envoient des projections sur deux aires analogues des aires auditives secondaires, le nidopallium caudomédian (NCM) et le mésopallium caudomédian (CMM). Enfin, les informations auditives arrivent au noyau

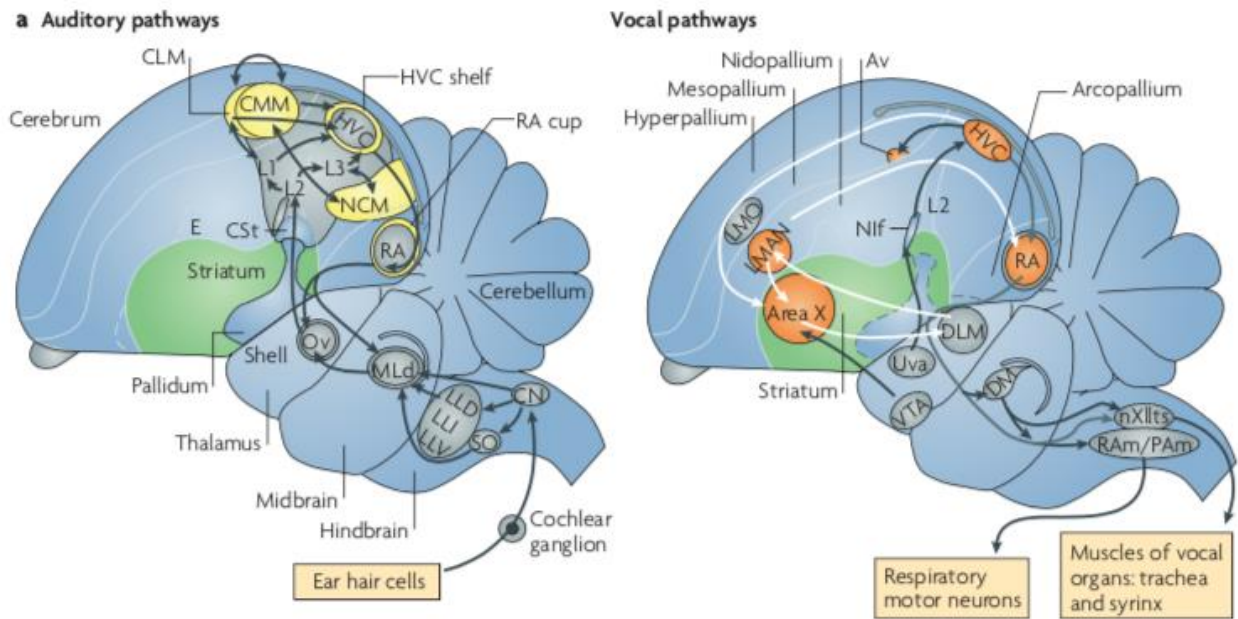


Figure 9 : Le cerveau des oiseaux chanteurs : régions et circuits spécialisés dans la perception, la production et l'apprentissage du chant.

Diagramme schématique de sections parasagittales du cerveau d'oiseau chanteur.

Les voies auditives (à gauche) indiquent le parcours des informations auditives depuis la cochlée jusqu'à l'HVC. En jaune sont représentées les régions montrant une forte augmentation de l'activité neuronale lorsque l'oiseau entend un chant : le nidopallium caudomédian (NCM) et mésopallium caudomédian (CMM) identifiés comme des analogues aux aires auditives secondaires des mammifères. Ces deux régions reçoivent des informations du champ L, subdivisé en trois parties et analogue au cortex auditif primaire des mammifères. Enfin le HVC (nom propre), un noyau sensori-moteur.

Les circuits (à droite) représentent les régions impliquées lors de la production et acquisition du chant de l'oiseau. Les flèches noires indiquent quelles sont les régions impliquées dans la production du chant tandis que les blanches présentent celles qui jouent un rôle dans l'acquisition du chant et le retour auditif. En orange sont représentées les régions actives lors de la production du chant : le noyau robustus de l'arcopallium (RA) qui est un noyau prémoteur, l'aire X qui est considérée comme analogue des ganglions de la base des mammifères et le noyau magnocellulaire latéral du nidopallium antérieur (LMAN).

Acronymes et abréviations:

Area X, Area X of the striatum; Av, avalanche; CLM, caudal lateral mesopallium; CN, cochlear nucleus; CST, caudal striatum; DLM, dorsal lateral nucleus of the medial thalamus; DM, dorsal medial nucleus of the thalamus; E, entopallium; L1, L2, L3, subdivisions of Field L; LLD, lateral lemniscus, dorsal nucleus; LLI, lateral lemniscus, intermediate nucleus; LLV, lateral lemniscus, ventral nucleus; LMAN, lateral magnocellular nucleus of the anterior nidopallium; LMO, lateral oval nucleus of the mesopallium; MLd, dorsal lateral nucleus of the mesencephalon; Nif, interfacial nucleus of the nidopallium; nXIIIts, tracheosyringeal portion of the nucleus hypoglossus (nucleus XII); Ov, ovoidalis; PAm, para-ambiguus; RA, robust nucleus of the arcopallium; RAmb, retroambiguus; SO, superior olive; Uva, nucleus uvulaeformis; VTA, ventral tegmental area.

HVC (nom propre), une région sensori-motrice qui est le point de départ de la voie motrice, qui passe par le noyau robustus de l'arcopallium (RA), un noyau prémoteur, et du circuit impliqué dans l'apprentissage du chant. L'aire X qui est considérée comme l'analogue des ganglions de la base des mammifères et le noyau magnocellulaire latéral du nidopallium antérieur (LMAN) sont engagés dans l'apprentissage du chant.

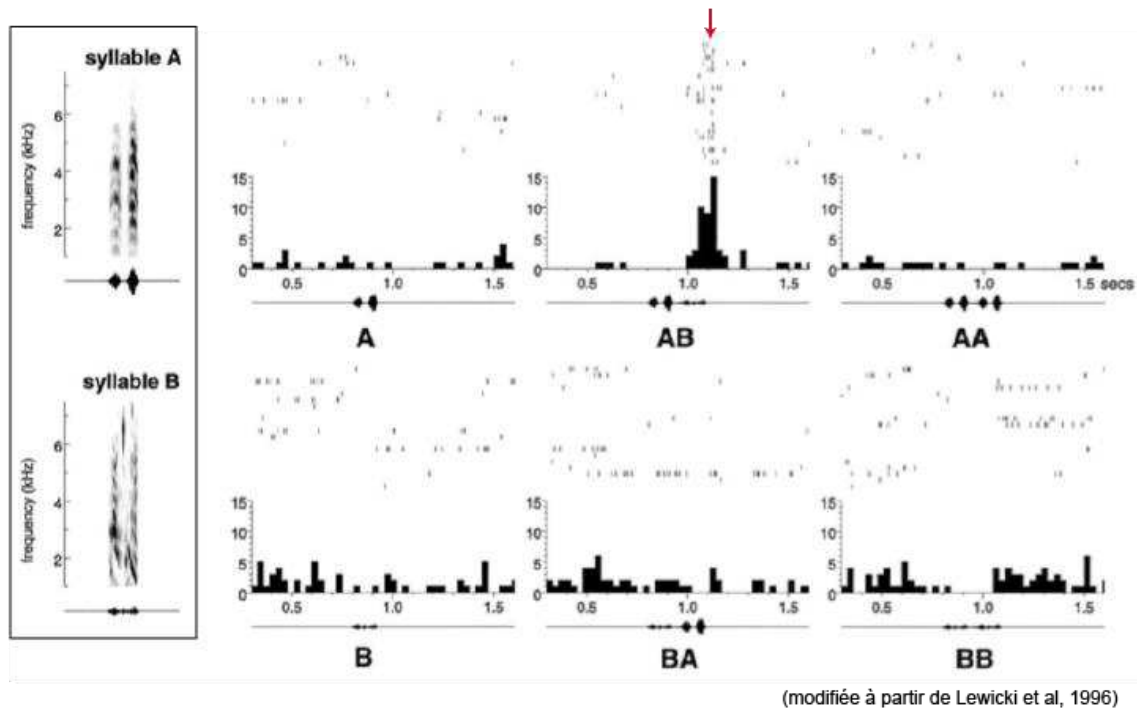
L'existence de telles régions permet de considérer les oiseaux chanteurs comme un modèle pertinent pour comprendre comment le système nerveux central intègre les informations sensorielles et motrices de façon à permettre aux individus de communiquer à travers des signaux vocaux.

A. Un noyau sensori-moteur, le HVC

Traitement auditif de l'ordre des signaux dans le chant

Le HVC (ancien acronyme désormais utilisé comme nom propre) est le point de départ des circuits impliqués dans la production et l'apprentissage du chant et reçoit des informations des régions auditives (**Figure 10**). Les toutes premières études conduites sur cette structure cérébrale ont permis d'établir le HVC comme une aire sensori-motrice. Une lésion bilatérale du HVC, chez l'adulte, entraîne une impossibilité pour l'oiseau de pouvoir chanter, démontrant le rôle crucial de cette structure dans la production du chant (Nottebohm et al. 1976). Quelques années plus tard sont publiées les premières données relatives aux propriétés auditives de neurones du HVC. Notons que le HVC contient des interneurones et trois types de neurones de projection (neurones projetant sur le noyau prémoteur RA (noyau robuste de l'arcopallium), sur l'aire X, aire considérée comme analogue des ganglions de la base des mammifères et sur le noyau Avalanche) et que ces premières études ont principalement enregistré des interneurones.

Les premiers résultats sont surprenants puisque d'emblée, il est observé une sélectivité des réponses auditives pour un type de stimulus donné, le propre chant de l'oiseau : lors de la diffusion du chant de l'oiseau (*bird's own song* ; BOS), le taux d'activité des neurones est plus important que lors de la diffusion du chant d'un congénère ou du chant d'un individu d'une autre espèce (Margoliash, 1986, 1983; Volman, 1993). Il a aussi été montré que changer l'ordre des syllabes dans le chant s'accompagne d'une diminution



(modifiée à partir de Lewicki et al, 1996)

Figure 10 : Neurone du HVC répondant sélectivement à la paire de syllabes AB.

(encadré à gauche) Spectrogrammes et oscillogrammes des deux syllabes, A et B présentées seules (A ou B), répétées (AA ou BB) ou associée en paire (AB ou BA) lors de l'enregistrement de l'activité d'un neurone du HVC de diamant mandarin.

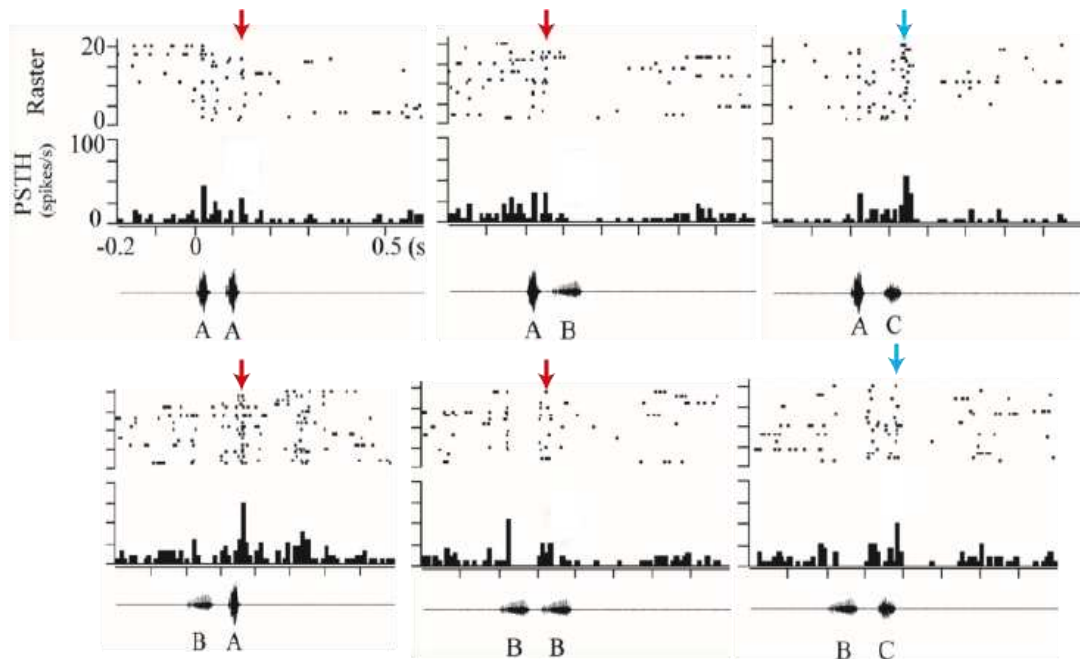
L'activité du neurone est représentée par un raster plot (chaque point représente un PA, chaque une ligne la répétition du stimulus) et un péri-stimulus histogramme (PSTH) (activité moyennée par classe de 50ms). L'oscillogramme du stimulus diffusé est présenté en dessous du PSTH.

La flèche rouge désigne l'activité lorsque le neurone répond significativement au stimulus.

On note que ce neurone répond fortement lors de la diffusion de la séquence AB. Si l'une des deux syllabes n'est pas diffusée ou si l'ordre des deux syllabes est inversé, le neurone ne répond pas.

des réponses neuronales (Margoliash 1983) : chez le bruant à couronne blanche, une espèce qui débute ses chants par deux sifflements, les enregistrements électrophysiologiques montrent que les neurones du HVC qui répondaient au 2^{ème} sifflement, ont une activité plus faible si on échange l'ordre de ces deux sifflements ou si on les diffuse séparément (Margoliash, 1983). Dans une étude ultérieure, conduite chez le diamant mandarin, il a été mis en évidence que l'activité des neurones du HVC varie suivant le nombre de syllabes du BOS présentées et leur ordre. Ces neurones s'activent plus fortement en réponse à deux syllabes qu'à chacune de ces deux syllabes diffusées séparément, ou encore à une syllabe entière qu'à cette même syllabe dont une portion est tronquée. Ces résultats suggèrent ainsi que les neurones du HVC expriment une intégration non linéaire des informations acoustiques sur un temps de plusieurs centaines de millisecondes (Margoliash and Fortune, 1992). Ces résultats ont été confortés par une étude ultérieure au cours de laquelle deux syllabes consécutives, notées A et B, ont été extraites des BOS de diamants mandarin (Lewicki and Arthur 1996). L'intensité des réponses neuronales à la diffusion de ces deux syllabes dans la séquence AB (présente dans le BOS) est plus forte que celle induite par la diffusion de ces deux syllabes séparément ainsi qu'à la somme de la réponse mesurée pour ces deux syllabes diffusées séparément, ou encore aux réponses induites par d'autres combinaisons de ces syllabes en AA, BB ou BA (**Figure 11**). Ces résultats ont renforcé l'hypothèse d'un codage en chaîne ou sériel du BOS où chaque neurone serait sensible à une syllabe ou à une séquence donnée du BOS (séquences AB, BC, CD, ABC, BCD, respectivement). Ainsi, lors de la diffusion du BOS, les neurones du HVC seraient successivement activés au fur et à mesure de l'enchaînement des différentes syllabes. En d'autres termes, la sélectivité des neurones du HVC pour le BOS reposerait sur une représentation du répertoire des éléments (syllabes et séquences de plusieurs syllabes) de ce chant au sein du HVC.

Beaucoup d'études se sont centrées sur les propriétés des neurones du HVC chez le diamant mandarin et le bruant à couronnes blanches dont les chants se composent d'un enchaînement linéaire de syllabes, sans varier d'un chant à l'autre. D'autres études plus récentes se sont intéressées à des espèces d'oiseaux chanteurs produisant des chants présentant une structure plus complexe que ceux de ces deux espèces. Ces études



(modifiée à partir de Nishikawa et al, 2008)

Figure 11: Exemple de neurone répondant lors de la diffusion de différentes paires de syllabes.

L'activité du neurone du HVC (chez le moineau du Japon) est représentée par un raster plot (un point est un PA et une ligne, une répétition du stimulus) et un péri-stimulus histogramme (PSTH) (activité moyennée par classe de 10ms). L'oscillogramme de la séquence de deux syllabes est placé en dessous du PSTH.

Les flèches rouges pointent l'activité du neurone pendant la diffusion de la syllabe A lorsqu'elle est précédée par une syllabe (A ou B), suivie par B, ou encore lorsqu'elle n'est plus diffusée après B.

Les flèches bleues pointent l'activité pendant l'activité du neurone pendant la diffusion de la syllabe C lorsqu'elle est précédée par une syllabe (A ou B).

On observe que la force des réponses d'un neurone à une syllabe dépend de ce qui précède et de sa place dans le chant

ont confirmé la sensibilité des neurones du HVC à l'ordre des syllabes chez le moineau du Japon, *Lonchura striata domestica*, (Nakamura and Okanoya, 2004), chez le bruant chanteur, *Melospiza melodia*, (Nealen and Schmidt, 2006), chez le canari (Lehongre and Del Negro 2011) : changer l'ordre des syllabes entraîne une baisse des réponses des neurones du HVC. Ce que suggèrent également ces études est le fait que le codage du BOS dans le HVC ne repose pas sur une représentation simple du répertoire de l'oiseau, mais plutôt sur un codage distribué : un même neurone peut être actif lors de la diffusion de différents types de chants (Nealen and Schmidt 2006; Nishikawa et al. 2008; Lehongre and Del Negro 2011). Par exemple, chez le moineau du Japon, la présentation d'une multitude de séquences composées à partir de deux syllabes chantées par l'oiseau lui-même a permis de mettre en évidence que des neurones peuvent être sensibles à une syllabe donnée indépendamment de la séquence dans laquelle elle est incluse (Nishikawa et al. 2008)(*Figure 11*).

B. Codage des probabilités de transition dans le HVC

Deux catégories d'oiseaux chanteurs peuvent être distinguées selon le degré de variabilité de l'ordre des syllabes constitutives des chants. Nous avons vu que certaines espèces telles que le diamant mandarin produisent toujours le même enchaînement de syllabes, alors que d'autres, telles que le moineau du Japon ou le canari, modifient ces enchaînements d'un chant à un autre. Pour autant, chez ces dernières espèces, l'ordre des syllabes constitutives des chants n'est pas aléatoire. Les probabilités de transition entre les syllabes des chants permettent de rendre compte de cette variabilité dans l'enchaînement des syllabes. Chez le diamant mandarin, l'ordre des syllabes restant fixe d'un chant à un autre, les probabilités de transition sont de 100% : la syllabe A précède la syllabe B dans 100% des cas, la syllabe précède la syllabe C dans 100% des cas, *etc.* Chez le canari en revanche, l'ordre des syllabes variant d'un chant à un autre, les probabilités de transition entre deux syllabes pourront être inférieures à 100%. De plus, deux types de probabilité de transitions peuvent être calculés : les probabilités de convergence et les probabilités de divergence. La probabilité de divergence pour la transition entre les syllabes « x-y » reflète la probabilité d'avoir « x » suivie par « y » (**A gauche**). La probabilité de convergence, à l'inverse, reflète la probabilité que « y » soit précédée

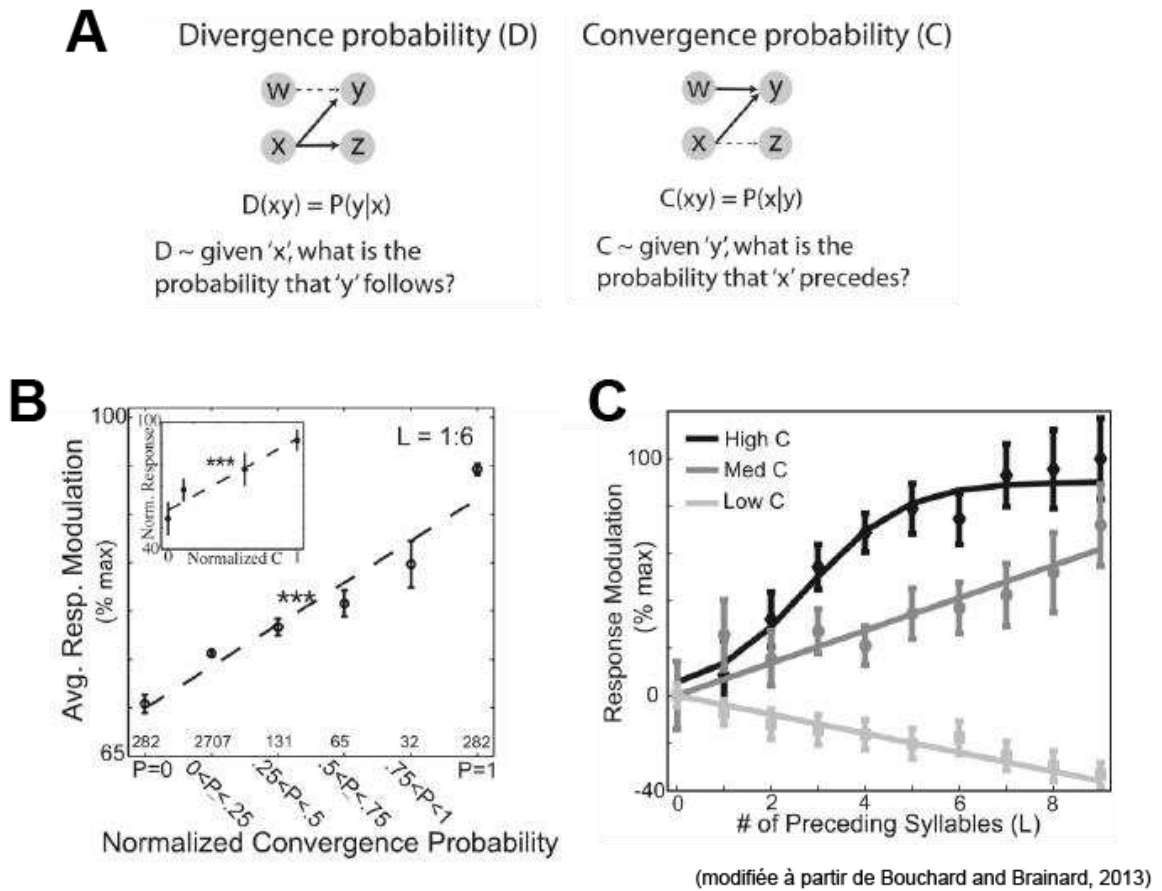


Figure 12: La force des réponses dépend des probabilités de convergence.

(A) Pour une transition donnée (xy) la probabilité de divergence (gauche) est la probabilité d'avoir « x » suivi par « y » et la probabilité de convergence (droite) est la probabilité d'avoir « y » précédé par « x ».

(B) Corrélation positive entre les probabilités de convergence et la force des réponses à la phrase qui suit (y pour la séquence xy) sur l'ensemble de la population neuronale du HVC enregistré (exemple de corrélation pour un neurone). La force des réponses est calculée à partir de la réponse maximale obtenue (% max).

(C) Force des réponses en fonction de la longueur de la séquence qui précède la syllabe (y pour xy, txy ou stxy, etc.) Trois catégories de force des probabilités ont été considérées : forte (High), moyenne (Med) et faible (Low).

La longueur de la séquence influe sur la force des réponses de manière différente selon les trois catégories établies.

par « x » (**A droite**). Il est important de noter qu'elles peuvent différer pour la même transition.

Une étude s'est penchée sur la question de l'influence de ces probabilités de transition dans le codage du propre chant de l'oiseau dans le HVC. Cette étude réalisée chez le moineau du Japon suggère que les réponses des neurones du HVC seraient corrélées avec les probabilités de convergence (Bouchard and Brainard, 2013). Dans cette étude, les stimuli utilisés lors de l'enregistrement de l'activité des neurones du HVC étaient non seulement le BOS, mais aussi un très grand nombre de séquences composées à partir des syllabes du BOS. Ces séquences variaient par leur longueur, *i.e.* nombre de syllabes, et leurs probabilités de convergence et de divergence calculées par rapport à la dernière syllabe de la séquence. Les résultats montrent une corrélation positive entre l'activité neuronale et les probabilités de convergence (**Figure 12 B**) : en normalisant le taux d'activité sur chaque site, pour une syllabe donnée et pour une séquence donnée, plus la probabilité de convergence vers cette syllabe ou séquence est forte, plus le taux d'activité neuronale est élevé. La longueur de la séquence précédant la syllabe cible peut impacter aussi sur la force des réponses neuronales à cette dite syllabe (**Figure 12 C**). Cette propriété des neurones du HVC pourrait contribuer à la discrimination des chants (Gentner and Hulse, 2000). Cependant, les probabilités de convergence n'expliqueraient qu'environ 25% de la variabilité des réponses neuronales, cela implique alors que d'autres paramètres jouent un rôle essentiel dans ce codage neuronal.

C. Codage moteur de l'organisation temporelle du chant

Les toutes premières études lésionnelles ont mis en évidence le rôle crucial joué par le HVC dans la production du chant (Nottebohm et al. 1976). Par la suite, les études électrophysiologiques ont mis en évidence le codage prémoteur des neurones du HVC (par exemple (McCasland 1987; McCasland and Konishi 1981; Hahnloser et al. 2002). Chez l'adulte, lorsque l'oiseau chante, les interneurones déchargent de façon soutenue tout au long du chant alors que les neurones de projection produisent de brèves bouffées de potentiels d'action, appelées « *bursts* », à un moment très précis du chant (Hahnloser et al. 2002). D'une syllabe à la suivante, les neurones mis en jeu diffèrent. L'hypothèse qui a été proposée suggère que la commande motrice du chant repose sur la mise en jeu successive et continue selon une certaine séquence de l'ensemble des neurones de

projection et notamment de ceux qui projettent vers RA (Hahnloser et al. 2002; Prather et al. 2008; Fujimoto et al. 2011; Katlowitz et al. 2018; Long et al. 2010). En faveur de cette idée, il a été montré que l'activité de 384 neurones de projection couvre 95% de la durée du chant (Lynch et al., 2016). Selon cette conception, le HVC jouerait le rôle « d'horloge », c'est-à-dire interviendrait principalement dans l'organisation temporelle du chant. Cette hypothèse est renforcée par le fait que le refroidissement du HVC induit une modification de la temporalité du chant, sans impacter les aspects spectraux (Long and Fee, 2008). Il faut toutefois mentionner le fait qu'une autre hypothèse a été proposée. Selon cette hypothèse, les neurones du HVC présenteraient une synchronisation de leur activité à certains moments essentiels du chant pour le contrôle des mécanismes mis en jeu au niveau périphérique (Amador et al., 2013). L'hypothèse étant que deux actes moteurs, ici

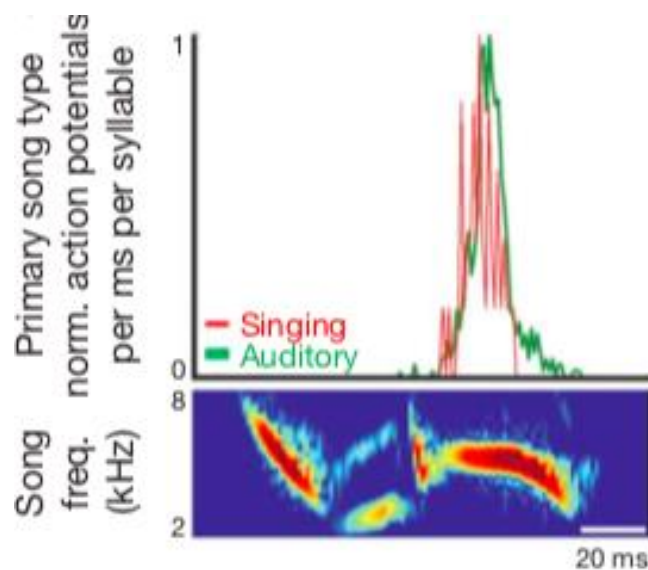


Figure 13: Activité audio-motrice d'un « neurone-miroir ».

Péristimulus histogramme représentant l'activité d'un neurone qui projette à l'aire X (HVCx) lors de l'écoute d'une note du chant (vert) ou chantée par l'oiseau (rouge). Le spectrogramme de la note est placé en-dessous.

(figure de Prather et al. 2008)

la production de deux syllabes distinctes, peut découler d'une seule commande motrice de par une continuité dans leurs paramètres physiques (pression et tension).

En ce qui concerne les neurones qui projettent vers l'Aire X (HVCx), des études ont montré qu'ils peuvent posséder les propriétés de neurones miroir (Prather et al. 2008) : chez le bruant des marais, l'activité auditive des neurones HVCx est un corollaire de l'activité motrice du chant (**Figure 13**). Lors de la production d'un chant, l'activité sensorielle des neurones est stoppée. L'activité auditive des neurones HVCx est ainsi liée à la perception des chants par l'oiseau.

De même, Vallentin et collaborateurs (2016) ont enregistré des interneurones du HVC rappelant une activité miroir observée chez le primate non-humain (Rizzolatti and Arbib, 1998). L'expérience consiste à modifier le premier modèle appris de chant AAAA (A étant une syllabe) des diamants juvéniles en introduisant une nouvelle syllabe, B et une nouvelle structure ABAB (alternance de syllabe) (**Figure 14**). Ce nouvel apprentissage se fait en deux étapes, la première où A est convenablement produite par l'oiseau au contraire de B. Lors de la deuxième plus tardive, les deux syllabes sont correctement produites par l'oiseau. Durant la première phase, les interneurones déchargent préférentiellement pendant la syllabe précédemment apprise (ici la syllabe A). La dernière phase montre de même un taux de décharge équivalent, mais cette fois-ci réparti sur les deux types de syllabes, A et B. Au contraire les courants excitateurs eux ne changent pas tout au long des différentes phases d'apprentissage. Cette inhibition a pour but de supprimer ainsi l'effet des entrées sensorielles sur les neurones prémoteurs pendant ces périodes, empêchant une plasticité accrue de la sortie motrice et permettant à l'oiseau d'avoir un chant stéréotypé. De ce fait, à travers cette étude, on peut supposer qu'altérer l'ordre des syllabes, en créant notamment de nouvelles transitions comme dans cette étude avec la transition AB, pourrait avoir un impact majeur sur les courants inhibiteurs plutôt qu'excitateur. L'ensemble de ces hypothèses vient de travaux réalisés chez deux espèces dont les chants sont relativement simples, le diamant mandarin et le bruant des marais. Peu d'études se sont intéressées au traitement de l'ordre dans le HVC, chez des espèces dont les chants ont des probabilités de transition entre les syllabes variables.

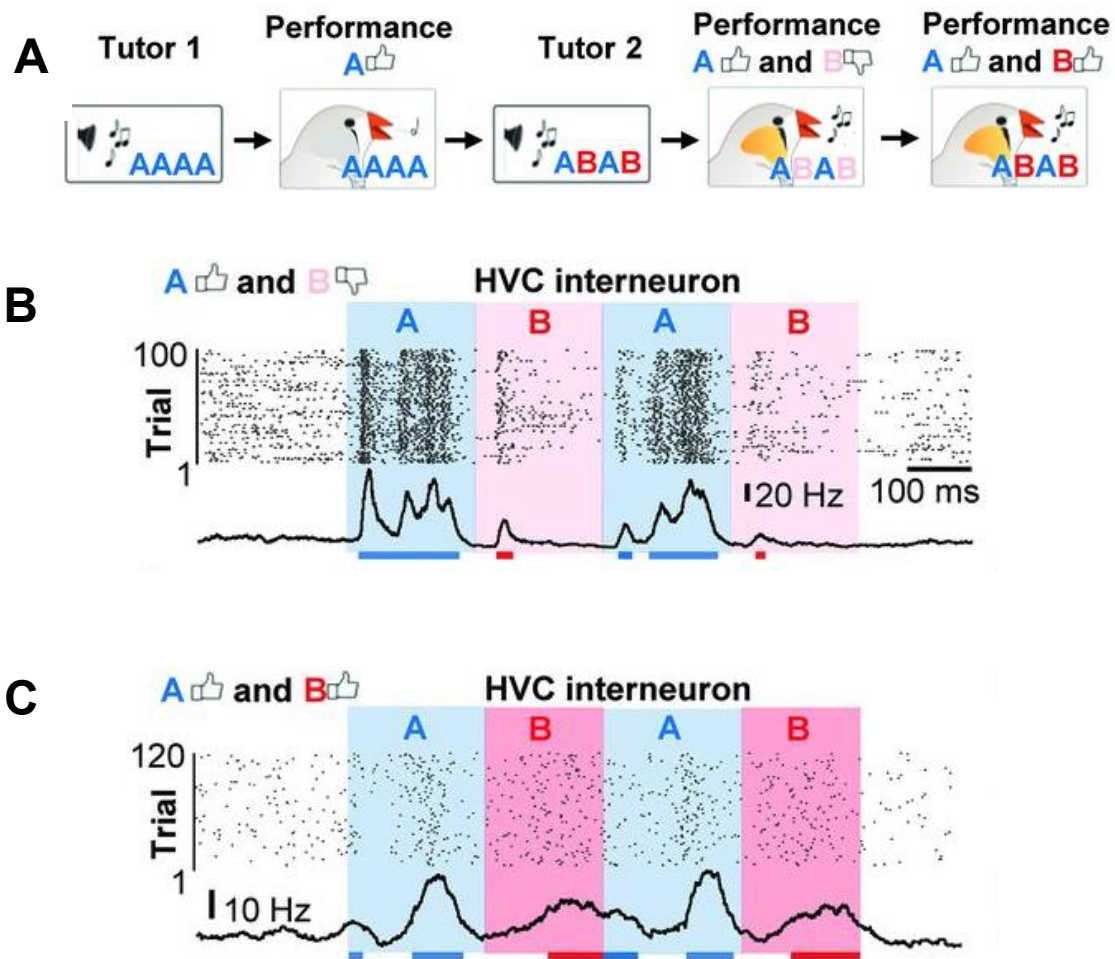


Figure 14: Apprentissage d'une nouvelle séquence comportant une nouvelle syllabe en deux étapes

(A) Schématisation de l'apprentissage en deux étapes de la séquence ABAB suite à un premier apprentissage d'une séquence AAAA.

(B) Raster plot présentant l'activité des interneurones durant la présentation de ABAB lorsque le juvénile produit une bonne copie du A et une mauvaise copie du B.

(C) Raster plot présentant l'activité des interneurones durant la présentation de ABAB lorsque le juvénile produit une bonne copie du A et une bonne copie du B

Sous les rasters : somme des potentiels d'action produits par le neurone. Les barres bleues et rouges horizontales représentent la période pendant laquelle le taux d'activité est significatif par rapport au niveau d'activité de base. On note que dans la première étape, les interneurones déchargent principalement lors de la diffusion du A, tandis que lorsque l'oiseau a copié correctement A et B, on ne distingue plus cette activation préférentielle.

Des études ont alors été réalisées chez le moineau du Japon (Fujimoto et al. 2011). Comme observé chez le diamant mandarin, lorsque l'oiseau chante, certains neurones qui projettent à l'aire X sont actifs avant la production d'une syllabe donnée (indépendamment de ce qui précède et suit). Mais d'autres ont une activité plus complexe : certains ne sont actifs que lors d'une transition donnée (ex : un neurone s'active lors de la transition F->B, mais pas lors de la transition F->A), d'autres sont actifs lors d'une certaine transition et en fonction de la production d'une troisième syllabe (le neurone s'active plus lors de la transition F->A si A est suivie de la syllabe B que lorsque A est suivie de la syllabe I). Cette étude présente un intérêt supplémentaire du fait que l'activité des neurones qui projettent à l'Aire X et qui sont présumés être potentiellement des neurones miroirs, a été non seulement enregistrée lorsque l'oiseau chantait, mais aussi lors de la diffusion du BOS. Pour un tiers des neurones, les changements d'activité observés lors de la diffusion du chant ont lieu environ au même moment que lorsque l'oiseau chante, c'est à dire, avant la diffusion d'une certaine syllabe ou au moment d'une certaine transition. Cela peut apparaître comme surprenant, étant donné que la syllabe attendue n'a pas encore été diffusée, mais se conçoit si l'on tient compte du fait que ces neurones peuvent avoir des propriétés de neurones miroirs.

Ces résultats suggèrent que le traitement de l'ordre mis en avant lors de la production, ayant notamment pour but le maintien du BOS, soit similaire à celui de lors de la perception du BOS. De ce fait, considérant les propriétés miroir des neurones de projection à l'aire X et interneurones, l'activité de ces neurones lors de la diffusion du BOS pourrait aussi dépendre du contexte dans lequel se situe la syllabe, autrement dit ce qui précède et suit, se basant sur des mécanismes de prédiction.

VI. Le système auditif des oiseaux chanteurs

Nous avons vu que les neurones du HVC jouent un rôle central dans le traitement des informations auditives. Le HVC est la cible directe et indirecte d'un ensemble de structures auditives. Il est alors légitime de s'interroger sur le fait que ces propriétés exprimées par les neurones du HVC pourraient déjà être présentes dans les structures cérébrales afférentes.

A. Aire auditive primaire

Chez l'oiseau chanteur, le champ L est considéré comme l'analogue du cortex auditif primaire des mammifères parce qu'il reçoit des informations de l'analogue du thalamus auditif, le noyau ovoïdalis. Cette aire est subdivisée en plusieurs régions d'entrée (L2a et le L2b) et de sortie (L1 et L3). Si comme chez les mammifères (Reale and Imig, 1980), il a été mis en évidence une organisation tonotopique au sein du champ L (Rübsamen and Dürscheidt 1986), les propriétés auditives des neurones varient entre les différentes sous-régions (Sen et al. 2001). Des neurones sont sensibles à une fréquence donnée, ou à certaines modulations d'amplitude de fréquences surtout au niveau du L2a et L2b, mais d'autres neurones peuvent présenter des réponses plus complexes qui ne peuvent pas se prédire sur la base de certains paramètres acoustiques (Sen et al. 2001). Lorsque la sélectivité des neurones pour le BOS a été mise en évidence dans le HVC, la question qui s'est posée a été de savoir si cette sélectivité n'a pas déjà émergé dans les régions auditives en amont. Des enregistrements dans le champ L ont montré que les neurones de cette région sont bien moins sensibles que ceux du HVC à des manipulations de l'organisation temporelle du chant (Lewicki and Arthur 1996; Amin et al. 2004). Inverser le chant ou seulement l'ordre des syllabes influe peu sur l'activité neuronale de la plupart des neurones du champ L, contrairement à ce qui est observé dans le HVC (Lewicki and Arthur 1996; Amin et al. 2004). Cela suggère que le codage des signaux acoustiques au niveau du champ L repose sur les paramètres acoustiques sans tenir compte du contexte dans lequel ces paramètres se trouvent, c'est-à-dire indépendamment de ce qui précède. Ces conclusions font écho à celles mises en avant chez les mammifères (Chechik and Nelken 2012; Lemus et al. 2009; Sloas et al. 2016). Pour changer l'activité neuronale, il faut modifier les caractéristiques spectrotemporelles au sein d'un élément acoustique, c'est-à-dire ici en diffusant par exemple la syllabe à l'envers, comme démontré chez le ouistiti en présentant un cri puis ce même cri à l'envers (Nagarajan et al., 2002; Wang et al., 1995).

B. Aire auditive secondaire

Entre le champ L et le noyau HVC se situent différentes régions auditives considérées comme de haut niveau, le Nidopallium Caudo-Médian (NCM), le Mésopallium Caudo-Médian (CMM) et le Mésopallium Caudo-Latéral (CLM).

Sélectivité pour les chants de l'espèce

Le Nidopallium Caudo-Médian ou NCM est la région qui a suscité le plus d'intérêt. Bien que des sons purs puissent induire des réponses auditives dans cette région (Müller and Leppelsack 1985), les neurones présentent des propriétés plus complexes que celles des neurones du Champ L. Dès la première étude ayant réalisé des enregistrements électrophysiologiques dans le NCM, il a été mis en évidence qu'un plus grand nombre de neurones répondaient plus fortement à des stimuli naturels (chants de congénères) qu'à des stimuli artificiels (sons purs, sons dans une large bande de fréquence) dans le NCM que dans le Champ L (Müller and Leppelsack 1985). Par la suite, ce qu'ont pointé de nombreuses études, basées sur l'enregistrement de l'activité des neurones ou sur l'activation de gènes immédiats précoces, comme le gène codant pour la protéine *Zenk* (acronyme des gènes précoces *zif268*, *Egr-1*, *NGF-1A*, and *Krox-24*) ou *cFos* est le fait que les neurones du NCM sont plus actifs lors de la diffusion de chants de l'espèce (consécifiques) que lors de la diffusion de ceux d'autres espèces (hétérosécifique) ou d'autres stimuli auditifs (du bruit blanc, son qui comporte toutes les fréquences du spectre (20Hz-20 kHz) ou un son pur) (Mello et al. 1992; Chew et al. 1996; Stripling et al. 1997; Bailey et al. 2002).

Effet de la familiarité

Les neurones du NCM ne montrent pas seulement une préférence pour les chants de l'espèce, ils peuvent également exprimer des réponses auditives qui montrent une discrimination entre des vocalisations (chants ou cris) familiers et non-familiers. Ils répondent différemment à des chants mémorisés suite à un conditionnement instrumental (tâche Go/no-Go ; **Figure 15**) (Schneider and Woolley, 2013; Thompson and Gentner, 2010) ou non (Menardy et al. 2012; 2014; Mouterde et al. 2017), par rapport à des chants qui n'ont jamais été présentés ou à des vocalisations d'individus familiers. Ces individus familiers ont par exemple partagé le contexte expérimental avec les sujets de l'étude pendant une longue période par rapport à celles d'individus qui n'avaient jamais été entendus (Menardy et al. 2012; 2014).

Dès les premières études portant sur le NCM, il a été montré sur la base de l'analyse de l'expression du gène *Zenk*, que la présentation répétée d'un même chant ou d'un même cri induisait une diminution de l'expression de ce gène (Mello et al. 1995). Cette

diminution peut s'observer jusqu'à 24h après la dernière diffusion du stimulus, ce qui indique la durabilité de cet effet (Mello et al. 1995). Elle n'est pas due à un « épuisement » des neurones : si un autre chant ou un autre cri, différent du premier, est diffusé, il est observé un regain d'activité à un niveau similaire à celui obtenu lors des premières diffusions du premier stimulus. Ce regain est qualifié de rupture d'habituation (Mello et al. 1995). Par la suite, des enregistrements électrophysiologiques ont également montré une diminution de l'activité neuronale au cours des présentations successives d'un chant conspécifique (Stripling et al. 1997; Stripling et al. 2001; Chew et al. 1995; Chew et al. 1996). Cette adaptation spécifique au stimulus, qualifiée d'habituation, s'observe dans certaines régions du système auditif des mammifères (Pérez-González et al. 2005), mais au sein du NCM, elle s'observe plus particulièrement pour les stimuli qui font répondre les neurones, c'est-à-dire les chants ou cris conspécifiques (Stripling et al. 1997). De nombreuses études ont exploité ce phénomène d'adaptation spécifique à un stimulus présenté de façon répétée et du regain d'activité faisant suite à la présentation d'un nouveau stimulus pour mettre en évidence une discrimination entre deux stimuli par le NCM (Chew et al. 1996; Smulders and Jarvis 2013; Lu and Vicario 2014; 2017). Le présumé est que s'il y a une rupture d'habituation lors de la diffusion d'un nouveau stimulus, les deux stimuli sont considérés comme différents l'un de l'autre. Il a ainsi pu être mis en avant une discrimination entre les chants de deux individus, entre le chant présenté à l'endroit et ce même chant présenté à l'envers (Chew et al. 1996) ou encore entre deux localisations pour un même chant (Smulders and Jarvis, 2013).

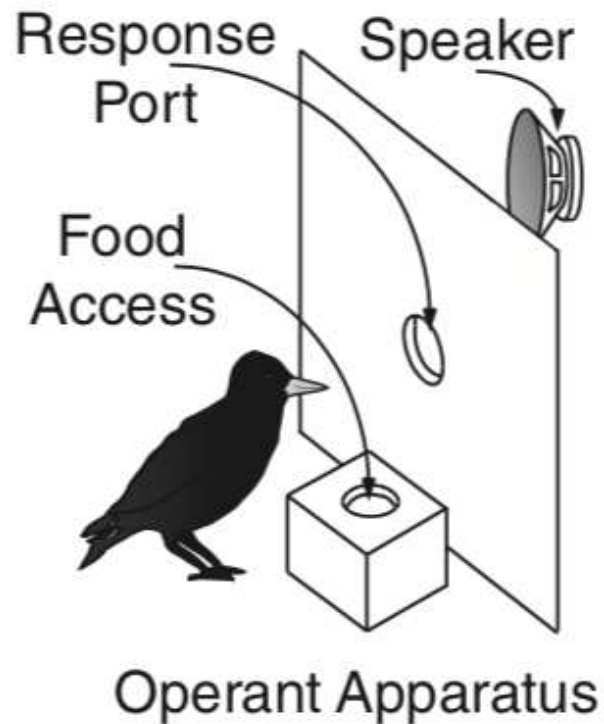


Figure 15: Schématisation de la tâche Go-No Go

À chaque essai, l'ordinateur sélectionne aléatoirement un chant. Lorsque le chant est diffusé, si le chant appartient à l'un des oiseaux connus par l'individu, l'oiseau doit s'approcher du trou dans les deux secondes (réponse Go), sinon il doit se tenir à distance (réponse No-go).

Un traitement non linéaire des paramètres acoustiques

Les réponses des neurones du NCM ne peuvent pas se prédire en se basant sur les propriétés acoustiques des signaux, comme cela peut s'observer dans le champ L (Sen et al. 2001; Grace et al. 2003; Mello and Ribeiro 1998). Chez le diamant mandarin, l'activité enregistrée lors de la présentation de stimuli auditifs qui possèdent les mêmes caractéristiques acoustiques en termes de fréquence et de modulation d'amplitude que les chants spécifiques induit des changements d'activité moindres que ceux observés lors de la présentation de chants de l'espèce (Sen et al. 2001; Grace et al. 2003). Chez le canari, la présentation de syllabes composées de deux éléments induit une activation neuronale, mesurée par le marquage de l'expression du gène Zenk, dans des régions du NCM qui diffèrent de celles activées lors de la présentation séparée de ces deux éléments : certaines régions ne montrent pas d'activation alors qu'elles sont activées lors de la diffusion de chacun des deux éléments (Mello and Ribeiro, 1998). En d'autres termes, l'addition des réponses neuronales enregistrées lors de la présentation de chacun des éléments n'est pas identique aux réponses neuronales obtenues lors de la présentation de ces mêmes éléments associés de façon à former une séquence (Mello and Ribeiro, 1998). En se basant sur le phénomène d'habituation, Ribeiro et ses collaborateurs ont apporté un argument supplémentaire en faveur de cette idée (Mello and Ribeiro, 1998). L'expérience a consisté en une présentation alternée de deux sifflements différents, espacés par 30 secondes pendant 7h (**Figure 16A**) ce qui induit une forte diminution de l'expression de Zenk. À la suite des 7h, ces deux mêmes sifflements ont été assemblés de façon à former une syllabe à deux éléments (les deux sifflements se succédaient l'un à l'autre) qui était présentée toutes les 30 secondes pendant 30 minutes. Ce changement a entraîné une réinduction de l'expression de Zenk (**Figure 16B**) suggérant que les neurones du NCM prennent en compte l'ensemble des paramètres acoustiques et notamment temporels comme un tout pour encoder des signaux acoustiques. Ces résultats suggèrent donc un traitement non linéaire des signaux acoustiques par le NCM. Ainsi, au vu de ce traitement non linéaire et de son implication dans la reconnaissance des chants des congénères, la question se pose de savoir si les neurones du NCM sont sensibles à l'ordre des syllabes au sein de chant ou de séquence. Au cours des répétitions, les neurones du NCM intègrent-ils à travers ses propriétés d'habituation, la séquence en tant qu'unique ensemble ou, chaque syllabe qui la composent individuellement ?

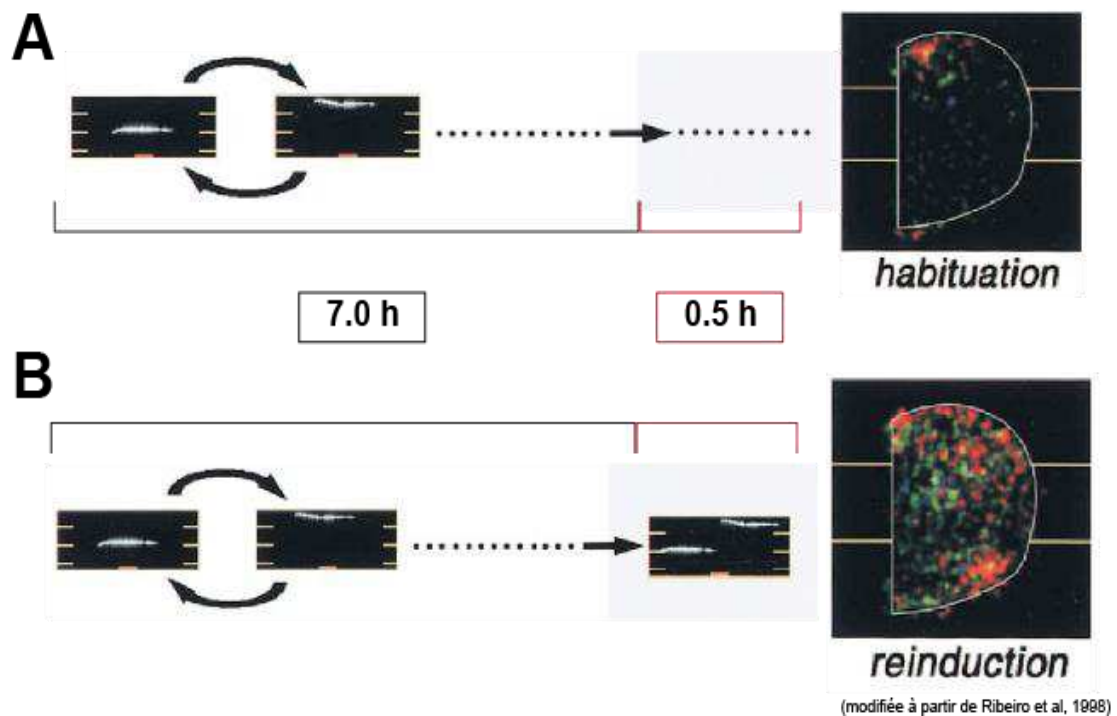


Figure 16: Intégration non linéaire des informations auditives.

- A) Marquage de l'expression du gène immédiat précoce Zenk dans le NCM après la diffusion pendant 7h30 minutes de deux notes alternativement toutes les 30 secondes.
- B) Marquage de l'expression du gène immédiat précoce Zenk dans le NCM après la diffusion pendant 7h00 minutes de deux notes alternativement, puis pendant les 30 minutes qui suivent d'une syllabe composée de ces deux notes.

On note que remplacer les deux notes présentées séparément par une syllabe composée de ces deux notes induit à nouveau l'expression du gène Zenk dans tout le NCM.

VII. Objectifs de l'étude

La question du traitement neuronal des signaux de communication vocaux est loin d'être résolue. Si de nombreuses études dans le domaine des neurosciences de l'audition se sont centrées sur le traitement des paramètres acoustiques tels que, par exemple, les fréquences, en comparaison, peu d'études ont cherché à déterminer comment sont encodées les séquences de signaux sonores au niveau cérébral. Pourtant, dans le langage humain, l'ordre des signaux joue un rôle fondamental dans la compréhension de ce qui a été produit.

Le but de ma thèse a été donc de comprendre comment et dans quelle région cérébrale l'ordre des signaux acoustiques qui s'enchaînent dans une séquence, est traité. Pour cela, j'ai étudié l'activité des neurones, lors de la diffusion de chants, dans deux régions cérébrales, le NCM et le HVC, chez deux espèces d'oiseaux chanteurs. D'un côté, le diamant mandarin, un modèle largement répandu dans les laboratoires, parce que les individus ont un chant unique et très stéréotypé. Cette espèce a l'avantage de chanter tout au long de l'année. De l'autre le canari, qui, a un chant complexe dans son organisation, mais qui est un oiseau saisonnier, limitant la période de l'année pendant laquelle les études pouvaient être réalisées.

A. Axe 1 : Étude du traitement de l'ordre des signaux acoustiques au sein d'une aire auditive, le NCM, chez le diamant mandarin.

La première étude a porté sur le traitement de l'ordre au niveau du NCM, chez le diamant mandarin parce que des études comportementales avaient montré, chez cette espèce, des capacités de discrimination parmi des chants composés de syllabes agencées selon des règles différentes. Au sein de cette région auditive, comme cela a été décrit plus haut, les neurones ont la particularité d'avoir des réponses qui diminuent au fur et à mesure qu'un stimulus est répété et qui augmentent à nouveau si un nouveau chant est présenté. Nous avons exploité cette propriété d'adaptation afin de savoir si les neurones du NCM étaient sensibles à l'ordre des syllabes dans les chants ou dans de courtes séquences agencées selon des règles différentes. Jusqu'à cette étude, peu de travaux avaient cherché à savoir si l'activité des neurones du NCM pouvait dépendre de la façon

dont les syllabes étaient agencées dans les chants. Cette étude a fait l'objet d'une publication dans la revue *The Journal of Neuroscience* (Cazala et al., 2019).

B. Axe 2 : Étude de la sensibilité à l'ordre des signaux acoustiques qui composent le propre chant de l'oiseau, au sein d'un noyau sensori-moteur, le HVC, chez le canari.

Depuis longtemps, des études ont montré que les neurones du HVC répondent différemment selon l'ordre des syllabes dans le chant diffusé. Cette propriété des neurones du HVC peut être mise en évidence si le chant diffusé est un de ceux produits par l'oiseau.

Afin d'aller plus loin, j'ai cherché à savoir sur quoi repose la sensibilité des neurones du HVC à l'ordre des syllabes dans le propre chant de l'oiseau. Pour cela, le modèle d'étude choisi a été le canari parce qu'il produit des chants complexes par la manière dont sont ordonnées les syllabes d'un chant à un autre. Notre approche a consisté, dans un premier temps, à mieux connaître la structure du chant de canari afin de déterminer sur quels aspects pourrait se baser le traitement de l'ordre des signaux dans le HVC.

Partie I :

TRAITEMENT DE L'ORDRE DES
SIGNAUX ACOUSTIQUES DANS UNE
AIRE AUDITIVE SECONDAIRE

Partie I : Traitement de l'ordre des signaux acoustiques dans une aire auditive secondaire

Article: Neuronal Encoding in a High-Level Auditory Area: From Sequential Order of Elements to Grammatical Structure.

Cazala A., Giret N., Edeline JM., Del Negro C.

Introduction

Encoder l'ordre des éléments au sein d'une séquence est nécessaire dans tout processus séquentiel. Dans le langage humain ou dans le chant des oiseaux chanteurs, les signaux auditifs sont organisés en séquences. Cependant, comment et où le cerveau traite ce paramètre reste peu compris.

Pour la vie sociale de l'oiseau chanteur, distinguer les chants entre les différents congénères est essentiel. Différentes stratégies sont mises en place pour discriminer les congénères dont, notamment, celles se basant sur une distinction dans l'agencement des syllabes (Suzuki et al. 2016; Ratcliffe and Weisman 1986). Or, les oiseaux chanteurs ne se contentent pas de discriminer l'ordre des syllabes sur la seule base d'un encodage temporel de chaque syllabe dans la séquence. Ils sont capables d'extraire les règles sous-jacentes à la construction des séquences. Suzuki et collaborateur (2017) ont pu mettre en avant, qu'inverser l'ordre entre deux cris associés une séquence revient à modifier les réponses comportementales de l'oiseau lors de son écoute. Chez le diamant mandarin, l'individu est capable de distinguer l'agencement de deux syllabes sur la base de règles grammaticales $(AB)^n$ et A^nB^n (Gentner and Margoliash, 2003; van Heijningen et al., 2009). Plusieurs études ont mis en évidence que les neurones du HVC sont sensibles à un changement de l'agencement des syllabes dans le propre chant de l'oiseau (Margoliash and Fortune 1992; Lewicki and Konishi 1995). Cependant, comment l'ordre des syllabes est codé au sein de chants conspécifiques ainsi que le rôle des régions télencéphaliques auditives dans les processus séquentiels est une question ouverte (Lu et Vicario, 2014 ;

Ono et al. 2016). Une de ces régions, le nidopallium caudo-médian (NCM) est un candidat idéal pour étudier cette discrimination. Les neurones du NCM présentent des réponses non linéaires : la réponse à une séquence de deux syllabes n'est pas égale à la somme des réponses à ces deux syllabes présentées séparément (Ribeiro et al. 1998). D'autre part, le NCM est connu depuis longtemps pour sa sensibilité au chant conspécifique : les neurones du NCM répondent plus fortement en réponse à des chants conspécifiques par rapport à d'autres signaux acoustiques, y compris des chants hétérospécifiques (Chew et al. 1996). De plus, les réponses neuronales sont modulées par l'expérience sociale, le taux d'activité étant plus élevé s'il s'agit d'un partenaire connu (Menardy et al., 2012).

Par ailleurs, les neurones du NCM présentent des propriétés d'habituation. Le taux d'activité des neurones du NCM diminue lors de la présentation successive d'un même stimulus. La diffusion d'un nouveau stimulus induit un regain de réponse neuronal permettant de conclure à une discrimination entre les deux stimuli présentés (Chew et al. 1995; Mello et al. 1995; Beckers and Gahr 2010; Menardy et al. 2012).

À partir de cette propriété, il nous est possible de tester si changer l'ordre au sein d'un chant conspécifique revient à diffuser un nouveau stimulus. De plus, à travers l'utilisation de séquences créées à partir de syllabes de chants conspécifiques, nous pourrions déterminer si les neurones du NCM sont sensibles à un changement de règles grammaticales. Ces deux grammaires sont organisées selon le schéma AABB et ABAB et sont discriminées par l'oiseau au niveau comportemental (A et B étant deux syllabes extraites de chants conspécifiques).

Résultats

Dans cette étude, nous avons effectué des enregistrements électrophysiologiques extracellulaires des neurones du NCM, chez des diamants mandarins mâles adultes, anesthésiés et éveillés, lors de la diffusion de chants et séquences de syllabes. Lors de ces enregistrements, tout comme décrit dans de précédentes études, on observe deux populations neuronales distinguables de par la forme de leur potentiel d'action (PA) et leur activité spontanée : les neurones larges ou BS pour *broad-spiking* et fins ou NS pour *narrow-spiking*. Ils se distinguent par leur schéma d'activation, les BS ayant une réponse phasique à des points précis dans le chant au contraire des NS qui ont une réponse

soutenue tout au long du chant. Seuls les BS montrent une sensibilité significative à un changement de l'ordre des syllabes constitutives d'un chant conspécifique. Non seulement les taux d'activité, mais aussi les patterns de réponses des neurones BS sont modifiés lors du changement de l'ordre des syllabes des chants présentés.

Dans les séquences, les syllabes sont agencées selon deux schémas, ABAB ou AABB. Ainsi entre ces deux séquences, seules les syllabes du milieu sont interverties. Suite à leur diffusion, des enregistrements ont été effectués chez des diamants mandarins anesthésiés et éveillés.

Chez les oiseaux anesthésiés, les séquences AABB et ABAB induisent une habituation des réponses neuronales suite à leurs diffusions répétées les séquences, mais ces répétitions s'accompagnent d'une diminution plus importante des réponses neuronales lors de la diffusion des séquences AABB que des séquences ABAB. Seuls les neurones BS détectent le changement AABB vers ABAB sans pour autant détecter le changement ABAB vers AABB. Les neurones NS ne détectent pas les changements d'ordre des syllabes dans ces séquences. Une analyse fine de l'intensité des réponses neuronales et de la reproductibilité des trains de potentiels d'action pour chaque syllabe nous a permis de mettre en évidence un codage local différent de ces syllabes et notamment de la première syllabe B dans les séquences AABB et ABAB. Cela illustre l'importance de considérer le contexte de production des syllabes dans une séquence.

L'étude menée chez les oiseaux éveillés a permis d'obtenir des résultats identiques à ceux enregistrés chez les sujets anesthésiés, suggérant que nos résultats ne dépendent pas du niveau d'attention de l'oiseau.

Conclusion

Ces résultats démontrent l'importance du contexte de chaque syllabe au sein de chants conspécifiques dans le NCM. Les neurones du NCM montrent une sensibilité à l'ordre temporel des syllabes dans les chants conspécifiques. Si jusqu'à présent de nombreuses études ont démontré l'importance d'un encodage des syllabes pour distinguer les chants conspécifiques dans le NCM, aucune n'a mis en avant l'importance de l'ordre des syllabes au sein de ces mêmes chants. Ce résultat fait écho aux études

comportementales ayant mis en avant une discrimination des congénères possible grâce à une sensibilité à l'ordre des syllabes au sein des chants (van Heijningen et al., 2009).

Cette étude suggère par ailleurs une sensibilité des neurones du NCM à la structure grammaticale de séquences de syllabe. Au-delà de coder simplement l'ordre des syllabes au sein de la séquence, l'agencement des syllabes suivant l'une ou l'autre des transitions impacte directement le niveau d'activité des neurones du NCM.

Nous proposons un modèle heuristique selon lequel la discrimination opérée par les neurones du NCM repose sur le traitement non linéaire des informations auditives en combinant une détection de changements locaux et globaux dans les séquences de syllabes.

Neuronal Encoding in a High-Level Auditory Area: From Sequential Order of Elements to Grammatical Structure

Aurore Cazala, Nicolas Giret, Jean-Marc Edeline, and Catherine Del Negro

Paris-Saclay Institute of Neuroscience, Centre National de la Recherche Scientifique, Unité Mixte de Recherche 9197, Paris-Saclay University, 91405 Orsay, France

Sensitivity to the sequential structure of communication sounds is fundamental not only for language comprehension in humans but also for song recognition in songbirds. By quantifying single-unit responses, we first assessed whether the sequential order of song elements, called syllables, in conspecific songs is encoded in a secondary auditory cortex-like region of the zebra finch brain. Based on a habituation/dishabituation paradigm, we show that, after multiple repetitions of the same conspecific song, rearranging syllable order reinstated strong responses. A large proportion of neurons showed sensitivity to song context in which syllables occurred providing support for the nonlinear processing of syllable sequences. Sensitivity to the temporal order of items within a sequence should enable learning its underlying structure, an ability considered a core mechanism of the human language faculty. We show that repetitions of songs that were ordered according to a specific grammatical structure (i.e., ABAB or AABB structures; A and B denoting song syllables) led to different responses in both anesthetized and awake birds. Once responses were decreased due to song repetitions, the transition from one structure to the other could affect the firing rates and/or the spike patterns. Our results suggest that detection was based on local differences rather than encoding of the global song structure as a whole. Our study demonstrates that a high-level auditory region provides neuronal mechanisms to help discriminate stimuli that differ in their sequential structure.

Key words: artificial grammar; auditory perception; multielectrode; sequence; songbirds

Significance Statement

Sequence processing has been proposed as a potential precursor of language syntax. As a sequencing operation, the encoding of the temporal order of items within a sequence may help in recognition of relationships between adjacent items and in learning the underlying structure. Taking advantage of the stimulus-specific adaptation phenomenon observed in a high-level auditory region of the zebra finch brain, we addressed this question at the neuronal level. Reordering elements within conspecific songs reinstated robust responses. Neurons also detected changes in the structure of artificial songs, and this detection depended on local transitions between adjacent or nonadjacent syllables. These findings establish the songbird as a model system for deciphering the mechanisms underlying sequence processing at the single-cell level.

Introduction

Learning to recognize temporal sequences is fundamental to sensory perception and social communication. Like speech and language, the songs of many species of songbirds consist of several distinct vocal elements called syllables, organized into sequences.

The sequential ordering of syllables results from a learning process (Lipkind et al., 2013). Modifying the temporal arrangement of vocal items may influence receiver's responses (Holland et al., 2000; Clucas et al., 2004; Gentner, 2008; Comins and Gentner, 2010; Briefer et al., 2013; Suzuki et al., 2016, 2017). Songbirds therefore have perceptual capacities for assessing the temporal ordering of vocal sounds.

Several basic processes for the coding of temporal sequences of items in the brain have recently been proposed (Dehaene et al., 2015). The encoding of transitions between two adjacent items, the merging of successive distinct elements into a single auditory object and other hypothetical mechanisms have been proposed (Kiggins et al., 2012; Dehaene et al., 2015). Songbirds are well-suited species for the investigation of how neuronal circuits encode the temporal sequences of vocal items as their brain contains

Received Oct. 25, 2018; revised April 24, 2019; accepted April 24, 2019.

Author contributions: A.C., N.G., and C.D.N. performed research; A.C., N.G., and C.D.N. analyzed data; N.G. and C.D.N. designed research; N.G. and J.-M.E. edited the paper; C.D.N. wrote the paper.

This work was supported by the Centre National de la Recherche Scientifique, the Idex Neuro-Saclay, and the University of Paris Sud. N.G. was supported by Idex Neuro-Saclay Postdoctoral Fellowship. A.C. was supported by French Minister of Research and Technology. We thank Christophe Pallier and Stanislas Dehaene for advice in the conception of the experimental design.

The authors declare no competing financial interests.

Correspondence should be addressed to Catherine Del Negro at catherine.del-negro@u-psud.fr.

<https://doi.org/10.1523/JNEUROSCI.2767-18.2019>

Copyright © 2019 the authors

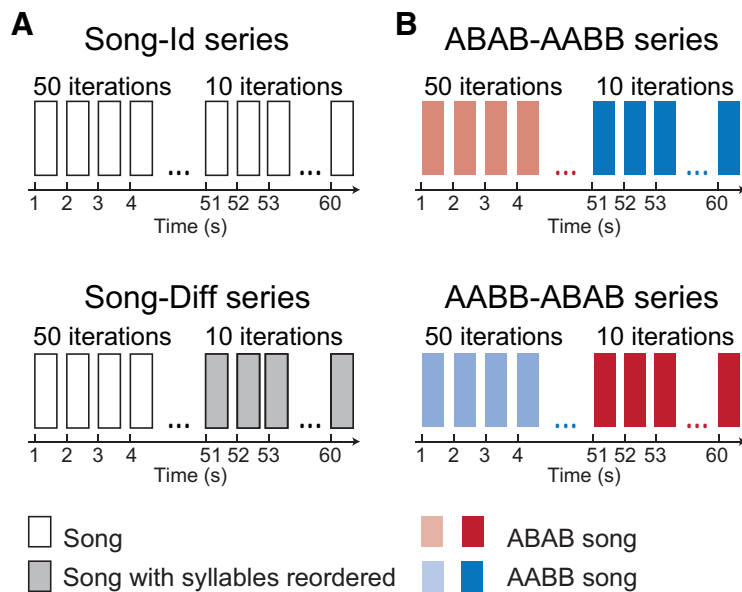


Figure 1. Schematic diagram of the series of songs used as stimuli. Except the Song-Id series, which consisted of 60 song iterations, the three other series consisted of 50 song iterations followed by 10 iterations of the same song in which syllables were reordered (Song-Diff series) or organized into a different structure (AABB-ABAB and ABAB-AABB series). Sets of two series were used: **A**, a Song-Id and its respective Song-Diff series built from the same syllables; **B**, an AABB-ABAB and an ABAB-AABB series built from distinct syllables.

many interconnected neural networks specialized in perception, learning, and production of vocal sounds, like humans. Sensitivity to temporal combinations of syllables has already been investigated in the set of brain nuclei specialized for singing and song learning (Margoliash, 1983; Doupe, 1997; Kojima and Doupe, 2008). For example, neurons in the nucleus HVC (used as a proper name) are sensitive to changes affecting the temporal structure of the bird's own song (Margoliash and Fortune, 1992; Lewicki and Konishi, 1995).

Much less is known about the encoding of temporal order of syllables within conspecific songs and the role of telencephalic auditory regions in sequence processing (Lu and Vicario, 2014; Ono et al., 2016). One of these regions, the caudomedial nidopallium (NCM), receives input from thalamo-recipient Field L and is a likely candidate area for discriminating songs according to syllable ordering. This is mainly because NCM neurons display nonlinear response properties. Indeed, neuronal responses to a multinote syllable cannot be explained by responses to the various notes presented separately (Ribeiro et al., 1998). Given these findings, it may be assumed that NCM neurons extend their nonlinear processing to sequences of contiguous syllables. In addition, NCM neurons show stimulus-specific adaptation where responses decrease with repeated presentation of the same song and exposure to a novel song reinstates responses (Chew et al., 1995; Mello et al., 1995; Beckers and Gahr, 2010; Menardy et al., 2012). This phenomenon has been interpreted as reflecting memory formation of the previously heard song. To date, most studies have only used songs produced by distinct birds as stimuli. In the present study, we provide evidence that a rearrangement in the temporal order of syllables is sufficient to reset neuronal responses after repeated presentations of a conspecific song.

Sensitivity to the order of items in a sequence may depend on the encoding of local transitions between adjacent or nonadjacent items or on its global structure (i.e., the items being grouped together as a “chunk”). Short sequences may present an oppor-

tunity to examine sensitivity to the local or global structure (Bekinschtein et al., 2009). The ability to learn the structure of sequences, considered as a core mechanism of human language faculty, has been investigated in both humans or nonhuman animals, including songbirds. In experiments, individuals were exposed to sound stimuli organized according to a structure based on artificial grammar, for example, (AB) n or a AnB n sequence (A and B denoting vocal sounds, with $n = 2$ or 3) (Fitch and Hauser, 2004; Friederici et al., 2006; Gentner et al., 2006; van Heijningen et al., 2009; Wilson et al., 2017; Wakita, 2019). Here, based on extracellular recordings in anesthetized and awake zebra finches, we show that NCM neurons responded differently to song stimuli organized into either an AABB or an ABAB structure.

Materials and Methods

Subjects and housing conditions. The subjects were 19 adult male zebra finches (*Taeniopygia guttata*), reared socially in the breeding colony of the Paris-Sud University. Birds were kept under a 12:12 light-dark cycle, with food and water *ad libitum*, and an ambient temperature

of 22°C–25°C. Experimental procedures were performed in compliance with national (JO 887–848) and European (86/609/EEC) legislation on animal experimentation, and following the guidelines used by the animal facilities of Paris-Sud University (Orsay, France), approved by the national directorate of veterinary services (#D91-429).

Auditory stimuli. As auditory stimuli, we first used natural zebra finch songs and variants of these songs that were built by manipulating syllable order. The set of songs included four distinct original zebra finch songs and three variants of these songs. Original songs came from a collection of songs previously recorded (sampling rate: 32 kHz) from adult male zebra finches that had lived in the laboratory's aviary years before the experiment. Subjects had never been exposed to these songs before the electrophysiological experiment. The four original songs were approximately the same length (mean \pm SD: 0.75 \pm 0.02 s). They consisted of five distinct syllables called A, B, C, D, and E. Both the delivery order of both syllables and silent gaps between syllables were manipulated to create song variants of the same length as the original ones. We made three variants of ABCDE songs by changing three or four of the four transitions between adjacent syllables: the BCEAD, EDCBA, or AECBD variants. Original songs and their respective variants were used to build sets of two series types (Fig. 1A). Each set included a series that consisted of 60 repeats of a given song (Song-Id series) and another series in which the same song was repeated 50 times followed by 10 presentations of a variant (Song-Diff series).

We also used four-element songs as stimuli. These artificial songs consisted of two element types, As and Bs, that were ordered in either the ABAB or the AABB structure. To build ABAB and AABB stimuli, we first selected 81 song syllables produced by 12 male zebra finches from our song database. The As and Bs were chosen to build ABAB or AABB stimuli of 0.70 \pm 0.30 s duration with 30–50 ms as intersyllable intervals, as typically found in zebra finch songs. Syllable duration ranged from 50 to 150 ms. All ABAB- or AABB-structured song stimuli started with the same introductory note. We created two series types: the ABAB-AABB or the AABB-ABAB series (Fig. 1B). They consisted of 50 repeats of either an ABAB or an AABB song followed by 10 presentations of the same song in which the serial positioning of the two central syllables A and B changed. In other words, the AABB and ABAB song stimuli of a given series were constructed from the same syllables A and B. We also built a third type of

series that consisted of only 60 repeats of an ABAB song. From the 81 syllables, we created a large set of ABAB-AABB or the AABB-ABAB series.

When a series was played back, the song stimulus was delivered at a rate of one per second.

Electrophysiological recordings. Neuronal activity in NCM was first recorded in anesthetized males while presenting the sets of songs that included either Song-Id and Song-Diff series or AABB-ABAB and ABAB-AABB series. Because NCM contains at least two populations of neurons that can be distinguished on the basis of their action potential (AP) width and their firing rate (FR) (Schneider and Woolley, 2013; Ono et al., 2016), we only analyzed the activity of very well isolated NCM single neurons (Fig. 2). Subsequently, we performed multiunit recordings in awake zebra finch males ($n = 5$) while presenting AABB-ABAB and ABAB-AABB series.

Acute recordings. Birds were anesthetized with isoflurane gas (in oxygen; induction: 3%, maintenance: 1.5%) that flowed through a small mask over the bird's beak. The bird was immobilized in a custom-made stereotaxic holder that allowed the head to be tilted at 45° and placed in a sound attenuation chamber. Lidocaine cream was applied to the skin. A window was opened in the inner skull layer, and small incisions were made in the dura. A multielectrode array of 8 or 16 tungsten electrodes (1–2 mΩ impedance at 1 kHz; Alpha Omega Engineering), which consisted of two rows of four (or eight) electrodes separated by 100 μm apart, with 100 μm between electrodes of the same row, was lowered to record extracellular activity. The array was positioned 0.3–0.5 mm lateral and 0.7–0.9 mm rostral to the bifurcation of the sagittal sinus in either the left or the right hemisphere, with a micromanipulator, as in previous studies (Stripling et al., 1997; Menardy et al., 2012, 2014). The probe was lowered very slowly until electrode tips reached 1200 μm below the brain surface. From 1200 to 1900 μm below the brain surface, auditory stimuli were delivered when the amplitude of AP waveforms recorded with at least one of the eight wires was clearly distinct from background noise. Recording sites were at least 100 μm apart to minimize the possibility that the neural activity recorded from two successive sites originated from the same single units. Electrode signals were amplified and filtered (gain 10,000; bandpass: 0.3–10 kHz; AlphaLab SnR, AlphaOmega Engineering) to extract multiunit activity. During recordings, voltage traces and APs were monitored in real time using the AlphaLab SnR software. Auditory stimuli were concomitantly recorded and digitized to precisely determine the onset of NCM responses with respect to the sound stimulus. While spiking activity was recorded, auditory stimuli were broadcasted through a speaker situated 30 cm from the bird. In the first experiment, the set of the two series of song stimuli, which included the Song-Id and Song-Diff series, was presented. The set of stimuli that consisted of both the ABAB-AABB and ABAB-AABB series was used in the second experiment. From one recording site to the following one, we changed the set of series used as auditory stimuli and the order of series. All stimuli had been normalized to achieve maximal amplitude of 70 dB (Audacity software) at the level of the bird's head. Spike sorting of neuronal activity was done offline.

Chronic recordings. Surgical procedures were similar as described above. To perform chronic recordings, we used a custom-built screw microdrive that allows a microelectrode array to be repositioned. We used arrays of eight electrodes (two rows of four electrodes separated by 100 μm apart; with a ground silver wire and a reference wire; 1–2 mΩ impedance at 1 kHz; AlphaOmega Engineering). Once the array was lowered into the brain to a depth of 1200 μm, the reference wire was inserted between the outer and the inner skull layers. The microdrive was secured to the skull using dental cement. Subjects were allowed to recover for a few days. In the sound-attenuation chamber, the implanted micro-

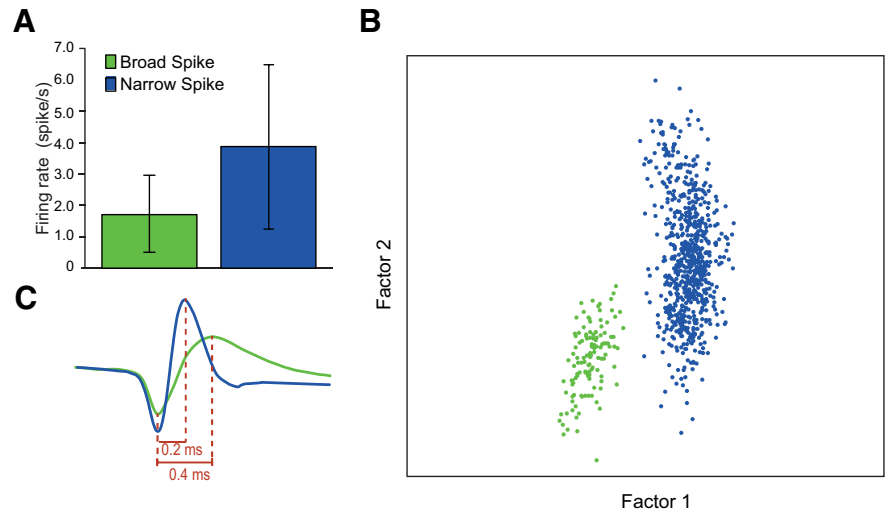


Figure 2. Two distinct cell types in NCM: BS and NS cells. **A**, Spontaneous FR (spike/s) of the populations of BS and NS neurons. Error bar indicates mean \pm SD. **B**, Representative example of spike clustering that distinguishes a BS (green) from an NS neuron (blue). **C**, Example of a spike waveform of these two neurons.

drive was connected through a commercial tether and head stage (AlphaOmega Engineering) to a mercury commutator located on the roof of the cage (Dragonfly Systems). An elastic thread built into the tether helped to support the weight of the implant. Subjects remained tethered during the experiment. The screw drive held the electrode array. Each full turn of the screw advanced the array by 200 μm. Before a recording session, we rotated the screw by half-turn to advance the microelectrode array in step as \sim 100 μm. Birds were not freely moving during the recording session. They were restrained with a jacket around their bodies. At least 24 h separated two recording sessions. As stimuli, we used one set of two series (ABAB-AABB or AABB-ABAB series) and also a third series that consisted of 60 repeats of an ABAB song (ABAB-ABAB series). From one recording session to the following one, we changed the set of series used as auditory stimuli and the order of series.

Data processing and analysis. In anesthetized birds, spike sorting was performed using the template-matching algorithm of the Spike2 software (version 8.0, Cambridge Electronic Design). In awake birds, spikes were detected by amplitude thresholding. Responses to stimuli were quantified by calculating average evoked FRs and Z scores. Evoked responses correspond to the FR during stimuli presentations. Z scores were measured as follows:

$$Z \text{ score} = \frac{\text{Evoked FR} - \text{Baseline FR}}{SD_{\text{Baseline FR}}}$$

The baseline FR was calculated for the last 200 ms period that preceded stimulus presentation. We calculated Z score values per block of 10 presentations, giving us 6 values per series (one per block of 10 iterations of the stimulus). The temporal pattern of responses evoked by both AABB and ABAB songs was quantified calculating the following: (1) the average FRs evoked by individual syllables contained in these songs; and (2) the spike-timing reliability coefficient (CorrCoef), which was used to quantify the iteration-to-iteration reliability of responses. It was computed per block of 10 stimulus iterations: it corresponds to the normalized covariance between each pair of AP trains and was calculated as follows:

$$\text{CorrCoef} = \frac{1}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{\sigma x_i x_j}{\sigma x_i \sigma x_j}$$

where N is the number of iterations and $\sigma x_i x_j$ is the normalized covariance at zero lag between spike trains x_i and x_j , where i and j are the iteration numbers. Spike trains x_i and x_j were previously convolved with a 10-ms-width Gaussian window. The CorrCoef was used because this index was not influenced by fluctuations of FR (Gaucher et al., 2013).

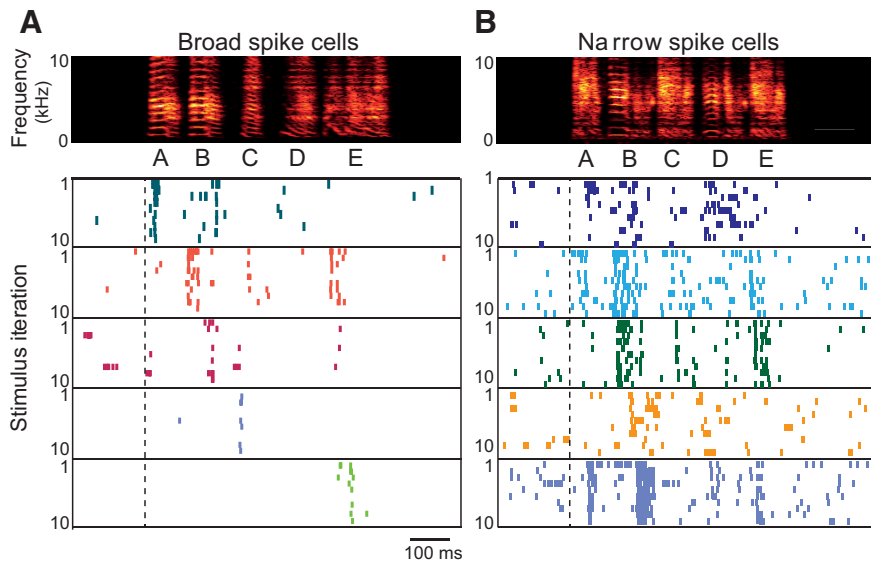


Figure 3. From one cell to another, song-driven spiking events can occur at different times during the presentation of a song stimulus. Neuronal responses of 5 BS and 5 NS cells are shown as raster plots (first 10 iterations). Top, The spectrogram of the song. BS neurons (**A**) produced more precise spike trains than NS neurons (**B**). Songs consisted of five distinct syllables (called from A to E). All the syllables of a given song can potentially elicit changes in spiking activity.

Statistical computations were performed in Statistica (version 8.0; StatSoft). FRs and *Z* score values were analyzed using repeated-measures ANOVA in a GLM. Several cofactors were included in the model, depending on the analysis: the block effect ($n = 6$), the series type (Song-Id vs Song-Diff; ABAB-AABB vs AABB-ABAB; ABAB-AABB vs ABAB-ABAB), the cell type (broad spike cells vs narrow spike cells), and the serial position of the syllable within the song ($n = 4$). We used planned contrast and Tukey HSD *post hoc* tests for assessing pairwise differences.

Histology. At the end of each experiment, the animal was killed with a lethal dose of pentobarbital and the brain quickly removed from the skull and placed in a fixative solution (4% PFA). Sections (100 μm) were cut on a vibratome to examine the location of multielectrode array penetration tracks.

Results

Experiment 1: repeated playbacks of a conspecific song followed by a change in the sequential order of syllables

Auditory responses in NCM decrease with song or call repetition, but a novel song or call stimulus reinstates responses (Chew et al., 1995; Stripling et al., 1997; Beckers and Gahr, 2010; Menardy et al., 2012). Various studies have exploited this property to characterize what is being encoded by the NCM neurons. Here, we assessed whether NCM neurons are sensitive to the auditory temporal context in which syllables occur by manipulating the temporal order of syllables. Such manipulation preserves the local and spectral structure within each syllable but alters what precedes each syllable. Transitions between adjacent syllables changed in all but one case (the transition B to C when ABCDE was reordered to BCEAD). We hypothesized that, if NCM neurons are only sensitive to the acoustic features of syllables, their responses should continue to decrease after changing the sequential order of syllables. On the contrary, if NCM neurons are also sensitive to what precedes syllables (i.e., depends on the auditory context), changing the temporal order of syllables within songs should reinstate robust neuronal responses.

Extracellular recordings from the NCM were performed in eight isoflurane-anesthetized adult males. Only well-isolated single units (Fig. 2), which responded to song playback with increased FRs relative to baseline activity, were selected. The

spiking activity of 205 auditory NCM neurons was analyzed. These units were from the dorsorostral portion (maximal depth 2000 μm) to the dorsocaudal portion as described by Menardy et al. (2012). They were driven by the playback of two series: the Song-Id and the Song-Diff series.

As described in previous studies (Meliza and Margoliash, 2012; Schneider and Woolley, 2013; Ono et al., 2016; Yanagihara and Yazaki-Sugiyama, 2016), responsive NCM neurons can be distinguished by the width of their AP and their spontaneous FR as follows: broad-spiking (BS) neurons (AP width: 0.50 ± 0.07 ms; FR: 1.81 ± 1.34 spikes/s; mean \pm SD, $n = 134$) and narrow-spiking (NS) neurons (AP width: 0.21 ± 0.05 ms, 4.00 ± 2.77 spikes/s; $n = 71$; Fig. 2). The pattern of song-evoked responses differed between BS and NS neurons. The responses of five BS cells and five NS cells to the first 10 iterations of a given song are displayed in Figure 3. The BS units showed phasic responses, firing only at precise restricted

points during the song stimulus. The NS units responded to the conspecific song with less accurate timing in their spike trains and continued to fire more or less during the entire song. As has been previously reported (Schneider and Woolley, 2013) and is illustrated in Figure 3, spiking events from one cell to another occurred at different time points in the song. All syllables can potentially drive spiking events from both BS and NS neurons, which supports the idea that neural coding of songs in the NCM relies on distributed population coding (Schneider and Woolley, 2013).

To assess NCM neuron sensitivity to repeated song exposure and to changes in syllable order, we calculated normalized *Z* score values per block of 10 iterations and performed a repeated-measures ANOVA on *Z* score values. Three cofactors, including block, series type, and cell type, were used. The time course of changes differed between the two series (interaction block \times series type, $F_{(5,1000)} = 2.40$, $p = 0.038$; Fig. 4A). Repetition of the same song 50 times decreased the response magnitude, but this did not differ between the two series (main effect of block factor, $F_{(5,1000)} = 9.01$, $p < 0.001$; from block 1 to block 5, *post hoc* HSD Tukey tests, all $p > 0.5$). Changing the syllable order within the song or extending its repetition differentially affected response magnitude (planned contrast of series type on the two last blocks, $F_{(1,200)} = 8.36$, $p = 0.004$; Song-Id series vs Song-Diff series on block 6, $p = 0.003$; Fig. 4A). Rearranging syllable order reinstated strong responses (Song-Diff series, block 5 vs block 6, $p = 0.003$).

The time course of responses and the sensitivity to syllable order rearrangement differed between the two cell types (interaction between the three factors: $F_{(5,1000)} = 2.91$, $p = 0.01$). NS cells did not show any modulation in their responses with stimulus repetition (Fig. 4B). A transient change in responses, however, occurred: from the first to the second presentation, FRs sharply declined (series Song-Id, first iteration mean \pm SD = 5.43 ± 0.61 spikes/s; second iteration mean \pm SD = 3.57 ± 0.63 spikes/s; paired *t* test, $t_{(70)} = 8.62$, $p < 0.001$; series Song-Diff: first iteration: 4.93 ± 0.51 spikes/s; second iteration: 3.2 ± 0.39 spikes/s; paired *t* test: $t_{(70)} = 11.73$, $p < 0.001$). FRs subsequently remained stable despite song repetition and syllable order rear-

repetition. In contrast, BS cells showed a gradual decrease in response magnitude when the same song was repeated (Fig. 4C; $F_{(5,500)} = 6.89$, $p < 0.0001$; *post hoc* tests for both series, block 1 vs 4 or 5, $p < 0.0001$). Reordering the syllables reset the FR to a high level ($F_{(1,100)} = 28.58$, $p < 0.0001$; *post hoc* test, block 5 vs 6, $p < 0.0001$); no change was observed when the order was kept the same (block 5 vs 6, $p = 0.99$). During the last block of song presentation, responses also differed between the two series ($p < 0.0001$). Of the 134 BS cells, 98 (73.1%) showed stronger responses when syllables were reordered. These results provide evidence that the majority of BS cells are sensitive to manipulation of the sequential ordering of syllables within songs. The Figure 4D displays the spiking activity of one representative BS neuron.

Interestingly, the mean FR of other BS cells did not increase when syllables were reordered, but responses were clearly affected by syllable reordering. Three such examples are presented in Figure 5; these also illustrate context dependency of responses. In the first example, the neuron showed peak activity at song onset during the repeated song (Fig. 5A, syllable A). Following syllable reordering (i.e., when syllable A was preceded by syllable E in the BCEAD song), the neuron no longer responded to this syllable (Fig. 5A, Song-Diff series). A peak in spiking activity was still observed when the syllable remained at the start of the ABCDE song (Fig. 5A, Song-Id series). Figure 5B shows the responses of another representative neuron. Rearranging the syllable order suppressed responses to syllable D (Fig. 5B, blue arrows, syllable D). Interestingly, these two neurons displayed in Figure 5A and B showed a greater response to another syllable when syllables were reordered (red arrows, syllable B in Fig. 5A; syllable E in Fig. 5B). Thus, the loss of response driven by a particular syllable could be accompanied by the reinstatement of robust responses to another syllable. The third example, which is displayed in Figure 5C, showed complete suppression of responses to syllable C after syllable reordering, even though this syllable remained at the same position within the song. A total of 12 cells displayed similar changes in discharge patterns. This suggests that a large fraction of BS neurons (110 of 134, 82%) encoded the auditory context in which syllables occurred through song repetitions either by increases in FR and/or changes in discharge patterns. Together, these results provide evidence that responses elicited by a multiple-syllable song cannot be reduced to responses evoked by individual syllables. This illustrates the sensitivity of NCM neurons to the auditory context in which syllables are presented.

Experiment 2: repeated playbacks of either an ABAB or AABB-structured song stimulus followed by a change in structure

To further examine the context dependency of NCM neurons, we used artificially AABB- or ABAB-structured song stimuli that

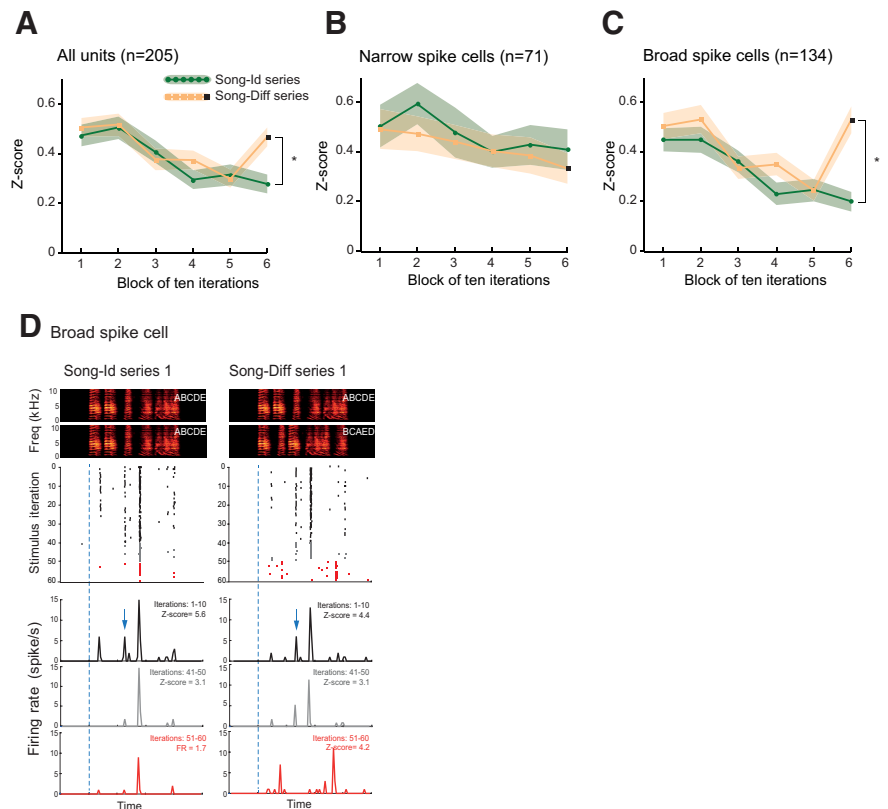


Figure 4. Reordering the syllables within a conspecific song affects the modulation of single-unit responses. Responses (Z score values) of the whole population of cells (**A**), the subset of NS cells (**B**), and the subset of BS cells (**C**) to repeated songs of either the Song-Diff (orange lines) or the Song-Id (green lines) series. Z score values were calculated per block of 10 song iterations. Changes in syllable order occurred at the beginning of the sixth block of the Song-Diff series. Thick line indicates mean. Shaded area represents SEM. * $p < 0.05$. **D**, Responses of a representative BS neuron to 50 presentations of a song consisting of five distinct syllables (ABCDE) followed by 10 presentations of the same song in which syllables are in the same order (Song-Id series) or arranged according to a new one (Song-Diff series; here BCAED). Neuronal responses are shown as raster plots (middle, 60 iterations) and peristimulus histograms (bottom) that are time-aligned with song spectrograms (top, the song repeated 50 times; bottom, the same song or the song with reordered syllables). Dotted lines indicate song onset.

differed in the serial positioning of the two central syllables A and B. The first and the fourth syllable remained at the same position. We aimed to examine how neurons responded to multiple repetitions of AABB and ABAB songs. Zebra finches are behaviorally able to discriminate between both songs (van Heijningen et al., 2009) and, based on previous studies, the NCM could provide a neural substrate for this ability. As suggested by the first experiment, if NCM neurons encode relationships between syllables, once responses are habituated, permuting the two central syllables A and B should reinstate larger responses.

At first, we performed neuronal recordings in anesthetized birds ($n = 7$). A total of 160 single units showed auditory responses to song stimuli. Both the AABB and ABAB songs induced a strong increase in FRs. From one neuron to another, spiking events occurred at different times in the songs (Fig. 6), and both syllables types could drive spiking activities. As shown in Figure 7A, responses to both song types were greatest during the first block of 10 iterations and showed a gradual decrease across successive blocks (main effect of block repetition, $F_{(5,790)} = 24.13$, $p < 0.001$; *post hoc* tests; block 1 vs 3, 4, and 5, for both series, $p < 0.01$ in all cases). However, lower responses were elicited by AABB compared with ABAB songs from the second block, suggesting that sequential structure of songs had an impact on responses (main effect of series type, $F_{(1,158)} = 9.25$, $p = 0.003$; blocks 2 and 5, $p < 0.002$). Neither the transition from AABB to

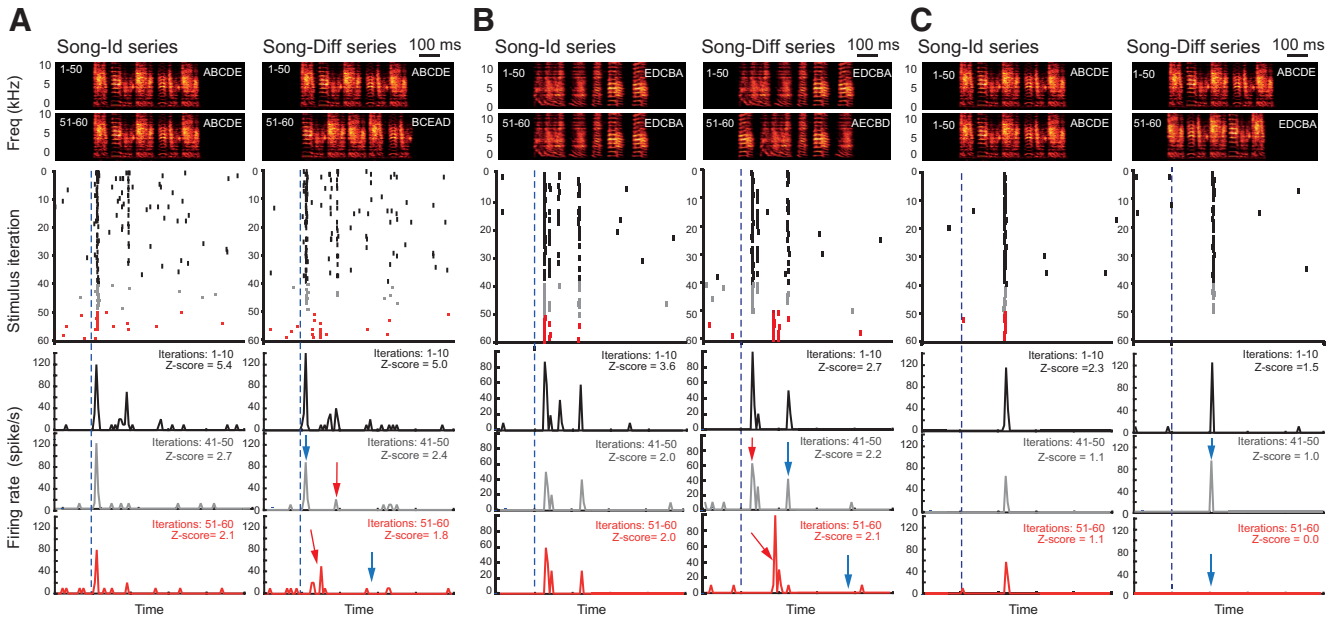


Figure 5. Reordering song syllables affects responses of three example BS cells. **A**, Spiking activity was suppressed after moving the syllable A from the first to the fourth position (ABCDE: initial order of syllables; BCEAD: new order in Song-Diff series; iterations: 51–60; blue arrows). There is reinstatement of responses to the syllable B after syllable reordering (red arrows). **B**, Reordering song syllables suppressed responses to the syllable D (EDCBA: initial order; AECBD: new order; blue arrows) and reinstated responses to syllable E (red arrows). **C**, Suppression of responses to the syllable C that remained to the same position after syllable reordering (ABCDE: initial order; EDCBA: new order; blue arrows). The preceding syllables were different. Neuronal responses are shown as raster plots (middle, 60 iterations) and peristimulus histograms (bottom) that are time-aligned with the song spectrograms. Spectrograms of songs used as stimuli: Top, During the first 50 iterations. Bottom, During the 10 last ones. Dotted blue lines indicate song onset.

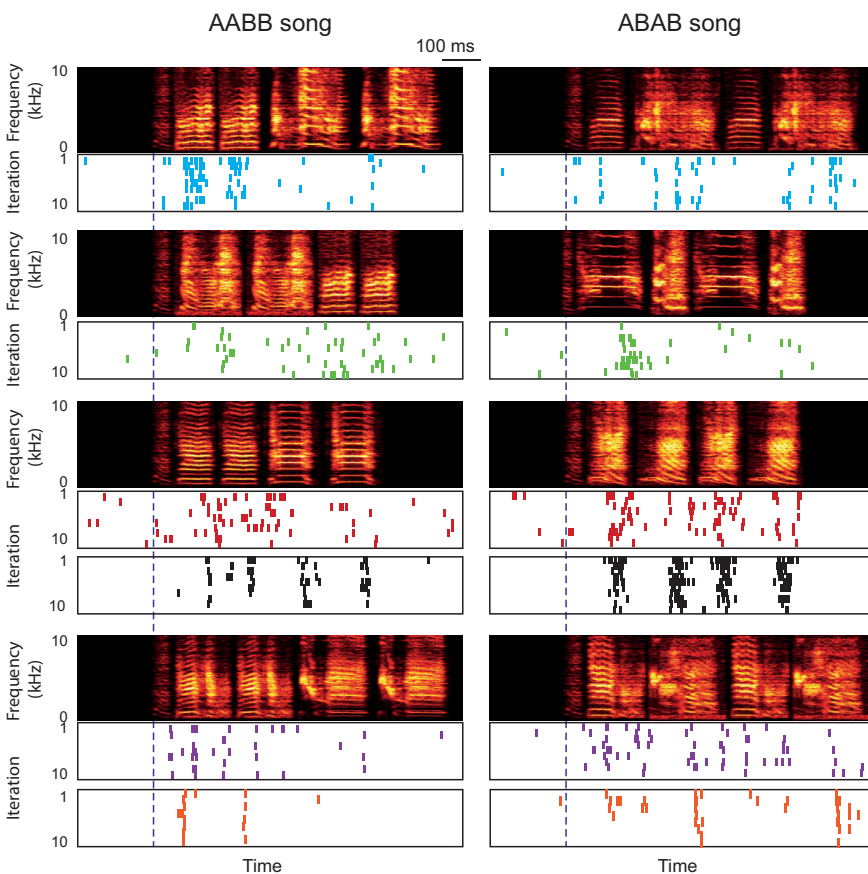


Figure 6. Responses of six example neurons to the playback of ABAB and AABB song stimuli. Neuronal responses are shown as raster plots (the first 10 iterations) that are time-aligned with the song spectrograms. Song-driven spiking events can occur at different times during both ABAB and AABB songs. Dotted blue lines indicate song onset. All the syllables of a given song can elicit changes in spiking activity.

ABAB nor vice versa was detected by the population of NCM cells as a whole.

Responses of both NS ($N = 68$) and BS cells ($N = 92$) decreased in magnitude with song repetition ($F_{(5,790)} = 24.13, p < 0.001$); however, a difference in the amount of decrease between AABB and ABAB songs was seen ($F_{(1,158)} = 9.25, p = 0.003; p < 0.002$). NS cells were not sensitive to changes in song structure once responses were habituated (Fig. 7B). In contrast, BS cells detected the transition from AABB to ABAB structure (“cell type” factor, $F_{(1,158)} = 6.12, p = 0.014$; *post hoc* test, block 5 vs 6; $p = 0.038$; Fig. 7C). Figure 7D illustrates a neuron showing this detection in its responses.

To assess whether responses could depend on behavioral state, we performed multiunit recordings in awake birds (recording sites $n = 52$ in 4 birds). On average, three (range 2–5) recording sessions (4.4 electrodes per recording session, range 2–8) per bird were performed with, on average, 5 d between two successive recording sessions (range 3–10). Most results are consistent with those obtained from anesthetized birds. Responses (Z score values) were the highest during the first block of song iterations and decreased during subsequent blocks (main effect of block repetition, $p < 0.001$). The AABB songs induced lower responses than ABAB songs (main effect of the song

type: $F_{(1,51)} = 11.9$, $p = 0.001$; from block 2 to block 5, *post hoc* tests: $p < 0.001$; Fig. 8A). After 50 song iterations, changing the song structure did not reinstate stronger responses (Fig. 8A). For example, changing from ABAB to AABB did not induce greater responses than continuing to repeat ABAB up to 60 times (Fig. 8A). The spiking activity of 18 single units from our recordings were further analyzed. As illustrated by the example presented in Figure 8B, BS cells showed an increase in FR when AABB songs were changed to ABAB. Thus, we confirm the results obtained from anesthetized birds that BS cells can detect the change in song structure from AABB to ABAB.

The difference in response habituation between AABB and ABAB songs and the differential effect of changing the song structure on responses led us to analyze how neurons respond to individual syllables contained in songs. We aimed to assess whether syllable-evoked responses express a context dependency and, if so, whether this context dependency is limited to local transitions between adjacent syllables. The first syllable B occurred in the following two different auditory contexts: (1) it was preceded by two iterations of the syllable A in AABB songs; and (2) it was preceded by only one in ABAB songs. The syllable pair AB, which included the first syllable B, was preceded by single syllable A in AABB songs while it occurred in silence in ABAB songs. If the auditory context does not affect syllable-evoked responses, similar responses to the first syllable B and the syllable pair AB (from syllable A to B) should be observed when AABB and ABAB songs are played back. On the contrary, differences between AABB and ABAB songs should be observed if the auditory context has an impact on responses.

Variations in syllable-evoked FRs differed between AABB and ABAB songs from the first to the fourth song syllable in both anesthetized and awake birds (“syllable position” effect: $F_{(3,465)} = 36.55$, $p < 0.001$ and $F_{(3,177)} = 8.52$, $p < 0.001$; interaction between “syllable position” and “song type” factors: $F_{(3,465)} = 24.72$, $p < 0.001$ and $F_{(3,177)} = 8.97$, $p < 0.001$; Fig. 9). As shown in Figure 9A and B, there was a gradual decrease in FR when syllables were organized according to the AABB structure (*post hoc* tests, from block 1 to 5, serial position 1 vs 3 and 4; 2 vs 3 and 4, all $p < 0.01$). In contrast, when syllables were alternated in ABAB songs, the FR first increased before it decreased and then increased again (Fig. 9A,B). Thus, FRs from one syllable to the next thus strongly differed between the two songs.

Responses driven by the first syllable B were weaker when AABB songs rather than ABAB songs were played back (*post hoc* test, $p < 0.01$). Changes in FR from syllable A to B differed. When syllable B followed two syllables A in AABB songs, the FR decreased (Fig. 9A,B, blocks 1 and 5). When syllable B followed a single A in ABAB songs, the FR increased (Fig. 9A,B, blocks 1 and 5). These results indicate that responses to the first syllable B

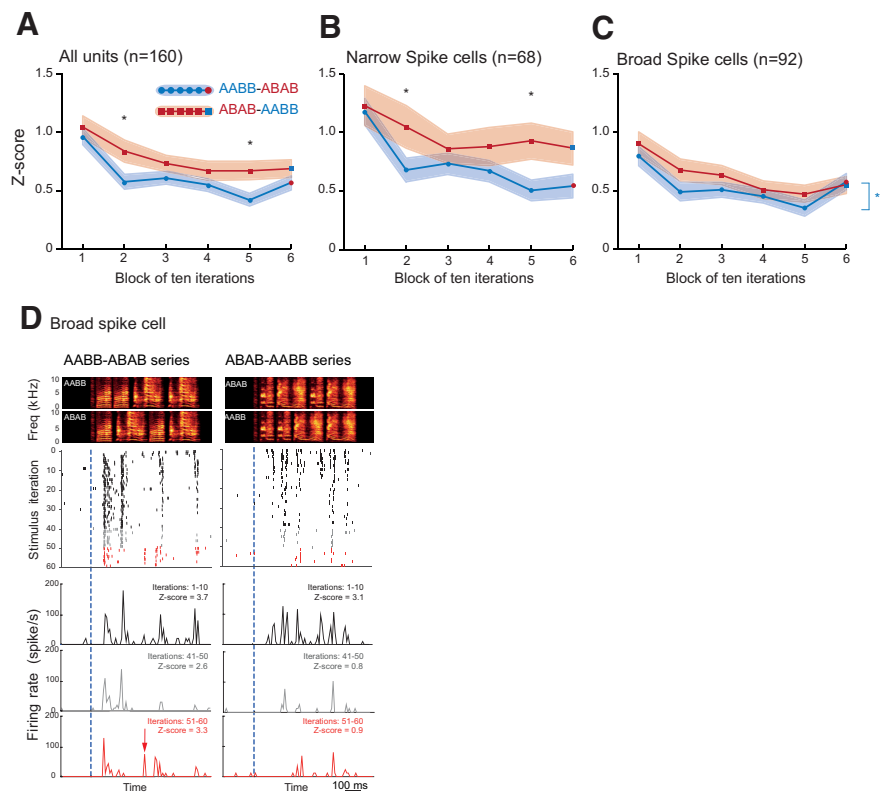


Figure 7. Responses driven by ABAB and AABB song stimuli in anesthetized birds. Responses (Z score values) of the whole population of cells (**A**), the subset of NS cells (**B**), and the subset of BS cells (**C**) across the successive blocks of AABB-ABAB and ABAB-AABB series. Z score values were calculated per block of 10 song iterations. BS cells detected the transition from AABB to ABAB structure. Thick line indicates mean. Shaded area represents SEM. $*p < 0.05$. **D**, Responses of a representative BS neuron to the playback of AABB and ABAB songs. Neuronal responses are shown as raster plots (middle, 60 iterations) and peristimulus histograms (bottom) that are time-aligned with song spectrograms. Spectrograms of songs used as stimuli: Top, During the first 50 iterations. Bottom, During the 10 last ones. Dotted blue lines indicate song onset. There is an increase in the response to the first syllable B (red arrow) when song structure changed from AABB to ABAB.

depended on the number of As that preceded it. Also, differences in responses between syllable A and the following syllable B depended on the auditory context in which the syllable pair AB occurred (silence vs preceded by the presentation of a syllable A).

A puzzling result, observed in both anesthetized and awake birds, was the large difference in response habituation between AABB and ABAB songs (Figs. 7A; 8A). The first syllable of AABB songs should have induced weaker responses than the first syllable of ABAB songs if the larger decrease in Z score values when AABB songs were played back resulted from fatigue. Instead, the first syllable A of AABB songs drove stronger responses than the first syllable A of ABAB songs in both anesthetized and awake birds (block 1 to 5, $p < 0.01$; in all cases). In contrast, the last syllable of AABB songs evoked much less spiking activity than the last syllable of ABAB songs both in anesthetized and awake birds (blocks 1 to 5, $p < 0.01$). These differences (larger responses to the first syllable and lower responses to the last syllable) between AABB and ABAB songs clearly illustrate the drastic decrease in FR during the AABB song. In addition, from the last syllable of a given AABB song iteration to the first syllable of the following one, the FRs partially recovered high values (Fig. 9A,B, blue dotted lines). This highlights another important difference between AABB and ABAB songs.

The impact of transition from one structure to the other one on syllable-evoked FRs of NS and BS cells in anesthetized conditions was subsequently assessed. In Figure 9C, responses elicited by central syllables during block 5 were artificially permuted to

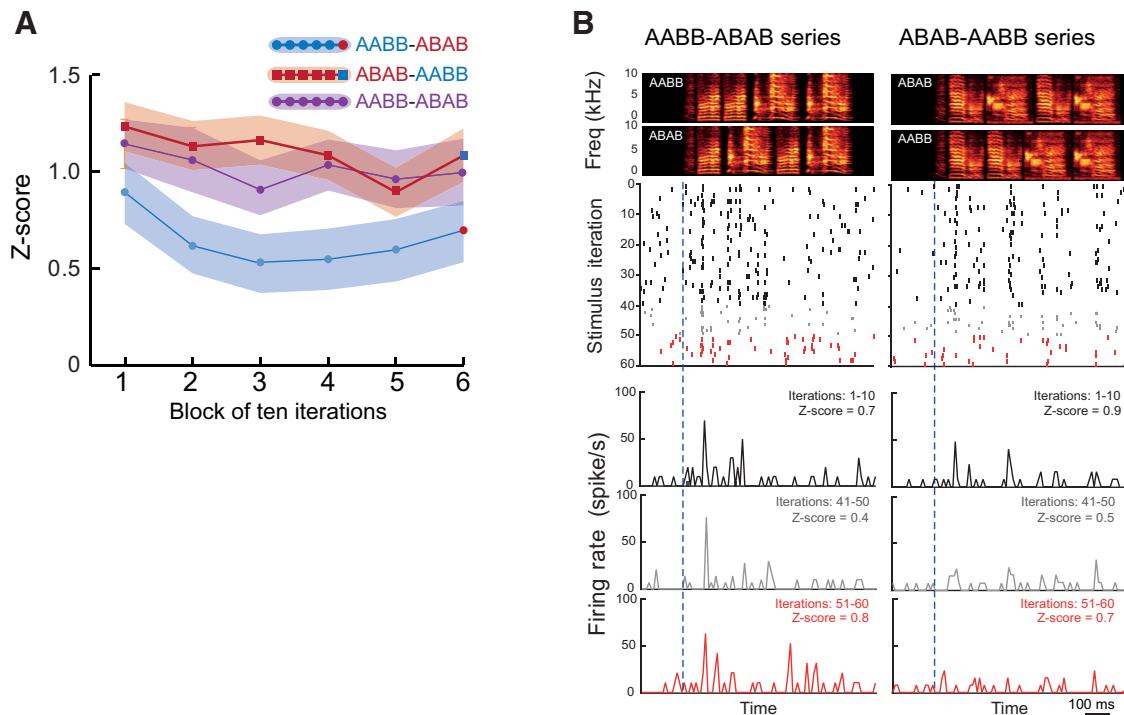


Figure 8. Responses driven by ABAB and AABB song stimuli in awake birds. **A**, Changes in multiunit activity across the successive blocks of AABB-ABAB (blue line) and ABAB-AABB (red line) series. A third series, the ABAB-ABAB series (purple line), was also played back. Thick line indicates mean. Shaded area represents SEM. **B**, Responses of a representative single unit (BS neuron) to the two structured songs. There is a reset in responses when the song structure changed from AABB to ABAB. Neuronal responses are shown as raster plots (middle, 60 iterations) and peristimulus histograms (bottom) that are time-aligned with song spectrograms (top). Spectrograms of songs used as stimuli: Top, During the first 50 iterations. Bottom, During the 10 last ones. Dotted blue lines indicate song onset. FR was averaged over the stimulus duration.

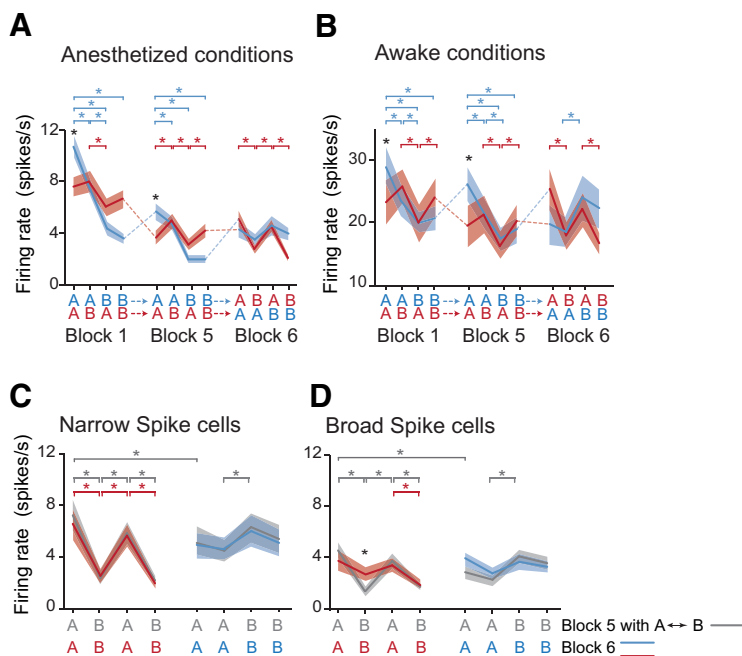


Figure 9. Variations in syllable-evoked FRs during song playbacks depend on song structure. In both anesthetized (**A**) and awake conditions (**B**), responses of the whole population of cells to the four song syllables of AABB (blue line) and ABAB (red line) songs. Thick line indicates mean. Shaded area represents SEM. Data collected before (blocks 1 and 5) and after (block 6) changes in song structure. There is a significant difference in response to the first syllable A between AABB and ABAB songs during blocks 1 and 5 (black $p < 0.001$) and the lack of this difference during block 6. Responses of NS (**C**) and BS (**D**) cells during blocks 5 and 6. Responses to the central syllables A and B during block 5 (in gray) are artificially permuted. Responses to AABB and ABAB songs are represented by blue and red, respectively. $*p < 0.05$, for the corresponding song structure (blue and red $*$ for AABB and ABAB songs, respectively).

show syllable-evoked responses according to the song structure used after the transition. Responses of NS cells during block 6 could be predicted by those elicited before permuting the central syllables (Fig. 9C). In contrast, BS cells showed a marked difference in their FR between blocks 5 and 6. More robust responses to the first syllable B after the transition from AABB to ABAB were evident (interaction between “block” and “song type” factors: $F_{(3,273)} = 3.11, p = 0.02$; block 5 vs 6; $p = 0.001$; Fig. 9D). In terms of the syllable pair AB that included the first syllable B, the FR of BS cells no longer differed between syllable A and B after transitions (Fig. 9D). These results clearly demonstrate that responses to song syllables did not strictly depend on acoustic identity of syllables but rather were influenced by the auditory context in which they occurred.

Responses of BS cells to the first song syllable A did not differ between AABB and ABAB songs after transitions (interaction between “block 5 and 6” and “song type” factors: $F_{(1,90)} = 7.8, p = 0.006$; block 5: $p < 0.001$; block 6: $p = 0.8$). In contrast, responses to the fourth syllable still differed between the two songs ($p < 0.001$). Responses to the first syllable A appeared to be influenced by the song structure as the

acoustic identity of both the first and the fourth syllables remained the same.

To evaluate to what extent the pattern of neuronal responses was affected by changes in song structure, we compared the precise timing of spike trains between two blocks of song presentations (block 4 vs 5 and block 5 vs 6) in awake birds. We computed the CorrCoef index (see Materials and Methods) from spike trains driven by each song syllable (Fig. 10A). CorrCoef values of ~ 0.2 were observed when spike trains recorded during blocks 4 and 5, before changing the song structure, were compared (Fig. 10B). These CorrCoef values represent a range usually seen for cortical neurons (Gaucher et al., 2013; Gaucher and Edeline, 2015). The CorrCoef values obtained for the second syllable reached a value close to 0 after altering the song structure (block 5 vs 6; Fig. 10C). This indicated a strong reduction in spike-timing reliability. Following transition from ABAB to AABB song, the second syllable of AABB songs (A) was followed by B in ABAB songs while the second syllable of ABAB songs (B) was preceded by two As in AABB songs. Therefore, changing the auditory context and particularly increasing the number of presentations of the syllable A before the syllable B had a strong impact on temporal reliability of responses to this syllable.

We also computed the CorrCoef index from spike trains driven by the syllable pair AB that included the first syllable B. This allowed us to examine whether the precise timing of APs elicited during this syllable pair was impacted by a change in the auditory context in which it occurred. We obtained CorrCoef values ~ 0.2 for the two songs before song transitions (block 4 vs 5; Fig. 10D). Following transition from one structure to the other (block 5 vs 6 and block 1 vs 6; Fig. 10D), CorrCoef values differed between the two transitions. The spike timing was less reliable after the transition from AABB to ABAB than from ABAB to AABB (block 1 vs 6, paired t test, $t_{(51)} = 2.70$, $p = 0.008$; block 5 vs 6, $t_{(51)} = 4.73$, $p < 0.001$; Fig. 10D). This, in addition to the results based on FR, indicate that the temporal reliability of neuronal discharges was markedly impacted by the transition from AABB to ABAB.

Discussion

Our results provide clear evidence that the neuronal responses of NCM neurons to playback of conspecific songs are sensitive to the song context in which syllables occur, and this sensitivity is expressed by changes in FRs and/or discharges patterns. The hypothetical outcomes of our study are summarized in Figure 11.

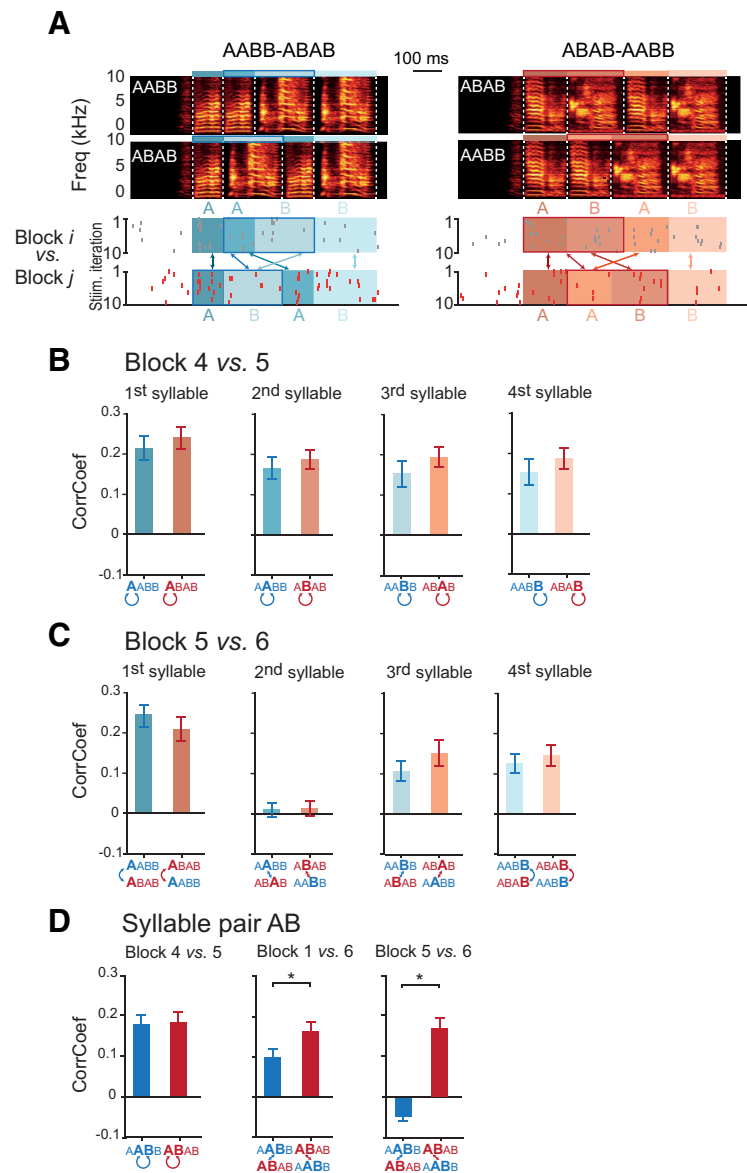


Figure 10. Effects of changes in song structure on temporal reliability of spike trains. **A**, Neuronal responses during two blocks of 10 song iterations are shown as raster plots that are time-aligned with song spectrograms. The CorrCoef index was computed by comparing the precise timing of spike trains between two blocks (*i* vs *j*) of song presentations (e.g., between iterations 1 to 10 of block *i* vs iterations 1 to 10 of block *j*). Computations were performed on spike trains elicited by individual syllables (AABB-ABAB and ABAB-AABB series: blue and red shaded areas on raster plots, respectively) or by the syllable pair AB that included the first syllable B (blue and red rectangles on raster plots). White dotted lines on spectrograms indicate individual syllables. Arrows above and below raster plots indicate the serial position of syllables before and after song transition, respectively. **B**, CorrCoef values (mean \pm SEM) for spike trains elicited by each of the four song syllables (in bold below histograms) of the AABB (in blue) and the ABAB (in red) songs during blocks 4 and 5. **C**, CorrCoef values for spike trains elicited by the same syllable (in bold below histograms) before (block 5) and after (block 6) changes in song structure. Arrows below histograms indicate the serial position of the syllable within the song before and after song transition. Moving the syllable from the second to the third position affected the temporal reliability of responses driven by this syllable. **D**, CorrCoef values for spike trains elicited by the syllable pair AB (in bold below histograms) during two blocks (4 and 5; 5 and 6; 1 and 6). * $p < 0.05$.

Sensitivity to the temporal order of syllables in conspecific songs

Our results show the repetition-induced decrease in responses, described by numerous studies both in mammals (Malmierca et al., 2014; Khouri and Nelken, 2015) and songbirds (Chew et al., 1995; Stripling et al., 1997; Phan et al., 2006; Beckers and Gahr, 2012; Smulders and Jarvis, 2013; Lu and Vicario, 2014). Once neuronal responses were decreased, the presentation of a novel song, generally sung by a distinct individual, resets the responses

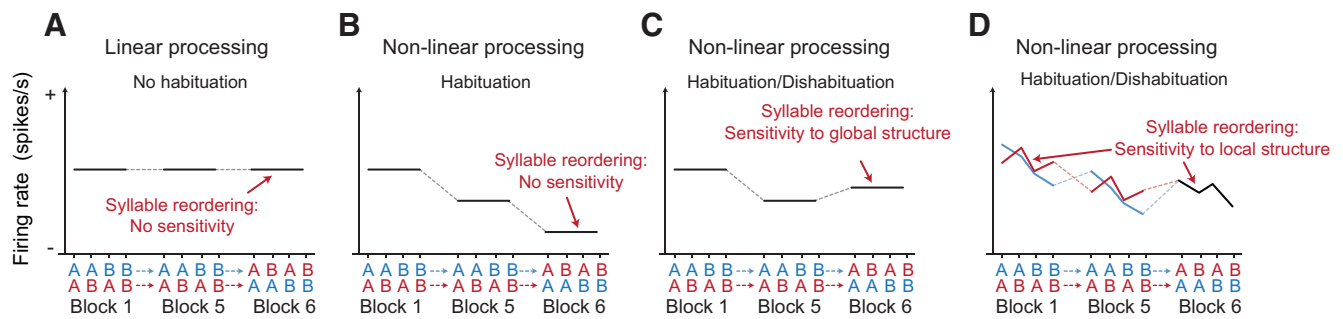


Figure 11. Hypothetical outcomes of our study. The different scenarios represent the responses of the whole population of cells to the presentation of AABB-ABAB and ABAB-AABB series during blocks 1, 5, and after changes in song structure (i.e., during block 6). **A**, Linearity of song responses. Neurons are only sensitive to certain acoustic features, and their responses are neither modulated by repetitions nor by changes in song structure (e.g., neurons of the songbird midbrain). **B–D**, Nonlinearity of song responses: repetition induces a decrease in responses. **B**, Neurons do not encode the temporal order of syllables within songs. **C**, Neurons encode the global structure of the song, syllables being grouped together as a “chunk.” Changing the song structure by permuting the two central syllables A and B reinstates stronger responses to all song syllables. **D**, Neurons encode the transitions between syllables. Changing the song structure by permuting the two central syllables A and B affects responses to these syllables. The sensitivity to local song structure is also detected in the initial responses during block 1. Responses to the first syllable B depends on the number of syllable A that precedes it. These responses are based on a set of results obtained for BS cells. NS cells did not show any changes in responses to song syllables after changes in song structure.

(Chew et al., 1995, 1996; Stripling et al., 1997; Smulders and Jarvis, 2013). This change in response magnitude represents a methodological approach to examine whether two stimuli that differ in certain aspects are encoded as similar or distinct auditory objects (Lu and Vicario, 2011; Moorman et al., 2011; Smulders and Jarvis, 2013). The reinstatement of strong responses when syllables were rearranged indicates that the temporal order of syllables within conspecific songs contributes to song identity.

To the best of our knowledge, no previous study has demonstrated that NCM neurons encode the sequential ordering of syllables within natural songs. Neurons in NCM have been previously found to be sensitive to the temporal order of sounds, for example, harmonic stacks in short or long streams (Lu and Vicario, 2014) and song elements in certain sequences of three elements (Ono et al., 2016). It was crucial to use conspecific songs as the activation of NCM neurons is largest when birds are exposed to this song type, as opposed to heterospecific songs or artificial stimuli (Mello and Clayton, 1994; Chew et al., 1996; Stripling et al., 1997; Van Meir et al., 2005). Behavioral studies have shown that the type of sounds used to build auditory stimuli and the way strings are segmented both influence the ability to detect changes in temporal sequences (Spierings et al., 2015; Neiworth et al., 2017).

Dependency on song context

Once the song stimulus has been repeated, larger responses to the novel arrangement of syllables are shown by BS cells rather than NS cells. Based on the width of the APs, these two cell types could correspond to excitatory (broad) and inhibitory (narrow) neurons (see the discussion in Schneider and Woolley, 2013). Our results therefore provide additional evidence that the two cell types differ in song processing (Meliza and Margoliash, 2012; Schneider and Woolley, 2013; Ono et al., 2016; Yanagihara and Yazaki-Sugiyama, 2016). Approximately three-fourths of BS cells showed increased responses when syllables were reordered; another subset of BS neurons ceased to fire at a given syllable when syllable order differed. A previous study reported that BS cells may fail to respond to a particular syllable included in a song but responded to this same syllable played in silence (Schneider and Woolley, 2013). By manipulating the temporal order of syllables in songs, we provide the first evidence that the responses to a given syllable depend on which syllables immediately precede it. As song responses in the NCM are nonlinear compared with song

responses in upstream auditory regions (Woolley, 2013), we assume that the extension of the nonlinear processing to contiguous syllables represents an emergent property of the NCM.

Syllables in songs were rearranged to alter transitions between adjacent syllables. In most cases, syllables were assigned to a different position than before (but see the example on Fig. 5C). Consequently, our song stimuli did not allow us to disentangle the following two possible interpretations for the response reinstatement: an encoding of either the transition between adjacent and nonadjacent syllables or the ordinal position of syllables in songs. However, both encodings require long-term integration of the sequential information that extends beyond individual syllables.

Sensitivity to grammatically structured songs

In both anesthetized and awake birds, NCM neurons (as a population) were sensitive to differences in song stimuli where two syllables were repeated twice or alternated. When the syllable B followed two As in AABB songs, the FR decreased. When the syllable B followed a single A in ABAB songs, the FR increased. Songbirds are behaviorally able to discriminate between differently structured sequences (Gentner et al., 2006; van Heijningen et al., 2009, 2013; Abe and Watanabe, 2011; Comins and Gentner, 2013, 2014). Our findings suggest that NCM may provide neural mechanisms to help discriminate between stimuli that differ in their sequential structure. We assume that the gradual decrease in FRs seen over the four syllables of AABB songs is a direct consequence of repetition of the first syllable. A study using XXY or YXX sequences (X and Y denoting song elements) as stimuli to examine responses in NCM of male zebra finches (Ono et al., 2016) found a gradual decrease over the three elements of XXY sequences (Ono et al., 2016, their Fig. 9).

Grammatically structured sequences have been used in many behavioral studies to determine whether nonhuman animals are able to learn the underlying structure of sequence sets (Berwick et al., 2011; Petkov and Wilson, 2012; Zaccarella and Friederici, 2017). This ability, which remains a matter of great debate (Beckers et al., 2017; Ghirlanda et al., 2017), was assessed by determining whether subjects detected the transition from one structure pattern to another or generalized the discrimination between stimuli to novel stimuli: in nonhuman primates (Fitch and Hauser, 2004; Uhrig et al., 2014; Wilson et al., 2015; Neiworth et al., 2017) and in songbirds (for review, see ten Cate, 2017). In our

study, we assessed detection of changes in grammatical structure by addressing this question at the neuronal level. Both the increase in responses of NCM neurons seen when AABB songs were rearranged into ABAB ones and the changes in temporal reliability of spike patterns seen when both transitions occurred provided support that encoding of relationships between syllables has occurred. However, most of the results highlight changes that are related to the first syllable B or the pair AB (which includes B), suggesting a detection of transitions based on local regularities/irregularities of songs rather than encoding of the global song structure (AABB vs ABAB). Behaviorally, zebra finches are able to generalize the discrimination between AABB and ABAB stimuli to novel song stimuli (van Heijningen et al., 2009). However, investigations have demonstrated that individuals do not learn the whole structure of the stimuli. Rather, they discriminate based on simpler differences, such as differences in the first (AA vs AB) or final two syllables (van Heijningen et al., 2013; Spierings and ten Cate, 2016), potentially by focusing on local transitions between two syllables (van Heijningen et al., 2009, 2013; Seki et al., 2013; Chen and ten Cate, 2015; Ghirlanda et al., 2017). Therefore, our experiments provide insights into the neuronal underpinnings of the behavioral ability to distinguish between AABB and ABAB structures, which has hitherto been unexplored at a cellular level in the avian and mammalian auditory system.

Despite the limited sensitivity of neurons to local song structure, NCM neurons could extend their encoding beyond contiguous syllables, as previously suggested (Lu and Vicario, 2014). The difference in BS cell responses to the first syllable A in AABB and ABAB songs (Fig. 9A,B), and the loss of response with a change in song structure supports this assumption.

Based on evidence provided by both experiments, our study demonstrates that an avian brain auditory region shows sensitivity to the sequential organization of sound elements in songs. Until recently, such sensitivity has been found mostly in the sensorimotor nucleus HVC, using the bird's own song as stimulus (Margoliash, 1986; Margoliash and Fortune, 1992; Lewicki and Konishi, 1995). The primary auditory area known as Field L shows less sensitivity than HVC neurons to manipulations in the order of syllables (Lewicki and Arthur, 1996). Sensitivity to the sequential order in which sounds occur could be a functional property that characterizes high-level auditory regions. However, very few studies have examined neuronal responses to changes in the ordering relationships of elements within sequences by recording single-unit activity in the auditory cortex of mammalian species (Weinberger and McKenna, 1988; McKenna et al., 1989; Kikuchi et al., 2017).

In conclusion, the present study provides new insights into properties of high-order auditory neurons by showing sensitivity to sequential organization of natural communication signals. Our findings establish the songbird as a model system for deciphering basic neural processes by which, not only ordering relationships, but also the global structure can be encoded.

References

- Abe K, Watanabe D (2011) Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nat Neurosci* 14:1067–1074.
- Beckers GJ, Gahr M (2010) Neural processing of short-term recurrence in songbird vocal communication. *PLoS One* 5:e11129.
- Beckers GJ, Gahr M (2012) Large-scale synchronized activity during vocal deviance detection in the zebra finch auditory forebrain. *J Neurosci* 32:10594–10608.
- Beckers GJ, Berwick RC, Okanoya K, Bolhuis JJ (2017) What do animals learn in artificial grammar studies? *Neurosci Biobehav Rev* 81:238–246.
- Bekinschtein TA, Dehaene S, Rohaut B, Tadel F, Cohen L, Naccache L (2009) Neural signature of the conscious processing of auditory regularities. *Proc Natl Acad Sci U S A* 106:1672–1677.
- Berwick RC, Okanoya K, Beckers GJ, Bolhuis JJ (2011) Songs to syntax: the linguistics of birdsong. *Trends Cogn Sci* 15:113–121.
- Briefer EF, Rybak F, Aubin T (2013) Does true syntax or simple auditory object support the role of skylark song dialect? *Anim Behav* 86:1131–1137.
- Chen J, ten Cate C (2015) Zebra finches can use positional and transitional cues to distinguish vocal element strings. *Behav Process* 117:29–34.
- Chew SJ, Mello C, Nottebohm F, Jarvis E, Vicario DS (1995) Decrements in auditory responses to a repeated conspecific song are long-lasting and require two periods of protein synthesis in the songbird forebrain. *Proc Natl Acad Sci U S A* 92:3406–3410.
- Chew SJ, Vicario DS, Nottebohm F (1996) A large-capacity memory system that recognizes the calls and songs of individual birds. *Proc Natl Acad Sci U S A* 93:1950–1955.
- Clucas BA, Freeberg TM, Lucas JR (2004) Chick-a-dee call syntax, social context, and season affect vocal responses of Carolina chickadees (*Poecile carolinensis*). *Behav Ecol Sociobiol* 57:187–196.
- Comins JA, Gentner TQ (2010) Working memory for patterned sequences of auditory objects in a songbird. *Cognition* 117:38–53.
- Comins JA, Gentner TQ (2013) Perceptual categories enable pattern generalization in songbirds. *Cognition* 128:113–118.
- Comins JA, Gentner TQ (2014) Auditory temporal pattern learning by songbirds using maximal stimulus diversity and minimal repetition. *Anim Cogn* 17:1023–1030.
- Dehaene S, Meyniel F, Wacongne C, Wang L, Pallier C (2015) The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. *Neuron* 88:2–19.
- Doupe AJ (1997) Song- and order-selective neurons in the songbird anterior forebrain and their emergence during vocal development. *J Neurosci* 17:1147–1167.
- Fitch WT, Hauser MD (2004) Computational constraints on syntactic processing in a nonhuman primate. *Science* 303:377–380.
- Friederici AD, Bahlmann J, Heim S, Schubotz RI, Anwander A (2006) The brain differentiates human and non-human grammars: functional localization and structural connectivity. *Proc Natl Acad Sci U S A* 103:2458–2463.
- Gaucher Q, Edeline JM (2015) Stimulus-specific effects of noradrenaline in auditory cortex: implications for the discrimination of communication sounds. *J Physiol* 593:1003–1020.
- Gaucher Q, Huetz C, Gourévitch B, Edeline JM (2013) Cortical inhibition reduces information redundancy at presentation of communication sounds in the primary auditory cortex. *J Neurosci* 33:10713–10728.
- Gentner TQ (2008) Temporal scales of auditory objects underlying bird-song vocal recognition. *J Acoust Soc Am* 124:1350–1359.
- Gentner TQ, Fenn KM, Margoliash D, Nusbaum HC (2006) Recursive syntactic pattern learning by songbirds. *Nature* 440:1204–1207.
- Ghirlanda S, Lind J, Enquist M (2017) Memory for stimulus sequences: a divide between humans and other animals? *R Soc Open Sci* 4:161011.
- Holland J, Dabelsteen T, Paris AL (2000) Coding in the song of the wren: importance of rhythmicity, syntax and element structure. *Anim Behav* 60:463–470.
- Khouri L, Nelken I (2015) Detecting the unexpected. *Curr Opin Neurobiol* 35:142–147.
- Kiggins JT, Comins JA, Gentner TQ (2012) Targets for a comparative neurobiology of language. *Front Evol Neurosci* 4:6.
- Kikuchi Y, Attaheri A, Wilson B, Rhone AE, Nourski KV, Gander PE, Kovach CK, Kawasaki H, Griffiths TD, Howard MA 3rd, Petkov CI (2017) Sequence learning modulates neural responses and oscillatory coupling in human and monkey auditory cortex. *PLoS Biol* 15:e2000219.
- Kojima S, Doupe AJ (2008) Neural encoding of auditory temporal context in a songbird basal ganglia nucleus, and its independence of birds' song experience. *Eur J Neurosci* 27:1231–1244.
- Lewicki MS, Arthur BJ (1996) Hierarchical organization of auditory temporal context sensitivity. *J Neurosci* 16:6987–6998.
- Lewicki MS, Konishi M (1995) Mechanisms underlying the sensitivity of songbird forebrain neurons to temporal order. *Proc Natl Acad Sci U S A* 92:5582–5586.
- Lipkind D, Marcus GF, Bemis DK, Sasahara K, Jacoby N, Takahashi M, Suzuki K, Feher O, Ravbar P, Okanoya K, Tchernichovski O (2013) Stepwise

- acquisition of vocal combinatorial capacity in songbirds and human infants. *Nature* 498:104–108.
- Lu K, Vicario DS (2011) Toward a neurobiology of auditory object perception: what can we learn from the songbird forebrain? *Curr Zool* 57:671–683.
- Lu K, Vicario DS (2014) Statistical learning of recurring sound patterns encodes auditory objects in songbird forebrain. *Proc Natl Acad Sci U S A* 111:14553–14558.
- Malmierca MS, Sanchez-Vives MV, Escera C, Bendixen A (2014) Neuronal adaptation, novelty detection and regularity encoding in audition. *Front Syst Neurosci* 8:111.
- Margoliash D (1983) Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. *J Neurosci* 3:1039–1057.
- Margoliash D (1986) Preference for autogenous song by auditory neurons in a song system nucleus of the white-crowned sparrow. *J Neurosci* 6:1643–1661.
- Margoliash D, Fortune ES (1992) Temporal and harmonic combination-sensitive neurons in the zebra finch's HVC. *J Neurosci* 12:4309–4326.
- McKenna TM, Weinberger NM, Diamond DM (1989) Responses of single auditory cortical neurons to tone sequences. *Brain Res* 481:142–153.
- Meliza CD, Margoliash D (2012) Emergence of selectivity and tolerance in the avian auditory cortex. *J Neurosci* 32:15158–15168.
- Mello CV, Clayton DF (1994) Song-induced ZENK gene expression in auditory pathways of songbird brain and its relation to the song control system. *J Neurosci* 14:6652–6666.
- Mello C, Nottebohm F, Clayton D (1995) Repeated exposure to one song leads to a rapid and persistent decline in an immediate early gene's response to that song in zebra finch telencephalon. *J Neurosci* 15:6919–6925.
- Menardy F, Touiki K, Dutrieux G, Bozon B, Vignal C, Mathevon N, Del Negro C (2012) Social experience affects neuronal responses to male calls in adult female zebra finches. *Eur J Neurosci* 35:1322–1336.
- Menardy F, Giret N, Del Negro C (2014) The presence of an audience modulates responses to familiar call stimuli in the male zebra finch forebrain. *Eur J Neurosci* 40:3338–3350.
- Moorman S, Mello CV, Bolhuis JJ (2011) From songs to synapses: molecular mechanisms of birdsong memory. *BioEssays* 33:377–385.
- Neiworth JJ, London JM, Flynn MJ, Rupert DD, Alldritt O, Hyde C (2017) Artificial grammar learning in tamarins (*Saguinus oedipus*) in varying stimulus contexts. *J Comp Psychol* 131:128–138.
- Ono S, Okanoya K, Seki Y (2016) Hierarchical emergence of sequence sensitivity in the songbird auditory forebrain. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 202:163–183.
- Petkov CI, Wilson B (2012) On the pursuit of the brain network for proto-syntactic learning in non-human primates: conceptual issues and neurobiological hypotheses. *Philos Trans R Soc Lond B Biol Sci* 367:2077–2088.
- Phan ML, Pytte CL, Vicario DS (2006) Early auditory experience generates long-lasting memories that may subserve vocal learning in songbirds. *Proc Natl Acad Sci U S A* 103:1088–1093.
- Ribeiro S, Cecchi GA, Magnasco MO, Mello CV (1998) Toward a song code: evidence for a syllabic representation in the canary brain. *Neuron* 21:359–371.
- Schneider DM, Woolley SM (2013) Sparse and background-invariant coding of vocalizations in auditory scenes. *Neuron* 79:141–152.
- Seki Y, Suzuki K, Osawa AM, Okanoya K (2013) Songbirds and humans apply different strategies in a sound sequence discrimination task. *Front Psychol* 4:447.
- Smulders TV, Jarvis ED (2013) Different mechanisms are responsible for dishabituation of electrophysiological auditory responses to a change in acoustic identity than to a change in stimulus location. *Neurobiol Learn Mem* 106:163–176.
- Spierings MJ, ten Cate C (2016) Zebra finches as a model species to understand the roots of rhythm. *Front Neurosci* 10:345.
- Spierings M, de Weger A, ten Cate C (2015) Pauses enhance chunk recognition in song element strings by zebra finches. *Anim Cogn* 18:867–874.
- Stripling R, Volman SF, Clayton DF (1997) Response modulation in the zebra finch neostriatum: relationship to nuclear gene regulation. *J Neurosci* 17:3883–3893.
- Suzuki TN, Wheatcroft D, Griesser M (2016) Experimental evidence for compositional syntax in bird calls. *Nat Commun* 7:10986.
- Suzuki TN, Wheatcroft D, Griesser M (2017) Wild birds use an ordering rule to decode novel call sequences. *Curr Biol* 27:2331–2336.e3.
- ten Cate C (2017) Assessing the uniqueness of language: animal grammatical abilities take center stage. *Psychon Bull Rev* 24:91–96.
- Uhrig L, Dehaene S, Jarraya B (2014) A hierarchy of responses to auditory regularities in the macaque brain. *J Neurosci* 34:1127–1132.
- van Heijningen CA, de Visser J, Zuidema W, ten Cate C (2009) Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proc Natl Acad Sci U S A* 106:20538–20543.
- van Heijningen CA, Chen J, van Laatum I, van der Hulst B, ten Cate C (2013) Rule learning by zebra finches in an artificial grammar learning task: which rule? *Anim Cogn* 16:165–175.
- Van Meir V, Boumans T, De Groof G, Van Audekerke J, Smolders A, Scheunders P, Sijbers J, Verhoye M, Balthazart J, Van der Linden A (2005) Spatiotemporal properties of the BOLD response in the songbirds' auditory circuit during a variety of listening tasks. *Neuroimage* 25:1242–1255.
- Wakita M (2019) Auditory sequence perception in common marmosets (*Callithrix jacchus*). *Behav Process* 162:55–63.
- Weinberger NM, McKenna TM (1988) Sensitivity of single neurons in auditory cortex to contour: toward a neurophysiology of music perception. *Music Perception* 5:355–389.
- Wilson B, Kikuchi Y, Sun L, Hunter D, Dick F, Smith K, Thiele A, Griffiths TD, Marslen-Wilson WD, Petkov CI (2015) Auditory sequence processing reveals evolutionarily conserved regions of frontal cortex in macaques and humans. *Nat Commun* 6:8901.
- Wilson B, Marslen-Wilson WD, Petkov CI (2017) Conserved sequence processing in primate frontal cortex. *Trends Neurosci* 40:72–82.
- Woolley SM (2013) The songbird auditory system. In: *Animal models of speech and language disorders* (Helekar SA, ed), pp 61–88. New York: Springer.
- Yanagihara S, Yazaki-Sugiyama Y (2016) Auditory experience-dependent cortical circuit shaping for memory formation in bird song learning. *Nat Commun* 7:11946.
- Zaccarella E, Friederici AD (2017) The neurobiological nature of syntactic hierarchies. *Neurosci Biobehav Rev* 81:205–212.

Partie II :

STRUCTURE GLOBALE DU CHANT DE
CANARI : ORGANISATION EN
SÉQUENCES RÉCURRENTES (*CHUNKS*)
ET IMPORTANCE DE LA POSITION

Partie II : Structure globale du chant de canari : Organisation en séquences récurrentes (*chunks*) et importance de la position

Introduction

Lorsqu'on écoute quelqu'un parler dans notre langue maternelle, on sait distinguer, sans aucune difficulté, les mots les uns des autres. En revanche, si la langue parlée nous est étrangère, on ne perçoit qu'un flot continu de sons. L'apprentissage du langage nous a permis d'apprendre à segmenter les mots et à regrouper différentes syllabes de façon à composer les mots de la langue. D'après des études réalisées chez des humains adultes et chez de jeunes enfants, pour parvenir à cet apprentissage, différents indices ou informations seraient pris en compte, parmi lesquels figure la détection des probabilités de transition entre les syllabes (Estes et al. 2007; Saffran et al. 1996). En effet, les probabilités entre les syllabes au sein des mots sont plus fortes qu'entre la syllabe de fin d'un mot et celle du début du mot suivant. Ainsi, les mots se définissent par le type de signaux acoustiques qui les compose, en fonction de l'ordre de ces signaux acoustiques, mais aussi des probabilités entre chacun des signaux.

Si le diamant mandarin ne chante qu'un seul chant dans lequel l'enchaînement des syllabes est fixe, d'autres espèces produisent des chants au sein desquels les syllabes s'enchaînent de façon plus variable. Chez le Moineau du Japon, ou chez le Canari, l'enchaînement des syllabes ou phrases constitutives des chants repose sur des probabilités de transition variables.

L'étude précédente a montré que changer l'ordre des syllabes a un impact sur les réponses des neurones (Lehongre and Del Negro 2011). Cependant, elle laisse en suspens la question de savoir sur quels types de traitement repose cette sensibilité à l'ordre des syllabes. Afin d'aborder cette question, nous avons porté notre attention sur la structure du chant de canari qui produit des chants au sein desquels l'ordre des syllabes est, dans une certaine mesure, variable.

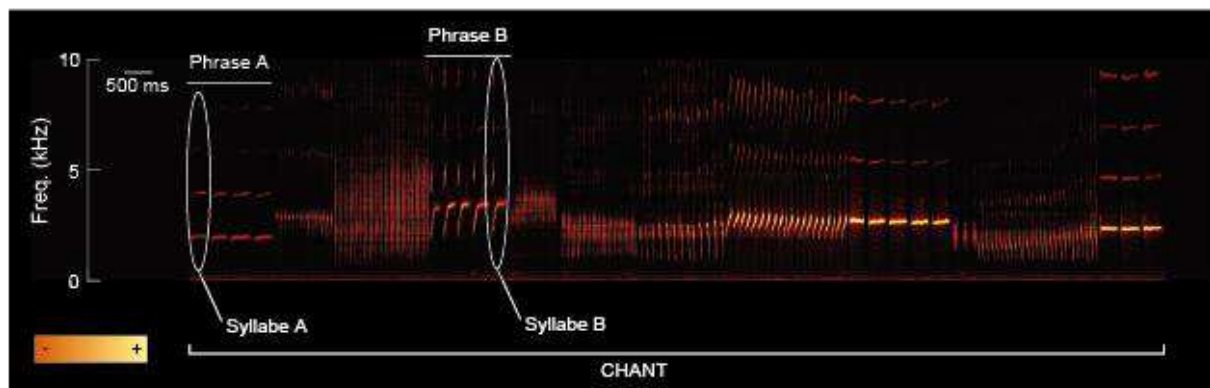


Figure 1: Spectrogramme d'un chant de canari adulte

Représentation d'un des chants d'un individu via un spectrogramme (fréquence (kHz) en ordonnée et durée (ms) en abscisse). L'énergie est représentée à travers un dégradé d'orange.

Le chant du canari est une succession de différentes phrases, chacune étant constituée de la répétition d'une même syllabe. La durée et le taux de répétitions par seconde des syllabes dans la phrase varient. Le canari produit des chants qui diffèrent par le type de phrases qui les composent ainsi que par leurs agencements.

C. La structure du chant chez le canari : plusieurs niveaux d'organisation

Parmi les espèces d'oiseaux chanteurs, certaines produisent des chants dont la structure peut présenter différents niveaux d'organisation (**Figure 1**). C'est le cas du chant du canari et notamment du canari domestique, qui est la forme domestiquée du Canari sauvage (*Serinus canaria*). Le canari est originaire des Iles Canaries. La structure du chant du canari domestique résulte de la sélection des oiseaux, par les éleveurs, depuis le 17^{ème} siècle.

1^{er} Niveau : la syllabe. Le chant du canari comme le chant de nombreuses espèces d'oiseaux chanteurs est constitué à partir d'un répertoire de signaux acoustiques complexes. Ces signaux acoustiques se distinguent les uns des autres par les paramètres acoustiques qui les caractérisent, et sont appelés, comme chez le diamant mandarin, des syllabes. Chaque individu possède un répertoire d'environ 15-20 types de syllabes qui diffèrent d'un individu à un autre. Parmi ces syllabes, environ 16% du répertoire d'un individu n'est chanté que par lui-même, les 84% restant étant partagé par un ou plusieurs congénères (Lehongre and Del Negro 2008; étude réalisée à partir des chants de 31 individus).

Outre le fait que certaines syllabes soient propres à l'oiseau, certaines sont considérées comme potentiellement porteuses d'informations nécessaires à la vie sociale de l'oiseau, nécessaires notamment à la reproduction et la défense du territoire. Par exemple, il a été montré que les chants constitués de syllabes parmi lesquelles figurent des syllabes composées de deux notes à rapide modulation de fréquence et répétées à un taux supérieur à 18/s sont préférées par les femelles (Vallet and Kreutzer 1995; Vallet et al. 1998). Ils induisent chez les femelles, plus de posture de sollicitation à la copulation que des chants n'en étant pas composés.

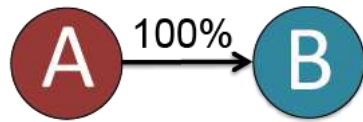
Il est important de préciser que les canaris domestiques sont, contrairement aux diamants mandarins, des oiseaux saisonniers. Ils modifient leur chant sur un rythme annuel (Leitner et al. 2001; Nottebohm et al. 1986; Nottebohm et al. 1987; Del Negro 2005). Les canaris adultes ne chantent pas ou peu en été. Ils commencent à produire des chants à l'automne, ces chants sont composés de syllabes dont les paramètres acoustiques (fréquence fondamentale par exemple) sont très variables (Alliende et al., 2017). Leur chant change tout au long de l'automne et de l'hiver et devient plus stable au printemps.

Les différentes syllabes produites présentent alors chacune une moindre variabilité de leurs paramètres acoustiques (notamment leur fréquence fondamentale). D'une année à l'autre, les canaris peuvent changer une partie (environ 30%) de leur répertoire de syllabes. Il n'est en revanche pas établi que cette modification du répertoire de syllabes, chaque année, témoigne d'un nouvel apprentissage à l'âge adulte ou si l'individu puise différemment chaque année dans un répertoire de syllabes auxquelles il a été exposé avant la période de maturité sexuelle (Lehongre et al. 2006; Lehongre et al. 2009).

2^{ème} niveau : la phrase. D'une espèce d'oiseaux chanteurs à une autre, la façon dont les syllabes s'enchaînent dans un chant diffère tels que des répétitions ou pas, l'existence de motifs c'est-à-dire d'enchaînement fixe de différentes syllabes qui peuvent être répétées ou revenir fréquemment. Chez le canari, chaque syllabe peut être produite une seule fois ou répétée de nombreuses fois pour constituer ce qui est appelé une phrase (**Figure 17**). De manière générale, les phrases durent autour de 800 ms. Les syllabes qui correspondent à de longs sifflements sont généralement répétées jusqu'à cinq fois dans une phrase. Au contraire, les syllabes constituées d'éléments à rapide modulation de fréquence seront répétées de nombreuses fois. D'un chant à un autre, si le nombre de répétitions d'une syllabe dans une phrase et la durée des phrases peuvent varier, les paramètres acoustiques de la syllabe qui est répétée plusieurs fois et la durée des silences sont stables.

3^{ème} niveau : le chant. Connaître le répertoire de phrases d'un canari ne suffit pas à définir comment un oiseau utilise son répertoire. En effet, les chants d'un même canari sont en moyenne constitués de 8 phrases (Lehongre et al. 2008), mais ce nombre peut varier de façon très importante. Certaines observations ont rapporté qu'un même canari peut produire des chants composés de trois jusqu'à une trentaine de phrases. À cela, il faut ajouter que, d'un chant à un autre pour un même individu, l'enchaînement des phrases n'est ni fixe, ni aléatoire.

FSA déterministe (dFSA)



FSA probabiliste (pFSA)

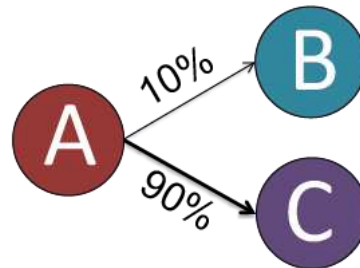


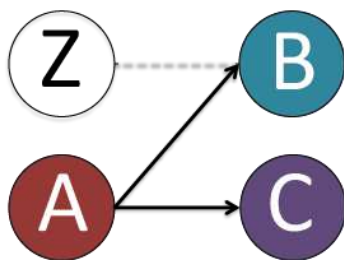
Figure 2 : Schématisation de deux types de grammaire, dFSA et pFSA

La grammaire d'automate à état fini (FSA) se subdivise en deux types. La force des probabilités est représentée par l'épaisseur de la flèche.

(Gauche) Schématisation d'une FSA déterministe (dFSA) : les transitions sont invariables ; les probabilités de transition sont de 1. Par exemple, la syllabe A a 100% de chance d'être suivie par la syllabe B.

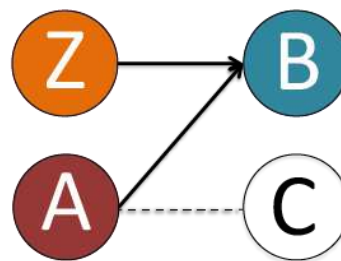
(Droite) Schématisation d'une FSA probabiliste (pFSA) : les transitions sont variables ; les probabilités de transition sont comprises entre 0 et 1. Par exemple, la syllabe A peut être suivie par deux syllabes différentes, 90% par B et 10% par C.

Probabilité de divergence (pDiv)



Sachant A, quelle est la probabilité que B suive?

Probabilité de convergence (pConv)



Sachant B, quelle est la probabilité que A précède?

Figure 3: Schématisation des probabilités de divergence et convergence

(Gauche) Probabilité de divergence (pDiv) : sachant A, la probabilité que B ou C suive.

(Droite) Probabilité de convergence (pConv) : sachant B, la probabilité que A ou Z ait précédé.

D. Enchaînement probabiliste des phrases dans les chants de canaris

Une autre raison pour laquelle le chant de canari a une structure complexe vient du fait que les chants ne sont pas une suite aléatoire des phrases du répertoire de l'individu. Pour définir comment les phrases s'organisent dans les chants, les chercheurs se sont tournés vers la linguistique, avec l'idée de déterminer notamment jusqu'à quel point les chants, de par leur structure, se comparent au langage humain.

Selon une conception actuelle (Kershbaum et al. 2014 ; Berwick et al. 2011), les types de grammaire utilisés pour construire les chants seraient simples. La construction des chants serait soumise à une grammaire dite régulière ou à état fini (*Finite-state automaton*, FSA) : les vocalisations se structurent selon un ensemble de règles simples à partir desquelles un automate peut construire, à partir d'état fini, les chants d'un individu donné. On observe deux principaux types de FSA (**Figure 2**). Le premier type selon lequel les probabilités de transition entre les syllabes ou phrases sont invariables est nommé FSA déterministe (dFSA ; **Figure 2 gauche**). Par exemple, la syllabe « A » est suivie dans 100% des cas par la syllabe « B ». Ce type de règles s'applique bien au chant du diamant mandarin pour lequel l'ordre des syllabes au sein de son chant est fixe. Le deuxième type, plus complexe, est dit probabiliste (pFSA ; **Figure 2 droite**) : les probabilités de transitions entre les syllabes ou phrases sont variables. La syllabe « A » peut, par exemple, être suivie dans 90% des cas par la syllabe « B » et dans 10% des cas par la syllabe « C ».

Pour une même transition AB, il existe deux types de probabilité (**Figure 3**) : la probabilité dite de divergence $p_{Div}(AB)$ (**Figure 3 gauche**) et celle de convergence $p_{Conv}(AB)$ (**Figure 3 droite**). En ce qui concerne la probabilité de divergence, c'est la probabilité, sachant A, d'avoir B (on se place au niveau de A, on prend en compte toutes les séquences de deux éléments commençant par A, et on calcule le nombre de fois que la séquence AB apparaît parmi toutes ces séquences). En ce qui concerne la probabilité de convergence, c'est la probabilité sachant B que A précède (on se place au niveau de B, on prend en compte toutes les séquences de deux éléments finissant par B, et on calcule le nombre de fois que la séquence AB apparaît parmi toutes ces séquences) lorsque l'on cherche à connaître les probabilités de syllabes menant à une syllabe cible (par exemple, sachant B quelles sont les probabilités d'avoir A avant).

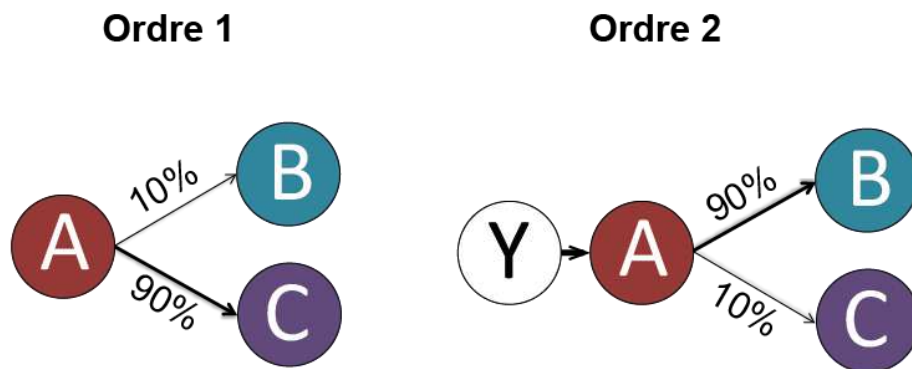


Figure 4: Exemple de probabilités de divergence d'une chaîne de Markov d'ordre 1 et 2

Ordre 1 : Probabilité que B ou C suive, sachant A (10% et 90% respectivement)

Ordre 2 : La probabilité que B ou C suive A dépend de ce qui précède A. Ici si Y précède A, la probabilité que B suive A est de 90% et non de 10%.

Ce type de grammaire pFSA est un exemple de processus Markovien c'est-à-dire faisant appel à une chaîne de Markov. Selon le processus Markovien appliqué aux probabilités de divergence, la probabilité d'occurrence de la future syllabe est déterminée non seulement par ce qui a précédé à plus ou moins long terme. On distingue plusieurs ordres dans les processus Markovien. L'ordre 1 correspond au précédent exemple où B suit A dans 90% des cas (**Figure 4 gauche**). Dans une chaîne de Markov d'ordre 2, les probabilités de transition entre la syllabe A et ce qui suit, la syllabe B ou C, dépend de ce qui précède la syllabe A. Par exemple, si avant la syllabe A, il y a la syllabe X alors les probabilités sont de 90% pour B et 10% pour C. Tandis que, si avant la syllabe A il y a la syllabe Y, les probabilités sont de 10% pour B et 90% pour C (**Figure 4 droite**). On peut alors s'intéresser aux probabilités de convergence ou de divergence d'ordre 1 à n , n pouvant théoriquement être infini (ce qui n'est évidemment pas le cas ni pour le chant des oiseaux ni pour le langage humain).

Chez certaines espèces d'oiseaux chanteurs, l'organisation des syllabes ou des phrases dans les chants suit un processus markovien. C'est par exemple le cas du moineau du Japon pour lequel l'enchaînement des syllabes suit un processus markovien d'ordre 1 (Fujimoto et al. 2011) sachant que des probabilités de transition entre certaines syllabes sont régies par un ordre supérieur à 1. De la même façon, chez le canari, qui a un répertoire plus important que le moineau du Japon, les enchaînements de phrases suivraient, pour la majorité d'entre eux, un processus markovien d'ordre 1. D'après Markowitz et al. 2013, certaines phrases (environ 10%) feraient cependant partie de séquences faisant appel à un processus markovien d'ordre supérieur à 1 et qui pourrait aller jusqu'à 4. L'agencement de ces syllabes serait alors gouverné par des dépendances à long terme qui s'étendraient sur plusieurs secondes (Markowitz et al. 2013). Cette étude met en avant les relations qui peuvent exister entre les différentes phrases du répertoire, et ainsi la complexité de la structure du chant de canari. Cependant, dans cette étude, l'accent est surtout mis sur les probabilités de transition (**Figure 5**) et certains aspects du chant n'ont pas été abordés. Cela conduit à penser que les chants de canari sont très variables dans leur composition et que l'étendue des séquences possibles est très grande.

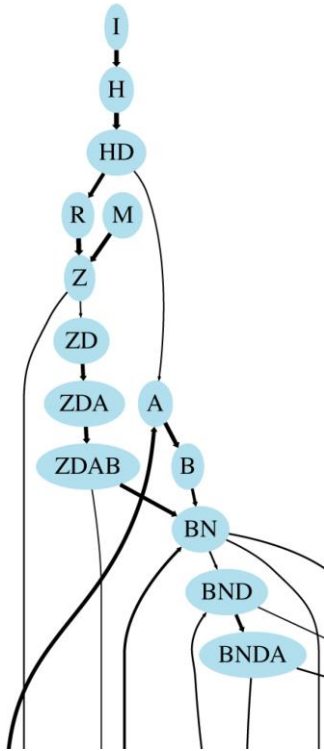


Figure 5 Représentation graphique (portion) du chant d'un canari selon les probabilités de convergence

Chaque phrase du chant est identifiée par une lettre. Les phrases pouvant s'enchaîner sont reliées par des flèches dont l'épaisseur est proportionnelle à la force des probabilités de divergence. Modifié de Markowitz et al. 2013.

E. But de l'étude

Le but de cette étude a été d'avancer dans la caractérisation de la structure des chants de canari. Pour cela, étant donné l'étude de Markowitz, nous avons pris en compte les probabilités de convergence et de divergence des séquences de plusieurs phrases présentes dans les chants (Markowitz et al., 2013). Cependant, ces probabilités ne résultent pas d'une prise en compte du nombre de fois que ces séquences apparaissent dans l'ensemble des chants. En effet, les probabilités de transition pour une séquence donnée sont calculées sur la base du nombre d'occurrences de chaque élément de part et d'autre de la transition. Nous avons également analysé le nombre d'occurrences de chacune des séquences de plusieurs phrases. Des études précédentes de l'équipe ont montré que, dans les chants de canaris, certaines phrases et certaines séquences de plusieurs phrases reviennent très fréquemment (Lehongre et al. 2008; Lehongre and Del Negro 2011).

Comme toute séquence, un chant présente un début et une fin. Si certaines phrases sont produites plus fréquemment et s'enchaînent préférentiellement avec d'autres phrases, on peut aussi supposer que ces phrases et/ou séquences de phrases pourraient avoir une position privilégiée dans les chants d'un même individu. Nous nous sommes intéressés à la position des différentes phrases et séquences de phrases dans les chants.

Enfin, étant donné les limites d'une description des chants de canari sur la seule base des probabilités, nous avons cherché une méthode qui apporte un nouvel aperçu de la structure du canari sur la façon dont s'enchaînent les phrases dans les chants sur la base des distances entre les différentes phrases.

Cette étude qui a été réalisée en collaboration avec Xavier Hinaut de l'équipe Mnemosyne (INRIA Bordeaux) nous a permis de mieux caractériser l'organisation globale sous la forme de *chunk* des chants de canaris et a constitué la première étape en perspective de l'étude de caractérisation des réponses neuronales à cette organisation particulière du chant (chapitre III).

Matériels et méthodes

F. Sujets et acquisition des chants

Les canaris utilisés pour l'ensemble des études étaient des adultes mâles (n=9) âgés de plus d'un an et provenant d'un élevage interne au laboratoire. L'enregistrement des chants a été réalisé entre les mois de mai et juillet pendant lesquels les canaris produisent leur chant le plus stéréotypé (Alliende et al., 2017; Leitner et al., 2001). Pour cela, les oiseaux sont placés dans une cage individuelle (l : 27 cm x L : 50 cm x h : 32 cm) installée dans un caisson d'isolation acoustique. Les oiseaux restent dans ce caisson pendant environ 1 semaine. Ils bénéficient d'un accès *ad libitum* à la nourriture et à l'eau et sous un régime de lumière de 12h de jour/12h de nuit. L'enregistrement des chants est réalisé avec un microphone Sennheiser (bande passante : 40-18000Hz) positionné dans le caisson à ~2 cm de la cage connecté à une interface audio (M-Audio Fast-track ultra 8R) elle-même reliée à un ordinateur. Les chants sont détectés automatiquement et enregistrés (taux d'échantillonnage de 22 050 Hz avec une taille d'échantillonnage de 16 bits) par le logiciel d'acquisition Sound Analysis Pro (Tchernichovski et al., 2000).

G. Analyse des chants

Création des répertoires individuels. Un chant, pour être considéré comme tel, doit être composé de plus de 2 phrases, et deux chants successifs sont séparés d'un silence minimum de 0,4s. Afin de réaliser une analyse des chants, le répertoire de phrases de chaque canari enregistré a été défini à partir des spectrogrammes (taille de la fenêtre : 1024 points, type de fenêtre : Hamming, échelle linéaire) visualisés à l'aide du logiciel Audacity (versions pour Macintosh et PC). Entre 50 et 300 chants ont été analysés pour chaque oiseau (moy=132). Plusieurs critères ont été pris en compte : les paramètres acoustiques et temporels qui les caractérisent, mais aussi le fait qu'elles apparaissent au moins une fois dans un autre contexte, c'est-à-dire précédées et suivies d'autres phrases que celles de la première fois. Chaque phrase est arbitrairement codée en utilisant une lettre suivie ou non d'un chiffre en fonction de la proximité acoustique de la phrase avec une autre du répertoire.

Structure des chants. Une fois le répertoire individuel de phrases établi, nous avons effectué un grand nombre d'analyses probabiliste sur l'ensemble des chants produits par chaque oiseau à l'aide de script personnalisé Python.

Nous avons calculé la probabilité d'occurrence de chaque phrase du répertoire de l'oiseau. Celle-ci correspond au nombre d'apparition de la phrase dans les chants, Ph^{tot} , par rapport au nombre total de chants, Ch^{tot} :

$$pOcc(ph) = Ph^{tot} / Ch^{tot}$$

Nous avons calculé la probabilité d'occurrence des séquences de phrases produites par l'oiseau. Le nombre de phrases incluses dans ces séquences varient de deux jusqu'au nombre maximum de phrases que peut contenir le chant produit par l'oiseau. Celle-ci correspond au nombre d'apparition de la phrase dans les chants, Seq^{tot} , par rapport au nombre total de chants, Ch^{tot} :

$$pOcc(seq) = Seq^{tot} / Ch^{tot}$$

Nous avons calculé les probabilités de divergence et de convergence pour chaque transition produite par l'oiseau. Une probabilité de transition correspond au rapport entre le nombre d'apparitions totales de la phrase par rapport à celui d'apparition dans la transition. Pour la transition x-y, la probabilité de divergence, $pDiv$, est le rapport entre le nombre d'apparitions de la phrase x, N^x , et le nombre d'apparitions de la transition x-y, N^{x-y} , la probabilité de convergence, $pConv$, est le rapport entre le nombre d'apparitions de la syllabe y, N^y , et le nombre d'apparitions de la transition x-y, N^{x-y} .

$$pDiv = N^{x-y} / N^x$$

$$pConv = N^{x-y} / N^y$$

Ces probabilités de transition ont été calculées pour des séquences de 1 à n phrases (où n varie selon la taille maximum des chants de l'individu).

Nous avons calculé la position des phrases selon deux méthodes : absolue et relative. La position absolue de la phrase x, PA_x , consiste à calculer le nombre de phrases dans chaque chant et d'attribuer simplement un nombre correspondant à sa position. Dans un chant ABCD, la position absolue de la phrase A, $PA_A=1$, $PA_B=2$, etc. Les chants ne faisant pas tous la même longueur, la dernière phrase peut être à une position différente selon la longueur maximum du chant.

La position relative consiste à normaliser chaque chant en fonction de sa longueur maximum. Quelle que soit la longueur du chant, il est borné entre 0 et 1. Pour calculer la

position relative de la phrase x , PR_x , il suffit de diviser sa position absolue, PA_x par le nombre maximum de phrases dans le chant, P_T :

$$PR_x = PA_x / P_T$$

Par exemple pour le chant ABCD, $P_T=4$, donc les positions relatives de A, B, C et D sont, respectivement $PR_A = 0,25$, $PR_B = 0,5$, $PR_C = 0,75$, $PR_D = 1$.

Représentation graphique des chants. Le *multidimensional scaling* (MDS) est une méthode de réduction de dimension. À partir d'une matrice de distances entre individus, ici des phrases (le début et la fin du chant étant considérés comme des phrases virtuelles), cette méthode permet de rendre compte de leurs positions relatives dans un repère euclidien à deux dimensions.

Partant de n points $x_1 \dots x_n \in \mathbb{R}^d$ dont seulement les distances $\|x_i - x_j\|$ sont connues, et d'une dimension voulue $N < d$, on souhaite trouver $y_1 \dots y_n \in \mathbb{R}^N$ qui respectent les distances précédentes, c'est-à-dire : $(\|x_i - x_j\|) \approx (\|y_i - y_j\|)$. Cela correspond donc au problème d'optimisation qui consiste à déterminer le min x_1, \dots, x_n

$$\sum_{i < j} (\|x_i - x_j\| - \delta_{i,j})^2 .$$

La matrice est construite selon le nombre de phrases qui sépare un couple de phrases données. Elle a été calculée dans le sens X1 vers X2 et X2 vers X1. En effet, une phrase pouvant se répéter, ces deux mesures peuvent ne pas être symétriques.

Par exemple sur la séquence **A**CD**C**D**C**D**A****B**, pour la paire **AB**

A -> **B** : 7 (1^{er} A et B), 1 (2^{ème} A et B)

B -> **A** : 1 (car il n'y a qu'un seul B)

Pour chaque paire, la médiane de chaque distance mesurée est calculée pour créer la matrice.

Par ailleurs, certaines paires de phrases peuvent ne pas exister, de ce fait il a été estimé une distance d'environ 1,5 fois la longueur moyenne des chants. Cette distance a pu être aussi déduite par égalité triangulaire afin de limiter l'impact de ces non-paires sur celles existantes.

Résultats

H. Occurrence des phrases et séquences de phrases

L'étude du répertoire de chaque individu indique qu'en moyenne, les oiseaux ont chanté $19,7 \pm 3,0$ (moyenne \pm SEM) types de phrases. Parmi ces phrases, certaines sont fréquemment chantées alors que d'autres le sont moins. La **Figure 6** présente les probabilités d'occurrence de chaque phrase dans l'ensemble des chants d'un individu. Cet oiseau produit 19 phrases différentes, certaines phrases (T2, K et X) sont produites dans plus de 80% des chants alors que d'autres (O1) le sont dans moins de 20% des chants. Chez cet oiseau, 11 phrases sur les 19 de son répertoire (soit 57,9%) apparaissent dans plus de 50% des chants, ce qui reflète les proportions que l'on retrouve chez l'ensemble des oiseaux étudiés ($56 \pm 3\%$). Ainsi, comme déjà mis en avant chez le canari, le choix des phrases qui composent son chant est différent d'une sélection aléatoire (Markowitz et al. 2013; Lehongre et al. 2008; Lehongre and Del Negro 2011). Après avoir identifié l'ensemble des phrases produites par chaque oiseau, nous avons recherché toutes les différentes combinaisons de n phrases (n, de 1 à 7).

Comme le montre la figure qui présente les résultats obtenus pour un individu, le nombre de séquences de n phrases produites par un oiseau est moins important que celui attendu si l'on considère toutes les combinaisons possibles (**Figure 7A et B**). Chez cet individu représentatif, dont le répertoire est composé de 19 phrases différentes de son répertoire, au lieu des 93024 combinaisons possibles, l'oiseau en produit seulement 63. Une analyse descriptive de l'agencement des phrases dans les différents chants montre que certaines séquences de phrases apparaissent dans plusieurs chants. Par exemple, on peut constater que ce même oiseau inclut fréquemment dans ses chants des séquences composées des phrases A1-B (**Figure 7C**). Pour affiner cette analyse, nous avons déterminé le pourcentage d'occurrence des combinaisons de deux phrases (**Figure 7D**) dans les 175 chants analysés. On constate que l'oiseau utilise certaines combinaisons de façon très fréquente : par exemple les séquences de phrases A1-B et B-C apparaissent dans plus de 40% des chants alors que de nombreuses autres séquences apparaissent dans moins de 3% des chants, autrement dit 1 seule fois. L'analyse des chants des différents individus utilisés dans cette étude, nous a montré que, sur l'ensemble des

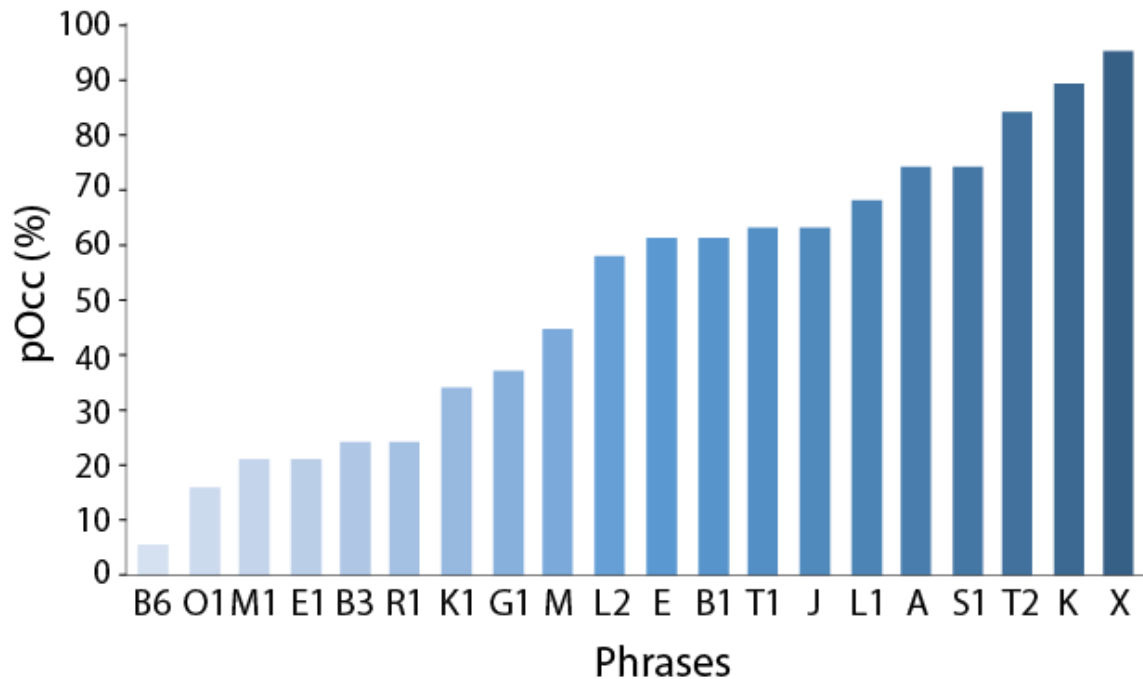


Figure 6 : Pourcentage d'occurrence des différentes phrases du répertoire d'un individu représentatif.

Chaque type de phrase a été codée par une lettre suivie ou non d'un chiffre. Les pourcentages d'occurrence des phrases sont basés sur le nombre de fois que la phrase a été chantée par rapport au nombre de chants analysés au total. Ces données ont été obtenues pour l'oiseau « Vert4 ». On peut noter que certaines phrases sont très peu fréquentes (ex : M1) tandis que d'autres sont présentes dans la plupart des chants (ex : X).

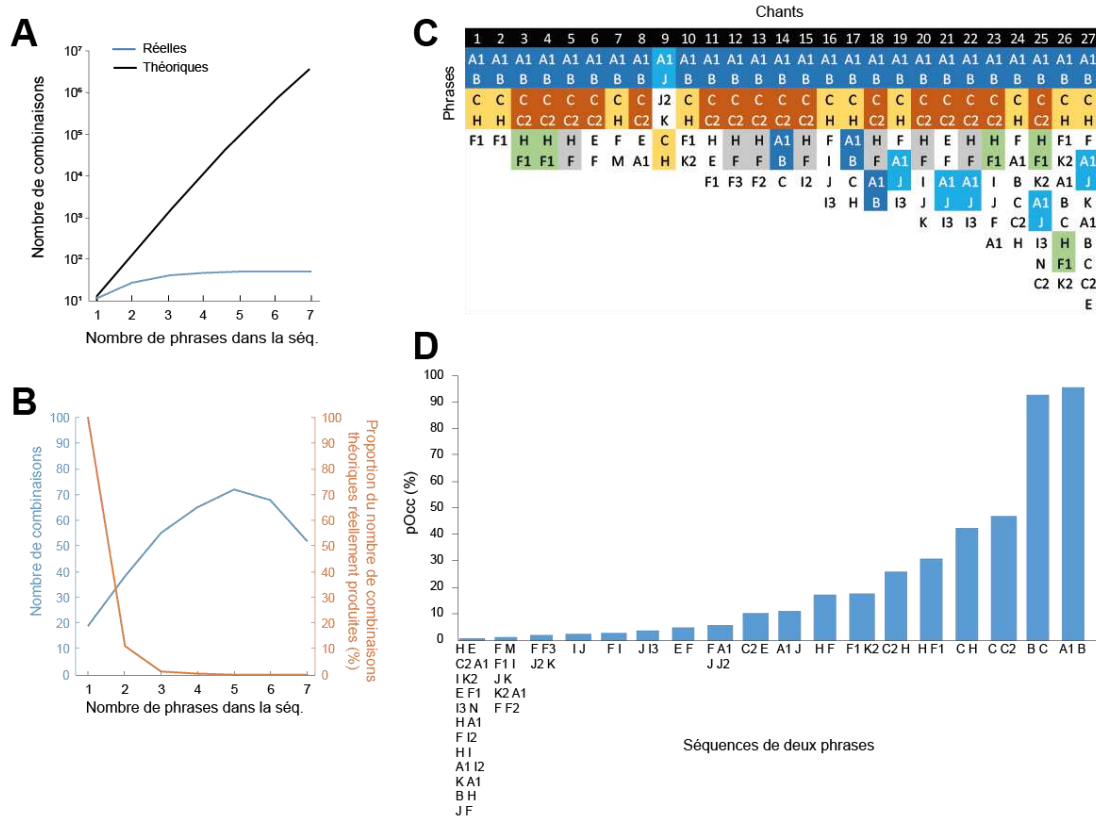


Figure 7 : Enchaînement des phrases selon des séquences chez un individu représentatif.

(A) Nombre de séquences réellement produites par un oiseau dont le répertoire est composé de 19 phrases (courbe bleue, même courbe représentée en B en échelle linéaire) et nombre de combinaisons théoriques possibles (courbe noire) en fonction du nombre de phrases présentes dans une séquence. Par exemple, pour des séquences de 2 et 3 phrases, le nombre de combinaisons théoriques est $19 \times 18 = 342$ ou encore $19 \times 18 \times 17 = 5814$, respectivement.

(B) Le nombre de combinaisons réellement produites par un oiseau (même courbe qu'en A) et la proportion de séquences réellement produites par rapport aux valeurs théoriques (courbe orange) : on observe le maximum pour les séquences contenant 5 phrases.

(C) Ce diagramme représente en colonne une sélection de 21 chants parmi les 175 chants (numérotation de 1 à 27) d'un oiseau représentatif. Chaque ligne correspond à une phrase identifiée par une lettre suivie ou non par un chiffre. Les séquences de deux phrases identiques sont représentées par une même couleur.

(D) Pourcentage d'occurrence des différentes séquences de deux phrases produites par l'oiseau dans les différents chants analysés.

canaris, chaque oiseau inclut dans 50% ou plus de leur chant au moins un enchaînement de deux, trois, voire même de quatre ou cinq phrases différentes (nombre moyen de séquences de 2 phrases \pm SEM : $2,8 \pm 1,3$). Ces séquences sont qualifiées récurrentes et avaient déjà été observées dans des études précédentes menées dans l'équipe (Del Negro, 2005; Lehongre et al., 2008).

I. Probabilités de transition

Ces séquences récurrentes ne se caractérisent pas seulement par leur occurrence. En calculant les probabilités de transition, divergence et convergence (voir Matériels et Méthodes) au sein de séquences de deux phrases, nous avons voulu déterminer si les séquences les plus récurrentes pouvaient être caractérisées par leurs probabilités de transition. Pour l'ensemble des oiseaux, on constate une corrélation positive entre l'occurrence des séquences et la probabilité soit de divergence, soit de convergence au sein de ces séquences (**Tableau 1; Figure 8**). Au plus une séquence de deux phrases apparaît souvent, au plus les probabilités entre les deux phrases est forte.

Les phrases constitutives de séquences de deux phrases pour lesquelles les probabilités de divergence et de convergence sont supérieures à 0,5 représentent en moyenne la moitié des phrases ($50,0 \pm 0,4\%$) des répertoires de chaque oiseau. En ce qui concerne les relations entre les probabilités de divergence et celles de convergence, on observe une corrélation significative que pour la moitié des oiseaux analysés et ce qui ne permet donc pas d'extraire un schéma global.

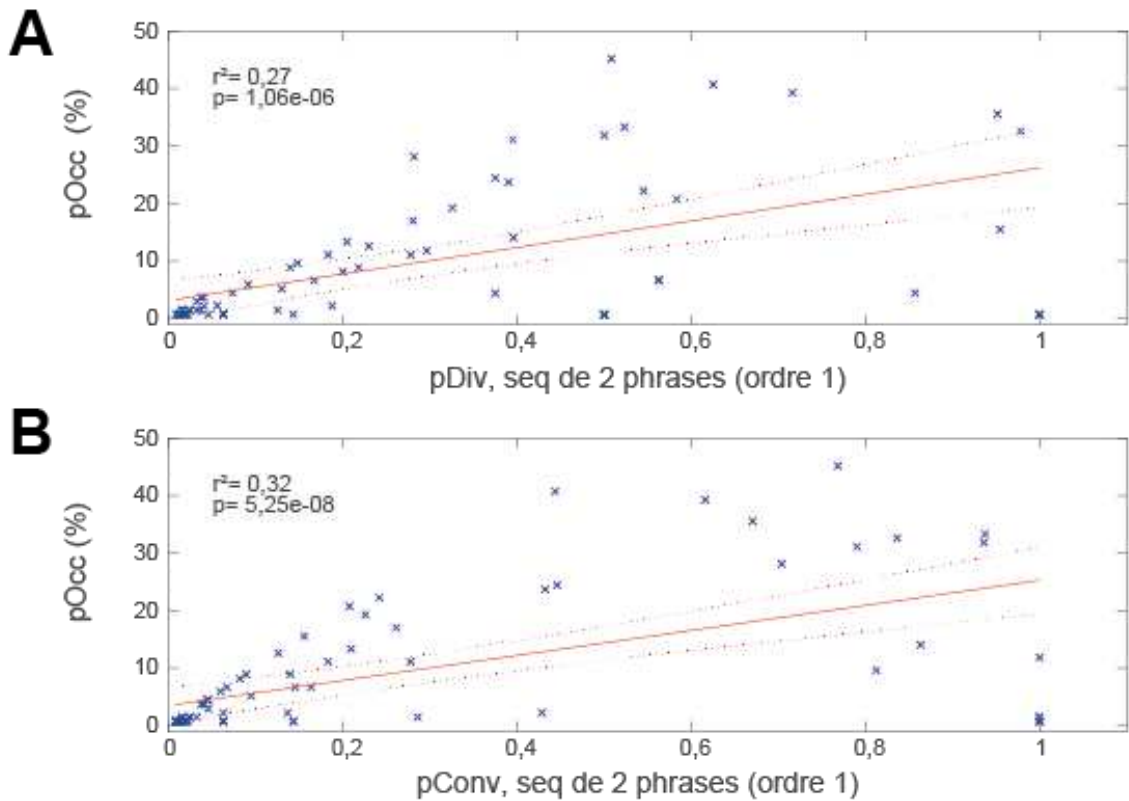


Figure 8: Corrélation entre les probabilités de transition et l'occurrence des séquences de deux phrases produites par un oiseau.

Point bleu : séquence de deux phrases produites par l'oiseau ; courbe en rouge : la courbe de tendance.

(A) Corrélation positive entre l'occurrence et les probabilités de divergence ($p < 0.001$)

(B) Corrélation positive entre l'occurrence et les probabilités de convergence ($p < 0.001$)

		oiseaux							
		1	2	3	4	5	6	7	8
pDiv	R ²	0,5	0,5	0,1	0,2	0,5	0,6	0,3	0,2
	p	***	***	***	***	***	***	***	***
pConv	R ²	0,6	0,5	0,2	0,2	0,5	0,5	0,3	0,2
	p	***	***	***	***	***	***	***	***

Tableau 1: Récapitulatif des indices de corrélation entre l'occurrence et les probabilités de transition

Ligne : valeurs du R² et du p ; pDiv : probabilité de divergence ; pConv : probabilité de convergence. ; colonne : oiseau. (***) équivaut à < 0.001)

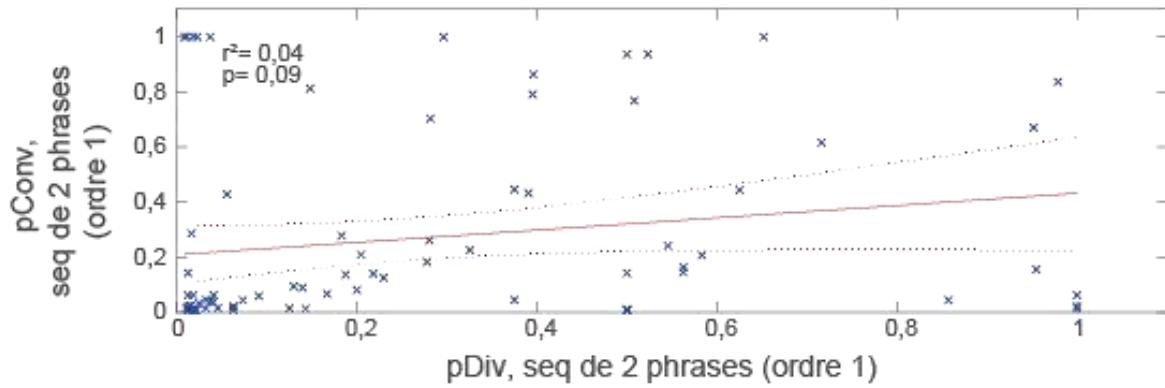


Figure 9: Absence de corrélation entre les probabilités de divergence et les probabilités de convergence pour les séquences de deux phrases (ordre 1).

Abcisse : probabilité de divergence (pDiv)

Ordonnée : probabilité de convergence (pConv)

Exemple de corrélation non significative entre les probabilités de convergence et de divergence pour l'ensemble des séquences de deux phrases produites par l'oiseau.

Point bleu : séquence de deux phrases ; courbe en rouge : courbe de tendance

		oiseaux							
		1	2	3	4	5	6	7	8
pDiv vs	R ²	0,323	0,36	0,005	0,002	0,335	0,295	0,036	0,056
pConv	p	***	***	ns	ns	***	***	ns	ns

Tableau 2: Récapitulatif des indices de corrélation entre les probabilités de divergence et les probabilités de convergence pour les séquences de 2 phrases pour les 8 oiseaux.

Ligne : valeurs du R² et du p; pDiv : probabilité de divergence ; pConv : probabilité de convergence. ; colonne : oiseau. (***) équivaut à <0.001 ; ns>0.05)

Parmi les séquences de deux phrases dont les transitions montrent de fortes probabilités de divergence ($>0,5$) et convergence ($>0,5$), on retrouve essentiellement les séquences avec les plus fortes occurrences, comme le démontre la corrélation précédemment présentée, mais pas seulement. Des probabilités de transition fortes au sein d'une séquence ne sont pas obligatoirement associées à une forte occurrence de cette même séquence. Certaines ont une occurrence faible et des probabilités de transition fortes, car les phrases qui la constituent, une fois chantées, sont nécessairement associées dans cette séquence donnée. De la même façon, deux séquences partageant une phrase commune peuvent être toutes deux fréquemment chantées, mais les probabilités de transition impliquant cette phrase commune seront inférieures ou égales à 0,5 pour au moins l'une des deux séquences. Par exemple, chez un des oiseaux, une des phrases, la phrase B34, arrive dans quasiment tous les chants (122 sur 124 chants) et peut être précédée par trois phrases différentes (parmi lesquelles les phrases B2 et D1). De ce fait, aucune des transitions avec la phrase B34 n'est forte ($\geq 0,5$; $p_{\text{Conv}}(B2, B34) = 0,35$; $p_{\text{Div}}(B2, B34) = 0,49$; $p_{\text{Conv}}(D1, B34) = 0,35$; $p_{\text{Div}}(D1, B34) = 0,42$). Pourtant, par exemple la séquence B2-B34 arrive dans 43 chants sur 124, la séquence D1-B34 arrive dans 45 chants. Ainsi, les probabilités de divergence et de convergence entre deux phrases consécutives ne caractérisent que, dans une certaine mesure, les séquences récurrentes.

J. Dépendances à long terme

Les résultats ont été obtenus par une analyse des séquences de deux phrases. Or, l'étude menée par Markowitz et collaborateurs (2013) propose que l'organisation des phrases dans les chants de canaris pourrait reposer, au moins en partie, sur des chaînes de Markov d'ordre supérieur 1, pouvant aller jusqu'à 4.

Une chaîne de Markov est une séquence de n éléments X , ici des phrases, dont les probabilités de la transition entre les phrases X_n et X_{n+1} , sont calculées en fonction de la phrase X_{n-y} où y varie selon l'ordre observé.

Nous nous sommes intéressés aux séquences de plus de deux phrases. En réitérant l'analyse de la corrélation entre l'occurrence et les probabilités de transition, les résultats pour les séquences de trois phrases diffèrent de ceux pour les séquences de deux phrases.

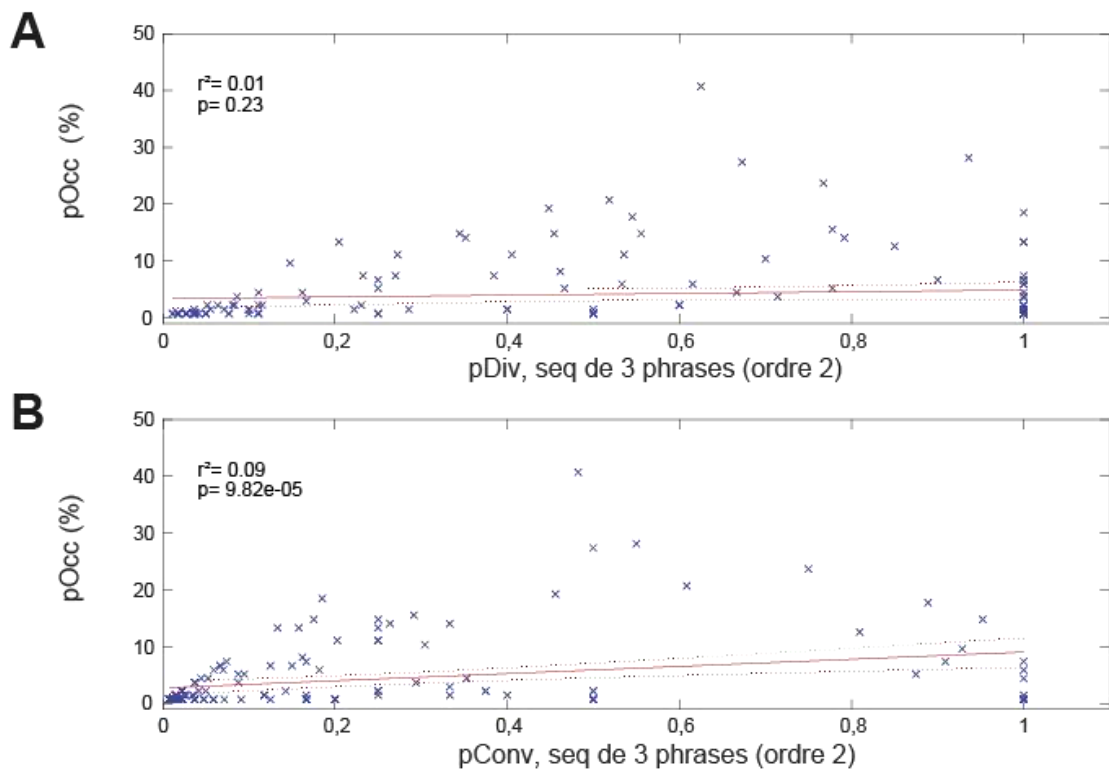


Figure 10: Corrélation entre les probabilités de transition et l'occurrence des séquences de trois phrases produites par un oiseau.

- (A) Pas de corrélation entre occurrence ($pOcc$) et probabilités de divergence ($pDiv$)
- (B) Corrélation positive entre occurrence ($pOcc$) et probabilités de convergence ($pConv$)

Point bleu : séquence de trois phrases ; Courbe en rouge : courbe de tendance

		oiseaux							
		1	2	3	4	5	6	7	8
pDiv	R ²	0,012	0,01	0,009	0,016	0,076	0,034	0,002	0,007
	p	ns	ns	ns	ns	ns	ns	ns	ns
pConv	R ²	0,128	0,16	0,095	0,057	0,38	0,19	0,068	0,177
	p	***	***	***	**	***	**	***	**

Tableau 3 : Récapitulatif des indices de corrélation entre les probabilités de transition et occurrence des séquences de trois phrases pour les 8 oiseaux.

Ligne : valeurs du R² et du p; pDiv : probabilité de divergence ; pConv : probabilité de convergence. ; colonne : oiseau. (***) équivaut à <0.001 ; ** équivaut à <0.01 ; ns>0.05)

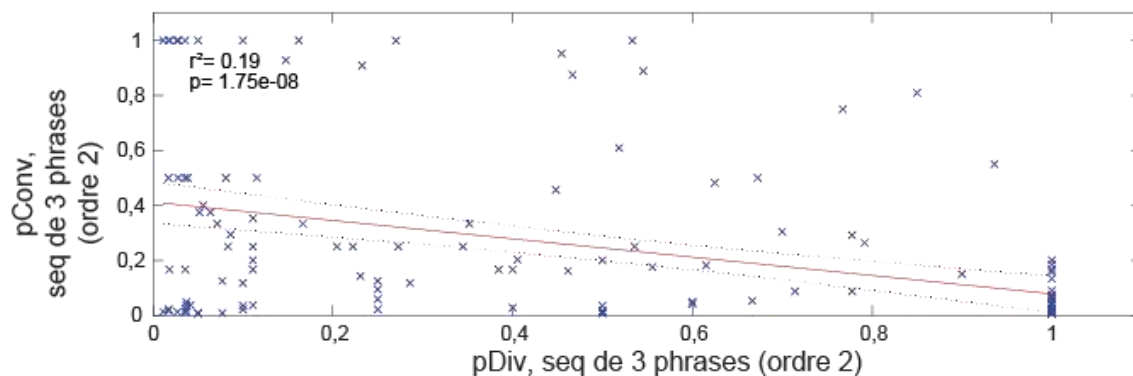


Figure 11 : Exemple d'absence de corrélation entre probabilités de divergence et probabilités de convergence pour les séquences de trois phrases.

Abscisse : probabilité de divergence (pDiv)

Ordonnée : probabilité de convergence (pConv)

Exemple de corrélation non significative entre les probabilités de convergence et de divergence pour l'ensemble des séquences de deux phrases produites par l'oiseau.

Point bleu : séquence de deux phrases ; courbe en rouge : courbe de tendance

		oiseaux							
		1	2	3	4	5	6	7	8
pDiv vs	R^2	0,27	0,04	0,19	0,084	0,03	0,11	0,15	0,06
pConv	p	***	**	***	**	ns	**	***	**

Tableau 4: Récapitulatif des indices de corrélation entre les probabilités de transition pour les séquences de trois phrases pour les 8 oiseaux.

Ligne : valeurs du R^2 et du p ; pDiv : probabilité de divergence ; pConv : probabilité de convergence. ; colonne : oiseau. (** équivaut à <0.01 ; ns >0.05)

Seule une corrélation entre l'occurrence et la probabilité de convergence apparaît significative (**Tableau 3; Figure 10**). Aucune corrélation entre l'occurrence et la probabilité de divergence n'est significative dans les séquences de trois phrases. De plus, au contraire des séquences de deux phrases, on constate une corrélation négative entre les probabilités de convergence et de divergence sur presque l'ensemble des oiseaux analysés (7 oiseaux sur 8) (**Tableau 4 ; Figure 11**). Ce résultat implique une différence selon que l'on considère l'ordre 1 ou 2 au sein d'une séquence et apporte des arguments supplémentaires dans le fait que le chant des canaris suit un processus markovien.

Chez un des oiseaux que nous avons enregistrés, l'agencement de certaines phrases dépendait d'un processus markovien d'ordre 3. En effet, dans l'exemple qui nous intéresse, on retrouve principalement deux phrases pouvant précéder ou suivre la séquence H-E. La séquence H-E est présente dans les chants et dans certains chants, elle est produite deux fois (588 fois pour 458 chants) et les probabilités de convergence et de divergence sont très fortes (0,99 et 0,95 respectivement). Elle peut être précédée soit par la phrase C, soit par la phrase Q, et être suivie soit par la phrase R, soit par la phrase N. Les probabilités de divergence de la séquence H-E indiquent que l'oiseau a une préférence pour chanter la séquence H-E-R plutôt que la séquence H-E-N (**Figure 12 gauche**). Cependant, on constate que ces probabilités de divergence sont impactées, si l'on considère les phrases qui précèdent la séquence H-E (**Figure 12 droite**). Si H-E est précédée de la phrase C, alors la phrase R suivra majoritairement la séquence H-E par rapport à la phrase N. Au contraire, si H-E est précédée de la phrase Q, alors la phrase N suivra majoritairement la séquence H-E par rapport à la phrase R. L'organisation des chants de canaris, voire plus précisément de certaines séquences récurrentes peut donc être liée à des dépendances à long terme.

K. Importance de la position des phrases

Comme tout événement séquentiel, le chant comporte un début et une fin. La question qui peut se poser alors est de savoir si toutes les phrases peuvent être au début ou seulement certaines, de même pour la fin.

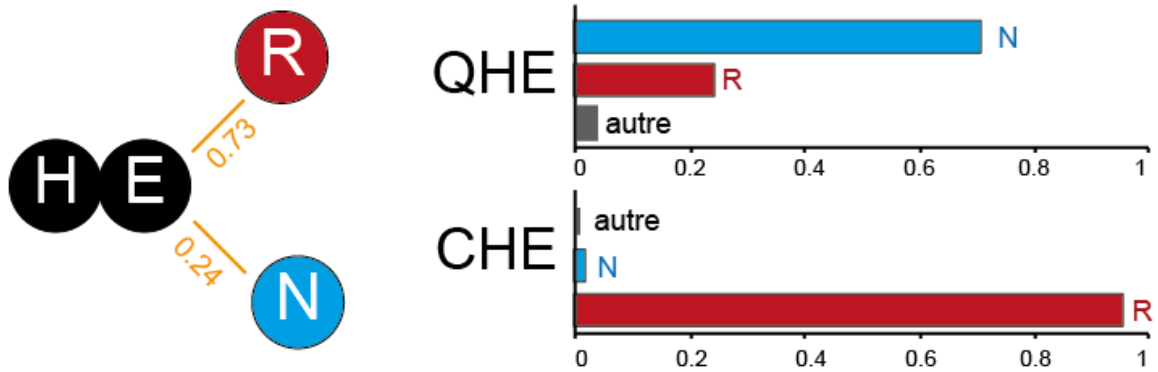


Figure 12: Exemple de dépendance à long terme

(Gauche) Probabilités de divergence des séquences H-E-R et H-E-N avec une prédominance de la séquence H-E-R.

(Droite) Impact de la phrase précédent la séquence HE, Q ou C, sur les probabilités de divergence de la transition E-N et E-R, précédemment observé.

On n'observe dans les chants de chaque oiseau enregistré, qu'ils ne débutent que par certains types de phrases, trois ou quatre différentes alors que le répertoire peut être d'environ 15-20 phrases. Si, en revanche, le nombre de phrases par lequel le chant finit est plus important, le calcul des probabilités de convergence prenant en compte la fin, nous indique que la transition pour certaines phrases avec la fin est supérieure à 0,5 indiquant que les chants peuvent finir préférentiellement par ces phrases. Ainsi, ces résultats suggèrent que certaines phrases peuvent se trouver préférentiellement au début ou à la fin.

Cela nous a amenés à analyser, pour chaque oiseau, les positions des différentes phrases dans les différents chants enregistrés. La position peut être calculée de façon absolue (PA) ou relative (PR) (voir Matériels et Méthodes). La position absolue nous permet de distinguer les phrases, qui, quelle que soit la longueur du chant, sont toujours à la même place. Par exemple, que le chant soit un enchaînement de trois ou vingt phrases, est-ce que la première phrase sera toujours identique ? La position relative, quant à elle, normalise la longueur de chaque chant à 1. Elle permet de déterminer la position de chaque phrase et, si, quel que soit le nombre de phrases que comporte le chant, la dernière phrase sera toujours identique.

Les exemples représentatifs de position présentés dans la **Figure 13** pour un oiseau permettent d'identifier assez facilement certaines phrases du début (B2, A1) et de la fin (D2, C1). On constate, par ailleurs, de fortes différences d'une phrase à une autre : certaines phrases présentent une variabilité très faible de leur positionnement dans le chant (par exemple B2, D2, C1), d'autres présentant une grande variabilité (par exemple B34, C3, B33). Il est intéressant de constater que les phrases présentant les variabilités de position les plus faibles sont celles par lesquelles l'oiseau débute et termine préférentiellement ses chants et que celles présentant les variabilités les plus fortes sont plutôt celles produites en milieu de chant. Sur l'ensemble des oiseaux, l'organisation des phrases entre le début et la fin du chant, selon leur position relative, n'est pas aléatoire.

Si certaines phrases ont une position préférentielle dans le chant d'un individu, on peut alors se poser la question de savoir si certaines séquences de phrases pourraient aussi avoir des positions préférentielles dans le chant. Pour répondre à cette question, nous avons réalisé la même analyse de position que précédemment en incluant les séquences les plus récurrentes identifiées, nommées ckX où X est un nombre (ck1) (**Figure 14**). Dans cet exemple, se basant sur les résultats

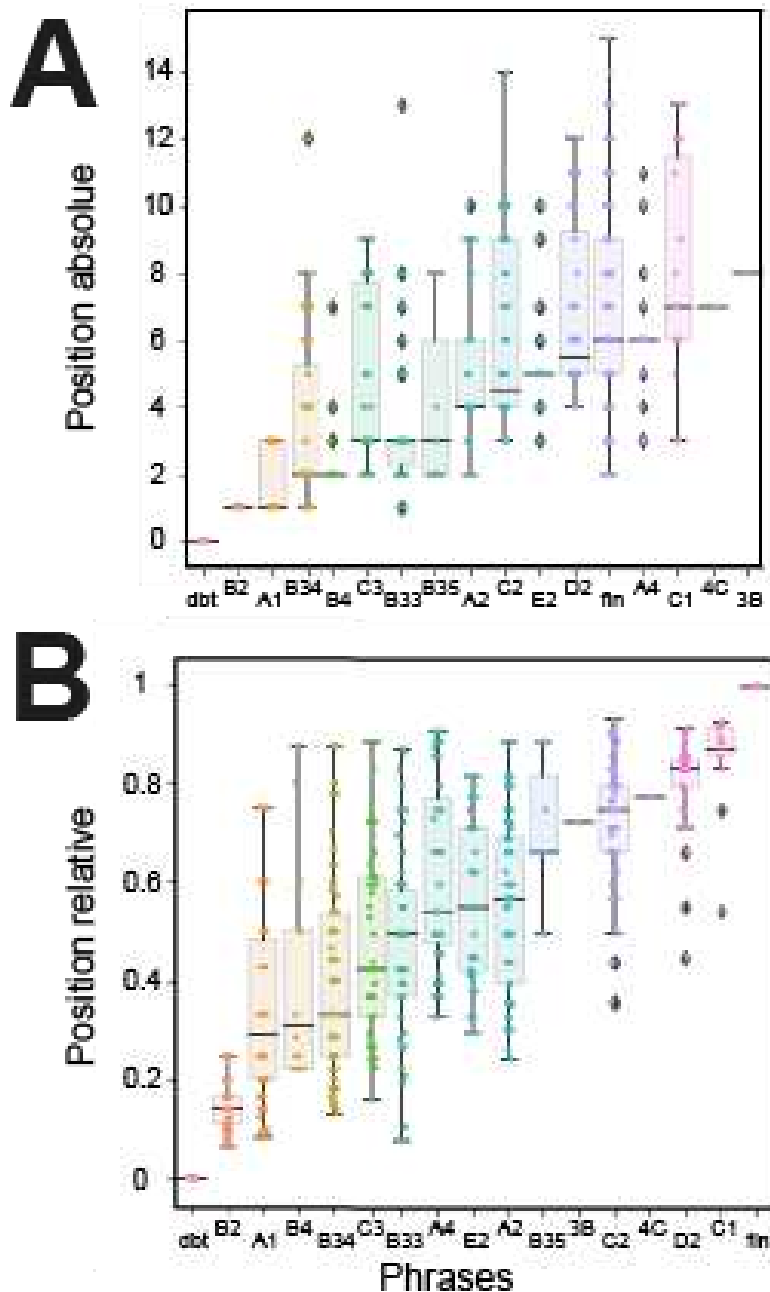


Figure 13: Positions des phrases dans le chant d'un oiseau

(A) Position absolue des phrases

(B) Position relative des phrases

Les phrases sont classées selon leur position du début (dbt) à la fin. Leur position est présentée sous la forme d'un box plot où chaque point à l'intérieur correspond à une position de la phrase dans le chant. Le box plot présente la variabilité de la position de la phrase dans le chant avec l'écart type (plus il est allongé, plus la position est variable). Les losanges représentent les phrases dont la position est en dehors du box plot.

Ligne noire : position médiane de la phrase.

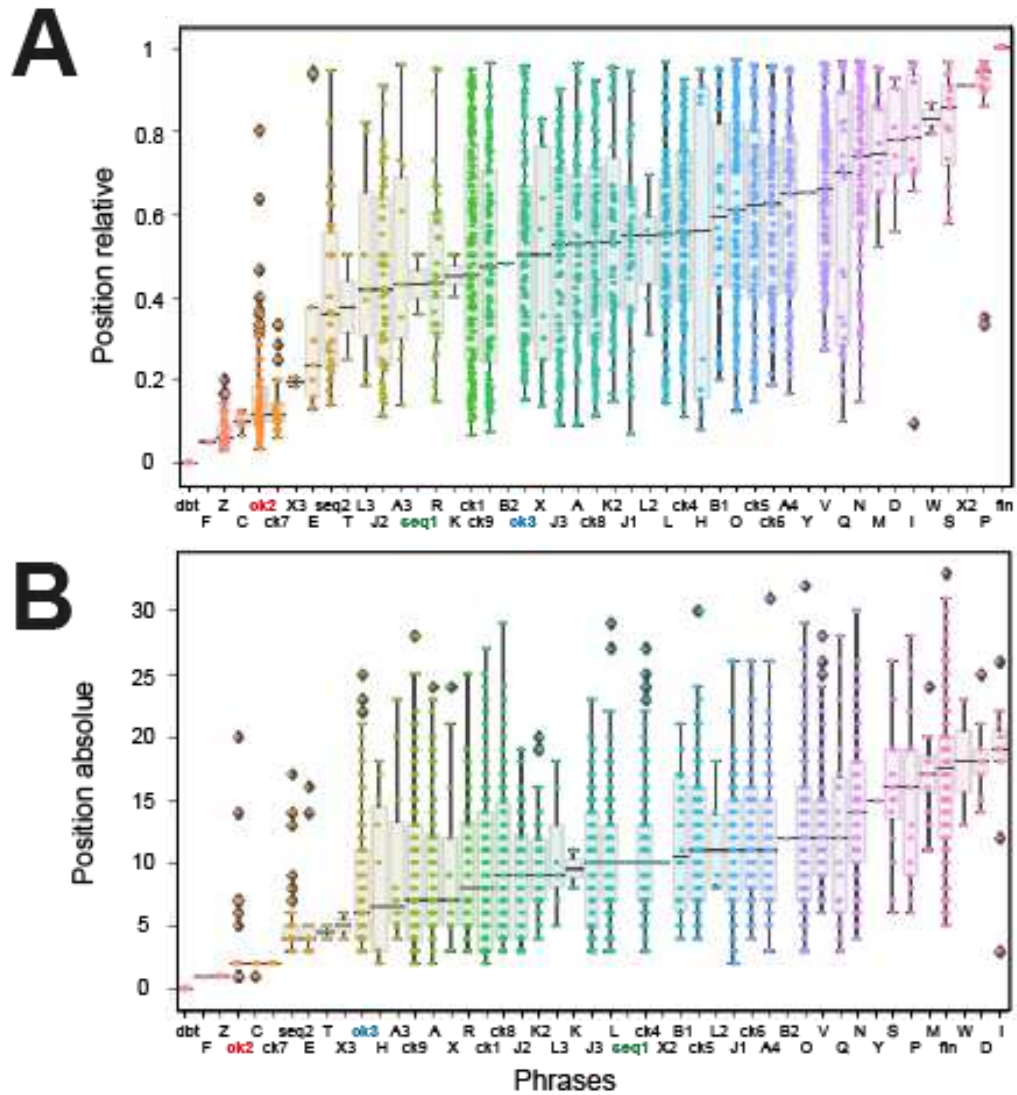


Figure 14: Position des phrases et chunks d'un oiseau

(A) Position relative des phrases et chunks dans le chant

(B) Position absolue des phrases et chunks dans le chant

Les séquences les plus récurrentes ont été remplacées par « ckX » où X correspond à un nombre attribué aléatoirement. On peut observer des phrases qui sont incluses présentes dans un chunk dans le cas où celles-ci seraient inclus au sein d'autres séquences.

ck2 (en rouge): chunk 2; séquence C-H-E-R

ck3 (en bleu) : chunk 3 ; séquence Q-H-E-N

seq1 (en vert) : C-H-E-N non identifiée comme chunk

seq2 (en noir) : Q-H-E-R non identifiée comme chunk

Ces séquences (ck2, ck3, seq1, seq2) sont au cœur d'une dépendance à long terme présentée plus haut

Ligne noire : position médiane de la phrase ou de la séquence

d'un oiseau précédemment observé dans la figure, on peut constater, que certains *chunks* ont des positions préférentielles dans les chants : les *chunks* 2 et 7 sont plutôt produits en début de chant tandis que le *chunk* 3 plutôt produit en milieu de chant. Il est par ailleurs important de noter que sur cet exemple représentatif, les séquences 1 et 2 (seq1 et seq2) correspondent aux séquences de phrases CHEN et QHER, et les *chunks* 2 et 3, les séquences de phrases CHER et QHEN, présentées plus haut (voir Dépendances à long terme).

Nous avons précédemment vu que les probabilités de divergence des séquences H-E-R et H-E-N changent suivant la phrase qui précède (C et Q). Or, on peut maintenant s'apercevoir que l'oiseau produit préférentiellement la phrase C au début des chants ($PR_C = 2$; $PR_C = 0,07$) et la phrase Q vers le milieu des chants ($PA_Q = 11$; $PR_Q = 0,40$). L'oiseau produit donc préférentiellement la séquence C-H-E-R au début de son chant. Pour autant, l'oiseau peut aussi produire la séquence C-H-E-N, même si l'occurrence et les probabilités de transition entre la séquence C-H-E et la phrase N sont plus faibles qu'avec la phrase R. Or si l'on prend en considération le paramètre de la position des phrases, on note que C-H-E-N n'arrive jamais en début de chant au contraire de C-H-E-R. La place de la phrase C semble ainsi impacter fortement la suite de la séquence H-E (**Figure 15**). Ainsi, ce qui peut être interprété comme des dépendances à long terme vient du fait que ce qui précède et suit la séquence H-E dépend de la position de H-E dans le chant.

En résumé, les différentes phrases ou séquences qui sont produites par un oiseau peuvent se trouver préférentiellement situées à certains points dans les chants.

L. Organisation des chants de canari selon les distances entre les phrases dans les chants.

Les analyses des chants de canaris ou d'autres espèces présentant des chants à l'organisation syntaxique complexe, se basent uniquement sur l'analyse des probabilités de transition entre les phrases et notamment essentiellement les probabilités de divergence. Or, comme le suggèrent les résultats précédents, les probabilités de transition ne permettent pas de rendre compte de l'organisation des chants de canaris, il faut aussi tenir compte des fréquences d'occurrence ainsi que des positions des phrases et séquences de phrases. Nous avons alors utilisé une nouvelle méthode, basée sur le calcul de distances entre les phrases.

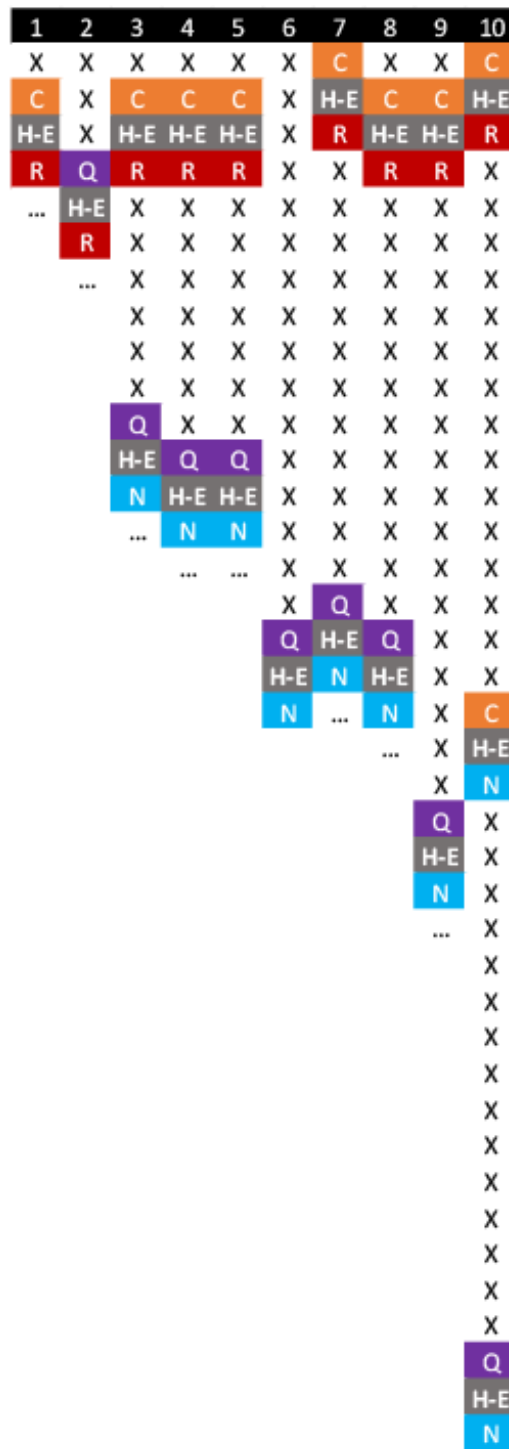


Figure 15: Dépendances à long terme dans le chant selon la place des phrases

Sélection de 10 chants parmi les 300 chants analysés pour l'oiseau correspondant (numérotation de 1 à 10).
 Colonne : chants.
 Ligne : phrase.
 Les phrases (C, Q, R et N) et séquences (H-E) ont été mises en couleur, les chants représentés en partie sont suivis de « ... ».

Les autres phrases ont été remplacées par un X.

On note que la phrase R n'est présente qu'en début de chant au contraire de la phrase N.
 Les phrases C et Q sont préférentiellement en début et milieu/fin de chant, respectivement.

L'analyse multidimensionnelle ou *multidimensional scaling* (MDS) permet de séparer les phrases sur un plan en deux dimensions, selon la distance qui sépare chaque paire possible, tout en considérant le début et la fin du chant comme des phrases virtuelles. Si les phrases sont fréquemment associées en séquence, alors la distance entre elles est réduite. Plus elles sont éloignées dans le chant, plus leur distance est grande. Pour comprendre les représentations graphiques susceptibles d'être engendrées par cette analyse MDS, nous avons dans un premier temps simulé des jeux de données (**Figure 16**). Si l'agencement est aléatoire, alors toutes les phrases sont à une distance équivalente entre elles, tout en étant éloignées du début et de la fin du chant (**Figure 16A**). Au contraire si l'enchaînement est toujours identique, on observe un cheminement sous la forme d'un arc de cercle entre le début et la fin du chant (**Figure 16B**). Enfin, si l'oiseau chante deux chants différents, soit ABCD et soit EFGH, on observe deux arcs de cercle clairement séparés entre le début et la fin du chant (**Figure 16C**).

La **Figure 17** présente un exemple de représentation d'un chant d'un canari obtenu par la méthode MDS. On note tout d'abord des groupes de phrases (**Figure 17**, cercles bleus). Ces groupes de phrases sont éloignés des autres signifiant que la distance entre ces phrases est réduite et qu'elle est plus importante avec les autres phrases. Les probabilités de transition entre ces phrases sont très fortes ($p_{Div}: 0,95 \pm 0,03$; $p_{Conv}: 0,96 \pm 0,03$). Ces deux groupes de phrases correspondent à deux séquences qui peuvent se répéter à divers endroits dans le chant.

On observe ensuite un cheminement entre le début et la fin du chant, certaines phrases étant plus proches du début ou de la fin (**Figure 17**, cercles verts). Cette deuxième observation suggère que certaines phrases parmi ces deux groupes ont des places préférentielles dans le chant. Au contraire des deux groupes de phrases éloignées (cercles bleus), ces phrases (cercles verts) ont des places moyennes faciles à distinguer. Ainsi, cette méthode permet de décrire l'organisation du chant de canari en prenant en compte la position des phrases dans les chants.

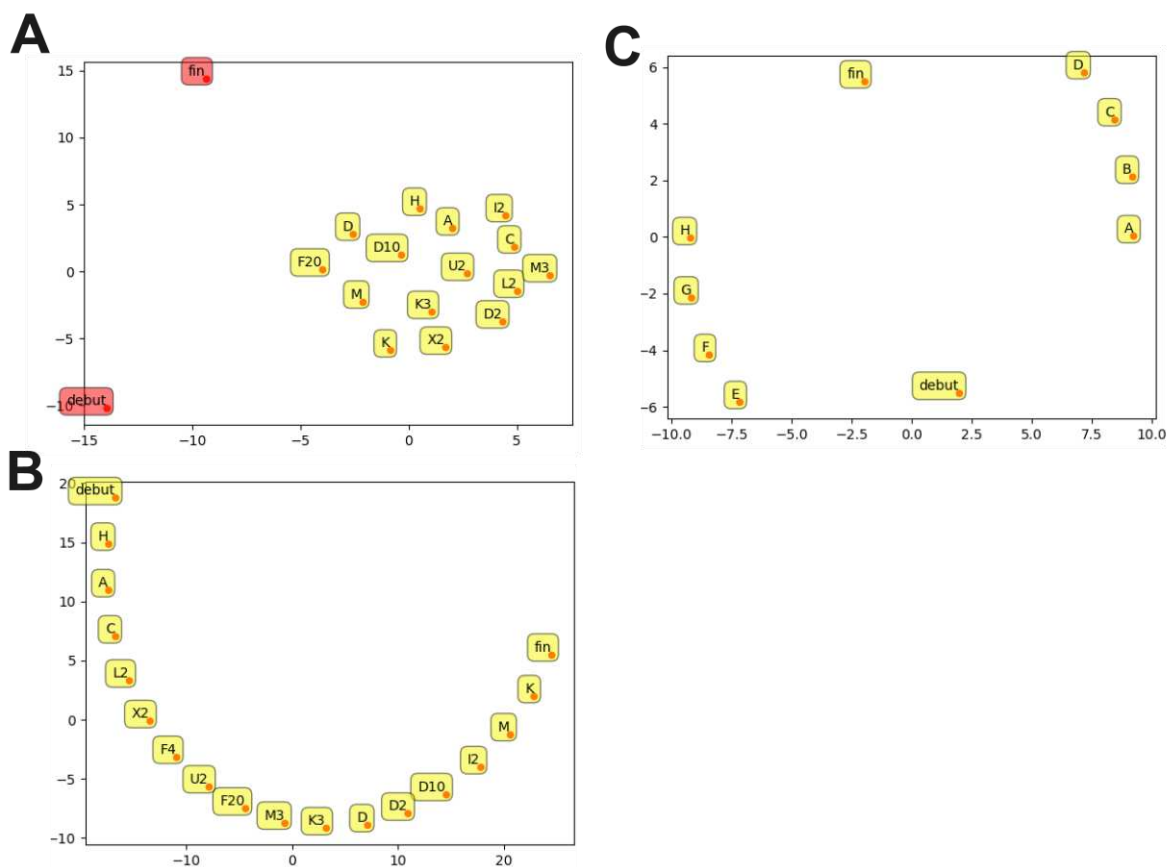


Figure 16: Représentation de l'organisation de chants théoriques via l'analyse MDS

Encadré jaune : les phrases ; point orange : la position de la phrase

En rouge : les éléments n'étant pas présents dans le même plan (début et fin en A).

L'analyse multidimensionnelle (MDS) permet de séparer à l'aide d'une matrice, chaque paire de phrases selon leur position dans les différents chants. Ici est présenté le MDS obtenu pour trois séries de chant :

- (A)** *Des chants dans lesquels la position de chaque phrase varie systématiquement d'un chant à un autre.*
- (B)** *Des chants avec un enchaînement de phrases stéréotypé, c'est-à-dire toujours le même chant*
- (C)** *Deux chants différents : soit ABCD, soit EFGH*

Pour chaque série, le nombre de chants est le même et correspond à celui de l'oiseau de la Fig. 33

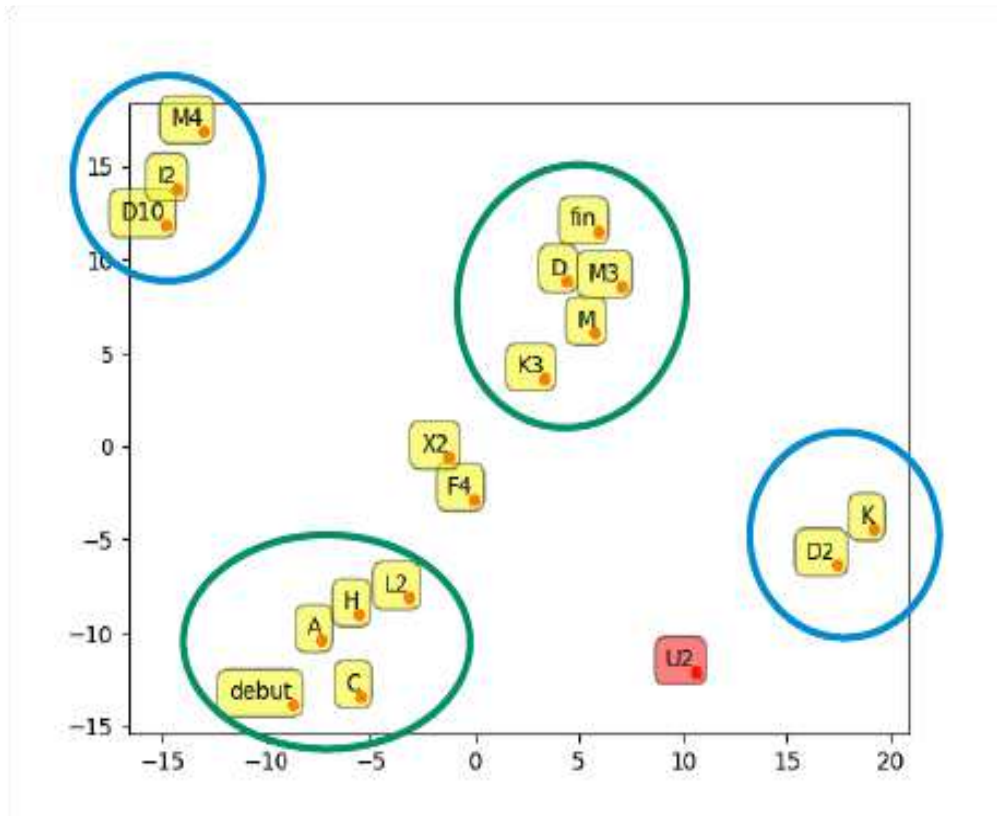


Figure 17: Représentation de l'organisation des chants d'un canari par la méthode MDS

Encadré jaune : les phrases ; point orange : la position de la phrase.

Cette représentation du chant associe différents aspects des trois modèles présentés sur la figure précédente. Les phrases qui sont proches sont celles qui forment des séquences. Certaines phrases sont proches du début et de la fin (cercle vert). Certaines phrases (cercle bleu) sont éloignées du cheminement central, ceci est lié au fait que les séquences qu'elles forment (D2-K et D10-I2-M4) sont présentes à différents endroits dans le chant et donc les distances avec les autres phrases du chant sont importantes. Ces deux séquences se trouvent rarement dans le même chant.

La phrase en rouge est une phrase peu produite.

Discussion

De nombreuses études se sont intéressées à la composition du répertoire des syllabes ou phrases des oiseaux chanteurs. Pourtant, le chant se caractérise non seulement par les syllabes ou phrases qui les composent, mais aussi par leur ordre d'agencement. Si la détermination de l'ordre des syllabes ou phrases dans les chants d'un individu est aisée chez des espèces au chant dit « simple » comme le diamant mandarin, où le chant est une suite de syllabes dont l'ordre est fixé, il en est tout autrement pour des espèces aux chants complexes telles que le canari. Le chant des canaris est composé de syllabes répétées formant des phrases. L'enchaînement flexible des phrases forme les chants. La complexité du chant du canari est notamment liée au fait qu'un canari modifie, d'un chant à un autre, non seulement la composition, c'est-à-dire les phrases qui constituent ses chants, mais aussi l'ordre de ses phrases. Les différentes analyses que nous avons effectuées nous ont permis de déterminer des séquences de phrases stéréotypées et/ou récurrentes dans les chants produits par un même oiseau.

La présence de telles séquences nous conduit à penser qu'il existe chez les canaris un niveau d'organisation supplémentaire s'intercalant entre le niveau « phrase » et le niveau « chant ». Un tel niveau a déjà été décrit chez le moineau du Japon. Ils produisent des enchaînements de 2 à 5 syllabes de façon groupée : les probabilités de transition entre les syllabes de tels enchaînements sont fortes alors qu'elles sont faibles avec d'autres syllabes. Ces séquences de syllabes fortement stéréotypées ont été qualifiées de « *chunk* » (Okanoya, 2004). Nous proposons l'hypothèse que les canaris pourraient organiser leur chant sous la forme de plusieurs séquences stéréotypées de syllabes, autrement dit de « *chunks* ».

Nos analyses des séquences de phrases suggèrent que les *chunks* se définissent selon plusieurs paramètres : le pourcentage d'occurrence dans les différents chants, les probabilités de convergence et de divergence des séquences de n phrases (n allant de 2 à 5) afin de prendre en compte les dépendances à long terme (Markowitz et al., 2013) et la position dans le chant. Nous avons aussi pu constater que, pris séparément, ces paramètres ne suffisent pas à identifier l'ensemble des séquences présentes dans le chant d'un oiseau. En effet, le pourcentage d'occurrence n'identifie pas les séquences peu présentes, mais pourtant stéréotypées avec des probabilités de transition très fortes. De

même, identifier les *chunks* à partir de l'analyse seule des probabilités de convergence et de divergence exclut les séquences partageant une ou plusieurs phrases communes.

Nous avons aussi pu constater que la prise en compte du paramètre position permet d'affiner la caractérisation de certaines séquences déterminées par le calcul des probabilités de convergence et de divergence. Certaines phrases ont une position stéréotypée tandis que d'autres peuvent être situées tout au long du chant. Parmi ces dernières, on distingue deux profils :

- des phrases pour lesquelles les probabilités de divergence et de convergence n'évoluent pas, quelle que soit leur position. Elles sont fréquemment incluses dans des *chunks* après une analyse des probabilités de transition au sein de séquences de deux phrases (ordre 1)
- des phrases qui en fonction de leur position ne sont pas associées aux mêmes phrases. Ces phrases peuvent être au sein de dépendances à long terme où leurs positions impactent les transitions qu'elles effectuent.

Les positions préférentielles des phrases dans les chants peuvent impacter la composition même des séquences (par exemple les séquences C-H-E-N, C-H-E-R, Q-H-E-N et Q-H-E-R). Une étude précédente a proposé que l'organisation des phrases dans le chant des canaris est réglée par des processus Markoviens d'ordre compris entre 1 et 4 (Markowitz et al., 2013). Nos analyses remettent quelque peu en cause cette idée si on introduit le paramètre supplémentaire de la position des phrases. Étonnamment, cette prise en compte de la position même des phrases dans les chants n'a, à ma connaissance, jamais été réalisée dans des études précédentes s'attachant à caractériser les règles qui régissent l'organisation des phrases dans le chant des canaris (Markowitz et al., 2013) ou d'autres espèces au chant complexe (Okanoya, 2004).

Cette étude de l'organisation des chants de canari nous a amenés à reconsidérer les méthodes utilisées pour évaluer cette organisation. Les méthodes utilisées dans la littérature reposent le plus souvent sur un calcul des probabilités de transition, et même le plus généralement des probabilités de divergence. Dans notre recherche de représentation graphique, nous avons exploré la possibilité d'avoir recours à la méthode MDS basée sur le calcul de distance entre les différentes phrases de chaque chant. Cette représentation inclut la position des phrases par rapport au début et à la fin du chant. Elle

permet aussi d'identifier des phrases qui sont fréquemment produites dans une même séquence, ce qui facilite l'identification de *chunks* auxquels elles pourraient appartenir. Cette représentation nous permet donc de visualiser la structure globale du chant.

Cependant, cette méthode MDS tel que nous l'avons calculée présente des limites. La première est liée à la non-prise en compte de phrases identiques répétées dans un même chant : lorsqu'un oiseau chante par exemple la séquence ABAC, la distance considérée entre C et A est fixée à 1, en négligeant donc la distance entre C et le premier A (qui serait de 3).

De plus, au contraire des distances entre les villes sur une carte de France par exemple, toutes les distances entre les phrases ne sont pas mesurables. En effet, l'oiseau ne produit pas toutes les paires qu'ils pourraient faire à partir de son répertoire (par exemple un oiseau peut chanter deux chants, ABC et CDE, mais jamais de chant incluant à la fois A, B, C, D et E). Pour pallier cette absence, nous avons fixé une valeur théorique équivalente à 1,5 fois la longueur du chant. Il existe donc au sein de notre matrice de distance des approximations qui peuvent modifier la représentation.

Enfin, cette représentation ne donne pas d'information sur la variance. On ne sait pas si une phrase présente à égale distance des marqueurs début et fin l'est, car elle est systématiquement produite au milieu du chant, ou si elle peut être à différents endroits dans le chant, le milieu du chant étant sa position moyenne.

Le MDS est donc une représentation qui permet de pouvoir mieux visualiser la structure globale du chant au contraire des autres représentations probabilistes, mais ne suffit pas, en l'état actuel, à renvoyer une image complète et facilement interprétable du chant des canaris. Une difficulté majeure pour obtenir une représentation fiable réside dans la multiplicité des paramètres à considérer pour caractériser le chant d'un individu. Il faut ajouter aussi que se pose la question de la hiérarchie entre les différents paramètres, autrement dit peut-on considérer que ce sont les probabilités de transition, la position, la fréquence d'occurrence, les dépendances à long terme qui définissent le mieux un chant ? Des études supplémentaires restent donc à être menées pour mieux comprendre ces règles d'organisation des phrases dans le chant des canaris.

Une question qui se pose aussi est de savoir si l'organisation des phrases dans un chant contribue à l'individualité. Dans une colonie de canaris, le taux de partage de phrases varie de 12 à 26% suivant les études (Lehongre et al., 2008; Waser and Marler, 1977), ces taux

diminuant très vite lorsque l'on considère des séquences de 2 à 6 phrases (Lehongre et al., 2008). L'identification des *chunks* pose la question de leur rôle pour transmettre un message relatif à l'identité de l'émetteur. Des études comportementales sont nécessaires pour éclaircir ce point.

Une autre question est de savoir dans quelle mesure ces *chunks* sont des éléments déterminants pour la représentation du propre chant (BOS) de l'oiseau. Les neurones du noyau HVC, structure sensori-motrice, présentent une sélectivité dans leurs réponses auditives pour le BOS. Dans le chapitre suivant, nous explorerons dans quelle mesure les neurones du HVC détectent des variations dans l'organisation des phrases dans les chants et plus particulièrement des *chunks* que nous avons identifiés.

En conclusion, le chant du canari, par le fait qu'il puisse être structuré selon différents niveaux d'organisation (syllabe, phrase, *chunk*, chant), que les phrases puissent se trouver dans une certaine mesure, à certaines places dans les chants, et qu'elles s'enchaînent selon des probabilités de transition révélant des dépendances à long terme, possède bien plus de points communs avec le langage humain que les espèces d'oiseaux chanteurs communément utilisés pour étudier les bases neurobiologiques du chant. En ce sens, le canari peut apparaître comme un modèle pertinent pour étudier les mécanismes susceptibles d'être impliqués dans le langage et notamment le traitement de l'ordre des signaux acoustiques.

Partie III :

ETUDE DU CODAGE NEURONAL DE
L'ORGANISATION SÉQUENTIELLE DU
PROPRE CHANT DE L'OISEAU DANS
UNE AIRE SENSORI-MOTRICE

Partie III : Etude du codage neuronal de l'organisation séquentielle du propre chant de l'oiseau dans une aire sensori-motrice

Introduction

À l'issue de la précédente partie consacrée à l'analyse des chants de canari, nous avons vu que les chants de chaque canari (Bird's own song, ou BOS) sont composés de différentes syllabes organisées en phrases, issues du répertoire de l'individu, dont l'agencement n'est pas aléatoire, mais est régi par des probabilités de convergence et de divergence d'ordre 1 et plus. Nous avons observé que les probabilités de transition entre deux phrases du répertoire varient largement indiquant que certaines phrases sont fortement associées en séquences tandis que d'autres ne le sont jamais. Cependant, les probabilités laissent de côté des informations importantes telles que l'occurrence de chaque phrase et de chaque séquence. Les analyses ont indiqué que certaines séquences reviennent très souvent. Au sein de ces séquences récurrentes qui apparaissent comme « stéréotypées », les probabilités de divergence et convergence peuvent être très fortes. Nous avons proposé que ces séquences de syllabes forment ce qu'on appelle un « *chunk* » et que ces *chunks* constituent un niveau supplémentaire d'organisation du chant du canari, entre le niveau phrase et le niveau chant. De ce fait, si l'on doit, chez le canari, caractériser l'organisation des chants produits par un individu donné, on doit considérer à la fois les probabilités de transition entre les phrases, mais aussi les *chunks*.

Si d'un chant à un autre, le type de phrases varie, leur place dans les chants n'est pas aléatoire. Certaines phrases et séquences sont préférentiellement au début du chant tandis que d'autres sont à la fin. Nos résultats indiquent donc que l'agencement des phrases dans les chants est régi par plusieurs aspects (probabilité, *chunking* et position). À partir de là, la question que nous nous sommes posée est de savoir dans quelle mesure ces aspects sous-tendent le traitement de l'ordre des phrases dans les chants au niveau cérébral.

Jusqu'à présent, un traitement de l'ordre des signaux dans les chants a surtout été étudié au niveau d'une région sensori-motrice, le HVC. Au niveau de cette région, les neurones répondent sélectivement plus au(x) chant(s) produit(s) par l'individu (BOS) qu'aux chants de congénères (Margoliash, 1983; Volman, 1993). Des é

comportementales suggèrent que le BOS est un chant qui jouerait un rôle dans les processus de discrimination et de reconnaissance. Le BOS serait perçu comme différent par rapport aux chants d'autres individus, que ce soit chez des espèces territoriales ou non, d'après des études réalisées sur le terrain ou en laboratoire (McArthur 1986; Gentner and Hulse 1997; Cynx and Nottebohm, 1990). Chez les espèces qui font du *song matching*, c'est-à-dire qui répondent à un chant par le même chant ou le plus proche, le BOS pourrait être utilisé pour évaluer dans quelle mesure les chants entendus diffèrent du BOS (Pytte and Suthers, 1999; Horn and Falls, 1996). Dans les études de laboratoire basées sur une tâche de conditionnement instrumental (Go No-Go), il a été pointé que les performances de discrimination sont bien meilleures lorsque les oiseaux ont à discriminer le BOS d'un chant d'un individu familier, par rapport à une discrimination entre deux chants d'individus de l'espèce (Cynx and Nottebohm, 1990; Scharff et al., 1998). Au-delà du traitement auditif du BOS, d'autres études suggèrent que le HVC a un rôle dans la perception auditive (revue dans Prather 2013; Mooney 2014). Chez les femelles, qui ne chantent pas, mais possèdent un HVC, une lésion du HVC induit un changement dans le degré de discrimination entre les chants, notamment entre les chants conspécifiques, mais aussi entre les chants conspécifiques et hétérospécifiques (Brenowitz et al. 1991; Del Negro et al. 1998). Chez les mâles, une lésion du HVC n'affecte pas la discrimination entre deux chants si celle-ci était déjà mise en place avant la lésion. En revanche, elle affecte l'acquisition de nouvelles discriminations (Gentner and Hulse, 2000).

Pour en revenir au traitement du BOS, lors de sa diffusion, de nombreuses études se sont penchées sur la sélectivité des réponses des neurones du HVC en faveur du BOS (revue de Mooney 2014). Modifier les paramètres acoustiques des syllabes constitutives du chant entraîne une diminution des réponses neuronales par rapport au BOS initial (Margoliash, 1986), mais surtout ce qu'ont montré les études, c'est l'importance de l'organisation temporelle du BOS. Diffuser le chant à l'envers, ce qui revient, à la fois, à inverser l'ordre des syllabes et inverser l'ordre des paramètres acoustiques de chaque syllabe, induit des réponses neuronales bien moins importantes que le BOS (Margoliash 1986; 1983; Lewicki and Arthur 1996; Coleman 2004; Nakamura and Okanoya 2004; Del Negro 2005; Nealen and Schmidt 2006). Ne changer que l'ordre des syllabes sans modifier leur structure acoustique (les caractéristiques spectrotemporelles) suffit aussi à induire une baisse des réponses par rapport au BOS, chez le diamant mandarin qui a un seul chant dont l'ordre est fixe (Lewicki and Arthur 1996; Coleman 2004) ou chez des espèces qui

ont plusieurs types de chants (Prather et al., 2008) dont l'ordre des syllabes dans les chants est variable (Lehongre and Del Negro, 2011; Nakamura and Okanoya, 2004; Nealen and Schmidt, 2006).

Cette sensibilité en faveur de la structure temporelle du BOS émerge au sein du système auditif (revue dans Prather and Mooney 2004), mais elle est plus exclusive au sein du HVC, les neurones pouvant ne pas répondre (ou peu) à d'autres chants que le BOS. Dans le champ L, qui est l'analogue du cortex primaire des mammifères, changer l'ordre affecte peu les réponses, en comparaison de ce qui est observé dans l'HVC (Lewicki and Arthur 1996). Dans le noyau interfacial du nidopallium (Nif), d'où proviennent principalement les informations auditives arrivant à l'HVC, les neurones répondent plus au BOS qu'au chant dont l'ordre a été modifié, mais ils répondent aux deux types de chant, contrairement aux neurones du HVC (Coleman, 2004; Janata and Margoliash, 1999).

Outre une sensibilité à l'ordre des syllabes dans le chant, l'analyse des réponses lors de la diffusion du BOS a montré que les neurones peuvent répondre spécifiquement à une syllabe (ils seront actifs lors la diffusion de celle-ci, quel que soit son contexte, par exemple, la syllabe A dans le contexte X-A et Y-A ou spécifiquement à une combinaison de syllabes. Si cette combinaison est modifiée, les réponses de ces mêmes neurones s'en retrouvent diminuées, voire supprimées (Lewicki and Arthur 1996; Margoliash 1986; Margoliash and Fortune 1992; Coleman 2004). À la suite de ces résultats, il a été émis l'hypothèse d'un codage en chaîne du BOS. En d'autres termes, la sélectivité des neurones du HVC pour le BOS reposerait sur un codage du répertoire des éléments (syllabes et séquences de plusieurs syllabes) de ce chant au sein du HVC, certains neurones étant sensibles à une syllabe donnée, d'autres à une autre syllabe, d'autres à une séquence de 2 syllabes du BOS, d'autres à une autre séquence de ces syllabes de 2. Lors de la diffusion du BOS, il y aurait une activation en série, selon une chaîne fixe, des différents neurones du HVC. Cette conception repose sur des résultats obtenus chez le diamant mandarin qui a un seul chant, composé de quelques syllabes, toujours émises selon le même ordre. Mais la sélectivité pour le BOS au niveau du HVC a également été mise en évidence chez des espèces aux chants variables, le moineau du Japon et le bruant chanteur et le canari (Nealen and Schmidt 2006; Nakamura and Okanoya 2004; Fujimoto et al. 2011; Del Negro 2005; Lehongre and Del Negro 2011). Ce que montrent ces études est le fait que les neurones pouvaient répondre à plusieurs syllabes ou à de nombreuses séquences différentes (Nealen and Schmidt 2006; Nakamura and Okanoya 2004; Fujimoto et al.

2011; Lehongre and Del Negro 2011) suggérant un codage du BOS distribué sur la population des neurones du HVC (Nishikawa et al. 2008; Allende et al. 2013).

La sensibilité à l'ordre temporel des syllabes, chez les espèces ayant un chant variable, pose la question de savoir quels types de traitement sous-tendent cette sensibilité. Chez le moineau du Japon, les chants de chaque individu sont constitués à partir d'une dizaine de syllabes différentes qui s'enchaînent selon des probabilités de transition d'ordre 1 ou d'ordre supérieur. Bouchard et Brainard (2013) ont alors cherché à savoir si les réponses dépendaient des probabilités de transition. Pour cela, ils ont enregistré l'activité des neurones lors de la diffusion de séquences composées d'un millier de syllabes produites par l'oiseau, mais combinées de façon pseudo-aléatoire. Puis, ils ont analysé les réponses à chacune des syllabes du répertoire en prenant en compte les probabilités de convergence et de divergence pour les séquences (d'une longueur de 1 à 9 syllabes) précédant la syllabe en question. Les résultats indiquent que plus la probabilité de convergence est forte, plus la réponse à la syllabe est forte et que la probabilité de convergence explique 28% de la variance dans la force des réponses (Bouchard and Brainard, 2013). Dans cette étude, ils se sont également intéressés aux probabilités de divergence, mais leurs résultats indiquent que la probabilité de divergence n'expliquerait que 8 % de la variance dans les réponses des neurones.

À partir de là, la question que nous nous sommes posée a été de savoir si, de la même façon, chez le canari, les réponses des neurones du HVC dépendaient des probabilités de convergence ou de divergence. Pour aborder cette question, nous n'avons pas suivi la méthode utilisée par Bouchard and Brainard, qui a consisté à diffuser de très longues séries de syllabes (d'un millier de syllabes). Nous avons, comme dans de nombreuses études (Doupe 1997; Amin et al. 2004; Nealen and Schmidt 2006; Prather et al. 2008; Lehongre and Del Negro 2011) manipulé l'ordre des syllabes d'un BOS, c'est-à-dire construit des chants artificiels à partir des phrases de ce BOS. Dans une précédente étude de notre équipe, réalisée chez le canari, l'ordre des phrases du BOS avait été manipulé, en changeant la position de toutes les phrases (Lehongre and Del Negro 2011). Les probabilités de convergence et de divergence de toutes les transitions étaient égales à zéro. Aucun des *chunks* constituant le BOS n'avait été respecté. Les réponses obtenues lors de la diffusion de ces chants étaient bien moins importantes que celles induites par la diffusion du BOS. Pour comprendre ce qui a induit ces changements dans les réponses, la nouvelle étude que nous avons réalisée a été basée sur la construction de chants

artificiels à partir du BOS, qui conservaient certains aspects de l'organisation du BOS. Étant donné que les chants de canari sont constitués de *chunks* au sein desquels les probabilités de convergence et de divergence sont fortes, nous avons diffusé deux types de versions du BOS. La première est constituée de versions du BOS dont l'ordre des phrases au sein de ces *chunks* a été changé (de façon à ce que les probabilités, initialement fortes, soient très faibles), mais qui conservent l'ordre des *chunks*. La deuxième catégorie, à l'inverse, consiste à créer des versions du BOS dans lequel l'ordre des phrases au sein des *chunks* a été maintenu (gardant les fortes probabilités intactes), mais dont l'ordre des *chunks* a été changé. Notre but a été de déterminer si la sensibilité des neurones du HVC à l'organisation temporelle du BOS porte sur l'ordre des phrases qui forment des *chunks*. Dans ce cas changer l'ordre des *chunks* ne devrait peu ou pas avoir d'impact sur les réponses. Si elle s'étend au-delà des *chunks*, dans ce cas, un effet sur les réponses devrait être observé.

Matériels et méthodes

A. Sujets

Les canaris utilisés pour l'ensemble des études étaient des adultes mâles (n=8) âgés de plus d'un an et provenant d'un élevage interne au laboratoire. L'enregistrement des chants a été réalisé entre les mois de mai et juillet pendant lesquels les canaris produisent des chants stéréotypés (Nottebohm et al. 1986 ; Leitner et al. 2001 ; Del Negro et al. 2005). Pour cela, les oiseaux ont été placés dans une cage individuelle (l : 27 cm x L : 50 cm x h : 32 cm) installée dans un caisson d'isolation acoustique. Les oiseaux sont restés dans ce caisson pendant environ 1 semaine. Ils ont bénéficié d'un accès *ad libitum* à la nourriture et à l'eau et ils étaient sous un régime de lumière de 13h de jour/11h de nuit.

B. Stimuli auditifs

Analyse des chants. Entre 50 et 300 chants ont été analysés pour chaque oiseau. Toutes les analyses décrites dans la partie II ont été réalisées à partir des chants des oiseaux de cette étude.

Sélection du BOS. À partir des analyses effectuées sur les chants, nous avons cherché à sélectionner un chant pour chaque individu, un BOS, qui puisse être considéré comme représentatif de ce que l'oiseau est capable de chanter. Nous avons sélectionné des BOS qui étaient constitués de trois séquences « *chunk* » composées d'un minimum de deux phrases.

Pour déterminer quels *chunks* étaient présents dans les chants d'un individu et quelle était la longueur de ces *chunks*, nous avons utilisé un programme écrit sous python. Pour des chants d'une longueur moyenne de 10 phrases, il est cherché les 3 ou 4 *chunks*, de taille 2, 3 ou 4 phrases qui sont produites le plus fréquemment. La **Figure 1** présente pour un oiseau les 10 séquences (motifs) les plus présents. On tient compte tout d'abord des motifs de taille 2, les 10 plus présents : on ne retient que ceux qui sont présents dans plus de 15 % des chants. Puis, on refait à nouveau cette opération pour les motifs de trois phrases (motif de taille 3). À partir de là, on prend en compte l'occurrence des motifs de

3 phrases et celle des motifs de 2 phrases qui le composent (la séquence ABP vient de AB en ajoutant P en dernière position, et de BP par ajout de A en première position). On calcule alors pour chaque séquence de trois phrases le rapport r :

$$r = \frac{Nocc(seq3ph)}{\max(Nocc(seq2ph))}$$

Avec $Nocc(seq3ph)$, le nombre d'occurrence du motif de trois phrases et $\max(Nocc(seq2ph))$, le nombre maximum d'occurrences d'un des motifs de deux phrases qui composent celui de trois phrases. On ne conserve les motifs de trois phrases que si le rapport r est $\geq 0,5$. Si des motifs de trois phrases ont entre eux une distance d'édition ≤ 2 (i.e. une seule lettre varie entre deux motifs) on ne garde que le *chunk* celui qui a l'occurrence la plus élevée. L'opération est renouvelée entre les motifs de trois phrases et quatre phrases. Si le rapport est $\geq 0,50$, on considère le motif de quatre phrases.

A

motifs de taille 2		
nb d'occurrences du motif		% nb de chant
47	AB	92
31	BP	61
27	TA	53
21	ZA	41
20	VE	39
20	PV	39
20	###ZF	39
20	ZF	39
18	###ZA	35
18	CT	35

B

motifs de taille 3					
nb d'occurrences du motif		% motif précédent			
31	ABP	Axx	66	xxP	100
25	TAB	Txx	53	xxB	93
19	ZAB	Zxx	40	xxB	90
19	BPV	xxV	61	Bxx	95
16	###ZAB				
16	FCT	Fxx	89		
14	CTA	Cxx	52	xxA	78
12	PVT	xxT	60		
9	ZFC	xxC	45		
9	###ZFC				

Figure 1 : Occurrence des motifs de deux phrases et trois phrases

(A) Occurrence des motifs de deux phrases. Les 10 premiers ont été sélectionnés sur la base d'un premier classement par nombre d'occurrences des motifs (colonne de gauche). Leur occurrence étant supérieure à 15% (colonne de droite) ils sont présélectionnés pour être des chunks. On attribue une couleur à chaque motif.

(B) Nombre d'occurrences (colonne 1) des motifs de trois phrases (colonne 2) à partir de ceux déjà observés dans le tableau en A (les couleurs sont à mettre en correspondance avec celles des précédents motifs de deux phrases). Les colonnes 3 et 5 indiquent de quels motifs de deux phrases, le motif de trois phrases peut provenir. Les colonnes 4 et 6 indiquent les rapports r en pourcentage pour chaque motif et sont sélectionnées seulement les motifs dont $r \geq 50\%$

Par exemple le motif ABP peut provenir du motif AB ou BP. Le rapport r pour ABP/AB vaut alors $31/47 = 0.66$

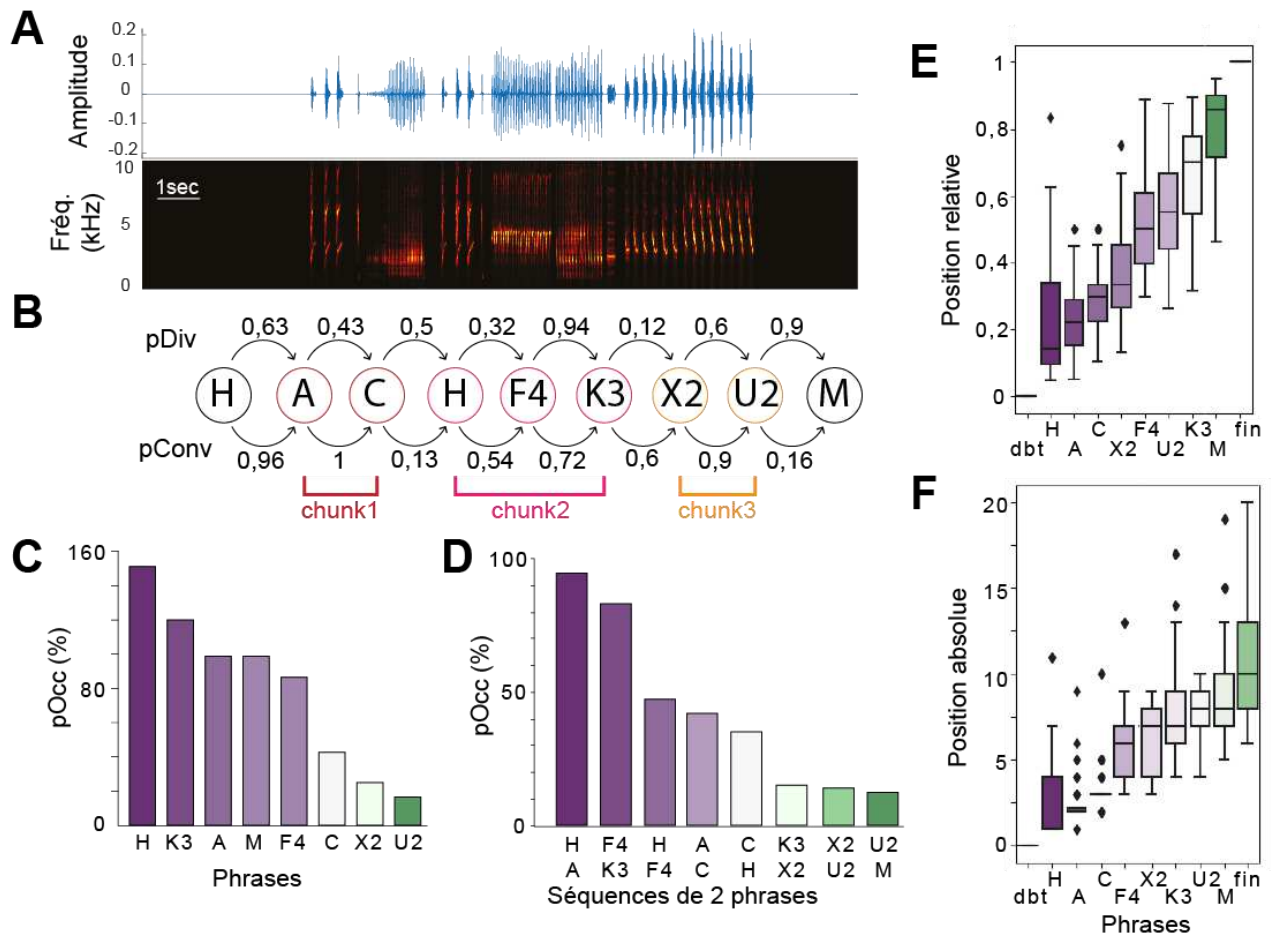


Figure 2 : Exemple de BOS sélectionné pour un oiseau

- (A) Spectrogramme et oscillogramme du BOS représentatif d'un oiseau. Chaque phrase est définie par un code sous la forme d'une lettre suivie ou non d'un chiffre. Les traits de couleurs délimitent les phrases entre elles (rouge pour le chunk1; rose pour le chunk2 et orange pour chunk3). La première et dernière phrase sont représentées par un trait blanc.
- (B) Probabilités de divergence (pDiv) et de convergence (pConv) pour chaque transition. On peut noter les transitions fortes dans F4-K3 et X2-U2.
- (C) Pourcentage d'occurrence de chaque phrase du BOS dans l'ensemble des chants analysés (n=61). La phrase H peut être produite plusieurs fois dans un même chant. Les phrases sont classées selon leur pourcentage d'occurrence (dégradé de violet vers le vert).
- (D) Pourcentage d'occurrence des séquences de deux phrases du BOS.
- (E) Position relative et absolue (F) de chaque phrase du BOS sélectionné (ligne centrale : médiane ; rectangle : quartiles ; losanges : valeurs hors distribution.)

Après avoir déterminé quels sont les *chunks* les plus produits par un oiseau, nous avons cherché dans la liste des chants, un chant qui soit constitué des 3 *chunks* parmi les plus présents (**Figure 2**). Nous avons également regardé les probabilités de divergence et de convergence des séquences de 2 phrases qui composent le BOS (**Figure 2B**) et la position de chacune des phrases dans les chants. Comme le montre la figure 2E et F, le BOS sélectionné pour cet oiseau, est constitué de phrases dont les positions relatives et absolues augmentent au fur et à mesure que l'on passe d'une phrase à la suivante.

C. Création des différentes versions modifiées du BOS.

À partir du BOS, il a été créé deux groupes de chants (**Figure 3**). À chaque *chunk* d'un BOS il a été attribué un chiffre : 1 pour le *chunk* 1, 2 pour le *chunk* 2 et 3 pour le *chunk* 3 (**Figure 14** & **Figure 15**).

Les chants du premier groupe ont été créés à partir du BOS dans lequel nous avons effectué un changement que nous avons appelé « local » : nous avons permuté deux phrases d'un *chunk* au sein du BOS, celles qui avaient des probabilités de transition fortes. Les BOS comportant trois *chunks*, il a été formé 3 chants : les chants L1, L2 et L3 avec, respectivement une permutation dans le *chunk* 1, le *chunk* 2 et le *chunk* 3 (**Figure 3**).

Les chants du deuxième groupe ont été créés à partir du BOS dans lequel nous avons changé l'ordre des *chunks*. L'ordre des phrases au sein des *chunk* était préservé. Dans un premier temps, il a été créé deux chants à partir de changements dits « globaux » : G231 et G312. Le chant G231 correspond donc à la succession des *chunks* 2, 3 et 1 tandis que le chant G312, aux *chunks* 3, 1 et 2. Ainsi, avec le BOS composé des *chunks* 1, 2 et 3 et les chants G231 et G312, chaque *chunk* est aux trois positions possibles.

Par la suite trois autres chants globaux ont été créés : les chants G132, G213 et G321, de façon à avoir toutes les combinaisons possibles (BOS, G231, G312, G132, G213 et G321).

Dans chaque chant créé, la première et la dernière phrase, symbolisées respectivement par la lettre *i* et *f*, n'ont pas été modifiées : elles ont gardé leur place. En conséquence, par rapport au BOS, les chants L1, L2, L3, G231 et G312 ont tous trois transitions entre deux phrases qui ont été modifiées. Dans les chants L1, L2, L3, les trois

transitions se suivent. Dans les chants G231 et G312, les trois transitions sont : entre la première phrase et le premier *chunk* de ce chant, entre la dernière phrase et le *chunk* qui précède, plus une autre transition (dans G231 : entre le *chunk* 3 et le *chunk* 1 ; dans G312 : à nouveau entre le *chunk* 3 et le *chunk* 1).

Chez un oiseau, nous avons également créé deux chants qui ont résulté du déplacement d'une des phrases.

Les changements apportés au BOS ont été réalisés à partir avec le logiciel Audacity à partir d'un script écrit en Python. L'intensité maximale de tous les types de chants a été réglée à une moyenne à la même valeur (-18 dB).

Parmi les chants qui ont été diffusés à tous les oiseaux se trouve également le chant d'un canari congénère qui n'avait jamais été entendu par les oiseaux de notre étude.

À partir de tous les différents types de chant, nous avons constitué des fichiers dans lesquels un chant donné était répété 10 fois, avec un intervalle inter-chant de 4 secondes.

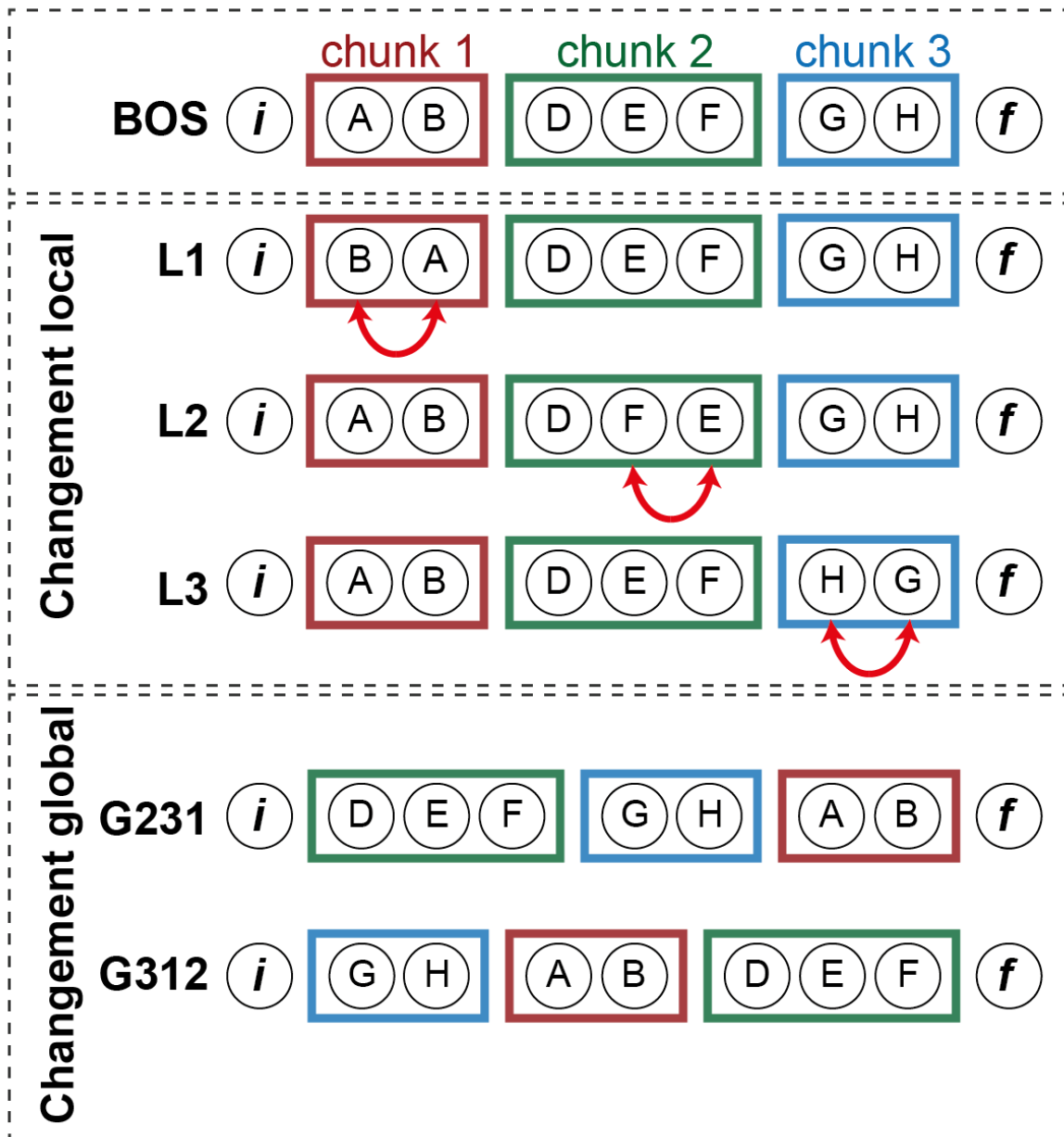


Figure 3 : Représentation schématique des chants créés à partir du BOS

(Haut) Le BOS composé de trois chunks. D'un oiseau à un autre, le nombre de phrases peut varier (ici un exemple : 2 phrases pour les chunks 1 et 3 et 3 phrases pour le chunk 2). Chaque phrase est représentée par une lettre. La première phrase (i) et la dernière (f) ne sont pas incluses dans les chunks et restent à leur position.

(Milieu) Changements locaux : 2 phrases d'un chunk sont permutées. L'ordre entre les chunks est identique à celui du BOS.

(Bas) Changements globaux : l'ordre des chunks est modifié. Exemple de chants créés : G231 et G312. L'ordre des phrases à l'intérieur des chunks n'est pas changé.

D. Enregistrements électrophysiologiques

Les enregistrements électrophysiologiques ont été réalisés sous anesthésie par inhalation d'un mélange gazeux composé d'oxygène et d'isoflurane (induction à 3% et maintenu à 1,5%). Pour cela, les oiseaux étaient placés dans une boîte à induction. Une fois l'oiseau endormi, il était positionné dans l'appareil stéréotaxique et fixé à l'aide de barres d'oreilles creuses. Après avoir retiré une partie des deux couches d'os du crâne et enlevé les méninges dans une zone restreinte de quelques centaines de microns (en largeur et en longueur), un peigne de 16 électrodes était descendu dans le HVC, d'après les coordonnées indiquées par l'atlas du cerveau de canari (Stokes et col, 1974). Les électrodes étaient organisées selon deux rangées de 8 électrodes (électrodes de tungstène ; impédance 1-2 M Ω à 1 kHz ; Alpha Omega Engineering) séparées entre elles de 100 μ m. La descente des électrodes était progressive : à partir d'une profondeur de 200-300 Microns (à partir de la surface du cerveau) et jusqu'à 900-1000 microns. L'activité dans le HVC se caractérise d'un point de vue électrophysiologique par une forte activité irrégulière et en bouffées. Les signaux des électrodes étaient amplifiés et filtrés (gain 10 000 ; passe-bande : 0,3-10 kHz ; taux d'échantillonnage : 44 kHz ; système AlphaLab SnR, AlphaOmega Engineering ®) pour extraire l'activité multi-unitaires. Au cours des enregistrements réalisés avec le logiciel Spike 2 (version 9.0, Cambridge Electronic Design ®), l'activité était surveillée en temps réel. Lorsque des potentiels d'action apparaissaient au-dessus du niveau de base (au moins trois fois le niveau de base), les chants étaient diffusés et l'activité multi unitaire était enregistrée. Puis, les électrodes étaient descendues d'au moins 100 μ m afin de s'assurer que d'autres neurones étaient enregistrés au site suivant. D'un site d'enregistrement à un autre, l'ordre de diffusion des chants changeait. Deux enceintes placées chacune à 30 cm de la tête de l'oiseau diffusaient les stimuli. Le microphone était relié au système d'enregistrement des signaux électrophysiologiques de façon à ce que les chants soient enregistrés en même temps et de façon synchrone avec l'activité neuronale. L'intensité des chants diffusés était ajustée, de façon à ce que les intensités maximales soient entre 70 et 75 dB.

E. Analyse des données électrophysiologiques

Une fois les enregistrements terminés, nous avons inspecté les enregistrements multi-unitaires effectués sous chacune des électrodes et à chaque site d'enregistrement afin de rechercher quels étaient les enregistrements dans lesquels des potentiels d'action dépassaient largement du niveau de base. Lorsque c'était le cas, nous avons extrait l'activité d'un ou de deux neurones sur la base de la forme de leur potentiel d'action, à partir du logiciel Spike 2. Pour chaque neurone extrait, l'activité était alors recueillie lors de la diffusion des différents chants, quantifiée par classe de 1 ms et stockée dans des fichiers.

Les réponses aux stimuli ont été quantifiées en calculant la moyenne du taux d'activité spontanée et la moyenne du taux d'activité évoquée par le stimulus. Le taux d'activité spontanée (*FR spontanée*) correspond au taux de décharge pendant les trois secondes de silence qui précédaient chaque chant. Le taux d'activité évoqué (*FR Evoquée*) a été calculé sur trois durées : i) celle du chant, ii) celle de chaque *chunk* et iii) celle de chaque phrase et chaque transition.

À partir de ces valeurs, nous avons calculé le Z-score qui est un indice normalisé, selon la formule :

$$Z\ score = \frac{FR\ Evoquée - FR\ spontanée}{SD_{FR\ spontanée}}$$

Le Z-score est calculé à partir des valeurs moyennes obtenues sur les 10 présentations du même chant. Nous avons donc un Z-score correspondant à chacun des chants, chacun des *chunks* de chaque chant, chacune des phrases et chacune des transitions.

Ces calculs ont été réalisés par un programme que j'ai écrit sous Matlab. Ce programme crée un raster plot et péri-stimulus histogramme (PSTH) à partir du nombre de potentiels d'action par classe de 10 ms.

Pour chaque neurone, nous avons mesuré la largeur moyenne de ses potentiels d'action à partir des formes données par le logiciel Spike2. La largeur du potentiel d'action (PA) et le taux d'activité spontanée permettent de distinguer deux types de neurones dans

le HVC: d'une part, les interneurons ont potentiel d'action d'une largeur $<0,4\text{ms}$ et un taux d'activité élevé et d'autre part, les neurones de projection (qui peuvent projeter à l'Aire X, au noyau RA ou au noyau Avalanche) qui ont un potentiel d'action d'une largeur $\geq 0,4\text{ms}$ et un taux d'activité plus faible (Del Negro, 2005; Del Negro and Edeline, 2001, 2002; Mooney, 2000).

F. Statistiques

Les analyses statistiques ont été menées avec le logiciel Statistica (version 8.0; StatSoft). Afin de savoir si le neurone répondait significativement, les valeurs de FR spontanées et évoquées ont été comparées avec un test de Student apparié. Afin de comparer les réponses aux différents chants, nous avons réalisé des ANOVA. Plusieurs facteurs ont été : l'effet « chant » (BOS vs. Locaux ; BOS vs. Globaux ; etc.), l'effet « *chunk* » (*chunk* 1 vs. *chunk* 2 vs. *chunk* 3), l'effet « position » (1, 2 ou 3), l'effet « type de phrases », et l'effet « type de cellules (neurones à PA fin vs. neurones à PA large). Pour comparer les résultats 2 à 2, nous avons utilisé un test post hoc, le test de Tukey HSD.

Une Analyse en composante principale a été réalisée afin de déterminer dans quelle mesure les réponses aux phrases du BOS dépendent de certains paramètres. Pour chacune des phrases du BOS diffusé, nous avons pris comme données, la valeur moyenne du Z-score sur l'ensemble des neurones de l'oiseau, des paramètres acoustiques qui caractérisent les phrases de la phrase (fréquence moyenne, durée de la phrase, largeur de la bande de fréquence, entropie), les probabilités de convergence et de divergence des séquences de 2, 3 et 4 phrases qui incluent la phrase (en dernière position), son occurrence moyenne et sa position relative moyenne dans les chants.

Résultats

A. Réponses aux phrases du BOS

Au cours de l'étude, nous avons enregistré l'activité des neurones du HVC chez 8 canaris anesthésiés. Comme stimulus, nous avons utilisé un BOS considéré comme représentatif des chants de l'individu. Ce BOS comporte entre 8 et 12 phrases (moyenne \pm SEM : $11,7 \pm 1,3$). Nous avons, au total, analysé l'activité de 333 neurones, chacun de ces neurones présente une augmentation significative de son activité par rapport au niveau de base (FR spontanée) à la présentation d'au moins une des phrases du BOS.

Les réponses ont été évaluées par le taux d'activité et le Z-score calculés sur plusieurs niveaux : la durée totale du chant, la durée de chacune des trois chunks qui constituent les chants, et enfin la durée de chacune des phrases.

Tout d'abord nous nous sommes focalisés sur les réponses au BOS. Comme précédemment observé (Del Negro, 2005; Lehongre and Del Negro, 2011), les réponses des neurones chez un même oiseau varient d'un neurone à un autre en termes de phrases qui font répondre du nombre et de type de phrases qui font répondre (**Figure 4**). Les résultats indiquent que chez la plupart des oiseaux, toutes les phrases ont fait répondre au moins un neurone (**Figure 5**). En moyenne, les neurones répondent à ($1,96 \pm 0,25$) phrases. Cependant, comme décrit dans des études antérieures (Sutter and Margoliash, 1994), certaines phrases font préférentiellement décharger un grand nombre de neurones parmi ceux enregistrés (**Tableau 1**). Chez tous les oiseaux, au moins une phrase du BOS fait répondre plus de 50% des neurones. Il est à noter que chez un oiseau, des enregistrements ont été réalisés dans le HVC dans les deux hémisphères. Ce sont les mêmes phrases qui font répondre un grand nombre de neurones au niveau des hémisphères droit et gauche. On ne note pas de corrélation entre la position moyenne de la phrase sur l'ensemble des chants enregistrés, ou dans le BOS, et l'activité neuronale (**Figure 6**).

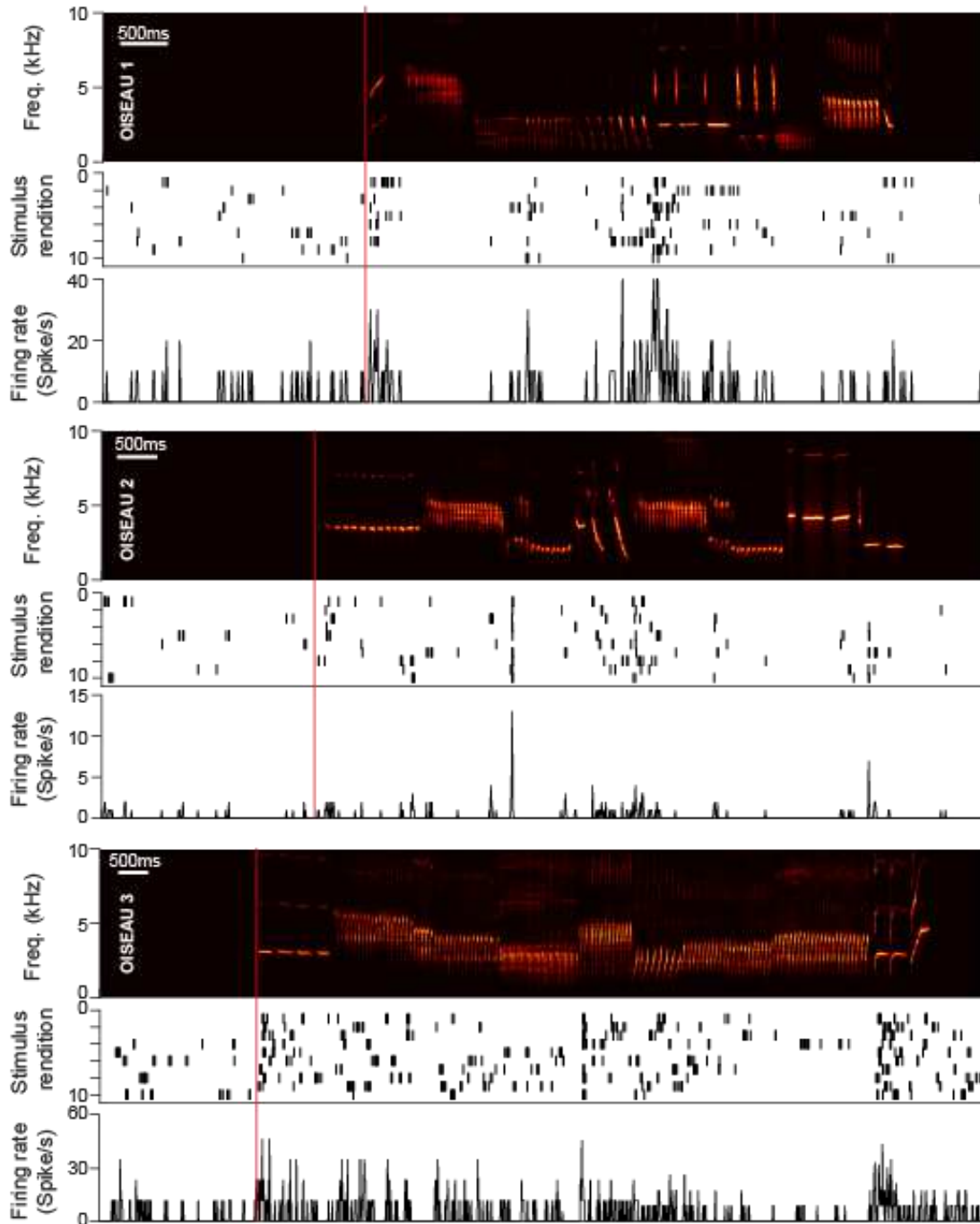


Figure 4 : Exemple de réponses de neurones du HVC lors de la diffusion du BOS, chez 3 oiseaux différents.

L'activité de ces trois neurones est représentée par un raster plot (un point est un PA, une ligne une répétition du chant, de 1 à 10) et le PSTH correspondant (activité moyennée sur 10ms). Au-dessus, le spectrogramme du chant. La ligne rouge indique le début de chaque chant.

oiseau	Pourcentage du nombre de neurones répondant significativement à la phrase											
	phrase 1	phrase 2	phrase 3	phrase 4	phrase 5	phrase 6	phrase 7	phrase 8	phrase 9	phrase 10	phrase 11	phrase 12
1	66	13	6	27	33	33	1	20	4			
2	23	13	27	13	57	23	0	20	7	3		
3	40	20	12	24	28	36	36	8				
4	71	10	39	44	21	25	36	8	31	33		
5	23	1	60	4	50	25	15	35	4	50	31	23
6	78	50	7	83	7	17	18	59	36	21	21	52
7	24	10	50	8	18	18	8	8	100			
8	33	36	1	30	30	52	21	27	39	12	33	

Tableau 1 : Pourcentage de neurones actifs lors de la diffusion de chaque phrase

En rouge : les phrases faisant répondre 50% ou plus de la population neuronale

Colonne : type de phrase ; Ligne : données par oiseau

Nous avons alors cherché à caractériser les réponses des neurones sur la base des caractéristiques des phrases (paramètres acoustiques, position, probabilités). En ce qui concerne les caractéristiques acoustiques ou l'occurrence, comme dans une étude précédente, nous n'avons trouvé aucune corrélation significative entre les différents paramètres acoustiques et le taux d'activité ou le Z-score (**Figure 7** et **Figure 9**).

Nous avons alors cherché à savoir si les réponses dépendaient des probabilités de convergence et/ou de divergence.

En ce qui concerne les probabilités de transition, de convergence et de divergence, nous avons tout d'abord pris en compte les différentes transitions dans le BOS (ex la transition AB dans un chant ABCD, puis la transition BC) et pour chacune, nous avons considéré la probabilité de convergence (C(AB)) et celle de divergence (D(AB)), d'abord séparément. L'analyse des réponses des neurones indique que durant une phrase, le Z-score, correspondant à une normalisation du taux de décharge des neurones, n'est pas significativement corrélé à la probabilité de divergence ni de convergence de la transition qui précède la phrase (**Figure 8**).

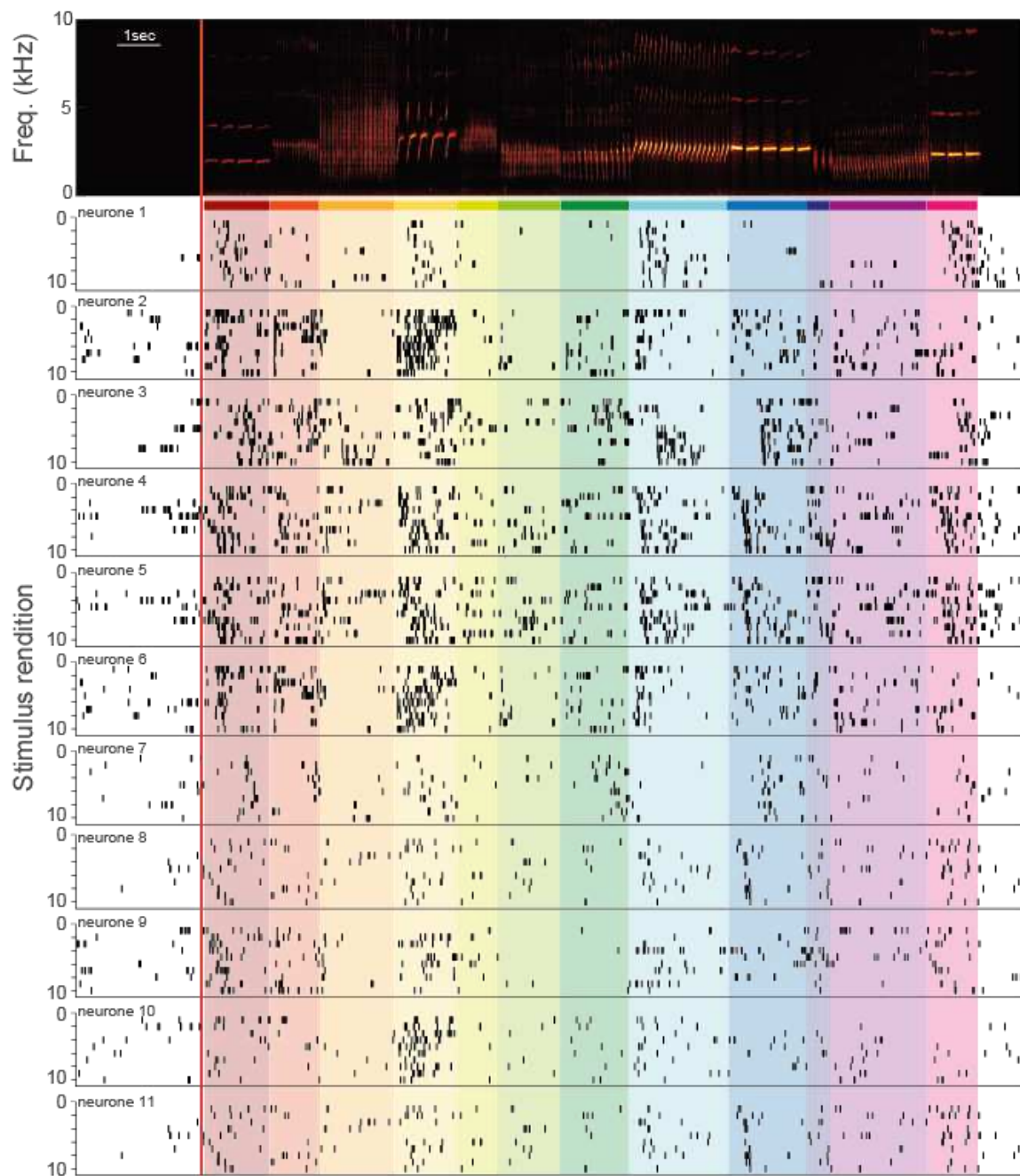


Figure 5 : Diversité dans les réponses des neurones enregistrés chez un même oiseau lors de la diffusion du BOS.

Raster de l'activité de 12 neurones du HVC (point : PA, ligne : une répétition du chant). La ligne rouge : début du chant. Spectrogramme du BOS au-dessus des rasters. Une couleur par phrase.

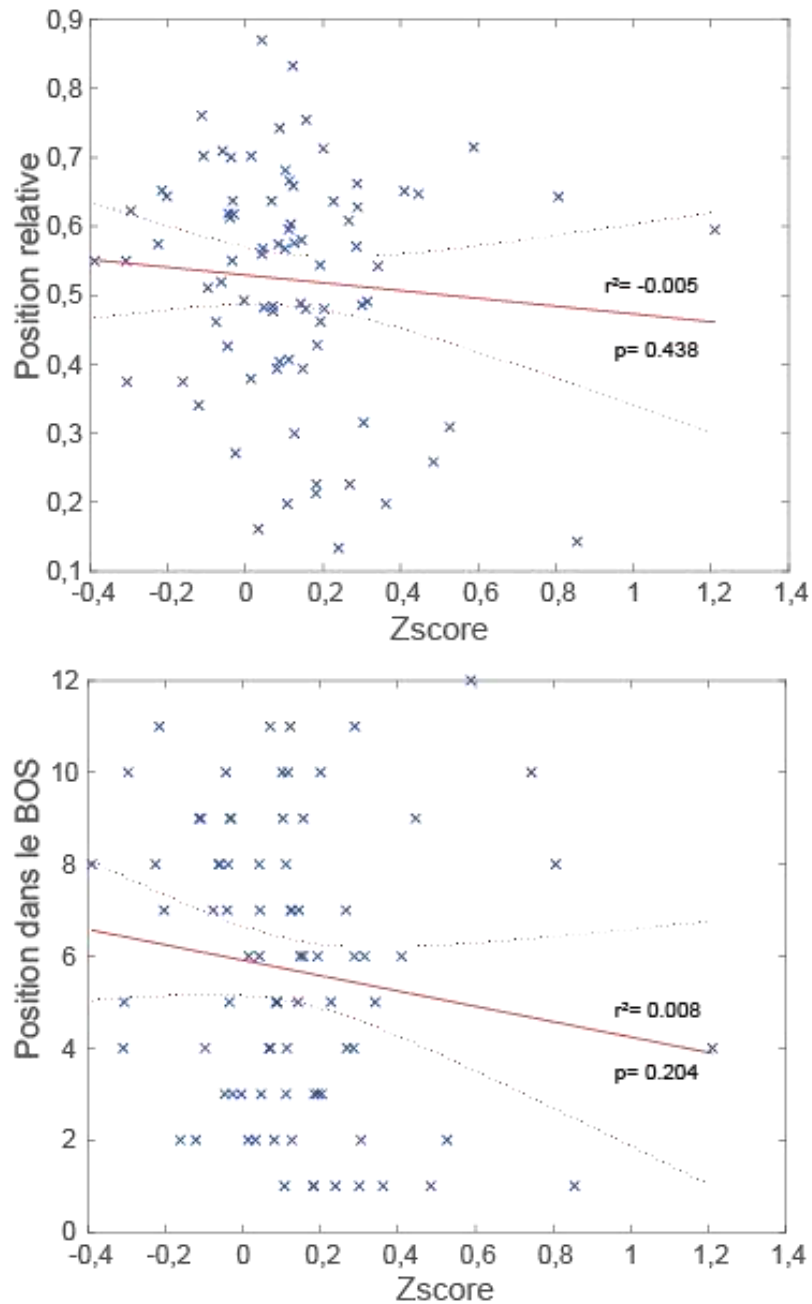


Figure 6 : Absence de corrélation entre la position des phrases et la force des réponses (Z-score) à ces phrases.

Point bleu : phrase (8 BOS diffusés ; un par oiseau) ; Ligne rouge : droite de régression.

(Haut) Aucune corrélation entre la position relative des phrases (normalisation de longueur de l'ensemble des chants analysés pour chaque oiseau à 1) et l'activité neuronale

(Bas) Aucune corrélation entre la position de la phrase dans le BOS et l'activité neuronale

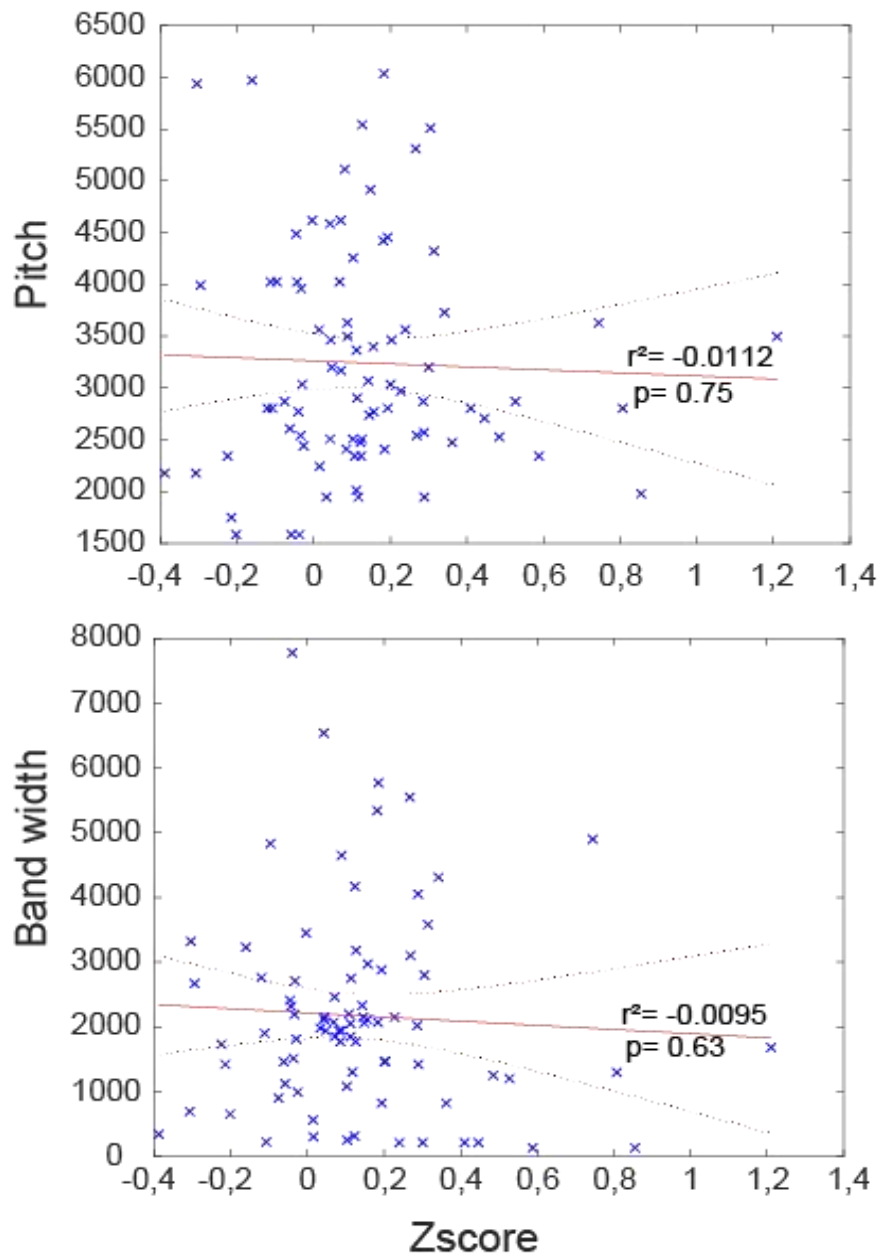


Figure 7 : Absence de corrélation entre la force des réponses (Z-score) et deux des paramètres acoustiques mesurés sur chacune des phrases des BOS diffusés.

Point bleu : phrase (8 BOS diffusés ; un par oiseau) ; Ligne rouge : droite de régression. Les valeurs de p indiquent l'absence de corrélation significative.

(Haut) La fréquence moyenne (pitch) des syllabes des différentes phrases en fonction du Z-score moyen obtenu pour chaque phrase.

(Bas) La largeur de la bande de fréquence (Bandwidth) des différentes phrases en fonction du Z-score moyen obtenu pour chaque phrase.

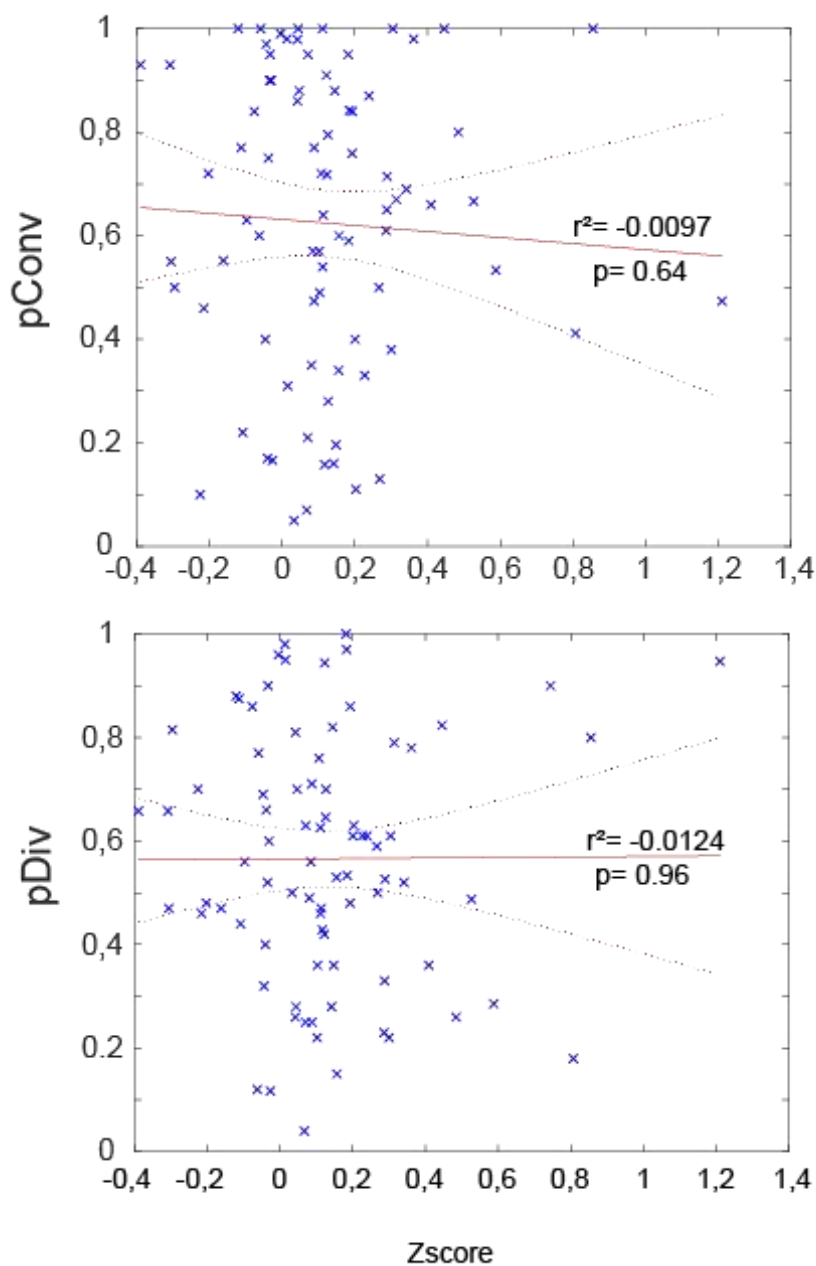


Figure 8 : Absence de corrélation entre les probabilités de transition et la force des réponses (Z-score).

Point bleu : phrase (8 BOS diffusés ; un par oiseau) ; Ligne rouge : droite de régression. Les valeurs de p indiquent que l'absence de corrélation significative.

(Haut) Pour chacune des transitions (ex : XY), valeur de la probabilité de convergence (pConv) en fonction de la valeur du Z-score à la phrase Y.

(Bas) Pour chacune des transitions (ex : XY), valeur de la probabilité de divergence (pDiv) en fonction de la valeur du Z-score à la phrase Y.

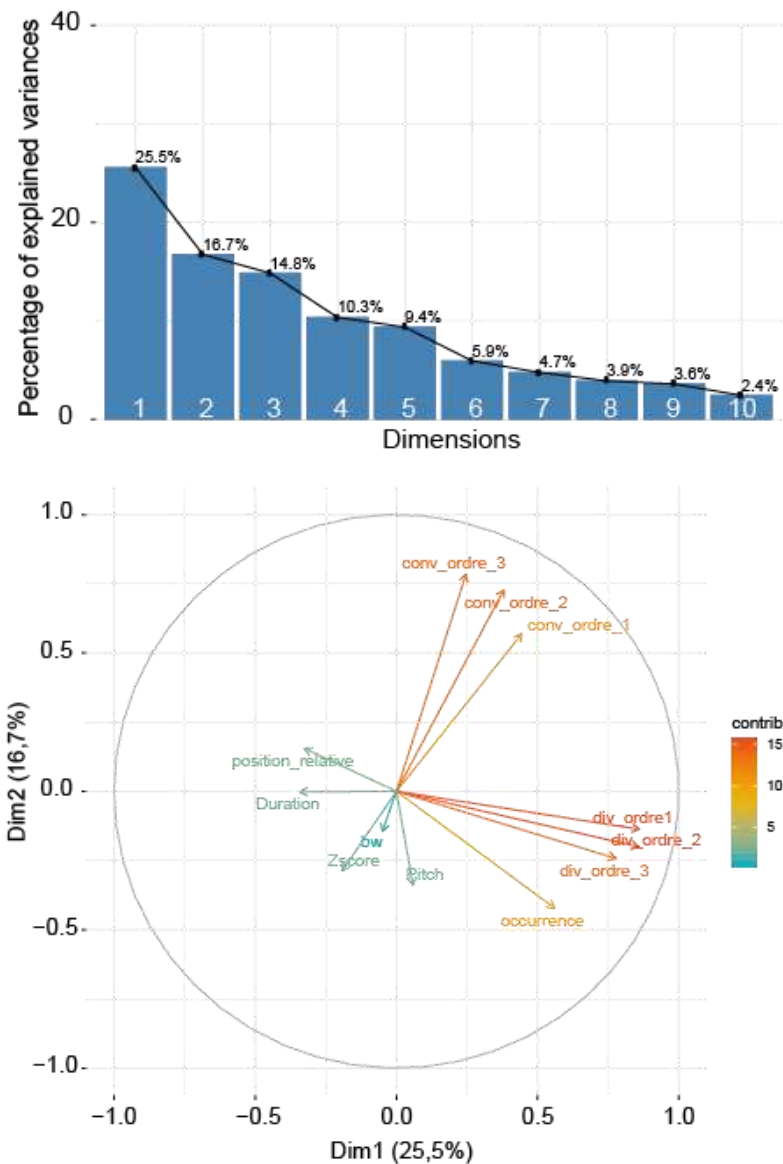


Figure 9 : Analyse en Composante Principale

Corrélation entre les 12 variables (Z-score, position relative, durée, largeur de la bande de fréquence (bw), bande de fréquence (pitch), occurrence, ainsi que les probabilités de divergence et de convergence de l'ordre 1 à 3)

(Haut) Valeurs des dimensions reliant les 12 variables. La dimension explique seulement 25,5% des résultats obtenus à travers l'ensemble des variables mesurées.

(Bas) Représentation cartographique des 12 variables en fonction des dimensions 2 (Dim2) et 1 (Dim1). On note que l'activité (Z-score) est peu corrélée avec l'ensemble des 11 autres variables.

La contribution de chaque variable à une dimension donnée qui les sépare est représentée par un dégradé du bleu vers le rouge.

Afin d'observer s'il est possible de corrélérer les réponses neuronales avec l'une des onze variables mesurées (position relative, durée, largeur de la bande de fréquence, bande de fréquence, occurrence, ainsi que les probabilités de divergence et de convergence de l'ordre 1 à 3) nous avons réalisé une analyse en composante principale (ACP). Aucune des variables mesurées n'est corrélée avec les réponses neuronales aux phrases. Les résultats indiquent que ces différents paramètres ne permettent pas d'expliquer comment les neurones répondent (**Figure 9**).

Sachant que le nombre de neurones répondant significativement à une phrase varie d'un neurone à un autre, nous avons considéré non plus la force des réponses (Z-score), mais le nombre de neurones qui répondent à chacune des phrases. Les résultats ont indiqué que les transitions avec une probabilité de convergence forte (supérieure à 0,5 ; $n=54/84$) ne font pas répondre plus de neurones que celles avec des probabilités de convergence faibles ($<0,5$ ($n=30/84$) : respectivement 25% ($\pm 4\%$) et 26% ($\pm 9\%$) des neurones. Il en est de même lorsqu'on considère les probabilités de divergence fortes ($>0,5$; $n=48/84$; 27% des neurones $\pm 5\%$; $<0,5$; $n=36/84$; 26% des neurones $\pm 6\%$). Sachant cela, nous avons pris en compte les deux types de probabilités, de divergence et de convergence, ensemble. Mais, nous nous sommes placés non pas, en prenant en compte chacune des transitions du BOS, mais chacune des phrases. Pour expliquer ce point, nous avons pris en compte la probabilité de convergence pour la transition qui précède la phrase et la probabilité de divergence pour la transition qui suit la phrase. Par exemple, pour la phrase B, dans le chant ABC..., nous avons considéré (C(AB)) et (D(BC)). Nous avons, pour chaque phrase des BOS, considéré ces deux probabilités, selon qu'elle était supérieure (+) ou inférieure à 0,5 (-) (**Figure 10A et B**). De cette façon, chaque phrase entre dans une des 4 catégories suivantes :

D(AB)+/ C(BC)+ : les probabilités de convergence et divergence sont fortes ($>0,5$), ce qui précède ou suit la phrase est prédictible, la phrase est au sein d'un *chunk*.

D(AB)-/ C(BC)- : il est difficile de prédire ce qui précède ou suit la phrase, la phrase n'est pas dans un *chunk*.

D(AB)+/ C(BC)- : ce qui suit la phrase est prédictible et ce qui la précède non, la phrase est au début d'un *chunk*.

D(AB)-/ C(BC)+ : ce qui précède la phrase est prédictible et ce qui la suit non, la phrase est une fin de *chunk*.

La répartition des 84 phrases entre ces quatre catégories est respectivement de 36, 10, 20 et 17 (**Figure 10C**). Nous avons distingué deux types de phrases, celles qui font répondre plus ou moins de 50% des neurones (**Figure 10D**). Comme le montre la **Figure 10**, si on regarde le profil des 14 phrases faisant répondre 50% ou plus de neurones, leur répartition entre les quatre catégories est différente de la répartition générale ($\text{Chi}^2=12,31$; $p<0.006$). Ces phrases sont majoritairement dans la catégorie div+/conv- : c'est-à-dire que, pour ces phrases en question, si on considère la phrase qui précède et la phrase qui suit, la probabilité d'avoir la phrase qui précède n'est pas forte, mais en revanche la probabilité d'avoir la phrase suivante est la plus forte, ce qui s'observe pour une phrase qui est au début d'un chunk. On observe aussi une faible représentation des phrases de la catégorie div+/conv+ qui était pourtant dominante si l'on observe le profil de toutes les phrases, ce qui indique que les phrases qui sont au sein d'un chunk ne sont pas celles qui ont le plus de chance de faire répondre les neurones. Ces résultats indiquent donc que les neurones du HVC pourraient préférentiellement répondre à un certain type de phrase. En revanche, les 69 phrases qui font répondre moins de la moitié du HVC, n'ont pas un profil type (Chi^2 $p>0.05$). De même si l'on sélectionne parmi elles, seulement les 32 phrases faisant répondre le moins les neurones ($<20\%$ de la population).

B. Importance du type de phrases qui précèdent

Dans le chant du canari, une même phrase peut être incluse dans différentes séquences au sein d'un ou plusieurs chants. Parmi les BOS sélectionnés dans cette étude, deux incluent la répétition d'une ou plusieurs phrases. Nous avons ainsi pu analyser si la réponse neuronale à une phrase donnée est sensible au contexte dans lequel elle se situe. Comme le montre l'exemple de la **Figure 11**, la force des réponses d'un neurone à une phrase donnée peut changer selon ce qui précède la phrase (**Figure 11**, phrase soulignée en bleu et rouge). Si on considère les probabilités de divergence et de convergence, elles ne permettent pas d'expliquer les différences dans les réponses à ces phrases (1^{ère} phrase, en bleue, 1^{ère} séquence vs. 2^{ème} séquence, $p\text{Div}(XY) : 0,49$ vs. $0,36$; $p\text{Conv}(XY) : 0,35$ vs. $0,2$; $p\text{Div}(TXY) : 0,95$ vs. $0,75$; $p\text{Conv}(TXY) : 0,36$ vs. $0,29$). Cela suggère que le contexte, en termes de place, d'identité de la ou des phrases précédentes, pourrait influencer sur la force des réponses. Les analyses nous ont également indiqué qu'une des deux phrases avaient fait répondre plus de 50% des neurones, mais seulement lorsqu'elle était dans la première séquence (phrase en rouge).

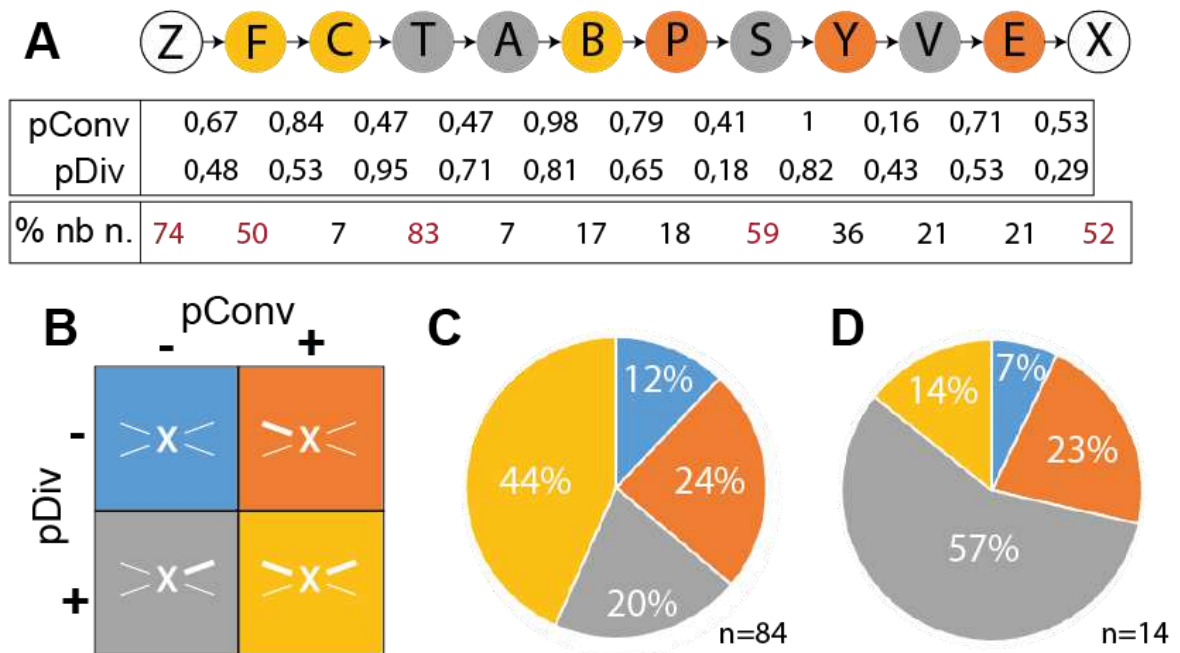


Figure 10: Caractérisation en termes de probabilités de transition des phrases qui font répondre plus de 50% des neurones.

- (A) Probabilité de convergence de la transition qui précède ($pConv$) et probabilité de divergence de la transition qui suit ($pDiv$) pour chacune des phrases du BOS de **la Figure 38**. Par exemple pour la phrase F : $pConv(ZF)$ et $pDiv(FC)$. Sous du tableau, le pourcentage de neurones répondant significativement à chacune des phrases (en rouge si $>50\%$). Les couleurs correspondent au profil de la phrase (B).
- (B) Quatre catégories de phrases ont été distinguées à partir de la valeur de la probabilité de divergence et de la probabilité de convergence ; (+) équivaut à $>0,5$ et (-) équivaut à $<0,5$
- (C) Répartition de l'ensemble des phrases dans les quatre catégories.
- (D) Répartition des phrases générant des réponses significatives d'au moins 50% de la population neuronale ans les quatre catégories.

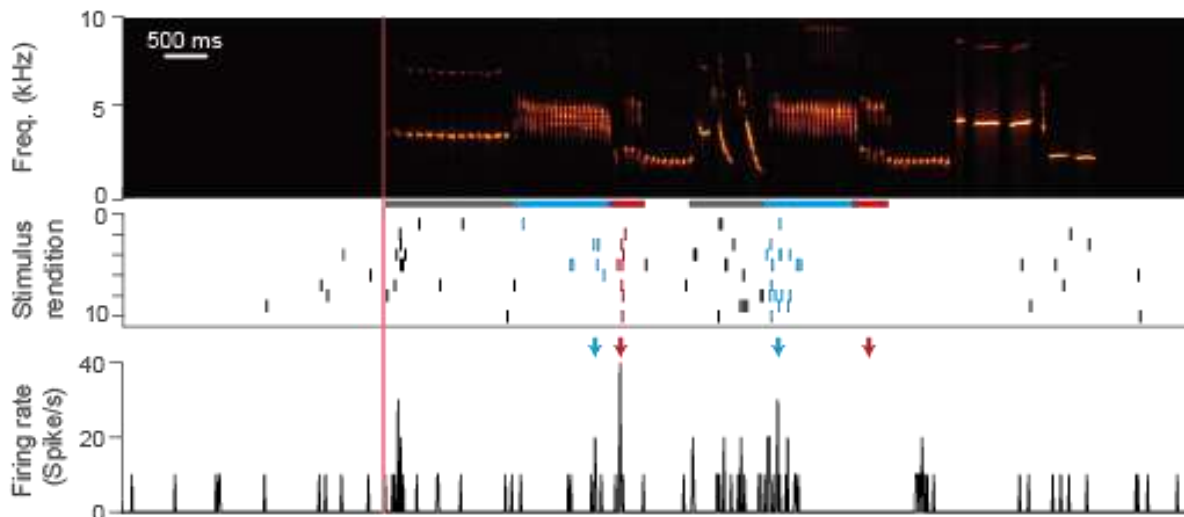


Figure 11 : Réponses neuronales lors de la diffusion d'un BOS composé de deux fois la même séquence de deux phrases.

Spectrogramme d'un BOS dans lequel une séquence de deux phrases (trait bleu et rouge) est répétée deux fois.

L'activité neuronale est représentée sous la forme d'un raster plot et d'un PSTH. La ligne rouge indique le début du chant. Les deux phrases ne sont pas précédées par la même phrase (trait gris). Selon où se trouve les deux phrases, elles n'évoquent pas les mêmes changements d'activité (flèches bleues et rouges).

En fonction de son contexte, les réponses aux deux phrases identiques sont différentes (les réponses associées sont notifiées par le même code couleur).

C. Réponses aux autres types de chant

Nous avons présenté d'autres chants que le BOS. Parmi eux figure un chant conspécifique. Comme attendu, les neurones du HVC répondent significativement moins par rapport au BOS (moy \pm SEM, BOS $0,72 \pm 0,02$, CONSPE= $0,32 \pm 0,03$; $p < 0,05$, $n = 210$).

Pour déterminer à quoi sont sensibles les neurones du HVC, nous avons créé deux types d'altérations du BOS, des altérations dites « locales » et d'autres dites « globales » (voir matériels et méthodes).

Un changement local consiste en une permutation entre deux phrases d'un *chunk* donné. Dans le BOS, les probabilités de transition entre ces deux phrases sont fortes (ordre 1 div : $0,63 \pm 0,04$; conv : $0,72 \pm 0,06$). Les BOS étant constitué de 3 *chunks*, nous avons trois versions du BOS (L1, L2, L3). Dans ces trois chants, suite à la permutation, trois transitions, dont celles entre les deux phrases permutées, sont égales à zéro. Ces chants ont été créés afin de savoir dans quelle mesure les réponses aux phrases dépendaient de l'ordre à l'intérieur des séquences.

Un changement global consiste au déplacement d'un ou plusieurs *chunks* tout en conservant l'agencement des phrases en leur sein. Chez tous les oiseaux nous avons diffusé deux chants obtenus après le déplacement d'un seul *chunk* : déplacement du *chunk* 1 après les *chunks* 2 et 3 (chant G231) ou du *chunk* 3 avant les *chunks* 1 et 2 (chant G312).

Pour chacun des deux chants créés, les première et dernière phrases, symbolisées respectivement par les lettres *i* et *f*, ne sont pas concernées par les modifications effectuées, ce sont les mêmes que celles du BOS et elles sont restées à leur place. Cela s'accompagne d'une baisse des probabilités de convergence et de divergence d'ordre 1 au niveau de 3 transitions (**Figure 12 ; Tableau 2**).

Ces deux chants ont été créés afin de savoir si, maintenir l'ordre et les probabilités des transitions au sein des *chunks*, quel que soient leur position, suffisaient à conserver les réponses telles qu'elles sont observées préalablement lors de la diffusion du BOS naturel. Si on considère le BOS, G231 et G312, chacun des 3 *chunks* est présent à chacune des 3 positions. Si l'on considère les 5 versions du BOS, en termes de transitions impactées, elles sont donc au nombre de trois que ce soient pour les chants L1, L2 et L3 et G231 et G312 (**Figure 12**).

Par ailleurs, il n'y a pas de différence statistique entre la moyenne des probabilités de transition entre les différentes versions du BOS (pour les transitions d'ordre 1, 2 et 3 ; $p > 0.05$; **Tableau 2**).

Probabilités de divergence pour chacun des chants						
Oiseau	BOS	L1	L2	L3	G231	G312
1	0,56	0,39	0,40	0,45	0,49	0,49
2	0,68	0,58	0,27	0,16	0,50	0,35
3	0,54	0,39	0,34	0,35	0,39	0,44
4	0,58	0,33	0,49	0,49	0,40	0,44
5	0,51	0,35	0,36	0,32	0,36	0,36
6	0,50	0,35	0,39	0,40	0,35	0,33
7	0,67	0,44	0,48	0,50	0,49	0,52
8	0,60	0,38	0,40	0,51	0,47	0,41
Moyenne	0,58	0,40	0,39	0,40	0,43	0,42
SEM	0,025	0,031	0,027	0,045	0,023	0,026

Probabilités de convergence pour chacun des chants						
Oiseau	BOS	L1	L2	L3	G231	G312
1	0,62	0,48	0,45	0,50	0,52	0,51
2	0,68	0,58	0,27	0,16	0,50	0,35
3	0,55	0,37	0,34	0,36	0,43	0,46
4	0,67	0,38	0,46	0,49	0,39	0,39
5	0,63	0,44	0,48	0,45	0,51	0,49
6	0,52	0,42	0,36	0,36	0,41	0,40
7	0,71	0,46	0,54	0,52	0,54	0,59
8	0,68	0,40	0,44	0,53	0,55	0,50
Moyenne	0,63	0,44	0,42	0,42	0,48	0,46
SEM	0,026	0,027	0,033	0,047	0,024	0,029

Tableau 2: Probabilités de transition au sein des différents chants diffusés :

Ligne : oiseau ; Colonne : des probabilités de divergence (haut ; $pDiv(xy)$) et de convergence (bas ; $pConv(xy)$) pour chacun des chants (BOS, L1, L2, L3, G231 et G312)

Lignes en rouge : moyenne et le SEM sur l'ensemble des huit oiseaux.

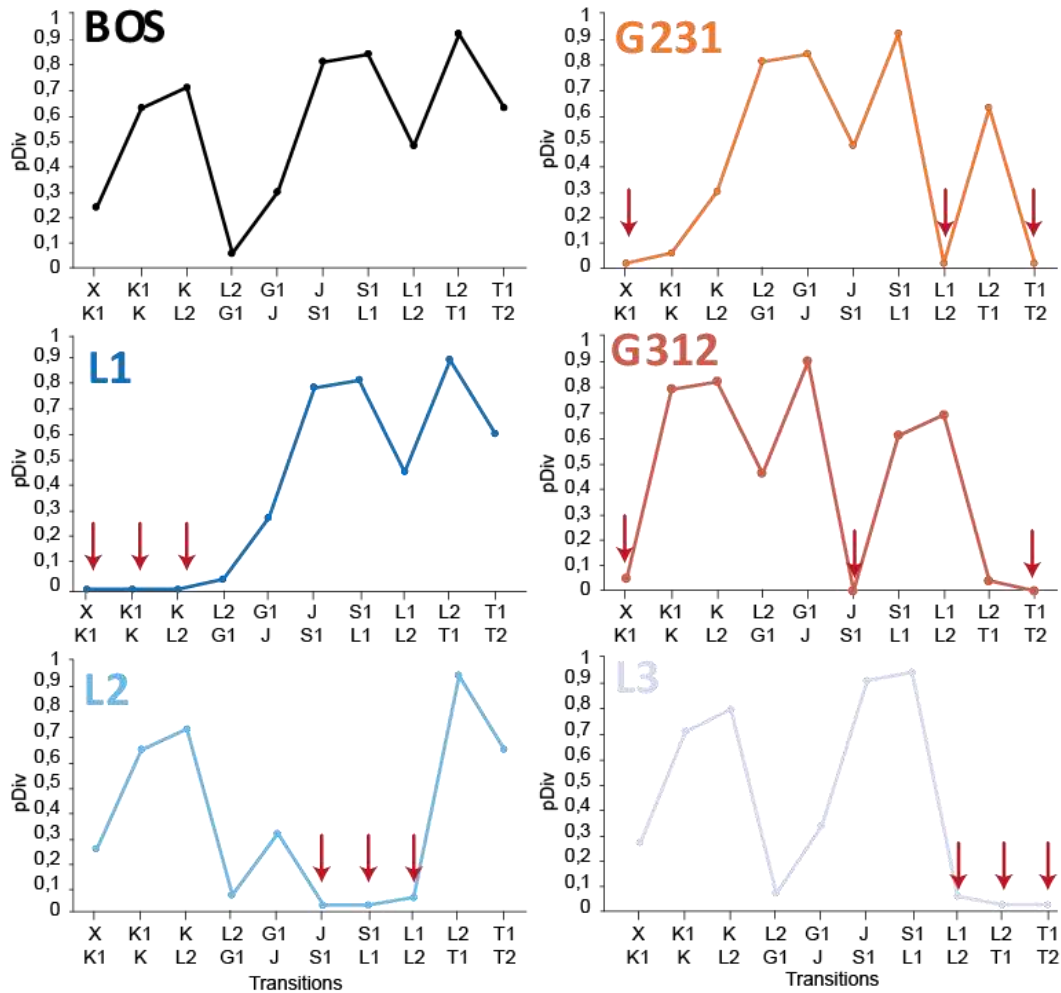


Figure 12 : Permuter deux phrases ou changer un chunk de place fait chuter trois probabilités de divergence, si on compare au BOS.

Valeurs des probabilités de divergence pour les différentes transitions présentes dans les chants diffusés chez un oiseau donné. Les flèches rouges indiquent les transitions altérées. Pour chaque chant, trois transitions sont touchées. Seule leur localisation évolue entre les chants.

Tout d'abord afin d'examiner l'impact des changements locaux et globaux sur les réponses des neurones ($n=206$), nous avons calculé les Z-scores sur la durée totale des chants. Les résultats indiquent que, contrairement aux chants L1, L2 et L3, la diffusion des stimuli G231 et G312 induit une diminution significative des réponses neuronales par rapport à celles obtenues à la diffusion du BOS (test t de Student ; globaux $p<0.05$; locaux $p>0.5$) (**Figure 13**). La comparaison des Z-scores entre les chants G231 et G312 indique une différence significative dans les réponses, le chant G312 impactant plus fortement les réponses que le chant G231 (t student $p<0.05$). Cela indique que les deux types de changement impactent différemment l'activité des neurones : le HVC est plus sensible à des changements dans la position des *chunks* que dans l'ordre des phrases.

Pour aller plus loin, nous avons alors calculé les Z-scores à partir de l'activité moyenne pendant chacun des *chunks*. La **Figure 14** présente, pour chacun des 3 *chunks*, la force des réponses des neurones au BOS et aux 5 autres versions du BOS ($n=206$). L'ANOVA indique un effet du type de chant sur les réponses pour le *chunk* 1 ($F_{5, 1025}=13,32$; $p<0,001$), le *chunk* 2 ($F_{5, 1025}=10,01$; $p<0,001$), mais pas pour le *chunk* 3 ($p=0,5$). En ce qui concerne les changements locaux, malgré le changement d'ordre au sein d'un *chunk* donné (*chunk* 1 pour L1, *chunk* 2 pour L2 et *chunk* 3 pour L3) qui affecte toutes les probabilités de transition au sein de ce *chunk*, les valeurs des Z-score ne diffèrent pas significativement pas celui du même *chunk* dans le BOS (tests post-hoc ; $p>0.05$) (**Figure 14**). Les *chunks* au sein desquels aucun changement dans l'ordre des phrases n'a eu lieu, ne montrent pas non plus de différences par rapport au BOS.

Les changements globaux quant à eux, s'accompagnent d'une diminution d'activité, pendant la diffusion d'au moins un *chunk*, si on compare par rapport au BOS : pendant le *chunk* 1 du chant G231 ($p=0,002$) et pendant les *chunks* 1 et 2 du chant G321 ($p<0,001$ et $p=0,001$, respectivement ; **Figure 14**). Ces résultats suggèrent que les réponses aux phrases du *chunk* 1 dépendent de ce qui précède. De plus, les réponses pendant le *chunk* 1 sont significativement moins importantes selon qu'il soit diffusé dans G312 ou G231 ($p=0,009$). Dans ces deux stimuli, le *chunk* 1 est précédé du *chunk* 3 et seulement ce qui précède le *chunk* 3 varie, ces résultats suggèrent que les réponses aux phrases du *chunk* 1 dépendent de ce qui précède le *chunk* 3. Cela suppose que la sensibilité aux phrases du *chunk* 1 résulte d'un effet du contexte sur le long terme. De même, dans le chant G312, la baisse des réponses pendant le *chunk* 2, toujours précédé comme dans le BOS par le *chunk* 1, suggère également un long effet du contexte sur la sensibilité aux phrases du *chunk* 2.

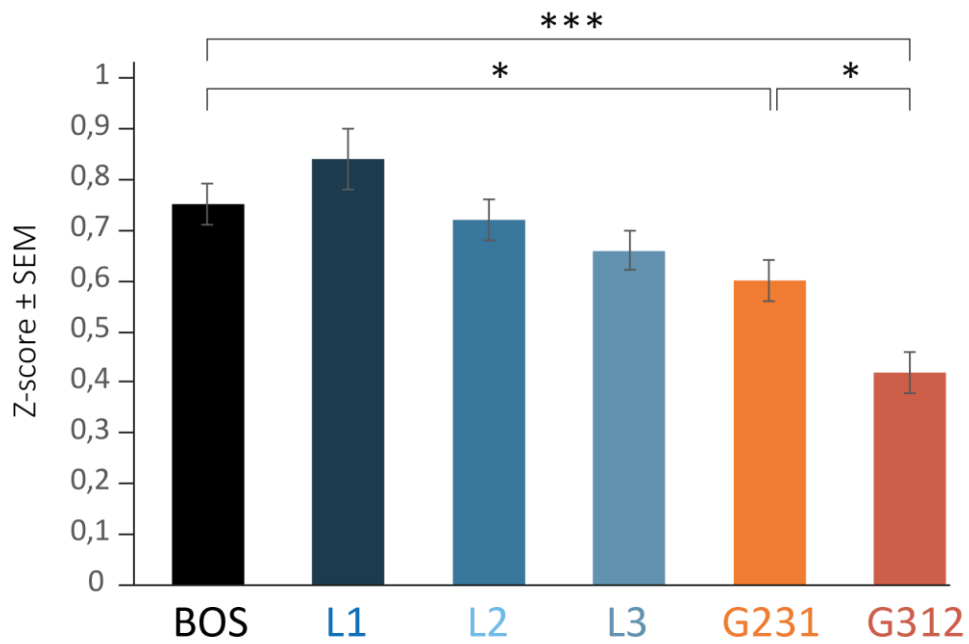


Figure 13 : Réponses moyennes des neurones pendant la durée des différents chants

Les histogrammes présentent les valeurs des Z-scores pour chaque chant diffusé (reprenant le code couleur de la figure précédente)

Seul les chants G231 et G312 induisent une diminution significative des réponses par rapport au BOS.

(n=206 *** équivaut à <math><0.001</math> ; ** <math><0.01</math> ; * <math><0.05</math>)

En revanche, ces résultats ne montrent pas de sensibilité des neurones aux phrases du *chunk 3*, étant donné l'absence de différences significatives avec les réponses du *chunk 3* au BOS.

Pour un nombre plus élevé de neurones ($n=333$), nous avons analysé les réponses aux chants BOS, G231 et G312. L'ANOVA indique un effet du facteur « type de chant » ($F_{2, 664}=4,85$; $p=0,008$), du facteur « position » ($F_{2, 664}=31,74$; $p<0,001$), et une interaction entre les deux facteurs ($F_{4, 1328}=15,29$; $p<0,001$). La **Figure 15** présente la force de ces réponses aux 3 *chunks* de ces trois chants selon leur position, afin de montrer l'évolution des réponses d'un *chunk* au suivant au sein de ces chants. On observe que les réponses au *chunk* en première position des 3 chants (*chunk 1* du BOS, *chunk 2* du chant G231 et *chunk 3* du chant G312), sont équivalentes.

En ce qui concerne le BOS, les réponses d'un *chunk* au suivant ne changent pas significativement (**Figure 15**). En revanche, on observe une chute significative des réponses du *chunk 3* au *chunk 1* dans G312 (position 1 vs. 2 ; $p<0,001$). Cette chute s'observe également pour G231 (position 2 vs. 3, $p<0,001$), mais elle est significativement moins importante que dans G312 (position 2 de G312 vs. position 3 de G231, $p<0,001$), sur les réponses au *chunk 1*. Pour un nombre plus important de neurones, on retrouve donc un long effet du contexte.

Dans les chants BOS, G231 et G312, les trois *chunks* sont chacun aux trois positions, mais certaines transitions ne sont présentes (exemple : la transition du *chunk 2* au *chunk 1*). Afin de déterminer si les autres transitions, que celles présentes dans ces 3 chants, peuvent affecter les réponses, nous avons créé trois autres stimuli globaux : G132, G213 et G321 qui ont été présentés à 3 oiseaux (nombre de neurones= 84). Tout d'abord, nous avons analysé l'activité sur la durée totale du chant. L'analyse statistique indique un effet du facteur chant ($F_{5,360}=5,39$; $p<0,001$). Comme le montre la figure 16, le chant G312 n'est pas le seul à induire une activité moins importante que celle obtenue lors de la diffusion du BOS, on observe également moins de réponses lors de la diffusion des chants G213 ($p=0,01$) et G321 ($p<0,001$). Concernant les réponses aux 3 *chunks* dans les trois nouveaux chants, G132, G231 et G321 (**Figure 16**), les résultats indiquent une interaction entre les facteurs chants et position ($F_{4,300}=9,57$; $p<0,001$) et une chute des réponses lorsque le *chunk 1* suit le *chunk 2*, que ce soit dans G213 ($p<0,001$) ou dans G321 ($p=0,006$).

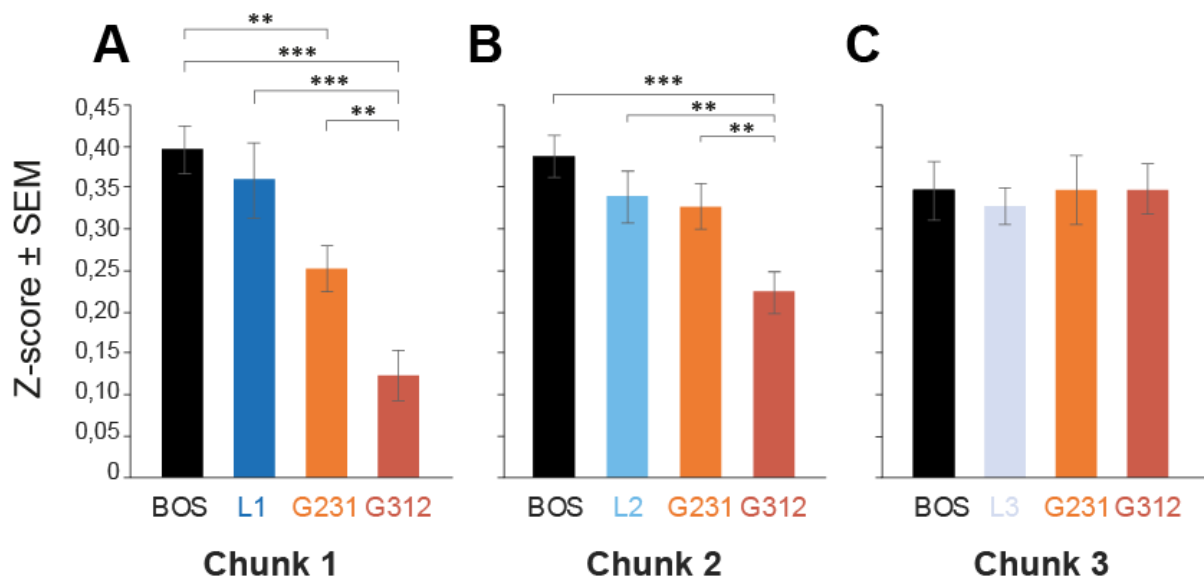


Figure 14 : Réponses moyennes des neurones pendant la durée des trois chunks des différents chants.

Quel que soit le chunk considéré, la permutation de deux phrases n'induit pas de changements dans les réponses par rapport au BOS. (n=206 ; *** équivaut à <0.001 ; ** <0.01).

Le *chunk* 1 induit donc moins de réponses que le *chunk* précédent, c'est-à-dire quand il est en position 2 ou 3, quel que soit le type de *chunk* qui le précède. On n'observe pas cela pour les deux autres *chunks*.

Ces résultats nous amènent deux informations importantes. Premièrement, un changement local n'entraîne pas d'altérations des réponses neuronales sur l'ensemble du *chunk* visé même quand l'ordre y est fortement modifié. Deuxièmement, un changement global peut affecter significativement les réponses à un ou plusieurs *chunks* quand bien même l'enchaînement des phrases en leur sein est préservé. Cependant, ce n'est pas toujours le cas.

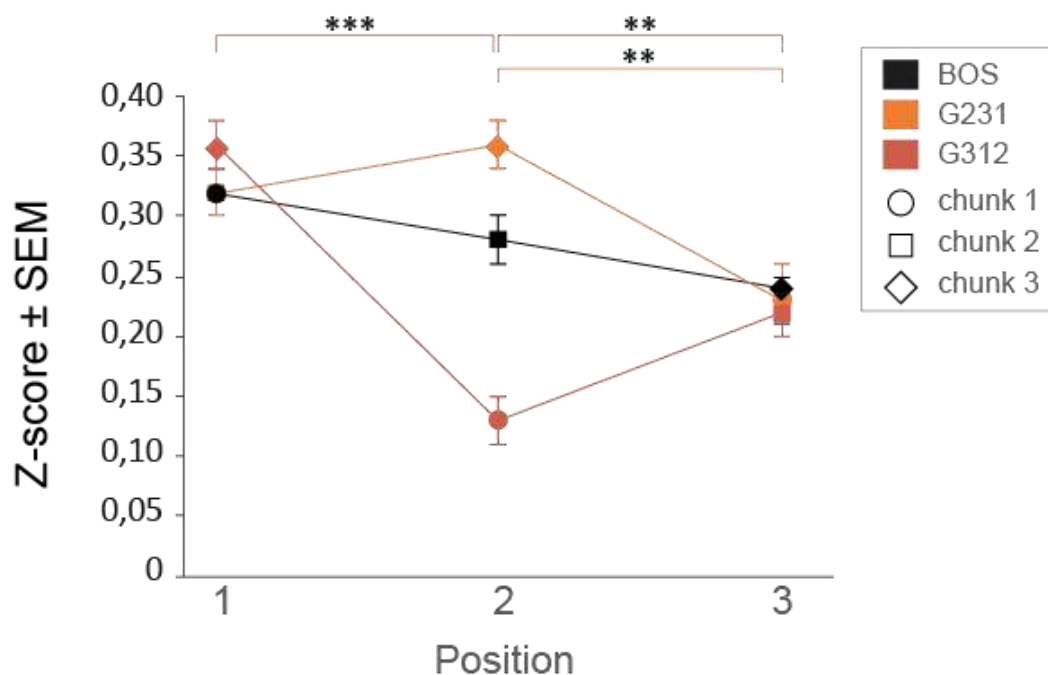


Figure 15 : Réponses neuronales aux chunks selon leur position dans les chants.

Réponses des neurones suite à la diffusion du BOS et des chants G231 et G312.

Les chunks dans le chant sont présentés selon leur ordre de diffusion (position 1, 2 ou 3) et symbolisés par un rond (chunk 1), un carré (chunk 2) ou un losange (chunk 3). La couleur correspond au chant diffusé (noir pour le BOS ; orange pour G231 ; rouge pour G312).

(n=333 ; *** équivaut à <0.001 ; ** <0.01)

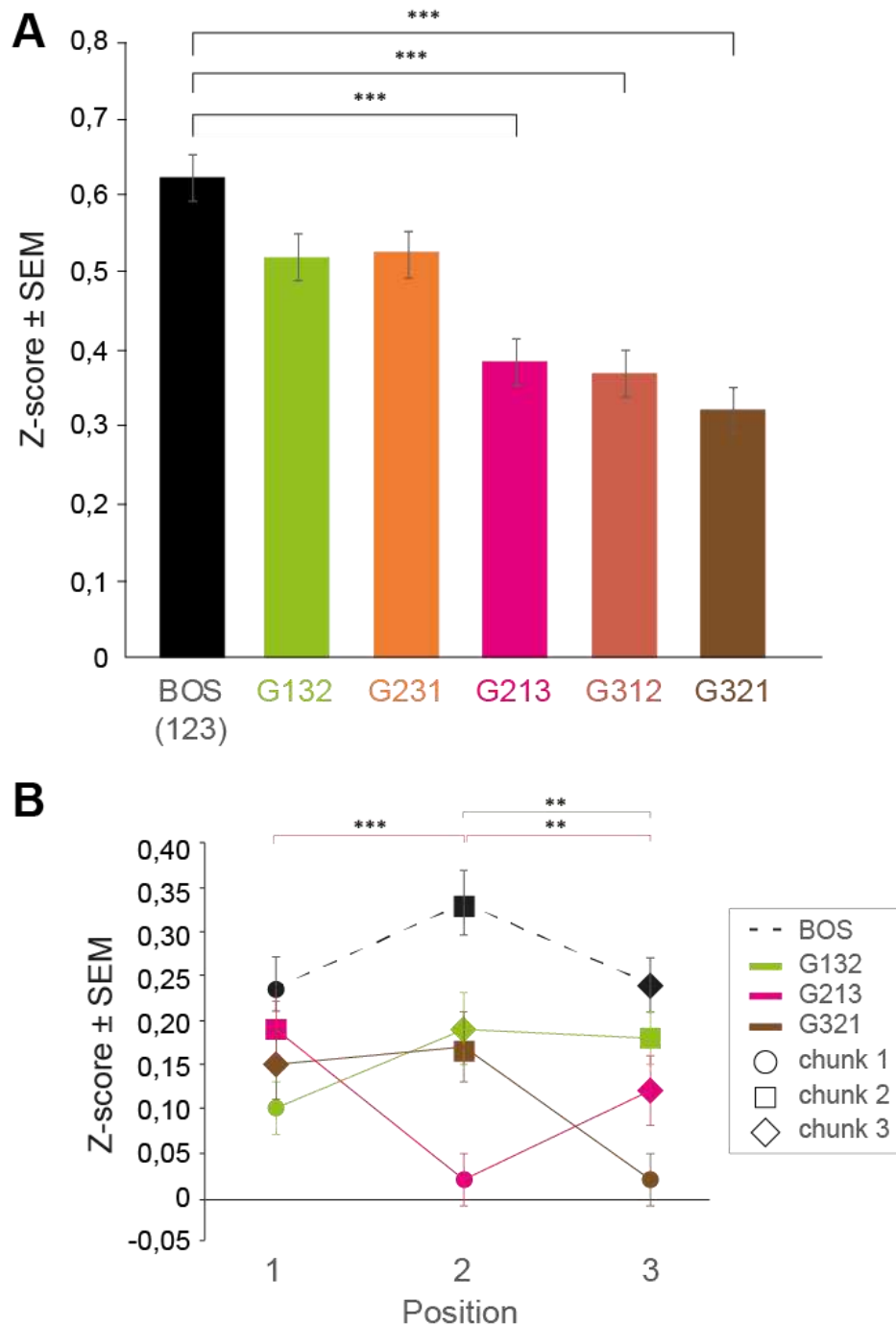


Figure 16 : Réponses neuronales lors de la diffusion d'autres types dans lesquels les chunks s'enchaînent différemment

(A) Réponses moyennes des neurones pendant la durée des chants créés à partir du déplacement des chunks. Les chants présentent l'ensemble des combinaisons possibles des trois chunks.

(B) Réponses neuronales pendant la durée de chacun des chunks des chants G132, G213 et G321.

Ligne pointillée : BOS

(n=84 ; *** équivaut à <0.001 ; ** <0.01)

D. Impact des changements globaux et locaux sur les réponses aux phrases

Nous avons voulu savoir quel impact avaient les changements dans l'ordre des phrases du BOS sur les réponses aux phrases.

La première phrase du BOS la phrase *I*, n'étant pas concernée par les changements effectués pour créer les différentes versions du BOS, ne montre aucune différence significative, quel que soit le chant diffusé (L1, L2, L3, G231, G312) par rapport au BOS.

Tout d'abord, nous avons cherché à déterminer l'impact d'un changement local, c'est-à-dire la permutation de deux phrases. Ces permutations conduisent à ce que trois transitions aient des probabilités égales à zéro pris en compte dans les analyses, les réponses à quatre phrases, les deux ayant permuté et celle qui respectivement précède et suit (**Figure 17**). Les résultats statistiques indiquent un effet significatif de ces permutations : un effet significatif des facteurs « chant » et « phrases » et une interaction significative entre ces deux facteurs pour les trois comparaisons (L1, L2 ou L3 vs. BOS). Que ce soit pour L1, L2 ou L3, les réponses à la 2^{ème} phrase, celle qui passe devant (phrase *j* sur la **Figure 17**), diffèrent des réponses à ces mêmes phrases dans le BOS ($p < 0,01$ dans les trois cas). Si on considère le chant L1, alors l'effet est une diminution des réponses. Si on considère L2 ou L3, il y a une augmentation des réponses. Ces résultats indiquent donc un effet de la permutation, changer le type de phrase impacte les réponses sur la première phrase déplacée. En revanche, on n'observe pas de différences en ce qui concerne la force des réponses à la phrase, qui en permutant passe derrière, (phrase *i* de la **Figure 17**) et à la phrase qui suit la permutation (phrase *j+1* sur la **Figure 17**), bien que les phrases qui les précédaient aient changé et que les probabilités soient égales à zéro. Les changements locaux n'affectent donc pas les réponses à toutes les phrases dont les probabilités de transitions ont été modifiées, mais seulement à une seule (celle qui a été permutée et est passée devant lors de la diffusion).

Nous avons également analysé les réponses à chacune des phrases des trois *chunks* des chants G231 et G312. Si pour un oiseau, le nombre de phrases dans le *chunk* considéré est de deux phrases, nous avons réalisé les analyses statistiques concernant l'ensemble des oiseaux, seulement sur les deux premières phrases du *chunk* en question afin d'intégrer l'ensemble des oiseaux dans nos analyses (**Figure 18**).

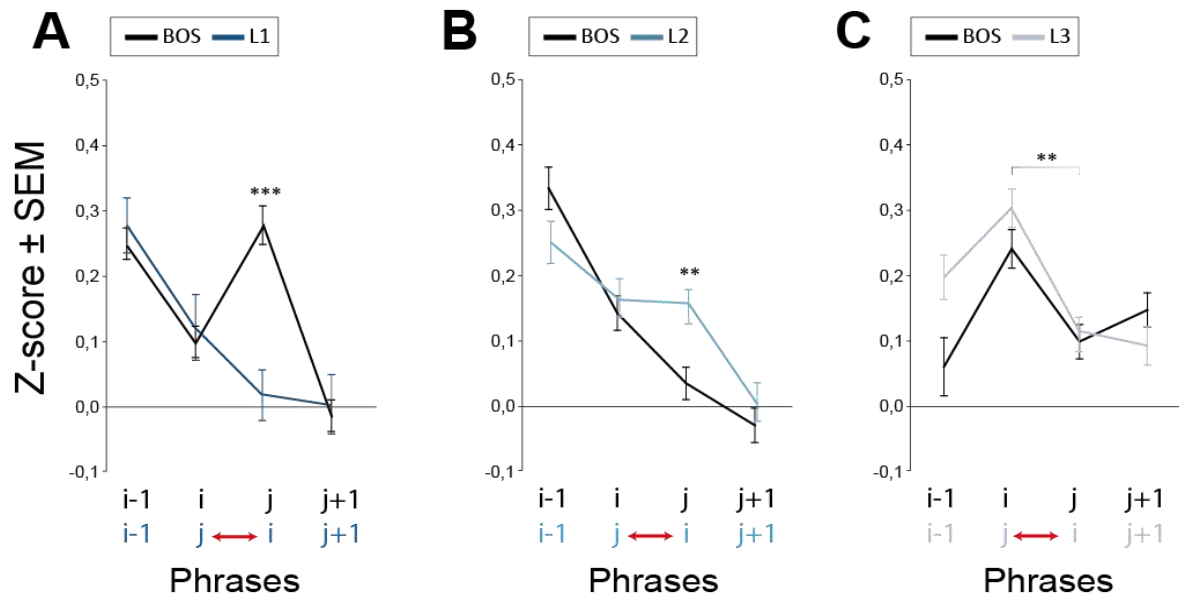


Figure 17 : Effet de la permutation de deux phrases sur les réponses neuronales

ligne noire : BOS ; lignes bleues : permutation des phrases *i* et *j* dans le chunk (chunk 1 : L1, chunk 2 : L2, chunk 3 : L3) ; flèche rouge : phrases permutées. Alignement des phrases à partir de la première nouvelle transition (*i*-1 – *i* pour le BOS et *i*-1 – *j* après permutation). On observe que seule la 3^{ème} phrase du BOS, celle qui après permutation passe en 2^{ème} position, est significativement impactée par le changement sur l'analyse des résultats sur l'ensemble des oiseaux ($n=332$; *** équivaut à <0.001 ; ** <0.01)

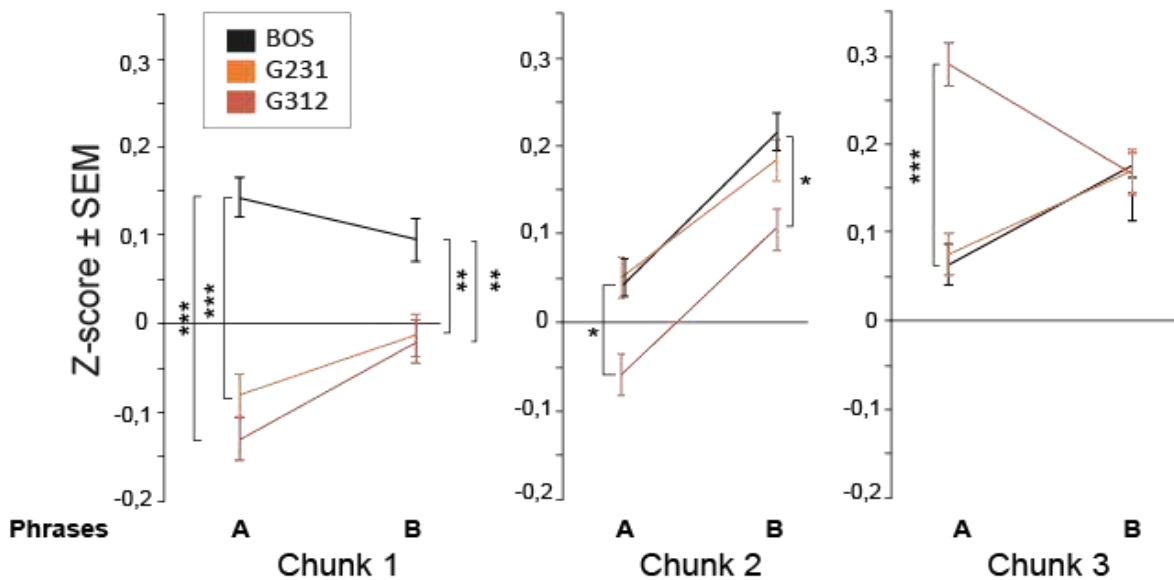


Figure 18 : Réponses neuronales aux phrases après un changement dans l'ordre des chunks

Chaque phrase est notifiée par une lettre et a été alignée selon le début de chaque chunk.

($n=332$; *** équivaut à <0.001 ; ** <0.01 ; * <0.05)

En ce qui concerne le *chunk 1*, les réponses aux deux premières phrases sont diminuées lorsque les chants G312 et G231 sont diffusés, comparativement aux réponses à ces mêmes phrases dans le BOS ($p < 0.005$; **Figure 18**).

Dans le *chunk 2*, les deux premières phrases de ce chunk induisent une activité plus faible lorsqu'elles sont dans G312 que dans le BOS ($p < 0.01$; **Figure 18**). Pourtant, le changement effectué dans l'ordre des *chunks* a eu lieu plusieurs phrases avant ces deux phrases, les probabilités des transitions d'ordre 1 qui précèdent sont les mêmes que celles du BOS.

Par ailleurs, en ce qui concerne le *chunk 3* du chant G312, l'activité à la première phrase du *chunk* est augmentée par rapport à celle obtenue lorsqu'elles sont dans le BOS ($p < 0.005$; **Figure 18**).

Ainsi, changer l'ordre des *chunks* peut avoir un impact sur les réponses jusqu'à deux phrases dans ces *chunks*. En ce qui concerne le changement de la position du *chunk 3*, du BOS à G312, ce changement a un impact sur au moins quatre phrases, deux phrases du *chunk 1* et deux phrases du *chunk 2*, ce qui révèle un effet durable sur l'activité des neurones de ce changement. L'ensemble de ces résultats indiquent donc que changer l'ordre des *chunks* dans le chant peut avoir des conséquences sur les réponses à d'autres phrases que la première, bien que cela n'affecte que les probabilités de convergence et de divergence de la transition qui précède la première phrase des *chunks* dont la position a été modifiée. En d'autres termes, alors que la phrase qui précède est la même et que les probabilités de convergence et de divergence d'ordre 1 sont les mêmes que dans le BOS, une phrase peut induire des réponses différentes, cela dépend de ce qui a été diffusé avant la phrase qui la précède.

En conclusion, conserver l'ordre des phrases au sein des *chunks* ne suffit pas à préserver les réponses aux phrases des *chunks* (**Figure 19**).

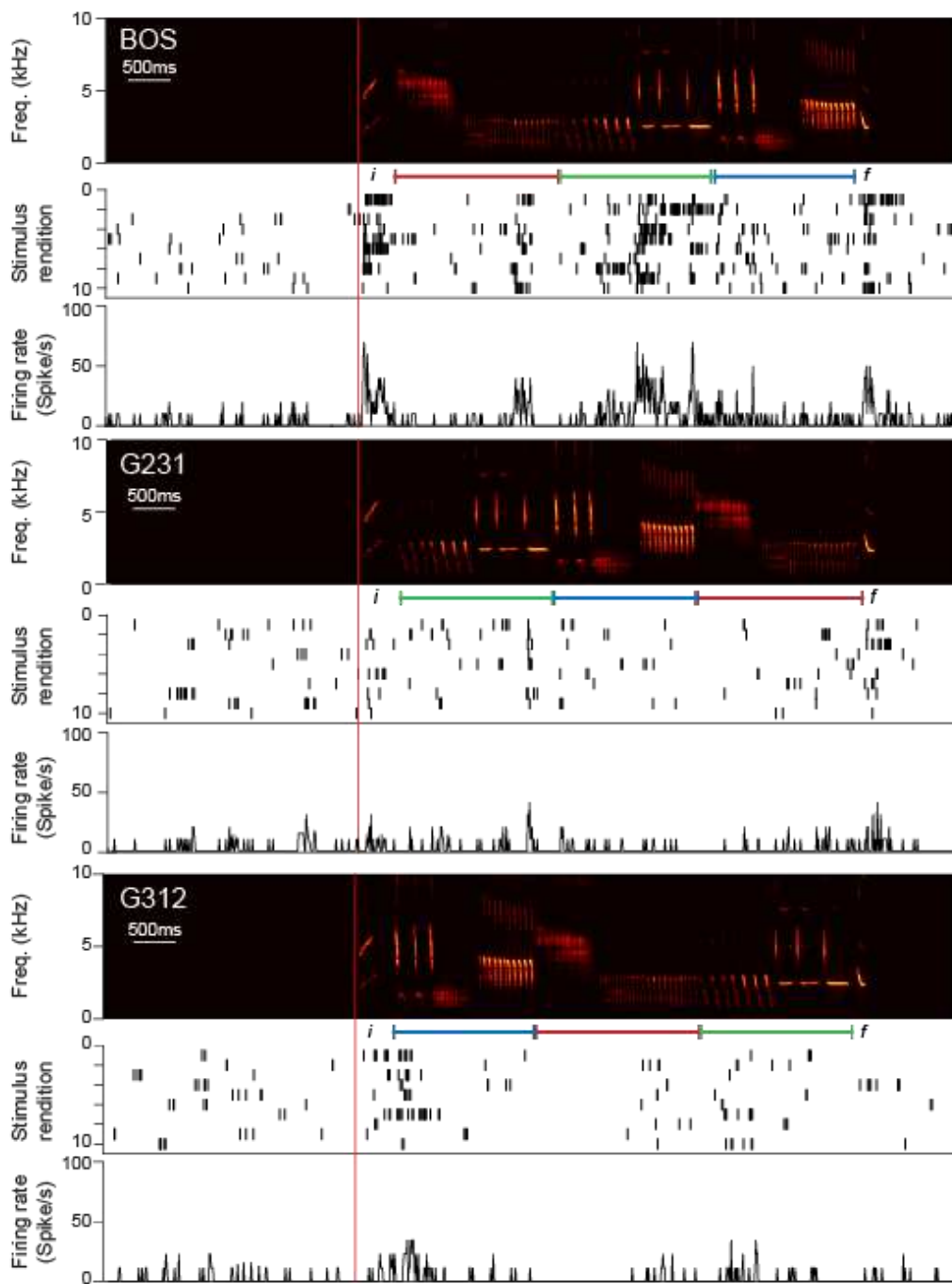


Figure 19 : Exemple de réponses d'un neurone suite à un changement dans l'ordre des chunks.
 Les chunks sont notifiés par des traits de couleurs (chunk 1 en rouge, chunk 2 en vert et chunk 3 en bleu). La première et dernière phrase sont notifiées par, respectivement la lettre *i* et *f*.
 On observe une diminution des réponses pendant les chants G231 et G312 notamment sur les phrases du chunk 1. Les chants sont représentés à travers un spectrogramme.

E. Étude des réponses à la phrase finale

Les résultats précédents ont indiqué qu'en ce qui concerne les phrases des chants L1, L2 et L3 qui n'avaient pas été permutés, on n'observe aucun effet de ces permutations, malgré le changement d'ordre des phrases qui précèdent. Ces phrases n'avaient pas changé de position et le nombre de phrases qui les précédaient n'avait pas changé. En ce qui concerne les chants G231 et G312, on observe un effet des changements effectués sur les réponses à certaines phrases. Mais dans ce cas, les changements ont impliqué à la fois un changement dans l'ordre des phrases qui précèdent, du nombre de phrases qui précèdent et un changement de position dans le chant. Ceci laissait à penser qu'il faut nécessairement changer la place d'une phrase et/ou le nombre de phrases qui précèdent pour modifier les réponses du HVC à son encontre. Or, dans les chants que nous avons présentés, la dernière phrase n'a jamais changé de place, et par là même le nombre de phrases qui la précède n'a jamais changé. Seul, l'ordre des phrases qui précèdent a changé. Cela nous a conduits à focaliser sur les réponses à la dernière phrase du chant, f. Nous avons sélectionné les neurones qui répondaient à cette dernière phrase ($Z_{score} \geq 0$).

Les résultats indiquent un effet du facteur « chant » ($F_{5,690}=6,09$ $p < 0,001$). On observe que permuter deux phrases dans le 2^{ème} chunk (L2 ; qui sont en moyenne séparés par 3,6 phrases de la phrase finale) ou dans le 3^{ème} chunk (L3 ; les deux phrases précèdent la phrase finale) induit une baisse significative de la réponse à la dernière phrase du chant, comparativement à celle obtenue pour le BOS (**Figure 20**). En revanche, changer la position de 2 phrases du 1^{er} chunk (qui sont en moyennes situées à 6,75 phrases), n'affecte pas les réponses à la dernière phrase. Cela indique une sensibilité à l'ordre des phrases qui précèdent, cet ordre peut s'étendre jusqu'à plus de 3 phrases, mais ne va pas jusqu'à plus de 6.

De plus comme observé précédemment, les résultats indiquent que les réponses sont moins importantes pour les chants G231 et G312 que pour le BOS et le chant L1. Nous avons alors calculé les probabilités de convergence et de divergence, de l'ordre 1 à 3 pour la transition précédant la phrase finale pour chacun des chants et chacun des oiseaux. La Figure 20 donne la valeur moyenne sur l'ensemble des oiseaux. Les résultats indiquent que, quel que soit le type de probabilité, convergence ou divergence, et l'ordre, les Z-scores ne varient pas en fonction des probabilités (les 6 valeurs du R^2 varient de 0,1 à 0,11).

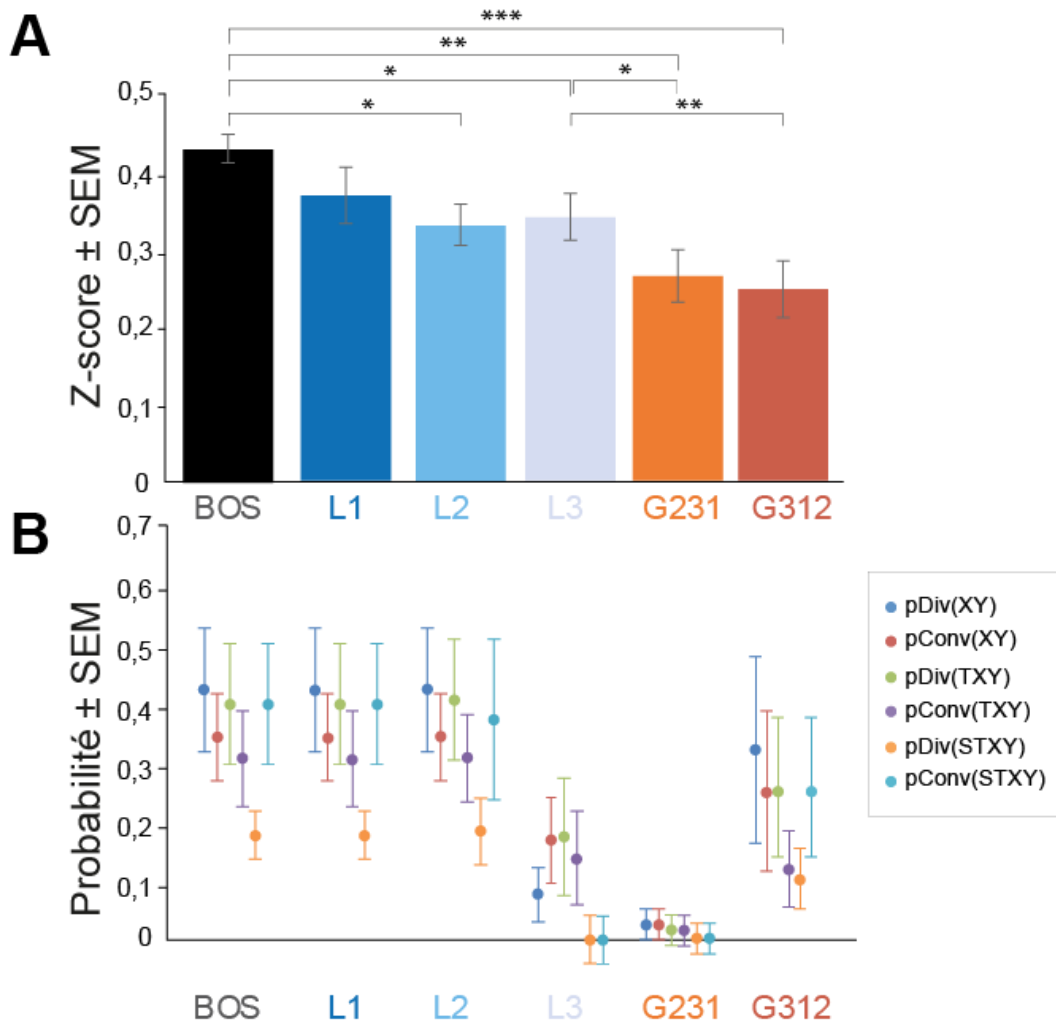


Figure 20: Réponses à la dernière phrase selon le chant diffusé et la force des probabilités de transition.

(Haut) Réponses neuronales à la dernière phrase de chacun des chants. La phrase est la même et n'a pas changé de position.

(Bas) Probabilités de convergence et divergence pour la transition précédant la phrase finale.

(n=208 ; *** équivaut à <math><0.001</math> ; ** <math><0.01</math> ; * <math><0.05</math>)

F. Étude des réponses lors d'une modification du contexte à partir d'une seule phrase

Afin de déterminer dans quelle mesure change le type de phrases qui précèdent influent sur les réponses à une phrase donnée, nous avons pour un oiseau, créé deux chants où une phrase a été déplacée artificiellement dans le BOS (nombre de neurones, $n=66$; **Figure 21**). Dans le premier cas, nous avons déplacé la phrase A de façon à la placer après une phrase qui peut la précéder dans les chants produits par l'oiseau (**Figure 22**). Dans le second cas, nous avons placé la phrase V et placée après une phrase derrière laquelle elle ne se trouve jamais dans les chants, la transition n'est jamais produite par l'oiseau (**Figure 23**).

La phrase A, dans le BOS (chant : ZFCTABPSYVEX) fait répondre 3% des neurones enregistrés et l'activité ne diffère pas du niveau de l'activité spontanée (Z-score moyen = -0,1). De même, lorsqu'elle est permutée avec T dans le chant L1 (chant : ZFCTBAPSYVEX), les mêmes résultats sont obtenus (3% des neurones ; Z-score = -0,1). Si elle est déplacée au début du chant après la phrase Z (chant : ZAFCTBSYVEX), la séquence ZA existe, 18% de la population neuronale répond significativement à cette phrase et on constate une augmentation significative de l'activité lors de cette phrase (Z-score : 0,4). Cette différence de Z-score est significative (chant vs. BOS : $t=6,261$, $p<0.001$). On retrouve exactement les mêmes résultats que dans le chant G231, la phrase A étant au début puisque c'est la première du *chunk* 2. En termes de probabilité, les probabilités de transition ne semblent pas suffire pas à expliquer cette différence de (probabilités : $D(TA)=0,71$, $C(TA)=0,47$; $D(ZA)=0,51$, $C(ZA)=0,36$) puisque les probabilités sont plus importantes pour la transition TA que pour la transition ZA.

Par ailleurs, il est intéressant de constater que les réponses à la phrase B ($n+3$), qui n'est plus précédée de A, sont différentes (dans BOS : Z-score = -0,01 ; dans le chant ZAFCTBSYVEX : Z-score = 0,09 ; chant vs. BOS : test de Wilcoxon $V=1$ $p<0.001$). De même pour les réponses à la phrase T ($n+2$) (dans BOS : Z-score = 1,9 ; dans le chant ZAFCTBSYVEX : Z-score = 1 ; chant vs. BOS : test de Wilcoxon $V=541$ $p<0.001$).

Il faut noter que la séquence TB est présente dans le chant L1, l'activité évoquée par la phrase B (chant : ZFCTBAPSYVEX) est similaire à celle obtenue dans le BOS (dans L1 : Z-score : -0,01). Dans le chant ZAFCTBSYVEX, la phrase B est à la même position que dans le BOS, mais la phrase A ne précède pas immédiatement, mais quatre phrases avant. Ce résultat laisse entendre une intégration du contexte sur le long terme.

Dans le deuxième cas, nous avons déplacé la phrase V en début de chant pour former le chant ZFCVTABPSYEX. La séquence CV n'a jamais été produite par l'oiseau. Dans le BOS, la phrase V s'accompagne d'une augmentation d'activité (Z-score= 0,32 ; 20% des neurones répondent significativement). Dans le nouveau contexte (les probabilités de transition sont à zéro et la place est également nouvelle, la phrase n'apparaissant jamais en début de chant), les neurones montrent une inhibition de leur activité (Z-score=-0,37 ; 6% des neurones ; chant vs. BOS ; test de Wilcoxon $V=510,5$ $p<0.001$). De même, on peut voir que ce changement impacte jusqu'à deux phrases après V (CVTA). Pendant la phrase T (n+1) qui suit, on constate une diminution significative de l'activité contrairement à ce qui est observé quand elle est dans le BOS (chant vs. BOS, Z-score: 0,53 vs. 1,87 ; test Wilcoxon $V=561$ $p<0,001$; nombre de neurones répondant significativement : 18% vs. 52% ; probabilités : $D(VT)=0,39$, $C(VT)=0,39$; $D(CT)=0,95$, $C(CT)=0,47$). Au contraire, on observe une augmentation d'activité pendant la phrase A (n+2) contrairement à ce qui est observé quand elle est dans le BOS (Z-score=-0,1 vs. 0,32; test de Wilcoxon $V=74$ $p<0.0$; 10% vs. 3% des neurones; $D(VTA)=0,6$; $C(VTA)=0,19$).

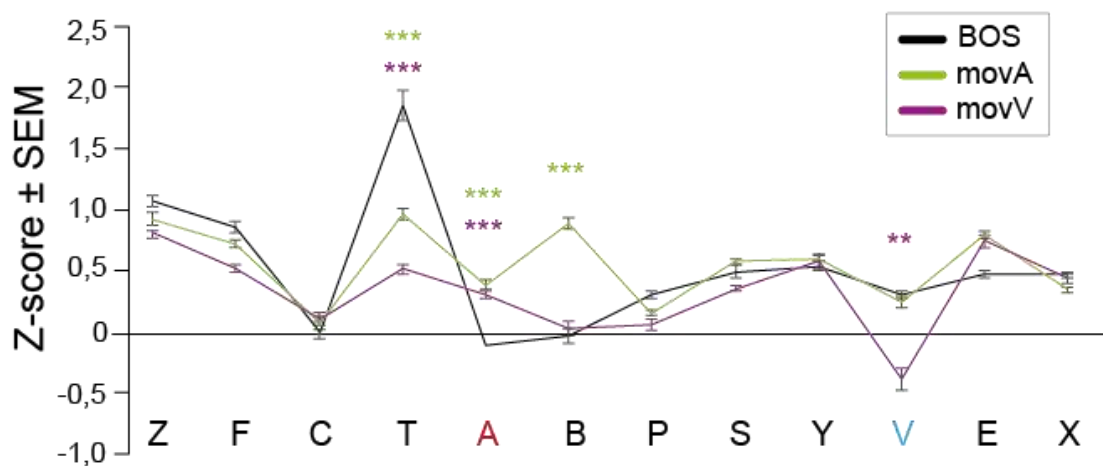


Figure 21: Réponses aux phrases du BOS avant et après déplacement d'une phrase

Réponses neuronales aux phrases du BOS d'un oiseau avant (BOS en noir) et après déplacement de la phrase A en rouge (movA en vert) ou V en bleue (movV en violet). Les phrases sont agencées selon leur ordre dans le BOS.

Agencement movA : Z A F C T B P S Y V E X

Agencement movV : Z F C V T A B P S Y E X

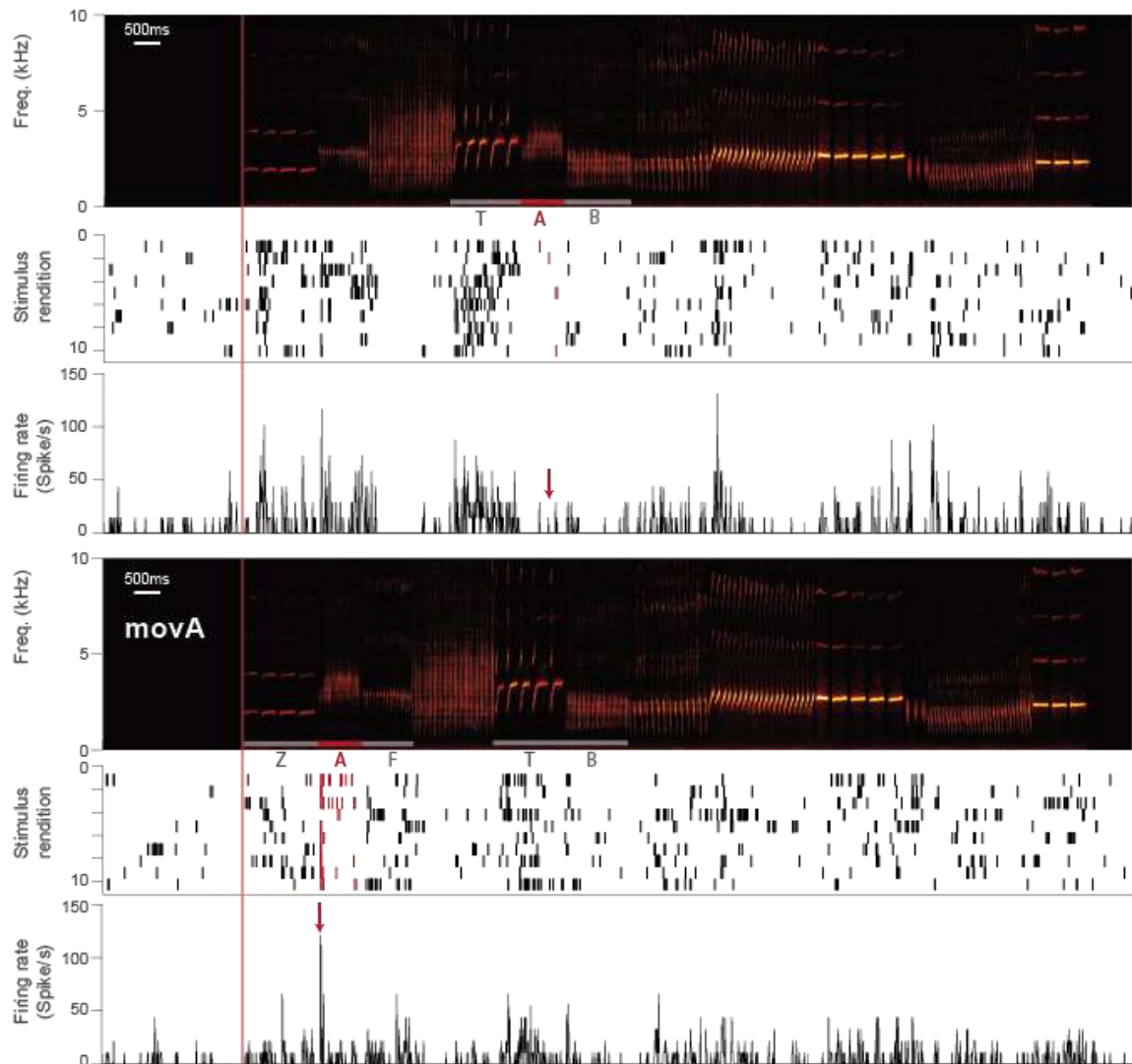


Figure 22: Effet du déplacement de la phrase A sur les des réponses d'un neurone

La phrase A (trait rouge) a été déplacée. Les phrases qui l'entourent sont indiquées par un trait gris.

(Haut) La phrase A dans le BOS induit peu d'activité

(Bas) Une fois déplacée au début du chant, après la phrase Z, elle induit une augmentation d'activité (flèches rouges).

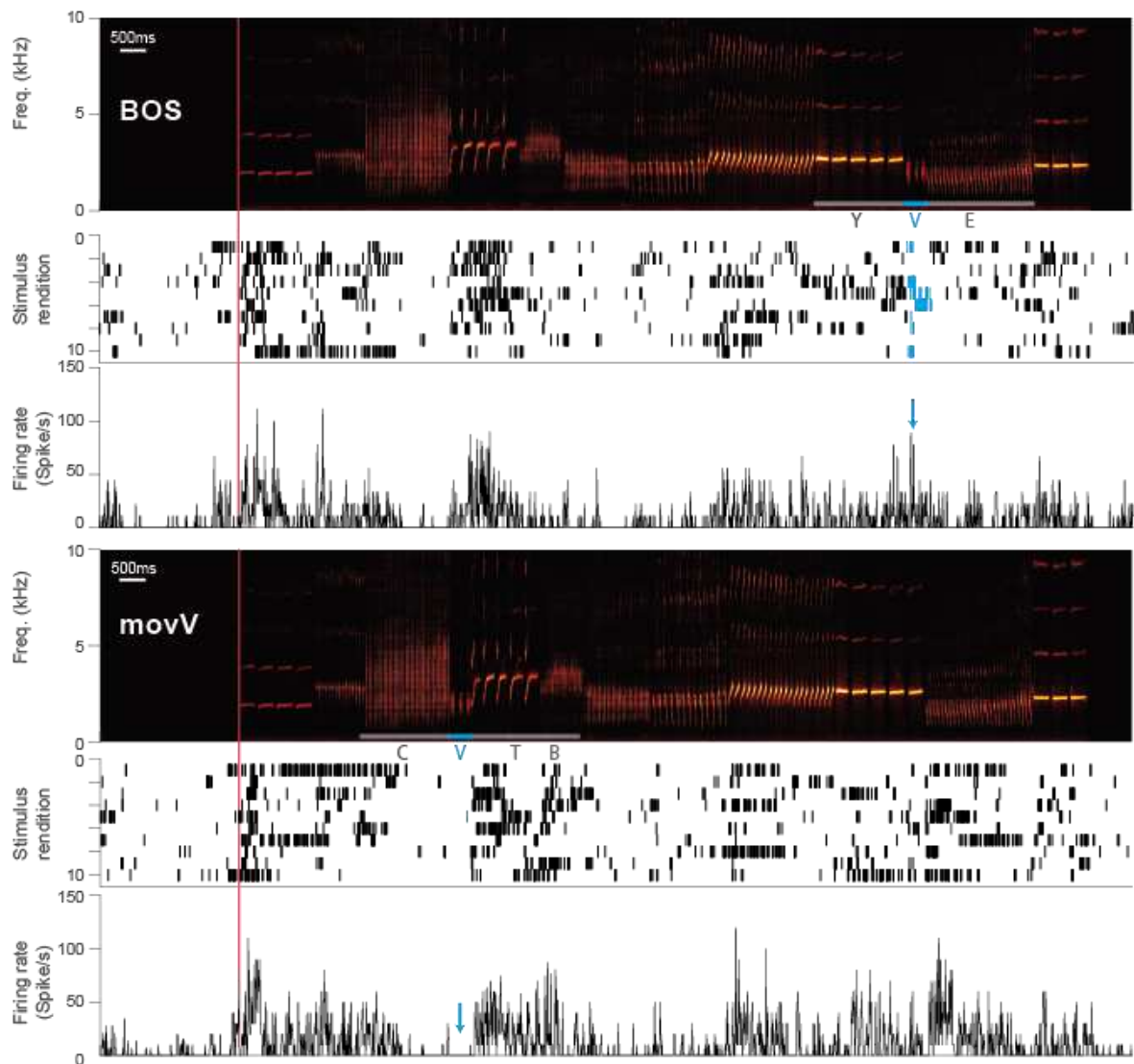


Figure 23 : Effet du déplacement de la phrase V sur les des réponses d'un neurone

La phrase V (trait bleu) a été déplacée. Les phrases qui l'entourent sont indiquées par un trait gris.

(Haut) La phrase V dans le BOS induit des réponses.

(Bas) Une fois déplacée au début du chant, après la phrase C, elle induit une diminution d'activité (flèches bleues).

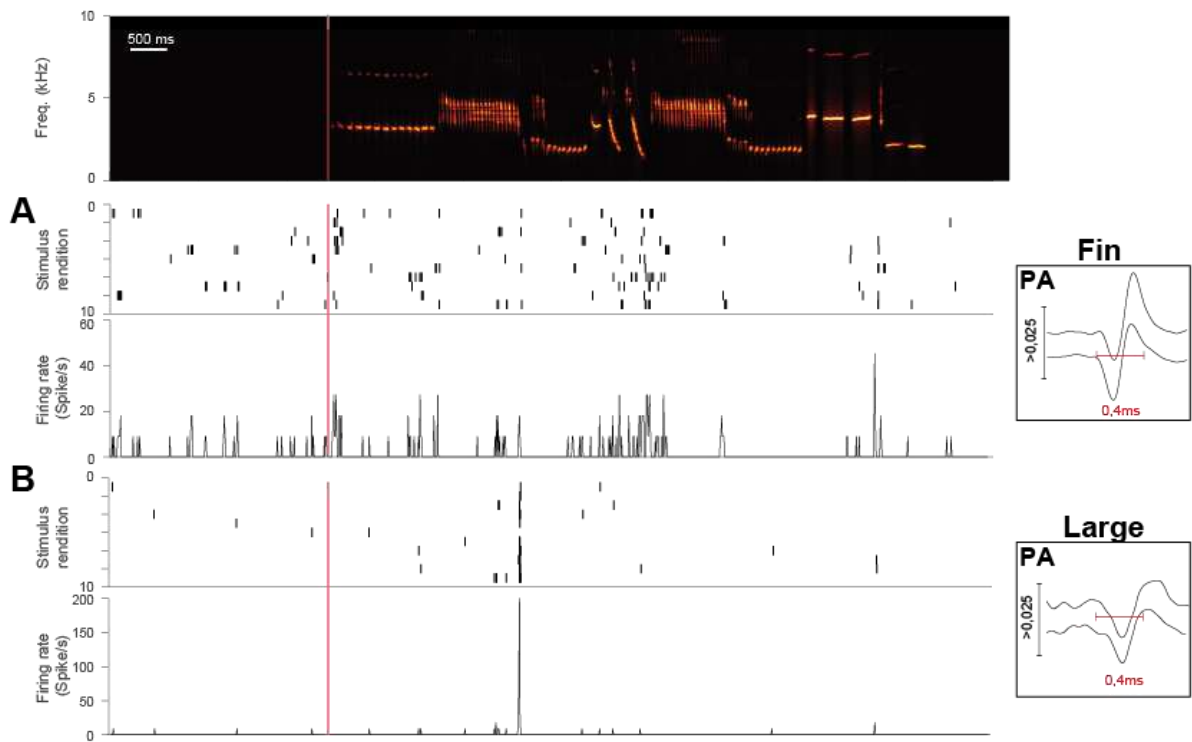


Figure 24 : Exemple de réponses d'un neurone à PA fin et d'un neurone à PA large enregistrés avec la même électrode

Haut : spectrogramme du BOS

Encadré à droite : forme moyenne des potentiels d'action (PA) des deux populations enregistrées

(A) Exemple de réponse d'un neurone à PA fin à un chant.

(B) Exemple de réponse d'un neurone à PA large à un chant.

On observe que ces deux neurones ont un profil d'activité différent.

En conclusion, changer le contexte d'une phrase, c'est-à-dire ce qui précède, peut modifier l'activité pendant plusieurs phrases

G. Réponses en fonction du type de neurones

Le HVC présente deux types de populations neuronales. Ils se distinguent à la fois par leur potentiel d'action (PA) fins ($<0,4\text{ms}$) et larges ($\geq 0,4\text{ms}$), mais aussi par leur taux d'activité. Les neurones à PA fin répondent de manière plus soutenue à plusieurs phrases du BOS tandis que les neurones à PA large ont une activité phasique, centrée sur une ou quelques phrases. La **Figure 56** présente un exemple de réponses d'un neurone à PA fin et d'un neurone à PA large enregistrés sous la même électrode. On observe que ces neurones répondent différemment.

En ce qui concerne le type de phrases du BOS qui font répondre, l'analyse des distributions n'a pas révélé de différence significative entre les deux populations basées sur le type de neurones, sur l'ensemble des oiseaux ($p > 0,05$).

Puis nous avons comparé les réponses aux différents types de chants entre les deux populations. Les résultats statistiques indiquent une interaction significative entre les facteurs « chant », « phrases » et « type de population » pour le chant L1 ($F_{1,7544}=6,38$ $p=0,0018$) et le chant L2 ($p < 0,001$). Des analyses nous indiquent que les neurones à PA fins répondent différemment à une des phrases qui a été permutée, alors que ce n'est pas le cas pour les neurones à PA larges (**Figure 25 ; Figure 26**). Cette différence entre les deux populations ne s'observe plus lorsque l'on considère les réponses au chant L3, ainsi que pour l'ensemble des changements globaux (G231 et G312).

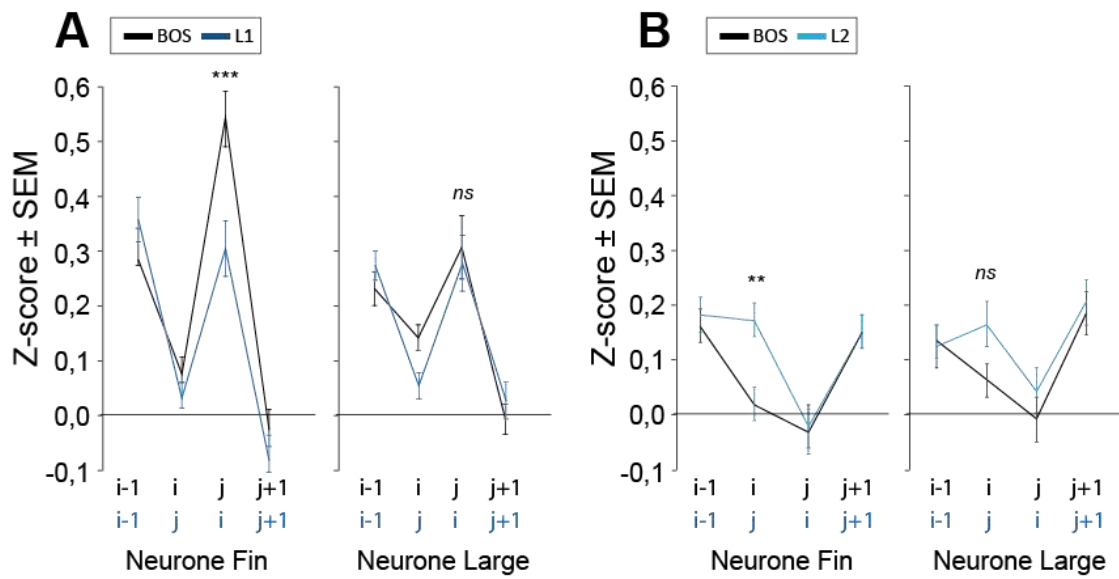


Figure 25 : Effet de la permutation de deux phrases sur les réponses des deux populations neurones, les neurones à PA fins et les neurones à PA larges.

Réponses moyennes aux phrases du BOS (lignes noires) et des chants L1 et L2 (lignes bleues). Quatre phrases ont été considérées : les phrases *i* et *j* qui ont été permuées, la phrase *i-1* qui précède et la phrase *j+1* qui suit. On observe que seules les réponses à la 3^{ème} phrase du BOS, la phrase *j*, ont été impactées par la permutation et, cela seulement pour la population des neurones à PA fins. (*n* fin=201 ; *n* large=132 *** équivaut à <0.001 ; ** <0.01 ; ns>0.05)

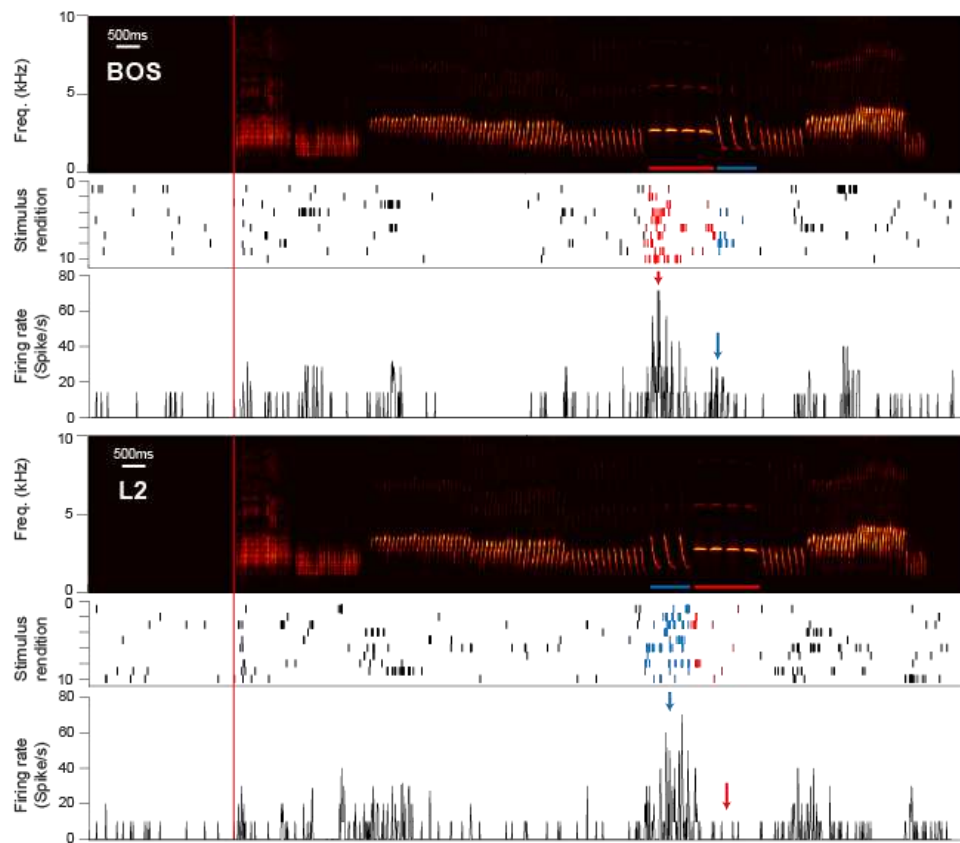


Figure 26: Exemple des réponses d'un neurone à PA fin suite à une permutation dans le chunk

L'activité d'un neurone à PA large ($PA < 0.4$ ms) est représentée par un raster plot et un PSTH (chant BOS : en haut ; chant L1 : en bas). Les deux phrases permutes : traits rouge et bleu. On note un changement dans les réponses si on considère le type de phrases.

Discussion

Comme montré précédemment, les neurones du HVC chez le Canari sont sensibles à l'ordre des phrases dans le BOS : changer l'ordre d'agencement des phrases dans ce chant conduit à des changements dans la force des réponses. Mais, ce que met en évidence cette étude est le fait qu'un changement dans les réponses à une phrase donnée du BOS peut résulter, non pas d'un changement dans le type de phrase qui immédiatement précède, mais au-delà, dans le type de phrases qui se situent bien avant. Ainsi, au niveau du HVC, les neurones présentent une sensibilité au contexte, qui résulte d'une intégration sur une longue durée, c'est à dire sur plus d'une seconde voir plus, des signaux acoustiques.

A. Réponses des neurones et probabilités de transition

Cette étude a été réalisée afin d'approfondir nos connaissances sur ce qui sous-tend la sensibilité des neurones du HVC à l'ordre des phrases. Dans une étude précédente réalisée chez le moineau du Japon (Bouchard and Brainard, 2013), il avait été mis en évidence que l'amplitude des réponses à une phrase donnée est corrélée avec la probabilité de convergence de la séquence qui précède cette réponse : plus la probabilité était forte, plus le neurone était actif. Le canari, tout comme le moineau du Japon, des chants variables pour lesquels il peut être calculé des probabilités de transition entre chaque élément des chants. Nous avons alors émis l'hypothèse selon laquelle l'activité des neurones du HVC à une phrase donnée pourrait dépendre des probabilités d'apparition de la séquence de phrases qui précédait.

La précédente analyse de la structure des chants de canari a mis en avant, au sein des chants, des séquences de phrases au sein desquelles les probabilités de transition, de convergence et de divergence sont toutes les deux fortes. Nous avons alors supposé que, si les neurones du HVC chez le canari étaient sensibles aux probabilités de transition, alors modifier l'ordre des phrases de façon à ce que les probabilités de transition soient égales à zéro, dans ces séquences que nous nommons chunks, devrait avoir un impact conséquent sur les réponses des neurones au chant. En d'autres termes, nous avons voulu savoir s'il suffisait d'altérer les transitions fortes, d'un point de vue probabiliste, entre

deux phrases pour que les neurones montrent un changement important dans leurs réponses. Selon la même perspective, si les neurones étaient sensibles aux probabilités de transition, modifier l'ordre des phrases en conservant les transitions fortes au sein des chunks, par l'intermédiaire d'un changement de l'ordre de ces chunks, devrait avoir un moindre impact sur les réponses que changer l'ordre des phrases au sein des chunks.

Cependant les résultats n'apportent pas d'arguments qui vont dans le sens de ces prédictions. Permuter des phrases au sein d'un chunk, ce qui crée des séquences n'existant pas dans les chants de l'individu (chants L1, L2 et L3), ne s'accompagne pas d'une baisse de l'activité moyenne sur la durée totale du chant, sur la durée totale des chunks et affecte seulement, sur l'ensemble des oiseaux, les réponses à l'une ou aux deux phrases qui ont été permutées. Au contraire de nos prédictions, changer la position d'un chunk, tel que déplacer le chunk 1 après les chunks 2 et 3 (chant G231) ou le chunk 3 avant les chunks 1 et 2 (chant G312), s'accompagne d'une baisse de l'activité moyenne calculée sur la durée totale du chant et sur la durée totale d'au moins un chunk, comparé aux périodes équivalentes lors de la diffusion du BOS. En ce qui concerne les réponses aux différentes phrases, on observe une diminution ou une augmentation pendant la durée de plusieurs phrases, allant jusqu'à quatre lorsque le chunk 3 est passé devant les chunks 1 et 2 (chant G312). Pourtant, les probabilités moyennes calculées sur l'ensemble des chants en prenant compte les séquences de 2, 3 ou 4 phrases ainsi que le nombre de transitions impactées, ne diffère pas d'un chant à l'autre. D'un point de vue du nombre de transitions qui ont été changées, dans tous les cas, trois transitions sont égales à zéro dans les différentes versions manipulées du BOS. Cependant, leur répartition diffère : dans un cas (chants L1, L2 et L3), elles sont regroupées dans un chunk, dans l'autre cas (chants G231 et G312), elles sont réparties à trois endroits différents dans le chant. Ainsi, si on compare aux réponses initiées par le BOS, un chant composé d'une longue séquence de 10 phrases au sein de laquelle trois transitions à la suite sont égales à zéro, a bien moins d'effet qu'un chant avec trois transitions égales à zéro qui sont entre des phrases aux extrémités des chunks suite à leur déplacement.

Des arguments supplémentaires à l'idée que les réponses des neurones ne dépendent pas ou peu des probabilités de transition peuvent être donnés. Parmi les BOS diffusés, un est constitué d'une même séquence de trois phrases répétées deux fois, les probabilités de convergence et de divergence entre les phrases sont fortes et donc identiques entre les deux séquences. Pourtant, un même neurone peut répondre à l'une

des phrases dans la première séquence et à une autre phrase lors de la deuxième fois (**Figure 11**). Par ailleurs, dans l'étude après avoir permuté deux phrases d'un chunk, prenons par exemple la séquence A-B-C-D dans laquelle les deux phrases B et C ont été permutées, constituant la séquence A-C-B-D, la transition avec la phrase qui suit, c'est-à-dire entre B et D, n'existe jamais dans les chants (chez aucun oiseau). Pourtant ce changement n'affecte pas les réponses à la phrase B, mais seulement celles à la phrase C (**Figure 17**). Les réponses de B dans la séquence A-C-B-D sont similaires à celles obtenues dans le BOS A-B-C-D. Enfin, sur la base des réponses aux différentes phrases du BOS, nous n'avons pas mis en évidence de corrélation directe entre les probabilités de divergence ou de convergence dans les séquences de deux phrases (ordre 1) ou de trois phrases (ordre 2) et la force des réponses (**Figure 8**). De plus, notre analyse en composante principale n'a mis en avant aucune corrélation forte entre les probabilités de divergence et de convergence ainsi que pour l'ensemble des paramètres de position et occurrence, avec l'activité neuronale (**Figure 9**).

Ainsi, de nombreux résultats laissent à penser que la force des réponses des neurones du HVC ne dépend pas ou peu des probabilités de transition, ce qui n'est pas en accord avec l'étude de Bouchard et Brainard, 2013. De plus, la seule corrélation mise en avant montre l'opposé de ce qui avait précédemment été suggéré. Pour expliquer cette différence, il peut être indiqué que cette étude a été réalisée chez une autre espèce que le canari, que les stimuli consistaient en de longues chaînes d'un millier de syllabes, ce qui évite la question de la position des syllabes par rapport au début et à la fin du chant. Cette question pourrait être déterminante dans la façon dont répondent les neurones du HVC.

B. Réponses des neurones et position des phrases dans le chant

Si les différences dans les réponses entre le BOS et les versions manipulées de ce chant ne reposent pas ou peu sur les probabilités de transition entre les phrases, la question qui se pose est de savoir sur quelles informations pourraient reposer les traitements effectués par l'HVC (et les régions en amont).

Lorsqu'on manipule l'ordre des éléments dans une séquence, cette manipulation peut conduire à changer la position des éléments. La position est un paramètre à prendre en considération. En effet, parmi les mécanismes qui ont été proposés comme susceptibles de sous-tendre le traitement de l'ordre des éléments dans une séquence, il y

aurait, d'une part, un basé sur les probabilités de transition qui lient un élément au suivant et d'autre part, un mécanisme basé sur la position de chacun des éléments dans la séquence (Endress et al. 2009). Dans cette étude, nous avons manipulé l'ordre des phrases. Nous avons changé seulement la position entre deux phrases dans les chants L1, L2 et L3 par rapport au BOS. Dans ces chants, les phrases qui n'ont pas changé de place n'induisent pas de réponses différentes de celles obtenues lors de la diffusion du BOS. En revanche, dans les chants G312, aucune phrase n'était à sa position par rapport à celle qu'elle avait initialement dans le BOS. De nombreuses phrases du chant G312 induisent des réponses qui diffèrent de celles lorsqu'elles sont dans le BOS (2 phrases du chunk1, 2 phrases du chunk 2 et une phrase du chunk 3). De même dans le chant G231, toutes les phrases avaient changé de position par rapport à celles occupées dans le BOS. Mais, cette fois-ci, moins de phrases montrent des réponses différentes de celles observées lors de leur diffusion dans le BOS (seulement 2 phrases du chunk 1).

Si on considère la position moyenne des phrases dans les chants enregistrés, les analyses présentées dans le chapitre précédent indiquent que la position des phrases n'est pas aléatoire, certaines sont au début, d'autres plutôt à la fin. La **Figure 13** du chapitre II montre que les BOS diffusés sont composés de phrases dont la position moyenne augmente au fur et à mesure que l'on se rapproche de la fin. Les phrases du début du chant et notamment du chunk 1 semblent avoir une position relative plus déterminée que celle des autres phrases. En ce qui concerne la position des chunks dans les chants, les analyses indiquent que le chunk 1 se trouve préférentiellement au début des chants et surtout toujours avant le chunk 2. Le chunk 2, quant à lui, peut également se trouver au début du chant lorsque le chunk 1 n'est pas produit. Or, ce qu'indiquent nos résultats est le fait que le chunk 1 est celui pour lequel l'impact du changement de position est toujours observé. Que ce soit lorsqu'il passe à la troisième position dans le chant G231 ou à la deuxième position dans G312, les neurones répondent moins aux phrases de ce chunk que lorsqu'il est en première position dans le BOS. De même, si on s'intéresse aux autres chants dans lesquels la position des chunks a été modifiée, c'est-à-dire les chants G321 et G213, le chunk 1 induit une activité moins importante que le chunk qui immédiatement précède. En revanche, on n'observe pas cela lorsque le chunk 1 est en première position dans G132, l'activité des neurones pendant ce chant ne diffère pas de celle enregistrée lors de la diffusion du BOS. Ces résultats suggèrent donc que les

neurones du HVC chez le canari pourraient être sensibles à la position des phrases du chunk 1, c'est-à-dire aux phrases du début des chants.

Jusqu'à présent, à notre connaissance, peu d'études d'électrophysiologie réalisées dans le HVC ou une autre région, chez l'oiseau ont cherché à savoir si les neurones pouvaient être sensibles à la position des éléments dans une séquence. Une étude comportementale suggère que les étourneaux sont capables de discriminer entre des séquences composées des mêmes éléments, appelés des motifs, en prenant en compte la position de ces éléments. Ils se baseraient plus particulièrement sur la position absolue des éléments plutôt que sur leur position relative (pour une séquence ABCD, la position absolue de B est d'être à la deuxième place ; la position relative est d'être après A)(Comins and Gentner, 2010). Des études chez des espèces dont le chant n'est pas variable, ont pu déjà mettre en avant le contrôle de la position des syllabes dans l'apprentissage (James and Sakata, 2017). Chez le diamant mandarin, des jeunes entendaient des chants dans lesquels les syllabes pouvaient être à différentes positions. Malgré cela, les analyses des chants montrent que les chants générés suite à ces expositions présentent un certain type de syllabe, en début, milieu et fin de chant. Cette étude suggère donc l'existence de mécanismes sensoriels et moteurs pour lesquels la position des syllabes dans le chant entre en jeu. Ainsi, la possibilité que la position des syllabes ou phrases puisse être encodée au niveau cérébral pourrait être une hypothèse à prendre en considération, d'autant plus dans le HVC, impliqué directement dans le circuit de la production du chant.

C. Réponses des neurones et *chunks*

Le chapitre précédent basé sur l'analyse de la structure des chants de canari nous a conduits à penser qu'un niveau supplémentaire entre celui de la phrase et celui du chant était à prendre en compte : celui des chunks. Étant donné que les résultats indiquent qu'affecter l'organisation des chunks a des effets moins importants que conserver l'organisation des chunks en changeant leur ordre, le traitement du BOS par les neurones du HVC (et les neurones en amont) n'apparaît pas reposer sur une sensibilité aux fortes transitions au sein de chunks. Sinon, des résultats inverses à ceux de notre étude auraient dû être obtenus.

Le fait d'utiliser lors de nos enregistrements électrophysiologiques, un système à 16 électrodes nous a permis d'enregistrer plusieurs dizaines de neurones chez chacun des

individus. En accord avec des études antérieures (Sutter and Margoliash, 1994), certaines phrases (14 sur les 84 phrases qui composaient les BOS) ont évoqué une augmentation d'activité chez plus de 50 % de neurones enregistrés. En moyenne, entre une et deux phrases sur la dizaine de celles qui composent le BOS. Les autres phrases évoquent des réponses d'un nombre plus réduit de neurones, voire seulement de quelques-uns. Un fait intéressant est que, lorsque chez un oiseau, des enregistrements ont été réalisés dans le HVC des hémisphères droit et gauche, ce sont les mêmes phrases qui font répondre la majorité des neurones enregistrés. Un tel constat nous a conduits à nous demander si ces phrases se distinguaient des autres ou si elles étaient à la suite d'une transition particulière. Nous avons alors observé les probabilités de transition pour la phrase et non plus pour une transition donnée. Considérant les probabilités des phrases fortes au sein d'un chunk, et faible avec ce qui l'entoure, cela nous a permis de catégoriser la phrase selon son profil si celle-ci se présente en début, milieu ou fin de chunk. Parmi ces 14 phrases, plus de phrases ont une probabilité de convergence, pour la transition, qui précède, inférieure à 50% et une probabilité de divergence, pour la transition qui suit, supérieure à 50%, que ce qui était attendu si on considère la répartition de l'ensemble total des phrases. Ce type de phrases correspond à celles que l'on peut trouver au début d'un chunk. Ainsi, les seules caractéristiques que nous avons mises en évidence pour ces phrases qui sont à l'origine d'une synchronisation de l'activité dans le HVC, sont relatives au chunk, laissant à penser que la structure en chunk du chant de canari pourrait contribuer, au moins en partie, au codage du BOS dans le HVC.

D. Traitements des informations relatives à l'agencement des phrases par les deux populations neuronales

Le HVC présente deux types de populations neuronales, une constituée de neurones à potentiel d'action (PA) large et une à potentiels d'action fins. Des enregistrements intracellulaires couplés à des marquages des neurones ont indiqué que les neurones à PA larges correspondaient aux neurones de projection excitateurs tandis que les neurones à PA fins correspondraient aux interneurones inhibiteurs. La largeur du potentiel d'action n'est pas la seule caractéristique qui distingue ces deux populations neuronales : les neurones à PA fins ont un taux d'activité plus élevé que les neurones à PA larges et ils répondent de manière soutenue à plusieurs syllabes du BOS alors que les

neurones à PA larges répondent par des augmentations phasiques d'activité à une ou quelques-unes des syllabes (Mooney 2000; Mooney et al. 2001).

Dans la littérature, quelques études basées sur des enregistrements intracellulaires se sont intéressées au traitement de l'ordre en mettant en perspective les enregistrements obtenus sur les deux populations neuronales. Que ce soit les interneurones ou les cellules de projection, les deux types présentent une sélectivité pour le BOS par rapport à ce chant à l'envers ou avec les syllabes placées dans l'ordre inverse. Mais, l'inhibition exercée par les interneurones sur les neurones de projection jouerait un rôle majeur dans la façon dont les neurones de projection répondent sélectivement au BOS par rapport à ce chant diffusé à l'envers ou avec un ordre inversé des syllabes, mais diffusé à l'endroit (Rosen and Mooney, 2006, 2003) ou à la diffusion d'une paire de syllabes AB par rapport à celle de la paire BA (ou de la syllabe A ou de la syllabe B, seule) (Lewicki and Konishi 1995).

Dans notre étude, en termes de traitement des phrases du BOS, on ne note pas de différences majeures dans les réponses des deux types de neurones. Les neurones à PA fins et à PA larges montrent majoritairement une sensibilité pour les mêmes phrases. En ce qui concerne les réponses aux différentes versions manipulées du BOS, on distingue quelques différences entre les deux populations neuronales. Seuls les neurones à PA fins montrent un effet de la permutation des deux phrases dans les chants L1 et L2 (**Figure 25**). En revanche, le changement de place des chunks dans les chants G231 et G312 affecte de la même façon les réponses de ces deux populations. Les neurones à PA fins apparaissent donc plus sensibles aux changements dans l'ordre de syllabes qui sont liées entre elles par des fortes transitions que les neurones à PA larges. Ces données renforcent l'idée que changer localement l'ordre de deux phrases liées entre elles par des transitions fortes a moins d'impact que changer l'ordre des chunks puisqu'ils indiquent que les neurones qui sont potentiellement des neurones de projection répondent aux chants L1 et L2 comme au BOS. Mais, sur la base de ce qui a été mis en évidence en ce qui concerne le rôle des neurones à PA fins, c'est à dire putativement des interneurones inhibiteurs dans la sensibilité des neurones de projection, il est difficile d'interpréter la différence dans la sensibilité des neurones des deux populations de neurones au changement d'ordre des phrases. Ce que nous proposons est l'idée que cette différence de sensibilité soit sous-tendue par l'existence de mécanismes différents d'intégration des informations provenant des régions auditives en amont.

E. Intégration à long-terme du type de phrases qui précèdent

Indépendamment de la question de déterminer si les réponses des neurones du HVC dépendent des probabilités de transition entre les phrases ou de leur position dans le chant, les résultats montrent que les réponses à une phrase peuvent dépendre du type de phrases qui a précédé, en ne limitant pas cela à la phrase qui a immédiatement précédé, mais au contraire en prenant en compte les autres phrases qui ont précédé. Cette idée repose sur plusieurs arguments. Tout d'abord, lorsqu'un BOS est composé d'une même séquence de deux phrases, il a été observé qu'un neurone répondait à la deuxième phrase dans une séquence et n'y répondait pas dans l'autre suggérant que les réponses dépendent de ce qui a précédé cette séquence. Ensuite, dans les chants G312, les neurones ne répondent pas aux phrases du chunk 2 comme ils y répondent lorsqu'elles sont dans le BOS. Pourtant ces phrases du chunk 2 sont précédées des phrases du chunk 1 dans les deux cas. Un autre exemple vient de l'analyse des réponses à la présentation de la phrase finale. Alors qu'elle n'a pas bougé de position, quel que soit le type de chant qui a été diffusé, les réponses obtenues dépendent des phrases qui précèdent. Alors que le changement effectué a eu lieu au moins 3 phrases avant, comme c'est le cas dans le chant L2, les neurones peuvent répondre moins à cette phrase. Les phrases étant en moyenne d'une durée de plusieurs centaines de millisecondes, cela suggère que les neurones intègrent les informations sur une longue période.

Enfin, nous avons, pour un oiseau, créé deux nouveaux chants où seule une phrase a été déplacée. Au contraire d'un chant local, il ne s'agit pas d'une permutation, mais d'un changement de contexte pour la phrase déplacée. Dans un premier cas, nous avons placé une phrase dans un contexte qui a déjà été produit par l'oiseau tandis que le deuxième nous avons placé une phrase dans un contexte jamais connu par l'oiseau. Nous avons pu constater que le contexte impacte sur les réponses à la phrase que le contexte soit connu ou non de l'oiseau (**Figure 21**). De plus, nous avons pu constater que les réponses ne sont pas seulement différentes pour la phrase cible, mais aussi pour les deux phrases qui suivent, ce qui sous-tend l'idée que la permutation ne suffit pas à changer suffisamment le contexte de la phrase pour induire une répercussion sur les réponses pour les phrases suivantes.

De tels résultats indiquent que, chez le canari, les neurones de l'HVC peuvent intégrer sur une longue période les informations relatives aux phrases qui précèdent.

Jusque-là, cela avait été mis en évidence chez le diamant mandarin, il avait été, en effet, montré que les réponses à une syllabe pouvaient dépendre des types de syllabes qui précèdent indiquant une intégration sur plusieurs centaines de millisecondes (Margoliash and Fortune, 1992). Chez le canari, ce temps d'intégration pourrait donc être plus long et prendre en compte les informations acoustiques sur plusieurs secondes.

Une étude récente a mis en avant cette capacité à intégrer les informations sur une longue période pour les neurones du HVC chez le canari. Mais, contrairement à notre étude, les mesures de l'activité neuronale en faisant appel à une technique d'imagerie calcique ont été réalisées lorsque l'oiseau chantait (Cohen et al., 2019). Ainsi, que ce soit sur le versant perceptif ou le versant moteur, l'intégration des informations sur une très longue période, une période de plusieurs secondes, pourrait être une des propriétés des neurones du HVC chez le canari.

F. Conclusion

En conclusion, l'ensemble des résultats de cette étude indique que les réponses à une phrase se définissent par bien plus que par sa seule identité acoustique, c'est-à-dire les paramètres acoustiques que la caractérise. Ils apportent des informations quant au traitement du HVC en rapport avec les règles qui sous-tendent l'agencement des phrases dans le chant de canari. Ils ne mettent pas en évidence que les réponses dépendent des probabilités de transition et que le traitement du chant est centré sur les chunks. Ils mettent en avant le fait qu'un changement dans les réponses à une phrase donnée du BOS peut résulter d'un changement dans le type des phrases qui se situent bien avant sa diffusion. Ils révèlent ainsi une intégration sur une longue période du contexte dans lequel se situent les phrases. Ils montrent particulièrement cela pour les phrases du premier chunk du BOS. Bien que les réponses ne diffèrent pas entre les trois chunk quand ils sont dans le BOS, leur traitement semble être néanmoins différent. Changer de position le premier chunk, ce qui équivaut également à changer le contexte dans lequel se trouvent les phrases de ce chunk, a un fort impact sur les réponses aux phrases de ce chunk ainsi qu'à celles qui suivent. Les résultats suggèrent donc que la représentation du BOS au niveau du HVC reposerait tout particulièrement sur le type de phrases qui débute fréquemment les chants.

DISCUSSION GÉNÉRALE

Discussion générale

L'objectif de cette thèse a été d'étudier comment sont encodés les chants en tant que séquences de syllabes et de phrases dans le cerveau des oiseaux chanteurs. Pour cela, nous nous sommes intéressés à deux aires cérébrales, une aire auditive analogue du cortex auditif secondaire, le NCM, et une aire sensori-motrice, le HVC. De plus, nous avons pris le parti d'aborder notre question principale chez deux espèces d'oiseaux chanteurs, le diamant mandarin et le canari.

Notre premier but a été de déterminer si l'ordre des syllabes dans le chant constitue un paramètre auquel les neurones du NCM répondent lors de la diffusion des chants. Cette question a été abordée chez le diamant mandarin. Pour aller plus loin et avancer dans la compréhension des mécanismes mis en jeu, étant donné que le HVC est une aire dont les neurones sont sensibles à un changement de l'ordre des syllabes dans le BOS, nous avons cherché à savoir sur quoi pouvait reposer le fait que les neurones du HVC répondent moins lorsque l'ordre des syllabes est modifié. Cette question a été le point de départ de l'étude réalisée chez le canari.

Aspects méthodologiques

La question qui peut se poser est pourquoi avoir étudié les réponses des neurones du NCM chez le diamant mandarin et celles des neurones du HVC chez le canari. Plusieurs raisons ont conduit à ce choix.

Tout d'abord, le diamant mandarin et le canari se différencient par de nombreux aspects, parmi lesquels figure tout particulièrement la plasticité de leur chant (Brenowitz and Beecher 2005). Chez le diamant mandarin, une fois que l'adulte produit un chant, celui-ci n'est pas soumis à des variations saisonnières (Dunn and Zann, 1996). Chez le canari, au contraire, le chant change d'une saison à l'autre : le répertoire se modifie à l'automne et à cette saison, les chants sont beaucoup moins stéréotypés d'un point de vue temporel (durée du chant, des phrases et des silences entre les phrases). Cela est dû à des changements hormonaux qui par l'intermédiaire de récepteurs au niveau cérébral conduisent à des réorganisations neuroanatomiques (Nottebohm et al. 1986; Kirn et al. 1989), mais aussi des changements dans les réponses des neurones lors de la diffusion de

chant (Remage-Healey et al. 2010; De Groof et al. 2013; Del Negro and Edeline 2002; Meitzen et al. 2007; Meitzen et al. 2009). Ainsi, au sein du HVC, moins de neurones montrent un changement dans leur activité lors de la diffusion du BOS ainsi que moins de neurones répondent sélectivement au BOS lorsque les canaris sont dans les conditions automnales (Del Negro, 2005). Tout cela nous a conduits à étudier, chez le canari, seulement les réponses des neurones du HVC lors de la diffusion de chants et uniquement au printemps. L'étude des réponses auditives au sein du NCM chez le canari aurait nécessité d'être également réalisée au printemps.

Une autre raison vient du fait que peu d'études suggéraient que les neurones du NCM pouvaient être sensibles à l'ordre des syllabes dans les chants notamment conspécifiques. Or, la plupart des études qui conduisaient à cette hypothèse avaient été réalisées chez le diamant mandarin (Schneider and Woolley 2013; Ono et al. 2016; Lu and Vicario 2014). Les diamants mandarins sont un modèle d'étude très utilisé dans le domaine des neurosciences s'intéressant au chant chez les oiseaux. Ils produisent un seul chant, propre à l'oiseau, au sein duquel l'ordre des syllabes est fixe. Ils le répètent un nombre variable de fois tout au long de la journée et tout au long de leur vie. Sur cette base, nous avons émis l'idée que ces multiples répétitions permettraient aux diamants mandarins de mémoriser le chant des congénères, non seulement sur la base des caractéristiques acoustiques des syllabes, mais aussi de l'agencement des syllabes. Comme le NCM était considéré comme impliqué dans la mémorisation des chants, cela nous a conduits à penser que les réponses des neurones du NCM pourraient dépendre de l'ordre des syllabes dans le chant. En revanche, chez le canari, aucun enregistrement électrophysiologique n'a été, jusqu'à présent, réalisé dans le NCM. Rien ne permettait de savoir si, les neurones du NCM chez le canari, comme chez le diamant mandarin, répondaient aux chants d'autres individus identiquement au BOS. Étant donné cela, réaliser chez le canari la même étude que celle qui a porté sur l'étude des réponses auditives des neurones du NCM chez le diamant mandarin aurait nécessité des analyses préalables (évaluation des propriétés d'habituation des neurones par exemple).

Il peut être également ajouté que des données comportementales montrent que les diamants mandarins sont capables de discriminer entre des chants qui varient de par l'ordre d'agencement des syllabes (Gentner and Hulse, 2000; Ratcliffe and Weisman, 1986). Chez le canari, à notre connaissance, aucune étude n'a montré cela. En conséquence, si les résultats avaient montré une insensibilité des neurones du NCM à

l'ordre d'agencement des phrases, nous n'aurions pas pu exclure la possibilité que ce soit lié à une potentielle incapacité des canaris à discriminer des chants de différents individus sur la seule base de l'organisation des phrases dans les chants. De plus, étant donné la complexité de l'organisation des phrases des chants de canaris, nous n'excluons pas la possibilité d'un traitement plus complexe des neurones du NCM de canari des neurones du NCM des diamants mandarins.

Si plusieurs raisons nous ont conduits à nous intéresser aux neurones du NCM chez le diamant mandarin, alors on peut alors se demander pourquoi avoir soulevé la question des mécanismes qui sous-tendent la sensibilité à l'ordre des syllabes au sein du HVC chez le canari. Une des principales raisons était qu'une étude précédente de l'équipe avait montré que les neurones du HVC de canaris répondent moins lorsque les chants créés à partir d'un BOS n'ont aucune des transitions entre les phrases que l'oiseau est capable de chanter (Lehongre and Del Negro 2011). Des études antérieures s'étaient déjà intéressées à la sensibilité à l'ordre des syllabes par les neurones du HVC chez le diamant mandarin (Margoliash and Fortune, 1992). Plus récemment, des études ont été réalisées chez le moineau du Japon, qui comme le canari et contrairement au diamant mandarin a des chants variables (Bouchard and Brainard 2013; Fujimoto et al 2011). Soulever cette question chez le canari s'inscrivait dans ce contexte avec l'idée supplémentaire de prendre en compte un aspect qui n'avait pas été considéré chez le moineau du Japon : le fait que certaines phrases et certaines séquences reviennent très fréquemment et pour certaines préférentiellement à une position donnée dans le chant.

Contribution des études à la littérature

De nombreuses études portant sur le NCM ont mis en évidence que le NCM est impliqué dans le traitement des chants de l'espèce (Chew et al. 1995; Mello et al. 1992; Stripling et al. 1997; Stripling et al. 2001). Lors de la diffusion de chants conspécifiques, les réponses neuronales sont plus fortes que celles obtenues lors de la diffusion d'autres stimuli acoustiques tels que des chants hétérospécifiques. À partir de là, un intérêt s'est porté sur la structure spectro-temporelle des chants pour déterminer à quoi sont sensibles les neurones d'un point de vue acoustique (Ribeiro et al. 1998; Terleph et al. 2006; Van Ruijssevelt et al. 2017), laissant de côté l'hypothèse selon laquelle l'ordre des syllabes en tant que tel pourrait influencer sur les réponses. Une étude a pourtant montré que l'activation des neurones du NCM lors de la diffusion de syllabes composées de deux notes qui se suivent, ne peut pas s'expliquer par l'activation des neurones induite par ces deux notes séparément (Ribeiro et al. 1998). De même, la réponse à une séquence de syllabes n'équivaut pas à la somme des réponses à chacune des syllabes de la séquence, diffusées indépendamment les unes des autres. Ceci laisse à penser que les neurones du NCM peuvent intégrer les informations acoustiques en prenant en compte celles qui précèdent. Cette intégration est non linéaire, comme celle des neurones du HVC. En d'autres termes, les neurones du NCM pourraient être sensibles au contexte acoustique dans lequel une syllabe est présentée. C'est cette hypothèse qui nous a conduits à apporter des arguments en faveur de l'idée que les réponses des neurones du NCM à un chant ne peuvent pas seulement s'expliquer par les réponses aux différentes syllabes de ce chant. Après avoir répété cinquante fois un chant et observé une baisse des réponses due à un phénomène d'habituation, changer l'ordre des syllabes dans le chant réinstaurait de fortes réponses. Si les neurones du NCM n'avaient été sensibles qu'à certains paramètres acoustiques ou qu'à certaines combinaisons de ces paramètres que l'on peut trouver dans les syllabes, changer l'ordre des syllabes n'aurait pas dû réinstaurer les réponses. Ce que l'on a aussi observé est le fait que des réponses à certaines phrases avaient totalement disparu alors qu'elles étaient encore présentes, si l'ordre des syllabes restait le même. Ces résultats indiquent donc que les réponses à une syllabe dans un chant ne peuvent pas être prédites à partir des réponses à cette syllabe dans ce même chant, mais avec un ordre des syllabes différent.

Au-delà de ces considérations, le NCM est une aire auditive considérée comme analogue du cortex auditif secondaire des mammifères. À partir de là, la question qui se pose est de savoir si une sensibilité à l'ordre des sons est une des caractéristiques du cortex auditif secondaire. Mais, peu d'études se sont jusqu'à présent intéressées à cette question chez les mammifères. Une de ces études montre que les réponses à un son donné dépendent du type de séquence dans lequel il se trouve : selon que les fréquences augmentent d'un son au suivant dans la séquence, qu'ils diminuent ou qu'ils soient mélangés, les réponses à ce son diffèrent (McKenna et al. 1989). Pour des séquences constituées de plusieurs sons identiques, la réponse ne dépend pas de la position, mais de son contexte, autrement dit, les sons qui précèdent (McKenna et al. 1989). Dans une autre étude où deux sons étaient présentés dans des séquences dont seul le premier son variait, il a été montré que les réponses des neurones au deuxième son dépendent du premier (Brosch and Scheich, 2008). Bien que dans ces études, les séquences ne soient pas comme les chants constitués d'une série de vocalisations complexes séparées par de très courts silences, mais des suites de sons purs séparés par des silences de plusieurs centaines de millisecondes, ces données permettent de supposer des caractéristiques communes entre le NCM des oiseaux chanteurs et le cortex auditif secondaire des mammifères (AII). Il peut être ajouté le fait que le phénomène d'habituation et celui d'adaptation spécifique à un stimulus (un stimulus est présenté de façon répétée alors qu'un deuxième stimulus est présenté plus rarement) s'observent dans le cortex auditif secondaire comme dans le NCM (Nieto-Diego and Malmierca, 2016).

Au niveau du HVC, des études antérieures avaient déjà montré que les neurones du HVC intègrent les informations spectrotemporelles sur plusieurs centaines de millisecondes. Notre étude basée sur l'analyse des réponses des neurones du HVC après une manipulation de l'ordre des phrases du BOS nous permet de montrer que cette intégration peut s'étendre sur une plus longue période. Les réponses à la dernière phrase du BOS en fonction de ce qui précède sont un exemple qui illustre cette idée. Les BOS qui ont été diffusés étaient composés de la succession de trois *chunks*, comprenant 2 ou 3 phrases chacun. Permuter deux phrases dans le 2^{ème} *chunk*, c'est-à-dire au moins trois phrases avant la phrase finale induit une baisse significative de la réponse à la dernière phrase du chant, comparativement à celle obtenue pour le BOS.

Notre étude des réponses des neurones du HVC avait pour but d'examiner sur quoi reposaient les réponses à une phrase donnée. Une étude précédente suggère que les réponses à une syllabe donnée dépendent des probabilités de convergence entre ladite syllabe et les séquences de phrases qui la précèdent (Bouchard and Brainard, 2013). Nos résultats n'apportent pas d'arguments qui vont dans ce sens. Ils mettent plutôt en avant une sensibilité à la position des phrases qui constituent le premier *chunk*. Changer de position le *chunk* 1 s'accompagne d'une baisse des réponses aux phrases de ce *chunk*, baisse qui peut s'étendre aux phrases du *chunk* 2 lorsque le chant présenté est le chant G312. Jusqu'à présent, aucune étude basée sur des enregistrements électrophysiologiques lors de la diffusion de chants n'avait montré que les réponses des neurones pouvaient dépendre de la position des syllabes.

Encodage de l'ordre des syllabes dans le chant au sein du NCM et du HVC.

Au sujet de l'hypothèse d'un encodage des chants en tant que *chunk* ou série de *chunks*, les résultats de nos études montrent que l'ordre des syllabes ou phrases influe les réponses des neurones, que ce soit dans le NCM ou dans le HVC. Un chant apparaît donc encodé par les syllabes ou phrases qui le composent, mais aussi par l'agencement de ces dernières. La rupture d'habituation des neurones du NCM observée lors du changement de l'ordre des syllabes suggère que le chant présenté de multiples fois a été encodé comme un tout ou, en d'autres termes, comme un *chunk*. De la même manière, l'étude réalisée dans le HVC avait pour but de savoir si le traitement du BOS repose sur la prise en compte d'un niveau intermédiaire entre celui des phrases et celui des chants, celui des *chunks*. Nous avons fait l'hypothèse que changer les *chunks* de place aurait moins d'impact sur les réponses des neurones que changer l'ordre dans ces *chunks* et que chaque *chunk* est traité quelque peu indépendamment du précédent ou du suivant.

Dans le NCM, afin de savoir si les séquences peuvent être traitées comme un *chunk*, nous avons répété la présentation de séquences courtes constituées selon deux schémas, ABAB ou AABB, suivies de la présentation de ces mêmes séquences, mais au sein desquelles les deux syllabes centrales ont été permutées. Ainsi, AABB est devenu ABAB et ABAB est devenu AABB. Sur la base d'une telle hypothèse, nous aurions dû avoir une

augmentation des réponses dans les deux cas, si ces séquences formaient un *chunk* (comme présenté sur la Figure 11 de l'article publié dans J. Neurosci.). Mais, nos résultats n'ont pas montré cela. Seul le passage de AABB à ABAB s'est accompagné d'une augmentation des réponses, due à une augmentation de réponses au premier B qui passait de la 3^{ème} à la 2^{ème} place. Ils suggèrent donc plutôt une détection d'un changement local, c'est-à-dire du passage de AA (dans AABB) à AB (dans ABAB), que d'un changement de la structure global de la séquence, c'est-à-dire du passage d'AABB à ABAB. En conséquence, les résultats suggèrent plutôt un encodage des transitions, particulièrement la première (AA dans AABB et AB dans ABAB) qu'un encodage du chant dans sa globalité.

Dans le HVC, nos résultats remettent en cause l'idée que les chants soient considérés comme une série de *chunks*, indépendants les uns des autres. Si cela était le cas, nous aurions dû observer que changer l'ordre des phrases au sein des *chunks* avait un impact plus important que changer les *chunks* de place. Mais nos résultats suggèrent le contraire. Le fait de permuter des phrases qui forment des transitions fortes a un impact sur les réponses à une de ces deux phrases, mais a peu d'impact si on considère l'ensemble des phrases du chant ou les changements dans les réponses qui sont observés quand l'ordre des *chunks* est modifié.

Questions soulevées par nos études.

Nos études soulèvent un certain nombre de questions.

En ce qui concerne les neurones du NCM, nos données montrent que les réponses à une syllabe varient en fonction de ce qui précède, mais aucune donnée ne nous permet de savoir si cela inclut juste la fin de la syllabe précédente, de l'ensemble de la syllabe qui précède ou encore au-delà.

En ce qui concerne les neurones de l'HVC chez le canari, on peut se demander quelle est l'activité de ces neurones quand l'oiseau chante. Une seule étude a été réalisée. Elle a fait appel à une technique d'imagerie calcique et a porté sur un seul type de neurones, ceux qui projettent à l'aire X (Cohen et al., 2019). D'un point de vue électrophysiologique, il reste toujours à déterminer quelle est l'activité des différents types de neurones lorsque le canari chante, si une forte activité s'observe

préférentiellement lorsque l'oiseau produit les phrases qui font répondre plus de 50% des neurones lorsqu'elles sont diffusées.

Il serait intéressant d'effectuer les études que nous avons réalisées chez la même espèce afin de comprendre la hiérarchie du traitement du contexte selon les différentes aires.

La mise en évidence d'un effet de l'ordre des éléments du chant sur les réponses neuronales a reposé sur un changement de cet ordre à partir d'un stimulus familier, devenu familier par la procédure d'habituation en ce qui concerne le NCM et résultant d'un apprentissage pour le propre chant de l'oiseau, en ce qui concerne le HVC. La question que l'on peut se poser alors est de savoir si présenter un chant inconnu et ce même chant dans lequel l'ordre a été changé, conduirait aux mêmes résultats que ceux que nous avons obtenus.

Enfin, notre étude pose la question du corrélat comportemental de cette propriété neuronale de sensibilité à l'ordre des syllabes ou phrases dans le chant chez le canari. On peut émettre l'hypothèse que celle-ci contribue à la reconnaissance interindividuelle et donc au maintien des relations sociales.

Conclusion

L'ensemble des résultats apportent de nouveaux arguments en faveur de l'idée selon laquelle les oiseaux peuvent être considérés comme un modèle d'étude des bases neurobiologiques du langage.

Le fait que le chant du Canari soit structuré selon différents niveaux d'organisation (syllabe, phrase, chunk, chant), que la position des phrases ne soit pas aléatoire ni fixe, que les phrases s'enchaînent selon des probabilités de transition qui révèlent des dépendances à long terme, indique que les chants peuvent posséder bien plus de points communs avec le langage humain que celui des espèces d'oiseaux chanteurs qui sont communément utilisés pour étudier les bases neurobiologiques du chant et qui sont souvent pris en compte pour mettre en avant les limites du chant des oiseaux comme modèle d'étude (Berwick et al., 2011).

Le fait que, chez une espèce d'oiseau chanteur, les neurones d'une aire auditive analogue à une aire corticale auditive chez les mammifères soient sensibles à l'ordre séquentiel des éléments diffusés, comme cela a été montré chez le primate non humain (Kikuchi et al., 2017) et chez l'humain (voir la revue, Dehaene et al. 2015) apporte un parallèle supplémentaire entre les oiseaux et les primates.

Enfin, une sensibilité au contexte pouvant s'étendre sur plusieurs phrases, c'est-à-dire plusieurs secondes, comme le montrent les neurones du HVC, permet d'envisager comment l'ordre des syllabes ou des mots du langage humain pourrait être traité au niveau neuronal.

Nos études apportent donc des arguments supplémentaires à l'idée que l'étude des propriétés des neurones chez les oiseaux chanteurs peut permettre de progresser dans la compréhension des mécanismes neuronaux sous-tendant le langage humain.

RÉFÉRENCES BIBLIOGRAPHIQUES

Références bibliographiques

- Alliende, J., Giret, N., Pidoux, L., Del Negro, C., Leblois, A., 2017. Seasonal plasticity of song behavior relies on motor and syntactic variability induced by a basal ganglia-forebrain circuit. *Neuroscience* 359, 49–68.
<https://doi.org/10.1016/j.neuroscience.2017.07.007>
- Alliende, J., Lehongre, K., Del Negro, C., 2013. A species-specific view of song representation in a sensorimotor nucleus. *J. Physiol.-Paris* 107, 193–202.
<https://doi.org/10.1016/j.jphysparis.2012.08.004>
- Amador, A., Perl, Y.S., Mindlin, G., Margoliash, D., 2013. Elemental gesture dynamics are encoded by song premotor cortical neurons. *Nature* 495, 59–64.
<https://doi.org/10.1038/nature11967>
- Amin, N., Grace, J.A., Theunissen, F.E., 2004. Neural response to bird's own song and tutor song in the zebra finch field L and caudal mesopallium. *J. Comp. Physiol. [A]* 190, 469–489. <https://doi.org/10.1007/s00359-004-0511-x>
- Arnold, K., Zuberbühler, K., 2006. Semantic combinations in primate calls: Language evolution. *Nature* 441, 303–303. <https://doi.org/10.1038/441303a>
- Aubin, T., Jouventin, P., 2002. Localisation features in penguin calls 205, 3793–3798.
- Bailey, D.J., Rosebush, J.C., Wade, J., 2002. The hippocampus and caudomedial neostriatum show selective responsiveness to conspecific song in the female zebra finch. *J. Neurobiol.* 52, 43–51. <https://doi.org/10.1002/neu.10070>
- Beckers, G.J.L., Gahr, M., 2010. Neural Processing of Short-Term Recurrence in Songbird Vocal Communication. *PLoS ONE* 5, e11129.
<https://doi.org/10.1371/journal.pone.0011129>
- Beecher, M.D., 1994. Song perception in the song sparrow: birds classify by song type but not by singer. *Anim. Behav.* 47, 1343–1351.
- Beecher, M.D., 1991. *Successes and failures of parent-offspring recognition*. Cambridge University Press.
- Bernal, S., Dehaene-Lambertz, G., Millotte, S., Christophe, A., 2010. Two-year-olds compute syntactic structure on-line: Two-year-olds compute syntactic structure on-line. *Dev. Sci.* 13, 69–76. <https://doi.org/10.1111/j.1467-7687.2009.00865.x>
- Berwick, R.C., Okanoya, K., Beckers, G.J.L., Bolhuis, J.J., 2011. Songs to syntax: the linguistics of birdsong. *Trends Cogn. Sci.* 15, 113–121.
<https://doi.org/10.1016/j.tics.2011.01.002>
- Bouchard, K.E., Brainard, M.S., 2013. Neural Encoding and Integration of Learned Probabilistic Sequences in Avian Sensory-Motor Circuitry. *J. Neurosci.* 33, 17710–17723. <https://doi.org/10.1523/JNEUROSCI.2181-13.2013>
- Brenowitz, E.A., Beecher, M.D., 2005. Song learning in birds: diversity and plasticity, opportunities and challenges. *Trends Neurosci.* 28, 127–132.
<https://doi.org/10.1016/j.tins.2005.01.004>
- Brenowitz, E.A., Nalls, B., Wingfield, J.C., Kroodsma, D.E., 1991. Seasonal changes in avian song nuclei without seasonal changes in song repertoire. *J. Neurosci.* 11, 1367–1374. <https://doi.org/10.1523/JNEUROSCI.11-05-01367.1991>
- Briefer, E., Aubin, T., Rybak, F., 2009. Response to displaced neighbours in a territorial songbird with a large repertoire. *Naturwissenschaften* 96, 1067–1077.
<https://doi.org/10.1007/s00114-009-0567-0>

- Brosch, M., Scheich, H., 2008. Tone-sequence analysis in the auditory cortex of awake macaque monkeys. *Exp. Brain Res.* 184, 349–361.
<https://doi.org/10.1007/s00221-007-1109-7>
- Cate, C.T., Slater, P.J.B., 1991. Song learning in zebra finches: how are elements from two tutors integrated? *Anim. Behav.* 42, 150–152. [https://doi.org/10.1016/S0003-3472\(05\)80617-7](https://doi.org/10.1016/S0003-3472(05)80617-7)
- Cazala, A., Giret, N., Edeline, J.-M., Del Negro, C., 2019. Neuronal Encoding in a High-Level Auditory Area: From Sequential Order of Elements to Grammatical Structure. *J. Neurosci.* 39, 6150–6161. <https://doi.org/10.1523/JNEUROSCI.2767-18.2019>
- Chechik, G., Nelken, I., 2012. Auditory abstraction from spectro-temporal features to coding auditory entities. *Proc. Natl. Acad. Sci.* 109, 18968–18973.
<https://doi.org/10.1073/pnas.1111242109>
- Chew, S.J., Mello, C., Nottebohm, F., Jarvis, E., Vicario, D.S., 1995. Decrements in auditory responses to a repeated conspecific song are long-lasting and require two periods of protein synthesis in the songbird forebrain. *Proc. Natl. Acad. Sci.* 92, 3406–3410. <https://doi.org/10.1073/pnas.92.8.3406>
- Chew, S.J., Vicario, D.S., Nottebohm, F., 1996. A large-capacity memory system that recognizes the calls and songs of individual birds. *Proc. Natl. Acad. Sci.* 93, 1950–1955. <https://doi.org/10.1073/pnas.93.5.1950>
- Chomsky, N., 2017. Language architecture and its import for evolution. *Neurosci. Biobehav. Rev.* 81, 295–300. <https://doi.org/10.1016/j.neubiorev.2017.01.053>
- Christophe, A., Millotte, S., Bernal, S., Lidz, J., 2008. Bootstrapping Lexical and Syntactic Acquisition. *Lang. Speech* 51, 61–75.
<https://doi.org/10.1177/00238309080510010501>
- Cohen, Y., Shen, J., Semu, D., Leman, D.P., Liberti, W.A., Perkins, L.N., Liberti, D.C., Kotton, D., Gardner, T.J., 2019. Hidden neural states underlie canary song syntax (preprint). *Neuroscience*. <https://doi.org/10.1101/561761>
- Coleman, M.J., 2004. Synaptic Transformations Underlying Highly Selective Auditory Representations of Learned Birdsong. *J. Neurosci.* 24, 7251–7265.
<https://doi.org/10.1523/JNEUROSCI.0947-04.2004>
- Collier, K., Bickel, B., van Schaik, C.P., Manser, M.B., Townsend, S.W., 2014. Language evolution: syntax before phonology? *Proc. R. Soc. B Biol. Sci.* 281, 20140263.
<https://doi.org/10.1098/rspb.2014.0263>
- Comins, J.A., Gentner, T.Q., 2010. Working memory for patterned sequences of auditory objects in a songbird. *Cognition* 117, 38–53.
<https://doi.org/10.1016/j.cognition.2010.06.009>
- Cynx, J., 1990. Experimental determination of a unit of song production in the zebra finch (*Taeniopygia guttata*). *J. Comp. Psychol.* 104, 3–10.
<https://doi.org/10.1037/0735-7036.104.1.3>
- Cynx, J., Nottebohm, F., 1990. Timbre Discrimination in Zebra Finch (*Taeniopygia guttata*) Song Syllables. 1990 104, 303–308.
- De Groof, G., Poirier, C., George, I., Hausberger, M., Van der Linden, A., 2013. Functional changes between seasons in the male songbird auditory forebrain. *Front. Behav. Neurosci.* 7. <https://doi.org/10.3389/fnbeh.2013.00196>
- Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., Pallier, C., 2015. The Neural Representation of Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees. *Neuron* 88, 2–19.
<https://doi.org/10.1016/j.neuron.2015.09.019>

- Del Negro, C., 2005. Selectivity of Canary HVC Neurons for the Bird's Own Song: Modulation by Photoperiodic Conditions. *J. Neurosci.* 25, 4952–4963. <https://doi.org/10.1523/JNEUROSCI.4847-04.2005>
- Del Negro, C., Edeline, J.-M., 2001. Differences in auditory and physiological properties of HVC neurons between reproductively active male and female canaries (*Serinus canaria*). *Eur. J. Neurosci.* 14, 1377–1389. <https://doi.org/10.1046/j.0953-816x.2001.01758.x>
- Del Negro, C., Gahr, M., Leboucher, G., Kreutzer, M., 1998. The selectivity of sexual responses to song displays: effects of partial chemical lesion of the HVC in female canaries. *Behav. Brain Res.* 96, 151–159. [https://doi.org/10.1016/S0166-4328\(98\)00009-6](https://doi.org/10.1016/S0166-4328(98)00009-6)
- Del Negro, Edeline J.-M., 2002. Sex and season influence the proportion of thin spike cells in the canary HVC. *NeuroReport*.
- Ding, N., Patel, A.D., Chen, L., Butler, H., Luo, C., Poeppel, D., 2017. Temporal modulations in speech and music. *Neurosci. Biobehav. Rev.* 81, 181–187. <https://doi.org/10.1016/j.neubiorev.2017.02.011>
- Doupe, A.J., 1997. Song- and Order-Selective Neurons in the Songbird Anterior Forebrain and their Emergence during Vocal Development. *J. Neurosci.* 17, 1147–1167. <https://doi.org/10.1523/JNEUROSCI.17-03-01147.1997>
- Dunn, A.M., Zann, R.A., 1996. Undirected Song Encourages the Breeding Female Zebra Finch to Remain in the Nest. *Ethology* 102, 540–548. <https://doi.org/10.1111/j.1439-0310.1996.tb01146.x>
- Eimas, P.D., 1975. Auditory and phonetic coding of the cues for speech: Discrimination of the [r-l] distinction by young infants. *Percept. Psychophys.* 18, 341–347. <https://doi.org/10.3758/BF03211210>
- Eimas, P.D., siqueland, jusczyk, vigorito, 1971. Speech Perception in Infants 171, 4.
- Endress, A.D., Nespore, M., Mehler, J., 2009. Perceptual and memory constraints on language acquisition. *Trends Cogn. Sci.* 13, 348–353. <https://doi.org/10.1016/j.tics.2009.05.005>
- Estes, K.G., Evans, J.L., Alibali, M.W., Saffran, J.R., 2007a. Can Infants Map Meaning to Newly Segmented Words?: Statistical Segmentation and Word Learning. *Psychol. Sci.* 18, 254–260. <https://doi.org/10.1111/j.1467-9280.2007.01885.x>
- Estes, K.G., Evans, J.L., Else-Quest, N.M., 2007b. Differences in the nonword repetition performance of children with and without specific language impairment: a meta-analysis. *J. Speech Lang. Hear. Res. JSLHR* 50, 177–195. [https://doi.org/10.1044/1092-4388\(2007/015\)](https://doi.org/10.1044/1092-4388(2007/015))
- Fisher, C., Klingler, S.L., Song, H., 2006. What does syntax say about space? 2-year-olds use sentence structure to learn new prepositions. *Cognition* 19–29.
- Fitch, W.T., 2004. Computational Constraints on Syntactic Processing in a Nonhuman Primate. *Science* 303, 377–380. <https://doi.org/10.1126/science.1089401>
- Friston, K., 2005. A theory of cortical responses. *Philos. Trans. R. Soc. B Biol. Sci.* 360, 815–836. <https://doi.org/10.1098/rstb.2005.1622>
- Fujimoto, H., Hasegawa, T., Watanabe, D., 2011. Neural Coding of Syntactic Structure in Learned Vocalizations in the Songbird. *J. Neurosci.* 31, 10023–10033. <https://doi.org/10.1523/JNEUROSCI.1606-11.2011>
- Garrido, M.I., Kilner, J.M., Stephan, K.E., Friston, K.J., 2009. The mismatch negativity: A review of underlying mechanisms. *Clin. Neurophysiol.* 120, 453–463. <https://doi.org/10.1016/j.clinph.2008.11.029>

- Gentner, T.Q., Hulse, S.H., 2000. Female European starling preference and choice for variation in conspecific male song. *Anim. Behav.* 59, 443–458.
<https://doi.org/10.1006/anbe.1999.1313>
- Gentner, T.Q., Hulse, S.H., 1997. Article No. ar980810 Perceptual mechanisms for individual vocal recognition in European starlings, *Sturnus vulgaris*.
- Gentner, T.Q., Margoliash, D., 2003. The Neuroethology of Vocal Communication: Perception and Cognition, in: Simmons, A.M., Fay, R.R., Popper, A.N. (Eds.), *Acoustic Communication*. Springer-Verlag, New York, pp. 324–386.
https://doi.org/10.1007/0-387-22762-8_7
- Goldin-Meadow, S., Yang, C., 2017. Statistical evidence that a child can create a combinatorial linguistic system without external linguistic input: Implications for language evolution. *Neurosci. Biobehav. Rev.* 81, 150–157.
<https://doi.org/10.1016/j.neubiorev.2016.12.016>
- Goller, F., Suthers, R.A., 1996. Role of syringeal muscles in controlling the phonology of bird song. *J. Neurophysiol.* 76, 287–300.
<https://doi.org/10.1152/jn.1996.76.1.287>
- Grace, J.A., Amin, N., Singh, N.C., Theunissen, F.E., 2003. Selectivity for Conspecific Song in the Zebra Finch Auditory Forebrain. *J. Neurophysiol.* 89, 472–487.
<https://doi.org/10.1152/jn.00088.2002>
- Griessmann, B., Naguib, M., 2002. Song Sharing in Neighboring and Non-Neighboring Thrush Nightingales (*Luscinia luscinia*) and its Implications for Communication. *Ethology* 108, 377–387. <https://doi.org/10.1046/j.1439-0310.2002.00781.x>
- Hahnloser, R.H.R., Kozhevnikov, A.A., Fee, M.S., 2002. An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* 419, 65–70.
<https://doi.org/10.1038/nature00974>
- Hausberger, M., 1997. Social influences on song acquisition and sharing in the European starling (*Sturnus vulgaris*), in: *Social Influences on Vocal Development*. Cambridge University Press, New York, NY, US, pp. 128–156.
<https://doi.org/10.1017/CBO9780511758843.008>
- Hauser, M.D., 2002. The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science* 298, 1569–1579.
<https://doi.org/10.1126/science.298.5598.1569>
- Hay, J.F., Pelucchi, B., Estes, K.G., Saffran, J.R., 2011. Linking sounds to meanings: Infant statistical learning in a natural language. *Cognit. Psychol.* 63, 93–106.
<https://doi.org/10.1016/j.cogpsych.2011.06.002>
- Holland, J., Dabelsteen, T., Paris, A.L., 2000. Coding in the song of the wren: importance of rhythmicity, syntax and element structure. *Anim. Behav.* 60, 463–470.
<https://doi.org/10.1006/anbe.2000.1529>
- Hurford, J.R., 2011. *The Origins of Grammar: Language in the Light of Evolution II*. OUP Oxford.
- James, L.S., Sakata, J.T., 2017. Learning Biases Underlie “Universals” in Avian Vocal Sequencing. *Curr. Biol.* 27, 3676–3682.e4.
<https://doi.org/10.1016/j.cub.2017.10.019>
- Janata, P., Margoliash, D., 1999. Gradual Emergence of Song Selectivity in Sensorimotor Structures of the Male Zebra Finch Song System. *J. Neurosci.* 19, 5108–5118.
<https://doi.org/10.1523/JNEUROSCI.19-12-05108.1999>
- Jusczyk, 1999. Finding and remembering words: some beginnings by english-learning infants. *Am. Psychol. Soc.*

- Katlowitz, K.A., Picardo, M.A., Long, M.A., 2018. Stable Sequential Activity Underlying the Maintenance of a Precisely Executed Skilled Behavior. *Neuron* 98, 1133-1140.e3. <https://doi.org/10.1016/j.neuron.2018.05.017>
- Kikuchi, Y., Attaheri, A., Wilson, B., Rhone, A.E., Nourski, K.V., Gander, P.E., Kovach, C.K., Kawasaki, H., Griffiths, T.D., Howard, M.A., Petkov, C.I., 2017. Sequence learning modulates neural responses and oscillatory coupling in human and monkey auditory cortex. *PLoS Biol.* 15, e2000219. <https://doi.org/10.1371/journal.pbio.2000219>
- Kirn, J.R., Clower, R.P., Kroodsma, D.E., Devoogd, T.J., 1989. Song-related brain regions in the red-winged blackbird are affected by sex and season but not repertoire size. *J. Neurobiol.* 20, 139–163. <https://doi.org/10.1002/neu.480200304>
- Kroodsma, D.E., Byers, B.E., 1991. The Function(s) of Bird Song. *Birds* 31, 318–328.
- Kuhl, P.K., 2014. Early Language Learning and the Social Brain. *Cold Spring Harb. Symp. Quant. Biol.* 79, 211–220. <https://doi.org/10.1101/sqb.2014.79.024802>
- Kuhl, P.K., 2010. Brain Mechanisms in Early Language Acquisition. *Neuron* 67, 713–727. <https://doi.org/10.1016/j.neuron.2010.08.038>
- Kuhl, P.K., 2004. Early language acquisition: cracking the speech code. *Nat. Rev. Neurosci.* 5, 831–843. <https://doi.org/10.1038/nrn1533>
- Kuhl, P.K., Tsao, F.-M., Liu, H.-M., 2003. Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proc. Natl. Acad. Sci.* 100, 9096–9101. <https://doi.org/10.1073/pnas.1532872100>
- Langus, A., Mehler, J., Nespors, M., 2017. Rhythm in language acquisition. *Neurosci. Biobehav. Rev.* 81, 158–166. <https://doi.org/10.1016/j.neubiorev.2016.12.012>
- Lawson, S.L., Fishbein, A.R., Prior, N.H., Ball, G.F., Dooling, R.J., 2018. Relative salience of syllable structure and syllable order in zebra finch song. *Anim. Cogn.* 21, 467–480. <https://doi.org/10.1007/s10071-018-1182-2>
- Lehongre, K., Aubin, T., Del Negro, C., 2009. Influence of social conditions in song sharing in the adult canary. *Anim. Cogn.* 12, 823–832. <https://doi.org/10.1007/s10071-009-0241-0>
- Lehongre, K., Aubin, T., Robin, S., Del Negro, C., 2008. Individual Signature in Canary Songs: Contribution of Multiple Levels of Song Structure. *Ethology* 114, 425–435. <https://doi.org/10.1111/j.1439-0310.2008.01486.x>
- Lehongre, K., Del Negro, C., 2011. Representation of the bird's own song in the canary HVC: contribution of broadly tuned neurons. *Neuroscience* 173, 93–109. <https://doi.org/10.1016/j.neuroscience.2010.11.015>
- Lehongre, K., Lenouvel, P., Draganoiu, T., Del Negro, C., 2006. Long-term effect of isolation rearing conditions on songs of an 'open-ended' song learner species, the canary. *Anim. Behav.* 72, 1319–1327. <https://doi.org/10.1016/j.anbehav.2006.03.025>
- Leitner, S., Voigt, C., Garcia-Segura, L.-M., Van't Hof, T., Gahr, M., 2001. Seasonal Activation and Inactivation of Song Motor Memories in Wild Canaries Is Not Reflected in Neuroanatomical Changes of Forebrain Song Areas. *Horm. Behav.* 40, 160–168. <https://doi.org/10.1006/hbeh.2001.1700>
- Lemus, L., Hernandez, A., Romo, R., 2009. Neural codes for perceptual discrimination of acoustic flutter in the primate auditory cortex. *Proc. Natl. Acad. Sci.* 106, 9471–9476. <https://doi.org/10.1073/pnas.0904066106>
- Lengagne, T., Aubin, T., Jouventin, P., Lauga, J., 2000. Perceptual salience of individually distinctive features in the calls of adult king penguins. *J. Acoust. Soc. Am.* 107, 508–516. <https://doi.org/10.1121/1.428319>

- Lewicki, M.S., 1996. Intracellular Characterization of Song-Specific Neurons in the Zebra Finch Auditory Forebrain. *J. Neurosci.* 16, 5854–5863. <https://doi.org/10.1523/JNEUROSCI.16-18-05854.1996>
- Lewicki, M.S., Arthur, B.J., 1996. Hierarchical Organization of Auditory Temporal Context Sensitivity. *J. Neurosci.* 16, 6987–6998. <https://doi.org/10.1523/JNEUROSCI.16-21-06987.1996>
- Lewicki, M.S., Konishi, M., 1995. Mechanisms underlying the sensitivity of songbird forebrain neurons to temporal order. *Proc. Natl. Acad. Sci.* 92, 5582–5586. <https://doi.org/10.1073/pnas.92.12.5582>
- Lipkind, D., Marcus, G.F., Bemis, D.K., Sasahara, K., Jacoby, N., Takahasi, M., Suzuki, K., Feher, O., Ravbar, P., Okanoya, K., Tchernichovski, O., 2013. Stepwise acquisition of vocal combinatorial capacity in songbirds and human infants. *Nature* 498, 104–108. <https://doi.org/10.1038/nature12173>
- Long, M.A., Fee, M.S., 2008. Using temperature to analyse temporal dynamics in the songbird motor pathway. *Nature* 456, 189–194. <https://doi.org/10.1038/nature07448>
- Long, M.A., Jin, D.Z., Fee, M.S., 2010. Support for a synaptic chain model of neuronal sequence generation. *Nature* 468, 394–399. <https://doi.org/10.1038/nature09514>
- Lu, K., Vicario, D.S., 2017. Familiar But Unexpected: Effects of Sound Context Statistics on Auditory Responses in the Songbird Forebrain. *J. Neurosci.* 37, 12006–12017. <https://doi.org/10.1523/JNEUROSCI.5722-12.2017>
- Lu, K., Vicario, D.S., 2014. Statistical learning of recurring sound patterns encodes auditory objects in songbird forebrain. *Proc. Natl. Acad. Sci.* 111, 14553–14558. <https://doi.org/10.1073/pnas.1412109111>
- Lynch, G.F., Okubo, T.S., Hanuschkin, A., Hahnloser, R.H.R., Fee, M.S., 2016. Rhythmic Continuous-Time Coding in the Songbird Analog of Vocal Motor Cortex. *Neuron* 90, 877–892. <https://doi.org/10.1016/j.neuron.2016.04.021>
- Margoliash, D., 1986. Preference for autogenous song by auditory neurons in a song system nucleus of the white-crowned sparrow. *J. Neurosci.* 6, 1643–1661. <https://doi.org/10.1523/JNEUROSCI.06-06-01643.1986>
- Margoliash, D., 1983. Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. *J. Neurosci.* 3, 1039–1057. <https://doi.org/10.1523/JNEUROSCI.03-05-01039.1983>
- Margoliash, Fortune, 1992. Temporal and Harmonic Combination-Sensitive Neurons in the Zebra Finch's HVC. *J. Neurosci.*
- Markov, 1913. An example of Statistical Study on the Text of Eugene Onegin illustrating the linking of events to a chain. *Izv. Imp Akad. Nauk VI*, 153–162.
- Markowitz, J.E., Ivie, E., Kligler, L., Gardner, T.J., 2013. Long-range Order in Canary Song. *PLoS Comput. Biol.* 9, e1003052. <https://doi.org/10.1371/journal.pcbi.1003052>
- Mathevon, N., Aubin, T., 1997. Reaction to conspecific degraded song by the wren *Troglodytes troglodytes*: Territorial response and choice of song post. *Behav. Processes* 39, 77–84. [https://doi.org/10.1016/S0376-6357\(96\)00046-0](https://doi.org/10.1016/S0376-6357(96)00046-0)
- McArthur, P.D., 1986. Similarity of playback songs to self song as a determinant of response strength in song sparrows (*Melospiza melodia*). *Anim. Behav.* 34, 199–207. [https://doi.org/10.1016/0003-3472\(86\)90024-2](https://doi.org/10.1016/0003-3472(86)90024-2)
- McCasland, J., 1987. Neuronal control of bird song production. *J. Neurosci.* 7, 23–39. <https://doi.org/10.1523/JNEUROSCI.07-01-00023.1987>

- McCasland, J.S., Konishi, M., 1981. Interaction between auditory and motor activities in an avian song control nucleus. *Proc. Natl. Acad. Sci.* 78, 7815–7819.
<https://doi.org/10.1073/pnas.78.12.7815>
- McKenna, T.M., Weinberger, N.M., Diamond, D.M., 1989. Responses of single auditory cortical neurons to tone sequences. *Brain Res.* 481, 142–153.
[https://doi.org/10.1016/0006-8993\(89\)90494-0](https://doi.org/10.1016/0006-8993(89)90494-0)
- Meitzen, J., Perkel, D.J., Brenowitz, E.A., 2007. Seasonal changes in intrinsic electrophysiological activity of song control neurons in wild song sparrows. *J. Comp. Physiol. A* 193, 677–683. <https://doi.org/10.1007/s00359-007-0222-1>
- Meitzen, J., Weaver, A.L., Brenowitz, E.A., Perkel, D.J., 2009. Plastic and Stable Electrophysiological Properties of Adult Avian Forebrain Song-Control Neurons across Changing Breeding Conditions. *J. Neurosci.* 29, 6558–6567.
<https://doi.org/10.1523/JNEUROSCI.5571-08.2009>
- Mello, C., Nottebohm, F., Clayton, D., 1995. Repeated exposure to one song leads to a rapid and persistent decline in an immediate early gene's response to that song in zebra finch telencephalon. *J. Neurosci.* 15, 6919–6925.
<https://doi.org/10.1523/JNEUROSCI.15-10-06919.1995>
- Mello, C.V., Vicario, D.S., Clayton, D.F., 1992. Song presentation induces gene expression in the songbird forebrain. *Proc. Natl. Acad. Sci.* 89, 6818–6822.
<https://doi.org/10.1073/pnas.89.15.6818>
- Mello, Ribeiro, 1998. ZENK protein regulation by song in the brain of songbirds. *J. Comp. Neurol.* 393, 426–438.
- Menardy, F., Giret, N., Del Negro, C., 2014. The presence of an audience modulates responses to familiar call stimuli in the male zebra finch forebrain. *Eur. J. Neurosci.* 40, 3338–3350. <https://doi.org/10.1111/ejn.12696>
- Menardy, F., Touiki, K., Dutrieux, G., Bozon, B., Vignal, C., Mathevon, N., Del Negro, C., 2012. Social experience affects neuronal responses to male calls in adult female zebra finches: Call familiarity affects neuronal responses in NCM. *Eur. J. Neurosci.* 35, 1322–1336. <https://doi.org/10.1111/j.1460-9568.2012.08047.x>
- Mooney, R., 2014. Auditory–vocal mirroring in songbirds. *Philos. Trans. R. Soc. B Biol. Sci.* 369, 20130179. <https://doi.org/10.1098/rstb.2013.0179>
- Mooney, R., 2000. Different Subthreshold Mechanisms Underlie Song Selectivity in Identified HVC Neurons of the Zebra Finch. *J. Neurosci.* 20, 5420–5436.
<https://doi.org/10.1523/JNEUROSCI.20-14-05420.2000>
- Mooney, R., Hoese, W., Nowicki, S., 2001. Auditory representation of the vocal repertoire in a songbird with multiple song types. *Proc. Natl. Acad. Sci.* 98, 12778–12783.
<https://doi.org/10.1073/pnas.221453298>
- Mouterde, S.C., Elie, J.E., Mathevon, N., Theunissen, F.E., 2017. Single Neurons in the Avian Auditory Cortex Encode Individual Identity and Propagation Distance in Naturally Degraded Communication Calls. *J. Neurosci.* 37, 3491–3510.
<https://doi.org/10.1523/JNEUROSCI.2220-16.2017>
- Müller, C.M., Leppelsack, H.-J., 1985. Feature extraction and tonotopic organization in the avian auditory forebrain. *Exp. Brain Res.* 59.
<https://doi.org/10.1007/BF00261351>
- Nagarajan, S.S., Cheung, S.W., Bedenbaugh, P., Beitel, R.E., Schreiner, C.E., Merzenich, M.M., 2002. Representation of Spectral and Temporal Envelope of Twitter Vocalizations in Common Marmoset Primary Auditory Cortex. *J. Neurophysiol.* 87, 1723–1737. <https://doi.org/10.1152/jn.00632.2001>

- Nakamura, K.Z., Okanoya, K., 2004. Neural correlates of song complexity in Bengalese finch high vocal center: *NeuroReport* 15, 1359–1363.
<https://doi.org/10.1097/01.wnr.0000125782.35268.d6>
- Nealen, P.M., Schmidt, M.F., 2006. Distributed and Selective Auditory Representation of Song Repertoires in the Avian Song System. *J. Neurophysiol.* 96, 3433–3447.
<https://doi.org/10.1152/jn.01130.2005>
- Nieto-Diego, J., Malmierca, M.S., 2016. Topographic Distribution of Stimulus-Specific Adaptation across Auditory Cortical Fields in the Anesthetized Rat. *PLOS Biol.* 14, e1002397. <https://doi.org/10.1371/journal.pbio.1002397>
- Nishikawa, J., Okada, M., Okanoya, K., 2008. Population coding of song element sequence in the Bengalese finch HVC. *Eur. J. Neurosci.* 27, 3273–3283.
<https://doi.org/10.1111/j.1460-9568.2008.06291.x>
- Nottebohm, F., Nottebohm, M.E., Crane, L., 1986. Developmental and seasonal changes in canary song and their relation to changes in the anatomy of song-control nuclei. *Behav. Neural Biol.* 46, 445–471. [https://doi.org/10.1016/S0163-1047\(86\)90485-1](https://doi.org/10.1016/S0163-1047(86)90485-1)
- Nottebohm, F., Nottebohm, M.E., Crane, L.A., Wingfield, J.C., 1987. Seasonal changes in gonadal hormone levels of adult male canaries and their relation to song. *Behav. Neural Biol.* 47, 197–211. [https://doi.org/10.1016/S0163-1047\(87\)90327-X](https://doi.org/10.1016/S0163-1047(87)90327-X)
- Nottebohm, F., Stokes, T.M., Leonard, C.M., 1976. Central control of song in the canary, *Serinus canarius*. *J. Comp. Neurol.* 165, 457–486.
<https://doi.org/10.1002/cne.901650405>
- Nowicki, S., Searcy, W.A., 2004. Song Function and the Evolution of Female Preferences: Why Birds Sing, Why Brains Matter. *Ann. N. Y. Acad. Sci.* 1016, 704–723.
<https://doi.org/10.1196/annals.1298.012>
- Okanoya, K., 2004. The Bengalese Finch: A Window on the Behavioral Neurobiology of Birdsong Syntax. *Ann. N. Y. Acad. Sci.* 1016, 724–735.
<https://doi.org/10.1196/annals.1298.026>
- Ono, S., Okanoya, K., Seki, Y., 2016. Hierarchical emergence of sequence sensitivity in the songbird auditory forebrain. *J. Comp. Physiol. A* 202, 163–183.
<https://doi.org/10.1007/s00359-016-1070-7>
- Pérez-González, D., Malmierca, M.S., Covey, E., 2005. Novelty detector neurons in the mammalian auditory midbrain. *Eur. J. Neurosci.* 22, 2879–2885.
<https://doi.org/10.1111/j.1460-9568.2005.04472.x>
- Perruchet, P., Pacton, S., 2006. Implicit learning and statistical learning: one phenomenon, two approaches. *Trends Cogn. Sci.* 10, 233–238.
<https://doi.org/10.1016/j.tics.2006.03.006>
- Perruchet, P., Poulin-Charronnat, B., Tillmann, B., Peereman, R., 2014. New evidence for chunk-based models in word segmentation. *Acta Psychol. (Amst.)* 149, 1–8.
<https://doi.org/10.1016/j.actpsy.2014.01.015>
- Perruchet, P., Vinter, A., 1998. PARSER: A Model for Word Segmentation. *J. Mem. Lang.* 39, 246–263. <https://doi.org/10.1006/jmla.1998.2576>
- Perruchet, P., Vinter, A., Pacteau, C., Gallego, J., 2002. The formation of structurally relevant units in artificial grammar learning. *Q. J. Exp. Psychol. Sect. A* 55, 485–503. <https://doi.org/10.1080/02724980143000451>
- Petkov, C.I., Jarvis, E.D., 2012. Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. *Front. Evol. Neurosci.* 4.
<https://doi.org/10.3389/fnevo.2012.00012>

- Petruszewycz, M., 1983. Chaînes de Markov et statistiques linguistiques. *Mots* 7, 85–95.
<https://doi.org/10.3406/mots.1983.1118>
- Prather, J., Mooney, R., 2004. Neural correlates of learned song in the avian forebrain: simultaneous representation of self and others. *Curr. Opin. Neurobiol.* 14, 496–502. <https://doi.org/10.1016/j.conb.2004.06.004>
- Prather, J.F., 2013. Auditory signal processing in communication: Perception and performance of vocal sounds. *Hear. Res.* 305, 144–155.
<https://doi.org/10.1016/j.heares.2013.06.007>
- Prather, J.F., Peters, S., Nowicki, S., Mooney, R., 2008. Precise auditory–vocal mirroring in neurons for learned vocal communication. *Nature* 451, 305–310.
<https://doi.org/10.1038/nature06492>
- Pytte, C.L., Suthers, R.A., 1999. A bird’s own song contributes to conspecific song perception. *Anim. Behav.* 7, 1–10.
- Ratcliffe, Weisman, 1986. Phrase Order Recognition by Brown-headed Cowbirds. *Anim. Behav.* 34, 100–105.
- Reale, R.A., Imig, T.J., 1980. Tonotopic organization in auditory cortex of the cat. *J. Comp. Neurol.* 192, 265–291. <https://doi.org/10.1002/cne.901920207>
- Remage-Healey, L., Coleman, M.J., Oyama, R.K., Schlinger, B.A., 2010. Brain estrogens rapidly strengthen auditory encoding and guide song preference in a songbird. *Proc. Natl. Acad. Sci.* 107, 3852–3857.
<https://doi.org/10.1073/pnas.0906572107>
- Ribeiro, S., Cecchi, G.A., Magnasco, M.O., Mello, C.V., 1998. Toward a song code: evidence for a syllabic representation in the canary brain. *Neuron* 21, 359–371.
[https://doi.org/10.1016/S0896-6273\(00\)80545-0](https://doi.org/10.1016/S0896-6273(00)80545-0)
- Ribeiro, S., Cecchi, G.A., Magnasco, M.O., Mello, C.V., n.d. Toward a Song Code: Evidence for a Syllabic Representation in the Canary Brain. *Neuron* 13, 1–10.
- Rizzolatti, G., Arbib, M.A., 1998. Language within our grasp. *Trends Neurosci.* 21, 188–194. [https://doi.org/10.1016/S0166-2236\(98\)01260-0](https://doi.org/10.1016/S0166-2236(98)01260-0)
- Rose, G.J., Goller, F., Gritton, H.J., Plamondon, S.L., Baugh, A.T., Cooper, B.G., 2004. Species-typical songs in white-crowned sparrows tutored with only phrase pairs. *Nature* 432, 753–758. <https://doi.org/10.1038/nature02992>
- Rosen, M.J., Mooney, R., 2006. Synaptic Interactions Underlying Song-Selectivity in the Avian Nucleus HVC Revealed by Dual Intracellular Recordings. *J. Neurophysiol.* 95, 1158–1175. <https://doi.org/10.1152/jn.00100.2005>
- Rosen, M.J., Mooney, R., 2003. Inhibitory and Excitatory Mechanisms Underlying Auditory Responses to Learned Vocalizations in the Songbird Nucleus HVC. *Neuron* 39, 177–194. [https://doi.org/10.1016/S0896-6273\(03\)00357-X](https://doi.org/10.1016/S0896-6273(03)00357-X)
- Rübsamen, R., Dörrscheidt, G.J., 1986. Tonotopic organization of the auditory forebrain in a songbird, the European starling. *J. Comp. Physiol. A* 158, 639–646.
<https://doi.org/10.1007/BF00603820>
- Saffran, J.R., Hauser, M., Seibel, R., Kapfhammer, J., Tsao, F., Cushman, F., 2008. Grammatical pattern learning by human infants and cotton-top tamarin monkeys. *Cognition* 107, 479–500. <https://doi.org/10.1016/j.cognition.2007.10.010>
- Saffran, J.R., Newport, E.L., Aslin, R.N., 1996. Word Segmentation: The Role of Distributional Cues. *J. Mem. Lang.* 35, 606–621.
<https://doi.org/10.1006/jmla.1996.0032>
- Saffran, J.R., Werker, J.F., Werner, L.A., 2007. The Infant’s Auditory World: Hearing, Speech, and the Beginnings of Language, in: Damon, W., Lerner, R.M. (Eds.),

- Handbook of Child Psychology. John Wiley & Sons, Inc., Hoboken, NJ, USA, p. chpsy0202. <https://doi.org/10.1002/9780470147658.chpsy0202>
- Santolin, C., Saffran, J.R., 2018. Constraints on Statistical Learning Across Species. *Trends Cogn. Sci.* 22, 52–63. <https://doi.org/10.1016/j.tics.2017.10.003>
- Scharff, C., Nottebohm, F., Cynx, J., 1998. Conspecific and heterospecific song discrimination in male zebra finches with lesions in the anterior forebrain pathway. 1998 10.
- Schneider, D.M., Woolley, S.M.N., 2013. Sparse and Background-Invariant Coding of Vocalizations in Auditory Scenes. *Neuron* 79, 141–152. <https://doi.org/10.1016/j.neuron.2013.04.038>
- Sen, K., Theunissen, F.E., Doupe, A.J., 2001. Feature Analysis of Natural Sounds in the Songbird Auditory Forebrain. *J. Neurophysiol.* 86, 1445–1458. <https://doi.org/10.1152/jn.2001.86.3.1445>
- Silva-Pereyra, J., Conboy, B.T., Klarman, L., Kuhl, P.K., 2007. Grammatical Processing without Semantics? An Event-related Brain Potential Study of Preschoolers using Jabberwocky Sentences. *J. Cogn. Neurosci.* 19, 1050–1065. <https://doi.org/10.1162/jocn.2007.19.6.1050>
- Sloas, D.C., Zhuo, R., Xue, H., Chambers, A.R., Kolaczyk, E., Polley, D.B., Sen, K., 2016. Interactions across Multiple Stimulus Dimensions in Primary Auditory Cortex. *eneuro* 3, ENEURO.0124-16.2016. <https://doi.org/10.1523/ENEURO.0124-16.2016>
- Smulders, T.V., Jarvis, E.D., 2013. Different mechanisms are responsible for dishabituation of electrophysiological auditory responses to a change in acoustic identity than to a change in stimulus location. *Neurobiol. Learn. Mem.* 106, 163–176. <https://doi.org/10.1016/j.nlm.2013.08.010>
- Stripling, R., Kruse, A.A., Clayton, D.F., 2001. Development of song responses in the zebra finch caudomedial neostriatum: Role of genomic and electrophysiological activities. *J. Neurobiol.* 48, 163–180. <https://doi.org/10.1002/neu.1049>
- Stripling, R., Volman, S.F., Clayton, D.F., 1997. Response Modulation in the Zebra Finch Neostriatum: Relationship to Nuclear Gene Regulation. *J. Neurosci.* 17, 3883–3893. <https://doi.org/10.1523/JNEUROSCI.17-10-03883.1997>
- Sutter, Margoliash, 1994. Global synchronous response to autogenous song in Zebra Finch HVC. *J. Neurobiol.* 72.
- Suzuki, T.N., 2014. Communication about predator type by a bird using discrete, graded and combinatorial variation in alarm calls. *Anim. Behav.* 87, 59–65. <https://doi.org/10.1016/j.anbehav.2013.10.009>
- Suzuki, T.N., Wheatcroft, D., Griesser, M., 2016. Experimental evidence for compositional syntax in bird calls. *Nat. Commun.* 7. <https://doi.org/10.1038/ncomms10986>
- Tchernichovski, O., Nottebohm, F., Ho, C.E., Pesaran, B., Mitra, P.P., 2000. A procedure for an automated measurement of song similarity. *Anim. Behav.* 59, 1167–1176. <https://doi.org/10.1006/anbe.1999.1416>
- Terleph, T.A., Mello, C.V., Vicario, D.S., 2006. Auditory topography and temporal response dynamics of canary caudal telencephalon. *J. Neurobiol.* 66, 281–292. <https://doi.org/10.1002/neu.20219>
- Theunissen, F.E., Shaevitz, S.S., 2006. Auditory processing of vocal sounds in birds. *Curr. Opin. Neurobiol.* 16, 400–407. <https://doi.org/10.1016/j.conb.2006.07.003>
- Thompson, J.V., Gentner, T.Q., 2010. Song Recognition Learning and Stimulus-Specific Weakening of Neural Responses in the Avian Auditory Forebrain. *J. Neurophysiol.* 103, 1785–1797. <https://doi.org/10.1152/jn.00885.2009>

- Vallentin, D., Kosche, G., Lipkind, D., Long, M.A., 2016. Inhibition protects acquired song segments during vocal learning in zebra finches. *Science* 351, 267–271. <https://doi.org/10.1126/science.aad3023>
- Vallet, E., Beme, I., Kreutzer, M., 1998. Two-note syllables in canary songs elicit high levels of sexual display. *Anim. Behav.* 55, 291–297. <https://doi.org/10.1006/anbe.1997.0631>
- Vallet, E., Kreutzer, M., 1995. Female canaries are sexually responsive to special song phrases. *Anim. Behav.* 49, 1603–1610. [https://doi.org/10.1016/0003-3472\(95\)90082-9](https://doi.org/10.1016/0003-3472(95)90082-9)
- Van Heijningen, C.A.A., de Visser, J., Zuidema, W., ten Cate, C., 2009. Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proc. Natl. Acad. Sci.* 106, 20538–20543. <https://doi.org/10.1073/pnas.0908113106>
- Van Ruijssevelt, L., Washington, S.D., Hamaide, J., Verhoye, M., Keliris, G.A., Van der Linden, A., 2017. Song Processing in the Zebra Finch Auditory Forebrain Reflects Asymmetric Sensitivity to Temporal and Spectral Structure. *Front. Neurosci.* 11, 549. <https://doi.org/10.3389/fnins.2017.00549>
- Volman, S., 1993. Development of neural selectivity for birdsong during vocal learning. *J. Neurosci.* 13, 4737–4747. <https://doi.org/10.1523/JNEUROSCI.13-11-04737.1993>
- Wang, X., Merzenich, M.M., Beitel, R., Schreiner, C.E., 1995. Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. *J. Neurophysiol.* 74, 2685–2706. <https://doi.org/10.1152/jn.1995.74.6.2685>
- Waser, M.S., Marler, P., 1977. Song learning in canaries. *J. Comp. Physiol. Psychol.* 91, 1–7. <https://doi.org/10.1037/h0077299>
- Werker, J.F., Lalonde, C.E., 1988. Cross-Language Speech Perception: Initial Capabilities and Developmental Change 24, 672–683.
- Williams, H., Staples, K., 1992. Syllable chunking in zebra finch (*Taeniopygia guttata*) song. *J. Comp. Psychol. Wash. DC* 106, 278–286.
- Wilson, B., Marslen-Wilson, W.D., Petkov, C.I., 2017. Conserved Sequence Processing in Primate Frontal Cortex. *Trends Neurosci.* 40, 72–82. <https://doi.org/10.1016/j.tins.2016.11.004>
- Wilson, B., Smith, K., Petkov, C.I., 2015. Mixed-complexity artificial grammar learning in humans and macaque monkeys: evaluating learning strategies. *Eur. J. Neurosci.* 41, 568–578. <https://doi.org/10.1111/ejn.12834>
- Wilson, S., LaDeau, S.L., Tøttrup, A.P., Marra, P.P., 2011. Range-wide effects of breeding- and nonbreeding-season climate on the abundance of a Neotropical migrant songbird. *Ecology* 92, 1789–1798. <https://doi.org/10.1890/10-1757.1>

ANNEXE

Annexe

Table des figures

<i>Figure 1 : Schématisation des trois types de grammaires : à état fini, régulière et libre du contexte.....</i>	15
<i>Figure 2 : Exemple de chaînes de Markov d'ordre 1 et 2.....</i>	20
<i>Figure 3 : Illustration des mécanismes de traitement de l'ordre.....</i>	25
<i>Figure 4: Chants de deux espèces d'oiseaux chanteurs.</i>	27
<i>Figure 5 : Effet du changement de l'ordre des syllabes dans le chant du tuteur au cours de la période d'apprentissage.</i>	29
<i>Figure 6: Position des différents types de syllabes dans les chants.</i>	31
<i>Figure 7 : Au cours de la période d'apprentissage, présenter des syllabes isolées ou des paires de syllabes influe sur la structure des chants.</i>	33
<i>Figure 8: L'ordre est porteur d'information.</i>	37
<i>Figure 9 : Le cerveau des oiseaux chanteurs : régions et circuits spécialisés dans la perception, la production et l'apprentissage du chant.</i>	39
<i>Figure 10 : Neurone du HVC répondant sélectivement à la paire de syllabes AB.</i>	41
<i>Figure 11: Exemple de neurone répondant lors de la diffusion de différentes paires de syllabes.....</i>	43
<i>Figure 12: La force des réponses dépend des probabilités de convergence.....</i>	45
<i>Figure 13: Activité audio-motrice d'un « neurone-miroir ».</i>	47
<i>Figure 14: Apprentissage d'une nouvelle séquence comportant une nouvelle syllabe en deux étapes.....</i>	49
<i>Figure 15: Schématisation de la tache Go-No Go.....</i>	54
<i>Figure 16: Intégration non linéaire des informations auditives.</i>	56
<i>Figure 1: Spectrogramme d'un chant de canari adulte</i>	80
<i>Figure 2 : Schématisation de deux types de grammaire, dFSA et pFSA</i>	83
<i>Figure 3: Schématisation des probabilités de divergence et convergence</i>	83
<i>Figure 4: Exemple de probabilités de divergence d'une chaîne de Markov d'ordre 1 et 2.....</i>	85

Figure 4 Représentation graphique (portion) du chant d'un canari selon les probabilités de convergence	87
Figure 5 : Pourcentage d'occurrence des différentes phrases du répertoire d'un individu représentatif.	93
Figure 6 : Enchaînement des phrases selon des séquences chez un individu représentatif.	94
Figure 7: Corrélacion entre les probabilités de transition et l'occurrence des séquences de deux phrases produites par un oiseau.	96
Figure 8: Absence de corrélacion entre les probabilités de divergence et les probabilités de convergence pour les séquences de deux phrases (ordre 1).	97
Figure 9: Corrélacion entre les probabilités de transition et l'occurrence des séquences de trois phrases produites par un oiseau.	99
Figure 10 : Exemple d'absence de corrélacion entre probabilités de divergence et probabilités de convergence pour les séquences de trois phrases.	100
Figure 11: Exemple de dépendance à long terme.....	102
Figure 12: Positions des phrases dans le chant d'un oiseau	104
Figure 13: Position des phrases et chunks d'un oiseau	105
Figure 14: Dépendances à long terme dans le chant selon la place des phrases.....	107
Figure 15: Représentation de l'organisation de chants théoriques via l'analyse MDS	109
Figure 16: Représentation de l'organisation des chants d'un canari par la méthode MDS.....	110
Figure 1: Occurrence des motifs de deux phrases et trois phrases.....	124
Figure 2: Exemple de BOS sélectionné pour un oiseau.....	125
Figure 3: Représentation schématique des chants créés à partir du BOS.....	128
Figure 4 : Exemple de réponses de neurones du HVC lors de la diffusion du BOS, chez 3 oiseaux différents.....	133
Figure 5: Diversité dans les réponses des neurones enregistrés chez un même oiseau lors de la diffusion du BOS.....	135
Figure 6: Absence de corrélacion entre la position des phrases et la force des réponses (Z-score) à ces phrases.	136

<i>Figure 7: Absence de corrélation entre la force des réponses (Z-score) et deux des paramètres acoustiques mesurés sur chacune des phrases des BOS diffusés.</i>	137
<i>Figure 8: Absence de corrélation entre les probabilités de transition et la force des réponses (Z-score).</i>	138
<i>Figure 9: Analyse en Composante Principale</i>	139
Figure 10: Caractérisation en termes de probabilités de transition des phrases qui font répondre plus de 50% des neurones.	142
<i>Figure 11 : Réponses neuronales lors de la diffusion d'un BOS composé de deux fois la même séquence de deux phrases.</i>	143
<i>Figure 12 : Permuter deux phrases ou changer un chunk de place fait chuter trois probabilités de divergence, si on compare au BOS.</i>	146
<i>Figure 13 : Réponses moyennes des neurones pendant la durée des différents chants</i>	148
<i>Figure 14 : Réponses moyennes des neurones pendant la durée des trois chunks des différents chants.</i>	150
<i>Figure 15: Réponses neuronales aux chunks selon leur position dans les chants.</i> ..	151
<i>Figure 16 : Réponses neuronales lors de la diffusion d'autres types dans lesquels les chunks s'enchaînent différemment</i>	152
<i>Figure 18 : Réponses neuronales aux phrases après un changement dans l'ordre des chunks</i>	154
<i>Figure 17 : Effet de la permutation de deux phrases sur les réponses neuronales .</i>	154
<i>Figure 19 : Exemple de réponses d'un neurone suite à un changement dans l'ordre des chunks.</i>	156
<i>Figure 21: Réponses aux phrases du BOS avant et après déplacement d'une phrase</i>	160
<i>Figure 22: Effet du déplacement de la phrase A sur les des réponses d'un neurone</i>	161
<i>Figure 23 : Effet du déplacement de la phrase V sur les des réponses d'un neurone</i>	162
<i>Figure 24 : Exemple de réponses d'un neurone à PA fin et d'un neurone à PA large enregistrés avec la même électrode</i>	163
Figure 25 : Effet de la permutation de deux phrases sur les réponses des deux populations neurones, les neurones à PA fins et les neurones à PA larges.	165

Figure 26: Exemple des réponses d'un neurone à PA fin suite à une permutation dans le chunk	166
--	-----

Table des tableaux :

<i>Tableau 1: Récapitulatif des indices de corrélation entre l'occurrence et les probabilités de transition</i>	96
<i>Tableau 2: Récapitulatif des indices de corrélation entre les probabilités de divergence et les probabilités de convergence pour les séquences de 2 phrases pour les 8 oiseaux.</i>	97
<i>Tableau 3 : Récapitulatif des indices de corrélation entre les probabilités de transition et occurrence des séquences de trois phrases pour les 8 oiseaux.</i>	99
<i>Tableau 4: Récapitulatif des indices de corrélation entre les probabilités de transition pour les séquences de trois phrases pour les 8 oiseaux.</i>	100
<i>Tableau 1: Pourcentage de neurones actifs lors de la diffusion de chaque phrase.</i>	134
<i>Tableau 2: Probabilités de transition au sein des différents chants diffusés :</i>	145

Titre : Codage neuronal de l'ordre des signaux acoustiques dans le chant des oiseaux chanteurs

Mots clés : Electrophysiologie ; oiseaux chanteurs ; audition ; syntaxe

Résumé : Le chant des oiseaux, comme le langage oral chez les humains, nécessite une reconnaissance fine des signaux acoustiques émis. Mais un chant d'oiseau n'est pas qu'une simple succession d'éléments sonores, appelés syllabes. L'agencement des syllabes suit des règles, et, d'un point de vue comportemental, les oiseaux peuvent discriminer les chants en fonction de cet agencement. Comme ils possèdent, de plus, un ensemble de régions cérébrales spécialisées dans la perception, la production et l'apprentissage du chant, ils sont un très bon modèle d'étude des mécanismes neuronaux impliqués dans le traitement de l'ordre temporel des signaux acoustiques. Au cours de cette thèse, cette question a été le cœur de deux études basées sur des enregistrements électrophysiologiques (extracellulaires) de l'activité des neurones lors de la diffusion de chants. La première étude a porté sur le traitement de l'ordre au niveau d'une aire analogue du cortex auditif secondaire des mammifères, le nidopallium caudo-médian (NCM), chez une espèce, le diamant mandarin (*Taeniopygia guttata*), capable de discriminer des chants suivant l'agencement des syllabes. Les neurones du NCM ont la particularité d'avoir des réponses qui diminuent lors des répétitions d'un stimulus et qui augmentent si un nouveau chant est présenté. En se basant sur cette propriété d'adaptation, nous avons montré qu'il n'est pas nécessaire de changer de chant, un changement de l'ordre des syllabes suffit à réinstaller des réponses. Changer l'ordre modifie le contexte acoustique dans lequel se trouve une syllabe (ce qui précède est différent), les neurones du NCM seraient donc sensibles à ce contexte. Pour aller plus loin, nous avons utilisé comme stimuli de courtes séquences composées de 2 syllabes différentes, A et B, organisées selon les structures grammaticales ABAB ou AABB. Les résultats montrent que les neurones détectent cette différence de structure,

apportant un argument supplémentaire à l'hypothèse d'un traitement de la façon dont les syllabes se succèdent dans un chant. La deuxième étude a porté sur le traitement de l'ordre des syllabes dans une aire sensori-motrice, le HVC, chez le canari (*Serinus canaria*). Les neurones du HVC ont déjà été identifiés comme sensibles à l'ordre des syllabes dans le propre chant de l'oiseau (*bird's own song* ou BOS). Le canari produit des chants complexes composés de phrases successives, elles-mêmes contenant des répétitions d'une même syllabe. L'agencement des phrases dans les chants est régi par des probabilités de transition et conduit à former des séquences stéréotypées et récurrentes dans les chants, les chunks. De plus, la position des phrases varie d'une phrase à l'autre. L'activité des neurones du HVC a été enregistrée lors de la diffusion d'un des BOS de l'oiseau composé de 3 chunks et de variantes : le BOS dans lequel seul l'ordre des phrases dans un *chunk* ou seul l'ordre des chunks est modifié. Les résultats montrent un impact plus fort des modifications de l'ordre des chunks que de l'ordre des phrases dans les chunks, sur les réponses neuronales. Contrairement à une étude chez une autre espèce d'oiseau, nos résultats n'apportent pas d'arguments en faveur de l'idée que la sensibilité à l'ordre repose sur un traitement des probabilités de transition entre les phrases. Ils suggèrent que la sensibilité des neurones s'étend sur plusieurs phrases, au-delà de la structure des chunks eux-mêmes et dépend de l'ordre des chunks dans le chant. Ces études montrent que les réponses des neurones à une syllabe ou à une phrase donnée dépendent de ce qui la précède, que ce soit au niveau d'une aire sensorimotrice ou d'une aire auditive de haut niveau. Au niveau du HVC, cette sensibilité au contexte peut s'étendre sur plusieurs phrases, et donc plusieurs secondes, permettant d'envisager comment l'ordre des mots du langage humain pourrait être traité au niveau neuronal.

Title: Neural coding of order of song sequences in songbirds

Keywords: Electrophysiology; songbirds; audition; syntax

Abstract: Songs of songbirds, such as speech in humans, requires fine recognition of emitted acoustic signals. But bird's song isn't only a succession of sound elements, called syllables. The order of syllables follows rules, and, from a behavioral point of view, birds can discriminate songs according to this order. Moreover, since they have a set of specialized brain regions in the perception, production and learning of singing, they are a major model for studying neural mechanisms involved in the processing of temporal order of acoustic signals. During this thesis, the focus was on two studies based on electrophysiological recordings (extracellular) of the neuron's activity during the song diffusion. The first study focused on the treatment of order in an analog area of the mammalian secondary auditory cortex, the caudomedial nidopallium (NCM), in one species, the zebra finch (*Taeniopygia guttata*), able to discriminate songs according to syllables order. NCM neurons responses decrease during iterations of a stimulus and increase if a new song is presented. Based on this adaptation property, we have shown that it's not necessary to change the song, a change in the order of syllables is sufficient to re-establish responses. Changing the order changes the acoustic context of the syllable (what comes before is different), so the NCM neurons would be sensitive to this context. To go further, we used as stimuli a short sequence composed of 2 different syllables, A and B, organized according to ABAB or AABB grammatical structures. The results show that the neurons detect this difference in structure, providing an additional argument to the hypothesis of a treatment depending on how the syllables follow each other in

a song. The second study focused on the treatment of the order of syllables in a sensorimotor area, the HVC (proper noun), in the canary (*Serinus canaria*). HVC neurons have already been identified as sensitive to syllable order in the bird's own song (BOS). The canary produces complex songs composed of successive phrases, themselves containing repetitions of the same syllable. The arrangement of the phrases in the songs depends on transition probabilities and leads to the formation of stereotyped and recurrent sequences, called chunks. In addition, the position of sentences varies from one sentence to another. The activity of the HVC neurons was recorded during the diffusion of one of the BOS composed of 3 chunks and variants: BOS in which only the order of the phrases in a chunk or only the order of the chunks, is changed. The results show a stronger impact on the neuronal responses, of changes in the order of the chunks than in the order of sentences in the chunks. At the contrary to a study of another species of bird, our results don't provide any arguments in favor of the idea that order sensitivity is based on a treatment of transition probabilities between phrases. They suggest that the sensitivity of neurons extends over several phrases, beyond the structure of the chunks themselves and depends on chunks order in the song. These studies show that responses of neurons to a given syllable or phrase depend on what precedes it, whether at the level of a sensorimotor area or a high-level auditory area. At the HVC level, this sensitivity to context can extend over several phrases, and therefore several seconds, to consider how the order of words of human language could be treated at the neuronal level.

