Temporal dynamics in the neural representation of sensorimotor tasks : investigation of the song-timing network in birds.

Fjola Hyseni

THÈSE PRÉSENTÉE

POUR OBTENIR LE GRADE DE

DOCTEUR DE L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE MATHEMATIQUES ET INFORMATIQUE

SPÉCIALITÉ : Informatique

Par *Fjola Hyseni*

Dynamique temporelle dans la représentation neuronale des tâches sensorimotrices : étude du réseau de contrôle du chant chez les oiseaux.

Sous la direction de : Nicolas P. ROUGIER
Co-directeur : Arthur LEBLOIS

Soutenue le 1er Décembre 2023

Membres du jury :

Mme. Adrienne FAIRHALL	Professeur	University of Washington	Rapporteure
M. Albert COMPTE	Professor	IDIBAPS	Rapporteur
Mme. Daniela VALLENTIN	Chargée de recherche	MPI, Seewiesen	Examinatrice
M. Hervé ROUAULT	Chargé de recherche	Université Aix Marseille	Examinateur
M. Alexander PITTI	Professeur	CY Cergy-Paris Université	Examinateur
M. Arthur LEBLOIS	Chargé de recherche	Université de Bordeaux	Co-directeur de these
M. Nicolas P. ROUGIER	Directeur de recherche	Université de Bordeaux	Directeur de these

Temporal dynamics in the neural representation of sensorimotor tasks: investigation of the song-timing network in songbirds.

Abstract:
Temporally precise movement patterns underlie many motor skills, yet the origin of temporal control in motor behaviors remains unclear. The zebra finch song system has shown to be an outstanding model to study temporal control and sequential neuronal activity in the order of tens to hundreds of milliseconds. Like human speech, birdsong relies on a tight muscle coordination, with its premotor nucleus, HVC, responsible for the precise control of song tempo. Current computational models of HVC rely on the synfire chain, a purely feedforward network model that can account for HVC sequential activity. Synfire chains are however not robust to noise and function for a narrow range of feedforward weights, thus requiring fine tuning during learning. On the contrary, attractor dynamics provide networks with robust functional properties that make them an alternative to feedforward models. Therefore, we propose that HVC neuronal dynamics may be modelled using a Ring Attractor with a narrow Gaussian connectivity profile, where recurrent connections allow the formation of an activity bump that remains stable across a wide range of weights.In the case of asymmetrical connectivity, the bump of activity moves across the network, generating sequential neuronal activity. We show that the width of the activity bump, and thus the duration of transient neuronal activation, can be decreased to reproduce the brief activity bursts of HVC neurons. Additionally, we reproduce a syllable duration plasticity experiment by implementing a reward covariance reinforcement learning rule in the network. Consistent with behavioral results, the change in duration is specific to the target syllable. Lastly, we investigate further with an EI network model of spiking neurons and show that with a more biologically plausible and precise model, we are able not only to reproduce HVC's fast spiking dynamics, but also perform with specificity a behavioral learning paradigm to modify syllable duration. These findings are confronted with behavioral results of daily duration changes in birds underdoing a Conditional Auditory Feedback protocol to adaptively change syllable duration.

Keywords: Timing, Songbirds, HVC, Ring Attractor, AdEx

Dynamique temporelle dans la représentation neuronale des tâches sensori-motrices : étude du réseau de contrôle du chant chez les oiseaux.

Résumé:
La précision temporelle des mouvements est à la base de nombreuses tâches motrices et actions innées, mais l'origine du contrôle temporel dans les comportements moteurs n'est toujours pas claire. Le système de chant du diamant mandarin s'est révélé être un excellent modèle pour étudier le contrôle temporel et l'activité neuronale séquentielle sur des durées de quelques dizaines à quelques centaines de millisecondes. Comme la parole humaine, le chant des oiseaux repose sur une coordination musculaire précise, le noyau prémoteur, HVC, étant responsable du contrôle du tempo du chant. Les modèles théoriques actuels de HVC reposent sur la chaîne de propagation de synchronie ('synfire chains'), un modèle de réseau purement feedforward qui peut rendre compte de l'activité séquentielle dans HVC. Cependant, les chaînes de propagation de synchronie ne sont pas robustes au bruit et fonctionnent pour une gamme étroite de poids synaptiques, et nécessitent donc un réglage fin pendant l'apprentissage. A l'inverse, la dynamique d'attracteurs confère aux réseaux des propriétés fonctionnelles robustes qui en font une alternative aux modèles de type "feedforward". Par conséquent, nous proposons que la dynamique neuronale de la HVC puisse être modélisée à l'aide d'un attracteur en anneau avec un profil de connectivité gaussien étroit, où les connexions récurrentes permettent la formation d'une bosse d'activité qui reste stable à travers une large gamme de poids. Dans le cas d'une connectivité asymétrique, la bosse d'activité se déplace à travers le réseau, générant une activité neuronale séquentielle. Nous montrons que la largeur de la bosse d'activité, et donc la durée de l'activation neuronale transitoire, peut être réduite pour reproduire les brèves bouffées d'activité observées dans les neurones de HVC. En outre, nous reproduisons une expérience de plasticité de la durée des syllabes en implémentant une règle d'apprentissage par renforcement dite 'de covariance' dans le réseau. Conformément aux résultats comportementaux, le changement de durée est spécifique à la syllabe cible. Enfin, nous poursuivons nos recherches avec un modèle de réseau Excitateur-Inhibiteur de neurones à potentiels d'action et montrons qu'avec un modèle plus plausible et plus précis sur le plan biologique, nous sommes capables non seulement de reproduire la dynamique neuronale de HVC, mais aussi d'expliquer les modifications de la durée des syllabes observées dans un paradigme d'apprentissage comportemental. Ces résultats sont confrontés aux résultats comportementaux des changements quotidiens de durée chez les oiseaux soumis à un protocole de rétroaction auditive conditionnelle pour modifier de manière adaptative la durée des syllabes.

Mots-clés: Timing, Oiseaux Chanteurs, HVC, Ring Attractor, AdEx

Résumé:

Dans ce manuscrit, nous visons à étudier l'organisation temporelle du contrôle moteur à une échelle inférieure à la seconde dans le modèle animal des oiseaux chanteurs à l'aide de la modélisation computationnelle. Dans le chapitre 1, nous présentons le concept plus large du temps dans la perception et l'action, la taxonomie du timing dans les neurosciences, les différentes échelles du timing et ses substrats anatomiques/neuraux. Parmi les substrats anatomiques, nous distinguons le noyau prémoteur des oiseaux chanteurs comme un modèle biologique particulièrement adapté de la synchronisation motrice à l'échelle de dizaines et de centaines de millisecondes. Contrairement aux mécanismes sous-jacents plus complexes des mammifères, les oiseaux chanteurs possèdent un ensemble spécifique de noyaux cérébraux interconnectés, connu sous le nom de "système de chant", responsable de l'acquisition et de l'exécution de la tâche sensorimotrice de la production du chant. Il s'agit d'une structure plus simple à étudier, tant sur le plan expérimental que sur le plan informatique. En outre, les processus de production et d'apprentissage du chant des oiseaux ressemblent à ceux de la parole humaine. C'est pourquoi, dans le chapitre 2, nous présentons tout d'abord l'anatomie et le comportement des oiseaux chanteurs dans le contexte de la production vocale et de l'apprentissage, ainsi que les similitudes avec les mammifères. Nous terminons le chapitre en nous concentrant uniquement sur la synchronisation motrice, sa plasticité et les noyaux impliqués dans son émergence chez l'oiseau chanteur. Dans l'ensemble, nous pensons que les chapitres 1 et 2 ont mis en évidence la pertinence et les avantages de l'utilisation des circuits de timing du chant chez les oiseaux chanteurs pour modéliser le controle temporel fin des actions à l'échelle inférieure à la seconde. Le chapitre suivant, le chapitre 3, présente un ensemble d'hypothèses et d'approches théoriques du timing moteur inférieur à la seconde, allant des principales classes de modèles utilisées chez différentes espèces (principalement les mammifères) aux modèles spécifiquement dédiés à la simulation de la dynamique du timing du chant chez l'oiseau. Nous essayons de positionner ces derniers dans l'une des classes plus larges et de fournir une correspondance entre un modèle spécifique et un modèle plus général. Enfin, nous incluons une section sur les systèmes dynamiques et les attracteurs, en tant que classe de modèles robustes qui peuvent être proposés pour modéliser la dynamique temporelle dans les noyaux contrôlant le timing du chant: HVC (utilisé comme un nom propre).

En bref, nous considérons que les circuits de timing du chant chez l'oiseau chanteur sont représentatifs du timing moteur à l'échelle inférieure à la seconde. Nous partons du principe que la dynamique observée chez l'oiseau chanteur pourrait permettre de mieux comprendre le contrôle temporel chez d'autres taxons, y compris chez l'humain, en particulier parce que la tâche à accomplir, le chant des oiseaux, est très similaire à la parole humaine. Pour étudier la timing moteur sub-seconde, nous utilisons une approche théorique et formulons deux questions principales:

- Quelle est la dynamique sous-jacente qui donne lieu au contrôle temporel précis d'une tâche sensorimotrice apprise ?
- Quels sont les mécanismes qui permettent la flexibilité comportementale du timing moteur ?

Le noyau prémoteur HVC a été modélisé soit comme une structure de timing autonome, soit comme un élément d'un système distribué, d'où émerge le timing. Dans les deux cas, l'hypothèse d'une structure de type chaîne, qui crée une génération de séquences, est avancée. Nous commençons le chapitre 5 en nous concentrant sur le modèle de chaîne Synfire, qui est considéré comme la norme dans la modélisation du noyau HVC. Nous étudions tout d'abord sa flexibilité temporelle en explorant la plage de poids synaptiques possibles qui permet la propagation de la chaîne synaptique, puis nous explorons la robustesse du modèle au bruit dans l'entrée et les poids synaptiques. Cette étude nous incite à rechercher un modèle plus robuste. Idéalement, ce modèle devrait refléter la structure présumée d'une aire corticale et de HVC, en incorporant des connexions récurrentes excitatrices et l'inhibition. En même temps, il devrait conserver une configuration en forme de chaîne pour tenir compte de la force d'entraînement (propagation en chaîne). Ces caractéristiques requises évoquent le concept d'un type particulier de modèle d'attracteur en anneau, capable de progresser à travers ses points fixes ou ses états d'attracteur.

Nous formulons cette proposition au chapitre 5, et progressons étape par étape pour (i) prouver que les propriétés revendiquées sont intégrées dans le modèle, (ii) montrer les différents mécanismes possibles *drive time*, c'est-à-dire faire en sorte que l'activité se propage dans le réseau, entraînant la génération de séquences, et enfin (iii) montrer qu'une architecture d'attracteur en anneau peut modéliser les caractéristiques les plus spécifiques du noyau HVC, c'est-à-dire le caractère éparse de l'activité relative au chant et la durée des bouffées de potentiels d'action. En outre, nous explorons cette architecture neuronale à deux niveaux de complexité neuronale, un modèle neuronal basé sur le taux de décharge et un modèle neuronal à potentiels d'action, ainsi que dans un réseau excitateur-inhibiteur (EI). Cela permet de s'assurer que nos résultats ne sont pas le fruit de l'approche simplifiée que nous suivons, mais qu'il s'agit plutôt d'une caractéristique qui peut être appliquée à différents modèles. Il permet également de montrer l'importance de l'architecture dans la conduite du comportement, ce qui est conforme à l'hypothèse d'un modèle d'horloge de population (D. V. Buonomano and Laje, 2010).

Cependant, nous ne manquons pas de reconnaître la pertinence plus importante d'un modèle de potentiels d'action pour simuler les spécificités de la dynamique de HVC, c'est-à-dire le nombre de potentiels d'action par bouffée et la durée des bouffées. La dynamique des bouffées dépend à la fois de la force synaptique et des propriétés intrinsèques du neurone, qui sont modélisées par l'adaptation déclenchée par les potentiels d'action dans un modèle de neurone adaptatif exponentiel 'intègre et décharge' (AdEx). Avec les modèles basés sur le taux de décharge et

AdEx, nous montrons également les résultats d'une inhibition simulée au niveau du HVC. Cela pourrait servir de prédiction et guider les études expérimentales.

En outre, le HVC peut être seul à contribuer au timing moteur, ou il peut être guidé par des zones cérébrales en amont : le noyau auditif NIf (nucleus interfacialis of the nidopallium) et/ou le noyau thalamique Uva (nucleus uvaeformis) (détaillé dans le chapitre 2). Alors que les études de lésion du NIf (Cardin et al., 2005) ont révélé que le NIf n'est pas nécessaire à la production de chant chez l'adulte, plusieurs études ont montré que l'Uva est très important dans la production de chant (Coleman and E. T. Vu, 2005) et une étude récente a montré qu'il pilote un ensemble particulier de neurones dans le HVC, ceux qui déchargent à la limite des syllabes (Moll et al., 2023). Étant donné qu'il ne semble pas y avoir de latéralisation (Long and M. S. Fee, 2008) dans la production du chant, il est probable que l'entrée Uva ait un effet sur le timing. En tant que tel, UVA envoie une entrée séquentielle atteignant les deux noyaux HVC (droit et gauche) à des moments distincts du chant et n'aurait aucun effet sur les séquences des bouffées, dont nous supposons toujours qu'elles sont générées au sein de l'HVC. Un modèle testant cette hypothèse est également ajouté à la section 5.4.

La présence d'une plasticité comportementale dans le chant des oiseaux a été révélée initialement par des expériences de changement de hauteur et de durée des syllabes à l'aide d'un paradigme de rétroaction auditive conditionnelle (Conditional Auditory Feedback) (détaillé au chapitre 2). Au chapitre 6, nous visons à reproduire numériquement ce paradigme comportemental pour inciter à un changement de durée dans les deux sens, c'est-à-dire pour allonger et raccourcir la durée de la syllabe cible, à l'aide d'une règle d'apprentissage par renforcement de la récompense-covariance. L'une des principales caractéristiques de cette expérience comportementale, qui représente un défi théorique, est le fait que seul le segment ciblé (syllabe) est affecté. Dans une représentation numérique simplifiée de ce comportement, un ensemble de neurones représente une syllabe. Dans une architecture de chaîne synchrone, un ensemble de neurones, et plus précisément deux couches de neurones non connectés, représentent le début et la fin de la syllabe et la mise à jour des poids ne se produit que dans les connexions entre ces deux couches, ce qui fait que les neurones des deux couches sont actives de manière plus proche dans le temps (raccourcissement de la durée) ou plus loin dans le temps (allongement de la durée). Comme les poids entre les autres couches ne sont pas mis à jour, il n'y a pas d'effet sur les autres syllabes/segments. Cependant, lors de l'utilisation de réseaux neuronaux connectés de manière récurrente, une diminution de la durée peut avoir lieu de différentes manières, et un neurone ne se projette pas uniquement vers l'avant, ce qui fait que ses autres synapses sont également mises à jour, ce qui aurait un effet sur de multiples "ensembles" de neurones, dont certains ne sont pas affectés au segment cible. On peut donc observer un effet sur d'autres syllabes, désigné sous le nom d'"interférence entre les syllabes".

Au chapitre 6, nous examinons tout d'abord si, avec l'architecture de l'attracteur en anneau et la règle d'apprentissage choisie, un changement de durée est pos-

sible. Ensuite, nous vérifions comment les poids sont modifiés pour obtenir ce changement. Nous avons d'abord émis l'hypothèse d'une modification positive locale des poids, similaire à celle observée dans la chaîne synchrone, dans le cas d'un raccourcissement de la durée. De plus, nous avons fait l'hypothèse que ce changement devrait être très étroitement focalisé au niveau des neurones pré- et postsynaptiques affiliés à la syllabe cible. De cette manière, aucun effet sur les autres syllabes ne serait observé. Avant de commencer les simulations d'apprentissage, nous avons testé cette hypothèse en ajoutant un tel changement à la matrice de poids et nous avons confirmé notre hypothèse. Cependant, il ne s'agirait pas d'un mécanisme robuste et comme nous le montrerons dans le chapitre 6, ce n'est pas le mécanisme que nous observons. En outre, des mécanismes d'apprentissage légèrement différents sont observés lorsque nous comparons 4 modèles différents, (i) basé sur le taux de décharge, (ii) Exponential Integrate-and-Fire (EIF), (iii) AdEx neuron model et (iv) réseau EI. Enfin, bien qu'avec toutes les versions de notre modèle, nous soyons capables d'acquérir un changement de durée sans interférence dans aucune des syllabes, nous observons un effet intriguant au niveau de l'exécution. Il s'agit d'un léger changement dans la syllabe suivante, qui s'estompe lorsque l'on fait la moyenne de plusieurs exécutions, dans les différentes simulations aec différentes initialisations. Pour confronter cette observation à des preuves expérimentales, nous analysons les données d'une expérience de décalage de durée, menée au laboratoire par un collègue. Dans la dernière section du chapitre 6, nous présentons ces données, qui révèlent comment la durée de la syllabe cible change au fil des jours et les effet collatéraux sur la syllabe suivante.

Le dernier chapitre (chapitre 7) est consacré à une discussion générale sur les implications de nos résultats et leur position dans la littérature. Dans une perspective plus large sur le timing, nous présentons un modèle d'attracteur en anneau, qui s'inscrit dans la classe des modèles d'horloges de population, comme une approche théorique du timing moteur subseconde. À l'instar des horloges de population, ce modèle offre un timing robuste et flexible. Dans la perspective de l'oiseau chanteur et des neurosciences théoriques, nous discutons de l'émergence possible d'un tel modèle structuré à partir de l'apprentissage initial du chant et de la manière dont nous pouvons reproduire numériquement l'apprentissage par imitation auprès d'un tuteur adulte, pendant la période sensorielle et sensorimotrice de l'apprentissage du chant. En outre, nous proposons que les chaînes de synfire et les attracteurs en anneau puissent en fait se situer dans un continuum. Enfin, nous suggérons une règle d'apprentissage supplémentaire pour expliquer le retour à la durée de base, observé après l'arrêt du CAF.

List of Publications

1. F. Hyseni, N. P. Rougier, et al. "Attractor dynamics drive flexible timing in birdsong." In: *International Conference on Artificial Neural Networks*. Springer. 2023, pp. 112–123. DOI: 10.1007/978-3-031-44198-1_10

2. F. Hyseni, N. Rougier, et al. "Comparative study of the synfire chain and ring attractor model for timing in the premotor nucleus in male Zebra Finches." In: *ESANN 2023-European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*. 2023, pp. 647–652. DOI: 10.14428/esann/2023.ES2023-120

Peer- Reviewed Conferences

1. F. Hyseni, N. Rougier, et al. "An Attractor Model of the Temporal Dynamics of the Songbird's Premotor Nucleus." In: *32nd Annual Computational Neuroscience Meeting (CNS)*. Leipzig, Germany, 2023

2. F. Hyseni, N. Rougier, et al. "Temporal Dynamics in an Attractor Model of the Songbird's Premotor Nucleus." In: *Computational and Systems Neuroscience (COSYNE)*. Lisbon, Portugal, 2022. URL: https://www.world-wide.org/cosyne-22/temporal-dynamics-attractor-model-songbirds-b261b11a/

Invited Talk

1. F. Hyseni, N. Rougier, et al. "A model of Sequence Timing in Songbird's Premotor Nucleus." In: *Student Conference on Biological Sciences (SCBS)*. Tirana, Albania, 2021. URL: 10.5281/zenodo.5877414

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisors, Nicolas Rougier and Arthur Leblois. I still remember our interview and the follow up email, which convinced me that this was the ideal position for me. However, given my medical background, it indeed required a considerable leap of faith on your part to extend this opportunity to me and I am immensely grateful for your trust. Your guidance has not only allowed me to grow within my particular field of study, but also to acquire a profound understanding on what it means to be a good and rigorous scientist.

I would like to extend my thanks to the members of the jury - Prof. Albert Compte, Prof. Adrienne Fairhall, Dr. Hervé Roualt, Dr. Daniela Vallentin and Dr. Alexander Pitti - for accepting the responsibility of reviewing and examining this thesis. Additionaly, I would like to express my appreciation the members of my PhD committe, Pierre-Yves Oudeyer and Nicolas Giret, with whom I had the pleasure to hold annual meetings to discuss the progress of my thesis. Their comments and encouragment over the years have been very valuable.

A special thanks to Roman Ursu, whose experimental contributions have been very important for the second part of this thesis, as detailed in Chapter 6. Our engaging discussions have provided more insight into the behavioral paradigm, and have guided our exploration of interference in the duration shift experiments. This resulted in very important findings. I would also like to express my gratitude to all the members of the songbird group - Remya, Eduarda, Carmen, Xavier, Nathan, Anindita - for the fruitful discussions.

Moreover, I have been very fortunate to be part of two amazing teams, Mnemosyne and Team 4 at the IMN, and I would like to thank both for welcoming me so warmly and for the opportunity to delve into the knowledge of their expertise. Through our weekly team meetings, I have gained extensive knowledge in various domains, of experimental and computational neuroscience.

An important aspect of my scientific journey in Bordeaux has been my engagement with open science. In this regard, I am particularly thankful to Eduarda Centeno for introducing me to BordeauxTea. Our collaboration culminated in an Open Science Workshop in October 2023, days before the submission of this manuscript. I would like to extend a special thank you to Nicolas for his support throughout the organization, as well as for his invaluable remarks and suggestions. I also express my gratitude to the Bordeaux Neurocampus and to all the organizers - Arthur, Deepshika, Yoni, Juan, Nathan, Chloé, Ankur, and Carmen - and speakers. I hope this workshop becomes a tradition that continues in the years to come.

A heartfelt thanks to the Mnemosyne people and my friends here in Bordeaux - Remya, Snigdha, Nikos, Naomi, Deepshika, Eduarda, Chloé, Hugo, Axel, Ankur, Melodie, Maeva, Adrien - for being like a second family these past few years. Getting to know you, sharing this journey, and gaining such brilliant friends has been an absolute pleasure. Snigdha, my teammate, roommate, and confidante, I will always cherish our long evening discussions over dinner or tea.

A loving thank you to my partner, Meti, for whom I have no words to properly convey my gratitude. You have been there at every step of the way, closely by my side, helpful, understanding, sometimes critical, encouraging and always loving, supportive and confident in me. Very briefly, it is wonderful to be so deeply understood. Also, thank you to my friends - Cansu, Tomi, Sukhi, Zeynep, Erta, Suard, Semi, Asjon, Tedi - who followed this journey remotely. Despite the distance, I have always felt your presence and immensely enjoyed travelling together.

Last, but not least, the warmest gratitude, to my family- my mother, Bukuri, my father, Sajet, and my brother, Sueld - thank you for being the best support system I could ever hope for. I cannot express how crucial your unwavering belief in me and your love has been throughout this journey. Thank you for your patience, encouragement, confidence and especially, Sueld, for always having the wisest advice.

It has been a beautiful experience.
Thank you.

Contents

1 Timing

"If you slow speech down too much, it becomes unintelligible and if you speed a musical piece too much, it ceases to be music. " D. Buonomano (2017)

Time is an ubiquitous concept, present in almost every aspect of our lives, linked both to perception and action, at many different scales (microseconds to millions of years). It is important when performing temporally precise movements, such as pressing the brakes at the right moment before hitting another car, and also when discussing age or history from years, decades or centuries ago. Time is linked to memory, as a key factor in remembering when an event took place and it is linked to future planning, in determining when an action should be performed. We implicitly rely on time when determining event order, which is a key factor in causality and associative learning. Time is crucial to communication through language and cognition, in general. Moreover, its essence can clearly be witnessed in some disciplines, such as music, dance, sports etc., where several dimensions of time are combined.

Its broad presence in our everyday lives and its conceptualization, is illustrated by the fact that the word *time* is currently the most used word in the English language. Such usage should imply an understanding of the concept of time, as broad and empirical as this may be. However, defining time is a different challenge. If we were to refer to dictionary definitions, time is a nonspatial continuum that is measured in terms of events which succeed one another from past through present to future (Merriem-Webster); or the entity measured in years, days, hours, minutes, seconds (Cambrigde). Both definitions rely on two different aspects of time, while not encompassing them both and not adding other important properties thereof. For the purpose of this manuscript, a richer definition of time is

crucial. Therefore, in this general introduction, we will try to present several field-specific definitions, in an attempt to build a broader background of time, after which we will focus on its role and representation in neuroscience.

To date, there are two main physical notions of time. The classical notion suggested by Newton, defines time as an 'absolute' and universal quantity that flows independently of objects and events in space. Classical mechanics and dynamical systems theory rely strongly on this notion of time. For instance, the state of an object is described with differential equations involving time derivatives (Strogatz, 1994). The other notion of time is the one suggested by Einstein (1915) through the theory of relativity. Time is referred to as a fourth dimension, which is locally defined in space and can shrink or expand depending on the surrounding objects/events.

Time is tightly linked to event order and thus the principle of causality. Despite the differences in both notions, in essence, the principle of causality, states that cause precedes effect and implies a directionality of time, from past (or present, cause) to present (future, effect). These notions and principles have had a big impact on how we perceive, conceptualize and address time (Rovelli, 2019).

In philosophy, several aspects of time are addressed. These include its definition and properties, its linearity, direction, relative or absolute nature and its perception. The main two philosophical notions of time are summarized by the theories of presentism and eternalism, which relate to the aforementioned physical notions (D. Buonomano and Rovelli, 2021). Presentism states that the present state of the world is real, the past has passed and the future is yet to come. On the other hand, eternalism suggests a four dimensional block (3 spatial dimensions + time), where past, present and future coexist and are equally real.

A long-standing philosophical debate is whether time is a measurable quantity or a notion purely based on our perception. An important theory in this regard is Bergson's theory (Bergson, 1889), which claims that time is an ever-changing and indivisible term that cannot be measured and even if measured, such measure would not reflect the reality of our perception. This suggests that the approaches we use to metricize time as we metricize space, although practical, do not offer an accurate description of time.

Although per definition time is a nonspatial continuum, we tend to "spatialize" it in our daily conversations (Casasanto and Boroditsky, 2008; Traugott, 1978). Indeed, we strongly associate time to space because the latter is a concrete and tangible concept. Across cultures and languages, many spatial metaphors or phrases are present, i.e. *length* of a speech, *shorter* duration and *back* in the past.

The spatial metaphors of time share similarities, but the vocabulary is sometimes not consistent between different languages and cultures (Boroditsky, 2018). For instance, in English we refer to time with the front/back vocabulary, where time is placed in a horizontal axis with the future in the front and the past in the back. However, in Aymara language, the future is placed in the back and the past in the front (Núñez and Sweetser, 2006). Moreover, other languages may place time in the vertical axis (up/down for future/past in Mandarin Chinese (Boroditsky,

Fuhrman, et al., 2011)) or even use cardinal directions (Australian Aboriginal community (Boroditsky and Gaby, 2010)). These findings show that there are different ways in which time is addressed and therefore perceived.

In neuroscience, several test batteries have been proposed and implemented to unravel the underlying neural substrates of time, both in physiological and pathological conditions. A common outcome of many of these studies, is that a multitude of cognitive processes are associated with time and temporal processing. For instance, time is not only present in tasks that explicitly or implicitly necessitate a temporal computation (interval discrimination, duration estimation, sequence generation), but it is also linked to metacognitive tasks such as understanding concepts, memory, time travel, envisioning future actions and planning.

Moreover, our perception of time "durée" (Bergson, 1889) is complex and subjective; it is not linear, but can expand or shrink based on a number of internal and external factors. Despite this, time in neuroscience is generally considered to be linearly evolving. Its accuracy is not claimed at all scales, but there are mechanisms operating at timescales which are as objective as our traditional clock, i.e. the circadian rhythm. On the other hand, mechanisms of time-telling at shorter scales, such as for instance milliseconds, are much less known. Experimental findings show neurons accurately tune to specific interval durations after training (Gouvêa et al., 2015; Kawai et al., 2015). These findings give rise to the question of whether there is an internal representation of time (clock-like) and if so, if it is universal or different across scales and modalities.

Due to the complexity of time in neuroscience, several aspects need to be discussed in detail. In the following sections, we will first present a taxonomy of timing based on evidence from the literature. Then, we will summarize briefly the different scales of time, alongside their underlying mechanisms. Lastly, we will focus on the proposed neural mechanisms of motor timing at the scale of tens to hundreds of milliseconds.

1.1 Taxonomy of Timing

Timing is crucial for a wide range of sensorimotor tasks. However, there are numerous uncertainties regarding the underlying mechanisms. For instance, sensory and motor timing may not rely on the same circuitry, there could be different mechanisms for different scales of timing (subsecond, suprasecond etc.) and the underlying mechanisms could be dedicated or intrinsic (R. Ivry and Schlerf, 2008). To address the broader question of timing, a series of relevant tasks has been identified, the collection of which has given rise to a series of experimental paradigms that have been implemented. According to Paton and D. V. Buonomano (2018), any task that is based on interval or duration, and requires a timing device to be solved, can be considered a timing tasks. Broadly, timing tasks are divided into three main categories (Grondin, 2010): (i) duration/interval discrimination, (ii) production and (iii) reproduction.

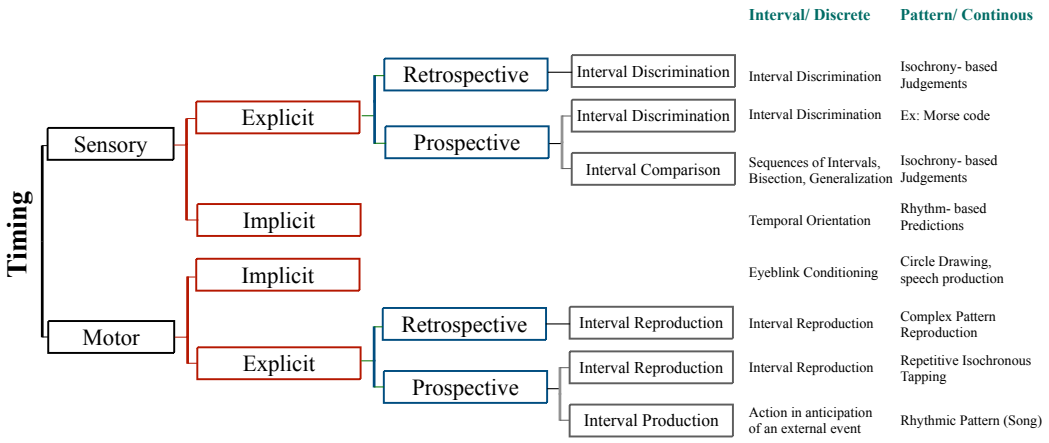


Figure 1.1: **Taxonomy of Timing.** A representation of the taxonomy of timing, compiled based the evidence provided throughout the years (Breska and R. B. Ivry, 2016; J. Coull and Nobre, 2008; Grondin, 2010; M. Mauk and D. Buonomano, 2004): namely sensory/motor, implicit/explicit, prospective/retrospective and interval/pattern timing. Each of the limits of these dimensions is shown to engage different mechanisms, hence bringing the necessity of separating it into subdomains when studying timing. Whereas the other domains represent separate dimensions, prospective/retrospective timing is a subdivision of explicit timing. On the side, different tasks, where different aspects of timing can be studied, are illustrated.

To position the investigated tasks and formulate more systematic hypotheses, a taxonomy of timing is necessary. A taxonomy of timing implies a good understanding of the shared and different mechanisms, underlying different tasks. As such, it is still an open and challenging aspect in neuroscience and there exists several proposals. J. Coull and Nobre (2008) proposes at least two dimensions and four timing sub-domains. The first dimension is that of perceptual and motor timing, whereas the second is implicit and explicit timing. Although these subdomains would lie in a continuum in their respective dimensions, specific neural mechanisms play a role in each of the two ends of the spectrum. A review of such findings in the neuroimaging (fMRI) literature is provided by J. Coull and Nobre (2008), who show that implicit and explicit timing recruit different circuitries. Moreover, they show and suggest that so do sensory and motor timing. A third dimension of timing is suggested by Grondin (2010), that of prospective and retrospective timing. On the other hand, Paton and D. V. Buonomano (2018) suggest a different second and third dimension. In their review, they summarize the timing dimensions as 3: (i) sensory and motor timing, (ii) subsecond and suprasecond and (iii) interval and pattern timing.

There are similarities and differences, originating from the field and tasks investigated, between these three proposed taxonomies. We consider all four proposed dimensions and illustrate each in figure 1.1 alongside their respective timing

task(s). Below, we detail the subdomains, alongside their experimental tasks and paradigms and we include evidence of the different mechanisms reported in each subdomain. Sub and suprasecond timing, also take place in the taxonomy. Although we do not include subsecond and suprasecond timing in the figure, we describe the scales of timing and their mechanisms in section (1.2).

1.1.1 Sensory and Motor Timing

The one common dimension of timing in the three aforementioned reviews is sensory and motor timing. The sensory subdomain of timing is referred to as temporal perception and concerns the perception and 'measurement' of a duration through our senses (which do not include a timing sense). It is a rather tricky concept, as it is dependent on a rather vast amount of external and internal factors, including emotional state (stress, excitement, dread), spatial position, attention, temperature, our actions etc. For instance, in the context of waiting for the bus or running to catch it, two possibly identical durations can be perceived as very different. However, more precise perceptions of time are also present, such as our perception of a day (circadian clock) or that of specific intervals, when trained. The latter has been shown by several studies across different species, which have shown neural tuning to specific time intervals following training ((Bueti, 2011; Bueti and D. V. Buonomano, 2014; D. V. Buonomano, Bramen, et al., 2009; Karmarkar and D. V. Buonomano, 2003; Nagarajan et al., 1998) (electric fish (Carlson, 2009), rats (Zhou et al., 2010), zebra finches (Margoliash, 1983)) (Paton and D. V. Buonomano, 2018)).

On the other hand, motor timing involves tasks where temporal control is necessary to perform an action or anticipate it (Paton and D. V. Buonomano, 2018). In the wild, this would mean escaping, starting to run instantly at the sight of a predator, at a football court hitting the ball at the exact time that the goalkeeper is distracted by someone else, in a piano class playing the right note at the right time etc. Another intuitive example of motor timing is human speech, where words are uttered in a sequence with small pauses in between. Longer pauses translate to commas in the written language and even longer pauses to dots, thus affecting the meaning of the sentence. For instance, a pause in the millisecond scale, can significantly alter sentence meaning: 'Let's eat grandma', insinuating the grandmother should be eaten (?!), as compared to 'Let's eat, grandma!', where she is being invited to join for dinner. Moreover, these speech pauses are affected by the speech rate and vary in duration, number and position in the sentence (Matzinger et al., 2020).

Although the description of sensory and motor timing above, may insinuate they are discrete points in a line, that is not the case. They rather lie in a continuum. However, tasks where one is more heavily present, are necessary to investigate them and the involved circuitry in detail. In such tasks, imaging studies have revealed they do not engage the same areas, with perceptual timing tasks, showing mostly correlation with activity in the anterior portion of SMA (pre-SMA), the right inferior frontal cortex and BG and motor timing tasks, with BG and a context-

dependent coactivation of the SMA (preSMA and/or SMA proper), inferior frontal cortex or cerebellum (J. Coull and Nobre, 2008).

1.1.2 Implicit and Explicit Timing

Implicit timing, refers to tasks where timing develops implicitly from the task in hand, as an automatic, unconscious processing of temporal information, often integral to motor coordination, habituation, or rhythmic activities. On the other hand, explicit timing necessitates a conscious awareness of timing, for instance in estimating how long an event lasted, or how much time needs to pass before performing an action. It requires attentional resources focused on timing. As such, explicit timing is crucial for tasks demanding active temporal judgments and predictions.

Implicit and explicit timing have been shown through behavioral and imaging studies to involve different mechanisms. In a review of fMRI studies (J. Coull and Nobre, 2008), it was shown that during explicit timing there is an activation of the basal ganglia, and a context-dependent co-activation of the prefrontal, premotor and cerebellar areas, whereas during implicit timing, the inferior parietal and premotor areas are active.

1.1.3 Retrospective and Prospective Timing

Another dimension important when designing experimental paradigms to study timing is retrospective and prospective timing. The former is dedicated to time that has passed and the latter to coming time. Retrospective timing, is associated with memory (Block and Zakay, 1997) and to test it, in experimental settings, the subject is asked after a particular interval has passed to report elapsed time. In contrast, for prospective timing, the subject is informed beforehand that a timing task is to take place and is asked to perform a timing-related judgement/action.

An example of retrospective timing in everyday life, could be reporting the amount of time you were stuck in traffic that day, whereas examples of prospective timing are estimating the amount of time before you enter an exam, or the time it takes to finish a presentation. Both prospective and retrospective timing are subject to a wide range of scales, even though retrospective timing studies are known to explore more orders of time as compared to prospective timing (studied mostly on brief intervals). In most taxa, prospective timing above a certain range, i.e. hours, days may not possible, as this would involve more complex cognitive mechanisms and elicit future planning.

1.1.4 Interval and Pattern Timing

Another dimension of timing entails its complexity, the two limits of which being interval (simple, 'absolute') and pattern/dynamic (complex, 'relative', 'beat-based') timing. The former refers to estimating/performing an action based on

the duration of a particular, isolated, event. The latter includes continuous time-keeping as in the case of drawing a circle (Spencer and R. B. Ivry, 2013), where timing is implicit and arises as a property of the performed action. It also includes temporal sequences, whereby time-keeping of the singular intervals, but also the overall sequence structure is required.

Behavioral (Grube et al., 2010), imaging (Teki et al., 2011) and computational (Hardy and D. V. Buonomano, 2016) approaches reveal that there are possibly different mechanisms involved in these two subdomains. An example is the presence of the cerebellar involvement only in interval timing (Breska and R. B. Ivry, 2016; Spencer and R. B. Ivry, 2013), but not in dynamic timing. This has been shown both through imaging studies and observed in impairments from cerebellar pathologies. However, there are exceptions as in the case of rhythmic finger tapping, which is associated with pattern timing, but is still affected by cerebellar pathologies, manifesting with higher temporal variability.

1.2 Timing at Different Temporal Scales

Temporal processing can occur at different timescales (Buhusi and Meck, 2005), from microseconds to a day, across at least 12 orders of magnitude (M. Mauk and D. Buonomano, 2004). Broadly, M. Mauk and D. Buonomano (2004) split the timescales into 4 main groups: circadian rhythms (King and Takahashi, 2000), seconds (Gibbon, 1977), milliseconds (Karmarkar and D. V. Buonomano, 2003) and microseconds (Carr, 1993). Upon observing the broad range of timescales illustrated in Fig.1.2, the question of the underlying mechanisms and whether they are similar arises.

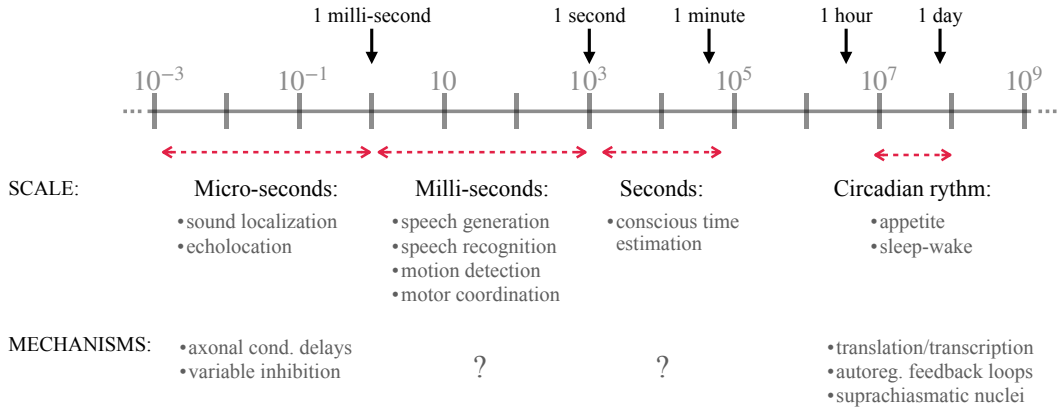


Figure 1.2: Temporal processing at different scales. At least 12 orders of magnitude are reported in human temporal processing. Adapted from M. Mauk and D. Buonomano (2004).

As we have briefly mentioned above different mechanisms are engaged in each. Evidence on the timescale of the circadian rhythm and the microsecond scale reveals their respective mechanisms. The former relies on a "period" gene (Reddy

et al., 1984), which encodes the transcription/ translation of a protein. The amount of this protein serves as a hourglass clock- sandclock with a cycle of a bit more than 24 hours (in humans). The suprachiasmatic nuclei are involved in this process. The latter (microsecond scales) engages in sound localization (delay of sound travel from one ear to the other) and echolocation (Covey and Casseday, 1999). Its mechanism relies on axonal conduction delays and variable inhibition (M. Mauk and D. Buonomano, 2004).

However, the sub- and suprasecond scale remain less clear, with evidence suggesting the latter may be more related to conscious timing and cognitive capacities (Buhusi and Meck, 2005). Time processing in this scale is often referred to as time estimation and recruits the prefrontal and parietal cortices. This is thought to be a very flexible timing mechanism. On the other hand, the tens to hundreds of millisecond scale, is considered 'automatic' (Lewis et al., 2003) and proposed to rely on more sophisticated and complex mechanisms (M. Mauk and D. Buonomano, 2004). It is crucial for speech and fine motor coordination. Moreover, it is closely linked to motor and premotor circuits and Lewis et al. (2003) suggest that it may 'keep time' using a temporal pattern generator mechanism. The differences in mechanism between these two scales have been revealed by pharmacological (Rammsayer, 1999), psychophysical (Karmarkar and D. V. Buonomano, 2007) and imaging studies (Lewis et al., 2003).

1.3 Proposed Neural Mechanisms

The mechanisms underlying timing remain a fundamental and broadly researched topic. There are several hypotheses, which are addressed by different computational models and experimental paradigms, and have their own supporting and violating body of evidence. Since there is no evidence regarding the existence of a dedicated organ or sense for time estimation, this means such estimation derives from internal mechanisms, involving some form of computation (Clark, 1998). In this regard, the two main hypotheses, comprising the two broader model classes are the dedicated and intrinsic model of timing (R. Ivry and Schlerf, 2008).

Dedicated Model

The first hypothesis claims that there exist dedicated/specialized neural mechanisms, involving one or several brain areas, that can perceive duration and execute timely behaviour. The underlying mechanism is hypothesized to be ubiquitous and invariant across conditions, modalities, task and scales (milliseconds to seconds) (Karmarkar and D. V. Buonomano, 2003; Zelaznik et al., 2002). This hypothesis was motivated by several reasons, including (i) its simple and intuitive nature, (ii) the observation that our sense of passage of time is comparable across modalities (visual, auditory) (Grondin and Rousseau, 1991) and (iii) the similarity in temporal variability (ratio between variability and mean duration) between perception and

production tasks (R. B. Ivry and Hazeltine, 1995). Moreover, arguments associating specific brain areas to dedicated timing systems are present. Notably, the cerebellum, basal ganglia, supplementary motor area and the prefrontal cortex have been proposed as candidate regions.

Dedicated models also include distributed time representations, where several areas interact to give rise to timing. Each of these areas can have a particular function, for instance, one being the pacemaker and the other the accumulator, as in the case of the striatal beat frequency model (SBF). In this model, a set of cortical neurons (timekeepers) that oscillate at regular, but distinct frequencies, are followed by striatal integrators that combine the information received with feedback and form the basis of interval timing (M. S. Matell and Meck, 2004; Meck et al., 2008).

Dedicated systems (Bueti, 2011; R. Ivry and Schlerf, 2008) can be split in three broad classes, the pacemaker-accumulator models, process-decay models, and oscillator/coincidence-detection models (M. S. Matell and Meck, 2000). The most commonly addressed models are clock based models, namely the pacemaker accumulator model (Creelman, 1962; Treisman, 1963), and the scalar expectancy theory (Gibbon, 1977). The former relies on a pacemaker, which generates neural pulses and an accumulator that assembles them, creating a linear measure of duration based on the number of accumulated ticks. The latter contains an attentional switch, a comparison step and a decision stage. A particular benefit of the scalar expectancy theory, which is necessary in all timing models at this scale, is the ability to reproduce the *scalar property of timing*. The scalar property of timing, states according to Weber's law, that the longer the duration, the higher the error in timing estimation will be.

Intrinsic Model

Intrinsic models of timing suggest that temporal information is not specialized and associated with (a) clock(s). It instead emerges from the instrinsic cellular and network dynamics of neural cirucuits and depends on the task demands (D. V. Buonomano, 2000; D. V. Buonomano and Merzenich, 1995). This hypothesis is motivated by (i) the importance and ubiquity of temporal computations, (ii) evidence showing that timing is specific to modalities (Burr et al., 2007) and tasks and (iii) the proposed analogy with spatial representations in the brain, which are vast (Paton and D. V. Buonomano, 2018). In support of the second point, time-keeping in different modalities (i.e. visual vs auditory) has been shown to recruit different circuitry (i.e. visual neural mechanisms vs auditory) (Burr et al., 2007; R. Ivry and Schlerf, 2008).

There exist several proposed mechanisms implementing the same conceptual idea. A possible mechanism of intrinsic timing (Pariyadath and D. Eagleman, 2007) suggests that duration may be encoded in the magnitude of neural activity and the amount of energy spent over time. Similar to the SET, here the intensity of the stimulus and attention affect temporal perception (D. M. Eagleman, 2008; Pariya-

dath and D. Eagleman, 2007). Other studies have proposed that timing is encoded in beta oscillations (Fujioka et al., 2012) or in patterns of activity throughout the neural network (D. V. Buonomano and Merzenich, 1995). An example of the latter is the state dependent network (SDN) (D. V. Buonomano and M. D. Mauk, 1994; Karmarkar and D. V. Buonomano, 2007), which proposes that any network could process temporal information, and represent different durations through unique spatial organizations. They propose that temporal selectivity is an intrinsic property of local neural circuits that relies on time-varying synaptic and neuronal properties, most notably short-term synaptic plasticity.

The applicability of either model (at least for perception) could map into the temporal range, where intrinsic models can be more suited for the subsecond and dedicated timing models for the suprasecond scale (R. Ivry and Schlerf, 2008).

Embodied Time

Embodiment of time refers to a tangible, physical structure, the observation of which provides an (rough) estimation of time. In an isolated setting, with nothing but one's self to keep track of time, walking or drawing or any kind of action can be used as a timing mechanism, especially when memory, reward signals or a clock provide cues on how much time has passed at the end of the 'trial'. Indeed, several studies have witnessed exactly this. They observe the appearance of 'superstitious' (stereotypical) behaviors occurring when animals perform a timing task pertaining a reward. These behaviors entail individually unique, repeated actions (or sequences of actions) that seem to be collateral and unrelated to the task. Such actions are interpreted as a way to keep track of time through more easily measurable metrics, i.e. the distance travelled (rats), or rotations/ turns (pigeons). These actions are then likely reinforced upon reward delivery and an association between them and the reward is created, making the animals repeat them in the subsequent trials.

Similar stereotypical behaviors have been observed in humans. In everyday activities, we draw associations between the mean duration of certain tasks and elapsed time, such as the time to drink a coffee, smoke, walk to work, or give a well-rehearsed presentation. For instance in the latter, we know 45 minutes have passed, probably not because we have an internal clock measuring each second passing by, or because we have neurons tuned to this specific duration. Instead, it is because after many rehearsals, we know approximately how long it generally takes us to get to any point of the presentation, based on the speed of speech. However, it is important to note the difference in the scale of time, as the mechanisms described here are in the suprasecond scales. Conversely, in the subsecond scale, evidence have shown that neural tuning to a certain duration occurs.

1.4 Neural Substrates of Motor Timing

Investigation of motor timing and its proposed neural mechanisms can be performed both in a physiological and pathological state, either experimentally induced or clinically present. A loop of knowledge between the anatomical substrates of timing and its clinical impairments emerges, as findings and observations from both symbiotically work to bring together a general picture of understanding, providing a clearer view of time in neuroscience. Based on the literature, no neurological or psychiatric disorder has timing deficits in every scale and none appears solely with timing deficits (J. T. Coull et al., 2011; Paton and D. V. Buonomano, 2018). This statement could suggest different timing mechanisms in different scales, and either distributed or intrinsic mechanisms of timing. As briefly mentioned above, several areas have been proposed as candidates for encoding and/or generation of timing tasks (defined in section 1.1), in the framework of each of the proposed neural mechanisms in section 1.3. These include the basal ganglia, the cerebellum, the premotor cortex, supplementary motor (SMA), Brocca and precentral gyrus (J. T. Coull et al., 2011). Some of these proposals originate from the human literature, behavioral observations in certain tasks, conditions or in lesions, presenting with disruptions in motor timing.

An example of a transfer of knowledge between clinics and neuroanatomy is the observation of the behaviour of patients with Parkinson's disease (PD), which has become an incentive for the investigation of motor timing's substrates in the striatum and the dopaminergic system. Other disorders affecting the basal ganglia and also presenting with altered perceptual or motor timing, include Huntington's disease and Tourette's. The most studied basal ganglia dysfunction, Parkinson's disease is characterized by motor deficits, including bradykinesia (slowness of movement). In PD, timing impairments are pronounced in explicit timing (section 1.1.2), both in the perceptual and motor subdomain. In the sensory domain they are present in the subsecond and second timescale, across multiple modalities (visual, auditory, tactile) and are correlated with disease severity (Artieda et al., 1992).

In the motor domain, timing impairments manifest with bradykinesia, with accuracy reduction in the sub- and suprasecond scale and higher timing variability (inconsistent across studies) (J. T. Coull et al., 2011). Moreover, a prominent finding is the overestimation of shorter duration and underestimation of longer ones, hence bringing them closer in the estimation. This is referred to as a 'migratory' effect and was present in the retrieval phase. Patients in the off-medicate state overestimated a shorter duration (8 s) and underestimated a longer one (21 s) (Malapani et al., 1998). Another study with nonmedicated patients showed overestimation of both shorter and longer durations, suggesting a role of the basal ganglia in determining the clock speed (Pastor et al., 1992). A very recent study supports the clock-speed hypothesis, by cooling the striatum with a thermoelectric device (Monteiro et al., 2023). They showed that cooler temperatures cause dilation, and warmer temperatures contraction, of both neural activity and temporal judgement.

However, the speed of this activity does not affect continuous kinematics. Another study revealed that the effects of striatum inactivation were greater for *suprasecond* timing (Kunimatsu et al., 2018), while those in the cerebellum for subsecond timing. Taken together, these studies suggest that the basal ganglia is a *mid-level controller*, with no detailed control on the sequence, but control on selection, link and modulation of actions (Graybiel, 1998; Monteiro et al., 2023; Park et al., 2020).

Cerebellum is another structure implicated in timing. Initially, it was associated solely with motor timing, since in the case of cerebellar ataxia, difficulty in the precise temporal control is observed. A number of studies claim a role for the cerebellum in timing in the *subsecond* range, in the sensory and motor subdomain (J. T. Coull et al., 2011), and in explicit and implicit timing (Breska and R. B. Ivry, 2016). Examples of motor tasks used to test timing in cerebellar patients, include interval reproduction, synchronization-continuation tapping task, circle production and the best known task: eyeblink conditioning test. Several hypotheses are made on the role of cerebellum. One hypothesis proposes a context- dependant role, emerging in the case of malfunction of the corticostriatal pathway (Merchant et al., 2013) or when processing at subsecond scale (Lewis et al., 2003) is needed. Another hypothesis is built in a review, where Breska and R. B. Ivry (2016) first present a series of studies highlighting the importance of the cerebellum in numerous timing tasks. They then show the tasks not associated with the cerebellum, including circle drawing, isochrony-based judgements and rhythm-based predictions. Such tasks can be grouped as dynamic or continuous tasks, which leads to their proposal that the cerebellum may be required when timing discrete intervals, but not when it is intrinsic to rhythmic or continuous dynamics.

Cortical motor areas are also involved in timing, and as the intrinsic hypothesis suggests, timing could be an intrinsic property of the cortical circuits. Studies have revealed the engagement of the primary visual (Gavornik and Bear, 2014) and auditory cortices (Brosch et al., 2005) in timing tasks including the respective modalities. Moreover, motor cortical areas are shown to be implicated in timed actions, such as controlling self initiated actions (Mushiake et al., 1991), and sequentially organizing multiple actions (Tanji, 2001). Hence, it is suggested that, beyond just representing features of motor output, motor regions may have intrinsic oscillatory dynamics or sequential dynamics (Mita et al., 2009; Shenoy et al., 2013) to act as their own pattern generators.

As briefly mentioned above, the aforementioned brain areas are likely linked in a distributed system. For instance, the basal ganglia (McFarland and Haber, 2001) and cerebellum (Prevosto et al., 2010) send input through the thalamus to the cortex. In rodents, these cortical areas are respectively the motor cortex and deeper cortical layers (Kaneko, 2013) and they might be specifically involved in the cortico-basal ganglia and the cortico-cerebellar loops. However, further research in necessary.

The findings presented above, although mainly in humans, contain and were supported by evidence found in other taxa. One such example are songbirds, which have a somewhat simpler and clearer timing mechanism. In song-producing

male zebra finches a premotor area, HVC, contains neurons firing in a time-locked manner to the song (R. Hahnloser et al., 2002). Manipulations on this area, with a cooling/heating paradigm, have shown, that both neural activity and song timing are a function of the temperature given to the nucleus, exhibiting constriction when the nucleus is heated, and dilation otherwise (Long and M. S. Fee, 2008). This effect is only seen in manipulations to this area and not on the subsequent motor area, which is line with the hypothesis of a population clock model, suggesting that the neuronal population of this area encodes the temporal dynamics of song production at the scale of tens to hundreds of milliseconds.

2 Songbirds and Birdsong: Vocalizations and Timing

2.1 Cognition and Anatomy in Birds

Birds have been long studied across a variety of different fields and subjects. They are particularly puzzling in the field of neuroscience due to their prominent cognitive skills. Traditionally, the evolution of cognition and brain development is represented as a function of telencephalic evolution and is thought to have occurred in a progressive pattern, culminating with the human cerebrum (Edinger et al., 1903). In this terminology, birds represented an earlier stage of evolution, because most of the avian telencephalon was considered a hypertrophied basal ganglia. However, the modern view of the vertebrate evolution is quite different (Jarvis et al., 2005; Reiner et al., 2004), showing that birds do not represent an earlier stage of evolution but rather an alternate one as shown in Fig. 2.1, all while having/exhibiting similar cognitive function and abilities to mammals.

Sophisticated Cognitive Functions

Cognitive abilities of birds have been investigated in a different light since the 1950s and appreciated as more complex as previously assumed (Jarvis et al., 2005).

Among these studies, Marler (1955) showed that chaffinches produce a range of different sounds/calls, which are adapted to different scenarios and functions, such as female attraction, rival male repelling, predator confusion and danger communication. Other research in pigeons showed that they can form higher order concepts to categorize objects (Lubow, 1974) and different styles of paintings (Monet vs Picasso) (S. Watanabe et al., 1995). Moreover, they can communicate using visual symbols (Epstein et al., 1980), can learn and store up to several hundred visual patterns over long periods of time (Fersen and Delius, 1989) and can lie (report incorrectly) to increase reward (Lanza et al., 1982). The latter was also shown in two species of flycatching birds, where they deceptively used alarm calls, i.e. 'cried wolf' (Munn, 1986). Another cognitive skill observed in magpies is object permanency, which they develop early on to use for food-storing purposes (Pollok et al., 2000). Magpies are also reported to recognize themselves in the mirror (Prior et al., 2008). Additionally, skills like sound localization in owls (Knudsen, 2002) and vocal production and learning (Nottebohm, 1972) in songbirds, parrots and hummingbirds are present. Parrots can even learn human words and communicate through them.

A multitude of studies have focused on what is reported to be the family with some of the most intelligent bird species, *corvids* (jays, crows). They exhibit sophisticated cognitive skills, some of which were previously reported only in primates. For instance, caledonian crows have an exceptional purpose-oriented (food retrieval) tool manipulation/modification capacity (bending wire into hooks) (Weir et al., 2002). Rook, Eurasian jay, and New Caledonian crow show as good a performance in a causal cognition task (Aesop fable), as seven-year-old children (Logan et al., 2014). Moreover, corvids have good spatial memory (Balda and Kamil, 1992), have episodic-like long term memory (spatiotemporal) (N. S. Clayton and Dickinson, 1998), adapt food-storing strategies (Emery and N. Clayton, 2001), generalize to guide behavior (Smirnova et al., 2015; Veit and Nieder, 2013) and can plan for the future (Raby et al., 2007).

Some of the cognitive skills reported above, such as self recognition in the mirror, episodic memory or future planning were thought to be present sparsely in mammals, or only in humans. However, not all birds, just like not all mammals have such sophisticated abilities. Taken together, this section aims to present the diverse and sophisticated cognitive skillset birds possess. Of particular importance for this manuscript is song production and learning, but also temporal representation (past and future).

Anatomy

Although they exhibit comparable capabilities, the anatomical structures of mammals and aves differ significantly. For instance, the architecture and organization of the avian telencephalon is quite different, appearing mistakingly more primitive, so much so that birds were previously though to only possess a hypertrophied striatum (Edinger et al., 1903). However, more recent neurochemical, histological,

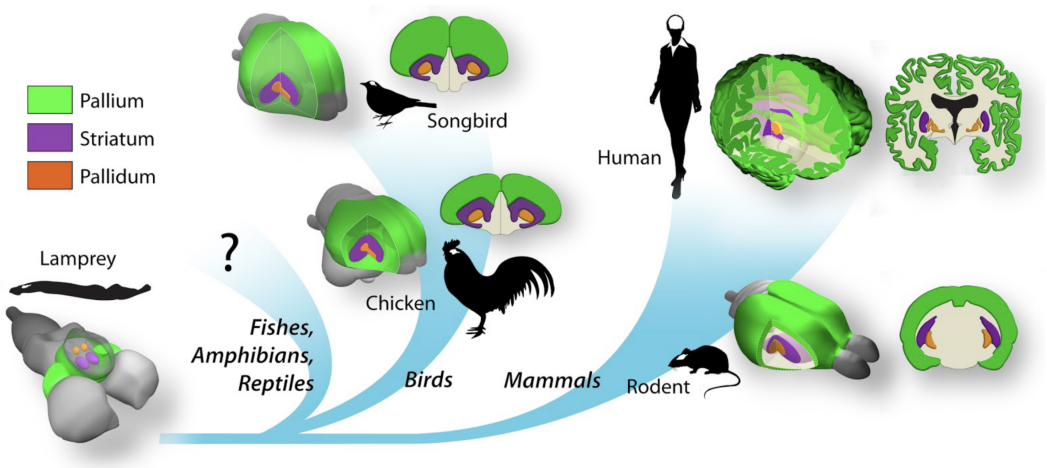


Figure 2.1: **Comparative evolution of the striatum and pallium in vertebrates.** The colormap on the left reflects the presence of each of the three domains, pallium, striatum and pallidum. From the figure, it is clear that the ratio of the brain mass devoted to the pallium increase in parallel in various vertebrates' taxa. This figure was taken to illustrate subsection 2.1 from Boraud et al. (2018)

embriological and genetic evidence (Reiner et al., 2004) have shown the presence of pallium, which is not laminated, but instead organized in nuclei. Based on what is observed behaviorally, it appears that a laminated, 6 layered cortex (as we humans possess) with pallial-cortical folding is not necessary for higher cognition and a simpler nuclear pallial organization (Güntürkün, 2011; Jarvis et al., 2005) could be and is also sufficient. Despite the distinction in presence and absence of lamination, the microarchitecture (dopaminergic projections, neurochemical, functional and electrophysiological similarities between prefrontal cortex and nidopallium caudolateral) and the allometric (relative size of forebrain) properties of the mammalian and avian forebrains are remarkably similar (Güntürkün, 2011).

In the light of more information about the avian anatomy, a group of researchers (Jarvis et al., 2005) presented a new nomenclature for the avian telencephalon. It concludes that the avian telencephalon is organized into three main, developmentally distinct domains: pallial, striatal (medial and lateral dorsal; ventral: nucleus accumbens and the olfactory tubercle) and pallidal domains (Jarvis et al., 2005). The pallium is organized into four subdivisions, hyperpallium, mesopallium, nidopallium and arcopallium and as in mammals, it comprises up to 75 percent of the telencephalic volume (measured on the sagittal series of pigeon and zebra finch brain sections) (Jarvis et al., 2005). The presence of such structures and such investigations suggest a convergent evolution (Boraud et al., 2018; Güntürkün, 2011) (Fig. 2.1) and significant similarities between birds and mammals.

While the evolution of these two classes has occurred independently (Jarvis et al., 2005) as their evolutionary lines separated almost 300 million years ago (Güntürkün, 2011), there is clear evidence of convergent evolution [1], both in cognition and neural architecture (Güntürkün, 2011).

2.2 What do songbirds teach us about vocalizations?

In the sections above we advocated for birds being a good model to study different aspects of cognition. Here, we try to convey that some birds, in particular, provide an invaluable model to study verbal social communication, vocal production and vocal learning. As a general introduction, birds are categorized into nonpasserine (chickens, ducks, eagles) and passerine (perching) birds. Passerines are futher divided into oscines (songbirds) and suboscines (manakins, flycatchers), where only oscines (sparrows, robins, finches, corvids, cowbirds, lyrebirds) are vocal learners, making them particularly interesting in studying vocal learning. Oscines can be open-ended or close-ended learners. This is defined, respectively, by the absence or presence of a critical period of plasticity in these birds. For instance, in closed-ended learners, such as zebra finches, learning can occur only within a particular (critical/ sensitive) period. Beyond this period, close- ended learners cannot learn any new song elements and continue to produce highly stereotyped vocalizations throughout their lives. On the other hand, open-ended learners, such as nightingales, European starlings and canaries, can learn new songs and vocalizations throughout their lives. In the sections below, we will focus on oscine birds, and more particularly, on zebra finches.

2.2.1 Vocal production

There are two main groups of vocalizations produced by oscine birds: calls and birdsong. The first, are brief and thought to be mostly innate, although learning/modification is possible in some species (Marler, 2004). They are considered means of social communication that convey a vast range of messages, including maintenance of social group coherence (contact and separation calls, *Stack* and *Tet*), flight, food, agression (*Wsst*), alarm (*Thuck*, mobbing and distress), breeding, mate attraction, territory defense etc. (Marler, 2004). They can also be combined in different ways and slightly modified to adapt different information.

On the other hand, birdsong is a set of ordered strings of sounds separated by brief silent intervals and lasts from a few seconds to many tens of seconds (A. J. Doupe and Kuhl, 1999). Like human speech, birdsong is a complex acoustic signal, produced by the flow of air during expiration through the tight coordination of vocal muscles. It is composed of notes or 'elements', which combine to form

[1] Convergent evolution describes the acquisition of the same biologic trait in unrelated lineages of organisms due to a similar selection pressure. (Güntürkün, 2011)

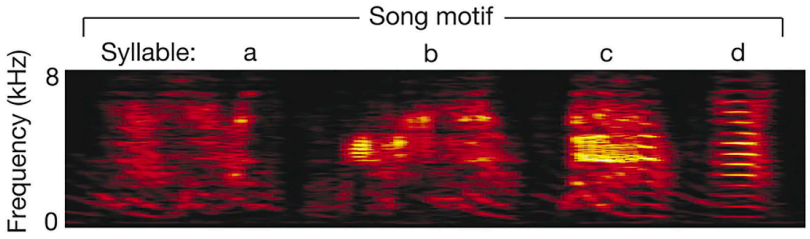


Figure 2.2: **Spectrogram of a birdsong, showing also the motif organization.**

syllables, further combining to build phrases or 'motifs'. A motif refers to a stereotyped sequence within a song (Fig. 2.2). The elements of birdsong rely on different timescales, with notes and syllables on the order of tens of milliseconds and syllables and phrases on the order of hundreds of milliseconds. These definitions are suitable for birds with highly stereotyped songs such as zebra finches.

However, they do not fully apply to other birds with more complex organization, such as bengalese finches and canaries. For instance, bengalese finches songs' are composed of notes, which organize in chunks that are not stereotypical, i.e. a note can be found in different chunks and some notes can be repeated multiple times. Furthermore, these chunks are sung in a probabilistic manner. On the other hand, in canaries, songs are organized in phrases, composed of trilled repetitions of single syllables. The transitions between different phrases relies on long-range complex syntactic rules.

Despite the particularities of different species, taken together, the similarities with speech production, temporal control, vocal muscle coordination and sensorimotor learning indicate similar neural mechanisms involved in speech and birdsong. While comparable to speech, stereotypical birdsong (zebra finch) is likely not comparable to language (Brainard and A. J. Doupe, 2002) in terms of the boundless, flexible and complex meaning language can convey. Birdsong typically conveys information regarding the bird's species, ancestry, age, identity and reproductive fitness. It is used in courtship and in intrasexual contexts, to repel rival males (Searcy and Andersson, 1986).

As stated above and as it will be detailed below, due to the similarities with speech, songbirds and in particular zebra finches with their simple syntax and song stereotypy are the most studied laboratory model to study vocal production and learning. The male zebra finch is a close-ended learner, with a highly stereotypical song, spanning 0.5 to 1 second, and consisting of repetitions of a single motif (Immelmann, 1969), composed of 3-7 syllables. Whereas only males can sing, both sexes engage in calling for communication.

2.2.2 Vocal learning

Vocal learning is a rare trait among the taxa, but it is widely common in about 4000 species of songbirds. Moreover, birdsong has a vocal learning process bearing an uncanny resemblance to speech acquisition in humans. These similarities lie on the vital importance of learning for song production, the necessity of both sound perception and production (sensory and sensorimotor period) during learning and the presence of a critical period in some species, like zebra finches (A. J. Doupe and Kuhl, 1999). Songbirds and humans learn how to sing and speak, respectively in a simplified two-staged process, composed of a sensory and a sensorimotor phase.

The sensory phase is characterized by listening to the tutor's song and building a memory of the song template. It is restricted to a sensitive (critical) period that spans a few weeks, it occurs quite quickly (with only a few minutes exposure to the tutor's song) and could result in a long- term memory of the tutor song. Due to the presence of an innate predisposition toward certain songs, usually the ones from the same species and closer relatives (Mooney, 2009a), there is certain variability in how this phase evolves (how quickly and how accurately). In the absence of tutor exposure, the critical period (Eales, 1985) is postponed. If the tutor continues to be absent, abnormal, 'isolate' songs emerge (Marler, 1970), but in 3-4 generations, these 'isolate songs' turn to species specific songs (Fehér et al., 2009b), implying a species- specific genetic predisposition and a possible innateness of the song.

In the sensorimotor phase, the bird starts producing vocalizations and relies on auditory feedback to match its own song to the tutor's song template. In zebra finches, as shown in Fig. 2.3, the sensory period, lasting until 60 days post hatch (dph) and the sensorimotor (30 - 90 dph) one overlap and further exposure is granted. In the sensorimotor period, the bird goes through two substages, the subsong and the plastic song, where the first resembles child babbling, with no apparent structure and the latter is a highly variable, highly plastic vocalization, comprised of phrases and becomes more and more similar to the tutor song every passing day. If this learning stage is compromised, for instance by post-exposure deafening, the bird is not able to learn the tutor song (Konishi, 1985). This is both due to the need for tutor sensory input and due to the need of sensory input coming in the form of a feedback from the bird itself.

If the first two stages remain uncompromised, song variability gradually reduces until vocal learning is finalized with the crystallized song, which is a stereotypical vocalization containing very little variability. The bird continues to receive auditory feedback to maintain the song similar to the template. When it is deafened after learning, there is visible song deterioration (K. W. Nordeen and E. J. Nordeen, 1992). Additionally, it has been shown that behavioral modifications, targeting pitch (Andalman and M. S. Fee, 2009) and duration (Ali et al., 2013) can be learned. Taken together, these evidence insinuate active maintenance of the song, i.e. the bird always improves the proximity of his song to the template through a continuous auditory feedback and that there is something in this auditory input

perception that prevents it from learning new song elements (Vallentin, Kosche, et al., 2016) and/or modifying existing ones. Moreover, in the context of chapter 1, we would like to emphasize that song learning comprises learning a set of features related to the tutor's song, one of which is the precise timing within and between syllables. This is what we referred to as pattern timing, in chapter 1.

Intrinsic constraints

If exposure to a tutor is what drives song learning, does this mean any song can be learned? There appear to be certain species specific constraints, some of which are related to respiratory circuits and some to genetic predisposition. A series of experiments have investigated the two, and observed that when songbirds are exposed to their parent's song, a conspecific tutor's song and a different species song, there are differences in the learning outcome (Konishi, 1985; Marler, 1997; Marler and S. Peters, 1977). Another important evidence is that isolate birds (birds without tutors) living in a colony, after 3 or 4 generations can give rise to a species typical song (Fehér et al., 2009b). Another study has shown that while syllable coding is experience dependent, temporal gap coding is innate (Araki et al., 2016b).

Adult Plasticity

Although close-ended learners do not learn new song elements in adulthood, they still have the ability to introduce slight modifications to their song (Tumer and Brainard, 2007; Warren et al., 2011). Behavioral adaptation has been demonstrated through a Conditional Auditory Feedback (CAF) paradigm in two song features, namely pitch and duration. CAF aims to increase or decrease the duration (pitch) of a target syllable. The targeted (pitch or durations) acoustic feature is computed live while the bird is singing, and feedback is delivered accordingly, i.e. when the bird sings outside a 'desired range', white noise is given. More specifically, white noise is presented as a perturbatory stimulus every time the target syllable's performance was lower (higher, depending on the protocol) than a dynamically changing threshold.

The threshold is set by acquiring a set of baseline distributions prior to starting the protocol and is changed based on the mean duration/pitch across trials. The initial threshold is set at approximately a third of the distribution. When subjected to such a protocol, the birds try to evade white noise by altering their songs and after several trials across days, the bird shifts the duration (pitch) distribution by about 2 to 3 standard deviations. In the duration protocol (Ali et al., 2013; Pehlevan et al., 2018), a specificity to the target syllable in the acquired modification was observed, i.e. no change was observed on any of the other syllables.

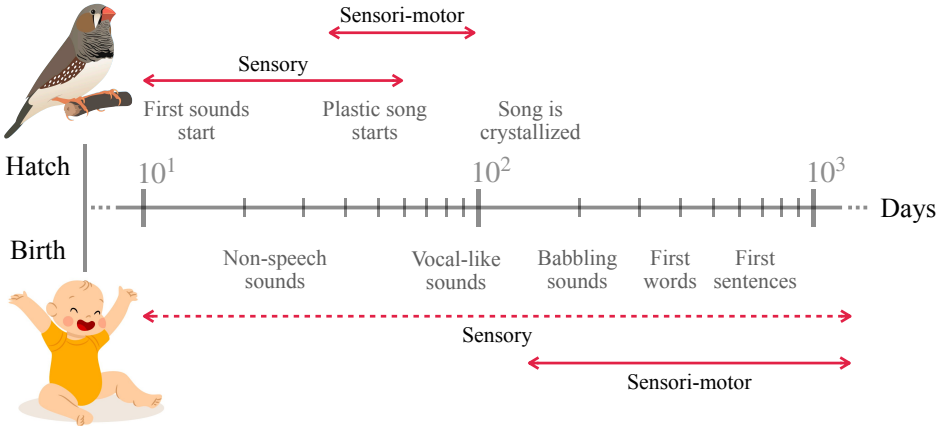


Figure 2.3: **Stages of vocal learning in zebra finches and humans, shown in a logarithmic scale.**

2.2.3 Parallels with Human Speech

As briefly stated above, there are similarities between birdsong and human speech. Broadly, they are both complex acoustic signals, composed of three main dimensions: time, frequency and intensity. During vocal production, speech and birdsong rely on air expiration generating a complex waveform, through the vocal cords in the larynx and the syrinx, respectively. In humans, the waveform is then modified (filtered) by the rest of the vocal tract (Stevens, 1994), whereas in birds there is evidence that the width of beak opening has a similar role and affects sound frequency (Westneat et al., 1993). The basic vocalization units (phonetic unit and note respectively) are grouped to form 'basic processing units', the syllables or phrases, which are then placed in order and separated by brief intervals (gaps). Such order is referred to as song syntax in the songbird literature and remotely corresponds to grammar. Lastly, they both rely on timing on different timescales, the one of the building units and the one of the sequence (motif, phrase, sentence).

Vocal learning in both taxa also presents with a lot of similarities. For instance, both taxa have critical periods, during which they learn through imitation of an adult tutor. They heavily rely on auditory feedback and several stages of learning (sensory, sensorimotor and motor) are present in both, as shown in Fig. 2.3. Moreover, the same gene, FoxP2, is linked to developmental vocal learning in humans (Lai et al., 2001) and songbirds (Haesler et al., 2004).

Several intrinsic constrains and predispositions in vocal learning have been observed in songbirds and humans (A. J. Doupe and Kuhl, 1999). Although some of these are attributed to the physical apparatus, the repertoire of sounds is still quite large, i.e. a wide range of vocalizations can be generated. Hence, other mechanisms could also be involved, such as reported innate sensory recognition and learning preferences (A. J. Doupe and Kuhl, 1999). Moreover, similar to *isolate* songs produced by songbirds, children exposed 'pidgin' languages and deaf chil-

dren not exposed to sign language (Goldin-Meadow and Mylander, 1998), develop some elements (words or gestures) and order them somewhat conforming to a rudimentary grammar.

Lastly, a social factor is also present. In the language literature, it has been referred to as a social gating mechanism, driving children to only learn sounds coming from human sources (Kuhl, 2004; Kuhl, 2007). Zebra finch, as a highly social songbird, also relies on social features (who is feeding them) to choose their tutor. Interestingly, they can override their innate conspecific song preferences and learn the song of a bengalese finch, if they are the ones feeding them (Immelmann, 1969).

2.3 Neural Networks for Song Perception and Production

Given the behavioral similarities detailed above, the birds present a good model to investigate neural dynamics underlying different tasks, which may present with seemingly more complex brain network in mammals and/or commonly used animal models. In particular, in mammals, upon studying a sensorimotor skill, such as speech production and its learning, complex interactions between cortical areas (primary and secondary motor cortex in rodents, motor, premotor and supplementary motor areas in primates) related to many different sensorimotor associations and motor skills are present (Dum and Strick, 2002). Conversely, in birds, the sensorimotor skill of song production and its learning has a dedicated set of interconnected brain nuclei known as the "song system", coordinating patterned breathing and vocal muscle activity necessary for vocalization. The simplicity of this system makes them an outstanding model to study the neural mechanisms of vocal learning and more generally, of sensorimotor learning (Brainard and A. J. Doupe, 2002).

The neural circuitry underlying this mechanism has three main building blocks: (i) the auditory pathway, responsible of the sensory acquisition of the song from tutor exposure and of auditory feedback to the system to maintain the crystallized song (Brainard and A. J. Doupe, 2002), (ii) the Anterior Forebrain Pathway (AFP) entailing initial song learning, subsong generation, maintenance and variability in adult song and (iii) the motor pathway responsible for the production of the adult song. The last two constitute the 'song system'.

2.3.1 Auditory Pathway

The communication between the auditory system and song system is crucial for a vocalization so heavily reliant on auditory feedback, such as birdsong. As shown in Fig.2.4, auditory input enters the song system through nucleus interfaciallis of the ndiopallium (NIf) and HVC. The anatomical connections, could help understand the sensorimotor integration during song learning, and moreover during maintenance and adaptation. For instance, area caudomedial nidopallium (NCM)

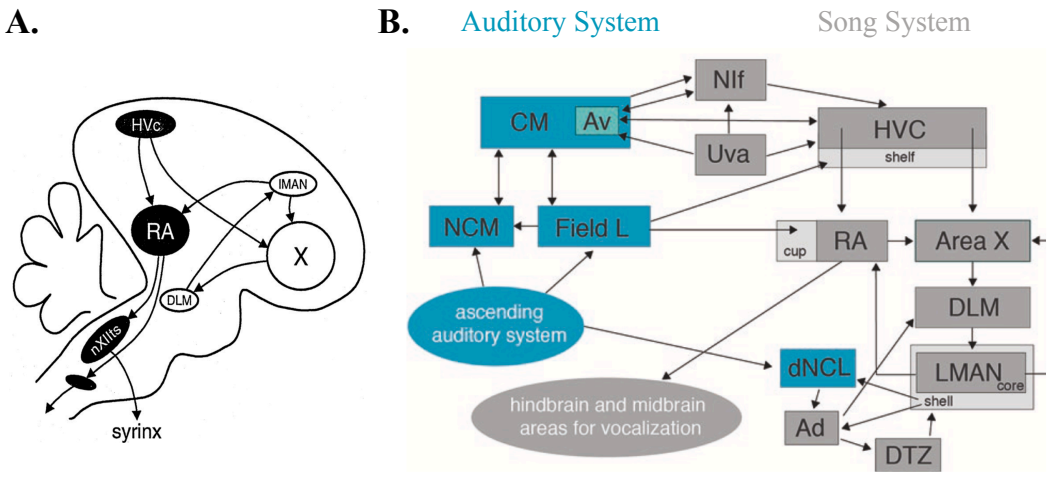


Figure 2.4: **Anatomy of the Auditory and Song System in Zebra Finches** (A) Simplified schema of the song system. (Luo and Perkel, 1999) (B) Summary of (most of) the neural substrates of song perception and production (Sakata and Yazaki-Sugiyama, 2020). Some parts of the ascending auditory pathway and the cerebellum; midbrain and hindbrain catecholaminergic areas are not included. The abbreviations used are detailed in Table 2.1.

is hypothesized to be affiliated with the formation and storage of tutor song memories, providing a template for comparison during song learning. On the other hand, nucleus avalanche (Av) transmits motor signals to the auditory system. Below, we briefly explain the anatomy and the homologies with mammals. The abbreviations used are detailed in Table 2.1.

Auditory information comes from the ascending auditory system (cochlea- hindbrain cochlear nuclei - auditory nerve- auditory midbrain) through the auditory thalamus nucleus Ov, which projects to Field L (analogous to the primary auditory cortex) and then to the higher auditory areas CM (caudal mesopallium) and NCM (caudo-medial nidopallium). It reaches the song system through NIf and HVC. Moreover, in the ventral CM, nucleus Avalanche (Av) receives input through the auditory thalamic nucleus uvaeformis (Uva) and has bidirectional connections with both NIf and HVC (Akutagawa and Konishi, 2010). A loop between the auditory and motor areas is formed. In this loop, HVC_{Av} neurons, have an important role in integrating motor-related signals with activity in the auditory system. Experimental evidence with ablations have confirmed the existence of this link during learning and adaptation of song timing (Roberts et al., 2017). The detailed pathways and their respective intersection with the song system are illustrated in Fig. 2.4.

Several of these structures are homologous to auditory structures in mammal. For instance, the ascending auditory pathway of songbirds is homologous to that observed in other vertebrates: *cochlea- hindbrain- midbrain-thalamus-cortex*. The avian auditory midbrain is homologous and functionally analogous to the mam-

malian central nucleus of the inferior colliculus (ICc). Moreover, the avian auditory thalamus Ov is homologous to the ventral division of the mammalian medial geniculate body. Field L2, which receives the most thalamic input in songbirds is homologous to layer IV of the primary auditory cortex in mammals. Other higher-order auditory areas included in the diagram are referred to as higher cortex (Woolley and Portfors, 2013).

Abbreviation	Structure name
Ad	Dorsal arcopallium
AFP	Anterior forebrain pathway
Area X	vocal portion of the basal ganglia
Av	Avalanche, a ventral region in CLM
CLM	Caudolateral mesopallium
DLM	Dorsal thalamic zone
DTZ	Dorsal thalamic zone
HVC	Used as proper name
HVC_{Av}	HVC cells that project to Av
HVC_{RA}	HVC cells that project to RA
HVC_{X}	HVC cells that project to X
LC	Locus coeruleus
LMAN	Lateral magnocellular nucleus of anterior nidopallium
NCM	Caudomedial nidopallium
NIf	Nucleus interfacialis of the nidopallium
Ov	Nucleus ovoidalis
PAG	Avian periaqueductal gray
RA	Robust nucleus of the arcopallium
Uva	Nucleus uvaeformis
VP	Ventral pallidum
VTA	Ventral tegmental area

Table 2.1: Abbreviations used to refer to avian brain structures.

2.3.2 Anterior Forebrain Pathway (AFP)

The anterior forebrain pathway (AFP) is a basal ganglia-thalamic- cortical circuit that indirectly connects HVC and RA. It contains the avian basal ganglia nucleus Area X, the medial portion of the dorsolateral thalamic nucleus (DLM), and the lateral magnocellular nucleus of the anterior nidopallium (LMAN). In oscine songbirds, AFP is essential for vocal learning and plasticity (A. Doupe et al., 2005). The hypothesized functions can be grouped into three, comparison with the song template, generation of motor variability, and computation of error or instructive signal.

Area X could be involved in template memorization or matching. This is supported by evidence showing that suppression of FoxP2 in area X during tutor exposure and sensorimotor learning, impairs imitation (Haesler et al., 2004). It also

is at the receiving end of a lot of dopaminergic projections from VTA and substantia nigra pars compacta (SNc), which implies a role for dopamine in song learning (Doya and Sejnowski, 1998), consistent with its role in motor learning in mammals.

On the other hand, the function of LMAN mostly lies on variability generation across the sensorimotor period and adulthood. It is considered the essential premotor nucleus for babbling (Aronov et al., 2008), as the HVC RA pathway is not strongly developed yet and exploration by LMAN drives its further consolidation. In subsong, only 10 percent of HVC_{RA} neurons burst rhythmically and they are mostly aligned to syllable onset (Okubo et al., 2015).

The role of AFP in song production grows lesser as crystallization occurs and it seems to be only important for the minimal variability present in adult song. For instance, area X lesions during the sensorimotor period, cause a failure in song stabilization and LMAN lesions cause the song to stabilize prematurely (Bottjer et al., 1984; Ölveczky et al., 2005; Scharff and Nottebohm, 1991; Sohrabji et al., 1990). Conversely, AFP lesions in adulthood have less disruptive effects. Experiments using a CAF paradigm, suggest that LMAN provides biased instructions, to avoid vocal errors (Andalman and M. S. Fee, 2009), consistent with a role in auditory feedback-dependent song plasticity. This is supported by the absence of song deterioration in deafened birds with a lesioned LMAN (Bottjer et al., 1984; Scharff and Nottebohm, 1991).

In figure 2.4, the connections within the AFP and the reset of the song system are illustrated. The initial projections come to the AFP from HVC through the X projection neurons. Then, the GABAergic neurons in area X send excitatory and inhibitory projections to the medial nucleus of the dorsolateral thalamus (DLM). This then sends projections to the lateral portion of the magnocellular nucleus of the anterior nidopallium (LMAN). LMAN send projections both to area X and RA.

The AFP is homologous to the basal ganglia-thalamic-cortical circuit in mammals, where area X is homologous to the mammalian basal ganglia (A. Doupe et al., 2005) and LMAN is functionally analogous to parts of the mammalian frontal cortex (Pfenning et al., 2014; Reiner et al., 2004). Moreover, the thalamic DLM neurons show thalamic properties and their projections to LMAN are glutamatergic, corresponding to the mammalian thalamocortical projections.

2.3.3 Motor Pathway

The motor pathway is a direct pathway connecting HVC to Robust Nucleus of the Arcopallium (RA). RA provides motor input to motor neurons in the tracheosyringeal portion of the hypoglossal motor nucleus (XII). The syringeal muscles are innervated and sound is produced. It also sends input to the respiratory premotor neurons in the ventral respiratory group, which contains the expiratory nucleus retroambigualis and the inspiratory nucleus parambigualis. These inputs, project back to the thalamus and HVC (Brainard and A. J. Doupe, 2013; Schmidt et al., 2011). The information transmission delay from HVC to RA is $\sim$5 ms (R. Hahnloser et al., 2002) and the one from HVC_{RA} to the output $\sim$ 15-20 ms (A. A. Kozhevnikov and

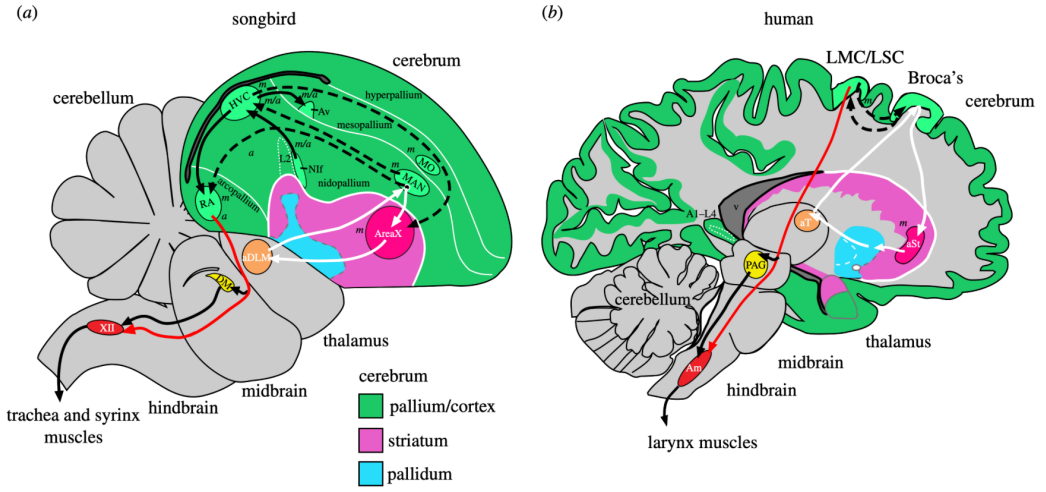


Figure 2.5: **Brain pathways involved in birdsong and speech production** In black arrows the motor pathway, in white arrows, the AFP. In dashed black, the connections between the AFP and motor pathway, and in red arrows the direct projection from vocal motor cortex to vocal motor neurons. *a* indicates auditory, *m* motor and *m*/*a* both auditory and motor neural activity. Some connections in the human brain are proposed based on known connectivity of adjacent brain regions in non-human primates. Taken from Pfenning et al. (2014). [2]

M. S. Fee, 2007) or 35 ms (Ali et al., 2013). As briefly mentioned above, this pathway develops later compared to the AFP and is not present during babbling and subsong. It only appears during plastic song. As its influence gets stronger and as input from LMAN grows weaker, there is a decrease in song variability until song crystallization is reached. The role of this pathway in song production is crucial, as lesions either severely disrupt it, or abolish it (Nottebohm et al., 1976; Scharff, Kirn, et al., 2000; Simpson and D. Vicario, 1990).

As we will detail below, HVC_{RA} neurons burst only once during the ∼1 second motif, with a brief duration of ∼10 ms. In contrast RA neurons fire more frequently, up to 10 times per motif, which suggests multiple HVC neurons send converging input to multiple RA and single HVC neurons diverging input to multiple RA neurons (Mooney, 2009a). RA bursts are on average uncorrelated to spectral features of the song (Leonardo and M. S. Fee, 2005). However, there is likely some structure there as well, as birds exposed to the same tutor were shown to have similar patterns of bursting activity (Mooney, 2009b). Also, an important feature of RA is its position at the convergence of HVC and LMAN input, suggesting that this

[2] A1–L4: primary auditory cortex—layer 4, Am: nucleus ambiguous, aSt: anterior striatum, LMC: laryngeal motor cortex, LSC: laryngealsomatosensory cortex, MN: motor neurons, MO: oval nucleus of the anterior mesopallium, v: ventricle space.

is the area where song variability and the biased exploration during learning is induced.

HVC is functionally analogous to the premotor cortex in mammals and RA is analogous to the vocal motor region of the primary motor cortex in mammals (Fujimoto et al., 2011; Hara et al., 2012). An illustration of the homologies and analogies in motor and anterior forebrain pathways between songbirds and humans is presented in Fig.2.5.

2.4 Anatomical Substrates of Motor Timing In Songbirds

2.4.1 A Local Motor Timing System: Premotor Nucleus HVC

HVC (proper name), formerly called the hyperstriatum ventrale, pars caudalis and then the high vocal center, is an avian nucleus functionally analogous to the premotor cortex in mammals. HVC receives auditory input and lies at the intersection of the motor and anterior forebrain pathway. Its position justifies its role in song perception (Margoliash, 1997), production (Nottebohm et al., 1976) and learning (Roberts et al., 2017). Interestingly, song perception (of the birds own song) and production incite similar firing patterns in HVC (Mooney, 2000; J. Prather et al., 2008). The implicated neurons (HVC_X) are referred to as auditory-vocal mirror neurons (J. Prather et al., 2008) and are similar to the visual- motor mirror neurons first in the premotor cortex of macaque monkeys (Gallese et al., 1996). They are thought to be important in social learning.

In the context of learning, in 'the AFP template comparison' hypothesis, HVC is proposed to be a source of live auditory input (Mooney, 2004; J. Prather et al., 2008) sent to area X, where it is compared to the song template stored in AFP. Other hypotheses of learning, suggest HVC provides a centrally generated efference copy that serves as an internal prediction of the feedback (Troyer and A. J. Doupe, 2000), or gives a global time representation, as a timing generator (M. Fee and Goldberg, 2011). HVC's role in learning is supported by multiple experimental studies, showing high dopamine concentrations in HVC while listening to a live tutor and inhibition of song learning upon their manipulation (Tanaka et al., 2018).

The main hypothesis we address in the thesis is the one related to HVC as an area important to song production and learning, in the temporal domain (Ali et al., 2013; R. Hahnloser et al., 2002; Pehlevan et al., 2018). The first evidence of such a role was provided by focal stimulation (E. Vu et al., 1994) in HVC, which disrupted song sequencing. Moreover, brief neuronal bursts in a particular subpopulation (HVC_{RA}), time-locked to the song were reported (R. Hahnloser et al., 2002). To follow the trail of HVC's role in motor timing, a cooling study was performed and as hypothesized, slowing down of both sequence and song was observed (Andalman, Foerster, et al., 2011; Long and M. S. Fee, 2008). Hence, HVC was positioned as local network encoding temporal control. However, other studies have hypothesized a

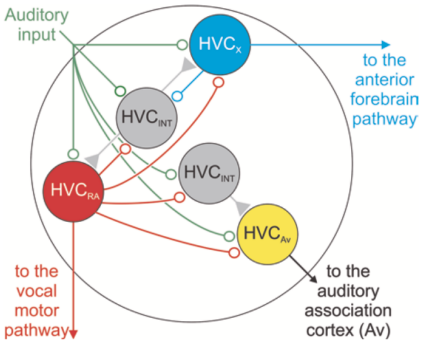


Figure 2.6: **The HVC Microcircuitry:** The connections within HVC and their outside projections. Taken from (Murphy et al., 2020)

distributed network, where Uva, brainstem and RA could be organized (in a loop) to control motor timing (Hamaguchi, Tanaka, et al., 2016).

HVC Microcircuitry

There are four main types of cells in HVC, HVC_{RA} neurons, HVC_X neurons, HVC_{Av} and interneurons (Fig. 2.6). As the name implies they project to RA, X (analogous to basal ganglia) and nucleus Avalanche respectively. Below, we detail important information for each of these subpopulations.

HVC_{Av} neurons project to nucleus Avalanche, which provides feedforward and feedback information between the auditory pathway and the song system, through bidirectional connections with HVC. HVC_{Av} cells exclusively receive input from HVC_{RA} cells, suggesting a possible role for them in song-timing. Ablation studies in juveniles have revealed problems with imitation of the song as a sequence and less visible problems in individual syllable learning. In deafened adults with Av lesions, temporal features were not affected, in contrast with spectral ones. Moreover, HVC_{Av} cell lesions interfere with the rate of learning and recovery (return) of normal song timing in a CAF paradigm (Roberts et al., 2017). Together, these findings reveal an important role of HVC_{Av} neurons in learning and adaptive modification of song timing.

HVC_X neurons are another class of neurons in HVC, which project to area X and furthermore activate the AFP, a basal ganglia pathway necessary for vocal plasticity. They fire quite regularly with moderate spike-frequency adaptation in response to a tonic input (Mooney and J. F. Prather, 2005). These neurons are considered *mirror neurons*, showing similar responses to syllable perception and production (J. Prather et al., 2008). During singing, multiple bursts of these neurons, precede similar elements of the syllable patterns by ∼40ms, suggesting a relation with temporal features. However, there seems to be no relation between their activity and song spectral features. Moreover, no auditory responses were present

in these neurons during singing (A. A. Kozhevnikov and M. S. Fee, 2007). Hence, although previous hypotheses have implicated them in transmission of auditory information (both spectral and temporal) to the AFP during singing, more recent evidence suggests that it rather transmits information about sequence timing (A. A. Kozhevnikov and M. S. Fee, 2007).

Possibly the only inhibitory source in HVC, interneurons fire densely and relatively continually during the song. Together with X projection HVC neurons, they fire in alternating phases of a local 30 Hz rhythm (Markowitz et al., 2015). Interneurons are connected to both RA and X projecting HVC neurons (Mooney and J. F. Prather, 2005), suggesting an impact in song production, which has been shown through pharmacological studies (blockade of GABAA conductances) (Kosche et al., 2015). Vallentin, Kosche, et al. (2016) also showed that they are significantly correlated with song imitation accuracy and their maturation is performance rather than age dependant. They futher hypothesize that their maturation has an effect on the critical period, and its inhibition could allow continued (adult) learning, like in open-ended learners.

HVC_{RA} neurons are likely the most important class in HVC for song production, since they project to the motor nucleus RA. However, as shown in Fig. 2.6 and detailed above, they lie inside a circuitry, so their activity is related to the activity of other neurons in HVC. HVC_{RA} emit bursts of one to several action potentials in response to a tonic input (Mooney and J. F. Prather, 2005) and burst only once in a motif during singing. They fire time-locked to the song, always at the same time across multiple renditions, with a jitter of less than a millisecond. The bursts contain 3-6 spikes and a duration of 10 ms (R. Hahnloser et al., 2002). This is illustrated in the experimental raster plot by R. Hahnloser et al., 2002, shown in Fig. 2.7. Moreover, their activity spans the whole song, creating a neuronal mapping thereof (Lynch et al., 2016).

2.4.2 A Distributed Motor Timing System in Songbirds?

In the section above, we have presented several experimental findings suggesting a role of HVC is song production and learning, focused on the temporal domain. Moreover, we have detailed the roles of individual populations, which support such a function. However, although HVC is mostly accepted as a timing-related area, how much of this role (motor timing) can be attributed solely to this nucleus, remains an open discussion. Other upstream areas, could be affecting HVC input and be the ones actually 'driving' time. Since during singing HVC is only affected by motor input, the possible contributing areas are limited and the proposed distributed timing mechanisms focus on the roles of NIf or Uva on HVC.

NIf (nucleus interface of the nidopallium) is a significant source of auditory input to HVC and also plays an important role during sleep (R. H. Hahnloser and M. S. Fee, 2007). It is located in Field L, which is the avian homologous of the primary auditory cortex. NIf precedes introductory notes by several milliseconds (McCasland, 1987) and hence may have a premotor role. A study, which pharma-

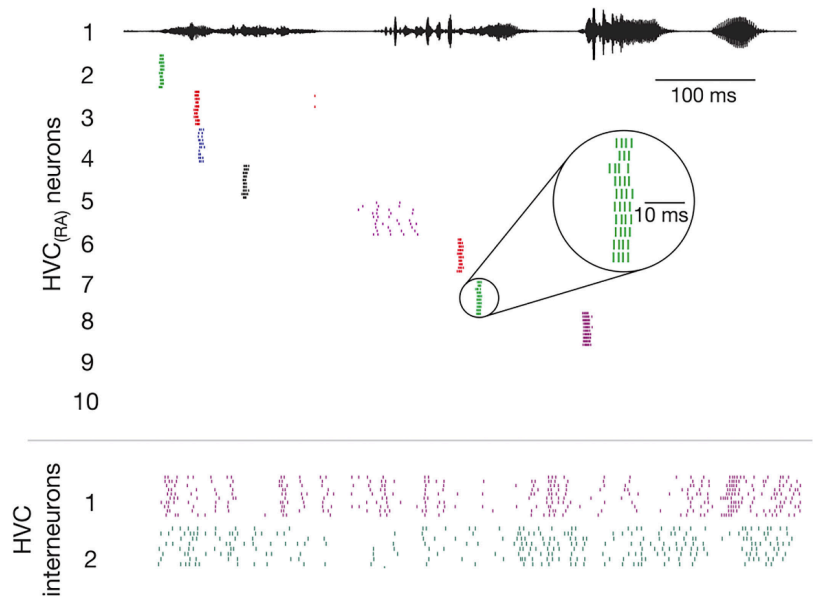


Figure 2.7: HVC_{RA} **(N = 8) time-locked activity to the song and** HVC_I **(N = 2) neurons.** Raster plots of several song renditions in a single bird. Each row is one rendition and 10 renditions are shown for each neuron. The activity is aligned to the onset of the nearest syllable. Taken from R. Hahnloser et al. (2002).

cologically inactivated NIf (Naie and R. Hahnloser, 2011) has shown that the effect is different in juvenile birds singing subsong, those singing plastic song and adult birds. In the latter, there is only a transitory effect, which decreases song stereotypy (by 10%) and destabilizes syllable sequences (Cardin et al., 2005; Naie and R. Hahnloser, 2011). These results suggest that although present, NIf is not essential for song production in adults.

On the other hand, several hypotheses have been proposed on the role of Uva in motor timing during song production as a possible higher structure, guiding the HVC sequence. More distinctively, Elmaleh et al. (2021) presented three hypotheses of Uva generated input affecting sequence production, (i) in a continuous pattern, (ii) in a more discrete pattern, i.e. at every syllable and (iii) not significantly, as local HVC connections are able to maintain sequence generation independently. In that study, in anesthetized birds, where sleep replay of the sequence is observed, they lesion Uva and observe that the sequence remains unaffected, revealing important information of network connectivity, although only during sleep. In another more recent study (Moll et al., 2023), Uva driven neurons were identified in HVC during singing. They are always active at syllable initiation or/and ending and constitute about 15% percent of the HVC_{RA} neurons. This suggests that Uva may be an upstream control of HVC. Due to its temporal pattern, it could likely have a synchronizing activity on the two HVC. This is consistent with previous

hypotheses, derived from Uva lesions (Coleman and E. T. Vu, 2005; H. Williams and D. S. Vicario, 1993) and stimulations (E. Vu et al., 1994).

2.4.3 Plasticity of Motor Timing

Upon discussing the roles of different pathways, we have heavily relied on AFP as a song *learning pathway*. As previous evidence have suggested, in a CAF experimental paradigm targeting syllable-level pitch adaptation, this pathway was indeed involved. This was clearly demonstrated through lesions in LMAN and area X, which compromised song plasticity. However, interestingly, this was not the case when duration was targeted for modification (Ali et al., 2013). Lesions in LMAN, MMAN and area X had very little to no effect on syllable modification, suggesting a different learning mechanism is present in the temporal domain.

To investigate this further, the authors performed recordings in HVC during the CAF protocol, which revealed that HVC activity reflects the temporal changes observed in the song (but not the spectral ones). Their recordings show a local stretching/shrinkage of the temporal structure at the level of a syllable targeted for modification. The duration of the song segment and its modification is a function of the speed of the activity propagation. Taken together with the very small effect of AFP, these findings suggest that the synaptic changes driving the change in activity propogation and furthemore interval length, are likely occurring at the level of HVC itself.

Hence, the role of HVC in motor timing and its plasticity provides an excellent framework to study the robustness and flexibility of motor timing. These features are critical in the performance and refinement of many motor skills (including, but not limited to vocal production), not only in songbirds, but also in humans.

3 Models of Subsecond Motor Timing in Neuroscience

In the first two chapters, we motivated the broader question and illustrated the characteristic properties of the biological model appropriate to address it. To briefly summarize, in chapter 1, we highlighted the importance of motor timing on the scale of tens to hundreds of milliseconds for a wide range of tasks, ranging from simple eyelid conditioning to complex tasks like playing the piano or speech production. Moreover, we defined timing tasks as those that require some sort of timing device to solve, like interval/duration discrimination, production and reproduction etc. Among those tasks, speech production is an example of pattern timing, which involves two different scales, one at the level of the syllables and one at the level of the sentence (sequence). Furthermore, in Chapter 2, we presented a biological model, songbirds as an example of vocal production very similar to humans'. As such, it highly depends on precise timing, encoded in the premotor nucleus HVC, and it is studied by a number of experimental investigations. Behavioral and electrophysiological studies have revealed important information about the underlying circuitry, however, some questions remain unanswered. In this setting, the benefit of computational modeling is significant. It can add supplementary information, while being able to investigate time-consuming and/or experimentally non-testable hypotheses. Moreover, it can guide future experimental paradigms through predictions, and can help understand the underlying mechanisms. In this chapter, we will first explore the general models of subsecond motor timing and then we will focus on the ones used for HVC. Lastly, we include a section on attractors as an alternative modeling approach.

3.1 General Models

As briefly discussed in Chapter 1, timing models can be split into two broad classes (i) dedicated or (ii) intrinsic systems (R. Ivry and Schlerf, 2008). A series of models pertaining to one or the other, have been designed. These include the pacemaker

accumulator models (Creelman, 1962), the scalar expectancy theory (Gibbon, 1977), multiple oscillator models (Buhusi and Meck, 2005; Church and Broadbent, 1991; M. S. Matell and Meck, 2004; Miall, 1989), multiple neuron models (Grossberg and Schmajuk, 1989; Moore et al., 1989; Staddon and Higa, 1999), ramping models (Gavornik, Shuler, et al., 2009; Reutimann et al., 2004; Simen et al., 2011), population clocks (D. V. Buonomano and M. D. Mauk, 1994; Liu and D. V. Buonomano, 2009) etc. Each of these classes of models has one or few representatives. We have focused only on the ones aiming to model motor timing (as opposed to perceptual timing), even though the same principles and sometimes models can apply. Table 3.1 illustrates the main representative(s) and characteristics of each of these model classes.

Pacemaker- Accumulator Model

The first models of timing on the scale of hundreds of milliseconds and seconds were pacemaker-accumulator models (Creelman, 1962; Treisman, 1963). The principle of these models is simple and similar to a man-made clock, with one pacemaker generating neural pulses and an accumulator assembling them. Such mechanism creates a linear measure of duration based on the number of accumulated ticks. The frequency of tick generation has been a subject of debate, pertaining the mechanism with which it is determined. For instance, a constant frequency would suggest a very objective/perfectly 'accurate' timing mechanism, similar to having an actual clock, which experiments and our personal experience demonstrate not to be the case. This clock is likely affected by a number of factors.

Another model in this class and by far, the most influential is referred to as the scalar-expectancy theory (Gibbon, 1977), containing (i) a clock compartment with an attentional switch, (ii) a comparison compartment between (a) the working memory, where the representation of the duration is transferred and (b) the reference standard in the long-term memory and (iii) a decision making compartment. Regarding the frequency of neural ticks, throughout the years evidence has shown that the rate of the pacemaker changes with respect to a multitude of external (stimulus intensity, pharmacological manipulations (Meck, 1983)) and internal factors (attention (Lejeune, 1998), fatigue, body temperature (Wearden and Penton-Voak, 1995), excitement, boredom, switch latencies). The standard pacemaker-accumulator models find little support in biology, because they are not very consistent with experimental findings and imply a dedicated system through a mechanism, the anatomical correlates of which remain unknown (Buhusi and Meck, 2005; M. Mauk and D. Buonomano, 2004).

Striatal Beat Frequency Model

The Striatal Beat Frequency (SBF) model, part of the multiple oscillator model class, is comprised of cortical 'time-keeping' neurons' oscillating across varied frequencies and striatal spiny neurons acting as coincidence detectors. This model

is designed to encompass the primary brain areas suggested to play a role in motor timing, specifically the frontal cortex and the basal ganglia (Buhusi and Meck, 2005; M. S. Matell and Meck, 2004). The SBF is reported to be consistent with recordings and behavioral findings (M. Matell et al., 2003) and it reproduces the scalar property of timing. However, the sequential activity observed in striatal neurons during timing tasks (Gouvêa et al., 2015) seems highly unlikely to play the role of a coincidence detector.

Multiple Neuron Models: Spectral timing/ Labeled line

The spectral timing (Grossberg and Schmajuk, 1989) or labeled line model, assumes a broad range of time constants distributed along a spectrum. This model was initially developed based on the neural characteristics of the hippocampus. However, it has also been used to model conditional response timing in the cerebellum, assuming different time constants for the granule cells (Bullock et al., 1994), or in a later model, differently timed calcium responses in Purkinje cells (Fiala et al., 1996). In this model, depending on the time interval between unconditional and conditional stimulus, different synapses are reinforced, either the ones with fast or slow membrane dynamics.

Ramping/ Climbing Model

Ramping models (Balci and Simen, 2016; Durstewitz, 2003; Simen et al., 2011) propose that time is encoded in the monotonic changes in firing rate and that actions are produced when the firing rates reach a threshold value. It is resemblant of a pacemaker-accumulator model, with tonic pulses (Paton and D. V. Buonomano, 2018). The generated patterns are highly similar to the ones observed in different areas of the brain during motor timing tasks. A large number of data shows such ramping activity in the prefrontal cortex (Niki and M. Watanabe, 1979), premotor and motor cortex (Mita et al., 2009; Murakami et al., 2014) and the parietal cortex (Jazayeri and Shadlen, 2015), suggesting that this is a highly probable underlying mechanism for motor timing. However, it is not clear whether ramping activity encodes time, or it, instead reflects motor preparation, and whether it is the output of upstream timing circuits.

Population Clock Model

Population clock models assume that time is both generated and represented (Karmarkar and D. V. Buonomano, 2007) in the dynamically changing population of neurons. First proposed in the context of the cerebellum (D. V. Buonomano and M. D. Mauk, 1994), this model suggests specific patterns of activity encode specific instances of time and trained output units recognise them to provide a readout (D. V. Buonomano and Laje, 2010). Hence, the clock is an emergent property of the network and the population clock speed covaries with the readout. This is consistent with

Class	*Models of Timing*	Mechanism	D/I
Pacemaker Accumulator	Pacemaker Accumulator Model (Creelman, 1962)	Clock (Pacemaker, Switch and Accumulator)	D
	Scalar Expectancy Theory (SET) (Gibbon, 1977)	Clock (Pacemaker, Switch and Accumulator), Memory, Comparison between the two	D
Multiple Oscillators	Connectionist Model (Church and Broadbent, 1991)	tracking the half phase of a series of oscillators (clock stage) and continually comparing the phase of these oscillators to a memory for the phases of these same oscillators on previously reinforced trials (memory stage)	D
	Beat Frequency (BF) (Miall, 1989)	Oscillations of Different Frequencies	D
	Striatal Beat Frequency (SBF) (Buhusi and Meck, 2005; M. S. Matell and Meck, 2004)	Oscillations of Different Frequencies and coincidence detectors: spiny neurons of the striatum	D
Multiple Neuron	Memory Decay (Staddon and Higa, 1999)	Multiple timescales, Gaussian activation model, Temporal decay model, coupled leaky integrators	D
	Spectral timing (Grossberg and Schmajuk, 1989), Labeled Line	neurons can respond to the CS with a broad range of time constants (hippocampus)	D
	Tapped delay line (Moore et al., 1989)	time-varying activity by arranging neurons in a chain so that each synapse adds a discrete delay to the signal	D
Ramping/ Climbing	Ramping based Models (Gavornik, Shuler, et al., 2009; Reutimann et al., 2004; Simen et al., 2011)	monotonic changes in firing rate, and that an actions are produced when the firing rates reaches a threshold value	D
Population Clock	Mauk cerebellum (D. V. Buonomano and M. D. Mauk, 1994)	Unique spatial pattern of activity in a neural network	I
	Synfire Chain (Abeles, 1982; Liu and D. V. Buonomano, 2009)	Sparse Population Clock, Feedforward Network	I

Table 3.1: **Computational Models of Timing.** The first three classes of timing are based on the review from (M. S. Matell and Meck, 2004). Further information included information comes from different reviews on motor timing in the subsecond scale, the main three being: (Addyman et al., 2016; M. S. Matell and Meck, 2004; Medina and M. D. Mauk, 2000; Paton and D. V. Buonomano, 2018). D stands for dedicated systems and I for intrinsic or distributed systems.

experimental findings (Bakhurin et al., 2017; Gouvêa et al., 2015; Long and M. S. Fee, 2008; Monteiro et al., 2023), which demonstrate temporal scaling. Hence, slowing or speeding up the temporal evolution of activity should translate to dilation or contraction of the readout behaviour.A perfect example of this occurrence is the effect of HVC cooling in songbirds (Long and M. S. Fee, 2008), which appears with a dilated song. Another study with the same cooling paradigm in the striatum (Monteiro et al., 2023) showed that cooler temperatures caused dilation, and warmer temperatures contraction, of both neural activity and time perception, judging one interval as longer (shorter) than it was when cooled (heated).

This property can also be captured by ramping models, through changing the slope of neuronal activity. Ramping and population clock models both have their advantages and shortcomings. However, due to the sequential nature of population clocks, i.e. the reliance of the later patterns from earlier ones, population clocks are considered better suited for pattern timing (Hardy and D. V. Buonomano, 2016), which also underlies speech and birdsong.

3.2 Models of HVC

One example of a sparse population clock, where each neuron is active only once during each trajectory, is the synfire chain, which is also one of the most important timing models in the songbird literature. Synfire chains were first introduced by Abeles (1982). A synfire chain is a feed-forward network of neurons composed of many layers (or pools), where each neuron in one pool excites the neurons of the next pool, and each neuron in the receiving pool is excited by many neurons of the previous pool. This generates a volley of spikes propagating synchronously (Abeles, 1982; Diesmann et al., 1999) from pool to pool. The particular structure with all-to-all connectivity between the pools/layers, referred to as a complete transmission line, was suggested by Griffith (1963). It can guarantee a fixed level of activity in a network of excitatory neurons. Similar structures were used to learn and reproduce complicated space-time patterns (Grossberg, 1969), for cortical (Ikegaya et al., 2004) and primary motor cortex (Prut et al., 1998) activity.

Upon investigating its properties and applicabilities, it soon became one of the main models tackling motor timing in songbirds' song production (R. Hahnloser et al., 2002; D. Jin et al., 2007). In this system, synfire chains are able to explain experimental recordings and behavioral findings. They can generate sparse bursting activity (R. Hahnloser et al., 2002) and upon tuning the noise, they can produce the observed syllable duration variability (Glaze and Troyer, 2012). Moreover, they account for behavioral adaptation. As explained in Chapter 2, in a CAF paradigm aiming to change the duration of a particular syllable in the song using white noise, the bird can change the duration of the target syllable without affecting the rest of the syllables' duration. In Pehlevan et al. (2018), the authors tried to model the learning specificity observed (Ali et al., 2013; Pehlevan et al., 2018), and compared the results from three selected models: synfire chains, a dynamic attractor model

(Laje and D. V. Buonomano, 2013) and a feedback stabilized recurrent neural network (Rajan et al., 2010; Sussillo and Abbott, 2009). They reported that only the synfire chain was able to reproduce the behavior with specificity to the target, whereas in the two other models an effect on other non target syllables (RNN) was present.

Another network architecture, similar to synaptic chains is proposed by Chang and D. Z. Jin (2009). Here, neural activity is sustained by external inputs, guided by strong global feedback inhibition and unidirectional excitation between groups or neurons. The model belongs to the pulse-coupled oscillators class. In the simple implementation of this model, a *race-to-spike* occurs every time a neuron spikes, during which global inhibition affects all neurons and excitation activates the "closer" neuron in the network. Hence, spike propagation becomes an "attractor" to which the dynamics flow from any initial conditions. In contrast to the synfire chains' limited stable regime, this model is robust in the *driven chain* regime for a large range of excitation and inhibition strengths. Moreover, in the multiple-chain scenario, a winner-take-all mechanism emerges, where only one chain is selected. Assuming each chain encodes one syllable of the zebra finch song, this could be a potential action selection mechanism.

Moreover, other models of HVC have tried to address HVC as part of a distributed network. Although previous cooling and heating experiments had shown that HVC could be in itself the timekeeper (local timing network) (Long and M. S. Fee, 2008), other evidence based on the relatively small amount of dilation (shrinking) after HVC heating (cooling), propose that a loop could be present, including HVC, Uva, the brainstem and RA (Hamaguchi, Tanaka, et al., 2016). In a comparison between the distributed and the local timing mechanism model simulations, they show that although both generate sparse sequential activity, only the distributed model through the temporal organization of active and silent phases of synaptic activity, generates detectable membrane potential correlations and coherence peaks between HVC neurons. This view challenges the traditional local synfire chain perspective, adding and emphasizing the role of collective, network-wide dynamics and hemispheric connectivity in generating the precise temporal sequences required for song. Other models emphasizing the importance of different structures of the loop have been proposed, such as the bilateral HVC synchronization model by Pang et al. (2022), where Uva guarantees interhemispheric synchrony.

3.3 Attractor Networks

Having described the general models of subsecond motor timing and the main ones used to model HVC, here we focus on an alternative model, attractor networks. In this section, we first briefly describe the dynamical systems theory and attractor networks, their architectures and applicability. Then, we focus on the ring model, its properties, applications and the reasons making it an alternative model for HVC.

Dynamical System and Attractors

Dynamical systems theory describes systems, including particle(s) or variable(s), the states of which change as a function of time while obeying a set of differential equations (Strogatz, 1994). They can be deterministic or stochastic. The state of a dynamical system is a point or vector in its state space and the set of states toward which the system converges over time is considered an attractor (Milnor, 2006). The central signatures of attractors, as summarized by Khona and I. Fiete (2022), include the lower dimensional attractor manifold, the convergence to the attractor state following perturbation and the long-time autonomous stability of the attractor states. The latter refers to isolation, no external input or activity, i.e. a completely self-sustained, autonomous system. However, in the context of a larger complex and interconnected network such as the brain, a system with untuned inputs is accepted as *effectively autonomous*.

In neuroscience, attractor networks can be used to describe and model a large number of cognitive tasks, such as short term and long term memory representation (Compte et al., 2000; Funahashi et al., 1989; Sharma et al., 2022), sequence generation (Laje and D. V. Buonomano, 2013), integration (Stringer et al., 2002), classification (Chaudhuri and I. Fiete, 2019) and decision making (X.-J. Wang, 2008). Different network architectures can give rise to point attractors (Hopfield, 1982), line, ring (Ben-Yishai et al., 1995b) and plane attractors (Burak and I. R. Fiete, 2009), cyclic attractors and chaotic attractors (Sompolinsky et al., 1988). These are split into discrete and continuous attractors, where the former consists of a set of discrete states where the network could flow to and the latter a continuous one.

Discrete attractors are mainly represented by the Hopfield model (Hopfield, 1982), where each discrete attractor corresponds to a particular memory pattern the network learns through a Hebbian-like learning rule and then can recall. As the network with the learned weights has stabilized to form attractor states, even when presented to a partial or noisy cue, the dynamics of the system drive it toward the stored pattern, effectively performing pattern completion. Hence, the network will converge to different attractor states (memories) depending on the initial conditions. The connectivity matrix is a symmetric pattern across the diagonal. Discrete models have also been used to model cortical up and down states, perceptual bistability (McWalter and McDermott, 2019; M. Wang et al., 2013), delay-period dynamics during a binary decision task in rodents (Inagaki et al., 2019) and recurrent dynamics in the olfactory system in flies (Lin et al., 2014).

On the other hand, continuous attractors neural networks (CANNs) refer to manifolds in the state space where each point within the region is an attractor, providing a continuum of stationary attractor states. In their recent review, Khona and I. Fiete (2022) make a compilation of the sufficient conditions to allow it: (i) The presence of nonlinear neurons with saturating responses or inhibition-dominated recurrent interactions and a uniform excitatory drive is necessary to keep network activity bounded (Amari, 1972; Ben-Yishai et al., 1995a; K. Zhang, 1996); (ii) strong weights, local excitation and broader inhibition to drive pattern formation (K.

Zhang, 1996), which give rise to the attractor states and (iii) continuous symmetry (translational or rotational invariance) in the weights. Examples of these models' applications in the brain include using line attractors (infinite set of fixed points) for oculomotor integration (Aksay et al., 2001; Arnold and Robinson, 1997), ring attractors for head direction cells (Kim et al., 2017; Taube et al., 1990), two-dimensional toroidal attractors in the grid cell system (Burak and I. R. Fiete, 2009; Trettel et al., 2019).

Ring Attractor

Ring attractors are continuous attractors, and are referred to also as line attractors, where ends meet or bump attractors (Ben-Yishai et al., 1995b; Compte et al., 2000; Hansel and Sompolinsky, 1998; K. Zhang, 1996). They are structured attractor networks, where the neurons are arranged in a ring-like topology and each neuron represents a particular (preferred) feature. The connectivity is characterized by symmetric synaptic weights, strong local excitation and broader inhibition, which helps maintain the localized bump and stabilizes it. Additionally, uniform excitatory input is provided to all neurons. This gives rise to a bump-like activity pattern centered in a particular location, and the totality of these locations in the state space forms a ring. Since the mid 1990s, ring attractors have been proposed to underlie the rodent's head direction system (Ben-Yishai et al., 1995b; K. Zhang, 1996) and more recent evidence has identified network motifs and dynamics in the Drosophila's head direction system (anterodorsal thalamus) corresponding to ring attractors (Kim et al., 2017). Supported by the suggestion that a possible underlying mechanism for motor timing could rely on strong internal connections capable of self-sustained activity (M. Mauk and D. Buonomano, 2004), the ring model can also be considered for motor timing. However, the ring model has stationary states and to be able to account for pattern motor timing, it would need to generate a drifting bump of activity. That is possible through a moving external input, neuronal adaptation (Hansel and Sompolinsky, 1998) or asymmetry in the weights. In chapter 5, we will detail these three mechanisms and propose one as suitable for modeling HVC dynamics.

4 Objectives and Overview of the Thesis

In this manuscript we aim to address subsecond motor timing in the animal model of songbirds using computational modeling. The preceding three chapters provide an introduction to the key concepts necessary to understand the broader context, songbirds and modeling approaches. In Chapter 1, we present the broader concept of time, the taxonomy of timing in neuroscience, the different scales of timing and the anatomical/neural substrates. Among the anatomical substrates, we single out the premotor nucleus in songbirds as an outstanding biological model of motor timing at the scale of tens to hundreds of milliseconds. In contrast to more complex underlying mechanisms in mammals, songbirds have a dedicated set of interconnected brain nuclei known as the *song system*, responsible for the acquisition and performance of the sensorimotor task of song production. This entails a simpler structure to study, both experimentally and computationally. Moreover, the processes of birdsong production and learning are resemblant to those of human speech. With this motivation, in Chapter 2, we present the anatomy, physiology and behavior of songbirds in the context of vocal production and learning. We end the chapter by focusing only on motor timing, its plasticity and the nuclei in songbirds hypothesized to be involved in its emergence.

Altogether, in Chapter 1 and 2, we have highlighted the relevance and benefits of using the song-timing circuitry in songbirds to model subsecond motor timing. The following chapter, chapter 3 presents a collection of hypotheses and computational approaches to subsecond motor timing, ranging from the main classes of models used across different species (mainly mammals), to models specifically dedicated to simulating the dynamics of song-timing. We position the latter to one of the larger classes and provide a mapping from a specific model to a more general one. Lastly, we include a section on dynamical systems and attractors, as a class of robust models which can be proposed to model temporal dynamics in HVC.

In brief, we consider the song timing circuitry in songbirds to be an example of subsecond motor timing. The dynamics observed in songbirds could lead to a wider understanding of temporal control in other taxa, including humans, particularly because the task at hand, birdsong, is so similar to human speech. To address subsecond motor timing, we use a computational approach and formulate two main questions:

- What are the underlying dynamics that give rise to the precise temporal control of a learned sensorimotor task?
- What are the mechanisms, by which behavioral flexibility of motor timing is possible?

Modeling The Underlying Mechanisms of Song Timing

Premotor nucleus HVC, has been modeled either as a standalone timing structure (D. Jin et al., 2007) or as an element within a distributed system (Hamaguchi, Tanaka, et al., 2016), from which timing emerges. In both cases, a chain-like structure, which creates sequence generation is hypothesized. We start Chapter 5 by focusing on the main example of a synaptic chain, the Synfire Chain model (Abeles, 1982; Diesmann et al., 1999), which is considered the golden standard in modeling HVC. Due to its structure and previous reports, we expect limited robustness and aim to first show that this is the case, before proceeding to another model. Hence, we propose investigating its timing flexibility by exploring the possible synaptic weight range that allows *chain propagation* and the model's robustness to noise in the input and synaptic weights. Furthermore we search for a more robust model, which should mirror the reported structure of a cortical area and HVC, incorporating recurrent connections and inhibition. At the same time, it should retain a chain-like configuration to account for feedforwardness (chain propagation). These required characteristics evoke the concept of a particular kind of a ring attractor model, one that is capable of progressing through its fixed points or attractor states. This would manifest with a moving bump of activity.

To explore this hypothesis, it is simpler to start with a rate based model and then if necessary move to a spiking neural network, which can explain the underlying dynamics more accurately. The rate based model provides results that up to a certain limit, should be mappable and can guide investigation in models entailing more neuronal complexities, such as spiking neurons. As the goal of exploring a different model lies on its robustness, (i) we will first prove that the proposed model is robust. Then (ii) we will explore the different mechanisms that can make the activity propagate in the network, causing a sequential activity as reported in HVC. As feedforwardness is hypothesized in the connections of the HVC_{RA} neurons, we add a β term, to make the connectivity pattern asymmetric. This will, on its own, allow bump propagation. Lastly (iii) we will investigate if a ring attractor architecture can capture even more specific features of HVC_{RA}, i.e. sparsity in activity and burst time duration (R. Hahnloser et al., 2002).

Following the implementation of a one population rate based model and addressing the above mentioned points, a more biologically realistic model may be necessary to address the number of spikes per burst and burst time more accurately. Moreover, the role of the neuronal dynamics in the network behavior should also be explored. For instance, D. Z. Jin et al. (2007) have previously shown that instrinsic bursting, with a conductance based model increases robustness in

a synfire chain model. An additional mechanism to achieve bursting dynamics is with an adaptation current, through an AdEx neuronal model (Gerstner and Brette, 2009). Hence, we expect that adaptation would also increase robustness in the ring model and investigate the relationship. Lastly, as the neurons responsible for song timing are inside a microcircuitry in HVC, it is important to model the other elements of the microcircuitry. Interneurons play an important role in motor timing (Kosche et al., 2015; Vallentin, Kosche, et al., 2016) and hence, their effect will be explored in an excitatory inhibitory (EI) network.

HVC may be a sole contributor to motor timing, or it could be guided by upstream brain areas: the auditory nucleus NIf (nucleus interfacialis of the nidopallium) and/or the thalamic nucleus Uva (nucleus uvaeformis) (detailed in Chapter 2). Whereas lesion studies in NIf (Naie and R. Hahnloser, 2011) have revealed that NIf is not necessary in adult song production, several studies have shown that Uva is highly important in song production (Coleman and E. T. Vu, 2005) and a recent study has shown that it drives a particular set of neurons in HVC, those firing at syllable boundaries (Moll et al., 2023). Since there appears to be no lateralization (Long and M. S. Fee, 2008) in song production, it is probable that Uva input has a synchronizing effect. As such, it would resemble a sequential input reaching the two HVCs at discrete times in the song and have no effect in the burst sequences, which we still hypothesize to be generated within HVC. The effect of such an input in the model and whether it synchronizes the activity of the 2 simulated HVCs, is included in Section 5.4.

Plasticity in Motor Timing

The presence of behavioral plasticity in birdsong has been revealed initially through pitch and duration shift experiments using a Conditional Auditory Feedback paradigm (Ali et al., 2013; Andalman and M. S. Fee, 2009) (detailed in Chapter 2). In Chapter 6, we aim to numerically simulate the protocol in order to incite duration shift in either direction, i.e. to lengthen and shorten the target syllable duration, using a reward-covariance reinforcement learning rule. A key feature of this behavioral experiment and a reported computational challenge, is the fact that only the targeted segment (syllable) is affected (Pehlevan et al., 2018). In a simplified numerical representation of this behavior, a set of neurons represents a syllable. Within a synfire chain architecture, a set of neurons, and more specifically two layers of connected neurons, represent the start and the end of the syllable and the weight update occurs only in the connections between these two layers, making the neurons of the two layers fire closer in time (duration shortening) or further in time (duration lengthening). As the weights between the other layers are not updated, there is no effect on any of the other syllables/segments. However, upon using recurrently connected neural networks, a duration decrease could take place in many different ways, and a neuron does not only project forward, making its other synapses also update. This would show an effect on multiple 'sets' of neu-

rons, some of which not mapped to the target segment. Hence, an effect in other syllables, referred to as *interference*, can be observed.

In Chapter 6, we first explore whether with the ring attractor architecture and the chosen learning rule, a shift in duration is possible. Next, we check how the weights are modified to acquire this shift. We first hypothesized a local positive weight change modification, similar to the one observed in the synfire chain, in the case of duration shortening. Moreover, we hypothesized that this change would have to be very narrowly focused at the level of the pre- and postsynaptic neurons affiliated with the target syllable. In such a way, no effect on the other syllables would be observed. Before starting the learning simulations, we tested this hypothesis, by adding such a change to the weight matrix and confirmed our hypothesis. However, this would not be a robust mechanisms and as we will show in chapter 6, this is not the mechanism we observe.

Moreover, with the same reasoning as presented above, the effect of different neuronal models to the learning is also to be addressed. For instance, does adaptation increase the rate of learning, or does it rather decrease it? What is the effect in learning, of an inhibitory population as compared to global inhibition? Most importantly, the question of whether the underlying mechanisms, i.e. the pattern of change in weights, is significantly different based on the neuronal model used.

Lastly, the specificity of learning needs to be investigated, whether it is present with a rate based model and if it is the same in different levels of neuronal complexity. This then needs to be confronted with previously reported experimental data and data collected in the lab. Hence, in chapter 6, we present the results of learning at the level of the target syllable, the presence or absence of interference and the analysis of experimental data, which have been acquired using a CAF paradigm to change syllable duration. As the averaged results have been already reported, the individual mechanisms each bird uses to change syllable duration, are of particular interest to us.

Implications and Perspectives

In the last chapter (Chapter 7), a general discussion on the implications of our results and their position in the literature takes place. From the wider timing perspective, we present a ring attractor model, taking place in the population clock model class, as a computational approach to subsecond motor timing. Like population clocks, this model provides robust and flexible timing. In the songbird and computational perspective, we discuss the possible emergence of such a structured model from initial learning and how we can numerically reproduce imitation learning from an adult tutor, during the sensory and sensorimotor period. Furthermore, we propose that the synfire chains and ring attractors, may in fact lie in a continuum. Lastly, we suggest an additional learning rule to account for the return to baseline duration, observed after the CAF is stopped.

5 A Robust Model of Motor Timing

Neural mechanisms of sub-second timing with millisecond precision can be investigated through the male zebra finch birdsong, where the premotor nucleus HVC is responsible for the precise control of motor timing, leading a tight coordination of vocal and respiratory muscles muscles that produce a stereotypical song. Given that birdsong is a learned sensorimotor skill, the nucleus HVC provides an excellent environment to study motor timing in such skills. Thus, modeling the observed dynamics of HVC can give further insights into the underlying mechanisms and generate important predictions. To some extent, such predictions might be useful not only for birdsong, but also for human speech and other sensori-motor skills.

Traditionally, HVC has been described with a network of excitatory neurons organized in a sequentially connected chain of neuronal populations, often referred to as *the synfire chain*, belonging to the class of population clock models. However, the purely feedforwad connectivity pattern of synfire chains does not appear compatible with the connectivity patterns observed experimentally in cortical networks. More specifically, unidirectional connectivity between groups of neurons is incompatible with: the high level of reciprocal connectivity typically observed

in the cortex (Perin et al., 2011), the strongly interconnected web (Kosche et al., 2015), and with structured excitation and inhibition in HVC. Additionally, as it has been previously shown (Pehlevan et al., 2018), synfire chain networks are sensitive to noise and require fine tuned synaptic strengths to avoid runaway excitation or decay.

In this chapter, we propose that the gradual propagation of an activity bump in HVC is driven by attractor dynamics. In particular, a ring attractor can drive a drifting activity bump with robust and resilient properties thanks to recurrent connections (Ben-Yishai et al., 1995b; Compte et al., 2000; Hansel and Sompolinsky, 1998; K. Zhang, 1996). However, it is unclear whether the ring attractor can account for the properties of HVC neuronal dynamics and the behavioral adaptation of the song timing.

The structure of this chapter is as follows. First, we review the caveats of synfire chains. Second, we introduce the ring attractor model as an alternative and investigate its limitations. We then reproduce experimental observations and explore the possible underlying timing mechanisms in HVC with our proposed model. We do this in three levels of increasing complexity: with a rate-based model (Wilson and Cowan, 1973), with a spiking neural network (SNN) model (Lapique, 1907) and with a two population, excitatory inhibitory (EI) network. We also make predictions about the effect of local inhibition (GABAA agonist and antagonist) in HVC. Lastly, based on recent findings, we also discuss a SNN model with synchronizing input coming from the thalamic nucleus, Uva.

5.1 Robustness of the State of the Art Model: Synfire Chain

Here we assess the robustness of propagating localized activity profiles in synfire chains. To do this, we implement a synfire chain model with adaptive exponential integrate and fire (AdEx) neurons.

The neuron model: AdEx

We use an AdEx neuronal model, as a cost-effective choice that can generate different types of spiking patterns: tonic activity, spike frequency adaptation, bursting activity, etc. Since we try to model HVC_{RA} neurons, shown to exhibit burst-like spiking patterns, AdEx proves to be an ideal neuron model. It relies on two main equations (Brette and Gerstner, 2005):

$$C\frac{dV_i}{dt} = -g_L\,(V_i - E_L) + g_L\,\Delta_T\,e^{\left(\frac{V-V_T}{\Delta_T}\right)} - w_i + I_{\text{syn}} + I_{\text{ext}} + \sqrt{\tau_n}\,\sigma_n\,\eta_i(t), \quad (5.1)$$

$$\tau_w\frac{dw_i}{dt} = a\,(V_i - E_L) - w_i + b\,\tau_w\,\delta(t - t_i), \quad (5.2)$$

which describe the dynamics of the membrane potential V_i and the adaptation current w_i of neuron i, respectively.

We first explain the membrane potential dynamics in Eq. (5.1), where C denotes the membrane capacitance. The first right-hand-side (rhs) term describes the leak term with g_L and E_L denoting the leak conductance and leak reversal potential, respectively. The second rhs term describes the exponential activation, with Δ_T being the slope factor of the activation and V_T the threshold potential. The third term in Eq. (5.1) describes the coupling to the adaptation current. Moreover, I_{syn} and I_{ext} are generally time-dependent and describe the synaptic and external currents, respectively. Lastly, $\eta_i(t)$ refers to the white noise, the amplitude of which is determined by the corresponding time constant τ_n and standard deviation σ_n.

In the adaptation current dynamics, Eq. (5.2), τ_w denotes the adaptation timescale. The first rhs term describes the coupling of adaptation current to the membrane potential, with a coefficient a. The second term describes the leak of the current, and the third term denotes an increase of the adaptation current by b at each spike time t_i of neuron i. More precisely, a spike time-point t_i is one that satisfies $V_i(t_i) > V_T$.

When the membrane potential crosses the threshold potential, $V_i > V_T$, the above parameters are updated:

$$V_i \to V_R, \tag{5.3}$$

$$w_i \to w_i + b, \tag{5.4}$$

where the membrane potential V is reset to the reset potential V_R, and the adaptation current w is updated according to Eq. (5.2). This system of equations tries to reproduce the biological behavior of neurons. For instance, the sodium activation curve is simulated through the slope factor Δ_T (sharpness of the spikes) and the calcium-dependent potassium channels though the spike triggered adaptation parameter b. The latter is under the assumption that calcium influx occurs mainly during an action potential (Gerstner and Brette, 2009).

The Synfire Chain

We build a chain of 90 layers where each layer comprises n neurons. Each neuron on layer l has an excitatory synaptic connection (weight) to every neuron of the next layer $l+1$ as illustrated in Fig.5.1. This is also known as an all-to-all feed-forward connectivity. Following the assumptions previously presented for the synfire chain, the mapping between network activity and time is layer specific, such that a time-point $t(l)$ is a function of only the activity of the neurons in layer l. Hence, the activity in each layer encodes the start and end of the interval of two consecutive layers, i.e. the first spike of the layer marks the beginning of the new song interval (syllable + gap, inside a motif) and the end of the previous one. The

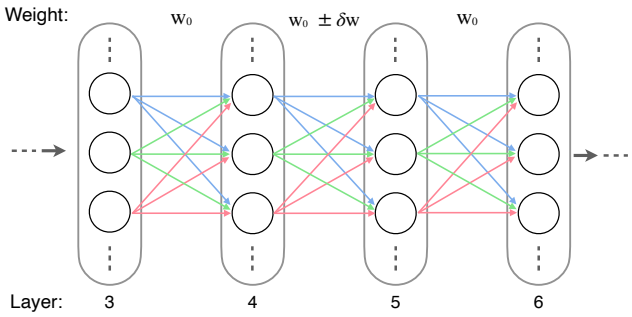


Figure 5.1: **Architecture of the Synfire Chain Network.** The Synfire Chain exhibits an all-to-all feed-forward connectivity. For the flexibility protocol, only the weights between the neurons of layer 4 and 5 are changed by δw. The flexibility is defined by the range of values δw can take.

synaptic input is a function of the synaptic weights and the postsynaptic potential (PSP), as presented below:

$$I_{\text{syn},i} = \sum_j W_{ij} \sum_k \Theta(t - t_j^k)\, e^{\frac{-(t-t_j^k)}{\tau_s}}, \tag{5.5}$$

where W_{ij} is the synaptic connectivity weight with i being the post-synaptic and j the pre-synaptic neurons, respectively. $\sum_j$ denotes a summation over all pre-synaptic neurons. Moreover, $\Theta(t - t_j^k)$ is a step function with t_j^k denoting the k^{th} spike of the j^{th} neuron and $\sum_k$ denotes a summation over all the spike times of neuron j. Lastly, τ_S denotes the synaptic time constant. For brevity, all the parameters of Eqs.(5.1)–(5.5) are summarized in Table 5.1.

Procedure of Simulations

Eqs. (5.1) (5.2) are integrated by using a forward time Euler scheme with a numerical time-step $\Delta t = 0.01$ ms for an overall duration of $\mathcal{T} = 150$ ms. Optimization runs were performed to verify that the chosen Δt does not alter the outcome. Parameter values for the simulations are detailed in Table 5.1.

Symbol	Parameter	Values	Symbol	Parameter	Values
N	number of neurons	3000	E_L	leak reversal potential	-70.6 mV
C	membrane capacitance	281 pF	g_L	leak conductance	30 nS
V_T	threshold	-50.4 mV	Δ_T	exponential slope factor	2 mV
Δt	timestep	0.01 ms	W_{ij}	weight, pre-j to post-i	1.4 mV
τ_w	adaptation time constant	100 ms	a	adaptation coupling parameter	- 0.5 nS
b	spike triggered adaptation	0.5 nA	I_{ext}	initial external input	12 nA
τ_n	time constant of the noise	10 ms	σ_n	st. deviation of noise	0.2 mV
τ_s	synaptic time constant	5 ms	$\mathcal{T}$	simulation duration	150 ms

Table 5.1: Values of the parameters used in Eqs.(5.1) - (5.2) for the Synfire Chain.

In the following subsections, we report simulation results of synfire chain networks for different number of neurons per layer, n. We initialize the system by applying an external current of $I_{\text{ext}} = 12$ nA only to the neurons of the layer $l = 1$ during the first 1 ms of each simulation. The feed-forward synaptic weights are set equal to each other:

$$W_{ij} = \begin{cases} \mathrm{w}_0 & \text{if } \; j \in l \;\&\; i \in l+1 \\ 0 & \text{otherwise,} \end{cases} \tag{5.6}$$

where, l denotes the layer index and w_0 is manually determined for each n in order to allow for activity propagation in the chain. In other words, Eq. (5.6) tells us that each neuron at the layer $l+1$ receives synaptic input only from the neurons of the previous layer l (Fig. 5.1).

In Fig. 5.2, the first two panels present the resultant activity. Moreover, in subsubsection 5.1 we test the flexibility of the network, i.e. the range of the synaptic weight between two specific layers (see Fig. 5.1) which can sustain the uniform activity propagation along the chain. Next, in section 5.1 we test the robustness of the network. Namely, we investigate how far the synfire chain network can sustain the activity propagation when it is under the influence of noisy perturbation. Detailed descriptions are provided in the corresponding sections.

Flexibility of Chain Propagation

We examine the possible connectivity range that the weights can explore across 10 synfire chains of varying sizes, i.e. $n = \{15, 30, 45, 60, 75, 90, 105, 130, 135, 150\}$ number of neurons per layer. As mentioned above, w_0 is determined separately for each n such that it allows for a uniform activity propagation along the chain. We find that w_0 values which allow for propagation decrease as n increases. Moreover, we numerically explore the behavior of the network when the weights between neurons of the layers 4 and 5 alone are modified, i.e. $W_{ij} = \mathrm{w}_0 + \delta\mathrm{w}$, whereas all weights of neurons from other layers l to $l+1$ (with $l \neq 4$) are kept at w_0. We focus on the limits of $\delta\mathrm{w}$ where the activity propagation stops/saturates in order to determine the upper/lower limits of weight flexibility. This range was previously referred to as *timing flexibility* (Pehlevan et al., 2018).

We find an absolute decrease in the flexibility range of the synaptic weights as the number of neurons per layer increases, and an almost constant percentage wise flexibility. Hence, in larger networks, the weights have to be fine-tuned within a narrow interval in order to sustain an activity propagation (Fig. 5.2 (B)). The network with largest n comprises a total of 13,500 neurons, a number that approaches the $\sim$20,000 song-related HVC_{RA} neurons reported (M. S. Fee et al., 2004; N. Wang et al., 2002).

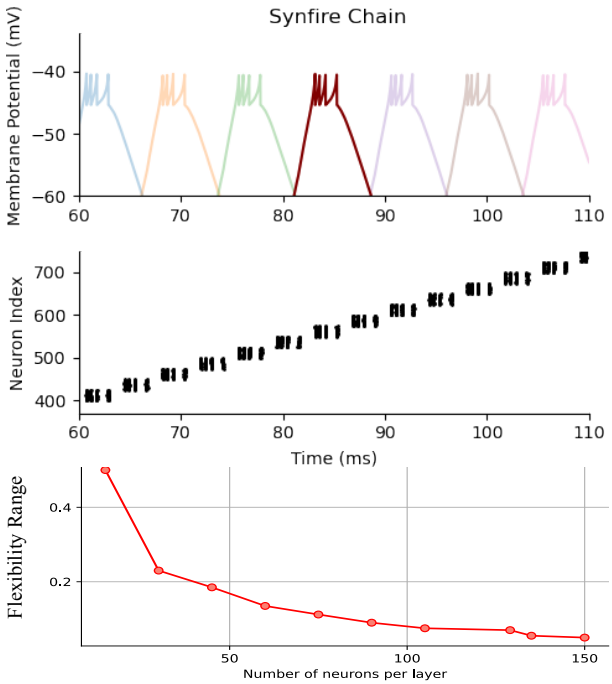


Figure 5.2: **Synfire Chain and the Flexibility Range in Different Network Sizes** Upper two panels: Spiking dynamics of the Synfire Chain AdEx model. Third panel: Timing flexibility range refers to the range of synaptic weight, which allow proper activity propagation in the chain. It is expressed as the ΔW range, of the synaptic weight W. The values of ΔW are computed for different number of neurons per layer in the synfire chain.

Robustness protocols

Although several studies have investigated the robustness of synfire chains, here we particularly study it in the context of perturbations to HVC dynamics. We design two numerical protocols to evaluate the robustness of the synfire chain model under possible noisy perturbations: (i) in the synaptic weights, W_{ij}, and (ii) in the external current, I_{ext}. They are to the network with 30 neurons per layer (N = 90). In both cases, binary noise is injected on a selected percentage of synaptic weights and neurons, respectively. The magnitude of noise is set based on a number of numerical explorations. For synaptic weights, it amounts to a value of 30% of the synaptic weights and for external current to 10% of the initial external input. We observe, in both protocols, that the activity propagation along the chain stops when the percentage of perturbed weights or neurons reaches a certain limit (shown in Fig. 5.3).

For investigation simplicity purposes, synaptic weight changes are set to be all negative and upon affecting 10 % of the synapses, we show in Fig. 5.3 (A) that in the synfire chain, propagation stops. The second protocol can also be taken as

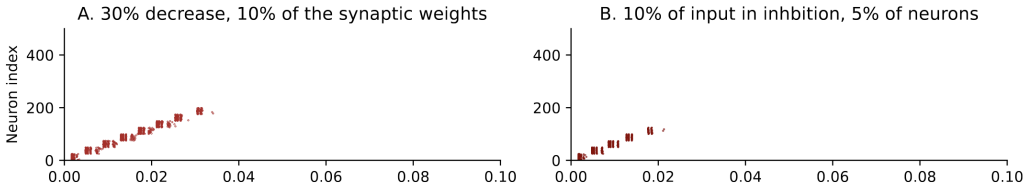


Figure 5.3: **Robustness of the Synfire Chain.** (A) Behavior of the synfire chain facing a decrease in synaptic weights. At 10% weight involvement, there is a stop in propagation. (B) Behavior of the synfire chain in face of a decrease in external input. There is a stop in propagation when 5% of the population is affected.

a proxy for local inhibition, where neurons although selected randomly based on their preferred timing (feature space), could be close in space. With a 10% decrease in external input, to 5% of the neurons, propagation stops as shown in Fig. 5.3 (B).

Discussion

Above, we present a synfire chain model with AdEx neurons. Although it is a very good model to explain the sequential activity in HVC and the neuronal dynamics, the robustness of such a model is debatable. We have present 3 protocols, advocating for a lack of robustness. The first protocol on testing the flexibility range has previously been shown in a chain of single integrate and fire neurons (Pehlevan et al., 2018). In our simulations, upon adding adaptation to the neuron model and multiple neurons per layer, we observe similar results, proving that only a limited range of weights allows proper propagation in the chain. This is important as it requires fine tuning of the weights and may be a limitation in learning, as the weights cannot change below or above a certain limit. The two other protocols show that the limit after which propagation in the chain breaks in response to a small noisy perturbation, is relatively small (10% for weights and 5% for neurons). These protocols can, however, be extended further, with different perturbation patterns and can be implemented in synfire chains with heterogenous weights.

5.2 The Rate-Based Ring Attractor Model of HVC

In this section, we investigate how accurately a ring model can describe and furthermore explain HVC network dynamics. In contrast to the synfire chain model above, where the neuronal activity is described via the membrane potential, here we describe the neuronal activity via the firing rate, i.e. a rate-based model. The mean firing rate, $m(x,t)$ is a variable averaged in space and time. The rate based model has less parameters than the spiking model, is easier to investigate and can provide valuable insights, which we use for the spiking neuron model in section 5.3. The structure of this section is presented in the following paragraphs.

In subsection 5.2.1 we introduce the rate-based model in a ring architecture, which has been studied in detail in different contexts (Hansel and Sompolinsky, 1998; K. Zhang, 1996). We follow the established formalism and use a connectivity profile with short-range excitation and long-range inhibition, allowing the spontaneous formation of a localized neuronal activity, i.e. a bump. Moreover, to acquire a a very *narrow* bump, we propose using a gaussian profile (instead of a cosine as in Hansel and Sompolinsky (1998)). Next, in section 5.2.2, we numerically investigate the parameter space of the connectivity profile for which the ring model leads to the bump formation. Then, in subsection 5.2.3, we focus on different mechanisms that can generate an activity bump which drifts along the ring. We initially revise cases where the external input I_{ext} is present, and then focus on cases when external input is absent, $I_{\text{ext}} = 0$. In the latter case, we discuss scenarios where the system is subject to adaptation currents or asymmetries in the connectivity profile. Both of these features can induce a drive in the system. Lastly, in subsection 5.2.4, we focus on the scenario with an asymmetric connectivity pattern. Here, we use similar techniques as in subsection 5.2.2 to study how the asymmetry of connectivity affects the robustness of the system, i.e. parameter space for which a drifting bump can exist. These results pave our way to proposing the ring attractor as a robust model for describing HVC dynamics.

5.2.1 Model and Methods

The Ring Model

We consider a neural population whose mean firing rate is expressed as $m(x,t)$, with $x \in [-\pi/2, \pi/2[$ being the position over a closed one-dimensional ring with period π, and t representing time. The firing rate $m(x,t)$ represents the averaged neuronal activity, and is governed by the following equation:

$$\tau \frac{d}{dt} m(x,t) = -m(x,t) + G\big[\, I_{\text{ext}}(x,t) + I_{\text{syn}}(x,t) + I_a(x,t) - T + \sqrt{\tau_n}\sigma_n \eta(x,t) \,\big], \tag{5.7}$$

where τ is the neuronal membrane time constant. On the rhs of Eq. (5.7), I_{ext} is the external constant input, I_{syn} the synaptic input and T represents the threshold. I_a is the adaptation current, and it is set to 0, unless otherwise stated. The last rhs

term is a zero-mean white noise which, for numerical purposes, is correlated in x with a Gaussian of standard deviation of $\pi/500$. For the nonlinear gain function $G[I]$, the simple semi-linear form is adopted:

$$G[I] = \begin{cases} 0 & I < 0 \\ I & 0 \leq I \leq 1 \\ 1 & I > 1. \end{cases} \tag{5.8}$$

In order to investigate the system numerically, we first discretize the position x in $N = 1000$ neuronal units. Then, we use the following expression for the synaptic input:

$$I_{\text{syn}}(x,t) = \sum_{x'} \frac{1}{N} W(x - x')\, m(x',t), \tag{5.9}$$

where $W(x - x')$ denotes the synaptic connectivity between the pre-x', and the post-x, synaptic neural positions. In other words, W is the weight matrix and $\sum_{x'}$ denotes a summation over all the pre-synaptic position indices.

Note that all the previous equations are identical with the ones derived a previous study (Hansel and Sompolinsky, 1998). The modification we introduce is at the level of W, which we choose of the following form:

$$W(x - x') = W_0 + W_2\, e^{-\frac{1}{2}\left(\frac{x - x' + \beta}{\sigma}\right)^2}, \tag{5.10}$$

where $(x - x' + \beta)$ is taken modulo π, $W_0 < 0$ stands for the global inhibition, $W_2 > 0$ the excitation factor and σ the width of the gaussian excitatory connectivity profile. We have also introduced a bias term, β, that makes the connectivity pattern asymmetric. As a consequence, the activity can be pushed to dynamically drift along the ring in one direction or the other, depending on the sign of β. A more detailed discussion on the role of this parameter is reported in section 5.2.3.

In our model, x *does not* represent the neuronal position in a spatial topology, as HVC micro-circuitry does not display spatio-temporal organization (see Fig. 5.4 (A)). Instead, x represents the neuronal position in the feature space, which is defined based on the preferred timing of each neuron (see Fig. 5.4 (B)).

Simulation

Eq. 5.7 is integrated by using a forward Euler scheme with a numerical time-step $\Delta t = 0.25$ ms. Multiple runs were performed to identify an appropriate value for Δt that does not alter the outcome. Unless otherwise stated, parameter values for the simulations are set as in Table 5.2.

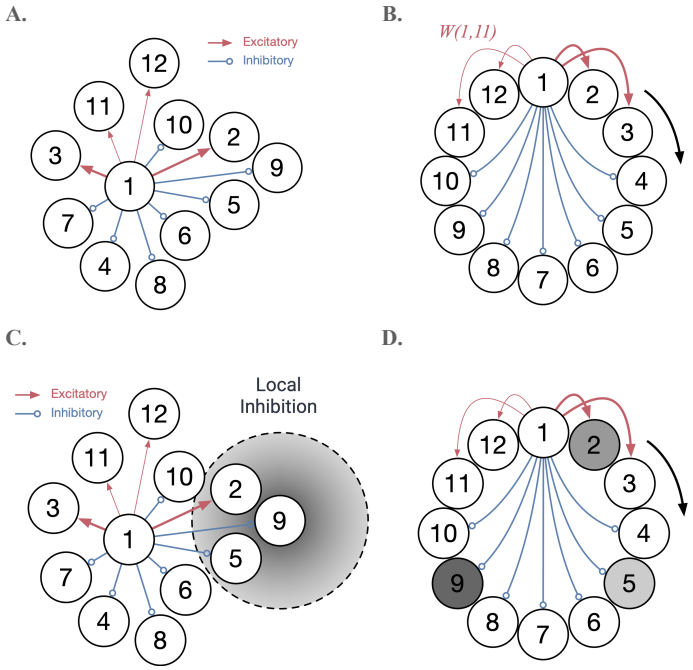


Figure 5.4: **Simplified illustration of the ring connectivity and consequences of local inhibition.** (A) The connectivity represents the physical location of neurons. (B) The same connectivity is represented using the preferred timing as a neighbourhood proxy for the placement of neurons. (C) Injection of a local inhibition in nucleus HVC. (D) The same local inhibition shown when neurons are ordered according to their preferred timing. In this spatial representation, the effect is not local, but rather distributed, which makes the network more robust.

Center of Mass of Activity (COM) and Bump speed

In order to characterize the system's dynamics, we define a "center of mass" (COM) of activity. The peak position of the bump, $C(t)$, is thus defined using the COM of activity over the whole population. $C(t)$ is computed based on equations of COM for systems with periodic boundary conditions (Bai and Breen, 2008).

We assume that the mapping between network activity and time is specific to the bump such that, at the time-point $t(x)$, the COM of the bump is at neuron position x, which is maximally activated. Thus, as the bump of activity moves across the ring, different neurons will be the most active at different time-points (Fig. 5.8). To acquire this, $C(t)$ is further discretized into $\bar{C}(t)$ and maps to the nearest discrete unit/neuron position:

$$\bar{C}(t) = \text{argmin}_i |C(t) - x_i|. \tag{5.11}$$

Symbol	Parameter	Values	Symbol	Parameters	Values
N	number of neurons	1000	T	threshold	0.9
Δt	timestep	0.25 ms	W_0	global inhibition	-5
$\mathcal{T}$	duration	2 s	W_2	excitation strength	28
τ	membrane time constant	10 ms	β	bias in the weight matrix	0.05
I_{ext}	external constant input	1.1	σ	Eq. (5.10)	0.067
τ_n	time constant of the noise	1 ms	σ_n	std. deviation of the noise	0.02

Table 5.2: Values of the parameters used in Eqs. (5.7) – (5.10).

The speed of the bump is then computed as the displacement (in the neuronal feature space) of the center of mass over time.

Syllable definition and duration

A syllable corresponds to a fixed segment of the ring. We divide, the discretized space of the ring into equally sized segments such that for s syllables and n neurons, syllable i is defined by $[\frac{i}{s}n, \frac{i+1}{s}n]$. The beginning and the end of syllable i are respectively defined by the timepoint the center of the activity bump $\bar{C}(t)$ crosses the lower and upper limit of the segment. Mean syllable duration in zebra finches has been reported to be 110 ± 56 ms (Glaze and Troye, 2006). We aim to simulate an 'accurate' syllable duration, which in our model, depends on the speed of the bump. Hence, the magnitude of drifting mechanism, in our case the bias, needs to be chosen accordingly (detailed in section 5.2.3).

5.2.2 Conditions for a Stationary Bump

Continuous attractor models require the fulfillment of a set of theoretical conditions to be able to converge into different attractor states (Khona and I. Fiete, 2022): nonlinear neurons, continuous symmetry in the weights and strong weights with short-range excitation and long-range inhibition. Moreover, phase analyses to investigate when the network sets in the different stationary states are necessary. In the ring model with a cosine symmetric connectivity profile, three main states were shown (Hansel and Sompolinsky, 1998): (i) a homogeneous state, where there is no activity in the network or it converges to a minimal homogeneous excitatory state; (ii) an amplitude instability, where the whole population fires at the saturation level; and (iii) the marginal state, where bump formation occurs. The activity bump appears as a "balance" between strong & localized neuronal interactions and the global inhibition. The different positions where the bump formation takes place make out the low dimensional ring attractor.

Here we investigate the parameter space which allows for the spontaneous formation of the bump in the rate-based model with a narrow Gaussian symmetric ($\beta = 0$) connectivity profile. In section 5.2.4 we also investigate the case of an asymmetric profile, $\beta > 0$. We perform numerical analyses to investigate the

behavior of the network by tuning the global inhibition, W_0, and the excitation parameters, namely the amplitude W_2 (5.5 (A)) and the width σ of the connectivity profile (5.5 (B)).

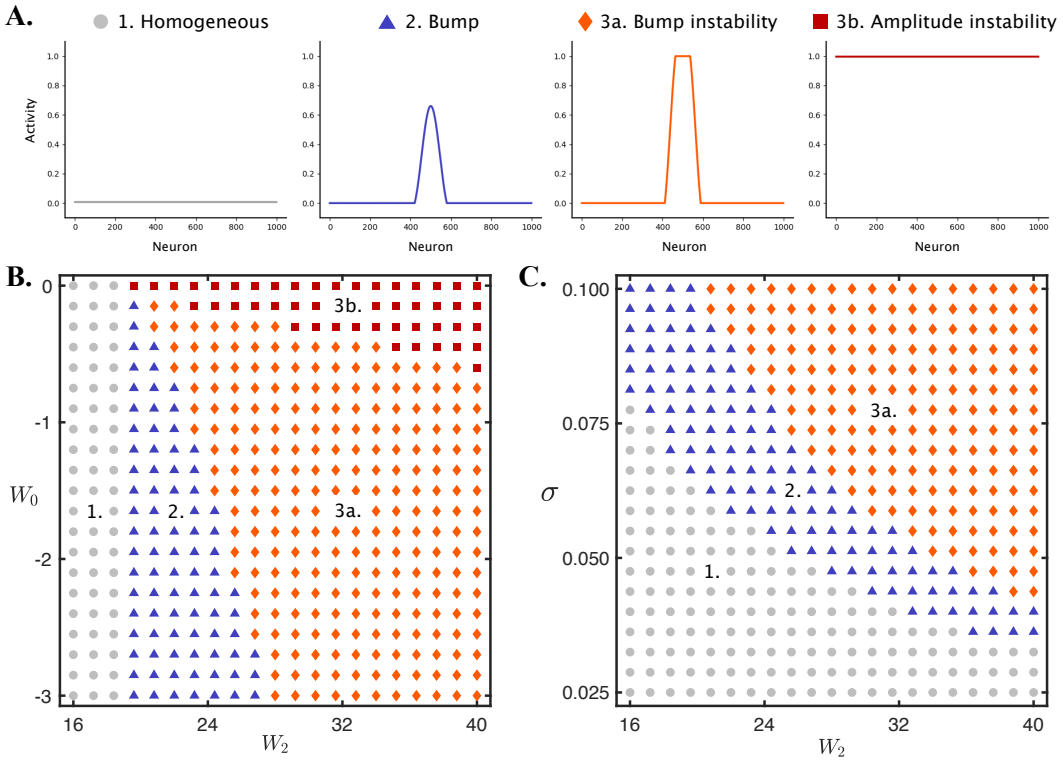


Figure 5.5: **Two numerical phase diagrams for the rate based model with a Gaussian symmetric connectivity profile.** (A) The four distinct asymptotic states of the system. Their corresponding regions are color coded in the phase diagrams. (B) The rate based model's phase diagram based on W_0 and W_2. (C) The rate based model's phase diagram based on σ and W_2. Parameters as in Table 5.2, except for $I_{\text{ext}} = 0.92$, $\beta = 0$, $\tau_n = 0$.

To identify the different asymptotic states and their regions in the parameter space, we start from a quiescent state which has only two consecutive neurons firing at the saturation level:

$$m(x_i, t = 0) = \begin{cases} 1 & \text{if } i = 499, 500 \\ 0 & \text{otherwise.} \end{cases} \tag{5.12}$$

We let the system run for a duration of $\mathcal{T} = 5$ s, a sufficient amount of time for the system to reach its asymptotic state. Similar to previous works (Aussel et al., 2017; Hansel and Sompolinsky, 1998), we observe the following 4 types of states (see Figure 5.5 (C)):

- In Region 1, the system converges to a homogeneous state.
- In Region 2, the system converges to a bump, i.e. the marginal state.
- In Region 3a, the system yields a saturated bump, i.e. bump instability.
- In Region 3b, the whole system saturates, i.e. amplitude instability.

The observed states are dependent on the parameters we investigate. In Fig. 5.5 (A), where bump width is kept constant, the transition out of the homogenous state depends solely on W2's magnitude. As the magnitude increases above a certain limit, the marginal state appears. The rest of the states are dependent on the interaction (ratio) of W_0 and W_2. In contrast in (B), where W_0 is kept constant, the transition out of the homogenous state relies on the interaction between the amplitude, W_2 and width σ of the excitation. Moreover, region 3a is dependent on the current activation function; if the activation function was changed to remove the saturation ($I \leq 1$), the peak of the bump would diverge to infinity.

5.2.3 Mechanisms Generating a Drifting Bump

In the presence of a symmetric ($\beta = 0$) connectivity and a constant and homogeneous input, upon stimulus presentation, the activity bump settles in one of the stationary states and remains there until another localized input or a high enough perturbation is exerted. However, since the purpose of this model is to generate sequential activity, we investigate how propagation across these states can be achieved. There are at least three ways to ensure bump propagation. These include an external drive in the form of a moving input stimulus and internal drives such as adaptation and asymmetric connections. The relationships between each drive and the population activity are described below and illustrated in Fig. 5.6 and 5.7.

External drive: Moving Input

In a cosine ring model, several moving activity profiles in response to changing stimulus feature, virtual rotation and locking to moving stimulus feature have been analytically investigated (Hansel and Sompolinsky, 1998). Here, we focus on the latter, where the stimulus feature changes with time. We simulate, in our Gaussian ring model, three different types of velocity profiles, (i) constant velocity, (ii) with acceleration and (iii) with an oscillatory velocity profile. We show that the population activity profile can lock to the given input and that in all the three cases, there is a quasi-linear relationship between the bump speed and input velocity. When modeling the effect of an external stimulus on the system, x_0 represents the feature for which the external input is maximal, i.e. the stimulus feature. We assume that the external input affects the system in the following general form (Hansel and Sompolinsky, 1998):

$$I_{ext}(x - x_0) = \mathcal{C}(1 - \epsilon + 4\epsilon\, e^{-\frac{1}{2}\left(\frac{x-x_0}{\sigma}\right)^2}), \tag{5.13}$$

where $\mathcal{C}$ is the maximal amplitude of the external input and ϵ stands for the stimulus tuning parameter. If $\epsilon = 0$, the input is homogeneous and if 0.5, it is tuned with a gaussian profile, to affect the nearby neurons and not those further away. Hence, in the all the simulation below, we take $\epsilon = 0.5$, in order to have an activity tuned to the stimulus. The velocity profiles we investigate are described below.

Constant Velocity: The input stimulus changes smoothly with time, with a constant velocity (Fig.5.6 (A)).

$$x_0(t) = \mathrm{v}_0 t. \tag{5.14}$$

Acceleration: presents with a changing speed profile of finite initial velocity and a constant acceleration (Fig.5.6 (B)), a_0:

$$x_0(t) = \mathrm{v}_0 t + \mathrm{a}_0 t^2. \tag{5.15}$$

Oscillatory Velocity : In this case, the velocity is oscillating, hence both continuous and changing (Fig.5.6 (C)). The peak position of the input current is thus of the following form:

$$x_0(t) = \frac{\pi}{2}\cos(\omega t). \tag{5.16}$$

The results show that the relationship between bump speed and input velocities in the three profiles is quasilinear. However, as (cumulative) input velocity in each profile increases it becomes harder for the bump to follow it. After a certain limit, after which no movement is detected. An example of such an event, is in the oscillating regime, where at a high frequency, the bump explores a much shorter trajectory than the input.

Intrinsic drive: Adaptation

Adaptation can make the bump move by generating a local, strong, delayed negative feedback, which suppresses localized activity (Hansel and Sompolinsky, 1998). This causes higher activity in the nearby unadapted region. Hence, adaptation not only generates different spiking patterns in the spiking neuron model (described above, AdEx), but also movement. The adaptation current obeys the dynamic equation below:

$$\tau_a \frac{dI_a(x,t)}{dt} = -I_a(x,t) + J_a\, m(x,t), \tag{5.17}$$

where τ_a is the time constant of adaptation and J_a its strength. We test the effects of changing J_a and observe that as it is increased, the bump moves faster. Also, unlike the other drives which can define the direction of bump movement, in this case, unless there is another asymmetry the bump can move in either direction (i.e. clockwise or anticlockwise).

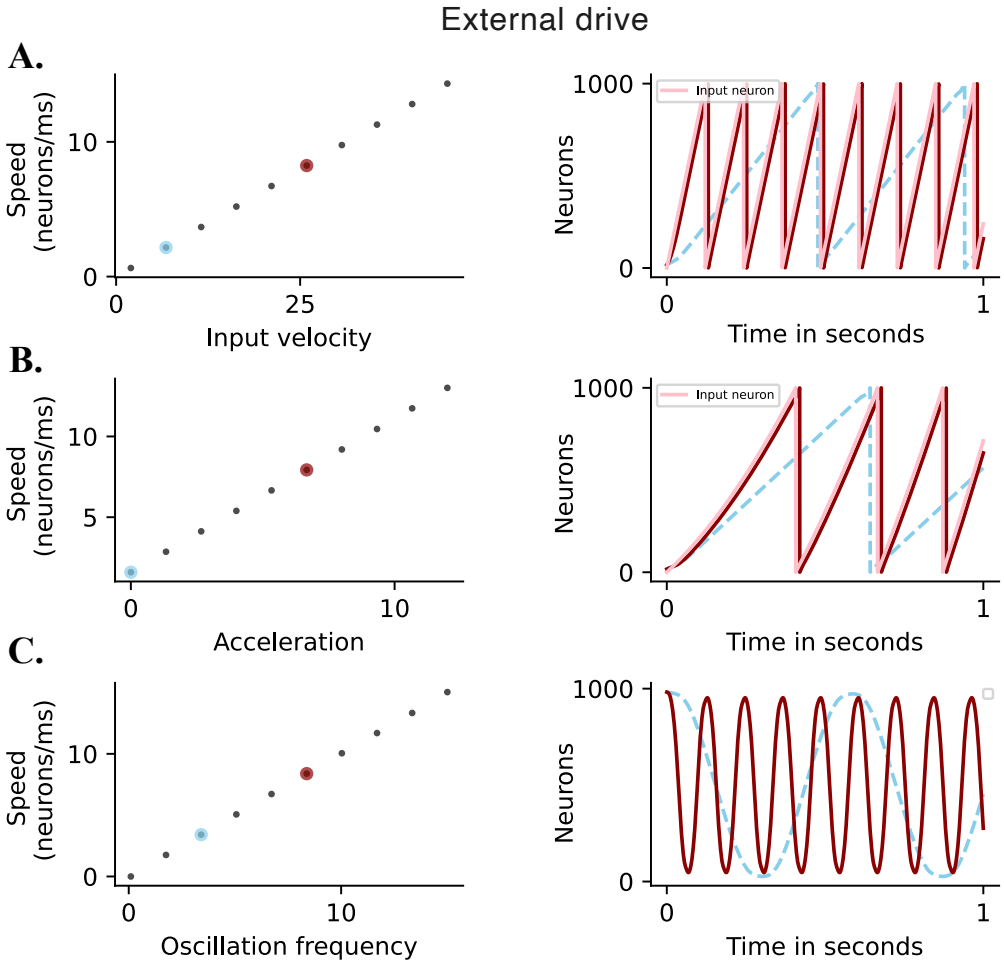


Figure 5.6: **Moving activity profile in response to different external inputs.** (A) The higher the velocity of the stimulus, the higher the speed of the bump, which follows it. (right) A depiction of two selected input velocities (color coded with left figure) and the population activity in time, represented by the center of mass of the bump. In pink the stimulus feature for the red velocity profile. (B) same as A but for accelerating velocity profile. (C) same as A and B, for the oscillatory speed profile.

Intrinsic drive: Bias

The bias term β was briefly introduced above and is present in Eq. 5.10, in the connectivity matrix. It provides asymmetry, ensuring network feed-forwardness and bump propagation. The sign of the bias determines the direction of propagation and the bump speed exhibits a quasilinear relationship with β, such that the higher the bias, the higher the speed of movement of the bump (Fig. 5.7 (B)). We choose the magnitude of the bias to acquire a constant speed, which alongside our

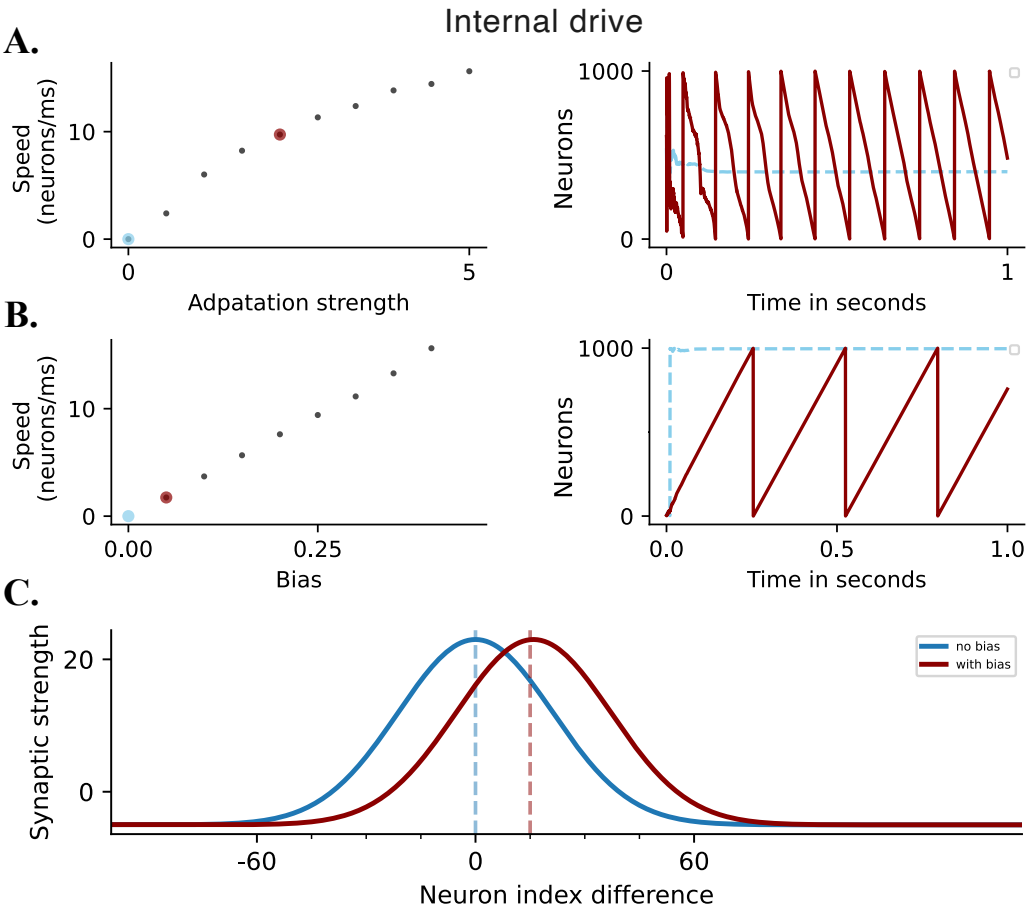


Figure 5.7: **Moving activity profile in response to different internal drives** (A) Response of the bump speed to different adaptation strengths is a quasi linear one. (right) A depiction of two selected input velocities (color coded with left figure) and the population activity in time, represented by the center of mass of the bump. (B) An asymmetric connectivity pattern makes the bump of activity move across the network. On the right, the population activity when $\beta = 0$ (blue) and when $\beta = 0.05$ are shown. In blue, there is no activity due to symmetric connections. (C) Connectivity profile for presynaptic neuron 0 with no bias and bias (color coded with figure B). This bias is the one we use for the rest of the results presented. Self-connections are set to 0 but not represented on the figure.

network size, allows a syllable duration of approximately 120 ms. The rest of the presented results are for an asymmetric connectivity profile and use a bias value of $\beta = 0.05$ (Fig. 5.7 (C)).

5.2.4 Asymmetric Connectivity Profile to Model HVC

The ring attractor model with an asymmetric connectivity profile gives rise to an activity bump, which propagates with a speed, defined by the bias β (as shown in Fig. 5.6). The transient population activation can be investigated by the width of the activity bump for one neuron. The width is affected by the neuronal model (rate) used, the width of the connectivity profile (σ) and the membrane time constant (τ). Regarding the second component σ, we have shown the possible values it can take for the system to remain in the bump regime (Fig.5.5) and we take the lowest of these values, for our chosen W_0 and W_2, to guarantee a narrow bump. To address the last contributing factor, we explore two values for the membrane time constant τ. With a $\tau = 1$ms, we can observe a bump width of $\sim$10 ms. Conversely, when $\tau = 1$ms the bump duration is $\sim$100ms. Despite the activity in HVC being much shorter than 100ms, we choose the latter value to stay consistent with previously reported membrane time constants. We illustrate the activity across time in Fig. 5.8 and use a Poisson process to generate spikes based on the firing rate.

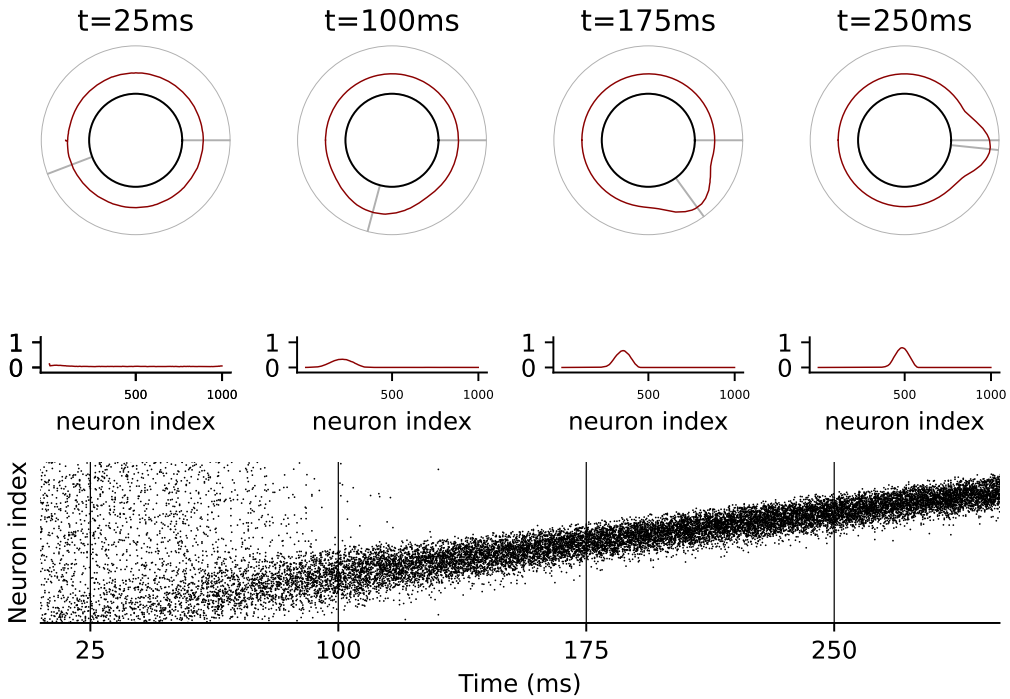


Figure 5.8: **Activity propagation in the ring.** The first two panels show the position of the bump of activity in the network at a particular point in time. The third panel serves to present the possible spiking pattern this rate network would be compatible with. They were generated using a homogeneous Poisson process.

To investigate how the bias term affects the parameter space where the bump is formed, we perform similar numerical analyses as in section 5.2.2. We obtain

four asymptotic states, and have two main observations, which help understand the effect of the bias. First, we observe in Fig. 5.9 (A), where $\beta = 0.05$, that the parameter space of the bump state is larger as compared to the system with $\beta = 0$ (Fig. 5.5). Second, in Fig. 5.9 (B), we see that as the magnitude of the bias increases, so does the region where the bump occurs, i.e. in regions where at a bias of 0, there is bump instability, at a bias of 0.05, there is a bump. We think this is due to the feedforwardness β induces, which constantly transfers the flux of activity to the nearby population. Hence, it does not allow for saturation in regions where a high enough W_2 would otherwise lead to a bump instability when $\beta = 0$. Taken together, these findings suggest that a $\beta > 0$, expands the parameter space where the bump is formed and sustained.

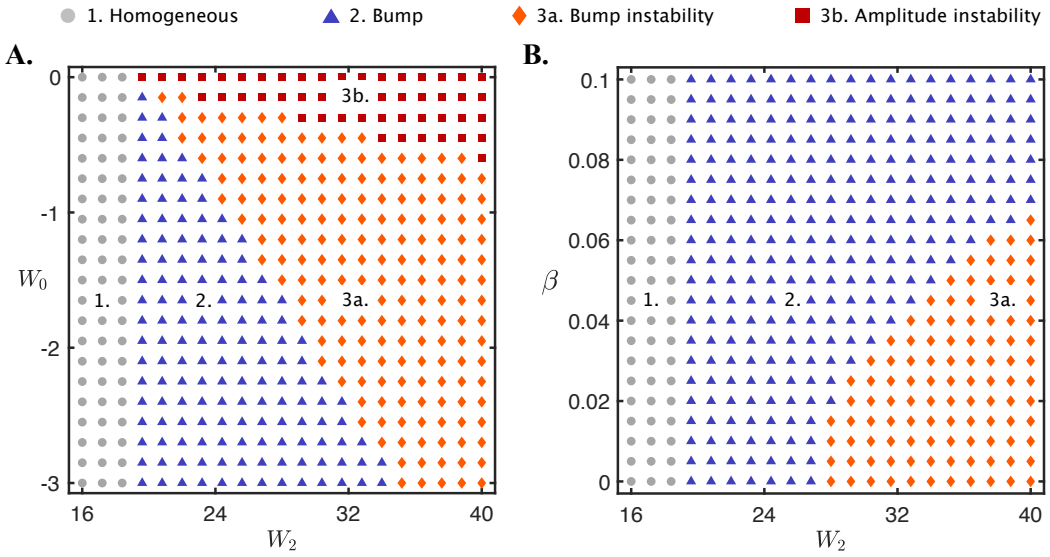


Figure 5.9: **Two numerical phase diagrams for the rate based model with an asymmetric connectivity profile, $\beta > 0$.** (A) The rate based model's phase diagram based on W_0 and W_2. (B) The rate based model's phase diagram based on on β and W_2. For this phase diagram, W_0 was chosen as a value of -3, correspondent to the last row of the the matrix in (A).

5.2.5 Discussion

Attractor dynamics have been used to model a wide range of cognitive processes including memory representation, sequence generation, decision making and integration. A group of criteria have been proposed to claim possible attractor dynamics in different networks in the brain (Khona and I. Fiete, 2022). HVC activity abides at least 4 out of 5 of these criteria, namely: (i) it possesses a low-dimensional set of states, timing in the song (1D) (R. Hahnloser et al., 2002), that corresponds to attractors in the state space, (ii) it exhibits robustness to perturbation and returns to the low-dimensional state after it. This was first shown by electrical stimulation (E. Vu

et al., 1994) in HVC. Although the stimulation perturbs song timing, once removed, the activity is quickly restored. (iii) Moreover, the states are invariant and persistent over time. In adult zebra finches, even in the absence of HVC main inputs (Nif and Uva) (Elmaleh et al., 2021; Naie and R. Hahnloser, 2011), HVC neuronal activity underlying song production persists. Also, HVC singing-related activity can be evoked outside singing, e.g. during sleep (Elmaleh et al., 2021), but it is partial. (iv) Lastly, the criterion of isometry, can be advocated by the neuronal activity of projection neurons and interneurons, being locked the song (Lynch et al., 2016). Song timing is driven by HVA_{RA} neurons, active at a $\sim 29-35$ms latency before feature (syllable onset/offset) appearance. As a continous activity spanning the song duration, it can be mapped to different instances in the song. Moreover, changing the distance in time between spikes, through heating and cooling experiments, has shown a mappable effect on the speed of the produced sequence (Long and M. S. Fee, 2008). The fifth condition, pertaining to anatomical and structural correlates, remains to be studied and investigated further. However, there are three reasons that make us speculate it may also apply to HVC. First, the neurons firing close in time, will likely have their connections strengthened due to synaptic plasticity and Hebbian learning. Hence, the underlying pattern of HVC_{RA} neurons would include reciprocal connectivity and rely on stronger connectivity between neurons firing at the same time in song. Secondly, as a local circuit encoding motor timing, HVC is expected to rely on regimes with strong internal connections capable of self-sustained activity (D. V. Buonomano and Laje, 2010). Thirdly, evidence from mammalian visual cortex show that neurons with similar tuning, exhibit stronger connections (Pattadkal et al., 2018). Moreover, in a recent model of the turtle's visual cortex, constrained by experimental data, the presence of rare strong synapses and many weaker synapses was shown. The former has a role in sequence propagation and the latter in its enhancement or disruption (Riquelme et al., 2023). These findings in the mammalian and turtle visual cortex may be similar for the premotor cortical area in songbirds.

In this section, we have shown the results of using a ring attractor to reproduce HVC dynamics. We have presented a robust model, functioning for a wide range of weights and have shown that we can generate a moving bump of activity. However, this model has the limitation of having a too long duration of the transient population activity. In this model, the duration of the burst of activity observed in HVC in zebra finches ($\sim$10 ms) (R. Hahnloser et al., 2002), can only be reproduced with artificially short neuronal time constants (1ms). Above, we have mentioned the three factors affecting the 'burst' duration: neuron model, width of the connectivity profile and membrane time constant. As we cannot modify the later two (section 5.2.4), it may be worth exploring a spiking neuronal model. Such model also guarantees a more accurate representation of the spiking dynamics in the network and the short-duration burst of activity spreading across the nucleus, which can be achieved by adding adaptation.

Some observations in the ring attractor lead to important questions in the songbird literature. For instance, activity propagation is possible through external

inputs, adaptation and asymmetric weights. The first could insinuate a higher structure driving the activity in HVC (Moll et al., 2023). The second, an intrinsic neuronal property could insinuate an effect on song timing which is not based on experience (Fehér et al., 2009a) (Araki et al., 2016a). It could also be a combination of intrinsic and experience based factors, presenting in the form of the bias in our model. This could be learned during sensory (auditory) learning from the tutor. Finally, songbirds are also known for being a good model for the neural mechanisms of vocal production in humans (A. J. Doupe and Kuhl, 1999). Therefore, similar neural mechanisms may underlie speech and song timing. The dynamics of cortical neurons driving speech production may thus also be accurately represented by the present model.

5.3 A Spiking Neural Network Model of HVC

As noted above, the use of a spiking neural networks (SNN) becomes indespensable when one aims towards a biologically plausible description of the HVC dynamics. Hence, in this section we investigate the conditions for which a ring attractor can be sustained in a SNN model. First, in section 5.3.1, we describe the neuronal dynamics with an adaptive and exponential (AdEx) model, as in the synfire chain model, section 5.1. However, in contrast to the synfire chain, the neurons here are placed in a periodic ring, exactly as in the rate model, section 5.2.

We investigate interactions with short-range excitation and long-range inhibition, where more specifically, we first focus on a simple rectangular connectivity profile in section 5.3.2, and then on a Gaussian connectivity profile in section 5.3.3. The Gaussian connectivity profile is the same as in the rate model in section 5.2. In the first case, we start with an exponential integrate and fire neuron and test the effects of adaptation, noise and synaptic delays on the system. In the second case, we investigate the conditions for which the Gaussian connectivity profile can allow for spontaneous bump formation in an AdEx neuronal model.

5.3.1 Model and Methods

We consider a ring population of $N = 3000$ AdEx integrate-and-fire neurons that are able to display bursting activity as it has been observed in HVC. The dynamics of each AdEx neuron $i \in \{1, 2, ..., N\}$ were previously presented in the Synfire Chain model (section 5.1). However, for completeness, we briefly review the corresponding equations below:

$$C\frac{dV_i}{dt} = -g_L\,(V_i - E_L) + g_L\,\Delta_T\,e^{\left(\frac{V-V_T}{\Delta_T}\right)} - w_i + I_{\text{syn}} + I_{\text{ext}} + \sqrt{\tau_n}\,\sigma_n\,\eta_i(t), \quad (5.18)$$

$$\tau_w\frac{dw_i}{dt} = a\,(V_i - E_L) - w_i + b\,\tau_w\,\delta(t - t_i), \quad (5.19)$$

where V_i and w_i are, respectively, the membrane potential and adaptation current. When the membrane potential crosses the threshold potential ($V > V_T$), the above parameters are updated:

$$V_i \to V_R = -45\text{ mV}, \quad (5.20)$$

$$w_i \to w_i + b, \quad (5.21)$$

where V_R is the reset potential. The update parameter in the adaptation current is the spike triggered adaptation, here set to $b = 0.01$ nA.

The synaptic input is defined as a function of the synaptic weights and the postsynaptic potential, and it reads:

$$I_{\text{syn}} = \sum_j W_{ij} \sum_k \Theta(t - t_j^k)\, e^{\frac{-(t-t_j^k)}{\tau_s}}, \tag{5.22}$$

where W_{ij} represents the synaptic weight between the pre-j and the post-i synaptic neurons. The single neuron dynamics are identical to those in the Synfire Chain model, section 5.1, where we have explained in detail every variable, function and parameter that is present in Eqs. (5.18)–(5.22).

In this section, we investigate the network dynamics of two types of connectivity profiles with short-range excitation and long-range inhibition. First, in section 5.3.2 we discuss the case with a symmetric rectangular connectivity profile, which reads:

$$\text{Rectangular:} \quad W_{ij} = \begin{cases} 0 & \text{if } i = j, \\ W_2 & \text{if } 1 \leq |i-j| \leq 10, \\ W_0 & \text{otherwise.} \end{cases} \tag{5.23}$$

In other words, Eq. (5.23) reflects that the neuron i will receive excitatory input, $W_2 > 0$, from its 20 nearest neighbors (in feature space) and an inhibitory input, $W_0 < 0$, from the rest of population. Autopses are set to 0.

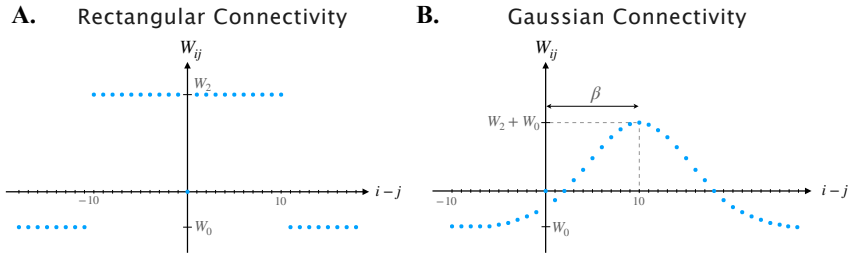


Figure 5.10: (A) Rectangular connectivity profile. (B) Gaussian connectivity profile.

In section 5.3.3, we then focus on a network with an asymmetric Gaussian connectivity profile, as in the rate-based ring model, section 5.2.4. More specifically, the corresponding equation reads:

$$\text{Gaussian:} \quad W_{ij} = \begin{cases} 0 & \text{if } i = j, \\ W_0 + W_2\, e^{-\frac{1}{2}\left(\frac{i-j+\beta}{\sigma}\right)^2} & \text{otherwise,} \end{cases} \tag{5.24}$$

where the bias, β, determines the asymmetry, and σ denotes the width of the profile.

Simulation

All simulations were run on macOS 13.2.1, with Python 3.9.7 and Brian2. The equations were integrated using an Euler scheme with a timestep $\Delta t = 0.1$ms. Multiple runs were performed to identify an appropriate value for Δt that does not alter the outcome. Parameter values for the simulations are detailed in Table 5.3, unless otherwise stated in specific sections.

Symbol	Parameter	Value	Symbol	Parameter	Value
N	number of neurons	3000	E_L	leak reversal potential	-70.6 mV
C	membrane capacitance	281 pF	g_L	leak conductance	30 nS
V_T	threshold	-50.4 mV	Δ_T	slope factor	2 mV
Δt	timestep	0.1 ms	W_0	global inhibition	-0.49 mV
W_2	excitatory factor	0.71 mV	β	bias	0.03
σ	width of connectivity profile	0.023	I_{ext}	external constant input	0.7 nA
τ_w	adaptation time constant	200 ms	a	adaptation coupling parameter	-0.5 nS
b	spike triggered adaptation	7 pA	$\mathcal{T}$	duration of simulation	300 ms
τ_{noise}	(time constant of the noise)	10 ms	σ_{noise}	standard deviation of noise	5 mV

Table 5.3: Values of the parameters used in Eqs.(5.18) - (5.22) and (5.24).

Center of Mass (COM) and Bump speed

Due to the architecture, size of the network and noise, several neurons may fire ($\sim$1-6) simultaneously. Hence, to acquire a one dimensional position vector of the neural activity across time, the firing rate of the population is needed. The position $C(t)$ (center of mass) of the bump can be computed as the maximum value of the population's mean firing rate. From the spike indices and times (extracted from Brian2), we compute a binary matrix (spike/ no spike), which is convolved in time and space using a two-dimensional filter with a Gaussian kernel of standard deviation 5 ms and 50 neurons respectively. The peak of the activity bump encodes time such that at time $t(x)$, neuron x is maximally activated. As the bump of activity moves across the network different neurons become more active progressively (Fig. 5.8). The center of mass $C(t)$, is further discretized into $\bar{C}(t)$ such as to coincide with the nearest unit position:

$$\bar{C}(t) = \text{argmin}_i |C(t) - x_i|. \tag{5.25}$$

The speed of the bump is then computed as the displacement (in the neuronal feature space) of the center of mass over time. Alternatively, the speed can be calculated using the amount of time between two neurons' first spikes and their distance in feature space.

Syllable definition and duration

A syllable corresponds to a fixed segment of the ring. We choose to simulate a syllable duration of 50 ms and for simplification, we chose equal size segments

such that for s syllables and n neurons, syllable i is defined by $[\frac{i}{s}n, \frac{i+1}{s}n]$. Syllable duration boundaries are defined based on the timepoints when the peak of the spike-computed activity bump $\hat{C}(t)$ crosses the lower and the upper limit of the segment.

5.3.2 Symmetric Rectangular Connectivity Profile

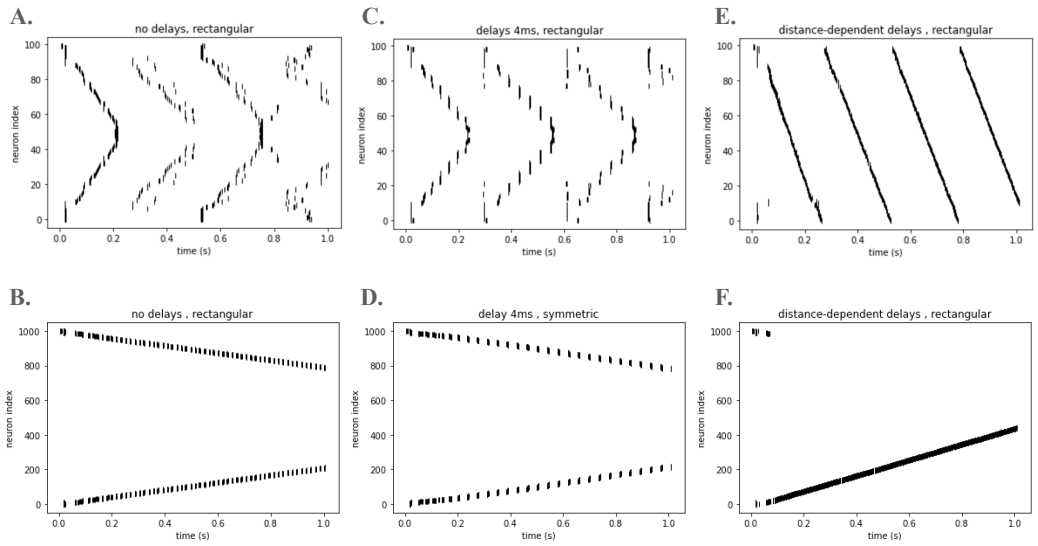


Figure 5.11: **Activity Propagation in a Symmetric Ring Model with a Rectangular Connectivity Profile and AdEx Neurons** (A,C,E) Network of a size of 100 neurons, across 3 different conditions: with no delays, with equal delays and with distance based delays. (B,D,F) same as the previous panels for a larger network (1000 neurons).

In this part, we describe an exploration conducted in the spiking model, using the symmetric rectangular connectivity profile as defined in Eq. (5.23). We investigate the system in the absence of noise. Due to the effect of adaptation (Hansel and Sompolinsky, 1998), we expect the formation of an activity bump that propagates in one direction in the ring. Instead, we observe two bumps of activity being formed and evolving in two different directions (Fig. 5.11). We illustrate the bump generation described in networks of two different sizes, with 100 and 1000 neurons. This effect may be a result of the mechanism of adaptation, making a local inhibition at the neuron level and a relative heightened activity in the adjacent non-adapted areas. Given that the neighboring regions in both directions are equivalent due to the network's architecture, it's likely that both have comparable activity levels, leading to the creation of two distinct bumps. However in the presence of slight asymmetry, i.e. through noise, we would instead observe only one propagating bump.

Next, we explore the effect of synaptic delays in the model. Constant delays show very similar results (Fig. 5.11 (C,D)), with a delay between spikes. However, there is still no asymmetry. In contrast, adding distance dependent delays would guarantee that. To be able to do that, we first need to map our one dimensional ring to a three dimensional space, which has the shape of a sphere and a radius of $\pi/2$. We make no assumptions about how the feature space relates to the actual space, other than it being random. Hence, each neuron $\{0, ..., N\}$ in the feature space is randomly assigned to one in the 3D space. We then calculate the Euclidean distance between all pairs of neurons and multiply the result by a factor of 2.5ms, to achieve a maximum delay of 7.9ms as reported previously. This adjustment aligns with experimental data that indicates conduction delays varying between 0 to 7.9ms (Egger et al., 2020). As expected, upon adding distance dependent delays a single bump, propagating in one direction, is formed. Distance dependent delays provide asymmetricity as undoubtedly one neuron will be closer in space than the other. The direction is determined by the closest neuron (Fig. 5.11 (E, F)).

5.3.3 Asymmetric Gaussian Connectivity Profile

In this part, we investigate an asymmetric gaussian connectivity profile as defined in Eq. (5.24). Although there could be similarities in behaviour between a rectangular and gaussian profile, it is important to investigate the latter as well. A reason for this is to be consistent with the connectivity profile in the rate model. Moreover, it is likely that the synaptic weights are 'graded' rather than binary, and lastly because of the central limit theorem, stating that the distribution of a random variable converges to a normal distribution as the number of samples increases.

We first report and illustrate the neural dynamics and then investigate the conditions where such dynamics are possible. When using parameters inspired from biological systems, we observe the formation of an activity profile that drifts along the ring where every active neuron bursts in 3-6 spikes for a typical duration of ≈ 10 ms (see Fig. 5.12). This matches experimentally-observed neuronal dynamics (R. Hahnloser et al., 2002).

5.3.4 Conditions for a Drifting Bump

In this part, we perform numerical analyses of the SNN model to investigate the parameter space for which a drifting bump forms due to neuronal interactions. Similar to the analysis of the rate-based model, we identify the distinct asymptotic states and their corresponding regions in the W_0 – W_2 space.

We run simulations of exponential integrate and fire neurons by combining Eqs. (5.18) (5.20) (5.22) & (5.24). Adaptation current dynamics, Eqs. (5.21) (5.21) is excluded here in order to allow easier translation/mapping to the rate model phase diagrams. However, inclusion of adaptation likely does not affect the phase dia-

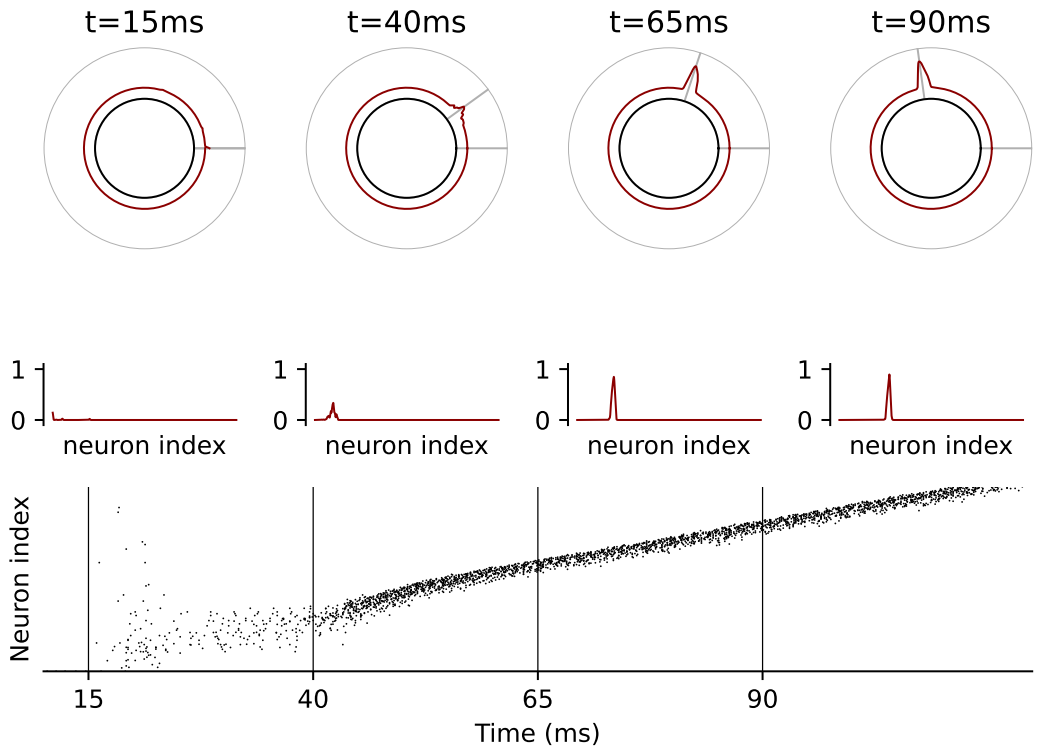


Figure 5.12: **Activity Propagation in an Asymmetric Ring Model with a Gaussian Connectivity Profile and AdEx Neurons** The first two panels show the position of the bump of activity (the computation of which is present in 5.3.1) in the network at a particular point in time. The third panel is the raster plot, showing the brief bursts of activity and the propagation in the network.

grams significantly; it would possibly only broaden the area of the stable regime (D. Jin et al., 2007).

The model was initialized with a constant input and a "pinch" input applied to only one neuron (neuron index, 500) for 15ms. Then it was run for a long enough time; a time after the initial transient formation has passed and the stable state is present. At the last time-step, we observed where the system settled and identified four distinct regimes:

- In Region 1 , the system converges to a homogeneous regime (Fig. 5.13 (A) in gray dots).
- In Region 2 , the system converges to a stable/stationary bump (Fig. 5.13 (A), in light blue dots).
- In Region 3, the system converges to a moving bump. Two different moving bumps are observed,
  - a "hopping" bump (purple triangles in Fig. 5.13 (A)) and
  - a continuously moving bump (blue triangles).

• In Region 4, the system diverges to bump or amplitude instability (orange and red).

Determining the boundary of the fourth region was challenging, so we decided to call anything above a certain maximum spike number a bump instability. However, while this is acceptable for our particular model, where the spike number should be below 10, it would need another criterion for a more general model. Perhaps as shown in the last panel of Fig. 5.13 (B) in red, the amplitude instability can be determined when the mean firing rate of the last timestep appears as saturated, with a flat top. This is also similar to the criterion used for the rate based model.

Moreover, the presence of a stationary bump and "hopping" bump state are interesting. The first, is likely dependent on a interaction between bias and excitation (W_2). Although the center of mass of the activity settles in one place, the bias seems to keep driving neuronal movement within the region itself. As excitation increases, the "hopping" bump appears. It is a state between stationary and bump state, and has characteristics of both. The center of mass moves in time, but it stays in one centered location for a prolonged time and a large number of neurons are involved ($\sim$ 100). Here, the interaction between bias and recurrent excitation, allows activity propagation. However, these profiles have not been further investigated in this manuscript.

The parameters are chosen from region 3b, where a continuously moving bump emerges.

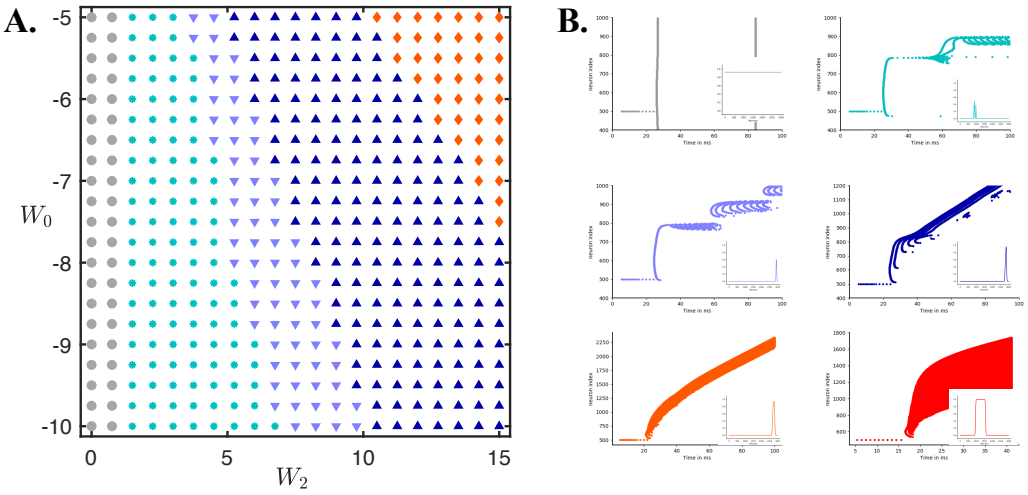


Figure 5.13: **Phase diagram of an Exponential Integrate-and-Fire neuron model in an asymmetric (with bias) Ring Attractor Architecture.** (A) The phase diagram of the five possible regimes. (B) The raster plots of each of the observed states and their respective activity profile in the last timestep of each simulation. The plots in (B) are color coded to the phase diagram.

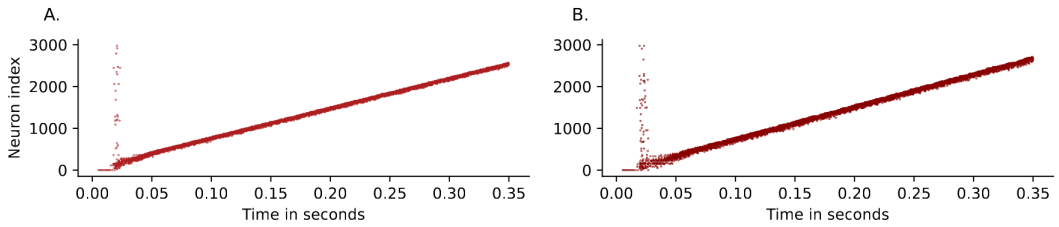


Figure 5.14: Robustness of the Ring Model. (A) Behavior of the synfire chain in face of a change in synaptic weights, i.e. 30% decrease. There is no stop in propagation, even when all connections are affected. (B) Behavior of the synfire chain in face of an external inhibition, amounting to 10% of I_{ext}. There is no stop in propagation when 50 percent of the population is affected.

Robustness

In addition, we use the same two robustness protocols reported in the Section 5.1 in order to also be able to make a comparison of the robustness of the model with the Synfire Chain. As it was illustrated in Fig. 5.3, with the administered binary noise magnitude in the synaptic weights and input, the Synfire Chain stops activity propagation even when a small proportion of synaptic weights and neurons are affected. In the case of the Ring Attractor, using the same amplitude as used in the synfire chain, even at an 100% synaptic and neuronal involvement, there is almost no effect (Fig. 5.14).

This can be expected from the recurrent architecture present in the ring, as compared to the purely feed-forward one in the synfire chain. Moreover, as the input affects randomly neurons in the network, their absence can easily be compensated. Lastly, unlike in the Synfire Chain, the weights can be pulled from a wide parameter range of inhibitory (W_0) and excitatory (W_2) values to bring forward the bump state (Fig. 5.13). In the case of a perturbation, such as one caused by noise or a robustness paradigm as we described, the network 'drowns' it across time and goes back to its marginal state.

5.3.5 Predictions from the Models

We designed an additional protocol to evaluate the robustness of the network under possible biological perturbations and the effect of inhibition administered directly at the HVC nucleus level. We simulate the local injection of a drug inducing the inhibition of neuronal activity (e.g. Muscimol), optogenetic stimulation/cholinergic agonist, or Gabazine injection, with effects that spread with a Gaussian spatial distribution into the network, as shown in Fig. 5.4 (C, D). We assume this external perturbation is constant in time and we add it to the main equation (5.18). For the simulations of Muscimol injection and optogenetic stimulation, the equation reads:

$$I_i = \alpha \, e^{-\frac{1}{2}(\frac{x-\mu}{\sigma_i})^2}., \tag{5.26}$$

where α is the amplitude and σ_i the width of distribution in space. In contrast, in the Gabazine simulation, as an antagonist of gabaergic receptors, the global inhibition parameter W_0 is affected. In this protocol, the additional input is a function of the neurons' location in the three dimensional space. To simulate a lateral inhibition/activation to HVC, we randomly choose a neuron from those most distant from the center of the sphere, and use it as the peak of the gaussian (input). Then, we compute the Euclidean distance between this neuron and each neuron in the sphere, which constitutes the $x - \mu$ line in Eq. (5.26). The outcome is a one dimensional array, which is then added to the main voltage equation.

We test different distribution widths (σ) and input amplitudes (α). Moreover, by changing the sign of α in Eq. (5.26), we are able to model two possible effects at the level of $GABA_A$ receptors; with a negative sign that of the $GABA_A$ Agonist, Muscimol, which enforces inhibition and with optogenetic stimulation, which serves to minimize the inhibition.

Inhbition: $GABA_A$ Agonist (Muscimol)

Muscimol is a drug that acts as an agonist at the $GABA_A$ receptors and it is widely used in experimental studies. We simulate its effects in the rate based model and in the spiking model and observe a visible effect in bump formation. Due to the high level of disruption, there is a delay in sequence initiation (Fig. 5.16 (A)). The amount of delay is linearly related to the amplitude and width of the inhibition. Moreover, although the sequence absorbs the perturbation and is maintained for the rest of the simulation, a decrease in the speed of propagation is present. Speed has a quasi linear relationship (shown in Fig. 5.15) with the amplitude and standard derivation, such as when one of the two increases, the speed of the bump decreases. The standard deviations in Fig. 5.15 are expressed as a function of the size of the system. The limit above which the bump does not recover is defined at $\sigma = 0.9$ (in a sphere with radius equal to $\pi/2$) and an amplitude $\alpha = -14$ nA.

Overall, these findings suggest potential predictions regarding HVC behavior in the face of such inhibition. The effect might be a delay in the initiation of the sequence (song), without any subsequent disruption in the propagation of the sequence, as depicted in Fig. 5.16. Nonetheless, bump speed propagation decreases (Fig. 5.15, middle), leading to a compromised timing, i.e. slightly stretched time.

It is particularly difficult to test the first hypothesis/ prediction of the model, as a song initiation time would need to be defined, from which the delay is to be computed. A possibility is to compute an average reaction or singing time in response to female presentation (directed singing) or with respect to the introductory notes. Subsequent changes in this reaction time, in the context of Muscimol inhibition, could then be studied.

Our second finding is easier to test, although a mapping of our amplitude values to drug dosages, is necessary. It could be possible that the allowed dosage maps to a very low amplitude in our simulation, reflecting in a very small speed decrease,

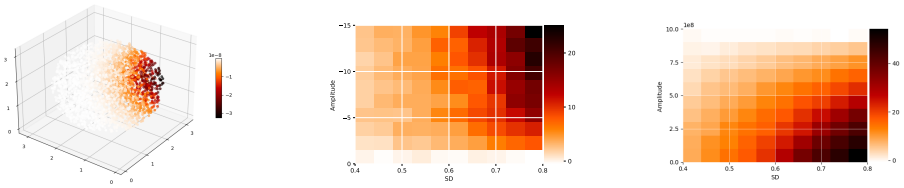


Figure 5.15: **Effect of Inhibition on the Speed of the Bump.** On the right, an illustration of the distribution of the inhibitory input in the 3D space is shown. In the middle and on the left, respectively, Muscimol's and Gabazine's effect on the bump speed are investigated, for different amplitudes and distributions (width). The amplitude is expressed in nA and is the coefficient of the gaussian, whereas the SD is expressed based on the size of the system (π). The colors represent the change in speed as a percentage of the inital (uninhibited) propagation speed.

i.e. very high doses could completely degrade the song (Isola et al., 2020). It is possible that a stretching effect similar to the one reported (Isola et al., 2020) in Bengalese finches upon muscimol injections is present.

Optogenetic activation

In our model (rate, SNN), in terms of the initial response to this input, we observe the same effect as with Muscimol- mimicking simulations. The bump takes longer to form, i.e. there is a delay in sequence initiation. In the spiking model, the network is more sensitive to higher amplitudes of excitation as compared to the rate model. Moreover, the speed profile has a nonlinear relationship to the input's size and/or magnitude. When the amplitude is kept constant and we explore different σ_i, there is a linear relationship (increase in speed) up to a certain limit, after which there is no bump formation. Upon crossing a second limit, the linear relationship reverts. This suggests that the excitation of such kind should be fine tuned and can at different distributions, provide different results. However, this remains to be further investigated.

Inhibition: $GABA_A$ Antagonist (Gabazine)

Gabazine is a drug that acts as an antagonist at the $GABA_A$ receptors, and results in a reduction of GABA-mediated synaptic inhibition, namely W_0. Our findings with the gabazine mimicking simulation (Fig. 5.17), show an effect in speed, which is quasi-linear and in the opposite direction, compared to the effect with Muscimol. Hence, when the amplitude or standard deviation of this inhibition is increased, the speed increases (Fig. 5.15, left).

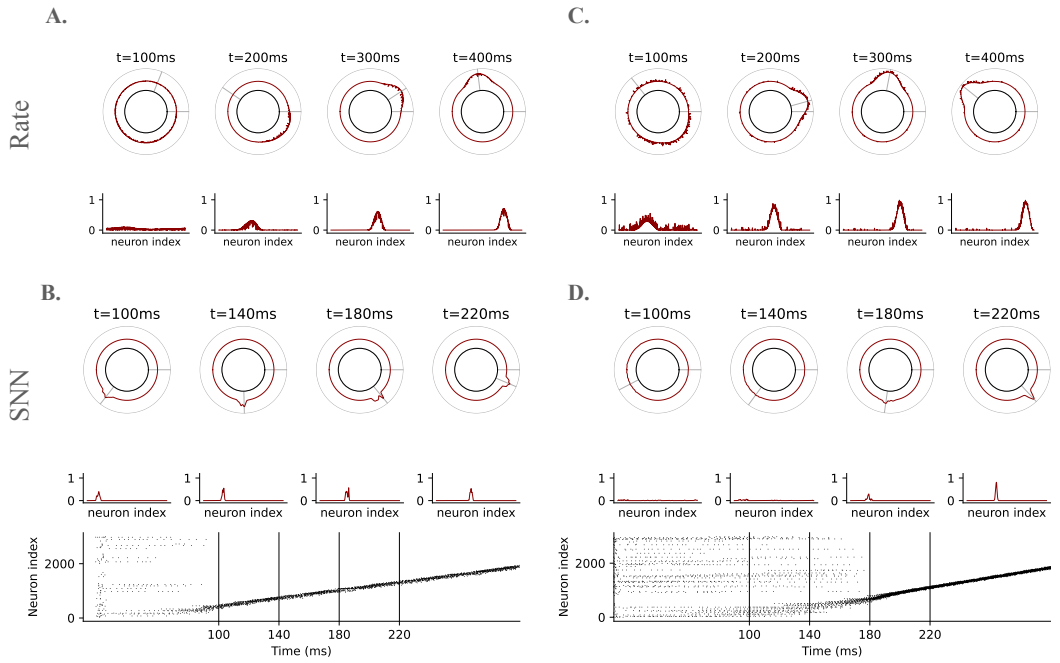


Figure 5.16: **How does inhibition affect bump propagation?** Here we present two example cases of the effect of moderate inhibition in the rate and SNN model. In (A,B) we illustrate a Muscimol-mimicking simulation, both in the rate model (A) and in the SNN (B). We observe a disruption in the beginning, but similar to a noisy perturbation, the model is able to recover after a while, although the inhibition is constant. (C,D) In these panels, we illustrate optogenetic stimulation-mimicking simulation, both in the rate model (C) and in the SNN (D). In both cases we observe only a disruption in the beginning.

5.3.6 Discussion

HVC neuronal recordings reveal a sparse network activity with bursts of 3-6 spikes lasting 6-10 ms, time-locked to the song. In this section, we have introduced a ring attractor model that captures the sequential neuronal activity of HVC while preserving the individual neuronal properties. We use spiking neurons and since previous evidence (Long, D. Z. Jin, et al., 2010) has shown that HVC bursts are mediated by calcium spikes, we add neuronal adaptation as a possible mechanism to capture this. More specifically, adaptation, modeled by a differential equation first presented in Brette and Gerstner (2005), can account for ionic channel activity, depolarizing currents (captured by parameter a in Eq.5.19) and calcium dependent potassium channels (captured by parameter b).

Our results provide evidence that an asymmetric ring attractor model with adaptive exponential integrate and fire neurons can explain the underlying mechanisms motor timing in songbirds. The model is robust, able to generate the same spiking/bursting activity and activity propagation in the network. In terms of

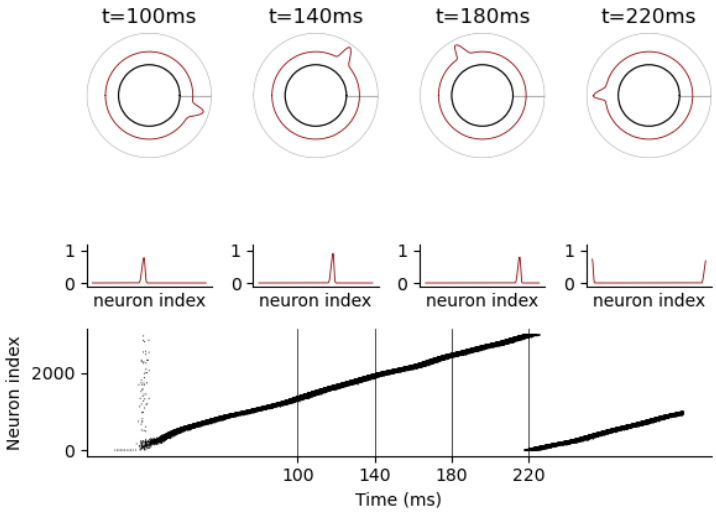


Figure 5.17: **How does inhibition affect bump propagation?** Illustration of a case of Gabazine-mimicking simulation in the SNN model.

robustness, we first compare our model with a traditional synfire chain, where weights have to be fine tuned to support propagation. We perform numerical simulations to compile phase diagrams, and show that the regime where the bump is formed, the marginal state, is quite broad. Hence, the weights do not need fine tuning. Moreover, we investigate their robustness with two protocols, described in Subsection 5.1. We show that while the synfire chain shows a high sensitivity to random partial decreases in external input and synaptic weights, the ring attractor remains mostly unaffected and is robust to noise, even at high levels of neuronal and synaptic involvement. We only witness a slight delay in activity initiation. This is consistent with the literature on attractors where noise is absorbed by the system, although it may slightly disrupt it in the beginning (Khona and I. Fiete, 2022).

A possible further exploration of our model is investigating behavioral adaptation, which entails learning to modify the duration of a syllable, without affecting the other syllables. This has been shown to be the case in songbirds (Ali et al., 2013). In chapter 6, we will simulate such a behavioral learning paradigm, and present the results.

5.4 Extended model with thalamic (Uva) Input: Synchrony?

A number of studies have investigated and focused on providing a better understanding of the role of Uva in song production. Several models of its function have

been proposed as a: (i) a synchronizing center for the two hemispheres (Ashmore, Renk, et al., 2008; Gibb et al., 2009; Schmidt, 2003; E. Vu et al., 1994), (ii) an intermediary transmitting important song-related information or feedback from the brainstem (Ashmore, Renk, et al., 2008; Hamaguchi, Tanaka, et al., 2016), and (iii) an initiation hub (Alonso et al., 2015) at discrete times. According to recent evidence (Moll et al., 2023), a percentage of HVC neurons are Uva-driven and fire at particular times during the song, namely at the beginning and/or end of the syllable. However, the exact role and significance of these neurons and the incoming input is still a topic of discussion.

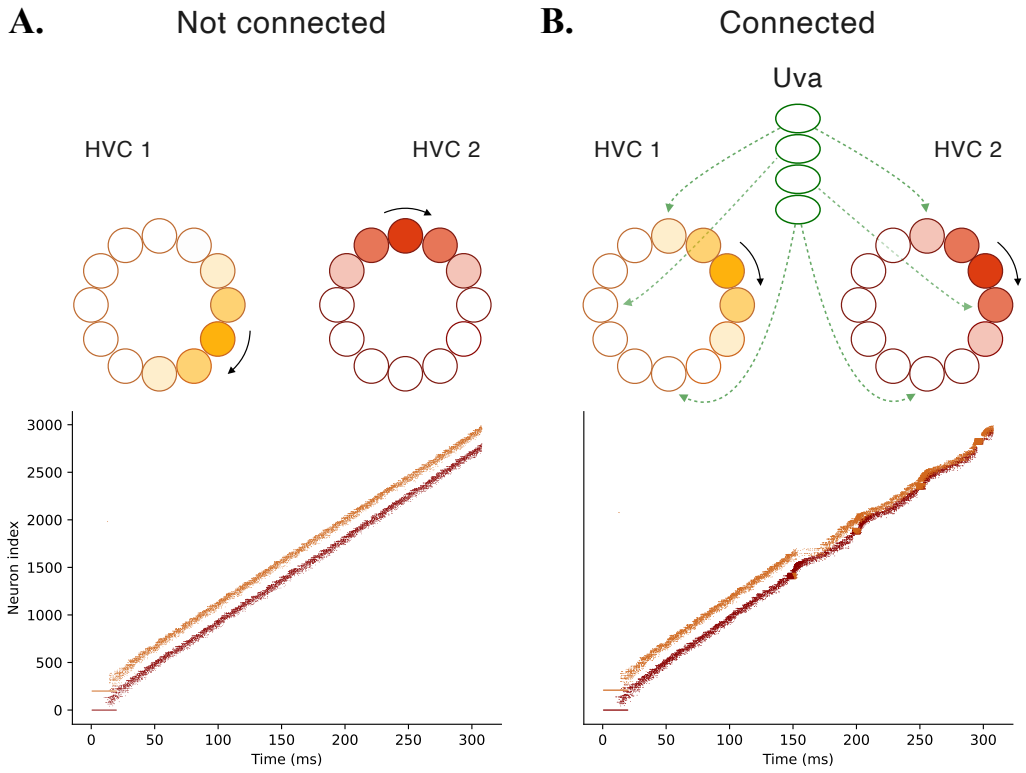


Figure 5.18: **Uva as a possible synchronizing nucleus between the HVCs of the two hemispheres.** In (A) HVCs of the two hemispheres are not connected. When the sequences of each are initiated at different times, they evolve in parallel, with the same speed and no convergence. In (B) the two HVCs receive input from Uva, which is able to make their respective sequences converge, making one's speed go faster and the other's slower, until they fire synchronously.

A prominent hypothesis is that Uva plays a synchronizing role between the two HVCs. Cooling studies have shown that the effect of such manipulation in either HVC manifest with song stretching. This suggests both HVCs are equally important and/or complementary in controlling song timing (Long and M. S. Fee, 2008). However, synaptic noise or other physiological factors could drive the two

HVCs to present with slightly different activity, which could affect timing accuracy. Hence, the information coming from the two HVCs could be, synchronized through a higher level structure, the thalamic nucleus Uva. The presence of such input at the syllable level (Moll et al., 2023) could be affected by respiration.

We simulate 2 spiking neural networks, with the same properties as described above, and add a sequential input representing the input originating from Uva. To reproduce experimental findings, only 15 % of neurons receive input from Uva (Moll et al., 2023) and these are the neurons firing at the beginning and end of each syllable. We first, investigate the behavior of the system in a high noise regime and observe synchronization. However, to be able to illustrate this effect better, we run a dedicated simulation where the starting input is different in the two HVCs, similar to what has been shown in Pang et al. (2022). As one sequence starts from neuron index 200, instead of neuron index 0, the signal is ahead of the other HVC's signal. Whereas in the case of no external intervention, these two signals would always be separate from one another, evolving in parallel (Fig. 5.18(A)), here, due to the presence of a synchronizing unit, after a short while the signals converge (Fig. 5.18(B)), proving a consistent input to RA.

5.5 An EI Network Model of HVC

In the previous sections, we investigated the behavior of the model in a one population model, in the presence of global inhibition and excitatory and inhibitory synaptic connections. However, interneurons likely have a significant importance on the produced dynamics. To investigate the effect of inhibition of a more complex form, we expand our model to a 2-population model. We consider an excitatory ($\mathcal{E}$) and an inhibitory ($\mathcal{I}$) population, both with single neuronal dynamics described by Eqs. (5.18) (5.20) (5.22) & (5.24), without adaptation currents. The only difference is that the inhibitory populations membrane time constant is twice that of the E population. The size of the ($\mathcal{E}$) and ($\mathcal{I}$) populations are 3000 and 600 neurons respectively, with a ratio of 5:1 (Fig. 5.19 (A)). The connections between the neurons in the excitatory population are guided by a ring attractor with a Gaussian connectivity profile, same as in the models presented above (Eq.(5.10). However, W_0, global inhibition is absent, as inhibition is now in a loop with the excitatory neurons.

All parameter values are same as in Table 5.3. Moreover, among themselves, the inhibitory neurons are not connected. The synaptic weights from the ($\mathcal{E}$) to I population are set 0.015 mV and a connection probability of 0.79. The synaptic weights from I to E are set to 0.49 mV, with a connection probability of 0.67. Connection probabilities were set based on previous findings (Kosche et al., 2015). The magnitude of synaptic weights (EI, IE) is set to enable proper activity propagation in the excitatory population and allow burst of $\sim$ 10ms duration, 4-6 spikes. Although, in the presented model, we do not explicitly propose a structure for the inhibition, there is an emergent structure originating from its interactions with the excitatory

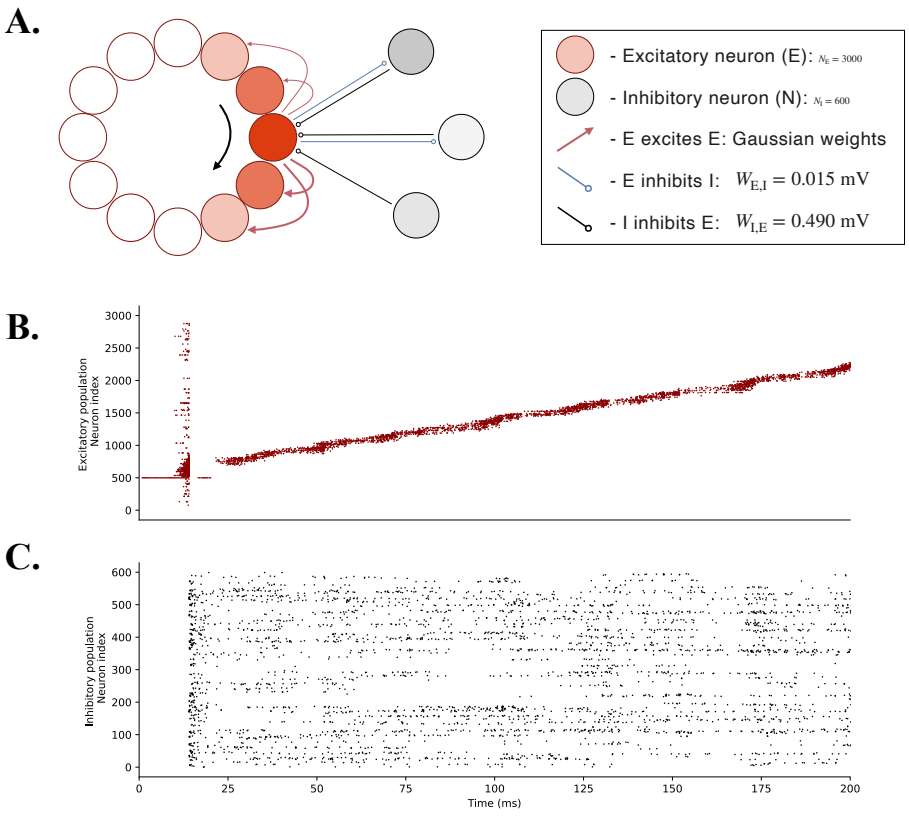


Figure 5.19: **Activity propagation in the EI model.** (A) Illustration of the network architecture. (B) Raster plot of the excitatory population. (C) Raster plot of the inhibitory population.

population. We see a spiking pattern in both populations (Fig. 5.19) consistent with recordings. Whereas we had set the parameters to acquire the spiking pattern of the E population, we presented no additional constrains for I. The mean firing rate of the inhibitory population is oscillating at ∼30Hz, matching the one reported from experiments (Markowitz et al., 2015).

6 A Flexible Model of Motor Timing: Behavioral Adaptation

Behavioral findings using the Conditional Auditory Feedback (CAF) paradigm (described in Ch. 2) have shown that when an interval (syllable + gap) in a songbird's song is targeted for lengthening (shortening), this occurs with specificity (Ali et al., 2013; Pehlevan et al., 2018). The songbird is able to change the duration of that target interval, without significantly affecting other interval durations. Hence, song length increases/decreases as an effect of the change in one interval. This supports one of the three previously proposed hypotheses. The other two are: (i) global modification, where all the elements of the sequence change (they all lengthen, when one lengthens) and (ii) interference, where the duration modification affects the other syllables, such as to keep the song length constant (i.e. when one syllable lengthens, the others shorten).

How the reported specificity in behavioral adaptation can be captured from a computational perspective poses an interesting question, which can provide insight on the underlying mechanisms. In a recent article (Pehlevan et al., 2018), three different models were explored: synfire chains, dynamic attractors (Laje and D. V. Buonomano, 2013) and feedback-stabilized RNNs (fsRNN) (Rajan et al., 2010; Sussillo and Abbott, 2009). It was shown that only synfire chains are able to conduct interval modification in both directions (shortening, lengthening), without causing any interference (effect on the other intervals). In contrast, with fsRNNs there is

interference in all the other intervals and in the attractor model there is interference in the adjacent ones. However, since our proposed attractor model, the ring model is able to capture and provide explanations for HVC_{RA} neuron behavior, we decide to further explore its behavior when presented with a simulated version of this learning paradigm. We use a reward covariance learning rule (R. Williams, 1992) and adapt the simulation to the experimental conditions presented (Ali et al., 2013). The used formalisms and adaptations are consistent with the computational methodology used in a previous study (Pehlevan et al., 2018). In this chapter, we present our findings at four levels of increasing complexity using first, a one population network with (i) rate-based neurons, (ii) exponential integrate and fire neurons (eIF), (iii) adaptive exponential integrate and fire (AdEx) neurons and (iv) a two population network (EI) with exponential integrate and fire neurons. Lastly, we make a comparison between our numerical findings and recent behavioral data collected in our lab.

6.1 Reinforcement Learning: Numerical Implementation of the CAF experimental Paradigm

We use a reward covariance learning rule (R. Williams, 1992) to simulate the CAF paradigm. The equations reads:

$$\Delta W_{ij} = \gamma \; R \; e_{ij} \tag{6.1}$$

$$\text{with} \qquad e_{ij} = \int_0^t \frac{dt'}{\tau_e} \, e^{-(t-t')/\tau_e} \, \eta_i(t') m_j(t'), \tag{6.2}$$

where ΔW_{ij} denotes the learning-induced change in the synaptic weight between the pre- (j), and postsynaptic neurons (i). γ denotes the learning rate, R the reward value and e_{ij} the eligibility trace. The eligibility trace (Eq. (6.2)) controls the synaptic change amplitude and flags whether a synapse is *eligible* for modification, based on recent activation of the plastic synapse. The eligibility trace used in our model is computed based on the target syllable [1], where (Eq. (6.2)) 0 is the start and t the end of the syllable. Moreover, it relies on a time constant τ_e, the noise of the postsynaptic neuron η_i and m_j the firing rate of the presynaptic neuron.

In our reinforcement learning simulations we set the learning rate to $\gamma = 0.0002$, and the eligibility time constant to $\tau_e = 35$ ms. γ was chosen to match learning rates observed in songbirds experiments and tau_e to match the delay between HVC signal and motor action (Ali et al., 2013). The reward R takes a value of 0 or 1 when the syllable is targeted for modification, and 0 when it is not targeted.

[1] In the rest of this chapter, we use the term syllable instead of interval to maintain consistency with the terminology used in Chapter 5. Moreover, we have defined no gaps in the model, so what we refer to as syllable, encompasses them.

If targeted for modification, the syllable duration at the end of each trial is compared to the running average of the target syllable duration. This is updated in each trial as shown in (6.3):

$$\bar{I}^{(\text{tar})} \leftarrow 0.995\bar{I}^{(\text{tar})} + 0.005I^{(\text{tar})}, \tag{6.3}$$

where $I^{(\text{tar})}$ denotes the current syllable duration and $\bar{I}^{(\text{tar})}$ the running average of the target syllable duration. When the syllable is targeted for lengthening, if $I^{(\text{tar})} > \bar{I}^{(\text{tar})}$, a reward is given ($R = 1$), and the weights are updated according to the learning rule (6.1). If $I^{(\text{tar})} < \bar{I}^{(\text{tar})}$, no reward is given, $R = 0$, and the synaptic weights remain unchanged.

The RL pseudoalgorithm when the syllable is targeted for lengthening

```
for i = 1 to 50 do
    run(model)
    duration ← find_syllable_duration()
    baseline[i] ← duration
end for

for i = 1 to 1000 do
    run(model)
    duration ← find_syllable_duration()
    duration_target← duration_target × 0.995 + duration × 0.005
    if duration > duration_target then
        ΔW ← R × γ × e
    else
        ΔW ← 0
    end if
    W ← W + ΔW
end for

for i = 1 to 50 do
    run(model)
    duration ← find_syllable_duration()
    postrl[i] ← duration
end for
```

Simulations

All other simulation parameters unrelated to the learning for each of the models (rate, eIF, AdEx), remain consistent with those detailed in their respective sessions in Chapter 5. Prior to implementing the learning rule, 50 trials are run to determine the baseline distribution of the syllables. Then, the learning trials are run. In

the rewarded trials, the weight matrix, W is updated by ΔW_{ij}. Lastly, 50 more trials are run to determine the distribution of durations with the new weight matrix W ($W_{initial} + \Delta W$). A simplified representation of the followed protocol is presented in the pseudoalgorithm. Lastly, the baseline and post-reinforcement (postrl) syllable distributions are compared using independent t-tests. Significance is reported when $p < 0.01$.

6.2 Learning in the Rate-Based Model

We first explore the effects of our learning rule and paradigm, with a one population model of rate based neurons (5.2). Rate based neurons models require less computational power as compared to spiking neurons, can give valuable intuition, and guide explorations in the spiking model. In this section we report the results of the presented reinforcement learning ('simulated' CAF) protocol.

6.2.1 Results

As explained above, we run 50 trials prior to and following learning, to determine duration distribution with the initial weights, and with the weights modified by the learning rule. With the used noise magnitude, we acquire a mean baseline duration of 118.6 ms, with a standard deviation of 0.75 ms.

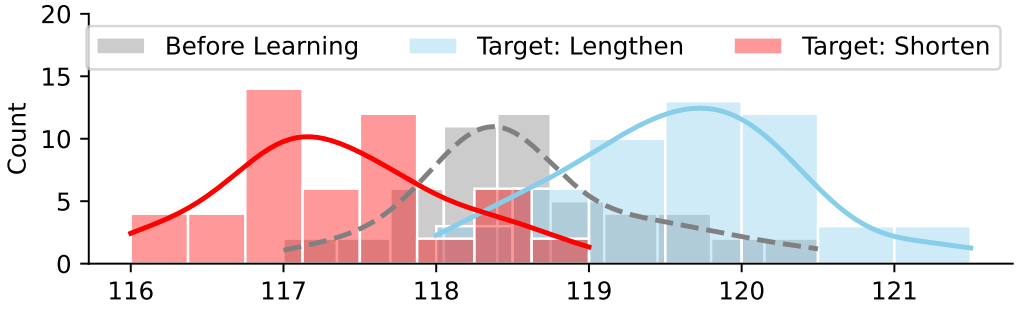


Figure 6.1: **Target syllable duration distribution before and after reinforcement learning (for shortening, lengthening) of the targeted syllable.** Effects of reinforcement learning in single run simulations of 1000 trials, aiming to change the duration of the syllable in the direction of lengthening (blue) or shortening (red). Baseline duration distribution in gray represent 50 trials prior to RL, and each of the two other distributions represents post-RL duration distributed (color coded).

Following the learning protocol for syllable shortening and lengthening, the respective mean syllable durations become 117.33 ms and 119.59 ms, with standard deviations of 0.7 and 0.75 ms. Hence, for a run consisting of 1000 learning trials, the difference between mean durations (pre and post learning) amounts to 1.27 ms in the case of syllable shortening and 0.98 ms for syllable lengthening. In other

words, the temporal shift is ∼1.3 and ∼1.7 standard deviations from the baseline distribution (Fig. 6.1).

To make the analysis more complete, we investigate the averaged results as well. We make the assumption that each run, executed with a distinct seed, simulates a different behavior and consider them as "simulated birds". Therefore, 10 runs are viewed as 10 simulated birds, and the mean results offer a measurement akin to those employed in behavioral studies. The mean value of the distribution from each run is taken for all the syllables, providing a 10 (runs) by 5 (syllables) matrix. Then, this distribution of means is compared with the average baseline distribution (10 x 5 matrix). We report that the single run results are similar to the averaged ones. In the latter, the averaged change for shortening and lengthening are respectively 0.98 and 1.17 ms. Upon increasing the number of trials from 1000 to 3000, we obtain a mean change of ∼1.6 ms for syllable lengthening and 1.79 ms for shortening. A slight difference between the magnitudes of learning in the two cases (shorten/lengthen) can be observed. It may indicate that syllable shortening is easier to perform as compared to lengthening.

To what extent is the modification specific to the target syllable?

To investigate the specificity of such a temporal shift, we analyzed the durations of the non-target syllables before and after learning, considering both individual runs and the averaged outcomes of the 10 runs. With the single runs, we conducted independent t-tests comparing syllable distributions before and after the learning and observed, in every seed, a temporal shift at the level of the target syllable ($p < 0.001$). In few seeds (10%), we observe an effect in the subsequent syllable, with a significance level of $p < 0.05$ and a smaller ($\sim \frac{1}{2}x$) t-value compared to one of the target syllable (x). In terms of change in mean duration, in this example case with interference, the target syllable changes by 0.94 ms and the following syllable by 0.4 ms. In the averaged results, no significant ($p < 0.05$) change in any other syllable, except for the targeted one is present. Taken together, these results show an effect at the target syllable level, but no discernible effect on the other syllables (Fig. 6.2).

Underlying Mechanisms of Behavioral Adaptation

The change in synaptic weights (ΔW) drives a change in population activity and bump speed, causing the temporal shift. Hence, the patterns of this change (ΔW) reveal important information about the underlying mechanisms of behavioral adaptation. Upon looking at the matrices of the ΔW at the end of learning, we observe the change is localized on a portion of the presynaptic neurons, but it is diffused across all postsynaptic neurons. However, even in the postsynaptic neurons, the main effect (in terms of magnitude) can be observed in the same localized population. These affected neurons are in fact the neurons correlated with the target syllable.

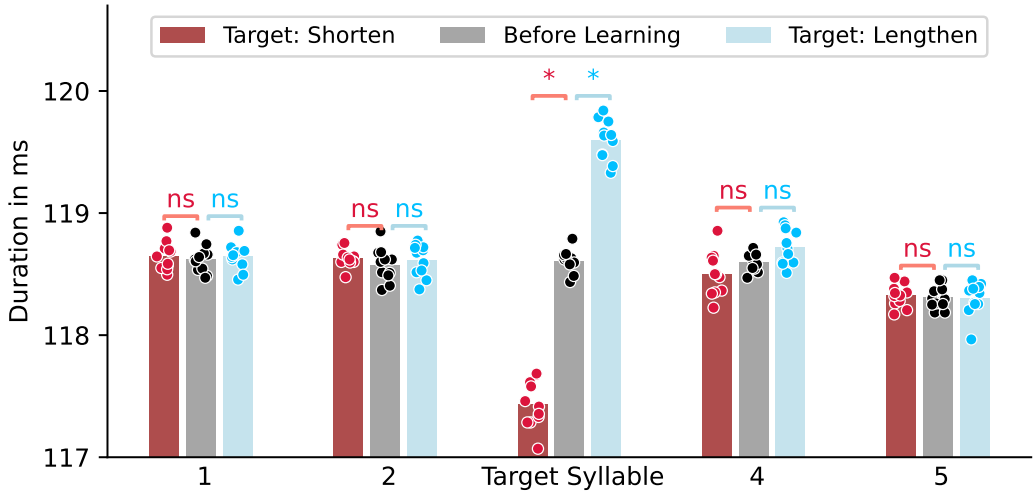


Figure 6.2: **Learning is specific to the targeted syllable.** 10 runs (of 1000 trials) of learning, aiming to achieve syllable duration shortening (red)/ lengthening (blue) are run. For each of the 10 runs in the three conditions (baseline/increase/decrease duration), a mean duration is computed from the respective duration distributions (10 mean values per condition) and these are shown as points in the bar plot. Baseline mean durations are displayed in gray. * stands for $p < 0.001$.

As it was described in Chapter 5 we have a network of 1000 neurons, equally split into 5 equal segments of 200 neurons, which each define the duration of one of the 5 syllable. The start and the end of the syllable is defined by the moment the center of mass of the bump crosses the 'start' neuron and the 'end' neuron. As the bump approaches the 'end' neuron, a bigger and bigger portion of it engages neurons associated with the following syllable. The amount of the neurons affected depends on the width of the bump. In rewarded trials, the connections emerging from these presynaptic neurons will be likewise strengthened. Based on these observations and for illustration (clarity) purposes, in Fig. 6.3(A,C), we show only these 250 neurons, where 0 and 200, are respectively, the neurons at the start and end of the target syllable.

In Fig. 6.3 (A,C), we show that, both when the syllable is targeted for shortening and lengthening, we observe an effect centered at the diagonal, with almost the same magnitude, but of different signs. For instance, in the single run, in the ΔW in Fig. 6.3 A, where the target is to shorten the syllable, we observe a decrease in the synaptic connections in the upper half of the diagonal and an increase in the lower half. It translates to a weakening of the connections from the presynaptic neuron 100 to the postsynaptic neurons prior (neuron 0 to 100) and a strengthening of the ones ahead (neuron 100 to 200). These changes modify the connectivity profile at the level of the involved presynaptic neurons (but not of the other presynaptic neurons), making the bump move faster in that area.

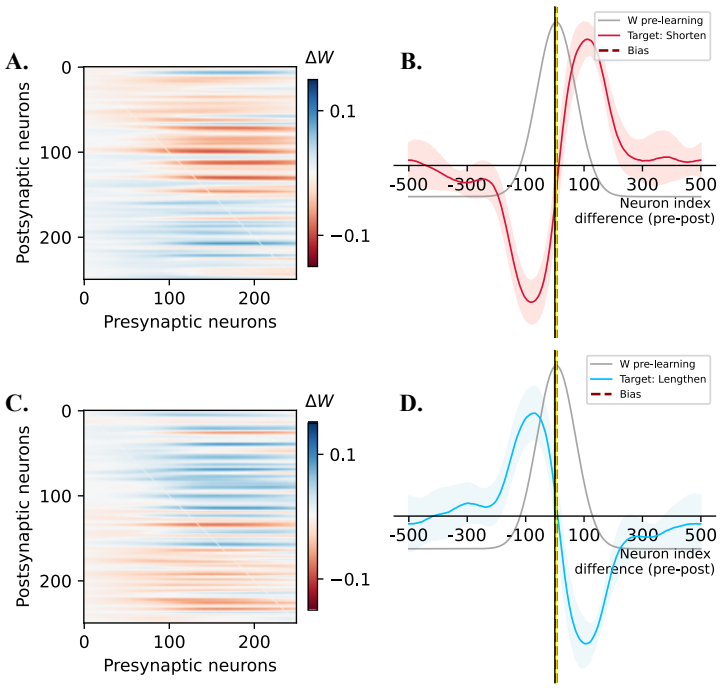


Figure 6.3: **Localized changes in synaptic weights, following learning.** A) ΔW of a single run (target: decrease syllable duration), zoomed in at the area of pre and post-synaptic neurons (n=200) encoding the target syllable and 50 neurons after. B) ΔW are rotated by 90 degrees, to see the synaptic weight change with respect to the diagonal. The mean of the ΔW across the presynaptic neurons encoding the target syllable, averaged across 10 runs. The thicker line represents the mean and the filled in area the standard deviations of each of the 10 runs' averaged ΔW(across the presynaptic neurons). C, D) Same as A and B for syllable duration lengthening. Note: The pre-learning W is scaled for illustration purposes.

Moreover, we investigate ΔW across different runs and present an average. First, since we saw in panel A and B, that the changes occurring are with respect to the diagonal, we rotate the ΔW matrix by 90 degrees to get a representation based on the diagonal. Then, we average across the columns (presynaptic neurons), to investigate the averaged effect of all involved presynaptic synaptic to the postsynaptic neurons. Lastly, we average the effects on the postsynaptic neurons across 10 runs. The results are illustrated in Fig. 6.3 (B,D). Consistent with what we reported at the single run level, in the case of syllable shortening, we can clearly see a synaptic weight decrease to the preceding postsynaptic neurons and an increase to the following ones. The flipped version with respect to the y axis, is seen in the other protocol (lengthening). Hence, the mechanisms involved are very similar.

The reported modification pattern is in fact a modification of the bias β, i.e. an increase. As described in the previous chapter, β grants feedforwardness to the

network, by defining a bump centered further ahead from the diagonal. It hence affects the speed of bump movement. We also showed that this relationship is quasilinear, i.e. an increase in bias, increases bump speed. The presented ΔW in Fig. 6.3 (B) shows an example of increasing the bias by two different mechanisms, during learning. If the synaptic weights to the neurons ahead of the ones defined by the bias are strengthened and the ones prior to it are weakened, an increase in bias occurs.

To better interpret and put into context the results of our learning rule in the synaptic weight change, we would like to summarize some important contributing factors, originating from the learning rule:

(i) The learning rule dictates an eligibility trace spanning from the start to the end of each syllable (Eq. (6.2)) and only the target syllable can be given a reward of 1.

(ii) Based on equation 6.2, the mean firing rate of only the presynaptic neurons active at the time of the target syllable, is taken to compute the eligibility trace. This corresponds to the positions, where the dynamic COM's discretized form (neuron) falls in the segment $[\alpha\frac{i}{s}n, \alpha\frac{i+1}{s}n]$ and a few following neurons, which are active due to the shape/width of the bump.

(iii) The reinforced synaptic weights can be with any of the postsynaptic neurons. However, even in this case, the effect is more pronounced in the regions corresponding to the target syllable encoding neurons. Other effects are averaged out.

6.3 Learning in the SNN

In this section, we present the results of the application of the same learning formalism (R. Williams, 1992) in the spiking neural network (6.1),(6.2). We will show our results with an exponential integrate and fire neuronal model, with an AdEx neuron model and in a two population (EI) network of exponential integrate and fire neurons. As already mentioned in Chapter 5, in Table 3.1, the size of the network used for the SNN is 3000 neurons, about 7 times less than the song-related HVC_{RA} population (20,000). With the current model, we get at most 10 neurons active at the same time (timestep), as compared to 100-200 reported in HVC.

Only three syllables are investigated, the one prior to the target, the target and the following syllable. This choice was guided by the results from our work in the previous section with the rate based model, which showed that when there is an effect in the other syllables, it is always in the following one. Everything in the learning rule is identical with the one presented in the rate model, except the mean firing rate, the learning rate γ, which equals 0.04 and the number of trials (200). Since no mean firing rate is present, it will either have to be computed based on the generated spikes or replaced by another parameter. We choose to do the former and explain below how it is computed.

Syllable Duration and Learning Signal

We compute the 'mean firing rate' by first generating a binary matrix from the spikes, and then convolving this matrix in time and space[2], with a Gaussian kernel of standard deviation 5 ms and 50 neurons respectively. This provides a narrow bump of activity in both dimensions, the center of mass of which indicates the activity peak. Just like in the rate model, the timepoints at which the center of mass of the bump crosses the boundary neurons, define the syllable duration. The same bump-like signal is also fed into the learning algorithm.

This choice was made considering this signal as similar to the neuronal activity in the presence of a larger network (20,000 neurons). As briefly mentioned above, with our current network size and propagation speed, about 5- 10 neurons are active at the same time-step (Δt), compared to $\sim$ 100- 200 active in HVC. Hence, the activity is less continuous than in HVC and reflects the stochasticity of the noise (with a spatial convolution, $\sigma = 2$ neurons), which was chosen to reproduce the duration variability reported from experiments. To compute syllable duration, a center of mass of activity is necessary. Assigning one of the 5- 10 neurons active as the most active or as the center of mass would not be precise enough in capturing the whole dynamics. In this context, a small- windowed convolution in space is necessary. Moreover, as the neurons have bursts of spikes, a temporal correlation with a 5 ms standard deviation (σ) in a 10 ms burst is useful. In the context of learning, due to stochasticity, the mean over only 10 active synapses, reflecting our

[2] In this chapter, the space we refer to is feature space, unless otherwise stated.

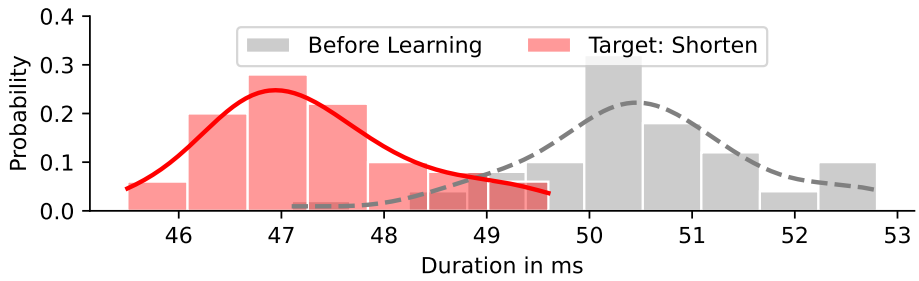
10 active neurons at one Δt, is less reliable than the total ΔW computed from 100 -200 synapses underging synaptic plasticity in HVC. One way to account for this is by smoothing the activity as described above.

6.3.1 Results

Shortening: Exponential Integrate-and-Fire Neuron

As explained above, we first run 50 trials prior to and following learning. We acquire a mean baseline duration of 50.5 ms, with a standard deviation of 1.1 ms. Following 200 trials of learning, targeting syllable shortening the mean syllable duration becomes 47.2 ms, with standard deviation of 0.95 ms (Fig. 6.4a). The difference is 3.29 ms, corresponding to a temporal shift of ∼3 standard deviations from the baseline distribution. Moreover, the mean change across 8 runs amounts to 3.14 ms, with the lowest run difference at 2.01 ms and the highest at 3.6 ms.

(a) Effect of reinforcement learning on duration distribution shift on a single run.



(b) Averaged effect of RL across runs.

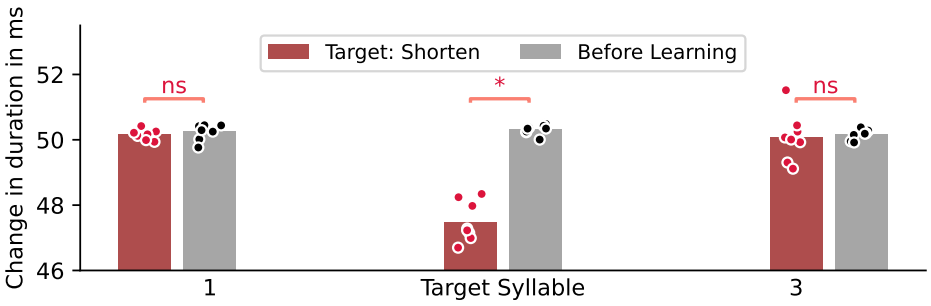


Figure 6.4: **Representation of the effect of learning in a behavioral adaptation paradigm as modeled by an Exponential Integrate and Fire Neuron Model** (a) Target syllable duration distribution before and after reinforcement learning. Duration distributions are computed as explained in section 6.1. (b) Adaptation is specific to the target syllable and no effects are observed in the adjacent and non-adjacent syllables on the averaged results of 10 runs. * stands for $p < 0.001$.

We follow the same protocol explained in section 6.2.1 to investigate the presence of interference. The results of 8 averaged runs show no interference in the adjacent syllables. However, at the level of individual runs, 3 out of 8 (37.5 %) have a significant interference ($p < 0.01$). In these three cases, the target syllable mean changes respectively by 3, 2 and 2 ms, whereas the following syllable changes by 0.97, -1.2 and 1.2 ms. The ratio of change is not constant in the three cases and it is not always in the same direction, i.e. the following syllable is increased or decreased. Such bidirectional change of the following syllable is likely the cause of averaging out, i.e. of no interference in the averaged results.

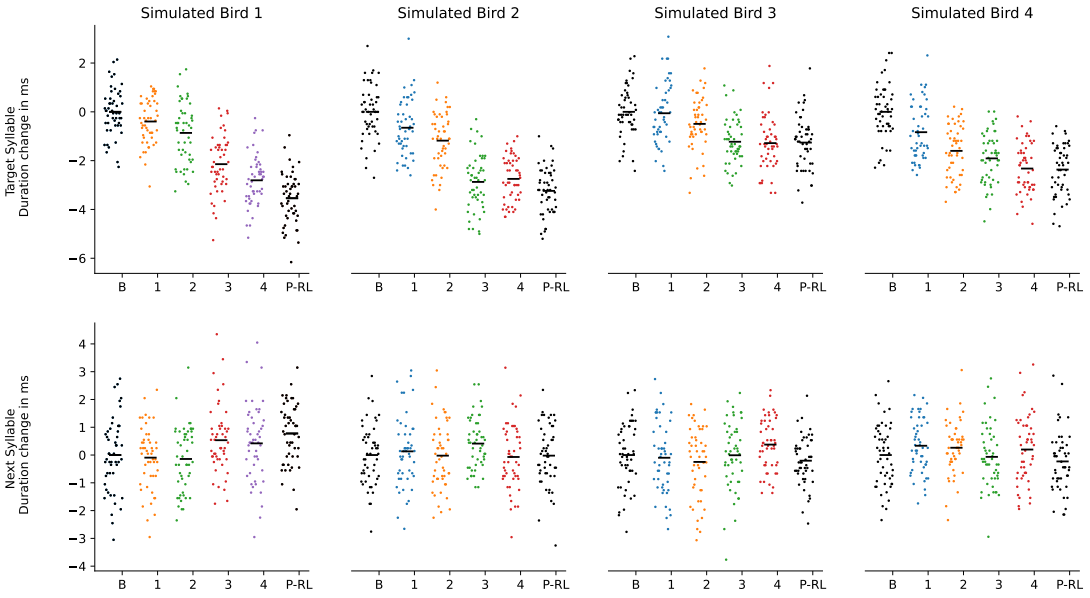


Figure 6.5: **Four simulated birds across learning (Exponential IF):** Change in duration across 200 trials of learning in the Exponential Integrate-and-Fire neuron model. The results from 4 seeds are shown, where the first and last bars respectively represent, the 50 baseline (B) trials and the 50 post RL (P-RL) trials. The 200 trials in the middle show the progression of learning in time. They are binned in 4, each containing 50 trials. In the first row the change in duration of the target syllable is shown, and in the second that of the following syllable. That is because, whenever we see an effect in the other syllables it is always in the next syllable. In black, baseline and post-RL distributions.

Upon looking at the ΔW, we observe a weight modification, quite different from the one witnessed in the rate-based model. In the single run ΔW, we do not see a clear effect centered on the diagonal and it becomes harder to interpret the effect by looking solely at the connectivity matrix in Fig.6.6. An observation of the averaged results provides more insight. The postsynaptic neurons affected

are rather distributed, with the largest effect observed at the level of the (i) self-connections of the bump and (ii) following it (up to 700 neurons after) (Fig.6.6, right). (i) The first observation is that of an increase in synaptic weights at the level of the diagonal, i.e. all connections to neighbouring neurons are strengthened. Neighbouring neurons are defined by the width of the activity bump in space. The increase in synaptic weights at this level likely reflects a Hebbian-like learning mechanism, and serves as a mechanism to conserve the bump's stationary state. Hence, it could slow down the bump slightly. It is a particularity of the effect of the learning rule in the spiking neural network and it is present in the AdEx model as well. It was not present in the rate based neuron model.

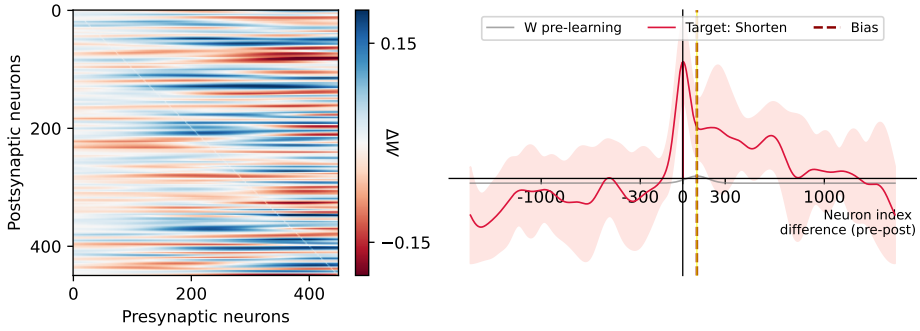


Figure 6.6: **How do the weights change?** Change of ΔW weights in a single run. Only the neurons affiliated with the target syllable are plotted (n = 450). (Right) The average mean across the 10 runs. The target is to decrease the duration of the syllable. In red the averaged change in synaptic weights, based on the distance from the presynaptic neurons, affiliated with the target syllable. In gray, the Gaussian connectivity profile of the presynaptic neurons to the postsynaptics.

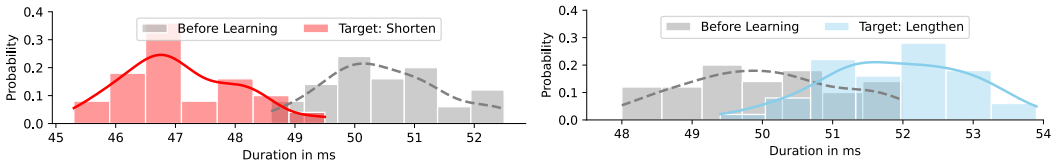
(ii) On the other hand, the increase in the weights after the bias, provides an addition to the bias, and thus, increases it. Such an increase in the bias consequently increases the speed of the bump/activity in the location defined by the presynaptic neurons involved. Just like we described in the rate, only the presynaptic neurons correlated with the target syllable are affected. However, there is quite a high variability in the weight changes, as compared to our previous results. This observation, taken together with the behavioral outcome revealing bidirectional interference in the single runs, drives us to speculate that the simulated 'birds' incorporate different mechanisms in adapting syllable duration to an external reward.

AdEx

In chapter 5, we state our reasons for choosing an AdEx neuron model. These reasons are mostly to reproduce biologically realistic dynamics and more specifically the burst like activity in HVC, driven by the calcium channel dynamics. In this chapter, we started with a simple exponential IF model and showed the effects of learning. Here, we explore whether adding an adaptation current to the neuron makes a difference in the learning rate or in the ΔW. A recent study (Kubo et al., 2023) has revealed additional implications of neural adaptation to learning in an artificial neural network, i.e. a facilitation of learning. This may or may not be the case in the ring attractor model. An important contributing factor is the role of adaptation in bump propagation, and furthermore its speed. There is an interplay between the speed caused by adaptation and the one by the bias.

Here, we investigate the experiment both in the direction of shortening and lengthening. In the case of the latter, the learning rate, has to be adapted and lowered in order to avoid interference. The underlying reasons of this difference between the two paradigms remain to be further investigated.

(a) Effect of reinforcement learning on duration distribution shift at a single run level.



(b) Averaged effect of reinforcement learning in 10 runs.

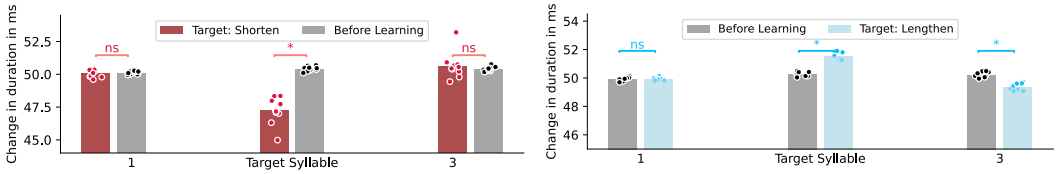


Figure 6.7: **Representation of the effect of learning in a behavioral adaptation in paradigm as modeled by an Adaptive Exponential Integrate and Fire Neuronal Model** (a) Target syllable duration before and after reinforcement learning. Duration distributions are computed as explained above in subsubsection 6.1. (b) In the case of syllable shortening, adaptation is specific to the target syllable and no effects are observed in the adjacent and non-adjacent syllables on the averaged results of 10 runs. In the case of lengthening, there is a significant effect also at the level of the following syllable. In the individual runs, this effect is observed in 50- 60% of the cases.

For the case presented in Fig. 6.7a (right), we acquire a mean baseline duration of 50.48 ms, with a standard deviation of 0.96 ms. Following the learning protocol with 200 trials for syllable shortening the mean syllable duration becomes 47.07 ms, with standard deviation of 0.96 ms. The change is 3.4 ms (Fig. 6.7),

corresponding to a temporal shift of ∼3 standard deviations from the baseline distribution. Moreover, the mean change across 10 runs amounts to 3.1 ms, with the lowest run difference at 1.8 ms and the highest at 5.6 ms.

No interference is observed in the averaged results, in case of syllable shortening. However, two of the 10 runs present with interference in the next syllable. The significance is $p < 0.01$, whereas the magnitude of change is in relation to the one of the target syllable is as follows. The target syllable mean change, in the two cases, is 3.29 ms and 5.6 ms and on the following syllable 0.8 ms and -2.9ms, respectively. A bidirectional effect on the next syllable's shift is also observed here, i.e. both lengthening and shortening alongside the shortening in the target syllable. We show 4 different runs across the learning trials in Fig. 6.8, to illustrate the effects at the level of the target and following syllable. In contrast, in the case of lengthening, interference is present at the level of the following syllable, in the opposite direction.

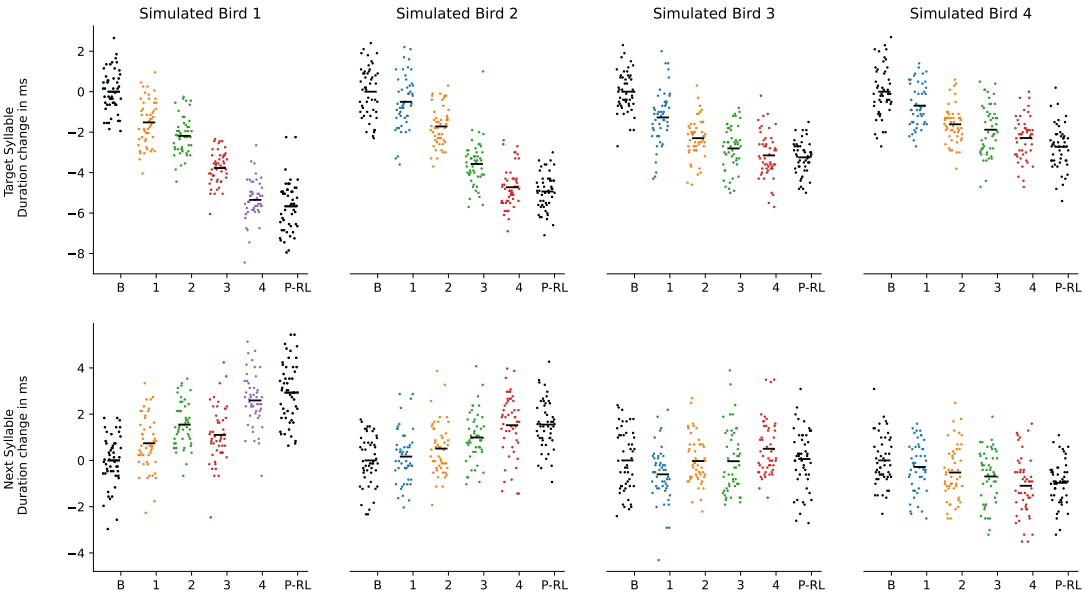


Figure 6.8: **Four simulated birds across learning (AdEx) to shorten the duration of a target syllable.** Change in duration across 200 trials of learning. Just like in Fig. 6.5 the results from 4 seeds. In the first row the change in duration of the target syllable is shown, and in the second that of the following syllable.

Moreover, an illustration of the final ΔW for a single run and the average across 10 runs are shown in Fig. 6.9. Notably, it is dissimilar from the rate ΔW and quite similar the Exponential IF. The postsynaptic neurons affected are relatively more localized, although they are still not at the level of the neurons encoding the target

syllable ($[-200 : 250]$). The same 'conservation' reinforcement at the level of the self-connections (bump), as in the eIF, is present. As in eIF, the mechanism driving the bump faster, is the increase in the area following the bias.

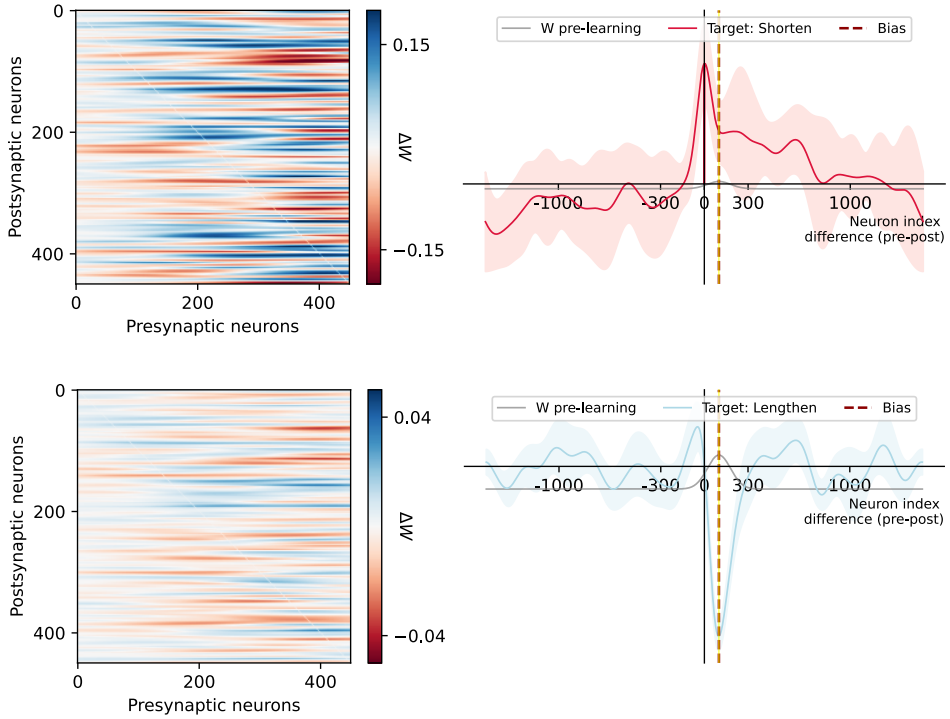


Figure 6.9: **How do the weights change in the AdEx neuronal model?** Change of ΔW weights in a single case and the average mean across the 10 runs. In the upper panel, the target is to decrease the duration of the syllable and in the lower panel, to increase it.

6.4 Learning in the EI Network Model

Furthermore, we investigate learning in the EI Network with eIF neurons. In contrast to the model presented in Chapter 5, here the connection probabilities between the two populations are set to 1. We use the same learning rule in the EI network first presented in the beginning of this chapter. The plasticity rule only affects the E-E synaptic weights. However, the changes in weights, do not only affect the excitatory population, but as they are in a loop, the inhibitory population as well.

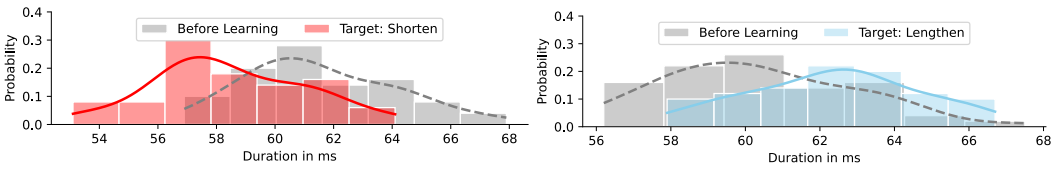
For the case presented in Fig. 6.10a (right), we acquire a mean baseline duration of 61.5 ms, with a standard deviation of 2.5 ms. Following the learning protocol of 300 trials for syllable shortening the mean syllable duration becomes 58.4 ms, with standard deviation of 2.5 ms. The change is 3.1 ms, corresponding to a temporal

shift of $\sim$ 1 standard deviations from the baseline distribution. Using the same protocol for syllable lengthening, the results are similar, as shown in Fig.6.10a (left).

Moreover, in the shortening experiment. the mean change across 12 runs amounts to 2.1 ms, with the lowest run difference at 1.1 ms and the highest at 3.1 ms. When the number of trials per run is increased to 500, we see a mean change of 3.4 ms, with a lowest run value at 2.76 and the highest at 4.3 ms. No interference is observed in either cases in the averaged results (Fig. 6.10b). However, in the 500 trials runs, one of the 5 runs presents with interference in the next syllable. The significance is $p < 0.01$, whereas the magnitude of change is half of the one observed in the target syllable. The target syllable change in mean is by 2.81 ms and the one on the following syllable at 1.4 ms.

On the left, the results of the lengthening experiment are shown. In terms of effect at the level of the target syllable, they are very similar. However, there is visible interference in the opposite direction of the target syllable change in the following syllable, i.e. the following syllable's duration decreases, as the target syllable's increases. The reason for such a difference between the two experiments remains unclear and is not further investigated in this manuscript.

(a) Effect of reinforcement learning on duration distribution shift at a single run level.



(b) Total effect of reinforcement learning across 10 runs.

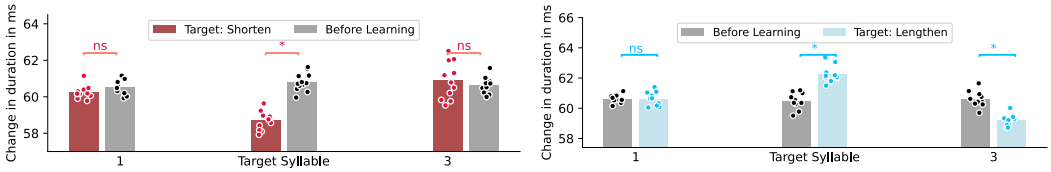


Figure 6.10: **Representation of the effect of learning in a behavioral adaptation in paradigm as modeled by an EI network of Exponential Integrate and Fire Neuronal Model** (a) Target syllable duration before and after reinforcement learning. Duration distributions are computed as explained above in subsubsection 6.1. (b) Adaptation is specific to the target syllable and no effects are observed in the adjacent and non-adjacent syllables on the averaged results of 10 runs.

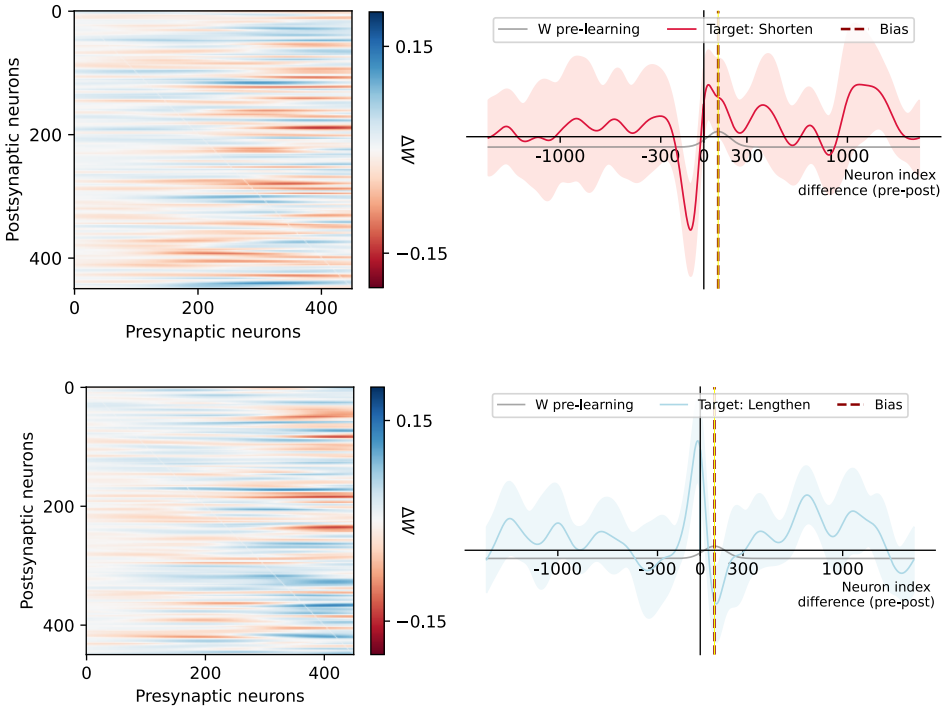


Figure 6.11: **How do the weights change in the EI network model?** Change of ΔW weights in a single case and the averaged mean across the 10 runs. In the upper panel, the target is to decrease the duration of the syllable and in the lower panel, to increase it..

The synaptic weights are altered with a different pattern (Fig. 6.11, upper panels). When compared to the one population eIF model, which is the same neuronal model, but does not include inhibition, the difference is clear. While the effect on the further away neurons is very noisy, the main effect is at the postsynaptic neurons preceding the diagonal. There is a clear minimum right before the diagonal/self-connections and a slight increase in weights following the it, with a peak (similar to the two previous ΔW) between the diagonal and the bias. The dip in activity, which can also be seen from the single run weights, slightly changes the connectivity shape. It is an alternative way of a relative bias increase (similar to one of the mechanisms in the rate model). Hence, it increases the speed of the bump at the location determined by the presynaptic neurons involved.

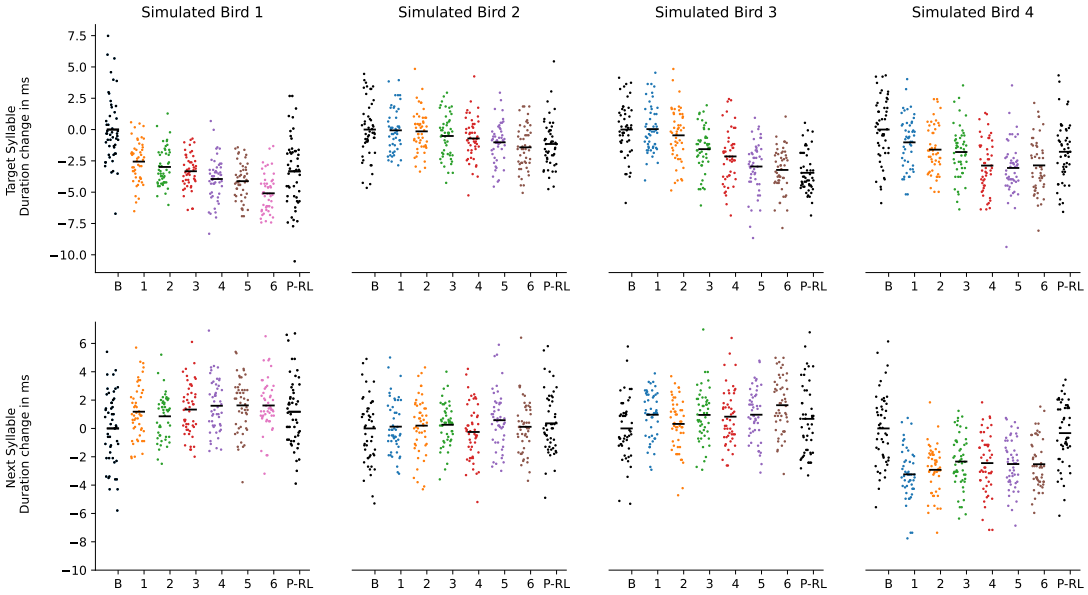


Figure 6.12: **Four simulated birds across learning to decrease syllable duration in an EI network (Exponential IF):** Change in duration across 300 trials of learning in the Exponential Integrate-and-Fire neuron model. The results from 4 seeds are shown. They are binned in 4, each containing 50 trials. In the first row the change in duration of the target syllable is shown, and in the second that of the following syllable.

6.5 Comparing Model Results to Experimental Data

There is significant evidence from the literature, that a Conditional Auditory Paradigm makes the bird shift its duration in both directions, left and right, i.e. to decrease or increase mean syllable duration (Ali et al., 2013; Pehlevan et al., 2018). Moreover, the effect of the CAF on the other syllables is investigated and an absence of interference in any of the syllables is reported. Our averaged results in all the models support this finding, but we do notice interference in a small percentage of the runs (10% for rate, 37.5 % for eIf and 20% AdEx and EI). They are always in the syllable following the target and the magnitude of change is smaller. As this effect is actually averaged out when we look at 10 different seeds, we wonder if that was also the case with the experimental data and what the magnitude of change in the next syllable for each bird across days is. Moreover, interference is present even in the averaged results when the syllable is targeted for lengthening. To address these questions, we use data previously collected in the lab in a CAF paradigm. The methodology and most of the results are reported in the the

thesis of Ursu (2023). However, the findings on the effect on the next syllable are a preliminary finding from the data and have previously not been reported.

6.5.1 Methods

Behavioral Paradigm

Adult zebra finches (aged >90) were used in the following experiments. They were housed in a 12h light/ dark cycle and provided food and water ad libidum. For the first three days the bird was singing without any reward. Based on the first song spectrograms, a syllable in each bird was chosen for modification (lengthening). In the fourth day, the conditional auditory feedback protocol was begun. A microphone captures the sound of the bird's song, which is then transferred to the amplifier and then to the software, through which the duration is computed based on the envelope of the song (Skocik and A. Kozhevnikov, 2013).

Whenever the duration of the chosen syllable does not cross a set threshold (set at the mean of the baseline distribution, which was computed from the first 3 baseline days), real-time audio feedback, in the form of a white noise is given with a latency of $\sim$5 ms for 100 ms. Moreover, the threshold is updated dynamically, based on the bird's performance in the last 200 renditions. If the fraction of durations that have not received white noise is $\geq$ 70%, then the threshold is updated (Ali et al., 2013; Ursu, 2023). This update is only made in the direction of learning. For instance, if the syllable is targeted for lengthening and the threshold is at a value x, it will be increased to $x + \Delta x$ if $\geq$ 70% of the durations where higher than x. However, it will never decrease, as such an update would not permit lengthening.

The CAF protocol was continued for a variable number of days (4-10), depending on how quickly/slowly the bird shifted the target duration distribution. A minimum of 4-5 days was set for each bird, but if the shift was less than 2 standard deviations at the end of 5 days, the CAF was continued until it reached that value. Moreover, we would like to specify that although in the rest of the manuscript we have referred to the duration as a syllable, in the experimental protocol the targeted element in the song is actually an interval encompassing both the syllable and the following gap.

Data analysis

Among the different birds undergoing the CAF protocol, we selected only the ones that had a interval following the target interval. In the cases, where the interval was at the end of the motif, the bird was excluded from the analysis. This is because we would like to see the adaptation both at the level of the target interval, and the following one. Moreover, as it was previously reported (Ursu, 2023) some birds did not achieve the temporal shift and some birds presented with an amplitude increase (N = 7) in the envelope, which was interpreted as a duration

change. However, as this is no pure temporal shift, these birds were also excluded from the analysis. Hence in the analysis presented below, a total of 8 birds (N = 8) were evaluated.

The extracted audio files were band-pass filtered at cutoff frequencies of 500 and 8000 Hz, then squared and low-pass filtered with a box car window of 10 ms (Ursu, 2023), providing the envelope. For each bird, we manually select a threshold, based on which the interval is segmented. Then, we manually label 100 - 200 motifs per bird, and use them as training data for the HVC (Nicholson, 2018) software, a supervised machine learning software used to perform automatic interval labelling. It then predicts the annotations for the rest of the motifs. After the prediction is complete, a text file for each song is exported, with the onset and offset of each interval and the respective label.[3]

6.5.2 Results

Behavioral data reveal interesting findings. The duration of the target interval and the following interval of the 8 birds presented was variable. We use the durations in the third day of baseline (Fig. 6.13) and compute the mean baseline duration for each bird and then average them across birds. The mean across birds is 123 ms for the target interval, with a mean standard deviation of 3.4 ms and 169 ms with a standard deviation of 4.4 ms for the following interval.[4] We present the durations with respect to the difference from the baseline duration's mean. The duration variability (1.5-4 % of mean durations) discussed above is also visible from these plots (Fig. 6.13).

In the last day of CAF the overall mean target interval duration is 145 ms (standard deviation of 5.7 ms) and that of the following interval 163 ms (standard deviation 3.8 ms). The difference ($\delta_{duration}$) is 22 ms (18%, 4.4σ) at the level of the target interval and 6 ms (3.5%, $< 2\sigma$) on the next interval. Upon looking at the individual effects from baseline to the last day of CAF, as a mean duration, we observe changes in interval lengthening, from 7 to 30% the interval duration and an effect in the following intervals from 1 to 8%.[5]

The mean change per day across birds for the target syllable is 3.46 ms and for the next syllable 0.84 ms. Using a one-sample t-test to compare the difference per

[3] A detailed description of the analysis steps and pipeline used in the lab, compiled by Remya Sankar: https://github.com/rsankar9/HVCHackathon/tree/main/HVC.

[4] Following the order in Fig.6.13, the birds had mean target intervals of 137, 163, 97, 142, 79, 165, 103 and 125 ms. Respective standard deviations are 3.3, 2.6, 4.1, 4.4 , 5.4, 2.5, 1.9, 3.2 ms.
The following intervals' mean durations were 194, 178, 136, 206, 145, 228, 187 and 75 ms. Respective standard deviations are 2.9, 3.5, 2.9, 4, 9.7, 5, 4 and 2.7 ms.

[5] Following the order in Fig.6.13, after learning, the birds had mean target intervals of 146.34, 180.55, 120.69, 162.8 , 103.55, 168.78, 118.95 and 161.78 ms. Respective standard deviations are 2.8, 4.9, 5.1, 7.2, 10.9, 4.2, 3.7 and 7.2 ms.
The following intervals' mean durations were 190.19, 176.1 , 125.15, 202.08, 140.1 , 215.14, 184.47, 72.16 ms. Respective standard deviations are 3.6, 3.6, 2.9, 4.8, 5.6, 3.5, 3.4 and 2.7 ms.

day with 0, we show a significant change for the target syllable, with a p value of $p = 0.009$ and for the next syllable with $p = 0.018$.

Moreover, in some of the birds, such as bird 2, 3, 6 the change in the following interval is very clear and in the opposite direction, whereas other birds such as 4 and 5 show more diverse effects (Fig. 6.13) across days. For instance, bird 5 shows an increase followed by a decrease during the CAF paradigm. The changes are considerably smaller in magnitude as compared to the ones in the target interval ($\sim$5 times less), but they are present, they are not always in the same direction across days and need to be explored more.

6.6 Discussion

In this chapter, we show the results after we have implemented a reinforcement learning rule into our models with the goal of simulating the CAF behavioral paradigm. Although the same parameters are used in the three spiking models, there are differences in mean and standard deviation. The mean is likely affected by the addition of the adaptation current (increases bump speed and decreases duration) and the addition of an inhibitory population, which appears to slow down the bump and increase duration. On the other hand, the variability (σ) is similar in eIF and AdEx ($\sim$1ms), but increases significantly in the EI network, amounting to $\sim$2.5ms. These variabilities correspond to a coefficient of variation of $\sim$2% (eIf, AdEx) and $\sim$4% (EI), similar to $\sim$1-10% CV range reported in young adult zebra finches (Glaze and Troyer, 2013; Glaze and Troyer, 2012). The increase of variability in the EI model could be related both to an effect of the I population, but also due to the increase in duration (61.5ms as compared to 50.5ms in the other two models). The latter could represent with the scalar property of timing, which states that a longer duration has a higher variability. However, no change in standard deviation was seen when the duration changed to 58 ms. It is probable that a bigger difference in mean (within model) is necessary to acquire a change in variability. This remains to be tested.

With respect to learning, in the 4 used models, we acquire a temporal shift of the distribution of the target interval, with an average of 1.8 ms (rate), 3.1 ms (eIF), 3.1 ms (AdEx) and 3.4 ms (EI) respectively. In the averaged results, no effect on the adjacent interval is present. However, at the single run level, we report interference in some of the runs (10- 38%, depending on the model used), with a temporal shift lesser than in the target interval.

In the rate based model, we implement learning in both directions, to shorten and lengthen the interval duration. We observe that the magnitude of change in interval duration, is higher when it is targeted for shortening as compared to lengthening. This could suggest that in such a protocol, it is easier to decrease interval duration. It appears to be the case in the behavioral paradigm as well (Ali et al., 2013; Pehlevan et al., 2018). In the three spiking models, we focus on explaining the mechanisms of interval shortening. However, for the AdEx and

EI network model, syllable lengthening is presented in the figures as well. Using the same parameters for syllable lengthening, we observed that in order to not get a very high interference at the next interval (in the lengthening experiment), the learning rate had to be decreased to $\gamma = 0.004$. Following this modification, less interference was observed, but it was still present, always at the level of the next syllable and in the opposite direction. These reasons drove us to look at the experimental data and investigate whether there is an observable change at the level of the syllable following the target one.

Hence, we confront the present results with experimental data, and observe slight interference present in most birds. At the end of learning, all the birds had a significant temporal shift of more than 2.8 standard deviations in the target interval. In the following interval, only two birds had a shift of more than 2 standard deviations. Based on the criteria put by the experimenter (Ursu, 2023), we could say 25% of the birds presented with interference in the next interval. As in previous reports (Pehlevan et al., 2018) and as in our simulations, the averaged effects on the following interval are not significant, accounting to about 1.4σ on average (affected by a 4σ change in Bird 3), whereas the effect in the other interval is on average 6.8σ.

At the end of the CAF, the effect on the next interval is at the opposite direction with that of the target, i.e. if the target interval is increasing, the following interval is decreasing. However, on a day by day basis, we observe changes in the same direction as well. These observations might hint at the existence of diverse learning mechanisms across birds and the influence of specific training days on interference. While the simulations seem to echo these findings, the limited sample of just eight birds restricts the strength of our analysis. As such, a more extensive sample would be beneficial.

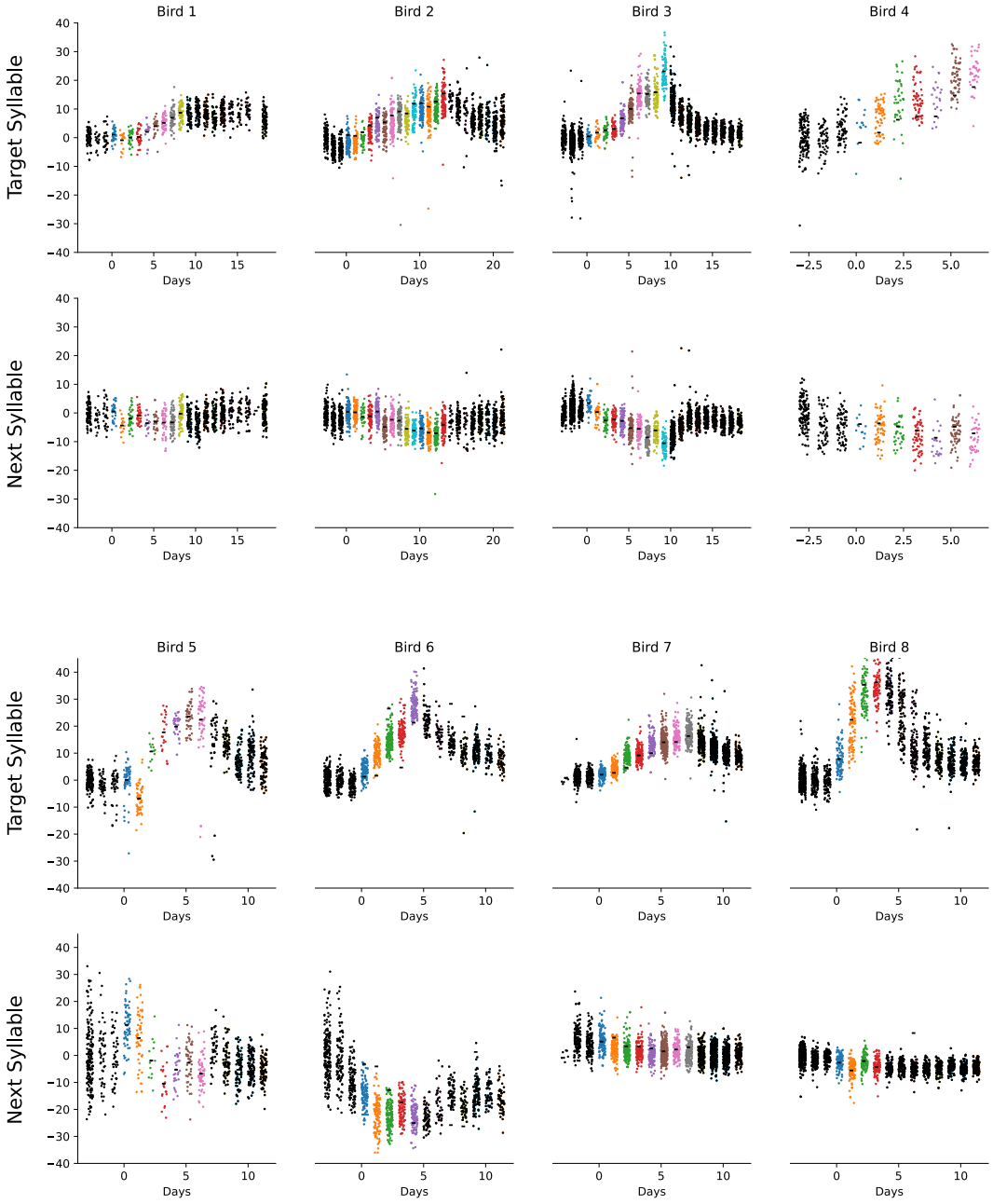


Figure 6.13: **Interval duration change across days of 8 birds undergoing CAF.** Each bin represents a day, each point the change in duration of one interval from the mean interval duration during baseline. In black, the baseline and return durations are represented. The durations of the other days are in colors, and the mean duration is shown by a black line. The x axes are aligned at the first day when CAF was started (day 0).

7 Discussion and Conclusions

In this chapter, we discuss the implications of our previously presented results and suggest their position in the literature of timing and songbirds. Although each individual results chapter is followed by its own discussion section, here, a more general insight on the implication of our proposal as a whole is compiled. Moreover, we rediscuss limitations and some methodological choices, the support in the literature that drove them, and the possible effect they may have had in our results. Several impromtu findings were acquired during the work, which could drive further experimental and numerical work. We try to formulate possible proposals in both directions. Lastly, we include a section on future perspectives.

7.1 A Robust and Flexible Model of Motor Timing using Attractors

In Chapter 5, we have presented a robust model of timing. We claim the robustness of our model based on the extensive literature on ring attractors, numerical analysis to find the parameter space allowing bump formation and perturbation protocols. We show that a ring architecture, with a gaussian connectivity profile, can give rise to neuronal dynamics consistent with those observed from electrophysiological recordings. Each neuron fires a burst of 3-6 spikes, with an approx-

imate duration of 10ms and activity propagation is guaranteed by a feedforward term, the bias. The speed of propagation can be adjusted by changing the bias term, demostrating temporal scaling.

Moreover, in Chapter 6, we focused on the flexibility of our model. We numerically implemented a well known behavioral paradigm, CAF, and showed that our model is flexible and can change the duration of a target syllable in response to a reward profile, implemented with reinforcement learning. Previous work has suggested that flexible time keeping requires a one to one mapping between synapses and the interval duration (Pehlevan et al., 2018). Our weight profiles present a widespread effect on multiple synapses, but by default due to the learning rule, it is concentrated on the synapses projecting from the presynaptic neurons affiliated to the target syllable. Moreover, we see that although a minimal effect is observed in the extremities, the synapses most reinforced are the ones with the postsynaptic neurons, also affiliated with the target syllable. This is mostly consistent with the suggestion of a one to one mapping.

It is challenging to construct a model that is both robust and flexible, as these two properties tend to be in opposite ends of the spectrum (Khona and I. Fiete, 2022). However, not only is this necessary in a general computational model, but it is even more in our specific model which aims to simulate a very particular area, with very clear dynamics, exhibiting both robustness and flexibility.

7.1.1 Rate and SNN Model

We started our numerical simulations with a coarse grained approach, using a rate based model. The first results on robustness, duration distribution and reinforcement learning are presented in this model. This helped us acquire an understanding of whether such an architecture would be able to able to simulate HVC dynamics and furthermore behavioral adaptation. However, although we used a very narrow Gaussian for the connectivity profile, the time constant of our rate model constrained the emergent bump width (narrowness), not allowing us to reach the observed 10 ms burst activity in HVC_{RA} neurons.

Alongside the motivation to investigate the effect of single neuron dynamics inside this architecture, this drove us to expand our simulations to a spiking neural network. Moreover, we added adaptation, to ensure the bursts observed in HVC. This provides a more biologically plausible model, constrained by experimental observations. Our results with the SNN are equivalent to the readout from the rate based model, both for robustness and flexibility. Moreover in the SNN, at the single run level, more nuanced results are observed, with slight interference at the following syllable after learning. This is consistent with previous experimental results (Ursu, 2023). However, the underlying learning mechanisms in the SNN are much harder to interpret and possibly require more investigations.

7.1.2 Relating the EI model with the Literature

Additionally, a model with two populations of excitatory and inhibitory neurons was investigated. At the microcircuitry level, there is a loop between HVC_{RA} and interneurons (Mooney and J. F. Prather, 2005), and moreover, a number of recent studies have highlighted the importance of inhibition in HVC sequence generation (Kosche et al., 2015). We used unstructured, non recurrently connected inhibition and studied at a very basic level the effects of such a population to our results. We tested the model with different parameters for excitation and inhibition, and fit those to experimental findings (Kosche et al., 2015). Our results are in line with the suggestion that HVC_{RA} activity is sufficient to drive structured activity in the interneuron network (30 Hz, as shown by Markowitz et al. (2015)). However, we did not investigate recurrent connections within the inhibitory neurons themselves and the differences appearing in the model in the case of a previously set structured inhibition. This could provide an interesting perspective for a more detailed computational model.

7.2 Biological Relevance and Interpretation

7.2.1 Motor Timing

A brief overview of the timing literature and the main proposed neural mechanisms and models were presented in chapter 1 and 3. Here, we would like to restate the role of songbirds in the motor timing literature, in the order of tens to hundreds of milliseconds and discuss the role and contribution of our study. Birdsong is characterized by a temporal structure, similar to human speech (A. J. Doupe and Kuhl, 1999), with both a syllable and sequence/song level temporal structure. In the taxonomy of timing (section 1.1), it would be referred to as belonging to subsecond, explicit, prospective, pattern motor timing. This temporal structure has been associated with premotor nucleus HVC through different experiments, including stimulation studies (E. Vu et al., 1994), extracellular recordings (R. Hahnloser et al., 2002) and cooling experiments in HVC (Long and M. S. Fee, 2008). HVC RA projecting neurons fire $\sim$ 10 ms bursts of 3-6 spikes, time-locked to the song that span the whole duration of the song. HVC is considered either as the main timing structure or as a crucial element of a distributed timing system.

Dedicated or Intrinsic Model of Motor Timing

In the light of the two broad classes of conceptual models of timing, the question arises whether HVC is a dedicated or an intrinsic model of timing. Numerous studies have shown temporal selectivity in nucleus HVC, either to specific intervals (Margoliash, 1983) or to more complex temporal features of song (like order and song, bird's own song) (Gentner and Margoliash, 2003; Lewicki and Konishi, 1995; Margoliash and Fortune, 1992; Mooney, 2000). With respect to motor timing, R. Hahnloser

et al. (2002) first showed the chain-like sequence in HVC during song production and cooling experiments in HVC revealed a causal link between sequence speed and song (Long and M. S. Fee, 2008). Taken together, these two aspects of HVC in sensory and motor timing, could show a more general role of HVC in temporal processing. This could drive us to consider it as a dedicated model of timing, however HVC is also crucial in other aspects song learning and song production (syntax (Y. S. Zhang et al., 2017)), not only with respect timing. Moreover, in terms of timing, we clearly see a role of HVC in song timing but no such role has been reported in other sequential motor behavior.

Hence, attributing HVC to intrinsic models of timing would be more intuitive. In this theoretical framework, time is encoded in the population activity, and is affected by intrinsic properties of the neurons and network such as time constants, ionic channels, neuronal adaptation, synapses etc. Population clocks are a subclass of intrinsic timing models, where time is encoded in dynamic changes of neural populations (D. V. Buonomano and M. D. Mauk, 1994; Medina and M. D. Mauk, 2000). The experimental studies in songbirds (R. Hahnloser et al., 2002; Long, D. Z. Jin, et al., 2010) support the position of HVC as a population clock, as HVC_{RA} neurons manifest with time-varying activity time-locked to the song. The model used in this manuscript, the ring model takes part in the population clock subclass of intrinsic models of motor timing. In the ring model, the dynamic changes in the population and more specifically in the bump encode timing.

Contribution of the Attractor Model

Our ring attractor model stands as a type of population clock models (Paton and D. V. Buonomano, 2018), with each attractor state representing a particular time in the song. In the emergence of each of these states, the individual neuronal and circuit dynamics are crucial, and the travel in the state space implies time. A previous model of the same class has been introduced by Laje and D. V. Buonomano (2013), called dynamic attractor. However, although this model accounts for robustness, the scalar property of timing and is an embodiment of the theoretical viewpoint of population clocks, another study (Pehlevan et al., 2018) has presented concerns regarding its flexibility to behavioral change. As an attractor, we show our model preserves the same robustness properties, while being flexible. Moreover, this is supported by a suggestion in the same study that flexibility could as well be a feature of effective feedforward networks, and an asymmetric Hopfield networks or other attractor models could provide similar results to a synfire chain. Taken together, our results show that this is possible with a ring model.

Localized Model of Motor Timing

HVC can be conceptually modeled as a local or distributed timing area. The first hypothesis considers HVC as the sole driving force of timing, and the second proposes that timing emerges from the interaction between multiple areas including

HVC (Elmaleh et al., 2021; Hamaguchi, Tanaka, et al., 2016). In our model, the external inputs reaching HVC, were considered to be on average homogenous and constant, which would make HVC a localized and *effectively autonomous* system. We focus only on modeling the sequential activation observed there. This activity is likely due to a functionally feedforward network and is generally modeled with synaptic chains. However, while being functionally feedforward, the underlying connectivity in HVC is recurrent (R. Hahnloser et al., 2002; Hamaguchi, Tanaka, et al., 2016; Yamashita et al., 2008).

A connection probability of 0.11 (Mooney and J. F. Prather, 2005) was reported between HVC_{RA} neurons. With a ring attractor model, we approach the suggested recurrent architecture with a narrow connectivity profile and through an asymmetric connectivity pattern, by adding a *bias* (β), we generate feedforwardness within it. The narrow connectivity profile is a Gaussian, the width of which is set based on the reported connection probabilities and the assumption that the connectivity must represent the sparse dynamics observed. Moreover, the added bias term makes the activity propagate in time, generating a sequence. The bias amplitude determines the speed and it is adapted to the sequence propagation speed in HVC. This is also set to simulate syllable durations in birdsong. Lastly, to our knowledge no autopses have been reported in HVC so we assign self-connections to 0 (as in Drew and Abbott (2003)).

We have also explored in Chapter 5 (section 5.3.2) rectangular connectivity pattern. We observe no reportably different dynamics. Hence, we assume that the chosen profile of a gaussian can be more biologically plausible, with the neurons sending variable (graded) synaptic weights to other neurons, rather than binary ones. In our proposed model, the rectangular profile could instead reflect an intermediary state between the random connections, prior to learning, and the final adult HVC network. This could map to observations in mammalian literature where sparsening (Komiyama et al., 2010) and temporal sharpening (Masamizu et al., 2014; A. J. Peters et al., 2014) of the neural activity occurs as motor learning progresses. Another connectivity profile to be considered, since the system is defined in a periodic ring is a von Misses (periodic normal distribution) connectivity profile. This can mostly be beneficial in exploring the system analytically.

Lastly, in the main models, inhibition is modeled as a non specific, global effect. Although experimental exploration have revealed patterns in inhibition, it is unclear whether it is a readout of the existing loop with HVC_{RA} neurons. This is explored with the EI network, both in terms of model dynamics and its flexibility.

Distributed Model of Motor Timing

Due to the highly connected position of HVC (projections from Uva, nucleus, Avalanche, CM, NIf and Field L) in the auditory pathway and the song system, it is important to consider the distributed hypothesis, investigate each of these structures and discuss their possible role in driving time. As we are focused on the temporal control of song, the question is whether these projections are present

during song production and to answer it, Vallentin and Long (2015) has shown that only motor inputs to HVC are present, possibly from Uva, NIf and other HVC_{RA} neurons. (i) We are already including the connections between HVC_{RA} neurons in the model. (ii) NIf has been shown not to be crucial in adult song production (Naie and R. Hahnloser, 2011), but numerous studies have shown Uva is and it is specifically involved in temporal control as well. (iii) Uva receives bilateral feedback projections from the brainstem nuclei and stimulation studies have revealed that focal stimulation in the brainstem inspiratory-related nucleus paraambigualis, can disrupt song timing (Ashmore, Wild, et al., 2005). Moreover, when Uva is lesioned bilaterally, permanent song disruption is observed (Coleman and E. T. Vu, 2005; H. Williams and D. S. Vicario, 1993). Recently, it has also been shown that the projections from thalamic nucleus Uva to HVC have a spatiotemporal structure (Moll et al., 2023). Altogether, these findings could suggest a role of Uva in interhemispheric synchronization (Coleman and E. T. Vu, 2005; Schmidt, 2003) .

In the light of these evidence, we add a model where Uva input is present to illustrate the effect that such an instructive signal would have on the dynamics. The parameters and model we create are based on recent experimental reports (Moll et al., 2023), where it was shown that 15% of HVC_{RA} neurons are *Uva-driven neurons*. The observed activity of Uva driven neurons at syllable onset/offset is in line with the observation of periods of interhemispheric synchronization time locked to syllable onset. In the model, the scalar input Uva is given only to the neurons present at syllable onset/offset. In a very noisy 'environment', or in the case of a delay in propagation in one of the HVCs, we observe that Uva input is able to quickly synchronize the activity. This shows the ability of our model to adapt to such an external input and that the model would be a biologically plausible explanation, even in the case of a distributed timing system.

However, in the learning mechanism, adding Uva would create significant challenges if the synaptic weights to the Uva driven neurons are plastic. If that is not the case, then Uva would just present some constrains in terms of how much the duration can be modified. This remains to be further explored, both computationally and experimentally.

7.2.2 Learning Rule for Behavioral Adaptation

Several models of reinforcement learning involving basal ganglia have been proposed in songbirds. The RL framework is also proposed for vertebrates' basal ganglia function. Although evidence has shown that learning in the temporal domain does not involve basal ganglia (Ali et al., 2013) and hence it is not modeled in this work, a general overview of the framework gives insight on the mechanisms involved.

The aforementioned models include template comparison (Doya and Sejnowski, 1998; M. Fee and Goldberg, 2011; I. R. Fiete, M. S. Fee, et al., 2007), efference copy (Troyer and A. J. Doupe, 2000) and inverse models (Hanuschkin et al., 2013) (review: Leblois and Darshan (2014)). The 'AFP comparison' hypothesis assumes that the song tem-

plate is stored in AFP and is compared in area X (Mooney, 2004; J. F. Prather et al., 2008; Sakata and Brainard, 2008) to the auditory information coming from HVC about the ongoing song. The result can then be directly transmitted through the AFP to the motor pathway (Doya and Sejnowski, 1998; Troyer and A. J. Doupe, 2000). However, this hypothesis assumes auditory feedback activity in the AFP, possibly coming from HVC, which is inconsistent with evidence showing no response to distorted auditory feedback in area X originating from HVC projections (A. A. Kozhevnikov and M. S. Fee, 2007; J. Prather et al., 2008) and LMAN (Leonardo, 2004), or within the basal ganglia (M. Fee and Goldberg, 2011).

An alternative hypothesis presented by (M. Fee and Goldberg, 2011), suggests that the song-quality evaluation signal, computed after comparison between the song template or tutor's memory and ongoing song (auditory cortex), is transmitted through neuromodulatory inputs such as dopamine, from the VTA to area X (Gale, Person, et al., 2008) as a reward prediction error (Gale and Perkel, 2010). Moreover, area X receives input from LMAN, accounting for variability and from HVC, which provides a global (fixed) representation of time. If activity in LMAN is correlated with good feedback, reward is provided and strengthening of the synaptic weights between HVC and X occurs. This then drives a temporal pattern of *bias* in LMAN in favor of improved song performance, through projections to DLM (loop: LMAN-X-DLM-LMAN). Their proposed learning rule reads: $\Delta W_{HVC_X} = \beta LHR$, where R represents the reward reaching the MSNs in area X from the VTA, and the eligibility trace is defined by H the activity of HVC neurons and L the activity of LMAN neurons. This hypothesis is in the same line with the computational model of I. R. Fiete, M. S. Fee, et al. (2007). It assumes the same actor-critic-experimenter RL scheme (Doya and Sejnowski, 1998), where LMAN is the experimenter, RA the actor and there is a scalar evaluation signal as a critic. Here, HVC sequence is assumed to be fixed, and only its mapping to RA is malleable, i.e. HVC RA synapses are plastic, whereas the synapses from LMAN to RA are considered empiric synapses, i.e. specialized for experimentation.

In both hypotheses, the role of area X as a critical area in learning is emphasized. This is consistent with behavioral adaptation in pitch change, where exploratory variability is provided by the thalamo-cortical circuit and area X is a key area of reinforcement learning. However, learning in the temporal domain is not well represented by this hypothesis, as it was shown that the AFP and the song-related basal ganglia circuits are not needed for temporal plasticity (Ali et al., 2013). Moreover, the findings of the study by Ali et al. (2013) suggest that the learning of these two features, namely pitch and timing, rely on two different circuitry and more specifically, learning in the time domain is expressed within the HVC itself. This suggests that the duration of the song segment is a function of the speed of the activity propagation and is supported by their recordings, which show a local stretching/shrinkage of the temporal structure at the level of a syllable targeted for modification. Below, we explain the interpretation of the RL rule in the context of learning in the temporal domain.

Biological Interpretation of the Learning Rule

In the light of these results, it is reasonable to assume that variability and the reward signal project to HVC, either directly or indirectly. In our approach, we use a reinforcement learning rule, first proposed by R. Williams (1992) and then used in the context of songbirds by I. R. Fiete and Seung (2006). This all is adapted to the ring model architecture. Our learning rule can be written as:

$$\Delta W_{\mathrm{HVC}_{RA}} = \gamma R e_{\mathrm{HVC}_{RA}} \quad \text{, with :} \tag{7.1}$$

$$e_{\mathrm{HVC}_{RA}} = \int_0^t \frac{dt'}{\tau_e} e^{-(t-t')/\tau_e} \eta_{\mathrm{HVC}_{RA}}(t') m_{\mathrm{HVC}_{RA}}(t'), \tag{7.2}$$

where R is a reward value of 0 and 1 and the weight changes are dependent on the eligibility trace $e_{HVC_{RA}}$. The eligibility trace controls the synaptic change amplitude and flags whether a synapse is *eligible* for modification, based on recent activation of the plastic synapse. Its value is increased during the coactivation of pre and postsynaptic activity, and decays exponentially with a time constant (τ_e), which should roughly match the maximum delay of the reinforcers and hence, link synaptic processes to behavior. Since HVC signal precedes motor (timing) activity and thus reward (white noise in CAF) by 35 ms (Ali et al., 2013), τ_e was chosen of this value (Pehlevan et al., 2018). Evidence of the presence of an eligibility trace is supported by several experimental studies in the striatum (Yagishita et al., 2014), cortex (He et al., 2015), and hippocampus (Brzosko et al., 2015) (review: (Gerstner, Lehmann, et al., 2018)). Taken together, theoretical and experimental evidence suggest that the eligibility trace is representative of neuromodulatory effect.

In the context of reward and/or surprise, and in the case of songbird learning, it could be provided to HVC through midbrain (A11: mesencephalic central gray) (Appeltants et al., 2000; Hamaguchi and Mooney, 2012) and/or PAG (Tanaka et al., 2018) dopaminergic projection neurons and/or norepinephrine from noradrenergic nucleus subceruleus (A6) (Appeltants et al., 2000).

The experimenter, is considered to be the noise reaching HVC, $\eta_{\mathrm{HVC}_{RA}}$ and could biologically map to the noisy synaptic activity within HVC itself or LMAN, which could reach HVC indirectly through A11 (Hamaguchi and Mooney, 2012) and/or Ad (Gale, Person, et al., 2008) . However, A11 cannot project to HVC both as a reward and as the experimenter, as likely they are mutually exclusive. The hypothesis of possible indirect projections through LMAN driving noise, is supported by the finding that a lesion in LMAN reduces learning rate and variability. Hence, LMAN may be a source of variability but possibly not be the only one.

The actor is the activity of the HVC_{RA} neurons, represented by $m_{HVC_{RA}}$ (in the rate based model).The threshold used to determine the presence or absence of white noise (negative reward) was updated on every trial, based on the average syllable duration, as it was done experimentally (Ali et al., 2013; Ursu, 2023).

Altogether, the learning rule reflects a biologically plausible mechanism of how the synapses in HVC are updated to modify syllable duration in response to a

reward profile, coming in the form of an *perturbatory* auditory feedback. It is in accordance with the proposed conceptual mechanism proposed and allows us to investigate timing adaptation as a separate process, independent of other behavioral components.

Confrontation of Simulation Results to Behavioral Data

When the findings following this learning approach are confronted to experimental data, we get interesting results.

(i) First, previous results (Ali et al., 2013; Pehlevan et al., 2018), have shown that the bird is able to adaptively modify motor timing of a particular interval, without showing any effects on the other intervals. This is a critical finding of behavioral flexibility and could translate to human speech and other kind of motor sequence production involved in pattern timing. It suggests that one "syllable" in a behavioral sequence can be modified independently of the rest of the sequence.

Our results with the rate based model support this finding. The model is able to learn to change the duration of a syllable in either direction (shortening, lengthening), without any interference on adjacent or nonadjacent syllables. However, these results was more nuanced when implemented in a SNN. In the case of shortening, the general averaged effect remained consistent, i.e. no interference was observed on other syllables. However, upon looking at single runs, in all the three models (exponential IF, AdEx and EI), we observed an effect at the level of the next syllable. This effect was averaged out across runs, which could also be the case in the presented experimental findings. Moreover, in the case of lengthening, interference was present at the level of the next syllable.

(ii) To test whether this is the case, and investigate if at any point during the learning process, there is an effect (either a daily effect or a small effect) in the following syllable, we analyzed data previously collected in the lab. As in our simulations, the averaged results show no significant interference in the other syllables. The change in magnitude is too small ($<2\sigma$) in comparison to that of the target syllable.

However, individual birds show different cases of interference present in the following syllable. Across the birds who have such effects (25%), the effect is consistently in the opposite direction. The presence of interference in some birds, and not in others could be due to different learning mechanisms, as different birds present variability (individuality) during learning. An example of such variability is the different learning rates. Previously presented results showing no interference, could be due to averaging effects or it might indicate that we need a larger sample size. Moreover, as in our simulations, an asymmetry in duration shift could also be possible. As in the original study both duration shortening and lengthening are grouped, this asymmetry is hard to assess. On the other hand, our data comprises of only duration lengthening, so a comparison between the two protocols is at this time, not possible.

7.3 Limitations

In this work, we start with a rate based model to test our hypothesis, we explore different kinds of neuronal complexity, while aiming for a low-level representation and biological plausibility. We do not use conductance based or a Hodgkin Huxley model, but we believe our assumptions are cost effective and biologically sufficient. Moreover, we notice that the observed mechanisms are conserved across the different complexity levels of our models. However, there are certain limitations to our model, that have not been captured by the model and could incite future work and/or could have affected our results.

7.3.1 Cellular and Network Properties

Primarily, we are modeling only one cell type, the HVC_{RA} neuron. This is based on evidence that these cells are associated with motor timing, shown both from recordings of time-locked activity to the song (R. Hahnloser et al., 2002) and from cooling/heating manipulations which have an effect on their burst time in the sequence, and consequently to the song too (Hamaguchi, Tanaka, et al., 2016; Long and M. S. Fee, 2008). However, interneurons and HVC_X neurons could also have a role in song production.

(i) For instance, bilateral inhibition in HVC of interneurons through Gabazine (Kosche et al., 2015) leads to a degradation in song structure, highlighting their importance in maintenance of activity. We tested this effect in our one population SNN model, where inhibition is modeled as a global variable. Upon administration of certain (detailed in Chapter 5) levels of inhibition, the sequence production degrades. Moreover, we show that the model is much more sensitive to removal of inhibition, in comparison to addition. This is because inhibition allows a localized bump formation and its removal could drive it to amplitude instability, which may map to what is occurring experimentally. Additionally, in one of the models we also added interneurons (Subsection 7.1.2).

(ii) HVC_X neurons are connected to both HVC_{RA} and interneurons, and could potentially have an effect on the resultant sequence. However, in our model, this effect was assumed negligible due to experimental findings showing that selective ablation of X projecting neurons had no effect on song production (Scharff, Kirn, et al., 2000).

In modeling HVC_{RA} neurons, an important factor is the burst activity observed, and whether this is driven by network dynamics and input or whether it is rather an intrinsic property of the neurons. From an experimental perspective, Long, D. Z. Jin, et al. (2010) have identified L-type calcium channels to have a role in generating or initiating bursting activity in HVC_{RA} neurons and have suggested that bursts are calcium-mediated. From a computational perspective, D. Z. Jin et al. (2007) have shown that an intrinsic burst mechanism increases the robustness of the synaptic chains used to model HVC.

In two different models (D. Z. Jin et al., 2007; Long, D. Z. Jin, et al., 2010), bursting neurons were simulated with two compartment model, a somatic and dendritic compartment, containing conductances for generating calcium spikes. However, these spikes were longer in duration than the HVC_{RA} burst, so D. Jin et al. (2007) propose that burst duration could be defined by either a shorter dendritic spike or by the contribution of strong somatic spike frequency adaptation. Another single compartment approach (Daou et al., 2013), constructed detailed Hodgkin-Huxley models of the three main model groups in HVC, and showed through experiments the importance of each conductance, highlighting the effect of calcium-activated potassium currents and an A-type potassium current for HVC_{RA}. Although our model does not use conductance based synapses, we use adaptation in the framework of an AdEx model, where the b parameter is assumed to account for the the effect of calcium dependent potassium channels.

Another important component driving network dynamics are synaptic delays. In our model with rectangular connectivity profile, we tested the effect of adding delays as a function of distance, taken from the distibution provided by Egger et al. (2020). Although in their model, the effect of delays is significant in activity propagation, it does not seem to convey the same importance in ours. However, further explorations in the Gaussian connectivity profile could be done, but we do not expect any changes. In the light of neural field theory, delays have shown to provide a travelling wave (Palkar et al., 2023), which translated to our model could have an effect on bump propagation speed. This has not been tested in this work.

7.3.2 HVC in Bengalese Finches and Canaries

Our findings are representative of a stereotypical, linear syntax, such as the courtship song of the male zebra finch. However, in the presence of branching, or transition probabilities, where the motif could either be ABCE or ABDE (where A, B, C, D and E stand for syllables) as observed in Bengalese finches and canaries, they may not be applicable at the same extent. That does not solely depend on the chosen computational approach, but more on the role of HVC. For instance, alongside its role in timing, experiments have shown that cooling HVC also affects transition probabilities (Y. S. Zhang et al., 2017) in Bengalese finches, suggesting that the precise burst time encodes transitions. How a ring attractor model could account for branching, based on spike-times is still unclear. We could imagine two or more rings sharing part of the same trajectory and then branching to one or the other ring. The branching mechanism could be probabilistic, and the transition markovian as in a previous model by D. Z. Jin (2009), where the synaptic chains would be replaced by rings.

7.4 Perspectives

7.4.1 A Continuum? Synaptic Chains to Asymmetric Ring Models

An interesting proposal is that the synaptic chains represented by synfire chains in HVC models, form a continuum with the asymmetric ring model. We compute phase diagrams by investigating the network's evolution after a high perturbation is provided. In the spiking model (exponential IF) with bias, alongside the three main states reported in the rate model, we observe a 'hopping' bump state, where spike propagation is not continuous but rather jump-like. It manifests as populations of neurons active at approximately the same time with some jitter due to noise. This pattern of activation is reminiscent of the synfire chain, where neurons are organized in layers.

This propagation/chain pattern is present at lower levels of excitation (W_2) and when the ratio between excitation and inhibition is low. We also observe the same hopping bump phase when the ratio between the bias and width of the connectivity profile is high. Although these observations have not been quantified, they show that a feedforward network with recurrence, displays similarity in different regimes with a purely feedforward architecture. A systematic investigation of the effect of connectivity width, could reveal that in fact synfire chains and attractors lie at two different ends of the same continuum.

7.4.2 Initial Learning in Unstructured Networks

In this study, we assume a gaussian connectivity profile within HVC neurons. This is based on experimental findings, showing the percentage of HVC_{RA} connections (Mooney and J. F. Prather, 2005) and the reported strongly interconnected web (Kosche et al., 2015), with structured excitation and inhibition. We also add a bias term in the connectivity matrix, which accounts for an effectively feedforward network. However, whether this structure emerges, or is learned during sensorimotor learning is a question we have not addressed. Based on the behavior observed in the chosen animal model (zebra finch), the connectivity pattern and bias could be learned during the sensorimotor phase with exposure to the tutor's song. Part of the observed sequential dynamics is also encoded in intrinsic properties (adaptation) of the neurons or relies on influence from the brainstem areas (respiratory-related information) and likely has a genetic component (Fehér et al., 2009b). These three factors can be interconnected and possibly pose constrains on the produced song and interact with the connectivity, so they should be considered as well.

Learning could drive the development of the neural sequences, through an external source, in this case the tutor, which repeats the same sequence and entrains the neurons in HVC to spike in a certain order (Amari, 1972; Nowotny et al., 2003). Another hypothesis suggests that the timing mechanism in HVC develops before song acquisition (Jun and D. Z. Jin, 2007). In this work, they use spike time de-

pendent plasticity (STDP), axon remodeling and synaptic decay to induce a self-organization process leading to a long synfire chain. A similar approach (I. R. Fiete, M. S. Fee, et al., 2007) uses STDP and heterosynaptic competition, which guarantees spatiotemporal distribution of the activity and consequently the unary sequential code in HVC. Moreover, in this study when giving temporally structured input, the learning is rapid. We would like to pursue a similar approach and observe whether with a similar learning rule the model can learn a simplified asymmetric ring architecture. In line with Jun and D. Z. Jin (2007)'s proposal, it could be that the recurrent architecture in HVC is learned before tutor exposure, and it is only the bias (speed) that is learned through supervised learning. It could also be that the initial structure is binary, as in the rectangle connectivity profile, and the gaussian and bias are both learned.

7.4.3 Return to Baseline Duration

Experimental evidence (Ali et al., 2013; Pehlevan et al., 2018; Ursu, 2023) have shown that following the CAF paradigm, the bird's syllable duration returns to baseline. This suggests a possible conservation mechanism, or as previously called *template reinforcement mechanism*, which rewards duration changes toward the baseline duration. The area responsible for such a return has been identified to be Av and its lesion prevents the bird to return to baseline. In the model we have presented, we could say we simulate a case where there is such a lesion and to account for its presence we would need to add another learning rule (Roberts et al., 2017).

The learning rule would be effective at the level of each syllable and would be added to the CAF learning rule. It reinforces the synapses based on where in the baseline normal distribution the current duration is. Hence, the further away from the mean it goes, the more the reward approaches 0. In the case of the addition of the reinforcement learning rule (CAF), the latter would overcome the conservation reward (due to its size) and minimize its effect, thus still allowing duration shift to occur. Once the reward from CAF is taken out, the conservation learning rule should be able to take the syllable duration to the baseline duration.

In a previous study, a template reinforcer learning rule was implemented (Pehlevan et al., 2018) in addition to the learning rule. No effect was observed in terms of increase in specificity, but an effect was observed in the return. There, the template reinforcer rule provides reward only after the mean of the duration distribution exceeds the mean of the baseline by 0.1ms, in the opposite direction of this excess. For instance, if mean duration distribution is higher than mean baseline distribution by at least 0.1ms, all duration values less then the mean duration distribution are awarded. This resembles a bimodal reward, drifting the syllable distribution to the mean every time it 'escapes' it. Although such a mechanism works, a scenario when the bird is rewarded every time the duration is within the baseline distribution, as we propose, can also be applicable.

List of Figures

List of Tables

Bibliography

1. M. Abeles. "Local Cortical Circuits: An Electrophysiological study." *Springer Berlin.*, 1982.
2. C. Addyman, R. M. French, and E. Thomas. "Computational models of interval timing". *Current Opinion in Behavioral Sciences* 8, 2016, pp. 140–146. DOI: 10.1016/j.cobeha.2016.01.004. URL: https://doi.org/10.1016/j.cobeha.2016.01.004.
3. E. Aksay, G. Gamkrelidze, H. S. Seung, R. Baker, and D. W. Tank. "In vivo intracellular recording and perturbation of persistent activity in a neural integrator". *Nature neuroscience* 4:2, 2001, pp. 184–193.
4. E. Akutagawa and M. Konishi. "New brain pathways found in the vocal control system of a songbird". *The Journal of Comparative Neurology* 518:15, 2010, pp. 3086–3100. DOI: 10.1002/cne.22383. URL: https://doi.org/10.1002/cne.22383.
5. F. Ali, T. M. Otchy, C. Pehlevan, A. L. Fantana, Y. Burak, and B. P. Ölveczky. "The Basal Ganglia Is Necessary for Learning Spectral, but Not Temporal, Features of Birdsong". *Neuron* 80:2, 2013, pp. 494–506. DOI: 10.1016/j.neuron.2013.07.049. URL: https://doi.org/10.1016/j.neuron.2013.07.049.
6. R. G. Alonso, M. A. Trevisan, A. Amador, F. Goller, and G. B. Mindlin. "A circular model for song motor control in Serinus canaria". *Frontiers in Computational Neuroscience* 9, 2015. DOI: 10.3389/fncom.2015.00041. URL: https://doi.org/10.3389/fncom.2015.00041.
7. S. Amari. "Learning patterns and pattern sequences by self-organizing", 1972.
8. A. S. Andalman and M. S. Fee. "A basal ganglia-forebrain circuit in the songbird biases motor output to avoid vocal errors". *Proceedings of the National Academy of Sciences* 106:30, 2009, pp. 12518–12523. DOI: 10.1073/pnas.0903214106. URL: https://doi.org/10.1073/pnas.0903214106.
9. A. S. Andalman, J. N. Foerster, and M. S. Fee. "Control of Vocal and Respiratory Patterns in Birdsong: Dissection of Forebrain and Brainstem Mechanisms Using Temperature". *PLoS ONE* 6:9, 2011. Ed. by S. J. Kiebel, e25461. DOI: 10.1371/journal.pone.0025461. URL: https://doi.org/10.1371/journal.pone.0025461.

10. D. Appeltants, P. Absil, J. Balthazart, and G. F. Ball. "Identification of the origin of catecholaminergic inputs to HVc in canaries by retrograde tract tracing combined with tyrosine hydroxylase immunocytochemistry". *Journal of chemical neuroanatomy* 18:3, 2000, pp. 117–133.
11. M. Araki, M. M. Bandi, and Y. Yazaki-Sugiyama. "Mind the gap: Neural coding of species identity in birdsong prosody." *Science*, 2016. DOI: 10.1126/science.aah6799.
12. M. Araki, M. M. Bandi, and Y. Yazaki-Sugiyama. "Mind the gap: Neural coding of species identity in birdsong prosody". *Science* 354:6317, 2016, pp. 1282–1287. DOI: 10.1126/science.aah6799. URL: https://doi.org/10.1126/science.aah6799.
13. D. Arnold and D. Robinson. "The oculomotor integrator: testing of a neural network model". *Experimental brain research* 113, 1997, pp. 57–74.
14. D. Aronov, A. S. Andalman, and M. S. Fee. "A Specialized Forebrain Circuit for Vocal Babbling in the Juvenile Songbird". *Science* 320:5876, 2008, pp. 630–634. DOI: 10.1126/science.1155140. URL: https://doi.org/10.1126/science.1155140.
15. J. Artieda, M. A. Pastor, F. Lacruz, and J. A. Obeso. "Temporal discrimination is abnormal in Parkinson's disease". *Brain* 115:1, 1992, pp. 199–210.
16. R. C. Ashmore, J. A. Renk, and M. F. Schmidt. "Bottom-Up Activation of the Vocal Motor Forebrain by the Respiratory Brainstem". *The Journal of Neuroscience* 28:10, 2008, pp. 2613–2623. DOI: 10.1523/jneurosci.4547-07.2008. URL: https://doi.org/10.1523/jneurosci.4547-07.2008.
17. R. C. Ashmore, J. M. Wild, and M. F. Schmidt. "Brainstem and Forebrain Contributions to the Generation of Learned Motor Behaviors for Song". *The Journal of Neuroscience* 25:37, 2005, pp. 8543–8554. DOI: 10.1523/jneurosci.1668-05.2005. URL: https://doi.org/10.1523/jneurosci.1668-05.2005.
18. A. Aussel, L. Buhry, and R. Ranta. "Stability conditions of Hopfield ring networks with discontinuous piecewise-affine activation functions". In: *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*. IEEE. 2017, pp. 3350–3355.
19. L. Bai and D. Breen. "Calculating center of mass in an unbounded 2D environment". *Journal of Graphics Tools* 13:4, 2008, pp. 53–60.
20. K. I. Bakhurin, V. Goudar, J. L. Shobe, L. D. Claar, D. V. Buonomano, and S. C. Masmanidis. "Differential encoding of time by prefrontal and striatal network dynamics". *Journal of Neuroscience* 37:4, 2017, pp. 854–870.
21. F. Balci and P. Simen. "A Decision Model of Timing." *Current Opinion in Behavioral Sciences* 8, 2016. DOI: 10.1016/j.cobeha.2016.02.002.
22. R. P. Balda and A. C. Kamil. "Long-term spatial memory in Clark's nutcracker, Nucifraga columbiana". *Animal Behaviour* 44:4, 1992, pp. 761–769.

23. R. Ben-Yishai, R. L. Bar-Or, and H. Sompolinsky. "Theory of orientation tuning in visual cortex." *Proceedings of the National Academy of Sciences* 92:9, 1995, pp. 3844–3848. DOI: 10.1073/pnas.92.9.3844. URL: https://doi.org/10.1073/pnas.92.9.3844.

24. R. Ben-Yishai, R. L. Bar-Or, and H. Sompolinsky. "Theory of orientation tuning in visual cortex." *Proceedings of the National Academy of Sciences* 92:9, 1995, pp. 3844–3848.

25. H. Bergson. *Essai sur les données immédiates de la conscience, Paris, PUF, coll.«* 1889.

26. R. A. Block and D. Zakay. "Prospective and retrospective duration judgments: A meta-analytic review". *Psychonomic bulletin & review* 4:2, 1997, pp. 184–197.

27. T. Boraud, A. Leblois, and N. P. Rougier. "A natural history of skills". *Progress in Neurobiology* 171, 2018, pp. 114–124.

28. L. Boroditsky. "Language and the construction of time through space". *Trends in neurosciences* 41:10, 2018, pp. 651–653.

29. L. Boroditsky, O. Fuhrman, and K. McCormick. "Do English and Mandarin speakers think about time differently?" *Cognition* 118:1, 2011, pp. 123–129.

30. L. Boroditsky and A. Gaby. "Remembrances of times East: absolute spatial representations of time in an Australian aboriginal community". *Psychological science* 21:11, 2010, pp. 1635–1639.

31. S. W. Bottjer, E. A. Miesner, and A. P. Arnold. "Forebrain Lesions Disrupt Development But Not Maintenance of Song in Passerine Birds". *Science* 224:4651, 1984, pp. 901–903. DOI: 10.1126/science.6719123. URL: https://doi.org/10.1126/science.6719123.

32. M. S. Brainard and A. J. Doupe. "Translating Birdsong: Songbirds as a Model for Basic and Applied Medical Research". *Annual Review of Neuroscience* 36:1, 2013, pp. 489–517. DOI: 10.1146/annurev-neuro-060909-152826. URL: https://doi.org/10.1146/annurev-neuro-060909-152826.

33. M. S. Brainard and A. J. Doupe. "What songbirds teach us about learning". *Nature* 417:6886, 2002, pp. 351–358. DOI: 10.1038/417351a. URL: https://doi.org/10.1038/417351a.

34. A. Breska and R. B. Ivry. "Taxonomies of timing: where does the cerebellum fit in?" *Current Opinion in Behavioral Sciences* 8, 2016, pp. 282–288. DOI: 10.1016/j.cobeha.2016.02.034. URL: https://doi.org/10.1016/j.cobeha.2016.02.034.

35. R. Brette and W. Gerstner. "Adaptive Exponential Integrate-and-Fire Model as an Effective Description of Neuronal Activity." *J. Neurophysiol. 94*, 2005.

36. M. Brosch, E. Selezneva, and H. Scheich. "Nonauditory events of a behavioral procedure activate auditory cortex of highly trained monkeys". *Journal of Neuroscience* 25:29, 2005, pp. 6797–6806.

37. Z. Brzosko, W. Schultz, and O. Paulsen. "Retroactive modulation of spike timing-dependent plasticity by dopamine". *eLife* 4, 2015. DOI: 10.7554/elife.09685. URL: https://doi.org/10.7554/elife.09685.

38. D. Bueti. "The Sensory Representation of Time". *Frontiers in Integrative Neuroscience* 5, 2011. DOI: 10.3389/fnint.2011.00034. URL: https://doi.org/10.3389/fnint.2011.00034.

39. D. Bueti and D. V. Buonomano. "Temporal Perceptual Learning". *Timing and Time Perception* 2:3, 2014, pp. 261–289. DOI: 10.1163/22134468-00002023. URL: https://doi.org/10.1163/22134468-00002023.

40. C. V. Buhusi and W. H. Meck. "What makes us tick? Functional and neural mechanisms of interval timing". *Nature Reviews Neuroscience* 6:10, 2005, pp. 755–765. DOI: 10.1038/nrn1764. URL: https://doi.org/10.1038/nrn1764.

41. D. Bullock, J. C. Fiala, and S. Grossberg. "A neural model of timed response learning in the cerebellum". *Neural Networks* 7:6-7, 1994, pp. 1101–1114. DOI: 10.1016/s0893-6080(05)80161-3. URL: https://doi.org/10.1016/s0893-6080(05)80161-3.

42. D. Buonomano. *Your brain is a time machine: The neuroscience and physics of time*. WW Norton & Company, 2017.

43. D. Buonomano and C. Rovelli. "Bridging the neuroscience and physics of time". *arXiv preprint arXiv:2110.01976*, 2021.

44. D. V. Buonomano. "Decoding temporal information: a model based on short-term synaptic plasticity". *Journal of Neuroscience* 20:3, 2000, pp. 1129–1141.

45. D. V. Buonomano and M. M. Merzenich. "Temporal information transformed into a spatial code by a neural network with realistic properties". *Science* 267:5200, 1995, pp. 1028–1030.

46. D. V. Buonomano, J. Bramen, and M. Khodadadifar. "Influence of the interstimulus interval on temporal processing and learning: testing the state-dependent network model". *Philosophical Transactions of the Royal Society B: Biological Sciences* 364:1525, 2009, pp. 1865–1873. DOI: 10.1098/rstb.2009.0019.

47. D. V. Buonomano and R. Laje. "Population clocks: motor timing with neural dynamics". *Trends in Cognitive Sciences* 14:12, 2010, pp. 520–527. DOI: 10.1016/j.tics.2010.09.002.

48. D. V. Buonomano and M. D. Mauk. "Neural Network Model of the Cerebellum: Temporal Discrimination and the Timing of Motor Responses". *Neural Computation* 6:1, 1994, pp. 38–55. DOI: 10.1162/neco.1994.6.1.38. URL: https://doi.org/10.1162/neco.1994.6.1.38.

49. Y. Burak and I. R. Fiete. "Accurate path integration in continuous attractor network models of grid cells". *PLoS computational biology* 5:2, 2009, e1000291.

50. D. Burr, A. Tozzi, and M. C. Morrone. "Neural mechanisms for timing visual events are spatially selective in real-world coordinates". *Nature neuroscience* 10:4, 2007, pp. 423–425.
51. J. A. Cardin, J. N. Raksin, and M. F. Schmidt. "Sensorimotor Nucleus NIf Is Necessary for Auditory Processing But Not Vocal Motor Output in the Avian Song System". *Journal of Neurophysiology* 93:4, 2005, pp. 2157–2166. DOI: 10.1152/jn.01001.2004. URL: https://doi.org/10.1152/jn.01001.2004.
52. B. A. Carlson. "Temporal-Pattern Recognition by Single Neurons in a Sensory Pathway Devoted to Social Communication Behavior". *The Journal of Neuroscience* 29:30, 2009, pp. 9417–9428. DOI: 10.1523/jneurosci.1980-09.2009. URL: https://doi.org/10.1523/jneurosci.1980-09.2009.
53. C. E. Carr. "Processing of Temporal Information in the Brain". *Annual Review of Neuroscience* 16:1, 1993, pp. 223–243. DOI: 10.1146/annurev.ne.16.030193.001255. URL: https://doi.org/10.1146/annurev.ne.16.030193.001255.
54. D. Casasanto and L. Boroditsky. "Time in the mind: Using space to think about time". *Cognition* 106:2, 2008, pp. 579–593. DOI: 10.1016/j.cognition.2007.03.004.
55. W. Chang and D. Z. Jin. "Spike propagation in driven chain networks with dominant global inhibition". *Physical Review E* 79:5, 2009. DOI: 10.1103/physreve.79.051917. URL: https://doi.org/10.1103/physreve.79.051917.
56. R. Chaudhuri and I. Fiete. "Bipartite expander Hopfield networks as self-decoding high-capacity error correcting codes". *Advances in neural information processing systems* 32, 2019.
57. R. M. Church and H. A. Broadbent. "A connectionist model of timing." *Neural network models of conditioning and action*, 1991, pp. 225–240.
58. A. Clark. "Time and mind". *The Journal of Philosophy* 95:7, 1998, pp. 354–376.
59. N. S. Clayton and A. Dickinson. "Episodic-like memory during cache recovery by scrub jays". *Nature* 395:6699, 1998, pp. 272–274.
60. M. J. Coleman and E. T. Vu. "Recovery of impaired songs following unilateral but not bilateral lesions of nucleus uvaeformis of adult zebra finches". *Journal of Neurobiology* 63:1, 2005, pp. 70–89. DOI: 10.1002/neu.20122. URL: https://doi.org/10.1002/neu.20122.
61. A. Compte, N. Brunel, P. S. Goldman-Rakic, and X.-J. Wang. "Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model". *Cerebral cortex* 10:9, 2000, pp. 910–923.
62. J. T. Coull, R.-K. Cheng, and W. H. Meck. "Neuroanatomical and neurochemical substrates of timing". *Neuropsychopharmacology* 36:1, 2011, pp. 3–25.
63. J. Coull and A. Nobre. "Dissociating explicit timing from temporal expectation with fMRI". *Current Opinion in Neurobiology* 18:2, 2008, pp. 137–144. DOI: 10.1016/j.conb.2008.07.011. URL: https://doi.org/10.1016/j.conb.2008.07.011.

64. E. Covey and J. H. Casseday. "TIMING IN THE AUDITORY SYSTEM OF THE BAT". *Annual Review of Physiology* 61:1, 1999, pp. 457–476. DOI: 10.1146/annurev.physiol.61.1.457. URL: https://doi.org/10.1146/annurev.physiol.61.1.457.

65. C. Creelman. "Human discrimination of auditory duration." *Journal of the Acoustical Society of America,34, 582–593*, 1962.

66. A. Daou, M. T. Ross, F. Johnson, R. L. Hyson, and R. Bertram. "Electrophysiological characterization and computational models of HVC neurons in the zebra finch". *Journal of Neurophysiology* 110:5, 2013, pp. 1227–1245. DOI: 10.1152/jn.00162.2013. URL: https://doi.org/10.1152/jn.00162.2013.

67. M. Diesmann, G. M.-O., and A. A. "Stable propagation of synchronous spiking in cortical neural networks." *Nature*, 1999.

68. A. Doupe, D. Perkel, A. Reiner, and E. Stern. "Birdbrains could teach basal ganglia research a new song". *Trends in Neurosciences* 28:7, 2005, pp. 353–363. DOI: 10.1016/j.tins.2005.05.005. URL: https://doi.org/10.1016/j.tins.2005.05.005.

69. A. J. Doupe and P. K. Kuhl. "BIRDSONG AND HUMAN SPEECH: Common Themes and Mechanisms". *Annual Review of Neuroscience* 22:1, 1999, pp. 567–631. DOI: 10.1146/annurev.neuro.22.1.567. URL: https://doi.org/10.1146/annurev.neuro.22.1.567.

70. K. Doya and T. J. Sejnowski. "A computational model of birdsong learning by auditory experience and auditory feedback", 1998, pp. 77–88.

71. P. J. Drew and L. F. Abbott. "Model of Song Selectivity and Sequence Generation in Area HVc of the Songbird". *Journal of Neurophysiology* 89:5, 2003, pp. 2697–2706. DOI: 10.1152/jn.00801.2002. URL: https://doi.org/10.1152/jn.00801.2002.

72. R. Dum and P. Strick. "Motor areas in the frontal lobe of the primate". *Physiology and Behavior* 77:4-5, 2002, pp. 677–682. DOI: 10.1016/s0031-9384(02)00929-0. URL: https://doi.org/10.1016/s0031-9384(02)00929-0.

73. D. Durstewitz. "Self-Organizing Neural Integrator Predicts Interval Times through Climbing Activity." *Journal of Neuroscience*, 2003. DOI: 10.1523/JNEUROSCI.23-12-05342.

74. D. M. Eagleman. "Human time perception and its illusions". *Current opinion in neurobiology* 18:2, 2008, pp. 131–136.

75. L. A. Eales. "Song learning in zebra finches: some effects of song model availability on what is learnt and when". *Animal Behaviour* 33:4, 1985, pp. 1293–1300. DOI: 10.1016/s0003-3472(85)80189-5. URL: https://doi.org/10.1016/s0003-3472(85)80189-5.

76. L. Edinger, A. Wallenberg, and G. M. Holmes. "Das Vorderhirn der Vögel." *Untersuchungen uber die vergleichende Anatomie des Gehirnes.* 20, 1903, pp. 343–426.

77. R. Egger, R. Tupikov, and M. E. et al. "Local Axonal Conduction Shapes the Spatiotemporal Properties of Neural Sequences." *Cell.*, 2020.

78. A. Einstein. "Die Feldgleichungen der Gravitation". *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften*, 1915, pp. 844–847.

79. M. Elmaleh, D. Kranz, A. Asensio, F. Moll, and M. Long. "Sleep replay reveals premotor circuit structure for a skilled behavior." *Neuron*, 2021. DOI: 10.1016/j.neuron.2021.09.021.

80. N. Emery and N. Clayton. "Effects of experience and social context on prospective caching strategies by scrub jays". *Nature* 416, 2001, pp. 443–446.

81. R. Epstein, R. P. Lanza, and B. F. Skinner. "Symbolic communication between two pigeons,(Columba livia domestica)". *Science* 207:4430, 1980, pp. 543–545.

82. M. Fee and J. Goldberg. "A hypothesis for basal ganglia-dependent reinforcement learning in the songbird". *Neuroscience* 198, 2011, pp. 152–170. DOI: 10.1016/j.neuroscience.2011.09.069. URL: https://doi.org/10.1016/j.neuroscience.2011.09.069.

83. M. S. Fee, A. A. Kozhevnikov, and R. H. Hahnloser. "Neural mechanisms of vocal sequence generation in the songbird". *Annals of the New York Academy of Sciences* 1016:1, 2004, pp. 153–170.

84. O. Fehér, H. Wang, S. Saar, P. P. Mitra, and O. Tchernichovski. "De novo establishment of wild-type song culture in the zebra finch." *Nature*, 2009. DOI: 10.1038/nature07994.

85. O. Fehér, H. Wang, S. Saar, P. P. Mitra, and O. Tchernichovski. "De novo establishment of wild-type song culture in the zebra finch". *Nature* 459:7246, 2009, pp. 564–568. DOI: 10.1038/nature07994. URL: https://doi.org/10.1038/nature07994.

86. L. von Fersen and J. D. Delius. "Long-term retention of many visual patterns by pigeons". *Ethology* 82:2, 1989, pp. 141–155.

87. J. C. Fiala, S. Grossberg, and D. Bullock. "Metabotropic Glutamate Receptor Activation in Cerebellar Purkinje Cells as Substrate for Adaptive Timing of the Classically Conditioned Eye-Blink Response". *The Journal of Neuroscience* 16:11, 1996, pp. 3760–3774. DOI: 10.1523/jneurosci.16-11-03760.1996. URL: https://doi.org/10.1523/jneurosci.16-11-03760.1996.

88. I. R. Fiete, M. S. Fee, and H. S. Seung. "Model of Birdsong Learning Based on Gradient Estimation by Dynamic Perturbation of Neural Conductances". *Journal of Neurophysiology* 98:4, 2007, pp. 2038–2057. DOI: 10.1152/jn.01311.2006. URL: https://doi.org/10.1152/jn.01311.2006.

89. I. R. Fiete and H. S. Seung. "Gradient Learning in Spiking Neural Networks by Dynamic Perturbation of Conductances". *Physical Review Letters* 97:4, 2006. DOI: 10.1103/physrevlett.97.048104. URL: https://doi.org/10.1103/physrevlett.97.048104.

90. H. Fujimoto, T. Hasegawa, and D. Watanabe. "Neural coding of syntactic structure in learned vocalizations in the songbird". *Journal of Neuroscience* 31:27, 2011, pp. 10023–10033.

91. T. Fujioka, L. J. Trainor, E. W. Large, and B. Ross. "Internalized timing of isochronous sounds is represented in neuromagnetic beta oscillations". *Journal of Neuroscience* 32:5, 2012, pp. 1791–1802.

92. S. Funahashi, C. J. Bruce, and P. S. Goldman-Rakic. "Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex". *Journal of neurophysiology* 61:2, 1989, pp. 331–349.

93. S. D. Gale and D. J. Perkel. "A Basal Ganglia Pathway Drives Selective Auditory Responses in Songbird Dopaminergic Neurons via Disinhibition". *The Journal of Neuroscience* 30:3, 2010, pp. 1027–1037. DOI: 10.1523/jneurosci.3585-09.2010. URL: https://doi.org/10.1523/jneurosci.3585-09.2010.

94. S. D. Gale, A. L. Person, and D. J. Perkel. "A novel basal ganglia pathway forms a loop linking a vocal learning circuit with its dopaminergic input". *The Journal of Comparative Neurology* 508:5, 2008, pp. 824–839. DOI: 10.1002/cne.21700. URL: https://doi.org/10.1002/cne.21700.

95. V. Gallese, L. Fadiga, L. Fogassi, and G. Rizzolatti. "Action recognition in the premotor cortex". *Brain* 119:2, 1996, pp. 593–609.

96. J. P. Gavornik and M. F. Bear. "Higher brain functions served by the lowly rodent primary visual cortex". *Learning & Memory* 21:10, 2014, pp. 527–533.

97. J. P. Gavornik, M. G. H. Shuler, Y. Loewenstein, M. F. Bear, and H. Z. Shouval. "Learning reward timing in cortex through reward dependent expression of synaptic plasticity". *Proceedings of the National Academy of Sciences* 106:16, 2009, pp. 6826–6831. DOI: 10.1073/pnas.0901835106. URL: https://doi.org/10.1073/pnas.0901835106.

98. T. Q. Gentner and D. Margoliash. "Neuronal populations and single cells representing learned auditory objects". *Nature* 424:6949, 2003, pp. 669–674. DOI: 10.1038/nature01731. URL: https://doi.org/10.1038/nature01731.

99. W. Gerstner and R. Brette. "Adaptive exponential integrate-and-fire model". *Scholarpedia* 4:6, 2009, p. 8427.

100. W. Gerstner, M. Lehmann, V. Liakoni, D. Corneil, and J. Brea. "Eligibility Traces and Plasticity on Behavioral Time Scales: Experimental Support of NeoHebbian Three-Factor Learning Rules". *Frontiers in Neural Circuits* 12, 2018. DOI: 10.3389/fncir.2018.00053. URL: https://doi.org/10.3389/fncir.2018.00053.

101. L. Gibb, T. Q. Gentner, and H. D. I. Abarbanel. "Brain Stem Feedback in a Computational Model of Birdsong Sequencing". *Journal of Neurophysiology* 102:3, 2009, pp. 1763–1778. DOI: 10.1152/jn.91154.2008. URL: https://doi.org/10.1152/jn.91154.2008.

102. J. Gibbon. "Scalar Expectancy Theory and Weber's Law in Animal Timing." *Psychological Review, 84, 279-325.*, 1977. DOI: 10.1037/0033-295X.84.3.279.

103. C. M. Glaze and T. W. Troyer. "Development of temporal structure in zebra finch song". *Journal of neurophysiology* 109:4, 2013, pp. 1025–1035.

104. C. M. Glaze and T. W. Troye. "Temporal Structure in Zebra Finch Song: Implications for Motor Coding." *Journal of Neuroscience*, 2006. DOI: 10.1523/JNEUROSCI.3387-05.2006.

105. C. M. Glaze and T. W. Troyer. "A Generative Model for Measuring Latent Timing Structure in Motor Sequences". *PLoS ONE* 7:7, 2012. Ed. by M. J. Coleman, e37616. DOI: 10.1371/journal.pone.0037616. URL: https://doi.org/10.1371/journal.pone.0037616.

106. S. Goldin-Meadow and C. Mylander. "Spontaneous sign systems created by deaf children in two cultures". *Nature* 391:6664, 1998, pp. 279–281.

107. T. S. Gouvêa, T. Monteiro, A. Motiwala, S. Soares, C. Machens, and J. J. Paton. "Striatal dynamics explain duration judgments". *eLife* 4, 2015. DOI: 10.7554/elife.11386. URL: https://doi.org/10.7554/elife.11386.

108. A. M. Graybiel. "The basal ganglia and chunking of action repertoires". *Neurobiology of learning and memory* 70:1-2, 1998, pp. 119–136.

109. J. Griffith. "On the stability of brain-like structures." *Biophys. J. 3:299-308.*, 1963.

110. S. Grondin. "Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions". *Attention, Perception, and Psychophysics* 72:3, 2010, pp. 561–582. DOI: 10.3758/app.72.3.561. URL: https://doi.org/10.3758/app.72.3.561.

111. S. Grondin and R. Rousseau. "Judging the relative duration of multimodal short empty time intervals". *Perception & psychophysics* 49:3, 1991, pp. 245–256.

112. S. Grossberg. "Some networks that can learn, remember, and reproduce any number of complicated space-time patterns." *I. J. Math. Mech., 19:53-91*, 1969.

113. S. Grossberg and N. A. Schmajuk. "Neural dynamics of adaptive timing and temporal discrimination during associative learning". *Neural Networks* 2:2, 1989, pp. 79–102. DOI: 10.1016/0893-6080(89)90026-9. URL: https://doi.org/10.1016/0893-6080(89)90026-9.

114. M. Grube, F. E. Cooper, P. F. Chinnery, and T. D. Griffiths. "Dissociation of duration-based and beat-based auditory timing in cerebellar degeneration". *Proceedings of the National Academy of Sciences* 107:25, 2010, pp. 11597–11601. DOI: 10.1073/pnas.0910473107. URL: https://doi.org/10.1073/pnas.0910473107.

115. O. Güntürkün. "The convergent evolution of neural substrates for cognition". *Psychological Research* 76:2, 2011, pp. 212–219. DOI: 10.1007/s00426-011-0377-9.

116. S. Haesler, K. Wada, A. Nshdejan, E. E. Morrisey, T. Lints, E. D. Jarvis, and C. Scharff. "FoxP2 expression in avian vocal learners and non-learners". *Journal of Neuroscience* 24:13, 2004, pp. 3164–3175.

117. R. Hahnloser, A. Kozhevnikov, and M. Fee. "An ultra-sparse code underlies the generation of neural sequences in a songbird." *Nature*, 2002. DOI: 10.1038/nature00974.

118. R. H. Hahnloser and M. S. Fee. "Sleep-related spike bursts in HVC are driven by the nucleus interface of the nidopallium". *Journal of neurophysiology* 97:1, 2007, pp. 423–435.

119. K. Hamaguchi and R. Mooney. "Recurrent Interactions between the Input and Output of a Songbird Cortico-Basal Ganglia Pathway Are Implicated in Vocal Sequence Variability". *The Journal of Neuroscience* 32:34, 2012, pp. 11671–11687. DOI: 10.1523/jneurosci.1666-12.2012. URL: https://doi.org/10.1523/jneurosci.1666-12.2012.

120. K. Hamaguchi, M. Tanaka, and R. Mooney. "A Distributed Recurrent Network Contributes to Temporally Precise Vocalizations". *Neuron* 91:3, 2016, pp. 680–693. DOI: 10.1016/j.neuron.2016.06.019. URL: https://doi.org/10.1016/j.neuron.2016.06.019.

121. D. Hansel and H. Sompolinsky. "Modeling feature selectivity in local cortical circuits." *Book Chapter*, 1998.

122. A. Hanuschkin, S. Ganguli, and R. H. R. Hahnloser. "A Hebbian learning rule gives rise to mirror neurons and links them to control theoretic inverse models". *Frontiers in Neural Circuits* 7, 2013. DOI: 10.3389/fncir.2013.00106. URL: https://doi.org/10.3389/fncir.2013.00106.

123. E. Hara, M. V. Rivas, J. M. Ward, K. Okanoya, and E. D. Jarvis. "Convergent differential regulation of parvalbumin in the brains of vocal learners". *PloS one* 7:1, 2012, e29457.

124. N. F. Hardy and D. V. Buonomano. "Neurocomputational models of interval and pattern timing". *Current Opinion in Behavioral Sciences* 8, 2016, pp. 250–257.

125. K. He, M. Huertas, S. Z. Hong, X. Tie, J. W. Hell, H. Shouval, and A. Kirkwood. "Distinct Eligibility Traces for LTP and LTD in Cortical Synapses". *Neuron* 88:3, 2015, pp. 528–538. DOI: 10.1016/j.neuron.2015.09.037. URL: https://doi.org/10.1016/j.neuron.2015.09.037.

126. J. J. Hopfield. "Neural networks and physical systems with emergent collective computational abilities." *Proceedings of the National Academy of Sciences* 79:8, 1982, pp. 2554–2558. DOI: 10.1073/pnas.79.8.2554. URL: https://doi.org/10.1073/pnas.79.8.2554.

127. F. Hyseni, N. Rougier, and A. Leblois. "A model of Sequence Timing in Songbird's Premotor Nucleus." In: *Student Conference on Biological Sciences (SCBS)*. Tirana, Albania, 2021. URL: 10.5281/zenodo.5877414.

128. F. Hyseni, N. Rougier, and A. Leblois. "An Attractor Model of the Temporal Dynamics of the Songbird's Premotor Nucleus." In: *32nd Annual Computational Neuroscience Meeting (CNS)*. Leipzig, Germany, 2023.

129. F. Hyseni, N. Rougier, and A. Leblois. "Comparative study of the synfire chain and ring attractor model for timing in the premotor nucleus in male Zebra Finches." In: *ESANN 2023-European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*. 2023, pp. 647–652. DOI: 10.14428/esann/2023.ES2023-120.

130. F. Hyseni, N. Rougier, and A. Leblois. "Temporal Dynamics in an Attractor Model of the Songbird's Premotor Nucleus." In: *Computational and Systems Neuroscience (COSYNE)*. Lisbon, Portugal, 2022. URL: https://www.world-wide.org/cosyne-22/temporal-dynamics-attractor-model-songbirds-b261b11a/.

131. F. Hyseni, N. P. Rougier, and A. Leblois. "Attractor dynamics drive flexible timing in birdsong." In: *International Conference on Artificial Neural Networks*. Springer. 2023, pp. 112–123. DOI: 10.1007/978-3-031-44198-1_10.

132. Y. Ikegaya, G. Aaron, R. Cossart, D. Aronov, I. Lampl, D. Ferster, and R. Yuste. "Synfire Chains and Cortical Songs: Temporal Modules of Cortical Activity". *Science* 304:5670, 2004, pp. 559–564. DOI: 10.1126/science.1093173. URL: https://doi.org/10.1126/science.1093173.

133. K. Immelmann. "Song development in the zebra finch and other estrildid finches." *(No Title)*, 1969, p. 61.

134. H. K. Inagaki, L. Fontolan, S. Romani, and K. Svoboda. "Discrete attractor dynamics underlies persistent activity in the frontal cortex". *Nature* 566:7743, 2019, pp. 212–217.

135. G. R. Isola, A. Vochin, and J. T. Sakata. "Manipulations of inhibition in cortical circuitry differentially affect spectral and temporal features of Bengalese finch song". *Journal of Neurophysiology* 123:2, 2020, pp. 815–830. DOI: 10.1152/jn.00142.2019. URL: https://doi.org/10.1152/jn.00142.2019.

136. R. Ivry and J. Schlerf. "Dedicated and intrinsic models of time perception." *Trends Cogn Sci.* 12(7), 2008, pp. 273–80. DOI: 10.1016/j.tics.2008.04.002.

137. R. B. Ivry and R. E. Hazeltine. "Perception and production of temporal intervals across a range of durations: evidence for a common timing mechanism." *Journal of Experimental Psychology: Human Perception and Performance* 21:1, 1995, p. 3.

138. E. D. Jarvis et al. "Avian brains and a new understanding of vertebrate brain evolution". *Nature Reviews Neuroscience* 6:2, 2005, pp. 151–159. DOI: 10.1038/nrn1606.

139. M. Jazayeri and M. N. Shadlen. "A neural mechanism for sensing and reproducing a time interval". *Current Biology* 25:20, 2015, pp. 2599–2609.

140. D. Jin, F. Ramazanoğlu, and H. Seung. "Intrinsic bursting enhances the robustness of a neural network model of sequence generation by avian brain area HVC." *J Comput Neurosci*, 2007.

141. D. Z. Jin. "Generating variable birdsong syllable sequences with branching chain networks in avian premotor nucleus HVC". *Physical Review E* 80:5, 2009. DOI: 10.1103/physreve.80.051902. URL: https://doi.org/10.1103/physreve.80.051902.

142. D. Z. Jin, F. M. Ramazanoğlu, and H. S. Seung. "Intrinsic bursting enhances the robustness of a neural network model of sequence generation by avian brain area HVC". *Journal of Computational Neuroscience* 23:3, 2007, pp. 283–299. DOI: 10.1007/s10827-007-0032-z. URL: https://doi.org/10.1007/s10827-007-0032-z.

143. J. K. Jun and D. Z. Jin. "Development of Neural Circuitry for Precise Temporal Sequences through Spontaneous Activity, Axon Remodeling, and Synaptic Plasticity". *PLoS ONE* 2:8, 2007. Ed. by S. Akbarian, e723. DOI: 10.1371/journal.pone.0000723. URL: https://doi.org/10.1371/journal.pone.0000723.

144. T. Kaneko. "Local connections of excitatory neurons in motor-associated cortical areas of the rat". *Frontiers in neural circuits* 7, 2013, p. 75.

145. U. R. Karmarkar and D. V. Buonomano. "Timing in the absence of clocks: encoding time in neural network states". *Neuron* 53:3, 2007, pp. 427–438.

146. U. R. Karmarkar and D. V. Buonomano. "Temporal Specificity of Perceptual Learning in an Auditory Discrimination Task". *Learning and Memory* 10:2, 2003, pp. 141–147. DOI: 10.1101/lm.55503.

147. R. Kawai, T. Markman, R. Poddar, R. Ko, A. L. Fantana, A. K. Dhawale, A. R. Kampff, and B. P. Ölveczky. "Motor Cortex Is Required for Learning but Not for Executing a Motor Skill". *Neuron* 86:3, 2015, pp. 800–812. DOI: 10.1016/j.neuron.2015.03.024. URL: https://doi.org/10.1016/j.neuron.2015.03.024.

148. M. Khona and I. Fiete. "Attractor and integrator networks in the brain." *Nat Rev Neurosci* 23, 2022, pp. 744–766. DOI: 10.1038/s41583-022-00642-0.

149. S. S. Kim, H. Rouault, S. Druckmann, and V. Jayaraman. "Ring attractor dynamics in the Drosophila central brain". *Science* 356:6340, 2017, pp. 849–853. DOI: 10.1126/science.aal4835. URL: https://doi.org/10.1126/science.aal4835.

150. D. P. King and J. S. Takahashi. "Molecular Genetics of Circadian Rhythms in Mammals". *Annual Review of Neuroscience* 23:1, 2000, pp. 713–742. DOI: 10.1146/annurev.neuro.23.1.713. URL: https://doi.org/10.1146/annurev.neuro.23.1.713.

151. E. I. Knudsen. "Instructed learning in the auditory localization pathway of the barn owl". *Nature* 417:6886, 2002, pp. 322–328.

152. T. Komiyama, T. R. Sato, and . K. Svoboda. "Learning-related fine-scale specificity imaged in motor cortex circuits of behaving mice". *Nature* 464:7292, 2010, pp. 1182–1186. DOI: 10.1038/nature08897.

153. M. Konishi. "Birdsong: From Behavior to Neuron". *Annual Review of Neuroscience* 8:1, 1985, pp. 125–170. DOI: 10.1146/annurev.ne.08.030185.001013. URL: https://doi.org/10.1146/annurev.ne.08.030185.001013.

154. G. Kosche, D. Vallentin, and M. A. Long. "Interplay of Inhibition and Excitation Shapes a Premotor Neural Sequence". *The Journal of Neuroscience* 35:3, 2015, pp. 1217–1227. DOI: 10.1523/jneurosci.4346-14.2015. URL: https://doi.org/10.1523/jneurosci.4346-14.2015.

155. A. A. Kozhevnikov and M. S. Fee. "Singing-related activity of identified HVC neurons in the zebra finch". *Journal of neurophysiology* 97:6, 2007, pp. 4271–4283.

156. Y. Kubo, E. Chalmers, and A. Luczak. "Biologically-inspired neuronal adaptation improves learning in neural networks". *Communicative and Integrative Biology* 16:1, 2023. DOI: 10.1080/19420889.2022.2163131. URL: https://doi.org/10.1080/19420889.2022.2163131.

157. P. K. Kuhl. "Early language acquisition: cracking the speech code". *Nature reviews neuroscience* 5:11, 2004, pp. 831–843.

158. P. K. Kuhl. "Is speech learning 'gated'by the social brain?" *Developmental science* 10:1, 2007, pp. 110–120.

159. J. Kunimatsu, T. W. Suzuki, S. Ohmae, and M. Tanaka. "Different contributions of preparatory activity in the basal ganglia and cerebellum for self-timing". *Elife* 7, 2018, e35676.

160. C. S. Lai, S. E. Fisher, J. A. Hurst, F. Vargha-Khadem, and A. P. Monaco. "A forkhead-domain gene is mutated in a severe speech and language disorder". *Nature* 413:6855, 2001, pp. 519–523.

161. R. Laje and D. V. Buonomano. "Robust timing and motor patterns by taming chaos in recurrent neural networks". *Nature Neuroscience* 16:7, 2013, pp. 925–933. DOI: 10.1038/nn.3405. URL: https://doi.org/10.1038/nn.3405.

162. R. P. Lanza, J. Starr, and B. Skinner. ""Lying" in the pigeon". *Journal of the Experimental Analysis of Behavior* 38:2, 1982, pp. 201–203.

163. L. Lapique. "Researches quantatives sur l'excitation electrique des nerfs traitee comme une polarization". *Journal of Physiology, Pathology and Genetics* 9, 1907, pp. 620–635.

164. A. Leblois and R. Darshan. "Basal Ganglia: Songbird Models". In: *Encyclopedia of Computational Neuroscience*. Springer New York, 2014, pp. 1–6. DOI: 10.1007/978-1-4614-7320-6_84-1. URL: https://doi.org/10.1007/978-1-4614-7320-6_84-1.

165. H. Lejeune. "Switching or gating? The attentional challenge in cognitive models of psychological time". *Behavioural Processes* 44:2, 1998, pp. 127–145. DOI: 10.1016/s0376-6357(98)00045-x. URL: https://doi.org/10.1016/s0376-6357(98)00045-x.

166. A. Leonardo. "Experimental test of the birdsong error-correction model". *Proceedings of the National Academy of Sciences* 101:48, 2004, pp. 16935–16940.

167. A. Leonardo and M. S. Fee. "Ensemble coding of vocal control in birdsong". *Journal of Neuroscience* 25:3, 2005, pp. 652–661.

168. M. S. Lewicki and M. Konishi. "Mechanisms underlying the sensitivity of songbird forebrain neurons to temporal order." *Proceedings of the National Academy of Sciences* 92:12, 1995, pp. 5582–5586. DOI: 10.1073/pnas.92.12.5582. URL: https://doi.org/10.1073/pnas.92.12.5582.

169. P. A. Lewis, R. Miall, S. Daan, and A. Kacelnik. "Interval timing in mice does not rely upon the circadian pacemaker". *Neuroscience letters* 348:3, 2003, pp. 131–134.

170. A. C. Lin, A. M. Bygrave, A. De Calignon, T. Lee, and G. Miesenböck. "Sparse, decorrelated odor coding in the mushroom body enhances learned odor discrimination". *Nature neuroscience* 17:4, 2014, pp. 559–568.

171. J. K. Liu and D. V. Buonomano. "Embedding Multiple Trajectories in Simulated Recurrent Neural Networks in a Self-Organizing Manner". *The Journal of Neuroscience* 29:42, 2009, pp. 13172–13181. DOI: 10.1523/jneurosci.2358-09.2009. URL: https://doi.org/10.1523/jneurosci.2358-09.2009.

172. C. J. Logan, S. A. Jelbert, A. J. Breen, R. D. Gray, and A. H. Taylor. "Modifications to the Aesop's Fable paradigm change New Caledonian crow performances". *PLoS One* 9:7, 2014, e103049.

173. M. A. Long and M. S. Fee. "Using temperature to analyse temporal dynamics in the songbird motor pathway". *Nature* 456:7219, 2008, pp. 189–194. DOI: 10.1038/nature07448. URL: https://doi.org/10.1038/nature07448.

174. M. A. Long, D. Z. Jin, and M. S. Fee. "Support for a synaptic chain model of neuronal sequence generation". *Nature* 468:7322, 2010, pp. 394–399. DOI: 10.1038/nature09514. URL: https://doi.org/10.1038/nature09514.

175. R. E. Lubow. "HIGH-ORDER CONCEPT FORMATION IN THE PIGEON 1". *Journal of the Experimental Analysis of Behavior* 21:3, 1974, pp. 475–483.

176. M. Luo and D. J. Perkel. "A GABAergic, strongly inhibitory projection to a thalamic nucleus in the zebra finch song system". *Journal of Neuroscience* 19:15, 1999, pp. 6700–6711.

177. G. Lynch, T. Okubo, A. Hanuschkin, R. Hahnloser, and M. Fee. "Rhythmic Continuous-Time Coding in the Songbird Analog of Vocal Motor Cortex." *Neuron*, 2016. DOI: 10.1016/j.neuron.2016.04.021.

178. C. Malapani, B. Rakitin, R. Levy, W. H. Meck, B. Deweer, B. Dubois, and J. Gibbon. "Coupled Temporal Memories in Parkinson's Disease: A Dopamine-Related Dysfunction". *Journal of Cognitive Neuroscience* 10:3, 1998, pp. 316–331. DOI: 10.1162/089892998562762. URL: https://doi.org/10.1162/089892998562762.

179. D. Margoliash. "Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow". *The Journal of Neuroscience* 3:5, 1983, pp. 1039–1057. DOI: 10.1523/jneurosci.03-05-01039.1983. URL: https://doi.org/10.1523/jneurosci.03-05-01039.1983.

180. D. Margoliash and E. Fortune. "Temporal and harmonic combination-sensitive neurons in the zebra finch's HVc". *The Journal of Neuroscience* 12:11, 1992, pp. 4309–4326. DOI: 10.1523/jneurosci.12-11-04309.1992. URL: https://doi.org/10.1523/jneurosci.12-11-04309.1992.

181. D. Margoliash. "Functional organization of forebrain pathways for song production and perception". *Journal of neurobiology* 33:5, 1997, pp. 671–693.

182. J. E. Markowitz, W. A. Liberti, G. Guitchounts, T. Velho, C. Lois, and T. J. Gardner. "Mesoscopic Patterns of Neural Activity Support Songbird Cortical Sequences". *PLOS Biology* 13:6, 2015. Ed. by J. Ashe, e1002158. DOI: 10.1371/journal.pbio.1002158. URL: https://doi.org/10.1371/journal.pbio.1002158.

183. P. Marler. "A comparative approach to vocal learning: Song development in white-crowned sparrows." *Journal of Comparative and Physiological Psychology* 71:2, Pt.2, 1970, pp. 1–25. DOI: 10.1037/h0029144. URL: https://doi.org/10.1037/h0029144.

184. P. Marler. "Bird Calls: Their Potential for Behavioral Neurobiology". *Annals of the New York Academy of Sciences* 1016:1, 2004, pp. 31–44. DOI: 10.1196/annals.1298.034. URL: https://doi.org/10.1196/annals.1298.034.

185. P. Marler. "Characteristics of some animal calls". *Nature* 176:4470, 1955, pp. 6–8.

186. P. Marler. "Three models of song learning: evidence from behavior". *Journal of neurobiology* 33:5, 1997, pp. 501–516.

187. P. Marler and S. Peters. "Selective vocal learning in a sparrow". *Science* 198:4316, 1977, pp. 519–521.

188. Y. Masamizu, Y. R. Tanaka, Y. H. Tanaka, R. Hira, F. Ohkubo, K. Kitamura, Y. Isomura, T. Okada, and M. Matsuzaki. "Two distinct layer-specific dynamics of cortical ensembles during learning of a motor task". *Nature Neuroscience* 17:7, 2014, pp. 987–994. DOI: 10.1038/nn.3739. URL: https://doi.org/10.1038/nn.3739.

189. M. Matell, W. Meck, and M. Nicolelis. "Integration of Behavior and Timing". In: *Functional and Neural Mechanisms of Interval Timing*. CRC Press, 2003. DOI: 10.1201/9780203009574.ch15. URL: https://doi.org/10.1201/9780203009574.ch15.

190. M. S. Matell and W. H. Meck. "Cortico-striatal circuits and interval timing: coincidence detection of oscillatory processes". *Cognitive Brain Research* 21:2, 2004, pp. 139–170. DOI: 10.1016/j.cogbrainres.2004.06.012. URL: https://doi.org/10.1016/j.cogbrainres.2004.06.012.

191. M. S. Matell and W. H. Meck. "Neuropsychological mechanisms of interval timing behavior". *BioEssays* 22:1, 2000, pp. 94–103. DOI: 10.1002/(sici)1521-1878(200001)22:1<94::aid-bies14>3.0.co;2-e. URL: https://doi.org/10.1002/(sici)1521-1878(200001)22:1%3C94::aid-bies14%3E3.0.co;2-e.

192. T. Matzinger, N. Ritt, and W. T. Fitch. "Non-native speaker pause patterns closely correspond to those of native speakers at different speech rates". *PLOS ONE* 15:4, 2020. Ed. by K. Miwa, e0230710. DOI: 10.1371/journal.pone.0230710. URL: https://doi.org/10.1371/journal.pone.0230710.

193. M. Mauk and D. Buonomano. "The Neural Basis of Temporal Processing." *Annual review of neuroscience* 27, 2004, pp. 307–40. DOI: 10.1146/annurev.neuro.27.070203.144247.

194. J. McCasland. "Neuronal control of bird song production". *The Journal of Neuroscience* 7:1, 1987, pp. 23–39. DOI: 10.1523/jneurosci.07-01-00023.1987. URL: https://doi.org/10.1523/jneurosci.07-01-00023.1987.

195. N. R. McFarland and S. N. Haber. "Organization of thalamostriatal terminals from the ventral motor nuclei in the macaque". *Journal of Comparative Neurology* 429:2, 2001, pp. 321–336.

196. R. McWalter and J. H. McDermott. "Illusory sound texture reveals multi-second statistical completion in auditory scene analysis". *Nature communications* 10:1, 2019, p. 5096.

197. W. H. Meck, T. B. Penney, and V. Pouthas. "Cortico-striatal representation of time in animals and humans". *Current opinion in neurobiology* 18:2, 2008, pp. 145–152.

198. W. H. Meck. "Selective adjustment of the speed of internal clock and memory processes." *Journal of Experimental Psychology: Animal Behavior Processes* 9:2, 1983, pp. 171–201. DOI: 10.1037/0097-7403.9.2.171. URL: https://doi.org/10.1037/0097-7403.9.2.171.

199. J. F. Medina and M. D. Mauk. "Computer simulation of cerebellar information processing". *Nature Neuroscience* 3:S11, 2000, pp. 1205–1211. DOI: 10.1038/81486. URL: https://doi.org/10.1038/81486.

200. H. Merchant, D. L. Harrington, and W. H. Meck. "Neural basis of the perception and estimation of time". *Annual review of neuroscience* 36, 2013, pp. 313–336.

201. D. S. Miall. "Beyond the schema given: Affective comprehension of literary narratives". *Cognition and Emotion* 3:1, 1989, pp. 55–78. DOI: 10.1080/02699938908415236. URL: https://doi.org/10.1080/02699938908415236.

202. J. W. Milnor. "Attractor". *Scholarpedia* 1:11, 2006, p. 1815.

203. A. Mita, H. Mushiake, K. Shima, Y. Matsuzaka, and J. Tanji. "Interval time coding by neurons in the presupplementary and supplementary motor areas". *Nature neuroscience* 12:4, 2009, pp. 502–507.

204. F. W. Moll, D. Kranz, A. C. Asensio, M. Elmaleh, L. A. Ackert-Smith, and M. A. Long. "Thalamus drives vocal onsets in the zebra finch courtship song". *Nature* 616:7955, 2023, pp. 132–136. DOI: 10.1038/s41586-023-05818-x. URL: https://doi.org/10.1038/s41586-023-05818-x.

205. T. Monteiro, F. S. Rodrigues, M. Pexirra, B. F. Cruz, A. I. Gonçalves, P. E. Rueda-Orozco, and J. J. Paton. "Using temperature to analyze the neural basis of a time-based decision". *Nature Neuroscience* 26:8, 2023, pp. 1407–1416. DOI: 10.1038/s41593-023-01378-5. URL: https://doi.org/10.1038/s41593-023-01378-5.

206. R. Mooney. "Different Subthreshold Mechanisms Underlie Song Selectivity in Identified HVc Neurons of the Zebra Finch". *The Journal of Neuroscience* 20:14, 2000, pp. 5420–5436. DOI: 10.1523/jneurosci.20-14-05420.2000. URL: https://doi.org/10.1523/jneurosci.20-14-05420.2000.

207. R. Mooney. "Neural mechanisms for learned birdsong". *Learning and Memory* 16:11, 2009, pp. 655–669. DOI: 10.1101/lm.1065209. URL: https://doi.org/10.1101/lm.1065209.

208. R. Mooney. "Neurobiology of song learning". *Current opinion in neurobiology* 19:6, 2009, pp. 654–660.

209. R. Mooney. "Synaptic Mechanisms for Auditory-Vocal Integration and the Correction of Vocal Errors". *Annals of the New York Academy of Sciences* 1016:1, 2004, pp. 476–494.

210. R. Mooney and J. F. Prather. "The HVC Microcircuit: The Synaptic Basis for Interactions between Song Motor and Vocal Plasticity Pathways". *The Journal of Neuroscience* 25:8, 2005, pp. 1952–1964. DOI: 10.1523/jneurosci.3726-04.2005. URL: https://doi.org/10.1523/jneurosci.3726-04.2005.

211. J. W. Moore, J. E. Desmond, and N. E. Berthier. "Adaptively timed conditioned responses and the cerebellum: A neural network approach". *Biological Cybernetics* 62:1, 1989, pp. 17–28. DOI: 10.1007/bf00217657. URL: https://doi.org/10.1007/bf00217657.

212. C. A. Munn. "Birds that 'cry wolf'". *Nature* 319:6049, 1986, pp. 143–145.

213. M. Murakami, M. I. Vicente, G. M. Costa, and Z. F. Mainen. "Neural antecedents of self-initiated actions in secondary motor cortex". *Nature neuroscience* 17:11, 2014, pp. 1574–1582.

214. K. Murphy, K. S. Lawley, P. Smith, and J. F. Prather. "New insights into the avian song system and neuronal control of learned vocalizations". *The neuroethology of birdsong*, 2020, pp. 65–92.

215. H. Mushiake, M. Inase, and J. Tanji. "Neuronal activity in the primate premotor, supplementary, and precentral motor cortex during visually guided and internally determined sequential movements". *Journal of neurophysiology* 66:3, 1991, pp. 705–718.

216. S. S. Nagarajan, D. T. Blake, B. A. Wright, N. Byl, and M. M. Merzenich. "Practice-Related Improvements in Somatosensory Interval Discrimination Are Temporally Specific But Generalize across Skin Location, Hemisphere, and Modality". *The Journal of Neuroscience* 18:4, 1998, pp. 1559–1570. DOI: 10.1523/jneurosci.18-04-01559.1998.

217. K. Naie and R. Hahnloser. "Regulation of learned vocal behavior by an auditory motor cortical nucleus in juvenile zebra finches." *J. Neurophysiol.*, 2011. DOI: 10.1152/jn.01035.2010.

218. D. Nicholson. *NickleDave/hybrid-vocal-classifier: annotated-high-level-finch*. 2018. DOI: 10.5281/ZENODO.1475481. URL: https://zenodo.org/record/1475481.

219. H. Niki and M. Watanabe. "Prefrontal and cingulate unit activity during timing behavior in the monkey". *Brain research* 171:2, 1979, pp. 213–224.

220. K. W. Nordeen and E. J. Nordeen. "Auditory feedback is necessary for the maintenance of stereotyped song in adult zebra finches". *Behavioral and Neural Biology* 57:1, 1992, pp. 58–66. DOI: 10.1016/0163-1047(92)90757-u. URL: https://doi.org/10.1016/0163-1047(92)90757-u.

221. F. Nottebohm. "The Origins of Vocal Learning". *The American Naturalist* 106:947, 1972, pp. 116–140. DOI: 10.1086/282756. URL: https://doi.org/10.1086/282756.

222. F. Nottebohm, T. M. Stokes, and C. M. Leonard. "Central control of song in the canary, Serinus canarius". *The Journal of Comparative Neurology* 165:4, 1976, pp. 457–486. DOI: 10.1002/cne.901650405. URL: https://doi.org/10.1002/cne.901650405.

223. T. Nowotny, M. I. Rabinovich, and H. D. Abarbanel. "Spatial representation of temporal information through spike-timing-dependent plasticity". *Physical Review E* 68:1, 2003, p. 011908.

224. R. E. Núñez and E. Sweetser. "With the future behind them: Convergent evidence from Aymara language and gesture in the crosslinguistic comparison of spatial construals of time". *Cognitive science* 30:3, 2006, pp. 401–450.

225. T. S. Okubo, E. L. Mackevicius, H. L. Payne, G. F. Lynch, and M. S. Fee. "Growth and splitting of neural sequences in songbird vocal development". *Nature* 528:7582, 2015, pp. 352–357. DOI: 10.1038/nature15741. URL: https://doi.org/10.1038/nature15741.

226. B. P. Ölveczky, A. S. Andalman, and M. S. Fee. "Vocal Experimentation in the Juvenile Songbird Requires a Basal Ganglia Circuit". *PLoS Biology* 3:5, 2005. Ed. by W. Schultz, e153. DOI: 10.1371/journal.pbio.0030153. URL: https://doi.org/10.1371/journal.pbio.0030153.

227. G. Palkar, J.-y. Wu, and B. Ermentrout. "The inhibitory control of traveling waves in cortical networks". *PLOS Computational Biology* 19:9, 2023, e1010697.

228. R. Pang, A. Duffy, D. Bell, Z. Torok, and A. Fairhall. "Precision motor timing via scalar input fluctuations", 2022. DOI: 10.1101/2022.05.18.492498. URL: https://doi.org/10.1101/2022.05.18.492498.

229. V. Pariyadath and D. Eagleman. "The effect of predictability on subjective duration". *PloS one* 2:11, 2007, e1264.

230. J. Park, L. T. Coddington, and J. T. Dudman. "Basal ganglia circuits for action specification". *Annual review of neuroscience* 43, 2020, pp. 485–507.

231. M. Pastor, J. Artieda, M. Jahanshahi, and J. Obeso. "Time estimation and reproduction is abnormal in Parkinson's disease". *Brain* 115:1, 1992, pp. 211–225.

232. J. J. Paton and D. V. Buonomano. "The Neural Basis of Timing: Distributed Mechanisms for Diverse Functions". *Neuron* 98:4, 2018, pp. 687–705. DOI: 10.1016/j.neuron.2018.03.045. URL: https://doi.org/10.1016/j.neuron.2018.03.045.

233. J. J. Pattadkal, G. Mato, C. van Vreeswijk, N. J. Priebe, and D. Hansel. "Emergent orientation selectivity from random networks in mouse visual cortex". *Cell reports* 24:8, 2018, pp. 2042–2050.

234. C. Pehlevan, F. Ali, and B. Ölveczky. "Flexibility in motor timing constrains the topology and dynamics of pattern generator circuits." *Nat Commun 9, 977*, 2018. DOI: 10.1038/s41467-018-03261-5.

235. R. Perin, T. K. Berger, and H. Markram. "A synaptic organizing principle for cortical neuronal groups". *PNAS*, 2011. DOI: 10.1073/pnas.1016051108.

236. A. J. Peters, S. X. Chen, and T. Komiyama. "Emergence of reproducible spatiotemporal activity during motor learning". *Nature* 510:7504, 2014, pp. 263–267. DOI: 10.1038/nature13235.

237. A. R. Pfenning et al. "Convergent transcriptional specializations in the brains of humans and song-learning birds". *Science* 346:6215, 2014. DOI: 10.1126/science.1256846. URL: https://doi.org/10.1126/science.1256846.

238. B. Pollok, H. Prior, and O. Güntürkün. "Development of object permanence in food-storing magpies (Pica pica)." *Journal of Comparative Psychology* 114:2, 2000, pp. 148–157. DOI: 10.1037/0735-7036.114.2.148.

239. J. Prather, S. Peters, S. Nowicki, and R. Mooney. "Precise auditory-vocal mirroring in neurons for learned vocal communication." *Nature*, 2008. DOI: 10.1038/nature06492.

240. J. F. Prather, S. Peters, S. Nowicki, and R. Mooney. "Precise auditory–vocal mirroring in neurons for learned vocal communication". *Nature* 451:7176, 2008, pp. 305–310.

241. V. Prevosto, W. Graf, and G. Ugolini. "Cerebellar inputs to intraparietal cortex areas LIP and MIP: functional frameworks for adaptive control of eye movements, reaching, and arm/eye/head movement coordination". *Cerebral Cortex* 20:1, 2010, pp. 214–228.

242. H. Prior, A. Schwarz, and O. Güntürkün. "Mirror-induced behavior in the magpie (Pica pica): evidence of self-recognition". *PLoS biology* 6:8, 2008, e202.

243. Y. Prut, E. Vaadia, H. Bergman, I. Haalman, H. Slovin, and M. Abeles. "Spatiotemporal Structure of Cortical Activity: Properties and Behavioral Relevance". *Journal of Neurophysiology* 79:6, 1998, pp. 2857–2874. DOI: 10.1152/jn.1998.79.6.2857. URL: https://doi.org/10.1152/jn.1998.79.6.2857.

244. C. R. Raby, D. M. Alexis, A. Dickinson, and N. S. Clayton. "Planning for the future by western scrub-jays". *Nature* 445:7130, 2007, pp. 919–921.

245. K. Rajan, L. Abbott, and H. Sompolinsky. "Stimulus-dependent suppression of chaos in recurrent neural networks". *Physical review e* 82:1, 2010, p. 011903.

246. T. H. Rammsayer. "Neuropharmacological evidence for different timing mechanisms in humans". *The Quarterly Journal of Experimental Psychology: Section B* 52:3, 1999, pp. 273–286.

247. P. Reddy, W. Zehring, D. Wheeler, V. Pirrotta, C. Hadfield, J. Hall, and M. Rosbash. "Molecular analysis of the period locus in Drosophila melanogaster and identification of a transcript involved in biological rhythms." *Cell.*, 1984. DOI: 10.1016/0092-8674(84)90265-4.

248. A. Reiner et al. "Revised nomenclature for avian telencephalon and some related brainstem nuclei". *The Journal of Comparative Neurology* 473:3, 2004, pp. 377–414. DOI: 10.1002/cne.20118.

249. J. Reutimann, V. Yakovlev, S. Fusi, and W. Senn. "Climbing Neuronal Activity as an Event-Based Cortical Representation of Time". *The Journal of Neuroscience* 24:13, 2004, pp. 3295–3303. DOI: 10.1523/jneurosci.4098-03.2004. URL: https://doi.org/10.1523/jneurosci.4098-03.2004.

250. J. L. Riquelme, M. Hemberger, G. Laurent, and J. Gjorgjieva. "Single spikes drive sequential propagation and routing of activity in a cortical network". *Elife* 12, 2023, e79928.

251. T. F. Roberts, E. Hisey, M. Tanaka, M. G. Kearney, G. Chattree, C. F. Yang, N. M. Shah, and R. Mooney. "Identification of a motor-to-auditory pathway important for vocal learning". *Nature Neuroscience* 20:7, 2017, pp. 978–986. DOI: 10.1038/nn.4563. URL: https://doi.org/10.1038/nn.4563.

252. C. Rovelli. *The order of time*. Penguin, 2019.

253. J. T. Sakata and M. S. Brainard. "Online contributions of auditory feedback to neural activity in avian song control circuitry". *Journal of Neuroscience* 28:44, 2008, pp. 11378–11390.

254. J. T. Sakata and Y. Yazaki-Sugiyama. "Neural circuits underlying vocal learning in songbirds". *The neuroethology of birdsong*, 2020, pp. 29–63.

255. C. Scharff and F. Nottebohm. "A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: implications for vocal learning". *The Journal of Neuroscience* 11:9, 1991, pp. 2896–2913. DOI: 10.1523/jneurosci.11-09-02896.1991. URL: https://doi.org/10.1523/jneurosci.11-09-02896.1991.

256. C. Scharff, J. R. Kirn, M. Grossman, J. D. Macklis, and F. Nottebohm. "Targeted Neuronal Death Affects Neuronal Replacement and Vocal Behavior in Adult Songbirds". *Neuron* 25:2, 2000, pp. 481–492. DOI: 10.1016/s0896-6273(00)80910-1. URL: https://doi.org/10.1016/s0896-6273(00)80910-1.

257. M. F. Schmidt. "Pattern of Interhemispheric Synchronization in HVc During Singing Correlates With Key Transitions in the Song Pattern". *Journal of Neurophysiology* 90:6, 2003, pp. 3931–3949. DOI: 10.1152/jn.00003.2003. URL: https://doi.org/10.1152/jn.00003.2003.

258. M. F. Schmidt, J. McLean, and F. Goller. "Breathing and vocal control: the respiratory system as both a driver and a target of telencephalic vocal motor circuits in songbirds". *Experimental Physiology* 97:4, 2011, pp. 455–461. DOI: 10.1113/expphysiol.2011.058669. URL: https://doi.org/10.1113/expphysiol.2011.058669.

259. W. A. Searcy and M. Andersson. "Sexual selection and the evolution of song". *Annual Review of Ecology and Systematics* 17:1, 1986, pp. 507–533.

260. S. Sharma, S. Chandra, and I. Fiete. "Content addressable memory without catastrophic forgetting by heteroassociation with a fixed scaffold". In: *International Conference on Machine Learning*. PMLR. 2022, pp. 19658–19682.

261. K. V. Shenoy, M. Sahani, and M. M. Churchland. "Cortical control of arm movements: a dynamical systems perspective". *Annual review of neuroscience* 36, 2013, pp. 337–359.

262. P. Simen, F. Balci, L. Souza, J. Cohen, and P. Holmes. "A Model of Interval Timing by Neural Integration." *The Journal of neuroscience : the official journal of the Society for Neuroscience* 31, 2011, pp. 9238–53. DOI: 10.1523/JNEUROSCI.3121-10.2011.

263. H. Simpson and D. Vicario. "Brain pathways for learned and unlearned vocalizations differ in zebra finches". *The Journal of Neuroscience* 10:5, 1990, pp. 1541–1556. DOI: 10.1523/jneurosci.10-05-01541.1990. URL: https://doi.org/10.1523/jneurosci.10-05-01541.1990.

264. M. Skocik and A. Kozhevnikov. "Real-time system for studies of the effects of acoustic feedback on animal vocalizations". *Frontiers in Neural Circuits* 6, 2013, p. 111.

265. A. Smirnova, Z. Zorina, T. Obozova, and E. Wasserman. "Crows spontaneously exhibit analogical reasoning". *Current Biology* 25:2, 2015, pp. 256–260.

266. F. Sohrabji, E. J. Nordeen, and K. W. Nordeen. "Selective impairment of song learning following lesions of a forebrain nucleus in the juvenile zebra finch". *Behavioral and Neural Biology* 53:1, 1990, pp. 51–63. DOI: 10.1016/0163-1047(90)90797-a. URL: https://doi.org/10.1016/0163-1047(90)90797-a.

267. H. Sompolinsky, A. Crisanti, and H.-J. Sommers. "Chaos in random neural networks". *Physical review letters* 61:3, 1988, p. 259.

268. R. M. C. Spencer and R. B. Ivry. "Cerebellum and Timing". In: *Handbook of the Cerebellum and Cerebellar Disorders*. Springer Netherlands, 2013, pp. 1201–1219. DOI: 10.1007/978-94-007-1333-8_52. URL: https://doi.org/10.1007/978-94-007-1333-8_52.

269. J. E. R. Staddon and J. J. Higa. "TIME AND MEMORY: TOWARDS A PACEMAKER-FREE THEORY OF INTERVAL TIMING". *Journal of the Experimental Analysis of Behavior* 71:2, 1999, pp. 215–251. DOI: 10.1901/jeab.1999.71-215. URL: https://doi.org/10.1901/jeab.1999.71-215.

270. K. Stevens. "Scientific substrates of speech production". *Introduction to communication sciences and disorders*, 1994, pp. 399–437.

271. S. Stringer, T. Trappenberg, E. Rolls, and I. Araujo. "Self-organizing continuous attractor networks and path integration: one-dimensional models of head direction cells". *Network: Computation in Neural Systems* 13:2, 2002, pp. 217–242.

272. S. H. Strogatz. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. Westview Press., 1994.

273. D. Sussillo and L. F. Abbott. "Generating coherent patterns of activity from chaotic neural networks". *Neuron* 63:4, 2009, pp. 544–557.

274. M. Tanaka, F. Sun, Y. Li, and R. Mooney. "A mesocortical dopamine circuit enables the cultural transmission of vocal behaviour". *Nature* 563:7729, 2018, pp. 117–120.

275. J. Tanji. "Sequential organization of multiple movements: involvement of cortical motor areas". *Annual review of neuroscience* 24:1, 2001, pp. 631–651.

276. J. S. Taube, R. U. Muller, and J. B. Ranck. "Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis". *Journal of Neuroscience* 10:2, 1990, pp. 420–435.

277. S. Teki, M. Grube, S. Kumar, and T. D. Griffiths. "Distinct Neural Substrates of Duration-Based and Beat-Based Auditory Timing". *The Journal of Neuroscience* 31:10, 2011, pp. 3805–3812. DOI: 10.1523/jneurosci.5561-10.2011. URL: https://doi.org/10.1523/jneurosci.5561-10.2011.

278. E. C. Traugott. "On the expression of spatio-temporal relations in language". *Universals of Human Language III*, 1978.

279. M. Treisman. “Temporal discrimination and the indifference interval: Implications for a model of the "internal clock".” *Psychological Monographs: General and Applied, 77(13), 1–31.*, 1963. DOI: 10.1037/h0093864.

280. S. G. Trettel, J. B. Trimper, E. Hwaun, I. R. Fiete, and L. L. Colgin. “Grid cell co-activity patterns during sleep reflect spatial overlap of grid fields during active behaviors”. *Nature neuroscience* 22:4, 2019, pp. 609–617.

281. T. W. Troyer and A. J. Doupe. “An Associational Model of Birdsong Sensorimotor Learning I. Efference Copy and the Learning of Song Syllables”. *Journal of Neurophysiology* 84:3, 2000, pp. 1204–1223. DOI: 10.1152/jn.2000.84.3.1204. URL: https://doi.org/10.1152/jn.2000.84.3.1204.

282. E. C. Tumer and M. S. Brainard. “Performance variability enables adaptive plasticity of ‘crystallized’adult birdsong”. *Nature* 450:7173, 2007, pp. 1240–1244.

283. R. Ursu. “Role of the cerebellum in avian song learning”. *Thesis*, 2023.

284. D. Vallentin, G. Kosche, D. Lipkind, and M. A. Long. “Inhibition protects acquired song segments during vocal learning in zebra finches”. *Science* 351:6270, 2016, pp. 267–271. DOI: 10.1126/science.aad3023. URL: https://doi.org/10.1126/science.aad3023.

285. D. Vallentin and M. A. Long. “Motor Origin of Precise Synaptic Inputs onto Forebrain Neurons Driving a Skilled Behavior”. *The Journal of Neuroscience* 35:1, 2015, pp. 299–307. DOI: 10.1523/jneurosci.3698-14.2015. URL: https://doi.org/10.1523/jneurosci.3698-14.2015.

286. L. Veit and A. Nieder. “Abstract rule neurons in the endbrain support intelligent behaviour in corvid songbirds”. *Nature communications* 4:1, 2013, p. 2878.

287. E. Vu, M. Mazurek, and Y. Kuo. “Identification of a forebrain motor programming network for the learned song of zebra finches.” *Journal of Neuroscience*, 1994. DOI: 10.1523/JNEUROSCI.14-11-06924.1994.

288. M. Wang, D. Arteaga, and B. J. He. “Brain mechanisms for simple perception and bistable perception”. *Proceedings of the National Academy of Sciences* 110:35, 2013, E3350–E3359.

289. N. Wang, P. Hurley, C. Pytte, and J. R. Kirn. “Vocal control neuron incorporation decreases with age in the adult zebra finch”. *Journal of Neuroscience* 22:24, 2002, pp. 10864–10870.

290. X.-J. Wang. “Decision making in recurrent neuronal circuits”. *Neuron* 60:2, 2008, pp. 215–234.

291. T. L. Warren, E. C. Tumer, J. D. Charlesworth, and M. S. Brainard. “Mechanisms and time course of vocal learning and consolidation in the adult songbird”. *Journal of neurophysiology* 106:4, 2011, pp. 1806–1821.

292. S. Watanabe, J. Sakamoto, and M. Wakita. "PIGEONS'DISCRIMINATION OF PAINTINGS BY MONET AND PICASSO". *Journal of the experimental analysis of behavior* 63:2, 1995, pp. 165–174.

293. J. H. Wearden and I. S. Penton-Voak. "Feeling the heat: body temperature and the rate of subjective time, revisited." *he Quarterly journal of experimental psychology. B, Comparative and physiological psychology* 48:2, 1995, pp. 129–141. URL: https://doi.org/10.1080/14640749508401443.

294. A. A. Weir, J. Chappell, and A. Kacelnik. "Shaping of hooks in New Caledonian crows". *Science* 297:5583, 2002, pp. 981–981.

295. M. W. Westneat, J. H. Long Jr, W. Hoese, and S. Nowicki. "Kinematics of birdsong: functional correlation of cranial movements and acoustic features in sparrows". *Journal of Experimental Biology* 182:1, 1993, pp. 147–171.

296. H. Williams and D. S. Vicario. "Temporal patterning of song production: Participation of nucleus uvaeformis of the thalamus". *Journal of Neurobiology* 24:7, 1993, pp. 903–912. DOI: 10.1002/neu.480240704. URL: https://doi.org/10.1002/neu.480240704.

297. R. Williams. "Simple statistical gradient-following algorithms for connectionist reinforcement learning." *Mach Learn*, 1992.

298. H. R. Wilson and J. D. Cowan. "A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue". *Kybernetik* 13:2, 1973, pp. 55–80.

299. S. M. Woolley and C. V. Portfors. "Conserved mechanisms of vocalization coding in mammalian and songbird auditory midbrain". *Hearing research* 305, 2013, pp. 45–56.

300. S. Yagishita, A. Hayashi-Takagi, G. C. Ellis-Davies, H. Urakubo, S. Ishii, and H. Kasai. "A critical time window for dopamine actions on the structural plasticity of dendritic spines". *Science* 345:6204, 2014, pp. 1616–1620. DOI: 10.1126/science.1255514. URL: https://doi.org/10.1126/science.1255514.

301. Y. Yamashita, M. Takahasi, T. Okumura, M. Ikebuchi, H. Yamada, M. Suzuki, K. Okanoya, and J. Tani. "Developmental learning of complex syntactical song in the Bengalese finch: A neural network model". *Neural Networks* 21:9, 2008, pp. 1224–1231. DOI: 10.1016/j.neunet.2008.03.003. URL: https://doi.org/10.1016/j.neunet.2008.03.003.

302. H. N. Zelaznik, R. M. C. Spencer, and R. B. Ivry. "Dissociation of explicit and implicit timing in repetitive tapping and drawing movements." *Journal of Experimental Psychology: Human Perception and Performance* 28:3, 2002, pp. 575–588. DOI: 10.1037/0096-1523.28.3.575. URL: https://doi.org/10.1037/0096-1523.28.3.575.

303. K. Zhang. "Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: A theory." *Journal of Neuroscience*, 1996.

304. Y. S. Zhang, J. D. Wittenbach, D. Z. Jin, and A. A. Kozhevnikov. "Temperature Manipulation in Songbird Brain Implicates the Premotor Nucleus HVC in Birdsong Syntax". *The Journal of Neuroscience* 37:10, 2017, pp. 2600–2611. DOI: 10.1523/jneurosci.1827-16.2017. URL: https://doi.org/10.1523/jneurosci.1827-16.2017.

305. X. Zhou, É. de Villers-Sidani, R. Panizzutti, and M. M. Merzenich. "Successive-signal biasing for a learned sound sequence". *Proceedings of the National Academy of Sciences* 107:33, 2010, pp. 14839–14844. DOI: 10.1073/pnas.1009433107. URL: https://doi.org/10.1073/pnas.1009433107.